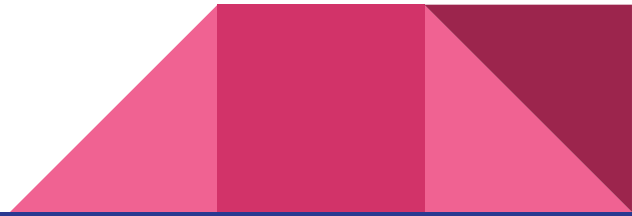
Data Acquisition & Management Final Project Presentation

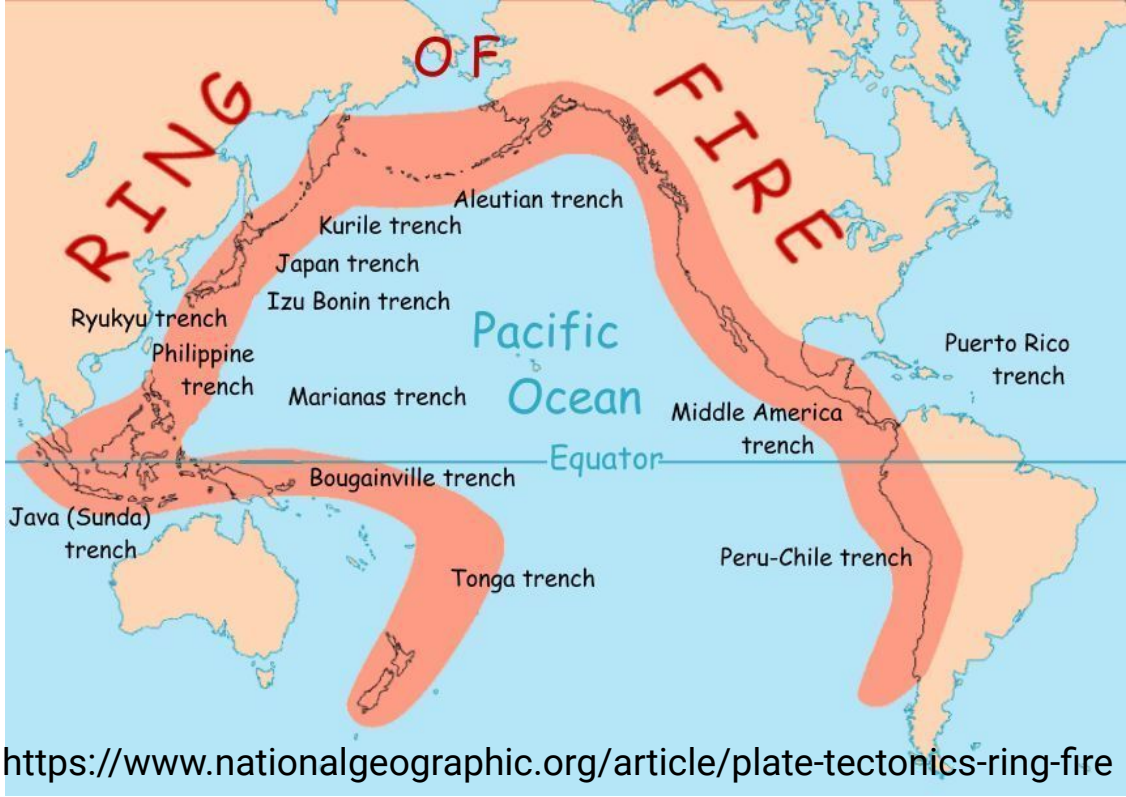
Marla Goodman

Table of Contents:



1. Introduction

- Destructive earthquakes: between ~ 5.5 - 8.9.
- Scientists expect ~ 17 major earthquakes in any given year.
- 80% - 90% of them will occur in the ‘Ring of Fire’



https://www.nationalgeographic.org/article/plate-tectonics-ring-fire

2. Research Questions

1. Which area(s) around the globe receive the most significant and/or frequent earthquakes, outside the “Ring of Fire?”

- Direct resources (advanced tech, money, humanitarian aid, better building materials)

2. How many ocean/coast earthquakes caused a subsequent tsunami? Is there a connection between earthquake magnitude/frequency and an ensuing tsunami?

