Script.R

bartoszwiernicki

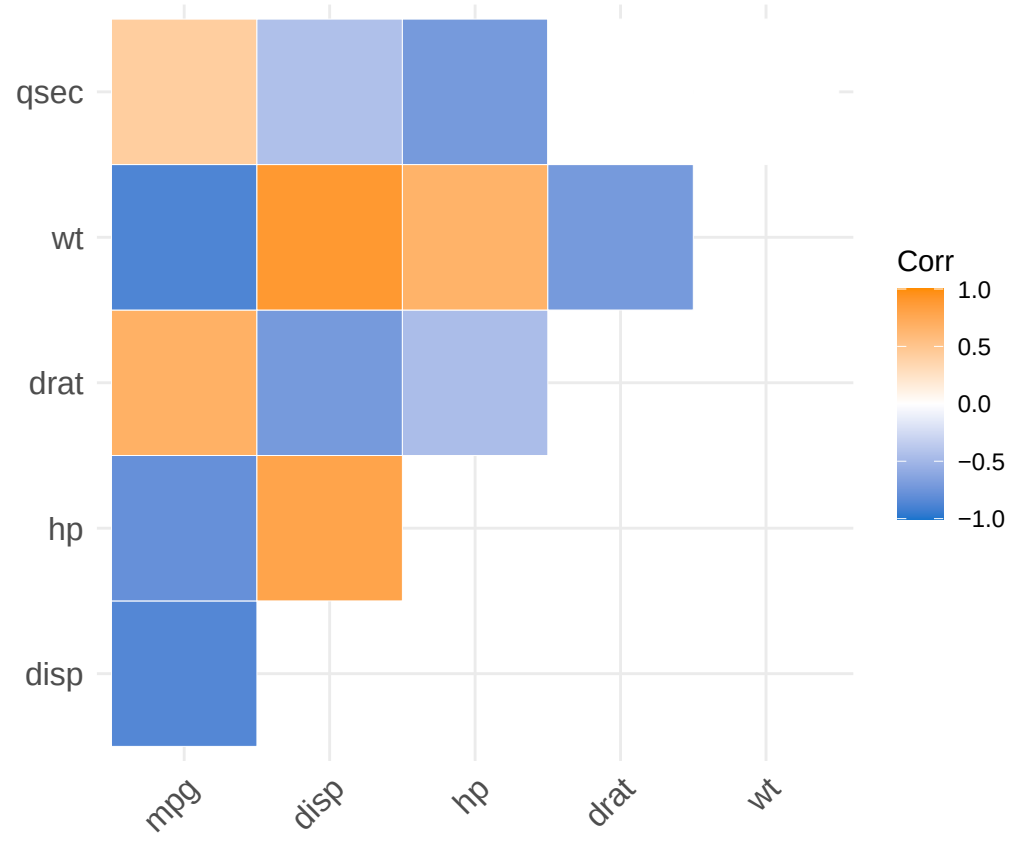
2023-08-21

```
# Loading the libraries ----
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(ggplot2)
library(ggpubr)
library(ggcorrplot)

# Load the data
data("mtcars")

# Select only numeric columns. In this example some of the columns were removed from the analysis
mtcars <- mtcars %>%  select_if(is.numeric) %>%
  select(-c(cyl, vs, am, gear, carb))

# Create correlations matrix and the p.value matrix using the ggcorrplot package
correlations <- cor(mtcars)
p.values.correlations <- cor_pmat(mtcars)


# Create the correlogram
correlogram <-
ggcorrplot(correlations,
             method='square',
             outline.col = "white",
             type='upper',
             sig.level = 0.05,
             insig = 'blank',
             colors = c("dodgerblue3", "white", "darkorange"),
             p.mat = p.values.correlations) +
  theme(axis.text.x = element_text(angle=45, hjust=1)) +
  labs(x=NULL, y=NULL)
```

