

SEIS 631 - Final Project Draft

Adnan Suri

4/24/2022

```
setwd("/Users/Home/Desktop/St Thomas/SEIS 631")
```

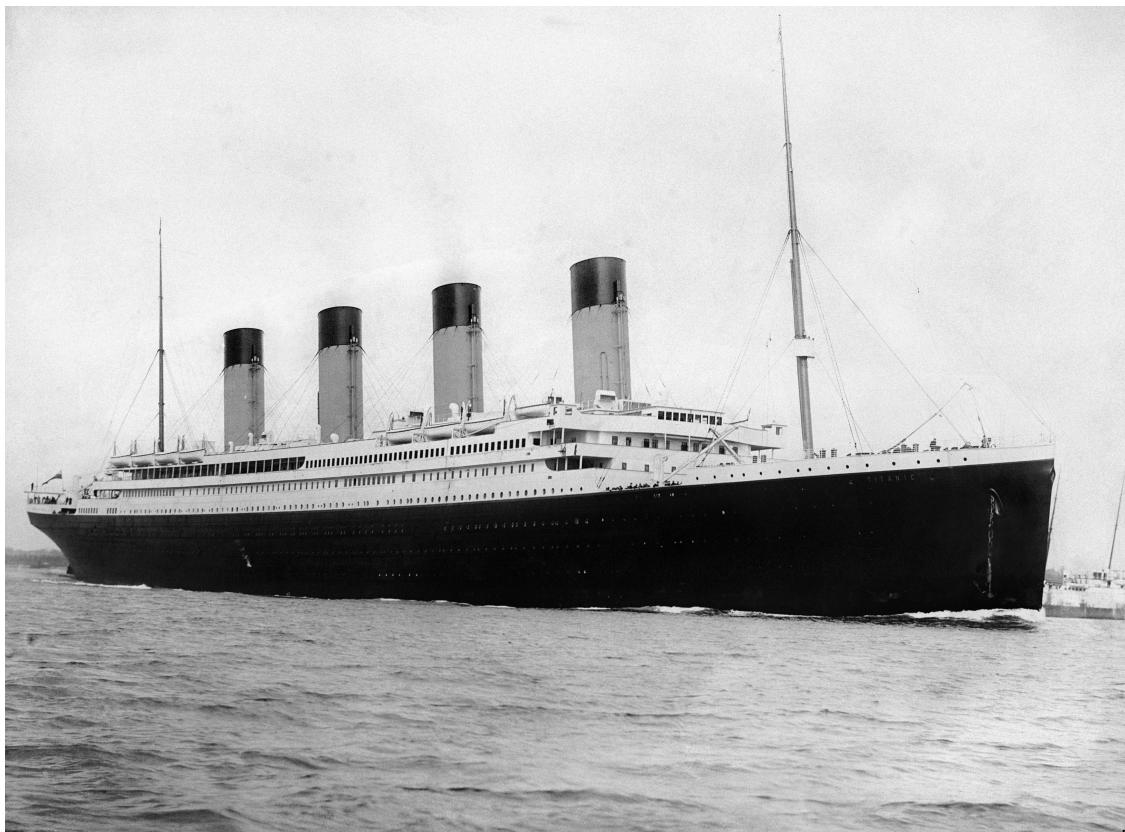


Figure 1: Titanic Ship

```
library(readr);
Titanic <- read_csv("titanic.csv")
```

```
## Rows: 891 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (5): Name, Sex, Ticket, Cabin, Embarked
## dbl (7): PassengerId, Survived, Pclass, Age, SibSp, Parch, Fare
##
```

```
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

References:

1. **Titanic Picture**
2. **SEIS-631 HW-3**
3. **Dataset** [https://github.com/elisabetta42/dataset_analysis/blob/master/titanic3.csv]
4. [<https://ggplot2.tidyverse.org>]
(<https://github.com/Adam-1792/631-rtopics>)
https://github.com/Adam-1792/Adnan_Suri--631-Final-Project/blob/main/.gitignore

Introduction

WHAT: Since I watched the Titanic movie in 1997, I have always wanted to learn more statistical status of the Titanic ship. I hypothesize that more males died vs. females during the event. Also, I would like to know the survival age range and perform further statistical analysis to compare the data, like the overall survival rate.

WHY: It is unfortunate that the Titanic ship sank and people died. People created a dataset to understand statistics and mainly apply those statistics for people's safety and a better world, which is very similar to the Covid dataset. Because of my interest in Titanic, I will use R language to learn more about titanic data.

HOW: My goal would be to visualize the data by using tidyverse for titanic data manipulation and visualization. I will be doing the data analysis by using the following:

MINIMAL: My goal would be to get more familiar with R, import titanic dataset into R, runs statistics using R functionality, especially ggplot, tidyvers, visuals where I can run the data with graphs, charts formats, and write the report in R Markdown and export in PDF.

AMBITIOUS: My next stretch goal would be to add more visuals and importing the dataset into Power BI (<https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-r-scripts>). Installing R script into BI desktop and create "Titanic Dashboard" to run different visuals with slicer option.

Topics From Class

Topic 1: Lecture 8: Tidy Data a.k.a Tidyverse approach to R.

```
## 1) Use R functionality to analyze the data using ggplot 2 (data visualization)
## 2) R functionality dplyr (data manipulation)
## 3) R functionality tidyr (data cleaning)
```

