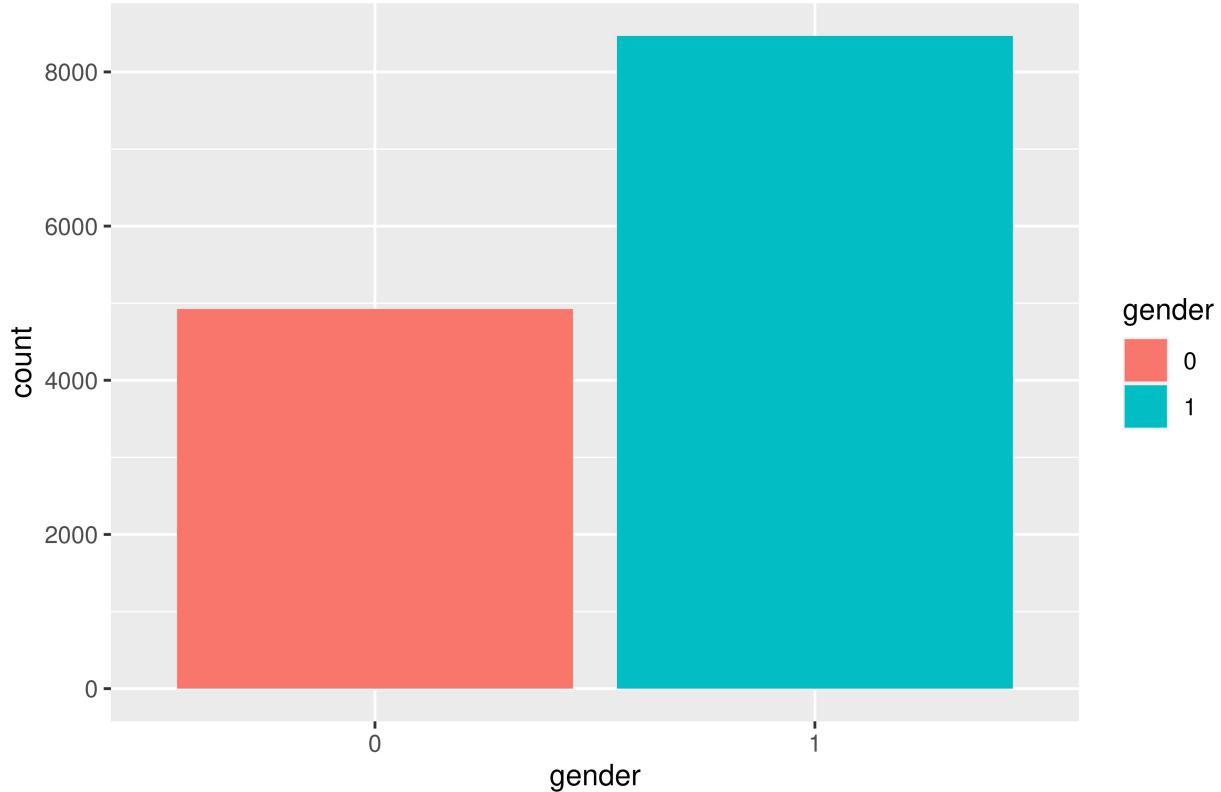


## Visualizations

```
#reading in the pre-preprocessed dataset  
  
data = read.csv("finalize_df.csv",sep = " ")
```

```
data %>% mutate(gender = as.factor(gender)) %>% ggplot(aes(gender)) + geom_bar(aes(fill=gender)) + ggti
```

