

# Closing Thoughts

**Vishal Patel**

Spring 2022

# What We Covered

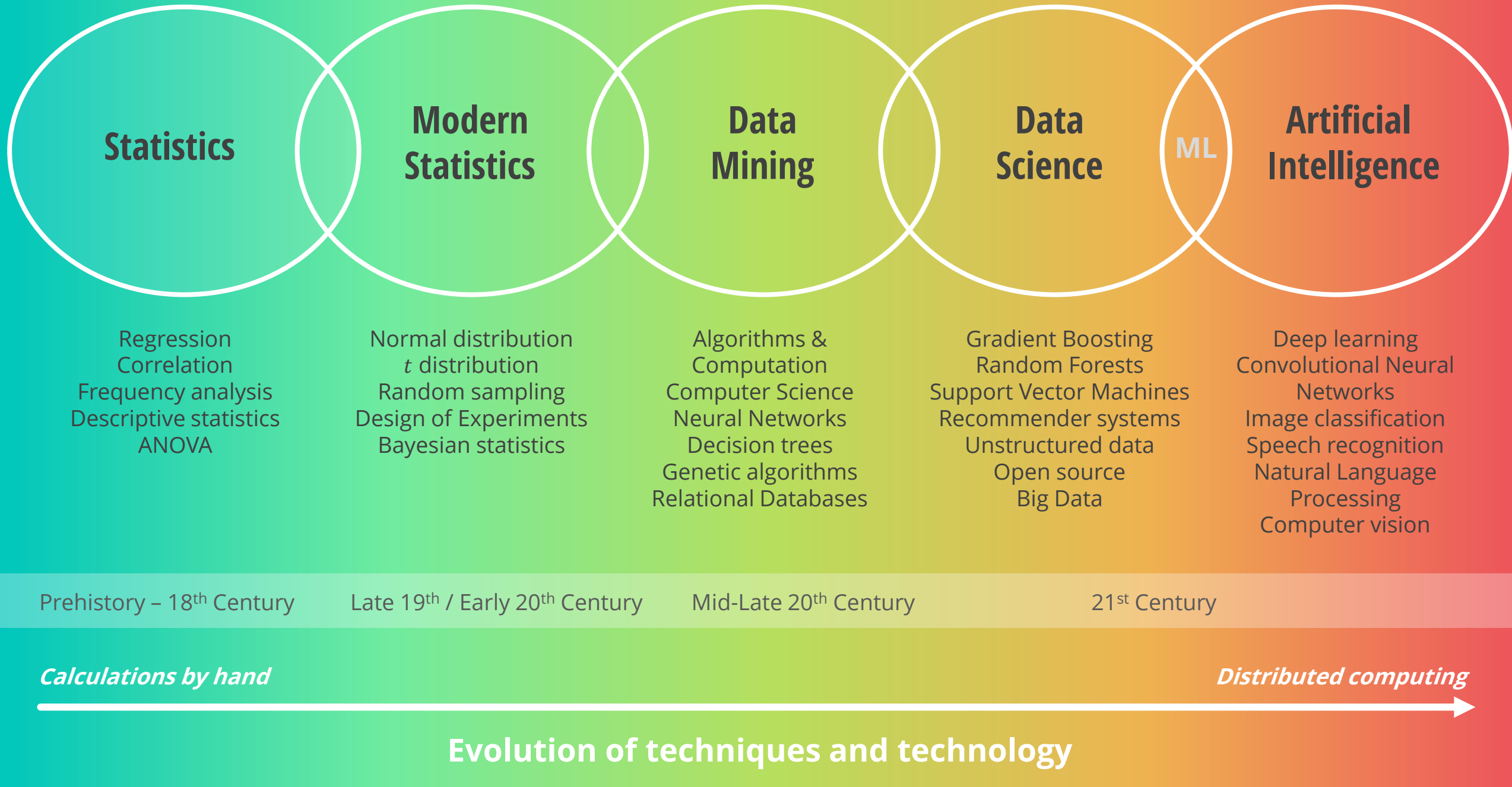
1. Introduction to Data Mining
2. The Data Science Process
3. Introduction to Regression
4. Linear Regression model
5. Build a model using partitioning
6. Decision trees
7. Classification Trees
8. Random Forests
9. Gradient Boosting Trees
10. Hyper-parameter optimization
11. Introduction to Neural Networks
12. Introduction to Clustering
13. Agglomerative Clustering
14. k-means Clustering
15. DBSCAN Clustering
16. PCA
17. Association Analysis (Apriori algorithm)
18. Collaborative Filtering
19. Data Wrangling

+ 20 Jupyter Notebooks

**Education:** What, when and why

**Training:** How





**“The last 10 years have been about building a world that is mobile-first.  
But in the next 10 years,  
we will shift to a world that is **AI-first.**”**

