

FML Final Exam

Praneeth Simha

2023-05-04

```
library(tidyverse)

## — Attaching core tidyverse packages — tidyverse
2.0.0 —
## ✓ dplyr      1.1.0      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.1      ✓ tibble     3.1.8
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr      1.0.1
## — Conflicts —
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## ⓘ Use the http://conflicted.r-lib.org/conflicted-package to force
all conflicts to become errors

library(dplyr)
library(tidyr)
library(ggplot2)
library(ggthemes)

## Warning: package 'ggthemes' was built under R version 4.2.3

#Installing the packages

library(readxl)
PUDL <- read.csv("C:/Users/ADMIN/Downloads/PUDL.csv")

#loading the real time dataset

# Check the data
str(PUDL)

## 'data.frame':   608564 obs. of  30 variables:
## $ rowid                : int  1 2 3 4 5 6 7 8 9 10 ...
## $ plant_id_eia         : int  3 3 3 7 7 7 7 8 8 8 ...
## $ plant_id_eia_label   : chr  "Barry" "Barry" "Barry"
## "Gadsden" ...
## $ report_date          : chr  "2008-01-01" "2008-01-
## 01" "2008-01-01" "2008-01-01" ...
## $ contract_type_code   : chr  "C" "C" "C" "C" ...
## $ contract_type_code_label : chr  "C" "C" "C" "C" ...
```

```
## $ contract_expiration_date : chr "2008-04-01" "2008-04-01" "" "2015-12-01" ...
## $ energy_source_code : chr "BIT" "BIT" "NG" "BIT" ...
## $ energy_source_code_label : chr "BIT" "BIT" "NG" "BIT" ...
## $ fuel_type_code_pudl : chr "coal" "coal" "gas" "coal" ...
## $ fuel_group_code : chr "coal" "coal" "natural_gas" "coal" ...
## $ mine_id_pudl : int 0 0 NA 1 2 3 NA 4 4 1 ...
## $ mine_id_pudl_label : int 0 0 NA 1 2 3 NA 4 4 1 ...
## $ supplier_name : chr "interocean coal" "interocean coal" "bay gas pipeline" "alabama coal" ...
## $ fuel_received_units : num 259412 52241 2783619 25397 764 ...
## $ fuel_mmbtu_per_unit : num 23.1 22.8 1.04 24.61 24.45 ...
## $ sulfur_content_pct : num 0.49 0.48 0 1.69 0.84 1.54 0 2.16 1.24 1.9 ...
## $ ash_content_pct : num 5.4 5.7 0 14.7 15.5 14.6 0 15.4 11.9 15.4 ...
## $ mercury_content_ppm : num NA NA NA NA NA NA NA NA NA NA ...
## $ fuel_cost_per_mmbtu : num 2.13 2.12 8.63 2.78 3.38 ...
## $ primary_transportation_mode_code : chr "RV" "RV" "PL" "TR" ...
## $ primary_transportation_mode_code_label : chr "RV" "RV" "PL" "TR" ...
## $ secondary_transportation_mode_code : chr "" "" "" "" ...
## $ secondary_transportation_mode_code_label : chr "" "" "" "" ...
## $ natural_gas_transport_code : chr "firm" "firm" "firm" "firm" ...
## $ natural_gas_delivery_contract_type_code : chr "" "" "" "" ...
## $ moisture_content_pct : num NA NA NA NA NA NA NA NA NA NA ...
## $ chlorine_content_ppm : num NA NA NA NA NA NA NA NA NA NA ...
## $ data_maturity : chr "final" "final" "final" "final" ...
## $ data_maturity_label : chr "final" "final" "final" "final" ...

# Explore the data
glimpse(PUDL)

## Rows: 608,564
## Columns: 30
## $ rowid <int> 1, 2, 3, 4, 5, 6, 7, 8,
```

