

Homework #1 – Carmen Schuler

Question #1

```

> #1.(a) read text file which creates a data frame called Su
> Su <- read.delim("G:/My Drive/Personal/Grad Program/CSC 587 - Adv Data Mining/Scripts/datamining-main/Rscripts/data/Su_raw_matrix.txt")
> #1.(b)
> mean(Su$Liver_2.CEL) #mean of Liver_2.CEL column
[1] 241.8246
> sd(Su$Liver_2.CEL) #standard deviation of Liver_2.CEL column
[1] 1133.352
> #1.(c)
> colMeans(Su) #mean or average of each column
  Brain_1.CEL   Brain_2.CEL Fetal_brain_1.CEL Fetal_brain_2.CEL Fetal_liver_1.CEL Fetal_liver_2.CEL   Liver_1.CEL
    204.9763     315.0924      198.3439       267.6551       209.8722       399.1482       160.8558
  Liver_2.CEL
    241.825
> colSums(Su) #sum of each column
  Brain_1.CEL   Brain_2.CEL Fetal_brain_1.CEL Fetal_brain_2.CEL Fetal_liver_1.CEL Fetal_liver_2.CEL   Liver_1.CEL
    2588031     3978357      2504290       3379413       2649846       5039645       2030966
  Liver_2.CEL
    3053278

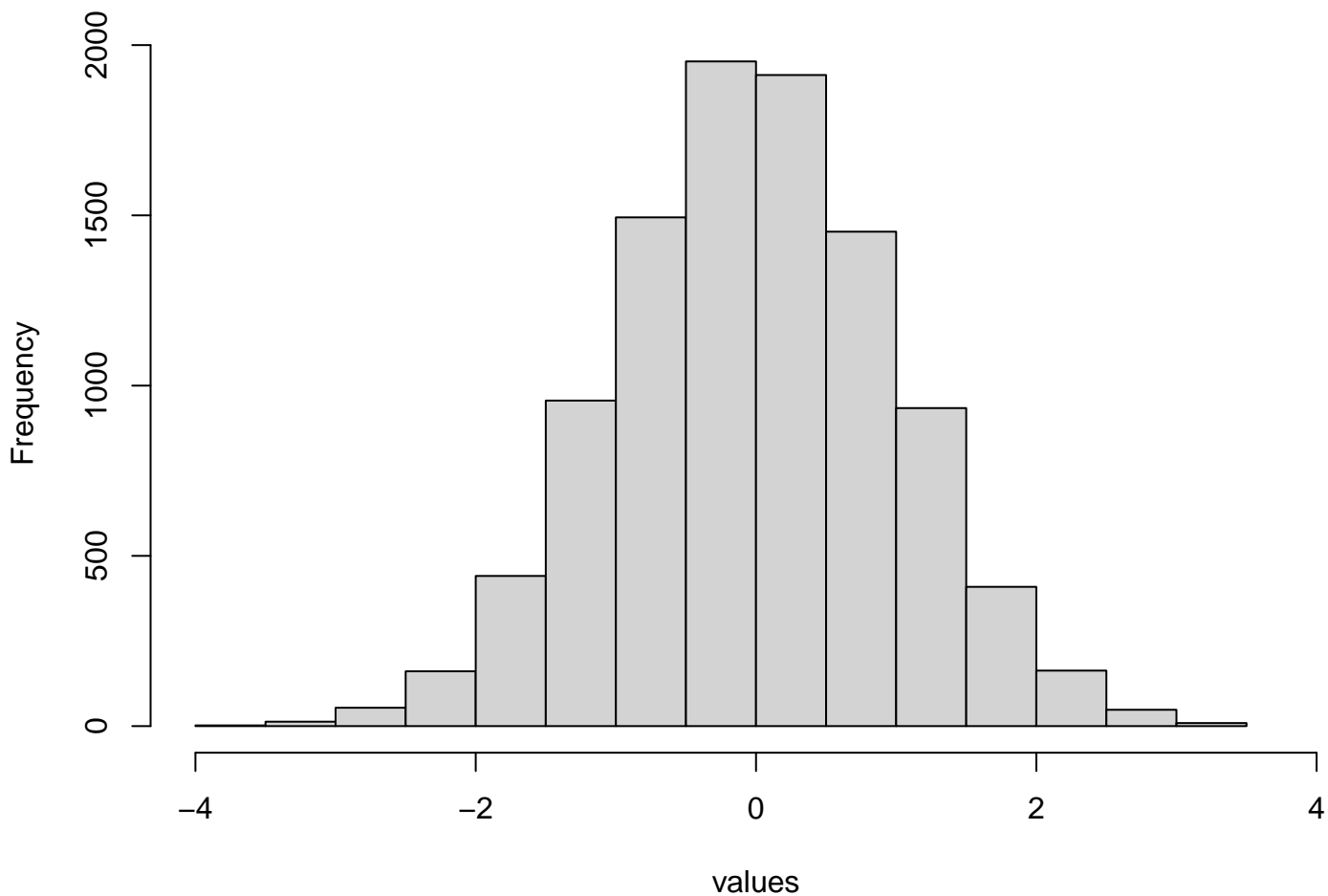
>

