Have All the Cost Estimates Already Been Done?

Data Science in Cost Analysis

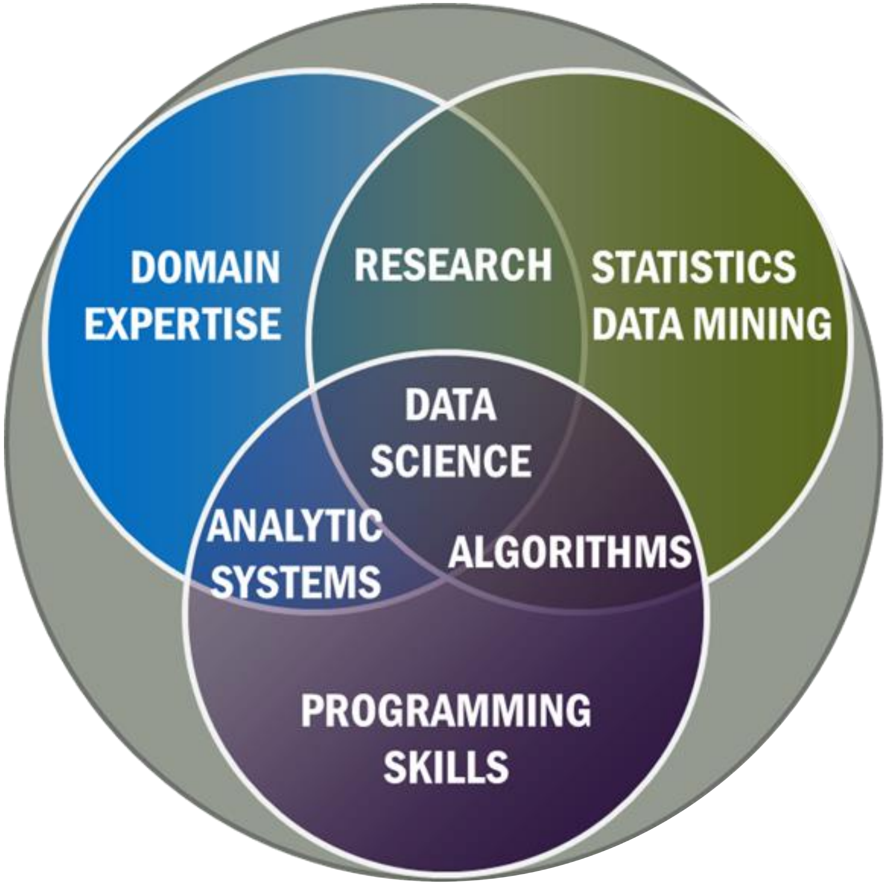
Jeremy Eden
ICEAA Washington Capital Luncheon Series - Arlington, VA
October 2018

Data Science

Data Science- What is it?

National Institute of Standards and Technology
U.S. Department of Commerce

The extraction of actionable knowledge directly from data through a process of discovery, hypothesis, and analytical hypothesis analysis



NIST Big Data Workgroup

Data Science – Who is using / will use it?

Data Science is/will soon be used by everyone (Even Cost Estimators!), but it is already used in many organizations for …

- Defense
- Fraud Detection
- Commerce
- Commercial Services
- Much More!

Booz | Allen | Hamilton

NATIONAL SECURITY AGENCY
UNITED STATES OF AMERICA

amazon

WALT DISNEY
Imagineering

AMERICAN EXPRESS

Google

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

Data Science – How good is it really?

