

# When Data Science Blows Your Mind

or, how to ask questions that make others nod their heads knowingly and pretend they had the same questions



## Presenters

# Tim

- + SVP - Modeling
- + Former AP Statistics and math teacher
- + Built a “sabermetric” model to set the starting lineup while coaching JV Baseball
- + The Analytical and Enthusiastic Emcee



## Presenters

# Mike

- + SVP - Modeling & Client Strategy
- + Former AVP - Financial Aid at Loyola Marymount
- + Chose an undergraduate program because he was able to opt out of a math requirement
- + The Influential and Resilient Energizer



# + today's agenda

---

1. You Are Already Doing Data Science
2. The Problem
3. Data Science to the Rescue
4. Questions You Should Be Asking

Data science can be intimidating.

Data science is about extracting signal from noise.

You are doing data science all the time.

## Too Sexy?



October 2012 Issue

DATA

# Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

**W**hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."



1

# The Problem

---

# How Aid Strategies are Sometimes Developed, but Definitely Not at *Your* Institution

This Year

	As need goes up 				
As GPA goes up 	\$5,000	\$6,000	\$7,000	\$8,000	\$9,000
	\$7,500	\$8,500	\$9,500	\$10,500	\$11,500
	\$10,000	\$11,000	\$12,000	\$13,000	\$14,000
	\$12,500	\$13,500	\$14,500	\$15,500	\$16,500
	\$15,000	\$16,000	\$17,000	\$18,000	\$19,000



## How Aid Strategies are Sometimes Developed, but Definitely Not at *Your* Institution

### Next Year

	As need goes up <span>—————→</span>				
As GPA goes up  ↓	\$5,000 <b>+\$500</b>	\$6,000 <b>+\$500</b>	\$7,000 <b>+\$500</b>	\$8,000 <b>+\$500</b>	\$9,000 <b>+\$500</b>
	\$7,500 <b>+\$500</b>	\$8,500 <b>+\$500</b>	\$9,500 <b>+\$500</b>	\$10,500 <b>+\$500</b>	\$11,500 <b>+\$500</b>
	\$10,000 <b>+\$500</b>	\$11,000 <b>+\$500</b>	\$12,000 <b>+\$500</b>	\$13,000 <b>+\$500</b>	\$14,000 <b>+\$500</b>
	\$12,500 <b>+\$500</b>	\$13,500 <b>+\$500</b>	\$14,500 <b>+\$500</b>	\$15,500 <b>+\$500</b>	\$16,500 <b>+\$500</b>
	\$15,000 <b>+\$500</b>	\$16,000 <b>+\$500</b>	\$17,000 <b>+\$500</b>	\$18,000 <b>+\$500</b>	\$19,000 <b>+\$500</b>

- Aid strategies are developed initially based on what “feels right” and then incrementally adjusted based on what “feels right” plus what “felt wrong” about the outcomes from last year.
- There’s too much data and complexity in a typical enrollment and yield cycle for our minds to process it and act on it in a systematic fashion.

