



**REvolution**  
computing

*We do the math*

## Introduction to R

David Smith, REvolution Computing  
LUGOD, Nov 16 2009



# Open-source statistical analysis and visualization



# What is R?

- Open-source software for advanced statistical analysis and data visualization
- Interpreted language *designed* for statistical programming
  - Virtually all statistical and predictive analytic methods available
  - Create analysis in a fraction of the time compared to C++, Python, SAS, SPSS, ...
- [www.r-project.org](http://www.r-project.org)

- Late 70's / Early '80s: S language invented at AT&T Bell Labs
- 1983-1992: S version 3
  - 1998: Chambers wins ACM Software Systems award for S
- 1993: R Gentleman and R Ihaka create R
  - 1995: Released under GPL v2
- 1997: R Core Group formed
- 2001: R 1.0.0 released

- Oct 2009: R 2.10.0 released
- 2M+ users (estimated)
  - Google, Facebook, Pfizer, Bank of America, New York Times
- 2000+ contributed packages
  - [cran.r-project.org](http://cran.r-project.org)
- Annual international user group meeting
- Commercially supported by Revolution Computing

# Momentum behind R


HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR TIMES TOPICS My Account Welcome, nonfamous Log Out Help TimesPeople

**The New York Times** **Business Computing** Search All NYTimes.com

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION ARTS STYLE TRAVEL JOBS REAL ESTATE AUTOS

Search Technology **Inside Technology** Bits Personal Tech »  
Internet Start-Ups Business Computing Companies Blog » Cellphones, Cameras, Computers and more

## Data Analysts Captivated by R's Power



Stuart Isett for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By **ASHLEE VANCE**  
Published: January 6, 2009

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

**Related**  
[Bits: R You Ready for R?](#)  
[The R Project for Statistical Computing](#)

R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly

**More Articles in Technology »**

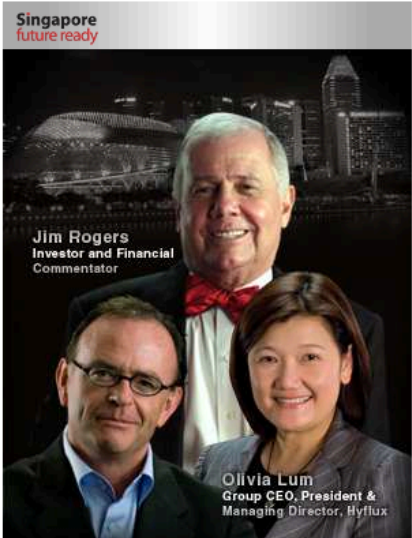
### Latest Technology Headlines

FROM NYTIMES.COM

- Is Palm's Comeback Losing Steam?
- Little Benefit Seen, So Far, in Electronic Patient Records
- Cisco Raises Bid for Tandberg of Norway
- YouTube + Univision = TuTube
- Obama Walks China's 'Great Firewall'

Powered by **LinkedIn**

**Singapore future ready**



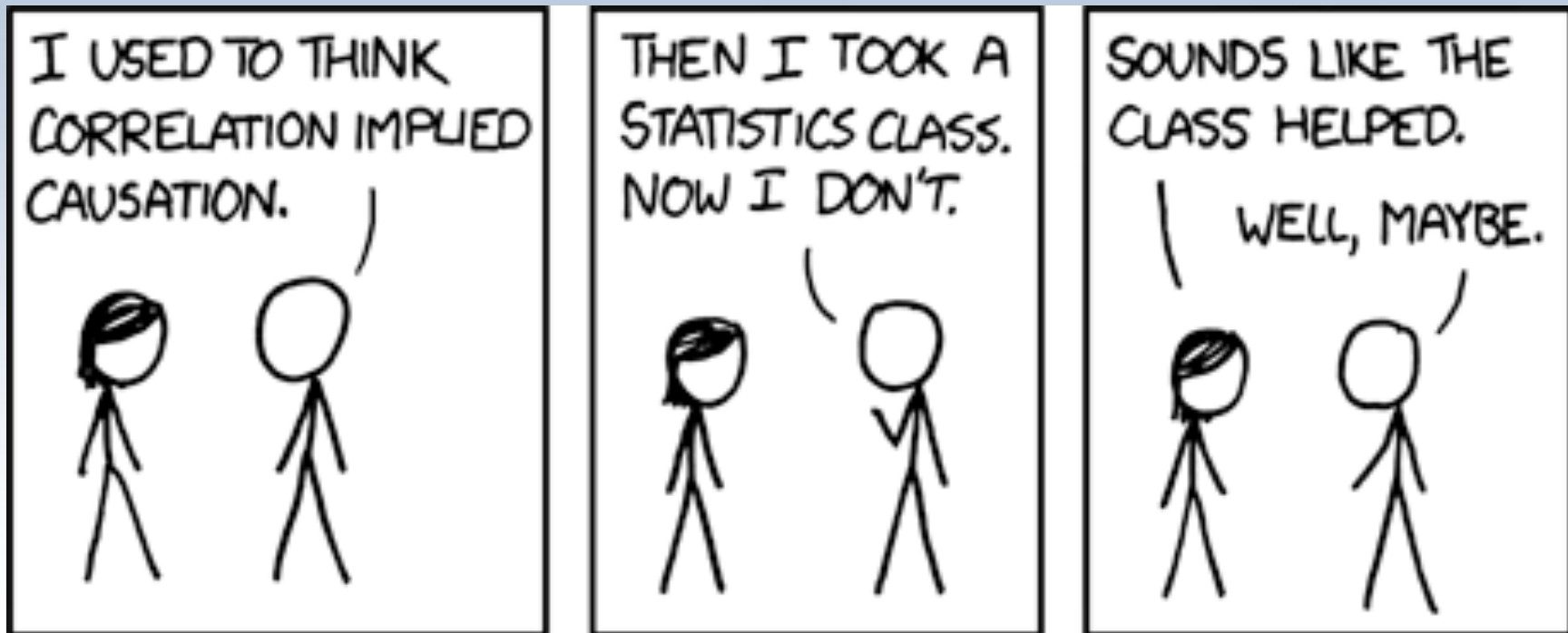
**Jim Rogers**  
Investor and Financial Commentator

