



STAT 3010 R PROJECT

Rebekah Sander

Abstract

This report provides an exploratory data analysis of the Passengers that were aboard the Titanic.
The data was collected on a sample of 2208 passengers and recorded different key variables.

10/12/2022

1. According to the data, there were 2208 passengers on the Titanic.

2.

Table 1: Data Dictionary of Analysis Variables for the Passengers Aboard the Titanic				
VARIABLE	LABEL	GENEAL TYPE	SPECIFIC TYPE	MEASUREMENT UNITS
Name	Names of the passengers	Categorical	Identifier	N/A
Survived	Passenger survival status	Categorical	Nominal	Died, Survived
Boarded	City of boarding	Categorical	Nominal	N/A
MWC	Man, Woman, Child	Categorical	Nominal	Man, Woman, Child
Age	Age of passenger	Quantitative	Discrete	Years
Adut_or_Chld	Adult or Child	Categorical	Nominal	Adult, Child
Sex	Sex of the passenger	Categorical	Nominal	Male, Female
Paid	Ticket Price	Quantitative	Discrete	Dollars
Class_Dept	Class Department of Passenger	Categorical	Ordinal	N/A
Class	Passenger class initial	Categorical	Ordinal	R, E, 1, 2, 3, V, D

3. Out of 2208 passengers, 1496 were reported to have died and 712 reported as survived. That is, 67.8% of the total passengers died and 32.2% of the total passengers survived.

Table 2:
Frequency(Proportions) of
Passengers that Survived
and Died, n=2208

	Overall (N=2208)
Survived	
Died	1496 (67.8%)
Survived	712 (32.2%)

Figure 1: Pie Chart of Passengers that survived and Died(n=2208)

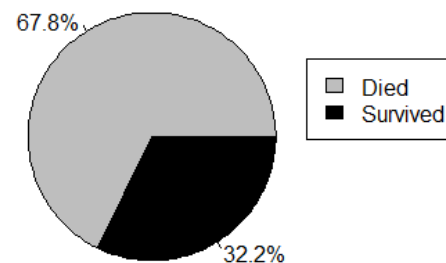
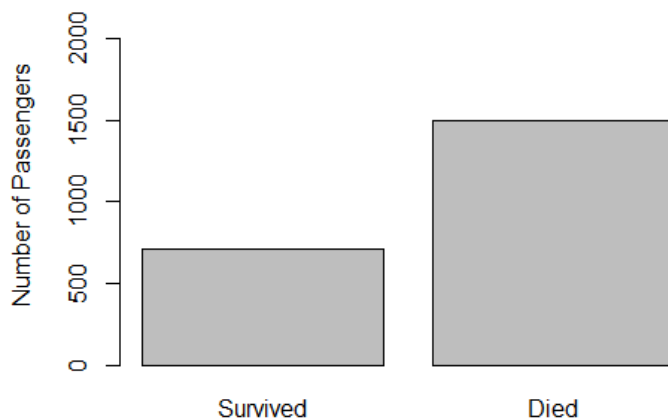


Figure 2: Bar Chart of Passengers that Survived and Died(n=2208)



```
sander_rebekah_rproject.R* x  Untitled1* x
#####
5- #####Frequency and proportion of the passengers that survived and died####
6- #####
7
8 survived_table<-table(Survived)
9 addmargins(Survived_table)
10
11 prop.table(survived_table)
12 round(100*prop.table(survived_table),digits=1)
13 apply(100*prop.table(survived_table),1,function(u)sprintf("%.1f%",u))
14
15 #table of Survived proportions using table1 package
16 library(table1)
17 table1::table1(~Survived, data=Titanic,
18               caption = "<h4><b>Table 2: Frequency(Proportions) of Passengers
19               that Survived and Died, n=2208</b></h4>")
20
21 #Pie Chart of Survived
22 pie(table(Survived), main="Figure 1: Pie chart of Survivor Status(n=2208)")
23 survived_labels<-round((table(Survived))/sum(table(Survived))*100,1)
24 survived_labels<-paste(survived_labels,"%",sep="")
25 greyscale<-c("grey","black")
26 pie(table(Survived), main="Figure 1: Pie chart of Passengers that
27   survived and Died(n=2208)", col=greyscale,
28   labels=survived_labels)
29 legend(1,0.5,c("Died", "Survived"), fill=greyscale)
30
31 #Bar plot of Survived
32 barplot(sort(table(Survived), increasing=true), main="Figure 2: Bar Chart of Passengers
33   that Survived and Died(n=2208)",
34   ylab="Number of Passengers", ylim=c(0,2208))
35
```

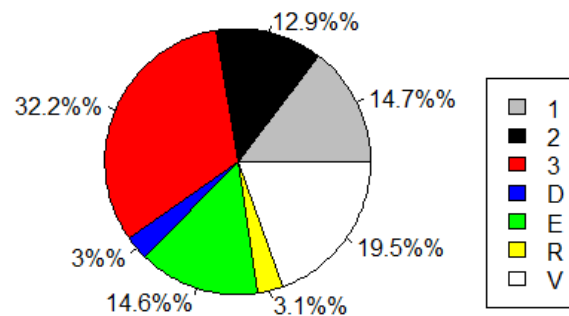
4. Out of 2208 Passengers:

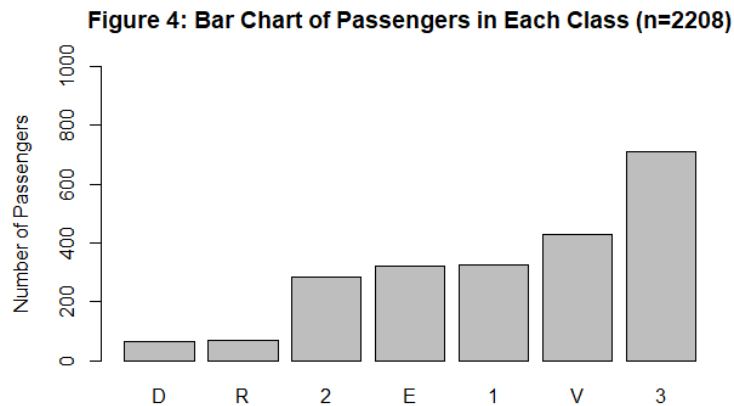
- 324 passengers were part of first class, making up 14.7% of all passengers.
- 285 passengers were part of second class, making up 12.9% of all passengers.
- 710 passengers were part of third class, making up 32.2% of all passengers.
- 66 passengers were part of deck crew, making up 3% of all passengers.
- 323 passengers were part of engineering crew, making up 14.6% of all passengers.
- 69 passengers were part of restaurant staff, making up 3.1% of all passengers.
- 431 passengers were part of victualling crew, making up 19.5% of all passengers.

