

Hmk4-Q3

Jialei Wang(jw3696)

11/12/2018

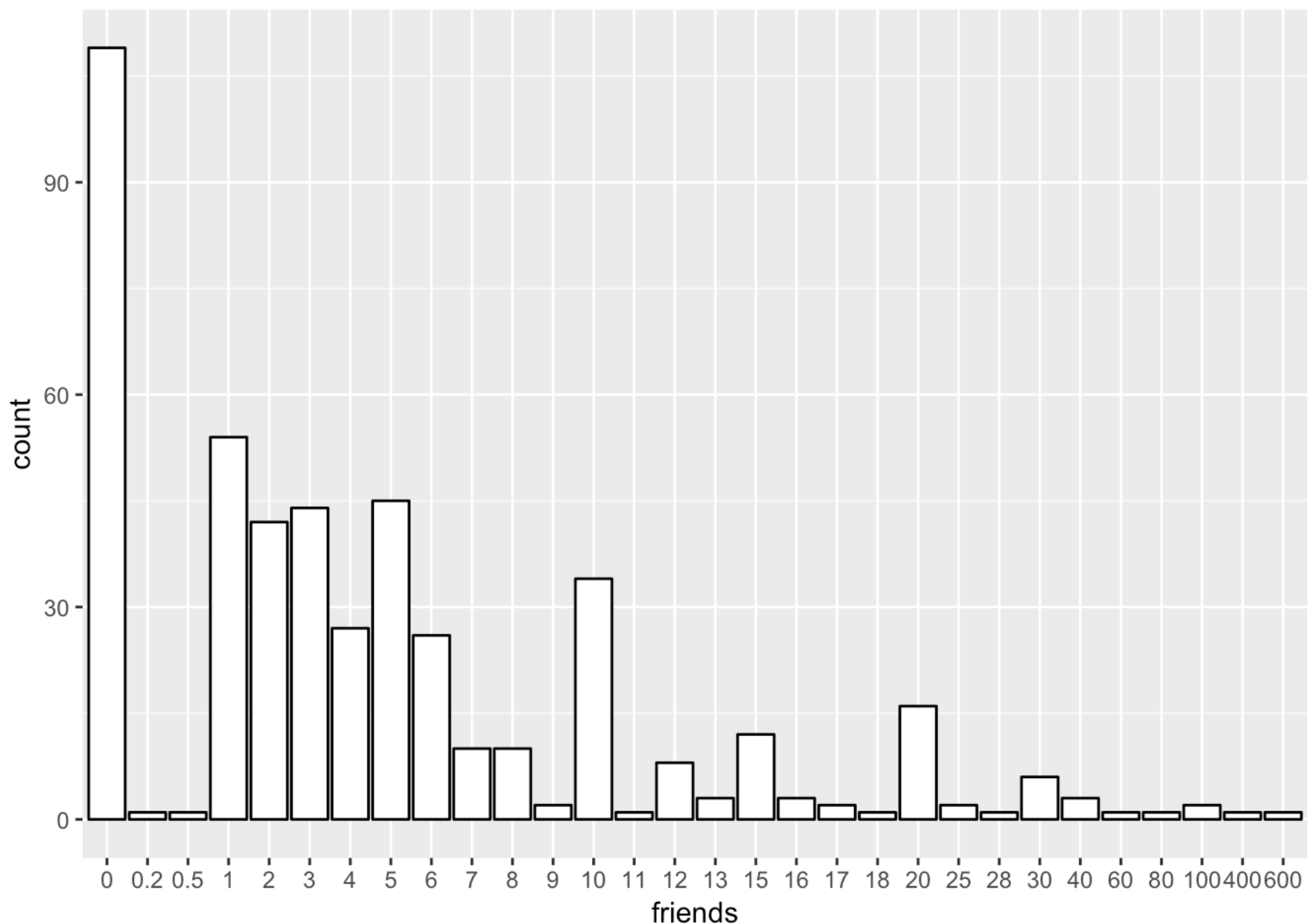
Import data

```
library(readr)
dataset <- read_csv("foreveralone.csv")
```

```
library(ggplot2)
library(tidyr)
library(tidyverse)
library(dplyr)
```

