**Data Wrangling on Coal Consumption**

I have retrieved data for coal consumption from <http://594442.youcanlearnit.net/coal.csv>

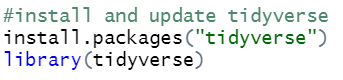
The main purpose is to prepare the dataset for analysis.

**Business Question**

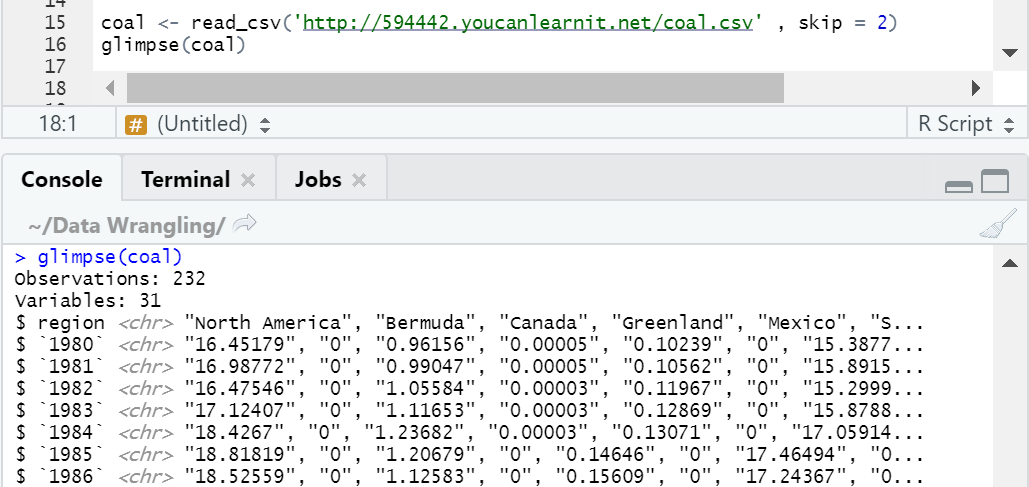
Identify trends in coal consumption from 1980 to 2010.

**Analysis**

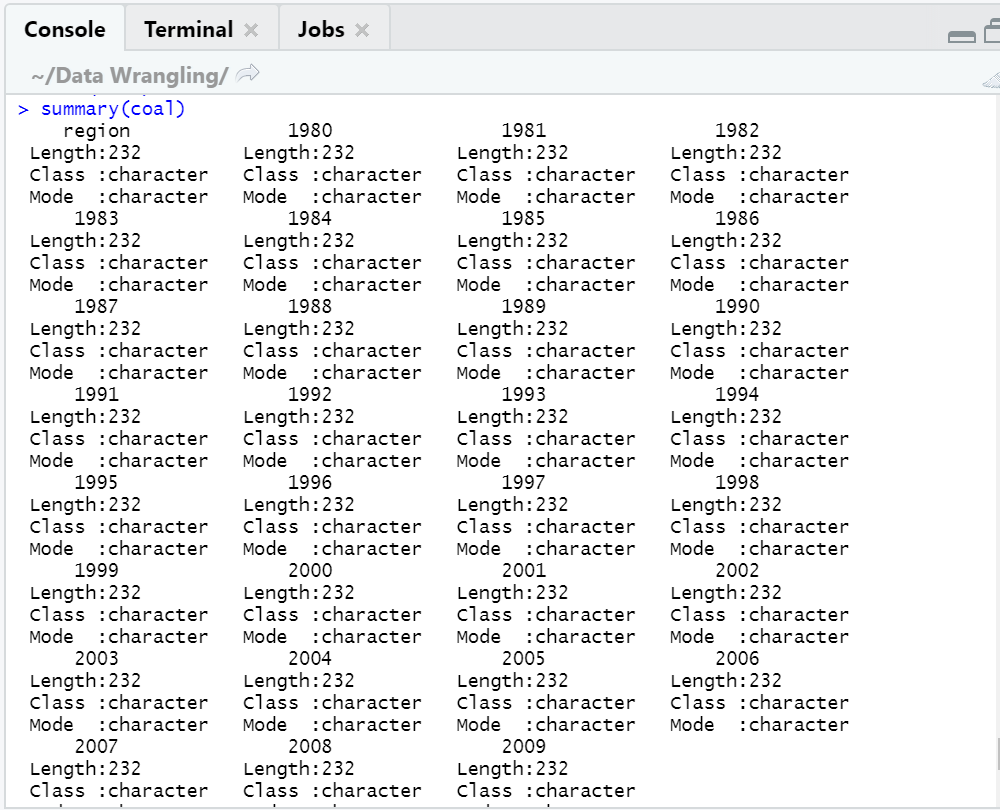
I am going to use tidyverse library to perform data wrangling operations



