

# Tutorials Of R Using Titanic Data

*Josemari Feliciano*

*7/10/2017*

## Tutorial Notes:

The Beauty of R Studio is the ability to have notebooks for data analysis. It allows us to run R-code inline similar to IPython Notebooks.

Think of it as a fancy chemistry notebook where you can run the the experiments in the notebook itself – but for data analysis! For this notebook, I will use the titanic data for analysis. This assumes that you have R and RStudio loaded.

**Note** : mosaic package is required for this if you want to run data yourself and not to simply view them. If you do not have this, run this code in your r-console independently:

```
install.packages("ggplot2")
```

## Getting started:

We probably want to load the titanic data first. We probably want to load the mosaic library out of the way. So we can accomplish both with the following script:

```
load("Titanic.Rdata")
library("mosaic")
```

Now that the data has been loaded, we probably want to see which data variables we will deal with! Below, we will use `names()` to print the variable names in our data to have them handy.

```
names(Titanic)
```

```
## [1] "Gender" "Age" "Name" "Fare" "Class" "Survived"
```

The `names()` function we just ran displayed the 6 variables within the Titanic Data which include Gender, Age, Name, Fare, Class and Survived.

Of course, you could have looked at the actual CSV file or RData file directly. But functions like `names()` are very useful when wrangling data such as JSON and similar data types.

```
typeof(Titanic)
```

```
## [1] "list"
```

```
sapply(Titanic,class)
```

```
##   Gender      Age      Name      Fare      Class  Survived
## "factor" "numeric" "factor" "numeric" "factor" "factor"
```

```
bargraph( ~ Class, data =Titanic)
```

