

# BharathiK\_HW1.R

Bharathi

2025-02-07

```
#Part 1
```

```
#Set the directory to retrieve the necessary data sets
getwd()
```

```
## [1] "C:/Users/Bharathi/OneDrive/Documents/Data Mining/Rscripts"
```

```
# Its in the correct directory
```

```
##(a) Load the data set and name it su
su = read.delim('Su_raw_matrix.txt')
```

```
#Read the data set
head(su)
```

	Brain_1.CEL	Brain_2.CEL	Fetal_brain_1.CEL	Fetal_brain_2.CEL
## 100_g_at	120.25	255.00	3.50	31.00
## 1000_at	583.60	885.40	253.65	293.35
## 1001_at	35.85	40.50	47.20	33.05
## 1002_f_at	17.60	19.85	11.10	23.10
## 1003_s_at	0.15	26.40	78.00	36.00
## 1004_at	34.70	77.75	-12.00	16.90
	Fetal_liver_1.CEL	Fetal_liver_2.CEL	Liver_1.CEL	Liver_2.CEL
## 100_g_at	6.50	-8.25	19.15	73.00
## 1000_at	201.20	433.75	134.25	251.15
## 1001_at	86.35	119.25	37.10	72.10
## 1002_f_at	38.75	94.60	452.10	662.50
## 1003_s_at	89.50	34.00	22.75	100.00
## 1004_at	121.75	24.25	30.00	54.35

```
##(b) To get the mean of Liver_2.CEL column
mean(su$Liver_2.CEL, na.rm=TRUE)
```

```
## [1] 241.8246
```

```
# The mean is 241.8246
```

```
# To get the standard deviation of Liver_2.CEL column
sd(su$Liver_2.CEL)
```

```
## [1] 1133.352
```

*# The standard deviation is 1133.352*

*##(c) To get the average and total values of each column.*

```
mean_su_column = colMeans(su[sapply(su, is.numeric)], na.rm=TRUE)
mean_su_column
```

```
##      Brain_1.CEL      Brain_2.CEL Fetal_brain_1.CEL Fetal_brain_2.CEL
##      204.9763       315.0924      198.3439       267.6551
## Fetal_liver_1.CEL Fetal_liver_2.CEL      Liver_1.CEL      Liver_2.CEL
##      209.8722       399.1482      160.8558       241.8246
```

```
colSums(su,na.rm=TRUE)
```

```
##      Brain_1.CEL      Brain_2.CEL Fetal_brain_1.CEL Fetal_brain_2.CEL
##      2588031        3978357      2504290        3379413
## Fetal_liver_1.CEL Fetal_liver_2.CEL      Liver_1.CEL      Liver_2.CEL
##      2649846        5039645      2030966        3053278
```

*#Part 2*

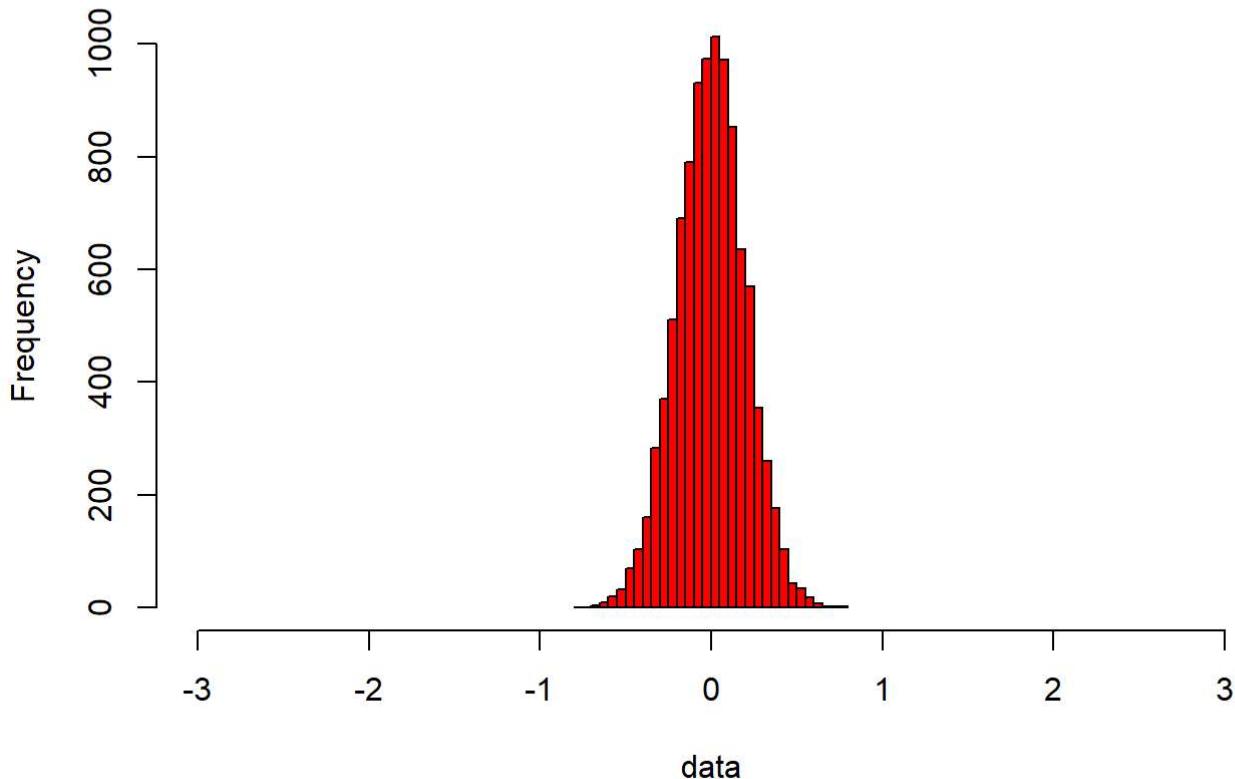
*# Generate 10000 numbers with mean=0, and Standard deviation=0.2*

```
data = rnorm(10000, mean=0, sd=0.2)
```

*#plot histogram with mean=0 & sigma=0.2*

```
hist(data, breaks=40, col="red", xlim=c(-3,3), ylim=c(0,1000),
     main="Histogram_0.2", col.main="blue")
```

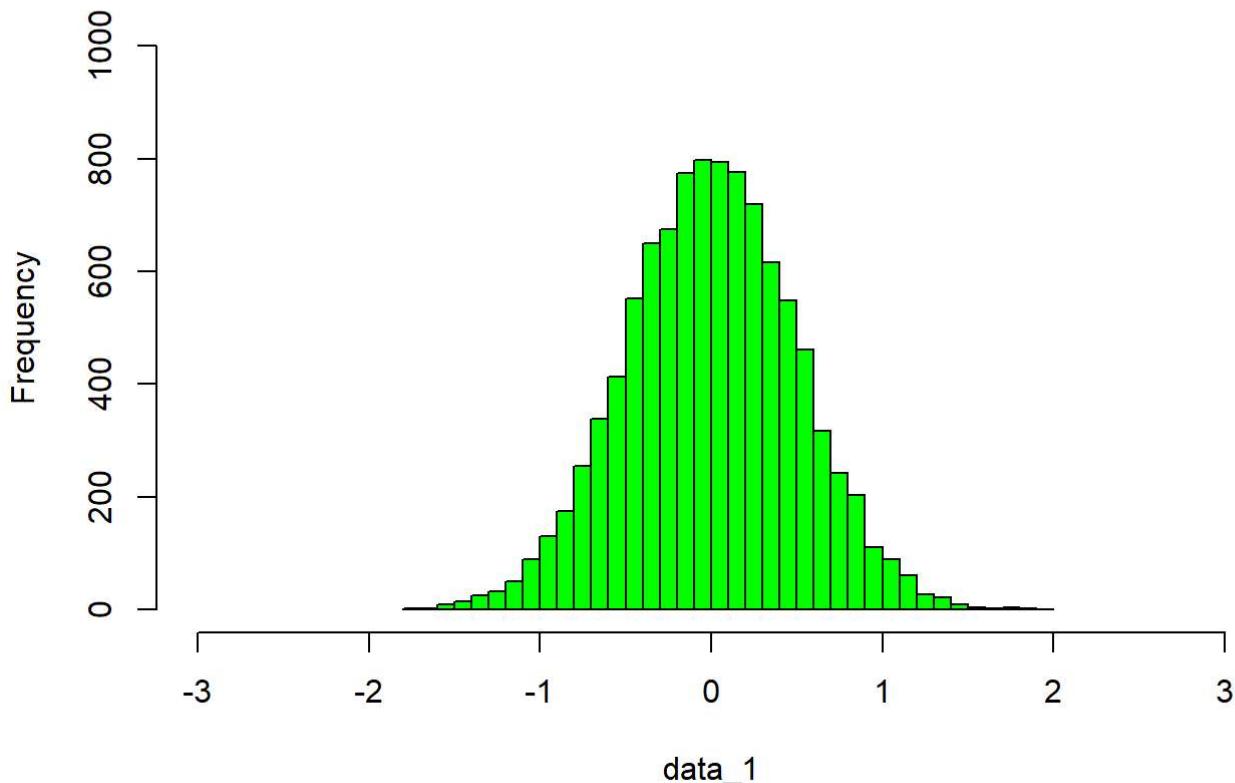
## Histogram\_0.2



```
# Generate 10000 numbers with mean=0, and Standard deviation=0.5
data_1 = rnorm(10000, mean=0, sd=0.5)

#plot histogram with mean=0 & sigma=0.5
hist(data_1, breaks=40, col="green", xlim=c(-3,3), ylim=c(0,1000),
      main="Histogram_0.5", col.main="blue")
```

