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.

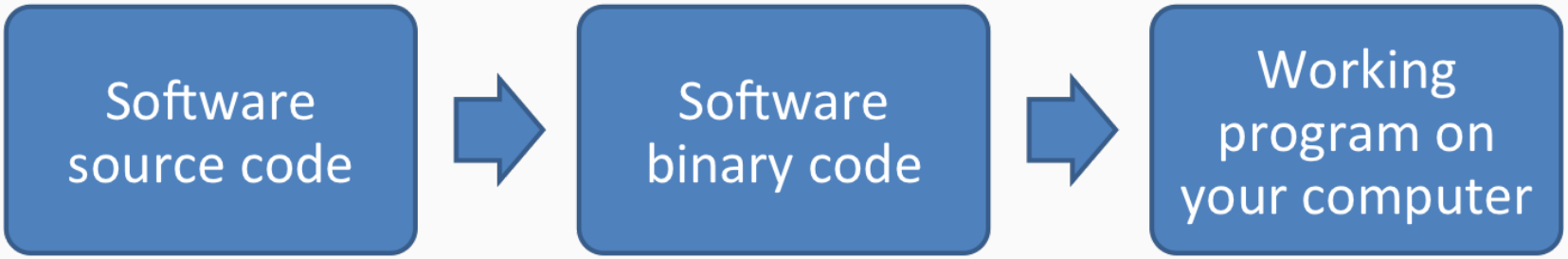
> *"The best thing about R is that it was developed by statisticians. The worst thing about R is that... it was developed by statisticians."*
> *-Bo Cowgill, Google, at the Bay Area R Users Group*

Other programming languages popular for statistical computing include:

- SAS
- SPSS
- Matlab
- Julia
- Python

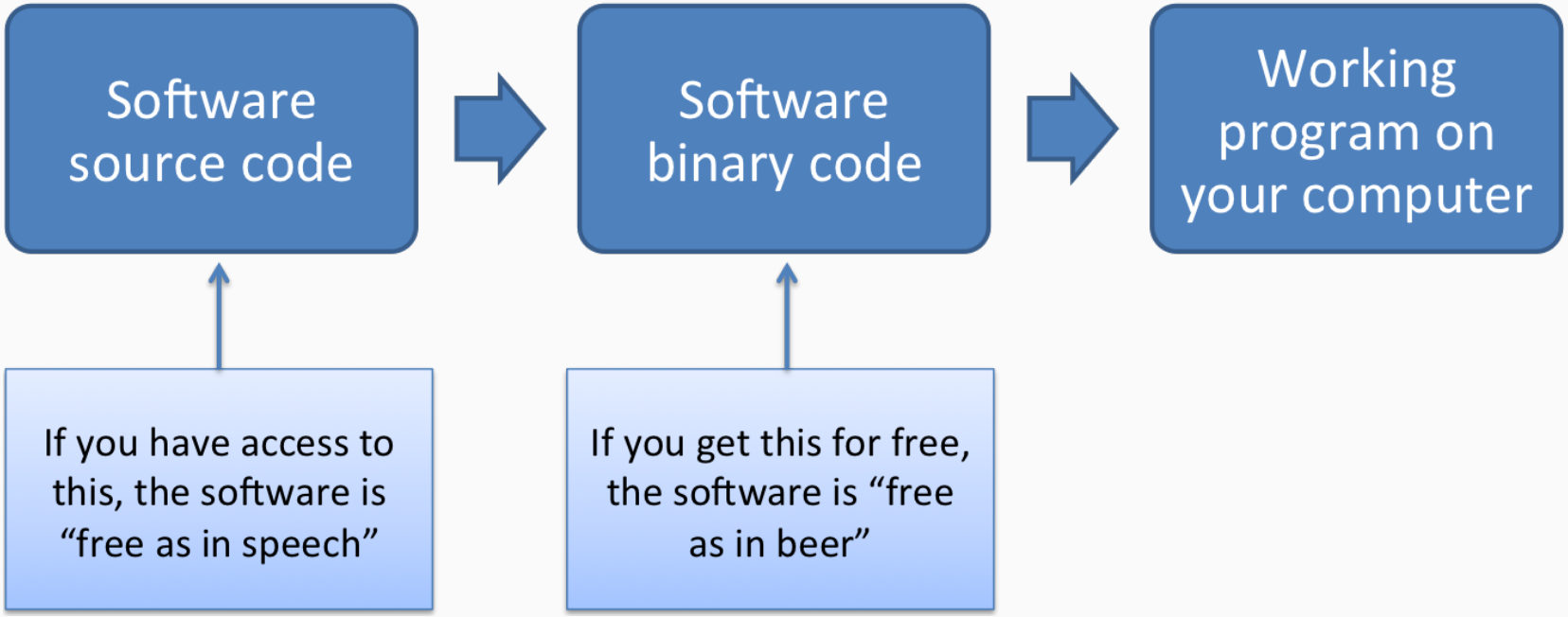
Free and open-source software

How a lot of software is created:



Free and open-source software

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

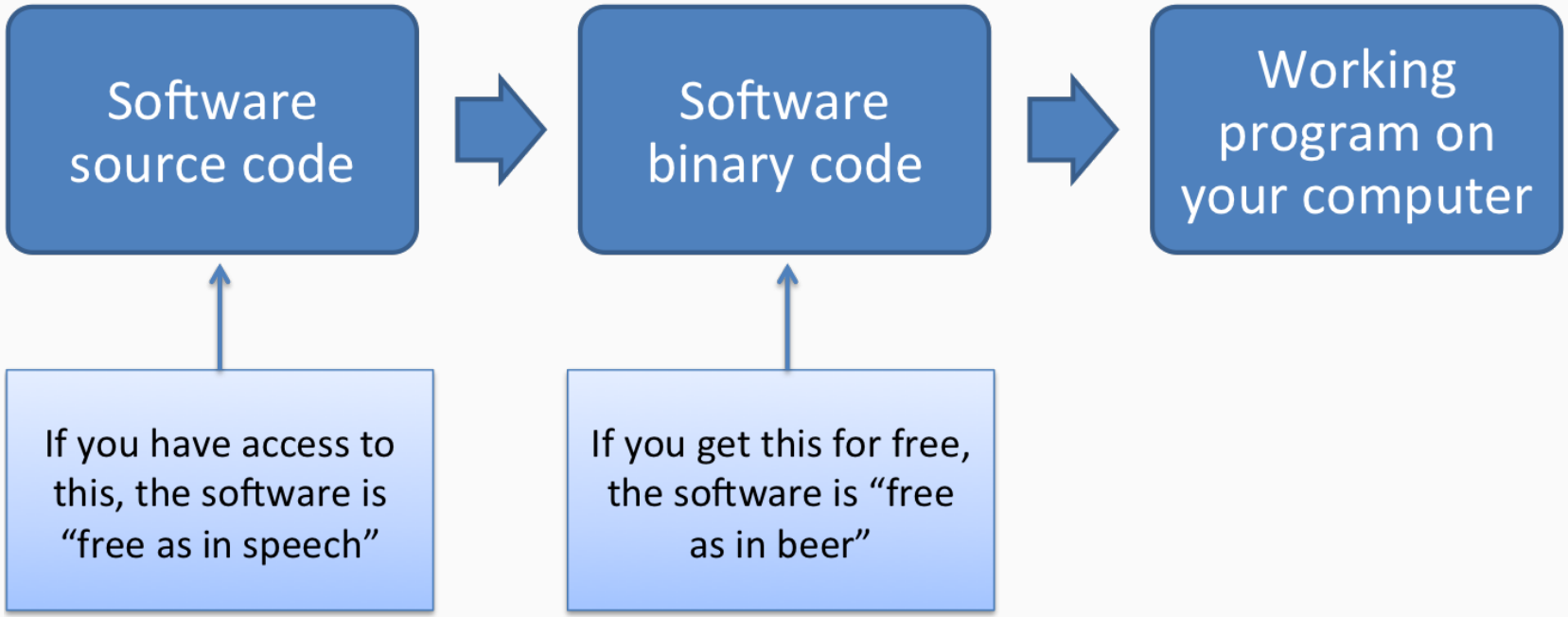


- **Gratis**: Free as in beer
- **Libre**: Free as in speech

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.