9, 10...	
## \$ plant_id_eia	<int> 3, 3, 3, 7, 7, 7, 7, 8,
8, 8,...	
## \$ plant_id_eia_label	<chr> "Barry", "Barry",
"Barry", "G...	
## \$ report_date	<chr> "2008-01-01", "2008-01-
01", "...	
## \$ contract_type_code	<chr> "C", "C", "C", "C", "S",
"S",...	
## \$ contract_type_code_label	<chr> "C", "C", "C", "C", "S",
"S",...	
## \$ contract_expiration_date	<chr> "2008-04-01", "2008-04-
01", "...	
## \$ energy_source_code	<chr> "BIT", "BIT", "NG",
"BIT", "B...	
## \$ energy_source_code_label	<chr> "BIT", "BIT", "NG",
"BIT", "B...	
## \$ fuel_type_code_pudl	<chr> "coal", "coal", "gas",
"coal"...	
## \$ fuel_group_code	<chr> "coal", "coal",
"natural_gas"...	
## \$ mine_id_pudl	<int> 0, 0, NA, 1, 2, 3, NA, 4,
4, ...	
## \$ mine_id_pudl_label	<int> 0, 0, NA, 1, 2, 3, NA, 4,
4, ...	
## \$ supplier_name	<chr> "interocean coal",
"interoceaa...	
## \$ fuel_received_units	<dbl> 259412, 52241, 2783619,
25397...	
## \$ fuel_mmbtu_per_unit	<dbl> 23.100, 22.800, 1.039,
24.610...	
## \$ sulfur_content_pct	<dbl> 0.49, 0.48, 0.00, 1.69,
0.84,...	
## \$ ash_content_pct	<dbl> 5.4, 5.7, 0.0, 14.7,
15.5, 14...	
## \$ mercury_content_ppm	<dbl> NA, NA, NA, NA, NA, NA,
NA, N...	
## \$ fuel_cost_per_mmbtu	<dbl> 2.135, 2.115, 8.631,
2.776, 3...	
## \$ primary_transportation_mode_code	<chr> "RV", "RV", "PL", "TR",
"TR",...	
## \$ primary_transportation_mode_code_label	<chr> "RV", "RV", "PL", "TR",
"TR",...	
## \$ secondary_transportation_mode_code	<chr> "", "", "", "", "", "",
"", "...	
## \$ secondary_transportation_mode_code_label	<chr> "", "", "", "", "", "",
"", "...	
## \$ natural_gas_transport_code	<chr> "firm", "firm", "firm",
"firm...	
## \$ natural_gas_delivery_contract_type_code	<chr> "", "", "", "", "", "",

```

"", "..."
## $ moisture_content_pct          <dbl> NA, NA, NA, NA, NA, NA,
NA, N...
## $ chlorine_content_ppm          <dbl> NA, NA, NA, NA, NA, NA,
NA, N...
## $ data_maturity                 <chr> "final", "final",
"final", "f...
## $ data_maturity_label           <chr> "final", "final",
"final", "f..."

```

The data set is mostly clean, although it has a number of variables with significant missing values. The next actions should be made to fix this problem. Remove any variables with a large number of missing values.

```

# col names with missing values
colnames(PUDL)[colSums(is.na(PUDL)) > 0]

## [1] "mine_id_pudl"          "mine_id_pudl_label"  "mercury_content_ppm"
## [4] "fuel_cost_per_mmbtu"  "moisture_content_pct" "chlorine_content_ppm"

# all missing values
all <- PUDL %>%
  summarise_all(funs(sum(is.na(.)))) %>%
  gather(key = "variable", value = "missing_values") %>%
  filter(missing_values > 0) %>%
  arrange(desc(missing_values))

## Warning: `funs()` was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with `tibble::lst()`: tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))

# remove the variables with missing values
PUDL <- PUDL %>%
  select(-all$variable)

# check the data
str(PUDL)

## 'data.frame':   608564 obs. of  24 variables:
## $ rowid                : int  1 2 3 4 5 6 7 8 9 10 ...
## $ plant_id_eia          : int  3 3 3 7 7 7 7 8 8 8 ...
## $ plant_id_eia_label    : chr  "Barry" "Barry" "Barry"
"Gadsden" ...
## $ report_date           : chr  "2008-01-01" "2008-01-
01" "2008-01-01" "2008-01-01" ...
## $ contract_type_code    : chr  "C" "C" "C" "C" ...