```

Question #2

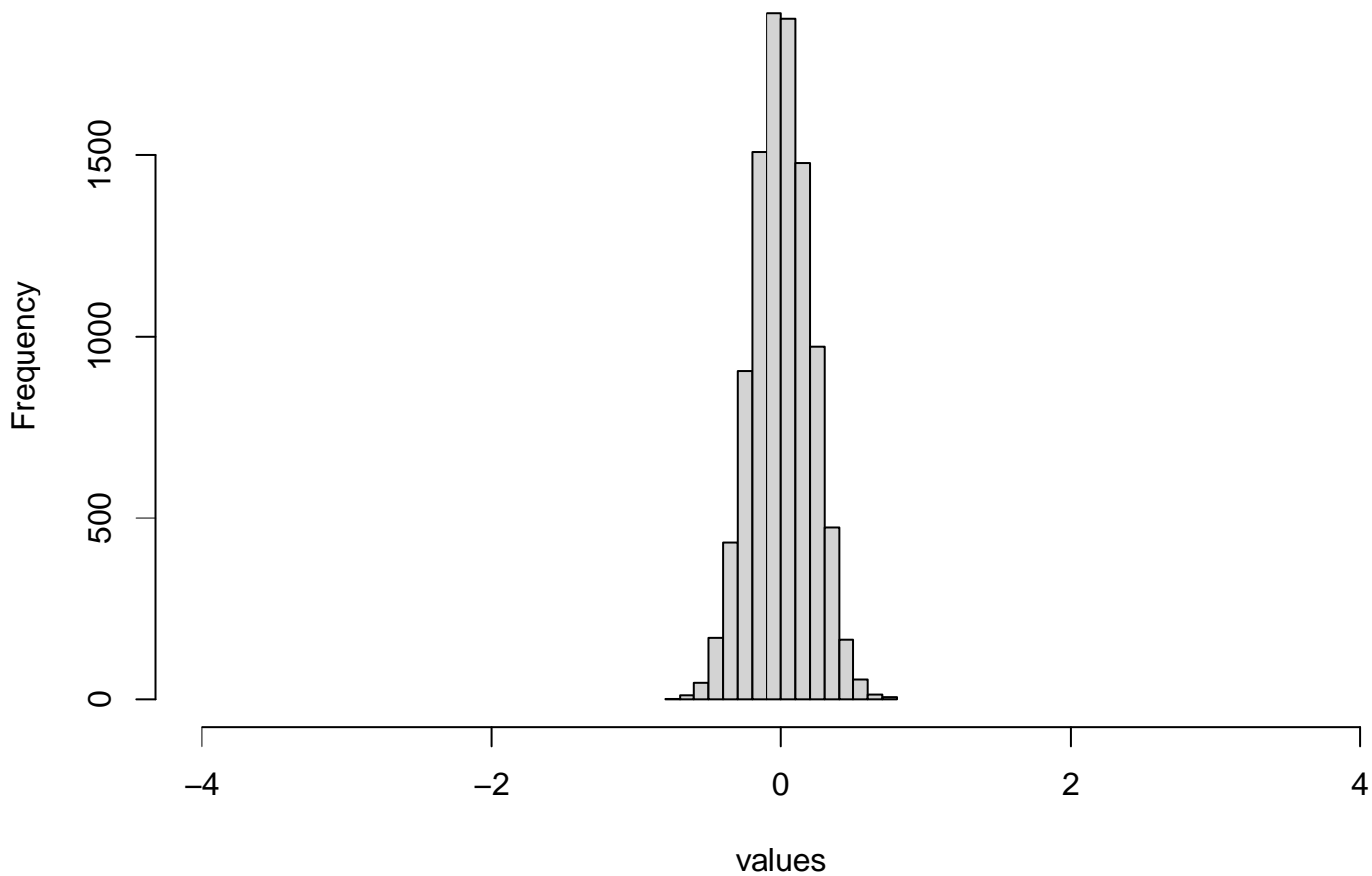
mean=0, sd=1

Histogram of values



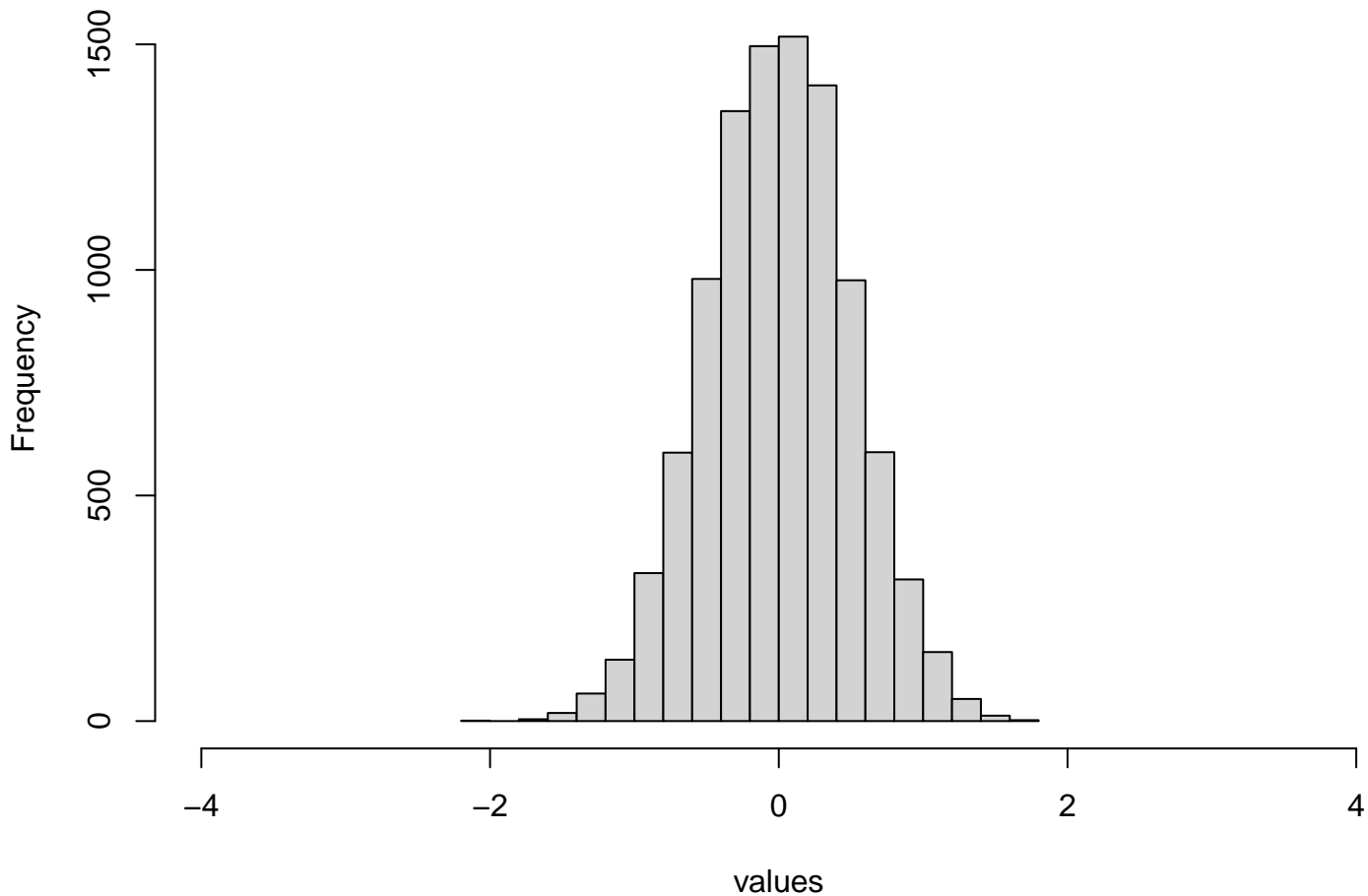
mean=0, sigma=0.2

Histogram of values



mean=0, sigma=0.5

Histogram of values



The standard deviation (or sigma) indicates how close the data is to the mean of the data. In these cases, we chose 10,000 random numbers in a normal distribution. In the first case (page #1), with the sigma at 1, the points are more spread out and further from the mean since the standard deviation is higher. In the second case (page #2) the sigma is 0.2, and the lowest of the three examples so there are more data points closer to the mean. For the third case (page #3) the sigma is 0.5 which shows that the data is further from the mean than the 0.2 sigma but not as far apart as the 1 sigma. Therefore, the higher the standard deviation, the more the data will be spread apart from the mean, and the lower the standard deviation, the closer the data points will be to the mean.

```

` `` {r}
#2.(a) and (b)
values = rnorm(10000, mean = 0, sd = 1) #changed sd from 1 then to 0.2 and then to 0.5 and ran again each time
hist(values, xlim = c(-4,4))
` ``

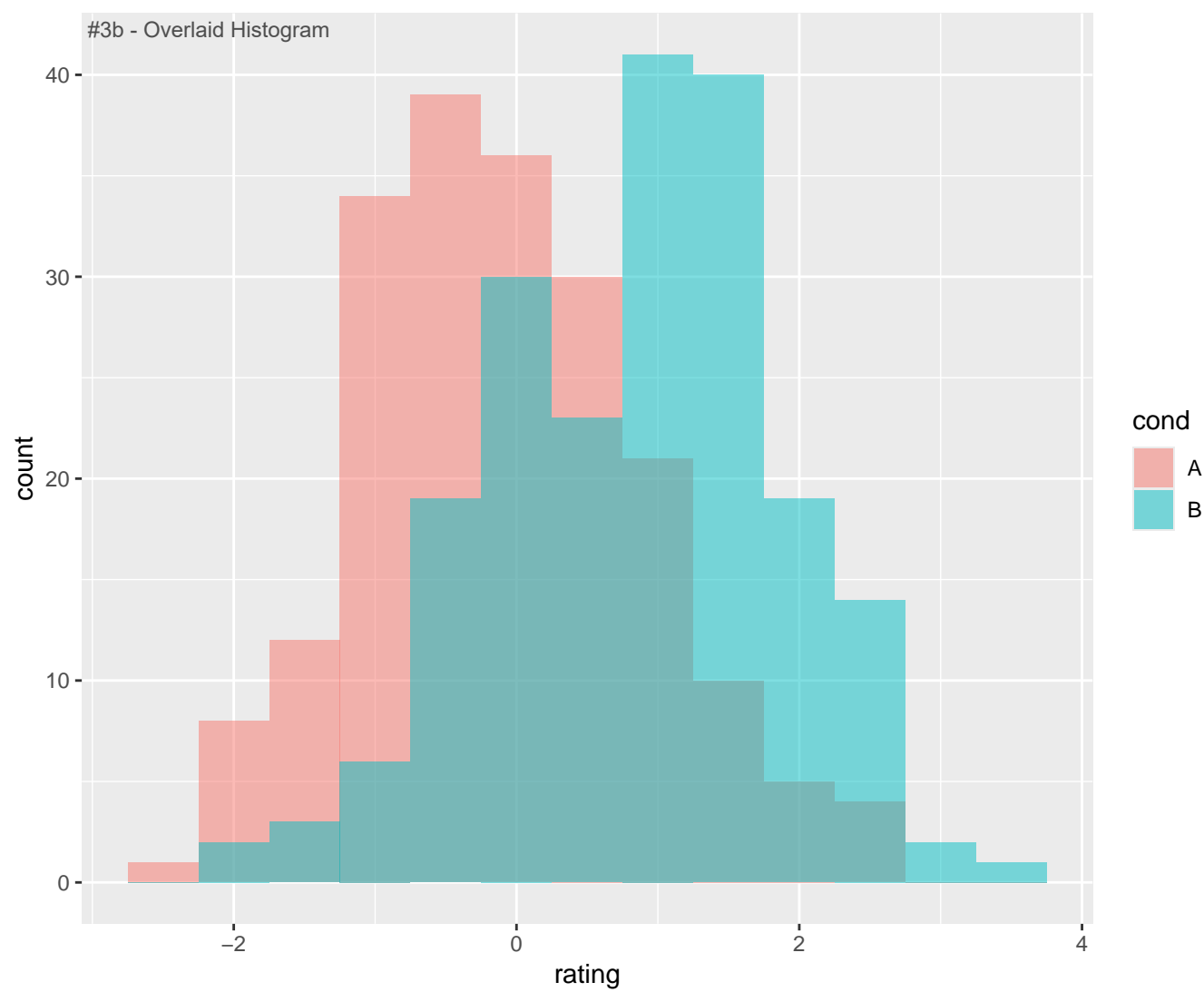
` `` {r}
#2.(a) and (b)
values = rnorm(10000, mean = 0, sd = 0.2)
hist(values, xlim = c(-4,4))
` ``

` `` {r}
#2.(a) and (b)
values = rnorm(10000, mean = 0, sd = 0.5)
hist(values, xlim = c(-4,4))
` ``

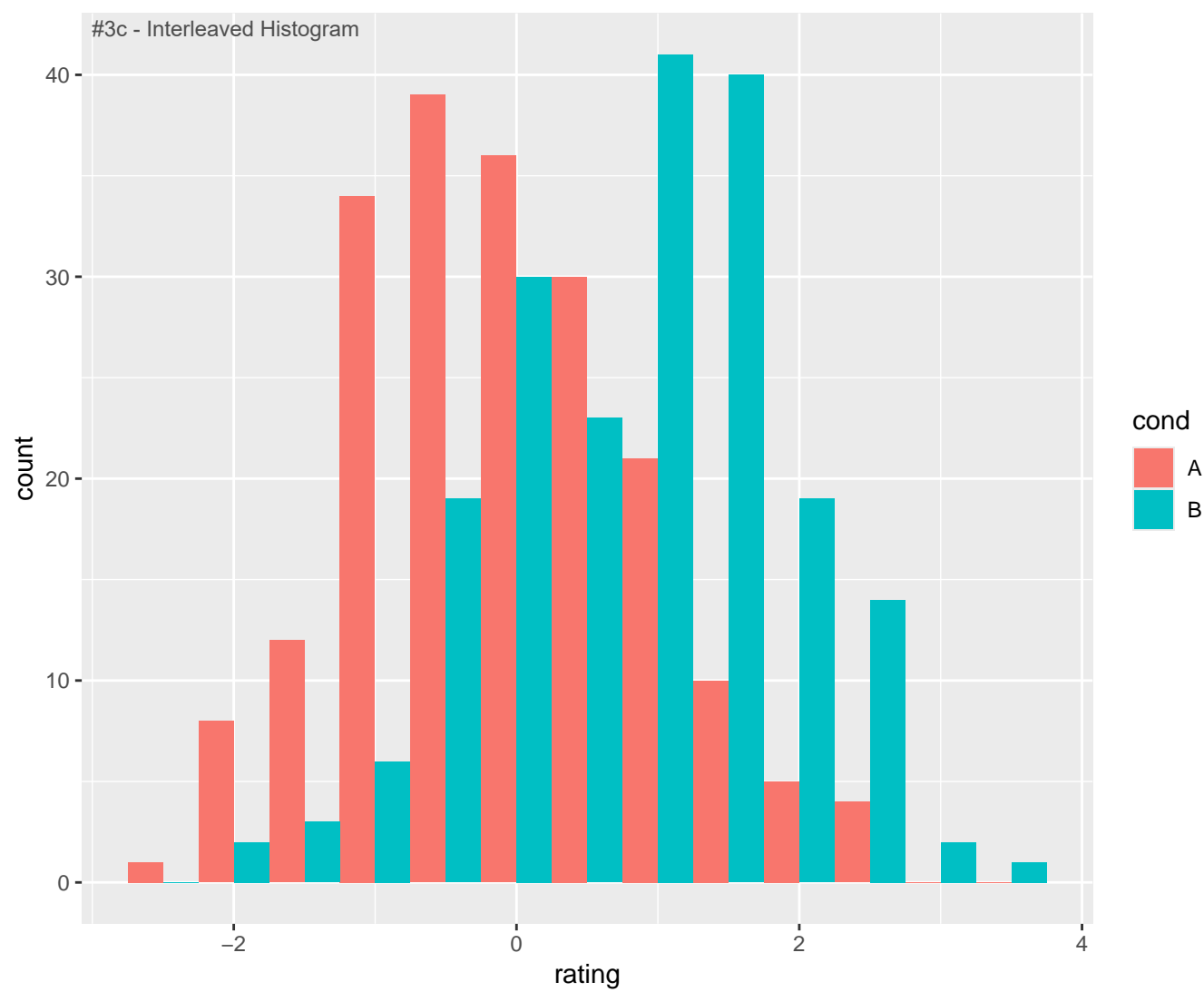
```

