# EarthQuakes VA

ANALYZE AND EXPLORE EARTHQUAKES OCCURRENCES, FREQUENCIES AND BEHAVIOR THROUGH A VISUAL ANALTYCS SYSTEM

# Introduction & Context

Our project purpose is to help our users to better understand the correlation between earthquakes that arises in the last 20 years aiming to provide an easy visualization of the data.

- Earthquakes have always been considered a natural disaster, bringing destruction and devastation in cities and countries.
- An earthquake could cause a very high number of victims, without counting the damages to the buildings.
- The study of the earthquakes is conduced starting from the data analysis and the study of the past earthquakes.
- Of particular importance could be the monitoring of the various earthquakes of a specific country during the years or the comparison of two adjacent countries to see how, also between near territory the earthquake occurrences changes.

# Dataset and preprocessing

The dataset is taken from the USGS, the United States Geological Survey that monitors and reports on earthquakes, their impacts and hazards, and conducts research on the causes and effects of earthquakes.

We selected as object of our analysis only those countries bounded in the Eurasia continent.

- **id**: earthquake identifier
- **latitude**: lat. of earthquake epicenter
- **longitude**: long. of earthquake epicenter
- **magnitude**: earthquake strength level
- **magType**: algorithm used to compute the mag.
- **depth**: distance in km between epicenter and ipocenter
- **gap**: azimuthal gap between stations
- **rms**: root-mean-squared arrival time (sec)
- **place**: country where the earhquake arised
- **type**: event type: ‘’earthquake’’ or ‘’quarry’’
- **status**: automatic or reviewed
- **nst**: number o stations used to compute earthquake location

# Dataset and preprocessing

The dataset is taken from the USGS, the United States Geological Survey that monitors and reports on earthquakes, their impacts and hazards, and conducts research on the causes and effects of earthquakes.

Preprocessing process removed some column of dataset not useful for our purposes.

“Place” column is standardized and adapted to the map countries.

Moreover, PCA algorithm is performed on dataset features in order to calculate two PCA components and store them as two new columns of the dataset:

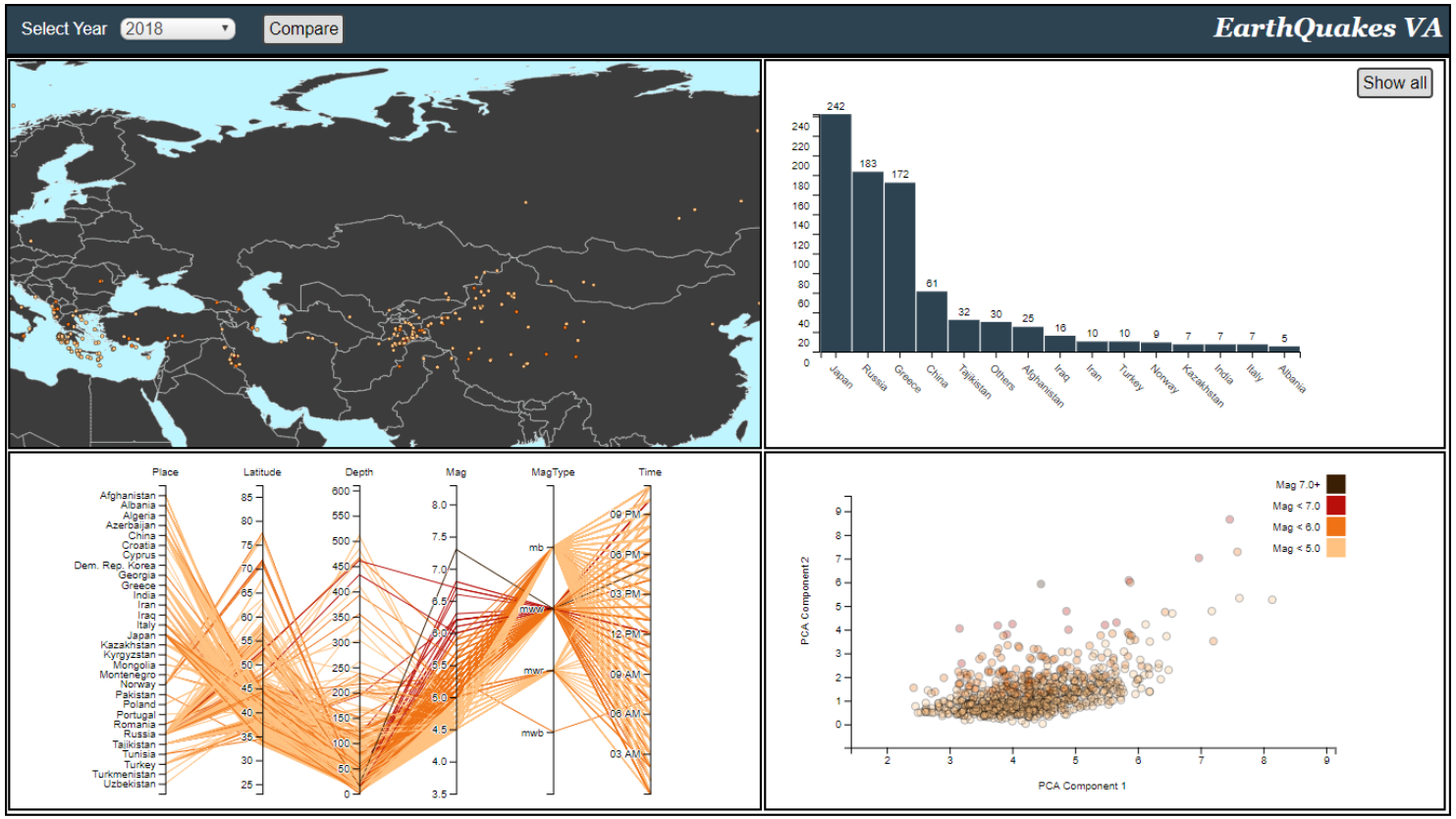
- **PCA_Component1**: the first component computed by PCA algorithm, the one with the highest variance.
- **PCA_Component2**: the second component computed by PCA algorithm, the one with the highest variance that is orthogonal to the first, so it is linearly uncorrelated with it.

