R Notebook

```
dataset <- c(-15,2,3,5,6,8,9,12)
zscore<-(dataset-mean(dataset))/sd(dataset)
zscore
```

```
## [1] -2.27496663 -0.21233022 -0.09099867  0.15166444  0.27299600  0.51565910
## [7]  0.63699066  1.00098532
```

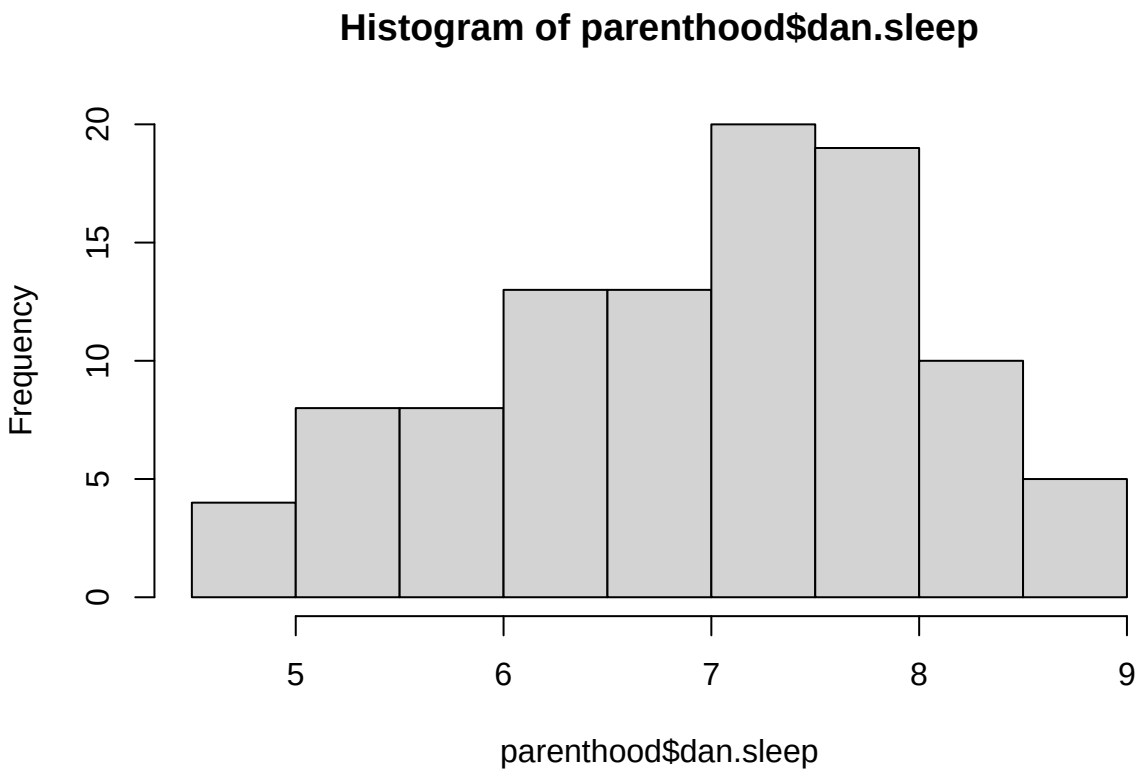
```
setwd("C:/Users/ushad/Documents/GitHub/BSE658/Module 3/Notebooks")

load( "parenthood.Rdata" )
head(parenthood)
```

```
##   dan.sleep baby.sleep dan.grump day
## 1      7.59      10.18        56   1
## 2      7.91      11.66        60   2
## 3      5.14       7.92        82   3
## 4      7.71       9.61        55   4
## 5      6.68       9.75        67   5
## 6      5.99       5.04        72   6
```

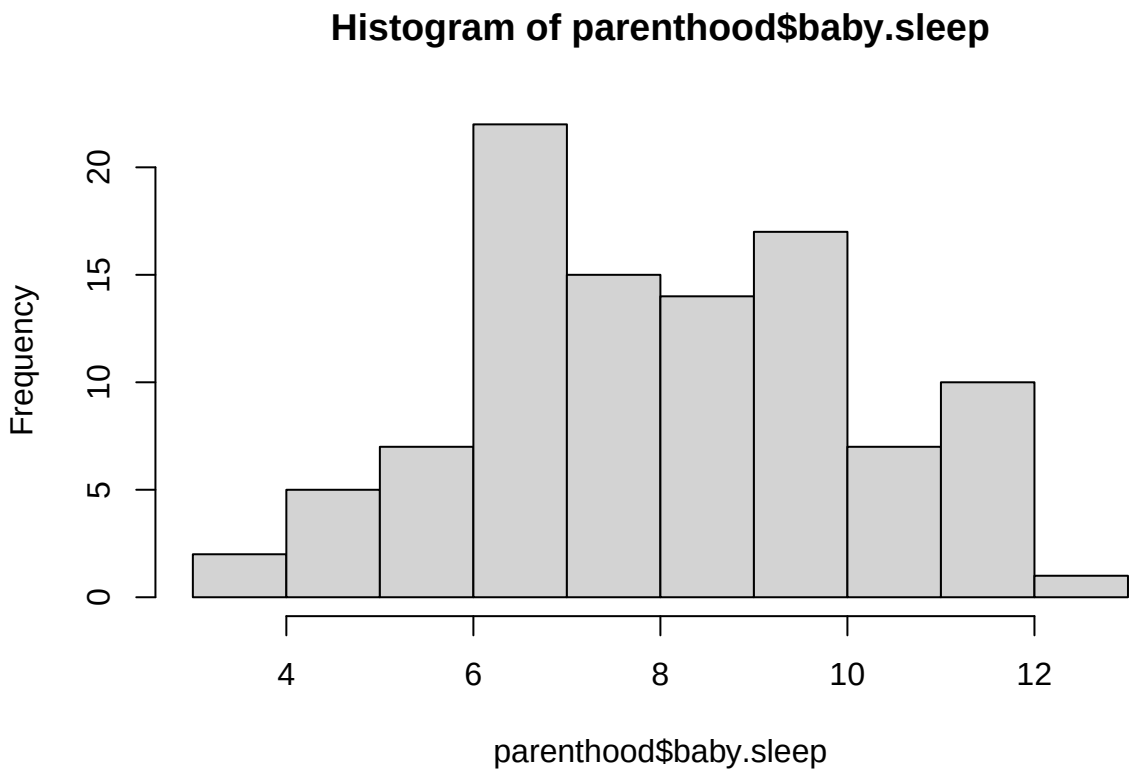
```
library(psych)
describe(parenthood)
```

```
##            vars   n  mean    sd median trimmed   mad   min    max range  skew
## dan.sleep     1 100  6.97  1.02   7.03    7.00  1.09  4.84   9.00  4.16 -0.29
## baby.sleep    2 100  8.05  2.07   7.95    8.05  2.33  3.25  12.07  8.82 -0.02
## dan.grump     3 100 63.71 10.05  62.00   63.16  9.64 41.00  91.00 50.00  0.43
## day           4 100 50.50 29.01  50.50   50.50 37.06  1.00 100.00 99.00  0.00
##            kurtosis   se
## dan.sleep     -0.72 0.10
## baby.sleep    -0.69 0.21
## dan.grump     -0.16 1.00
## day           -1.24 2.90
```

