

## Assignment 2 MATH 208 (Question 2)

Parsa Yadollahi  
260869949  
MATH 208 - Assignment 2

### Question 2

a)

```
data(Diabetes)
df_diabetes = as_tibble(Diabetes)
df_grouped = df_diabetes %>%
  group_by(group) %>%
  summarise_all(list(
    Avg=mean,
    Med=median
  )) %>%
  pivot_longer(
    col=contains('_'),
    names_to = "Measure") %>%
  pivot_wider(
    id_cols = Measure,
    names_from = group
  ) %>%
  arrange(
    desc(Measure)
  )

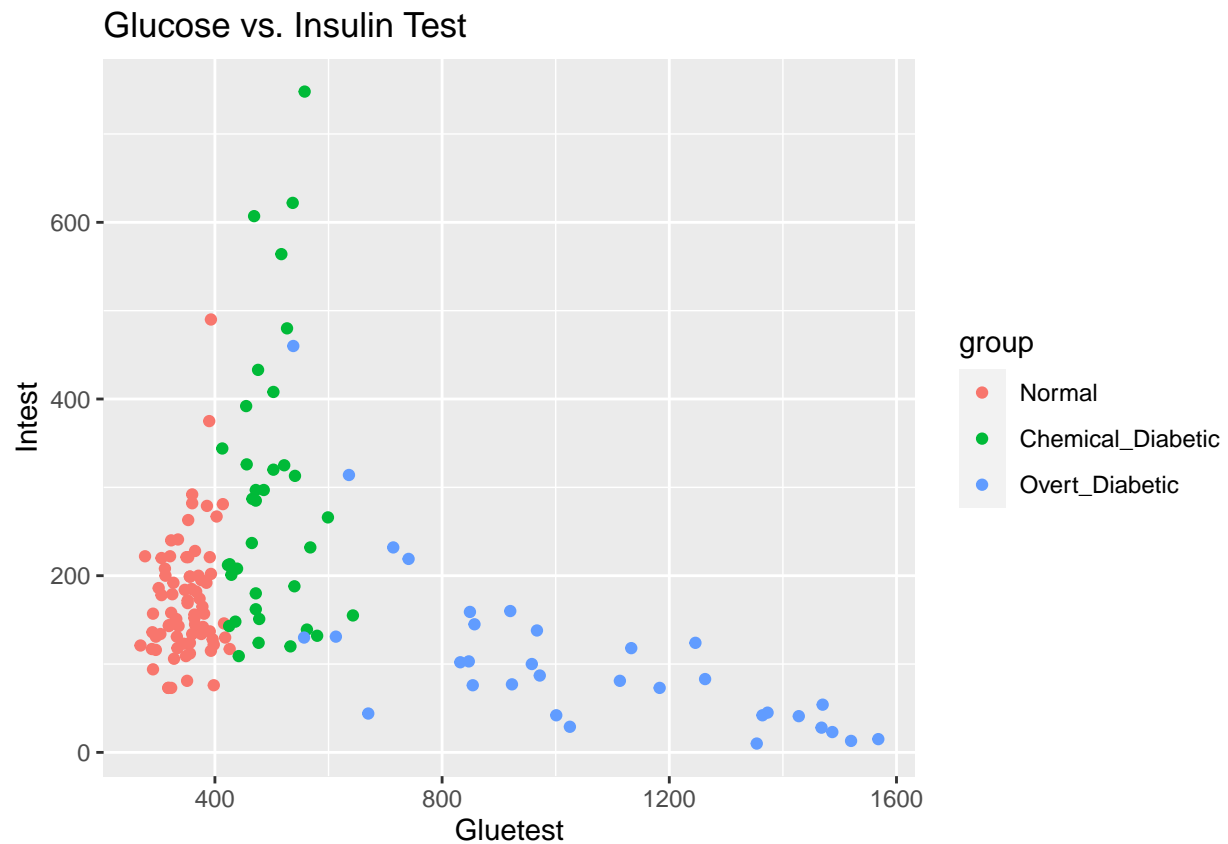
df_grouped
```

