WK1: Loading Data Into a Data Frame

Dirk Hartog

2023-09-03

OVERVIEW

The article and the subsequent data that I choose to work with is titled "Congress Today Is Older Than It's Ever Been: OK, boomer? More like boomer, OK!" publised on the FiveThirtyEight.com. The articl can be found here

The article describes some basic statistics, identifies and visualizes trends in the ages of the memebers of The House of Representatives and the Senate in the US starting with the 66th Congress (1919 - 1921) to the 118th Congress (2023-2025).

```
# loading in the necessary libraries
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   3.4.3     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 become errors
```

Reading in the data via link to the raw data on github

```
congress_data <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/congress-demogr

head(congress_data)
```

```
##   congress start_date chamber state_abbrev party_code                bioname
## 1       82 1951-01-03   House           ND        200    AANDAHL, Fred George
## 2       80 1947-01-03   House           VA        100 ABBITT, Watkins Moorman
## 3       81 1949-01-03   House           VA        100 ABBITT, Watkins Moorman
## 4       82 1951-01-03   House           VA        100 ABBITT, Watkins Moorman
## 5       83 1953-01-03   House           VA        100 ABBITT, Watkins Moorman
## 6       84 1955-01-03   House           VA        100 ABBITT, Watkins Moorman
##   bioguide_id   birthday cmltv_cong cmltv_chamber age_days age_years generation
## 1     A000001 1897-04-09          1             1    19626  53.73306       Lost
## 2     A000002 1908-05-21          1             1    14106  38.62012   Greatest
## 3     A000002 1908-05-21          2             2    14837  40.62149   Greatest
```

```
## 4     A000002 1908-05-21          3             3    15567  42.62012   Greatest
## 5     A000002 1908-05-21          4             4    16298  44.62149   Greatest
## 6     A000002 1908-05-21          5             5    17028  46.62012   Greatest
```

Subsetting the data

I wanted to filter the data to include the parties, chamber of congress, and age years in each of the Congressional periods for the last 20 years. I used age_years to represent the age of the member of congress.

- congress: The number of the Congress that this member's row refers to
- start_date: First day of a Congress
- age_years: In the data age_years was calculated first by calculating age_days: start_date minus birthday. Then taking age_days and dividing by 365.25
- chamber: The chamber a member of Congress sat in: Senate or House
- party_code: A code that indicates a member's party
- generation: Generation the member belonged to, based on the year of birth

The code below is how I filtered and renamed the columns.

```
congress_sub <- congress_data %>% filter(start_date >= "2003-01-03") %>%
  select(congress, start_date, age_years, chamber, party_code, generation)

names(congress_sub) <- c("congress_served", "start_date", "age_years", "chamber", "party", "generation")

dim(congress_sub)
```

```
## [1] 6002    6
```

Changing the data types for easier manipulation

- Converting the columns *chamber, generation* to *factor* data type
- Converting *party* to a *character* data type
- Converting *start_date* to *date* data type

```
congress_sub$chamber <- as.factor(congress_sub$chamber)
congress_sub$generation <- as.factor(congress_sub$generation)
congress_sub$party <- as.character(congress_sub$party)
congress_sub$start_date <- as.Date(congress_sub$start_date, tryFormats = "%Y-%m-%d")

glimpse(congress_sub)
```

```
## Rows: 6,002
## Columns: 6
## $ congress_served <int> 108, 109, 110, 111, 108, 109, 110, 111, 112, 108, 109,~
## $ start_date      <date> 2003-01-03, 2005-01-03, 2007-01-03, 2009-01-03, 2003-~
## $ age_years       <dbl> 64.52293, 66.52430, 68.52293, 70.52430, 60.12320, 62.1~
## $ chamber         <fct> House, House, House, House, House, House, House, House~
## $ party           <chr> "100", "100", "100", "100", "100", "100", "100", "100"~
## $ generation      <fct> Silent, Silent, Silent, Silent, Silent, Silent, Silent~
```

