

Overview of “Auto” Dataset from “ISLR” Package

Julian Hatwell

February 29, 2016

This document provides a brief overview of the Auto dataset in the ISLR R package.

```
##           mpg           cylinders      displacement      horsepower
##  Min.      : 9.00      Min.      :3.000      Min.      : 68.0      Min.      : 46.0
##  1st Qu.:17.00      1st Qu.:4.000      1st Qu.:105.0      1st Qu.: 75.0
##  Median :22.75      Median :4.000      Median :151.0      Median : 93.5
##  Mean   :23.45      Mean   :5.472      Mean   :194.4      Mean   :104.5
##  3rd Qu.:29.00      3rd Qu.:8.000      3rd Qu.:275.8      3rd Qu.:126.0
##  Max.   :46.60      Max.   :8.000      Max.   :455.0      Max.   :230.0
##
##           weight      acceleration           year           origin
##  Min.      :1613      Min.      : 8.00      Min.      :70.00      Min.      :1.000
##  1st Qu.:2225      1st Qu.:13.78      1st Qu.:73.00      1st Qu.:1.000
##  Median :2804      Median :15.50      Median :76.00      Median :1.000
##  Mean   :2978      Mean   :15.54      Mean   :75.98      Mean   :1.577
##  3rd Qu.:3615      3rd Qu.:17.02      3rd Qu.:79.00      3rd Qu.:2.000
##  Max.   :5140      Max.   :24.80      Max.   :82.00      Max.   :3.000
##
##
##           name
##  amc matador      : 5
##  ford pinto       : 5
##  toyota corolla   : 5
##  amc gremlin      : 4
##  amc hornet       : 4
##  chevrolet chevette: 4
##  (Other)          :365
```

From the summary, and the associated help (not shown), the following observations can be made:

The dataframe contains 392 rows and 9 columns.

Correlogram of Auto

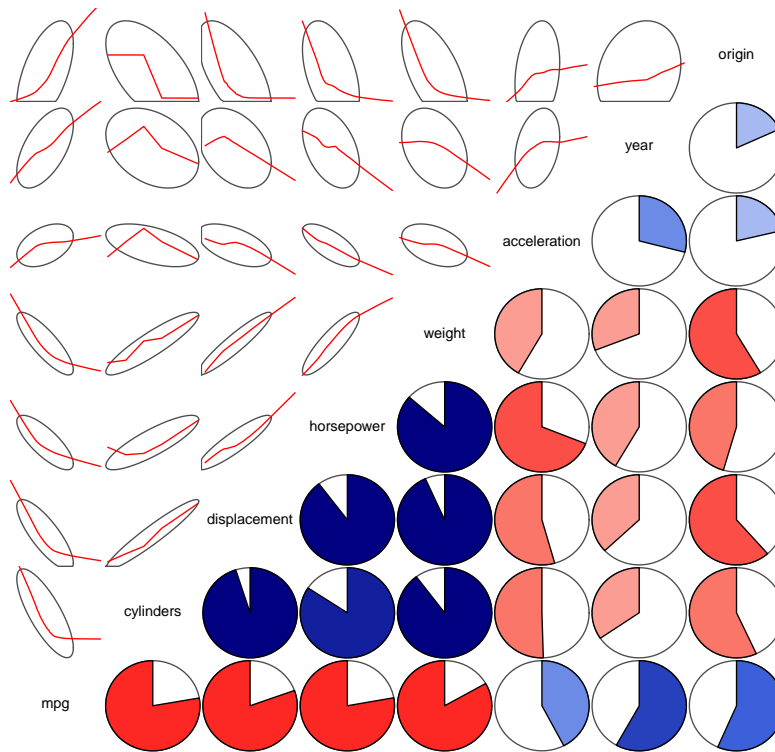


Figure 1: multi-variate comparisons

```
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.66463  -0.13786   0.05178   0.53133   3.05319
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   8.4245     0.7449   11.31  <2e-16 ***
## cylinders    -1.6345     0.1508  -10.84  <2e-16 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 254.40  on 390  degrees of freedom
## AIC: 258.4
##
## Number of Fisher Scoring iterations: 6
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5813  -0.3162   0.1239   0.4695   3.4264
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.585837   0.521290   10.71  <2e-16 ***
## displacement -0.032723   0.003265  -10.02  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 233.16  on 390  degrees of freedom
## AIC: 237.16
##
## Number of Fisher Scoring iterations: 6
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2025  -0.3445   0.0614   0.5228   2.6467
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  9.29885    1.00692   9.235  <2e-16 ***
## horsepower  -0.09675    0.01080  -8.958  <2e-16 ***

