[Download as PDF](#)

URL to view Results	[[Click Here](#)]

Response Summary:

Acquire Worksheet

Goal: Identify appropriate data sources, analyze the data, identify data types, variables, list assumptions about the data
Objectives: Students will identify and acquire data from appropriate data sources
Outcomes: Data for the current visualization challenge

1. Student Information *

First Name	Samruddhi
Last Name	Tawade
Course (e.g. CGT 270-001)	CGT 27000
Term (e.g. F2019)	F2021

2. Email Address *
stawade@purdue.edu

3. Visualization Assignment *
- Training Data

Generate

4. Identify appropriate data sources: is the data publicly available? What search methods were used? *

Data source 1	Earthquakes in 1910-2017, Turkey @alpko https://www.kaggle.com/caganseval/earthquake?select=earthquake.csv (earthquakes, datasets)
Data source 2	Significant Earthquakes, 1965-2016 https://www.kaggle.com/usgs/earthquake-database (earthquakes, datasets)
	Magnitude 6+ Earthquakes

Data source 3	Magnitude 6+ Earthquakes Tableau

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *

All of the data is structured in the form of a table in a spreadsheet. Because all three data sets are in the form of a list, and come from many different sources, these data would be tertiary.

6. Data types: what types of data are in the data? How are they stored? What is the access to the data (API, JSON, txt, csv, etc.)? What structure holds the data (data base, spreadsheet, etc.)? *

This data is quantitative because it is numerical. It is also nominal because there is no order for how it has been placed in the chart. The data is stored in a chart on a spreadsheet, and the datasets from Kaggle are stored in a .csv format. The data from Tableau is stored on an Excel spreadsheet.

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *

The data variables include Time, Date, Magnitude, Latitude, Longitude, Location, and Depth, among many others. The parameters would be the mean magnitude of all the earthquakes that have occurred, as well as the average depth, time, latitude, and longitude. They are all independent variables.

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *

Some dates are unavailable in the Significant Earthquakes database. My audience would be the general public, and anyone who is interested in researching strong earthquakes.

Generate

9. What real life behavior does the data reflect? Does it show patterns of activity, regularity of events, a timeline, population data, etc? Explain. *

The earthquake datasets show patterns of activity, as well as regularity and frequency. The datasets measure the magnitude, which demonstrates activity, as well as the date and time which depict frequency and regularity.

11. What are the weaknesses of the data source? Is it likely that the source will be available in the future? Is the data complete? What is the quality of the data? Is it specific to your needs for. the current project? Is the data in the format you need? Are there missing data? Explain. *

One of the datasets contains missing data for the dates that the earthquakes occurred on. Some datasets also contain magnitudes that are lower than 6.0, or are specific to only one region instead of the entire world. Other than these

are specific to only one region instead of the entire world. Other than these weaknesses, the datasets seem to have no faults.

12. What information is emphasized? What is the central focus of the data? Explain. *

The data generally emphasizes the magnitude of the earthquake, as well as the location, date, and time. These four categories are usually first in the the chart, which emphasizes their importance,

13. At what level of granularity is the data provided? Is the data summarized, or do you have access to the raw data? Is the data categorized or is the data in a format that allows you to create your own categories, etc. Explain. *

The data is categorized in a chart, but I can still use these categories to create my own. I do not have access to the raw data, and there is no summary of the data, only a list.

14. What is the scope of the data? What topics can be covered using the data? Is there a time range/frame? Is the data for a specific area/discipline/demographic etc.? Explain. *

Topics that pertain to the frequency in relation to the location of the earthquake, tectonic plate activity, and magnitude in relation to location of the earthquake can be covered. This data would be ideal for drawing comparisons between the categories. The datasets have a range from 1910 to 2017, and one of the datasets is specific to Turkey.