

# **ACADGILD**

# SESSION 6: Visualization & Plotting Assignment 1

Submitted by: Munmun Ghosal Login Id: munmun55@gmail.com (M):+91-8007178659

## Data Analytics

## Table of Contents

. Problem Statement	3
Solution	=

#### 1. Problem Statement

1. Import the Titanic Dataset from the following link:

https://drive.google.com/file/d/1JTJCjdGuUxzKXYlwOavwovB01k6FWg3r/view?ts=5b42ea10

Perform the below operations:

- a) Pre-process the passenger names to come up with a list of titles that represent families and represent using appropriate visualization graph.
- b) Represent the proportion of people survived by family size using a graph.
- c) Impute the missing values in Age variable using Mice library, create two different graphs showing Age distribution before and after imputation

#### 2. Solution

#### **Import the Titanic Dataset**

The R-script for the given problem is as follows:

#### The output of the R-Script (from Console window) is given as follows:

```
Console Terminal ×
> library("readr")
> # Import Data Set ; Titanic
> TitanicData <- read.csv("E:/munmun_acadgild/acadgild data analytics/supporting files/titanic3.csv")
> View(TitanicData)
 str(TitanicData)
                 1309 obs. of 14 variables:
'data.frame':
 $ pclass : int 1 1 1 1 1 1 1 1 1 1 1 ...
$ survived : int 1 1 0 0 0 1 1 0 1 0 ...
$ pclass
            : Factor w/ 1307 levels "Abbing, Mr. Anthony",..: 22 24 25 26 27 31 46 47 51 55 ...
: Factor w/ 2 levels "female", "male": 1 2 1 2 1 2 1 2 1 2 ...
 $ name
 $ sex
             : num 29 0.917 2 30 25 ...
 $ age
 $ sibsp
             : int 0111101020...
             : int 0222200000..
 $ parch
            : Factor w/ 929 levels "110152","110413",..: 188 50 50 50 50 125 93 16 77 826 ...
 $ ticket
$ fare : num 211 152 152 152 152 ...

$ cabin : Factor w/ 187 levels "","A10","A11",..: 45 81 81 81 81 151 147 17 63 1 ...

$ embarked : Factor w/ 4 levels "","C","Q","S": 4 4 4 4 4 4 4 4 4 2 ...

$ boat : Factor w/ 28 levels "","I","10","11",..: 13 4 1 1 1 14 3 1 28 1 ...
$ body : int NA NA NA 135 NA NA NA NA 22 ...
$ home.dest: Factor w/ 370 levels "","?Havana, Cuba",..: 310 232 232 232 232 238 163 25 23 230 ...
> psych::describe(TitanicData)
             vars
                          mean
                                    sd median trimmed
                                                            mad min
                                                                           max
                                                                                  range
                                                                                          skew kurtosis
pclass
               1 1309
                          2.29
                                  0.84
                                          3.00
                                                           0.00 1.00
                                                                          3.00
                                                                                   2.00 -0.60
                                                                                                           0.02
                                                   2.37
                                                                                                   -1.32
survived
                2 1309
                          0.38
                                  0.49
                                          0.00
                                                   0.35
                                                           0.00 0.00
                                                                          1.00
                                                                                   1.00
                                                                                          0.49
                                                                                                   -1.77
                                                                                                           0.01
                3 1309 653.69 377.31 653.00
name*
                                                 653.62 484.81 1.00 1307.00 1306.00
                                                                                          0.00
                                                                                                   -1.2010.43
sex*
                4 1309
                          1.64
                                  0.48
                                          2.00
                                                   1.68
                                                           0.00 1.00
                                                                          2.00
                                                                                   1.00 -0.60
                                                                                                   -1.64
                                                                                                           0.01
                5 1046
                         29.88
                                 14.41
                                         28.00
                                                  29.39
                                                          11.86 0.17
                                                                                  79.83
                                                                                          0.41
                                                                                                    0.13
age
                                                                         80.00
sibsp
                6 1309
                          0.50
                                  1.04
                                          0.00
                                                   0.27
                                                           0.00 0.00
                                                                          8.00
                                                                                   8.00
                                                                                         3.84
                                                                                                   19.93
                                                                                                           0.03
                7 1309
                          0.39
                                          0.00
                                                                                   9.00 3.66
                                                                                                           0.02
                                  0.87
                                                   0.18
                                                           0.00 0.00
                                                                          9.00
                                                                                                   21.42
parch
 ticket*
                8 1309 464.60 278.04 460.00
                                                 465.23 379.55 1.00
                                                                        929.00
                                                                                 928.00 -0.01
                                                                                                   -1.33
                                                                                                           7.68
                                 51.76
                                                         10.24 0.00
fare
                9 1308
                         33.30
                                        14.45
                                                  21.57
                                                                        512.33
                                                                                 512.33 4.36
                                                                                                   26.87
                                                                                                           1.43
                                                                                186.00 2.10
cabin*
               10 1309
                         23.04
                                 47.82
                                          1.00
                                                  10.17
                                                           0.00 1.00
                                                                        187.00
                                                                                                    3.14
                                                                                                           1.32
                          3.49
embarked*
               11 1309
                                  0.82
                                          4.00
                                                   3.61
                                                           0.00 1.00
                                                                          4.00
                                                                                   3.00 - 1.13
                                                                                                   -0.51
                                                                                                           0.02
boat*
               12 1309
                          5.97
                                  8.00
                                         1.00
                                                   4.29
                                                           0.00 1.00
                                                                         28.00
                                                                                  27.00 1.42
                                                                                                    0.64
                                                                                                           0.22
body
               13 121 160.81
                                97.70 155.00
                                                160.34 130.47 1.00
                                                                        328.00
                                                                                 327.00
                                                                                         0.09
                                                                                                   -1.28
                                                                                                           8.88
home.dest*
              14 1309 113.16 124.56 54.00
                                                  98.99 78.58 1.00
                                                                       370.00 369.00 0.59
                                                                                                   -1.19
                                                                                                          3.44
```