```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 276.51  on 390  degrees of freedom
## AIC: 280.51
##
## Number of Fisher Scoring iterations: 6
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.47456  -0.29696   0.04822   0.38160   2.93790
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 11.763284   1.162411  10.120  <2e-16 ***
## weight      -0.004119   0.000413  -9.974  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 228.31  on 390  degrees of freedom
## AIC: 232.31
##
## Number of Fisher Scoring iterations: 6
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00445  -1.05483  -0.05066   1.09337   1.72398
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.52830    0.71043  -6.374 1.84e-10 ***

```

```

## acceleration 0.29193 0.04539 6.431 1.26e-10 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 543.43 on 391 degrees of freedom
## Residual deviance: 492.71 on 390 degrees of freedom
## AIC: 496.71
##
## Number of Fisher Scoring iterations: 4
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.88994 -0.96401 -0.00187  0.95881  1.88399
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -20.20537    2.53754  -7.963 1.68e-15 ***
## year         0.26595    0.03337   7.969 1.59e-15 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 543.43 on 391 degrees of freedom
## Residual deviance: 466.32 on 390 degrees of freedom
## AIC: 470.32
##
## Number of Fisher Scoring iterations: 4
##
## Call:
## glm(formula = fmla1, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2814 -0.8559 -0.2318  0.8419  1.5374
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)

```

```

## (Intercept) -2.4859      0.2846  -8.735  <2e-16 ***
## origin      1.6704      0.1919   8.705  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 426.69  on 390  degrees of freedom
## AIC: 430.69
##
## Number of Fisher Scoring iterations: 4
##
## Call:
## glm(formula = fmla, family = binomial, data = df1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4277  -0.1061   0.0080   0.2123   3.1631
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -17.154875   5.763805  -2.976 0.002917 **
## cylinders    -0.162589   0.423195  -0.384 0.700835
## displacement  0.002095   0.012034   0.174 0.861789
## horsepower   -0.041019   0.023872  -1.718 0.085750 .
## weight       -0.004315   0.001140  -3.784 0.000154 ***
## acceleration  0.016065   0.141462   0.114 0.909582
## year          0.429459   0.075225   5.709 1.14e-08 ***
## origin        0.477339   0.362014   1.319 0.187314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 157.54  on 384  degrees of freedom
## AIC: 173.54
##
## Number of Fisher Scoring iterations: 8

```



