

Riley Zidel

Question 1

```
su = read.delim("Su_raw_matrix.txt")
```

```
mean(su$Liver_2.CEL)
```

```
sd(su$Liver_2.CEL)
```

```
colMeans(su)
```

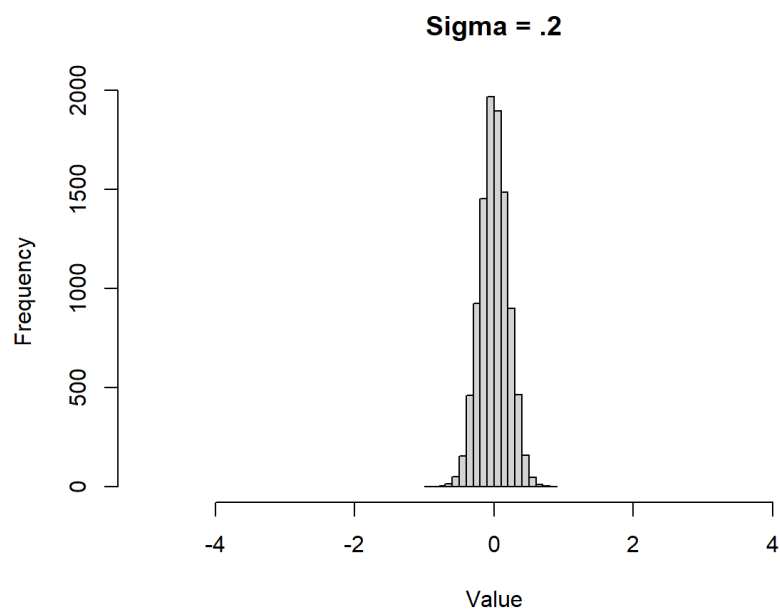
```
colSums(su)
```

Question 2 -----

```
x <- rnorm(10000, 0, .x)
```

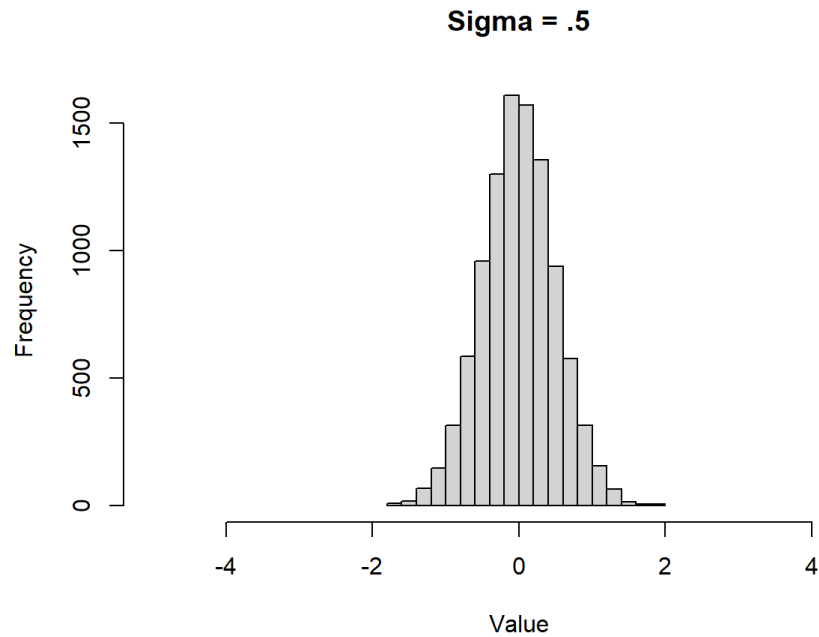
```
y <- rnorm(10000, 0, .5)
```

```
hist(x, main = "Sigma = .2", xlim = c(-5,5), xlab = "Value")
```



Riley Zidel

```
hist(y, main = "Sigma = .5", xlim = c(-5,5), xlab = "Value")
```



When sigma is .2 the histogram is much more narrow and taller. This is because

the standard deviation is smaller, which implies that the values deviate from

the mean less, causing them to be more closely bunched together.