# Visualization

The project comes up with a header to filter data by year and it shows through some graphs and a map the earthquake arose for that year in Eurasia.

Starting the project, this overview appears to the users: they can explore the map, read the frequencies of earthquake through the bar, filter the parallel chart and the scatterplot to focus on particular data.

Moreover, if one wants to focus on a specific country, he can select through the map a country (or two if he wants to compare them) and visualize more detailed info of the features of that country.



# Map and parallel coordinates chart

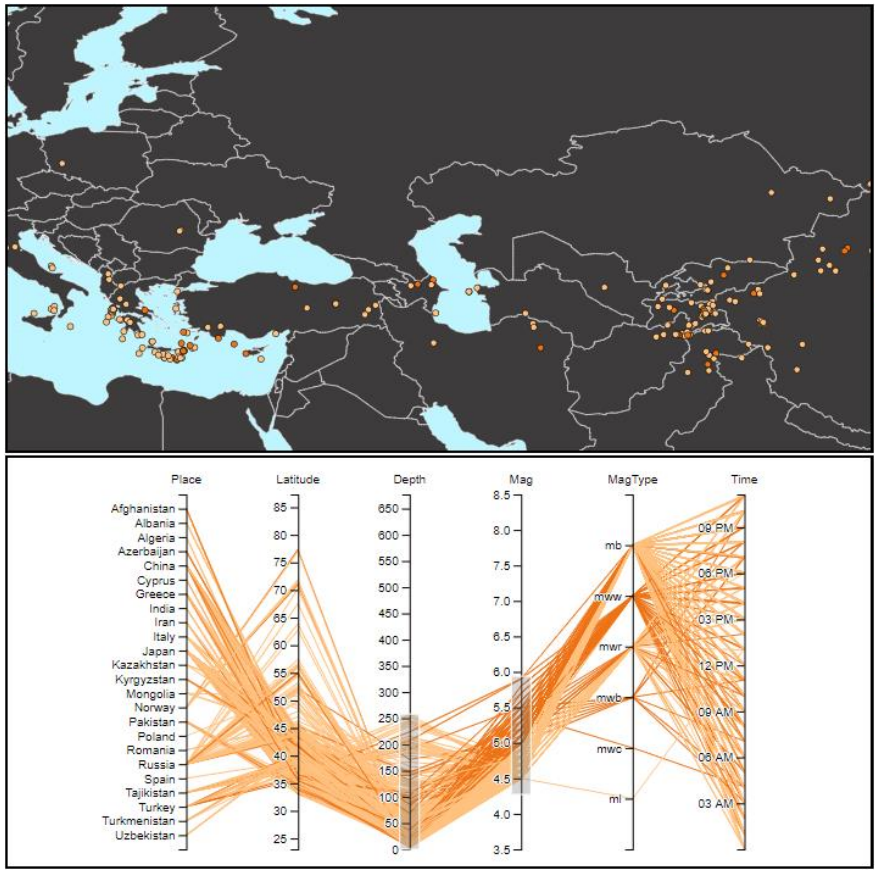
These graphs are two of the main visual elements of the project. They fill both the overview page and the comparison page, and through parallel filtering, the map points are updated.

