

Дескриптивна статистика

З курс, статистика, Шкляр Ірина Володимирівна

Завдання 4, варіант 9

1.

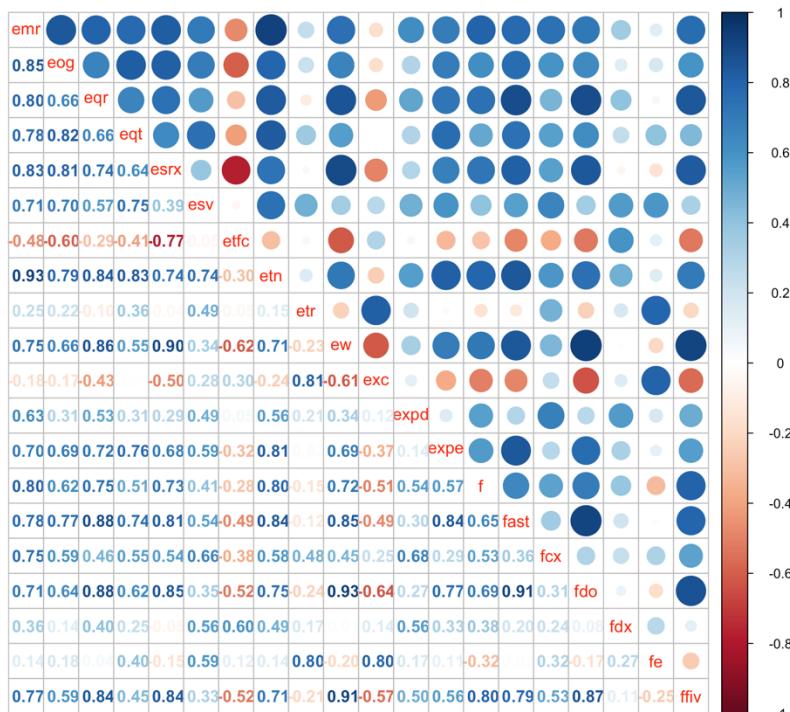
```
> filenames=list.files(path="/Users/irynashkliar/Downloads/tables", full.names=TRUE)
>
> datalist = lapply(filenames,function(x){x0<-read.csv(file=x,header=F)[,c(1,6)];
colnames(x0)<-c("data", unlist(strsplit(x,"[.]"))[2]);x0})
>
> y<-Reduce(function(x,y) {merge(x,y,by="data")}, datalist)
```

2.

```
> M <- cor(y[,-1])
>
> install.packages('corrplot')
>
> library(corrplot)
```

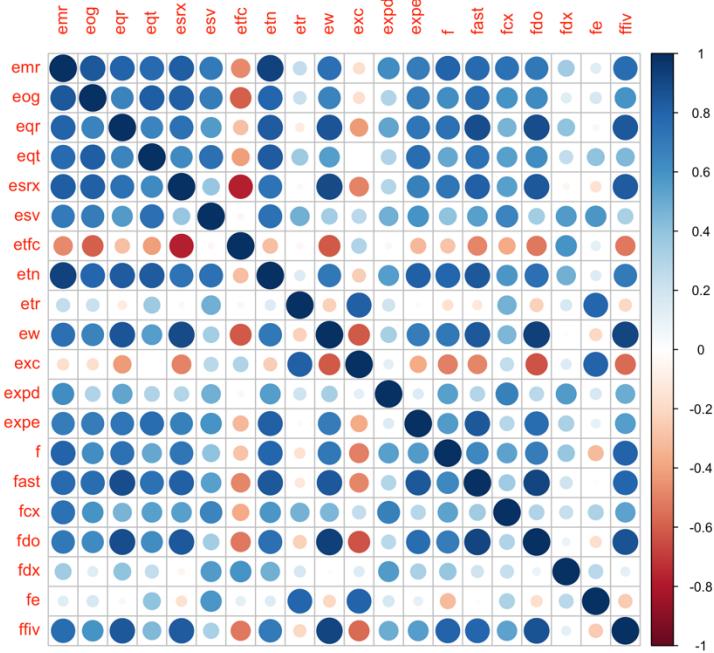
Числовий формат:

```
> corrplot.mixed(M)
```



