

# **Intro Stats/ Data Science**

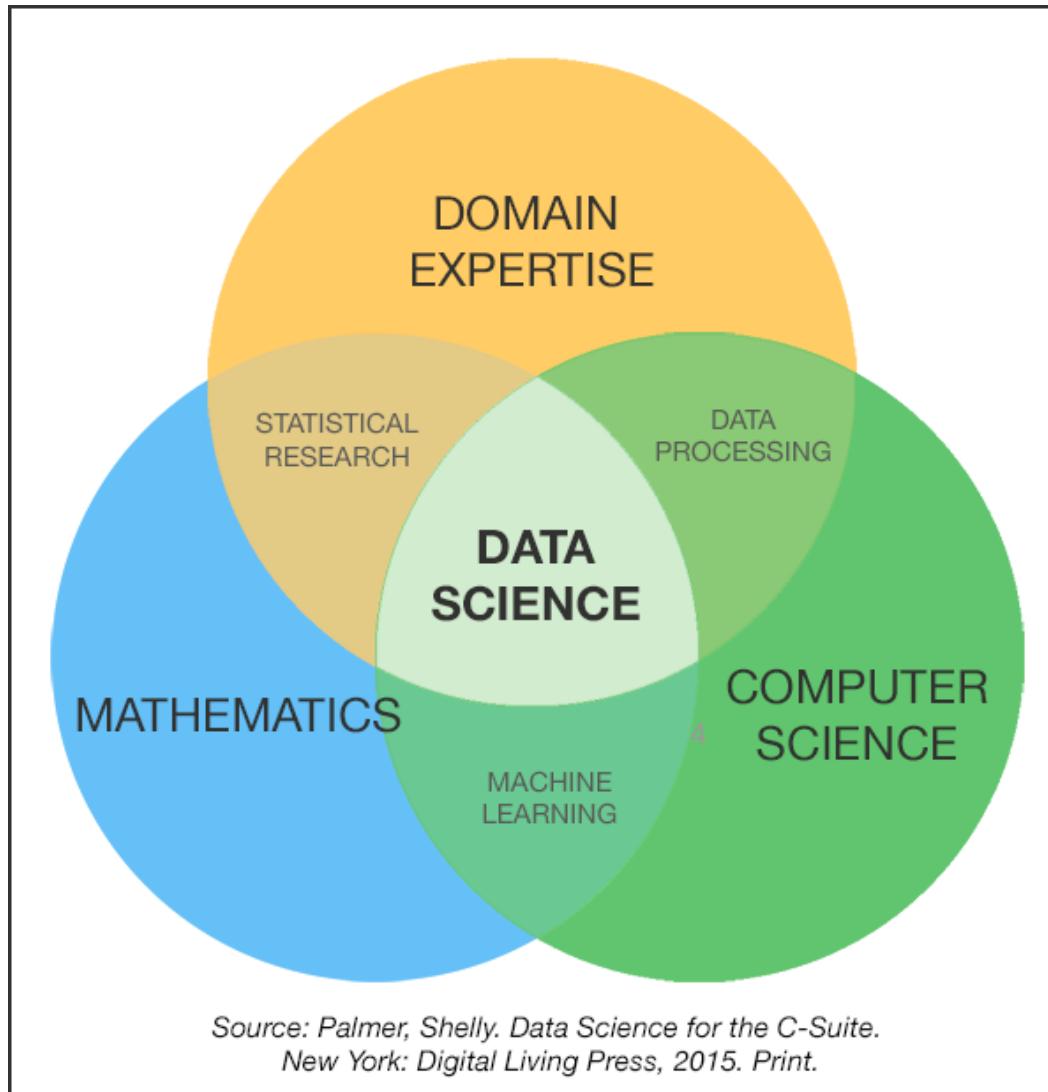
# Where are we?



