

(* Иерархический кластерный анализ *)

(* Загрузить пакет *)

```
Needs["HierarchicalClustering`"]
```

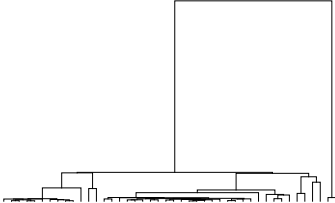
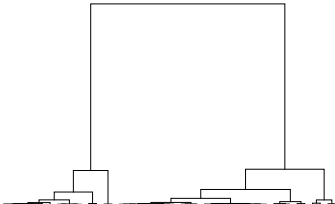
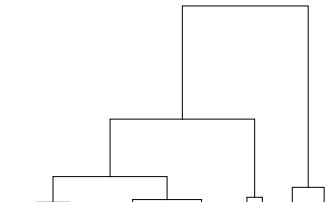
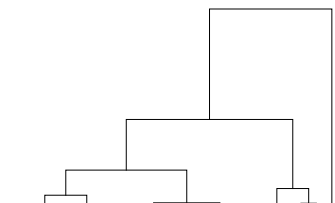
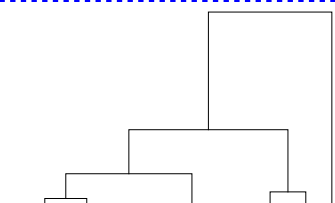
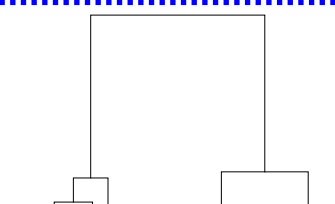
(* Подготовить данные *)

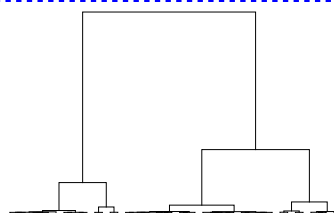
```
d = Import[FileNameJoin[{NotebookDirectory[], "org1.xlsx"}]][[1]]
```

```
d // TableForm
```

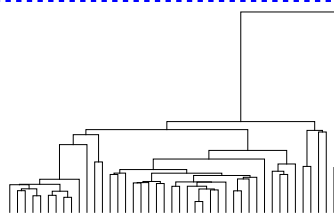
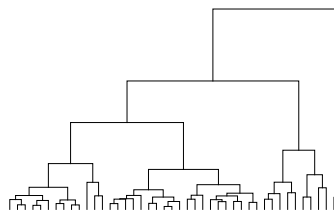
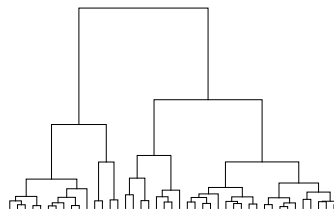
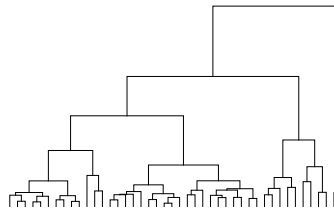
| Company | CAPEX | SG | |
|-------------------------------------------------|------------|-------------|---|
| ПАО "Газпром" | 0.116853 | 0.102006 | ! |
| ПАО АНК "Башнефть" | 0.0672134 | 0.0578289 | ! |
| ПАО "Татнефть" | 0.0901148 | 0.024508 | ! |
| НК "Роснефть - Дагнефть" | 0.106811 | -0.0113245 | { |
| ОАО "Саратовнефтегаз" | 0.0347499 | -0.0215618 | ! |
| "Волгограднефтегеофизика" | 0.152515 | -0.270714 | ! |
| "Саратовнефтегеофизика" | 0.00080903 | -0.116149 | - |
| ОАО НПО "Буровая техника" | 0.00269278 | -0.161388 | - |
| "Башнефтегеофизика" | 0.234207 | 0.279626 | ! |
| ОАО "Балаковорезинотехника" | 0.0177528 | -0.181726 | (|
| АО "Газпром промгаз" | 0.00565302 | -0.367363 | (|
| ПАО "Северсталь" | 0.0749816 | -0.0774239 | ! |
| ОАО "Белорецкий металлургический комбинат" | 0.00456162 | -0.00463725 | ! |
| ПАО "ГАЙСКИЙ ГОК" | 0.139497 | 0.0119947 | ! |
| ПАО "Комбинат Южуралникель" | 0.00312661 | -0.181248 | - |
| ОАО "Рузхиммаш" | 0.00508106 | 0.0461081 | (|
| ОАО "Новотроицкий цементный завод" | 0.0500763 | -0.0113405 | { |
| ОАО "Горнозаводскцемент" | 0.0421824 | 0.00600758 | ! |
| ОАО "Соликамский магниевый завод" | 0.0400043 | -0.192963 | - |
| ОАО Усинскгеонефть | 0.00267117 | 0.209295 | (|
| ОАО "Новолипецкий металлургический комбинат" | 0.0795882 | -0.10259 | ! |
| ОАО «Муромтепловоз» | 0.006731 | 0.268483 | ! |
| ОАО «Ковровский электромеханический завод» | 0.0351641 | 0.210661 | ! |
| ОАО "Новатэк" | 0.159848 | 0.413252 | ! |
| ПАО "Транснефть" | 0.111675 | 0. | ! |
| ПАО "Алроса" | 0.125251 | 0.116815 | ! |
| ОАО "Норильский никель" | 0.0988855 | -0.0468478 | ! |
| ОАО "Магнитогорский металлургический комбинат" | 0.0751104 | -0.121998 | - |
| ПАО "Мечел" | 0.0612617 | -0.193257 | - |
| ОАО "Полюс золото" | 0.190736 | -0.11042 | - |
| ПАО "Распадская" | 0.0836338 | 0.00429549 | - |
| ПАО "Корпорация ВСМПО-АВИСМА" | 0.0804434 | 0.167119 | ! |
| ОАО "Рн Холдинг" | 0.121867 | -0.0754498 | ! |
| ОАО НК "Роснефть" | 0.144296 | 0.319586 | ! |
| ОАО "Славнефть-Мегионнефтегаз" | 0.102013 | -0.0168618 | ! |
| ОАО «Алтайкрайгазсервис» | 0.0669874 | -0.0584479 | (|
| ПАО "Кокс" | 0.101528 | -0.0583756 | - |
| "Коммунарковский рудник" | 0.155107 | 0.602174 | ! |
| ОАО "Красноярскнефтепродукт" | 0.0300638 | -0.15284 | ! |
| ОА "Новосибирскнефтегаз" | 0.0291121 | -0.173198 | - |
| ОАО "Оргэнергогаз" | 0.0168173 | -0.241071 | ! |
| АО "Приморскуголь" | 0.135955 | 0.238498 | ! |
| АО "Сибирьгазсервис" | 0.0241752 | 0.140283 | ! |
| ОАО "Якутская топливно-энергетическая компания" | 0.0437165 | 0.0534615 | ! |