**Olivia Lum**  
Group CEO, President & Managing Director, Hyflux

RECOMMEND  
TWITTER  
E-MAIL  
SEND TO PHONE  
PRINT  
SINGLE PAGE  
REPRINTS  
SHARE

THIS

# What is Statistics?



<http://xkcd.com/552/>

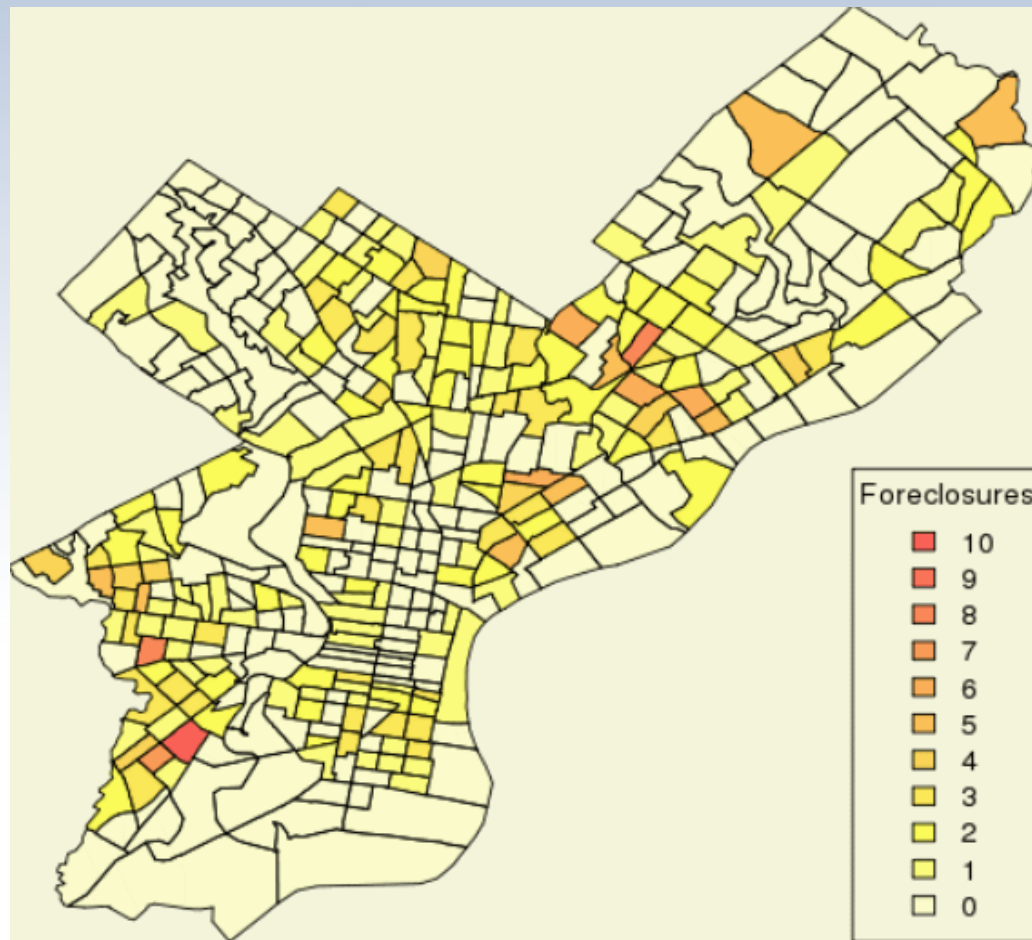
# REvolution Computing: The “R” Company

- REvolution R
  - Free, high-performance binary distribution of R
- REvolution R Enterprise
  - Subscription-based version, bundled with proprietary extensions
  - Fully supported, validated
- R Consulting and Training
- R Community Development
  - R Evangelism ([blog.revolution-computing.com](http://blog.revolution-computing.com))
  - R community portal
  - Technical and financial contributions to R Project



