finding0411_JM

```
setwd("/Users/jingrumu/Documents/Spring2018/DataMiningCup2018/raw_data")
library(data.table)
items=fread("items.csv",header=T)
train=fread("train.csv",header=T)
prices=fread("prices.csv",header=T)
```

Date

Treat date numerically, set 2017-10-01 as the starting date.

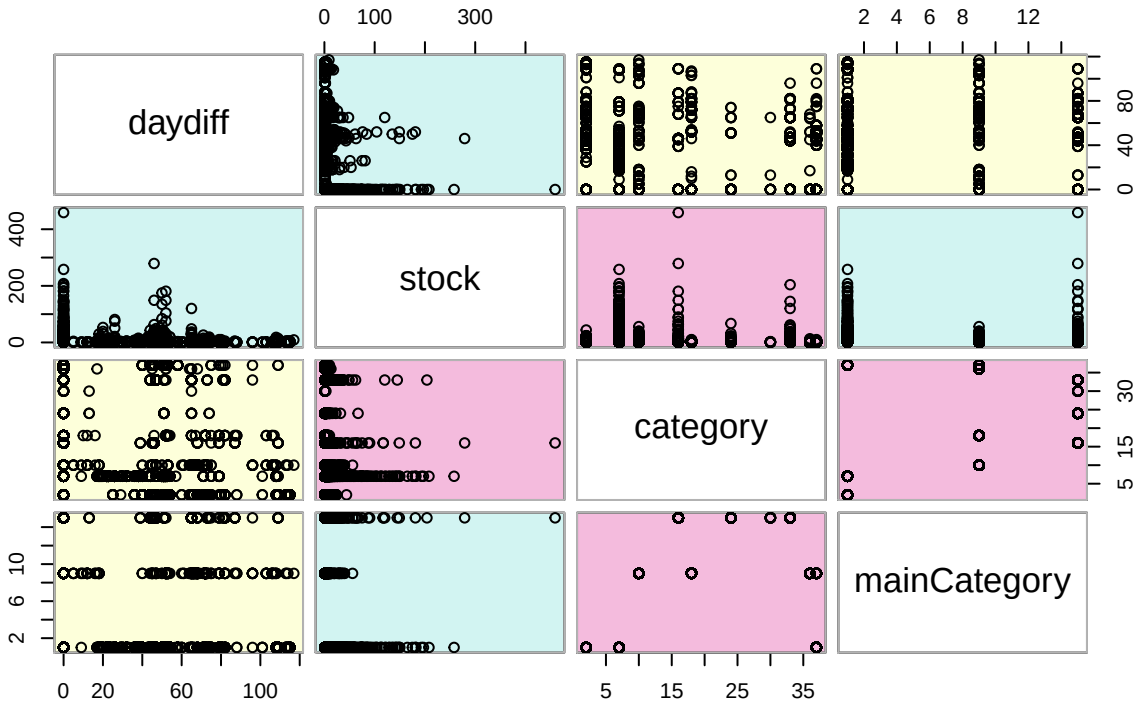
```
library(stringr)
date=str_replace(items$releaseDate, "-", "")
date=as.numeric(str_replace(date, "-", ""))
items$date=date
day.diff=as.Date(items$releaseDate, format="%Y-%m-%d")-
  as.Date("2017-10-01", format="%Y-%m-%d")
items$daydiff=as.numeric(day.diff)
```

