HW2

Prisma Lopez

2/5/2019

```r
# time consuming method:
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------- tidyverse

## v ggplot2 3.2.1      v purrr   0.3.2
## v tibble  2.1.3      v dplyr   0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------------ tidyverse_confli
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
probes <- read_csv("HW2_data.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   prognosis = col_character()
## )

## See spec(...) for full column specifications.
```

```r
probesG = probes %>% filter(prognosis=='GOOD') %>% select(c(1:30))
probesP = probes %>% filter(prognosis=='POOR') %>% select(c(1:30))
meanG=apply(probesG,2,mean)
meanP=apply(probesP,2,mean)

diff=meanP-meanG
grouped=sort(diff,decreasing=TRUE) # sorted in decreasing order of mean value

write.csv(grouped, file = 'hw2p1q2b.csv')
f=read.csv('hw2p1q2b.csv')
colnames(f) <- c("probe","mean diff")

#f=f %>% mutate(uORd=`mean diff`<0)
f$Type<-NA
f[f$`mean diff` <0, ][, "Type"] <- "DOWN"
f[f$`mean diff` >0, ][, "Type"] <- "UP"
write.csv(f, file = 'hw2p1q2b.csv')

#install.packages("lattice")
library(lattice)
```

```r
# refer to:https://rcompanion.org/rcompanion/d_02.html
#bartlett.test(37086_at ~ prognosis,data=probes) # can't on multiple columns at a time


# For method refer to: https://rpubs.com/kaz_yos/1204
## Multiple t-tests for all probes vs. prognosis:
#probes$prognosis[probes$prognosis=='GOOD'] <- 2
