

AdmitSee Analytics

Mike Yung

Context



Part I: The Model

- Can we build a model that predicts a student's chances* of being admitted into college?



Part II: The Essay

- What insights can we glean from the Common App essay?

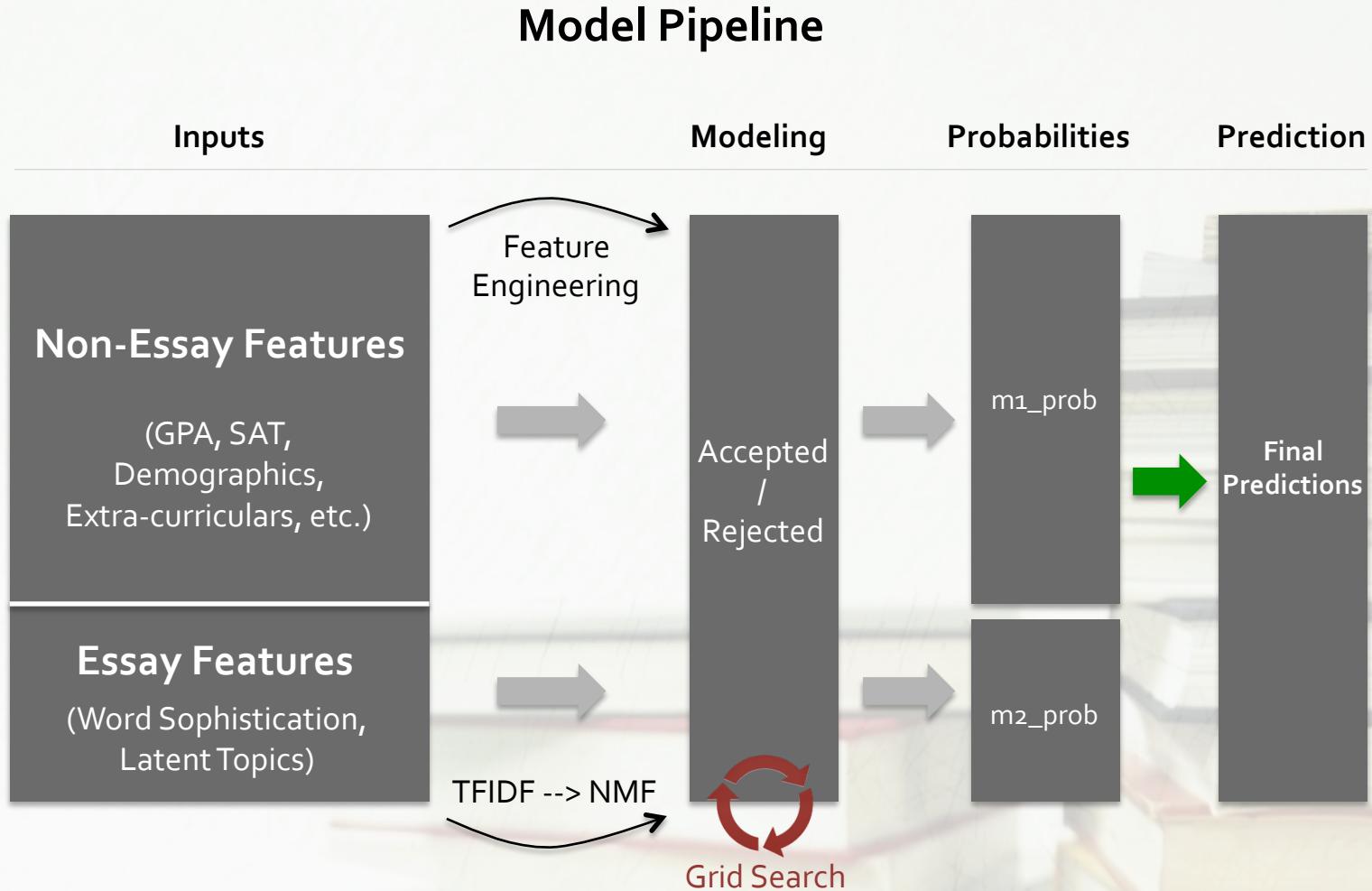
*There are some resources that 'calculate' your chances based on your GPA, SAT, and demographics, but none (at least publicly available) that take into account detailed factors such as specific extracurriculars, academic trajectory, the Common App essay etc.

