ELDER RESEARCH
DATA SCIENCE • MACHINE LEARNING • AI

Leading a Successful Data Science Initiative

Gerhard Pilcher

Elder Research, INC.

Headquarters: 300 West Main Street, Suite 301
Charlottesville, VA 22903
434.973.7673 | Fax: 434.973.7673
OTHER LOCATIONS IN Arlington, VA | Linthicum, MD | Raleigh, NC

Before we get started…

Questions

- Please use the questions tab to enter questions throughout the webinar (reference slide number if applicable)
- Questions will be answered during the Q&A session at the end

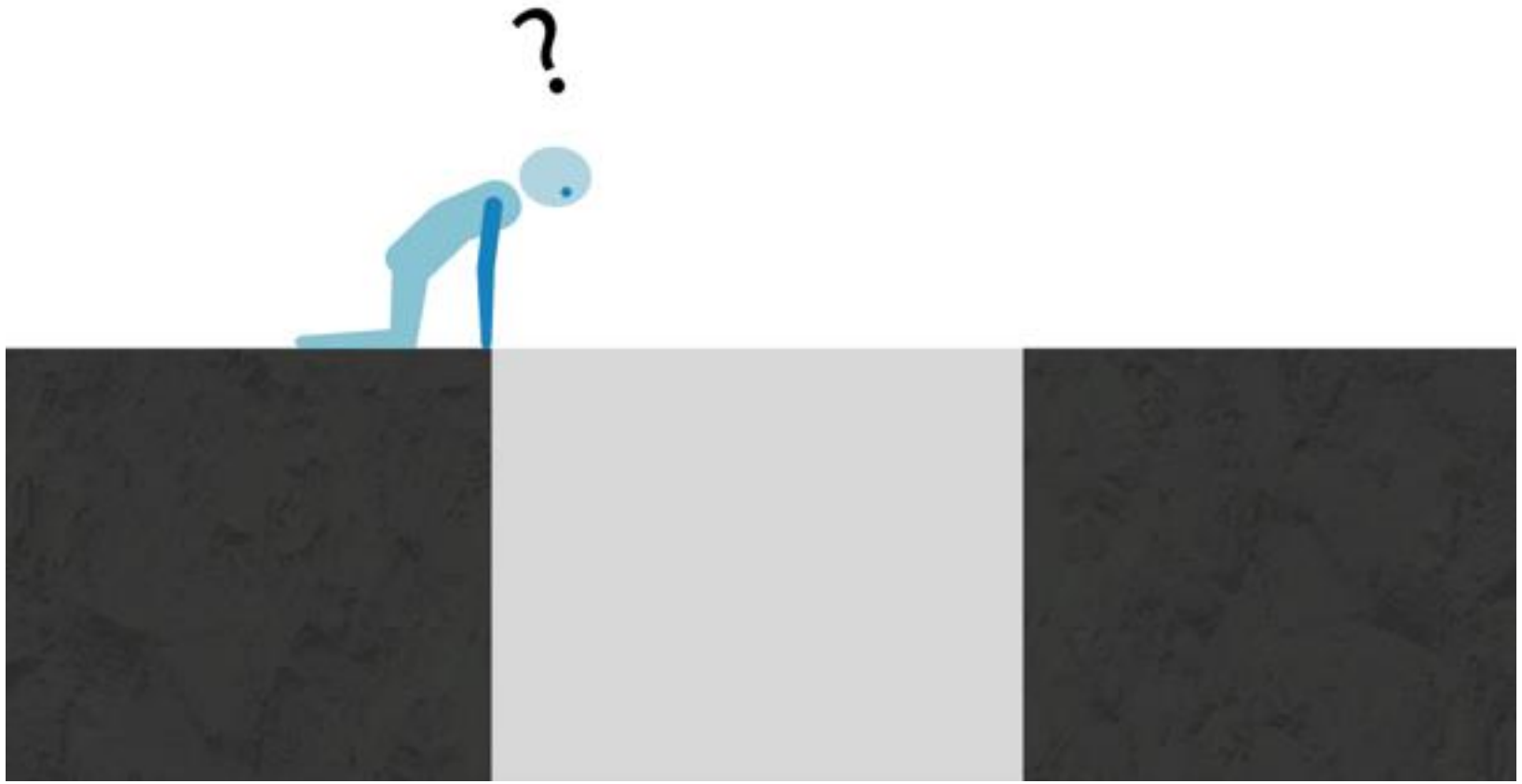
Slides

- Slides will be posted on the webinar website next week or email paul.derstine@elderresearch.com to request a copy

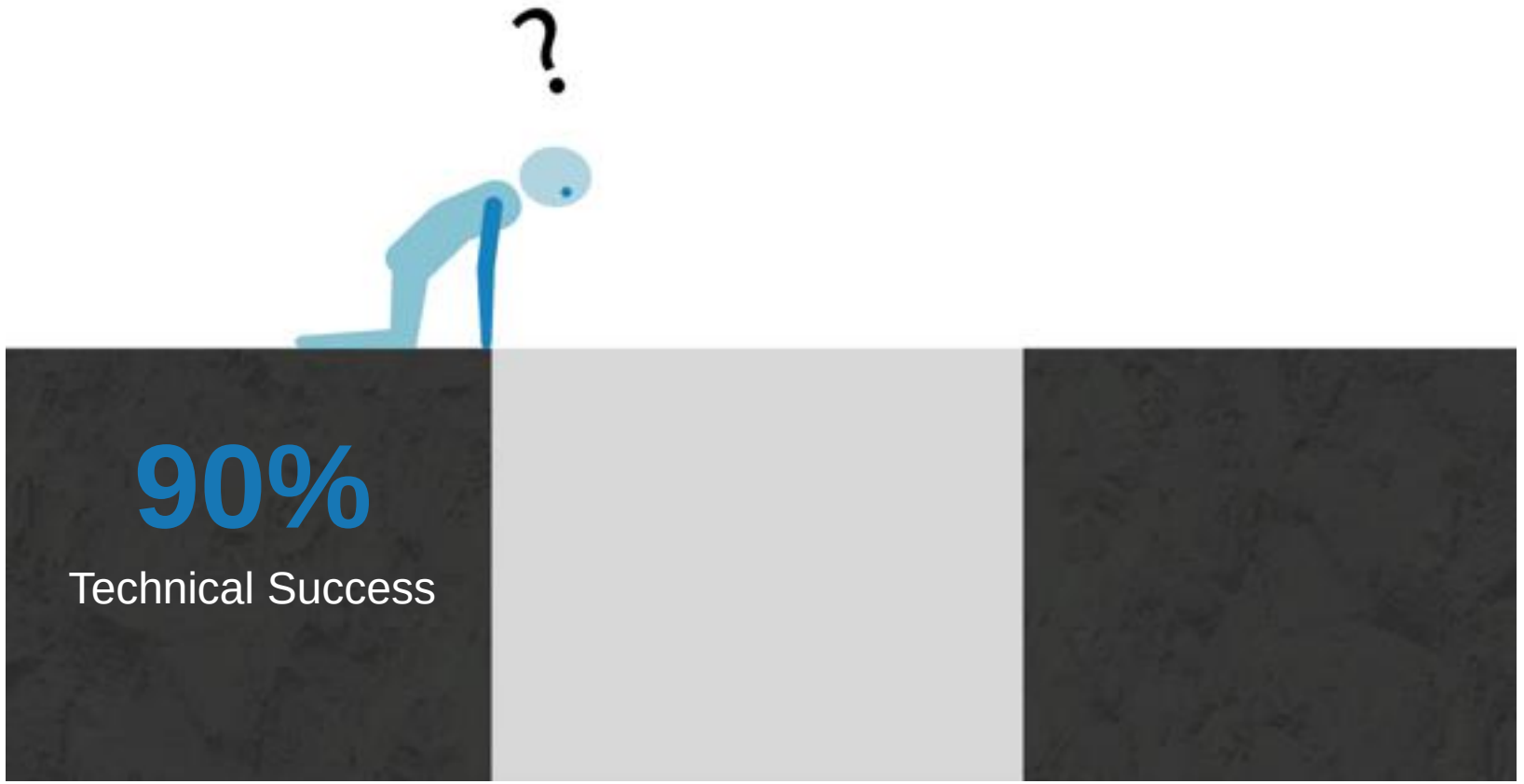
Problem during the webinar?

