Homework 2

Jasmine Nakayama

February 14, 2018

Load packages:

```
library(car)
library(dplyr)
library(ggplot2)
```

Question 1: What kind of R object is the Davis dataset?

```
class(car::Davis)
```

```
## [1] "data.frame"
```

Question 2: How many observations are in the Davis dataset?

```
car::Davis%>%
  summarize(nObs=n())
```

```
##   nObs
## 1  200
```

Question 3: For reported weight, how many observations have a missing value?

```
sum(is.na(car::Davis$repwt))
```

```
## [1] 17
```

Question 4: How many observations have no missing values?

```
car::Davis%>%
filter(complete.cases(.))%>%
  summarize(n())
```

```
##   n()
## 1 181
```

Question 5: Create a subset containing only females. How many females are in this subset?

```
#create subset with only females
females<-car::Davis%>%
filter(sex=='F')
#count number in subset
summarize(females,n())
```

```
##   n()
## 1 112
```

Question 6: What is the average BMI for these individuals?

```
#create new variable with height in meters
davishtmeter<-car::Davis%>%
mutate(htmeter=height/100)
#create new variable for BMI ("the ratio of the weight in kilograms divided by the square of the
 height in meters")"
davisbmi<-davishtmeter%>%
  mutate(bmi=weight/htmeter^2)
#obtain average BMI
davisbmi%>%
  summarize(mean(bmi))
```

```
##   mean(bmi)
## 1  24.70096
```

Question 7: How do these individuals fall into the BMI categories (what are the frequencies and relative %'s)?

