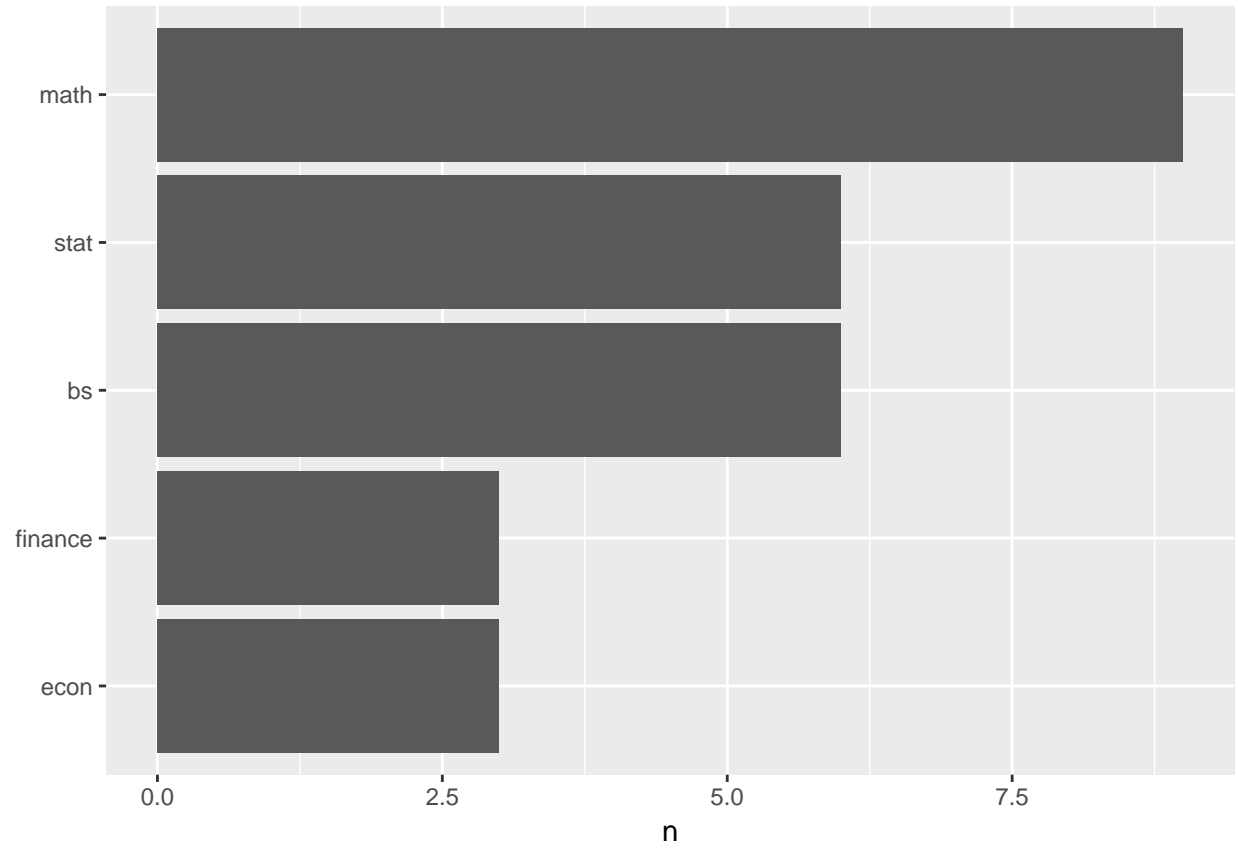


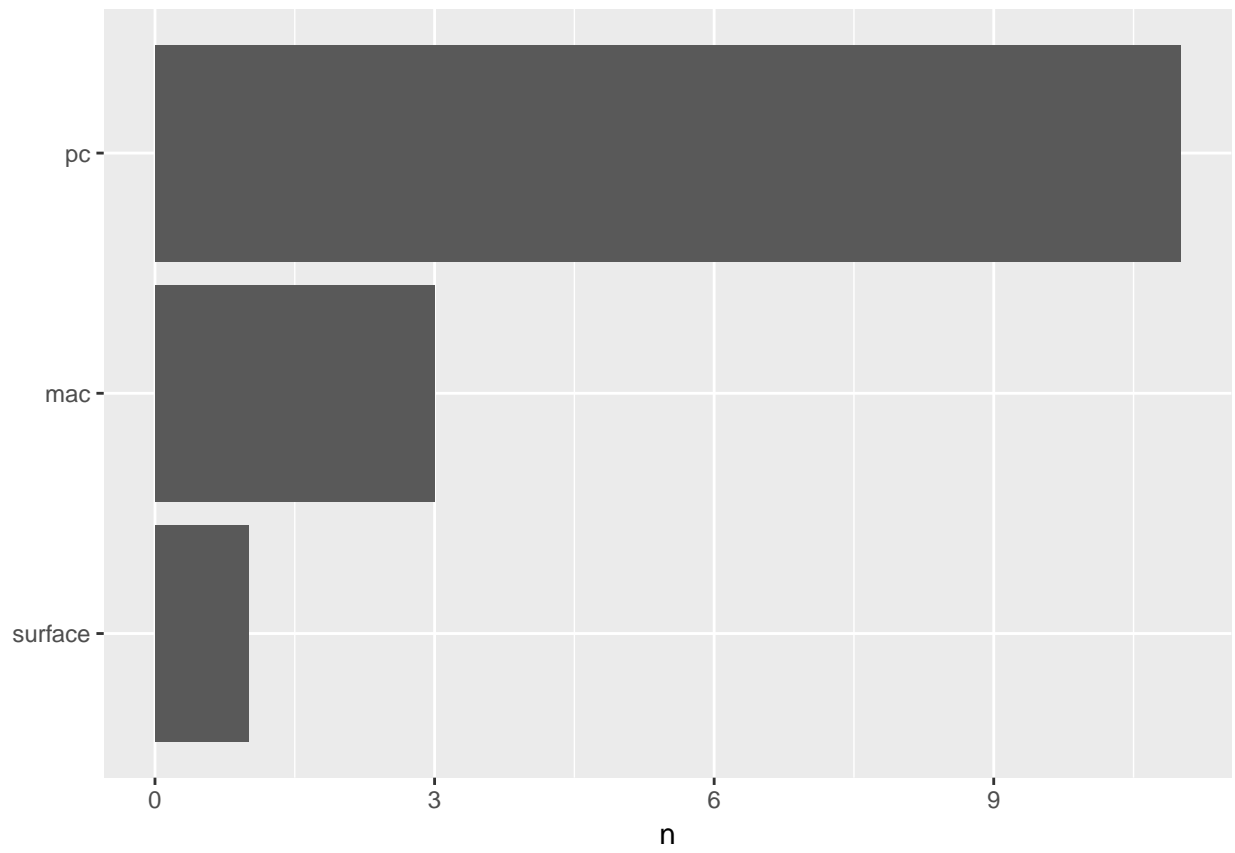
R Notebook

Problem 2: Class Data