- Again, to direct resources appropriately

3. Getting The Data

- NOAA: Kaggle.com by Chandra Shkehar (https://www.kaggle.com/shekpaul/major-earthquakes-noaa?select=NOAA+Earthquaqe+since+1600.csv).
- USGS:

```
from bs4 import BeautifulSoup
-Import bs4 as bs
-import urllib.request
-usgs=urllib.request.urlopen('https://www.usgs.gov/natural-hazards/earthquake-hazards/science/20-largest-earthquakes-world?qt-science_center_objects=0#qt-science_center_objects').read()
-souped = bs.BeautifulSoup (usgs, 'lxml')
-table = souped.find('table')
-table_rows = table.find_all('tr')
-for tr in table_rows:
      td = tr.find_all('td')
       v = [data.text for data in td]
      print(v)
-challenge
```

4. Data Tinkering

- **NOAA:**
  -Dropped irrelevant columns
  -1960 research (noaa)
  -Still left with almost 2,500 rows

- **USGS:**
  -Named and dropped columns
  -transform data from object to datetime in order to get just the year

	Year	Tsunami	Name	Magnitude	Deaths	Injuries	Damage $Mill
3029	1960	Yes	PERU: AREQUIPA,CHUQUIBAMBA,CARAVELI,COTAHUASI	7.8	63.0	200.0	NaN
3030	1960	NaN	PERU: LIMA,NAZCA,ICA,HUANCAVELIC,PALPA,HUAITARA	7.0	63.0	NaN	NaN

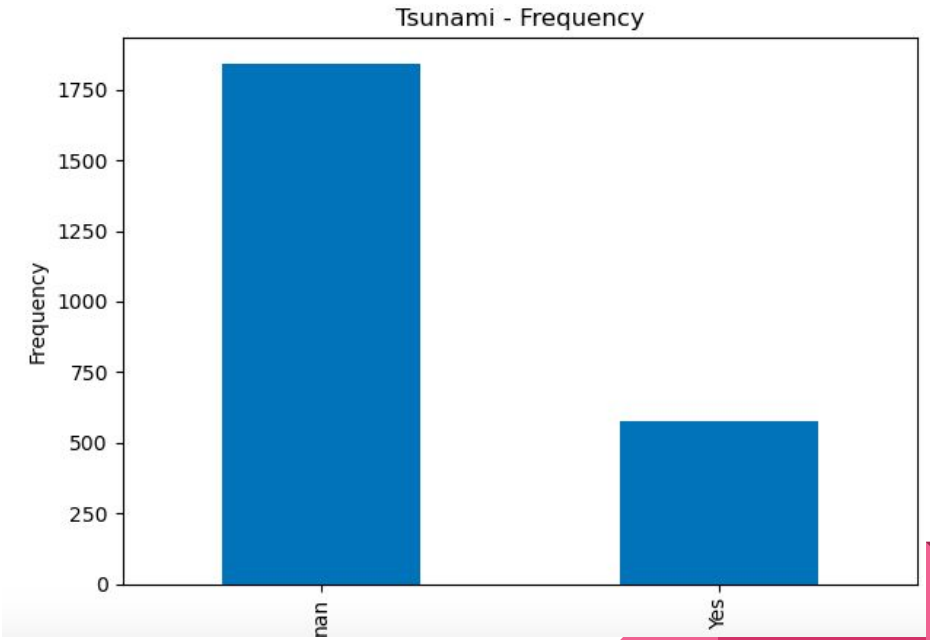
	Mag	Location	Year
0	9.5	Bio-Bio, Chile	1960
1	9.2	Southern Alaska	1964

5. Brief Exploratory Data Analysis: Tsunami

-noaa['Tsunami'].describe()

count	578
unique	1
top	Yes
freq	578

-tsun=(noaa['Tsunami'].value_counts(dropna=False))
-tsun.plot.bar(title='Tsunami - Frequency')