– Sundar Pichai



**“With artificial intelligence  
we are **summoning the demon.**”**

– Elon Musk





Boston Dynamics

**I, FOR ONE, WELCOME OUR**

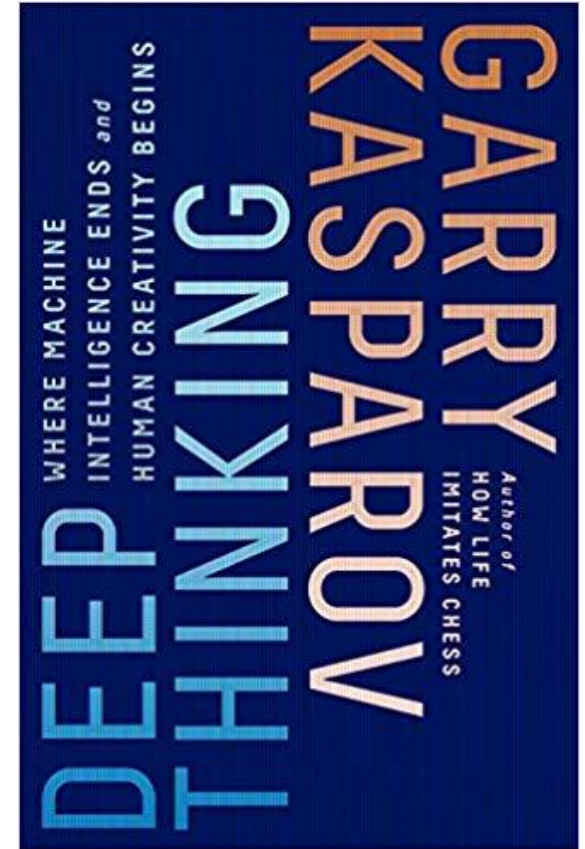


**NEW AI OVERLORDS**

memegenerator.net



Good riddance, you might imagine. But the worries about operator-less elevators were quite similar to the concerns we hear today about driverless cars. In fact, I learned something surprising when I was invited to speak to the Otis Elevator Company in Connecticut in 2006. The technology for automatic elevators had existed since 1900, but people were too uncomfortable to ride in one without an operator. It took the 1945 strike and a huge industry PR push to change people's minds, a process that is already repeating with driverless cars. The cycle of automation, fear, and eventual acceptance goes on.





1

**Jobs**

2

**Ethics**



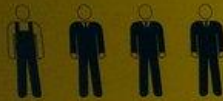
# Ein Bauer ernährt zusätzlich ...

Durch die zunehmende Mechanisierung ist folgende Entwicklung möglich geworden:



Bauer Städter

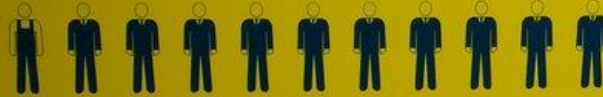
## 1900



### +3

Ein Bauer ernährt sich und zusätzlich drei Städter.

## 1950



### +10

Ein Bauer ernährt sich und zusätzlich zehn Städter.

## 2005



### +142

Ein Bauer ernährt sich und zusätzlich 142 Städter.

# The number of people one farmer could feed

Museum of Bread Culture, Germany



waze



@waze

Follow



We're working with the [@LADOTofficial](#) and our [#WazeCommunity](#) to keep the Waze Map updated with road closures (currently at 110) & open shelters (16) to help Southern Californians navigate safely around the [#wildfires](#). Please help us spread the word.

**Joel Rubin**  [@joelrubin](#)

"The Los Angeles Police Department asked drivers to avoid navigation apps, which are steering users onto more open routes — in this case, streets in the neighborhoods that are on fire." [fw.to/EMJxt2E](#)

7:59 AM - 7 Dec 2017





**Tay** was an artificial intelligence chatbot that was originally released by Microsoft Corporation via Twitter on March 23, 2016.

It caused subsequent controversy when the bot began to post inflammatory and offensive tweets through its Twitter account, forcing Microsoft to shut down the service only 16 hours after its launch.

It was soon replaced with **Zo**.



(a) Three samples in criminal ID photo set  $S_c$ .



(b) Three samples in non-criminal ID photo set  $S_n$ .



Racial features purportedly associated with criminality, from Wu and Zhang (2016)

Smile detector!

"[T]he authors find that their algorithm can classify criminal faces with 90% accuracy ... algorithm finds that criminals have shorter distances between the inner corners of the eyes, smaller angles between the nose and the corners of the mouth and higher curvature to the upper lip"

# Social and Economical Bias

1. Errors (and biases) propagate quickly
2. Far reaching consequences
3. 'Bias' is not just a technical issues, it's a **socio-technical** issue
4. There's no silver bullet-solution
  1. What do we mean by "bias"?
  2. Who gets to decide what's fair?

