

# Homework 2

*Vicki Hertzberg and Melinda Higgins*

*2/14/2018*

## ***Due Date is 21 February 2018***

This homework is meant to further your `dplyr` and `ggplot2` skills.

First, install the package

- `car`

### **Installing the `car` package**

We found some hiccups when we were designing this homework. With a little sleuthing, we were able to figure out that some of the issues related to installing the package and dependent package called `quantreg`. So before you install `car` use the following R commands:

- `install.packages("quantreg", dependencies=TRUE)`
- `install.packages("car", dependencies=TRUE)`

You might get this question in the console:

"Do you want to install from sources the package which needs compilation" followed by a prompt for you to respond yes or no, which looks like

y/n:

Usually when you see this prompt in RStudio, y is a good default response. However when installing `quantreg` and `car`, we found that if you answered n to the prompts, all will work well. (*answering y here leads to other issues you can avoid for now... we don't want you to descend into R purgatory, LOL*)

### **The Data - Davis dataset in the `car` package**

The `Davis` dataset in the `car` package contains data on the measured and reported heights and weights of men and women engaged in regular exercise. *[For more information, type `?car::Davis` in the Console to bring up the HELP pages on the `Davis` dataset in the `car` package.]*

Use tools within the `dplyr` package as much as possible to answer the following questions.

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

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

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

**Question 4: How many observations have no missing values? (*HINT: find complete cases*)**

---

Create a subset containing only females.

**Question 5: How many females are in this subset?**

---

That last question was an opportunity for you to show-off your `dplyr` confidence.

Now return to the overall dataset with both males and females.

Body mass index is one way to quantify the amount of tissue mass (muscle, fat, and bone) in an individual, then categorize that person as *underweight*, *normal weight*, *overweight*, or *obese* according to that value.

We calculate the BMI as the **ratio of the weight in kilograms divided by the square of the height in meters**, and the categorization based on BMI is as follows:

### BMI Categories

Category	BMI range (kg/m <sup>2</sup> )
Underweight	<18.5
Normal	18.5 to <25
Overweight	25 to <30
Obese	30 or higher

Create the BMI variable and then a variable to depict BMI category. Note that the `height` variable is in centimeters, and `weight` is in kg. You need to create the BMI variable using the correct formula.

Now answer these questions:

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

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

---

### Test your graphing skills using `ggplot2`

Using the `Davis` dataset from the `car` package, create the following graphics/figures using `ggplot()` and associated `geom_xxx()` functions.

**Question 8: Create a histogram of BMI.**

*What do you notice about the distribution (any outliers or skewness)?*

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

*Remember to remove any outliers if needed*

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

*(note: the y-axis should be counts)*