Hosting Data Packages via ‘drat’: A Case Study with Hurricane Exposure Data

Data reproducibility session, UseR! 2017

Brooke Anderson, Colorado State University
✉: brooke.anderson@colostate.edu
🐦: @gbwanderson
.ribbon: github.com/geanders

Dirk Eddelbuettel, Debian and R Projects, Ketchum Trading
✉: edd@debian.org
🐦: @eddelbuettel
.ribbon: github.com/eddelbuettel
Goal: Create a publicly-accessible package with historical data (1988–present) on county-level exposure to Atlantic basin tropical storms in the United States (with exposure based on distance to storm track, rain, wind, floods, or tornadoes).
Motivation

“Data" package

hurricaneexposedata
Hosted with drat.

“Code" package

hurricaneexposure
Available through CRAN.
Motivation

"Data" package

hurricaneexposedata
Hosted with drat.

"Code" package

hurricaneexposure
Available through CRAN.

Challenges

1. Host a package with drat.

2. Connect a CRAN package with a data package hosted with drat.
Motivation

“Data" package

hurricaneexposuredata
Hosted with drat.

“Code" package

hurricaneexposure
Available through CRAN.

Challenges

1. Host a package with drat.

2. Connect a CRAN package with a data package hosted with drat.
Hosting a package with drat

R repositories and drat
An R package repository is a network-accessible directory structure, and drat helps you build and maintain your own.

Extensive resources
- Drat Basics for Package Users
- Drat Basics for Package Authors
- Drat FAQs
Packages hosted on drat repositories

For package users, how is a drat package different?

- “Not at all”, really.
- Tell R about the additional repository via
  - repos argument directly in `install.packages()` et al.,
  - setting another `repos` field in `options()`, or
  - using drat helper functions `addRepo()` or `drat:::add()`
- Full details about `repos` in `help(download.packages)`

For package maintainers, how is a drat package different?

- Need to create a personal repository and make it available to users
- Can use functions in the drat package to manage this repository
How is GitHub involved?

- Hosting on GitHub particularly easy because
  - Every GitHub repo has an (optional) web presence
  - A GitHub username plus `drat` is a unique URL
- So if we assume `drat` as the (GitHub) repository name, then we can reference it with just the username:
  ```
  addRepo("geanders")
  ```
Coding a CRAN package to use a drat package

"Data" package

'hurricaneexposedata'
Host with drat.

"Code" package

'hurricaneexposure'
Available through CRAN.

Challenges

1. Host a package with drat.
2. Connect a CRAN package with a data package hosted with drat.
“Unfortunately it is out of our area of expertise to do deep level troubleshooting on an individual’s computer to find the cause of the issue. In this case I would recommend using a colleague’s computer and see if that is successful in creating a submission.”

– Example of recent non-R technical support
Technical support experience, R

1. Search GitHub
1. Technical support experience, R

2. Brooke Anderson
@gbwanderson

Can anyone point me to a CRAN package that imports a drat repo package? I'm trying to fix `build_win` build errors. @eddelbuettel #rstats

11:16 PM - 17 Jan 2017

RETTWEETS  LIKES
2         2
Technical support experience, R

1. Search GitHub

2. Brooke Anderson
   @gbwanderson

   Can anyone point me to a CRAN package that imports a drat repo package? I'm trying to fix `build_win` build errors. @eddelbuettel #rstats

   RETWEETS  |
   -------------
   2

   LIKES  |
   -------------
   2

   11:16 PM - 17 Jan 2017

3. Hosting Data Packages via drat: A Case Study with Hurricane Exposure Data
   by G. Brooke Anderson, Dirk Eddelbuettel
How hard is it to code a CRAN package to “depend” on a drat package?

- Not very
- The DESCRIPTION file has an `Additional_repositories` field
  - Optional packages can reside on `Additional_repositories`
  - Mandatory packages (`Imports`, `Depends`, `LinkingTo`) cannot
  - But `Suggests` can point there – feature we use

- Repos created with drat are a good use case for `Additional_repositories`
- But also ensure you test for presence of optional package
How hard is it to code a CRAN package to “depend” on a drat package?

- Need to add code in the CRAN package to do three jobs:
  - Add infrastructure in the DESCRIPTION file
  - Customize behavior when the package is loaded
  - Condition code based on availability of drat package (functions, vignette, examples)
Add infrastructure in the DESCRIPTION file

- Include the name of the data package in the “Suggests" field.
- Give the address of the drat repository with the data package in the “Additional_repositories" field.

For example, the hurricaneexposure DESCRIPTION file includes:

Suggests: hurricaneexposuredata (>= 0.0.2)
Additional_repositories: https://geanders.github.io/drat
Coding a CRAN package to use a drat data package

Customize behavior when the package is loaded

- These steps can be included in “zzz.R" (or similar) file, by writing `.onLoad` and `.onAttach` functions.
- Check if the user has the data package and create an environment that stores the result.
- If the user does not have the data package, give a start-up message telling the user how to install it.
Coding a CRAN package to use a drat data package

Condition code based on availability of data package

- Condition code in functions that need access to the data package (e.g., helper function can be written in the package).
- Condition code in documentation examples to only run if the data package is available.
- Condition vignette to only evaluate code chunks if the data package is available.
Example: Conditioning code in the vignette
A code chunk like the following can be included at the start of the vignette to condition the evaluation of code chunks on having the data package available:

```r
hasData <- requireNamespace("hurricaneexposuredata", quietly = TRUE)
if (!hasData){
  knitr::opts_chunk$set(eval = FALSE)
  msg <- paste("Note: Examples in this vignette require that the",
                "`hurricaneexposuredata` package be installed. The",
                "system currently running this vignette does not",
                "have that package installed, so code examples",
                "will not be evaluated."")
  msg <- paste(strwrap(msg), collapse = "\n")
  message(msg)
}
```
Article in The R Journal


Example R packages

- hurricaneexposure: geanders/hurricaneexposure
- hurricaneexposuredata: geanders/hurricaneexposuredata

drat documentation

- Drat Basics for Package Users
- Drat Basics for Package Authors

Acknowledgments

Thanks to Martin Morgan and Roger Peng for helpful suggestions. This work was supported in part by a grant from the National Institute of Environmental Health Sciences (R00ES022631)