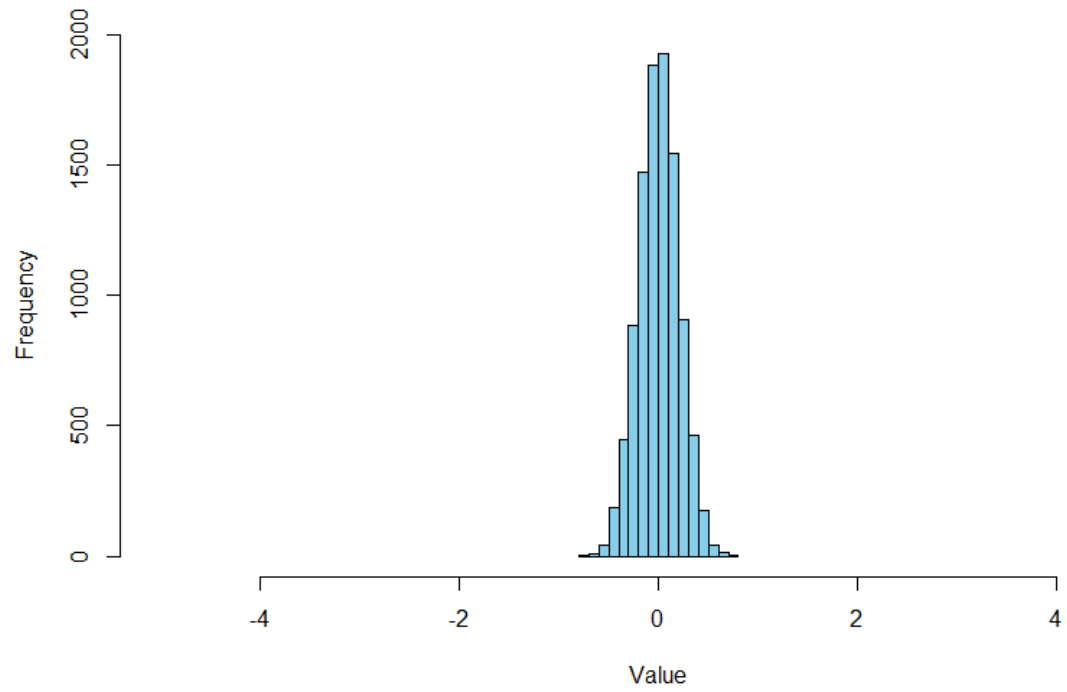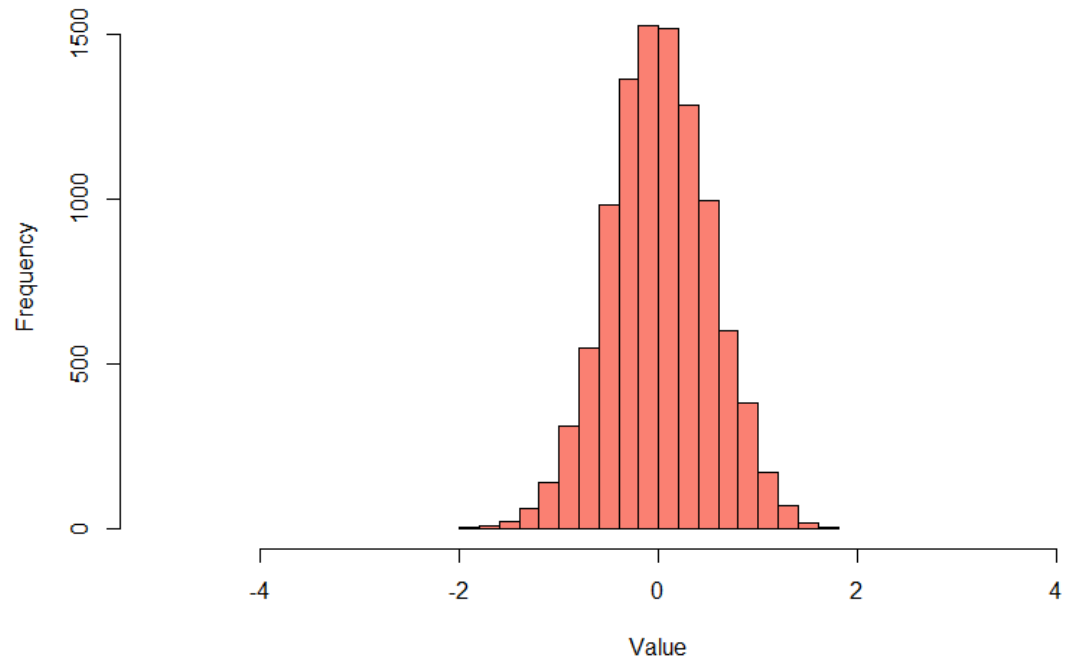Histogram where mean = 0, and sigma = .2)

Histogram where mean = 0, and sigma = .5)

We can see that most of the data stays centered around 0. However, when the standard deviation increased to 0.5, the spread of the data became much wider. The tails of the distribution now extend close to -2 and 2, whereas with a smaller standard deviation, the data was more tightly clustered around 0. This shows that increasing the standard deviation increases the variability of the data while keeping the mean the same.
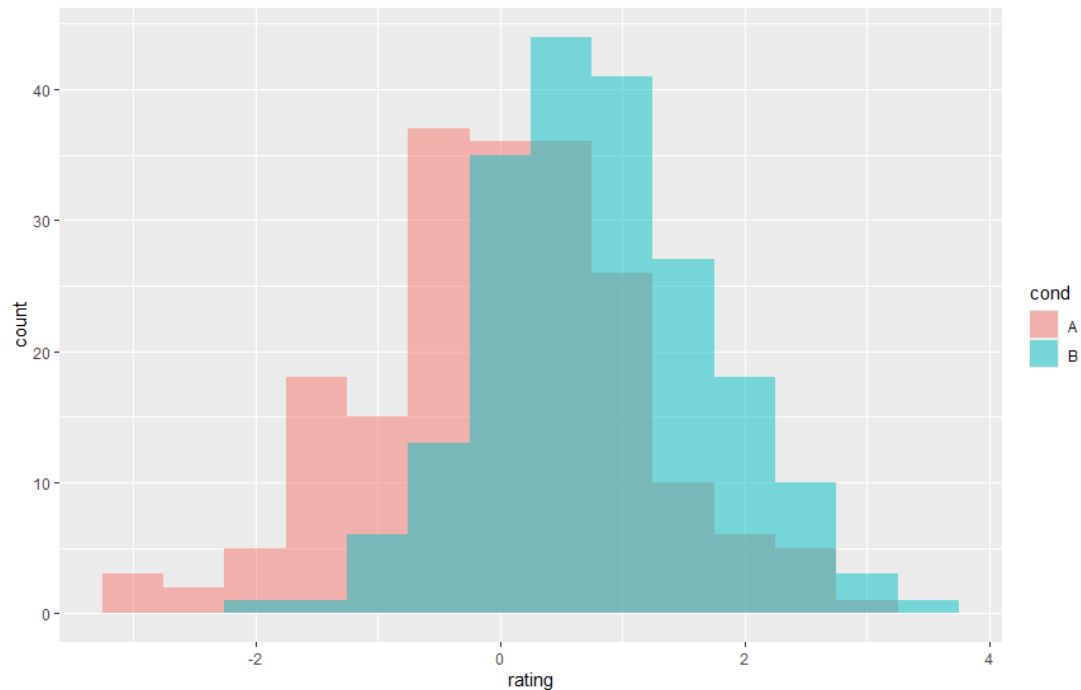
Shows the hisotgrams for both condtion A and B overlapping eachother on the same plot. This way, we can easily compare the similarities and differences between both the condition A and B
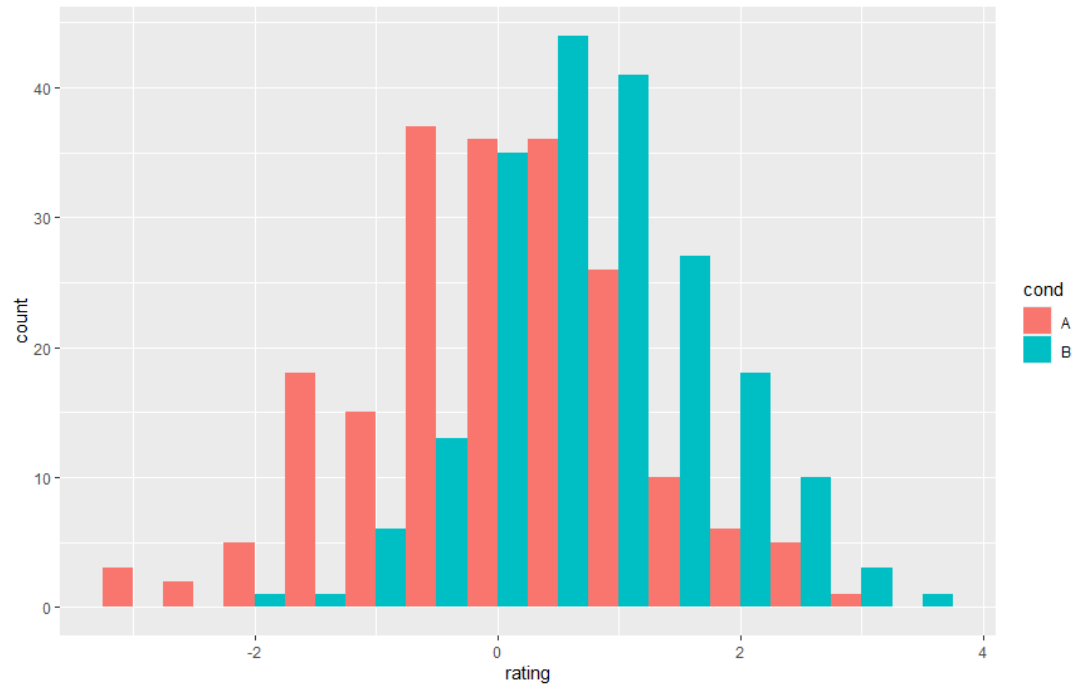
Shows the histograms in a side by side manner instead of them overlapping. This made it easier for me to compare the frequencies however, I preferred the overlapping method more because it was easier for me to visualize quickly.
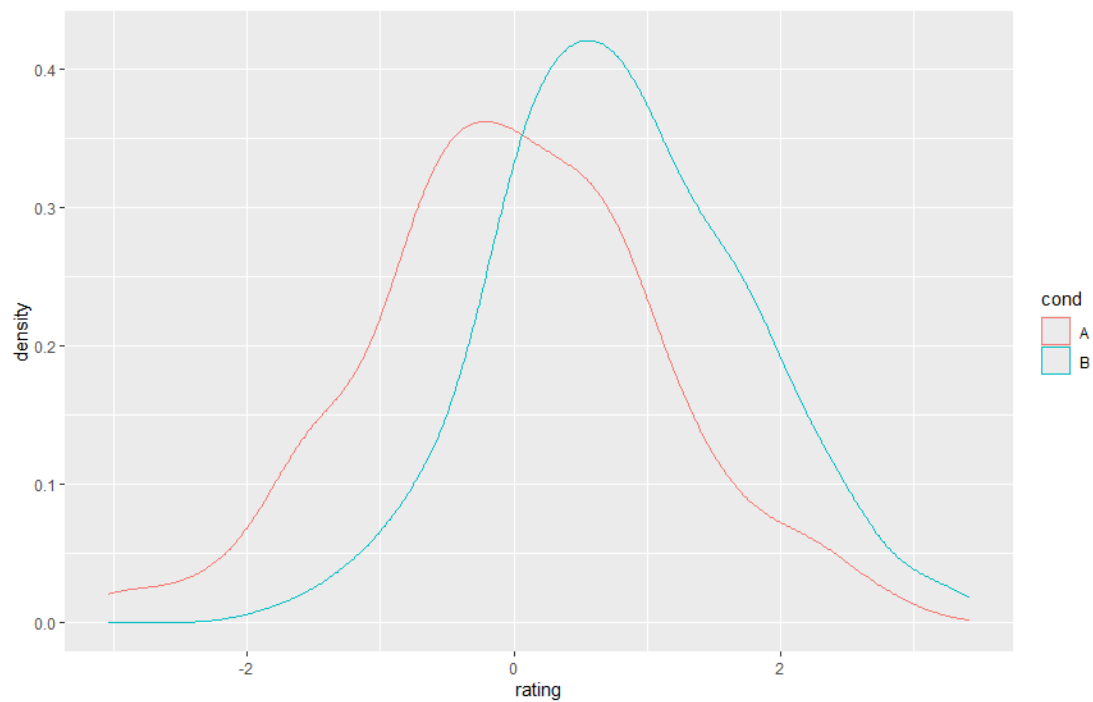
This density plot shows a smooth curve to represent the overall density between the two groups. The smooth curve made it really easy to identify where the majority of the distribution laid.
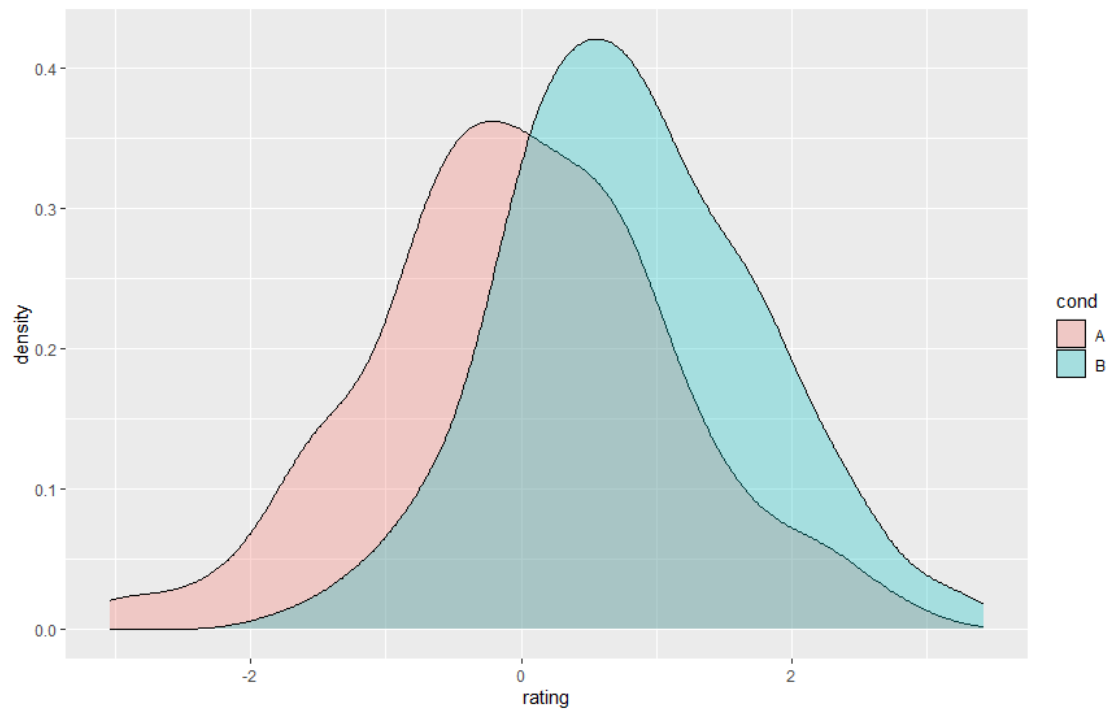
Very similar to the previous plot however, now we see the transparency between the two density plots. This allowed me to more easily see where the overlap is and where Condition B peaks at a higher rating than condition B
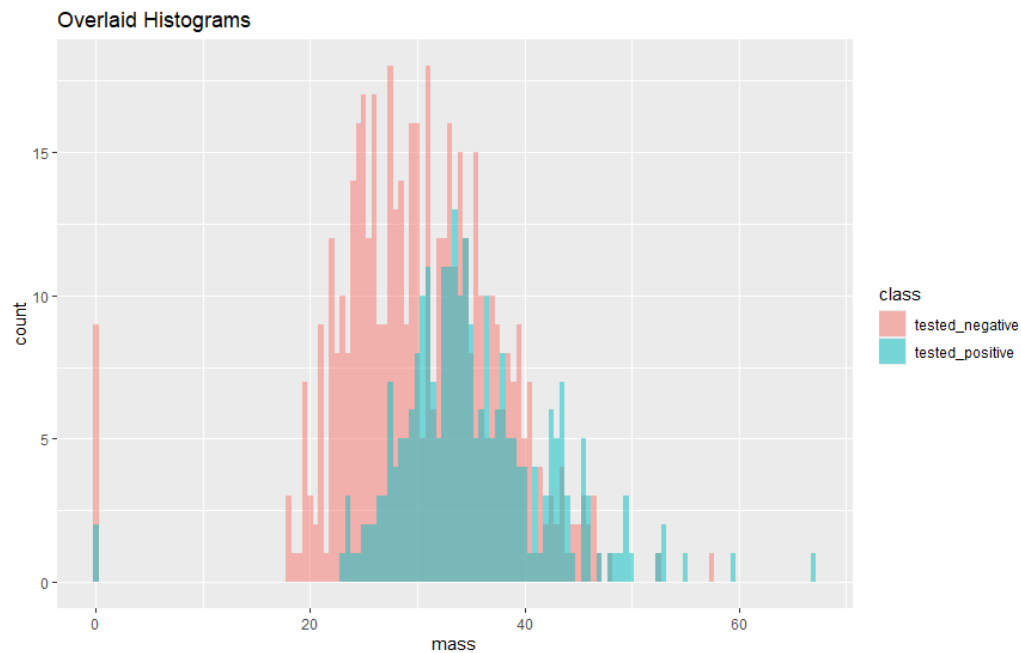
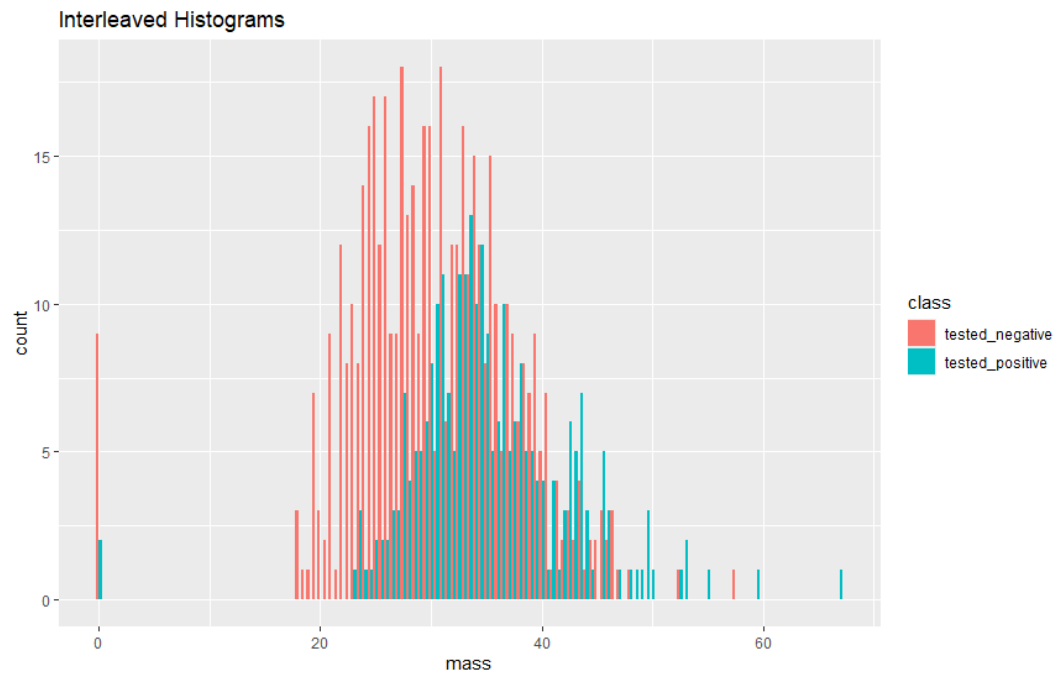This is an overlaid histogram but this time for the diabetes data. Here we once again see the two histograms overlapping which makes it easy to see the similarities and differences between the people who tested negative and positive.

