Exercise 9

Den seje gruppe

3/31/2021

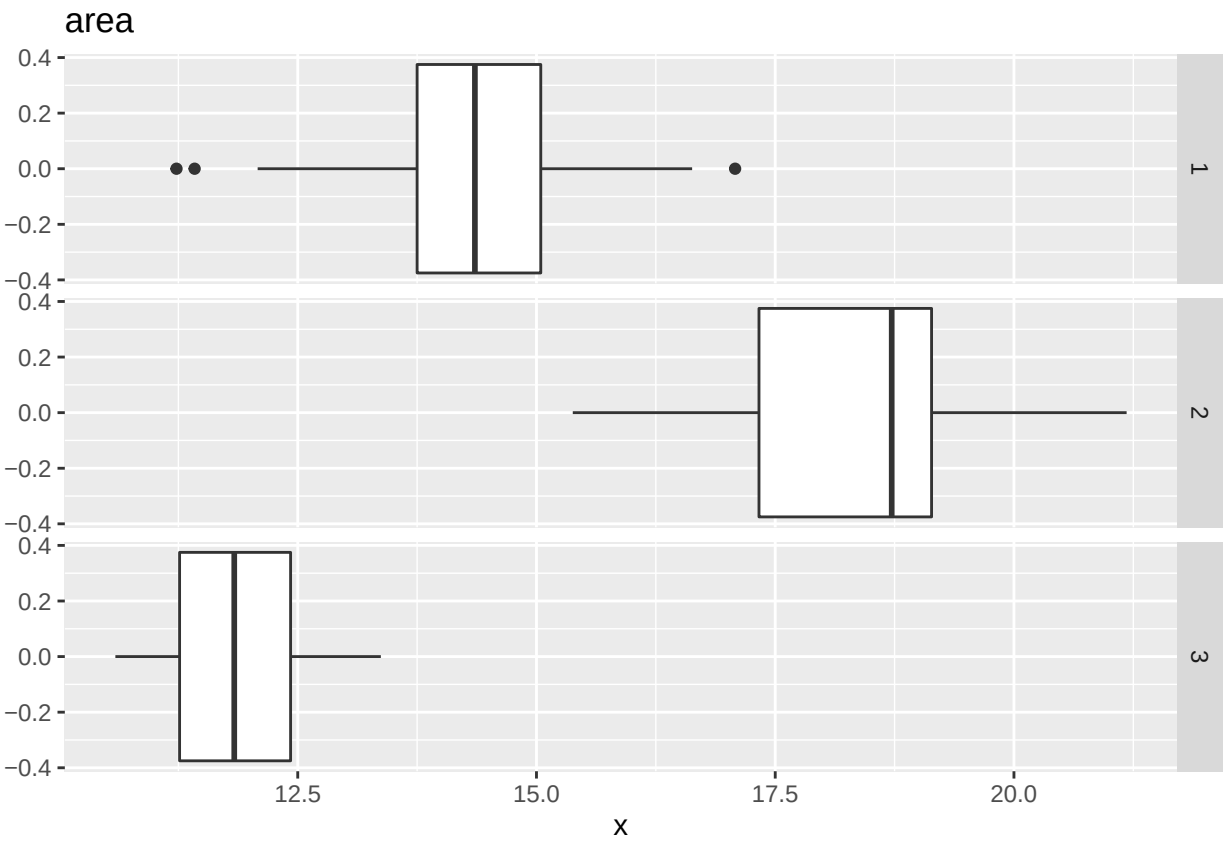
(a) Data preprocessing

```
df <- read.csv("seeds_dataset.csv")[-1]

names(df) <- c("area", "perim", "compact", "len_k", "width", "asym", "len_kg", "class")

normalized <- scale(df)
```