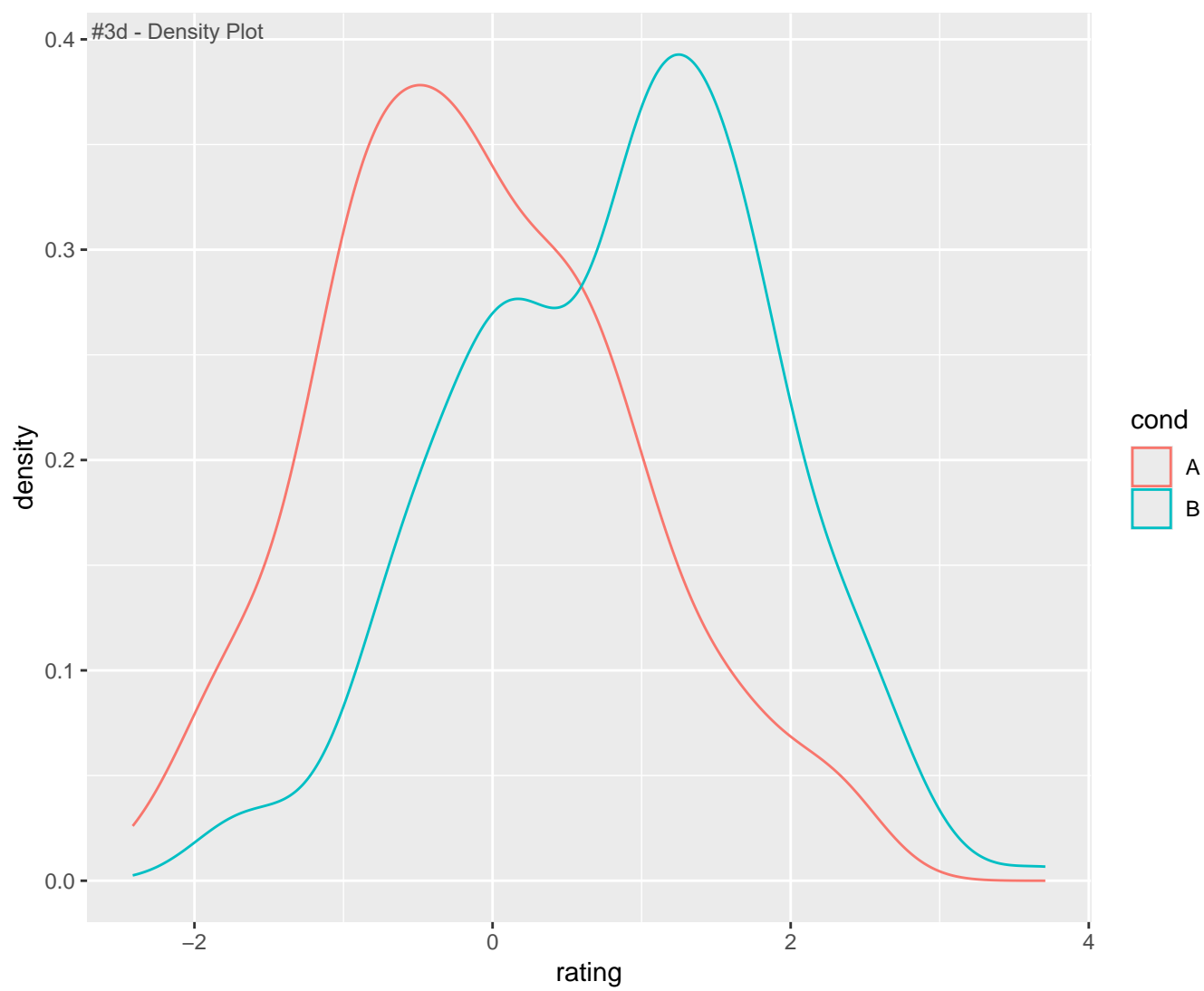

Question #3

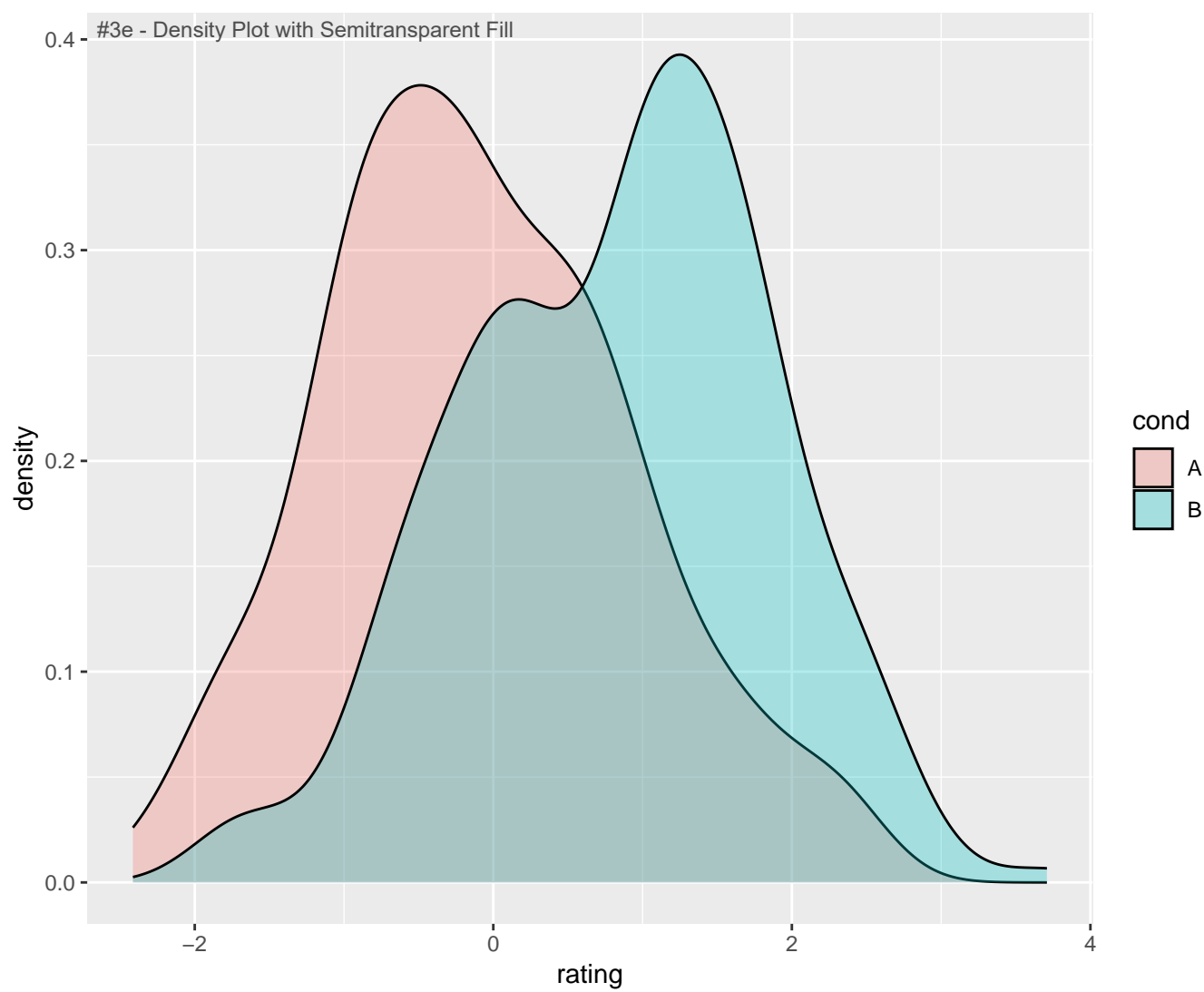
#3b - Overlaid Histogram



#3c - Interleaved Histogram







```
` ``{r}
```

```
library(ggplot2)
```

```
#3.(a) #dat data file
```