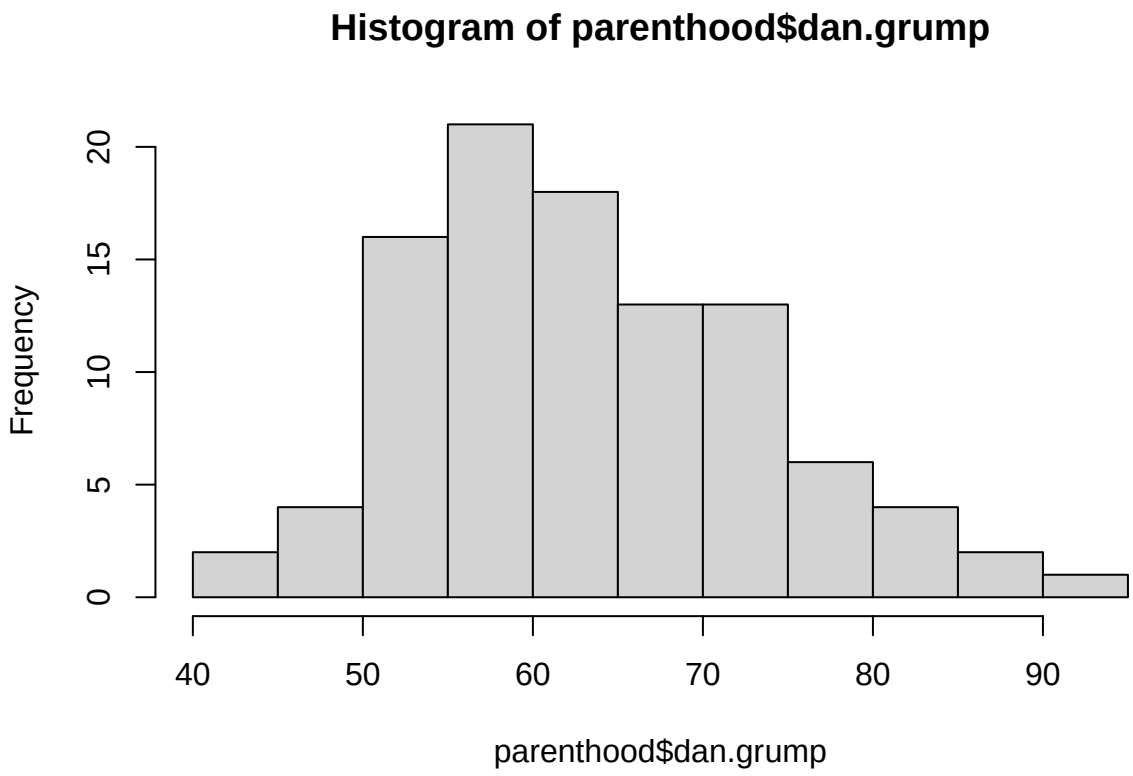
```
hist(parenthood$dan.sleep)
```



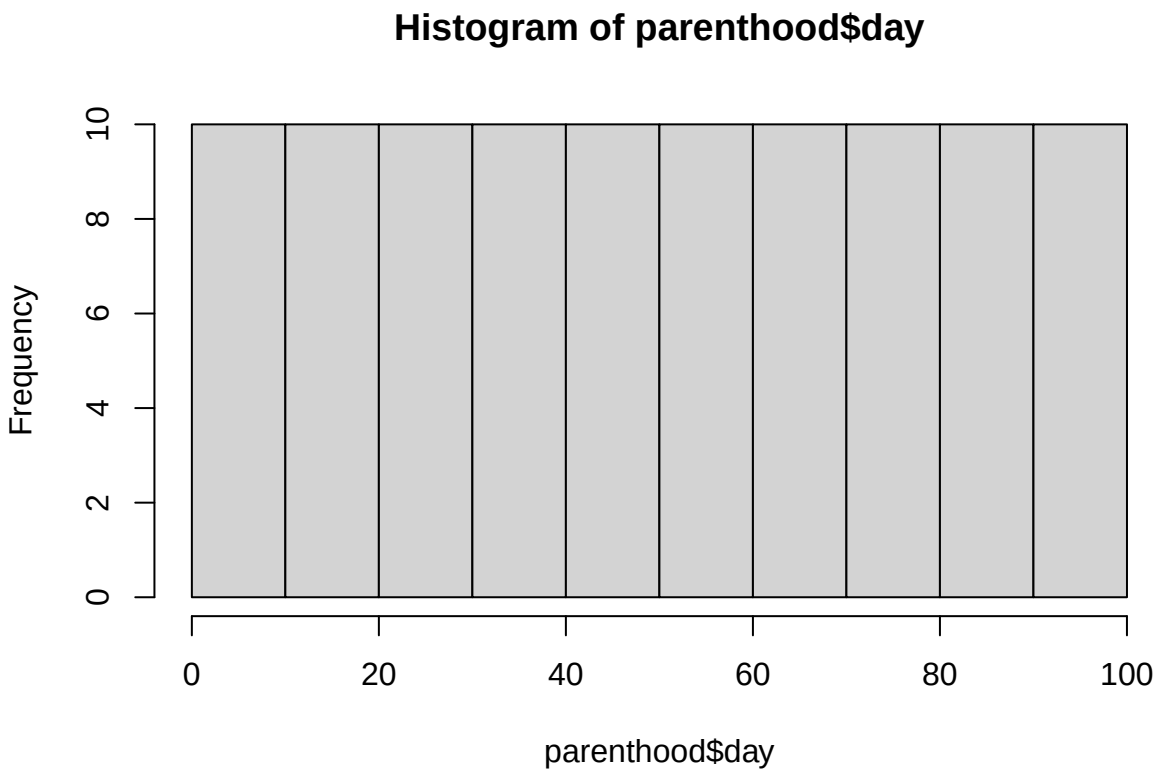

```
hist(parenthood$baby.sleep)
```




```
hist(parenthood$dan.grump)
```




```
hist(parenthood$day)
```



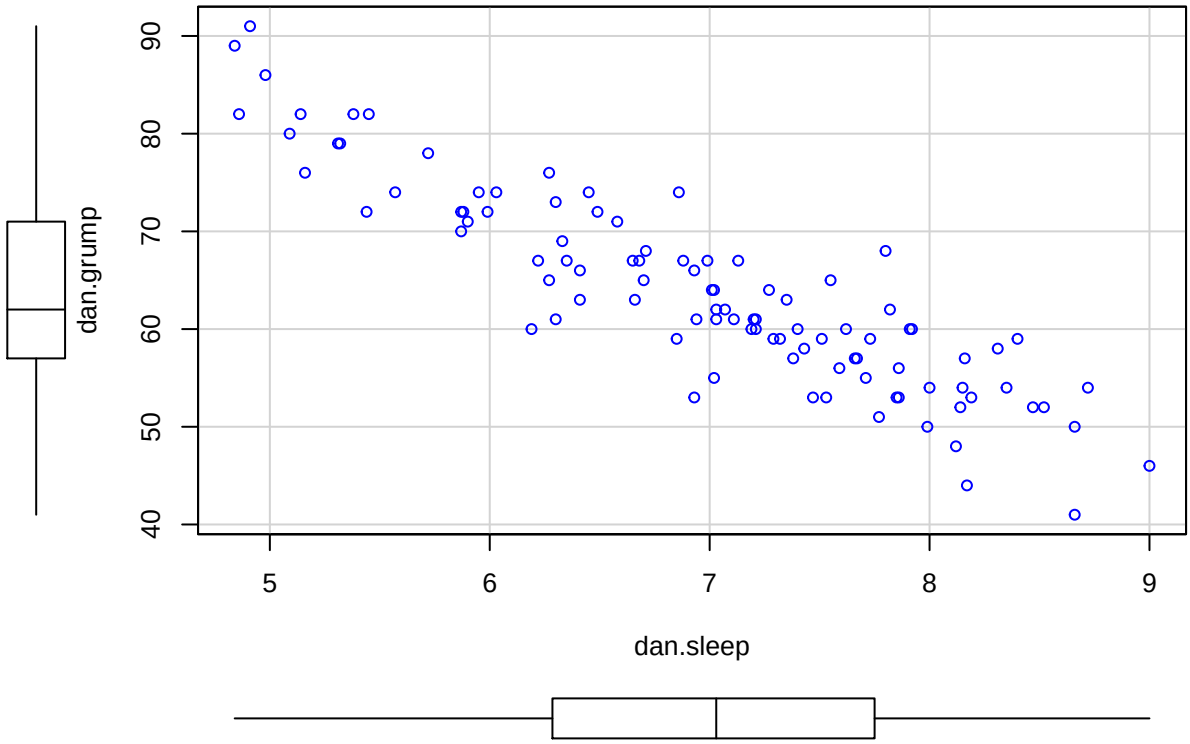

```
#install.packages("car")
#install.packages("Rcpp")
library(car) #does not knit if unhashed
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

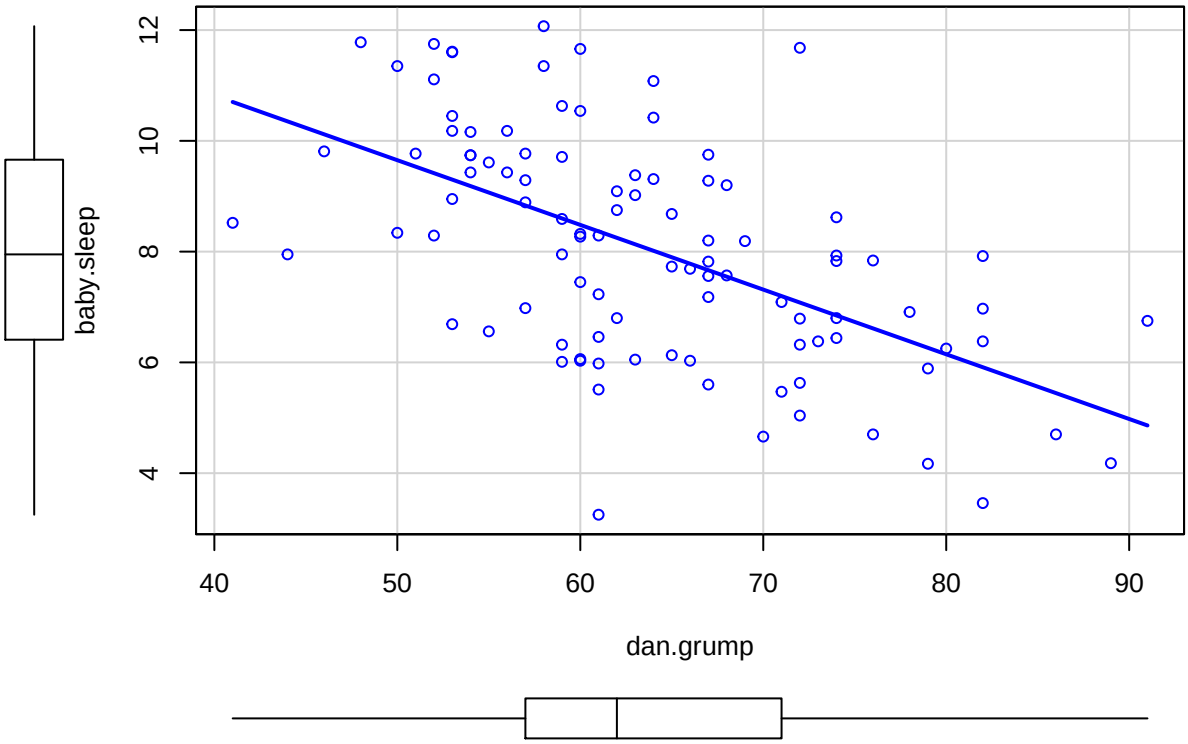
```
## The following object is masked from 'package:psych':
##
##     logit
```