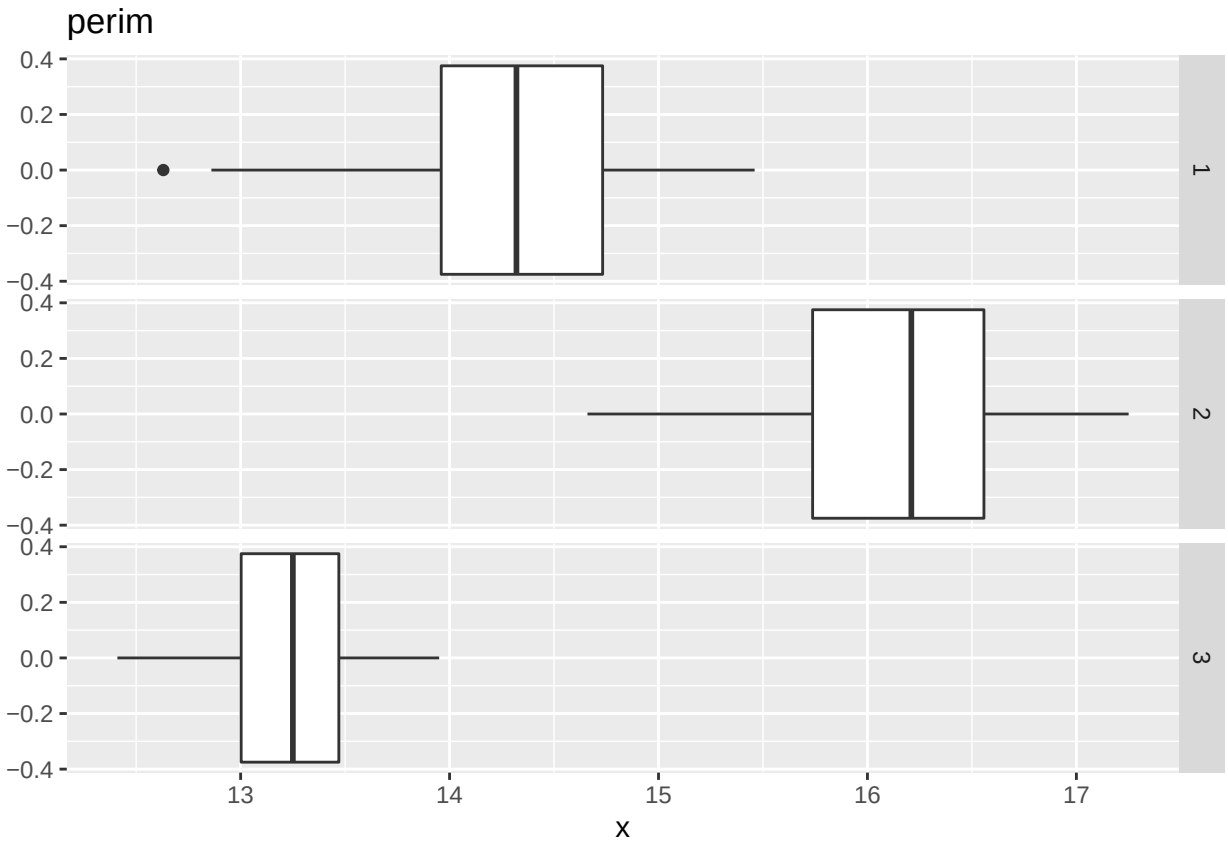
```
# pairs(df, lower.panel = NULL)

boxplotter <- function(x) {
  ggplot(df, aes(x)) +
    geom_boxplot() +
    facet_grid(vars(class))
}


i = 0
lapply(df[-8], function(x) {
  i <<- i + 1
  boxplotter(x) +
    ggtitle(names(df)[i])
  })
```