Part 1: The Model

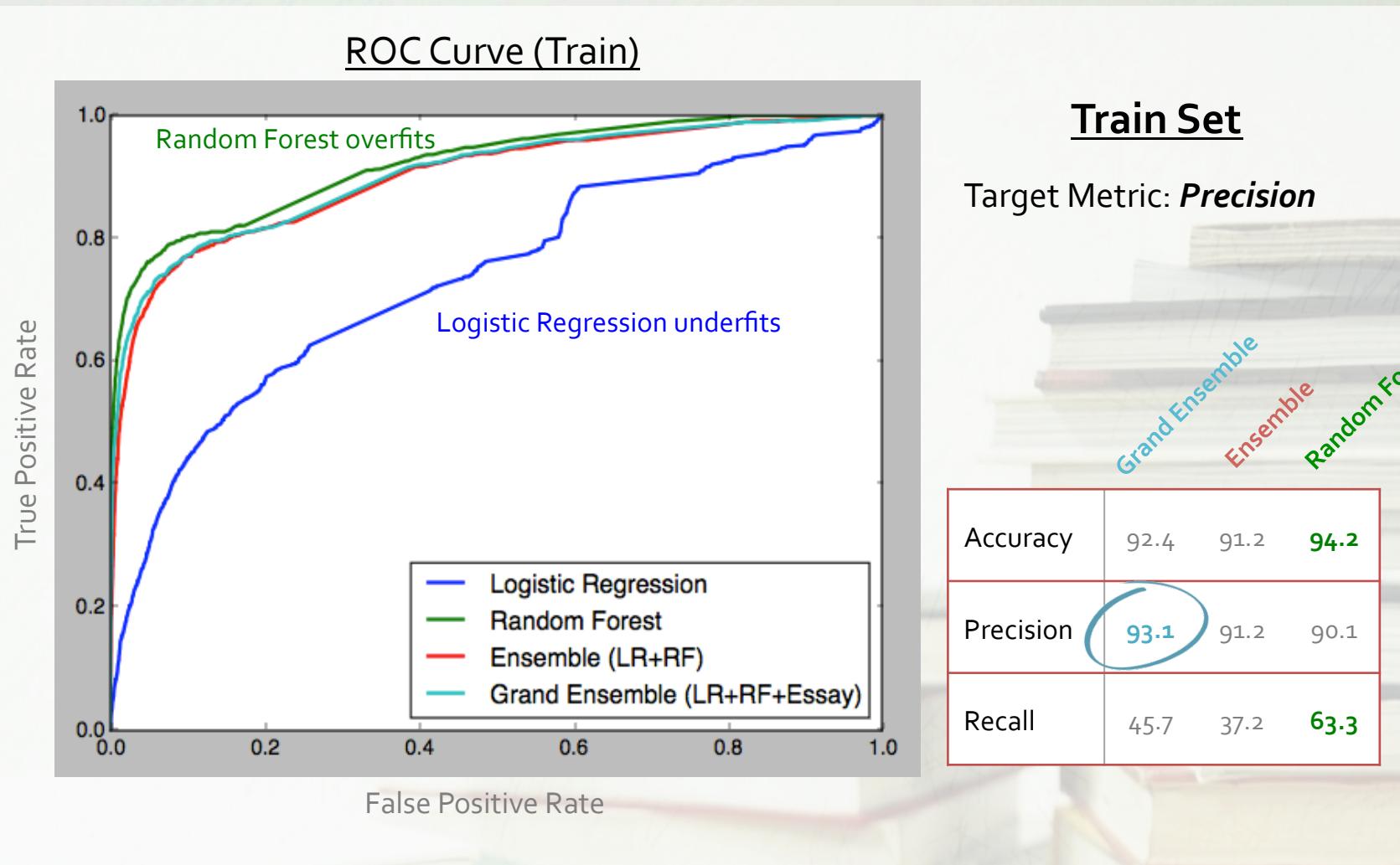
Can we build a model that predicts a student's chances of being admitted into college?