Google Maps

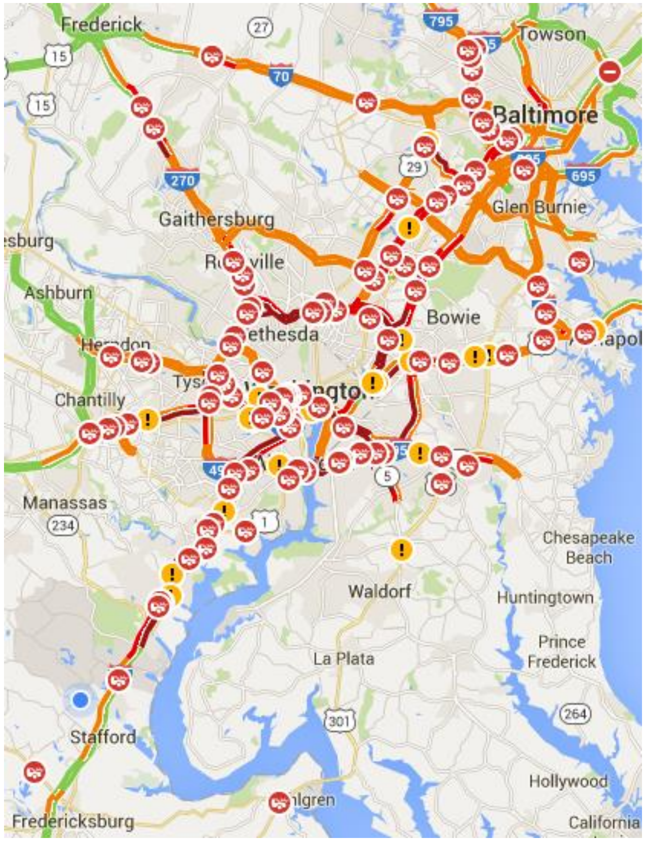
Traffic Predictions

- One of the most accurate traffic systems ever

The Data Science

- Collects data from road sensors and local transportation departments
- Over 2 billion monthly active users that receive and share data on traffic conditions

Recently achieved data analysis threshold allowing traffic to be predicted on any road, at any time of the day, and is now focused on any weather conditions



Data Science – How good is it really?

Disneyworld

Burned out lightbulbs

- Creates appearance of poor maintenance
- Not appealing to guests
- Can ruin illusions and special effects

The Data Science

- Data collected and every socket in every light and attraction in parks and resorts
- History of bulbs in that socket

Generally changes out lightbulbs during overnight maintenance the evening BEFORE the bulbs burn out

Data Science – How good is it really?

Stanford/s Lucile Packard Children's Hospital

Jenny Frankovich, attending physician

- Young girl had lupus and her kidneys were shutting down
- Some also developed blood clots which can be prevented with drugs, but those carry high risks
- Not sure if drugs should be administered or not, what to do?

Lucile Packard
Children's Hospital
Stanford

The Data Science

- How many lupus patients?
- How many with same symptoms as patients?
- How many of those patients had a clot?

Patient was treated for clots and made a full recovery

Data Science – How good can it get?

Wired Magazine- June 2008

Chris Anderson -The End of Theory: The Data Deluge Makes The Scientific Method Obsolete

Scientific Method: “Correlation is not Causation”

- No conclusions on basis of correlation between x and y
- Must understand underlying connections for a model
- Hard to test and experiment unknowns

Data Science Method: “Correlation is enough”

- Computing power and large amounts of data at the problem
- Find patterns in data
- All questions can be answered even if we don’t know why

The World’s Cheapest Car | 23 Hot Summer Gadgets
Get Ready for the Google Phone
WIRED
16.07
JUL.2008
Do we need THEORY?
How do we find DARK MATTER?
Is Earth safe from ASTEROIDS?
Where will TERRORISTS strike next?
How can you win any LAWSUIT?
Who will be the next PRESIDENT?
How will we grow enough FOOD?
How can we protect BUSINESS from risk?
The End of Science
The quest for knowledge used to begin with grand theories.
Now it begins with massive amounts of data. Welcome to the Petabyte Age.

1976 George Box - Statistician - Journal of the American Statistical Association

"All models are wrong, but some are useful.“

2008 Peter Norvig - Google’s research director - O'Reilly Emerging Technology Conference

"All models are wrong, and increasingly you can succeed without them."

Data Science – You're Freaking me out!

Lucile Packard Children's Hospital Stanford

Stanford/s Lucile Packard Children's Hospital

Abandoned the program

- Concern that they didn't understand why some people had clots and others didn't even if they could predict who would and wouldn't develop them correctly

Google

Organic (unpaid) search results

"…we don't know why one website is better than another one"

- If the statistics of incoming links say one website is better than another, that's good enough
- Can match ads to content without knowing anything about either
- Can translate languages without knowing either

Data Science and Cost Estimating

Data Science & Cost Estimating – Who is doing it?

