

Yurong Jiang

Homework03

```
#1
library(dplyr)

##
## 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

library(magrittr)
library(ggplot2)

congress = read.csv("congress.csv")

congress %>%
  filter(congress == 113) %>%
  summarise(mean_age = mean(age))

##   mean_age
## 1 57.62904

#2.
congress%>%
  filter(as.Date(termstart,format = '%Y-%m-%d') >= 1947-01-01) %>%
  group_by(party)%>%
  summarise(total_number = n())

## # A tibble: 4 x 2
##   party total_number
##   <fct>      <int>
## 1 D          5568
## 2 I           27
## 3 ID           2
## 4 R          4763

#3.
congress %>%
  filter(congress == 92) %>%
  group_by(party,chamber) %>%
  summarise(total_number = n())

## # A tibble: 5 x 3
## # Groups:   party [?]
##   party chamber total_number
##   <fct> <fct>      <int>
## 1 D     house       258
## 2 D     senate       56
```

```
## 3 I      senate      4
## 4 R      house     186
## 5 R      senate     45
```

```
#4.
congress %>%
  filter(congress == 112 & chamber == 'senate') %>%
  group_by(party) %>%
  summarise(max_age = max(age),
            min_age = min(age),
            avg_age = mean(age))
```

```
## # A tibble: 3 x 4
##   party max_age min_age avg_age
##   <fct>   <dbl>   <dbl>   <dbl>
## 1 D      90.3    38.2    62.8
## 2 I      69.3    68.9    69.1
## 3 R      78.8    39.6    60.8
```

