STAT 231: Shiny Project

Overview

In collaboration with your Power Up Group (PUG), you will create an interactive Shiny application that explores a topic of interest. The focus of this project is data wrangling, interactive data visualization, and effective communication. Thus, the question(s) you address need not be complex, and the answers might involve only basic summary statistics. But, the dataset(s) you work with should provide a data ingestion and/or wrangling challenge. That is, tidying up the data and getting it into the format needed to produce even simple statistics of interest may be complex. The interactivity of the Shiny application should enhance the message you're trying to convey.

This project is deliberately open-ended to allow you to explore your creativity and interests. There are only two main rules that must be followed:

(1) *Your project must be centered around data.* Preferably, you will work with large, complex and/or messy data. Alternatively, you may work with a dataset that isn't very messy, but is challenging to obtain (e.g., a challenging web scraping task).

(2) *You must create an interactive Shiny application.* The interactivity of the visualizations should enhance the message you're trying to convey. Each group member will create at least one component of the Shiny application.

Learning Objectives

The learning objectives of this project are to demonstrate your ability to:

1. Identify a set of questions that can be addressed with data available to you

2. Wrangle a large, messy dataset (including gathering, reshaping, and cleaning the data) into a format necessary to answer the question(s) at hand

3. Deploy interactive data visualizations using Shiny

4. Effectively communicate results via visualizations and oral presentation

5. Effectively collaborate with your peers and identify the value in teamwork

Timeline and Grade Contributions

All deliverables must be submitted by 9:00 PM EDT on the dates provided at the appropriate submission location. Adherence to these deadlines is required to help us keep on pace for the semester.

Checkpoint	Due Date	Credit	Submission Location
Initial proposal	9/28	5 pts	Issue in GitHub (group repo)
Revised proposal	10/1	10 pts	Issue in GitHub (group repo)
Wrangled dataset(s)	10/8	30 pts	GitHub (group repo)
Shiny App	10/15	50 pts	GitHub (group repo)
Presentation	10/15	30 pts	During class (recorded or live)
Reflection	10/16	10 pts	Gradescope

Components

Initial Proposal

The initial proposal is due by 9:00 PM on Monday, September 28th. We will use class time on Thursday, September 24th to begin the initial proposal.

The proposal should contain the following content:

1. Title: the title of your project

2. Purpose: Describe the general topic/phenomena you want to explore, as well as some questions that you hope to address.

3. Data: Identify one or more data sources that could be used in the project. What form is the data in (downloadable csv file? needs to be scraped from web?)? What do you imagine will be challenging about ingesting and/or wrangling the data?

4. Shiny app: Describe some visualizations, tables, and/or other components you envision including in your Shiny app. What will the interactive components be?

Feel free to include more than one direction if you would like feedback on different ideas you have. Part of the process here is learning how to refine questions, and how to evaluate whether you'll actually be able to answer the question you set out to answer.

Revised Proposal

The revised proposal is due by 9:00 PM on Thursday, October 1st. At this point, you should have landed on a finalized plan for what data you'll be using and details about the vision for your Shiny app. The revised proposal should be in the same format as the initial proposal (given above), and it should incorporate any feedback received on the initial proposal.

Wrangled Dataset(s)

The wrangled dataset(s) is/are due by 9:00 PM on Thursday, October 8th. The (reproducible!) code to create the dataset(s) should be saved in a .R or .Rmd file within your group repository and named "data_wrangling". I will be running the code, so make sure your repository is organized (so I can easily find the data_wrangling file!) and make sure the code is reproducible. The most common reproducibility error I've encountered with student work is that students have an object saved in their local environment that is not defined in the code. A good way to check the reproducibility of your code is to run it from a completely clean environment.

At this stage, your data should be in the format needed to create the visualizations and summaries you'll present in the Shiny app. You want to have as little wrangling as possible within your Shiny app program, so it's best to save your wrangled datasets as permanent files that you can load in at the top of the Shiny app program. You may have more than one wrangled dataset – for instance, if there are some outputs in your Shiny app that require the data to be in long format and other outputs in your Shiny app that require the data to be in wide format, then you'll want to create both a long format dataset and a wide format dataset in your "data_wrangling" file.

