Day Four: Data Fundamentals and Intro to RStudio Environment

SDS 192: Introduction to Data Science

Lindsay Poirier Statistical & Data Sciences, Smith College

Spring 2022

```
global_landslide <- read.csv("https://data.nasa.gov/api/views/dd9e-wu2v/rows.csv")
```

1. Identify a unique key in this dataset. Check whether this unique key repeats.

```
# Check whether the unique key you've identified repeats
```

2. Calculate the total fatality count in this dataset and total injury count in this dataset. Calculate the percentage of NA entries in each of these variables.

```
# Calculate the total fatality count and total injury count

# Calculate the % NA values
```

3. Uncomment and complete the code below to generate a new column with a newspaper headline for each row in the dataset. Your headline should include at least five variables from the dataset, concatenated with narrative text.

```
# Generate a new column with a newspaper headline for each landslide

#global_landslide$headline <- paste()
```

4. Check the possible values in `landslide_size`. Factor this variable, setting the levels from smallest to largest. Table the unique values in `landslide_size` and `landslide_size-factored`.

```
# Check the possible values in landslide_size

# Uncomment below and factor landslide_size
#global_landslide$landslide_size_factored <-

# Compare the outputs when you run the table() function with `landslide_size` vs. with global_landslide
```