251Project

AJ, Chase, Chloe

Shark Analysis

Hello World

Read in Data and Clean

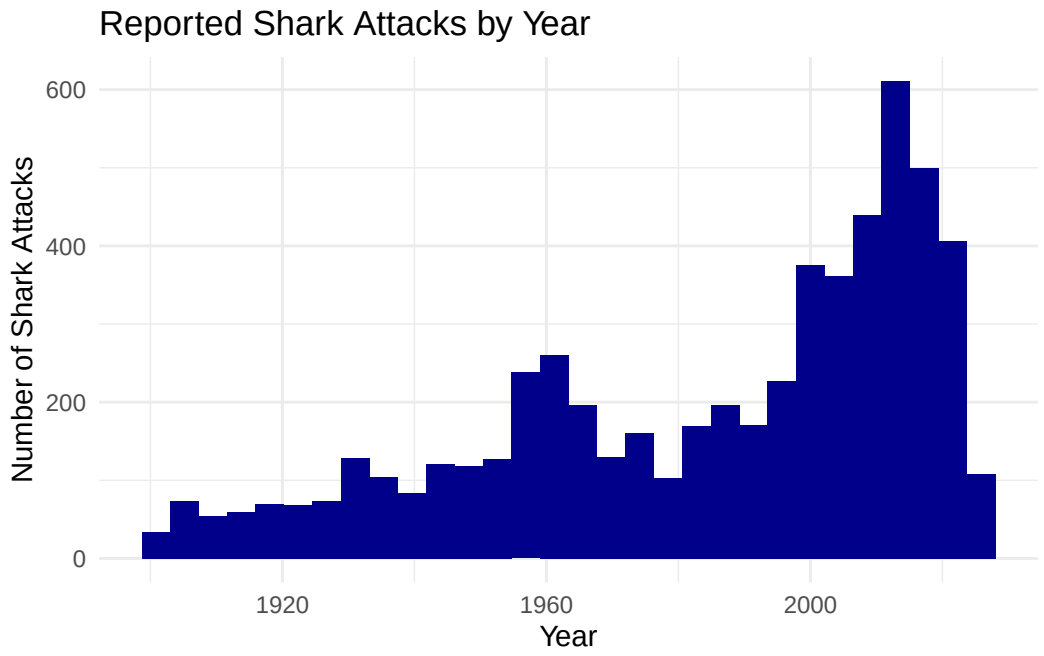
```
sharks <- read_excel("sharks.xlsx") %>%
  select(1:14) %>%
  rename(Fatal = 'Fatal Y/N') %>%
  mutate(Year = as.numeric(Year),
         Sex = case_when(
           str_detect(tolower(Sex), "m") ~ "Male",
           str_detect(tolower(Sex), "f") ~ "Female",
           TRUE ~ NA_character_),
     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

EDA

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

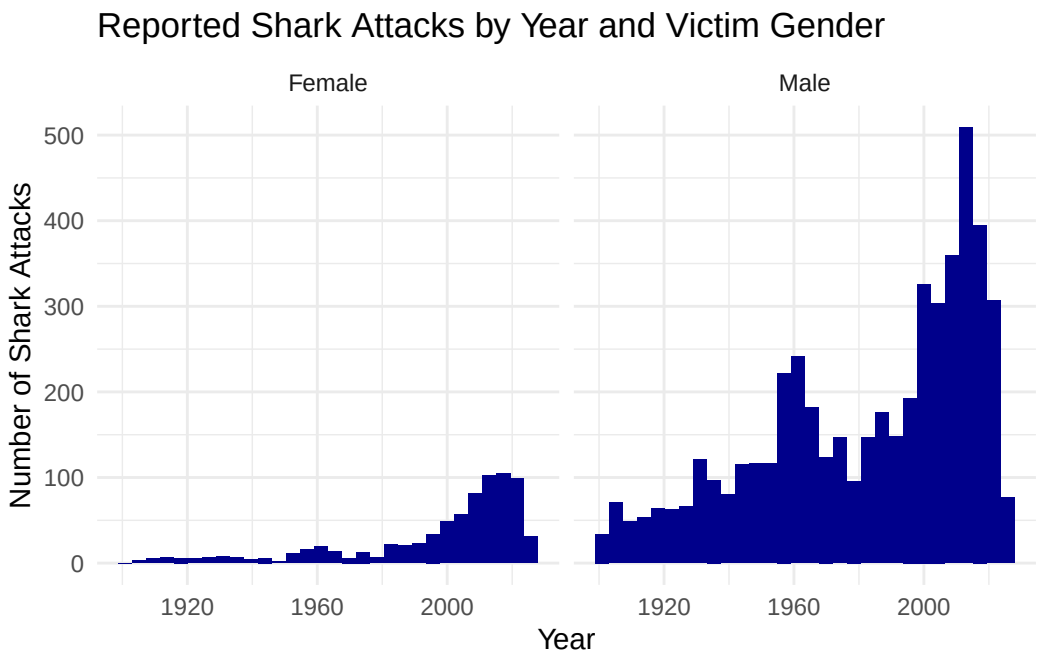
```
       x = "Year",
       y = "Number of Shark Attacks")
```

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

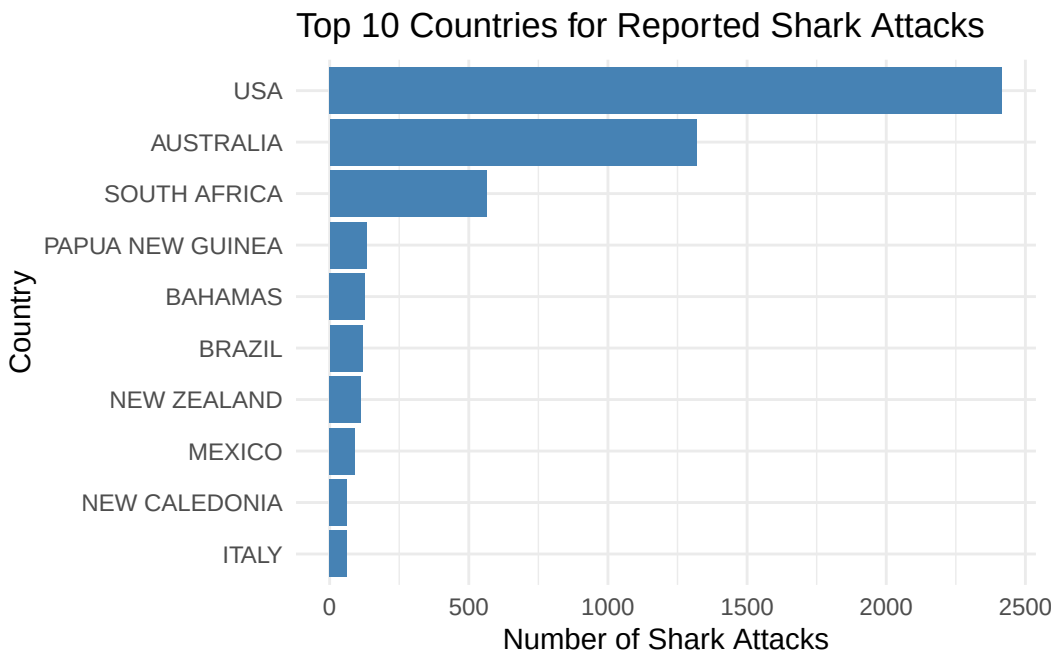



