

# Going Digital: Work Faster and Smarter, Not Harder

Sean Willems

Haslam Chair in Supply Chain Analytics,  
University of Tennessee

**velocity**

CONNECT + ACCELERATE + INNOVATE


# MITx Is Transforming Education



**VERIFIED** ✓

MITx  
Supply Chains for  
Manufacturing I  
**15.762x**


Upcoming  
Starts: March 10, 2020

**VERIFIED** ✓

MITx  
Supply Chains for  
Manufacturing II  
**15.763x**

Upcoming  
Starts: May 5, 2020



## MITx Has Taught Me

- ▶ We can democratize content
  - No less than 4,000 students per course offering
  - 155 countries
  - Median age 30
  - 25% age 25 and under
  - 15% age 41 and older
  - 15% high school diploma or less, 30% with advanced degrees
- ▶ No video should be longer than 12 minutes
- ▶ Everyone listens at 1.5x speed
- ▶ We cannot achieve same learning objectives we achieve on campus, but we can still achieve significant benefits



# Nature Of Analytics Is Evolving

## COSC 526 - Introduction to Data Mining

3 Credit Hours

A comprehensive introduction to the field of data mining. Topics covered include data preprocessing, predictive modeling (e.g., decision trees, SVM, Bayes, K-nearest neighbors, bagging, boosting), model evaluation techniques, clustering (hierarchical, partitional, density-based), classification, association analysis, and anomaly detection. Case studies from text mining, electronic commerce, social science, and bioinformatics are covered. All programming projects are student-designed (no standard packages permitted).

*Recommended Background: Programming proficiency in languages such as C, C++, or Java. Knowledge of scripting languages such as Perl or Python is very beneficial.*

*WORK HARDER*



## ANTH 449 - Big-data Anthropology

3 Credit Hours

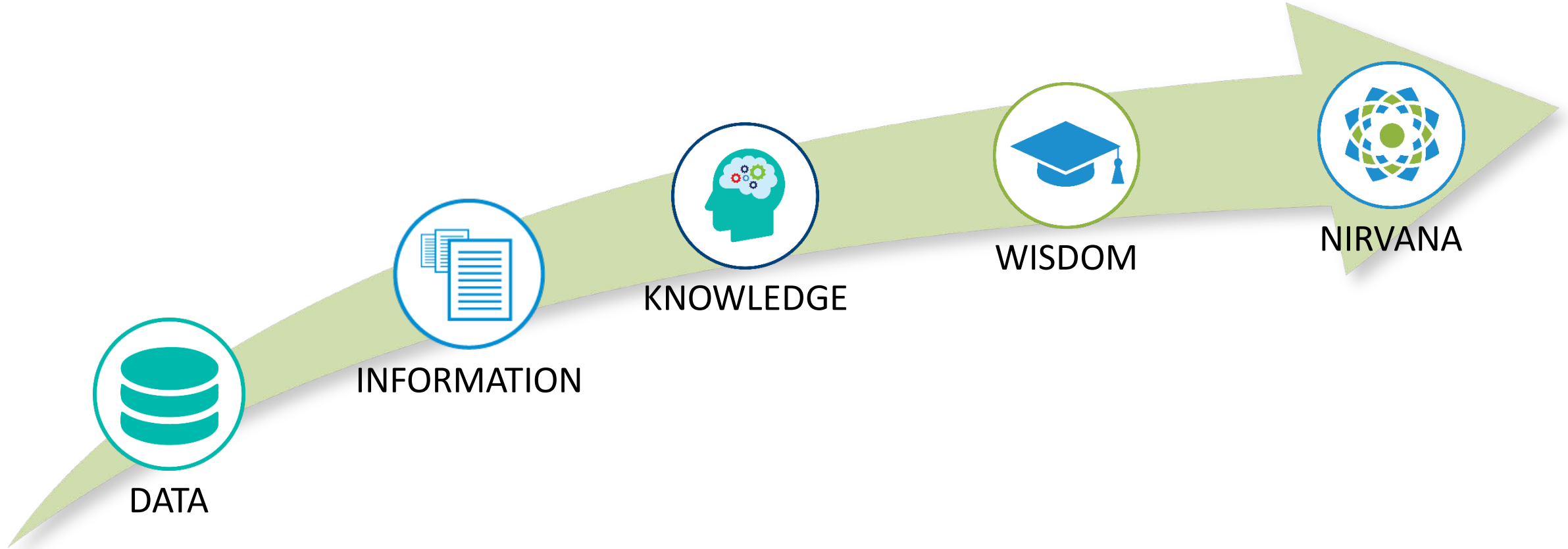
Big-data research is now a major part of the social sciences, including anthropology. Course combines lectures with practical tutorials in computational approaches using digital data on cultural change, from ancient to contemporary social media. Students are encouraged to bring a laptop, but no programming experience is required.

