2016 Election: Revisited

STAT 545 Project by Travis Benedict and Jordan Pflum
November 30, 2018

Start

O V E R V I E W

2016 Election Background
Brief overview of the 2016 Election and statistical policial leans commonly employed

Data Overview
Describe data sets and predictors contained within.

Exploratory Analysis
Initial Exploration of Data

Project Goal
Central guiding question of project

Variable Selection
Explanation of techniques and methods used

Application and Comparison
Application to Ordinal Regression and comparison to other models

Conclusion
Concluding thoughts

Q&A

2016 Election Background

Historical Information

Candidates

The Republican ticket was held by businessman **Donald Trump** while the Democratic ticket was held by former Secretary of State **Hillary Clinton**

Result

While Donald Trump **lost the popular vote** to Hillary Clinton by more than 2.8 million votes, he **won 30 states** and a decisive **304 electoral votes** compared to 227, becoming the **47th president**

Message

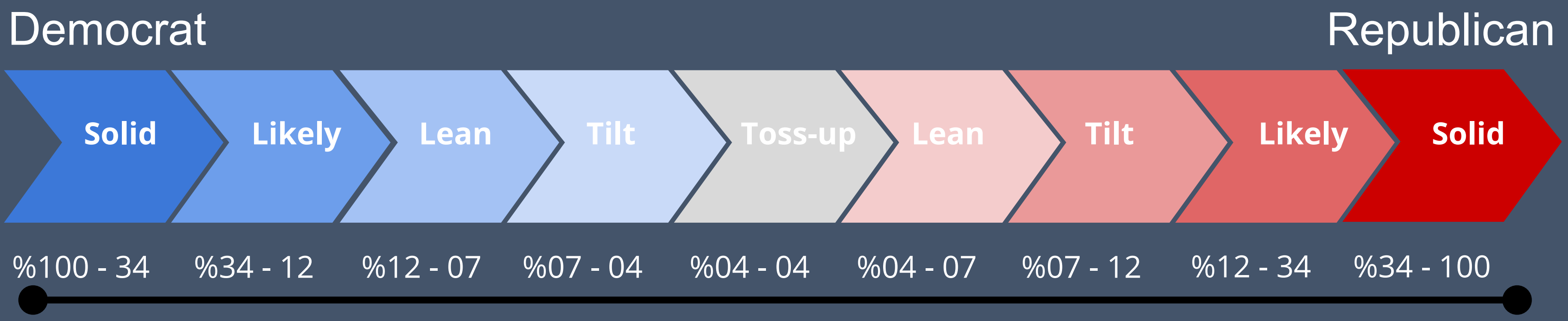
Donald Trump’s populist, nationalist campaign, promising to “**Make America Great Again**”, starkly contrasted Clinton’s expansion and promotion of **racial**, **LGBTQ**, and **women’s rights**

Key Demographics

Trump’s appeal to **white working-class** voters outside major cities in **pivotal manufacturing states** proved to be a key factor.

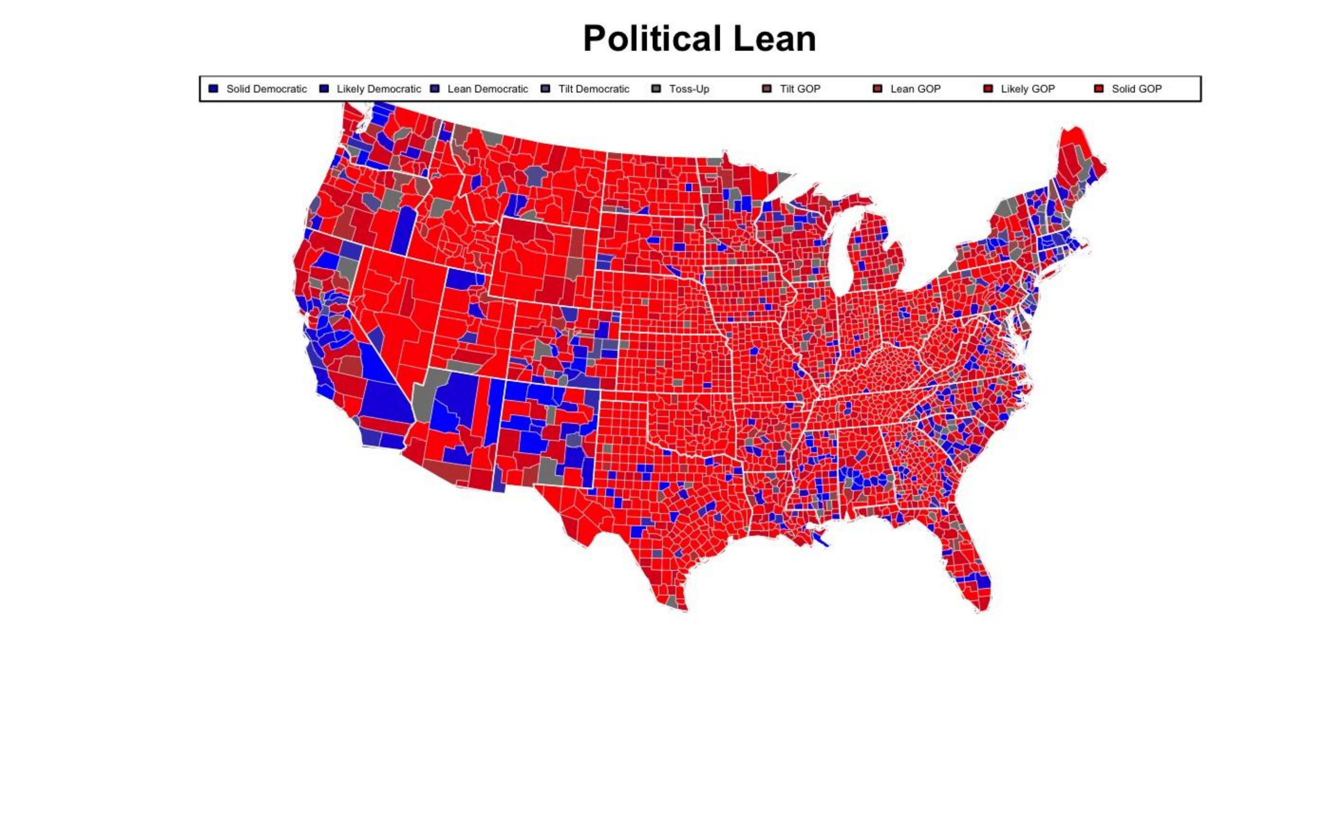
2016 Election Background

Statistical Political Lean



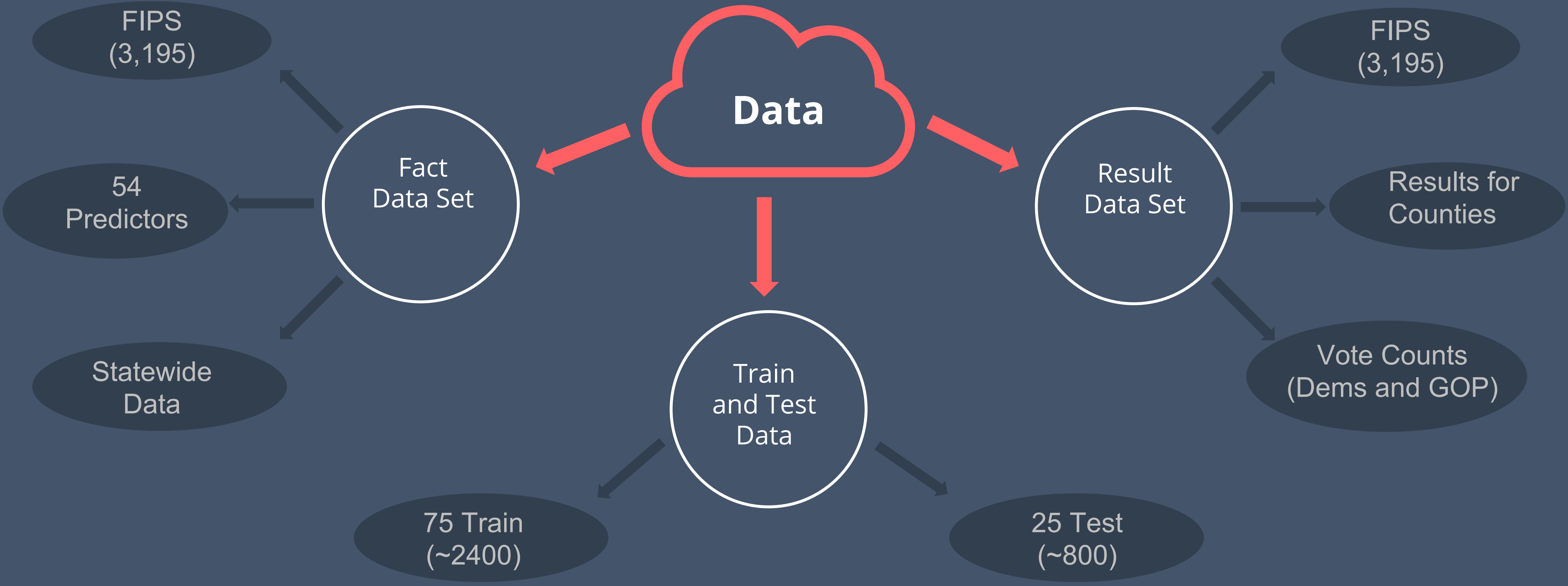
2016 Election Background

Statistical Political Lean Map



Data Overview

Data Sets Description



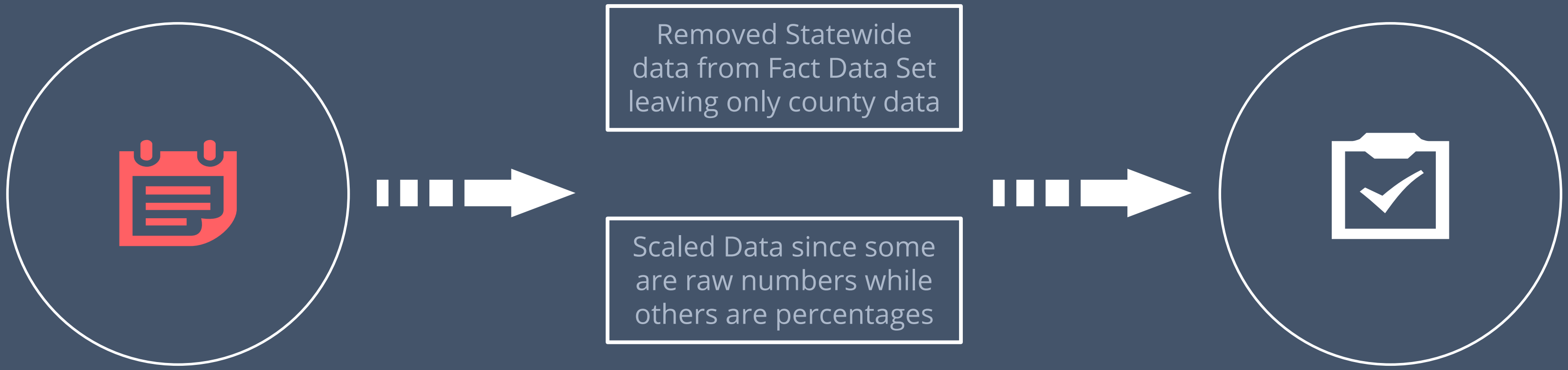
Data Overview

Fact Data Set Predictors (Examples)

Predictor Code	Predictor Description
RHI725214	Hispanic or Latino, percent, 2014
POP815213	Language other than English spoken at home, pct age 5+, 2009-2013
EDU685213	Bachelor's degree or higher, percent of persons age 25+, 2009-2013
INC910213	Per capita money income in past 12 months (2013 dollars), 2009-2013
PVY020213	Persons below poverty level, percent, 2009-2013
PST045214	Population, 2014 estimate
AGE295214	Persons under 18 years, percent, 2014
AGE775214	Persons 65 years and over, percent, 2014
SEX255214	Female persons, percent, 2014
RHI225214	Black or African American alone, percent, 2014
VET605213	Veterans, 2009-2013