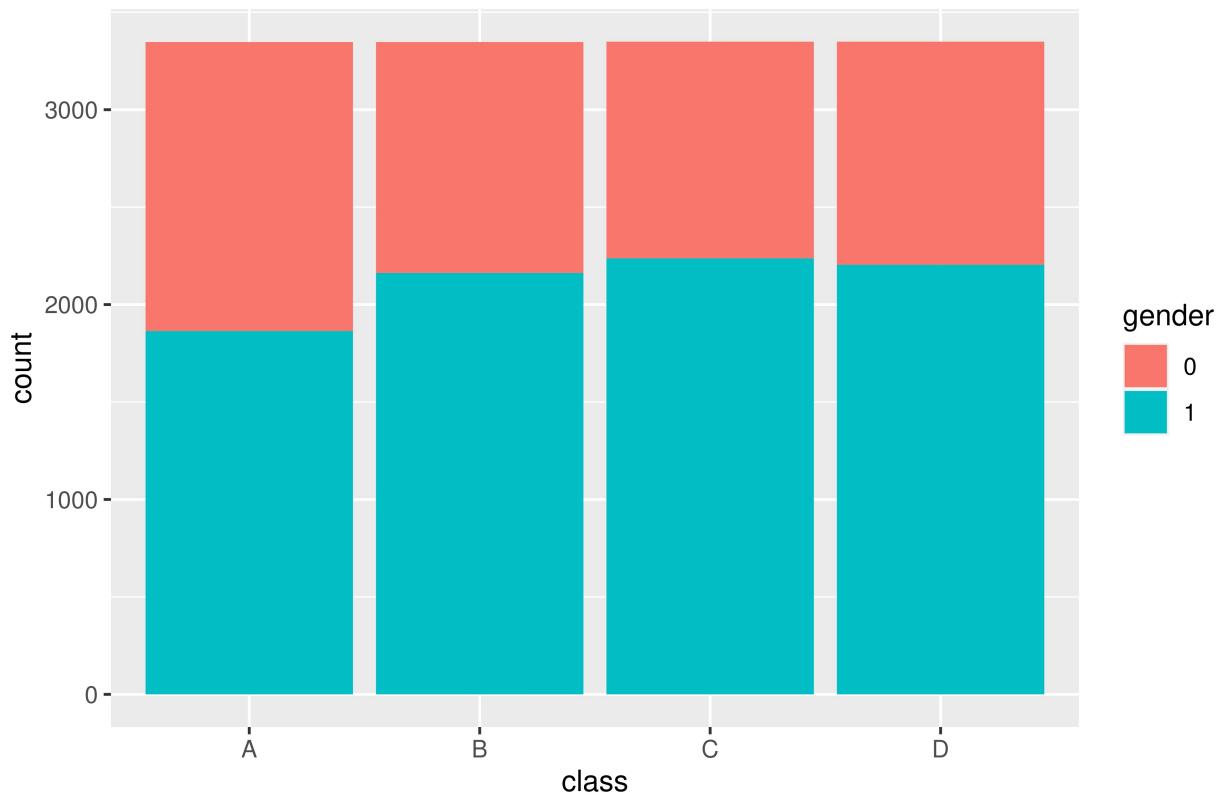
Count of Female vs Male



```
#bar plots
```

