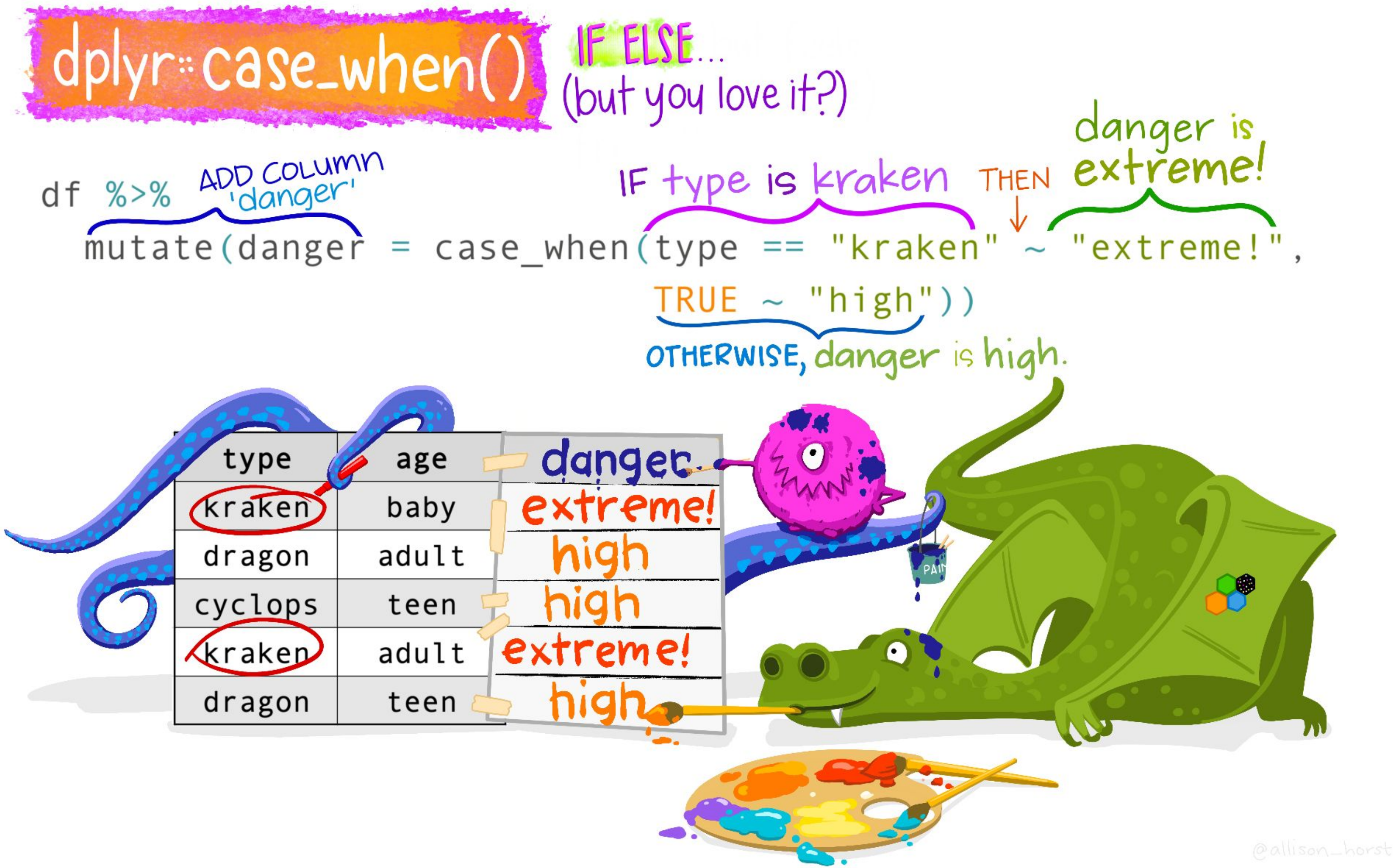
Linear Regression II

EDS 232
Instructor: Mateo Robbins
Winter 2023

MEDS

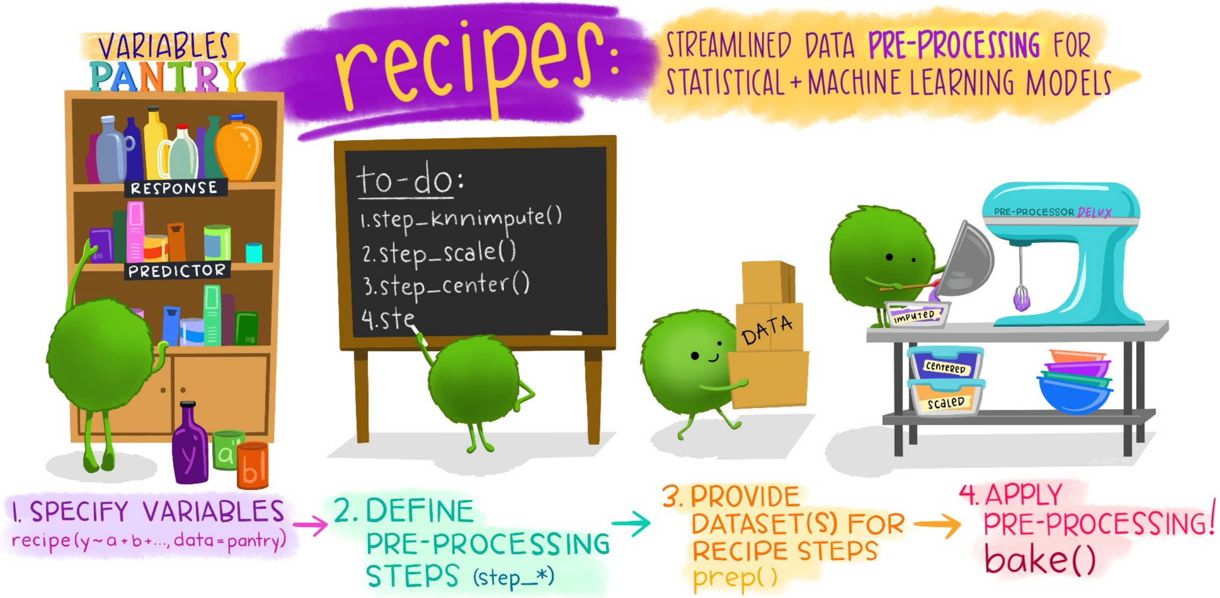
Following up from last time...



More context for tidymodels

Model preprocessing

A recipe prepares your data for modeling



A recipe is a specification of intent

- Recipes allow us to specify the preprocessing steps to carry out on the data to prepare it for modeling

A recipe is a specification of intent

- Recipes allow us to specify the preprocessing steps to carry out on the data to prepare it for modeling
  - Encoding predictors
  - Computing polynomials of degree n...

Recipe code

```
example_recipe <- recipe(outcome ~ predictor, data = your_data) %>%
  step_integer(all_predictors(), zero_based = TRUE)
```

Prepping and baking your recipe

`prep()`: executes the transformations specified in your recipe onto some data (usually training data)

`bake()`: takes a prepped recipe and applies it to a data set (usually the test)

ML Modeling Process

- **Define**
  - **Specify what type of model will it be. Define some parameters.**
- **Fit**
  - **Capture patterns from provided data. The heart of the modeling process.**
- **Predict**
  - **Use the model.**
- **Evaluate**
  - **How accurate are the model’s predictions?**

Specifying a model in tidymodels

- Specify the model based on mathematical structure
  - Linear regression, random forest, KNN, etc.
- Specify an engine for fitting the model
  - Usually which software package to use
- Declare mode of the model (when required)
  - Numeric outcome = regression, qualitative outcome = classification

Specifying a model in tidymodels

```
example_spec <- linear_reg() %>%
  set_engine("lm") %>%
  set_mode("regression")
```

Workflow Bundles Model Components

```
example_wf <- workflow() %>%
  add_recipe("some_recipe") %>%
  add_model("some_spec")
```