## Histogram\_0.5



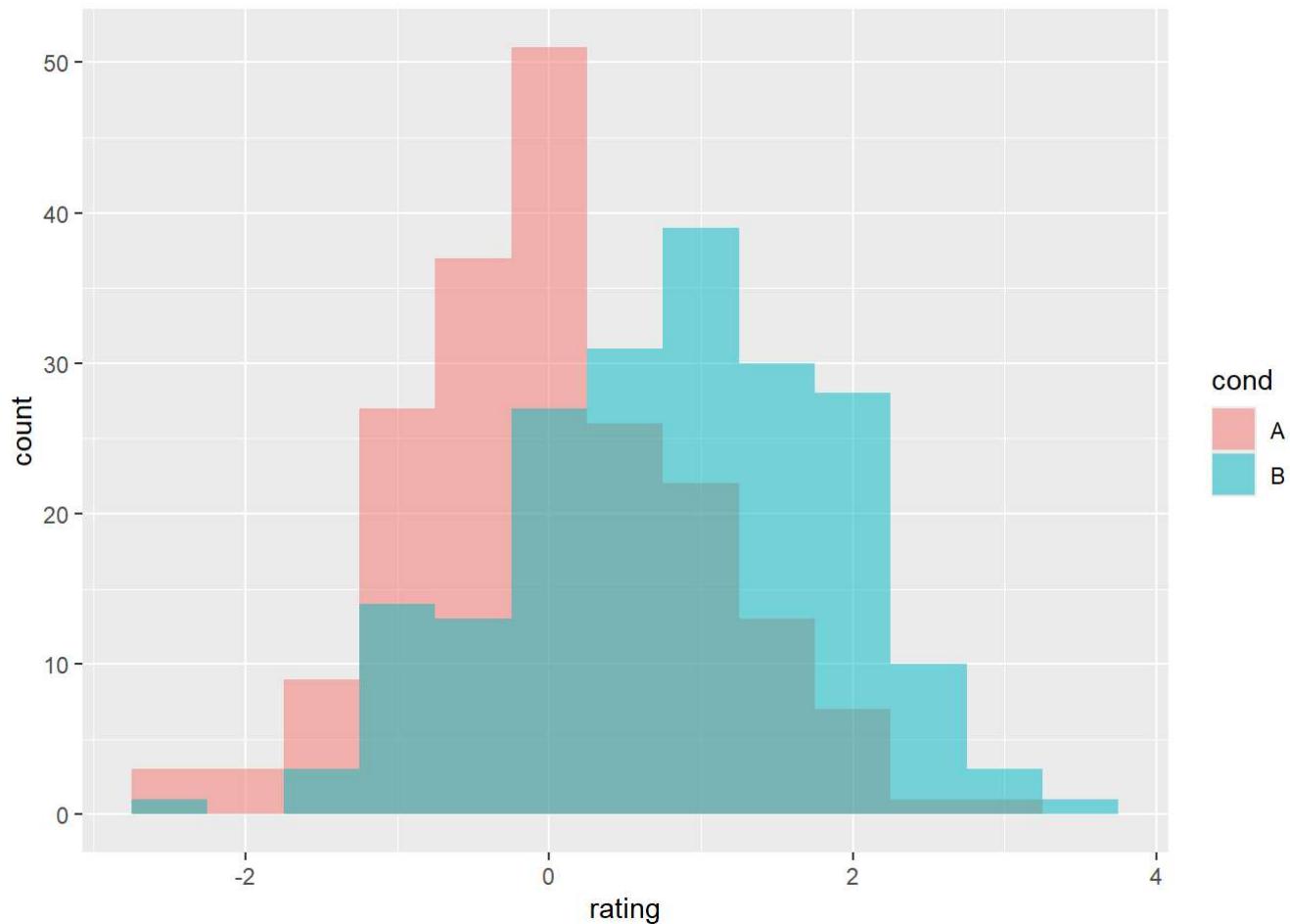
```
#Histogram for sigma 0.2 is much taller and narrower than the histogram for sigma 0.5
#This difference is due to standard deviation being smaller which will cause the data points to
cluster tightly around the mean.
```

```
#Part 3
```

```
#Loading the required package
library(ggplot2)

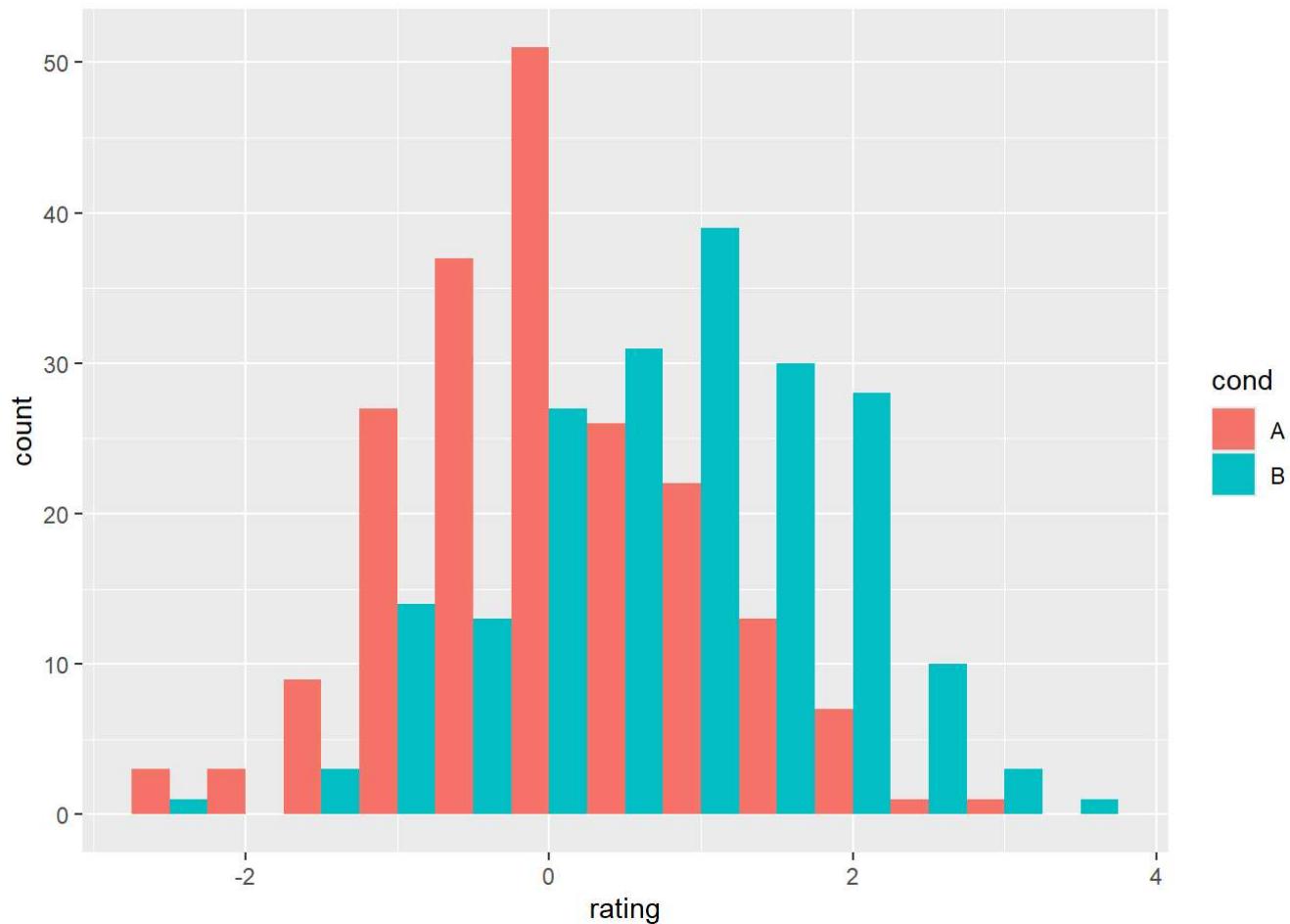
#(a)
dat <- data.frame(cond = factor(rep(c("A","B"), each=200)),
                  rating = c(rnorm(200),rnorm(200, mean=.8)))

#(b) Overlaid Histogram
ggplot(dat, aes(x=rating, fill=cond)) +
  geom_histogram(binwidth=.5, alpha=.5, position="identity")
```