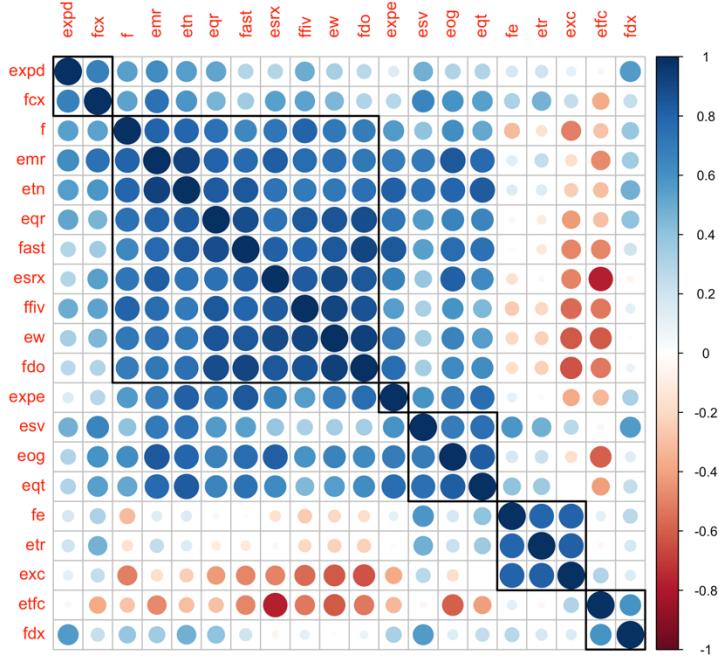
Карта кореляцій:

> corrplot(M)



3.

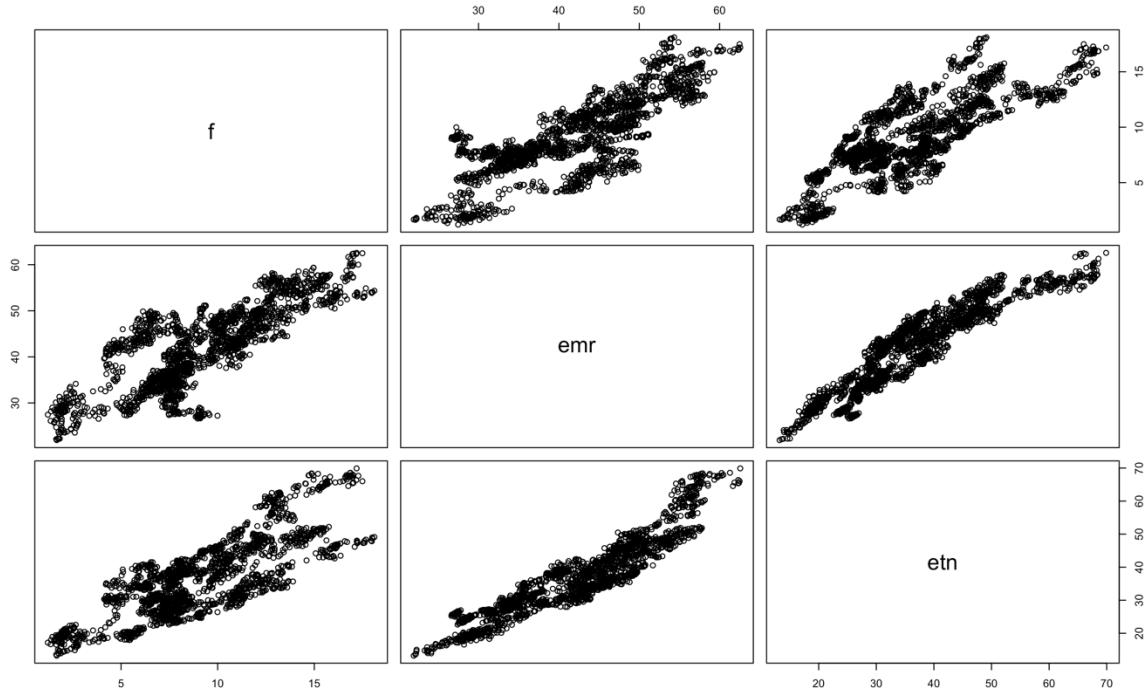
> corrplot(M,order="hclust",addrect=6)



4.