```
scatterplot( dan.grump ~ dan.sleep, data = parenthood, regLine = FALSE, smooth = FALSE)
```



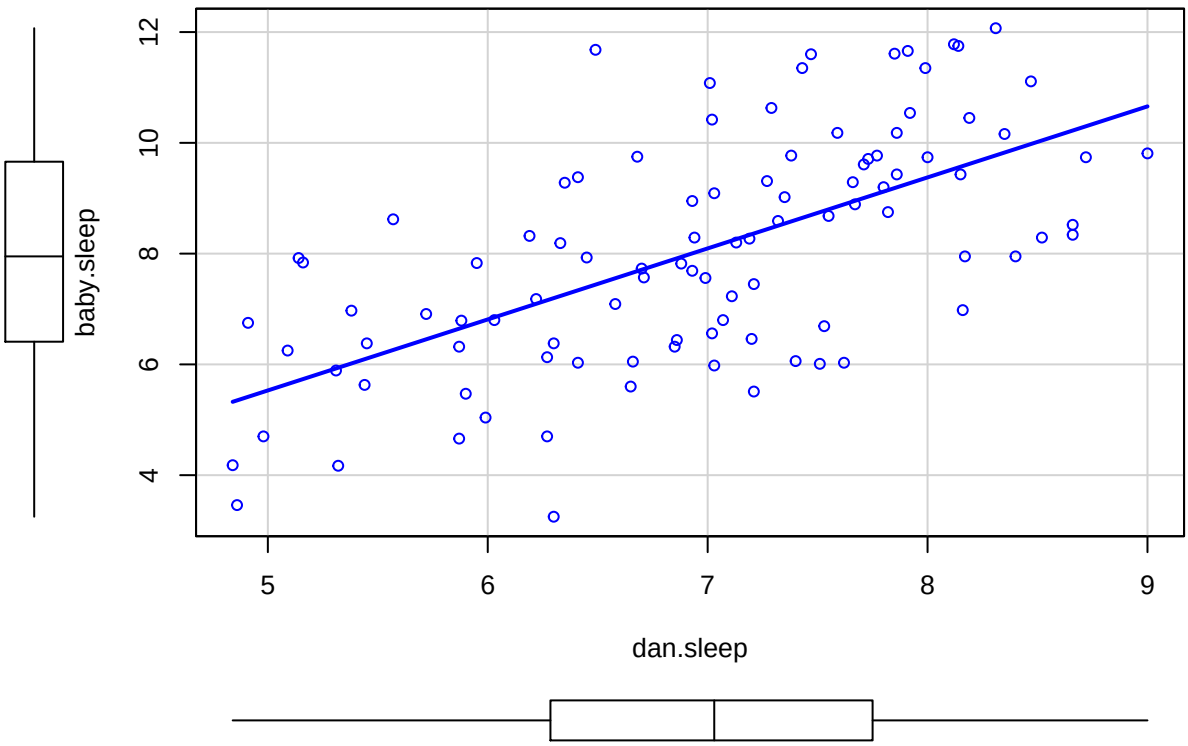

```
#scatterplot
```

```
scatterplot(baby.sleep ~ dan.grump, data=parenthood, smooth=FALSE)
```



baby sleep is positively correlated with dan sleep, dan sleep is negatively correlated with dan grump, dan grump is negatively correlated with baby sleep

```
scatterplot(baby.sleep ~ dan.sleep, data=parenthood, smooth=FALSE)
```




```
cor(x = parenthood$dan.sleep, y = parenthood$dan.grump)
```

```
## [1] -0.903384
```

```
cor(parenthood) #gives matrix of values
```

```
##              dan.sleep  baby.sleep   dan.grump         day
## dan.sleep   1.00000000  0.62794934 -0.90338404 -0.09840768
## baby.sleep  0.62794934  1.00000000 -0.56596373 -0.01043394
## dan.grump  -0.90338404 -0.56596373  1.00000000  0.07647926
## day        -0.09840768 -0.01043394  0.07647926  1.00000000
```

```
load("C:/Users/ushad/Documents/GitHub/BSE658/Module 3/Notebooks/work.Rdata")
#cor(work) shows error as work has non numeric data
```

```
library(lsr)
correlate(parenthood)
```

```
##
## CORRELATIONS
## ============
## - correlation type:  pearson
## - correlations shown only when both variables are numeric
```

```
##
##            dan.sleep baby.sleep dan.grump    day
## dan.sleep          .      0.628    -0.903 -0.098
## baby.sleep     0.628          .    -0.566 -0.010
## dan.grump     -0.903     -0.566         .  0.076
## day           -0.098     -0.010     0.076      .
```

```
correlate(parenthood, corr.method="spearman")
```

```
##
## CORRELATIONS
## ============
## - correlation type:  spearman
## - correlations shown only when both variables are numeric
##
##            dan.sleep baby.sleep dan.grump    day
## dan.sleep          .      0.645    -0.887 -0.089
## baby.sleep     0.645          .    -0.587  0.003
## dan.grump     -0.887     -0.587         .  0.056
## day           -0.089      0.003     0.056      .
```

```
correlate(work)
```

```
##
## CORRELATIONS
## ============
## - correlation type:  pearson
## - correlations shown only when both variables are numeric
##
##           hours  tasks   pay    day weekday   week day.type
## hours         .  0.800 0.760 -0.049       .  0.018        .
## tasks     0.800      . 0.720 -0.072       . -0.013        .
## pay       0.760  0.720     .  0.137       .  0.196        .
## day      -0.049 -0.072 0.137      .       .  0.990        .
## weekday       .      .     .      .       .      .        .
## week      0.018 -0.013 0.196  0.990       .      .        .
## day.type      .      .     .      .       .      .        .
```

```
cor(work$hours, work$pay)
```

```
## [1] 0.7604283
```

```
load( "C:/Users/ushad/Documents/GitHub/BSE658/Module 3/Notebooks/effort.Rdata" )
effort
```

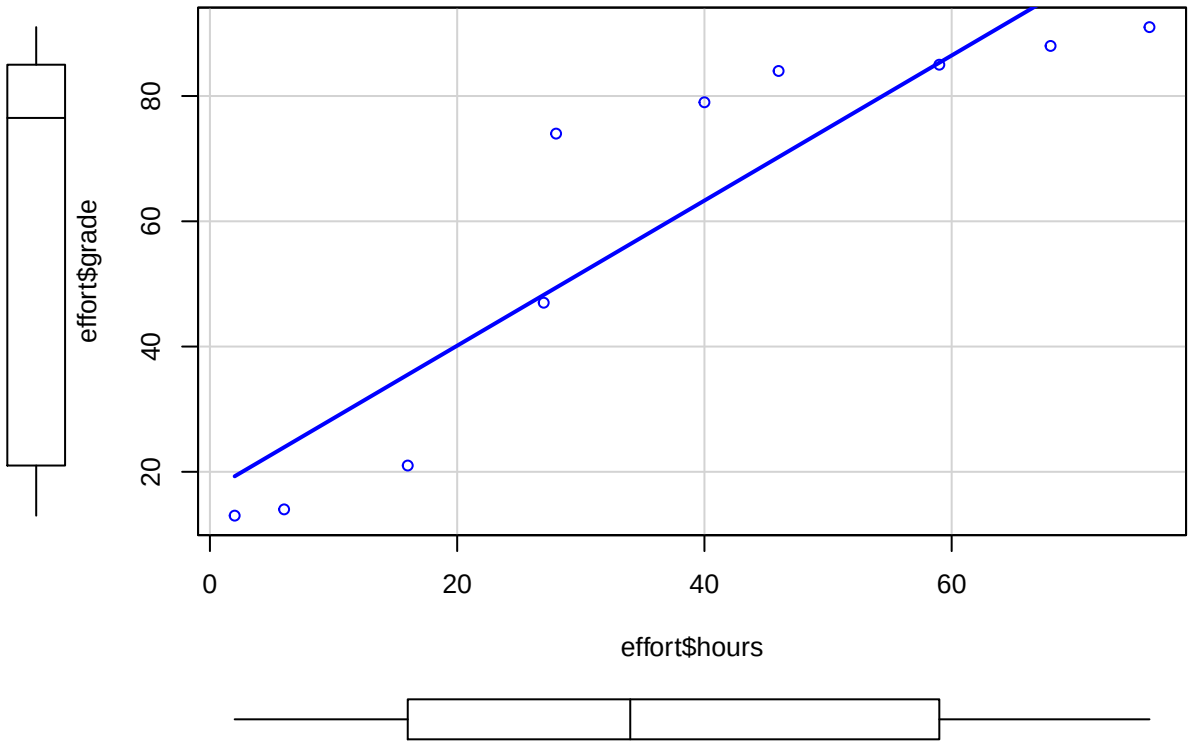
```
##    hours grade
## 1      2    13
## 2     76    91
## 3     40    79
## 4      6    14
## 5     16    21
```

