Data Visualization Lab Assignment

Muneeb Iftikhar

2026-02-26

Dataset Description

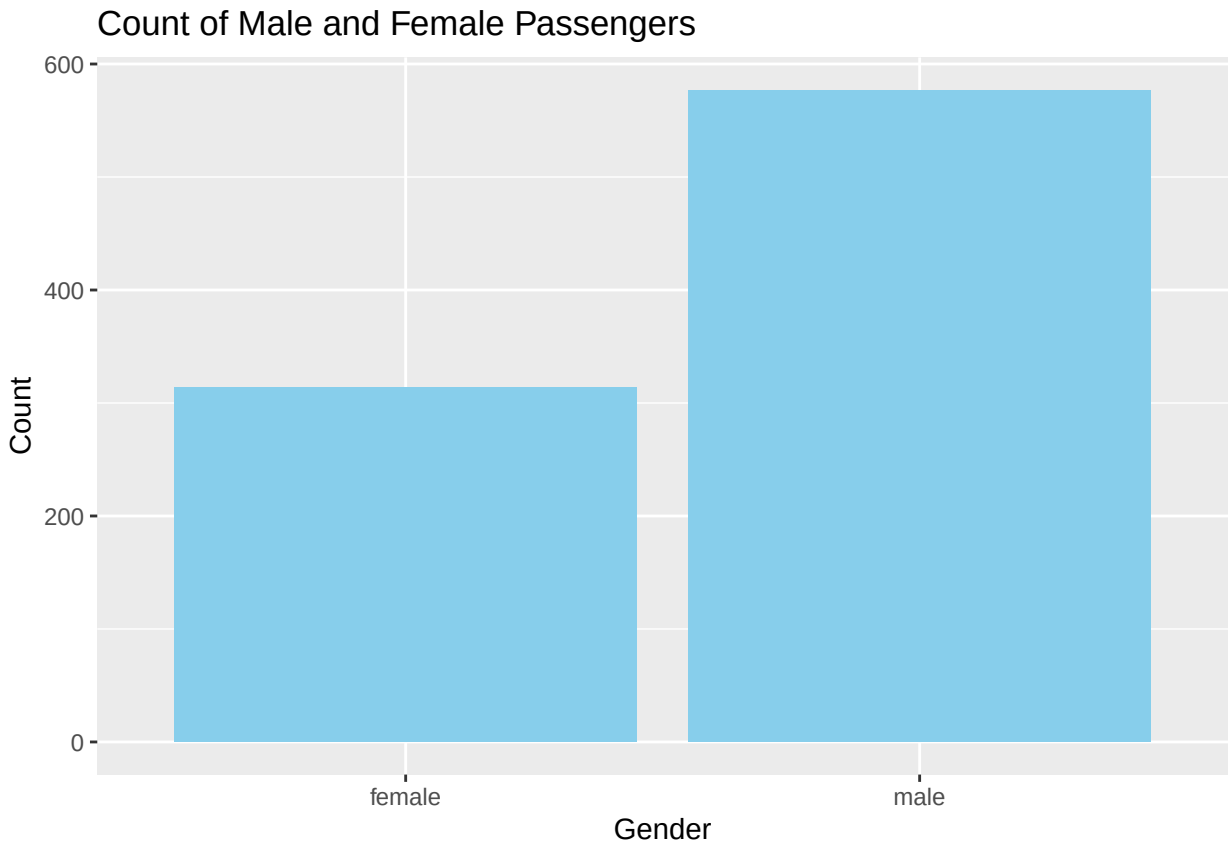
The Titanic dataset contains passenger information including survival status, passenger class, age, gender, fare, and embarkation port.

- Total Observations: 891
- Variables: PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked

```
data <- read.csv("Titanic-Dataset.csv")

# Handle missing values
data$Age[is.na(data$Age)] <- median(data$Age, na.rm=TRUE)
data$Embarked[is.na(data$Embarked)] <- "S"
```

1. Bar Chart – Gender Count

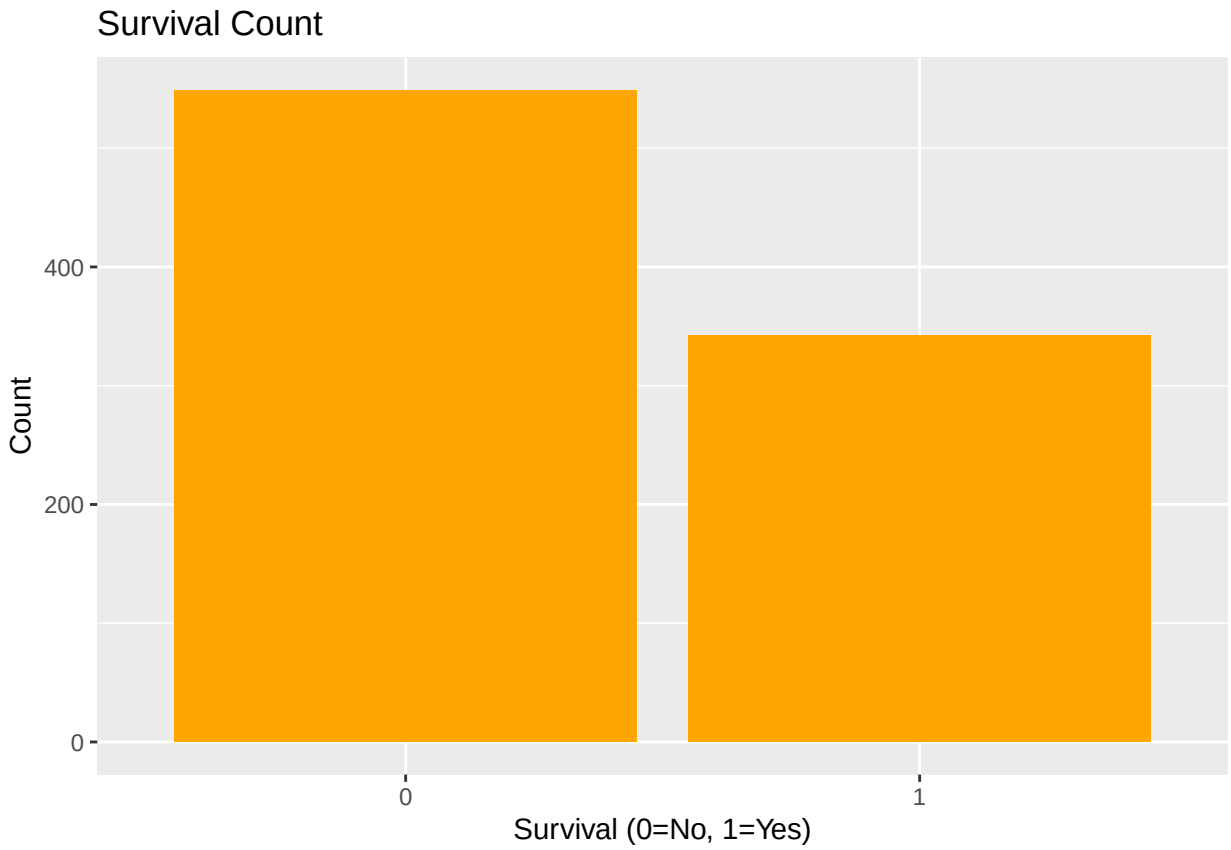
```
ggplot(data, aes(x=Sex)) +
  geom_bar(fill="skyblue") +
  labs(title="Count of Male and Female Passengers",
       x="Gender", y="Count")
```



Interpretation: Male passengers were higher in number compared to female passengers.

