Dimensonality reduction

Ran command install.packages("ggplot2") install.packages("lattice") install.packages("caret")

```
library(ggplot2)
library(lattice)
library(caret)
library(tidyverse)
## Warning: package 'tidyverse' was built under R version 4.2.3
## Warning: package 'readr' was built under R version 4.2.3
## Warning: package 'forcats' was built under R version 4.2.3
## — Attaching core tidyverse packages -
                                                             – tidyverse 2.0.0 —
## √ dplyr 1.1.0 √ readr
## √ forcats 1.0.0

√ stringr

                                    1.5.0
## √ lubridate 1.9.2
                        √ tibble
                                    3.1.8
## √ purrr
              1.0.1
                        √ tidyr
                                    1.3.0
## — Conflicts —
                                                       – tidyverse_conflicts() —
## X dplyr::filter() masks stats::filter()
## X dplyr::lag() masks stats::lag()
```

Data exploration

become errors

· Converted EMISSIONS into factor

X purrr::lift() masks caret::lift()

 Since we are prediction emissions based on other variable I had converted emissions with a cutoff point 250, greater than 200 to be 1, otherwise to be 0 also converted engine size ad factor for logistic regression

i Use the]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to

```
fuelData<- read.csv("Fuel_Consumption_2000-2022.csv")
str(fuelData)</pre>
```

```
## 'data.frame': 22556 obs. of 13 variables:
##
                 $ YEAR
  $ MAKE
                        "ACURA" "ACURA" "ACURA" ...
##
                  : chr
  $ MODEL
                 : chr "1.6EL" "1.6EL" "3.2TL" "3.5RL" ...
##
  $ VEHICLE.CLASS : chr "COMPACT" "MID-SIZE" "MID-SIZE" ...
##
                 : num 1.6 1.6 3.2 3.5 1.8 1.8 1.8 3 3.2 1.8 ...
  $ ENGINE.SIZE
##
##
  $ CYLINDERS
                 : int 4466444664...
  $ TRANSMISSION : chr "A4" "M5" "AS5" "A4" ...
##
                  : chr "X" "X" "Z" "Z" ...
## $ FUEL
## $ FUEL.CONSUMPTION: num 9.2 8.5 12.2 13.4 10 9.3 9.4 13.6 13.8 11.4 ...
##
  $ HWY..L.100.km. : num 6.7 6.5 7.4 9.2 7 6.8 7 9.2 9.1 7.2 ...
## $ COMB..L.100.km. : num 8.1 7.6 10 11.5 8.6 8.2 8.3 11.6 11.7 9.5 ...
  $ COMB..mpg. : int 35 37 28 25 33 34 34 24 24 30 ...
##
##
  $ EMISSIONS
                  : int 186 175 230 264 198 189 191 267 269 218 ...
```

summary((fuelData\$YEAR))

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 2000 2006 2012 2012 2017 2022
```

```
fuelData$MAKE <- as.factor(fuelData$MAKE)
fuelData$EMISSIONS <- as.factor (ifelse (fuelData$EMISSIONS>250,1,0))
#fuelData$ENGINE.SIZE <- as.factor(fuelData$ENGINE.SIZE)
#fuelData$CYLINDERS<-as.factor(fuelData$CYLINDERS)
fuelData$EMISSIONS</pre>
```

```
##
       [37] 0 0 0 0 0 1 0 1 0 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 1 0 0 0 1 1 1 1 1 1 1 1 1
##
      ##
     ##
##
     [145] 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
     ##
##
     ##
     [253] 0 0 0 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##
     ##
##
     ##
     ##
     [433] 1 1 0 0 0 0 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 0 1 0 0 1 1 0
##
     ##
     ##
     ##
     ##
     [613] 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1
##
     ##
     ##
     ##
     [757] 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
##
     ##
     ##
     ##
     ##
##
     ##
    ##
    [1045]\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0
    ##
##
    [1117] 1 0 0 1 1 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0
    ##
##
    ##
    ##
    ##
    ##
##
    [1369] 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1
     \begin{smallmatrix} 1405 \end{smallmatrix} ] \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{
##
##
    [1441] 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1
##
    ##
    ##
##
    ##
    ##
    ##
    ##
    [1765] 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0
##
```

