MPG in popular car models

john myers

2/23/2022

We will use the mpg dataset in R to explore the relationships between engine features and mpg. This will give us insight on how mpg is affected by different features of the car. We will also look at newer cars versus older cars to see if there is a change in fuel efficiency.

**library**(tidyverse)

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

## v ggplot2 3.3.5 v purrr 0.3.4

## v tibble 3.1.6 v dplyr 1.0.8

## v tidyr 1.2.0 v stringr 1.4.0

## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --

## x dplyr::filter() masks stats::filter()

## x dplyr::lag() masks stats::lag()

Now import the dataset from RStudio.

data <- mpg

head(data)

## # A tibble: 6 x 11

## manufacturer model displ year cyl trans drv cty hwy fl class

## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>

## 1 audi a4 1.8 1999 4 auto(l5) f 18 29 p compa~

## 2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compa~

## 3 audi a4 2 2008 4 manual(m6) f 20 31 p compa~

## 4 audi a4 2 2008 4 auto(av) f 21 30 p compa~

## 5 audi a4 2.8 1999 6 auto(l5) f 16 26 p compa~

## 6 audi a4 2.8 1999 6 manual(m5) f 18 26 p compa~

Checking for null values.

data %>% is.na() %>% colMeans()

## manufacturer model displ year cyl trans

## 0 0 0 0 0 0

## drv cty hwy fl class

## 0 0 0 0 0

Comparing MPG highway with number of cylinders in the engine.There is a decrease in mpg when the car has a greater number of cylinders.

data %>%

ggplot(aes(cyl,hwy)) + geom\_point(size=3, color = "blue") + geom\_point() + geom\_smooth(method = 'lm') + labs(title = "Highway mpg versus number of cylinders")

## `geom\_smooth()` using formula 'y ~ x'

Chart, line chart

Description automatically generated

This chart shows that both transmission types get better mpg on the highway than in the city. Manuel transmissions tend to get better mpg versus their automatic counterpart. Manuel transmissions do not have the highest mpg in the city or highway category.

data %>%

group\_by(trans) %>%

summarise(hwy = mean(hwy), cty = mean(cty)) %>%

gather(key, val, -trans) %>%

ggplot(aes(y=trans,x=val)) + geom\_col(position = 'dodge') +

facet\_wrap(~key, scales = 'free\_x') +

labs(title = "mpg based on transmission type")

Chart

Description automatically generated

This shows the transmission type versus the highway mpg with the color on the Blue scale representing year of the car. You can see that the newer the car, the better the mpg when transmissions are compared with transmissions of the same type. Volkswagon had one transmission type not follow that pattern.

data %>%

ggplot(aes(cty, hwy, color = year)) +

geom\_point(size = 3, alpha = 0.5) +

geom\_smooth() +

facet\_wrap(~manufacturer) +

labs(title = "Transmission\_versus\_Highway\_mpg\_with\_year\_of\_car")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 24

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 4

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at 24

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 1

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition

## number 0

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : There are other near

## singularities as well. 4

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 13

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 2

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 3.3901e-017

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 4

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at 13

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 2

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition

## number 3.3901e-017

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : There are other near

## singularities as well. 4

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 10.995

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 1.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 1.01

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : span too small. fewer

## data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at

## 10.995

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius

## 1.005

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition

## number 0

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : There are other near

## singularities as well. 1.01

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : at 10.995

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : radius 2.5e-005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 10.995

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 0.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 1.01

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : zero-width neighborhood. make span bigger

## Warning: Computation failed in `stat\_smooth()`:

## NA/NaN/Inf in foreign function call (arg 5)

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : at 12.995

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : radius 2.5e-005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 12.995

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 0.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 1.01

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : zero-width neighborhood. make span bigger

## Warning: Computation failed in `stat\_smooth()`:

## NA/NaN/Inf in foreign function call (arg 5)

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 15.99

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 1.01

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 1.0201

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : span too small. fewer

## data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at

## 15.99

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 1.01

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition

## number 0

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : There are other near

## singularities as well. 1.0201

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : pseudoinverse used at 17.985

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : neighborhood radius 2.015

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : reciprocal condition number 4.1478e-017

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =

## parametric, : There are other near singularities as well. 4.0602

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at

## 17.985

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius

## 2.015

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition

## number 4.1478e-017

## Warning in predLoess(object$y, object$x, newx = if

## (is.null(newdata)) object$x else if (is.data.frame(newdata))

## as.matrix(model.frame(delete.response(terms(object)), : There are other near

## singularities as well. 4.0602

Chart, scatter chart

Description automatically generated

Now lets take a look into Volkswagen hwy mpg versus transmission.The graph show that older models of Volkswagon had better mpg in auto(4), which means automatic 4 cylinders, and in manual(5).

data %>%

filter(manufacturer == 'volkswagen') %>%

ggplot(aes(trans, hwy, color = year)) +

geom\_point(size = 3, alpha = 0.5) +

labs(title = "volkswagen transmission versus mpg")

Chart, scatter chart

Description automatically generated

You can tell that the older models have better fuel efficiency and that a higher highway mpg translates into a higher city mpg.

data %>%

filter( manufacturer == 'volkswagen') %>%

ggplot(aes(cty, hwy, color = year)) +

geom\_point(size = 3, alpha = 0.5) +

geom\_smooth(method='lm') +

labs(title = "volkswagen transmission versus mpg")

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

Conclusion:

The greater number of cylinders in an engine the less the mpg of the car. Volkswagen trended by decreasing mpg with a newer car and the rest of the car companies saw an upwards trend in mpg for a newer car. This could be due to the newer Volkswagen being a heavier type of vehicle. That would be an area for further study.