

Is the gender pay gap closing?

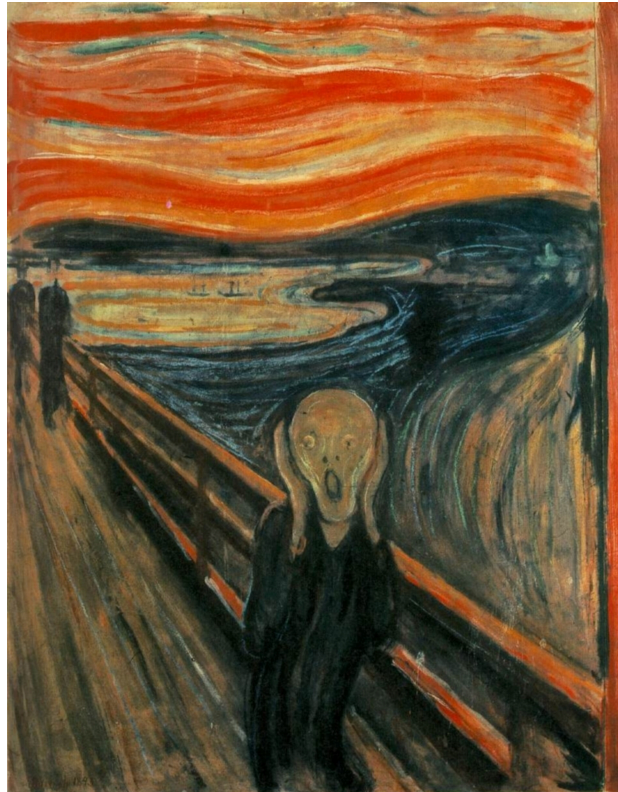
Week 6, Lecture 12

<https://live.ds306.org>

October 8, 2026

Midterm next Thursday

Any questions? 😊



Munch scream painting

From Tuesday

- We used earnings data to forecast the payoff from a master's degree.
- An indicator let us fit a separate line for each degree.
- The interaction coefficient tested whether the earnings gap was growing.

Today's questions

1. Is the gender pay gap closing?
2. Where is the gender pay gap largest?

We'll investigate these questions with the table and plotting skills we've been practicing.

The gender pay gap

- The **gender pay gap** is the difference in typical earnings between women and men. We will measure typical earnings using the median.
- Women earn less than men in many countries around the world, although the size of the gap varies. [International Labour Organization](#)
- The gap has persisted for decades. In the U.S., it has narrowed since the 1980s, with relatively little change over the past two decades. [Pew Research Center](#)
- Differences in occupations, experience, family responsibilities, and discrimination can all contribute to the gap.

Q1: Is the gender pay gap closing?

Your prediction

In 1991, women with bachelor's degrees earned about **62%** of what men with bachelor's degrees earned.

Here we are measuring median annual earnings.

What do you think that percentage was in 2024?

Enter a number between 0 and 100.

The historical earnings table

```
history |> filter(year == 1991) |> head(4)
```

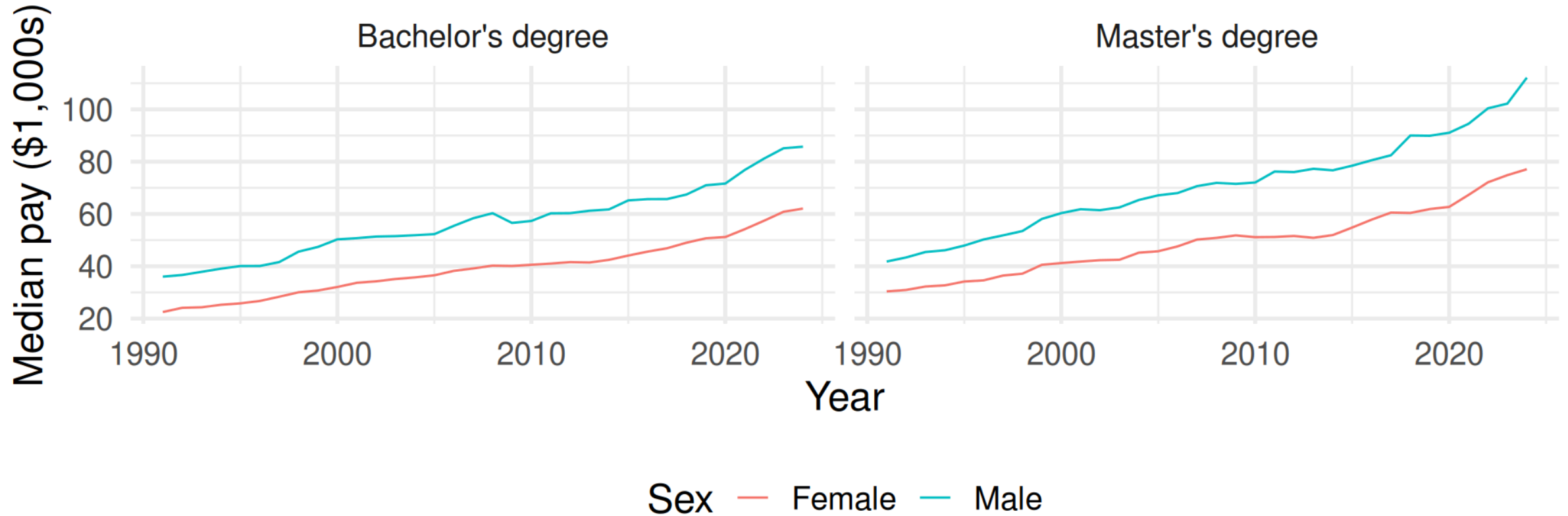
- Workers aged 25+ across occupations, from 1991 through 2024.
- Each row contains median annual earnings for one year, sex, and education level.
- pay is in thousands of that year's dollars.
- Source: [Census Table P-20](#)

Keep the college graduates

```
college_history <- history |>
  filter(education %in% c("Bachelor's degree", "Master's degree"))
college_history |> head(4)
```

The two groups are bachelor's and master's degree holders.

The first plot we want



Plot earnings over time

```
college_history |>
  ggplot(aes(_____, _____, color = _____)) +
  _____ +
  facet_wrap(~ _____) +
  labs(x = "Year", y = "Median pay ($1,000s)", color = "S
```

- Plot median earnings against year, with a different color for each sex.
- Connect each group's earnings over time with a line.
- Make separate panels for bachelor's and master's degree

Measure the gap relatively

Pay has risen for both groups. A dollar gap of \$10,000 means something different when earnings are \$30,000 than when they are \$100,000.

We will measure women's median earnings as a fraction of men's.

Suppose median earnings double for both groups. What happens to the dollar gap? What happens to women's earnings as a fraction of men's earnings? Which would you plot across these decades?

The two numbers we need

```
college_history |>  
  filter(year == 1991, education == "Bachelor's degree")
```

We need both earnings amounts on the same row to calculate Female / Male.

The table we want

```
# A tibble: 2 × 4
  year education      Female      Male
<dbl> <chr>      <dbl> <dbl>
1  1991 Bachelor's degree    22.5    36.0
2  2024 Bachelor's degree    62.0    85.7
```

How would you turn the previous table into this one? Describe the steps or name the command you would use.

Reshape the earnings

Open the current fold and fill in the two column names.

```
history_work <- history |>
  filter(education %in% c("Bachelor's degree", "Master's
    degree"))

  pivot_wider(names_from = _____, values_from = _____)

history_work |> head(3)
```

- Keep bachelor's and master's earnings.
- Put male and female pay on the same row for each year and education level.
- Calculate women's pay as a fraction of men's pay.

Calculate the relative earnings

A ratio of 0.80 means women's median earnings are 80% of men's.

```
history_work <- college_history |>
  pivot_wider(names_from = sex, values_from = pay) |>

  mutate(ratio = _____)

history_work |> select(year, education, ratio) |> head(3)
```

- Keep bachelor's and master's earnings.
- Put male and female pay on the same row for each year and education level.
- Calculate women's pay as a fraction of men's pay.

Plot the gap over time

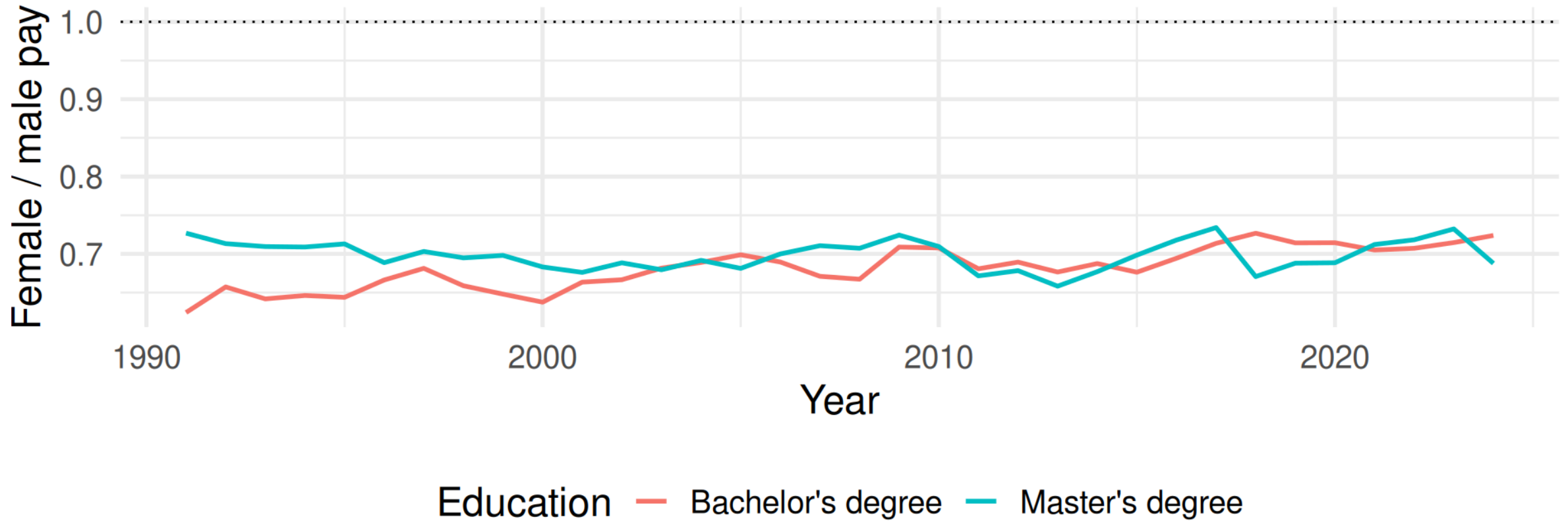
The completed `history_gap` table is available.

```
history_gap |>
  ggplot(aes(x = year, y = _____, color = _____)) +
  geom_line() +

  geom_hline(yintercept = 1, linetype = "dotted") +
  labs(x = "Year", y = "Female / male pay",
       color = "Education")
```

Plot the female/male earnings ratio over time for both education levels. Submit your plot. State whether the gap has closed by 2024.

What happened from 1991 to 2024?



Interpret the change

```
history_gap |>  
  filter(year %in% c(1991, 2024)) |>  
  select(year, education, ratio)
```

How did the gap change for each education level? Did more education correspond to a smaller gap in 2024? Use the ratios in your answer.

When would earnings reach parity?

ratio = 1 means equal median earnings.

Could we use a straight line to estimate when that happens?

Choose an education group. What would you regress on what? How would you use the fitted line to estimate the year of equal earnings?

Fit the bachelor's trend

```
bachelor_history <- history_gap |>
  filter(education == "Bachelor's degree") |>
  mutate(time = year - 1991)

history_model <- lm(ratio ~ time, data = bachelor_history)
coef(history_model)
```

Each unit of time is one year after 1991.

Solve for the year

The fitted line is approximately:

$$\text{earnings ratio} = 0.6430 + 0.002339 \times (\text{year} - 1991).$$

Set the earnings ratio to 1 and solve for the year. How far beyond our last observation is that?

Ask R for the same answer

```
intercept <- coef(history_model)[1]  
slope <- coef(history_model)[2]  
1991 + (1 - intercept) / slope
```

The fitted trend reaches 1 around **2144**.

Does the chosen period matter?

```
recent_history_model <- lm(ratio ~ time,  
  data = bachelor_history |> filter(year >= 2000))  
coef(recent_history_model)
```

This fit uses 2000–2024. Use its coefficients to get another parity year. Would you report either year as a precise prediction? Explain.

Q1: What did we find?

- The bachelor's earnings gap narrowed from 1991 to 2024.
- In 2024, women's median earnings were about 72% of men's for bachelor's graduates and 69% for master's graduates.
- Our fitted bachelor's trend reaches equal median earnings around 2144.

What else would you want to know about the workers in these groups?

Q2: Where is the gender pay gap largest?

ACS data

```
acs |> select(sex, age, degree, pay) |> head(4)
```

- Each row describes one worker surveyed in 2024.
- These are full-time computer/math workers, ages 25–50, with a BA or MA.
- pay is annual wage income in thousands of 2024 dollars.

Start with the overall gap

What are the median earnings for men and women in this sample?

```
summary_work <- acs |>
  group_by(_____) |>
  summarize(n = n(), median_pay = median(_____), .groups = 'drop')
summary_work
```

- Group the workers by sex and report n and median_pay for each group.
- Report the difference in median earnings in dollars.

The overall result

acs_summary

The median earnings gap in this sample is about **\$20,000/year**.

Does the gap remain within the same job and degree?

Let's investigate software developers with bachelor's degrees.

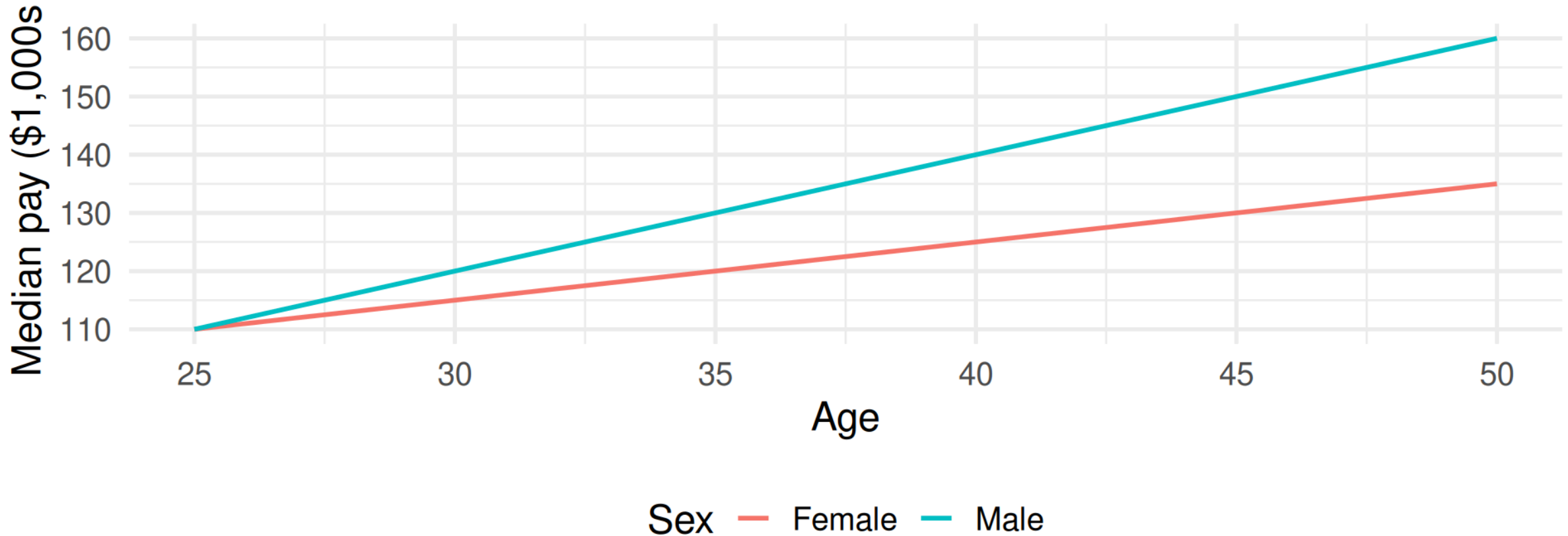
```
occupation_names |> filter(occupation == "Software developers")
```

occupation_code connects the worker records to these occupation names.

Could the gap differ by age?

After keeping one job and degree, do you expect a similar pay gap among workers in their twenties and workers in their forties? Explain. What would you plot to check it?

The age plot we want



We want median earnings at each age among software developers with bachelor's degrees.

