

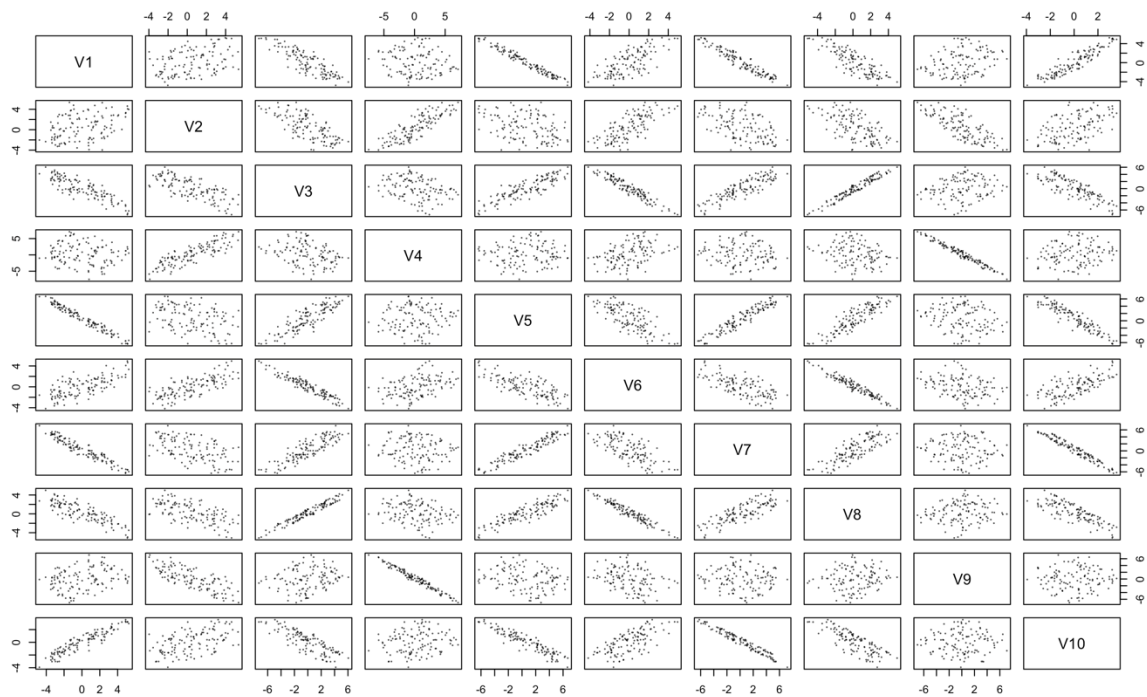
Дескриптивна статистика

3 курс, статистика, Шкляр Ірина Володимирівна

Завдання 5, варіант 9

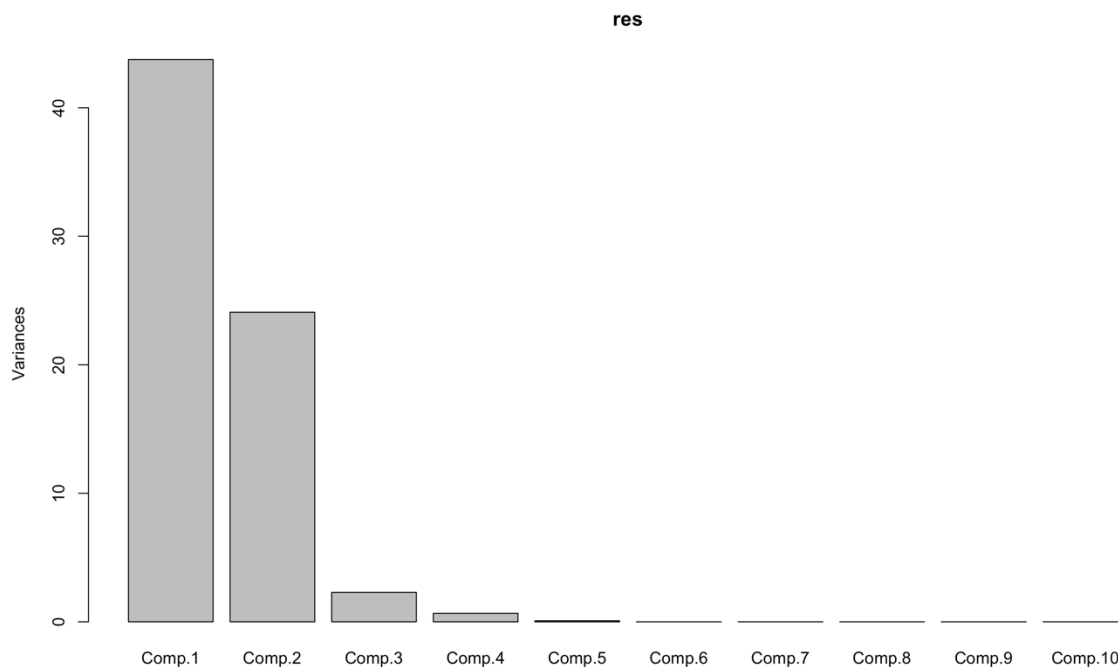
```
> z<-read.table("/Users/irynashkliar/Downloads/multi/F9t.txt",header=F)
>
> pairs(z,cex=0.1)
```

