

STAT123_Assignment2

Eric Huber

14/02/2022

```
data("Titanic")
NHLDData =read.csv("NHLDData.csv")
?Titanic

## starting httpd help server ... done
```

Question 1:

(1.a) Create (and print out) a table which contains the adult passengers (of all classes and genders) who survived

```
AdultPas = as.table(Titanic[1:4, ,2,2])
AdultPas

##           Sex
## Class  Male Female
##  1st     57    140
##  2nd     14     80
##  3rd     75     76
##  Crew   192     20
```

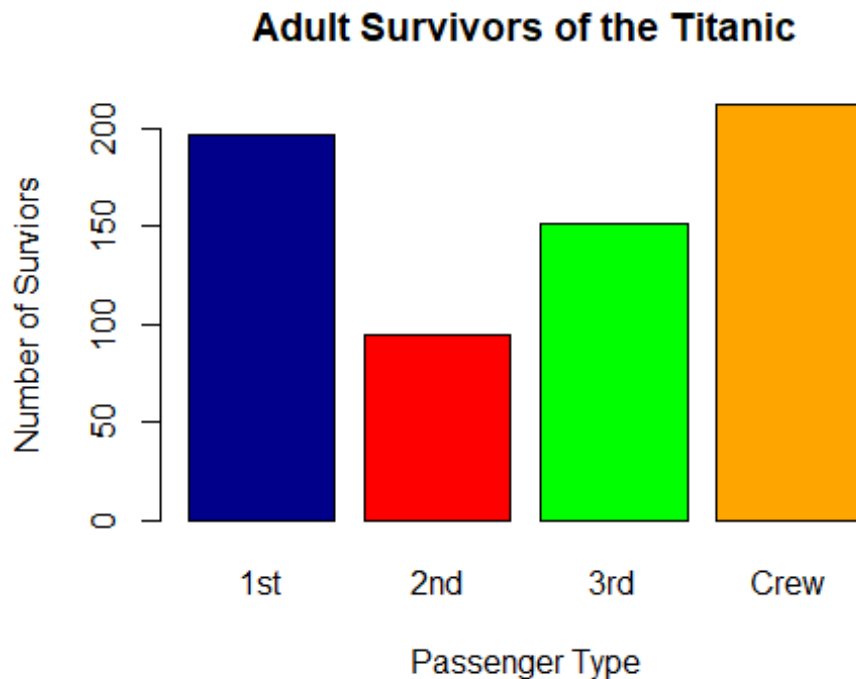
(1.b) Create (and print out) a vector called survived which contains the adult passengers who survived. Hint: You may need to use rowSums() on your answer from part (a).

```
survived = rowSums(AdultPas)
survived

##  1st  2nd  3rd Crew
## 197   94 151  212
```

(1.c) Create a barplot displaying the survived vector. Make sure to include a main title and to label your x-axis. Also, make sure that each bar is a different colour.

```
barplot(survived, main = "Adult Survivors of the Titanic", xlab = "Passenger
Type", ylab = "Number of Survivors", col = c("darkblue", "red", "green",
"orange"))
```



(1.d) What does the bar graph imply about the survival of adult passengers based on class?

It implies that the Crew had the most survivors followed closely by the 1st class and that the 3rd class had more survivors than the 2nd class.

(1.e) Create (and print out) a vector called died which contains the adult passengers who did not survive.

```
died = c(rowSums(as.table(Titanic[1:4,1:2,2,1])))
died
```

```
## 1st 2nd 3rd Crew
## 122 167 476 673
```

(1.f) Create (and print out) a vector called percent.Survived which contains the percentage of adult passengers who survived in each class.

```
total = sum(survived)

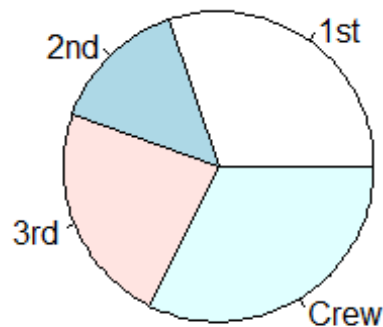
percent.Survived = (survived/total)*100
percent.Survived
```

```
##      1st      2nd      3rd      Crew
## 30.12232 14.37309 23.08869 32.41590
```

(1.g) Create a pie chart that displays the percent.Survived data. Be sure to include a main title for your pie chart.

```
pie(percent.Survived, main= "Percent of Adult Passengers that Survived")
```

Percent of Adult Passengers that Survived



(1.h) What does the pie chart imply about the survival of the adult passengers based on class? Does this imply something different than the bar graph did? If yes, why?