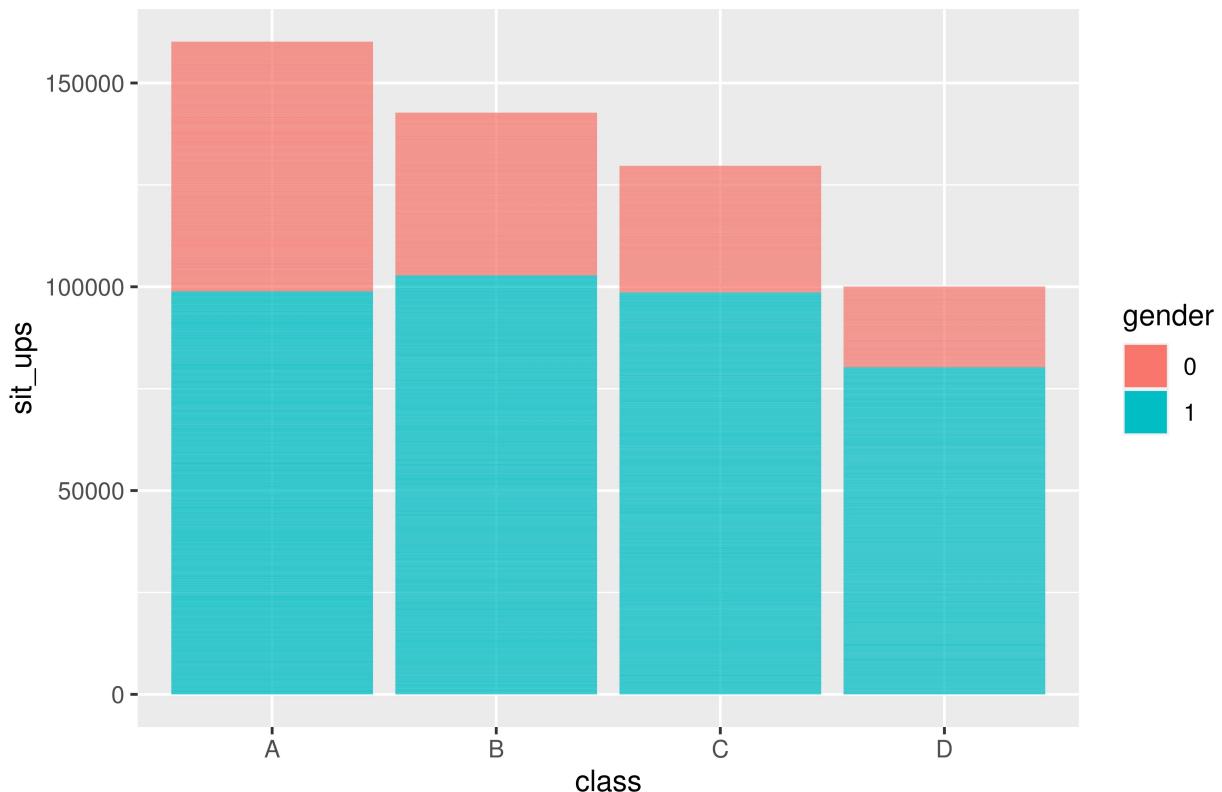
```
data %>% mutate(gender = factor(gender)) %>% ggplot(aes(class)) + geom_bar(aes(fill=gender)) +ggttitle("T
```

## Response Variable by Gender and Class



```
data %>% mutate(class = as.factor(class), gender = as.factor(gender)) %>% ggplot(aes(x=class, y=sit_ups))
```

