

Homework 2: NRSG 741

Tommy Flynn^{*,a}

^aEmory School of Nursing 1520 Clifton Rd Atlanta GA 30322

Abstract

This assignment applies the dplyr and ggplot2 packages to work with the Davis dataset in the car package, which contains data on the measured and reported heights and weights of men and women engaged in regular exercise. The associated GitHub repository can be found at https://github.com/tommyflynn/N741_Homework/tree/master/Flynn_HW_02.

Using dplyr

Questions 1 & 2:

```
# 1. What kind of R object is the Davis dataset?
davis <- car::Davis
# to print class inline, use `r paste(class(davis))`

# 2. How many observations are in the Davis dataset? Q2: to print number of
# observations, use `r count(davis)`
```

Answer 1: The Davis dataset is the “data.frame” class of R object.

Answer 2: There are 200 observations.

Question 3:

```
# 3. For reported weight, how many observations have a missing value? Use
# the is.na function to filter non-missing values from the repwt variable,
# then show inline with `r count(repmissing)`
repmissing <- filter(davis, is.na(repwt))
```

Answer: Reported weight has 17 missing values

Question 4:

```
# 4. How many observations have missing values? HINT: find the complete
# rows... then show inline with `r completeobs`
completeobs <- count(na.omit(davis))
```

Answer: The Davis dataset has 181 complete observations.

*Corresponding Author
Email address: tjflynn@emory.edu (Tommy Flynn)

Question 5:

```
# 5. Create a subset containing only females. How many females are in this
# subset? create a subset of all female obs, print #obs inline with `r
# count(femaleset)
femaleset <- davis %>% subset(sex == "F")
```

Answer: A subset containing only females from the Davis data has 112 observations.

Question 6:

```
# 6. What is the average BMI for these individuals? create a new data.frame
# with height unit transformation (cm -> m), calculated BMI, and categorized
# BMI.
davisBMI <- davis %>% mutate(sqrmetHT = (height/100) * (height/100)) %>% mutate(BMI = weight/sqrmetHT) %>%
  mutate(BMIclass = if_else(BMI < 18.5, "1. Underweight", if_else(BMI < 25,
    "2. Normal", if_else(BMI < 30, "3. Overweight", "4. Obese", "Missing"),
    "Missing"))))
# Print for Q6 inline with, `r mean(davisBMI$BMI)`
```

Answer: The mean BMI is 24.7009556.

Question 7:

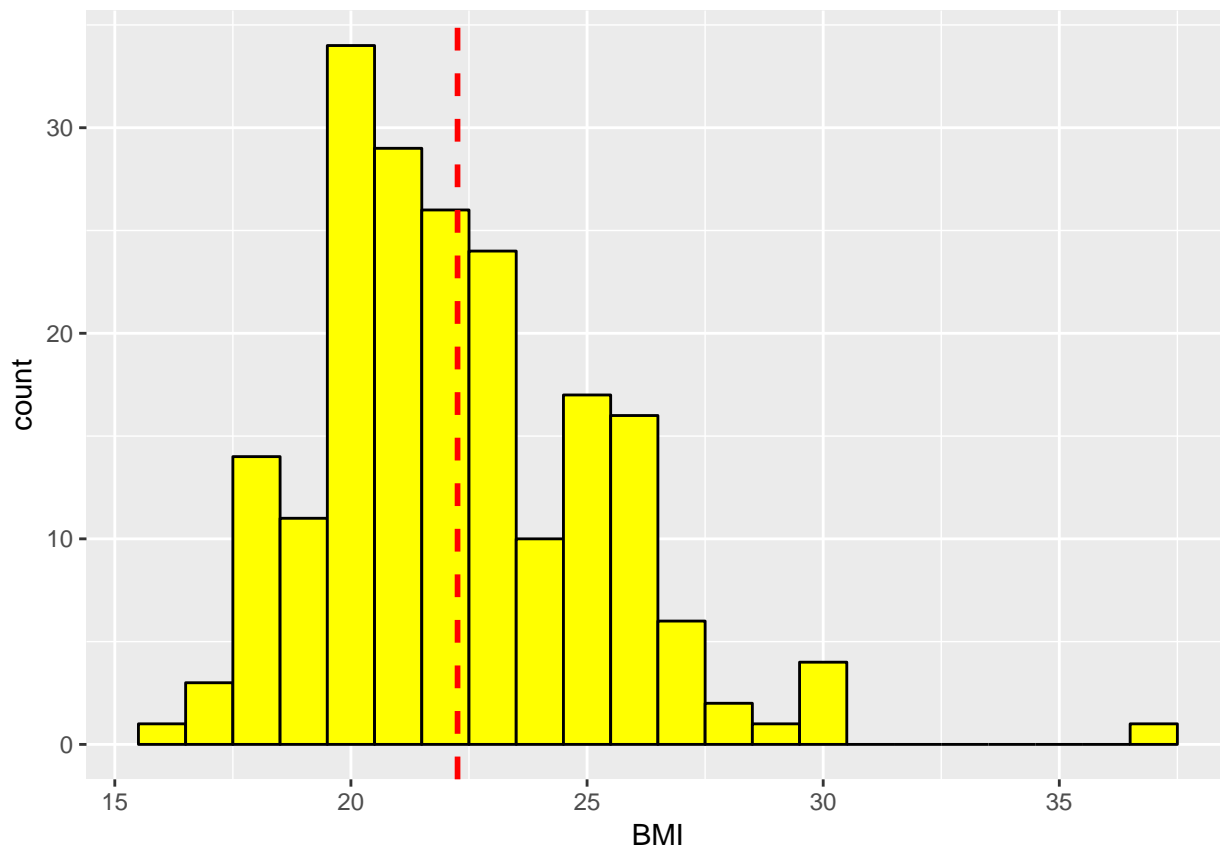
```
# 7. How do these individuals fall into the BMI categories (what are the
# frequencies and relative %'s)? Use janitor tably() and knitr kable()
# functions to make it pretty
davisBMI %>% janitor::tabyl(BMIclass) %>% knitr::kable()
```