How does the gap change with age?

```
software_work <- acs |>
  _____

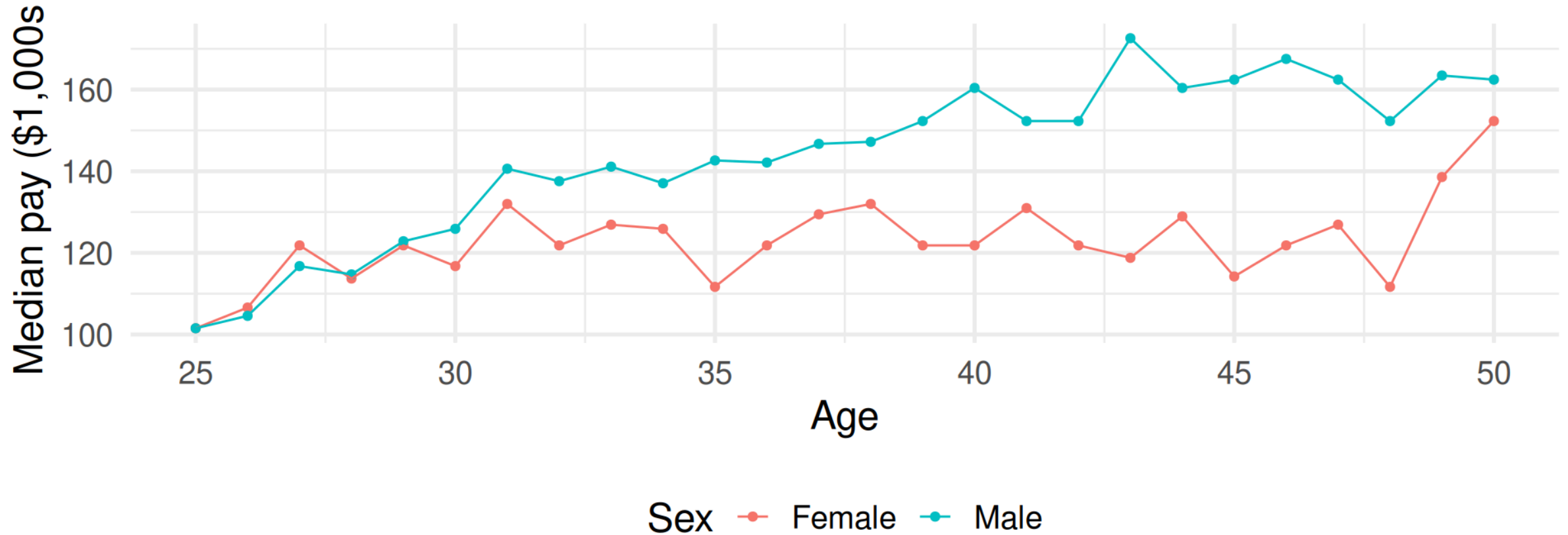
age_work <- software_work |>
  _____

age_work |>
  _____

# Compare pay gaps in the two age groups.
# Your answer:
```

- Join occupation_names to acs and keep software developers with bachelor's degrees.
- Calculate the number of workers and median pay for each sex and age. Call the columns n and median_pay.
- Plot median pay by

What do the age medians show?



Same job, same degree

same_job_medians

- At ages 25–29, both groups have median earnings of about **\$112k**.
- At ages 40–49, the medians are **\$122k for women** and **\$159k for men**.

What could explain this pattern?

Among software developers with bachelor's degrees, median earnings are similar for younger men and women, but farther apart for older workers.

What could be driving this pattern? Suggest an explanation and describe what additional data you would use to investigate it.

What if everyone works 40 hours?

Use `software_bachelors`, our completed one-job, one-degree table.
Investigate ages 25–29 and 40–49, keeping only 40-hour workers.

```
matched_hours_work <- software_bachelors |>
  _____

matched_gap_work <- matched_hours_work |>
  _____

matched_gap_work
```

- Keep 40-hour workers aged 25–29 or 40–49. Label the two groups `age_group` using `ifelse()`.
- Report `n` and `median_pay` for each sex and `age_group`.
- Put Male and

What changed when we held hours fixed?

same_job_hours_gap

How would you describe these results to a student about to graduate? Use the earnings gaps in both age groups.

Model mean pay by age

Our plots have summarized median pay. The linear regression we learned on Tuesday models **mean pay**.

We can fit one mean-pay line for men and one for women among these software developers with bachelor's degrees.

Use `software_bachelors` again, which includes all work schedules.

The interaction coefficient will tell us how their slopes differ.

How would you write one regression formula that fits both lines?
What would its response, age variable, and indicator be?

Set up the indicator

```
age_data <- software_bachelors |>
  mutate(age0 = age - 25,
         female = ifelse(sex == "Female", 1, 0))
age_data |> select(sex, age, age0, female) |> head(4)
```

- female = 0 selects men; female = 1 selects women.
- age0 = 0 means age 25.

Fit the two lines together

$$\text{pay} = \beta_0 + \beta_1 \text{age0} + \beta_2 \text{female} + \beta_3 \text{age0 female} + \epsilon.$$

```
age_work_model <- lm(pay ~ _____, data = age_data)
coef(age_work_model)
```

- Fit pay with separate intercepts and age slopes for men and women.
- Display the coefficients and identify the coefficient measuring the slope difference.

Read the coefficients

```
coef(age_model)
```

An analyst says: “The coefficient on female is nearly zero, so this model shows no earnings gap.”

Do you agree? Use predicted earnings at ages 25 and 45 in your answer.

The two fitted lines

$$\begin{aligned}\text{Men:} \quad & 135.748 + 2.922 \times \text{age0} \\ \text{Women:} \quad & 136.018 + 0.546 \times \text{age0}\end{aligned}$$

All amounts are in thousands of dollars.

Add the fitted values

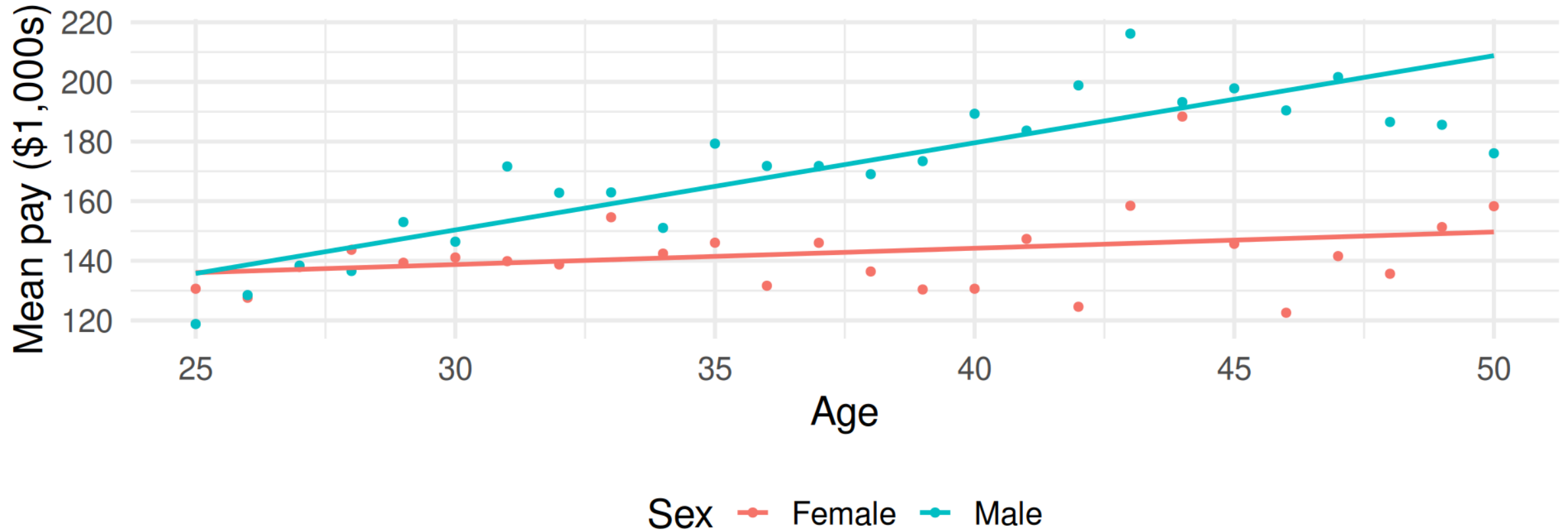
```
age_fitted <- software_bachelors |>
  group_by(sex, age) |>
  summarize(n = n(), mean_pay = mean(pay), .groups = "drop") |>
  mutate(age0 = age - 25, female = ifelse(sex == "Female", 1, 0)) |>

  modelr::add_predictions(age_model, var = "fit")

age_fitted |> select(sex, age, mean_pay, fit) |> head(3)
```

`add_predictions()` adds a `fit` column, as in Tuesday's forecast plot.

Mean earnings and fitted lines



Test whether the slopes differ

```
summary(age_model)$coefficients
```

Which row tests equal slopes? What does its estimate tell us about the age pattern? What does its p-value tell us?

Is the gap larger among the highest earners?

Return to all 27,028 workers in acs. Their median earnings gap is about \$20,000.

Do you expect a similar dollar gap among low earners and high earners? What would you look at beyond the median to check your prediction?

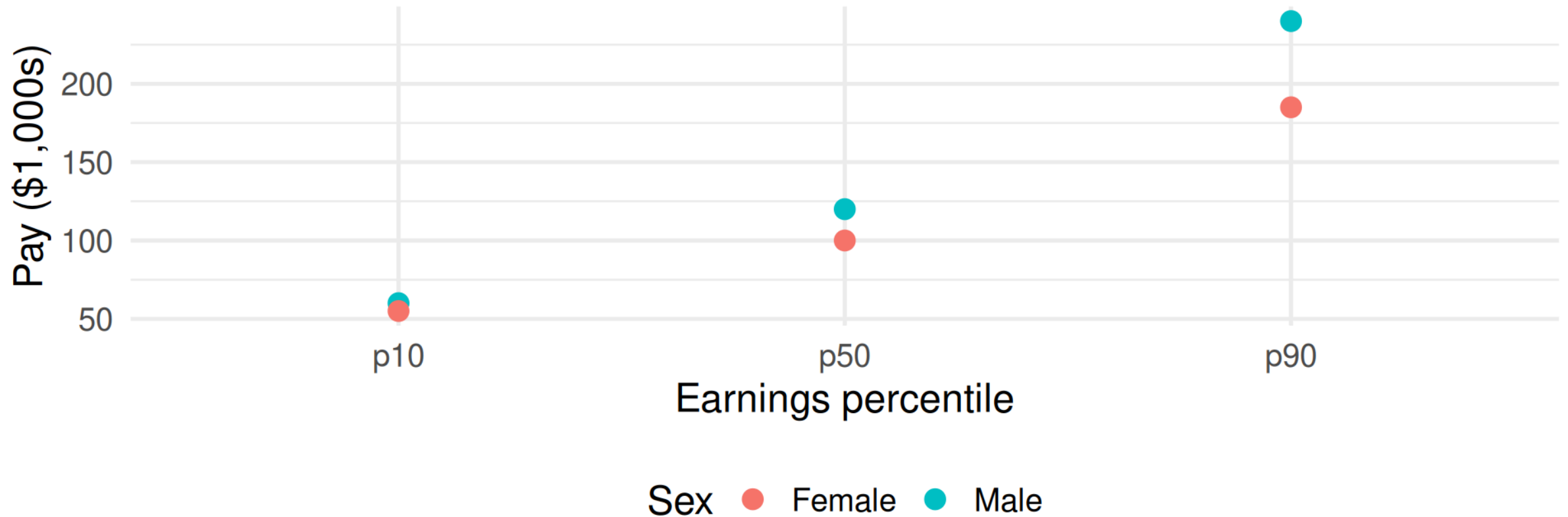
Look beyond the middle

- The 10th percentile describes earnings near the lower end.
- The 50th percentile is the median.
- The 90th percentile describes earnings near the upper end.

`quantile()` finds these values. `probs = 0.9` asks for the 90th percentile.

```
quantile(acs$pay, probs = c(0.1, 0.5, 0.9))
```


The earnings plot we want



Each percentile needs an earnings value for men and one for women.

Where is the dollar gap largest?

```
percentiles_work <- acs |>
  _____

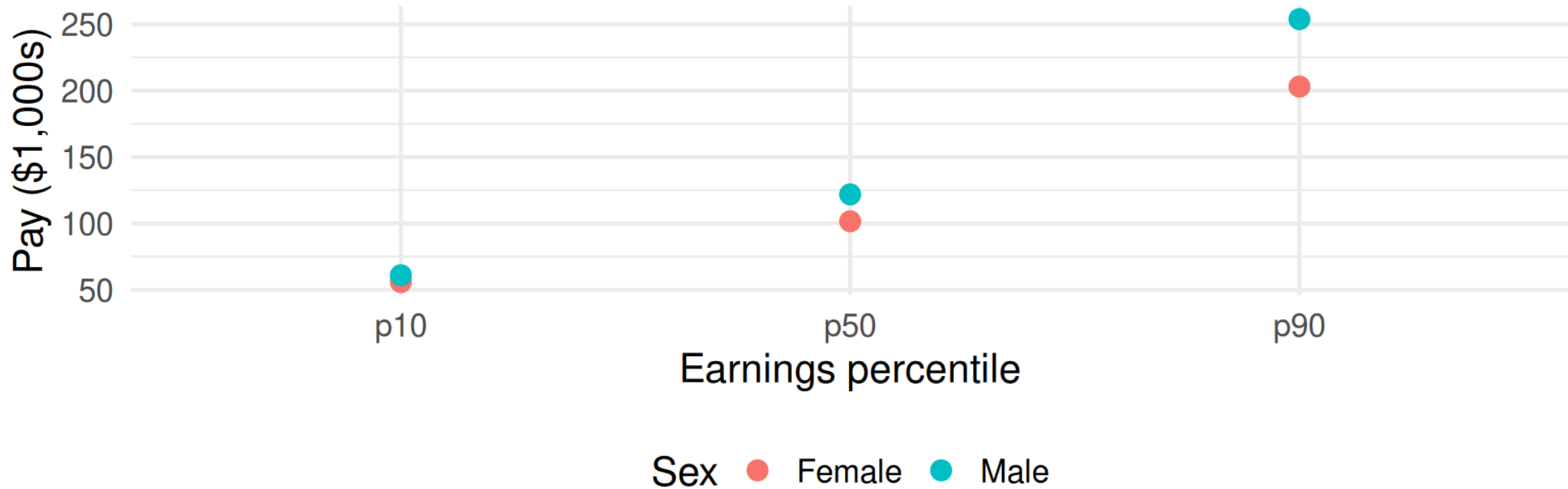
percentiles_long_work <- percentiles_work |>
  _____

percentiles_long_work |>
  _____

# Where is the dollar gap largest? Use values from your t
# Your answer:
```

- Calculate p10, p50, and p90 of pay separately for men and women.
- Reshape the three percentile columns into rows named percentile and pay.
- Plot earnings by percentile, distinguishing men and women. Label the pay units.

The gap across the earnings distribution



A report describes the earnings gap as \$20,000. Write a more informative one-sentence summary using this plot.

Additional investigations

The next examples are available if time remains.

What might explain the pattern?

We also know each worker's degree, usual hours, and whether they work in the public or private sector. Which would you investigate first, and what would you calculate or plot?

Could education explain it?

Do men and women in this sample have the same mix of bachelor's and master's degrees?

```
degree_work <- table(acs$sex, acs$degree)

prop.table(degree_work, margin = _____)
```

- Calculate the conditional degree distribution within each sex.
- Report the fraction with master's degrees for women and for men.

Education in our sample

```
prop.table(degree_table, margin = 1)
```

About **40% of women** and **32% of men** hold master's degrees.

Look within each degree

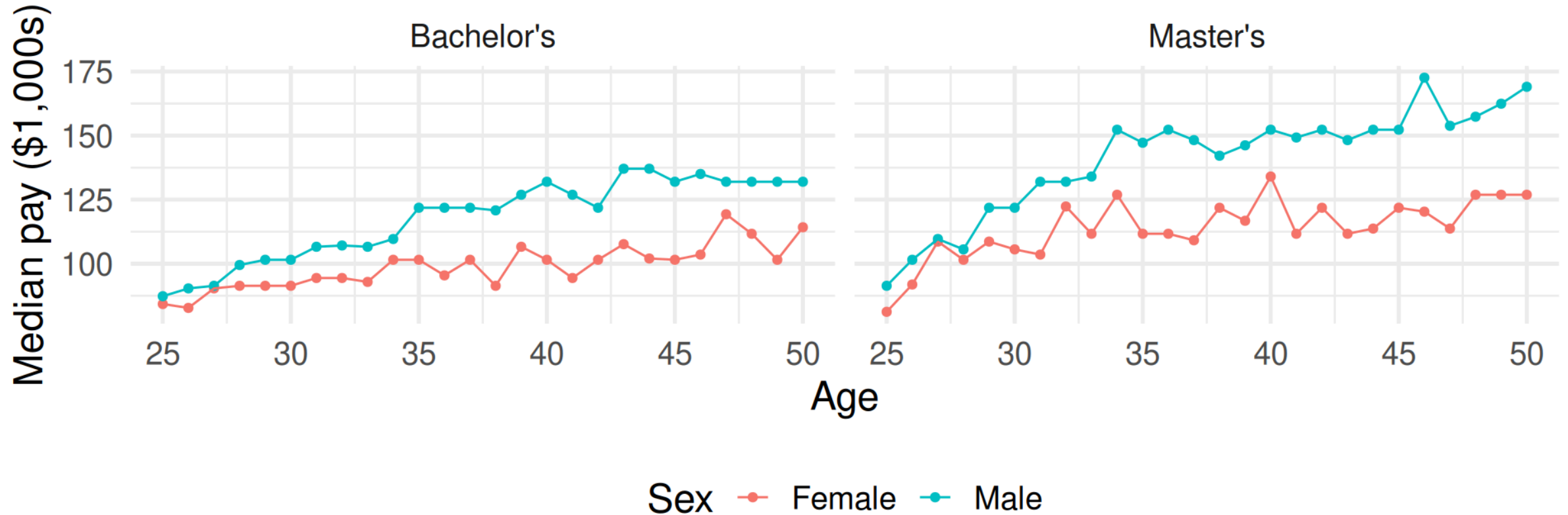
Use the full acs table to make age plots separately for bachelor's and master's graduates. What needs to change in the grouping and plot?

Add one grouping variable

```
degree_age_medians <- acs |>  
  group_by(degree, sex, age) |>  
  summarize(median_pay = median(pay), .groups = "drop")
```

```
degree_age_medians |>  
  ggplot(aes(age, median_pay, color = sex)) +  
  geom_point() + geom_line() + facet_wrap(~ degree) +  
  labs(x = "Age", y = "Median pay ($1,000s)", color = "Sex")
```

Does the pattern persist?



The gap is larger among older workers within both degree groups.

Could working hours explain it?

