STOWERS INSTITUTE®
FOR MEDICAL RESEARCH

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

An introduction to reproducible research using RStudio, knitr and git

Jeff Johnston (jjj@stowers.org)
Zeitlinger Lab
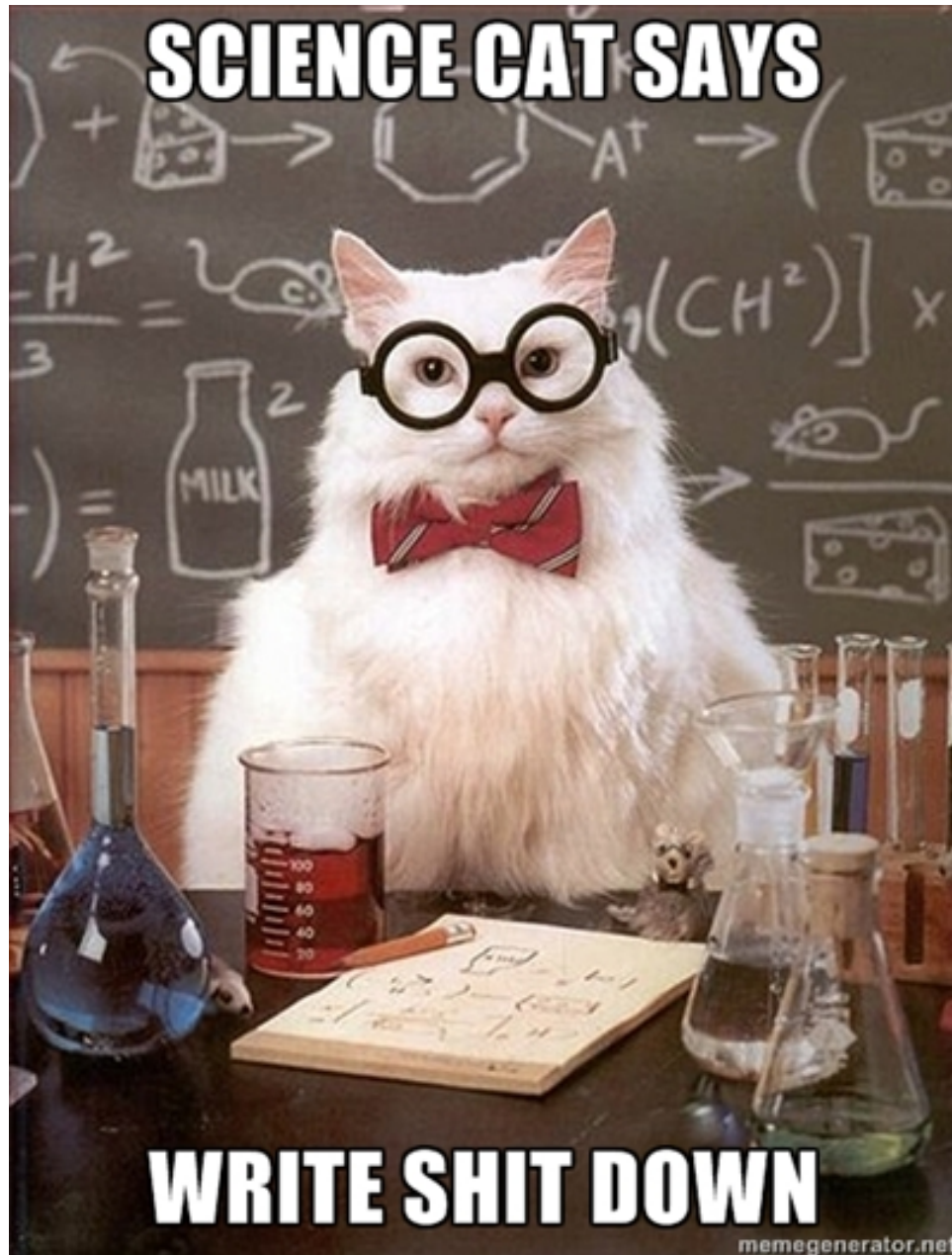Stowers Institute for Medical Research
Kansas City, MO

BioC
July 19th, 2013

What are these tools?

- **RStudio**: an integrated development environment (IDE) for R

- **knitr**: an R package designed to make literate programming easier

- **git**: a version control system for tracking changes to text and software code

The first (and most important) lesson

Literate Programming

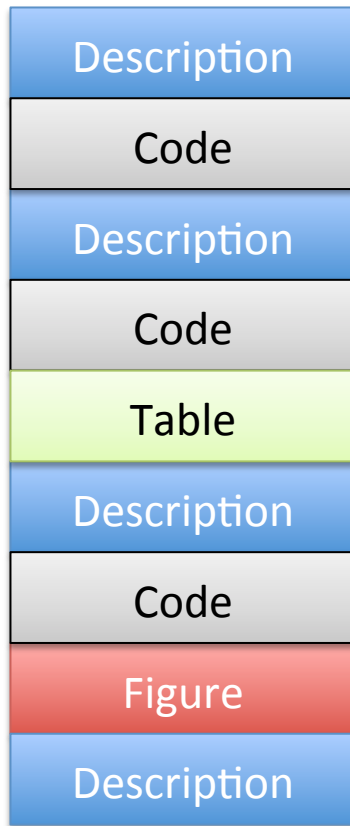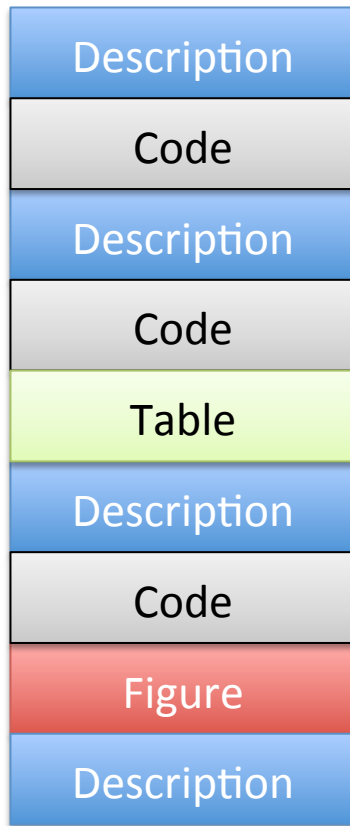| Description |
| Code |
| Description |
| Code |
| Table |
| Description |
| Code |
| Figure |
| Description |

In literate programming, an analytical document is composed of a descriptive narrative "woven" together with software code and computed results.

| Description |
| Code |
| Description |
| Code |
| Table |
| Description |
| Code |
| Figure |
| Description |

Advantages

- A single document both describes and performs the analysis

- Enforces reproducibility

- Gets you in the habit of Good Science

Version Control

- Think of Microsoft Word's *Track Changes* feature

- Version control allows you to:
  - Save the "state" of your project at any time
  - Keep a log of all your changes
  - Return to or review previous versions of your analyses
  - More easily work with others on the same project
  - *Lots* more!

Agenda

- Start a new version-controlled analysis project in **RStudio**

- Learn about knitr, markdown, tables, figures and citations

- Perform a quick bioinformatics analysis as a **knitr** document

- Occasionally "save" our progress with **git**