5. Brief Exploratory Data Analysis: Location (noaa)

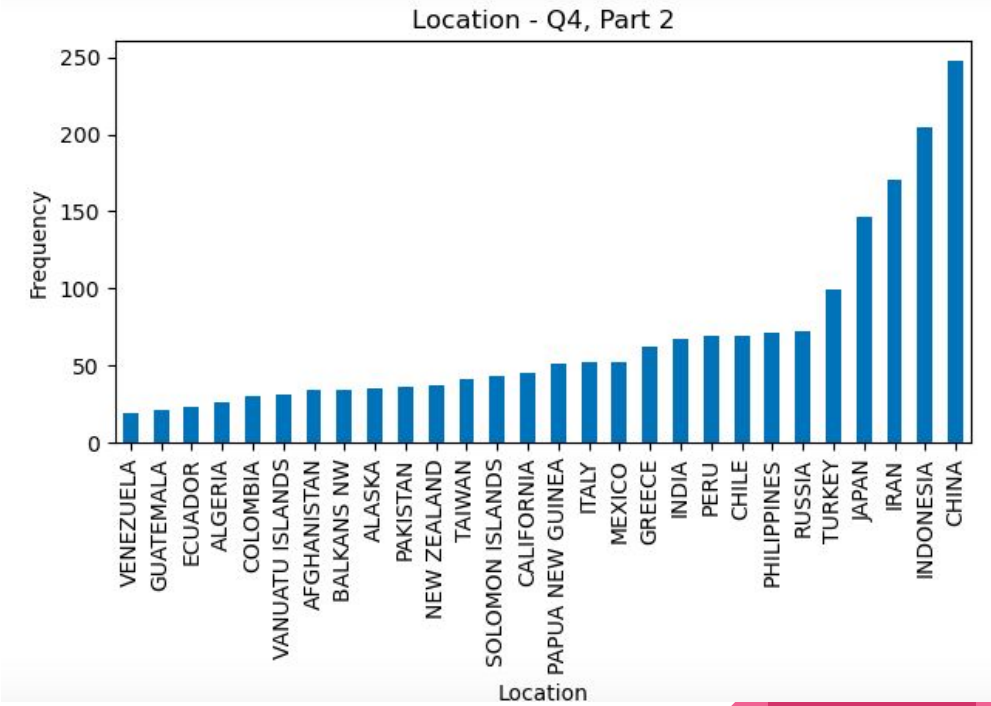
-location[0].describe()

count	2419
unique	203
top	CHINA
freq	248

-loc=location[0].value_counts().sort_values()

-quarter4=loc.iloc[150:203] to get the most frequent

-plotted second part of quarter 4



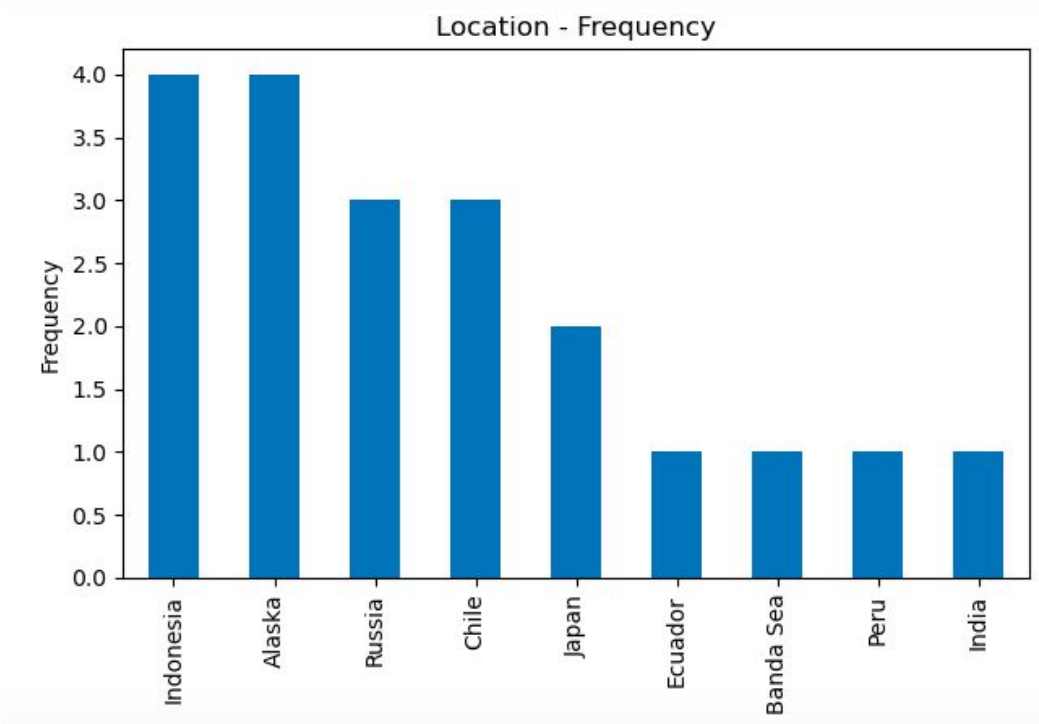
Many of these values make sense in that they intersect the 'Ring of Fire'.

