

# Visualization Project: Both Republicans And Democrats Have an Age Problem

Ethan Rubenstein

## Background

The purpose of this report is to replicate the graph entitled *Average Age of New Members of Congress*, which was featured in Nate Silver's article *Both Republicans And Democrats Have an Age Problem*. The article, published in 2014, made note that the average age of congressional lawmakers was at a record height of nearly 60 years old. The dataset Silver uses draws its data from three sources: GovTrack.us, the Sunlight Foundation, and the New York Times' Congress API. It encompasses the general biographical information of Congress members from 1947 to 2013, such as party affiliation, name, birthday, and age. The data graphic that will be recreated in this report seeks to represent the major trends in the average age of Congress members by only looking at new Congress members, as incumbent members are much more likely to be older, given that they have served multiple terms. The graphic also breaks down the trend between party, showing how their mean Congressional ages have changed relative to each other.

## Data Wrangling

```
library(tidyverse)
# The zoo library will be used to calculate moving averages using the rollmean() function
require(zoo)
# Read in dataset
congress = read.csv("congress-terms.csv")
# Preview data
head(congress)
```

```
##   congress chamber bioguide  firstname  middlename  lastname  suffix  birthday
## 1      80      house M000112   Joseph    Jefferson  Mansfield      1861-02-09
## 2      80      house D000448   Robert     Lee        Doughton      1863-11-07
## 3      80      house S000001   Adolph    Joachim    Sabath       1866-04-04
## 4      80      house E000023   Charles   Aubrey     Eaton       1868-03-29
## 5      80      house L000296   William   Lewis     Lewis       1868-09-22
## 6      80      house G000017   James     A.         Gallagher   1869-01-16
##   state party incumbent termstart  age
## 1    TX      D          Yes 1947-01-03 85.9
## 2    NC      D          Yes 1947-01-03 83.2
## 3    IL      D          Yes 1947-01-03 80.7
## 4    NJ      R          Yes 1947-01-03 78.8
## 5    KY      R          No  1947-01-03 78.3
## 6    PA      R          No  1947-01-03 78.0
```

Since we want to visualize the average age of new members of Congress, let's filter out all incumbent members:

```
new_congress = congress %>%
  filter(incumbent == "No")
head(new_congress,3)
```

```
##   congress chamber bioguide  firstname  middlename  lastname  suffix  birthday
## 1      80      house L000296   William   Lewis     Lewis       1868-09-22
## 2      80      house G000017   James     A.         Gallagher   1869-01-16
## 3      80      house C000590   Howard    Aldridge   Coffin       1877-06-11
##   state party incumbent termstart  age
## 1    KY      R          No  1947-01-03 78.3
## 2    PA      R          No  1947-01-03 78.0
## 3    MI      R          No  1947-01-03 69.6
```

## Republican dataframe

Now that the dataframe consists only of new members of Congress, we must split the data into Republicans and Democrats. To do this, we will make two different dataframes. For now, though, let's focus on the Republican dataframe:

```
new_republicans = new_congress %>%
  filter(party == "R")
```

```
##   congress chamber bioguide  firstname  middlename  lastname  suffix  birthday
## 1      80      house L000296   William   Lewis     Lewis       1868-09-22
## 2      80      house G000017   James     A.         Gallagher   1869-01-16
## 3      80      house C000590   Howard    Aldridge   Coffin       1877-06-11
##   state party incumbent termstart  age
## 1    KY      R          No  1947-01-03 78.3
## 2    PA      R          No  1947-01-03 78.0
## 3    MI      R          No  1947-01-03 69.6
```

```
## # A tibble: 3 x 2
##   `new_republicans$congress` avg_age
##   <int>         <dbl>
## 1             80         47.3
## 2             81         50.1
## 3             82         50.0
```

