



R news and tutorials contributed by (573) R bloggers

- [Home](#)
- [About](#)
- [RSS](#)
- [add your blog!](#)
- [Learn R](#)
- [R jobs](#)
- [Contact us](#)

Welcome!

Follow @rbloggers {28K fo

Here you will find daily **news and tutorials about R**, contributed by over 573 bloggers.

There are many ways to **follow us -**

[By e-mail:](#)

Your e-mail here

 Subscribe

 25169 readers

BY FEEDBURNER

[On Facebook:](#)

R blogg...

 30K likes

 Like Page

 Be the first of your friends to like this

If you are an R blogger yourself you are invited to [add your own R content feed to this site](#) (Non-English R bloggers should add themselves- [here](#))

[Jobs for R-users](#)

- [Senior Statistician/Statistician \(>\\$100/year\) @ Shinjuku-ku, Tōkyō-to, Japan](#)
- [Junior Data Scientist @ Farnham, England](#)
- [Vehicle Valuations Manager @ Farnham, England](#)
- [Manager – Quantitative Analytics @ London, United Kingdom](#)
- [Pharmacometrics](#)

[Bootcamp @ Wellesley, Massachusetts, US](#)

Popular Searches

- [web scraping](#)
- [heatmap](#)
- [maps](#)
- [shiny](#)
- [twitter](#)
- [alt=](#)
- [boxplot](#)
- [hadoop](#)
- [time series](#)
- [animation](#)
- [ggplot2](#)
- [trading](#)
- [latex](#)
- [Ggplot](#)
- [PCA](#)
- [excel](#)
- [finance](#)
- [quantmod](#)
- [RStudio](#)
- [googlevis](#)
- [rattle](#)
- [how to import image file to R](#)
- [market research](#)
- [eclipse](#)
- [knitr](#)
- [rmdr](#)
- [coplot](#)
- [tutorial](#)
- [SQL](#)
- [map](#)

Recent Posts

- [The Unbearable Insolence of Prime Numbers or \(Playing to be Ulam\)](#)
- [23 new R jobs from around the world \(2016-02-10\)](#)
- [Finding the K in K-means by Parametric Bootstrap](#)
- [In case you missed it: January 2016 roundup](#)
- [Analysis: Clinton backed by Big Money; Sanders by Small](#)
- [Craft http calls cleverly with curlconverter](#)
- [The Easiest Way to Learn ggplot2](#)
- [Data Visualization Course Notes](#)
- [Clusters of Texts](#)
- [Durban R Users](#)

[Group Meetup: 24](#)

[February 2016 @](#)

[The Green Door](#)

- [Clusters of \(French\) Regions](#)
- [New Release of RStudio \(v0.99.878\)](#)
- [An Introduction to Time Series with JSON Data](#)
- [EARL London 2016 – Call for Abstracts](#)
- [Using SVG graphics in blog posts](#)

Other sites

- [Statistics of Israel](#)
- [SAS blogs](#)
- [Jobs for R-users](#)

More on Exploring Correlations in R

August 28, 2012

By [Stephen Turner](#)

Like Share 18 Tweet Share 14

(This article was first published on [Getting Genetics Done](#), and kindly contributed to [R-bloggers](#))

About a year ago I wrote a post about producing [scatterplot matrices in R](#). These are handy for quickly getting a sense of the correlations that exist in your data. Recently someone asked me to pull out some relevant statistics (correlation coefficient and p-value) into tabular format to publish beside a scatterplot matrix. The built-in `cor()` function will produce a correlation matrix, but what if you want p-values for those correlation coefficients? Also, instead of a matrix, how might you get these statistics in tabular format (variable i , variable j , r , and p , for each i - j combination)? Here's the code (you'll need the `PerformanceAnalytics` package to produce the plot).

```

1  ## Correlation matrix with p-values. See http://goo.gl/nahmV f
2  cor.prob <- function (X, dfr = nrow(X) - 2) {
3    R <- cor(X, use="pairwise.complete.obs")
4    above <- row(R) < col(R)
5    r2 <- R[above]^2
6    Fstat <- r2 * dfr / (1 - r2)
7    R[above] <- 1 - pf(Fstat, 1, dfr)
8    R[row(R) == col(R)] <- NA
9    R
10 }
11
12 ## Use this to dump the cor.prob output to a 4 column matrix
13 ## with row/column indices, correlation, and p-value.
14 ## See StackOverflow question: http://goo.gl/fCUcQ
15 flattenSquareMatrix <- function(m) {
16   if( (class(m) != "matrix") | (nrow(m) != ncol(m))) stop("Mus
17   if(!identical(rownames(m), colnames(m))) stop("Row and colum
18   ut <- upper.tri(m)
19   data.frame(i = rownames(m)[row(m)[ut]],
20             j = rownames(m)[col(m)[ut]],
21             cor=t(m)[ut],
22             p=m[ut])
23 }
24
```

```

25 # get some data from the mtcars built-in dataset
26 mydata <- mtcars[, c(1,3,4,5,6)]
27
28 # correlation matrix
29 cor(mydata)
30
31 # correlation matrix with p-values
32 cor.prob(mydata)
33
34 # "flatten" that table
35 flattenSquareMatrix(cor.prob(mydata))
36
37 # plot the data
38 library(PerformanceAnalytics)
39 chart.Correlation(mydata)

