midwest
architecture community
collaboration

MIDWEST ARCHITECTURE COMMUNITY COLLABORATION 2020

NOVEMBER 5, 2020

MACC 2020: Adaptable architecture: building resilience in a time of change

MACC MISSION

- The Midwest Architecture Community Collaboration's (MACC) purpose is to bring all domains of architecture together to share information and techniques of interest to all of us. It is our shared belief that through collaboration, we can better understand and promote the significance of architecture to business success.

midwest architecture community collaboration

MODELING THE REAL WORLD WITH ACCURACY AND PRECISION

- Bob Grogan/Picture goes here

- Why this topic and why me?
- Trained a Materials Scientist/Engineer
- Preferred Applied Mathematics
- Unapologetic Architect
- Engineering leader

midwest architecture community collaboration

RESILIENCE

- The American Psychological Association (2014) defines **resilience** as "the process of adapting well in the face of adversity, trauma, tragedy, threats or even significant sources of stress
- **Resilience Theory** argues that it's not the nature of adversity that is most important, but how we deal with it. When we face adversity, misfortune, or frustration, **resilience** helps us bounce back. It helps us survive, recover, and even thrive in the face and wake of misfortune – but that's not all there is to it.

THE PROBLEM SPACE

What I Hope to Cover

- We often drive toward a precisely toward an answer that is inaccurate or at least misguided
- Applying methods of problem analysis and an architectural perspective should inform data science efforts
- **We want to maximize the work not done**

What I Hope to Avoid

- Ethical implications
- Opinions of data science algorithms or tools
- Use cases around politics or consumer behavior
- The word Kafka

ARCHITECTURE AND DATA SCIENCE

Domain Architecture

- Theorized Parameters of Significance
- Business Implications
- Data Quality
- Data Characterization

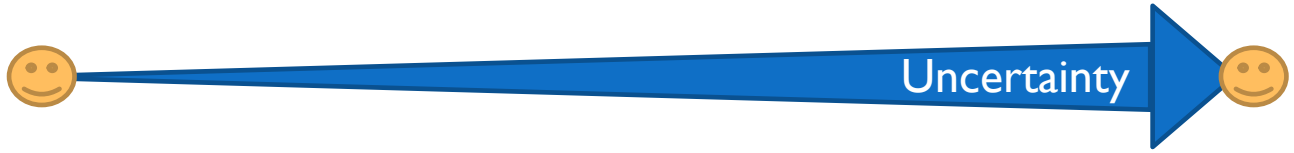
Technical Architecture

- Data streaming cloud hosted big data
- Supervised vs Unsupervised
- Python vs R
- Jupyter Notebooks

