Comparing grade outcomes in a second statistics course based on introductory statistics curricula

Mark Greenwood, Professor of Statistics

June 29, 2018

Acknowledgments

- Charlie Carpenter (MSU Undergraduate Math-Stat Option, Biostatistics Graduate student at CU-Denver)
- Jenny Green (Associate Professor of Statistics)

STAT 217: Intermediate Statistical Concepts at MSU

- Required for a wide variety of majors on campus
- 8-10 sections of 35 students per year ((re-)growing in recent years)
- Pre-requisite: STAT 216 or equivalent
  - Increasing rates of AP Statistics and students meeting pre-requisites from other institutions
  - Varying versions of STAT 216 (consensus, randomization)
- STAT 217 is about 20% of the size of STAT 216

History of 217Q: Intermediate Statistical Concepts at MSU

- Before 2009
  - Using some of Moore and McCabe and supplemental materials
  - Minitab(?)
- Spring 2009 to Summer 2013
  - Using last chapters of Deveaux, Velleman, and Bock's *Stats: Data and Models*
  - Deviated from DVB in ANOVA parametrization, depth of coverage on many topics
  - R (and sometimes R-commander)

My book criteria for STAT 217

- Integrate R via R-studio from first day, show power of software in examples
  - Stay up to date with changes in R packages
- Agnostic to previous statistics experience (randomization, consensus, AP, other institutions)
- Emphasize scope of inference
- Real, messy data sets with information on data collection and units for all variables, examples of both randomized and randomly sampled data
- Reasonably priced
- Cover only topics I want covered

STAT 217: Fall 2013 - Present

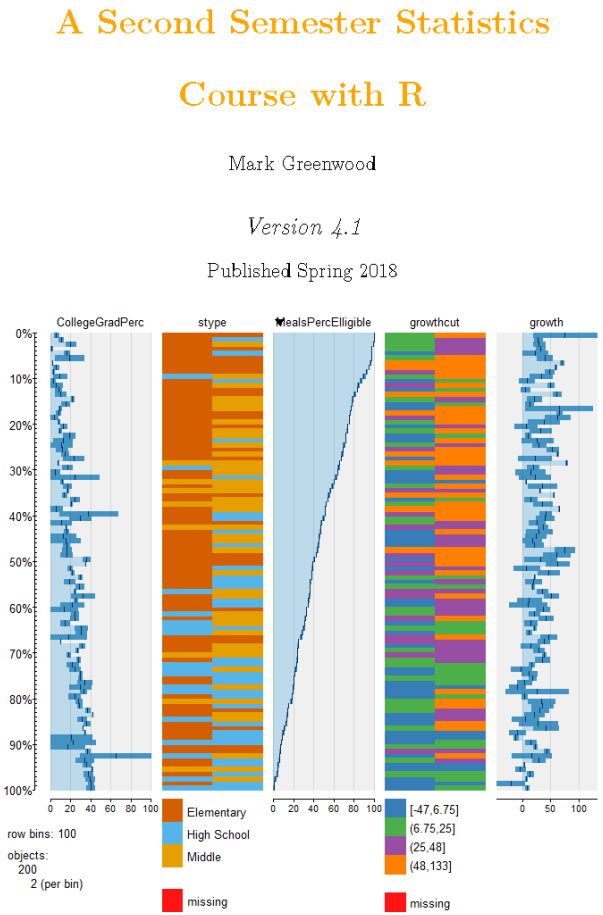


Figure 1: Cover of book

STAT 217: Fall 2013 - Present

- *A Second Semester Statistics Course with R (Versions 0.1 to 4.0)* by Greenwood (and Banner for V. 0.1-3)
  - Locally printed
  - Published via MSU Scholarworks (https://scholarworks.montana.edu/xmlui/handle/1/2999): pdf
  - And http://www.math.montana.edu/courses/s217/documents/_book/: html and pdf versions
    - Bookdown format (V 4+)

STAT 217 Topics covered:

- Scope of inference
- 2-sample mean (parametric, permutation, and bootstrap inferences)
  - Introduce R, R-studio, and loops, basic plotting
  - Review hypothesis testing and confidence intervals
- One-Way ANOVA (permutation and parametric F-test) with follow-up tests
  - Begin working with indicator variables
- Two-Way ANOVA with interactions
- Chi-square tests (permutation and parametric χ^2)
- Simple linear regression inference (parametric and permutation)
  - Cover log-transformations and prediction intervals
- Multiple linear regression
  - All quantitative predictors (multicollinearity and model selection including AICs)
  - One categorical predictor with interactions

Favorite R packages for STAT 217

- `beanplot` (Kampstra, 2008)
  - Improvement of stripcharts and boxplots
- `mosaic` (Pruim, Kaplan, Horton, 2017)
  - Permutation and bootstrapping "code"
  - Philosophy: All code is formula based, `__(Y~X,data=...)`
- `car` (Fox and Weisburg, 2011)
  - `scatterplot`, `vif` and `Anova` (Type II ANOVA tests)
- `effects` (Fox, 2003)
  - `plot(allEffects(modelname))`
- `tableplot` (Tennekes and de Jonge, 2017); corrplot (Wei and Simko, 2017)
- 2-WAY ANOVA interaction plotting function:

```
source("http://www.math.montana.edu/courses/s217/documents/intplot.R")
```

Reproducible Research methods

- Rmarkdown (http://rmarkdown.rstudio.com/authoring_basics.html)
- All sections discuss it, instructors allowed to require it

