

# Análise de Correlação

*Max Pereira*

*24/04/2020*

## Teste de Correlação

O teste de correlação é usado para avaliar a associação entre duas ou mais variáveis

### Correlação de Pearson

O coeficiente de correlação de Pearson ( $r$ ), também chamado de correlação linear ou  $r$  de Pearson, é um grau de relação entre duas variáveis quantitativas e mostra o grau de correlação através de valores situados entre -1 e 1.

### Valor-p (nível de significância)

se o valor-p é menor que 5%, então a correlação entre  $x$  e  $y$  é significativa.

### Conjunto de Dados (Dataset mtcars)

```
dados <- mtcars
head(dados)
```

##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Pretende-se verificar a correlação entre as variáveis “mpg” e “wt”

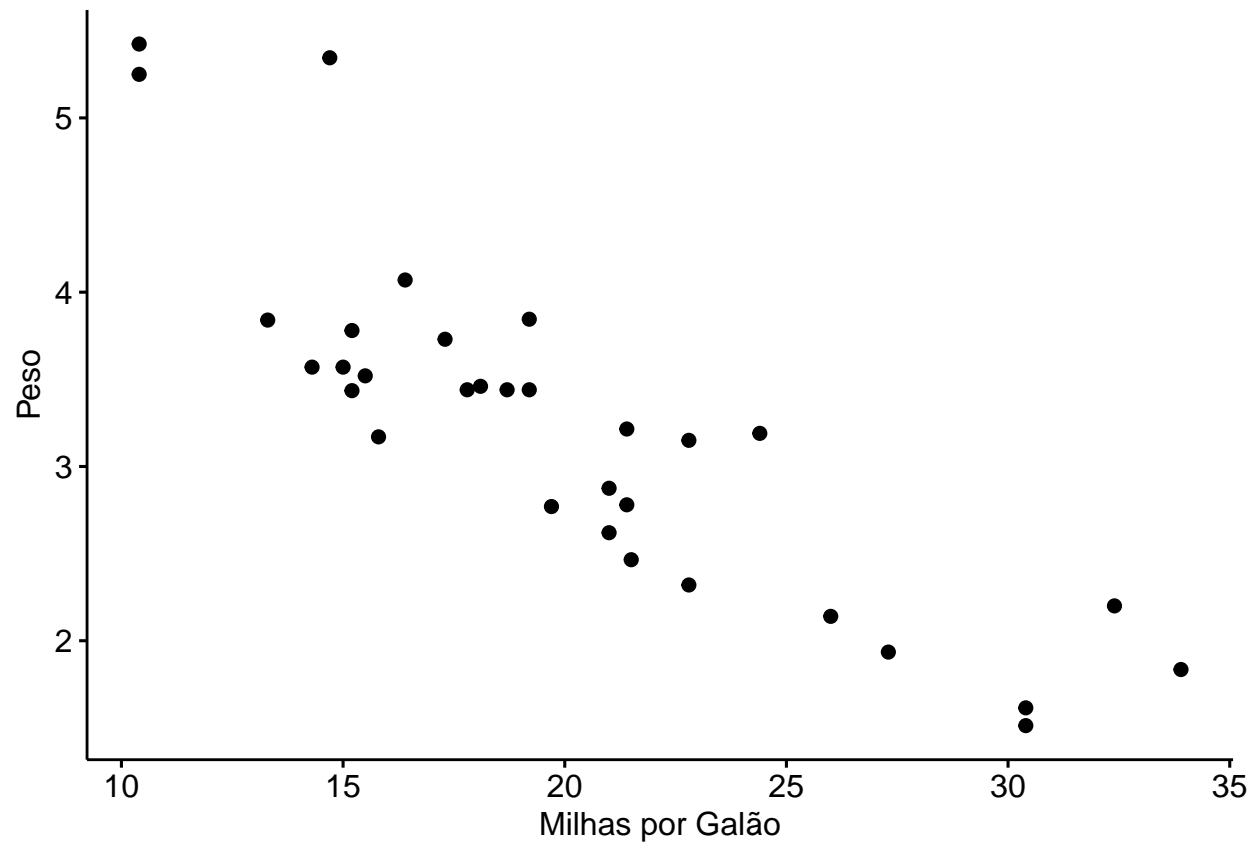
### Visualizando os dados

```
library(ggpubr)
```

```
## Loading required package: ggplot2
```