Riley Zidel

#Question 3 -----

```
library(ggplot2)
```

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

#Overlaid Histogram

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

#Interleaved Histogram

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

#Density plot

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

#Density plot with semitransparent fill

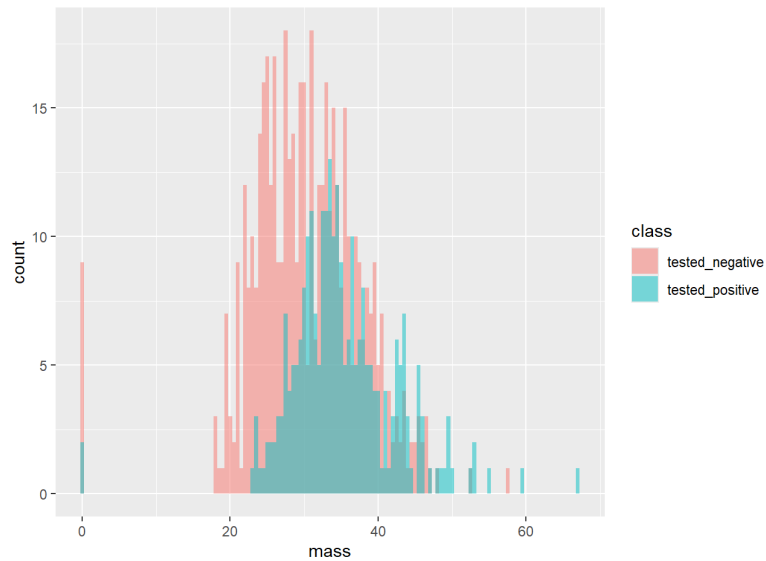
```
ggplot(dat, aes(x=rating, fill=cond)) + geom_density(alpha=.3)
```

```
diabetes <- read.csv("diabetes_train.csv")
```

