

Project 1 Instructions

Vignettes

Vignettes are explanations of some concept, package, etc. with text, code, and output interweaved. We already know how to make them with R Markdown!

Project Objectives

This project involves creating **two** vignettes (each an HTML file with a table of contents) about reading in data of a certain type (JSON and XML) and exploring it (summaries and graphs - more details on this later). You should create a github repo to house all of the work you do for this project.

Vignette Content Details

You are going to create two vignettes, one for JSON data and one for XML data. The components that must be present in each vignette include:

- Your code chunks should be shown in the final document unless they are set up chunks or other behind the scenes things that aren't important.
- Describe your type of data (JSON or XML). What is it, where does it get used, and why is it a good way to store data? This should be detailed enough that someone that hasn't seen that type of idea would have a good idea what they are dealing with. You should link to references where applicable.
- Discussion of possible packages/functions that are available for reading your type of data into R. (There are three major packages for JSON data and two major packages for XML data.) Choose one (for each data type) and explain why you've chosen it.
- Find a dataset of the appropriate type (JSON or XML) and describe where you found the data, how the data was collected, what the variables are, etc. You should have different data sets for each vignette.
 - The data you read in should have at least two categorical variables and two quantitative variables.
- Read in the data set describing the options your package's functions allow.
- Perform basic exploratory data analysis that reveals a meaningful idea that you would reasonable want to investigate further. Not all things reported need to show something meaningful (i.e. graphs that show no relationship are fine) but you should end up with a solid lead that you would pursue further.
 - At some point you should create a useful function(s) to do something meaningful with the data or customize the way you read the data in.
 - You should create a new variable.
 - You should create some contingency tables and numeric summaries by some of your categorical variables
 - You should create some plots (at least a side-by-side bar plot, side-by-side box plots, and scatter plots with coloring)

When you are finished, you should post a link to your repository as your submission.

Rubric for Grading (each vignette is out of 100)

Item	Points	Notes
General use of Good Programming Practices	15	All code should be indented, follow naming conventions, etc.
Use of headings, table of contents, chunk options, etc.	5	Worth either 0, 3, or 5
Data type (JSON or XML) description	15	Worth either 0, 5, 10, or 15 points
Discussion of relevant R packages	10	Worth either 0, 4, 7, or 10
Description of Data Set	15	Worth either 0, 5, 10, or 15 points
Custom function (best practices should be used)	10	Worth either 0, 4, 7, or 10
Creation of relevant new variable(s)	5	Worth either 0, 3, or 5
Calculation of relevant numeric & graphical summaries	25	Worth either 0, 5, ..., or 25 points

Notes on grading:

- For items worth say 0, 5, 10, or 15 points, we will generally move you down one level for each syntax, logical, or other error present in the code. The same holds true for missing a required item or lacking in a description.