```
## 6     28    74
## 7     27    47
## 8     59    85
## 9     46    84
## 10    68    88
```

```
cor( effort$hours, effort$grade )
```

```
## [1] 0.909402
```

```
scatterplot(effort$hours, effort$grade, regLine = TRUE, smooth = FALSE)
```




```
cor(effort$hours, effort$grade, method="spearman")
```

```
## [1] 1
```

```
load( "C:/Users/ushad/Documents/GitHub/BSE658/Module 3/Notebooks/anscombesquartet.Rdata" )
cor(X1, Y1 )
```

```
## [1] 0.8164205
```

```r
cor(X2, Y2 )
```

```
## [1] 0.8162365
```

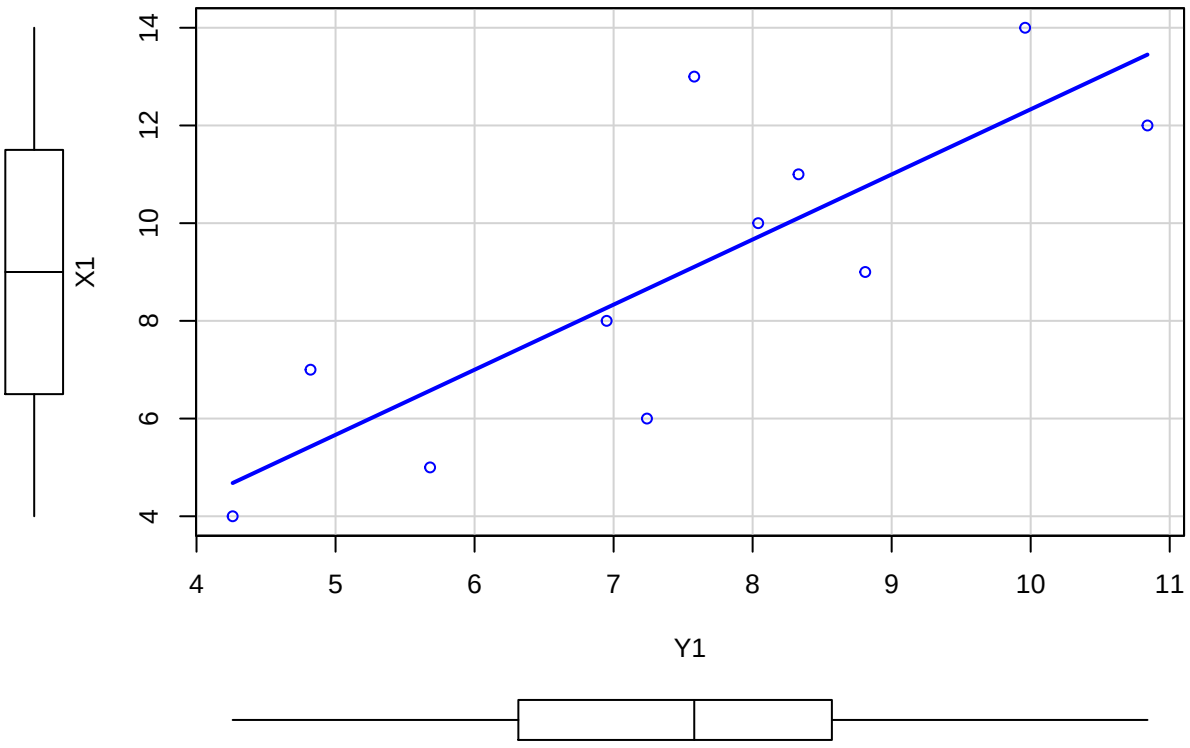
```r
cor(X3, Y3)
```

```
## [1] 0.8162867
```

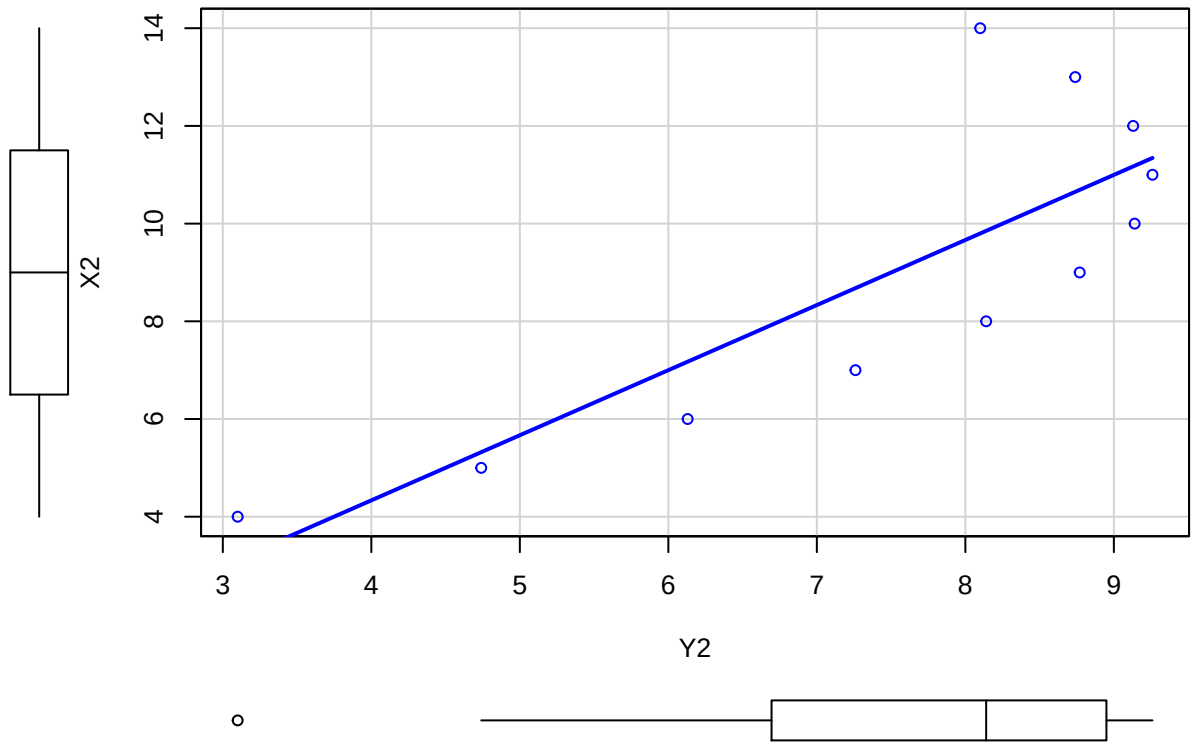
```r
cor(X4, Y4)
```

