STAT 412/612 HOMEWORK TIDYING DATA

Instructions:

1) Execute the following installations (separately and one at a time):

**install.packages("devtools")** (Remember do not send installation code to Rmarkdown)

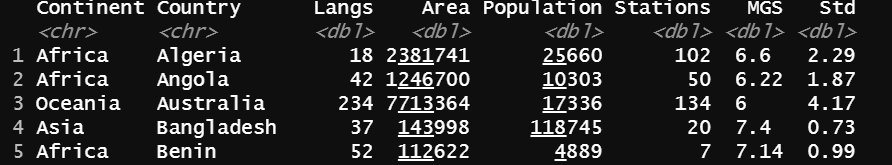
**devtools::install\_github("jvcasillas/untidydata")**

Enter the r code **untidydata** and then select **language diversity.**

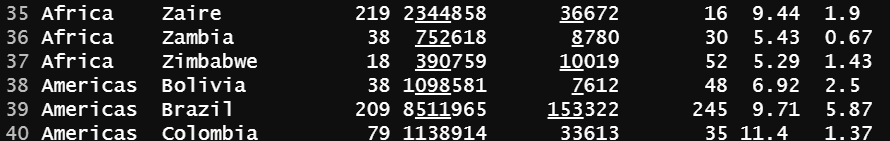
You should produce the following r data call of  **untidydata::language\_diversity**

2) The data set is too long (444 rows) and should be wider in order to become more useful and understandable. Use and show R code to produce a table that is wider data table version of the original data table.

A partial printing of the wider data table is given below:



3) Now use and show R code that will arrange the Continents in alphabetical order. **Print the first 40 rows**. A partial table showing rows 35 – 40 are given below.



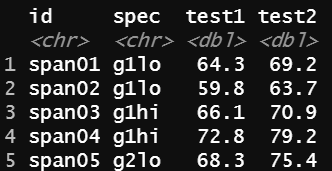
4) Now use and show R code that will only produce Asian countries in the data table. There are 16 Asian countries.

5) Enter the r code **untidydata** and then select **pre\_post**

You should produce the following r data call of  **untidydata::pre\_post**

Produce the **pre\_post** table that shows all 16 rows.

6) For the **pre\_post** data table, Use and show R code that will eliminate the under scores for the variable **spec.** A partial table is given below:

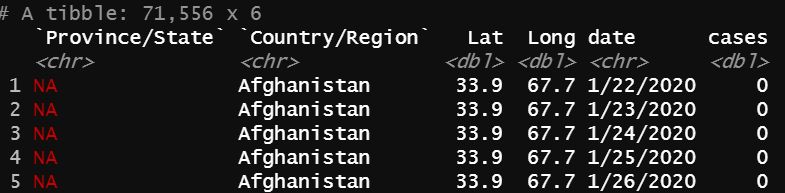


7) Load the given link below and then create an excel file for the covid- 19 data that is given.

**https://raw.githubusercontent.com/CSSEGISandData/COVID19/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series/time\_series\_covid19\_confirmed\_global**.

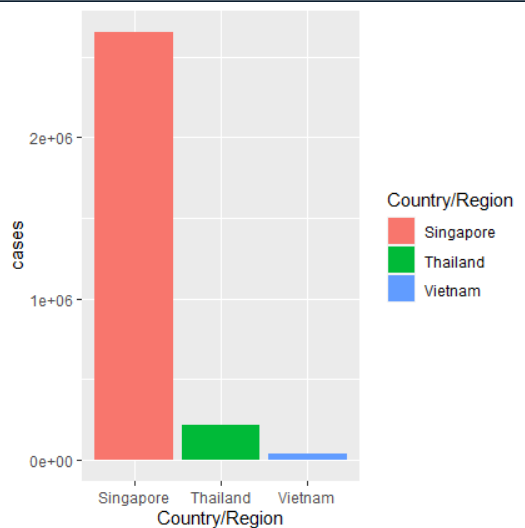
Use and show R code that imports the excel spread sheet into R. Assign your imported data the variable **covid**.

8) The **covid** data table is too wide. Apply and show R tyding code that will create a longer data table and then print 25 rows of the new data table. A partial table is given below.



9) Now use and show R code that changes your table from number 8 and will produce a table that only shows data for the countries Vietnam, Thailand, and Singapore. When printed, the first ten rows automatically shows data for Singapore.

10) Now show and use R code to produce the bar graph given below that shows approximate counts for the countries Vietnam, Thailand and Singapore



11) Now show and use R code that will give accurate counts for Covid-19 cases for all there countries. (The number of cases for the country of Singapore is 2,653,210)