ICEAA
amazon
Walt Disney
Imagineering
Booz | Allen | Hamilton
DAU

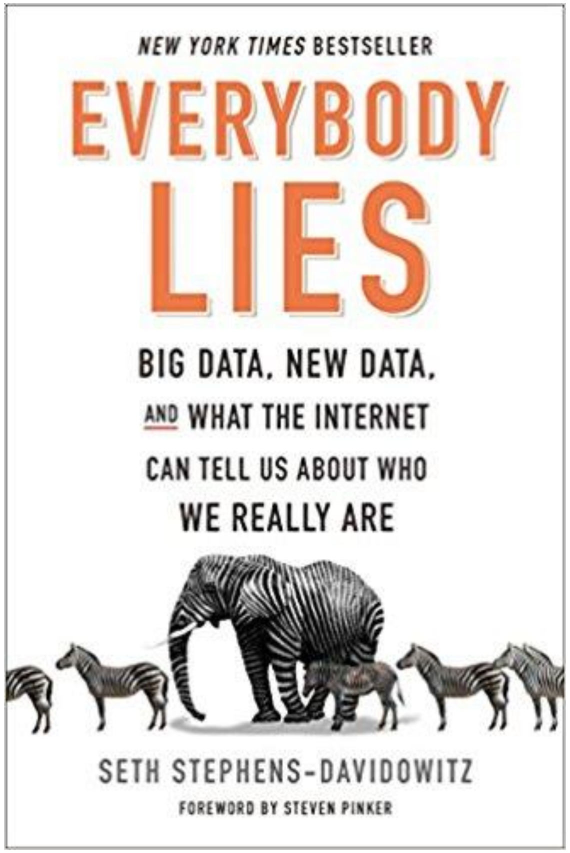
Organizations using data science for other disciplines are also using it for cost estimating. Additionally, the ICEAA community and it's members have been used some data science techniques already ….

- Construction
- Software Cost Estimation
- General Cost Estimation
- Defense Cost Estimation
- Computing and Network Structures … and more

Data Science & Cost Estimating – I want a Cost Estimating Example!

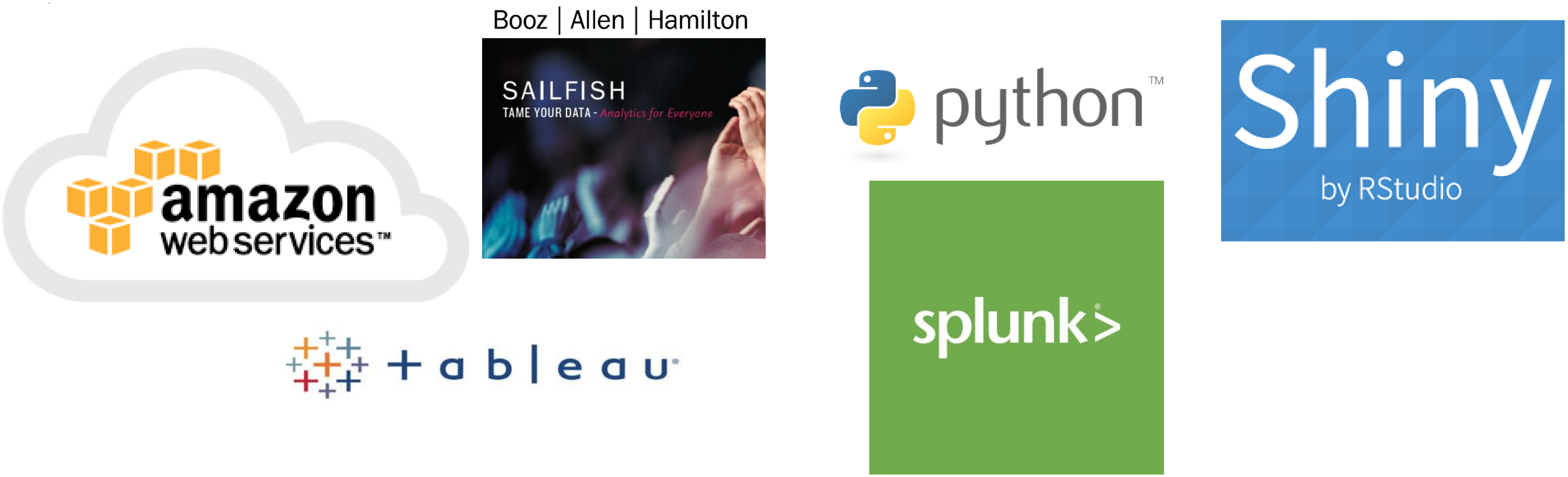
L2 Inc Clinic on Digital Impacting Consumer Behavior

- Seth Stephens-Davidowitz - Presentation of “Google is Digital Truth Serum” (https://www.youtube.com/watch?v=TS2der4Ag_s&t=2s)
- Jeff Seder
  - Leaves Wall Street and uses Data Science to find the most cost effective Race Horse
  - Creates databases on horse nostril size, fast twitch muscles, size of left ventricle of heart vs price
  - Advises client to buy his own horse at Auction for $300k (minimum price)
  - Horse was American Pharoh, wins the triple crown



The debate rages on why the left ventricle of a race horse can indicate a winner

Data Science & Cost Estimating – What tools are used?



