Introduction to R for Data Management and Analysis

Marcel Ramos

Thursday, June 20, 2019

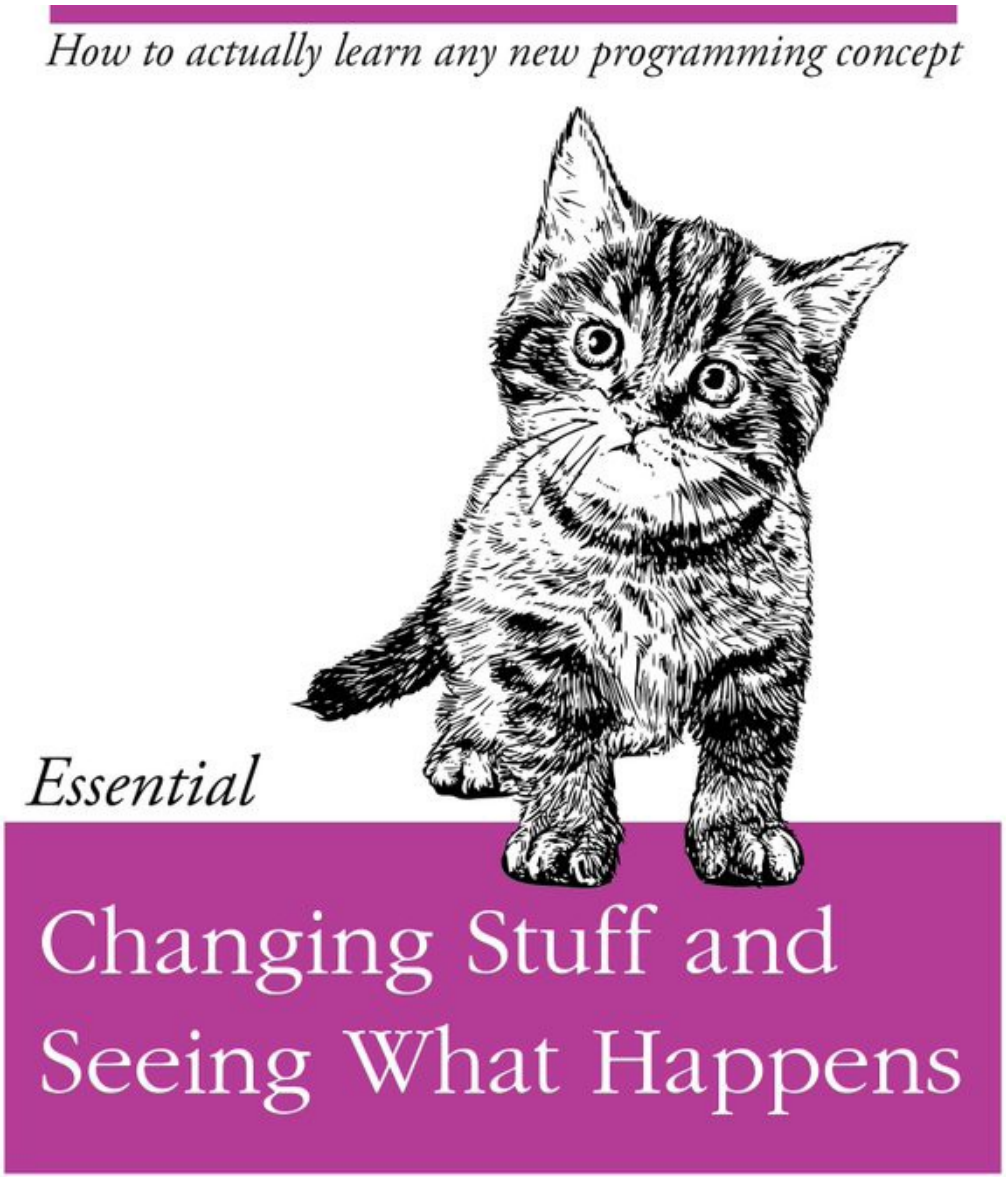
Announcements

- Last in-person class
- Next online classes
  - Data Analysis workflow
  - Reporting and reproducibility
- Review session on Tuesday?

Notes on Tuesday's lecture

- Questions on the material
- Breakdown bigger problems into small manageable steps

Working in R


How to actually learn any new programming concept
Essential
Changing Stuff and
Seeing What Happens

Notes on Tuesday's lecture (cont.)