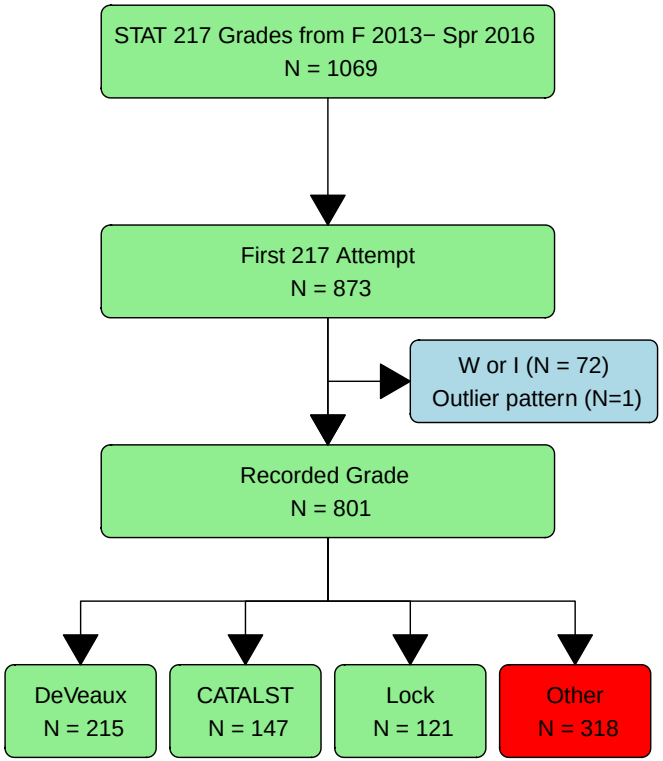
STAT 216Q: Introduction to Statistics

- Interested in impacts of STAT 216 curricula and student performance in STAT 216 on performance in STAT 217
  - Also control for overall student GPA before taking STAT 217
- See Hildreth, Robison-Cox, and Schmidt (2018) "Comparing Student Success and Understanding in Introductory Statistics Under Consensus and Simulation-Based Curricula" for details of STAT 216 at MSU in this timeframe

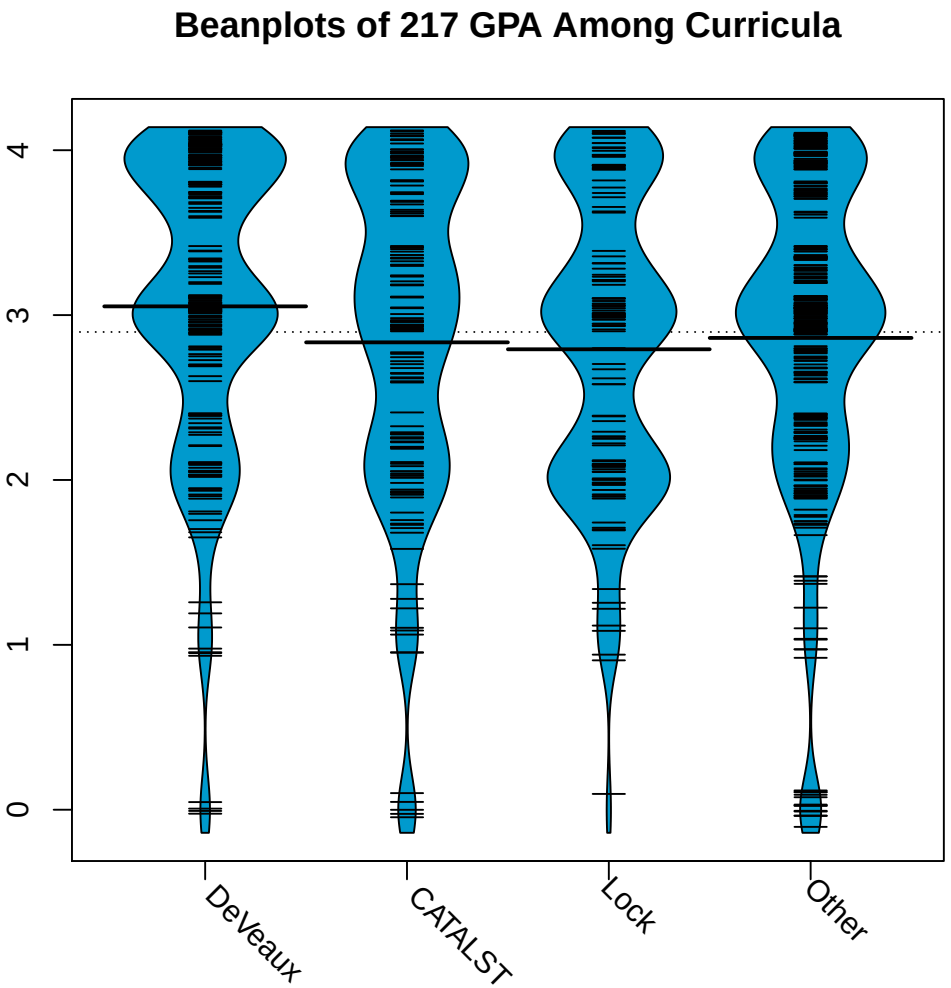
Versions of STAT 216

- CATALST
  - Randomization to teach inference
  - Version with TinkerPlots
  - MSU-specific version inspired by CATALST (with customized Shiny apps)
- DVB
  - Consensus curriculum
- Lock(^{5})
  - Randomization with some parametric and web applets