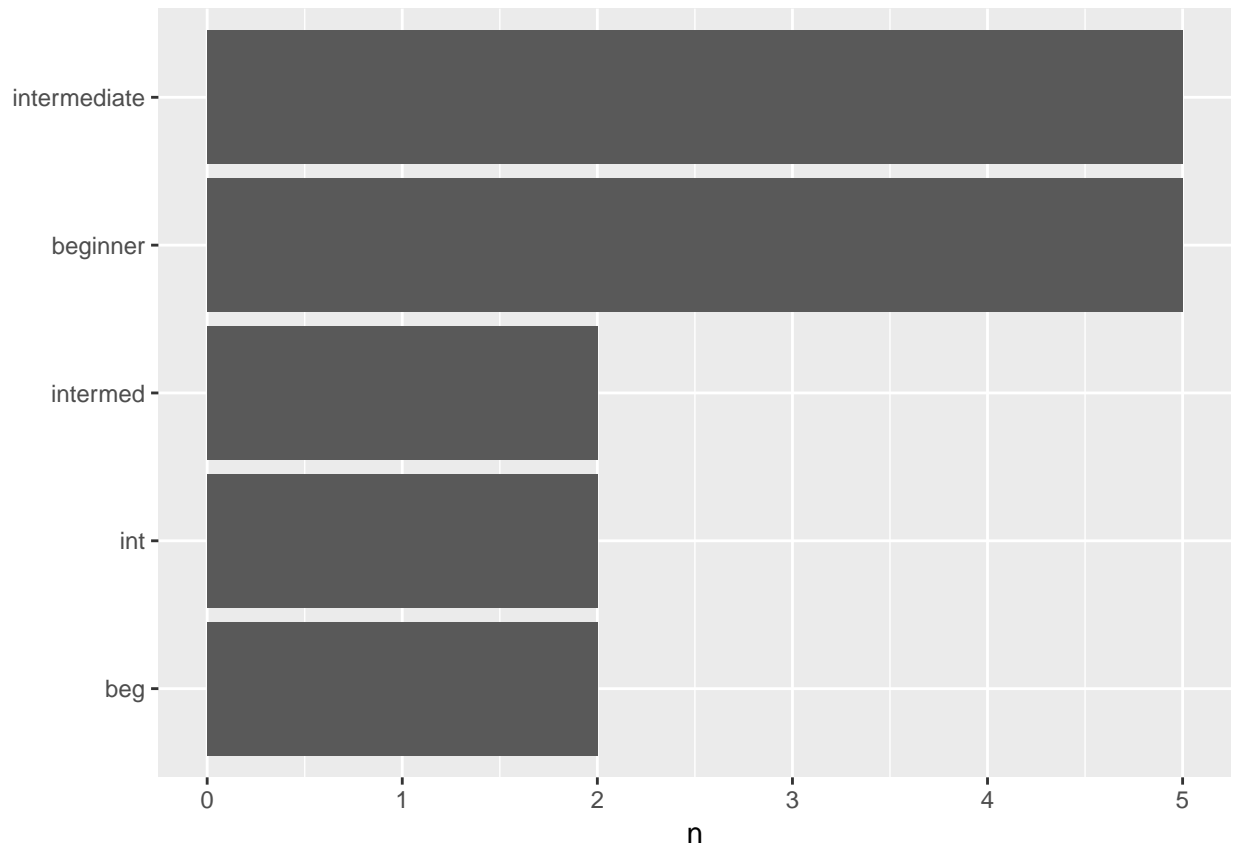
math
finance
stat
econ

From the plot, we can get that most students have bachelor degree with background at math.