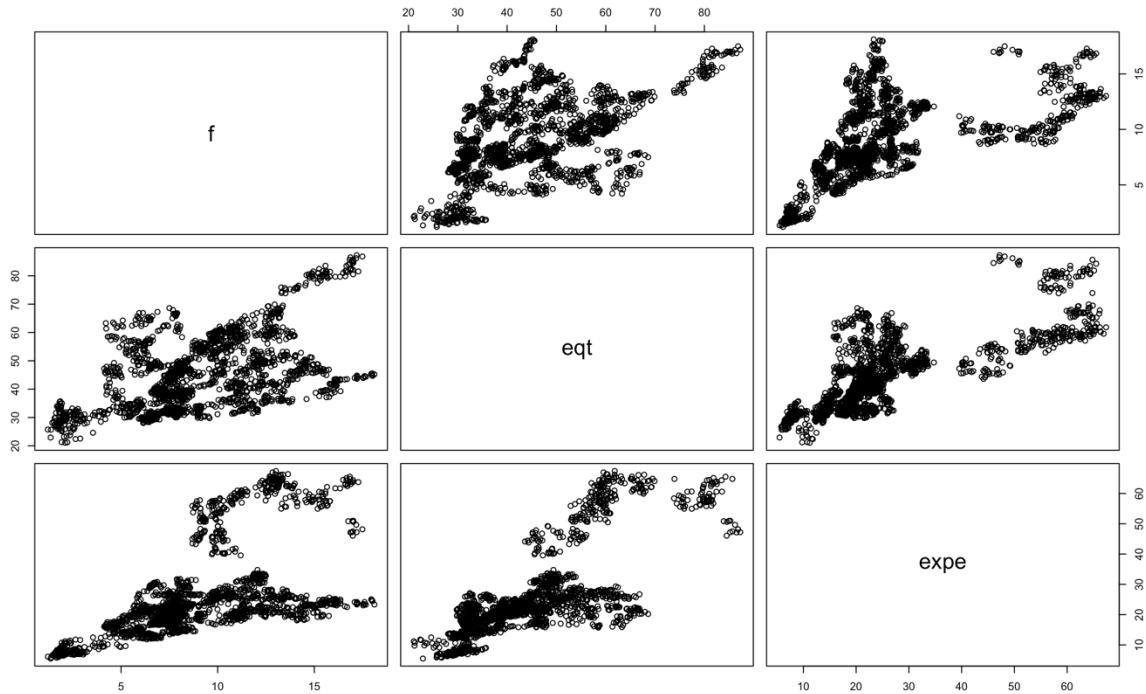
1)

```
> pairs(y[,c("f","emr","etn")]) ----- one cluster
```



2)

```
> pairs(y[,c("f","eqt","expe")]) ----- three clusters
```



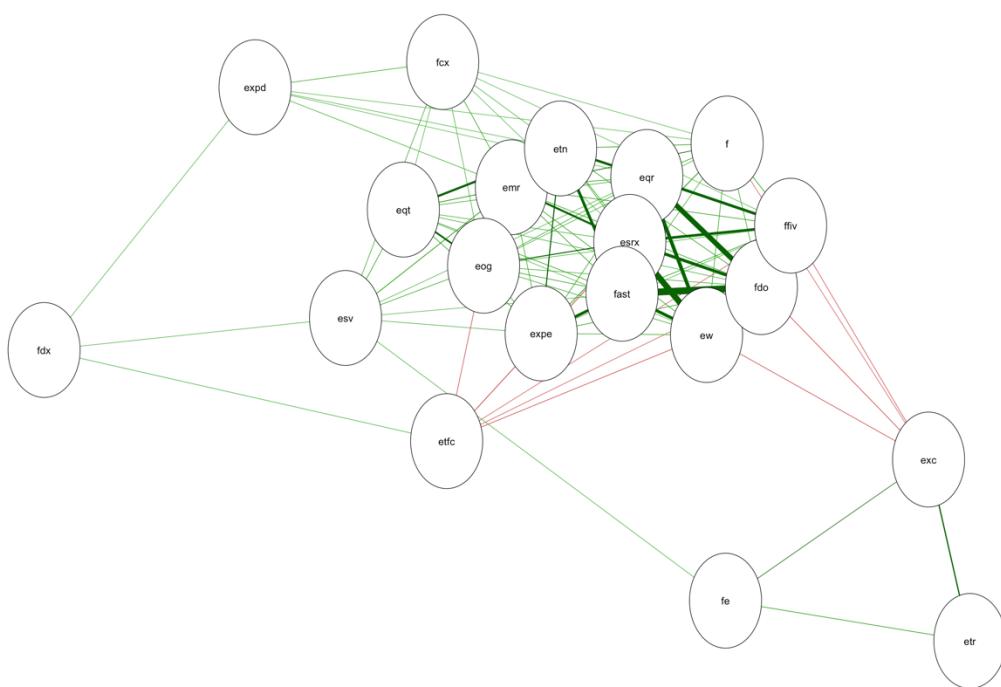
Тут доцільно використовувати кореляції Пірсона, оскільки видно залежності між фірмами (в обох випадках).

