Increasing Reproducibility by Freezing R

Jeremy Stephens
Vanderbilt Biostatistics
Statistical Computing Series

November 16, 2012

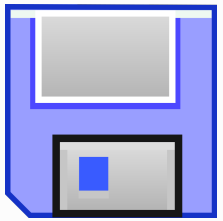
What Is Reproducibility?

Measure of how easy it is to reproduce something

Why Be Reproducible?

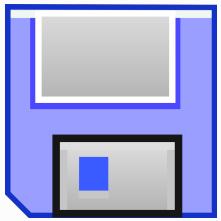
- Decrease job security (j/k)
- Fewer mistakes
- Science!
- Proof of due diligence

Programming Reproducibility Basics



Saving your code

Sharing your tools



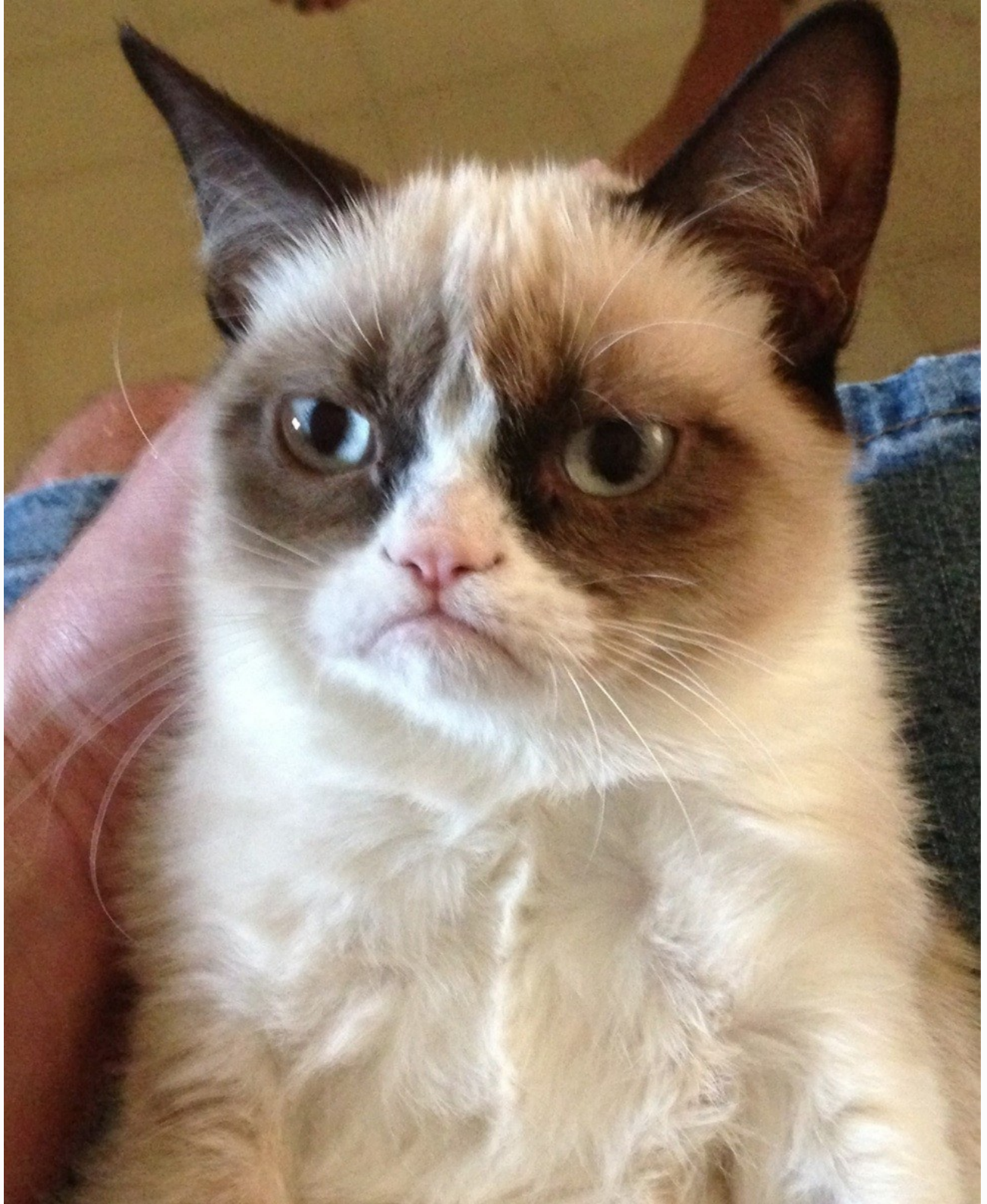
Saving Your Code

- Script files
- Sweave / knitr for reports
- Packaging

Sharing Your Tools

- R
- CRAN for packages

CHANGE
Change



But change is good, right?

Freezing R

- Advantages
  - Insulation from changes
  - Guarantee same behavior
  - Ability to upgrade when you choose
- Disadvantages
  - No benefits from bug fixes or new features
  - Harder to share code

How?

- Install multiple versions of R
- Set R version to use for each project
- Local R package directories

Renv (pronounced “arr env”)

- Command-line tool
- Lives in your home folder
- Manage multiple versions of R
- Switch R versions automatically

Renv commands

- Renv global
- Renv local
- Renv shell
- Renv install

Freezing Packages

- Use project .Rprofile to specify local library path
- install.packages() to the local path

Workflow for new projects

```
$ mkdir my_project
$ cd my_project
$ Renv local 2.15.2
$ mkdir .Rlib
$ echo ".First <- function() {
.libPaths('.Rlib') }"
$ R
```

Retrofitting old projects

- Find modification time of your code files
- Install version of R that was active at that time (check out the archive page on CRAN)
- Install old packages via `R CMD INSTALL`

Tips for Upgrading

- Testing (testthat)
- Comparing results (diff, meld, kompare)