# Midterm

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyv
erse 1.2.1 ──
```

```
## ✔ ggplot2 3.2.1      ✔ purrr   0.3.2
## ✔ tibble  2.1.3      ✔ dplyr   0.8.3
## ✔ tidyr   1.0.0      ✔ stringr 1.4.0
## ✔ readr   1.3.1      ✔ forcats 0.4.0
```

```
## ── Conflicts ──────────────────────────────────────────── tidyverse_c
onflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(fivethirtyeight)

data(congress_age)
```
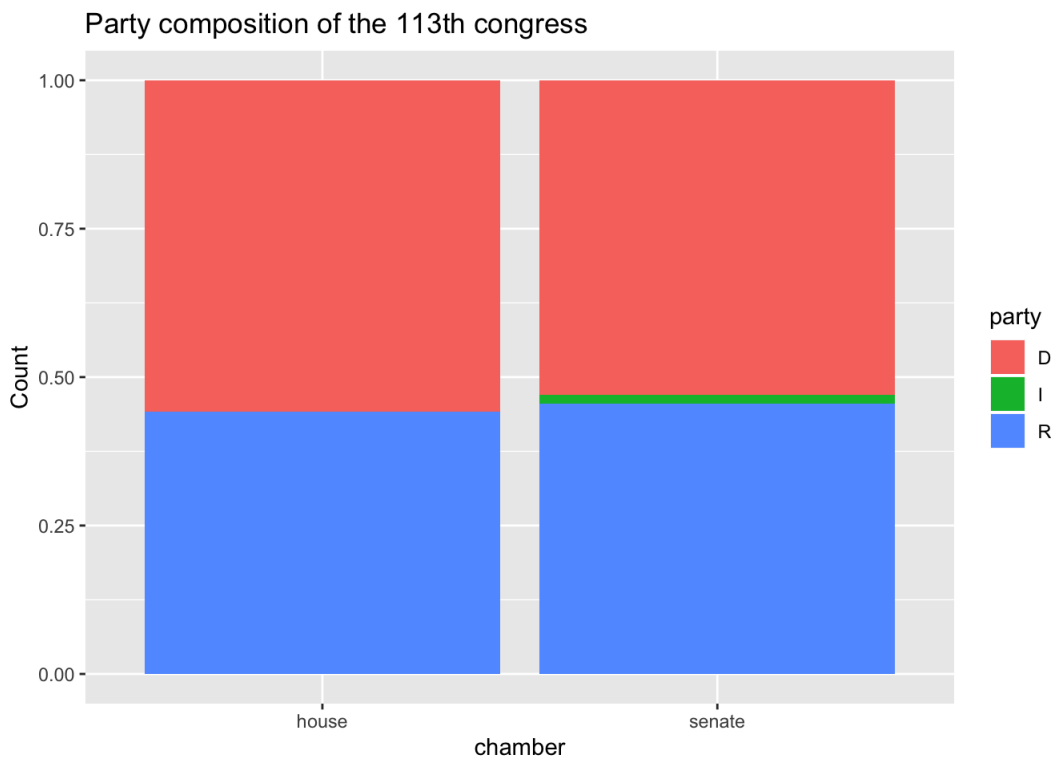
```
head(congress_age)
```

```
## # A tibble: 6 x 13
##   congress chamber bioguide firstname middlename lastname suffix birthday
##      <int> <chr>   <chr>    <chr>     <chr>      <chr>    <chr>  <date>
## 1       80 house   M000112  Joseph    Jefferson  Mansfie… <NA>   1861-02-09
## 2       80 house   D000448  Robert    Lee        Doughton <NA>   1863-11-07
## 3       80 house   S000001  Adolph    Joachim    Sabath   <NA>   1866-04-04
## 4       80 house   E000023  Charles   Aubrey     Eaton    <NA>   1868-03-29
## 5       80 house   L000296  William   <NA>       Lewis    <NA>   1868-09-22
## 6       80 house   G000017  James     A.         Gallagh… <NA>   1869-01-16
## # … with 5 more variables: state <chr>, party <chr>, incumbent <lgl>,
## #   termstart <date>, age <dbl>
```

```
dir <- filter(congress_age, party == "D" | party == "I" | party == "R")
ggplot(dir, aes(x = chamber, fill = party)) + geom_bar(position = "fill") + ggtitle("Party composition of th
e 113th congress") + xlab("chamber") + ylab("Count")
```