Annual pay might differ because people work different amounts.

hours records usual weekly hours; weeks records weeks worked.

An approximate hourly rate is:

$$\text{hourly pay} = \frac{1000 \times \text{pay}}{\text{hours} \times \text{weeks}}.$$

Calculate the hourly earnings

```
hours_work <- acs |>
  mutate(hourly_pay = _____) |>

  group_by(sex) |>
  summarize(median_hours = median(hours),
            median_hourly = median(hourly_pay), .groups =
hours_work
```

- Calculate hourly_pay from pay, hours, and weeks, with pay converted to dollars.
- Report median weekly hours and hourly earnings by sex.

What did hours account for?

```
hours_summary |> select(sex, median_hours, median_hourly)
```

Do these results support working hours as the main explanation for the annual earnings gap? Cite a number in your answer.

Does the workplace matter?

sector identifies private employment or public employment (local, state, or federal government).

We want median earnings by sex, separately for each sector.

What would you change in the overall summary to make this table?
What plot would you use to show the four medians?

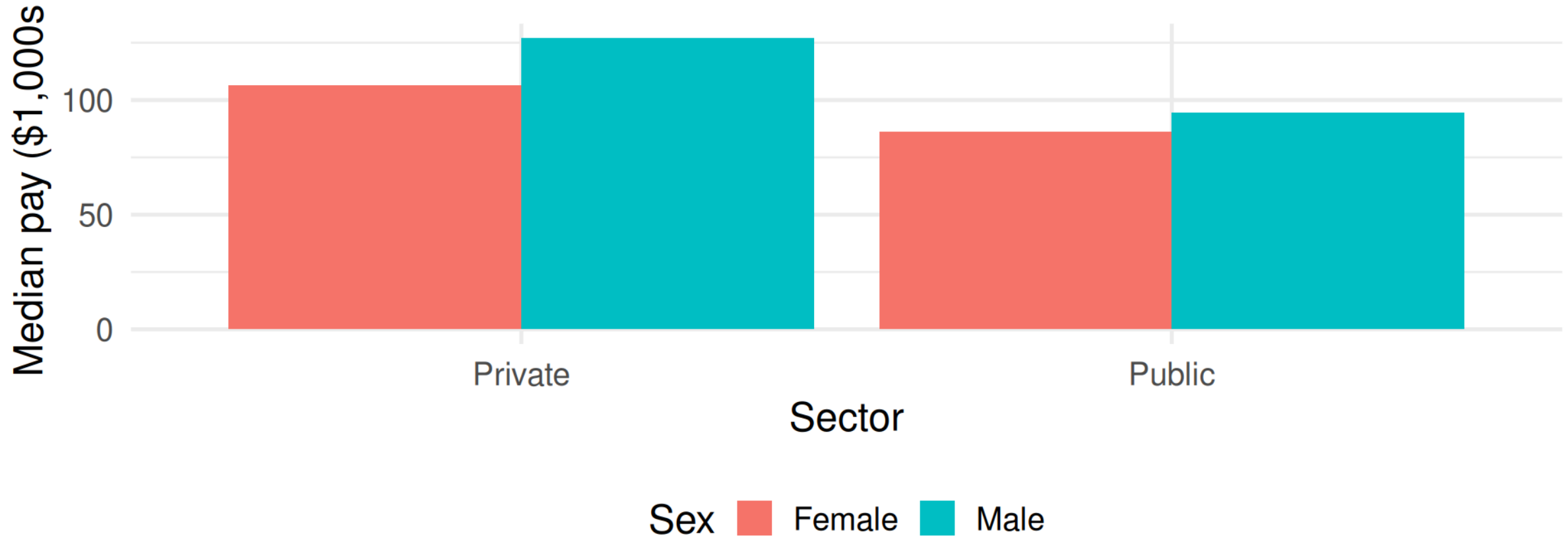
Submit the workplace plot

```
# Write the grouped summary and call it sector_work.
```

```
# Make a bar chart or a point plot.
```

Calculate median pay for men and women within each sector. Make and submit a plot showing all four medians. Which sector has the larger dollar gap?

Public and private employment



The gap is about **\$20,000** in private employment and **\$8,000** in public employment.

What would you investigate next?

Suggest one further analysis that would help explain the smaller public-sector gap. Use a variable we have and describe the table or plot you would make.

Optional: does the gap differ by occupation?

We have the occupation codes. The names are in a separate table.

```
occupation_names |> head(4)
```

Let's investigate software developers, systems analysts, operations research analysts, and computer support specialists.

Join, summarize, reshape, rank

```
occupation_work <- acs |>
  left_join(occupation_names, by = "occupation_code") |>
  filter(occupation %in% c("Software developers", "Computer
    "Operations research analysts", "Computer support spe

# Continue the pipeline with a grouped summary.

# Reshape, calculate gap = Male - Female, and arrange the
```

- Calculate median_pay for each occupation and sex.
- Put Male and Female median_pay on the same row for each occupation.
- Calculate gap as Male minus Female and sort from largest to smallest.

The occupation results

```
occupation_gap |> select(occupation, gap)
```

What would you report?

Write one sentence about the historical trend, one about the same-job age pattern, and one about the earnings distribution. Include a number and suggest one next investigation.

What did we practice?

- **Predict** a pattern before calculating it.
- **Investigate** with grouped summaries, reshaping, joins, and plots.
- **Interpret** ratios and the two lines in an indicator model.
- **Communicate** a finding using concrete results.
- **Iterate** by checking another possible explanation.

Data and next week

- [Historical median earnings](#), from Census Table P-20.
- [2024 ACS workers](#) and [occupation names](#).
- [Source and preparation notes](#).
- Midterm review: Tuesday, October 13.
- Midterm: Thursday, October 15.

Week 6, Lecture 12: The gender pay gap

Closed October 8, 2026 at 3:44 PM

34 class questions · 1176 anonymous responses · 0 student questions

The following slides preserve what students submitted during class. No names or login information are included.

What do you think that percentage was in 2024?

Enter a number between 0 and 100.

76 anonymous responses

Other responses

76

Student responses

1. 100
2. 20
3. 40
4. 40
5. 55
6. 67
7. 67
8. 68
9. 70

Student responses

1. 70
2. 71
3. 72
4. 72
5. 72
6. 72
7. 73
8. 73
9. 74

Student responses

1. 75
2. 75
3. 75
4. 75
5. 75
6. 75
7. 75
8. 75
9. 75

Student responses

1. 75
2. 75
3. 75
4. 75
5. 76
6. 78
7. 80
8. 80
9. 80

Student responses

1. 80
2. 80
3. 80
4. 80
5. 80
6. 80
7. 80
8. 80
9. 80

Student responses

1. 80
2. 81
3. 81
4. 82
5. 82
6. 82
7. 82
8. 83
9. 83

Student responses

1. 84
2. 84
3. 85
4. 85
5. 85
6. 85
7. 85
8. 85
9. 85

Student responses

1. 85
2. 88
3. 88
4. 89
5. 90
6. 90
7. 90
8. 90
9. 90

Student responses

1. 92
2. 95
3. 95
4. 99

- Plot median earnings against year, with a different color for each sex. - Connect each group's earnings over time with a line. - Make separate panels for bachelor's and master's degree holders.

75 anonymous responses

Other responses

75

Student responses

1. college_history
2. college_history college_history |> ggplot(aes(x=year, y=pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay (\$1,000s)", color = "Sex")
3. college_history college_history |> ggplot(aes(y=pay, x=year, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay (\$1,000s)", color = "Sex")
4. college_history |> ggplot(aes(x = time , y = pay , color = sex)) + geom_line() + facet_wrap(~ pay) + labs(x = "Year", y = "Median pay (\$1,000s)", color = "Sex")
5. college_history |> ggplot(aes(x = time, y = pay, color = red)) + geom_line() + facet_wrap(~ degree) + labs(x = "Year", y = "Median pay (\$1,000s)", color = "Sex")
6. college_history |> ggplot(aes(x = time, y = pay, color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay (\$1,000s)", color = "Sex")
7. college_history |> ggplot(aes(x = time, y = pay)) + geom_line(aes(color = sex)) + facet_grid(~ education) + labs(x = "Year", y = "Median pay (\$1,000s)", color = "Sex")

Student responses

1. `college_history |> ggplot(aes(x = time, y = pay)) + geom_line(aes(color = sex)) facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year , y = pay)) + geom_line(aes(color = sex)) + facet_grid(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x = year , y = pay)) + geom_line(aes(color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay , color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay, color = sex )) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x = year, y = pay, color = sex, group = sex)) + geom_line() + facet_wrap(~ education) + labs( x = "Year", y = "Median pay ($1,000s)", color = "Sex" )`

Student responses

1. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line(aes(color = sex)) + facet_grid(~ education)`

Student responses

1. `college_history |> ggplot(aes(x = year, y = pay, color = sex)) + geom_line() + facet_wrap(~education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay, color = wrap)) + geom_line(aes(color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x = year, y = pay) + geom_line(aes(color = sex)) + facet_grid(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay) + geomline(aes(color = sex)) facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay)) + geom_line(aes(color = sex)) + facet_grid(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x = year, y = pay)) + geom_line(aes(color = sex)) + facet_grid(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(x = year, y = pay)) + geom_line(aes(color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay)) + geom_line(aes(color = sex)) + facet_grid(~education)`
3. `college_history |> ggplot(aes(x = year, y = pay))+ geom_line(aes(color = sex)) + facet_grid(education)`
4. `college_history |> ggplot(aes(x = year, y =pay)) + geom_line(aes(color = sex)) + facet_grid(~ education)`
5. `college_history |> ggplot(aes(x == time, y == pay)) + geom_line(aes(color = sex)) + facet_grid(~ education) labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(x= year, y = pay)) + geom_line(aes(color = sex)) facet_wrap(~ education) labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
7. `college_history |> ggplot(aes(x=year, y = pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(x=year, y= pay) + geom_line(aes(color = sex)) + facet_wrap(~ education) + # labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x=year, y=pay, color =sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(x=year, y=pay)) + geom_line(aes(color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(year, earnings, color = sex)) + geom_line() + facet_wrap(~ degree) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(year, median_earnings, color = sex)) + geom_line() + facet_wrap(~ degree) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(year, median_earnings, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(year, median, color = sex)) + geom_line() + facet_wrap(~ degree) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(year, pay, color = gender)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
3. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
6. `college_history |> ggplot(aes(year, pay, color = sex)) + geom_line(aes(color = sex)) + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `college_history |> ggplot(aes(year, pay)) + geom_line(aes(color = sex)) + facet_grid(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
2. `college_history |> ggplot(aes(x = year, y = pay)) + geom_line(aes(color = sex)) + facet_grid(~ education)`
3. `college_history |> ggplot(aes(x = year, y = median_earnings, color = sex)) + geom_line() + facet_wrap(~ degree_level) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
4. `college_history |> ggplot(aes(x = year, y = pay, color = sex )) + geom_line() + facet_wrap(~ education) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`
5. `college_history |> ggplot(aes(x = year, y = pay)) + geom_line(aes(color = sex)) + facet_grid(~ education)`
6. `college_history |> ggplot(aes(year, earnings, color = sex)) + geom_point() + geom_line() + facet_wrap(~ degree) + labs(x = "Year", y = "Median pay ($1,000s)", color = "Sex")`

Student responses

1. `ggplot(paygap, aes(x = year, y = median_earnings, color = sex, group = sex)) + geom_line() + geom_point() + facet_wrap(~ degree) + labs( x = "Year", y = "Median earnings", color = "Sex" )`

Suppose median earnings double for both groups. What happens to the dollar gap? What happens to women's earnings as a fraction of men's earnings? Which would you plot across these decades?

0 anonymous responses

No responses were submitted.

How would you turn the previous table into this one? Describe the steps or name the command you would use.

76 anonymous responses

Use pivot wider / pivot_wider

55

Responses recommending widening the table (pivot wider) to reshape from long to wide format.

Use summarize / aggregate

7

Responses suggesting summarizing or aggregating the data as the step to change the table.

Other responses

6

Compute a ratio variable

4

Response proposing to create a new ratio variable rather than reshaping the table.

Response themes for question 4

Use pivot longer / pivot_longer

4

Response recommending pivoting to a longer format, specifying which columns supply names and values.

Student responses

1. First pivot wider, with categories coming from sex and pay values coming from pay. Then divide
2. I would group by year, pivot sex wider
3. i would use pivot wider, names_from = gender, values_from = pay. then you can use mutate to create a new column with a gap
4. I would use pivot_wider
5. I would use pivot_wider() in R to turn the gender values into separate columns (Female and Male), while keeping year and education as they are.
6. I would use pivot_wider() so the Female and Male values become separate columns on the same row for each year and education level.
7. I would use pivot_wider() to turn the sex values into separate Female and Male columns, with their earnings as the values.
8. I'll use pivot_wider function. column_from = sex and value_from = pay.
9. One way that I would turn the previos table into this one would be to use pivot wider, as we need to go wider.

Student responses

1. pivot_wider
2. pivot longer
3. pivot wider
4. pivot wider
5. pivot wider
6. pivot wider
7. pivot wider
8. pivot wider
9. pivot wider

Student responses

1. pivot wider
2. pivot wider
3. Pivot wider
4. pivot wider by sex values to female and male
5. pivot wider to form new columns for female pay and male pay, so each entry has pay for both sexes.
6. Pivot wider with gender being the col name and median pay being the values
7. pivot wider, column names from sex, value from pay
8. pivot wider, value = pay, names from sex
9. Pivot Wider!!! HEHEHE I supposed I would pivot to use female and male as new columns. Afterwards I would filter out desired years.

Student responses

1. pivot_longer
2. Pivot_longer but keep year and education in place, so that female and male move from values in the table to column titles
3. pivot_wide
4. pivot_wide
5. pivot_wide with names_by sex and value_by pay
6. pivot_wider
7. pivot_wider
8. pivot_wider shows the wider table of the information we need
9. pivot_wider() create separate columns for female and male earnings, then use mutate() to calculate the female-to-male earnings ratio

Student responses

1. Pivt wider, names from sex and values from pay
2. Take the sex column and pivot wider over two new columns.
3. To do this we would need to use pivot_wider so that it is split into two columns, one for each gender.
4. To turn it into this table, you use the pivot_wider function to make female and male into columns.
5. use piviot widers
6. use pivot wider
7. use pivot wider
8. use pivot wider
9. use pivot wider to combine

Student responses

1. Use pivot_wider with Male and Female as the new columns and pay as the data
2. Use pivot_wider, with names_from "sex", and values_from "pay"
3. use pivot_wider() to make sex into columns.
4. Use pivot_wider() to turn the sex values into separate Female and Male columns containing median earnings, keeping year and education as the row identifiers.
5. We would use pivot wider to make the table have the male and female as columns
6. we would use pivot_wider so that the sex groups become two separate columns, names_to would Female/Male, values from sex
7. wider
8. wider
9. would want to pivot wider on year

Student responses

1. You could group by year, then pivot wider to get the pay for each gender in the same row.

Student responses

1. group by year, pivot wider
2. I would reshape the data from long to wide format using pivot_wider
3. summarize
4. Use left_join and pivot_wider
5. we first need to turn the female row with its male row into columns using year and education
6. You could add a column using the mutate function and then summarize the results using summarize function
7. You would have to group by year. and pivot wider

Student responses

1. 1. make a wider plot 2. calculate the percentage in mutate
2. change the == year
3. filter by year and pivot
4. transpose, filter and summarize
5. use pivot_wider()
6. You would change the table to wide data

Student responses

1. inner_join
2. Make a ratio variable
3. pivot wider then mutate(pay_gap = male/female)
4. pivot wider, names_from = c(male, female), values_from pay

Student responses

1. groupby education, select sex, and pivot longer
2. pivot_longer, with names from sex and values from pay
3. Pivot_wider()
4. You would need to pivot the table to change the format using pivot_longer

Put male and female pay on the same row for each year and education level.

71 anonymous responses

Other responses

71

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold mutate(ratio = Female / Male) history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = _____, values_from = _____) # endfold history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> pivot_wider(names_from = sex, values_from = pay) |> mutate(pay_ratio = Female / Male) history_work |> head(3)

Student responses

1. `# fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = gender, values_from = earnings) # endfold history_work |> head(3)`
2. `# fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)`
3. `# fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)`

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> mutate(ratio = Female/Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio= Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> # endfold mutate(ratio = Female / Male) history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = earnings) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = mediam) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) history_work <- history_work |> mutate(pay_ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold history_work |> mutate(ratio = Female / Male) # ratio is the fraction of women's pay compared to men

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold mutate(ratio = Female/Male) history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold |> mutate(pay_gap = Female/Male) history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) # endfold mutate(ratio = Female/Male) history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male * 100) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. `# fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male)`
2. `# fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) history_work |> head(3)`
3. `# fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)`

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(Ratio = Female/Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> # endfold mutate(ratio = Female / Male) history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(pay_ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female/Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ration = Female / Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(fraction = Female/Male) # endfold # Again this was created by doing female/male history_work |> head(3) # The calculation of male to female was created with female/male
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> mutate(ratio = Female/Male) # endfold history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> mutate(ratio= Female/Male) # endfold history_work |> head(3)
2. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> mutate(ratio = Female/Male) history_work |> head(3)

Student responses

1. # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay)|> mutate(ratio=female/male) # endfold history_work |> head(3)
2. history # fold: [Keep the college graduates] history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> # endfold # fold: [Put male and female earnings on the same row] pivot_wider(names_from = sex, values_from = pay) |> mutate(ratio = Female / Male) # endfold history_work |> head(3)
3. history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> pivot_wider(names_from = sex, values_from = earnings) |> mutate(pay_ratio = Female / Male) history_work |> head(3)
4. history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> pivot_wider(names_from = sex, values_from = median_pay)

Student responses

1. `history_work <- history |> filter(education %in% c("Bachelor's degree", "Master's degree")) |> pivot_wider(names_from = sex, values_from = pay) |> mutate(women_fraction = Women / Men) history_work |> head(3)`
2. `pivot_wider(names_from = sex, values_from = earnings)`

Calculate women's pay as a fraction of men's pay.

