Lab Assignment #3 (Due: 11:00am, Wed, Sep 4)

If you cannot finish it by 11am, please talk to me. I'll give you an extension. No Penalty!!

Write Your Name Here

<< Attention >> You can use this file for our lab assignment. Please edit as you wish. (You should delete unnecessary parts and add your work.)

1. The Titanic (Data: Titanic)

```
> attach(Titanic)
```

(a) Find the frequencies.

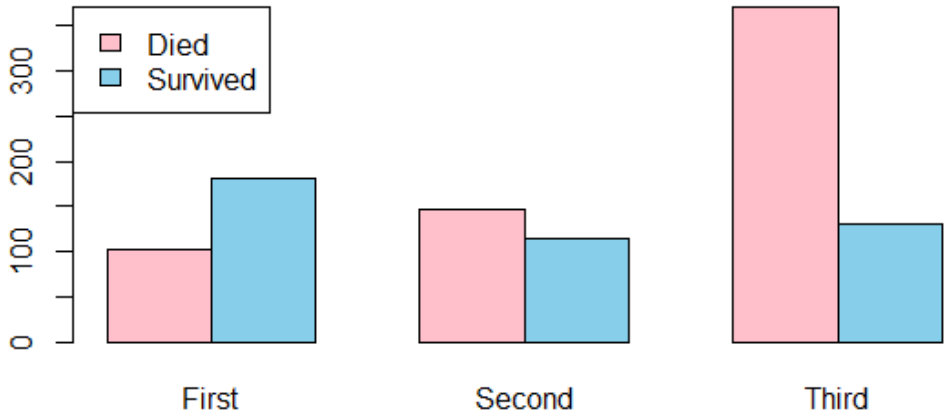
```
> table (Status, Class)
          Class
Status     First Second Third
  Died       103    146   370
  Survived   181    115   131
```

- Complete the table.

		Ticket Class		
		First Class	**Second Class**	**Third Class**
Status	**Died**	**103**	**146**	**370**
	Survived	**181**	**115**	**131**

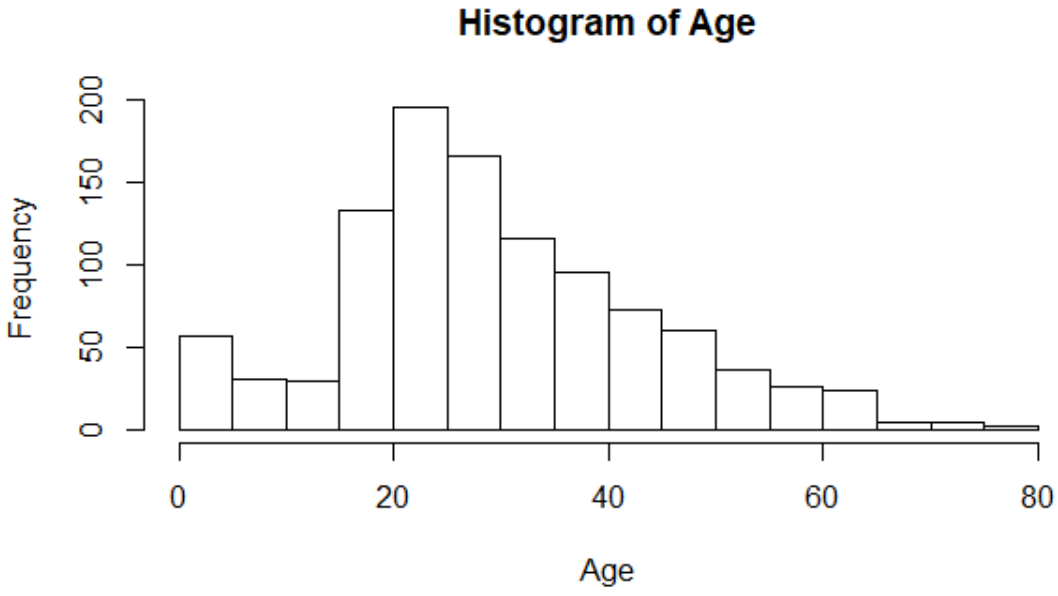
(b) Let's compare the death/survival rates of passengers in each ticket class (i.e. relationship between the passenger's survival status and the passenger's ticket class).

```
> barplot(table(Status, Class), beside=TRUE, col=c("pink", "skyblue"))
> legend("topleft", legend=c("Died", "Survived"), fill=c("pink", "skyblue"))
```



