Milestone_6

Charles James

2023-10-22

```r
titanic = read.csv("titanic.csv")
colnames(titanic)
```

```
##  [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"
##  [6] "Age"         "SibSp"       "Parch"       "Ticket"      "Fare"
## [11] "Cabin"       "Embarked"
```

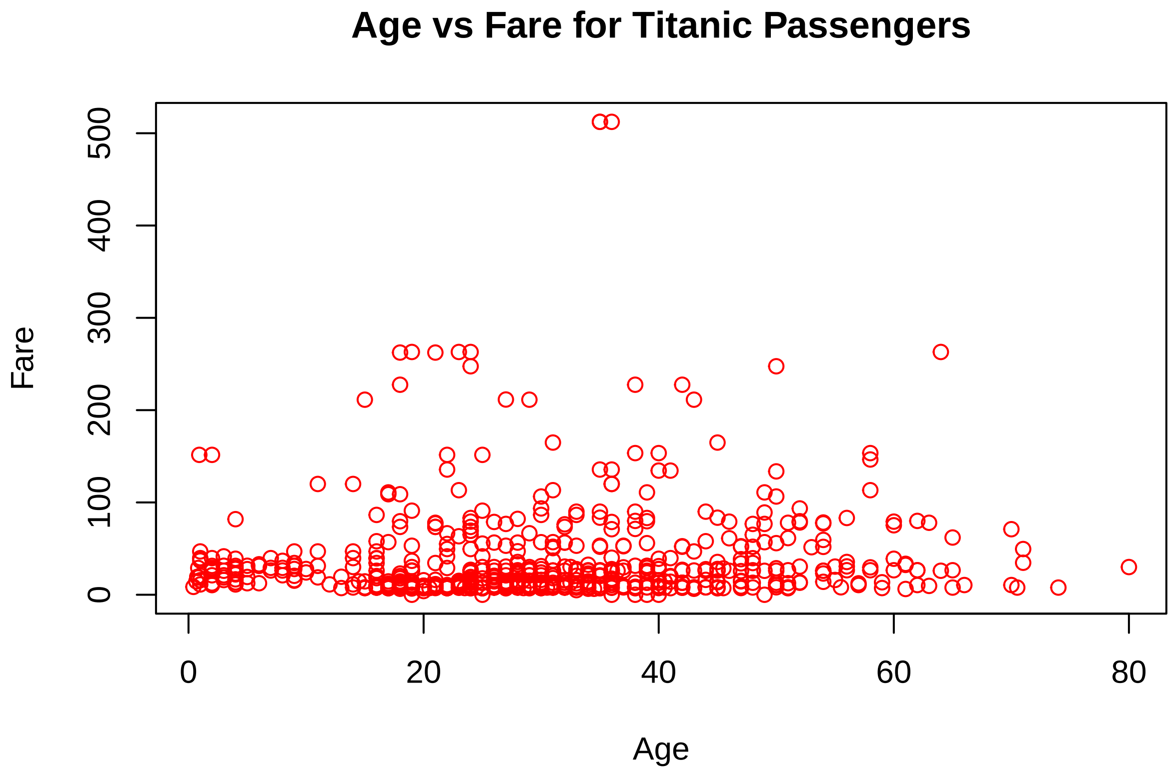
```r
dim(titanic)
```

```
## [1] 891  12
```

Age vs Ticket Price

```r
titanic <- na.omit(titanic)
age = titanic$Age
fare = titanic$Fare

plot(age, fare,
     xlab = "Age",
     ylab = "Fare",
     main = "Age vs Fare for Titanic Passengers",
     col = "red"
     )
```



```
cor(age, fare)
```

```
## [1] 0.09606669
```

We can see that there isn't much correlation, as the data points are spread out from each other. The majority of the data points are between 0 - 100 on the Y-Axis which indicates the price for the ticket. This means that the vast majority of people paid less than $100 for their ticket. Even still there are still to many data points scattered around that region to form any significant correlation between the two variables. Using R, the correlation coefficient is .09, which supports what the histogram shows.