```
## [1] 0.8165214
```

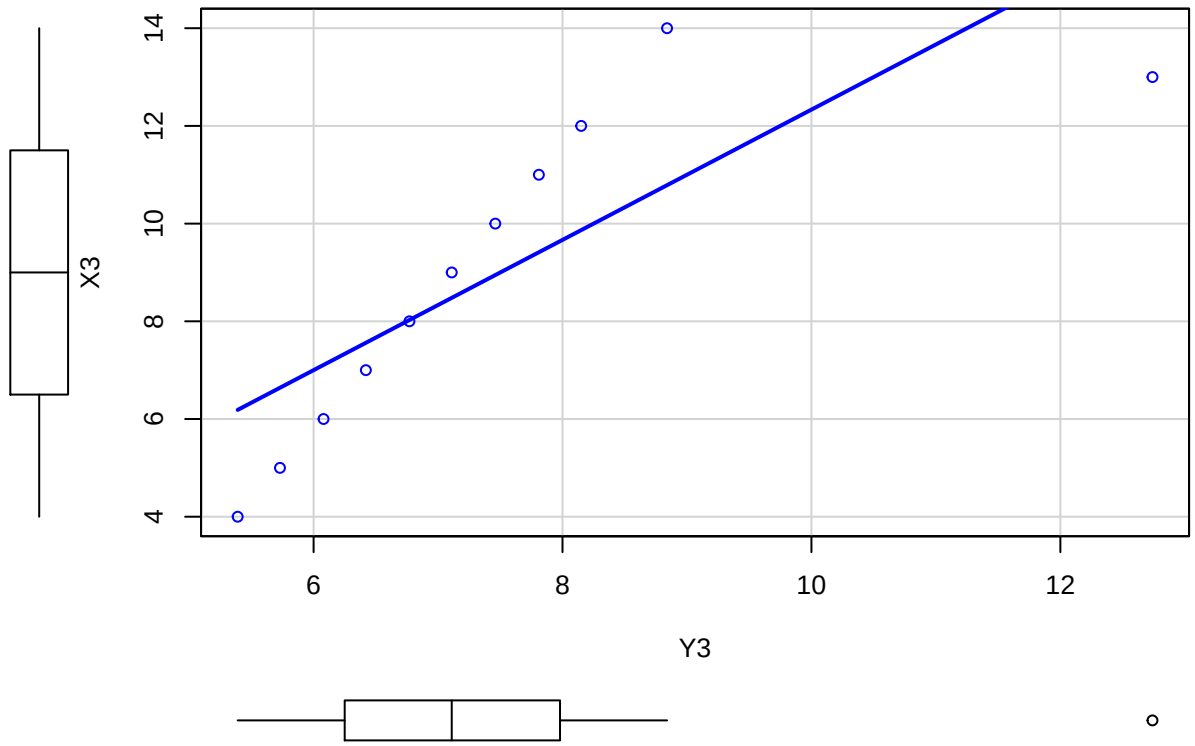
```r
scatterplot(X1~Y1, smooth=FALSE)
```



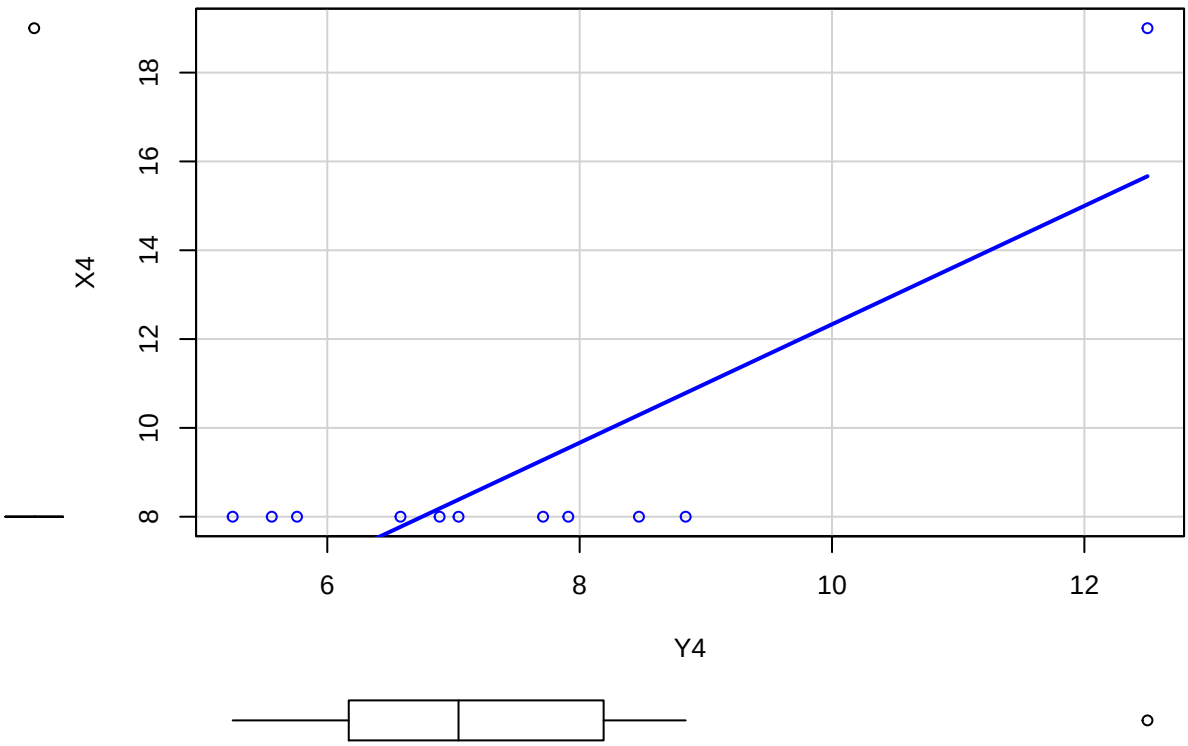

```r
scatterplot(X2~Y2, smooth=FALSE)
```




```
scatterplot(X3~Y3, smooth=FALSE)
```




```
scatterplot(X4~Y4, smooth=FALSE)
```




```
load( "C:/Users/ushad/Documents/GitHub/BSE658/Module 3/Notebooks/parenthood2.Rdata" )
print( parenthood2 )
```

```
##     dan.sleep baby.sleep dan.grump day
## 1        7.59         NA        56   1
## 2        7.91      11.66        60   2
## 3        5.14       7.92        82   3
## 4        7.71       9.61        55   4
## 5        6.68       9.75        NA   5
## 6        5.99       5.04        72   6
## 7        8.19      10.45        53   7
## 8        7.19       8.27        60   8
## 9        7.40       6.06        NA   9
## 10       6.58       7.09        71  10
## 11       6.49      11.68        72  11
## 12       6.27       6.13        65  12
## 13       5.95       7.83        74  13
## 14       6.65       5.60        67  14
## 15       6.41       6.03        66  15
## 16       6.33       8.19        69  16
## 17       6.30       6.38        73  17
## 18       8.47      11.11        52  18
## 19       7.21       5.51        61  19
## 20       7.53       6.69        53  20
## 21       8.00       9.74        54  21
```

```
## 22       7.35       9.02       63  22
## 23       6.86       6.44       74  23
## 24       7.86       9.43       56  24
## 25       4.86       3.46       82  25
## 26       5.87       6.32       72  26
## 27       8.40       7.95       NA  27
## 28         NA       7.69       66  28
## 29       7.21       7.45       60  29
## 30       6.99         NA       67  30
## 31       8.17       7.95       44  31
## 32       7.85         NA       53  32
## 33       6.27       4.70       76  33
## 34       8.66       8.52       41  34
## 35       4.98       4.70       86  35
## 36       6.19       8.32       60  36
## 37       6.41       9.38       63  37
## 38       4.84       4.18       89  38
## 39       7.03       5.98       61  39
## 40       7.66       9.29       57  40
## 41       7.51         NA       59  41
## 42       7.92      10.54       60  42
## 43       8.12      11.78       48  43
## 44       7.47      11.60       53  44
## 45       7.99      11.35       50  45
## 46       5.44       5.63       72  46
## 47       8.16       6.98       57  47
## 48       7.62       6.03       60  48
## 49       5.87       4.66       70  49
## 50       9.00       9.81       46  50
## 51       8.31      12.07       58  51
## 52       6.71       7.57       68  52
## 53       7.43      11.35       58  53
## 54       5.90         NA       71  54
## 55       8.52       8.29       52  55
## 56       6.03         NA       74  56
## 57       7.29         NA       59  57
## 58       7.32       8.59       59  58
## 59       6.88       7.82       67  59
## 60       6.22       7.18       67  60
## 61       6.94       8.29       61  61
## 62       7.01      11.08       64  62
## 63         NA       6.46       61  63
## 64         NA       3.25       61  64
## 65         NA       9.74       54  65
## 66       7.82       8.75       62  66
## 67       8.14      11.75       52  67
## 68       7.27       9.31       64  68
## 69         NA       7.73       65  69
## 70       7.55       8.68       NA  70
## 71       7.38       9.77       57  71
## 72       7.73       9.71       59  72
## 73       5.32         NA       79  73
## 74       7.86      10.18       53  74
## 75       6.35       9.28       NA  75
```

```
## 76       7.11         NA        61  76
## 77       5.45       6.38        82  77
## 78       7.80       9.20        68  78
## 79       7.13       8.20        67  79
## 80       8.35      10.16        54  80
## 81       6.93       8.95        53  81
## 82         NA       6.80        62  82
## 83       8.66       8.34        50  83
## 84       5.09       6.25        NA  84
## 85       4.91         NA        NA  85
## 86       7.03       9.09        62  86
## 87       7.02      10.42        64  87
## 88         NA       8.89        57  88
## 89       8.15       9.43        54  89
## 90       5.88       6.79        NA  90
## 91         NA       6.91        78  91
## 92       6.66       6.05        63  92
## 93       6.85         NA        59  93
## 94       5.57       8.62        74  94
## 95       5.16       7.84        76  95
## 96         NA       5.89        79  96
## 97       7.77       9.77        51  97
## 98       5.38       6.97        82  98
## 99       7.02       6.56        55  99
## 100      6.45       7.93        74 100
```

```
describe( parenthood2 )
```

```
##            vars   n  mean    sd median trimmed   mad   min    max range  skew
## dan.sleep     1  91  6.98  1.02   7.03    7.02  1.13  4.84   9.00  4.16 -0.33
## baby.sleep    2  89  8.11  2.05   8.20    8.13  2.28  3.25  12.07  8.82 -0.09
## dan.grump     3  92 63.15  9.85  61.00   62.66 10.38 41.00  89.00 48.00  0.38
## day           4 100 50.50 29.01  50.50   50.50 37.06  1.00 100.00 99.00  0.00
##            kurtosis   se
## dan.sleep     -0.73 0.11
## baby.sleep    -0.59 0.22
## dan.grump     -0.31 1.03
## day           -1.24 2.90
```

```
#n=91 for dan sleep, n=89 for baby sleep, n=92 for dan grump, which means dan sleep has 9, baby sleep h
```

```
cor(parenthood2) #gives NA in matrix cells because there are missing values in all headers
```

```
##            dan.sleep baby.sleep dan.grump day
## dan.sleep          1         NA        NA  NA
## baby.sleep        NA          1        NA  NA
## dan.grump         NA         NA         1  NA
## day               NA         NA        NA   1
```

```
correlate(parenthood2)
```

```
##
```

```
## CORRELATIONS
## ============
## - correlation type:  pearson
## - correlations shown only when both variables are numeric
##
##            dan.sleep baby.sleep dan.grump    day
## dan.sleep          .      0.615    -0.903 -0.077
## baby.sleep     0.615          .    -0.568  0.058
## dan.grump     -0.903     -0.568         .  0.006
## day           -0.077      0.058     0.006      .
```