# What can you do with R?

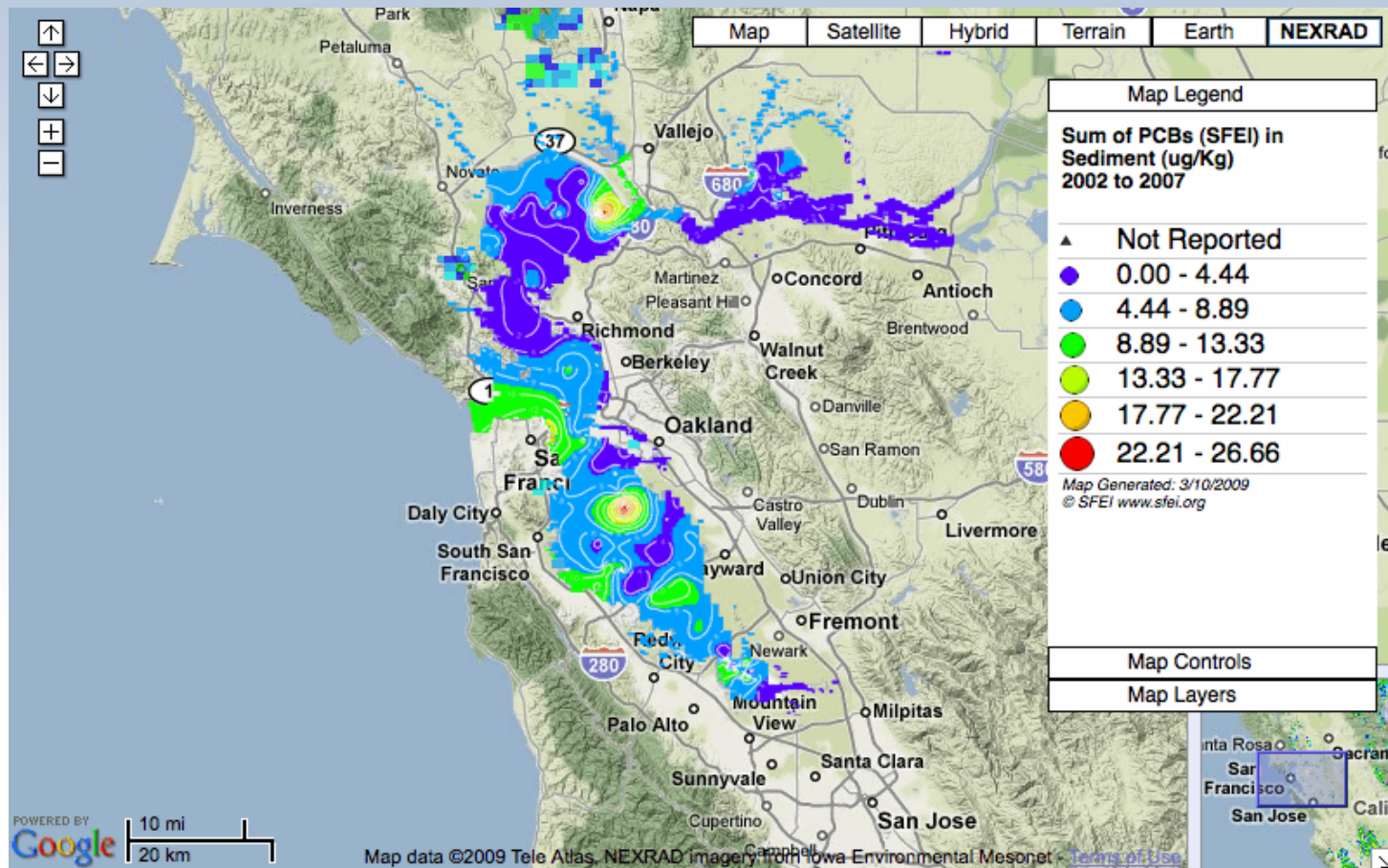
- Mash-up messy data sources to analyze the foreclosure crisis



From O'Reilly's *Data Mash-ups in R*.