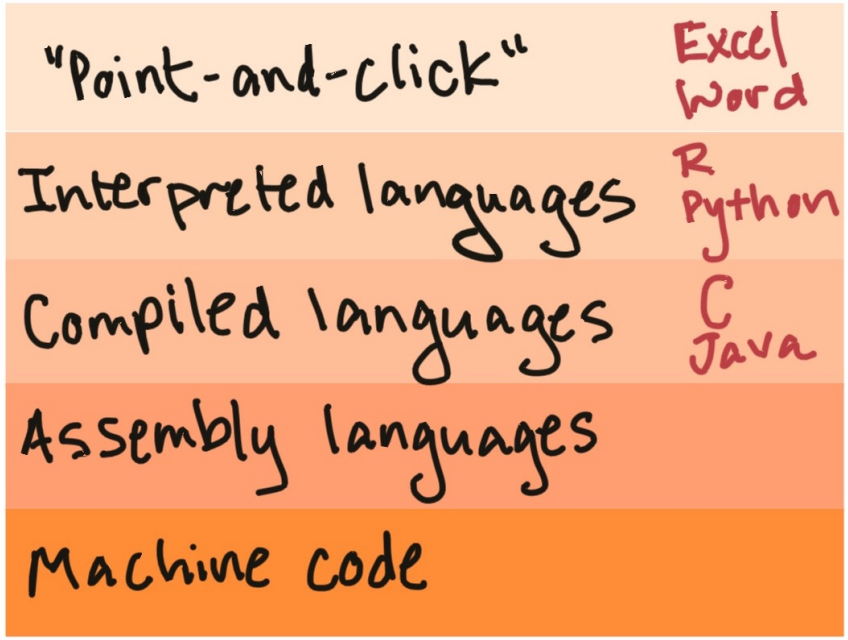
Free and open-source software

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

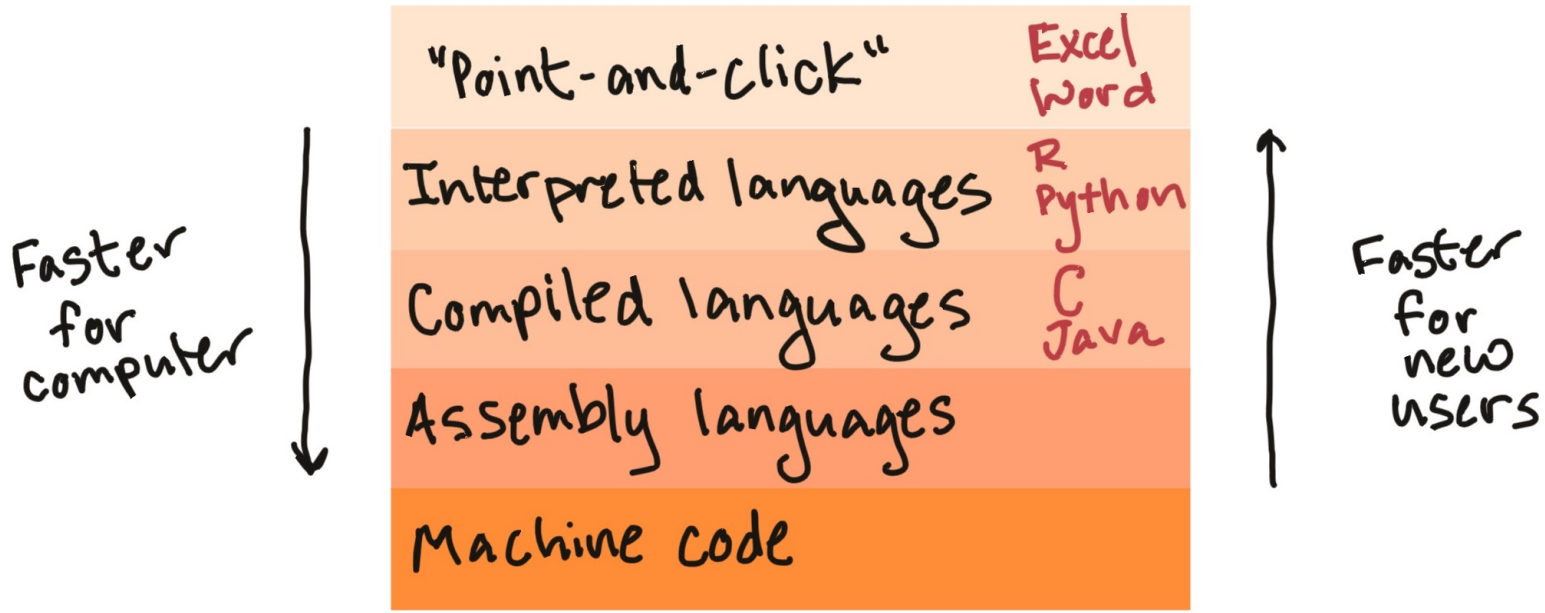
> "***Open-source,*** *generalizable, and scalable bioinformatics tools*"

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

Interpreted languages



Interpreted languages



Interpreted languages