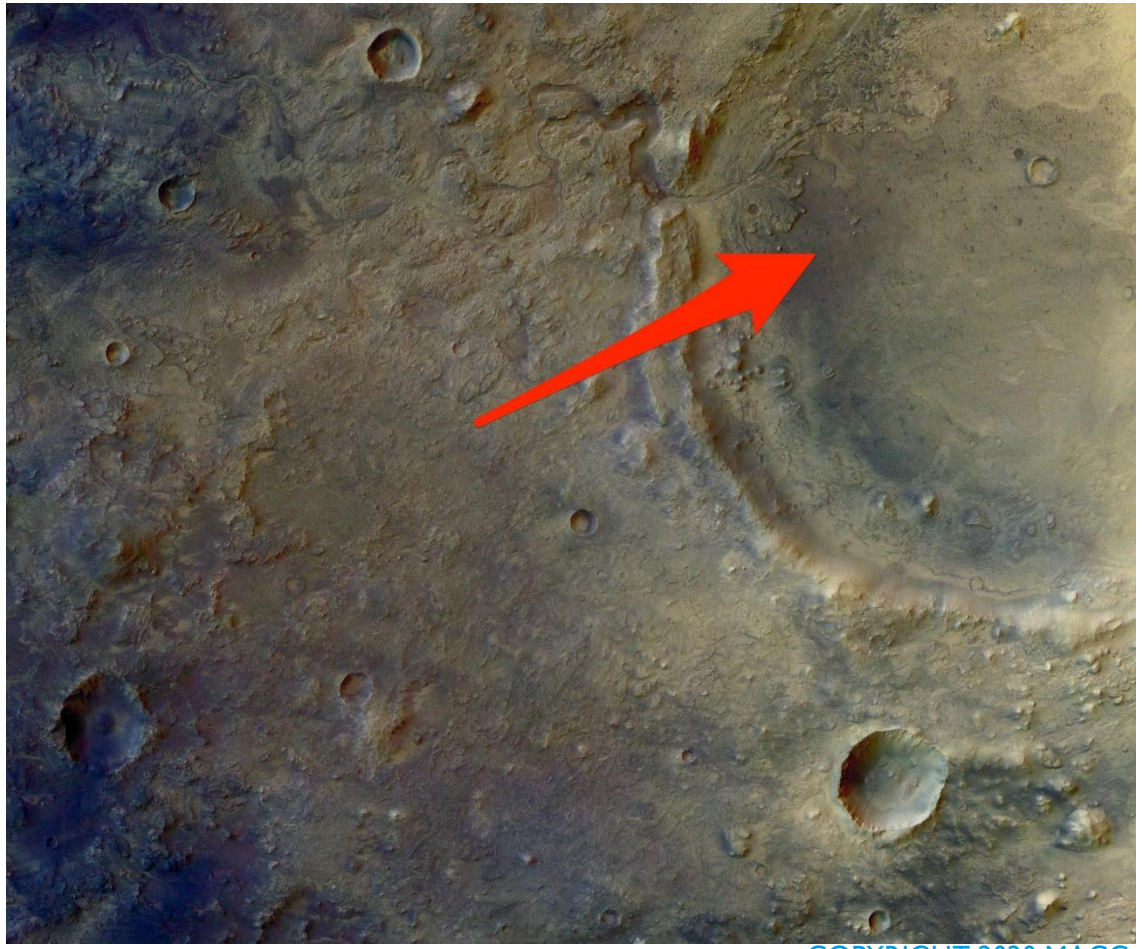
REMINDER OF ACCURACY AND PRECISION

Qualifying the Nature of the Outcome

Chance that a cloud data pipeline will produce a result



Firewall | Ingest | Transform | Persist | Model



midwest architecture community collaboration

A BRIEF ORIGIN STORY- APPLIED MATHEMATICS

- Gaussian diffusion experiment
- Parameter trends (linear, logarithmic, exponential)
- Development of specific formulation of the real world

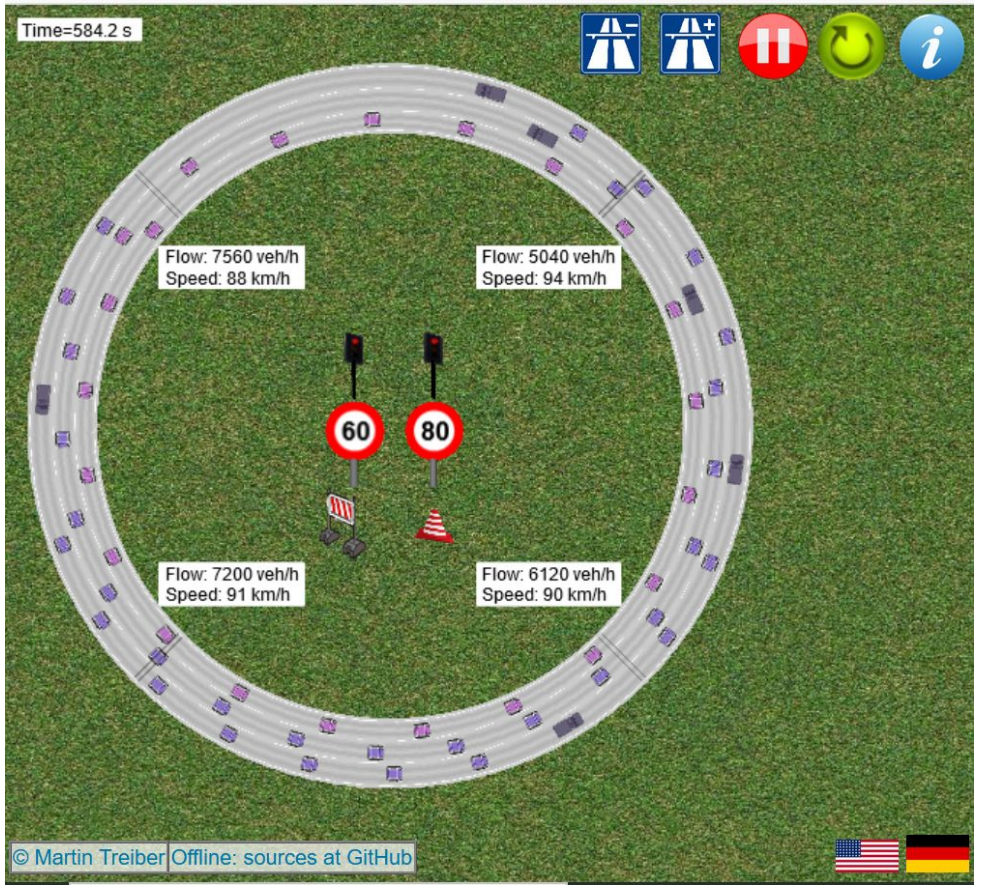
Equation

Asymptotes

midwest architecture community collaboration

TRAFFIC HARMONICS

- After years of intelligent automatons running models on supercomputers, a team in Japan put 22 cars on a road
- Theoretical models with simple rules can easily create car crashes