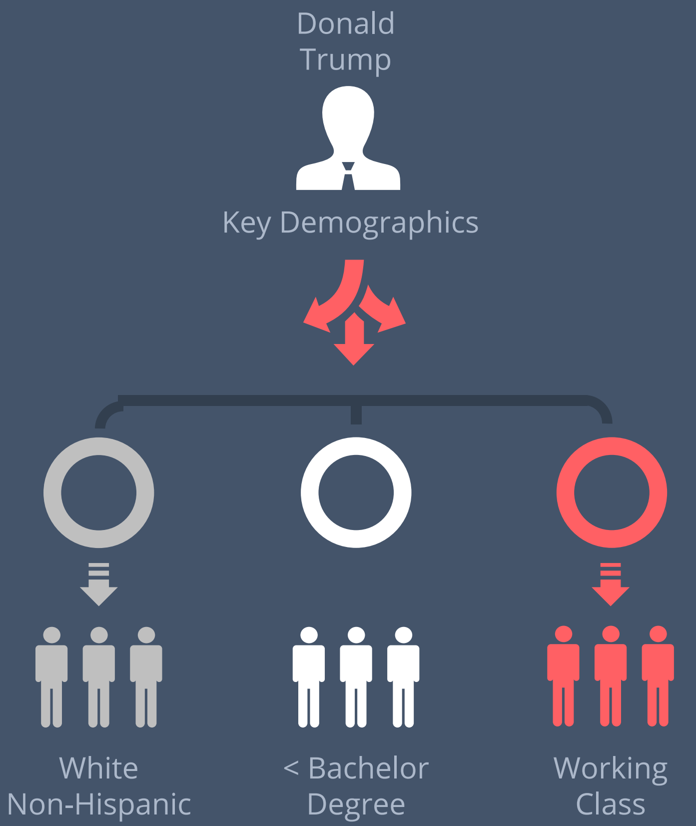
Data Overview

Cleaning the Data



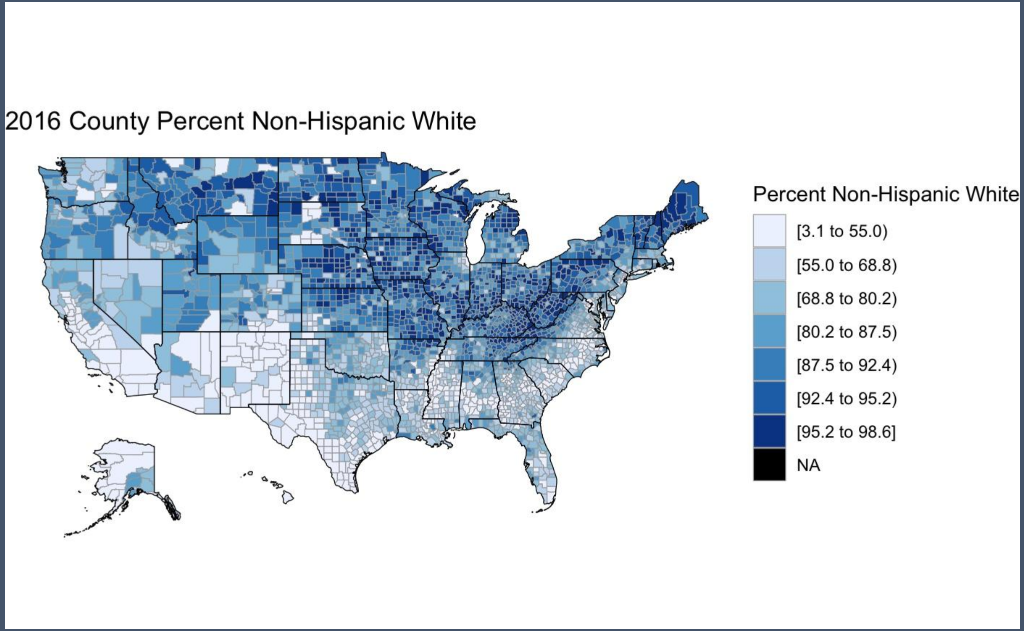
Explanatory Analysis

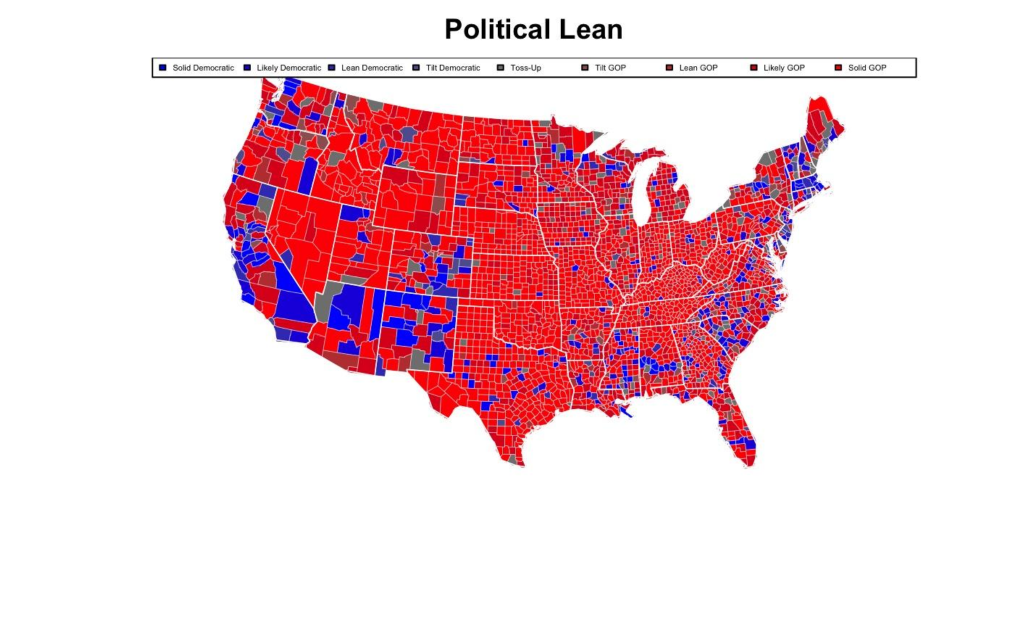
Trump's Base



Explanatory Analysis

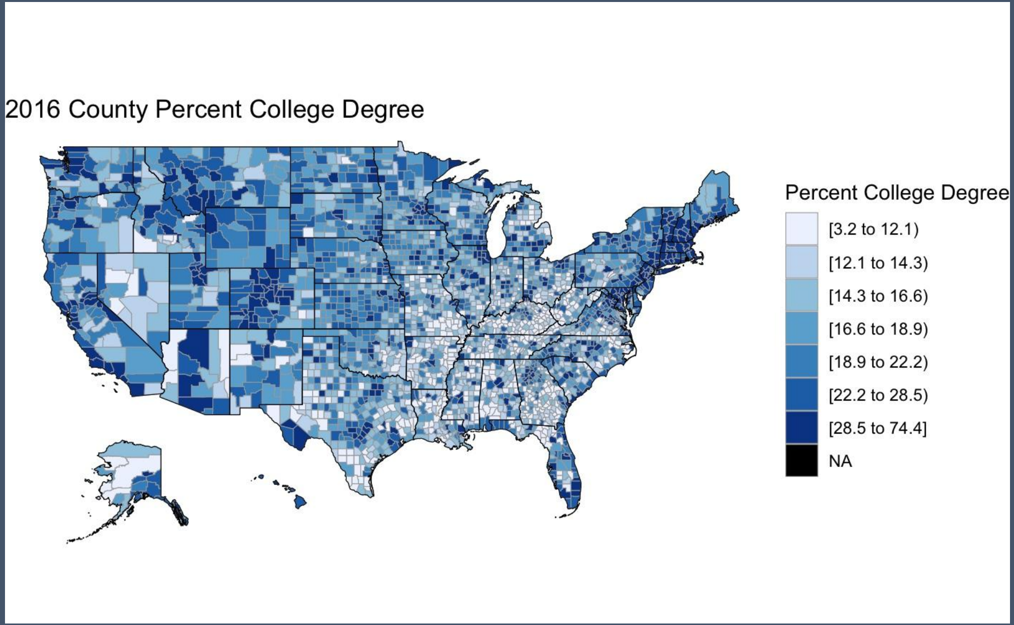
White Non-Hispanic

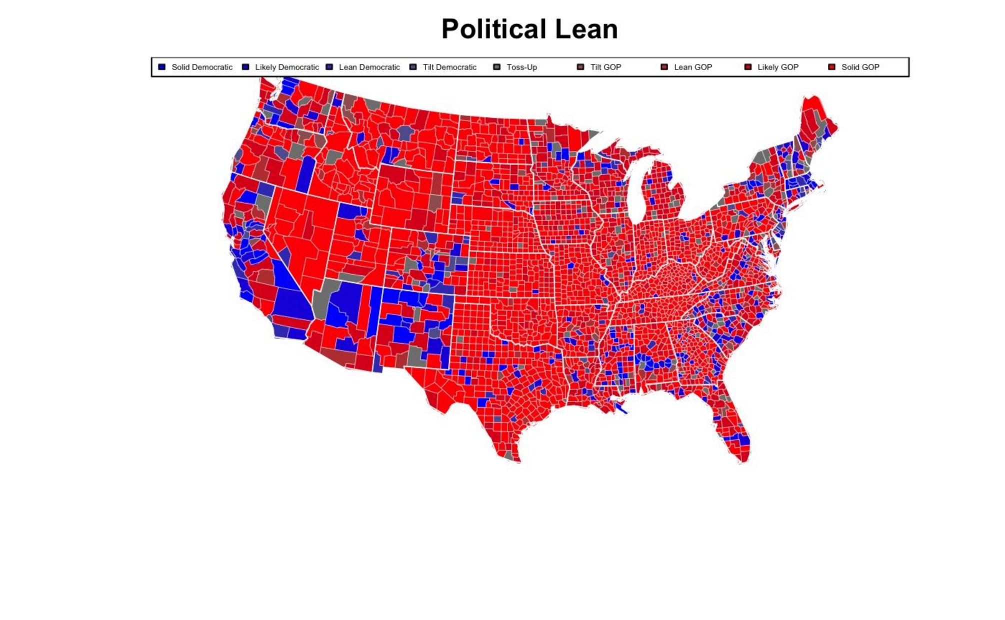




Explanatory Analysis

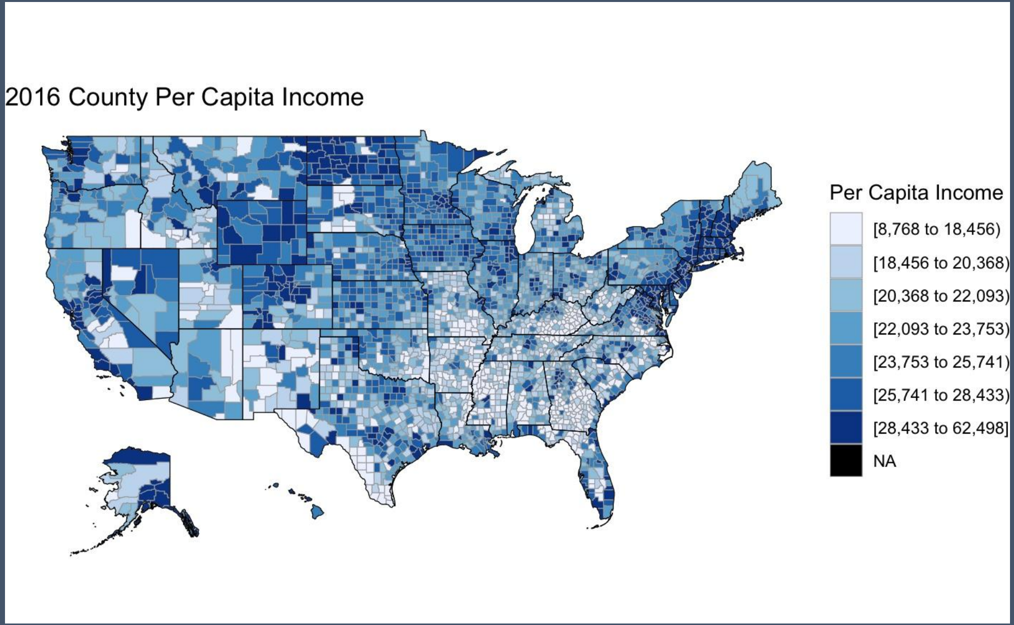
< Bachelor Degree

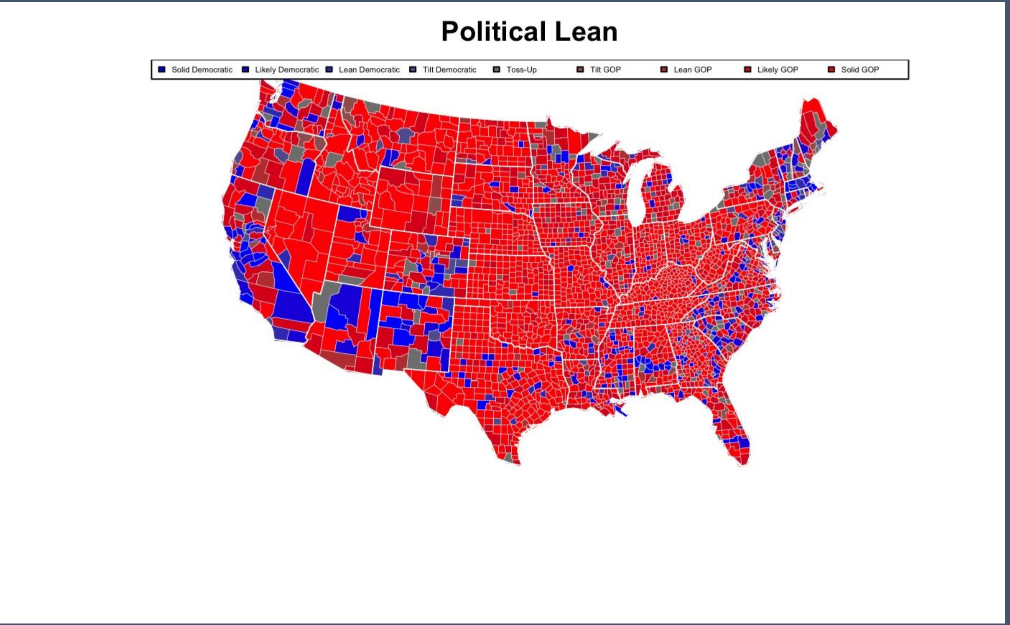