```
##
    ##
    ##
##
    ##
    ##
##
     \begin{smallmatrix} 2053 \end{smallmatrix} ] \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{
##
   ##
##
    ##
    ##
    ##
##
    ##
    ##
   ##
    [2449] 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 0
##
   ##
   ##
##
    [2557] 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 1
##
   ##
##
    ##
    ##
##
    ##
    [2845] 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 1 1 0 0
##
##
   ##
   ##
    ##
   [3025] 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1
##
##
    [3097] 1 1 1 1 1 1 1 0 1 0 0 0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1
##
##
    ##
    ##
    ##
   ##
##
    ##
    ##
   [3421] 0 0 0 0 0 0 1 0 0 0 1 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0
##
    ##
    [3493] 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 0 0 1 0 1 0 1 0 0 0
##
    ##
##
    ##
    [3637] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0
```

```
##
    ##
    [3745] 0 0 0 1 1 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 1 1 1 0 0 1 1
##
##
    ##
    ##
##
    ##
   ##
##
    [3997] 1 1 1 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 1
##
    ##
    ##
    [4105] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
##
    ##
    ##
   ##
    [4285] 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 1 0 0 1 0 0 1 0 1 1 1 1 1 0 1 1 1
##
   ##
   ##
##
    [4393] 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 1 0 0 1
##
    \begin{smallmatrix} 4465 \end{smallmatrix} ] \hspace*{0.2cm} 0 \hspace*{0.2cm} 0 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 0 \hspace*{0.2cm} 0 \hspace*{0.2cm} 0 \hspace*{0.2cm} 0 \hspace*{0.2cm} 1 \hspace*{0.2cm} 0 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 0 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 0 \hspace*{0.2cm} 1 \hspace*{0.2cm} 1 \hspace*{0.2cm} 0 \hspace*{0
##
##
    ##
    ##
##
    ##
    [4645] 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0
    [4681] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0
##
   ##
##
   ##
    ##
    ##
##
    ##
##
    ##
##
    ##
   ##
##
    ##
##
   [5257] 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0
##
    ##
    ##
    [5365] 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
##
##
    ##
    [5473] 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 1 0 1 1
```

```
##
      ##
      ##
##
      ##
      ##
##
      ##
      ##
      ##
##
      [5869] 0 0 0 1 1 0 0 0 0 0 0 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##
      [5941] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
##
##
      ##
      ##
       \begin{smallmatrix} [6049] \end{smallmatrix} 0 \hspace{0.1cm} 1 \hspace{0.1cm} 0 \hspace{0.1cm} 1 \hspace{0.1c
      ##
      ##
      ##
      ##
##
      ##
      ##
##
      ##
      ##
##
      ##
      ##
      ##
      ##
      [6625] 1 0 0 0 1 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0
##
##
      ##
##
      ##
##
      ##
      ##
      ##
      ##
##
       \begin{smallmatrix} [6985] \end{smallmatrix} 0 \hspace{.08in} 0 \hspace{.08in} 0 \hspace{.08in} 1 \hspace{.08in} 0 \hspace{.08in} 1 \hspace{.0in} 1 \hspace{.08in} 1 \hspace{.08in
##
      ##
      ##
      [7093] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1
      ##
      ##
      ##
##
      ##
      ##
```

```
##
[7381] 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0
##
##
##
##
##
##
##
[7597] 0 0 0 0 0 1 1 0 1 1 1 1 0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 1 1 0 0 0 0 0 0 0
##
##
[7669] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 1 1 1 0 0 1 0 1 1 0
[7705] 0 1 0 1 1 0 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 0 0
##
##
##
##
##
##
[7921] 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
##
##
##
[8029] 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1 1 0 0 1
##
##
##
##
##
##
##
##
[8281] 0 0 1 0 1 0 1 1 1 1 0 0 1 1 1 1 0 0 0 1 0 1 0 0 0 1 0 0 1 1 1 1 1 0 0 1 1 1
##
##
##
[8389] 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 1 1 1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0
##
##
##
##
##
##
##
##
##
##
[8785] 1 0 0 1 0 0 0 1 1 0 0 1 0 0 0 1 1 0 0 1 0 0 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0
##
##
##
[8857] 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0
##
##
##
##
##
##
##
##
```