Table 3:
Frequency(Proportions)
of Passengers in Each
Class, n=2208

	Overall (N=2208)
Class	
1	324 (14.7%)
2	285 (12.9%)
3	710 (32.2%)
D	66 (3.0%)
E	323 (14.6%)
R	69 (3.1%)
V	431 (19.5%)

Figure 3: Pie Chart of Passengers in Each Class(n=2208)





```

35 #####
36 #####Frequency and proportion of the passengers in each class###
37 #####
38 #####
39
40 class_table<-table(class)
41 addmargins(class_table)
42 prop.table(class_table)
43 round(100*prop.table(class_table),digits=1)
44 apply(100*prop.table(class_table),1,function(u)sprintf("%.1f%%",u))
45
46 #table of class proportions using table1 package
47 library(table1)
48 table1::table1(~class, data=titanic, caption = "<h4><b>Table 3: Frequency(Proportions) of Passengers
49 in Each Class, n=2208</b></h4>")
50
51 #Pie Chart of Class
52 pie(table(class), main="Figure 3: Pie Chart of Passengers in each Class(n=2208)")
53 class_labels<-round((table(class))/sum(table(class))*100,1)
54 class_labels<-paste(class_labels,"%",sep="")
55 colorscale<-c("grey","black","red","blue","green","yellow","white")
56 pie(table(class), main="Figure 3: Pie Chart of Passengers in Each Class(n=2208)", col=colscale,
57 labels=class_labels)
58 legend(1.5,0.5,c("1", "2", "3", "D", "E", "R", "V"), fill=colscale)
59
60 #Barplot of Class
61
62 barplot(sort(table(class), increasing=true), main="Figure 4: Bar Chart of Passengers in Each Class (n=2208)",
63 ylab="Number of Passengers", ylim=c(0,1000))
64
65

```

Class code key:

- 1- First Class
- 2- Second Class
- 3- Third Class
- D- Deck Crew
- E- Engineering Crew
- R- Restaurant Staff
- V- Victualling Crew

5. Out of 2208 total passengers, 124 were children, 1652 were men, and 432 were women. This means 5.6% of passengers were children, 74.8% of passengers were men, and 19.6% of passengers were women.

Table 4:
Frequency(Proportions) of
Men, Women, and Children
(n=2208)

	Overall (N=2208)
MWC	
Child	124 (5.6%)
Man	1652 (74.8%)
Woman	432 (19.6%)

Figure 5: Pie Chart of Men, Women,
and Children (n=2208)

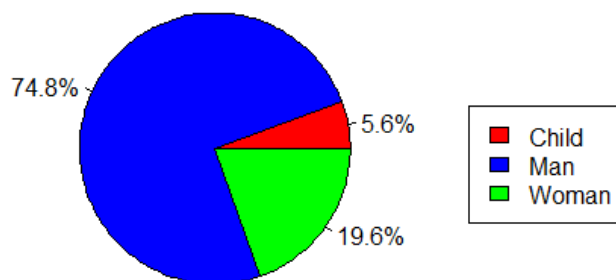
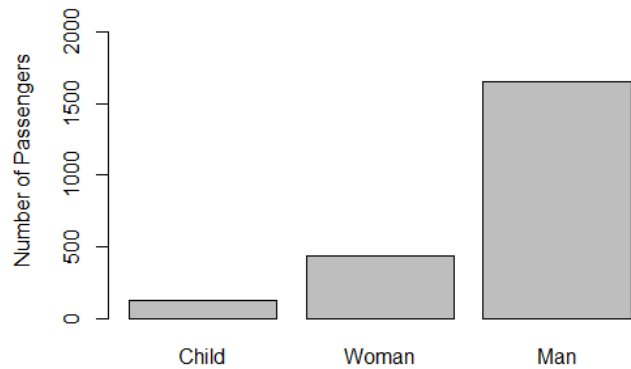


Figure 6: Bar Chart of Men, Women, and Children (n=2208)



```
sander_rebekah_project.R* x  Untitled1* x
Source on Save  Run  Up  Down

66 > #####
67 > ####Frequency and proportion of the men, women, and children###
68 > #####
69
70 MWC_table<-table(MWC)
71 addmargins(MWC_table)
72 prop.table(MWC_table)
73 round(100*prop.table(MWC_table),digits=1)
74 apply(100*prop.table(MWC_table),1,function(u)sprintf("%.1f%%",u))
75
76 #table of MWC proportions using table1 package
77 library(table1)
78 table1::table1(~MWC, data=Titanic, caption =
79   "<h4><b>Table 4: Frequency(Proportions) of Men,
80     women, and Children (n=2208)</b></h4>")
81
82 #Pie Chart of MWC
83 pie(table(MWC), main="Figure 5: Pie Chart of Men, women,
84   and Children (n=2208)")
85
86 MWC_labels<-round((table(MWC))/sum(table(MWC))*100,1)
87 MWC_labels<-paste(MWC_labels,"%",sep="")
88 colorscale<-c("red","blue","green")
89 pie(table(MWC), main="Figure 5: Pie chart of Men, women,
90   and Children (n=2208)", col=colourscale,
91   labels=MWC_labels)
92 legend(1.5,0.25,c("child", "Man", "woman"), fill=colourscale)
93
94 #Barplot of MWC
95 barplot(sort(table(MWC), increasing=true), main="Figure 6: Bar Chart of Men, women,
96   and Children (n=2208)",
97   ylab="Number of Passengers", ylim=c(0,2208))
98
```

6. Out of 2208 total passengers, 85 passengers' class departments were not recorded. In first class: 274 were just passengers, 3 were part of the H&W Guarantee Group, and 36 were servants. In second class: 252 were just passengers, 6 were part of the H&W Guarantee Group, 8 were musicians, and 4 were servants. There were 656 third class passengers. And the crew passengers consisted of: 58 of deck crew, 8 of Titanic Officers, 320 of engineering crew, 69 of restaurant staff, 424 of victualling crew, and 5 of postal clerk.
- There were 3,000 Harland and Wolff employees who worked on the construction of the Titanic. 9 of those 3,000 were selected to travel on the Titanic as representatives of H&W and were known as the H&W Guarantee Group.
 - According to the data, there were 8 musicians aboard the Titanic and they traveled as second class passengers.