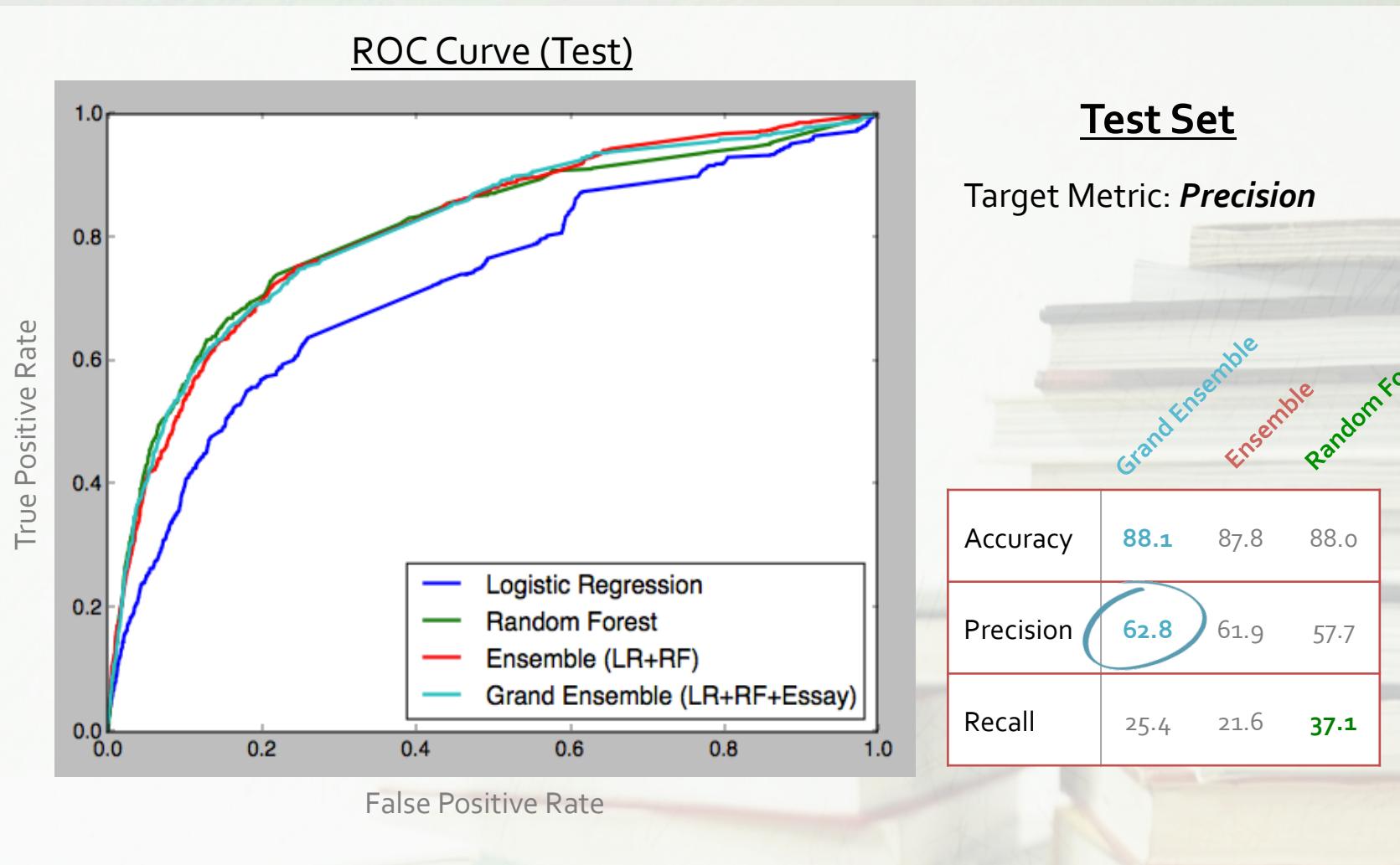
Building an Ensemble Model



Evaluating the Model



Evaluating the Model



Interpreting the Model

Logistic Regression Model

Variable	$e^{\text{Coefficient}}$
Leader	2.26
Student Gov	1.69
Varsity Sport	1.58
Sports Captain	1.29
Award	1.22
Community Service	1.21
SAT Score	1.0003
SAT Times Taken	0.39

How to Interpret

- If you aren't already in a **leadership** position, taking one will **more than double** your odds of being admitted
- If you gain an **additional 100 points** on your **SAT**, you can **increase** your odds of being admitted by **3%**
- If you **take the SATs one more time**, you can **reduce** your odds of being admitted by **61%**

*Note: this is only a subset of all variables used in the model

Part 2: The Essay

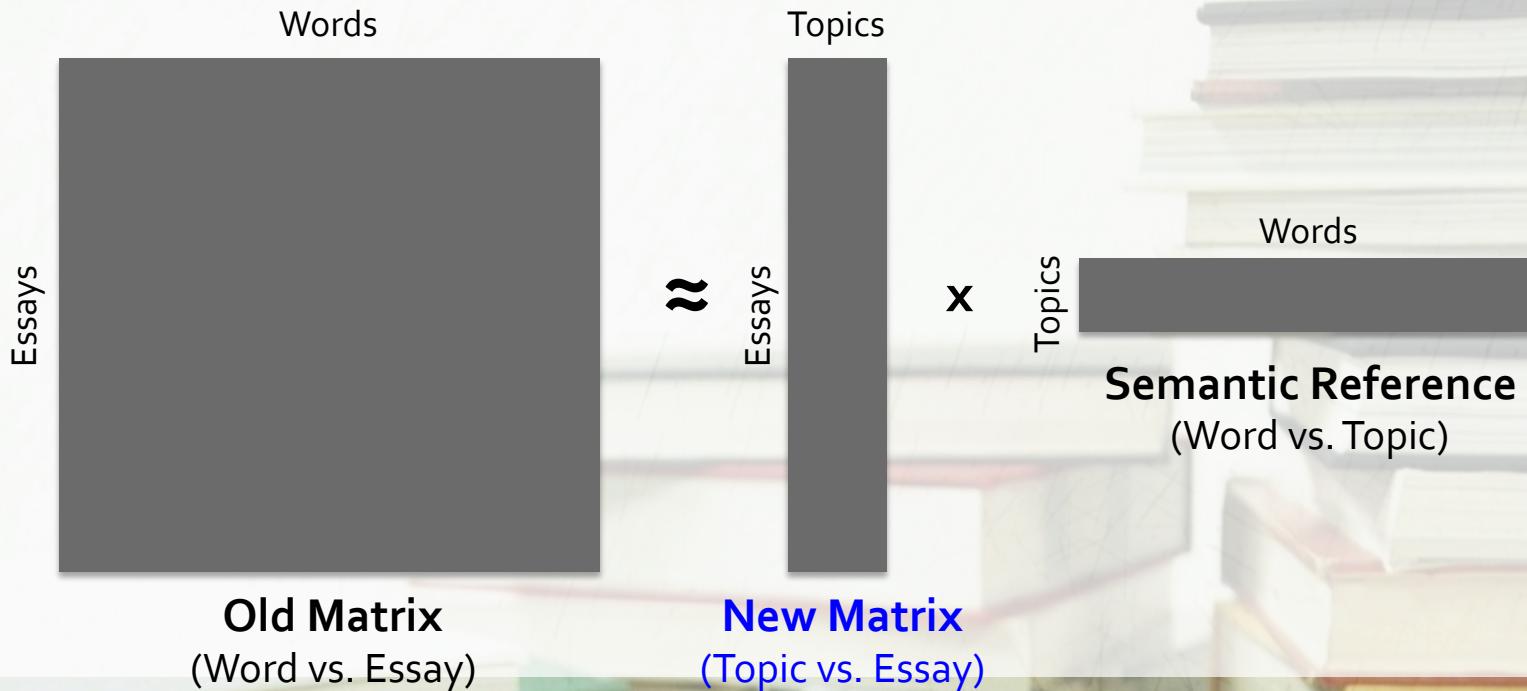
What insights can we glean from the Common App essay?