Explanatory Analysis

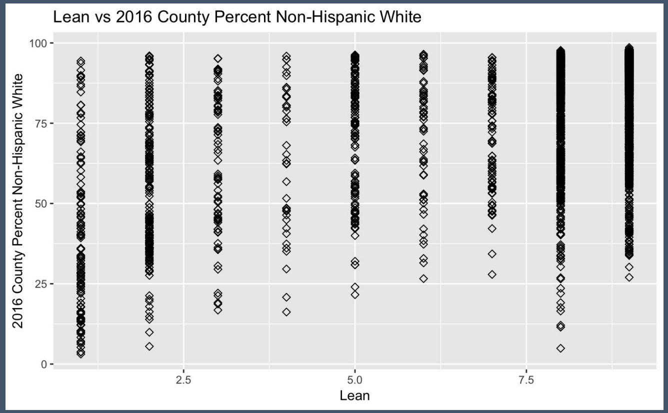
Working Class

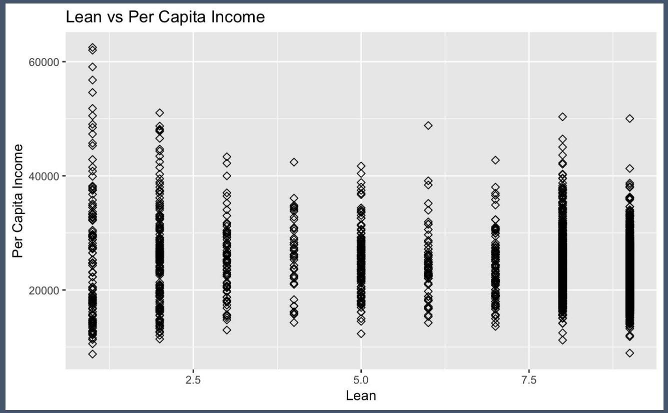


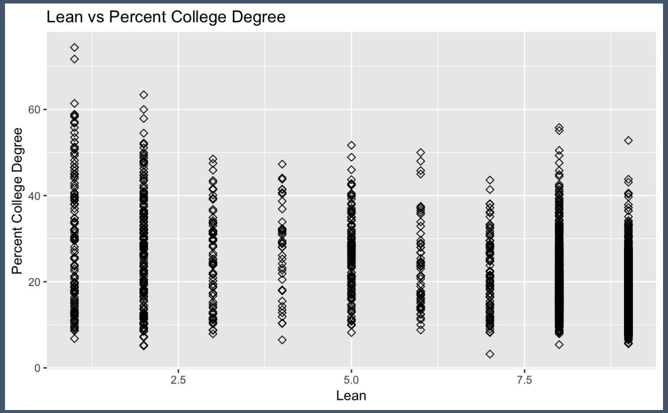


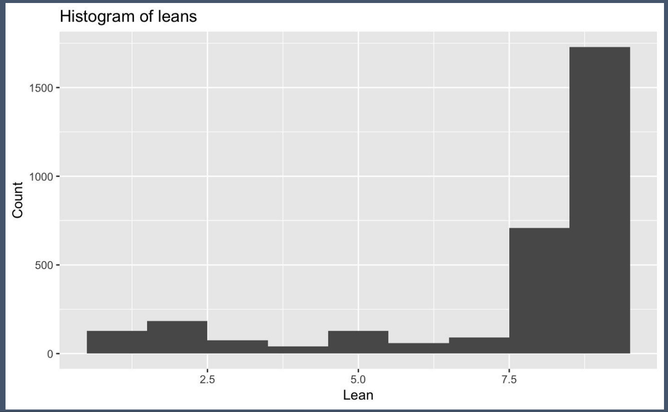
Explanatory Analysis

Scatter Plots and Histogram







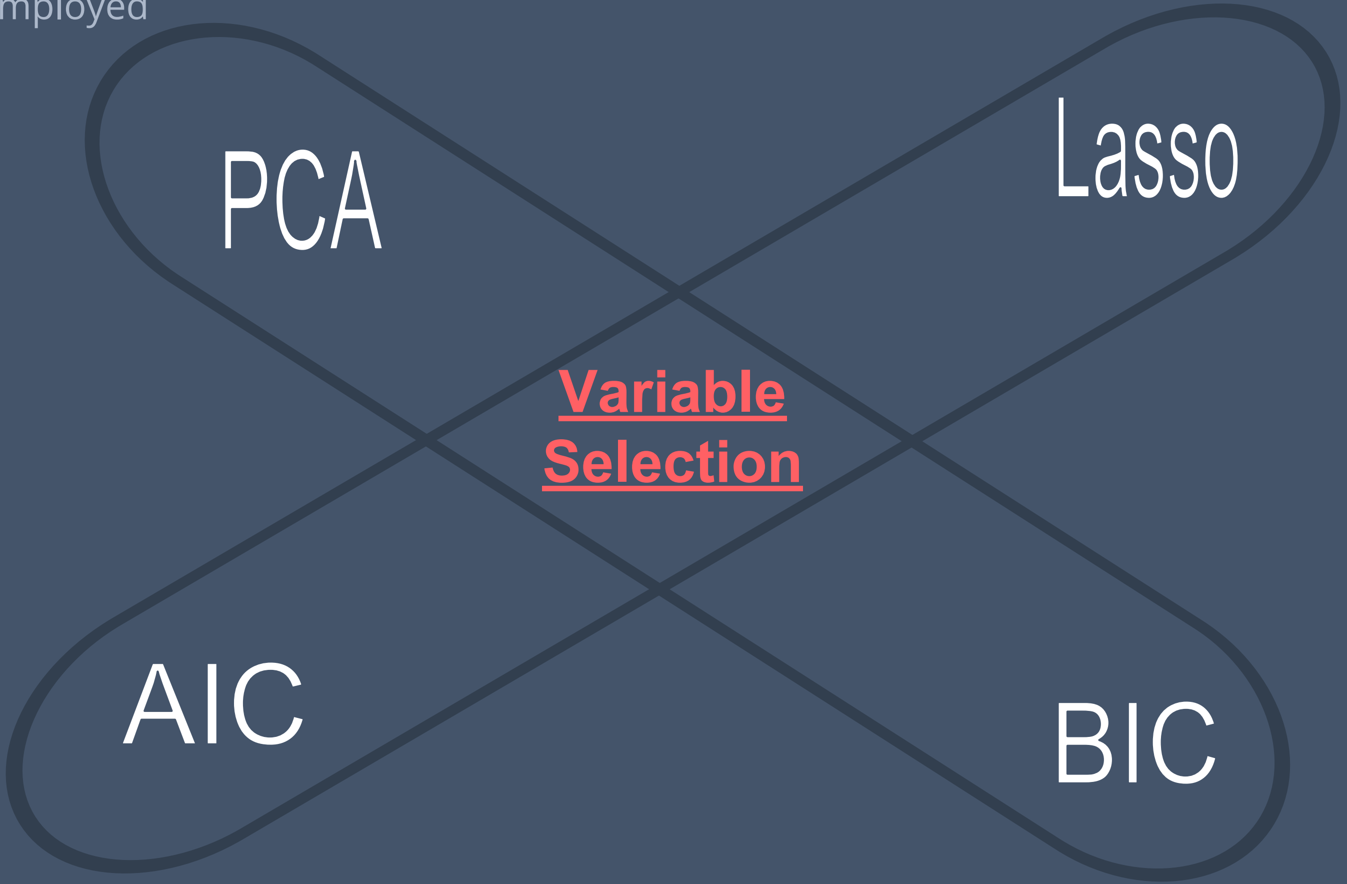


Central Question

Does the probability associated with a logistic regression outcome provide insight for ordinal regression?