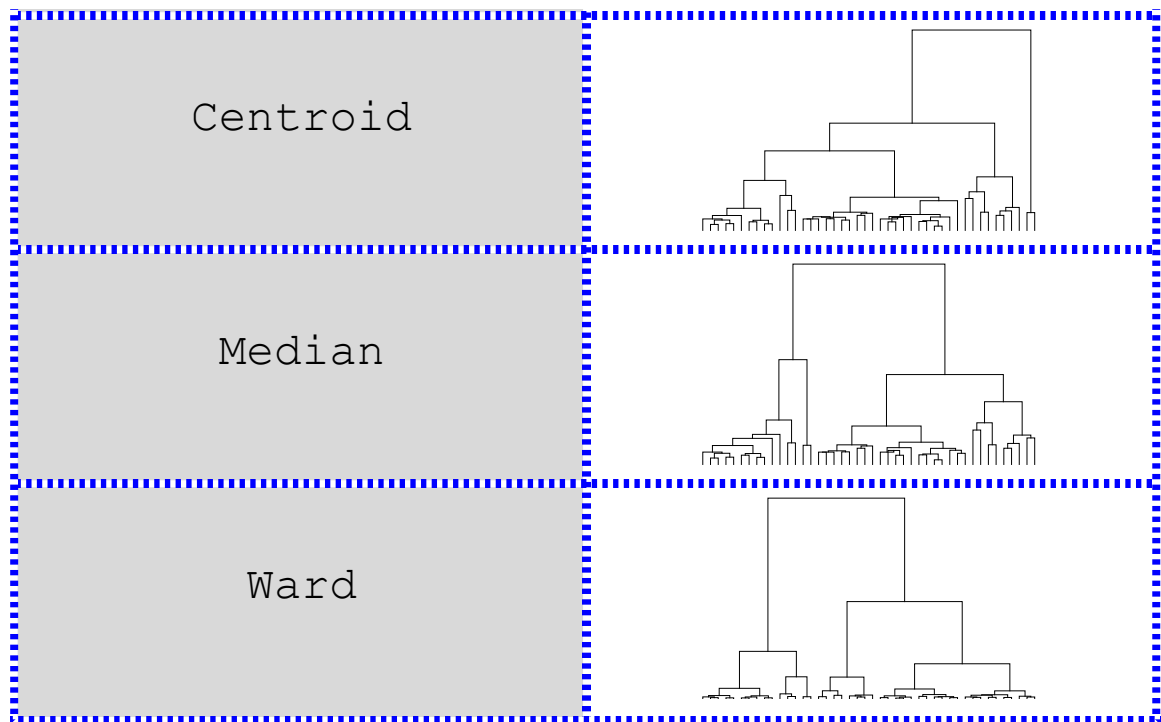
```
{#, DendrogramPlot[ddData, DistanceFunction -> Automatic, Linkage -> #]} & /@  
{"Single", "Average", "Complete",  
"WeightedAverage", "Centroid", "Median", "Ward"} //  
Grid[Prepend[#, {"Table", "Dendrogram(EuclideanDistance)"}], Frame -> All,  
FrameStyle -> Directive[Dashing[2], Thickness[3], Blue], ItemSize -> 20,  
Background -> {1 -> LightGray, 1 -> LightGray}, ItemStyle -> Large] &
```