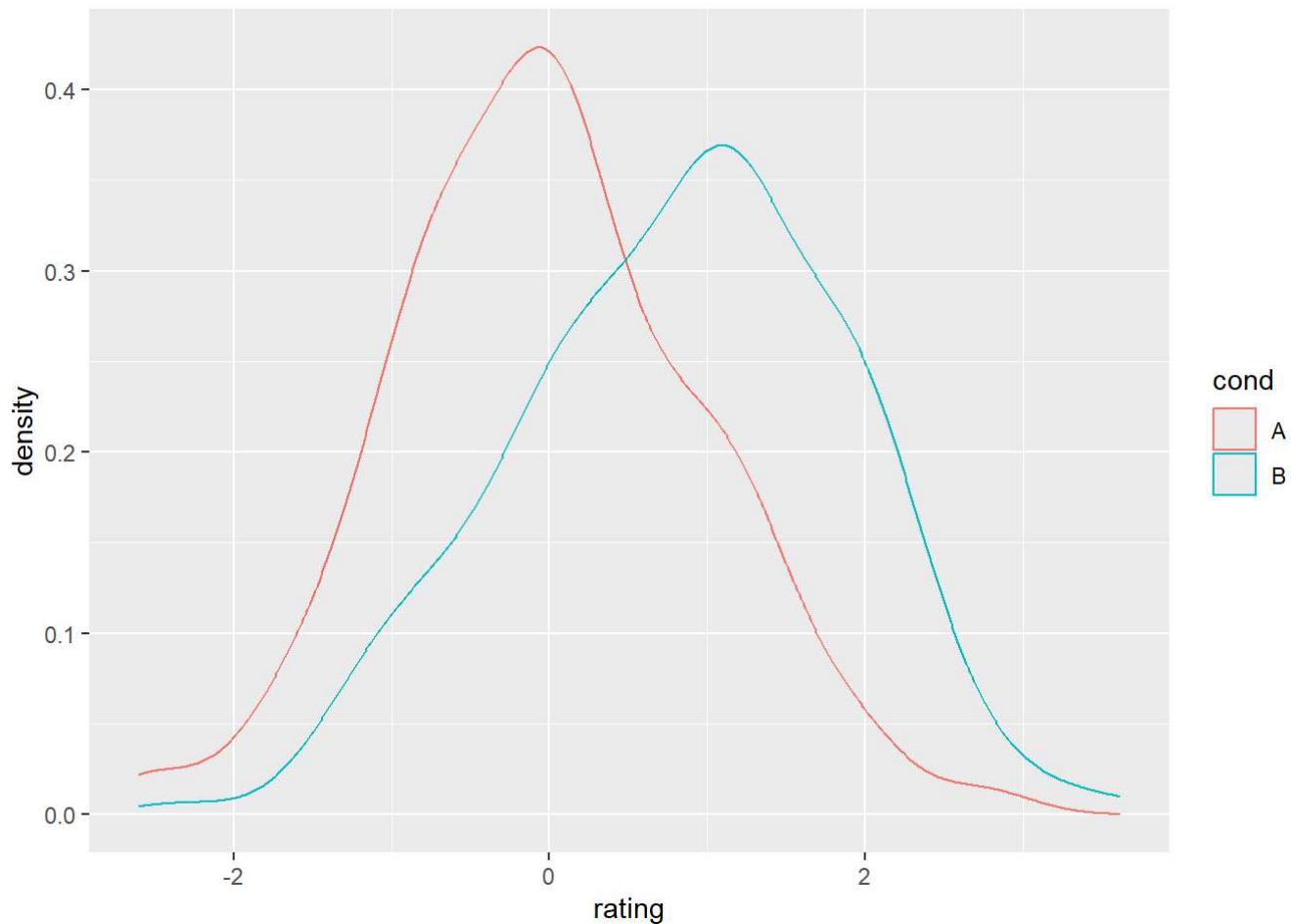


#(c) Interleaved histograms

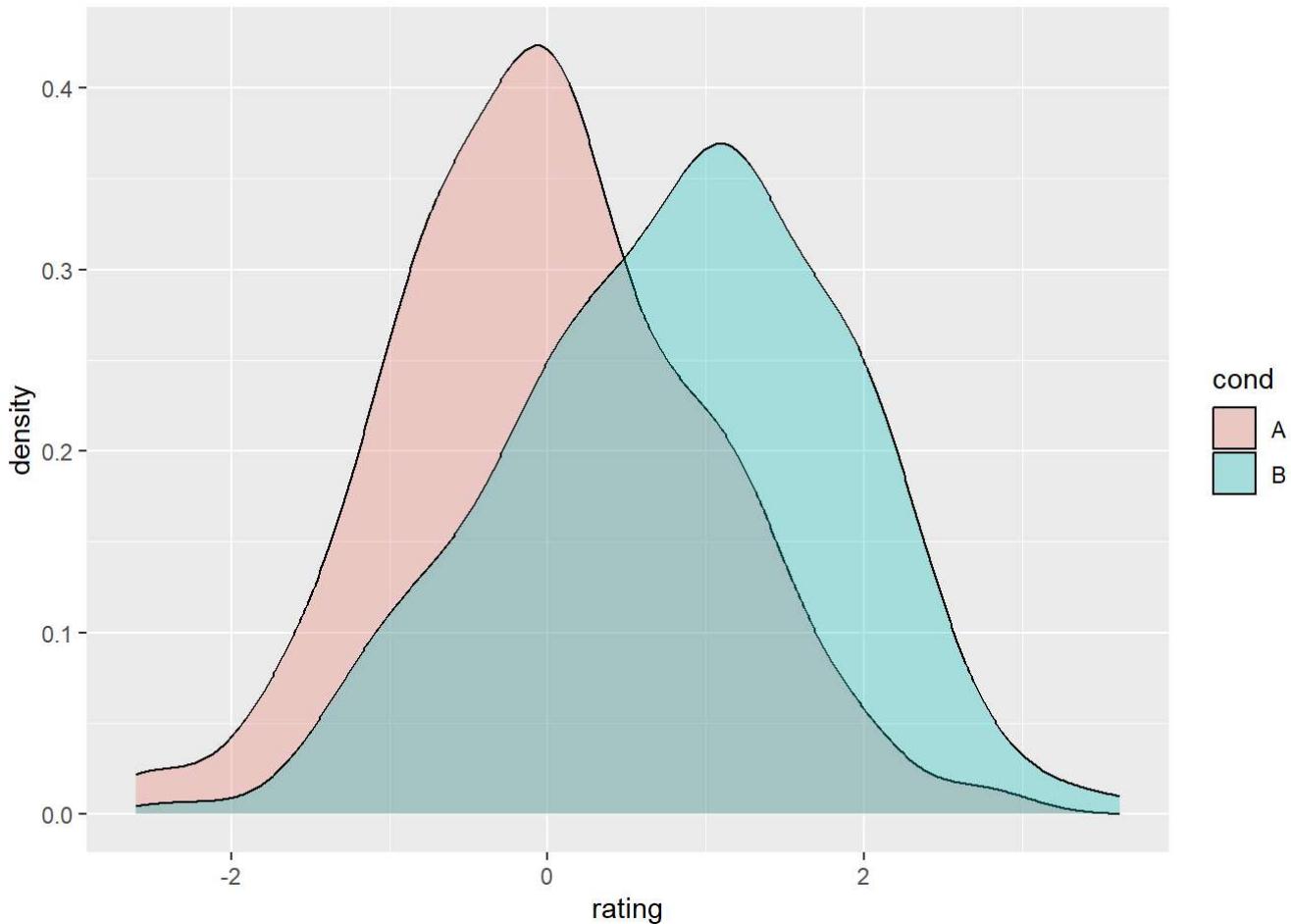
```
ggplot(dat, aes(x=rating, fill=cond)) + geom_histogram(binwidth=.5, position="dodge")
```



```
#(d) Density plots  
ggplot(dat, aes(x=rating, colour=cond)) + geom_density()
```



```
#(e) Density plots with semitransparent fill  
ggplot(dat, aes(x=rating, fill=cond)) + geom_density(alpha=.3)
```

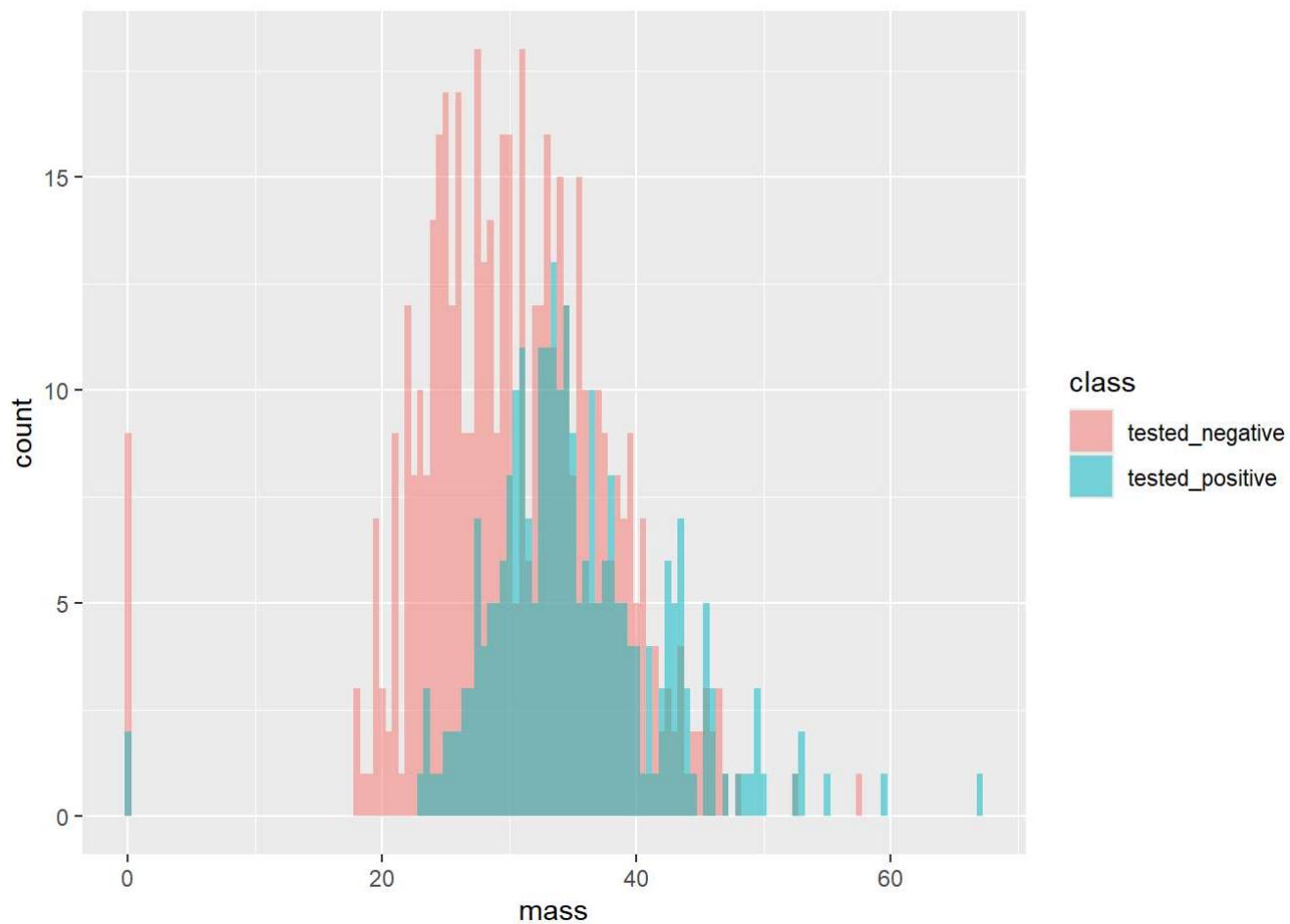


```
#(f) To plot all 4 types of histogram
```