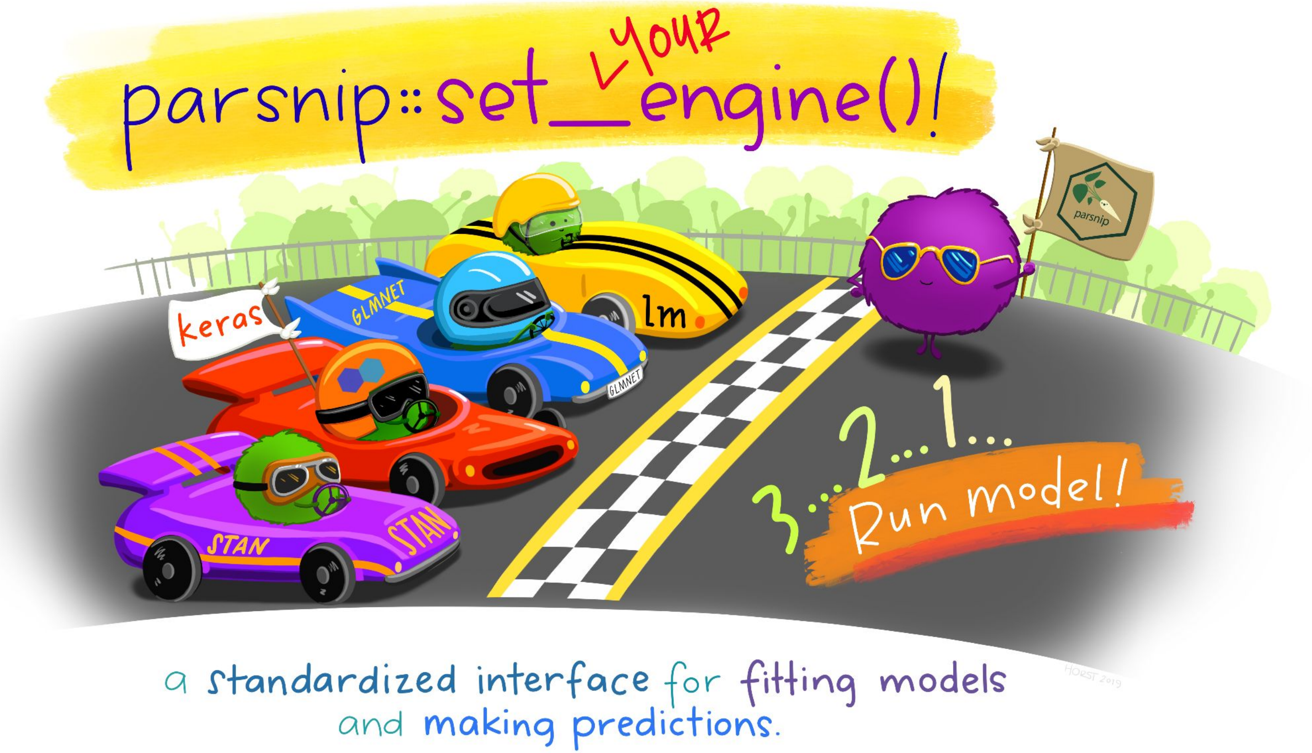
Why use a workflow()?

- `Simplify workspace`
- `Prepping and fitting with a single call`
- `Simplified custom tuning`
- `Postprocessing`

Why use a workflow()?

- **`Simplify workspace`**
- **`Prepping and fitting with a single call`**
- *`Simplified custom tuning`*
- *`Postprocessing`*

parsnip()


parsnip::set_your engine()!
keras
GLMNET
lm
GLMNET
STAN
STAN
parsnip
3...2...1...
Run model!
a standardized interface for fitting models
and making predictions.