```

```
## $ contract_type_code_label      : chr "C" "C" "C" "C" ...
## $ contract_expiration_date      : chr "2008-04-01" "2008-04-
01" "" "2015-12-01" ...
## $ energy_source_code            : chr "BIT" "BIT" "NG" "BIT"
...
## $ energy_source_code_label      : chr "BIT" "BIT" "NG" "BIT"
...
## $ fuel_type_code_pudl          : chr "coal" "coal" "gas"
"coal" ...
## $ fuel_group_code              : chr "coal" "coal"
"natural_gas" "coal" ...
## $ supplier_name                : chr "interocean coal"
"interocean coal" "bay gas pipeline" "alabama coal" ...
## $ fuel_received_units          : num 259412 52241 2783619
25397 764 ...
## $ fuel_mmbtu_per_unit          : num 23.1 22.8 1.04 24.61
24.45 ...
## $ sulfur_content_pct           : num 0.49 0.48 0 1.69 0.84
1.54 0 2.16 1.24 1.9 ...
## $ ash_content_pct              : num 5.4 5.7 0 14.7 15.5 14.6
0 15.4 11.9 15.4 ...
## $ primary_transportation_mode_code : chr "RV" "RV" "PL" "TR" ...
## $ primary_transportation_mode_code_label : chr "RV" "RV" "PL" "TR" ...
## $ secondary_transportation_mode_code : chr "" "" "" "" ...
## $ secondary_transportation_mode_code_label : chr "" "" "" "" ...
## $ natural_gas_transport_code     : chr "firm" "firm" "firm"
"firm" ...
## $ natural_gas_delivery_contract_type_code : chr "" "" "" "" ...
## $ data_maturity                 : chr "final" "final" "final"
"final" ...
## $ data_maturity_label           : chr "final" "final" "final"
"final" ...
```

2. Ensure that the variables have the right attributes. For example, numerical or categorical.

```
# attributes
sapply(PUDL, class)

##               rowid
##               "integer"
##           plant_id_eia
##               "integer"
##       plant_id_eia_label
##               "character"
##           report_date
##               "character"
##       contract_type_code
##               "character"
## contract_type_code_label
##               "character"
```

```

##          contract_expiration_date
##          "character"
##          energy_source_code
##          "character"
##          energy_source_code_label
##          "character"
##          fuel_type_code_pudl
##          "character"
##          fuel_group_code
##          "character"
##          supplier_name
##          "character"
##          fuel_received_units
##          "numeric"
##          fuel_mmbtu_per_unit
##          "numeric"
##          sulfur_content_pct
##          "numeric"
##          ash_content_pct
##          "numeric"
##          primary_transportation_mode_code
##          "character"
##          primary_transportation_mode_code_label
##          "character"
##          secondary_transportation_mode_code
##          "character"
##          secondary_transportation_mode_code_label
##          "character"
##          natural_gas_transport_code
##          "character"
##          natural_gas_delivery_contract_type_code
##          "character"
##          data_maturity
##          "character"
##          data_maturity_label
##          "character"

```

3. To ensure that both the data, and the analysis are unique to each student, randomly sample about 2% of your data using a random 4-digit number as the seed to sample the data. Use 75% of the sampled data as the training set, and the rest as the test set (if needed). This should yield a training set of about 9000 and a test of about 3000.

```

# set seed
set.seed(1122)

# sample the data
sampled <- PUDL %>%
  sample_frac(0.02)

# split the data

```

```

train <- sampled %>%
  sample_frac(0.75)

test <- sampled %>%
  anti_join(train)

## Joining with `by = join_by(rowid, plant_id_eia, plant_id_eia_label,
## report_date, contract_type_code, contract_type_code_label,
## contract_expiration_date, energy_source_code, energy_source_code_label,
## fuel_type_code_pudl, fuel_group_code, supplier_name, fuel_received_units,
## fuel_mmbtu_per_unit, sulfur_content_pct, ash_content_pct,
## primary_transportation_mode_code, primary_transportation_mode_code_label,
## secondary_transportation_mode_code,
## secondary_transportation_mode_code_label,
## natural_gas_transport_code, natural_gas_delivery_contract_type_code,
## data_maturity, data_maturity_label)`

# check the data
str(train)

## 'data.frame': 9128 obs. of 24 variables:
## $ rowid : int 347804 399073 534470
386975 506950 98750 384101 377369 309829 152164 ...
## $ plant_id_eia : int 1927 54817 7296 55350
55340 2718 52120 2965 703 1830 ...
## $ plant_id_eia_label : chr "Riverside" "Ponderosa
Pine Energy Ptrs" "State Line Combined Cycle" "Dresden Energy Facility" ...
## $ report_date : chr "2014-11-01" "2016-05-
01" "2019-12-01" "2015-12-01" ...
## $ contract_type_code : chr "S" "S" "S" "S" ...
## $ contract_type_code_label : chr "S" "S" "S" "S" ...
## $ contract_expiration_date : chr "" "" "" "" ...
## $ energy_source_code : chr "NG" "NG" "NG" "NG" ...
## $ energy_source_code_label : chr "NG" "NG" "NG" "NG" ...
## $ fuel_type_code_pudl : chr "gas" "gas" "gas" "gas"
...
## $ fuel_group_code : chr "natural_gas"
"natural_gas" "natural_gas" "natural_gas" ...
## $ supplier_name : chr "various suppliers
(natural gas only)" "etc" "various (natural gas spot purchases only)"
"pacific summit" ...
## $ fuel_received_units : num 547258 18496 645082
432494 9903 ...
## $ fuel_mmbtu_per_unit : num 1.053 0.994 1.037 1.078
1.025 ...
## $ sulfur_content_pct : num 0 0 0 0 0 0.71 0 0 0.89
0.6 ...
## $ ash_content_pct : num 0 0 0 0 0 12.4 0 0 8.5
10.5 ...
## $ primary_transportation_mode_code : chr "PL" "PL" "PL" "PL" ...

