Real Time Earthquake Analysis

Department of Applied Data Science

San Jose State University

DATA 230: Data Visualization

Professor: Andrew H. Bond

May 17, 2022

By

Ahaz Bhatti

011109033

Table of Content

1.Introduction **3**

1.1 Purpose of this Document **3**

1.2 Intended Audience **3**

1.3 Scope **3**

1.4 About The dataset **4**

1.4.1 Definitions **4**

2. Background and Objectives **6**

3. Architecture and High level design **7**

4.Organization **8**

4.1 Project group **8**

4.2 Customer **8**

5. Development Process **8**

6. Deliverables **12**

7. Github **12**

8. Charts **13**

9. Dashboards **23**

10. References **26**

1.Introduction

1.1 Purpose of this Document

The purpose of this document is to provide a detailed project description of Real Time Earthquake data analysis, which is designed to assess earthquakes around the world by visualizing earthquake patterns and statistics. This document includes details about organization, roles, deliverables, project risks, time plans and financial plans.

1.2 Intended Audience

This document shall be used in all phases of the project as a guideline. Intended audiences of this project are all project stakeholders:

- Professor Andrew Bond
- TA Ritanjali Jena
- Ahaz Bhatti

1.3 Scope

This document shows the project plan for the analysis of Real Time Earthquake data using Tableau. The document includes the purpose of the project, organization plan for the project, data collection, and applications used. The data for the project includes information about different types of earthquakes which occur all around the world, so the scope of data is available all throughout the world and not just the United States. The data only goes back 30 days and refreshes every minute. The live data is collected from the United States Geological Survey website.

1.4 About The dataset

1.4.1 Definitions

The dataset contains following fields

Brief about the Earthquake event terms/ variables:

- Depth- Depth of the event in kilometers. The depth where the earthquake begins to rupture. This depth may be relative to the WGS84 geoid, mean sea-level, or the average elevation of the seismic stations which provided arrival-time data for the earthquake location.
- DepthError- Uncertainty of reported depth of the event in kilometers.
- Dmin- Horizontal distance from the epicenter to the nearest station (in degrees). 1 degree is approximately 111.2 kilometers. In general, the smaller this number, the more reliable is the calculated depth of the earthquake.
- Gap- The largest azimuthal gap between azimuthally adjacent stations (in degrees). In general, the smaller this number, the more reliable the calculated horizontal position of the earthquake. Earthquake locations in which the azimuthal gap exceeds 180 degrees typically have large location and depth uncertainties.
- HorizontalError- Uncertainty of reported location of the event in kilometers.
- Id- A unique identifier for the event. This is the current preferred id for the event, and may change over time. See the "ids" GeoJSON format property.

- Latitude- Decimal degrees latitude. Negative values for southern latitudes.
- LocationSource- The network that originally authored the reported location of this event.
- Longitude- Decimal degrees longitude. Negative values for western longitudes.
- Mag- The magnitude for the event. Earthquake magnitude is a measure of the size of an earthquake at its source.
- MagError- Uncertainty of reported magnitude of the event.
- MagNst- The total number of seismic stations used to calculate the magnitude for this earthquake.
- MagSource- Network that originally authored the reported magnitude for this event.
- MagType- The method or algorithm used to calculate the preferred magnitude for the event.
- Net- The ID of a data contributor. Identifies the network considered to be the preferred source of information for this event.
- Nst- The total number of seismic stations used to determine earthquake location.
- Place- Textual description of named geographic region near to the event. This may be a city name, or a Flinn-Engdahl Region name. If there is no nearby city within 300 kilometers (or if the nearby cities database is unavailable for some reason), the Flinn-Engdahl (F-E) seismic and geographical regionalization scheme is used.
- Rms-The root-mean-square (RMS) travel time residual, in sec, using all weights. This

parameter provides a measure of the fit of the observed arrival times to the predicted arrival times for this location. Smaller numbers reflect a better fit of the data.

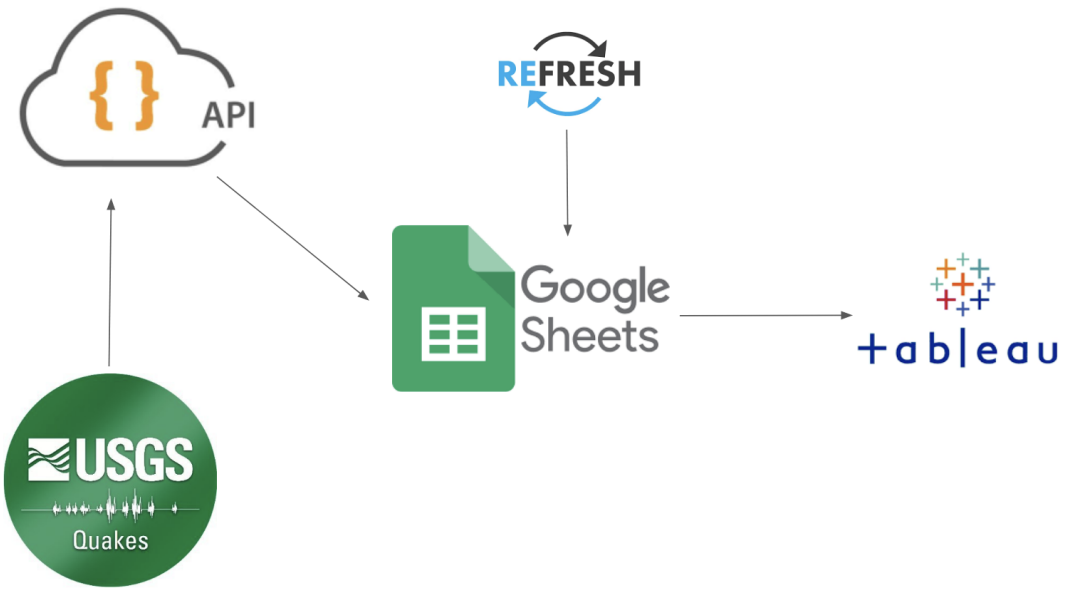
- Status- Indicates whether the event has been reviewed by a human. Status is either automatic or reviewed. Automatic events are directly posted by automatic processing systems and have not been verified or altered by a human. Reviewed events have been looked at by a human. The level of review can range from a quick validity check to a careful reanalysis of the event.
- Time- Time when the event occurred. Times are reported in milliseconds since the epoch (1970-01-01T00:00:00.000Z), and do not include leap seconds. In certain output formats, the date is formatted for readability.
- Type- Type of seismic event.
- Updated- Time when the event was most recently updated. Times are reported in milliseconds since the epoch

2. Background and Objectives

An earthquake is a vibration or shaking that occurs due to a shift or movement of the earth layer originating from the base of the earth's surface. Earthquakes are recorded from Seismometers. The variable for Seismometers is called Nst, and is on the visuzation, how many seismic stations there are in total for an earthquake location. According to the United States Geological Survey website the USGS monitors and reports earthquakes, assesses their consequences and risks, and conducts specialized research on earthquake sources and effects. These efforts are part of the

National Earthquake Hazards Reduction Program (NEHRP), a four-agency collaboration established by Congress. This data is updated every minute and can go back up to 30 days. The dataset contains historical data as well for earthquakes. There is also an option to select real time data only for the past hour, day, week and month.

3. Architecture and High level design



The data for this Real Time Earthquake Analysis comes from one source

1. United States Geological Survey - This is an official website of the government and the data is updated every minute
2. API - This website is an open source website which contains more detailed data about Earthquakes and contains for 4 API's depending on the duration of real time data needed

The data was collected from the USGS website, and what I mean is I took the API from the USGS website and it was in xls format.

After the initial data collection, all the data was imported into Google Sheets and because this is real time data, I was not able to use tableau prep builder to fix some data issues. However, within Tableau I Capitalized all columns and renamed some columns such as (mag) to (Magnitude) and changed (depth) to (Depth km) to understand which measurement is being used. Some changes I would fix in tableau prep builder is remove rows with many missing values and remove outliers. Finally, I was able to use this live data imported into Tableau for real time data analysis and visualization of the data.

