Midterm I

Fall 2018

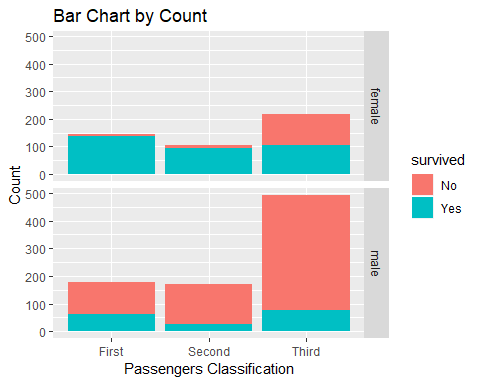
Due: Wednesday, October 17

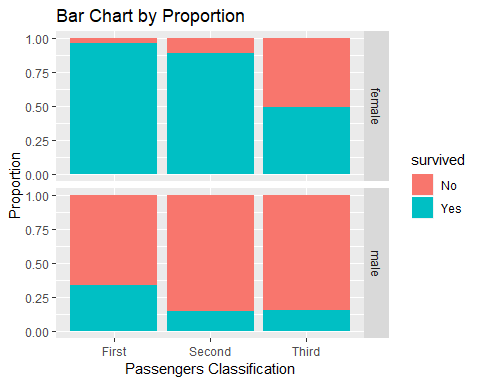
## I. Summary of all the variables in the tibble

## pclass survived name sex age   
## First :323 No :809 Length:1309 female:466 Min. : 0.1667   
## Second:277 Yes:500 Class :character male :843 1st Qu.:21.0000   
## Third :709 Mode :character Median :28.0000   
## Mean :29.8811   
## 3rd Qu.:39.0000   
## Max. :80.0000   
## NA's :263   
## sibsp parch ticket fare   
## Min. :0.0000 Min. :0.000 Length:1309 Min. : 0.000   
## 1st Qu.:0.0000 1st Qu.:0.000 Class :character 1st Qu.: 7.896   
## Median :0.0000 Median :0.000 Mode :character Median : 14.454   
## Mean :0.4989 Mean :0.385 Mean : 33.295   
## 3rd Qu.:1.0000 3rd Qu.:0.000 3rd Qu.: 31.275   
## Max. :8.0000 Max. :9.000 Max. :512.329   
## NA's :1   
## cabin embarked boat body   
## Length:1309 C :270 Length:1309 Min. : 1.0   
## Class :character Q :123 Class :character 1st Qu.: 72.0   
## Mode :character S :914 Mode :character Median :155.0   
## NA's: 2 Mean :160.8   
## 3rd Qu.:256.0   
## Max. :328.0   
## NA's :1188   
## home.dest   
## Length:1309   
## Class :character   
## Mode :character   
##   
##   
##   
##

1. Define each of the following types of variables and give an example from the Titanic tibble:
   * Categorical nominal
   * Categorical ordinal
   * Numerical
2. Explain some of the challenges that exist with missing observations.

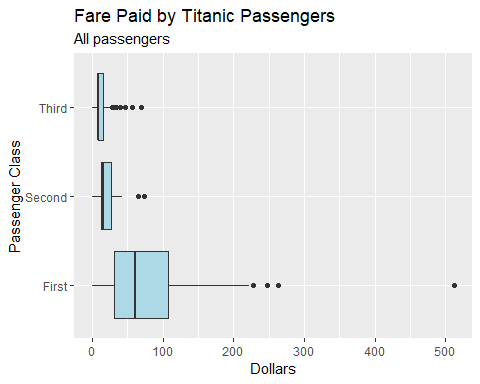
## II. Covariation among passenger class, survival, and gender





1. Why do we use bar charts to look for covariation with these variables?
2. What conclusions do you reach from these graphs that would help us predict who survived?
3. Write the R code that creates these two difference sets of bar charts.

## III. Comparison of distributions of fares by passenger class



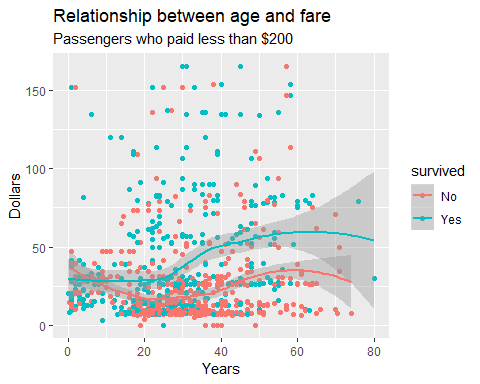
1. Explain the parts of the violin plots.
2. Compare the distributions of the fares of the different passenger classes.
   * Location
   * Scale
   * Symmetry
   * Outliers
3. Write the R code that creates the above boxplot plot.

## IV. Comparison of means and medians

## # A tibble: 3 x 3  
## pclass median mean  
## <ord> <dbl> <dbl>  
## 1 First 60 87.5  
## 2 Second 15.0 21.2  
## 3 Third 8.05 13.3

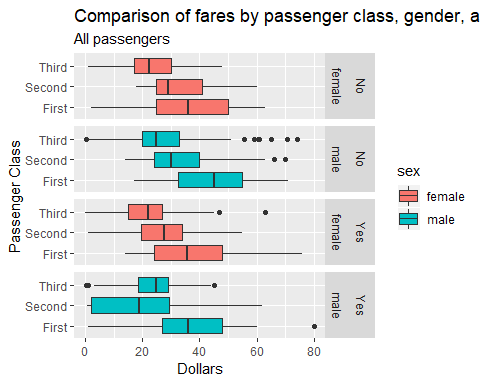
1. Compare the means and the medians of the fares paid by each passenger class.
2. Do you prefer to use the means or the medians? Why?
3. How do missing observations complicate our calculations?
4. Write the code that creates this table.

## V. Covariation between age and fare



1. Describe the relationship between fare and age for those who survived and those that didn’t.
2. Write the R code that will create this diagram.

## VI. Age by passenger class, gender, and survival



1. Find some patterns from the boxplot that would help you predict whether a given passenger survived or didn’t survive.
2. Write the R code that constructs this graph.