```
congress_age %>% arrange(desc(age))
```

```
## # A tibble: 18,635 x 13
##    congress chamber bioguide firstname middlename lastname  suffix
##       <int> <chr>   <chr>    <chr>     <chr>      <chr>     <chr>
##  1      107 senate  T000254  J.        Strom      Thurmond  <NA>
##  2      106 senate  T000254  J.        Strom      Thurmond  <NA>
##  3      111 senate  B001210  Robert    C.         Byrd      <NA>
##  4      105 senate  T000254  J.        Strom      Thurmond  <NA>
##  5      113 senate  L000123  Frank     R.         Lautenb…  <NA>
##  6      104 senate  T000254  J.        Strom      Thurmond  <NA>
##  7       86 senate  G000418  Theodore  Francis    Green     <NA>
##  8      112 senate  I000025  Daniel    K.         Inouye    <NA>
##  9      103 senate  T000254  J.        Strom      Thurmond  <NA>
## 10      113 house   H000067  Ralph     M.         Hall      <NA>
## # … with 18,625 more rows, and 6 more variables: birthday <date>,
## #   state <chr>, party <chr>, incumbent <lgl>, termstart <date>, age <dbl>
```

J. Strom Thurmond is the oldest.

```
congress_age %>% count(incumbent)
```

```
## # A tibble: 2 x 2
##   incumbent     n
##   <lgl>     <int>
## 1 FALSE      2937
## 2 TRUE      15698
```

```
obama <- filter(congress_age, congress == 113, party == "R" | party == "D")
```

```
onehundred <- filter(congress_age, congress == 100)

onehundred %>% count(incumbent)
```

```
## # A tibble: 2 x 2
##   incumbent     n
##   <lgl>     <int>
## 1 FALSE        63
## 2 TRUE        481
```

There were 63 new congresspeople in the 100th Congress.

```
congress_age %>% group_by(chamber) %>% summarize_at(vars(age), funs(mean))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
## # A tibble: 2 x 2
##   chamber   age
##   <chr>   <dbl>
## 1 house    52.4
## 2 senate   57.3
```
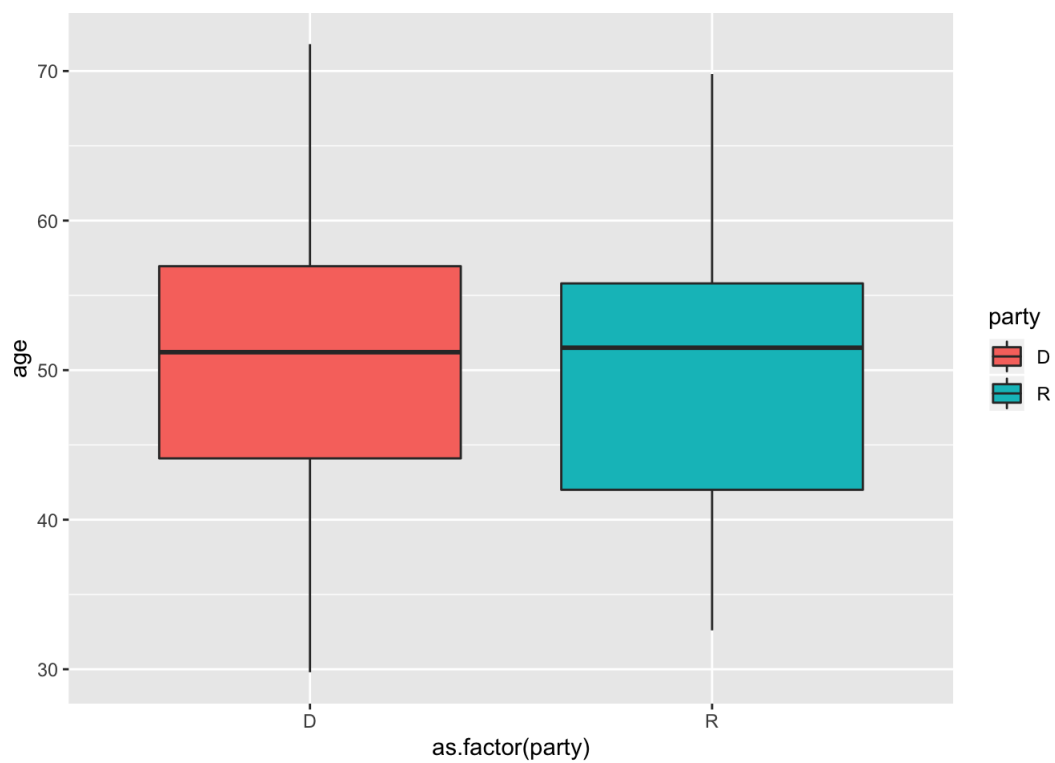
The Senate is older.