4.Organization

4.1 Project group

It is an individual work done by Ahaz Bhatti (011109033), San Jose State University

4.2 Customer

The target customers are listed below:

1. Professor Andrew Bond
2. TA Ritanjali Jena/Students

5. Development Process

The process for the project starts from collecting the data. For analyzing we needed realtime data about Earthquakes. For this purpose, the data was collected from the United States Geological Survey which contained all the data from the government database about earthquakes which happened over the past 30 days.

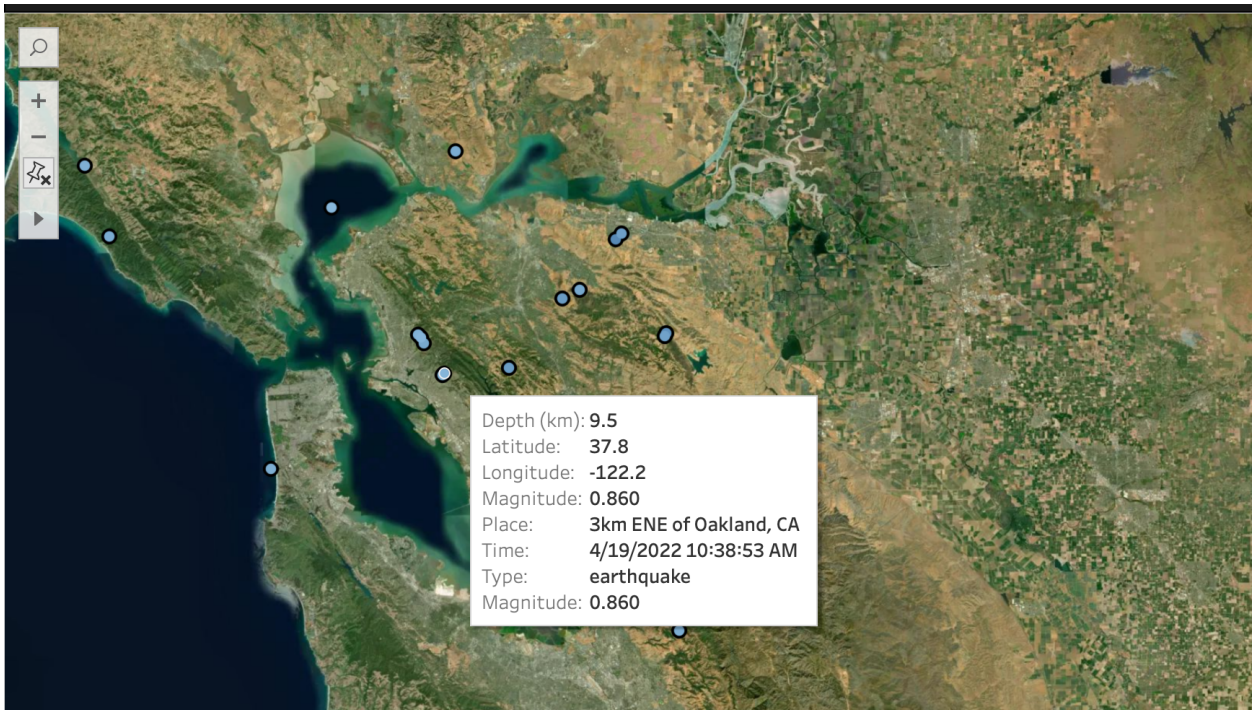
Once the APPI was collected the individual xls file was imported to Google Sheets using the import function. Google Sheets is a spreadsheet just like excel and real time data is stored inside this sheet and updated every minute.

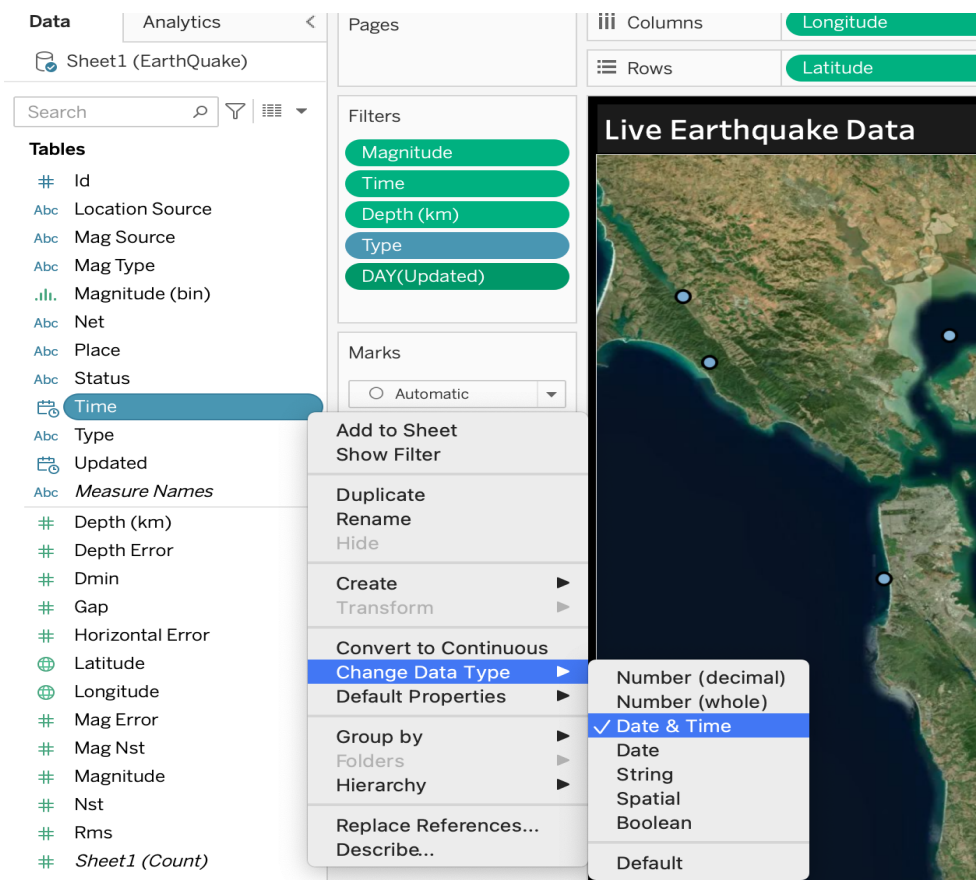
The data types for the columns in each variable were changed according to the data it contained, for altering the data type for the Time column from string to Date, and so on.

Many rows contained null values, it made some rows meaningless, however, the data was still efficient. Some of the rows that contain null values were HorizontalError, DepthError, MagError, MagNst and more. However, the lack of missing values did not affect the th visualizations.

P	Q	R	S
horizontalError	depthError	magError	magNst
0.35	0.74		1
9.5	1.9	0.098	10
	0.4		
0.35	0.78	0.125	17
6	5.3	0.049	122
0.27	0.89		1
	1.1		
0.29	0.73		1
12.5	1.9	0.113	26
	0.5		
	0.6		
0.34	0.74	0.135	25
0.24	0.38	0.1	6
0.6	0.85	0.1005295983	5
0.55	0.85	0.204	8
0.49	1.94		1
	0.4		
0.36	1.15	0.17	2
9.5	1.9	0.089	12
14.6	23.4	0.091	31
1.5	1.9	0.038	92
	0.4		