The pie chart tells us that if you are a crew member you have the highest percent chance of surviving, followed closely by 1st class, 3rd class and finally 2nd class. This implies largely the same thing, it implies that you have highest percent chance of surviving if you were a crew member, followed by 1st, 3rd and 2nd class.

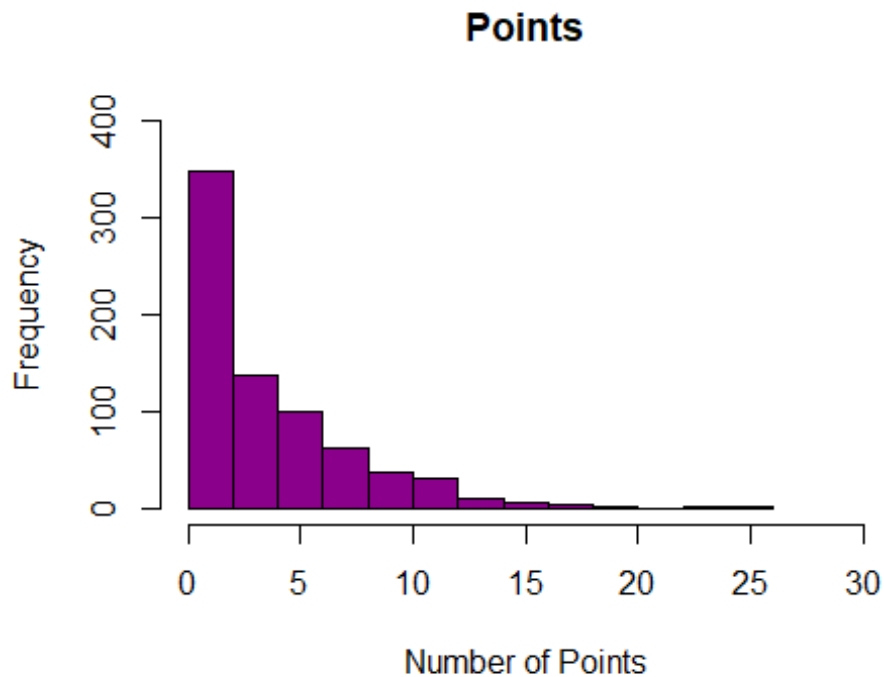
Question 2:

(2.a) Create (but do not print) a vector called `points` containing the number of points for each player (the variable `P` in the data set).

```
points = c(NHLDData$P)
```

(2.b) Create a histogram displaying the distribution of this variable. Be sure to have both a main title and a title on your x-axis. Also, make sure that the scale on the x axis goes to 30 and the scale on the y axis goes to 400.

```
hist(points,  
      main= "Points",  
      xlab="Number of Points",  
      xlim = c(0,30),  
      ylim = c(0,400),  
      col="darkmagenta")
```



(2.c) Describe the shape of the distribution (symmetric, left-skewed, right-skewed).

The distribution is unimodal, asymmetric and extremely right skewed.

(2.d) What is an appropriate measure of the center of the distribution (mean or median), why?

I believe the median would be a more appropriate due to the fact that it isn't impacted by outliers in our data like the mean calculation would be.

(2.e) Compute the appropriate center value and the corresponding measures of variability

```
median(points) #Center Value
```

```
## [1] 3
```

```
summary(points)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   1.000   3.000   3.826   6.000   26.000
```

```
IQR(points) # Interquartile Range
```

```
## [1] 5
```

```
var(points) # Variance
```

```
## [1] 14.78332
```

```
sqrt(var(points)) # Standard Deviation
## [1] 3.844908
range(points) # Range
## [1] 0 26
```

Question 3:

Consider the following sample of points from the NHL data set:

```
GivenSample = c(3,1,8,5,3,1,2,0,5,2,0,1,3,3,2,1,11,0,6,1)
```

(3.a) Create a stemplot of the distribution of the sample.

```
stem(GivenSample)

##
## The decimal point is 1 digit(s) to the right of the |
##
## 0 | 000111112223333
## 0 | 5568
## 1 | 1
```

(3.b) Does the distribution resemble the one seen in question 2? Explain why there might be some differences.

The distribution does resemble the one seen earlier but there are some differences. Since we are taking a small sample from the entire set some of the data will not be represented. The range in question 3 is also a lot smaller due to this reason, as it will be less likely to select the outliers in our data.