Exploration of MPG dataset
by Eugeniah Arthur

About dataset

The dataset is the mpg data in R . It consists of 234 rows and 11 variables. Below are the variables in red with brief descriptions by them.

Manufacturer, Model , displ(engine displacement, in litres(L)) ,year(year of manufacture),cyl (number of cylinders), trans(type of transmission), drv (f = front-wheel drive, r = rear wheel drive, 4 = 4wd), cty (city miles per gallon), hwy (highway miles per gallon), fl (fuel type), class("type" of car)

Questions to be answered after this exploration

- What is the typical Engine Displacement of a car? Are there any extreme values ?

- What is the typical engine displacement for each type of wheel ?
- What is the typical engine displacement for each class of car ? Is the spread stable, if not is there a transformation that will stabilize the spread?
- What is the most efficient wheel type for within city transport in terms of mpg?Identify the specific manufacturer and model of car that is unusual in that category .

- What is the most efficient four wheel car for within city transport ?

What is the typical Engine Displacement of a car? Are there any extreme values ?

Engine displacement of a car is right skewed. Most engine displacement are smaller sized. It's distribution is bimodal.

Median: 3.3

Inner fence:[-0.9, 7.9]

Outer fence:[-4.2, 11.2]

There is no datapoint that fall outside the inner fence. Hence, there is no outlier or unusual data point. The typical engine displacement is 3.3 litres.

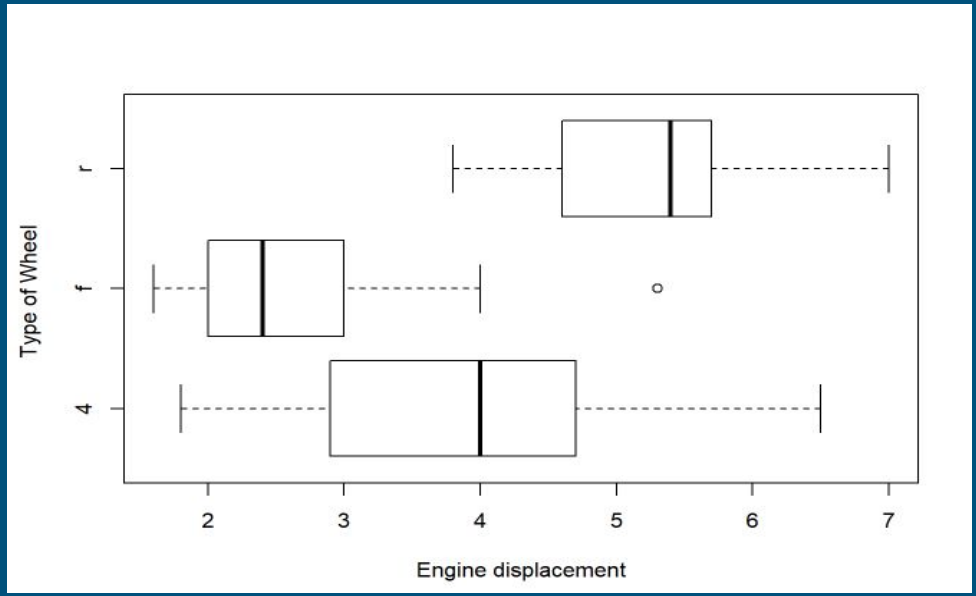
```
## 1 | 2: represents 1.2
##  leaf unit: 0.1
##             n: 234
##    22    1. | 66666888888888888999
##    62    2* | 00000000000000000000022222444444444444
##   100    2. | 55555555555555555557777777788888888
##   (27)   3* | 000000001111113333333334444
##   107    3. | 555556677788888888999
##    86    4* | 00000000000000022224
##    66    4. | 66666666667777777777777777
##    38    5* | 00222223333344444444
##    17    5. | 67777777799
##     6    6* | 0122
##     2    6. | 5
##     1    7* | 0
```

What is the typical engine displacement for each type of wheel ?

A parallel boxplot shows that the engine displacement of the various categories is slightly skewed. The median is a better description of the typical engine displacement for the various wheel type.

The typical Engine displacement is therefore 4L for 4wd, 2.4L for front wheel drive and 5.4L for rear view drive.

Front wheel drive has the least engine displacement.



