Titanic

Carolina Montoya Bran

2023-10-22

Titanic History

The Titanic was a British luxury passenger liner that tragically sank on its maiden voyage in 1912 after colliding with an iceberg. The disaster resulted in the loss of over 1,500 lives, making it one of the deadliest maritime disasters in history. The Titanic's story has captivated the world and continues to be a subject of fascination.

Survivors' demographics played a significant role in the Titanic's history. Let's explore some intriguing information about survivors categorized by age, class, and sex:

Install and load the Titanic dataset

```
install.packages("datasets")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
## Warning: package 'datasets' is a base package, and should not be updated
```

```
library(datasets)
data(Titanic)
```

Data exploration

```
library(knitr)
kable(Titanic)
```

Class	Sex	Age	Survived	Freq
1st	Male	Child	No	0
2nd	Male	Child	No	0
3rd	Male	Child	No	35
Crew	Male	Child	No	0
1st	Female	Child	No	0
2nd	Female	Child	No	0
3rd	Female	Child	No	17
Crew	Female	Child	No	0
1st	Male	Adult	No	118
2nd	Male	Adult	No	154
3rd	Male	Adult	No	387
Crew	Male	Adult	No	670
1st	Female	Adult	No	4
2nd	Female	Adult	No	13
3rd	Female	Adult	No	89

Class	Sex	Age	Survived	Freq
Crew	Female	Adult	No	3
1st	Male	Child	Yes	5
2nd	Male	Child	Yes	11
3rd	Male	Child	Yes	13
Crew	Male	Child	Yes	0
1st	Female	Child	Yes	1
2nd	Female	Child	Yes	13
3rd	Female	Child	Yes	14
Crew	Female	Child	Yes	0
1st	Male	Adult	Yes	57
2nd	Male	Adult	Yes	14
3rd	Male	Adult	Yes	75
Crew	Male	Adult	Yes	192
1st	Female	Adult	Yes	140
2nd	Female	Adult	Yes	80
3rd	Female	Adult	Yes	76
Crew	Female	Adult	Yes	20

```
str(Titanic)
```

```
##  'table' num [1:4, 1:2, 1:2, 1:2] 0 0 35 0 0 0 17 0 118 154 ...
##  - attr(*, "dimnames")=List of 4
##   ..$ Class   : chr [1:4] "1st" "2nd" "3rd" "Crew"
##   ..$ Sex     : chr [1:2] "Male" "Female"
##   ..$ Age     : chr [1:2] "Child" "Adult"
##   ..$ Survived: chr [1:2] "No" "Yes"
```

Search for missing values on the dataset

```
any_missing <- any(is.na(Titanic))

if (any_missing) {
  print("The dataset contains missing values.")
} else {
  print("The dataset does not contain missing values.")
}
```

```
## [1] "The dataset does not contain missing values."
```