The titanic dataset is shown as follows:

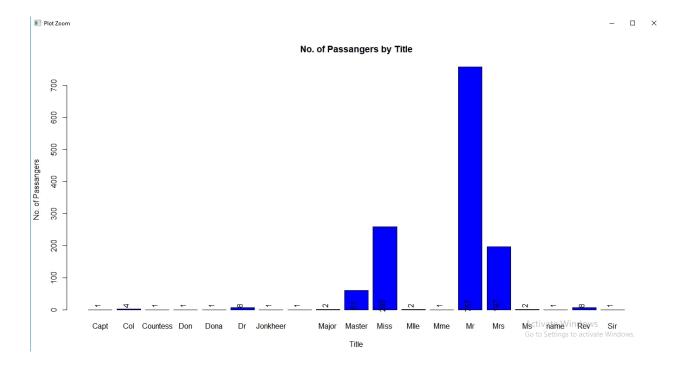
1	name Allen, Miss. Elisabeth Walton	sex	age	sibsp		pclass survived name sex age sibsp parch ticket fare cabin membarked boat body home.dest												
					100		fare	cabin	embarked	boat	body	home.dest						
	The state of the s	female	29.0000	0	_		211.3375		S	2	NA							
	Allison, Master. Hudson Trevor	male	0.9167	1	_	113781	151.5500	C22 C26	S	11	NA	Montreal, PQ / Chesterville, OI						
0	Allison, Miss. Helen Loraine	female	2.0000	1	2	113781	151.5500	C22 C26	S		NA	Montreal, PQ / Chesterville, O						
0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, O						
0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	2	113781	151.5500	C22 C26	S		NA	Montreal, PQ / Chesterville, O						
1	Anderson, Mr. Harry	male	48.0000	0	0	19952	26.5500	E12	S	3	NA	New York, NY						
-1	Andrews, Miss. Kornelia Theodosia	female	63.0000	1	0	13502	77.9583	D7	S	10	NA	Hudson, NY						
0	Andrews, Mr. Thomas Jr	male	39.0000	0	0	112050	0.0000	A36	S		NA	Belfast, NI						
1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0000	2	0	11769	51.4792	C101	S	D	NA	Bayside, Queens, NY						
0	Artagaveytia, Mr. Ramon	male	71.0000	0	0	PC 17609	49.5042		c		22	Montevideo, Uruguay						
0	Astor, Col. John Jacob	male	47.0000	1	0	PC 17757	227.5250	C62 C64	c		124	New York, NY						
1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18.0000	1	0	PC 17757	227.5250	C62 C64	C	4	NA	New York, NY						
1	Aubart, Mme. Leontine Pauline	female	24.0000	0	0	PC 17477	69.3000	B35	С	9	NA	Paris, France						
1	Barber, Miss. Ellen "Nellie"	female	26.0000	0	0	19877	78.8500		S	6	NA							
1	Barkworth, Mr. Algernon Henry Wilson	male	80.0000	0	0	27042	30.0000	A23	S	В	NA	Hessle, Yorks						
0	Baumann, Mr. John D	male	NA	0	0	PC 17318	25.9250		S		NA	New York, NY						
0	Baxter, Mr. Quigg Edmond	male	24.0000	0	1	PC 17558	247.5208	B58 B60	C		NA	Montreal, PQ						
1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.0000	0	1	PC 17558	247.5208	B58 B60	C	6	NA	Montreal, PQ						
1	Bazzani, Miss. Albina	female	32.0000	0	0	11813	76.2917	D15	c	8	NA							
0	Beattie, Mr. Thomson	male	36.0000	0	0	13050	75.2417	C6	С	A	NA	Winnipeg, MN						
1	Beckwith, Mr. Richard Leonard	male	37.0000	1	1	11751	52.5542	D35	S	5	NA	New York, NY						
1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0000	1	1	11751	52.5542	D35	S	5	NA	New York, NY						
	0 1 1 0 0 1 1 1 1 1 0 0 0 0 1 1 1 1 1 0 0 0 0 1	0 Artagaveytia, Mr. Ramon 0 Astor, Col. John Jacob 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) 1 Aubart, Mme. Leontine Pauline 1 Barber, Miss. Ellen 'Nellie' 1 Barkworth, Mr. Algernon Henry Wilson 0 Baumann, Mr. John D 0 Baxter, Mr. Quigg Edmond 1 Baxter, Mrs. James (Helene DeLaudeniere Chaput) 1 Bazzani, Miss. Albina 0 Beattie, Mr. Thomson 1 Beckwith, Mr. Richard Leonard	O Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 1 Anderson, Mr. Harry male 1 Anderson, Mr. Harry male 1 Andrews, Miss. Kornelia Theodosia female 0 Andrews, Mr. Thomas Jr male 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 0 Artagaveytia, Mr. Ramon male 0 Astor, Col. John Jacob male 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 1 Aubart, Mme. Leontine Pauline female 1 Barber, Miss. Ellen 'Nellie' female 1 Barkworth, Mr. Algernon Henry Wilson male 0 Baumann, Mr. John D male 0 Baxter, Mr. Guigg Edmond male 1 Baxter, Mrs. James (Helene Delaudeniere Chaput) female 1 Bazzani, Miss. Albina female 0 Beattie, Mr. Thomson male	0         Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000           1         Anderson, Mr. Harry         male         48,0000           1         Andrews, Miss. Kornelia Theodosia         female         63,0000           0         Andrews, Miss. Kornelia Theodosia         male         39,0000           1         Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         53,0000           0         Astor, Col. John Jacob         male         47,0000           1         Astor, Col. John Jacob (Madeleine Talmadge Force)         female         18,0000           1         Aubart, Mme. Leontine Pauline         female         24,0000           1         Barker, Miss. Ellen "Nellie"         female         26,0000           1         Barkworth, Mir. Algernon Henry Wilson         male         80,0000           0         Baumann, Mr. John D         male         A/4,0000           1         Baster, Mr. Quigg Edmond         male         24,0000           1         Baszter, Mrs. James (Helene Delaudeniere Chaput)         female         30,0000           1         Baszter, Mr. Shina         female         32,0000           1         Baszter, Mr. Thomson         male         37,0000	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1           1 Anderson, Mr. Harry         male         48,0000         0           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1           0 Andrews, Miss. Kornelia Theodosia         male         39,0000         0           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         53,0000         2           0 Artagaveytia, Mr. Ramon         male         71,0000         0           1 Astor, Col. John Jacob         male         47,0000         1           1 Astor, Mrs. John Jacob (Madeleine Talmadge Force)         female         18,0000         1           1 Barker, Miss. Ellen "Nellie"         female         26,0000         0           1 Barker, Mr. Algernon Henry Wilson         male         26,0000         0           0 Baumann, Mr. John D         male         NA         0           0 Baxter, Mr. Quigg Edmond         male         24,0000         0           1 Bazter, Mrs. James (Helene Delaudeniere Chaput)         female         32,0000         