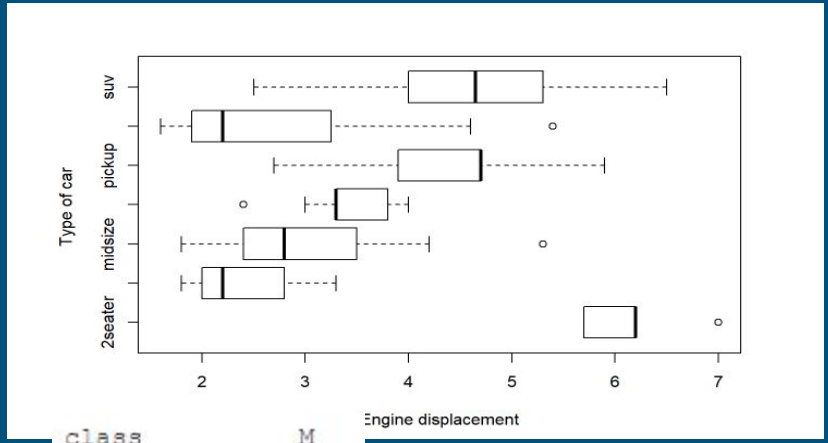

```
## # A tibble: 3 x 2
##   drv       M
##   <chr> <dbl>
## 1 4       4
## 2 f       2.4
## 3 r       5.4
```

What is the typical engine displacement for each class of car ?

A look at the parallel boxplot shows that the engine displacement of the different class of car is skewed and have outliers. The median is a good measure of the typical engine displacement. The spread of the groups are very different.

A 2 seater car has an average (typical) engine displacement of 6.2 L, compact (2.2 liters) , midsize(2.8liters)) , minivan(3.3liters)) , pickup (4.7 liters) , subcompact(2.2 liters),SUV(4.65liters).

Compact cars have the least engine displacement and 2 seater cars have the highest typical engine displacement. The typical for minivan is the same as the typical of engine displacement without any categories.



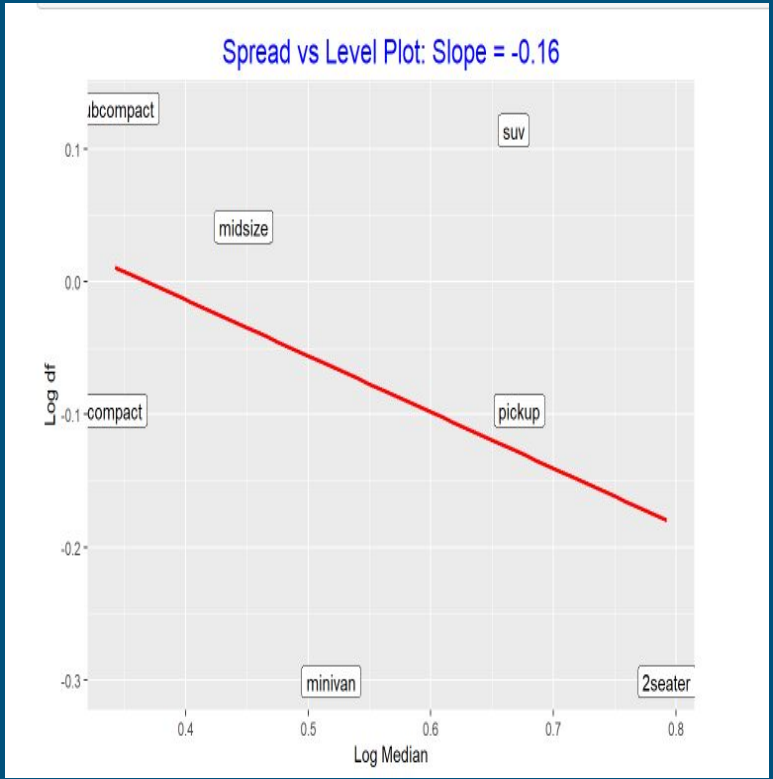
class	M
<chr>	<dbl>
2seater	6.2
compact	2.2
midsize	2.8
minivan	3.3
pickup	4.7
subcompact	2.2
suv	4.65

Spread Vs level plot

From the previous plot, the variation of the different batches differ.

The spread vs level plot shows a transformation may improve the spread in the data.
There exist a negative relationship between the spread and the levels of the class of car. Thus, as the levels increases the spread tends to decrease.