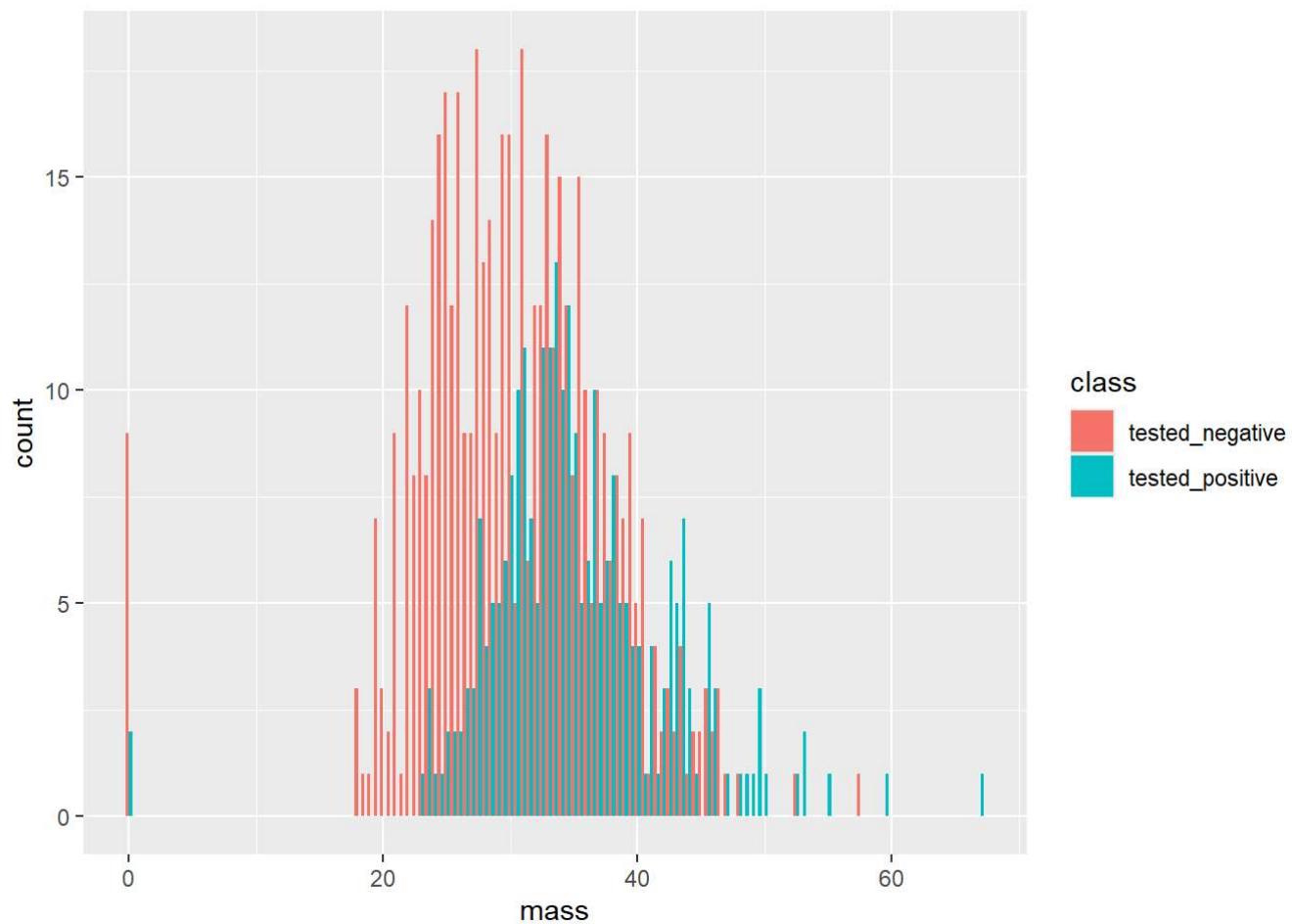
```
# Load in the data set
data.file <- file.path('diabetes_train.csv')
diabetes <- read.csv(data.file, header = TRUE, sep = ',')
head(diabetes)
```

```
##   preg  plas pres skin insu mass pedi age       class
## 1     6   148    72   35     0 33.6 0.627  50 tested_positive
## 2     1     85    66   29     0 26.6 0.351  31 tested_negative
## 3     8   183    64     0     0 23.3 0.672  32 tested_positive
## 4     1     89    66   23   94 28.1 0.167  21 tested_negative
## 5     0   137    40   35  168 43.1 2.288  33 tested_positive
## 6     5   116    74     0     0 25.6 0.201  30 tested_negative
```

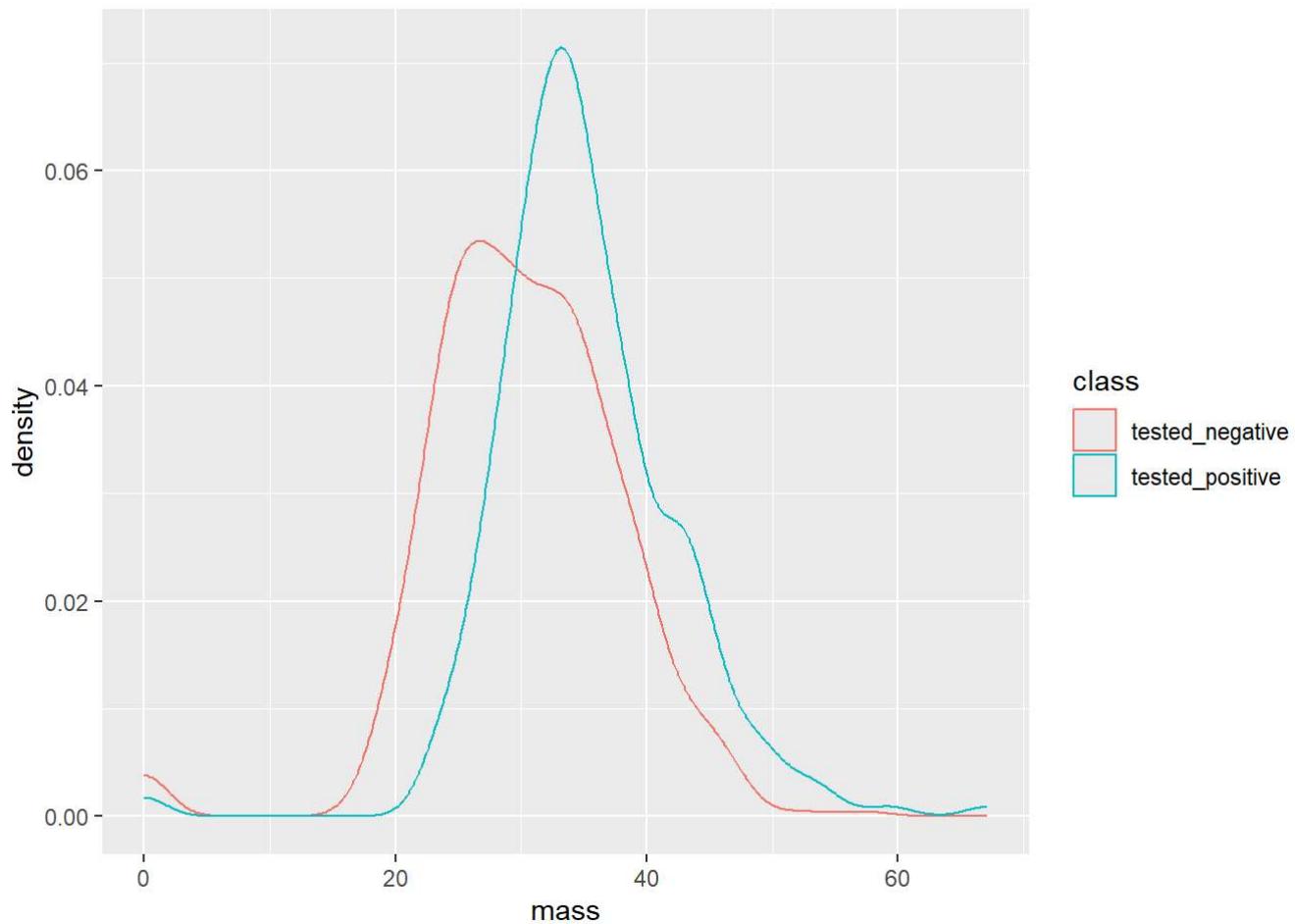
```
#Overlaid Histogram
ggplot(diabetes, aes(x=mass, fill=class)) +
  geom_histogram(binwidth=.5, alpha=.5, position="identity")
```



