R we f#¢$ed?

Why We Don't Understand Risk, and How it Dooms Us All

Strangers with Candy?

- Your six-year-old daughter is lost in the city. What should she do?

www.flickr.com/photos/keyphotographics/2185720078

REvolution computing
We do the math

Seven things you should know about risk

REvolution computing
We do the math

1:

Risk

= Probability

× Impact

REvolution computing

We do the math

Models: not just for dressing up and falling down

www.flickr.com/photos/edgun/2624354242

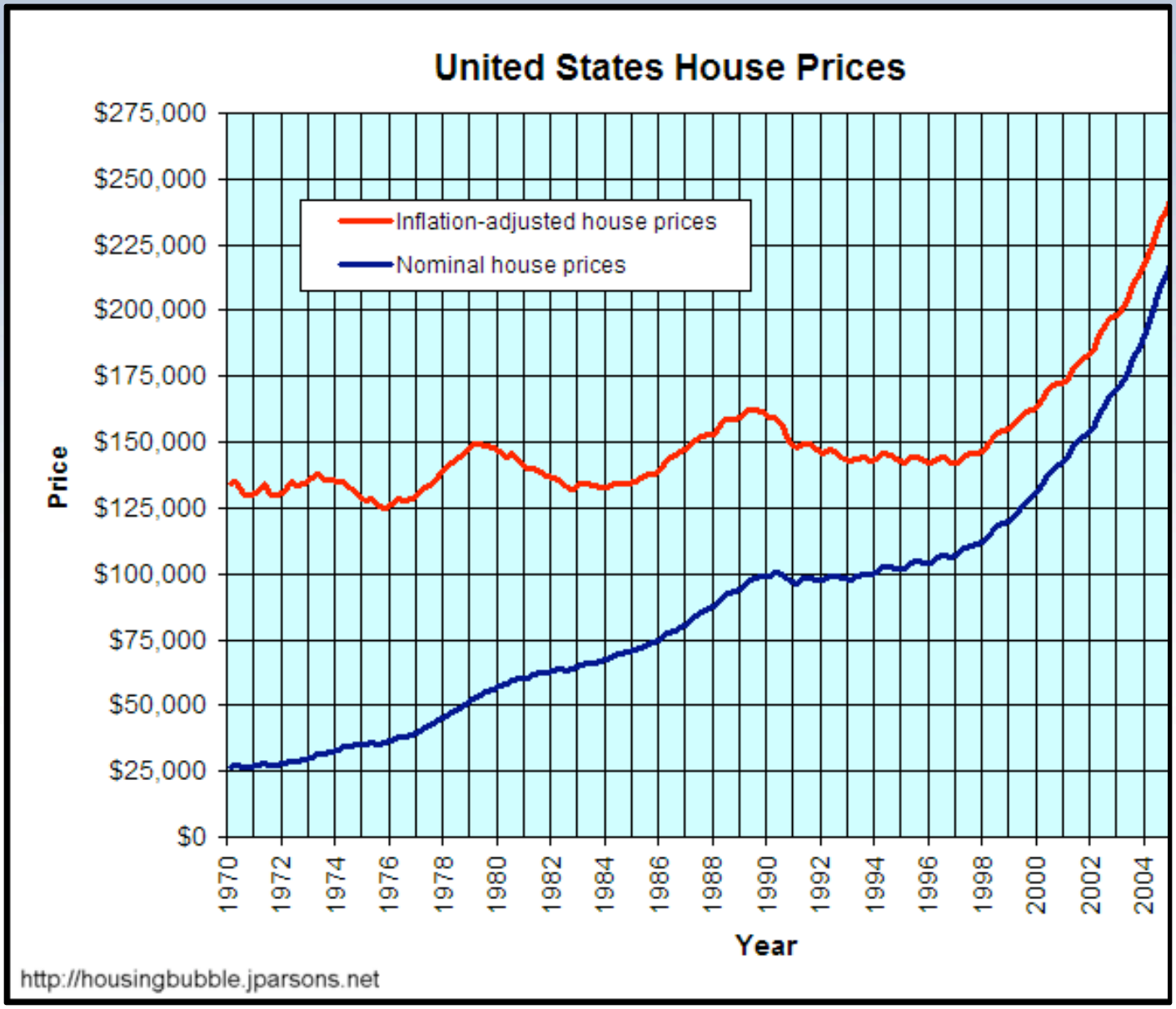
REvolution computing
We do the math

2: All models are wrong, but some are useful.

