

# Investigating the effect of Class on Surviving the Titanic

Protim Ganguly

18 September 2018

## Question of Interest

Was the socioeconomic Class of passengers on the Titanic a predictor of whether a passenger survived or not ?

Yes the social economic class of passengers on the Titanic was a predictor for the chances of survival of a passenger.

## Load the libraries needed.

The libraries needed are 'tidyverse' and 'titanic'. If you are running this code on your own pc you will have to install them, if you are running this in a lab they are already installed.

Load up the libraries as follows:

```
library(tidyverse)

## -- Attaching packages -----
## ----- tidyverse 1.2.1 --

## v ggplot2 3.0.0      v purrr  0.2.5
## v tibble  1.4.2      v dplyr  0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts -----
## ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(titanic)
```

The data are contained in a package called 'titanic'. The dataset itself is called `titanic_train`.

## Subjective Impressions

The key variable of interest 'Survival' (representing whether a passenger survived or not) is coded as 0 and 1. To make the analysis clearer to interpret create a new variable which recoded the 0 and 1 to No and Yes respectively.

```
passengers <- titanic_train %>%
  mutate(Survived = ifelse(Survived == 0, "No", "Yes"))
```

There are three levels of the (categorical) variable Pclass namely 1st, 2nd and 3rd (coded as 1,2,3).

A table of the proportion of survivors by Class (with clearer labels) is as follows:

Task: Create the table of summaries needed by inserting the relevant r chunk

```
passengers %>% select(Pclass, Survived) %>% table()

##           Survived
## Pclass   No Yes
##      1    80 136
##      2    97  87
##      3   372 119
```

Task: Create the table of corresponding percentages by inserting the relevant r chunk

```
passengers %>% select(Pclass, Survived) %>% table() %>% prop.table(margin = 1)

##           Survived
## Pclass           No           Yes
##      1 0.3703704 0.6296296
##      2 0.5271739 0.4728261
##      3 0.7576375 0.2423625
```

Task: add some text here to interpret the results from the tables you have just created.

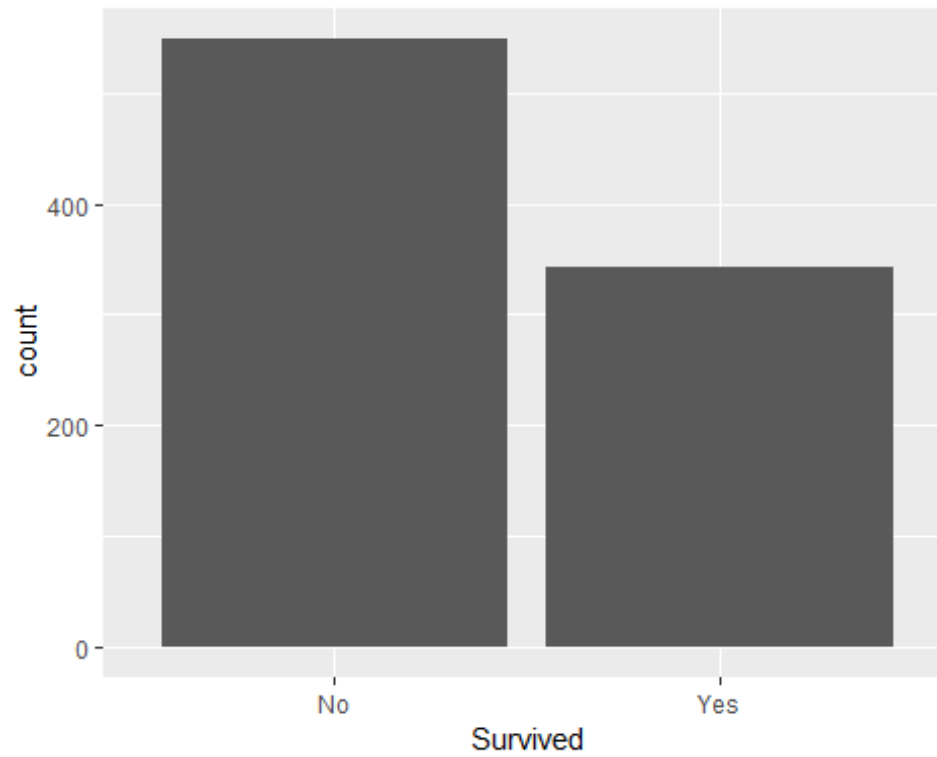
From the Frequency table we can see that around 136 people out of 216 had survived from Class 1, 87 people out of 184 from Class 2, 119 people out of 491 from Class 3. The general trend we can see is that the survival chances were highest from Class 1 and decreased as we went down the order of the class.