*(RE) Prerequisite(s): 130 or 137 or 210.*



*WORK SMARTER*

## Five-Stage Analytics Hierarchy



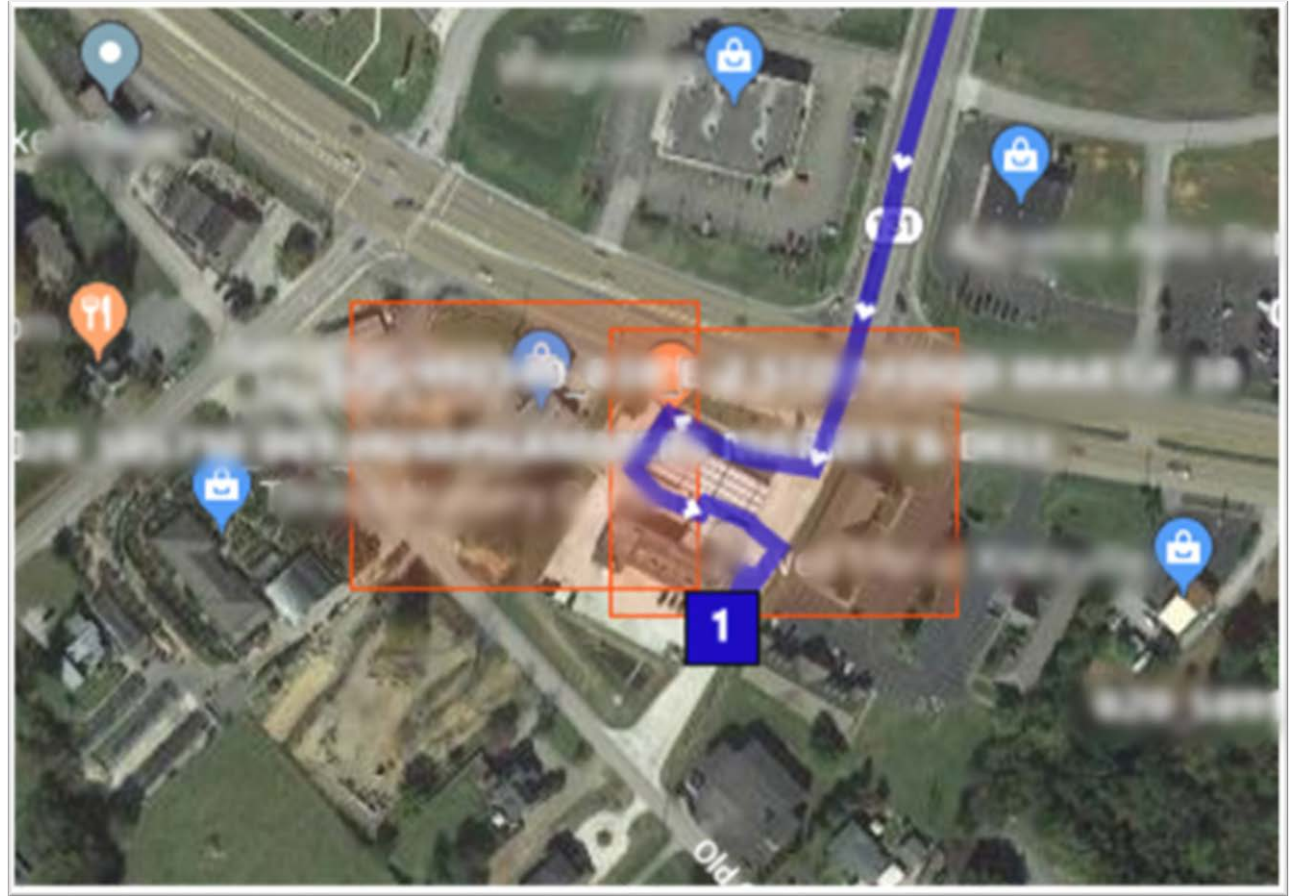
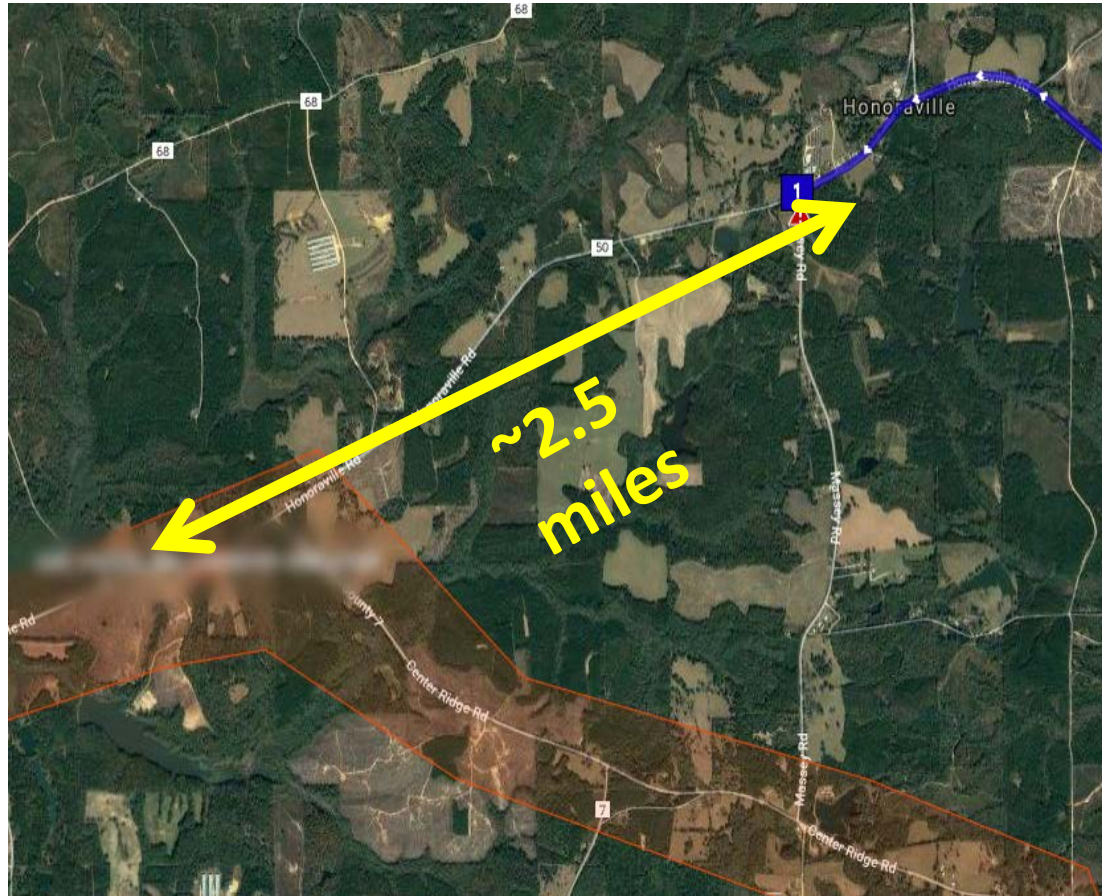
## Data Is Still Not Easy

- ▶ Two IoT-enabled machines at major defense manufacturer
  - Machines are identical
  - More than 1,000 IoT devices on each machine
  - 30 GB data stream
- ▶ Machine 1
  - 1 is operational, 0 is idle
- ▶ Machine 2
  - 1 is idle, 0 is operational





# Automated Data Collection Has Flaws



# Information Requires Processing

- ▶ RFID scans at auto plant
  - 30% scans missing from database
  - Median number of scans per vehicle ~10
  - Maximum number of scans per vehicle exceeds 2,000
- ▶ Geo-fence data for trucking company
  - 25% of stops have zero geo-fence data
  - 50% of stops are missing some geo-fence data
- ▶ Do we ignore these realities? Do we fill in the blanks?