```

```

## $ primary_transportation_mode_code_label : chr "PL" "PL" "PL" "PL" ...
## $ secondary_transportation_mode_code : chr "" "" "" "" ...
## $ secondary_transportation_mode_code_label: chr "" "" "" "" ...
## $ natural_gas_transport_code : chr "interruptible" "firm"
"firm" "firm" ...
## $ natural_gas_delivery_contract_type_code : chr "interruptible" "firm"
"firm" "firm" ...
## $ data_maturity : chr "final" "final" "final"
"final" ...
## $ data_maturity_label : chr "final" "final" "final"
"final" ...

str(test)

## 'data.frame': 3043 obs. of 24 variables:
## $ rowid : int 589950 94332 60791
455738 318497 159008 35578 578731 412563 366491 ...
## $ plant_id_eia : int 997 3631 50489 1745
55206 6094 56500 6264 55328 6823 ...
## $ plant_id_eia_label : chr "Michigan City" "Sam
Rayburn" "PPG Powerhouse C" "Trenton Channel" ...
## $ report_date : chr "2021-07-01" "2009-07-
01" "2008-12-01" "2017-11-01" ...
## $ contract_type_code : chr "S" "S" "S" "S" ...
## $ contract_type_code_label : chr "S" "S" "S" "S" ...
## $ contract_expiration_date : chr "" "" "2008-12-01" ""
...
## $ energy_source_code : chr "NG" "DFO" "NG" "NG" ...
## $ energy_source_code_label : chr "NG" "DFO" "NG" "NG" ...
## $ fuel_type_code_pudl : chr "gas" "oil" "gas" "gas"
...
## $ fuel_group_code : chr "natural_gas"
"petroleum" "natural_gas" "natural_gas" ...
## $ supplier_name : chr "arch" "exxonmobil" "jp
morgan" "various (natural gas spot purchases only)" ...
## $ fuel_received_units : num 16307 72 35378 3846
1382365 ...
## $ fuel_mmbtu_per_unit : num 1.06 5.84 1.03 1.01 1.03
...
## $ sulfur_content_pct : num 0 0.35 0 0 0 3.1 0 4.18
0 3.04 ...
## $ ash_content_pct : num 0 0 0 0 0 11.6 0 9.61 0
8.5 ...
## $ primary_transportation_mode_code : chr "PL" "TR" "" "PL" ...
## $ primary_transportation_mode_code_label : chr "PL" "TR" "" "PL" ...
## $ secondary_transportation_mode_code : chr "" "TR" "" "" ...
## $ secondary_transportation_mode_code_label: chr "" "TR" "" "" ...
## $ natural_gas_transport_code : chr "interruptible" ""
"firm" "firm" ...
## $ natural_gas_delivery_contract_type_code : chr "interruptible" "" ""

```



```

"firm" ...
## $ data_maturity           : chr  "final" "final" "final"
"final" ...
## $ data_maturity_label     : chr  "final" "final" "final"
"final" ...