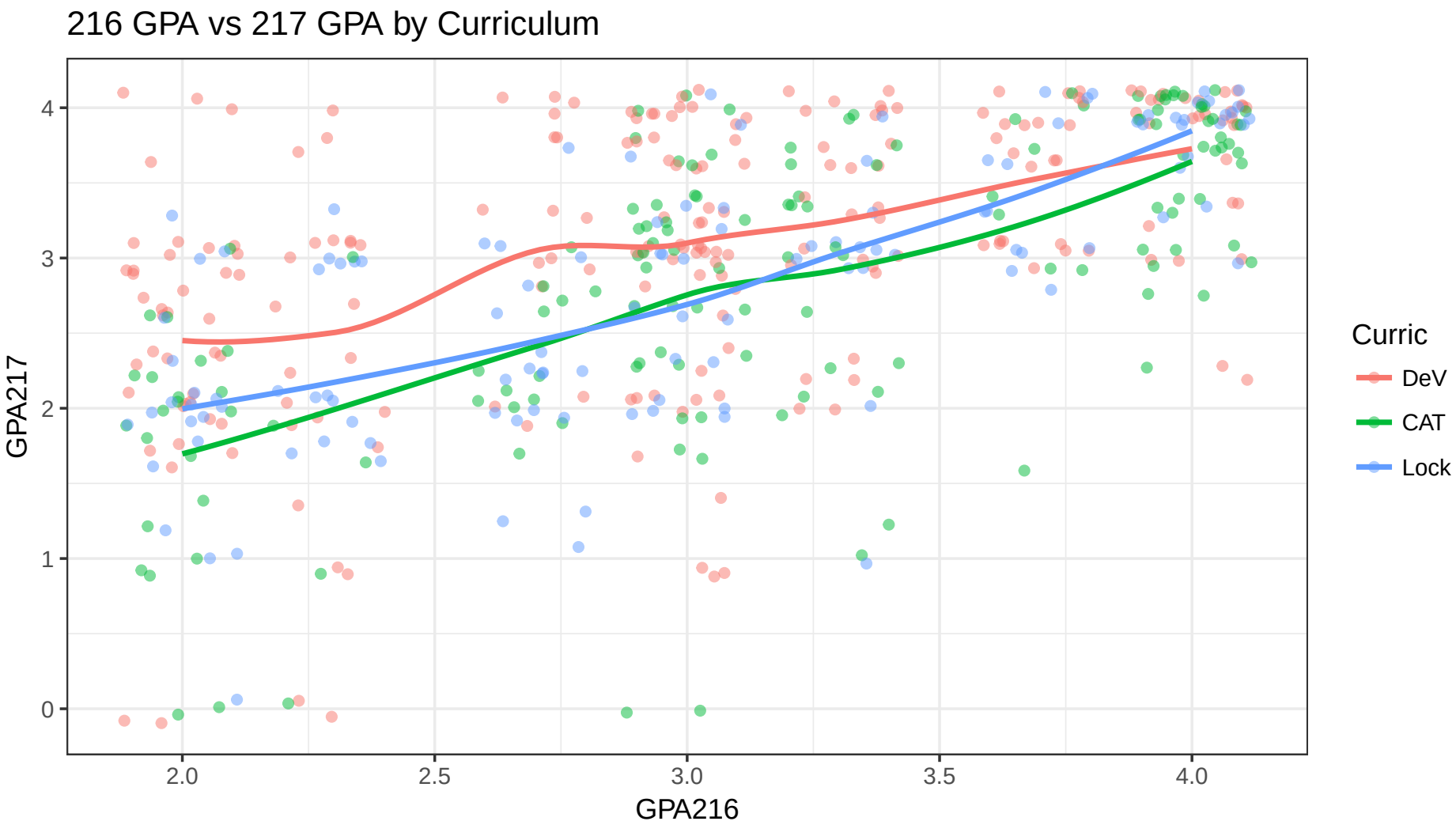
Inclusion Diagram for 217 Grade data



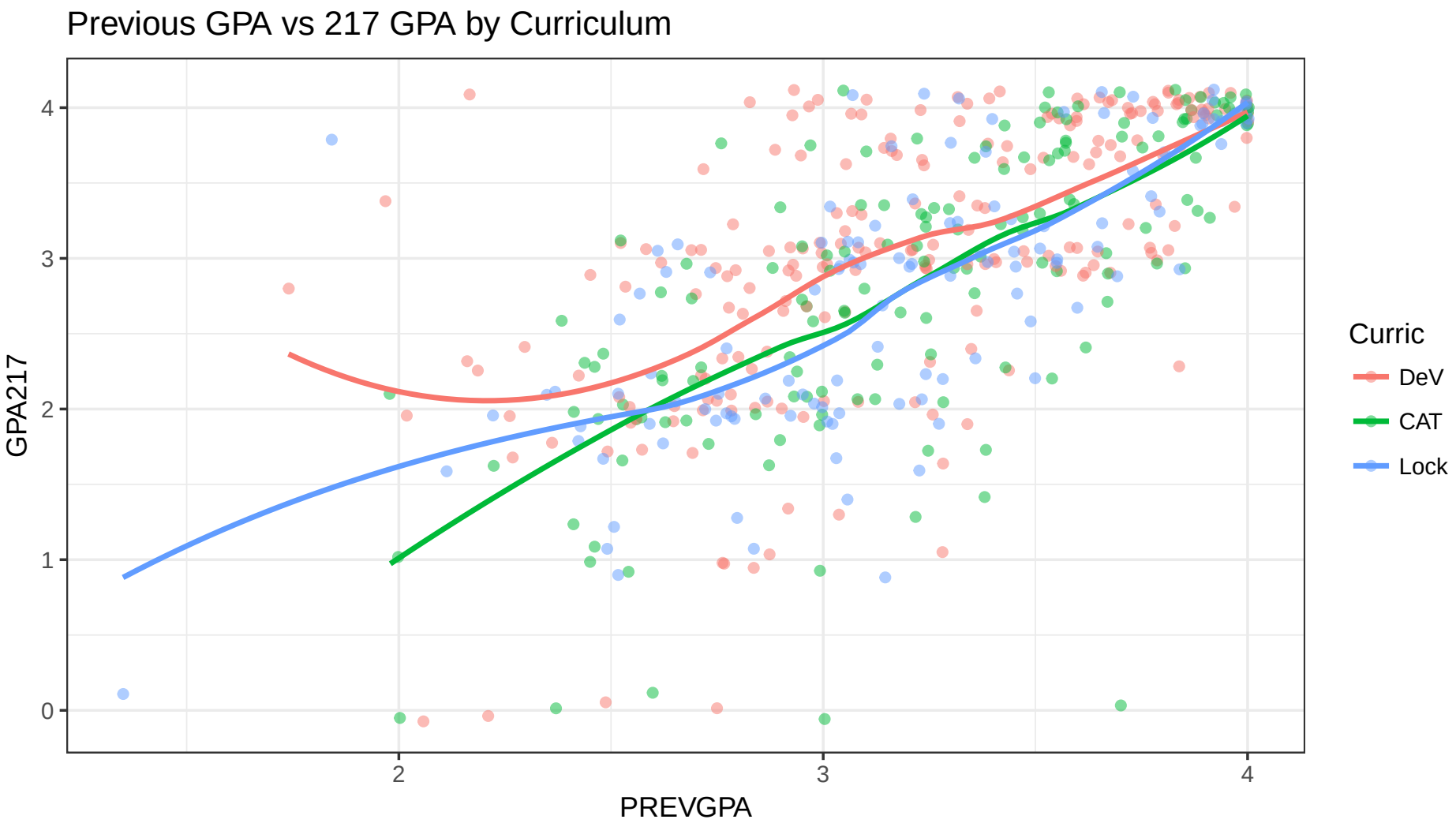
STAT 217 Grades by curriculum:



Scatterplot of 216 and 217 grades by curricula



Scatterplot of Previous GPA with 217 Grade by curricula



Censored regression model for 217 GPA

- $y_i^\star = \beta_0 + \beta_1 x_{1i} + ... + \epsilon_i$
- with $\epsilon_i \sim N(0, \sigma^2)$ and we observe

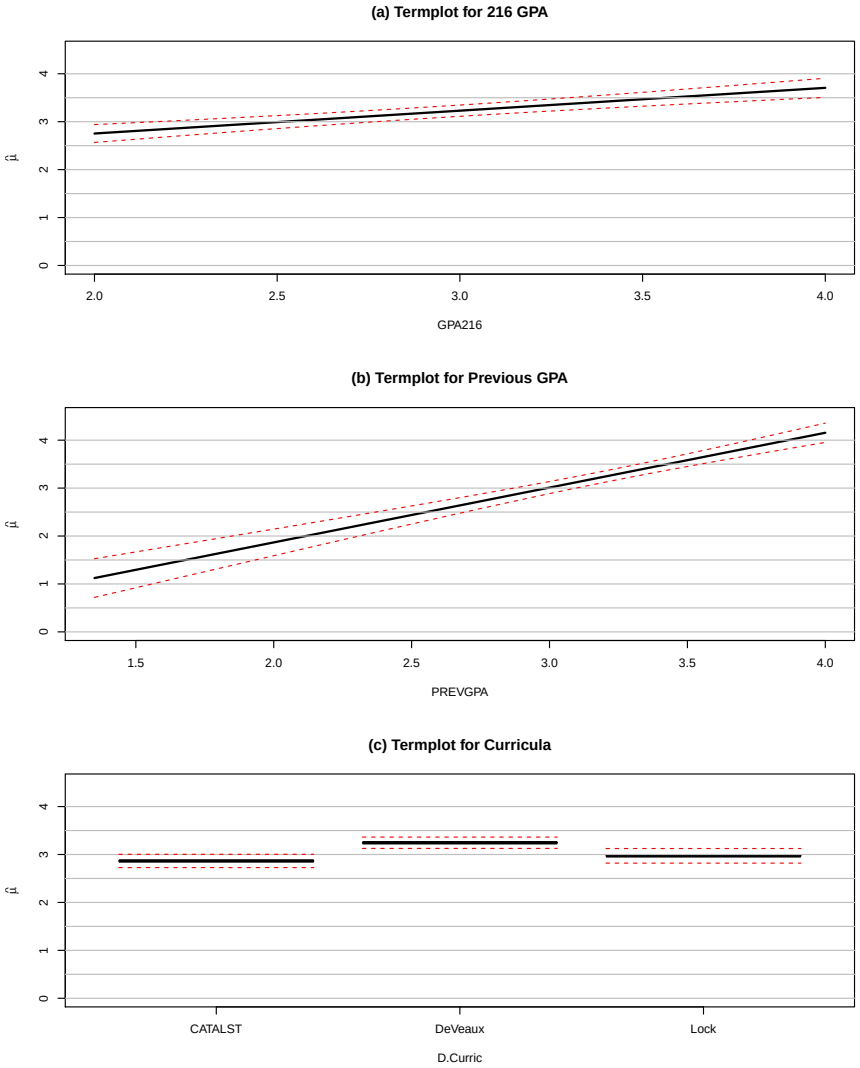
$$y_i = \begin{cases} y_i^\star & \text{if } y_i^\star < 4 \\ 4 & \text{if } y_i^\star \geq 4 \end{cases} \tag{1}$$

- Modeled using tobit link for right censoring in VGAM (Yee, 2015)
- Predictors: 216 GPA, Previous GPA, Curricula (3 levels), and 216 GPA and Curricula interaction
- LRT Interaction (216 Grade:Curricula) test: $\chi^2_2 = 3.9$, p-value=0.141
- Similar results when treating 217 grade as an ordinal response, with and without accounting for semester and section in semester

Additive model, censored grade response

- LRT Curricula: $\chi^2_2 = 18.9$, p-value$<$0.0001
- 216 GPA slope: 0.476, SE=0.0766, p-value$<$0.0001
- Previous GPA slope: 1.144, SE=0.104, p-value$<$0.0001
- Pairwise comparisons of curricula:
  - CATALST (vs DVB): -0.38, SE=0.091, p-value$<$0.0001
  - Lock (vs DVB): -0.273, SE=0.095, p-value=0.004
  - Lock (vs CATALST): 0.107, SE=0.103, p-value=0.299

