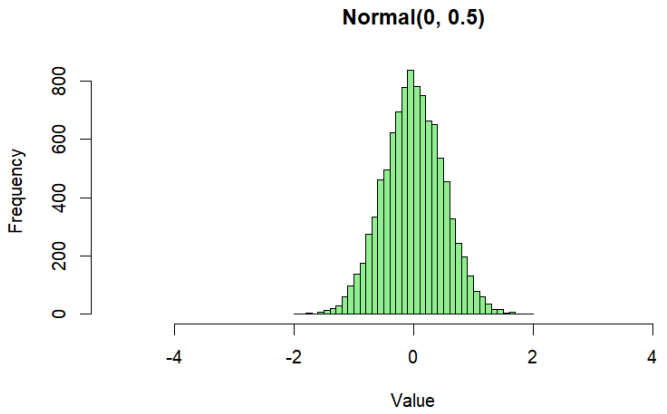
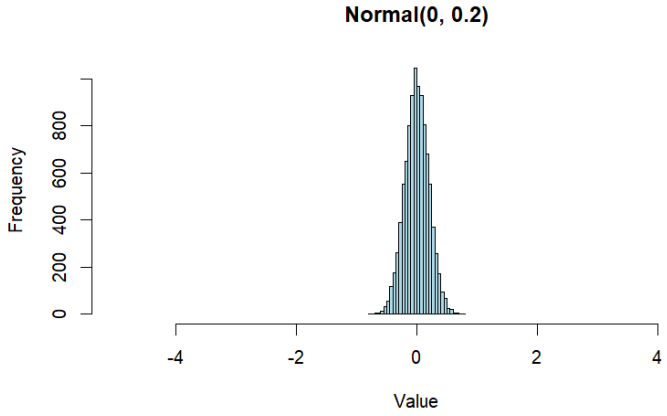
HOMEWORK 1 BY: Abdulhakim Abayu