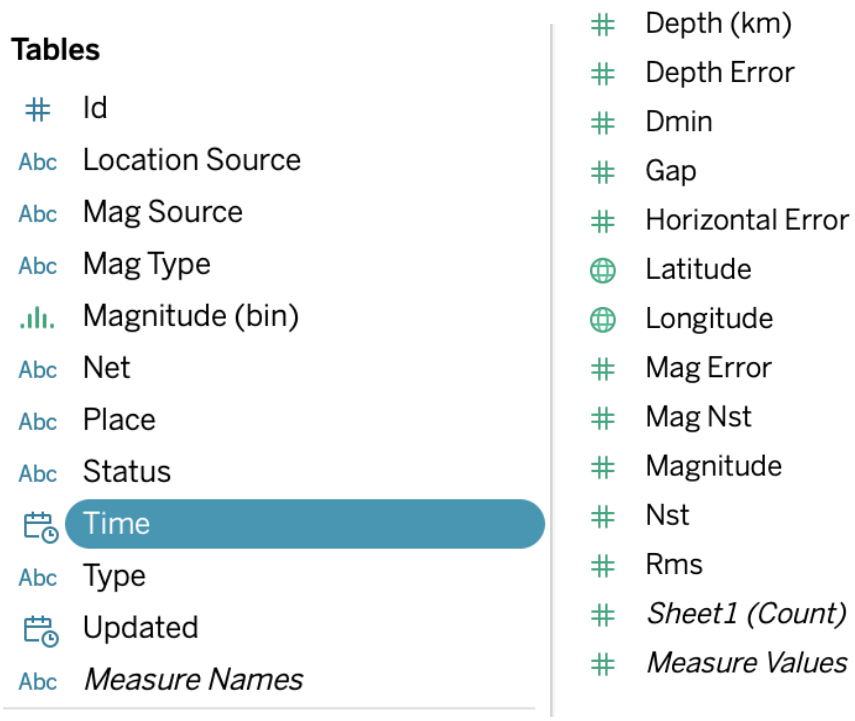
Initially, the time field was not working to showcase what time the earthquake occurred. The value for it was coming out blank when I marked it for my map visualization. Shortly, after I noticed the Time variable was in string format and it was not working properly. So, first I changed the data type to time and afterwards I selected date and time for the time variable, so the date and time would both show.


Depth (km): 9.5
Latitude: 37.8
Longitude: -122.2
Magnitude: 0.860
Place: 3km ENE of Oakland, CA
Time: 4/19/2022 10:38:53 AM
Type: earthquake
Magnitude: 0.860

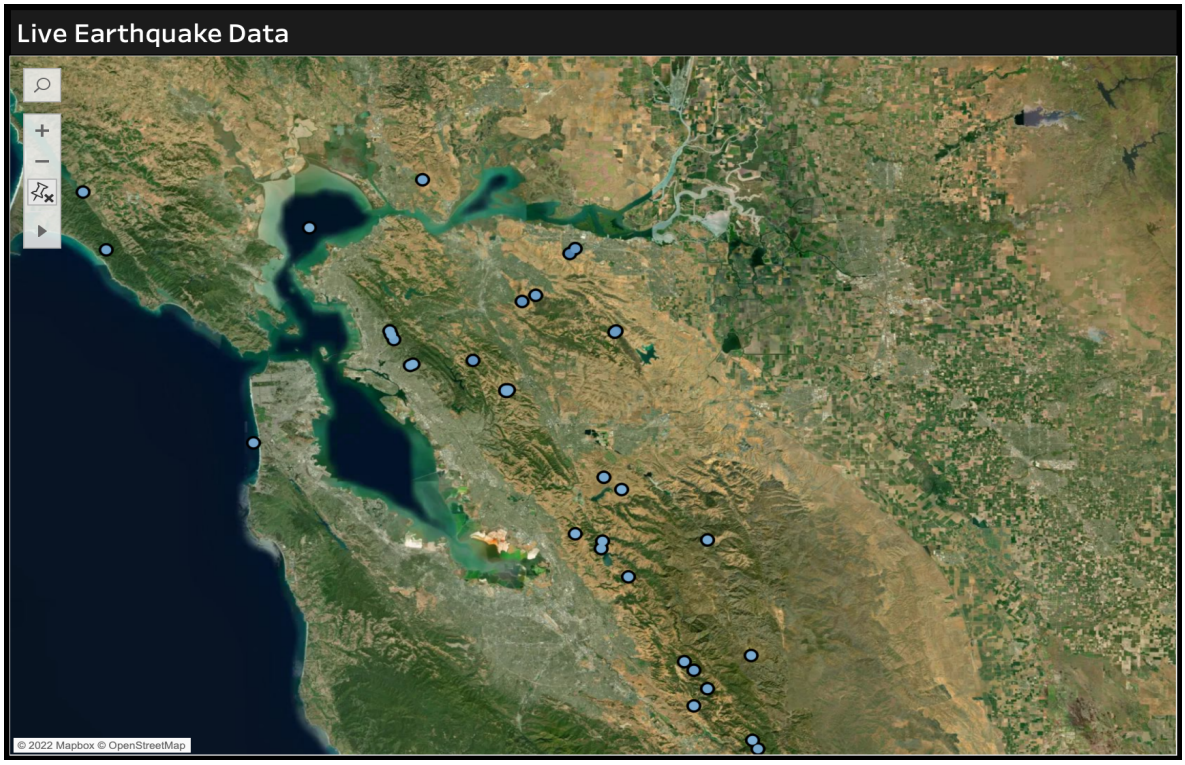

Data
Analytics
Sheet1 (EarthQuake)
Search
Tables
Id
Location Source
Mag Source
Mag Type
Magnitude (bin)
Net
Place
Status
Time
Type
Updated
Measure Names
Depth (km)
Depth Error
Dmin
Gap
Horizontal Error
Latitude
Longitude
Mag Error
Mag Nst
Magnitude
Nst
Rms
Sheet1 (Count)
Pages
Filters
Magnitude
Time
Depth (km)
Type
DAY(Updated)
Marks
Automatic
Add to Sheet
Show Filter
Duplicate
Rename
Hide
Create
Transform
Convert to Continuous
Change Data Type
Default Properties
Group by
Folders
Hierarchy
Replace References...
Describe...
Number (decimal)
Number (whole)
Date & Time
Date
String
Spatial
Boolean
Default
Columns
Longitude
Rows
Latitude
Live Earthquake Data

Now the dataset was ready to be used for visualizations. However, the dataset contained 22 variables, and the names were confusing, I had to make sure formatted the column names correctly so it could be understandable while making the visualizations

The figure below shows the columns and how all of them are capitalized and how names such as mag has been renamed to Magnitude, this made it easier to recognize the columns when visualizing.



Once all the data was ready it was time to perform analysis and create visualizations in Tableau. The figure below shows the sample of my first visualization with real time data.



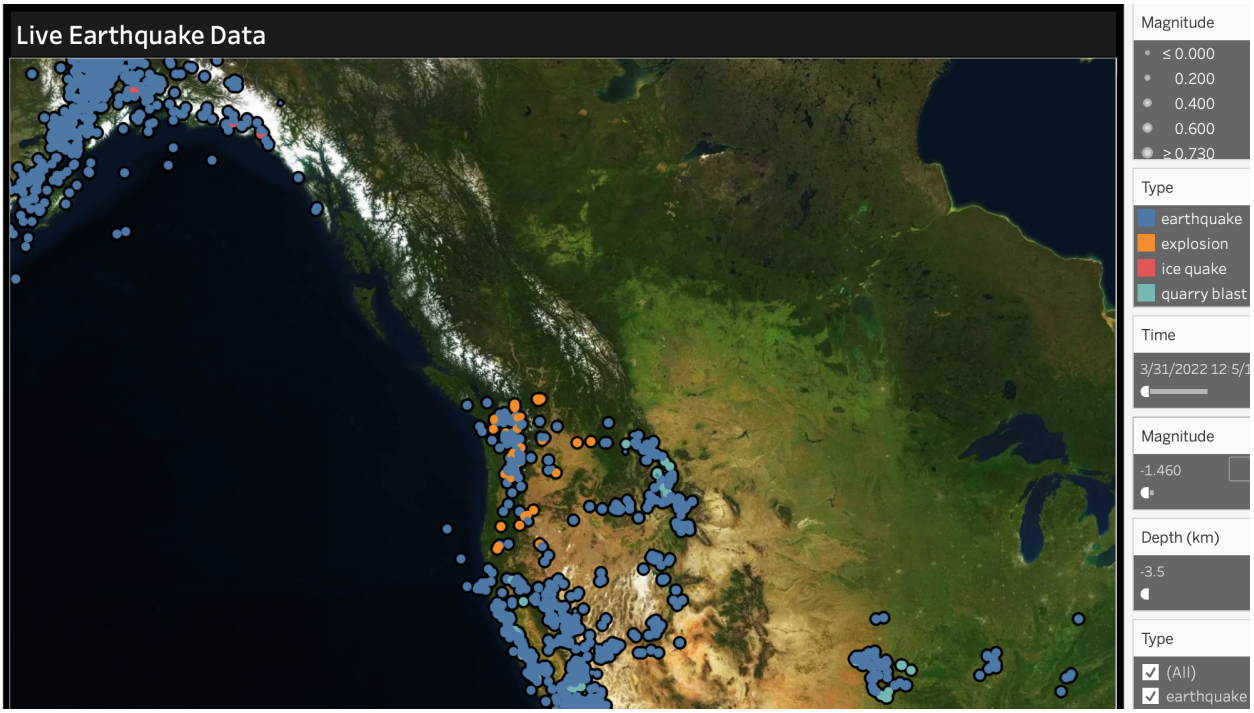
6. Deliverables

The most important deliverable of this project is to provide 3 important dashboards for overall distributed analysis. A project presentation is also given. And the Final Project report will be submitted. All the data processing and visualization are to be delivered.