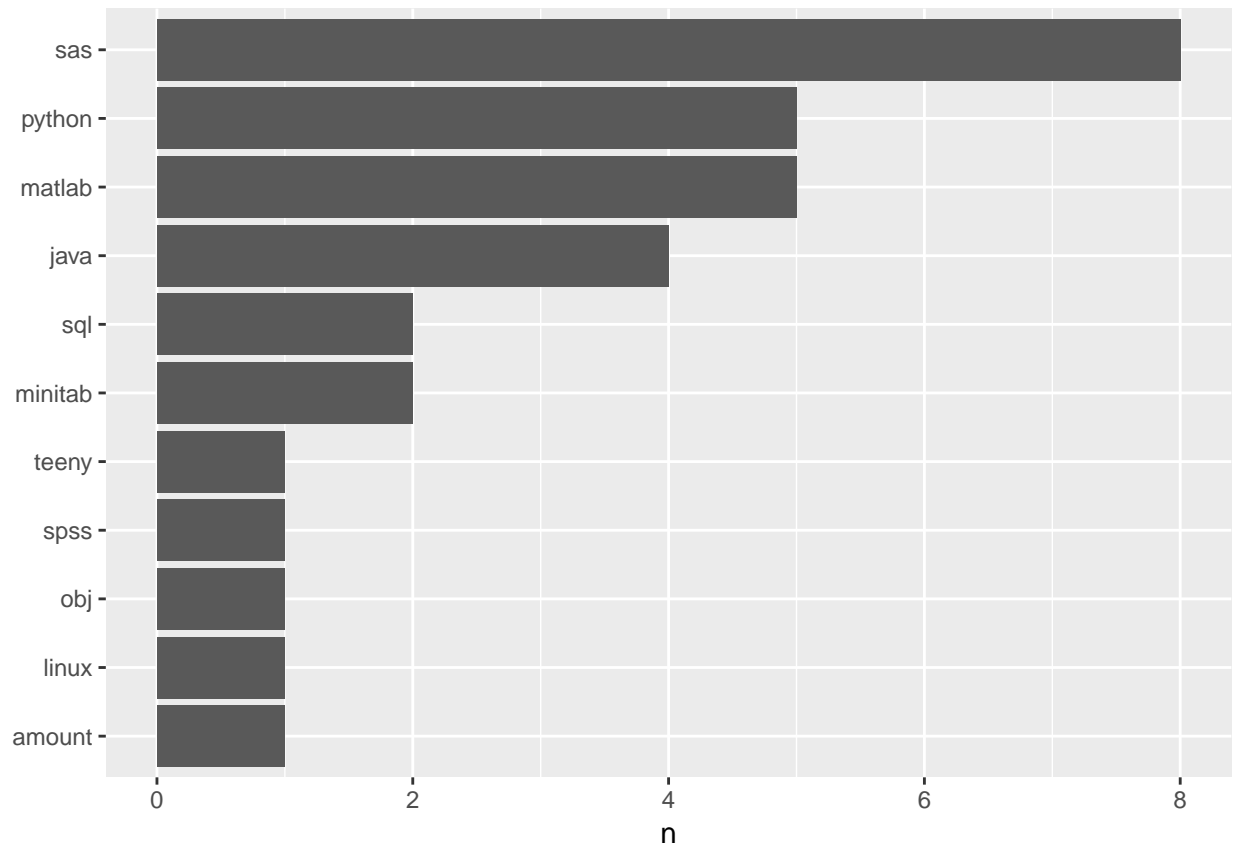
pc
mac

From the plot, we can get most students use PC system.



intermediate beginner

From the graph, we can get most students think themselves as intermediate.



