

# Kaggle\_project1\_Titanic: Learning from Disaster

*Ioana Dragomirescu*

*April 3, 2017*

Practice Skills: Binary Classification, R

According to Data Science Solutions book the workflow should cover:

1. Question or problem definition
2. Acquire training and testing data -have a first look
3. Prepare and clean the data
4. Analyze, identify patterns, explore the data
5. Model, predict and solve the problem
6. Visualize, report and present the problem solving steps and the final solution
7. Submit the results

## 1. Question or problem definition

### 1. 1. Sad story behind data:

RMS Titanic was in 1912 the most amazingly large British Ship, one of the three Olympic class ocean liners operated by the White Star Line. It sank in the North Atlantic Ocean in the early morning of 15 April 1912, after colliding with an iceberg around 11:40 p.m. during her maiden voyage from Southampton to New York City. Books were written about this tragedy, romantic films or documentary were made, behind the scenes stories were told.

Starting his journey, Titanic had 2224 peoples on board. Everything in first class was luxurious, in fact that was the idea: to distinguish the boat from other english liners through luxurious services. This is way many famous people were on board: the American millionaire John Jacob Astor IV and his wife Madeleine Force Astor, industrialist Benjamin Guggenheim, Macy's owner Isidor Straus and his wife Ida, Denver millionairess Margaret "Molly" Brown, Sir Cosmo Duff Gordon and his wife, couturière Lucy (Lady Duff-Gordon). Also aboard the ship were the White Star Line's managing director J. Bruce Ismay and Titanic's designer Thomas Andrews (who by the way died in this disaster). But only 833 were First Class Passengers the rest 614 in Second Class and 1,006 in Third Class, all for a capacity of 2,453. It clearly had advanced safety features such as watertight compartments and remotely activated watertight doors, however there it only carried enough lifeboats for 1,178 people.

The ship left Southampton on 10 April 1912, it called at Cherbourg in France and Queenstown in Ireland and then headed west to New York. On 14 April, four days into the crossing and about 375 miles (600 km) south of Newfoundland, she hit an iceberg at 11.40 p.m. ship's time. The collision caused the ship's hull plates to buckle inwards along her starboard (right) side and opened five of her sixteen watertight compartments to the sea; the ship gradually filled with water. Meanwhile, passengers and some crew members were evacuated in lifeboats, many of which were launched only partially loaded. At 2.20 a.m., the ship broke apart and foundered, one thousand people were still aboard. Just under two hours after Titanic sank, RMS Carpathia arrived at the scene, where she brought aboard an estimated 705 survivors.

One of the most controversial issues examined by the inquiries was the role played by SS Californian, which had been only a few miles from Titanic but had not picked up her distress calls or responded to her signal rockets. Californian had warned Titanic by radio of the pack ice that was the reason Californian had stopped for the night, but was rebuked by Titanic's senior wireless operator, Jack Phillips.

So let's recap the data without any emotion:

- Started his voyage on April 10, 1912
- 2224 peoples on board, 1502 died
- 20 lifeboats, only enough for 1178 people
- Titanic hit an iceberg on 14.04 at 11.40 p.m.
- 2.20 am, the ship broke apart and foundered
- 2 hours later SS Carpathia saves an estimated number of 705 survivors
- The wreck of Titanic was discovered in 1985

We used Wikipedia as bibliographical reference for the story above.

## 1.2. Kaggle Competition Description

*The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.*

*One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.*

*In this challenge, we ask you to complete the analysis of what sorts of people were likely to survive. In particular, we ask you to apply the tools of machine learning to predict which passengers survived the tragedy.*

## 2. Acquire training and testing data

Download the data (<https://www.kaggle.com/c/titanic/data> (<https://www.kaggle.com/c/titanic/data>)).

The data has been split into two groups:

1. training set (train.csv)
2. test set (test.csv) \

*The test set should be used to see how well your model performs on unseen data. For the test set, we do not provide the ground truth for each passenger. It is your job to predict these outcomes. For each passenger in the test set, use the model you trained to predict whether or not they survived the sinking of the Titanic.*

We will need particular libraries/packages which have to be installed.

```
suppressPackageStartupMessages(library(ggplot2))
suppressPackageStartupMessages(library(data.table))
suppressPackageStartupMessages(library(Matrix))
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(e1071))
suppressPackageStartupMessages(library(party))
suppressPackageStartupMessages(library(randomForest))
suppressPackageStartupMessages(library(rpart))
suppressPackageStartupMessages(library(knitr))
suppressPackageStartupMessages(library(caret))
suppressPackageStartupMessages(library(rpart))
suppressPackageStartupMessages(library(rpart.plot))
suppressPackageStartupMessages(library(rattle))
suppressPackageStartupMessages(library(lattice))
suppressPackageStartupMessages(library(PerformanceAnalytics))

set.seed(3433)
setwd('C:/Users/ss_cr/')

# Import the train and test data from Computer

training <- read.csv(file="train.csv",na.strings=c("NA","#DIV/0!",""), header=TRUE)

testing1 <- read.csv(file="test.csv",na.strings=c("NA","#DIV/0!",""), header=TRUE)

is.data.frame(training)
```

```
## [1] TRUE
```

```
is.data.frame(testing1)
```

```
## [1] TRUE
```

```
dim(training)
```

```
## [1] 891 12
```

```
dim(testing1)
```

```
## [1] 418 11
```

The training set has 891 entries with 12 predictors and the test set has 418 entries, which means that exactly 418 predictions should be submitted in a csv file plus a header row.

In the testing database the Survived/Died column is missing.

### 3. Cleaning and Preparing the Data

You are now cleaning the data by doing the following:

1. Added missing Age and Fare values
2. Converted categorical variables to dummy variables
3. Created new variables to better fit a model

```
# missing age
```

```
training$Fare[is.na(testing1$Fare)] <- mean(testing1$Fare, na.rm=TRUE)  
training$Age[is.na(training$Age)] <- median(training$Age, na.rm=TRUE)
```

In the cabin variable we find a lot of missing values, near 77% of the registers, so we won't use it. Then, in the ticket variable, at first glance, there's nothing relevant so it will be removed too.

```
print(training$Cabin)
```

##	[1]	<NA>	C85	<NA>	C123
##	[5]	<NA>	<NA>	E46	<NA>
##	[9]	<NA>	<NA>	G6	C103
##	[13]	<NA>	<NA>	<NA>	<NA>
##	[17]	<NA>	<NA>	<NA>	<NA>
##	[21]	<NA>	D56	<NA>	A6
##	[25]	<NA>	<NA>	<NA>	C23 C25 C27
##	[29]	<NA>	<NA>	<NA>	B78
##	[33]	<NA>	<NA>	<NA>	<NA>
##	[37]	<NA>	<NA>	<NA>	<NA>
##	[41]	<NA>	<NA>	<NA>	<NA>
##	[45]	<NA>	<NA>	<NA>	<NA>
##	[49]	<NA>	<NA>	<NA>	<NA>
##	[53]	D33	<NA>	B30	C52
##	[57]	<NA>	<NA>	<NA>	<NA>
##	[61]	<NA>	B28	C83	<NA>
##	[65]	<NA>	<NA>	F33	<NA>
##	[69]	<NA>	<NA>	<NA>	<NA>
##	[73]	<NA>	<NA>	<NA>	F G73
##	[77]	<NA>	<NA>	<NA>	<NA>
##	[81]	<NA>	<NA>	<NA>	<NA>
##	[85]	<NA>	<NA>	<NA>	<NA>
##	[89]	C23 C25 C27	<NA>	<NA>	<NA>
##	[93]	E31	<NA>	<NA>	<NA>
##	[97]	A5	D10 D12	<NA>	<NA>
##	[101]	<NA>	<NA>	D26	<NA>
##	[105]	<NA>	<NA>	<NA>	<NA>
##	[109]	<NA>	<NA>	C110	<NA>
##	[113]	<NA>	<NA>	<NA>	<NA>
##	[117]	<NA>	<NA>	B58 B60	<NA>
##	[121]	<NA>	<NA>	<NA>	E101
##	[125]	D26	<NA>	<NA>	<NA>
##	[129]	F E69	<NA>	<NA>	<NA>
##	[133]	<NA>	<NA>	<NA>	<NA>
##	[137]	D47	C123	<NA>	B86
##	[141]	<NA>	<NA>	<NA>	<NA>
##	[145]	<NA>	<NA>	<NA>	<NA>
##	[149]	F2	<NA>	<NA>	C2
##	[153]	<NA>	<NA>	<NA>	<NA>
##	[157]	<NA>	<NA>	<NA>	<NA>
##	[161]	<NA>	<NA>	<NA>	<NA>
##	[165]	<NA>	<NA>	E33	<NA>
##	[169]	<NA>	<NA>	B19	<NA>
##	[173]	<NA>	<NA>	A7	<NA>
##	[177]	<NA>	C49	<NA>	<NA>
##	[181]	<NA>	<NA>	<NA>	F4
##	[185]	<NA>	A32	<NA>	<NA>
##	[189]	<NA>	<NA>	<NA>	<NA>

## [193]	<NA>	F2	B4	B80
## [197]	<NA>	<NA>	<NA>	<NA>
## [201]	<NA>	<NA>	<NA>	<NA>
## [205]	<NA>	G6	<NA>	<NA>
## [209]	<NA>	A31	<NA>	<NA>
## [213]	<NA>	<NA>	<NA>	D36
## [217]	<NA>	<NA>	D15	<NA>
## [221]	<NA>	<NA>	<NA>	<NA>
## [225]	C93	<NA>	<NA>	<NA>
## [229]	<NA>	<NA>	C83	<NA>
## [233]	<NA>	<NA>	<NA>	<NA>
## [237]	<NA>	<NA>	<NA>	<NA>
## [241]	<NA>	<NA>	<NA>	<NA>
## [245]	<NA>	C78	<NA>	<NA>
## [249]	D35	<NA>	<NA>	G6
## [253]	C87	<NA>	<NA>	<NA>
## [257]	<NA>	B77	<NA>	<NA>
## [261]	<NA>	<NA>	E67	B94
## [265]	<NA>	<NA>	<NA>	<NA>
## [269]	C125	C99	<NA>	<NA>
## [273]	<NA>	C118	<NA>	D7
## [277]	<NA>	<NA>	<NA>	<NA>
## [281]	<NA>	<NA>	<NA>	<NA>
## [285]	A19	<NA>	<NA>	<NA>
## [289]	<NA>	<NA>	<NA>	B49
## [293]	D	<NA>	<NA>	<NA>
## [297]	<NA>	C22 C26	C106	B58 B60
## [301]	<NA>	<NA>	<NA>	E101
## [305]	<NA>	C22 C26	<NA>	C65
## [309]	<NA>	E36	C54	B57 B59 B63 B66
## [313]	<NA>	<NA>	<NA>	<NA>
## [317]	<NA>	<NA>	C7	E34
## [321]	<NA>	<NA>	<NA>	<NA>
## [325]	<NA>	C32	<NA>	D
## [329]	<NA>	B18	<NA>	C124
## [333]	C91	<NA>	<NA>	<NA>
## [337]	C2	E40	<NA>	T
## [341]	F2	C23 C25 C27	<NA>	<NA>
## [345]	<NA>	F33	<NA>	<NA>
## [349]	<NA>	<NA>	<NA>	C128
## [353]	<NA>	<NA>	<NA>	<NA>
## [357]	E33	<NA>	<NA>	<NA>
## [361]	<NA>	<NA>	<NA>	<NA>
## [365]	<NA>	<NA>	D37	<NA>
## [369]	<NA>	B35	E50	<NA>
## [373]	<NA>	<NA>	<NA>	<NA>
## [377]	<NA>	C82	<NA>	<NA>
## [381]	<NA>	<NA>	<NA>	<NA>
## [385]	<NA>	<NA>	<NA>	<NA>

## [389]	<NA>	<NA>	B96 B98	<NA>
## [393]	<NA>	D36	G6	<NA>
## [397]	<NA>	<NA>	<NA>	<NA>
## [401]	<NA>	<NA>	<NA>	<NA>
## [405]	<NA>	<NA>	<NA>	<NA>
## [409]	<NA>	<NA>	<NA>	<NA>
## [413]	C78	<NA>	<NA>	<NA>
## [417]	<NA>	<NA>	<NA>	<NA>
## [421]	<NA>	<NA>	<NA>	<NA>
## [425]	<NA>	<NA>	<NA>	<NA>
## [429]	<NA>	E10	C52	<NA>
## [433]	<NA>	<NA>	E44	B96 B98
## [437]	<NA>	<NA>	C23 C25 C27	<NA>
## [441]	<NA>	<NA>	<NA>	<NA>
## [445]	<NA>	A34	<NA>	<NA>
## [449]	<NA>	C104	<NA>	<NA>
## [453]	C111	C92	<NA>	<NA>
## [457]	E38	D21	<NA>	<NA>
## [461]	E12	<NA>	E63	<NA>
## [465]	<NA>	<NA>	<NA>	<NA>
## [469]	<NA>	<NA>	<NA>	<NA>
## [473]	<NA>	D	<NA>	A14
## [477]	<NA>	<NA>	<NA>	<NA>
## [481]	<NA>	<NA>	<NA>	<NA>
## [485]	B49	<NA>	C93	B37
## [489]	<NA>	<NA>	<NA>	<NA>
## [493]	C30	<NA>	<NA>	<NA>
## [497]	D20	<NA>	C22 C26	<NA>
## [501]	<NA>	<NA>	<NA>	<NA>
## [505]	B79	C65	<NA>	<NA>
## [509]	<NA>	<NA>	<NA>	<NA>
## [513]	E25	<NA>	<NA>	D46
## [517]	F33	<NA>	<NA>	<NA>
## [521]	B73	<NA>	<NA>	B18
## [525]	<NA>	<NA>	<NA>	C95
## [529]	<NA>	<NA>	<NA>	<NA>
## [533]	<NA>	<NA>	<NA>	<NA>
## [537]	B38	<NA>	<NA>	B39
## [541]	B22	<NA>	<NA>	<NA>
## [545]	C86	<NA>	<NA>	<NA>
## [549]	<NA>	<NA>	C70	<NA>
## [553]	<NA>	<NA>	<NA>	<NA>
## [557]	A16	<NA>	E67	<NA>
## [561]	<NA>	<NA>	<NA>	<NA>
## [565]	<NA>	<NA>	<NA>	<NA>
## [569]	<NA>	<NA>	<NA>	C101
## [573]	E25	<NA>	<NA>	<NA>
## [577]	<NA>	E44	<NA>	<NA>
## [581]	<NA>	C68	<NA>	A10

## [585]	<NA>	E68	<NA>	B41
## [589]	<NA>	<NA>	<NA>	D20
## [593]	<NA>	<NA>	<NA>	<NA>
## [597]	<NA>	<NA>	<NA>	A20
## [601]	<NA>	<NA>	<NA>	<NA>
## [605]	<NA>	<NA>	<NA>	<NA>
## [609]	<NA>	C125	<NA>	<NA>
## [613]	<NA>	<NA>	<NA>	<NA>
## [617]	<NA>	<NA>	F4	<NA>
## [621]	<NA>	D19	<NA>	<NA>
## [625]	<NA>	D50	<NA>	D9
## [629]	<NA>	<NA>	A23	<NA>
## [633]	B50	<NA>	<NA>	<NA>
## [637]	<NA>	<NA>	<NA>	<NA>
## [641]	<NA>	B35	<NA>	<NA>
## [645]	<NA>	D33	<NA>	A26
## [649]	<NA>	<NA>	<NA>	<NA>
## [653]	<NA>	<NA>	<NA>	<NA>
## [657]	<NA>	<NA>	<NA>	D48
## [661]	<NA>	<NA>	E58	<NA>
## [665]	<NA>	<NA>	<NA>	<NA>
## [669]	<NA>	C126	<NA>	B71
## [673]	<NA>	<NA>	<NA>	<NA>
## [677]	<NA>	<NA>	<NA>	B51 B53 B55
## [681]	<NA>	D49	<NA>	<NA>
## [685]	<NA>	<NA>	<NA>	<NA>
## [689]	<NA>	B5	B20	<NA>
## [693]	<NA>	<NA>	<NA>	<NA>
## [697]	<NA>	<NA>	C68	F G63
## [701]	C62 C64	E24	<NA>	<NA>
## [705]	<NA>	<NA>	<NA>	E24
## [709]	<NA>	<NA>	C90	C124
## [713]	C126	<NA>	<NA>	F G73
## [717]	C45	E101	<NA>	<NA>
## [721]	<NA>	<NA>	<NA>	<NA>
## [725]	E8	<NA>	<NA>	<NA>
## [729]	<NA>	<NA>	B5	<NA>
## [733]	<NA>	<NA>	<NA>	<NA>
## [737]	<NA>	B101	<NA>	<NA>
## [741]	D45	C46	B57 B59 B63 B66	<NA>
## [745]	<NA>	B22	<NA>	<NA>
## [749]	D30	<NA>	<NA>	E121
## [753]	<NA>	<NA>	<NA>	<NA>
## [757]	<NA>	<NA>	<NA>	B77
## [761]	<NA>	<NA>	<NA>	B96 B98
## [765]	<NA>	D11	<NA>	<NA>
## [769]	<NA>	<NA>	<NA>	<NA>
## [773]	E77	<NA>	<NA>	<NA>
## [777]	F38	<NA>	<NA>	B3



