Model Planning and Building

Code

For this phase of the project I want use the the data I scraped from Pew Research Centers website that has the immigration information of congresspeople in the House and the Senate. One question I want to answer is weather there is a corelation between a congressperson being a immigrant and their party and how they vote on bills.

115th: https://www.pewresearch.org/fact-tank/2018/08/21/immigrants-or-children-of-immigrants-make-up-at-least-12-of-congress/ (https://www.pewresearch.org/fact-tank/2018/08/21/immigrants-or-children-of-immigrants-make-up-at-least-12-of-congress/)

116th: https://www.pewresearch.org/fact-tank/2019/01/24/in-116th-congress-at-least-13-of-lawmakers-are-immigrants-or-the-children-of-immigrants/ (https://www.pewresearch.org/fact-tank/2019/01/24/in-116th-congress-at-least-13-of-lawmakers-are-immigrants-or-the-children-of-immigrants/)

Here I imported the data from the tables on the two websites and created two tibbles with it.

```
library("tidyverse")
library("rvest")
library("dplyr")
library("knitr")

purl("deliverable1.Rmd", output = "part1.r") # produces r source fro
m rmd
source("part1.r") # executes the source


congress115 <- read_html("https://www.pewresearch.org/fact-tank/2018
/08/21/immigrants-or-children-of-immigrants-make-up-at-least-12-of-c
ongress/")
congress116 <- read_html("https://www.pewresearch.org/fact-tank/2019
/01/24/in-116th-congress-at-least-13-of-lawmakers-are-immigrants-or-
the-children-of-immigrants/")
```

```
names115 <- congress115 %>% html_nodes("tr > td:nth-child(1)") %>% h
tml_text()

generation115 <- congress115 %>% html_nodes("tr > td:nth-child(2)")
%>% html_text()

birthplace115 <- congress115 %>% html_nodes("tr > td:nth-child(3)")
%>% html_text()

office115 <- congress115 %>% html_nodes("tr > td:nth-child(4)") %>%
html_text()

party115 <- congress115 %>% html_nodes("tr > td:nth-child(5)") %>% h
tml_text()

states115 <- congress115 %>% html_nodes("tr > td:nth-child(6)") %>%
html_text()

c115Tibble <- tibble(Name = names115, generation = generation115, Pa
rent_or_Members_birthplace = birthplace115, Chamber = office115, Par
ty = party115, State = states115)



names116 <- congress116 %>% html_nodes("tr > td:nth-child(1)") %>% h
tml_text()

generation116 <- congress116 %>% html_nodes("tr > td:nth-child(2)")
%>% html_text()

birthplace116 <- congress116 %>% html_nodes("tr > td:nth-child(3)")
%>% html_text()

office116 <- congress116 %>% html_nodes("tr > td:nth-child(5)") %>%
html_text()

party116 <- congress116 %>% html_nodes("tr > td:nth-child(4)") %>% h
tml_text()

states116 <- congress116 %>% html_nodes("tr > td:nth-child(6)") %>%
```

```
html_text()

c116Tibble <- tibble(Name = names116, generation = generation116, Pa
rent_or_Members_birthplace = birthplace116, Chamber = office116, Par
ty = party116, State = states116)
```

Cleaning data:

```
c115Tibble$Party <- substr(c115Tibble$Party, 0, 1)
c116Tibble$Party <- substr(c116Tibble$Party, 0, 1)


c115Tibble <- mutate_if(c115Tibble,
is.character, str_replace_all, pattern = "First", replacement = "1")

c115Tibble <- mutate_if(c115Tibble,
is.character, str_replace_all, pattern = "Second", replacement = "2"
)


as.numeric(as.character(c115Tibble$generation))




c116Tibble <- mutate_if(c116Tibble,
is.character, str_replace_all, pattern = "Immigrant", replacement =
"1")

c116Tibble <- mutate_if(c116Tibble,
is.character, str_replace_all, pattern = "Child", replacement = "2")


as.numeric(as.character(c116Tibble$generation))



c115Tibble <- mutate_if(c115Tibble,
is.character, str_replace_all, pattern = "Representative", replaceme
nt = "house")
```

```
c115Tibble <- mutate_if(c115Tibble,
is.character, str_replace_all, pattern = "Senator", replacement = "s
enate")

c116Tibble <- mutate_if(c116Tibble,
is.character, str_replace_all, pattern = "Representative", replaceme
nt = "house")

c116Tibble <- mutate_if(c116Tibble,
is.character, str_replace_all, pattern = "Senator", replacement = "s
enate")
```

Relaced the full word for the persons party with just the first letter to match existing data.