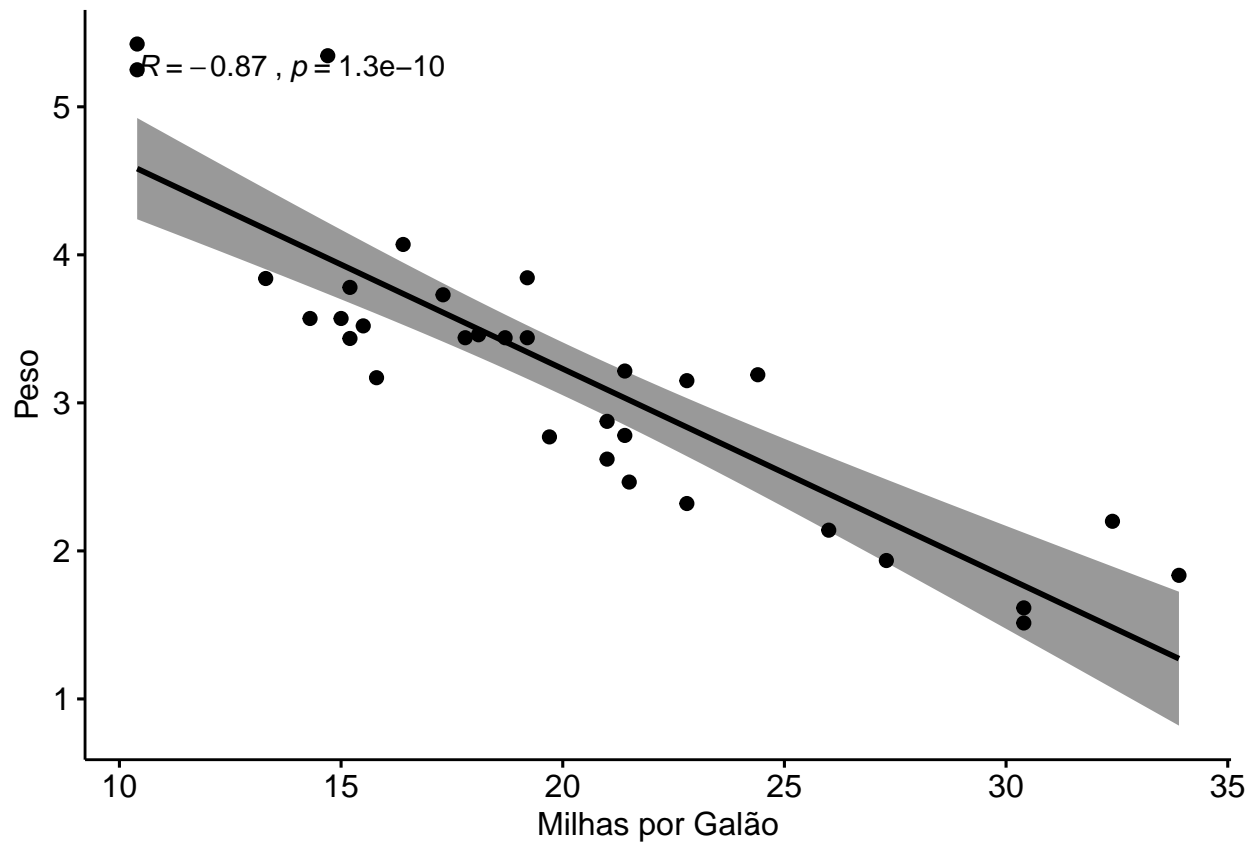
```
## Loading required package: magrittr
```

```
ggscatter(dados, x = "mpg", y = "wt",
          xlab = "Milhas por Galão", ylab = "Peso")
```



Acrescentando parâmetros estatísticos

```
ggscatter(dados, x = "mpg", y = "wt",  
          add = "reg.line", conf.int = TRUE,  
          cor.coef = TRUE, cor.method = "pearson",  
          xlab = "Milhas por Galão", ylab = "Peso")
```



