























# An intro to the R tidyverse: dplyr,ggplot2, and some Twitter conference data visualization

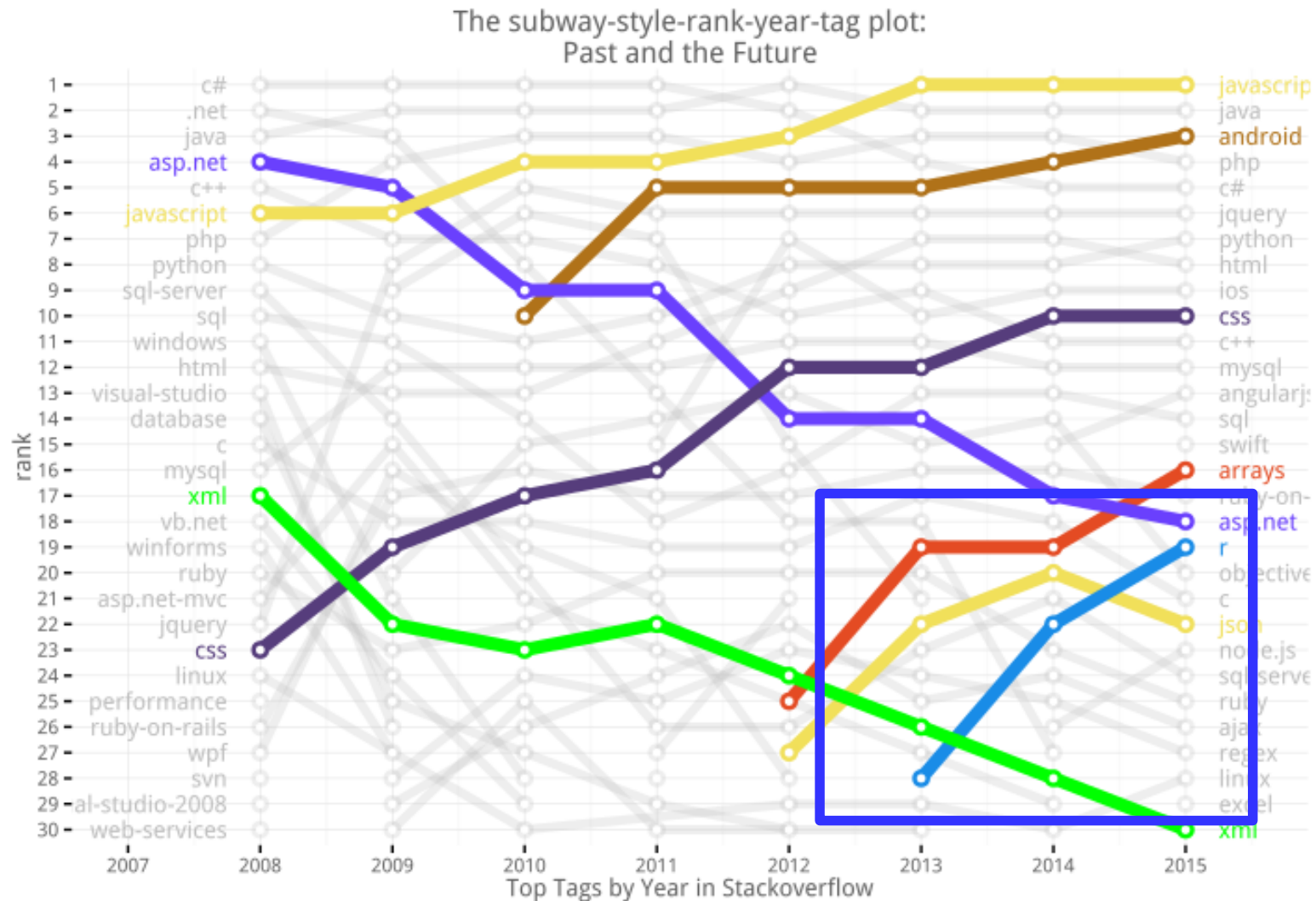
Dr. Thomas E. Keller  
@tek\_keller



# R, bolstered by interest in data-science, has grown

Language Rank	Types	Spectrum Ranking
1. C	  	100.0
2. Java	  	98.1
3. Python	 	98.0
4. C++	  	95.9
5. R		87.9
6. C#	  	86.7
7. PHP		82.8
8. JavaScript	 	82.2
9. Ruby	 	74.5
10. Go	 	71.9

