PROJECT REPORT

About the Project

I took the data from the NOAA website. The data contains the rainfall data of the month June 2015, from different weather stations across the world, spanning over 144 countries. It also contains information about pressure, temperature, elevation, etc. This data was available in from of a PDF and had missing values in various columns

Work flow of the Project

- First I converted the PDF file into an MS-Excel file from pdf to excel converter available online. This caused the data to get highly messed up in the Excel file.
- The data was loaded into R and various functions were implemented to clean the data.
- I choose to use a NoSQL database for storing the data. My choice of NoSQL database was MongoDB (The main intension here was to test and use my knowledge of NoSQL Databases)
- After storing the data in MongoDB, I ran a few database queries to find some meaningful insights and to check if the data is loaded properly in the database.
- At last I did some exploratory analysis in R, by plotting some useful charts and tables.

Data analysis queries in R

Q1 Countries which had the maximum average (of stations) rainfall.

Q2 Bar plot of top 10 or 20 countries having maximum rainfall

Q3 Station recorded most rain.

Q4 Amount of rainfall at the highest elevation.

Q5 Average rainfall in the world for the month June 2015

Q6 Station which recorded most days on which the rainfall was 1mm or more.

Q7 Plot between vapour pressure and temperature

Q8 Plot between elevation and pressure

How to run the code

Please run the scripts in this order.

Data Cleaning script ---------- > MongoDB script ---------- > Data Analysis script

Insights from the analysis

- The graph between Vapour Pressure and Temperature shows that both follows a similar relationship. As the temperature goes up vapour pressure in the atmosphere also increase.
- The graph between pressure and elevation depicts a linearly inverse relationship between them. At higher elevation pressure is low and at low elevations the atmospheric pressure is high.

Footnote

The Units of different columns are given below

COLUMN	MEANING	VALUES IN
Elevation	Elevation of the station	Meters
Station_Pressure	Atmospheric pressure at the station	Hectopascals
Sea_Level_Pressure	Pressure at sea level near the station	Hectopascals
Mean_temp	Mean temperature at the station	Degree Celsius
Mean_Vapour_P	Mean vapour pressure at the station	Kilo Pascal
Days_1mm	Number of days on which there was 1mm of rainfall or more in one month	Millimetres
Total_Rain	Total rainfall at the station in one month	Millimetres
Total_Sunshine	Total number of hours of sunshine at the station in one month	Hours

(There are a total of 4 files including this Project Report)