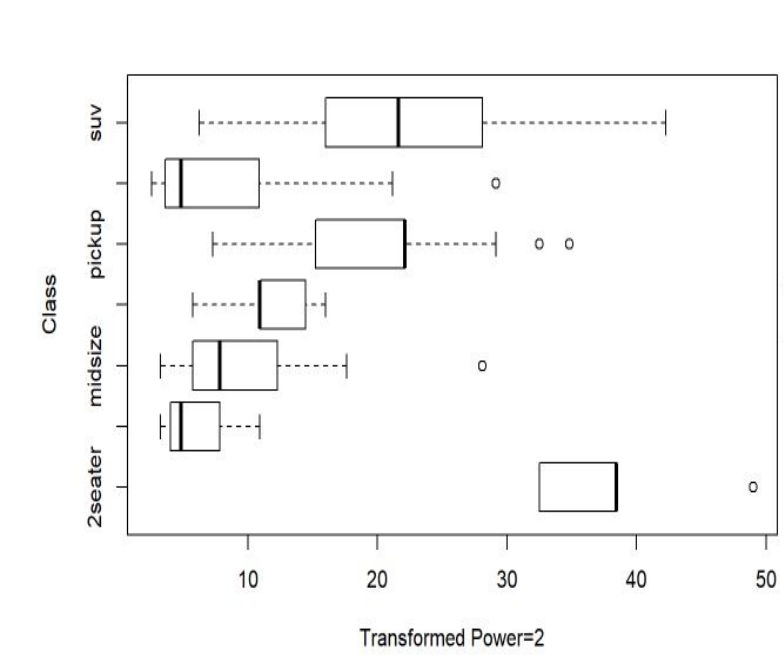
A transformation of
1-(- 0.16) = 1.16 approximately 2 may improve the spread.

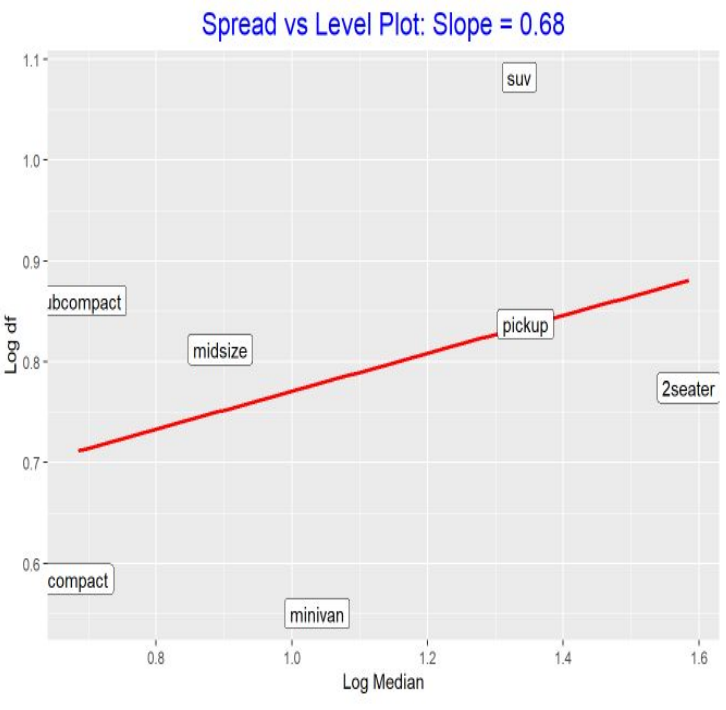



```
# A tibble: 7 x 5
  class         M    df log.M  log.df
  <chr>     <dbl> <dbl> <dbl>   <dbl>
1 2seater     6.2  0.5   0.792 -0.301
2 compact     2.2  0.800 0.342 -0.0969
3 midsize     2.8  1.1   0.447  0.0414
4 minivan     3.3  0.5   0.519 -0.301
5 pickup      4.7  0.8   0.672 -0.0969
6 subcompact  2.2  1.35  0.342  0.130
7 suv         4.65 1.30  0.667  0.114
```

Transformation to stabilize spread.

After transforming the engine displacement variable by squaring it .
The spread vs level plot shows an improvement though it is not very obvious from the boxplot of the batches.
Expressing the reexpressed data using a transformation of 1-0.68=0.32 approx 0.5

