Exploring the Weight Loss Datasets

Ted Laderas

May 12, 2016

In this exercise you will be conducting an exploratory data analysis (EDA) on two datasets. In both datasets, an experimental weight loss drug was given in a double-blinded study to a set of patients. There are two groups of patients. One group received treatment and one group received a placebo. The patients were weighed before treatment, and then, at a second clinic visit, weighed again after treatment. Dataset A was conducted at a university clinic in Southwest Portland, and Dataset B was conducted at a nursing home in Detroit, Michigan.

Your goal is two-fold. You will use Shiny to visualize the data in multiple ways and identify possible issues in the data. Then you will collect the plots that support your case.

Getting Started

1) Once you have cloned the git repository (https://github.com/laderast/shinyEDA), open the `shinyEDA.Rproj` file using Rstudio.
2) Read the documentation for each dataset (A and B). Does the documentation give you any clues about how to approach the data?
3) Open the `global.R` file. The majority of your work will be done where it reads `##Place Data Processing Steps Here`. If necessary, you will be using dplyr commands (such as `mutate()` and `filter()`) to manipulate each dataset during your EDA.
4) To run the Shiny App, use the `Run App` button in the right corner of the source window. Start exploring each tab and see the different ways you can visualize the data.
5) EDA is an iterative process. Try filtering and transforming the data to see what effect it has on the visualization
6) Be ready to talk people through your process and how you discovered issues with data. If necessary, you can save the plots by right-clicking and saving them.

Dataset A

For dataset A, do you believe that weight loss occurred more in the treatment group versus the non-treatment group? What does the data look like? Are there any issues with the data? Do you believe the differences are real? Show your pattern of thinking. Use the available information in the data dictionary to guide your exploration in the data. What aspects of the data do each of the visualizations highlight?

If you need to remove data, script your filtering process and place it in `global.R`. Make sure your filtered dataset is called `dataset`, so you can continue to use the explorer. If you need to calculate new variables to visualize, make sure that you save them as variables in `dataset`.

Dataset B

If you have time, tackle the second dataset, `datasetB`, by loading it into the Shiny App. Given your conclusion for step 1, does dataset B support that conclusion or not?

Combining Datasets

Can you combine the information from dataset B and dataset A in a meaningful way? Do you have to transform the data to compare them? How would you combine them into a single dataset? Are there any large differences between the two datasets? What differences are there?

If you get stuck

We are here to help. Be sure to use your post-its to signal if you need help. If you really get stuck, there is an answer key talking you through the process in the `answers/` folder in the app. But you will learn way more by exploring the data and staying curious.