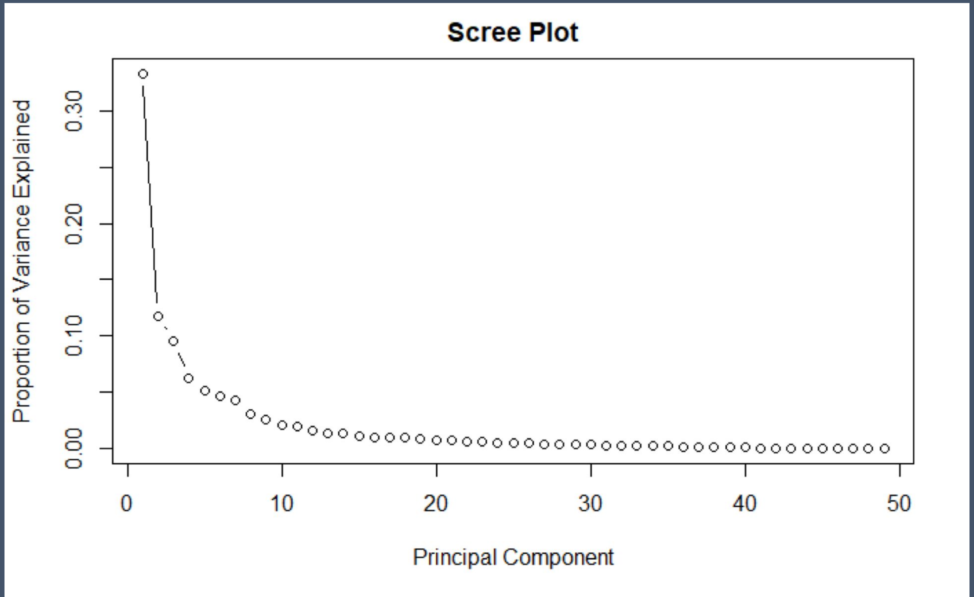
Variable Selection

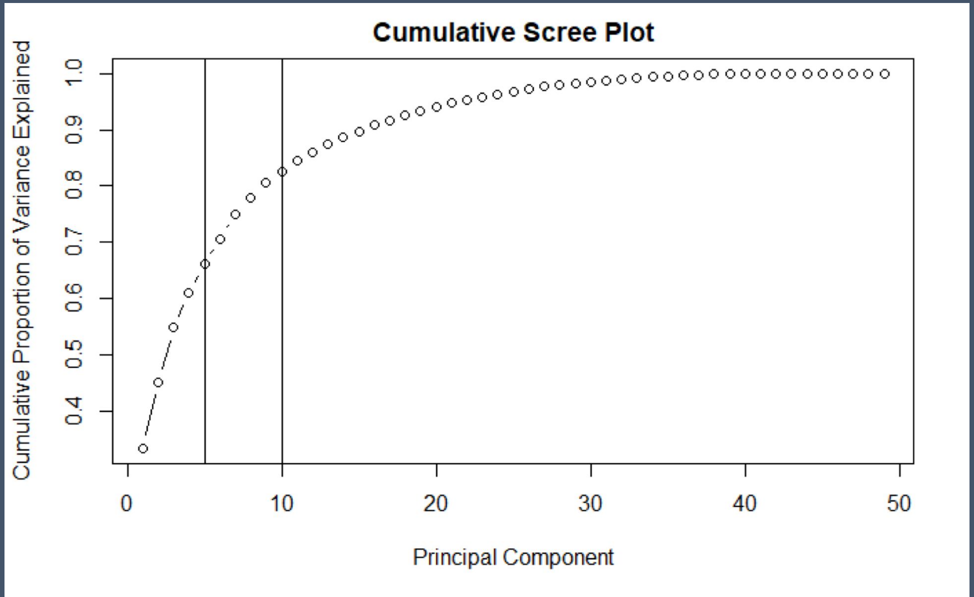
Techniques Employed



Variable Selection

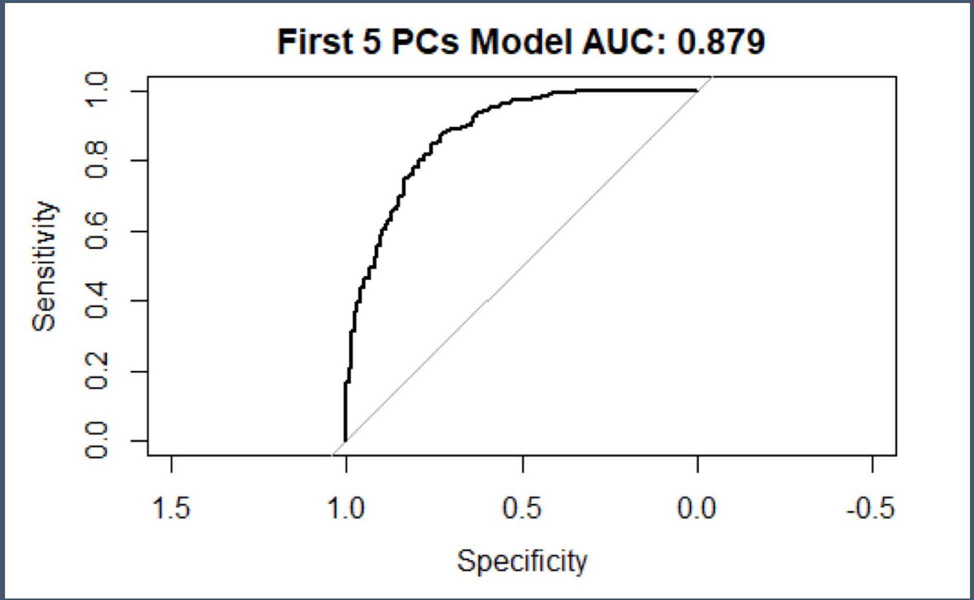
Method: PCA

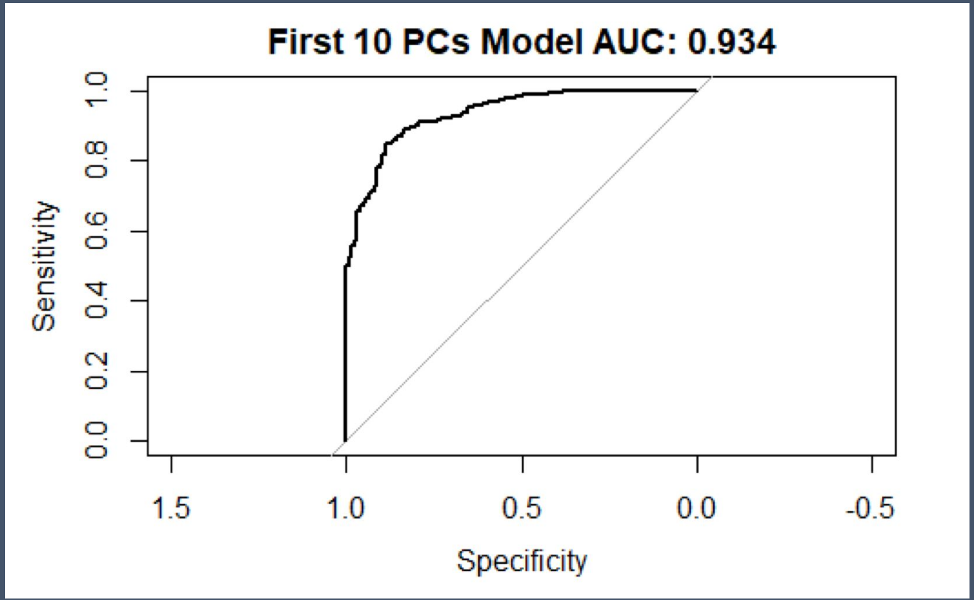




Variable Selection

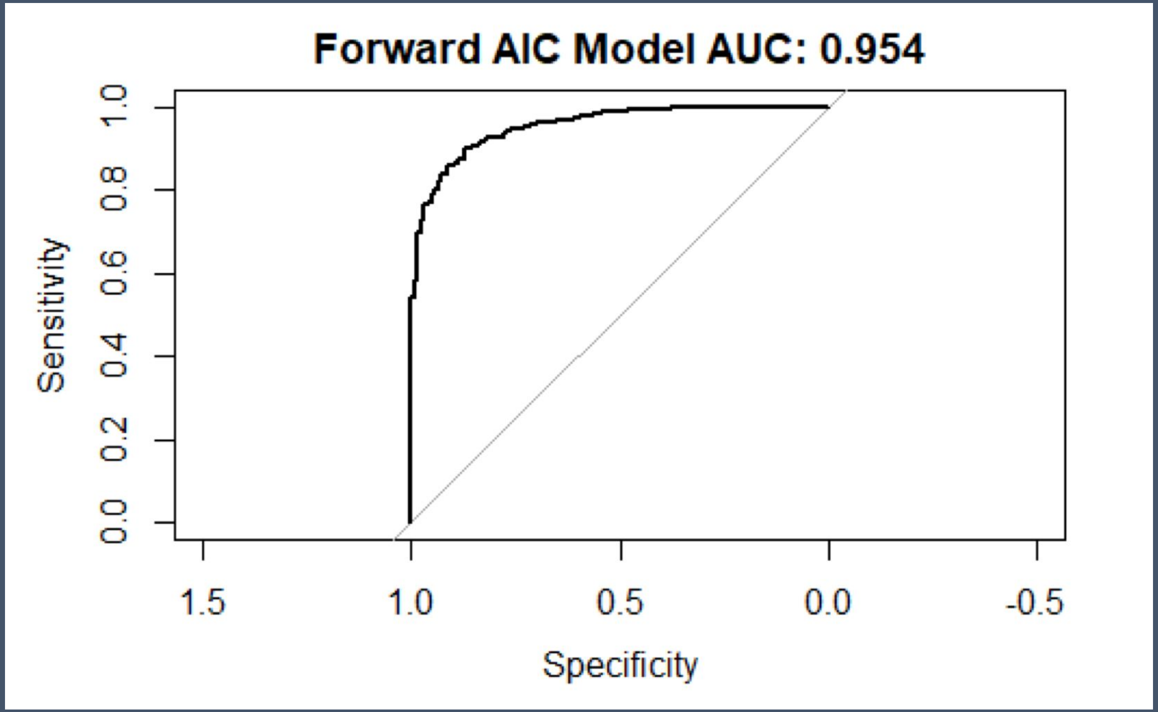
Method: PCA (cont.)





Variable Selection

Method: Forward AIC

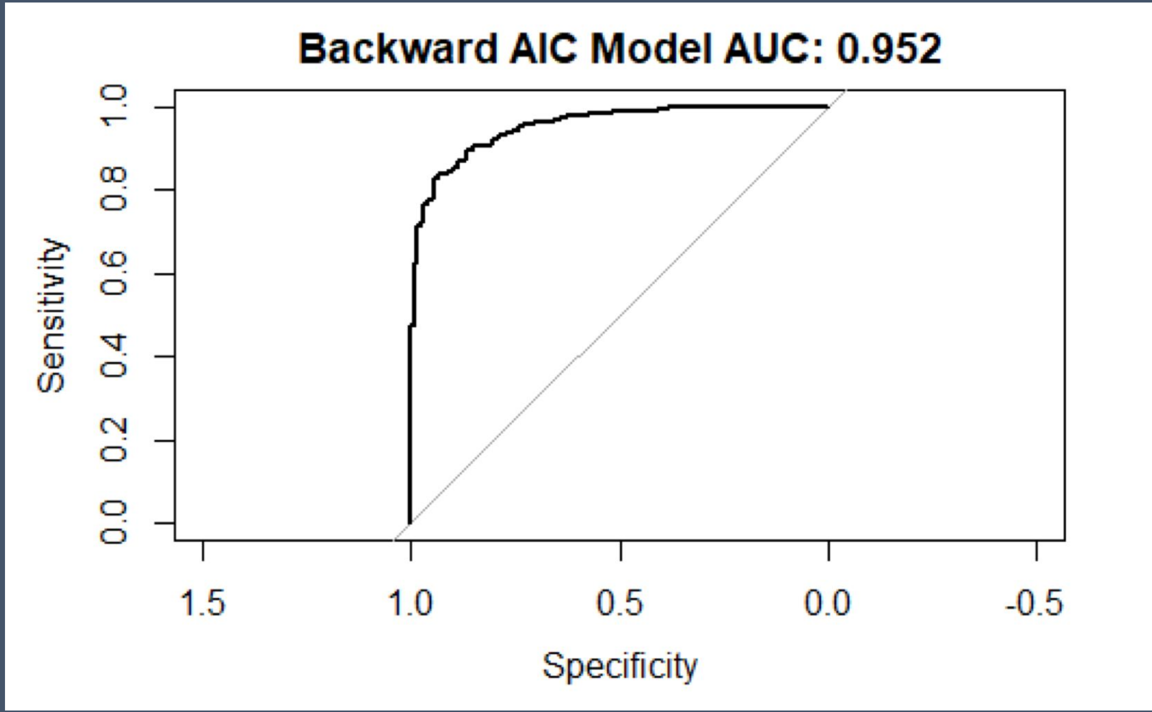


Variables Selected

"(Intercept)" "RHI825214" "EDU685213" "RHI225214" "HSG096213" "PST120214" "AGE295214" "SEX255214" "HSG495213" "INC910213" "POP715213" "SBO015207" "HSD310213" "AGE775214" "LFE305213" "INC110213" "POP645213" "RHI625214" "SBO215207" "SBO315207" "AGE135214" "SBO515207"

Variable Selection

Method: Backward AIC

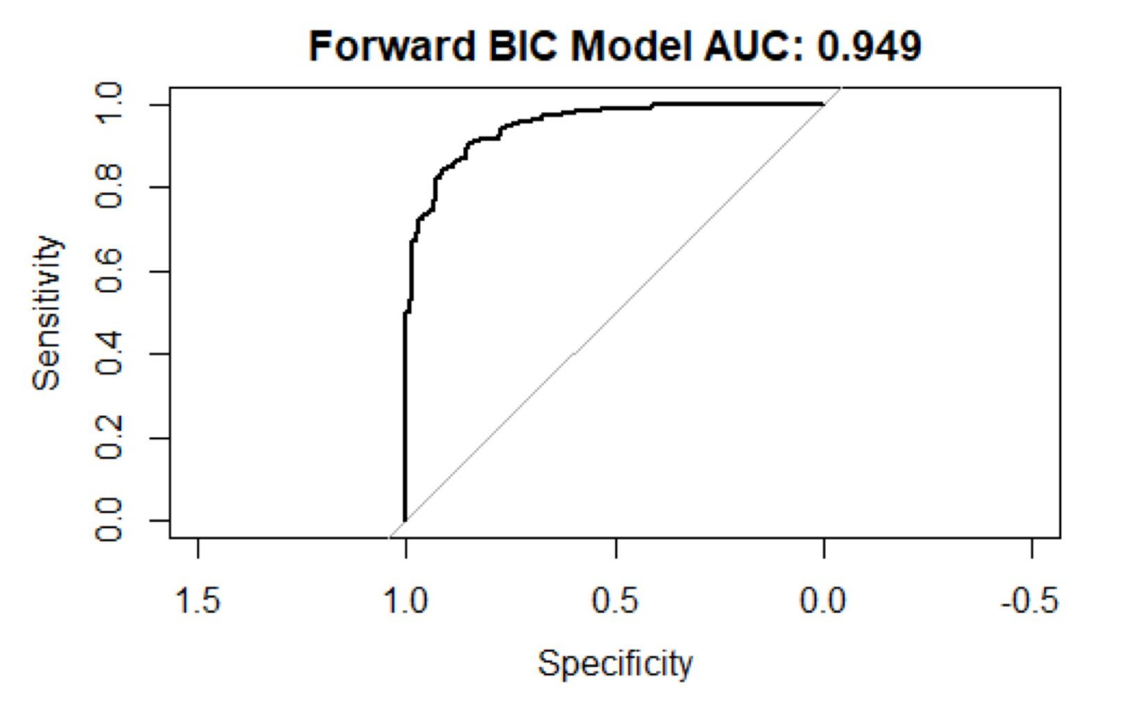


Variables Selected

"(Intercept)" "PST120214" "POP010210" "AGE135214" "AGE295214" "AGE775214" "RHI225214" "RHI425214" "RHI625214" "RHI725214" "RHI825214" "POP715213" "EDU635213" "EDU685213" "VET605213" "LFE305213" "HSG010214" "HSG096213" "HSG495213" "HSD410213" "HSD310213" "INC910213" "INC110213" "BZA010213" "NES010213" "SBO001207" "SBO315207" "SBO215207" "SBO015207" "RTN130207" "AFN120207”

Variable Selection

Method: Forward BIC

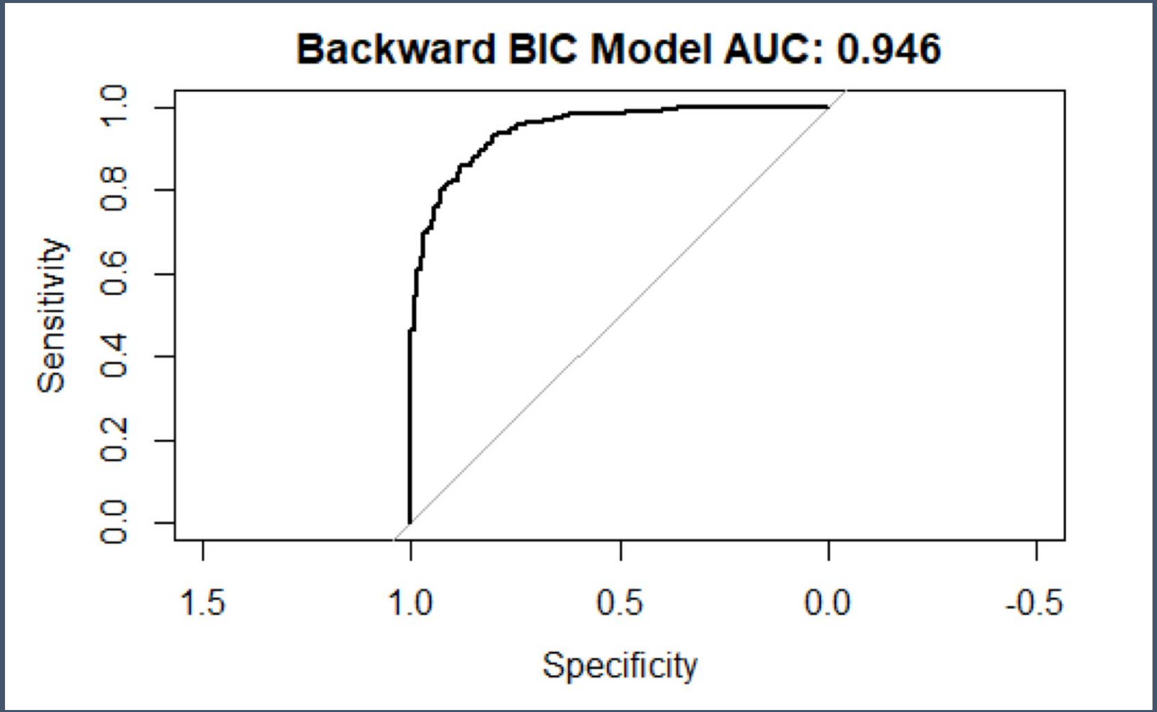


Variables Selected

"(Intercept)" "RHI825214" "EDU685213" "RHI225214" "HSG096213" "PST120214" "AGE295214" "SEX255214" "HSG495213" "INC910213" "POP715213" "SBO015207" "HSD310213" "AGE775214" "LFE305213"