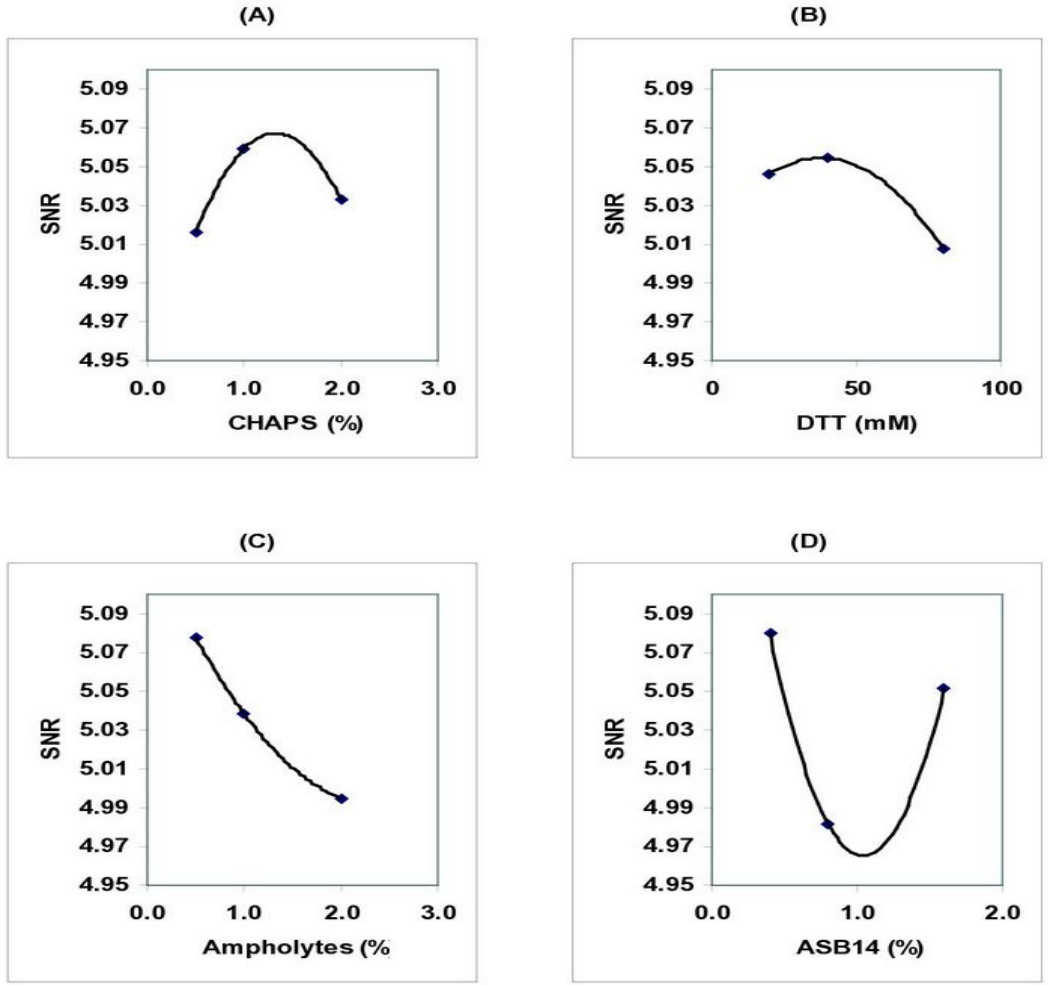


https://traffic-simulation.de/ring.html

midwest architecture community collaboration

TAGUCHI DESIGN – SMALL DATA

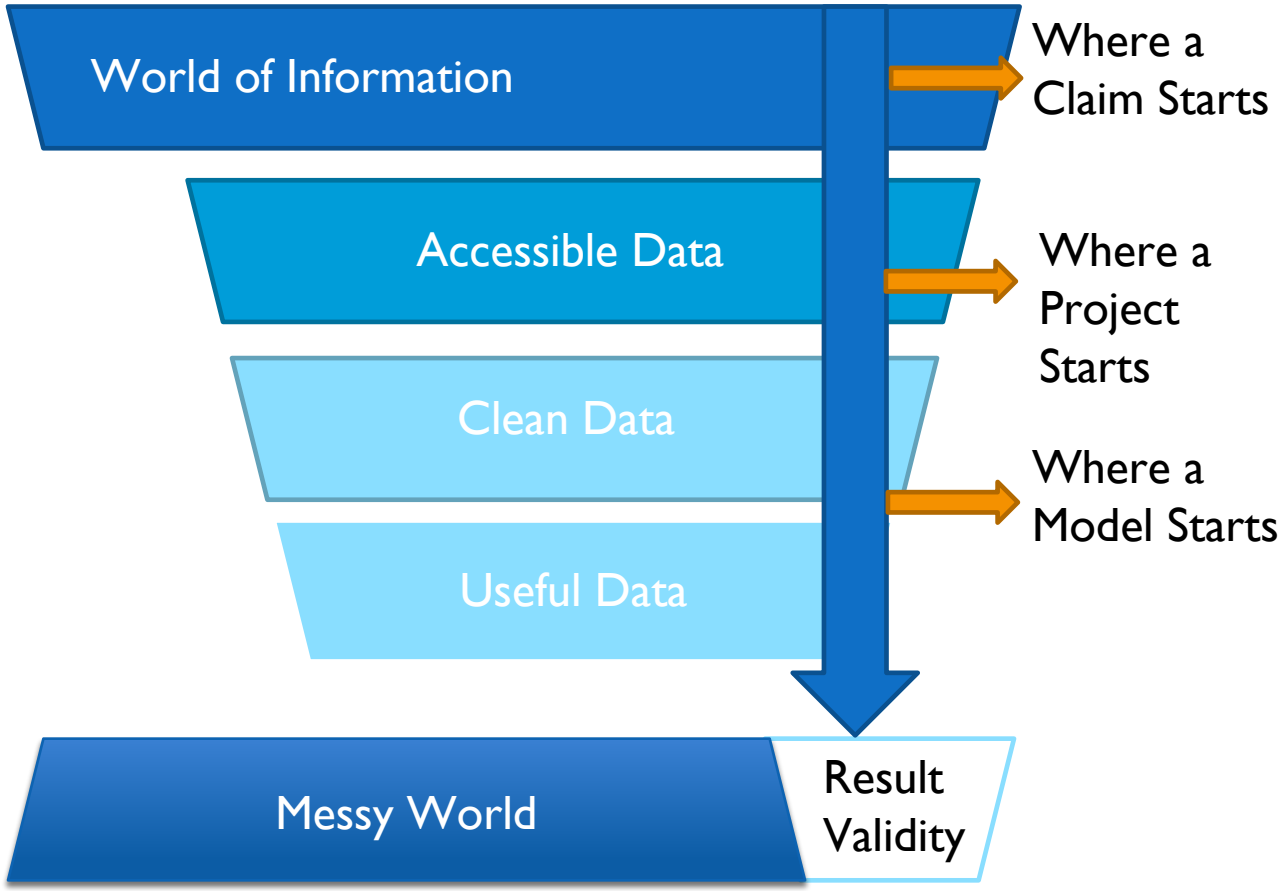
- Taguchi Design- the anti- big data
- Trying to judge signal from noise by small sets of data with independent variables
- Produce a matrix to develop trends