The original figure utilized two-term rolling averages instead of just the average age of new members per Congressional session. The `rollmeanr()` function from the `zoo` package will be used to calculate two-term rolling mean for new Republicans:

```
new_republicans$rolling_avg_age = rollmeanr(new_republicans$avg_age, 2, fill=NA)
head(new_republicans,3)
```

```
## # A tibble: 3 x 3
##   `new_republicans$congress` avg_age rolling_avg_age
##   <int>         <dbl>         <dbl>
## 1             80         47.3             NA
## 2             81         50.1             48.7
## 3             82         50.0             50.1
```

Note that the original figure omits the 80th Congress (inaugurated 1947), as it uses two-term rolling averages. This code does the same. The rolling average is the mean of the current and previous Congress' average age. Therefore, `rolling_avg_age` must lag by 1, meaning the 80th Congress is not directly included in the final figure, but still effects the two-term rolling average age of the 81st Congress (inaugurated 1949).

In the previous step, the `group_by()` function reduced the number of variables, most of which are irrelevant for producing the original figure. However, we still need the inauguration date of each Congress. Luckily, we can do this easily by creating a new variable, called `inaugurated`, and assign it a sequence. The sequence will start at 1947, the year of the 80 Congress' inauguration, to 2013, the 113th Congress' inauguration, and increment by two, the length of one Congressional term.

```
new_republicans$inaugurated = seq(1947, 2013, by = 2)
head(new_republicans,3)
```

```
## # A tibble: 3 x 4
##   `new_republicans$congress` avg_age rolling_avg_age inaugurated
##   <int>         <dbl>         <dbl>         <dbl>
## 1             80         47.3             NA         1947
## 2             81         50.1             48.7         1949
## 3             82         50.0             50.1         1951
```

With the two-term rolling average age and the inauguration date of each Congress, we now have the data that we need to replicate the Republican trendline of the graph.

## Democrat dataframe

Now, we will simply rerun the above code for new Democratic members of Congress, just changing the dataframe reference from `new_republicans` to `new_democrats`.

```
new_democrats = new_congress %>%
  filter(party == "D")
```

```
##   congress chamber bioguide  firstname  middlename  lastname  suffix  birthday
## 1      80      house L000296   William   Lewis     Lewis       1868-09-22
## 2      80      house G000017   James     A.         Gallagher   1869-01-16
## 3      80      house C000590   Howard    Aldridge   Coffin       1877-06-11
##   state party incumbent termstart  age
## 1    KY      R          No  1947-01-03 78.3
## 2    PA      R          No  1947-01-03 78.0
## 3    MI      R          No  1947-01-03 69.6
```

```
## # A tibble: 3 x 2
##   `new_democrats$congress` avg_age
##   <int>         <dbl>
## 1             80         42.5
## 2             81         48.0
## 3             82         46.8
```

The original figure utilized two-term rolling averages instead of just the average age of new members per Congressional session. The `rollmeanr()` function from the `zoo` package will be used to calculate two-term rolling mean for new Democrats:

```
new_democrats$rolling_avg_age = rollmeanr(new_democrats$avg_age, 2, fill=NA)
head(new_democrats,3)
```

```
## # A tibble: 3 x 3
##   `new_democrats$congress` avg_age rolling_avg_age
##   <int>         <dbl>         <dbl>
## 1             80         42.5             NA
## 2             81         48.0             45.3
## 3             82         46.8             47.4
```

Note that the original figure omits the 80th Congress (inaugurated 1947), as it uses two-term rolling averages. This code does the same. The rolling average is the mean of the current and previous Congress' average age. Therefore, `rolling_avg_age` must lag by 1, meaning the 80th Congress is not directly included in the final figure, but still effects the two-term rolling average age of the 81st Congress (inaugurated 1949).