5.

```
> install.packages("vctrs")
>
> install.packages("ggplot2")
>
> install.packages("qgraph")
>
> library(qgraph)
```

На мою думку, на цьому варіанті найкраще видно залежність:

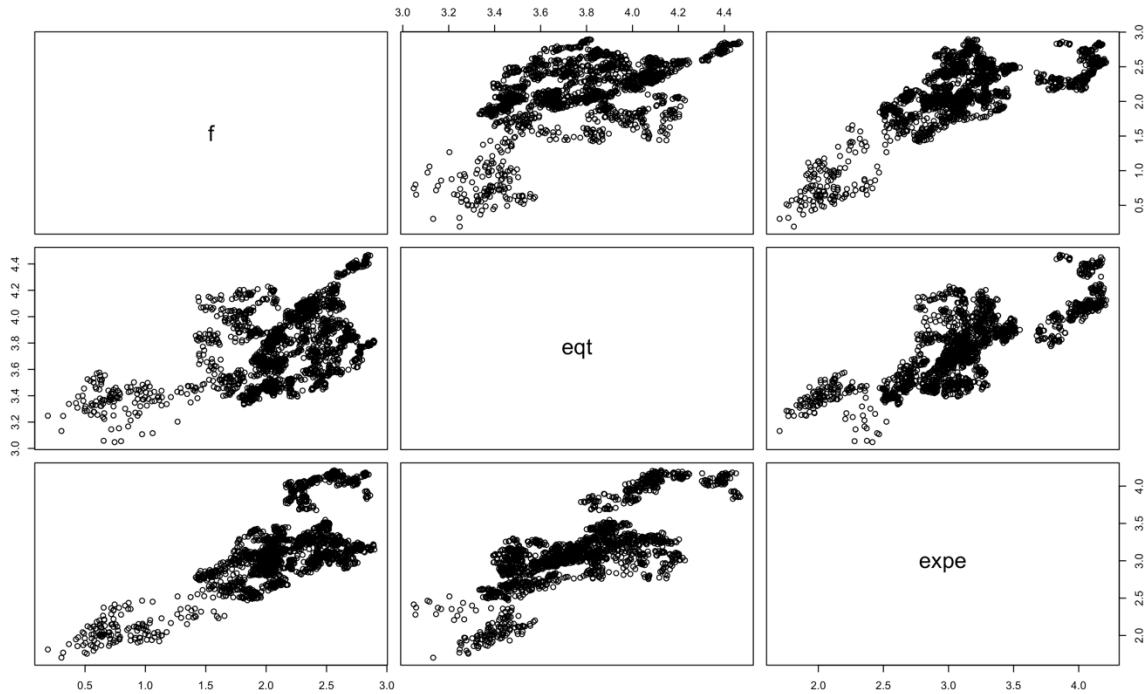
```
> qgraph(M,layout="spring",threshold=0.5,labels=colnames(M))
```



6.

Я вибрала три фірми з різних кластерів, які використовувались в 4 пункті.

