Shark Attack Analysis

AJ, Chase, Chloe

Data Set

- General information about our data set
- Information about the columns
- Link

EDA

Read in Data and Clean

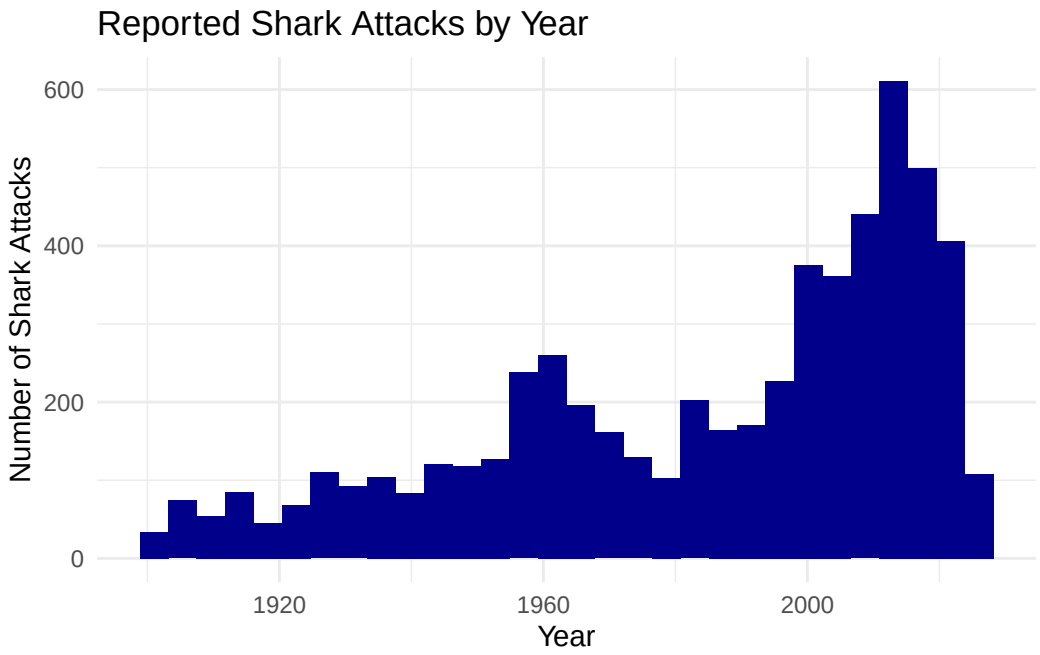
```
sharks <- read_excel("sharks.xlsx") %>%
  select(1:14) %>%
  rename(Fatal = 'Fatal Y/N') %>%
  mutate(Year = as.numeric(Year),
         Sex = case_when(
           str_detect(tolower(Sex), "m") ~ "Male",
           str_detect(tolower(Sex), "f") ~ "Female",
           TRUE ~ NA_character_),
         Fatal = case_when(
           str_detect(tolower(Fatal), "n") ~ "No",
           str_detect(tolower(Fatal), "y") ~ "Yes",
           TRUE ~ NA_character_)) %>%
  filter(Year > 1900)
```

Here we read in the shark attack data and clean the variables of greatest interest. We make sure all upper and lower case versions for "Male" and "Female" are counted for "Sex" and all cases of "Yes" and "No" are counted for "Fatal". We also ensure that year is counted numerically and filter for only those shark attacks in the last 100 years because the data before that period is not well kept.