# Numbers don't lie!!

---

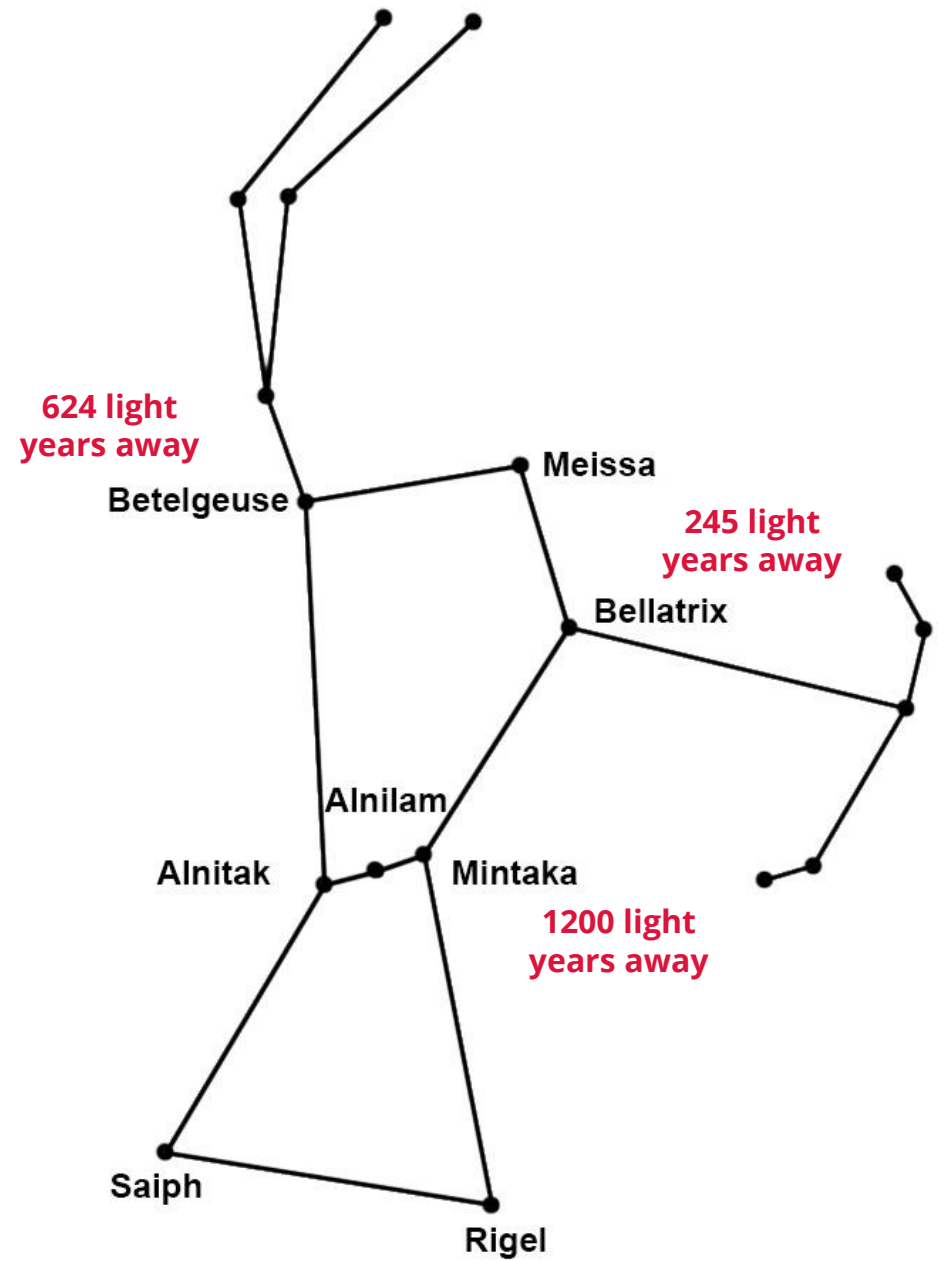
**If you torture the data long enough,  
it will confess to anything.**