Making the persons generation a number instead of first/second and child/immigrant.

Everyone who is not a first or 2nd generation immigrant was given the generation number 3 because I do not have data for how many generations non recently immigrated congresspeople have been Americans.

Switch “Representative” to “house” and “Senator” to “senate” to match existing data.

Setting the generation in the politician tibble to the correct values. Ended up doing this manually after spending days trying to figure out a better way to do this.

```
polititon_with_generation <- merge(c115Tibble, c116Tibble)

#changing the name column so its the same as the other data frames s
o I can merge the generation numbers into those tables
polititon_with_generation$Name <- gsub("\\,.*","", polititon_with_
generation$Name)
colnames(polititon_with_generation)[colnames(polititon_with_genera
tion) == "Name"] <- "last_name"

#replacing the generation number in politician with the number from
polititon_with_generation

#which(politician$last_name=="Clarke")
#gets the row number. had to use since search wont show the actual r
ow number.
```

```
politician[4,"generation"]=2
politician[5,"generation"]=2
politician[6,"generation"]=2

politician[128,"generation"]=2
politician[129,"generation"]=2
politician[130,"generation"]=2

politician[1436,"generation"]=2
politician[1437,"generation"]=2
politician[1438,"generation"]=2

politician[95,"generation"]=2
politician[96,"generation"]=2
politician[96,"generation"]=2

politician[1439,"generation"]=2
politician[1440,"generation"]=2
politician[1441,"generation"]=2

politician[116,"generation"]=2
politician[117,"generation"]=2
politician[118,"generation"]=2

politician[47,"generation"]=2
politician[48,"generation"]=2
politician[49,"generation"]=2


politician[275,"generation"]=1
politician[276,"generation"]=1
politician[277,"generation"]=1

politician[248,"generation"]=2
politician[249,"generation"]=2
politician[250,"generation"]=2

politician[223,"generation"]=2
politician[224,"generation"]=2
politician[225,"generation"]=2
```

```
politician[204,"generation"]=2
politician[205,"generation"]=2
politician[206,"generation"]=2

politician[269,"generation"]=2
politician[270,"generation"]=2
politician[271,"generation"]=2

politician[1486,"generation"]=2
politician[1487,"generation"]=2
politician[1488,"generation"]=2

politician[198,"generation"]=2
politician[199,"generation"]=2
politician[200,"generation"]=2

politician[309,"generation"]=2
politician[310,"generation"]=2
politician[311,"generation"]=2

politician[323,"generation"]=2
politician[324,"generation"]=2
politician[325,"generation"]=2

politician[1500,"generation"]=2
politician[1501,"generation"]=2
politician[1502,"generation"]=2

politician[1492,"generation"]=2
politician[1493,"generation"]=2
politician[1494,"generation"]=2

politician[378,"generation"]=2
politician[379,"generation"]=2
politician[380,"generation"]=2

politician[391,"generation"]=1
politician[392,"generation"]=1
politician[393,"generation"]=1

politician[484,"generation"]=2
politician[485,"generation"]=2
```

```
politician[486,"generation"]=2

politician[510,"generation"]=2
politician[511,"generation"]=2
politician[512,"generation"]=2

politician[452,"generation"]=2
politician[453,"generation"]=2
politician[454,"generation"]=2

politician[1546,"generation"]=2
politician[1547,"generation"]=2
politician[1548,"generation"]=2

politician[548,"generation"]=2
politician[549,"generation"]=2
politician[550,"generation"]=2

politician[1538,"generation"]=2
politician[1539,"generation"]=2
politician[1540,"generation"]=2

politician[1535,"generation"]=1
politician[1536,"generation"]=1
politician[1537,"generation"]=1

politician[530,"generation"]=2
politician[531,"generation"]=2
politician[532,"generation"]=2

politician[641,"generation"]=1
politician[642,"generation"]=1
politician[643,"generation"]=1

politician[699,"generation"]=2
politician[700,"generation"]=2
politician[701,"generation"]=2

politician[682,"generation"]=2
politician[683,"generation"]=2
politician[684,"generation"]=2
```

```
politician[704,"generation"]=1
politician[705,"generation"]=1
politician[706,"generation"]=1

politician[781,"generation"]=1
politician[782,"generation"]=1
politician[783,"generation"]=1

politician[773,"generation"]=2
politician[774,"generation"]=2
politician[775,"generation"]=2

politician[1600,"generation"]=2
politician[1601,"generation"]=2
politician[1602,"generation"]=2

politician[869,"generation"]=2
politician[870,"generation"]=2
politician[871,"generation"]=2

politician[1612,"generation"]=1
politician[1613,"generation"]=1
politician[1614,"generation"]=1

politician[927,"generation"]=2
politician[928,"generation"]=2
politician[929,"generation"]=2

politician[1630,"generation"]=2
politician[1631,"generation"]=2
politician[1632,"generation"]=2

politician[1645,"generation"]=2
politician[1646,"generation"]=2
politician[1647,"generation"]=2

politician[1648,"generation"]=2
politician[1649,"generation"]=2
politician[1650,"generation"]=2

politician[1083,"generation"]=1
politician[1084,"generation"]=1
```

```
politician[1085,"generation"]=1

politician[1159,"generation"]=2
politician[1160,"generation"]=2
politician[1161,"generation"]=2

politician[1658,"generation"]=2
politician[1659,"generation"]=2
politician[1660,"generation"]=2

politician[1148,"generation"]=2
politician[1149,"generation"]=2
politician[1150,"generation"]=2

politician[1680,"generation"]=2
politician[1681,"generation"]=2
politician[1682,"generation"]=2

politician[1165,"generation"]=1
politician[1166,"generation"]=1
politician[1167,"generation"]=1

politician[1176,"generation"]=2
politician[1177,"generation"]=2
politician[1178,"generation"]=2

politician[1219,"generation"]=2
politician[1220,"generation"]=2
politician[1221,"generation"]=2

politician[1696,"generation"]=2
politician[1697,"generation"]=2
politician[1698,"generation"]=2

politician[1278,"generation"]=1
politician[1279,"generation"]=1
politician[1280,"generation"]=1

politician[1312,"generation"]=2
politician[1313,"generation"]=2
politician[1314,"generation"]=2
```

```
politician[1717,"generation"]=2
politician[1718,"generation"]=2
politician[1719,"generation"]=2


summary(politician)
```