What's in an Essay?



Non-Negative Matrix Factorization



What's in an Essay?



Non-Negative Matrix Factorization

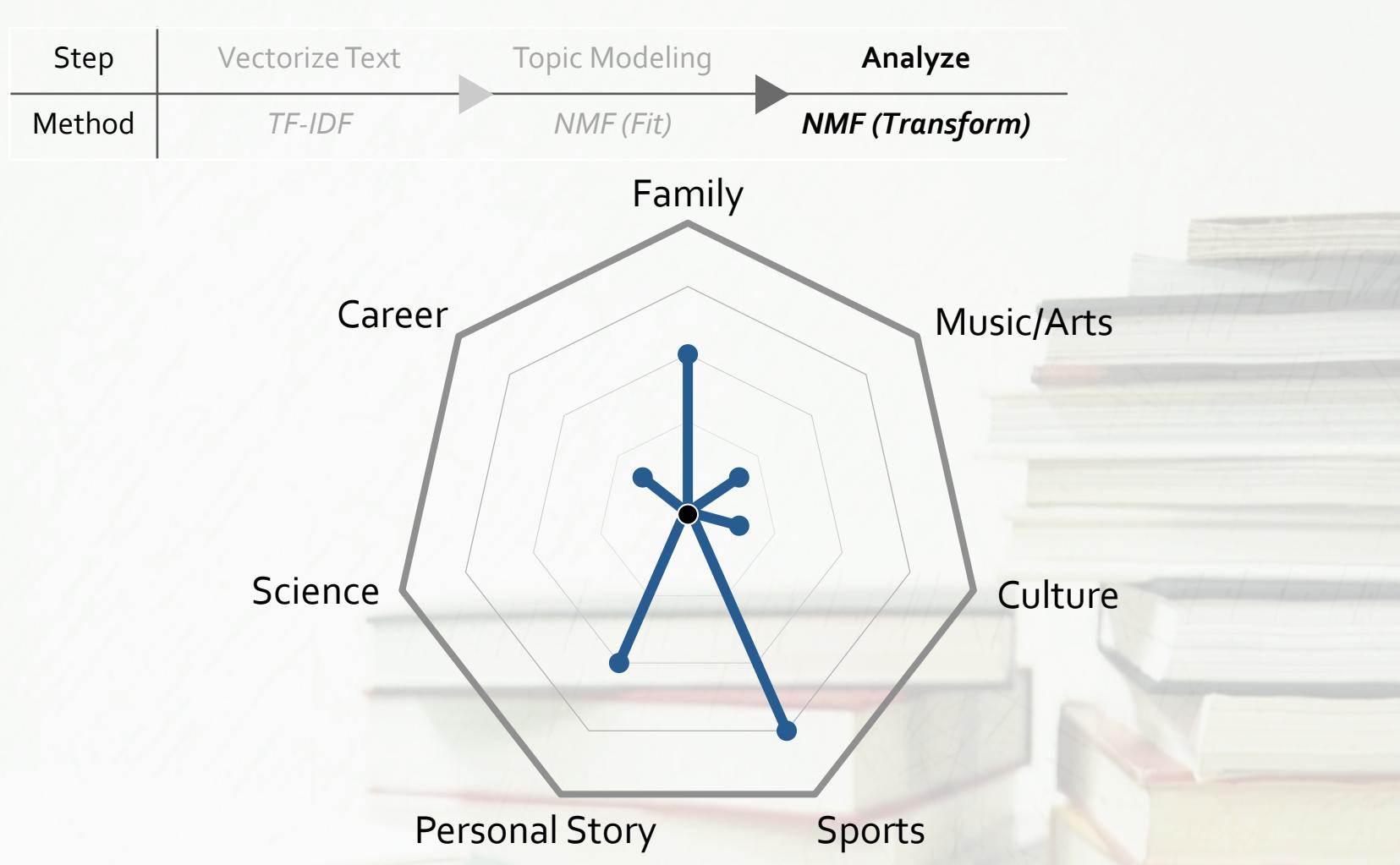
Words

Topics

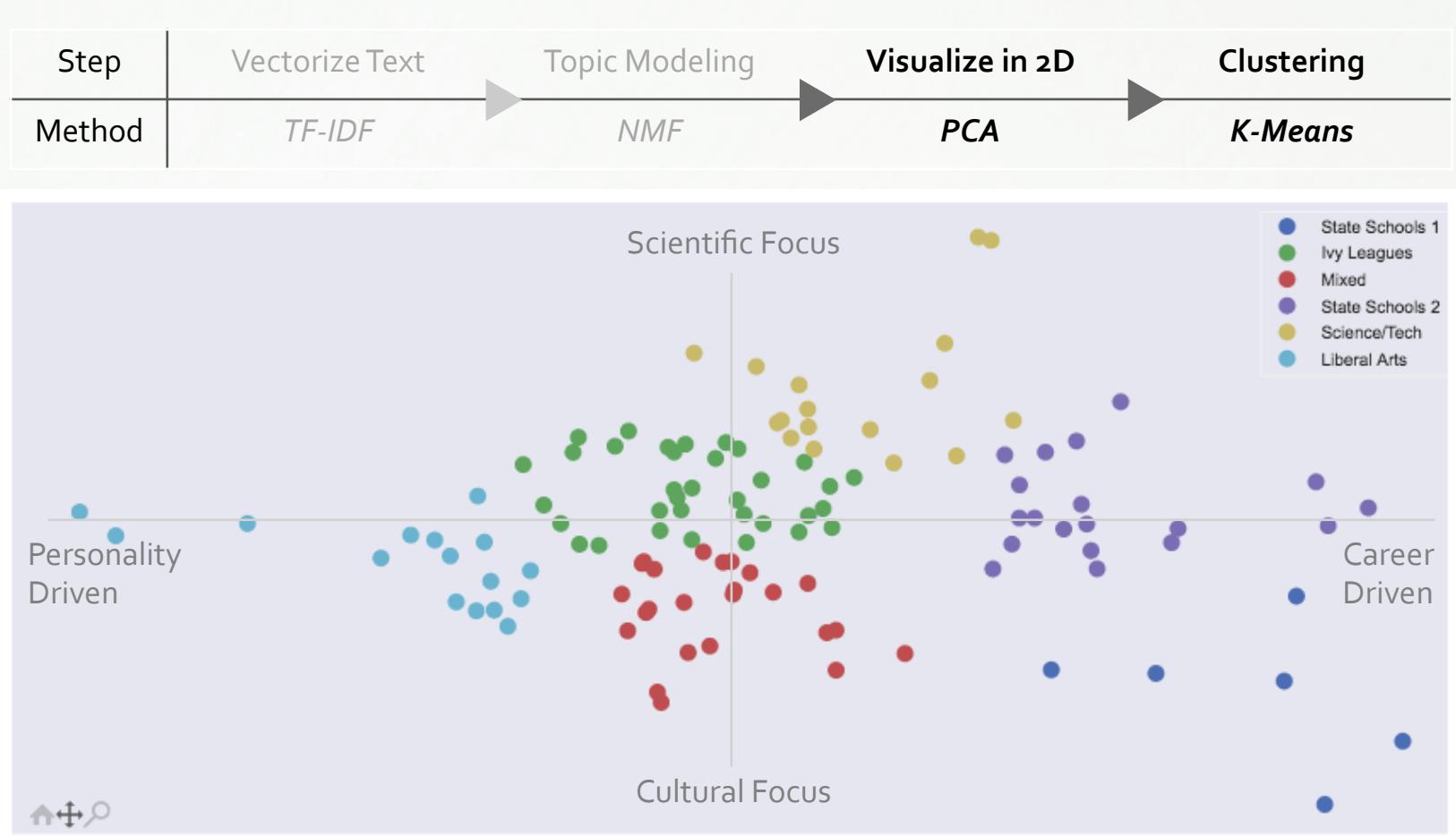
mother	father	family	parent	sister	...	Family
music	play	piano	perform	theater	...	Music/Arts
culture	world	language	travel	american	...	Culture
team	game	coach	player	season	...	Sports
research	science	computer	technology	math	...	Personal/Story
work	education	career	success	community	...	Science
						Career

Semantic Reference
(Word vs. Topic)