Variable Selection

Method: Backward BIC

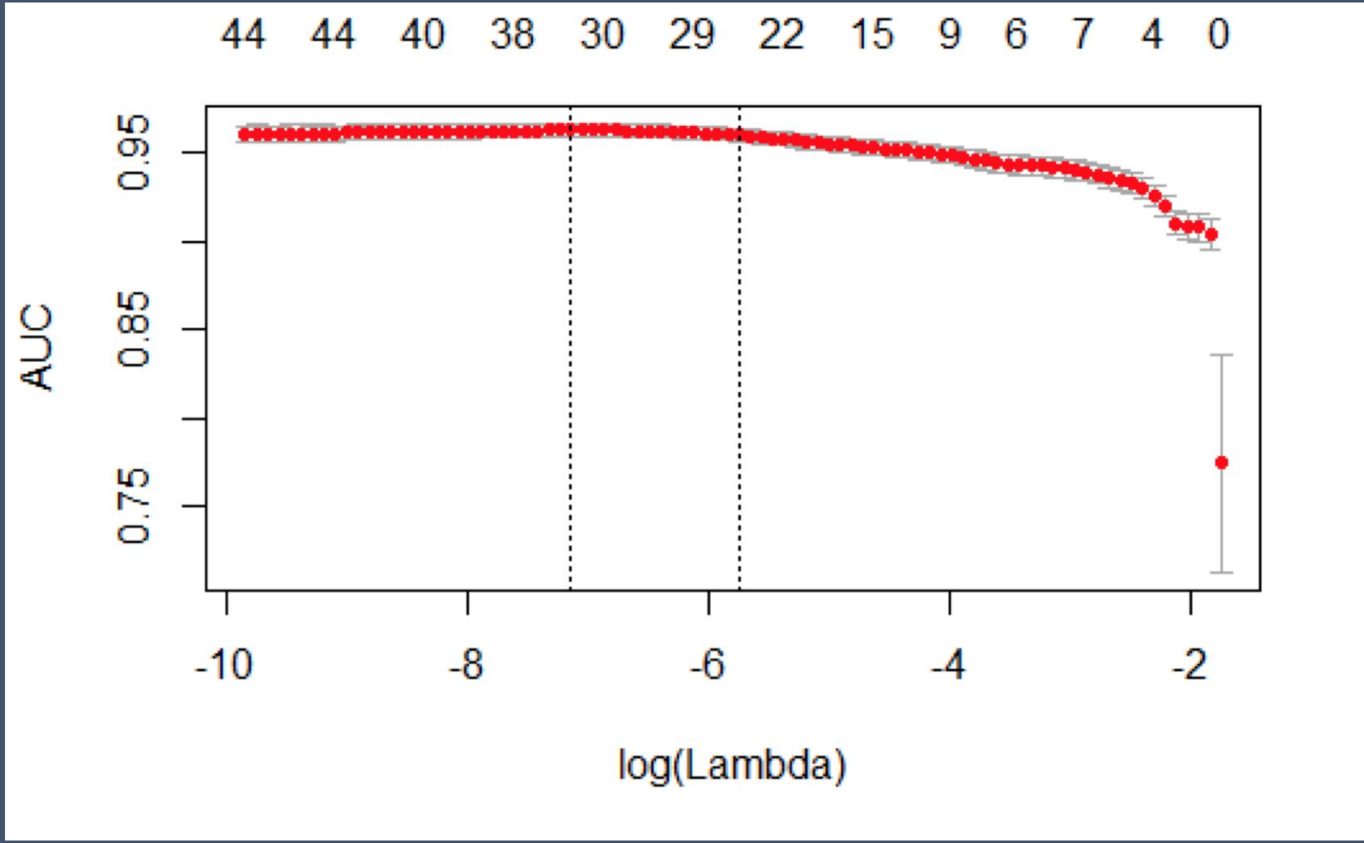


Variables Selected

"(Intercept)" "AGE295214" "RHI825214" "POP715213" "EDU685213" "LFE305213" "HSG010214" "HSG096213" "HSG495213" "HSD410213" "HSD310213" "INC910213" "SBO015207"

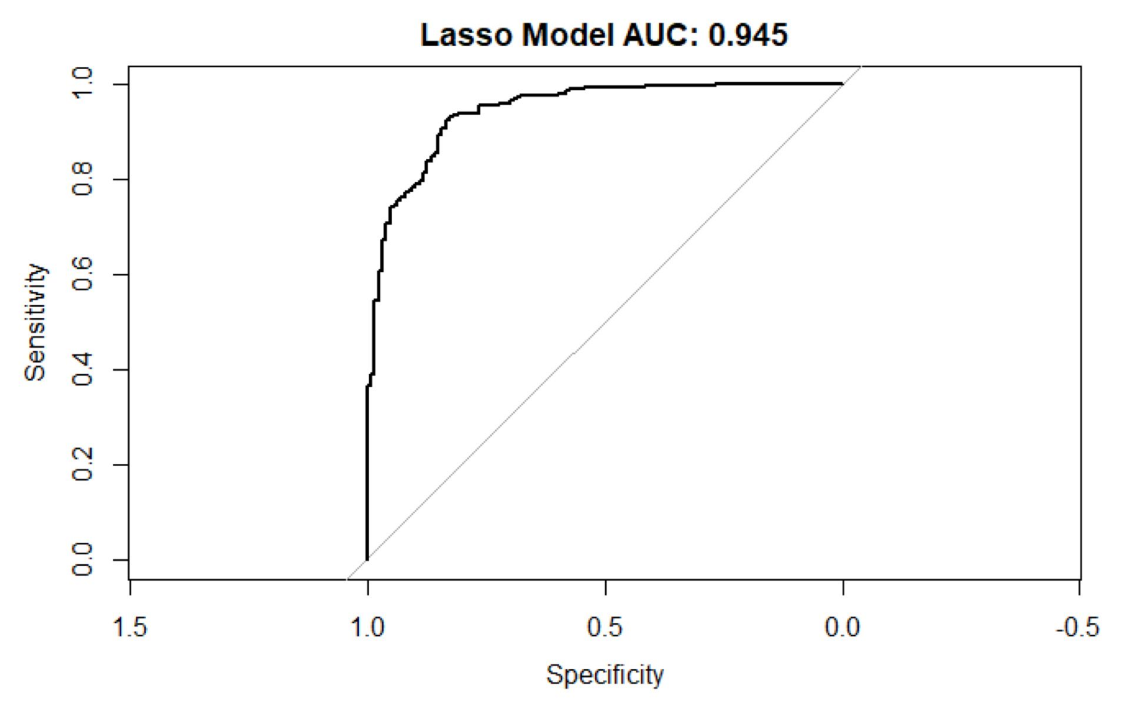
Variable Selection

Lasso Grid Search



Variable Selection

Method: Lasso

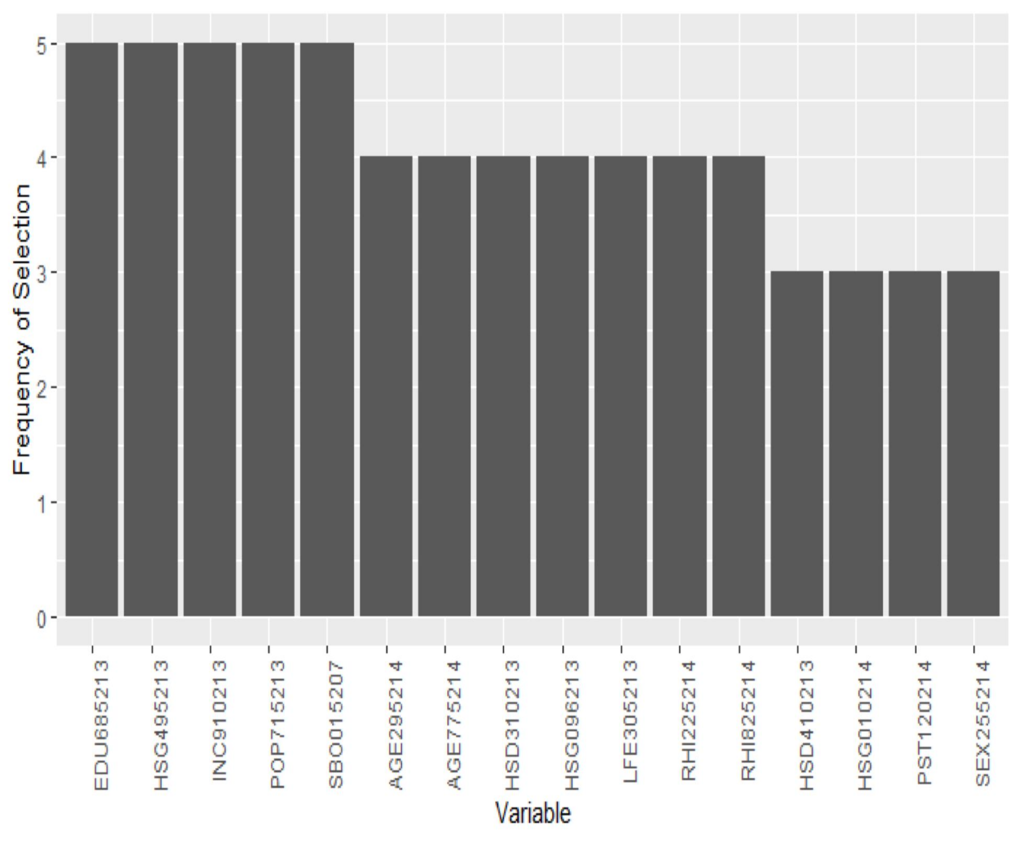


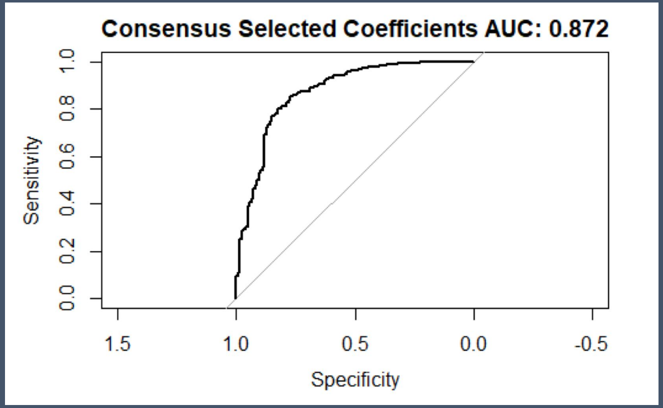
Variables Selected

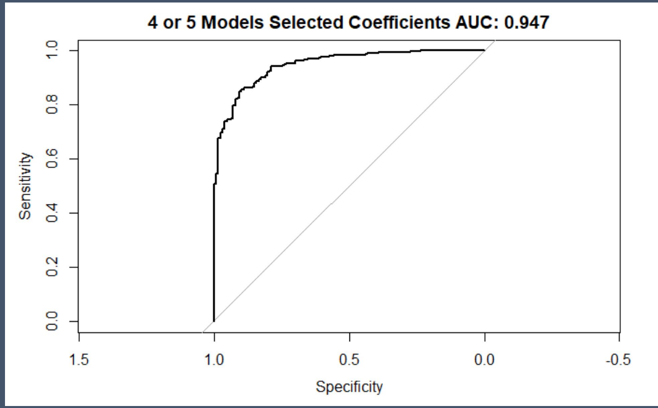
"PST045214" "POP010210" "AGE775214" "SEX255214" "RHI125214" "RHI225214" "RHI325214" "RHI725214" "POP715213" "POP645213" "POP815213" "EDU685213""VET605213" "HSG010214" "HSG495213" "HSD410213" "INC910213" "PVY020213" "BZA010213" "SBO115207" "SBO515207" "SBO015207" "MAN450207" "WTN220207" "BPS030214"