Shark Attacks Each Year

```
sharks %>%
  filter(!is.na(Sex)) %>%
  ggplot(aes(x = Year)) +
  geom_histogram(fill = "darkblue") +
  theme_minimal() +
  labs(title = "Reported Shark Attacks by Year",
       x = "Year",
       y = "Number of Shark Attacks")
```

```
`stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```

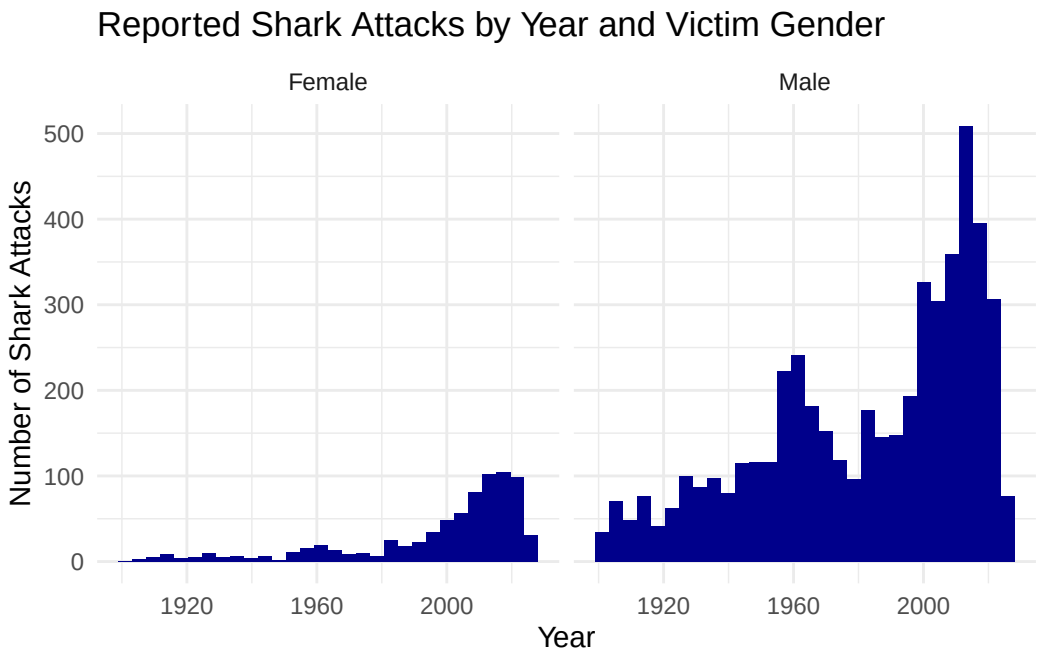


First of all, we want to look at the general distributions of shark attacks over the last 100 years. We can see that the number of reported shark attacks has risen greatly over this period of time. It appears that the growth has accelerated in the last 10-20 years. We wonder if this truly reflects an increase in the number of attacks or if it has more to do with the increase in access to technology for reporting.

Shark Attacks by Gender

```r
sharks %>%
  filter(!is.na(Sex)) %>%
  ggplot(aes(x = Year)) +
  geom_histogram(fill = "darkblue") +
  facet_wrap(~Sex) +
  theme_minimal() +
  labs(title = "Reported Shark Attacks by Year and Victim Gender",
       x = "Year",
       y = "Number of Shark Attacks")
```

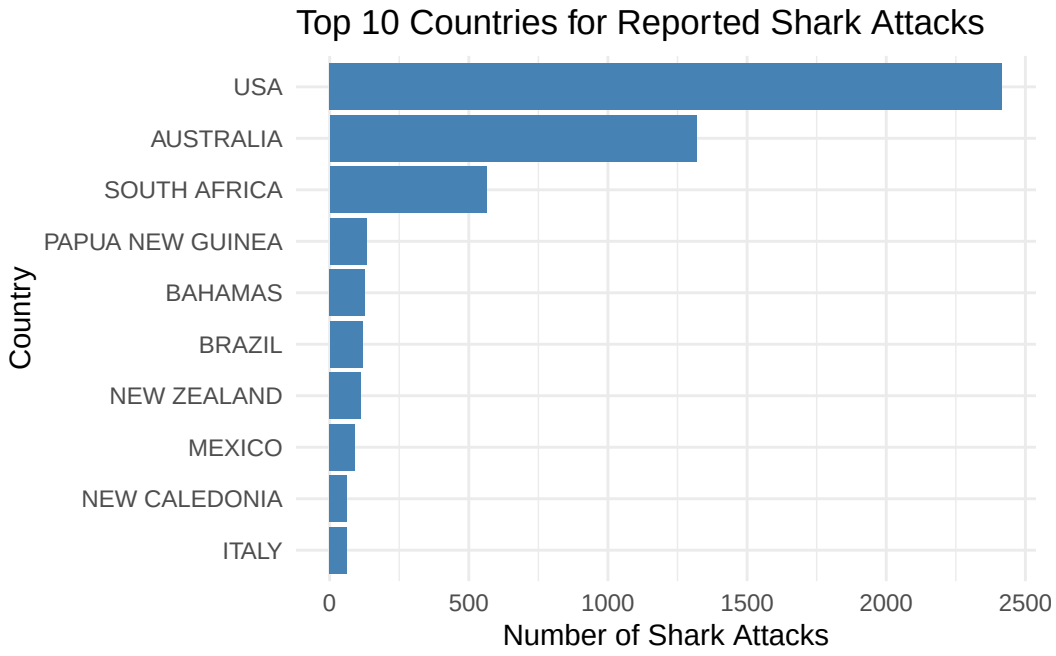
```
`stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```