| Table | Dendrogram(EuclideanDistance) |
|-----------------|--------------------------------------------------------------------------------------|
| Single |  |
| Average |  |
| Complete |  |
| WeightedAverage |  |
| Centroid |  |
| Median |  |

| | |
|------|------------------------------------------------------------------------------------|
| Ward |  |
|------|------------------------------------------------------------------------------------|

```
{#, DendrogramPlot[ddData, DistanceFunction → ManhattanDistance,
  Linkage → #]} & /@ {"Single", "Average", "Complete",
  "WeightedAverage", "Centroid", "Median", "Ward"} //
Grid[Prepend[#, {"Table", "Dendrogram(ManhattanDistance)"}], Frame → All,
  FrameStyle → Directive[Dashing[2], Thickness[3], Blue], ItemSize → 20,
  Background → {1 → LightGray, 1 → LightGray}, ItemStyle → Large] &
```

| | |
|-----------------|--------------------------------------------------------------------------------------|
| Table | Dendrogram(ManhattanDistance) |
| Single |  |
| Average |  |
| Complete |  |
| WeightedAverage |  |



```
po = DistanceMatrix[ddData, DistanceFunction → EuclideanDistance];  
% // MatrixForm
```

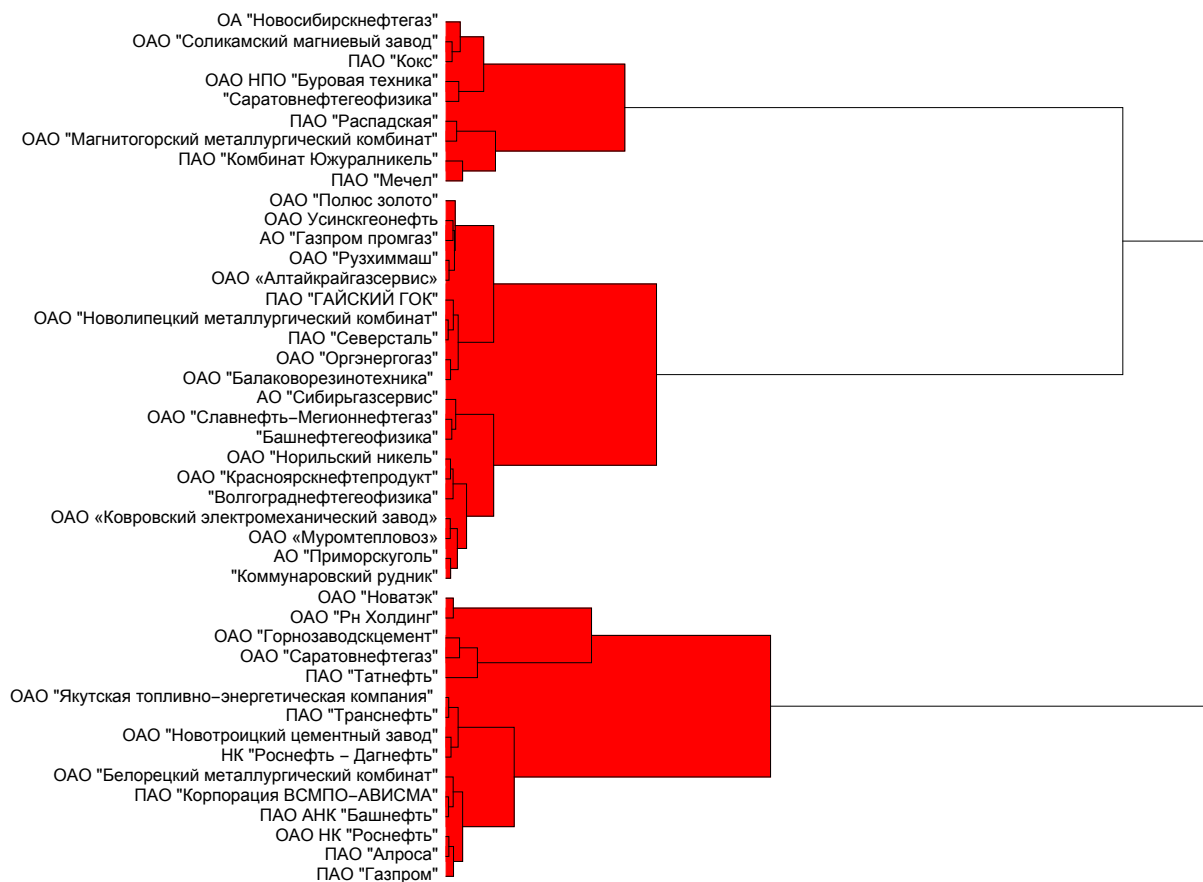
```
aggl = DirectAgglomerate[po, ddNames, Linkage → "Ward"]
```

```
Cluster[Cluster[Cluster[Cluster[Cluster[ПАО "Газпром",
  Cluster[ПАО "Алроса", ОАО НК "Роснефть", 0.276779, 1, 1], 0.745454, 1, 2],
  Cluster[Cluster[ПАО АНК "Башнефть", ПАО "Корпорация ВСМПО-АВИСМА", 0.24968,
    1, 1], ОАО "Белорецкий металлургический комбинат", 0.711362, 2, 1],
  1.66043, 3, 3], Cluster[Cluster[НК "Роснефть - Дагнефть",
    ОАО "Новотроицкий цементный завод", 0.49592, 1, 1],
  Cluster[ПАО "Транснефть", ОАО "Якутская топливно-энергетическая компания" ,
    0.263254, 1, 1], 1.20912, 2, 2], 6.84242, 6, 4],
Cluster[Cluster[ПАО "Татнефть", Cluster[ОАО "Саратовнефтегаз",
  ОАО "Горнозаводскцемент", 1.35085, 1, 1], 3.14562, 1, 2],
Cluster[ОАО "РН Холдинг", ОАО "Новатэк", 0.72408, 1, 1],
14.6129, 3, 2], 32.6012, 10, 5],
Cluster[Cluster[Cluster[Cluster[Cluster[
