pt.R

Pranshu

Wed Jun 01 01:10:05 2016

```
setwd("~/Research/Diabetes analysis2/NewDiabetes")
pima.indians.diabetes <- read.csv("C:/Users/IBM_ADMIN/Desktop/pima-indians-diabetes.d
ata", header=FALSE)
names(pima.indians.diabetes)<-c('TimesPregnant','GlucoseLV','DiastolicBP','TricepsThi
ckness','SerumInsulin','BMI','Heridarymarkup','Age','Classification')
View(pima.indians.diabetes)
library(ggplot2)
```

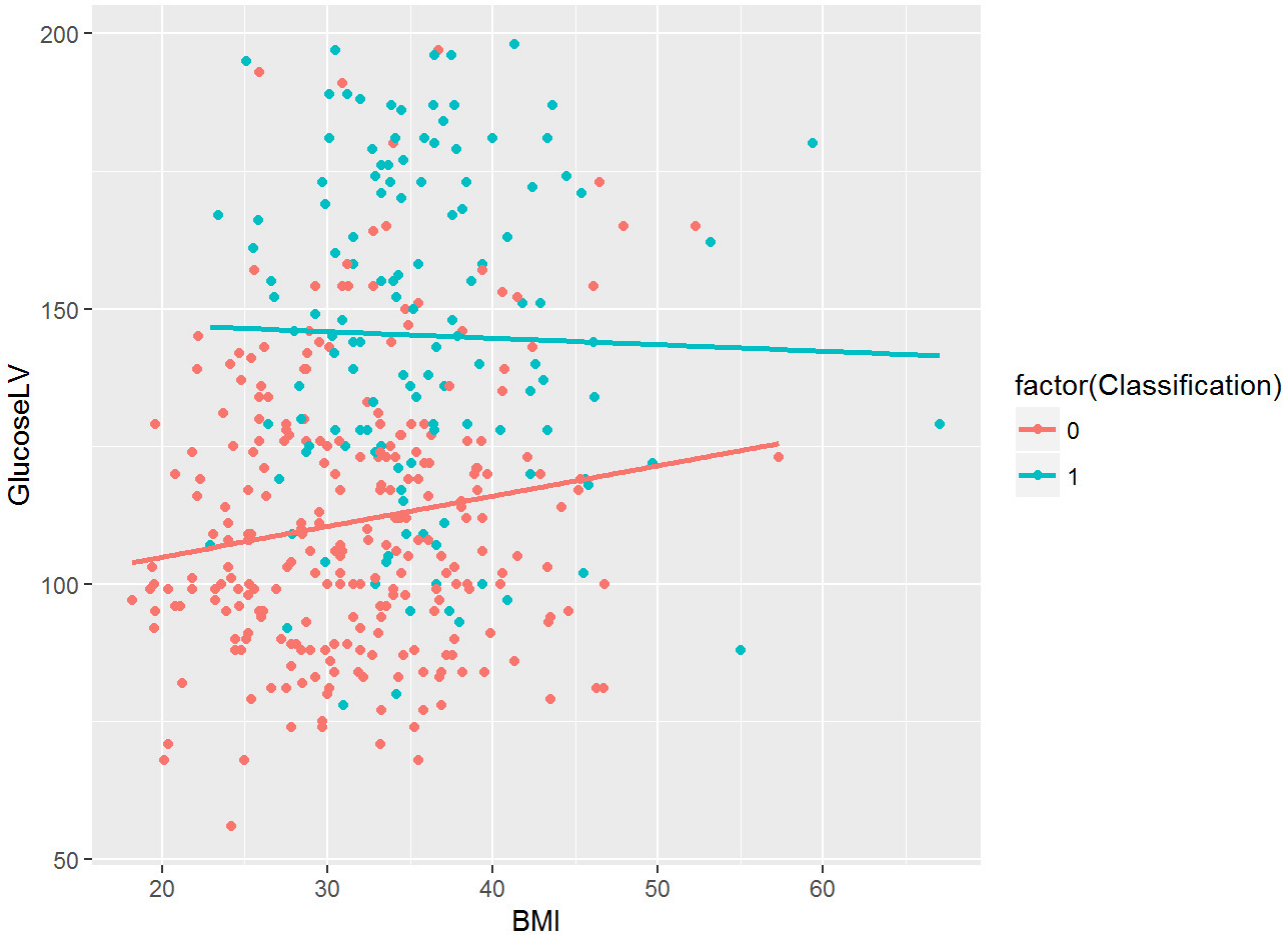
```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
pima.indians.diabetes$DiastolicBP[pima.indians.diabetes$DiastolicBP==0]<-NA
pima.indians.diabetes$GlucoseLV[pima.indians.diabetes$GlucoseLV==0]<-NA
pima.indians.diabetes$TricepsThickness[pima.indians.diabetes$TricepsThickness==0]<-NA
