Module6Assign1

Adwoa Kessie

2/26/2022

library(readr)  
library(ggplot2)  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.2

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble 3.1.5 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v purrr 0.3.4 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(tidymodels)

## Warning: package 'tidymodels' was built under R version 4.1.2

## Registered S3 method overwritten by 'tune':  
## method from   
## required\_pkgs.model\_spec parsnip

## -- Attaching packages -------------------------------------- tidymodels 0.1.4 --

## v broom 0.7.9 v rsample 0.1.1   
## v dials 0.0.10 v tune 0.1.6   
## v infer 1.0.0 v workflows 0.2.4   
## v modeldata 0.1.1 v workflowsets 0.1.0   
## v parsnip 0.1.7 v yardstick 0.0.9   
## v recipes 0.1.17

## Warning: package 'dials' was built under R version 4.1.2

## Warning: package 'infer' was built under R version 4.1.2

## Warning: package 'modeldata' was built under R version 4.1.2

## Warning: package 'parsnip' was built under R version 4.1.2

## Warning: package 'recipes' was built under R version 4.1.2

## Warning: package 'rsample' was built under R version 4.1.2

## Warning: package 'tune' was built under R version 4.1.2

## Warning: package 'workflows' was built under R version 4.1.2

## Warning: package 'workflowsets' was built under R version 4.1.2

## Warning: package 'yardstick' was built under R version 4.1.2

## -- Conflicts ----------------------------------------- tidymodels\_conflicts() --  
## x scales::discard() masks purrr::discard()  
## x dplyr::filter() masks stats::filter()  
## x recipes::fixed() masks stringr::fixed()  
## x dplyr::lag() masks stats::lag()  
## x yardstick::spec() masks readr::spec()  
## x recipes::step() masks stats::step()  
## \* Search for functions across packages at https://www.tidymodels.org/find/

trucks <- read\_csv("~/UNCW/BAN 502/Module 6/trucks.csv")

## Rows: 4000 Columns: 3

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## dbl (3): Driver\_ID, Distance, Speeding

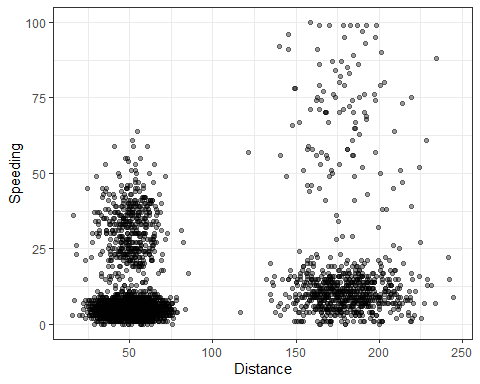
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

ggplot(trucks, aes(Distance, Speeding)) +  
 geom\_point(alpha = 0.4) + theme\_bw()



## Yes I can see some natural clusters being created here. There are 3 definite clusters, but could also be one more where the speed and distance are both high.

We also scale the data. This is critical for quantitative data to ensure that no variable (particularly a variable with large values, skews the data and the resulting clusters).

trucks\_cleaned = scale(trucks)   
summary(trucks\_cleaned)

## Driver\_ID Distance Speeding   
## Min. :-1.7314 Min. :-1.1319 Min. :-0.7821   
## 1st Qu.:-0.8657 1st Qu.:-0.5759 1st Qu.:-0.4903   
## Median : 0.0000 Median :-0.4248 Median :-0.3444   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.8657 3rd Qu.:-0.1947 3rd Qu.:-0.1255   
## Max. : 1.7314 Max. : 3.1560 Max. : 6.5127

#scale works by calculating the mean and standard deviation of the entire variable, then scales each element by subtracting the mean and dividing by the standard deviation

set.seed(64)  
kclust = kmeans(trucks\_cleaned, centers = 2) #run k-means clustering with k = 3  
kclust #view results