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

```
#3.(b) Overlaid histogram - shows both conditions together and where the counts overlap
```

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

```
#3.(c) Interleaved histogram - shows both conditions but now has them separate to see how the counts compare
```

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

```
#3.(d) Density plots - similar to the histogram but it in a smooth curve rather than bars
```

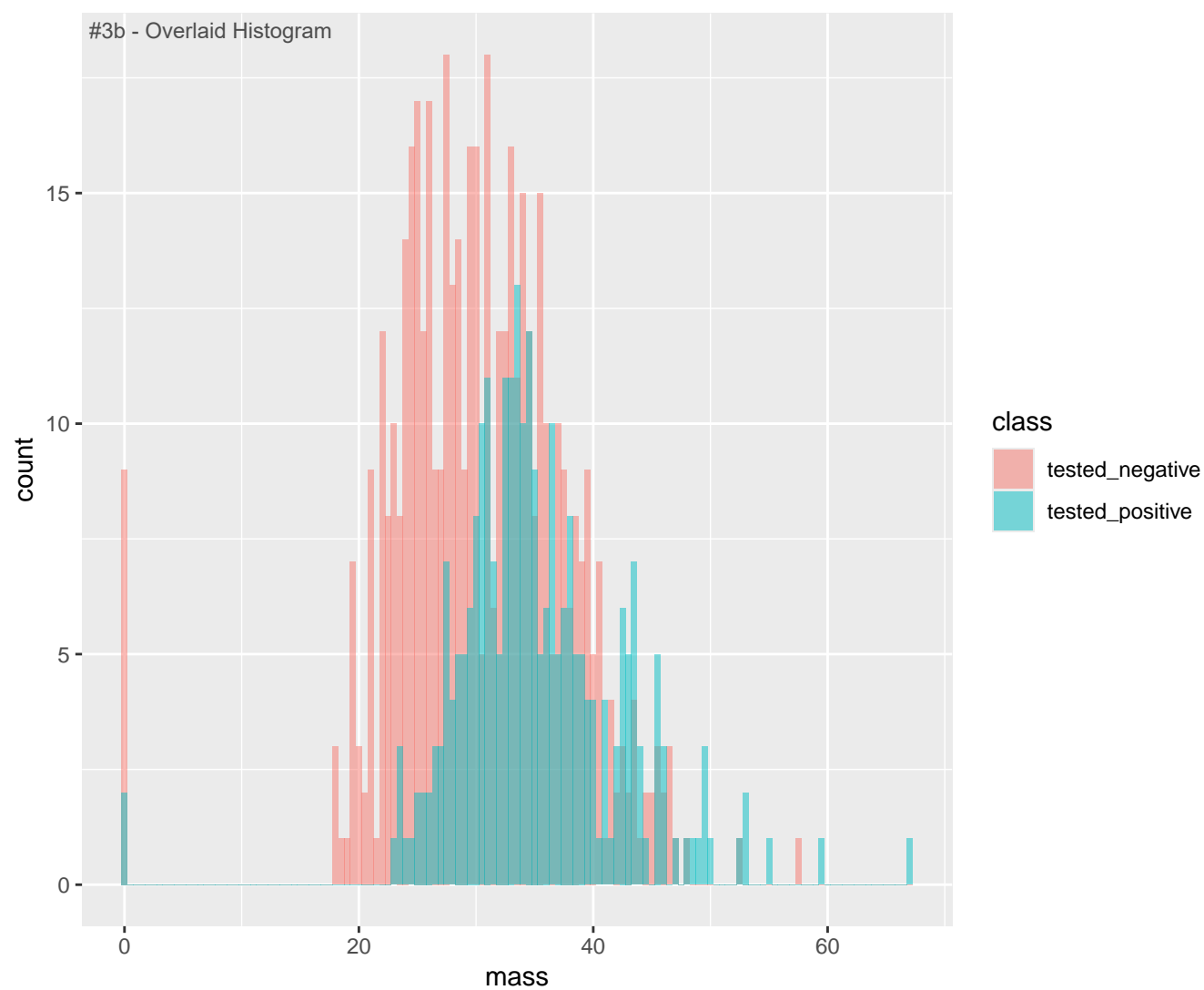
```
ggplot(dat, aes(x = rating, colour = cond)) + geom_density()
```

```
#3.(e) Density plot with semitransparent fill - has the smooth curve as the density plot but now each condition is filled in by color so you can better see where they overlap
```

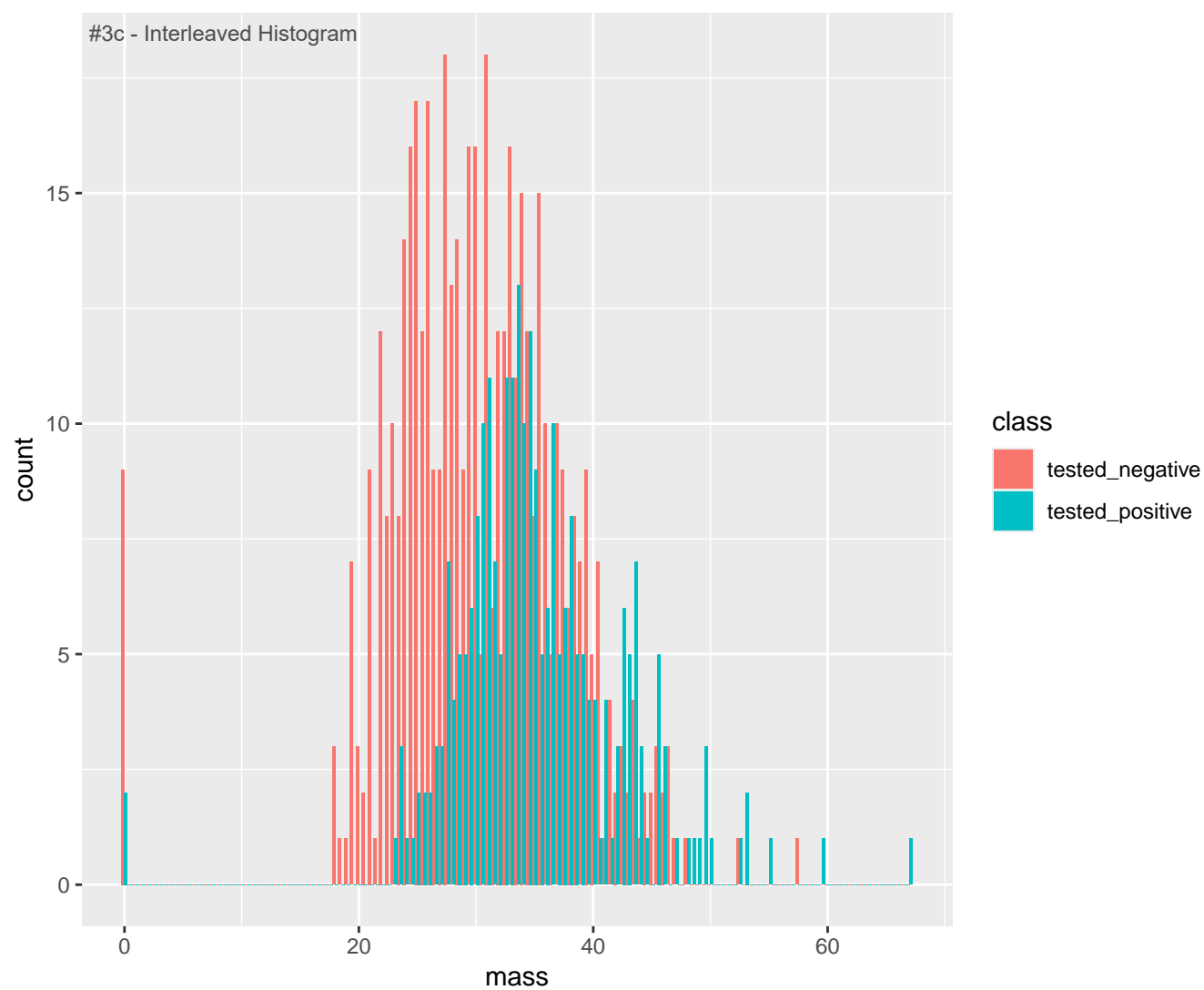
```
ggplot(dat, aes(x = rating, fill = cond)) + geom_density(alpha = .3)
```

```
` ``
```