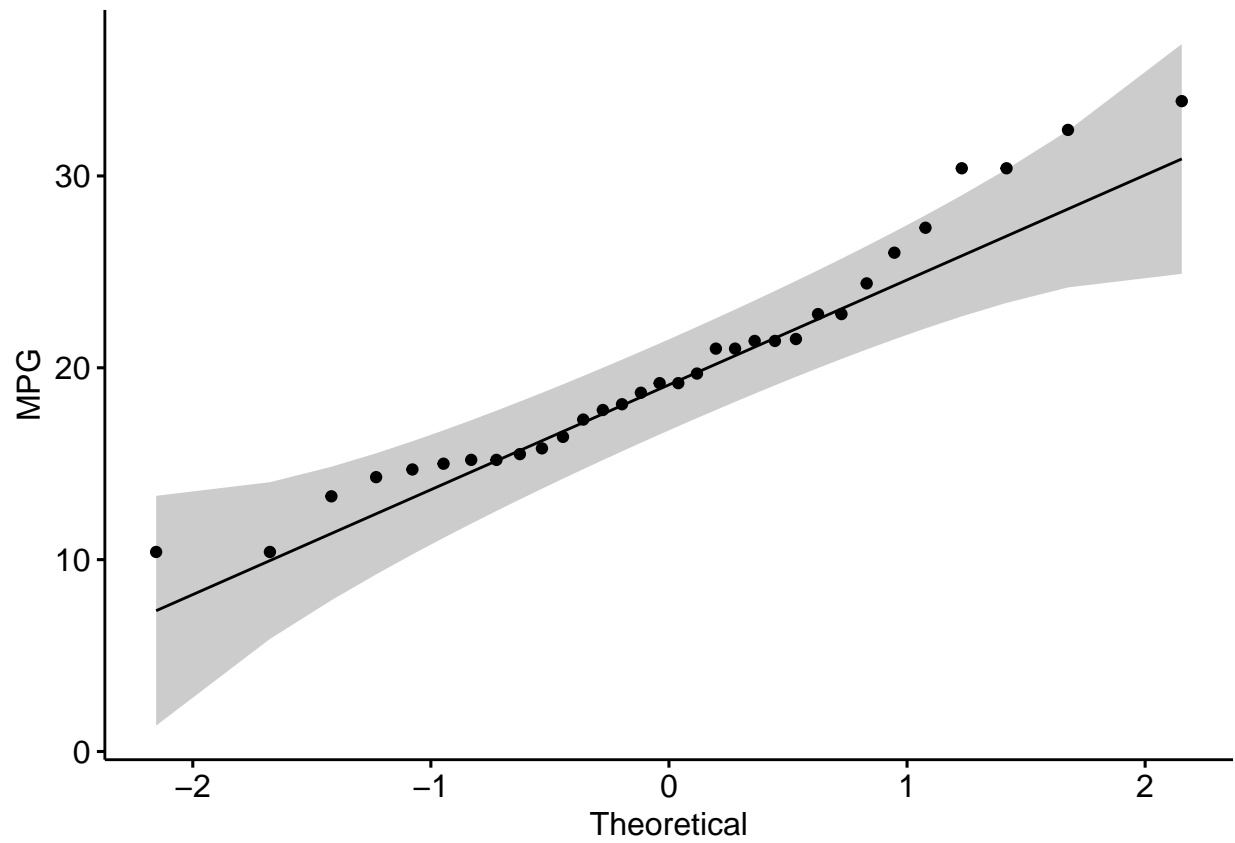
Analisando o gráfico é possível verificar uma covariação linear.

## Inspeção visual da normalidade dos dados usando Q-Q plots

Esse gráfico mostra a correlação entre uma amostra e uma distribuição normal.

