STAT 331 Final Project

Item	Due	Points
Phase 1: Group Selection	Thursday, May 11 by 4:00pm	5
Phase 2: Research Due	Thursday, May 25 by 4:00pm	10
Phase 3: Final Submission Due	Monday, June 12 by 5:00pm	80

In this group assignment you will complete a comprehensive report exploring both new statistics and new features of R. You may choose your groups, but there must be 3 people in each group if possible.

Shiny, by RStudio, is a web application framework for R. With Shiny apps we have the ability to bring the power of R to people who don't know how to program and don't really need to. Explore the gallery at shiny.rstudio.com to get a sense of what's possible with Shiny.

Shiny Apps can be especially useful and appropriate for exploring/analyzing data in a more interactive way.

Your task in this project consists of building a Shiny app to explore/analyze a rich data set or collection of data sets of your choice.

As an example to jumpstart your thoughts, consider www.gapminder.org (under gapminder world). This is a wonderful data visualization tool that you should aim to emulate in at least some respects: being able to choose which variables to look at and being able to look at as many as 4+ variables at a time by using a scatterplot, color, and size. However, you should also incorporate analysis/inference into your app as well (i.e. confidence intervals, hypothesis testing, regression, and/or ANOVA).

Another possible source of data sets is https://blog.rstudio.org/2014/07/23/new-data-packages/

You must use at least 2 packages presented on in the video project. At least 1 of these 2 must be something other than dplyr/magrittr or ggplot2. You're encouraged to use more than 2, but 2 is the minimum.

Phase 1: Group Selection

Please submit a document to PolyLearn as a group that contains your group member names. Only one group member needs to submit something to Poly Learn.

Phase 2: Research/Draft

Submit a document created in R Notebook that outlines your research for the project. This should include an explanation of the statistical concepts addressed, an outline of your approach in R (ie, packages/functions), and any references used (including books, websites, and data sources).

Two things your draft must have:

1) A description of your data (variables and size) for my approval
2) A draft template/drawing of how you envision your Shiny app looking/working (this will be separate, submitted in class)

Keep in mind the grading rubric as you are creating your draft.

Phase 3: Final Submission

Your final project submission will consist of two things:

1) The Shiny App you've built and all relevant (data and code) files (in a .zip folder)
2) A brief R Notebook (HTML or PDF) document explaining how to use your app

While I'm asking for a separate report on how to use your app, you should include helpful tips and instructions within the app itself as well. Check with me if you have questions or concerns.

Your app should include functionality for exploring the data graphically and numerically, as well as provide some statistical tools for analysis. It might work well to put each of these in its own tab. Users should be able to explore different variables with different visualizations and summaries using menus and/or widgets in the app.

Grading Rubric

At each phase of the project, a 20% penalty will be applied for any late submissions.

____/5 Phase 1: Group Selection

____/5 Group members identified

____/10 Phase 2: Research

____/2 Explanation of the statistical concepts

____/2 Outline the approach in R (ie, packages/functions)

____/2 References any outside sources used, including books, websites, and data sources

____/2 Description of the data to be used

____/2 Template/diagram of Shiny app appearance

____/85 Phase 3: Final Submission

____/3 Introduction to the project (Report)

____/3 Explanation of the statistical concepts (Report)

____/3 Appropriately divided into sections (Report)

____/3 Produces plots to illustrate results (Report)

____/3 Plots are adjusted to publication quality (Report)

____/3 Provides readable and commented R code (App)

____/3 Explains results (Report)

____/3 Provides an overall conclusion (Report)

____/6 Overall quality (organization, grammar, explanation, etc.) (Report)

____/10 Clean, user-friendly interface with simple instructions for use (App)

____/15 Provides ample visualization options for exploring data graphically (App)

____/15 Provides ample summarization options for exploring data numerically (App)

____/15 Provides options for analysis of various univariate and bivariate relationships in data (App)

____/100 Final Grade