## [781]	<NA>	B20	D6	<NA>
## [785]	<NA>	<NA>	<NA>	<NA>
## [789]	<NA>	B82 B84	<NA>	<NA>
## [793]	<NA>	<NA>	<NA>	<NA>
## [797]	D17	<NA>	<NA>	<NA>
## [801]	<NA>	<NA>	B96 B98	<NA>
## [805]	<NA>	<NA>	A36	<NA>
## [809]	<NA>	E8	<NA>	<NA>
## [813]	<NA>	<NA>	<NA>	B102
## [817]	<NA>	<NA>	<NA>	<NA>
## [821]	B69	<NA>	<NA>	E121
## [825]	<NA>	<NA>	<NA>	<NA>
## [829]	<NA>	B28	<NA>	<NA>
## [833]	<NA>	<NA>	<NA>	E49
## [837]	<NA>	<NA>	<NA>	C47
## [841]	<NA>	<NA>	<NA>	<NA>
## [845]	<NA>	<NA>	<NA>	<NA>
## [849]	<NA>	C92	<NA>	<NA>
## [853]	<NA>	D28	<NA>	<NA>
## [857]	<NA>	E17	<NA>	<NA>
## [861]	<NA>	<NA>	D17	<NA>
## [865]	<NA>	<NA>	<NA>	A24
## [869]	<NA>	<NA>	<NA>	D35
## [873]	B51 B53 B55	<NA>	<NA>	<NA>
## [877]	<NA>	<NA>	<NA>	C50
## [881]	<NA>	<NA>	<NA>	<NA>
## [885]	<NA>	<NA>	<NA>	B42
## [889]	<NA>	C148	<NA>	
## 147 Levels:	A10 A14 A16 A19 A20 A23 A24 A26 A31 A32 A34 A36 A5 A6 ... T			

```
print(training$Ticket)
```

##	[1]	A/5 21171	PC 17599	STON/O2. 3101282
##	[4]	113803	373450	330877
##	[7]	17463	349909	347742
##	[10]	237736	PP 9549	113783
##	[13]	A/5. 2151	347082	350406
##	[16]	248706	382652	244373
##	[19]	345763	2649	239865
##	[22]	248698	330923	113788
##	[25]	349909	347077	2631
##	[28]	19950	330959	349216
##	[31]	PC 17601	PC 17569	335677
##	[34]	C.A. 24579	PC 17604	113789
##	[37]	2677	A./5. 2152	345764
##	[40]	2651	7546	11668
##	[43]	349253	SC/Paris 2123	330958
##	[46]	S.C./A.4. 23567	370371	14311
##	[49]	2662	349237	3101295
##	[52]	A/4. 39886	PC 17572	2926
##	[55]	113509	19947	C.A. 31026
##	[58]	2697	C.A. 34651	CA 2144
##	[61]	2669	113572	36973
##	[64]	347088	PC 17605	2661
##	[67]	C.A. 29395	S.P. 3464	3101281
##	[70]	315151	C.A. 33111	CA 2144
##	[73]	S.O.C. 14879	2680	1601
##	[76]	348123	349208	374746
##	[79]	248738	364516	345767
##	[82]	345779	330932	113059
##	[85]	SO/C 14885	3101278	W./C. 6608
##	[88]	SOTON/OQ 392086	19950	343275
##	[91]	343276	347466	W.E.P. 5734
##	[94]	C.A. 2315	364500	374910
##	[97]	PC 17754	PC 17759	231919
##	[100]	244367	349245	349215
##	[103]	35281	7540	3101276
##	[106]	349207	343120	312991
##	[109]	349249	371110	110465
##	[112]	2665	324669	4136
##	[115]	2627	STON/O 2. 3101294	370369
##	[118]	11668	PC 17558	347082
##	[121]	S.O.C. 14879	A4. 54510	237736
##	[124]	27267	35281	2651
##	[127]	370372	C 17369	2668
##	[130]	347061	349241	SOTON/O.Q. 3101307
##	[133]	A/5. 3337	228414	C.A. 29178
##	[136]	SC/PARIS 2133	11752	113803
##	[139]	7534	PC 17593	2678
##	[142]	347081	STON/O2. 3101279	365222

## [145]	231945	C.A. 33112	350043
## [148]	W./C. 6608	230080	244310
## [151]	S.O.P. 1166	113776	A.5. 11206
## [154]	A/5. 851	Fa 265302	PC 17597
## [157]	35851	SOTON/OQ 392090	315037
## [160]	CA. 2343	371362	C.A. 33595
## [163]	347068	315093	3101295
## [166]	363291	113505	347088
## [169]	PC 17318	1601	111240
## [172]	382652	347742	STON/O 2. 3101280
## [175]	17764	350404	4133
## [178]	PC 17595	250653	LINE
## [181]	CA. 2343	SC/PARIS 2131	347077
## [184]	230136	315153	113767
## [187]	370365	111428	364849
## [190]	349247	234604	28424
## [193]	350046	230080	PC 17610
## [196]	PC 17569	368703	4579
## [199]	370370	248747	345770
## [202]	CA. 2343	3101264	2628
## [205]	A/5 3540	347054	3101278
## [208]	2699	367231	112277
## [211]	SOTON/O.Q. 3101311	F.C.C. 13528	A/5 21174
## [214]	250646	367229	35273
## [217]	STON/O2. 3101283	243847	11813
## [220]	W/C 14208	SOTON/OQ 392089	220367
## [223]	21440	349234	19943
## [226]	PP 4348	SW/PP 751	A/5 21173
## [229]	236171	4133	36973
## [232]	347067	237442	347077
## [235]	C.A. 29566	W./C. 6609	26707
## [238]	C.A. 31921	28665	SCO/W 1585
## [241]	2665	367230	W./C. 14263
## [244]	STON/O 2. 3101275	2694	19928
## [247]	347071	250649	11751
## [250]	244252	362316	347054
## [253]	113514	A/5. 3336	370129
## [256]	2650	PC 17585	110152
## [259]	PC 17755	230433	384461
## [262]	347077	110413	112059
## [265]	382649	C.A. 17248	3101295
## [268]	347083	PC 17582	PC 17760
## [271]	113798	LINE	250644
## [274]	PC 17596	370375	13502
## [277]	347073	239853	382652
## [280]	C.A. 2673	336439	347464
## [283]	345778	A/5. 10482	113056
## [286]	349239	345774	349206
## [289]	237798	370373	19877

## [292]	11967	SC/Paris 2163	349236
## [295]	349233	PC 17612	2693
## [298]	113781	19988	PC 17558
## [301]	9234	367226	LINE
## [304]	226593	A/5 2466	113781
## [307]	17421	PC 17758	P/PP 3381
## [310]	PC 17485	11767	PC 17608
## [313]	250651	349243	F.C.C. 13529
## [316]	347470	244367	29011
## [319]	36928	16966	A/5 21172
## [322]	349219	234818	248738
## [325]	CA. 2343	PC 17760	345364
## [328]	28551	363291	111361
## [331]	367226	113043	PC 17582
## [334]	345764	PC 17611	349225
## [337]	113776	16966	7598
## [340]	113784	230080	19950
## [343]	248740	244361	229236
## [346]	248733	31418	386525
## [349]	C.A. 37671	315088	7267
## [352]	113510	2695	349237
## [355]	2647	345783	113505
## [358]	237671	330931	330980
## [361]	347088	SC/PARIS 2167	2691
## [364]	SOTON/O.Q. 3101310	370365	C 7076
## [367]	110813	2626	14313
## [370]	PC 17477	11765	3101267
## [373]	323951	PC 17760	349909
## [376]	PC 17604	C 7077	113503
## [379]	2648	347069	PC 17757
## [382]	2653	STON/O 2. 3101293	113789
## [385]	349227	S.O.C. 14879	CA 2144
## [388]	27849	367655	SC 1748
## [391]	113760	350034	3101277
## [394]	35273	PP 9549	350052
## [397]	350407	28403	244278
## [400]	240929	STON/O 2. 3101289	341826
## [403]	4137	STON/O2. 3101279	315096
## [406]	28664	347064	29106
## [409]	312992	4133	349222
## [412]	394140	19928	239853
## [415]	STON/O 2. 3101269	343095	28220
## [418]	250652	28228	345773
## [421]	349254	A/5. 13032	315082
## [424]	347080	370129	A/4. 34244
## [427]	2003	250655	364851
## [430]	SOTON/O.Q. 392078	110564	376564
## [433]	SC/AH 3085	STON/O 2. 3101274	13507
## [436]	113760	W./C. 6608	29106

## [439]	19950	C.A. 18723	F.C.C. 13529
## [442]	345769	347076	230434
## [445]	65306	33638	250644
## [448]	113794	2666	113786
## [451]	C.A. 34651	65303	113051
## [454]	17453	A/5 2817	349240
## [457]	13509	17464	F.C.C. 13531
## [460]	371060	19952	364506
## [463]	111320	234360	A/S 2816
## [466]	SOTON/O.Q. 3101306	239853	113792
## [469]	36209	2666	323592
## [472]	315089	C.A. 34651	SC/AH Basle 541
## [475]	7553	110465	31027
## [478]	3460	350060	3101298
## [481]	CA 2144	239854	A/5 3594
## [484]	4134	11967	4133
## [487]	19943	11771	A.5. 18509
## [490]	C.A. 37671	65304	SOTON/OQ 3101317
## [493]	113787	PC 17609	A/4 45380
## [496]	2627	36947	C.A. 6212
## [499]	113781	350035	315086
## [502]	364846	330909	4135
## [505]	110152	PC 17758	26360
## [508]	111427	C 4001	1601
## [511]	382651	SOTON/OQ 3101316	PC 17473
## [514]	PC 17603	349209	36967
## [517]	C.A. 34260	371110	226875
## [520]	349242	12749	349252
## [523]	2624	111361	2700
## [526]	367232	W./C. 14258	PC 17483
## [529]	3101296	29104	26360
## [532]	2641	2690	2668
## [535]	315084	F.C.C. 13529	113050
## [538]	PC 17761	364498	13568
## [541]	WE/P 5735	347082	347082
## [544]	2908	PC 17761	693
## [547]	2908	SC/PARIS 2146	363291
## [550]	C.A. 33112	17421	244358
## [553]	330979	2620	347085
## [556]	113807	11755	PC 17757
## [559]	110413	345572	372622
## [562]	349251	218629	SOTON/OQ 392082
## [565]	SOTON/O.Q. 392087	A/4 48871	349205
## [568]	349909	2686	350417
## [571]	S.W./PP 752	11769	PC 17474
## [574]	14312	A/4. 20589	358585
## [577]	243880	13507	2689
## [580]	STON/O 2. 3101286	237789	17421
## [583]	28403	13049	3411

## [586]	110413	237565	13567
## [589]	14973	A./5. 3235	STON/O 2. 3101273
## [592]	36947	A/5 3902	364848
## [595]	SC/AH 29037	345773	248727
## [598]	LINE	2664	PC 17485
## [601]	243847	349214	113796
## [604]	364511	111426	349910
## [607]	349246	113804	SC/Paris 2123
## [610]	PC 17582	347082	SOTON/O.Q. 3101305
## [613]	367230	370377	364512
## [616]	220845	347080	A/5. 3336
## [619]	230136	31028	2659
## [622]	11753	2653	350029
## [625]	54636	36963	219533
## [628]	13502	349224	334912
## [631]	27042	347743	13214
## [634]	112052	347088	237668
## [637]	STON/O 2. 3101292	C.A. 31921	3101295
## [640]	376564	350050	PC 17477
## [643]	347088	1601	2666
## [646]	PC 17572	349231	13213
## [649]	S.O./P.P. 751	CA. 2314	349221
## [652]	231919	8475	330919
## [655]	365226	S.O.C. 14879	349223
## [658]	364849	29751	35273
## [661]	PC 17611	2623	5727
## [664]	349210	STON/O 2. 3101285	S.O.C. 14879
## [667]	234686	312993	A/5 3536
## [670]	19996	29750	F.C. 12750
## [673]	C.A. 24580	244270	239856
## [676]	349912	342826	4138
## [679]	CA 2144	PC 17755	330935
## [682]	PC 17572	6563	CA 2144
## [685]	29750	SC/Paris 2123	3101295
## [688]	349228	350036	24160
## [691]	17474	349256	1601
## [694]	2672	113800	248731
## [697]	363592	35852	17421
## [700]	348121	PC 17757	PC 17475
## [703]	2691	36864	350025
## [706]	250655	223596	PC 17476
## [709]	113781	2661	PC 17482
## [712]	113028	19996	7545
## [715]	250647	348124	PC 17757
## [718]	34218	36568	347062
## [721]	248727	350048	12233
## [724]	250643	113806	315094
## [727]	31027	36866	236853
## [730]	STON/O2. 3101271	24160	2699

## [733]	239855	28425	233639
## [736]	54636	W./C. 6608	PC 17755
## [739]	349201	349218	16988
## [742]	19877	PC 17608	376566
## [745]	STON/O 2. 3101288	WE/P 5735	C.A. 2673
## [748]	250648	113773	335097
## [751]	29103	392096	345780
## [754]	349204	220845	250649
## [757]	350042	29108	363294
## [760]	110152	358585	SOTON/O2 3101272
## [763]	2663	113760	347074
## [766]	13502	112379	364850
## [769]	371110	8471	345781
## [772]	350047	S.O./P.P. 3	2674
## [775]	29105	347078	383121
## [778]	364516	36865	24160
## [781]	2687	17474	113501
## [784]	W./C. 6607	SOTON/O.Q. 3101312	374887
## [787]	3101265	382652	C.A. 2315
## [790]	PC 17593	12460	239865
## [793]	CA. 2343	PC 17600	349203
## [796]	28213	17465	349244
## [799]	2685	345773	250647
## [802]	C.A. 31921	113760	2625
## [805]	347089	347063	112050
## [808]	347087	248723	113806
## [811]	3474	A/4 48871	28206
## [814]	347082	364499	112058
## [817]	STON/O2. 3101290	S.C./PARIS 2079	C 7075
## [820]	347088	12749	315098
## [823]	19972	392096	3101295
## [826]	368323	1601	S.C./PARIS 2079
## [829]	367228	113572	2659
## [832]	29106	2671	347468
## [835]	2223	PC 17756	315097
## [838]	392092	1601	11774
## [841]	SOTON/O2 3101287	S.O./P.P. 3	113798
## [844]	2683	315090	C.A. 5547
## [847]	CA. 2343	349213	248727
## [850]	17453	347082	347060
## [853]	2678	PC 17592	244252
## [856]	392091	36928	113055
## [859]	2666	2629	350026
## [862]	28134	17466	CA. 2343
## [865]	233866	236852	SC/PARIS 2149
## [868]	PC 17590	345777	347742
## [871]	349248	11751	695
## [874]	345765	P/PP 3381	2667
## [877]	7534	349212	349217

```
## [880] 11767                230433                349257
## [883] 7552                    C.A./SOTON 34068      SOTON/OQ 392076
## [886] 382652                  211536                112053
## [889] W./C. 6607              111369                370376
## 681 Levels: 110152 110413 110465 110564 110813 111240 111320 ... WE/P 5735
```

```
training$Cabin<-NULL
training$Ticket<-NULL
```

We add a new column with the Title extracted from the Name column. This feature engineering process will add additional information for each passenger regarding the social status of the passenger. Then we group all titles beside the classical one in only one category.

Many variables are categorical and we have to transform them to numeric using factor.

```
training$Title <- gsub('(.*, )|(\\.\\.\\.*)', '', training$Name)
different_titles<-c('Don', 'Rev', 'Capt', 'Col', 'Dr', 'Major', 'Jonkheer', 'Sir', 'Lady', 'the Countess' )

training$Title[training$Title %in% different_titles] <- 'Titles'

training$Title<-as.factor(training$Title)

training$Sex<-as.factor(training$Sex)

levels(training$Sex)<-c("Female", "Male")

training$Survived<-factor(training$Survived, levels=c(1,0))

levels(training$Survived)<-c("Survived", "Died")
```

Adding a new column with AgeG category: Child (<16), Young (16,25), Adult (25,55), Old (>55)

```
training$AgeG <- training$Age
training$AgeG <- ifelse(training$AgeG<=16, "Child", ifelse(training$AgeG<=25,
"Young", ifelse(training$AgeG<=55, "Adult", ifelse(training$AgeG>55, "Old", "UnknownAge"))))
```

## 4. Analyze, identify patterns, explore the data

Let's explore the data frame, have a first look at the passenger distribution

```
table(training$Sex, training$Title)
```



```
##
##           Master Miss Mlle Mme  Mr Mrs  Ms Titles
##   Female      0  182    2   1   0 125    1      3
##   Male       40    0    0   0 517    0    0     20
```

As you can see there were at least 20 male wearing a title on Titanic.

```
table(training$Sex, training$AgeG)
```

```
##
##           Adult Child Old Young
##   Female    183    49   9    73
##   Male     367    51  31   128
```

In the training set, the number of adults men was almost twice the number of adult women. The situation is almost the same for young passengers, yet the number of children was more or less the same. However, there were much more old men (31) on the Titanic compared to women(9). 27 of those old men died, 8 old women survived.

```
table(training$AgeG, training$Sex, training$Survived)
```

```
## , , = Survived
##
##
##           Female Male
##   Adult     138   68
##   Child      33   22
##   Old         8    4
##   Young     54   15
##
## , , = Died
##
##
##           Female Male
##   Adult       45  299
##   Child       16   29
##   Old          1   27
##   Young       19  113
```

Clearly, most people were travelling in second and third class. The first class was in fact the only class where the number of people surviving was larger than the number of people dying.

```
table(training$Pclass, training$Survived)
```

```
##
##      Survived Died
##    1      136   80
##    2       87   97
##    3      119  372
```

```
table(training$AgeG, training$Survived)
```

```
##
##      Survived Died
##  Adult      206  344
##  Child       55   45
##   Old       12   28
##  Young       69  132
```

Most of the people that died in the disaster were young people, older than 16 and adults, younger than 55.