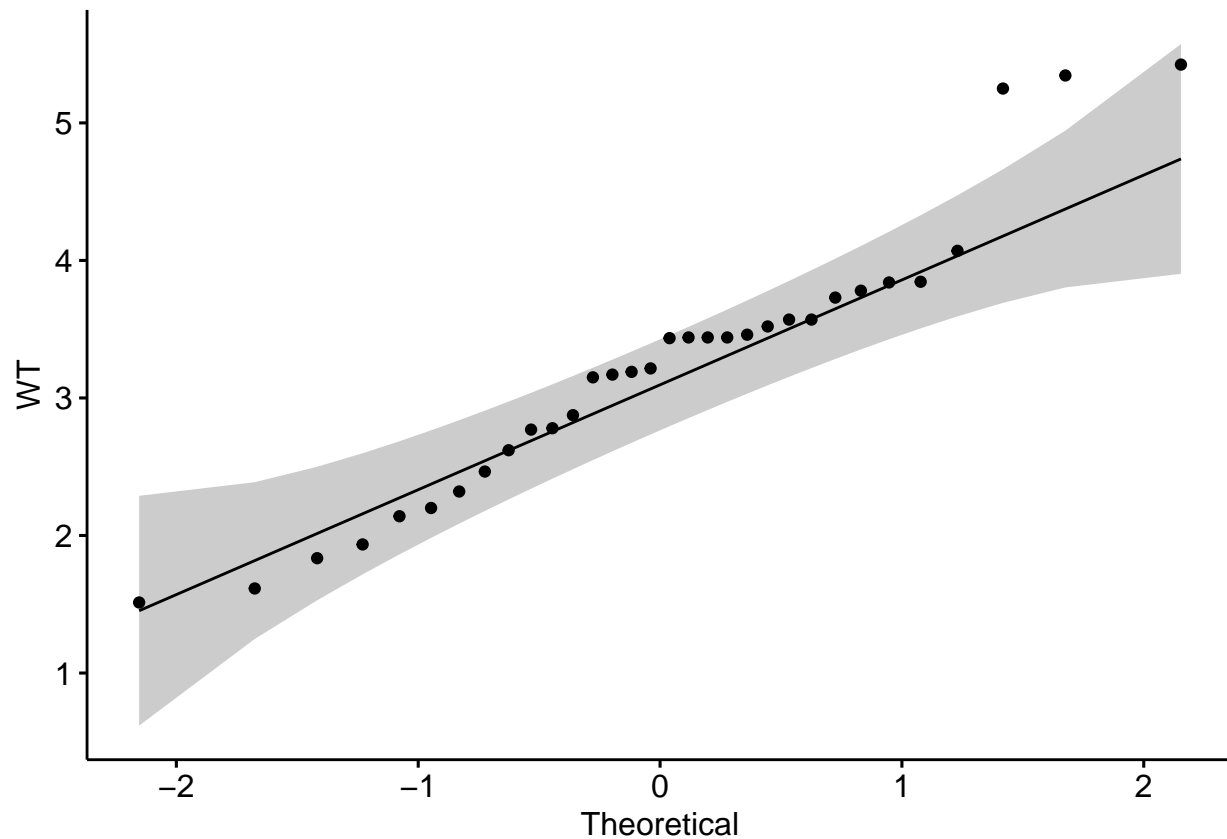
Variável “mpg”

```
ggqqplot(dados$mpg, ylab = "MPG")
```



Variável “wt”

```
ggqqplot(dados$wt, ylab = "WT")
```



**Importante!** Se os dados não estiverem normalmente distribuídos é recomendável usar outro método de correlação.

## Teste de correlação de Pearson

```
res <- cor.test(dados$wt, dados$mpg, method = "pearson")
res

##
## Pearson's product-moment correlation
##
## data: dados$wt and dados$mpg
## t = -9.559, df = 30, p-value = 1.294e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9338264 -0.7440872
## sample estimates:
## cor
## -0.8676594
```

Acessando os valores calculados pela função `cor.test()`: `p.value` e `estimate`

```
res$p.value
```

```
## [1] 1.293959e-10
```

```
res$estimate
```

```
##          cor  
## -0.8676594
```

## Matriz de Correlação

```
dados2 <- mtcars[, c(1,3,4,5,6,7)]  
head(dados2)
```

```
##           mpg disp  hp drat   wt  qsec  
## Mazda RX4      21.0  160 110 3.90 2.620 16.46  
## Mazda RX4 Wag  21.0  160 110 3.90 2.875 17.02  
## Datsun 710      22.8  108  93 3.85 2.320 18.61  
## Hornet 4 Drive  21.4  258 110 3.08 3.215 19.44  
## Hornet Sportabout 18.7  360 175 3.15 3.440 17.02  
## Valiant        18.1  225 105 2.76 3.460 20.22
```

```
mtx_cor <- cor(dados2)  
round(mtx_cor, 2)
```

```
##           mpg disp  hp drat   wt  qsec  
## mpg      1.00 -0.85 -0.78 0.68 -0.87 0.42  
## disp    -0.85  1.00 0.79 -0.71 0.89 -0.43  
## hp      -0.78 0.79  1.00 -0.45 0.66 -0.71  
## drat     0.68 -0.71 -0.45  1.00 -0.71 0.09  
## wt      -0.87 0.89 0.66 -0.71  1.00 -0.17  
## qsec     0.42 -0.43 -0.71 0.09 -0.17  1.00
```

Se o dataset contém valores em branco (missing) use: `cor(dados2, use = "complete.obs")`