pima.indians.diabetes$SerumInsulin[pima.indians.diabetes$SerumInsulin==0]<-NA
pima.indians.diabetes$BMI[pima.indians.diabetes$BMI==0]<-NA
pima.indian.diabetes2<-na.omit(pima.indians.diabetes)
summary(pima.indian.diabetes2)
```

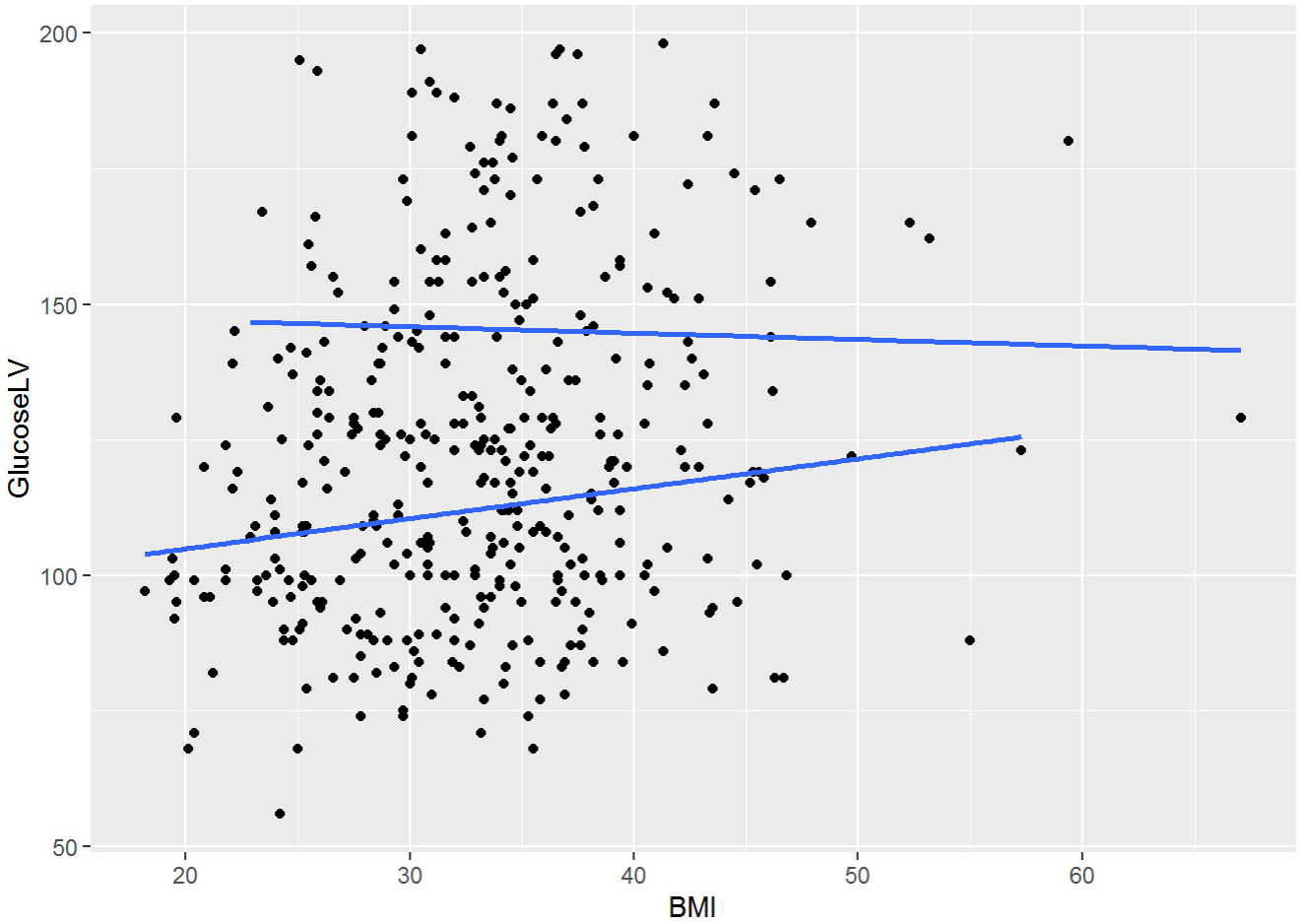
```
##  TimesPregnant      GlucoseLV      DiastolicBP     TricepsThickness
##  Min.   : 0.000   Min.   : 56.0   Min.   : 24.00   Min.   : 7.00
##  1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 62.00   1st Qu.:21.00
##  Median : 2.000   Median :119.0   Median : 70.00   Median :29.00
##  Mean   : 3.301   Mean   :122.6   Mean   : 70.66   Mean   :29.15
##  3rd Qu.: 5.000   3rd Qu.:143.0   3rd Qu.: 78.00   3rd Qu.:37.00
##  Max.   :17.000   Max.   :198.0   Max.   :110.00   Max.   :63.00
##   SerumInsulin         BMI        Heridarymarkup       Age
##  Min.   : 14.00   Min.   :18.20   Min.   :0.0850   Min.   :21.00
##  1st Qu.: 76.75   1st Qu.:28.40   1st Qu.:0.2697   1st Qu.:23.00
##  Median :125.50   Median :33.20   Median :0.4495   Median :27.00
##  Mean   :156.06   Mean   :33.09   Mean   :0.5230   Mean   :30.86
##  3rd Qu.:190.00   3rd Qu.:37.10   3rd Qu.:0.6870   3rd Qu.:36.00
##  Max.   :846.00   Max.   :67.10   Max.   :2.4200   Max.   :81.00
##  Classification
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3316
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
View(pima.indian.diabetes2)
ggplot(pima.indian.diabetes2, aes(x=Age,y=BMI,col=factor(Classification)))+geom_point
()
```

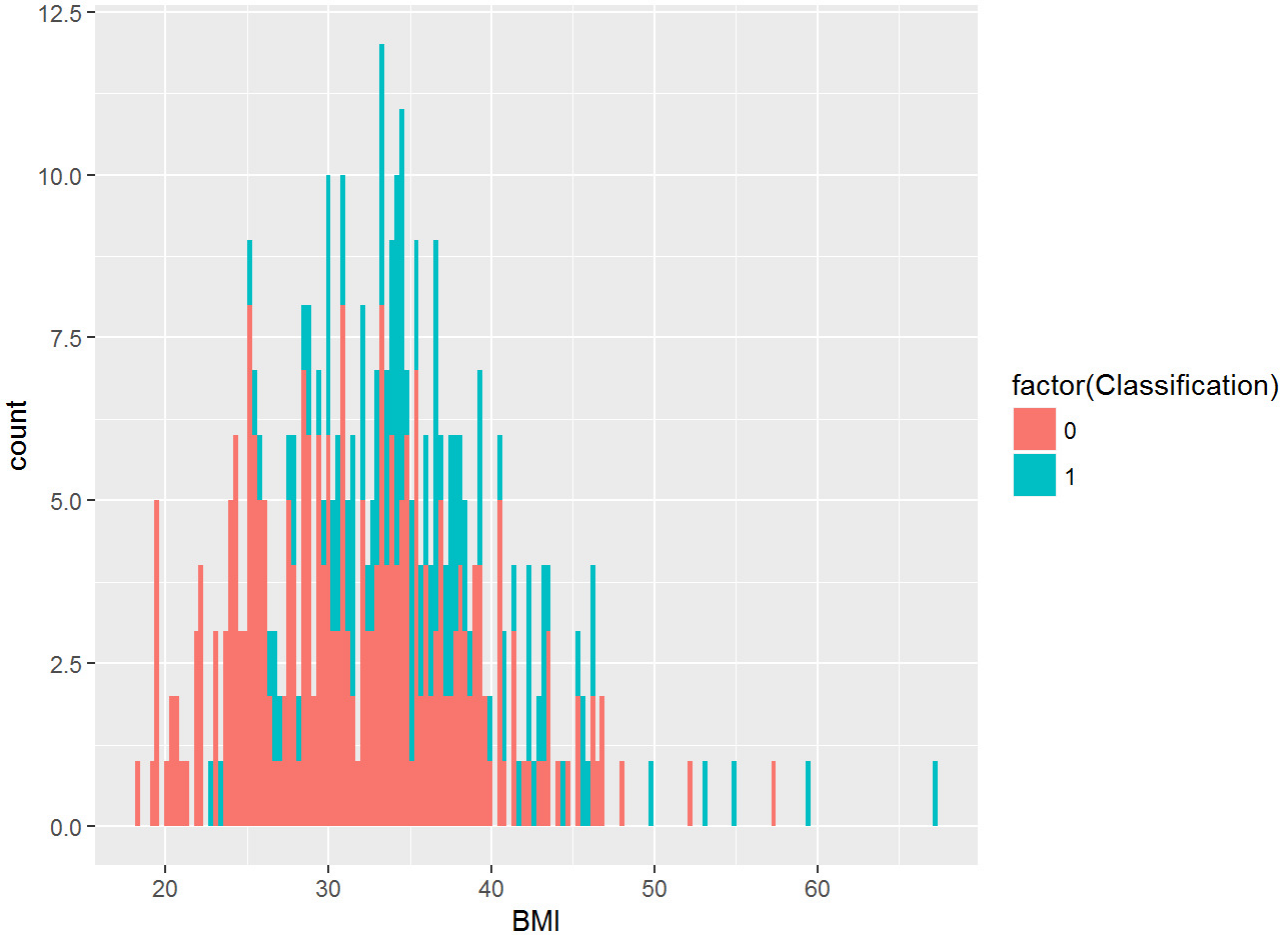

```
ggplot(pima.indian.diabetes2, aes(x=BMI,y=GlucoseLV,col= factor(Classification)))+geo