- Simply refresh your browser screen

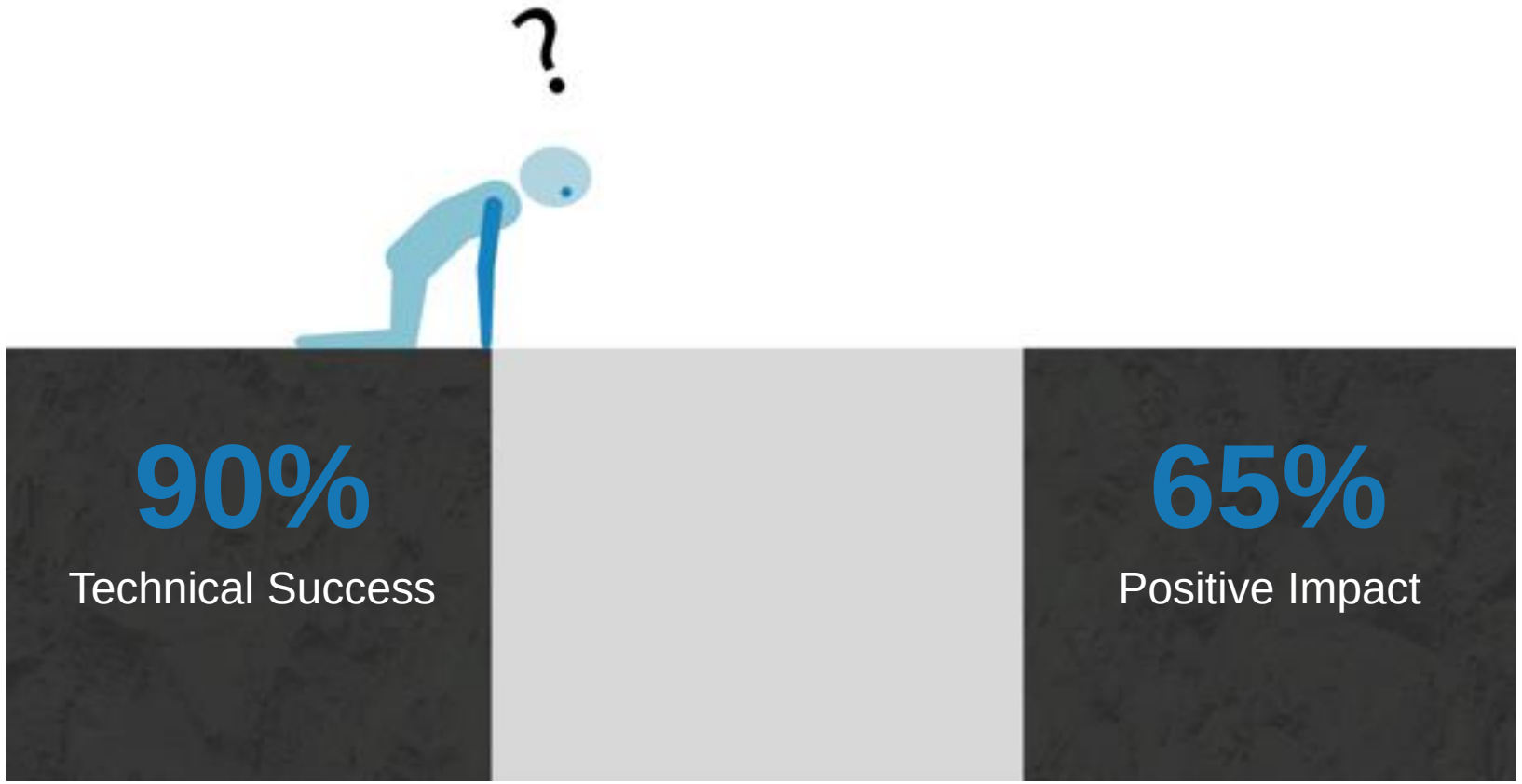
The Success Gap



The Success Gap



The Success Gap



Data Science, AI, and Machine Learning

DATA SCIENCE

https://www.techopedia.com/definition/30202/data-science

Data Science is a broad field that refers to the collective processes, theories, concepts, tools, and technologies that enable the review, analysis, and extraction of valuable knowledge and information from raw data.

AI

https://www.forbes.com/.../the-key-definitions-of-artificial-intelligence-ai-that-explain...

Feb 14, 2018 - Merriam-Webster defines **artificial intelligence** this way: A branch of computer science dealing with the simulation of intelligent behavior in computers. The capability of a machine to imitate intelligent human behavior.

MACHINE LEARNING

https://www.techemergence.com/what-is-machine-learning/

“Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.”

What is a Model?

$$E_i = \bar{g}_{Na} m_i^3 h_i E_{Na} + \bar{g}_K n_i^4 E_K + \bar{g}_L E_L + \sum_{j \neq i} G_{synapse[j,i]} E_{j,i}$$

- Techniques help build a model
- A model is the resulting set of equations

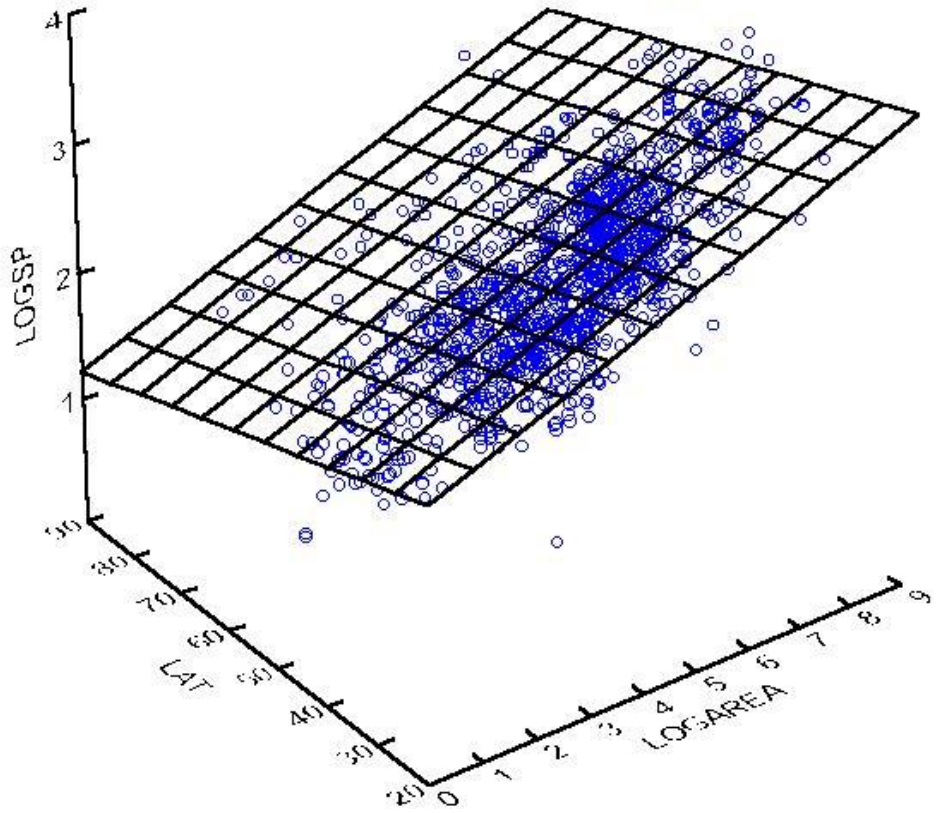
$$C \frac{\partial V_i}{\partial t} + g_i (V_i - E_i) = I_i^{applied}(t)$$

$$M_{12} = M_{21} = I_{Z_2} + m_2 \left(\frac{l_2^2}{4} + \frac{1}{2} l_1 l_2 c_2 \right), M_{22} = I_{Z_2} + m_2 \frac{l_2^2}{4}$$

$$M_{11} = I_{Z_1} + I_{Z_2} + m_1 \frac{l_1^2}{4} + m_2 \left(l_1^2 + \frac{l_2^2}{4} + l_1 l_2 c_2 \right)$$

$$M(q) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

$$y' = (m_1 * x_1) + (m_2 \circ * x_2) + b$$



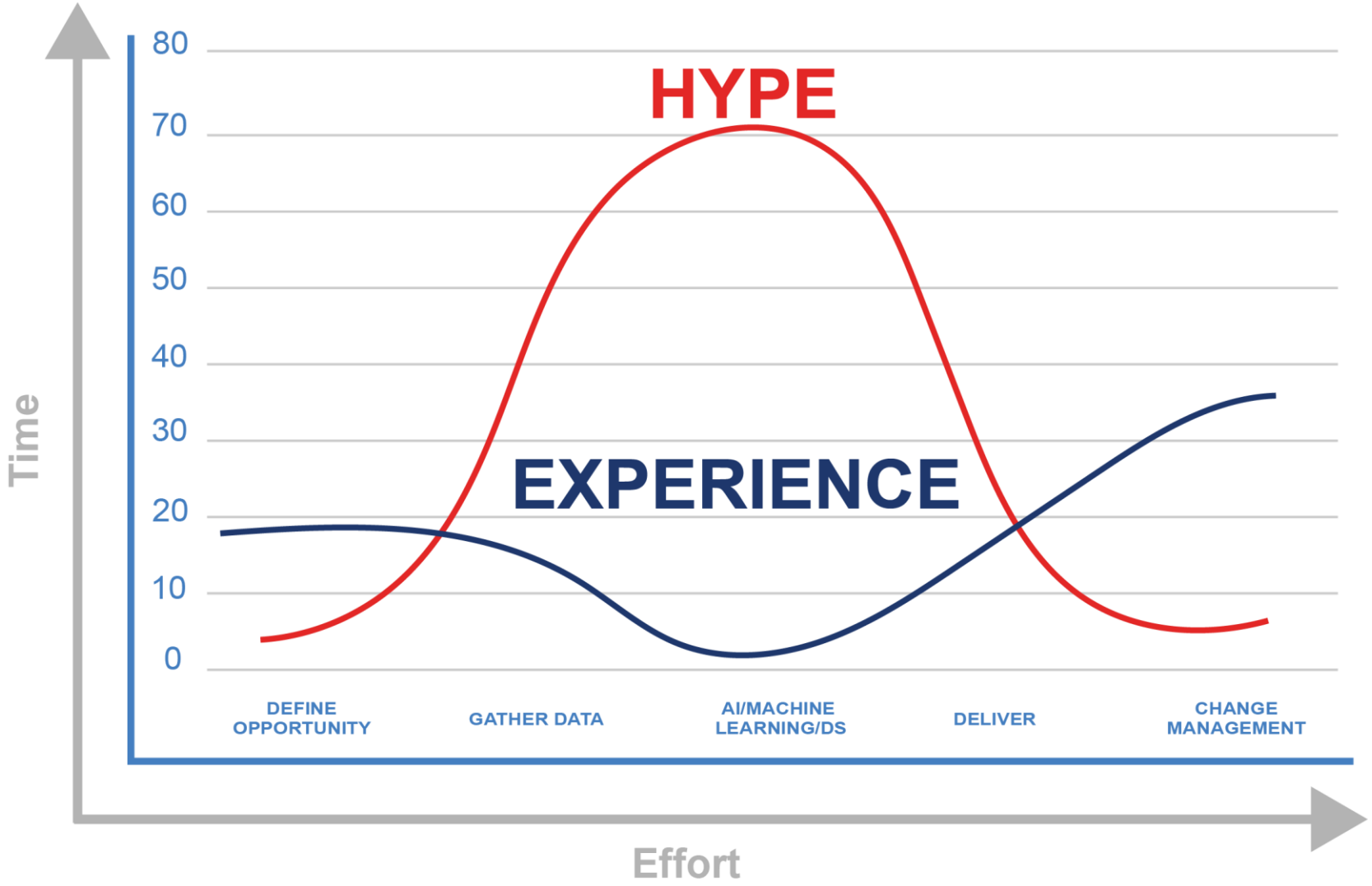
$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = u + D(t)$$

$$\frac{\partial m_i}{\partial t} = \alpha_m (V_i)(1 - m_i) - \beta_m (V_i) m_i$$