## Total Sit Ups by Gender and Class



```
#feature dataset

features_data = data[,-c(2,12)] #12 is response variable and 2 is gender

#Renaming feature names for the plots

feature_names = c("age", "height in cm", "weight in kg", "body fat in %","distolic blood pressure","sys

#Distribution plot for all features by gender

mylist = list()

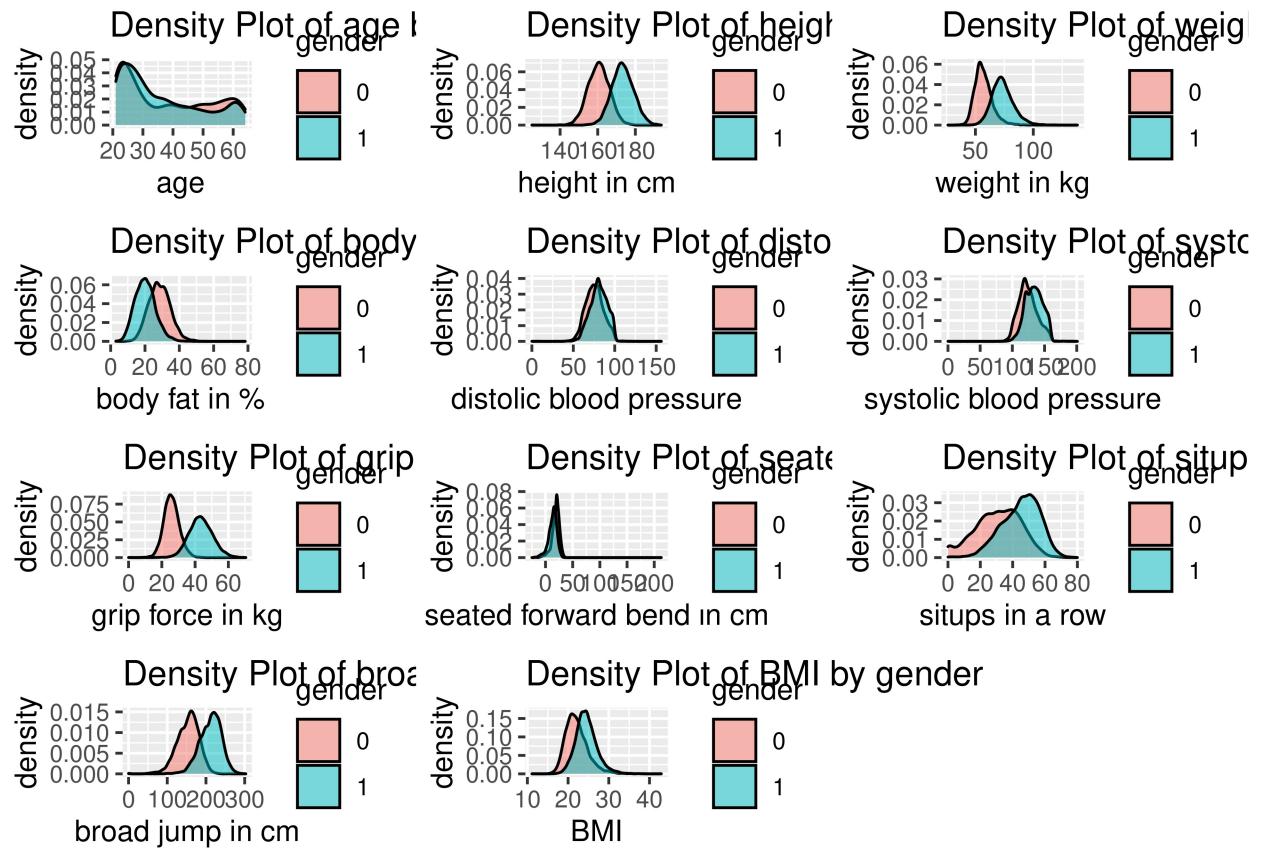
for(i in 1:ncol(features_data)){

  a_plot = data %>% mutate(gender = as.factor(gender)) %>% ggplot(aes_string(features_data[,i],fill="gen

  mylist[[i]] = a_plot

}

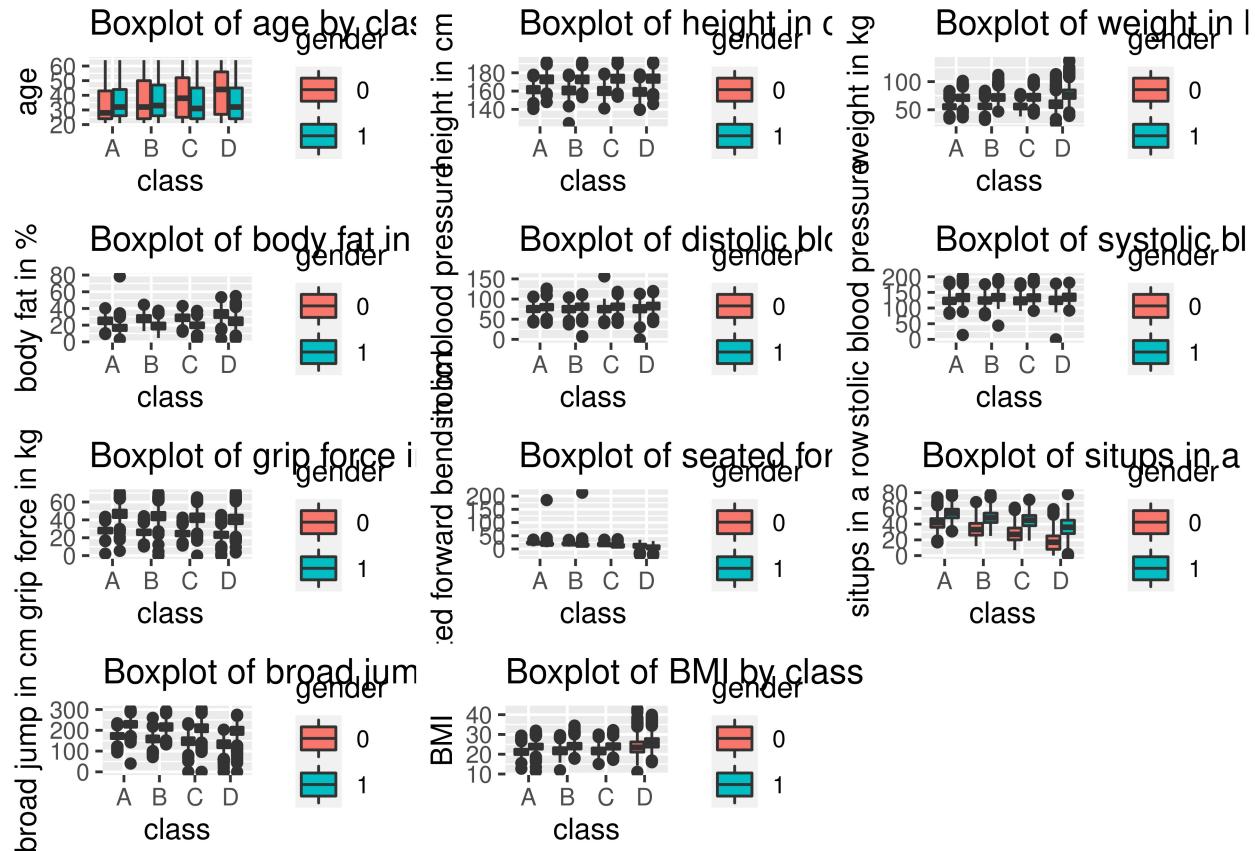
grid.arrange(grobs = mylist,3)
```