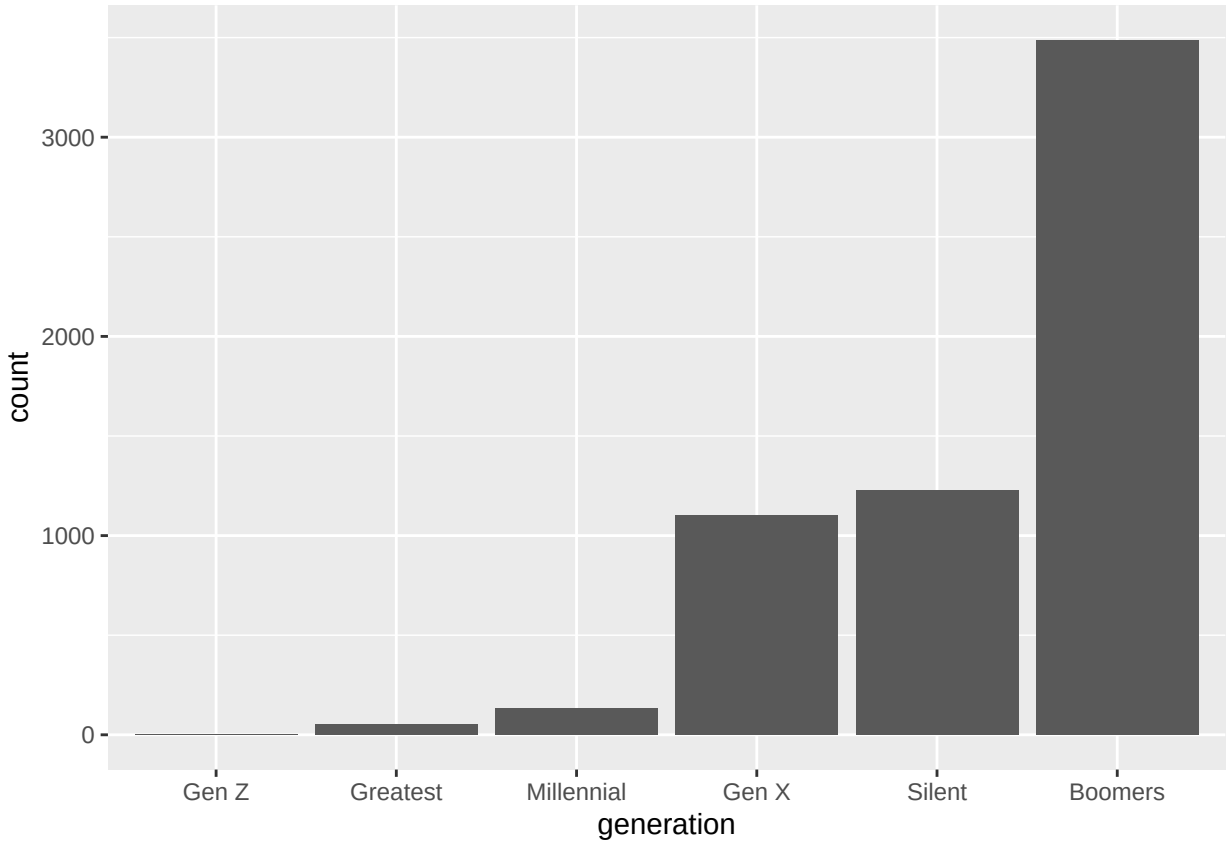
Replace the party code with the actual name of the party

```
congress_sub$party[congress_sub$party == "100"] <- "Democrat"
congress_sub$party[congress_sub$party == "200"] <- "Republican"
congress_sub$party[congress_sub$party =="328"] <- "Inependent"
```