Parallel allows to user to brush its axes in order to select a range of values of interest.

The map responds to parallel changes updating the points.

This interaction enable an easy filtering of the data from any feature of the dataset.

Also the parallel filtering will influence all the other graphs of the project.



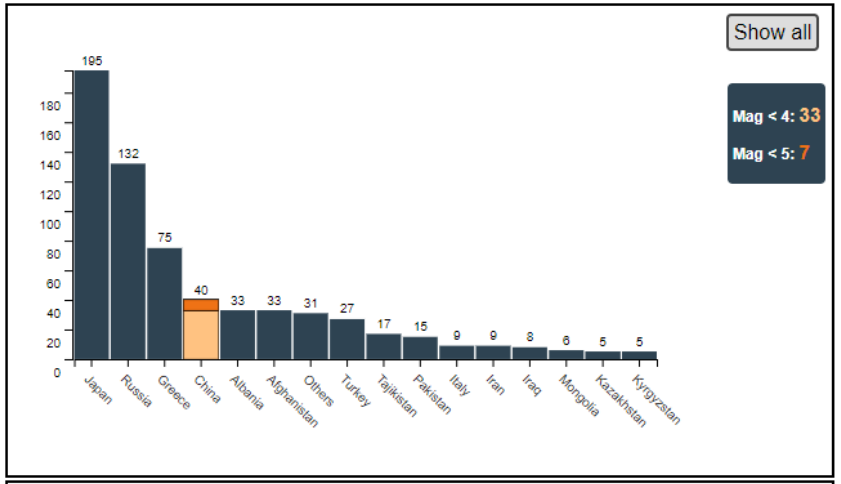
# Bar chart and scatter plot

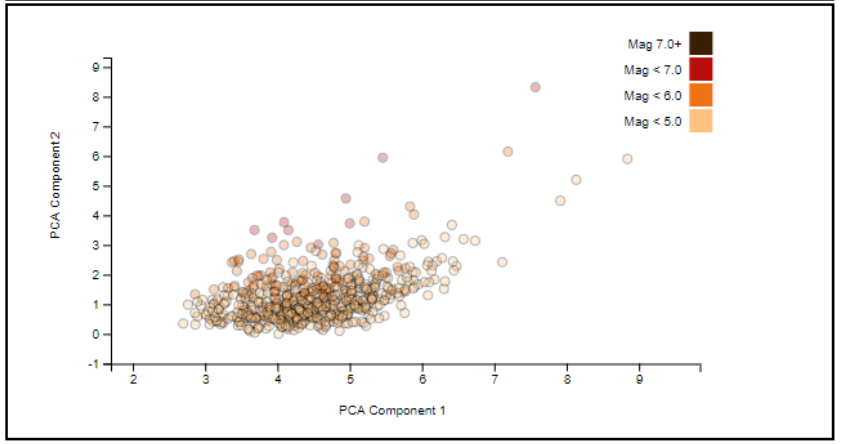
These are the other two important element of the overview and through them the user can again filter and deeper analyze the data.

The barchart shows the frequencies of earthquakes in each country.
To better analyze them, the user can click on a bar to analyze percentages of each magnitude level.

Also countries with few occurrences are grouped together creating “others” labels, and the “show all” button allows to display them.