## Matriz de correlação com valores-p (p-value)

A função `rcorr()[Hmisc]` pode ser usada para calcular os níveis de significância. A função retorna ambos os coeficientes de correlação e valores-p.

```
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
res2 <- rcorr(as.matrix(dados2))
res2
```

```
##      mpg  disp    hp  drat    wt  qsec
## mpg   1.00 -0.85 -0.78  0.68 -0.87  0.42
## disp -0.85  1.00  0.79 -0.71  0.89 -0.43
## hp   -0.78  0.79  1.00 -0.45  0.66 -0.71
## drat  0.68 -0.71 -0.45  1.00 -0.71  0.09
## wt   -0.87  0.89  0.66 -0.71  1.00 -0.17
## qsec  0.42 -0.43 -0.71  0.09 -0.17  1.00
##
## n= 32
##
##
## P
##      mpg    disp    hp    drat    wt    qsec
## mpg              0.0000 0.0000 0.0000 0.0000 0.0171
## disp 0.0000          0.0000 0.0000 0.0000 0.0131
## hp   0.0000 0.0000          0.0100 0.0000 0.0000
## drat 0.0000 0.0000 0.0100          0.0000 0.6196
## wt   0.0000 0.0000 0.0000 0.0000          0.3389
## qsec 0.0171 0.0131 0.0000 0.6196 0.3389
```

Extraindo os coeficientes de correlação

```
res2$r
```

```
##      mpg      disp      hp      drat      wt      qsec
## mpg   1.0000000 -0.8475514 -0.7761684  0.68117191 -0.8676594  0.41868403
## disp -0.8475514  1.0000000  0.7909486 -0.71021393  0.8879799 -0.43369788
## hp   -0.7761684  0.7909486  1.0000000 -0.44875912  0.6587479 -0.70822339
## drat  0.6811719 -0.7102139 -0.4487591  1.00000000 -0.7124406  0.09120476
## wt   -0.8676594  0.8879799  0.6587479 -0.71244065  1.0000000 -0.17471588
## qsec  0.4186840 -0.4336979 -0.7082234  0.09120476 -0.1747159  1.00000000
```

Extraindo os valores-p

```
res2$P
```

```
##      mpg      disp      hp      drat      wt
## mpg              NA 9.380328e-10 1.787835e-07 1.776240e-05 1.293958e-10
## disp 9.380328e-10          NA 7.142679e-08 5.282022e-06 1.222311e-11
## hp   1.787835e-07 7.142679e-08          NA 9.988772e-03 4.145827e-05
## drat 1.776240e-05 5.282022e-06 9.988772e-03          NA 4.784260e-06
## wt   1.293958e-10 1.222311e-11 4.145827e-05 4.784260e-06          NA
```

```
## qsec 1.708199e-02 1.314404e-02 5.766253e-06 6.195826e-01 3.388683e-01
##      qsec
## mpg  1.708199e-02
## disp 1.314404e-02
## hp   5.766253e-06
## drat 6.195826e-01
## wt   3.388683e-01
## qsec      NA
```