midwest architecture community collaboration

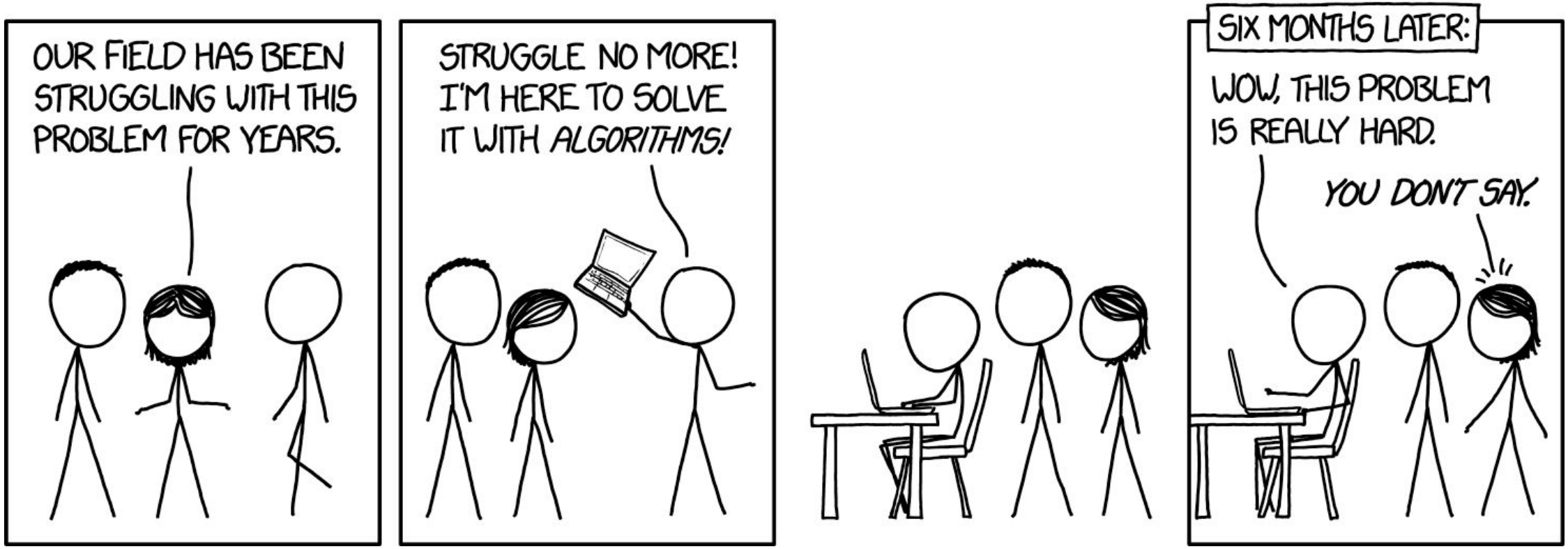
DATA SCIENCE- BY COMPARISON

- Start with the real world (credit transactions, average global temperatures, clickstreams)
- Aggregate and analyze the data and apply algorithms to find patterns
- Monitor the accuracy of the model and refine (learn)
- Teach machines to act in the face of uncertainty (AI)



midwest architecture community collaboration

THE APPROACH IS SUPPLEMENTAL NOT MAGIC

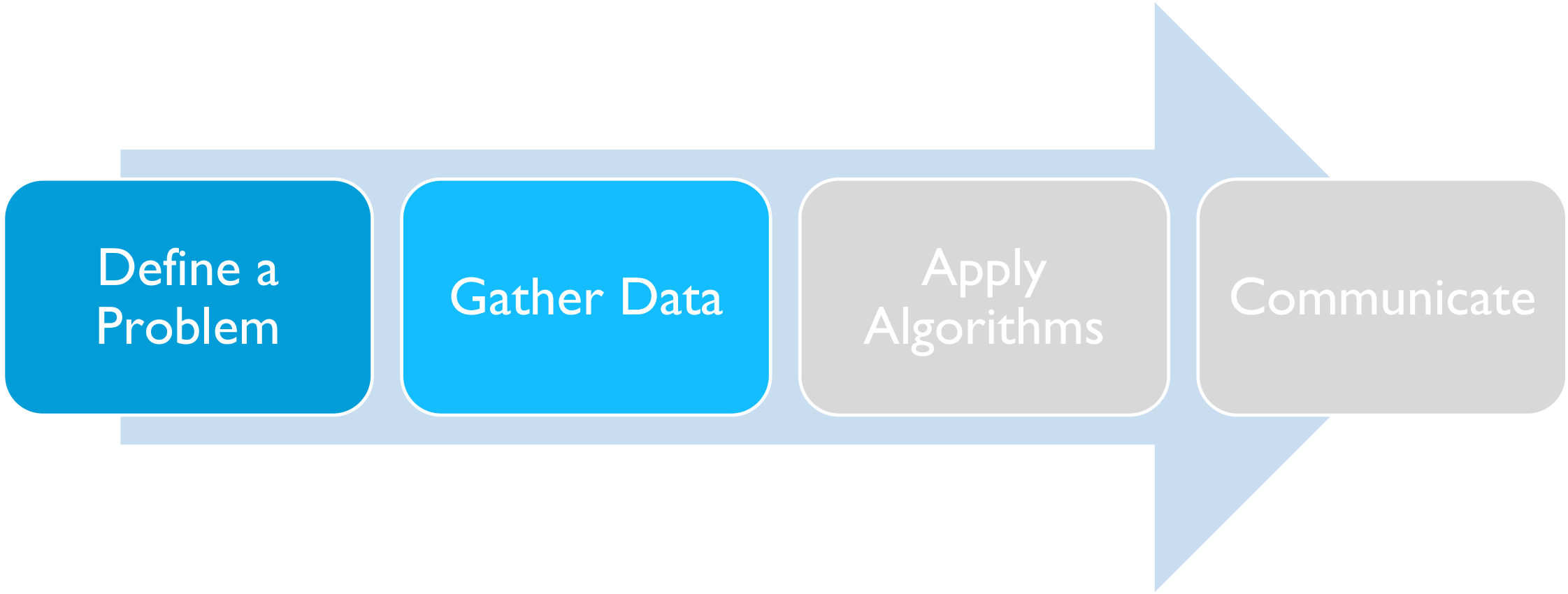


https://xkcd.com/1831/

"We TOLD you it was hard."
"Yeah, but now that I'VE tried, we KNOW it's hard."