BMIclass	n	percent
1. Underweight	18	0.090
2. Normal	143	0.715
3. Overweight	35	0.175
4. Obese	4	0.020

Graphs with ggplot2

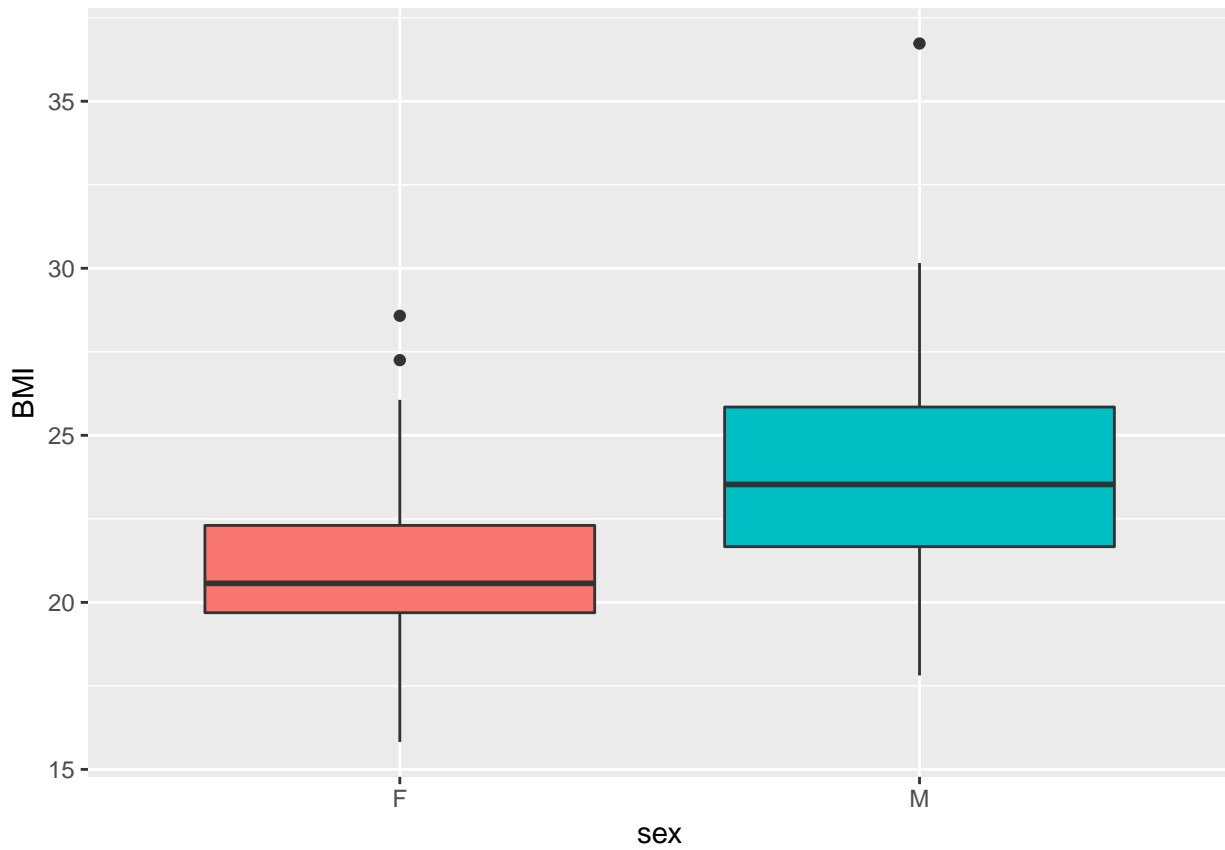
Question 8:

```
# 8. Create a histogram of BMI.
davisBMI %>% filter(!is.na(BMI)) %>% filter(BMI < 50) %>% ggplot(aes(BMI)) +
  geom_histogram(colour = "black", fill = "yellow", position = "dodge", binwidth = 1) +
  geom_vline(aes(xintercept = mean(BMI, na.rm = T)), colour = "red", linetype = "dashed",
    size = 1)
```