Pseudocode

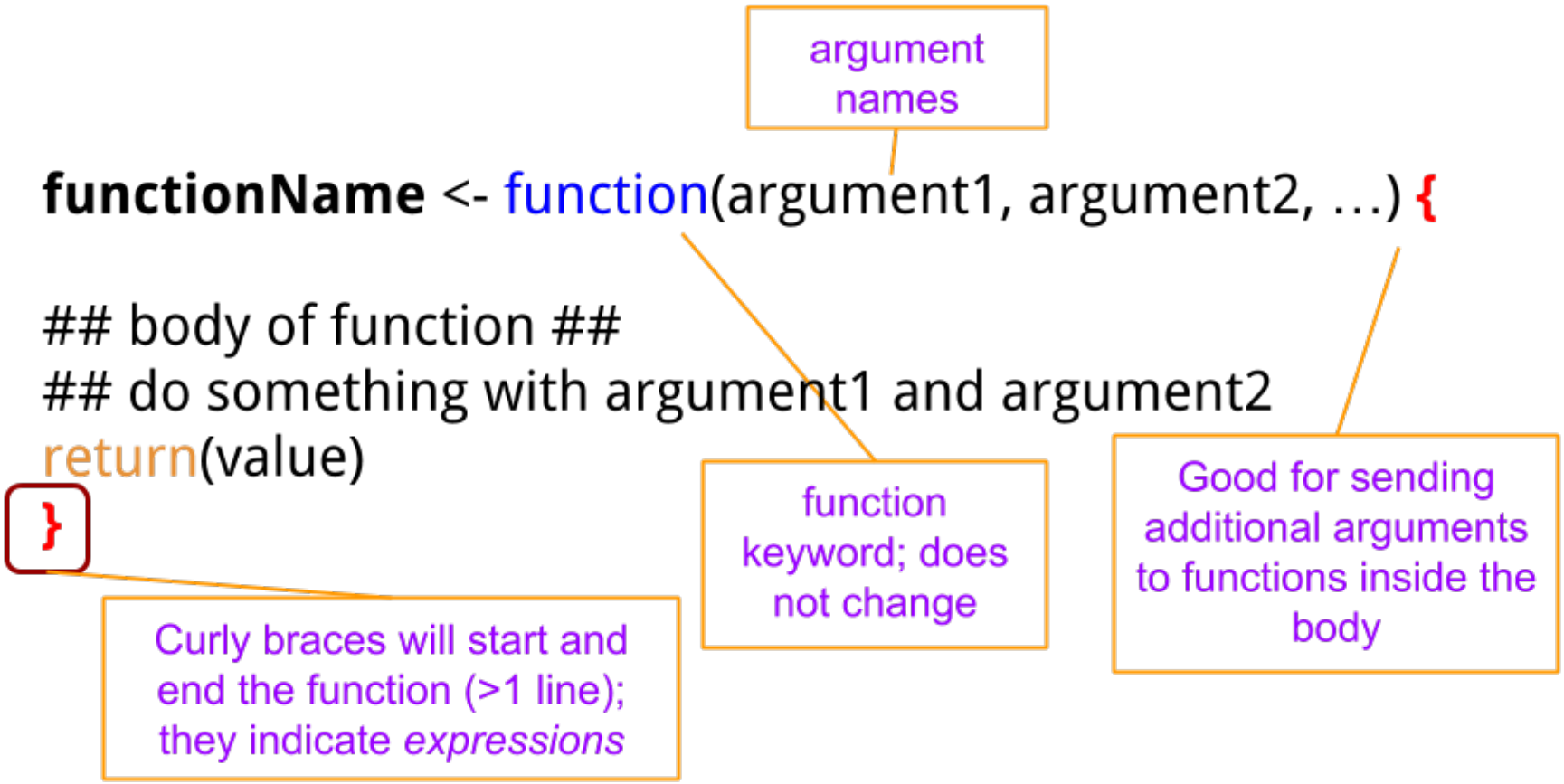
- `for` loop structure

```
for (variable in vector) {
    # < enter code here >
}
```

- `function` structure

```
functionname <- function(argument1 = "default1",
                         argument2 = "default2") {
    myresult <- anotherfunction(argument1, argument2)
        return(myresult)
}
```

More on functions



function(arglist) expr
return(value)

