Ethan Behar

CSC-423

Assignment 4

Data Set: This data set I obtained from <http://www.pdc.org/iweb/volcano_deadliest.jsp>. I was interested in dealing with data that came from volcanoes. Google was my starting point on the search for this data.

Code and Comments:

> install.packages("ggplot2") – installs package needed for the project

> library(ggplot2) – loads the library we will use

> data<-read.csv("volcanoDataSet.csv") –reads the data from the file into ‘data’

> data – checks that the data was loaded corrected

> attach(data) – allows user to use data like individual data members

> qplot(Date,Deaths,col=as.factor(TypeOfDeath)) – plots the data on a graph with a legend

Description of Data: One thing I would to point out is that the data did come with locations but for simplicity I didn’t include the locations in the graph. The locations can be found in the data file. The data from this website tells some interesting points about volcanoes. In this set majority of people died from ashes in the air and filling their lungs. These high deaths due to ashes are located near water and have an average elevation over 2,000 ft. This could explain why so many people died from ashes. With that altitude of the volcanoes and wind coming from the water, it seems ideal for ash to flow throughout the air so easily. Starvation was also another devastating factor that killed about 100,000 people in two locations. Both of these volcanoes are located on islands. Due to the location it seemed the lava flow right down the mountain sides and into the plains in the area. These lands around the volcanoes were most likely farm land and when they erupted they destroyed the civilizations crops resulting in major starvation. Another interesting fact is that the volcanoes that started tsunamis are located in Asia near the pacific islands. The eruptions must have been so strong that it caused the ocean to react violently with tsunamis. The last thing I would like to point out is that mudflows from volcanoes haven’t killed a person since 1919 in this data. Because this data is limited I do not want to extrapolate on the mudflows for there might be other data saying otherwise.