

# 4 MYTHS ABOUT DATA SCIENCE

---

DAN S REZNIK, DIRECTOR

DATA SCIENCE CONSULTING LTD



# CONTENTS

- Artificial Intelligence is Intelligence
- Machine Learning is Learning
- Useful Analytics is Predictive
- Data Science is Science

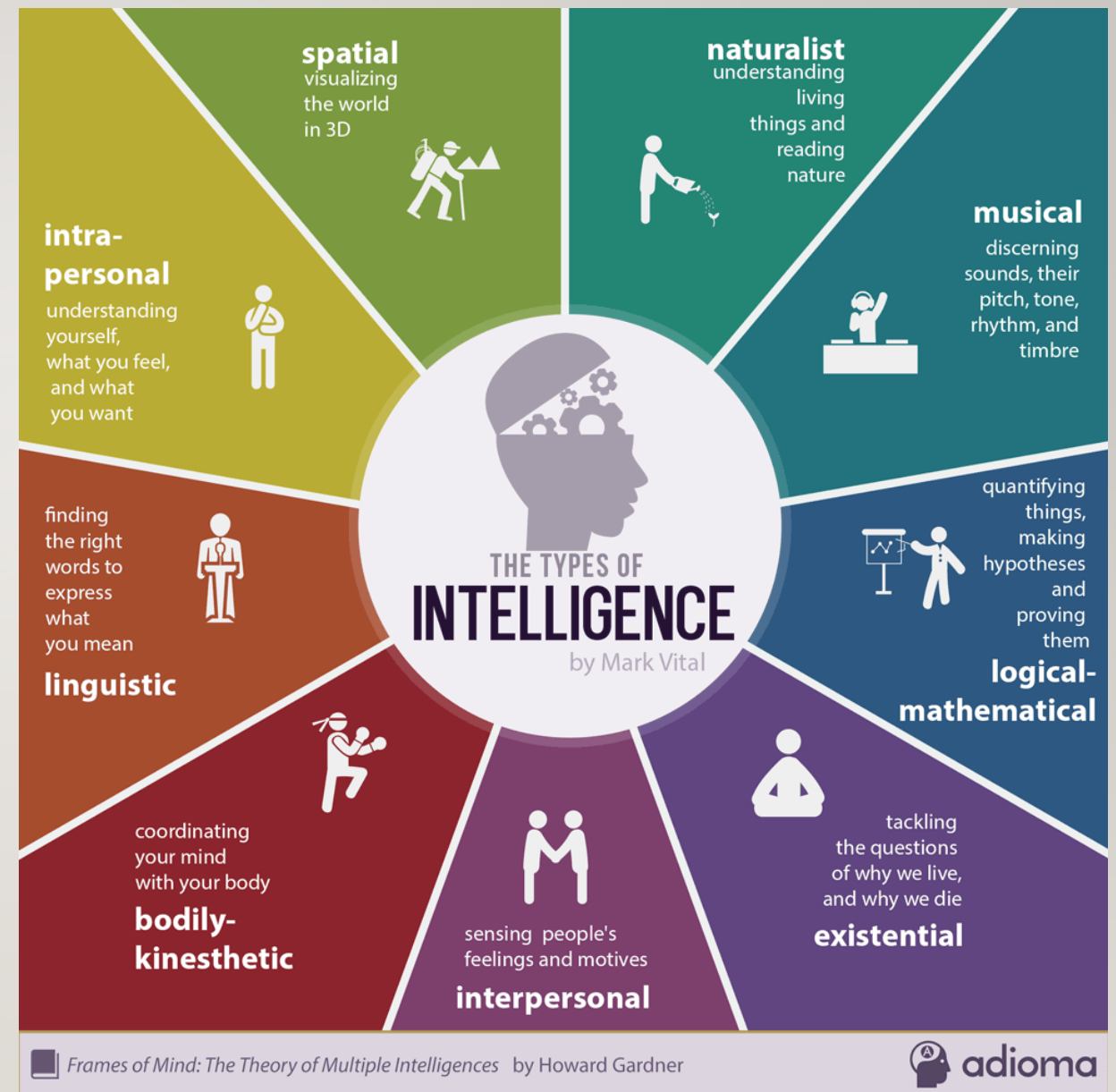
THE CLOWN SHOW

# MYTH I: ARTIFICIAL INTELLIGENCE IS INTELLIGENCE





# TYPES OF INTELLIGENCE



# COGNITIVE THEORIES (Information-Processing)

Rehearsal

Forgetting

Complex  
Cognition

Perception

## EXECUTIVE PROCESSES (Mechanisms of Change)

Encoding

### ATTENTION

- Selective
- Divided
- Sustained
- Executive

### 1 ENCODING (Getting)

- Rehearsal
- Deep-Processing
- Elaboration
- Constructing Images
- Organization
  - Outlining
  - Hierarchical Arrangement
  - Tabulation
  - Chunking

## MEMORY

### STAGE MODEL

- Sensory
- Short-Term
- Long-Term

## PROCESSES

### 3 RETRIEVAL

#### CUES

- Recall
- Recollection
- Recognition
- Relearning

### 2 STORAGE (Retention)

#### Memory

- Sensory
- Short-Term
- Working
- Long-Term

### FORGETTING (Reasons Why We Forget)

- Retrieval Failure
- Interference
- Failure to Store
- Motivated Forgetting

# INTELLIGENCE

---

## ARTIFICIAL

- Parse digital data
- Achieve specific goals and tasks
- Adapt
- Applications
  - Games
  - OCR, voice-to-text
  - Object, face recognition
  - Autonomous cars

## HUMAN

- Understanding content
- Awareness of self, other, and context
- Learning, applying
- Emotional intelligence / theory of mind
- Reasoning / problem-solving
- Planning, Creating
- Critical thinking, rejecting
- Joking / Loving / Giving



