

Tabular data Assignment

Tabular Data Assignment

In this Assignment, you will create different data visualizations of a public data set of your choice from the data.boston.gov website. After our introduction to R and ggplot2 using 311 data in class, you will explore the ggplot grammar of graphics further in this assignment.

1. Find a suitable data set from data.boston.gov and download it in the .csv format. You should pay attention to the structure of your data set, to make it useful for the following explorations. The data set should contain multiple variables including **time/date variables**, **numerical variables** (which are not just administrative codes), and (probably the most common) **categorical variables**.
2. Create a new Rstudio project, which will be the place for all R scripts you write and use during this course. Create a data subfolder in your studio project directory to help you organize your files. Copy the .csv file into this folder.
3. Import the file using the tidyverse `read_csv()` function (create a new R script). This function will automatically interpret the variable data structure and create an appropriate tibble data frame. Use ggplot2 to visualize structure of the dataset and its internal relationships. What do the different variables stand for? How does the city use these variables? How do you think the data set was generated, by whom?
4. Pay attention to picking appropriate chart types and color scales for your analysis. Use appropriate titles and labels. As you analyze your data set, you will also explore the features of ggplot2 - its different layout options (geom_), scales (e.g. linear or logarithmic), coordinate systems (e.g. cartesian or polar), and color mappings (also under scales, quantitative, qualitative etc.). You should also consider layout options such as facets, theme settings or placement of the legend.
5. Create a R-markup (.rmd) file that introduces the data set, explains your analysis, and presents your findings using the charts you created.

Submission

Turn in your paper as a knitted .html file (the result of a compiled .rmd) over blackboard, by Wednesday, February 27. Make sure that the html file includes all charts and visuals.

Grading

Your assignment will be graded on the basis of:

- Your choice of an appropriate data set that allows a variety of explorations, your understanding of its purpose and structure (demonstrated through your writeup in the .rmd file)
- Your acquired skills of using ggplot to produce well-formed plots with appropriately labeled axes and titles.
- Your understanding of visual variables