# What can you do with R?

- Find a clean place to surf in the Bay Area



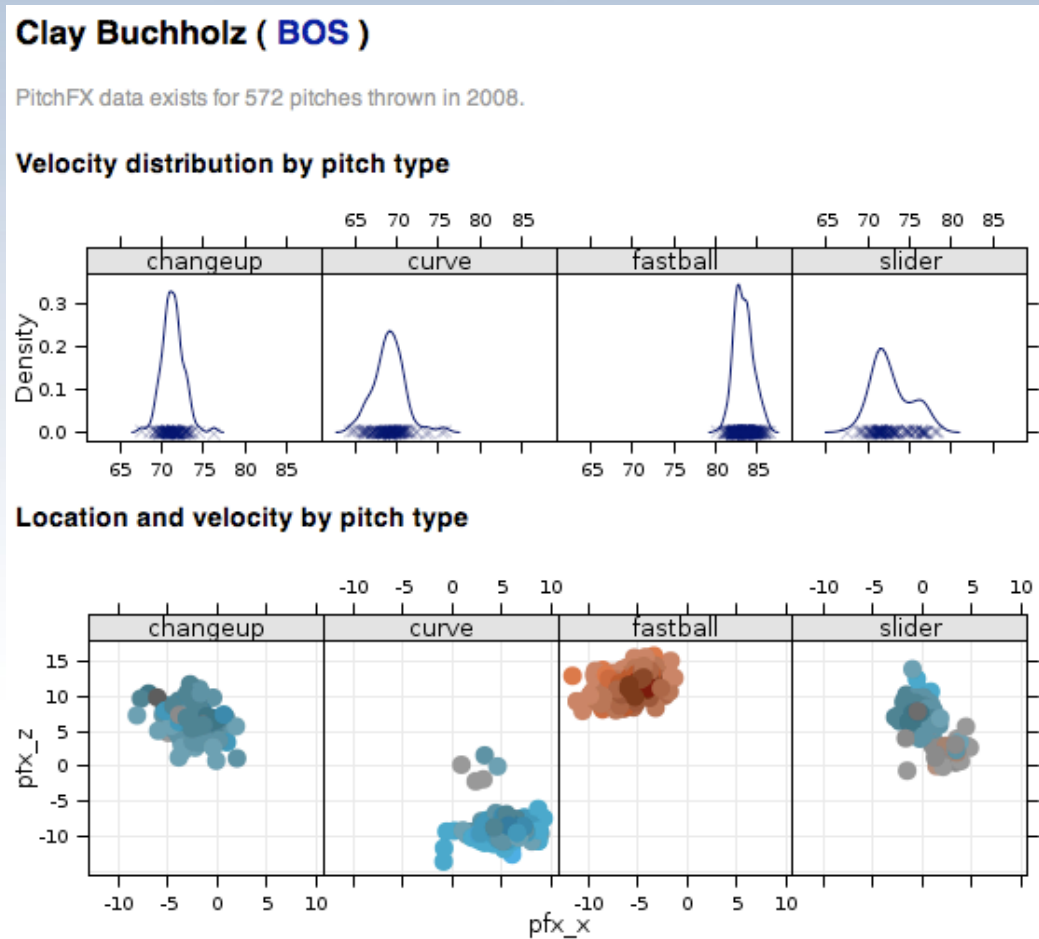
LUGOD, Nov 16 2009

John Oram, a scientist at the [San Francisco Estuary Institute](#) (SFEI) uses R to collect and monitor environmental data from the waters and wetlands of the Bay Area

10

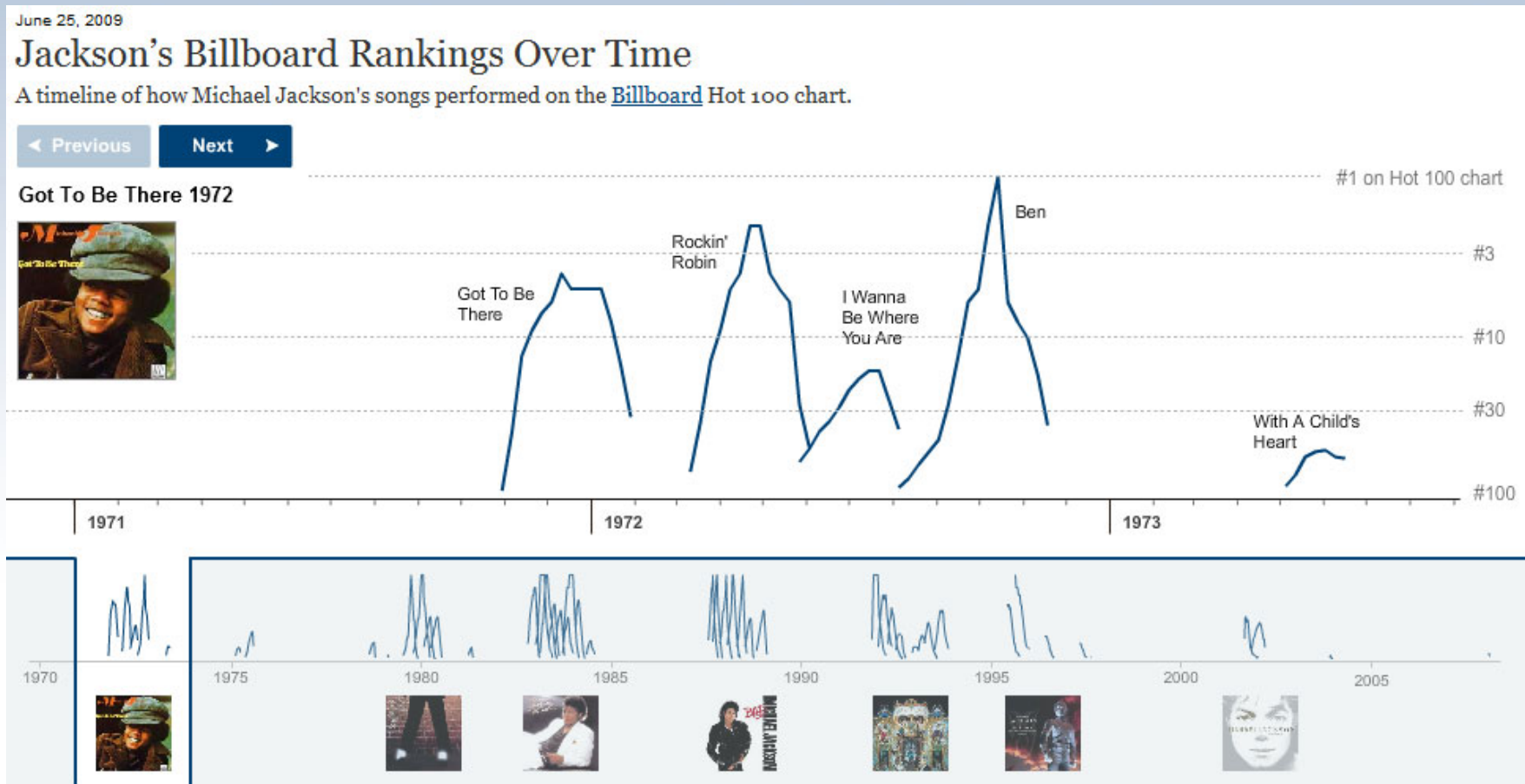
# What can you do with R?