Visualizing the data

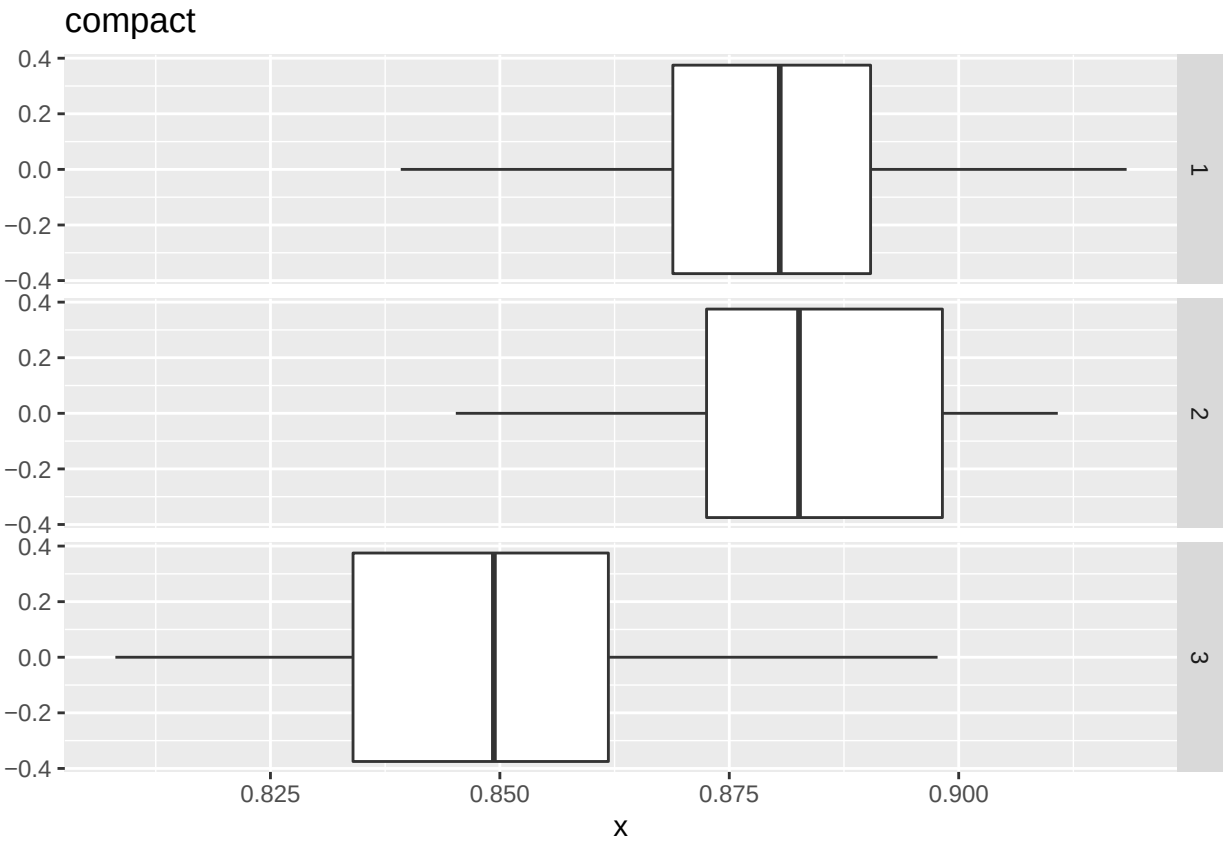
```
## $area
```