In the previous step, the `group_by()` function reduced the number of variables, most of which are irrelevant for producing the original figure. However, we still need the inauguration date of each Congress. Luckily, we can do this easily by creating a new variable, called `inaugurated`, and assign it a sequence. The sequence will start at 1947, the year of the 80 Congress' inauguration, to 2013, the 113th Congress' inauguration, and increment by two, the length of one Congressional term.

```
new_democrats$inaugurated = seq(1947, 2013, by = 2)
head(new_democrats,3)
```

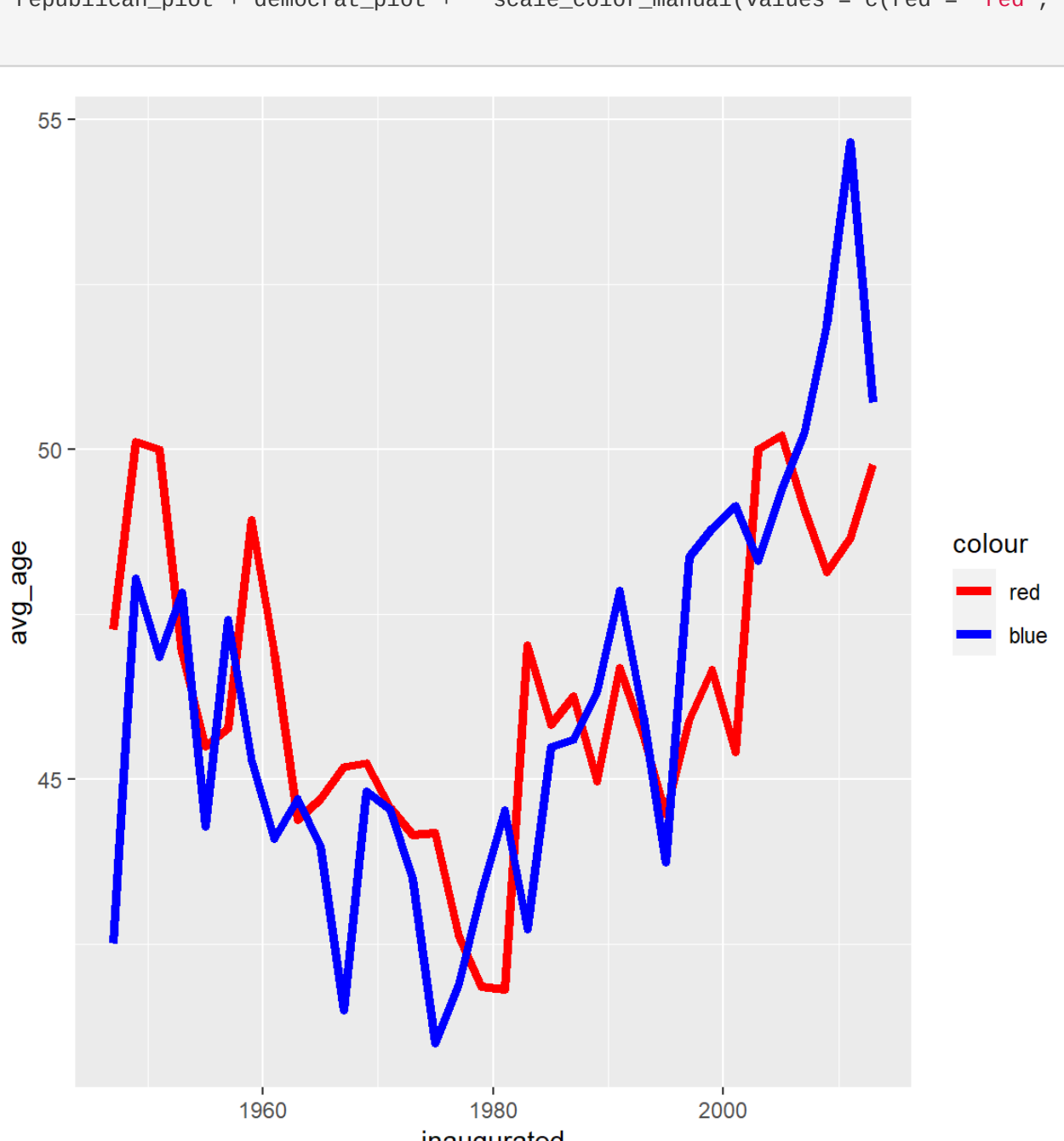
```
## # A tibble: 3 x 4
##   `new_democrats$congress` avg_age rolling_avg_age inaugurated
##   <int>         <dbl>         <dbl>         <dbl>
## 1             80         42.5             NA         1947
## 2             81         48.0             45.3         1949
## 3             82         46.8             47.4         1951
```

