HW6 : Working with Data

Problems

Two datasets are from the World Bank DataBank. The first one (**WB_Growth**) contains information of economic growth by country, starting from 1960. The second one (**WB_Regions**) contains a categorization of countries into regions.

1. Read two datasets into R. (For the first dataset skip first 4 rows and name it *dat1*, name the second dataset *dat2* and drop the column *SpecialNotes* from it). If a region is missing change the value of it to *Unknown*.
2. The dataset *dat1* contains one extra row. Remove that row using the column *Country.Code*.
3. Write a function $AllNA()$ that verifies if a vector is entirely consists of NAs. Apply that function to the two datasets to remove columns containing only NAs.
4. Merge obtained two datasets into the dataset *dat*.
5. For each country calculate the mean economic growth since 1960 (use the fact that columns containing economic growth numbers start with X). Save that information in the column *Growth* in the *dat* dataset.
6. For each region calculate the mean economic growth in that region. Add a column that will show the date and time of analysis performed (only first row of this column contains information, other rows are blank). Write this information into the file *Statistics.csv*. (Make sure that the output file includes row names).

Extra Problems

7. Remove all rows from the dataset *dat* having *NAN* value in the column *Growth*.
8. Sort the dataset *dat* in increasing order using the column *Growth*. Reinitialize the row names of the sorted dataset.
9. Find the row name of *Armenia* in the sorted *dat* dataset.
10. Write a function that calculates how many letter *A*s (the case does not matter) there are in a string. Apply this function to country names of the dataset *dat* and keep this information in the column *numA*.