

LEARN TO MASTER
BIG DATA

MASTER OF SCIENCE IN
PREDICTIVE ANALYTICS
ONLINE PROGRAM
[DOWNLOAD BROCHURE](#)


NORTHWESTERN
UNIVERSITY

SCHOOL OF
PROFESSIONAL
STUDIES

REGISTER NOW

Newsletters
Web Seminars
Videos
Blogs
Radio Shows and more

information

management

Welcome, Philip | [My Account](#) | [Log Out](#)[White Papers](#) | [Web Seminars](#) | [Newsletters](#) | [eBooks](#)

- [Big Data & Analytics](#)
- [Data Management](#)
- [MDM & Data Governance](#)
- [Infrastructure](#)
- [Info Strategy & Leadership](#)
- [BI & Data Discovery](#)
- [Mobility](#)
- [web seminars & white papers](#)
- [resource center](#)

Blog

Cheat Sheets for Data Science

by [Steve Miller](#)

DEC 23, 2014 4:46pm ET

[Print](#)[Email](#)[Reprints](#)[Comment \(1\)](#)[Twitter](#)[LinkedIn](#)[Facebook](#)[Google+](#)

I received an email a few weeks back from a company called DataCamp announcing a comprehensive cheat sheet for the [R data.table package](#). Knowing I'm a data.table proselyte, the co-founder asked if I'd take a look and offer suggestions – and perhaps, I suspect, mention it in my blog.

I'm a big R fan and have been an enthusiastic [data.table devotee](#) since I discovered it several years ago. I'm also in frequent contact with package originator Matthew Dowle, and agree with more than a few in the R community that data.table's been a game-changer for elevating R's appeal to data scientists.

The DataCamp cheat sheet's well done and quite handy, covering most functionality in a single page. It certainly helps that the package serves functions – in this case data management, access, grouping and summarization – that are limited in scope. Still, the URL to the page is front and center on my notebook.

Ten years ago, I carried multiple cheat sheets in my backpack. Now, with all the support available in flexibly-queryable online documentation and support sites like [Stack Overflow](#), the need for cheat sheets isn't as acute. I maintain, however, close ties with a few.

The ubiquitous [R Reference Card](#) by Tom Short has been my trusty companion for ten years. I'm impressed with how the author managed to catch the guts of “core R” in such a compact space. Every time I pull it out, I refresh my memory on R minutiae I'd either forgotten – or never knew. Of course, the R ecosystem's now so large that you almost need cheat sheets for cheat sheets. [R Task Views](#) to the rescue. For graphics, there're cheat sheets for [R's ggplot implementation](#) that are also pertinent to the Python library of the same name.

Before I started with first S and then R about 14 years ago, my platform of choice for data management and statistical analysis was SAS. Now, though SAS is a much smaller arrow in my DS quiver, I nonetheless use it and language clone [WPS](#) semi-regularly. The timeless [SAS Cheat Sheet by David Franklin](#) makes it much easier for me to re-connect the SAS dots. Once this gets me back in the game, I simply Google my questions for the online doc answers. One change in my SAS arsenal not reflected in Franklin's page is Proc SQL, which I prefer to data step programming.

Just as I go back and forth between R and SAS for statistical analysis, so also do I migrate between Python and R for more comprehensive data science needs. And [Python Basics](#) provides a quick refresher on important basic syntax to get me going. One lament is that Python dictionaries aren't covered, but for those I just pull up archived code.

As [I've noted in previous](#) blogs, Python programming for analytics/data science has changed considerably in recent years with the ascent of community-provided libraries such as numpy, scipy, sci-kit and pandas. Given the array-orientation promoted by those packages, Python DS code can look as much like R as it does core Python.

And with these libraries now a cornerstone of analytics work in Python, it's not surprising there are cheat sheets for them as well. A spartan but still quite useful reference for numpy, scipy and pandas is [available here](#). I use the splendid [10 Minutes to pandas](#) ipython notebook at least every other month. Indeed, the [pandas documentation](#) is so good, I hardly ever now open originator Wes McKinney's well-written book, [Python for Data Analysis](#).

Finally, I maintain my scikit-learn sanity by calling on [Peekaboo's machine learning](#) reference. Can't tell either the R or scikit ML players without a scorecard.

These cheat sheets are just the tip of the DS iceberg, a small, select sampling of what's available for data science. I welcome reader input on others they find useful for their data management and analytics work.