```
##
 ##
 ##
##
 ##
 ##
##
 ##
 ##
##
 ##
 ##
 ##
##
 ##
 [9685] 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##
 ##
 [9793] 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1
##
 ##
 ##
##
 [9901] 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0
##
 ##
 [9973] 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 1 0 0 0 0
## [10117] 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
## [10153] 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 1 0 0 0 1 1 1 1 1 0 0 0 0 0 1 1 0 0
## [10189] 0 1 0 0 1 1 1 1 1 0 0 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 1
## [10297] 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0
## [10369] 0 0 0 0 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
## [10405] 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 1 1 1 0 0 0 0 0 1 0 0 0
## [10657] 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0 1 0 0 0 0 0
## [10693] 0 1 1 1 0 0 0 0 1 1 1 0 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1
## [10837] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0
## [10909] 0 0 0 0 1 1 0 1 0 0 0 0 0 0 1 1 1 0 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0
```

```
 \hbox{ \#\# [11233] 0 0 0 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 1 1 } 
## [11377] 0 0 1 1 1 1 1 0 0 1 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 1 0 0 0 0 1 1 1 1 1 1
 \hbox{ \#\# [11413] 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 1 1 1 1 1 1 0 1 0 1 1 1 1 }  
## [11449] 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
## [11737] 1 1 0 1 1 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0
## [11809] 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 0 0 1 1 0 0 0 0 0
## [11989] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0
## [12817] 0 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 1 0 0
```

```
## [12997] 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [13033] 0 0 0 0 1 1 0 1 0 0 0 0 1 0 0 0 1 1 1 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0
## [13285] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 1 1 0 0 0 1 1 1 1 1 1 1
## [13429] 0 0 0 0 0 0 0 0 1 0 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1
## [13573] 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
## [13825] 0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 0
## [13933] 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1
## [14041] 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0
## [14365] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0
## [14581] 1 1 1 1 0 0 0 1 1 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 0 0
```

```
## [14869] 1 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0
## [15013] 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1
## [15049] 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [15265] 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
## [15373] 0 1 0 0 1 1 0 1 0 1 1 0 1 0 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0
## [15409] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0
 \hbox{ \#\# [15481] 0 0 0 0 0 0 1 1 1 1 0 0 1 1 1 1 1 0 1 0 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 } 
## [15517] 1 1 1 1 0 0 0 1 1 0 1 1 0 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 1 1 0 1 1 0 0
## [15553] 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 0
## [15661] 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 0
## [16381] 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0
## [16453] 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 0 1 0
```

```
## [16633] 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 0 1 1 1 0 1 1 1 1
## [16849] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
## [17245] 1 1 1 1 1 0 0 0 0 0 0 1 0 0 1 0 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1
## [17353] 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
## [17389] 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
## [17641] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
## [18469] 0 0 1 1 0 0 1 1 1 1 1 0 0 0 1 0 0 1 1 1 1 1 1 0 0 0 0 1 0 0 0 1 1 1 1 1 0 0
## [18865] 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1
## [19081] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 1 0 0 0 0 0 0 0 0
## [19513] 1 0 0 0 0 1 1 1 1 1 1 1 0 0 1 1 0 0 1 1 1 1 1 0 0 0 1 0 0 1 1 1 1 0 0 0 1 0
## [19585] 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0
## [19621] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 0
## [19693] 0 1 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1
## [19765] 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 0 1 1 1 1 1 1 1 1 0 0 1 0 0 0 1 1 0 0 0 1 1
## [20017] 1 1 1 1 1 1 0 0 1 0 0 0 0 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1
## [20161] 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 1 0 0 0 0 1 1 0 0 1 1 1
```

```
## [20305] 1 0 1 0 1 0 0 0 0 1 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1
## [20773] 0 0 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
## [21493] 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 0 0
## [21637] 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [21709] 1 0 0 0 0 1 1 0 0 1 1 0 0 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0
```