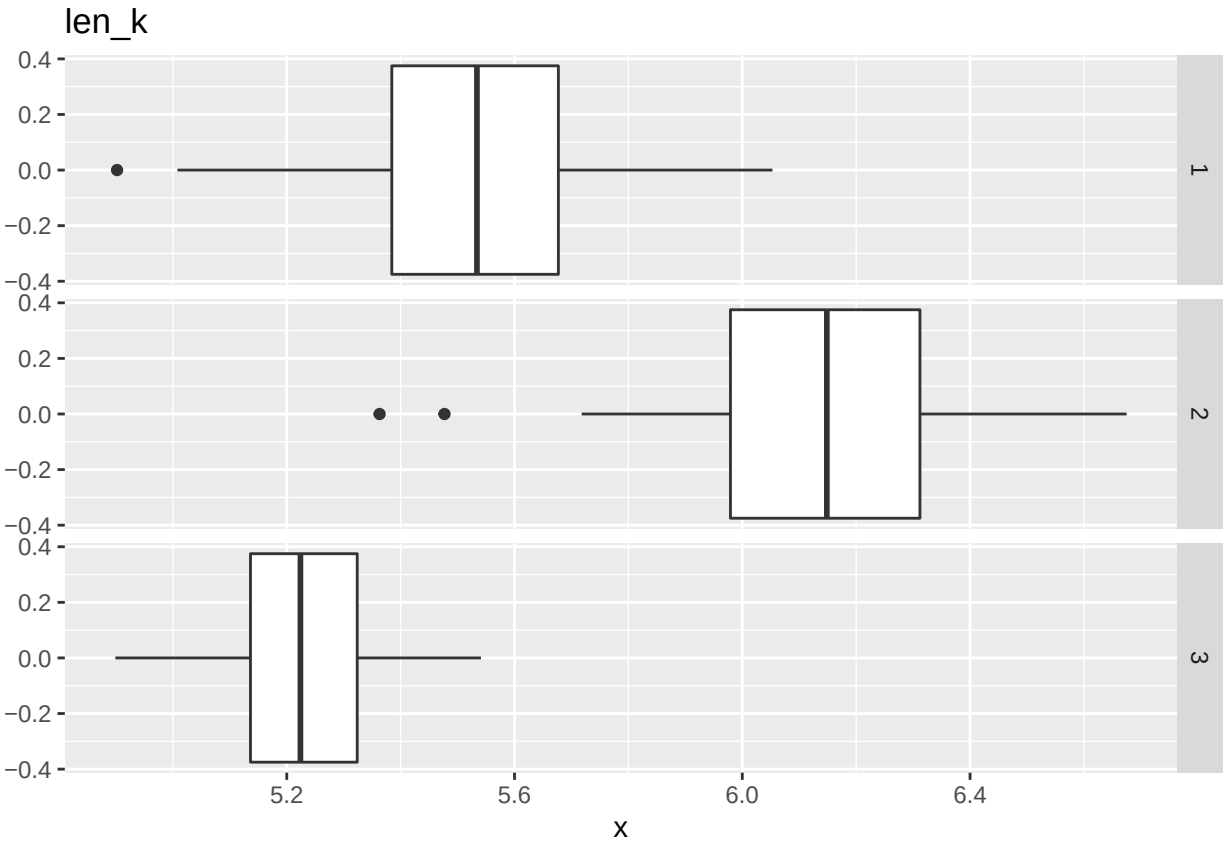

```
##
## $perim
```



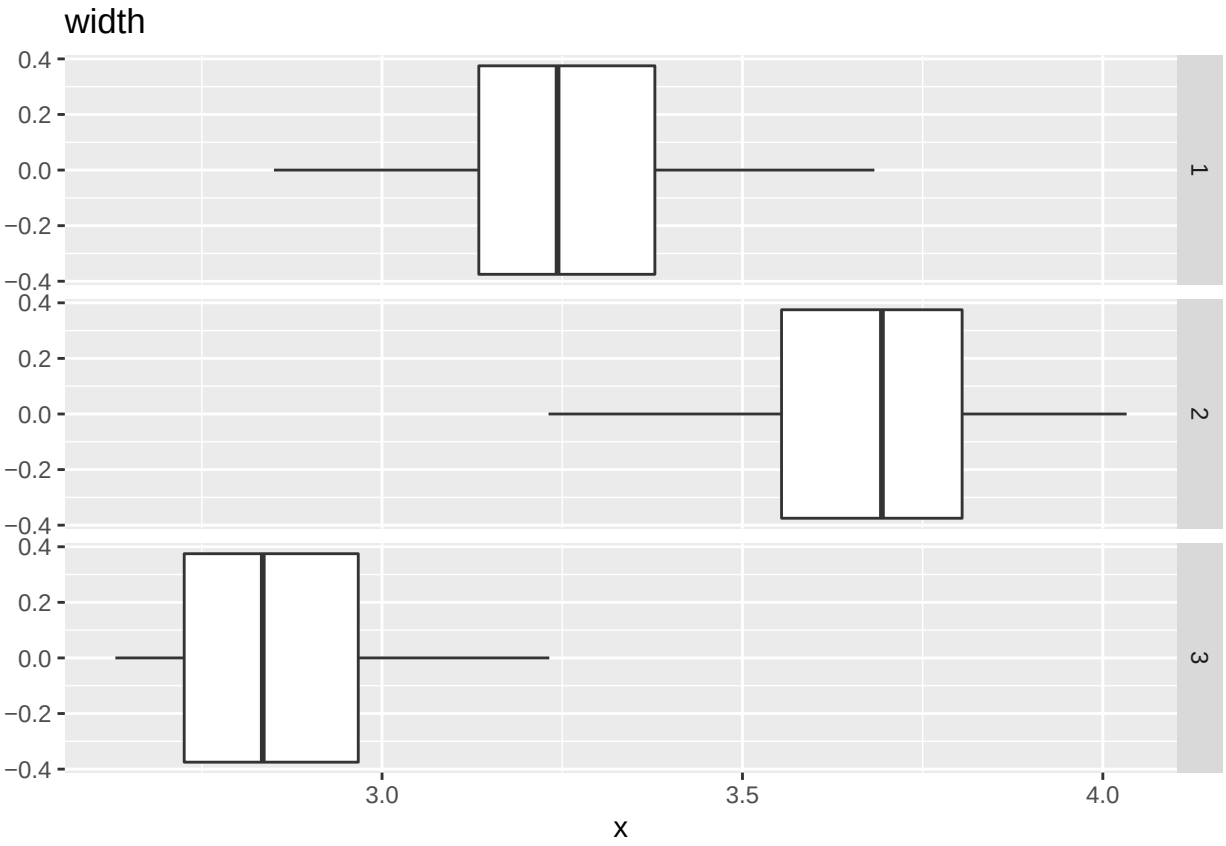

```