7. Github

https://github.com/ahazbhatti/Earthquake-Analysis-Tableau

8. Charts



The are 4 types of quakes

Type			
earthquake	explosion	ice quake	quarry blast
3,392	37	6	65

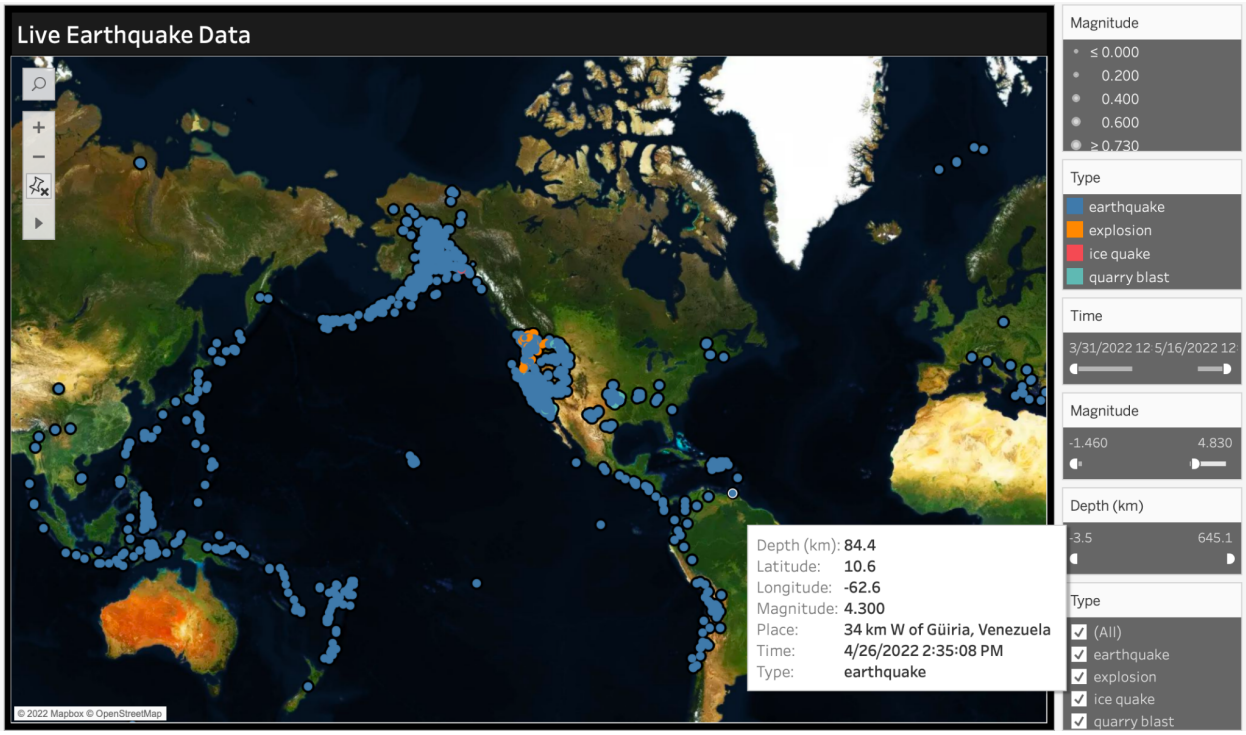
The standard is the earthquake which happens when the Earth's surface is violently shaken. Movements in Earth's outermost layer are causing the shaking. In the last 30 days there were 3,392 earthquakes around the world.

Next up is quarry blast, this occurs in mining when a chemical and physical procedure occurs through the use of explosives. In the last 30 days there were 65 quarry blasts around the world.

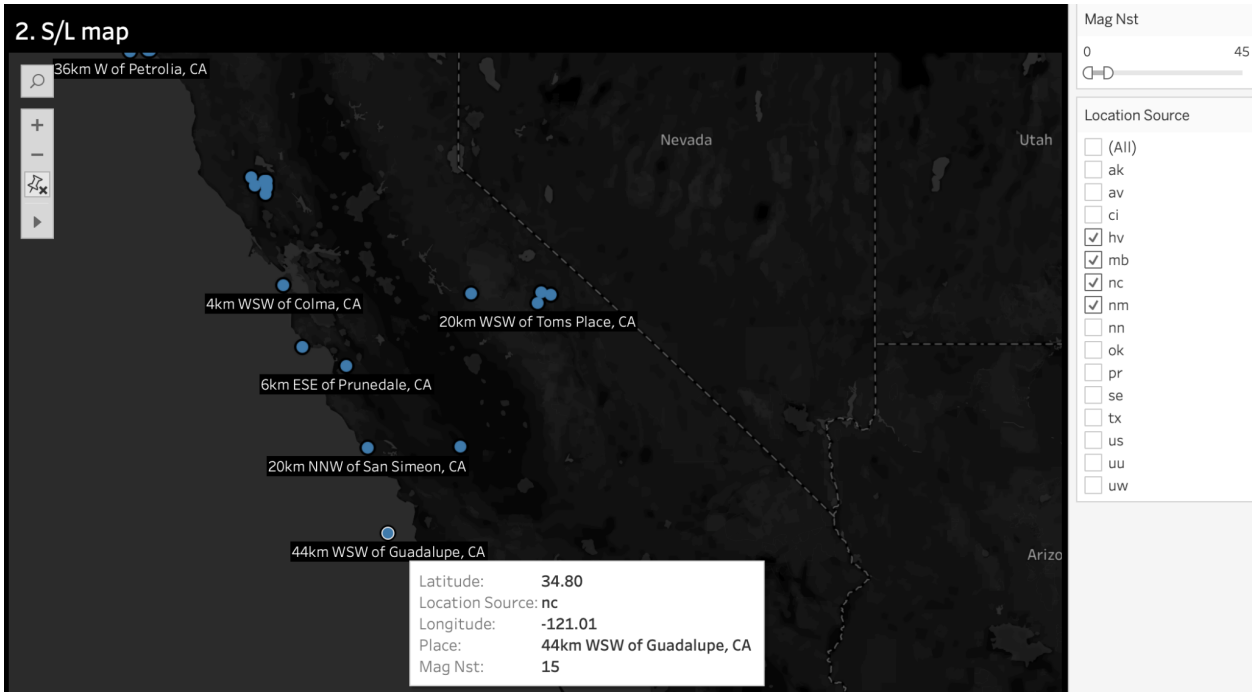
Then we have an explosion. An explosion is a rapid volume expansion accompanied by a

massive outward release of energy, usually accompanied by high temperatures and the release of high-pressure gasses. In the last 30 days there were 37 quarry blasts around the world.

