2208-351-351m-451-w07a-p1-Correlation-PairwisePlots-Code.R

frenchrh

2022-10-11

```r
# 2008-351-351m-451-w06b-p3-PairwiseCorrelationExample.R
# R. H. French, Sameera Naline Venkat, Raymond Weiser, license CC by SA 4.0


# Library in the package and check the documentation

library('psych')
?psych
?psych::pairs.panels

# You can also use the ggpairs function in the GGally package

library(GGally)
```

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

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

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

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
?GGally::ggpairs

# We'll use Fischer's Iris Dataset
# This dataset comes in base-r

?iris
```
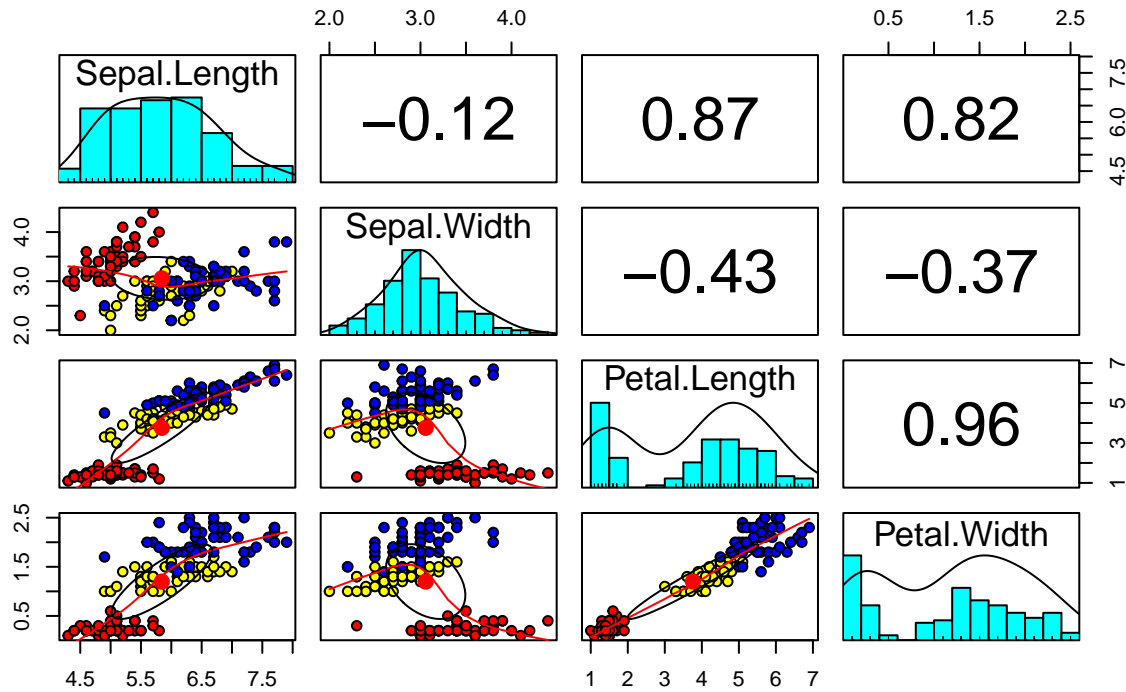

```r
# Lets look at pairwise correlation among the variables
# A standard first step in EDA

pairs.panels(
  iris[1:4],
  bg = c("red", "yellow", "blue")[iris$Species],
  pch = 21,
  main = "Fisher Iris data by Species"
)
```

Fisher Iris data by Species