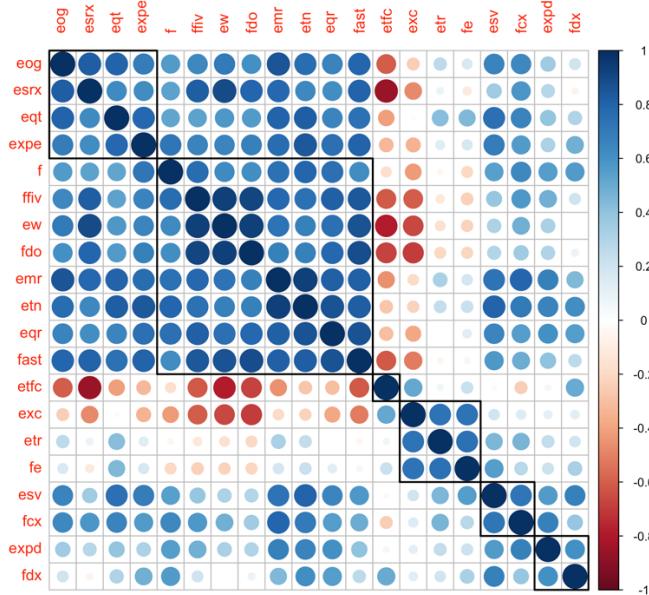
```
> H <- y[,-1]
>
> W <- log(H)
>
> pairs(W[,c("f","eqt","expe")])
```



Я вважаю, що більш доцільно використовувати кореляції Пірсона для цін акцій, тому що краще видно залежність.

Зайдемо кластери для log за допомогою кореляції Пірсона:

```
> W01 <- cor(W,method = "pearson")
>
> corrplot(W01,order="hclust",addrect=6)
```

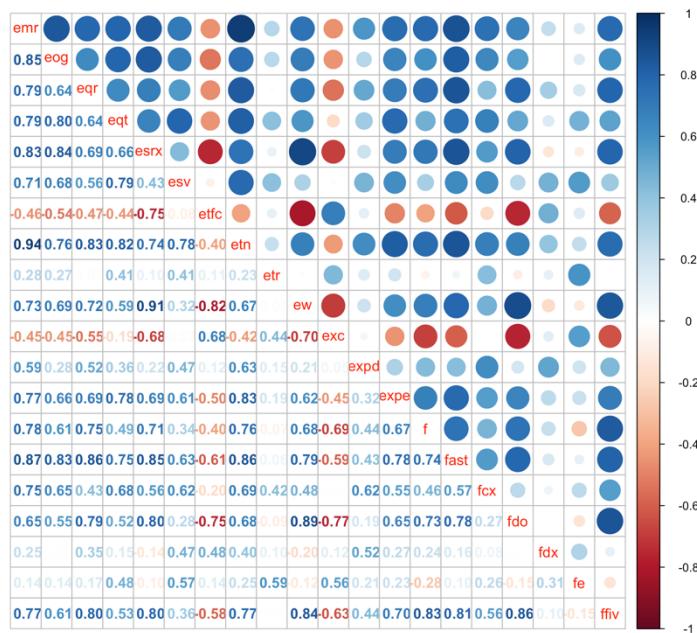


Дуже схожим виявився великий кластер, майже всі фірми в ньому зійшлися з фірмами в кластері в третьому пункті.

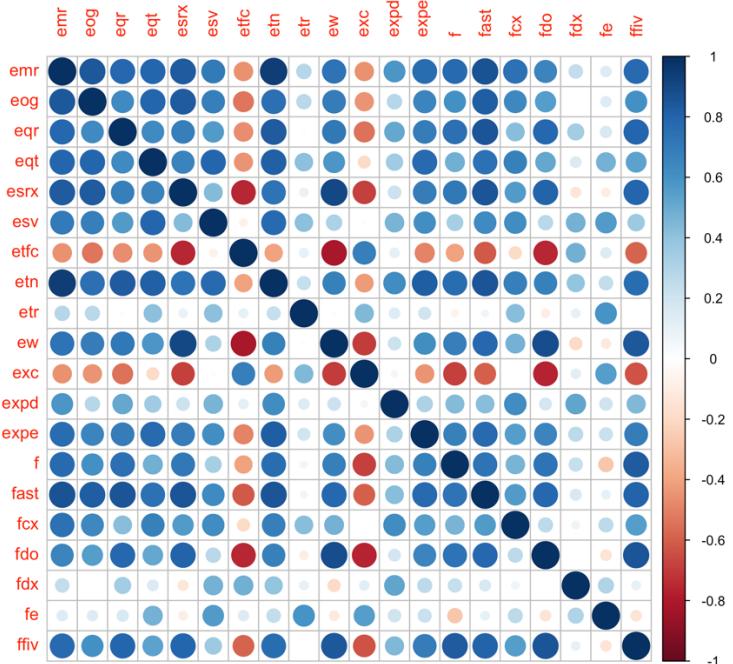
7.