##
## $compact
```



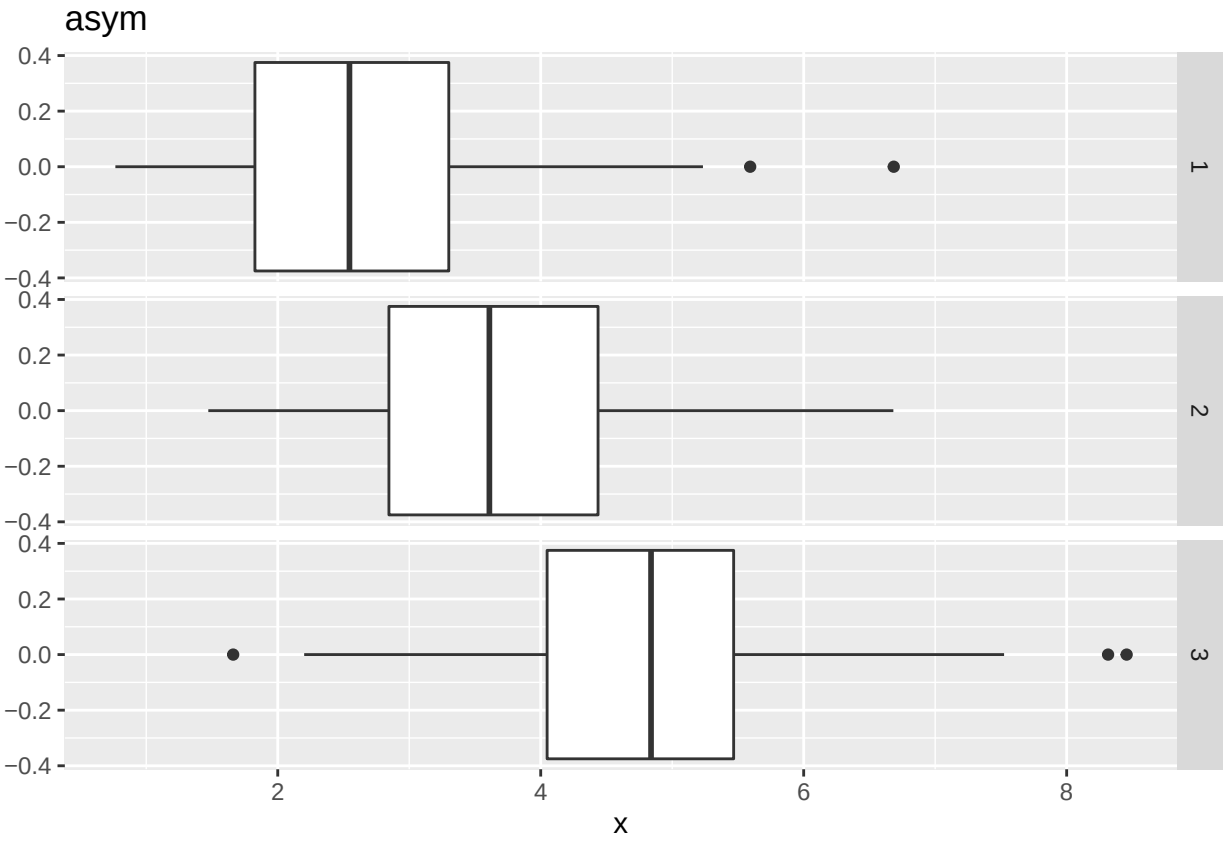

```
##
## $len_k
```




```
##
## $width
```



