STAT 413/613 Final Project

Project Instructions and Rubric

Richard Ressler

2020-09-05

Contents

1 Description

For the final project you will work in teams of 2 to 4 people to complete this assignment *using tools from the class.* Students enrolled in 413 will create a Shiny App for data analysis. Students in 613 will create a Shiny App for both data analysis and interactive statistical analysis/modeling.

This is a rapid development project, a common event for data scientists. Consider it a two month-long "Hackathon" where collaboration and focus are key to success.

1.1 Groups

This is a group project. Everyone must join a two to four person group. Groups may include people from from either course (413 or 613). You are expected to form groups on your own through rapid collaboration across email, chat, phone, Zoom, etc.. The Canvas Collaborations spreadsheet has everyone's email and GitHub ID. The assignment will be a group repo in GitHub and a group assignment in Canvas.

Submit an assignment response on Canvas by September 22nd with the members of your group and the general topic. Also update the spreadsheet in Canvas collaborations with your group and topic.

1.2 Project Goal

This is *not* a data analysis project to answer a hypothesis *per se*, like you may have done in 412/612.

Each group is to rapidly build a Shiny App to enable someone to investigate your data set and conduct their own graphical analysis and statistical analysis/modeling (across a limited set of options) to obtain and interpret multiple answers (based on their chosen options). In the course of doing so, you will follow a thread that produces an answer based on the choices you adopt for your vignette and oral presentation, but that is secondary to enabling choices in the App.

1.3 Topic

Each Group will select a topic of interest agreed on by the group. Recent topics have included apps on COVID-related data, social media data, DC crime data, Florida incarceration data, small business loans, or even DC water management. There is no restriction other than it **cannot be the same topic (or at least data set) you worked on for STAT 412/612 or another course at AU**.

1.4 References

Identify at least three references from either academic papers or widely-available authoritative sources that provide context for your topic and use case. Identify the papers/sources in your vignette. You may also refer to them in your oral presentation to set context.

1.5 Data Source

Each group should select as a data source, a web-based site that is the primary owner or authoritative provider of original data (**not** a secondary source of cleaned data like Kaggle.com).

The data set should include recent (2018 or later) data with a reasonable number of records. While it is expected that every data set will require some cleaning, be careful to focus the number of variables of interest so as to minimize the amount of effort spent cleaning and tidying the data.

1.6 Use Case

Given a topic and data source, each group should define a Use Case for its app. The Use Case should specify "who" the typical user is expected to be and the typical questions your App should help the user answer. Specifying a use case allows you to make some assumptions about the skills and knowledge of the typical user to focus your App development and shape your user interface design.

I expect the Use Case and App scope (number of questions) and scale (number of App elements) to be greater for groups of four people than for groups of two.

1.7 Caveats on Scope and Scale for Rapid Development

There is limited time to produce the final deliverables. **Do NOT try to "Solve World Hunger" or "Boil the Ocean".** A key part of data science is shaping or focusing a project to produce rapid results against specific requirements.

While the scope and scale will be greater for groups of four people than for groups of two, the evaluation rubric is not about scale and scope *per se*.

A successful project will cover all the elements in the rubrics so as to produce a **"minimum viable product"** that *works* and *supports the specified use case.*

1.8 GitHub and Group Collaboration

There are multiple methods available for collaboration.

- GitHub. Once your group is identified, I will create a repo for it as an assignment on GitHub. Group members will receive a link with an invite to join. Each Group Member will have write privileges by default. All written deliverables should be managed within GitHub using appropriate branches, pushes, pull requests, and merge methods as agreed upon by your group.
- I will create groups in Canvas so you can use the chat and files function as well as set up meetings on your own using Zoom. Only members of the group (plus the TA and me) will have access.
- Zoom. If you sign in with your AU email using the `@american.edu` version (not the `@student.american.edu`), you should be able to schedule meetings as you wish.

1.9 External Collaboration

- Do not discuss your group project with other groups or classes.
- You may ask me questions at any time.
- **Recommend scheduling a session with me before Thanksgiving break**, of about 15 to 30-minutes, to discuss your group's approach and progress. It's a good time to level expectations and discuss questions or feedback on your project plan.

2 Project Deliverables, Due Dates, and Grading Rubrics

2.1 Deliverables

There are four Deliverables for the Final Project.