## How Do You Attain Knowledge

### Small Data Mindset

Understand the problem's underlying drivers

Examine outliers

Ask decision makers how they solve problem today

Start small and scale

Anthropological case study

### Big Data Mindset

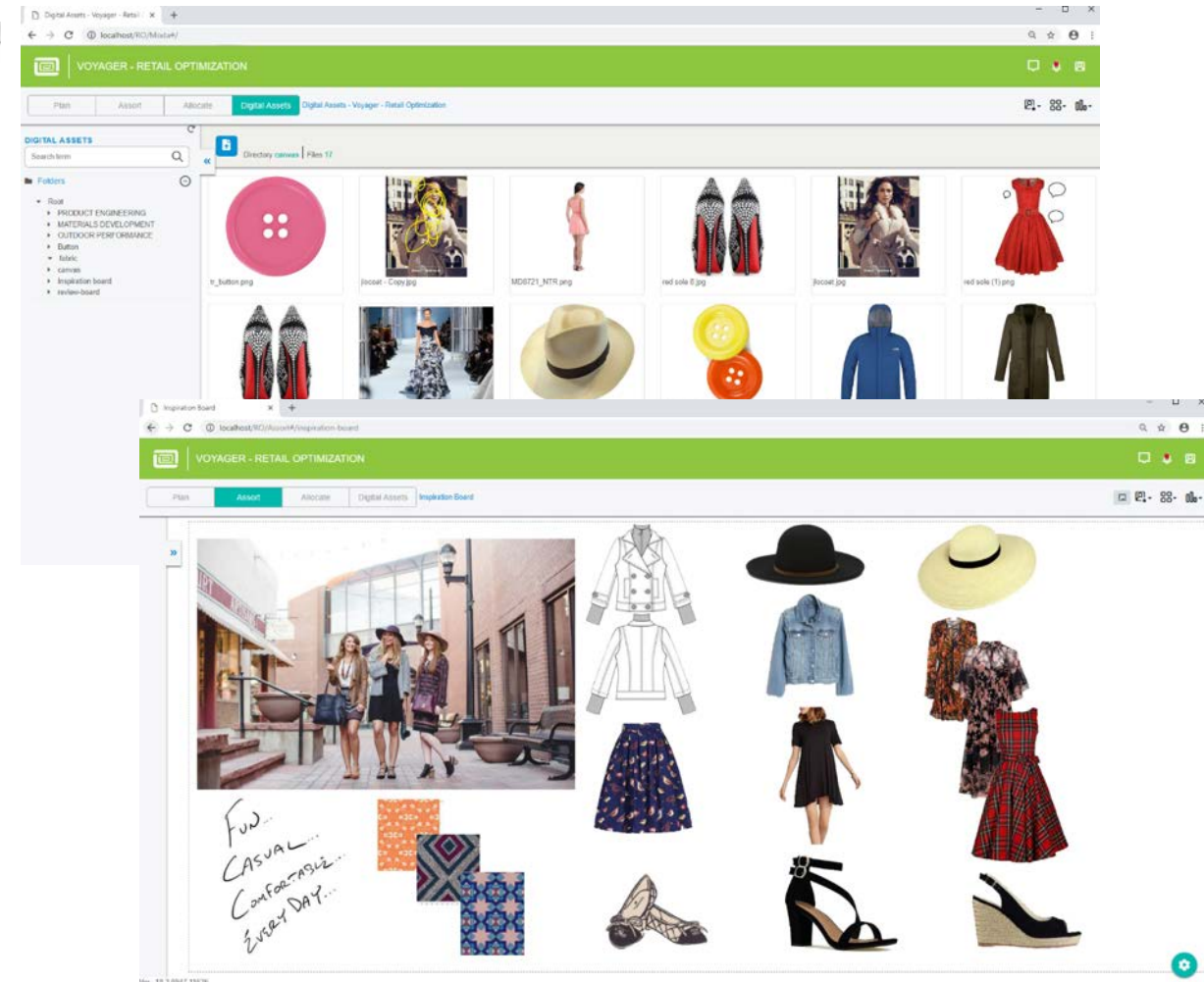
Determine what is statistically significant

Outliers wash out

Infer how decision makers solve problem today

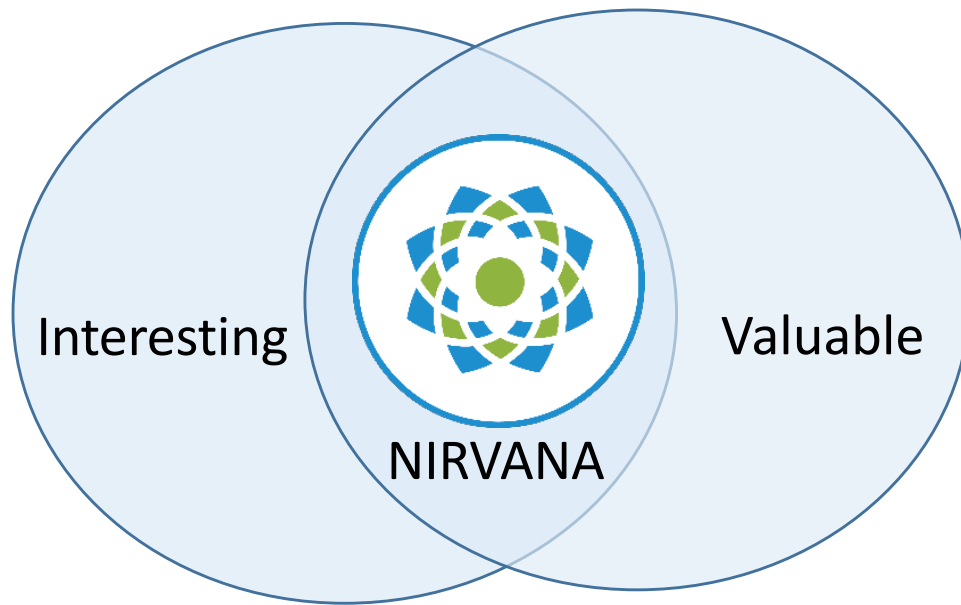
Start large

Clinical trial



## Attaining Nirvana Going Forward

Successful Decision Making =  $\frac{(\text{\#of Decisions})}{(\text{Decision Timeliness})} \times (\text{Decision Quality}) \times (\text{Decision Impact})$



