REGRESSION MODEL COURSE PROJECT

Oluwadare, Margaret

7/15/2020

EXECUTIVE SUMMARY

This report is made towards the completion of Coursera the Regression Models course on the Data Science Specialization by Johns Hopkins University.

In this project we will explore some features that affect fuel consumption in miles per gallon (MPG) answering some questions about the nature of transmission (labelled as 'am').The dataset is of a collection of cars (mtcars - Motor Trend Car Road Tests), and we are interested in exploring the relationship between a set of variables. In particularly we want answer two major questions:

• Is an automatic or manual transmission better for MPG? • Quantifying how different is the MPG between automatic and manual transmissions?

We are going to estimate the relationship between type of transmission and other independant variables, such as weight (wt), 1/4 miles/time (qsec), along with miles per gallon (MPG), which will be our outcome.

Using simple linear regression model and multiple regression model we conclude that manual transmission cars when compared against automatic transmission cars adjusted by number of cylinders, gross horspower and weight gets a factor of 1.8 more miles per gallon. This implies it goes more further.

DATA DESCRIPTION The 'mtcars' data set was extracted from the 1974 Motor Trend US magazine, which comprises of 32 observations and 11 variables. We will use regression modelling and exploratory analysis to show how transmission (am) feature affect the miles per fallon (MPG) feature. The dataset "mtcars" is located in the package "dataset". Below is a description of the variables

mpg: Miles per US gallon cyl: Number of cylinders disp: Displacement (cubic inches) hp: Gross horsepower drat: Rear axle ratio wt: Weight (lb / 1000) qsec: 1 / 4 mile time vs: V/S am: Transmission (0 = automatic, 1 = manual) gear: Number of forward gears carb: Number of carburetors

EXPLORATORY DATA ANALYSIS OF THE DATA

We load in the data set, perform the necessary data transformations and look at the descriptive of the data.

```
attach(mtcars)
View(mtcars)
str(mtcars)
```

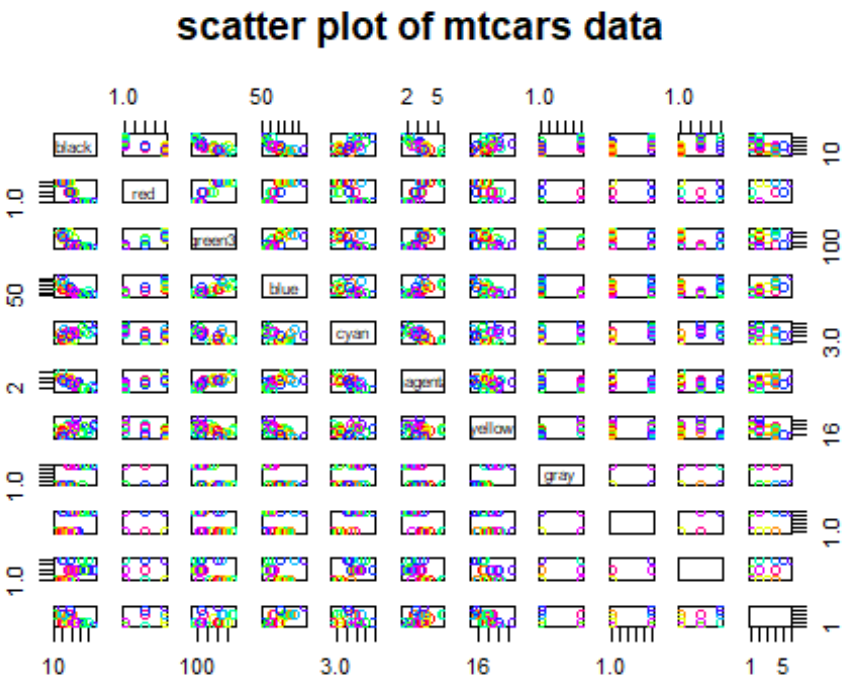
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
### CONVERT CATEGORICAL TO FACTORS
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am, labels = c('Auto','Manual')) #### assign label
values
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "Auto","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

Now that we are all set, let's explore the various relationships between variables of interest and others. As a star, we plot the relationships between all the variables of the dataset.

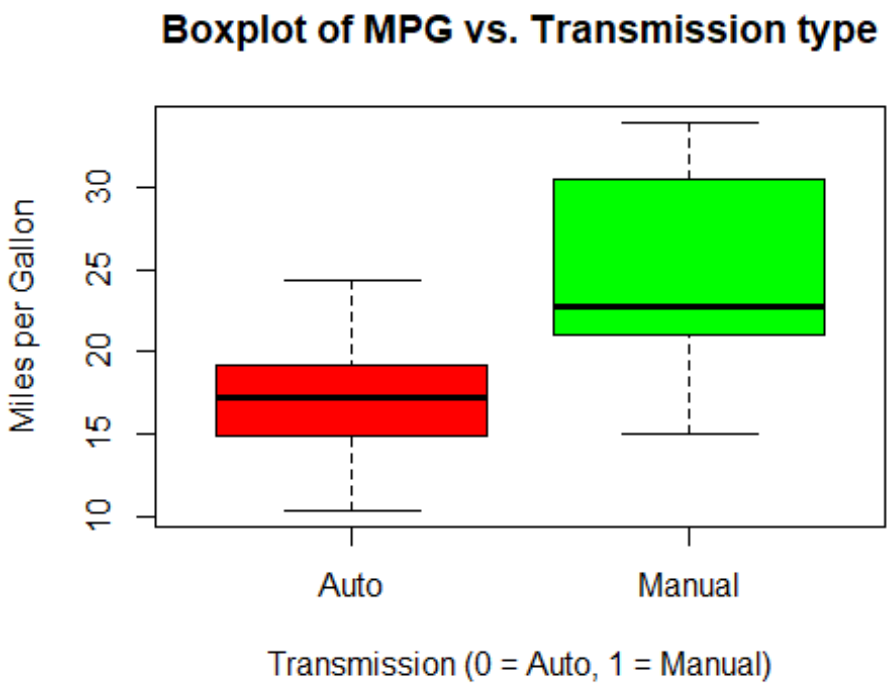
```
#Scatter plot matrix for mtcars dataset
pairs(mpg ~ ., data = mtcars, main = "scatter plot of mtcars data", col = rai
nbow(11), labels = palette())
```



From the plot, there is strong correlation between mpg and other varaibles. We will use regressional analysis investigate this relationship.

