

HW Assignment 1

Ahmed Hassan

2024-08-31

R Markdown Introduction

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

In this R Markdown Document I am loading the data of the ages of congress members from modern times all the way back to the 1800s The article I got this data from was Congress Today Is Older Than It's Ever Been By Geoffrey Skelley.

Find the original article here: <https://fivethirtyeight.com/features/aging-congress-boomers/> raw data csv: https://raw.githubusercontent.com/fivethirtyeight/data/master/congress-demographics/data_aging_congress.csv

In this article Geoffrey is arguing that boomers are the oldest members in congress since the past couple century. I picked this article because I believe this goes with the global issue we are facing today of an ageing population, declining birth rates, and increase in cost of living. This is a topic I have been following closely and following whats been happening from country to country that are facing this issue. There is also the case that there is a lot of money to be made from the insider trading and lobbying when being a part of congress. A final point I would like to mention is that we increased life expectancy due to vaccines, nursing, hygiene, and antibiotics. So it's no wonder the baby boomer generation account for half the members of congress.

Skelly in the articles talks about how half of congress being boomers can affect policy. For example, he mentions that the median age for house of representative members is 59 years old while the median age for senators is 65 years old. I would say the reason for that age difference between house members and senators is because its a lot harder to stay as a house member than a senator. A senators term is 6 years, while a house term is 2 years. Therefore a person who is in the house of representative is constantly campaigning throughout their term. Besides my conclusion, Skelly argues that these members of congress would pass policy that reflect based on what their age group since the older population tend to vote more than the younger population.

The author mentioned that

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/congress-demographics/data_aging_
congress_age_Data <- read.csv(url)

#print(congress_age_Data)
```

Conclusion

Here I omitted all the columns except for generation and age_years. The other columns such as the state_abbrev, "party_code", "bioname", "bioguide_id" are redundant information to look at the data for the congress people's ages. What I wanted to here was take the count of all members of congress by generation. Some generations pre 50 states of America I want to omit because the United States was a much smaller country then and didn't have as high of a population count. So we won't have as much members in the house. the only generations I would keep as part of my data are boomers, Gen x, Gen Z, Millennial, Missionary, and Silent generation. These are a better comparison to see what happened in the last century. I filtered data as well through the data from the years 2000-2020. My findings for the filtered data and unfiltered data showed really no change in the average age. One thing I can do better is group by the name of the members of congress and take their median age during their stay at congress. But, I don't think this data really truly represents the declining birth rate and the aging population. The author of the article mentioned that members of congress were always older by default. Another problem is that millennials and Gen Z have not reached their 50's or 60's yet. A final problem I would like to add is that there is also the fact that the incumbent who is running has more experience, better name recognition among the voters giving that person a more likely chance of winning reelection. We can see that with the Presidency since the time Washington was in office that it was very rare for a president to lose the election on his second term in office. There is also the case of a higher life expectancy. I would like to do more analysis on the data as the author did as soon as I learn more how to work with data like this

```
## # A tibble: 10 x 3
##   generation average_age count
##   <chr>          <dbl> <int>
## 1 Boomers        53.1  5108
## 2 Gen X          44.8  1130
## 3 Gen Z          26.0    1
## 4 Gilded         82.6   15
## 5 Greatest       52.1  7147
## 6 Lost           53.4  4732
## 7 Millennial     36.2   133
## 8 Missionary     57.9  4768
## 9 Progressive    68.1   485
## 10 Silent        54.0  5601
```

```
## # A tibble: 6 x 3
##   generation average_age count
##   <chr>          <dbl> <int>
```

```
## 1 Boomers      57.3  3762
## 2 Gen X        45.0  1114
## 3 Gen Z        26.0    1
## 4 Greatest     80.6   73
## 5 Millennial   36.2   133
## 6 Silent       68.1  1466
```

I would like to add a data set that shows the total births in the US. I am interested in comparing this dataset with the age of congress in the future

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats   1.0.0    v readr     2.1.5
## v ggplot2   3.5.1    v stringr  1.5.1
## v lubridate 1.9.3    v tibble   3.2.1
## v purrr     1.0.2    v tidyr    1.3.1
## -- 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
```

```
library(openintro)
```

```
## Warning: package 'openintro' was built under R version 4.3.3
```

```
## Loading required package: airports
```

```
## Warning: package 'airports' was built under R version 4.3.3
```

```
## Loading required package: cherryblossom
```

```
## Warning: package 'cherryblossom' was built under R version 4.3.3
```

```
## Loading required package: usdata
```

```
## Warning: package 'usdata' was built under R version 4.3.3
```

```
data('present', package='openintro')
```

```
present
```

```
## # A tibble: 63 x 3
##   year    boys  girls
##   <dbl> <dbl> <dbl>
## 1  1940 1211684 1148715
## 2  1941 1289734 1223693
## 3  1942 1444365 1364631
## 4  1943 1508959 1427901
## 5  1944 1435301 1359499
## 6  1945 1404587 1330869
## 7  1946 1691220 1597452
## 8  1947 1899876 1800064
## 9  1948 1813852 1721216
## 10 1949 1826352 1733177
## # i 53 more rows
```

```
present <- present %>%
  mutate(total_births = boys + girls)

ggplot(data = present, aes(x = year, y = total_births)) +
  geom_line()
```