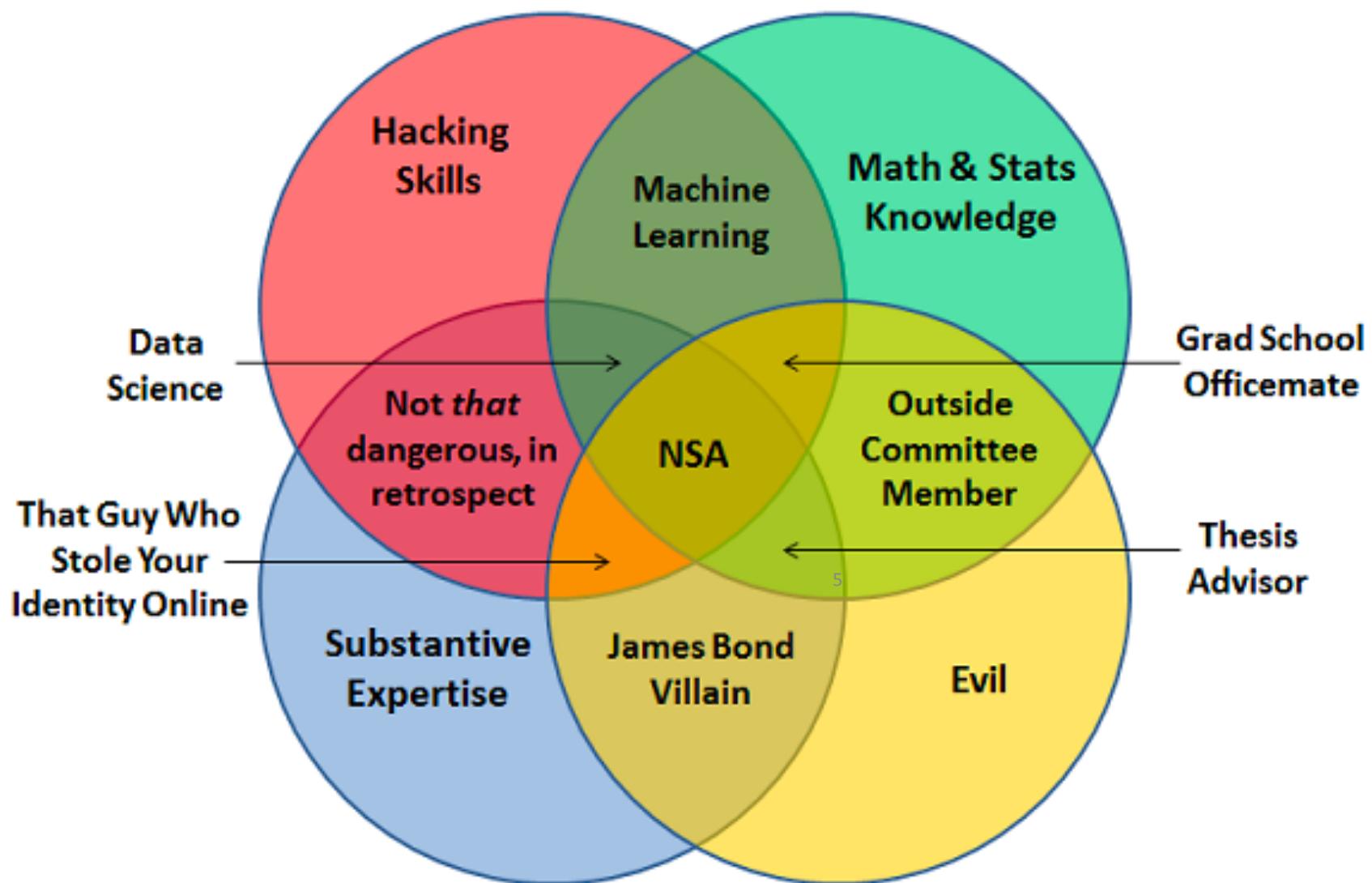
# Data Science



# Data Science?

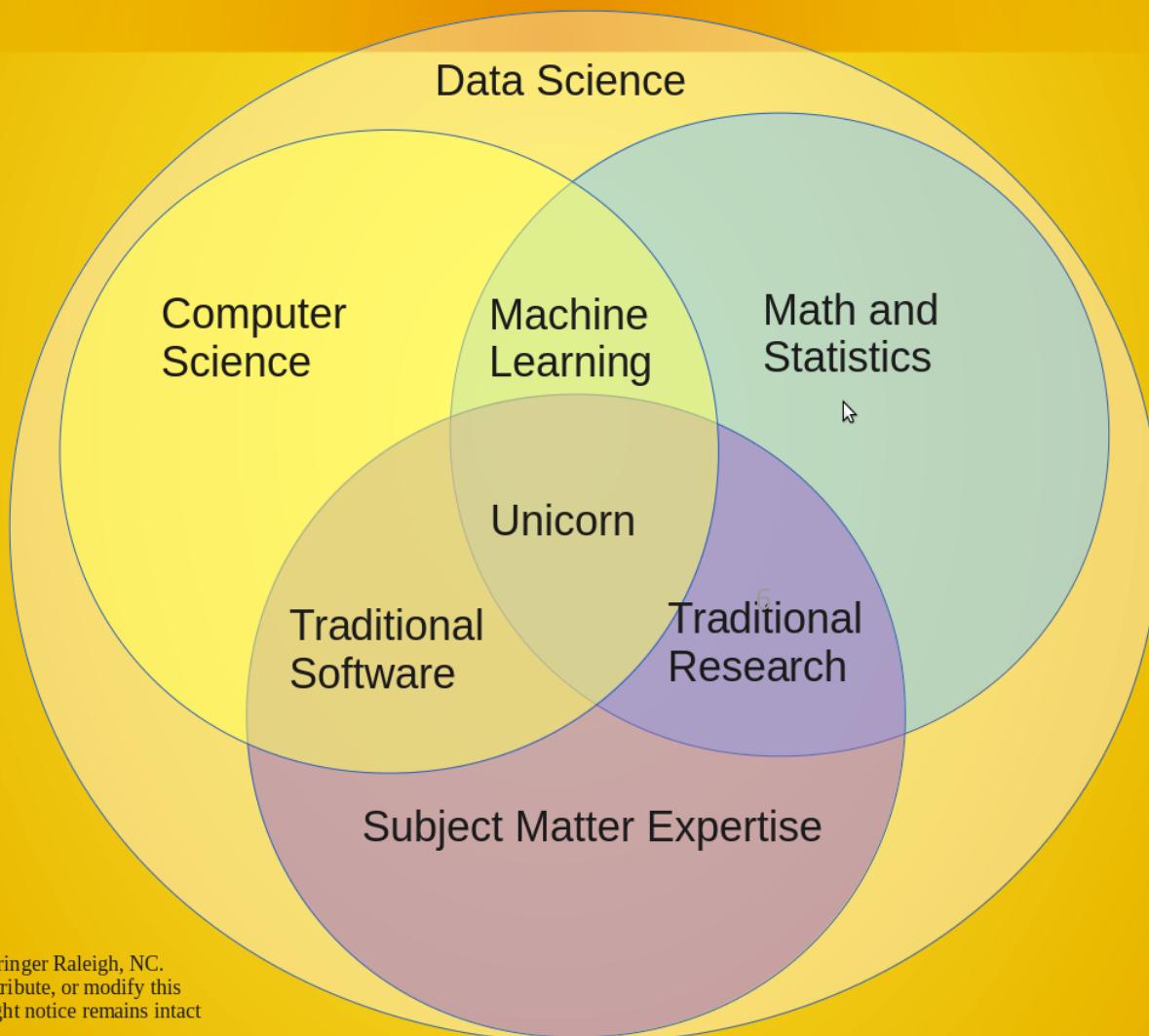


# Or?

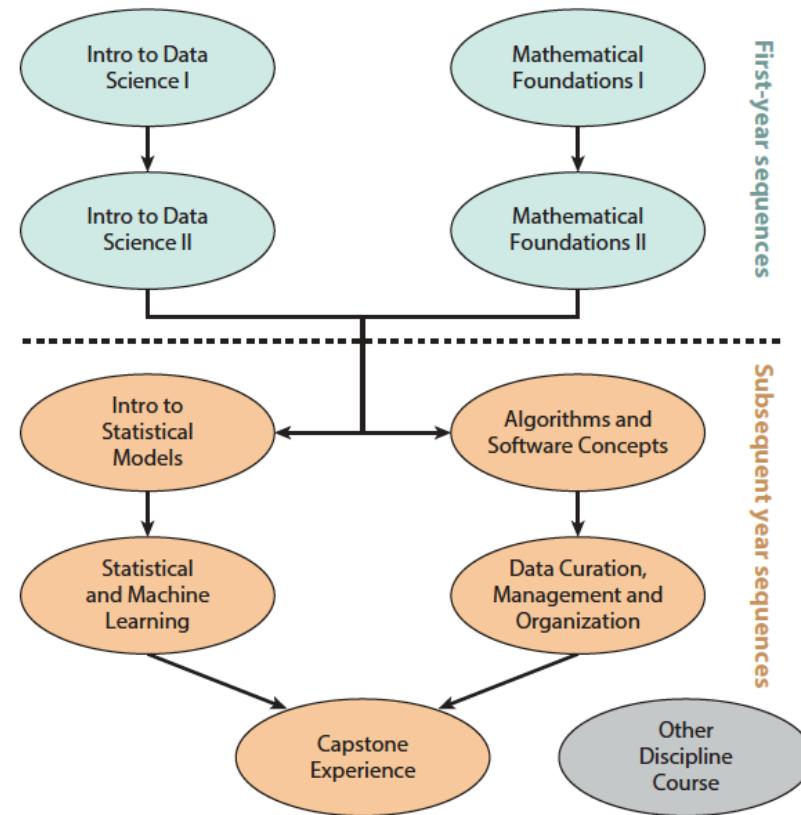


# Data Science!!

## Data Science Venn Diagram v2.0



# Possible path



**Guidelines:** <http://www.annualreviews.org/doi/full/10.1146/annurev-statistics-060116-053930>

**Course