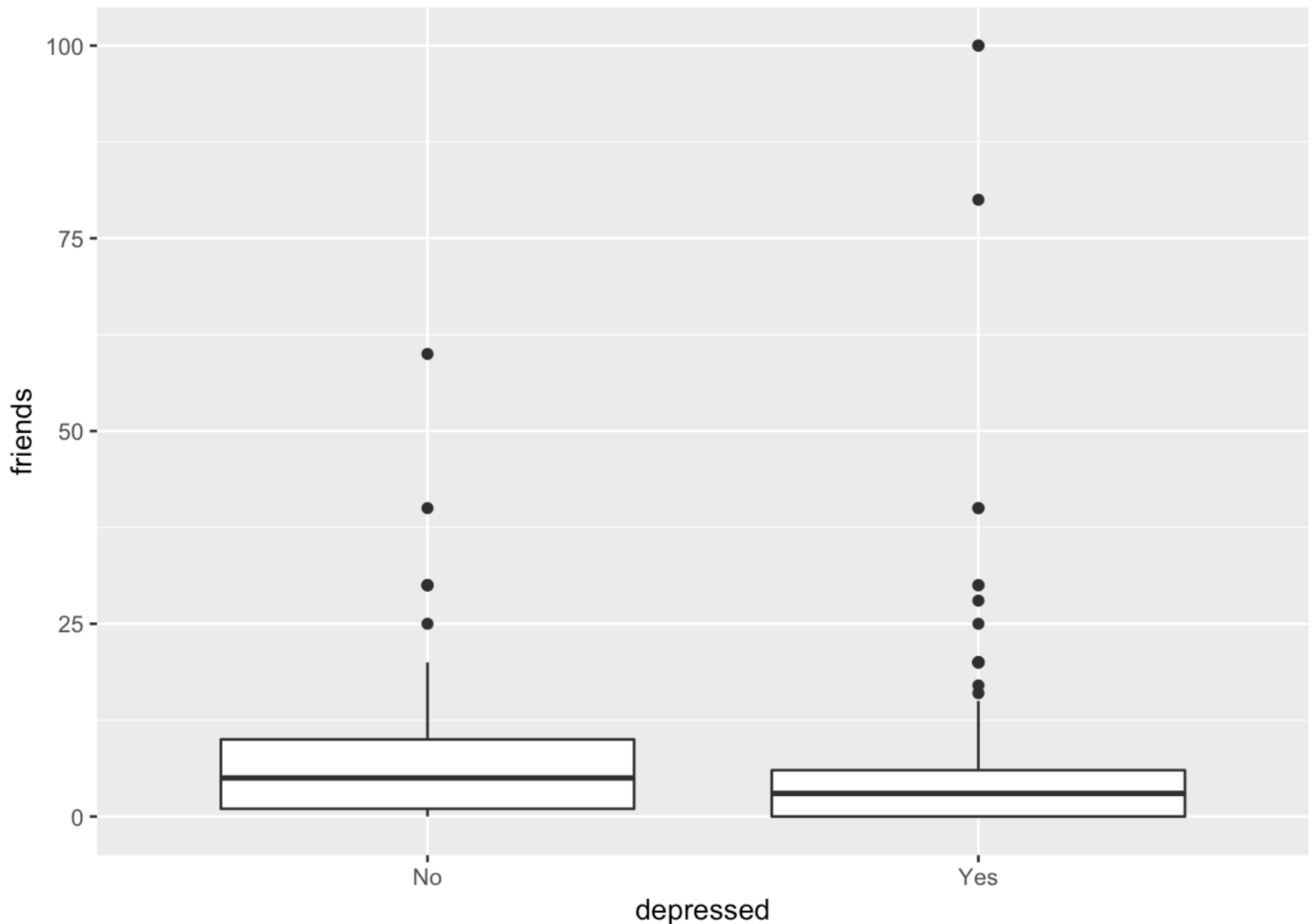
I will analyse the “friends” variable in the data set.

```
friends <- dataset %>% select(friends)
friends$friends <- as.factor(friends$friends)
g <- ggplot(friends,aes(x=friends)) + geom_bar(color = "black", fill="white")
g
```



From the above graph, I can see 2 impossible values (0.2, 0.5). In the final project, I will take them as 1 friend. Also the bars for 10, 20 are taller than bars around them. In addition, values over 30 are all multiples of 10. It shows that people tend to answer multiples of 10 when being asked to write a number.