G. E. P. Box, Statistician

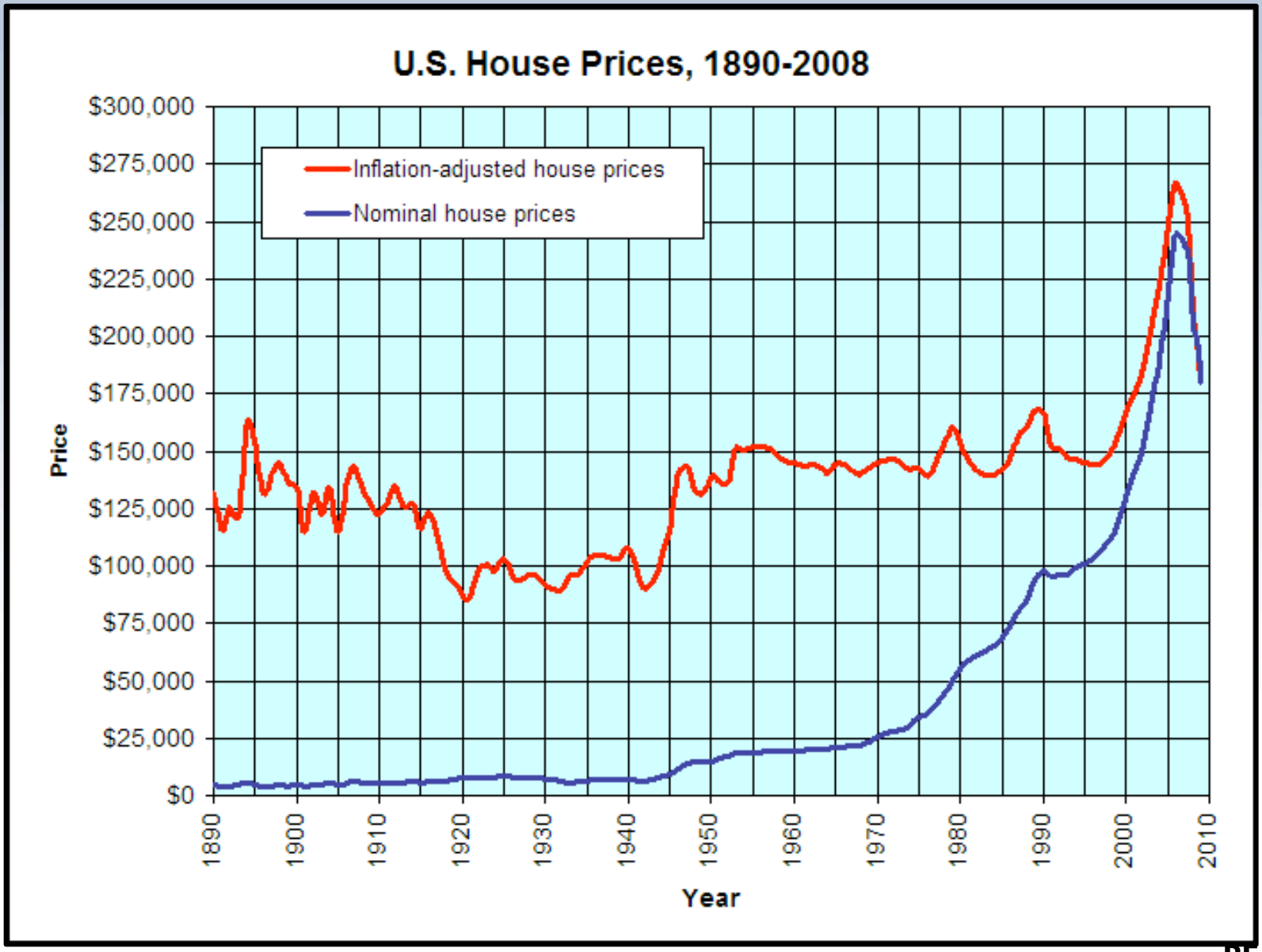
REvolution computing
We do the math

Housing prices only ever go up!



REvolution computing
We do the math

Er, well…



Robert Shiller, www.irrationalexuberance.com

REvolution computing
We do the math

#3: Coincidences are not surprising

Jessica Utts: http://anson.ucdavis.edu/~utts/AmerStat2003.pdf

These things could never happen!

- Two submarines crash in the ocean
- Meteor strikes an airplane
- Cosmic rays make a computation fail

"Everything we see has about a zero probability. Calculating these probabilities after the fact is kind of meaningless."

– Peter Westfall

REvolution computing
We do the math

#4: Conditional probability: you're doing it wrong

Jessica Utts: http://anson.ucdavis.edu/~utts/AmerStat2003.pdf

Talking on your cell phone while driving is totally safe

AAA (2001):

- 1.5% of drivers in accidents were using a cellphone
- 10.9% of drivers were distracted by a passenger
- OMG! Passengers are more dangerous than cellphones?

#5: Surveys and experiments probably have bias

REvolution computing
We do the math

“Sure officer, I’ll tell you exactly what a dumbass I just was.”

- Further critique by Magliozzi & Magliozzi (2001): this is **volunteered** information
- http://tinyurl.com/R-aaa-study

