Investigating the effect of Class on Surviving the Titanic

Ryan Maguire, id=18464854

21 October, 2019

## Question of Interest

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

Yes of course, the wealthier people in 1st class, had more power and money therefore their lives were valued more than the poor who resided in the lower classes and this is further proved by the data represented in this document.

### 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 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- 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 iteslf 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 %>%  
 group\_by( Pclass, Survived) %>%  
 summarise (n = n()) %>%  
 mutate(freq = n / sum(n))

## # A tibble: 6 x 4  
## # Groups: Pclass [3]  
## Pclass Survived n freq  
## <int> <chr> <int> <dbl>  
## 1 1 No 80 0.370  
## 2 1 Yes 136 0.630  
## 3 2 No 97 0.527  
## 4 2 Yes 87 0.473  
## 5 3 No 372 0.758  
## 6 3 Yes 119 0.242

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

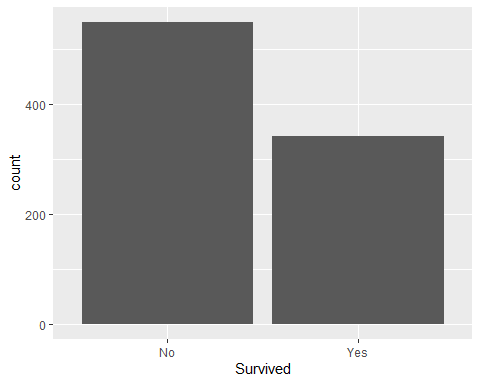
From the table above we can see that 1st class is the only class where more people actually survived than died, with over a 63% survival rate, while in second class this drops to a 47% survival rate and in 3rd class it drops to an abysmal 25% survival rate. This data further shows how your standing in life would have affected your survival chances aboard the Titanic.

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 accoordingly. 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

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")



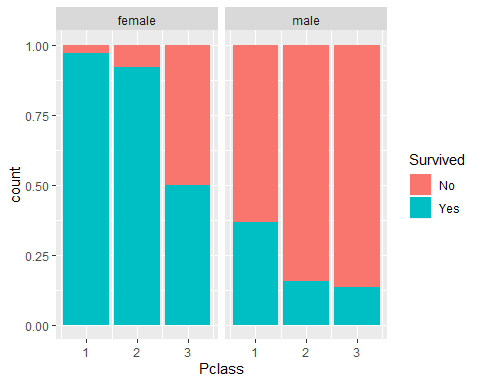
Task: Once you have created the plot write your interpretation here based on these plots.

The first plot clearly shows that the majority of people aboard the Titanic unfortunately lost their lives around 470 died vs around the 270 who survived, our second plot which plots the survival among the classes shows that 3rd class has the lowest with 25% and 1st class has the highest with around 62.5%, this disparity is to be expected due to the way these classes were treated aboard the Titanic with 1st class passengers having the priority of the lifeboats when the ship went down .2nd class resides in between both of these with 50% survival rate which was to be expected.

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.

ggplot(data=passengers, aes(Pclass),(Sex))+  
geom\_bar (aes(fill=Survived), position="fill") +  
facet\_grid(~Sex)



# 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).

After analyzing the data I could conclusively say that gender and class had huge roles in whether you survived or not aboard the titanic, this is shown superbly in the chart above, which shows over 90% of 1st class females survived while only 40% of 1st class males survived and In terms of class only 50% of 3rd class females survived from our data vs the 90% of 1st class females and this is the same story when we look at second class or males. Therefore, class and gender are extremely valid predictors to the survival rates of different groups aboard the Titanic.

Knit your markdown saves the Word file and Turn it in on the course Blackboard page once the deadline date has been set.