```
# From TidyVerse/Google R Style Guide
#   Spacing
#     Place spaces around all binary operators (=, +, -, <-, etc.).
#       Exception: Spaces around ='s are optional when passing parameters in a function call.
#     Do not place a space before a comma,
#       but always place one after a comma
```
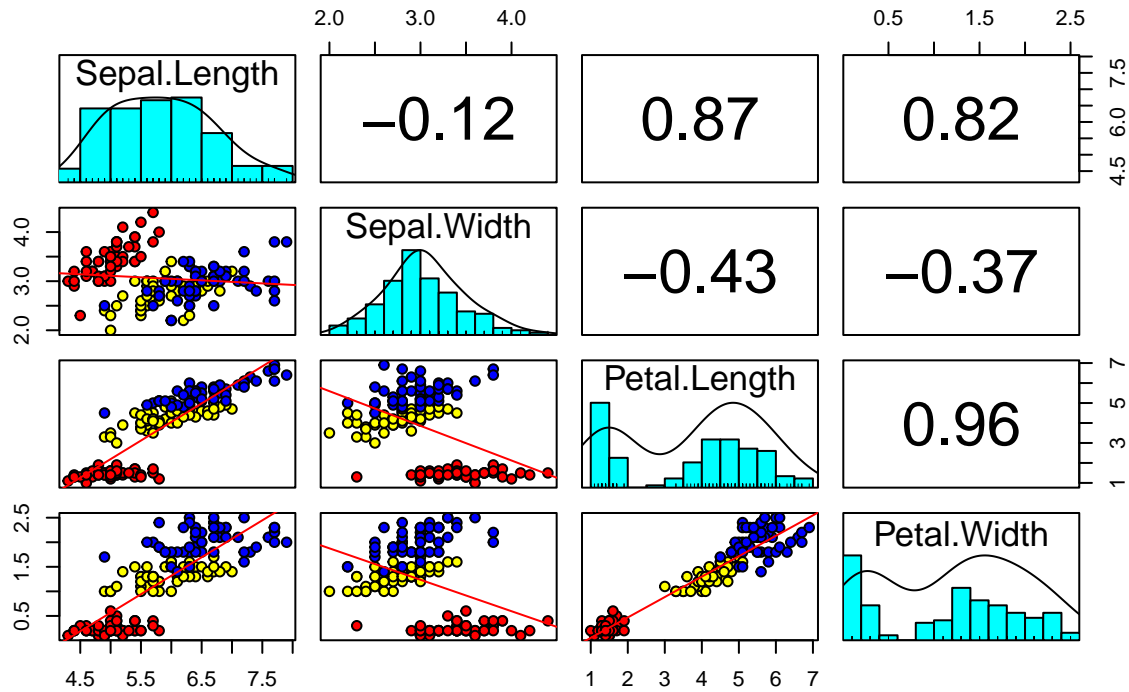
```
# We don't like the ellipse or the red dot.
# Not useful for our work.

pairs.panels(
  iris[1:4],
  bg = c("red", "yellow", "blue")[iris$Species],
  pch = 21,
  main = "Fisher Iris data by Species",
  smooth = FALSE,
  lm = TRUE,
  ellipses = FALSE
)
```

Fisher Iris data by Species


```
# But the lm fit line, misleads your eye, and misinforms us
# Lets get rid of it. with lm = FALSE