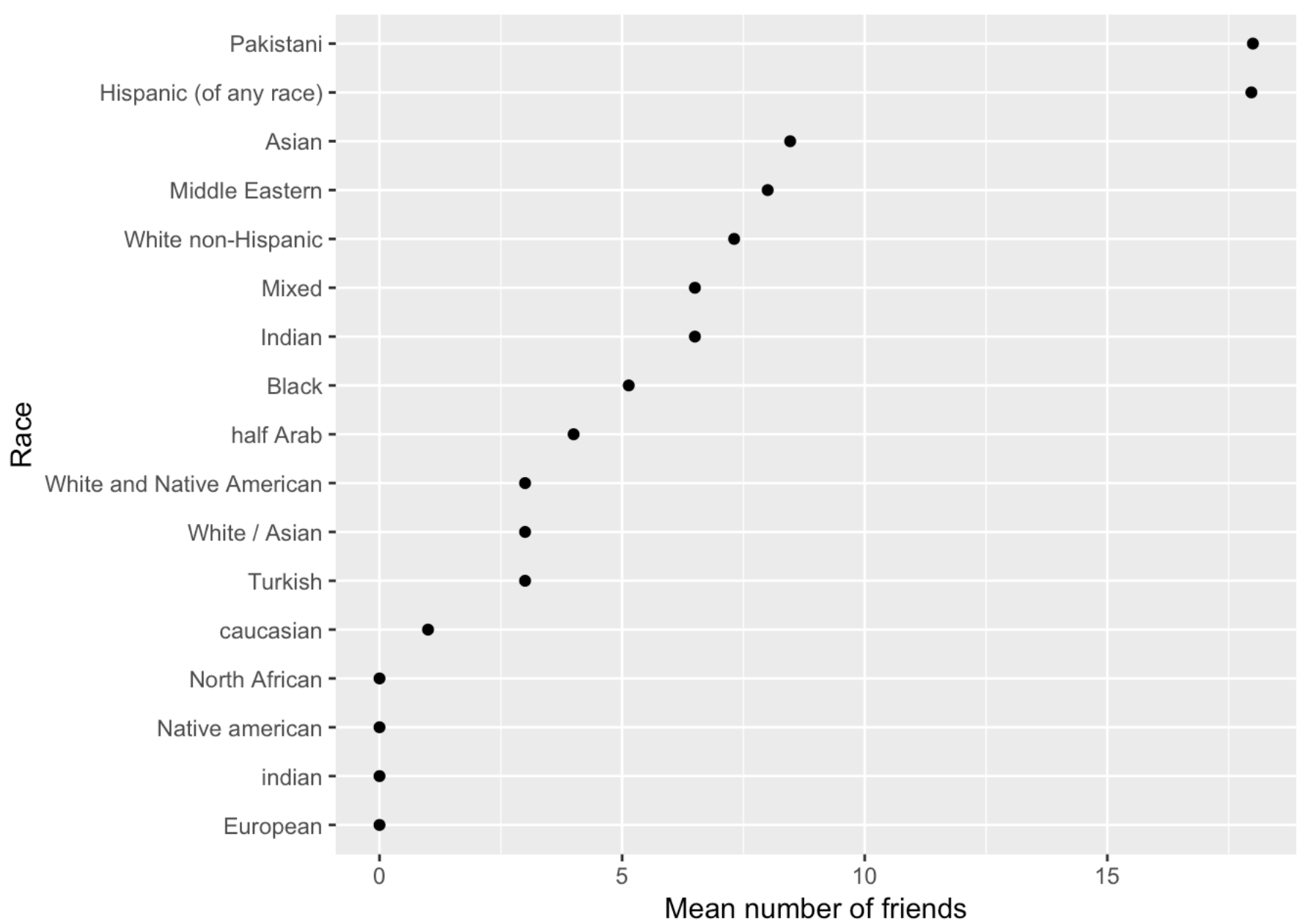
```
frddep <- dataset %>% select(friends, depressed)
frddep <- frddep[frddep$friends <= 100,]
g1<-ggplot(frddep, aes(x=depressed, y=friends)) + geom_boxplot()
g1
```



At first I thought a lack of friends might lead to a higher chance of depression. However from the plot, I discovered that number of friends does not seem to have a relationship with depression. Here for clarity I removed 400,600 friends which belongs to (No Depression) Category which means not depressed group has more extreme values. But the median and the hinges are close.

```
frdrace <- dataset %>% filter(!str_detect(race, "First") & race != "helicopterkin")
reps <- frdrace$race[str_detect(frdrace$race, fixed('mix', ignore_case=TRUE))]
reps = c(reps, "Multi")
reps2 = c("Half Asian half white", "white and asian")
frdrace$race[frdrace$race %in% reps] <- "Mixed"
frdrace$race[frdrace$race %in% reps2] <- "White / Asian"
frdrace <- frdrace %>% group_by(race) %>% summarise(Mean = mean(friends))

g2 <- ggplot(frdrace, aes(x = Mean, y = fct_reorder(race, Mean)))+geom_point() + ylab(
  "Race") + xlab("Mean number of friends")
g2
```



From the above graph we can see that among all races, Pakistani have the most friends. While North African, Native American, Indian, European all have average friend number 0. This may be caused by people from these races tend to not answer this question. In such case, 0 can either represent 0 friends or missing value, this makes it harder for us to analyze this data.