

# Putting the R in NICAR!

(AKA AN INTRODUCTION TO R)

Annie Waldman and Ryann Grochowski Jones, ProPublica  
[annie.waldman@propublica.org](mailto:annie.waldman@propublica.org) | @AnnieWaldman  
[ryann.jones@propublica.org](mailto:ryann.jones@propublica.org) | @ryanngro

**SLIDES LINK:** <https://goo.gl/Cv2Qv3>

**R CODE LINK:** <https://goo.gl/vb3IMU>

**TIPSHEET:** <https://goo.gl/0i0j3X>

# SO WHAT EXACTLY IS R?

- In the 1980s, Bell Labs developed a new statistical software called “S.”
- AT&T went public with the software (“S-PLUS”), but that cost \$\$.
- In 1994, two researchers (Ross Ihaka and Robert Gentleman) developed the first S-like program that wasn’t S, and they called it R.
- Unlike S, R would turn into the statistical language of the people, meaning it would be free and open-source.
- Although R is similar to S-PLUS, it has become something in its own right.
- Over the years, the program has extended beyond the statistics community, and now is used by numerous fields from epidemiologists and research doctors to quants, political scientists, and of course, data journos!





# WHY SHOULD JOURNOS USE R?

- Because it's not hard. Although there is a learning curve, once you get to the first plateau, you'll find that it's much, much better than the alternatives.
- Can deal with millions of data points and huge files (much better than excel).
- Because it's a programming language, you can save your history to make your data manipulation reproducible for your colleagues AND the data community at large.
- Eventually you will hit a wall with other programs, but with R, the range of complexity will allow you to try and eventually master more advanced statistical techniques such as regressions and modeling. The environment encourages experimentation!