```
## # A tibble: 10 x 4
##   Measure      Normal Chemical_Diabetic Overt_Diabetic
##   <chr>      <dbl>      <dbl>      <dbl>
## 1 sspg_Med   105          223          320
## 2 sspg_Avg   114          209.          319.
## 3 relwt_Med   0.95          1.06          0.98
## 4 relwt_Avg   0.937          1.06          0.984
## 5 instest_Med 157          252.           83
## 6 instest_Avg 173.          288           106
## 7 glutest_Med 353          476.           972
## 8 glutest_Avg 350.          494.          1044.
## 9 glufast_Med  90          99.5           203
## 10 glufast_Avg 91.2          99.3           218.
```

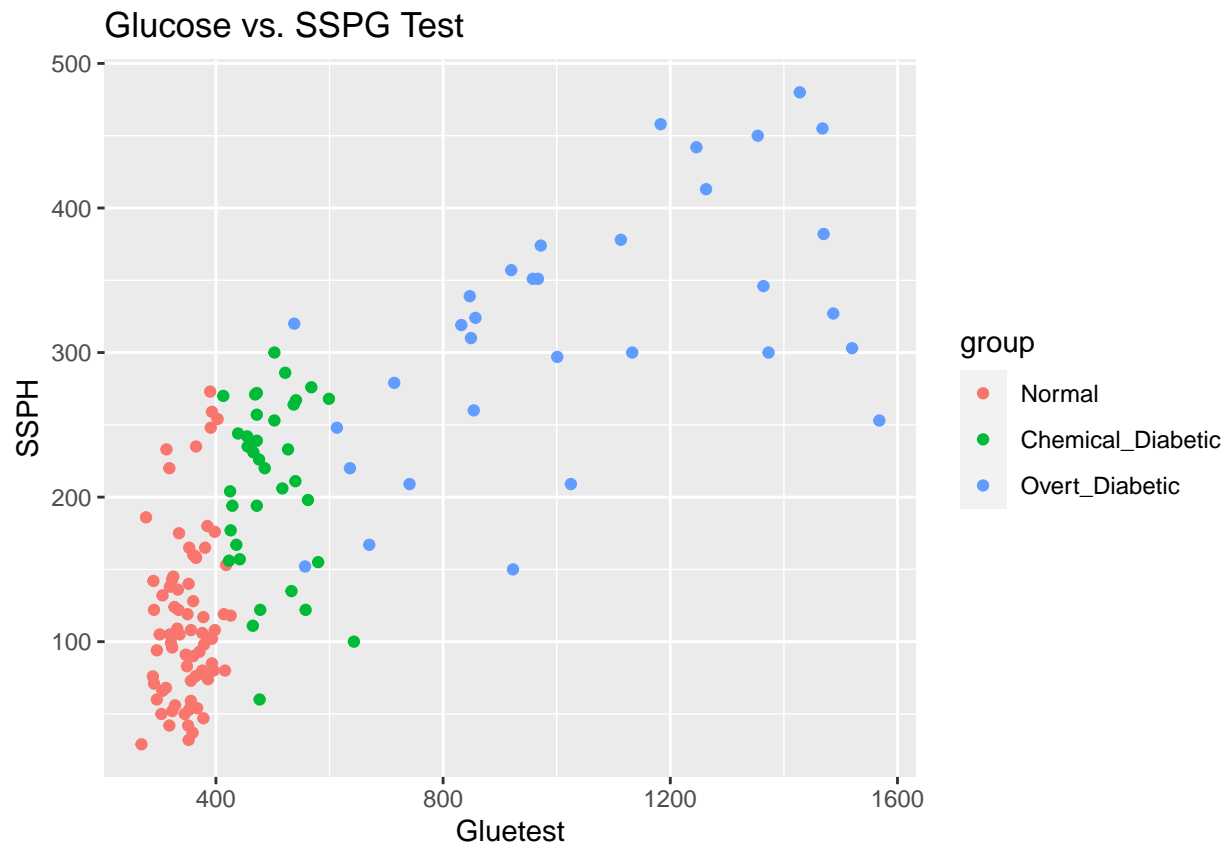
The variable that seem to differentiate amongst the different types of diabetes is glutest, intest, glufast and sspg

b)