## K-means clustering with 2 clusters of sizes 1991, 2009  
##   
## Cluster means:  
## Driver\_ID Distance Speeding  
## 1 -0.8642756 0.1055964 0.03207659  
## 2 0.8565319 -0.1046503 -0.03178919  
##   
## Clustering vector:  
## [1] 1 2 2 1 1 2 2 1 1 2 1 2 2 1 1 2 2 1 1 2 1 2 1 2 1 2 1 1 2 1 1 1 2 2 1 1 1  
## [38] 1 2 1 2 2 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 2 2 2 1 1 1  
## [75] 2 2 1 1 2 1 2 1 1 2 2 1 1 2 1 2 1 1 2 1 2 2 1 2 1 1 1 1 1 2 2 2 2 1 1 2 2  
## [112] 1 1 1 2 2 1 1 1 2 1 1 1 1 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 1 2 2 2 2 2 2 1 1  
## [149] 2 2 2 1 2 1 1 2 1 2 1 2 1 2 1 1 2 2 2 2 1 2 1 2 2 1 2 2 1 1 1 1 1 1 1 2 2  
## [186] 2 1 2 1 2 2 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 1 1 2 2 2 2 2 1 2 2 1 2 1 2 1 2  
## [223] 1 1 1 2 1 2 1 2 2 2 1 2 2 2 1 1 1 2 2 2 1 1 2 2 1 1 1 2 2 1 2 1 2 2 2 2 1  
## [260] 2 1 2 2 1 1 1 2 1 1 1 1 2 2 1 1 2 1 1 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 1 1 1  
## [297] 1 1 2 2 1 2 2 1 1 2 2 2 2 2 1 2 2 1 2 1 2 2 1 1 2 1 1 1 2 1 1 2 1 2 2 2 1  
## [334] 1 1 1 2 2 2 1 1 1 2 1 2 2 2 2 2 1 2 1 2 1 1 2 1 2 2 1 1 1 2 1 1 1 1 2 2 1  
## [371] 1 2 1 2 1 1 1 1 1 1 2 2 1 1 1 2 1 2 2 2 2 1 1 1 2 2 2 2 1 2 1 2 1 1 1 1 1  
## [408] 1 1 1 1 2 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2 1 2 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2  
## [445] 2 1 1 1 2 2 2 1 1 2 1 2 2 2 2 2 2 1 1 1 2 2 1 2 2 2 2 1 1 1 2 2 2 1 2 1 1  
## [482] 1 1 2 1 2 1 2 2 1 2 1 1 2 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 1 1 1 2 1 1 1 2 1  
## [519] 2 2 2 2 1 2 1 1 2 1 1 1 2 1 1 2 2 2 1 1 1 2 2 2 2 1 1 2 2 1 1 2 1 1 2 2 1  
## [556] 1 1 1 1 1 1 2 1 2 2 2 1 2 2 2 1 2 1 1 2 1 1 1 2 1 2 2 1 2 2 2 2 2 1 1 2 2  
## [593] 1 2 2 1 2 2 1 2 2 1 1 1 2 2 1 2 2 2 1 1 2 2 1 1 1 2 2 2 1 2 1 2 1 2 2 2 1  
## [630] 1 1 1 2 2 1 2 1 1 2 1 2 2 2 2 1 1 1 2 1 1 2 2 1 1 2 2 1 2 1 2 2 1 2 2 1 2  
## [667] 1 2 2 1 1 2 2 2 2 1 1 1 2 2 1 1 2 1 1 1 1 2 1 2 1 1 1 2 2 2 1 2 1 2 1 2 2  
## [704] 1 2 2 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 2 1 1 2 2 1 1 2 1 2 2 2 1 1 2 1 1 2 2  
## [741] 2 1 2 1 2 1 2 2 2 1 1 1 1 2 1 2 1 2 1 1 1 1 2 2 2 2 1 2 2 1 2 1 1 1 2 1 2  
## [778] 1 1 2 2 1 1 1 1 2 1 2 2 2 2 1 1 2 2 1 1 1 1 2 1 1 2 2 1 2 1 2 1 2 2 1 2 1  
## [815] 2 2 2 1 2 1 2 2 2 1 1 2 2 2 1 2 2 1 1 1 1 2 1 2 2 2 2 1 1 1 2 1 2 2 2 1 2  
## [852] 1 2 2 2 1 2 1 2 2 2 2 1 2 1 1 2 1 1 2 1 2 2 1 2 2 2 2 1 2 1 2 2 1 1 1 1 2  
## [889] 2 2 2 1 2 1 2 1 1 2 2 1 2 2 2 1 1 1 2 1 1 2 2 2 1 2 1 1 1 2 1 2 2 1 1 2 2  
## [926] 2 1 1 1 1 2 1 1 1 2 1 2 1 2 2 1 2 1 1 2 1 2 1 2 2 1 2 1 1 1 1 1 1 2 2 2 2  
## [963] 2 2 1 2 1 1 2 2 2 2 1 2 2 1 1 1 1 2 2 1 2 1 2 1 1 2 2 1 1 1 1 2 2 2 2 2 1  
## [1000] 2 1 2 1 1 1 1 1 2 2 2 2 2 1 2 2 1 1 2 2 2 2 1 2 1 2 2 1 1 1 1 1 1 2 2 1 1  
## [1037] 2 1 2 1 2 1 1 1 2 2 1 2 2 1 1 1 2 1 2 1 1 2 2 2 1 2 1 2 2 1 2 2 2 1 2 2 2  
## [1074] 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1 1 2 2 1 1 1 1 2 2 1 1 1 2 2 2  
## [1111] 1 2 1 2 2 1 2 1 2 1 1 1 1 2 1 2 2 2 1 2 2 1 2 1 2 2 1 2 2 2 2 1 2 2 1 1 2  
## [1148] 1 2 1 2 2 2 1 1 2 1 1 1 1 1 2 2 1 2 2 2 1 2 2 2 1 2 1 1 1 1 2 1 2 1 2 1 2  
## [1185] 2 2 2 1 2 1 1 1 2 1 2 2 1 2 1 1 1 1 1 2 2 2 1 2 2 1 1 2 2 1 2 2 1 1 2 1 2  
## [1222] 1 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 2 1 1 2 2 1 2 2 1 2 2 2 1 1 2 1 1 2  
## [1259] 2 2 2 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 2 2 2 2 1 1 2 2 1 1 1 1 1 1 2 1 2 2  
## [1296] 2 2 1 1 2 1 1 2 1 2 1 2 2 1 2 1 2 2 1 2 1 1 1 2 1 2 1 2 2 2 2 1 1 2 2 2 1  
## [1333] 2 1 2 1 2 1 2 2 2 1 2 1 2 2 1 1 1 2 1 1 2 1 2 1 2 2 1 2 2 1 2 2 2 1 2 2 1  
## [1370] 2 2 2 2 2 1 2 1 1 1 2 1 1 1 2 2 1 2 2 1 2 1 1 2 1 1 1 2 1 1 1 2 1 2 1 1 1  
## [1407] 2 1 2 2 2 1 2 1 2 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 2 2 1 2 1 1 1 1 1 2 2 1 2  
## [1444] 2 2 1 2 1 1 1 2 2 1 2 2 1 2 1 2 2 2 2 2 1 1 2 2 1 2 1 1 1 1 2 2 1 2 1 2 1  
## [1481] 2 1 2 2 1 2 1 1 2 2 2 2 2 2 1 2 1 2 1 1 2 2 2 2 1 1 1 2 2 2 1 2 2 1 1 1 2  
## [1518] 1 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 2 2 1 1 2 2 2 1 2 1 2 2 1 1 2 2 2 2 2 2 1  
## [1555] 2 2 2 1 2 2 2 2 1 2 2 1 1 1 1 1 1 2 1 2 2 2 1 2 1 1 2 2 1 2 1 1 2 2 2 1 2  
## [1592] 1 2 1 2 1 2 2 1 2 1 2 2 2 1 2 2 1 2 2 2 2 1 2 2 1 1 1 2 2 1 2 1 2 2 1 2 2  
## [1629] 1 1 2 1 2 1 1 2 1 2 2 2 1 2 2 2 1 1 2 2 2 1 2 1 1 1 2 1 2 1 2 1 2 1 2 2 1  
## [1666] 1 2 2 1 1 2 1 1 1 1 2 1 1 2 1 1 1 2 1 1 2 2 1 1 2 2 2 2 2 2 1 2 2 2 1 2 2  
## [1703] 1 2 2 1 2 2 2 1 1 2 1 2 2 1 1 2 1 2 2 2 1 2 1 1 2 2 1 2 1 2 2 2 2 2 1 2 1  
## [1740] 2 2 1 2 1 2 1 2 1 2 1 2 1 1 2 2 1 2 1 1 2 2 1 2 2 2 2 2 2 1 1 2 2 1 2 1 2  
## [1777] 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 1 1 1 2 1 1 1 2 2 2 2 1 1 2 2 1 1 2 1 1 2 2  
## [1814] 2 2 1 1 2 1 1 2 2 2 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2 2 1 2 1 1 1 2 1 1 2  
## [1851] 2 2 1 2 2 1 1 1 2 2 1 1 1 2 2 2 2 2 1 2 1 1 2 1 2 1 2 1 1 1 2 1 1 1 2 1 1  
## [1888] 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2 1  
## [1925] 1 1 2 1 2 2 1 2 2 2 2 1 1 2 2 1 1 2 2 1 2 1 2 2 2 1 1 1 2 2 2 1 1 2 2 2 1  
## [1962] 2 1 2 1 2 1 1 1 2 2 1 1 1 2 1 2 2 2 2 1 2 2 1 2 2 1 1 1 2 