Leadership

Business Focus

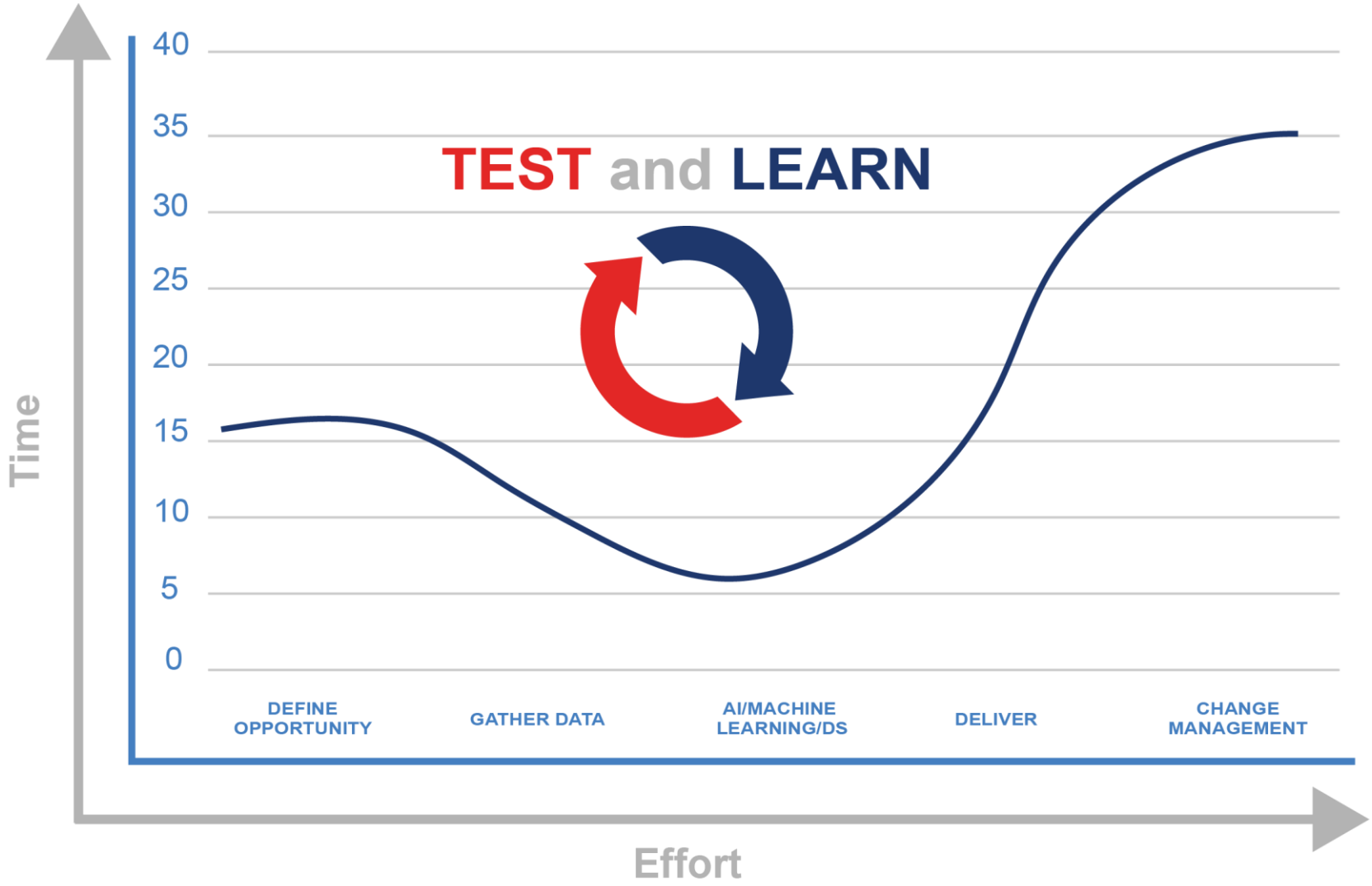
Business Focus



Bias Alert

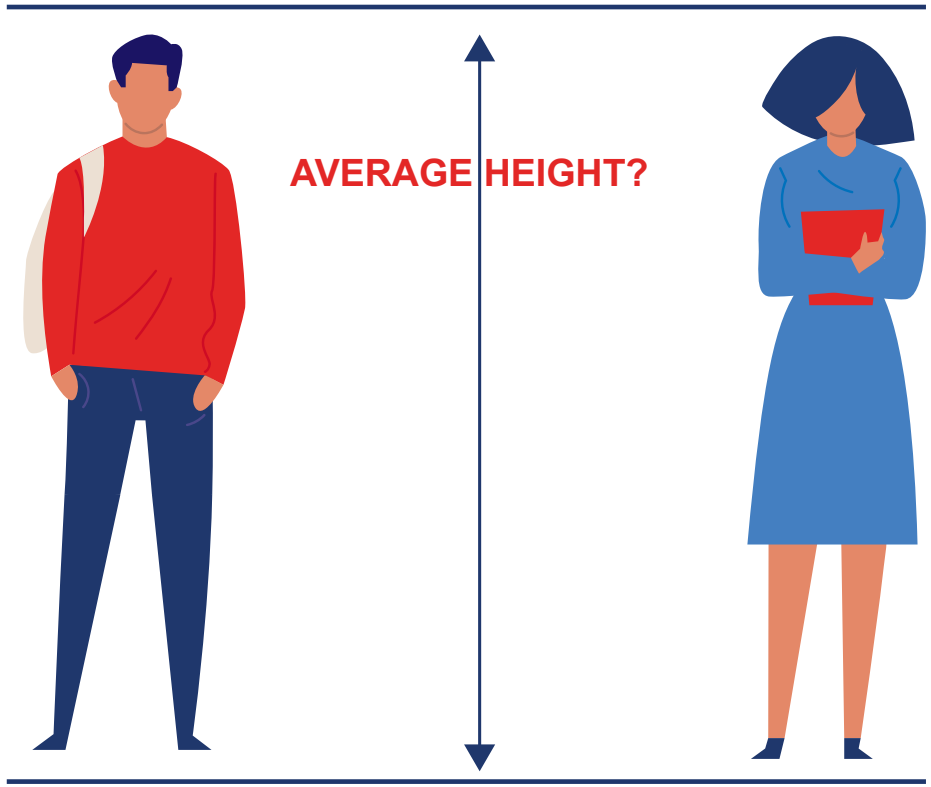
Leadership

Iterative Process



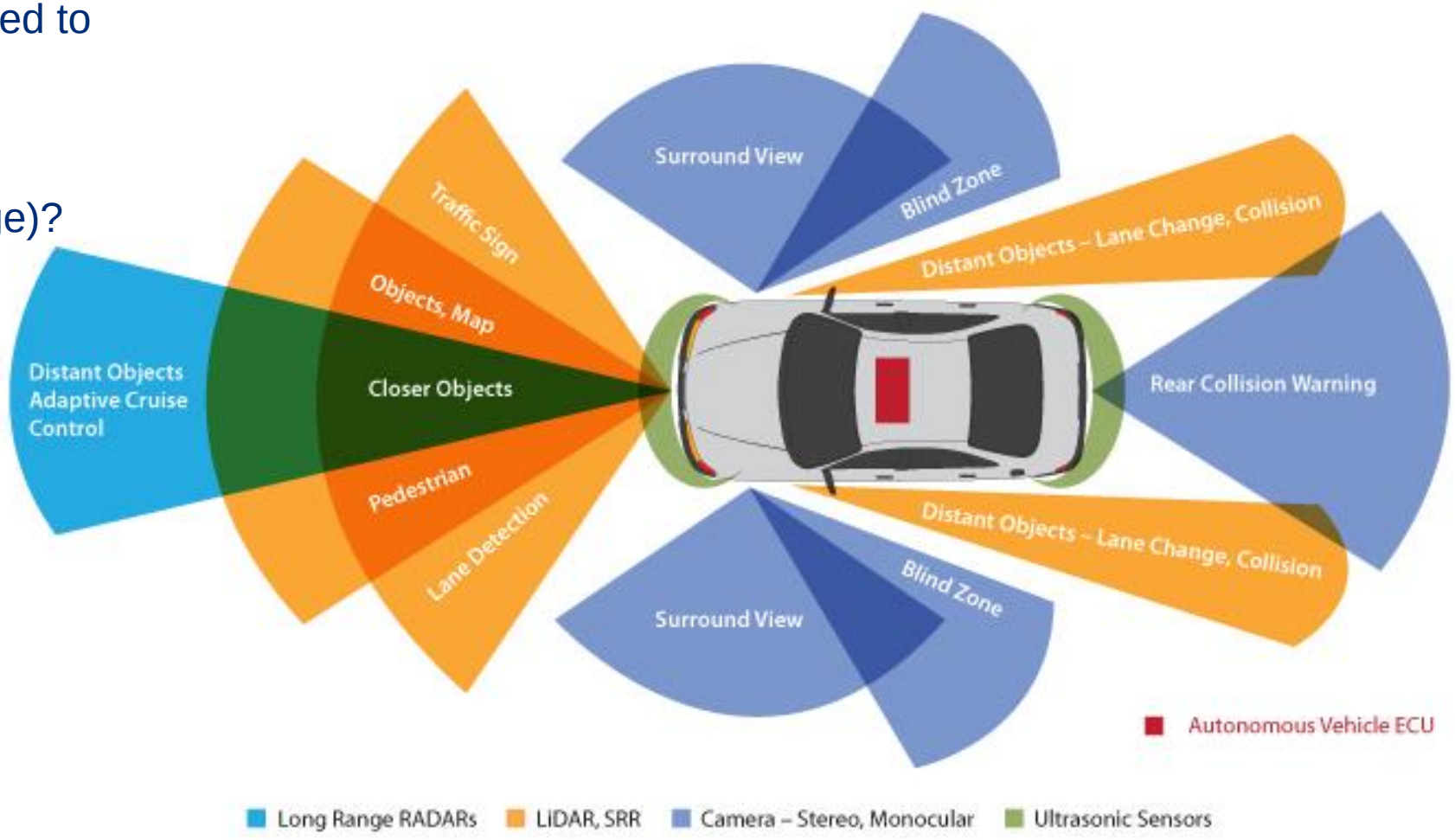
Baseline

Need a way to measure progress



Autonomous Vehicles

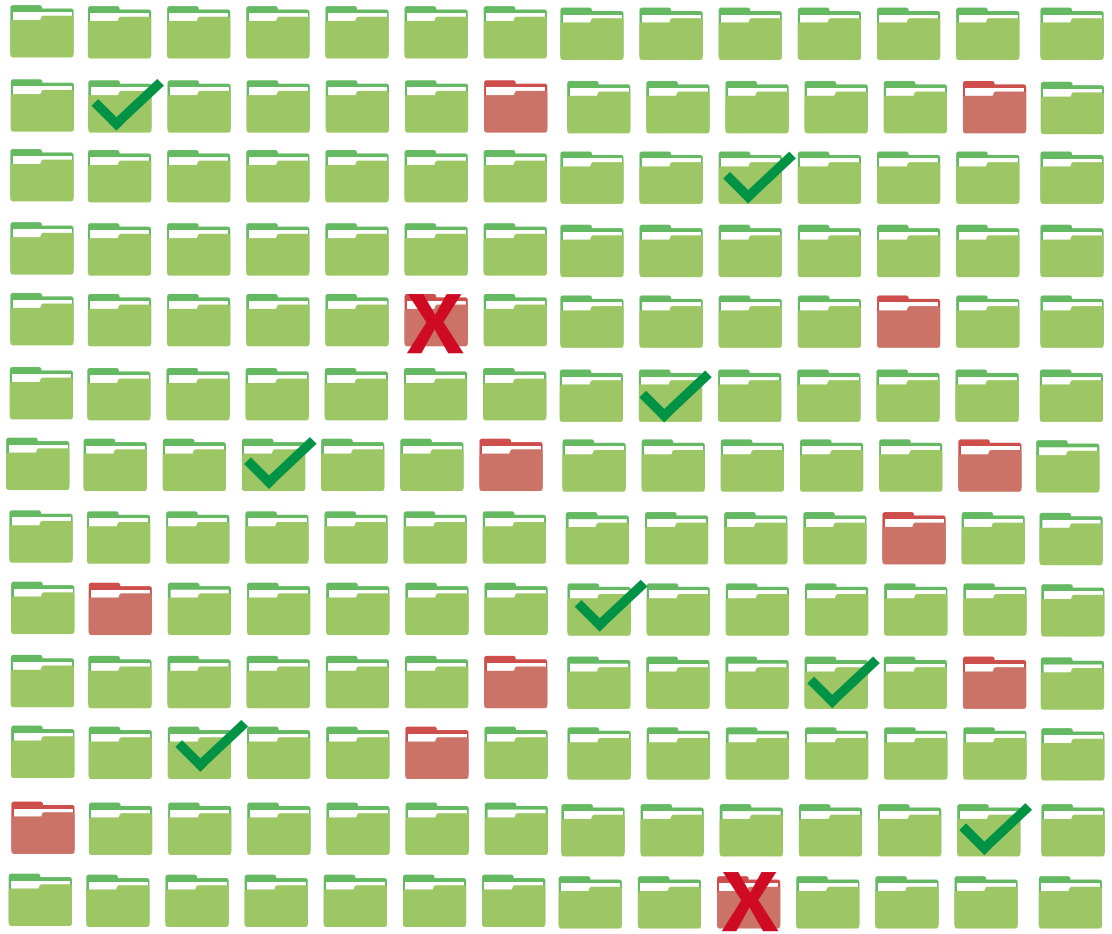
- Improving (compared to baseline)?
- Fair (non-discriminatory)?
- Adequate (coverage)?
- Proactive (new schemes)?