```
## [22069] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 1 0 0 1 1 1 1 1 0 0 0 0 0
## [22177] 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 0 1 1 1 1 1 1 0 1 0 1 0 0 0 0 1 1 1 1 1 1 0 0 0
## [22249] 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
## [22321] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [22501] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 0 0 0
## [22537] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
## Levels: 0 1
```

- Count NA
- We have 22556 observations with 13 variables
- Train 18000, test is the test 4556

```
colSums(is.na(fuelData))
```

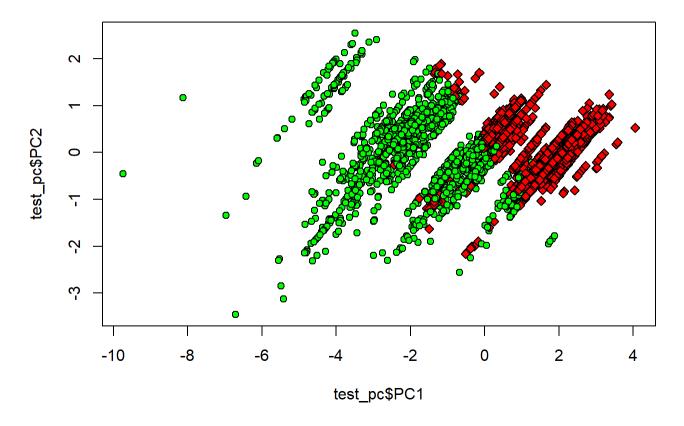
```
##
                YEAR
                                   MAKE
                                                     MODEL
                                                               VEHICLE.CLASS
##
                   0
        ENGINE.SIZE
                             CYLINDERS
                                             TRANSMISSION
                                                                         FUEL
##
##
                                                                            0
                   0
                                                                  COMB..mpg.
  FUEL.CONSUMPTION
                        HWY..L.100.km.
                                          COMB..L.100.km.
##
                                      0
##
           EMISSIONS
##
##
                   0
```

```
i <- sample(1:22566, 18000, replace = FALSE)
train <-fuelData[i,]
test<-fuelData[-i,]
set.seed(1234)
pca_out <-preProcess(train[, c(5,6,10,11)], method = c("center","scale","pca") )
pca_out</pre>
```

```
## Created from 17994 samples and 4 variables
##
## Pre-processing:
## - centered (4)
## - ignored (0)
## - principal component signal extraction (4)
## - scaled (4)
##
## PCA needed 2 components to capture 95 percent of the variance
```

- PCA suggested only 2 components as Variables
- Now plot the PCA
- build GLM using logistic regression and pca components

```
train_pc <- predict(pca_out, train[, c(5,6,10,11)])
test_pc <- predict(pca_out, test[,])
plot(test_pc$PC1, test_pc$PC2, pch=c(23,21,22)[unclass(test_pc$EMISSIONS)],bg=c("red","gree
n","blue")[unclass(test_pc$EMISSIONS)])</pre>
```



```
train_df <- data.frame(train_pc$PC1, train_pc$PC2, train$EMISSIONS)
test_df <- data.frame(test_pc$PC1, test_pc$PC2, test$EMISSIONS)
library(class)
set.seed(1234)
glm1 = glm(EMISSIONS~train_pc$PC1+train_pc$PC2, data = train, family = binomial)</pre>
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
probs <- predict(glm1, newdata = test , type = "response")</pre>
```

Warning: 'newdata' had 4562 rows but variables found have 18000 rows

```
pred<-ifelse(probs> 0.5, 1,0)
acc1<- mean(pred == as.integer(test$EMISSIONS))

## Warning in pred == as.integer(test$EMISSIONS): longer object length is not a
## multiple of shorter object length</pre>
```

```
summary(glm1)
```

```
##
## Call:
## glm(formula = EMISSIONS ~ train_pc$PC1 + train_pc$PC2, family = binomial,
      data = train)
##
##
## Deviance Residuals:
##
      Min
                1Q Median
                                 3Q
                                         Max
## -4.3966 -0.2252 -0.0207 0.1024
                                      2.6367
##
## Coefficients:
##
               Estimate Std. Error z value Pr(>|z|)
                0.29000 0.03243 8.942 <2e-16 ***
## (Intercept)
## train pc$PC1 -3.30594  0.05853 -56.479  <2e-16 ***
## train pc$PC2 -2.00396  0.06309 -31.765  <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## (Dispersion parameter for binomial family taken to be 1)
##
      Null deviance: 24805.0 on 17993 degrees of freedom
##
## Residual deviance: 7050.1 on 17991 degrees of freedom
    (6 observations deleted due to missingness)
##
## AIC: 7056.1
##
## Number of Fisher Scoring iterations: 8
```

