Contributing to R packages and projects can be a rewarding way to give back to the tools you use and to improve your own programming skills in the process. In this talk, Kara will discuss some of the varied ways to contribute to existing projects. Anyone, whether a seasoned programmer or someone brand new to R, can make useful contributions to R packages. Kara will draw on her experience working on ggplot2 to offer strategies for finding your way in an unfamiliar codebase, and to give insights into the relationship between maintainers and contributors.

The statistical modelling ecosystem in R can resemble the wild west. However, we could resolve many of the peculiarities of the current modelling ecosystem by differentiating between individual models and families of models. Using penalized regression as an example, Alex will show how modelling tasks can be more intuitively expressed in code that treats models and model families as distinct classes with distinct methods. This “statistical type safety” is a key component of any successful grammar of modelling, reducing cognitive load on data analysts and providing a target interface for researchers developing new methods. Alex will discuss current work on the modelling ecosystem in R, the problem of technical debt when using multiple R packages, and best practices for new modelling packages.

and show how these first attempts at a grammar of modelling result in an interface that is intuitive for users and also a target interface for researchers developing new methods. I will discuss the technical debt incurred by the current modelling ecosystem in R and steps towards a pit of success for the modelling ecosystem, including

(The form says you’re only looking for lightning talks at the moment, but if there’s an opportunity to speak a little longer I’d love to share some of the fascinating work on modelling that’s currently going down).

He will also discuss

ongoing be a fundamental component

Statistical modelling in R is notoriously laborous. Each modelling technique lives in its own package, has its own (sometimes idiosyncratic) interface and see of methods. In this talk, I argue that we need to differentiate between statistical models (KNN with k = 5) and statistical model families (KNN). I will argue that models and model families should be represented by their own objects with distinct classes, and distinct methods acting on those classes. I will discuss ongoing efforts to standardize methods for fitting and interacting with statistical models, and show how these first attempts at a grammar of modelling result in an interface that is intuitive for users and also a target interface for researchers developing new methods. I will discuss the technical debt incurred by the current modelling ecosystem in R and steps towards a pit of success for the modelling ecosystem, including

- the pit of the success: we ain't there yet

- my background: running kaggle workshops for students just getting into R and Python

- models and model families (motivating example: LASSO)

- distinct objects with distinct classes

- why current interfaces are conceptually confusing and need revision

- ongoing work to standardize model construction and usage (recipes, rsample, tidyposterior, broom)

- interface: researchers need to be able to target it. many connections and intermediate pieces that haven't been standardized (what broom tries to do). broom interesting: standardize in one place or in each home package: only good if everyone follows community standards.

- best practices from modelling

- big thing part 1: specifying and fitting the model

- big thing part 2: interacting with the model. Many intermediate objects

- some particular parts of modelling that we aren’t going to touch on toward since they already have DSLs that I think work: Bayesian stuff, deep learning stuff

Alex Hayes is a senior at Rice University studying statistics. He is particularly interested in improving modelling interfaces in R.

Alex Hayes

alexpghayes@gmail.com

Rice University

On modelling in R

# Towards a grammar of modelling in R

# Intuitive modelling in R via statistical type safety

## Abstract

Recent developments in the R ecosystem, particularly the advent of the tidyverse, have res

Statistical modelling in R is notoriously laborous. Each modelling technique lives in its own package, has its own (sometimes idiosyncratic) interface and see of methods. In this talk, I argue that we need to differentiate between statistical models (KNN with k = 5) and statistical model families (KNN). I will argue that models and model families should be represented by their own objects with distinct classes, and distinct methods acting on those classes. I will discuss ongoing efforts to standardize methods for fitting and interacting with statistical models, and show how these first attempts at a grammar of modelling result in an interface that is intuitive for users and also a target interface for researchers developing new methods. I will discuss the technical debt incurred by the current modelling ecosystem in R and steps towards a pit of success for the modelling ecosystem, including

- the pit of the success: we ain't there yet

- my background: running kaggle workshops for students just getting into R and Python

- models and model families (motivating example: LASSO)

- distinct objects with distinct classes

- why current interfaces are conceptually confusing and need revision

- ongoing work to standardize model construction and usage (recipes, rsample, tidyposterior, broom)

- interface: researchers need to be able to target it. many connections and intermediate pieces that haven't been standardized (what broom tries to do). broom interesting: standardize in one place or in each home package: only good if everyone follows community standards.

- best practices from modelling

Yes I would like to apply for a scholarship.