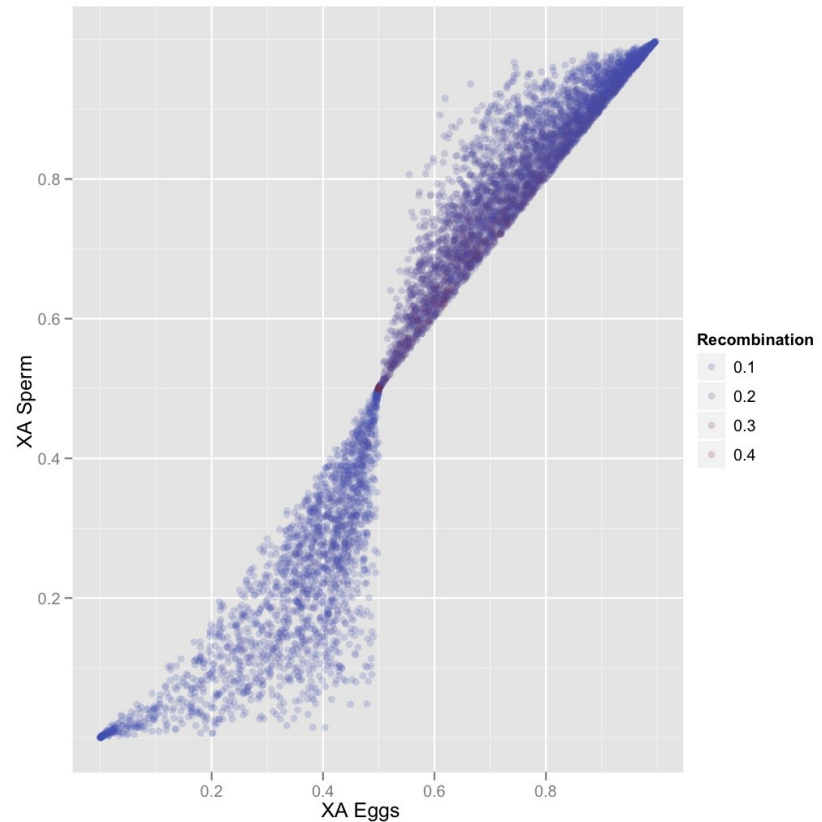
# R gets on the subway



# Why [R]?

---

- Advanced techniques
- A powerful research tool
- Repeatable code
- Price
- PRICE
- PRICE