With the two-term rolling average age and the inauguration date of each Congress, we now have the data that we need to replicate both the Republican and Democrat trendlines of the original graph.

## Data Visualization

To start recreating the graph, we will create two line charts using `geom_line()`. One line chart will be the `republican_plot` which will contain a line chart of the Republican trendline while the other plot will be the `democrat_plot`. We can then create a `ggplot()` object and add both plots to it, so that both charts are overlaid on the same figure. We will also set the color values and line thickness at this point with `scale_color_manual()` and the `size` parameter.

```
republican_plot = geom_line(data = new_republicans,
  aes(x = inaugurated, y = avg_age, color = "red"),
  size = 1.7)
democrat_plot = geom_line(data = new_democrats,
  aes(x = inaugurated, y = avg_age, color = "blue"),
  size = 1.7)
plot = ggplot() + republican_plot + democrat_plot + scale_color_manual(values = c(red = "red", blue = "blue"))
plot
```

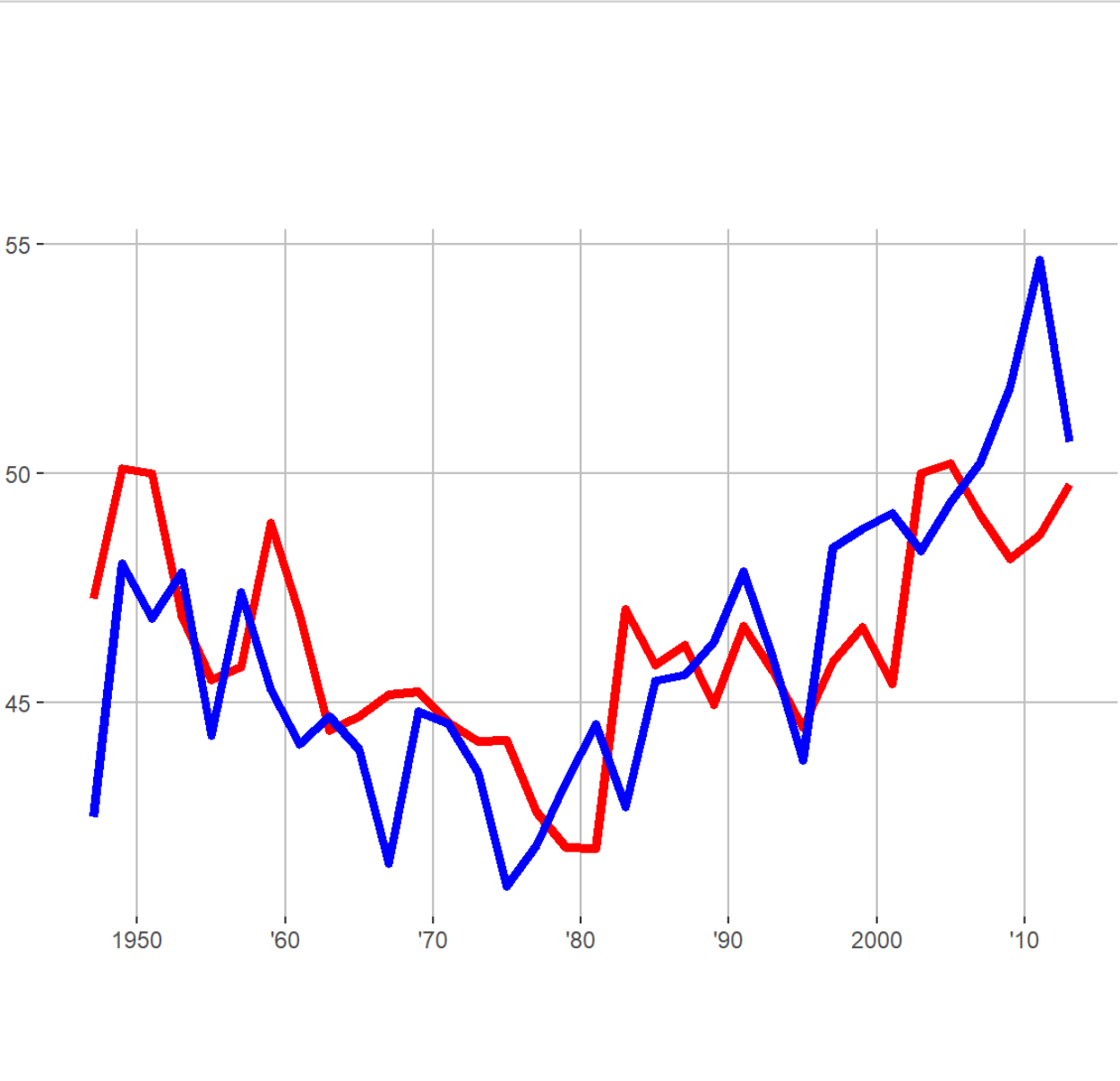


#### The original plot has grid lines and has a more rectangular aspect ratio. `theme(aspect.ratio = .64)` will make the x-axis longer than the y-axis, while

`scale_x_continuous(breaks = seq(1950, 2015, by = 10), labels=c("1950", "'60", "'70", "'80", "'90", "'2000", "'10"))` creates breaks every 10 steps, starting at 1950 and ending at 2015, allowing for us to implement gridlines with the `panel` parameter of `theme()`.

Additionally, the original graph does not have traditional axis labels or a legend, so we can remove those features with `labs(x=NULL, y=NULL)` and `guides(colour="none")`.