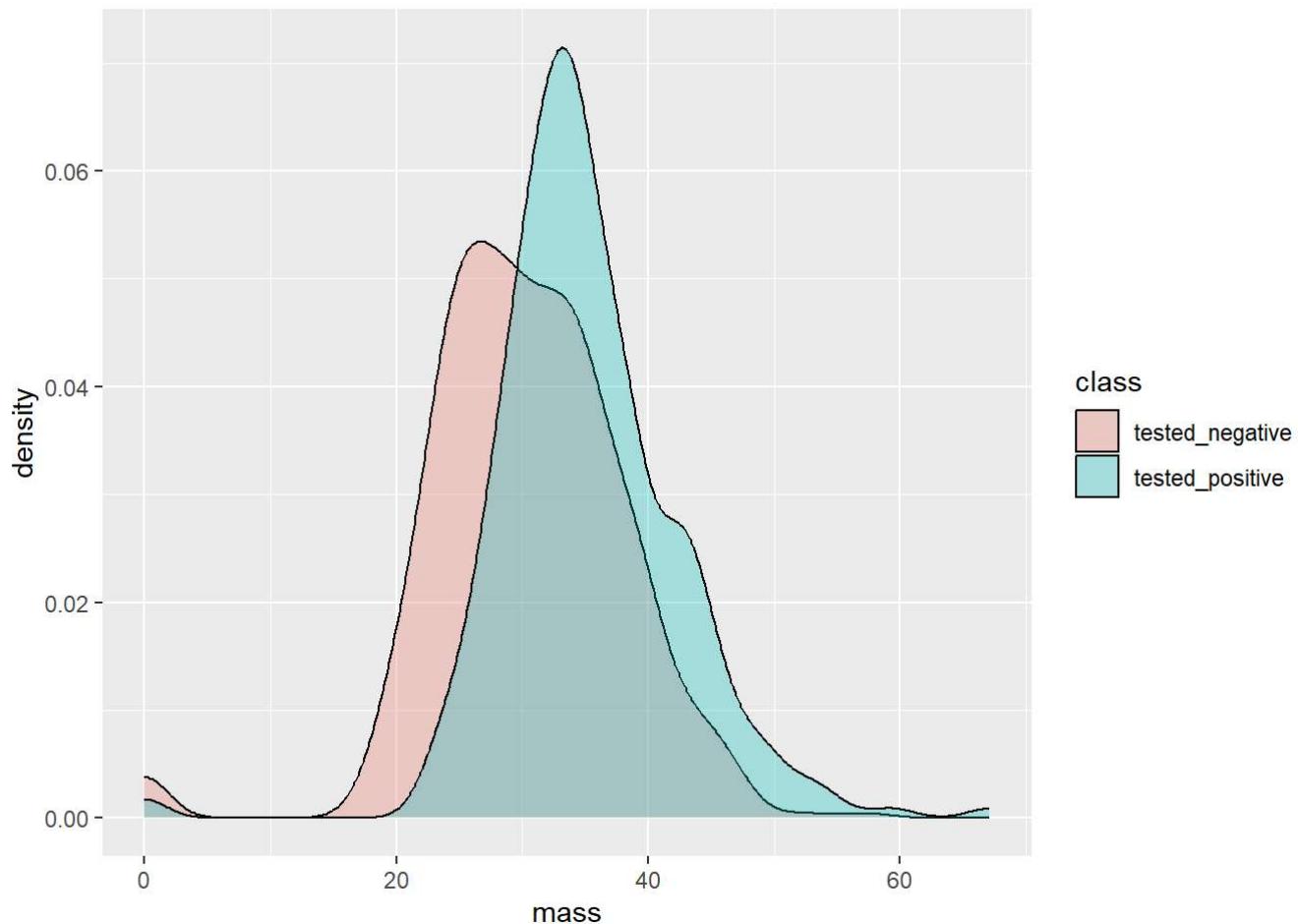
```
#Interleaved Histogram  
ggplot(diabetes, aes(x=mass, fill=class)) + geom_histogram(binwidth=.5, position="dodge")
```



```
#Density plots  
ggplot(diabetes, aes(x=mass, colour=class)) + geom_density()
```



```
#Density plots with semitransparent fill  
ggplot(diabetes, aes(x=mass, fill=class)) + geom_density(alpha=.3)
```



#Part 4

```
#Load in the data set
data_4 <- file.path('titanic.csv')
passengers <- read.csv(data_4, header = TRUE, sep = ',')
head(passengers)
```

```

##   X PassengerId Survived Pclass
## 1 0          1       0     3
## 2 1          2       1     1
## 3 2          3       1     3
## 4 3          4       1     1
## 5 4          5       0     3
## 6 5          6       0     3
##
##                                     Name      Sex Age SibSp Parch
## 1           Braund, Mr. Owen Harris   male  22    1    0
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female 38    1    0
## 3           Heikkinen, Miss. Laina female 26    0    0
## 4           Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35    1    0
## 5           Allen, Mr. William Henry   male  35    0    0
## 6           Moran, Mr. James        male  NA    0    0
##
##      Ticket      Fare Cabin Embarked
## 1 A/5 21171 7.2500            S
## 2 PC 17599 71.2833           C85
## 3 STON/O2. 3101282 7.9250            S
## 4 113803 53.1000           C123
## 5 373450  8.0500            S
## 6 330877  8.4583            Q

```

```

#Load Library
library(tidyverse)

```

```

## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓forcats   1.0.0     ✓ stringr   1.5.1
## ✓ lubridate 1.9.4     ✓ tibble    3.2.1
## ✓ purrr    1.0.4     ✓ tidyverse  1.3.1
## — Conflicts ————— tidyverse_conflicts() —
## ✘ dplyr::filter() masks stats::filter()
## ✘ dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

```

```

#(a)
passengers %>% drop_na() %>% summary()

```

```

##      X          PassengerId       Survived        Pclass
## Min.   : 0.0   Min.   : 1.0   Min.   :0.0000  Length:714
## 1st Qu.:221.2 1st Qu.:222.2  1st Qu.:0.0000  Class  :character
## Median :444.0  Median :445.0  Median :0.0000  Mode   :character
## Mean   :447.6  Mean   :448.6  Mean   :0.4062
## 3rd Qu.:676.8 3rd Qu.:677.8  3rd Qu.:1.0000
## Max.   :890.0  Max.   :891.0  Max.   :1.0000
##      Name          Sex          Age        SibSp
## Length:714      Length:714      Min.   : 0.42  Min.   :0.0000
## Class  :character  Class  :character  1st Qu.:20.12 1st Qu.:0.0000
## Mode   :character  Mode   :character  Median :28.00  Median :0.0000
##                               Mean   :29.70  Mean   :0.5126
##                               3rd Qu.:38.00 3rd Qu.:1.0000
##                               Max.   :80.00  Max.   :5.0000
##      Parch         Ticket        Fare        Cabin
## Min.   :0.0000  Length:714      Min.   : 0.00  Length:714
## 1st Qu.:0.0000  Class  :character  1st Qu.: 8.05  Class  :character
## Median :0.0000  Mode   :character  Median : 15.74  Mode   :character
## Mean   :0.4314
## 3rd Qu.:1.0000
## Max.   :6.0000
##      Embarked
## Length:714
## Class  :character
## Mode   :character
## 
## 
## 
```

*#The %>% operator is used to pipe the passengers dataset into the drop\_na() function, which removes any rows with missing values.*

*#Then the passenger dataet is pipe to summary function to provide statistics for each column*

*#(b)*

```
passengers %>% filter(Sex == "male")
```

```
##   X PassengerId Survived Pclass
## 1  0         1       0     3
## 2  4         5       0     3
## 3  5         6       0     3
## 4  6         7       0     1
## 5  7         8       0     3
## 6 12        13      0     3
## 7 13        14      0     3
## 8 16        17      0     3
## 9 17        18      1     2
## 10 20       21      0     2
## 11 21       22      1     2
## 12 23       24      1     1
## 13 26       27      0     3
## 14 27       28      0     1
## 15 29       30      0     3
## 16 30       31      0     1
## 17 33       34      0     2
## 18 34       35      0     1
## 19 35       36      0     1
## 20 36       37      1     3
## 21 37       38      0     ?
## 22 42       43      0     3
## 23 45       46      0     3
## 24 46       47      0     3
## 25 48       49      0     3
## 26 50       51      0     3
## 27 51       52      0     3
## 28 54       55      0     1
## 29 55       56      1     1
## 30 57       58      0     3
## 31 59       60      0     3
## 32 60       61      0     3
## 33 62       63      0     1
## 34 63       64      0     3
## 35 64       65      0     1
## 36 65       66      1     3
## 37 67       68      0     3
## 38 69       70      0     3
## 39 70       71      0     2
## 40 72       73      0     2
## 41 73       74      0     3
## 42 74       75      1     3
## 43 75       76      0     3
## 44 76       77      0     3
## 45 77       78      0     3
## 46 78       79      1     2
## 47 80       81      0     ?
## 48 81       82      1     3
## 49 83       84      0     1
## 50 86       87      0     3
## 51 87       88      0     3
```