2 2 2 2 2 2 1 1  
## [1999] 2 1 2 1 2 2 2 2 1 1 1 1 1 1 1 2 1 1 2 2 2 1 2 1 1 2 2 2 1 2 2 1 1 1 2 2 2  
## [2036] 2 1 2 1 2 2 2 1 2 1 1 2 1 2 2 2 2 1 1 2 1 1 1 1 1 2 1 1 2 1 1 1 1 2 2 1 2  
## [2073] 2 1 2 1 2 2 2 2 2 2 1 2 1 2 1 2 1 1 2 1 1 2 1 2 2 1 1 1 2 1 2 1 1 1 1 1 2  
## [2110] 1 2 2 2 2 2 1 2 2 2 2 2 1 1 1 1 2 2 1 1 1 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2  
## [2147] 2 2 2 1 1 2 2 1 2 1 2 2 2 1 1 2 1 2 1 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 1 2 2  
## [2184] 1 2 1 2 2 1 2 2 1 2 1 2 2 2 2 2 2 1 2 1 2 1 2 1 2 2 1 2 2 2 1 1 1 1 1 2 1  
## [2221] 1 2 2 1 2 2 2 2 1 1 2 2 2 1 1 2 1 1 1 2 2 1 2 1 2 2 2 1 2 1 1 2 2 1 1 2 1  
## [2258] 2 1 1 1 1 1 1 2 1 1 2 2 1 1 1 2 2 2 2 2 2 1 1 2 2 1 2 1 2 1 1 2 1 2 1 2 2  
## [2295] 2 1 2 2 2 1 2 1 2 2 1 2 1 1 1 1 1 2 1 1 2 1 2 2 2 1 1 1 2 1 2 2 2 2 1 2 1  
## [2332] 2 2 1 1 1 1 2 1 2 2 1 1 1 2 1 2 2 2 1 1 2 1 2 2 1 2 2 1 1 1 1 1 1 2 1 2 2  
## [2369] 1 2 2 1 2 2 2 2 1 2 2 2 1 2 1 2 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 2 2  
## [2406] 2 2 2 2 2 1 2 1 2 2 1 1 2 2 1 1 1 2 2 2 1 2 1 2 1 2 2 2 1 1 2 2 2 2 1 1 1  
## [2443] 1 2 2 1 1 1 1 2 2 1 1 2 2 1 1 1 2 2 1 2 1 1 1 1 2 1 1 1 2 2 1 2 2 2 1 2 1  
## [2480] 2 1 1 2 1 1 2 1 2 1 2 2 1 2 2 1 1 1 2 2 1 2 2 1 1 1 1 2 1 2 1 2 2 1 1 2 2  
## [2517] 1 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1 2 1 1 1 1  
## [2554] 2 1 2 1 2 2 1 2 1 1 2 2 2 2 1 1 2 1 2 1 1 2 2 1 2 2 2 2 1 2 2 1 1 2 2 2 1  
## [2591] 2 2 1 2 2 1 1 1 1 2 1 1 1 2 1 1 1 2 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 2 1 1 2  
## [2628] 1 2 1 1 1 2 1 1 1 2 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 2 1 2 2 2 1 1 1 1 2 2 2  
## [2665] 1 2 1 2 2 2 2 1 2 2 2 2 1 2 2 1 1 2 2 2 1 2 2 1 2 1 1 1 1 1 2 1 2 2 2 1 2  
## [2702] 1 2 2 1 1 1 2 2 2 1 2 1 2 2 2 2 1 1 2 2 1 1 2 1 1 1 2 1 1 2 1 2 2 1 1 2 1  
## [2739] 1 1 1 2 1 1 1 1 1 1 2 2 2 2 2 1 2 1 1 2 2 1 2 2 1 2 1 2 2 1 1 1 2 2 2 1 1  
## [2776] 2 1 2 1 1 1 1 2 2 1 2 1 2 1 2 2 1 2 1 1 1 2 2 1 1 2 2 2 2 2 1 2 1 1 2 1 2  
## [2813] 2 2 2 2 1 1 1 2 1 2 2 1 2 2 2 1 1 2 1 2 2 1 2 2 2 2 1 2 2 1 2 2 2 2 2 2 1  
## [2850] 1 1 1 1 2 1 1 2 1 1 2 2 1 1 2 2 2 1 2 2 1 1 1 2 2 2 1 2 2 1 2 2 1 1 2 2 1  
## [2887] 2 2 1 2 1 1 2 1 2 2 1 1 1 2 1 1 2 2 2 2 2 2 1 2 1 1 1 2 1 1 2 2 1 2 2 2 1  
## [2924] 1 2 1 1 1 2 1 2 1 2 1 2 2 1 2 2 2 2 2 2 2 1 1 1 2 1 2 1 1 1 1 1 1 1 2 2 1  
## [2961] 1 2 2 1 1 2 1 1 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 1 1 2 1 1 1  
## [2998] 2 1 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 1 2 2 1 1 1 2 2 2 2 1 1 1 1 1 1 1 2  
## [3035] 2 2 2 2 2 1 2 1 2 2 1 1 2 2 1 1 2 2 1 2 1 2 2 1 2 2 1 1 2 2 1 2 1 1 2 2 1  
## [3072] 2 1 2 2 2 2 1 2 2 1 2 1 1 1 2 2 2 2 2 2 2 2 1 2 1 1 1 1 1 2 1 1 1 2 2 1 1  
## [3109] 1 2 2 1 1 2 1 2 2 1 1 2 1 2 2 1 1 2 1 1 1 2 2 2 1 2 2 2 2 2 2 1 2 1 2 1 1  
## [3146] 2 2 1 1 2 2 1 2 2 1 2 2 2 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 2 2 1 2 1 1  
## [3183] 2 2 1 1 1 1 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 2  
## [3220] 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 2 2 1 2 1 2 2 2 2 1 2 1 1 2 1 2 2 1 2 2 2 2  
## [3257] 1 2 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 2 1 2 1 2 1 2 1 2 1  
## [3294] 2 2 1 1 2 1 1 1 2 1 2 1 2 1 1 1 2 1 1 2 2 2 2 1 2 1 1 2 2 2 1 2 1 1 2 1 1  
## [3331] 1 2 1 1 2 1 2 1 1 1 1 2 1 2 1 1 1 1 2 1 1 2 1 2 1 1 1 2 1 1 2 1 1 1 2 1 1  
## [3368] 2 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 2 2 1 2 2 1 2 1 2 1  
## [3405] 1 2 1 2 2 1 1 1 1 1 1 2 1 1 2 2 2 1 1 2 2 1 1 1 2 1 1 1 1 1 2 2 1 2 2 1 2  
## [3442] 1 1 1 1 1 2 2 1 2 1 2 2 1 2 2 1 1 2 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 2 1 1 1  
## [3479] 1 2 2 1 2 1 1 2 1 1 1 2 1 1 1 2 2 2 1 2 1 1 2 1 1 1 1 1 1 1 2 2 2 2 2 2 1  
## [3516] 1 1 1 2 2 1 1 1 2 2 1 2 1 1 2 2 1 1 2 1 1 1 2 1 1 2 1 1 1 2 2 2 2 1 2 2 1  
## [3553] 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 2 1 2 1 2 2 1 2 2 1 1 1 2 2 2 2 1 1 2 1 2 1  
## [3590] 1 1 1 1 1 1 1 1 1 2 1 1 2 2 2 1 2 1 2 1 2 1 1 1 1 2 2 1 1 2 1 2 1 1 1 1 1  
## [3627] 2 1 1 1 1 2 1 2 1 1 1 1 2 2 2 2 1 1 1 1 2 1 1 1 1 2 2 2 1 1 2 2 2 1 2 1 1  
## [3664] 1 2 1 2 2 1 1 2 1 1 2 2 1 1 1 2 2 2 1 1 1 2 2 2 1 1 1 2 1 2 1 1 1 2 2 2 1  
## [3701] 1 1 2 1 1 1 1 1 2 1 2 1 2 2 2 2 1 2 1 1 2 1 1 1 2 1 1 1 1 2 2 1 1 2 2 1 1  
## [3738] 2 1 2 2 2 2 1 2 1 2 2 1 1 2 1 1 1 2 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 2  
## [3775] 2 1 1 1 1 1 2 2 1 2 2 1 2 2 2 1 1 1 1 2 2 1 1 2 2 2 1 2 1 1 2 1 1 1 2 1 1  
## [3812] 1 2 1 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 2 2 1 1 1 1 2 1 2 1 1 1 1 1  
## [3849] 1 2 2 2 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 1 1 2 1 2 1 2 1 2 1 1 2  
## [3886] 2 1 1 2 1 1 1 2 1 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 1 1 2 1  
## [3923] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 2 2 1 2 1 1  
## [3960] 1 1 1 1 1 2 2 1 1 1 2 2 2 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1  
## [3997] 1 2 2 1  
##   
## Within cluster sum of squares by cluster:  
## [1] 4952.508 4035.092  
## (between\_SS / total\_SS = 25.1 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