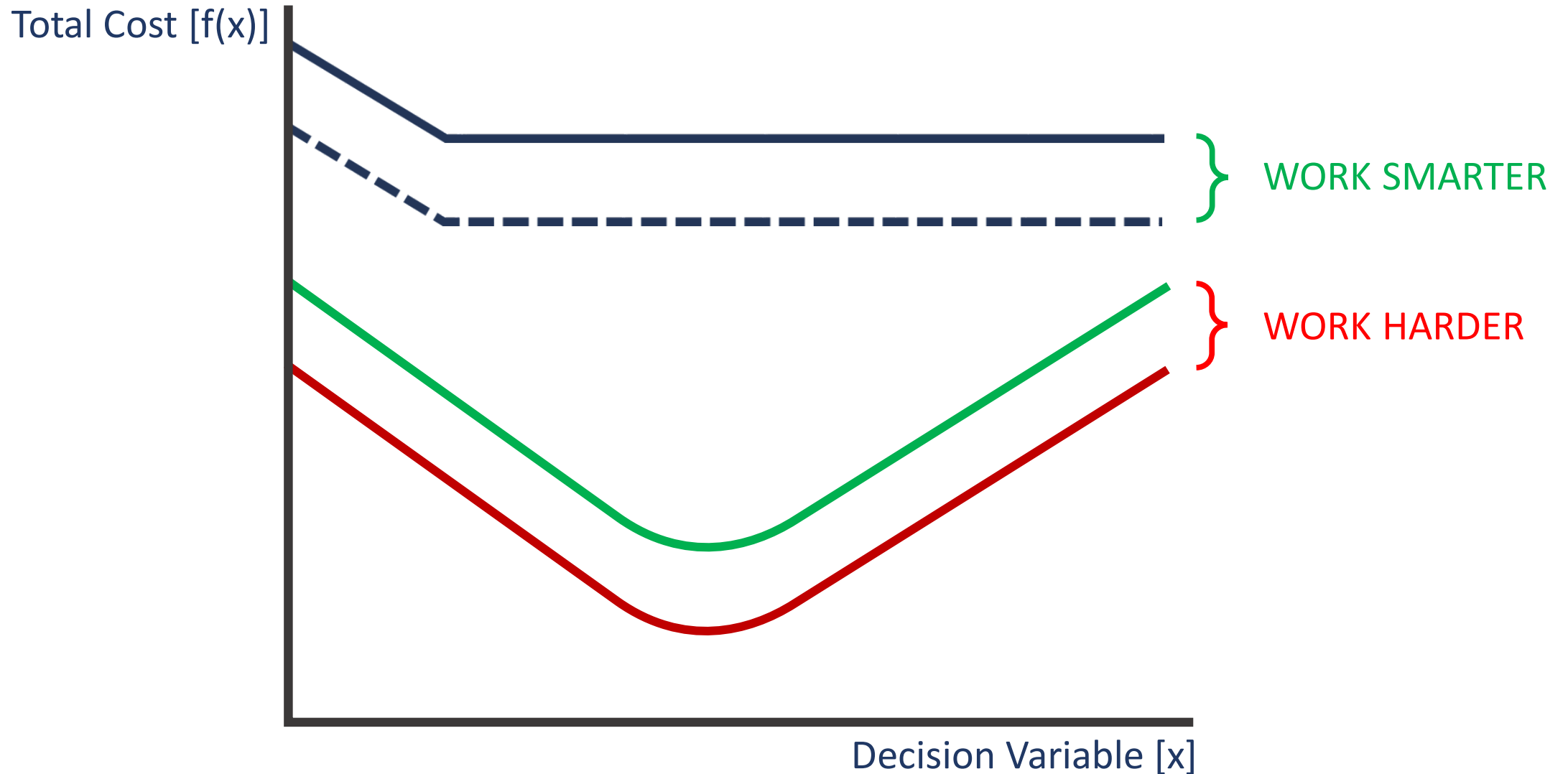
Frequency

Correctness,  
acceptability

Opportunity  
Cost/Benefit

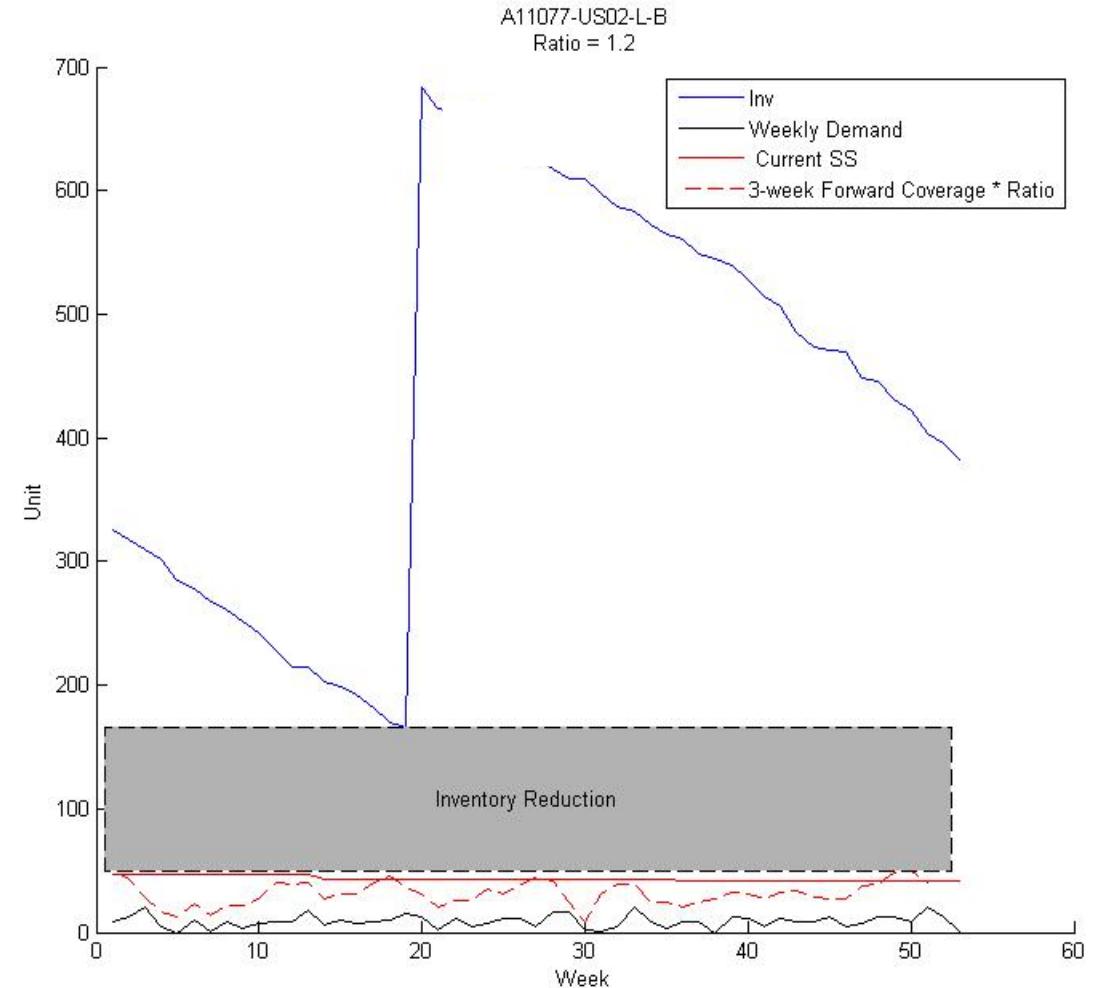


## How I Like to Teach Optimization to MIT Students

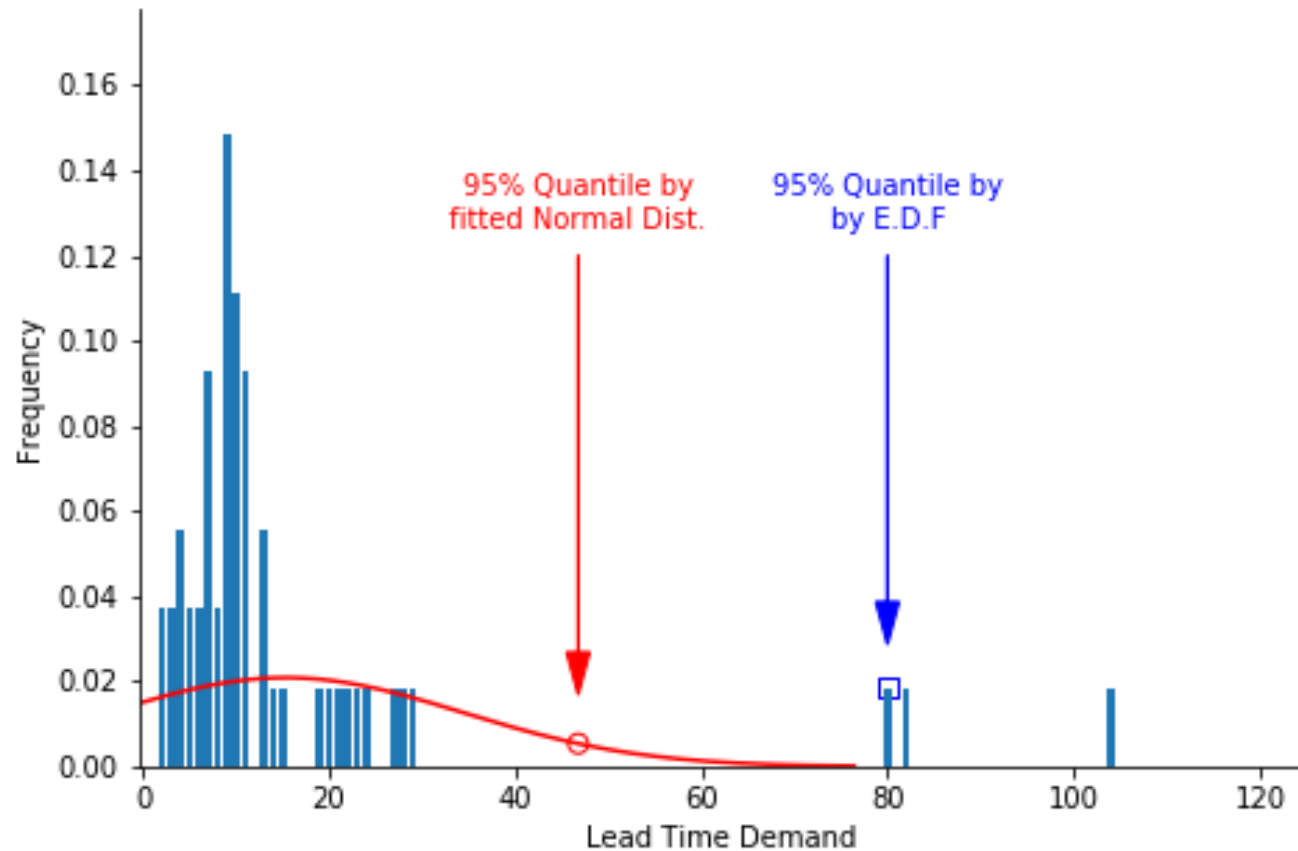


# Identification of Safe Inventory Reductions

- ▶ 10k SKU-Locations
- ▶ Filter built into decision support tools, with a fudge factor of 1.2