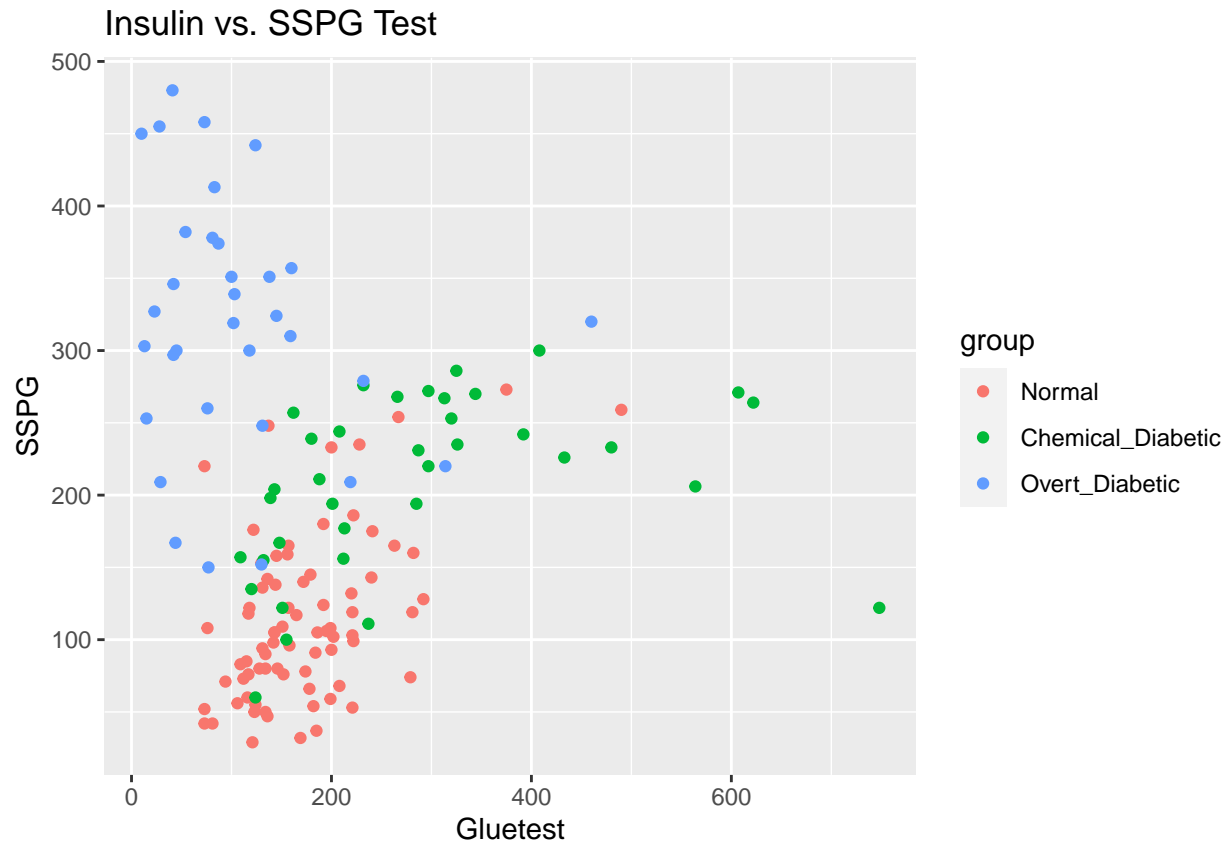
```
df_test = df_diabetes %>% group_by(group)
ggplot(df_diabetes, aes(x=glutest, y=instest, col = group)) + geom_point() +
  labs(title="Glucose vs. Insulin Test", x="Glutetest", y="Intest")
```



```
ggplot(df_diabetes, aes(x=glutest, y = sspg, col = group)) + geom_point() +
  labs(title="Glucose vs. SSPG Test", x="Glutetest", y="SSPH")
```



```