© Pixar, Inc

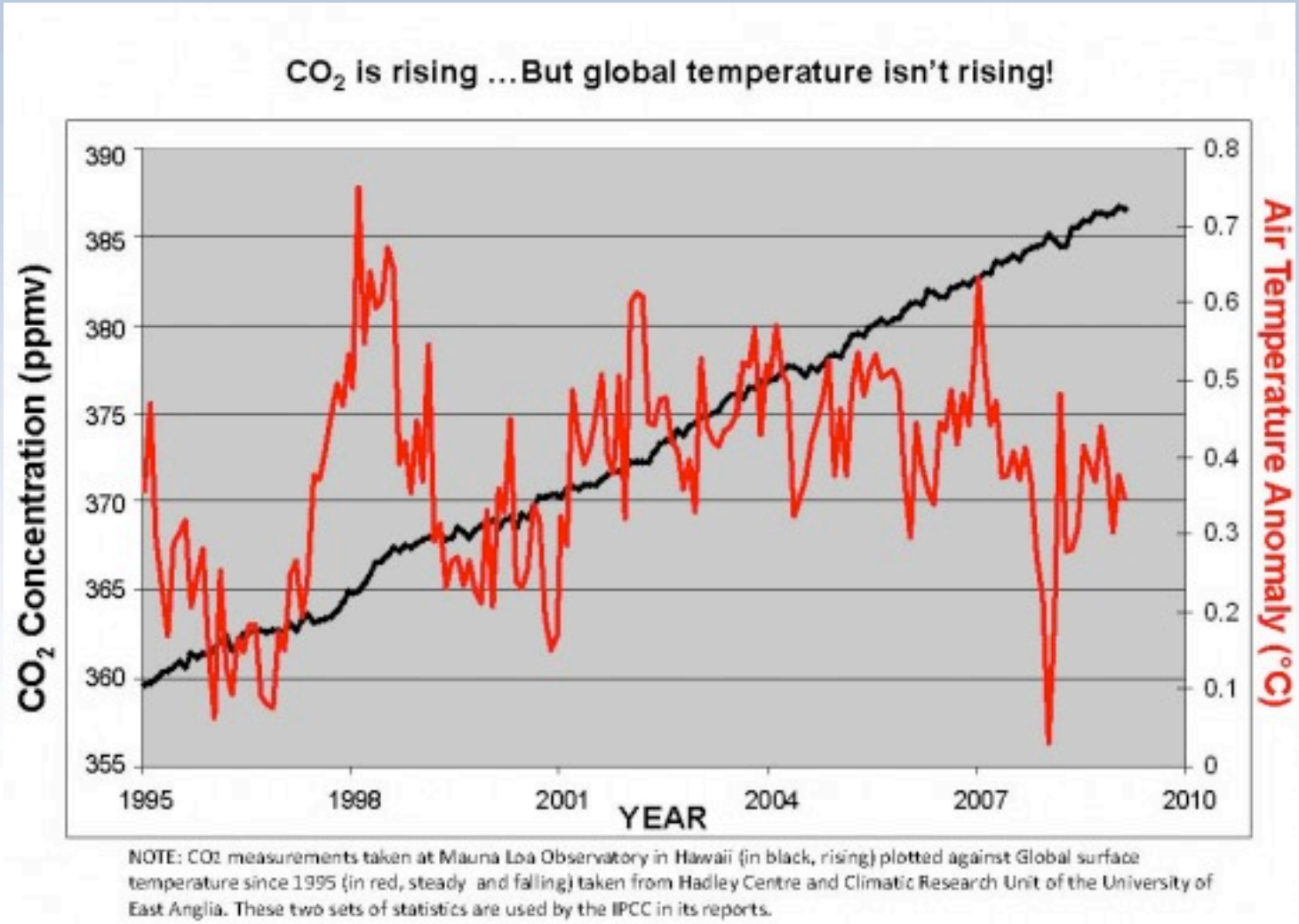
REvolution computing
We do the math

#6: The plural of anecdote is not data

Frank Kotsonis

REvolution computing
We do the math

Global warming is totally bunk!



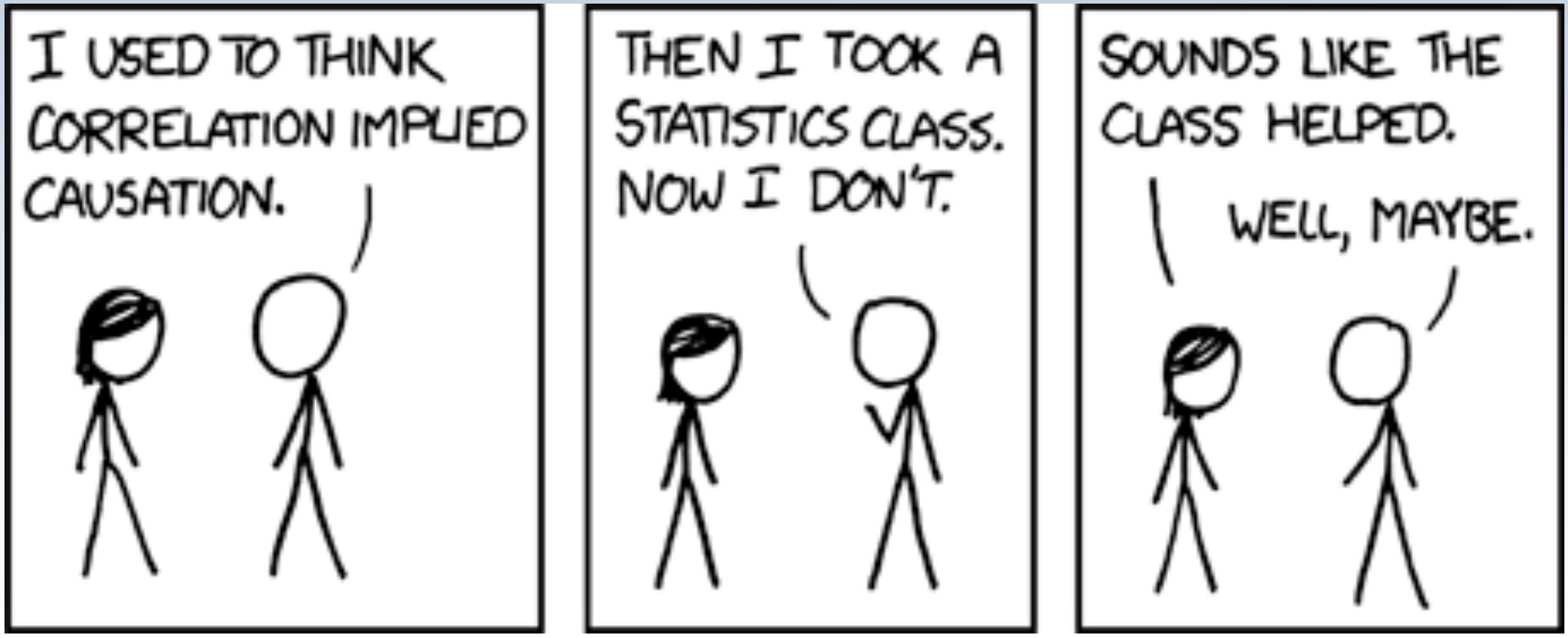
NOTE: CO2 measurements taken at Mauna Loa Observatory in Hawaii (in black, rising) plotted against Global surface temperature since 1995 (in red, steady and falling) taken from Hadley Centre and Climatic Research Unit of the University of East Anglia. These two sets of statistics are used by the IPCC in its reports.

http://flowingdata.com/2009/07/20/important-data-please-act-responsibily/

REvolution computing
We do the math

#7: Correlation doesn't imply causation

Suggestively-waggling eyebrows



http://xkcd.com/552/

REvolution computing
We do the math

Open-source statistical analysis and visualization

REvolution computing

We do the math

What can you do with R?