- ▶ Help planners identify \$3.4 million worth of **actionable inventory** (from a selection of \$10 million inventory)



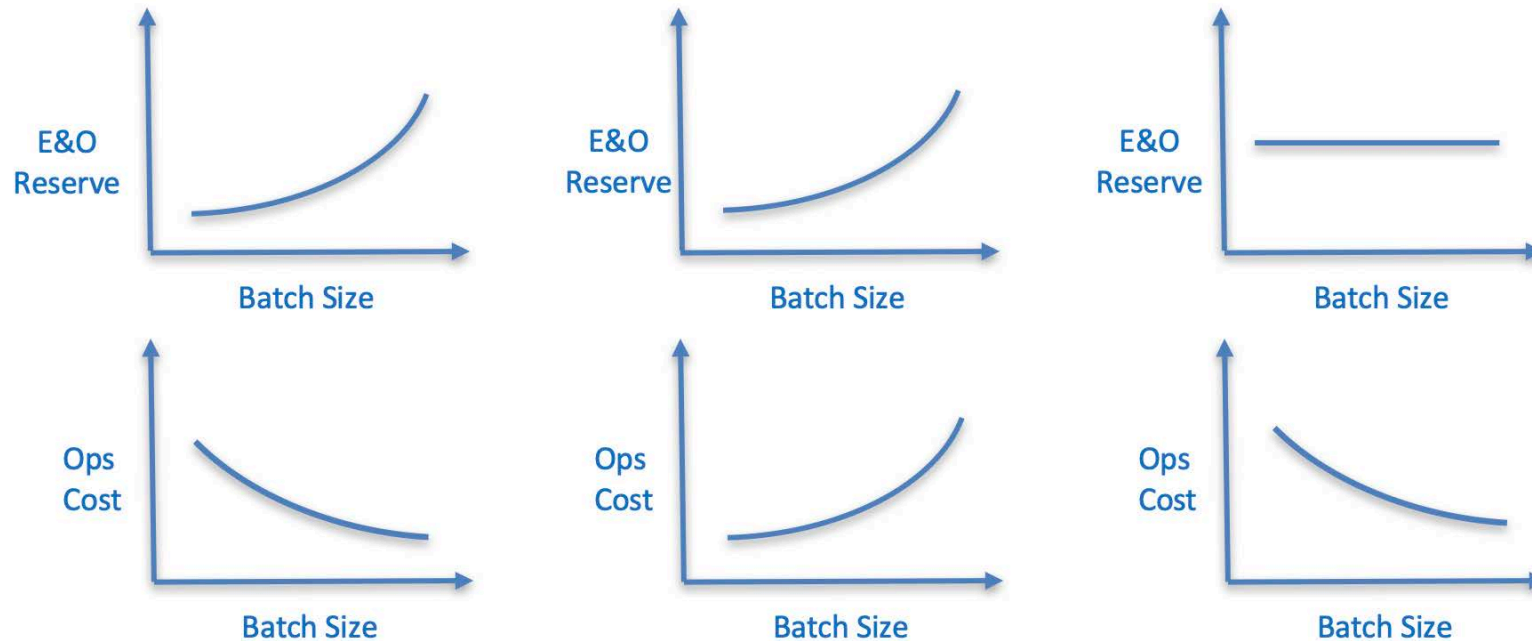
# Don't Approximate, Calculate Empirical Distribution Function



Identifying when lead time is non-negligible and i.i.d. normal distribution isn't a good fit.