## 572	881	882	0	3			
## 573	883	884	0	2			
## 574	884	885	0	3			
## 575	886	887	0	2			
## 576	889	890	1	1			
## 577	890	891	0	3			
##					Name	Sex	Age SibSp Parch
## 1					Braund, Mr. Owen Harris	male	22.00 1 0
## 2					Allen, Mr. William Henry	male	35.00 0 0
## 3					Moran, Mr. James	male	NA 0 0
## 4					McCarthy, Mr. Timothy J	male	54.00 0 0
## 5					Palsson, Master. Gosta Leonard	male	2.00 3 1
## 6					Saunderscock, Mr. William Henry	male	20.00 0 0
## 7					Andersson, Mr. Anders Johan	male	39.00 1 5
## 8					Rice, Master. Eugene	male	2.00 4 1
## 9					Williams, Mr. Charles Eugene	male	NA 0 0
## 10					Fynney, Mr. Joseph J	male	35.00 0 0
## 11					Beesley, Mr. Lawrence	male	34.00 0 0
## 12					Sloper, Mr. William Thompson	male	28.00 0 0
## 13					Emir, Mr. Farred Chehab	male	NA 0 0
## 14					Fortune, Mr. Charles Alexander	male	19.00 3 2
## 15					Todoroff, Mr. Lalio	male	NA 0 0
## 16					Uruchurtu, Don. Manuel E	male	40.00 0 0
## 17					Wheadon, Mr. Edward H	male	66.00 0 0
## 18					Meyer, Mr. Edgar Joseph	male	28.00 1 0
## 19					Holverson, Mr. Alexander Oskar	male	42.00 1 0
## 20					Mamee, Mr. Hanna	male	NA 0 0
## 21					Cann, Mr. Ernest Charles	male	21.00 0 0
## 22					Kraeff, Mr. Theodor	male	NA 0 0
## 23					Rogers, Mr. William John	male	NA 0 0
## 24					Lennon, Mr. Denis	male	NA 1 0
## 25					Samaan, Mr. Youssef	male	NA 2 0
## 26					Panula, Master. Juha Niilo	male	7.00 4 1
## 27					Nosworthy, Mr. Richard Cater	male	21.00 0 0
## 28					Ostby, Mr. Engelhart Cornelius	male	65.00 0 1
## 29					Woolner, Mr. Hugh	male	NA 0 0
## 30					Novel, Mr. Mansouer	male	28.50 0 0
## 31					Goodwin, Master. William Frederick	male	11.00 5 2
## 32					Sirayanian, Mr. Orsen	male	22.00 0 0
## 33					Harris, Mr. Henry Birkhardt	male	45.00 1 0
## 34					Skoog, Master. Harald	male	4.00 3 2
## 35					Stewart, Mr. Albert A	male	NA 0 0
## 36					Moubarek, Master. Gerios	male	NA 1 1
## 37					Crease, Mr. Ernest James	male	19.00 0 0
## 38					Kink, Mr. Vincenz	male	26.00 2 0
## 39					Jenkin, Mr. Stephen Curnow	male	32.00 0 0
## 40					Hood, Mr. Ambrose Jr	male	21.00 0 0
## 41					Chronopoulos, Mr. Apostolos	male	26.00 1 0
## 42					Bing, Mr. Lee	male	32.00 0 0
## 43					Moen, Mr. Sigurd Hansen	male	25.00 0 0
## 44					Stanaff, Mr. Ivan	male	NA 0 0
## 45					Moutal, Mr. Rahamin Haim	male	NA 0 0

## 566		Balkic, Mr. Cerin male 26.00	0	0
## 567		Carlsson, Mr. Frans Olof male 33.00	0	0
## 568		Vander Cruyssen, Mr. Victor male 47.00	0	0
## 569		Gustafsson, Mr. Alfred Ossian male 20.00	0	0
## 570		Petroff, Mr. Nedelio male 19.00	0	0
## 571		Laleff, Mr. Kristo male NA	0	0
## 572		Markun, Mr. Johann male 33.00	0	0
## 573		Banfield, Mr. Frederick James male 28.00	0	0
## 574		Sutefall, Mr. Henry Jr male 25.00	0	0
## 575		Montvila, Rev. Juozas male 27.00	0	0
## 576		Behr, Mr. Karl Howell male 26.00	0	0
## 577		Dooley, Mr. Patrick male 32.00	0	0
##	Ticket	Fare	Cabin	Embarked
## 1	A/5 21171	7.2500		S
## 2	373450	8.0500		S
## 3	330877	8.4583		Q
## 4	17463	51.8625	E46	S
## 5	349909	21.0750		S
## 6	A/5. 2151	8.0500		S
## 7	347082	31.2750		S
## 8	382652	29.1250		Q
## 9	244373	13.0000		S
## 10	239865	26.0000		S
## 11	248698	13.0000	D56	S
## 12	113788	35.5000	A6	S
## 13	2631	7.2250		C
## 14	19950	263.0000	C23 C25 C27	S
## 15	349216	7.8958		S
## 16	PC 17601	27.7208		C
## 17	C.A. 24579	10.5000		S
## 18	PC 17604	82.1708		C
## 19	113789	52.0000		S
## 20	2677	7.2292		C
## 21	A./5. 2152	8.0500		S
## 22	349253	7.8958		C
## 23	S.C./A.4. 23567	8.0500		S
## 24	370371	15.5000		Q
## 25	2662	21.6792		C
## 26	3101295	39.6875		S
## 27	A/4. 39886	7.8000		S
## 28	113509	61.9792	B30	C
## 29	19947	35.5000	C52	S
## 30	2697	7.2292		C
## 31	CA 2144	46.9000		S
## 32	2669	7.2292		C
## 33	36973	83.4750	C83	S
## 34	347088	27.9000		S
## 35	PC 17605	27.7208		C
## 36	2661	15.2458		C
## 37	S.P. 3464	8.1583		S
## 38	315151	8.6625		S
## 39	C.A. 33111	10.5000		S

## 560	350026	14.1083	S		
## 561	28134	11.5000	S		
## 562	233866	13.0000	S		
## 563	PC 17590	50.4958	A24	S	
## 564	345777	9.5000	S		
## 565	347742	11.1333	S		
## 566	349248	7.8958	S		
## 567	695	5.0000 B51 B53 B55	S		
## 568	345765	9.0000	S		
## 569	7534	9.8458	S		
## 570	349212	7.8958	S		
## 571	349217	7.8958	S		
## 572	349257	7.8958	S		
## 573	C.A./SOTON	34068	10.5000	S	
## 574	SOTON/OQ	392076	7.0500	S	
## 575		211536	13.0000	S	
## 576		111369	30.0000	C148	C
## 577		370376	7.7500	Q	

```
# The filter function filters out a dataset which contains only male in the 'Sex' Column  
#(c)  
passengers %>% arrange(desc(Fare))
```

```
##          X PassengerId Survived Pclass
## 1      258         259       1     1
## 2      679         680       1     1
## 3      737         738       1     1
## 4      27          28       0     1
## 5      88          89       1     1
## 6      341         342       1     1
## 7      438         439       0     1
## 8      311         312       1     1
## 9      742         743       1     1
## 10     118         119       0     1
## 11     299         300       1     1
## 12     380         381       1     ?
## 13     557         558       0     1
## 14     700         701       1     1
## 15     716         717       1     1
## 16     527         528       0     1
## 17     377         378       0     1
## 18     689         690       1     1
## 19     730         731       1     ?
## 20     779         780       1     1
## 21     318         319       1     1
## 22     856         857       1     1
## 23     268         269       1     1
## 24     332         333       0     1
## 25     609         610       1     1
## 26     297         298       0     1
## 27     305         306       1     1
## 28     498         499       0     1
## 29     708         709       1     1
## 30     31          32       1     1
## 31     195         196       1     1
## 32     269         270       1     1
## 33     325         326       1     1
## 34     373         374       0     1
## 35     319         320       1     1
## 36     337         338       1     1
## 37     334         335       1     1
## 38     660         661       1     1
## 39     390         391       1     ?
## 40     435         436       1     1
## 41     763         764       1     1
## 42     802         803       1     1
## 43     215         216       1     1
## 44     393         394       1     1
## 45     659         660       0     1
## 46     306         307       1     1
## 47     550         551       1     1
## 48     581         582       1     1
## 49     698         699       0     1
## 50     307         308       1     1
## 51     505         506       0     1
```