May be the title did not count if you want to get on a boat!

```
table(training$Title, training$Survived)
```

```
##
##      Survived Died
##  Master       23   17
##   Miss      127   55
##   Mlle        2    0
##   Mme         1    0
##   Mr         81  436
##   Mrs         99   26
##   Ms          1    0
##  Titles        8   15
```

This analysis of the data can be also done by summarizing some of the passengers characteristics and also visualizing some of the fact. Data Analysis: according to Age variable

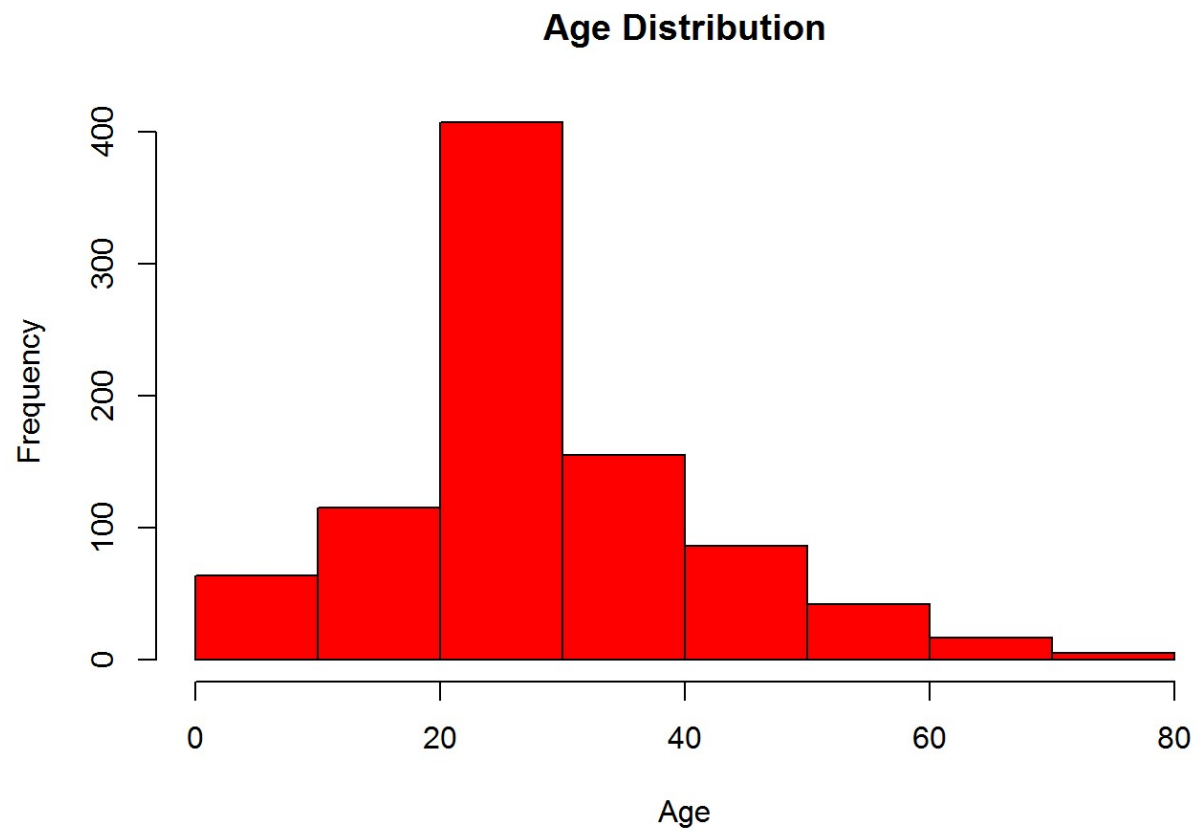
```
summary(training$Age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.42   22.00   28.00   29.36   35.00   80.00
```

We can see that for the variable Age there are 117 missing values. We can consider filling this NA values with the median value of all ages.

We can analyze the age distribution of all passengers using a histogram

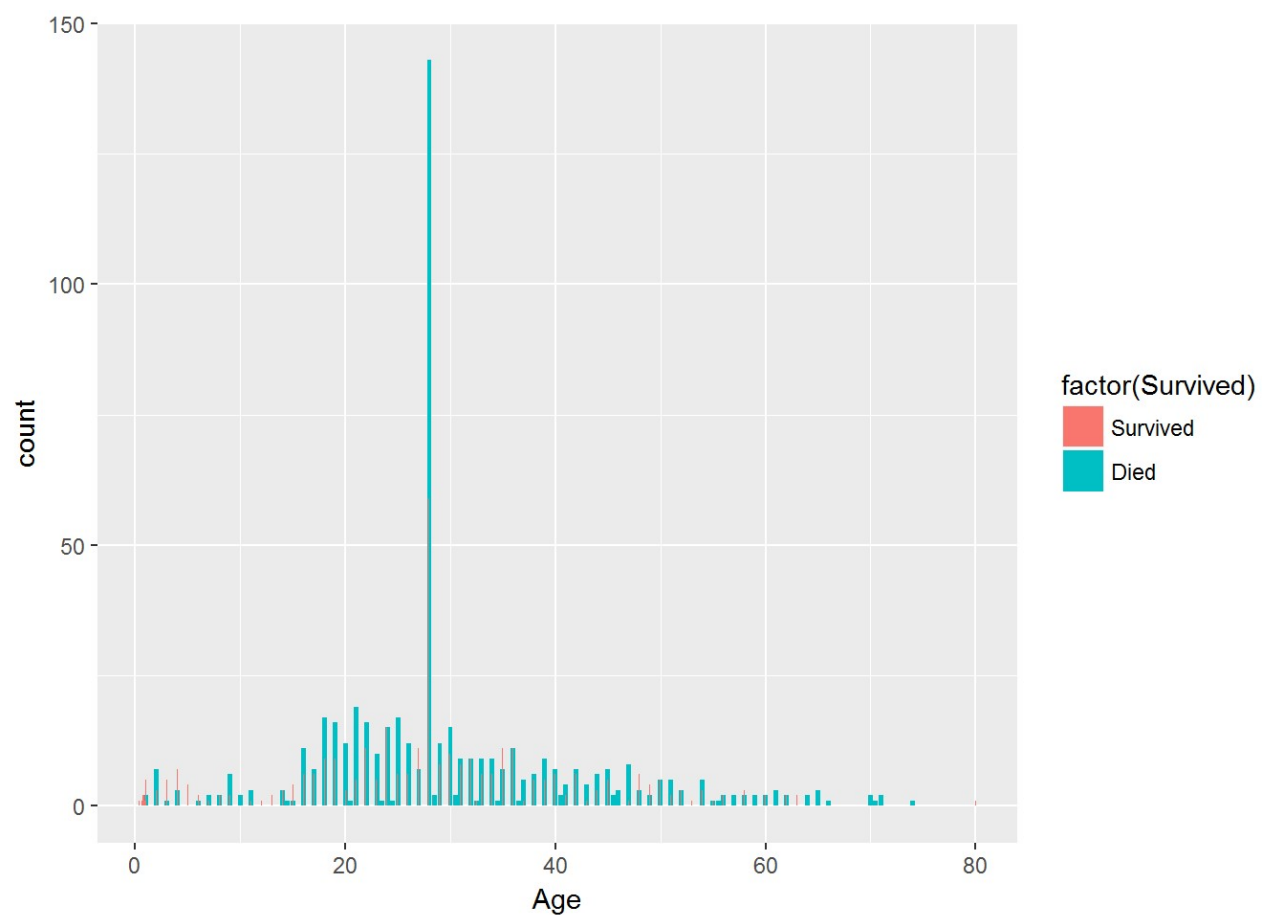
```
hist(training$Age, col=c("red"), main="Age Distribution", xlab="Age")
```



We can also use ggplot2 to visualize the Age distribution compared with the survived attribute

```
ggplot(training[1:891,], aes(x = Age, fill = factor(Survived))) +  
  geom_bar(stat='count')
```

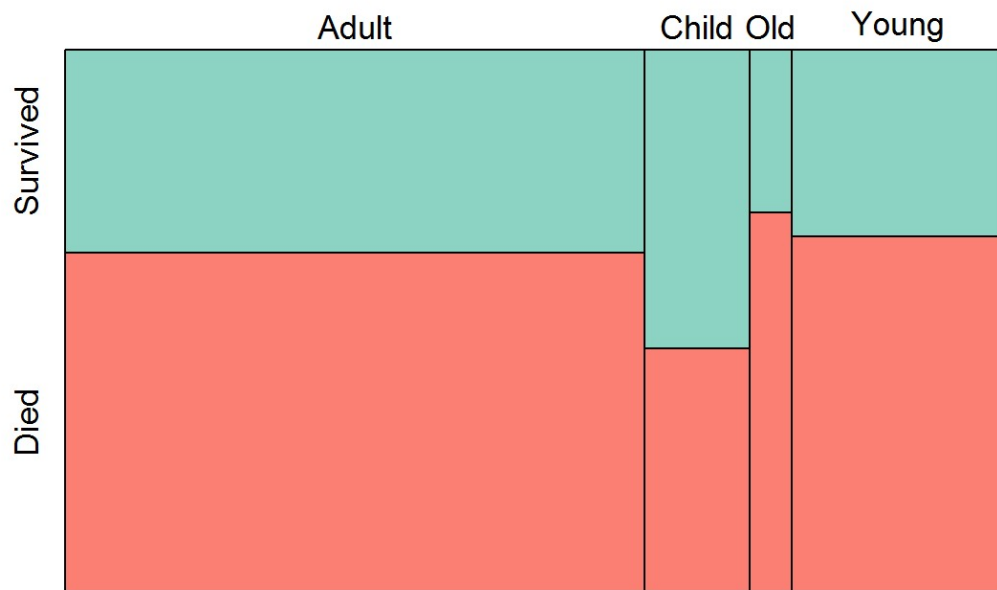
```
## Warning: position_stack requires non-overlapping x intervals
```



The plot is not very clear, so we can use the AgeG predictor instead.

```
mosaicplot(training$AgeG ~ training$Survived, main="Passenger Survival by Age",  
            color=c("#8dd3c3", "#fb8073"), shade=FALSE, xlab="", ylab="",  
            off=c(0), cex.axis=1.1)
```

## Passenger Survival by Age



```
#dev.off()
```

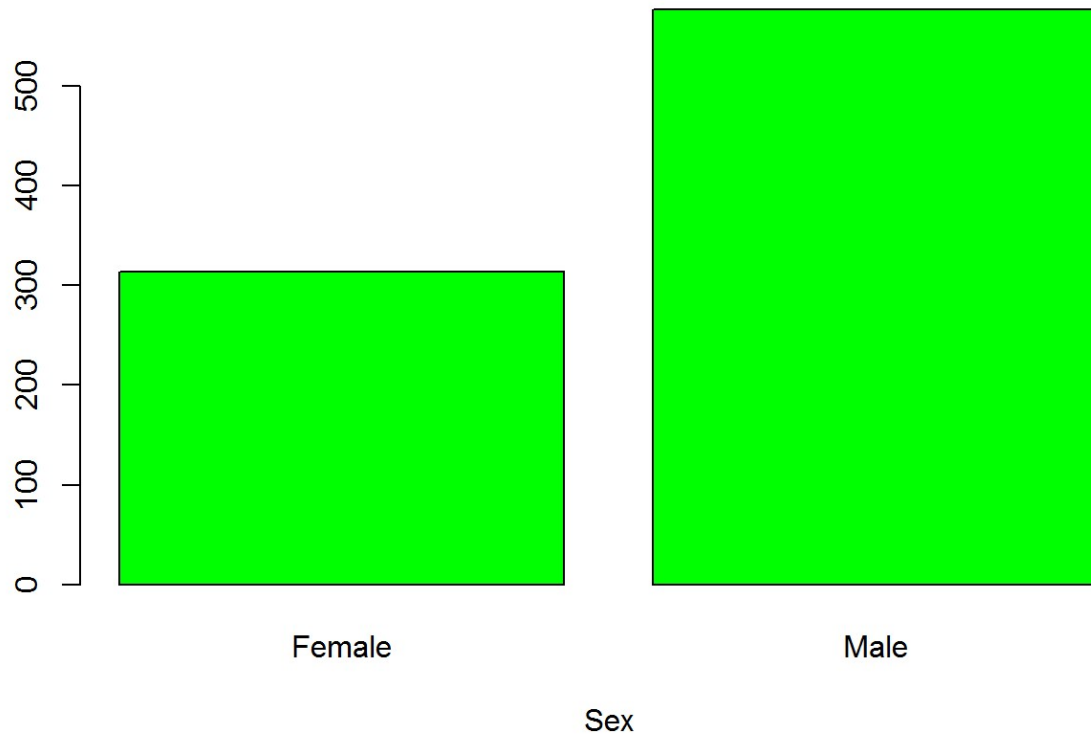
Data Analysis : according to Sex variable

```
summary(training$Sex)
```

```
## Female   Male  
##    314    577
```

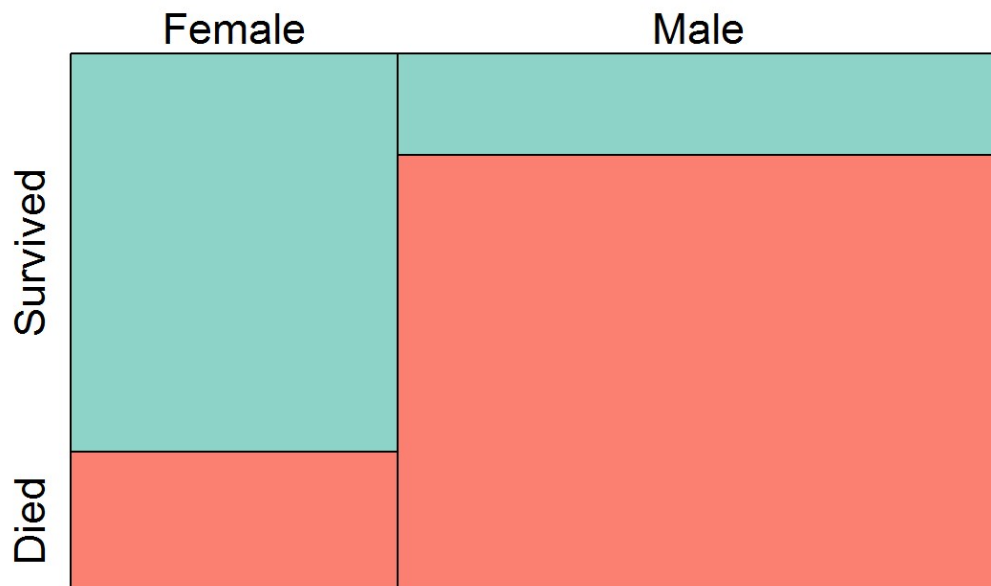
```
barplot(table(training$Sex), col=c("green"), main="Sex Distribution among Passa  
nger", xlab="Sex")
```

## Sex Distribution among Passanger



```
mosaicplot(training$Sex ~ training$Survived, main="Passenger Survival by Sex",  
            color=c("#8dd3c7", "#fb8072"), shade=FALSE, xlab="", ylab="",  
            off=c(0), cex.axis=1.4)
```

## Passenger Survival by Sex



```
dev.off()
```

```
## null device  
##           1
```

We can see that being a women clearly increased the chances of surviving! Almost 75% of women survived and almost 75% of men died. You can also see that being a child influenced the rate of surviving.

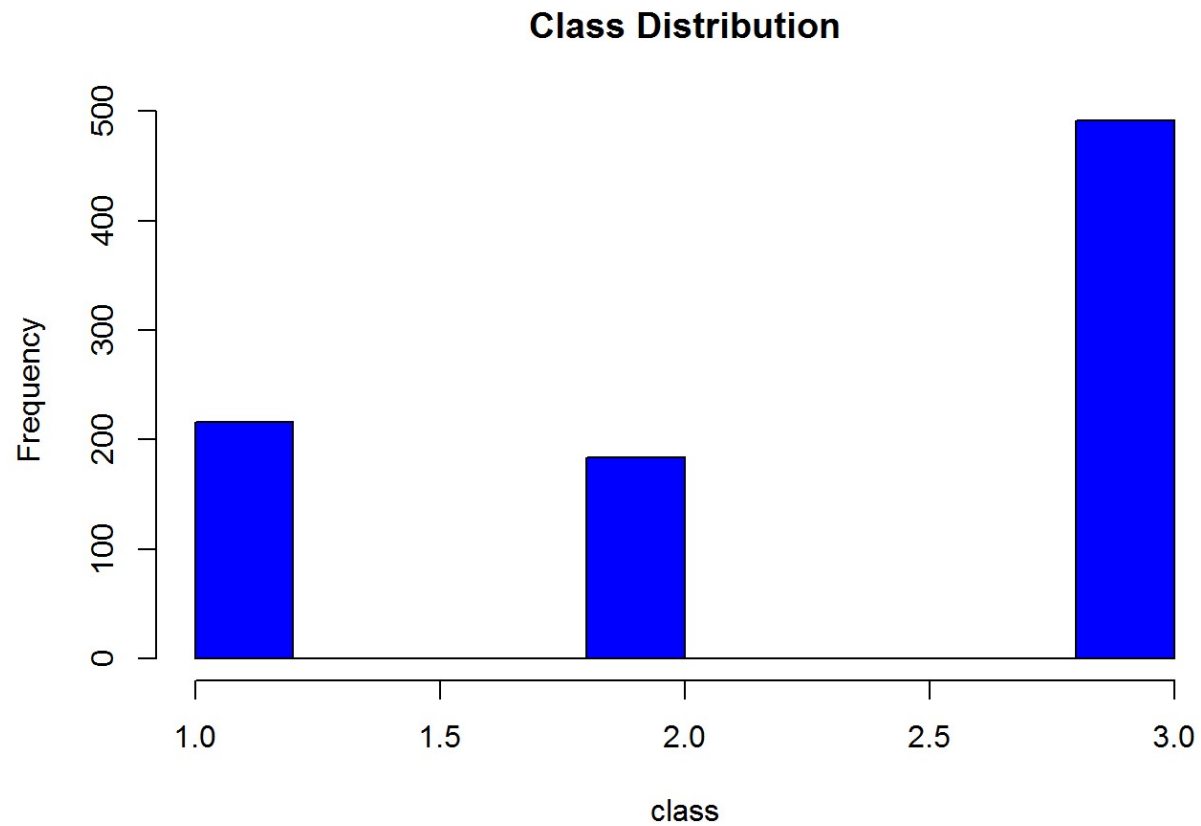
Data Analysis : according to class variable

Most of the passengers were travelling in 3rd class!

```
summary(training$Pclass)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##  1.000   2.000   3.000   2.309   3.000   3.000
```

```
hist(training$Pclass, col=c("blue"), main="Class Distribution", xlab="class")
```



```
table(training$Parch, training$Survived)
```

```
##  
##      Survived Died  
## 0         233  445  
## 1          65   53  
## 2          40   40  
## 3           3    2  
## 4           0    4  
## 5           1    4  
## 6           0    1
```

It looks like the nb of Siblings has an influence on the death rate.

```
table(training$SibSp, training$Survived)
```



```
##
##      Survived Died
##    0         210  398
##    1         112   97
##    2          13   15
##    3           4   12
##    4           3   15
##    5           0    5
##    8           0    7
```

We are now ready to build a model which will make predictions!