correlogram



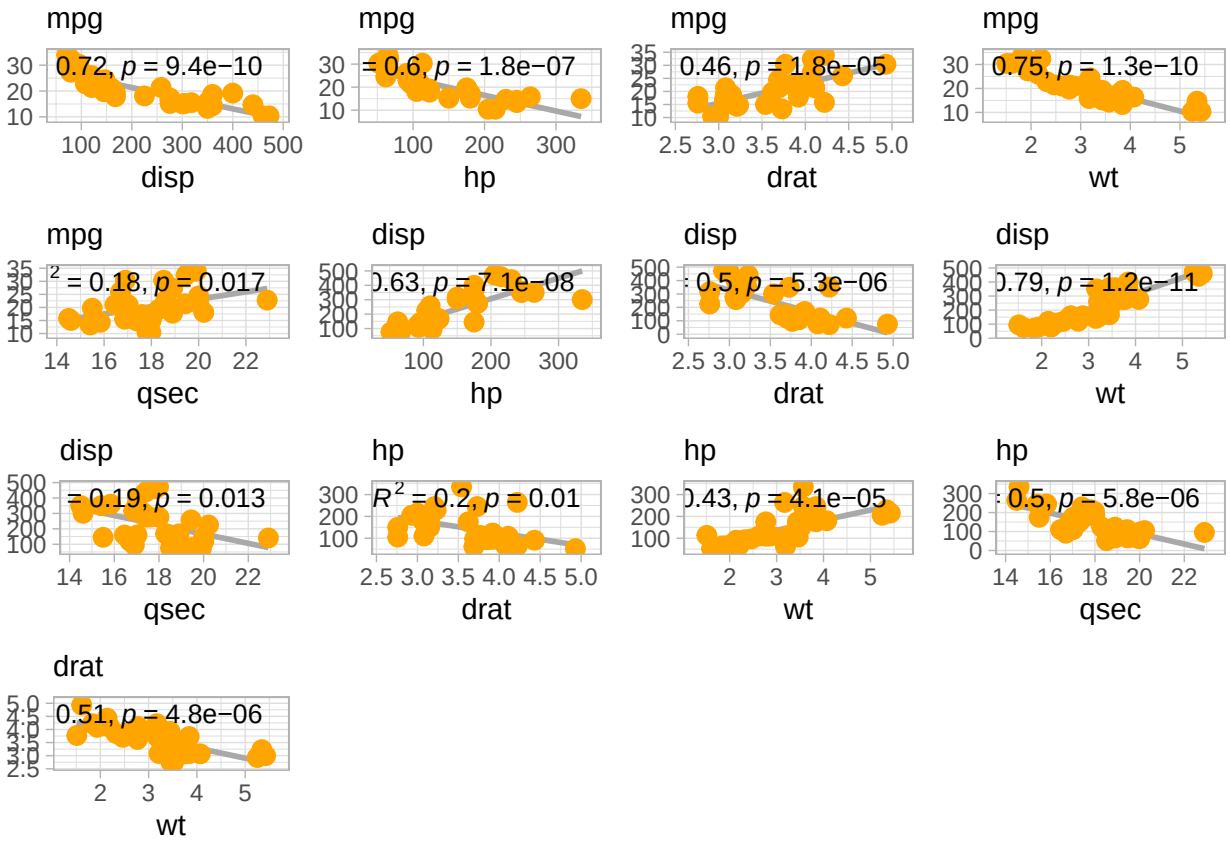

```r
# Correlation plots ----

# Change the p.value matrix into data frame
p.values.correlations <- as.data.frame(p.values.correlations)

# Create a long dataframe with the p.values
p.values.correlations.sel <-
  p.values.correlations %>%
  mutate(another = rownames(.)) %>%
  pivot_longer(cols = -another,
               values_to = 'pvalue',
               names_to = 'first') %>%
  filter(pvalue !=0) %>%
  filter(pvalue <=.05) %>%
  mutate(combined = str_c(first, another, sep = '_')) %>%
  mutate(sorted_words = sapply(strsplit(combined, "_"),
                               function(x) paste(sort(x), collapse = "_"))) %>%
  distinct(sorted_words, .keep_all = T) %>%
  select(-combined, -sorted_words)

# Function to create regression plots
my_plot <- function(source, df) {
  x_var <- paste0(p.values.correlations.sel.item[1,2])
  y_var <- paste0(p.values.correlations.sel.item[1,1])
```

```
  ggplot(data = df, aes(x=.data[[x_var]], y=.data[[y_var]]))+
    geom_smooth(aes(x=.data[[x_var]], y=.data[[y_var]]), method = 'lm', se=FALSE, color='darkgray', for
    geom_point(size=3, color='orange') +
    theme_light()+
    #my_theme()+
    theme(
          legend.position = 'right',
          legend.title = element_blank()) +
    scale_y_continuous(expand = c(0.1, 0.1)) +
    scale_x_continuous(expand = c(0.1, 0.1)) +
    labs(title = NULL,
         subtitle = y_var, y=NULL) +
    stat_cor(aes(label = paste(after_stat(rr.label), after_stat(p.label), sep = "*`,`~")),
             label.y= Inf, label.x = Inf, vjust = 1, hjust = 1.1, size = 3.5)
}

# Creating a list of plots
plots.corr <- list()
for (i in (1:nrow(p.values.correlations.sel))){