0           1 Baztani, Miss. Albina         female         36,0000         0           1 Beckwith, Mr. Richard Leonard         male         37,0000         1 <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1         2           1 Anderson, Mr. Harry         male         48,0000         0         0         0           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1         0           0 Andrews, Miss. Kornelia Theodosia         female         39,0000         0         0           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         39,0000         2         0           0 Artagaveytia, Mr. Ramon         male         71,0000         0         0           0 Astor, Col, John Jacob         male         47,0000         1         0           1 Astor, Mrs. John Jacob (Madeleine Talmadge Force)         female         18,0000         1         0           1 Aubart, Mme. Leontine Pauline         female         24,0000         0         0         0           1 Barker, Miss. Ellen "Nellie"         female         26,0000         0         0         0           1 Barker, Mr. Algernon Henry Wilson         male         80,0000         0         0         0           0 Baumann, Mr. John D         male         A4,000         0         1         1           1 Baxter, Mr. S. Jamse (Helene Delaudeniere Cha</td> <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female 25,0000         1         2 113781           1 Anderson, Mr. Harry         male 48,0000         0         0 19952           1 Andrews, Miss. Kornelia Theodosia         female 63,0000         1         0 13502           0 Andrews, Mr. Thomas Jr         male 39,0000         0         0 112050           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female 53,0000         2         0 11769           0 Artagaveytia, Mr. Ramon         male 71,0000         0         0 PC 17609           0 Attor, Col. John Jacob         male 47,0000         1         0 PC 17757           1 Astor, Mrs. John Jacob (Madeleine Talmadge Force)         female 18,0000         1         0 PC 17777           1 Aubart, Mme. Leontine Pauline         female 24,0000         0         0 PC 17477           1 Barker, Miss. Ellen "Nellie"         female 80,0000         0         0 27042           1 Barker, Mr. Algemon Henry Wilson         male 80,0000         0         0 PC 17318           0 Baumann, Mr. John D         male 24,0000         0         1 PC 17558           1 Baxter, Mr. Quigg Edmond         male 24,0000         0         1 PC 17558           1 Bazter, Mr. S. James (Helner Detaudeniere Chaput)         female 32,0000         0         1 PC 17</td> <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1         2         113781         151,5500           1 Anderson, Mr. Harry         male         48,0000         0         0         19952         26,5500           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1         0         13502         77,9583           0 Andrews, Mr. Thomas Jr         male         39,0000         0         0         112050         0,0000           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         53,0000         2         0         11769         11769         11769         11769         1479502           0 Attagaveytia, Mr. Ramon         male         47,0000         0         0         PC 177679         227,5252           1 Astor, Col. John Jacob         male         47,0000         1         0         PC 177572         227,5252           1 Aubart, Mme, Leontine Pauline         female         18,0000         1         0         PC 177577         227,5250           1 Barker, Miss, Ellen "Nellie"         female         24,0000         0         0         PC 17477         69,3000           1 Barker, Mr. Algernon Henry Wilson         male         80,0000         0</td> <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25.0000 1 2 113781 151.5500 C22 C26 1 Anderson, Mr. Harry male 48.0000 0 0 19952 26.5500 E12 1 Anderson, Mr. Harry male 48.0000 1 0 19952 26.5500 E12 1 Andrews, Miss. Kornelia Theodosia female 63.0000 1 0 0 13502 77.9583 D7 0 Andrews, Mr. Thomas Jr male 39.0000 0 0 112050 0.0000 A36 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 53.0000 2 0 11769 51.4792 C101 0 Artagaveytia, Mr. Ramon male 71.0000 0 0 PC 17609 49.5042 0 Astor, Col. John Jacob male 47.0000 1 0 PC 17757 227.5250 C62 C64 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18.0000 1 0 PC 17777 227.5250 C62 C64 1 Aubart, Mme. Leontine Pauline female 24.0000 0 0 PC 17477 69.3000 B35 1 Barber, Miss. Ellen 'Nellie' female 26.0000 0 0 19877 78.8500 1 Barker, Mr. Algernon Henry Wilson male 80.0000 0 0 27042 30.0000 A23 0 Baumann, Mr. John D male NA 0 PC 17318 25.9520 0 Baster, Mr. Quigg Edmond male 24.0000 0 1 PC 17588 247.5208 B58 B60 1 Baszter, Mrs. James (Helene Delaudeniere Chaput) female 30.0000 0 1 PC 17558 247.5208 B58 B60 1 Bazten, Mrs. Albina female 30.0000 0 0 18813 76.2917 D15 0 Beattie, Mr. Thomson male 36.0000 1 1 11 11751 52.5542 D35</td> <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25,0000 1 2 113781 151.5500 C22 C26 S 1 Anderson, Mr. Harry male 48,0000 0 0 19952 26,5500 E12 S 1 Andrews, Miss. Kornelia Theodosia female 63,0000 1 0 13502 77,9583 D7 S 0 Andrews, Miss. Kornelia Theodosia female 63,0000 0 0 112050 0.0000 A36 S 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 53,0000 2 0 11769 51,4792 C101 S 0 Artagaveytia, Mr. Ramon male 71,0000 0 0 P C 17609 49,5042 C C 0 Astor, Col. John Jacob male 47,0000 1 0 P C 17757 227,5250 662 C64 C 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18,0000 1 0 P C 17757 227,5250 662 C64 C 1 Aubart, Mme. Leontine Pauline female 24,0000 0 0 P C 17477 69,3000 