# Characterization of Trade-offs Batch Size, Operational Cost, and E&O Reserve

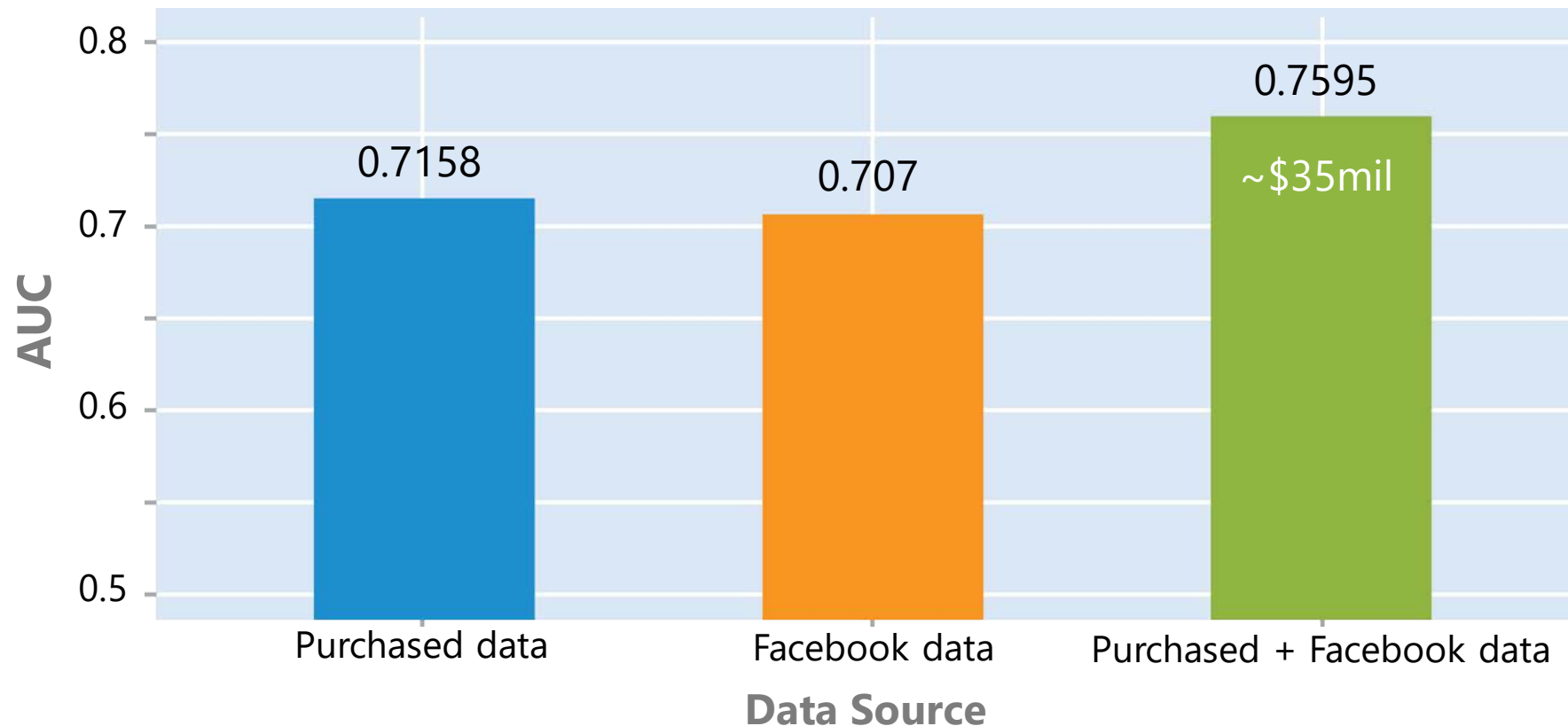


- Flexibility to include: stochastic demand, uncertain yield, production lead time, product shelf-life, and etc.

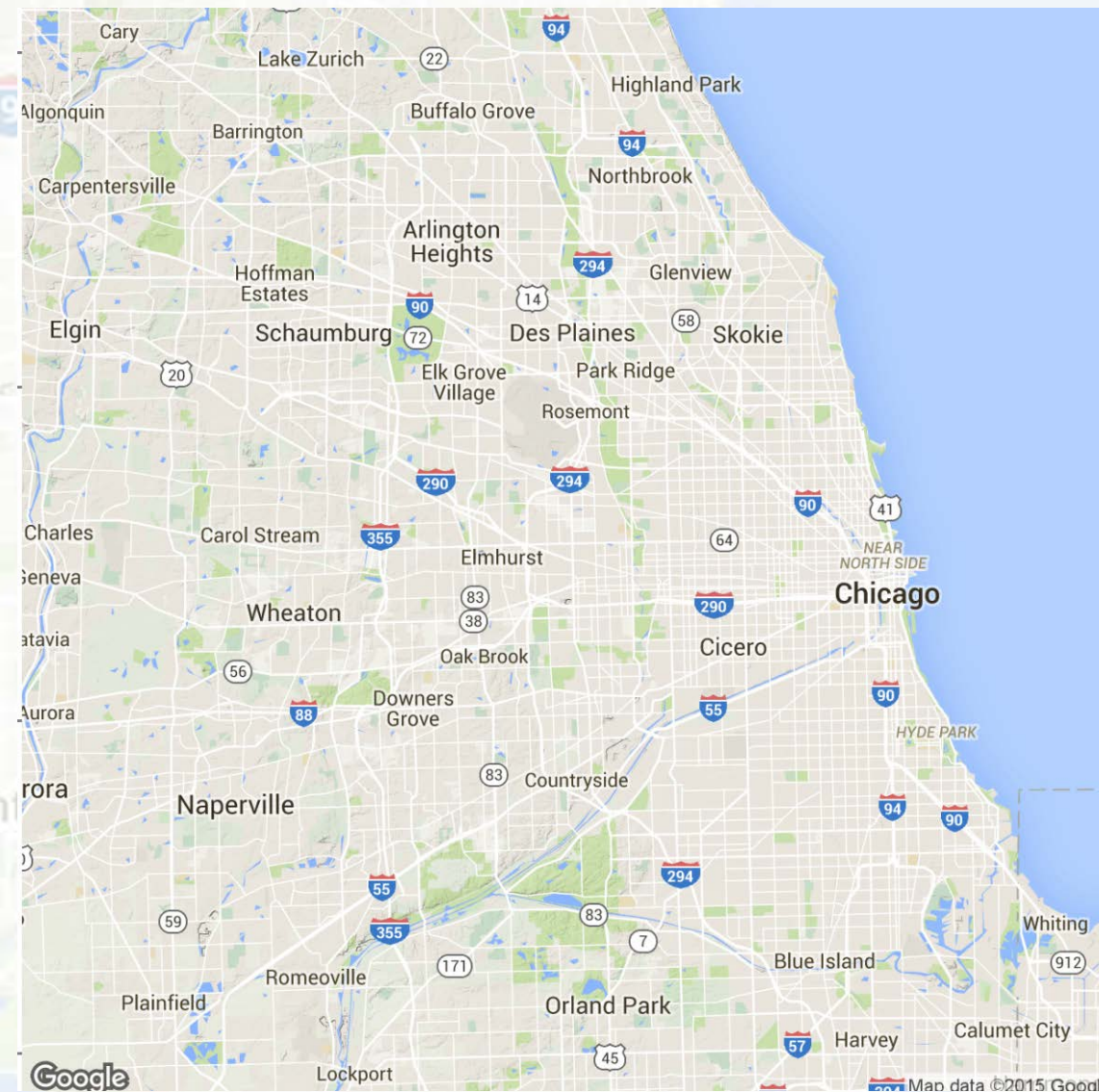
## Big Data Example from Michel Ballings

- Predict for Coca Cola Refreshments which prospects are most likely to become customers, so their call center (~5000 calls/week) can prioritize which prospects to call.