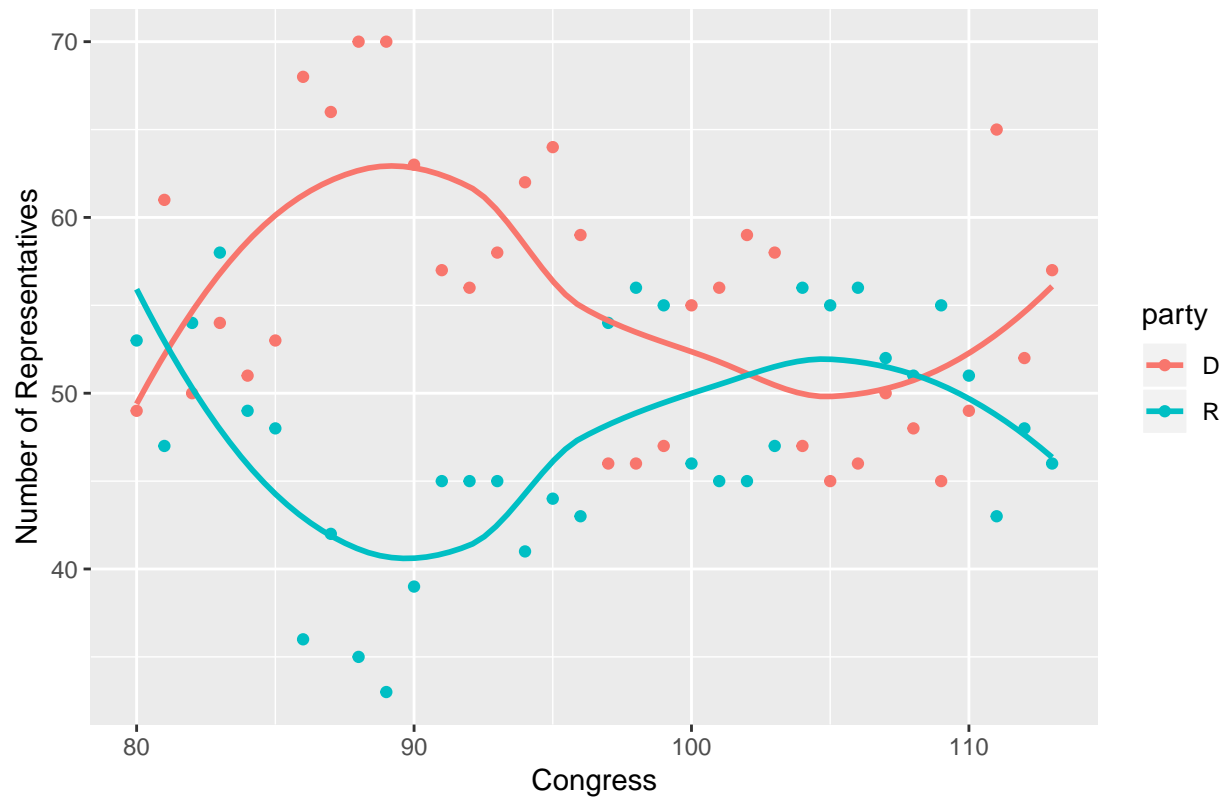
```
#5
congress %>%
  filter(congress == 113 & chamber == 'house' & party == 'D') %>%
  group_by(state) %>%
  summarise(number = n()) %>%
  arrange(-number) %>%
  summarise(highest_state = first(state),
            highest_num = first(number))
```

```
## # A tibble: 1 x 2
##   highest_state highest_num
##   <fct>          <int>
## 1 CA              38
```

```
#6
data6 = congress %>%
  filter(chamber == 'senate' & (party == 'D' | party == 'R')) %>%
  group_by(party, congress) %>%
  summarise(number = n())
ggplot(data6, aes(x=congress, y=number, color = party))+
  geom_point()+
  geom_smooth(se=F)+
  ggtitle("Trend of Senate Representatives")+
  ylab("Number of Representatives")+
  xlab("Congress")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Trend of Senate Representatives



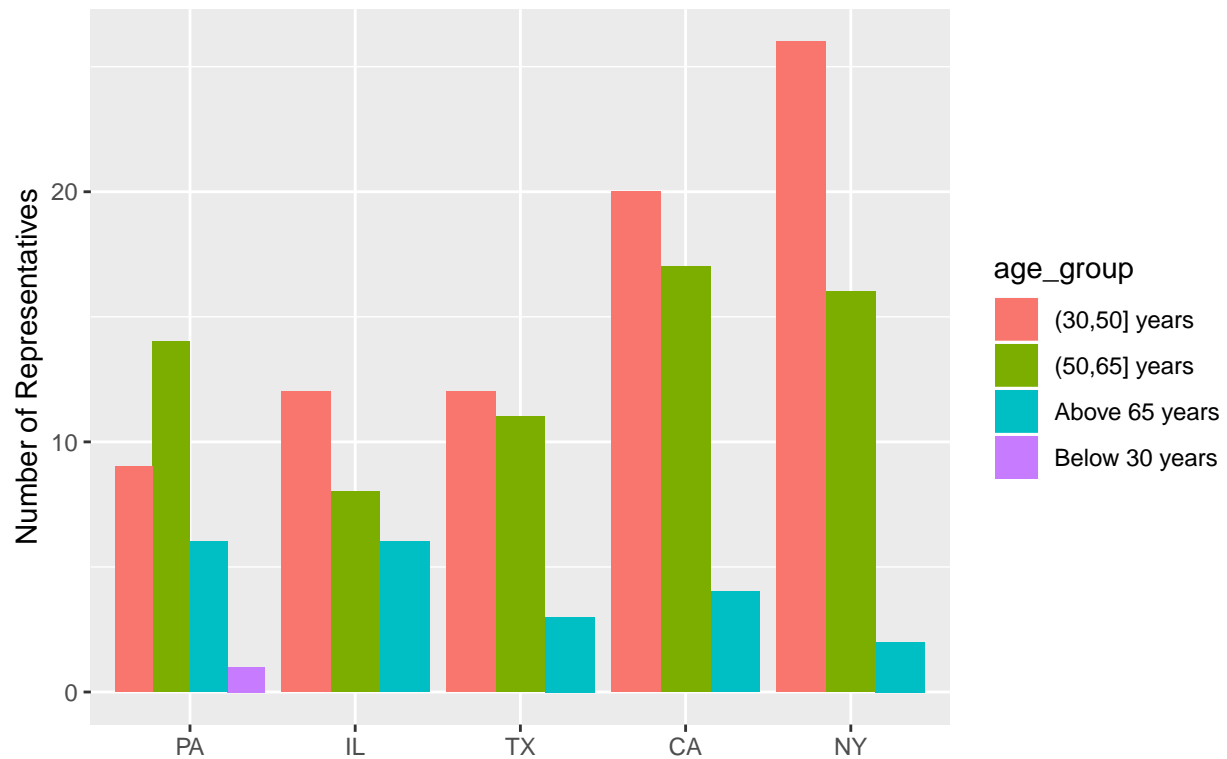
#7.