Time to create some barcharts. Some of the code will be given, some you will have to copy from the example file given and adapt accordingly. Hint, look at the plots relating to gender and adapt them by replacing Gender with Class (i.e. Pclass variable).

## Bar chart of survival overall

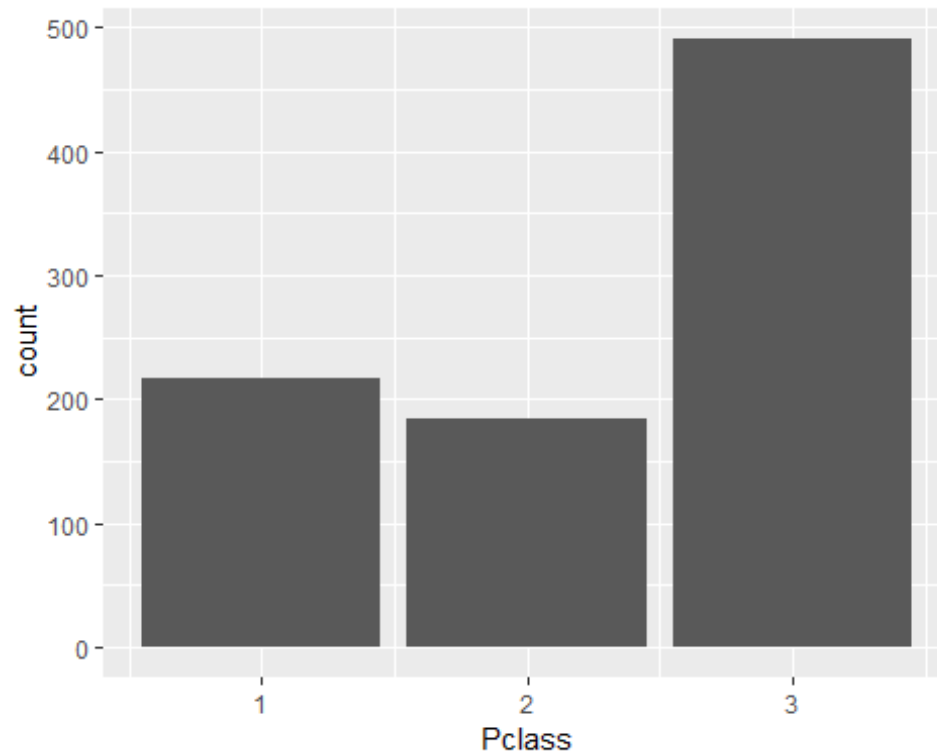
Task: Create a bar chart of survival overall by inserting the relevant r chunk

```
ggplot(passengers, aes(x = Survived)) +
  geom_bar()
```



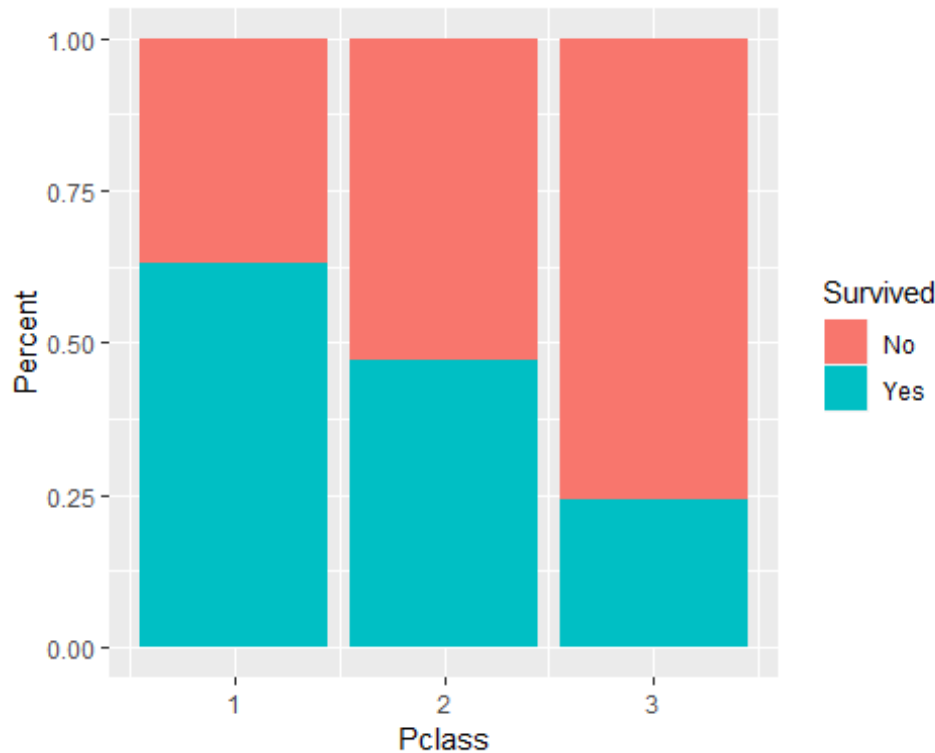
### Plot bar chart of survival by Class

```
ggplot(passengers, aes(x = Pclass)) +  
  geom_bar()
```