Relationship with others

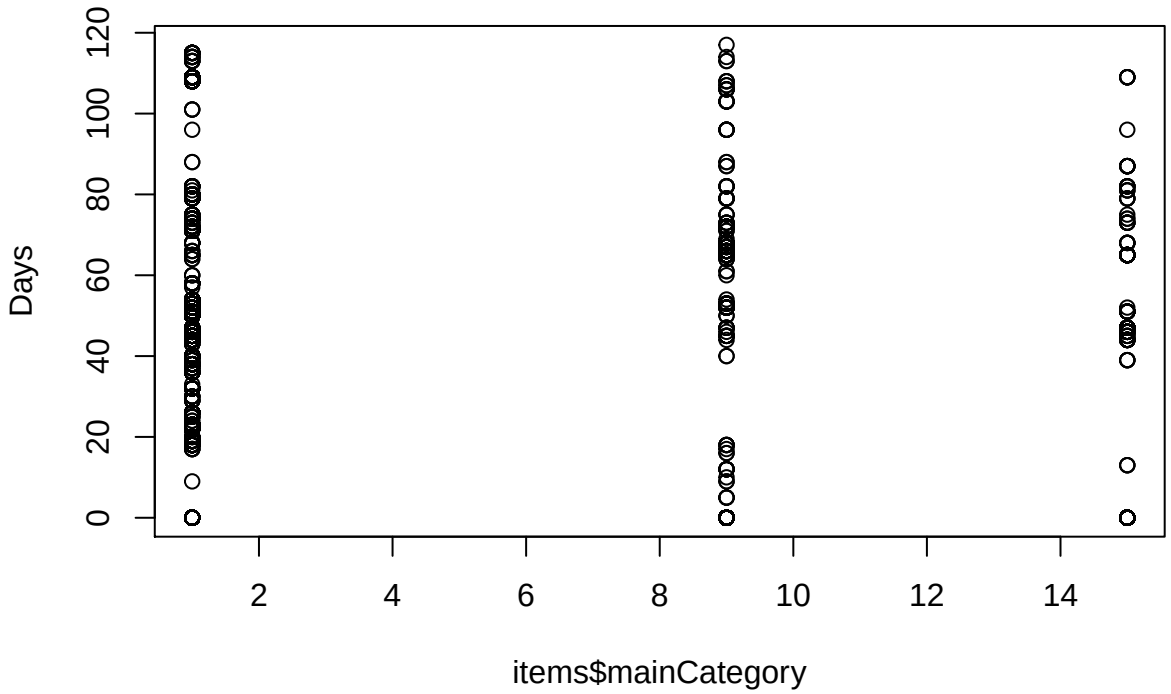
```
library(gclus)
```

```
## Loading required package: cluster
```

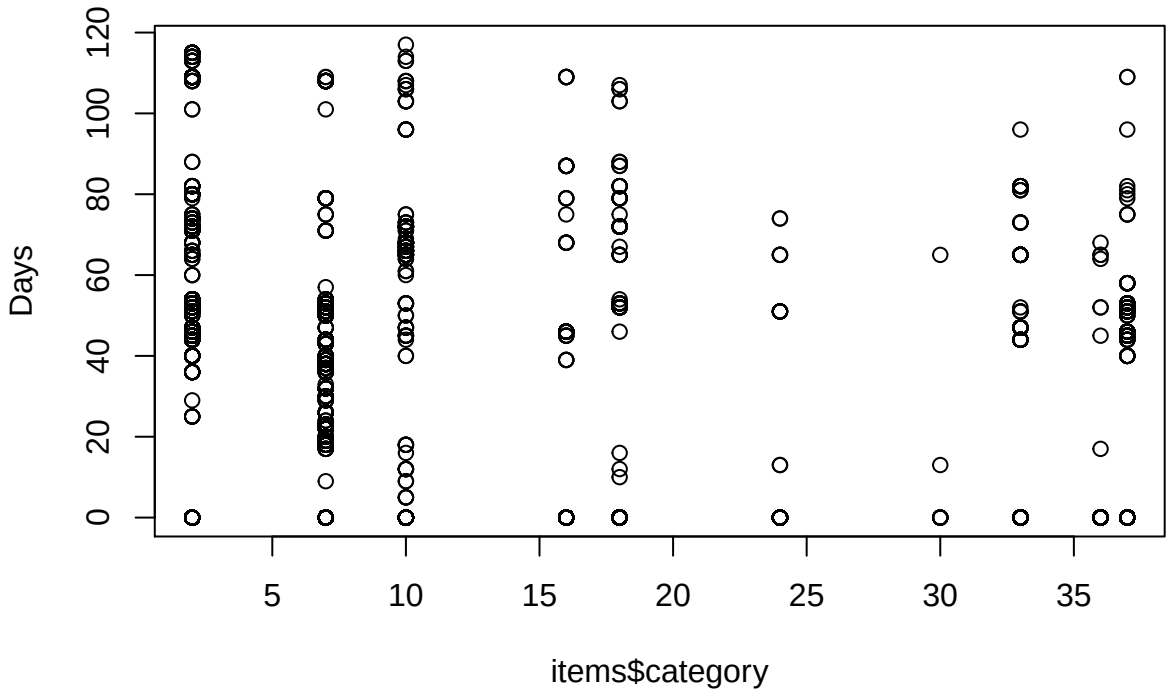
```
dta <- items[,c(6,7,9,12)] # get data
dta.r <- abs(cor(dta)) # get correlations
dta.col <- dmat.color(dta.r) # get colors
# reorder variables so those with highest correlation
# are closest to the diagonal
dta.o <- order.single(dta.r)
cpairs(dta, dta.o, panel.colors=dta.col, gap=.5,
main="Variables Ordered and Colored by Correlation" )
```




```
# plot with MainCategory
plot(items$mainCategory,day.diff,ylab="Days")
```




```
# plot with category
plot(items$category,day.diff,ylab="Days")
```




```
# plot with stock
plot(items$stock, day.diff, ylab="Days")
```