But there are some that don't intersect this ring, yet still have a significant number of earthquakes. I will analyze these places for my research questions.

5. Brief Exploratory Data Analysis: Location (usgs)

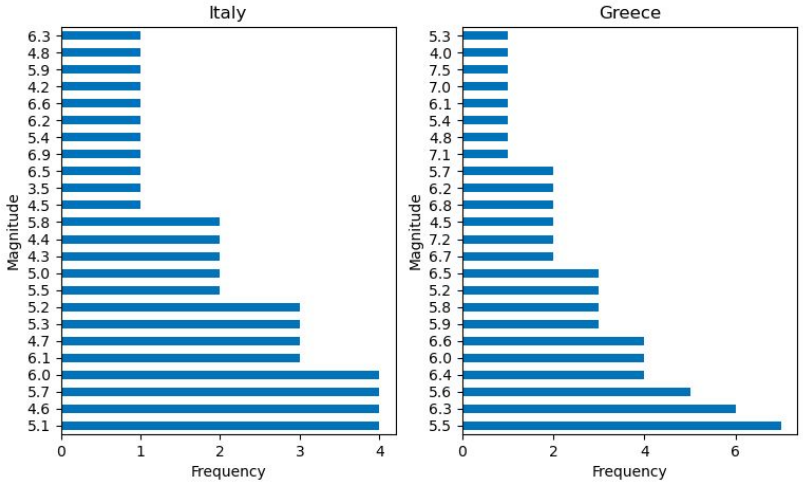
-Plotting by frequency (value_counts())

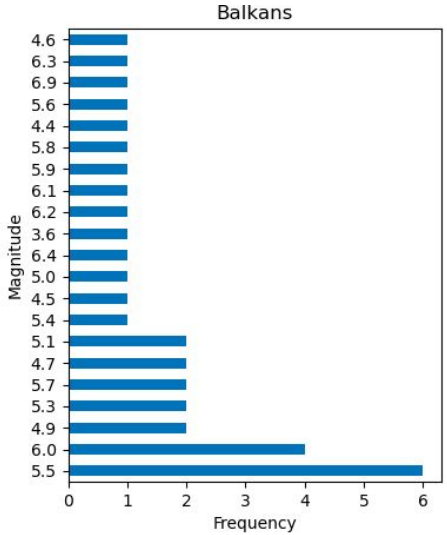
-Not much to go off

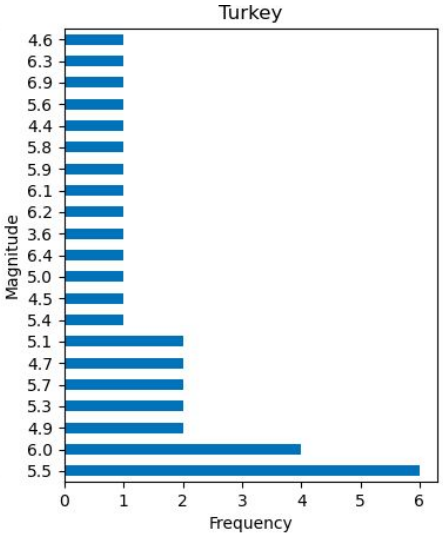


6. Analysis For Research Questions: Q1

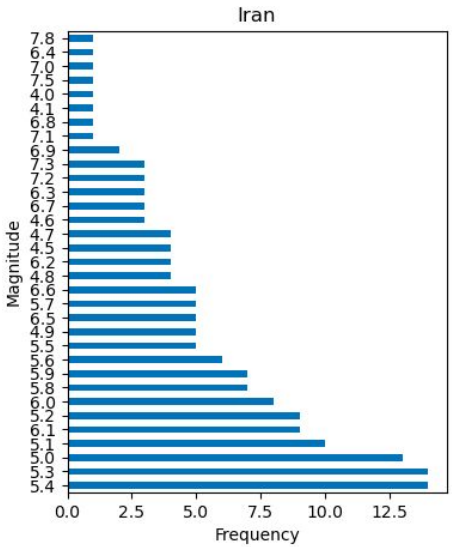
- Usgs data not too helpful
- chart1=(a['Magnitude'].value_counts())
- chart1.plot.barh(title='Italy')
- Google Maps Screenshot → → → → → →