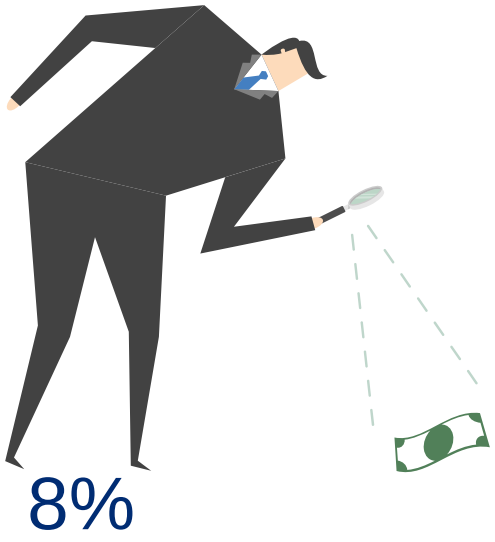
NXP Semiconductors

Baseline – Random Selection

0.01%

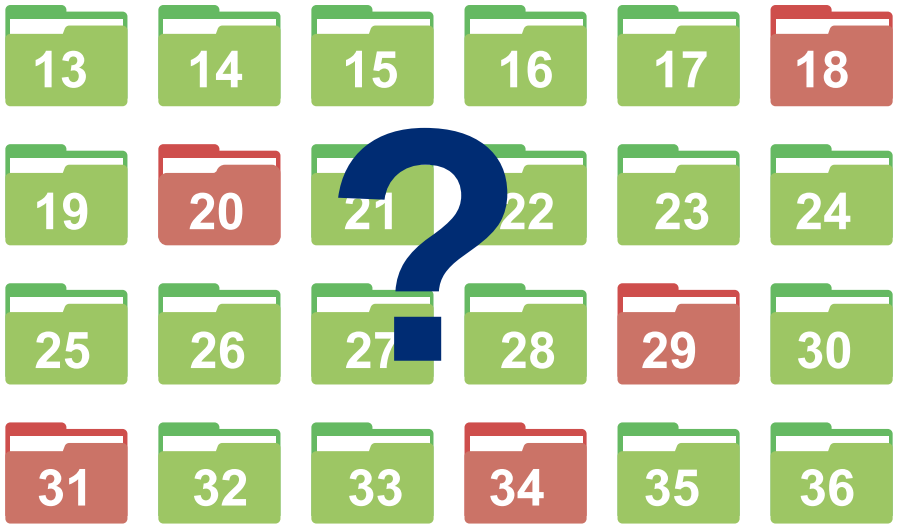


Period 1 – *Human Rules*



Selection Bias

THRESHOLD



Low Ranking Cases
NOT REVIEWED

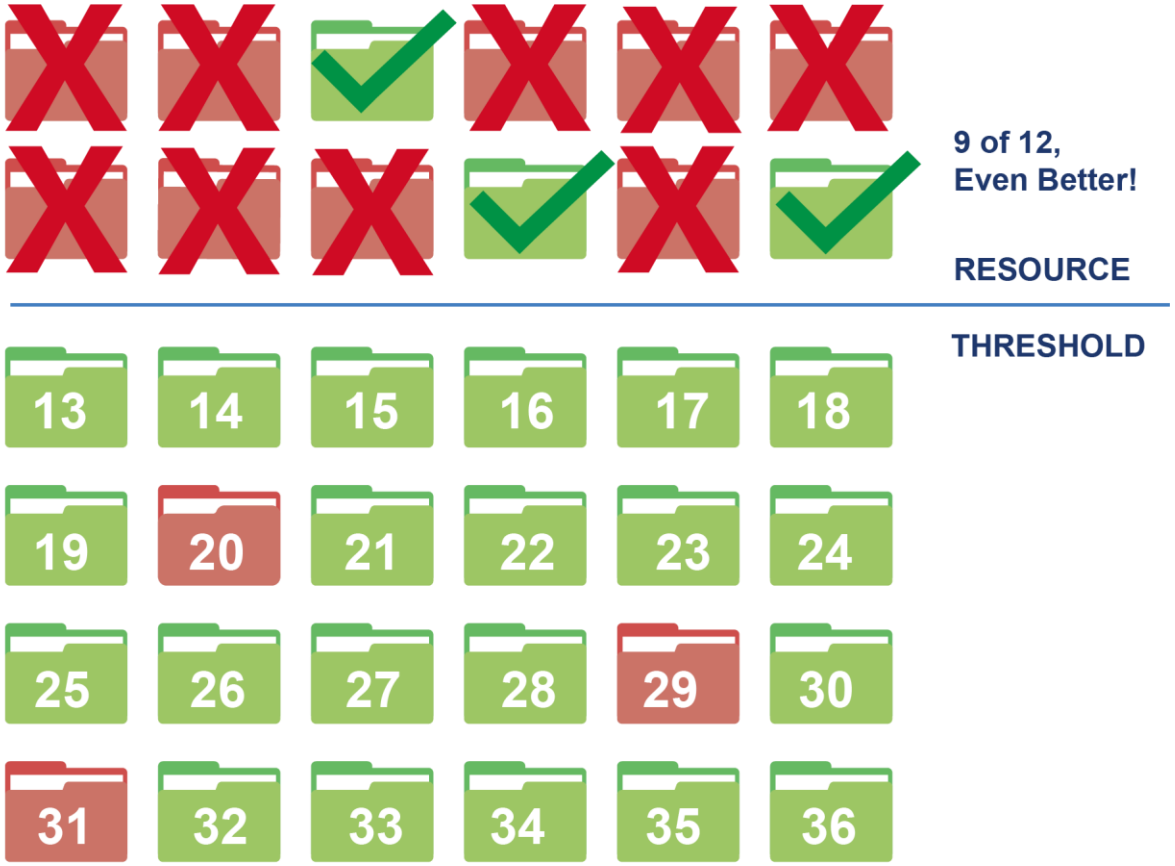
Assumed Negative
But Unknown

Period 2 – *Data Driven Rules*

58%

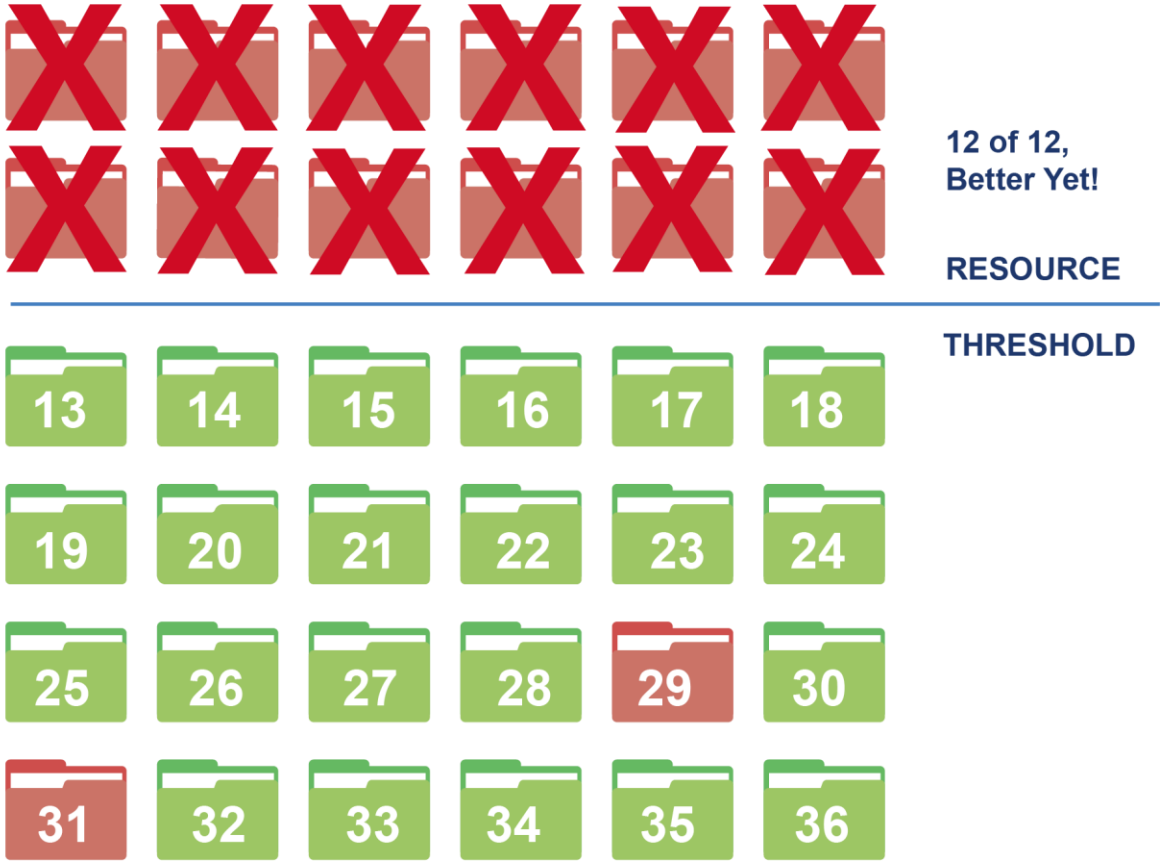
Period 3 – *Rebuild Model With Period 2 Data*