B35 C 1 Barker, Mrs. Ellen 'Nellie' female 26,0000 0 0 19877 78,8500 S 1 Barker, Mrs. Algernon Henry Wilson male 80,0000 0 0 27042 30,0000 A23 S 0 Baumann, Mr. John D male NA 0 0 P C 17318 25,9250 S 5 D Barker, Mr. Quigg Edmond male 24,0000 0 1 P C 17589 247,5208 B58 B60 C 1 Baxter, Mrs. James (Helene DeLaudeniere Chaput) female 50,0000 0 1 10 13813 76,2917 D15 C D Beattie, Mr. Thomson male 36,0000 0 0 13050 75,2417 C6 C C 1 Beatein, Mr. Richard Leonard male 37,0000 1 1 11 11751 52,5542 D35 S</td> <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25,0000 1 2 113781 151,5500 C22 C26 5 1 Anderson, Mr. Harry male 48,0000 0 0 19952 26,5500 E12 5 3 3 1 Andrews, Miss. Kornelia Theodosia female 63,0000 1 0 13502 77,9583 D7 5 10 0 Andrews, Mrs. Thomas Jr male 39,0000 0 0 112050 0,0000 A36 5 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 53,0000 2 0 11769 51,4792 C101 5 D 0 Artagaveytia, Mr. Ramon male 71,0000 0 0 PC 17609 49,5042 C C 101 5 D 0 Attor, Col. John Jacob male 47,0000 1 0 PC 17757 227,5250 62,6264 C 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18,0000 1 0 PC 17757 227,5250 62,6264 C 4 1 Aubart, Mme. Leontine Pauline female 24,0000 0 0 PC 17477 69,3000 B35 C 9 9 1 Barber, Miss. Ellen "Nellie" female 26,0000 0 0 19877 78,8500 S 6 6 1 Barkworth, Mrs. Algernon Henry Wilson male 80,0000 0 0 PC 17588 27,5250 S 5 B 0 Baumann, Mr. John D male NA 0 0 PC 17758 27,5250 B8 B60 C 1 Baxter, Mr. Quigg Edmond male 24,0000 0 1 PC 17588 27,5250 B8 B60 C 1 Baxter, Mrs. James (Helner DeLaudeniere Chaput) female 32,0000 0 1 PC 17580 27,5250 B8 B60 C 6 1 Bazzani, Miss. Albina female 32,0000 0 0 13050 75,2417 C6 C A 1 Bestein, Mrs. Fichard Leonard male 36,0000 1 1 111751 52,5542 D35 S 5</td> <td>0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1         2         113781         151,5500         C22 C26         S         NA           1 Anderson, Mr. Harry         male         48,0000         0         0         19952         26,5500         E12         S         3         NA           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1         0         13502         77,9583         D7         S         10         NA           0 Andrews, Mr. Thomas Jr         male         39,0000         0         0         11769         0,0000         A36         S         NA           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         53,0000         2         0         11769         51,4792         C101         S         D         NA           0 Attagaveytia, Mr. Ramon         male         47,0000         1         0         PC 17699         49,5042         C         22         2         A1364         A1400,Mrs. John Jacob (Madeleine Talmadge Force)         female         18,0000         1         0         PC 17757         227,5250         G62 C64         C         4         NA           1 Barte, Mrs. John D.         female         <t< td=""></t<></td>	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1         2           1 Anderson, Mr. Harry         male         48,0000         0         0         0           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1         0           0 Andrews, Miss. Kornelia Theodosia         female         39,0000         0         0           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         39,0000         2         0           0 Artagaveytia, Mr. Ramon         male         71,0000         0         0           0 Astor, Col, John Jacob         male         47,0000         1         0           1 Astor, Mrs. John Jacob (Madeleine Talmadge Force)         female         18,0000         1         0           1 Aubart, Mme. Leontine Pauline         female         24,0000         0         0         0           1 Barker, Miss. Ellen "Nellie"         female         26,0000         0         0         0           1 Barker, Mr. Algernon Henry Wilson         male         80,0000         0         0         0           0 Baumann, Mr. John D         male         A4,000         0         1         1           1 Baxter, Mr. S. Jamse (Helene Delaudeniere Cha	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female 25,0000         1         2 113781           1 Anderson, Mr. Harry         male 48,0000         0         0 19952           1 Andrews, Miss. Kornelia Theodosia         female 63,0000         1         0 13502           0 Andrews, Mr. Thomas Jr         male 39,0000         0         0 112050           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female 53,0000         2         0 11769           0 Artagaveytia, Mr. Ramon         male 71,0000         0         0 PC 17609           0 Attor, Col. John Jacob         male 47,0000         1         0 PC 17757           1 Astor, Mrs. John Jacob (Madeleine Talmadge Force)         female 18,0000         1         0 PC 17777           1 Aubart, Mme. Leontine Pauline         female 24,0000         0         0 PC 17477           1 Barker, Miss. Ellen "Nellie"         female 80,0000         0         0 27042           1 Barker, Mr. Algemon Henry Wilson         male 80,0000         0         0 PC 17318           0 Baumann, Mr. John D         male 24,0000         0         1 PC 17558           1 Baxter, Mr. Quigg Edmond         male 24,0000         0         1 PC 17558           1 Bazter, Mr. S. James (Helner Detaudeniere Chaput)         female 32,0000         0         1 PC 17	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1         2         113781         151,5500           1 Anderson, Mr. Harry         male         48,0000         0         0         19952         26,5500           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1         0         13502         