STAT 331 Final Project

Item	Due	Points
Phase 1: Group Selection & Data	Friday, November 2 by 5:00pm	5
Phase 2: Visualizations Due	Friday, November 9 by 5:00pm	20
Phase 3: Initial App with Graphs	Friday, November 16 by 5:00pm	20
Phase 4: Final Submission Due	Monday, December 10 by 5:00pm	85

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 & Data

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.

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

Your data set(s) should include at least 2 categorical variables of interest and at least 2 quantitative variables of interest.

This must include:

1) A description of your data (variables and size) for my approval

Phase 2: Visualizations

Submit a document created in R Notebook that gives multiple examples of the visualizations you will include in your app using your data.

Things to keep in mind:

1) At least some of these plots will be interactive in the app (not in this R Notebook), so be sure to include examples using different sets of variables
2) You may want to use the leaflet package in your app to map certain data. In this R Notebook you may use leaflet or other mapping packages (maps, ggmap, etc.) to give examples.

Phase 3: Initial App with Graphs

Submit a first draft of your app with the following:

1) The visualizations you created in Phase 2
2) Interactive functionality added to the visualizations
3) This app should be uploaded to shinyapps.io. Your submission for this phase should consist of a document with your group member names and a link to your app on shinyapps.io.

Phase 4: Final Submission

Your final project submission will consist of two things:

1) Upload your completed app to shinyapps.io
2) Submit on Poly Learn a .zip folder containing the following:
    a. The shiny app you've built and all relevant (data and code) files
    b. An R Notebook (HTML or PDF) document explaining how to use your app as well as some conclusions about your data. **A link to your app on shinyapps.io must be included in this document.**

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.

Again, consult the rubric at the end of this document for guidance on what your app should do and include.

- Descriptive statistics for the variables of interest should be accessible in your app.
- Graphical summaries (your visualizations from Phase 2/3) for the variables of interest should be accessible in your app
- At least one inferential component is included: confidence interval(s), hypothesis test(s), regression(s), etc.

Grading Rubric

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

____/5 Phase 1: Group Selection & Data

____/1 Group members identified

____/4 Research and data described

____/20 Phase 2: Visualizations

____/10 Multiple examples of multiple types of graphs

____/10 Graphs are well labeled and organized

____/20 Phase 3: Initial App with Graphs

____/10 Working app with no errors

____/10 Graphs are interactive

____/85 Phase 4: 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)

____/130 Final Grade