pairs.panels(
  iris[1:4],
  bg = c("red", "yellow", "blue")[iris$Species],
  pch = 21,
  main = "Fisher Iris data by Species",
  smooth = FALSE,
  lm = FALSE,
  ellipses = FALSE
)
```
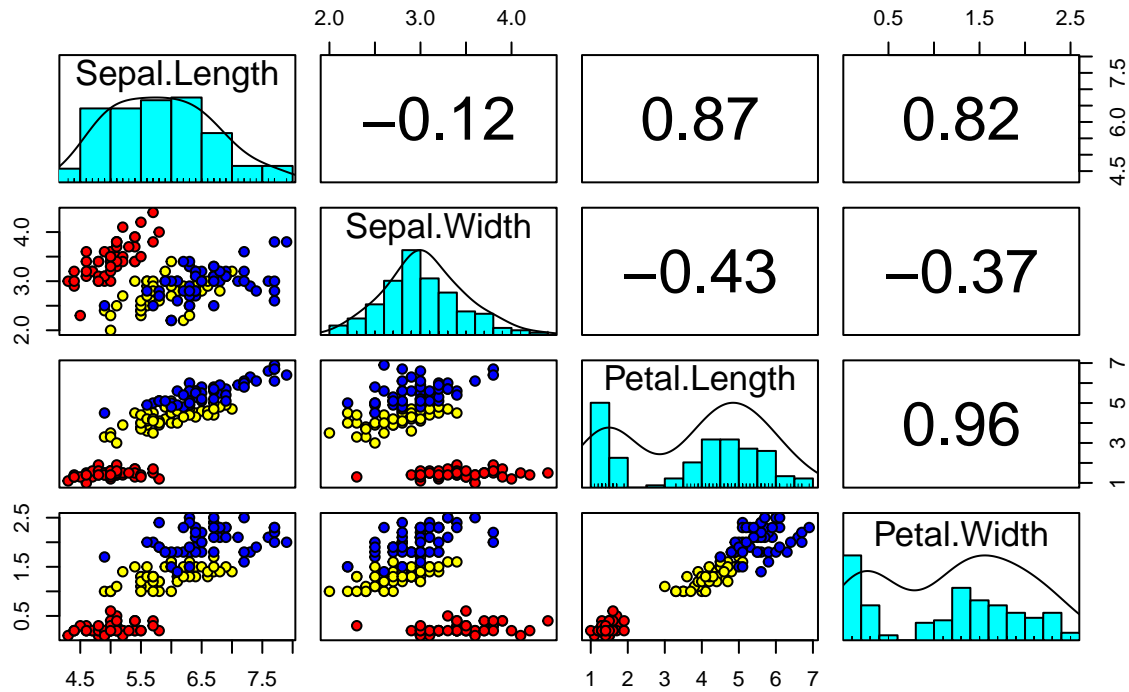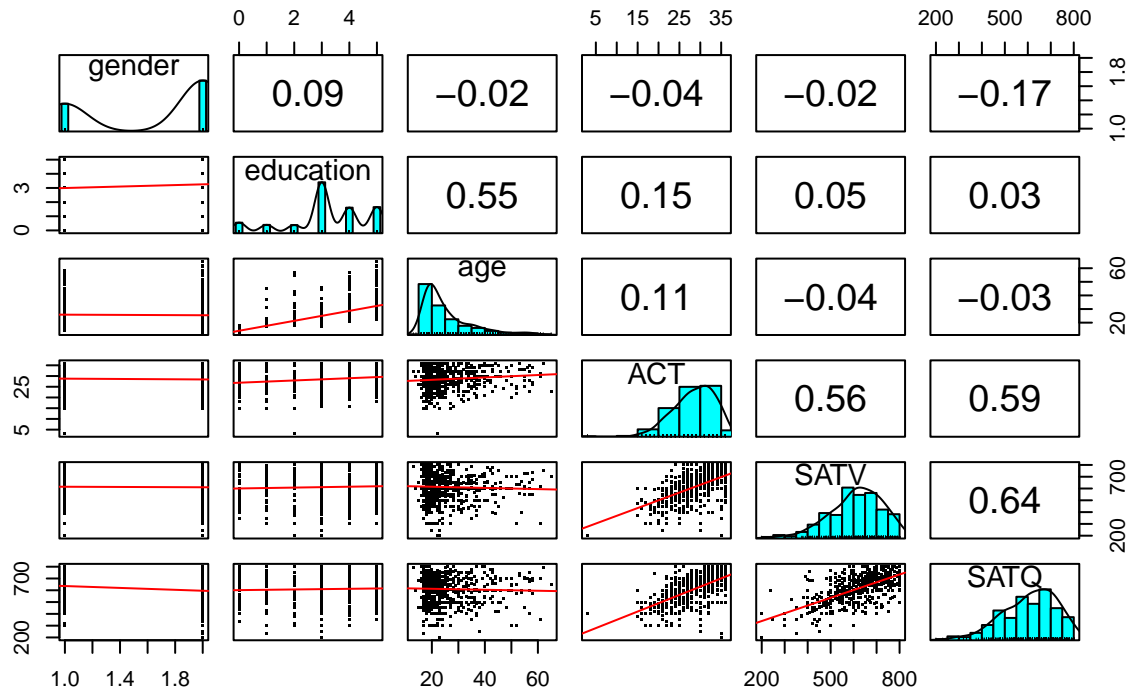
Fisher Iris data by Species