Install the ggplot2 package and transform dataset into a dataframe

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
Titanic_df <- as.data.frame(Titanic)
```

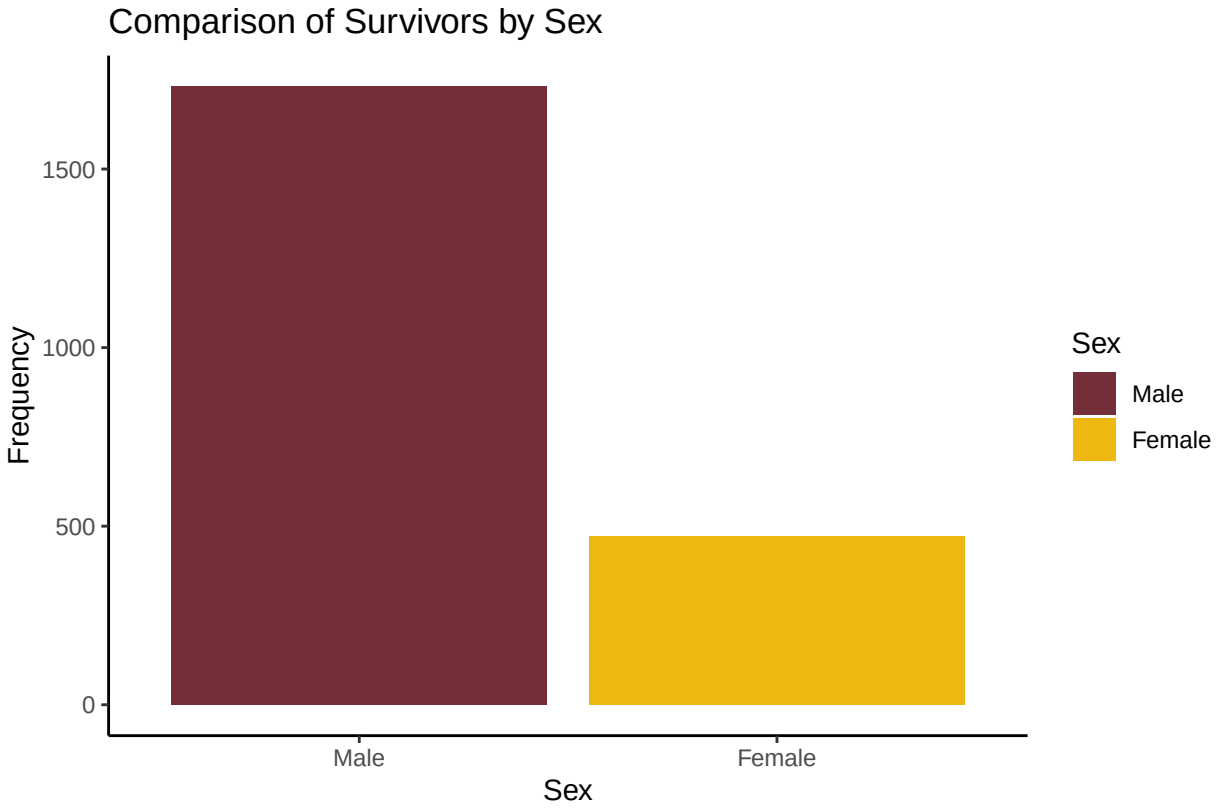
Visual Voyage: Aesthetic Insights

I have created three visually appealing graphs to compare the variables included in the Titanic dataset, namely sex, age, and class.

To ensure accuracy and reliability, it's important to note that these plots were generated using the Titanic dataset provided by R's datasets collection. However, in the event that you encounter any discrepancies or variations in the displayed data, please take them into consideration.

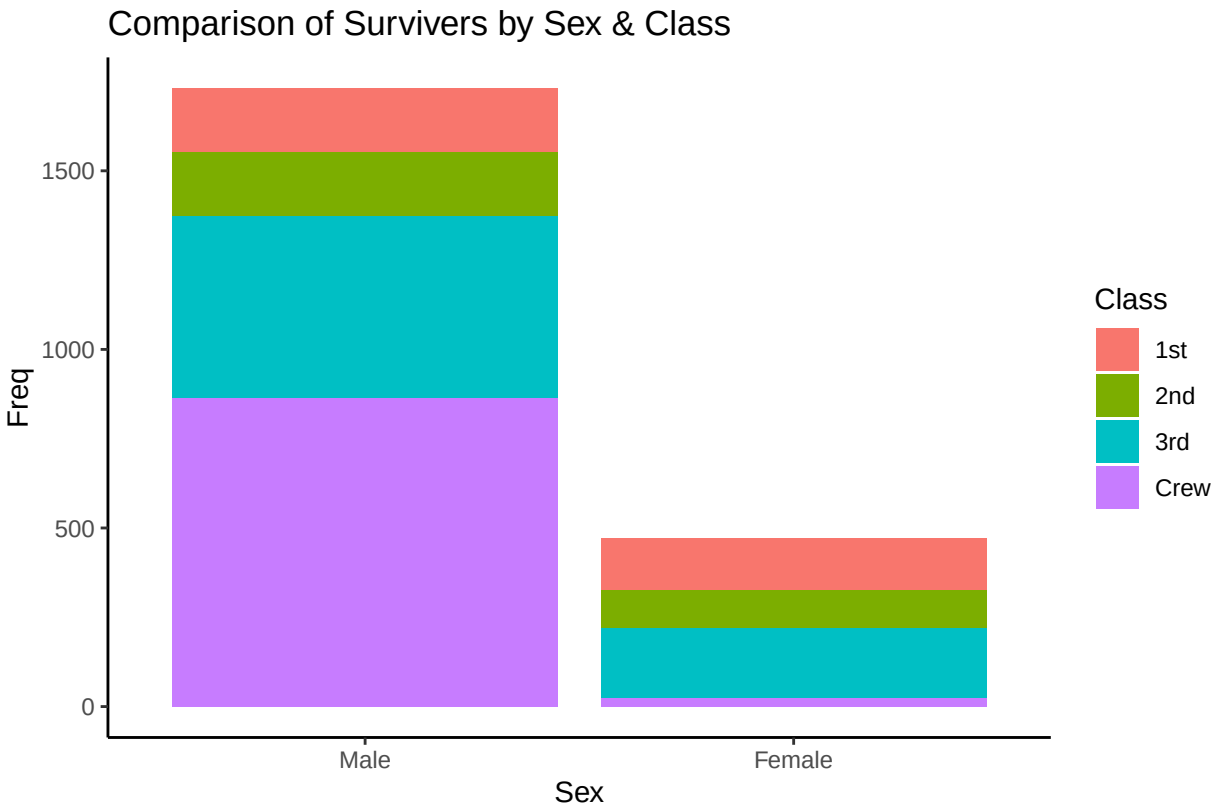
```
graph <- ggplot(data = Titanic_df) +
  geom_bar(mapping = aes(x = Sex, y = Freq, fill = Sex), stat = "identity") +
  ggtitle("Comparison of Survivors by Sex") +
  labs(y = "Frequency", caption = "*The label 'Frequency' indicates the rate of each variable described
  scale_fill_manual(values = c("#722F37", "#Efb810")) +
  theme_classic()