```

explore-correlations.r hosted with ❤ by GitHub [view raw](#)

The `cor()` function will produce a basic correlation matrix. 12 years ago [Bill Venables provided a function on the R help mailing list](#) for replacing the upper triangle of the correlation matrix with the p-values for those correlations (based on the known relationship between t and r). The `cor.prob()` function will produce this matrix.

Finally, the `flattenSquareMatrix()` function will “flatten” this matrix to four columns: one column for variable i , one for variable j , one for their correlation, and another for their p-value (thanks to [Chris Wallace on StackOverflow](#) for helping out with this one).

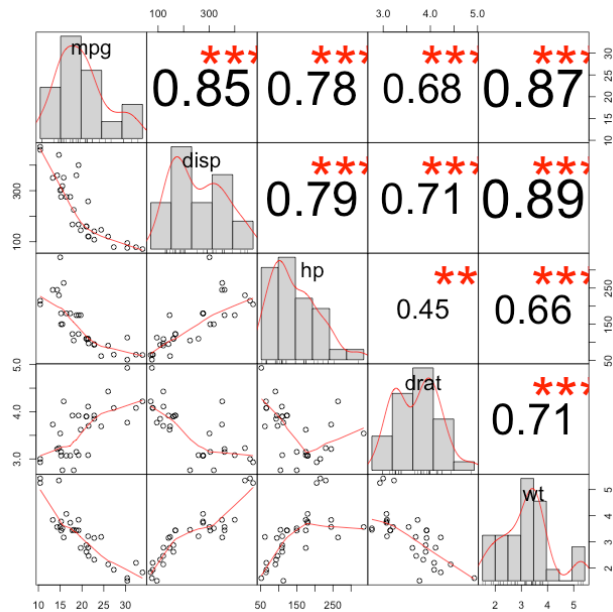
```

1 > cor(mydata)
2           mpg      disp      hp      drat      wt
3 mpg  1.0000000 -0.8475514 -0.7761684  0.6811719 -0.8676594
4 disp -0.8475514  1.0000000  0.7909486 -0.7102139  0.8879799
5 hp   -0.7761684  0.7909486  1.0000000 -0.4487591  0.6587479
6 drat  0.6811719 -0.7102139 -0.4487591  1.0000000 -0.7124406
7 wt   -0.8676594  0.8879799  0.6587479 -0.7124406  1.0000000
8
9 > cor.prob(mydata)
10          mpg      disp      hp      drat
11 mpg      NA  9.380327e-10  1.787835e-07  1.776240e-05  1.29
12 disp -0.8475514      NA  7.142679e-08  5.282022e-06  1.22
13 hp   -0.7761684  7.909486e-01      NA  9.988772e-03  4.14
14 drat  0.6811719 -7.102139e-01 -4.487591e-01      NA  4.78
15 wt   -0.8676594  8.879799e-01  6.587479e-01 -7.124406e-01
16
17 > flattenSquareMatrix(cor.prob(mydata))
18      i  j      cor      p
19 1  mpg disp -0.8475514 9.380327e-10
20 2  mpg  hp -0.7761684 1.787835e-07
21 3  disp hp  0.7909486 7.142679e-08
22 4  mpg drat 0.6811719 1.776240e-05
23 5  disp drat -0.7102139 5.282022e-06
24 6   hp drat -0.4487591 9.988772e-03
25 7  mpg  wt -0.8676594 1.293958e-10
26 8  disp  wt  0.8879799 1.222322e-11
27 9   hp  wt  0.6587479 4.145827e-05
28 10 drat  wt -0.7124406 4.784260e-06

```

explore-correlations-output.txt hosted with ❤ by GitHub [view raw](#)

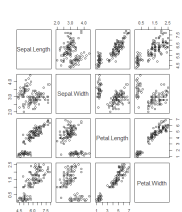
Finally, the `chart.Correlation()` function from the `PerformanceAnalytics` package produces a very nice scatterplot matrix, with histograms, kernel density overlays, absolute correlations, and significance asterisks (0.05, 0.01, 0.001):



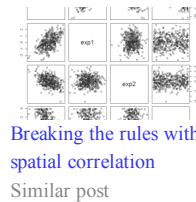
Getting Genetics Done by Stephen Turner is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.