Censored regression model termplots



Conclusions

- Some differences in 217 performance based on curricula but impacts of previous performance levels is much more pronounced (how well they did mattered more than what curriculum they used)
- Cautions: Student selection into CATALST/ active learning/ TEAL rooms/ schedule (MWF vs TTh) could be biasing comparisons

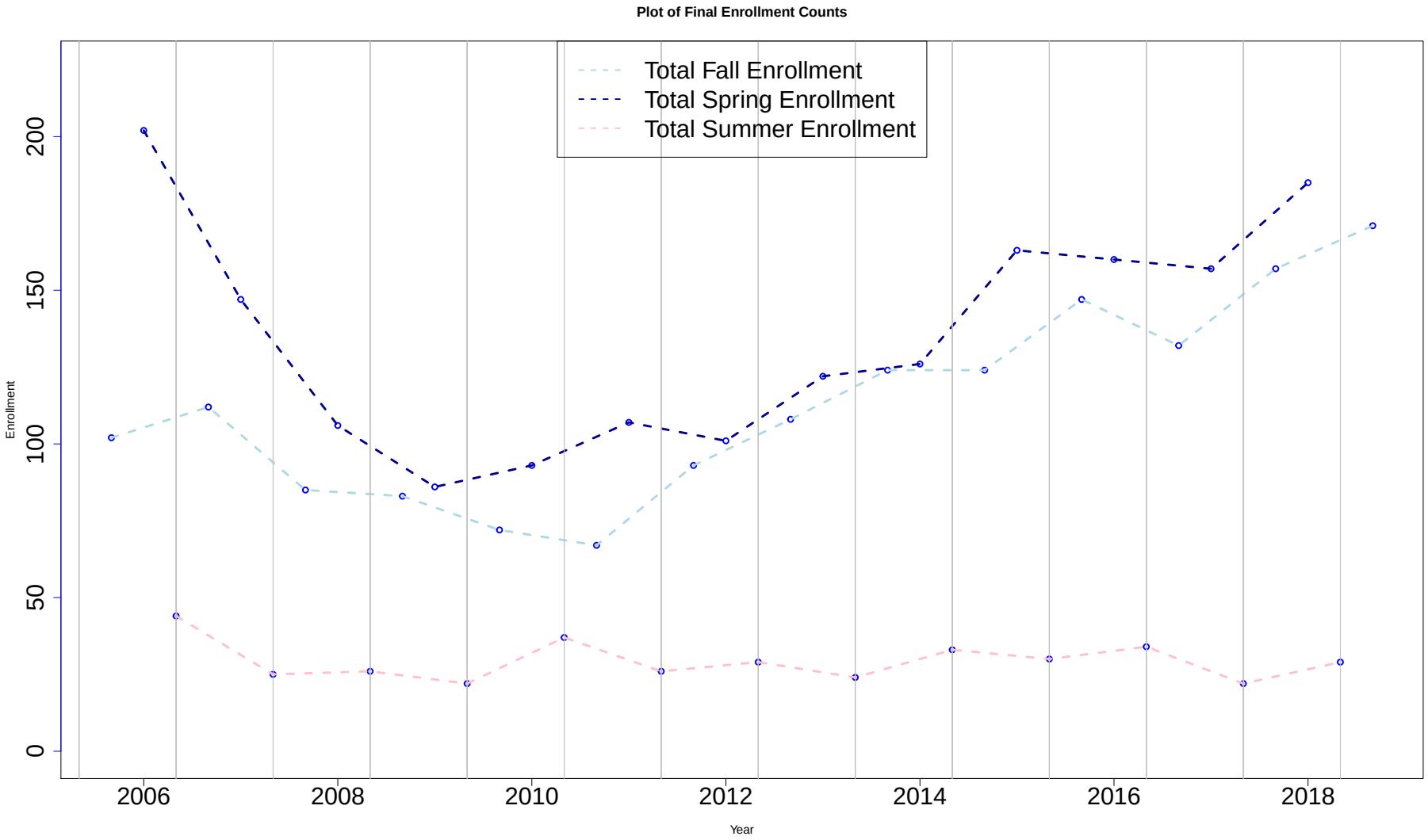
Next work

- Impacts of 216 curricula on "liking" of statistics on first day of 217 or rates of taking 217
- Update data set to try to obtain additional students that took 217 from this cohort of 216 students - might increase the "Lock" part of the data set
- Explore AP/non-MSU/non-216 pre-requisite students vs MSU 216 for 217 performance (data on "how" meeting pre-requisite was inconsistent)
- Censored response mixed model to account for semester and section in semester impacts

