

Homework 5

Due Wednesday Nov 4, 2020

2020-10-12

For each assignment, turn in by the due date/time. Late assignments must be arranged prior to submission. In every case, assignments are to be typed neatly using proper English in Markdown.

This week, we spoke about Exploratory Data Analysis and plotting. To begin the homework, we will as usual, start by loading, munging and creating tidy data sets. In this homework, our goal is to create informative (and perhaps pretty) plots showing features or perhaps deficiencies in the data.

Problem 1

Work through the Swirl “Exploratory_Data_Analysis” lesson parts 1 - 10. If you need some review of ggplot, see the tutorial on Rstudio.cloud.

Problem 2

Create a new R Markdown file within your local GitHub repo folder (file->new->R Markdown->save as).

The filename should be: HW5_lastname, i.e. for me it would be HW5_Settlage

You will use this new R Markdown file to solve the following problems.

Problem 3

Using tidy concepts, get and clean the following data on education from the World Bank.

http://databank.worldbank.org/data/download/Edstats_csv.zip

How many data points were there in the complete dataset? In your cleaned dataset?

Choosing 2 countries, create a summary table of indicators for comparison.

Problem 4

Using *base* plotting functions, create a single figure that is composed of the first two rows of plots from SAS’s simple linear regression diagnostics as shown here: <https://support.sas.com/rnd/app/ODSGraphics/examples/reg.html>. Demonstrate the plot using suitable data from problem 3.

Problem 5

Recreate the plot in problem 3 using ggplot2 functions. Note: there are many extension libraries for ggplot, you will probably find an extension to the ggplot2 functionality will do exactly what you want.

Problem 6

Finish this homework by pushing your changes to your repo.

Only submit the .Rmd and .pdf solution files. Names should be formatted HW5_lastname_firstname.Rmd and HW5_lastname_firstname.pdf