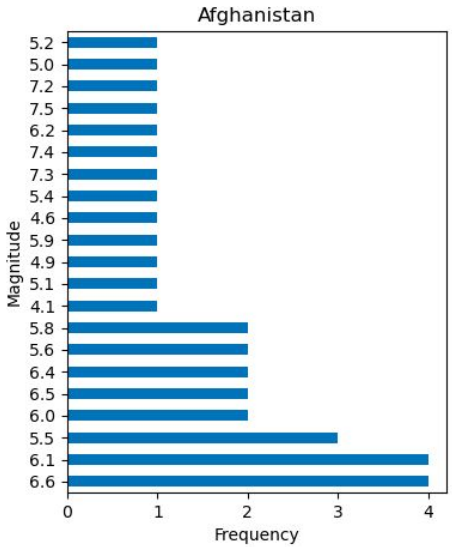


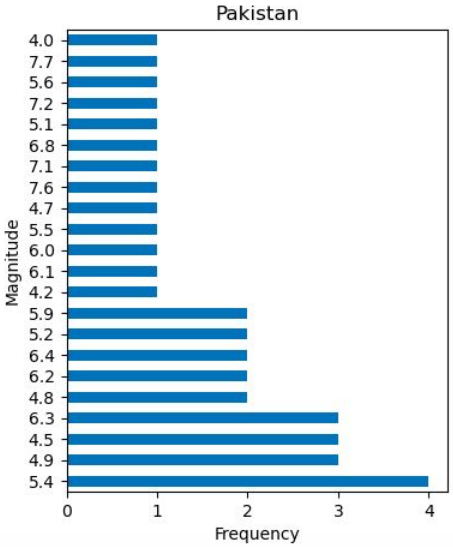


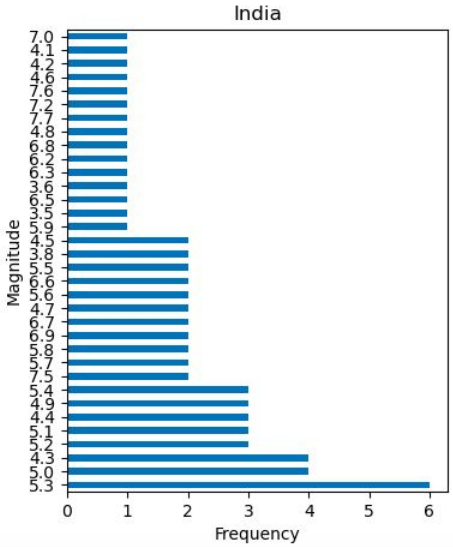


6. Analysis For Research Questions: Q1 (Cont.)









The eight areas see frequent and significant earthquakes.

6. Analysis For Research Questions: Q2

- tsun_df.groupby('Place').count().sort_values(by='Tsunami').tail(15) →
- Most places with 10 tsunamis or above intersect the 'Ring of Fire'. Japan has the most at 91 tsunamis.
- Maybe a correlation, causation not determined
- Problem - example of China
- Problem with Italy, Greece, the Balkans, Turkey, Iran, Pakistan, and India: most have tsunami frequency of 1 or 2, Turkey and Greece have tsunami frequencies of 5 and 18 respectively.

Place	Year	Tsunami
TAIWAN	11	11
PERU	12	12
NEW CALEDONIA	13	13
NEW ZEALAND	13	13
MEXICO	18	18
GREECE	18	18
VANUATU ISLANDS	19	19
PHILIPPINES	19	19
ALASKA	28	28
PAPUA NEW GUINEA	28	28
SOLOMON ISLANDS	31	31
CHILE	40	40
RUSSIA	52	52
INDONESIA	54	54
JAPAN	91	91

IRAN	1	1
PAKISTAN	2	2
PANAMA	1	1
CHINA	2	2

Answers/Conclusions

1. “The **Alpide earthquake belt**...accounts for about 17 percent of the world's largest earthquakes” -https://www.usgs.gov/faqs/where-do-earthquakes-occur?qt-news_science_products=0#qt-news_science_products
2. https://wn.com/alpide_belt/wikipedia → → → → -->
3. The data did identify another earthquake hotspot that I had not previously been aware of. Looks like scientists and other analysts got there first though.
4. Tsunami question could not be answered with the data chosen
5. USGS data set of top 20 largest earthquakes was not very useful
6. For future earthquake analysis, pick other data sets than usgs
7. For future tsunami analysis, I would get more data sets and do more research related to tsunamis