2)

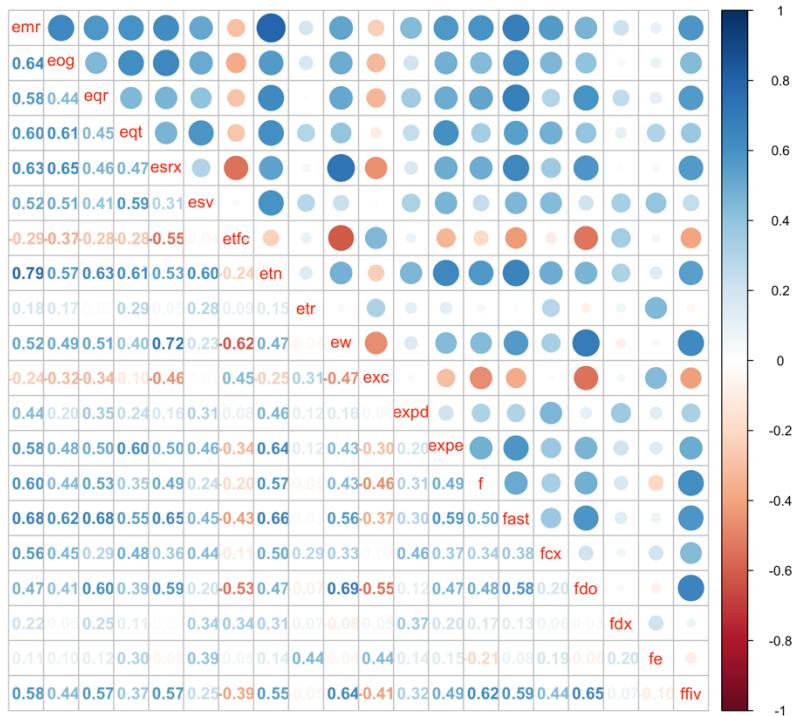
```
> M <- cor(y[,-1],method = "spearman")
>
> K <- cor(y[,-1],method = "kendall")
>
> corrplot.mixed(M)
```



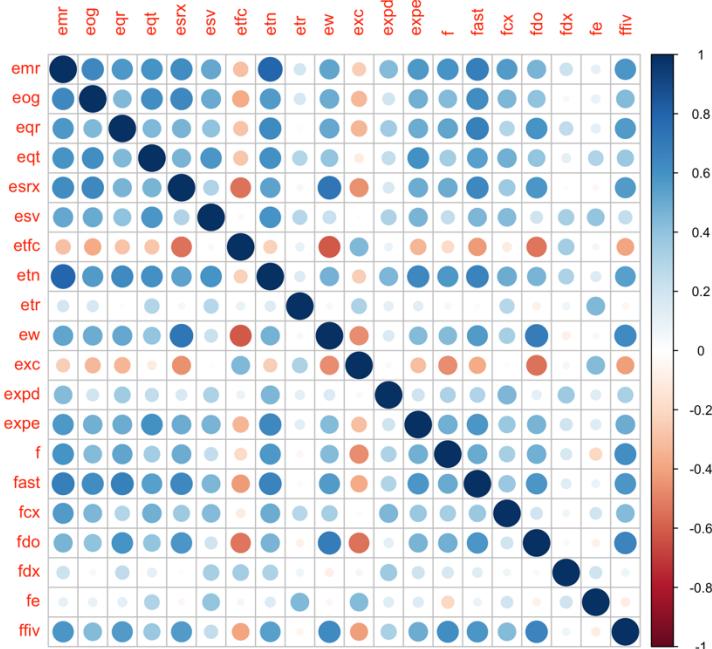
```
> corrplot(M)
```



```
> corrplot.mixed(K)
```