Маємо діаграму розсіювання пар:



Виразної геометричної структури даних не помітно, тому використаємо метод головних компонент:

```
> res<-princomp(z)
>
> plot(res)
```

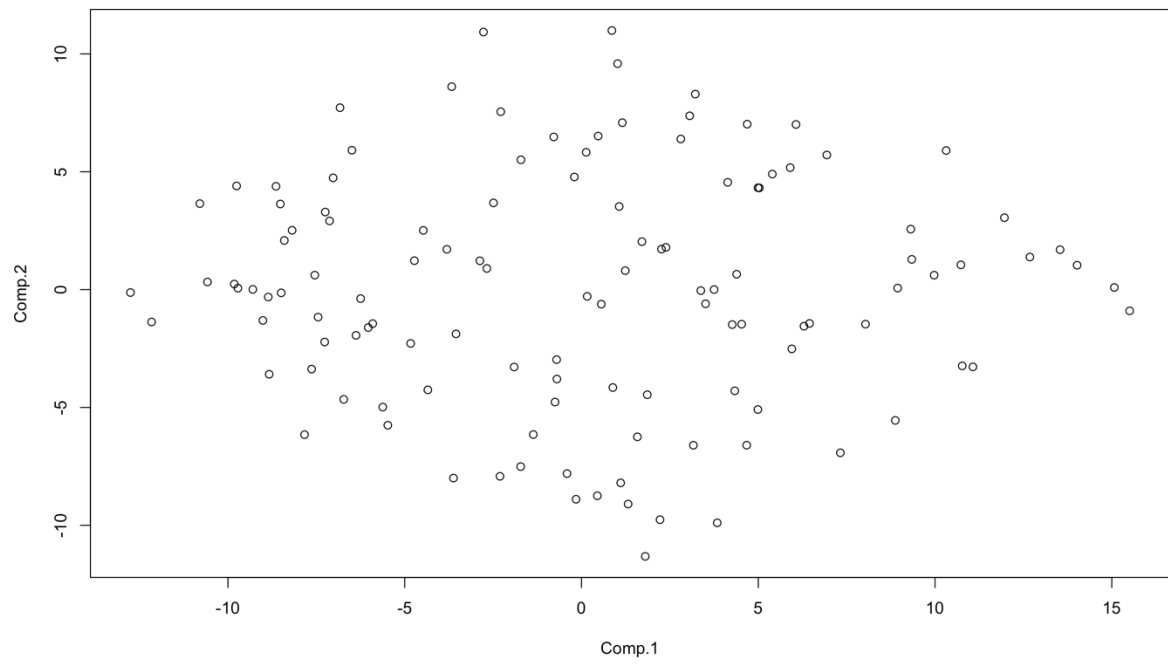


```
> summary(res)
Importance of components:
              Comp.1      Comp.2      Comp.3      Comp.4
Standard deviation  6.6149146  4.9082314  1.51243113  0.811654809
Proportion of Variance 0.6174019  0.3399144  0.03227533  0.009295274
Cumulative Proportion 0.6174019  0.9573163  0.98959159  0.998886866
              Comp.5      Comp.6      Comp.7 Comp.8 Comp.9
Standard deviation  0.280875609  5.038853e-08  4.620431e-09      0      0
Proportion of Variance 0.001113134  3.582472e-17  3.012205e-19      0      0
Cumulative Proportion 1.000000000  1.000000e+00  1.000000e+00      1      1
              Comp.10
Standard deviation      0
Proportion of Variance      0
Cumulative Proportion      1
>
```