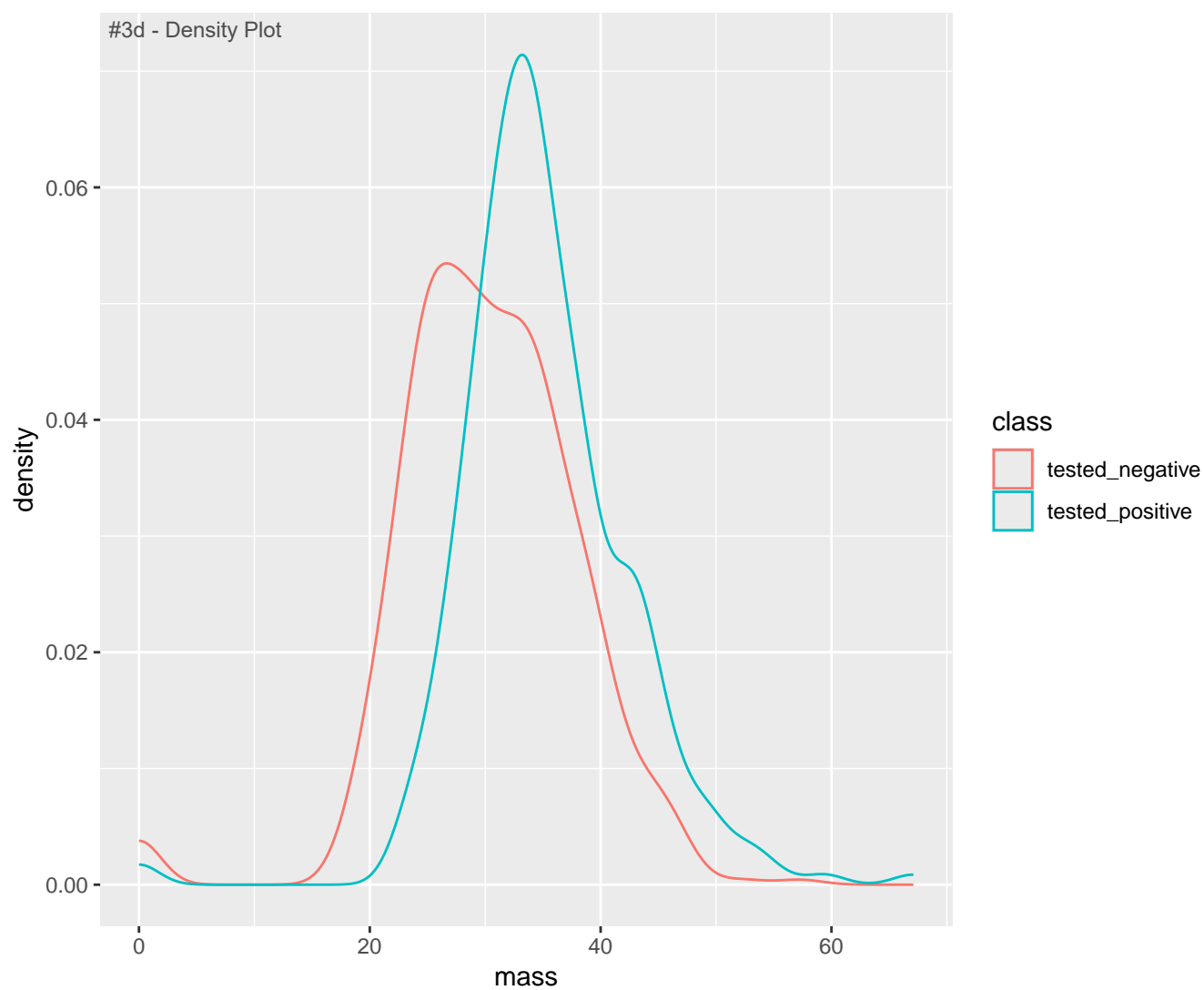
#3b - Overlaid Histogram



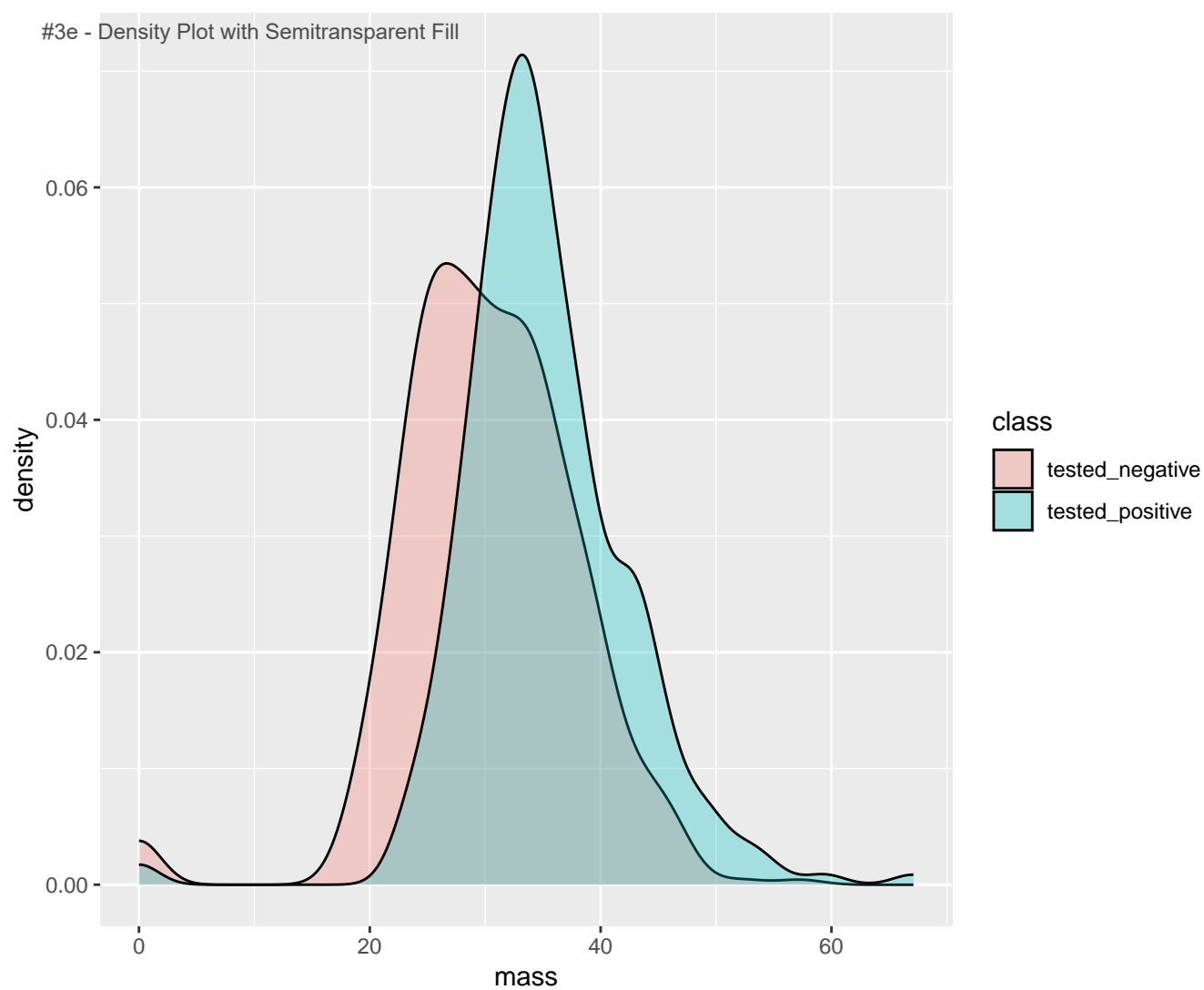
#3c - Interleaved Histogram



#3d - Density Plot



#3e - Density Plot with Semitransparent Fill



```
` ``{r}
```

```
library(ggplot2)
```

```
#3.(a) #diabetes data file
```

```
diabetes <- read.csv("G:/My Drive/Personal/Grad Program/CSC 587 - Adv Data Mining/Scripts/datamining-main/Rscripts/data/diabetes_train.csv")
```

```
#3.(b) Overlaid histogram - shows both classes together and where the counts overlap
```

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

```
#3.(c) Interleaved histogram - shows both classes but now has them separate to see how the counts compare
```

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

```
#3.(d) Density plots - similar to the histogram but it in a smooth curve rather than bars
```

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

```
#3.(e) Density plot with semitransparent fill - has the smooth curve as the density plot but now each class is filled in by color so you can better see where they overlap
```

```
ggplot(diabetes, aes(x = mass, fill = class)) + geom_density(alpha = .3)
```

```
` ``
```

Question #4

```
> library(tidyr)
> #4 titanic data file
> passengers <- read.csv("G:/My Drive/Personal/Grad Program/CSC 587 - Adv Data Mining/Scripts/datamining-main/Rscripts/data/titanic.csv")
> #4.(a)
```

```
> passengers %>% drop_na() %>% summary() #This operation removes any rows that have missing values and then creates
a summary for each column of the newly filtered data
```

X	PassengerId	Survived	Pclass	Name	Sex	Age
Min. : 0.0	Min. : 1.0	Min. :0.0000	Length:714	Length:714	Length:714	Min. : 0.42
1st Qu.:221.2	1st Qu.:222.2	1st Qu.:0.0000	Class :character	Class :character	Class :character	1st Qu.:20.12
Median :444.0	Median :445.0	Median :0.0000	Mode :character	Mode :character	Mode :character	Median :28.00
Mean :447.6	Mean :448.6	Mean :0.4062				Mean :29.70
3rd Qu.:676.8	3rd Qu.:677.8	3rd Qu.:1.0000				3rd Qu.:38.00
Max. :890.0	Max. :891.0	Max. :1.0000				Max. :80.00