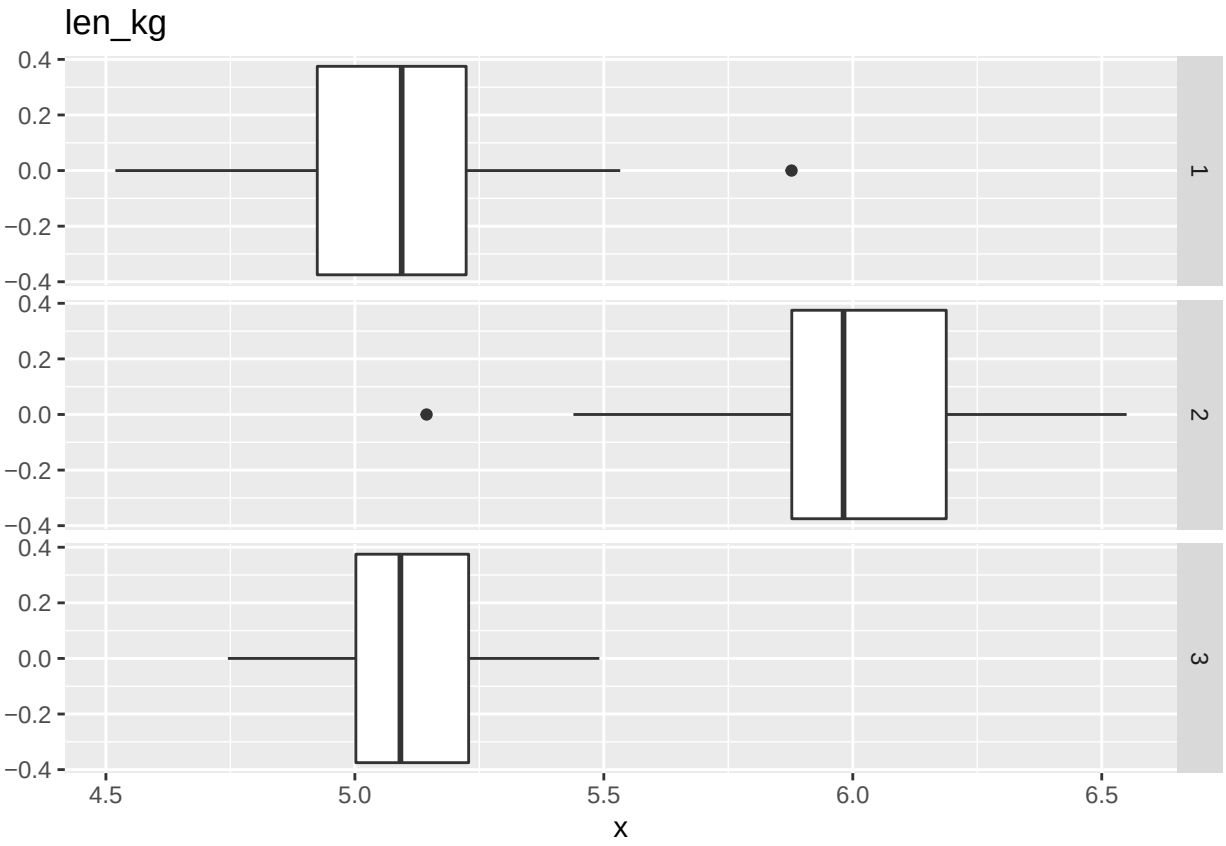

```
##
## $asym
```




```
##
## $len_kg
```