AUC comparison of acquired data vs Facebook data







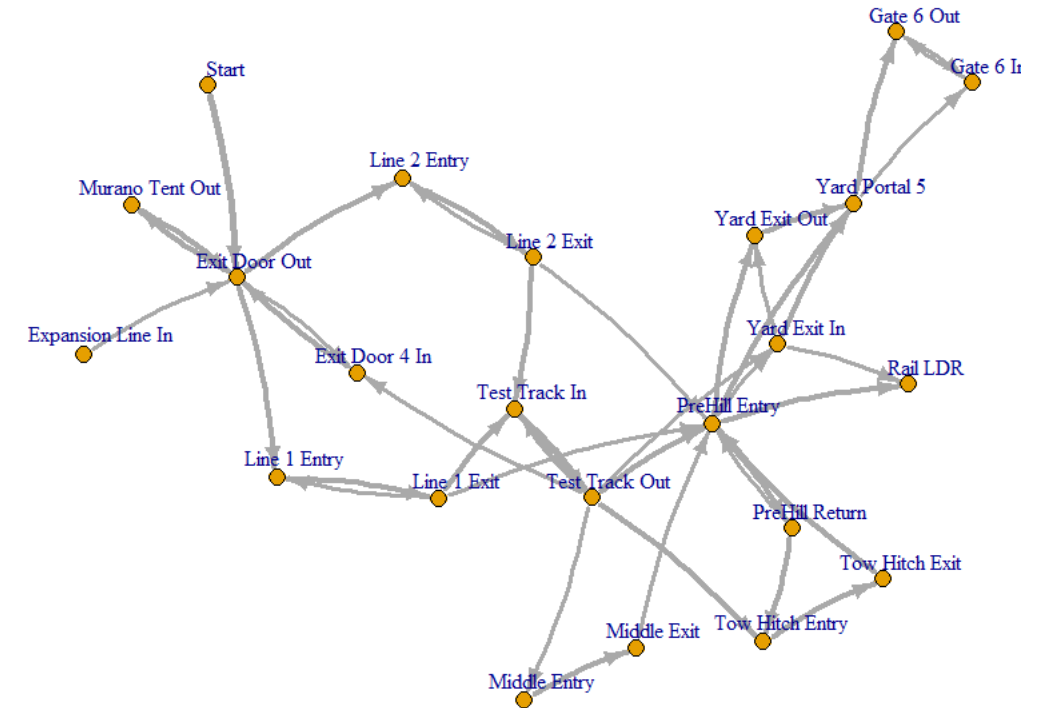
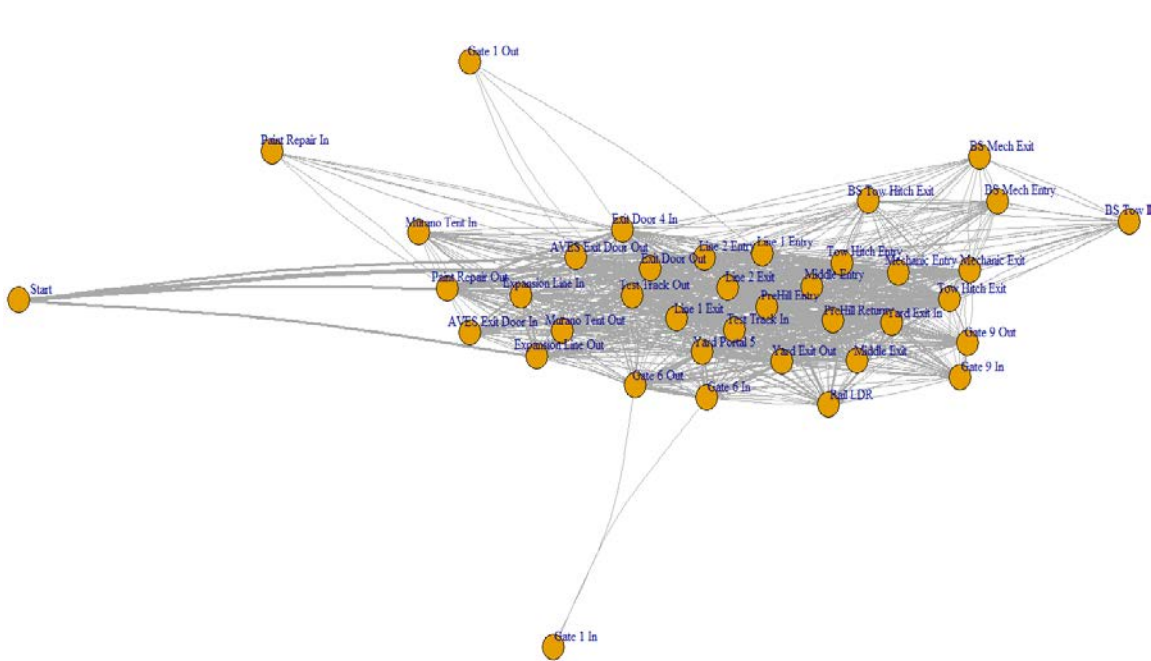