2. Survival Count

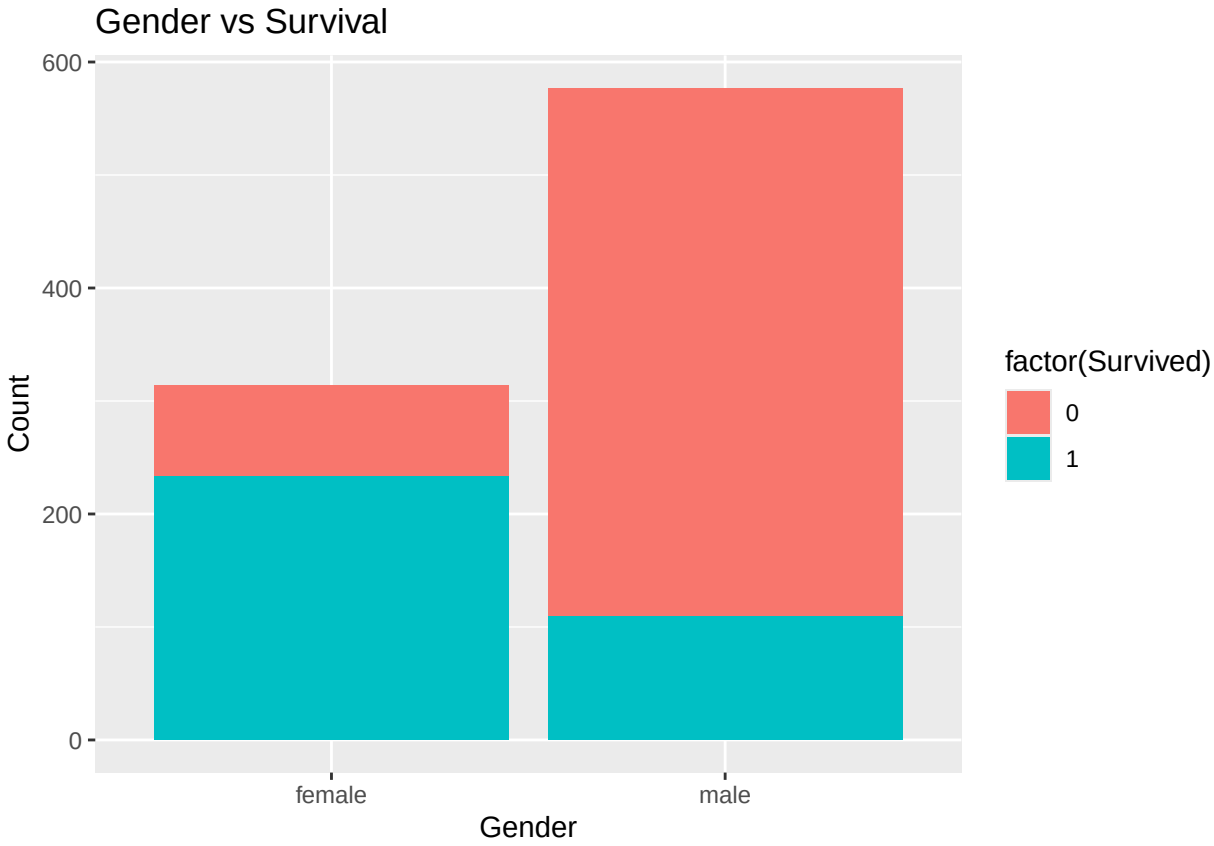
```
ggplot(data, aes(x=factor(Survived))) +
  geom_bar(fill="orange") +
  labs(title="Survival Count",
       x="Survival (0=No, 1=Yes)", y="Count")
```



Interpretation: More passengers did not survive compared to survivors.

3. Gender vs Survival

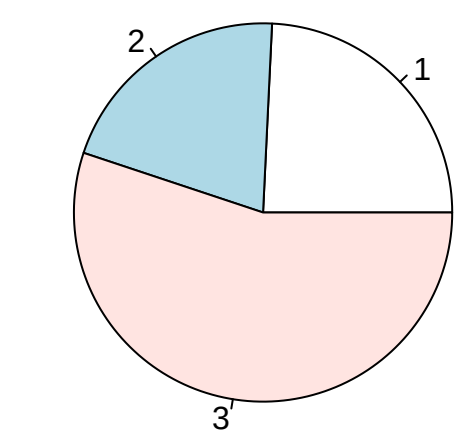
```
ggplot(data, aes(x=Sex, fill=factor(Survived))) +
  geom_bar() +
  labs(title="Gender vs Survival",
       x="Gender", y="Count")
```



Interpretation: Females had higher survival proportion.

4. Passenger Class Pie Chart

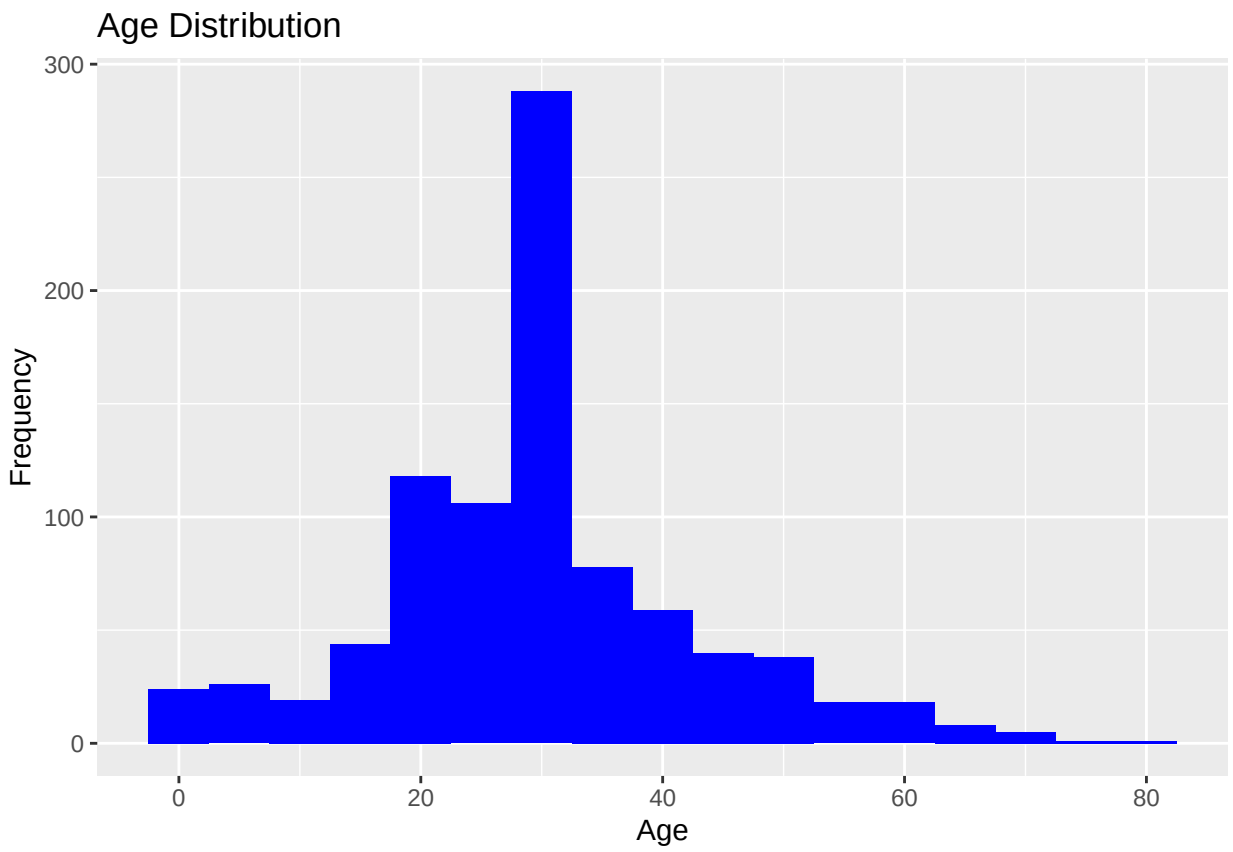
```
class_count <- table(data$Pclass)
pie(class_count, main="Passenger Class Distribution")
```



Interpretation: Most passengers were from third class.