Our varaible of interest is transmission type(am) on mpg, therefore we will plot boxplots of the variable mpg on transmission (see appendix). This plot shows that mpg increases when the transmission is manual.

```
#Boxplot of MPG vs. AM
boxplot(mpg ~ am, data = mtcars, col = (c("red","green")), xlab = "Transmissi
on (0 = Auto, 1 = Manual)", ylab = "Miles per Gallon", main = "Boxplot of MPG
vs. Transmission type" )
```



REGRESSION ANALYSIS To investigate our varaible we will build linear regression models based on the variables and try to find out the best model fit and making comparrison with out main model using anova. Analysis of residuals and diagnosis will also be performed.

MODEL BUILDING AND SELECTION

Considering our pairs plot where several variables has high correlation with mpg, an initial model with all the variables as predictors will be performed first. Stepwise model selection to select significant predictors for the final model is carried out. This is taken care by the step method which runs linear model multiple times to build multiple regression models and select the best variables from them using both forward selection and backward elimination methods by the AIC algorithm. The code is given below.

```
linmod <- lm(mpg ~ ., data = mtcars) #regressing mpg with other features
summary(linmod)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.87913   20.06582   1.190   0.2525
```

```
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## amManual     1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb2       -0.97935    2.31797  -0.423   0.6787
## carb3        2.99964    4.29355   0.699   0.4955
## carb4        1.09142    4.44962   0.245   0.8096
## carb6        4.47757    6.38406   0.701   0.4938
## carb8        7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

```
bestmod <- step(linmod, direction = "both") ##selecting the best model
```

```
## Start:  AIC=76.4
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - carb  5   13.5989 134.00 69.828
## - gear  2    3.9729 124.38 73.442
## - am    1    1.1420 121.55 74.705
## - qsec  1    1.2413 121.64 74.732
## - drat  1    1.8208 122.22 74.884
## - cyl   2   10.9314 131.33 75.184
## - vs    1    3.6299 124.03 75.354
## <none>               120.40 76.403
## - disp  1    9.9672 130.37 76.948
## - wt    1   25.5541 145.96 80.562
## - hp    1   25.6715 146.07 80.588
##
## Step:  AIC=69.83
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##        Df Sum of Sq    RSS    AIC
## - gear  2    5.0215 139.02 67.005
## - disp  1    0.9934 135.00 68.064
## - drat  1    1.1854 135.19 68.110
## - vs    1    3.6763 137.68 68.694
## - cyl   2   12.5642 146.57 68.696
```

```
## - qsec  1    5.2634 139.26 69.061
## <none>              134.00 69.828
## - am    1   11.9255 145.93 70.556
## - wt    1   19.7963 153.80 72.237
## - hp    1   22.7935 156.79 72.855
## + carb  5   13.5989 120.40 76.403
##
## Step:  AIC=67
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - drat  1    0.9672 139.99 65.227
## - cyl   2   10.4247 149.45 65.319
## - disp  1    1.5483 140.57 65.359
## - vs    1    2.1829 141.21 65.503
## - qsec  1    3.6324 142.66 65.830
## <none>              139.02 67.005
## - am    1   16.5665 155.59 68.608
## - hp    1   18.1768 157.20 68.937
## + gear  2    5.0215 134.00 69.828
## - wt    1   31.1896 170.21 71.482
## + carb  5   14.6475 124.38 73.442
##
## Step:  AIC=65.23
## mpg ~ cyl + disp + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - disp  1    1.2474 141.24 63.511
## - vs    1    2.3403 142.33 63.757
## - cyl   2   12.3267 152.32 63.927
## - qsec  1    3.1000 143.09 63.928
## <none>              139.99 65.227
## + drat  1    0.9672 139.02 67.005
## - hp    1   17.7382 157.73 67.044
## - am    1   19.4660 159.46 67.393
## + gear  2    4.8033 135.19 68.110
## - wt    1   30.7151 170.71 69.574
## + carb  5   13.0509 126.94 72.095
##
## Step:  AIC=63.51
## mpg ~ cyl + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - qsec  1     2.442 143.68 62.059
## - vs    1     2.744 143.98 62.126
## - cyl   2    18.580 159.82 63.466
## <none>              141.24 63.511
## + disp  1     1.247 139.99 65.227
## + drat  1     0.666 140.57 65.359
## - hp    1    18.184 159.42 65.386
```

```
## - am    1    18.885 160.12 65.527
## + gear  2     4.684 136.55 66.431
## - wt    1    39.645 180.88 69.428
## + carb  5     2.331 138.91 72.978
##
## Step:  AIC=62.06
## mpg ~ cyl + hp + wt + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - vs    1     7.346 151.03 61.655
## <none>              143.68 62.059
## - cyl   2    25.284 168.96 63.246
## + qsec  1     2.442 141.24 63.511
## - am    1    16.443 160.12 63.527
## + disp  1     0.589 143.09 63.928
## + drat  1     0.330 143.35 63.986
## + gear  2     3.437 140.24 65.284
## - hp    1    36.344 180.02 67.275
## - wt    1    41.088 184.77 68.108
## + carb  5     3.480 140.20 71.275
##
## Step:  AIC=61.65
## mpg ~ cyl + hp + wt + am
##
##        Df Sum of Sq    RSS    AIC
## <none>              151.03 61.655
## - am    1     9.752 160.78 61.657
## + vs    1     7.346 143.68 62.059
## + qsec  1     7.044 143.98 62.126
## - cyl   2    29.265 180.29 63.323
## + disp  1     0.617 150.41 63.524
## + drat  1     0.220 150.81 63.608
## + gear  2     1.361 149.66 65.365
## - hp    1    31.943 182.97 65.794
## - wt    1    46.173 197.20 68.191
## + carb  5     5.633 145.39 70.438

bestmod

##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Coefficients:
## (Intercept)         cyl6         cyl8           hp           wt     amManu
al
##    33.70832     -3.03134     -2.16368     -0.03211     -2.49683      1.809
21
```

The best model obtained from the above computations consists of the variables, cyl(with respect to vehicles with 6 and 8 cylinders), wt and hp as confounders and am as the independent variable. Details of the model are in the summary(bestmod) code below. We observe that the Adjusted R^2 value is 0.84. Therefore we can conclude that more than 84% of the variability is explained by the last model in 'bestmod'.