Deliverable	Points	Submission	Due Dates
Project Plan	10	Submit HTML and .Rmd on GitHub and a note in Canvas	10/21
Vignette	20	Submit HTML and .Rmd on GitHub and a note in Canvas	12/8
Shiny App	20	Submit Complete Repo on GitHub and a note in Canvas	12/8
Oral Demonstration	20	Delivered in Class	12/8

2.2 Project Plan (10)

- Write a one to two page report in R Markdown. Include the following elements:
  - App Title and Names of your group members (1)
  - Description of the problem to be analyzed in the Shiny App (1)
  - Proposed data sources. (1)
  - Concept for the Shiny App
    - Overall Layout (1)
    - Data Analysis and Visualization with User Choices and Results (2)
    - Statistical Modeling with User Choices and Results (2)
      - If group has only 413 students, the points go to Data Analysis and Visualization
  - Allocation of Responsibilities for the team (1)
  - Project Steps/Schedule (1)

2.3 Shiny App (20)

- If any groups are only 413 students, they should have a second data visualization tab if they do not want to create a Statistical Analysis/Model.

2.3.1 Rubric

Element	Criteria	Pts
Overall App	App Works. Has appropriate Theme. User-friendly Appearance	2
EDA Layout	Well-Organized and Easy to Use	2
EDA Controls	Sufficient, Flexible, and Clearly Labeled Inputs	2
EDA Outputs	Appropriate, Accurate and Clearly Labeled Results	2
Statistical Analysis Layout	Well-Organized and Easy to Use	2
Statistical Analysis Controls	Sufficient, Flexible, and Clearly Labeled Inputs	2
Statistical Analysis Outputs	Appropriate, Accurate and Clearly Labeled Results	2
GitHub Repo Organization	Proper Folder Structure	1
GitHub Repo Commits	Regular Updates on GitHub, Effective use of Branches	2
GitHub Repo Division of Labor	Team Effort on GitHub	2
GitHub Repo README.md	Clear and complete	1
	Total	**20**

2.4 Vignette (10)

The vignette should be a few pages written in R Markdown and compiled to HTML.
It should describe the problem the App is designed to solve and then show the reader how to solve it using the App.

2.4.1 Rubric

Element	Criteria	Pts
YAML Header	Title, Author, Date, and Output	1
Use Case	Describes the problem the App is designed to solve and by whom	1
Required Packages	Describes required packages and versions used in the App	1
Data Source and Structure	Identifies the default data source. Describes the general structure of the data.	2
EDA Inputs, Controls, and Outputs	Briefly walks through the Use Case describing how to solve the problem and interpret results	2
Statistical Analysis Inputs, Controls, and Outputs	Briefly Walks through the Use Case describing how to solve the problem and interpret results	2
References	At least three relevant to the App	1
	Total	10

2.5 Demonstration (20 Pts)

- Each group will make a *15-minute* oral demonstration to the class using *only the Shiny App* (no slides).
  - Walk through the use case in the vignette using the app.
  - **Everyone must participate as part of the demonstration.** If there is an unavoidable conflict and you cannot participate live on Wednesday 5:30PM, please coordinate with your group and me for alternative arrangements, e.g., recording your portion of the demonstration or presenting live to me at another time.
  - No slides or R code should be shown during the demonstration.
- When not presenting, students will fill out a peer-evaluation assessment on each of the other demonstrations.

2.5.1 Rubric

Element	Criteria	Pts
Introduction		2
	Conveys the context of the problem the app is designed to address Identifies the specific use case: Who is the intended user and what kinds of insights should they be able to get from using the app.	
Presentation of Data Sources		4
	Identifies the data source to include the “as of”date Describes how the data was obtained Summarizes how the data was cleaned and tidy’ed Identifies any assumptions or limitations about the data	

Element	Criteria	Pts
Demonstration of Data Analysis		3
	Follows a logical order Explains how to make choices and the rationales for choices Explains how to interpret the results	
Demonstration of Statistical Analysis/Modeling		3
	If only 413 students, points go to Presentation of Data Analysis Flow Follows a logical order Explains how to make choices and the rationales for choices Explains how to interpret the results	
Summation		3
	Summarizes the Use Case Summarizes the Data Sources Summarizes the Analysis and Results	
Demonstrator Performance	**Individual Grades** Demonstrates knowledge of material Oral presentation is clear, confident, and concise Answers to any questions are responsive	3
Peer Assessments		2
	Could you follow the flow of the demonstration for the use case? Did the demonstration convince you of the App's utility for the use case?	
	Total	**20**

3 Summary

This is a rapid development style project. Please plan early with your group to ensure no duplication of effort, yet ensuring all elements are covered for each deliverable. If you have questions about any aspects of the instructions, the rubrics, or expectations, please let me know right away. Good luck.