Graphing

```
averages <- add_column(averages, politician$generation)
colnames(averages)[colnames(averages) == "politician$generation"] <-
"generation"



 averages$generation <- as.factor(averages$generation)
# levels(averages$generation)
#
# averages$agree_pct <- as.factor(averages$agree_pct)
# levels(averages$agree_pct)

# avg_pct_vote_by_gen <- ggplot(averages, aes(generation, agree_pct)
) + geom_bar(stat="identity", width=.5, color="green") + theme(axis.
text.x = element_text(angle = 0))
avg_pct_vote_by_gen <- ggplot(averages, aes(x=generation, y=agree_pc
t)) + stat_summary(fun.y="mean", geom="bar", color="green") + ylab("
agree percentage") +ggtitle("politician generation vs. agree percent
age")
avg_pct_vote_by_gen
```

Here we can see that first generation congresspeople - those who immigrated here themselves, are the least likely to agree with Trump. Third generation congressprople are the most likely to agree with Trump.

I would like to be able to predict how likely it is for congresspeople whos families have been living in the country for several or multiple generations. The above graph gives an idea of the trend.

Modeling

```
cor(as.numeric(averages$generation), averages$agree_pct)
model <- lm(averages$agree_pct ~ as.numeric(averages$generation) + a
verages$predicted_agree)
summary(model)
plot(model)
```

```
head(predict(model, data.frame(generation = 4)))
```