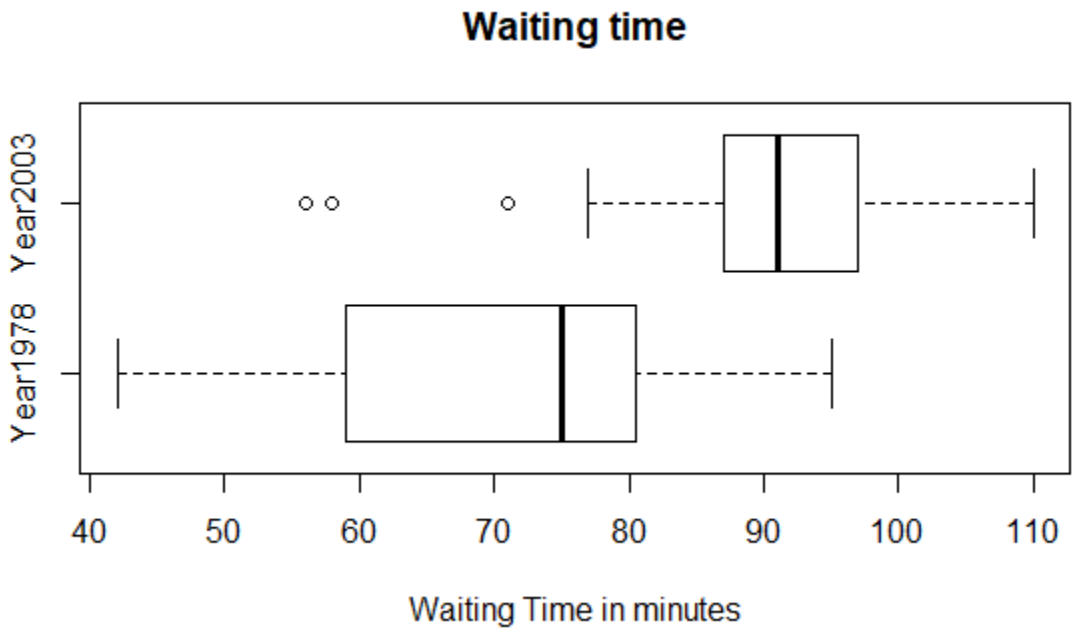
Lab Assignment #4 (Due: 11:00am, Thu, Sep 6)

Write Your Name Here

1. Old Faithful Geyser Eruptions (**Data: OldFaithful**)

(a) Let's compare the waiting times (i.e. inter-eruption times) between year 1978 (before the Coso Earthquake) and year 2003 (after the Coso Earthquake).

```
> attach(OldFaithful)
> boxplot(Minutes~Year, horizontal=T, main="Waiting time", xlab="Waiting Time in minutes" )
```



(b) Calculate the mean waiting time, the five-number summary of waiting time times for each year so that we can compare them.

```
> tapply(Minutes, Year, summary)
$`Year1978`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   42.0    59.0    75.0    71.0    80.5    95.0

$Year2003
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  56.00   87.00   91.00   91.19   97.00  110.00
```

(c) Scientists have investigated whether a 1998 earthquake lengthened the time between eruptions at Old Faithful. Do you think that waiting times (inter-eruption times) tend to be longer or shorter in one year than the other year? Explain why or why not using the results in previous parts.

The median in 1978 is 75.0 less than 91.0 which is the median in 2003

Therefore, the waiting time in 1978 is shorter than waiting time in 2003.

(d) Calculate the standard deviation and the IQR of waiting times for each year so that we can compare them**.**

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

```
> tapply(Minutes, Year, IQR)
Year1978 Year2003
    21.5     10.0
```

(e) Park rangers and visitors would also appreciate if Old Faithful was as reliable as its name implies. Do you think waiting times tend to be more consistent in one year than the other year? Explain why or why not using the results in previous parts.

From part (d), we already know IQR in 1978 larger than IQR in 2003. This show the waiting time shrink more in 2013. Therefore, the waiting time in 2003 tend to be more consistent than the waiting time in 1978.