Topic 2: Chapter 2, Summarizing the data.

```
## 1) Examine the numerical data using scatterplot, histogram, standard deviation, boxplot, quaterlies
## 2) Analyze data by considering categorical data;
# tables and barplots - class Hw-3
```

Topic 3: Distribution of random variables try with R

```
## 1) Normal distribution - plot(x, qnorm(x))
## 2) Geometric distribution - plot(x, gnorm(x))
## 3) Binomial distribution - plot(x, bnorm(x))
## 4) Negative binomial distribution - plot(x, pnorm(x))
```

Topic 4: Introduction to linear regression try with R

```
## 1) Understand Titanic data and define the correlation  
## 2) Draw different residual plots and analyse the data
```

Topic 5: R Shinny app

```
## 1) If time permits than will create a dashboard using R Shinny app  
## 2) Analyze the data and visualization
```

```
*****
```

Topic 2: Chapter 2, Summarizing the data There are 891 observations and 12 variables. It is a tibble: 891 x 12 with 7 - hold the numeric values with decimal points and 5

```
titanic <- read.table('titanic.csv', header = TRUE, sep = ",")  
summary(titanic)
```

```
##   PassengerId      Survived      Pclass       Name  
##   Min. : 1.0      Min. :0.0000  Min. :1.000  Length:891  
##   1st Qu.:223.5    1st Qu.:0.0000  1st Qu.:2.000  Class :character  
##   Median :446.0    Median :0.0000  Median :3.000  Mode  :character  
##   Mean   :446.0    Mean   :0.3838  Mean   :2.309  
##   3rd Qu.:668.5    3rd Qu.:1.0000  3rd Qu.:3.000  
##   Max.  :891.0    Max.  :1.0000  Max.  :3.000  
##  
##           Sex          Age      SibSp      Parch  
##   Length:891      Min. : 0.42  Min. :0.000  Min. :0.0000  
##   Class :character  1st Qu.:20.12  1st Qu.:0.000  1st Qu.:0.0000  
##   Mode  :character  Median :28.00  Median :0.000  Median :0.0000  
##  
##           Mean   :29.70  Mean   :0.523  Mean   :0.3816  
##  
##           3rd Qu.:38.00  3rd Qu.:1.000  3rd Qu.:0.0000  
##  
##           Max.  :80.00  Max.  :8.000  Max.  :6.0000  
##  
##           NA's   :177  
##           Ticket        Fare      Cabin      Embarked  
##   Length:891      Min. : 0.00  Length:891      Length:891  
##   Class :character  1st Qu.: 7.91  Class :character  Class :character  
##   Mode  :character  Median :14.45  Mode  :character  Mode  :character  
##  
##           Mean   :32.20  
##  
##           3rd Qu.:31.00  
##  
##           Max.  :512.33  
##
```