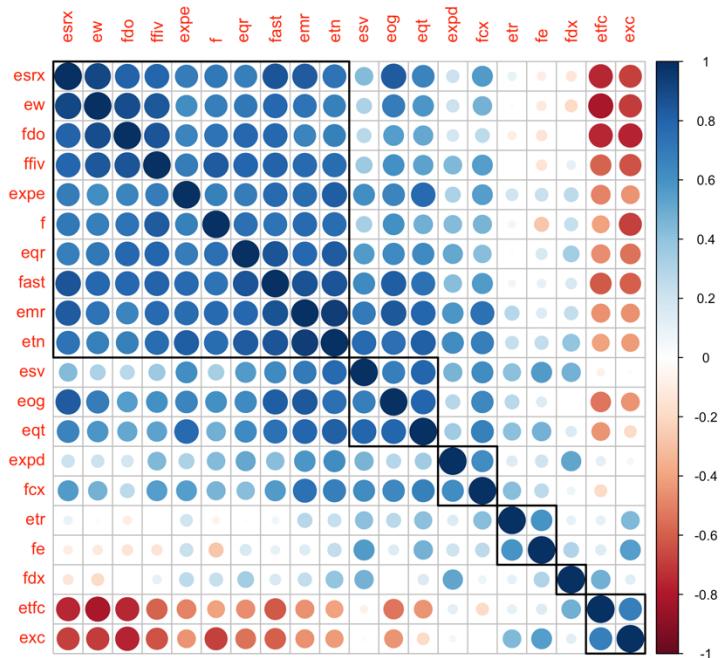


> corrplot(K)

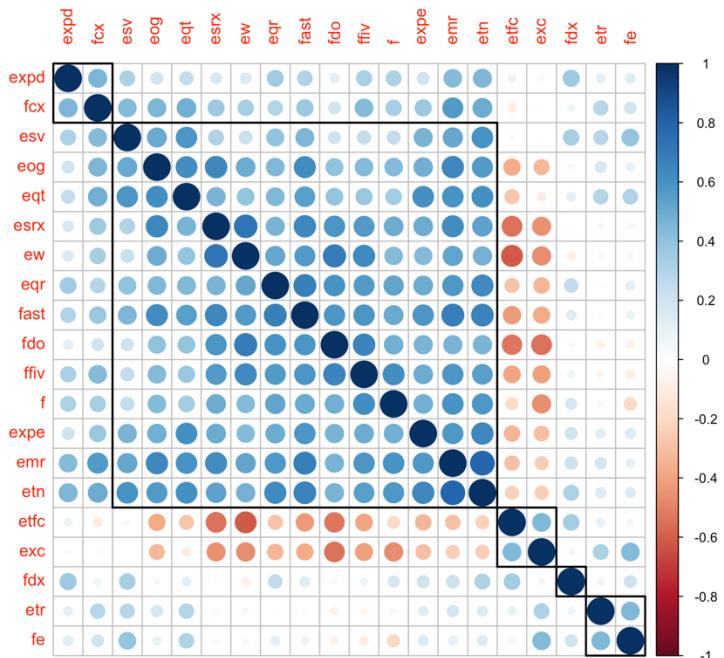


3)

> corrplot(M,order="hclust",addrect=6)