I'm done with OldFaithful file

```
> detach(OldFaithful)
```

2. The Titanic (Data: Titanic)

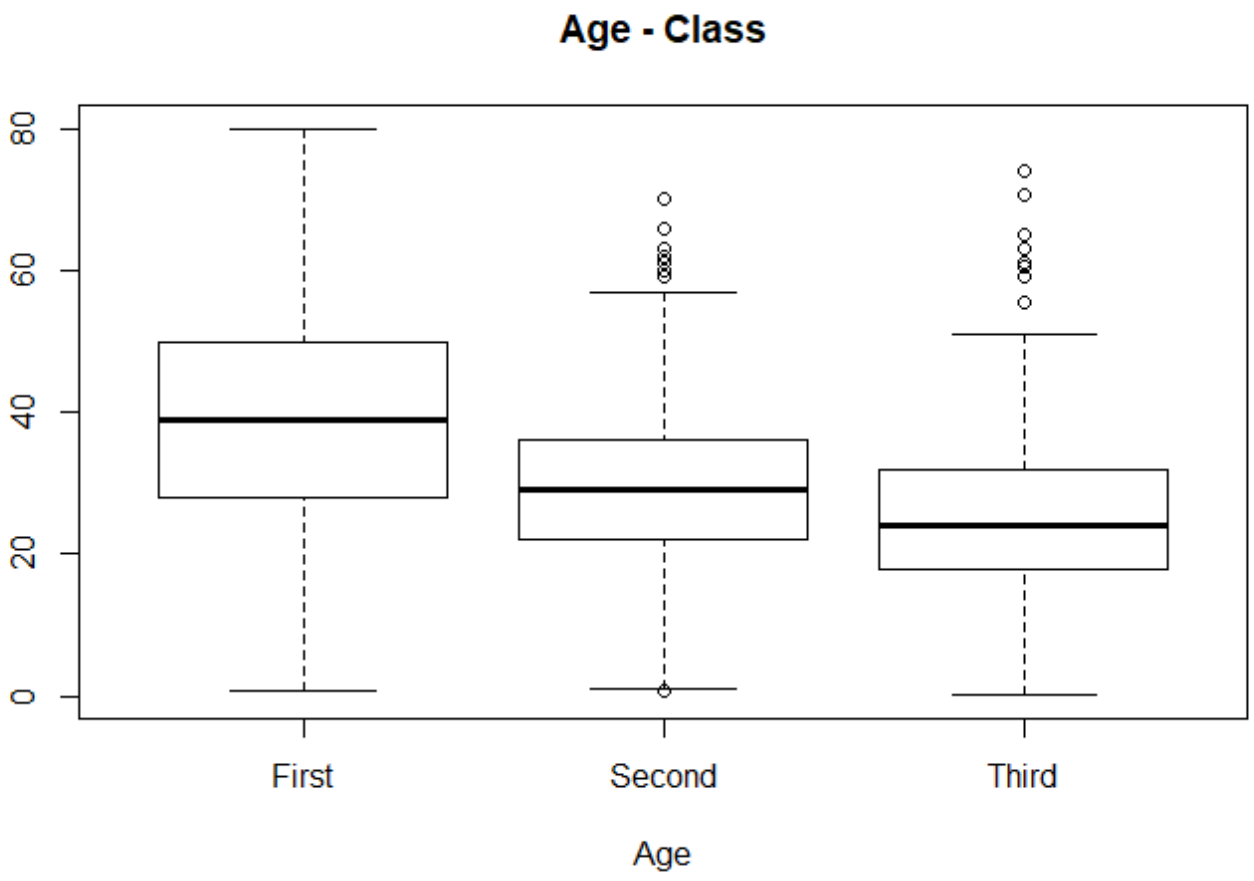
On April 15, 1912, on her maiden voyage, the Titanic collides with an iceberg and sank. The ship was luxurious but did not have enough lifeboats for the 2,224 passengers and crew. As a result of the collision, 1,502 people died. The ship had three classes of passengers. The level of luxury and the price of the ticket varied with the class, with first class being the most luxurious.

- First, import the data file "**Titanic.csv**" into RStudio.
- Then, attach the data file

```
> attach(Titanic)
```

(a) We compared the ages passengers among the First, Second, and Third classes. Draw a (side-by-side) boxplot so that we can compare the ages of three groups.

```
> boxplot(Age~Class, main="Age - Class", xlab="Age" )
```



(b) We can also calculate the median ages (or any statistics) of passengers for all combinations of ticket class and gender.

- Calculate the median age for each group. Paste the R command inside a text box.

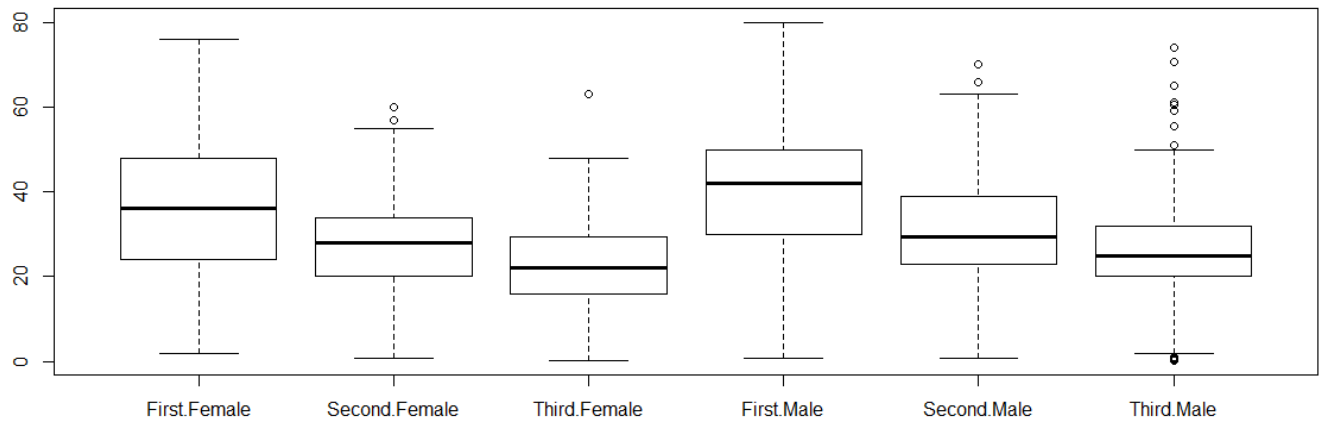
```
> tapply(Age, Class:Gender, median)
 First:Female    First:Male Second:Female   Second:Male  Third:Female    Third:Male
         36.0          42.0          28.0          29.5          22.0          25.0
```

- Report the median ages in the table below.

		Ticket Class		
		First Class	**Second Class**	**Third Class**
Gender	**Female**	**36.0**	**28.0**	**22.0**
	Male	**42.0**	**29.5**	**25**

(c) Similarly, draw (side-by-side) boxplots for ages for all combinations of ticket class and gender (six groups).

```
> boxplot(Age ~ Class:Gender)
```



(d) Tell me what you see in the plot in part (c). (There is no right or wrong answer,)

If we compare by class, all classes, the male is older than female

If we compare by gender, first class is oldest, second class in middle, third class is youngest

```
> detach(Titanic)
```