## 884	481	482	0	2
## 885	597	598	0	3
## 886	633	634	0	1
## 887	674	675	0	?
## 888	732	733	0	2
## 889	806	807	0	1
## 890	815	816	0	1
## 891	822	823	0	1
##				Name
## 1				Ward, Miss. Anna
## 2				Cardeza, Mr. Thomas Drake Martinez
## 3				Lesurer, Mr. Gustave J
## 4				Fortune, Mr. Charles Alexander
## 5				Fortune, Miss. Mabel Helen
## 6				Fortune, Miss. Alice Elizabeth
## 7				Fortune, Mr. Mark
## 8				Ryerson, Miss. Emily Borie
## 9				Ryerson, Miss. Susan Parker "Suzette"
## 10				Baxter, Mr. Quigg Edmond
## 11				Baxter, Mrs. James (Helene DeLaudeniere Chaput)
## 12				Bidois, Miss. Rosalie
## 13				Robbins, Mr. Victor
## 14				Astor, Mrs. John Jacob (Madeleine Talmadge Force)
## 15				Endres, Miss. Caroline Louise
## 16				Farthing, Mr. John
## 17				Widener, Mr. Harry Elkins
## 18				Madill, Miss. Georgette Alexandra
## 19				Allen, Miss. Elisabeth Walton
## 20				Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)
## 21				Wick, Miss. Mary Natalie
## 22				Wick, Mrs. George Dennick (Mary Hitchcock)
## 23				Graham, Mrs. William Thompson (Edith Junkins)
## 24				Graham, Mr. George Edward
## 25				Shutes, Miss. Elizabeth W
## 26				Allison, Miss. Helen Loraine
## 27				Allison, Master. Hudson Trevor
## 28				Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
## 29				Cleaver, Miss. Alice
## 30				Spencer, Mrs. William Augustus (Marie Eugenie)
## 31				Lurette, Miss. Elise
## 32				Bissette, Miss. Amelia
## 33				Young, Miss. Marie Grice
## 34				Ringhini, Mr. Sante
## 35				Spedden, Mrs. Frederic Oakley (Margareta Corning Stone)
## 36				Burns, Miss. Elizabeth Margaret
## 37				Frauenthal, Mrs. Henry William (Clara Heinsheimer)
## 38				Frauenthal, Dr. Henry William
## 39				Carter, Mr. William Ernest
## 40				Carter, Miss. Lucile Polk
## 41				Carter, Mrs. William Ernest (Lucile Polk)
## 42				Carter, Master. William Thornton II
## 43				Newell, Miss. Madeleine

```

## 876                                Betros, Mr. Tannous
## 877                                Leonard, Mr. Lionel
## 878                                Harrison, Mr. William
## 879                                Tornquist, Mr. William Henry
## 880                                Parkes, Mr. Francis "Frank"
## 881                                Johnson, Mr. William Cahoon Jr
## 882                                Cunningham, Mr. Alfred Fleming
## 883                                Campbell, Mr. William
## 884                                Frost, Mr. Anthony Wood "Archie"
## 885                                Johnson, Mr. Alfred
## 886                                Parr, Mr. William Henry Marsh
## 887                                Watson, Mr. Ennis Hastings
## 888                                Knight, Mr. Robert J
## 889                                Andrews, Mr. Thomas Jr
## 890                                Fry, Mr. Richard
## 891                                Reuchlin, Jonkheer. John George

```

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
## 1	female	35.00	0	0	PC 17755	512.3292	
## 2	male	36.00	0	1	PC 17755	512.3292	B51 B53 B55
## 3	male	35.00	0	0	PC 17755	512.3292	B101
## 4	male	19.00	3	2		19950 263.0000	C23 C25 C27
## 5	female	23.00	3	2		19950 263.0000	C23 C25 C27
## 6	female	24.00	3	2		19950 263.0000	C23 C25 C27
## 7	male	64.00	1	4		19950 263.0000	C23 C25 C27
## 8	female	18.00	2	2	PC 17608	262.3750	B57 B59 B63 B66
## 9	female	21.00	2	2	PC 17608	262.3750	B57 B59 B63 B66
## 10	male	24.00	0	1	PC 17558	247.5208	B58 B60
## 11	female	50.00	0	1	PC 17558	247.5208	B58 B60
## 12	female	42.00	0	0	PC 17757	227.5250	
## 13	male	NA	0	0	PC 17757	227.5250	
## 14	female	18.00	1	0	PC 17757	227.5250	C62 C64
## 15	female	38.00	0	0	PC 17757	227.5250	C45
## 16	male	NA	0	0	PC 17483	221.7792	C95
## 17	male	27.00	0	2		113503 211.5000	C82
## 18	female	15.00	0	1		24160 211.3375	B5
## 19	female	29.00	0	0		24160 211.3375	B5
## 20	female	43.00	0	1		24160 211.3375	B3
## 21	female	31.00	0	2		36928 164.8667	C7
## 22	female	45.00	1	1		36928 164.8667	
## 23	female	58.00	0	1	PC 17582	153.4625	C125
## 24	male	38.00	0	1	PC 17582	153.4625	C91
## 25	female	40.00	0	0	PC 17582	153.4625	C125
## 26	female	2.00	1	2		113781 151.5500	C22 C26
## 27	male	0.92	1	2		113781 151.5500	C22 C26
## 28	female	25.00	1	2		113781 151.5500	C22 C26
## 29	female	22.00	0	0		113781 151.5500	
## 30	female	NA	1	0	PC 17569	146.5208	B78
## 31	female	58.00	0	0	PC 17569	146.5208	B80
## 32	female	35.00	0	0	PC 17760	135.6333	C99
## 33	female	36.00	0	0	PC 17760	135.6333	C32
## 34	male	22.00	0	0	PC 17760	135.6333	
## 35	female	40.00	1	1		16966 134.5000	E34

## 868	male	19.00	0	0	365222	6.7500
## 869	female	18.00	0	0	365226	6.7500
## 870	male	34.00	0	0	3101264	6.4958
## 871	male	18.00	1	0	3101267	6.4958
## 872	male	43.00	0	0	C 7075	6.4500
## 873	male	34.50	0	0	2683	6.4375
## 874	male	61.00	0	0	345364	6.2375
## 875	male	33.00	0	0	695	5.0000
## 876	male	20.00	0	0	2648	4.0125
## 877	male	36.00	0	0	LINE	0.0000
## 878	male	40.00	0	0	112059	0.0000
## 879	male	25.00	0	0	LINE	0.0000
## 880	male	NA	0	0	239853	0.0000
## 881	male	19.00	0	0	LINE	0.0000
## 882	male	NA	0	0	239853	0.0000
## 883	male	NA	0	0	239853	0.0000
## 884	male	NA	0	0	239854	0.0000
## 885	male	49.00	0	0	LINE	0.0000
## 886	male	NA	0	0	112052	0.0000
## 887	male	NA	0	0	239856	0.0000
## 888	male	NA	0	0	239855	0.0000
## 889	male	39.00	0	0	112050	0.0000
## 890	male	NA	0	0	112058	0.0000
## 891	male	38.00	0	0	19972	0.0000

## Embarked

## 1	C
## 2	C
## 3	C
## 4	S
## 5	S
## 6	S
## 7	S
## 8	C
## 9	C
## 10	C
## 11	C
## 12	C
## 13	C
## 14	C
## 15	C
## 16	S
## 17	C
## 18	S
## 19	S
## 20	S
## 21	S
## 22	S
## 23	S
## 24	S
## 25	S
## 26	S
## 27	S

```
## 860      S
## 861      S
## 862      S
## 863      S
## 864      S
## 865      S
## 866      Q
## 867      Q
## 868      Q
## 869      Q
## 870      S
## 871      S
## 872      S
## 873      C
## 874      S
## 875      S
## 876      C
## 877      S
## 878      S
## 879      S
## 880      S
## 881      S
## 882      S
## 883      S
## 884      S
## 885      S
## 886      S
## 887      S
## 888      S
## 889      S
## 890      S
## 891      S
```

*#The arrange function arranges the dataset in decending order according to the 'Flare' column*

```
 #(d)
passengers %>% mutate(FamSize = Parch + SibSp)
```

```
##   X PassengerId Survived Pclass
## 1 0         1       0     3
## 2 1         2       1     1
## 3 2         3       1     3
## 4 3         4       1     1
## 5 4         5       0     3
## 6 5         6       0     3
## 7 6         7       0     1
## 8 7         8       0     3
## 9 8         9       1     3
## 10 9        10      1     2
## 11 10       11      1     3
## 12 11       12      1     1
## 13 12       13      0     3
## 14 13       14      0     3
## 15 14       15      0     3
## 16 15       16      1     2
## 17 16       17      0     3
## 18 17       18      1     2
## 19 18       19      0     3
## 20 19       20      1     3
## 21 20       21      0     2
## 22 21       22      1     2
## 23 22       23      1     3
## 24 23       24      1     1
## 25 24       25      0     3
## 26 25       26      1     ?
## 27 26       27      0     3
## 28 27       28      0     1
## 29 28       29      1     3
## 30 29       30      0     3
## 31 30       31      0     1
## 32 31       32      1     1
## 33 32       33      1     3
## 34 33       34      0     2
## 35 34       35      0     1
## 36 35       36      0     1
## 37 36       37      1     3
## 38 37       38      0     ?
## 39 38       39      0     3
## 40 39       40      1     3
## 41 40       41      0     3
## 42 41       42      0     2
## 43 42       43      0     3
## 44 43       44      1     2
## 45 44       45      1     3
## 46 45       46      0     3
## 47 46       47      0     3
## 48 47       48      1     3
## 49 48       49      0     3
## 50 49       50      0     3
## 51 50       51      0     3
```