#Distribution plot for all features by response variable (class)

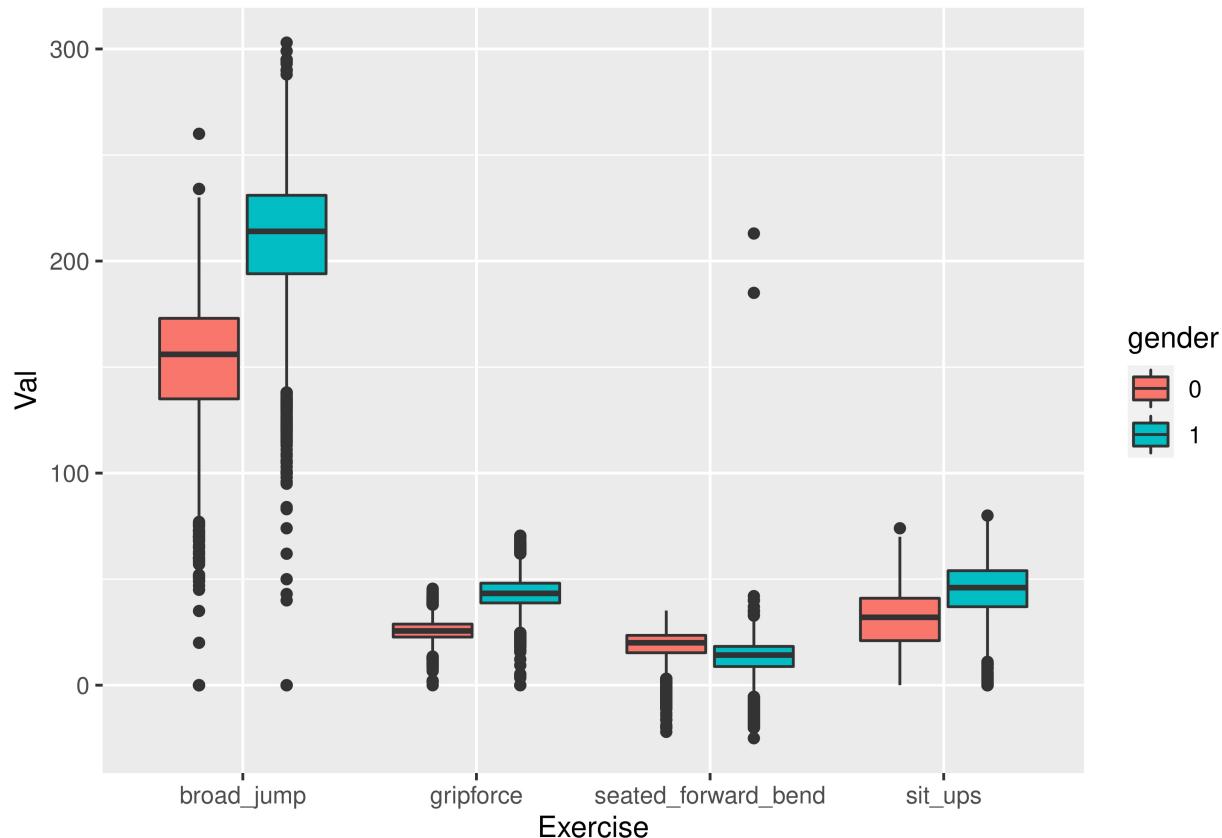
```
boxplot_list = list()

for(i in 1:ncol(features_data)){
  boxplot_list[[i]] = data %>% mutate(class = as.factor(class), gender = as.factor(gender)) %>%
    ggplot(aes_string("class", features_data[,i], fill = "gender")) + geom_boxplot() + ylab(feature_names[i])
}
grid.arrange(grobs=boxplot_list, 3)
```