## What Are Customers Most Brand Loyal To?





## Using RFID Scans to Identify Dominant Flow Paths



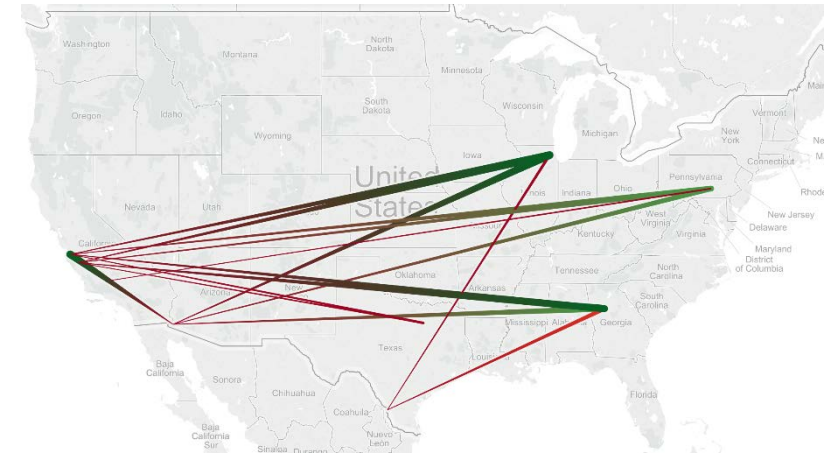
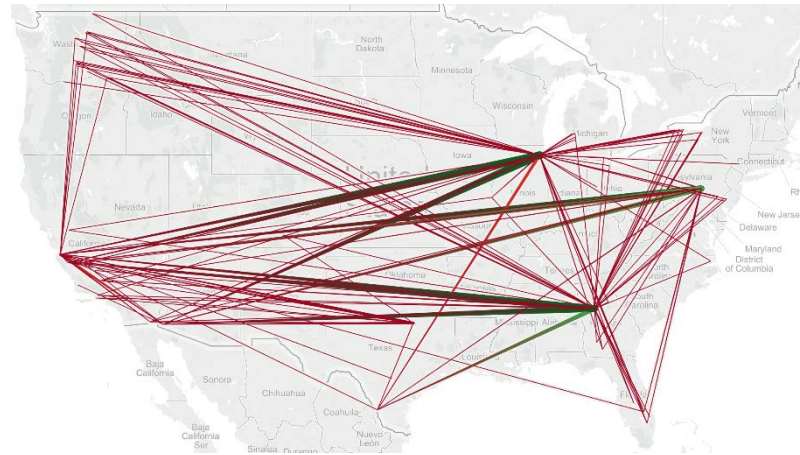
# WORK HARDER



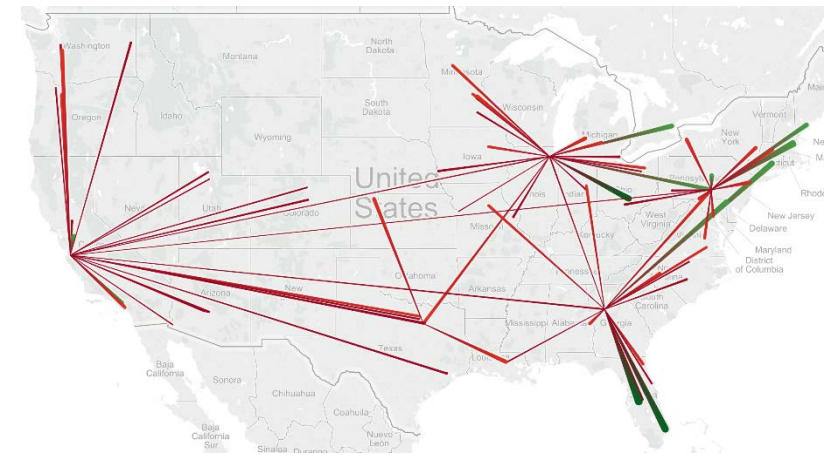
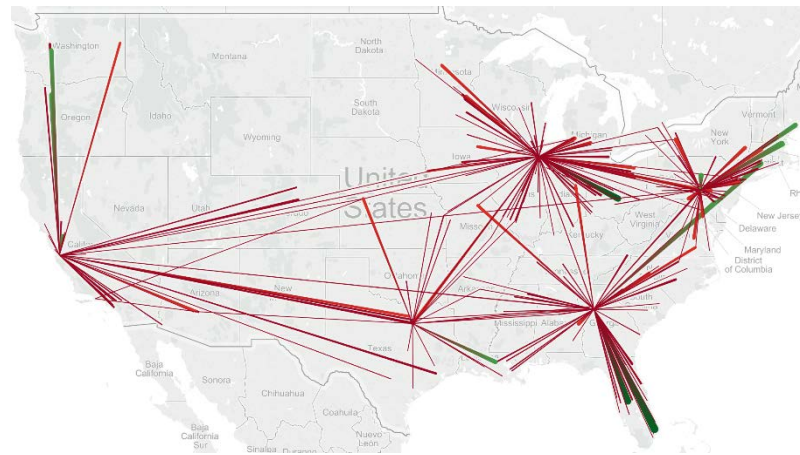
# WORK SMARTER

# Identifying Profitable Backhaul Lanes

**Raw materials**



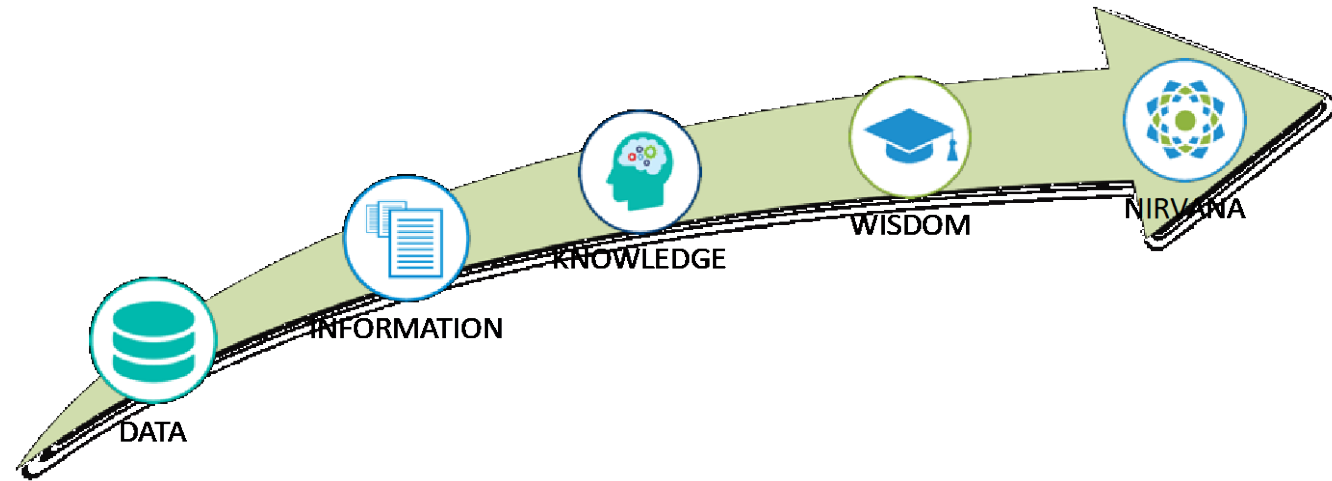
**Finished goods**





# What Do These Examples Share In Common?

- ▶ Working harder versus smarter
  - Mindset
  - Choice
- ▶ Five stage analytics hierarchy
  - Each stage is necessary
  - Each stage has challenges
- ▶ Two ways to attain knowledge
  - Small data mindset
  - Big data mindset
- ▶ Nirvana is possible
  - Intersection of valuable and interesting
  - Improve the current state
  - Simplify and accelerate
  - Communicate more, educate more





QUESTIONS?

---

THANK YOU