ggplot(df_diabetes, aes(x=instest, y = sspg, col = group)) + geom_point() +  
  labs(title="Insulin vs. SSPG Test", x="Glutest", y="SSPG")
```

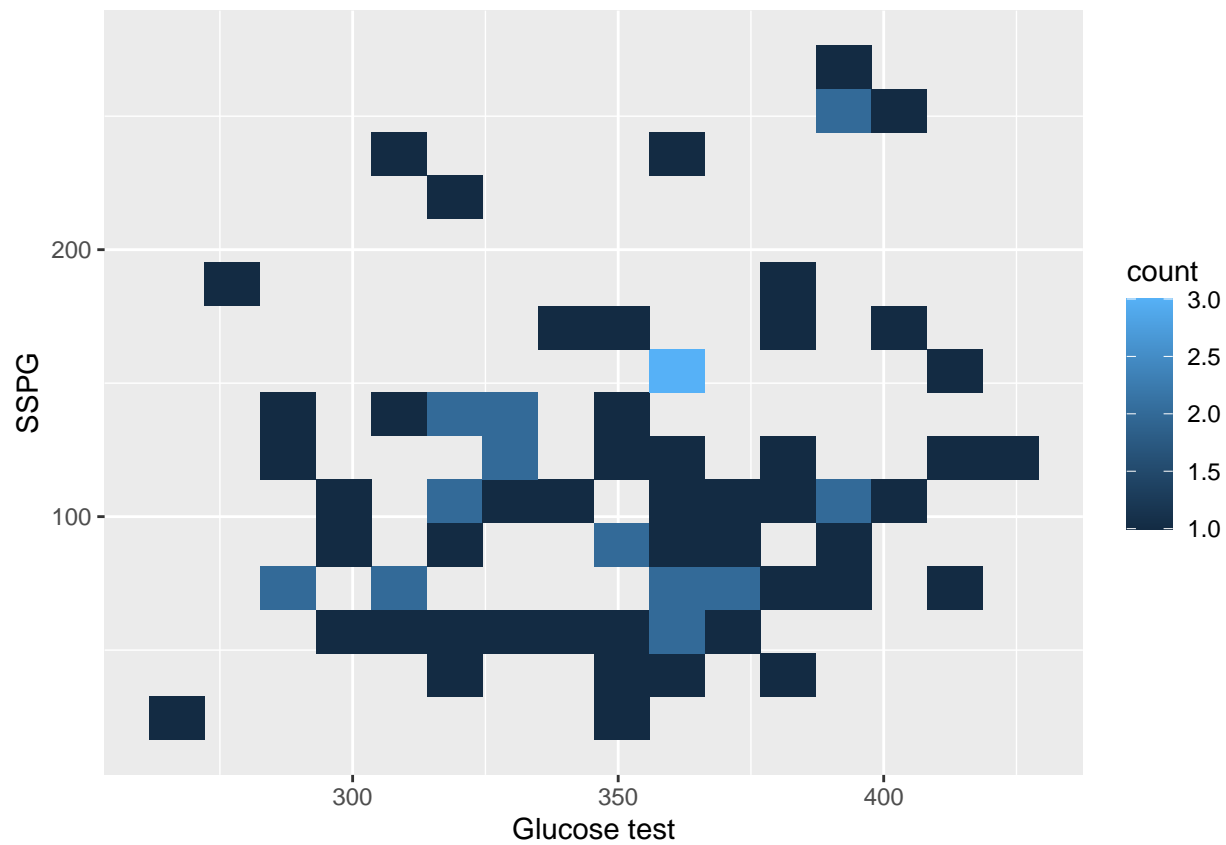


Glucose vs. SSPG test seem to return the greatest distinction looking at the plots above

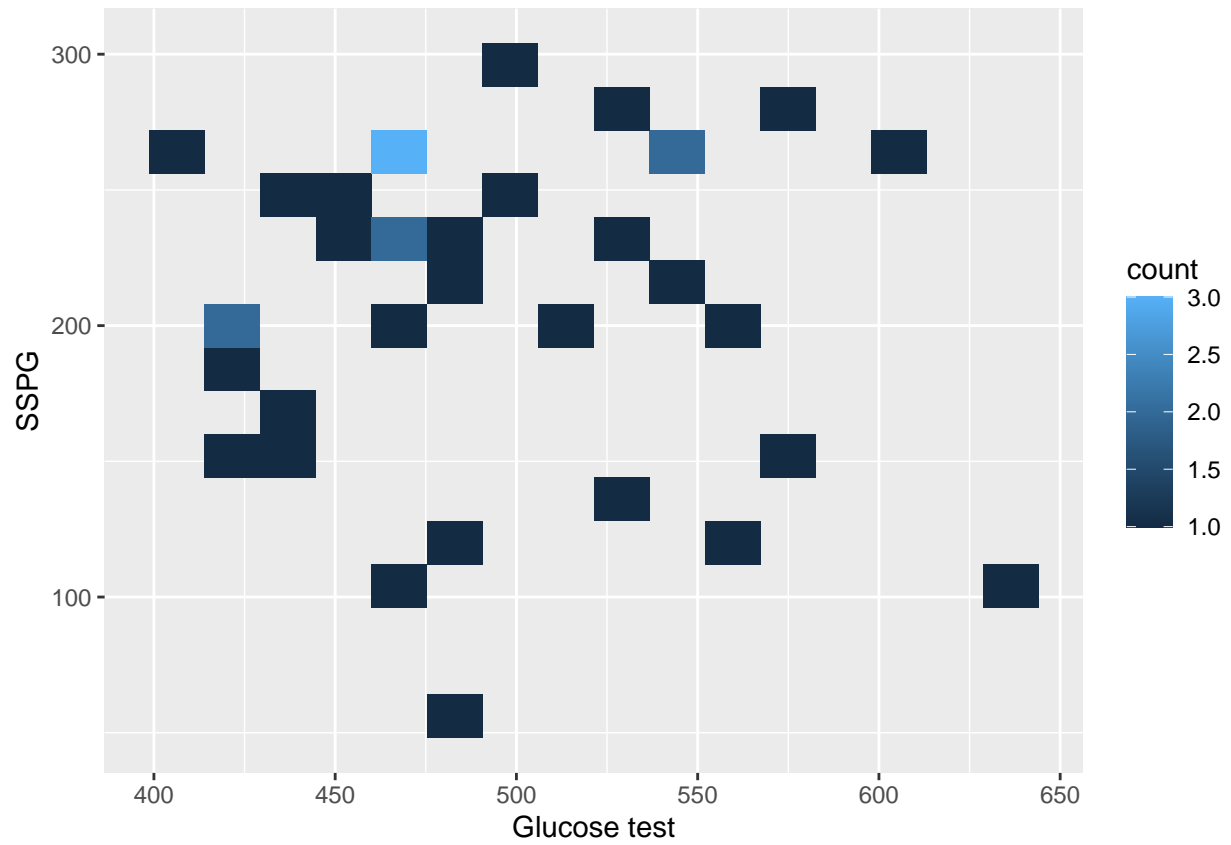
c)

```
# Group data
group_normal = df_diabetes[df_diabetes$group == "Normal",]