Table 5: Frequency(Proportions) of Passengers in each Class Department

Class_Dept	Overall (N=2208)
	85 (3.8%)
1st Class Passenger	274 (12.4%)
1st Class PassengerH&W Guarantee Group	3 (0.1%)
1st Class PassengerServant	36 (1.6%)
2nd Class Passenger	252 (11.4%)
2nd Class PassengerH&W Guarantee Group	6 (0.3%)
2nd Class PassengerMusician	8 (0.4%)
2nd Class PassengerServant	4 (0.2%)
3rd Class Passenger	656 (29.7%)
Deck Crew	58 (2.6%)
Deck CrewTitanic Officers	8 (0.4%)
Engineering Crew	320 (14.5%)
Restaurant Staff	69 (3.1%)
Victualling Crew	424 (19.2%)
Victualling CrewPostal Clerk	5 (0.2%)


```
sander_rebekah_rproject.R* x Untitled1* x
Source on Save Run Source
99
100 #####
101 #####Frequency of the passengers in each class Department###
102 #####
103
104 ClassDep_table<-table(Class_Dept)
105 addmargins(ClassDep_table)
106
107 prop.table(ClassDep_table)
108 round(100*prop.table(ClassDep_table),digits=1)
109 apply(100*prop.table(ClassDep_table),1,function(u)sprintf("%.1f%%",u))
110
111 #table of Class_Dept proportions using table1 package
112 library(table1)
113 table1::table1(~Class_Dept, data=Titanic, caption =
114               "<h4><b>Table 5: Frequency(Proportions) of Passengers in
115               each Class Department</b></h4>")
116
117 #Pie Chart of Class_Dept
118 pie(table(Class_Dept), main="Figure 7: Pie Chart of each Class
119     Department(n=2208)")
120
121 #Barplot of Class_Dept
122 barplot(sort(table(Class_Dept), increasing=true), main="Figure 8: Bar Chart of
123     Class Department(n=2208)", ylab="Number of Passengers", ylim=c(0,2208))
124
125
126
101:15 Frequency of the passengers in each class Department R Script
```

7. The distribution of the age of the passengers is slightly skewed right. Due to outliers, it is more meaningful to use the median of 29 years old as the center of the spread and the IQR of 15 years to describe the dispersion of the data.

**Table 6: Descriptive Statistics
of Passenger Age**

	Overall (N=2208)
Age	
Mean (SD)	30.1 (12.0)
Median(IQR)	29.0(15.0)
Min,Max	0.0800,74.0
Q1,Q3	22.0,37.0
Sample Size	2205
Missing	3 (0.1%)

Figure 7: Histogram of Age of Passengers

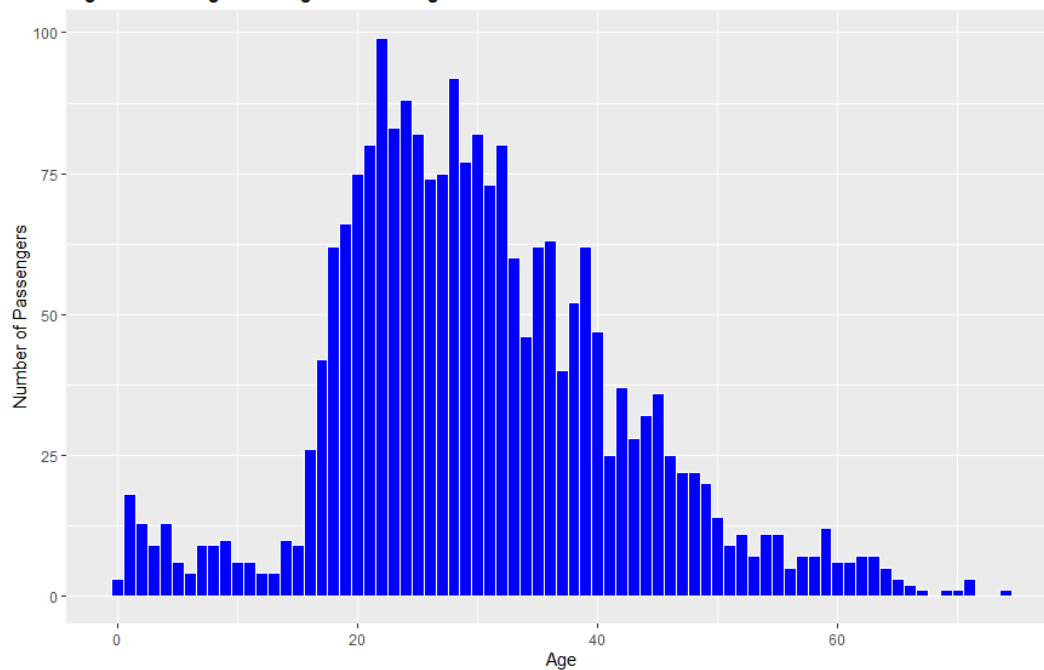
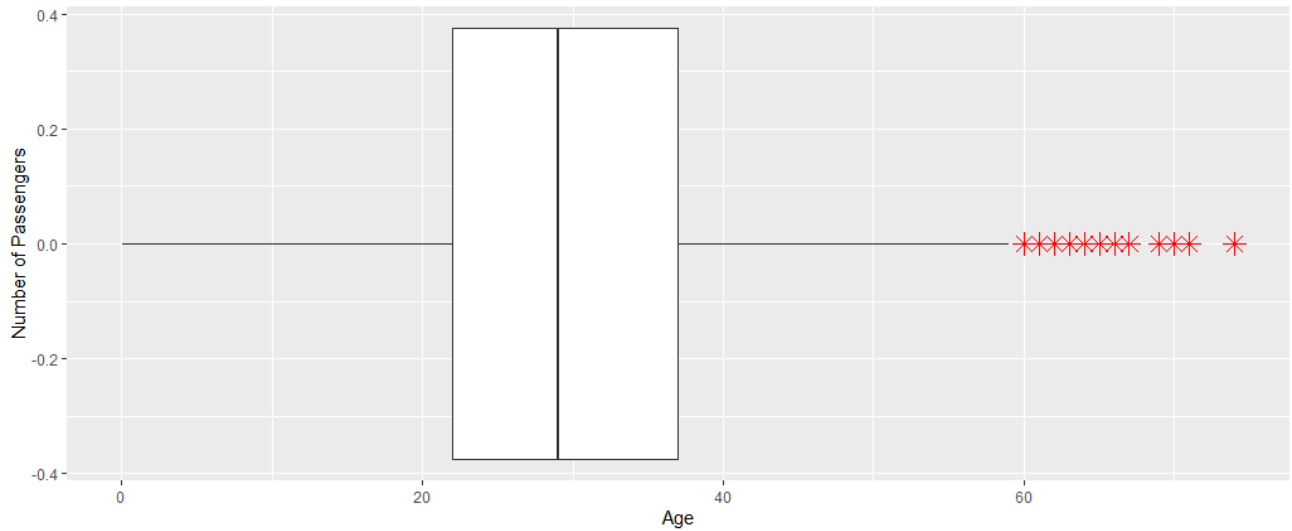