m_point()+geom_smooth(method="lm",se=FALSE)
```




```
ggplot(pima.indian.diabetes2,aes(x=BMI,y=GlucoseLV,Col=factor(Classification)))+geom_
point()+geom_smooth(method="lm",se=FALSE)
```




```
ggplot(pima.indian.diabetes2,aes(x=BMI,fill=factor(Classification)))+geom_histogram(b
inwidth = 0.3)
```




```
summary(pima.indian.diabetes2)
```

```
##  TimesPregnant      GlucoseLV      DiastolicBP     TricepsThickness
##  Min.   : 0.000   Min.   : 56.0   Min.   : 24.00   Min.   : 7.00
##  1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 62.00   1st Qu.:21.00
##  Median : 2.000   Median :119.0   Median : 70.00   Median :29.00
##  Mean   : 3.301   Mean   :122.6   Mean   : 70.66   Mean   :29.15
##  3rd Qu.: 5.000   3rd Qu.:143.0   3rd Qu.: 78.00   3rd Qu.:37.00
##  Max.   :17.000   Max.   :198.0   Max.   :110.00   Max.   :63.00
##    SerumInsulin         BMI        Heridarymarkup       Age
##  Min.   : 14.00   Min.   :18.20   Min.   :0.0850   Min.   :21.00
##  1st Qu.: 76.75   1st Qu.:28.40   1st Qu.:0.2697   1st Qu.:23.00
##  Median :125.50   Median :33.20   Median :0.4495   Median :27.00
##  Mean   :156.06   Mean   :33.09   Mean   :0.5230   Mean   :30.86
##  3rd Qu.:190.00   3rd Qu.:37.10   3rd Qu.:0.6870   3rd Qu.:36.00
##  Max.   :846.00   Max.   :67.10   Max.   :2.4200   Max.   :81.00
##  Classification
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3316
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.2.5
```

```
##
## Attaching package: 'dplyr'
```