Task: Create a stacked barchart of survival by Class by inserting the relevant r chunk.

```
ggplot(data=passengers, aes(Pclass))+  
  geom_bar(aes(fill=Survived), position="fill") +  
  ylab("Percent")
```



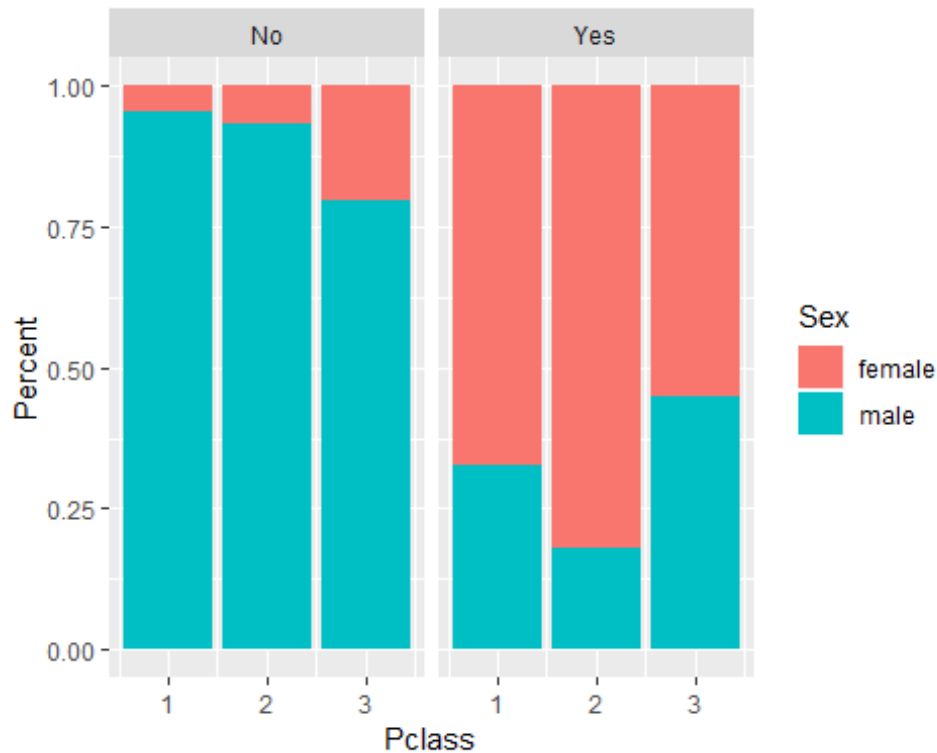
Task: Once you have created the plot write your interpretation here based on these plots.

The normal barchart of survival by Class doesn't give an accurate picture of the actual dataset. However the stacked bar chart provides a much more better visual representation of the actual dataset where we can see that majority of the people from Class 1 survived followed by people from Class 2 and then the least proportion of people from class 3 survived compared to all the other classes.

Bonus question. How would you create a stacked bar chart of survival by Class and Gender ? (Hint `facet_grid()` will be useful).

Task: Create the faceted barchart by inserting the relevant r chunk.

```
passengers %>% drop_na() %>%  
  ggplot(aes(x = Pclass)) +  
  geom_bar(aes(fill=Sex), position="fill") +  
  ylab("Percent") +  
  facet_grid(~Survived)
```



## Conclusion

Task: Write a short conclusion of whether you think Class is a useful predictor of whether a person survived the titanic and the role Gender plays in addition. Knit the file as a Word document (using the Knit icon above).

We can see that the total number of people who died in titanic were mostly from Class 3 and then Class 2 followed by class 1. This shows that yes class was a useful predictor in the survival rate of the people travelling in Titanic. Also we can see that most of the passengers who survived were from class 3 strangely. But we can see a greater variation in survival rate gender wise. Gender Wise a greater number of Males died from Class 1 strangely enough(since Class 1 paid more ticket Fare) decreasing towards Class 3. Among the people who survived even though more females survived than males but it was roughly spread out across all classes of ticket fare.