

StatR 502 Final Project

Gregor Thomas

Due dates:

- Project proposals must be submitted by midnight on Monday, February 15. (See the end of the document for proposal guidelines).
- Final projects will be due midnight on Wednesday, March 16 (the week following the last class).

Project requirements:

Your final project is to produce a small, fully-documented R package. The package should be data-focused, in that it includes data, a vignette that presents an analysis of the data, and functions that support the analysis.

At a minimum, your package needs to:

- include at least two non-trivial, original R functions
- include a dataset
- include documentation for
 - all exported R functions,
 - the data,
 - the package itself
- pass `devtools::check` or R CMD CHECK with no errors (warnings and notes can be acceptable)
- include a vignette as described below

Your package vignette should present a brief data analysis. It should demonstrate the regression modeling techniques learned this quarter. Your vignette should be a narrative walk through of your data and analysis. It should

- describe the data
- **state an analysis objective**
- describe any necessary data cleaning
- explore the data graphically (1-3 plots should be sufficient, focused on your objective)
- select an appropriate model and variables
- consider transformations, outliers, etc.
- refine the model
- interpret the model

This is a long list, but many of the items can be dealt with briefly. You do not need to show *everything* you try. Start with a description of a simple model, go into detail about a final model, and give an overview of how you got there and any dead-ends that didn't make it into the final model. Use your package's functions to keep your code clear and concise.

Things I **do not** want to see in your vignette include:

- verbose output from stepwise regression (or similar): if you iterate through models, manually or automatically, tell me how in a sentence or two, and then show me your final model.
- data transformation code: this can be “hidden” in the Rmd source document

What I **do** want to see in your vignette:

- Enough description about of the data and subject area that I know what's going on and why it matters.
- A clear statement of your goals in the analysis.
- Conclusions you draw from the models
- A quick discussion of what you would do if this were a longer/bigger project. ###
Project Submission

Projects are to be submitted as `.tar.gz` bundles as produced by `devtools::build` (with the default `binary = FALSE` option).

Project Proposal:

Your project proposal should be a **short** description of the data you would like to include in your package. Describe the data source, identify the response variable and research question, and give an overview of the covariates.

Try to pick data that you are interested in, either professionally or personally. Do find something that has several covariates. I'm reluctant to set strict requirements, but I would hope to see at least 5 predictors, with both categorical and continuous variables present.

As always, seek help if you're having trouble! Use the final project discussion board and we can talk about anything from technical problems to help finding interesting data to settling on a research question.