Creating a Package in R

Chris Woodard
for Tampa R User's Group November 2019 meeting

R Packages

- Redistributable bundles of code, data, settings, assets and documentation

- Familiar with ggplot2, dplyr, leaflet, and others

- Distributed through CRAN, GitHub, or manually when exported as .tgz files

- Can be created using RStudio

R Package Components

- Code

- Data, Raw Data

- Documentation

- Incidental assets

Creating A Package

R Package Tools

- *devtools* package

- *usethis* package

- *roxygen*

- RStudio

What Will It Do?

Linear Regression!

- Really? Why?

- Would be nice if lm() could fit in a Magritte pipeline by fitting in with the %>% operator

- Would be nice if we could fold in several kinds of output

dplyr Filtering and Subsetting

```
> cyl6 <- filter(mtcars, cyl == 6)
> select(cyl6, mpg, wt, hp)
   mpg    wt  hp
1 21.0 2.620 110
2 21.0 2.875 110
3 21.4 3.215 110
4 18.1 3.460 105
5 19.2 3.440 123
6 17.8 3.440 123
7 19.7 2.770 175
>
```

dplyr Filtering and Subsetting with Pipelines (%>%)

```
> mtcars %>% filter(cyl == 6) %>% select(mpg, hp, wt)
   mpg  hp    wt
1 21.0 110 2.620
2 21.0 110 2.875
3 21.4 110 3.215
4 18.1 105 3.460
5 19.2 123 3.440
6 17.8 123 3.440
7 19.7 175 2.770
```

Our Linear Regression Pkg

- instead of lm(formula, data, options)…

- ***linear_regression(data, formula)***

- returns augmented data frame

  - data in formula + fitted dv + residuals

  - correlations, coefficients, etc. as attributes

<u>Our</u> Linear Regression Pkg

- Can be called as:

  - *result <- data %>% linear_regression(formula)*

- *result* is a data frame with two extra columns: one for fitted and one for residuals; has correlations, coefficients, etc. as attributes.

Code & Things

Resources

- https://support.rstudio.com/hc/en-us/articles/200532317-Writing-Package-Documentation

- https://stackoverflow.com/questions/2310409/how-can-i-document-data-sets-with-roxygen

- https://r-pkgs.org