75%

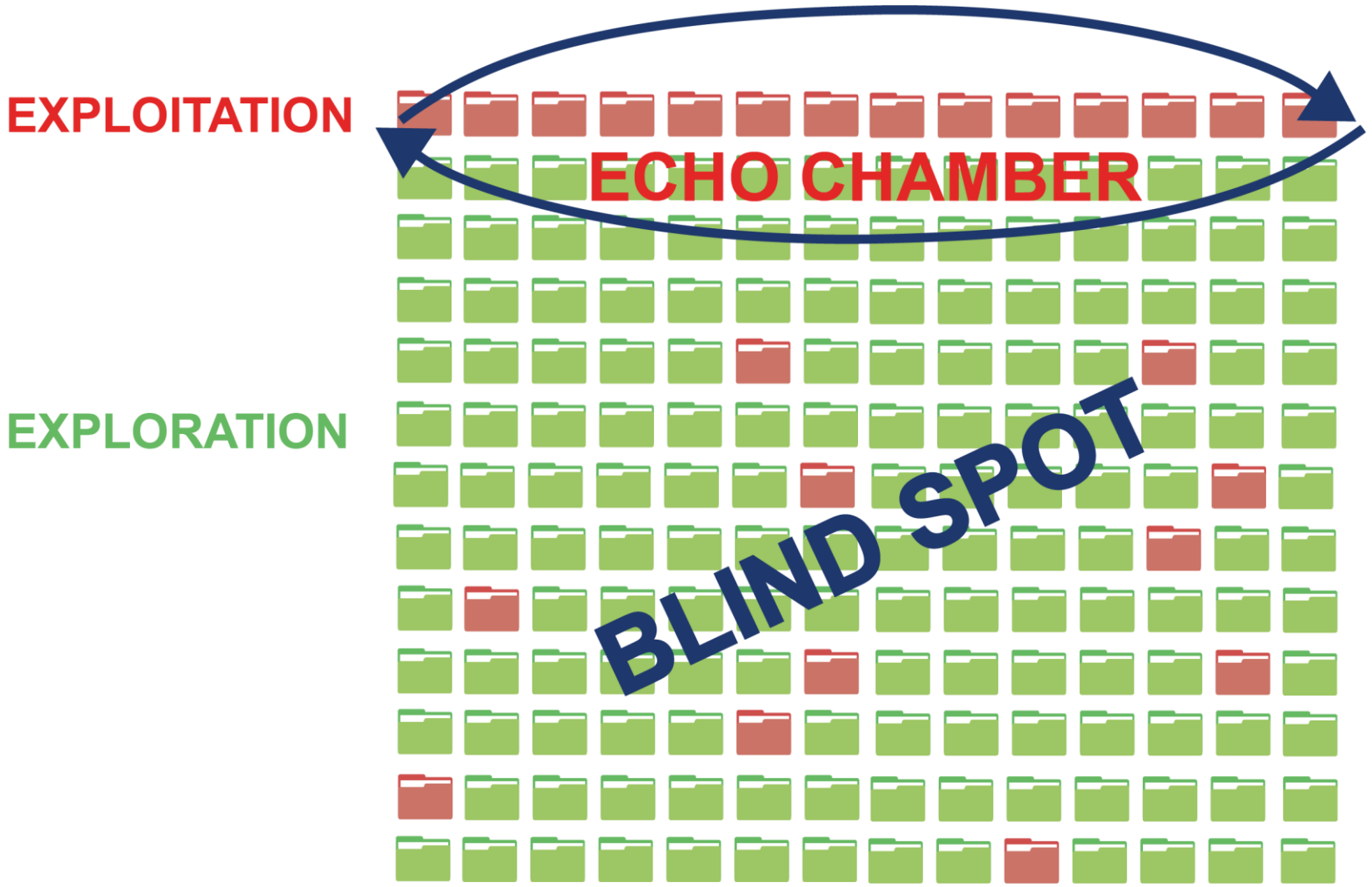


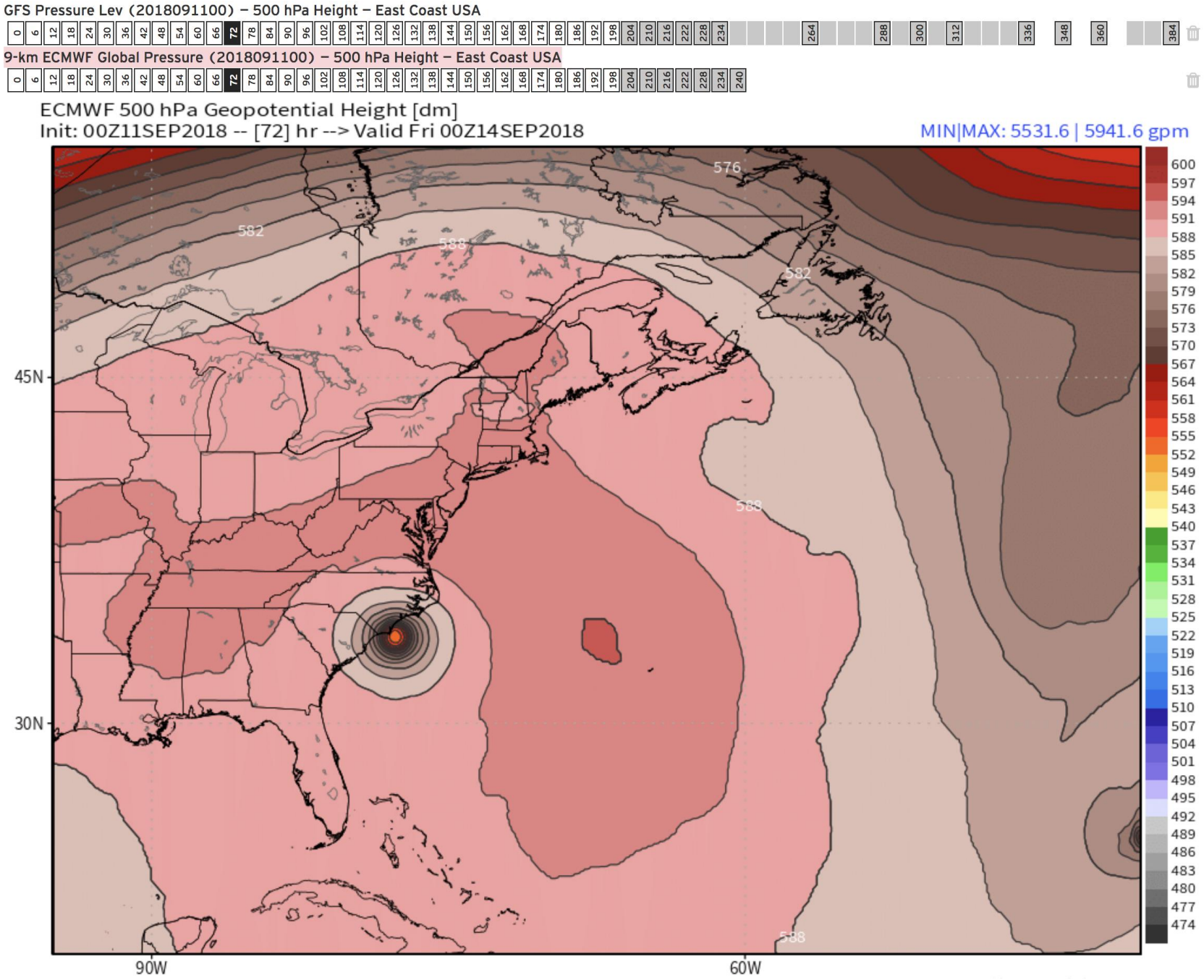
Period 4 – *Rebuild Model With Period 3 Data*

