

Creating an R Data Package

Math 241

Week 3

Step 1 Create a Github Repository called ‘RDatapackage’, and clone it to your machine (using Github Desktop).

Step 2 + Create package with `usethis::create_package("~/filepath/RDatapackage")`.

- Commit your changes to Github **BUT**.
 - Only commit the DESCRIPTION and NAMESPACE.
- Push your changes- can you see them on Github?

Step 3: Open the `.gitignore` file using a text editor of choice, and add the following lines:

- `.gitignore`
- `.Rbuildignore`

Step 4: Add the raw data and transform it into the clean form I want to share in the package.

- Run the function `usethis::use_data_raw()` to create a `data-raw` folder in the R project and the file `DATASET.R`.
- Put the raw data files in the `data-raw` folder.
- Use the file `DATASET.R` to load and wrangle the raw data.
- At the bottom of the wrangling file, include the following code to create a clean `.Rda` file:

```
# Swap seattle_bike for the name of your data
usethis::use_data(seattle_bike, overwrite = TRUE)
```

- Run the code in `DATASET.R` to create a new folder called `data` that contains the cleaned data file.

Step 5: Create the documentation/help file for the data.

- Run `usethis::use_r("insert_data_name")` to create a script file.
- Add `roxygen` comments that document the dataset.
 - See the [Object Documentation Chapter](#) of *R packages*.
- Run `devtools::document()`, which creates the `man` folder and an `Rd` help file.
- Run `?insert_data_name` in the console to view the help file.

roxygen comments

```
#' Portland park trees.
#'
#' A dataset of all the trees in 200 parks in Portland, OR.  Collected as part of the
#'
#' @format A data frame with 15856 rows and 34 variables:
#' \describe{
#'   \item{user_id}{ID}
#'   \item{inventory_date}{Date of data collection}
#'   \item{species}{Species of the tree.  All dead trees were listed as "unknown"}
#'   \item{common_name}{Common name of the tree}
#'   \item{dbh}{Diameter at breast height (4.5' above ground)}
#'   \item{condition}{Trees were rated as good, fair, poor, or dead. These general rat
#'   \item{tree_height}{Height from the ground to the live top of the tree. For dead t
#'   \item{crown_width_ns}{North to South canopy width}
#'   \item{crown_width_ew}{East to West canopy width}
#'   \item{crown_base_height}{Height from the ground to the lowest live canopy.}
#'   \item{collected_by}{Whether data were collect by staff or volunteer}
#'   \item{park}{Park where tree is located}
#'   \item{scientific_name}{Scientific name of the tree}
#'   \item{family}{Family of the tree}
#'   \item{genus}{Genus of the tree}
#'   \item{functional_type}{Categorical variable with groups: Broadleaf Deciduous (BD),
#'   \item{mature_size}{Categorical variable with groups: Large (L), Medium (M), and S
#'   \item{native}{Whether or not the tree is native}
#'   \item{edible}{Categorical variable of edible trees}
#'   \item{nuisance}{Categorical variable indicating if it is a nuisance species}
#'   \item{structural_value}{Monetary value of replacing the tree and the benefits tha
#'   \item{carbon_storage_lb}{The amount of carbon (in lbs.) that is bound up in both
#'   \item{carbon_storage_value}{The monetary value associated with tree effects on at
#'   \item{carbon_sequestration_lb}{The amount of carbon (in lbs.) removed from the at
#'   \item{carbon_sequestration_value}{The monetary value of carbon ($129.72/ton), est
#'   \item{stormwater_ft}{The amount (cubic feet) of avoided stormwater runoff because
#'   \item{stormwater_value}{The monetary value of stormwater runoff that is avoided a
#'   \item{pollution_removal_oz}{The amount (oz.) of air pollution that is removed fra
#'   \item{pollution_removal_value}{The monetary value associated with tree effects on
#'   \item{total_annual_benefits}{Sum of the annual benefits}
#'   \item{origin}{Origin of the tree}
#'   \item{species_factoid}{Additional information about the tree}
#'   \item{longitude}{Longitude}
#'   \item{latitude}{Latitude}
#' }
#' @source \url{https://www.portlandoregon.gov/parks/article/433143}
"pdxTrees"
```



Output help file

pdxTrees (pdxTrees)

R Documentation

Portland park trees.

Description

A dataset of all the trees in 200 parks in Portland, OR. Collected as part of the Urban Forestry Tree Inventory Project.

Usage

pdxTrees

Format

A data frame with 15856 rows and 34 variables:

user_id	ID
inventory_date	Date of data collection
species	Species of the tree. All dead trees were listed as "unknown"
common_name	Common name of the tree
dbh	Diameter at breast height (4.5' above ground)
condition	Trees were rated as good, fair, poor, or dead. These general ratings reflect whether or not a tree is likely to continue contributing to the urban forest (good and fair trees) or whether the tree is at or near the end of its life (poor and dead trees).
tree_height	Height from the ground to the live top of the tree. For dead trees, total height was measured.

- Notice that you need to include the dataset name as a string in the bottom of the R script. If you forget this, the help file for that dataset won't be created!

Step 6: Edit the [DESCRIPTION file](#) and consider creating a Readme file. The code `usethis::use_readme_md(open = interactive())` generates a skeleton Readme file.

Step 7: Delete any extraneous files or list them in `.Rbuildignore`.

Step 8: Run checks with `devtools::check(document = FALSE)` to make sure the package compiles without errors or warnings. Run `devtools::load_all()` to simulates the process of building, installing, and attaching the package if you want to check its contain.

- Fix any errors or warnings. (Note: it will still compile under warnings.)

Step 9: Try installing the package with `devtools::install()`.

Step 10/Always: Push the changes to the GitHub repository.