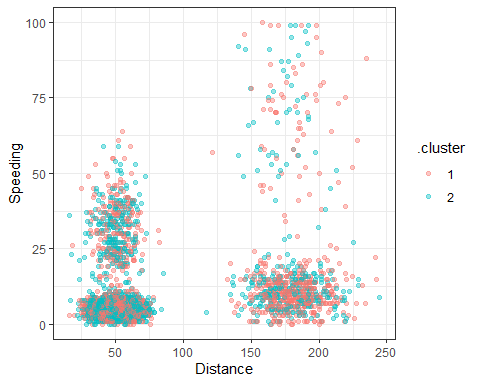
Add the cluster assignment to the dataset

trucks = augment(kclust, trucks)  
trucks

## # A tibble: 4,000 x 4  
## Driver\_ID Distance Speeding .cluster  
## <dbl> <dbl> <dbl> <fct>   
## 1 3423311935 71.2 28 1   
## 2 3423313212 52.5 25 2   
## 3 3423313724 64.5 27 2   
## 4 3423311373 55.7 22 1   
## 5 3423310999 54.6 25 1   
## 6 3423313857 41.9 10 2   
## 7 3423312432 58.6 20 2   
## 8 3423311434 52.0 8 1   
## 9 3423311328 31.2 34 1   
## 10 3423312488 44.3 19 2   
## # ... with 3,990 more rows