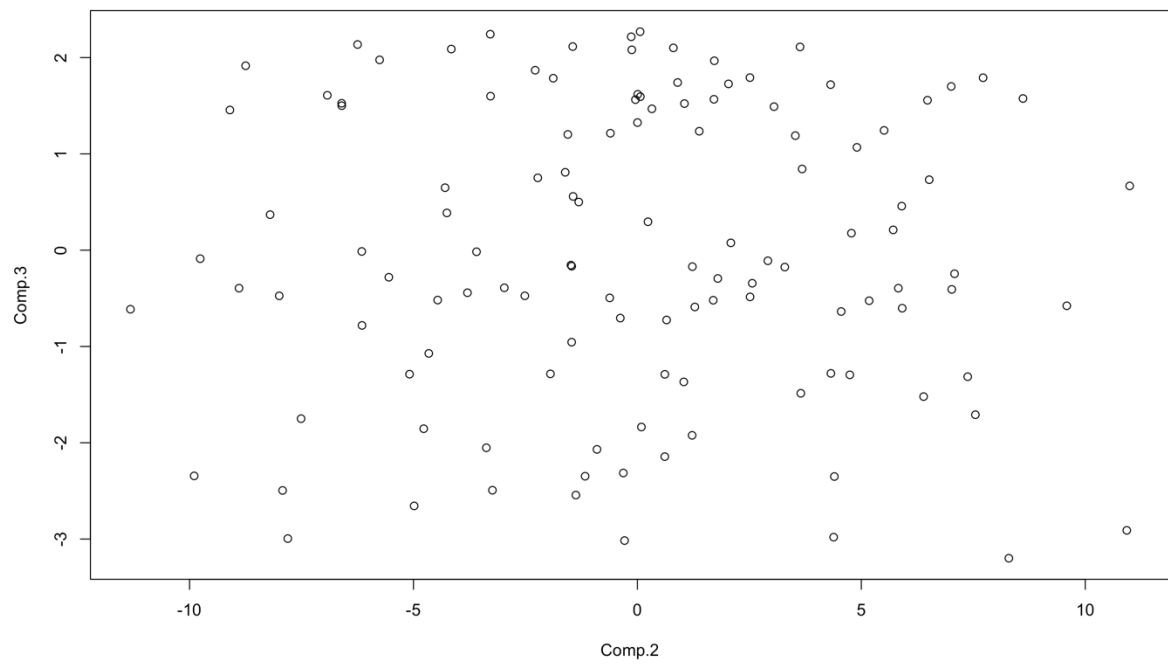
Бачимо, що дуже виділяється головна компонента, також друга і третя. Тому модель з перших трьох компонент пояснює 96% дисперсії даних.

Подивимось на попарні діаграми розсіювання для наших трьох вибраних компонент:

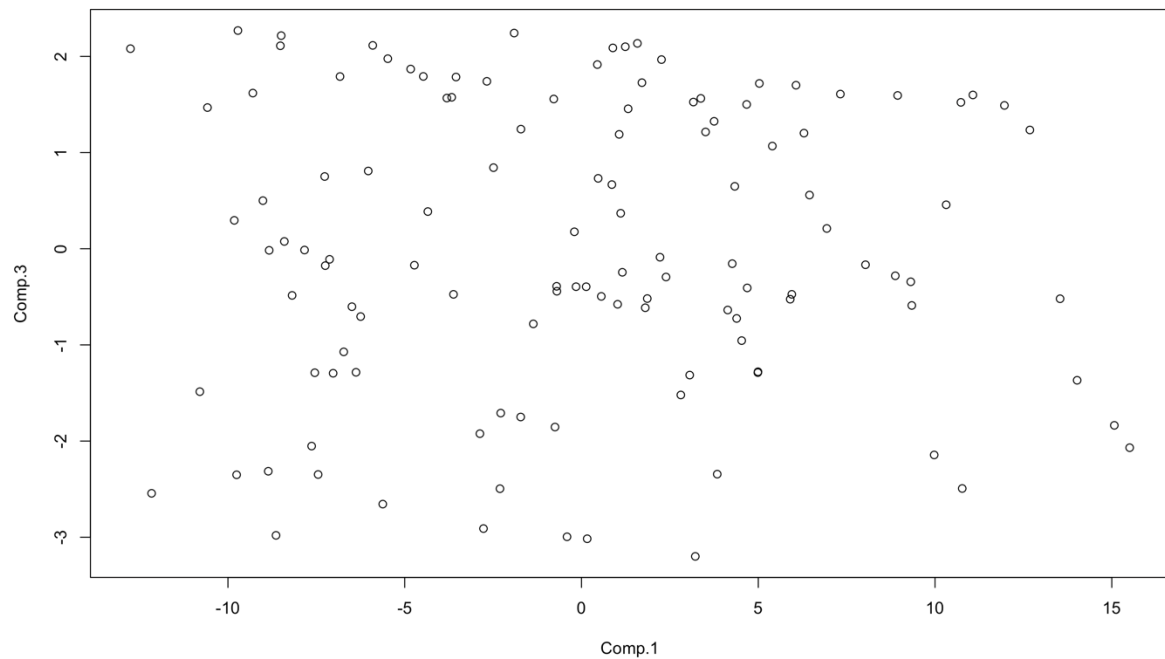
```
> plot(res$scores[,1:2])
```



```
> plot(res$scores[,2:3])
```

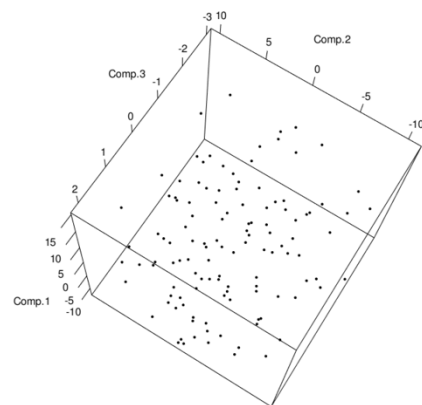
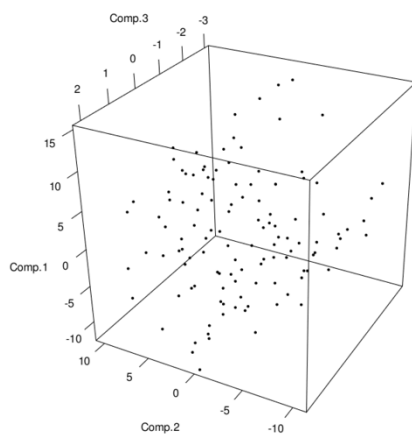


```
> plot(res$scores[,c(1,3)])
```



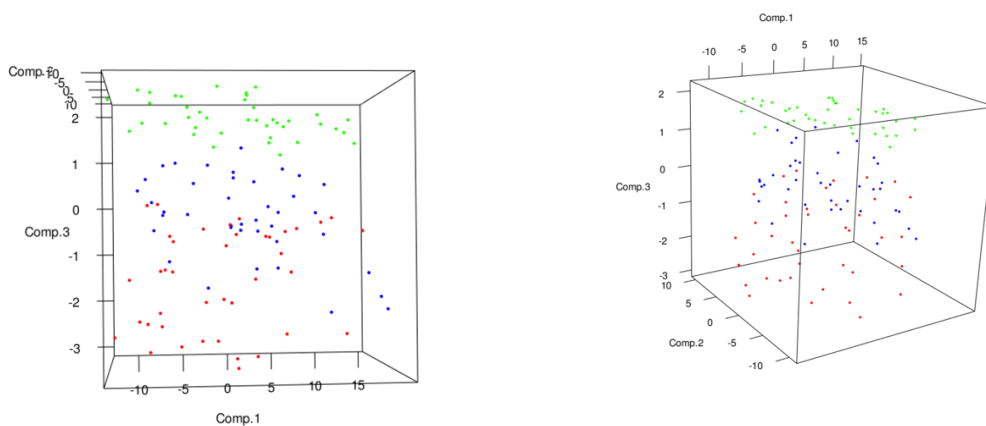
Далі щоб завантажити rgl, я спочатку завантажила XQuartz, і потім вже завантажився і rgl:

```
> library(rgl)
>
> plot3d(res$scores[,1:3])
```



