R from start to finish: Organizing your dissertation work with a reproducibility mindset using R and RStudio

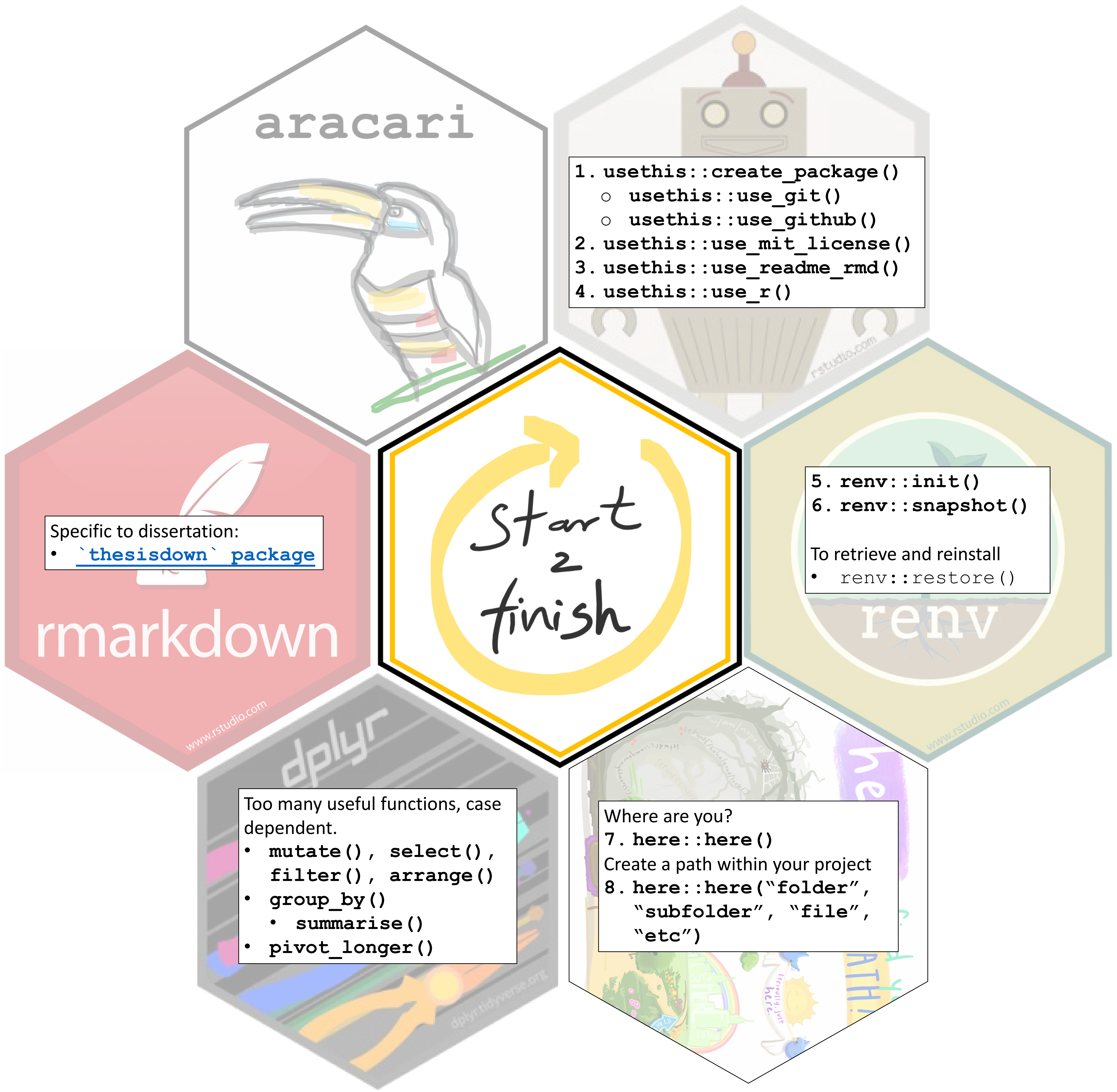
Javiera Rudolph @javirudolph

Introduction: Have you ever experienced running old code and having it break? Or found the code associated to a scientific article, and when trying to understand it or run it, realize it is almost impossible to figure out?
As an ecologist, and as many of us, I started my journey with R and data analysis as a self-directed adventure, first learning to code, and later realizing about the importance of reproducibility. Particularly associated with R programming, there is an overwhelming number of resources for reproducible research. This poster is meant to be a resource, a short guide and starting point for setting up a reproducible workflow in R.

- **Git and GitHub** – you don't need these two for setting up a project in R but maintaining version control is highly recommended and fundamental for reproducibility. This means that there is a history for your code and analysis. Connecting Git and GitHub to RStudio is system dependent, a good resource for this process can be found in happygitwithr.com
- **Build Packages** – R packages are great structures to organize your work. Building a package allows you to keep everything together, forces you to document your code, and follow certain conventions. You can dive deep into R packages with Hadley Wickham's book, or for a quick intro check my tutorial at: https://github.com/javirudolph/rladiesgnv

The tools:

1. **'usethis' package:** Helps with setup and automates tasks related package or project development. Very chatty package, will tell you what it is doing and ask for input.
   - 'roxygen2' package works with `usethis` for creating descriptive documentation of your package and functions.
2. **'renv' package:** Keep track of the packages used and the package version in a log, which can be tracked with version control and then restored in different machines. No packages breaking your analysis.
3. **'here' package:** if you start your scripts with 'setwd()' you need the **here** package. Avoid setting directory and file paths specific to your machine, making sure they are relative to the working project or package. Very useful when having .Rmd files in subfolders.
4. **`dplyr` package:** work with your data in R using scripts and avoid clicks in spreadsheet. Scripts are reproducible and provide a history of how you've manipulated the data. Avoid merged cells and colors in spreadsheets, have the first row of headers and if possible, make every row a different observation, with columns being different variables.
5. **`rmarkdown` package:** write your reports and even your manuscript with R markdown, which you can also track changes of it with git and avoid multiple 'final' versions. You can use various templates associated to journals and even use it to write and format your dissertation.
6. **`aracari` package:** just as an example, this package is in progress and corresponds to some of my dissertation work understanding animal movement and seed dispersal. It was built following the process outlined in this poster.



```
1. usethis::create_package()
   o usethis::use_git()
   o usethis::use_github()
2. usethis::use_mit_license()
3. usethis::use_readme_rmd()
4. usethis::use_r()
```

Specific to dissertation:
- `thesisdown` package

```
5. renv::init()
6. renv::snapshot()
```

To retrieve and reinstall
- `renv::restore()`

Too many useful functions, case dependent.
- `mutate(), select(), filter(), arrange()`
- `group_by()`
  - `summarise()`
- `pivot_longer()`

Where are you?

```
7. here::here()
```

Create a path within your project

```
8. here::here("folder",
   "subfolder", "file",
   "etc")
```

Check this project's site for additional resources at:
https://javirudolph.github.io/start2finish/

UF Department of Biology
UNIVERSITY of FLORIDA