Fitting a model

- Parsnip allows us to be indifferent to the interface of the underlying model
  - Simply use fit() to estimate the model

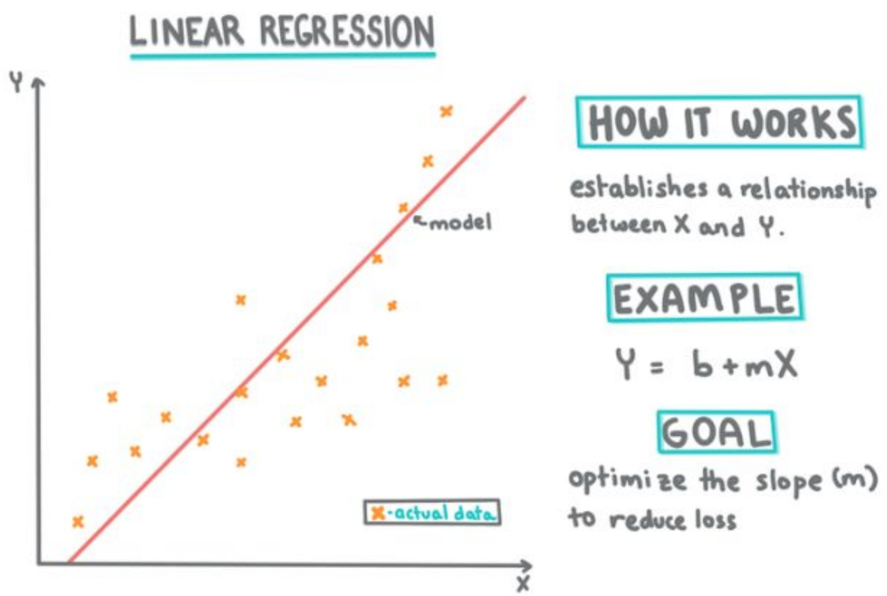
Parsnip is tidy

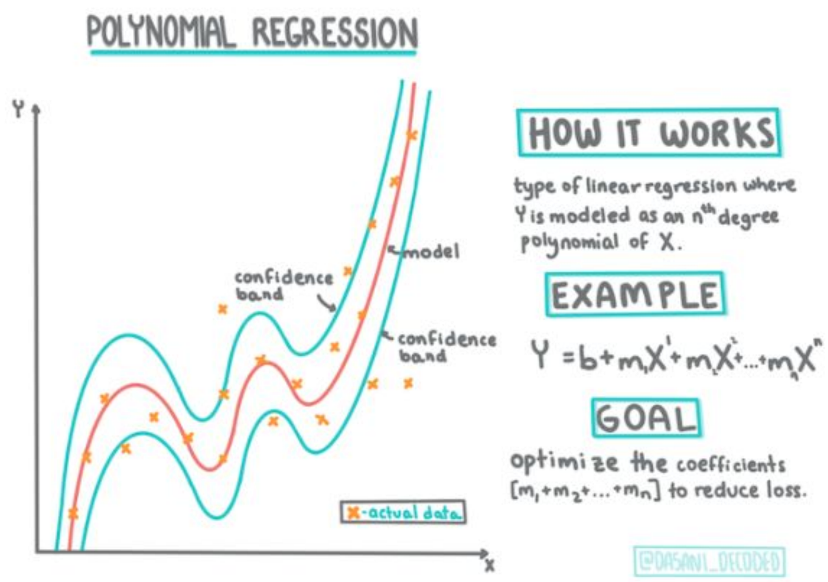
Table 3.1: Heterogeneous argument names for different modeling functions.

Function	Package	Code
`lda()`	MASS	`predict(object)`
`glm()`	stats	`predict(object, type = "response")`
`gbm()`	gbm	`predict(object, type = "response", n.trees)`
`mda()`	mda	`predict(object, type = "posterior")`
`rpart()`	rpart	`predict(object, type = "prob")`
various	RWeka	`predict(object, type = "probability")`
`logitboost()`	LogitBoost	`predict(object, type = "raw", nIter)`
`pamr.train()`	pamr	`pamr.predict(object, type = "posterior")`

Using Model Results

Today's Lab…more pumpkins





Infographic by Dasani Madipalli