- Compare baseball player performance



# R saves time for the New York Times

- Published 3 hours after Jackson's death:





## Getting Started with R on Ubuntu (Karmic)

```
$ sudo apt-get revolution-r
```

```
$ R
```

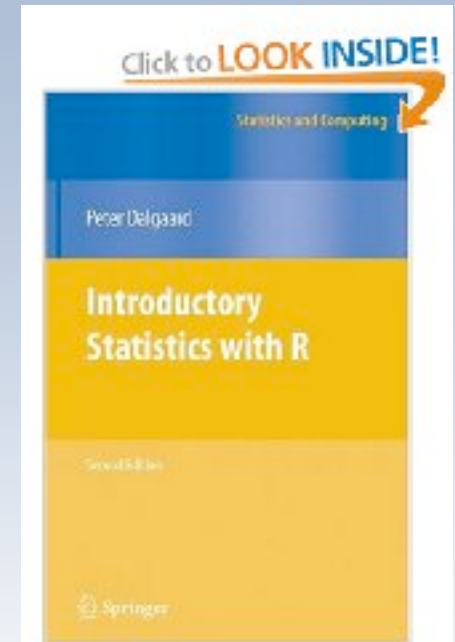
- R is interactive: type statements at the prompt, and the result is printed:

```
> x <- rnorm(10, mean=1, sd=2)
```

```
> mean(x)
```

```
[1] 1.695564
```

- *Introduction to R* (free)
  - 20-minute tutorial: Appendix A
- *Introductory Statistics with R*
  - Peter Dalgaard
- *R in a Nutshell* (O'Reilly)
- Other resources for beginners:
  - [www.revolution-computing.com/community/resources.php](http://www.revolution-computing.com/community/resources.php)
  - [blog.revolution-computing.com/beginner-tips/](http://blog.revolution-computing.com/beginner-tips/)



## ESS: An emacs-based GUI for R

- `sudo apt-get install ess`
- Start R in Emacs: **M-x R**
- <http://ess.r-project.org>
- ESS Reference Card:
  - <http://ess.r-project.org/refcard.pdf>

Let's do a simulation!

- Is this your birthday?