Figure 8: Boxplot of Age of Passengers



```

126
127 #####
128 ### Distribution of the age of passengers ###
129 #####
130
131 # Descriptive Statistics
132 mean(Age, na.rm=TRUE)
133 median(Age, na.rm=TRUE)
134 min(Age, na.rm=TRUE)
135 max(Age, na.rm=TRUE)
136 range(Age, na.rm=TRUE)
137 quantile(Age, na.rm=TRUE)
138 quantile(Age, na.rm=TRUE, seq(0,1,0.1))
139 IQR(Age, na.rm=TRUE)
140 var(Age, na.rm=TRUE)
141 sd(Age, na.rm=TRUE)
142
143 #Calculate the Summary Statistics
144 summary(Age)
145
146 # Table of Statistics using table1
147 library(table1)
148 table1::label(Age) <- "Age of Passenger"
149 table1::table1(~Age, data=Titanic, render.continuous=c(."Mean (SD)",
150               .="Median(IQR)", .="Min,Max", .="Q1,Q3", "Sample Size" = "N"),
151               caption = "<h4><b>Table 6: Descriptive Statistics of Passenger Age</b></h4>")
152
153 #Histogram of Age of Passengers
154 hist(Age, main="Figure 7: Histogram of Age of Passengers (n=2208)",
155      xlab="Age", ylab="Number of Passengers")
156
157 #Boxplot of Age
158 boxplot(Age, main="Figure 8: Boxplot of Age of Passengers (n=2208)",
159        ylab="Age")
160
161
162 #Working in the Tidy verse with ggplot2
163 library(ggplot2)
164
165 #Histogram
166 age_hist <- ggplot(Titanic, aes(x=Age)) + geom_histogram(fill="blue",
167               color="white", binwidth=1)
168 print(age_hist + labs(title="Figure 7: Histogram of Age of Passengers",
169               y="Number of Passengers", x="Age"))
170
171 #Boxplot
172 age_boxplot <- ggplot(Titanic, aes(x=Age)) + geom_boxplot(outlier.colour = "red",
173               outlier.shape = 8, outlier.size=4)
174 print(age_boxplot + labs(title="Figure 8: Boxplot of Age of Passengers",
175               y="Number of Passengers", x="Age"))
176
177

```

8. Ticket Prices are skewed right. A median of \$14.4 best describes what the majority of passengers paid for a ticket. And an IQR of 23.1 is the best measure of spread. The highest amount paid for a ticket was \$512.

**Table 7: Descriptive Statistics
of Ticket Price**

	Overall (N=2208)
Paid	
Mean (SD)	33.0 (51.9)
Median(IQR)	14.4(23.1)
Min,Max	0,512
Q1,Q3	7.90,31.0
Sample Size	1318
Missing	890 (40.3%)

Figure 9: Histogram of Ticket Prices

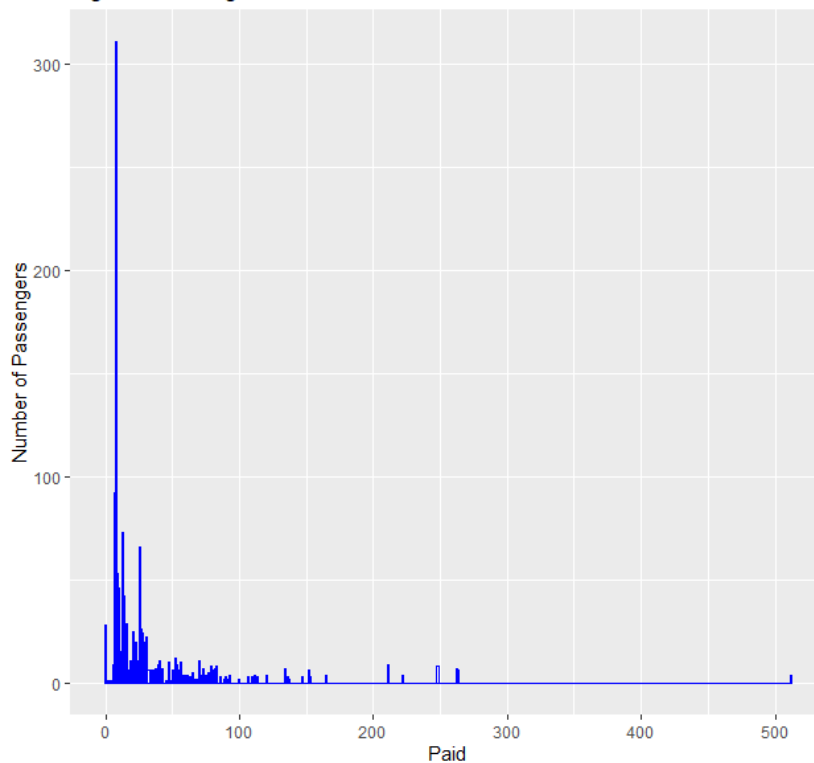
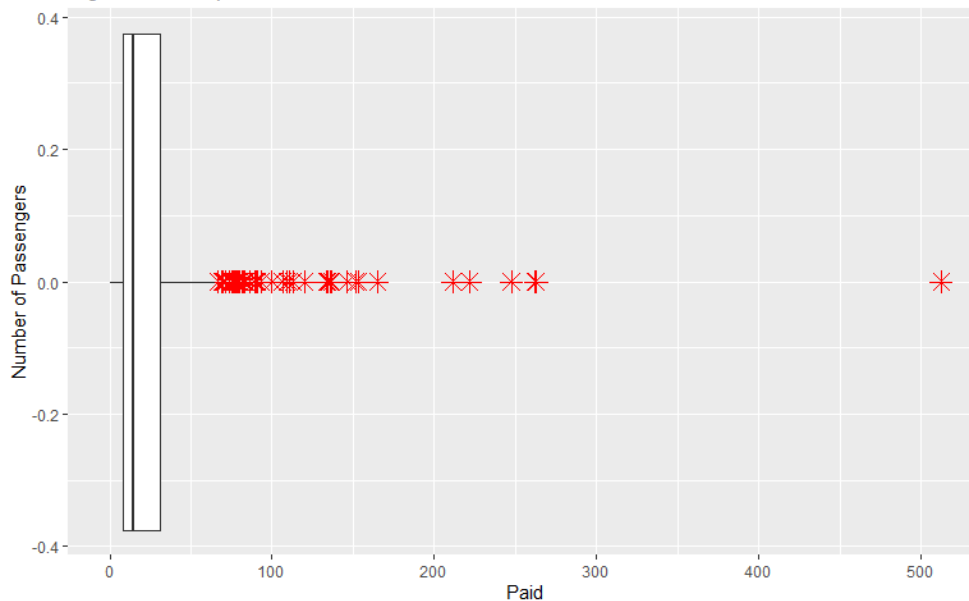