## 5. Model, predict and solve the problem

Partitioning the data into training and testing data: Data are divided into training data (60%) and testing data (40%).

```
inTrain <- createDataPartition(y=training$Survived, p=0.6, list=FALSE)
myTraining <- training[inTrain, ]
myTesting <- training[-inTrain, ]
dim(myTraining); dim(myTesting)
```

```
## [1] 536  12
```

```
## [1] 355  12
```

## Building the model - A model with a continuous response- ANOVA

We are presenting here more model that could be build fore predictions.

The rpart function allows the fitting of a classification tree. The minsplit value which determines the minimal number of observations per leaf that must exist in a node in order for a split to be attempted is here 20 (the default value). The complexity parameter responsible for saving computer time is 0.02. With anova splitting this means that the overall R-squared must increase by 0.02 at each step.

Each node shows

- the predicted value,
- the percentage of observations in the node.

```
Fit1 <- rpart(Survived ~ Sex+Age+Pclass+Fare+SibSp, data = myTraining, method
="anova", control = rpart.control(minsplit=20, cp=0.02), )
summary(Fit1)
```

```

## Call:
## rpart(formula = Survived ~ Sex + Age + Pclass + Fare + SibSp,
##       data = myTraining, method = "anova", control = rpart.control(minsplit =
20,
##       cp = 0.02))
##       n= 536
##
##           CP nsplit rel error      xerror      xstd
## 1 0.28265555      0 1.0000000 1.0056273 0.02071155
## 2 0.08277997      1 0.7173444 0.7213648 0.04215957
## 3 0.02978817      2 0.6345645 0.6419333 0.04135301
## 4 0.02471935      3 0.6047763 0.6555156 0.04272550
## 5 0.02433511      4 0.5800570 0.6434945 0.04246197
## 6 0.02000000      5 0.5557219 0.6330517 0.04250716
##
## Variable importance
##   Sex Pclass   Fare   Age  SibSp
##   48    19    19     9     4
##
## Node number 1: 536 observations,      complexity param=0.2826556
##   mean=1.615672, MSE=0.2366201
##   left son=2 (198 obs) right son=3 (338 obs)
##   Primary splits:
##     Sex      splits as LR,      improve=0.28265560, (0 missing)
##     Pclass < 2.5      to the left, improve=0.12024490, (0 missing)
##     Fare   < 10.48125 to the right, improve=0.09669118, (0 missing)
##     SibSp  < 2.5      to the left, improve=0.02404139, (0 missing)
##     Age    < 8.5      to the left, improve=0.01984063, (0 missing)
##   Surrogate splits:
##     Fare < 77.6229 to the right, agree=0.664, adj=0.091, (0 split)
##     Age  < 6.5      to the left, agree=0.640, adj=0.025, (0 split)
##     SibSp < 0.5      to the right, agree=0.634, adj=0.010, (0 split)
##
## Node number 2: 198 observations,      complexity param=0.08277997
##   mean=1.277778, MSE=0.2006173
##   left son=4 (107 obs) right son=5 (91 obs)
##   Primary splits:
##     Pclass < 2.5      to the left, improve=0.26430660, (0 missing)
##     SibSp  < 2.5      to the left, improve=0.09925558, (0 missing)
##     Fare   < 36.6875 to the right, improve=0.08193980, (0 missing)
##     Age    < 31.5      to the right, improve=0.03578393, (0 missing)
##   Surrogate splits:
##     Fare < 25.69795 to the right, agree=0.823, adj=0.615, (0 split)
##     Age  < 28.5      to the right, agree=0.652, adj=0.242, (0 split)
##     SibSp < 1.5      to the left, agree=0.591, adj=0.110, (0 split)
##
## Node number 3: 338 observations,      complexity param=0.02978817
##   mean=1.813609, MSE=0.1516491

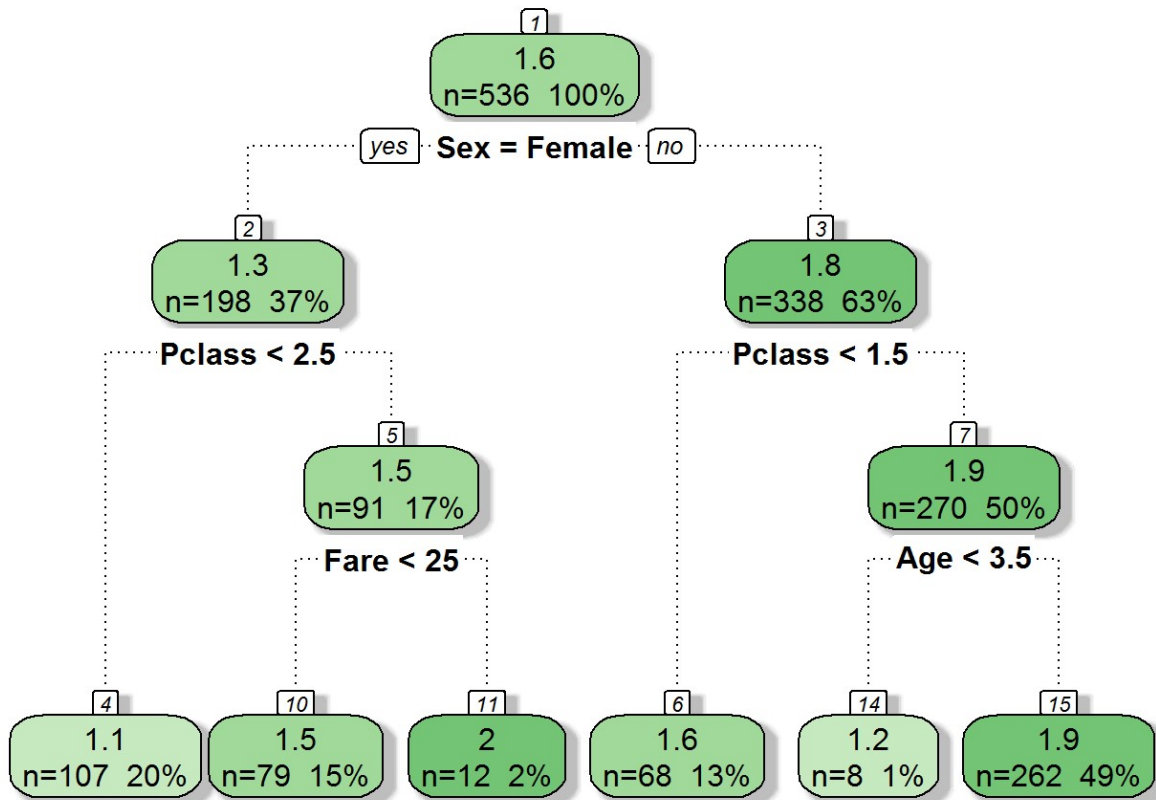
```

```

## left son=6 (68 obs) right son=7 (270 obs)
## Primary splits:
## Pclass < 1.5 to the left, improve=0.07370614, (0 missing)
## Fare < 26.14375 to the right, improve=0.06537612, (0 missing)
## Age < 3.5 to the left, improve=0.05078001, (0 missing)
## SibSp < 0.5 to the right, improve=0.01634379, (0 missing)
## Surrogate splits:
## Fare < 26.14375 to the right, agree=0.870, adj=0.353, (0 split)
## Age < 44.5 to the right, agree=0.811, adj=0.059, (0 split)
##
## Node number 4: 107 observations
## mean=1.065421, MSE=0.06114071
##
## Node number 5: 91 observations, complexity param=0.02433511
## mean=1.527473, MSE=0.2492453
## left son=10 (79 obs) right son=11 (12 obs)
## Primary splits:
## Fare < 24.80835 to the left, improve=0.13607590, (0 missing)
## SibSp < 2.5 to the left, improve=0.08034047, (0 missing)
## Age < 37.5 to the left, improve=0.03633721, (0 missing)
## Surrogate splits:
## SibSp < 2.5 to the left, agree=0.945, adj=0.583, (0 split)
##
## Node number 6: 68 observations
## mean=1.602941, MSE=0.2394031
##
## Node number 7: 270 observations, complexity param=0.02471935
## mean=1.866667, MSE=0.1155556
## left son=14 (8 obs) right son=15 (262 obs)
## Primary splits:
## Age < 3.5 to the left, improve=0.100484400, (0 missing)
## Fare < 7.74375 to the right, improve=0.017851610, (0 missing)
## SibSp < 2.5 to the left, improve=0.009691096, (0 missing)
## Pclass < 2.5 to the left, improve=0.002717560, (0 missing)
##
## Node number 10: 79 observations
## mean=1.455696, MSE=0.2480372
##
## Node number 11: 12 observations
## mean=2, MSE=0
##
## Node number 14: 8 observations
## mean=1.25, MSE=0.1875
##
## Node number 15: 262 observations
## mean=1.885496, MSE=0.1013927

```

```
fancyRpartPlot(Fit1)
```



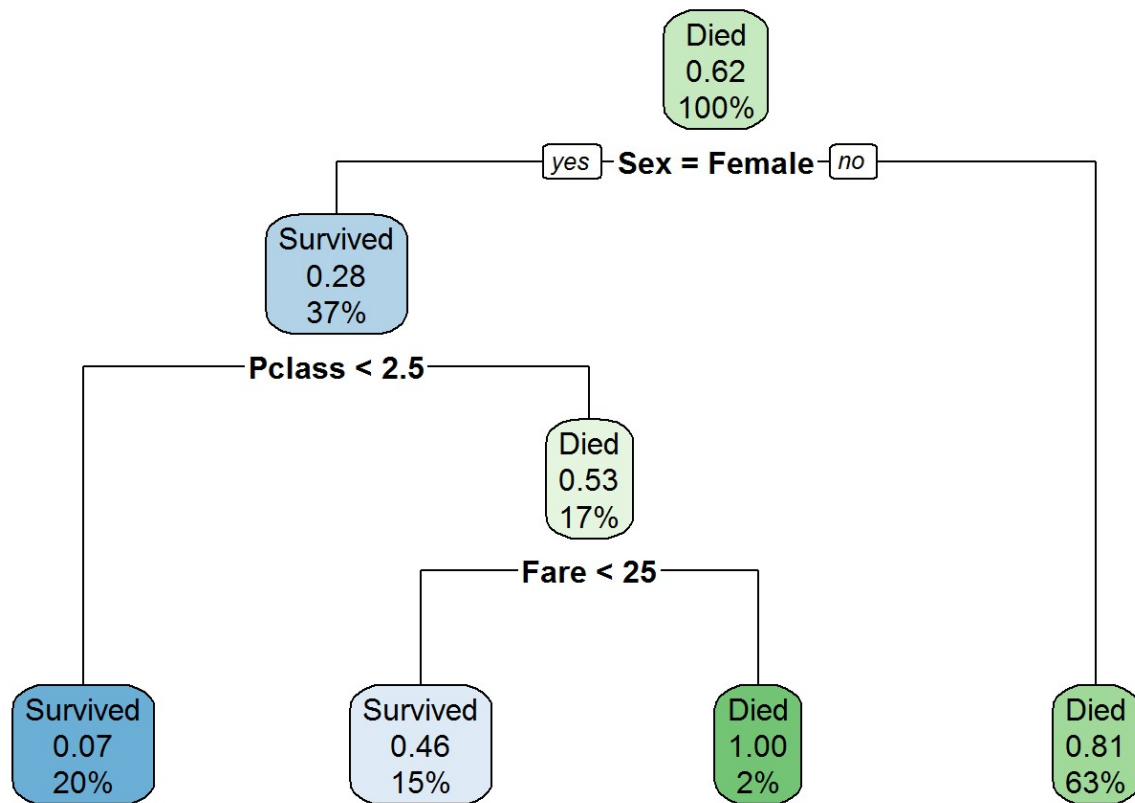
Rattle 2017-Apr-24 13:13:41 johnson

# Building the model - A model with binary response

Each node shows

- the predicted class (died or survived),
- the predicted probability of survival,
- the percentage of observations in the node

```
Fit2 <- rpart(Survived ~ Sex+Age+Pclass+Fare+SibSp, data = myTraining, cp=0.02)
rpart.plot(Fit2)
```



## Building the model- gbm and random forests and compare

The trainControl functions regarding the computational nuances of the train function will be considered the cross validation method and the number of k-folds is 3.

```
objControl <- trainControl(method="cv", number=3)
```

```
Fit3 <- train(Survived ~ Sex+Age+Pclass+Fare+SibSp, trControl=objControl, data = myTraining, method="gbm", verbose=FALSE, metric="Accuracy", train.fraction=0.5)
```

```
## Loading required package: gbm
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:caret':  
##  
##   cluster
```

```
## Loading required package: splines
```

```
## Loading required package: parallel
```

```
## Loaded gbm 2.1.1
```

```
## Loading required package: plyr
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then  
## dplyr:  
## library(plyr); library(dplyr)
```

```
## -----
```

```
##  
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:modeltools':  
##  
##   empty
```

```
## The following objects are masked from 'package:dplyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0,  
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify  
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0,
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0,
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0,
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0,
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0,
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

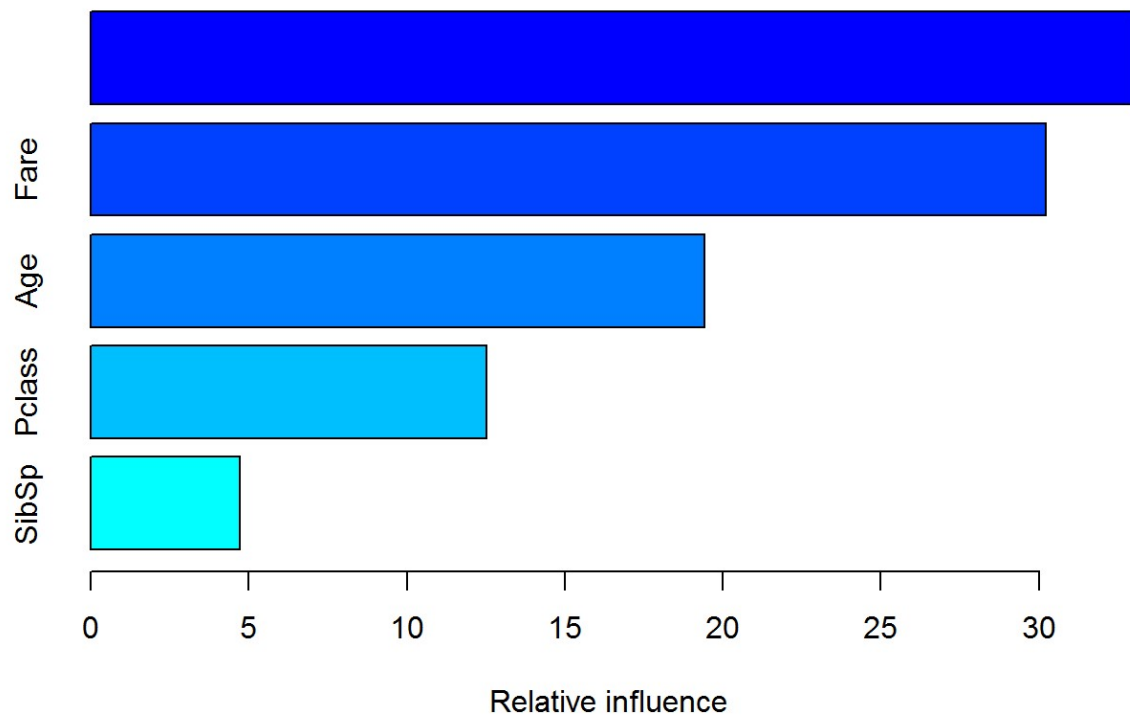
```
## Warning in gbm.fit(x = structure(c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
## 0, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
## 0, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
## 0, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
## Warning in gbm.fit(x = structure(c(0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1,
## 1, : Parameter 'train.fraction' of gbm.fit is deprecated, please specify
## 'nTrain' instead
```