5. Age Histogram

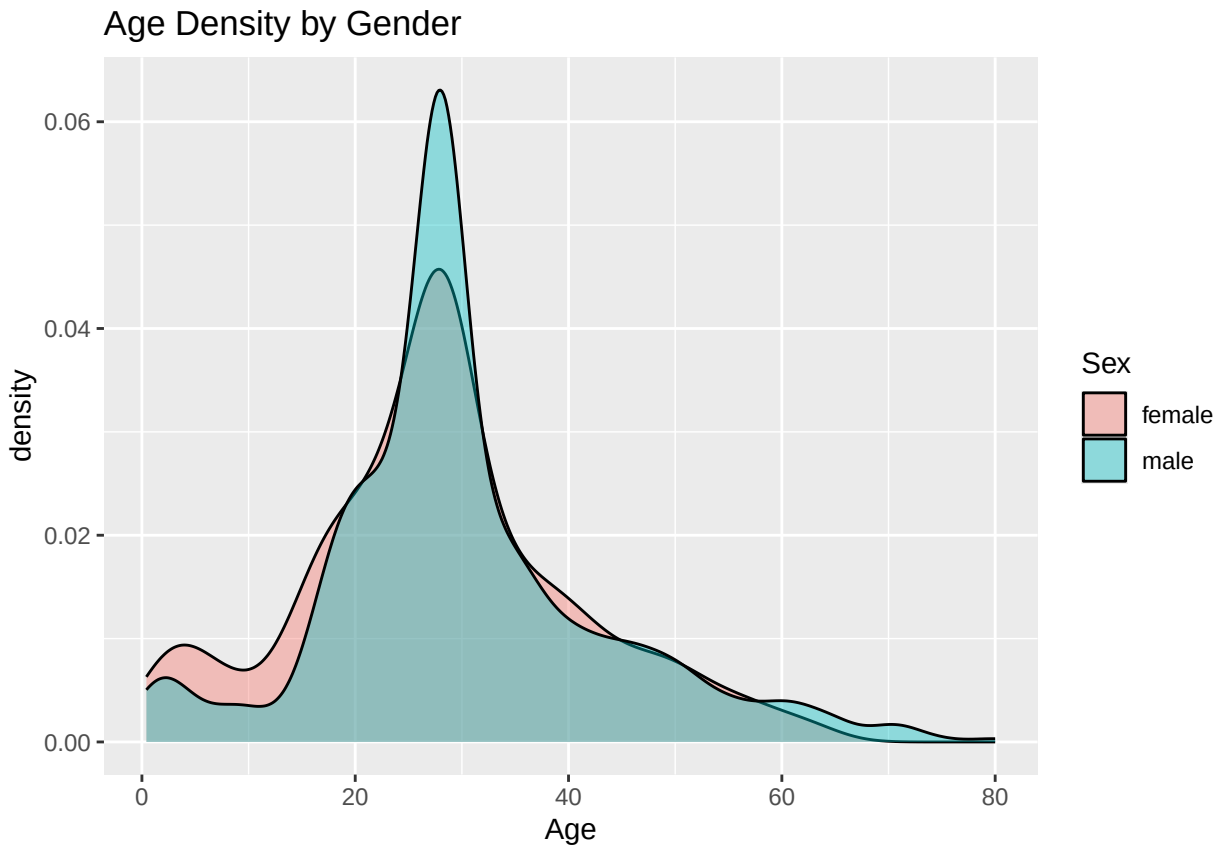
```
ggplot(data, aes(x=Age)) +
  geom_histogram(binwidth=5, fill="blue") +
  labs(title="Age Distribution",
       x="Age", y="Frequency")
```



Interpretation: Most passengers were between 20–40 years.

6. Density Plot

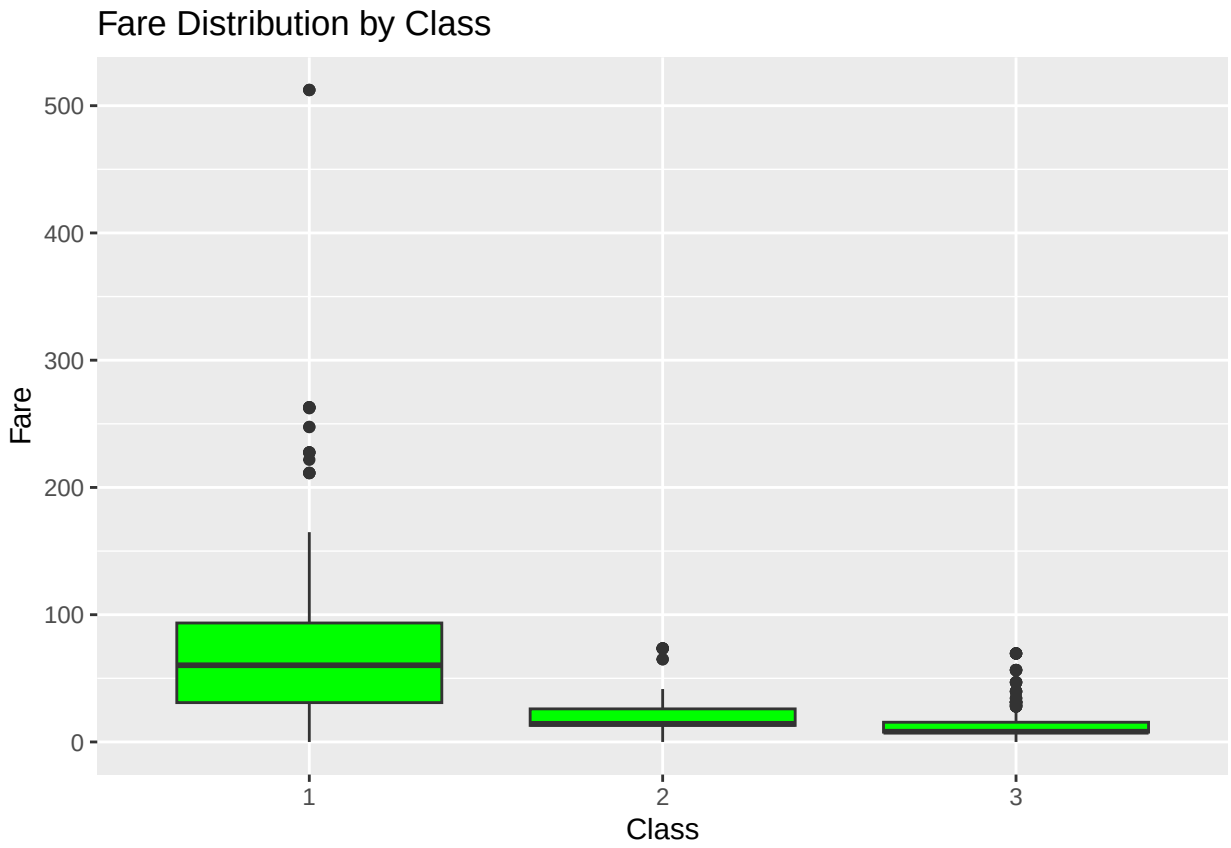
```
ggplot(data, aes(x=Age, fill=Sex)) +
  geom_density(alpha=0.4) +
  labs(title="Age Density by Gender")
```



Interpretation: Similar age distribution across genders.