# R Cran – core binaries & host of the zillions of libraries for specialized stats and visualizations



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

## The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

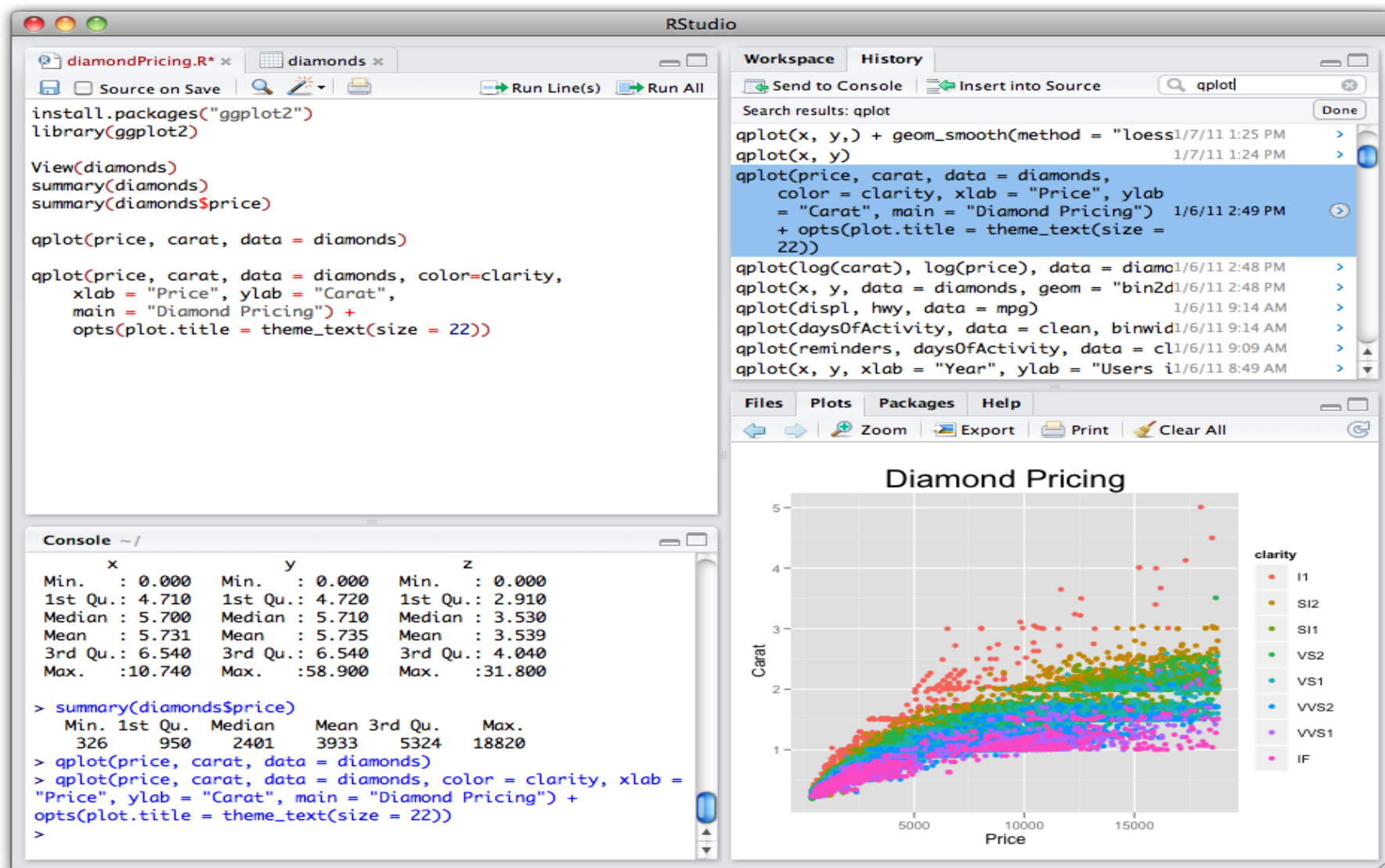
- The latest release (Tuesday 2016-06-21, Bug in Your Hair) [R-3.3.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