Figure 10: Boxplot of Ticket Prices



```

176
177
178 #####
179 ### Distribution of the ticket prices ###
180 #####
181
182 #Calculate the Summary Statistics
183 summary(Paid)
184
185 # Table of Statistics using table1
186 library(table1)
187 table1::label(Paid) <- "Age of Passenger"
188 table1::table1(~Paid, data=Titanic, render.continuous=c(., "Mean (SD)", ., "Median (IQR)",
189 ., "Min, Max", ., "Q1, Q3", "Sample Size" = "N"),
190 caption = "<h4><b>Table 7: Descriptive Statistics of Ticket Price</b></h4>")
191
192 #Histogram of Age of Passengers
193 hist(Paid, main="Figure 9: Histogram of Ticket Prices (n=2208)",
194 xlab="Age", ylab="Number of Passengers")
195
196
197 #Boxplot of Age
198 boxplot(Paid, main="Figure 10: Boxplot of Ticket Prices (n=2208)",
199 ylab="Age")
200
201
202 #Working in the Tidy verse with ggplot2
203 library(ggplot2)
204
205 #Histogram
206 paid_hist <- ggplot(Titanic, aes(x=Paid)) + geom_histogram(fill="white", color="blue", binwidth=1)
207 print(paid_hist + labs(title="Figure 9: Histogram of Ticket Prices", y="Number of Passengers", x="Paid"))
208
209 #Boxplot
210 paid_boxplot <- ggplot(Titanic, aes(x=Paid)) + geom_boxplot(outlier.colour = "red", outlier.shape = 8, outlier.size=4)
211 print(paid_boxplot + labs(title="Figure 10: Boxplot of Ticket Prices", y="Number of Passengers", x="Paid"))
212
213
214

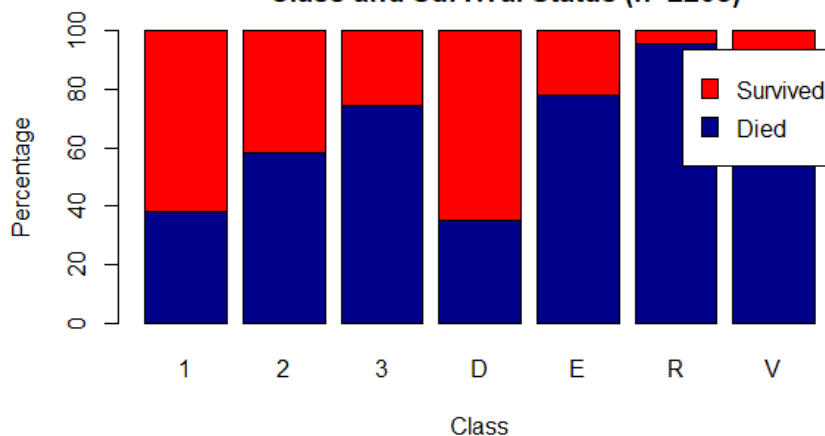
```

9. Observing the 100% stacked bar chart, we can see that first class had more survivors than second class and second class had more survivors than third class. From this observation, I believe there is potentially a relationship between a passenger's class and if they survived. This could be due to who people place a higher importance on. Regarding crew, the deck crew had the most survivors and this could be due to their location at the time of the sinking.

Table 8: Contingency Table of a Passenger's Class and Survival Status

	Died (N=1496)	Survived (N=712)	Overall (N=2208)
Class			
1	123 (8.2%)	201 (28.2%)	324 (14.7%)
2	166 (11.1%)	119 (16.7%)	285 (12.9%)
3	530 (35.4%)	180 (25.3%)	710 (32.2%)
D	23 (1.5%)	43 (6.0%)	66 (3.0%)
E	251 (16.8%)	72 (10.1%)	323 (14.6%)
R	66 (4.4%)	3 (0.4%)	69 (3.1%)
V	337 (22.5%)	94 (13.2%)	431 (19.5%)

Figure 11: 100% Stacked Bar Chart of Passenger's Class and Survival Status (n=2208)



```

sander_rebekah_rproject.R* x
Source on Save Run
215
216 #####
217 #####How is class related to survival###
218 #####
219 class_survived_table<-table(class, Survived)
220 class_survived_table
221 addmargins(class_survived_table)
222 prop.table(class_survived_table)
223 round(100*prop.table(class_survived_table),digits=1)
224 apply(100*prop.table(class_survived_table),1,function(u) sprintf("%.1f%%",u))
225
226 #using table1
227 library(table1)
228 table1::label(class)<-"Passenger Class"
229 table1::label(Survived)<-"Passenger Survival"
230 table1::table1(~Class | Survived, data=Titanic, render.continuous=c(."Mean (SD)",
231                               .="Median(IQR)",."Min,Max", .="Q1,Q3","Sample Size" = "N"),
232               caption = "<h4><b>Table 8: Contingency Table of a Passenger's
233                           Class and Survival Status</b></h4>")
234
235 #100% bar chart
236 surv_class_table<-table(Survived, Class)
237 percentage<-apply(surv_class_table,2,function(x){x*100/sum(x,na.rm=T)})
238 barplot(percentage,main="Figure 11: 100% Stacked Bar Chart of Passenger's
239       Class and Survival Status (n=2208) ",
240       xlab="Class",ylab="Percentage",col=c("darkblue","red"),
241       legend=rownames(surv_class_table))
242
243

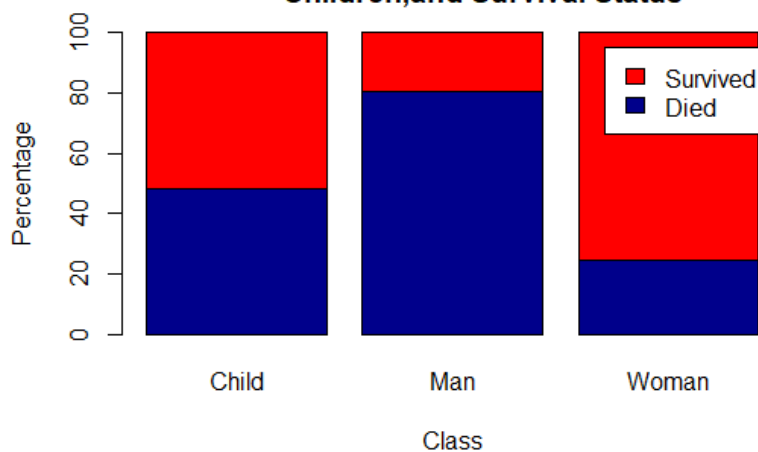
```

10. Observing the 100% stacked bar chart, we are able to see that there were more passengers that were women who survived, following the women are the child passengers. From these observations, there could potentially be a relationship between if the passenger was a man, woman, or child and if they survived. These observations could be due to the time period, women and children were most likely of higher priority.