```
cor(parenthood2, use = "complete.obs") #not same as correlate function
```

```
##              dan.sleep baby.sleep   dan.grump         day
## dan.sleep   1.00000000  0.6394985 -0.89951468  0.06132891
## baby.sleep  0.63949845  1.0000000 -0.58656066  0.14555814
## dan.grump  -0.89951468 -0.5865607  1.00000000 -0.06816586
## day         0.06132891  0.1455581 -0.06816586  1.00000000
```

```
cor(parenthood2, use = "pairwise.complete.obs") #gives result as same as correlate function
```

```
##              dan.sleep  baby.sleep    dan.grump          day
## dan.sleep   1.00000000  0.61472303 -0.903442442 -0.076796665
## baby.sleep  0.61472303  1.00000000 -0.567802669  0.058309485
## dan.grump  -0.90344244 -0.56780267  1.000000000  0.005833399
## day        -0.07679667  0.05830949  0.005833399  1.000000000
```

```
mean(x=parenthood2$baby.sleep, na.rm=TRUE)
```

```
## [1] 8.114494
```

```
#install.packages("ggcorrplot") does not knit without hashing
library(ggcorrplot)
```

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

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
data(mtcars)
?data
```

```
## starting httpd help server ... done
```

```r
summary(mtcars) #there are 11 columns
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```r
head(mtcars) #lets compute only for first 4 columns
```

```
##                   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
```

```r
newcor<-cor(mtcars[,1:4],mtcars[,1:4])
newcor
```

```
##             mpg        cyl       disp         hp
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475
## disp -0.8475514  0.9020329  1.0000000  0.7909486
## hp   -0.7761684  0.8324475  0.7909486  1.0000000
```

```r
correlate(mtcars[,1:4])
```

```
##
## CORRELATIONS
## ============
## - correlation type:  pearson
## - correlations shown only when both variables are numeric
##
```

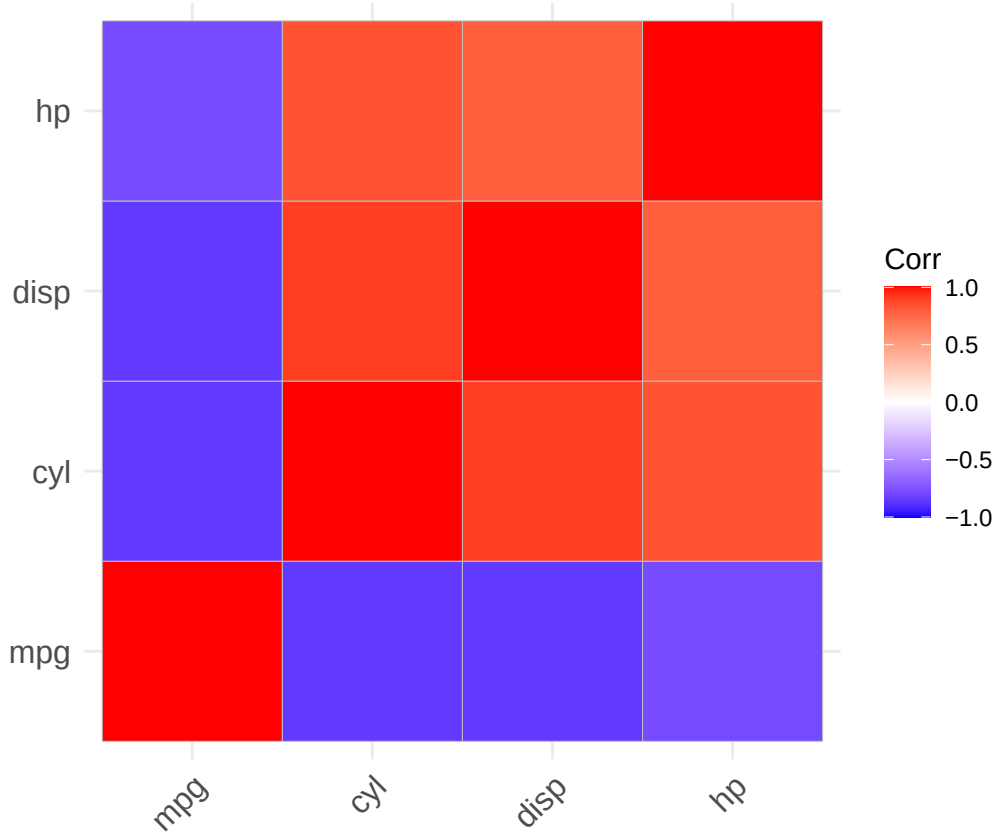
```
##         mpg    cyl   disp     hp
## mpg       . -0.852 -0.848 -0.776
## cyl  -0.852      .  0.902  0.832
## disp -0.848  0.902      .  0.791
## hp   -0.776  0.832  0.791      .
```

```
pval<-cor_pmat(mtcars[,1:4]) #p-value
pval
```

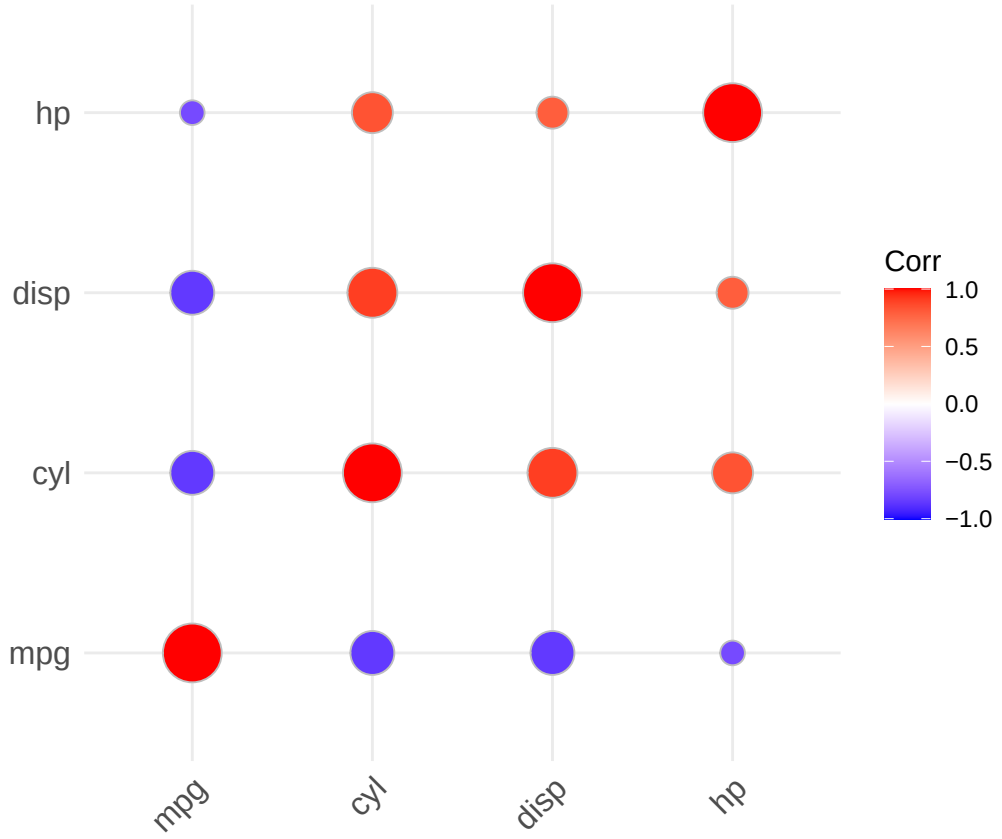
```
##                mpg          cyl         disp           hp
## mpg  0.000000e+00 6.112687e-10 9.380327e-10 1.787835e-07
## cyl  6.112687e-10 0.000000e+00 1.802838e-12 3.477861e-09
## disp 9.380327e-10 1.802838e-12 0.000000e+00 7.142679e-08
## hp   1.787835e-07 3.477861e-09 7.142679e-08 0.000000e+00
```

```
#what does head do? my round is not working
```