  Cluster["Коммунаровский рудник", АО "Приморскуголь", 0.452958, 1, 1],
  Cluster[ОАО «Муромтепловоз», ОАО «Ковровский электромеханический завод»,
    0.419996, 1, 1], 1.13066, 2, 2],
  Cluster["Волгограднефтегеофизика", Cluster[ОАО "Красноярскнефтепродукт",
    ОАО "Норильский никель", 0.440979, 1, 1], 0.702057, 1, 2], 2.05466, 4, 3],
  Cluster[Cluster["Башнефтегеофизика", ОАО "Славнефть-Мегионнефтегаз",
    0.572687, 1, 1], АО "Сибирьгазсервис", 0.967247, 2, 1], 4.77743, 7, 3],
Cluster[Cluster[Cluster[ОАО "Балаковорезинотехника" , ОАО "Оргэнергогаз",
  0.44835, 1, 1], Cluster[Cluster[ПАО "Северсталь",
    ОАО "Новолипецкий металлургический комбинат", 0.21089, 1, 1],
  ПАО "ГАЙСКИЙ ГОК", 0.725848, 2, 1], 1.21402, 2, 3], Cluster[
  Cluster[Cluster[ОАО «Алтайкрайгазсервис», ОАО "Рузхиммаш", 0.294086, 1, 1],
  Cluster[АО "Газпром промгаз", ОАО Усинскгеонефть, 0.693699, 1, 1],
    0.838083, 2, 2], ОАО "Полюс золото", 0.916598, 4, 1],
  4.77803, 5, 5], 21.1432, 10, 10], Cluster[
Cluster[Cluster[ПАО "Мечел", ПАО "Комбинат Южуралникель", 1.65677, 1, 1],
  Cluster[ОАО "Магнитогорский металлургический комбинат",
    ПАО "Распадская", 1.06142, 1, 1], 4.96333, 2, 2], Cluster[
  Cluster["Саратовнефтегеофизика", ОАО НПО "Буровая техника", 1.27394, 1, 1],
  Cluster[Cluster[ПАО "Кокс", ОАО "Соликамский магниевый завод", 0.603882,
    1, 1], ОА "Новосибирскнефтегаз", 1.41913, 2, 1], 3.77601, 2, 3],
  17.9557, 4, 5], 67.9985, 20, 9], 142.229, 15, 29]
```

```

itog = DendrogramPlot[aggl, LeafLabels → Automatic,
  Orientation → Right, HighlightLevel → 3, HighlightStyle → Red]

```



```

splitCluster1 = ClusterSplit[aggl, 3]

firstCluster = splitCluster1[[1]]
secondCluster = splitCluster1[[2]]
thirdCluster = splitCluster1[[3]]

firstClusterList = Cases[firstCluster, _String, All]

firstClustMean =
  Mean /@ Rest /@ Select[dd, MemberQ[firstClusterList, First[#]] &] // Mean

secondClusterList = Cases[secondCluster, _String, All]

secondClustMean =
  Mean /@ Rest /@ Select[dd, MemberQ[secondClusterList, First[#]] &] // Mean

thirdClusterList = Cases[thirdCluster, _String, All]