group_chemical = df_diabetes[df_diabetes$group == "Chemical_Diabetic",]
group_overt = df_diabetes[df_diabetes$group == "Overt_Diabetic",]

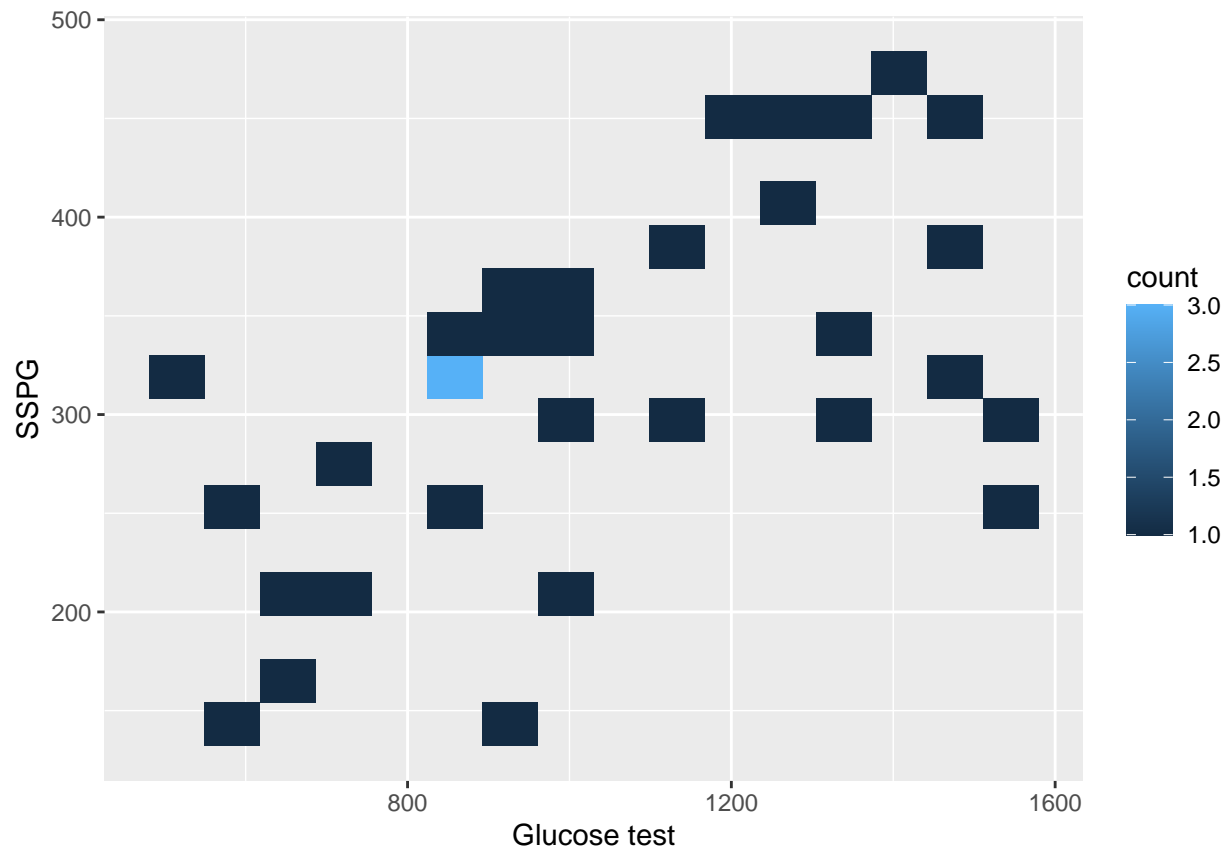
# Histograms
ggplot(group_normal, aes(x=glutest, y=sspg)) + geom_bin2d(bins=15) +
labs(x="Glucose test", y="SSPG")
```



```
ggplot(group_chemical, aes(x=glutest, y=sspg)) + geom_bin2d(bins=15) +  
labs(x="Glucose test", y="SSPG")