```

## [1] "cylinders"
##
##      0      1

```

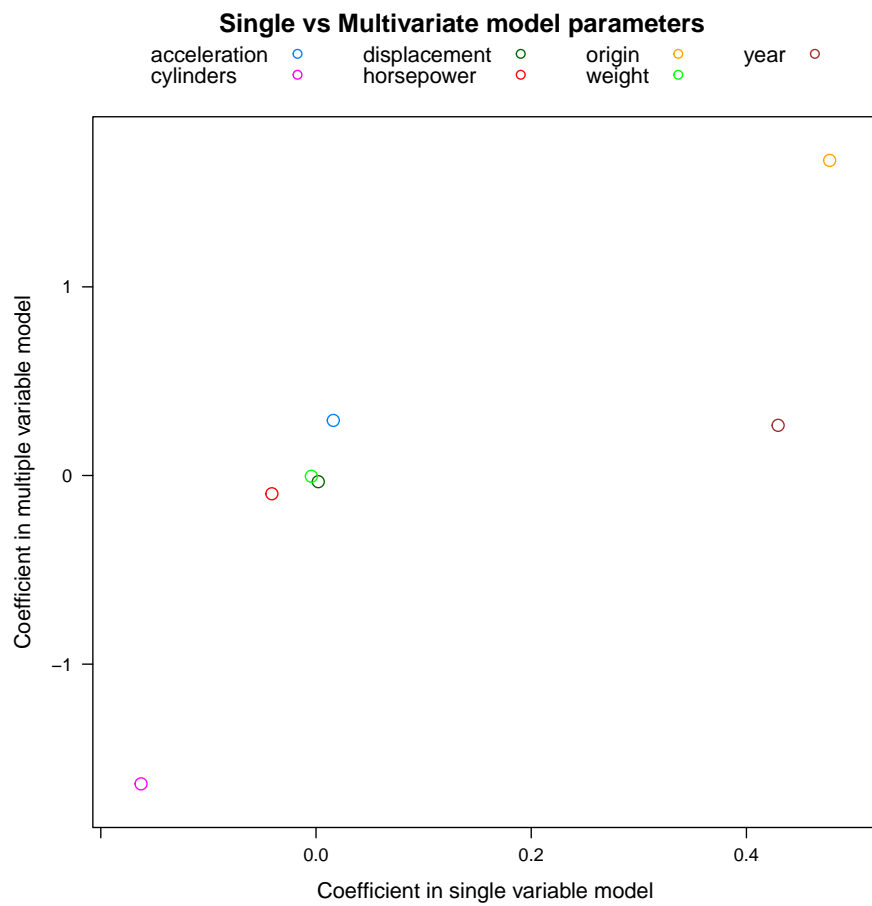


Figure 2: multi-variate comparisons

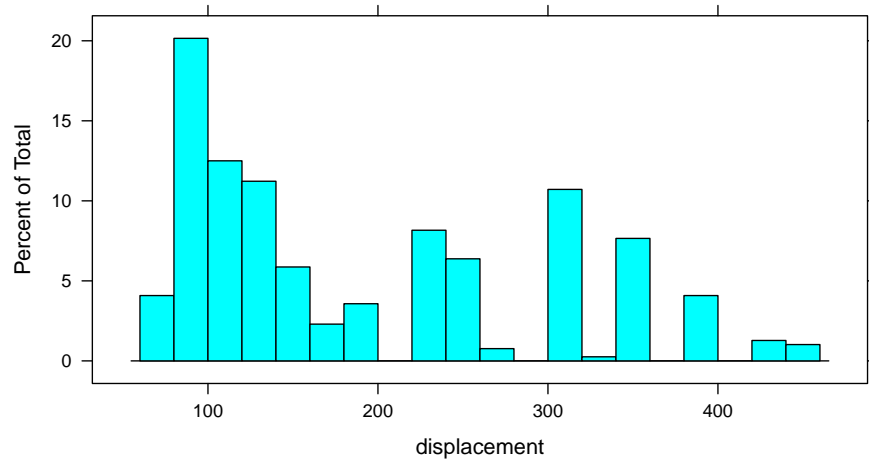


Figure 3: Histogram of the displacement variable

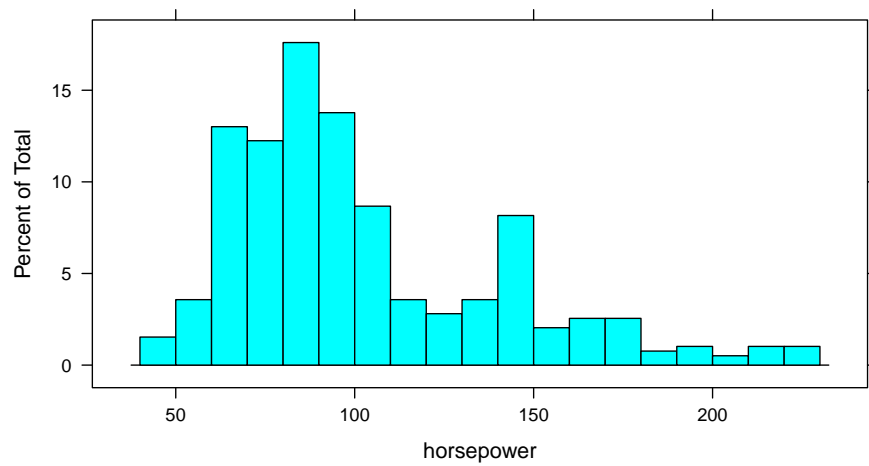