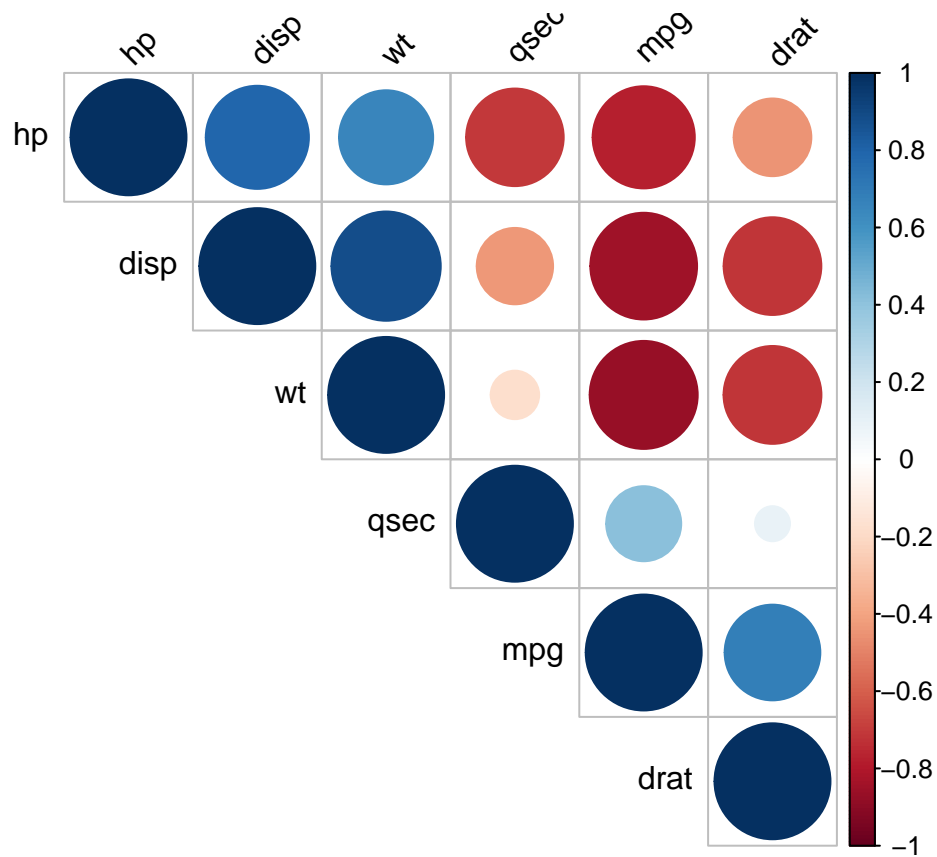
## Visualizando a matriz de correlação

Gráfico de correlação com a função `corrplot()`[`corrplot`] O parâmetro `type` pode ter os seguintes valores: “upper”, “lower”, “full”

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

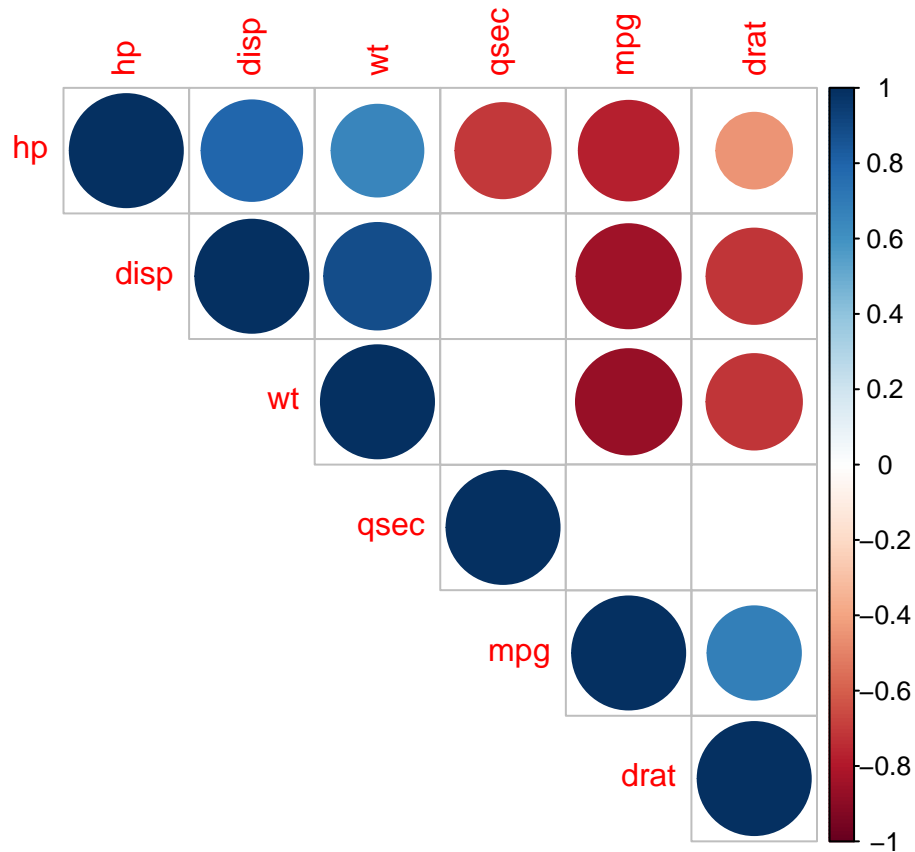
```
corrplot(mtx_cor, type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 45)
```





## Combinando os coeficientes de correlação e teste de significância

```
corrplot(res2$r, type="upper", order="hclust",  
         p.mat = res2$P, sig.level = 0.01, insig = "blank")
```



## Construindo scatter plots

A função `char.Correlation()` [PerformanceAnalytics] pode ser usada para mostrar uma coleção de gráficos de uma matriz de correlação

```
library(PerformanceAnalytics)
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend
```

```
chart.Correlation(dados2, histogram = TRUE, pch = 19)
```