```
dim(titanic)
```

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

```
length(dim(titanic))
```

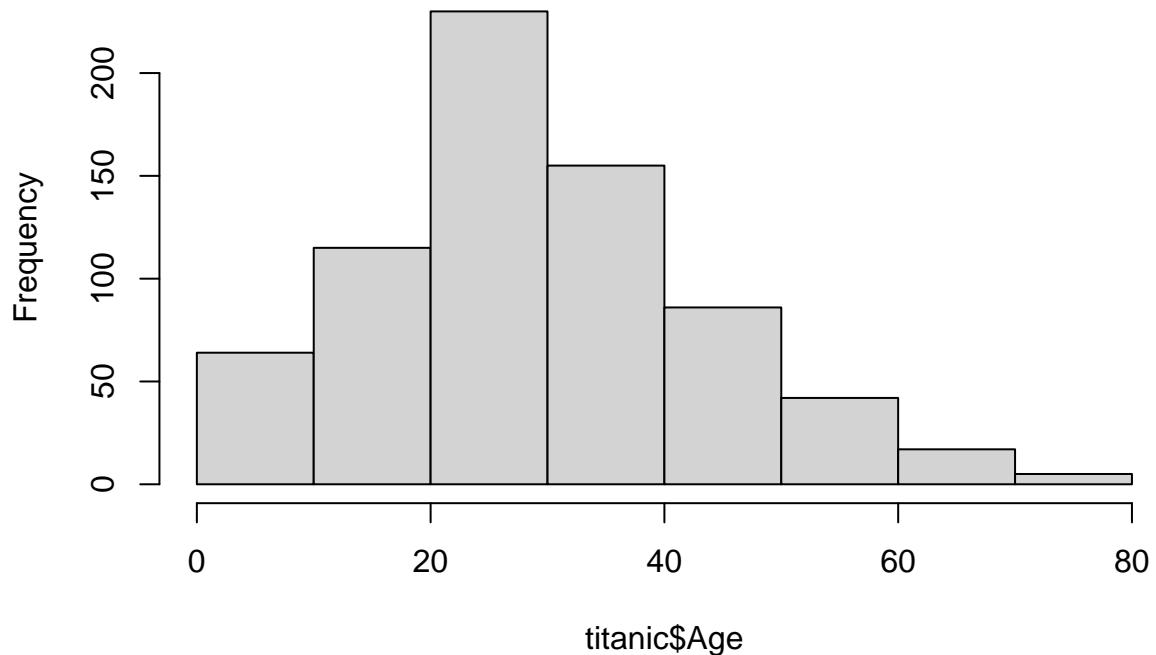
```
## [1] 2
```

```
names(titanic)