Table 9: Contingency Table of the Men, Women, Children and Survival Status

	Died (N=1496)	Survived (N=712)	Overall (N=2208)
MWC			
Child	60 (4.0%)	64 (9.0%)	124 (5.6%)
Man	1331 (89.0%)	321 (45.1%)	1652 (74.8%)
Woman	105 (7.0%)	327 (45.9%)	432 (19.6%)

Figure 12: 100% Stacked Bar Chart of the Men, Women, Children, and Survival Status




```

sander_rebekah_rproject.R*
Source on Save Run
244
245 #####
246 #####How is MWC related to survival###
247 #####
248 MWC_survived_table<-table(MWC, Survived)
249 MWC_survived_table
250 addmargins(MWC_survived_table)
251 prop.table(MWC_survived_table)
252 round(100*prop.table(MWC_survived_table),digits=1)
253 apply(100*prop.table(MWC_survived_table),1,function(u) sprintf("%.1f%%",u))
254
255 #using table1
256 library(table1)
257 table1::label(MWC)<- "man, woman, child"
258 table1::label(Survived)<- "Passenger Survival"
259 table1::table1(~MWC | Survived, data=Titanic, render.continuous=c(."Mean (SD)",
260 .="Median(IQR)",."Min,Max", .="Q1,Q3","Sample Size" = "N"),
261 caption = "<h4><b>Table 9: Contingency Table of the Men, women,
262 children and Survival Status</b></h4>")
263
264 #100% bar chart
265 surv_MWC_table<-table(Survived, MWC)
266 percentage<-apply(surv_MWC_table,2,function(x){x*100/sum(x,na.rm=T)})
267 barplot(percentage,main="Figure 12: 100% Stacked Bar Chart of the Men, women,
268 children, and Survival Status",xlab="Class",ylab="Percentage",
269 col=c("darkblue","red"), legend=rownames(surv_MWC_table))
270

```

11. The distribution of age of the passengers that died is more skewed right than the distribution of age of the passengers who survived. Both distributions are similar in shape though. Since both have outliers, the median is best to describe the center and both distributions have a median of 29 years. IQR only varies by 1 between the two distributions.

Table 10: Descriptive Statistics of Age by Survived Status

	Died (N=1496)	Survived (N=712)	Overall (N=2208)
Age			
Mean (SD)	30.5 (11.6)	29.5 (12.7)	30.1 (12.0)
Median(IQR)	29.0(15.0)	29.0(16.0)	29.0(15.0)
Min,Max	0.330,74.0	0.0800,64.0	0.0800,74.0
Q1,Q3	22.0,37.0	22.0,38.0	22.0,37.0
Sample Size	1493	712	2205
Missing	3 (0.2%)	0 (0%)	3 (0.1%)

Figure 13: Side by Side Histogram of Age by Survived

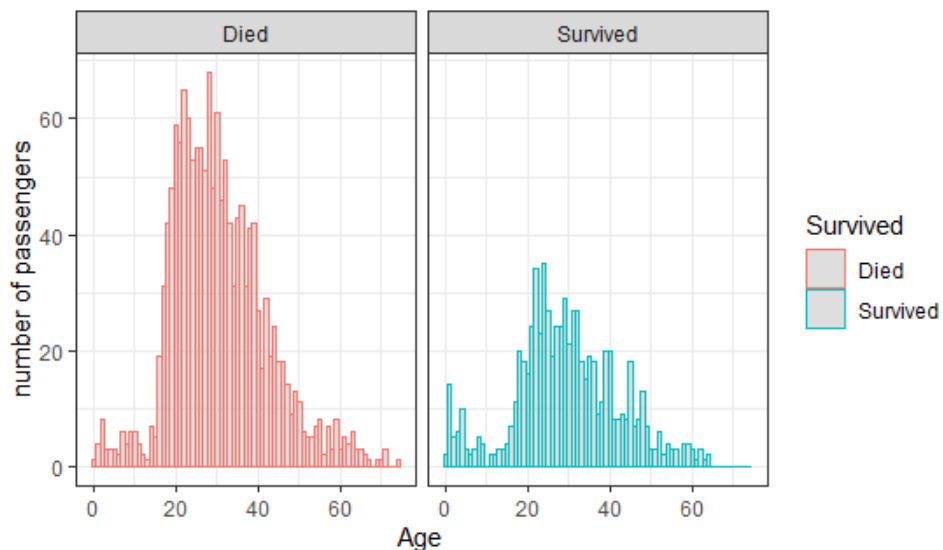
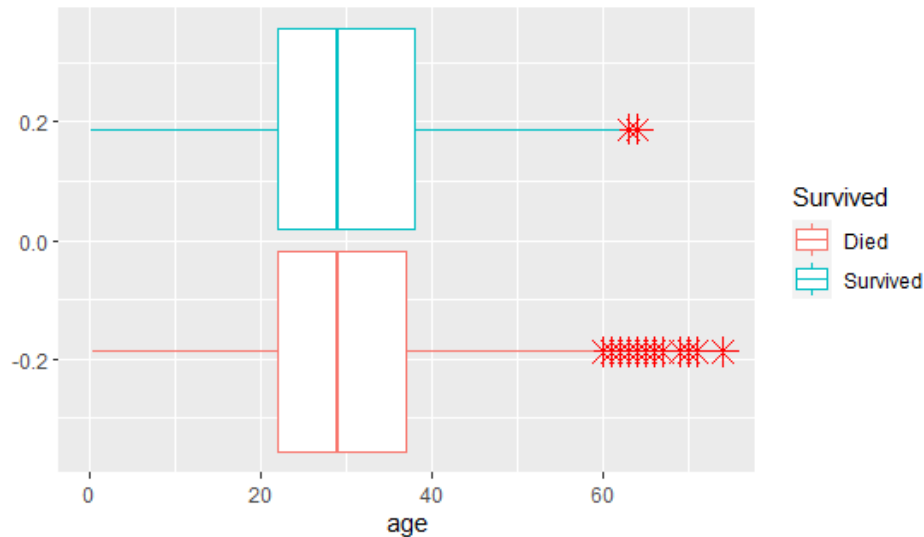