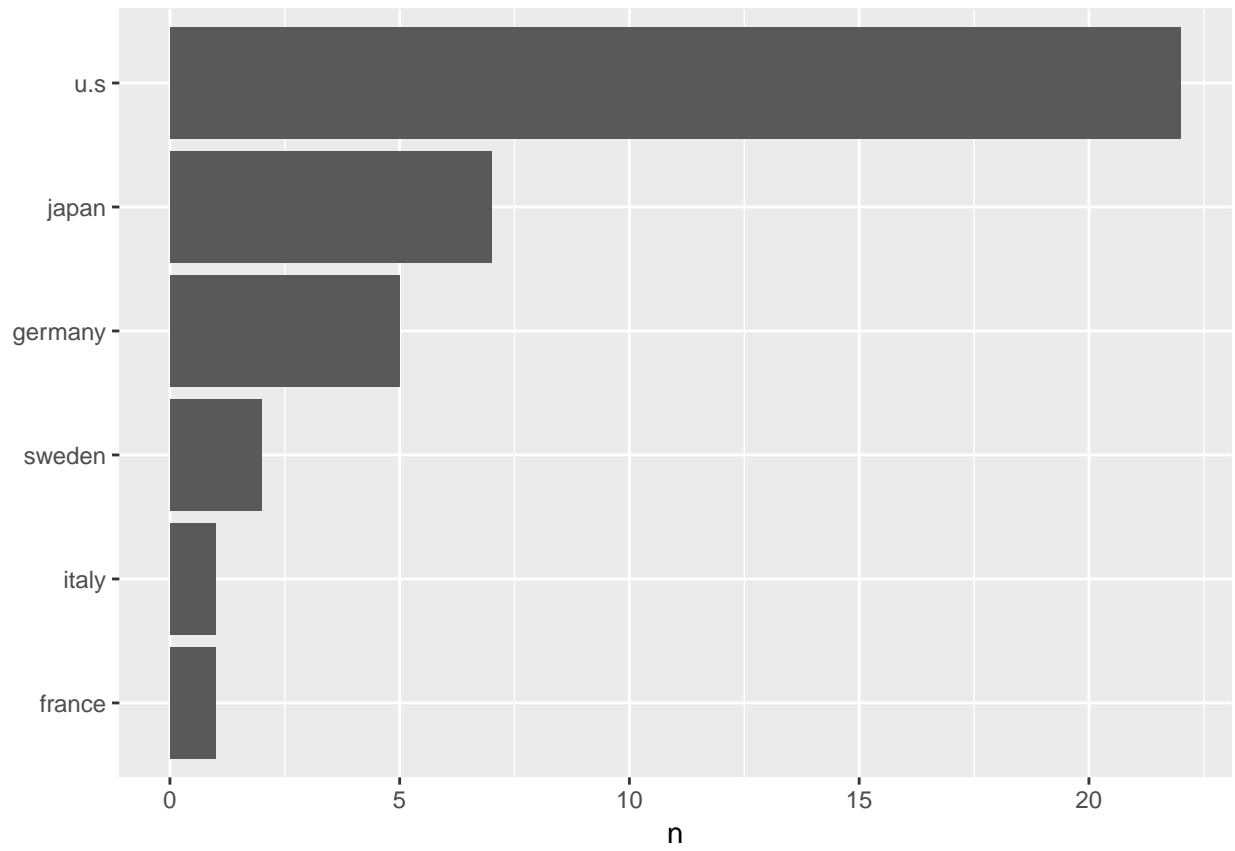


From the graph, we can get 8 students have learnt SAS. Python and Matlab ranked second behind SAS.