7. Boxplot (Fare by Class)

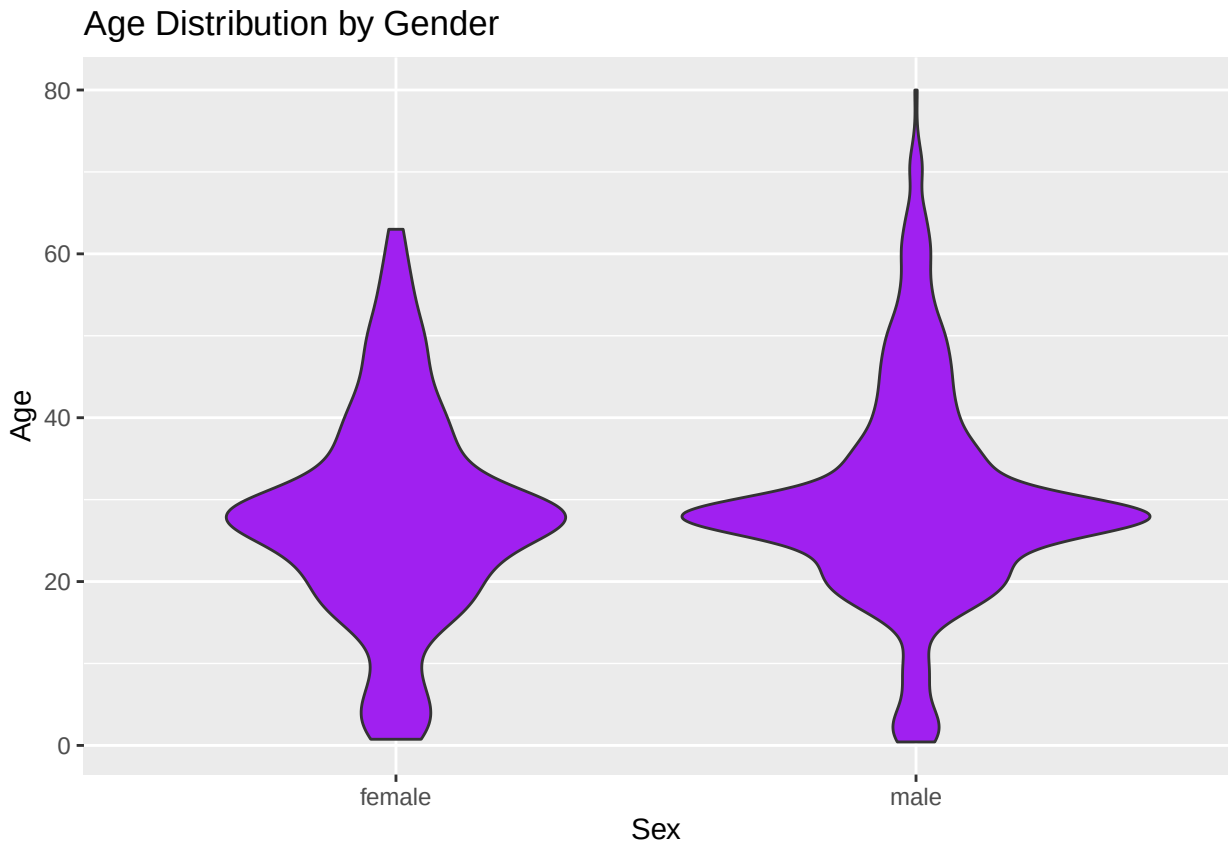
```
ggplot(data, aes(x=factor(Pclass), y=Fare)) +
  geom_boxplot(fill="green") +
  labs(title="Fare Distribution by Class",
       x="Class", y="Fare")
```



Interpretation: First-class passengers paid highest fares.

8. Violin Plot

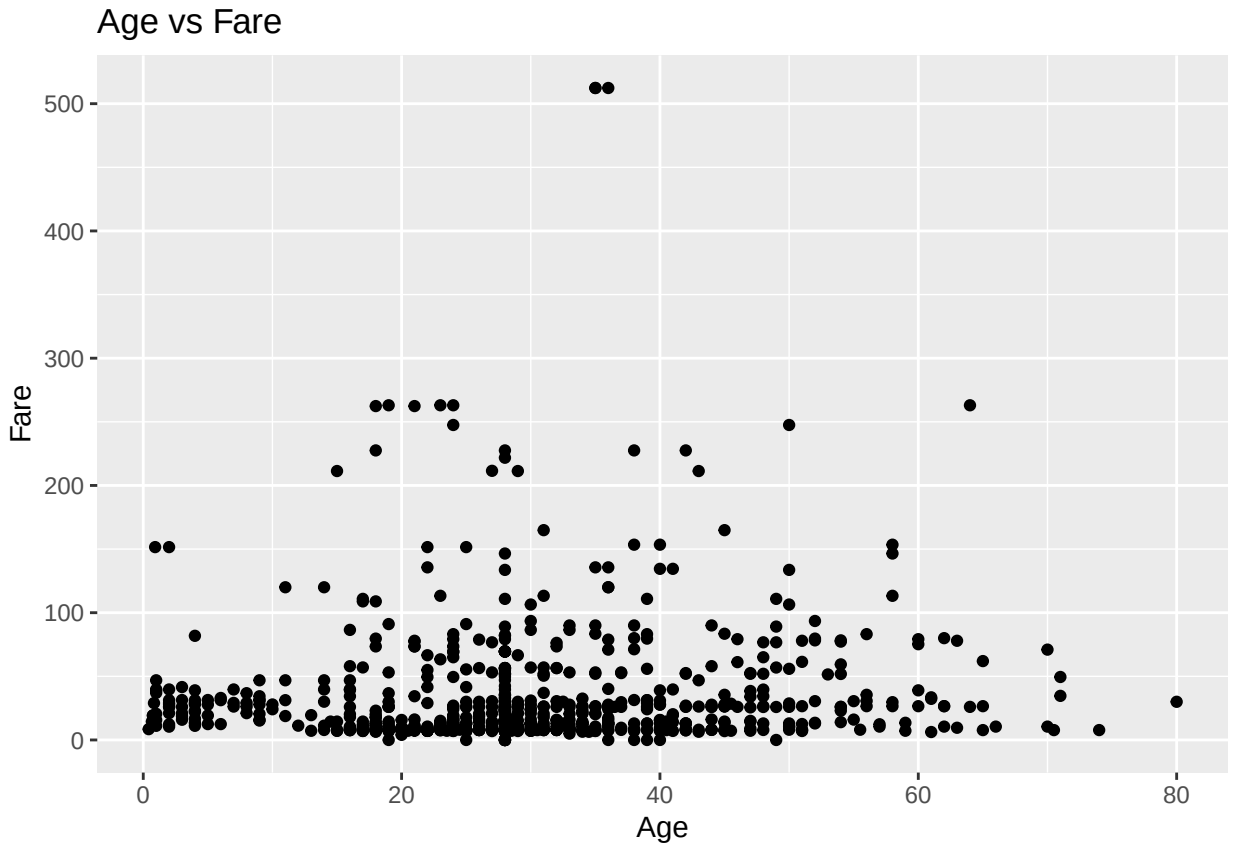
```
ggplot(data, aes(x=Sex, y=Age)) +
  geom_violin(fill="purple") +
  labs(title="Age Distribution by Gender")
```



Interpretation: Age spread similar for both genders.

9. Scatter Plot

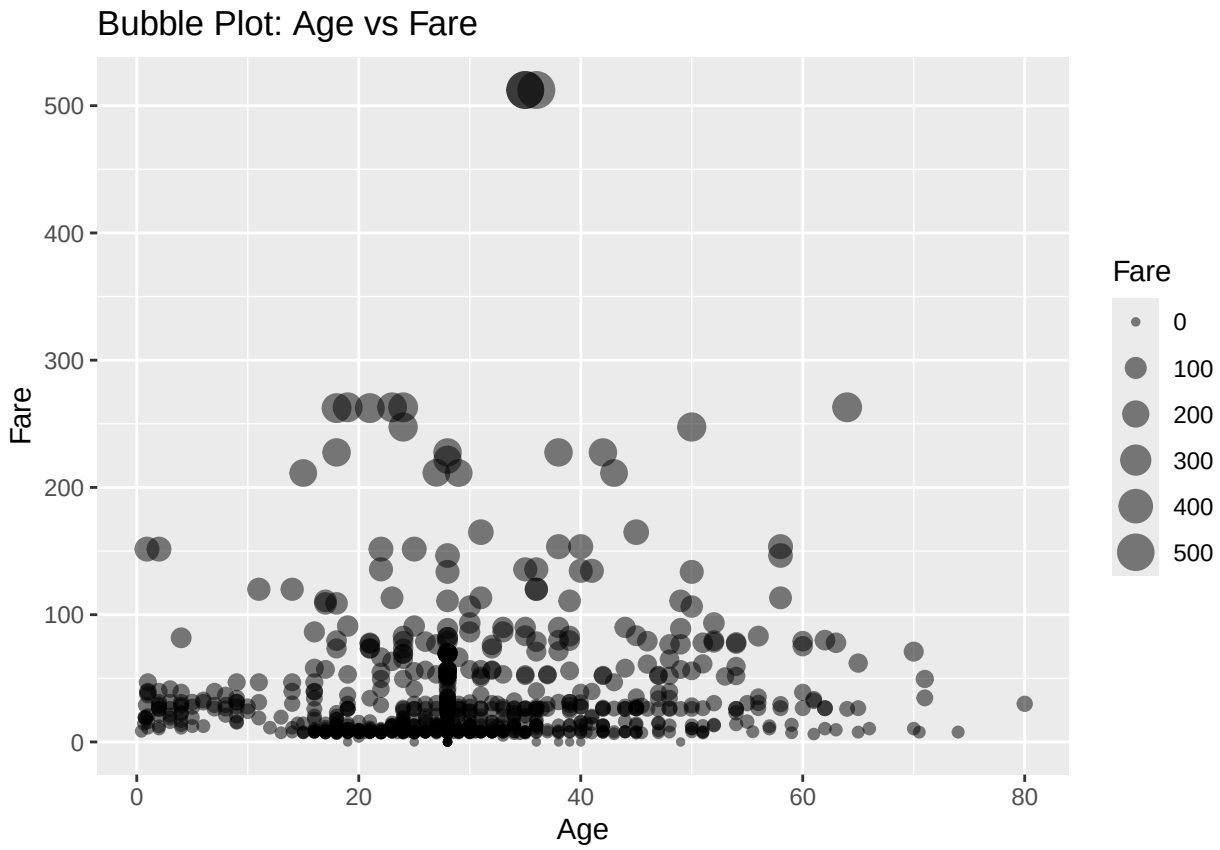
```
ggplot(data, aes(x=Age, y=Fare)) +
  geom_point() +
  labs(title="Age vs Fare")
```



