

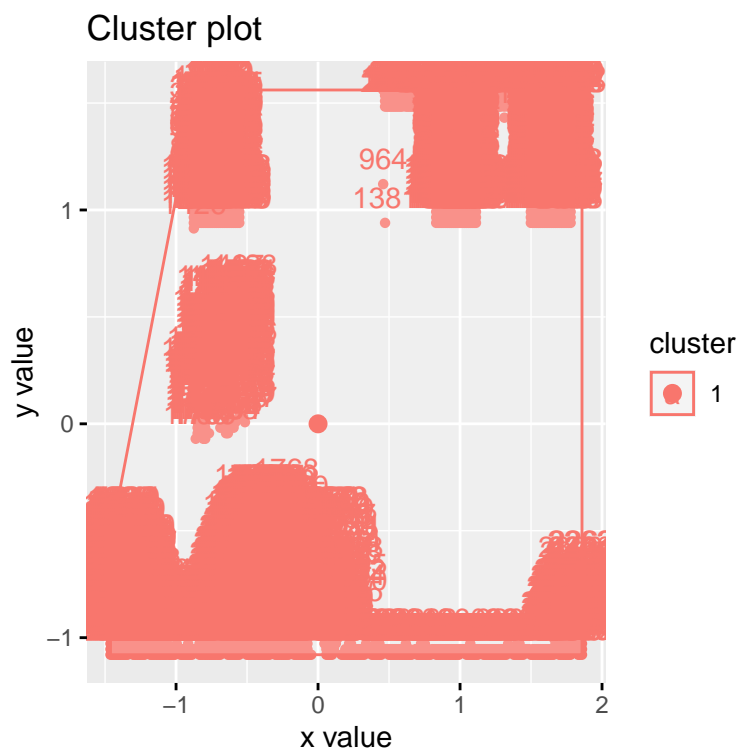
Clustering

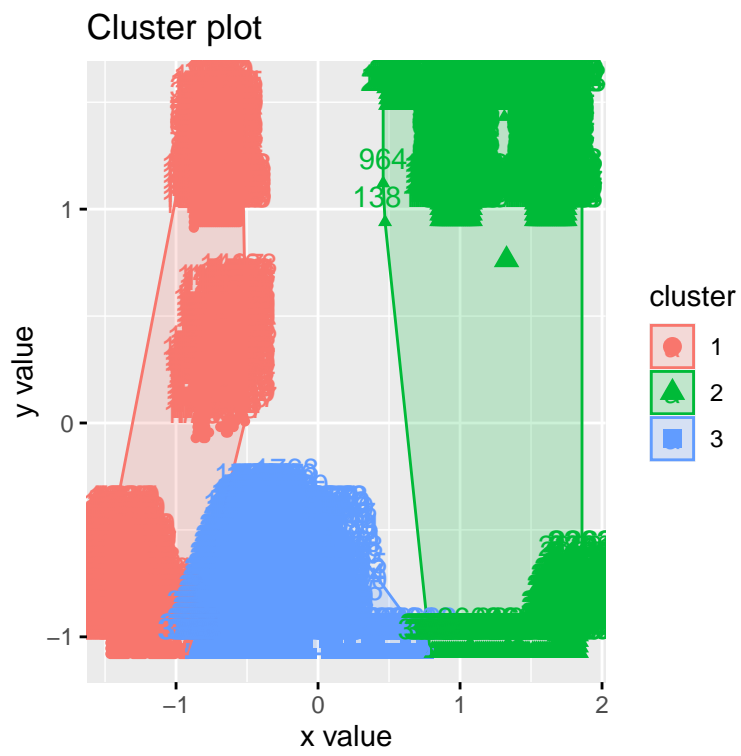
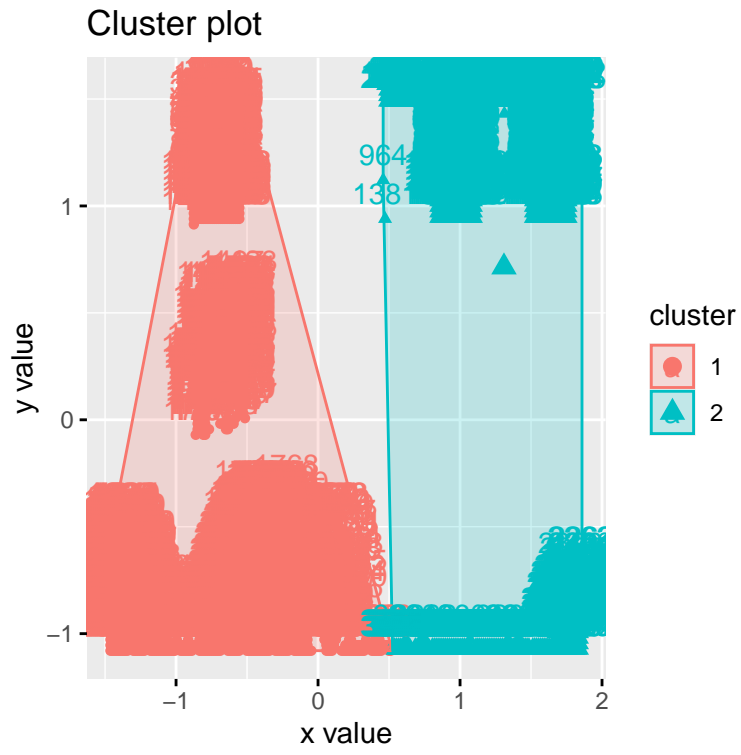
Stephen Smitshoek

