

Introductory Data Science Course Project: Retrieving Data from an API and Conducting an EDA




Justin Post
✉ jbpost2@ncsu.edu

Department of Statistics, NC State University

Purpose

Creating meaningful projects in a data science course can be a time consuming task. The purpose of this beyond session is to provide example project instructions from a graduate level introductory data science course using the tidyverse that could be used to create a similar assessment for your class along with example student submissions. The purpose of the project is to give students practice writing coherent R functions to query an API (think online database), return the data in a usable format, and summarize the data in a meaningful way. The project discussed here requires the topics of: R Markdown, Function Writing, Data Manipulation, and Summarizing Data. Students should be familiar with the common steps of conducting an exploratory data analysis in R.

Project Requirements

- 
- 
-  (optional) This is the `posterdown_betterland` template for the {posterdown} package! I was inspired by the twitter thread of [Mike Morrison](#) and wanted to apply the `#betterposter` concept to the reproducible (yet simple to use) functionality of the {posterdown} package ([Thorne 2019](#)). If you're not an R user don't sweat as you do **NOT** need to use it at all! Feel free to use only the Markdown functionality of this package :)

Objectives

1. Pick a template layout.
2. Write/ create your poster content distraction free.

Make better posters with RMarkdown + posterdown.



Methods

I will show here how to include poster elements that may be useful, such as an equation using mathjax:

$$E = mc^2$$

To reference a citation you can add your `.bib` file to the working directory and name it in the YAML metadata or generate an automated one as done here, then you only need to reference the label value in the `.bib` file. For example this package is built on top of the wonderful {pagedown} package and I will cite it at the end of this sentence using this in the rmd `{@R-pagedown}` ([Xie et al. 2022](#)).

To get a better understanding of how to include features like these please refer to the {posterdown} [wiki](#).

Now on to the results!

Results

Here you may have some figures to show off, bellow I have made a scatterplot with the infamous Iris dataset and I can even reference to the figure automatically like this, `Figure \@ref(fig:irisfigure)`, [Figure 1](#).