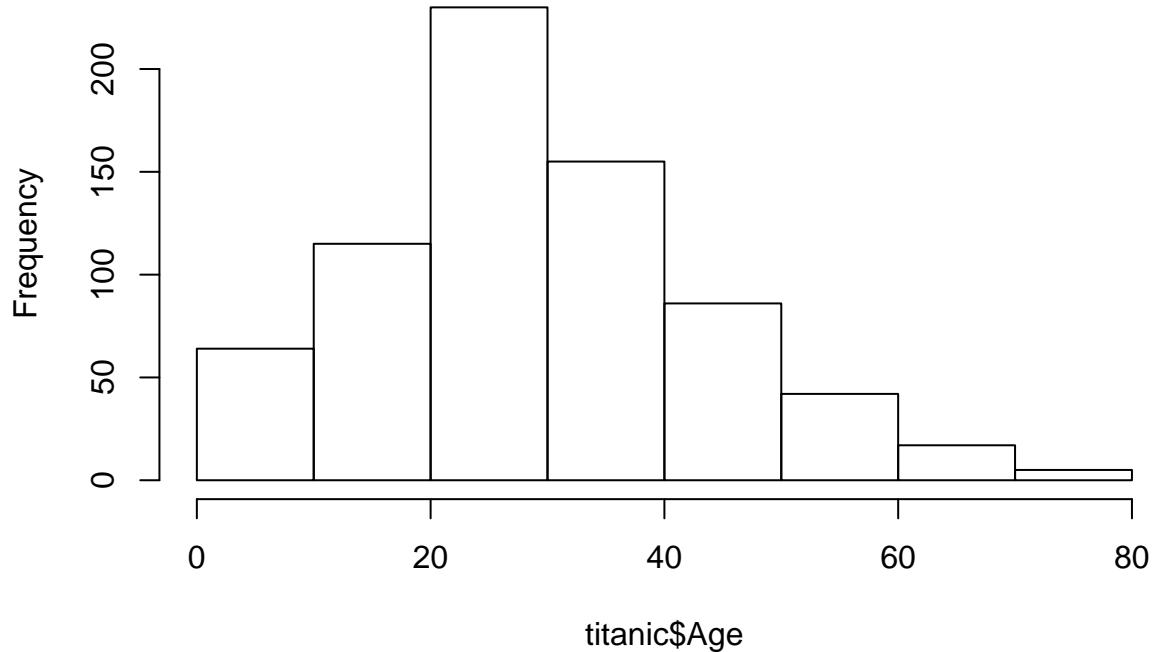
## [1] "PassengerId"  "Survived"      "Pclass"       "Name"        "Sex"
## [6] "Age"          "SibSp"        "Parch"       "Ticket"      "Fare"
## [11] "Cabin"        "Embarked"

plot(hist(titanic$Age))
```

Histogram of titanic\$Age



Histogram of titanic\$Age



```
head(titanic)
```

```
##   PassengerId Survived Pclass
## 1           1       0     3
## 2           2       1     1
## 3           3       1     3
## 4           4       1     1
## 5           5       0     3
## 6           6       0     3
##
##                                     Name      Sex Age SibSp Parch
## 1 Braund, Mr. Owen Harris    male    22    1    0
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38    1    0
## 3 Heikkinen, Miss. Laina    female  26    0    0
## 4 Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35    1    0
## 5 Allen, Mr. William Henry   male   35    0    0
## 6 Moran, Mr. James        male   NA    0    0
##
##            Ticket      Fare Cabin Embarked
## 1 A/5 21171 7.2500          S
## 2 PC 17599 71.2833         C85
## 3 STON/O2. 3101282 7.9250          S
## 4 113803 53.1000        C123
## 5 373450 8.0500          S
## 6 330877 8.4583          Q
```

```
tail(titanic)
```

```
##   PassengerId Survived Pclass          Name      Sex
## 886         886       0     3 Rice, Mrs. William (Margaret Norton) female
## 887         887       0     2 Montvila, Rev. Juozas male
## 888         888       1     1 Graham, Miss. Margaret Edith female
## 889         889       0     3 Johnston, Miss. Catherine Helen "Carrie" female
## 890         890       1     1 Behr, Mr. Karl Howell male
## 891         891       0     3 Dooley, Mr. Patrick male
##   Age SibSp Parch     Ticket   Fare Cabin Embarked
## 886 39     0    5 382652 29.125      Q
## 887 27     0    0 211536 13.000      S
## 888 19     0    0 112053 30.000     B42      S
## 889 NA     1    2 W./C. 6607 23.450      S
## 890 26     0    0 111369 30.000     C148      C
## 891 32     0    0 370376  7.750      Q
```