#probes$prognosis[probes$prognosis=='POOR'] <- 1
#probes$prognosis <- as.numeric(probes$prognosis)
ts = lapply(probes[,1:30], function(x) t.test(x ~ probes$prognosis, var.equal = TRUE))

ts.t <- sapply(ts, '[[', 'statistic') # t-score
ts.p <- sapply(ts, '[[', 'p.value') # p-value

#tst.sort=sort(ts.t,decreasing=TRUE)
tsp.sort=sort(ts.p,decreasing=FALSE)


write.csv(tsp.sort, file = 'hw2p1q2c.csv')
p=read_csv("hw2p1q2c.csv")


## Warning: Missing column names filled in: 'X1' [1]


## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   x = col_double()
## )

colnames(p) = c("probe","pvalue")
write.csv(tsp.sort, file = 'hw2p1q2c.csv')

f4=p[1:5,1:2]
write.csv(f4, file = 'hw2p1q2cTOP.csv')

library(lattice)

# Top 5 probes:
#34683_at
histogram(~`34683_at` | prognosis,data=probes,layout=c(1,2))
```
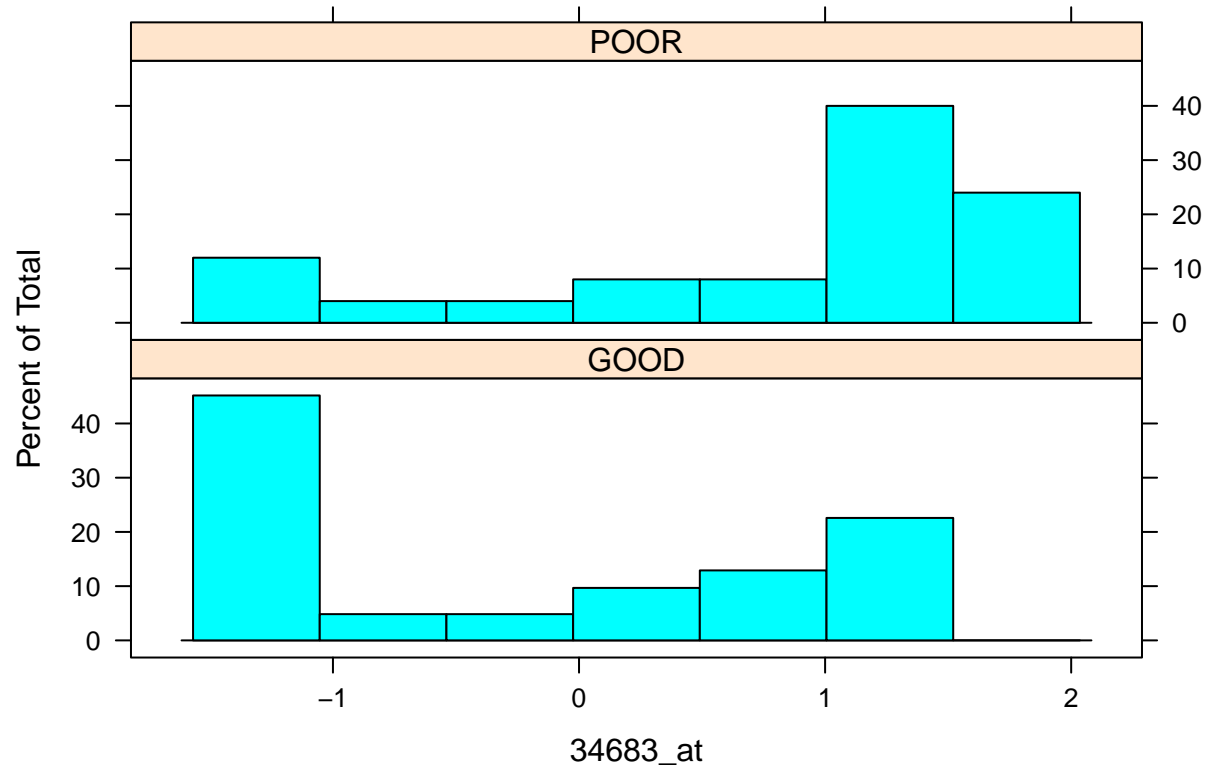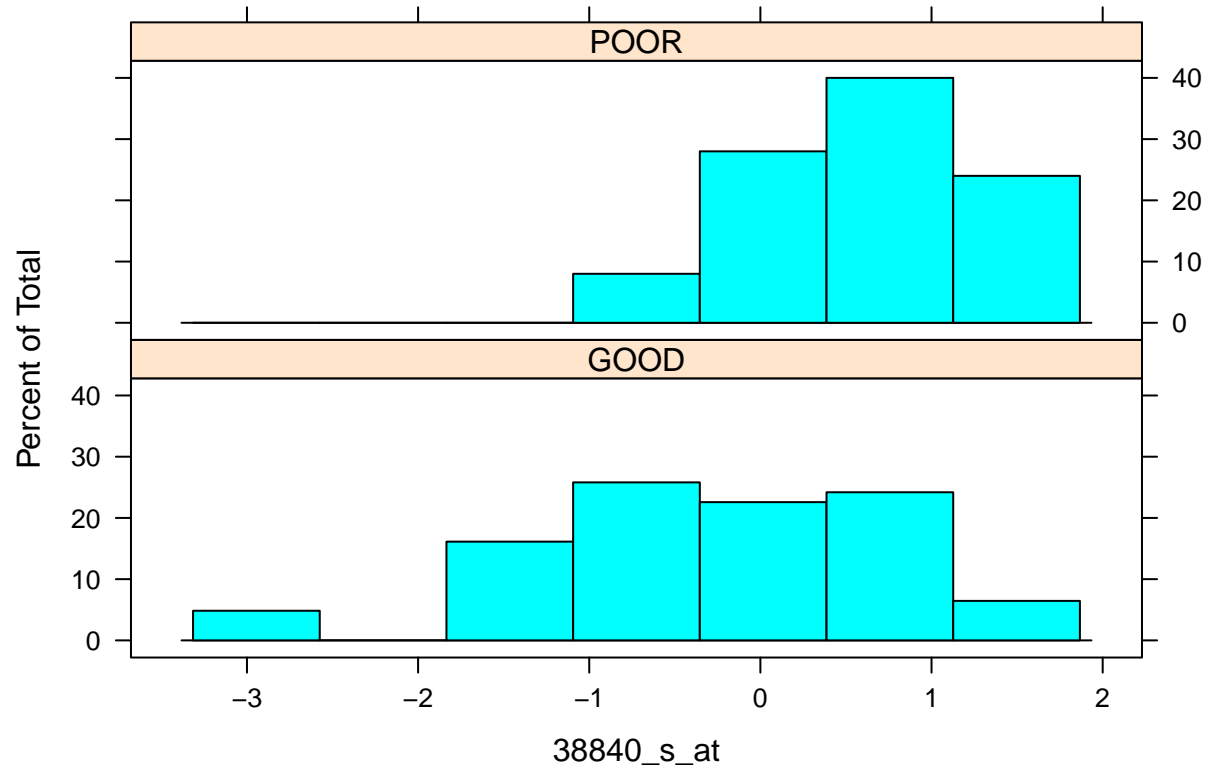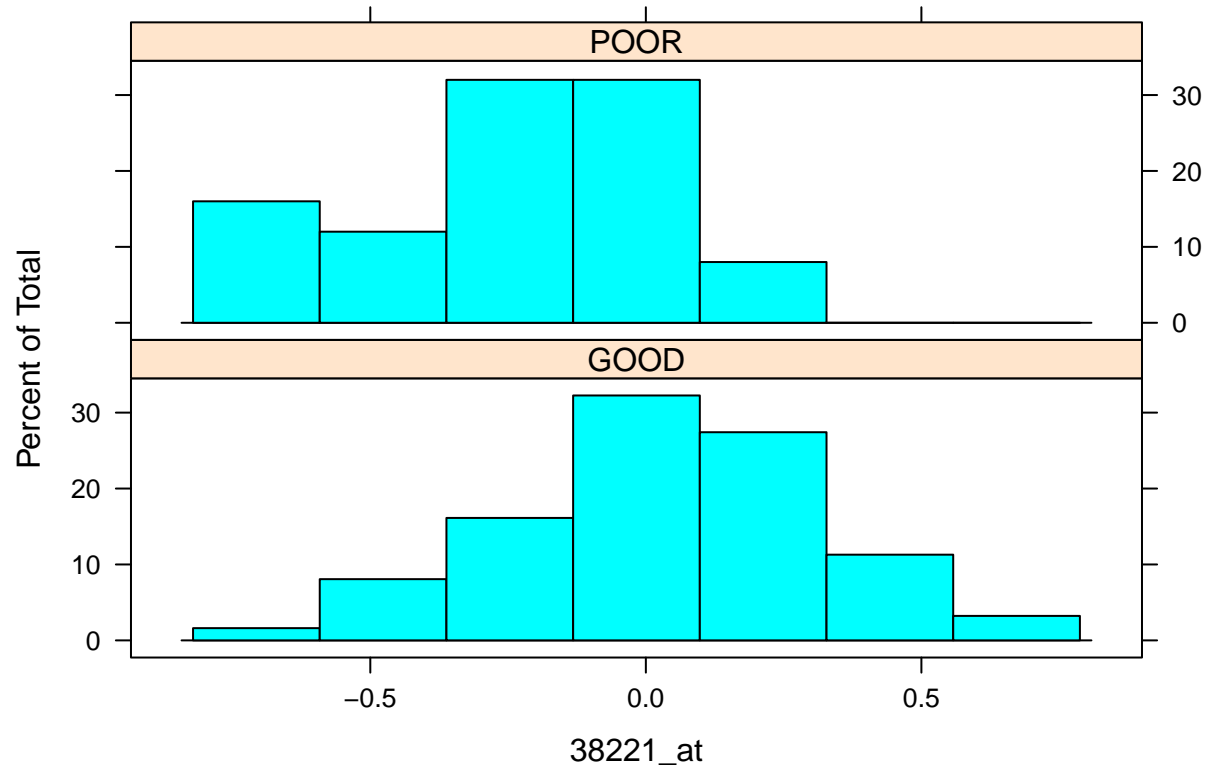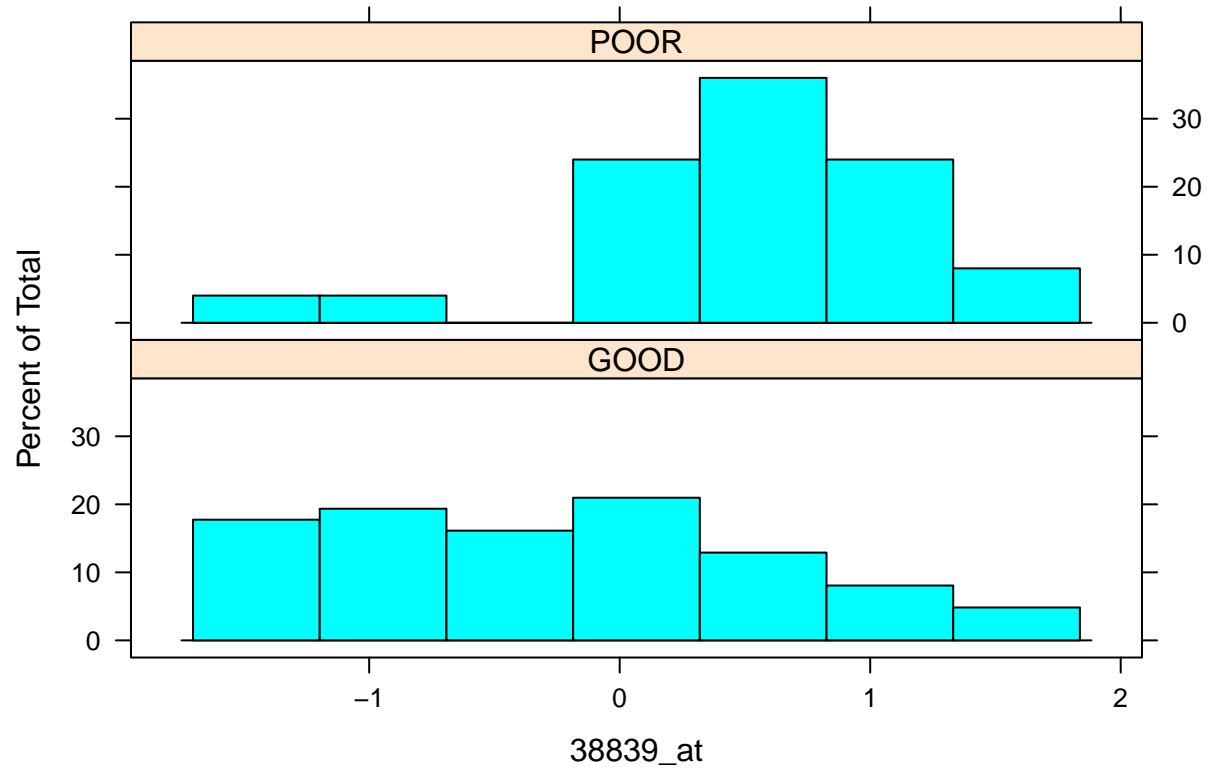
```r
#38840_s_at
histogram(~`38840_s_at` | prognosis,data=probes,layout=c(1,2))
```

```
#38221_at
histogram(~`38221_at` | prognosis,data=probes,layout=c(1,2))
```

```
#38839_at
histogram(~`38839_at` | prognosis,data=probes,layout=c(1,2))
```

```
#32607_at
histogram(~`32607_at` | prognosis,data=probes,layout=c(1,2))
```