Plot the clusters

ggplot(trucks, aes(Distance, Speeding, color = .cluster)) +  
 geom\_point(alpha = 0.4) + theme\_bw()

 ## The cluster on the left has more drivers from cluster 2, while the one on the right has more drivers from cluster 1

Perform k-means clustering with a pre-specified number of clusters.

set.seed(412)  
clusts =   
 tibble(k = 1:8) %>% #try from 1 to 8 clusters  
 mutate(  
 kclust = map(k, ~kmeans(trucks\_cleaned, .x)),  
 tidied = map(kclust, tidy),  
 glanced = map(kclust, glance),  
 augmented = map(kclust, augment, trucks\_cleaned)  
 )  
  
clusts

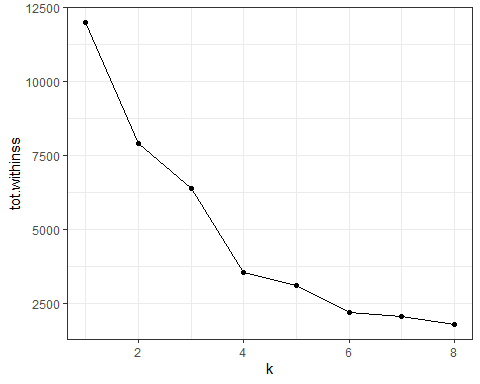
## # A tibble: 8 x 5  
## k kclust tidied glanced augmented   
## <int> <list> <list> <list> <list>   
## 1 1 <kmeans> <tibble [1 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 2 2 <kmeans> <tibble [2 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 3 3 <kmeans> <tibble [3 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 4 4 <kmeans> <tibble [4 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 5 5 <kmeans> <tibble [5 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 6 6 <kmeans> <tibble [6 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 7 7 <kmeans> <tibble [7 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>  
## 8 8 <kmeans> <tibble [8 x 6]> <tibble [1 x 4]> <tibble [4,000 x 4]>

Create relevant objects

clusters =  
 clusts %>%  
 unnest(cols = c(tidied))  
  
assignments =   
 clusts %>%   
 unnest(cols = c(augmented))  
  
clusterings =   
 clusts %>%  
 unnest(cols = c(glanced))

Because we are clustering across multiple variables (more than 2 or 3) it’s very difficult to plot the clusters in a meaningful way. However, we can look at a plot to see the performance of the clusters.

ggplot(clusterings, aes(k, tot.withinss)) +  
 geom\_line() +  
 geom\_point() + theme\_bw()

 In the plot above, we are looking for the “elbow”. This corresponds to the “best” number of clusters. For this data, 4 or 5 clusters would be appropriate. 4 is the best.

set.seed(412)  
kclust = kmeans(trucks\_cleaned, centers = 4) #run k-means clustering with k = 3  
kclust #view results