```
summary.data.frame(titanic)
```

```
##   PassengerId      Survived      Pclass          Name
## Min.   : 1.0   Min.   :0.0000   Min.   :1.000   Length:891
## 1st Qu.:223.5 1st Qu.:0.0000  1st Qu.:2.000   Class  :character
## Median :446.0  Median :0.0000  Median :3.000   Mode   :character
## Mean   :446.0  Mean   :0.3838  Mean   :2.309
## 3rd Qu.:668.5 3rd Qu.:1.0000  3rd Qu.:3.000
## Max.   :891.0  Max.   :1.0000  Max.   :3.000
##
##   Sex            Age        SibSp      Parch
## Length:891      Min.   : 0.42  Min.   :0.0000  Min.   :0.0000
## Class  :character 1st Qu.:20.12  1st Qu.:0.0000  1st Qu.:0.0000
## Mode   :character Median :28.00  Median :0.0000  Median :0.0000
##                   Mean   :29.70  Mean   :0.523   Mean   :0.3816
##                   3rd Qu.:38.00  3rd Qu.:1.000   3rd Qu.:0.0000
##                   Max.   :80.00  Max.   :8.000   Max.   :6.0000
##                   NA's   :177
##
##   Ticket        Fare        Cabin      Embarked
## Length:891      Min.   : 0.00  Length:891      Length:891
## Class  :character 1st Qu.: 7.91  Class  :character  Class  :character
## Mode   :character Median :14.45  Mode   :character  Mode   :character
##                   Mean   :32.20
##                   3rd Qu.:31.00
##                   Max.   :512.33
##
```

```
summary(titanic$Age)
```

```
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
##   0.42  20.12  28.00  29.70  38.00  80.00  177
```

```
table(titanic$Survived)
```

```

##          0      1
## 549 342

table(titanic$Sex)

##          female    male
##        314     577

table(titanic$Pclass)