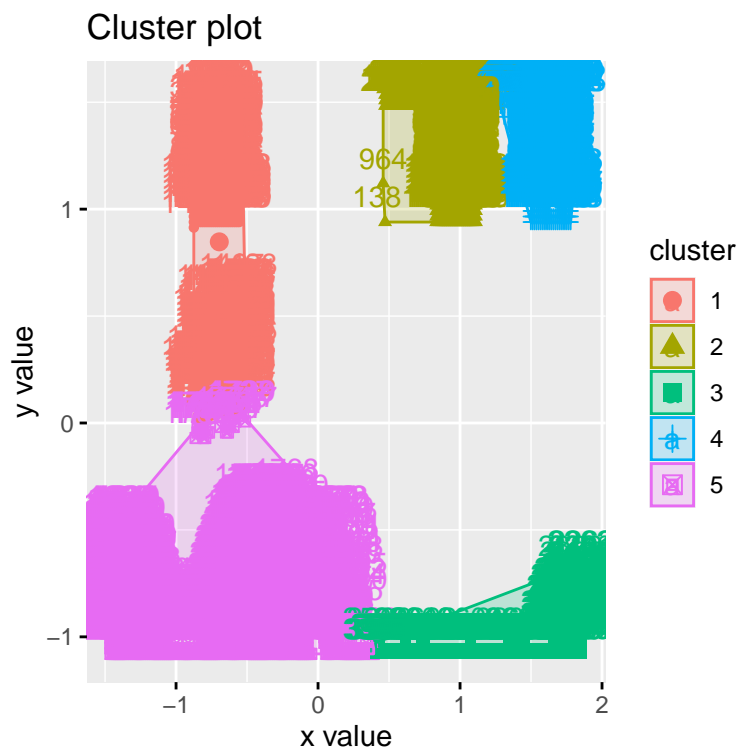
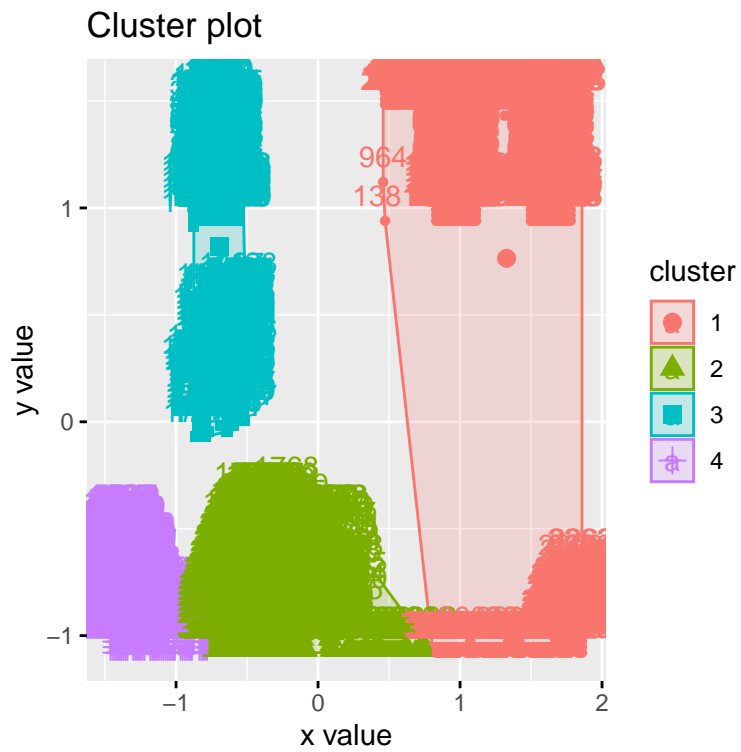
03/06/2022