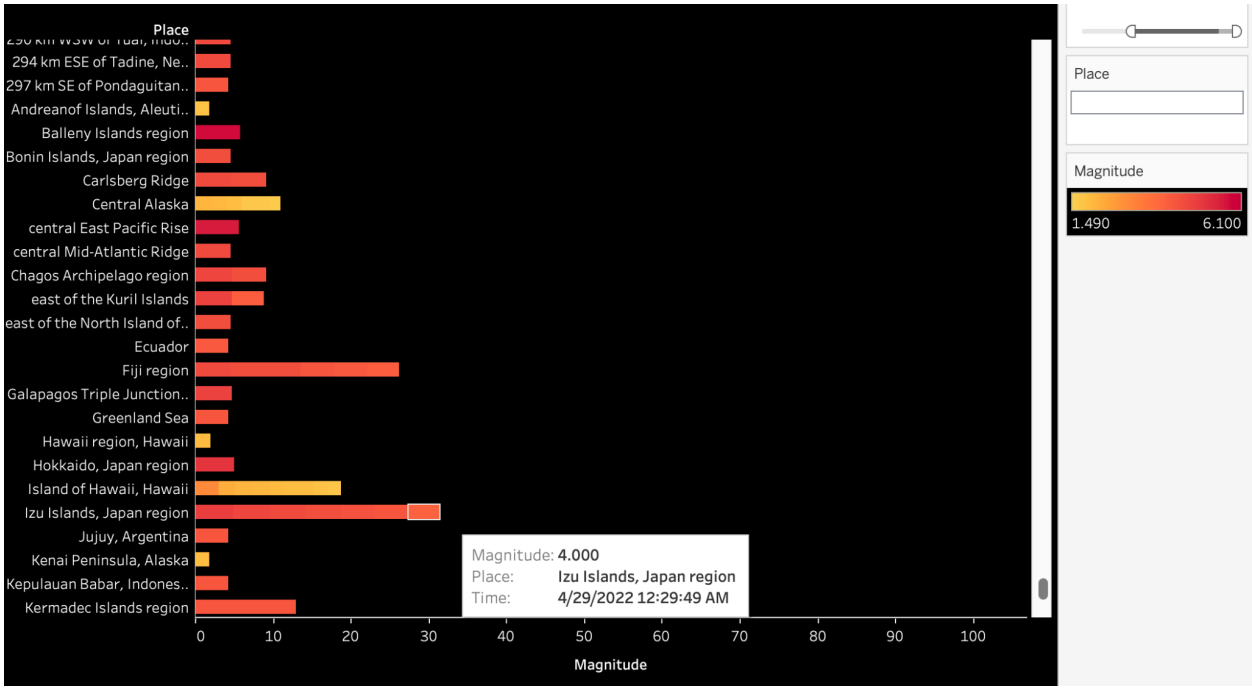
Lastly, we have ice quakes, this occurs when an ice field or other huge expanse of ice fractures due to lateral pressure, it causes a rapid shaking or sequence of abrupt motions. In the last 30 days there were 6 quarry blasts around the world.



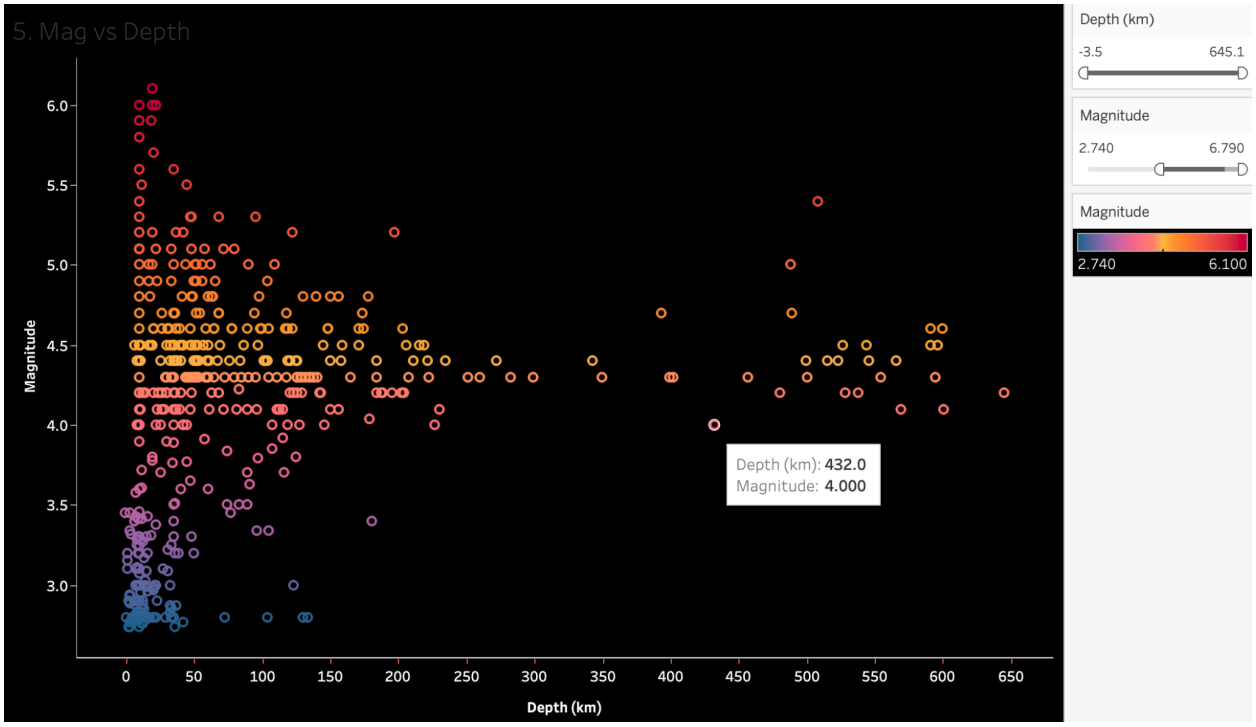
This map contains live earthquake data, it shows all the earthquakes around the world and their location and place. You can also see each location's magnitude for the earthquakes it experienced. You can also see the depth of the earthquake and change the time it occurred as well, same goes for magnitude and depth. One thing to note is that because this is real time, moving around may lag to get updated data.



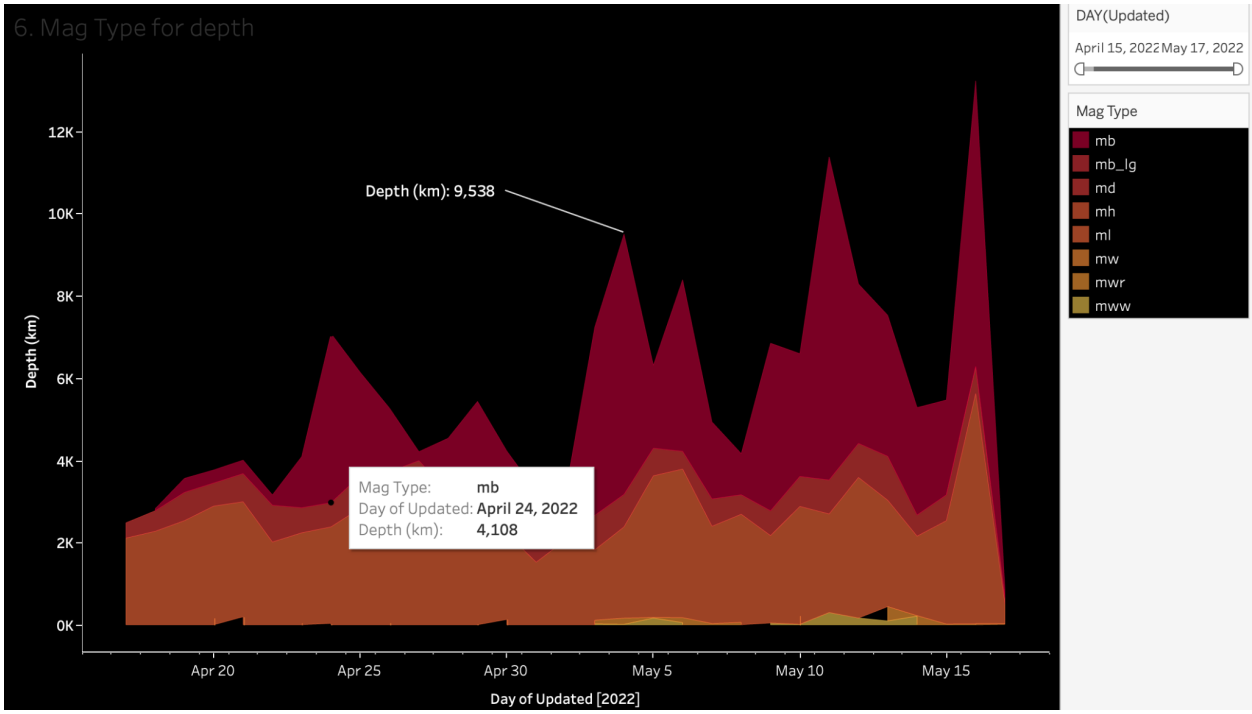
This map contains each location's Nst for its Magnitude. Nst's are seismic stations, these stations tell us where the earthquake happened and catches its magnitude. For example in the image above, 15 stations captured the earthquake for Guadalupe in California. Nst stations are all around the world to detect earthquakes. The more Nst's you have the more accurate readin you can get.