```

Visualizing the Data

```

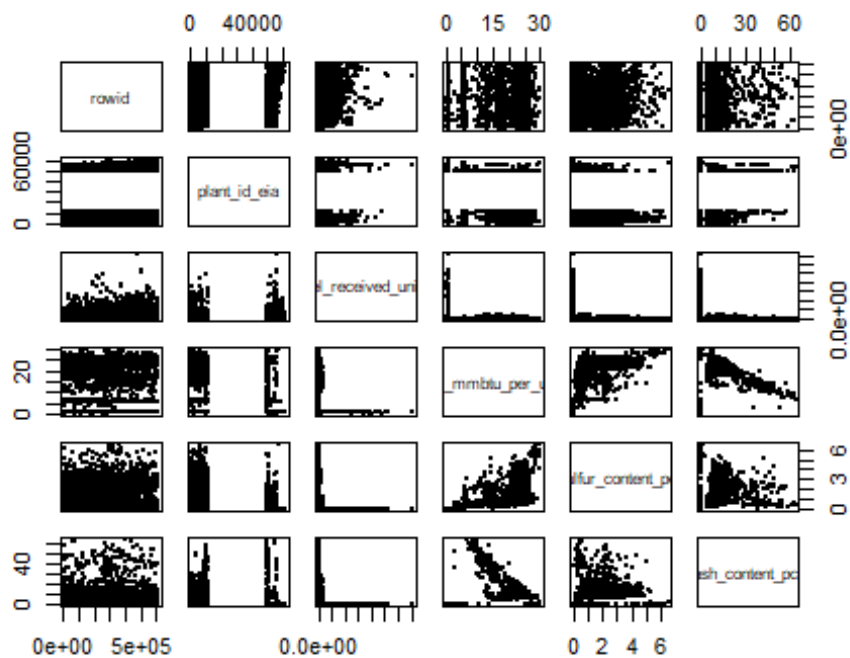
# visualize the data scatterplot matrix
numValues <- sapply(train, is.numeric)
numValues

##                                rowid
##                                TRUE
##                   plant_id_eia
##                                TRUE
##                   plant_id_eia_label
##                                FALSE
##                   report_date
##                                FALSE
##                   contract_type_code
##                                FALSE
##                   contract_type_code_label
##                                FALSE
##                   contract_expiration_date
##                                FALSE
##                   energy_source_code
##                                FALSE
##                   energy_source_code_label
##                                FALSE
##                   fuel_type_code_pudl
##                                FALSE
##                   fuel_group_code
##                                FALSE
##                   supplier_name
##                                FALSE
##                   fuel_received_units
##                                TRUE
##                   fuel_mmbtu_per_unit
##                                TRUE
##                   sulfur_content_pct
##                                TRUE
##                   ash_content_pct
##                                TRUE
##                   primary_transportation_mode_code
##                                FALSE
##                   primary_transportation_mode_code_label
##                                FALSE
##                   secondary_transportation_mode_code
##                                FALSE

```

```
## secondary_transportation_mode_code_label
##                                     FALSE
##               natural_gas_transport_code
##                                     FALSE
## natural_gas_delivery_contract_type_code
##                                     FALSE
##                               data_maturity
##                                     FALSE
##               data_maturity_label
##                                     FALSE

pairs(train[,numValues], pch = 19, cex = 0.5)
```



Clustering K-means

clustering

```
# k-means clustering
set.seed(9149)
numValues <- sapply(train, is.numeric)
kmeans <- kmeans(train[,numValues], centers = 3)
kmeans

## K-means clustering with 3 clusters of sizes 546, 8477, 105
##
## Cluster means:
##      rowid plant_id_eia fuel_received_units fuel_mmbtu_per_unit
## 1 362256.8    38176.42      1945453.5      1.0311245
## 2 298051.3    17040.97       86050.3      9.4429755
## 3 368379.0     32911.59     5113157.8     0.9828476
```

[illegible]

[illegible]

[illegible]

[illegible]

[3590] 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 1 2 2 1 2 2 2
2 2 2
[3627] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2
2 2 2
[3664] 2 1 2 2 2 2 2 2 2 2 2 3 2 2 2
2 2 2
[3701] 2 2 2 1 2 1 2 2 2 2 2 2 2 2
2 1 2
[3738] 2 2 2 2 2 2 2 2 2 2 3 2 1 2
2 2 2
[3775] 2 1 2 2 2 2
2 2 2
[3812] 1 2 1 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 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[3886] 2 2 2 3 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[3923] 2 1 2 2 2 2 2 2 2 2 2 2 2 1
2 2 2
[3960] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2
2 2 2
[3997] 2
2 1 2
[4034] 2
2 2 2
[4071] 2
1 1 2
[4108] 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2
2 2 2
[4145] 2 2 2 2 2 1 2
1 2 1
[4182] 2 1 2 2 2 2 2 2 2 2
2 2 2
[4219] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
2 3 2
[4256] 2 2 2 2 2 1 2
2 2 2
[4293] 2
2 2 2
[4330] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 2 1 1 2 1 2 2 2 2 2 2 2
2 2 2
[4367] 2 2 2 2 2 1 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 1 2 2 2 2 2 2 2 3 2 2 2 2
2 1 1
[4404] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[4441] 2
2 1 2
[4478] 2 2 2 2 1 2 2 2 2 1 2 2 1 2
2 2 2

[illegible]

[illegible]