```
summary(Fit3)
```



```
##           var    rel.inf
## SexMale SexMale 33.118634
## Fare      Fare 30.230075
## Age       Age 19.416798
## Pclass    Pclass 12.525261
## SibSp     SibSp  4.709233
```

```
Fit4<-train(Survived ~ Sex+Age+Pclass+Fare+SibSp, trControl=objControl, data =
myTraining, verbose=FALSE, metric="Accuracy",method="rf")
summary(Fit4)
```

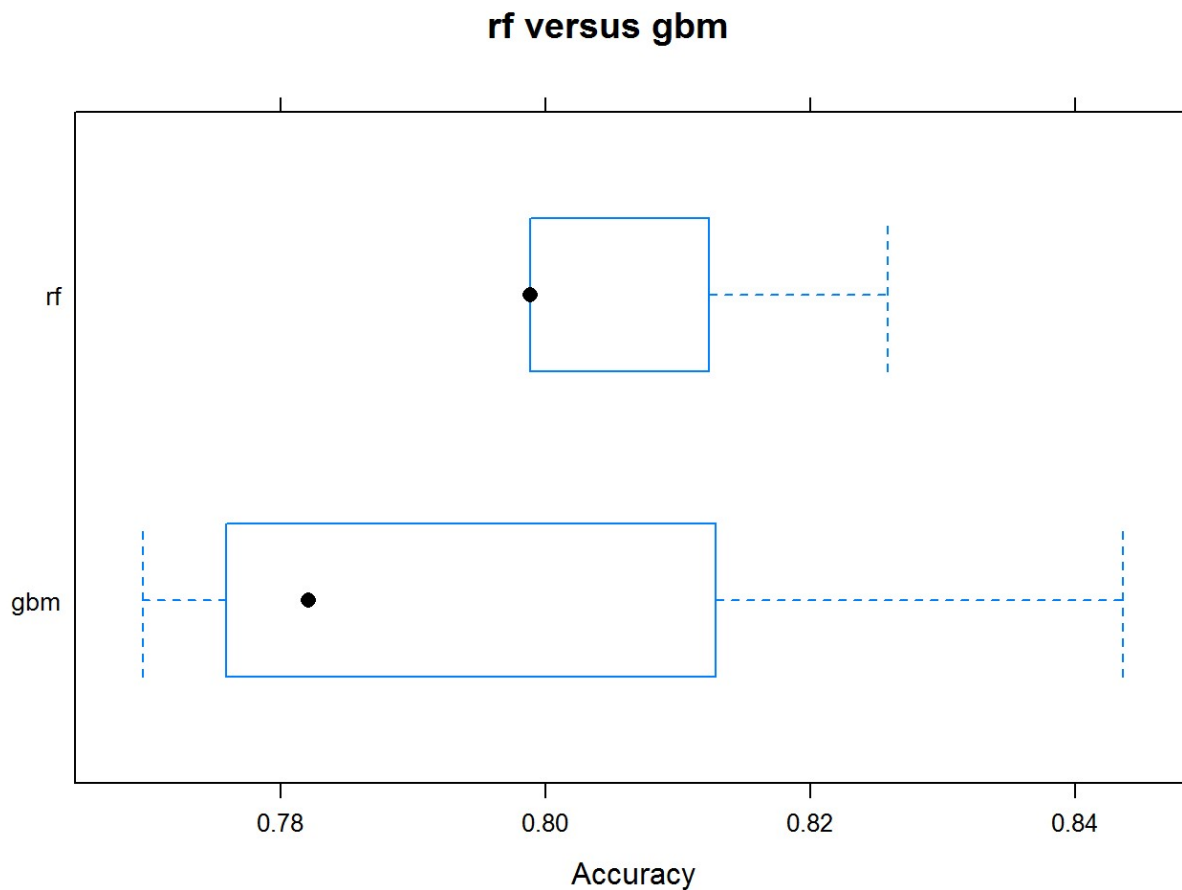


##	Length	Class	Mode
## call	5	-none-	call
## type	1	-none-	character
## predicted	536	factor	numeric
## err.rate	1500	-none-	numeric
## confusion	6	-none-	numeric
## votes	1072	matrix	numeric
## oob.times	536	-none-	numeric
## classes	2	-none-	character
## importance	5	-none-	numeric
## importanceSD	0	-none-	NULL
## localImportance	0	-none-	NULL
## proximity	0	-none-	NULL
## ntree	1	-none-	numeric
## mtry	1	-none-	numeric
## forest	14	-none-	list
## y	536	factor	numeric
## test	0	-none-	NULL
## inbag	0	-none-	NULL
## xNames	5	-none-	character
## problemType	1	-none-	character
## tuneValue	1	data.frame	list
## obsLevels	2	-none-	character

```
rValues34<-resamples(list(rf=Fit4,gbm=Fit3))
summary(rValues34)
```

```
##
## Call:
## summary.resamples(object = rValues34)
##
## Models: rf, gbm
## Number of resamples: 3
##
## Accuracy
##      Min. 1st Qu. Median   Mean 3rd Qu.  Max. NA's
## rf  0.7989  0.7989 0.7989 0.8079  0.8124 0.8258    0
## gbm 0.7697  0.7759 0.7821 0.7985  0.8128 0.8436    0
##
## Kappa
##      Min. 1st Qu. Median   Mean 3rd Qu.  Max. NA's
## rf  0.5613  0.5637 0.5661 0.5844  0.5960 0.6259    0
## gbm 0.4937  0.5112 0.5287 0.5597  0.5928 0.6569    0
```

```
bwplot(rValues34,metric="Accuracy", main="rf versus gbm")
```



## Building the prediction

The accuracy of the random forest model is larger.

```
predict4 <- predict(Fit4, newdata=myTesting)
predict3 <- predict(Fit3, newdata=myTesting)
identical(levels(predict4), levels(myTesting$Survived))
```

```
## [1] TRUE
```

```
levels(predict4)
```

```
## [1] "Survived" "Died"
```

```
levels(myTesting$Survived)
```

```
## [1] "Survived" "Died"
```

# Analyzing the error for the random forest model

There are two predicted classes: Survived and Died. The classifier made a total of 356 predictions (dim of myTesting). The classifier predicted for the random forest model: Survived 134, Died 222. In reality 133 Survived and 223 Died. The accuracy, which measures how often the classifier is correct in this case: 0.823.

```
CF4<-confusionMatrix(predict4, myTesting$Survived)
CF3<-confusionMatrix(predict3, myTesting$Survived)
CF3
```

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction Survived Died
##   Survived      89   16
##    Died       47  203
##
##              Accuracy : 0.8225
##              95% CI : (0.7787, 0.8609)
##   No Information Rate : 0.6169
##   P-Value [Acc > NIR] : < 2.2e-16
##
##              Kappa : 0.6076
##  McNemar's Test P-Value : 0.0001571
##
##              Sensitivity : 0.6544
##              Specificity : 0.9269
##              Pos Pred Value : 0.8476
##              Neg Pred Value : 0.8120
##              Prevalence : 0.3831
##              Detection Rate : 0.2507
##   Detection Prevalence : 0.2958
##              Balanced Accuracy : 0.7907
##
##              'Positive' Class : Survived
##
```

```
CF4
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Survived Died
##   Survived      102   22
##   Died          34  197
##
##           Accuracy : 0.8423
##           95% CI : (0.8001, 0.8786)
##   No Information Rate : 0.6169
##   P-Value [Acc > NIR] : <2e-16
##
##           Kappa : 0.6606
## Mcnemar's Test P-Value : 0.1416
##
##           Sensitivity : 0.7500
##           Specificity : 0.8995
##           Pos Pred Value : 0.8226
##           Neg Pred Value : 0.8528
##           Prevalence : 0.3831
##           Detection Rate : 0.2873
##           Detection Prevalence : 0.3493
##           Balanced Accuracy : 0.8248
##
##           'Positive' Class : Survived
##
```

Accuracy, sensitivity, specificity can be analysed. These are the metrics corresponding to maximum accuracy that could be achieved by the model. We will use the threshold accuracy value while scoring the model in the test set.

In my opinion both the gbm and the random forest model can be used.

## Use model 4 to predict the test data set

Cleaning the test data set

```
dim(testing1)
```

```
## [1] 418  11
```

```

testing1$Title <- gsub('(.*, )|(\\.\\.\\.*)', '', testing1$Name)
different_titles_test<-c('Don', 'Rev','Capt', 'Col', 'Dr','Major', 'Jonkheer', 'Sir', 'Lady', 'the Countess' )

testing1$Title[testing1$Title %in% different_titles_test] <- 'Titles'

testing1$Fare[is.na(testing1$Fare)] <- mean(testing1$Fare, na.rm=TRUE)
testing1$Age[is.na(testing1$Age)] <- median(testing1$Age, na.rm=TRUE)
testing1$AgeG <- testing1$Age
testing1$AgeG <- ifelse(testing1$AgeG<=16, "Child", ifelse(testing1$AgeG<=25,
"Young", ifelse(testing1$AgeG<=55, "Adult", ifelse(testing1$AgeG>55, "Old", "UnknownAge"))))

testing1$Title<-as.factor(testing1$Title)

testing1$Sex<-as.factor(testing1$Sex)

levels(testing1$Sex)<-c("Female", "Male")

testing1$Sex<-as.factor(testing1$Sex)

testing1$Cabin<-NULL
testing1$Ticket<-NULL
print(testing1)

```

##	PassengerId	Pclass
## 1	892	3
## 2	893	3
## 3	894	2
## 4	895	3
## 5	896	3
## 6	897	3
## 7	898	3
## 8	899	2
## 9	900	3
## 10	901	3
## 11	902	3
## 12	903	1
## 13	904	1
## 14	905	2
## 15	906	1
## 16	907	2
## 17	908	2
## 18	909	3
## 19	910	3
## 20	911	3
## 21	912	1
## 22	913	3
## 23	914	1
## 24	915	1
## 25	916	1
## 26	917	3
## 27	918	1
## 28	919	3
## 29	920	1
## 30	921	3
## 31	922	2
## 32	923	2
## 33	924	3
## 34	925	3
## 35	926	1
## 36	927	3
## 37	928	3
## 38	929	3
## 39	930	3
## 40	931	3
## 41	932	3
## 42	933	1
## 43	934	3
## 44	935	2
## 45	936	1
## 46	937	3
## 47	938	1

## 48	939	3
## 49	940	1
## 50	941	3
## 51	942	1
## 52	943	2
## 53	944	2
## 54	945	1
## 55	946	2
## 56	947	3
## 57	948	3
## 58	949	3
## 59	950	3
## 60	951	1
## 61	952	3
## 62	953	2
## 63	954	3
## 64	955	3
## 65	956	1
## 66	957	2
## 67	958	3
## 68	959	1
## 69	960	1
## 70	961	1
## 71	962	3
## 72	963	3
## 73	964	3
## 74	965	1
## 75	966	1
## 76	967	1
## 77	968	3
## 78	969	1
## 79	970	2
## 80	971	3
## 81	972	3
## 82	973	1
## 83	974	1
## 84	975	3
## 85	976	2
## 86	977	3
## 87	978	3
## 88	979	3
## 89	980	3
## 90	981	2
## 91	982	3
## 92	983	3
## 93	984	1
## 94	985	3
## 95	986	1
## 96	987	3

## 97	988	1
## 98	989	3
## 99	990	3
## 100	991	3
## 101	992	1
## 102	993	2
## 103	994	3
## 104	995	3
## 105	996	3
## 106	997	3
## 107	998	3
## 108	999	3
## 109	1000	3
## 110	1001	2
## 111	1002	2
## 112	1003	3
## 113	1004	1
## 114	1005	3
## 115	1006	1
## 116	1007	3
## 117	1008	3
## 118	1009	3
## 119	1010	1
## 120	1011	2
## 121	1012	2
## 122	1013	3
## 123	1014	1
## 124	1015	3
## 125	1016	3
## 126	1017	3
## 127	1018	3
## 128	1019	3
## 129	1020	2
## 130	1021	3
## 131	1022	3
## 132	1023	1
## 133	1024	3
## 134	1025	3
## 135	1026	3
## 136	1027	3
## 137	1028	3
## 138	1029	2
## 139	1030	3
## 140	1031	3
## 141	1032	3
## 142	1033	1
## 143	1034	1
## 144	1035	2
## 145	1036	1



## 146	1037	3
## 147	1038	1
## 148	1039	3
## 149	1040	1
## 150	1041	2
## 151	1042	1
## 152	1043	3
## 153	1044	3
## 154	1045	3
## 155	1046	3
## 156	1047	3
## 157	1048	1
## 158	1049	3
## 159	1050	1
## 160	1051	3
## 161	1052	3
## 162	1053	3
## 163	1054	2
## 164	1055	3
## 165	1056	2
## 166	1057	3
## 167	1058	1
## 168	1059	3
## 169	1060	1
## 170	1061	3
## 171	1062	3
## 172	1063	3
## 173	1064	3
## 174	1065	3
## 175	1066	3
## 176	1067	2
## 177	1068	2
## 178	1069	1
## 179	1070	2
## 180	1071	1
## 181	1072	2
## 182	1073	1
## 183	1074	1
## 184	1075	3
## 185	1076	1
## 186	1077	2
## 187	1078	2
## 188	1079	3
## 189	1080	3
## 190	1081	2
## 191	1082	2
## 192	1083	1
## 193	1084	3
## 194	1085	2

## 195	1086	2
## 196	1087	3
## 197	1088	1
## 198	1089	3
## 199	1090	2
## 200	1091	3
## 201	1092	3
## 202	1093	3
## 203	1094	1
## 204	1095	2
## 205	1096	2
## 206	1097	1
## 207	1098	3
## 208	1099	2
## 209	1100	1
## 210	1101	3
## 211	1102	3
## 212	1103	3
## 213	1104	2
## 214	1105	2
## 215	1106	3
## 216	1107	1
## 217	1108	3
## 218	1109	1
## 219	1110	1
## 220	1111	3
## 221	1112	2
## 222	1113	3
## 223	1114	2
## 224	1115	3
## 225	1116	1
## 226	1117	3
## 227	1118	3
## 228	1119	3
## 229	1120	3
## 230	1121	2
## 231	1122	2
## 232	1123	1
## 233	1124	3
## 234	1125	3
## 235	1126	1
## 236	1127	3
## 237	1128	1
## 238	1129	3
## 239	1130	2
## 240	1131	1
## 241	1132	1
## 242	1133	2
## 243	1134	1

## 244	1135	3
## 245	1136	3
## 246	1137	1
## 247	1138	2
## 248	1139	2
## 249	1140	2
## 250	1141	3
## 251	1142	2
## 252	1143	3
## 253	1144	1
## 254	1145	3
## 255	1146	3
## 256	1147	3
## 257	1148	3
## 258	1149	3
## 259	1150	2
## 260	1151	3
## 261	1152	3
## 262	1153	3
## 263	1154	2
## 264	1155	3
## 265	1156	2
## 266	1157	3
## 267	1158	1
## 268	1159	3
## 269	1160	3
## 270	1161	3
## 271	1162	1
## 272	1163	3
## 273	1164	1
## 274	1165	3
## 275	1166	3
## 276	1167	2
## 277	1168	2
## 278	1169	2
## 279	1170	2
## 280	1171	2
## 281	1172	3
## 282	1173	3
## 283	1174	3
## 284	1175	3
## 285	1176	3
## 286	1177	3
## 287	1178	3
## 288	1179	1
## 289	1180	3
## 290	1181	3
## 291	1182	1
## 292	1183	3

## 293	1184	3
## 294	1185	1
## 295	1186	3
## 296	1187	3
## 297	1188	2
## 298	1189	3
## 299	1190	1
## 300	1191	3
## 301	1192	3
## 302	1193	2
## 303	1194	2
## 304	1195	3
## 305	1196	3
## 306	1197	1
## 307	1198	1
## 308	1199	3
## 309	1200	1
## 310	1201	3
## 311	1202	3
## 312	1203	3
## 313	1204	3
## 314	1205	3
## 315	1206	1
## 316	1207	3
## 317	1208	1
## 318	1209	2
## 319	1210	3
## 320	1211	2
## 321	1212	3
## 322	1213	3
## 323	1214	2
## 324	1215	1
## 325	1216	1
## 326	1217	3
## 327	1218	2
## 328	1219	1
## 329	1220	2
## 330	1221	2
## 331	1222	2
## 332	1223	1
## 333	1224	3
## 334	1225	3
## 335	1226	3
## 336	1227	1
## 337	1228	2
## 338	1229	3
## 339	1230	2
## 340	1231	3
## 341	1232	2

## 342	1233	3
## 343	1234	3
## 344	1235	1
## 345	1236	3
## 346	1237	3
## 347	1238	2
## 348	1239	3
## 349	1240	2
## 350	1241	2
## 351	1242	1
## 352	1243	2
## 353	1244	2
## 354	1245	2
## 355	1246	3
## 356	1247	1
## 357	1248	1
## 358	1249	3
## 359	1250	3
## 360	1251	3
## 361	1252	3
## 362	1253	2
## 363	1254	2
## 364	1255	3
## 365	1256	1
## 366	1257	3
## 367	1258	3
## 368	1259	3
## 369	1260	1
## 370	1261	2
## 371	1262	2
## 372	1263	1
## 373	1264	1
## 374	1265	2
## 375	1266	1
## 376	1267	1
## 377	1268	3
## 378	1269	2
## 379	1270	1
## 380	1271	3
## 381	1272	3
## 382	1273	3
## 383	1274	3
## 384	1275	3
## 385	1276	2
## 386	1277	2
## 387	1278	3
## 388	1279	2
## 389	1280	3
## 390	1281	3