This is a chart showing horizontal bars of all the places that experienced an earthquake. Each place has a bar showing the magnitude of the earthquake, longer the bar, bigger the magnitude for the earthquake. The chart also contains the time of each earthquake respective to its place. There's many places where earthquakes occurred, users can also filter through place the are looking for in the right hand corner.

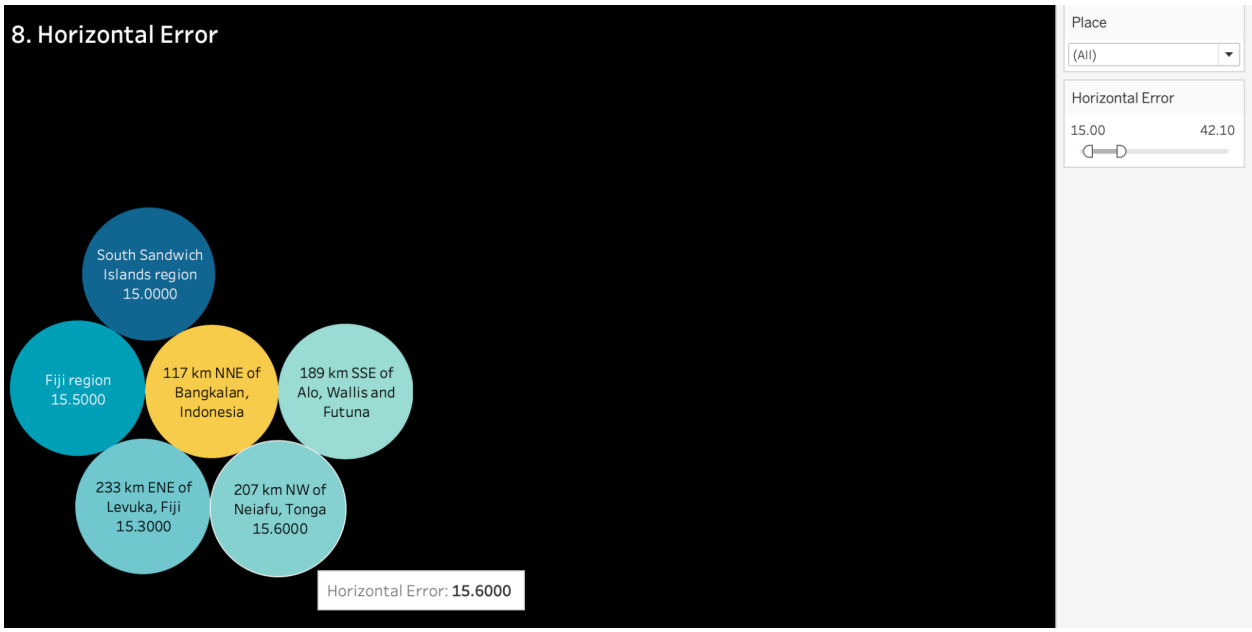


This chart contains scatter plots for Magnitude and Depth for an earthquake. Depth can be controlled on the right hand side, magnitude can be controlled as well. In the chart you can see an earthquake of 4.0 magnitude occurred at a depth of 432.0 kilometers. The colors also represent how intense the earthquake is, if the magnitude is high then the color will be more red, and lighter the magnitude it will be light blue.

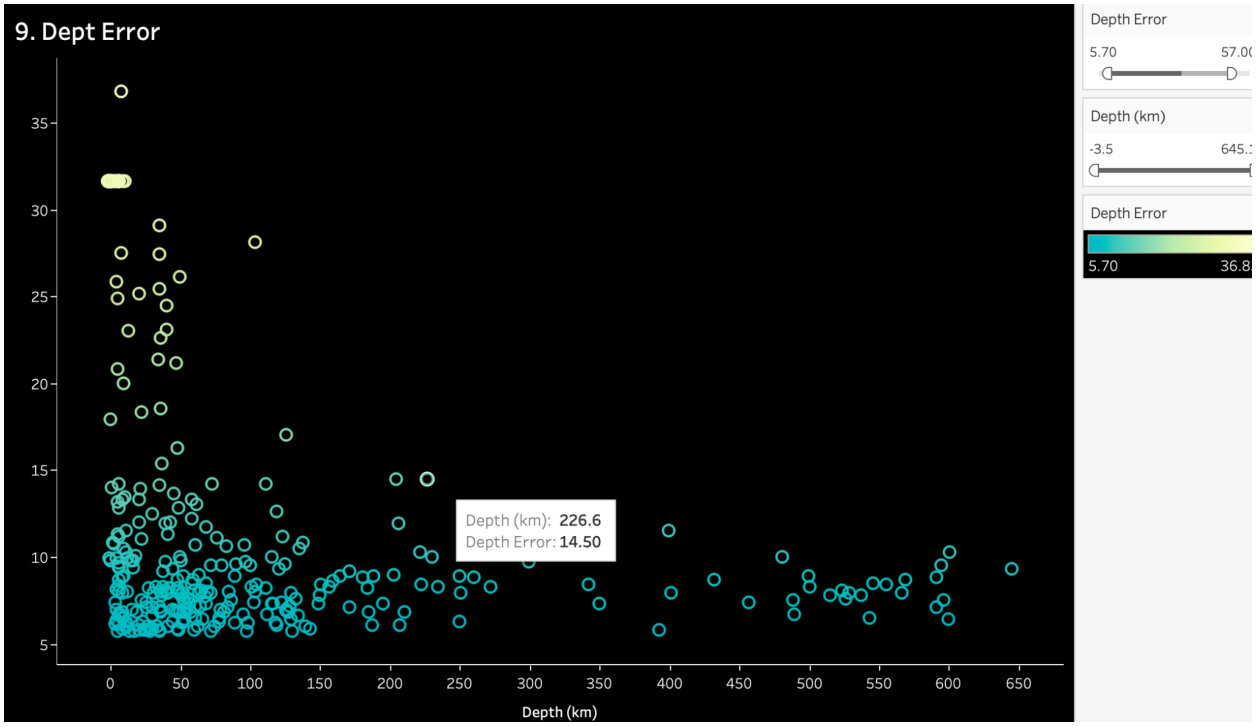


This is an area chart containing the depth of the earthquake for the past 30 days. The area is updated everyday and the days can be changed to see a specific depth for a specific day. One thing to note is that the area contains the type of magnitude that was used to calculate the magnitude of the earthquake. Earthquakes that contain a high magnitude are calculated by mb, and the lower ones with mb_lg, md and so on, depending on the magnitude.