## K-means clustering with 4 clusters of sizes 1457, 694, 1522, 327  
##   
## Cluster means:  
## Driver\_ID Distance Speeding  
## 1 -0.87423847 -0.4887583 -0.30533176  
## 2 -0.07710531 1.9525505 -0.01664689  
## 3 0.86802507 -0.4861001 -0.29880377  
## 4 0.01878405 0.2963155 2.78654635  
##   
## Clustering vector:  
## [1] 4 3 3 1 1 3 1 1 4 3 4 4 3 1 4 3 4 1 1 3 1 4 1 3 4 4 1 1 3 4 4 4 4 4 1 4 1  
## [38] 4 4 4 4 3 1 4 1 1 1 4 4 1 3 1 3 4 4 4 4 1 1 4 1 1 3 4 1 4 4 4 4 4 3 1 4 4  
## [75] 3 4 4 1 4 4 3 1 1 3 3 1 1 3 4 3 1 1 4 1 4 3 4 3 4 1 1 1 1 4 3 4 4 1 1 3 4  
## [112] 1 1 1 3 3 4 1 1 4 1 4 4 1 4 3 3 3 4 1 4 4 4 3 4 4 4 3 3 1 3 4 4 4 3 4 1 4  
## [149] 3 3 4 4 3 4 1 4 1 1 1 4 4 3 1 4 3 3 4 4 1 3 1 3 3 1 4 4 4 1 4 1 1 4 4 3 4  
## [186] 3 4 3 1 4 1 3 1 3 3 4 1 4 4 4 4 4 4 3 1 4 1 4 3 3 3 4 3 1 3 3 1 3 1 4 4 3  
## [223] 4 1 4 3 4 3 1 3 4 3 4 4 3 3 1 4 1 3 4 4 4 4 4 3 1 1 1 3 4 4 4 1 3 3 4 4 1  
## [260] 4 4 4 3 1 4 1 4 4 1 1 4 4 3 4 4 3 1 4 3 3 4 4 3 3 3 4 3 3 3 3 4 4 4 1 4 1  
## [297] 4 1 4 4 1 4 4 1 1 4 3 3 3 3 1 4 3 4 3 4 4 3 4 4 3 4 1 1 3 1 4 4 1 4 4 4 4  
## [334] 4 4 1 4 3 4 4 1 4 4 4 4 3 3 3 4 4 4 4 4 4 4 4 4 4 4 1 4 4 3 4 4 1 4 3 4 4  
## [371] 4 4 1 3 4 1 4 4 1 1 3 4 1 1 4 3 4 3 3 4 3 1 1 1 4 3 4 4 4 3 1 4 4 1 4 4 1  
## [408] 4 1 1 4 3 3 4 3 3 3 1 4 3 4 1 3 3 4 3 3 1 4 4 1 3 1 1 3 3 3 3 3 3 3 3 3 3  
## [445] 3 1 4 4 4 3 4 4 1 4 1 4 3 3 3 4 4 4 1 4 4 4 1 4 3 4 3 1 4 4 3 4 4 4 3 4 1  
## [482] 1 1 3 1 3 1 3 1 1 1 1 1 3 1 1 1 1 1 1 3 3 3 3 3 3 3 1 3 1 1 1 3 1 1 1 3 1  
## [519] 3 1 3 3 1 3 1 1 3 1 1 1 3 1 1 3 3 3 1 1 1 3 3 3 3 1 1 3 3 1 1 1 1 1 3 3 1  
## [556] 1 1 1 1 1 1 3 1 3 3 3 1 3 3 3 1 3 1 1 3 1 1 1 3 1 3 3 1 3 1 3 3 3 1 1 3 3  
## [593] 1 3 3 1 3 3 1 3 3 1 1 1 3 3 1 3 3 1 1 1 3 3 1 1 1 3 3 3 1 3 1 3 1 3 3 3 1  
## [630] 1 1 1 3 3 1 1 1 1 3 1 3 3 3 3 1 1 1 3 1 1 3 3 1 1 3 3 1 3 1 3 3 1 3 1 1 3  
## [667] 1 3 3 1 1 3 3 3 3 1 1 1 3 3 1 1 1 1 1 1 1 3 1 3 1 1 1 1 3 1 1 3 1 3 1 1 3  
## [704] 1 3 3 1 1 1 1 1 1 1 1 3 1 3 1 1 3 3 3 1 1 3 3 1 1 3 1 3 3 3 1 1 3 1 1 3 3  
## [741] 3 1 3 1 3 1 3 3 3 1 1 1 1 3 1 3 1 3 1 1 1 1 3 3 3 3 1 3 3 1 3 1 1 1 3 1 3  
## [778] 1 1 3 3 1 1 1 1 3 1 3 3 3 3 1 1 3 3 1 1 1 1 3 1 1 3 3 1 3 1 3 1 3 3 1 3 1  
## [815] 3 3 3 1 3 1 3 3 3 1 1 3 3 3 1 3 3 1 1 1 1 3 1 3 3 3 3 1 1 1 3 1 3 3 3 1 3  
## [852] 1 3 3 3 1 3 1 3 3 3 3 1 1 1 1 3 1 1 3 1 3 3 1 3 3 3 3 1 3 1 3 3 1 1 1 1 3  
## [889] 3 3 3 1 3 1 3 1 1 3 3 1 3 3 3 1 1 1 3 1 1 3 3 3 1 3 1 1 1 3 1 3 3 1 1 3 3  
## [926] 3 1 1 1 1 3 1 1 1 3 1 3 1 3 3 1 3 1 1 3 1 3 1 3 3 1 3 1 1 1 1 1 1 3 3 3 3  
## [963] 3 3 1 3 1 1 3 3 3 3 1 3 3 1 1 1 1 3 3 1 3 1 3 1 1 3 3 1 1 1 1 1 3 3 3 3 1  
## [1000] 3 1 3 1 1 1 1 1 3 3 3 3 3 1 3 3 1 1 3 3 3 3 1 3 1 3 3 1 1 1 1 1 1 3 3 1 1  
## [1037] 3 1 3 1 3 1 1 1 3 3 1 3 3 1 1 1 3 1 3 1 1 3 3 3 1 3 1 3 3 1 3 3 3 1 3 3 3  
## [1074] 3 3 3 3 3 1 3 1 3 1 3 1 3 3 3 3 3 3 3 3 3 1 1 3 3 1 1 1 1 3 3 1 1 1 3 3 3  
## [1111] 1 3 1 3 3 1 3 1 3 1 1 1 1 3 1 3 3 3 1 3 3 1 3 1 3 3 1 3 3 3 3 1 3 3 1 1 1  
## [1148] 1 3 1 3 3 3 1 1 3 1 1 1 1 1 3 3 1 3 3 3 1 3 3 3 1 3 1 1 1 1 3 1 3 1 3 1 3  
## [1185] 3 3 3 1 3 1 1 1 3 1 3 3 1 3 1 1 1 1 1 3 3 3 1 3 3 1 1 3 3 1 3 3 1 1 3 1 1  
## [1222] 1 3 1 3 3 3 3 3 3 1 1 3 3 3 3 1 3 1 3 3 1 1 3 3 1 3 3 1 3 3 3 1 1 3 1 1 3  
## [1259] 3 3 3 1 1 1 1 1 1 1 3 1 3 1 1 3 1 1 1 3 3 3 3 1 1 3 3 1 1 1 1 1 1 3 1 3 3  
## [1296] 3 3 1 1 3 1 1 3 1 3 1 3 3 1 3 1 3 3 1 3 1 1 1 3 1 3 1 3 3 3 3 1 1 3 3 3 1  
## [1333] 3 1 3 1 3 1 3 3 3 1 3 1 3 3 1 1 1 3 1 1 3 1 3 1 3 3 1 3 3 1 3 3 3 1 3 3 1  
## [1370] 3 3 3 3 3 1 3 1 1 1 3 1 1 1 3 3 1 3 3 1 3 1 1 1 1 1 1 3 1 1 1 3 1 3 1 1 1  
## [1407] 3 1 3 3 3 1 3 1 3 1 1 3 1 1 1 3 1 1 1 1 1 1 1 3 3 3 1 3 1 1 1 1 1 3 3 1 3  
## [1444] 3 3 1 1 1 1 1 3 3 1 3 3 1 3 1 3 3 3 3 3 1 1 3 3 1 3 1 1 1 1 3 3 1 3 1 3 1  
## [1481] 3 1 3 3 1 3 1 1 3 3 3 3 3 3 1 3 1 3 1 1 3 3 3 3 1 1 1 3 3 3 1 3 3 1 1 1 3  
## [1518] 1 3 3 1 3 3 1 3 3 3 1 1 1 3 3 1 3 3 1 1 3 3 3 1 3 1 3 3 1 1 3 3 3 3 3 3 1  
## [1555] 3 3 3 1 3 3 3 3 1 3 3 1 1 1 1 1 1 3 1 3 3 3 1 3 1 1 3 3 1 3 1 1 3 3 3 1 3  
## [1592] 1 3 1 3 1 3 3 1 3 1 3 3 3 1 3 3 1 3 3 3 3 1 3 3 1 1 1 3 3 1 3 1 3 3 1 3 3  
## [1629] 1 1 3 1 3 1 1 3 1 3 3 3 1 3 3 3 1 1 3 3 3 1 3 1 1 1 3 1 3 1 3 1 3 1 3 3 1  
## [1666] 1 3 3 1 1 3 1 1 1 1 3 1 1 3 1 1 1 1 1 1 3 3 1 1 3 3 3 3 3 3 1 3 3 3 1 3 3  
## [1703] 1 3 3 1 3 3 3 1 1 3 1 3 3 1 1 1 1 3 3 3 1 3 1 1 3 3 1 3 1 1 3 3 3 3 1 3 1  
## [1740] 3 3 1 3 1 3 1 3 1 1 1 3 1 1 3 3 1 3 1 1 3 3 1 3 3 3 3 3 3 1 1 3 3 1 3 1 3  
## [1777] 3 3 3 3 1 3 3 3 3 3 1 1 3 3 3 1 1 1 3 1 1 1 3 3 3 3 1 1 3 3 1 1 3 1 1 3 3  
## [1814] 3 3 1 1 3 1 1 3 3 3 3 3 1 1 1 3 1 1 1 3 1 1 1 1 1 1 3 3 1 3 1 1 1 3 1 1 3  
## [1851] 3 3 1 3 3 1 1 1 3 3 1 1 1 3 3 3 3 3 1 3 1 1 3 1 3 1 3 1 1 1 3 1 1 1 3 