Figure 4: Histogram of the horsepower variable

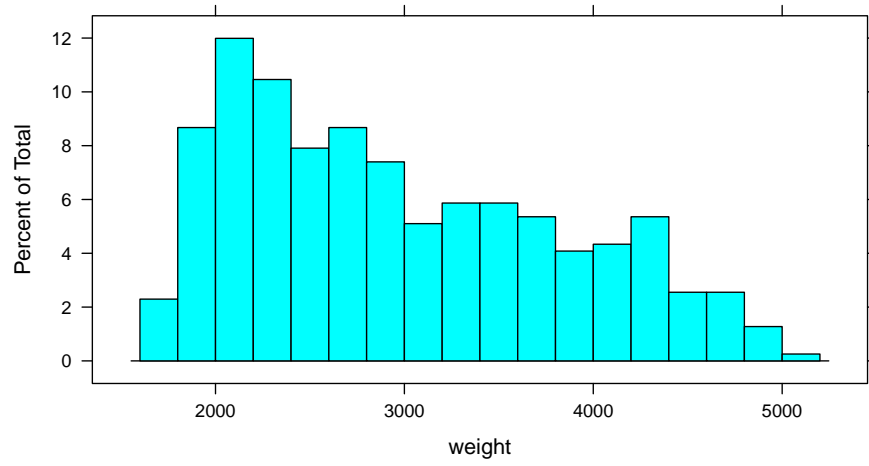


Figure 5: Histogram of the weight variable

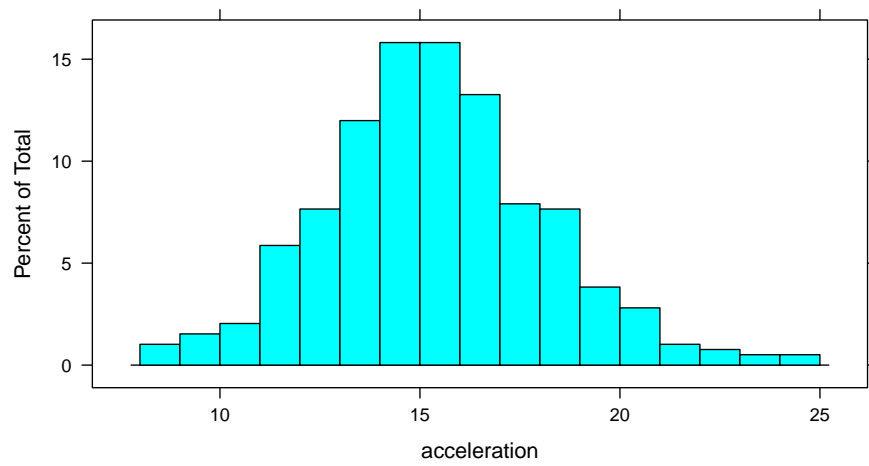


Figure 6: Histogram of the acceleration variable