```

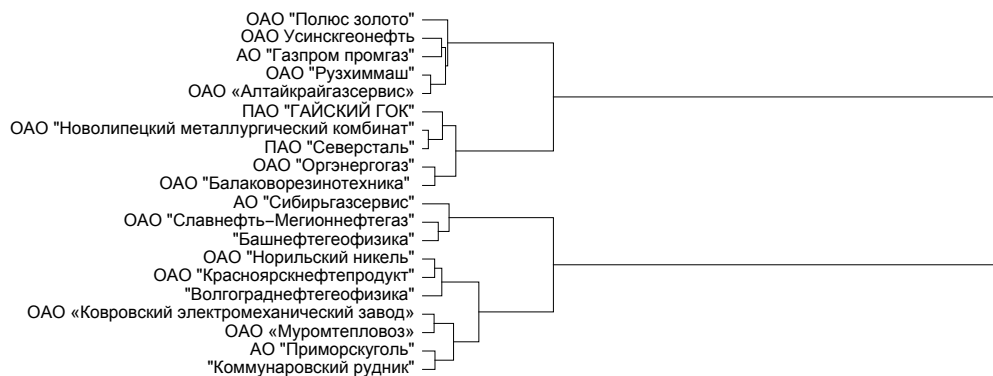
```
thridClustMean =  
  Mean /@ Rest /@ Select[dd, MemberQ[thirdClusterList, First[#]] &] // Mean
```

-1.15424

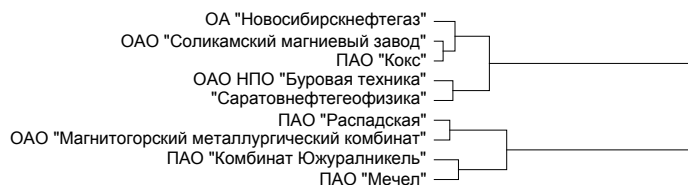
```
DendrogramPlot[firstCluster, LeafLabels → Automatic, Orientation → Right]
```



```
DendrogramPlot[secondCluster, LeafLabels → Automatic, Orientation → Right]
```

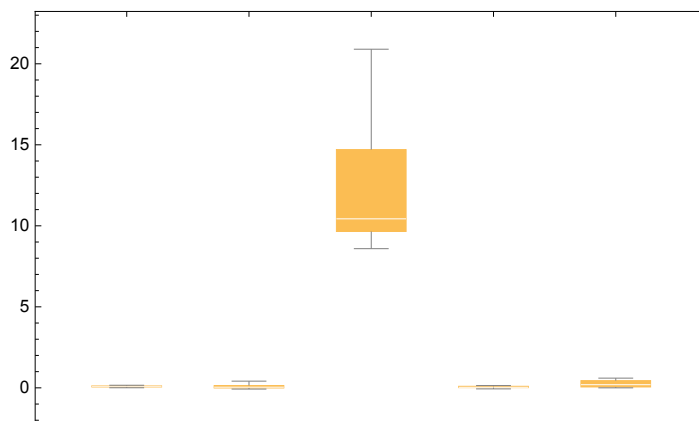


```
DendrogramPlot[thirdCluster, LeafLabels → Automatic, Orientation → Right]
```

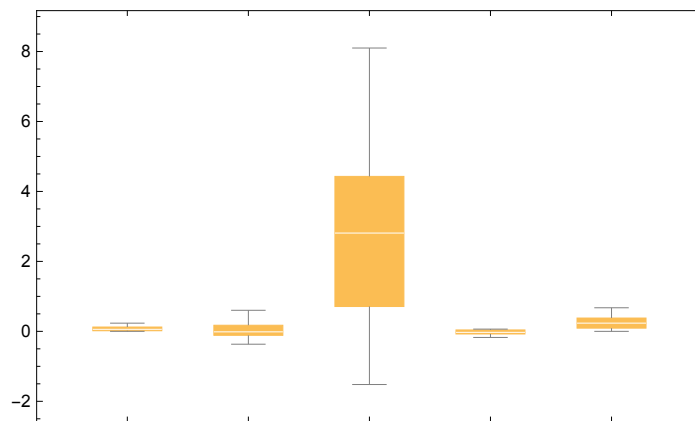


(* сравнение каждого кластера по всем признакам (подряд 3 кластера)*)

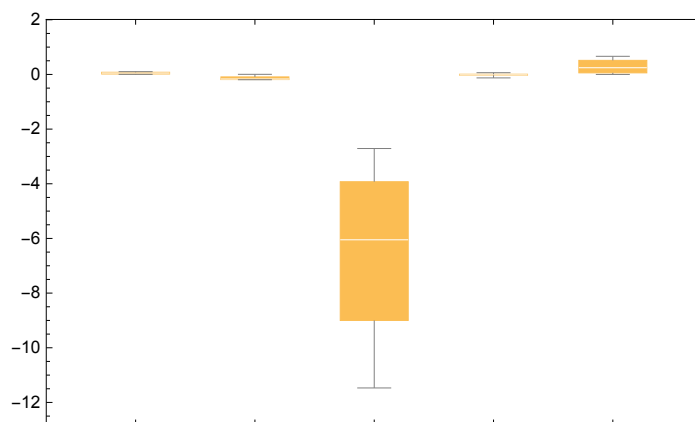
```
BoxWhiskerChart[firstClustKMean[ [All, #]] & /@ Range[5]]
```



```
BoxWhiskerChart[secondClustKMean[[All, #]] & /@ Range[5]]
```