We also wanted to take a look at the general trend for the number of reported shark attacks by year, split up among the genders. Although they follow a somewhat similar trend, the number of male victims is magnitudes larger than the number of female victims each year. Even though the number victims by gender differ greatly, we can still compare the proportions for these groups.

Shark Attacks by Country

```r
sharks %>%
  group_by(Country) %>%
  count() %>%
  arrange(desc(n)) %>%
  head(10) %>%
  ggplot(aes(x = n, y = reorder(Country, n))) +
  geom_col(fill = "steelblue") +
  labs(title = "Top 10 Countries for Reported Shark Attacks",
       x = "Number of Shark Attacks",
       y = "Country") +
  theme_minimal()
```

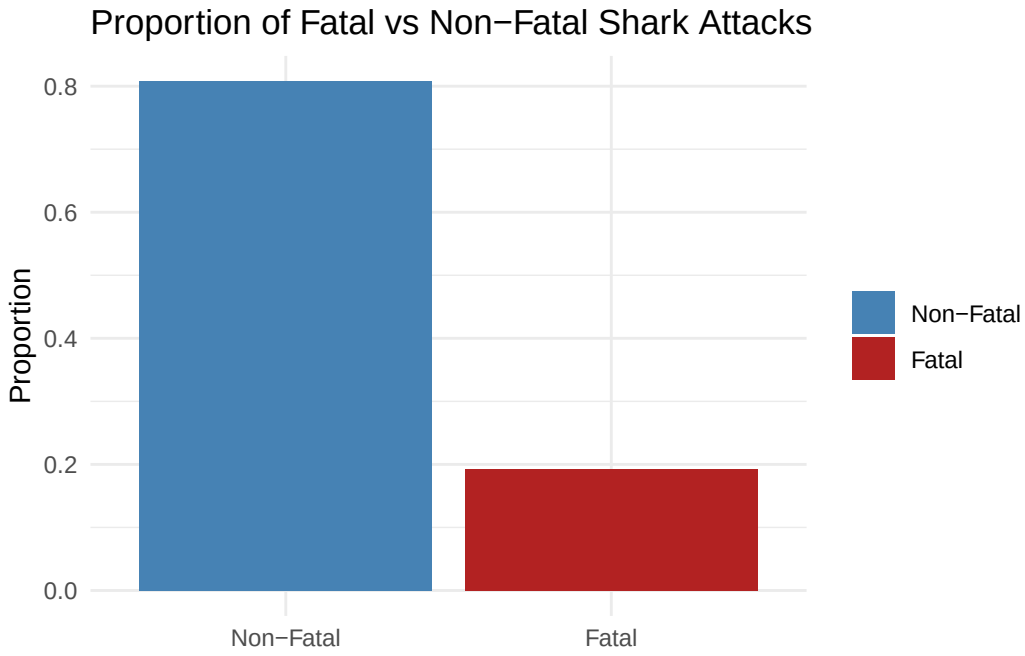


We also looked at the number of shark attacks per country. We focused on the nations with the largest number of reported shark attacks, noting that by far, "USA", "Australia" and "South Africa" lead the world. This is good to note as a potential variable of interest. It also gives us some sort of an idea where our observations are coming from.