# ATLAS

---

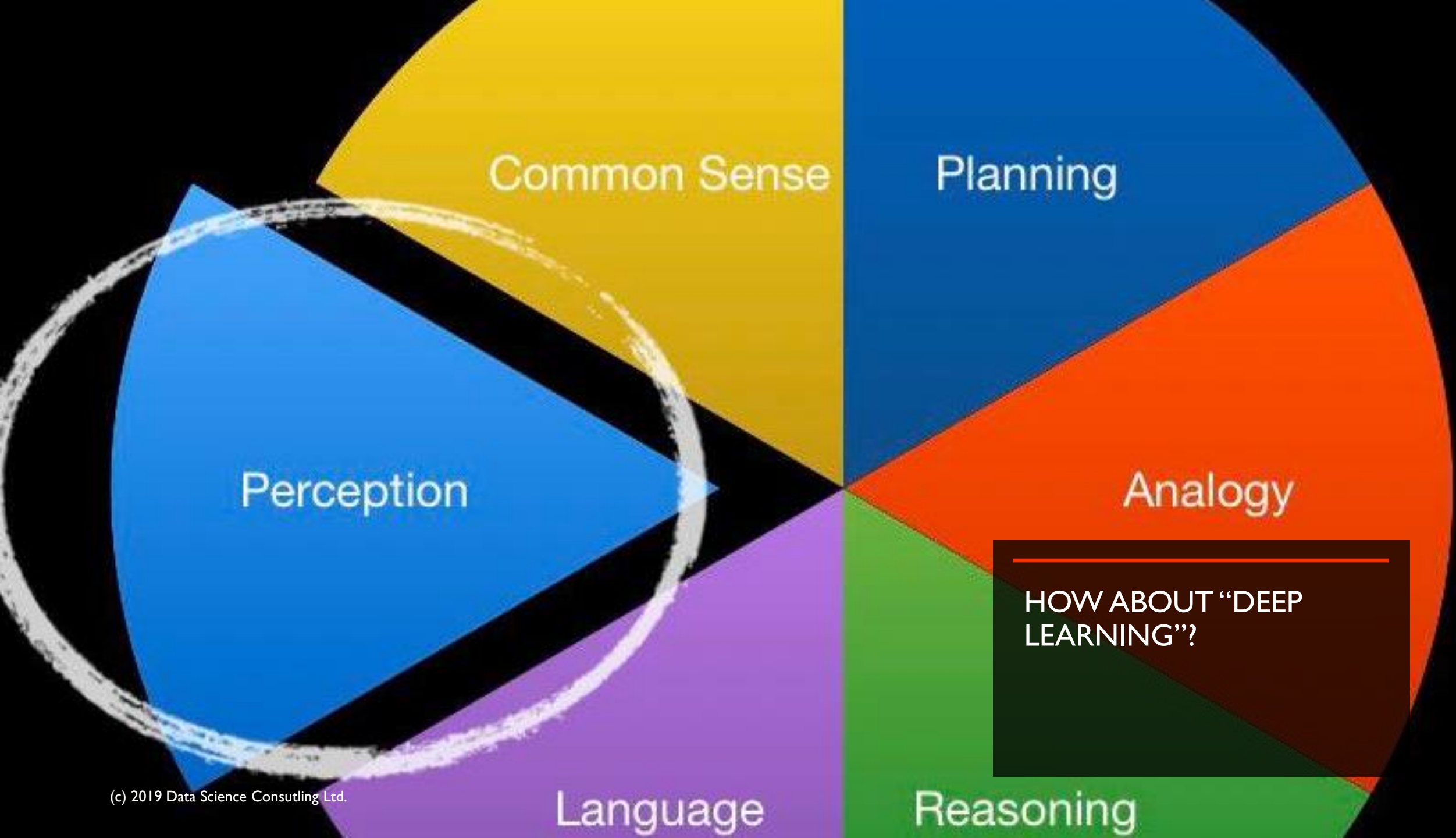




# AUTONOMOUS CARS

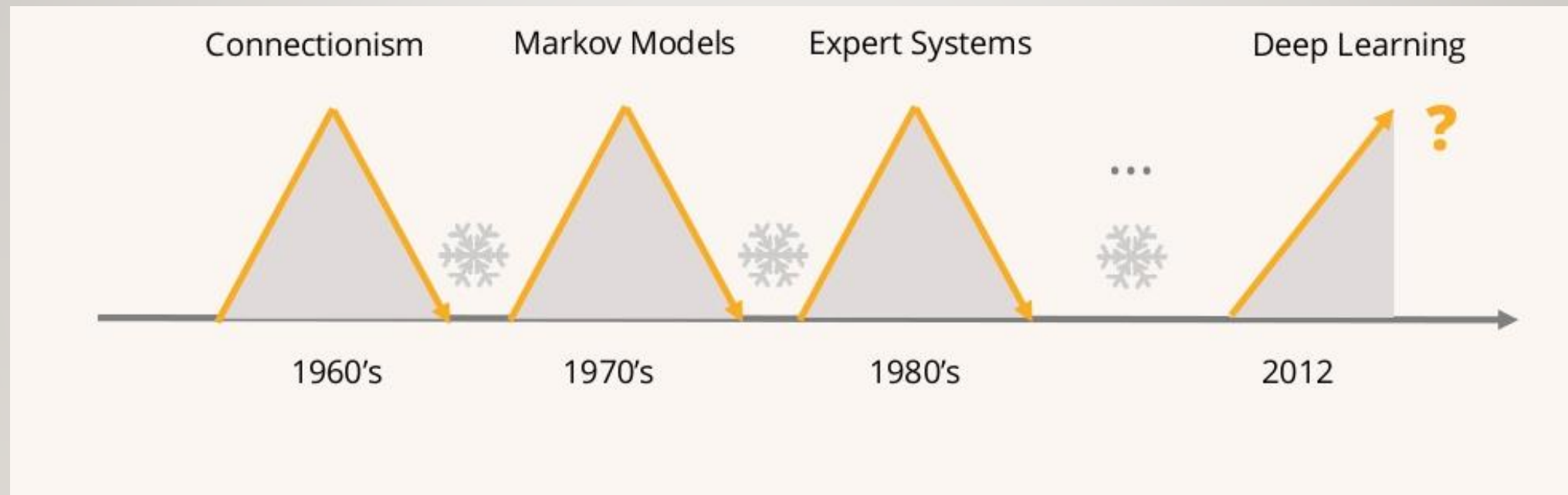
---





# AI “WINTERS”

---





# GOL

---





# MYTH 2: MACHINE LEARNING IS LEARNING

---



# LEARNING

---

## MACHINE

- Estimate parameters
- Hierarchically split
- Reinforce correct behavior
- Regress network weights
- Cluster

## HUMAN

- Cognitive: recall, calculate, discuss, analyze, problem solve, etc.
- Psychomotor: dance, swim, play football, dive, drive, ride, etc.
- Affective: To like something or someone, love, appreciate, fear, hate, worship, etc.



