Day 4 Pset

Cecilia Sui and all other TAs

Day 4 Outline

1. Descriptive Statistics
2. Introduction to `ggplot`
3. Simple linear regression

ggplot2 Basics

1. Find all the variables that negatively correlates with TFR from the **worldTFR** dataset. Can you offer a reasoning of the relationships?

2. Pick two variables from Q1 and use the package ggplot2 to plot them and TFR separately, where the two variables you choose should be on your x-axis, and TFR should be on the y-axis. Add appropriate labels. Compare the two plots and explain the differences.

3. Create a new dataframe named df1, such that the dataframe has two columns: Year and avg_TFR. The column Year should list the years with no duplicates, and the column TFR should store the average TFR of all countries during that specific year. Use the package ggplot2 to visualize the relationship between Year and avg_TFR. Describe what you see.

Linear regression

1. Find a dataset you think interesting. Load and do basic cleanings (dealing with NA values, check the data types, etc.).

2. Run a single linear regression, and report the results in Tables and Figures.