

Clustering Example 1: Five Data Points

Amy Wagaman, Nathan Carter

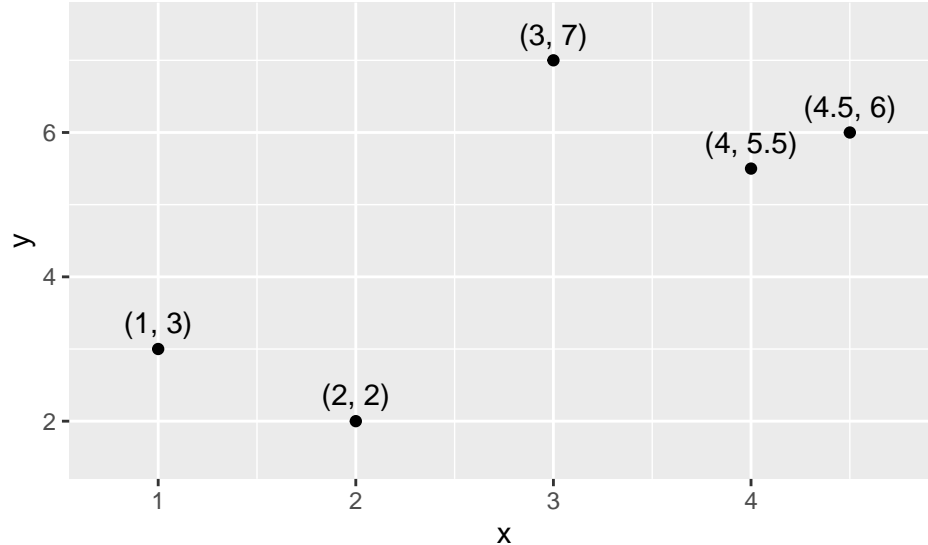
July 2020

Load necessary library.

```
library(ggformula)
```

Generate 5 data points and graph them.

```
x <- c(4,2,4.5,3,1)
y <- c(5.5,2,6,7,3)
Label <- c("(4, 5.5)", "(2, 2)", "(4.5, 6)", "(3, 7)", "(1, 3)")
mydata <- data.frame(x, y, Label)
ggplot(mydata,
  aes(x=x, y=y, label=Label))+
  xlim(0.75, 4.75)+ylim(1.5, 7.5)+geom_point()+
  geom_text(aes(label=Label),hjust = 0.5,vjust = -0.75)
```



Convert to a data frame and compute the distance matrix.

```
mydata <- data.frame(x,y)
mydata
```

```
##      x  y
## 1 4.0 5.5
## 2 2.0 2.0
```

```
## 3 4.5 6.0
## 4 3.0 7.0
## 5 1.0 3.0
```

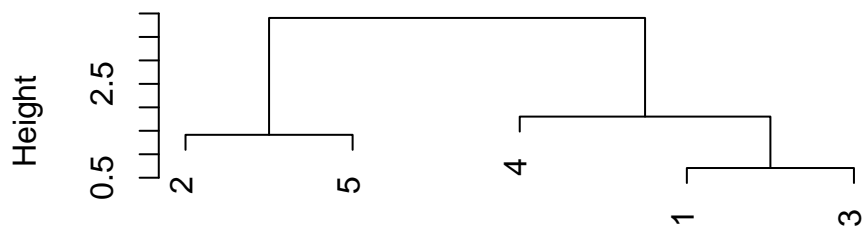
```
round(dist(mydata),2)
```

```
##      1      2      3      4
## 2 4.03
## 3 0.71 4.72
## 4 1.80 5.10 1.80
## 5 3.91 1.41 4.61 4.47
```

Perform hierarchical clustering with single linkage.

```
plot(hclust(dist(mydata), method = "single"))
```

Cluster Dendrogram



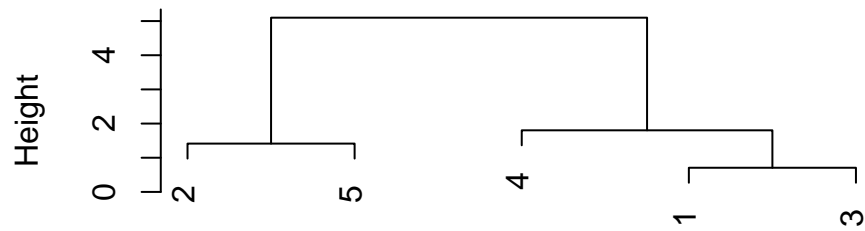
```
dist(mydata)
hclust (*, "single")
```

Perform hierarchical clustering with complete linkage.

The clusters are the same as in the previous case, but there are minor changes to the heights of the bars in the dendrogram.

```
plot(hclust(dist(mydata), method = "complete"))
```

Cluster Dendrogram

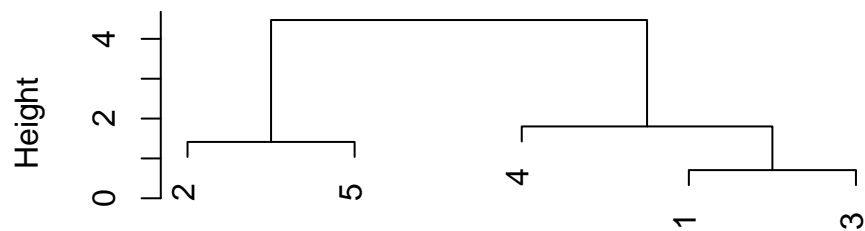


```
dist(mydata)
hclust (*, "complete")
```

Perform hierarchical clustering with average linkage.

```
plot(hclust(dist(mydata), method = "average"))
```

Cluster Dendrogram

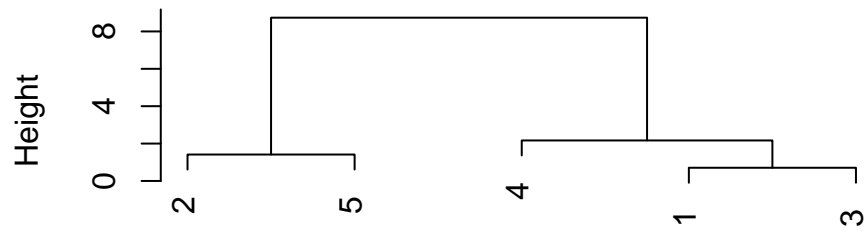


```
dist(mydata)
hclust (*, "average")
```

Perform hierarchical clustering with Ward's method.

```
plot(hclust(dist(mydata), method = "ward.D"))
```

Cluster Dendrogram



```
dist(mydata)  
hclust (*, "ward.D")
```