Peer Group Project Part 1: Shiny App

In collaboration with your Peer Group, you will create an interactive Shiny application that explores a topic of interest to your group.

This project will focus on data wrangling, interactive data visualization, and effective communication. The question(s) you address need not be complex, and the answers might involve only basic summary statistics. However, 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 three 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.
3. **Each group member must create at least one component of the Shiny application.**

Learning objectives

The learning objectives of the Shiny App 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; and
6. demonstrate awareness of ethical considerations related to data acquisition, management, and communication.

Components

Initial proposal (Thursday, September 30)

The initial proposal should be submitted as a new issue to GitHub titled **Proposal**.

The proposal should contain the following content:

- **Title**: The title of your project
- **Purpose**: Describe the general topic/phenomena you want to explore, as well as some questions that you hope to address.
- **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?
- **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 (Thursday, October 7)

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 a new comment in the same GitHub issue, should be in the same format as the initial proposal (given above), and should incorporate any feedback received on the initial proposal.

Wrangled dataset (Thursday, October 14)

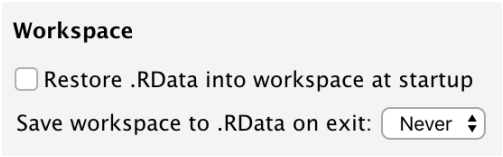
The (reproducible!) code to create the dataset(s) should be saved in an R or Rmd file within your group repository and named "data-wrangling." I will be running the code, so make sure your repository is organized (so I can easily find the "data-wrangling" file) and be sure your code is reproducible, readable, organized logically, and documented with informative comments.

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 little or no wrangling code 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.

A note on reproducibility
A common error with student work is having an object saved in your local environment that is not defined in the code. To help prevent this, you should never save the ".RData" file when prompted, and you should occasionally clear your environment and re-run your code.

You can change your workspace settings in **Tools** → **Global Options..** → **General**. Make sure the box next to "Restore .RData into workspace at startup" is *unchecked* and "Save workspace to .RData on exit:" is set to "*Never*".



Shiny application (Thursday, October 21)

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 (Thursday, October 21)

Each group will present their Shiny application to the class in a 6-7-minute oral presentation. 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. Communication is key as a data scientist. In their book *Build a Career in Data Science*, Emily Robinson and Jacqueline Nolis emphasize the importance of communication. Here are just three quotes:

> *"employers are first and foremost looking for evidence that you can code and communicate about data"* (page 59)
>
> *"Much of a data scientist's job is conveying information to nontechnical peers"* (page 141)
>
> *"A data scientist needs to be able to communicate. Over and over, people we interviewed for the book mentioned that their success came from communicating their work effectively."* (page 280)

You want to show you can communicate your results clearly (with the audience in mind) and concisely. If your audience cannot understand your results or 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, nor should you 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 (Friday, October 22)

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 (available in the projects folder and on Moodles). Check out the rubric early on in the project to see what qualities contribute to being a good team member!

Timeline

All project components except the final reflection will be due on Thursdays by 10pm ET so that I may review your material on Fridays.

Activity	**Timeline**
Initial proposal	Thursday, September 30 by 10pm ET
Revised proposal	Thursday, October 7 by 10pm ET
Wrangled dataset	Thursday, October 14 by 10pm ET
Shiny app and presentation	Thursday, October 21 by start of class
Peer group reflection	Friday, October 22 by 10pm ET