- Improve your career prospects

Dice.com: "Is There an R in Your Future?"
http://tinyurl.com/R-dice-video

REvolution computing
We do the math

What is R?

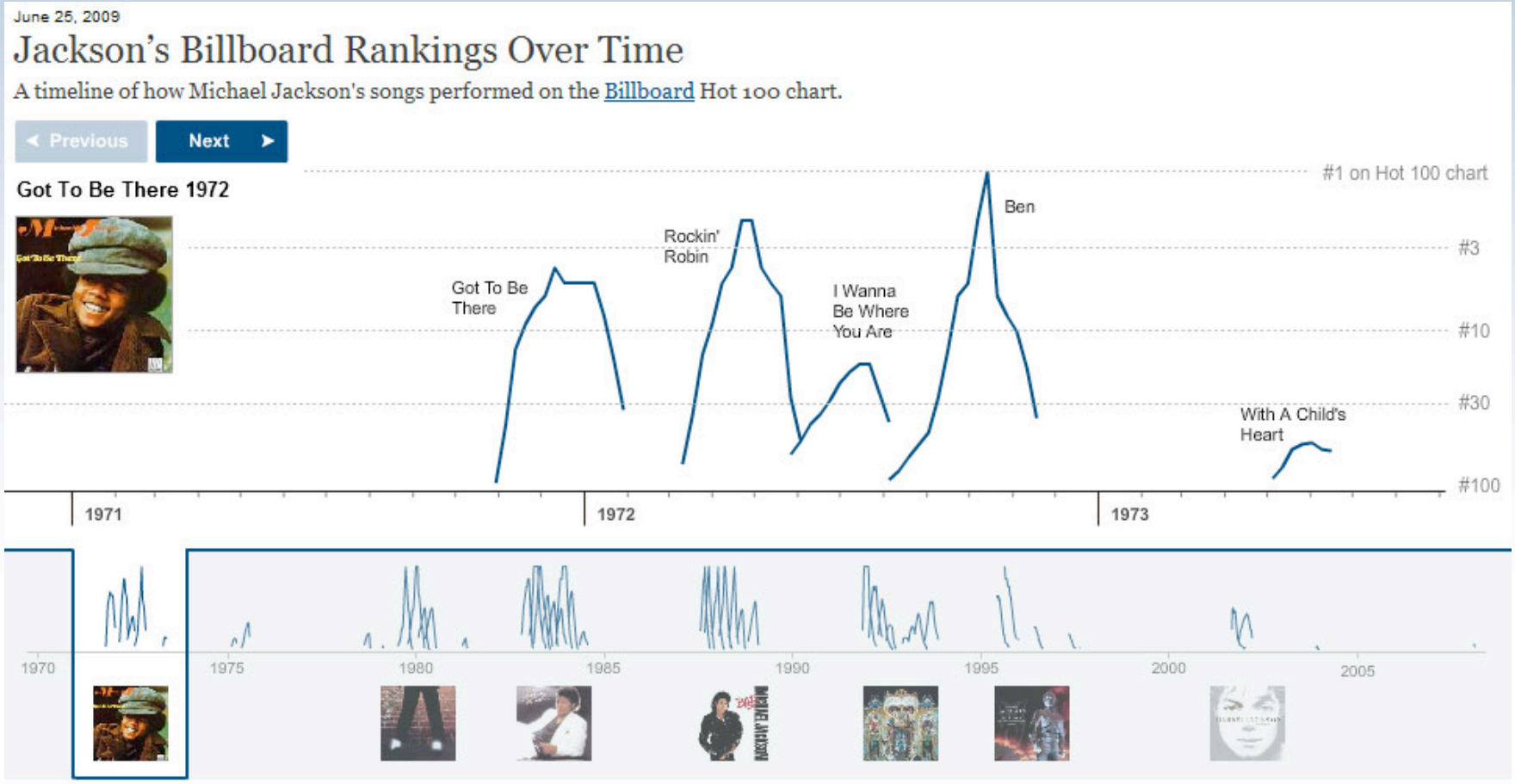
- R is open-source software for statistical computing and data visualization.
- R is an extensible, interactive programming language.
- R has a vast library of mathematical and statistical algorithms.
- R has a world-wide community of 1-2 million users and developers.
- R runs on Windows, Mac, Linux and Unix systems.

Who uses R?