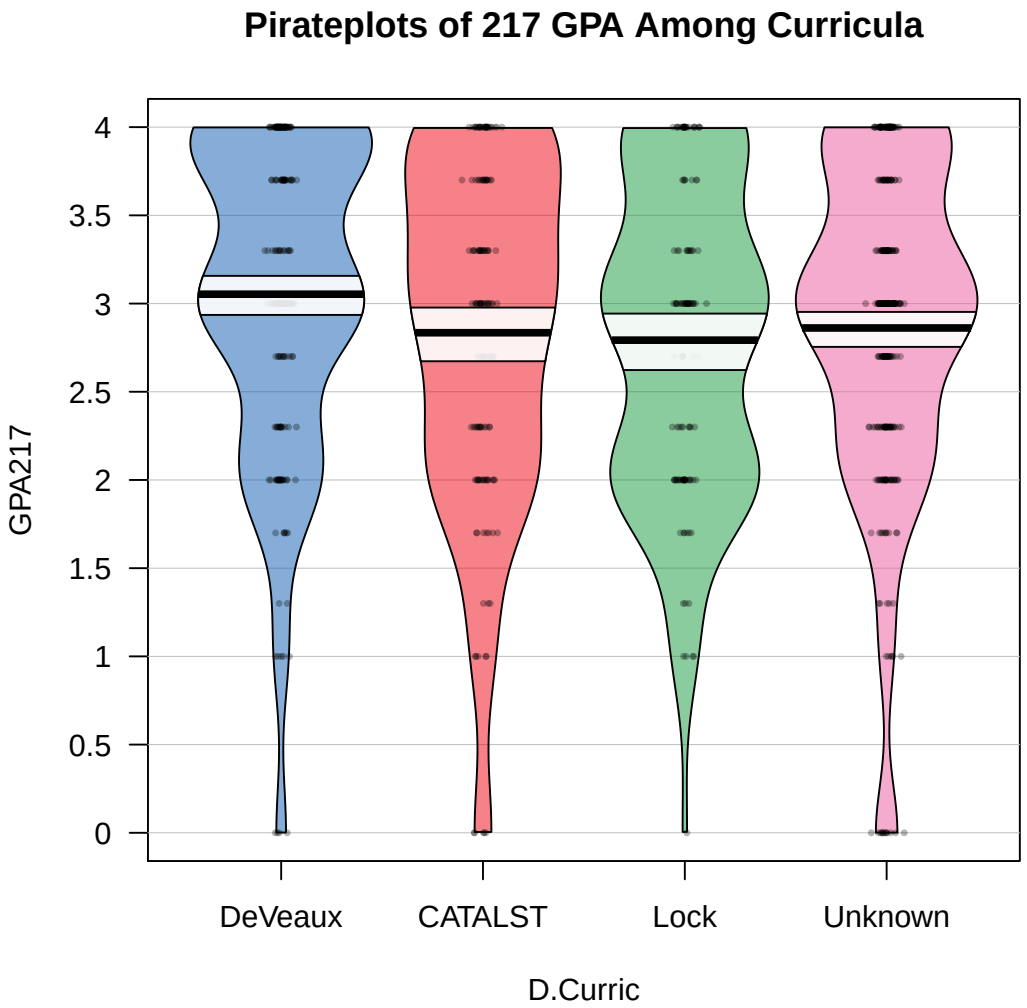
Citations:

- Fox, J. (2003) Effect Displays in R for Generalised Linear Models, *Journal of Statistical Software*, 8(15), 1-27.
- Fox, J. and Weisberg, S. (2011) *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage.
- Hildreth, L., Robison-Cox, J., and Schmidt, J. (2018) Comparing Student Success and Understanding in Introductory Statistics Under Consensus and Simulation-Based Curricula, *Statistical Education Research Journal*, 17(1), 103-120.
- Kampstra, P. (2008) Beanplot: A Boxplot Alternative for Visual Comparison of Distributions, *Journal of Statistical Software*, Code Snippets, 28(1). 1-9.
- Phillips, N. (2017) yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R". R package version 0.1.5.
- Pruim, R., Kaplan, D., and Horton, N. (2017) The mosaic Package: Helping Students to 'Think with Data' Using R, *The R Journal*, 9(1):77-102.
- Tennekes, M. and de Jonge, E. (2017) tabplot: Tableplot, a Visualization of Large Datasets. R package version 1.3-1.
- Wei, T. and Simko, V. (2017) R package "corrplot": Visualization of a Correlation Matrix (Version 0.84).
- Yee, T. (2015). *Vector Generalized Linear and Additive Models: With an*

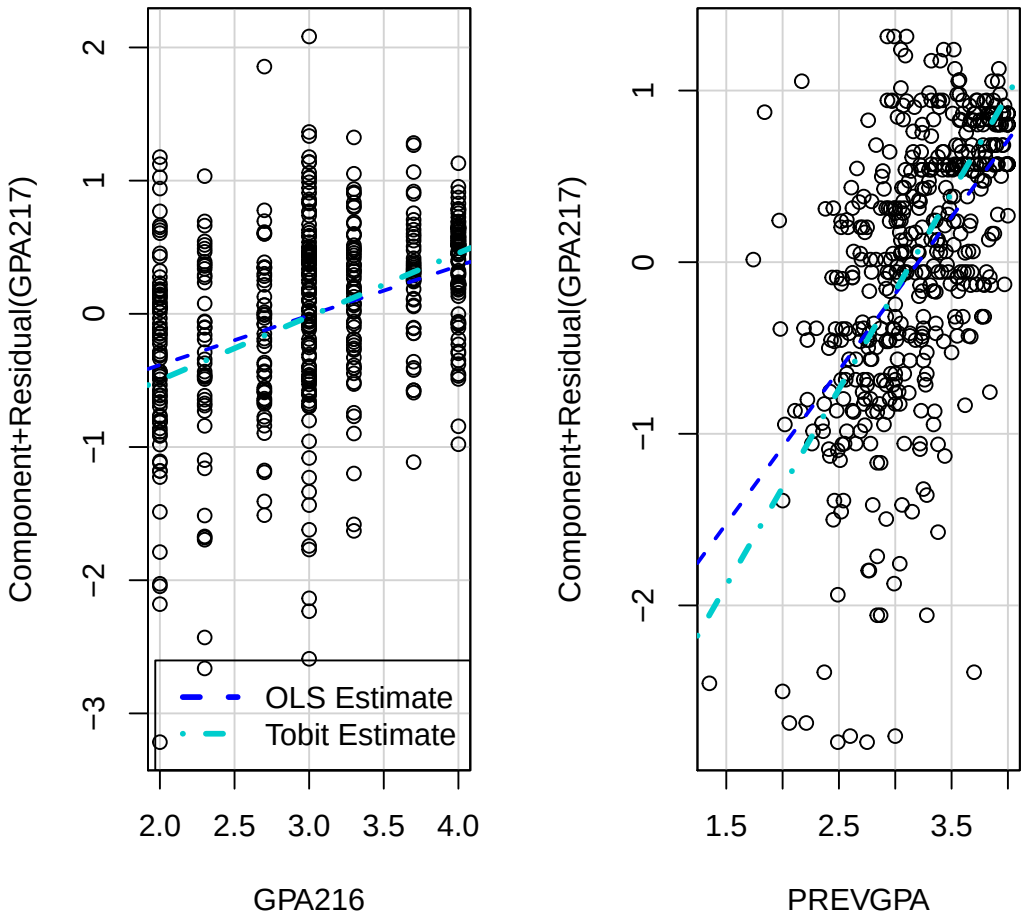
STAT 217 Enrollment Trends: By Semester



Or with Pirate Plots (from yarrr)