Variable Selection

Selected Coefficients

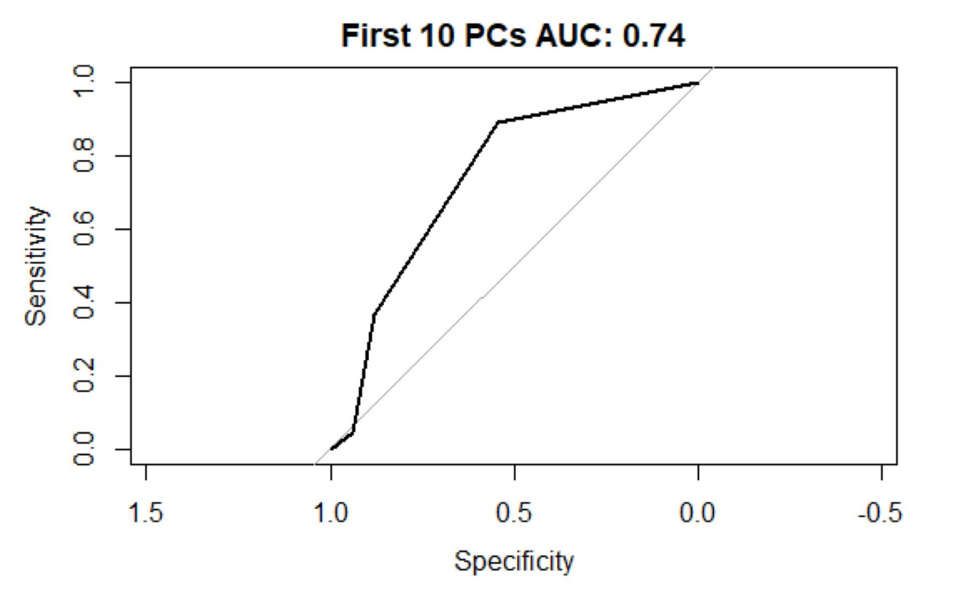






Application to Ordinal Regression

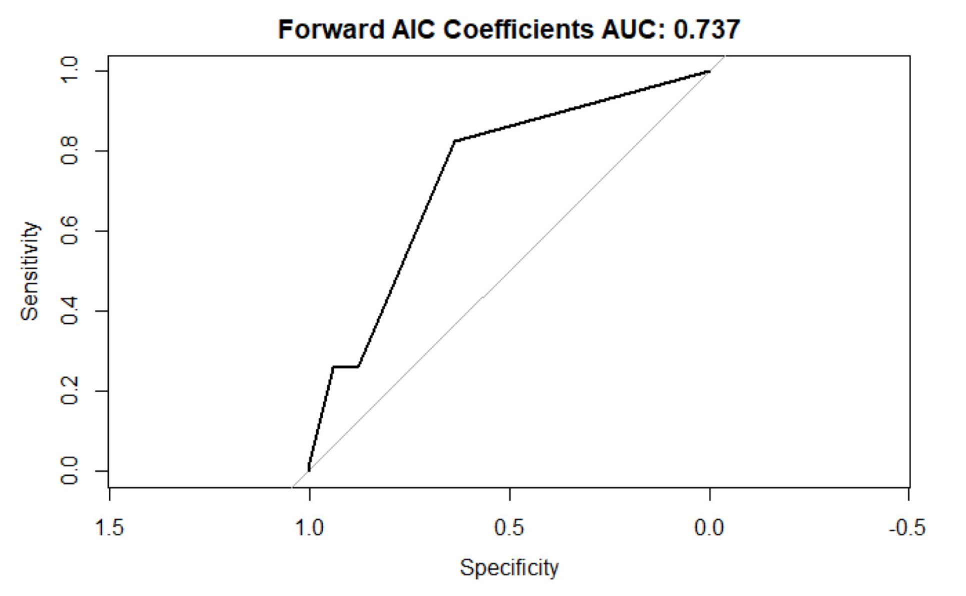
(Cumulative Logistic) PCA



	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	18	5	2	2	0	0	0	0	0
2	11	24	6	0	6	1	2	12	0
8	2	15	13	12	14	7	19	81	36
9	2	2	2	1	5	4	6	78	397

Application to Ordinal Regression

(Cumulative Logistic) Forward AIC



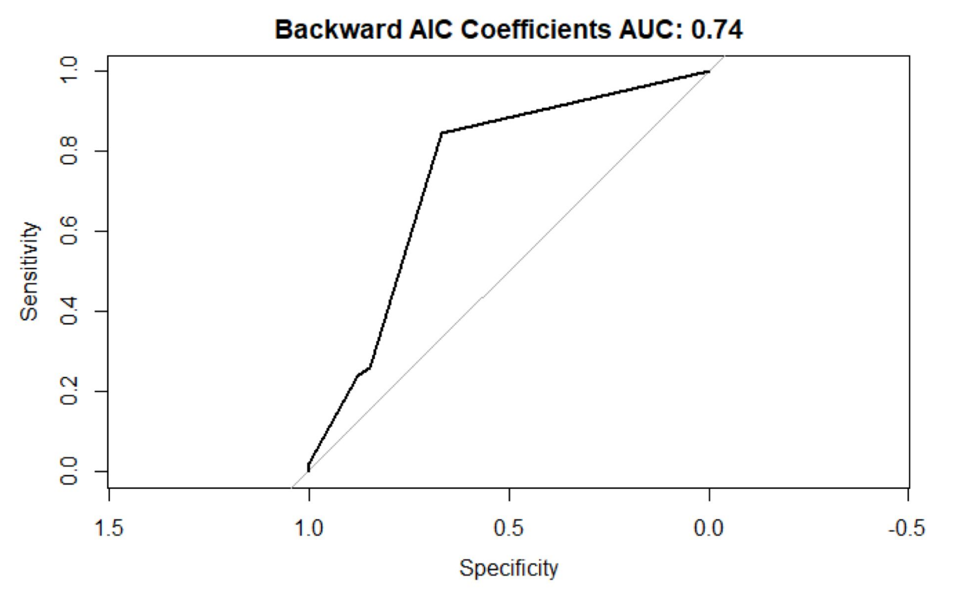
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	21	8	2	0	1	0	0	0	0
2	8	26	9	4	3	2	3	14	1
5	2	0	0	0	0	0	0	1	0
8	2	11	12	10	19	7	19	93	42
9	0	1	0	1	2	3	5	63	390

Variables Selected

"(Intercept)" "RHI825214" "EDU685213" "RHI225214" "HSG096213" "PST120214" "AGE295214" "SEX255214" "HSG495213" "INC910213" "POP715213" "SBO015207" "HSD310213" "AGE775214" "LFE305213" "INC110213" "POP645213" "RHI625214" "SBO215207" "SBO315207" "AGE135214" "SBO515207"

Application to Ordinal Regression

(Cumulative Logistic) Backward AIC



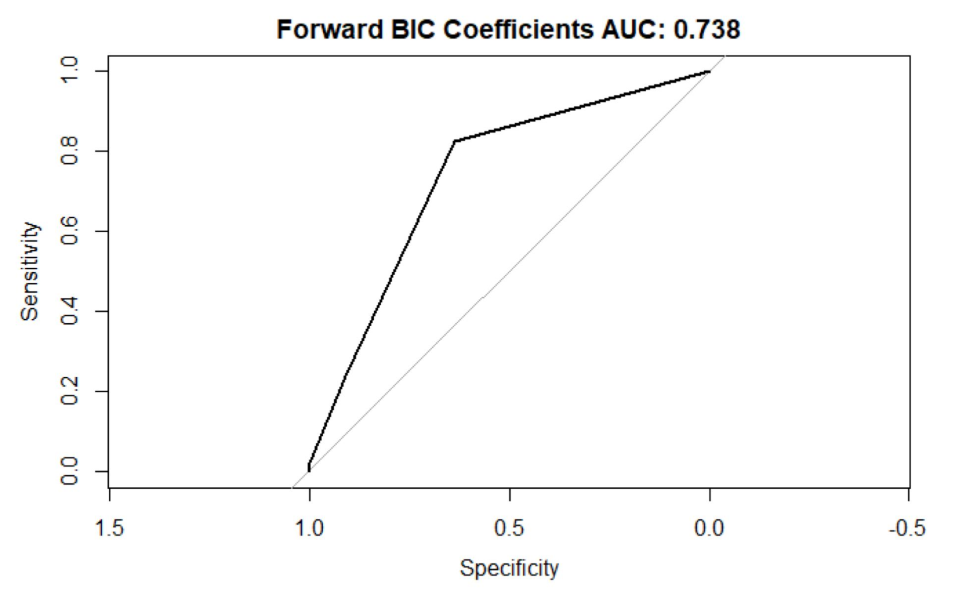
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	22	7	2	0	1	0	0	0	0
2	6	27	10	2	3	2	4	13	2
5	1	1	1	2	1	0	0	2	0
8	4	10	10	10	19	8	19	99	50
9	0	1	0	1	1	2	4	57	381

Variables Selected

"(Intercept)" "PST120214" "POP010210" "AGE135214" "AGE295214" "AGE775214" "RHI225214" "RHI425214" "RHI625214" "RHI725214" "RHI825214" "POP715213" "EDU635213" "EDU685213" "VET605213" "LFE305213" "HSG010214" "HSG096213" "HSG495213" "HSD410213" "HSD310213" "INC910213" "INC110213" "BZA010213" "NES010213" "SBO001207" "SBO315207" "SBO215207" "SBO015207" "RTN130207" "AFN120207”

Application to Ordinal Regression

(Cumulative Logistic) Forward BIC



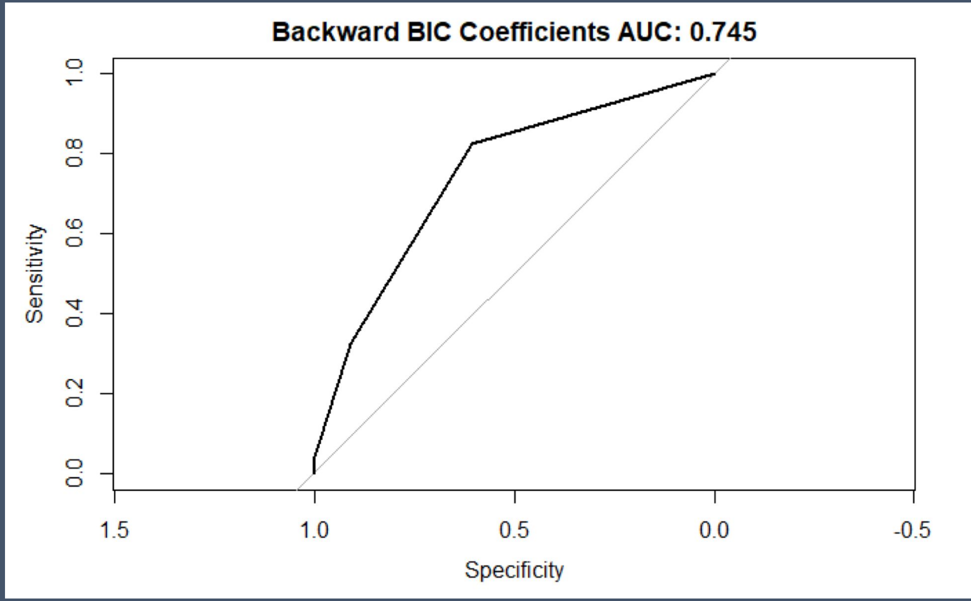
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	21	8	2	0	1	0	0	0	0
2	9	27	8	5	3	1	2	14	1
8	3	10	12	9	18	7	21	92	41
9	0	1	1	1	3	4	4	65	391

Variables Selected

"(Intercept)" "RHI825214" "EDU685213" "RHI225214" "HSG096213" "PST120214" "AGE295214" "SEX255214" "HSG495213" "INC910213" "POP715213" "SBO015207" "HSD310213" "AGE775214" "LFE305213"

Application to Ordinal Regression

(Cumulative Logistic) Backward BIC



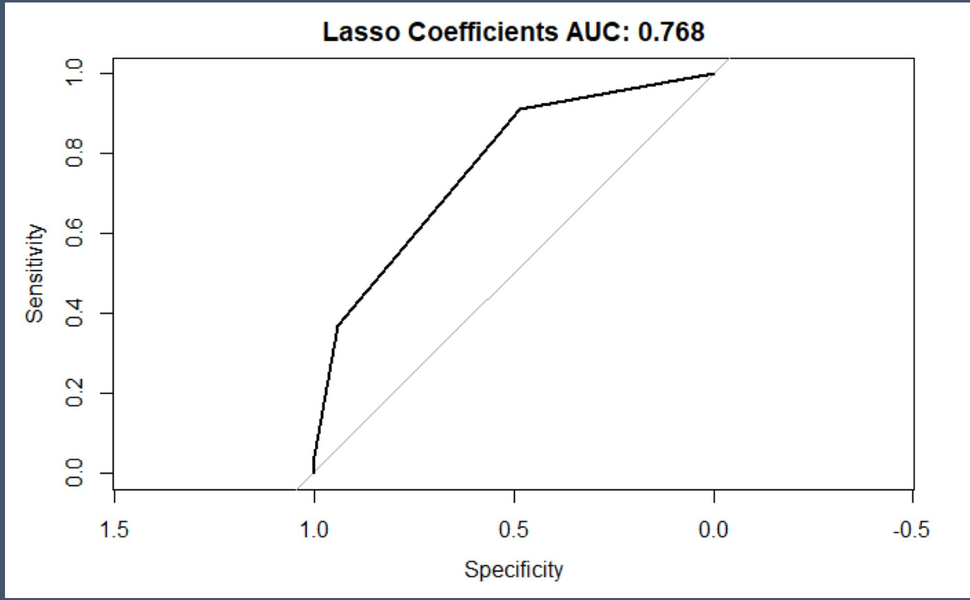
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	20	8	2	0	1	0	0	0	0
2	10	23	9	4	2	1	3	11	2
8	3	13	11	10	20	7	19	96	42
9	0	2	1	1	2	4	5	64	389

Variables Selected

"(Intercept)" "AGE295214" "RHI825214" "POP715213" "EDU685213" "LFE305213" "HSG010214" "HSG096213" "HSG495213" "HSD410213" "HSD310213" "INC910213" "SBO015207"