  p.values.correlations.sel.item <- p.values.correlations.sel[i,]

  plot_i <-
  my_plot(source = p.values.correlations.sel.item, df = mtcars)

  plots.corr[[paste0(unique(p.values.correlations.sel.item$another), '_',
              unique(p.values.correlations.sel.item$first))]] <- plot_i
}

# Saving the plots to the disc.
plot.names <- names(plots.corr)
lapply(seq_along(plots.corr), function(x) {
  plot <- plots.corr[[x]]
  name <- plot.names[x]

  ggsave(plot = plot,
         filename = paste0('./Plots/Correlations/',name, '.png'),
         width = 80,
         height = 80, units = 'mm')
})
```

```
## [[1]]
## [1] "./Plots/Correlations/mpg_disp.png"
##
## [[2]]
## [1] "./Plots/Correlations/mpg_hp.png"
##
## [[3]]
## [1] "./Plots/Correlations/mpg_drat.png"
##
## [[4]]
## [1] "./Plots/Correlations/mpg_wt.png"
##
## [[5]]
## [1] "./Plots/Correlations/mpg_qsec.png"
```

```
##
## [[6]]
## [1] "./Plots/Correlations/disp_hp.png"
##
## [[7]]
## [1] "./Plots/Correlations/disp_drat.png"
##
## [[8]]
## [1] "./Plots/Correlations/disp_wt.png"
##
## [[9]]
## [1] "./Plots/Correlations/disp_qsec.png"
##
## [[10]]
## [1] "./Plots/Correlations/hp_drat.png"
##
## [[11]]
## [1] "./Plots/Correlations/hp_wt.png"
##
## [[12]]
## [1] "./Plots/Correlations/hp_qsec.png"
##
## [[13]]
## [1] "./Plots/Correlations/drat_wt.png"
```

```r
# Arranging the plots
regressions <-
ggarrange(plotlist = plots.corr)

regressions
```


mpg
0.72, p = 9.4e−10
disp
mpg
= 0.6, p = 1.8e−07
hp
mpg
0.46, p = 1.8e−05
drat
mpg
0.75, p = 1.3e−10
wt
mpg
= 0.18, p = 0.017
qsec
disp
0.63, p = 7.1e−08
hp
disp
0.5, p = 5.3e−06
drat
disp
0.79, p = 1.2e−11
wt
disp
= 0.19, p = 0.013
qsec
hp
R2 = 0.2, p = 0.01
drat
hp
0.43, p = 4.1e−05
wt
hp
0.5, p = 5.8e−06
qsec
drat
0.51, p = 4.8e−06
wt