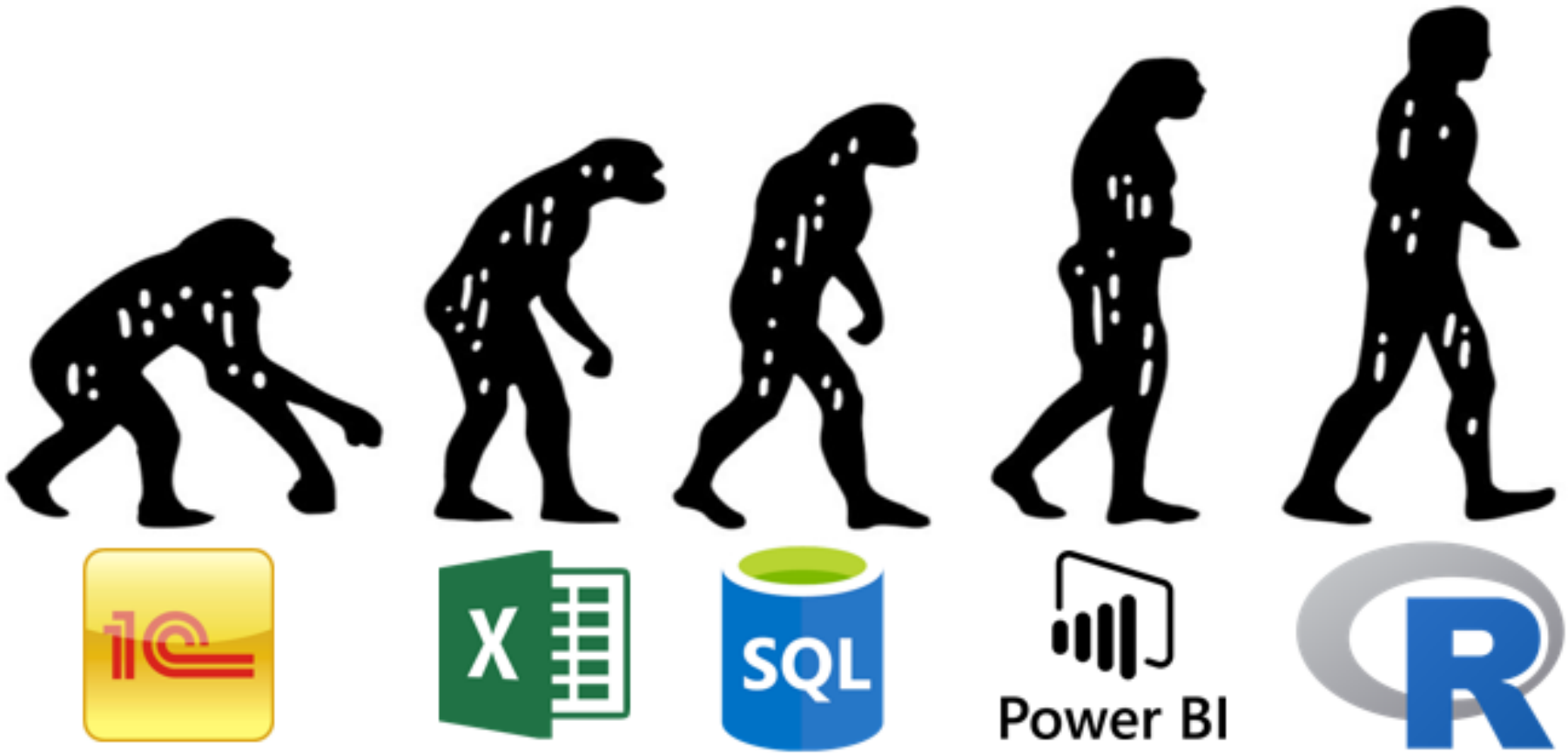
- The body of the function will include the operations to

Notes on Tuesday's lecture (cont..)

- Functions are powerful tools
- Minimize errors
- Create a set of operations to achieve a goal
- Easy to write
  - Predictable input
  - Predictable output
- Loops are useful but are not easily extensible

Why don't we use Excel?

Ranking Statistical Software



Mini Review Session

- Zero-level R Tutorial

Common Errors and Troubleshooting

- R Basics Chapter

Data Analysis

- R is particularly good at statistics
- Packages with new methods get published faster
- Extensibility is an MAJOR advantage compared to other software

The basics

- Frequency tables
- Calculating odds ratios
- `relevel`

Frequencies

- `gmodels` package
- `CrossTable` function
- `table` function

Statistical Tests

- `chisq.test` function
  - categorical 2x2
- `fisher.test` function
  - categorical with correction
- `t.test` function
  - categorical (2 levels) & continuous

Useful functions to apply on model objects

- Functions that work on `lm` class objects
  - `summary`
  - `fitted`
  - `resid`
  - `predict`