# FACE SURVEILLANCE

---





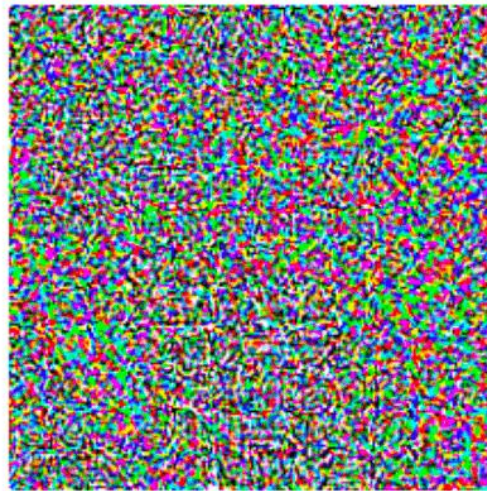
# DEEP LEARNING

---



“panda”  
57.7% confidence

+ .007 ×



“nematode”  
8.2% confidence

=



“gibbon”  
99.3 % confidence



# MYTH 3: USEFUL ANALYTICS IS PREDICTIVE

---



# ANALYTICS

---

## DESCRIPTIVE

- Describe Past
- Visualize Past

## PREDICTIVE

- Model past's parameters
- Project trends
- Project distribution
- Project multinomial fit
- Stationarity