Comparison of OLS vs Censored Regression estimates



WDF rates not impacted by 216 curricula

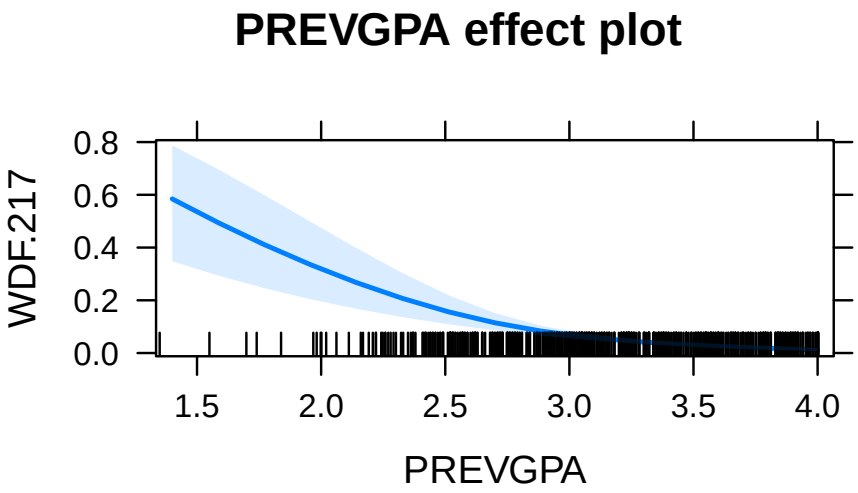
- Includes W students in data set (now $N = 515$)
- Logistic mixed model, accounting for semester and section in semester (similar results for logistic model)

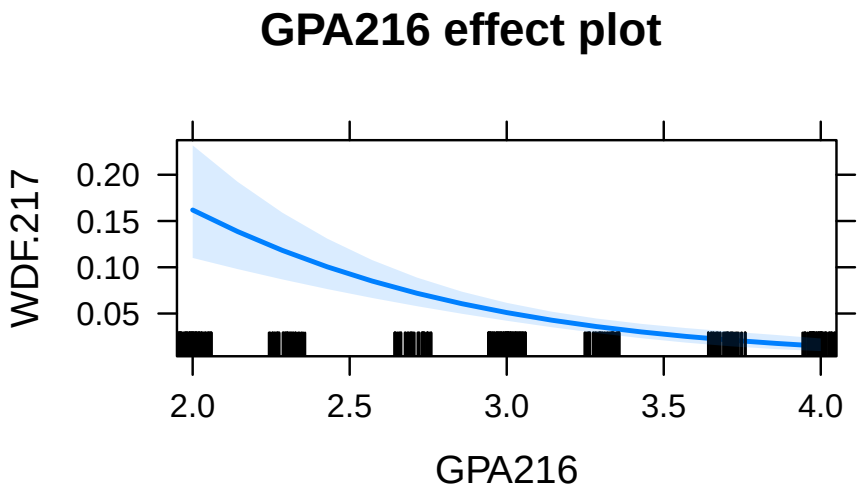
	Chisq	Df	Pr(>Chisq)
PREVGPA	18.673874	1	0.0000155
GPA216	13.254357	1	0.0002719
D.Curric	1.663231	2	0.4353454
GPA216:D.Curric	1.394054	2	0.4980637

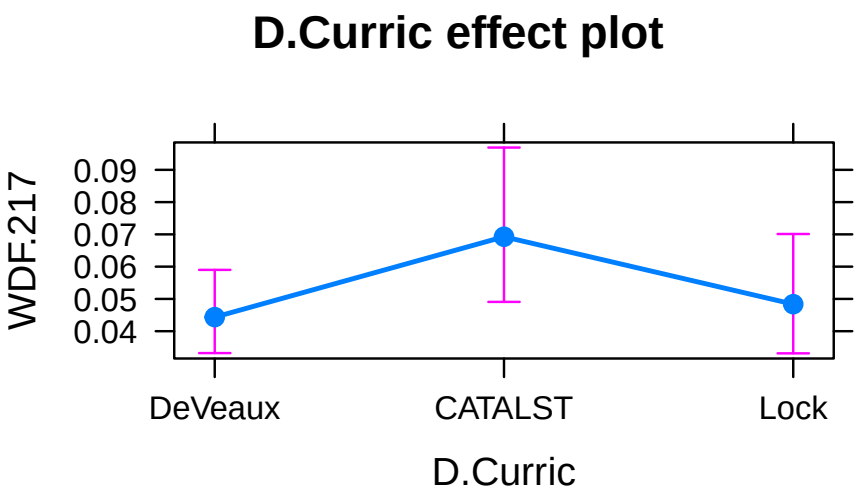
217 WDF rates not impacted by 216 curricula

	Chisq	Df	Pr(>Chisq)
PREVGPA	19.317829	1	0.0000111
GPA216	13.932791	1	0.0001895
D.Curric	1.666666	2	0.4345984

