

ESM 244 Assignment 1

All parts due by Tuesday (1/29) at 2pm. Please note: You will submit your individual efforts for Tasks 1 and 2 in a single .Rmd (and knitted html) to GauchoSpace by 2pm on 1/29. Final project (Shiny) groups will submit a single hard copy for Task 3 in class on Tuesday 1/29.

Task 1 (individual):

#TidyTuesday - space launches!

Tidy Tuesday is a new awesome project launched by RStudio, the R for Data Science (R4DS) community (follow @RStudio, @r4ds and #TidyTuesday on twitter) to get people wrangling and visualizing data each week. See more information about #TidyTuesdays on github here (check the ReadMe): <https://github.com/rfordatascience/tidytuesday>

You're asked to participate (a little late) in #TidyTuesday for the week of 1/15, which involves data for space launches. Information on the data:

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-01-15>

- a. Download the files **launches.csv** and **agencies.csv** from GauchoSpace
- b. Create a new R project, and drop the .csv files into the project folder
- c. Create a new well-organized Rmarkdown document (knit to html) in which you:
 - Read in and wrangle the data (as needed)
 - Visualize the data in a way that you think is interesting and informative. **Finalize** your graph (or table/figure - you choose). Everyone's wrangling and visualization can (and should) be different! You get to choose everything about how you're presenting the data (e.g., Will I visualize data for all launches or a subset? All agencies or just a few? Categorized by year, success/failure, agency, or something else? You pick!)
 - Update your markdown doc so that the only things appearing in your knitted document are your code and perfectly formatted final graph (no messages, warnings, etc.)

Task 2 (individual):

PCA for raw fruit and vegetable nutrients

For Task 2, use PCA to explore multivariate nutrient data (from USDA) for **raw** fruits and vegetables (dataset **usda_nutrients.csv** on GauchoSpace). In the same markdown document you created for Task 1, add well-organized and annotated code (which will only show the code and final output when knitted to html) to complete the following:

- Wrangle the data to only include food groups 'Vegetables and Vegetable Products' and 'Fruits and Fruit Juices'

- Only retain observations for which the short description (ShortDescrip) contains the string 'RAW'. **Hint:** see function `str_detect()` in the `stringr` package (it's in the tidyverse)
- Perform principle components analysis based on variables from 'Protein_g' through 'Zinc_mg'. If you haven't added or removed columns from the original data frame, those are columns 9:30.
- Create a nice biplot showing both individual points (unlabeled), variables (labeled, repelled), with the two food groups (Fruits, Vegetables) differentiated by color. You'll probably want to use `factoextra::fviz_pca_biplot` with customization. You do not need to include a figure caption. For more examples, [HERE](#) is a good starting point.
- In text below your graph, describe several general trends that you notice from the biplot re: relationships between variables (just pick 2 - 3 pairs, and describe what the biplot tells you about the correlation) and between the food groups.

Submitting Tasks 1 and 2:

- **Knit your final .Rmd for Tasks 1 and 2 to html. It should only include code, final graphics, and requested text (no messages, warnings, etc.). Upload your final .Rmd and html to GauchoSpace at the link provided.**

Task 3 (submit one hard copy per group in class on Tuesday 1/29):

Final project outline: Shiny app

The goal of Task 3 is to help you get started with your final project - a Shiny app in R. Please see the final requirements and rubric posted on GauchoSpace. **Note:** your app design will change over the course of creating it. That's OK. This is a starting point to get you rolling.

I strongly recommend checking out RStudio's Shiny gallery: <https://shiny.rstudio.com/gallery/> for inspiration and ideas.

For Task 3, prepare an organized document (bulletpoints encouraged) that includes:

- App summary:** A 2 - 3 sentence "big picture" summary of what you want your app to do
- Data:** A ½ page detailed description of the data you'll use in your app, including citations as appropriate. Some things to consider: What variables exist? How many observations? Are the data currently in tidy format?
- Widgets:** Briefly describe (½ page) the 3 (or more) widgets that you will include in your app. What type of widgets will you include (e.g. slider, radio buttons, etc. - see the Widgets gallery at the link above for possibilities)? What will be the inputs for each widget?
- Outputs:** Briefly describe (½ page) the 3 (or more) final outputs (tables, graphs, maps, figures, etc.) that will be produced based on widget selections.