Figure 14: Side by Side Boxplot of Age by Survived



```
sander_rebekah_rproject.R x
Source on Save
Code Tools
Run
Source

272
273 #####
274 ####Distribution of Age comparison between Survived and Died####
275 #####
276 #Descriptive Stats
277 library(table1)
278 table1::label(Age) <- "Age of Passenger"
279 table1::label(Survived) <- "Survived Status"
280 table1::table1(~Age|Survived, data=Titanic, render.continuous=c(."Mean (SD)",
281 .="Median(IQR)", .="Min,Max", .="Q1,Q3", "Sample Size" = "N"),
282 caption = "<h4><b>Table 10: Descriptive Statistics of Age by
283 Survived Status</b></h4>")
284
285 #Histogram
286 library(ggplot2)
287 survived_age_hist <- ggplot(Titanic, aes(x=Age, color=Survived)) +
288   geom_histogram(fill="grey", alpha=0.5, binwidth = 1) +
289   facet_grid(~Survived) + theme_bw()
290 print(survived_age_hist + labs(title="Figure 13: Side by Side Histogram of Age by
291 Survived", x="Age", y="number of passengers"))
292
293 #Boxplots
294 survived_age_box <- ggplot(Titanic, aes(x=Age, color=Survived)) +
295   geom_boxplot(outlier.colour = "red", outlier.shape = 8, outlier.size = 4)
296 print(survived_age_box + labs(title = "Figure 14: Side by Side Boxplot of Age by
297 Survived", x="age"))
298
```

12. The distribution of ticket prices of those who died and those who survived are both skewed right. With both the survived group and the died group, there are outliers, so the median is the best representation of center. It is important to take notice though, that the people who died have a median ticket price of 10.5 and the median ticket price of those who survived is 26. In addition, the highest ticket price paid was 512, which is significantly different than most ticket prices, and that passenger is reported to have survived.

Table 11: Descriptive Statistics of Ticket Price by Survived Status

	Died (N=1496)	Survived (N=712)	Overall (N=2208)
Paid			
Mean (SD)	22.9 (34.0)	49.6 (69.1)	33.0 (51.9)
Median(IQR)	10.5(18.1)	26.0(46.6)	14.4(23.1)
Min,Max	0,263	0,512	0,512
Q1,Q3	7.85,26.0	11.3,57.9	7.90,31.0
Sample Size	819	499	1318
Missing	677 (45.3%)	213 (29.9%)	890 (40.3%)

Figure 15: Side by Side Histogram of Ticket Price by Survived Status

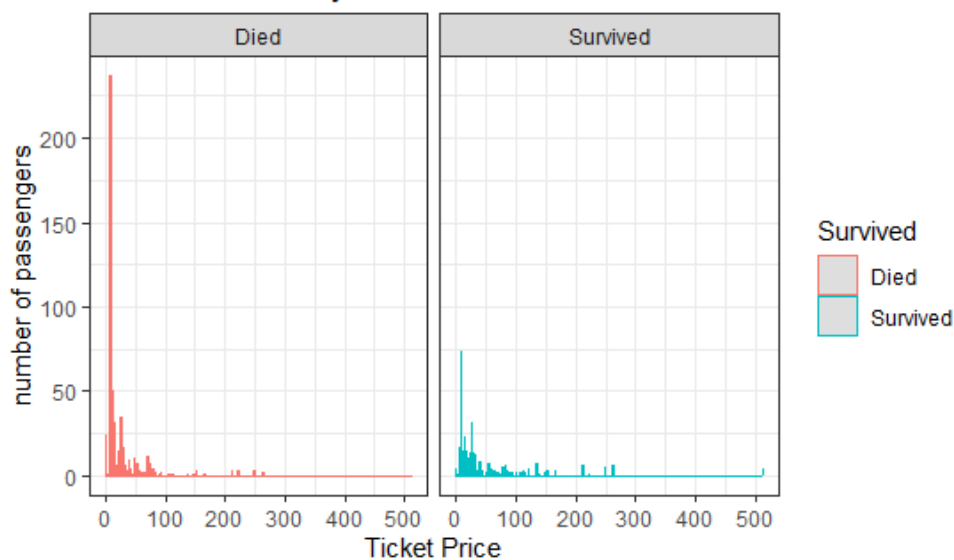
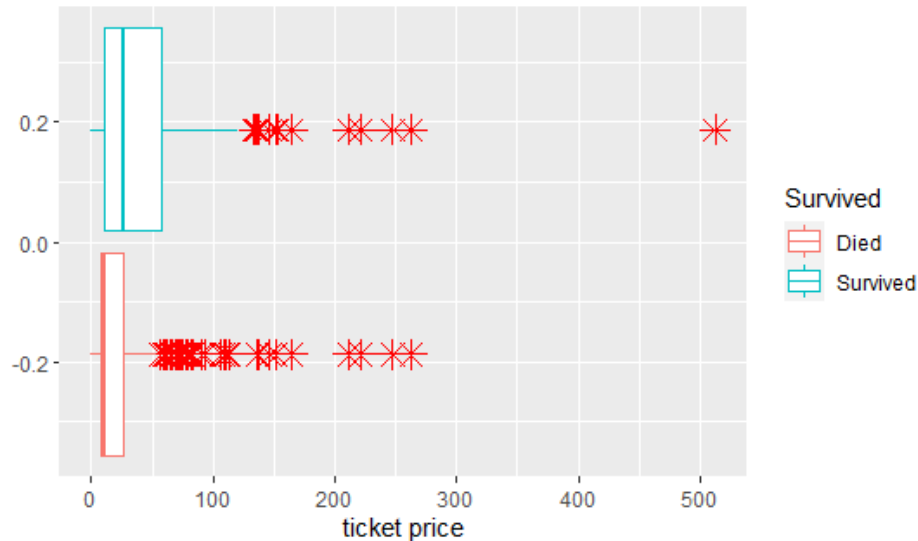