##          1      2      3
## 216 184 491

table(titanic$Pclass) ->pclass
prop.table(pclass)*100

##          1      2      3
## 24.24242 20.65095 55.10662

round(pclass, digits=1) ->pclass
pclass

##          1      2      3
## 216 184 491

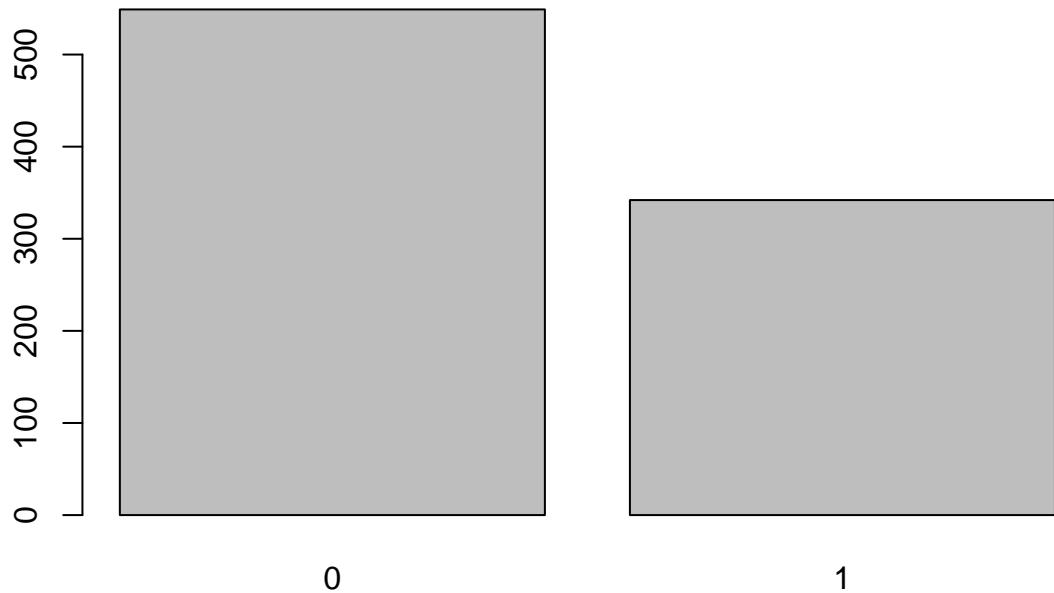
gender_survived <- table(titanic$Sex, titanic$Survived)
mosaicplot(gender_survived)

```

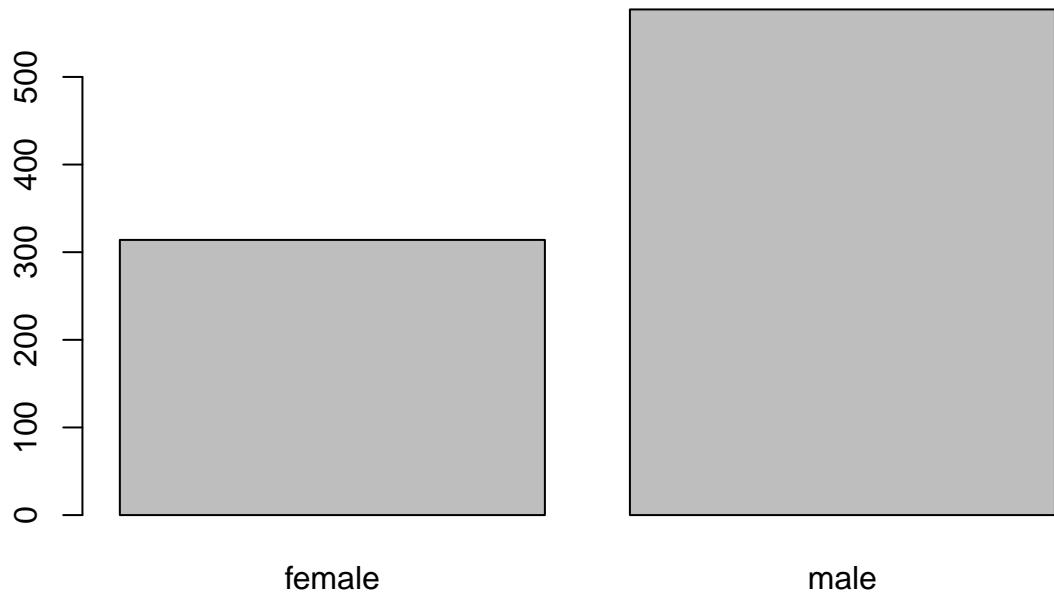
gender_survived



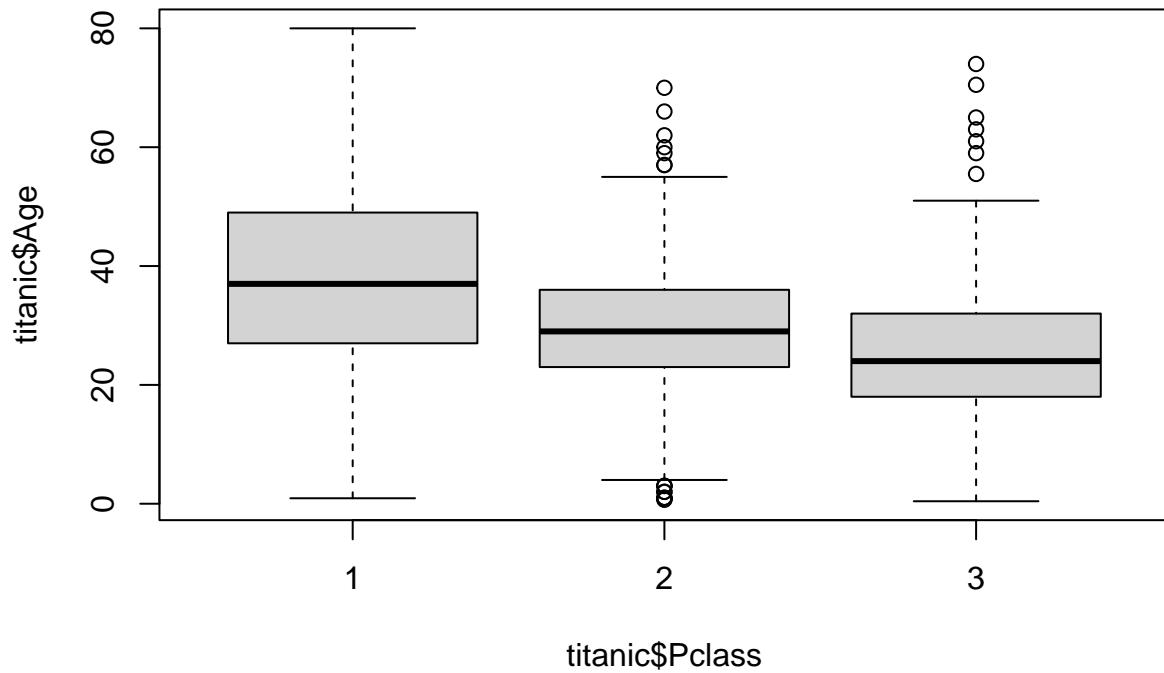
```
barplot(table(titanic$Survived))
```