### Questions About R

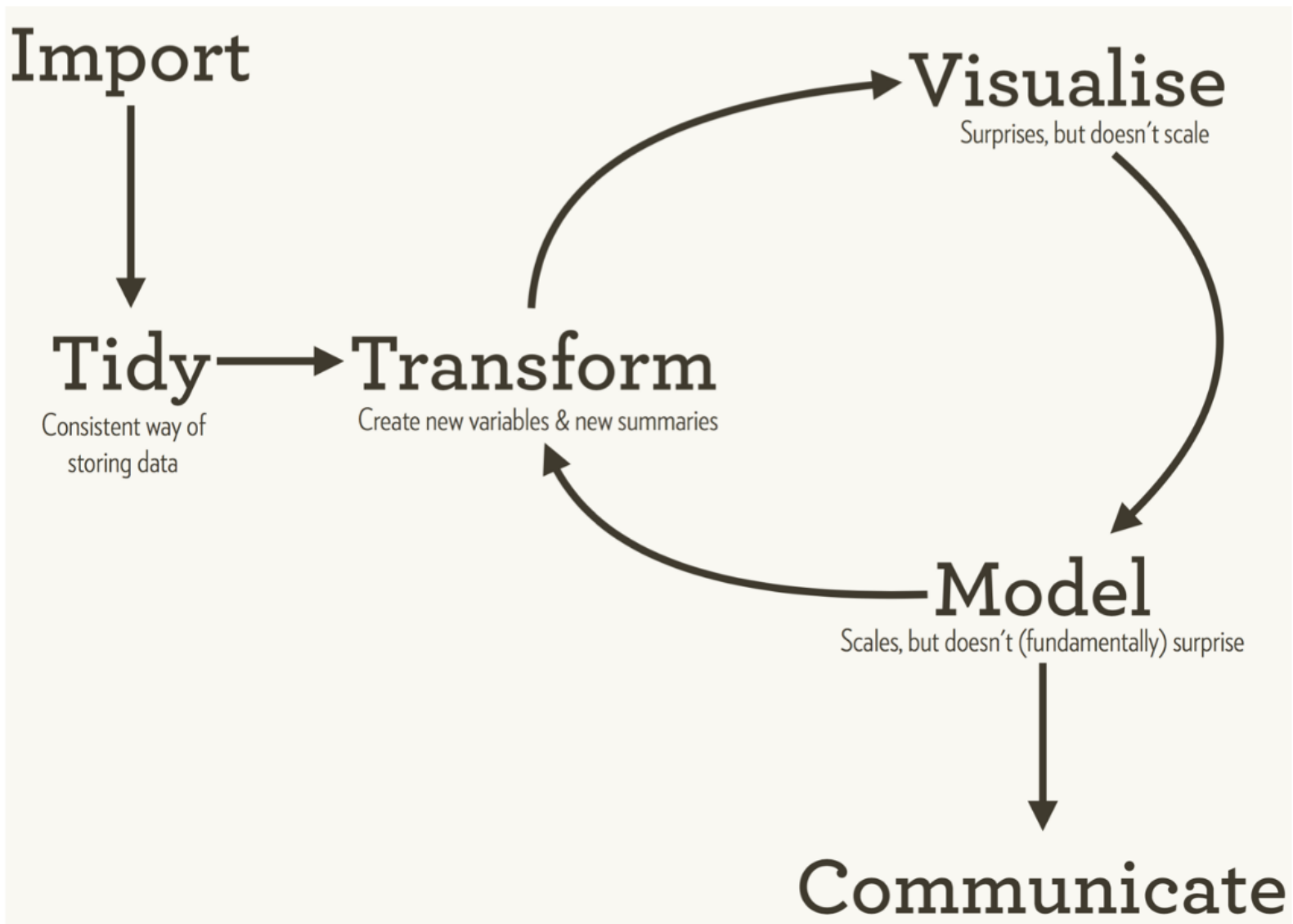
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

# RStudio

- <http://rstudio.org/>



# Statistical analysis cycle (Wickham)





# Twitter Mining with R

- Key libraries
  - twitteR, streamR to download from twitter
    - ROAuth to actually get authorized to use your account
    - twitteR is fine for short term things, but for large, sustained analyses use streamR (downloading 100k stream of #rio2016, etc)
      - TwitteR is limited to 1 week window of twitter search API
  - Tidytext, wordcloud, lubridate, ggraph
    - Breaking down tweets into words and doing sentiment analysis, in a “tidy” fashion
    - Lubridate for date parsing (it is a nightmare in R)
    - Ggraph -plot networks with ggplot

# Resources

- <http://r4ds.had.co.nz/>
- <http://docs.ggplot2.org/current/>
- The ggplot2 book (v2 is out)!
- <https://github.com/hadley/ggplot2-book>
- #rstats on twitter is great, not huge egos, helpful
- Stack overflow