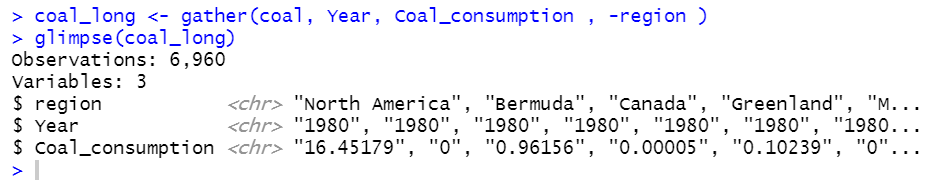
Now let’s import data and have a look at the variables



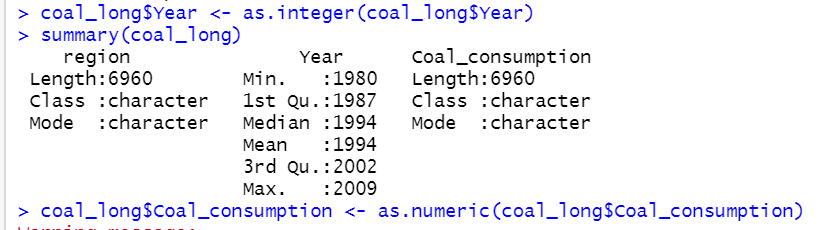
Summary of dataset:



The dataset is wide. A good rule of thumb is to create wide dataset to long. I am going to use [gather] function.

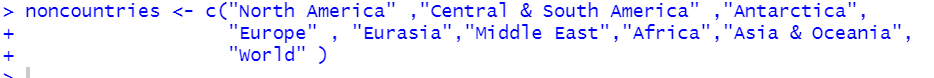


All the variables are of Character datatype, I will convert year as integer datatype and coal\_consumption = numeric.