Interpretation: No strong relationship between age and fare.

10. Bubble Plot

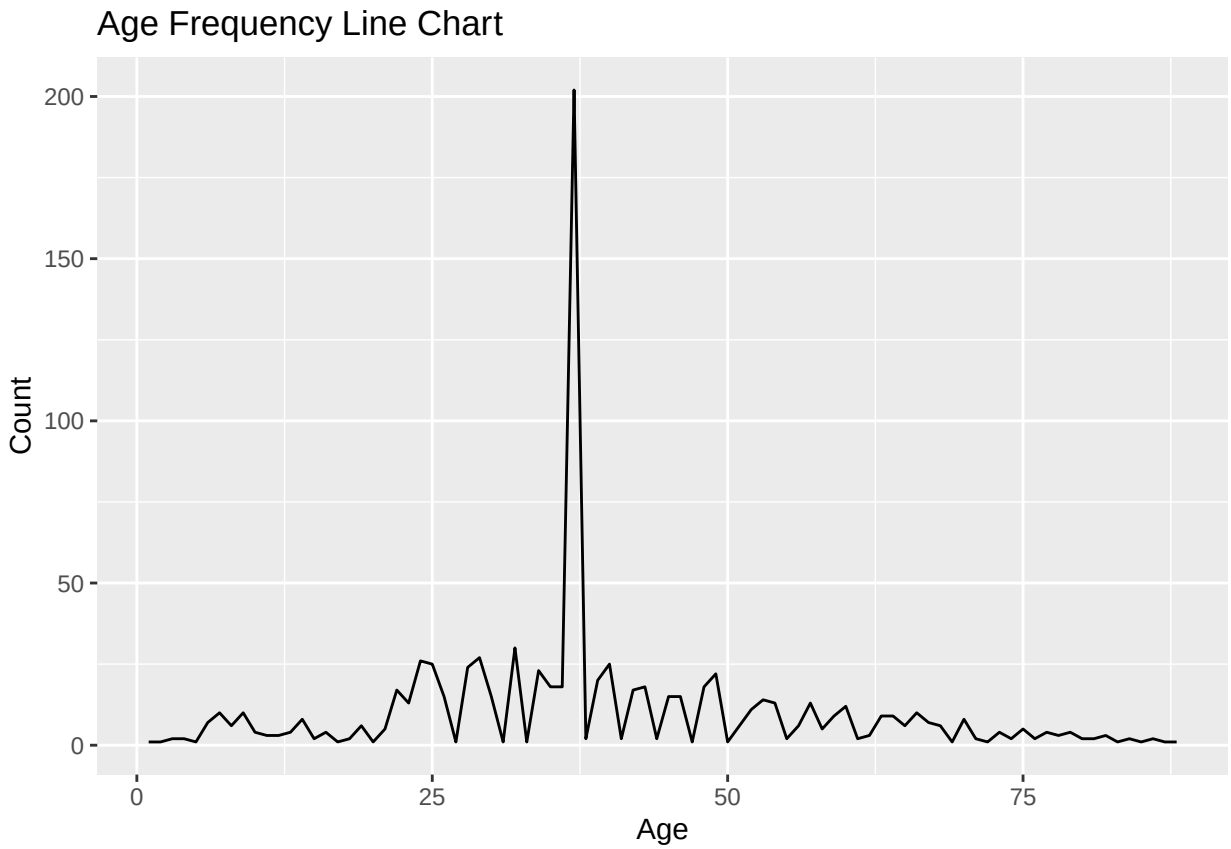
```
ggplot(data, aes(x=Age, y=Fare, size=Fare)) +
  geom_point(alpha=0.5) +
  labs(title="Bubble Plot: Age vs Fare")
```



Interpretation: Larger bubbles represent higher fares.

11. Line Chart

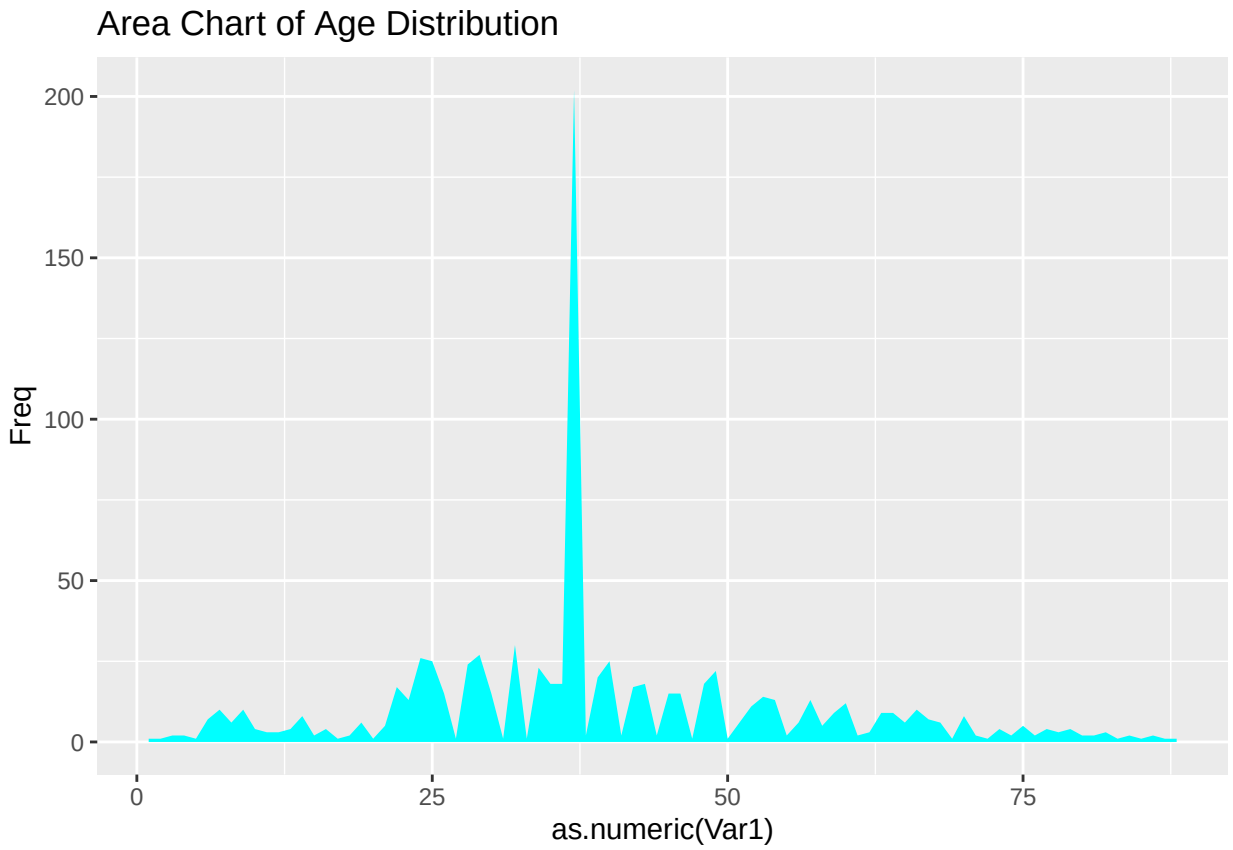
```
age_table <- as.data.frame(table(data$Age))

ggplot(age_table, aes(x=as.numeric(Var1), y=Freq)) +
  geom_line() +
  labs(title="Age Frequency Line Chart",
       x="Age", y="Count")
```



Interpretation: Certain ages appear more frequently.