Тут явного розподілення на кластери не видно, тому застосуємо кольорову 3д діаграму:

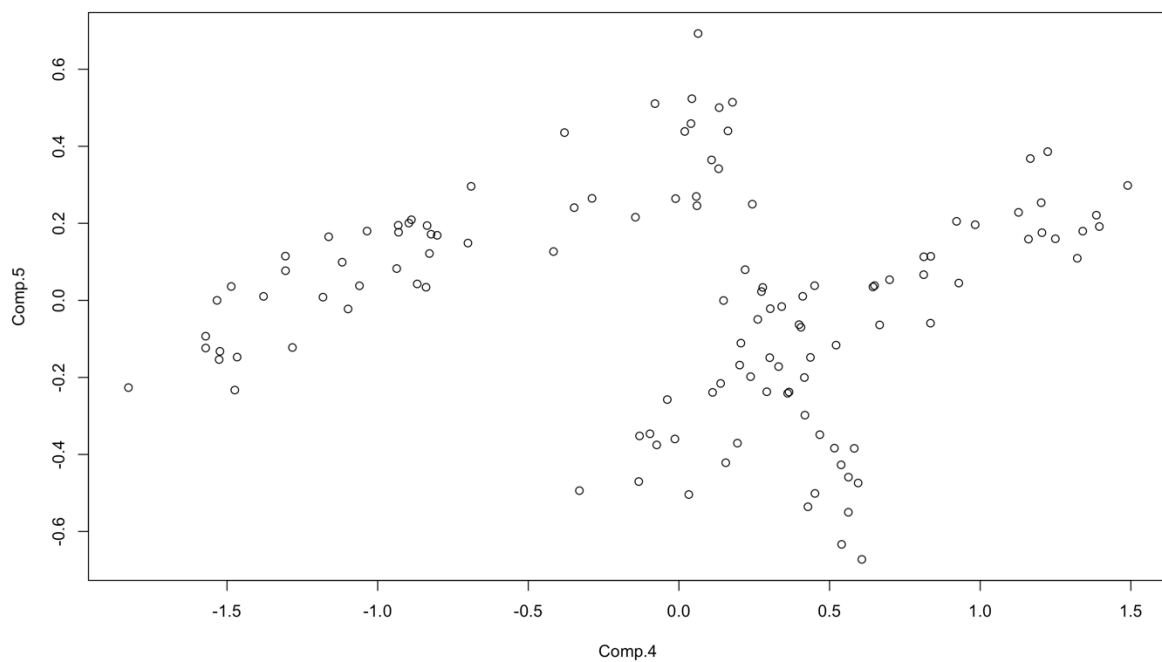
```
> plot3d(res$scores[,1:3],col=c("red","blue","green"))
```



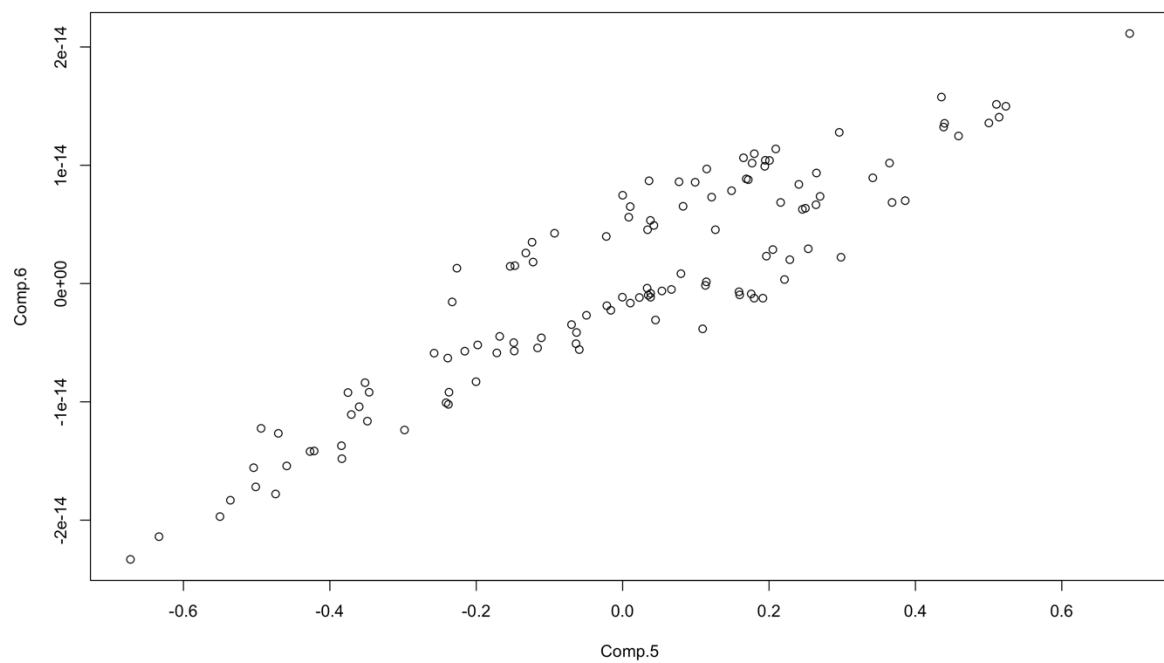
Дані розбиваються на три кластери, але на відміну від прикладу з книжки, тут їх без кольорів не видно. Також вони майже не перетинаються між собою, і не утворюють якоїсь явної структури – вийшли просто множини точок.

