

Lab_07

Fantastic Four

A couple of quick thoughts on dataset: * Lahman or Retrosheet baseball stats– some of these are not “tidy” based on experience * Weather data from somewhere– could also find data with “holes” * Google public datasets– lots of them out there * Use an API to snag our own data? (scott could run this) * Go check out catalog.data.gov * scott’s favorite so far: <https://www.kaggle.com/noaa/seismic-waves/data> * airline fleet data, might need some tidying <https://www.kaggle.com/traceyvanp/airlinefleet/data>

“Who is Hiring us?”

Fantastic Four is being hired by the Pacific Tsunami Museum in Hilo, Hawaii to look at tsunami data. There are two reasons that this museum is hiring the team. First, they would like to map out all the tsunamis in the Pacific in a creative way using the data we have. This could be economic impact of more recent tsunamis (since that part of the data is available) or the relationship between magnitude and source of the waves. Second, they would like to know if there is a better way to advise or educate the people of Hawaii about future tsunamis.

Questions and Importance

There are two questions we have set out to answer. The first is:

What are the patterns in tsunamis that come to Hawaii?

The reason we are looking at this question is to find the patterns of tsunamis that come to Hawaii. This is important because it can help us better prepare the community for impact (via insurance, warning, education) and furthermore minimize the damage to our beloved Hawaiian Islands.

The second question to look at is:

What are the patterns in tsunamis in the Pacific and how to they relate to ones recorded in Hawaii?

This question is important because by being able to see relationships between waves in Hawaii and other places in the Pacific, we can put together our resources in order to better prepare and educate society in order to minimize impact.

“A breif description of the dataset you wish to analyze next week”

The dataset Fantastic Four will be using will be a tsunami dataset from

<https://www.kaggle.com/noaa/seismic-waves/data>

This is an NOAA dataset of tsunami data for recordable history. A preview of the column headers of the raw data is the following:

```
## # A tibble: 2,582 x 45
## # ... with 2,582 more rows, and 45
## #   variables: SOURCE_ID <chr>,
## #   YEAR <chr>, MONTH <chr>, DAY <chr>,
## #   HOUR <chr>, MINUTE <chr>,
## #   CAUSE <chr>, VALIDITY <chr>,
## #   FOCAL_DEPTH <chr>,
## #   PRIMARY_MAGNITUDE <chr>,
## #   REGION_CODE <chr>, COUNTRY <chr>,
## #   `STATE/PROVINCE` <chr>,
## #   LOCATION <chr>, LATITUDE <chr>,
## #   LONGITUDE <chr>,
## #   MAXIMUM_HEIGHT <chr>,
## #   MAGNITUDE_ABE <chr>,
## #   MAGNITUDE_IIDA <chr>,
## #   INTENSITY_SOLOVIEV <chr>,
## #   WARNING_STATUS <chr>,
## #   MISSING <chr>,
## #   MISSING_ESTIMATE <chr>,
## #   INJURIES <chr>,
## #   INJURY_ESTIMATE <chr>,
## #   FATALITIES <chr>,
## #   FATALITY_ESTIMATE <chr>,
## #   DAMAGE_MILLIONS_DOLLARS <chr>,
## #   DAMAGE_ESTIMATE <chr>,
## #   HOUSES_DAMAGED <chr>,
## #   HOUSE_DAMAGE_ESTIMATE <chr>,
## #   HOUSES_DESTROYED <chr>,
## #   HOUSE_DESTRUCTION_ESTIMATE <chr>,
## #   ALL_MISSING <chr>,
## #   MISSING_TOTAL <chr>,
## #   ALL_INJURIES <chr>,
## #   INJURY_TOTAL <chr>,
## #   ALL_FATALITIES <chr>,
## #   FATALITY_TOTAL <chr>,
## #   ALL_DAMAGE_MILLIONS <chr>,
## #   DAMAGE_TOTAL <chr>,
## #   ALL_HOUSES_DAMAGED <chr>,
## #   HOUSE_DAMAGE_TOTAL <chr>,
## #   ALL_HOUSES_DESTROYED <chr>,
## #   HOUSE_DESTRUCTION_TOTAL <chr>
```

This data contains date, magnitude, location, depth, time, and damage attributes for tsunamis dating back to 2000 BC. This data set needs tidying for the following reasons:

- * The region we are looking at is the Pacific

- * There are a LOT of values are N/A and we need to evaluate/get rid of tsunamis that do not give significant data.

- * For economic impacts, it would be important to look at more recent data, which would be more realistic

- * Simplify columns (total damage, date/time, etc.)

Subsequent Questions and Importance:

Contributions

- Lindsay:
- Lexie:
- Li:
- Scott: