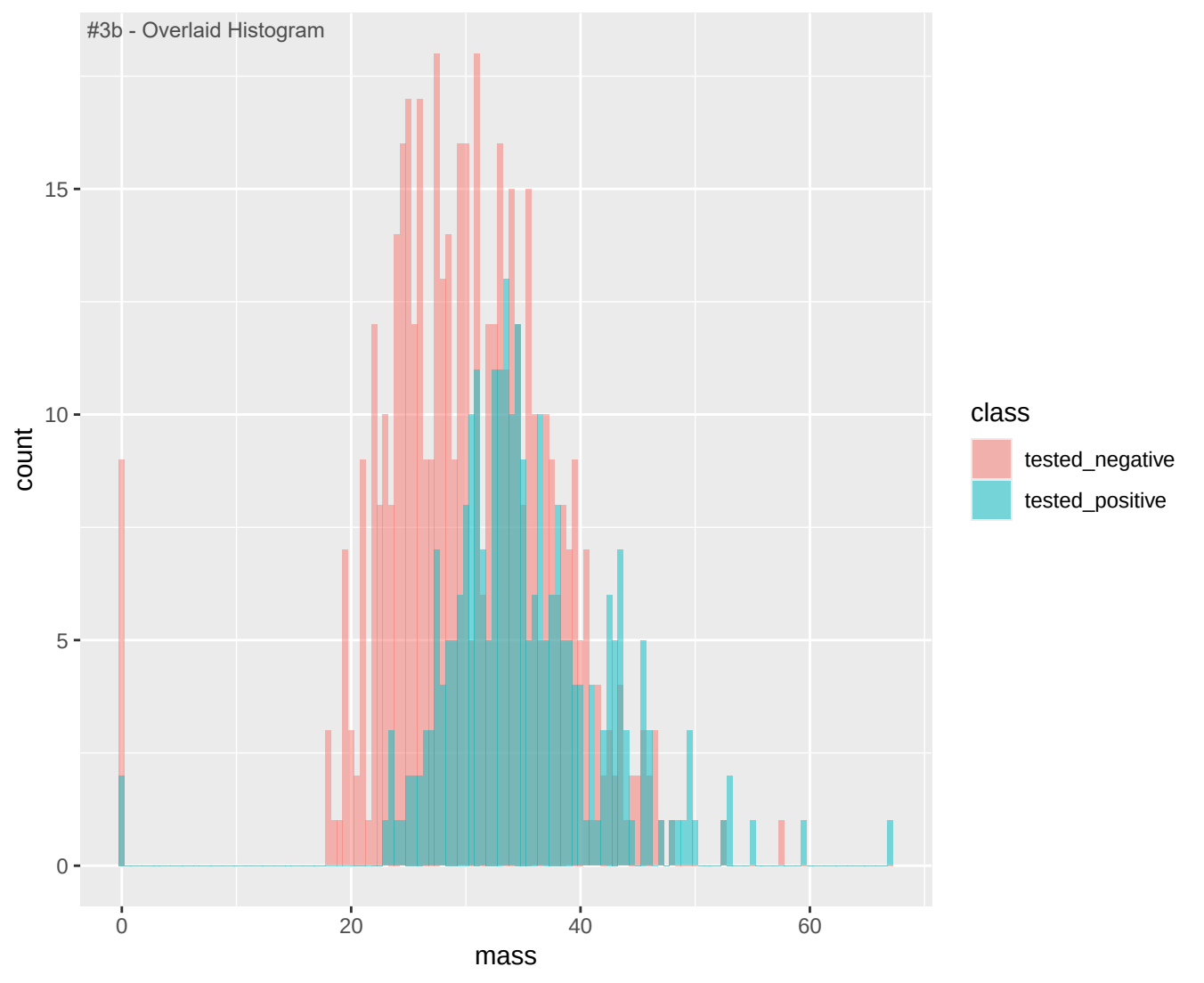
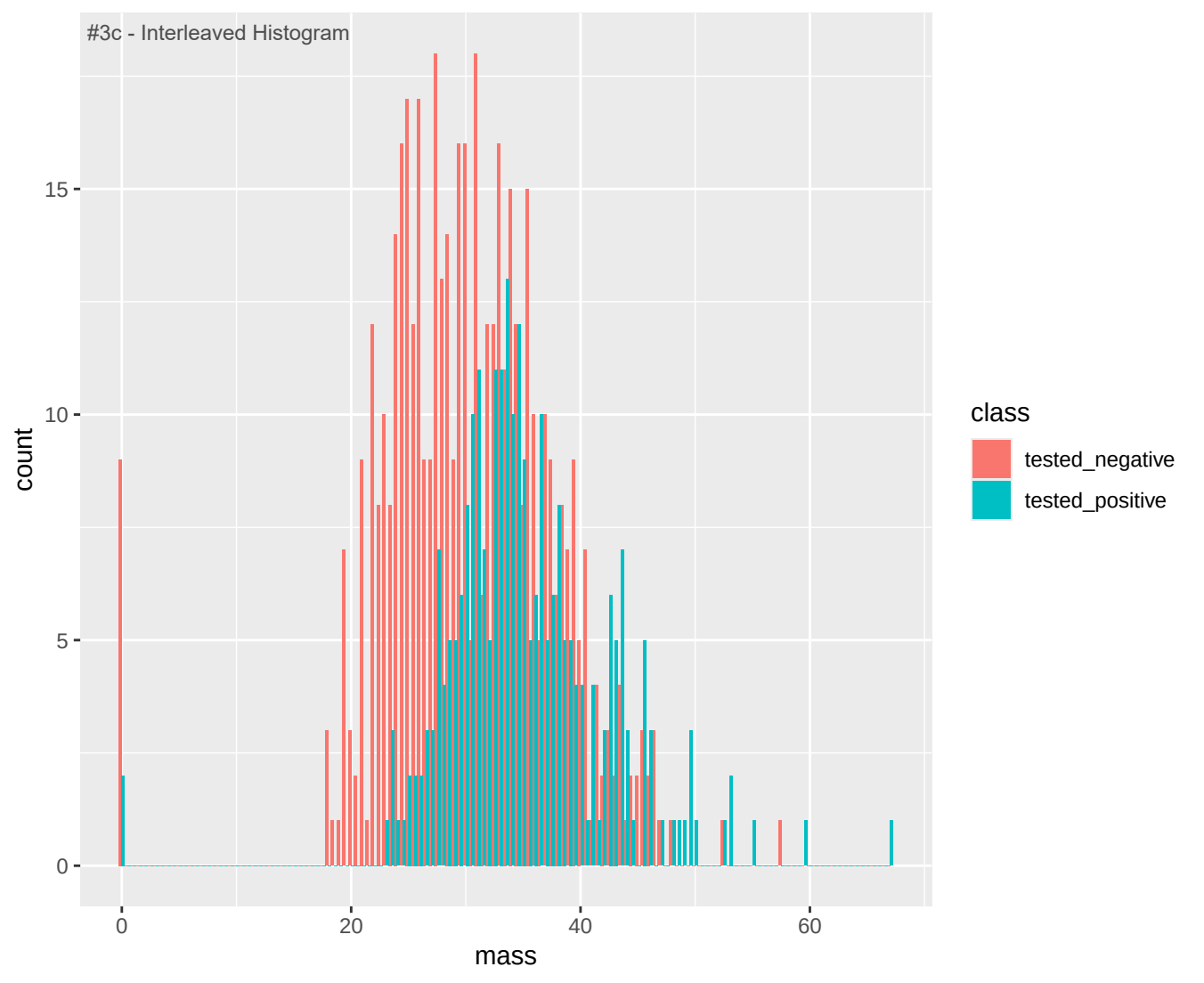
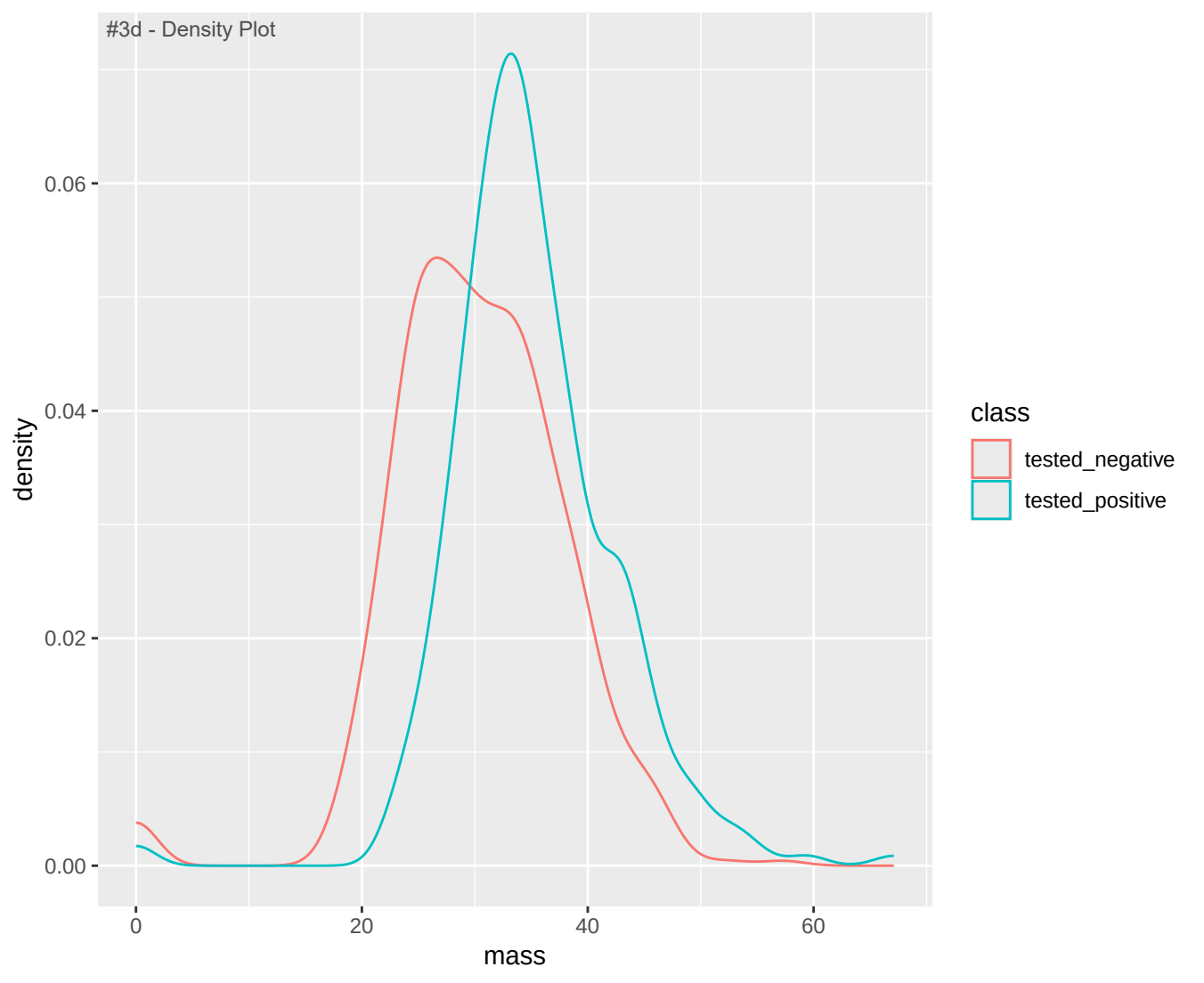
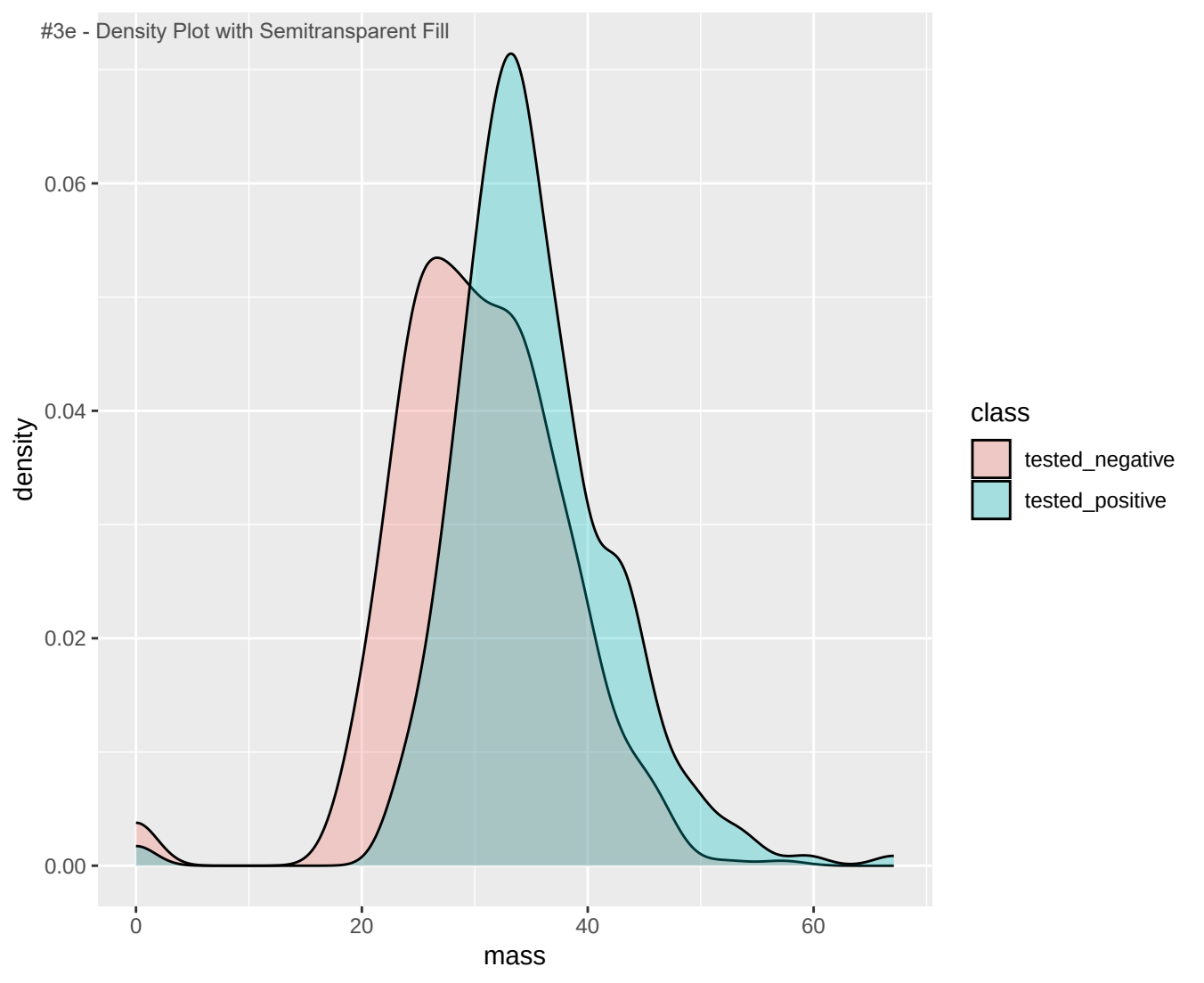

#3b - Overlaid Histogram
count
15
10
5
0
0
20
40
60
mass
class
tested_negative
tested_positive


#3c - Interleaved Histogram
count
15
10
5
0
0
20
40
60
mass
class
tested_negative
tested_positive


#3d - Density Plot
density
0.06
0.04
0.02
0.00
0
20
40
60
mass
class
tested_negative
tested_positive


#3e - Density Plot with Semitransparent Fill
density
0.06
0.04
0.02
0.00
0
20
40
60
mass
class
tested_negative
tested_positive

```{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)
```