midwest architecture community collaboration

A SIMPLIFIED DATA SCIENCE PROCESS



midwest architecture community collaboration

CORE EXAMPLES (TO BE DISCUSSED)

Google Maps Arrival Prediction

- Virtually unlimited data set
- Largely an interesting public service
- Usually over small distances
- Precision is important

UPS Package Delivery

- Data set limited to delivery machinery
- Lots of money at stake
- Virtually unlimited permutations
- Accuracy is critical precision less so

Is a bigger *n* greater in value?

HOW TO NOT USE DATA IN A SOLUTION

Swipe Text Solution

- Rely on existing hardware and keyboards
- Build a new user interaction model
- Apply a mathematical curve fit
- Find the closest fit match*
- Add curves as users accept suggestions

**From a database of all curves*

Results

- Users heavily engage with the new design
- New needs were uncovered, and the feature undergoes regular updates
- **Now every misspelling anyone else has ever swiped appears as an option while common words are missing**

midwest architecture community collaboration

ADDING A LITTLE ARCHITECTURE

Findings (as a user)

- *Big n is not necessarily greater than small n*
- *A better "trained" model can produce less accurate results with greater precision*

Guidelines to Improve Accuracy

- **Adding a "most likely" suggestion with a different data set**
- **Check the most common 1000 English words first**

BACK TO OUR EXAMPLES

Google Maps Arrival Prediction

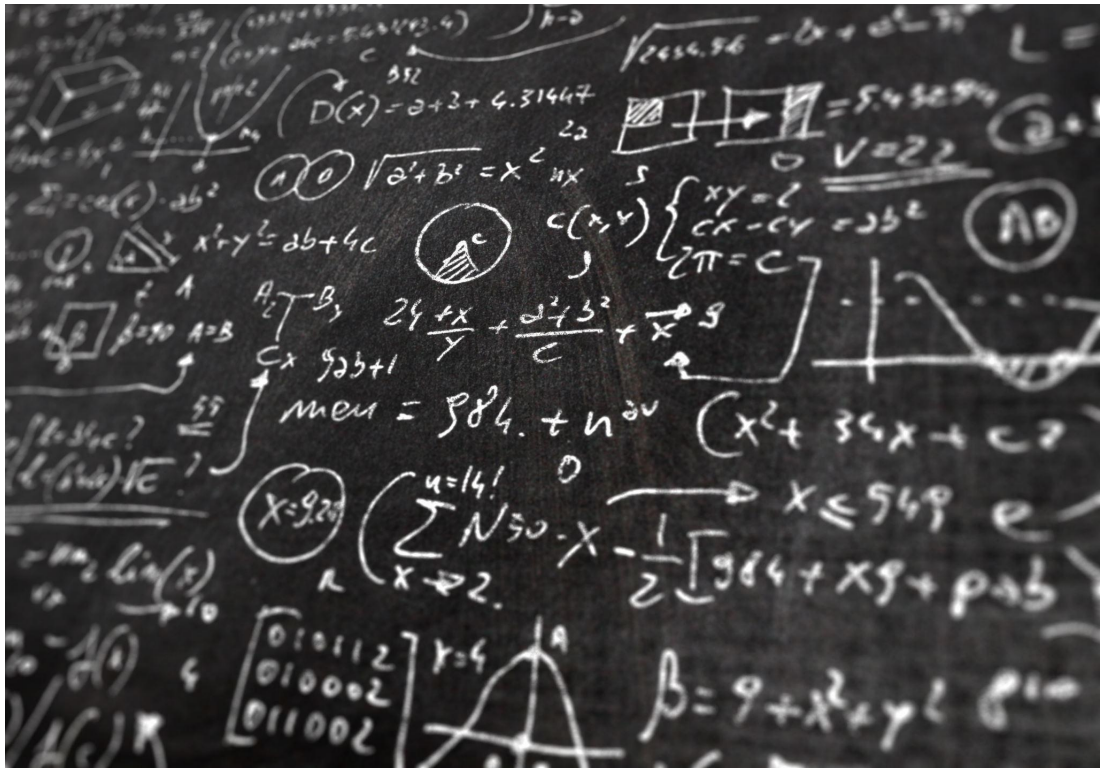
- Map searches include international destinations that suggest walking across the ocean
- More seriously: inclusion of side roads without taking condition into account led to tragedy
- Once the novelty wore off, users were less likely to engage- **rate the parking, disable location tracking**

