Outlier test

Joan Verdú

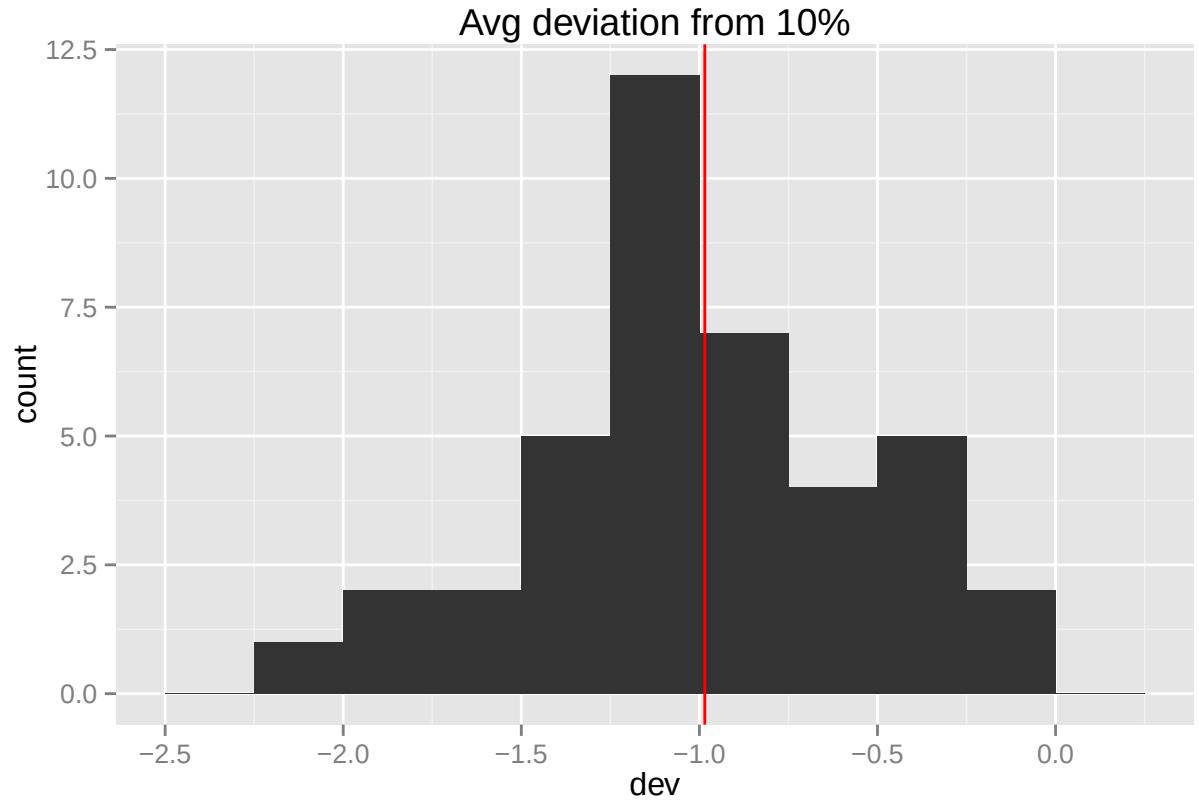
Monday, April 20, 2015

Exploration from 'heavytails_output.csv' file, considering outlier as abs(return/volatility)>1.645, thus, 10% probability tail.

```
# libraries
library(ggplot2)

# Load data
data<-read.csv('heavytails_output.csv')
data<-data[,c('net_id','vertex_id','return','volatility')]

# Extract day
data$date<-substr(data$net_id,12,21)
data$date<-as.Date(data$date,"%Y-%m-%d")

# Outlier classification
data$outlier<-abs(data$return/data$volatility)>1.645

# Total tickers
n_tickers<-length(unique(data$vertex_id))
# Days by ticker
n_days<-aggregate(date~vertex_id,data=data,function(x) length(unique(x)))
# Outliers by ticker
out<-aggregate(outlier~vertex_id,data=data,sum)
# Percent outliers by ticker
out$percent<-out$outlier/n_days$date
# Deviation from theoretical 10%
out$dev<-100*(out$percent-0.1)

summary(out$dev)
```