Riley Zidel

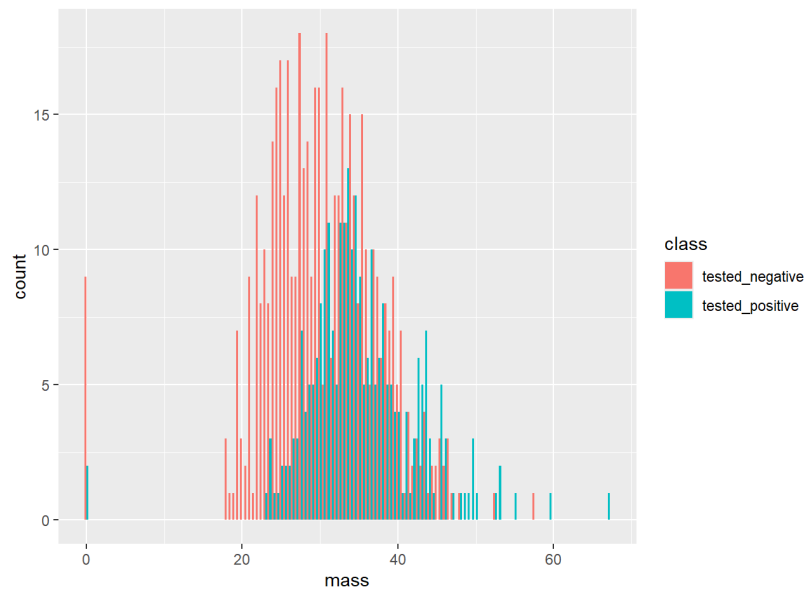
#Overlaid Histogram

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



#Interleaved Histogram

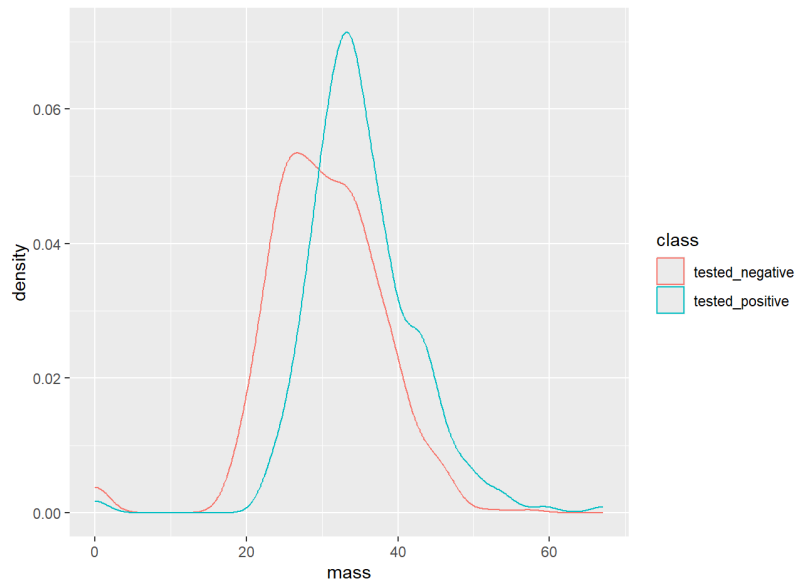
```
ggplot(diabetes, aes(x=mass, fill=class)) + geom_histogram(binwidth=.5, position="dodge")
```



Riley Zidel

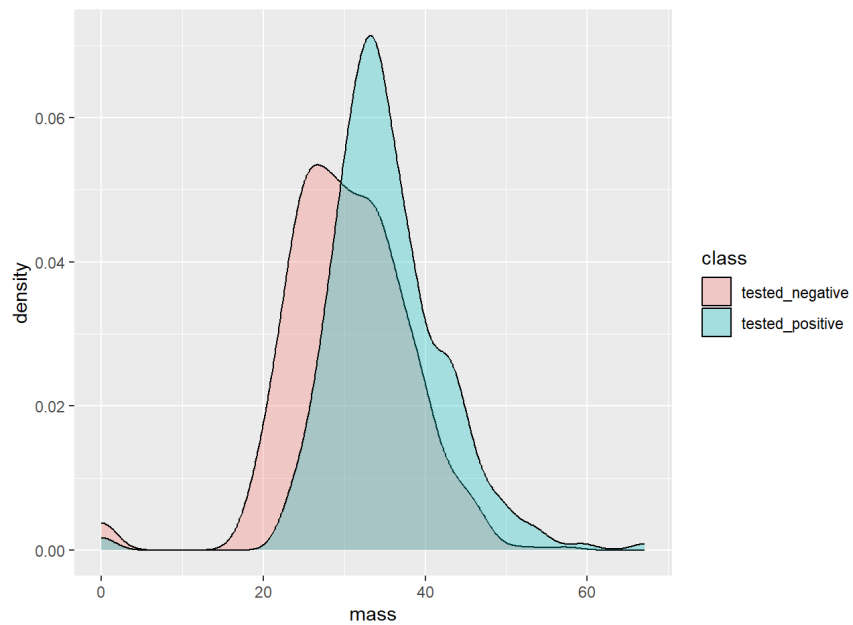
#Density plot

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



#Density plot with semitransparent fill

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



Riley Zidel

#Question 4 -----

```
library("tidyverse")
```

```
passengers = read.csv("titanic.csv")
```

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

```
# The above ignores any rows with NA under age, then prints the summary of the data
```

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

```
# The above filters and prints only the rows in which the sex of the row is "male"
```

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

```
# The above sorts the rows by fare, with the highest fare at the top and lowest at the bottom
```

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

```
# The above creates a new column called FamSize, which is equal to Parch + SibSp
```

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

```
# The above groups the passengers by sex, then displays a summary of the average
```

```
# fare and number survived for each group.
```

Riley Zidel

#Question 5 -----

```
quantile(diabetes$skin, c(.10, .30, .50, .60))
```