Like Share 18 Tweet Share 14

Related



y_1, \dots, y_n
Why an inverse-Wishart prior may not be such a good idea
In "R bloggers"



Breaking the rules with spatial correlation
Similar post

Scatterplot matrices in R

In "R bloggers"

18 Tweet 14
Like Share

To leave a comment for the author, please follow the link and comment on their blog: [Getting Genetics Done](#).

R-bloggers.com offers [daily e-mail updates](#) about R news and [tutorials](#) on topics such as: [Data science](#), [Big Data](#), [R jobs](#), visualization ([ggplot2](#), [Boxplots](#), [maps](#), [animation](#)), programming ([RStudio](#), [Sweave](#), [LaTeX](#), [SQL](#), [Eclipse](#), [git](#), [hadoop](#), [Web Scraping](#)) statistics ([regression](#), [PCA](#), [time series](#), [trading](#)) and more...

If you got this far, why not [subscribe for updates](#) from the site?
Choose your flavor: [e-mail](#), [twitter](#), [RSS](#), or [facebook](#)...

Like Share 18 Tweet Share 14

Comments are closed.

Search & Hit Enter

Recent popular posts

- [Finding the K in K-means by Parametric Bootstrap](#)
- [The Easiest Way to Learn ggplot2](#)

- [New Release of RStudio \(v0.99.878\)](#)

Most visited articles of the week

1. [Installing R packages](#)
2. [How to Learn R](#)
3. [Using apply, sapply, lapply in R](#)
4. [What has Kaggle learned from 2 million machine learning models?](#)
5. [In-depth introduction to machine learning in 15 hours of expert videos](#)
6. [How to Make a Histogram with Basic R](#)
7. [New Release of RStudio \(v0.99.878\)](#)
8. [Scatterplots](#)
9. [Computing and visualizing PCA in R](#)

Sponsors




MANGOSOLUTIONS
data analysis that delivers




Leading the way in
R Consulting, Training,
Data Analytics &
Application
Development

Learn R By Doing
www.DataCamp.com

Start Free Course



Free, open-source **D3.js** charting
for **RStudio** and **Shiny**



[Plotly: collaborative, publication-quality graphing.](#)



DATA SCIENCE
Certification
Training
(R, SAS & Excel)

GET CERTIFIED
simplilearn

R training
R consulting



QUANTIDE
knowledge from data


New in RStudio Server Pro
Collaboration | Multiple Sessions | R Versions



Free Evaluation

Intro and advanced
R courses for just \$19!


Redeem Promocode



udemy

Highland Statistics Ltd

Zero Inflated Models & GLMM
Beginner's Guide to GAM
Beginner's Guide to GLM & GLMM
Beginner's Guide to GAMM



Werden Sie zum Expe[R]ten mit der
R-Akademie von



eoda
daten » wissen » nutzen

Beratung | Software
Training | Lösungen



STATISTICS
VIEWS

Bringing Statistics Together

BECOME A DATA
SCIENTIST



NYC DATA SCIENCE
ACADEMY

12 Weeks to A New Career

#ODSC

BOSTON MAY 20-22


TICKETS ON SALE NOW

Download a FREE Chapter Today!

Parallel Computing for Data Science
With Examples in R, C++ and CUDA

SAVE 25%
on All R Books

Promo Code
CZP40



CRC Press
Taylor & Francis Group
www.crcpress.com

STATWORX

Consulting
Schulung
Data Mining



Mehr erfahren





[Contact us](#) if you wish to help support R-bloggers, and place your banner here.

[Jobs for R users](#)

- [Senior Statistician/Statistician \(>\\$100/year\) @ Shinjuku-ku, Tōkyō-to, Japan](#)
- [Junior Data Scientist @ Farnham, England](#)
- [Vehicle Valuations Manager @ Farnham, England](#)
- [Manager – Quantitative Analytics @ London, United Kingdom](#)
- [Pharmacometrics Bootcamp @ Wellesley, Massachusetts, US](#)
- [Data Scientist for TIBCO \(>\\$100K/year\)](#)
- [Senior Financial Analyst – Consumer Marketing Analytics @ Seattle, US](#)

Search & Hit Enter

[Full list of contributing R-bloggers](#)

[R-bloggers](#) was founded by [Tal Galili](#), with gratitude to the [R](#) community.

Is powered by [WordPress](#) using a [bayotasan.com](#) design.

Copyright © 2016 **R-bloggers**. All Rights Reserved. [Terms and Conditions](#) for this website