SibSp	Parch	Ticket	Fare	Cabin	Embarked
Min. :0.0000	Min. :0.0000	Length:714	Min. : 0.00	Length:714	Length:714
1st Qu.:0.0000	1st Qu.:0.0000	Class :character	1st Qu.: 8.05	Class :character	Class :character
Median :0.0000	Median :0.0000	Mode :character	Median : 15.74	Mode :character	Mode :character
Mean :0.5126	Mean :0.4314		Mean : 34.69		
3rd Qu.:1.0000	3rd Qu.:1.0000		3rd Qu.: 33.38		
Max. :5.0000	Max. :6.0000		Max. :512.33		

```
> #4.(b)
```

```
> passengers %>% filter(Sex == "male") #This operation filters the passengers to show only the male passengers
```

X	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	
1	0	1	0	3	Braund, Mr. Owen Harris	male	22.00	1	0	A/5 21171	7.2500	
2	4	5	0	3	Allen, Mr. William Henry	male	35.00	0	0	373450	8.0500	
3	5	6	0	3	Moran, Mr. James	male	NA	0	0	330877	8.4583	
4	6	7	0	1	McCarthy, Mr. Timothy J	male	54.00	0	0	17463	51.8625	E46
5	7	8	0	3	Palsson, Master. Gosta Leonard	male	2.00	3	1	349909	21.0750	
6	12	13	0	3	Saunderscock, Mr. William Henry	male	20.00	0	0	A/5. 2151	8.0500	
7	13	14	0	3	Andersson, Mr. Anders Johan	male	39.00	1	5	347082	31.2750	
8	16	17	0	3	Rice, Master. Eugene	male	2.00	4	1	382652	29.1250	
9	17	18	1	2	Williams, Mr. Charles Eugene	male	NA	0	0	244373	13.0000	
10	20	21	0	2	Fynney, Mr. Joseph J	male	35.00	0	0	239865	26.0000	
11	21	22	1	2	Beesley, Mr. Lawrence	male	34.00	0	0	248698	13.0000	D56
12	23	24	1	1	Sloper, Mr. William Thompson	male	28.00	0	0	113788	35.5000	A6
13	26	27	0	3	Emir, Mr. Farred Chehab	male	NA	0	0	2631	7.2250	
14	27	28	0	1	Fortune, Mr. Charles Alexander	male	19.00	3	2	19950	263.0000	C23 C25 C27
15	29	30	0	3	Todoroff, Mr. Lalio	male	NA	0	0	349216	7.8958	
16	30	31	0	1	Uruchurtu, Don. Manuel E	male	40.00	0	0	PC 17601	27.7208	
17	33	34	0	2	Wheadon, Mr. Edward H	male	66.00	0	0	C.A. 24579	10.5000	
18	34	35	0	1	Meyer, Mr. Edgar Joseph	male	28.00	1	0	PC 17604	82.1708	
19	35	36	0	1	Holverson, Mr. Alexander Oskar	male	42.00	1	0	113789	52.0000	
20	36	37	1	3	Mamee, Mr. Hanna	male	NA	0	0	2677	7.2292	
21	37	38	0	?	Cann, Mr. Ernest Charles	male	21.00	0	0	A./5. 2152	8.0500	
22	42	43	0	3	Kraeff, Mr. Theodor	male	NA	0	0	349253	7.8958	
23	45	46	0	3	Rogers, Mr. William John	male	NA	0	0	S.C./A.4. 23567	8.0500	
24	46	47	0	3	Lennon, Mr. Denis	male	NA	1	0	370371	15.5000	

25	48	49	0	3	Samaan, Mr. Youssef male	NA	2	0	2662	21.6792	
26	50	51	0	3	Panula, Master. Juha Niilo male	7.00	4	1	3101295	39.6875	
27	51	52	0	3	Nosworthy, Mr. Richard Cater male	21.00	0	0	A/4. 39886	7.8000	
28	54	55	0	1	Ostby, Mr. Engelhart Cornelius male	65.00	0	1	113509	61.9792	B30
29	55	56	1	1	Woolner, Mr. Hugh male	NA	0	0	19947	35.5000	C52
30	57	58	0	3	Novel, Mr. Mansouer male	28.50	0	0	2697	7.2292	
31	59	60	0	3	Goodwin, Master. William Frederick male	11.00	5	2	CA 2144	46.9000	
32	60	61	0	3	Sirayanian, Mr. Orsen male	22.00	0	0	2669	7.2292	
33	62	63	0	1	Harris, Mr. Henry Birkhardt male	45.00	1	0	36973	83.4750	C83
34	63	64	0	3	Skoog, Master. Harald male	4.00	3	2	347088	27.9000	
35	64	65	0	1	Stewart, Mr. Albert A male	NA	0	0	PC 17605	27.7208	
36	65	66	1	3	Moubarek, Master. Gerios male	NA	1	1	2661	15.2458	
37	67	68	0	3	Crease, Mr. Ernest James male	19.00	0	0	S.P. 3464	8.1583	
38	69	70	0	3	Kink, Mr. Vincenz male	26.00	2	0	315151	8.6625	
39	70	71	0	2	Jenkin, Mr. Stephen Curnow male	32.00	0	0	C.A. 33111	10.5000	
40	72	73	0	2	Hood, Mr. Ambrose Jr male	21.00	0	0	S.O.C. 14879	73.5000	
41	73	74	0	3	Chronopoulos, Mr. Apostolos male	26.00	1	0	2680	14.4542	
42	74	75	1	3	Bing, Mr. Lee male	32.00	0	0	1601	56.4958	
43	75	76	0	3	Moen, Mr. Sigurd Hansen male	25.00	0	0	348123	7.6500	F G73
44	76	77	0	3	Staneff, Mr. Ivan male	NA	0	0	349208	7.8958	
45	77	78	0	3	Moutal, Mr. Rahamin Haim male	NA	0	0	374746	8.0500	
46	78	79	1	2	Caldwell, Master. Alden Gates male	0.83	0	2	248738	29.0000	
47	80	81	0	?	Waelens, Mr. Achille male	22.00	0	0	345767	9.0000	
48	81	82	1	3	Sheerlinck, Mr. Jan Baptist male	29.00	0	0	345779	9.5000	
49	83	84	0	1	Carrau, Mr. Francisco M male	28.00	0	0	113059	47.1000	
50	86	87	0	3	Ford, Mr. William Neal male	16.00	1	3	W./C. 6608	34.3750	
51	87	88	0	3	Slocovski, Mr. Selman Francis male	NA	0	0	SOTON/OQ 392086	8.0500	
52	89	90	0	3	Celotti, Mr. Francesco male	24.00	0	0	343275	8.0500	
53	90	91	0	3	Christmann, Mr. Emil male	29.00	0	0	343276	8.0500	
54	91	92	0	3	Andreasson, Mr. Paul Edvin male	20.00	0	0	347466	7.8542	
55	92	93	0	1	Chaffee, Mr. Herbert Fuller male	46.00	1	0	W.E.P. 5734	61.1750	E31
56	93	94	0	3	Dean, Mr. Bertram Frank male	26.00	1	2	C.A. 2315	20.5750	
57	94	95	0	3	Coxon, Mr. Daniel male	59.00	0	0	364500	7.2500	
58	95	96	0	3	Shorney, Mr. Charles Joseph male	NA	0	0	374910	8.0500	
59	96	97	0	1	Goldschmidt, Mr. George B male	71.00	0	0	PC 17754	34.6542	A5
60	97	98	1	1	Greenfield, Mr. William Bertram male	23.00	0	1	PC 17759	63.3583	D10 D12
61	99	100	0	2	Kantor, Mr. Sinai male	34.00	1	0	244367	26.0000	
62	101	102	0	3	Petroff, Mr. Pastcho ("Pentcho") male	NA	0	0	349215	7.8958	
63	102	103	0	1	White, Mr. Richard Frasar male	21.00	0	1	35281	77.2875	D26
64	103	104	0	3	Johansson, Mr. Gustaf Joel male	33.00	0	0	7540	8.6542	
65	104	105	0	3	Gustafsson, Mr. Anders Vilhelm male	37.00	2	0	3101276	7.9250	
66	105	106	0	3	Mionoff, Mr. Stoytcho male	28.00	0	0	349207	7.8958	
67	107	108	1	3	Moss, Mr. Albert Johan male	NA	0	0	312991	7.7750	
68	108	109	0	3	Rekic, Mr. Tido male	38.00	0	0	349249	7.8958	
69	110	111	0	1	Porter, Mr. Walter Chamberlain male	47.00	0	0	110465	52.0000	C110
70	112	113	0	3	Barton, Mr. David John male	22.00	0	0	324669	8.0500	
71	115	116	0	3	Pekoniemi, Mr. Edvard male	21.00	0	0	STON/O 2. 3101294	7.9250	

72	116	117	0	3	Connors, Mr. Patrick male 70.50	0	0	370369	7.7500	
73	117	118	0	2	Turpin, Mr. William John Robert male 29.00	1	0	11668	21.0000	
74	118	119	0	1	Baxter, Mr. Quigg Edmond male 24.00	0	1	PC 17558	247.5208	B58 B60
75	120	121	0	2	Hickman, Mr. Stanley George male 21.00	2	0	S.O.C. 14879	73.5000	
76	121	122	0	3	Moore, Mr. Leonard Charles male NA	0	0	A4. 54510	8.0500	