Fatality of Shark Attacks

```r
sharks %>%
  filter(!is.na(Fatal)) %>%
  count(Fatal) %>%
```

```
  mutate(prop = n / sum(n)) %>%
  ggplot(aes(x = Fatal, y = prop, fill = Fatal)) +
  geom_col() +
  scale_x_discrete(labels = c("Yes" = "Fatal", "No" = "Non-Fatal")) +
  scale_fill_manual(values = c("Yes" = 'firebrick', "No" = "steelblue"),
                    labels = c("Yes" = "Fatal", "No" = "Non-Fatal")) +
  labs(title = "Proportion of Fatal vs Non-Fatal Shark Attacks",
       x = NULL,
       y = "Proportion",
       fill = NULL) +
  theme_minimal()
```

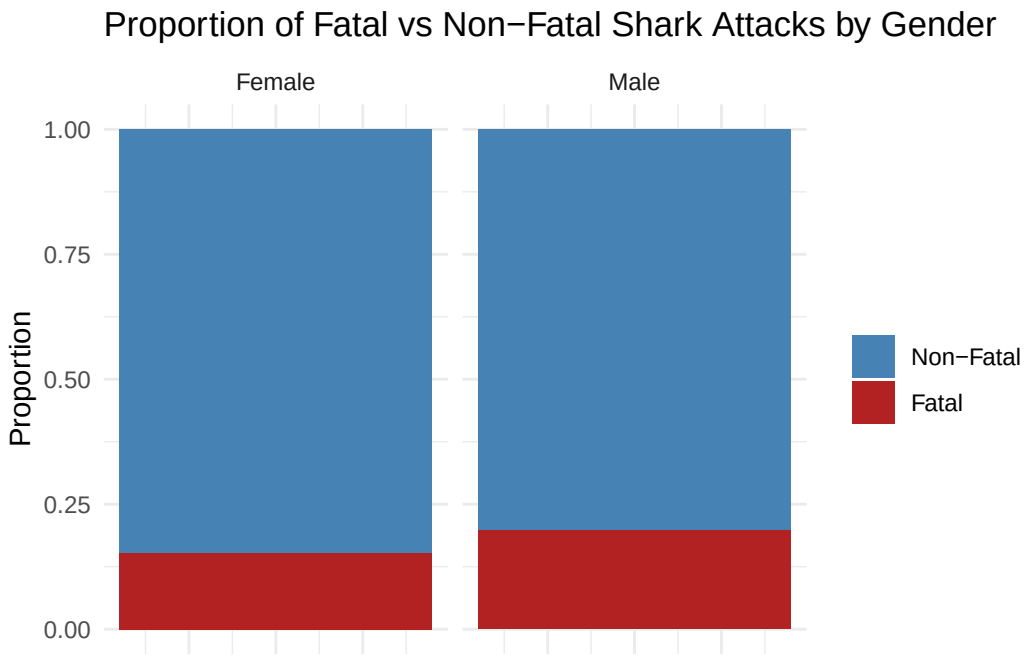


Now we take a look at our variable of interest, fatality. By summing over all of the reported attacks in the last century we can get a rough estimate of the proportion of shark attacks that result in the death of the victim. A little less than 20% of the shark attacks we have access to have been fatal.

Fatality by Gender

```
sharks %>%
  filter(!is.na(Fatal)) %>%
  filter(!is.na(Sex)) %>%
```

```r
group_by(Sex) %>%
count(Fatal) %>%
mutate(prop = n / sum(n)) %>%
ggplot(aes(x = 1, y = prop, fill = Fatal)) +
geom_col() +
facet_wrap(~Sex) +
scale_fill_manual(values = c("Yes" = 'firebrick', "No" = "steelblue"),
                  labels = c("Yes" = "Fatal", "No" = "Non-Fatal")) +
labs(title = "Proportion of Fatal vs Non-Fatal Shark Attacks by Gender",
     x = NULL,
     y = "Proportion",
     fill = NULL) +
theme_minimal() +
theme(axis.text.x = element_blank(),
      axis.ticks.x = element_blank())
```



Ultimately, our goal is to understand if Gender has an effect on the fatality of shark attacks. This graph allows us to quickly get an idea of what those proportions look like. It appears that of the reported attacks, male victims have a higher mortality rate than female victims. Once we preform our full analysis we will be better able to tell if this is a significant difference, and potentially begin to pontificate on why this is the case.

Research Question

- Your proposed research question
- why you chose that research question
- and how the data will permit you to answer it.