# DASHBOARDS

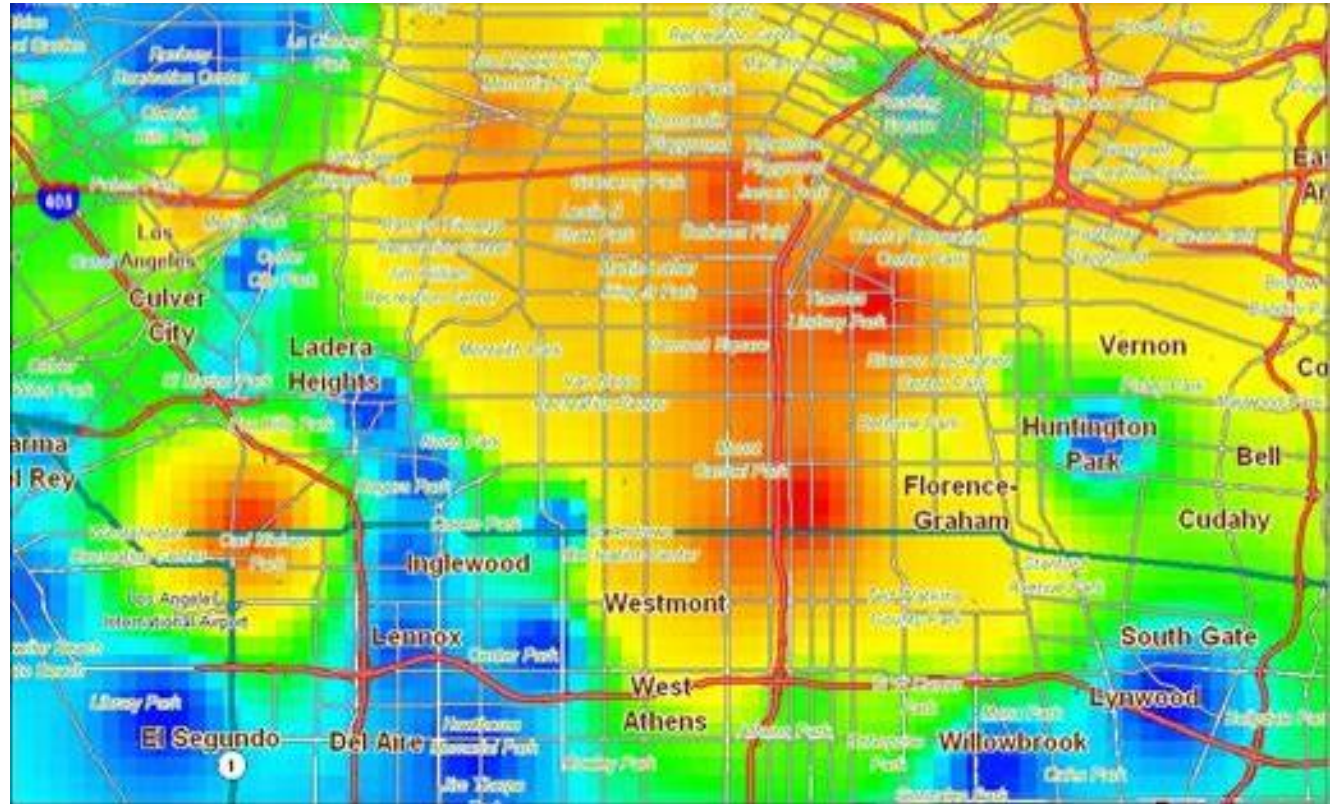
---





# CRIME PREDICTION

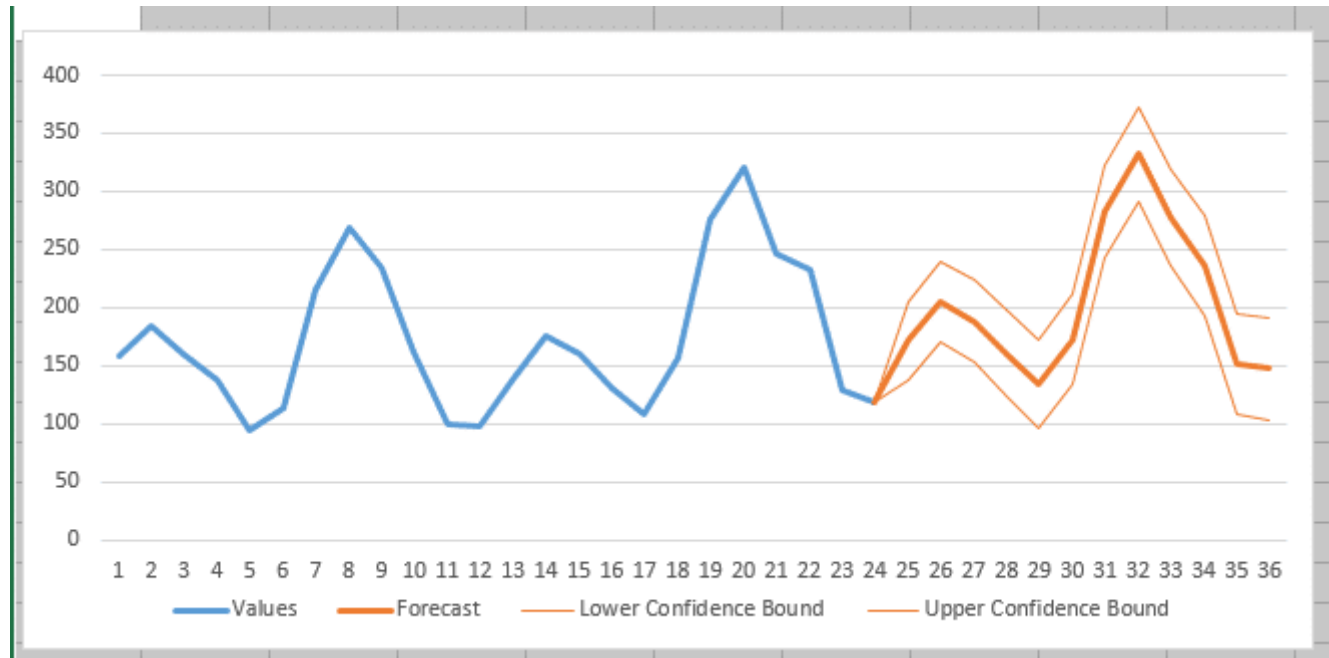
---



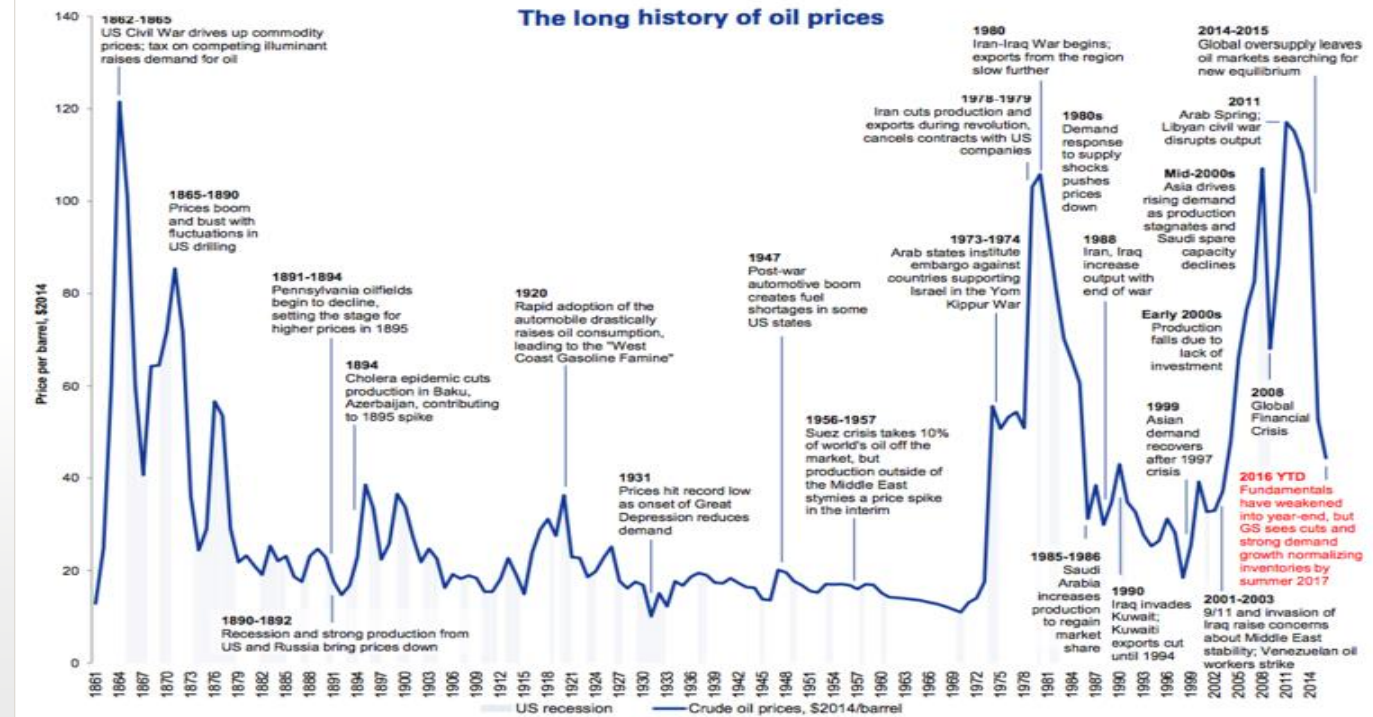


# TIME-SERIES FORECASTING

---



# TIME-SERIES FORECASTING



An earlier version of this chart appeared on pg. 16 of Top of Mind Issue #52: OPEC and Oil Opportunities.

Note: 2016 price shown is YTD average as of Dec. 19, 2016.

Source for data: BP, NBER/Federal Reserve Bank of St. Louis, Haver Analytics.

Source for annotations: ©James Hamilton, "Historical Oil Shocks," University of California, San Diego, February 2011; various news sources; Goldman Sachs Global Investment Research.