Embarked

1	S
2	S
3	Q
4	S
5	S
6	S
7	S
8	Q
9	S
10	S
11	S
12	S
13	C
14	S
15	S
16	C
17	S
18	C
19	S
20	C
21	S
22	C
23	S
24	Q
25	C
26	S
27	S
28	C
29	S
30	C
31	S
32	C
33	S
34	S
35	C
36	C
37	S
38	S
39	S
40	S
41	C

```

42      S
43      S
44      S
45      S
46      S
47      S
48      S
49      S
50      S
51      S
52      S
53      S
54      S
55      S
56      S
57      S
58      S
59      C
60      C
61      S
62      S
63      S
64      S
65      S
66      S
67      S
68      S
69      S
70      S
71      S
72      Q
73      S
74      C
75      S
76      S

```

```
[ reached 'max' / getOption("max.print") -- omitted 501 rows ]
```

```
> #4.(c)
```

```
> passengers %>% arrange(desc(Fare)) #This operation arranges the passengers by the Fare amount in descending order (largest to smallest)
```

	X	PassengerId	Survived	Pclass	Name	Sex	Age
1	258	259	1	1	Ward, Miss. Anna	female	35.00
2	679	680	1	1	Cardeza, Mr. Thomas Drake Martinez	male	36.00
3	737	738	1	1	Lesurer, Mr. Gustave J	male	35.00
4	27	28	0	1	Fortune, Mr. Charles Alexander	male	19.00
5	88	89	1	1	Fortune, Miss. Mabel Helen	female	23.00
6	341	342	1	1	Fortune, Miss. Alice Elizabeth	female	24.00
7	438	439	0	1	Fortune, Mr. Mark	male	64.00
8	311	312	1	1	Ryerson, Miss. Emily Borie	female	18.00

9	742	743	1	1	Ryerson, Miss. Susan Parker "Suzette" female 21.00
10	118	119	0	1	Baxter, Mr. Quigg Edmond male 24.00
11	299	300	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput) female 50.00
12	380	381	1	?	Bidois, Miss. Rosalie female 42.00
13	557	558	0	1	Robbins, Mr. Victor male NA
14	700	701	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force) female 18.00
15	716	717	1	1	Endres, Miss. Caroline Louise female 38.00
16	527	528	0	1	Farthing, Mr. John male NA
17	377	378	0	1	Widener, Mr. Harry Elkins male 27.00
18	689	690	1	1	Madill, Miss. Georgette Alexandra female 15.00
19	730	731	1	?	Allen, Miss. Elisabeth Walton female 29.00
20	779	780	1	1	Robert, Mrs. Edward Scott (Elisabeth Walton McMillan) female 43.00
21	318	319	1	1	Wick, Miss. Mary Natalie female 31.00
22	856	857	1	1	Wick, Mrs. George Dennick (Mary Hitchcock) female 45.00
23	268	269	1	1	Graham, Mrs. William Thompson (Edith Junkins) female 58.00
24	332	333	0	1	Graham, Mr. George Edward male 38.00
25	609	610	1	1	Shutes, Miss. Elizabeth W female 40.00
26	297	298	0	1	Allison, Miss. Helen Loraine female 2.00
27	305	306	1	1	Allison, Master. Hudson Trevor male 0.92
28	498	499	0	1	Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25.00
29	708	709	1	1	Cleaver, Miss. Alice female 22.00
30	31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie) female NA
31	195	196	1	1	Lurette, Miss. Elise female 58.00
32	269	270	1	1	Bissette, Miss. Amelia female 35.00
33	325	326	1	1	Young, Miss. Marie Grice female 36.00
34	373	374	0	1	Ringhini, Mr. Sante male 22.00
35	319	320	1	1	Spedden, Mrs. Frederic Oakley (Margaretta Corning Stone) female 40.00
36	337	338	1	1	Burns, Miss. Elizabeth Margaret female 41.00
37	334	335	1	1	Frauenthal, Mrs. Henry William (Clara Heinsheimer) female NA
38	660	661	1	1	Frauenthal, Dr. Henry William male 50.00
39	390	391	1	?	Carter, Mr. William Ernest male 36.00
40	435	436	1	1	Carter, Miss. Lucile Polk female 14.00
41	763	764	1	1	Carter, Mrs. William Ernest (Lucile Polk) female 36.00
42	802	803	1	1	Carter, Master. William Thornton II male 11.00
43	215	216	1	1	Newell, Miss. Madeleine female 31.00
44	393	394	1	1	Newell, Miss. Marjorie female 23.00
45	659	660	0	1	Newell, Mr. Arthur Webster male 58.00
46	306	307	1	1	Fleming, Miss. Margaret female NA
47	550	551	1	1	Thayer, Mr. John Borland Jr male 17.00
48	581	582	1	1	Thayer, Mrs. John Borland (Marian Longstreth Morris) female 39.00
49	698	699	0	1	Thayer, Mr. John Borland male 49.00
50	307	308	1	1	Penasco y Castellana, Mrs. Victor de Satode (Maria Josefa Perez de Soto y Vallejo) female 17.00
51	505	506	0	1	Penasco y Castellana, Mr. Victor de Satode male 18.00
52	537	538	1	1	LeRoy, Miss. Bertha female 30.00
53	544	545	0	1	Douglas, Mr. Walter Donald male 50.00
54	520	521	1	1	Perreault, Miss. Anne female 30.00
55	820	821	1	1	Hays, Mrs. Charles Melville (Clara Jennings Gregg) female 52.00

56	291	292	1	1	Bishop, Mrs. Dickinson H (Helen Walton)	female	19.00
57	484	485	1	1	Bishop, Mr. Dickinson H	male	25.00
58	224	225	1	1	Hoyt, Mr. Frederick Maxfield	male	38.00
59	245	246	0	1	Minahan, Dr. William Edward	male	44.00
60	412	413	1	?	Minahan, Miss. Daisy E	female	33.00
61	486	487	1	1	Hoyt, Mrs. Frederick Maxfield (Jane Anne Forby)	female	35.00
62	453	454	1	1	Goldenberg, Mr. Samuel L	male	49.00
63	849	850	1	1	Goldenberg, Mrs. Samuel L (Edwiga Grabowska)	female	NA
64	257	258	1	1	Cherry, Miss. Gladys	female	30.00
65	504	505	1	1	Maioni, Miss. Roberta	female	16.00
66	759	760	1	?	Roths, the Countess. of (Lucy Noel Martha Dyer-Edwards)	female	33.00
67	62	63	0	1	Harris, Mr. Henry Birkhardt	male	45.00
68	230	231	1	1	Harris, Mrs. Henry Birkhardt (Irene Wallach)	female	35.00
69	310	311	1	1	Hays, Miss. Margaret Bechstein	female	24.00
70	835	836	1	1	Compton, Miss. Sara Rebecca	female	39.00
71	879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.00
72	34	35	0	1	Meyer, Mr. Edgar Joseph	male	28.00
73	375	376	1	1	Meyer, Mrs. Edgar Joseph (Leila Saks)	female	NA
74	445	446	1	1	Dodge, Master. Washington	male	4.00
75	61	62	1	?	Icard, Miss. Amelie	female	38.00
76	829	830	1	1	Stone, Mrs. George Nelson (Martha Evelyn)	female	62.00

	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	0	PC 17755	512.3292		C
2	0	1	PC 17755	512.3292	B51 B53 B55	C
3	0	0	PC 17755	512.3292	B101	C
4	3	2	19950	263.0000	C23 C25 C27	S
5	3	2	19950	263.0000	C23 C25 C27	S
6	3	2	19950	263.0000	C23 C25 C27	S
7	1	4	19950	263.0000	C23 C25 C27	S
8	2	2	PC 17608	262.3750	B57 B59 B63 B66	C
9	2	2	PC 17608	262.3750	B57 B59 B63 B66	C
10	0	1	PC 17558	247.5208	B58 B60	C
11	0	1	PC 17558	247.5208	B58 B60	C
12	0	0	PC 17757	227.5250		C
13	0	0	PC 17757	227.5250		C
14	1	0	PC 17757	227.5250	C62 C64	C
15	0	0	PC 17757	227.5250	C45	C
16	0	0	PC 17483	221.7792	C95	S
17	0	2	113503	211.5000	C82	C
18	0	1	24160	211.3375	B5	S
19	0	0	24160	211.3375	B5	S
20	0	1	24160	211.3375	B3	S
21	0	2	36928	164.8667	C7	S
22	1	1	36928	164.8667		S
23	0	1	PC 17582	153.4625	C125	S
24	0	1	PC 17582	153.4625	C91	S
25	0	0	PC 17582	153.4625	C125	S

26	1	2	113781	151.5500	C22 C26	S
27	1	2	113781	151.5500	C22 C26	S
28	1	2	113781	151.5500	C22 C26	S
29	0	0	113781	151.5500		S
30	1	0	PC 17569	146.5208	B78	C
31	0	0	PC 17569	146.5208	B80	C
32	0	0	PC 17760	135.6333	C99	S
33	0	0	PC 17760	135.6333	C32	C
34	0	0	PC 17760	135.6333		C
35	1	1	16966	134.5000	E34	C
36	0	0	16966	134.5000	E40	C
37	1	0	PC 17611	133.6500		S
38	2	0	PC 17611	133.6500		S
39	1	2	113760	120.0000	B96 B98	S
40	1	2	113760	120.0000	B96 B98	S
41	1	2	113760	120.0000	B96 B98	S
42	1	2	113760	120.0000	B96 B98	S
43	1	0	35273	113.2750	D36	C
44	1	0	35273	113.2750	D36	C
45	0	2	35273	113.2750	D48	C
46	0	0	17421	110.8833		C
47	0	2	17421	110.8833	C70	C
48	1	1	17421	110.8833	C68	C
49	1	1	17421	110.8833	C68	C
50	1	0	PC 17758	108.9000	C65	C
51	1	0	PC 17758	108.9000	C65	C
52	0	0	PC 17761	106.4250		C
53	1	0	PC 17761	106.4250	C86	C
54	0	0	12749	93.5000	B73	S
55	1	1	12749	93.5000	B69	S
56	1	0	11967	91.0792	B49	C
57	1	0	11967	91.0792	B49	C
58	1	0	19943	90.0000	C93	S
59	2	0	19928	90.0000	C78	Q
60	1	0	19928	90.0000	C78	Q
61	1	0	19943	90.0000	C93	S
62	1	0	17453	89.1042	C92	C
63	1	0	17453	89.1042	C92	C
64	0	0	110152	86.5000	B77	S
65	0	0	110152	86.5000	B79	S
66	0	0	110152	86.5000	B77	S
67	1	0	36973	83.4750	C83	S
68	1	0	36973	83.4750	C83	S
69	0	0	11767	83.1583	C54	C
70	1	1	PC 17756	83.1583	E49	C
71	0	1	11767	83.1583	C50	C
72	1	0	PC 17604	82.1708		C

```

73 1 0 PC 17604 82.1708 C
74 0 2 33638 81.8583 A34 S
75 0 0 113572 80.0000 B28
76 0 0 113572 80.0000 B28