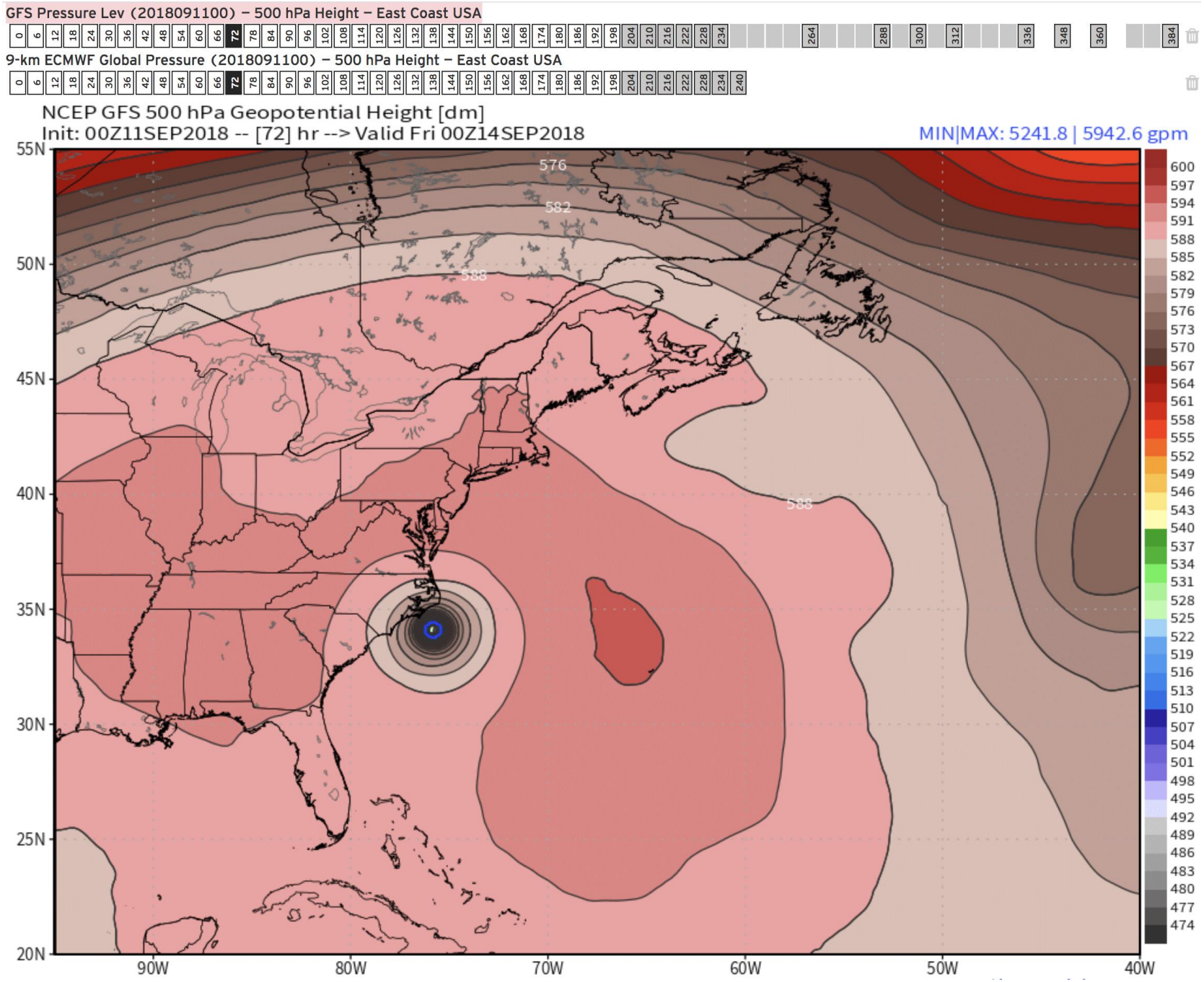
100%



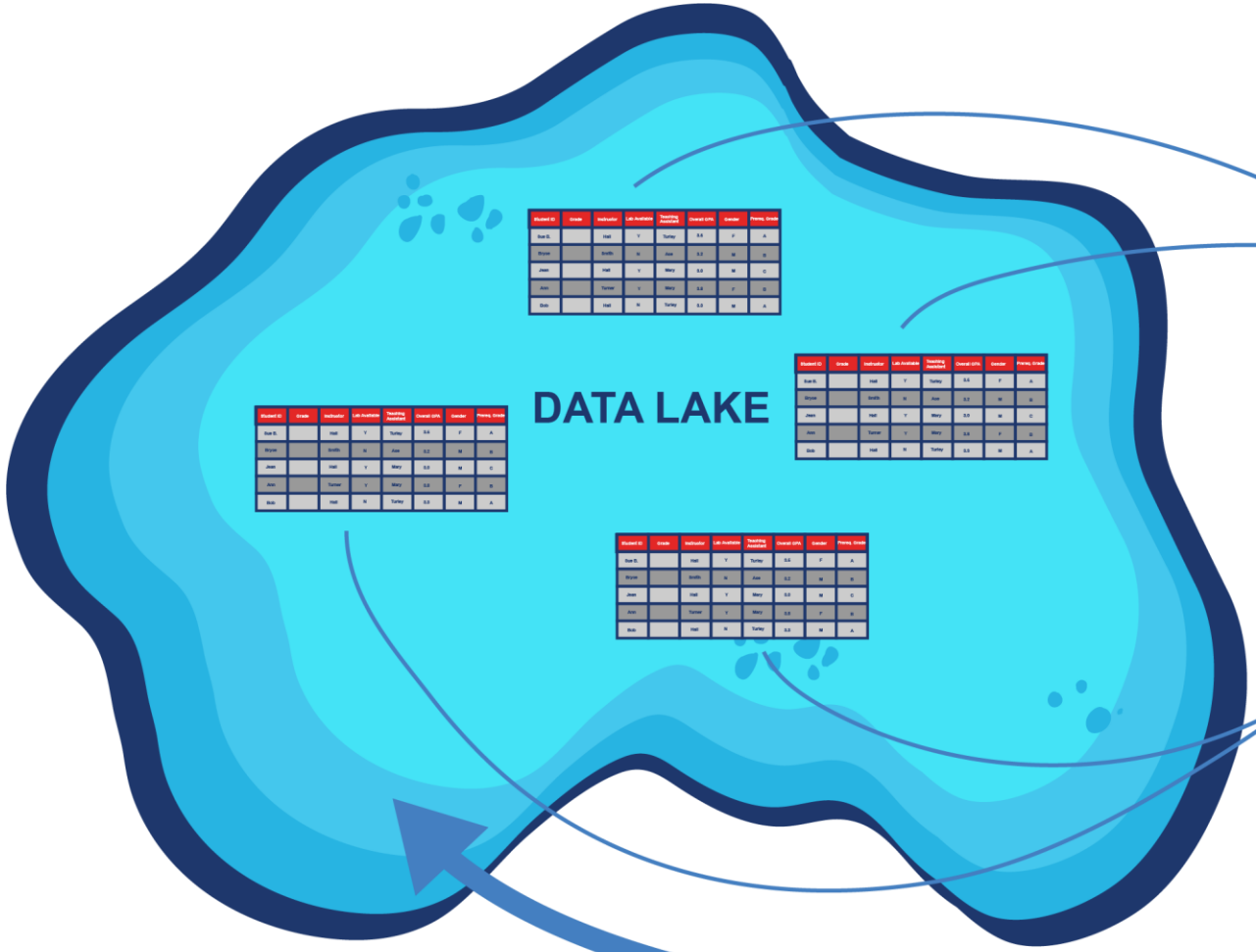
Reinforcement Bias




GFS Pressure Lev (2018091100) – 500 hPa Height – East Coast USA
9-km ECMWF Global Pressure (2018091100) – 500 hPa Height – East Coast USA
ECMWF 500 hPa Geopotential Height [dm]
Init: 00Z11SEP2018 -- [72] hr --> Valid Fri 00Z14SEP2018
MIN|MAX: 5531.6 | 5941.6 gpm
45N
30N
90W
60W
576
582
588
600
597
594
591
588
585
582
579
576
573
570
567
564
561
558
555
552
549
546
543
540
537
534
531
528
525
522
519
516
513
510
507
504
501
498
495
492
489
486
483
480
477
474


GFS Pressure Lev (2018091100) – 500 hPa Height – East Coast USA
9-km ECMWF Global Pressure (2018091100) – 500 hPa Height – East Coast USA
NCEP GFS 500 hPa Geopotential Height [dm]
Init: 00Z11SEP2018 -- [72] hr --> Valid Fri 00Z14SEP2018
