Name : Tanishq Thuse

Year : SY

Div : CS(AI)-B

Roll no : 60

PRN : 12310237

Subject : Assignment-9 DV

Assignment-9 :

Write observations on the inference from the above statistics.

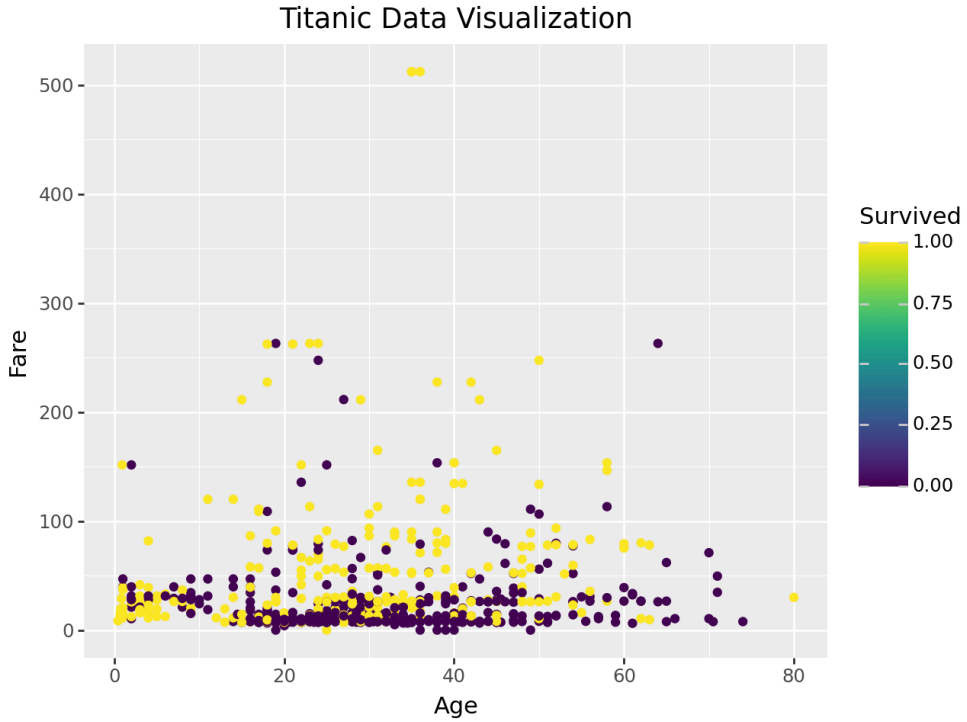
(The Inferences are mentioned at the end of the document)

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
path = "/content/titanic.csv"
df = pd.read_csv(path)
```

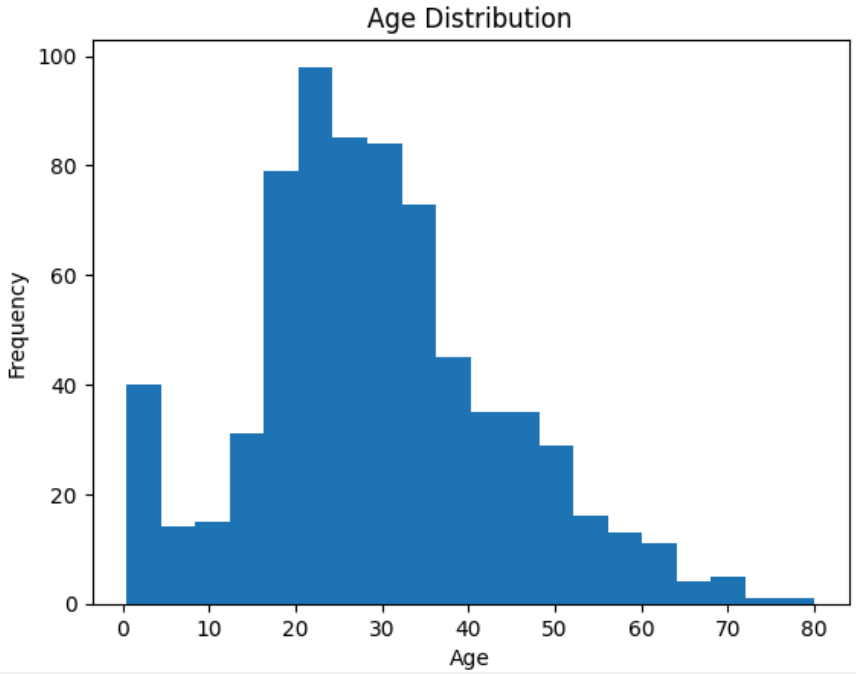
```
(ggplot(df)
 + aes(x='Age', y='Fare', color='Survived')
 + geom_point()
 + labs(title='Titanic Data Visualization', x='Age', y='Fare', color='Survived')
)
```