```

##      3      3      1
##      4     20    179
##      5      1      2
##      6     72     11
##      8    100      3
## [1] "year"
##
##          0      1
##      70    22      7
##      71    16     11
##      72    20      8
##      73    34      6
##      74    11     15
##      75    19     11
##      76    20     14
##      77    15     13
##      78    21     15
##      79    13     16
##      80      1     26
##      81      3     25
##      82      1     29
## [1] "origin"
##
##          0      1
##      1    173     72
##      2     14     54
##      3      9     70
## [1] "name"
##
##                                     0 1
##      amc ambassador brougham      1 0
##      amc ambassador dpl           1 0
##      amc ambassador sst           1 0
##      amc concord                   1 1
##      amc concord d/l               1 0
##      amc concord dl 6              1 0
##      amc gremlin                   4 0
##      amc hornet                    4 0
##      amc hornet sportabout (sw)    1 0
##      amc matador                   5 0
##      amc matador (sw)              2 0
##      amc pacer                     1 0
##      amc pacer d/l                 1 0
##      amc rebel sst                 1 0
##      amc spirit dl                  0 1

```

##	audi 100 ls	0 1
##	audi 100ls	1 1
##	audi 4000	0 1
##	audi 5000	1 0
##	audi 5000s (diesel)	0 1
##	audi fox	0 1
##	bmw 2002	0 1
##	bmw 320i	1 0
##	buick century	2 0
##	buick century 350	1 0
##	buick century limited	0 1
##	buick century luxus (sw)	1 0
##	buick century special	1 0
##	buick electra 225 custom	1 0
##	buick estate wagon (sw)	2 0
##	buick lesabre custom	1 0
##	buick opel isuzu deluxe	0 1
##	buick regal sport coupe (turbo)	1 0
##	buick skyhawk	1 0
##	buick skylark	1 1
##	buick skylark 320	1 0
##	buick skylark limited	0 1
##	cadillac eldorado	0 1
##	cadillac seville	1 0
##	capri ii	0 1
##	chevroelt chevelle malibu	1 0
##	chevrolet bel air	1 0
##	chevrolet camaro	0 1
##	chevrolet caprice classic	3 0
##	chevrolet cavalier	0 1
##	chevrolet cavalier 2-door	0 1
##	chevrolet cavalier wagon	0 1
##	chevrolet chevelle concours (sw)	1 0
##	chevrolet chevelle malibu	2 0
##	chevrolet chevelle malibu classic	2 0
##	chevrolet chevette	0 4
##	chevrolet citation	0 3
##	chevrolet concours	1 0
##	chevrolet impala	4 0
##	chevrolet malibu	2 0
##	chevrolet malibu classic (sw)	1 0
##	chevrolet monte carlo	1 0
##	chevrolet monte carlo landau	2 0
##	chevrolet monte carlo s	1 0
##	chevrolet monza 2+2	1 0

##	chevrolet nova	3 0
##	chevrolet nova custom	1 0
##	chevrolet vega	2 1
##	chevrolet vega (sw)	1 0
##	chevrolet vega 2300	0 1
##	chevrolet woody	0 1
##	chevy c10	1 0
##	