```
print(paste("glm1 has accuracy = ",acc1))
```

```
## [1] "glm1 has accuracy = NA"
```

- Try logistic regression on the non-reduced data
- · set ground truth

```
library(class)
set.seed(1234)
glm2 = glm(EMISSIONS~ HWY..L.100.km.+COMB..L.100.km.+ENGINE.SIZE+CYLINDERS, data = train, fam
ily = binomial)
```

pred

##

##

1

0 2424 1451 96 1897

2

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
probs <- predict(glm2, newdata = test , type = "response")</pre>
pred<-ifelse(probs> 0.5, 1,0)
acc2<- mean(pred == as.integer(test$EMISSIONS))</pre>
summary(glm2)
##
## Call:
## glm(formula = EMISSIONS ~ HWY..L.100.km. + COMB..L.100.km. +
      ENGINE.SIZE + CYLINDERS, family = binomial, data = train)
##
##
## Deviance Residuals:
      Min
                1Q Median 3Q
##
                                         Max
## -4.3903 -0.2246 -0.0204 0.1051
                                      2.7184
##
## Coefficients:
##
                  Estimate Std. Error z value Pr(>|z|)
## (Intercept) -24.10398 0.43855 -54.963 < 2e-16 ***
## HWY..L.100.km. 1.18967 0.07228 16.460 < 2e-16 ***
## COMB..L.100.km. 0.91262 0.06765 13.490 < 2e-16 ***
                              0.07035 4.677 2.91e-06 ***
## ENGINE.SIZE
                  0.32903
                0.44349 0.05336 8.312 < 2e-16 ***
## CYLINDERS
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## (Dispersion parameter for binomial family taken to be 1)
##
##
      Null deviance: 24805.0 on 17993 degrees of freedom
## Residual deviance: 7036.2 on 17989 degrees of freedom
##
    (6 observations deleted due to missingness)
## AIC: 7046.2
##
## Number of Fisher Scoring iterations: 8
print(paste("glm2 has accuracy = ",acc2))
## [1] "glm2 has accuracy = 0.0210434020166594"
table(pred, as.integer(test$EMISSIONS))
##
```

```
Accuracy from model is around 2% Confusion matrix shows the distribution of prediction where correct and
```

incorrect being around 0.50 margin.

LDA

```
library(MASS)
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##
       select
lda1<-lda(EMISSIONS~ HWY..L.100.km.+COMB..L.100.km.+ENGINE.SIZE+CYLINDERS,data =train)
lda1$means
    HWY..L.100.km. COMB..L.100.km. ENGINE.SIZE CYLINDERS
## 0
          7.385158
                           9.047314
                                       2.486813 4.721348
## 1
          10.758776
                          13.409739
                                       4.386665 7.194905
```