```
summary(bestmod)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

With the above result, we'll perform anova to compare aganist our initial model which will uses am as a predictor variable only, and the best model that was found through performing stepwise selection.

```
#Anova
initmodel <- lm(mpg ~ am, data = mtcars)
initmodel
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Coefficients:
## (Intercept)     amManual
##      17.147        7.245
```

```
anova(initmodel, bestmod)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is significant we will conclude that the variables cyl, hp and wt do contribute to the accuracy of the model.

INFERENCE With the result above we perform a t-test on normality assumption for transmission (am) and from the result, we see that the manual and automatic transmissions are significantly different.

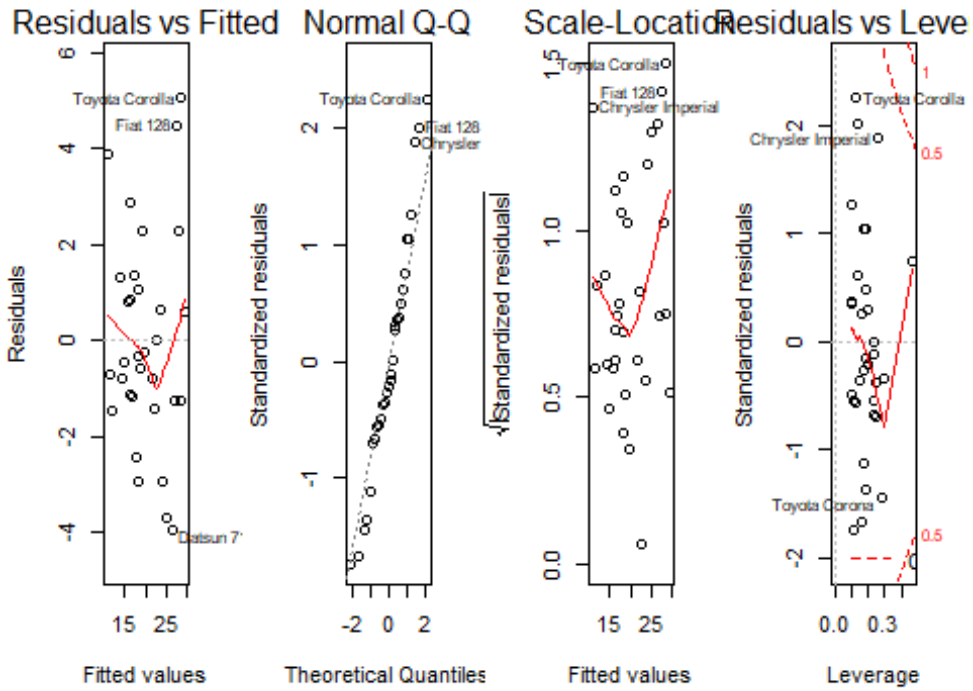
```
t.test(mpg ~ am, data = mtcars)

##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
##   mean in group Auto mean in group Manual
##             17.14737             24.39231
```

RESIDUALS AND DIAGNOSIS RESULTS We examine residual plots of our regression model and perform diagnostics to uncover outliers in the data set.The following observations can be inferred from our results:

1. Outliers are observed in the top right corners of the plot.
2. The Scale-Location plot points are scattered in a constant band pattern, implying constant variance.
3. The points in the residuals vs. fitted plot are randomly scattered on the plot verifying the independence condition.
4. The Normal Q-Q plot consists of the points falling on the line indicating normality of residuals.

With the above observation, we compute some regression diagnostics of our model to find out the leverage points as depicted below. We compute top five points in each case of influence measures.From the result, we notice that our analysis was correct, as the same cars are mentioned in the residual plots

```
par(mfrow = c(1, 4))
plot(bestmod)
```



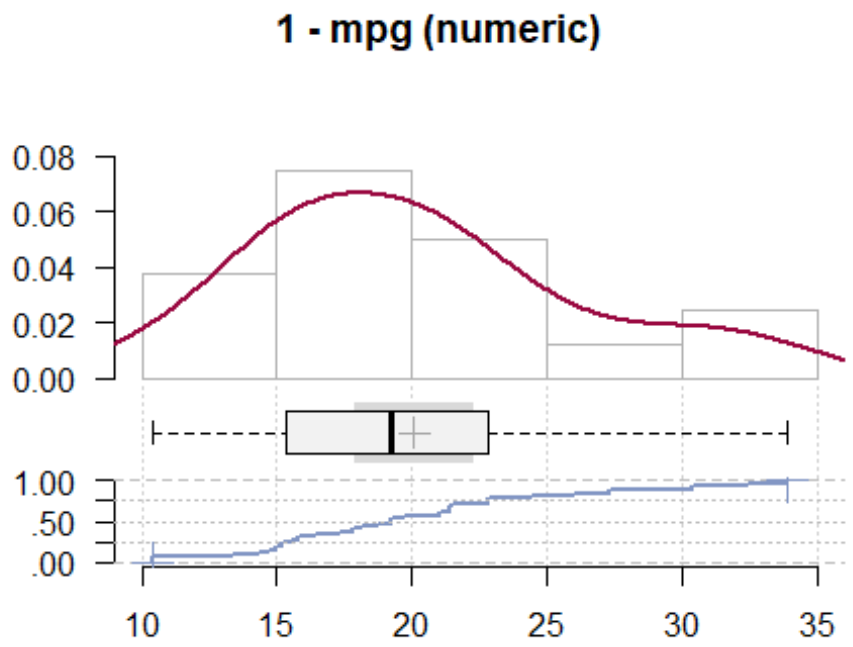