Scatterplot displays each element of the dataset according to the first two PCA components, helping to point out outliers and clusters.
Brushing a rectangle area on it will point out the selected point both on the map and on the parallel chart, changing the data color.



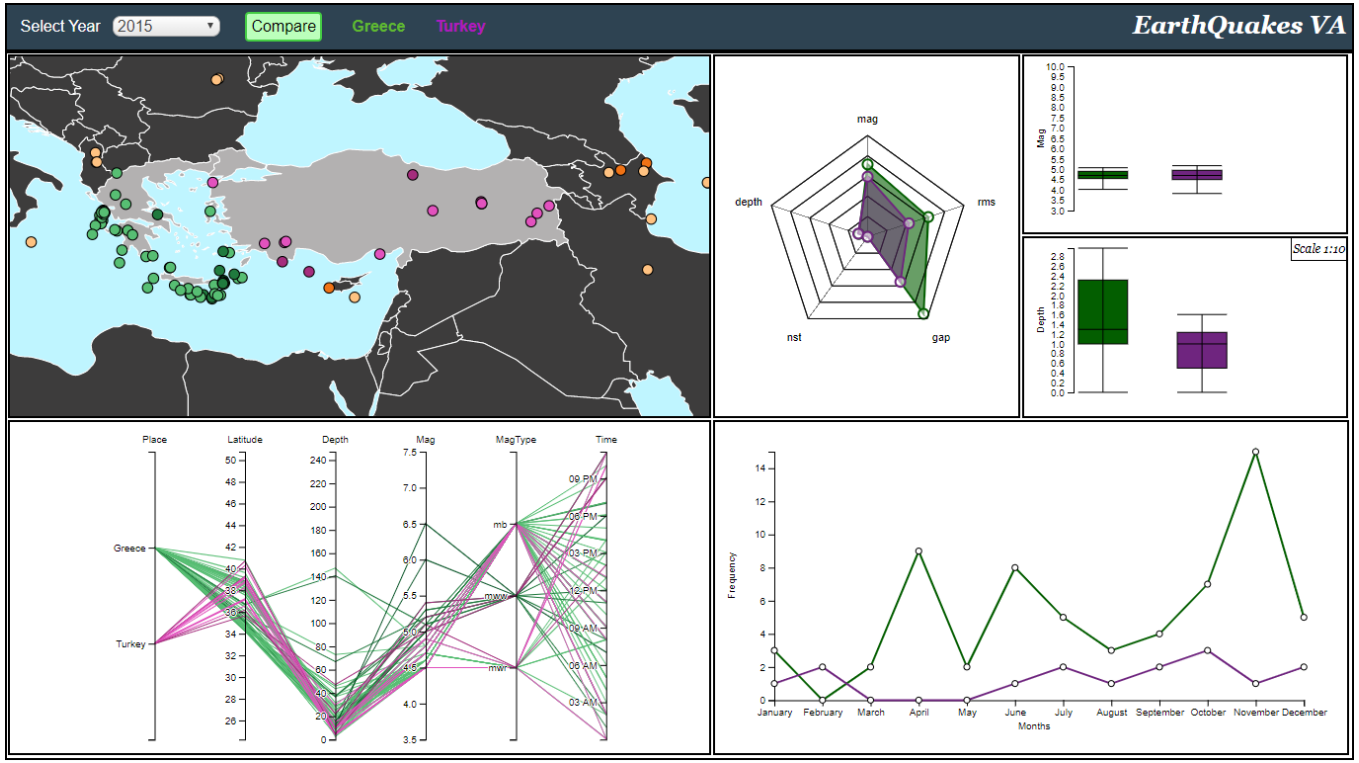



# Country selection and comparison mode

The second aspect of our project allows the user to select a country to focus on its earthquake and analyze it during the years. Moreover, he can choose a secondo country to directly compare both them.

We can see here that clicking on the map on a particular country, it will update parallel data. So in this sense there is a bidirectional interaction between these two graphs.

Parallel chart still works as before. Now, filtering range of data will update just the earthquakes of the selected country.

