

Scatterplot Matrices

Posted on January 31, 2013 by moonheadsing in R bloggers | 0 Comments

[This article was first published on [Learning Omics » R](#), and kindly contributed to R-bloggers]. (You can report issue about the content on this page [here](#))

Want to share your content on R-bloggers? [click here](#) if you have a blog, or [here](#) if you don't.

 Share

 Tweet

Scatterplot matrices are a great way to roughly determine if you have a linear correlation between multiple variables. This is particularly helpful in pinpointing specific variables that might have similar correlations to your genomic or proteomic data. If you already have data with multiple variables, load it up as described here.

If not, no worries because R comes with some various presaved datasets for practice (some are more interesting than others. To view these datasets, input the following.

```
data()
```

For this tutorial, we will be looking at the datasets **trees** and **ChickWeight**. First, load or open these datasets.

```
data(trees)
data(ChickWeight)
```

To see the actual data contained by these datasets, just write the title of the dataset.

```
trees
ChickWeight
```

- The **trees** dataset seems to contain three columns of measurements: Girth, Height and Volume.
- The **ChickWeight** dataset seems to involve little chicks getting fed different diets and being weighed at various time points.

To find out more information about the datasets and to confirm our observations, put a question mark before the title of the dataset.

```
?trees
?dataset
```

Now, you ready for the scatterplot?

```
pairs(trees)
```

 Search R-bloggers..

 Your e-mail here

52793 readers

Follow @rbloggers



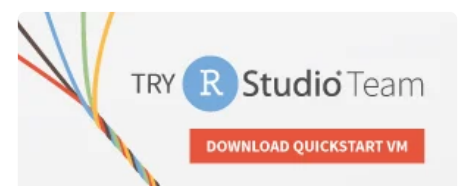
R bloggers

84K followers

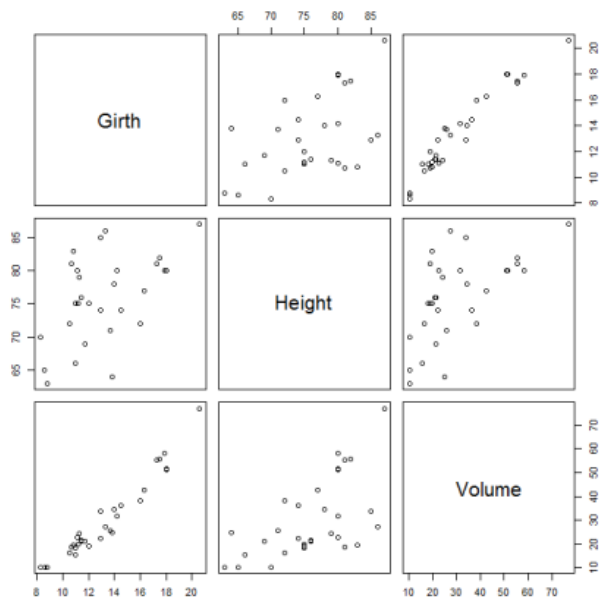
Most viewed posts (weekly)

Which data science skills are important
(\$50,000 increase in salary in 6-months)
PCA vs Autoencoders for Dimensionality Reduction
Better Sentiment Analysis with sentiment.ai
Self-documenting plots in ggplot2
5 Ways to Subset a Data Frame in R
How to write the first for loop in R
How to Calculate a Cumulative Average in R

Sponsors



-->



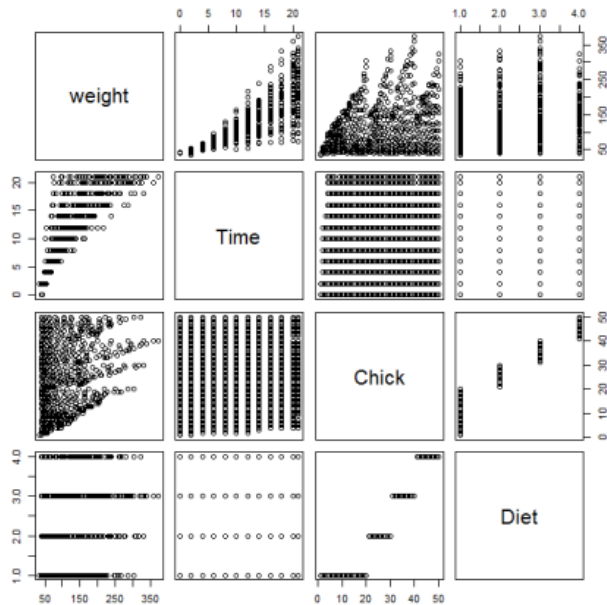
Dataset Trees Scatterplot Matrix

This is an example of a scatterplot matrix. The variables are written in a diagonal line from top left to bottom right. Then each variable is plotted against each other. For example, the middle square in the first column is an individual scatterplot of Girth and Height, with Girth as the X-axis and Height as the Y-axis. This same plot is replicated in the middle of the top row. In essence, the boxes on the upper right hand side of the whole scatterplot are mirror images of the plots on the lower left hand.