Application to Ordinal Regression

(Cumulative Logistic) Lasso



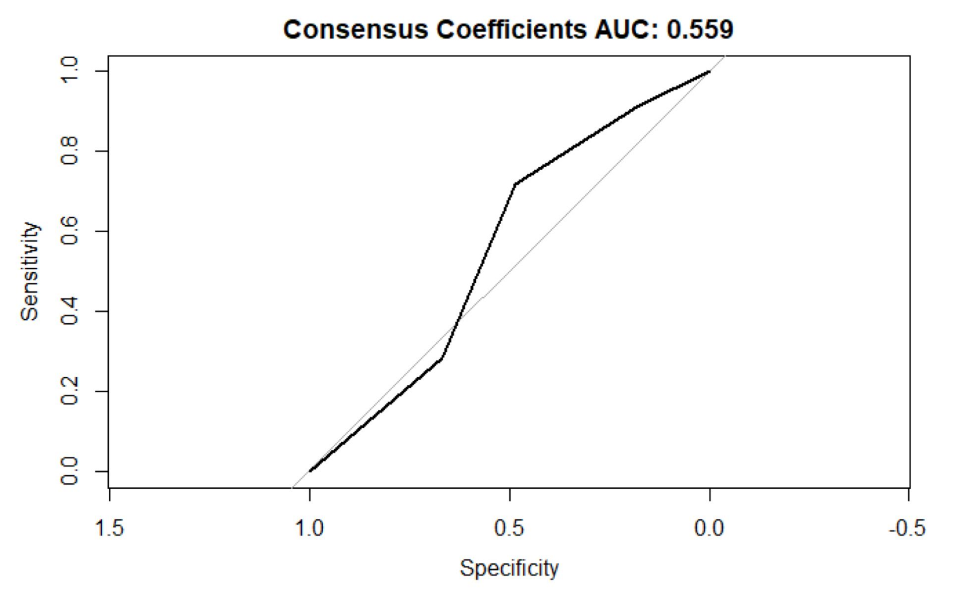
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	16	4	2	0	1	0	0	1	0
2	15	25	10	4	3	2	3	7	0
8	2	15	10	10	17	7	17	87	43
9	0	2	1	1	4	3	7	76	390

Variables Selected

"PST045214" "POP010210" "AGE775214" "SEX255214" "RHI125214" "RHI225214" "RHI325214" "RHI725214" "POP715213" "POP645213" "POP815213" "EDU685213""VET605213" "HSG010214" "HSG495213" "HSD410213" "INC910213" "PVY020213" "BZA010213" "SBO115207" "SBO515207" "SBO015207" "MAN450207" "WTN220207" "BPS030214"

Application to Ordinal Regression

(Cumulative Logistic) Consensus



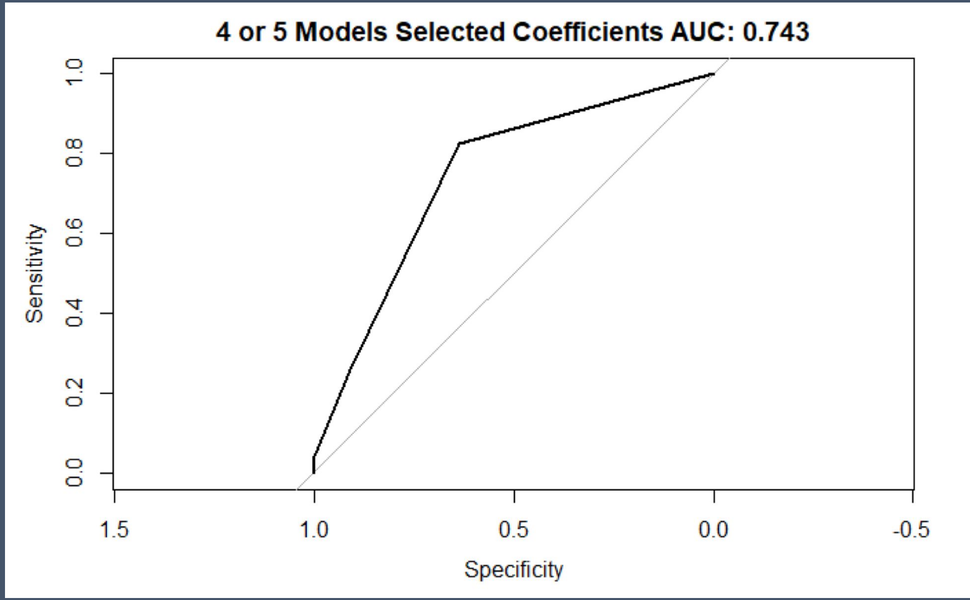
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	6	4	3	0	1	0	0	0	0
2	10	9	0	3	1	0	0	6	1
8	6	20	11	6	9	3	7	41	22
9	11	13	9	6	14	9	20	124	410

Variables Selected

"EDU685213" "HSG495213" "INC910213" "POP715213" "SBO015207"

Application to Ordinal Regression

(Cumulative Logistic) 4 out of 5 models



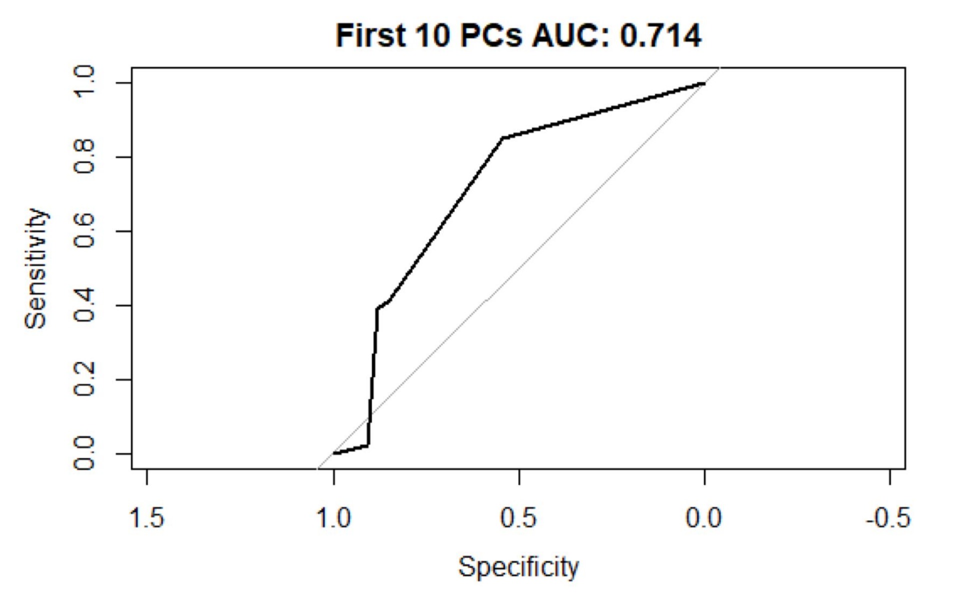
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	21	8	2	0	1	0	0	0	0
2	9	26	9	5	3	1	2	13	2
8	3	10	11	9	19	7	20	95	37
9	0	2	1	1	2	4	5	63	394

Variables Selected

"EDU685213" "HSG495213" "INC910213" "POP715213" "SBO015207" "AGE295214" "AGE775214" "HSD310213" "HSG096213" "LFE305213" "RHI225214" "RHI825214"

Application to Multinomial Regression

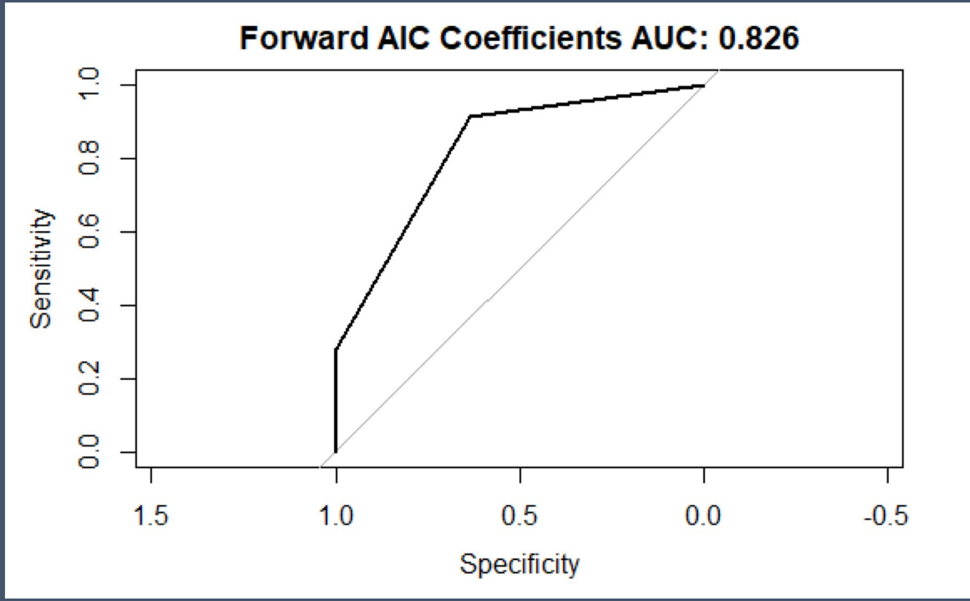
(Baseline Categorical Logit Model) PCA



	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	18	7	1	0	0	1	0	1	0
2	10	20	8	2	7	0	2	10	0
5	1	1	0	0	1	0	1	2	1
8	1	17	13	12	14	7	17	83	30
9	3	1	1	1	3	4	7	75	402

Application to Multinomial Regression

(Baseline Categorical Logit Model) Forward AIC



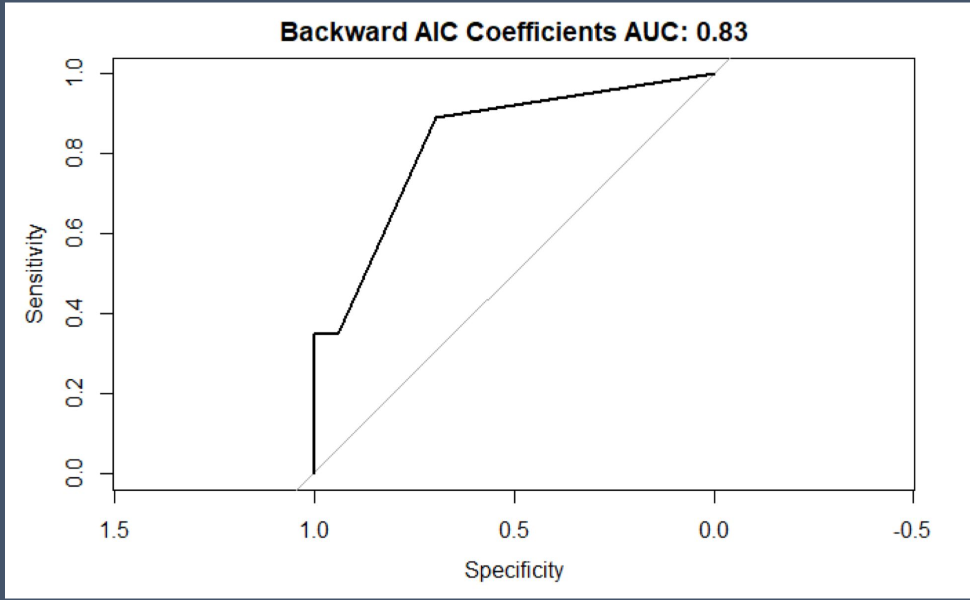
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	21	4	3	0	1	2	0	0	0
2	12	29	7	3	3	0	3	13	2
3	0	0	1	0	0	0	0	1	0
4	0	0	0	1	0	0	0	0	0
5	0	4	1	3	3	1	1	5	2
6	0	0	1	0	0	0	0	0	0
8	0	8	9	6	16	4	19	88	30
9	0	1	1	2	2	5	4	64	399

Variables Selected

"(Intercept)" "RHI825214" "EDU685213" "RHI225214" "HSG096213" "PST120214" "AGE295214" "SEX255214" "HSG495213"
"INC910213" "POP715213" "SBO015207" "HSD310213" "AGE775214" "LFE305213" "INC110213" "POP645213"
"RHI625214" "SBO215207" "SBO315207" "AGE135214" "SBO515207"