Problem 3

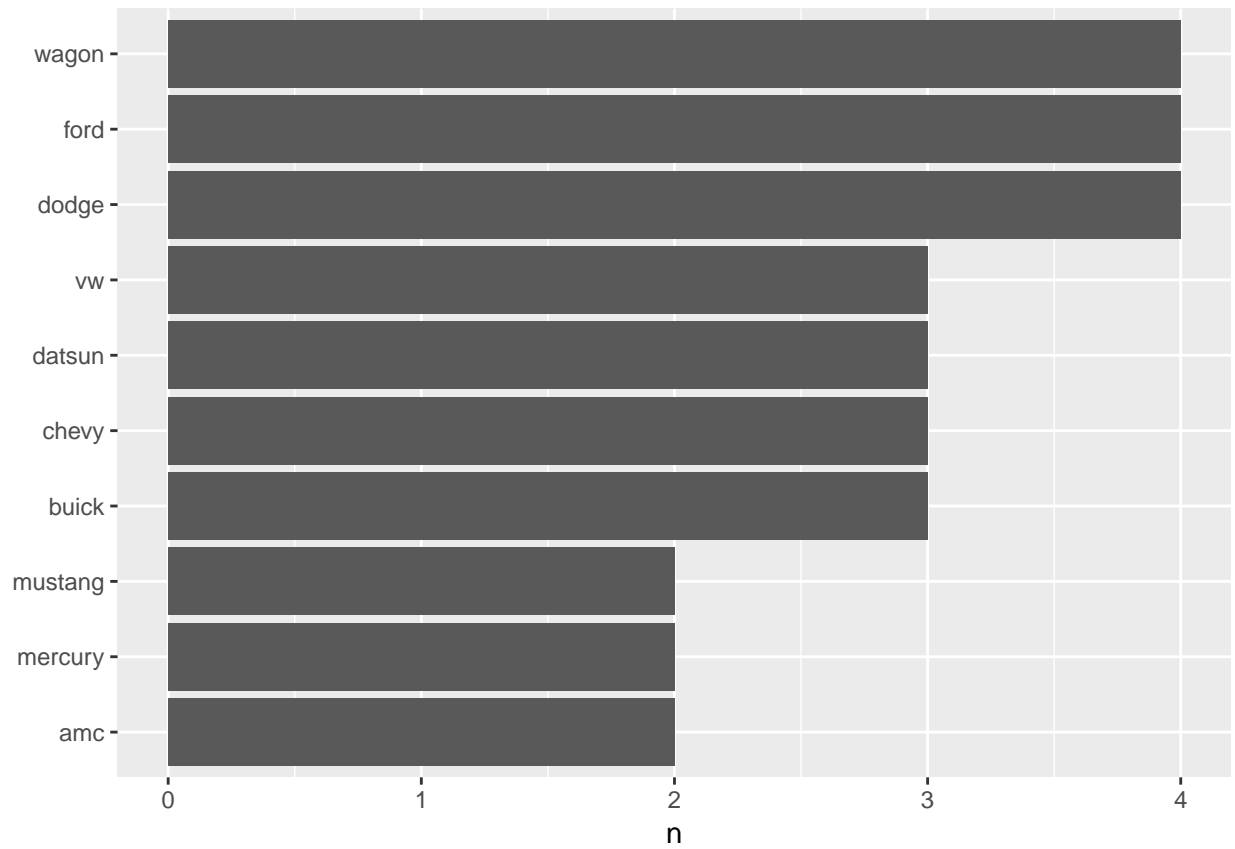
The data set is downloaded at <http://lib.stat.cmu.edu/DASL/Datafiles/Cars.html>.