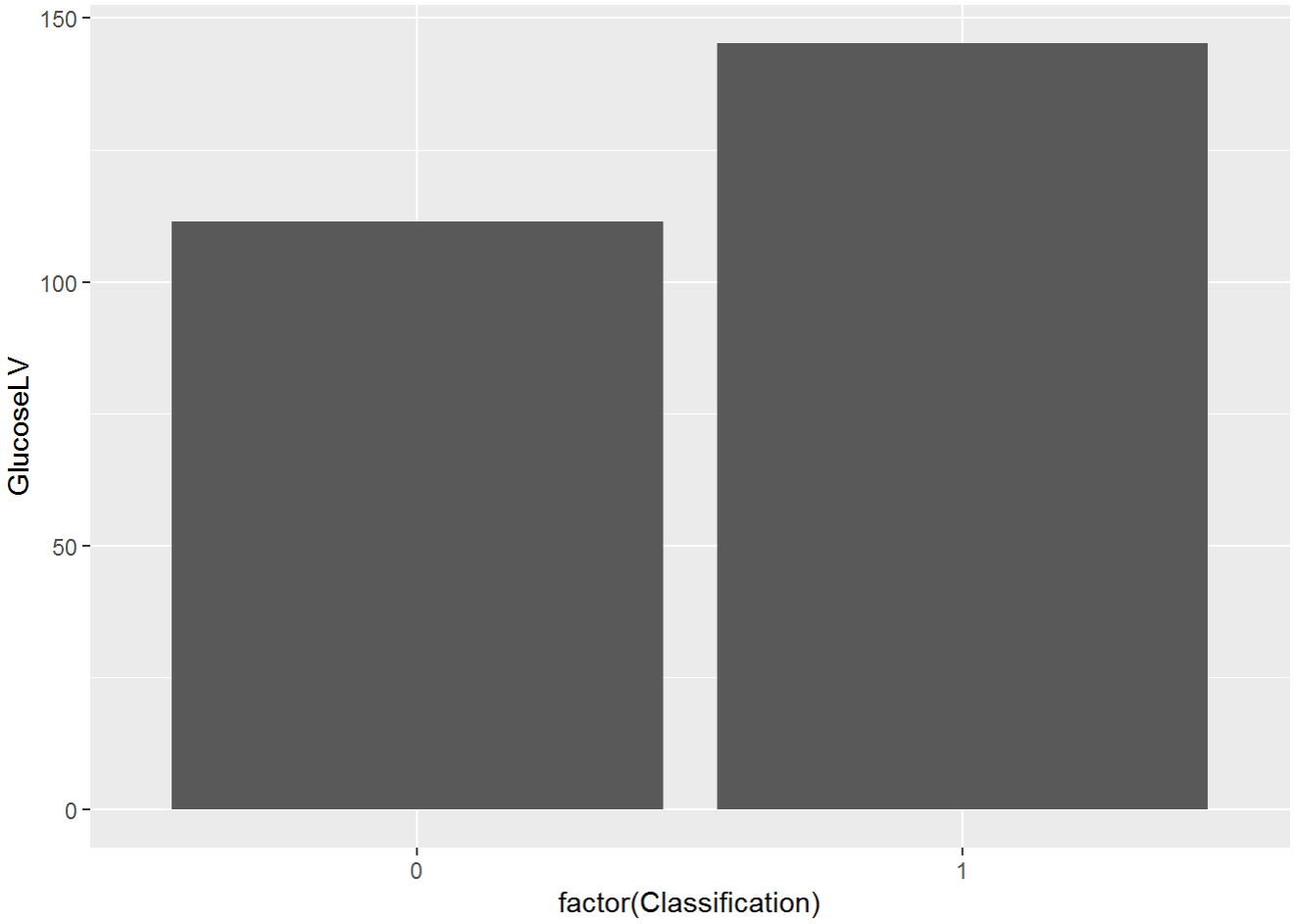
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

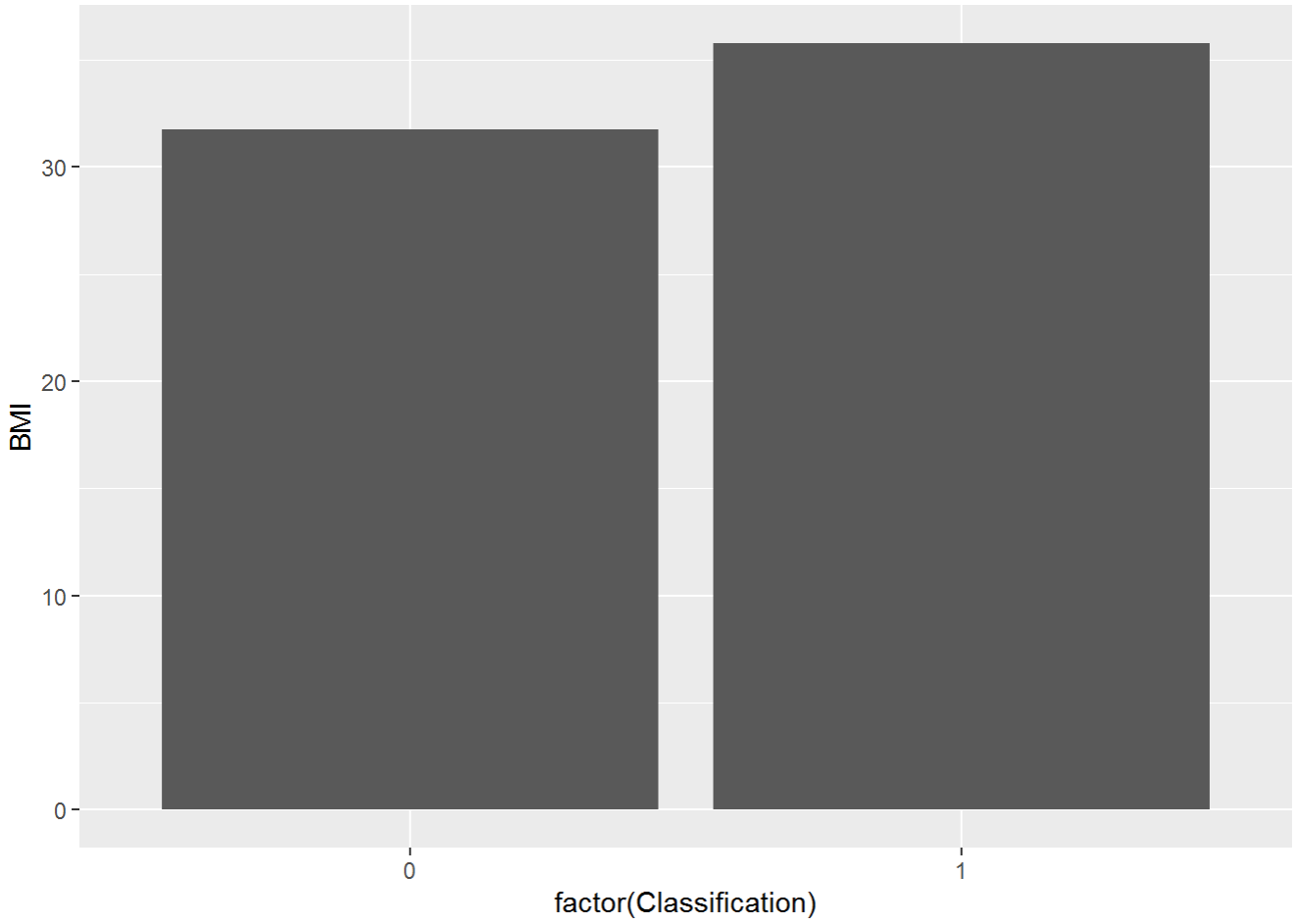
```
pima.indian.diabetes3<-pima.indian.diabetes2 %>% group_by(Classification)%>% summaris
e_each(funs(mean),GlucoseLV,DiastolicBP,SerumInsulin,BMI,TricepsThickness,Age)
View(pima.indian.diabetes3)
str(pima.indian.diabetes3)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    2 obs. of  7 variables:
##  $ Classification  : int  0 1
##  $ GlucoseLV       : num  111 145
##  $ DiastolicBP     : num  69 74.1
##  $ SerumInsulin    : num  131 207
##  $ BMI             : num  31.8 35.8
##  $ TricepsThickness: num  27.3 33
##  $ Age             : num  28.3 35.9
##  - attr(*, "na.action")=Class 'omit'  Named int [1:376] 1 2 3 6 8 10 11 12 13 16 .
..