Category	BMI range (kg/m2)
Underweight	<18.5
Normal	18.5 to <25
Overweight	25 to <30
Obese	30 or higher

```
#create new categories for BMI ranges
davisbmicat<-davisbmi%>%
  mutate(bmicat=if_else(bmi<18.5, "Underweight",
                   if_else(bmi<25, "Normal",
                     if_else(bmi<30, "Overweight", "Obese","Missing"),
                   "Missing"),
               "Missing"))
#get frequencies & relative percentages

freq<-davisbmicat%>%
count(bmicat)

davisbmicat%>%
count(bmicat)%>%
mutate(percentage=prop.table(freq$n))
```
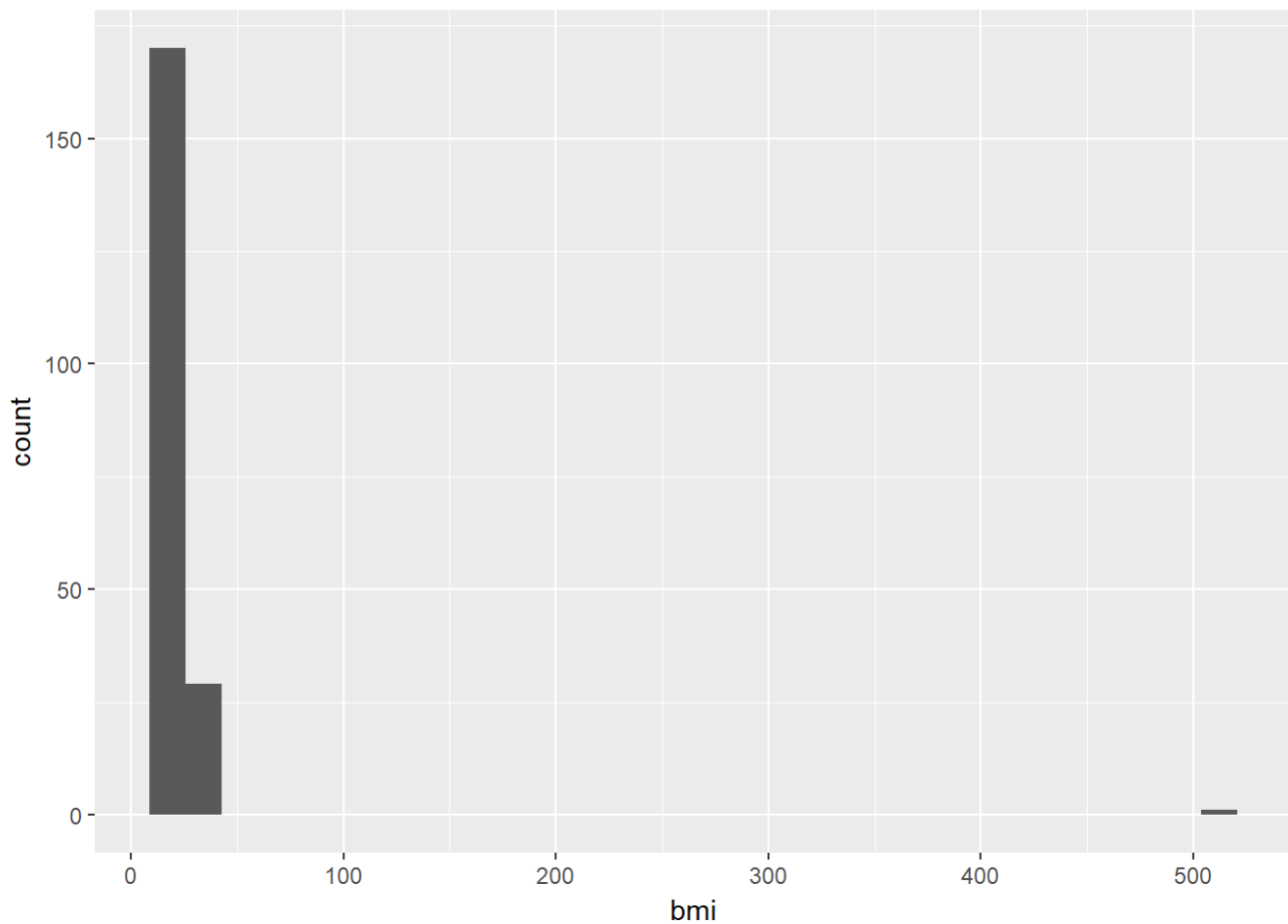
```
## # A tibble: 4 x 3
##   bmicat          n percentage
##   <chr>       <int>      <dbl>
## 1 Normal        143      0.715
## 2 Obese           4     0.0200
## 3 Overweight     35      0.175
## 4 Underweight    18     0.0900
```
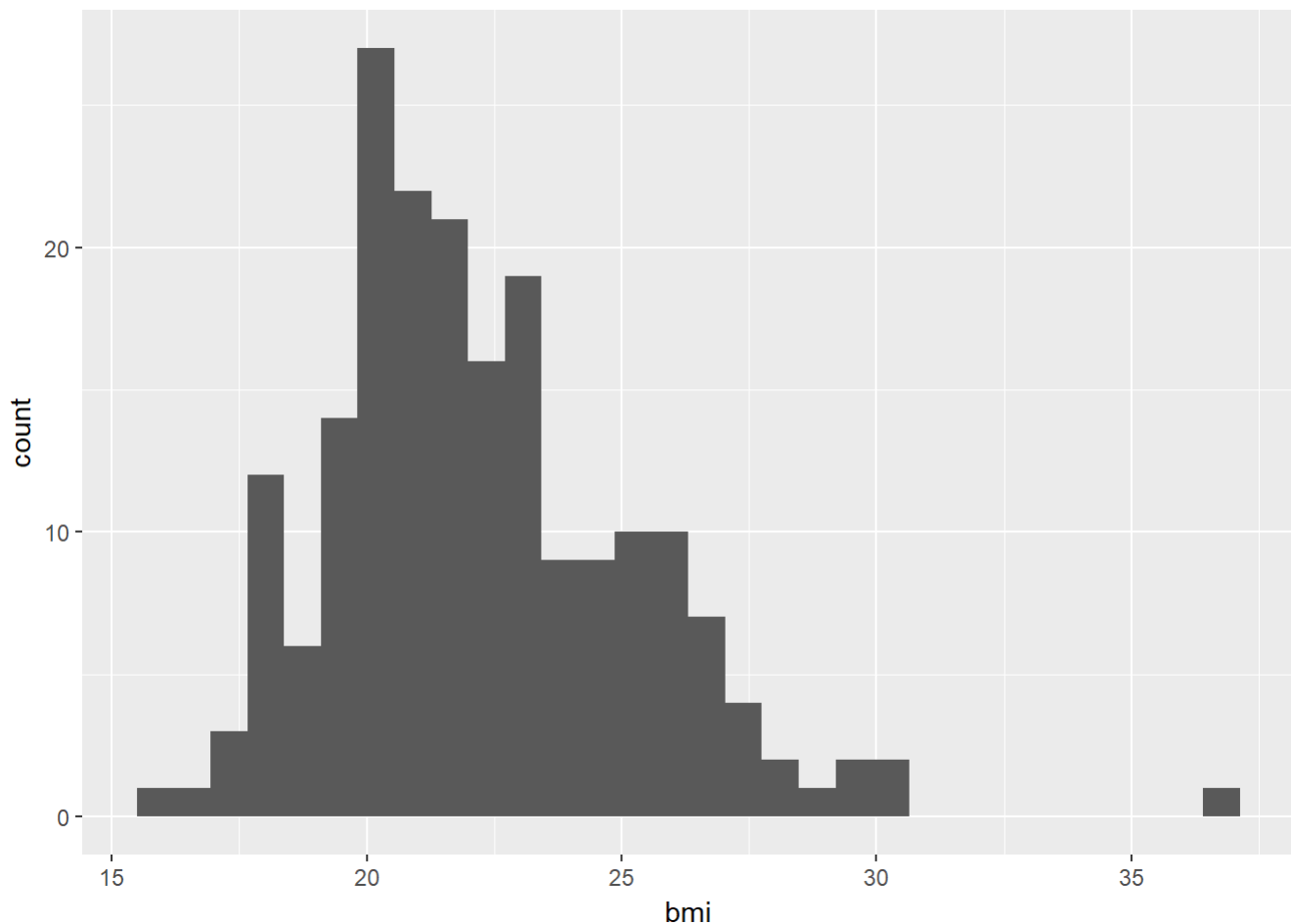
Question 8: Create a histogram of BMI.

```
davisbmicat%>%
ggplot()+geom_histogram(aes(bmi))
```