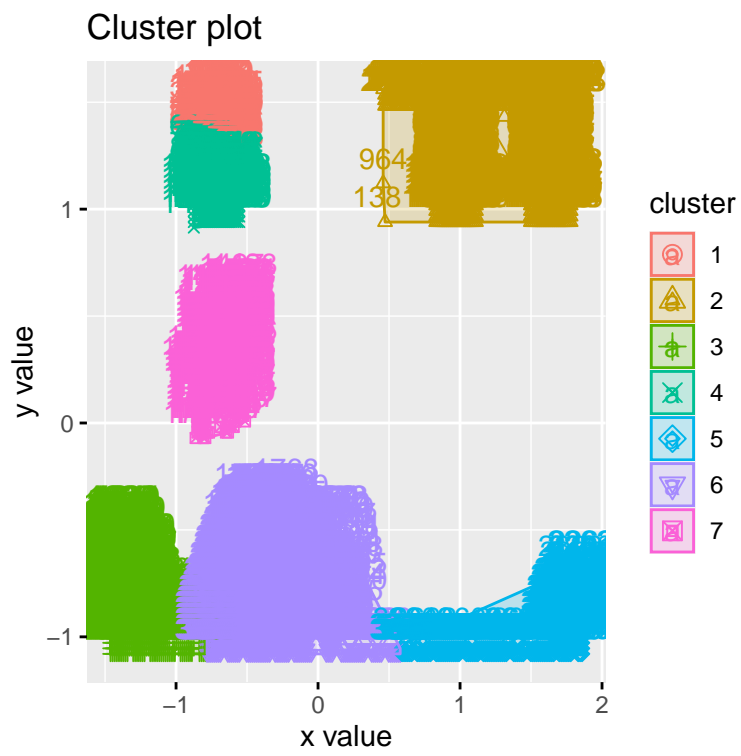
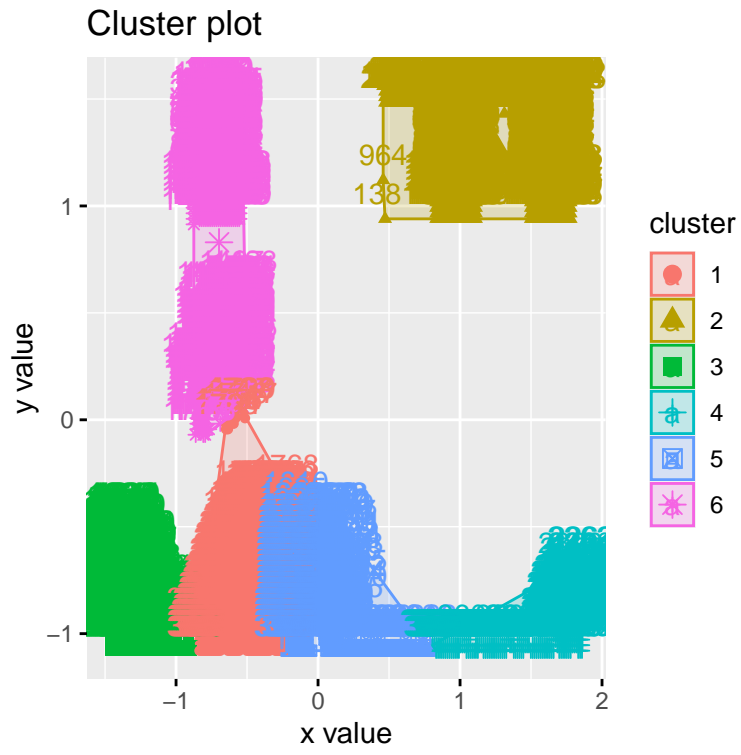
Welcome! Want to learn more? See two factoextra-related books at <https://goo.gl/ve3WBa>