January

6



## Simulating birthdays

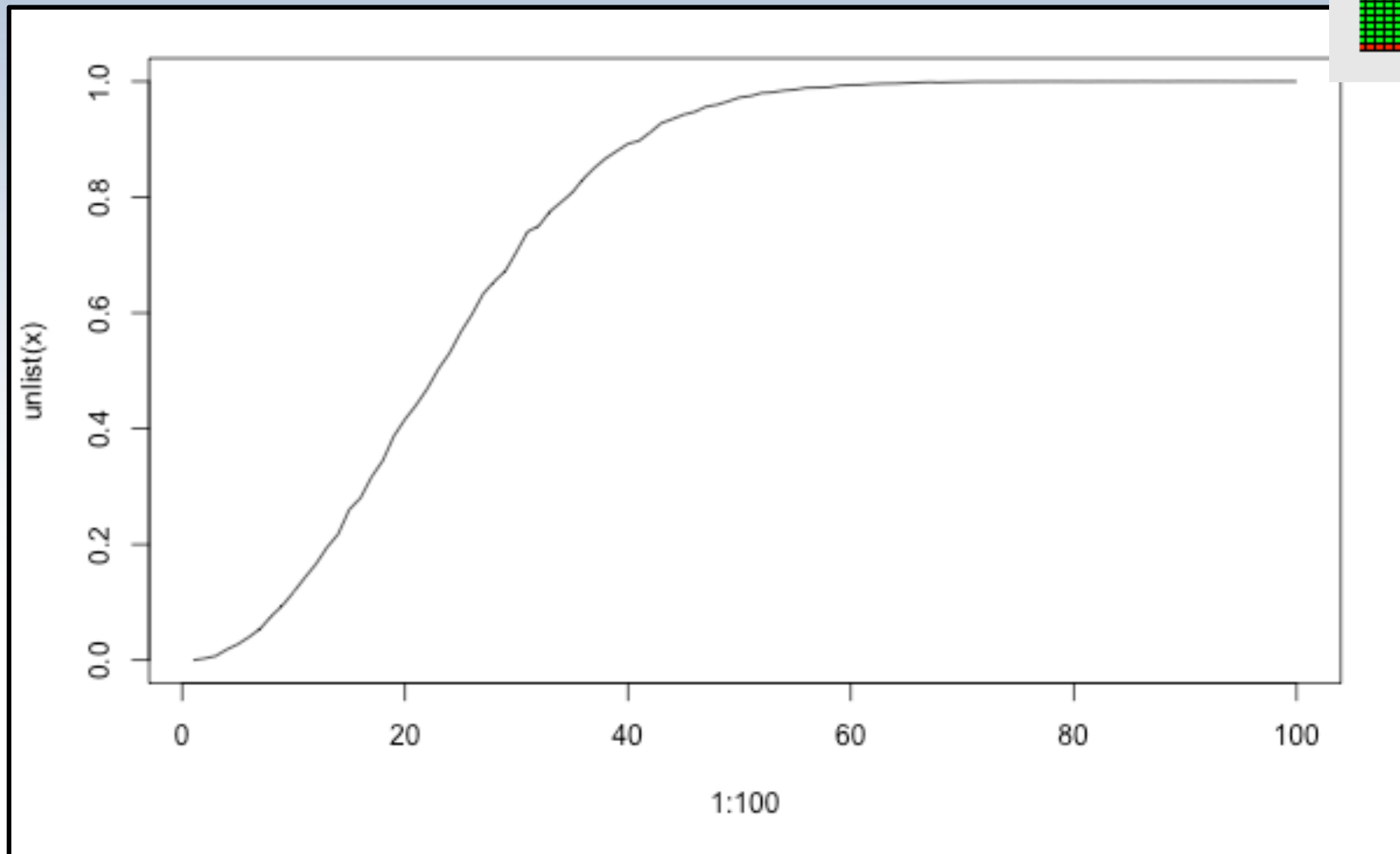
- A simple simulation:

```
birthday <- function(n) {  
  ntests <- 10000  
  pop <- 1:365  
  anydup <- function(i)  
    any(duplicated(  
      sample(pop, n, replace=TRUE)))  
  sum(sapply(seq(ntests), anydup)) / ntests  
}  
  
x <- foreach (j=1:100) %dopar% birthday (j)
```