```
leverage <- hatvalues(bestmod)
tail(sort(leverage),5)

##       Mazda RX4 Wag   Chrysler Imperial       Toyota Corona Lincoln Contin
ental
##           0.2496110           0.2611168           0.2777872           0.29
36819
##       Maserati Bora
##           0.4713671

influential <- dfbetas(bestmod)
tail(sort(influential[,6]),5)

##        Camaro Z28    Toyota Corolla Chrysler Imperial          Fiat 128
##         0.08398495        0.28853987        0.35074579        0.42920432
##     Toyota Corona
##         0.73054020
```

CONCLUSION Based on the observations from our best fit model, we can conclude the following

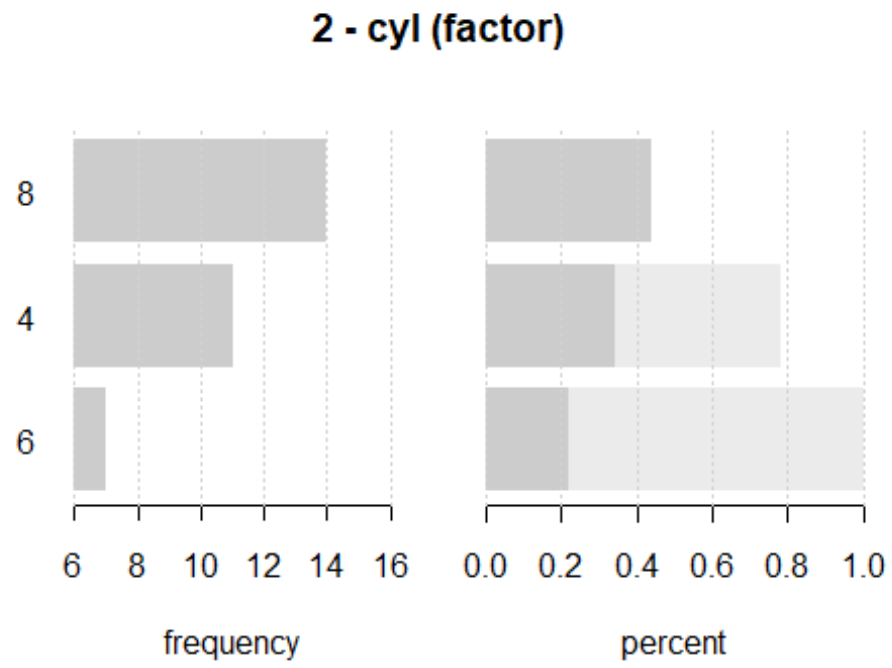
1. Cars with Manual transmission get more miles per gallon compared aganist cars with Automatic transmission. (1.8 adjusted by hp, cyl, and wt).
2. mpg will decrease by 2.5 for every 1000 lb increase in wt. mpg decreases negligibly with increase of hp.
3. mpg decrease by a factor of 3 and 2.2 respectively (adjusted by hp, wt, and am) with increased number of cylinders (cyl)from 4 to 6 and 8.

You can also embed plots, for example:

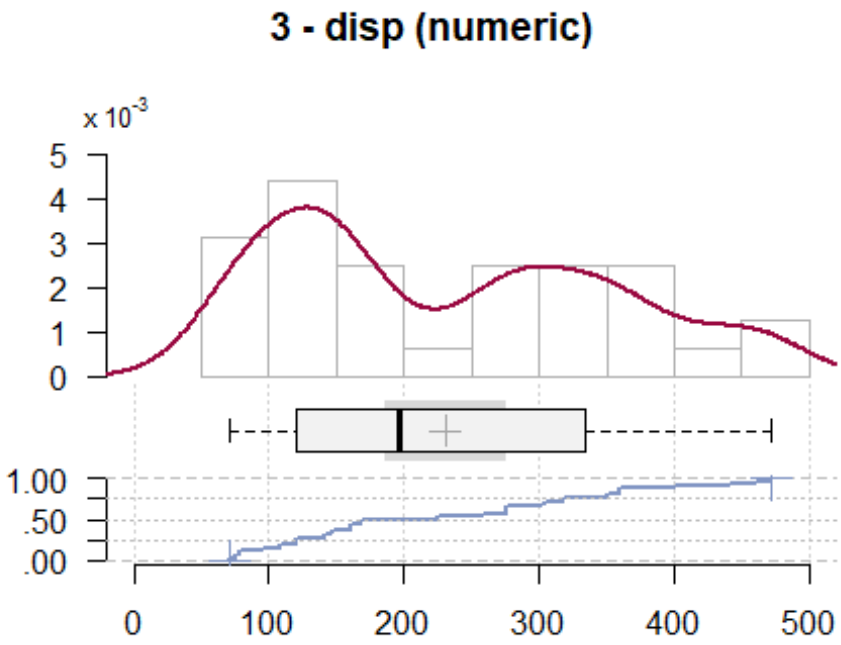
```
## Warning: package 'DescTools' was built under R version 3.6.3

## Warning in GetCOMAppHandle("Word.Application", option = "lastWord",
## existing = FALSE, : RDCOMClient is not available. To install it use:
## install.packages('RDCOMClient', repos = 'http://www.stats.ox.ac.uk/pub/RWi
n/')