Techs Strategic to BI  
(c) 2019 Dresner Advisory

[https://twitter.com/gp\\_pulipaka/status/1178379976514494470?s=20](https://twitter.com/gp_pulipaka/status/1178379976514494470?s=20)





# MYTH 4: DATA SCIENCE IS SCIENCE

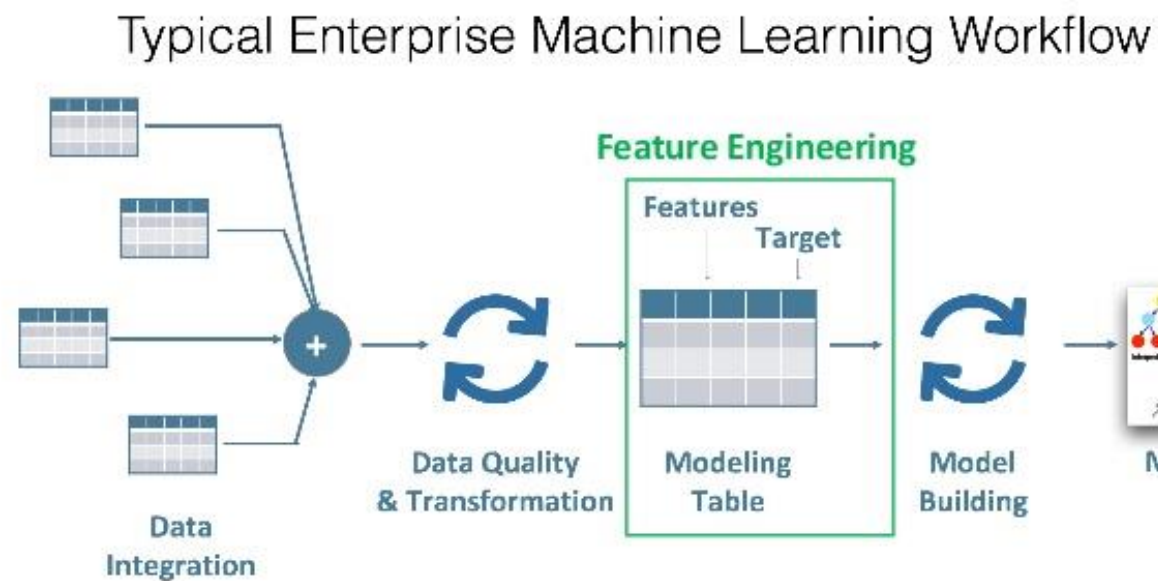