# Birthday Simulation

```
> x <- foreach (j=1:100) %dopar% birthday (j)
> plot(1:100, unlist(x), type="l")
```

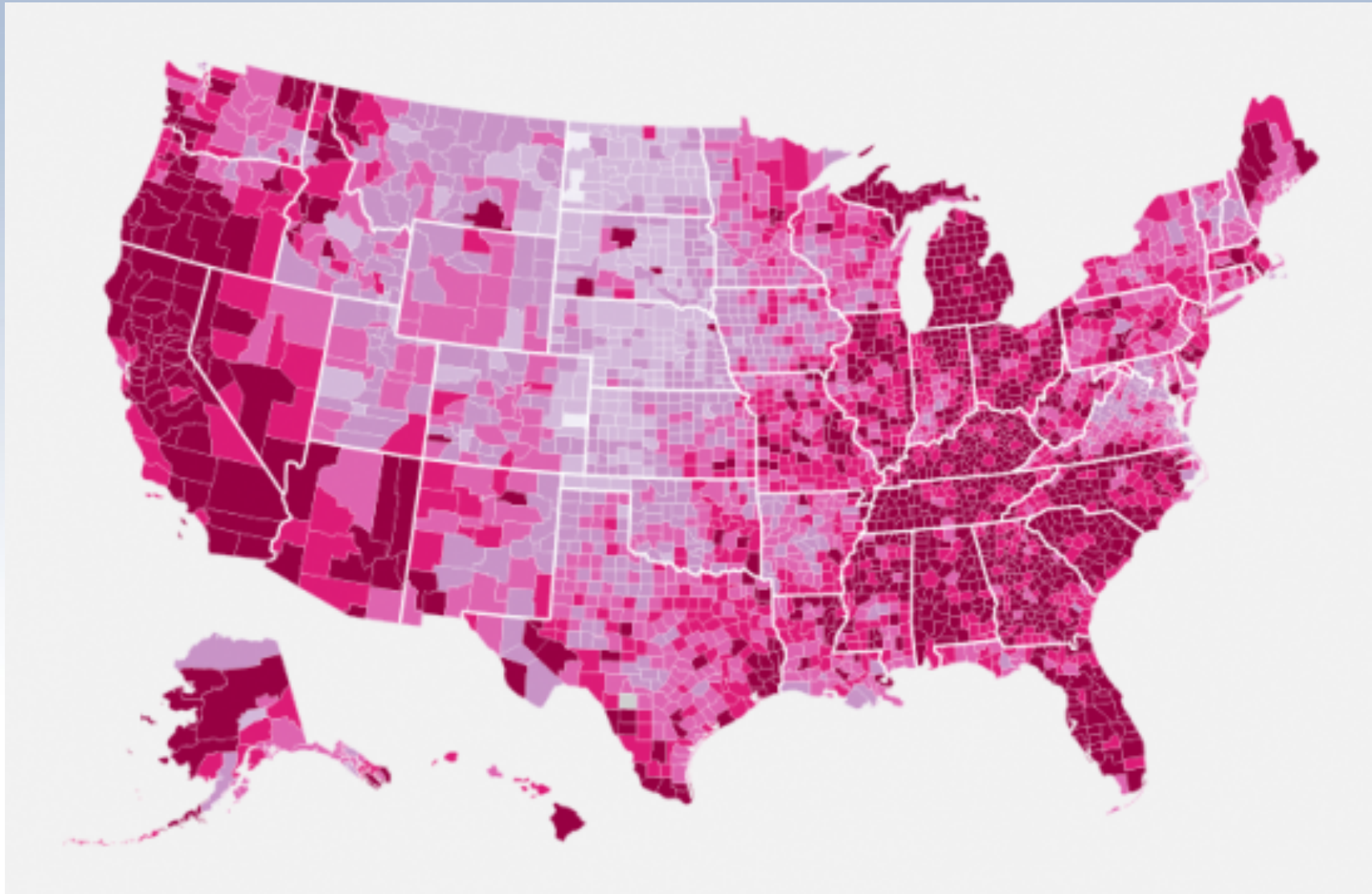
CPU Usage



## More Examples

- Plotting unemployment data on a map
- Displaying financial data as a calendar heatmap
- Analyzing Twitter hashtags