*From Darrell Huff's book *How to Lie With Statistics* (1954)*

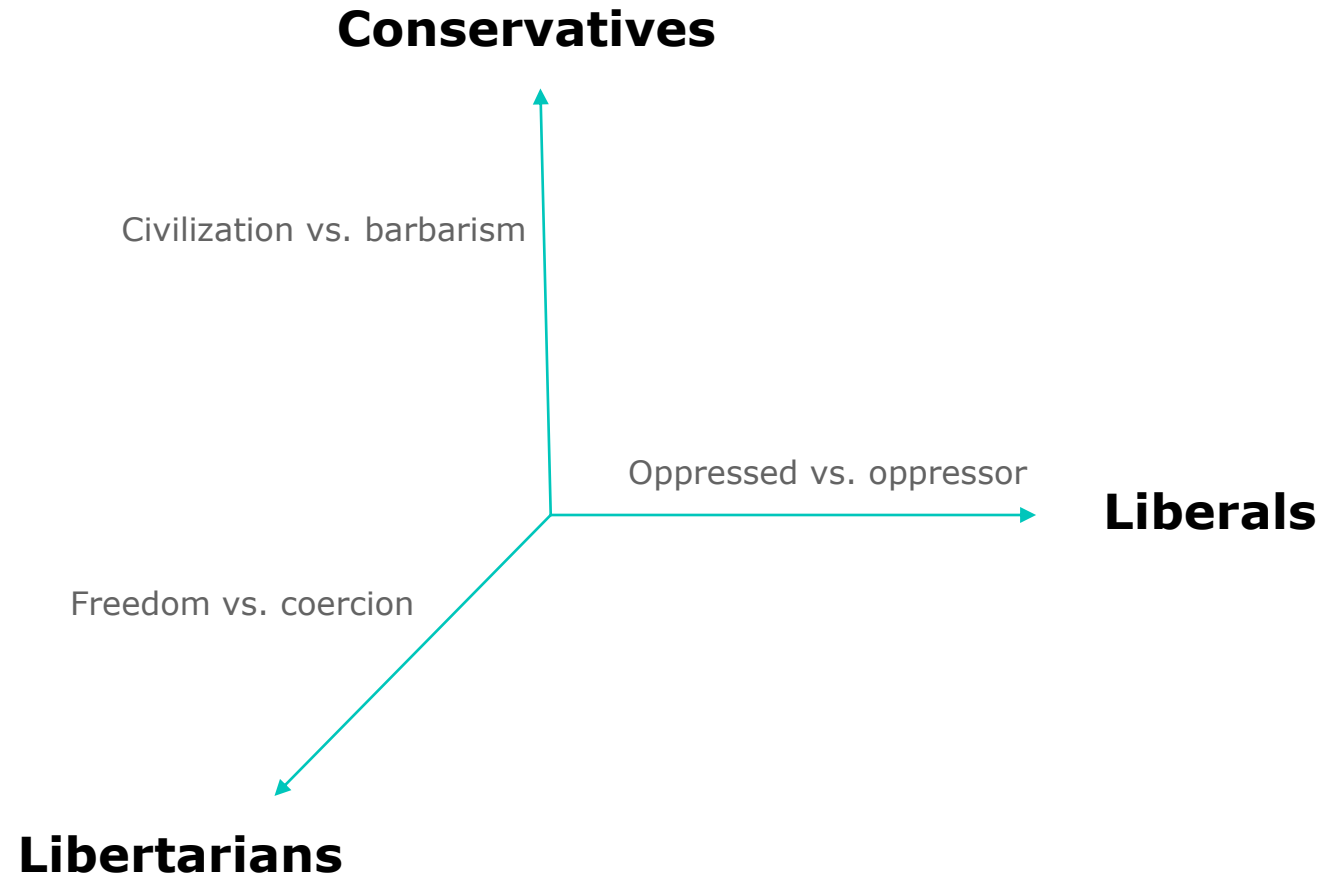
**Truth is much too complicated  
to allow anything but approximations.**

*John von Neumann*





# Arnold Kling's *The Three Languages Of Politics*





Mulla Nasrudin

**Someone saw Nasrudin searching for something on the ground.**

**“What have you lost?, he asked.**

**“My key”, said the Mulla.**

**So they both went down on their knees and looked for it.**

**After some time the man asked: “Where exactly did you drop it?”**

**“In my own house”.**

**“Then why are you looking here?”**

**“There is more light here than inside my own house”.**

**– Sufi parable [Idris Shah, citing Mulla Nasrudin]**

# **WORKING WITH AVAILABLE LIGHT**

**There is no substitute for  
getting to **know your data.****

– Witten and Frank

# Qualities of an Ideal [Data Scientist]

1. Curious, but skeptical
2. Critical thinking skills
3. Logical reasoning ability
4. Creativity
5. Strategic thinking
6. Ability to read the facts and understand assumptions
7. Understands necessary information and how to act on it
8. Engages with people to understand root of the problem
9. Ability to apply appropriate [data science] tools



“Before starting any project, I ask: What’s the geology, what’s the geomorphology, what’s the history, where does the wind come from, where does the sun come from, what are the shadow patterns, what’s the drainage system, what’s the flora?

...I try to be a **contextual** architect.” – **Glenn** Murcutt

Age of  
**CONTENT**



Age of  
**CONTEXT**

Contextual Data Science

# THANK YOU!!

[vishal@derive.io](mailto:vishal@derive.io)