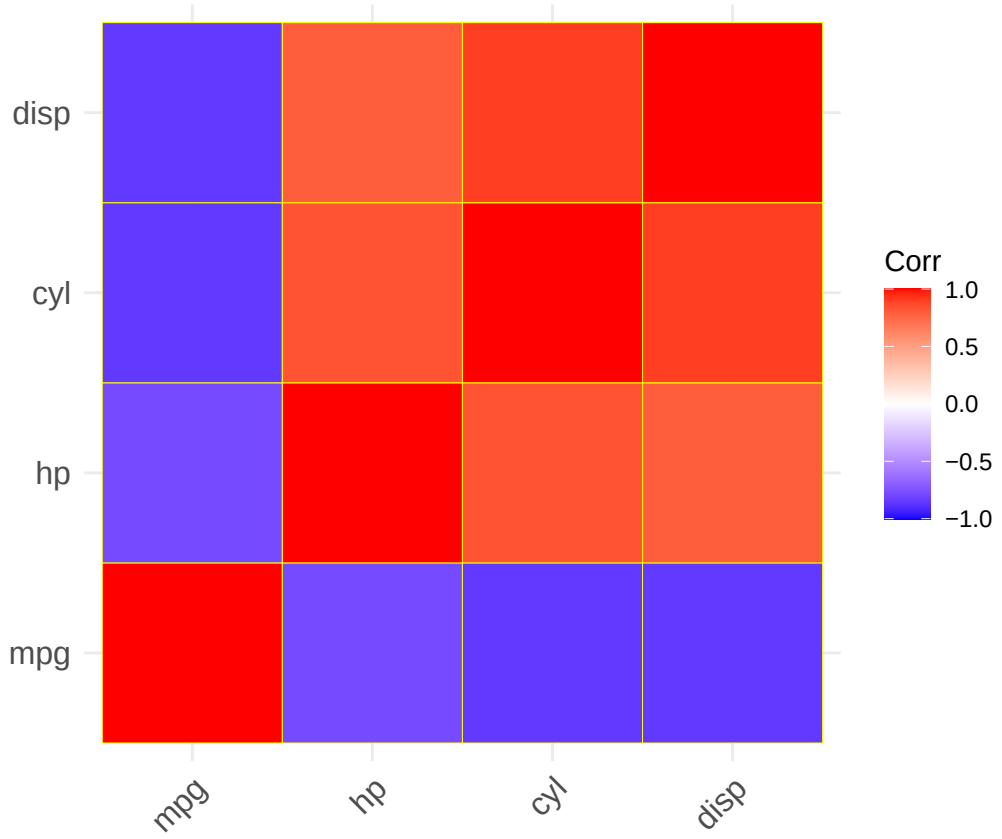
```
ggcorrplot(newcor) #does not take result of correlate function, only cor function as correlate function
```




```
ggcorrplot(newcor, method="circle")
```

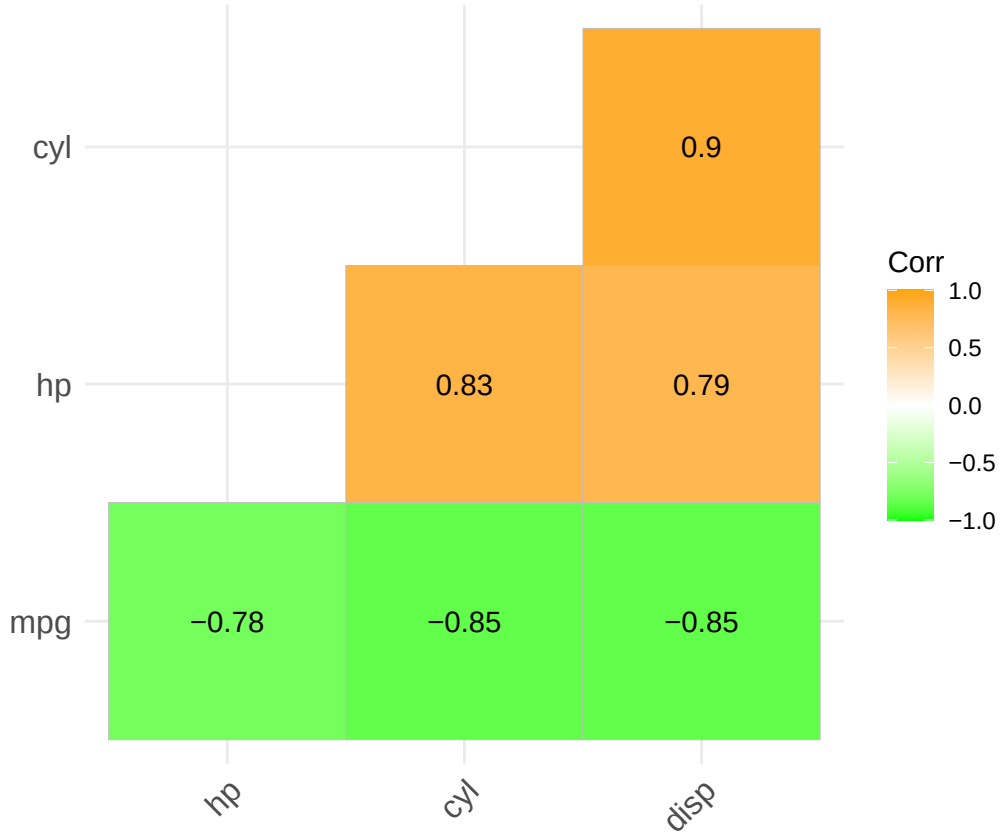



```
ggcorrplot(newcor, hc.order = TRUE, outline.col = "yellow") #what is heirarchial clustering?
```



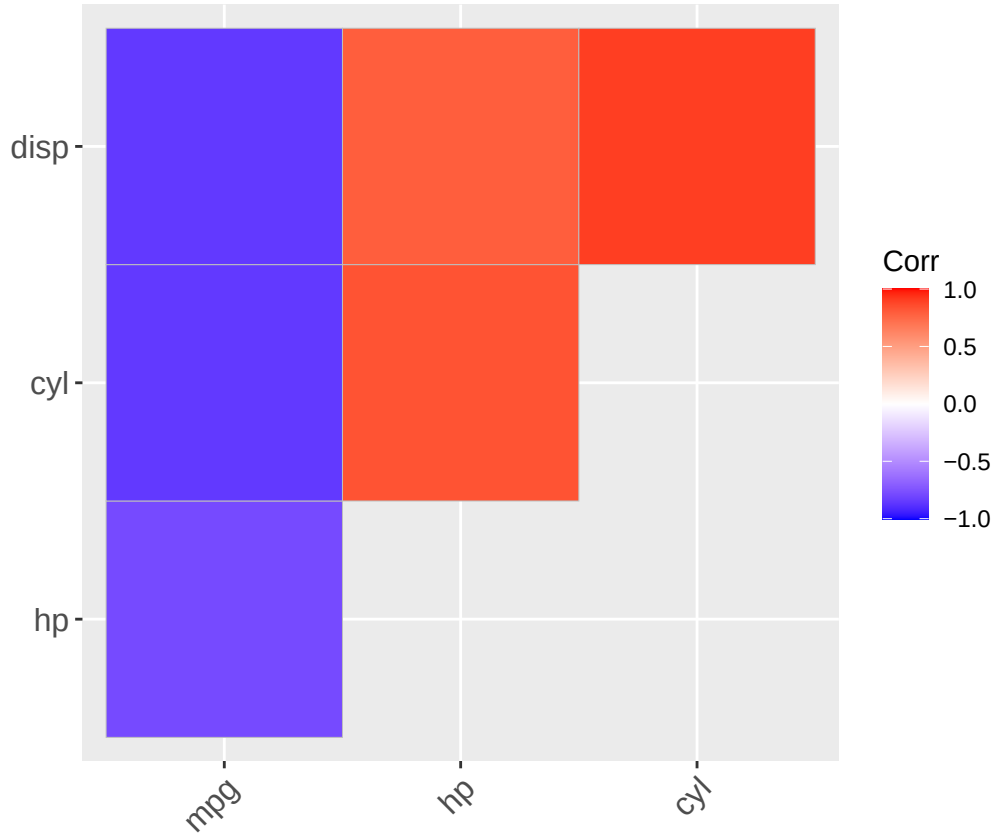

```
#i think it arranges the squares in hot-cold order
```

```
ggcorrplot(newcor, hc.order=TRUE, type="lower", colors=c("green","white","orange"), lab=TRUE)
```



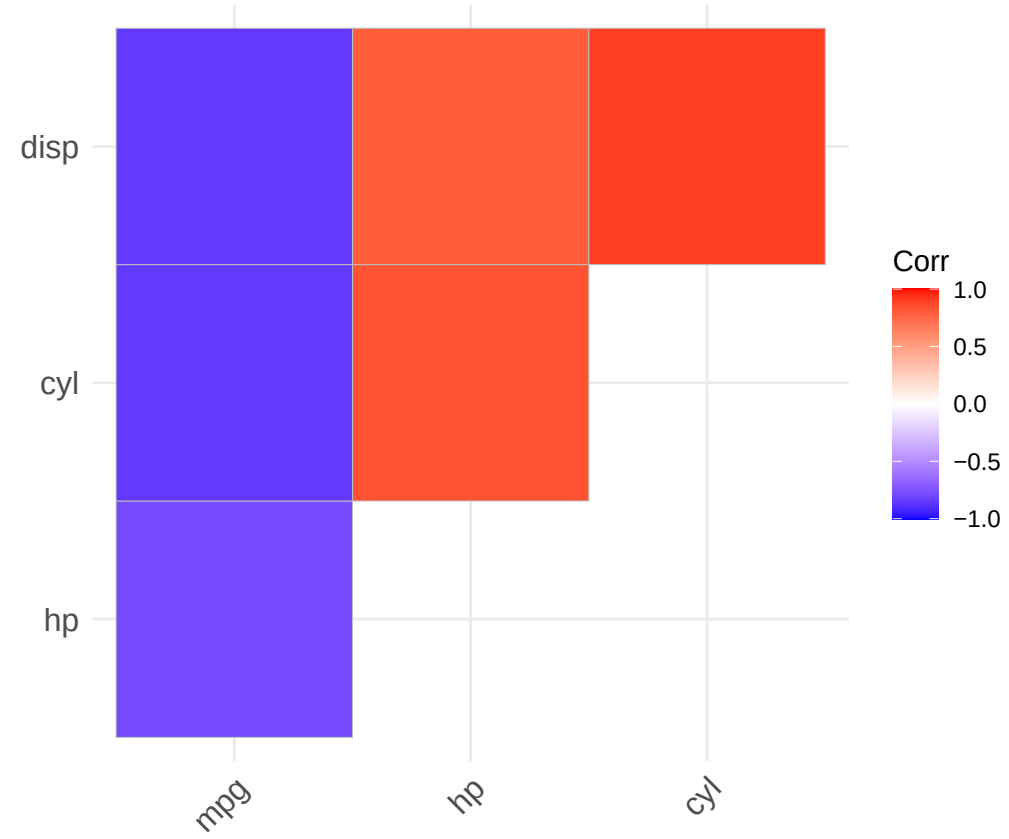

```
#3 temp colors, one for +ve 1, one for -ve 1, one for 0
```

```
ggcorrplot(newcor, hc.order=TRUE, type="upper",ggtheme = ggplot2::theme_gray)
```




```
#shows only non redundant values
#ggtheme changes bg color
```

```
ggcorrplot(newcor, hc.order=TRUE, type="upper", p.mat=pval) #no insignificant values
```



