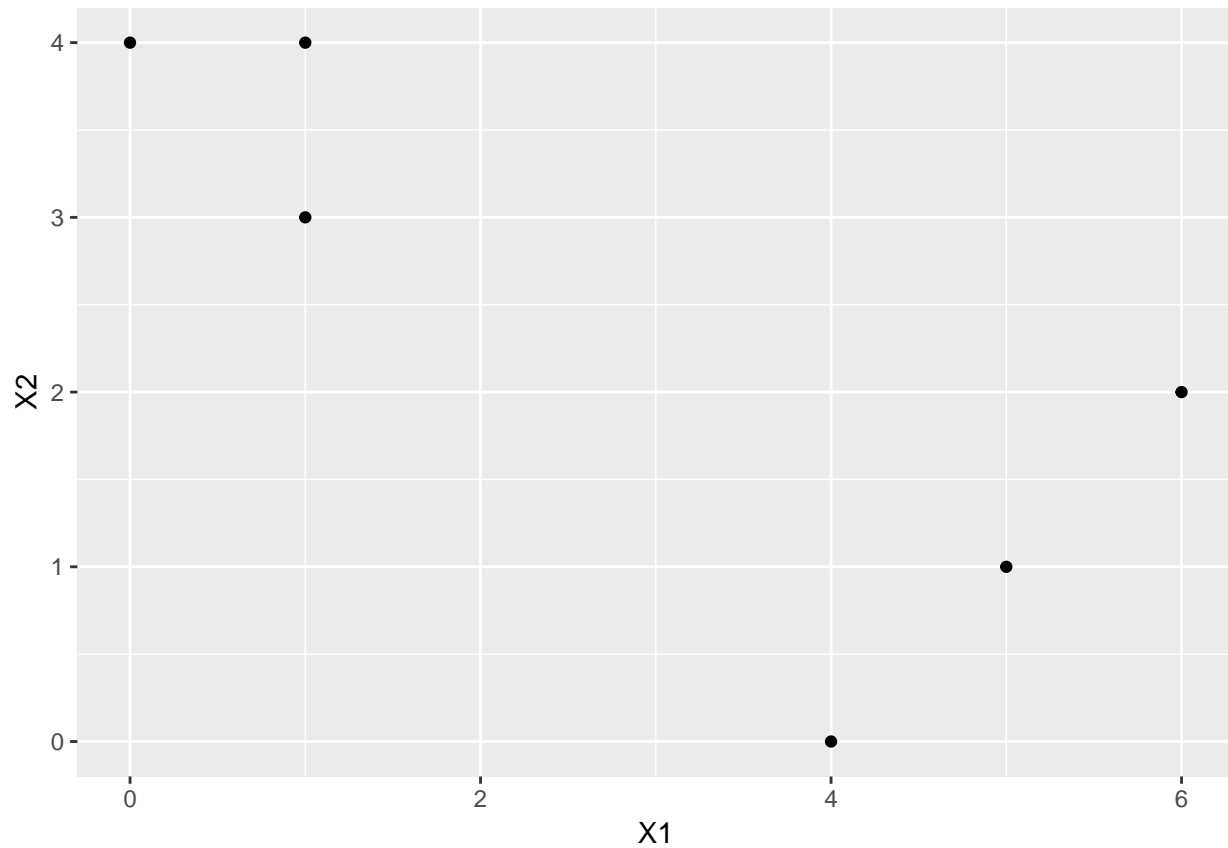# Week 8 Assignment

David Russo

3/9/2017

10.7 #3

- a)

```
obs <- data.frame(
  X1 = c(1, 1, 0, 5, 6, 4),
  X2 = c(4, 3, 4, 1, 2, 0)
)

obs %>%
  ggplot(aes(x = X1, y = X2)) +
  geom_point()
```


- b)

```
# set random number seed for replicable results
set.seed(2017)

# assign the labels
obs$cluster <- sample(rep(c(1, 2), each = 3), 6, replace = FALSE)
```

```
# print the labels
obs
```

```
##   X1 X2 cluster
## 1  1  4       2
## 2  1  3       1
## 3  0  4       1
## 4  5  1       1
## 5  6  2       2
## 6  4  0       2
```

- c)

```
centroids <-
  obs %>%
  dplyr::group_by(cluster) %>%
  dplyr::summarise(X1_mean = round(mean(X1), 2), X2_mean = round(mean(X2), 2)) %>%
  dplyr::mutate(centroid = paste(X1_mean, X2_mean, sep = ", ")) %>%
  dplyr::select(cluster, centroid) %>%
  as.data.frame()

centroids
```

```
##   cluster centroid
## 1       1  2, 2.67
## 2       2  3.67, 2
```