MIN|MAX: 5241.8 | 5942.6 gpm
576
582
588
588
55N
50N
45N
40N
35N
30N
25N
20N
90W
80W
70W
60W
50W
40W
600
597
594
591
588
585
582
579
576
573
570
567
564
561
558
555
552
549
546
543
540
537
534
531
528
525
522
519
516
513
510
507
504
501
498
495
492
489
486
483
480
477
474

Capturing Data Value

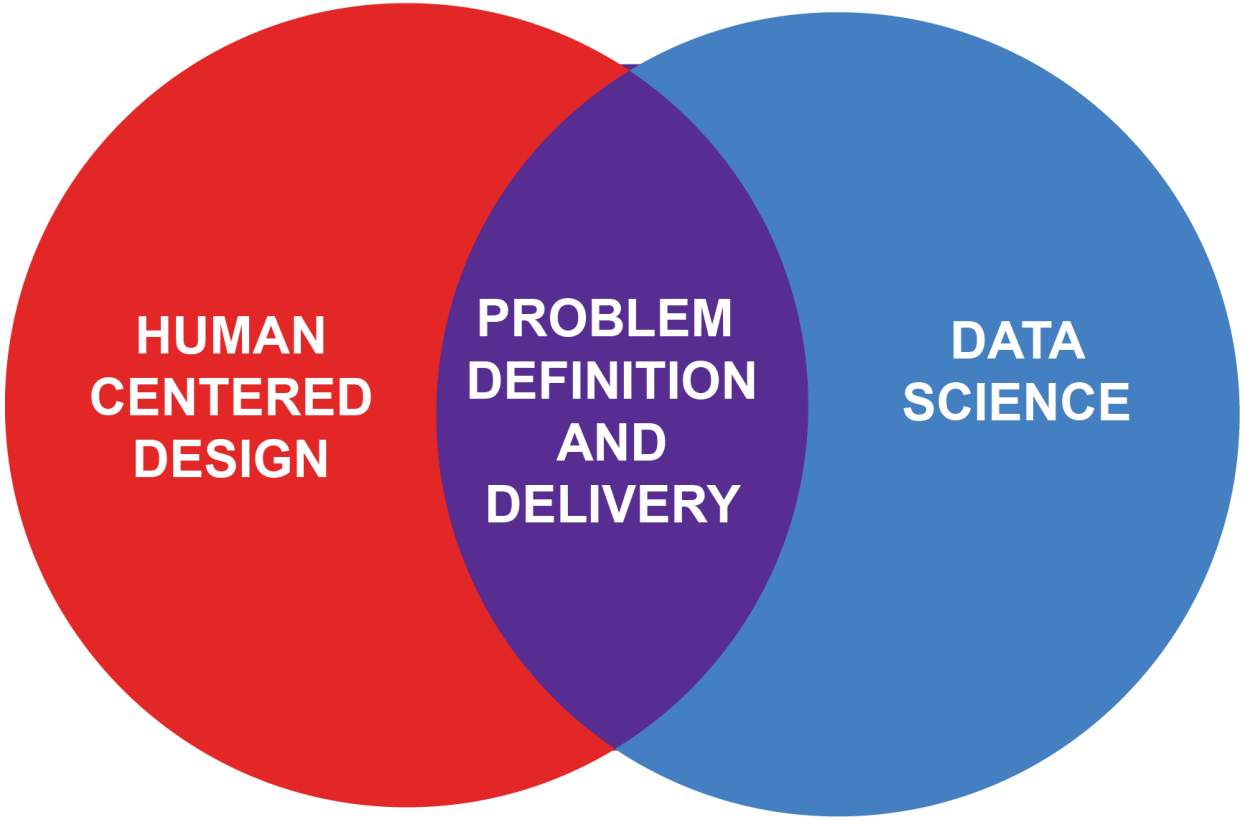


Student ID	Grade	Instructor	Lab Available	Teaching Assistant	Overall GPA	Gender	Prereq. Grade
Sue B.		Hall	Y	Turley	3.5	F	A
Bryce		Smith	N	Ace	3.2	M	B
Jean		Hall	Y	Mary	3.0	M	C
Ann		Turner	Y	Mary	3.8	F	B
Bob		Hall	N	Turley	3.3	M	A

Leadership

The ”Who”

Where Human Centered Design and Science Collide

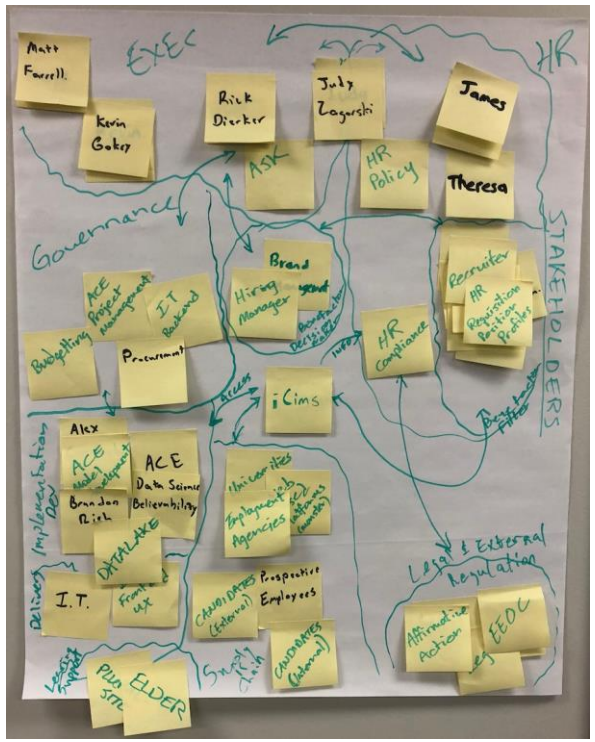


Gleicher's Formula for Change

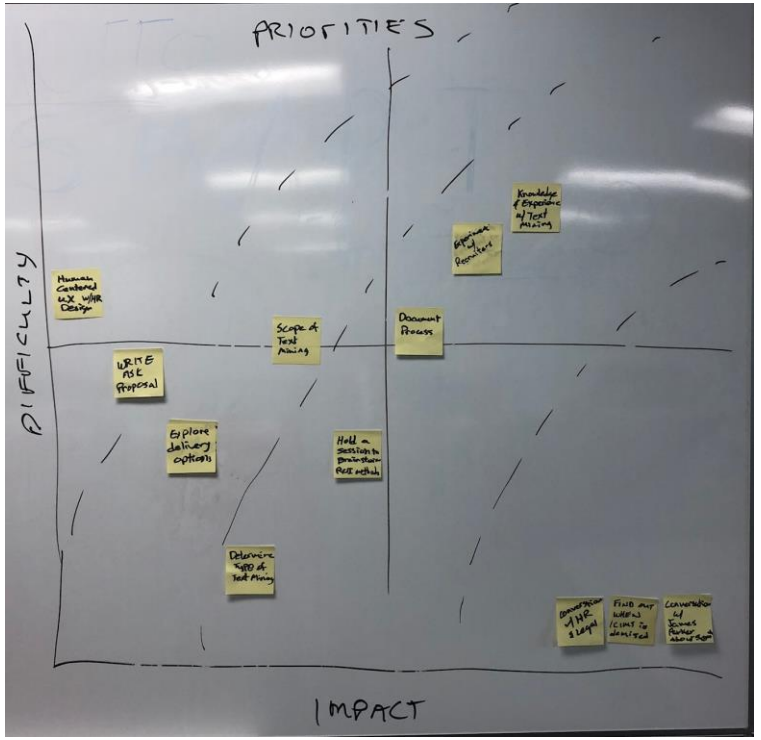
Dissatisfaction x Vision x First Steps > Resistance

