

Managing and deploying R packages across an
organisation
– Cat-herding 101

Mike K Smith
(Pfizer UK Ltd.)

Or...

How to get oldR with style

**How many of you have
R-3.5.1 installed?**

How many kept their
previous installation
(separate)?

**How many kept their
libraries of packages
for each version separate?**

Your R installation process (may) be like



**If you had to re-run an analysis
you did >10 years ago (using R),
*could you?***

Our R on HPC GRID



**We have R-1.9.1 available on
the HPC GRID (2004)**

R on colleagues desktops



Desktop R

- **We qualify and release versions of R for desktop, with a “snapshot” set of packages.**
- God alone knows whether colleagues get their R from this version (or CRAN).
- And who knows what packages they have, what versions or where they got them.

Desktop R

- We qualify and release versions of R for desktop, with a “snapshot” set of packages.
- **God alone knows whether colleagues get their R from this version (or CRAN).**
- And who knows what packages they have, what versions or where they got them.

Desktop R

- We qualify and release versions of R for desktop, with a “snapshot” set of packages.
- God alone knows whether colleagues get their R from this version (or CRAN).
- **And who knows what packages they have, what versions or where they got them.**

My job



How do I (try to) help?

- **Set up and maintain miniCRAN “snapshot” of packages for a given release.**
 - **Including CRAN, Github and internal packages**
- Following installation of (base) R, colleagues use installation script to install ALL packages.
 - That way, they don't have to run `install.packages()`, just call `library(...)`
- Provide functions that allow colleagues to check, update, add, install packages to reconcile what they have compared to our miniCRAN list.
- Ensure that desktop and HPC GRID package list matches.
 - Develop code on desktop, run for reproducibility on GRID.

How do I (try to) help?

- Set up and maintain miniCRAN “snapshot” of packages for a given release.
 - Including CRAN, Github and internal packages
- **Following installation of (base) R, colleagues use installation script to install ALL packages.**
 - That way, they don’t have to run `install.packages()`, just call `library(...)`
- Provide functions that allow colleagues to check, update, add, install packages to reconcile what they have compared to our miniCRAN list.
- Ensure that desktop and HPC GRID package list matches.
 - Develop code on desktop, run for reproducibility on GRID.

How do I (try to) help?

- Set up and maintain miniCRAN “snapshot” of packages for a given release.
 - Including CRAN, Github and internal packages
- Following installation of (base) R, colleagues use installation script to install ALL packages.
 - That way, they don’t have to run `install.packages()`, just call `library(...)`
- **Provide functions that allow colleagues to check, update, add, install packages to reconcile what they have compared to our miniCRAN list.**
- Ensure that desktop and HPC GRID package list matches.
 - Develop code on desktop, run for reproducibility on GRID.

How do I (try to) help?

- Set up and maintain miniCRAN “snapshot” of packages for a given release.
 - Including CRAN, Github and internal packages
- Following installation of (base) R, colleagues use installation script to install ALL packages.
 - That way, they don’t have to run `install.packages()`, just call `library(...)`
- Provide functions that allow colleagues to check, update, add, install packages to reconcile what they have compared to our miniCRAN list.
- **Ensure that desktop and HPC GRID package list *matches*.**
 - **Develop code on desktop, run for reproducibility on GRID.**

Colleagues range in attitudes...



“I don’t care, just give me the packages I need to get my work done...”

“Get out my way, I know what I’m doing...”

Yes, I know other options are available

- Microsoft R Open (MRO) + MRAN + checkpoint()
 - I LIKE the concept.
 - BUT Github, internal packages?
- packrat
 - Yes, BUT... consistency in packages ACROSS analyses AS WELL AS reproducibility.
 - Desktop vs HPC GRID
- RStudio Package Manager
 - Somewhere between packrat and miniCRAN and MRAN?

But ultimately...

- No matter what we do, we need to:
 - be able to move quickly without breaking things
 - push packages out in a controlled manner
 - square up desktop and HPC GRID packages
 - be able to support ALL colleagues
 - Research – latest, greatest, **NOW**.
 - Regulatory work – controlled, qualified.
 - facilitate reproducibility
 - Potentially for **YEARS**...

Cats herded