content:** [http://www.annualreviews.org/doi/suppl/10.1146/annurev-statistics-060116-053930/suppl\\_file/st04\\_de\\_veaux\\_supmat.pdf](http://www.annualreviews.org/doi/suppl/10.1146/annurev-statistics-060116-053930/suppl_file/st04_de_veaux_supmat.pdf)

# Infer—not first course

- Baumer — Unified approach to inference



- Why not just simulate?
  - Start with simulation
  - Central Limit Theorem
  - Bootstrap
- Multivariate?



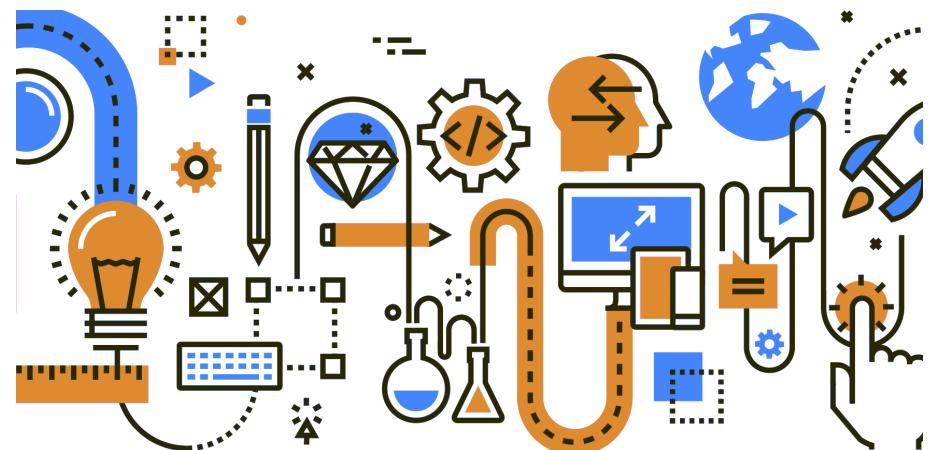
# CMU

- Data Science and General Education
  - Intro stats for non - computer people
  - Reasoning with Data
  - Community Yale/NUS



# Duke

- Multivariate – Hooray
- Tidyverse?
- Model-based
- Real Projects
- Collaboration – teams
- Communication



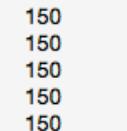
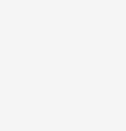
# Macalester