Application to Multinomial Regression

(Baseline Categorical Logit Model) Backward AIC



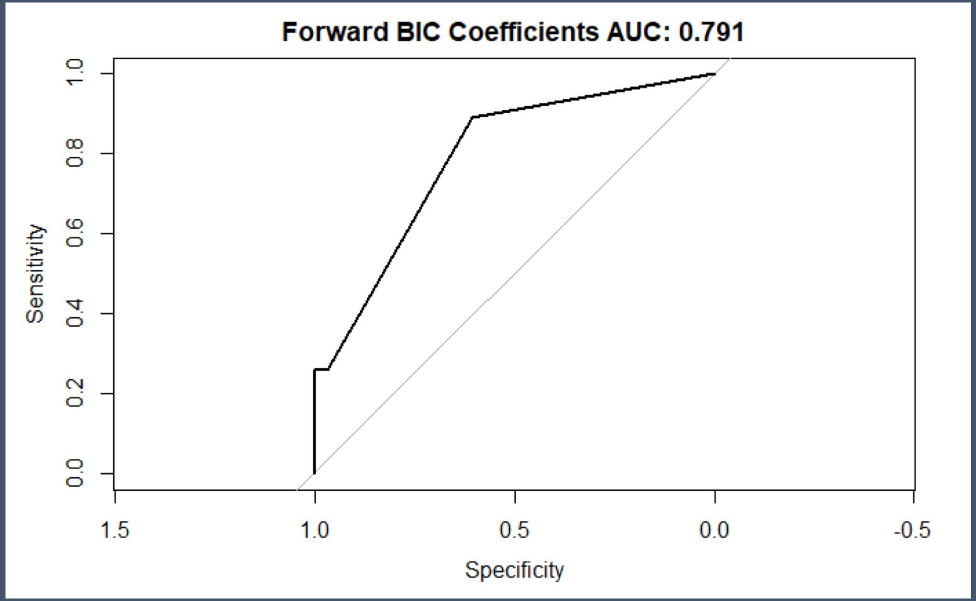
Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
23	5	0	0	1	1	0	0	0
8	25	7	3	3	1	5	11	2
2	0	1	1	0	0	0	0	0
0	0	2	1	0	0	0	0	0
0	5	4	1	5	1	1	8	0
0	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
0	9	8	8	14	7	18	86	42
0	1	0	1	2	2	3	66	389

Variables Selected

"(Intercept)" "PST120214" "POP010210" "AGE135214" "AGE295214" "AGE775214" "RHI225214" "RHI425214" "RHI625214" "RHI725214" "RHI825214" "POP715213" "EDU635213" "EDU685213" "VET605213" "LFE305213" "HSG010214" "HSG096213" "HSG495213" "HSD410213" "HSD310213" "INC910213" "INC110213" "BZA010213" "NES010213" "SBO001207" "SBO315207" "SBO215207" "SBO015207" "RTN130207" "AFN120207"

Application to Multinomial Regression

(Baseline Categorical Logit Model) Forward BIC



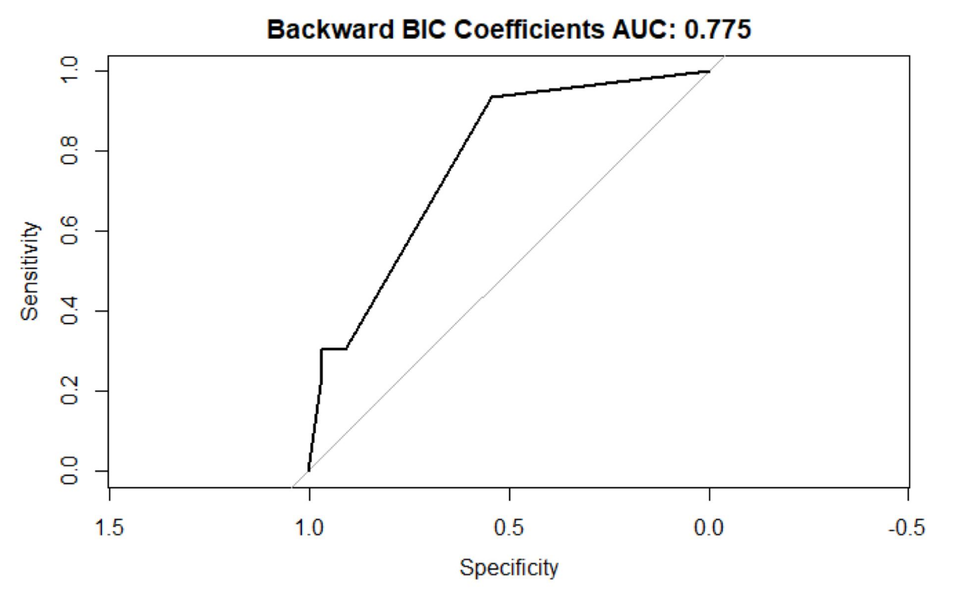
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	20	5	2	0	1	0	0	0	0
2	12	29	8	5	3	2	3	13	2
3	0	0	1	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0
5	0	4	1	3	1	0	1	2	0
8	0	7	11	5	18	5	19	89	36
9	0	1	0	2	2	5	4	67	395

Variables Selected

"(Intercept)" "RHI825214" "EDU685213" "RHI225214" "HSG096213" "PST120214" "AGE295214" "SEX255214" "HSG495213" "INC910213" "POP715213" "SBO015207" "HSD310213" "AGE775214" "LFE305213"

Application to Multinomial Regression

(Baseline Categorical Logit Model) Backward BIC



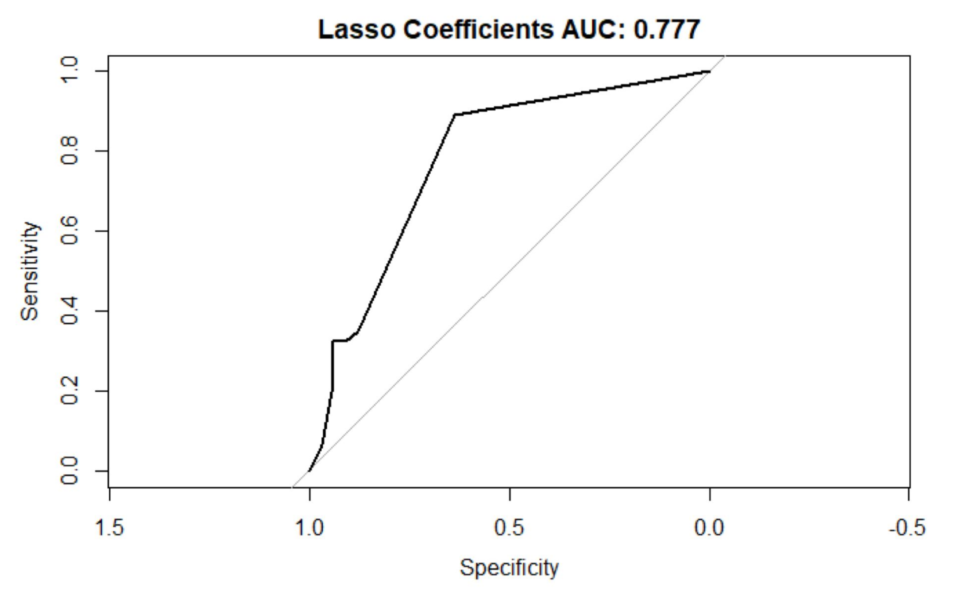
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	18	3	2	1	1	0	0	1	0
2	12	29	10	3	4	1	4	11	2
3	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0
5	0	3	2	3	1	1	3	3	0
8	1	10	9	6	16	5	15	87	35
9	0	1	0	2	3	5	5	69	396

Variables Selected

"(Intercept)" "AGE295214" "RHI825214" "POP715213" "EDU685213" "LFE305213" "HSG010214" "HSG096213"
"HSG495213" "HSD410213" "HSD310213" "INC910213" "SBO015207"

Application to Multinomial Regression

(Baseline Categorical Logit Model) Lasso



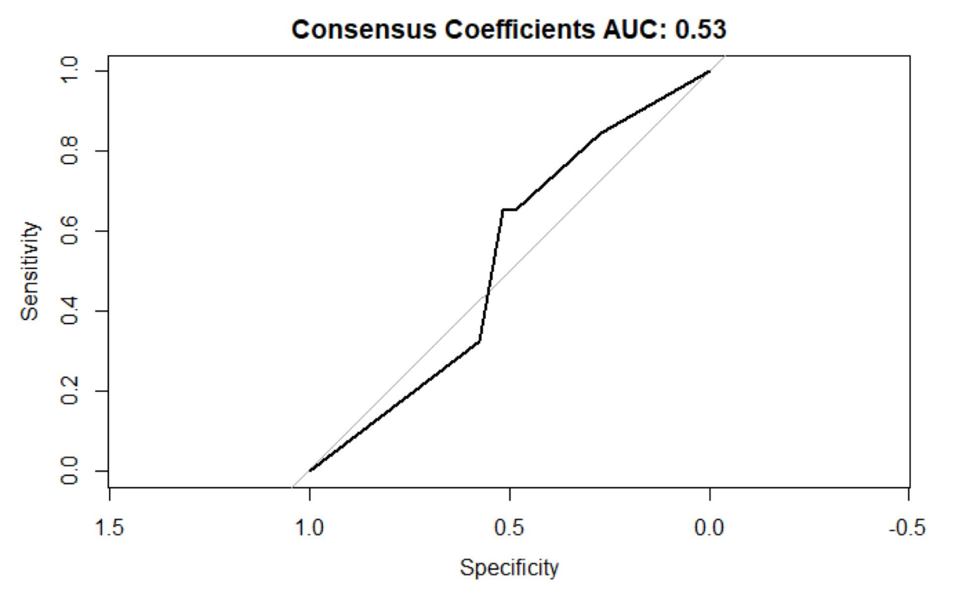
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	21	5	1	0	1	1	0	0	0
2	8	25	8	3	4	1	2	7	0
3	1	1	0	0	0	0	1	0	0
4	1	0	2	1	0	0	0	0	0
5	0	4	2	2	2	1	2	6	2
7	0	1	0	1	0	0	0	0	0
8	1	7	9	7	14	6	15	74	30
9	1	3	1	1	4	3	7	84	401

Variables Selected

"PST045214" "POP010210" "AGE775214" "SEX255214" "RHI125214" "RHI225214" "RHI325214" "RHI725214" "POP715213" "POP645213" "POP815213" "EDU685213""VET605213" "HSG010214" "HSG495213" "HSD410213" "INC910213" "PVY020213" "BZA010213" "SBO115207" "SBO515207" "SBO015207" "MAN450207" "WTN220207" "BPS030214"

Application to Multinomial Regression

(Baseline Categorical Logit Model) Consensus



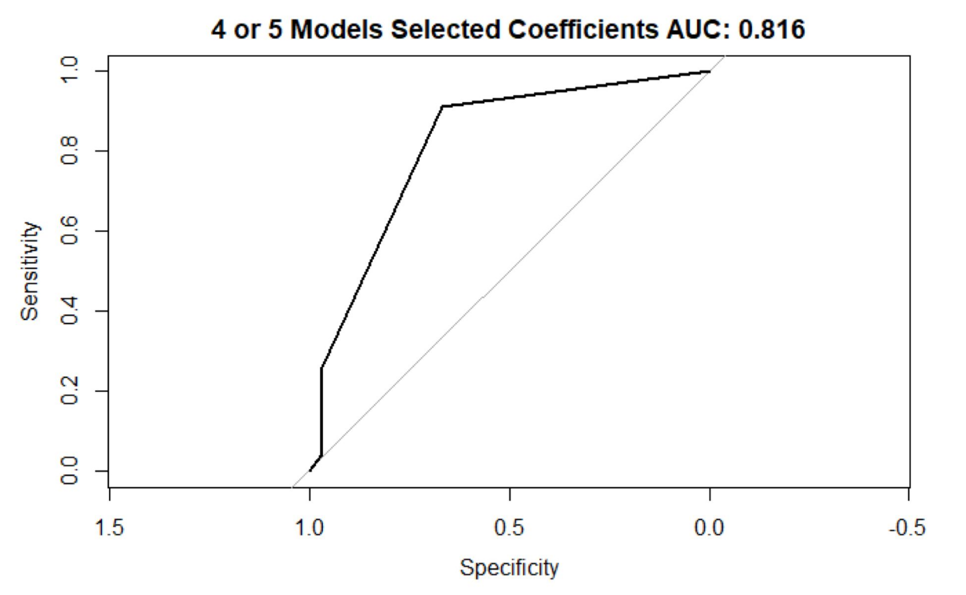
	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	9	7	2	0	1	0	0	1	0
2	7	9	2	3	2	0	0	4	2
5	1	0	0	0	1	0	0	1	0
8	2	15	9	6	7	3	6	30	12
9	14	15	10	6	14	9	21	135	419

Variables Selected

"EDU685213" "HSG495213" "INC910213" "POP715213" "SBO015207"

Application to Multinomial Regression

(Baseline Categorical Logit Model) 4 out of 5 models



	Solid Democratic	Likely Democratic	Lean Democratic	Tilt Democratic	Toss-Up	Tilt GOP	Lean GOP	Likely GOP	Solid GOP
1	22	4	2	1	1	0	0	1	0
2	10	30	9	3	3	1	3	11	2
3	0	0	1	0	0	0	0	0	0
5	0	3	1	2	1	0	0	3	1
8	0	7	10	7	17	6	19	91	32
9	1	2	0	2	3	5	5	65	398

Variables Selected

"EDU685213" "HSG495213" "INC910213" "POP715213" "SBO015207" "AGE295214" "AGE775214" "HSD310213" "HSG096213" "LFE305213" "RHI225214" "RHI825214"

Conclusion

Conventional Wisdom

Conventional Wisdom about the 2016 election says that Trump won because of white non-college educated voters. Our top models seem to confirm this.

Principal Components

Using the first 10 principal components performs arguably just as well as other top variable selection methods for logistic regression but performs worse for the multiple category methods. Likely due to the variance lost.

Variable Selection

For this data there is no “best” variable selection method. Seemingly the best way to select variables is to use multiple methods and get the most selected variables from each of those.

Ordinal vs Multinomial

Though the election results can be thought of as ordered, the multinomial regression models were able to more accurately predict the political lean of a county.

QUESTIONS TIME!

Have a question? Ask NOW!