UPS Package Delivery

- Very specific feedback loop enables accuracy to improve without *Data*
- Improved quality leads to specific partnerships producing better outcomes
- Time is money- **No more signing for packages**

HOW TO EVALUATE A SOLUTION OUTCOME

- Education technology is littered with *well designed, well intentioned* failure
- Roots of many of these failures trace back to better defining the scope of the possible
- This does not mean less ambitious goals! Just possibly less *technical* solutions



midwest architecture community collaboration

IMPROVING STUDENT DEGREE COMPLETION

Course Sherpa

- Promised to monitor trends in student enrollment and completion
- Result: informed the school that key classes were being offered in the wrong season
- Convincing professors to offer the courses in the Fall increased student completion by double digit percentages and reduce time to completion by almost a year

Knewton

- **Promised to precisely determine the knowledge gaps that lead to a student failing to complete higher level coursework and prescribe remediation based on a directed knowledge graph**
- **Recently sold for scrap**

midwest architecture community collaboration

BUT WHY DID IT FAIL?

Specific Reasons

- Illustrations were powerful but rare
- Relied on users to enter data with no other purpose
- Educating someone is difficult

Generalization

For data, there is a hierarchy of reliability that is orthogonal to durability and importance of results

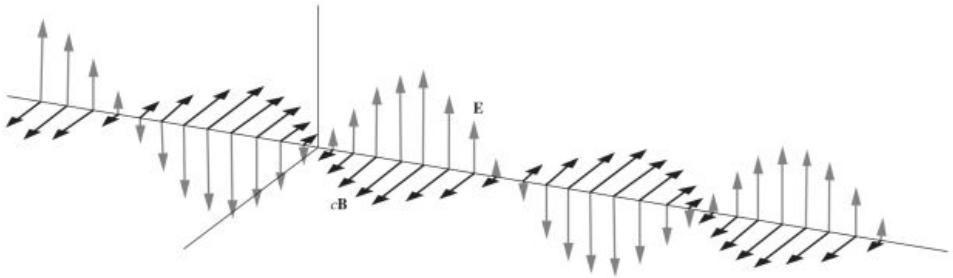
(jumping ahead due to availability of tools and data)

QUALIFYING THE ATTRIBUTES

How fluid is the measure?

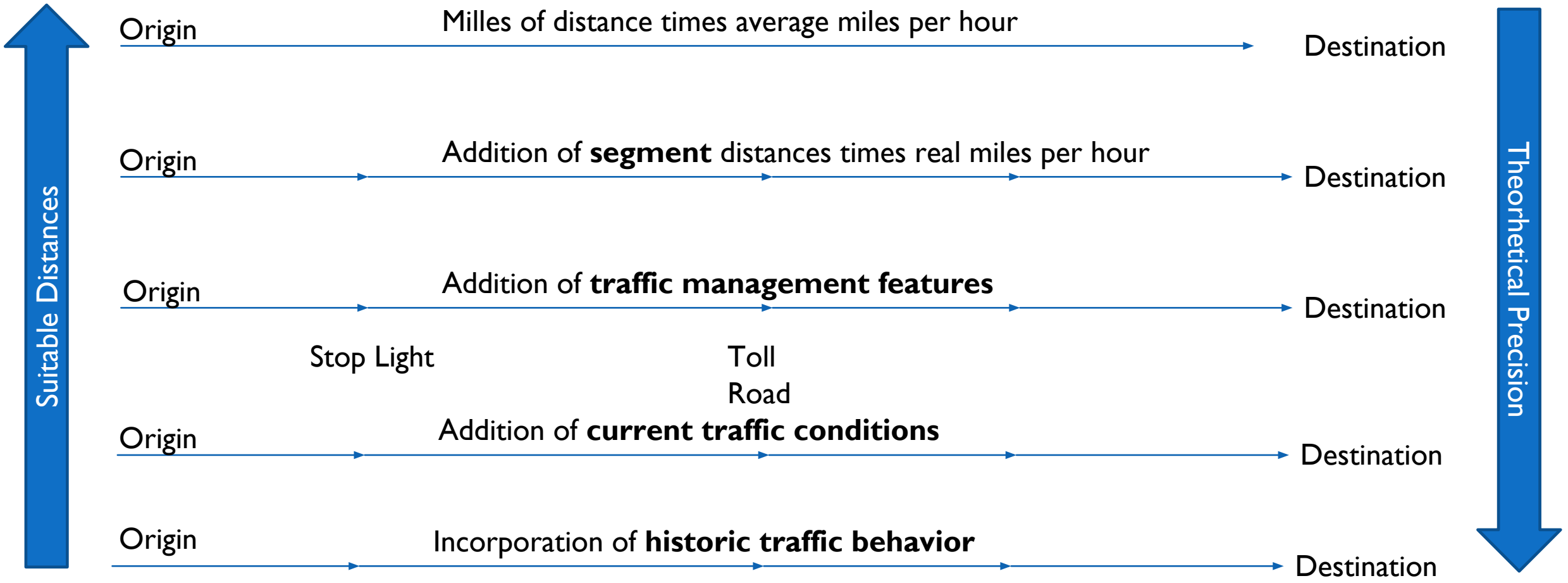
How influential are outside factors?