```
# Looks Better

# Now Lets look at a new dataset of self-reported SAT scores
# This data comes in the psych package

data(sat.act)
?sat.act

pairs.panels(
  sat.act,
  pch = ".",
  main = "Self-reported SAT scores",
  smooth = FALSE,
  lm = TRUE,
  ellipses = FALSE
)
```
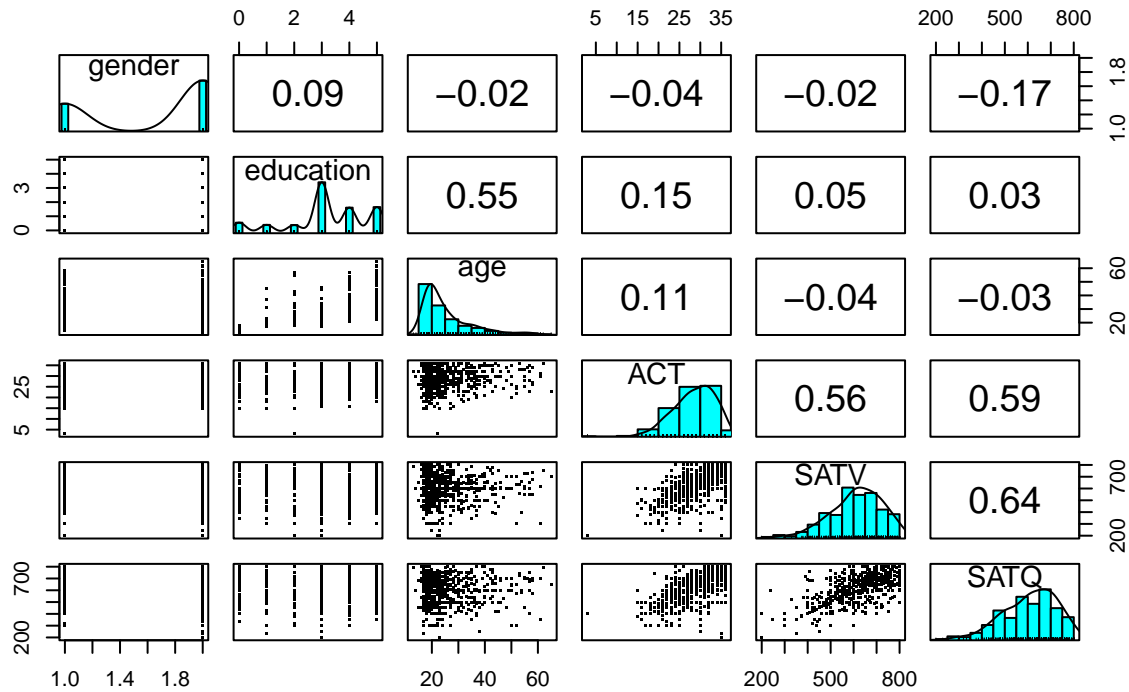
Self−reported SAT scores


```r
# by default pairs.panel uses the Pearson Correlation Coefficient
# its set in the method parameter

pairs.panels(
  sat.act,
  pch = ".",
  method = 'pearson',
  main = "Pearson Correlations: Self-reported SAT scores",
  smooth = FALSE,
  lm = FALSE,
  ellipses = FALSE
)
```
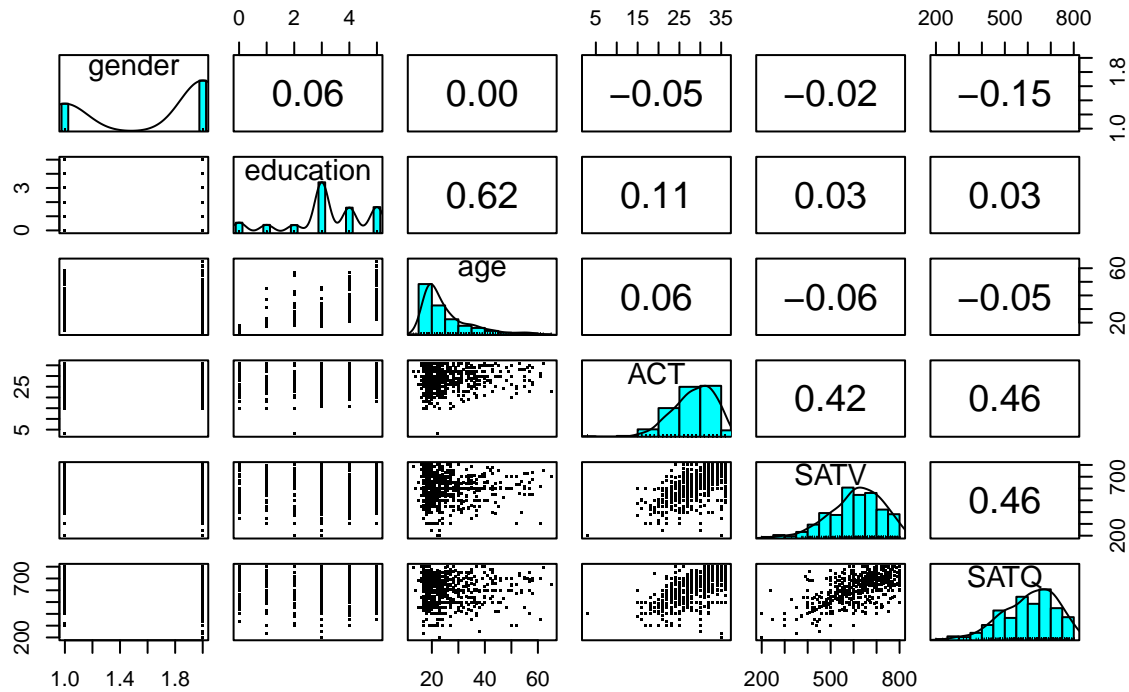
Pearson Correlations: Self−reported SAT scores