```
barplot(table(titanic$Sex))
```

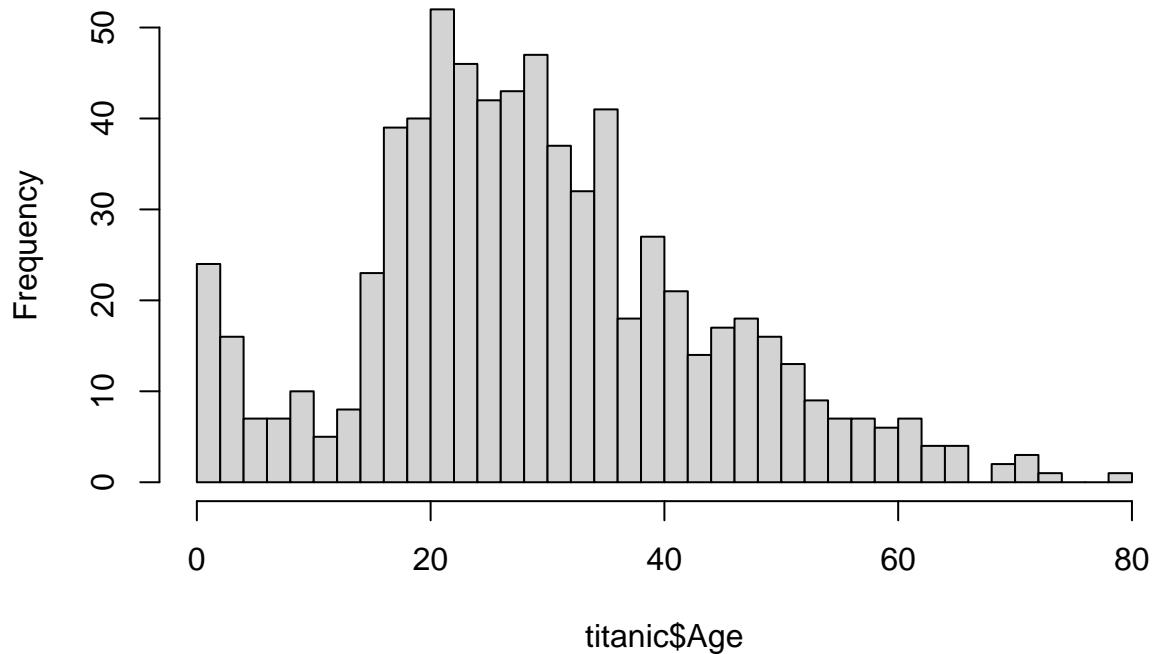


```
boxplot(titanic$Age ~ titanic$Pclass)
```



```
hist(titanic$Age, breaks = 50)
```

Histogram of titanic\$Age



Conclusion