How universal is the value set?



midwest architecture community collaboration

UCCESSIVE APPROXIMATIONS



midwest architecture community collaboration

BACK TO OUR CORE EXAMPLES

Google Maps Arrival Prediction

- Validity of prediction can change instantly
- Value of the prediction varies widely

UPS Package Delivery

TYING BACK TO RESILIENCE

- **Focus on defining the fastest path to a result that is valuable, repeatable, and extensible**
- The real world will progress despite the models
- Massive tools and data sets have created revolutionary opportunities
- Understanding the basic domain constraints should shape a better solution
- The architecture principle of modularity still rules

ANY QUESTIONS?

midwest architecture community collaboration

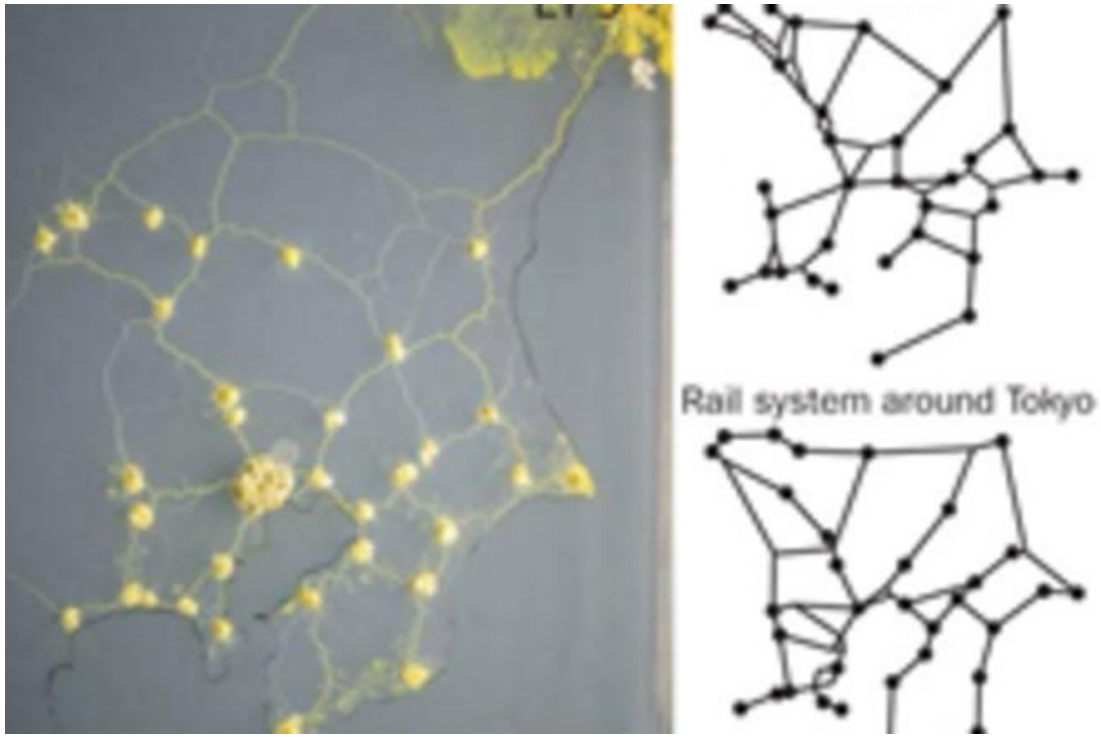
APPENDIX

SOME ADDITIONAL ANALOGIES

midwest architecture community collaboration

SLIME MOLD IMITATES TOKYO SUBWAY

- Given a set of food sources with distances matching key locations in Tokyo, slime mold efficiency results in a matching map



midwest architecture community collaboration

MODELS ARE PART OF AN ECOSYSTEM

Input	Model	Output
How easily can data be obtained?	Can variations and gaps be mitigated?	What actions may be taken based on this data?
How matched in the input to the output?	How well understood are the dimensions?	How do I communicate actual findings well?
How frequently is there a material change?	Can the results be appropriately qualified?	How could this be used as a tool for guidance?

Will this need to hold up in court?

midwest architecture community collaboration

REPRODUCIBILITY CRISIS

midwest architecture community collaboration