Here we have the same thing as the previous trial however, data is much more intricate and varies quite a bit. I still preferred the overlapping method more because it was easier for me to personally identify similarities and differences.
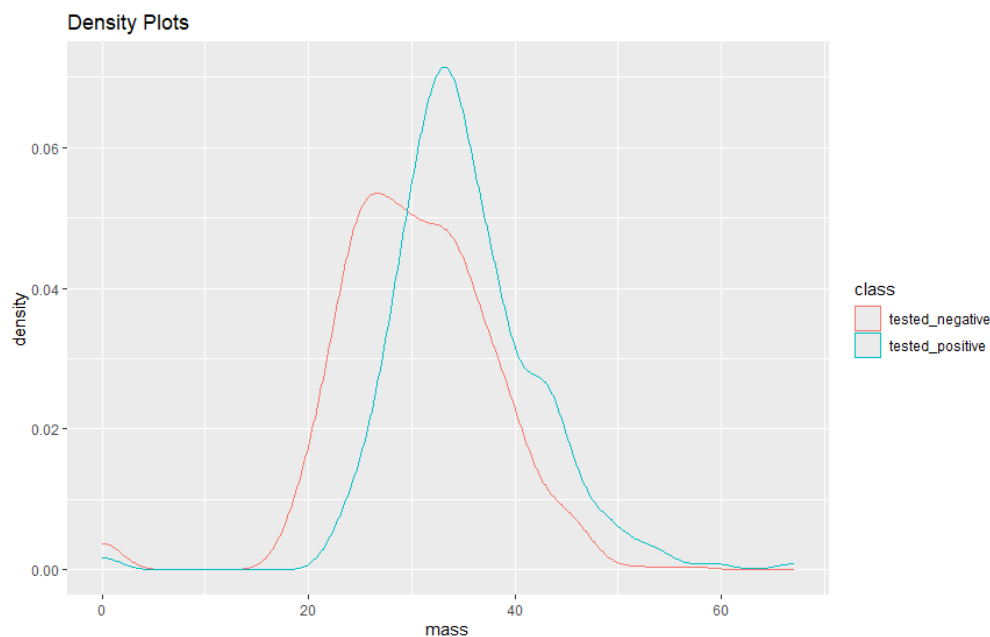
This is a density plot for the diabetes data that demonstrates the people who tested negative and positive. We can quickly identify that right around the 35 mass mark held the most positive tests.
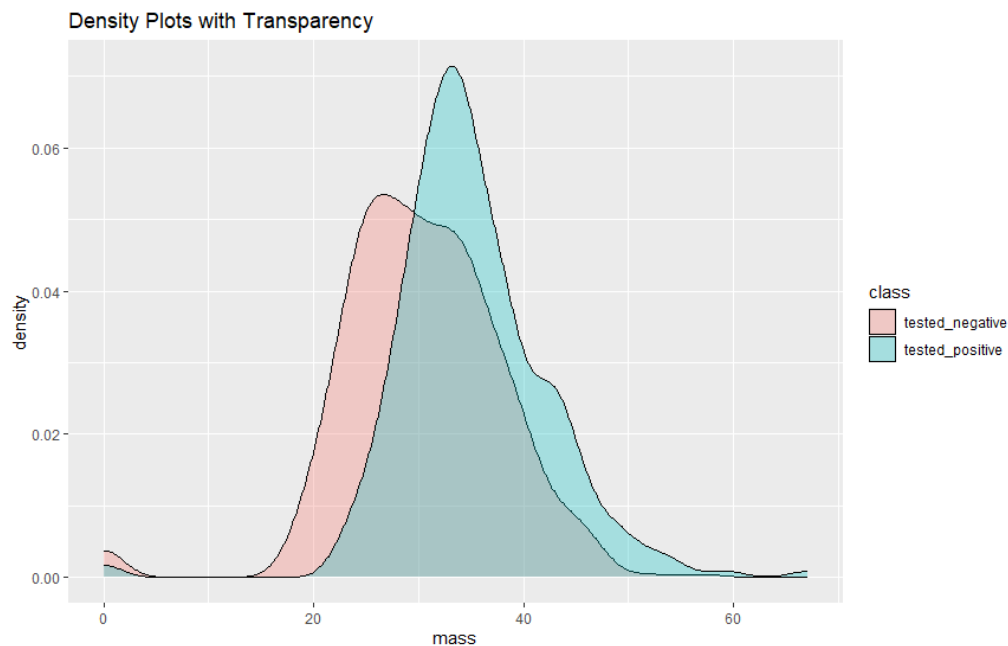
This is the overlapping density plot that more easily shows where the overlap is. Im starting to realize that I really like the overlapping histograms and density plots because its quick for me to compare similarities and differences. I am wondering if there is a way to make the overlapping point a different color so that way its not just both of them muted out.