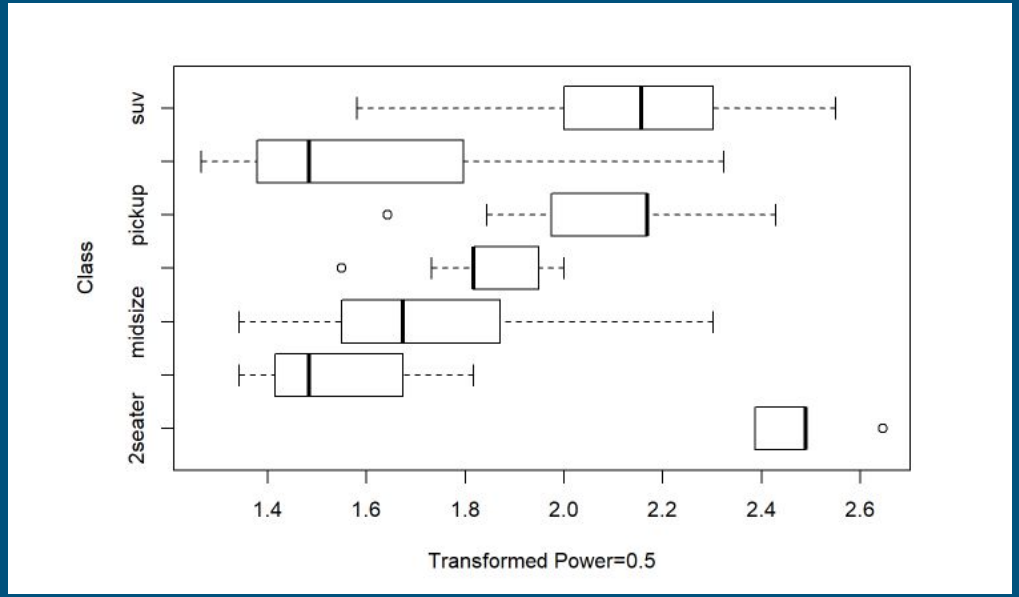




Final Re expression

The final transformation where the transformed was transformed again to a power of 0.5 seems to do a better job .

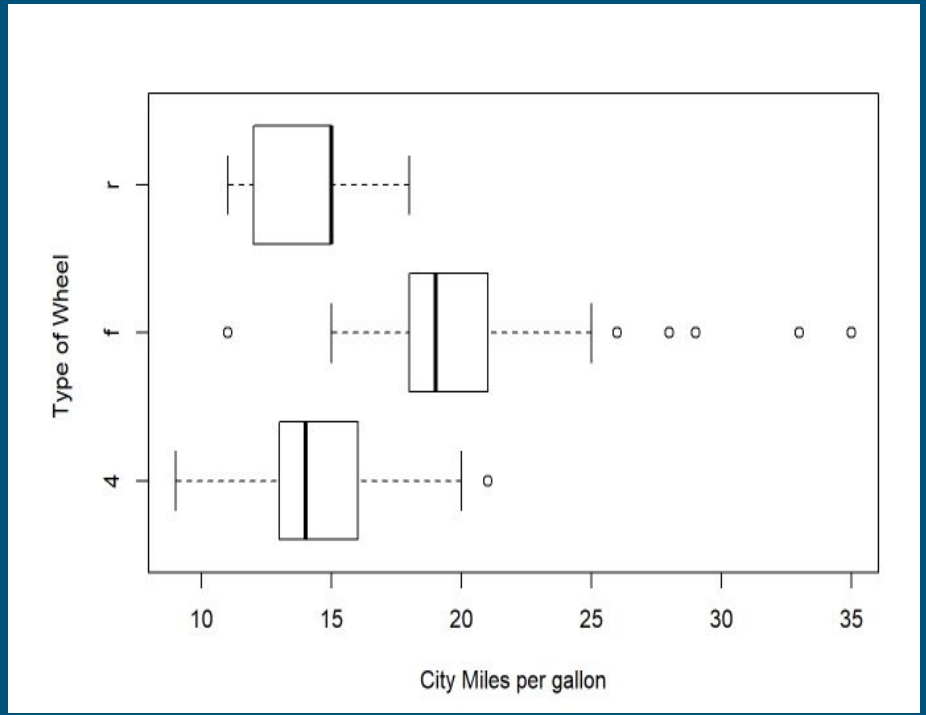
The boxplot shows that the variation of the engine displacement within the different class of car has reduced tremendously. This is an improvement on the original.



What is the most efficient wheel type for within city transport in terms of mpg?

For this, a boxplot of the variable city miles per gallon is plotted based on the type of wheels.

It is realised that the variation of the types of wheels are similar. However, front wheel drive had lots of outliers. Among the three types of Wheels, the 4wd has the least typical value followed by the rear wheel drive. The front wheel drive had the highest typical value. Therefore, On the average Front wheel drive is the most efficient Type of wheel for within city transportation.