## 391	1282	1
## 392	1283	1
## 393	1284	3
## 394	1285	2
## 395	1286	3
## 396	1287	1
## 397	1288	3
## 398	1289	1
## 399	1290	3
## 400	1291	3
## 401	1292	1
## 402	1293	2
## 403	1294	1
## 404	1295	1
## 405	1296	1
## 406	1297	2
## 407	1298	2
## 408	1299	1
## 409	1300	3
## 410	1301	3
## 411	1302	3
## 412	1303	1
## 413	1304	3
## 414	1305	3
## 415	1306	1
## 416	1307	3
## 417	1308	3
## 418	1309	3

##	Name	Sex
## 1	Kelly, Mr. James	Male
## 2	Wilkes, Mrs. James (Ellen Needs)	Female
## 3	Myles, Mr. Thomas Francis	Male
## 4	Wirz, Mr. Albert	Male
## 5	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	Female
## 6	Svensson, Mr. Johan Cervin	Male
## 7	Connolly, Miss. Kate	Female
## 8	Caldwell, Mr. Albert Francis	Male
## 9	Abraham, Mrs. Joseph (Sophie Halaut Easu)	Female
## 10	Davies, Mr. John Samuel	Male
## 11	Ilieff, Mr. Ylio	Male
## 12	Jones, Mr. Charles Cresson	Male
## 13	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	Female
## 14	Howard, Mr. Benjamin	Male
## 15	Chaffee, Mrs. Herbert Fuller (Carrie Constance Toogood)	Female
## 16	del Carlo, Mrs. Sebastiano (Argenia Genovesi)	Female
## 17	Keane, Mr. Daniel	Male
## 18	Assaf, Mr. Gerios	Male
## 19	Ilmakangas, Miss. Ida Livija	Female
## 20	Assaf Khalil, Mrs. Mariana (Miriam)"	Female

## 21	Rothschild, Mr. Martin	Male
## 22	Olsen, Master. Artur Karl	Male
## 23	Flegenheim, Mrs. Alfred (Antoinette)	Female
## 24	Williams, Mr. Richard Norris II	Male
## 25	Ryerson, Mrs. Arthur Larned (Emily Maria Borie)	Female
## 26	Robins, Mr. Alexander A	Male
## 27	Ostby, Miss. Helene Ragnhild	Female
## 28	Daher, Mr. Shedid	Male
## 29	Brady, Mr. John Bertram	Male
## 30	Samaan, Mr. Elias	Male
## 31	Louch, Mr. Charles Alexander	Male
## 32	Jefferys, Mr. Clifford Thomas	Male
## 33	Dean, Mrs. Bertram (Eva Georgetta Light)	Female
## 34	Johnston, Mrs. Andrew G (Elizabeth Lily" Watson)"	Female
## 35	Mock, Mr. Philipp Edmund	Male
## 36	Katavelas, Mr. Vassilios (Catavelas Vassilios)"	Male
## 37	Roth, Miss. Sarah A	Female
## 38	Cacic, Miss. Manda	Female
## 39	Sap, Mr. Julius	Male
## 40	Hee, Mr. Ling	Male
## 41	Karun, Mr. Franz	Male
## 42	Franklin, Mr. Thomas Parham	Male
## 43	Goldsmith, Mr. Nathan	Male
## 44	Corbett, Mrs. Walter H (Irene Colvin)	Female
## 45	Kimball, Mrs. Edwin Nelson Jr (Gertrude Parsons)	Female
## 46	Peltomaki, Mr. Nikolai Johannes	Male
## 47	Chevre, Mr. Paul Romaine	Male
## 48	Shaughnessy, Mr. Patrick	Male
## 49	Bucknell, Mrs. William Robert (Emma Eliza Ward)	Female
## 50	Coutts, Mrs. William (Winnie Minnie" Treanor)"	Female
## 51	Smith, Mr. Lucien Philip	Male
## 52	Pulbaum, Mr. Franz	Male
## 53	Hocking, Miss. Ellen Nellie""	Female
## 54	Fortune, Miss. Ethel Flora	Female
## 55	Mangiavacchi, Mr. Serafino Emilio	Male
## 56	Rice, Master. Albert	Male
## 57	Cor, Mr. Bartol	Male
## 58	Abelseth, Mr. Olaus Jorgensen	Male
## 59	Davison, Mr. Thomas Henry	Male
## 60	Chaudanson, Miss. Victorine	Female
## 61	Dika, Mr. Mirko	Male
## 62	McCrae, Mr. Arthur Gordon	Male
## 63	Bjorklund, Mr. Ernst Herbert	Male
## 64	Bradley, Miss. Bridget Delia	Female
## 65	Ryerson, Master. John Borie	Male
## 66	Corey, Mrs. Percy C (Mary Phyllis Elizabeth Miller)	Female
## 67	Burns, Miss. Mary Delia	Female
## 68	Moore, Mr. Clarence Bloomfield	Male
## 69	Tucker, Mr. Gilbert Milligan Jr	Male

## 70	Fortune, Mrs. Mark (Mary McDougald)	Female
## 71	Mulvihill, Miss. Bertha E	Female
## 72	Minkoff, Mr. Lazar	Male
## 73	Nieminen, Miss. Manta Josefina	Female
## 74	Ovies y Rodriguez, Mr. Servando	Male
## 75	Geiger, Miss. Amalie	Female
## 76	Keeping, Mr. Edwin	Male
## 77	Miles, Mr. Frank	Male
## 78	Cornell, Mrs. Robert Clifford (Malvina Helen Lamson)	Female
## 79	Aldworth, Mr. Charles Augustus	Male
## 80	Doyle, Miss. Elizabeth	Female
## 81	Boulos, Master. Akar	Male
## 82	Straus, Mr. Isidor	Male
## 83	Case, Mr. Howard Brown	Male
## 84	Demetri, Mr. Marinko	Male
## 85	Lamb, Mr. John Joseph	Male
## 86	Khalil, Mr. Betros	Male
## 87	Barry, Miss. Julia	Female
## 88	Badman, Miss. Emily Louisa	Female
## 89	O'Donoghue, Ms. Bridget	Female
## 90	Wells, Master. Ralph Lester	Male
## 91	Dyker, Mrs. Adolf Fredrik (Anna Elisabeth Judith Andersson)	Female
## 92	Pedersen, Mr. Olaf	Male
## 93	Davidson, Mrs. Thornton (Orian Hays)	Female
## 94	Guest, Mr. Robert	Male
## 95	Birnbaum, Mr. Jakob	Male
## 96	Tenglin, Mr. Gunnar Isidor	Male
## 97	Cavendish, Mrs. Tyrell William (Julia Florence Siegel)	Female
## 98	Makinen, Mr. Kalle Edvard	Male
## 99	Braf, Miss. Elin Ester Maria	Female
## 100	Nancarrow, Mr. William Henry	Male
## 101	Stengel, Mrs. Charles Emil Henry (Annie May Morris)	Female
## 102	Weisz, Mr. Leopold	Male
## 103	Foley, Mr. William	Male
## 104	Johansson Palmquist, Mr. Oskar Leander	Male
## 105	Thomas, Mrs. Alexander (Thamine Thelma)"	Female
## 106	Holthen, Mr. Johan Martin	Male
## 107	Buckley, Mr. Daniel	Male
## 108	Ryan, Mr. Edward	Male
## 109	Willer, Mr. Aaron (Abi Weller)"	Male
## 110	Swane, Mr. George	Male
## 111	Stanton, Mr. Samuel Ward	Male
## 112	Shine, Miss. Ellen Natalia	Female
## 113	Evans, Miss. Edith Corse	Female
## 114	Buckley, Miss. Katherine	Female
## 115	Straus, Mrs. Isidor (Rosalie Ida Blun)	Female
## 116	Chronopoulos, Mr. Demetrios	Male
## 117	Thomas, Mr. John	Male
## 118	Sandstrom, Miss. Beatrice Irene	Female



## 119	Beattie, Mr. Thomson	Male
## 120	Chapman, Mrs. John Henry (Sara Elizabeth Lawry)	Female
## 121	Watt, Miss. Bertha J	Female
## 122	Kiernan, Mr. John	Male
## 123	Schabert, Mrs. Paul (Emma Mock)	Female
## 124	Carver, Mr. Alfred John	Male
## 125	Kennedy, Mr. John	Male
## 126	Cribb, Miss. Laura Alice	Female
## 127	Brobeck, Mr. Karl Rudolf	Male
## 128	McCoy, Miss. Alicia	Female
## 129	Bowenur, Mr. Solomon	Male
## 130	Petersen, Mr. Marius	Male
## 131	Spinner, Mr. Henry John	Male
## 132	Gracie, Col. Archibald IV	Male
## 133	Lefebvre, Mrs. Frank (Frances)	Female
## 134	Thomas, Mr. Charles P	Male
## 135	Dintcheff, Mr. Valtcho	Male
## 136	Carlsson, Mr. Carl Robert	Male
## 137	Zakarian, Mr. Mapriededer	Male
## 138	Schmidt, Mr. August	Male
## 139	Drapkin, Miss. Jennie	Female
## 140	Goodwin, Mr. Charles Frederick	Male
## 141	Goodwin, Miss. Jessie Allis	Female
## 142	Daniels, Miss. Sarah	Female
## 143	Ryerson, Mr. Arthur Larned	Male
## 144	Beauchamp, Mr. Henry James	Male
## 145	Lindeberg-Lind, Mr. Erik Gustaf (Mr Edward Lingrey)"	Male
## 146	Vander Planke, Mr. Julius	Male
## 147	Hilliard, Mr. Herbert Henry	Male
## 148	Davies, Mr. Evan	Male
## 149	Crafton, Mr. John Bertram	Male
## 150	Lahtinen, Rev. William	Male
## 151	Earnshaw, Mrs. Boulton (Olive Potter)	Female
## 152	Matinoff, Mr. Nicola	Male
## 153	Storey, Mr. Thomas	Male
## 154	Klasen, Mrs. (Hulda Kristina Eugenia Lofqvist)	Female
## 155	Asplund, Master. Filip Oscar	Male
## 156	Duquemin, Mr. Joseph	Male
## 157	Bird, Miss. Ellen	Female
## 158	Lundin, Miss. Olga Elida	Female
## 159	Borebank, Mr. John James	Male
## 160	Peacock, Mrs. Benjamin (Edith Nile)	Female
## 161	Smyth, Miss. Julia	Female
## 162	Touma, Master. Georges Youssef	Male
## 163	Wright, Miss. Marion	Female
## 164	Pearce, Mr. Ernest	Male
## 165	Peruschitz, Rev. Joseph Maria	Male
## 166	Kink-Heilmann, Mrs. Anton (Luise Heilmann)	Female
## 167	Brandeis, Mr. Emil	Male

## 168	Ford, Mr. Edward Watson	Male
## 169	Cassebeer, Mrs. Henry Arthur Jr (Eleanor Genevieve Fosdick)	Female
## 170	Hellstrom, Miss. Hilda Maria	Female
## 171	Lithman, Mr. Simon	Male
## 172	Zakarian, Mr. Ortin	Male
## 173	Dyker, Mr. Adolf Fredrik	Male
## 174	Torfa, Mr. Assad	Male
## 175	Asplund, Mr. Carl Oscar Vilhelm Gustafsson	Male
## 176	Brown, Miss. Edith Eileen	Female
## 177	Sincock, Miss. Maude	Female
## 178	Stengel, Mr. Charles Emil Henry	Male
## 179	Becker, Mrs. Allen Oliver (Nellie E Baumgardner)	Female
## 180	Compton, Mrs. Alexander Taylor (Mary Eliza Ingersoll)	Female
## 181	McCrie, Mr. James Matthew	Male
## 182	Compton, Mr. Alexander Taylor Jr	Male
## 183	Marvin, Mrs. Daniel Warner (Mary Graham Carmichael Farquarson)	Female
## 184	Lane, Mr. Patrick	Male
## 185	Douglas, Mrs. Frederick Charles (Mary Helene Baxter)	Female
## 186	Maybery, Mr. Frank Hubert	Male
## 187	Phillips, Miss. Alice Frances Louisa	Female
## 188	Davies, Mr. Joseph	Male
## 189	Sage, Miss. Ada	Female
## 190	Veal, Mr. James	Male
## 191	Angle, Mr. William A	Male
## 192	Salomon, Mr. Abraham L	Male
## 193	van Billiard, Master. Walter John	Male
## 194	Lingane, Mr. John	Male
## 195	Drew, Master. Marshall Brines	Male
## 196	Karlsson, Mr. Julius Konrad Eugen	Male
## 197	Spedden, Master. Robert Douglas	Male
## 198	Nilsson, Miss. Berta Olivia	Female
## 199	Baimbrigge, Mr. Charles Robert	Male
## 200	Rasmussen, Mrs. (Lena Jacobsen Solvang)	Female
## 201	Murphy, Miss. Nora	Female
## 202	Danbom, Master. Gilbert Sigvard Emanuel	Male
## 203	Astor, Col. John Jacob	Male
## 204	Quick, Miss. Winifred Vera	Female
## 205	Andrew, Mr. Frank Thomas	Male
## 206	Omont, Mr. Alfred Fernand	Male
## 207	McGowan, Miss. Katherine	Female
## 208	Collett, Mr. Sidney C Stuart	Male
## 209	Rosenbaum, Miss. Edith Louise	Female
## 210	Delalic, Mr. Redjo	Male
## 211	Andersen, Mr. Albert Karvin	Male
## 212	Finoli, Mr. Luigi	Male
## 213	Deacon, Mr. Percy William	Male
## 214	Howard, Mrs. Benjamin (Ellen Truelove Arman)	Female
## 215	Andersson, Miss. Ida Augusta Margareta	Female
## 216	Head, Mr. Christopher	Male

## 217	Mahon, Miss. Bridget Delia	Female
## 218	Wick, Mr. George Dennick	Male
## 219	Widener, Mrs. George Dunton (Eleanor Elkins)	Female
## 220	Thomson, Mr. Alexander Morrison	Male
## 221	Duran y More, Miss. Florentina	Female
## 222	Reynolds, Mr. Harold J	Male
## 223	Cook, Mrs. (Selena Rogers)	Female
## 224	Karlsson, Mr. Einar Gervasius	Male
## 225	Candee, Mrs. Edward (Helen Churchill Hungerford)	Female
## 226	Moubarek, Mrs. George (Omine Amenia" Alexander)"	Female
## 227	Asplund, Mr. Johan Charles	Male
## 228	McNeill, Miss. Bridget	Female
## 229	Everett, Mr. Thomas James	Male
## 230	Hocking, Mr. Samuel James Metcalfe	Male
## 231	Sweet, Mr. George Frederick	Male
## 232	Willard, Miss. Constance	Female
## 233	Wiklund, Mr. Karl Johan	Male
## 234	Linehan, Mr. Michael	Male
## 235	Cumings, Mr. John Bradley	Male
## 236	Vendel, Mr. Olof Edvin	Male
## 237	Warren, Mr. Frank Manley	Male
## 238	Baccos, Mr. Raffull	Male
## 239	Hiltunen, Miss. Marta	Female
## 240	Douglas, Mrs. Walter Donald (Mahala Dutton)	Female
## 241	Lindstrom, Mrs. Carl Johan (Sigrid Posse)	Female
## 242	Christy, Mrs. (Alice Frances)	Female
## 243	Spedden, Mr. Frederic Oakley	Male
## 244	Hyman, Mr. Abraham	Male
## 245	Johnston, Master. William Arthur Willie""	Male
## 246	Kenyon, Mr. Frederick R	Male
## 247	Karnes, Mrs. J Frank (Claire Bennett)	Female
## 248	Drew, Mr. James Vivian	Male
## 249	Hold, Mrs. Stephen (Annie Margaret Hill)	Female
## 250	Khalil, Mrs. Betros (Zahie Maria" Elias)"	Female
## 251	West, Miss. Barbara J	Female
## 252	Abrahamsson, Mr. Abraham August Johannes	Male
## 253	Clark, Mr. Walter Miller	Male
## 254	Salander, Mr. Karl Johan	Male
## 255	Wenzel, Mr. Linhart	Male
## 256	MacKay, Mr. George William	Male
## 257	Mahon, Mr. John	Male
## 258	Niklasson, Mr. Samuel	Male
## 259	Bentham, Miss. Lilian W	Female
## 260	Midtsjo, Mr. Karl Albert	Male
## 261	de Messemaeker, Mr. Guillaume Joseph	Male
## 262	Nilsson, Mr. August Ferdinand	Male
## 263	Wells, Mrs. Arthur Henry (Addie" Dart Trevaskis)"	Female
## 264	Klasen, Miss. Gertrud Emilia	Female
## 265	Portaluppi, Mr. Emilio Ilario Giuseppe	Male

