

Preliminaries in R

What is R?

What is R?

- A programming language good for data analysis / statistics
- A base package of some software with many available user-created add-ons (packages)
- Free and open-source software
- An interpreted language

R is currently popular in a number of fields, including:

- Statistics / Biostatistics
- Machine learning
- Data journalism
- Ecology
- Financial engineering
- Bioinformatics

What is R?

R is a programming language popular for statistical computing.

“The best thing about R is that it was developed by statisticians. The worst thing about R is that... it was developed by statisticians.”

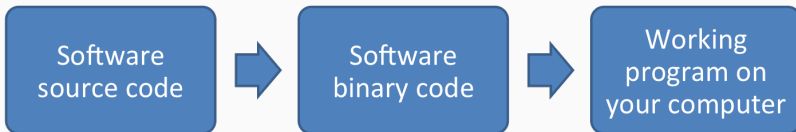
-Bo Cowgill, Google, at the Bay Area R Users Group

Other programming languages popular for statistical computing include:

- SAS
- SPSS
- Matlab
- Julia
- Python

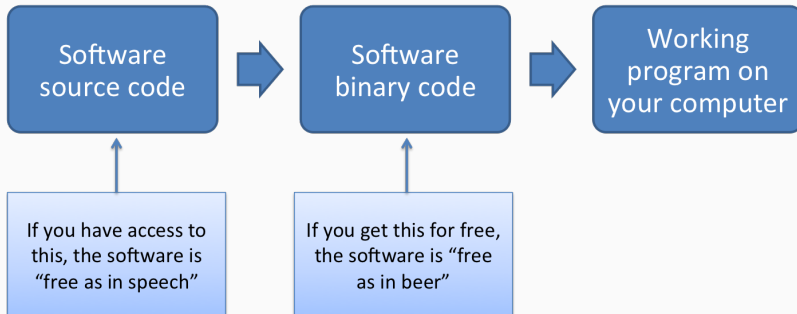
Free and open-source software

How a lot of software is created:



Free and open-source software

A basic sketch of how software can be “free”:

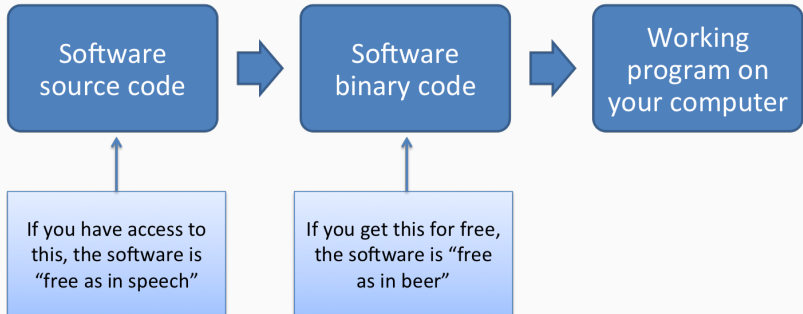


- **Gratis:** Free as in beer
- **Libre:** Free as in speech

Free and open-source software

With open-source software (free as in speech), you can:

- Check out the code to figure out how the software works
- Share the code (and software) with other people
- Make any changes you want to the code



Free and open-source software

“Despite its name, open-source software is less vulnerable to hacking than the secret, black box systems like those being used in polling places now. That’s because anyone can see how open-source systems operate. Bugs can be spotted and remedied, deterring those who would attempt attacks. This makes them much more secure than closed-source models like Microsoft’s, which only Microsoft employees can get into to fix.”

Woolsey and Fox. *To Protect Voting, Use Open-Source Software*. New York Times. August 3, 2017.

Funding agencies are starting to ask for grant proposals to develop open-source tools. For example, a recent call from the NIH asks for:

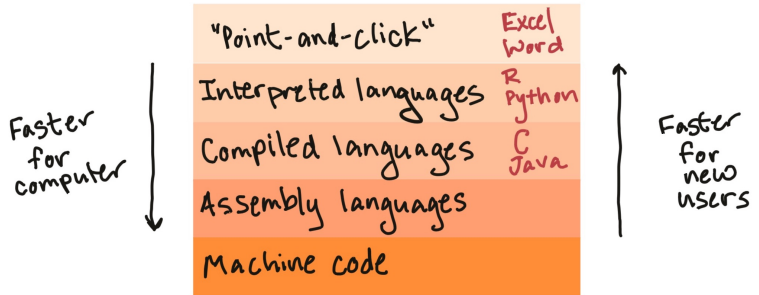
*“**Open-source**, generalizable, and scalable bioinformatics tools”*

NIH RFA-RM-17-012: “Metabolomics Data Analysis and Interpretation Tools (U01)”

Interpreted languages

"Point-and-click"	Excel Word
Interpreted languages	R Python
Compiled languages	C Java
Assembly languages	
Machine code	

Interpreted languages



Interpreted languages

Can get
the computer
to do more



"Point-and-click"	Excel Word
Interpreted languages	R Python
Compiled languages	C Java
Assembly languages	
Machine code	



Can figure
out how to do
something
more quickly
the first time

RStudio vs. R

In this class, I'm having you download both R and RStudio. It's helpful for you to know the difference between the two.

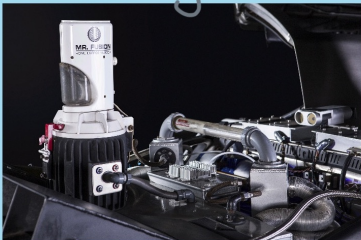
R provides the engine, while **RStudio** provides a nice place to work while using that engine (the leather interior, say).

RStudio vs. R



RStudio
Version: 1.2.1335

"Engine"



"User environment"



RStudio vs. R



Note: When you open **R**, it does have a user interface, it's just not as nice as RStudio's.

RStudio (the software) is an integrated development environment (IDE) for R. You download it separately from R, but it's a “nicer” way to work in R.

This IDE includes:

- An interface with “panes” for key tasks you'll be doing (e.g., one pane with the R console, one for scripts, one to view graphs)
- Code highlighting
- Version control (git) and interface with GitHub
- Tools for Shiny web app development
- Tools for R package development

RStudio (the company) is a leader in the R community. Currently, the company:

- Develops and freely provides the RStudio IDE
- Provides excellent resources for learning and using R (e.g., cheatsheets, free online books)
- Is producing some of the most-used R packages
- Employs some of the top people in R development
- Is a key member of The R Consortium (others include Microsoft, IBM, and Google)

If do not already have them, you will need to download and install both R and RStudio.

- Go to <https://cran.r-project.org> and download the latest version of R for your system. Install.
- Go to the [RStudio download page](#) and download the latest version of RStudio Desktop for your system. Install.
- Defaults should be fine for everything.