45 anonymous responses

Other responses

45

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = _____) # endfold history_work |> select(year, education, ratio) |> head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male * 100) # endfold history_work |> select(year, education, ratio)
|> head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = female / male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = female / male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female / Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/ Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/ Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. `# fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)`
2. `# fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)`
3. `# fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)`

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Student responses

1. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
2. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio = Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)
3. # fold: [Keep college graduates and put their earnings side by side] history_work <- college_history |> pivot_wider(names_from = sex, values_from = pay) |> #
endfold # fold: [Calculate women's pay as a fraction of men's pay] mutate(ratio =Female/Male) # endfold history_work |> select(year, education, ratio) |>
head(3)

Plot the female/male earnings ratio over time for both education levels. Submit your plot. State whether the gap has closed by 2024.

58 anonymous responses

Other responses

58

Student responses

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

Student responses

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

Student responses

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

Student responses

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

Student responses

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.

Student responses

1. # fold: [Plot the earnings ratio for each education level] history_gap |> ggplot(aes(x = year, y = ratio, color = education)) + geom_line() + # endfold # fold: [Mark equal earnings and label the axes] geom_hline(yintercept = 1, linetype = "dotted") + labs(x = "Year", y = "Female / male pay", color = "Education")
2. # fold: [Plot the earnings ratio for each education level] history_gap |> ggplot(aes(x = year, y = ratio, color = education)) + geom_line() + # endfold # fold: [Mark equal earnings and label the axes] geom_hline(yintercept = 1, linetype = "dotted") + labs(x = "Year", y = "Female / male pay", color = "Education") # endfold
3. history_gap |> ggplot(aes(x = year, y = ratio, color = education)) + geom_line(linewidth = 1.2) + geom_hline(yintercept = 1, linetype = "dashed", color = "gray50") + scale_y_continuous(labels = scales::label_number(accuracy = 0.01)) + scale_color_brewer(palette = "Dark2") + labs(x = "Year", y = "Female / male pay ratio", color = "Education", title = "Female-to-male earnings ratio over time") + theme_minimal() + theme(legend.position = "top", plot.titl

Student responses

1. i made the y = ratio the color = education
2. no, the gap has not closed by 2024.
3. No, the gap has not closed since neither line reaches 1.
4. The gap has closed between masters and bachelors but not from male and female
5. the gap has grown a little looking at 2024, but mostly stayed the same.
6. The gap has not closed by 2024. It essentially has remained almost static.
7. The gap has not closed, women still make ~70% of what men make
8. The gap is not really narrowing, it seems to be flat around 70%.
9. The gap isnt changing

Student responses

1. The pay gap appears to be very similar for the different education levels, and it appears to be stable at about 0.7.

How did the gap change for each education level? Did more education correspond to a smaller gap in 2024? Use the ratios in your answer.

0 anonymous responses

No responses were submitted.

Choose an education group. What would you regress on what? How would you use the fitted line to estimate the year of equal earnings?

78 anonymous responses

Regress earnings ratio on year

37

Suggests modeling the female-to-male earnings ratio against year and solving the fitted line for ratio = 1 to estimate the equal-earnings year.

Identify an education group

19

Mentions which education group to analyze (e.g., college graduates) without specifying regression details or estimating method.

Express uncertainty

6

States uncertainty or inability to answer; no modeling choice or education group given.

Ambiguous: regress ratio or gender/pay

5

Unclear statements that mention regressing ratio but mix up predictors (e.g., regress ratio on gender or regress bachelors on ratio) or lack clarity about variables.

Response themes for question 9

Other responses	4
Regress pay on education and/or year	4
Mentions regressing pay (income) on education level or year, rather than explicitly using the earnings ratio or solving for equal-earnings year.	
Specify linear model	3
Proposes using a straight line (linear regression) but does not state which education group or how to find the equal-earnings year.	

Student responses

1. bachelor's as group. Would regress pay ratio with time. By taking it when equals 1
2. Create a regression line for ratio against time, and we would want to see when the regression line reaches 1.
3. For bachelor's degree holders, regress the female/male earnings ratio on year
4. For masters student you could regress ratio on year and see when it will reach one
5. I select the bachelor's degree group because there seemed to be more of a change. I would use a linear regression because it seems to change approximately linearly. I would use the slope and intercept to calculate the estimated earning.
6. I would choose Bachelors as my education group, and I would regress the ratio of pay between men and women. I would try to see if there is a negative relationship that trends towards 1
7. I would choose the bachelor's degree group and regress the female-to-male earnings ratio on year. Then I would use the fitted regression line to find the year when the predicted ratio equals 1.
8. I would regress earnings ratio on year to get a fitted line to estimate the year where the ratio is equal to 1.

Student responses

1. I would regress female_pay/male_pay by year, to see if the ratio has a change related to time.
2. I would regress on ratio and use the fitted line to estimate the year of equal earnings
3. I would regress on ratio. If the predicted ratio is 1, then the year of equal earnings is estimated.
4. I would regress on year to predict pay ratio
5. I would regress ratio (the pay gap) on year, then set the resulting equation to 1 and solve.
6. I would regress ratio on time for a master degree.
7. I would regress ratio on year, and I would find where the fitted line reaches 1 on the y axis in order to estimate the year of equal earnings.
8. I would regress the female-to-male earnings ratio on year. Then I would set the predicted ratio equal to 1.
9. I would regress the female/male earnings ratio on year for one education group, then use the fitted line to find the year when the predicted ratio reaches 1.

Student responses

1. i would regress the ratio compared to the year, showing a line of best fit and finding where its y val hits 1 so the ratio is one
2. I would regress the ratio on time with just bachelors. When the intercept + slope = 1 is the tie when they are estimated equal earnings
3. I would regress the ratio over time to see its change. I would use the fitted line to estimate when the ratio will get to one
4. I would regress using the ratio on year. Use the fitted line until ratio is = 1.
5. I'm choosing the Bachelor's group and use a linear regression to regress pay ratio by year and predict pay gap by year.
6. Lets go with bachelors since thats the most common. I think I would regress on ratio over time and we would essentially regress until ratio equals 1
7. Master's degree, regress ratio on year, look at the slope and intercept to predict the year when ratio=1
8. regress on the ratio, the fitted line , the slope can be used to make a prediction
9. Regress pay ratio on year. Predict when median pay would be equal by finding the year when the regression equals one.

Student responses

1. regress ratio on time, then use model to predict when the ratio will equal 1
2. regress ratio on year and see when the line crosses the horizontal line at 1
3. Regress ratio on year. Use the fitted line until ratio = 1
4. regress the pay ratio gap to time. fitted line would show future predictions of the ratio until it intersected with 1
5. Regress the wage gap on time for bachelor's degrees, for example. We can see what time needs to be for the ratio to be one.
6. relate ratio to year for bachelor's degree using a linear model. Use the fitted line to determine the predicted year where the ratio is 1.
7. run a model on ratio x time, then given the estimated regression line, we could see when that regression line equals 1
8. Use bachelor degrees. Run a linear regression on the gap
9. we would have to regress ratio on year?

Student responses

1. You can regress ratio as the Y and time as the X. Then you could your fitted line, and solve for the time (year) that the value of the fitted line equals 1.
2. you would regress the ratio onto year.

Student responses

1. BA
2. bachelor's---regress on earning ratio?
3. Bachelors
4. Bachelors, I would regress on pay
5. Bachelors, regress on pay, by looking at the patterns
6. College graduates
7. For bachelor degree, the ratio is growing faster, which means the gap between female and male is closing to 1
8. For bachelors degree, I would regress the ratio on the year
9. I choose bachelor, I would use something like $\ln(\text{ratio} \sim t)$ and estimate earnings.

Student responses

1. i choose bachelors, i would regress average pay by gender group who only have a bachelors. i would see when the two lines intersect to see when the gender gap closed/will close
2. I would choose bachelors and regress on pay per year
3. I would choose the Bachelor's degree for female. I would use past earnings to predict the future.
4. I would choose the bachelor's degree group and regress the female-to-male earnings ratio on year. Then I would use the fitted regression line to find the year when the predicted ratio reaches 1, since that represents equal median earnings.
5. i would regress sex on pay
6. I would regress the ratio based on bachelors degree candidates and see how the ratio has changed between male and females
7. I'll choose Bachelor's Degree, and I'd regress pay on sex. The fitted line would estimate the gap based on the year.
8. If i chose bachelors, run linear regression based on pay to see if there is a trend
9. Looking at the education group of Master's Degree, I would regress to see when the ratio is predicted to equal one.

Student responses

1. Using bachelor's degrees, I would regress pay on year and use the final linear equation to see what the predicted year is when we reach a ratio of 1.

Student responses

1. bachlor's I would regress pay on gender and time
2. HAVing a headache rn, i cant think
3. i would regress pay (y) on education (x). i would use the fitted line to estimate the year of equal earnings but setting line regresison to 1
4. idk
5. unsure
6. We want to predict what the ratio of earnings between men and women is based off whether they have a bachelor's degree. Use the fitted line to draw it up to see when it hits 1 according to the model

Student responses

1. group 2
2. i would regress earnings ratio on year so i can see how the ratio changes over time
3. i would regress ratio on gender or pay
4. on different edu group
5. regressing the pay ratio itself on year to estimate the year of equal earnings using the fitted line

Student responses

1. bachelor's degree. I would regress on ratio ~ year. I'll use the R-squared value to see the correlation.
2. I would regress the effect of the classification of gender and maybe year after graduation to measure the earninngap
3. I'd choose masters. I'd regress on the difference in pay, and check when that reaches 0. Could also do it on the pay ratio, and check when ratio = 1.
4. We would test to see if there is a difference between the two groups by using a statistical test. If the ratio = 1 for both of the groups, we can conclude that the gap doesn't exist.

Student responses

1. education level and income
2. I think I will run regression based on the their sex and education and predict their payment
3. i would regress on pay on the year
4. I would regress on the pay, separated for females and males.

Student responses

1. a straight line can be used. if i choose my education group to be bachelor degree holders
2. Linear regression to find equation, then solve for the year when $y=1$
3. Use linear regration

Set the earnings ratio to 1 and solve for the year. How far beyond our last observation is that?

80 anonymous responses

Far future year around 2143–2144	54
Responses giving a calendar year in the mid-2100s as the solution (explicit years 2143 or 2144).	
Years beyond last observation in decades/centuries	8
Response stating how long after the last observation this occurs, expressed as a duration (about 153 years after 1991).	
Year stated without context or ambiguous	5
Very short or ambiguous numeric responses that give a number without explicit year context.	
Algebraic expression for calculation	4
Responses giving the algebraic formula or calculation to obtain the year without a final numeric simplification.	

Response themes for question 10

Cautionary or qualitative comment about reliability	4
Responses that state the predicted year and explicitly note the long extrapolation or reliability concerns.	
Decimal-year precision around 2143.63	3
Responses providing a more precise decimal-year estimate near 2143.63.	
Other responses	2

Student responses

1. ~2143
2. 21
3. 2143
4. 2143
5. 2143
6. 2143
7. 2143
8. 2143
9. 2143

Student responses

1. 2143
2. 2143
3. 2143
4. 2143 is the year
5. 2143 which is 119 years past our last observation
6. 2143, about 120 years from now
7. 2143, about 120 years from the end of the data
8. 2143, it is around 120 years later beyond 2024
9. 2143.62932877

Student responses

1. 2143.63, so far beyond
2. 2144
3. 2144
4. 2144
5. 2144
6. 2144
7. 2144
8. 2144
9. 2144

Student responses

1. 2144
2. 2144
3. 2144
4. 2144
5. 2144
6. 2144, 120 years after the last observation
7. 2144, 120 years past last observed
8. 2144, subtract the last year in data from 2144
9. 2144, which is 120 years passed our last observation

Student responses

1. about 2100
2. About 2143. This is about 120 years after our last observation
3. around 2144
4. At 2144, which is almost 120 years after our last observation.
5. giving year $\approx$ 2144. That is about 120 years beyond the last observation (2024).
6. part way through the year 2143
7. The earnings ratio will hit one in 2143.
8. The estimated parity year is 2144
9. The model predicts parity around 2144, about 120 years beyond the last observation in 2024.

Student responses

1. The resulting year is around 2143. That's far beyond now.
2. the year 2144
3. The year is about 2143.
4. the year is predicted to be 2144.
5. The year would be 2143.
6. the year would be 2144, or 120 years after our last prediction.
7. We are estimated to reach a ratio of 1 in 2,143. This is about 128 years beyond our last observation.
8. year 2143
9. year 2144

Student responses

1. \$143.63 beyond the last observation
2. $1 = .6430 + .002339(x-1991)$ 2. $.357 = .002339(x-1991)$
3. 119 years
4. 120 years beyond last obs
5. 123 years, 2144
6. About 153 years after 1991
7. It is projected to hit 1 in 2143, 152 years after 1991.
8. like 153 years, really beyond

Student responses

1. 10
2. 2143 would be the year the gender pay gap closes. its about 152 years away from 1991
3. 2144. 153 years later than 1991
4. it would be $0.6430 + 0.002339 \times (2025 - 1991)$, I think it would be aorund 90%
5. So the model predicts the earnings ratio reaches 1 around the year 2144. If the last observation was 2016, then this is about $2144 - 2016 = 128$ years beyond the last observation.

Student responses

1. $((1 - 0.6430)/0.002339) + 1991 = \text{year}$
2. $((1 - 0.6430)/0.002339) + 1991 = 2144$
3. $(1 - 0.643)/0.002339 + 1991$
4. $0.6430 + 0.002339 \times (\text{year} - 1991) = 1$ solve and get year=2144, that so far away

Student responses

1. about 120 years beyond our last observation in 2024. This is a large extrapolation, so it assumes the straight-line trend continues far outside the observed data,
2. not for a very very very long time. like a ridiculously long time.
3. The model predicts earnings parity around 2144, which is approximately 120 years after our last observation in 2024. This is a long extrapolation, so the prediction may not be very reliable.
4. The model predicts equal earnings around 2144, which is about 120 years after the last observation in 2024.

Student responses

1. 2143.63
2. Set $1 = 0.6430 + 0.002339 \cdot (\text{year} - 1991)$. So $\text{year} - 1991 = 0.357 / 0.002339$, which is about 153, so year is about 2144. Our last observation is 2024, so that's roughly 120 years beyond the data.
3. Setting ratio to 0, $\text{year} - 1991 = 152$

This fit uses 2000–2024. Use its coefficients to get another parity year. Would you report either year as a precise prediction? Explain.

0 anonymous responses

No responses were submitted.

- Group the workers by sex and report n and median_pay for each group. - Report the difference in median earnings in dollars.

77 anonymous responses

Other responses

77

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(____) |> summarize(n = n(), median_pay = median(____), .groups = "drop") # endfold summary_work summary_work <- work |> group_by(sex) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE)) summary_work max(summary_work\$median_pay) - min(summary_work\$median_pay)
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1] #median pay difference
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay 100/120 mp[2] = mp[1] # median pay difference
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp [1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2] - mp[1] # median pay difference
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop",) # endfold summary_work\$median_pay 101.525 - 121.830

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[1] - mp[2] summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work 102 - 122
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay 100/120 mp[2] - mp[1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay 100/120 mp[2] - mp[1] summary_work

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # A tibble: 2 × 3 # sex n median_pay # <chr> <int> <dbl> #1 Female 7447 102. #2 Male 19581 122. mp <- summary_work\$median_pay
mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # the difference is 20 summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold difference <- summary_work\$median_pay[2] - summary_work\$median_pay[1] difference
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold difference = median_pay summary_work

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold difference = summary_work\$median_pay["Female"] - summary_work\$median_pay["Male"] summary_work difference
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay 100/120 mp[2] - mp[1] summary_work
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay summary_work mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
5. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay mp[2] - mp[1] summary_work
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp <- summary_work\$median_pay summary_work mp[2] - mp[1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold mp<-summary_work\$median_pay mp[2]-mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold np <- summary_work\$median_pay np[2] - np[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
5. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay 100/120
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work #the difference in median earnings is 20 dollars
5. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2]-mp[1]
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay |> mp[2] - mp[1] #median difference
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp<- summary_work\$median_pay mp[2]-mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work summary_work |> summarize(difference = median_pay[2] - median_pay[1]) summary_work |> summarize(difference = median_pay[sex == "Male"] - median_pay[sex == "Female"])
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work gap <- summary_work\$median_pay[summary_work\$sex == "Male"] - summary_work\$median_pay[summary_work\$sex == "Female"] gap
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2] - mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work mp <-summary_work\$median_pay mp[2] - mp[1]
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work np<-summary_work\$median_pay np[2]-np[1]
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work summary_work\$median_pay[2] - summary_work\$median_pay[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold summary_work\$median_pay[1] - summary_work\$median_pay[2]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold print(((122-102)*1000) summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold print(122-102) summary_work
4. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") mp[2] - mp[1] summary_work
5. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")|> # endfold mp <- summary_work\$median_pay mp[2]-mp[1]

Student responses

1. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay)) #
endfold summary_work #mp <- summary_work\$median_pay #mp[2] - mp[1]
2. # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups =
"drop") # endfold summary_work
3. # fold: [Calculate each group's sample size and median pay] summary_work\$median_pay
4. acs # fold: [Calculate each group's sample size and median pay] summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay),
.groups = "drop") # endfold summary_work mp <- summary_work\$median_pay mp[2] - mp[1]
5. summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(earnings, na.rm = TRUE), .groups = "drop") summary_work

Student responses

1. `summary_work <- acs |> group_by(sex) |> summarize( n = n(), median_pay = median(earnings), .groups = "drop" ) summary_work summary_work |> summarize( gap_dollars = median_pay[sex == "Male"] - median_pay[sex == "Female"] )`
2. `summary_work <- acs |> group_by(sex) |> summarize( n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop" ) summary_work`
3. `summary_work <- acs |> group_by(sex) |> summarize( n = n(), median_pay = median(pay) ) summary_work summary_work |> summarize(difference = max(median_pay) - min(median_pay))`
4. `summary_work <- acs |> group_by(sex) |> summarize( n = n(), median_pay = median(pay), .groups = "drop" ) summary_work$median_pay median_pay[2] - median_pay[1]`
5. `summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") summary_work summary_work |> summarize(gap_dollars = (median_pay[sex == "Male"] - median_pay[sex == "Female"]) * 1000)`

Student responses

1. `summary_work <- acs |> group_by(sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") summary_work summary_work$median_pay[2] - summary_work$median_pay[1]`

After keeping one job and degree, do you expect a similar pay gap among workers in their twenties and workers in their forties? Explain. What would you plot to check it?