Be sure your code is readable, organized logically, and documented with informative comments.

Shiny Application

The final Shiny application will be due on Thursday, October 15th. Each team member needs to take the lead on at least one component of the Shiny app. One way to make this clear is to utilize the tab or navigation list format and have each member take the lead on coding one tab or panel of content.

You will be assessed on both a group and individual basis for this portion of the project. In a comment at the top of your shiny program, please indicate specifically which parts each team member took the lead on and/or contributed to, if it is not otherwise clear in the program.

Presentation

Each group will present their Shiny application to the class in a 6-7-minute oral presentation on Thursday, October 15th. Each group will have the option of presenting live to the class or preparing a pre-recorded video presentation to view during class. You will be assessed on both a group and individual basis for this portion of the project.

An effective oral presentation is an integral part of this project. Whether you choose to pursue a career in academia, industry or government, the ability to communicate clearly is of paramount importance. As a data scientist, the burden of proof is on you to convince your audience that what you are saying is true. If your audience cannot understand your results or their interpretations, then the technical merit of your project is irrelevant.

The intended audience for this presentation is our actual audience: a class of data science students. Your goal should be to convey to the audience a clear understanding of your topic, along with a basic understanding of your project, and how well the Shiny app addresses the question(s) you posed. You should ***not*** tell us everything that you did, or show a bunch of things that you tried that didn't work well. There are always exceptions, but your talk should probably follow this general structure: (brief) background on topic, define question(s) of interest, explain data source(s), display Shiny app (including demonstrating the app's functionality and the main messages it's conveying), concluding remarks.

Reflection

The reflection will be completed individually, and consists of a series of questions designed to help you reflect on what worked well within your group and what could be improved upon for next time. One of the questions will involve assessing yourself and your team members using the American Association of Colleges and Universities' Teamwork Value Rubric (posted on Moodle). Check out the rubric early on in the project to see what qualities contribute to being a good team member!

Assessment Criteria

Wrangled Dataset(s) (30 points)

	Score				
	Excellent	Good	Satisfactory	Poor	Unacceptable
Code Functionality (20 points)	18-20 points: The code is completely functional and responds correctly producing the correct outputs.	15-17 points: The program is mostly functional and responds correctly producing the correct outputs in most cases. There are minor problems with the program implementation.	10-14 points: The code is marginally functional with numerous errors and/or incomplete code sections.	5-9 points: The code is minimally functional with significant portions of the code missing or incomplete.	0-4 points: The code is not functional, producing no correct outputs, or was not attempted.
Code Readability (5 points)	5 points: The code is extremely well organized, properly formatted, and easy to follow.	4 points: The code is reasonably easy to read. There are minor formatting problems.	3 points: The code readable only with significant effort. There is little to no proper formatting.	1-2 points: The code is poorly organized and difficult to read. There is little to no consistency in formatting.	0 points: The code is readable only by the author or someone extremely knowledgeable with its layout and purpose.
Code Documentation (5 points)	5 points: The code is extremely well documented. Comments are completely consistent with the associated code. There are no spelling errors.	4 points: The code is reasonably well documented. There are minor formatting omissions that would have improved users understanding of code purpose. There may be limited spelling errors.	3 points: The code is marginally documented. There are significant portions of the code that are not documented or documented incorrectly. There are significant spelling errors that detract from the documentation.	1-2 points: The code is poorly documented. There are minimal comments and/or the comments are incorrect.	0 points: The code is not documented.