## 266	Lyntakoff, Mr. Stanko	Male
## 267	Chisholm, Mr. Roderick Robert Crispin	Male
## 268	Warren, Mr. Charles William	Male
## 269	Howard, Miss. May Elizabeth	Female
## 270	Pokrnic, Mr. Mate	Male
## 271	McCaffry, Mr. Thomas Francis	Male
## 272	Fox, Mr. Patrick	Male
## 273	Clark, Mrs. Walter Miller (Virginia McDowell)	Female
## 274	Lennon, Miss. Mary	Female
## 275	Saade, Mr. Jean Nassr	Male
## 276	Bryhl, Miss. Dagmar Jenny Ingeborg	Female
## 277	Parker, Mr. Clifford Richard	Male
## 278	Faunthorpe, Mr. Harry	Male
## 279	Ware, Mr. John James	Male
## 280	Oxenham, Mr. Percy Thomas	Male
## 281	Oreskovic, Miss. Jelka	Female
## 282	Peacock, Master. Alfred Edward	Male
## 283	Fleming, Miss. Honora	Female
## 284	Touma, Miss. Maria Youssef	Female
## 285	Rosblom, Miss. Salli Helena	Female
## 286	Dennis, Mr. William	Male
## 287	Franklin, Mr. Charles (Charles Fardon)	Male
## 288	Snyder, Mr. John Pillsbury	Male
## 289	Mardirosian, Mr. Sarkis	Male
## 290	Ford, Mr. Arthur	Male
## 291	Rheims, Mr. George Alexander Lucien	Male
## 292	Daly, Miss. Margaret Marcella Maggie""	Female
## 293	Nasr, Mr. Mustafa	Male
## 294	Dodge, Dr. Washington	Male
## 295	Wittevrongel, Mr. Camille	Male
## 296	Angheloff, Mr. Minko	Male
## 297	Laroche, Miss. Louise	Female
## 298	Samaan, Mr. Hanna	Male
## 299	Loring, Mr. Joseph Holland	Male
## 300	Johansson, Mr. Nils	Male
## 301	Olsson, Mr. Oscar Wilhelm	Male
## 302	Malachard, Mr. Noel	Male
## 303	Phillips, Mr. Escott Robert	Male
## 304	Pokrnic, Mr. Tome	Male
## 305	McCarthy, Miss. Catherine Katie""	Female
## 306	Crosby, Mrs. Edward Gifford (Catherine Elizabeth Halstead)	Female
## 307	Allison, Mr. Hudson Joshua Creighton	Male
## 308	Aks, Master. Philip Frank	Male
## 309	Hays, Mr. Charles Melville	Male
## 310	Hansen, Mrs. Claus Peter (Jennie L Howard)	Female
## 311	Cacic, Mr. Jego Grga	Male
## 312	Vartanian, Mr. David	Male
## 313	Sadowitz, Mr. Harry	Male
## 314	Carr, Miss. Jeannie	Female

## 315	White, Mrs. John Stuart (Ella Holmes)	Female
## 316	Hagardon, Miss. Kate	Female
## 317	Spencer, Mr. William Augustus	Male
## 318	Rogers, Mr. Reginald Harry	Male
## 319	Jonsson, Mr. Nils Hilding	Male
## 320	Jefferys, Mr. Ernest Wilfred	Male
## 321	Andersson, Mr. Johan Samuel	Male
## 322	Krekorian, Mr. Neshan	Male
## 323	Nesson, Mr. Israel	Male
## 324	Rowe, Mr. Alfred G	Male
## 325	Kreuchen, Miss. Emilie	Female
## 326	Assam, Mr. Ali	Male
## 327	Becker, Miss. Ruth Elizabeth	Female
## 328	Rosenshine, Mr. George (Mr George Thorne)"	Male
## 329	Clarke, Mr. Charles Valentine	Male
## 330	Enander, Mr. Ingvar	Male
## 331	Davies, Mrs. John Morgan (Elizabeth Agnes Mary White)	Female
## 332	Dulles, Mr. William Crothers	Male
## 333	Thomas, Mr. Tannous	Male
## 334	Nakid, Mrs. Said (Waika Mary" Mowad)"	Female
## 335	Cor, Mr. Ivan	Male
## 336	Maguire, Mr. John Edward	Male
## 337	de Brito, Mr. Jose Joaquim	Male
## 338	Elias, Mr. Joseph	Male
## 339	Denbury, Mr. Herbert	Male
## 340	Betros, Master. Seman	Male
## 341	Fillbrook, Mr. Joseph Charles	Male
## 342	Lundstrom, Mr. Thure Edvin	Male
## 343	Sage, Mr. John George	Male
## 344	Cardeza, Mrs. James Warburton Martinez (Charlotte Wardle Drake)	Female
## 345	van Billiard, Master. James William	Male
## 346	Abelseth, Miss. Karen Marie	Female
## 347	Botsford, Mr. William Hull	Male
## 348	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	Female
## 349	Giles, Mr. Ralph	Male
## 350	Walcroft, Miss. Nellie	Female
## 351	Greenfield, Mrs. Leo David (Blanche Strouse)	Female
## 352	Stokes, Mr. Philip Joseph	Male
## 353	Dibden, Mr. William	Male
## 354	Herman, Mr. Samuel	Male
## 355	Dean, Miss. Elizabeth Gladys Millvina""	Female
## 356	Julian, Mr. Henry Forbes	Male
## 357	Brown, Mrs. John Murray (Caroline Lane Lamson)	Female
## 358	Lockyer, Mr. Edward	Male
## 359	O'Keefe, Mr. Patrick	Male
## 360	Lindell, Mrs. Edvard Bengtsson (Elin Gerda Persson)	Female
## 361	Sage, Master. William Henry	Male
## 362	Mallet, Mrs. Albert (Antoinette Magnin)	Female
## 363	Ware, Mrs. John James (Florence Louise Long)	Female

## 364	Strilic, Mr. Ivan	Male
## 365	Harder, Mrs. George Achilles (Dorothy Annan)	Female
## 366	Sage, Mrs. John (Annie Bullen)	Female
## 367	Caram, Mr. Joseph	Male
## 368	Riihivouri, Miss. Susanna Juhantytar Sanni""	Female
## 369	Gibson, Mrs. Leonard (Pauline C Boeson)	Female
## 370	Pallas y Castello, Mr. Emilio	Male
## 371	Giles, Mr. Edgar	Male
## 372	Wilson, Miss. Helen Alice	Female
## 373	Ismay, Mr. Joseph Bruce	Male
## 374	Harbeck, Mr. William H	Male
## 375	Dodge, Mrs. Washington (Ruth Vidaver)	Female
## 376	Bowen, Miss. Grace Scott	Female
## 377	Kink, Miss. Maria	Female
## 378	Cotterill, Mr. Henry Harry""	Male
## 379	Hipkins, Mr. William Edward	Male
## 380	Asplund, Master. Carl Edgar	Male
## 381	O'Connor, Mr. Patrick	Male
## 382	Foley, Mr. Joseph	Male
## 383	Risien, Mrs. Samuel (Emma)	Female
## 384	McNamee, Mrs. Neal (Eileen O'Leary)	Female
## 385	Wheeler, Mr. Edwin Frederick""	Male
## 386	Herman, Miss. Kate	Female
## 387	Aronsson, Mr. Ernst Axel Algot	Male
## 388	Ashby, Mr. John	Male
## 389	Canavan, Mr. Patrick	Male
## 390	Palsson, Master. Paul Folke	Male
## 391	Payne, Mr. Vivian Ponsonby	Male
## 392	Lines, Mrs. Ernest H (Elizabeth Lindsey James)	Female
## 393	Abbott, Master. Eugene Joseph	Male
## 394	Gilbert, Mr. William	Male
## 395	Kink-Heilmann, Mr. Anton	Male
## 396	Smith, Mrs. Lucien Philip (Mary Eloise Hughes)	Female
## 397	Colbert, Mr. Patrick	Male
## 398	Frolicher-Stehli, Mrs. Maxmillian (Margaretha Emerentia Stehli)	Female
## 399	Larsson-Rondberg, Mr. Edvard A	Male
## 400	Conlon, Mr. Thomas Henry	Male
## 401	Bonnell, Miss. Caroline	Female
## 402	Gale, Mr. Harry	Male
## 403	Gibson, Miss. Dorothy Winifred	Female
## 404	Carrau, Mr. Jose Pedro	Male
## 405	Frauenthal, Mr. Isaac Gerald	Male
## 406	Nourney, Mr. Alfred (Baron von Drachstedt)""	Male
## 407	Ware, Mr. William Jeffery	Male
## 408	Widener, Mr. George Dunton	Male
## 409	Riordan, Miss. Johanna Hannah""	Female
## 410	Peacock, Miss. Treasteall	Female
## 411	Naughton, Miss. Hannah	Female
## 412	Minahan, Mrs. William Edward (Lillian E Thorpe)	Female

## 413	Henriksson, Miss. Jenny Lovisa							Female
## 414	Spector, Mr. Woolf							Male
## 415	Oliva y Ocana, Dona. Fermina							Female
## 416	Saether, Mr. Simon Sivertsen							Male
## 417	Ware, Mr. Frederick							Male
## 418	Peter, Master. Michael J							Male
##	Age	SibSp	Parch	Fare	Embarked	Title	AgeG	
## 1	34.50	0	0	7.82920	Q	Mr	Adult	
## 2	47.00	1	0	7.00000	S	Mrs	Adult	
## 3	62.00	0	0	9.68750	Q	Mr	Old	
## 4	27.00	0	0	8.66250	S	Mr	Adult	
## 5	22.00	1	1	12.28750	S	Mrs	Young	
## 6	14.00	0	0	9.22500	S	Mr	Child	
## 7	30.00	0	0	7.62920	Q	Miss	Adult	
## 8	26.00	1	1	29.00000	S	Mr	Adult	
## 9	18.00	0	0	7.22920	C	Mrs	Young	
## 10	21.00	2	0	24.15000	S	Mr	Young	
## 11	27.00	0	0	7.89580	S	Mr	Adult	
## 12	46.00	0	0	26.00000	S	Mr	Adult	
## 13	23.00	1	0	82.26670	S	Mrs	Young	
## 14	63.00	1	0	26.00000	S	Mr	Old	
## 15	47.00	1	0	61.17500	S	Mrs	Adult	
## 16	24.00	1	0	27.72080	C	Mrs	Young	
## 17	35.00	0	0	12.35000	Q	Mr	Adult	
## 18	21.00	0	0	7.22500	C	Mr	Young	
## 19	27.00	1	0	7.92500	S	Miss	Adult	
## 20	45.00	0	0	7.22500	C	Mrs	Adult	
## 21	55.00	1	0	59.40000	C	Mr	Adult	
## 22	9.00	0	1	3.17080	S	Master	Child	
## 23	27.00	0	0	31.68330	S	Mrs	Adult	
## 24	21.00	0	1	61.37920	C	Mr	Young	
## 25	48.00	1	3	262.37500	C	Mrs	Adult	
## 26	50.00	1	0	14.50000	S	Mr	Adult	
## 27	22.00	0	1	61.97920	C	Miss	Young	
## 28	22.50	0	0	7.22500	C	Mr	Young	
## 29	41.00	0	0	30.50000	S	Mr	Adult	
## 30	27.00	2	0	21.67920	C	Mr	Adult	
## 31	50.00	1	0	26.00000	S	Mr	Adult	
## 32	24.00	2	0	31.50000	S	Mr	Young	
## 33	33.00	1	2	20.57500	S	Mrs	Adult	
## 34	27.00	1	2	23.45000	S	Mrs	Adult	
## 35	30.00	1	0	57.75000	C	Mr	Adult	
## 36	18.50	0	0	7.22920	C	Mr	Young	
## 37	27.00	0	0	8.05000	S	Miss	Adult	
## 38	21.00	0	0	8.66250	S	Miss	Young	
## 39	25.00	0	0	9.50000	S	Mr	Young	
## 40	27.00	0	0	56.49580	S	Mr	Adult	
## 41	39.00	0	1	13.41670	C	Mr	Adult	
## 42	27.00	0	0	26.55000	S	Mr	Adult	

## 43	41.00	0	0	7.85000	S	Mr Adult
## 44	30.00	0	0	13.00000	S	Mrs Adult
## 45	45.00	1	0	52.55420	S	Mrs Adult
## 46	25.00	0	0	7.92500	S	Mr Young
## 47	45.00	0	0	29.70000	C	Mr Adult
## 48	27.00	0	0	7.75000	Q	Mr Adult
## 49	60.00	0	0	76.29170	C	Mrs Old
## 50	36.00	0	2	15.90000	S	Mrs Adult
## 51	24.00	1	0	60.00000	S	Mr Young
## 52	27.00	0	0	15.03330	C	Mr Adult
## 53	20.00	2	1	23.00000	S	Miss Young
## 54	28.00	3	2	263.00000	S	Miss Adult
## 55	27.00	0	0	15.57920	C	Mr Adult
## 56	10.00	4	1	29.12500	Q	Master Child
## 57	35.00	0	0	7.89580	S	Mr Adult
## 58	25.00	0	0	7.65000	S	Mr Young
## 59	27.00	1	0	16.10000	S	Mr Adult
## 60	36.00	0	0	262.37500	C	Miss Adult
## 61	17.00	0	0	7.89580	S	Mr Young
## 62	32.00	0	0	13.50000	S	Mr Adult
## 63	18.00	0	0	7.75000	S	Mr Young
## 64	22.00	0	0	7.72500	Q	Miss Young
## 65	13.00	2	2	262.37500	C	Master Child
## 66	27.00	0	0	21.00000	S	Mrs Adult
## 67	18.00	0	0	7.87920	Q	Miss Young
## 68	47.00	0	0	42.40000	S	Mr Adult
## 69	31.00	0	0	28.53750	C	Mr Adult
## 70	60.00	1	4	263.00000	S	Mrs Old
## 71	24.00	0	0	7.75000	Q	Miss Young
## 72	21.00	0	0	7.89580	S	Mr Young
## 73	29.00	0	0	7.92500	S	Miss Adult
## 74	28.50	0	0	27.72080	C	Mr Adult
## 75	35.00	0	0	211.50000	C	Miss Adult
## 76	32.50	0	0	211.50000	C	Mr Adult
## 77	27.00	0	0	8.05000	S	Mr Adult
## 78	55.00	2	0	25.70000	S	Mrs Adult
## 79	30.00	0	0	13.00000	S	Mr Adult
## 80	24.00	0	0	7.75000	Q	Miss Young
## 81	6.00	1	1	15.24580	C	Master Child
## 82	67.00	1	0	221.77920	S	Mr Old
## 83	49.00	0	0	26.00000	S	Mr Adult
## 84	27.00	0	0	7.89580	S	Mr Adult
## 85	27.00	0	0	10.70830	Q	Mr Adult
## 86	27.00	1	0	14.45420	C	Mr Adult
## 87	27.00	0	0	7.87920	Q	Miss Adult
## 88	18.00	0	0	8.05000	S	Miss Young
## 89	27.00	0	0	7.75000	Q	Ms Adult
## 90	2.00	1	1	23.00000	S	Master Child
## 91	22.00	1	0	13.90000	S	Mrs Young



## 92	27.00	0	0	7.77500	S	Mr Adult
## 93	27.00	1	2	52.00000	S	Mrs Adult
## 94	27.00	0	0	8.05000	S	Mr Adult
## 95	25.00	0	0	26.00000	C	Mr Young
## 96	25.00	0	0	7.79580	S	Mr Young
## 97	76.00	1	0	78.85000	S	Mrs Old
## 98	29.00	0	0	7.92500	S	Mr Adult
## 99	20.00	0	0	7.85420	S	Miss Young
## 100	33.00	0	0	8.05000	S	Mr Adult
## 101	43.00	1	0	55.44170	C	Mrs Adult
## 102	27.00	1	0	26.00000	S	Mr Adult
## 103	27.00	0	0	7.75000	Q	Mr Adult
## 104	26.00	0	0	7.77500	S	Mr Adult
## 105	16.00	1	1	8.51670	C	Mrs Child
## 106	28.00	0	0	22.52500	S	Mr Adult
## 107	21.00	0	0	7.82080	Q	Mr Young
## 108	27.00	0	0	7.75000	Q	Mr Adult
## 109	27.00	0	0	8.71250	S	Mr Adult
## 110	18.50	0	0	13.00000	S	Mr Young
## 111	41.00	0	0	15.04580	C	Mr Adult
## 112	27.00	0	0	7.77920	Q	Miss Adult
## 113	36.00	0	0	31.67920	C	Miss Adult
## 114	18.50	0	0	7.28330	Q	Miss Young
## 115	63.00	1	0	221.77920	S	Mrs Old
## 116	18.00	1	0	14.45420	C	Mr Young
## 117	27.00	0	0	6.43750	C	Mr Adult
## 118	1.00	1	1	16.70000	S	Miss Child
## 119	36.00	0	0	75.24170	C	Mr Adult
## 120	29.00	1	0	26.00000	S	Mrs Adult
## 121	12.00	0	0	15.75000	S	Miss Child
## 122	27.00	1	0	7.75000	Q	Mr Adult
## 123	35.00	1	0	57.75000	C	Mrs Adult
## 124	28.00	0	0	7.25000	S	Mr Adult
## 125	27.00	0	0	7.75000	Q	Mr Adult
## 126	17.00	0	1	16.10000	S	Miss Young
## 127	22.00	0	0	7.79580	S	Mr Young
## 128	27.00	2	0	23.25000	Q	Miss Adult
## 129	42.00	0	0	13.00000	S	Mr Adult
## 130	24.00	0	0	8.05000	S	Mr Young
## 131	32.00	0	0	8.05000	S	Mr Adult
## 132	53.00	0	0	28.50000	C	Titles Adult
## 133	27.00	0	4	25.46670	S	Mrs Adult
## 134	27.00	1	0	6.43750	C	Mr Adult
## 135	43.00	0	0	7.89580	S	Mr Adult
## 136	24.00	0	0	7.85420	S	Mr Young
## 137	26.50	0	0	7.22500	C	Mr Adult
## 138	26.00	0	0	13.00000	S	Mr Adult
## 139	23.00	0	0	8.05000	S	Miss Young
## 140	40.00	1	6	46.90000	S	Mr Adult