83 anonymous responses

Accumulation widens gap over time

13

Students believe small initial differences compound over careers, producing a larger pay gap in older age groups.

Gap decreases when role and degree fixed

13

Students expect a smaller pay gap if workers keep the same job and degree, attributing similarity in pay to identical roles or qualifications.

Gap may widen due to sexism

11

Students suggest the pay gap could grow with age because of gender-based raises or discriminatory practices favoring men.

Removing confounders reduces gap

11

Students argue that controlling for confounding variables (by holding job and degree constant) should reduce observed pay differences between age groups.

Response themes for question 13

Affirms question without detail**9**

Very short responses that simply answer yes or give minimal affirmation without explanation or plotting suggestions.

Depends on job, degree, and other factors**9**

Students emphasize conditionality: outcomes vary by job, degree, and other factors rather than a single uniform effect.

Older workers earn more**9**

Students expect higher pay for older workers, citing experience or age-related pay increases as reasons for a larger gap.

Plot medians or lines by age and gender**8**

Students recommend plotting median earnings against age or using separate lines/groups for men and women to visualize pay differences.

Student responses

1. I expect a bigger pay gap among workers in their forties
2. I expect a gap smaller in the 20s compared to the 40s, as workers in their 40s may still be dealing with worse subconscious effects from the older larger pay gap
3. I expect a larger one, because when staying longer, people know each other better and those with better skills will be paid more. I would try to plot the pay gap using box plot
4. I expect a larger pay gap between those in their forties because there's more opportunity for increases between 20 and 40. I would use a histogram to look at it
5. I expect the pay gap to decrease, but I don't expect it to close. I anticipate that the gap will increase as age increases. I would plot pay gap vs age
6. I would expect the pay gap to be larger for workers in their forties because differences can build over time, and I would plot pay against age with separate lines or groups for men and women.
7. I would imagine that the pay gap would increase over time, since there's a greater variation of pay at senior positions as opposed to entry-level positions. Differences in family dynamics would also factor

Student responses

1. I'd expect a larger pay gap among workers in their forties, because differences in promotions, experience, and career breaks could accumulate over time.
2. no I would expect a greter pay gap in their forties because there are more factors that may affect a carreer
3. No, I would expect the gap to be smaller for workers in their twenties and larger for older workers. Early on people in the same job/degree start more similarly, but differences in promotions, raises, and time out for caregiving can accumulate with age.
4. probably smaller gap among workers in their 20s and a larger gap among workers in their 40s, because it accumulate with time.
5. Would expect pay gap to differ, since people in better higher income jobs in 20s eventually make way way more than people in worse low income jobs.
6. Yes because they increase at the same rate. Plot the amount people in the same field earn after years in their field and compare the rates

Student responses

1. I expect that the pay gap would decrease
2. I expect the gap will be smaller because you are focusing on just one area instead of a wide ranging list of jobs
3. I suspect it will be smaller, but will not converge to 1. Men were accepted in STEM fields earlier, thus the ratio of men to women in stem fields is likely higher in higher age groups. People in the industry tend to make more money earlier, thus my expectation.
4. I think the gap could differ by age and also the gap would decrease when the role and degree is fixed
5. I think the gap will be smaller, since female dominated carrers do tend to pay less, but also there probably is still some gap
6. I think the pay gap will decrease
7. I think the pay gap will get smalller.
8. I would imagine that the pay gap would shrink because a big difference in the pay gap between young and old workers is their job.
9. pay gap will stay smaller, i would filter job and degree lok at each correlated plot it out male vs female and se if that expands or doesnt

Student responses

1. probably a gap smaller when they are older because then it is much more about how good you are at your job
2. Probably, because there may be different levels of that particular job. Like SWE I vs SWE II for example. I would check the actual role.
3. Smaller since it is the same roles
4. The pay gap will likely decrease. There will be less education disparity and less skill differences.

Student responses

1. different pay gap
2. I don't expect a similar pay gap among workers in their twenties and those in forties due to difference in working experience. I would plot a scatter plot to check it.
3. I don't think the pay gap would be the same. Workers in their forties may have a larger pay gap because differences in promotions, career progression, and work experience can accumulate over time. I would plot average salary by age group and gender, keeping the job and degree the same, to compare the gaps.
4. I expect a similar pay gap because workers in their forties typically have more tenure and are more likely to be promoted. i would plot a ggplot
5. I think that the pay gap might increase as age increases because interns or entry level positions will probably pay similarly between genders but older people will get higher positions and with that some sexism could come into play when deciding salaries and such

Student responses

1. I think the age gap remains the same or even gets wider because men are often given promotions over women. I would filter the data by job and degree and then create a line plot with each line representing a different age group.
2. I think the pay gap will get bigger as people age as men may be getting larger raises than women. I would plot a line plot, plotting pay gap by year colored by sex
3. I think the pay gap will get bigger for workers in their forties as men are more likely to get raises and pay bumps due to sexism.
4. I think the pay gap would be larger because people with more experience are paid better and generally, women have kids, take maternity leave, which hurts their chances of getting a promotion and staying with the company for that long. I would visualize this with a boxplot
5. i would expect the increase to be more, because males might be more likely to get promoted. I would plot median pay by sex over different age groups.
6. I would not expect a similar gap. I would expect the gap to be smaller for younger people because there were more barriers for women in the past

Student responses

1. Hopefully it'll get smaller because it'll get rid of confounding variables that could affect the pay difference.
2. I expect the pay gap will get smaller as people age, because raises and promotions may close the gap.
3. I think the gap could differ by ages, as older women are less likely to take time off for childcare once they are past that point, which is a big "argument" for the pay gap. To check it, I would plot age ranges against each other.
4. i think the gap will increase as workers get older because men will receive higher pay raises.
5. i would expect an enlarger paygap, and I have graph two line plots to check it
6. I wouldn't expect the same because experience could be a confounding factor. To check it, I would plot median pay by sex separately for workers in their twenties and workers in their forties, while keeping job and degree fixed.
7. If we removed confounds, the gap would most likely decrease.
8. no I dont expect a pay gap, i would use a line plot to plot the difference
9. probably higher pay for older people due to bonuses and promotions they may have received throughout their career

Student responses

1. The pay gap should be smaller among workers in their twenties because workers in their 40s are probably more likely to be male.
2. Yes, we can plot the both of them out

Student responses

1. I think the larger the age is, the larger the gap would be. Because as women grow old they would be discriminated more due to family issues? I would plot line chart to check it
2. I think the pay gap would be larger for people in their forties.
3. I would expect a similar pay gap because both would be on an even playing field in terms of degree. I would plot the pay over time.
4. i would expect it to remain the same given data tveve seen that the pay gap hasnt really changed. i would plot pay gap vs years of experience to check
5. I would expect the pay gap to be different between age groups. The people in their 40s properly have a bigger salary difference.
6. yeas
7. yes
8. yes
9. Yes. Plot filtering on ages

Student responses

1. Depends on the job and degree but overall I would say, other factors do matter
2. I expect certain job the pay gap would be enlarged. and at their forties, I expect pay gap is larger, plot, histogram
3. I expect the pay gap among workers in their twenties and workies in their forties to be slightly different, since it may be more likely for employers of new hires to attempt to close the pay gap
4. I expect the pay gap could widen because if men are better positioned, they could be the ones to get promotions. To check this, I would create separate groups for age and plot this
5. I think keeping only 1 job and degree depending on the job, i think the pay gap would differ, some fields
6. I think that the gap will remain relatively the same and decrease maybe slightly. I think that people just entering the workforce will be paid more similarly but that this gap will widen as they go further on in their career.
7. I think the pay gap will stay relatively the same, we should expect that these pay gaps will remain constant. I would do a scatter plot again

Student responses

1. It will probably differ by age since older women are less likely to take maternity leave etc, and have more experience. to make a plot, filter by age as well on a line.
2. No I dont because life doesn't change and the way promotions happen might play a role in how the gap changes as well. using a plot would be a good way to check this

Student responses

1. I think people in 30s will have a larger pay than people in 20s as they probably have more experience which probably makes them get paid way more
2. I think workers in their 40s will have a larger pay gap as salaries can grow in different rates.
3. i would expect a larger pay gap for older people
4. No, the younger people might get paid more for more complex jobs but the older people have probably been there longer
5. No. People in their forties are going to have more experience and their pay will reach a maximum amount
6. older workers may have higher pays since they have more experience, we can check by boxplot
7. Yes I think when the age is larger, the gap should be larger
8. Yes, because people earn promotions and raises over time, so older workers are likely to earn more. A line plot with median pay for workers over time, with one line for 20s workers and one line for 40s workers would show if there was a trend.
9. Yes, the people have more experience will earn more

Student responses

1. I do not expect a similar pay gap would make a grouped boxplot of earnings by sex, comparing workers in their twenties and forties while keeping the job and degree the same
2. I don't expect the pay gap to be exactly the same. Workers in their forties may have a larger gender pay gap because of differences in work experience and career progression. I would use a boxplot to compare men's and women's earnings in their twenties and forties, keeping the same occupation and education level.
3. I expect a larger gender pay gap among workers in their forties because earnings may increase differently with age. I would plot median earnings against age.
4. I expect that the gap will be smaller among younger workers, as the gap is closing over time. I would plot these in two different graphs to see if this is true
5. I think there will be a gap for sure because you should expect to get paid more the longer you work in the industry. I think that the pay gap will get smaller for gender with age.

Student responses

1. I would expect a pay gap between workers in twenties and forties because workers in their forties have had more time to climb up the ladder and get a higher paying job with more experience. use a bar plot or line graph to show the difference in trends
2. I would expect the pay gap to be larger in their 40s because that is the time when jobs allow a huge pay gap between positions, in your 20s there is more of a baseline analyst salary. I would check it with 2 scatter plots.
3. No, I don't think they would have the same pay.

- Join `occupation_names` to `acs` and keep software developers with bachelor's degrees. - Calculate the number of workers and median pay for each sex and age. Call the columns `n` and `median_pay`. - Plot median pay by age, distinguishing men and women. Label the axes and include the pay units. - Add a code comment describing how the gap differs between workers in their twenties and forties.

83 anonymous responses

Other responses

83

Student responses

1. `# acs |> select(occupation)|> print()` # fold: [Find software developers with a BA] `software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's")` # endfold `software_work |> print()` # fold: [Summarize earnings by sex and age] `age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay))` # endfold # fold: [Plot the age-specific medians] `age_work |> ggplot() + geom_line(aes(x=age,y=median_pay, color = sex))` # endfold # Compare pay gaps in the two age groups. # Your answer: MY CODE COMMENT # The gap begins between 30s and 45 and widens, # then shrinks after 45

Student responses

1. `# Find software developers with a BA software_work <- acs |> left_join(occupation_names, by = "occupation") |> filter(occupation_name == "Software developers", degree == "Bachelor's") # Summarize earnings by sex and age age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # Plot the age-specific medians age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (USD)", color = "Sex") # The gap appears smaller in workers' twenties and larger in their forties.`

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation") |> filter(occupation_name == "Software developers", education == "Bachelor's degree") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(earnings, na.rm = TRUE), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age (years)", y = "Median annual pay (dollars)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # The pay gap between men and women is smaller among workers in their # twenties and larger among workers in their forties.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> ____ # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> ____ # endfold # fold: [Plot the age-specific medians] age_work |> ____ # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> ____ # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> ____ # endfold # fold: [Plot the age-specific medians] age_work |> ____ # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> ____ # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> ____ # endfold # fold: [Plot the age-specific medians] age_work |> ____ # endfold # Compare pay gaps in the two age groups. # Your answer: # Filter to software developers with the same degree software_work <- acs |> filter(occupation == "Software developers", education == "Bachelor's degree", age >= 20, age < 50) |> mutate(age_group = if_else(age < 30, "20s", "40s")) |> filter(age < 30 | age >= 40) # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age_group) |> summarize(median_earnings = median(earnings, na.rm = TRUE), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age_group, y = median_earnings, fill = sex)) + geom_col(position = "dodge") + labs(x = "Age Group", y = "Median Earnings", title = "Median Earnings by Age Group and Sex") # endfold

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> _____ # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> _____ # endfold # fold: [Plot the age-specific medians] age_work |> _____ # endfold developers <- acs |> left_join(occupations, by = "occupation_code") |> filter(occupation == "Software developers", degree == "Bachelor's") age_summary <- developers |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop") ggplot(age_summary, aes(x = age, y = median_pay, color = sex)) + geom_line() + geom_point() + labs(x = "Age", y = "Median annual pay (\$)", color = "Sex") # Compare pay gaps in the two age groups. # Your answer: # The gender pay gap is smaller among workers in their twenties # and larger among workers in their forties.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> filter(education == ba) # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> ____ # endfold # fold: [Plot the age-specific medians] age_work |> ____ # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> inner_join(occupation_names, by = "occupation_code") |> filter(occupation == "Software developers", degree == "Bachelor's") age_medians <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # endfold age_medians |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + geom_point() + labs(x = "Age", y = "Median pay (\$1,000s)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: # The gender pay gap is smaller among workers in their twenties # and larger among workers in their forties.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> inner_join(occupation_names) |> filter(occupation == "Software Developer", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: Pay gap differs

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> inner_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (\$)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: The pay gap between men and women is relatively small in their twenties, but widens significantly as workers reach their forties.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = 'occupation_code') |> filter(education == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarise_(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occ_code") |> filter(occ_name == "Software developers", education == "Bachelor's degree") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age (years)", y = "Median annual pay (USD)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: In their twenties, men's and women's median pay is close; # by their forties, men's median pay is noticeably higher, so the gap widens with age.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occ") |> filter(occ_title == "Software developers", educ == "Bachelor's degree") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(incearn, na.rm = TRUE), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + geom_point() + labs(x = "Age (years)", y = "Median annual earnings (US dollars)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: Among workers in their twenties, men's and women's median pay # is close together, so the gap is small. By their forties the lines have # pulled apart, and men earn noticeably more than women at the median. # (Check this against your plot: the wording should match what you actually see.)

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(degree == "Bachelor's" & occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: Pay gaps gradually become larger when people grow older.

Student responses

1. `# fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(degree == "Bachelor's",
occupation == "Software developer") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n
= n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() #
endfold # Compare pay gaps in the two age groups. # Your answer:`

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(education == "Bachelor's", occupation == "Software developers") |> # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(age, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: Pay gap increases

Student responses

1. `# fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(occupation == "Software developers", education == "Bachelor's degree") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay ($)") # endfold # Compare pay gaps in the two age groups. # Your answer: # The gender pay gap may be different between workers # in their twenties and forties.`

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation") |> filter(occupation_name == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + geom_point() + labs(x = "Age", y = "Median pay (dollars)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: The pay gap between men and women can differ between workers in their twenties and forties.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by="occupation_code") |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(y = median_pay, x = age, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The gap appears to widen as people get older.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(education == "Bachelor's", occupation == "Software developers") |> # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: In their twenties, women and men has similar median pay, so the gap is small

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software Developers") # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # fold: [Plot the age-specific medians] age_work |> geom_line(aes(x = age, y = median_pay, color = sex)) # Compare pay gaps in the two age groups. # Your answer: the gap of twenties is lower than forties
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bacehlor's", occpuation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ____ # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's" occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n=n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:gap is bigger in 40s

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The pay gap is usually small in workers' twenties but becomes much larger in their forties.
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") |> # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold age_work # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") |> # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by_(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> _____ # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") software_work # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex,age) |> summarize(n=n(),median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x=age, y = median_pay)) _____ # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree=="Bachelor's degree") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n= n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's degree", occupation == "Software developers") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developer's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's") |> filter(occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() # endfold # Compare pay gaps in the two age groups. # Your answer: The gap is much smaller among workers in their twenties than workers in their forties.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == 'Software developers', degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: In their 20's the gap is very small/not there while in there 40's the gap is fairly large

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developer", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The gap appears to increase as age increases

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developer", education == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) > summarize(median_pay = median(pay)) software_work # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: They differ as the pay gap is smaller for the older age group.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers") |> filter(education == "Bachelor's degree") |> # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software_bachelors", degree == "Bachelor's") |> print() # endfold # fold: [Summarize earnings by sex and age] ##age_work <- software_work |> _____ # endfold # # fold: [Plot the age-specific medians] # age_work |> _____ # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> # endfold filter(degree == "Bachelor's", occupation == "Software developers") |> # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) _____ # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> _____ # endfold # fold: [Plot the age-specific medians] age_work |> _____ # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x=age, y=median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The pay gap seemed to increase from people in their twenties to 40s

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", education == "Bachelor's") |> # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: the gap is bigger

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The gap is greater between workers the older they are.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex,age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # There is a way bigger gap for workers in their forties than people in their twenties # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age)|> summarize(n = n(),median_pay= median(pay)) # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x =age, y=median_pay, color = sex))+ geom_line() # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> _____ # endfold # fold: [Plot the age-specific medians] age_work |> _____ # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers") |> filter(degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: the gap increases into the 40s

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> print() |> filter(degree == "Bachelor's", occupation == "Software developers") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: the gap in the twenties is not that big, almost identical, but it starts to get wider in the 40s to the late 40s, then has a sharp increase near the 50s

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + gg_line(aes(x=age, y=median_pay, color=sex)) + labs(x = "Age", y = "Pay (in dollars)", color = "Gender") # endfold # Compare pay gaps in the two age groups. # Your answer: pay gaps start to increase as you get older

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The gap closes as age increases
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(education == "Bachelor's", occupation == "Software developers") # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay), color = sex) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names)|> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age)|> summarize(n=n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot()+ geom_line(aes(x=age, y =median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:The gap increases as age increases
2. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names)|> filter(degree == "Bachelor's", occupation=="Software developers")> print # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex,age)|> summarize(n=n()), median_pay=median(pay) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot()+ geom_line(aes(x=age, y=median_pay), color=sex) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names)|> filter(degree == "Bachelor's", occupation == "Software developers") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex,age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: the gap widens as age grows older. the gap is small for 20s, and large for 40s

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names)|> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age)|> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex))+ geom_line() # Compare pay gaps in the two age groups. # Your answer: the gap increases for workers in their forties

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x=age, y=median_pay, color=sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: for women it stays around the same for their 20s and 40s, for men, pay increases

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> inner_join(occupation_names)|> filter(degree=="Bachelor's", occupation=="Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex,age)|> summarize(n=n(),median_pay=median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + # endfold labs(x = "Age", y = "Median pay(unit: 1000\$)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: as age is larger, the gap is also larger

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", degree == "Bachelor's") |> print() # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x=age,y=median_pay,color=sex)) + geom_line() # endfold # Compare pay gaps in the two age groups. # Your answer: #As age increases, so does the gender pay gap

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names)|> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age)|> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer: The gap is bigger in people's 40s than their 20s.