77,9583           0 Andrews, Mr. Thomas Jr         male         39,0000         0         0         112050         0,0000           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         53,0000         2         0         11769         11769         11769         11769         1479502           0 Attagaveytia, Mr. Ramon         male         47,0000         0         0         PC 177679         227,5252           1 Astor, Col. John Jacob         male         47,0000         1         0         PC 177572         227,5252           1 Aubart, Mme, Leontine Pauline         female         18,0000         1         0         PC 177577         227,5250           1 Barker, Miss, Ellen "Nellie"         female         24,0000         0         0         PC 17477         69,3000           1 Barker, Mr. Algernon Henry Wilson         male         80,0000         0	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25.0000 1 2 113781 151.5500 C22 C26 1 Anderson, Mr. Harry male 48.0000 0 0 19952 26.5500 E12 1 Anderson, Mr. Harry male 48.0000 1 0 19952 26.5500 E12 1 Andrews, Miss. Kornelia Theodosia female 63.0000 1 0 0 13502 77.9583 D7 0 Andrews, Mr. Thomas Jr male 39.0000 0 0 112050 0.0000 A36 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 53.0000 2 0 11769 51.4792 C101 0 Artagaveytia, Mr. Ramon male 71.0000 0 0 PC 17609 49.5042 0 Astor, Col. John Jacob male 47.0000 1 0 PC 17757 227.5250 C62 C64 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18.0000 1 0 PC 17777 227.5250 C62 C64 1 Aubart, Mme. Leontine Pauline female 24.0000 0 0 PC 17477 69.3000 B35 1 Barber, Miss. Ellen 'Nellie' female 26.0000 0 0 19877 78.8500 1 Barker, Mr. Algernon Henry Wilson male 80.0000 0 0 27042 30.0000 A23 0 Baumann, Mr. John D male NA 0 PC 17318 25.9520 0 Baster, Mr. Quigg Edmond male 24.0000 0 1 PC 17588 247.5208 B58 B60 1 Baszter, Mrs. James (Helene Delaudeniere Chaput) female 30.0000 0 1 PC 17558 247.5208 B58 B60 1 Bazten, Mrs. Albina female 30.0000 0 0 18813 76.2917 D15 0 Beattie, Mr. Thomson male 36.0000 1 1 11 11751 52.5542 D35	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25,0000 1 2 113781 151.5500 C22 C26 S 1 Anderson, Mr. Harry male 48,0000 0 0 19952 26,5500 E12 S 1 Andrews, Miss. Kornelia Theodosia female 63,0000 1 0 13502 77,9583 D7 S 0 Andrews, Miss. Kornelia Theodosia female 63,0000 0 0 112050 0.0000 A36 S 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 53,0000 2 0 11769 51,4792 C101 S 0 Artagaveytia, Mr. Ramon male 71,0000 0 0 P C 17609 49,5042 C C 0 Astor, Col. John Jacob male 47,0000 1 0 P C 17757 227,5250 662 C64 C 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18,0000 1 0 P C 17757 227,5250 662 C64 C 1 Aubart, Mme. Leontine Pauline female 24,0000 0 0 P C 17477 69,3000 B35 C 1 Barker, Mrs. Ellen 'Nellie' female 26,0000 0 0 19877 78,8500 S 1 Barker, Mrs. Algernon Henry Wilson male 80,0000 0 0 27042 30,0000 A23 S 0 Baumann, Mr. John D male NA 0 0 P C 17318 25,9250 S 5 D Barker, Mr. Quigg Edmond male 24,0000 0 1 P C 17589 247,5208 B58 B60 C 1 Baxter, Mrs. James (Helene DeLaudeniere Chaput) female 50,0000 0 1 10 13813 76,2917 D15 C D Beattie, Mr. Thomson male 36,0000 0 0 13050 75,2417 C6 C C 1 Beatein, Mr. Richard Leonard male 37,0000 1 1 11 11751 52,5542 D35 S	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25,0000 1 2 113781 151,5500 C22 C26 5 1 Anderson, Mr. Harry male 48,0000 0 0 19952 26,5500 E12 5 3 3 1 Andrews, Miss. Kornelia Theodosia female 63,0000 1 0 13502 77,9583 D7 5 10 0 Andrews, Mrs. Thomas Jr male 39,0000 0 0 112050 0,0000 A36 5 1 Appleton, Mrs. Edward Dale (Charlotte Lamson) female 53,0000 2 0 11769 51,4792 C101 5 D 0 Artagaveytia, Mr. Ramon male 71,0000 0 0 PC 17609 49,5042 C C 101 5 D 0 Attor, Col. John Jacob male 47,0000 1 0 PC 17757 227,5250 62,6264 C 1 Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18,0000 1 0 PC 17757 227,5250 62,6264 C 4 1 Aubart, Mme. Leontine Pauline female 24,0000 0 0 PC 17477 69,3000 B35 C 9 9 1 Barber, Miss. Ellen "Nellie" female 26,0000 0 0 19877 78,8500 S 6 6 1 Barkworth, Mrs. Algernon Henry Wilson male 80,0000 0 0 PC 17588 27,5250 S 5 B 0 Baumann, Mr. John D male NA 0 0 PC 17758 27,5250 B8 B60 C 1 Baxter, Mr. Quigg Edmond male 24,0000 0 1 PC 17588 27,5250 B8 B60 C 1 Baxter, Mrs. James (Helner DeLaudeniere Chaput) female 32,0000 0 1 PC 17580 27,5250 B8 B60 C 6 1 Bazzani, Miss. Albina female 32,0000 0 0 13050 75,2417 C6 C A 1 Bestein, Mrs. Fichard Leonard male 36,0000 1 1 111751 52,5542 D35 S 5	0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)         female         25,0000         1         2         113781         151,5500         C22 C26         S         NA           1 Anderson, Mr. Harry         male         48,0000         0         0         19952         26,5500         E12         S         3         NA           1 Andrews, Miss. Kornelia Theodosia         female         63,0000         1         0         13502         77,9583         D7         S         10         NA           0 Andrews, Mr. Thomas Jr         male         39,0000         0         0         11769         0,0000         A36         S         NA           1 Appleton, Mrs. Edward Dale (Charlotte Lamson)         female         53,0000         2         0         11769         51,4792         C101         S         D         NA           0 Attagaveytia, Mr. Ramon         male         47,0000         1         0         PC 17699         49,5042         C         22         2         A1364         A1400,Mrs. John Jacob (Madeleine Talmadge Force)         female         18,0000         1         0         PC 17757         227,5250         G62 C64         C         4         NA           1 Barte, Mrs. John D.         female <t< td=""></t<>						