Open Thoughts on Analytics

JOIN THE DISCUSSION

(1) [Comment](#)

SEE MORE IN

[Big Data/Analytics](#)

RELATED TAGS

[Big Data](#),

Comments (1)

Thank you for sharing.

I have a lot to plow through now!

Posted by Marcel M | Friday, January 09 2015 at 4:18AM ET

Add Your Comments:

Add your comments here.

☐ Notify me when other readers comment on this article.

[Click here to receive notifications without commenting](#)

POST ►

CANCEL

information | **BIG DATA**
management | IN PRACTICE

Expert commentary and case studies
to help you employ the right
strategy, structure and staff
to **leverage big data**

► SUBSCRIBE TO THE MONTHLY NEWSLETTER

Videos



[60 Seconds Smarter: E&M Guidelines Explained](#)

Most Read Most Emailed

- [10 Internet of Things \(IoT\) Trends & Considerations](#)
- [Data Acceleration: Turning Technology Into Solutions](#)
- [10 Tech Trends Through 2015](#)
- [Next Shift: From Big Data to Deep Data](#)
- [Data Science And Statistics: Colleges Must Evolve](#)

Featured Topics

- [BIG DATA](#)
- [BUSINESS INTELLIGENCE](#)
- [DATA INTEGRATION](#)
- [DATA MGMT](#)
- [DATA MODELING](#)
- [DATA QUALITY](#)
- [FINANCIAL SERVICES](#)
- [GOVERNANCE, RISK & COMPLIANCE](#)
- [HEALTH CARE](#)

Slideshows



[10 Big Data Career Killers](#)



[Gartner's Top 10 Cloud Myths](#)



[Gartner's Top 10 Strategic Tech Trends for 2015](#)



[8 Principles for Adopting Casework for Data](#)



[Debunking 5 of the Biggest Data Myths](#)



[Governance](#)[Five Habits of
Data-Driven
People](#)[More Slide Shows »](#)

Current Buzz on Analytics






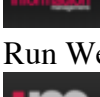




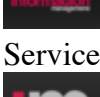

Twitter













Facebook



LinkedIn

-  [about 2 days ago](#) we were checking out Get Ready for BI Change bit.ly/14MbHk3
-  [about 2 days ago](#) we were checking out Cloud vs. Enterprise CRM Software Price War Coming bit.ly/1xzw0Ik
-  [about 2 days ago](#) we were checking out Obama Touts Well-Wired Iowa Town as Model for City-Run Web bit.ly/14LVone
-  [about 3 days ago](#) we were checking out Obama Cybersecurity Plan Could Change Healthcare Processes bit.ly/159Q0uL
-  [about 3 days ago](#) we were checking out Enterprise Marketing's Mobile Mind Shift bit.ly/14FVxsd
-  [about 4 days ago](#) we were checking out Cloud Security Concerns Shift From IT to Boardroom bit.ly/1FWbSun
-  [about 5 days ago](#) we were checking out Large Enterprises Embrace Disaster Recovery as a Service (DRaaS) bit.ly/1ya8LtE
-  [about 6 days ago](#) we were checking out Data Acceleration: Turning Technology Into Solutions bit.ly/14SDAXX
-  [about 6 days ago](#) we were checking out Next User Experience Battleground: Mobile Customer Experience bit.ly/1FNn4t4
-  [about 9 days ago](#) we were checking out CNA Using Big Data to Fight Insurance Fraud bit.ly/1y0PSI3

- 
[about 10 days ago](#) we were checking out Five Social Data Marketing Resolutions bit.ly/1FvHZ3M
- 
[about 10 days ago](#) we were checking out Personal Healthcare Data: Patients Will Take Control bit.ly/1AMBkhj
- 
[about 11 days ago](#) we were checking out Next Shift: From Big Data to Deep Data bit.ly/1xT9YUz
- 
[about 11 days ago](#) we were checking out Global Tech Market Looking Better For 2015 bit.ly/143GKY1
- 
[about 11 days ago](#) we were checking out Data Science And Statistics: Colleges Must Evolve bit.ly/1FpdFrs
- 
[about 11 days ago](#) we were checking out Flash Storage In the Data Center Catching On bit.ly/14qb1B4
- 
[about 11 days ago](#) we were checking out CNA Using Big Data to Fight Insurance Fraud bit.ly/14pO6Wl
- 
[about 13 days ago](#) we were checking out Patient Monitoring Market: \$5 Billion by 2020 bit.ly/1FgARbm
- 
[about 13 days ago](#) we were checking out Why Once-Successful Companies Fail bit.ly/1vYZoHn
- 
[about 13 days ago](#) we were checking out Average Mobile App Earns "D" Grade From Customers bit.ly/1vTsrMv