```
sharks %>%
  filter(!is.na(Sex)) %>%
  ggplot(aes(x = Year)) +
  geom_histogram(fill = "darkblue") +
  facet_wrap(~Sex) +
  theme_minimal() +
  labs(title = "Reported Shark Attacks by Year and Victim Gender",
       x = "Year",
       y = "Number of Shark Attacks")
```

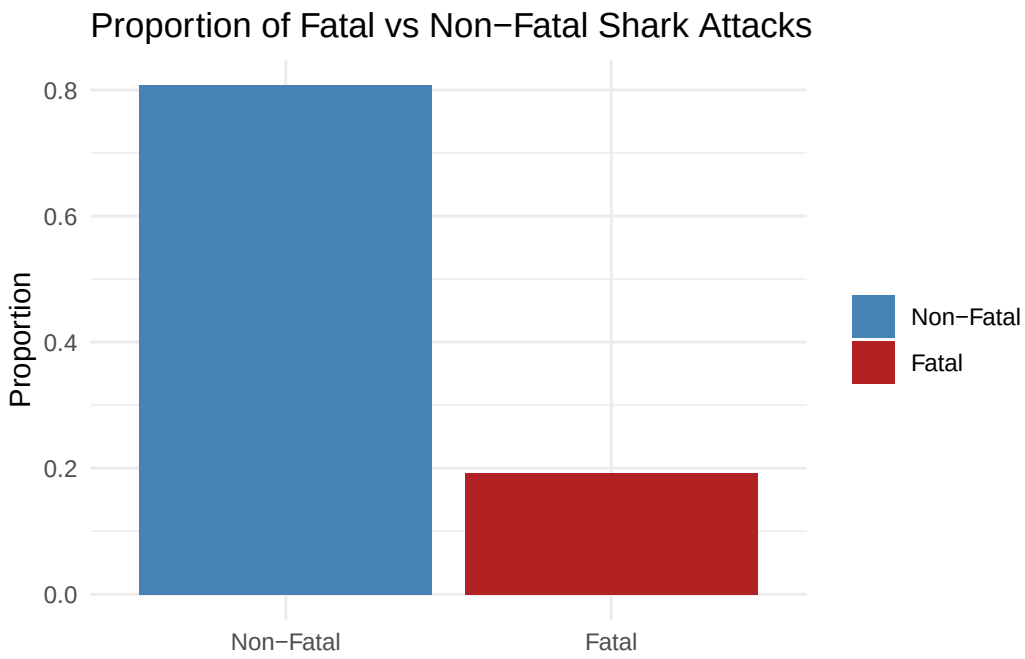
```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