a) Pre-process the passenger names to come up with a list of titles that represent families and represent using appropriate visualization graph.

#### The R-script for the given problem is as follows:

#### The output of the R-Script (from Console window) is given as follows:

```
> # Convert Name as character
 > TitanicData$Name <- as.character(TitanicData$Name)
 > # Extract the title from passenger names
> # Extract the title from passenger names
| Title 
 > TitanicData$subTitle <- gsub("\\..*", "", TitanicData$Name)
> TitanicData$Title <- gsub(".*\\ ", "", TitanicData$subTitle)</pre>
> table(TitanicData$Title) # Count of Titles
                       Capt
                                                                           Col Countess
                                                                                                                                                                            Don
                                                                                                                                                                                                                                                                                 Dr Jonkheer
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Miss
                                                                                 4
                                                                                                                              1
                                                                                                                                                                                                                                   1
                                                                                                                                                                                                                                                                                                                                                                                                                                                                               61
                                Mr
                                                                           Mrs
                                                                                                                                Ms
                                                                                                                                                                                                                           Sir
                                                                          197
> # Plot a bar-graph showing Number of Passengers by Title
         Title <- barplot(table(TitanicData$Title),
+ main = "No. of Passangers by Title", xlab = "Title",
+ ylab = "No. of Passangers", col = "Blue")
> text(Title, 0,table(TitanicData$Title), pos = 3, srt = 90)
```