[6365] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2
2 2 1
[6402] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[6439] 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2
2 2 2
[6476] 2 2 1 2 2 2 2 2 1 2
2 2 2
[6513] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2
2 2 2
[6550] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[6587] 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2
2 2 2
[6624] 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 3 2 2 2 2 2 2 1 2 2 2 2 2 2
1 2 2
[6661] 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 1 2
[6698] 2 1 2 2 2 2 2 2 2 2 2
2 2 1
[6735] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2
2 2 2
[6772] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2
2 2 2
[6809] 2 1 2 2 2 2 2 2 2
2 2 1
[6846] 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2
2 2 2
[6883] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2
2 2 1
[6920] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[6957] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2
2 2 2
[6994] 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[7031] 2 2 2 2 2 2 2 1 2 1
2 2 2
[7068] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 3 2 2 2 2 2 2 1 2 1 2 1 2 1
2 2 2
[7105] 2
2 2 2
[7142] 2 1 2 2 2 2 2
3 2 2
[7179] 1 2
1 2 2
[7216] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2
1 2 2
[7253] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2
2 2 2

[illegible]

##	[8215]	3	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	3	2	2	2	2	2	2	2	2	2	2	2
		2	3	2																											
##	[8252]	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
		1	2	2																											
##	[8289]	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2
		1	3	2																											
##	[8326]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2
		2	2	2																											
##	[8363]	2	2	2	2	1	2	1	2	2	3	2	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2
		2	2	2																											
##	[8400]	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	3	2	2
		2	2	2																											
##	[8437]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2
		2	2	2																											
##	[8474]	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2
		2	2	2																											
##	[8511]	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	1	2
		2	2	2																											
##	[8548]	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	1
		2	2	2																											
##	[8585]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2
		2	1	2																											
##	[8622]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
		2	2	2																											
##	[8659]	2	2	3	2	2	2																								

```
## Within cluster sum of squares by cluster:
## [1] 2.560048e+14 5.053238e+14 3.187239e+14
## (between_SS / total_SS = 79.9 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [2] "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"

# aggregate the data
aggregate(train[,numValues], by = list(kmeans$cluster), mean)

##   Group.1   rowid plant_id_eia fuel_received_units fuel_mmbtu_per_unit
## 1      1 362256.8   38176.42         1945453.5         1.0311245
## 2      2 298051.3   17040.97          86050.3         9.4429755
## 3      3 368379.0   32911.59        5113157.8         0.9828476
##   sulfur_content_pct ash_content_pct
## 1      0.0000000      0.000000
## 2      0.5487767      3.937276
## 3      0.0000000      0.000000
```

Visualizing the data

```
ggplot(train, aes(y = kmeans$cluster)) +
  geom_bar(aes(fill = kmeans$cluster), position = "dodge") +
  theme_economist() +
  theme(plot.title = element_text(hjust = 0.5))

## Warning: The following aesthetics were dropped during statistical
## transformation: fill
## i This can happen when ggplot fails to infer the correct grouping
## structure in
## the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
## variable into a factor?
```


[illegible]

[1185] 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 3 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2
2 2 2
[1222] 2 1 2 2 2 1 2
2 2 2
[1259] 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[1296] 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
2 2 2
[1333] 2
2 2 2
[1370] 2 1 2 2 2 2 2
2 2 2
[1407] 2
2 2 2
[1444] 2 1 2 1 2 2 2 2 2
2 2 2
[1481] 2 2 2 2 2 2 1 1 2
2 2 2
[1518] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 1 2
2 2 1
[1555] 2 2 2 1 1 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2
2 2 2
[1592] 2 1 3 2 2 2 2 2 2 2 2 2
2 2 2
[1629] 2 1 2 2 2 2 1 2 2 2 2 1 2
2 2 2
[1666] 2 2 2 2 2 2 2 2 3 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2
2 2 2
[1703] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2
2 2 2
[1740] 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 3 2 2 2 2 2 1 2 1 2 2
2 2 2
[1777] 2 2 2 2 2 2 2 2 3 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[1814] 2 2 2 2 2 2 2 1 2 3 3 2
2 2 2
[1851] 2 2 2 1 2 1 2 2
2 2 2
[1888] 2
1 2 2
[1925] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 3 2 2 2 2 2 2
2 2 1
[1962] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2
2 2 2
[1999] 2 2 2 2 2 2 2 1 2
2 2 2
[2036] 2 1 2 2 2 2 2
2 2 2
[2073] 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 1