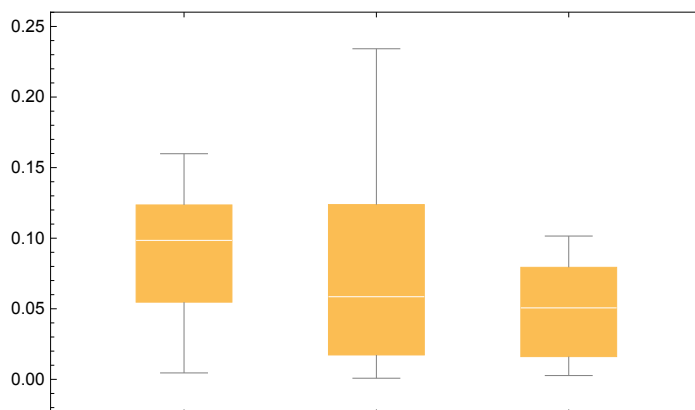


```
BoxWhiskerChart[thirdClustKMean[[All, #]] & /@ Range[5]]
```

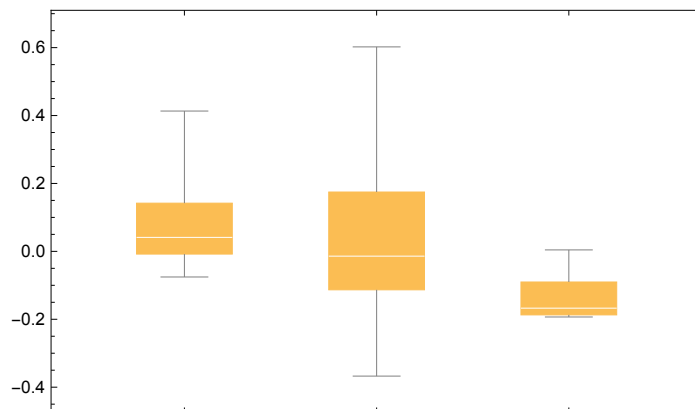


(* сравнение 3-х кластеров по одному признаку (поряд 5 признаков)*)

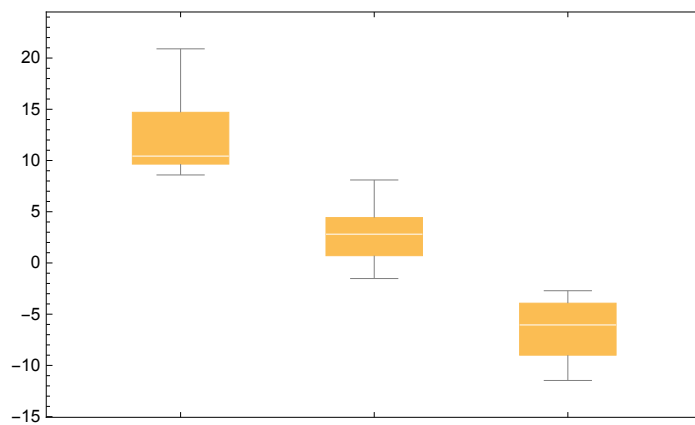
```
BoxWhiskerChart[{firstClustKMean[[All, 1]],  
secondClustKMean[[All, 1]], thirdClustKMean[[All, 1]]}]
```



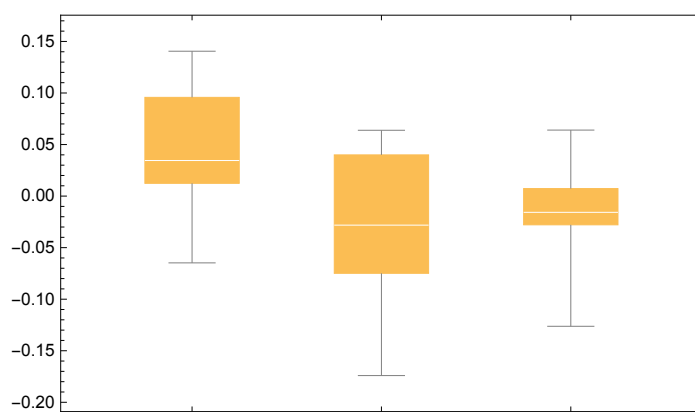

```
BoxWhiskerChart[{firstClustKMean[[All, 2]],  
secondClustKMean[[All, 2]], thirdClustKMean[[All, 2]]}]
```