These are just a few of a large selection of tools available. Many tools have the advantage of…

- Open source technology
- Low cost or free licensing
- Wide acceptance by stakeholders
  - Example: R is Department of Defense (DoD) Approved

Data Science & Cost Estimating– You're STILL Freaking me out!

How can Data Science be used for cost estimating when you can't provide actuals or a methodology for models?

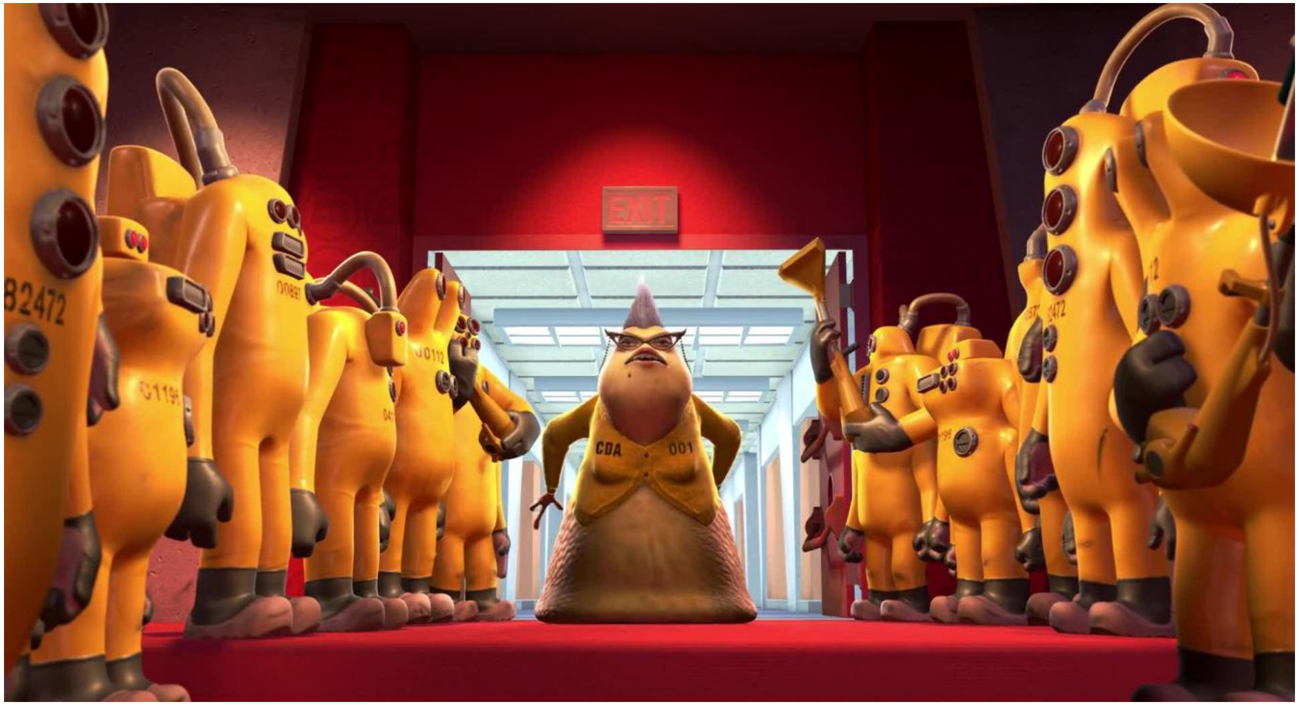
- We can do this because "correlation is enough" when we are applying a large amount of data and a lot of computing power at the problem
- We can "answer the question without knowing exactly why"

…but you will because of gaps in the data, unreliable data, outdated data, and other reasons…

- There are already a number of techniques for filling, cleaning, updating, and even RATING data
- Remember "all models are wrong, and we can succeed without them" as long as the accuracy is equivalent or better to what models would produce

In time, data science cost estimating will be MUCH faster and more accurate

ICEAA & Data Science – Run! ICEAA will call the CDA!



