Data 512 Final Project
Gautam Moogimane

PREDICTING EARTHQUAKES

Is it even possible?

CLOSE, BUT NO CIGAR

WHY?

- Complexity – Too many factors affect the tectonic plates under the surface, making it hard to pinpoint when they will shift.

- Environment – These events happen deep below the earths surface, making it almost impossible to monitor effectively.

- Lack of signals – Scientists have tried to find precursory signals before a quake, but the distance, and not knowing where to look, complicate matters.

- History does not help – Trying to predict based on historical data is never accurate, and can only reflect probability of occurrence.

CAN MACHINE LEARNING HELP?

Yes – It is an upcoming field that is making advances everyday, and opens up entirely new ways to attack the problem.

Yes – With sufficient data and improvements in technology we might be able to narrow the prediction window, and improve the accuracy close to a 100%.

No – To effectively predict we need to pinpoint the exact time, location and magnitude, which cannot be achieved using current machine learning.

No – Statistics is all about probabilities, and no model can predict with a 100% accuracy all the time.

THE AIM IS NOT TO REVOLUTIONIZE, BUT TO EXPERIMENT AND UNDERSTAND

Research questions

- What would the approximate magnitude of an earthquake in the US in 2019 be?

- Which country has historically had the most number of earthquakes?

DATA

The data has been downloaded entirely from the

NOAA – list of significant earthquakes database public site

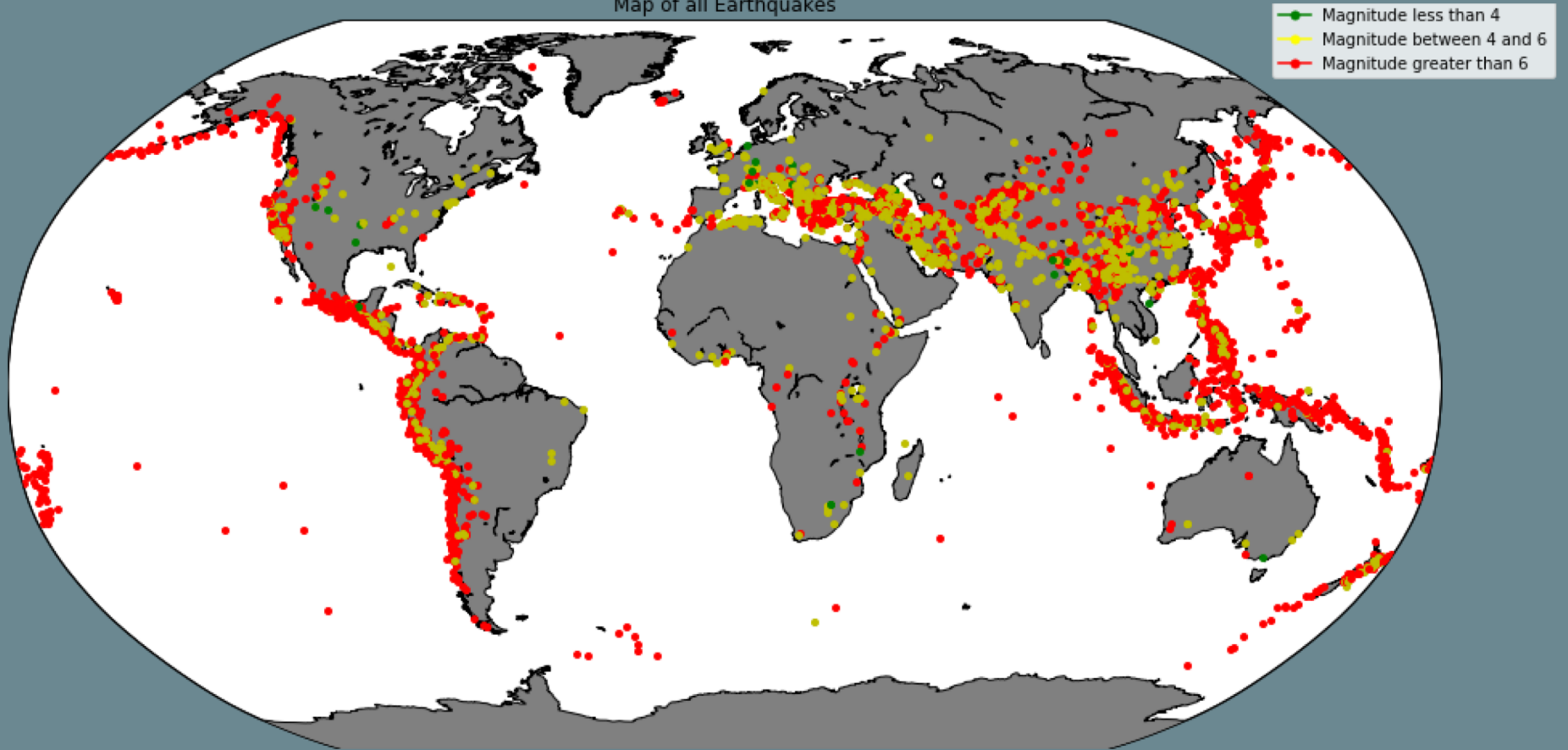https://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1

This list is historically relevant, but only contains significant earthquakes post 2150 BC.

The other comprehensive data site is hosted by USGS, but the data is post 1900 and there is a cap on the downloaded rows.

http://earthquake.usgs.gov/earthquakes/search/

Map of all Earthquakes

Legend:
- Magnitude less than 4
- Magnitude between 4 and 6
- Magnitude greater than 6

RESEARCH QUESTION - 1

What would the approximate magnitude of an earthquake in the US in 2019 be?

5.953

RESEARCH QUESTION - 2

Which country has historically had the most number of significant earthquakes?

China leads the pack with **559** big ones.

CONCLUSION

- The grouping of earthquakes on the map clearly highlights the two most damaging clusters, one along the North and South American west coast and other being around the Pacific ring of fire.

- None of the models tested were able to predict future earthquakes with a high degree of accuracy. The best score achieved was around 45% on the test data using random forest regression.

- The data has not been easy to source. Many sites have a limit on the number of rows that can be downloaded, and there is no central database to query.