chevy c20	1 0
##	chevy s-10	0 1
##	chrysler cordoba	1 0
##	chrysler lebaron medallion	0 1
##	chrysler lebaron salon	1 0
##	chrysler lebaron town @ country (sw)	1 0
##	chrysler new yorker brougham	1 0
##	chrysler newport royal	1 0
##	datsum 1200	0 1
##	datsum 200-sx	0 1
##	datsum 200sx	0 1
##	datsum 210	0 2
##	datsum 210 mpg	0 1
##	datsum 280-zx	0 1
##	datsum 310	0 1
##	datsum 310 gx	0 1
##	datsum 510	0 1
##	datsum 510 (sw)	0 1
##	datsum 510 hatchback	0 1
##	datsum 610	1 0
##	datsum 710	0 2
##	datsum 810	1 0
##	datsum 810 maxima	0 1
##	datsum b-210	0 1
##	datsum b210	0 1
##	datsum b210 gx	0 1
##	datsum f-10 hatchback	0 1
##	datsum pl510	0 2
##	dodge aries se	0 1
##	dodge aries wagon (sw)	0 1
##	dodge aspen	2 0
##	dodge aspen 6	1 0
##	dodge aspen se	1 0
##	dodge challenger se	1 0
##	dodge charger 2.2	0 1
##	dodge colt	0 3
##	dodge colt (sw)	0 1
##	dodge colt hardtop	0 1

##	dodge colt hatchback custom	0 1
##	dodge colt m/m	0 1
##	dodge coronet brougham	1 0
##	dodge coronet custom	1 0
##	dodge coronet custom (sw)	1 0
##	dodge d100	1 0
##	dodge d200	1 0
##	dodge dart custom	1 0
##	dodge diplomat	1 0
##	dodge magnum xe	1 0
##	dodge monaco (sw)	1 0
##	dodge monaco brougham	1 0
##	dodge omni	0 1
##	dodge rampage	0 1
##	dodge st. regis	1 0
##	fiat 124 sport coupe	0 1
##	fiat 124 tc	0 1
##	fiat 124b	0 1
##	fiat 128	0 2
##	fiat 131	0 1
##	fiat strada custom	0 1
##	fiat x1.9	0 1
##	ford country	1 0
##	ford country squire (sw)	2 0
##	ford escort 2h	0 1
##	ford escort 4w	0 1
##	ford f108	1 0
##	ford f250	1 0
##	ford fairmont	0 1
##	ford fairmont (auto)	1 0
##	ford fairmont (man)	0 1
##	ford fairmont 4	1 0
##	ford fairmont futura	0 1
##	ford fiesta	0 1
##	ford futura	1 0
##	ford galaxie 500	3 0
##	ford gran torino	3 0
##	ford gran torino (sw)	2 0
##	ford granada	1 0
##	ford granada ghia	1 0
##	ford granada gl	1 0
##	ford granada l	1 0
##	ford ltd	2 0
##	ford ltd landau	1 0
##	ford maverick	3 1

##	ford mustang	1 0
##	ford mustang gl	0 1
##	ford mustang ii	1 0
##	ford mustang ii 2+2	0 1
##	ford pinto	2 3
##	ford pinto (sw)	1 0
##	ford pinto runabout	1 0
##	ford ranger	0 1
##	ford thunderbird	1 0
##	ford torino	1 0
##	ford torino 500	1 0
##	hi 1200d	1 0
##	honda accord	0 2
##	honda accord cvcc	0 1
##	honda accord lx	0 1
##	honda civic	0 3
##	honda civic (auto)	0 1
##	honda civic 1300	0 1
##	honda civic 1500 gl	0 1
##	honda civic cvcc	0 2
##	honda prelude	0 1
##	maxda glc deluxe	0 1
##	maxda rx3	1 0
##	mazda 626	0 2
##	mazda glc	0 1
##	mazda glc 4	0 1
##	mazda glc custom	0 1
##	mazda glc custom l	0 1
##	mazda glc deluxe	0 1
##	mazda rx-4	1 0
##	mazda rx-7 gs	0 1
##	mazda rx2 coupe	1 0
##	mercedes-benz 240d	0 1
##	mercedes-benz 280s	