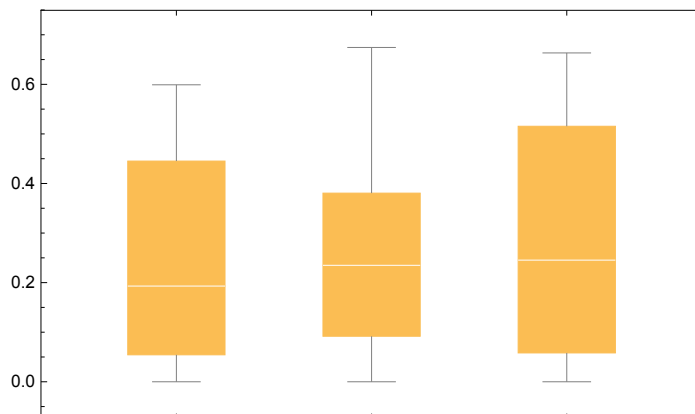
```
BoxWhiskerChart[{firstClustKMean[[All, 3]],  
secondClustKMean[[All, 3]], thirdClustKMean[[All, 3]]}]
```



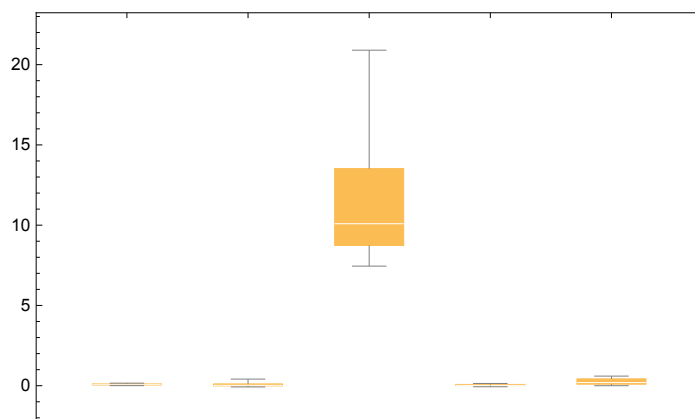
```
BoxWhiskerChart[{firstClustKMean[[All, 4]],  
secondClustKMean[[All, 4]], thirdClustKMean[[All, 4]]}]
```



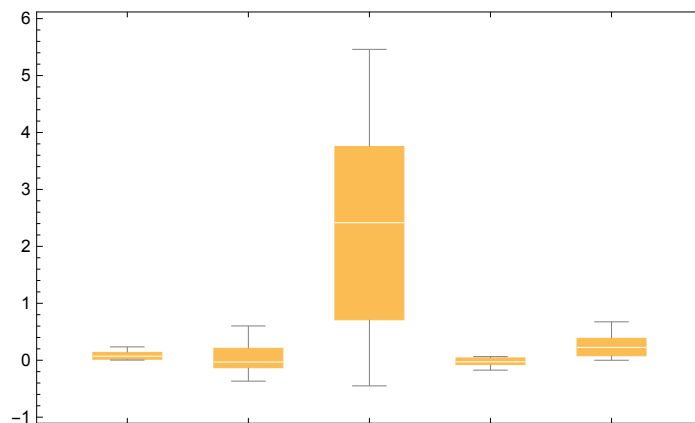
```
BoxWhiskerChart[{firstClustKMean[[All, 5]],
  secondClustKMean[[All, 5]], thirdClustKMean[[All, 5]]}]
```



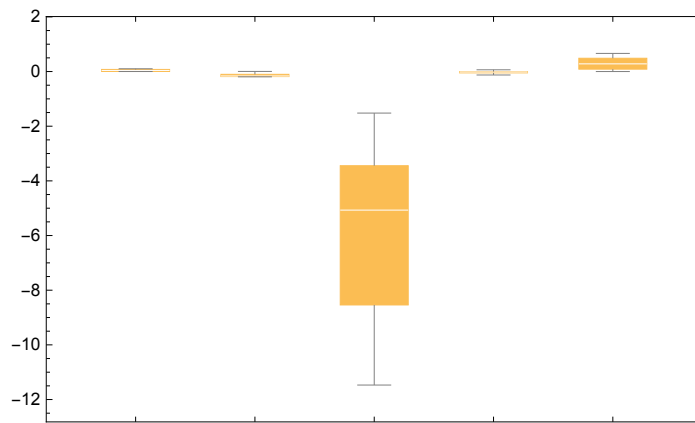
```
firstIer = Rest /@ Select[dd, MemberQ[firstClusterList, First[#]] &];
secondIer = Rest /@ Select[dd, MemberQ[secondClusterList, First[#]] &];
thirdIer = Rest /@ Select[dd, MemberQ[thirdClusterList, First[#]] &];
(* сравнение каждого кластера по всем признакам (порядк 3 кластера) *)
BoxWhiskerChart[firstIer[[All, #]] & /@ Range[5]]
```



```
BoxWhiskerChart[secondIer[[All, #]] & /@ Range[5]]
```