#gender by situps, broadjumps, seated bend, and gripforce

```
data %>% mutate(gender = as.factor(gender)) %>%
  pivot_longer(gripforce:broad_jump,names_to = "Exercise",values_to = "Val") %>%
  ggplot(aes(Exercise,Val,fill=gender)) + geom_boxplot()
```



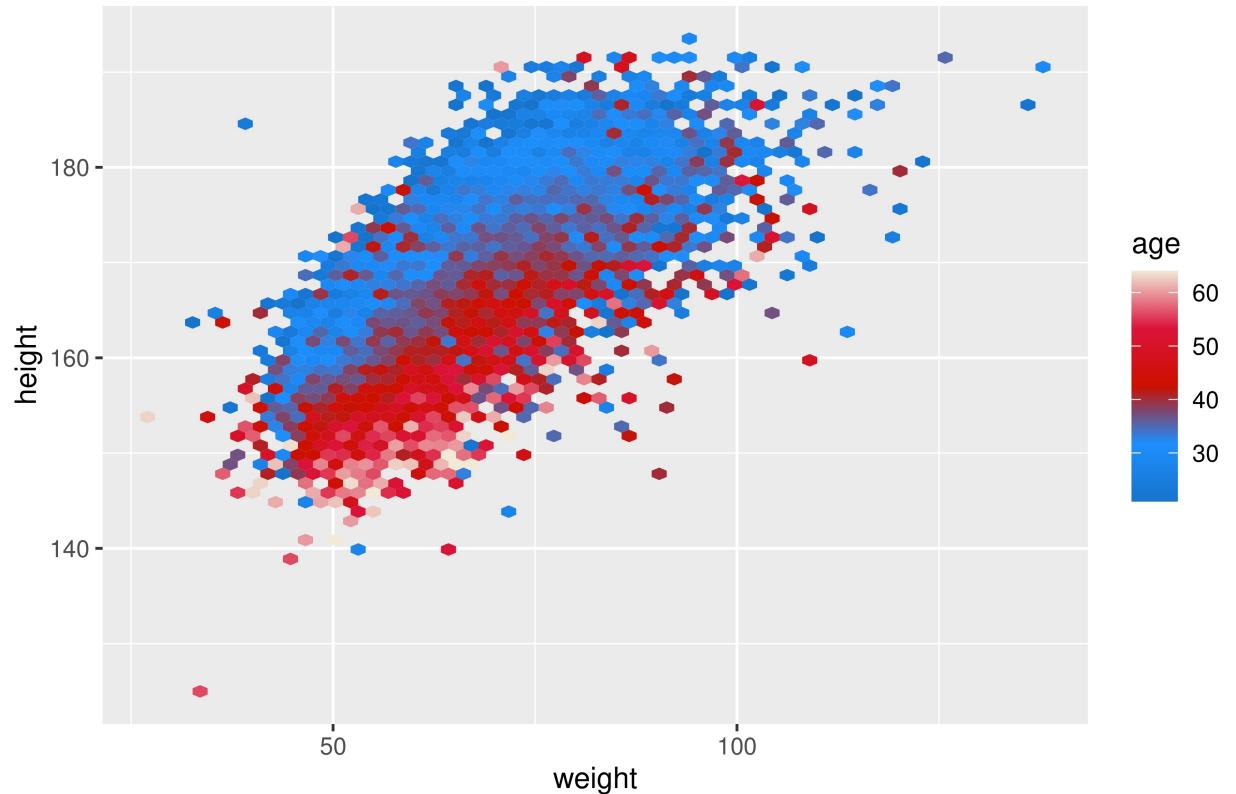
```
library(thewitcher)
```

```
## Warning: package 'thewitcher' was built under R version 4.1.2
```

```
#height vs weight by age
```

```
data %>% ggplot(aes(weight,height,z=age)) + stat_summary_hex(bins=60) + scale_fill_witcher(option="skel")
```