Shiny Application (50 points)

	Score				
	Excellent	Good	Satisfactory	Poor	Unacceptable
Code					
Functionality (10 points)	9-10 points: The code is completely functional and responds correctly producing the correct outputs.	7-8 points: The program is mostly functional and responds correctly producing the correct outputs in most cases. There are minor problems with the program implementation.	4-6 points: The code is marginally functional with numerous errors and/or incomplete code sections.	2-3 points: The code is minimally functional with significant portions of the code missing or incomplete.	0-1 point: The code is not functional, producing no correct outputs, or was not attempted.
Readability (5 points)	5 points: The code is extremely well organized, properly formatted, and easy to follow.	4 points: The code is reasonably easy to read. There are minor formatting problems.	3 points: The code readable only with significant effort. There is little to no proper formatting.	1-2 points: The code is poorly organized and difficult to read. There is little to no consistency in formatting.	0 points: The code is readable only by the author or someone extremely knowledgeable with its layout and purpose.
Documentation (5 points)	5 points: The code is extremely well documented. Comments are completely consistent with the associated code. There are no spelling errors.	4 points: The code is reasonably well documented. There are minor formatting omissions that would have improved users understanding of code purpose. There may be limited spelling errors.	3 points: The code is marginally documented. There are significant portions of the code that are not documented or documented incorrectly. There are significant spelling errors that detract from the documentation.	1-2 points: The code is poorly documented. There are minimal comments and/or the comments are incorrect.	0 points: The code is not documented.
Content					
Shiny application components (20 points)	18-20 points: Interactivity enhances the data visualization, and all graphical variable types are well-suited for the type and scale of data they represent, and there are clear labels and legends. The purpose of the component(s) is clear.	15-17 points: Interactivity enhances the data visualization and all graphical variable types used are well-suited for the type and scale of data they represent. Labeling and legends not completely clear. The purpose of the component(s) is clear.	10-14 points: Interactivity neither enhances nor detracts from the data visualization and most graphical variable types used are well-suited for the type and scale of data they represent. Labeling and legends not completely	5-9 points: Interactivity detracts from the data visualization and most graphical variable types are well-suited for the type and scale of data they represent. Unclear labeling and/or legends. The purpose of the component(s) is unclear.	0-4 points: No interactivity included.

			clear. The purpose of the component(s) is not completely clear.		
Shiny application cohesiveness* (10 points) team-level	9-10 points: The overall application provides a smooth user experience and an aesthetically-pleasing interface.	7-8 points: Some of the components of the application seem disjointed OR the interface is not welcoming (too basic or too overwhelming)	4-6 points: Many of the components of the application seem disjointed OR the interface is not welcoming (too basic or too overwhelming)	2-3 points: Many of the components of the application seem disjointed AND the interface is not welcoming (too basic or too overwhelming)	0-1 point: All of the components of the application seem disjointed AND the interface is not welcoming (too basic or too overwhelming)

* Cohesiveness will be scored for the team as a whole, while the other categories will be assessed individually.

Presentation (30 points)

	Score				
	Excellent	Good	Satisfactory	Poor	Unacceptable
Team					
Organization (10 points)	9-10 points: Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is clearly and consistently observable and is skillful, and makes the content of the presentation cohesive	8 points: Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is clearly and consistently observable within the presentation	7 points: Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is intermittently observable within the presentation	3-6 points: Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is barely detectable within the presentation	0-2 points: Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is not observable within the presentation
Content (10 points)	9-10 points: Central message is compelling (precisely stated, appropriately repeated, memorable, and strongly supported).	8 points: Central message is clear and consistent with supporting material.	7 points: Central message is basically understandable but is not memorable and/or not supported.	3-6 points: Central message can be deduced, but is not explicitly stated in the presentation.	0-2 points: Central message is unclear from the presentation.
Individual					
Language & Delivery* (10 points)	9-10 points: Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation compelling, and speaker appears polished and confident. Language in presentation is appropriate to audience.	8 points: Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation interesting, and speaker appears comfortable. Language in presentation is appropriate to audience.	7 points: Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation understandable, and speaker appears tentative. Language in presentation is appropriate to audience.	3-6 points: Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) detract from the understandability of the presentation, and speaker appears uncomfortable. Language in presentation is appropriate to audience.	0-2 points: Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) detract from the understandability of the presentation, and speaker appears uncomfortable. Language in presentation is not appropriate to audience.

* Language & Delivery will be scored individually, while the other categories will be assessed for the team as a whole.