(c) Let's take a look at the distribution of passengers' ages using a histogram.

```
> hist(Age)
```



(d) Calculate the mean age and the median age of the passengers. Also, find the age of the youngest and the oldest passengers. **Note:** See "R Handout1" for help.

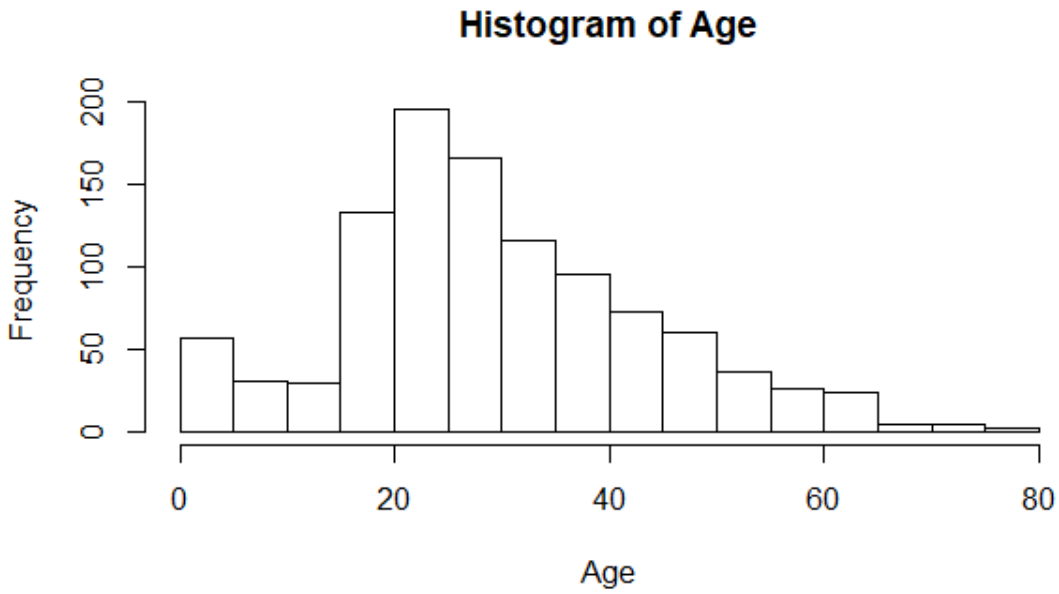
- Paste the R command and the results here inside a text box. **Hint:** You can type `summary(Age)` to calculate the sample statistics (5-number summary and mean).

```
> summary (Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1667 21.0000 28.0000 29.8811 39.0000 80.0000
```

Another way

```
> mean(Age)
[1] 29.88113
> median(Age)
[1] 28
> min(Age)
[1] 0.1667
> max(Age)
[1] 80
```

(e) Answer it based on the results in (c) and (d). Write sentences.



- Describe the distributional shape: Symmetric, Right-Skewed, or Left-Skewed.

  The shape is Right-Skewed.

- Are there more young passengers or older passengers?

  Yes, there are. The shape is Right-Skewed which show that the data concentrate to the left more than the right. The left side of the graph represent younger age passengers.

- What are the ages of the youngest and the oldest passengers?

  The youngest passenger is 2 years old.

  The oldest passenger is 80 years old.

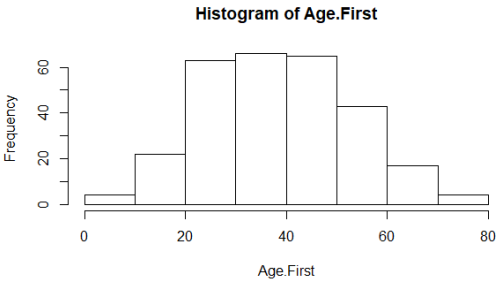
- What is the typical age of passengers?

