

# Statistics 360: Advanced R for Data Science

## Group Project

Brad McNeney



## Group projects

- ▶ You will do a final project in groups of two or three.
- ▶ The project and final exam will be worth 50% of your mark. I will do a poll in the first class to get feedback on the breakdown.
- ▶ For the project you will create an R package that implements a statistical method.
- ▶ Same method for all. Will do an in-class poll to choose one of the following:
  1. Multivariate Adaptive Regression Splines (MARS) – mainly for prediction
  2. Penalized logistic regression – mainly for inference
- ▶ The main fitting function will be developed in the class and lab exercises throughout the first half of the course.
- ▶ You will need to add “methods”, tests and documentation.
  - ▶ A grading rubric will be circulated around reading break.

# User interface

- ▶ The main fitting function and its output will have a similar interface to the `lm()` and `glm()` functions in R.
- ▶ Function arguments include a formulas to specify models, data and parameters that control the fitting.
- ▶ Output contains all the user needs to make predictions or inference.
- ▶ Write “methods” to do predictions, plots, inference, etc. Will implement as many of those for `lm()` as are relevant and practical.

```
methods(class="lm")
```

```
## [1] add1          alias          anova          case.names     coerce
## [6] confint        cooks.distance deviance       dfbeta         dfbetas
## [11] drop1          dummy.coef     effects       extractAIC     family
## [16] formula        hatvalues     influence      initialize     kappa
## [21] labels        logLik        model.frame    model.matrix   nobs
## [26] plot          predict       print         proj          qr
## [31] residuals     rstandard    rstudent      show          simulate
## [36] slotsFromS3   summary      variable.names vcov
## see '?methods' for accessing help and source code
```