Day 5 Pset

Cecilia Sui

1/14/2022

Day 5 Outline:

1. Packages and Help Pages
2. Introduction to dplyr
3. Introduction to ggplot2
4. Tips on Using R Markdown

dplyr Basics

We will use the resources dataset to do some exercises with dplyr

1. Select the first 10 rows and last 3 rows. Output the rows together as the same dataframe.
2. Select all rows where Year is 1951 and regime is non-negative.
3. Order the rows by Year in descending order. Break ties using the column oil in ascending order.
4. Select the 100th to 110th rows, and find the row with the highest value for metal.
5. Randomly select 10 rows with replacement, and standardize the columns oil and metal.
6. Select columns: cty_name, year, regime, oil, illit, and life and store them into a new dataframe.
7. Add a new column named new_oil using where the values are the product of standardized oil and regime type.
8. Calculate the **average** oil level for **each year** from 1966 to 1997.
9. Calculate the **median** metal value for **each** regime type from -10 to +10.

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.