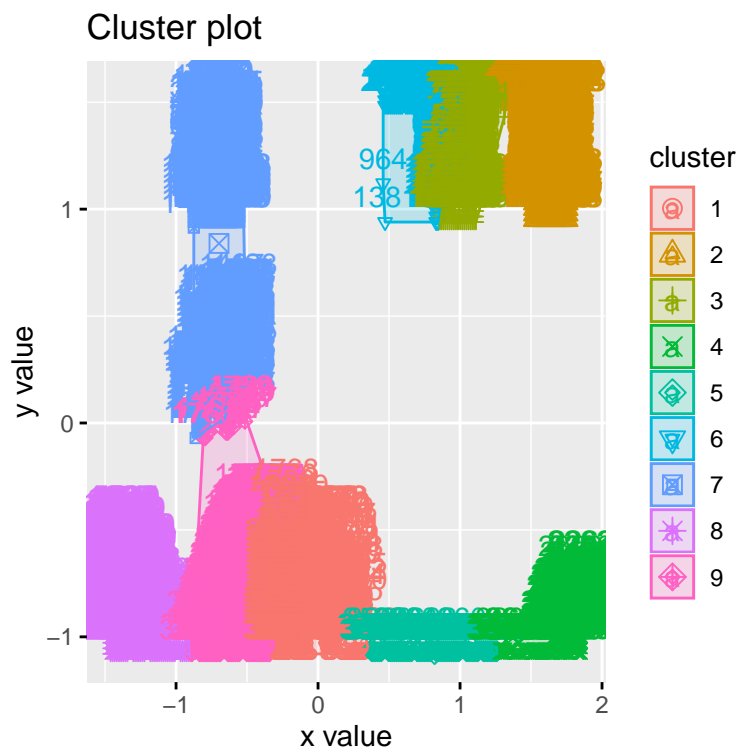
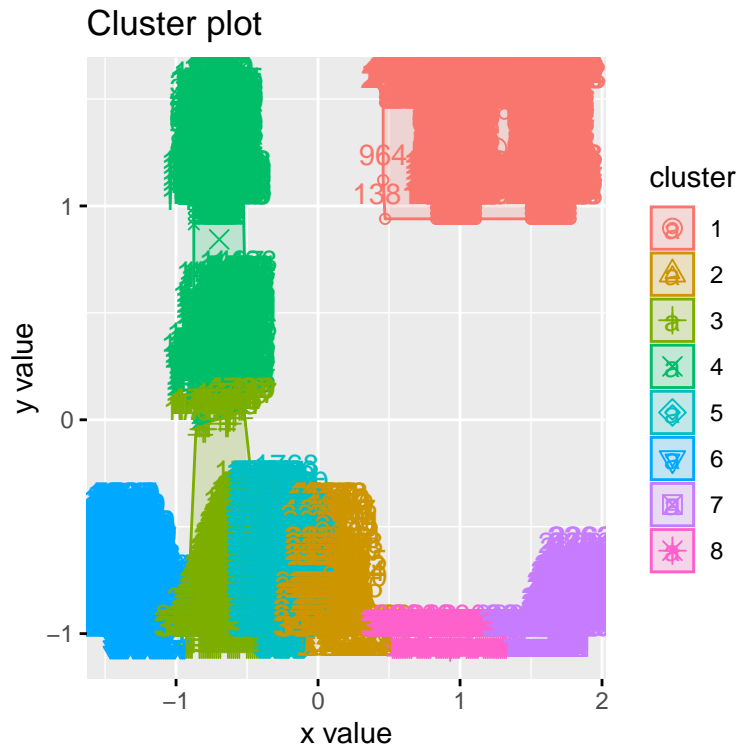
Clustered Data