## ------------------------------------------------------------------------
----
## Describe mtcars (data.frame):
##
## data frame:  32 obs. of  11 variables
##      32 complete cases (100.0%)
##
##   Nr  ColName  Class    NAs  Levels
##   1   mpg      numeric  .
##   2   cyl      factor   .    (3): 1-4, 2-6, 3-8
##   3   disp     numeric  .
##   4   hp       numeric  .
##   5   drat     numeric  .
##   6   wt       numeric  .
##   7   qsec     numeric  .
##   8   vs       factor   .    (2): 1-0, 2-1
##   9   am       factor   .    (2): 1-Auto, 2-Manual
##   10  gear     factor   .    (3): 1-3, 2-4, 3-5
##   11  carb     factor   .    (6): 1-1, 2-2, 3-3, 4-4, 5-6, ...
##
##
## ------------------------------------------------------------------------
----
## 1 - mpg (numeric)
##
##   length       n     NAs  unique      0s    mean  meanCI'
##       32      32       0      25       0  20.091  17.918
##           100.0%    0.0%            0.0%          22.264
##
##      .05     .10     .25  median     .75     .90     .95
##   11.995  14.340  15.425  19.200  22.800  30.090  31.300
##
##    range      sd   vcoef     mad     IQR    skew    kurt
##   23.500   6.027   0.300   5.411   7.375   0.611  -0.373
##
## lowest : 10.4 (2), 13.3, 14.3, 14.7, 15.0
## highest: 26.0, 27.3, 30.4 (2), 32.4, 33.9
##
## ' 95%-CI (classic)
```



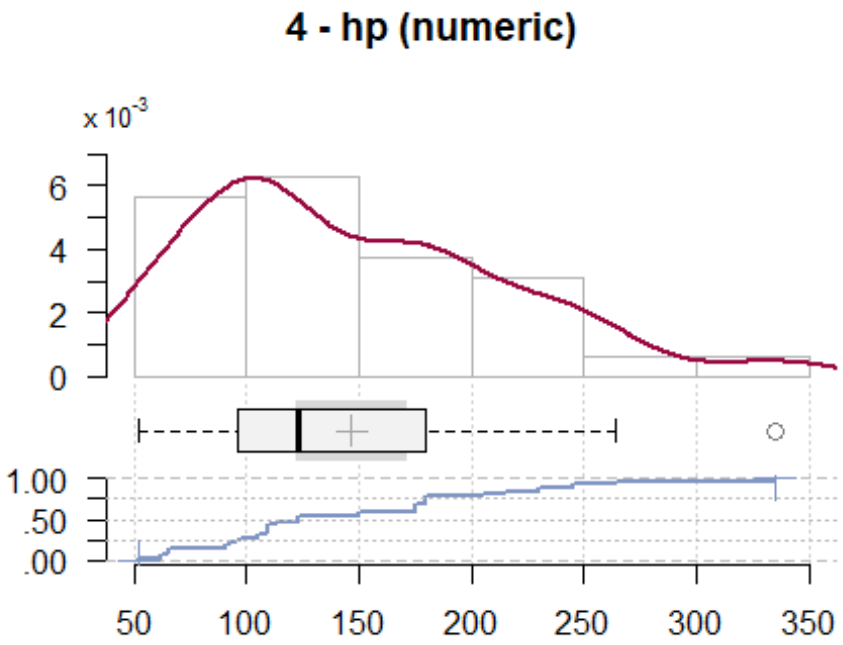

```
## ------------------------------------------------------------------------
----
## 2 - cyl (factor)
##
##   length      n    NAs unique levels  dupes
##       32     32      0      3      3      y
##          100.0%   0.0%
##
##    level  freq   perc  cumfreq  cumperc
## 1      8    14  43.8%       14    43.8%
## 2      4    11  34.4%       25    78.1%
## 3      6     7  21.9%       32   100.0%
```



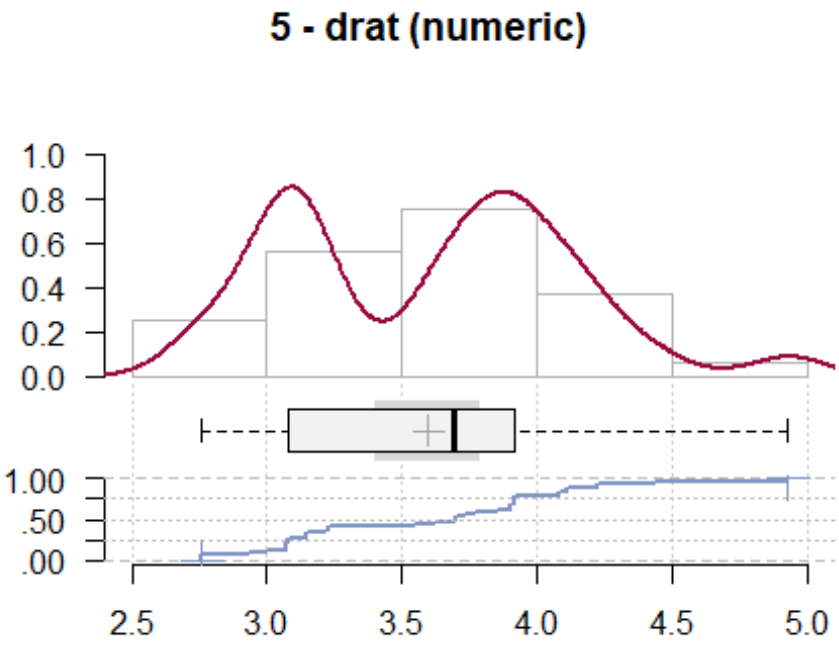