```
noninc <- filter(obama, incumbent == "FALSE")
```

```
ggplot(obama, aes(x = as.factor(party), y = age, fill = party)) +
  geom_boxplot()
```
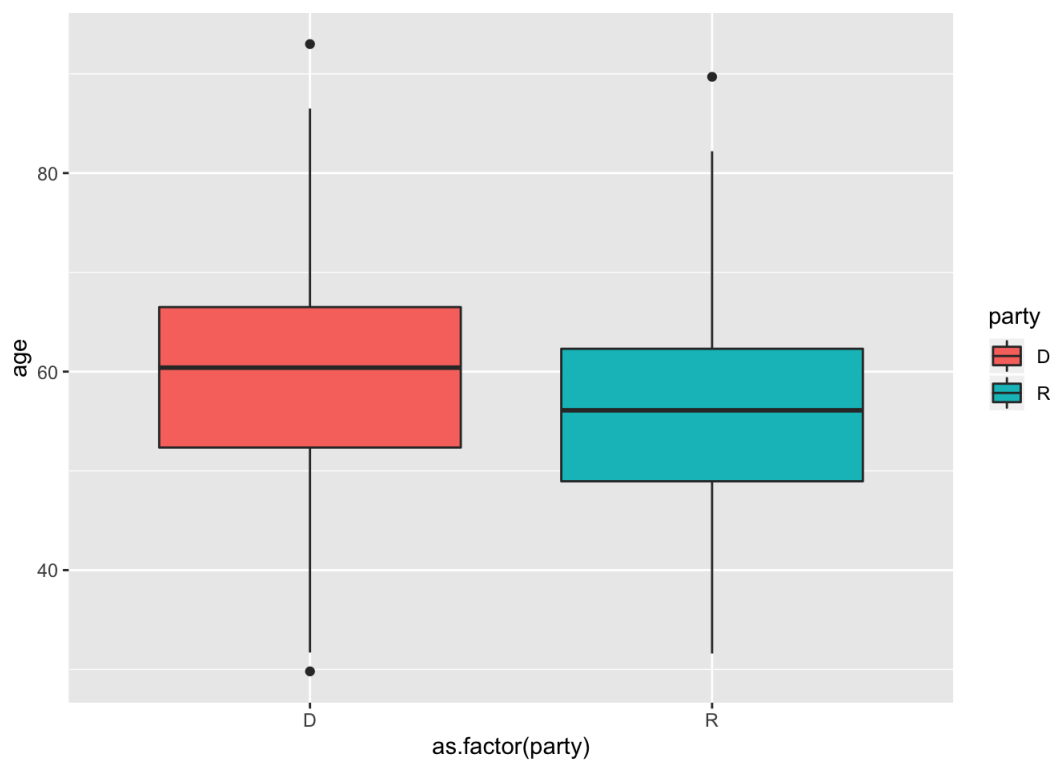


```
ggplot(noninc, aes(x = as.factor(party), y = age, fill = party)) +
  geom_boxplot()
```



```
obama %>% group_by(party) %>% summarize_at(vars(age), funs(mean))
```

```
## # A tibble: 2 x 2
##   party   age
##   <chr> <dbl>
## 1 D      59.6
## 2 R      55.8
```

```
noninc %>% group_by(party) %>% summarize_at(vars(age), funs(mean))
```

```
## # A tibble: 2 x 2
##   party   age
##   <chr> <dbl>
## 1 D      50.7
## 2 R      49.8
```

The criticism is valid because the average age is in the mid-to-late's for both parties, which is pretty high age considering that most new ideas come from people of younger ages. This criticism is not equally valid for both parties because the average age for Democrats is 59.6 for Democrats while the average age for Republicans is 55.8, so although there is not a huge difference, there is still a slight difference in the ages which could allow for slightly more new ideas to be part of the political conversation compared to Democrats. However, for non-incumbents, the ages were pretty much the same for both parties, with the Republican age being a year younger.