1 1  
## [1888] 3 3 3 3 3 3 3 1 3 1 3 3 3 1 3 3 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 3 1 3 1  
## [1925] 1 1 3 1 3 3 1 3 3 3 1 1 1 3 3 1 1 3 3 1 3 1 1 3 3 1 1 1 3 3 3 1 1 3 3 3 1  
## [1962] 3 1 3 1 3 1 1 1 3 3 1 1 1 3 1 3 1 3 3 1 3 1 1 3 3 1 1 1 3 3 3 3 3 3 3 1 1  
## [1999] 3 1 3 1 3 3 3 3 1 1 1 1 1 1 1 3 1 1 3 3 3 1 3 1 1 3 3 3 1 3 3 1 1 1 3 3 3  
## [2036] 3 1 3 1 3 3 3 1 3 1 1 3 1 3 3 3 3 1 1 3 1 1 1 1 1 3 1 1 3 1 1 1 1 3 3 1 3  
## [2073] 3 1 1 1 3 3 3 3 3 3 1 3 1 3 1 3 1 1 3 1 1 3 1 3 3 1 1 1 3 1 3 1 1 1 1 1 3  
## [2110] 1 3 1 3 3 3 1 3 3 3 3 3 1 1 1 1 3 3 1 1 1 1 3 3 3 3 1 1 3 3 1 3 3 1 3 3 3  
## [2147] 3 3 3 1 1 3 3 1 3 1 3 3 3 1 1 3 1 1 1 3 3 3 3 3 1 3 1 3 3 3 3 3 3 1 1 3 3  
## [2184] 1 3 1 3 3 1 3 3 1 3 1 3 3 3 3 3 3 1 3 1 3 1 3 1 3 3 1 3 3 3 1 1 1 1 1 3 1  
## [2221] 1 3 3 1 3 3 3 3 1 1 3 3 3 1 1 3 1 1 1 1 3 1 3 1 3 3 3 1 3 1 1 3 3 1 1 3 1  
## [2258] 3 1 1 1 1 1 1 3 1 1 3 3 1 1 1 3 3 3 3 3 3 1 1 3 3 1 3 1 3 1 1 3 1 3 1 3 3  
## [2295] 3 1 3 3 3 1 3 1 3 3 1 1 1 1 1 1 1 3 1 1 3 1 3 3 1 1 1 1 3 1 3 3 3 3 1 1 1  
## [2332] 3 3 1 1 1 1 3 1 3 3 1 1 1 3 1 3 3 3 1 1 3 1 3 3 1 3 3 1 1 1 1 1 1 3 1 3 1  
## [2369] 1 3 3 1 3 1 3 3 1 3 3 3 1 3 1 3 3 3 1 1 3 3 3 1 3 3 3 1 3 3 3 3 1 3 1 3 3  
## [2406] 3 3 3 3 3 1 3 1 3 3 1 1 3 3 1 1 1 3 3 3 1 3 1 3 1 3 3 3 1 1 3 3 3 3 1 1 1  
## [2443] 1 3 3 1 1 1 1 3 3 1 1 3 3 1 1 1 1 3 1 3 1 1 1 1 3 1 1 1 3 3 1 3 3 3 1 3 1  
## [2480] 3 1 1 3 1 1 3 1 3 1 3 3 1 3 3 1 1 1 3 3 1 3 3 1 1 1 1 3 1 3 1 3 3 1 1 3 3  
## [2517] 1 1 3 1 3 3 1 3 1 3 1 3 1 3 3 3 1 3 3 1 3 3 3 1 3 3 3 3 3 3 3 1 3 1 1 1 1  
## [2554] 1 1 3 1 3 3 1 3 1 1 3 3 3 3 1 1 3 1 3 1 1 3 3 1 3 3 3 3 1 3 3 1 1 3 3 3 1  
## [2591] 3 3 1 3 3 1 1 1 1 3 1 1 1 3 1 1 1 3 3 1 3 3 3 3 1 3 3 3 3 3 1 3 3 3 1 1 3  
## [2628] 1 3 1 1 1 3 1 1 1 3 3 1 1 1 1 3 3 1 1 1 1 3 1 1 1 3 1 3 3 3 1 1 1 1 3 1 3  
## [2665] 1 3 1 3 3 3 3 1 3 3 3 1 1 3 3 1 1 3 3 1 1 3 3 1 3 1 1 1 1 1 3 1 3 3 3 1 3  
## [2702] 1 3 3 1 1 1 3 3 3 1 3 1 3 1 3 1 1 1 3 3 1 1 3 1 1 1 3 1 1 3 1 3 3 1 1 3 1  
## [2739] 1 1 1 3 1 1 1 1 1 1 3 3 1 3 3 1 3 1 1 3 3 1 3 3 1 3 1 3 3 1 1 1 3 3 3 1 1  
## [2776] 3 1 3 1 1 1 1 3 3 1 3 1 3 1 3 3 1 3 1 1 1 3 3 1 1 3 3 3 3 3 1 3 1 1 3 1 3  
## [2813] 3 3 3 3 1 1 1 3 1 3 1 1 3 3 3 1 1 3 1 3 3 1 3 3 3 3 1 3 3 1 3 3 1 3 3 3 1  
## [2850] 1 1 1 1 3 1 1 3 1 1 1 3 1 1 3 3 3 1 3 3 1 1 1 3 3 3 1 3 3 1 3 3 1 1 3 3 1  
## [2887] 3 3 1 3 1 1 3 1 3 1 1 1 1 3 1 1 3 3 3 3 3 3 1 3 1 1 1 3 1 1 3 3 1 3 3 3 1  
## [2924] 1 3 1 1 1 3 1 3 1 3 1 3 3 1 3 3 3 3 3 3 3 1 1 1 1 1 3 1 1 1 1 1 1 1 3 3 1  
## [2961] 1 3 3 1 1 3 1 1 3 1 3 3 1 1 3 3 3 3 3 3 3 3 3 1 3 3 1 1 1 3 1 1 1 3 1 1 1  
## [2998] 3 1 1 3 3 3 3 3 3 3 1 1 3 3 3 3 3 1 1 1 3 3 1 1 1 3 3 3 3 1 1 1 1 1 1 1 3  
## [3035] 3 3 3 3 1 1 1 1 3 3 1 1 3 3 1 1 3 3 1 3 1 3 3 1 3 3 1 1 3 3 1 3 1 1 3 3 1  
## [3072] 3 1 3 3 3 3 1 3 3 1 3 1 1 1 3 3 3 3 3 3 3 3 1 3 1 1 1 1 1 3 1 1 1 3 3 1 1  
## [3109] 1 3 3 1 1 3 1 3 3 1 1 3 1 3 3 1 1 3 1 1 1 3 1 3 1 3 3 3 3 3 3 1 3 1 3 1 1  
## [3146] 3 3 1 1 3 3 1 3 3 1 3 3 3 1 1 1 3 1 3 1 1 1 1 1 3 1 1 1 1 3 1 3 3 1 3 1 1  
## [3183] 3 3 1 1 1 1 1 3 3 1 3 1 3 3 3 3 3 3 4 4 4 4 4 4 2 4 4 4 4 4 4 2 4 2 4 4 4  
## [3220] 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 4 4 4 4 4 2 4 2 4 4 4 4 4 4  
## [3257] 4 4 4 4 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 4 2 4 4 4 4 4 4 4 2 4 4 4 2  
## [3294] 4 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 2 4 4 4 4 4 2 2 2 2 2 2 2 2 2 2  
## [3331] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3368] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3405] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3442] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3479] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3516] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3553] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3590] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3627] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3664] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3701] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3738] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3775] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3812] 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3849] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3886] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3923] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3960] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [3997] 2 2 2 2  
##   
## Within cluster sum of squares by cluster:  
## [1] 647.0885 889.8846 713.9785 1288.1896  
## (between\_SS / total\_SS = 70.5 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