```
## ------------------------------------------------------------------------
----
## 3 - disp (numeric)
##
##    length        n      NAs   unique       0s     mean   meanCI'
##        32       32        0       27        0  230.722  186.037
##             100.0%     0.0%              0.0%           275.407
##
##       .05      .10      .25   median      .75      .90      .95
##    77.350   80.610  120.825  196.300  326.000  396.000  449.000
##
##     range       sd    vcoef      mad      IQR     skew     kurt
##   400.900  123.939    0.537  140.476  205.175    0.382   -1.207
##
## lowest : 71.1, 75.7, 78.7, 79.0, 95.1
## highest: 360.0 (2), 400.0, 440.0, 460.0, 472.0
##
## ' 95%-CI (classic)
```



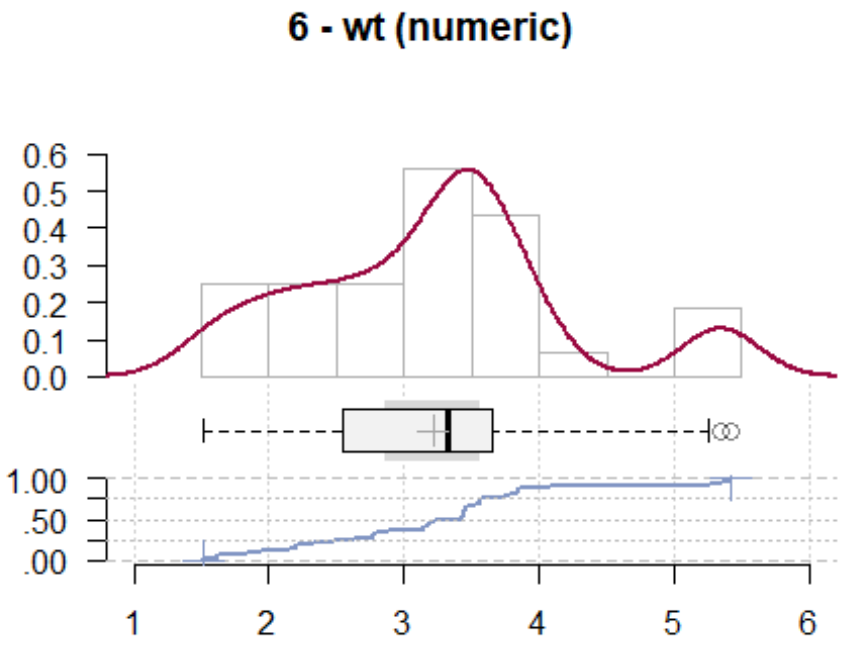

```
## ------------------------------------------------------------------------
----
## 4 - hp (numeric)
##
##   length       n    NAs  unique      0s    mean  meanCI'
##       32      32      0      22       0  146.69  121.97
##           100.0%   0.0%            0.0%          171.41
##
##      .05     .10    .25  median     .75     .90     .95
##    63.65   66.00  96.50  123.00  180.00  243.50  253.55
##
##    range      sd  vcoef     mad     IQR    skew    kurt
##   283.00   68.56   0.47   77.10   83.50    0.73   -0.14
##
## lowest : 52.0, 62.0, 65.0, 66.0 (2), 91.0
## highest: 215.0, 230.0, 245.0 (2), 264.0, 335.0
##
## ' 95%-CI (classic)
```



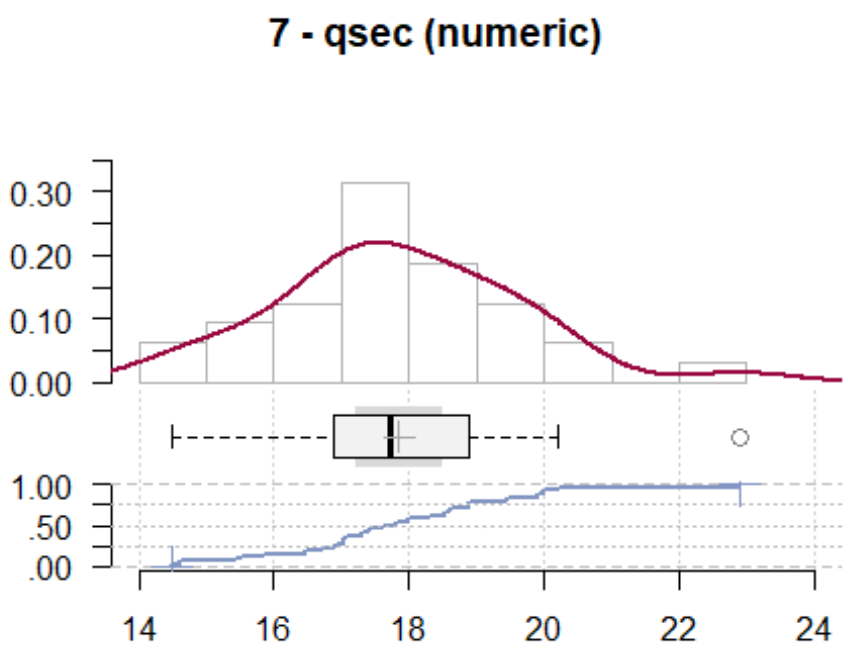

```
## ------------------------------------------------------------------------
----
## 5 - drat (numeric)
##
##   length       n     NAs  unique      0s    mean   meanCI'
##       32      32       0      22       0  3.5966   3.4038
##           100.0%    0.0%            0.0%           3.7893
##
##      .05     .10     .25  median     .75     .90      .95
##   2.8535  3.0070  3.0800  3.6950  3.9200  4.2090   4.3145
##
##    range      sd   vcoef     mad     IQR    skew     kurt
##   2.1700  0.5347  0.1487  0.7042  0.8400  0.2659  -0.7147
##
## lowest : 2.76 (2), 2.93, 3.0, 3.07 (3), 3.08 (2)
## highest: 4.08 (2), 4.11, 4.22 (2), 4.43, 4.93
##
## ' 95%-CI (classic)
```



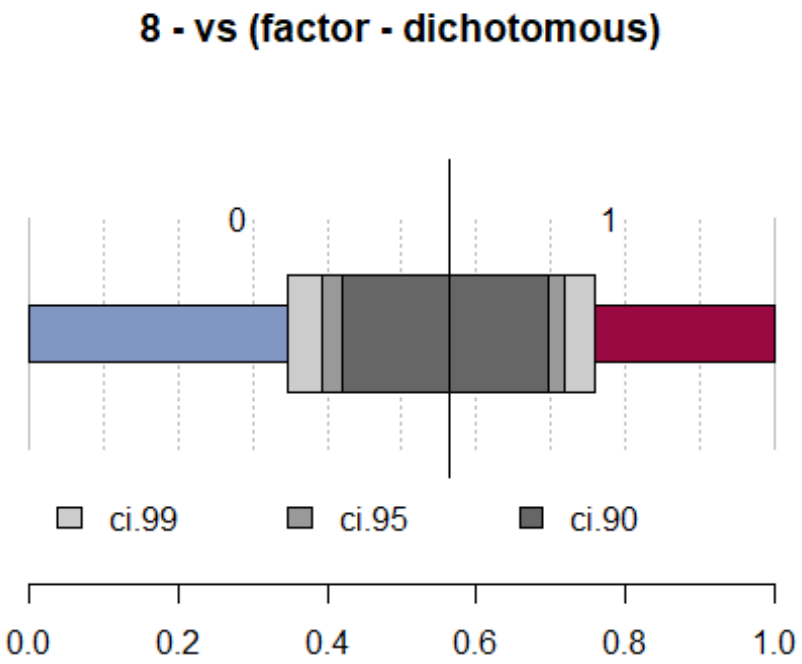