12. Area Chart

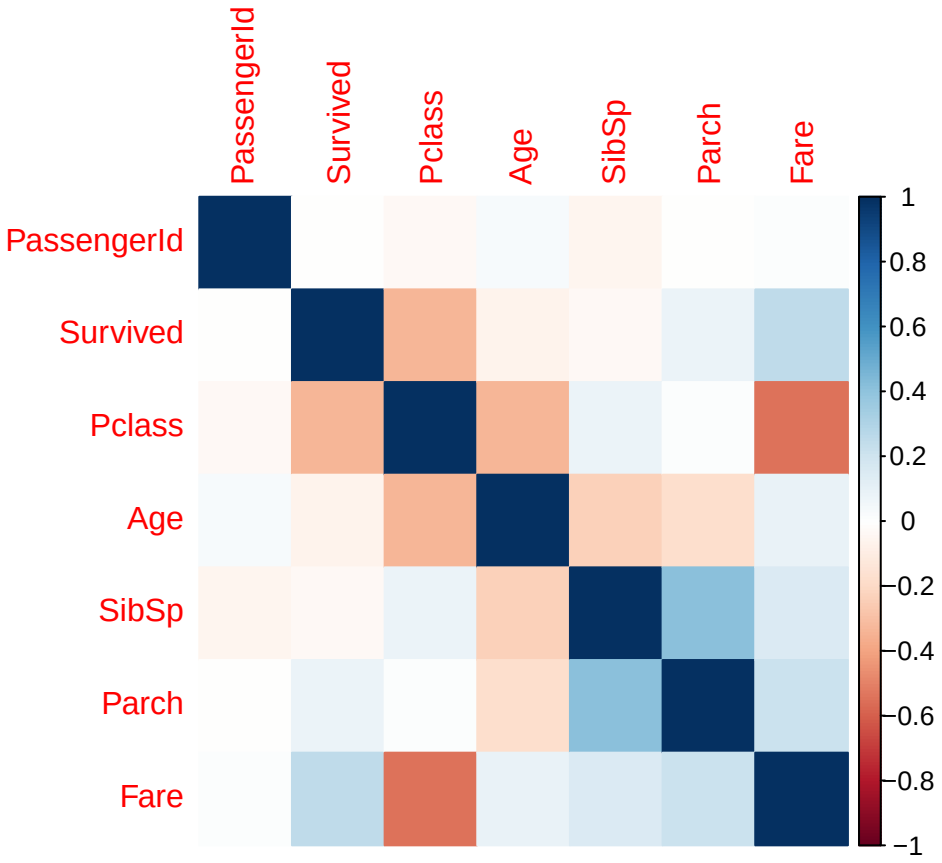
```
ggplot(age_table, aes(x=as.numeric(Var1), y=Freq)) +
  geom_area(fill="cyan") +
  labs(title="Area Chart of Age Distribution")
```



Interpretation: Shows overall age concentration.

13. Correlation Heatmap

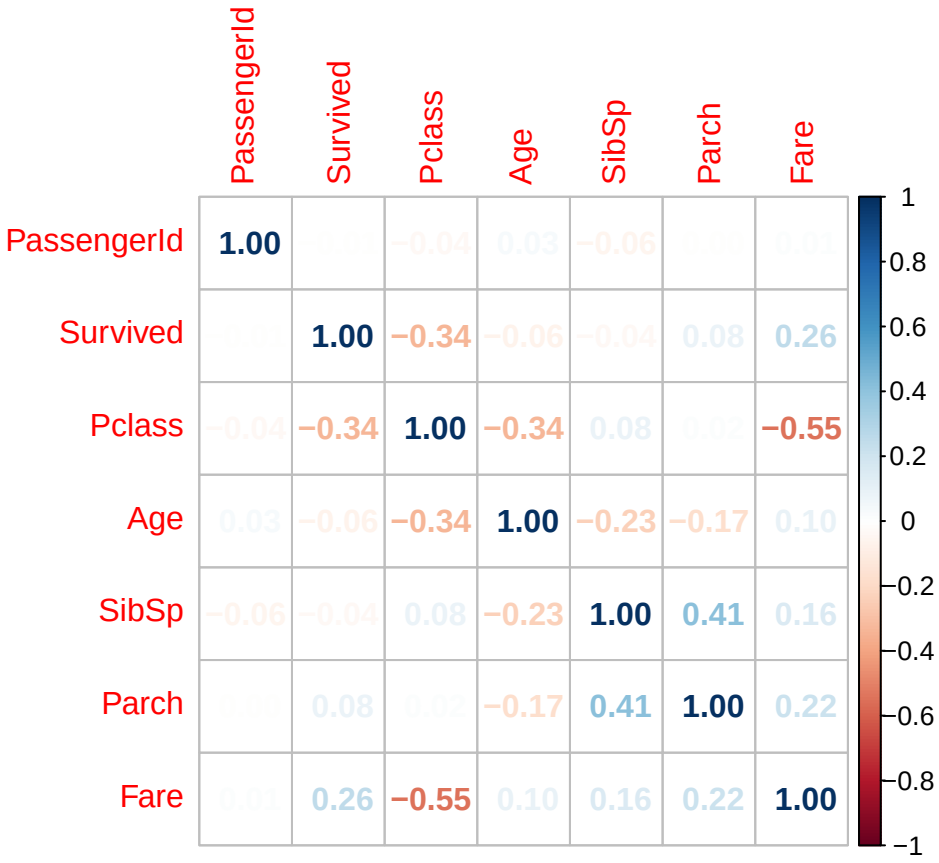
```
num_data <- data[,sapply(data,is.numeric)]
cor_matrix <- cor(num_data)
corrplot(cor_matrix, method="color")
```



Interpretation: Survival relates with class and fare.

14. Correlation Numbers

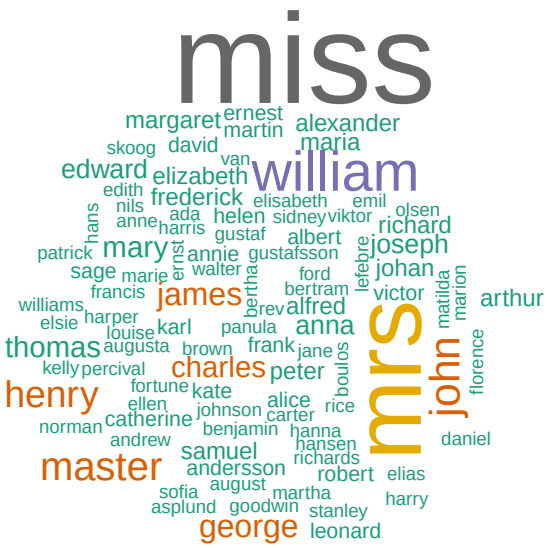
```
corrplot(cor_matrix, method="number")
```



Interpretation: Displays exact correlation values.