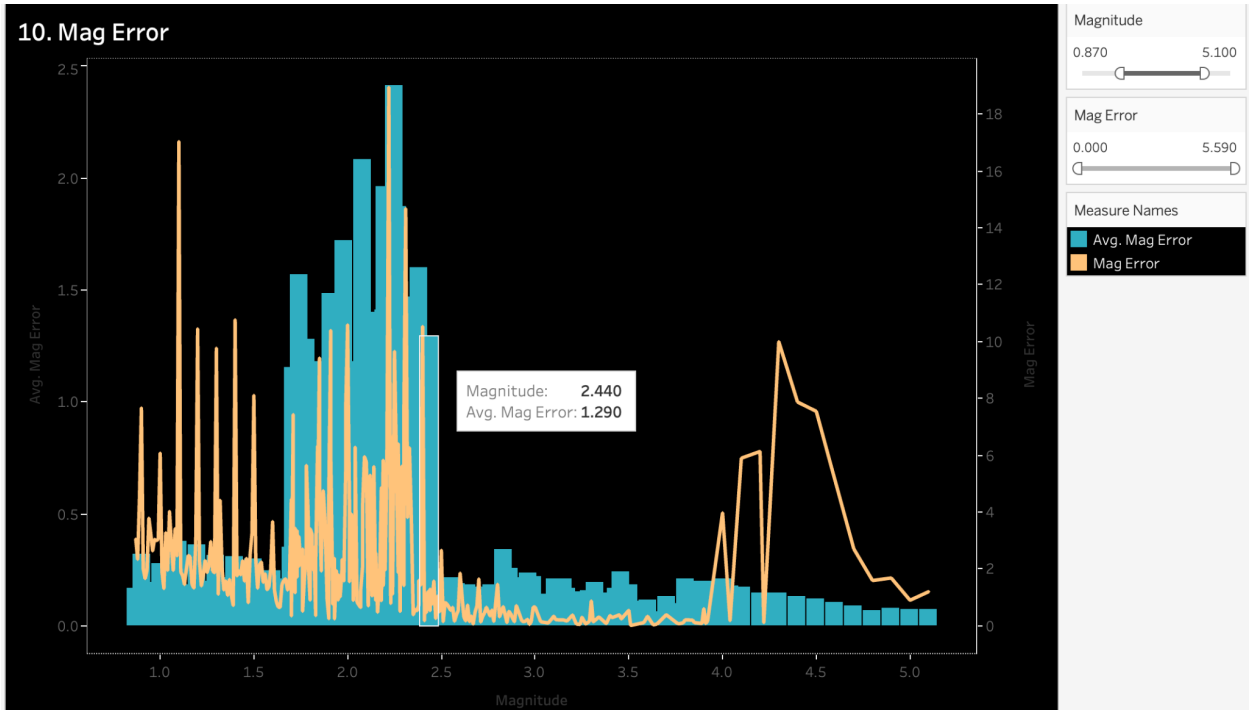
The figure above is a pie chart containing information whether the earthquake was reviewed by a human or not. This basically tells us whether the earthquake was reviewed by a human or automatic. Automatic means that the event was processed by a system and not a human, meaning the human had not verified the earthquake. If it is reviewed then the human has looked at the earthquake.



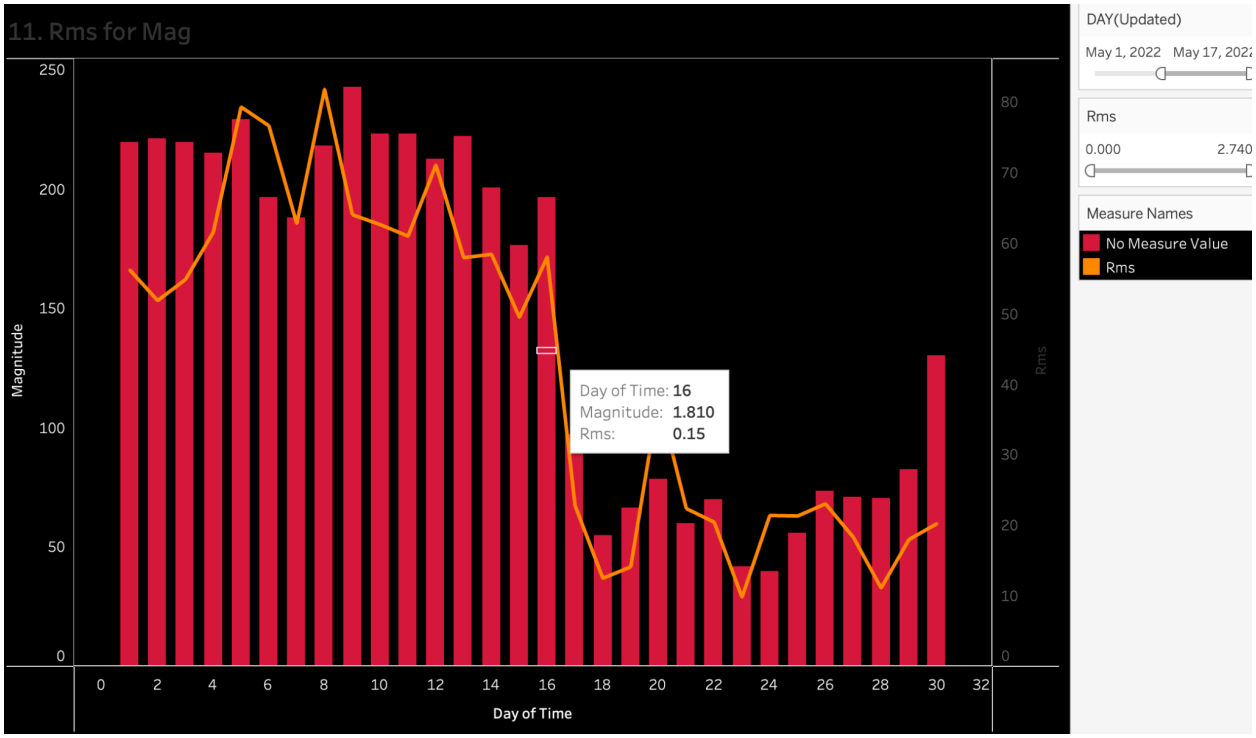
This is a packed bubble chart and contains information about the Horizontal Error for each location that experienced an earthquake. The bubble contains the city which experienced an earthquake and also contains the horizontal error when you hover it, if the bubble is small and doesn't contain it. Horizontal error is the uncertainty in the earthquake when it occurs. As you can see the graph Tonga had an uncertainty of 15.60.



This chart is a scatter plot. In this chart we can see the Depth Error for the Depth of an earthquake. You have an option on the right to filter through the Depth Error and Depth for the event. Depth Error is basically the uncertainty when the depth was recorded in kilometers. So in the image above you can see that an event which had a depth of 226.6 had an uncertainty of 14.50 in the graph.

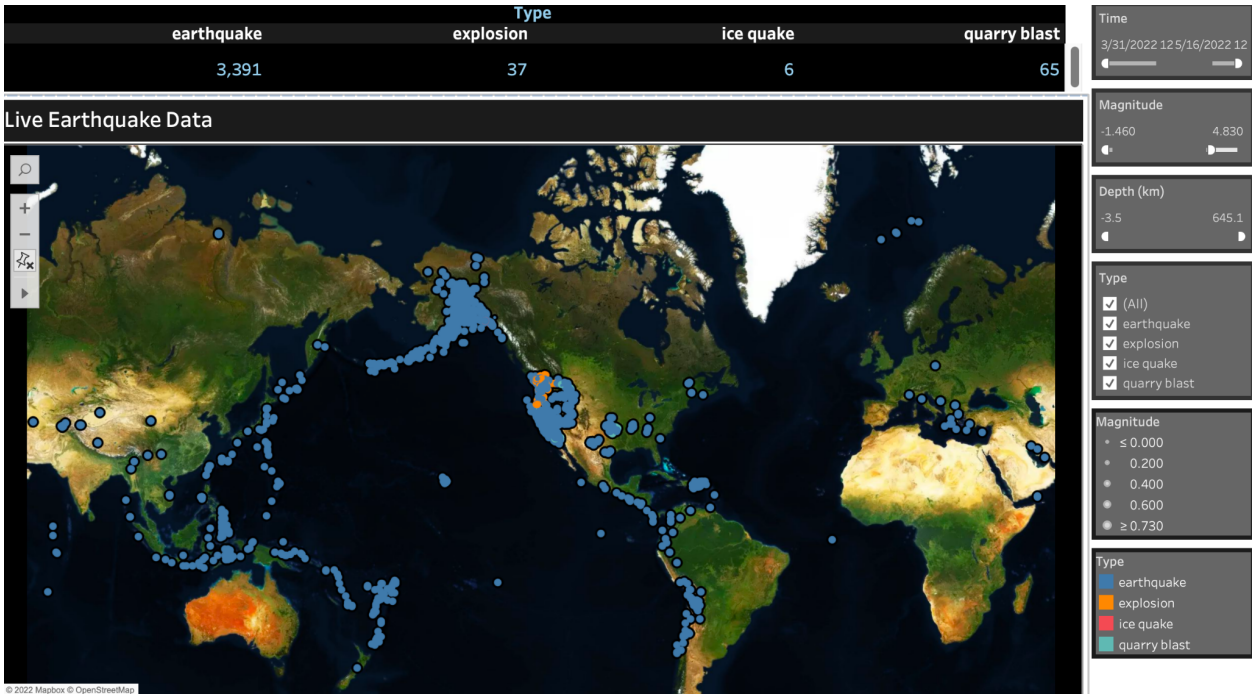


This is a dual combination chart. In this chart we can see the Magnitude Error for the Magnitude of an earthquake. You have an option on the right to filter through the Magnitude Error and Magnitude for the event. Magnitude Error is basically the uncertainty when the Magnitude was recorded. So in the image above you can see that an event which had a magnitude of 2.4 has an average magnitude error of 1.29, which is the uncertainty.