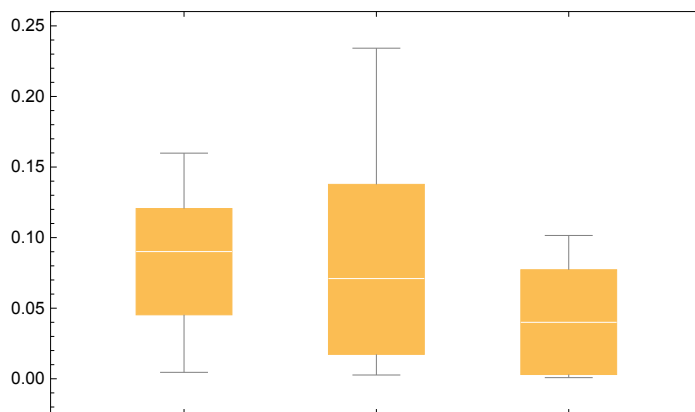


```
BoxWhiskerChart[thirdIer[[All, #]] & /@ Range[5]]
```

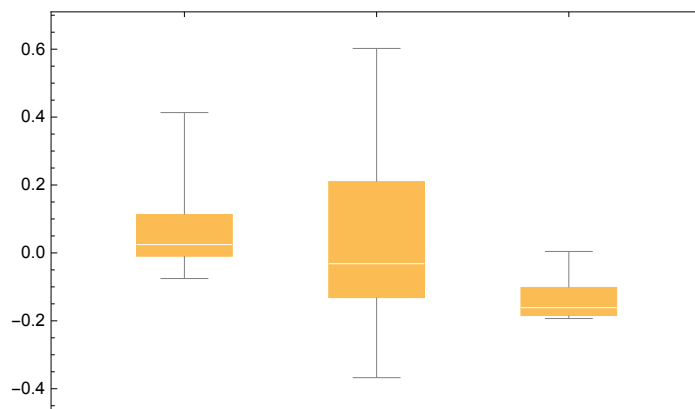


(* сравнение 3-х кластеров по одному признаку (поряд 5 признаков)*)

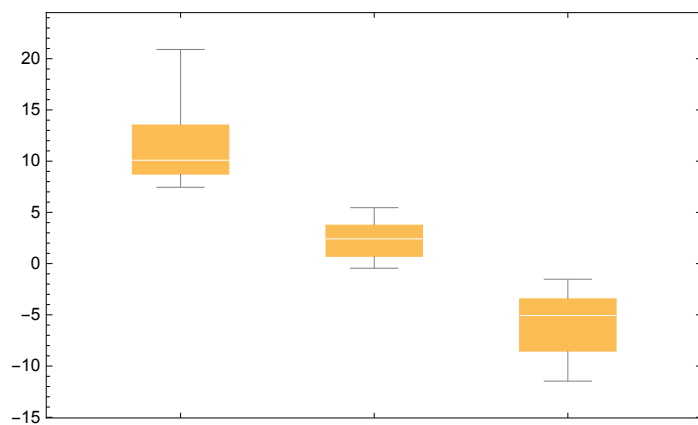
```
BoxWhiskerChart[{firstIer[[All, 1]], secondIer[[All, 1]], thirdIer[[All, 1]]}]
```



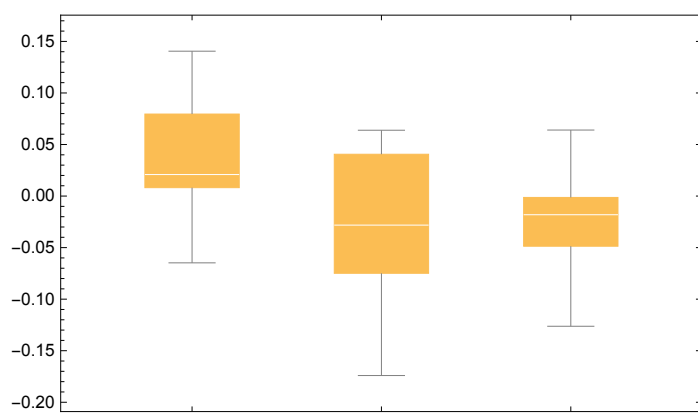
```
BoxWhiskerChart[{firstIer[[All, 2]], secondIer[[All, 2]], thirdIer[[All, 2]]}]
```



```
BoxWhiskerChart[{firstIer[[All, 3]], secondIer[[All, 3]], thirdIer[[All, 3]]}]
```



```
BoxWhiskerChart[{firstIer[[All, 4]], secondIer[[All, 4]], thirdIer[[All, 4]]}]
```



```
BoxWhiskerChart[{firstIer[[All, 5]], secondIer[[All, 5]], thirdIer[[All, 5]]}]
```