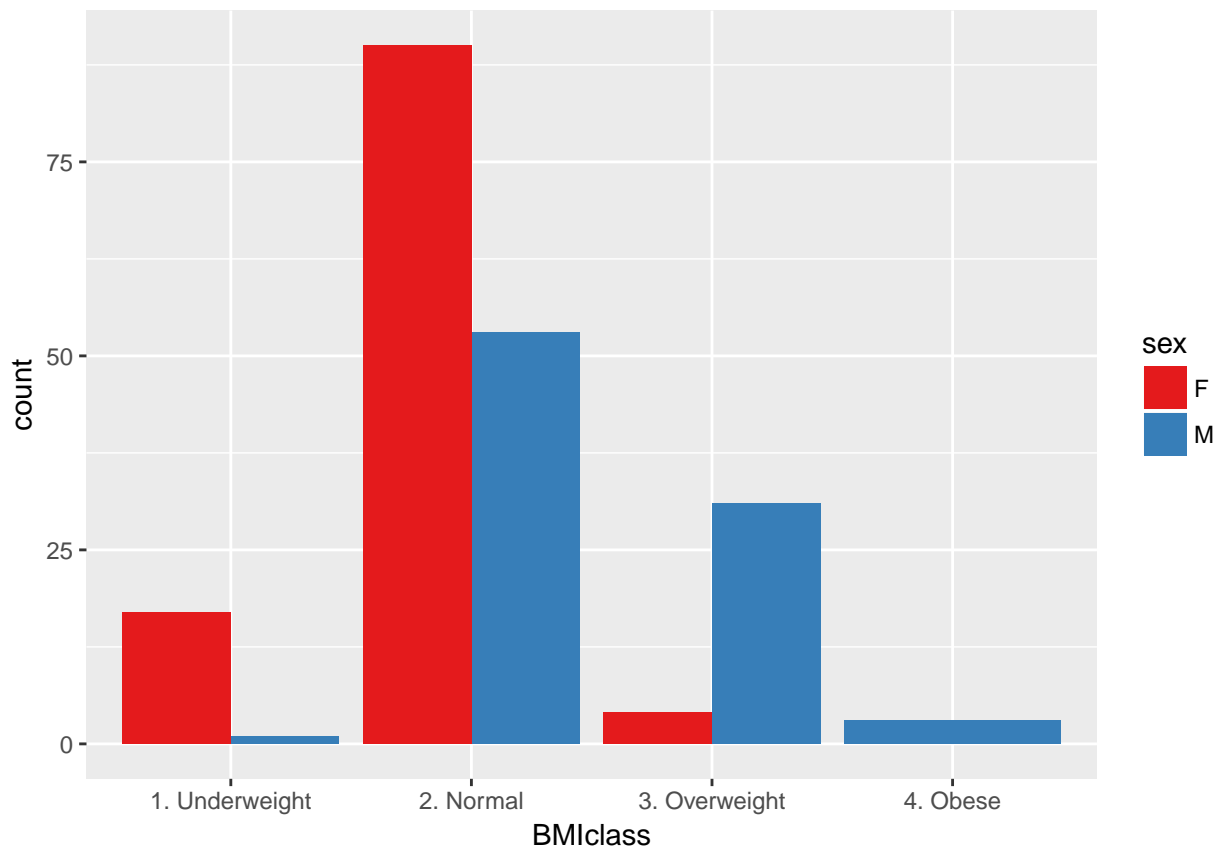
Question 9:

```
# 9. Create side-by-side boxplots of BMI distribution by gender.  
davisBMI %>% filter(BMI < 50) %>% filter(!is.na(BMI)) %>% filter(!is.na(sex)) %>%  
  ggplot(aes(x = sex, y = BMI, fill = sex)) + geom_boxplot() + guides(fill = FALSE)
```



Question 10:

```
davisBMI %>% filter(BMI < 50) %>% filter(!is.na(BMIclass)) %>% filter(!is.na(sex)) %>%
  ggplot(aes(BMIclass, fill = sex)) + geom_bar(position = "dodge") + scale_fill_brewer(palette = "Set1")
```



THE END