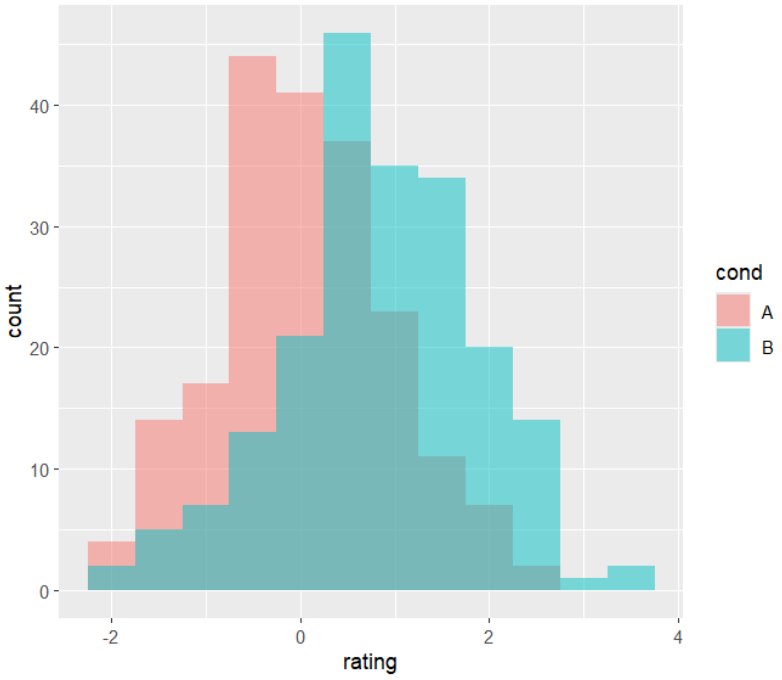
Q2.A&B



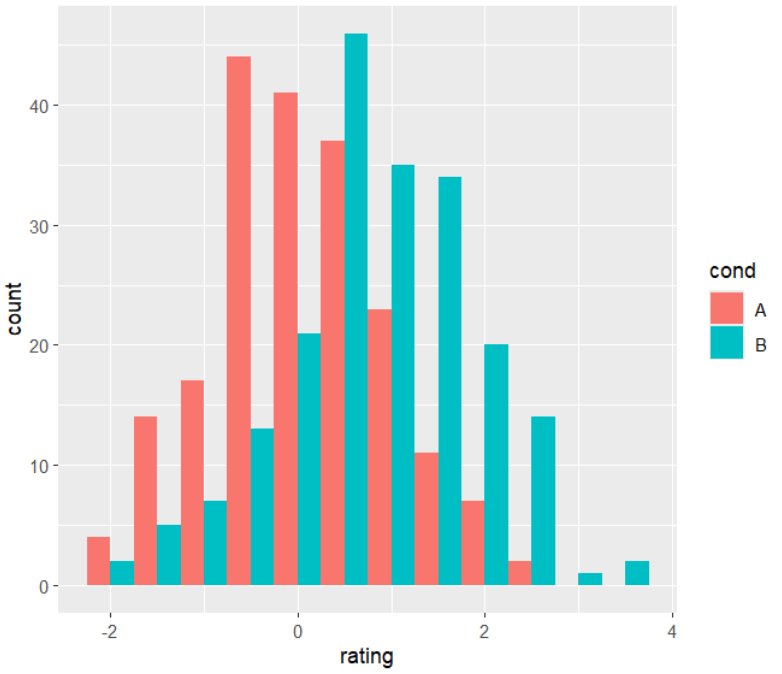


In the histogram with the normal (0,0.2) it is very tell and narrow and values are tightly cluster around 0 and this is because of the standard deviation is 0.2 and the mean 0, which mean the data has very low variability, on the other hand, the histogram with normal (0,0.5) is wider and shorter compared to the 0.2 chart, however values are still centered at 0 since the mean is still the same, but they are spread out across the x-axis and this is because the ST is larger.