## [2110]	2	2	2	1	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																									
## [2147]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2
2	2	2																									
## [2184]	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																									
## [2221]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2
2	2	1																									
## [2258]	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	1
2	2	2																									
## [2295]	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2
2	1	2																									
## [2332]	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	2	2																									
## [2369]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2
2	2	2																									
## [2406]	2	2	2	2	2	3	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2
2	2	3																									
## [2443]	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	1
2	2	2																									
## [2480]	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2
2	2	2																									
## [2517]	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2	2	3	2	2
2	2	2																									
## [2554]	2	1	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	3	2
2	2	1																									
## [2591]	2	2	3	2	2	2	2	2	3	2	2	1	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2
2	2	2			</																						

[illegible]

```
## [556] 2 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3
3 3 3
## [593] 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 2 3 2 3 3 3 3 3
3 3 3
## [630] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 3
3 3 3
## [667] 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2
3 3 2
## [704] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 3 3 3 3 3 3
3 3 3
## [741] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [778] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [815] 3 2 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 2 3 3 2 3 3 3 3 3 2 3 3 2 2 3 3 3
3 3 3
## [852] 3 3 3 3 3 3 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [889] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3
3 3 3
## [926] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3
3 3 3
## [963] 3 3 3 2 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3
3 3 3
## [1000] 3 3 3 3 3 3 2 3 3 3 2 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3
3 3 3
## [1037] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [1074] 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 2 3 3 3 3
3 2 3
## [1111] 2 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3
3 3 3
## [1148] 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3
3 3 3
## [1185] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 2 3 3 3 3 2 3 3 3 3 3 3 3
3 3 3
## [1222] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [1259] 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [1296] 3 3 3 2 3 3 3 3 3 3 3 3 2 3 3 3 2 3 2 3 3 3 3 3 2 3 3 2 3 2 2 3 2
3 3 3
## [1333] 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 2 3 2 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [1370] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3
3 3 3
## [1407] 3 1 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3
3 3 3
## [1444] 3 3 3 3 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
```

[illegible]

## [2406]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																								
## [2443]	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																								
## [2480]	3	3	3	3	3	3	2	3	3	3	2	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3
3	3	3																								
## [2517]	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3
3	3	3																								
## [2554]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
3	3	3																								
## [2591]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3
3	3	3																								
## [2628]	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																								
## [2665]	3	3	3	2	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																								
## [2702]	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
3	3	3																								
## [2739]	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
3	3	3																								
## [2776]	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																								
## [2813]	3	3	3	2	3	2	3	2	3	3	3	1	3	3	3	3	3	3	3	3	2	3	2	3	3	3
3	3	2																								
## [2850]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	
3	3	3																								
## [2887]	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	1	3	3	3	3	3
2	3	3																								
## [2924]	3	3	2	3	3	2	3																			

[illegible]

```
## [4256] 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4293] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4330] 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 2 3 2 3 3 2 2 3 2 3 3 3 3 3  
3 3 3  
## [4367] 3 3 3 3 3 2 3 3 3 3 3 3 3 3 1 3 3 3 3 3 2 3 3 3 3 3 3 1 3 3 3 3  
3 2 2  
## [4404] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 2 3 2 3 3 3 3 3 3 3 3  
3 3 3  
## [4441] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 2 3  
## [4478] 3 3 3 3 2 3 3 3 3 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4515] 3 2 3 3 2 3 3 3 3 2 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4552] 3 3 3 3 3 3 2 3 3 2 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
2 3 3  
## [4589] 3 3 3 3 3 2 2 3 3 3 2 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4626] 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4663] 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 2 3  
## [4700] 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4737] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 2 3  
3 3 3  
## [4774] 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
2 3 3  
## [4811] 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 2 3 3 3 3  
3 3 3  
## [4848] 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4885] 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 2 3 3 2 3 3 3  
3 3 3  
## [4922] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 2 1 3 3 3 3 3  
3 3 3  
## [4959] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3  
## [4996] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 2  
3 3 3  
## [5033] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 2 3 3 3  
2 3 3  
## [5070] 3 3 3 3 3 3 3 1 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3  
3 3 3  
## [5107] 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 2 3  
## [5144] 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 2
```

[illegible]