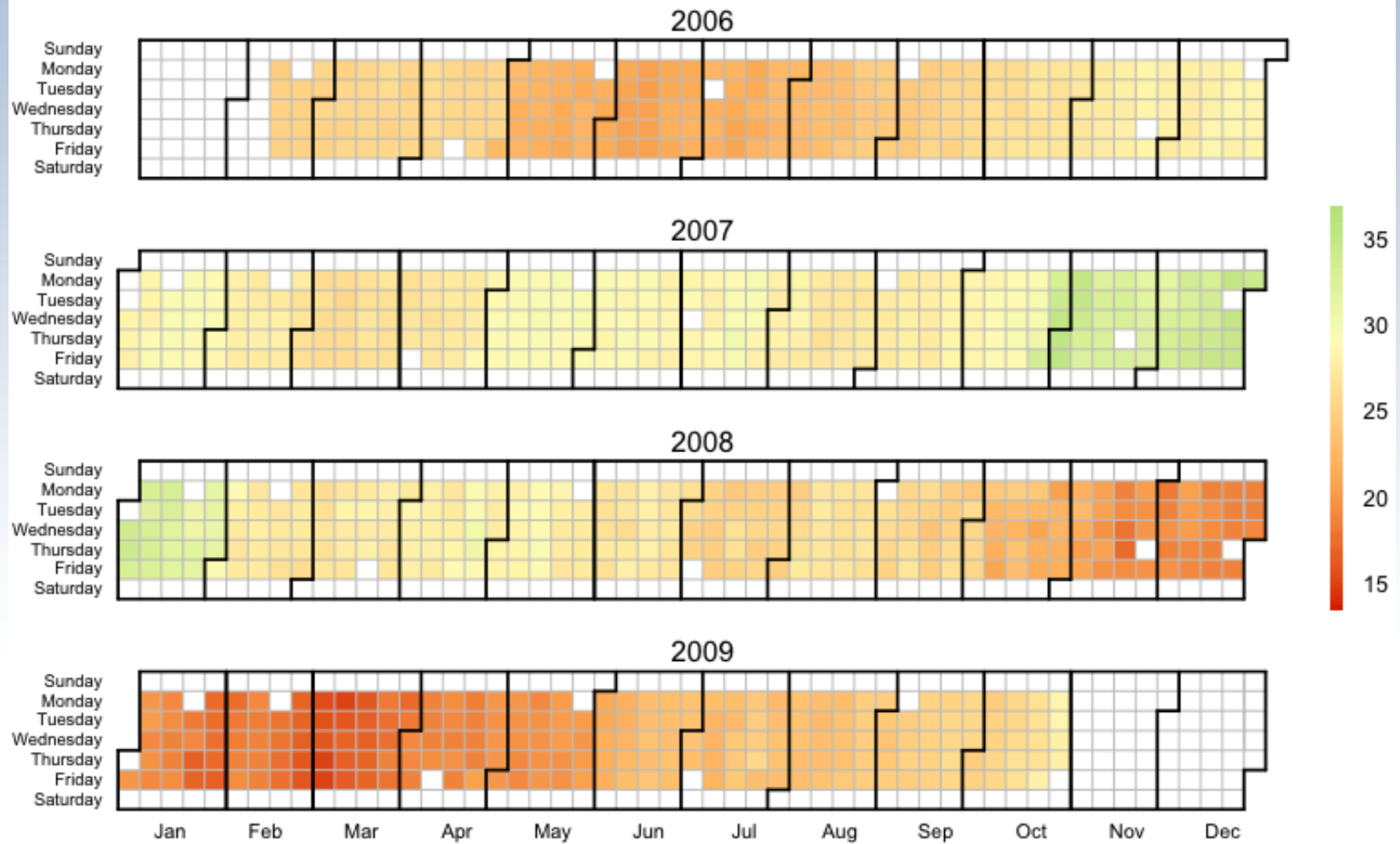
# Unemployment data map



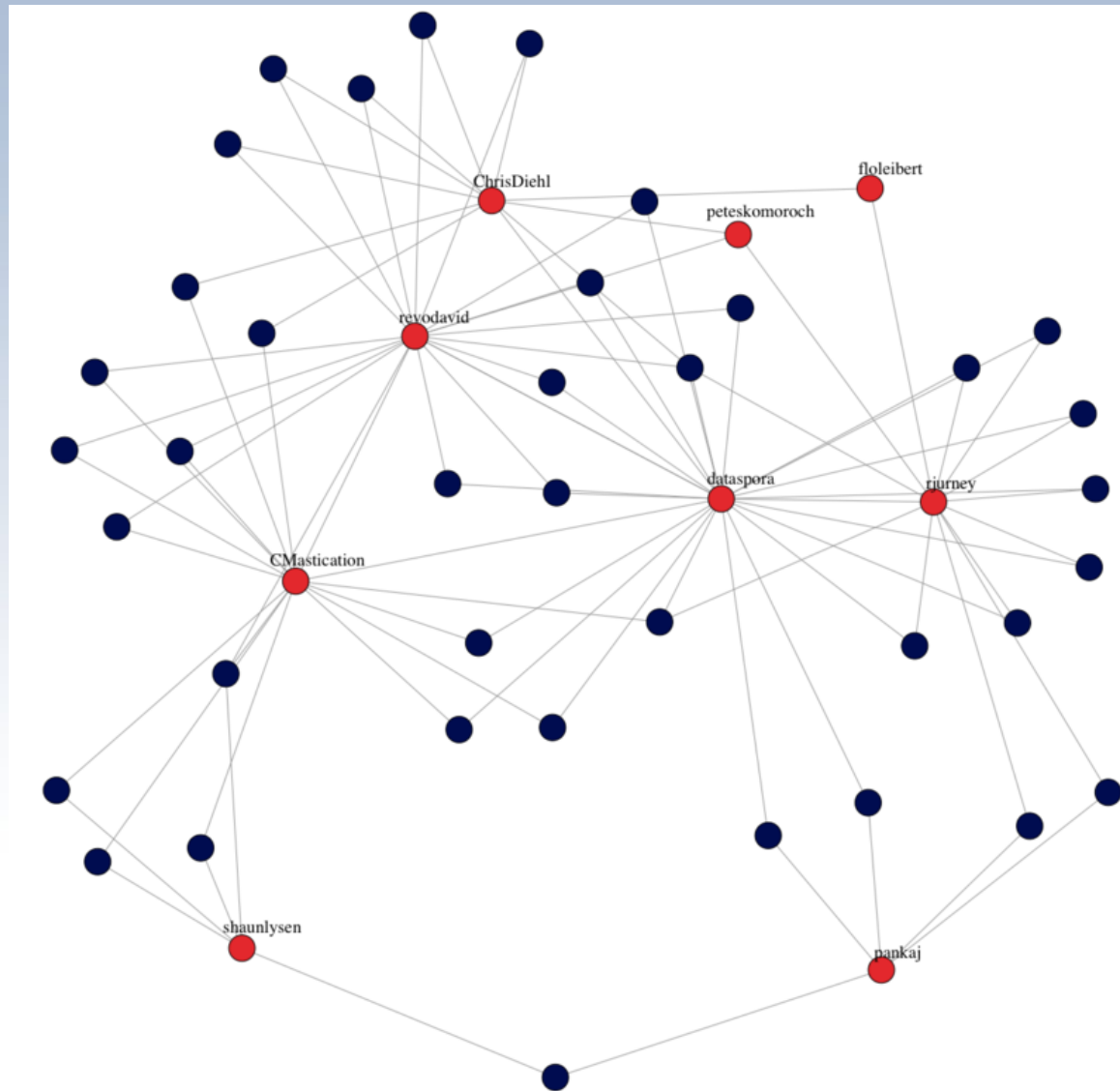


# Calendar heat map

## Calendar Heat Map of MSFT Adjusted Close



# Social network analysis with Twitter



# Thank You!

- David Smith
  - [david@revolution-computing.com](mailto:david@revolution-computing.com), @revodavid
- *Revolutions*, the R blog
  - [blog.revolution-computing.com](http://blog.revolution-computing.com)
- R Project
  - [www.r-project.org](http://www.r-project.org)
  
- Links for this talk:
  - [tinyurl.com/lugod-nov09](http://tinyurl.com/lugod-nov09)