Model Comparison Test=1

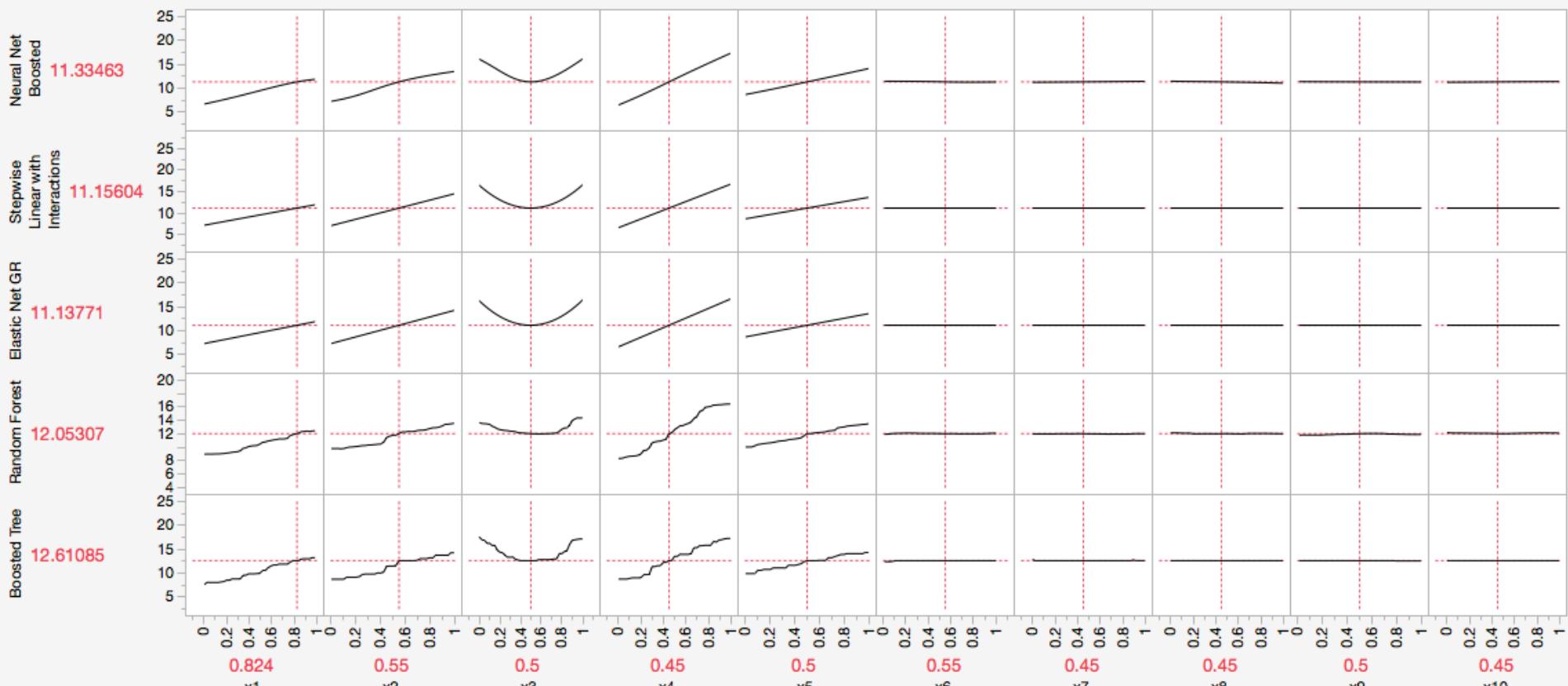
► Predictors

▼ Measures of Fit for y

Predictor	Creator	.2.4.6.8	RSquare	RASE	AAE	Freq
Neural Net Boosted	Neural		0.8730	1.3925	1.0899	150
Stepwise Linear with Interactions	Fit Least Squares		0.8748	1.3826	1.0981	150
Elastic Net GR	Fit Generalized Adaptive Elastic Net		0.8759	1.3763	1.0812	150
Random Forest	Bootstrap Forest		0.7459	1.9697	1.5389	150
Boosted Tree	Boosted Tree		0.8464	1.5312	1.2063	150