- R is used by anyone who needs visualize or extract information from data:
  - statisticians, scientists, data analysts, data miners, quantitative analysts, modelers, forecasters
- R is used to analyze and present data at organizations like:
  - Google, Bank of America, Pfizer, Facebook, Shell, the *New York Times*, the Food and Drug Administration
- R is used as a free alternative to products like:
  - SAS, SPSS, Stata, Statistica, S-PLUS.
- R is used by academics to implement the latest advances in statistics and predictive analytics

R saves time for the New York Times

- Published 3 hours after Jackson's death:

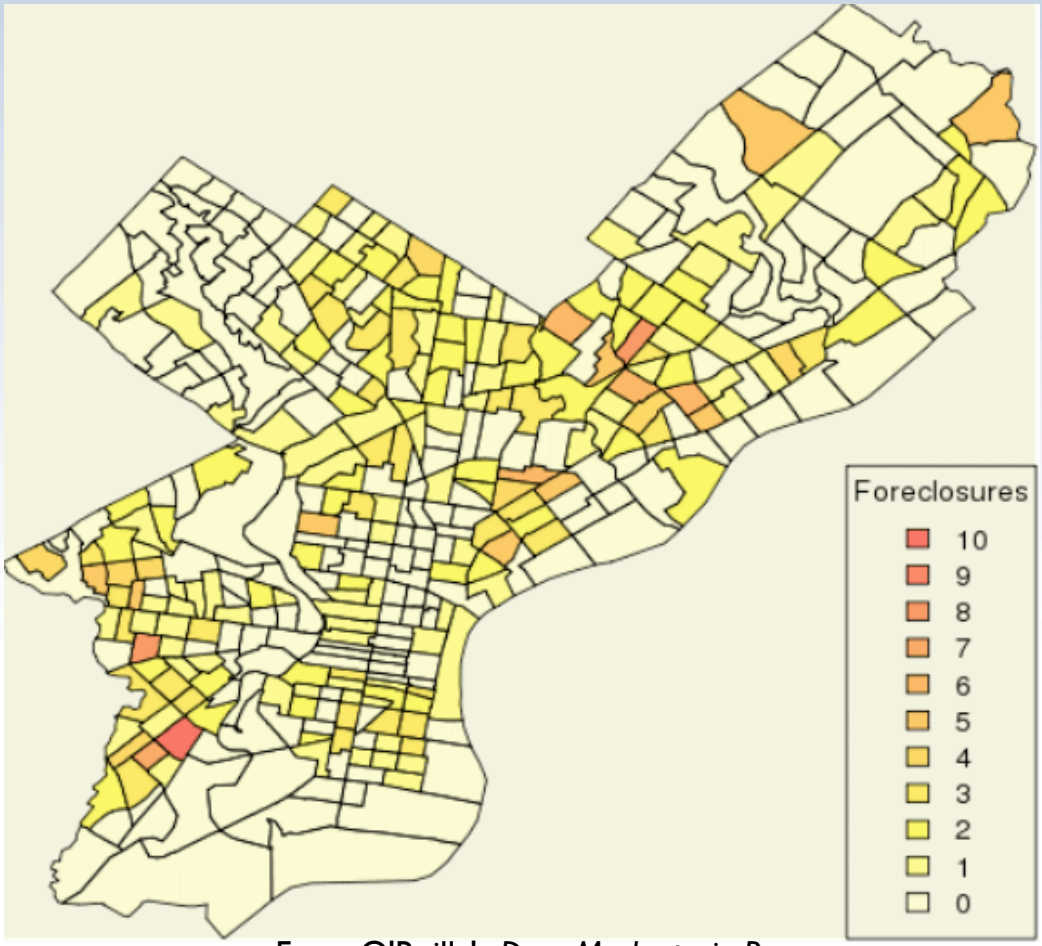


nyt.com, June 25 2009

REvolution computing
We do the math

What can you do with R?

