IMT 573: Problem Set 1

A First Data Science Report - Soutions

Ali Qazi

Due: June 27, 2021

Collaborators:

Instructions: Before beginning this assignment, please ensure you have access to R and RStudio; this can be on your own personal computer or on the IMT 573 R Studio Cloud.

1. Download the `01_ps_firstreport.Rmd` file from Canvas or save a copy to your local directory on RStudio Cloud. Supply your solutions to the assignment by editing `01_ps_firstreport.Rmd`.
2. Replace the "YOUR NAME HERE" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.
3. Be sure to include well-documented (e.g. commented) code chucks, figures, and clearly written text chunk explanations as necessary. Any figures should be clearly labeled and appropriately referenced within the text. Be sure that each visualization adds value to your written explanation; avoid redundancy – you do no need four different visualizations of the same pattern.
4. Collaboration on problem sets is fun and useful, and we encourage it, but each student must turn in an individual write-up in their own words as well as code/work that is their own. Regardless of whether you work with others, what you turn in must be your own work; this includes code and interpretation of results. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.
5. All materials and resources that you use (with the exception of lecture slides) must be appropriately referenced within your assignment.
6. Remember partial credit will be awarded for each question for which a serious attempt at finding an answer has been shown. Students are emph{strongly} encouraged to attempt each question and to document their reasoning process even if they cannot find the correct answer. If you would like to include R code to show this process, but it does not run without errors you can do so with the `eval=FALSE` option as follows:

```
a + b # these object don't exist
# if you run this on its own it with give an error
```

7. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit`, download and rename the knitted PDF file to `ps1_YourLastName_YourFirstName.pdf`, and submit the PDF file on Canvas.

Setup In this problem set you will need, at minimum, the following R packages.

```
# Load standard libraries
library(tidyverse) # This library gives us access to all the functions we will use
library(datasets)
```

Problem 1: Load a Dataset

Choose a dataset from the `datasets` package in R. Recall, this is the same package that was used in this module's lab exercise. In the lab, we used the `faithful` dataset. Choose a different dataset for this problem. Load the data in R and provide a brief description of the data contained. textcolor{red}{Hint: you might find the code `library(help = "datasets")` useful here.}

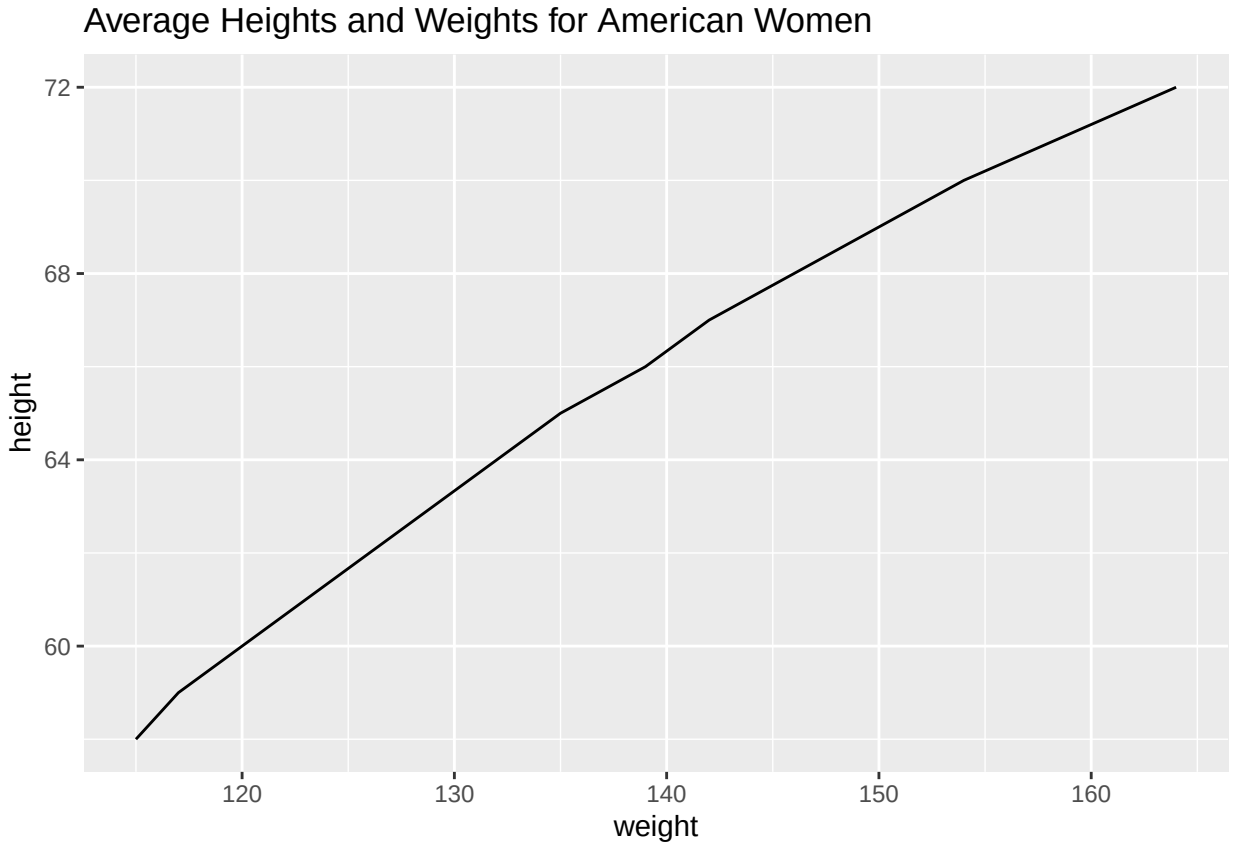
library(help = "datasets")

```
data("women") #load data to workspace
```

Problem 2: Create a Visual of the Data

Create one visualization using the data chosen in Problem 1.

```
p1 <- ggplot() + geom_line(aes(y = height, x = weight) , women)
p1 + labs(title = "Average Heights and Weights for American Women")
```



Problem 3: Reflection

As we enter this course, we will engage in consistent programming for data science tasks and interpretation. Please provide a short paragraph that discusses your level of comfort with data programming in R. Be sure to highlight any areas in which you anticipate needing additional support. This information will help your instructional team provide better examples and feedback throughout this course.

I am somewhat comfortable with R. I learned R on my own a few years back and will need to polish my skills but I believe with the right support I will be okay in R. As long as support is there and it is helpful I believe I will be okay.