```
## ------------------------------------------------------------------------
----
## 6 - wt (numeric)
##
##    length        n      NAs   unique       0s     mean    meanCI'
##        32       32        0       29        0  3.21725   2.86448
##             100.0%     0.0%              0.0%            3.57002
##
##       .05      .10      .25   median      .75      .90       .95
##   1.73600  1.95550  2.58125  3.32500  3.61000  4.04750   5.29275
##
##     range       sd    vcoef      mad      IQR     skew      kurt
##   3.91100  0.97846  0.30413  0.76725  1.02875  0.42315  -0.02271
##
## lowest : 1.513, 1.615, 1.835, 1.935, 2.14
## highest: 3.845, 4.07, 5.25, 5.345, 5.424
##
## ' 95%-CI (classic)
```



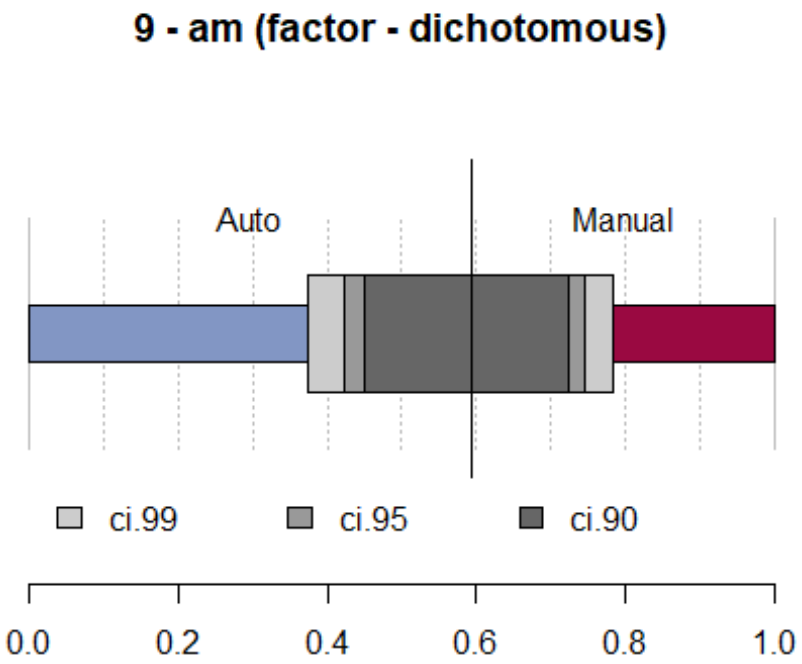

```
## ------------------------------------------------------------------------
----
## 7 - qsec (numeric)
##
##    length        n      NAs   unique       0s     mean   meanCI'
##        32       32        0       30        0  17.8488  17.2045
##             100.0%     0.0%              0.0%           18.4930
##
##       .05      .10      .25   median      .75      .90      .95
##   15.0455  15.5340  16.8925  17.7100  18.9000  19.9900  20.1045
##
##     range       sd    vcoef      mad      IQR     skew     kurt
##    8.4000   1.7869   0.1001   1.4159   2.0075   0.3690   0.3351
##
## lowest : 14.5, 14.6, 15.41, 15.5, 15.84
## highest: 19.9, 20.0, 20.01, 20.22, 22.9
##
## ' 95%-CI (classic)
```



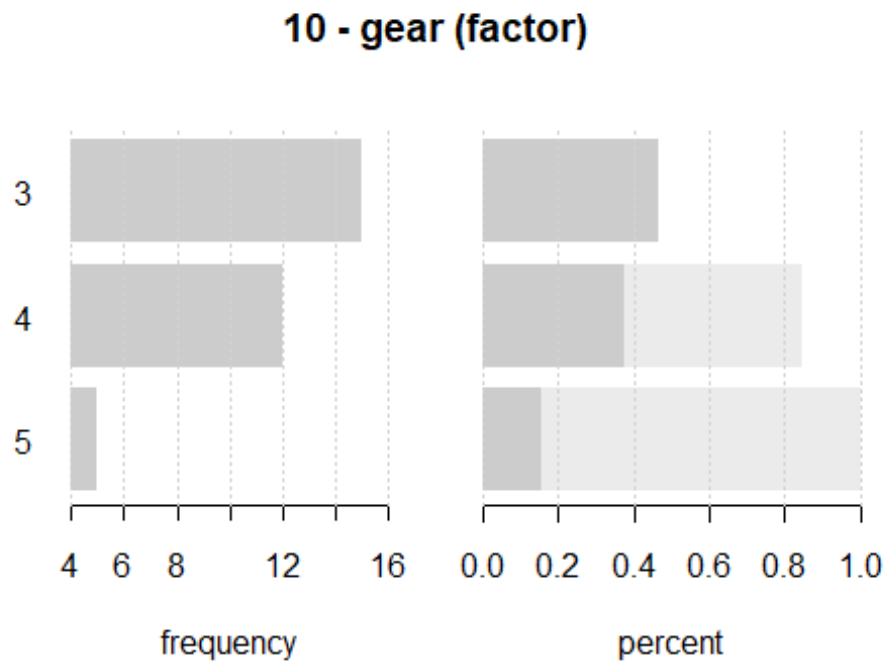