```
/usr/local/lib/python3.10/dist-packages/plotnine/layer.py:364: PlotnineWarning: geom_point : Removed 177 row
```

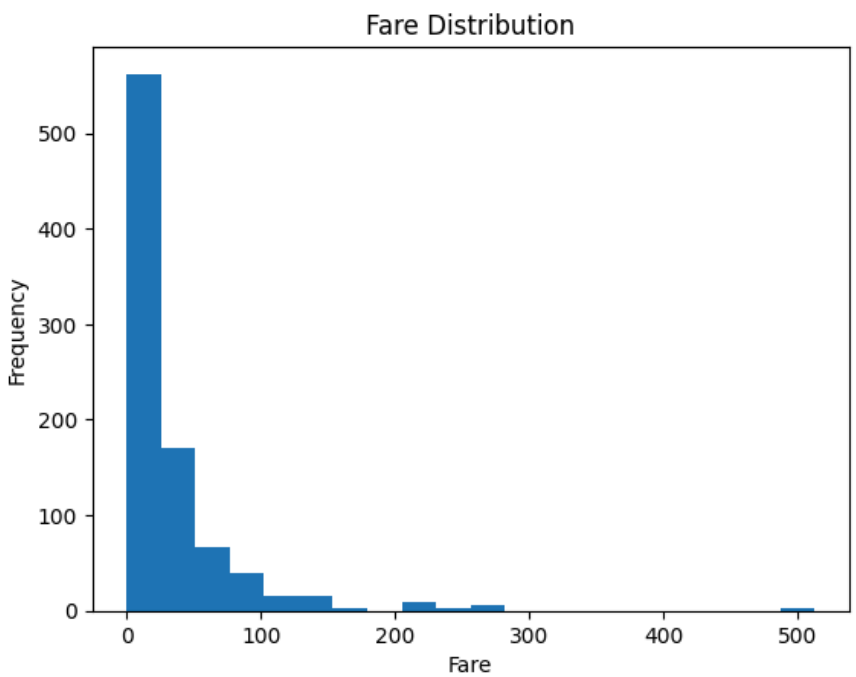



```
import matplotlib.pyplot as plt
# Histogram of Age
plt.hist(df['Age'].dropna(), bins=20)
```

```
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Age Distribution')
plt.show()
```



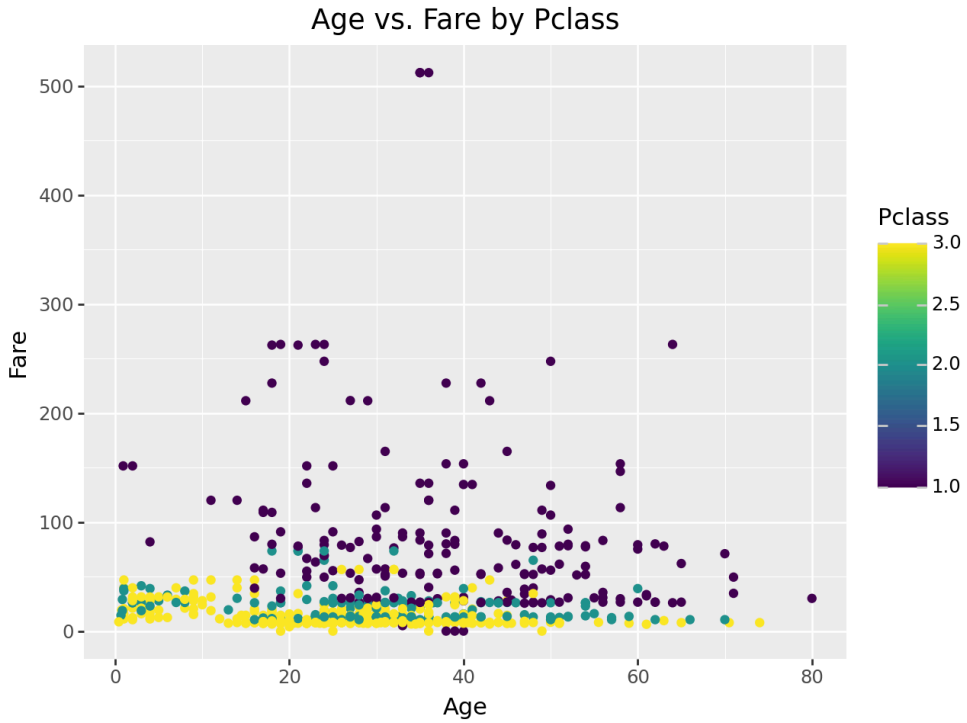

```
import matplotlib.pyplot as plt
# Histogram of Fare
plt.hist(df['Fare'].dropna(), bins=20)
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.title('Fare Distribution')
plt.show()
```



```
# ggplot for Age vs. Fare with different color based on Pclass
from plotnine import *

(ggplot(df)
 + aes(x='Age', y='Fare', color='Pclass')
 + geom_point()
 + labs(title='Age vs. Fare by Pclass', x='Age', y='Fare', color='Pclass')
)
```

```
/usr/local/lib/python3.10/dist-packages/plotnine/layer.py:364: PlotnineWarning: geom_point : Removed 177 row
```



General Insights

Observations based on the provided statistics:

1. Age and Fare Relationship:
   - The scatter plot of Age vs. Fare suggests a weak positive correlation, especially for passengers with higher fares, where age tends to be slightly higher.
   - Younger passengers with a wider range of fares are observed in comparison with older passengers, who generally have higher fares.
2. Survival and Age/Fare:
   - The color coding by 'Survived' in the Age vs. Fare scatter plot suggests a potentially more balanced survival rate for passengers with lower fares and diverse ages.
   - Further analysis might be needed to establish a definitive trend between age, fare, and survival.
3. Pclass and Age/Fare:
   - The scatter plot of Age vs. Fare colored by 'Pclass' indicates a clear distinction in fare for passengers belonging to different classes.
   - Passengers in higher Pclasses tend to have higher fares, and this group spans a greater range of ages.
4. Age and Fare Distribution:
   - The histograms of Age and Fare show a large portion of passengers in a younger age group (15-40).

- Fares are skewed towards lower values, with some passengers paying significantly higher fares.

Further Analysis:

- Explore survival rate across different age ranges.
- Analyze the relationship between Pclass, Age, Fare, and Survival.
- Determine correlation coefficients for Age and Fare.
- Conduct statistical tests to confirm insights about survival based on age and fare.