Table 1: Confusion Matrix Lloyd

1	2	3
1	60	0
60	10	2
9	0	68



k-means

```
c_lloyd <- kmeans(df[-8], 3, algorithm = "Lloyd")
c_macqueen <- kmeans(df[-8], 3, algorithm = "MacQueen")
c_forgy <- kmeans(df[-8], 3, algorithm = "Forgy")
c_har_won <- kmeans(df[-8], 3)

test <- data.frame(lloyd = c_lloyd$cluster,
                   macqueen = c_macqueen$cluster,
                   forgy = c_forgy$cluster,
                   har_won = c_har_won$cluster,
                   class = df$class)

table(test$lloyd, test$class) %>%
  kbl(caption = "Confusion Matrix Lloyd", booktabs = T)
```

```
table(test$macqueen, test$class) %>%
  kbl(caption = "Confusion Matrix MacQueen", booktabs = T)
```

Table 2: Confusion Matrix MacQueen

1	2	3
57	10	0
12	0	70
1	60	0

Table 3: Confusion Matrix Forgy

1	2	3
60	10	2
1	60	0
9	0	68

```
table(test$forgy, test$class) %>%
  kbl(caption = "Confusion Matrix Forgy", booktabs = T)
```

```
table(test$har_won, test$class) %>%
  kbl(caption = "Confusion Matrix Hartigan-Wong", booktabs = T)
```

```
cor(test)
```

```
##               lloyd   macqueen      forgy    har_won      class
## lloyd     1.0000000 -0.4578636  0.5369136  0.4578636  0.4335852
## macqueen -0.4578636  1.0000000  0.4867055 -1.0000000  0.4186104
## forgy     0.5369136  0.4867055  1.0000000 -0.4867055  0.8104053
## har_won   0.4578636 -1.0000000 -0.4867055  1.0000000 -0.4186104
## class     0.4335852  0.4186104  0.8104053 -0.4186104  1.0000000
```

```
kbl(cor(test), caption = "Correlation Matrix of k-means algorithms", booktabs = T)
```

Table 4: Confusion Matrix Hartigan-Wong

1	2	3
1	60	0
12	0	70
57	10	0

Table 5: Correlation Matrix of k-means algorithms

	lloyd	macqueen	forgy	har_won	class
lloyd	1.0000000	-0.4578636	0.5369136	0.4578636	0.4335852
macqueen	-0.4578636	1.0000000	0.4867055	-1.0000000	0.4186104
forgy	0.5369136	0.4867055	1.0000000	-0.4867055	0.8104053
har_won	0.4578636	-1.0000000	-0.4867055	1.0000000	-0.4186104
class	0.4335852	0.4186104	0.8104053	-0.4186104	1.0000000