```
## ------------------------------------------------------------------------
----
## 8 - vs (factor - dichotomous)
##
##   length      n    NAs unique
##       32     32      0      2
##          100.0%   0.0%
##
##    freq   perc  lci.95  uci.95'
## 0    18  56.2%   39.3%   71.8%
## 1    14  43.8%   28.2%   60.7%
##
## ' 95%-CI (Wilson)
```



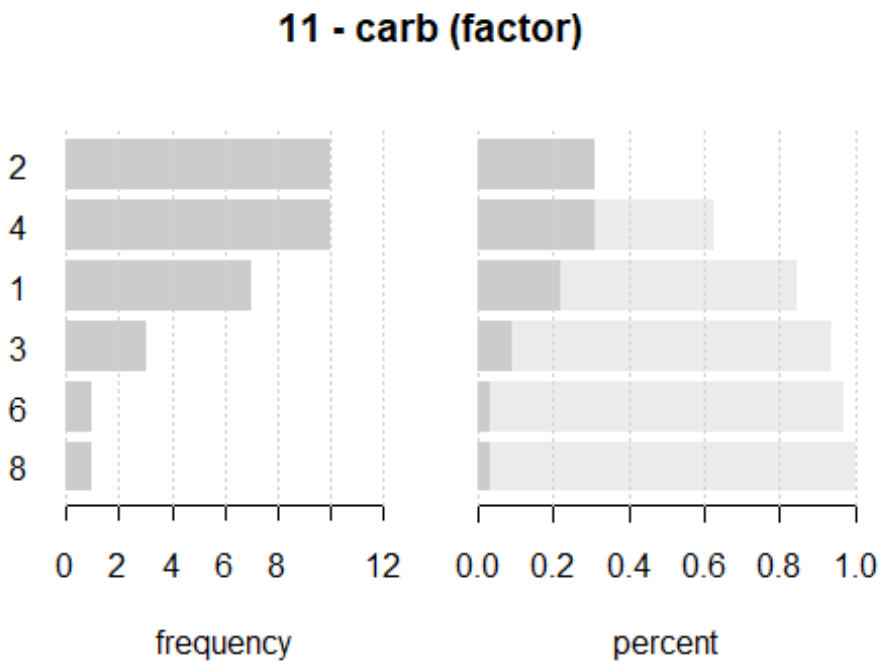

```
## ------------------------------------------------------------------------
----
## 9 - am (factor - dichotomous)
##
##   length      n    NAs unique
##       32     32      0      2
##          100.0%   0.0%
##
##         freq   perc  lci.95  uci.95'
## Auto      19  59.4%   42.3%   74.5%
## Manual    13  40.6%   25.5%   57.7%
##
## ' 95%-CI (Wilson)
```

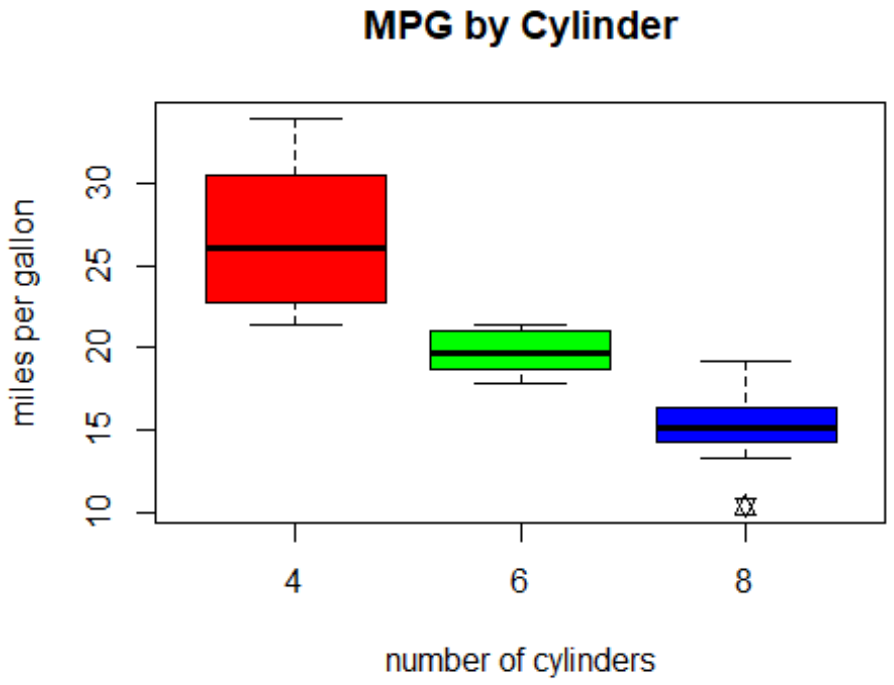



```
## ------------------------------------------------------------------------
----
## 10 - gear (factor)
##
##   length      n    NAs unique levels  dupes
##       32     32      0      3      3      y
##          100.0%   0.0%
##
##    level  freq   perc  cumfreq  cumperc
## 1      3    15  46.9%       15    46.9%
## 2      4    12  37.5%       27    84.4%
## 3      5     5  15.6%       32   100.0%
```




```
## ------------------------------------------------------------------------
----
## 11 - carb (factor)
##
##   length      n    NAs unique levels  dupes
##       32     32      0      6      6      y
##          100.0%   0.0%
##
##    level  freq   perc  cumfreq  cumperc
## 1      2    10  31.2%       10    31.2%
## 2      4    10  31.2%       20    62.5%
## 3      1     7  21.9%       27    84.4%
## 4      3     3   9.4%       30    93.8%
## 5      6     1   3.1%       31    96.9%
## 6      8     1   3.1%       32   100.0%
```


11 - carb (factor)
2
4
1
3
6
8
0 2 4 6 8 12
frequency
0.0 0.2 0.4 0.6 0.8 1.0
percent


MPG by Cylinder
miles per gallon
10 15 20 25 30
4 6 8
number of cylinders