Результатів ці підрахунки не дали, тому я подивилася інші три компоненти: 4, 5, 6.

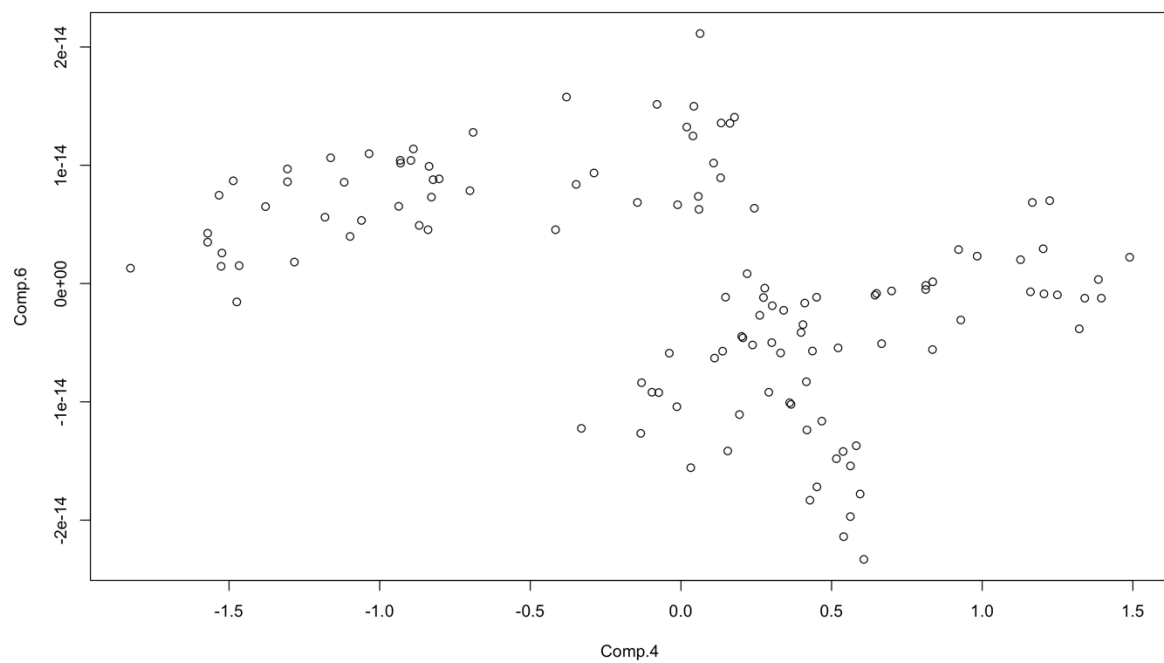
```
> plot(res$scores[,4:5])
```



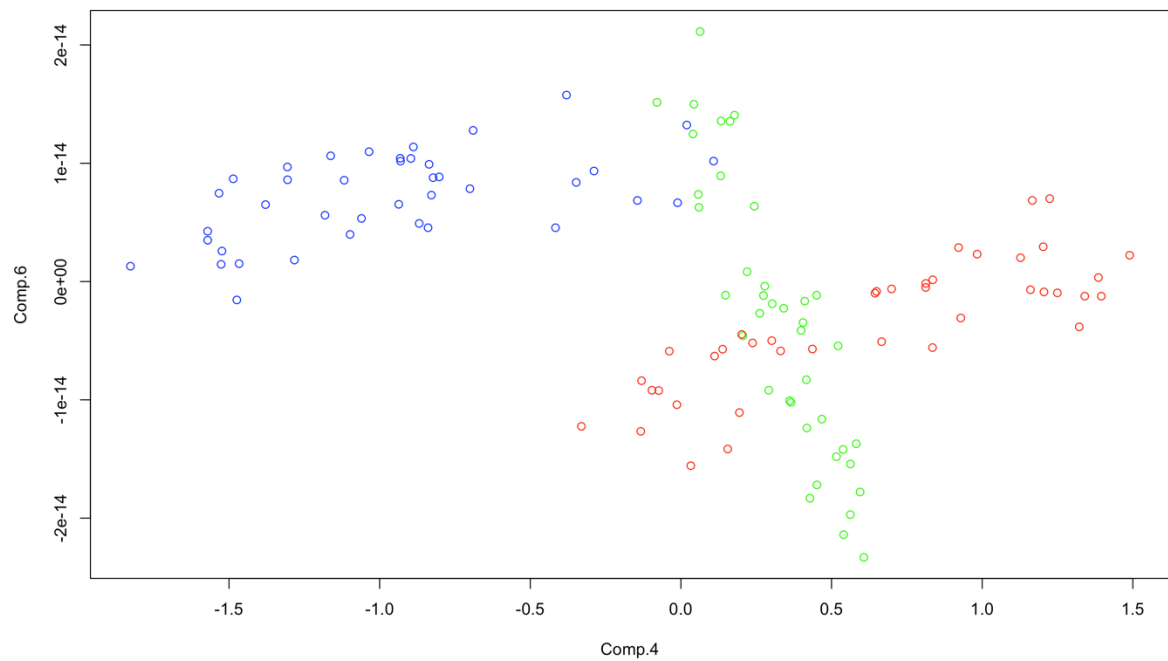
```
> plot(res$scores[,5:6])
```



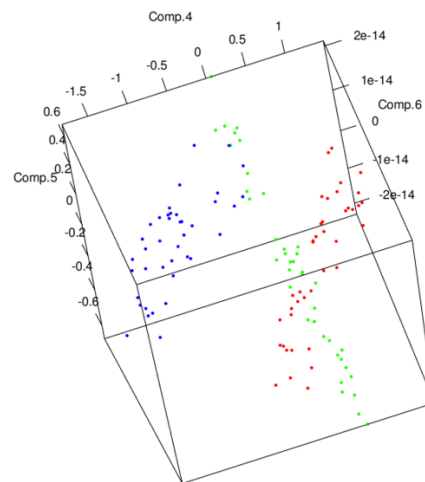
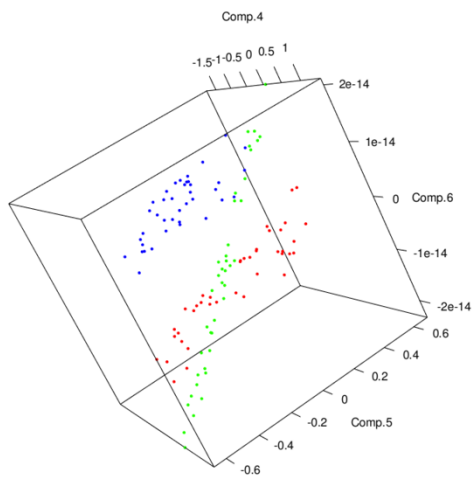
```
> plot(res$scores[,c(4,6)])
```



```
> plot(res$scores[,c(4,6)],col=c("red","blue","green"))
```



```
> library(rgl)
>
> plot3d(res$scores[,4:6],col=c("red","blue","green"))
```



Тут вже добре видно розділення на кластери. Третій кластер знаходиться на діагоналі кубу (зелений). Другий (синій) стикається з третім, але зовсім не стикається з першим (червоним). Також перший перетинає третій.