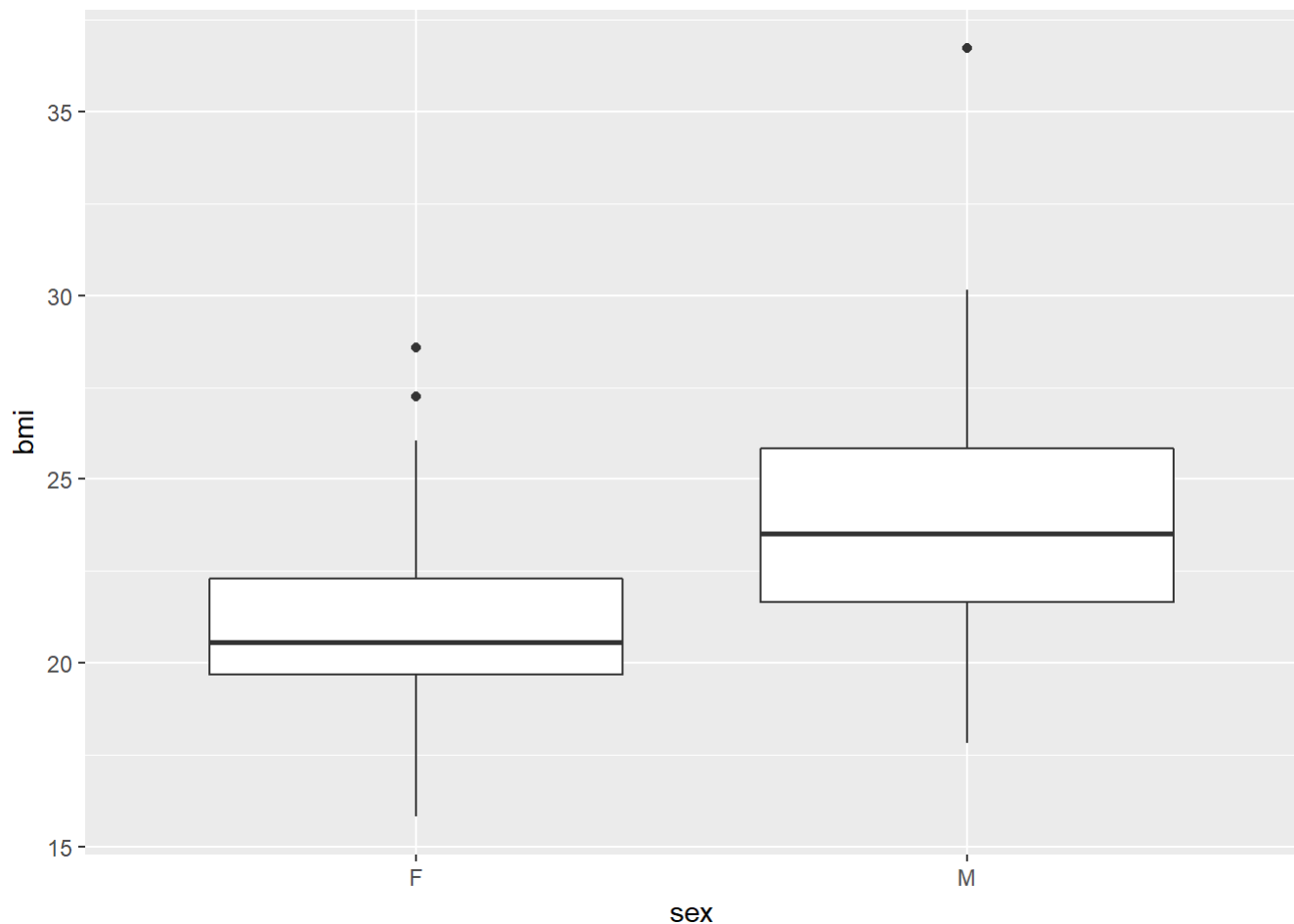There is one outlier whose BMI is over 500.

```
#remove outlier
newdavisbmicat<-davisbmicat%>%
  filter(bmi<100)
#look at new histogram
newdavisbmicat%>%
ggplot()+geom_histogram(aes(bmi))
```

Question 9: Create side-by-side boxplots of the BMI distributions by gender

```
#create boxplot of BMI distribution divided by gender
newdavisbmicat%>%
ggplot(aes(x=sex, y=bmi, fill=bmi))+geom_boxplot()
```

Question 10: Create a clustered bar chart of the BMI categories by gender

```
newdavisbmicat%>%
ggplot(aes(x=bmicat, fill=sex))+geom_bar(position="dodge")
```