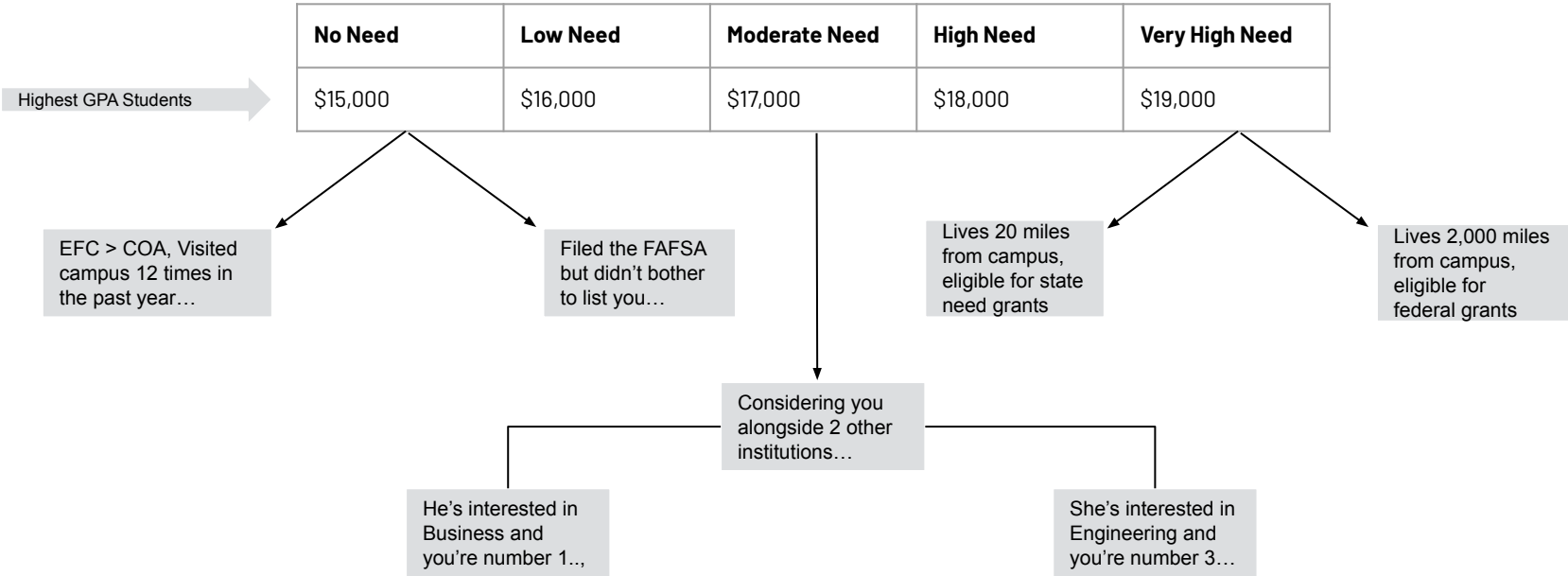
2

# Data Science to the Rescue

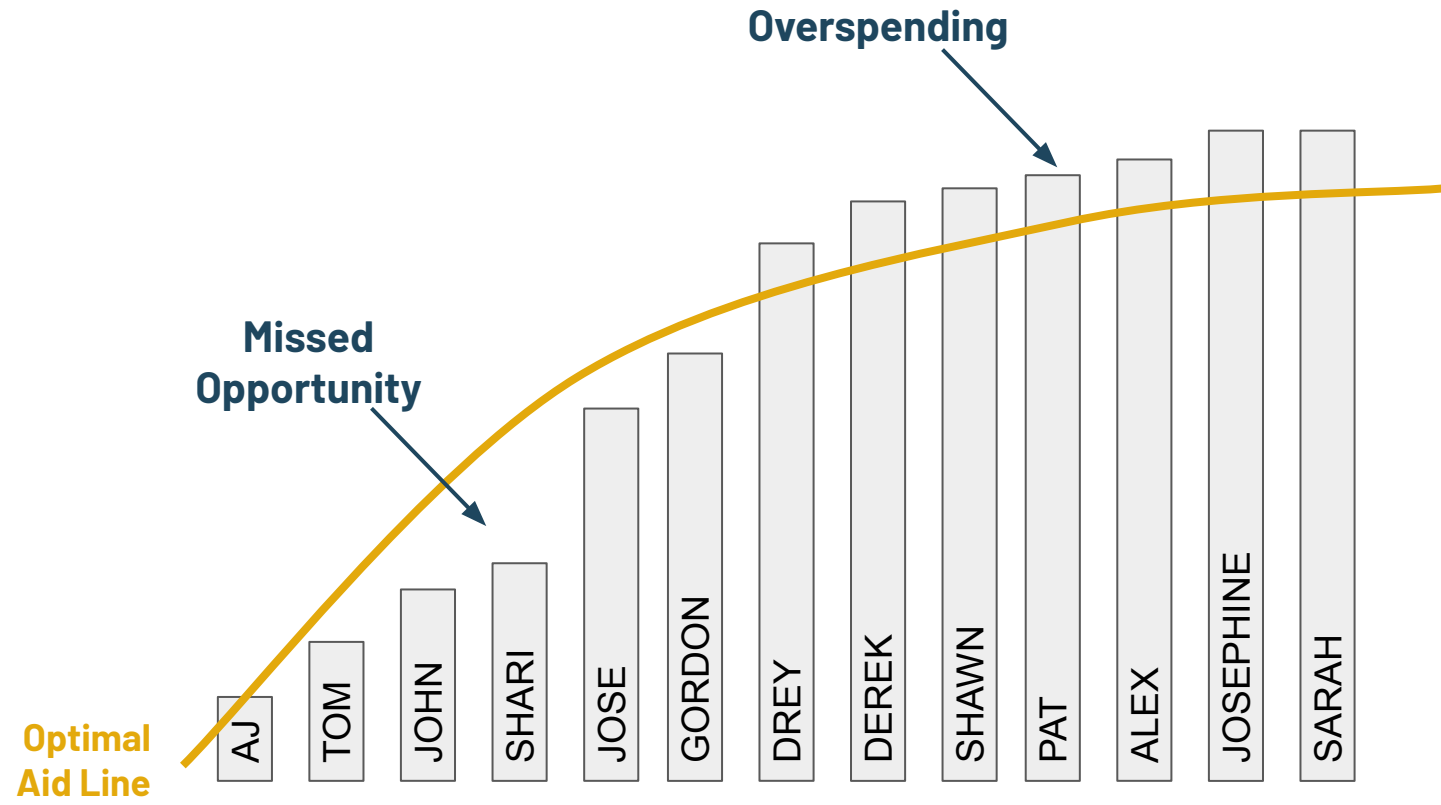
---

# There is a limit to what our human minds can process and develop

While it's often true that High School GPA and financial need are both negatively