Project: Deaths by Natural Disaster

Collaborators: Bonnie-jo Barnaby & Charles Lindner

Sources:

- World Health Organization Global Health Observatory
  - http://apps.who.int/gho/data/node.sdg.13-1-data?lang=en
    - WHO_Disaster_Data.xlsx
- International Red Cross and Red Crescent Federation
  - http://data.ifrc.org/fdrs/data-download
    - RC_Disaster_Data.csv
    - RC_Disaster_Codes.xlsx

Project Report

E (xtract): Original data sources and how the data was formatted

Sources:

- World Health Organization Global Health Observatory
  - http://apps.who.int/gho/data/node.sdg.13-1-data?lang=en
    - WHO_Disaster_Data.xlsx
- International Red Cross and Red Crescent Federation
  - http://data.ifrc.org/fdrs/data-download
    - RC_Disaster_Data.csv
    - RC_Disaster_Codes.xlsx
- Validated both reported mortality rates from the WHO and Red Cross were based on 1:100,000 people for normalization of data.

T (ransform): What data cleaning or transformation was required

Cleanup process for REDCC

- Created dataframe from CSV file
  - Code to show list of headers from full file
  - Selected only the columns we needed for analysis
    - Country
    - Years – 2012 to 2016
    - GDP
    - LifeExp
    - Population
    - Child Mortality
    - Maternal Mortality
- Filtered out data to only include years that match WHO data (2012-2016)
- Grouped by Country and determined the mean for the following data columns

    - GDP
    - LifeExp
    - Population
    - Child Mortality
    - Maternal Mortality
- Added a new column “Year” which represents the averages of 2012-2016
- Exported as CSV file

Cleanup process for WHO

- Determine sheets from excel file we needed to pull data from (data-text and Country sheets)
- Created two dataframes from excel file (data-text and Country)
    - List of headers for each dataframe
    - Selected columns we wanted
        - From data-text
            - Years – 2012 to 2016
            - Country
            - Gender
        - From Country
            - Country
            - Mortality
            - Land Area
            - Region
    - Renamed column headers
- Merged both dataframes on “Country” name column
- Filtered column “Gender” to get combined morality rate for both sexes by Country
- Exported as CSV file

L (oad): The final database, tables/collections and why this was chosen

- Loaded final production into relational database (Postgres) titled “Mortality”.
    - Selected relational database because data already existed in relational database from sources selected.
- Final tables or collections used in production database
    - Red Clean Table
        - Imported from red_clean.csv
    - WHO Clean Table
        - Imported from red_clean.csv
- Created SQL statement to provide following output:
    - Join the tables on Country column
    - Reorder columns for convenient viewing
    - Included monetary output of GDP column and formatted per billions of dollars
    - Included decimal rounding for mortality columns
    - Excluded null values for mortality column