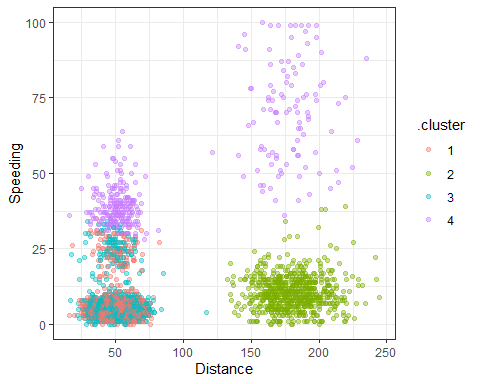
Add the cluster assignment to the dataset

trucks = augment(kclust, trucks)  
trucks

## # A tibble: 4,000 x 4  
## Driver\_ID Distance Speeding .cluster  
## <dbl> <dbl> <dbl> <fct>   
## 1 3423311935 71.2 28 4   
## 2 3423313212 52.5 25 3   
## 3 3423313724 64.5 27 3   
## 4 3423311373 55.7 22 1   
## 5 3423310999 54.6 25 1   
## 6 3423313857 41.9 10 3   
## 7 3423312432 58.6 20 1   
## 8 3423311434 52.0 8 1   
## 9 3423311328 31.2 34 4   
## 10 3423312488 44.3 19 3   
## # ... with 3,990 more rows

Plot the clusters

ggplot(trucks, aes(Distance, Speeding, color = .cluster)) +  
 geom\_point(alpha = 0.4) + theme\_bw()



## Cluster 2 is the most distinct cluster as I can see all the driver in that cluster are grouped together. That means with more distance travelled, these drivers were not overspeeding much.

## Cluster 4 is also distinct, but groups in 2 parts on the graph.

## Cluster 1 and Cluster 3 are kind of mixed with each other.