[6106] 3 2 3 3 3 3 3 3 3 2 3 3 3
3 3 3
[6143] 3 3 2 3
2 3 3
[6180] 3 2 3 3 3 3 2
3 3 2
[6217] 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 2 3 3 3 3 2 3 3 2 2 3 3 3 3 3 3 3 3 3
3 3 3
[6254] 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3
3 3 3
[6291] 3 1 3 3 3 3
3 3 3
[6328] 3 2 3 3 3 3 3 3 3 3 3 1
2 3 3
[6365] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3
3 3 2
[6402] 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[6439] 2 2 3 3 3 3 3 3 2 3 3 3 3 3 3 3 2 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[6476] 3 3 2 3 3 3 3 3 2 3
3 3 3
[6513] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 2 3 3 3
3 3 3
[6550] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[6587] 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 2 3 3 3 3 3 3
3 3 3
[6624] 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 2 3 3 3 3 1 3 3 3 3 3 3 2 3 3 3 3 3 3
2 3 3
[6661] 3 3 3 3 3 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3
3 2 3
[6698] 3 2 3 3 3 3 3 3 3 3 3 3
3 3 2
[6735] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 2 3 3 3 3 3 3 3 3 3 2 3 2 3 3 3 3
3 3 3
[6772] 3 3 3 3 3 3 3 3 2 3 2 3 3 3 2 3
3 3 3
[6809] 3 2 3 3 3 3 3 3 3 3
3 3 2
[6846] 3 3 3 2 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 3 3
3 3 3
[6883] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 2 3 3 3 3 3 3 3 2 3 3 3 3
3 3 2
[6920] 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[6957] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3
3 3 3
[6994] 3 3 3 3 3 2 3 3 3 3 2 3 3 3 3 3 2 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3
3 3 3

[7031] 3 3 3 3 3 3 3 2 3 2
3 3 3
[7068] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 1 3 3 3 3 3 3 2 3 2 3 2 3 2
3 3 3
[7105] 3
3 3 3
[7142] 3 2 3 3 3 3 3
1 3 3
[7179] 2 3
2 3 3
[7216] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 2 3 2 3 3 3 3 3 3 3 3 3 3
2 3 3
[7253] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 2 3
3 3 3
[7290] 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 3 3 3 3 3 3
3 3 3
[7327] 3 2 3 3 3 3 3 3
3 3 2
[7364] 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3
3 3 3
[7401] 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[7438] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 1
3 3 3
[7475] 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3
3 3 1
[7512] 3 3 3 3 3 2 3 1 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[7549] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 2 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[7586] 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 2 3
3 3 3
[7623] 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 2 3 3 3 3 3
3 3 3
[7660] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 3 3 1 3 3 3 3 3 3 2 3 3 3 3 3
3 3 3
[7697] 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3
3 3 3
[7734] 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 1 3 3 3 2
3 3 3
[7771] 3
3 3 3
[7808] 3 3 3 3 2 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3
3 3 1
[7845] 3 3 2 3 2 3
3 3 3
[7882] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
[7919] 3 3 2 3 3 3 3 3 3 3 1 3
3 3 3

## [7956]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																									
## [7993]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																									
## [8030]	3	3	3	3	2	3	3	3	2	3	3	2	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3
2	3	3																									
## [8067]	3	3	3	3	3	2	3	1	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																									
## [8104]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																									
## [8141]	2	3	3	3	2	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3																									
## [8178]	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	2
3	2	3																									
## [8215]	1	3	3	3	3	3	3	3	3	3	3	3	3	2	2	3	3	1	3	3	3	3	3	3	3	3	3
3	1	3																									
## [8252]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	3	3																									
## [8289]	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	2	3	3	2	3	3	3	3	3	3	3
2	1	3																									
## [8326]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	3	3	3	3	3	3	3
3	3	3																									
## [8363]	3	3	3	3	2	3	2	3	3	1	3	3	3	3	3	2	3	3	3	3	2	3	3	3	3	3	3
3	3	3																									
## [8400]	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	1	3
3	3	3																									
## [8437]	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3
3	3	3			</																						

```

## [8881] 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3
2 3 2
## [8918] 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 2 2
3 3 3
## [8955] 1 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 2 3 3 3 3 2 3 3 3
3 3 3
## [8992] 3 3 3 3 3 1 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3
3 3 3
## [9029] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 3 2 3 3 3 3 3 3 3 3 2 3
3 3 3
## [9066] 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3
## [9103] 3 3 2 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3
##
## Within cluster sum of squares by cluster:
## [1] 3.187239e+14 2.560048e+14 5.053238e+14
## (between_SS / total_SS = 79.9 %)
##
## Available components:
##
## [1] "cluster"          "centers"          "totss"            "withinss"
"tot.withinss"
## [6] "betweenss"       "size"             "iter"             "ifault"

```

KNN and K Means

```

# Length of k means
length(kmeans$cluster)

## [1] 9128

# Length of knn
length(knn)

## [1] 3043

```

The k means and KNN are similar but the K means is better than KNN .