#### b) Represent the proportion of people survived by family size using a graph.

#### The R-script for the given problem is as follows:

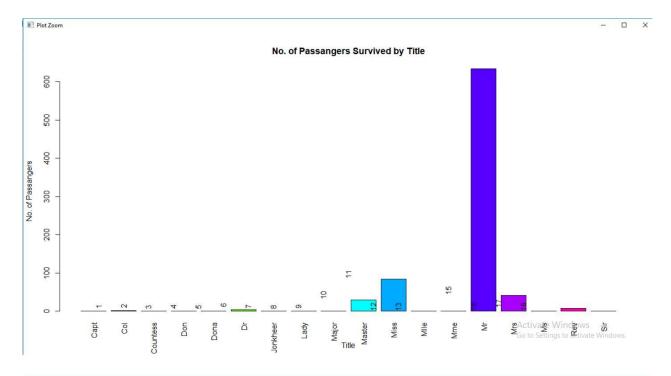
```
x <- table(TitanicData$Survived, TitanicData$Title) # table for survived and died
                              # 0 for survived and 1 for died
Х
p <- x[1,] # number of passengers survived
р
prop <- round(p*100/sum(p),1) # proportion of passangers survived
prop
# in barchart format
#for number of Passengers
barplot(p,
    main = "No. of Passangers Survived by Title",
    xlab = "Title",
    ylab = "No. of Passangers", col = rainbow(length(p)), las
=3) text(p, pos = 3, srt = 90)
# for percentage of passengers
barplot(prop, main = "No. of Passangers by Title", xlab = "Title",
    ylab = "Proportion of Passangers", col = c("Blue", "Red"),
    legend = rownames(prop), ylim=c(0, 100), las = 3)
text(prop, pos = 3, srt = 90)
# in Pie Chart format
pie_chart <- pie(p, labels = p, main = " No.of passengers of Survival by Family", col =
rainbow(length(p)), cex = 1)
```

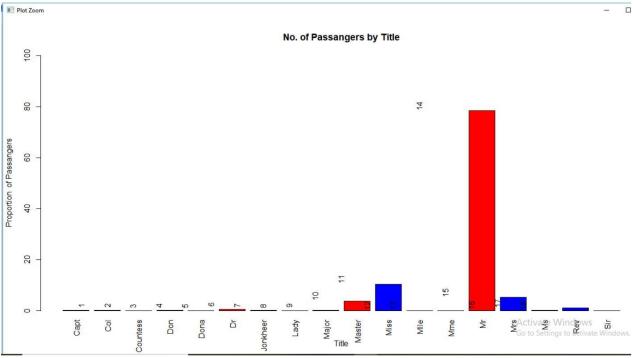
```
legend("topright", names(p), cex= 0.5, fill = rainbow(length(p)))
pie(prop, labels = prop, main = " Proportion of Survival by
    Family", col = rainbow(length(prop)), cex = 1)
legend("topright", names(prop), cex= 0.5, fill = rainbow(length(prop)))
```

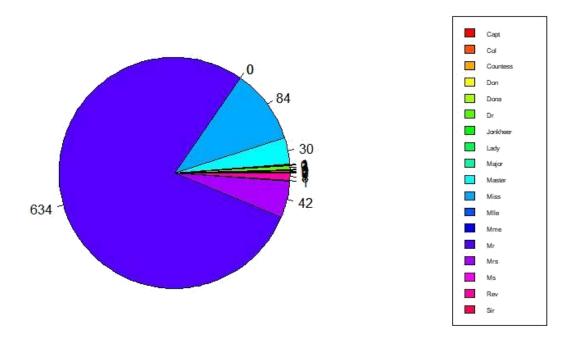
#### The output of the R-Script (from Console window) is given as follows:

```
> x <- table(TitanicData$Survived, TitanicData$Title) # table for survived
and died
                                                            # 0 for survived and 1
> X
for died
     Capt Col Countess Don Dona Dr Jonkheer Lady Major Master Miss Mlle Mme
Mr Mrs Ms Rev Sir
  0
       1
            2
                      0
                           1
                                     4
                                              1
                                                    0
                                                          1
                                                                 30
                                                                       84
                                                                                  0
634
      42
          1
       0
                           0
                                1
                                     4
                                              0
                                                    1
                                                          1
                                                                 31 176
                                                                             2
                                                                                  1
  1
            2
                      1
123 155 1
> p <- x[1,]# number of passengers survived
> p
    Capt
                Col Countess
                                    Don
                                            Dona
                                                        Dr Jonkheer
                                                                         Lady
Major
        Master
                     Miss
                            0
       1
                 2
                                      1
                                                0
                                                         4
                                                                    1
                                                                            0
1
        30
                   84
    Mlle
                                    Mrs
                                                       Rev
                                                                 sir
               Mme
                           Mr
                                              Ms
                 0
                                     42
       0
                          634
                                                1
                                                         8
                                                                   0
> prop <- round(p*100/sum(p),1) # proportion of passangers survived</pre>
> prop
               Col Countess
                                            Dona
                                                        Dr Jonkheer
    Capt
                                   Don
                                                                         Lady
        Master
                     Miss
Major
                          0.0
                                   0.1
                                             0.0
                                                      0.5
                                                                 0.1
                                                                          0.0
     0.1
               0.2
                   10.4
0.1
          3.7
    Mlle
               Mme
                           Mr
                                   Mrs
                                              Ms
                                                      Rev
                                                                 Sir
                                              0.1
                                                                 0.0
     0.0
               0.0
                         78.4
                                   5.2
                                                      1.0
> # in barchart format
>
                                       # for number of Passangers
 barplot(p,
          main = "No. of Passangers Survived by Title",
          xlab = "Title",
+
          ylab = "No. of Passangers", col = rainbow(length(p)), las =3)
 text(p, pos = 3, srt = 90)
 barplot(prop,
                                       # for percentage of passangers
>
          main = "No. of Passangers by Title", xlab = "Title",
ylab = "Proportion of Passangers", col = c("Blue","Red"),
+
           legend = rownames(prop), ylim=c(0, 100), las = 3)
 text(prop, pos = 3, srt = 90)
> # in Pie Chart format
> pie_chart <- pie(p, labels = p, main = " No.of passengers of Survival by</pre>
Family",
                    col = rainbow(length(p)), cex = 1)
> legend("topright", names(p), cex= 0.5, fill = rainbow(length(p)))
```

```
>>
> pie(prop, labels = prop, main = " Proportion of Survival by Family",
+col = rainbow(length(prop)), cex = 1)
> legend("topright", names(prop), cex= 0.5, fill = rainbow(length(prop)))
```