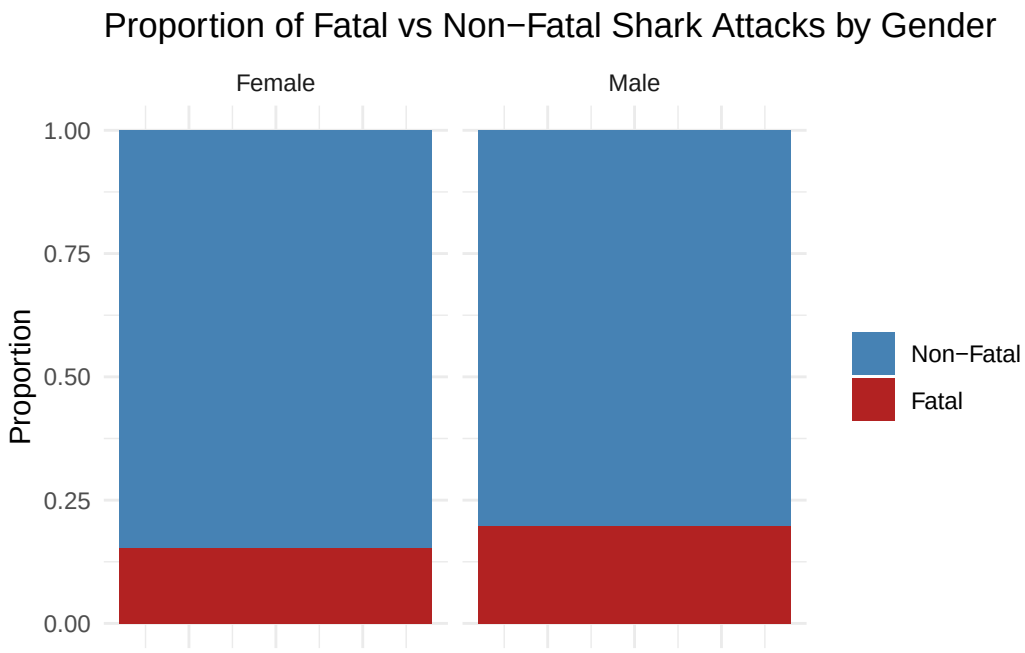

```
sharks %>%
  group_by(Country) %>%
  count() %>%
  arrange(desc(n)) %>%
  head(10) %>%
  ggplot(aes(x = n, y = reorder(Country, n))) +
  geom_col(fill = "steelblue") +
  labs(title = "Top 10 Countries for Reported Shark Attacks",
       x = "Number of Shark Attacks",
       y = "Country") +
  theme_minimal()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




```
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()
```




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


Proportion of Fatal vs Non−Fatal Shark Attacks by Gender
Female
Male
1.00
0.75
0.50
0.25
0.00
Proportion
Non−Fatal
Fatal