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(occupation == "Software developers", degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (USD)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: There definitely is a gap in pay between Males and Females. Males tend to have a greater median pay than females as the age increases

Student responses

1. # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot() + geom_line(aes(x=age, y=median_pay, color = sex)) # endfold # Compare pay gaps in the two age groups. # Your answer:

Student responses

1. acs # fold: [Find software developers with a BA] software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(occupation == "Software developers") |> filter(degree == "Bachelor's") # endfold # fold: [Summarize earnings by sex and age] age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians] age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Pay") # endfold # comment # Compare pay gaps in the two age groups. # Your answer: The pay gap is very simliar # It differs a lot

Student responses

1. `acs %>% left_join(occupation_names, by = "occupation_code") %>% filter(occupation == "Software developers", degree == "Bachelor's") %>% group_by(sex, age) %>% summarize(n = n(), median_pay = median(pay, na.rm = TRUE)) %>% ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (USD)", color = "Sex") # Compare pay gaps in the two age groups. # The pay gap looks smaller in workers' twenties and larger in their forties.`
2. `age_pay <- acs |> inner_join(occupation_names, by = "occupation_code") |> filter(occupation == "Software developers", degree == "Bachelor's") |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) age_pay |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay ($1,000s)", color = "Sex") # The pay gap is smaller for workers in their twenties and larger for workers in their forties.`

Student responses

1. answer: `software_work <- acs |> left_join(occupation_names, by = "occ1990") |> filter(occupation == "Software Developers", degree == "Bachelor's")`
`age_work <- software_work |> group_by(sex, age) |> summarize( n = n(), median_pay = median(income), .groups = "drop" )`
`age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs( x = "Age", y = "Median pay (US dollars)", color = "Sex" )`
The pay gap is usually small in workers' twenties but becomes much larger in their forties.

Student responses

1. `occupation_names |> inner_join(acs, by = "occupation") |> filter(occupation_name == "Software developers", degree == "Bachelor's") |> group_by(sex, age) |> summarize(n = n(), median_pay = median(income), .groups = "drop") |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (USD)", color = "Sex") # The gap appears smaller when workers are in their twenties compared to when they are in their forties, and the gap gradually gets wider as time goes on`
2. `software_pay <- acs %>% left_join(occupation_names, by = "occupation") %>% filter(occ_name == "Software developers", degree == "Bachelor's") %>% group_by(sex, age) %>% summarize(n = n(), median_pay = median(income), .groups = "drop") # The pay gap is usually smaller in workers' twenties and larger in their forties. ggplot(software_pay, aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (US dollars)", color = "Sex")`

Student responses

1. `software_work <- acs |> inner_join(occupation_names, by = "occupation") |> filter(occupation_name == "Software developers", education == "Bachelor's degree")`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")`
`age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay ($)", color = "Sex")` # Compare pay gaps in the two age groups. # Your answer: bigger gap in 40s

Student responses

1.

```
software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter( occupation == "Software developers", degree == "Bachelor's degree" )
age_work <- software_work |> group_by(sex, age) |> summarize( n = n(), median_pay = median(pay), .groups = "drop" )
age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_point() + geom_line() + labs( x = "Age", y = "Median pay ($1,000s)", color = "Sex" ) # Compare pay gaps in the two age groups.
# Your answer: # The gender pay gap is small among workers in their # twenties but larger among workers in their forties. # Compare pay gaps in the two age groups. # Your answer:
```

Student responses

1. `software_work <- acs |> left_join(occupation_names, by = "occupation_code") |> filter(occupation == "Software developers", degree == "Bachelor's")`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")`
`age_work |> ggplot(aes(age, median_pay, color = sex)) + geom_line() + labs(x = "Age (years)", y = "Median pay ($1,000s)", color = "Sex")` # Compare pay gaps in the two age groups. # Your answer: # In their twenties, men and women have similar median pay # (about \$112k each at ages 25-29). In their forties, the gap is much # larger: about \$159k for men vs. \$122k for women, a gap of roughly \$37k.

Student responses

1. `software_work <- acs |> left_join(occupation_names) |> filter( education == "Bachelor's degree", occupation == "Software developers") |> print()`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")`
`age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay ($1,000s)", color = "Sex")`
Compare pay gaps in the two age groups. # Your answer: In their twenties, the lines are close together and the gap is small. # By their forties, the lines pull apart and men's median pay is clearly higher than women's.

Student responses

1. `software_work <- acs |> left_join(occupation_names) |> filter( occupation == "Software developers", education == "Bachelor's degree" )`
`age_work <- software_work |> group_by(sex, age) |> summarize( n = n(), median_pay = median(earnings, na.rm = TRUE), .groups = "drop" )`
`age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs( x = "Age (years)", y = "Median annual pay ($)", color = "Sex" )`
Hypothesis: the pay gap is smaller in the twenties # and larger in the forties.
2. `software_work <- acs |> left_join(occupation_names) |> filter( occupation == "Software developers", education == "Bachelor's" )`
fold: [Summarize earnings by sex and age]
`age_work <- software_work |> group_by(sex, age) |> summarize( n = n(), median_pay = median(pay))`
endfold
fold: [Plot the age-specific medians]
`age_work |> ggplot()+ geom_line(aes(x=age, y = median_pay, color = sex))`
endfold
Compare pay gaps in the two age groups. # Your answer:

Student responses

1. `software_work <- acs |> left_join(occupation_names) |> filter(degree == "Bachelor's", occupation == "Software developers")`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")`
`age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay (US dollars)", color = "Sex")` # endfold # fold: [Plot the age-specific medians]
`age_work |> ggplot() + geom_line(aes(x = age, y = median_pay, color = sex()))` # endfold # Compare pay gaps in the two age groups. # Your answer: The gap looks smaller in workers' twenties and larger in their forties.
2. `software_work <- acs |> left_join(occupation_names) |> filter(occupation == "Software developers", education == "Bachelor's")`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay))`
`age_work |> ggplot() + geom_lines(aes(x = age, y = median_pay, color = sex))`

Student responses

1. `software_work <- acs |> left_join(occupation_names) |> filter(occupation_name == "Software developers", education == "Bachelor's degree")`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop")`
`age_work |> ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + geom_point() + labs(x = "Age", y = "Median pay ($)", color = "Sex")` # The gender pay gap is smaller in workers' twenties and larger in their forties.

Student responses

1. `software_work <- acs |> left_join(occupations, by = "occupation_code") |> filter(degree == "Bachelor's", occupation_code == 1021) # fold: [Summarize earnings by sex and age]`
`age_work <- software_work |> group_by(sex, age) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Plot the age-specific medians]`
`ggplot(aes(x = age, y = median_pay, color = sex)) + geom_line() + labs(x = "Age", y = "Median pay ($1,000s)", color = "Sex") # endfold # Compare pay gaps in the two age groups. # Your answer: The pay gap between men and women is smaller in their twenties than in their forties, with men earning more than women in both age groups.`

Among software developers with bachelor's degrees, median earnings are similar for younger men and women, but farther apart for older workers. What could be driving this pattern? Suggest an explanation and describe what additional data you would use to investigate it.

0 anonymous responses

No responses were submitted.

- Keep 40-hour workers aged 25-29 or 40-49. Label the two groups `age_group` using `ifelse()`.
 - Report `n` and `median_pay` for each sex and `age_group`.
 - Put Male and Female medians on the same row for each `age_group` and calculate gap in dollars as Male minus Female.
-

77 anonymous responses

Other responses

77

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(earnings, na.rm = TRUE), .groups = "drop") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> select(age_group, sex, median_pay) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> _____ # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> _____ # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(age < 30, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop") matched_gap_work <- matched_work |> select(age_group, sex, median_pay) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) matched_gap_work matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(age < 30, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> select(age_group, sex, median_pay) |> tidyr::pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = (Male - Female) * 1000) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(weekly_hours == 40, age %in% c("25-29", "40-49"))
|> mutate(age_group = if_else(age == "25-29", "25-29", "40-49")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <-
matched_hours_work |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") |> pivot_wider(names_from = sex,
values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, age %in% c("25-29", "40-49")) |> mutate(age_group = ifelse(age == "25-29", "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")
endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> select(age_group, sex, median_pay) |> tidyr::pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = male - female) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, age %in% c(25:29, 40:49)) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "40-49")) matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, age >= 25, age <= 49) |> mutate(age_group = ifelse(age >= 25 & age <= 29, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay), .groups = "drop")
endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> select(sex, age_group, median_pay) |> tidyr::pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, age >= 25, age <= 49) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "40-49")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work |> pivot_wider(names_from = sex, values_from = median_pay)

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40,) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite_old")))|> filter(age_group != "quite_old") # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarise(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
3. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "old")))) |> filter(!age_group == "old") |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay)) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29") ifelse(between(age, 40, 49), "40-49", "old")) |> select(age, age_group) |> print() # endfold matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(income), .groups = "drop") |> tidyr::pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female) matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29"), ifelse(between(age, 40, 49), "40-49", "quite old")) |> select(age, age_group) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "old"), ifelse(between(age, 40, 49), "40-49", "quite old")) |> filter(age_group != "quite old") |> select(age, age_group) |> print() # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n=n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) |> filter(age_group != "quite old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sec) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "ancient")))) |> filter(age_group %in% ("25-29", "40-49")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(sex, age_group) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "old")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex)|> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age,25,29),"25-29", ifelse(between(age,40,49),"40-49", "quite old")) # endfold)) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age,25,29), "25-29", ifelse(between(age,40,49),"40-49", "quite old")))) |> filter(age_group != "quite old") |> select(age,age_group) |> print() # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40,49)"40-49", "quite old"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "old")))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = if_else((between(age, 25, 29)), "25-29", if_else((between(age, 40, 49)), "40-49", "Old")))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by_(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "old"))) |> select(age_group == "25-29" || age_group == "40-49") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", NA)))) |> filter(!is.na(age_group)) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49","old"))) |> filter(age_group != "old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25,29), "25-29", ifelse(between(age, 40,49), "40-49", "old")))) |> filter(age_group != "old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age,25,29),"25-29","old")) matched_hours_work # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group(if_else(age, (between, 25, 29)), "25-29", if_else(age,(between, 40, 49)), "40-49")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25- 29", ifelse(between(age, 40, 49), "40-49", "quite_old"))) |> filter(age_group != "quite_old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old")))) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = if_else(between(age,25,29), "25-29", ifelse(between(age,40,49),"40-49", "quite old"))) |> filter(age_group != "quite_old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = if_else(between(age, 25, 29), "25-29", if_else(between(age, 40, 89), "40-49", "quite old"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n= n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", filter(between(age, 40, 49), "40-49", "quite old"))) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = if_else(between(age, 25, 29), "25-29", if_else(between(age, 40, 49)), "40-49", "quite old")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = if_else(between(age, 25, 29), "25-29", if_else(between(age, 40, 49), "40-49", "quite old"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "old")) |> select(sex, age_group, pay) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) |> filter(age_group != "quite old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old")))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) |> filter(age_group != "quite old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age,25,29),"25-29","old"))|> select(age, age_group) |> print() # matched_gap_work <- matched_hours_work |> # ____ # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate (age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40,49), "40-49", "old"))) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age,sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group == if_else(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old")))) |> filter(age_group != "quite old") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> print() # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40)|> mutate(age_group = ifelse(between(age,25,29), "25-29", ifelse(between(age,40,49), "40-49", "quite old"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex)|> summarize(n=n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40)|> mutate(age_group = ifelse(between(age,25,29), "25-29", "old"), ifelse(between(age, 40, 49), "40-49", "quite old")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40)|> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex)|> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40)|> mutate(age_group = ifelse(between(age,25,29), "25-29", ifelse(between(age,40,49), "40-49", "quite old")))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(age_group, sex)|> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours >= 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(young = (age >= 25 & age <= 29)) |> mutate(old = (age >= 40 & age <= 49)) |> mutate(age_group = if_else(young, "Young", "Old")) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours >= 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quiet old"))) |> filter(age_group != 'quite_old') |> select(age, age_group) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours_worked == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop") # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female) # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours==40)|> mutate(age_group=ifelse(between(age,25,29),"25-29", filter(between(age,40,49),"40-49") "quite old")) |>filter(age_group !=(quite old)) print # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> # endfold matched_gap_work
2. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = if_else(between(age, 25, 29), "25-29", if_else(between(age, 40, 49), "40-49"), "quite old")) |> filter_age_group != "quite old") |> select(age, age_group) |> print() # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work

Student responses

1. # fold: [Summarize 40-hour workers by age group] matched_hours_work <- software_bachelors |> filter(hours==40)|> mutate(age_group=ifelse(between(age,25,29), "25-29",ifelse(between(age,40,49),"40-49","rest"))) # endfold # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |>group_by(sex,age_group)|> summarize(n=n(),median_pay=median(pay)) # endfold matched_gap_work
2. matched_gap_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") |> select(age_group, sex, median_pay) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = (Male - Female) * 1000) matched_gap_work

Student responses

1. `matched_gap_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay)) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") matched_gap_work <- matched_hours_work |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = 1000 * (Male - Female)) matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") matched_gap_work <- matched_hours_work |> select(-n) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") matched_gap_work <- matched_hours_work |> select(-n) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(age < 30, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") matched_gap_work <- matched_hours_work |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age < 30, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay, na.rm = TRUE), .groups = "drop") matched_gap_work <- matched_hours_work |> pivot_wider(id_cols = age_group, names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female) matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40, (age >= 25 & age <= 29) | (age >= 40 & age <= 49)) |> mutate(age_group = ifelse(age >= 25 & age <= 29, "25-29", "40-49")) |> group_by(age_group, sex) |> summarise(n = n(), median_pay = median(income), .groups = "drop") matched_gap_work <- matched_hours_work |> select(age_group, sex, median_pay) |> tidyr::pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female)`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40, age %in% c(25:29, 40:49)) |> mutate(age_group = ifelse(age < 30, "25-29", "40-49"))
matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(earnings, na.rm = TRUE), .groups = "drop") |> pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> rename(Male = median_pay_Male, Female = median_pay_Female) |> mutate(gap = Male - Female) matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40, age >= 25, age <= 29 | age >= 40, age <= 49) |> mutate(age_group = ifelse(age < 30, "25-29", "40-49")) |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work <- matched_hours_work |> pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female) # endfold matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40, age >= 25, age <= 49) |> filter(age <= 29 | age >= 40) |> mutate(age_group = ifelse(age <= 29, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(pay), .groups = "drop") matched_gap_work <- matched_hours_work |> pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female)`
`matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = if_else(between(age, 25, 29), "25-29", "40-49"))`
`matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize( n = n(), median_pay = median(income), .groups = "drop" ) |> select(age_group, sex, median_pay) |> tidyr::pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female)`
`matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40, between(age, 25, 29) | between(age, 40, 49)) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", "40-49"))`
`matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarize( n = n(), median_pay = median(income), .groups = "drop" ) |> tidyr::pivot_wider( names_from = sex, values_from = c(n, median_pay) ) |> mutate(gap = median_pay_Male - median_pay_Female)`
`matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old")))) |> filter(age_group %in% c("25-29", "40-49")) matched_gap_work <- matched_hours_work |> group_by(age_group, sex) |> summarise(n = n(), median_pay = median(pay), .groups = "drop") |> pivot_wider(names_from = sex, values_from = c(n, median_pay)) |> mutate(gap = median_pay_Male - median_pay_Female) matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old"))) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> group_by(sex, age_group) |> summarize(n = n(), median_pay = median(pay)) # endfold matched_gap_work`

Student responses

1. `matched_hours_work <- software_bachelors |> filter(hours == 40) |> mutate(age_group = ifelse(between(age, 25, 29), "25-29", ifelse(between(age, 40, 49), "40-49", "quite old")))) # fold: [Calculate each age group's pay gap] matched_gap_work <- matched_hours_work |> _____ # endfold matched_gap_work`
2. `matched_hours_work <- software_bachelors |> filter(usual_hours == 40, age %in% 25:29 | age %in% 40:49) |> mutate(age_group = ifelse(age %in% 25:29, "25-29", "40-49")) |> group_by(age_group, sex) |> summarize(n = n(), median_pay = median(income), .groups = "drop") matched_gap_work <- matched_hours_work |> select(age_group, sex, median_pay) |> pivot_wider(names_from = sex, values_from = median_pay) |> mutate(gap = Male - Female)`

Student responses

1. `software_bachelors %>% filter(usual_hours == 40, age %in% 25:29 | age %in% 40:49) %>% mutate(age_group = ifelse(age %in% 25:29, "25-29", "40-49"))
%>% group_by(age_group, sex) %>% summarize(n = n(), median_pay = median(income), .groups = "drop") %>% tidyr::pivot_wider(names_from = sex,
values_from = c(n, median_pay)) %>% mutate(gap = median_pay_Male - median_pay_Female)`

How would you describe these results to a student about to graduate? Use the earnings gaps in both age groups.