Exploratory visualizations

1. Bar graph: Visualizing the total amount each generation has been represented in congress over the past 20 years.

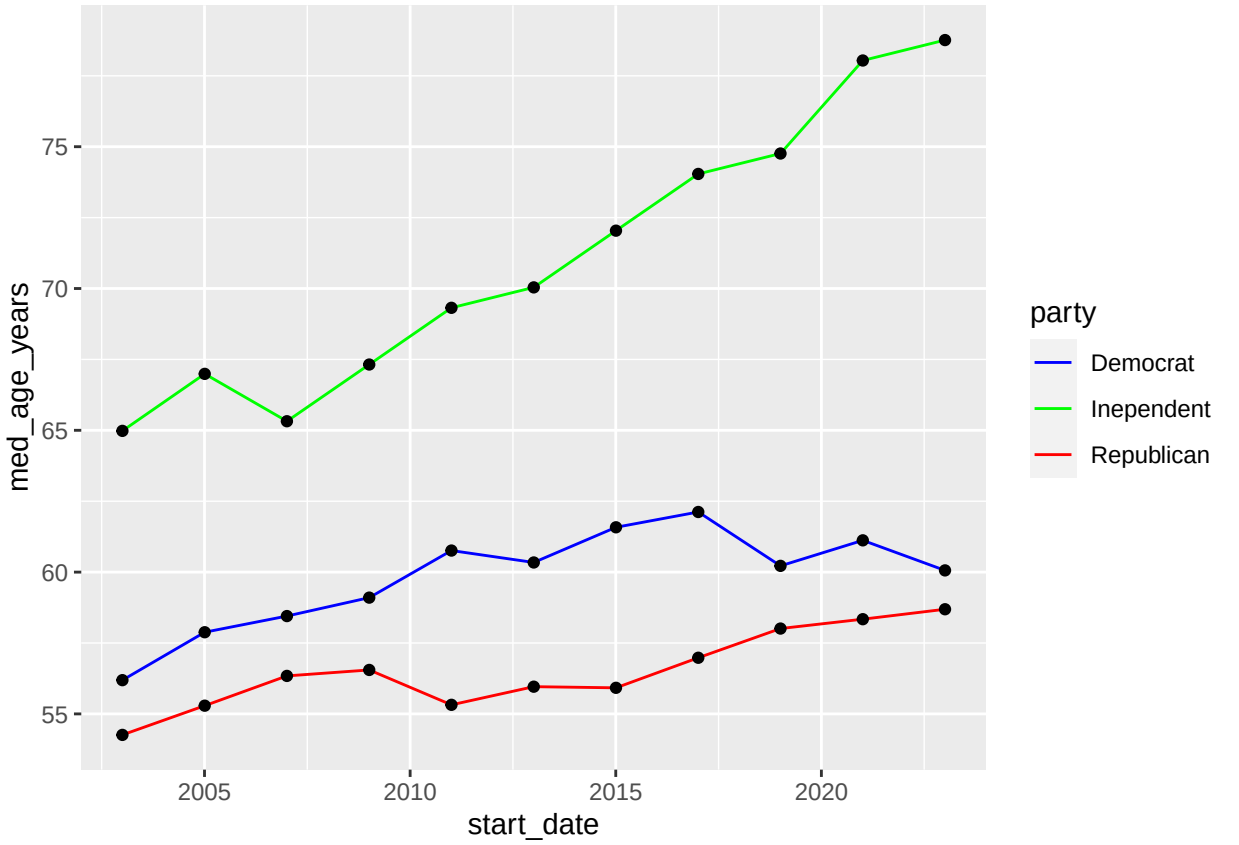
```
congress_sub %>%
  mutate(generation = generation %>% fct_infreq() %>% fct_rev()) %>%
  ggplot(aes(x = generation)) +
  geom_bar()
```



2. Line plot: Plotting trends of the median age in each party at the start of each congress from 2003 - 2023

```
plot2 <- congress_sub %>% group_by(start_date, party) %>% summarize(med_age_years = round(median(age_ye
```

```
ggplot(data = plot2, mapping = aes(x = start_date, y = med_age_years)) +
  geom_line(aes(color = party)) +
  geom_point() +
  scale_color_manual(values=c('Blue','Green', 'Red'))
```



CONCLUSIONS

In the article it mentioned several reasons for the increased age of the US Congress. Although not an exhautive list some of these reasons include

- The US population has also gotten older
  - Increased life expectancy
  - Fewer births leading to less younger people
- Members of congress are more likely to win re-election for a second term
  - Older individuals tend to vote for candidates similar to them
  - Older members of Congress tend to focus on issues that impact older individuals
- Baby Boomers represent the majority of Congress members today (48%)

Other areas that would be interesting to visualize are the bills introduced by the memebers of congress and the age of their coinstiuents.