15. Word Cloud

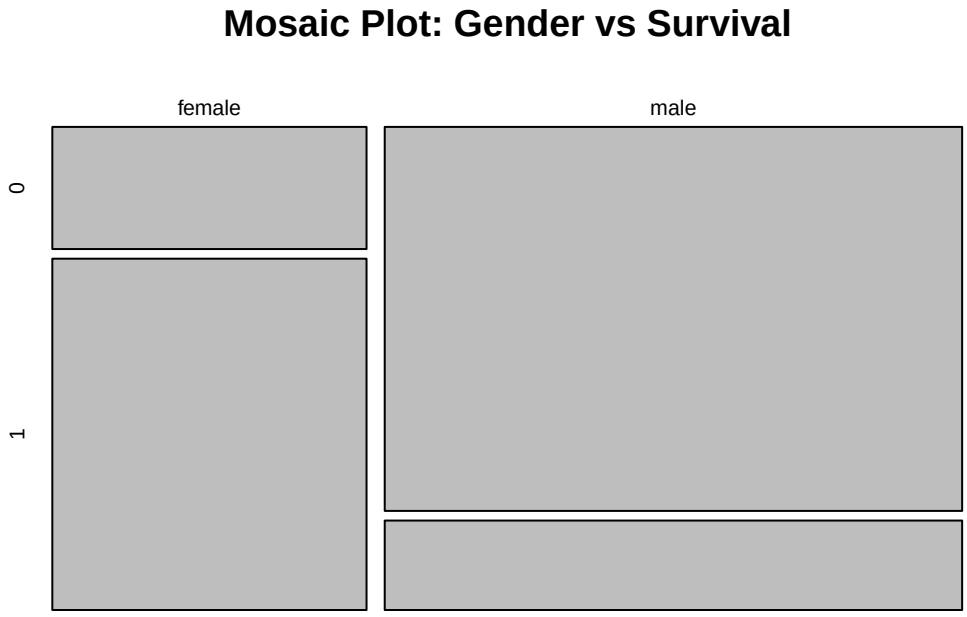
```
wordcloud(data$Name, max.words=100,
          colors=brewer.pal(8,"Dark2"))
```



Interpretation: Shows frequent words in passenger names.

16. Mosaic Plot

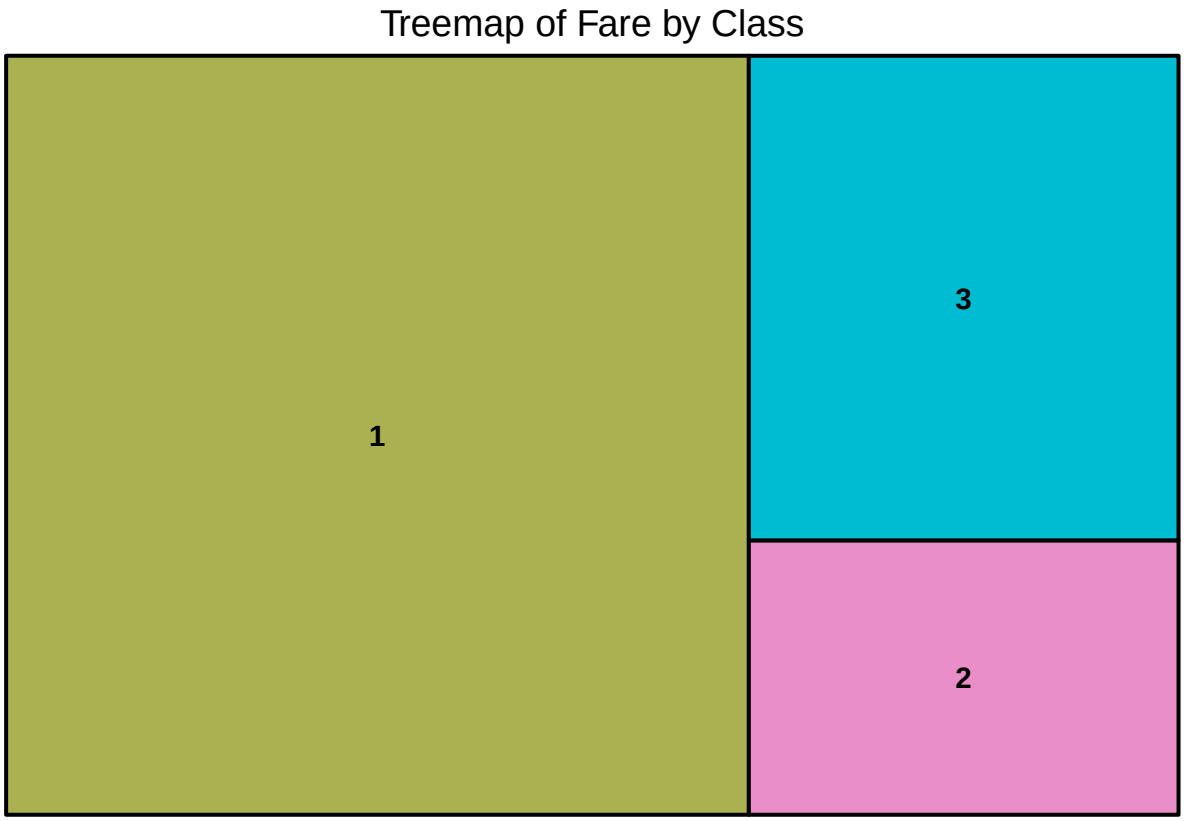
```
mosaicplot(table(data$Sex, data$Survived),
           main="Mosaic Plot: Gender vs Survival")
```



Interpretation: Females had higher survival rate.

17. Treemap

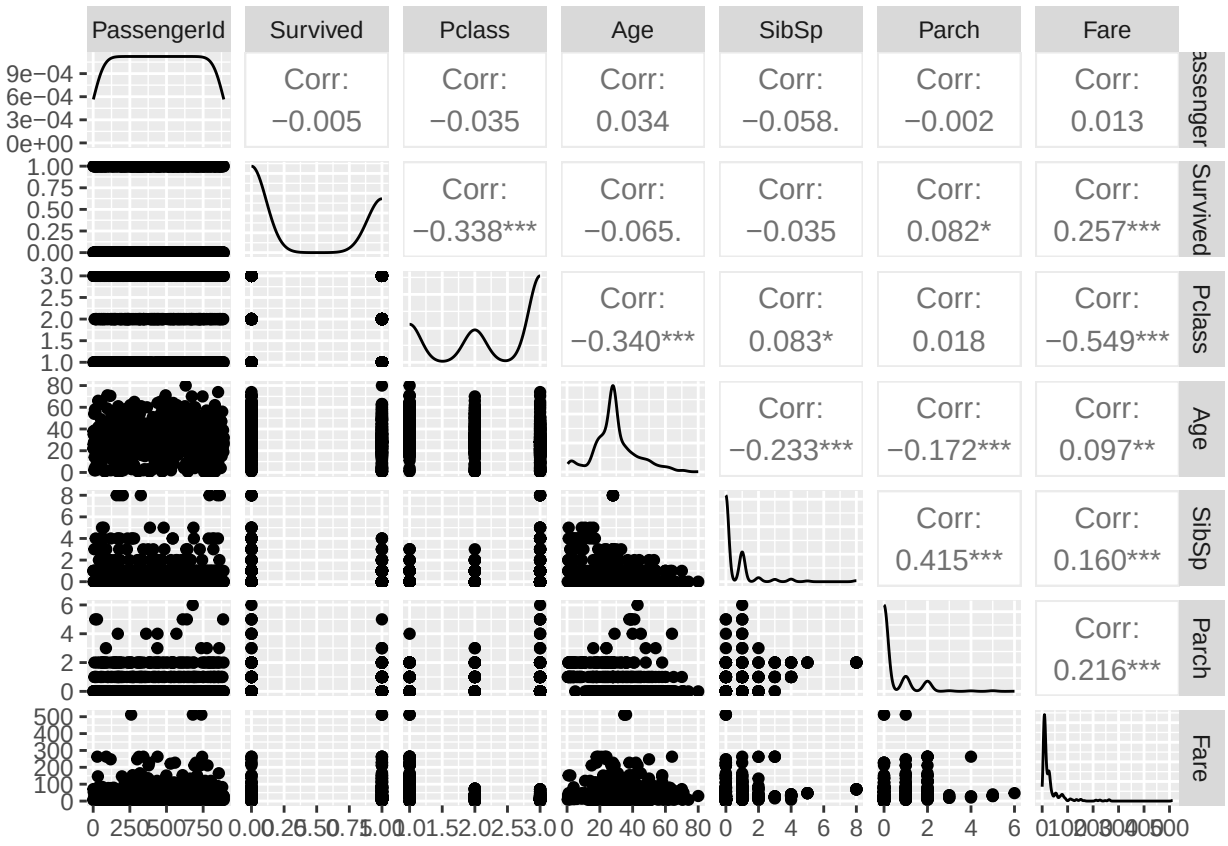
```
treemap(data,
        index="Pclass",
        vSize="Fare",
        title="Treemap of Fare by Class")
```



Interpretation: First class contributes most fare.

