Mini-Project 1
Math 241

Spring 2020

Due: Thursday February 27th at 8:30 am

Project Requirements/Goals

The following tasks should be done as a group.

- Determine what data you want to wrap into an R package. Claim your data source in the group assignments sheet.
  - **No two groups can create a package of the same data.** First come, first serve.
  - If the data are from the internet, post a link to the data source.
- Create an R data package.
- Write a short (~ 500 words +/ 100 words) blog post. The post should:
  - Introduce the data and address what/who the data represent.
  - Include at least one well-chosen example.
  - Communicate uncertainty when appropriate.
  - Include 1 (at most 2) data visualization(s).
- Decide if you would like to share the post on the course website.
  - If so, determine the `name` of the url you would like me to use to host your package (e.g., https://github.com/Reed-Statistics/INSERT_NAME).
  - I will then clone/download your private group repo and place it in the public repo.

- Be ready to give a 5 minute introduction to your R data package in class.
  - You should create 2-5 Google Slides.
  - Add your slides to this presentation.

The following task should be completed individually:

- Complete the following form to tell me whether or not you would like your group's blog post to be made public and to provide feedback of the division of labor.

Group Assignments

- You can find your group assignment here.
- I have also created a group repository on GitHub for collaborating on your R data package.

Rubric

I will grade the data package and blog post by accessing them directly in your group GitHub repo. Make sure it is all finalized (and extraneous materials removed) by the due date.

When I assess your data package, I will be looking for the following:

- The data include at least 4 variables and at least 30 observations.
- There are no errors or fatal warnings when I check the package.

- There is a complete and accurate DESCRIPTION file.
- The help file (i.e. data dictionary) is informative and clear and it cites the data source.
- The readme file gives me a quick idea of how I might want to use or explore these data.

When I assess your blog post, I will be looking for the following:

- The items bulleted above for what the blog should include.
- Any data visualizations make appropriate use of geoms and their aesthetics, scale, and context and are also clear and engaging.
- The post is well organized and well-written.
- That the post is creative and engaging.
- Any statistical arguments are clear and accurate.
- The writing is for a non-technical audience.

When I assess your 5 minute presentation, I will be looking for the following:

- Each group member participated.
- A clear and concise description of the data.
- A hook to get people interested in using the data themselves.