Katherine Angell
Springboard: Introduction to R
July 8, 2019

Capstone Project: Data Cleaning Description

- Seperated the date column into day, month, and year columns and removed the duplicate year column that already existed
- Removed the comma that remained in the day column after separation
- Renamed all of the columns I'll be focusing on for analysis because they were all in a format with spaces instead of "_" in between words
- Tried to replace NA values in columns where it made sense to try to impute--for the total number of survivors and the total number dead from a given incident I imputed the mean so that the column can be used for analysis
- Created new column of total migrants involved in the incident by adding total reported dead and missing with the total number of survivors
- Rounded the entire dataframe to 2 digits because the mean values that replaced NA values had 6-7 decimals