Above data shows a clear positive relativity between EMISSION and the rest of the predictors 0 classified vehicles with emission less than 250, and 1 is more than 250

```
lda_pred<- predict(lda1, newdata = test,type = "class")
lda_pred$class</pre>
```

```
##
  ##
  [38] 1 1 1 0 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 1 1 1 1
##
  ##
 ##
 ##
##
 ##
 [297] 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1
##
 [334] 1 1 1 1 0 0 0 0 0 1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 0
##
##
 ##
 [445] 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 1
##
##
 ##
 [556] 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 1 0 0 1 1
##
 ##
##
 [630] 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 1 0 1
##
 ##
##
 [778] 1 1 1 1 1 1 0 0 0 1 0 0 1 1 0 0 1 1 1 0 1 1 0 0 0 1 1 1 1 1 1 0 0 1 0 1 1 0
##
##
 [852] 0 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 1 0 0 0 0 0
##
##
 ##
 ##
## [1000] 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 1 0 0 1 1 1 1 0 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0
## [1037] 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 1 1 0 1 1 1 1 1 0 0 0 0 0 1 1 0 0
## [1148] 0 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1
## [1185] 0 0 0 0 1 1 0 0 1 1 1 1 1 0 1 1 0 0 0 0 0 1 0 1 0 1 0 0 1 1 1 1 1 1 1 1 1
## [1259] 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0
## [1333] 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0
## [1481] 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
## [1555] 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 1 1 0 0 1 1 1 0 0 0 0
 \hbox{ \#\# [1629] 1 1 1 0 0 0 0 0 0 0 1 0 1 1 1 1 0 1 1 1 1 0 0 0 1 0 0 1 1 0 0 0 0 1 1 1 1 1 1 } 
## [1666] 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 1 0 1 1 1 1 1 0 0 0 1 0 1 0
## [1703] 0 0 0 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 0 1 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 0
## [1851] 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0
```

```
## [2221] 1 0 0 0 0 0 1 0 0 0 0 0 1 1 0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0
## [2332] 0 1 0 0 0 0 0 1 0 1 1 0 0 0 1 0 0 0 1 0 0 0 1 1 1 0 1 0 1 1 1 0 0 0 0 1
## [2406] 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [2554] 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0
## [2591] 0 0 0 0 1 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0
## [2628] 0 1 0 1 1 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
## [2665] 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 1 0 0 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0
## [2702] 0 0 0 0 1 1 1 0 0 0 1 1 0 0 0 0 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0
## [2739] 0 0 0 0 1 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1
## [2776] 1 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0
## [2813] 0 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
## [2887] 0 0 0 1 0 1 1 1 1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 1 1 1 1 0 0 1 1 0 0 0 0 0 1 1 0
## [3072] 1 1 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 1 1 1 1 1 0 0 0 0
## [3109] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 1 0 0 0 0
## [3220] 1 0 0 0 1 1 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0
## [3257] 0 0 0 1 0 0 0 0 1 1 0 0 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 0 0 1 1 1
## [3294] 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0
## [3442] 1 0 1 1 1 1 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 1 1 0 0 1 0 0 0 1 0 1 1 0
## [3516] 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
## [3590] 1 0 0 1 1 0 0 0 1 0 1 1 0 0 0 1 1 0 0 0 1 1 0 1 0 1 0 1 0 0 1 1 1 1 1 1 1
## [3627] 1 1 0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 0 1 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 0
## [3664] 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 1 1 1 0 1 0 0 0 1 0 1 1 0
```

```
## [3812] 0 1 1 0 1 1 0 0 0 0 0 1 1 0 0 1 0 0 1 0 0 0 1 0 0 0 0 1 1 1 1 1 0 1 1 0 1
## [3849] 1 1 0 1 1 1 0 0 1 0 0 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 1 1 1 1 1 1 0 1 0 0 0
## [3886] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 0 1 1 1 0 0 0 1 0 1 1 1 1 0 0 1 0 0 0
## [3997] 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 0 0 1 1 1 0 0 0 0 1 1 1 0 0 0 0 0 0
## [4108] 0 0 1 0 1 1 1 1 1 1 0 1 0 1 1 1 0 0 0 1 0 1 1 0 0 0 0 0 0 0 1 0 0 0 1 1 1 1 1 1
## [4145] 1 0 0 0 1 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 1 0 1 1 1 1 1 1 1 0
## [4256] 1 1 1 1 1 0 1 0 1 0 0 1 1 1 0 1 1 0 0 0 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 1 0
## [4330] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1 1 0 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0
## [4441] 1 0 1 1 1 0 0 0 1 0 1 1 1 1 1 1 1 1 0 1 0 0 1 0 0 1 0 0 0 0 0 1 1 1 1 1 0 0
## [4515] 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
## [4552] 0 0 0 1 1 0 0 0 0 0 0
## Levels: 0 1
```

Summary: Predicted result shows an accuracy around 89% on the unreduced data. The LCA has not has as high as an accuracy rate comparing to PDA. where it was 95% accuracy using two predictors. Logistic regression wasn't able to predict anything practically using the reduced data

```
mean(lda_pred$class == test$EMISSIONS)

## [1] 0.9013591
```

22 of 22