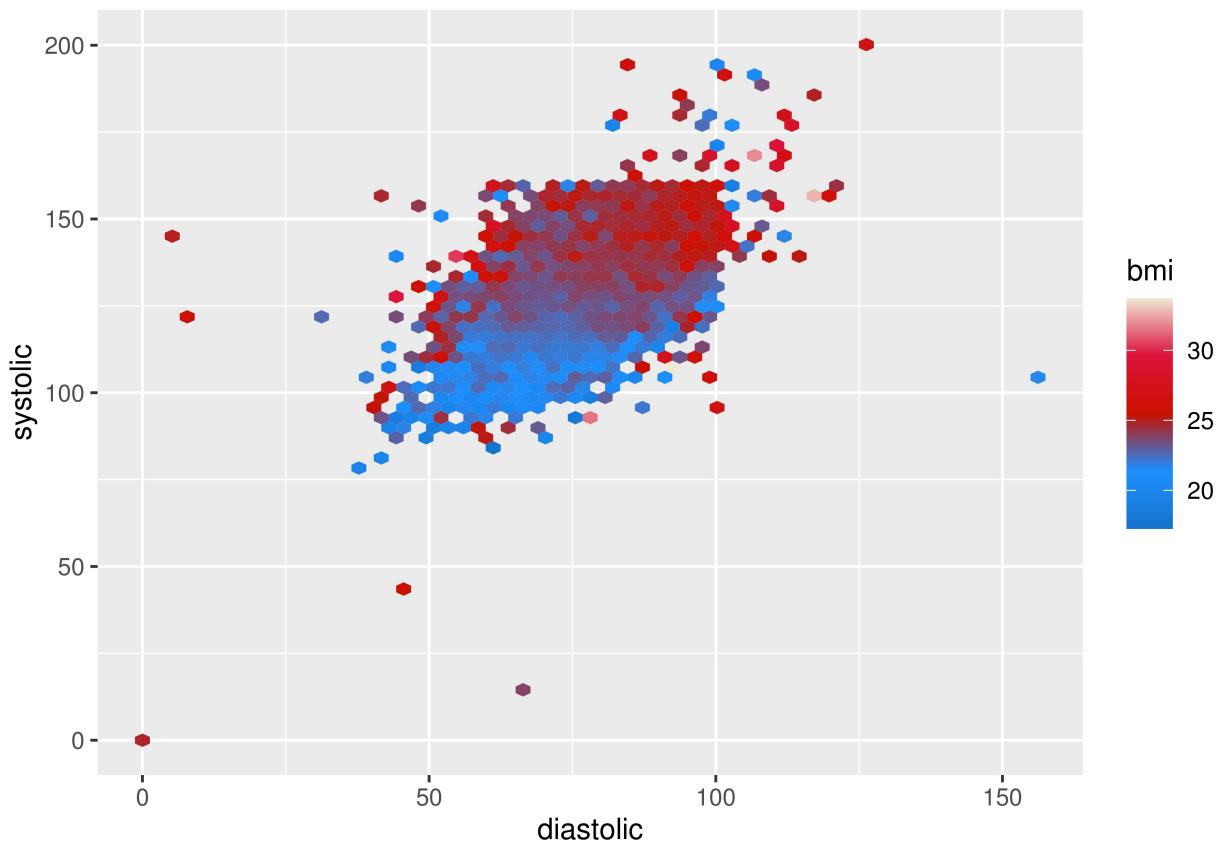
## Weight vs Height by Age



#most of the overweight ppl are in younger than 30

#systolic diastolic bmi

```
data %>% ggplot(aes(diastolic,systolic,z=bmi)) + stat_summary_hex(bins=60) + scale_fill_witcher(option=
```



correlation plot

```
corr = cor(data[,-c(2,12)])
ggcorrplot(corr, hc.order = TRUE,
  ggtheme = ggplot2::theme_dark(), colors = c("#d41945", "white", "#1eabb3")
)
```

