Fernanda Montoya
JHU Biostatistics PhD Placement Assessment – Data Science
8/6/2024

Github Repo link: https://github.com/fernanda-montoya/DSplacement

- This repository contains the code and pdf referenced in this doc, along with the same answers to both of the questions below in the commented sections of the code.

Question 1

This code webscrapes a wikipedia article on natural disasters and their death toll and then plots data relating to the death toll for a natural disaster throughout the 20th and 21st centuries.

The resulting plot is seen below where we note the death toll across time and allow the color of the points to correspond to the natural disaster type. The key for the disaster type is located to the right of the plot.
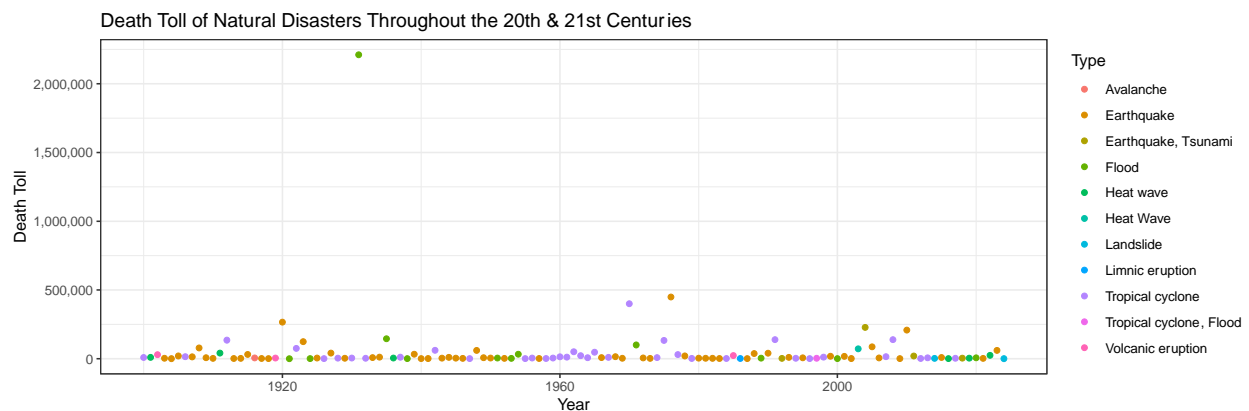
Question 2

How does the performance of the algorithm depend on e?

The user input for e adjust how large or small of a step is taken when performing the gradient descent and thus, we see that when running the function and changing only the e parameter, convergence to b happens more slowly for larger step sizes.

When does the algorithm fail and why?

This can algorithm fail for a couple of different reasons. One is if the number of iterations that the gradient descent completes is too low, the algorithm will not converge to the right

value of b as it cannot compute enough values for b to get close to the convergence criteria. Additionally, if the number of iterations is too large with a step size that is also too large (e.g. e = 2 and iter = 500), the convergence criteria in the function is not met and fails to find b.