Topic Distribution of a Sample Essay



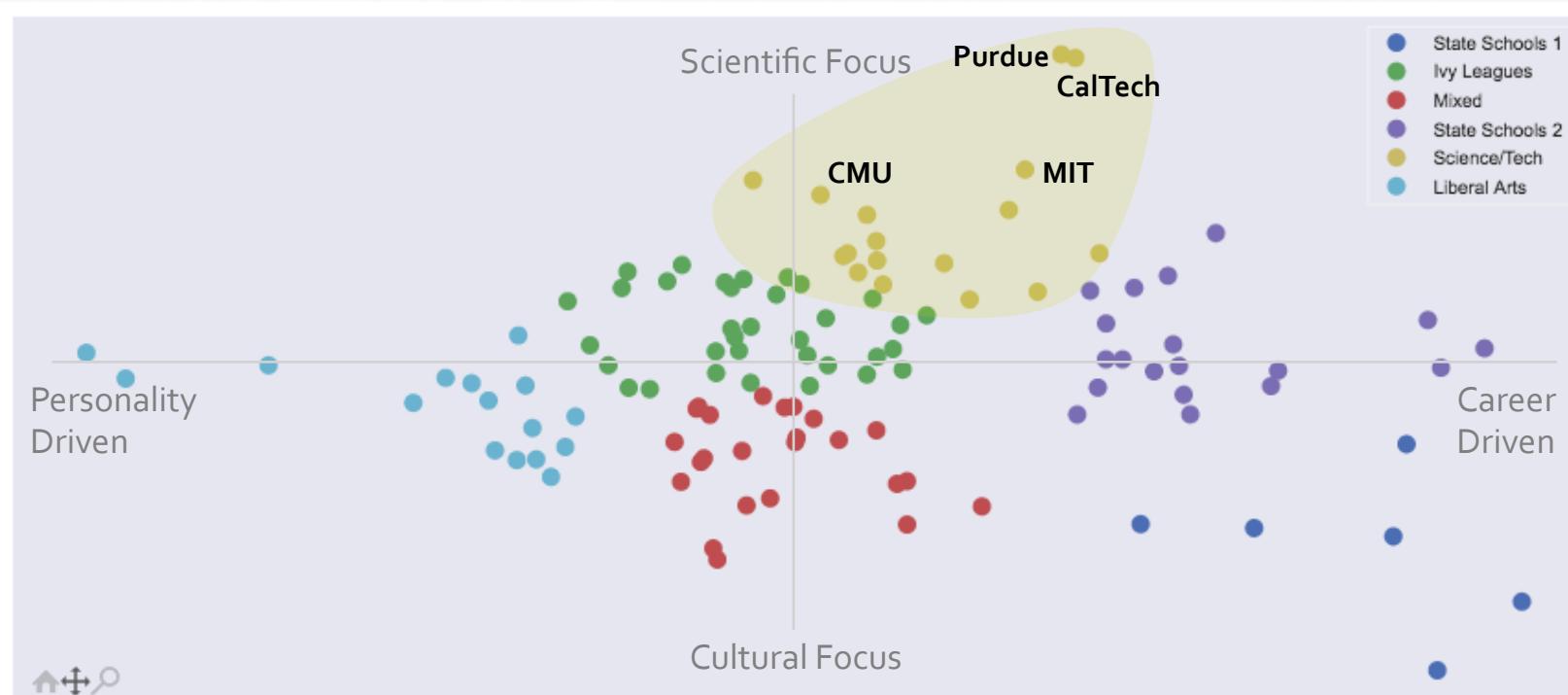
A 2-D Representation of College Essays



*Each point represents a college/university

A 2-D Representation of College Essays

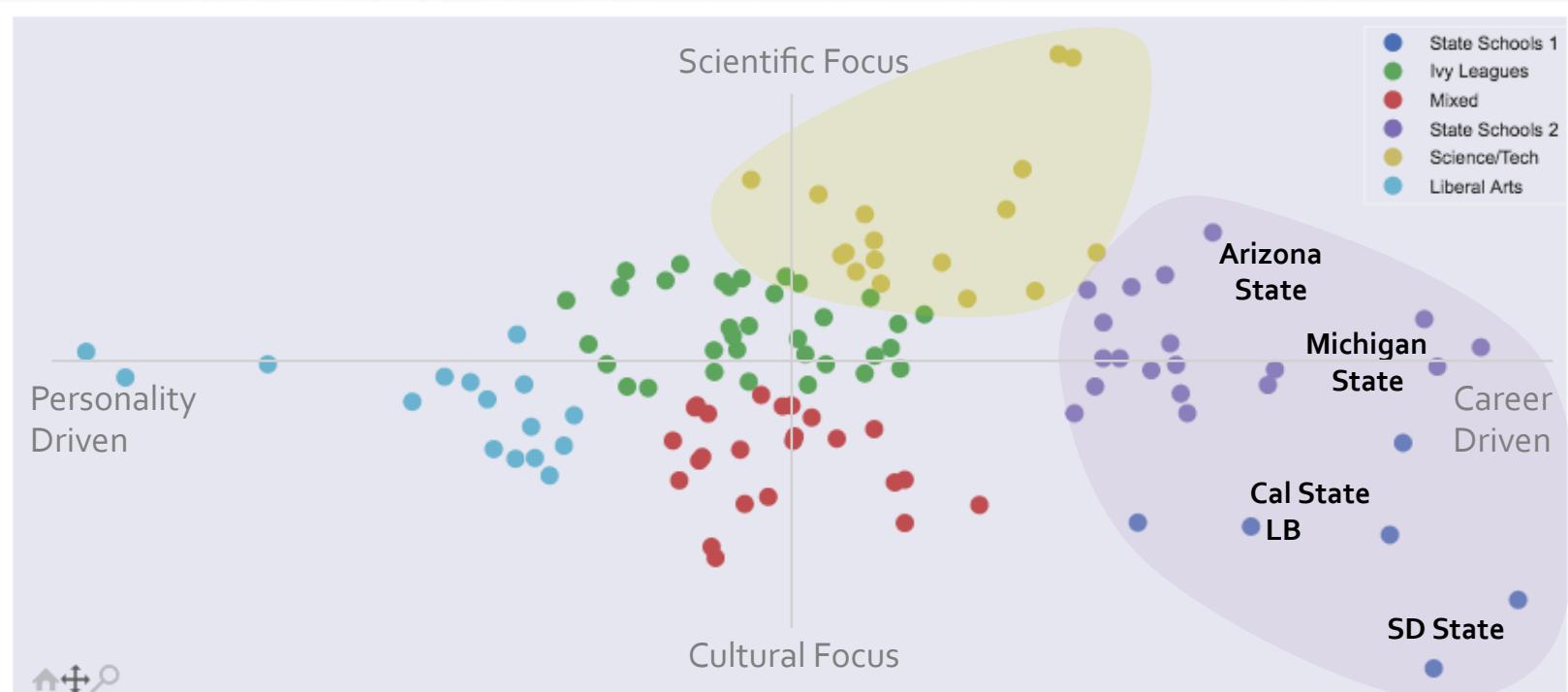