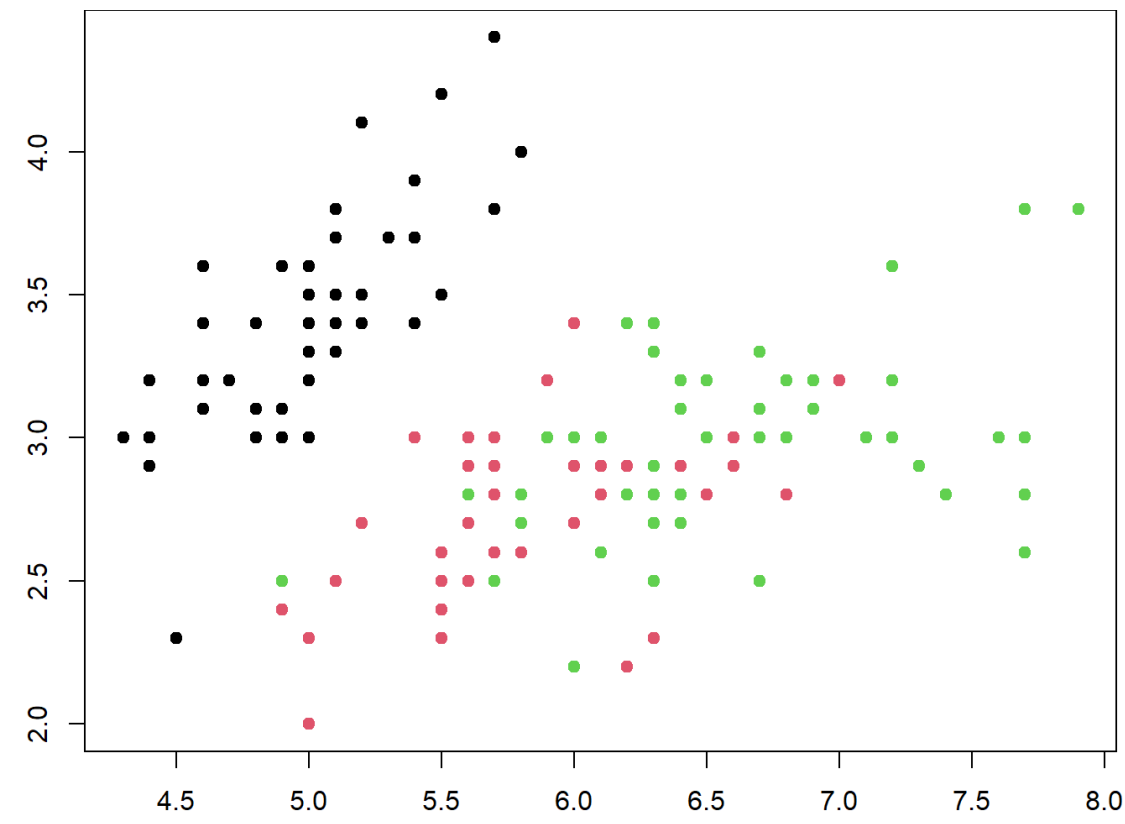


Figure 1: Here is a caption for the figure. This can be added by using the "fig.cap" option in the r code chunk options, see [this link](#) from the legend himself, [Yihui Xie](#).

Maybe you want to show off some of that fancy code you spent so much time on to make that figure, well you can do that too! Just use the `echo=TRUE` option in the r code chunk options, [Figure 2!](#)

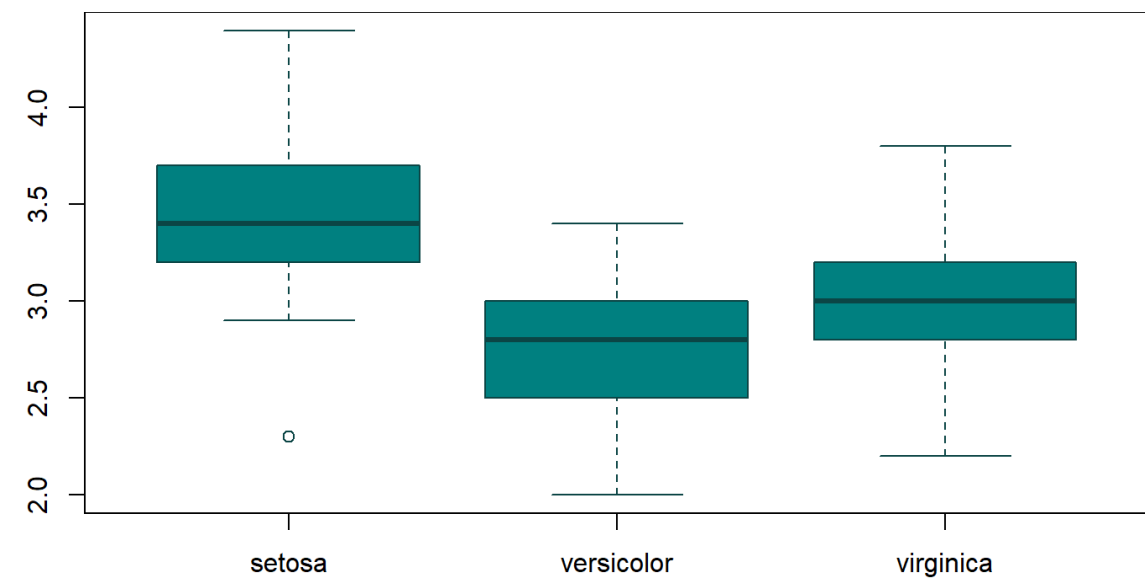


Figure 2: Boxplots, so hot right now!

How about a neat table of data? See, [Table 1](#):