217 WDF rates not impacted by 216 curricula







Details on organization of book

- Ch 1: (R)e-Introduction to Statistics
  - Introduction to all methods to be covered and when to use them
  - 2-sample mean test and CI, parametric equal variance
  - Permutation test and bootstrap CI for the difference in the means
  - Basic R-studio-based R introduction
  - In class activities require installing and then using R-studio
- Ch 2: One-Way ANOVA
  - Overall F-test and permutation test
  - Tukey's HSD
  - Cell-means and reference coding
  - Normality and equal variance diagnostic plots
- Ch 3: Two-Way ANOVA
  - Interaction plots (my function)
  - Additive and interaction models
  - Testing for interactions and then main effects using F-tests
  - car's Anova function for Type II tests

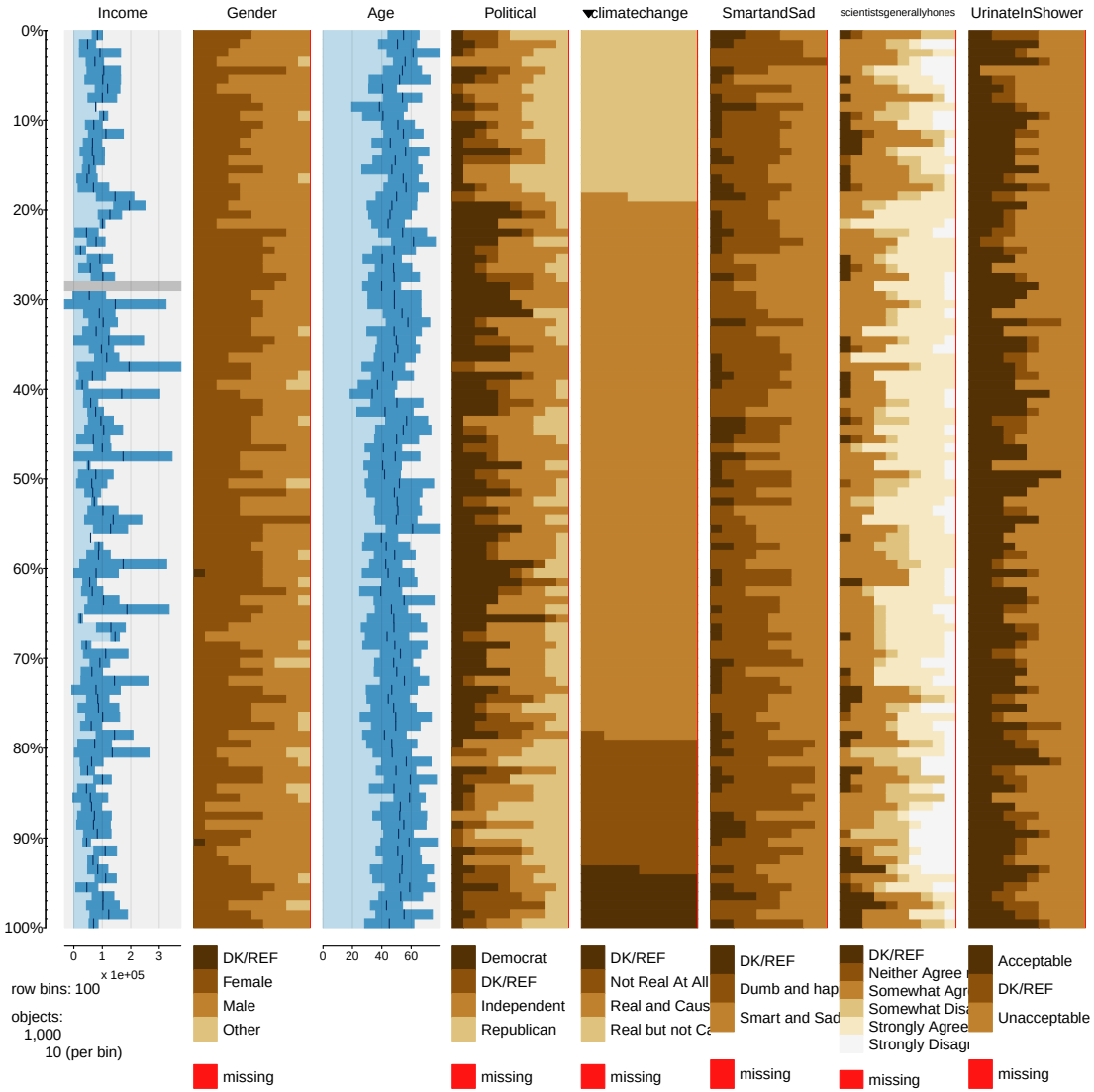
Organization of book

- Ch 4: Chi-square tests
  - Contingency tables, mosaic plots with standardized residuals
  - Distinguishing between independence and homogeneity tests
  - Connecting Chi-square distribution with permutation distribution
- Ch 5: Correlation and SLR
  - Bootstrap CI for the correlation coefficient
  - Scatterplots and SLR interpretations
- Ch 6: SLR inference
  - Permutation and parametric test
  - Confidence intervals for slopes
  - Diagnostics including Cook's D
  - log-transformations for x, y, or both
  - Confidence and Prediction intervals

Organization of book

- Ch 7: Multiple Linear Regression
  - multiple quantitative predictors, then with 1 categorical and 1 interaction
  - Multi-collinearity (VIFs and GVIFs)
  - Revisit diagnostics
  - Overall F-test, t-tests for individual slopes, t or F-test for interactions
  - Model selection using adjusted R-sq and AIC
- Ch 8: Case studies
  - Real data sets from *Biology Letters* analyzed using techniques from the semester
  - Attempted to show the power of the suite of methods to handle different types of problems and review identifying which method to use

Tableplot of Pulse of Nation, sorted on climatechange



Without non-response, sorted on SmartandSad