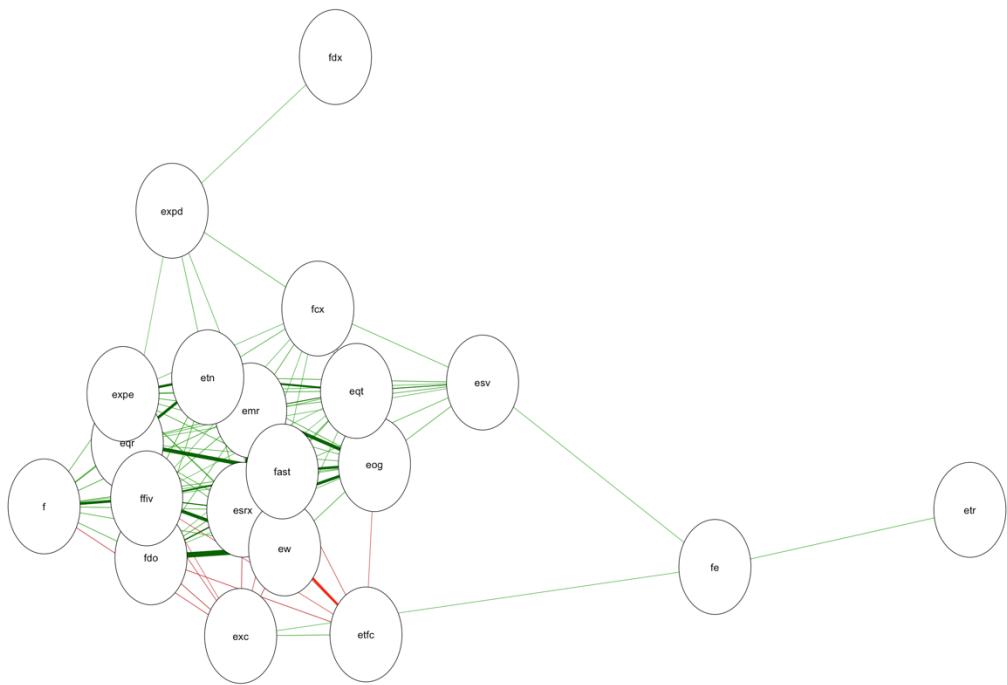


```
> corrplot(K,order="hclust",addrect=5)
```

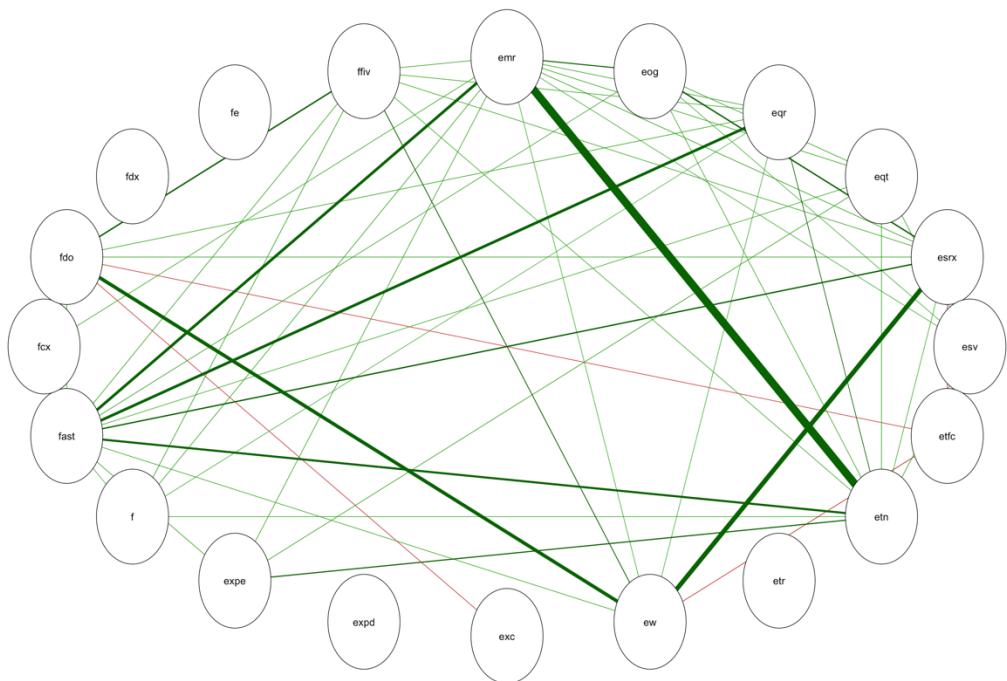


5)

```
> qgraph(M,layout="spring",threshold=0.5,labels=colnames(M))
```



```
> qgraph(K,threshold=0.5,labels=colnames(K))
```



Найбільш адекватними для опису залежностей будуть кореляції Пірсона, тому що вони рахуються саме по цінам акцій (коєфіцієнти кореляції Спірмана рахуються за рангами, а Кендалла – за кількістю узгоджених пар).