##   .. ..- attr(*, "names")= chr [1:376] "1" "2" "3" "6" ...
```

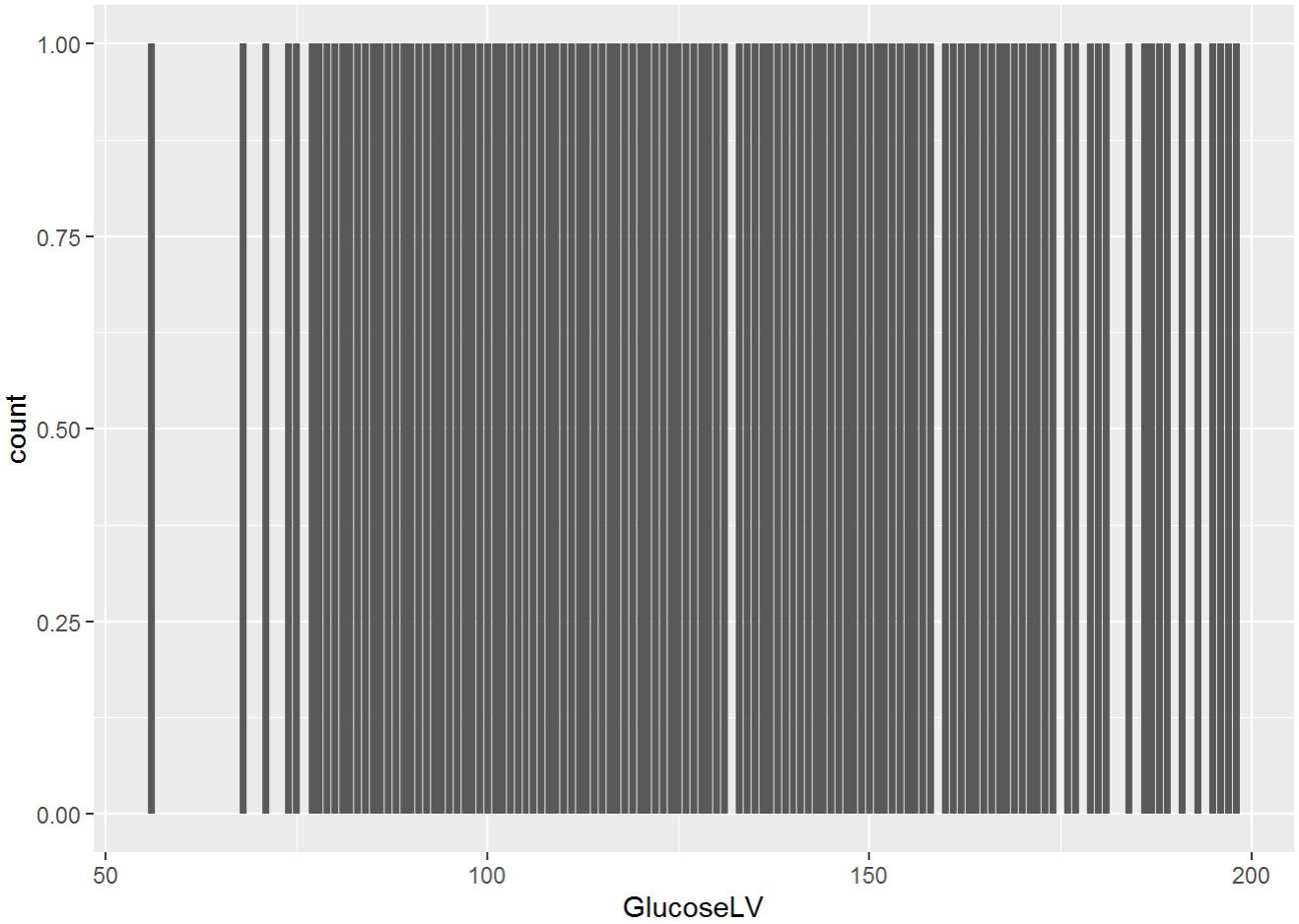
```
ggplot(pima.indian.diabetes3,aes(x=factor(Classification),y=GlucoseLV))+geom_bar(stat
="identity")
```




```
ggplot(pima.indian.diabetes3,aes(x=factor(Classification),y=BMI))+geom_bar(stat="iden
tity")
```




```
ggplot(pima.indian.diabetes2,aes(x=GlucoseLV,fill=Classification))+geom_bar(position
="fill")
```


1.00
0.75
0.50
0.25
0.00
count
50
100
150
200
GlucoseLV