- Dissatisfaction with current situation = business opportunity (Unsure which part of market and trade spend is effective)
- Vision = Data Science (Set of models to optimize market mix)
- First Steps = How to deliver the solution to decision makers (information needs of trade allocation are different than needs of digital marketing)

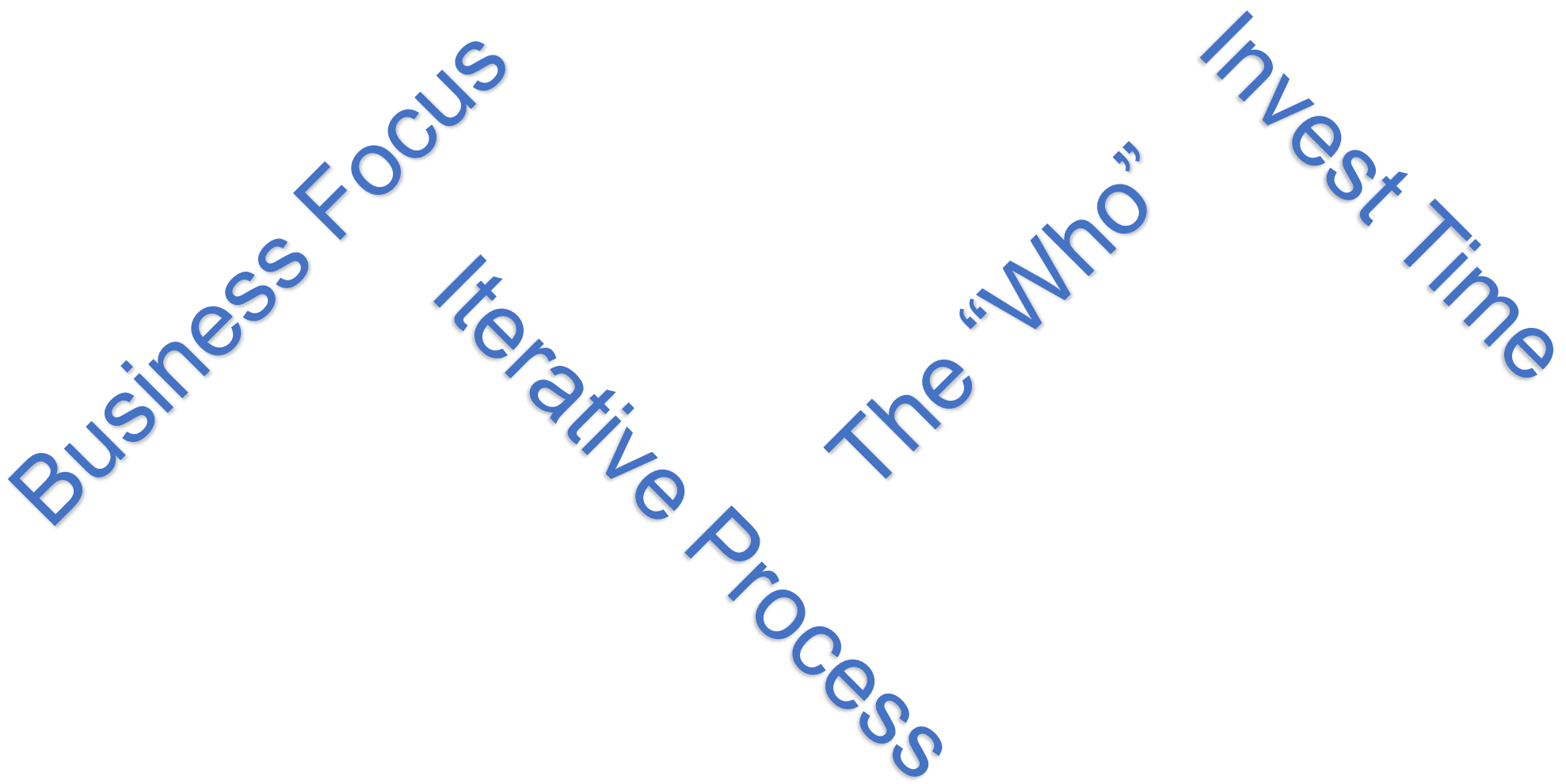
Human Centered Design



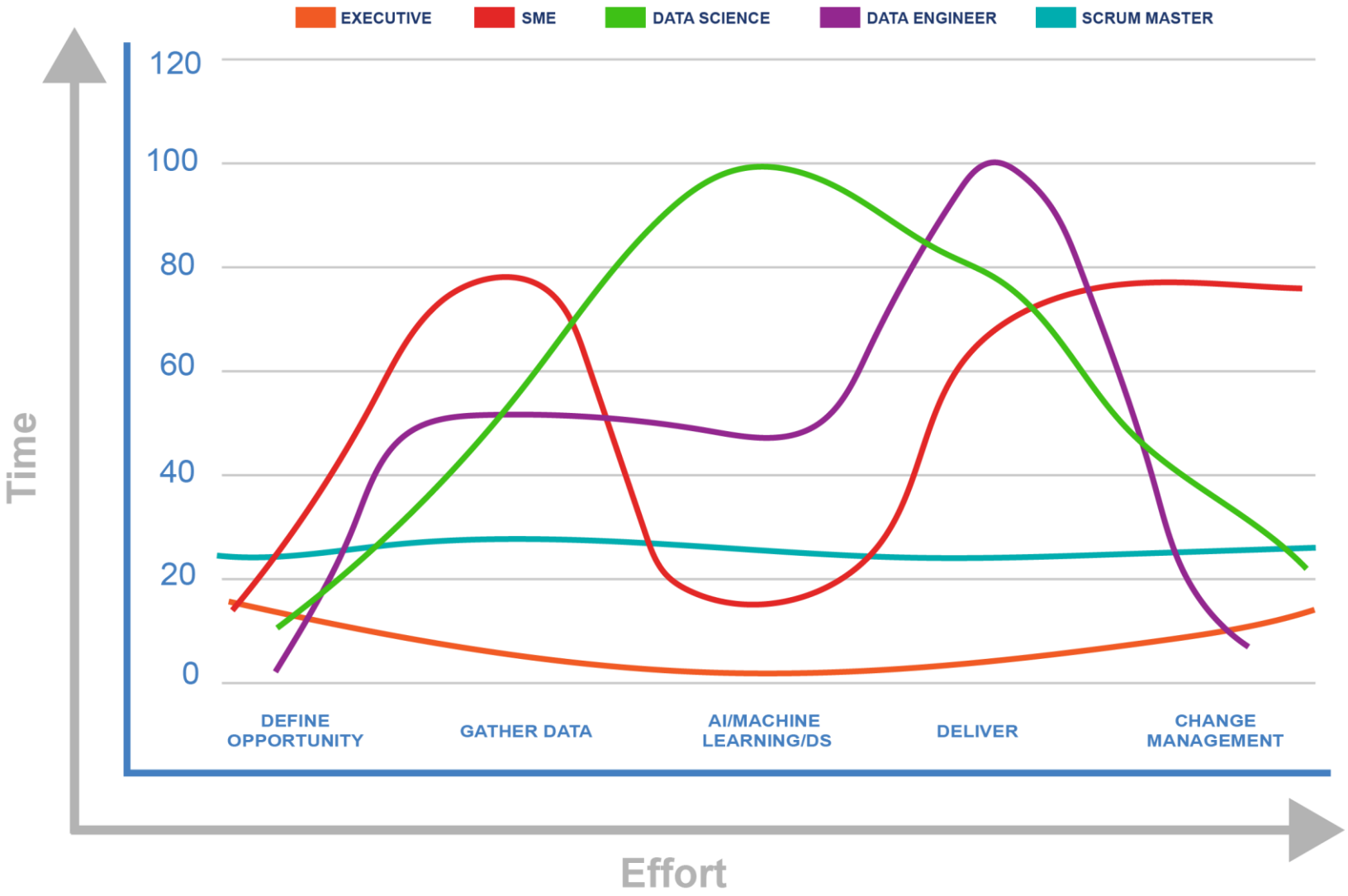
LUMA INSTITUTE℠



Leadership



Invest Time


EXECUTIVE
SME
DATA SCIENCE
DATA ENGINEER
SCRUM MASTER
120
100
80
60
40
20
0
Time
DEFINE OPPORTUNITY
GATHER DATA
AI/MACHINE LEARNING/DS
DELIVER
CHANGE MANAGEMENT
Effort

Invest Time

POLICE OFFICER
LOS ANGELES POLICE
LAPD

LOS ANGELES POLICE DEPARTMENT
BUREAUS AND AREAS
VALLEY BUREAU
WEST BUREAU
CENTRAL BUREAU
SOUTH BUREAU
POLICE STATION

Your questions?

Gerhard Pilcher
Elder Research, Chief Executive Officer

Gerhard enjoys predictive analytics and data mining, especially related to the areas of Fraud Detection, Financial Risk Management, and Health Care outcomes using various analytical methods, working with people, leading change, and timely management of complex projects. His work experience spans both private and government sectors including international experience.

Gerhard teaches at Georgetown University as an adjunct faculty member in the Math and Statistics Masters degree program. He also is an instructor for the three-day SAS Business Knowledge Series course "Data Mining: Principles and Best Practices" and been invited to teach at international conferences. Gerhard currently serves on the Institute for Advanced Analytics Advisory Board and George Washington University Masters in Science in Business Analytics Advisory Board.

Gerhard has extensive industry experience in government oversight, financial, construction and telecommunication industries both as a business owner and executive. He is a recognized expert in three dimensional roadway modeling and automated machine guidance using Global Positioning Satellite systems and has presented to various agencies including the Transportation Research Board.