1 0
##	mercedes benz 300d	0 1
##	mercury capri 2000	0 1
##	mercury capri v6	1 0
##	mercury cougar brougham	1 0
##	mercury grand marquis	1 0
##	mercury lynx l	0 1
##	mercury marquis	1 0
##	mercury marquis brougham	1 0
##	mercury monarch	1 0
##	mercury monarch ghia	1 0
##	mercury zephyr	1 0

##	mercury zephyr 6	1 0
##	nissan stanza xe	0 1
##	oldsmobile cutlass ciera (diesel)	0 1
##	oldsmobile cutlass ls	0 1
##	oldsmobile cutlass salon brougham	1 1
##	oldsmobile cutlass supreme	1 0
##	oldsmobile delta 88 royale	1 0
##	oldsmobile omega	1 0
##	oldsmobile omega brougham	0 1
##	oldsmobile starfire sx	0 1
##	oldsmobile vista cruiser	1 0
##	opel 1900	0 2
##	opel manta	0 2
##	peugeot 304	0 1
##	peugeot 504	1 3
##	peugeot 504 (sw)	1 0
##	peugeot 505s turbo diesel	0 1
##	peugeot 604sl	1 0
##	plymouth 'cuda 340	1 0
##	plymouth arrow gs	0 1
##	plymouth champ	0 1
##	plymouth cricket	0 1
##	plymouth custom suburb	1 0
##	plymouth duster	2 1
##	plymouth fury	1 0
##	plymouth fury gran sedan	1 0
##	plymouth fury iii	3 0
##	plymouth grand fury	1 0
##	plymouth horizon	0 1
##	plymouth horizon 4	0 1
##	plymouth horizon miser	0 1
##	plymouth horizon tc3	0 1
##	plymouth reliant	0 2
##	plymouth sapporo	0 1
##	plymouth satellite	1 0
##	plymouth satellite custom	1 0
##	plymouth satellite custom (sw)	1 0
##	plymouth satellite sebring	1 0
##	plymouth valiant	2 0
##	plymouth valiant custom	1 0
##	plymouth volare	1 0
##	plymouth volare custom	1 0
##	plymouth volare premier v8	1 0
##	pontiac astro	0 1
##	pontiac catalina	3 0

##	pontiac catalina brougham	1 0
##	pontiac firebird	1 0
##	pontiac grand prix	1 0
##	pontiac grand prix lj	1 0
##	pontiac j2000 se hatchback	0 1
##	pontiac lemans v6	1 0
##	pontiac phoenix	0 2
##	pontiac phoenix lj	1 0
##	pontiac safari (sw)	1 0
##	pontiac sunbird coupe	0 1
##	pontiac ventura sj	1 0
##	renault 12 (sw)	0 1
##	renault 12tl	0 1
##	renault 5 gtl	0 1
##	saab 99e	0 1
##	saab 99gle	1 0
##	saab 99le	0 2
##	subaru	0 2
##	subaru dl	0 2
##	toyota carina	1 0
##	toyota celica gt	0 1
##	toyota celica gt liftback	1 0
##	toyota corolla	0 5
##	toyota corolla 1200	0 2
##	toyota corolla 1600 (sw)	0 1
##	toyota corolla liftback	0 1
##	toyota corolla tercel	0 1
##	toyota corona	0 4
##	toyota corona hardtop	0 1
##	toyota corona liftback	0 1
##	toyota corona mark ii	0 1
##	toyota cressida	0 1
##	toyota mark ii	2 0
##	toyota starlet	0 1
##	toyota tercel	0 1
##	toyota corona mark ii (sw)	0 1
##	triumph tr7 coupe	0 1
##	volkswagen rabbit	0 1
##	volkswagen 1131 deluxe sedan	0 1
##	volkswagen 411 (sw)	1 0
##	volkswagen dasher	0 3
##	volkswagen jetta	0 1
##	volkswagen model 111	0 1
##	volkswagen rabbit	0 2
##	volkswagen rabbit custom	0 1