c) Impute the missing values in Age variable using Mice library, create two different graphs showing Age distribution before and after imputation

#### The R-script for the given problem is as follows:

```
p + geom_bar(position = "stack") + theme(axis.text.x = element_text(angle = 90)) +
       labs(title = "Counts of Title with Age Groups")
       library(mice)
       # All variables should be either factor or numeric.
       library(dplyr)
       str(TitanicData)
       dat <- TitanicData[,-13]
       str(dat)
       dat <- dat %>% mutate(agecat = as.factor(agecat), Title = as.factor(Title)) # convert as
       str(dat) # Check the data set
       # Now the data set is ready for imputation using library mice. called earlier
       init = mice(dat, maxit=5)
       meth = init$method
       predM = init$predictorMatrix
       # below variable are not required for predicting the age
       predM[, c("PassengerId", "Name", "Ticket", "Cabin", "Embarked")]=0
       # specify method for imputing the missing value
       meth[c("Age")]="norm"
       set.seed(1)
       # impute the missing values
       imputed = mice(dat, method=meth, predictorMatrix=predM, m=5)
       imputed <- complete(imputed)</pre>
       # check for missings in the imputed dataset
       sapply(imputed, function(x) sum(is.na(x)))
       # Title and Age Group after imputation
       library(ggplot2)
       p <- ggplot(data = imputed,
              mapping = aes(Title, fill = agecat))
       p + geom_bar(position = "stack")+theme(axis.text.x = element_text(angle = 90)) +
       labs(title = "Counts of Title with Age Groups")
The output of the R-Script (from Console window) is given as follows:
 library(readr)
 TitanicData <- within(TitanicData,
                               agecat <- NA
                               agecat[Age>=0 & Age<=25] <- "Low"
                               agecat[Age>=26 & Age<=40] <- "Middle"
agecat[Age>=41] <- "High"</pre>
                             })
 head(TitanicData)
  Pclass Survived
                                                                          Name
                                                                                    Sex
Age SibSp Parch Ticket
                                Fare
                                         Cabin
                                           Allen, Miss. Elisabeth Walton female
        1
                   1
29.0000
                     0 24160 211.3375
                                                 В5
                   1
                                          Allison, Master. Hudson Trevor
                                                                                  male
```

1

2 113781 151.5500 c22 c26

0.9167

```
Allison, Miss. Helen Loraine female
2.0000
                  2 113781 151.5500 C22 C26
           1
       1
                              Allison, Mr. Hudson Joshua Creighton
30.0000
            1
                   2 113781 151.5500 C22 C26
                 O Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female
                  2 113781 151.5500 c22 c26
25.0000
            1
                                                                       male
                                                Anderson, Mr. Harry
48.0000
            0
                     19952 26.5500
                   0
                                          E12
  Embarked Boat Body
                                           destination
                                                               SubTitle
                                                                        Title
agecat
              2
                                          St Louis, MO
                                                            Allen, Miss
         S
                  NA
                                                                          Miss
1
Middle
2
             11
                  NA Montreal, PQ / Chesterville, ON Allison, Master Master
         S
Low
3
         S
                  NA Montreal, PQ / Chesterville, ON
                                                          Allison, Miss
                                                                          Miss
Low
                  135 Montreal, PQ / Chesterville, ON
4
         S
                                                            Allison, Mr
                                                                            Mr
Middle
5
                  NA Montreal, PQ / Chesterville, ON
                                                           Allison, Mrs
         S
                                                                           Mrs
Low
6
              3
                                          New York, NY
                                                           Anderson, Mr
                                                                            Mr
High
> count <- table(TitanicData$agecat, TitanicData$Title)</pre>
> count
```

Capt Col Countess Don Dona Dr Jonkheer Lady Major Master Miss Mlle Mme Mr Mrs Ms Rev Sir 5 0 0 0 1 2 0 11 0 High 0 129 0 5 65 1 0 0 0 0 1 0 0 0 53 138 2 0 Low 1 211 37 0 0 0 Middle 0 0 1 1 1 1 1 0 0 0 61 0 0 238 1 3 0 68 > library(ggplot2) > p <- ggplot(data = TitanicData,</pre>

mapping = aes(Title, fill = agecat))
> p + geom\_bar(position = "stack") + theme(axis.text.x = element\_text(angle = agecat)) 90)) + labs(title = "Counts of Title with Age Groups")