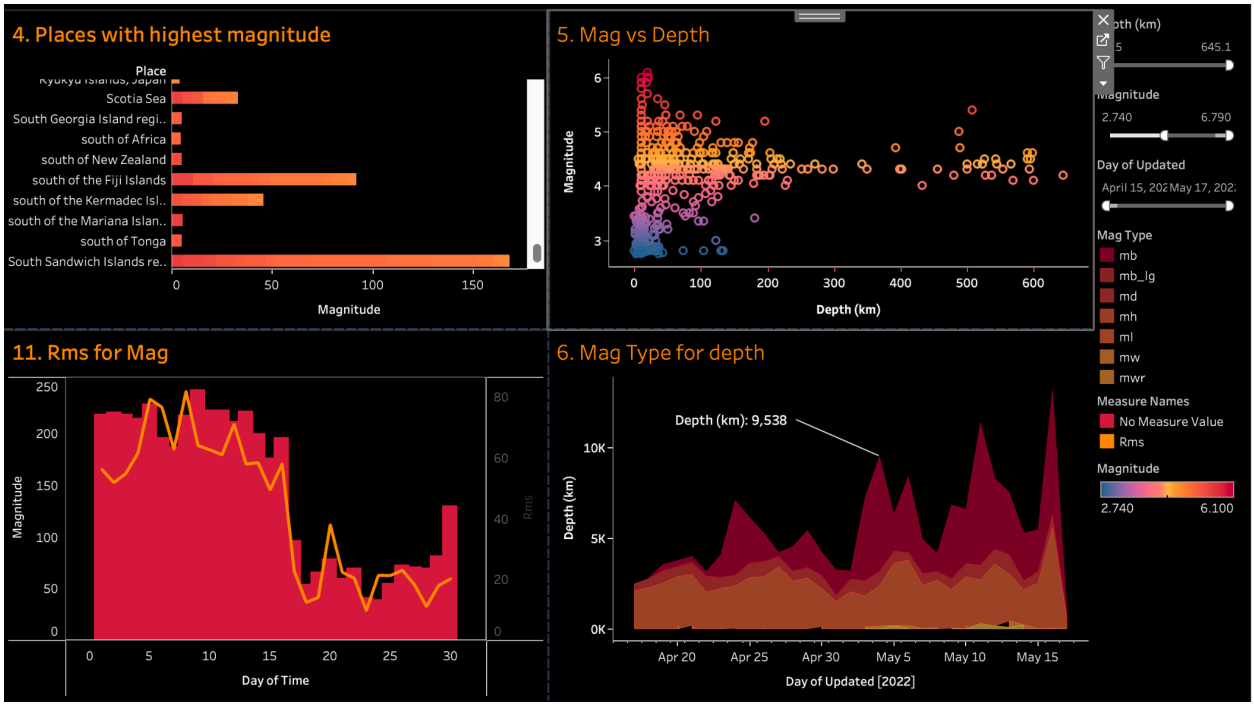


This chart is also a dual combination chart and shows the root mean square of the event's magnitude. On the right hand side you can also filter through the days to see a specific rms for a specific day. Root mean square tells the match of the observed arrival times to the projected arrival times for the site which is measured by this parameter. Smaller the number the better. The rms for magnitude of 1.8 is 0.15 which means it's a good fit.

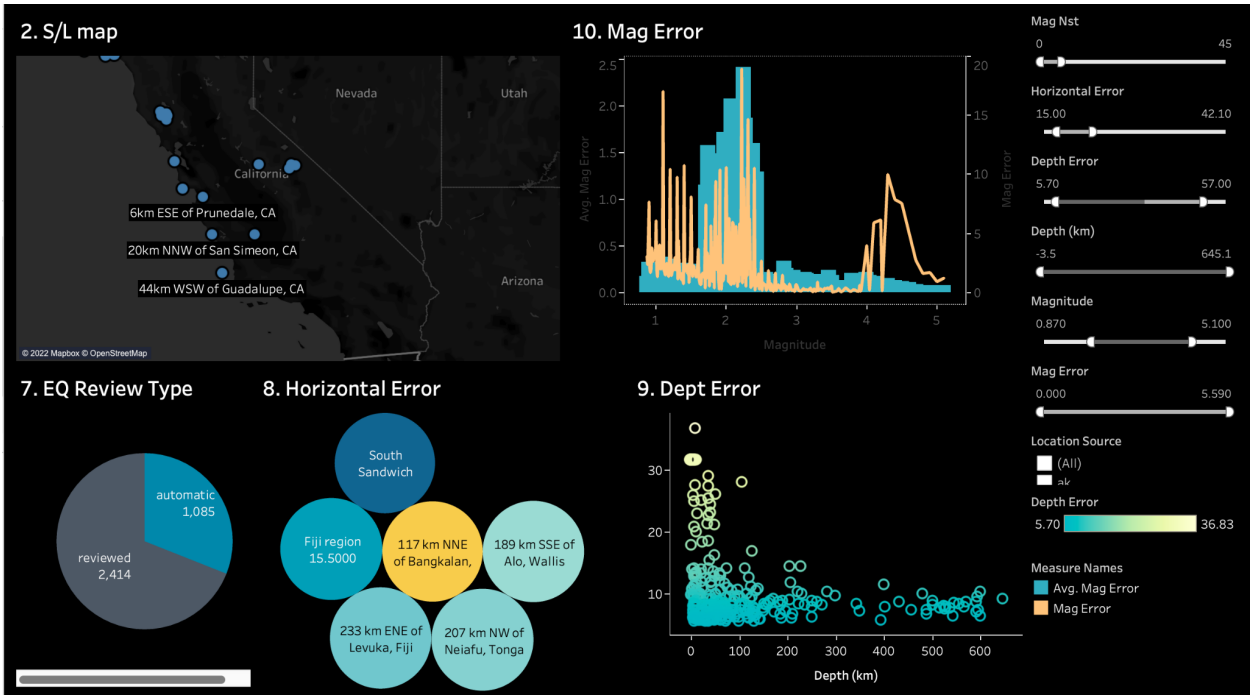
9. Dashboards



This is my first dashboard and it contains the charts for the types of quakes on the top. The map shows the overall information for data analysts to examine different types of quakes and filter through metrics of them. As we can see most of the earthquakes happen in the southwest region of the United States and many of them also happen in Alaska including ice quakes and other types of quakes in other regions. You can also filter through which type of quake you want to see on the map, and which depth you would like to see. Specific days can also be selected and which magnitude of the event you are looking for.



This dashboard shows the overall comparisons for the metrics of the earthquakes. This dashboard contains 4 charts, each chart contains useful information for earthquakes. These charts have been explained above, however, this whole dashboard is interactive, on the right hand side the filters change all the charts. If you change the magnitude and day, all charts will change to showcase the relevant information the user is looking for. The dashboard also has an earthquake type theme and stays consistent for the whole dashboard, this provides the user an easy and understandable interface to work with.



This is my last dashboard and it contains the statistics of the earthquakes. All the charts contain metrics of the quakes, and these measurements can be changed with the filters on the right. The whole dashboard is interactive meaning, when any filter is changed on the right hand side, the whole dashboard will change, users can get the most information out of this dashboard for any type of event they want to examine, and they can see the metrics for it. The theme is blue to match statistics, to give an easy and professional interface to work with.

10. References

https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php

https://www.kaggle.com/code/farazrahman/earthquakes-induced-quakes-know-the-difference/notebook

https://www.usgs.gov/faqs/why-are-we-having-so-many-earthquakes-has-naturally-occurring-earthquake-activity-been#:~:text=The%20National%20Earthquake%20Information%20Center,or%20approximately%2055%20per%20day.