18. Pair Plot

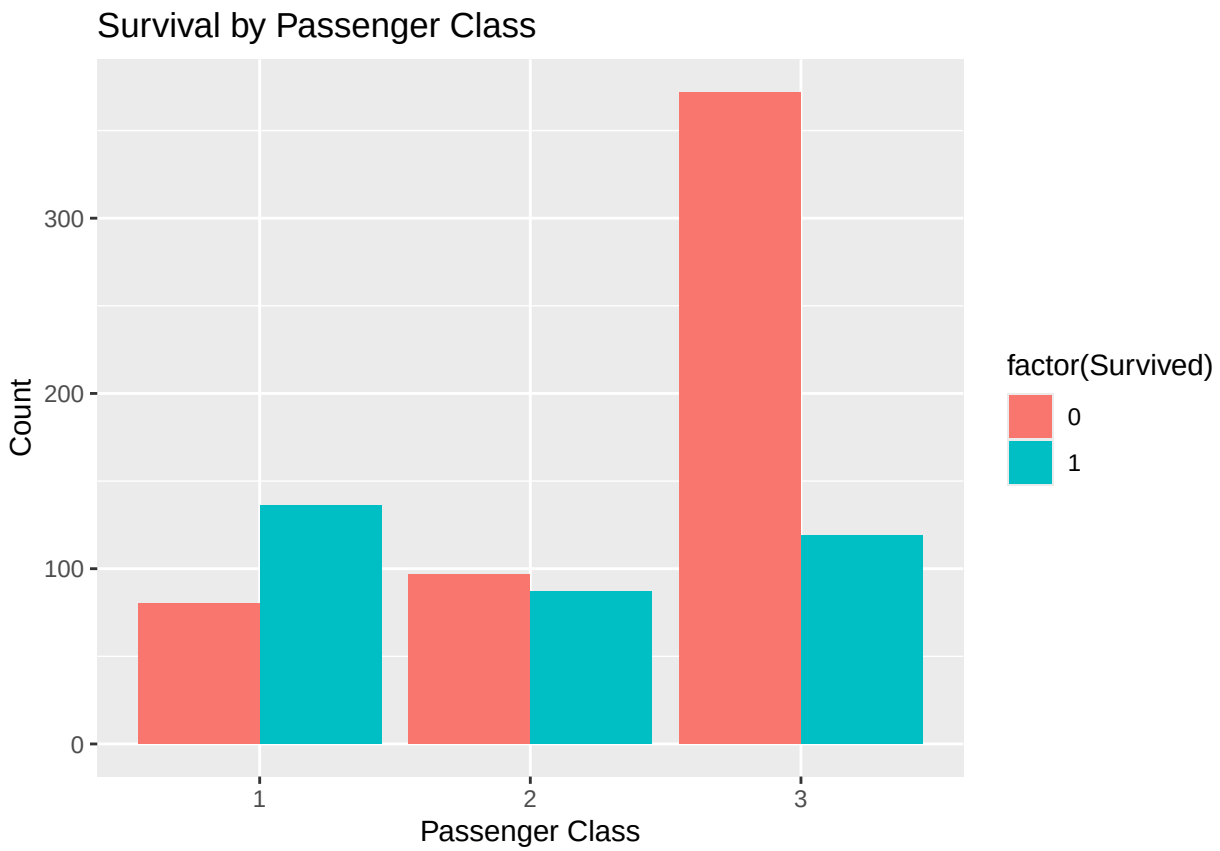
```
ggpairs(num_data)
```



Interpretation: Shows relationships among numeric variables.

19. Survival by Class

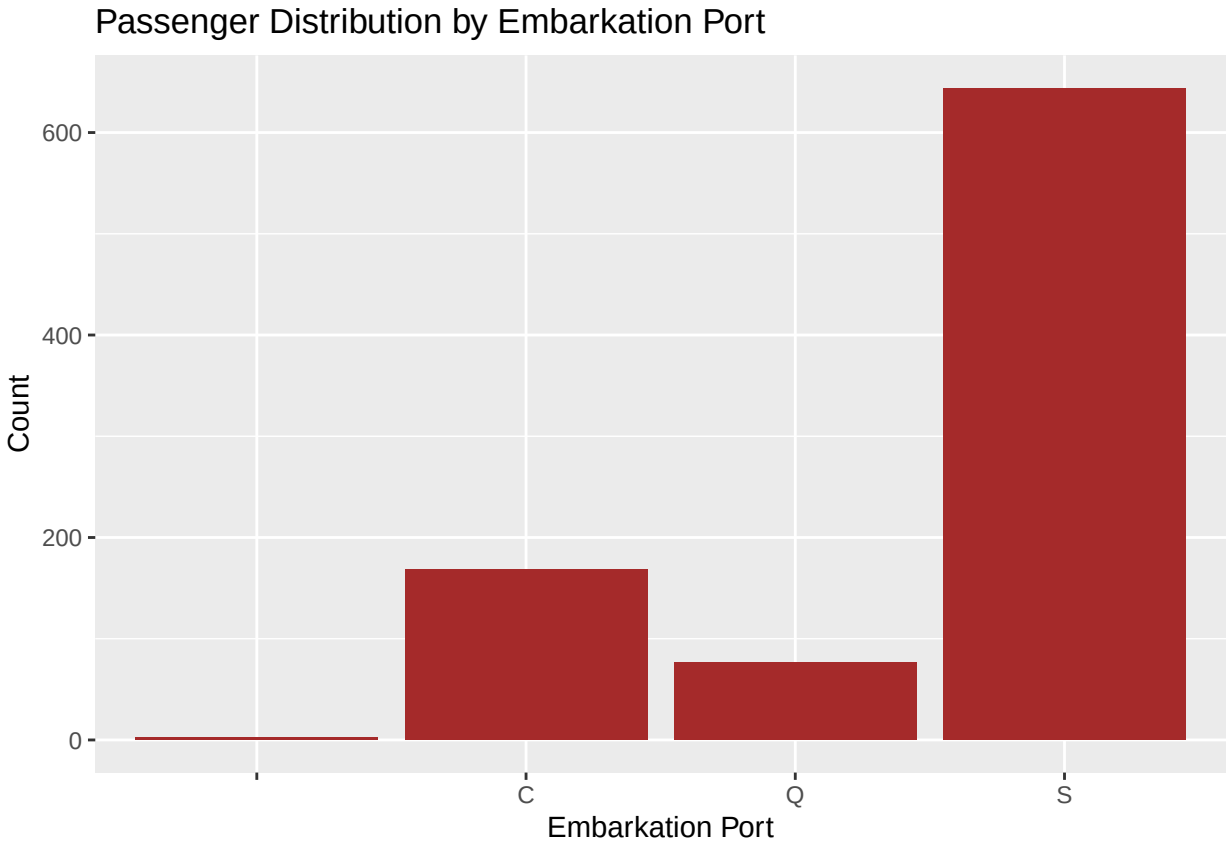
```
ggplot(data, aes(x=factor(Pclass), fill=factor(Survived))) +
  geom_bar(position="dodge") +
  labs(title="Survival by Passenger Class",
       x="Passenger Class", y="Count")
```



Interpretation: First-class passengers survived more.

20. Embarked Distribution

```
ggplot(data, aes(x=Embarked)) +
  geom_bar(fill="brown") +
  labs(title="Passenger Distribution by Embarkation Port",
       x="Embarkation Port", y="Count")
```



Interpretation: Most passengers embarked from port S.

Conclusion

The analysis shows that gender and passenger class significantly influenced survival. Females and first-class passengers had better survival rates.