Identifying Front Wheel Cars with unusual city mpg

From the previous slide, it was realized that the most efficient type of wheel for driving within the city was front wheel drive. I would like to identify the cars within this group that have unusual mpg within city
Inner fence : [13.5 ,25.5] Outer fence: [9, 30]
The type of cars that fell outside the innerfence
Are shown in the output.

Volkswagen New beetle(1999 model) manual transmission had the most efficient within city Mpg among the 8 extreme models.

Furthemore, Dodge Caravan(2008 model) is the least efficient front wheel drive .
Its mpg is 11 miles per gallon.

```
## # A tibble: 9 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 dodge        carav~   3.3  2008     6 auto~ f        11    17 e     mini~
## 2 honda        civic    1.6  1999     4 manu~ f        28    33 r     subc~
## 3 honda        civic    1.8  2008     4 manu~ f        26    34 r     subc~
## 4 toyota       corol~   1.8  1999     4 manu~ f        26    35 r     comp~
## 5 toyota       corol~   1.8  2008     4 manu~ f        28    37 r     comp~
## 6 toyota       corol~   1.8  2008     4 auto~ f        26    35 r     comp~
## 7 volkswagen   jetta    1.9  1999     4 manu~ f        33    44 d     comp~
## 8 volkswagen   new b~   1.9  1999     4 manu~ f        35    44 d     subc~
## 9 volkswagen   new b~   1.9  1999     4 auto~ f        29    41 d     subc~
```

Identifying Four wheel drive with unusual City MPG

The four wheel drive had the least efficient city mpg. However, from the boxplot, it is realised that there is an unusual mpg in the group.

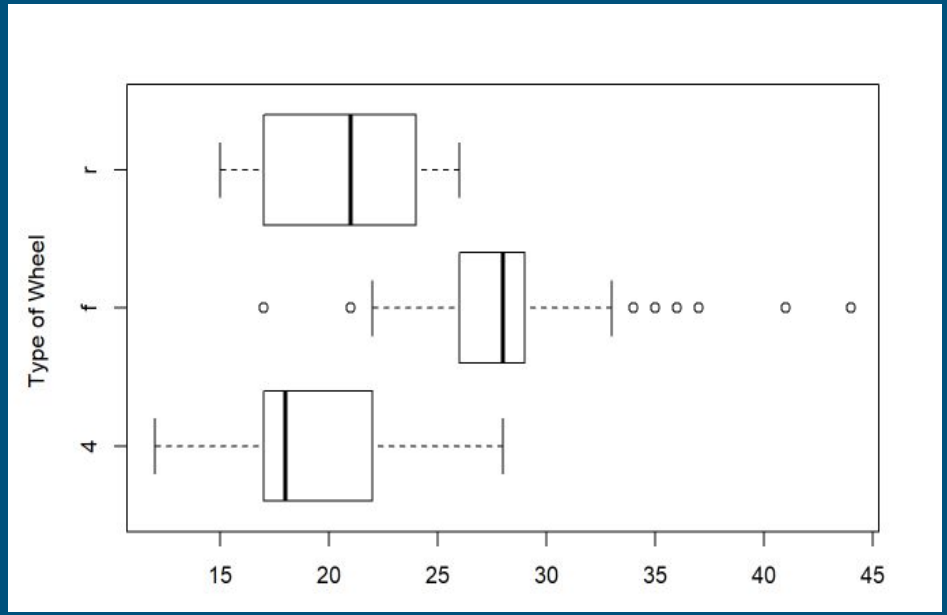
Innerfence=[8.5, 20.5]
Outer fence =[4,25]

Subaru impreza awd(1999 model) is the most Efficient 4 wheel drive for within city movement With an unusual mpg of 21

```
# A tibble: 1 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 subaru       impre~   2.2  1999     4 auto~ 4        21    26 r     subc~
```

What is the most efficient wheel type for highway transport in terms of mpg?

- 4 wheel drives have lower miles per gallon .
- Rear wheel drives have a symmetric shape with moderate highway mpg.
- The front wheel drive has the
- least variation withad typical mpg of around 28
- There are unusual highway mpg for the front wheel drive.
- Front wheel cars has the most
- efficient highway mpg.



Identifying Front Wheel Cars with unusual highway mpg

Inner fence:[21.5, 33.5]
outerfence=[17, 38]
The dodge caravan(2008&1999)
Are the least efficient front wheel cars on the highway in terms of mpg.