Follow @infomgmt

9,942 followers

[Analytics](#)

- [Hail to the Chiefs! Meet the CDO, CAO and CSO](#)
- [IBM: New Mainframe Boasts Improved Ability to Perform Analytics](#)
- [Data Acceleration: Turning Technology Into Solutions](#)
- [CNA Using Big Data to Fight Insurance Fraud](#)

[Business Intelligence](#)

- [Get Ready for BI Change](#)
- [UPS ORION Advanced Analytics Case Study](#)
- [Business Intelligence Analysts as Architects](#)

- [The Three C's of Self-Service BI](#)

Customer Experience

- [Insurance Executives See IT Role In Customer Satisfaction](#)
- [Cloud Security Concerns Shift From IT to Boardroom](#)
- [Next User Experience Battleground: Mobile Customer Experience](#)
- [Average Mobile App Earns "D" Grade From Customers](#)

Open Source

- [CNA Using Big Data to Fight Insurance Fraud](#)
- [Hadoop Isn't for Everyone](#)
- [China Plans Foreign Technology Purge, Shift to Homegrown Suppliers](#)
- [8 Hadoop Big Data Predictions for 2015](#)

Predictive Analytics

- [Know More, Sell More With Data Science in 2015](#)
- [Internet of People, Not Things, Is Next Frontier](#)
- [The MIT Analytics Imperative](#)
- [Google's Big Data to Help Auction.com Predict Homebuying Trends](#)

Data Governance

- [CIO Priorities: Strategic Architecture vs Quick Project Wins](#)
- [Obamacare Dental Healthcare Data Doesn't Add Up](#)
- [Target Wants Data Breach Bank Claims Dismissed](#)
- [Creating the Data Governance Killer App](#)

Data Integration

- [Big Data Requires Integration Technology](#)
- [Actuate, Oracle, Red Hat, Others Announce New or Updated Products](#)
- [How Healthcare CIOs View the Data Analytics Landscape](#)
- [Headache Number Two: Remedies for Data Migration Pain](#)

Data Management

- [Large Enterprises Embrace Disaster Recovery as a Service \(DRaaS\)](#)
- [Personal Healthcare Data: Patients Will Take Control](#)
- [Why Once-Successful Companies Fail](#)
- [Top Healthcare Technology Stories of 2014](#)
- [HOME](#)
- [About Us](#)
- [Contact Us](#)
- [Content Licensing](#)
- [Advertise with Us](#)

- [Customer Service](#)
- [Feedback](#)
- [My Account](#)
- [Site Map](#)
- [Privacy Policy](#)
- [Editorial Submissions](#)

[**sourcemedi**](#)

[**corporate site**](#)

banking

- [American Banker](#)
- [Bank Technology News](#)
- [American Banker Magazine](#)
- [Credit Union Journal](#)

MORTGAGES

- [National Mortgage News](#)

PAYMENTS

- [PaymentsSource](#)
- [Collections & Credit Risk](#)
- [ISO & Agent](#)

capital markets

- [Mergers & Acquisitions](#)
- [Asset Securitization Report](#)
- [Leveraged Finance News](#)
- [Private Placement Letter](#)
- [Traders Magazine](#)
- [Securities Technology Monitor](#)

MUNICIPAL FINANCE

- [The Bond Buyer](#)

accounting

- [Accounting Today](#)
- [Tax Pro Today](#)

HEALTHCARE & BENEFITS

- [Employee Benefit News](#)
- [Employee Benefit Adviser](#)
- [Health Data Management](#)
- [Insurance Networking News](#)
- [Information Management](#)

INVESTMENT ADVISORY

- [Financial Planning](#)
- [On Wall Street](#)
- [Bank Investment Consultant](#)
- [Money Management Executive](#)
- [Investment Mgt. Mandate Pipeline](#)

© 2015 [SourceMedia](#). All rights reserved.