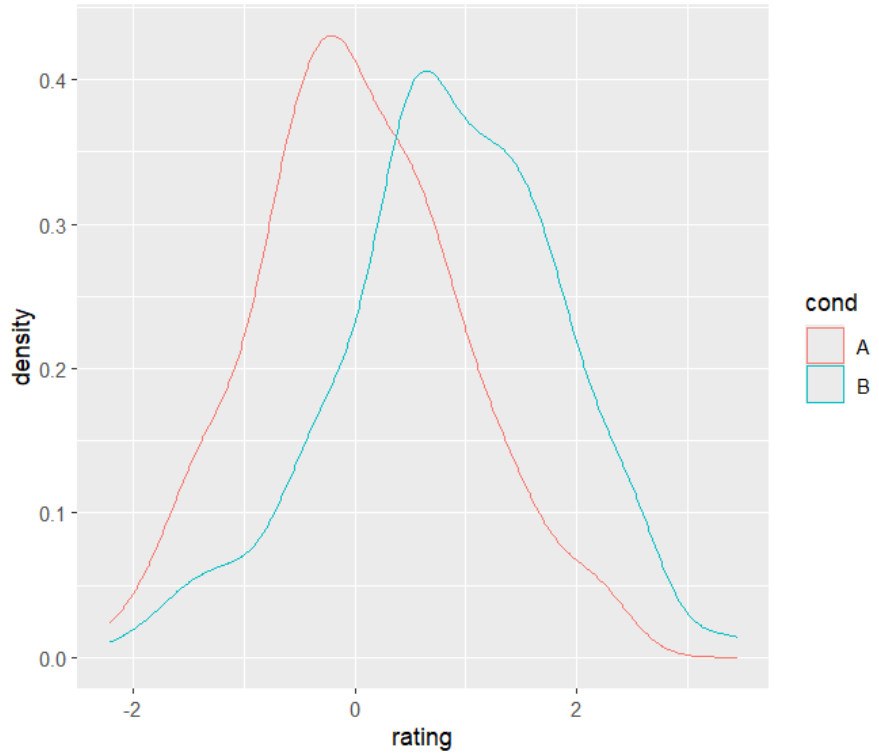
Q3.B


count
40
30
20
10
0
-2
0
2
4
rating
cond
A
B

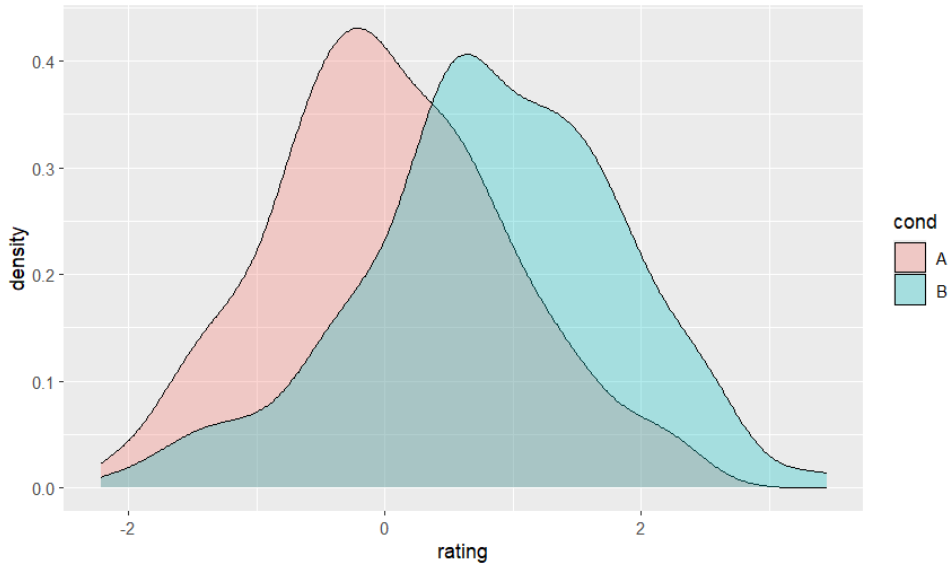
Q3.C


count
40
30
20
10
0
-2
0
2
4
rating
cond
A
B

Q3.D

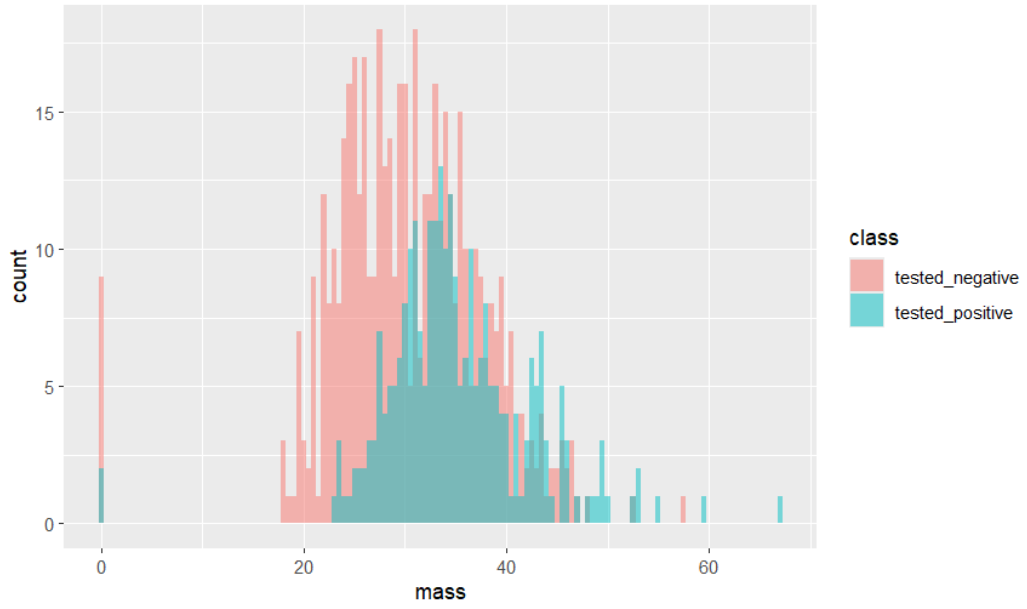

density
0.4
0.3
0.2
0.1
0.0
-2
0
2
rating
cond
A
B

Q3.E

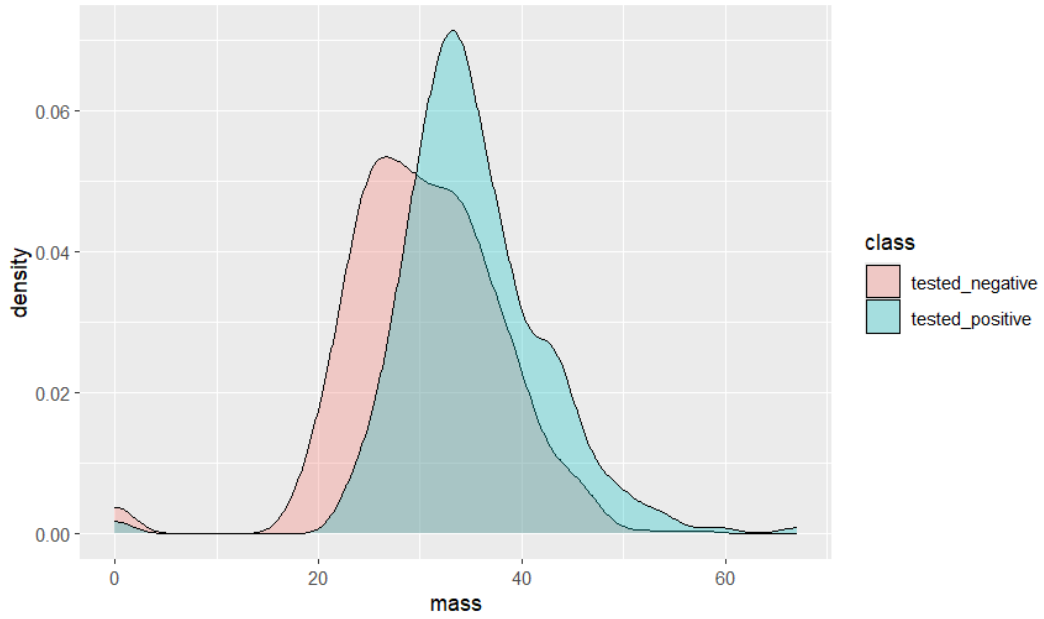

density
0.4
0.3
0.2
0.1
0.0
-2
0
2
rating
cond
A
B

Q3.F

overlaid histogram



Density plots with semitransparent fill



Q4.A

```
passengers %>% drop_na() %>% summary()
```

In this function, we are removing all rows with missing values from any column, and then we are showing a summary of statistics like min, max, mean, and median

Q4.B

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

In this function, we are selecting only the row where the passenger is male

Q4.C

```
passengers %>% arrange(desc(Fare))
```

In this function, we are sorting rows by a column with the highest fare first

Q4.D

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

In this function, we are adding a new variable using the mutate

Q4.E

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

In this function, we are splitting the data into males and females using the groupby (sex) and then getting summaries for the average fare per group and the total survivors in each group