```
## volkswagen rabbit custom diesel      0 1
## volkswagen rabbit l                  0 1
## volkswagen scirocco                  0 1
## volkswagen super beetle              0 1
## volkswagen type 3                    0 1
## volvo 144ea                           1 0
## volvo 145e (sw)                       1 0
## volvo 244dl                           1 0
## volvo 245                             1 0
## volvo 264gl                           1 0
## volvo diesel                          0 1
## vw dasher (diesel)                   0 1
## vw pickup                             0 1
## vw rabbit                             0 2
## vw rabbit c (diesel)                  0 1
## vw rabbit custom                      0 1
```

```
mpg01 <- factor(ifelse(Auto$mpg > median(Auto$mpg),
  1, 0))
df <- Auto %>% mutate(mpg01 = mpg01, name = factor(name),
  origin = factor(origin), cylinders = factor(cylinders),
  year = factor(year)) %>% select(-mpg)
```

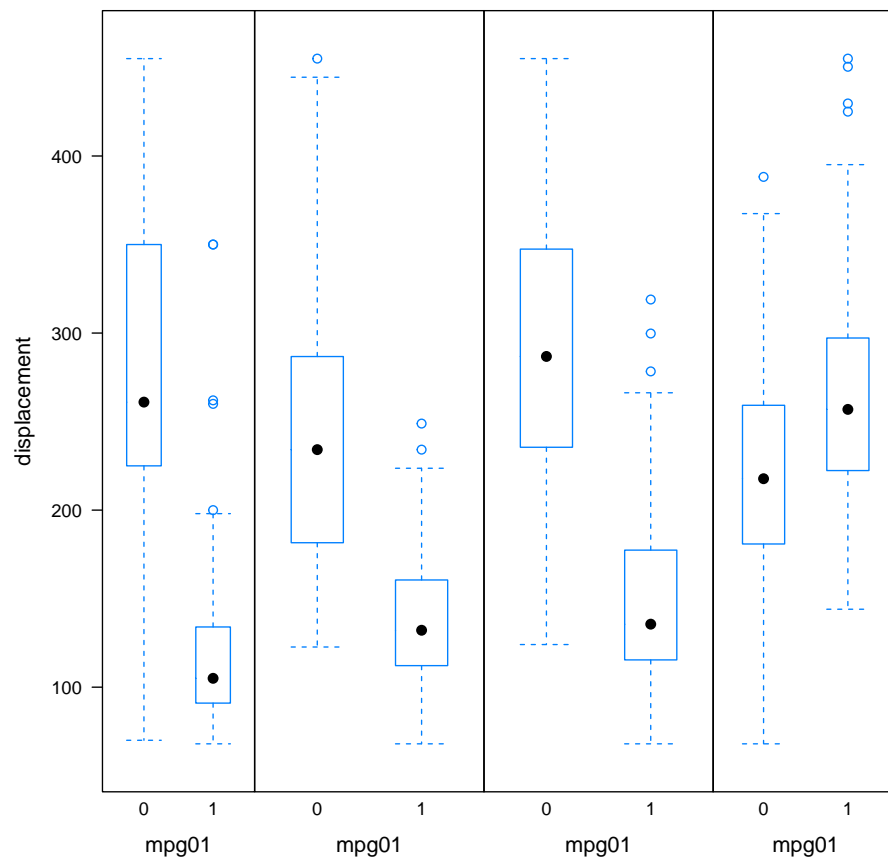


Figure 7: Boxplot of the dependent variable mpg01 by each factor variable

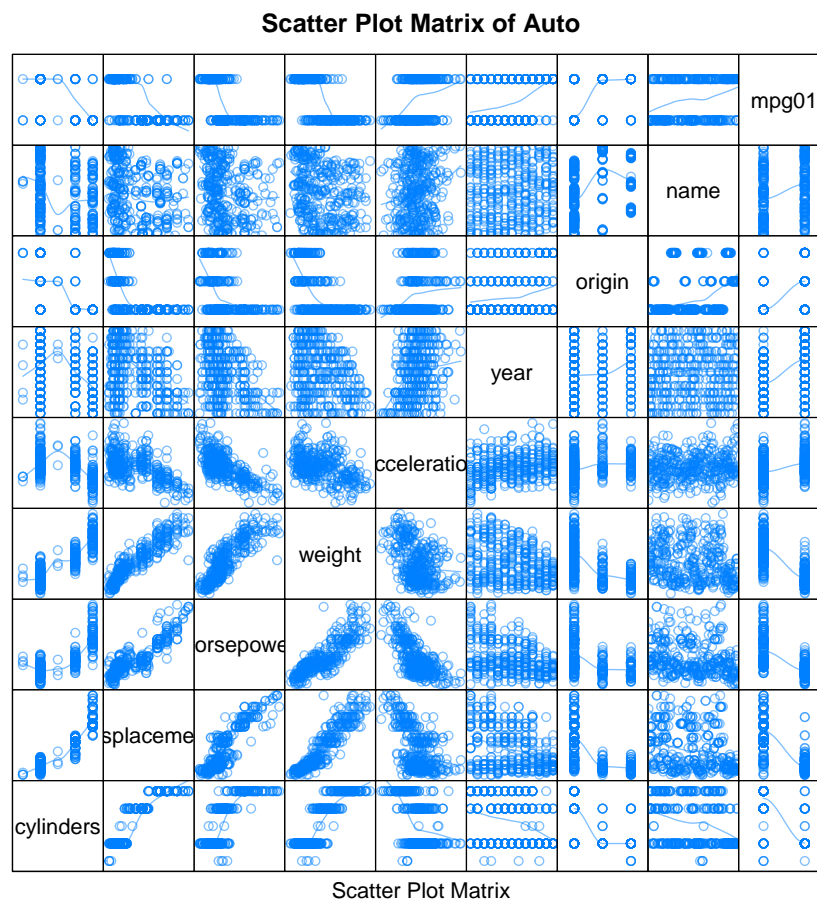


Figure 8: multi-variate comparisons