The data set is about measurements on 38 1978-79 model automobiles. The gas mileage in miles per gallon as measured by Consumers' Union on a test track. Other values as reported by automobile manufacturer. The Cohen's d tests the difference of MPG (Miles per gallon, a measure of gas mileage) between cars produced by U.S. and that produced by other countries, including France, Germany, Italy, Japan and Sweden.



U.S.
japan
germany

From the graph, we can get more than half of the cars been surveyed are manufactured by U.S..



chevy
wagon
buick vw datsun
ford
dodge

From the graph, we can get that wagon, ford and dodge are most popular.

Problem 4

Account has been registered at arc.vt.edu..

```
library(data.table)
library(pander)
Survey_Dat <- fread(input = "C:/Users/Echo/Desktop/2017 Fall/Statistical Package/HW8/Echo-Wei-HW/08_text.txt")
colnames(Survey_Dat) <- c("Major", "Platform", "Level", "Programming")

library(tidytext)
library(dplyr)
library(tidytext)

##### Major #####
library(ggplot2)
par(frow=c(1, 2))

c1 <- Survey_Dat[, 1]
c1_df <- data.frame(line = 1:14, text = c1)
c1_df %>%
  unnest_tokens(word, Major) %>%
  count(word, sort = TRUE) %>%
  filter(n > 2) %>%
  mutate(word = reorder(word, n)) %>%
```

```

ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

library(wordcloud)
c1_df %>%
  unnest_tokens(word, Major) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 200))
##### Platform #####
c2 <- Survey_Dat[, 2]
c2_df <- data.frame(line = 1:14, text = c2)
c2_df %>%
  unnest_tokens(word, Platform) %>%
  count(word, sort = TRUE) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

c2_df %>%
  unnest_tokens(word, Platform) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 200))
##### R Level #####
data(stop_words)
c3 <- Survey_Dat[, 3]
c3_df <- data.frame(line = 1:14, text = c3)
c3_df %>%
  unnest_tokens(word, Level) %>%
  anti_join(stop_words) %>%
  count(word, sort = TRUE) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

c3_df %>%
  unnest_tokens(word, Level) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 200))

data(stop_words)
c4 <- Survey_Dat[, 4]
c4_df <- data.frame(line = 1:14, text = c4)
c4_df %>%
  unnest_tokens(word, Programming) %>%
  anti_join(stop_words) %>%
  count(word, sort = TRUE) %>%
  mutate(word = reorder(word, n)) %>%

```

```

ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

c4_df %>%
  unnest_tokens(word, Programming) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 200))
Car_dat <- fread(input = "C:/Users/Echo/Desktop/2017 Fall/Inference Fundamentals/Homework/HW4/Cohen_Data")
Cou <- Car_dat[, 1]
Cou_df <- data.frame(line = 1:38, text = Cou)
Cou_df %>%
  unnest_tokens(word, Country) %>%
  anti_join(stop_words) %>%
  count(word, sort = TRUE) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

Cou_df %>%
  unnest_tokens(word, Country) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 200))
Ca <- Car_dat[, 2]
Ca_df <- data.frame(line = 1:38, text = Ca)
Ca_df %>%
  unnest_tokens(word, Car) %>%
  anti_join(stop_words) %>%
  count(word, sort = TRUE) %>%
  filter(n > 1) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

Ca_df %>%
  unnest_tokens(word, Car) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 200))

```