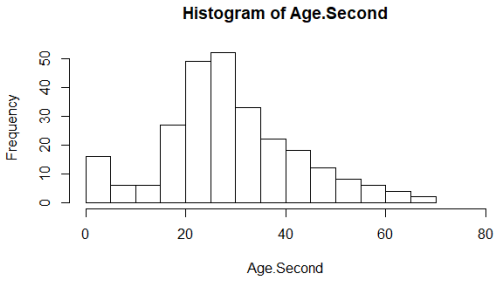
  The typical age of passengers is 28 years old

(f) Let's compare the typical age of passengers in each ticket class using histograms.

```
> Age.First = Age[ Class=="First" ]
> Age.Second = Age[ Class=="Second" ]
> Age.Third = Age[ Class=="Third" ]

> hist(Age.First, xlim=c(0, 80))
> hist(Age.Second, xlim=c(0, 80))
> hist(Age.Third, xlim=c(0, 80))
```





g) Let's compare the typical age of passengers in each ticket class.

- Calculate the five-number summary in each ticket class. Type the following

.

```
> tapply(Age, Class, summary)
```

- Copy the R commands and the result inside a text box.

```
> tapply(Age, Class, summary)
$`First`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.9167 28.0000 39.0000 39.1599 50.0000 80.0000

$Second
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6667 22.0000 29.0000 29.5067 36.0000 70.0000

$Third
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1667 18.0000 24.0000 24.8164 32.0000 74.0000
```

(h) Answer it based on the results in (f) and (g). Write sentences.

- Describe the distributional shape (Symmetric, Right-Skewed, or Left-Skewed) of each histogram.

  The histogram of Age and First class ticket is Right-Skewed
  The histogram of Age and Second class ticket is Right-Skewed
  The histogram of Age and Third class ticket is Right-Skewed

- What is the typical age of passengers in each ticket class? Do you think that passengers of a particular class are older (or younger) than passengers in the other classes? Explain why or why not.
  Note: We'll learn whether there is a significance difference in ages in the later chapter.

  The typical age of passengers in First class ticket is 39 years old
  The typical age of passengers in Second class ticket is 29 years old
  The typical age of passengers in Third class ticket is 24 years old

Looking at typical age of passenger in all three classes, we can see the age of third class ticket is younger than the second. The second class is younger than the first class. This means data of third class ticket concentrate around 24 years old while second class concentrate around 29 years old and first class concentrate around 39 years old. So I can conclude the passenger who buy third class ticket is the younger than the passenger who buy second class ticket, and the passengers who buy first class ticket is the oldest.

(i) Let's compare the typical ages of passengers who survived and passengers who died.

- Copy the R commands and the result inside a text box.
- What is the typical age of passengers who died? What is the typical age of passengers who survived? Do you think that survived passengers are younger or older than the passengers who died? Explain why or why not.

```
> tapply(Age, Status, summary)
$`Died`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3333 21.0000 28.0000 30.5454 39.0000 74.0000

$Survived
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1667 20.0000 28.0000 28.9182 38.0000 80.0000
```

Finish Titanic file

```
> detach (Titanic)
```

(a) We have two categorical variables. Find the frequencies.

- Paste the R command and the result inside a text box.
- Complete the table.

```
> attach(Commercial)
> table(Gender, Identify)
        Identify
Gender    No Yes
  Female 114  41
  Male    50  95
```

		Brand Awareness	
		Can Identify	**Cannot Identify**
Gender	**Female**	**41**	**114**
	Male	**95**	**50**

(b) Let's investigate the relationship between the gender of a viewer and the viewer's brand awareness.

