

R-bloggers

R news and tutorials contributed by (376) R bloggers

- Home
- About
- add your blog!
- Contact us
- RSS

WELCOME!

Here you will find daily **news and tutorials about R**, contributed by over 300 bloggers. You can subscribe for e-mail updates:

8468 readers

BY FEEDBURNER

And get updates to your Facebook: **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](#))

TOPICS OF INTEREST

ggplot2 r-project graphics programming rstats R Language Books visualization Data R-english events tutorial LaTeX finance Big Data packages sweave computing stocks R Programming Software applications Bioinformatics Simulation ggplot "R" Chemometrics Python Rmedia BMS Extensions plot twitter English cran random stats blog Quant finance Tutorials lattice SAS linux Announcements rblogs Climate Trends Probability Rcpp science video R bloggers predictive analytics

RECENT POSTS

Fun with the googleVis Package for R
Coefficient Plots in R
Simple and heuristic optimization
igraph and SNA: an

Merging Multiple Data Files into One Data Frame

April 24, 2011

By [Tony Cookson](#)

[Like](#) [2 likes](#). [Sign Up](#) to see what your friends like.

(This article was first published on [Coffee and Econometrics in the Morning](#), and kindly contributed to [R-bloggers](#))

We often encounter situations where we have data in multiple files, at different frequencies and on different subsets of observations, but we would like to match them to one another as completely and systematically as possible. In R, the `merge()` command is a great way to match *two* data frames together.

Just read the two data frames into R

```
mydata1 = read.csv(path1, header=T)
mydata2 = read.csv(path2, header=T)
```

Then, merge

```
myfulldata = merge(mydata1, mydata2)
```

As long as `mydata1` and `mydata2` have at least one common column with an identical name (that allows matching observations in `mydata1` to observations in `mydata2`), this will work like a charm. It also takes three lines.

What if I have 20 files with data that I want to match observation-to-observation? Assuming they all have a common column that allows merging, I would still have to read 20 files in (20 lines of code) and `merge()` works two-by-two... so I could merge the 20 data frames together with 19 merge statements like this:

```
mytempdata = merge(mydata1, mydata2)
mytempdata = merge(mytempdata, mydata3)
.
.
.
mytempdata = merge(mytempdata, mydata20)
```

That's tedious. You may be looking for a simpler way. If you are, I wrote a function to solve your woes called `multmerge()`. * Here's the code to define the function:

```
multmerge = function(mypath){
  filenames=list.files(path=mypath, full.names=TRUE)
  datalist = lapply(filenames,
    function(x){read.csv(file=x,header=T)})
  Reduce(function(x,y) {merge(x,y)}, datalist)
```

After running the code to define the function, you are all set to use it. The function takes a path. This path should be the name of a folder that contains **all of the files** you would like to read and merge together **and only those files you would like to merge**. With this in mind, I have two tips:

TOP 7 ARTICLES OF THE WEEK

1. The HUGE edge of Matlab over R is the Matlab Coder Toolbox. Converting R script code to C++/C?
2. Access data quickly and easily: `data.table` package
3. R 2.15.1 includes performance improvements inspired by `dataframe` package
4. Updating R but keeping your installed packages
5. Bayesian Nonparametrics in R
6. Why R is Hard to Learn
7. How do I Create the Identity Matrix in R?

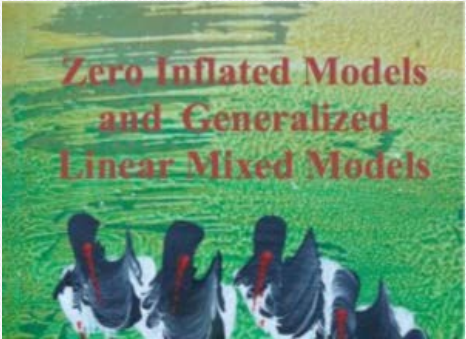
SPONSORS

| | ||| | | |||

MANGOSOLUTIONS

data analysis that delivers

[mango-solutions.com](#)



Zero Inflated Models and Generalized Linear Mixed Models with R. Zuur, Saveliev, Ieno (2012).

amateur's dabbling
SYTYCD — where are
these terrific dancers
come from?
Wrap-up on Blogging
with R Markdown and
tumblr
igraph and structured
text exploration
What kind of profitable
financial forecasting
models and algorithms
do R professional users
focus on?
Trying for a baby?
Here's how long it might
take.
Horizon Plot Already
Available
multiplicitiesoffreedom
demonstrates Chaos
Theory in Excel. If...
Bayesian credible
intervals in the
mainstream medical
literature
Impact of hardware
characteristics on
detectable fault behavior
Decline Effect in
Linguistics?
The inner workings of R
objects

OTHER SITES

Forex blogs
SAS blogs

POPULAR SEARCHES

ggplot
RStudio
hadoop
ggplot2
googleVis
heatmap
quantmod
trading
sweave
animation
Web Scraping
PCA
finance
knitr
map
maps
Tutorial
Excel
boxplot
Latex
GIS
Rcmdr
eclipse
twitter
ecdf
SQL
time series
ANOVA

1. Before you use this function, my suggestion is to create a new folder in a short directory (for example, the path for this folder could be "C://R//mergeme") and save all of the files you would like to merge in that folder.
2. In addition, make sure that the column that will do the matching is formatted the same way (and has the same name) in each of the files.

Suppose you saved your 20 files into the mergeme folder at "C://R//mergeme" and you would like to read and merge them. To use my function, you use the following syntax:


```
mymergeddata = multmerge("C://R//mergeme")
```


After running this command, you have a fully merged data frame with all of your variables matched to each other. Of course, most of the details in matching and merging data come down to making sure that the common column is specified correctly, but given that, this function can save you a lot of typing.

*Maybe a function like this exists out there already, but I think it is entertaining to write helpful functions. I also trust the functions more if I write them myself and test them on my own problems.

To **leave a comment** for the author, please follow the link and comment on his blog: [Coffee and Econometrics in the Morning](#).

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

 Like

 2 likes. Sign Up to see what your friends like.

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

Tags: Data, useful functions

Comments are closed.



Quantide: statistical consulting and training.

regression
Time%20Series

R-bloggers was founded by [Tal Galili](#), with gratitude to the R community.
Is powered by [WordPress](#) using a [bavotasan.com](#) design.
Copyright © 2012 **R-bloggers**. All Rights Reserved.