Volkswagen new beetle(1999) and Volkswagen Jetta is the most efficient front wheel car for the highway.

```
## # A tibble: 11 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 dodge        cara~   3.3  2008     6 auto~ f        11    17 e     mini~
##  2 dodge        cara~   3.8  1999     6 auto~ f        15    21 r     mini~
##  3 honda        civic   1.8  2008     4 manu~ f        26    34 r     subc~
##  4 honda        civic   1.8  2008     4 auto~ f        25    36 r     subc~
##  5 honda        civic   1.8  2008     4 auto~ f        24    36 c     subc~
##  6 toyota       coro~   1.8  1999     4 manu~ f        26    35 r     comp~
##  7 toyota       coro~   1.8  2008     4 manu~ f        28    37 r     comp~
##  8 toyota       coro~   1.8  2008     4 auto~ f        26    35 r     comp~
##  9 volkswagen   jetta   1.9  1999     4 manu~ f        33    44 d     comp~
## 10 volkswagen   new ~   1.9  1999     4 manu~ f        35    44 d     subc~
## 11 volkswagen   new ~   1.9  1999     4 auto~ f        29    41 d     subc~
```

Summary and Conclusion

- With respect to the data;
- The typical engine displacement of a car is 3.3L.
- Four wheel drive, front wheel drive and rear wheel drive have 4L, 2.4 L and 5.4L respectively.
- A 2 seater car has an average (typical) engine displacement of 6.2 L, compact (2.2 L) , midsize(2.8L)) , minivan(3.3L)) , pickup (4.7 L) , subcompact(2.2 L),SUV(4.65L).

- The engine displacement data had to be transformed to reduce difference in variability between different classes of cars.

- The most efficient type of wheel for within city driving is a front wheel drive car. Volkswagen New beetle(1999 model) manual transmission is the most efficient within city car in terms of mileage .Dodge Caravan(2008 model) is the least efficient front wheel drive car.

- Subaru impreza awd(1999 model) is the most efficient 4 wheel drive for within city movement having an unusual mpg of 21
- The most efficient type of wheel for highway driving is the front wheel car . Volkswagen new beetle(1999) and Volkswagen Jetta is the most efficient FWD car for the highway.

Appendix and R code

Finding Medians and parralel plots

```
library(LearnEDAfunctions)
attach(mpg)
median(displ)
summarize(group_by(mpg, class),
M=median(displ, na.rm=TRUE))
boxplot(displ~ class, horizontal =
TRUE,
data = mpg, xlab = "Engine
displacement",
ylab = "Type of car ")
summarize(group_by(mpg, drv),
M=median(displ, na.rm=TRUE))
boxplot(displ~ drv, horizontal = TRUE,
data = mpg, xlab = "Engine
displacement",
ylab = "Type of Wheel ")
stem.leaf(displ,0.1)
```

Finding inner and outer fence

```
vals=lval(displ)
vals
fs=vals$spreads[2] #fourth
Spread
stp=1.5*fs
#stp
lifim=vals$lo[2]-stp
#upper inner fence
uifim=vals$hi[2]+stp
#lower outer fence
lofim=vals$lo[2]-2*stp
#upper outer fence
uofim=vals$hi[2]+2*stp
innerfence=list(lifim,uifim)
innerfence
outerfence=list(lofim,uofim)
outerfence
# FInding extreme salaries
w=which(displ>=uifim)
mpg[w,]
```

Finding outliers in 4WD

```
vals=lval(d4wd)
vals
fs=vals$spreads[2] #fourth
Spread
stp=1.5*fs
stp
lifim=vals$lo[2]-stp
#upper inner fence
uifim=vals$hi[2]+stp
#lower outer fence
lofim=vals$lo[2]-2*stp
#upper outer fence
uofim=vals$hi[2]+2*stp
innerfence=list(lifim,uifim)
innerfence
outerfence=list(lofim,uofim)
outerfence
# Finding extreme salaries
w=which(d4wd>=uifim)
s4wd[w,]
```

Transforamtion and reexpression of engine displacement to reduce spread differences

```
mpg1=mpg
spread_level_plot(mpg1,displ,class)
mpg1$displc=(mpg1$displ)^(2)
boxplot(displc ~ class, horizontal =
TRUE,
data = mpg1, xlab = "Transformed
Power=2",
ylab = "Class")
spread_level_plot(mpg1,displc,class)
mpg1$displcc=(mpg1$displ)^(0.5)
boxplot(displcc ~ class, horizontal =
TRUE,
data = mpg1, xlab = "Transformed
Power=0.5",
ylab = "Class")
```