```
plot = plot + theme(aspect.ratio = .64) +
  theme(
    panel.background = element_rect(fill = NA),
    panel.grid.major = element_line(colour = "grey"),
    panel.ontop = FALSE
  ) + labs(x=NULL, y=NULL) + scale_x_continuous(breaks = seq(1950, 2015, by = 10),
    labels=c("1950", "'60", "'70", "'80", "'90", "'2000", "'10")) + guides(colour="none")
plot
```



The final step to replicate the graph is to add a title and several inline labels. `ggtitle` is one way that we can set the title and subtitle, and `geom_text` allows us to label the two charts without needing a legend to the left or the right of the figure.

`scale_y_continuous(breaks = seq(40, 60, by = 5), labels=c("40", "45", "50", "55", "60 yrs"))` creates a new y-axis scaling, which we can use to make custom tickmarks like 60yrs. However, the new scaling causes some problems, such as `ylim()` no longer working. To fix this, we can just add a `geom_text` with position `y=border` to make sure that the graph extends to that region.

```
plot = plot + ggtitle("Average Age of New Members of Congress", subtitle =
  "Rolling two-term average at start of term, 1949-2013") +
  geom_text(aes(y = 49, x = 1966), label = "Republicans", cex = 3.8, color="red") +
  geom_text(aes(y = 54, x = 1964), label = "Democrats", cex = 3.8, color="blue") +
  geom_text(aes(y = 60, x = 2004), label = "") + scale_y_continuous(breaks = seq(40, 60, by = 5),
    labels=c("40", "45", "50", "55", "60 yrs")) + geom_text(aes(y = 40, x = 1950), label = "")
plot
```



We have now reproduced the original plot from *Both Republicans And Democrats Have an Age Problem*.