A SAS-to-R Success Story

Beth Atkinson

February 3, 2018

MAYO CLINIC

Situation

MAYO CLINIC

Who we are

- Large Division: 300 staff in 3 locations
  - Cancer Center Statistics, Clinical Statistics, Computational Genomics, Medical Informatics
  - 100's small projects <=> large grants
- Historically
  - SAS Shop since 1974
  - Splus/R used occasionally by statisticians

MAYO CLINIC

Issue

- In 2014: Re-negotiated SAS license
- Led to the question of whether we can
  - Decrease our SAS dependency
  - Better serve our clients

MAYO CLINIC

Assessment

Infrastructure

- 3 Linux systems (Centos 6)
- 3 versions of R
- Common package repositories
  - CRAN, Bioconductor, Github, local
- 470 R users

MAYO CLINIC

Challenges

- Established SAS infrastructure
  - Training, macros, workflows
- R Infrastructure strain, rapid growth/changes
  - Package, R version management
- Time constraints (leaders, users)
- Different learning styles, levels of R knowledge
- Newer R users:
  - Data scientists, informatics specialists (NLP, Bioinformatics), IT, MDs, lab techs

MAYO CLINIC

Personal Barriers

- SAS known entity
  - Trust issues, misconceptions
  - Personal code repository
  - "My team uses SAS"
- Rapid change
  - This *used* to work
  - The answer changed!
  - Now *this* is a better way to do X

MAYO CLINIC

Personal Barriers

- Overwhelming number of packages, choices
  - SAS more stable
- Inconsistent levels of R documentation
  - Less standardized
  - "Just Google it" issues
  - Cryptic error messages, debugging challenges

MAYO CLINIC

Plan

MAYO CLINIC

Infrastructure

- Division $
- Identified group of R champions
  - Interested in different facets of R, excited by R
  - R stewards
  - Additional IT support
- RStudio Server Pro
- Identified popular local SAS macros, translated to R
  - New packages: arsenal (CRAN), dart (internal, data retrieval)
  - Took advantage of fresh start

MAYO CLINIC

Education/Support

- Education, multiple voices & formats:
  - Expanded new employee training: R time = SAS time
  - Seminars, videos, drop-in sessions
- R is fun!
  - "Anything [SAS/R] can do, [R/SAS] can do better"
- Support:
  - On call support, help desk
  - Distribution lists, regular R tips & updates

MAYO CLINIC

Reasons to change

- Motivation:
  - Survival class
  - Data challenges
  - Bookclubs (data mining, hierarchical models)
  - Markdown
  - Shiny

MAYO CLINIC

Results

MAYO CLINIC

Did anything change?

- More infrastructure
- People motivated to learn
  - Phil Bowsher (Shiny) came in June - 80 in room, 93 online
- Larger base of "experts"
- More primarily using R

MAYO CLINIC

Why still using SAS?

- Time, cost constraints
  - Small project pressures
- Low motivation
  - Change is hard, takes investment
- Data retrieval
- Data cleaning
- Certain tools painful in R
  - Mixed Effects, GEE

MAYO CLINIC

Plans for upcoming year

- Mock projects - putting it all together
  - Add tools to toolkit
  - Creating clean, tidy data
- User requested education
  - Recent query identified “function writing 101”, “how to find help”

MAYO CLINIC

Lessons learned

- Need to motivate project leads
- Encourage creativity, fun
- Important to ask for feedback
- Target presentations to different types of R users
- Larger conversation - what skills do we need in the future?

MAYO CLINIC