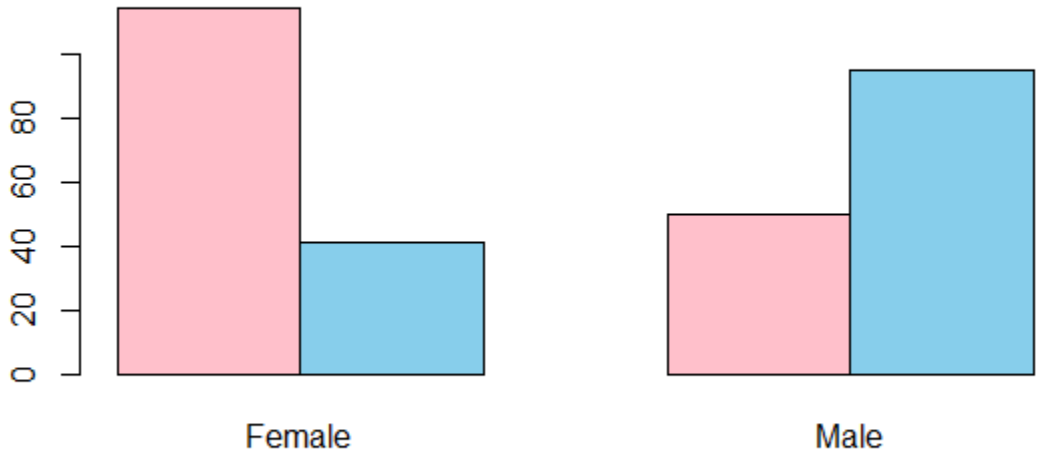
- Draw a side-by-side bar plot. The category names of gender (Male and Female) should appear in the horizontal axis.
- Add colors in the plot.
- Add a legend in the plot.
- Paste the R commands inside a text box.Paste the side-by-side bar plot here.

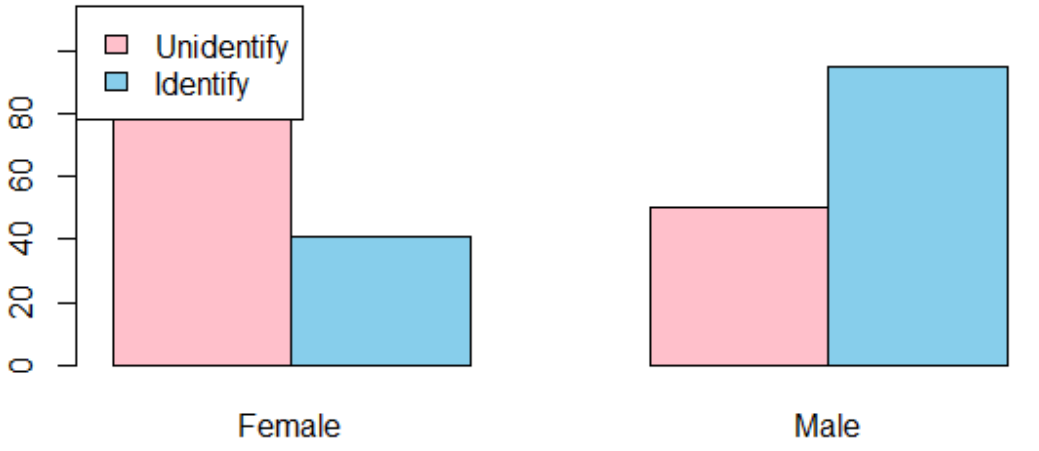
```
> barplot(table(Identify, Gender), beside=TRUE, col=c("pink", "skyblue"))
> legend("topleft", legend=c("Unidentify", "Identify"), fill=c("pink", "skyblue"))
```

The legend cover a part of the graph so I copy both graph here:

The top graph does not have legend so we can see the high

1 graph has legend.





(c) Discuss the relationship between the gender of a viewer and the viewer's brand awareness using the results in (a) and (b).

From the result in (a) and (b) I can see:

Number of men identify the brand name is 95/145 = 65.5%

Number of women identify the brand name is 41/155 = 65.5%

In conclusion, I can say both male and female have the same rate of awareness brand which is 65.5%. This number show that many of them do watch television advertisement (65%), however, there are another part of them (35%) cannot identify the brand.