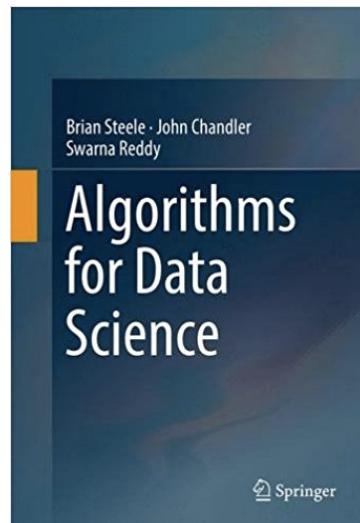
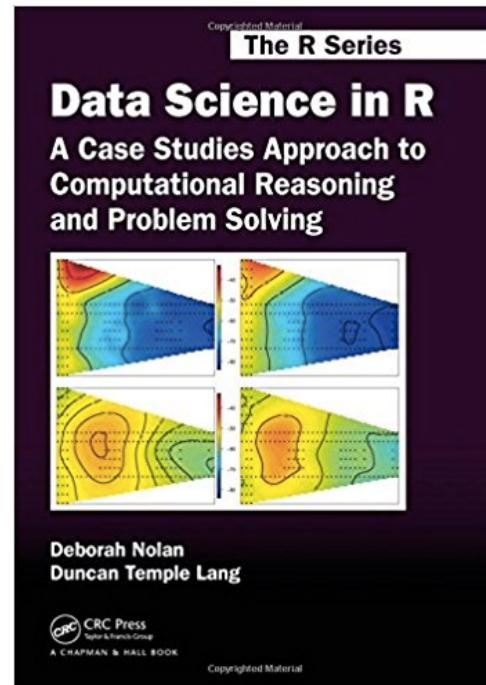
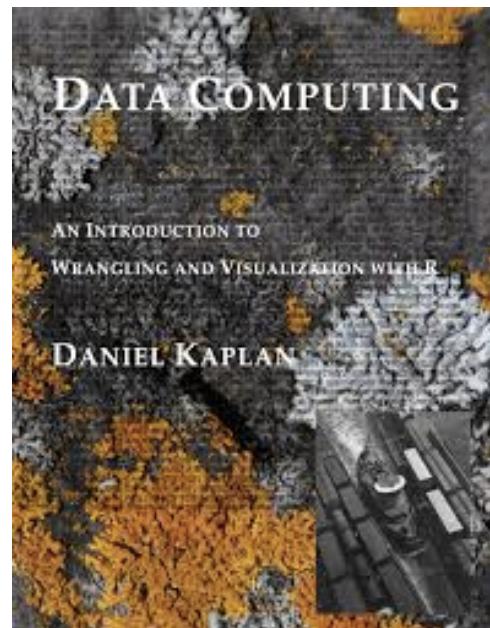
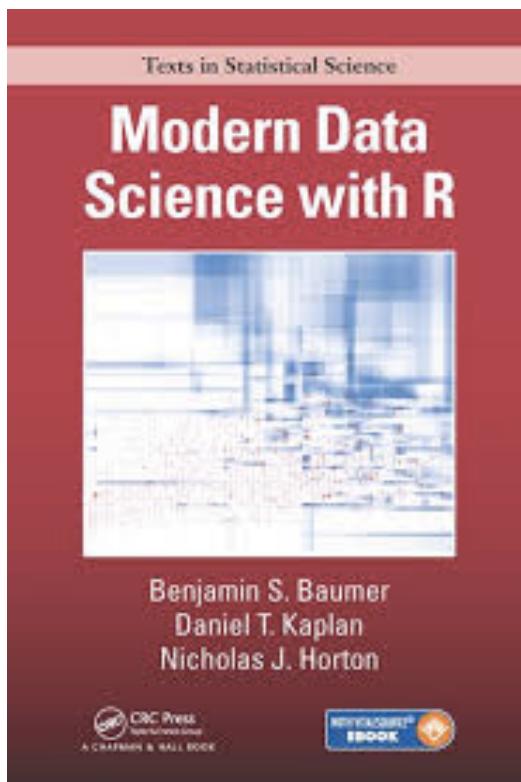
▼ Profiler Test=1

► Prediction Profiler

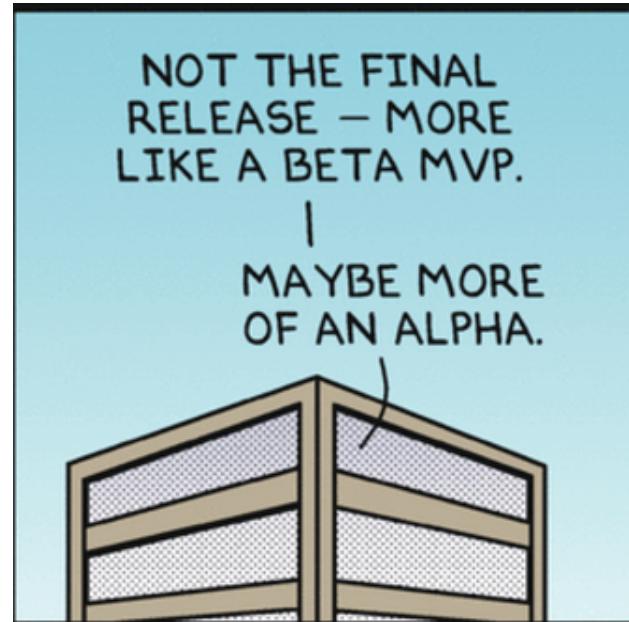


The figure displays a 5x10 grid of plots representing the prediction profiler for five different models. The rows are labeled on the left with the model names and their respective RSquare values: Neural Net Boosted (0.8730), Stepwise Linear with Interactions (0.8748), Elastic Net GR (0.8759), Random Forest (0.7459), and Boosted Tree (0.8464). The columns represent predictors x1 through x10. Each plot shows the relationship between a predictor and the predicted value, with other predictors set to their mean values. The axes for each plot are scaled differently, reflecting the nature of the data for each predictor.

# Books



# Where are we?



# Now what?



[deveaux@williams.edu](mailto:deveaux@williams.edu)

Thank you!!