Cluster #1: Colleges with STEM Focus



*Each point represents a college/university

A 2-D Representation of College Essays

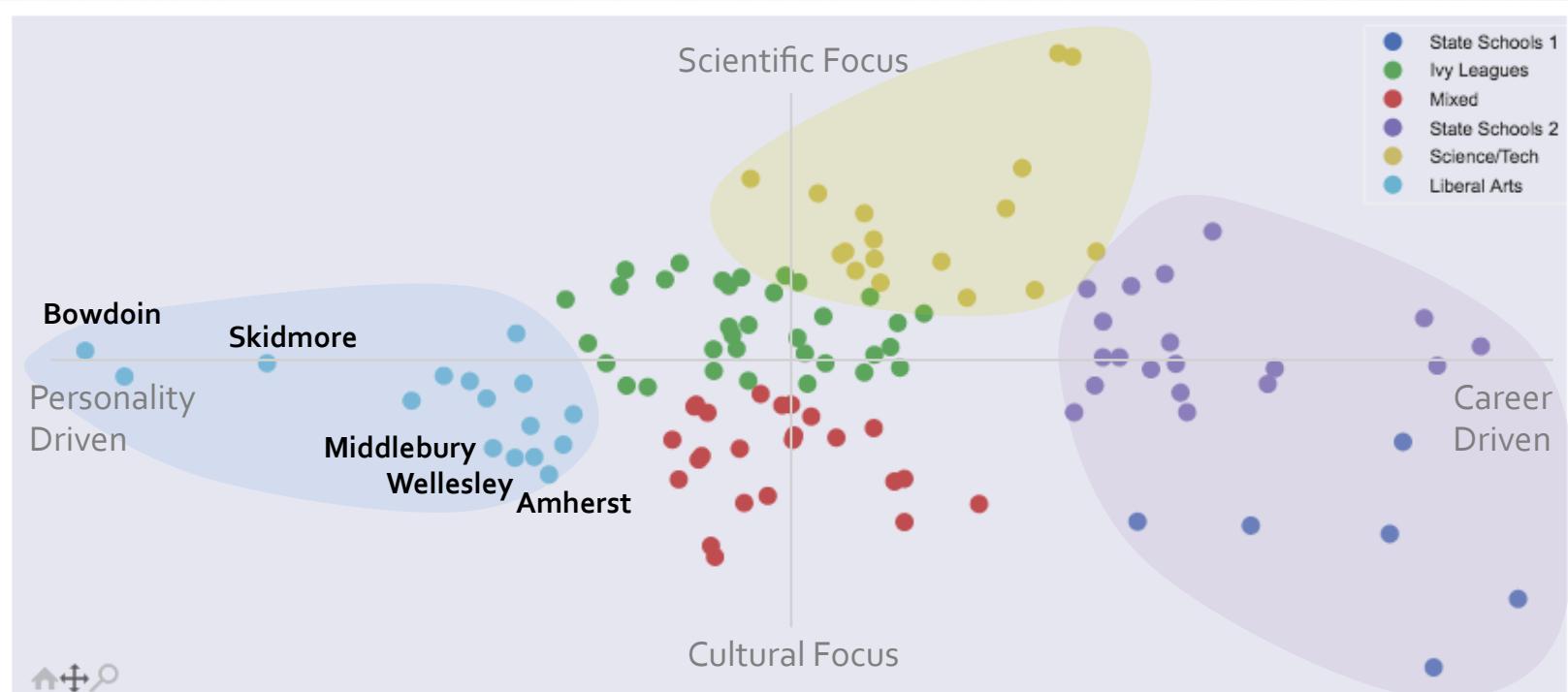
Cluster #2: State Schools, Large Institutions