```

```
[ reached 'max' / getOption("max.print") -- omitted 815 rows ]
```

```
> #4.(d)
```

```
> passengers %>% mutate(FamSize = Parch + SibSp) #This operation creates a new column called FamSize, which is calculated by adding the Parch and SibSp columns
```

	X	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
1	0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
2	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599
3	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
4	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803
5	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450
6	5	6	0	3	Moran, Mr. James	male	NA	0	0	330877
7	6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463
8	7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909
9	8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742
10	9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736
11	10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549
12	11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783
13	12	13	0	3	Saunderscock, Mr. William Henry	male	20.0	0	0	A/5. 2151
14	13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082
15	14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406
16	15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706
17	16	17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652
18	17	18	1	2	Williams, Mr. Charles Eugene	male	NA	0	0	244373
19	18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31.0	1	0	345763
20	19	20	1	3	Masselmani, Mrs. Fatima	female	NA	0	0	2649
21	20	21	0	2	Fynney, Mr. Joseph J	male	35.0	0	0	239865
22	21	22	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698
23	22	23	1	3	McGowan, Miss. Anna "Annie"	female	15.0	0	0	330923
24	23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788
25	24	25	0	3	Palsson, Miss. Torborg Danira	female	8.0	3	1	349909
26	25	26	1	?	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)	female	38.0	1	5	347077
27	26	27	0	3	Emir, Mr. Farred Chehab	male	NA	0	0	2631
28	27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950
29	28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NA	0	0	330959
30	29	30	0	3	Todoroff, Mr. Lalio	male	NA	0	0	349216
31	30	31	0	1	Uruchurtu, Don. Manuel E	male	40.0	0	0	PC 17601
32	31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	NA	1	0	PC 17569
33	32	33	1	3	Glynn, Miss. Mary Agatha	female	NA	0	0	335677
34	33	34	0	2	Wheadon, Mr. Edward H	male	66.0	0	0	C.A. 24579
35	34	35	0	1	Meyer, Mr. Edgar Joseph	male	28.0	1	0	PC 17604
36	35	36	0	1	Holverson, Mr. Alexander Oskar	male	42.0	1	0	113789
37	36	37	1	3	Mamee, Mr. Hanna	male	NA	0	0	2677
38	37	38	0	?	Cann, Mr. Ernest Charles	male	21.0	0	0	A./5. 2152
39	38	39	0	3	Vander Planke, Miss. Augusta Maria	female	18.0	2	0	345764

40	39	40	1	3	Nicola-Yarred, Miss. Jamila	female	14.0	1	0	2651
41	40	41	0	3	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40.0	1	0	7546
42	41	42	0	2	Turpin, Mrs. William John Robert (Dorothy Ann Wonnacott)	female	27.0	1	0	11668
43	42	43	0	3	Kraeff, Mr. Theodor	male	NA	0	0	349253
44	43	44	1	2	Laroche, Miss. Simonne Marie Anne Andree	female	3.0	1	2	SC/Paris 2123
45	44	45	1	3	Devaney, Miss. Margaret Delia	female	19.0	0	0	330958
46	45	46	0	3	Rogers, Mr. William John	male	NA	0	0	S.C./A.4. 23567
47	46	47	0	3	Lennon, Mr. Denis	male	NA	1	0	370371
48	47	48	1	3	O'Driscoll, Miss. Bridget	female	NA	0	0	14311
49	48	49	0	3	Samaan, Mr. Youssef	male	NA	2	0	2662
50	49	50	0	3	Arnold-Franchi, Mrs. Josef (Josefine Franchi)	female	18.0	1	0	349237
51	50	51	0	3	Panula, Master. Juha Niilo	male	7.0	4	1	3101295
52	51	52	0	3	Nosworthy, Mr. Richard Cater	male	21.0	0	0	A/4. 39886
53	52	53	1	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49.0	1	0	PC 17572
54	53	54	1	2	Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkinson)	female	29.0	1	0	2926
55	54	55	0	1	Ostby, Mr. Engelhart Cornelius	male	65.0	0	1	113509
56	55	56	1	1	Woolner, Mr. Hugh	male	NA	0	0	19947
57	56	57	1	2	Rugg, Miss. Emily	female	21.0	0	0	C.A. 31026
58	57	58	0	3	Novel, Mr. Mansouer	male	28.5	0	0	2697
59	58	59	1	2	West, Miss. Constance Mirium	female	5.0	1	2	C.A. 34651
60	59	60	0	3	Goodwin, Master. William Frederick	male	11.0	5	2	CA 2144
61	60	61	0	3	Sirayanian, Mr. Orsen	male	22.0	0	0	2669
62	61	62	1	?	Icard, Miss. Amelie	female	38.0	0	0	113572
63	62	63	0	1	Harris, Mr. Henry Birkhardt	male	45.0	1	0	36973
64	63	64	0	3	Skoog, Master. Harald	male	4.0	3	2	347088
65	64	65	0	1	Stewart, Mr. Albert A	male	NA	0	0	PC 17605
66	65	66	1	3	Moubarek, Master. Gerios	male	NA	1	1	2661
67	66	67	1	2	Nye, Mrs. (Elizabeth Ramell)	female	29.0	0	0	C.A. 29395
68	67	68	0	3	Crease, Mr. Ernest James	male	19.0	0	0	S.P. 3464
69	68	69	1	3	Andersson, Miss. Erna Alexandra	female	17.0	4	2	3101281
70	69	70	0	3	Kink, Mr. Vincenz	male	26.0	2	0	315151
71	70	71	0	2	Jenkin, Mr. Stephen Curnow	male	32.0	0	0	C.A. 33111

	Fare	Cabin	Embarked	FamSize
1	7.2500		S	1
2	71.2833	C85	C	1
3	7.9250		S	0
4	53.1000	C123	S	1
5	8.0500		S	0
6	8.4583		Q	0
7	51.8625	E46	S	0
8	21.0750		S	4
9	11.1333		S	2
10	30.0708		C	1
11	16.7000	G6	S	2
12	26.5500	C103	S	0
13	8.0500		S	0
14	31.2750		S	6

15	7.8542		S	0
16	16.0000		S	0
17	29.1250		Q	5
18	13.0000		S	0
19	18.0000		S	1
20	7.2250		C	0
21	26.0000		S	0
22	13.0000	D56	S	0
23	8.0292		Q	0
24	35.5000	A6	S	0
25	21.0750		S	4
26	31.3875		S	6
27	7.2250		C	0
28	263.0000	C23 C25 C27	S	5
29	7.8792		Q	0
30	7.8958		S	0
31	27.7208		C	0
32	146.5208	B78	C	1
33	7.7500		Q	0
34	10.5000		S	0
35	82.1708		C	1
36	52.0000		S	1
37	7.2292		C	0
38	8.0500		S	0
39	18.0000		S	2
40	11.2417		C	1
41	9.4750		S	1
42	21.0000		S	1
43	7.8958		C	0
44	41.5792		C	3
45	7.8792		Q	0
46	8.0500		S	0
47	15.5000		Q	1
48	7.7500		Q	0
49	21.6792		C	2
50	17.8000		S	1
51	39.6875		S	5
52	7.8000		S	0
53	76.7292	D33	C	1
54	26.0000		S	1
55	61.9792	B30	C	1
56	35.5000	C52	S	0
57	10.5000		S	0
58	7.2292		C	0
59	27.7500		S	3
60	46.9000		S	7
61	7.2292		C	0

```

62 80.0000      B28      0
63 83.4750      C83      S      1
64 27.9000      S      5
65 27.7208      C      0
66 15.2458      C      2
67 10.5000      F33      S      0
68  8.1583      S      0
69  7.9250      S      6
70  8.6625      S      2
71 10.5000      S      0
[ reached 'max' / getOption("max.print") -- omitted 820 rows ]
> #4.(e)
> passengers %>% group_by(Sex) %>% summarise(meanFare = mean(Fare), numSurv = sum(Survived)) #This operation groups the data by Sex (female and male),
    then creates a mean or average of the Fare amount for each Sex and the number that survived for each Sex
# A tibble: 2 × 3
  Sex    meanFare numSurv
<chr>    <dbl>    <int>
1 female    44.5     233
2 male     25.5     109

>

```

Question #5


```
> #5
> quantile(diabetes$skin, probs = c(0.1, 0.3, 0.5, 0.6)) #Calculates the 10th, 30th, 50th, and 60th percentiles of the skin attribute
10% 30% 50% 60%
  0  10  23  27

>
```