```



```
ggplot(group_overt, aes(x=glutest, y=sspg)) + geom_bin2d(bins=15) +  
labs(x="Glucose test", y="SSPG")
```

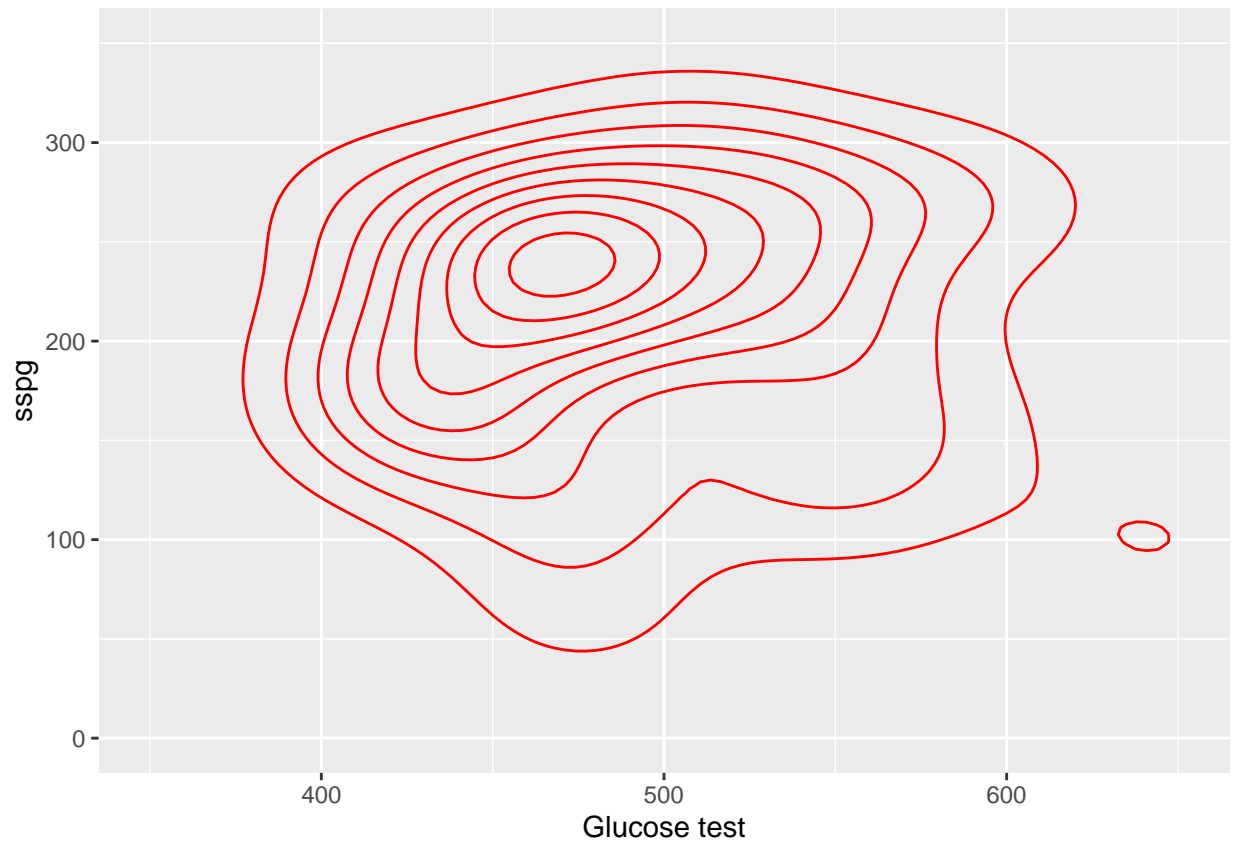


```
# contour plot
ggplot(group_normal, aes(x=glutest,y=sspg)) + geom_density_2d(col="red") +
  labs(x="Glucose test", y="sspg") + xlim(c(200,500)) + ylim(c(-50,350))
```

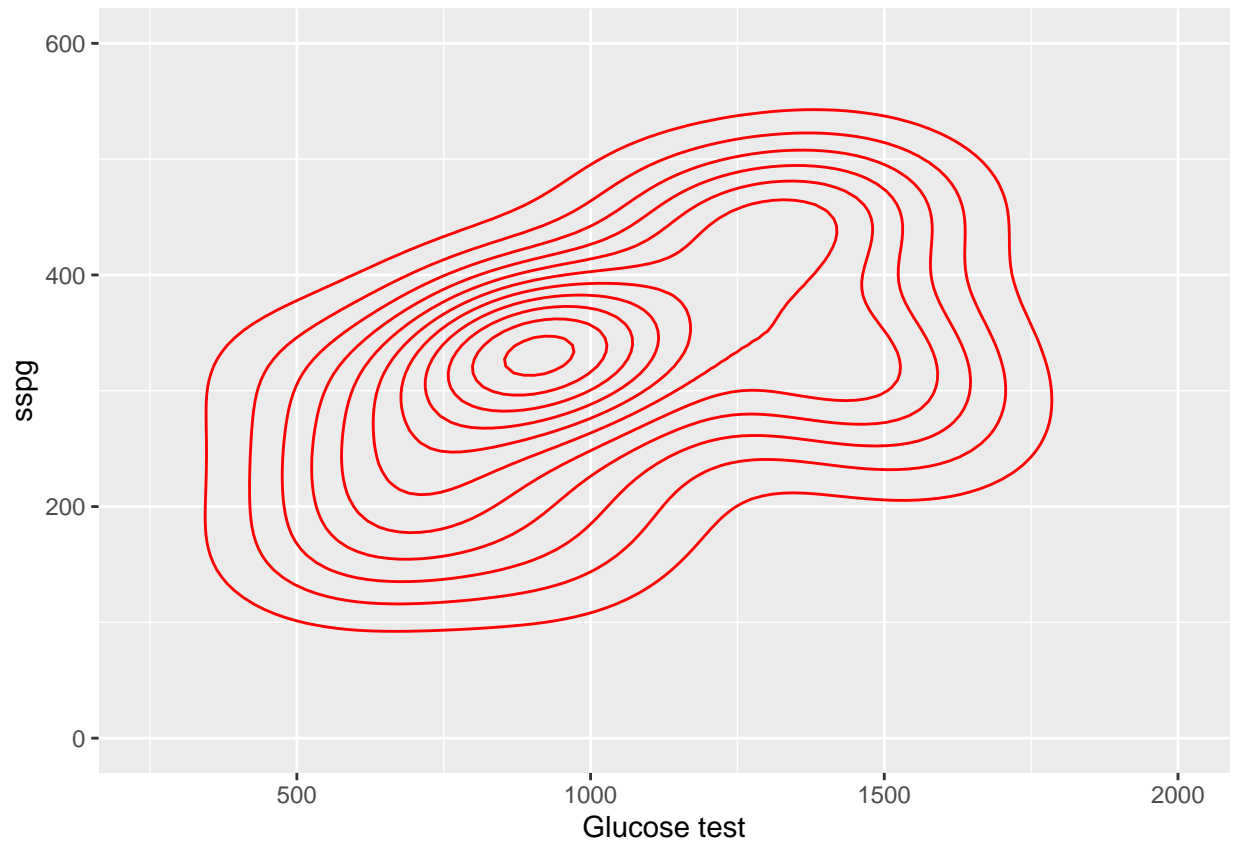


```
ggplot(group_chemical, aes(x=glutest,y=sspg)) + geom_density_2d(col="red") +  
  labs(x="Glucose test", y="sspg") + xlim(c(350,650)) + ylim(c(0,350))
```





```
ggplot(group_overt, aes(x=glutest,y=sspg)) + geom_density_2d(col="red") +  
  labs(x="Glucose test", y="sspg") + xlim(c(250,2000)) + ylim(c(0,600))
```



Yes, I find the plots provide useful summaries of the differences in the distributions.