*Each point represents a college/university

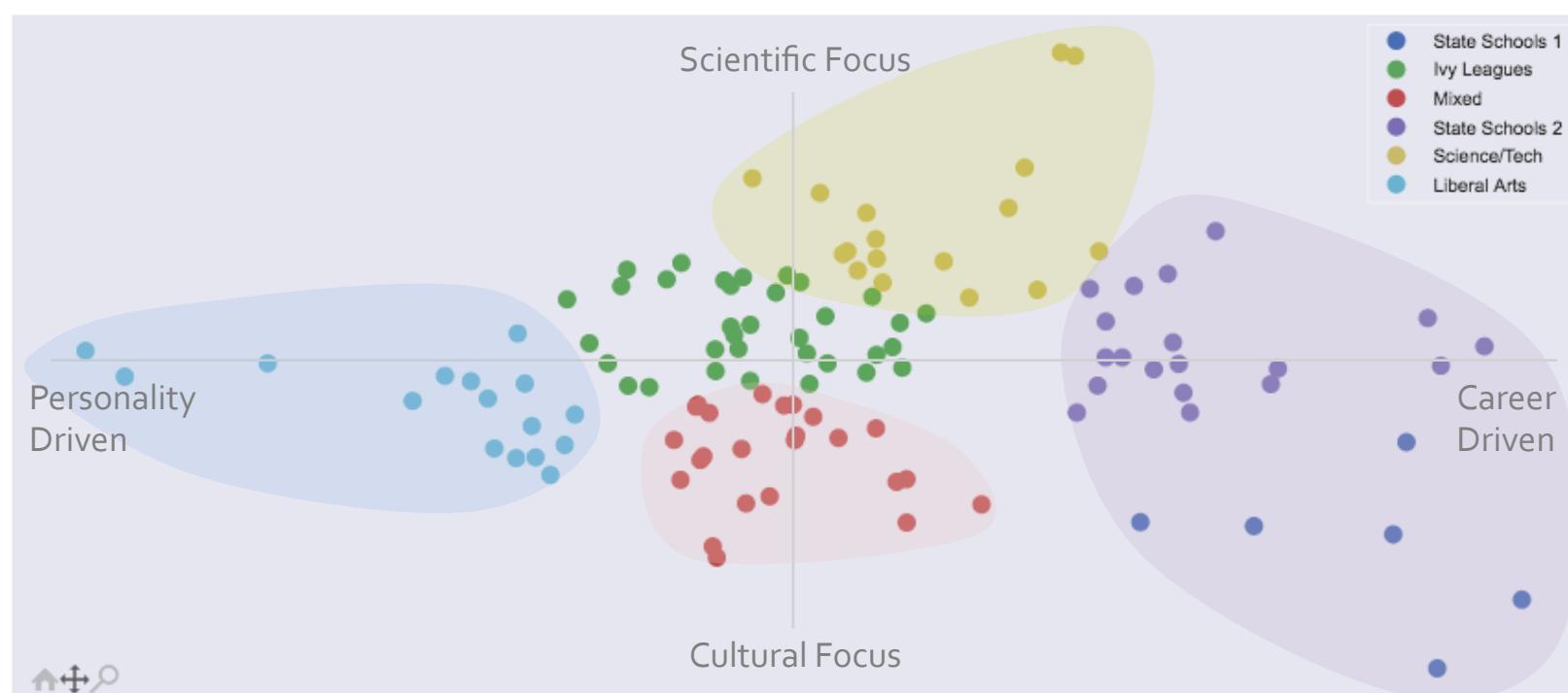
A 2-D Representation of College Essays

Cluster #3: Liberal Arts Schools



A 2-D Representation of College Essays

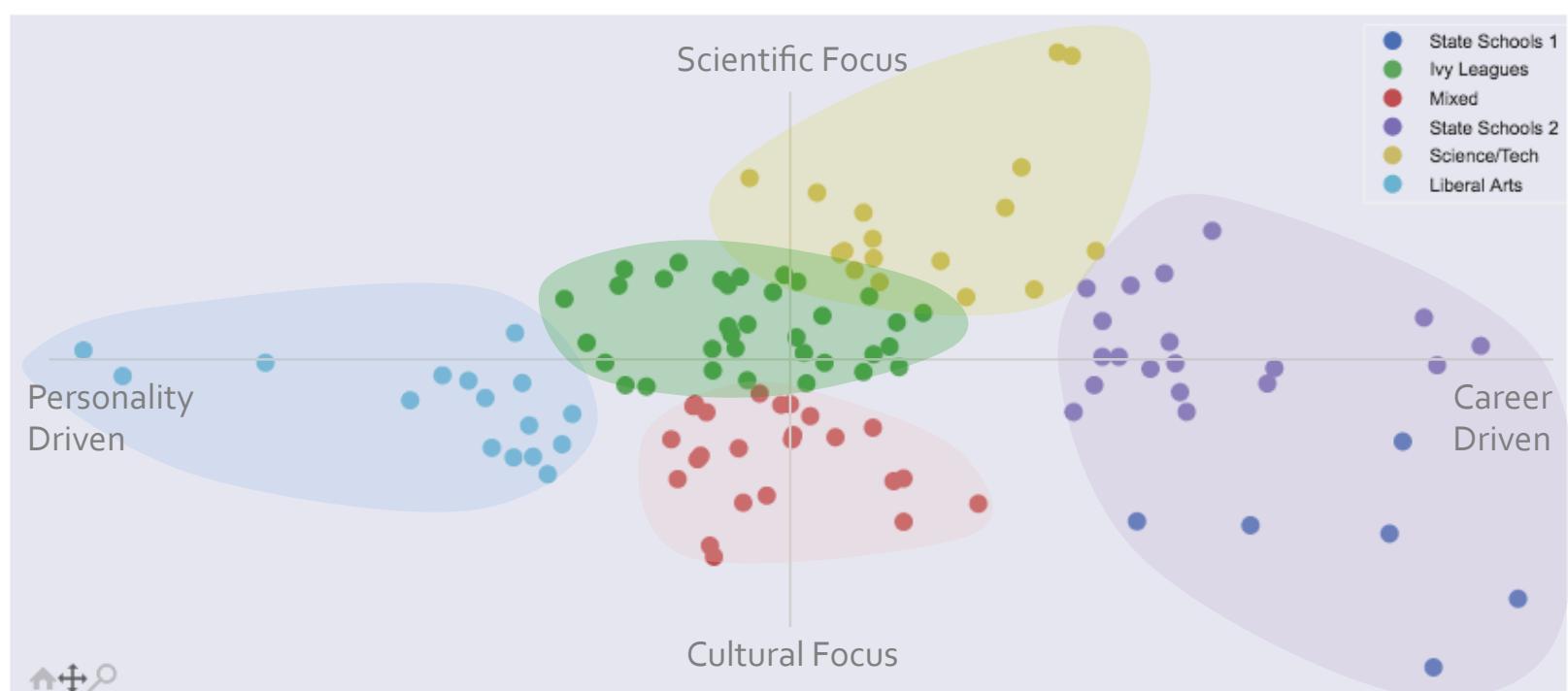
Cluster #4: A Mixed Bag



*Each point represents a college/university

A 2-D Representation of College Essays

Cluster #5: Ivy League



*Each point represents a college/university

Final Thoughts

- Similar Essay Recommender
 - Currently being deployed by AdmitSee into its product
 - Demo available post-presentation
- Limitation of data
 - Only enough to model ‘top school’ admittance
 - Can predict on a school-level with more data
- Further NLP exploration
 - N-gram (bigrams, trigrams, etc.)
 - PoS tagging
 - Latent Dirichlet Allocation (LDA)