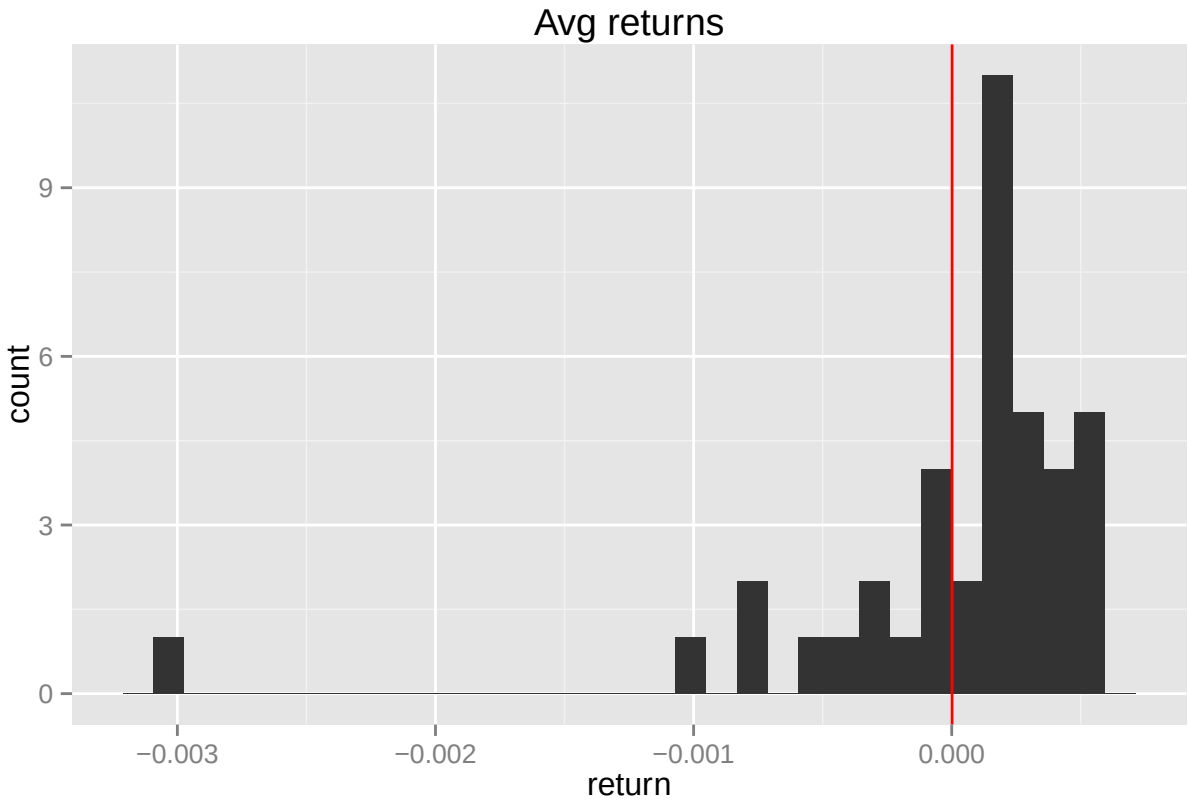
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.2400 -1.2000 -1.0800 -0.9850 -0.6890 -0.0097
```

```
qplot(main="Avg deviation from 10%",dev, data=out, geom="histogram",binwidth=0.25)+
    geom_vline(colour="red",xintercept = mean(out$dev))
```




```
# Mean returns by ticker
ret<-aggregate(return~vertex_id,data=data,mean)
qplot(main="Avg returns",return, data=ret,geom="histogram")+
    geom_vline(colour="red",xintercept = mean(ret$return))
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```


Avg returns
count
9
6
3
0
−0.003
−0.002
−0.001
0.000
return