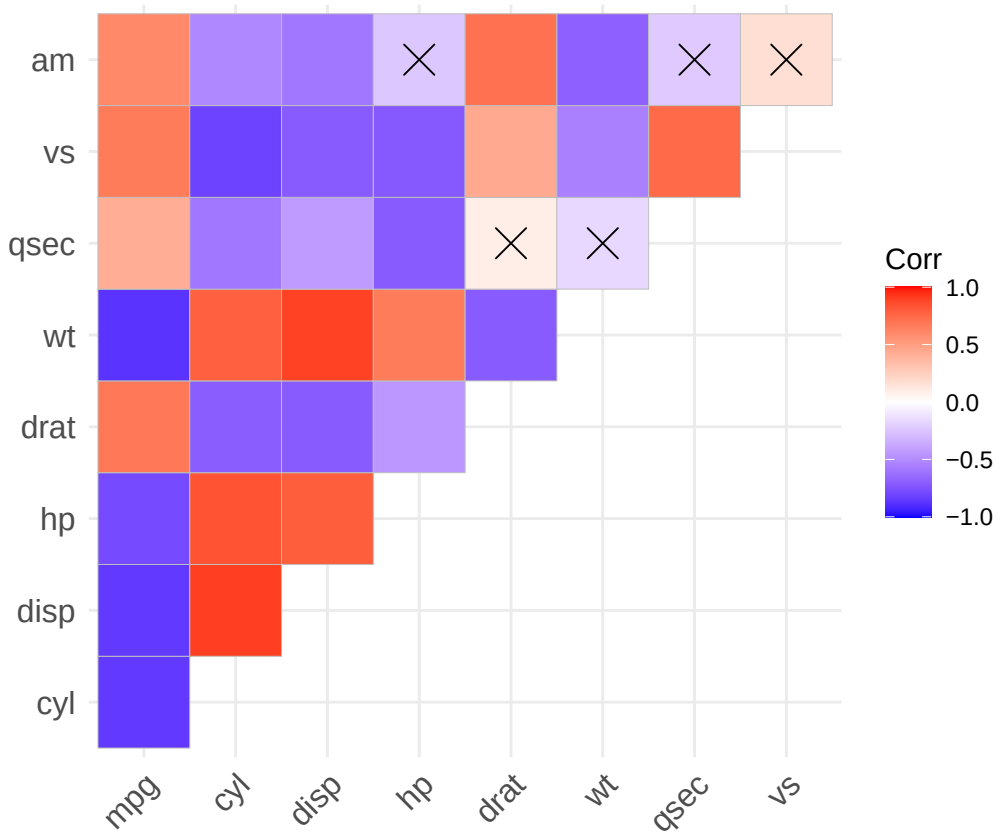

```
#took larger data to demo insignificant p values
corsev<-cor(mtcars[,1:9])
corsev
```

```
##             mpg        cyl       disp         hp        drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
##             qsec         vs         am
## mpg   0.41868403  0.6640389  0.5998324
## cyl  -0.59124207 -0.8108118 -0.5226070
## disp -0.43369788 -0.7104159 -0.5912270
## hp   -0.70822339 -0.7230967 -0.2432043
## drat  0.09120476  0.4402785  0.7127111
## wt   -0.17471588 -0.5549157 -0.6924953
## qsec  1.00000000  0.7445354 -0.2298609
## vs    0.74453544  1.0000000  0.1683451
## am   -0.22986086  0.1683451  1.0000000
```

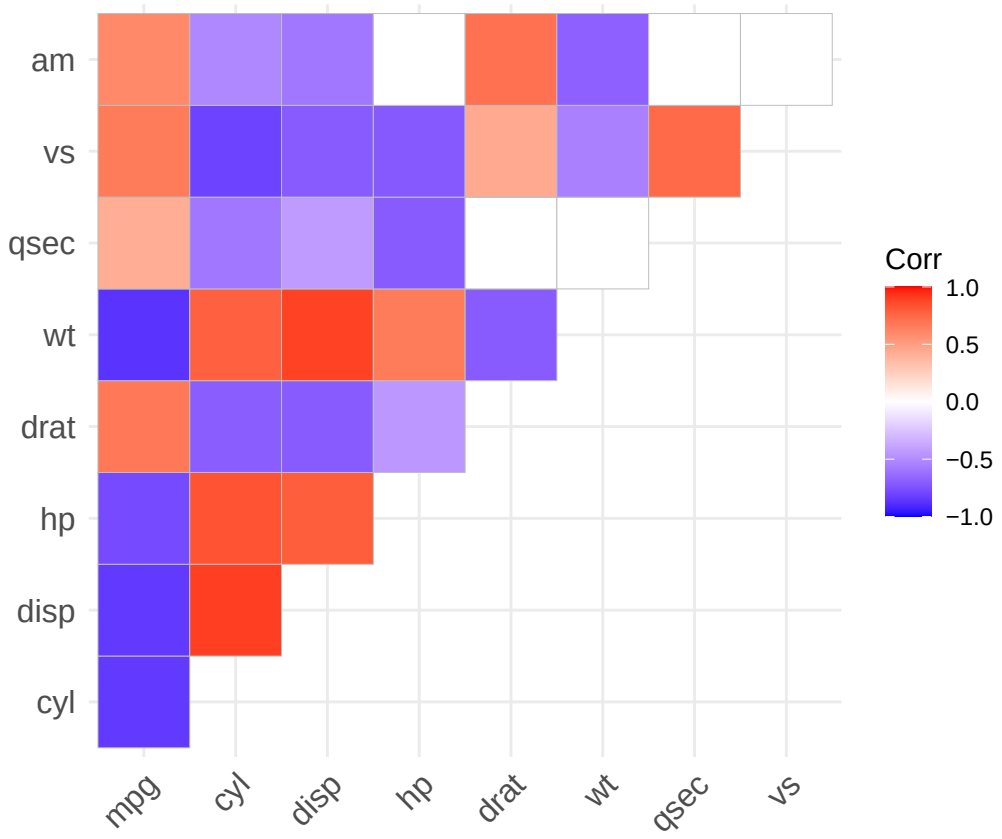
```
corsevp<-cor_pmat(mtcars[,1:9])
corsevp
```

```
##                mpg          cyl         disp           hp         drat
## mpg  0.000000e+00 6.112687e-10 9.380327e-10 1.787835e-07 1.776240e-05
## cyl  6.112687e-10 0.000000e+00 1.802838e-12 3.477861e-09 8.244636e-06
## disp 9.380327e-10 1.802838e-12 0.000000e+00 7.142679e-08 5.282022e-06
## hp   1.787835e-07 3.477861e-09 7.142679e-08 0.000000e+00 9.988772e-03
## drat 1.776240e-05 8.244636e-06 5.282022e-06 9.988772e-03 0.000000e+00
## wt   1.293959e-10 1.217567e-07 1.222320e-11 4.145827e-05 4.784260e-06
## qsec 1.708199e-02 3.660533e-04 1.314404e-02 5.766253e-06 6.195826e-01
## vs   3.415937e-05 1.843018e-08 5.235012e-06 2.940896e-06 1.167553e-02
## am   2.850207e-04 2.151207e-03 3.662114e-04 1.798309e-01 4.726790e-06
##                wt         qsec           vs           am
## mpg  1.293959e-10 1.708199e-02 3.415937e-05 2.850207e-04
## cyl  1.217567e-07 3.660533e-04 1.843018e-08 2.151207e-03
## disp 1.222320e-11 1.314404e-02 5.235012e-06 3.662114e-04
## hp   4.145827e-05 5.766253e-06 2.940896e-06 1.798309e-01
## drat 4.784260e-06 6.195826e-01 1.167553e-02 4.726790e-06
## wt   0.000000e+00 3.388683e-01 9.798492e-04 1.125440e-05
## qsec 3.388683e-01 0.000000e+00 1.029669e-06 2.056621e-01
## vs   9.798492e-04 1.029669e-06 0.000000e+00 3.570439e-01
## am   1.125440e-05 2.056621e-01 3.570439e-01 0.000000e+00
```

```
ggcorrplot(corsev, p.mat=corsevp, type="upper")
```



```
ggcorrplot(corsev, p.mat=corsevp, type="upper",insig = "blank")
```




```
?describe
```