0 anonymous responses

No responses were submitted.

How would you write one regression formula that fits both lines? What would its response, age variable, and indicator be?

75 anonymous responses

Unspecified or placeholder formula

18

Provides a noninformative or placeholder regression expression without specifying variables or structure.

Misformatted or incorrect formula usage

11

Provides a formula or function call that mixes arguments or uses incorrect syntax for lm, suggesting confusion about proper formula structure.

Specify interaction model formula

11

Gives an explicit regression formula using an interaction between age and a group indicator (sex) to fit separate lines for each group.

Model the difference between groups

8

Proposes modeling the difference in central tendency between groups (e.g., using difference in median pay) as part of the regression to fit both lines.

Response themes for question 18

Additive model without interaction

7

Suggests a model including age and a group indicator as additive terms (no interaction), implying common slope but different intercepts.

Explicit variable choices for model

7

Specifies which variable is the response, a transformed or centered age predictor, and a binary indicator for sex (female = 1, male = 0).

Include multiple predictors

7

States a need to include more than one predictor (implicitly suggesting age and a group indicator) to capture both lines.

Response themes for question 18

Use grouping or plotting command

6

Recommends splitting or faceting the data or using an indicator conceptually rather than specifying a regression formula (mentions commands like facet or using an indicator to separate lines).

Student responses

1. ?
2. . use interaction and indictaor
3. Add an indicator variable, for example 0 if woman, 1 if man.
4. by creatibg 2 slopes and 2 intecrcpets anfd create a true anf false for it, yuse an interactiion term and an indicator
5. create your regression line, then for each predictor, also include an indicator variable that adds/subtracts from that predictor if that idicator is 1
6. `ggplot(aes(x = age, y = mean_pay, color = sex) + geom_line()`
7. I would regress with sex and job occupuation of software develper,, to predict the payment
8. I would use software as a 0/1 variable, and work hours a numeric one, to fit a regression that reflects all of them
9. Include an indicator variable, and indicator term where one represents an offset between slopes, and one is the offset between intercepts of the two regressions.

Student responses

1. interaction
2. `lm(~)`
3. model difference between group with indicator
4. pay
5. the same as simple linear regression
6. Use an indicator variable to show whether age is in the right age group or not. and then use that if else statement to differentiate between the two age groups
7. Use the *. The indicator is gender, median pay is the response, and age is what is multiplied
8. we would use the code `lm`, and its age variable would be in groups, the response in a decreasing line, and indicator by the differences in median pay.
9. $y = a + bx + cl$

Student responses

1. include 2 indicator variables
2. `lm(median_pay ~ age sex, software_developers)`
3. response would be the yearly pay, age would be how many years they worked
4. use indicator variables to have different intercepts and slopes for men and women.
5. we can use indicator variable to indicate the slope and intercept changes
6. we could regress pay on age_group for each gender
7. We could use `lm` to make a linear model or `geom_line` or `abline`
8. We would include the interaction coefficient
9. We would use an indicator variable that is set to 0 or 1 depending on which line we are thinking about and then work from there

Student responses

1. You would need to use an indicator that gives a value of 0 or 1 depending on if someone is classified as male or female. In this we do an almost interaction between two predictors that we want to use and then can use that to get different intercepts and slopes for the relevant group we care about.
2. You would use indicator variables with the second regression variable being an additional slope on the line but multiplied by the indicator variable.

Student responses

1. `age_model <- lm(pay ~ age0 * female, data = age_data) coef(age_model)`
2. Fit `lm(pay ~ age0 * female, data = age_data)`. The response is pay, the age variable is age0 (age – 25), and the indicator is female
3. I would fit `pay ~ age * female`. The response is pay (mean annual pay), the age variable is age, the indicator is female (1 for women, 0 for men), and the interaction is `age:female`. That interaction captures how the women's slope differs from the men's slope — I expect it to be negative since the gap widened with age in the medians.
4. `lm(pay ~ age * gender, data = software_bachelors)`
5. `lm(pay ~ age * sex, data = software_bachelors)`
6. `pay ~ age * woman` response is pay, age variable is age, indicator is women.
7. $\text{pay} = \beta_0 + \beta_1 \text{age}_0 + \beta_2 \text{female} + \beta_3 \text{age}_0 \text{female} + \epsilon$.
8. Regress pay on age, a female indicator (1 for women, 0 for men), and age times female, so the response is pay, the age variable is age, the indicator is female, and the interaction coefficient shows how much women's slope differs from men's.

Student responses

1. The response would be pay, the age variable would be age, and the indicator would be sex, with an interaction like $\text{pay} \sim \text{age} * \text{sex}$ so men and women can have different slopes.
2. Use a multiple regression with an indicator to show both lines
3. use one linear model with an interaction between age and gender, response is pay and age variable is just age

Student responses

1. $\hat{\text{pay}} = \beta_1 * \text{age} + \beta_2 * \text{sex} + \beta_3 * \text{sex} * \text{age}$ Where sex is an indicator variable.
2. Indicator variables we can assign them values 0 or 1. If we want to see the effect of women or men, we can change the indicator variable to 1. If not, we set it to 0 so we can only look at the effect of the other variable.
3. no idea
4. not sure
5. $\text{pay} = B_0 + B_1(\text{age}) + B_2(\text{sex}) + B_3(\text{age} * \text{sex})$
6. response variable would be pay and indicator variable would be sex.
7. response would be pay, age variable would be age, and indicator variable would be sex which would be 1 or 0 to fit a line for both sexes.
8. You use an interaction coefficient with age, so sex and age are the predictors

Student responses

1. A binary variable for age group for the impact on intercept, then a binary indicator for age group that interacts with year for the impact on the slope of pay.
2. Im not sure honestly
3. Indicator variable for whether it is a female or male. The variable itself would be median_pay and graph in by age
4. `lm(median_pay ~ sex + age, data = software_bachelors)`
5. `lm(pay ~ age + female, software_bachelors)`
6. `lm(pay ~ age + sex)` sex would be the indicator variable
7. response is pay, age is age, indicator is sex.

Student responses

1. I would use pay as the response variable, age minus 25 as the predictor, and female as an indicator variable
2. Im not sure
3. `lm(salary ~ age * woman, data = software_bachelors)`
4. The regression formula would include an x variable and a y variable so both could be included on the line.
5. The response is annual earnings, and the age variable is age in years. With women as the reference group, the indicator is 1 for men and 0 for women.
6. use annual pay as the response variable, age minus 25 as the age variable, and a female indicator equal to 1 for women and 0 for men
7. you would have to separate the slopes and intercepts and use an indicator variable to label women as 1

Student responses

1. Call a linear regression function on pay compared to year and sex, with both as predictors
2. I would have to take in multiple predictors
3. I would select the degree, age, and pay and regress them
4. I would write a regression on age predicting pay with a indicator variable on sex to account for both regression lines in one linear model
5. i would write one regression formula by including multiple predictors or use a placeholder formula
6. using the mean of both
7. would have to do multiple regression. not sure on the rest

Student responses

1. To fit both lines you could use an indicator variable
2. use the lm function
3. We can use an indicator variable for sex
4. you use a command to split into two groups
5. You use a command to split into two groups and then plot individual regression lines from both groups.
6. you use a command to split into two groups like facet

- Fit pay with separate intercepts and age slopes for men and women. - Display the coefficients and identify the coefficient measuring the slope difference.

74 anonymous responses

Other responses

74

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age * female, data = age_data) coef(age_work_model) summary(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age * female, data = age_data) summary(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age * female, data = age_data) coef(age_work_model) summary(# fold: [Allow each group its own intercept and slope] age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age * female, data = age_data) coef(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age * gender, data = age_data) coef(age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age * sex, data = age_data) coef(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age, data = age_data) coef(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age, data = age_data) coef(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age*0, data = age_data) coef(age_work_model) summary(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age*sex, data = age_data) coef(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female , data = age_data) coef(age_work_model) # endfold
summary(age_work_model)
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female , data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
7. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
summary(age_work_model)

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model)
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary (age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model)
7. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) age_work_model |> summary() # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
7. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
summary(age_work_model)
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
7. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) #coefficient
measuring slope difference is # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) summary(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 + female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
7. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0 + female, data = age_data) summary(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model)
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model)
summary(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model) # endfold
6. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model) # endfold
7. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0*female, data = age_data) coef(age_work_model) # endfold

Student responses

1. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ age0+female+age0*female, data = age_data) coef(age_work_model) # endfold
2. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ sex * age, data = age_data) coef(age_work_model) # endfold
3. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~ sex, data = age_data) coef(age_work_model) # endfold
4. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~age0 * female, data = age_data) summary(age_work_model) # endfold
5. # fold: [Allow each group its own intercept and slope] age_work_model <- lm(pay ~sex * age, data = age_data) summary(age_work_model) # endfold
6. age_data <- software_bachelors |> mutate(age0 = age - 25, female = ifelse(sex == "Female", 1, 0)) age_model <- lm(pay ~ age0 * female, data = age_data) coef(age_model) # endfold
7. age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)
8. age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)

Student responses

1. `age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model)` This expands to: `pay ~ age0 + female + age0:female`
2. `age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold`
3. `age_work_model <- lm(pay ~ age0 * female, data = age_data) coef(age_work_model) # endfold`
4. `combined_model <- lm(pay ~ age0 * female, data = software_bachelors) coef(combined_model)`
5. `fit <- lm(pay ~ age0 * female, data = age_data) summary(fit)`
6. `pay_model <- lm(pay ~ age * sex, data = software_bachelors) coef(pay_model)`

An analyst says: "The coefficient on female is nearly zero, so this model shows no earnings gap." Do you agree? Use predicted earnings at ages 25 and 45 in your answer.

0 anonymous responses

No responses were submitted.

Which row tests equal slopes? What does its estimate tell us about the age pattern? What does its p-value tell us?

0 anonymous responses

No responses were submitted.

Do you expect a similar dollar gap among low earners and high earners? What would you look at beyond the median to check your prediction?

78 anonymous responses

High earners larger dollar gap

49

Student predicts high earners will exhibit a larger dollar wage gap (states high earners have higher gap).

Low earners smaller gap

9

Student expects a smaller gap among lower-income people compared to higher-income people.

Ambiguous/unspecified larger gap

5

Student asserts a larger gap but does not specify whether it applies to low or high earners or provide checks beyond the median.

Expect similar gap / same data

4

Student expects the dollar gap to be similar across earners or unchanged because the underlying data are the same.

Response themes for question 22

Check percentiles/means beyond median	3
Student emphasizes comparing percentiles, quartiles, means, ratios, and the full distribution (grouping by wage levels) in addition to the median to assess differences between low and high earners.	
High earners smaller gap	3
Student predicts the dollar gap will be smaller among high earners.	
Unsure / lack of prediction	3
Student expresses uncertainty and does not commit to a prediction or specific checks beyond the median.	
Other responses	2

Student responses

1. Based on our knowledge that the pay gap gets larger the longer you work, I would expect the gap to be smaller around low earners and larger around high earners.
2. bigger gaps
3. bigger. we could look at 75th percentile for high earners and 25th percentile for low earners
4. For higher earners, I expect there to be a higher wage gap as mostly men are likely in the top positions to be making that much money. I believe I would also begin to look at the mean now that outliers are less prevalent when we group the wages together.
5. For higher earners, I expect there to be a higher wage gap as mostly top positions get paid a lot more money.
6. High earners will have a greater dollar gap
7. high earners will have a higher wage gap
8. Higher gap among high earners, since the absolute amounts of pay is higher.
9. I expect a bigger difference among high earners

Student responses

1. I expect a bigger gap among high earners
2. I expect a different dollar gap among low earners and high earners. Beyond the median, I would also look at percentiles and mean.
3. I expect a higher gap among high learners because people can earn really a little without too much difference
4. I expect a higher gap for high earners because theres more males that earn higher wages. I would look at the standard deviation and range of earnings to check my answer.
5. I expect a higher gap for high earners, I would look at the standard deviation to see the variance
6. i expect a higher gap with the high earners, because even if the ratio is the same, the dollar amount changes
7. I expect a higher wage gap for higher earners
8. I expect a larger gap among high earners because they are able to pay more, making the gap bigger.
9. I expect higher gap among high earners. I would look at mean and quantiles other than median to check this

Student responses

1. I expect that gap to be bigger between high earners.
2. I expect the dollar gap to be larger among high earners.
3. I expect the gap to be larger in higher earners. Beyond the median, we can look at the people with the highest and lowest salaries.
4. I expect there to be a higher dollar gap for high earners, because there is much more room to earn more than there is to earn less (assuming at minimum a living wage). The q1 and q3 values could show this.
5. I guess more on high earners, ratio
6. I think at the lower end of the distribution the gap will be small because there isn't much room to decrease past a job's base starting rate, while the higher end of the distribution the gap will be larger
7. I think bigger gap between high earners
8. I think high earners would have a higher pay gap and among low earners the gap tends to be smaller. Beyond the median, I would look at the types of jobs and the degrees each job requires within each earning category.

Student responses

1. I think people make less money has the less earning gap, there is no that much thing to differentiate them, but i think the as the payment increases, the payment gap increases
2. I think that there will be a higher wage gap between those that make more money and it will be a lower gape between those that make less. This is largely due to the fact that high earners are likely later in their career and we saw that the top of our data was male dominated in terms of salary. We should also check to see what percentage is male vs female in these categories
3. I think that there will be a lower wage gap for lower earners and a higher wage gap for higher earners. I would make a graph separated by high and low with lines for men and women and pay.
4. I think the gap is smaller for low earners and way bugger for high earners. I would try to graph and look at differnce between those two groups

Student responses

1. I think the gap would change based on which group we are observing. for example for low earners, i think the gap would be smaller, but for high earners the pay can fluctuate a bit more so we might want to use percentiles to observe that prediction
2. I think there is going to be a larger wage gap because c suite is male dominated
3. I would expect a gap amongst high earners. Women were historically disadvantaged from entering positions of authority, which tend to be better paid. I would look at the breakdown of professions for this.
4. i would expect a general similar gap maybe a slightly higher gap in high earners, i would look at the jobs catergorized by high and low pay
5. I would expect a larger dollar gap among high earners but a larger percentage gap about low earners. To check my prediction, I would make residuals
6. i would expect bigger wage gap for high earning. you would have to group by wage level and compare means/medians of groups
7. i would expect higher earners to have higher gaps since more men are in high earning jobs on average. to check i would group by salary bands

Student responses

1. I would expect the gap to be higher in the high-earning careers.
2. I would expect the pay gap to be larger among high earners, and I would compare different percentiles of men's and women's pay, not just the median.
3. I would expect there to be a high gap between high earners and low earners
4. i would expect there to be a higher wage gap among high earners. I would look at the ratio
5. I would not necessarily expect the same dollar gap for low earners and high earners. I would expect the gap could be larger among higher earners. I would look at other parts of the pay distribution, not just the median — especially the 25th percentile and 75th percentile (and possibly the 90th percentile). Comparing these percentiles by sex would show whether the gap changes between lower- and higher-paid workers.
6. I would predict a larger dollar gap among high earners, I would compare mens and wormenr enaring at 10 50 and 090 percentile
7. I'd expect a larger dollar gap among high earners, because even the same percentage gap represents more dollars at higher salaries.

Student responses

1. no i would not expect a similar dollar gap among low and high earners. incomes tend to fan out, so dollar gap usually widens significantly among high earners
2. No, I would expect a larger dollar gap among high earners and a smaller one among low earners. Pay is more spread out at the top, so even a similar percentage gap becomes more dollars. I would look beyond the median at quartiles/percentiles and the whole distribution for men and women.
3. the higher the earners the bigger the gap
4. there will probably be a higher wage gap as high earners are generally older and we've already seen that the gap increases with age.
5. top 75th percentile vs first 25th percentile. The gap will be larger among top earners.