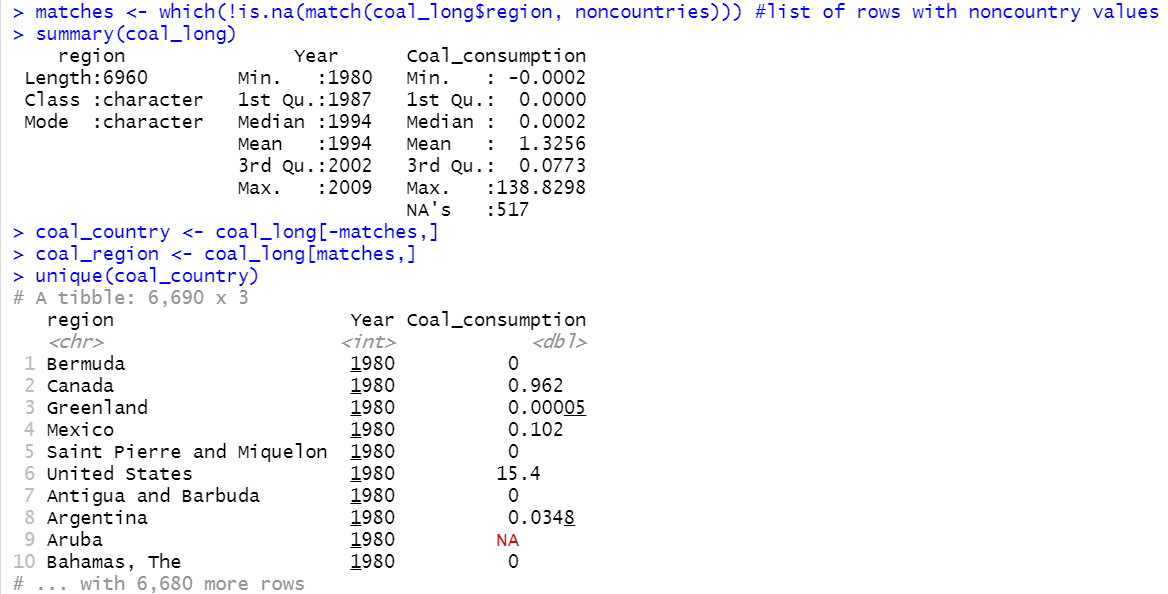


Now, I can see that the regions variables has to be classified as it consists of many values which are countries, world and continents. For our analysis, we will simplify it to continent levels and call them regions. I can also see world as a row. We will keep this record for now and will not delete it.

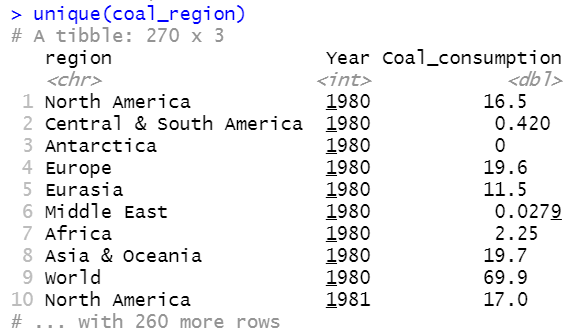




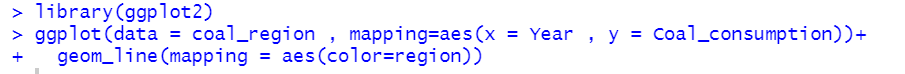
**Data preparation**



Now we have prepared data for analysis.



**Let’s use ggplot2 library to visualize**



**Insights**

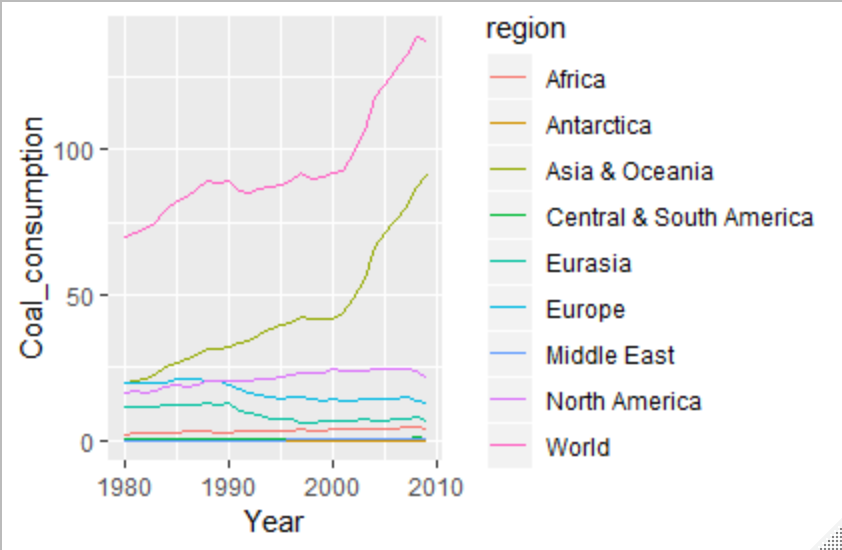


Figure 1: Plot for coal consumption trends across regions

From the line plot, we can see the trends in coal consumption in different regions of the world. From figure 1, we can see that the overall coal consumption has increased over the years.

Asia & Oceania has highest coal consumption and thus the trend line for coal consumption has peaked despite other regions having low/ consumption.