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

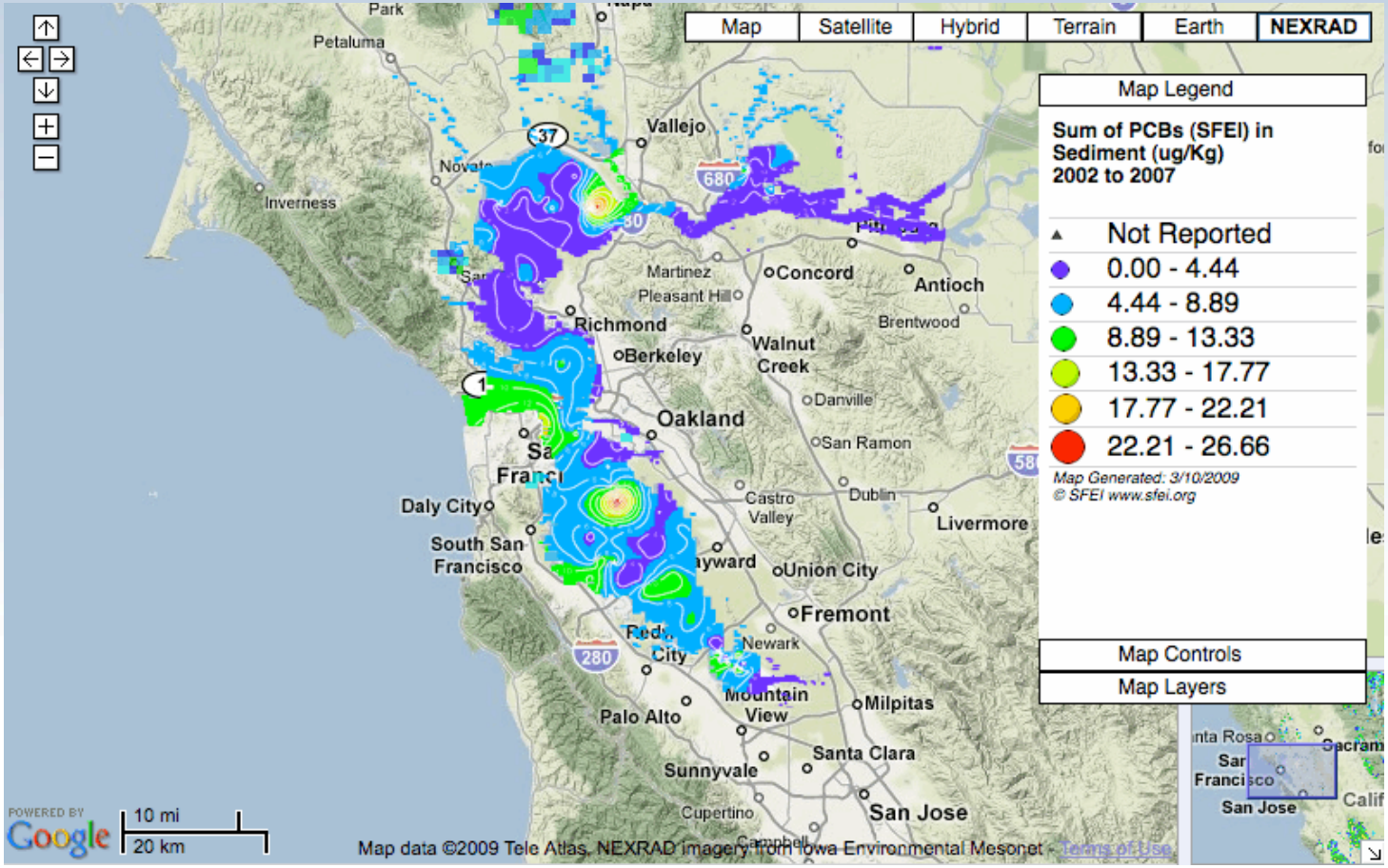


From O'Reilly's *Data Mash-ups in R.*

REvolution computing
We do the math

What can you do with R?

- Find a clean place to surf in the Bay Area

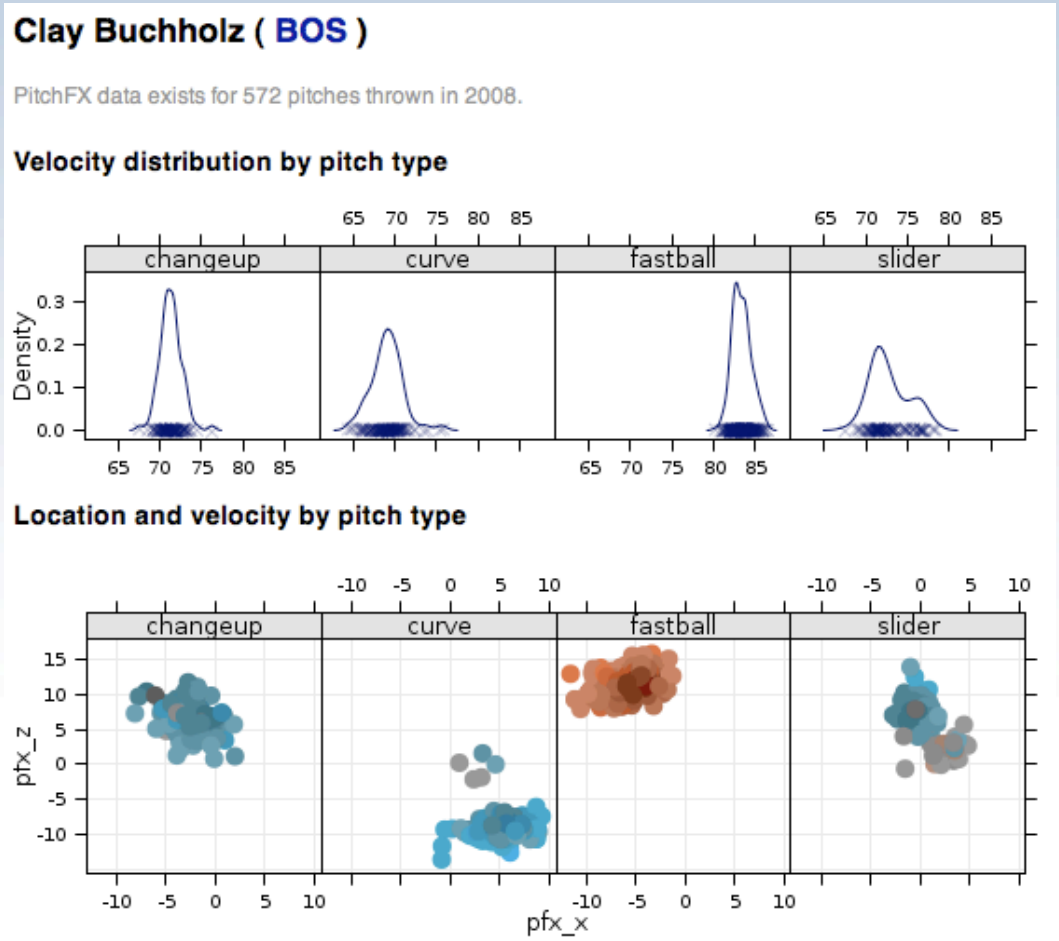


John Oram, a scientist at the [San Francisco Estuary Institute](#) (SFEI) uses R to collect and monitor environmental data from the waters and wetlands of the Bay Area

REvolution computing
We do the math

What can you do with R?