print(graph)
```



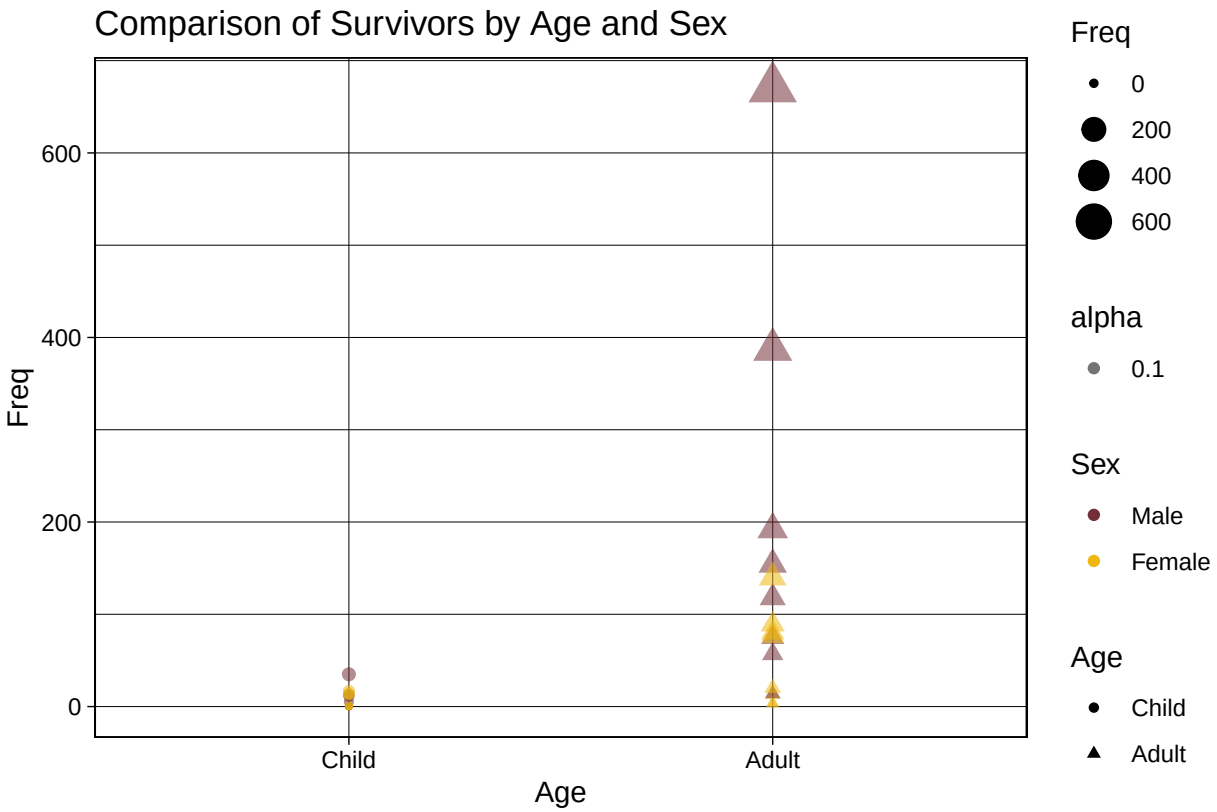
*The label 'Frequency' indicates the rate of each variable described on the graph

```
graph_1 <- ggplot(data = Titanic_df) +
  geom_bar(mapping = aes(x = Sex, y = Freq, fill = Class), stat = "identity") +
  ggtitle("Comparison of Survivers by Sex & Class") + labs(y = "Freq", caption = "*The label 'Freq' sta
print(graph_1)
```



he label 'Freq' states for Frequency and it indicates the rate of each variable described on the graph

```
graph_2 <- ggplot(data = Titanic_df) +
  geom_point(mapping = aes(x = Age, y = Freq, shape = Age, size = Freq, alpha = 0.1, color = Sex)) +
  scale_color_manual(values = c("#722F37", "#Efb810")) +
  ggtitle("Comparison of Survivors by Age and Sex") + labs(y = "Freq", caption = "*The label 'Freq' sta
print(graph_2)
```



: label 'Freq' states for Frequency and it indicates the rate of each variable described on the graph

Conclusion

The objective of this project was to explore an intriguing topic while executing some data analysis procedures. Besides, I have a particular passion for exploring historical events. Hope you enjoyed!