4.

A) This step is passengers %>% drop_na() %>% summary()

It filters out the missing data then summarizes what's cleaned

```
       X            PassengerId       Survived         Pclass            Name              Sex
 Min.   :  0.0    Min.   :  1.0    Min.   :0.0000   Length:714       Length:714        Length:714
 1st Qu.:221.2    1st Qu.:222.2    1st Qu.:0.0000   Class :character  Class :character  Class :character
 Median :444.0    Median :445.0    Median :0.0000   Mode  :character  Mode  :character  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
```

B) This step is passengers %>% filter(Sex == "male")

it filters the rows so it's just data on men will say. The output is pretty confusing; I'm not 100% sure if I did it right.

```
        Cabin Embarked
1                    S
2                    S
3                    Q
4       E46          S
5                    c
```

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

arranges the data so its sorted by Fare in descending order worked very well.

```
     Sex   Age SibSp Parch   Ticket     Fare          Cabin Embarked
1 female 35.00     0     0 PC 17755 512.3292                       C
2   male 36.00     0     1 PC 17755 512.3292   B51 B53 B55         C
3   male 35.00     0     0 PC 17755 512.3292         B101          C
4   male 19.00     3     2    19950 263.0000 C23 C25 C27           S
5 female 23.00     3     2    19950 263.0000 C23 C25 C27           S
6 female 24.00     3     2    19950 263.0000 C23 C25 C27           S
7   male 64.00     1     4    19950 263.0000 C23 C25 C27           S
```

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

combines the sum of Parch and SibSp into FamSize The result was pretty cool. You are able to see what the size of the family is, additionally it provides details on cabin and the fare.

```
            Ticket    Fare   Cabin Embarked FamSize
1        A/5 21171  7.2500               S        1
2        PC 17599 71.2833    C85         C        1
3 STON/O2. 3101282  7.9250               S        0
4          113803 53.1000    C123        S        1
5          373450  8.0500               S        0
6          330877  8.4583               Q        0
7           17463 51.8625    E46         S        0
8          349909 21.0750               S        4
9          347742 11 1333               c        2
```

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

Calculated the number of women and man that survived

```
  Sex      meanFare numSurv
  <chr>       <dbl>    <int>
1 female      44.5      233
2 male        25.5      109
```

5.

```
> #Step 5, calculating the quantiles for the skin attribute
> quantile(diabetes$skin, probs = c(0.10, 0.30, 0.50, 0.60))
10% 30% 50% 60%
  0  10  23  27
>
```