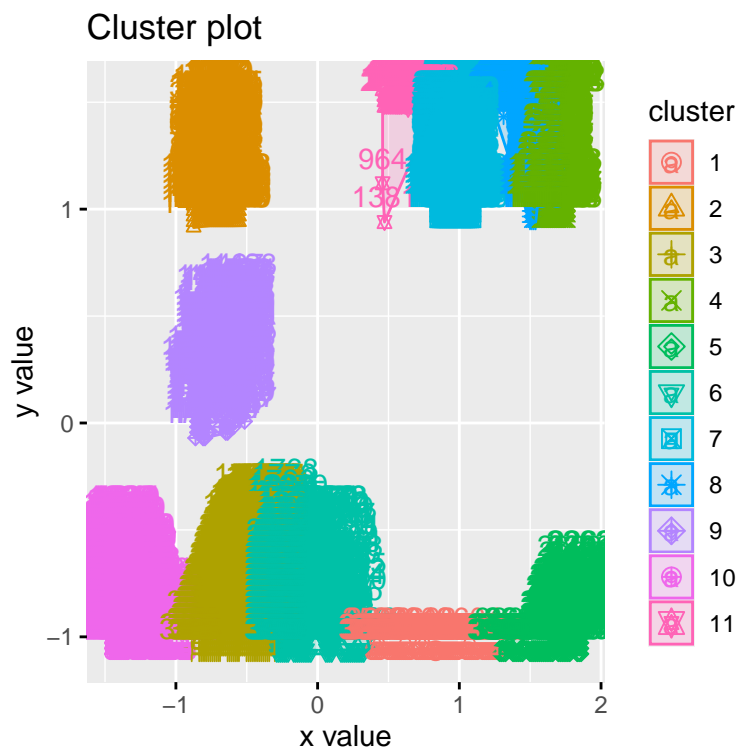
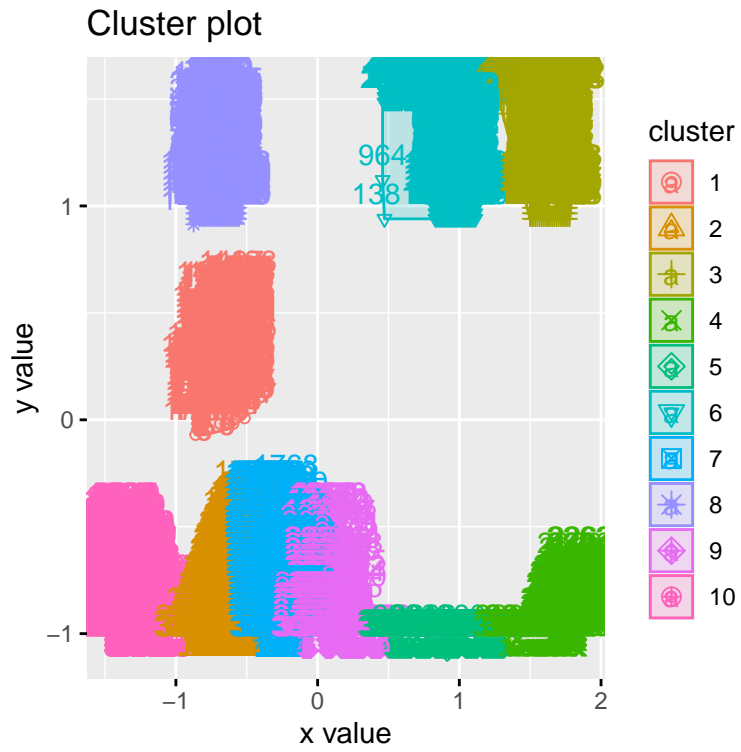




