

Problem Set 4

Shuai Yuan

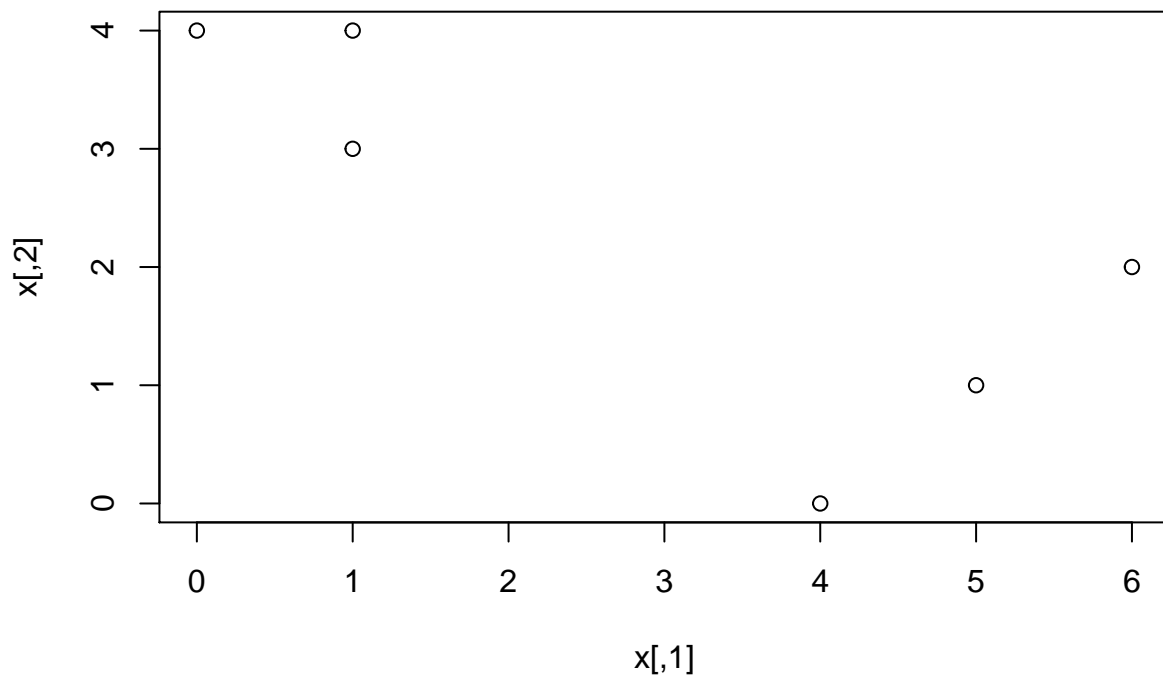
March 1, 2020

Q1

```
x <- cbind(c(1, 1, 0, 5, 6, 4), c(4, 3, 4, 1, 2, 0))
```

1.1

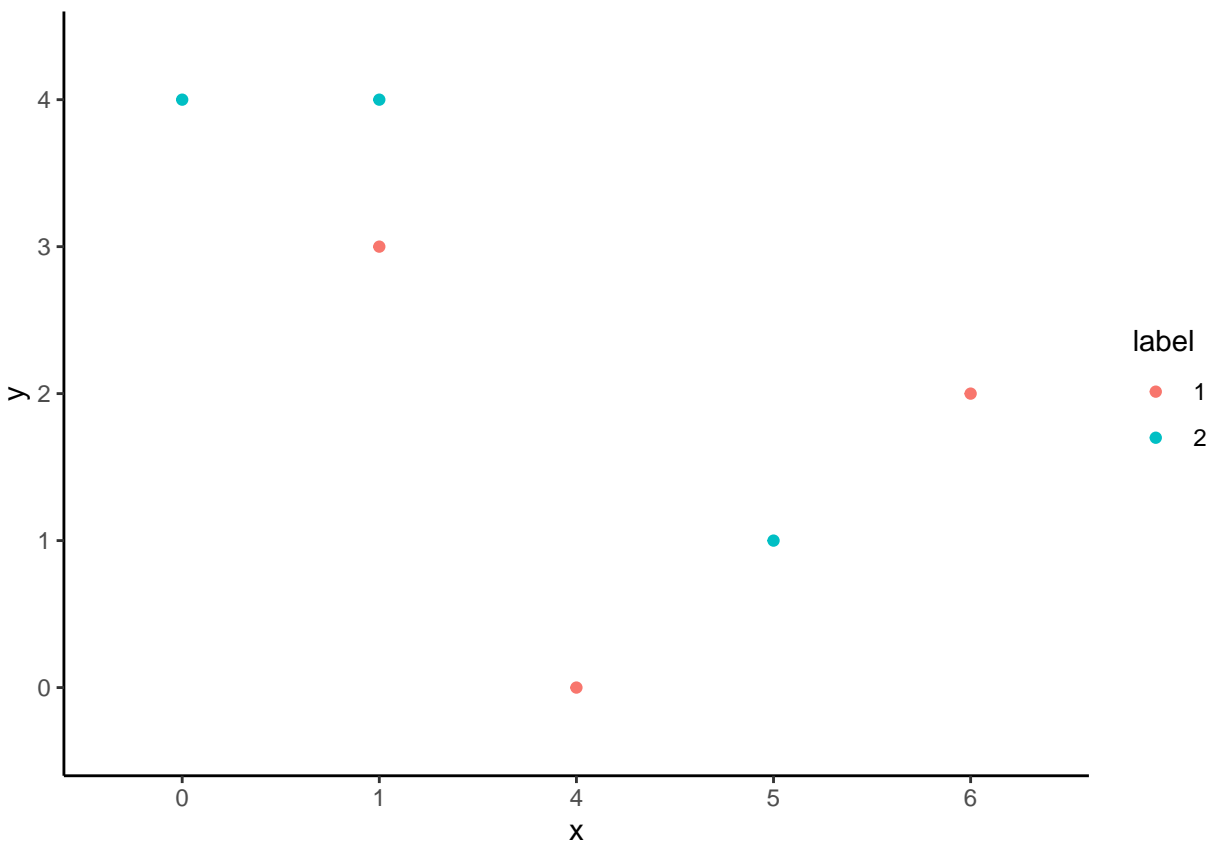
```
plot(x)
```



1.2

```
set.seed(1345)
label <- sample(c('1','2'), size = nrow(x), replace = TRUE)
cluster <- cbind(x, label)%>%
  data.frame()%>%
  rename(x = V1,
         y = V2)
```

```
cluster%>%
  ggplot(aes(x, y, color = label)) +
  geom_point()
```



1.3

```
cluster<- cluster%>%
  mutate(x = as.numeric(x))%>%
  mutate(y = as.numeric(y))%>%
  mutate(label = as.numeric(label))

cluster%>%
  group_by(label)%>%
  summarize(x = mean(x),
            y = mean(y))%>%
  kable()
```

label	x	y
1	3.333333	2.666667
2	2.333333	4.000000

1.4

```
assign<- function(data){
  centroid<- data%>%
```

```

  group_by(label)%>%
  summarize(x = mean(x),
            y = mean(y))
data <- data%>% mutate(label = ifelse(
  ((data$x - centroid$x[1])^2 + (data$y - centroid$y[1])^2)<
  ((data$x - centroid$x[2])^2 + (data$y - centroid$y[2])^2),
  1,2))
  return(data)
}

kable(assign(cluster))

```

x	y	label
2	5	2
2	4	2
1	5	2
4	2	1
5	3	1
3	1	1