Discussion 2-3

Ying Lin Zhao

Code

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(haven)
library(scales)
```

```
Attaching package: 'scales'

The following object is masked from 'package:purrr':

    discard

The following object is masked from 'package:readr':

    col_factor
```

```r
micro <- read_dta("cps_00001.dta")

filtered <- micro |>
  # Restrict to people working more than 50 weeks per year
  # Based on documentation, 6 is coded for those who worked
  # between 50-52 weeks
  filter(wkswork2 == 6) |>
  # Restrict to people earning between 0 and 99999998
  filter(incwage > 0 & incwage < 99999998)

summarized <- filtered |>
  group_by(year) |>
  summarise(
    p10 = Hmisc::wtd.quantile(x = incwage, weights = asecwt, probs = 0.1),
    p50 = Hmisc::wtd.quantile(x = incwage, weights = asecwt, probs = 0.5),
    p90 = Hmisc::wtd.quantile(x = incwage, weights = asecwt, probs = 0.9)
  )

pivoted <- summarized |>
  pivot_longer(
    cols = c("p10", "p50", "p90"),
    names_to = "quantity",
    values_to = "income"
  )

inflation <- read_csv("https://info3370.github.io/data/inflation.csv")
```

```
Rows: 77 Columns: 2
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (2): year, inflation_factor

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
joined <- pivoted |>
  left_join(inflation, by = join_by(year))

adjusted_data <-
  joined |>
  mutate(income = income * inflation_factor)
```

```r
adjusted_data |>
  ggplot(
    mapping = aes(x = year, y = income, color = quantity)
  ) +
  labs(
    title = "Inequality of Wage Income from 1962 to 2022",
    subtitle = "By the 10th, 50th, and 90th percentiles",
    x = "Year"
  ) +
  scale_y_continuous(
    name = "Income \n (Adjusted to Inflation Based on 2023       Dollars)",
    labels = label_dollar()
  ) +
  scale_color_discrete(
    name = "Percentile of\nDistribution",
    labels = c("10th", "50th", "90th")
  ) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
  )+
  geom_line()
```