Student responses

1. I assume the gap will be smaller on lower income people as opposed to higher income people
2. I expect a lower dollar gap among low income earners. This is why I would check the ratio of earnings instead of the raw dollar amount.
3. I would expect the gap among low earners to be closer
4. low earners should have a lower dollar gap.
5. Low earners will have lower gap while high earners will have higher gap. Yes we should look beyond median
6. no, I think the dollar gap among low earners is smaller. I would separate high earners from low earners and check the average difference
7. The gap should be lower for low earners because the pay limits the variance. I would check the IQR or some form of variance
8. the gap should be smaller for lower earners because they haven't been promoted or stratified in any way to justify a gap. i would look at different percentiles in earners
9. The gap should be smaller for lower wage workers group compared to higher wage works. I would make two regression lines showing differences in the two different income groups and then analyze from there

Student responses

1. I would expect a different gap for low and high earners. High earners are exceptions and likely do not follow a formulaic salary path while low wage earners do
2. I would look at p-value and coefficients
3. No I think in high income workers will not have this kind of differicen
4. the gap should be smaller
5. 如果高分位数的差距更大 那就说明收入越高，性别工资差距可能越明显。

Student responses

1. i expect a similar gap
2. I think there will be a similar gap as there was for the median because it's the same data.
3. similar
4. The gap would be about the same, and we would look at the mean and standard deviation as well.

Student responses

1. No. I would filter the pay into two groups, low and high, and compare median pay differences
2. Yes
3. yes,

Student responses

1. I expect the gap among high earners to be smaller
2. The gap is not higher amongsts the largfest earners. I look at variance.
3. The gap should be bigger for high earners because they already have a very strong skill set, so they already get paid a lot so its a bigger dollar amount

Student responses

1. expect smaller gap among low earners than high earners. check ratios within group
2. no I do not expect similar dollar gap
3. unsure

- Calculate p10, p50, and p90 of pay separately for men and women. - Reshape the three percentile columns into rows named percentile and pay. - Plot earnings by percentile, distinguishing men and women. Label the pay units. - Add a code comment stating where the dollar gap is largest, using values from your table.

70 anonymous responses

Other responses

70

Student responses

1. # Find p10, p50, and p90 for each sex pay_percentiles <- software_bachelors |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10), p50 = quantile(pay, 0.50), p90 = quantile(pay, 0.90)) # Turn the percentile columns into rows pay_long <- pay_percentiles |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # Plot the earnings at each percentile pay_long |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Percentile", y = "Pay (\$1,000s)", color = "Sex") # Where is the dollar gap largest? Use values from your table. # Your answer: 90

Student responses

1. # fold: [Calculate p10, p50, and p90 of pay separately for men and women] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10, na.rm = TRUE), p50 = quantile(pay, 0.50, na.rm = TRUE), p90 = quantile(pay, 0.90, na.rm = TRUE)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, fill = sex)) + geom_col(position = "dodge") + labs(title = "Earnings by Percentile for Men and Women", x = "Percentile", y = "Annual Earnings (\$)", fill = "Gender") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: # The dollar gap is largest at the 90th percentile (p90). # For example, if p90 is \$120,000 for men and \$85,000 for women (difference = \$35,000), # compared to smaller absolute differences at p50 and p10.

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] pay_pct <- your_data %>% group_by(gender) %>% summarise(p10 = quantile(payload, 0.10, na.rm = TRUE), p50 = quantile(payload, 0.50, na.rm = TRUE), p90 = quantile(payload, 0.90, na.rm = TRUE)) %>% pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") pay_pct ggplot(pay_pct, aes(x = percentile, y = pay, color = gender, group = gender)) + geom_point(size = 2) + geom_line() + labs(x = "Percentile", y = "Pay (dollars)", color = "Gender") # Where is the dollar gap largest? Use values from your table. # Your answer: its largest when the pay is hgiher, gap is largest in 0.9 percentile

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile" values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geomplot(aes(x=percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile has the biggest gap

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> ____ # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> ____ # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ____ # endfold pay_summary <- software_bachelors |> group_by(sex) |> summarise(p10 = quantile(pay, 0.10, na.rm = TRUE), p50 = quantile(pay, 0.50, na.rm = TRUE), p90 = quantile(pay, 0.90, na.rm = TRUE)) pay_summary pay_percentiles <- pay_summary |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") pay_percentiles pay_percentiles |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_point() + geom_line() + labs(x = "Pay percentile", y = "Annual pay (\$)", color = "Sex") pay_percentiles |> group_by(percentile) |> summarise(gap = max(pay) - min(pay)) |> arrange(desc(gap)) # Where is the dollar gap largest? Use values from your table. # Your answer: # The largest dollar gap is at the percentile with the highest gap # in the table above.

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> ____ # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to="percentile" values="pay")|>print # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ____ # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:
2. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay,0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: # largest at 90th

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) # endfold # fold: [Turn the percentile columns into rows]
percentiles_long_work <- percentiles_work |> summarize(q10 = quantile(pay,0.1), q50 = quantile(pay,0.5), q90 = quantile(pay,0.9)) # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> pivot_longer(p10:p90, values_from = "pay",names_from = "percentile") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> mutate(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9),) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = percentile, values_to = pay) # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarise(p10 = quantile(pay, 0.10, na.rm = TRUE), p50 = quantile(pay, 0.50, na.rm = TRUE), p90 = quantile(pay, 0.90, na.rm = TRUE)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex)) + geom_point() + labs(x = "Percentile", y = "Annual Pay", color = "Sex") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: At p90.

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_point(size = 2) + geom_line() + labs(x = "Percentile", y = "Pay (dollars)") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:yes

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot()+ geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> _____ # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: Largest at 90 percentile

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: The dollar gap is largest at the p90 percentile.

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: the 90th percentile has the biggest gap

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9),) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90 is the largest

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_from = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes (x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .10), p50 = quantile(pay, .50), p90 = quantile(pay, .90)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x= percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex)) + geom_point() # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: It is the largest at the 90th percentile

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x=percentile, y=pay, color=sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90 percentile has the biggest gap

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile , y= pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: the dollar gap is the largest for 90th percentile of 50

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10, na.rm = TRUE), p50 = quantile(pay, 0.50, na.rm = TRUE), p90 = quantile(pay, 0.90, na.rm = TRUE)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Percentile", y = "Pay (\$)") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: The dollar gap is the largest at percentile 90%

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10), p50 = quantile(pay, 0.50), p90 = quantile(pay, 0.90)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Percentile", y = "Pay (\$)", color = "Sex") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: The dollar gap is largest at p90 because

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, 0.10), p50 = quantile(payload, 0.50), p90 = quantile(payload, 0.90), .groups = "drop") # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Percentile", y = "Pay (dollars)", color = "Sex") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: The dollar gap is largest at the percentile where the difference between the male and female pay values in the table is greatest.

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10), p50 = quantile(pay, 0.50), p90 = quantile(pay, 0.90), .groups = "drop") # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Percentile", y = "Pay (\$)", color = "Sex") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(vlaue_from = p10:p90, names_from = sex values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(q10 = quantile(pay, .1), q50 = quantile(pay, .5), q90 = quantile(pay, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = q10:q90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: the gap is largest in the 90th percentile

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x=percentile, y=pay, color= sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile has the biggest gap

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile has the biggest gap

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile has the biggest gap, going from 203 to 254

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarise(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: the dollar gap is the largest in the 90% percentile

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .3), p90 = quantile(pay, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to="percentile" values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile has largest gap

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y =pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9),) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer (p10:p90, names_to = "percentile", values_to = "pay") |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10, na.rm=True), p50 = quantile(pay, 0.50, na.rm=True), p90 = quantile(pay, 0.90, na.rm=True)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(q10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(q10 = quantile(pay, 0.1), q50 = quantile(pay, 0.5), q90 = quantile(pay, 0.9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(q10:q90) |> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() +geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold) # Where is the dollar gap largest? Use values from your table. # Your answer: 90th percentile has the largest at about 50k per year.

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .10), p50 = quantile(payload, .50), p90 = quantile(payload, .90)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x=percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(p10 = quantile(pay, .1), p50 = quantile (pay, .5), p90 = quantile (pay, .9))|> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, values_to = "pay", names_to = "percentile")|> print() # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color =sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90 is the largest

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(q10 = quantile(pay, .1), q50 = quantile(pay, .5), q90 = quantile(pay, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(values_from = p10:p90, names_from = sex) # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> _____ # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(q10 = quantile(pay, .1), q50 = quantile(pay, .5), q90 = quantile(pay, .9)) # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(q10:q90, values_from = "percentile", names_from = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex))+ geom_point() # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(q10 = quantile(pay, 0.1), q50 = quantile(pay, 0.5), q90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols= c(q10, q50, q90), names_to="percentile", values_to="pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex,)) + geom_point() # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: q90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarise(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: At p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(q10 = quantile(pay, 0.10), q50 = quantile(pay, 0.50), q90 = quantile(pay, 0.90)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90) # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> _____ # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot()+geom_point(aes(x=percentile, y=pay, color=sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, .1), p50 = quantile(payload, .5), p90 = quantile(payload, .9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90, values_to = "pay", names_to = "percentile") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90 is large

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(p10 = quantile(pay, .1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(names_to = "percentile", values_to = "pay") # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x=percentile, y = pay, color = sex)) + geom_line() # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex)|> summarize(p10=quantile(pay, 0.1), p50=quantile(pay, 0.5), p90=quantile(pay, 0.9)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(p10:p90,names_to="percentile",values_to="pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot()+ geom_line(aes(x=percentile,y=pay,color=sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: p90

Student responses

1. # fold: [Find p10, p50, and p90 for each sex] percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(values_from = p10:p90) # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot() + geom_point(aes(x = percentile, y = pay, color = sex)) # endfold # Where is the dollar gap largest? Use values from your table. # Your answer:

Student responses

1.

```
pay_percentiles <- wages %>% group_by(sex) %>% summarize( p10 = quantile(pay, 0.10, na.rm = TRUE), p50 = quantile(pay, 0.50, na.rm = TRUE), p90 = quantile(pay, 0.90, na.rm = TRUE) ) %>% pivot_longer( cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay" ) pay_percentiles # Replace the numbers below with the values from your table. # The dollar gap is largest at p90 because men's pay minus women's pay is biggest there. ggplot(pay_percentiles, aes(x = percentile, y = pay, color = sex, group = sex)) + geom_point(size = 2) + geom_line() + labs( x = "Percentile", y = "Pay ($)", color = "Sex" )
```

Student responses

1. `percentile_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") percentile_work |> ggplot(aes(x = percentile, y = pay, color = sex)) + geom_point(size = 3) + labs(x = "Earnings percentile", y = "Pay ($1,000s)", color = "Sex") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: # The dollar gap is largest at the 90th percentile.`
2. `percentiles <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10), p50 = quantile(pay, 0.50), p90 = quantile(pay, 0.90)) |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") percentiles |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_point() + geom_line() + labs(x = "Percentile", y = "Pay ($1,000s)", color = "Sex") # The dollar gap is largest at the 90th percentile.`

Student responses

1.

```
percentiles_work <- acs |> group_by(sex) |> summarize( p10 = quantile(earnings, 0.10, na.rm = TRUE), p50 = quantile(earnings, 0.50, na.rm = TRUE), p90 = quantile(earnings, 0.90, na.rm = TRUE), .groups = "drop" ) percentiles_long_work <- percentiles_work |> pivot_longer( cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay" ) percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Earnings percentile", y = "Annual earnings ($)", color = "Sex") largest_gap <- percentiles_long_work |> pivot_wider(names_from = sex, values_from = pay) |> mutate(gap = Male - Female) |> slice_max(gap, n = 1, with_ties = FALSE) cat(sprintf( "# Largest gap at %s: men $%.0f, women $%.0f; gap $%.0f\n", largest_gap$percentile, largest_gap$Male, largest_gap$Female, largest_gap$gap ))
```


Student responses

1. `percentiles_work <- acs |> group_by(sex) |> summarize( p10 = quantile(pay, .1), p50 = quantile(pay, .5), p90 = quantile(pay, .9) )`
`percentiles_long_work <- percentiles_work |> pivot_longer( p10:p90, names_to = "percentile", values_to = "pay" )` # fold: [Turn the percentile columns into rows]
`percentiles_long_work <- percentiles_work |> pivot_longer( p10:p90, names_to = "percentile", values_to = "pay" )` # endfold
fold: [Plot the earnings at each percentile]
`percentiles_long_work |> ggplot() + geom_point(aes( x = percentile, y = pay, color = sex))` # endfold
Where is the dollar gap largest? Use values from your table.
Your answer: p90

Student responses

1.

```
percentiles_work <- acs |> group_by(sex) |> summarize( p10 = quantile(pay, 0.1, na.rm = TRUE), p50 = quantile(pay, 0.5, na.rm = TRUE), p90 = quantile(pay, 0.9, na.rm = TRUE) ) # Turn the percentile columns into rows
percentiles_long_work <- percentiles_work |> pivot_longer( p10:p90, names_to = "percentile", values_to = "pay" ) # Plot the earnings at each percentile
percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_point() + geom_line() + labs(y = "Pay (dollars)", x = "Percentile")
```

Student responses

1.

```
percentiles_work <- acs |> group_by(sex) |> summarize( p10 = quantile(pay, 0.1, na.rm = TRUE), p50 = quantile(pay, 0.5, na.rm = TRUE), p90 = quantile(pay, 0.9, na.rm = TRUE), .groups = "drop" ) percentiles_long_work <- percentiles_work |> tidyr::pivot_longer( cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay" ) percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex)) + geom_point(size = 3) + labs( x = "Earnings percentile", y = "Annual earnings (thousands of dollars)", color = "Sex" ) # endfold # Where is the dollar gap largest? Use values from your table. # the largest gap is 50 men earn 250 thousand and women earn 200 thousand
```

Student responses

1. `percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9)) |> print() # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold # fold: [Plot the earnings at each percentile] percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(y = "Pay (dollars)", x = "Percentile") # endfold # Where is the dollar gap largest? Use values from your table. # Compare men and women at p10, p50, and p90 in percentiles_work; the largest difference is the largest dollar gap.`

Student responses

1.

```
percentiles_work <- acs |> group_by(sex) |> summarize( p10 = quantile(pay, 0.1), p50 = quantile(pay, 0.5), p90 = quantile(pay, 0.9), .groups = "drop" )
percentiles_long_work <- percentiles_work |> pivot_longer( cols = p10:p90, names_to = "percentile", values_to = "pay" ) percentiles_long_work |> ggplot(aes(x
= percentile, y = pay, color = sex)) + geom_point() + labs( x = "Earnings percentile", y = "Pay ($1,000s)", color = "Sex" ) # Your answer: # The dollar gap is largest
at the 90th percentile.
```

Student responses

1. `percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, 0.10, na.rm = TRUE), p50 = quantile(pay, 0.50, na.rm = TRUE), p90 = quantile(pay, 0.90, na.rm = TRUE)) percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_point(size = 2) + geom_line() + labs(x = "Percentile", y = "Pay (dollars)")` # Where is the dollar gap largest? Use values from your table. # Your answer: The dollar gap is largest at p90.

Student responses

1. `percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(pay, probs = 0.10, na.rm = TRUE), p50 = quantile(pay, probs = 0.50, na.rm = TRUE), p90 = quantile(pay, probs = 0.90, na.rm = TRUE)) # endfold # fold: [Turn the percentile columns into rows] percentiles_long_work <- percentiles_work |> pivot_longer(cols = c(p10, p50, p90), names_to = "percentile", values_to = "pay") # endfold percentiles_long_work |> ggplot(aes(x = percentile, y = pay, color = sex, group = sex)) + geom_line() + geom_point() + labs(x = "Percentile", y = "Pay ($)", color = "Sex") # endfold # Where is the dollar gap largest? Use values from your table. # Your answer: It's biggest for p90`

Student responses

1. `percentiles_work <- acs |> group_by(sex) |> summarize(p10 = quantile(payload, probs = 0.1), p50 = quantile(payload, probs = 0.5), p90 = quantile(payload, probs = 0.9), .groups = "drop")`
`percentiles_long_work <- percentiles_work |> pivot_longer(c(p10, p50, p90), names_to = "percentile", values_to = "pay")`
`percentiles_long_work |> ggplot(aes(percentile, pay, color = sex)) + geom_point() + labs(x = "Percentile", y = "Annual pay ($1,000s)", color = "Sex")` # Where is the dollar gap largest? Use values from your table. # Your answer # The gap is largest at the 90th percentile, about \$51k # (vs. about \$20k at the median and \$5k at the 10th percentile).

**A report describes the earnings gap as \$20,000.
Write a more informative one-sentence summary
using this plot.**

0 anonymous responses

No responses were submitted.

We also know each worker's degree, usual hours, and whether they work in the public or private sector. Which would you investigate first, and what would you calculate or plot?

0 anonymous responses

No responses were submitted.

- Calculate the conditional degree distribution within each sex. - Report the fraction with master's degrees for women and for men.

0 anonymous responses

No responses were submitted.

Use the full `acs` table to make age plots separately for bachelor's and master's graduates. What needs to change in the grouping and plot?

0 anonymous responses

No responses were submitted.

- Calculate hourly_pay from pay, hours, and weeks, with pay converted to dollars. - Report median weekly hours and hourly earnings by sex.

0 anonymous responses

No responses were submitted.

**Do these results support working hours as the main explanation for the annual earnings gap?
Cite a number in your answer.**

0 anonymous responses

No responses were submitted.

What would you change in the overall summary to make this table? What plot would you use to show the four medians?

0 anonymous responses

No responses were submitted.

Calculate median pay for men and women within each sector. Make and submit a plot showing all four medians. Which sector has the larger dollar gap?

0 anonymous responses

No responses were submitted.

Suggest one further analysis that would help explain the smaller public-sector gap. Use a variable we have and describe the table or plot you would make.

0 anonymous responses

No responses were submitted.

- Calculate median_pay for each occupation and sex. - Put Male and Female median_pay on the same row for each occupation. - Calculate gap as Male minus Female and sort from largest to smallest.

0 anonymous responses

No responses were submitted.

Write one sentence about the historical trend, one about the same-job age pattern, and one about the earnings distribution. Include a number and suggest one next investigation.

0 anonymous responses

No responses were submitted.