Tidy model results with broom

- Use the `broom` package to clean up results from model functions
  - `tidy` - model coefficients
  - `augment` - fitted/residual values and more
  - `glance` - model level statistics

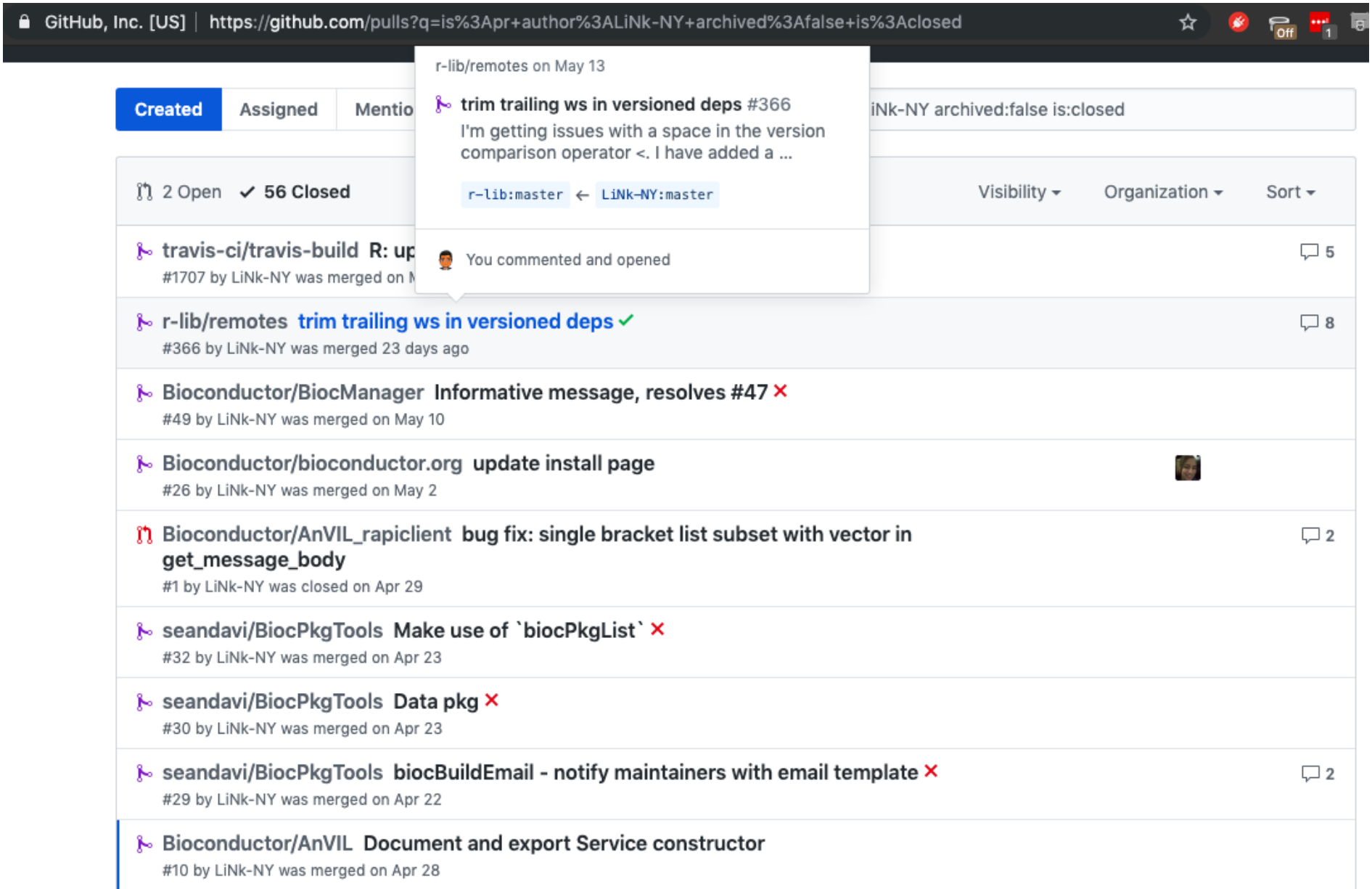
Linear Regression

- `lm` function
- UCLA walk-through

Logistic Regression and Odds Ratios

- `glm` function
- `Odds Ratio` calculation
- UCLA tutorial

Community driven development



GitHub assignment (next week)

- Signup on https://github.com/
- Look for the assignment to be posted under https://github.com/CUNYSPHcode/
- Fork the repository (will contain an `.Rmd` file)
- Upload your `.Rmd` file with the answers
- Create a pull request to submit your `.Rmd` file