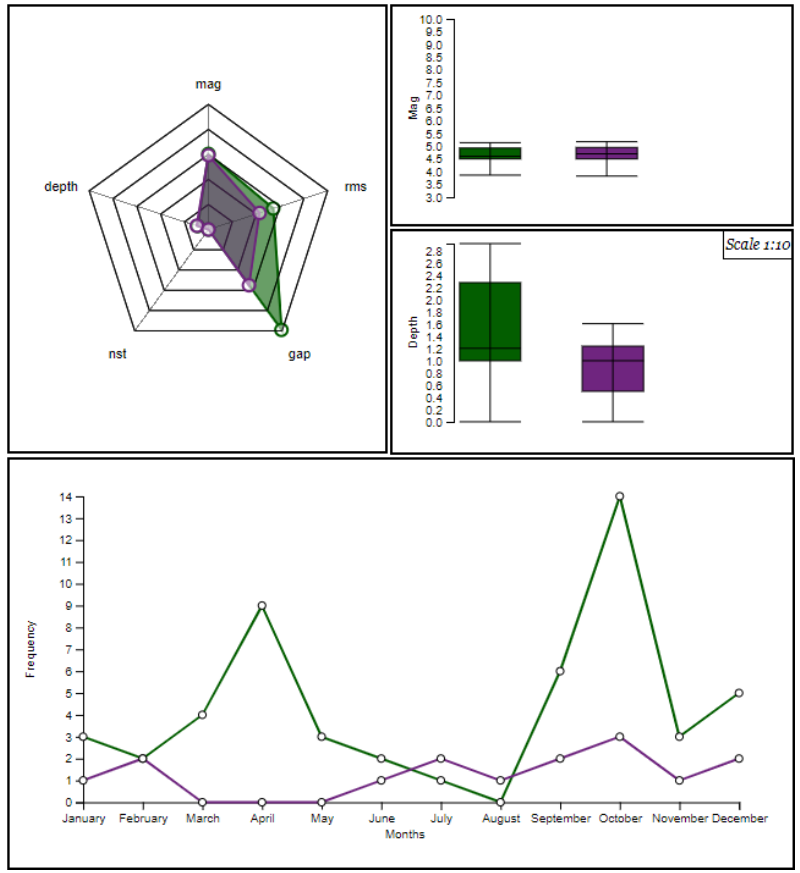


# Radar chart, line and boxplot

The right side of these mode does not directly influence other graphs but is used to show data in different aspects and from the point of view of different features.

The radarchart shows an area for each selected country: each feature is represented by an axe and the more the area cover an axe the closer the computed mean will be to the maximum value of that feature.

The lineplot show the trend of countries' earthquake frequencies during the selected year over each month.



# Analytics and interactions

The project offers many ways to users to directly interact with data letting visual elements interact each other. Also many computation algorithm are used to facilitate earthquake behavior comprehension

- **Header**: let the user to select a year between 2000 and 2020. Also he can check button for the comparison mode.
- **Map**: offers possibility to pan and zoom over the countries of Eurasia. Then, by clicking on one of them it will update data displayed in the page, interacting with all the graphs.
- **Scatterplot**: let the user to brush an area to focus on only part of data. Doing this, the parallel chart and the map will change their data color to point out which are the ones just selected with scatterplot.
- **Parallel**: it offers possibility to filter out data based on different feature of the dataset to focus on specific range of values. Doing this, all other graphs update their data according to the selected range of value. So, what emerges is also a bidirectional interaction between parallel and map and another one with parallel and scatterplot.
- **Barchart**: as already shown, clicking on a bar of the chart, it will show magnitude percentage for that country to better understand the entity of earthquakes.

# Analytics and interactions

The project offers many ways to users to directly interact with data letting visual elements interact each other. Also many computation algorithm are used to facilitate earthquake behavior comprehension

Let's now focus on main computation efforts performed by the project.

- The one, that immediately appears visible to the user, is the computation of the frequencies, both those of countries for years or month, both those of countries for magnitudes and also the ranking of most hit countries to give an order to the barchart.
- To compute the radarchart areas for each country, for each feature is computed a mean based on each earthquake of the filtered dataset.
- There are then some scaling computation to adapt data with the graphs and also the computation of quantiles for the boxplot each time the data changed.
- Finally other computation, less visible, like ranges for axes that are dinamically updated with the data changes.