correlated with probability of enrollment, that is hardly ever the complete story.



# Optimized financial aid strategies are continuous, adaptable functions

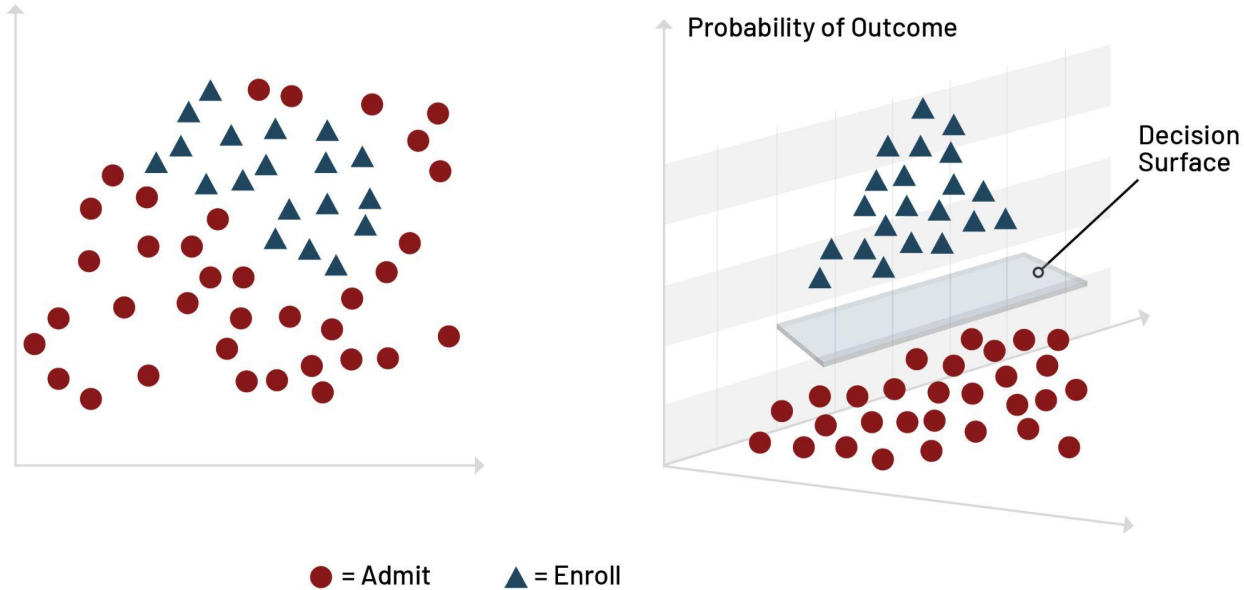


The traditional approach to financial aid optimization is...

Basic AF.