## 141	10.00	5	2	46.90000	S	Miss Child
## 142	33.00	0	0	151.55000	S	Miss Adult
## 143	61.00	1	3	262.37500	C	Mr Old
## 144	28.00	0	0	26.00000	S	Mr Adult
## 145	42.00	0	0	26.55000	S	Mr Adult
## 146	31.00	3	0	18.00000	S	Mr Adult
## 147	27.00	0	0	51.86250	S	Mr Adult
## 148	22.00	0	0	8.05000	S	Mr Young
## 149	27.00	0	0	26.55000	S	Mr Adult
## 150	30.00	1	1	26.00000	S	Titles Adult
## 151	23.00	0	1	83.15830	C	Mrs Young
## 152	27.00	0	0	7.89580	C	Mr Adult
## 153	60.50	0	0	35.62719	S	Mr Old
## 154	36.00	0	2	12.18330	S	Mrs Adult
## 155	13.00	4	2	31.38750	S	Master Child
## 156	24.00	0	0	7.55000	S	Mr Young
## 157	29.00	0	0	221.77920	S	Miss Adult
## 158	23.00	0	0	7.85420	S	Miss Young
## 159	42.00	0	0	26.55000	S	Mr Adult
## 160	26.00	0	2	13.77500	S	Mrs Adult
## 161	27.00	0	0	7.73330	Q	Miss Adult
## 162	7.00	1	1	15.24580	C	Master Child
## 163	26.00	0	0	13.50000	S	Miss Adult
## 164	27.00	0	0	7.00000	S	Mr Adult
## 165	41.00	0	0	13.00000	S	Titles Adult
## 166	26.00	1	1	22.02500	S	Mrs Adult
## 167	48.00	0	0	50.49580	C	Mr Adult
## 168	18.00	2	2	34.37500	S	Mr Young
## 169	27.00	0	0	27.72080	C	Mrs Adult
## 170	22.00	0	0	8.96250	S	Miss Young
## 171	27.00	0	0	7.55000	S	Mr Adult
## 172	27.00	0	0	7.22500	C	Mr Adult
## 173	23.00	1	0	13.90000	S	Mr Young
## 174	27.00	0	0	7.22920	C	Mr Adult
## 175	40.00	1	5	31.38750	S	Mr Adult
## 176	15.00	0	2	39.00000	S	Miss Child
## 177	20.00	0	0	36.75000	S	Miss Young
## 178	54.00	1	0	55.44170	C	Mr Adult
## 179	36.00	0	3	39.00000	S	Mrs Adult
## 180	64.00	0	2	83.15830	C	Mrs Old
## 181	30.00	0	0	13.00000	S	Mr Adult
## 182	37.00	1	1	83.15830	C	Mr Adult
## 183	18.00	1	0	53.10000	S	Mrs Young
## 184	27.00	0	0	7.75000	Q	Mr Adult
## 185	27.00	1	1	247.52080	C	Mrs Adult
## 186	40.00	0	0	16.00000	S	Mr Adult
## 187	21.00	0	1	21.00000	S	Miss Young
## 188	17.00	2	0	8.05000	S	Mr Young
## 189	27.00	8	2	69.55000	S	Miss Adult

## 190	40.00	0	0	13.00000	S	Mr Adult
## 191	34.00	1	0	26.00000	S	Mr Adult
## 192	27.00	0	0	26.00000	S	Mr Adult
## 193	11.50	1	1	14.50000	S	Master Child
## 194	61.00	0	0	12.35000	Q	Mr Old
## 195	8.00	0	2	32.50000	S	Master Child
## 196	33.00	0	0	7.85420	S	Mr Adult
## 197	6.00	0	2	134.50000	C	Master Child
## 198	18.00	0	0	7.77500	S	Miss Young
## 199	23.00	0	0	10.50000	S	Mr Young
## 200	27.00	0	0	8.11250	S	Mrs Adult
## 201	27.00	0	0	15.50000	Q	Miss Adult
## 202	0.33	0	2	14.40000	S	Master Child
## 203	47.00	1	0	227.52500	C	Titles Adult
## 204	8.00	1	1	26.00000	S	Miss Child
## 205	25.00	0	0	10.50000	S	Mr Young
## 206	27.00	0	0	25.74170	C	Mr Adult
## 207	35.00	0	0	7.75000	Q	Miss Adult
## 208	24.00	0	0	10.50000	S	Mr Young
## 209	33.00	0	0	27.72080	C	Miss Adult
## 210	25.00	0	0	7.89580	S	Mr Young
## 211	32.00	0	0	22.52500	S	Mr Adult
## 212	27.00	0	0	7.05000	S	Mr Adult
## 213	17.00	0	0	73.50000	S	Mr Young
## 214	60.00	1	0	26.00000	S	Mrs Old
## 215	38.00	4	2	7.77500	S	Miss Adult
## 216	42.00	0	0	42.50000	S	Mr Adult
## 217	27.00	0	0	7.87920	Q	Miss Adult
## 218	57.00	1	1	164.86670	S	Mr Old
## 219	50.00	1	1	211.50000	C	Mrs Adult
## 220	27.00	0	0	8.05000	S	Mr Adult
## 221	30.00	1	0	13.85830	C	Miss Adult
## 222	21.00	0	0	8.05000	S	Mr Young
## 223	22.00	0	0	10.50000	S	Mrs Young
## 224	21.00	0	0	7.79580	S	Mr Young
## 225	53.00	0	0	27.44580	C	Mrs Adult
## 226	27.00	0	2	15.24580	C	Mrs Adult
## 227	23.00	0	0	7.79580	S	Mr Young
## 228	27.00	0	0	7.75000	Q	Miss Adult
## 229	40.50	0	0	15.10000	S	Mr Adult
## 230	36.00	0	0	13.00000	S	Mr Adult
## 231	14.00	0	0	65.00000	S	Mr Child
## 232	21.00	0	0	26.55000	S	Miss Young
## 233	21.00	1	0	6.49580	S	Mr Young
## 234	27.00	0	0	7.87920	Q	Mr Adult
## 235	39.00	1	0	71.28330	C	Mr Adult
## 236	20.00	0	0	7.85420	S	Mr Young
## 237	64.00	1	0	75.25000	C	Mr Old
## 238	20.00	0	0	7.22500	C	Mr Young

##	239	18.00	1	1	13.00000	S	Miss Young
##	240	48.00	1	0	106.42500	C	Mrs Adult
##	241	55.00	0	0	27.72080	C	Mrs Adult
##	242	45.00	0	2	30.00000	S	Mrs Adult
##	243	45.00	1	1	134.50000	C	Mr Adult
##	244	27.00	0	0	7.88750	S	Mr Adult
##	245	27.00	1	2	23.45000	S	Master Adult
##	246	41.00	1	0	51.86250	S	Mr Adult
##	247	22.00	0	0	21.00000	S	Mrs Young
##	248	42.00	1	1	32.50000	S	Mr Adult
##	249	29.00	1	0	26.00000	S	Mrs Adult
##	250	27.00	1	0	14.45420	C	Mrs Adult
##	251	0.92	1	2	27.75000	S	Miss Child
##	252	20.00	0	0	7.92500	S	Mr Young
##	253	27.00	1	0	136.77920	C	Mr Adult
##	254	24.00	0	0	9.32500	S	Mr Young
##	255	32.50	0	0	9.50000	S	Mr Adult
##	256	27.00	0	0	7.55000	S	Mr Adult
##	257	27.00	0	0	7.75000	Q	Mr Adult
##	258	28.00	0	0	8.05000	S	Mr Adult
##	259	19.00	0	0	13.00000	S	Miss Young
##	260	21.00	0	0	7.77500	S	Mr Young
##	261	36.50	1	0	17.40000	S	Mr Adult
##	262	21.00	0	0	7.85420	S	Mr Young
##	263	29.00	0	2	23.00000	S	Mrs Adult
##	264	1.00	1	1	12.18330	S	Miss Child
##	265	30.00	0	0	12.73750	C	Mr Adult
##	266	27.00	0	0	7.89580	S	Mr Adult
##	267	27.00	0	0	0.00000	S	Mr Adult
##	268	27.00	0	0	7.55000	S	Mr Adult
##	269	27.00	0	0	8.05000	S	Miss Adult
##	270	17.00	0	0	8.66250	S	Mr Young
##	271	46.00	0	0	75.24170	C	Mr Adult
##	272	27.00	0	0	7.75000	Q	Mr Adult
##	273	26.00	1	0	136.77920	C	Mrs Adult
##	274	27.00	1	0	15.50000	Q	Miss Adult
##	275	27.00	0	0	7.22500	C	Mr Adult
##	276	20.00	1	0	26.00000	S	Miss Young
##	277	28.00	0	0	10.50000	S	Mr Adult
##	278	40.00	1	0	26.00000	S	Mr Adult
##	279	30.00	1	0	21.00000	S	Mr Adult
##	280	22.00	0	0	10.50000	S	Mr Young
##	281	23.00	0	0	8.66250	S	Miss Young
##	282	0.75	1	1	13.77500	S	Master Child
##	283	27.00	0	0	7.75000	Q	Miss Adult
##	284	9.00	1	1	15.24580	C	Miss Child
##	285	2.00	1	1	20.21250	S	Miss Child
##	286	36.00	0	0	7.25000	S	Mr Adult
##	287	27.00	0	0	7.25000	S	Mr Adult

##	288	24.00	1	0	82.26670	S	Mr Young
##	289	27.00	0	0	7.22920	C	Mr Adult
##	290	27.00	0	0	8.05000	S	Mr Adult
##	291	27.00	0	0	39.60000	S	Mr Adult
##	292	30.00	0	0	6.95000	Q	Miss Adult
##	293	27.00	0	0	7.22920	C	Mr Adult
##	294	53.00	1	1	81.85830	S	Titles Adult
##	295	36.00	0	0	9.50000	S	Mr Adult
##	296	26.00	0	0	7.89580	S	Mr Adult
##	297	1.00	1	2	41.57920	C	Miss Child
##	298	27.00	2	0	21.67920	C	Mr Adult
##	299	30.00	0	0	45.50000	S	Mr Adult
##	300	29.00	0	0	7.85420	S	Mr Adult
##	301	32.00	0	0	7.77500	S	Mr Adult
##	302	27.00	0	0	15.04580	C	Mr Adult
##	303	43.00	0	1	21.00000	S	Mr Adult
##	304	24.00	0	0	8.66250	S	Mr Young
##	305	27.00	0	0	7.75000	Q	Miss Adult
##	306	64.00	1	1	26.55000	S	Mrs Old
##	307	30.00	1	2	151.55000	S	Mr Adult
##	308	0.83	0	1	9.35000	S	Master Child
##	309	55.00	1	1	93.50000	S	Mr Adult
##	310	45.00	1	0	14.10830	S	Mrs Adult
##	311	18.00	0	0	8.66250	S	Mr Young
##	312	22.00	0	0	7.22500	C	Mr Young
##	313	27.00	0	0	7.57500	S	Mr Adult
##	314	37.00	0	0	7.75000	Q	Miss Adult
##	315	55.00	0	0	135.63330	C	Mrs Adult
##	316	17.00	0	0	7.73330	Q	Miss Young
##	317	57.00	1	0	146.52080	C	Mr Old
##	318	19.00	0	0	10.50000	S	Mr Young
##	319	27.00	0	0	7.85420	S	Mr Adult
##	320	22.00	2	0	31.50000	S	Mr Young
##	321	26.00	0	0	7.77500	S	Mr Adult
##	322	25.00	0	0	7.22920	C	Mr Young
##	323	26.00	0	0	13.00000	S	Mr Adult
##	324	33.00	0	0	26.55000	S	Mr Adult
##	325	39.00	0	0	211.33750	S	Miss Adult
##	326	23.00	0	0	7.05000	S	Mr Young
##	327	12.00	2	1	39.00000	S	Miss Child
##	328	46.00	0	0	79.20000	C	Mr Adult
##	329	29.00	1	0	26.00000	S	Mr Adult
##	330	21.00	0	0	13.00000	S	Mr Young
##	331	48.00	0	2	36.75000	S	Mrs Adult
##	332	39.00	0	0	29.70000	C	Mr Adult
##	333	27.00	0	0	7.22500	C	Mr Adult
##	334	19.00	1	1	15.74170	C	Mrs Young
##	335	27.00	0	0	7.89580	S	Mr Adult
##	336	30.00	0	0	26.00000	S	Mr Adult

## 337	32.00	0	0	13.00000	S	Mr Adult
## 338	39.00	0	2	7.22920	C	Mr Adult
## 339	25.00	0	0	31.50000	S	Mr Young
## 340	27.00	0	0	7.22920	C	Master Adult
## 341	18.00	0	0	10.50000	S	Mr Young
## 342	32.00	0	0	7.57920	S	Mr Adult
## 343	27.00	1	9	69.55000	S	Mr Adult
## 344	58.00	0	1	512.32920	C	Mrs Old
## 345	27.00	1	1	14.50000	S	Master Adult
## 346	16.00	0	0	7.65000	S	Miss Child
## 347	26.00	0	0	13.00000	S	Mr Adult
## 348	38.00	0	0	7.22920	C	Mrs Adult
## 349	24.00	0	0	13.50000	S	Mr Young
## 350	31.00	0	0	21.00000	S	Miss Adult
## 351	45.00	0	1	63.35830	C	Mrs Adult
## 352	25.00	0	0	10.50000	S	Mr Young
## 353	18.00	0	0	73.50000	S	Mr Young
## 354	49.00	1	2	65.00000	S	Mr Adult
## 355	0.17	1	2	20.57500	S	Miss Child
## 356	50.00	0	0	26.00000	S	Mr Adult
## 357	59.00	2	0	51.47920	S	Mrs Old
## 358	27.00	0	0	7.87920	S	Mr Adult
## 359	27.00	0	0	7.75000	Q	Mr Adult
## 360	30.00	1	0	15.55000	S	Mrs Adult
## 361	14.50	8	2	69.55000	S	Master Child
## 362	24.00	1	1	37.00420	C	Mrs Young
## 363	31.00	0	0	21.00000	S	Mrs Adult
## 364	27.00	0	0	8.66250	S	Mr Adult
## 365	25.00	1	0	55.44170	C	Mrs Young
## 366	27.00	1	9	69.55000	S	Mrs Adult
## 367	27.00	1	0	14.45830	C	Mr Adult
## 368	22.00	0	0	39.68750	S	Miss Young
## 369	45.00	0	1	59.40000	C	Mrs Adult
## 370	29.00	0	0	13.85830	C	Mr Adult
## 371	21.00	1	0	11.50000	S	Mr Young
## 372	31.00	0	0	134.50000	C	Miss Adult
## 373	49.00	0	0	0.00000	S	Mr Adult
## 374	44.00	0	0	13.00000	S	Mr Adult
## 375	54.00	1	1	81.85830	S	Mrs Adult
## 376	45.00	0	0	262.37500	C	Miss Adult
## 377	22.00	2	0	8.66250	S	Miss Young
## 378	21.00	0	0	11.50000	S	Mr Young
## 379	55.00	0	0	50.00000	S	Mr Adult
## 380	5.00	4	2	31.38750	S	Master Child
## 381	27.00	0	0	7.75000	Q	Mr Adult
## 382	26.00	0	0	7.87920	Q	Mr Adult
## 383	27.00	0	0	14.50000	S	Mrs Adult
## 384	19.00	1	0	16.10000	S	Mrs Young
## 385	27.00	0	0	12.87500	S	Mr Adult

```

## 386 24.00      1      2 65.00000      S Miss Young
## 387 24.00      0      0  7.77500      S   Mr Young
## 388 57.00      0      0 13.00000      S   Mr  Old
## 389 21.00      0      0  7.75000      Q   Mr Young
## 390  6.00      3      1 21.07500      S Master Child
## 391 23.00      0      0 93.50000      S   Mr Young
## 392 51.00      0      1 39.40000      S   Mrs Adult
## 393 13.00      0      2 20.25000      S Master Child
## 394 47.00      0      0 10.50000      S   Mr Adult
## 395 29.00      3      1 22.02500      S   Mr Adult
## 396 18.00      1      0 60.00000      S   Mrs Young
## 397 24.00      0      0  7.25000      Q   Mr Young
## 398 48.00      1      1 79.20000      C   Mrs Adult
## 399 22.00      0      0  7.77500      S   Mr Young
## 400 31.00      0      0  7.73330      Q   Mr Adult
## 401 30.00      0      0 164.86670      S Miss Adult
## 402 38.00      1      0 21.00000      S   Mr Adult
## 403 22.00      0      1 59.40000      C Miss Young
## 404 17.00      0      0 47.10000      S   Mr Young
## 405 43.00      1      0 27.72080      C   Mr Adult
## 406 20.00      0      0 13.86250      C   Mr Young
## 407 23.00      1      0 10.50000      S   Mr Young
## 408 50.00      1      1 211.50000      C   Mr Adult
## 409 27.00      0      0  7.72080      Q Miss Adult
## 410  3.00      1      1 13.77500      S Miss Child
## 411 27.00      0      0  7.75000      Q Miss Adult
## 412 37.00      1      0 90.00000      Q   Mrs Adult
## 413 28.00      0      0  7.77500      S Miss Adult
## 414 27.00      0      0  8.05000      S   Mr Adult
## 415 39.00      0      0 108.90000      C Dona Adult
## 416 38.50      0      0  7.25000      S   Mr Adult
## 417 27.00      0      0  8.05000      S   Mr Adult
## 418 27.00      1      1 22.35830      C Master Adult

```

```

testing1$Survived <- predict(Fit4, newdata=testing1)
testing1$Survived<-ifelse(testing1$Survived=="Survived",1,0)

```

### Writing the submission file!

```

solution <- data.frame(PassengerID = testing1$PassengerId, Survived = testing1
$Survived)
write.csv(solution, file = "dragomirescu.csv", row.names = FALSE)

```