```
# Now plot the Kendal correlations

pairs.panels(
  sat.act,
  pch = ".",
  method = 'kendal',
  main = "Kendal Correlations: Self-reported SAT scores",
  smooth = FALSE,
  lm = FALSE,
  ellipses = FALSE
)
```

Kendal Correlations: Self-reported SAT scores

	gender		education		age		ACT		SATV		SATQ
gender		0.06		0.00		−0.05		−0.02		−0.15	
education				0.62		0.11		0.03		0.03	
age						0.06		−0.06		−0.05	
ACT								0.42		0.46	
SATV										0.46	
SATQ											

```r
# Now plot the Spearman correlations

pairs.panels(
  sat.act,
  pch = ".",
  method = 'spearman',
  main = "Spearman Correlations: Self-reported SAT scores",
  rug = TRUE,
  smooth = FALSE,
  lm = FALSE,
  ellipses = FALSE
)
```

Spearman Correlations: Self−reported SAT scores
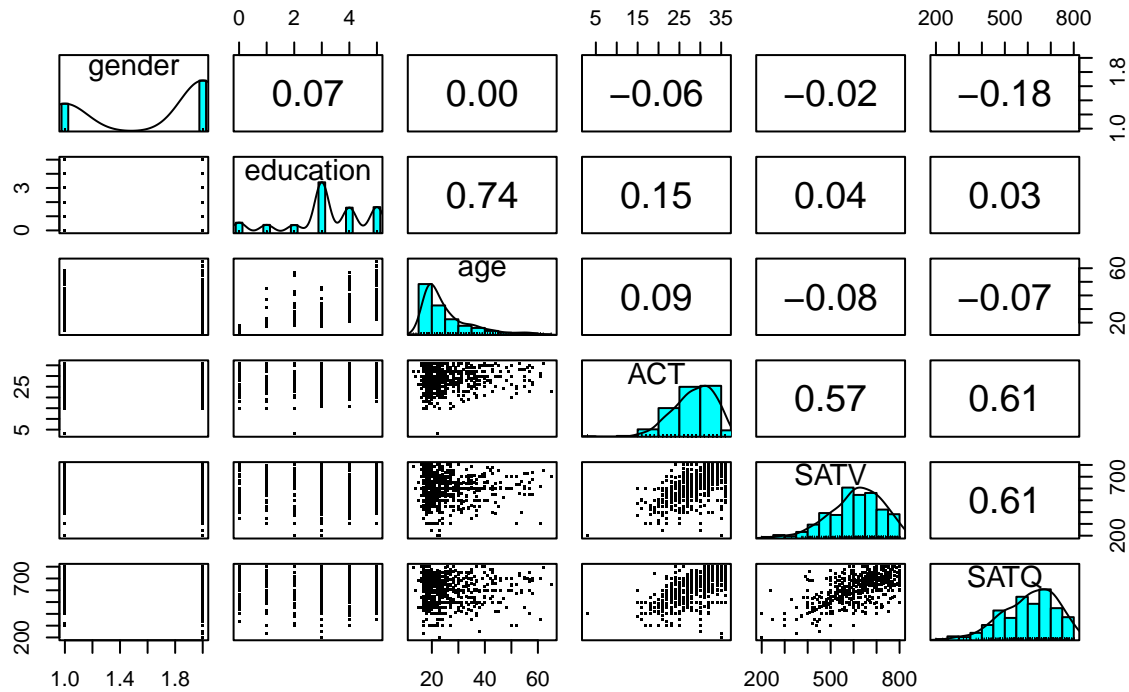

```
# You can learn more about Correlation coefficients
# https://en.wikipedia.org/wiki/Correlation_and_dependence


# cite: Dr. Rojiar Haddadian, SDLE Research Center, CWRU
```