- Compare baseball player performance

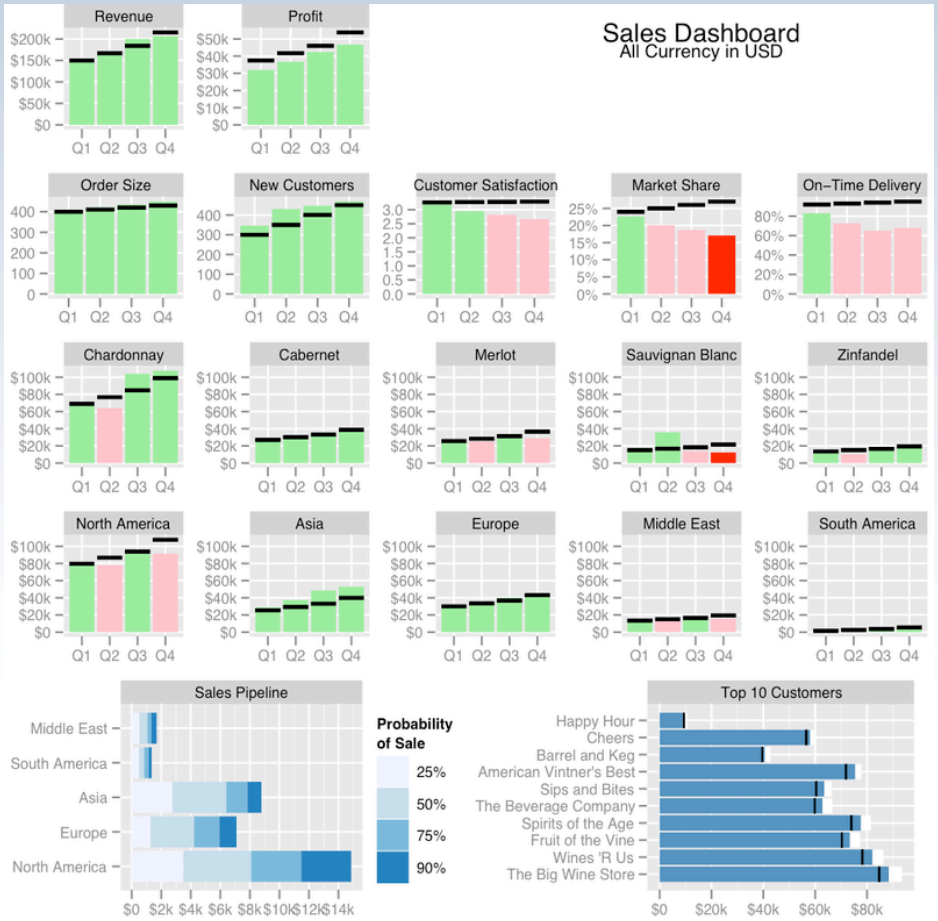


PitchFX Viewer, by Mike Driscoll: labs.dataspora.com/gameday/

REvolution computing
We do the math

What can you do with R?

- Build a Web dashboard, fast.



From the "Learning R" blog, "R: From zero to Web 2.0 in six weeks"

REvolution computing
We do the math

What can you do with R?

- Show that a tipping point in gay marriage support is near.

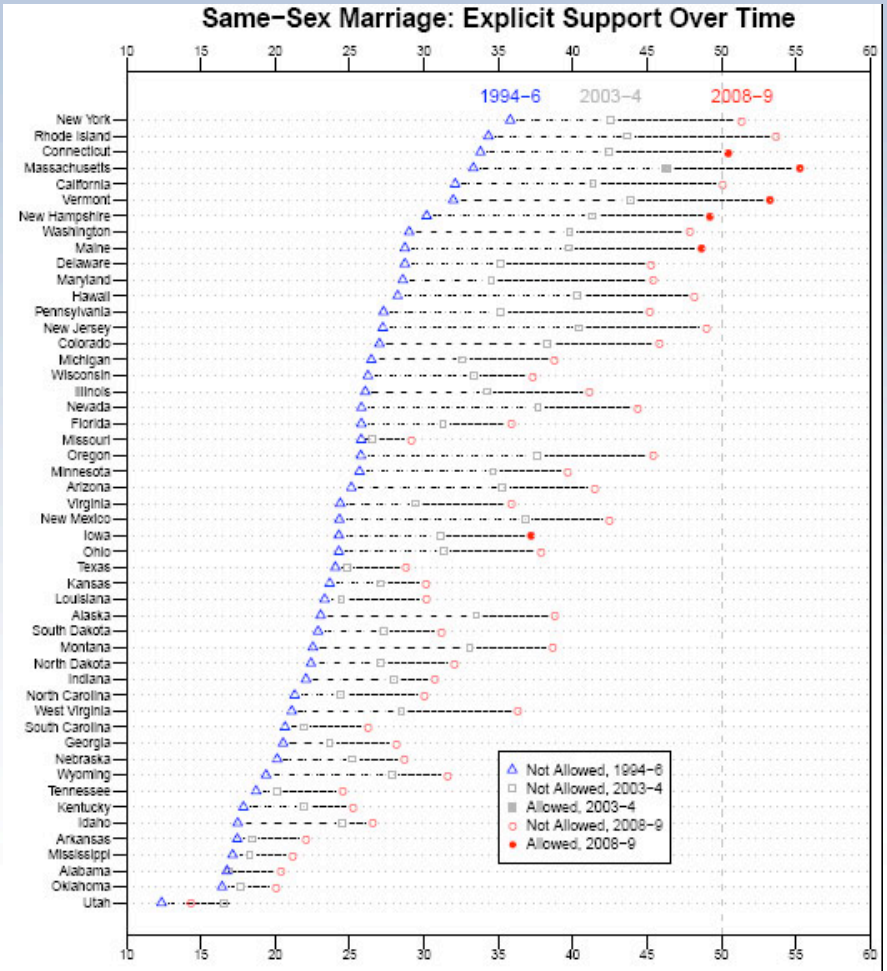


Figure 6: *Same-Sex Marriage Opinion and Policy Over Time (Online Appendix only).* Opinion is estimated using subsets of the poll data from the years indicated. States are ordered by opinion in 1994-6. Note that approximately as much change has occurred in the last four years (solid lines) as the previous eight (dashed lines) and that states with higher levels of early support changed the most. Policy is as of June 2009.