Disney Monsters Inc.

Wait! There is no need for ICEAA to be alarmed or to call the CDA (Cost Decontamination Unit)!

Over the last few years ICEAA has said it is a priority to:

- Advance the cost estimating industry by supporting new fields like Agile and Data Science
- Enable the growth of the cost estimating profession and ICEAA membership
- Revise reference material and training opportunities as needed to include the latest techniques

ICEAA & Data Science – How can ICEAA participate?

ICEAA has an opportunity to grow and serve it's membership at the same time through data science cost estimation …

- Include sections on data science cost estimating techniques in the Cost Estimating Book of Knowledge (CEBoK) or create an **Data Science Cost Estimating Book of Knowledge (DSCEBoK)**
- Include specific data science cost estimating in the currently offered classroom training or offer data science versions of the classroom training
- Create a Data Science Certified Cost Estimator/Analyst (DSCCA) credential (or a specialty credential similar to the Parametric Specialty Certification) that compliments the existing Certified Cost Estimator/Analyst (CCEA) along with the requirements guidance to sustain the credential

ICEAA

DATA SCIENCE

CEBoK®

A Data Science Cost Estimator Certification would standardize the baseline knowledge required to conduct cost estimates using data science tools/techniques and further expand ICEAA

Ok, I'm in! How Do I Get Started?

You can obtain data science skills and incorporate them in your cost estimating today by…

- Instead of just focusing on models you can use, try focusing on data
- Try a tool like Python or R (remember R is free and DoD approved) for analysis
- Attend formal or informal data science training to learn some common techniques and tools
- Capture data/favorite data sites whenever possible even if it seems unrelated to what you are doing

Seth Stephens-Davidowitz - "There are left ventricles out there."

Open the ICEAA app on your phone RIGHT NOW, go to this session, take the survey, tell them them how INCREDIBLE this presentation was, and you want to know when ICEAA will offer a Data Science Cost Estimating Certification

Lame! Lame! Lame! How Do I Get Started TODAY?

Here are a list of resources you can go to TODAY that will enable you to start using Data Science in your cost estimates TONIGHT …

Learn R

HARVARD UNIVERSITY

https://online-learning.harvard.edu/course/data-science-r-basics

Capture data/favorite data sites whenever possible even if it seems unrelated to what you are doing at the time

DATA.GOV

https://www.data.gov/

amazon web services

https://aws.amazon.com/datasets/

CENTRAL INTELLIGENCE AGENCY

THE WORLD FACTBOOK

https://www.cia.gov/library/publications/the-world-factbook/

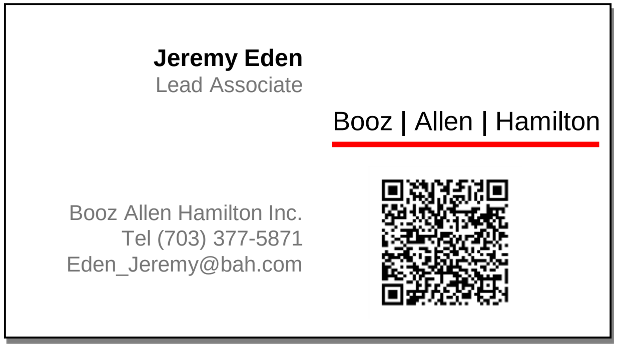
Government of Canada

https://open.canada.ca/en

EU Open Data Portal

http://data.europa.eu/euodp/en/data/

Questions



Jeremy Eden
Lead Associate

Booz | Allen | Hamilton

Booz Allen Hamilton Inc.
Tel (703) 377-5871
Eden_Jeremy@bah.com

Sources

"The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," 2008.
https://www.wired.com/2008/06/pb-theory/

"Big Data Not a Cure All In Medicine," 2015.
https://www.npr.org/2015/01/05/375201444/big-data-not-a-cure-all-in-medicine

Barker, Laura and Wilson, Josh, "Integrating Cost Estimating and Data Science Methods in R," The ICEAA Professional Workshop, June 2016.
http://www.iceaaonline.com/ready/wp-content/uploads/2016/06/PA11-ppt-Integrating-Methods-in-R.pdf

Stephens-Davidowitz, Seth, "Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are"
https://www.amazon.com/Everybody-Lies-Internet-About-Really/dp/0062390856