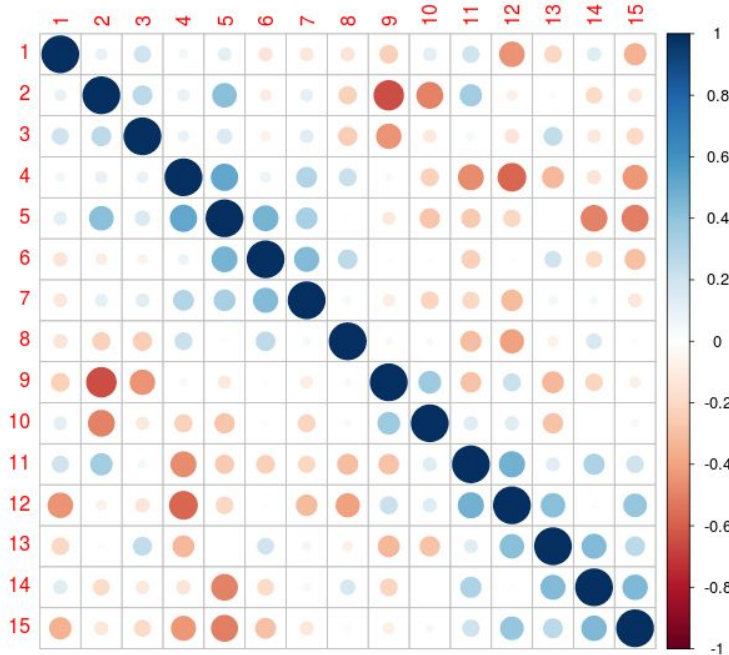
Basic AND FAULTY

## Data science can tell us at what point a person makes a decision



Term you can use to sound smart: **HYPERPARAMETER TESTING**

## Data science can summarize large amounts of data in order to uncover patterns



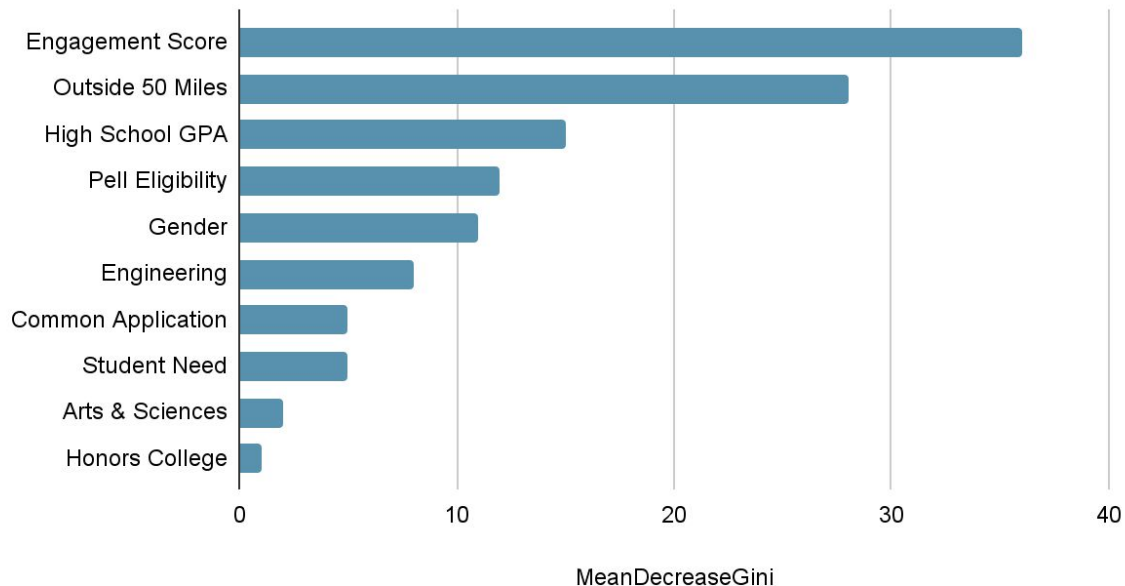
### *Variable List (Illustrative):*

1. Out of State Student
2. Independent High School
3. Female
4. Household Income > \$200k
5. Honors Student
6. Nursing Major
7. Early Applicant
8. Within 50 miles of campus
9. Household Income < \$75k
10. Business Major
11. Undecided Major
12. Pell Eligible
13. Biology Major
14. Common Application
15. Campus Visitor

**Term you can use to sound smart: CORRELATION MATRIX**

# Data science can tell us not just WHAT drives behavior, but HOW MUCH it drives behavior

Variable Importance Plot



**Term you can use to sound smart: BOOTSTRAPPING**



## How We Got Here

---

**1.** **Abraham de Moivre**  
—The Empirical Rule (1738)

**2.** **John Tukey**  
—Identifying Outliers (1972)

**3.** **George Box**  
“All models are wrong,  
some are useful” (1976)

## Where We Are Now

- + Computers are really good at conducting repetitive tasks until meaningful relationships are uncovered, humans are not.
- + Data science is a rapidly growing, highly interconnected field that is working to make computers even smarter in the future.
- + Humans are really good at interpreting findings within the context of an external environment, computers are not.
- + Data collection, interpretation, and evaluation are the most powerful tools we have for making informed, intentional decisions in an environment of uncertainty and risk.

3

# Asking the Right Questions

---

# Questions

- + Does this make sense to me?
- + What is the effect size?
- + Are there any interaction effects?
- + Is this effect linear?
- + This may be statistically significant, but is it *substantively* significant?

4

# Questions + Answers

---

# Thank You!



[CarnegieHigherEd.com](https://CarnegieHigherEd.com)