## 884 883	884	0	2
## 885 884	885	0	3
## 886 885	886	0	3
## 887 886	887	0	2
## 888 887	888	1	1
## 889 888	889	0	?
## 890 889	890	1	1
## 891 890	891	0	3
##			Name
## 1			Braund, Mr. Owen Harris
## 2			Cumings, Mrs. John Bradley (Florence Briggs Thayer)
## 3			Heikkinen, Miss. Laina
## 4			Futrelle, Mrs. Jacques Heath (Lily May Peel)
## 5			Allen, Mr. William Henry
## 6			Moran, Mr. James
## 7			McCarthy, Mr. Timothy J
## 8			Palsson, Master. Gosta Leonard
## 9			Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
## 10			Nasser, Mrs. Nicholas (Adele Achem)
## 11			Sandstrom, Miss. Marguerite Rut
## 12			Bonnell, Miss. Elizabeth
## 13			Saundercock, Mr. William Henry
## 14			Andersson, Mr. Anders Johan
## 15			Vestrom, Miss. Hulda Amanda Adolfina
## 16			Hewlett, Mrs. (Mary D Kingcome)
## 17			Rice, Master. Eugene
## 18			Williams, Mr. Charles Eugene
## 19			Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)
## 20			Masselmani, Mrs. Fatima
## 21			Fynney, Mr. Joseph J
## 22			Beesley, Mr. Lawrence
## 23			McGowan, Miss. Anna "Annie"
## 24			Sloper, Mr. William Thompson
## 25			Palsson, Miss. Torborg Danira
## 26			Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)
## 27			Emir, Mr. Farred Chehab
## 28			Fortune, Mr. Charles Alexander
## 29			O'Dwyer, Miss. Ellen "Nellie"
## 30			Todoroff, Mr. Lalio
## 31			Uruchurtu, Don. Manuel E
## 32			Spencer, Mrs. William Augustus (Marie Eugenie)
## 33			Glynn, Miss. Mary Agatha
## 34			Wheadon, Mr. Edward H
## 35			Meyer, Mr. Edgar Joseph
## 36			Holverson, Mr. Alexander Oskar
## 37			Mamee, Mr. Hanna
## 38			Cann, Mr. Ernest Charles
## 39			Vander Planke, Miss. Augusta Maria
## 40			Nicola-Yarred, Miss. Jamila
## 41			Ahlin, Mrs. Johan (Johanna Persdotter Larsson)
## 42			Turpin, Mrs. William John Robert (Dorothy Ann Wonnacott)
## 43			Kraeff, Mr. Theodor

## 876					Najib, Miss. Adele Kiamie "Jane"		
## 877					Gustafsson, Mr. Alfred Ossian		
## 878					Petroff, Mr. Nedelio		
## 879					Laleff, Mr. Kristo		
## 880					Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)		
## 881					Shelley, Mrs. William (Imanita Parrish Hall)		
## 882					Markun, Mr. Johann		
## 883					Dahlberg, Miss. Gerda Ulrika		
## 884					Banfield, Mr. Frederick James		
## 885					Sutefall, Mr. Henry Jr		
## 886					Rice, Mrs. William (Margaret Norton)		
## 887					Montvila, Rev. Juozas		
## 888					Graham, Miss. Margaret Edith		
## 889					Johnston, Miss. Catherine Helen "Carrie"		
## 890					Behr, Mr. Karl Howell		
## 891					Dooley, Mr. Patrick		
##	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
## 1	male	22.00	1	0	A/5 21171	7.2500	
## 2	female	38.00	1	0	PC 17599	71.2833	C85
## 3	female	26.00	0	0	STON/O2. 3101282	7.9250	
## 4	female	35.00	1	0		113803	53.1000
## 5	male	35.00	0	0		373450	8.0500
## 6	male	NA	0	0		330877	8.4583
## 7	male	54.00	0	0		17463	51.8625
## 8	male	2.00	3	1		349909	21.0750
## 9	female	27.00	0	2		347742	11.1333
## 10	female	14.00	1	0		237736	30.0708
## 11	female	4.00	1	1		PP 9549	16.7000
## 12	female	58.00	0	0		113783	26.5500
## 13	male	20.00	0	0		A/5. 2151	8.0500
## 14	male	39.00	1	5		347082	31.2750
## 15	female	14.00	0	0		350406	7.8542
## 16	female	55.00	0	0		248706	16.0000
## 17	male	2.00	4	1		382652	29.1250
## 18	male	NA	0	0		244373	13.0000
## 19	female	31.00	1	0		345763	18.0000
## 20	female	NA	0	0		2649	7.2250
## 21	male	35.00	0	0		239865	26.0000
## 22	male	34.00	0	0		248698	13.0000
## 23	female	15.00	0	0		330923	8.0292
## 24	male	28.00	0	0		113788	35.5000
## 25	female	8.00	3	1		349909	21.0750
## 26	female	38.00	1	5		347077	31.3875
## 27	male	NA	0	0		2631	7.2250
## 28	male	19.00	3	2		19950	263.0000
## 29	female	NA	0	0		330959	7.8792
## 30	male	NA	0	0		349216	7.8958
## 31	male	40.00	0	0		PC 17601	27.7208
## 32	female	NA	1	0		PC 17569	146.5208
## 33	female	NA	0	0		335677	7.7500
## 34	male	66.00	0	0		C.A. 24579	10.5000
## 35	male	28.00	1	0		PC 17604	82.1708

## 868	male	31.00	0	0	PC	17590	50.4958	A24
## 869	male	NA	0	0		345777	9.5000	
## 870	male	4.00	1	1		347742	11.1333	
## 871	male	26.00	0	0		349248	7.8958	
## 872	female	47.00	1	1		11751	52.5542	D35
## 873	male	33.00	0	0		695	5.0000	B51 B53 B55
## 874	male	47.00	0	0		345765	9.0000	
## 875	female	28.00	1	0	P/PP	3381	24.0000	
## 876	female	15.00	0	0		2667	7.2250	
## 877	male	20.00	0	0		7534	9.8458	
## 878	male	19.00	0	0		349212	7.8958	
## 879	male	NA	0	0		349217	7.8958	
## 880	female	56.00	0	1		11767	83.1583	C50
## 881	female	25.00	0	1		230433	26.0000	
## 882	male	33.00	0	0		349257	7.8958	
## 883	female	22.00	0	0		7552	10.5167	
## 884	male	28.00	0	0	C.A./SOTON	34068	10.5000	
## 885	male	25.00	0	0	SOTON/OQ	392076	7.0500	
## 886	female	39.00	0	5		382652	29.1250	
## 887	male	27.00	0	0		211536	13.0000	
## 888	female	19.00	0	0		112053	30.0000	B42
## 889	female	NA	1	2	W./C.	6607	23.4500	
## 890	male	26.00	0	0		111369	30.0000	
## 891	male	32.00	0	0		370376	7.7500	C148

## Embarked FamSize

## 1	S	1	
## 2	C	1	
## 3	S	0	
## 4	S	1	
## 5	S	0	
## 6	Q	0	
## 7	S	0	
## 8	S	4	
## 9	S	2	
## 10	C	1	
## 11	S	2	
## 12	S	0	
## 13	S	0	
## 14	S	6	
## 15	S	0	
## 16	S	0	
## 17	Q	5	
## 18	S	0	
## 19	S	1	
## 20	C	0	
## 21	S	0	
## 22	S	0	
## 23	Q	0	
## 24	S	0	
## 25	S	4	
## 26	S	6	
## 27	C	0	

```

## 860      C      0
## 861      S      2
## 862      S      1
## 863      S      0
## 864      S     10
## 865      S      0
## 866      S      0
## 867      C      1
## 868      S      0
## 869      S      0
## 870      S      2
## 871      S      0
## 872      S      2
## 873      S      0
## 874      S      0
## 875      C      1
## 876      C      0
## 877      S      0
## 878      S      0
## 879      S      0
## 880      C      1
## 881      S      1
## 882      S      0
## 883      S      0
## 884      S      0
## 885      S      0
## 886      Q      5
## 887      S      0
## 888      S      0
## 889      S      3
## 890      C      0
## 891      Q      0

```

*#The mutate function creates a new column called 'FamSize' by adding 'Parch' column and 'SibSp'  
#Parch column represents the number of parents or children abroad  
#The 'SibSp' column represents the number of siblings or spouse abroad.  
#Combining these 2 columns will provide the family size for each passenger*

```

#(e)
passengers %>% group_by(Sex) %>% summarise(meanFare = mean(Fare), numSurv = sum(Survived))

```

```

## # A tibble: 2 × 3
##   Sex     meanFare numSurv
##   <chr>     <dbl>    <int>
## 1 female     44.5     233
## 2 male       25.5     109

```

```
#The group_by function groups the data by Sex  
#The summarise function calculates the mean fare for each sex and the total number of survivors  
for each sex.  
#The mean fare for females and males were 44.5 and 25.5 respectively  
#The number of survivors were higher in female(233) than in male(109)
```

#Part 5

```
quantile(diabetes$skin, probs = c(.1, .3, .5, .6))
```

```
## 10% 30% 50% 60%  
## 0 10 23 27
```

```
#The skin attributes for 10% percentile is 0, 30% percentile is 10, 50% percentile is 23 and 60%  
percentile is 27
```