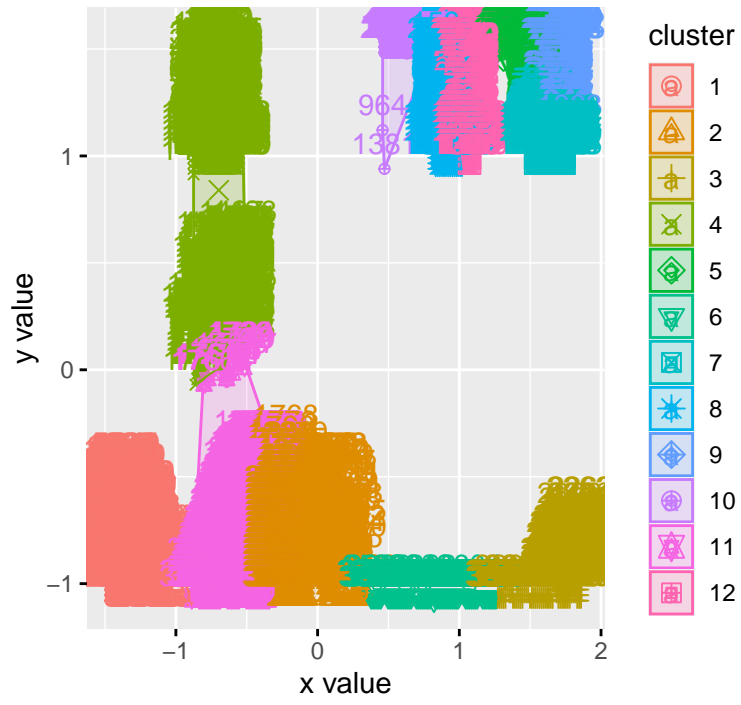




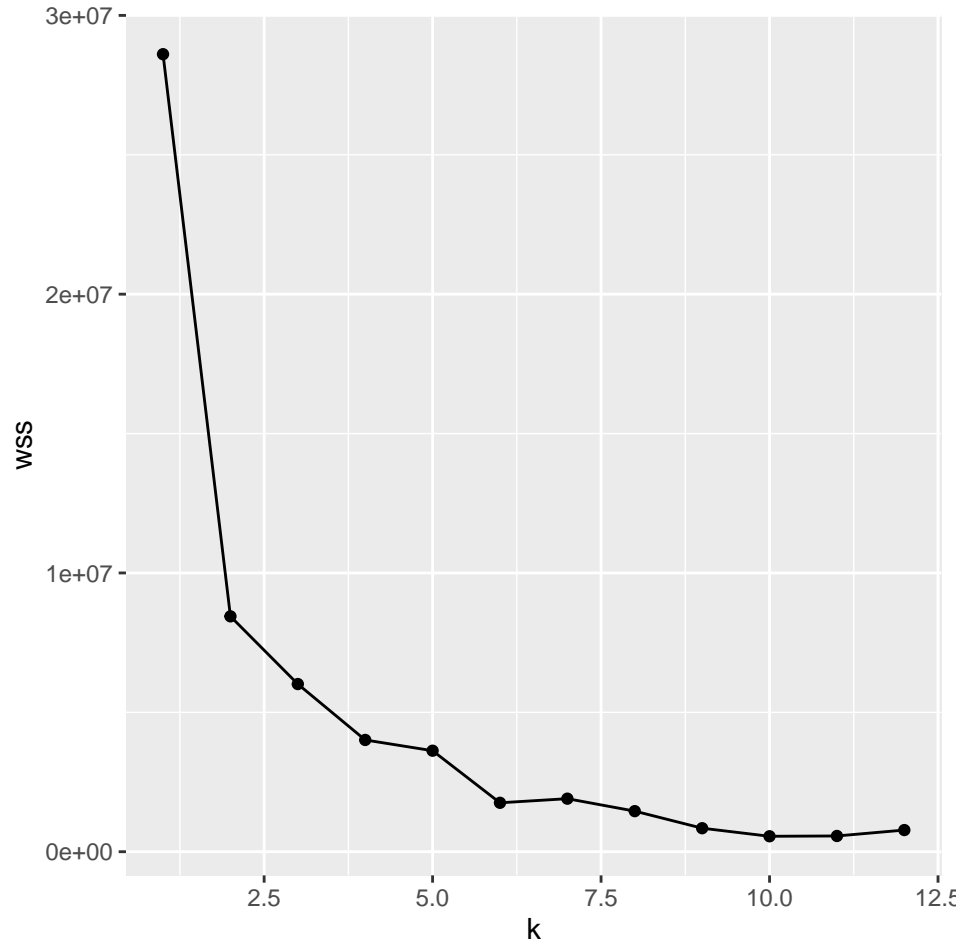




Cluster plot



Number of Clusters



The 'elbow' in the above graph occurs when k is two and so that is number of cluster that should be used for this data.

Code

```
library(ggplot2)
library(class)
library(caTools)
library(factoextra)

setwd("C:\\Users\\sksmi\\PeytoAccess\\Personal\\Bellevue\\DSC520\\dsc520")

clustering_data <- read.csv("data\\clustering-data.csv")

wss <- data.frame()
for (k in 1:12) {
  km_cluster <- kmeans(clustering_data, k)
  print(fviz_cluster(km_cluster, data=clustering_data))
  wss <- rbind(wss, c(k, km_cluster$tot.withinss))
}
colnames(wss) <- c("k", "wss")

ggplot(data=wss, aes(x=k, y=wss)) + geom_point() + geom_line()
```