STAT218 Course Project

Overview

- This project is to be completed using R Markdown and at each step, submitted as a PDF file into the appropriate Assignment folder on Quercus. You should be inserting code in the appropriate code chunks throughout this file.
- While it is my preference that you directly knit to PDF and I will have a Zoom screen-share tech session with anyone that asks, you may knit to HTML or Word and then convert it to PDF.
- There are **three** steps to completing your project; details to follow at the end of this document.
- You will be marked based solely on what is visible in the submitted document. If you have code that is commented out in your .rmd file and therefore does not display in the PDF document, it will not be marked. While I do not require you to submit your .rmd file, you need to be able to supply it to me upon request.
- Your course project, in its final form, is due at **11:59 p.m. on Friday, November 19** and must be submitted in the Assignments-Course Project folder on Quercus.

Academic Integrity

- This project is to represent your own work and should be written up independently. That being said:
  - it is okay to help eachother find a good data set;
  - it is okay to ask your classmates general questions in R such as "What is the command to get residuals?" or "How do I change the axis labels in a scatterplot?";
  - it is okay to ask your classmates where in the notes you can find similar examples to certain questions;
  - it is explicitly **not** okay to send your exact code to a classmate who asks you how you did a question. Point them instead towards the correct function or a similar problem in the notes;
  - if you are getting an error in your code that is preventing you from knitting your project and it requires someone to look at your work, please send it to myself or your TA to look at. In order to keep working on your project while waiting for your me or your TA to respond, it may be possible to comment out the part that is causing issues and continue working.

Some Useful Tips

- Please do **not** spend hours being frustrated working through a problem because there is an error in your code. While you should spend some time trying to find your error, a good resource is Google. If, however, you are finding yourself spending a lot of time on one error, please contact me. I may ask what you have tried so far, and I will try to help as much as possible. If you have not done anything, I may direct you to the course materials or some other resource.

- Knit often. After everything that you do (each question part or every line of typeset math you do), knit your file. This will help you find your error immediately. It is better to narrow your error down to a few lines than several hundred. As soon as you have an error, your whole project will fail to load so you must find it quickly.
- You should attempt the parts (and their sub-parts) as we progress through the course. Once your data set has been approved, there are many parts you can do immediately. A benefit to doing it in pieces is that you can get even more practice with concepts and coding as we move through the course material.
- Do **not** delete or comment out my text; it is necessary for the file to be comprehensible. If there's an R code chunk you don't use, or there's a spot where I say to delete my comment and write your answer, and you don't use it, that is a clear sign you have missed something.
- Summary: Start early, knit often, ask for help **BEFORE** you waste a lot of time.

Detailed Instructions

Your project contains **three steps** (2 official and one optional) and is worth a total of **50 marks**.

Step 1: The first step in your project is selecting a data set. To see the some of data sets "built-in" R, you can use the package MASS and type data(), i.e.:

```
library(MASS)
data()
```

We need to ensure the data set is suitable, so in this step you will select data set and then do some calculations to make sure you have a data set that you can work with for the rest of the project. This is **Part A** in your course project scaffold and is worth **5 marks**. **No two students can use the same data set.** Whoever submits their Part A successfully first (including all the necessary calculations and comments) gets to use that data set; anyone submitting after (and requesting to use the same data set) will be told the data set cannot be used. This step is due **October 18 at 11:59pm** on Quercus. You will receive feedback about your data set within approximately one week of submission. **Late** submissions will be accepted up to a maximum of 48 hours late (October 20 at 11:59pm), with a penalty of **1 mark per day**. Beyond 48 hours, you forfeit all 5 marks and will still be required to find a suitable data set in order to proceed.

If you are required to choose another data set, at the instruction of myself or your TA, you will be given 72 hours to recomplete this step; if it is submitted beyond 72 hours of feedback, the late penalty of **1 mark per day** will apply. Beyond 72 hours, you forfeit all 5 marks. Again, you must submit as many times as necessary until you have a suitble data set or risk losing many more marks.

Step 2: This is an **optional** step. In this step, assuming you have successfully completed Part A, you are to complete and submit Part B of your project. This interim step, which includes parts of the final submission, gives us a chance to give you some feedback and checking you are on the right track in terms of coding, commenting your code and showing your work. It is worth **5 bonus marks**. If you would like to complete this step, your project containing Parts A and B should be submitted by **October 25 at 11:59pm** on Quercus. You will receive feedback about your work thus far within approximately one week of submission. Late submissions will **not** be accepted.

Step 3: In this final step, you submit your entire project. This means, in addition to Part A, Parts B-E should be completed. Remember you should be knitting to PDF. This step is due **November 19 at 11:59pm** on Quercus. You may resubmit your final course project as many times as needed up until November 19 at 11:59pm. After this, a new Assignment folder will be constructed for **late submissions** where you may only submit once. Please read the course syllabus regarding late submissions. For each day late, to a maximum of **5 days**, **5%** (of the entire grade of your project) will be deducted. **No** submissions will be accepted after November 24 at 11:59pm.