In this scatterplot, it is probably safe to say that there is a correlation between Girth and Volume (Go data! Confirming the obvious) because the plot looks like a line. There is probably less of a correlation between Height and Girth in addition to Height and Volume. More statistical analyses would be needed to confirm or deny this.

Now for ChickWeight.

```
pairs(ChickWeight)
```



Dataset ChickWeight Scatterplot

This scatterplot matrix is unfortunately not as clean as the last plot because it contains discrete data points for Time, Chick and Diet. However, much can still be extracted from this scatterplot matrix (think about BS exercises you might have done for English or Art) about experimental design and possible outcomes.

Managed RStudio Infrastructure



STATWORX

Data Science Service

Data Science | Consulting | Development | Training



Save 40% on

MANNING

books with code: nlrblog40

R + 

Learn R (for ecology)

Our ads respect your privacy. Read our Privacy Policy page to learn more.

Contact us if you wish to help support R-bloggers, and place **your banner here**.

- Scatterplots related to Time are evenly distributed into columns or rows, suggesting that data was actually collected in a regimented fashion. (As in, data was collected at the times it should have been for all the Chick samples).
- There were about 50 chicks. The first 20 were on diet 1 and then the next three groups of 10 were given diet 2, 3 or 4.
- Looking at Row 4, Column 1, there is a possibility that chicks on diet 3 gained more weight than chicks on diets 1, 2 or 4.
- Looking at Row 2, Column 1, it seems that chicks weighed about the same amount at the beginning of the experiment but variation increased as time passed on. In general, there is an increase in weight.

There you have it!

In conclusion,

- Scatterplot matrices are **good** for determining rough linear correlations of metadata that contain continuous variables.
- Scatterplot matrices are **not so good** for looking at discrete variables.

Related

Taking a Subset of a Data Frame in R

I just wrote a new chapter for my students describing how to subset a data frame in R. The full text is available at
December 12, 2016
In "R bloggers"

A *simple* introduction to ggplot2 (for plotting your data!)

If you've ever been totally confused by ggplot2 and what it is or how it works, my intention is that this short tutorial simplifies it
April 14, 2022
In "R bloggers"

R Tutorial: Visualizing Multivariate Relationships in Large Datasets

R Tutorial: Visualizing multivariate relationships in Large Datasets A tutorial by D.M. Wiig In two previous blog posts I
February 6, 2017
In "R bloggers"



Share



Tweet

To **leave a comment** for the author, please follow the link and comment on their blog:

[Learning Omics » R.](#)

R-bloggers.com offers **daily e-mail updates** about R news and tutorials about [learning R](#) and many other topics. [Click here if you're looking to post or find an R/data-science job.](#)

Want to share your content on R-bloggers? [click here](#) if you have a blog, or [here](#) if you don't.

← Previous post

Next post →

Recent Posts

extinction minus one

A zsh Helper Script For Updating macOS

RStudio Daily Electron + Quarto CLI Installs

Predictive Analytics Models in R

repoRter.nih: a convenient R interface to the NIH RePORTER Project API

Markov Chain Introduction in R

Dual axis charts – how to make them and why they can be useful

Monte Carlo Analysis in R

Stock Market Predictions Next Week

Capture errors, warnings and messages

{golem} 0.3.2 is now available

Convert column to categorical in R

Which data science skills are important

(\$50,000 increase in salary in 6-months)

A prerelease version of Jupyter Notebooks

and unleashing features in JupyterLab

Markov Switching Multifractal (MSM) model using R package

Dashboard Framework Part 2: Running Shiny in AWS Fargate with CDK

Jobs for R-users

Junior Data Scientist / Quantitative economist

Senior Quantitative Analyst

R programmer

Data Scientist – CGIAR Excellence in

Agronomy (Ref No: DDG-

R4D/DS/1/CG/EA/06/20)

Data Analytics Auditor, Future of Audit Lead

@ London or Newcastle

python-bloggers.com (python/data-science news)

Explaining a Keras _neural_ network

predictions with the-teller

Object Oriented Programming in Python –

What and Why?

Dunn Index for K-Means Clustering Evaluation

Installing Python and Tensorflow with Jupyter

Notebook Configurations

How to Get Twitter Data using Python

Visualizations with Altair

Spelling Corrector Program in Python

Full list of contributing R-bloggers

Archives

Select Month



Other sites

SAS blogs