```
congress$age_group = ifelse(as.integer(congress$age)<=30, "Below 30 years",
                           ifelse(as.integer(congress$age) <= 50, "(30,50] years",
                                   ifelse(as.integer(congress$age) <= 65, "(50,65] years","Above 65 years"))
```

#8

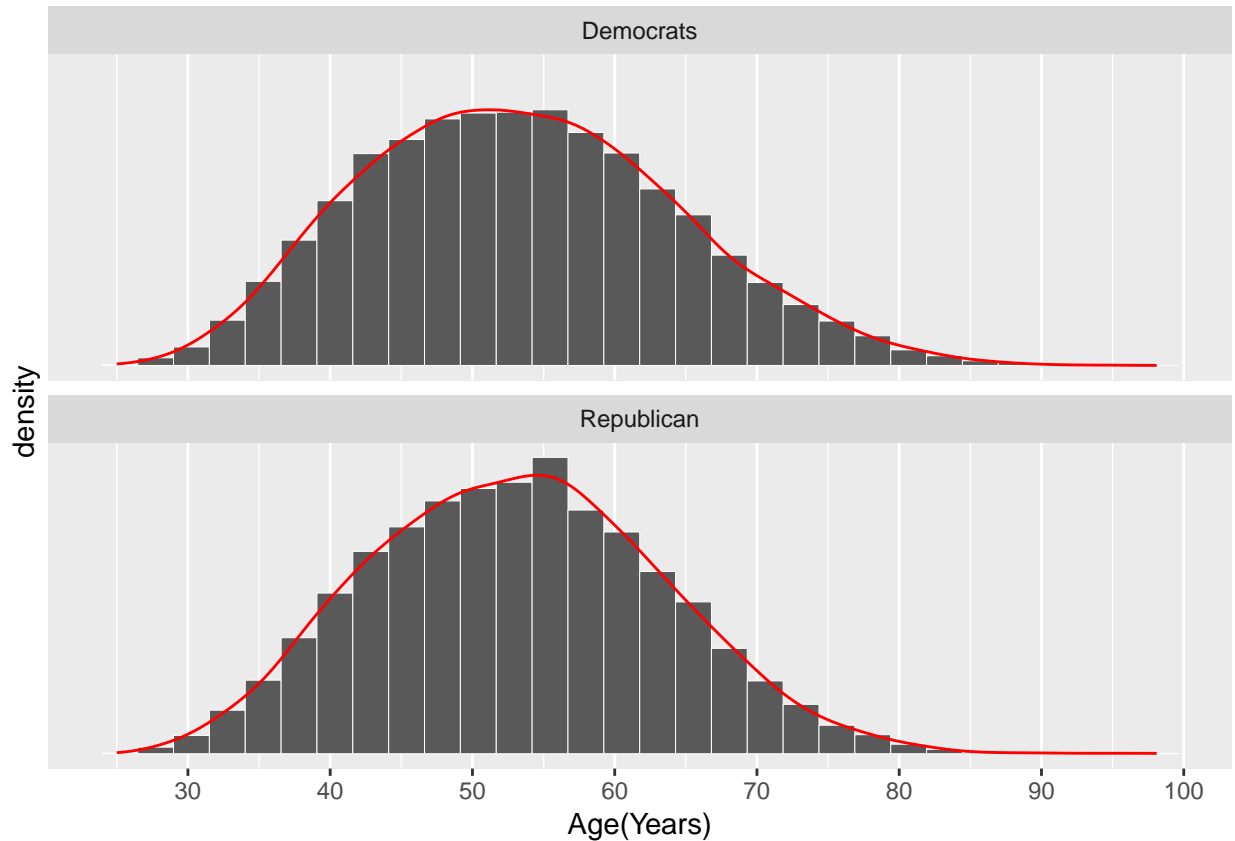
```
congress %>%
  filter(congress == 90 & (state == 'CA'|state == 'NY'|state == 'TX'|state == 'PA'|state == 'IL')) %>%
  group_by(state, age_group) %>%
  summarise(number_con = n()) %>%
  ggplot(aes(x = reorder(state, number_con), y = number_con, fill = age_group)) +
  geom_col(position = 'dodge') +
  xlab('') +
  ylab('Number of Representatives') +
  ggtitle('Age Distribution for Congress 90')
```

Age Distribution for Congress 90



```
#9
data = congress
levels(data$party) = c("AL", "Democrats", "I", "ID", "L", "Republican")
data %>%
  filter(party %in% c("Democrats", "Republican")) %>%
  ggplot(aes(x = age, y = ..density..)) +
    geom_histogram(color = 'white', size = 0.01) +
    geom_line(stat = "density", color = 'red') +
    facet_wrap(~party, nrow = 2) +
    scale_x_continuous(name= "Age(Years)", breaks = seq(30, 100, 10), labels = seq(30, 100, 10)) +
    scale_y_continuous(name = "density", breaks = NULL)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



10

Key point: 1. It's important to know that communication, attention and time should be paid to understanding the context for the need to communicate. 2. We should always be able to distinguish exploratory and explanatory analysis. Exploratory analysis is what you do to understand the data and figure out what might be noteworthy or interesting to highlight to others. Explanatory analysis is taking the time to turn the data into information that can be consumed by an audience: the two pearls. 3. Knowing you are communicating whom. 4. Knowing you want your audience to know what. 5. Understanding how can you use data to help make your point. 6. Figuring out what tone you want your communication to set. 7. Understanding what data is available that will help make my point. 8. If you know exactly what it is you want to communicate, you can make it fit the time slot you're given, even if it isn't the one for which you are prepared. 9. Big Idea has three components: It must articulate your unique point of view; It must convey what's at stake; and it must be a complete sentence. 10. When it comes to explanatory analysis, being able to concisely articulate exactly who you want to communicate to and what you want to convey before you start to build content reduces iterations and helps ensure that the communication you build meets the intended purpose.