The chart is based on a hierarchical model of polling data by Jeff Lax and Justin Philips using the glmer function in R

REvolution computing
We do the math

Let's do a simulation!

- Is this your birthday?

January

6

REvolution computing

We do the math

Simulating birthdays

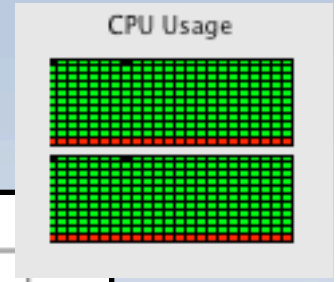
- A simple simulation:

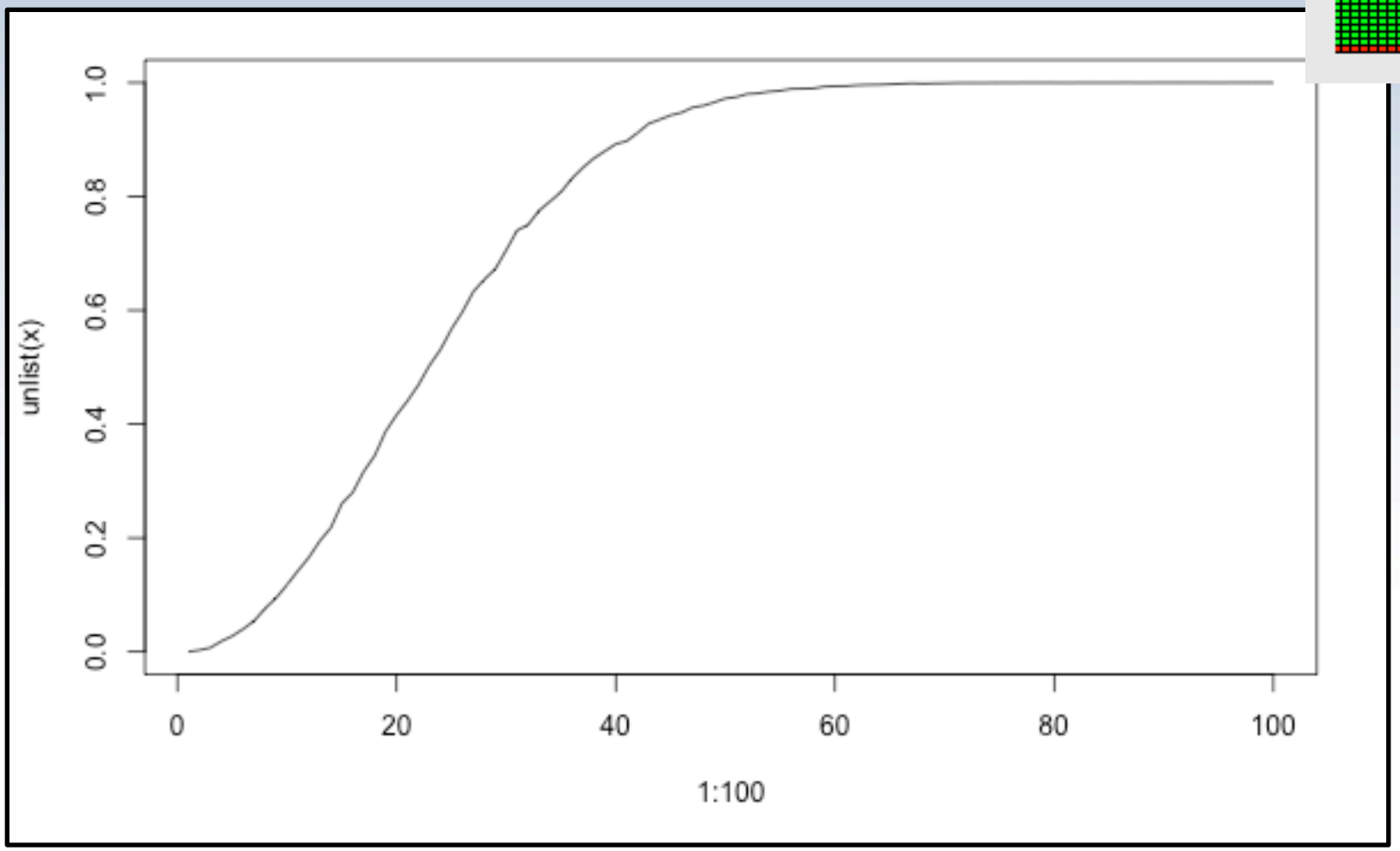
```
birthday <- function(n) {
  ntests <- 10000
  pop <- 1:365
  anydup <- function(i)
    any(duplicated(
      sample(pop, n, replace=TRUE)))
  sum(sapply(seq(ntests), anydup)) / ntests
}

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")
```





REvolution computing
We do the math

Thank You!

- David Smith
  - david@revolution-computing.com, @revodavid
- Danese Cooper
  - danese@revolution-computing.com
- REvolution Computing
  - Booth 405: Free Monkeys at 6PM!
  - www.revolution-computing.com
- *Revolutions*, the R blog
  - blog.revolution-computing.com
- R Project
  - www.r-project.org

REvolution computing
We do the math