GROUP 3
TOPIC: SHARK ATTACKS

Project Proposal
November 3, 2022

1. Project dataset and Inspiration

The topic chosen for our exploratory data analysis is based on shark attacks. The dataset we will be mainly using (found on Kaggle: https://www.kaggle.com/datasets/mysarahmadbhat/shark-attacks?select=attacks.csv) documents globally from 1900 to 2018 reported shark attacks. These reported incidents include categories such as the activity the attacked was doing, the location they were attacked, the date & time, and if the attack was fatal.

We choose this specific dataset file because it has a good assortment of data per reported incident. For the entirety of the collected data there are over 6,000 incidents documented and when narrowed down to what we will use for most of our analysis, we still will have around 1,000. The idea to analyze data on shark attacks is unique and interesting to our team, and ultimately, we would like to know which areas of the ocean we should avoid.

As expected for these globally reported incidents, the data contains many null values and is inconsistent in the format for most of the categories. One other user who also used this dataset replaced the data with "unidentified" instead of dropping the data. However, it still may could make sense to note that we will be excluding some of the reports based on lack of information for our specific analysis. Another study, completed by Kyra Gray, can be reviewed http://rstudio-pubs-static.s3.amazonaws.com/235161_31605f684ff2479890c1eac7d927cec8.html on this specific data set. In addition to the EDA completed, the report gives more insight to the actual data collection resources surrounding this international shark attack file.

2. Research Questions and Prediction

The questions we will be researching for our analysis are the following:

- Where do reported shark attacks happen most in the world?
- What reported activities are being done during a shark attack?
- Was the attack reported fatal?
- What reported month of year do shark attacks occur most often?

3. Visualizations

1. **Bar** – Most common species of sharks reported
2. **Line** – Time of year shark attacks occur and/or time of day
3. **Donut** – Reported Activities – maybe do two (2) pie charts, one for marked fatal and non-fatal by activity
4. **Map** – Global Locations of Shark attacks for top 7 most reported and another map for just the US locations
5. **Violin** - *possibly*

4. Color Palette

5. Roles and Responsibilities

Name	Role	Responsibilities
Brenna Wallace	Lead Presentation & Group	Research Question, Visuals, Slides, Write up, Collate Presentation
Omar Zaher	Lead Data Cleaning & Visuals	Research Question, Visuals, Slides, Write up, Format visuals
Jennifer Wilkerson	Lead Regression & Write up	Research Question, Visuals, Slides, Write up, Collate Write up