Extra credit

INFO 2950 - Spring 2023

Lina Liu

5/10/23

Setup

Load packages and data:

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- 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(usmap)
library(ggplot2)

#| label: load-data
tuesdata <- tidytuesdayR::tt_load('2023-05-09')
```

```
--- Compiling #TidyTuesday Information for 2023-05-09 ----
--- There are 2 files available ---
--- Starting Download ---
```

```
    Downloading file 1 of 2: `childcare_costs.csv`
    Downloading file 2 of 2: `counties.csv`

--- Download complete ---
```

```
tuesdata <- tidytuesdayR::tt_load(2023, week = 19)
```

```
--- Compiling #TidyTuesday Information for 2023-05-09 ----
--- There are 2 files available ---
--- Starting Download ---


    Downloading file 1 of 2: `childcare_costs.csv`
    Downloading file 2 of 2: `counties.csv`

--- Download complete ---
```

```
childcare_costs <- tuesdata$childcare_costs
counties <- tuesdata$counties
```

Extra credit

Research Question: What is the labor force participation of mothers who have children throughout the United States in the year 2018? How does this compare to 2008?

```
labor_part_2018 <- childcare_costs |>
  # select relevant columns to use for analysis
  select(county_fips_code, study_year, flfpr_20to64_under6, flfpr_20to64_6to17) |>
  # filter so we only use the year 2018
  filter(study_year == "2018") |>
  # did this justttt in case if there's n/a values
  drop_na() |>
  # combine the column for labor participation of mothers with children under the age
  # of 6 alongside the labor participation of mothers with children 6 to 17.
  pivot_longer(
    cols = -c("county_fips_code", "study_year"),
```

```
    names_transform = parse_number,
    values_to = "pct"
  ) |>
  # get rid of the name column, which is kindaaa unnecessary :p
  select(-name) |>
  # because there's two rows of the same area (one for the labor participation
  # percentage of mothers with children under 6 age group and the other one for
  # the 6-17 age group), I decided to average them together to get the total
  # labor participation
  group_by(county_fips_code) |>
  summarize(pct = mean(pct))
```

```
# I followed this tutorial for the US map:
# https://jtr13.github.io/cc19/different-ways-of-plotting-u-s-map-in-r.html

plot_usmap(regions = "counties")  +
  labs(title = "Labor Force Participation of Mothers throughout the United States",
       subtitle = "For the year 2018",
       caption = "Source: National Database of Childcare Prices") +
  theme(panel.background = element_blank())
```