- e. **Sketch your user interface.** Produce a simple illustration (digital recommended) of your app's user interface. Each tab should have its own "slide" in your illustration. An example is included below (yours will be different - this is just to give you an idea of the type of thing I'd expect).

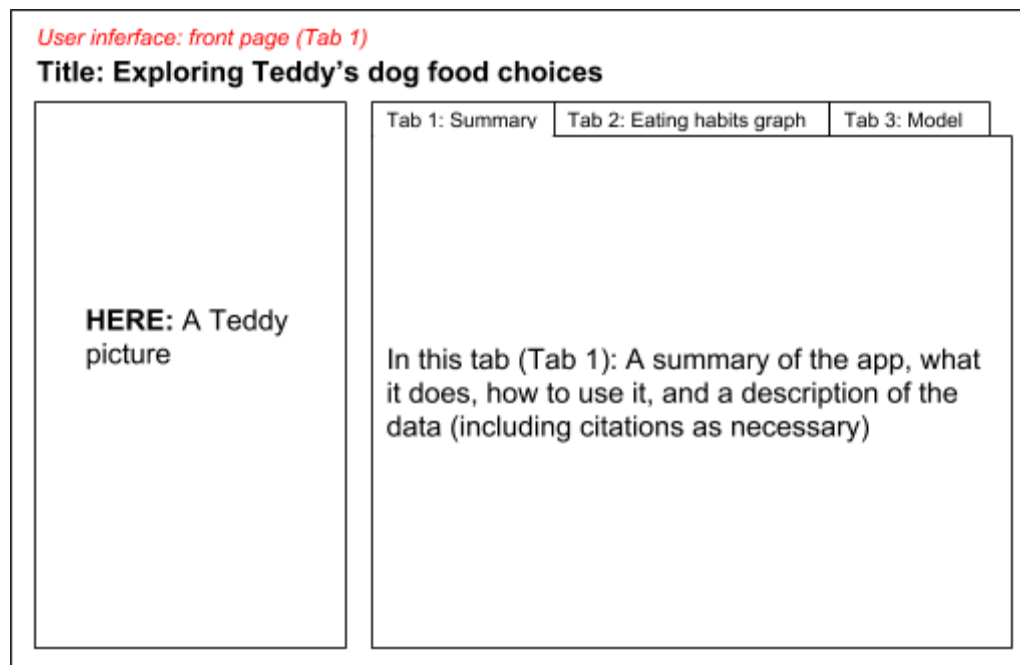
Note: the more detailed and thoughtful you are in preparing this app outline, the easier your life will be when actually making it.

Task 4 (each Shiny app group):

Create your Shiny app repo on GitHub, add me as a collaborator, and fill out the Shiny app google sheet (link below)

- a. Create a repo on github for your final app, and add all group members and me (allisonhorst) as collaborators. Include a ReadMe with a tentative title for your app, and all group member names. **Note:** If you are using confidential data (e.g. from a GP client), you should make your repo private.
- b. Fill out the google spreadsheet [HERE](#) with a tentative app title, group member names, and a brief description.

Example app sketch (for Task 3):



User interface: Exploring eating habits (Tab 2)

Title: Exploring Teddy's dog food choices

Sidebar Panel

Widget 1: Select time

Type: Date slider

Input: Select timespan to include in analysis

Widget 2: Select meal

Type: Radio buttons

Input: Meal (breakfast, lunch, dinner)

Tab 1: Summary

Tab 2: Eating habits graph

Tab 3: Model

Output visualization: Daily choices

A stacked column graph showing which foods Teddy chose for that meal with data existing for the selected time span, categorized by day of the week

Graph of daily food preferences here

User interface: Food choice model (Tab 3)

Title: Exploring Teddy's dog food choices

Sidebar Panel

Text description of model (inputs, outputs, assumptions, type of model, etc.)

Tab 1: Summary

Tab 2: Eating habits graph

Tab 3: Model

Output: Graph of model predictions + table

A graph of predicted choices by multinomial logistic regression, and a finalized table of model outputs

Graph of predicted probabilities for food choices

Table of model outputs