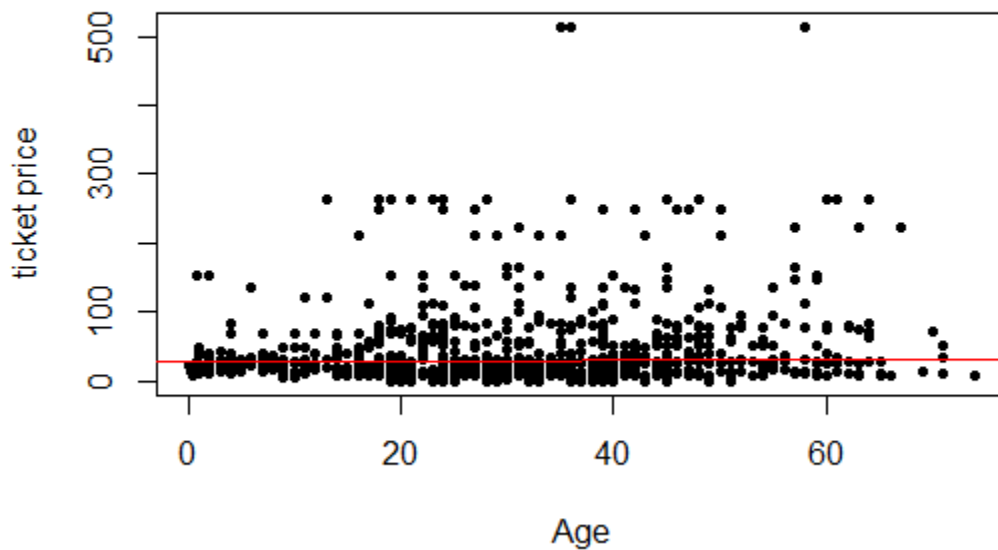
Figure 16: Side by Side Boxplot of Ticket Price by Survived Status



```
sander_rebekah_rproject.R* x
Source on Save Run
299 #####
300 #####Distribution of Ticket Price comparison between Survived and Died####
301 #####
302 #####
303 #Descriptive Stats
304 library(table1)
305 table1::label(Paid) <- "Ticket Price"
306 table1::label(Survived) <- "Survived Status"
307 table1::table1(~Paid|Survived, data=Titanic, render.continuous=c(."Mean (SD)",
308 .="Median(IQR)", .="Min,Max", .="Q1,Q3", "Sample Size" = "N"),
309 caption = "<h4><b>Table 11: Descriptive Statistics of Ticket Price
310 by Survived Status</b></h4>")
311
312 #Histogram
313 library(ggplot2)
314 survived_paid_hist <- ggplot(Titanic, aes(x=Paid, color=Survived)) +
315   geom_histogram(fill="grey", alpha=0.5, binwidth = 1) + facet_grid(~Survived) + theme_bw()
316 print(survived_paid_hist + labs(title="Figure 15: Side by Side Histogram of Ticket
317 Price by Survived Status", x="Ticket Price", y="number of passengers"))
318 #Boxplots
319 survived_paid_box <- ggplot(Titanic, aes(x=Paid, color=Survived)) +
320   geom_boxplot(outlier.colour = "red", outlier.shape = 8, outlier.size = 4)
321 print(survived_paid_box + labs(title = "Figure 16: Side by Side Boxplot of Ticket
322 Price by Survived Status", x="ticket price"))
323
324
```

13. The correlation coefficient of age and amount paid is 0.18. This tells us there is a weak positive correlation between age and ticket price. From this observation, we cannot conclude that there is a relationship between age and ticket price.

Figure 17: Scatterplot of Age and Ticket Price



```
sander_rebekah_project.R* x
Source on Save
Run
323
324
325
326 #####
327 #####Relationship between age and ticket price###
328 #####
329
330 # Scatter Plot
331 plot(Age, Paid, main="Figure 17: Scatterplot of Age and Ticket Price",
332      xlab="Age",ylab="ticket price", pch=20)
333
334 # regression line
335 abline(lm(Age~Paid),col="red")
336
337 #correlation coefficient
338 cor(Age, Paid, use="complete.obs")
339
340
```

14.

Table 12: Frequency of Each Passenger's Class (New Class Variable)

	Overall (N=2208)
Class Department	
Crew	889 (40.3%)
First	324 (14.7%)
Second	285 (12.9%)
Third	710 (32.2%)

```
sander_rebekah_rproject.R* x
339
340
341 #####
342 ###New Class Variable Creation and Frequency Table###
343 #####
344 #variable creation
345 Titanic$Department[class=="D"]<-"Crew"
346 Titanic$Department[class=="E"]<-"Crew"
347 Titanic$Department[class=="V"]<-"Crew"
348 Titanic$Department[class=="R"]<-"Crew"
349 Titanic$Department[class=="1"]<-"First"
350 Titanic$Department[class=="2"]<-"Second"
351 Titanic$Department[class=="3"]<-"Third"
352 Titanic$Department<-factor(Titanic$Department)
353
354 #Frequency Table
355 library(table1)
356 table1::label(Titanic$Department)<-"Class Department"
357 table1::table1(~Titanic$Department, data=Titanic,
358               caption = "<h4><b>Table 12: Frequency of Each Passenger's Class
359                       (New Class Variable)</b></h4>")
360
```

15.

Table 13: Contingency Table of Each Passenger's Age by Survived Status (New Age Variable)

	Died (N=1496)	Survived (N=712)	Overall (N=2208)
Age group			
Baby	16 (1.1%)	27 (3.8%)	43 (1.9%)
Child	36 (2.4%)	31 (4.4%)	67 (3.0%)
Teenager	63 (4.2%)	28 (3.9%)	91 (4.1%)
Young Adult	383 (25.6%)	170 (23.9%)	553 (25.0%)
Middle Adult	510 (34.1%)	231 (32.4%)	741 (33.6%)
Older Adult	424 (28.3%)	201 (28.2%)	625 (28.3%)
Senior	61 (4.1%)	24 (3.4%)	85 (3.8%)
Missing	3 (0.2%)	0 (0%)	3 (0.1%)

```

360
361
362 #####
363 #####New Age Variable Creation and Contingency Table###
364 #####
365
366 Titanic$ageGRP<-cut(Age,breaks=c(0,3,12,17,24,34,54,Inf),
367                      labels=c("Baby", "Child", "Teenager", "Young Adult", "Middle Adult",
368 Titanic$ageGRP<-factor(Titanic$ageGRP)
369
370 #Contingency Table
371 library(table1)
372 table1::label(Titanic$ageGRP)<-"Age group"
373 table1::label(Survived)<-"Survived Status"
374 table1::table1(~Titanic$ageGRP|Survived, data=Titanic,
375                caption = "<h4><b>Table 13: Contingency Table of Each Passenger's Age
376                by Survived Status (New Age Variable)</b></h4>")
377
378
379

```


- 16.** Through this analysis, many interesting things are to be said about the passengers of the Titanic. The comparisons of the Survived variable and multiple other variables were especially interesting to see. I believe that a lot of statistics make sense when looking at the numbers and knowing what kind of world it was. It was clear that being a woman, a child, or just being wealthy would higher the chances of being one of the survived.

In addition, it was interesting to take note of the ages that were aboard the Titanic. There were more middle adults than any other age group. This could be due to still being somewhat young but also having some money.