---





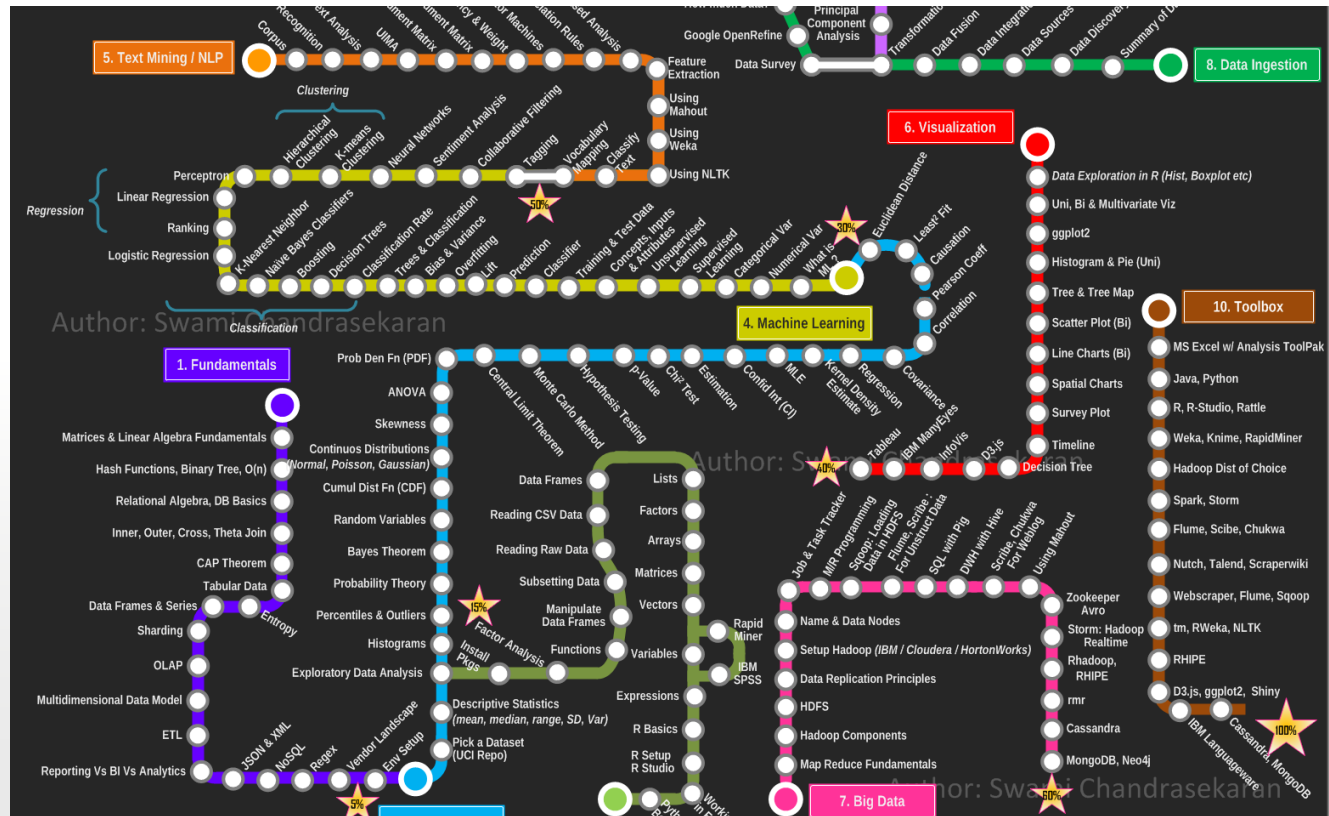
# WRANGLE FEATURES

---



H<sub>2</sub>O.ai

---





# MESTRE CALCETEIRO

---



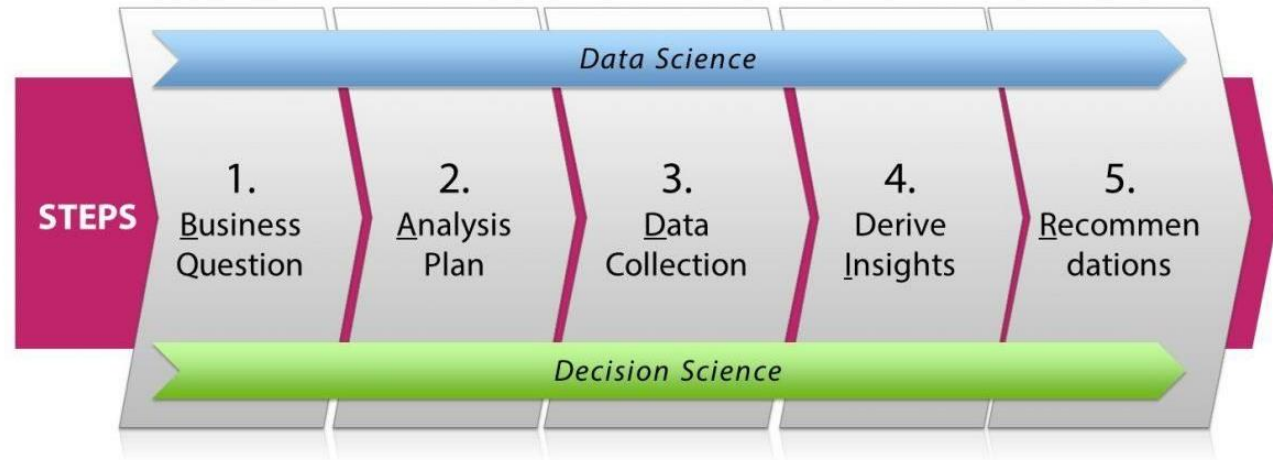


# CABINET MAKING



# START WITH PROBLEM NOT TECH

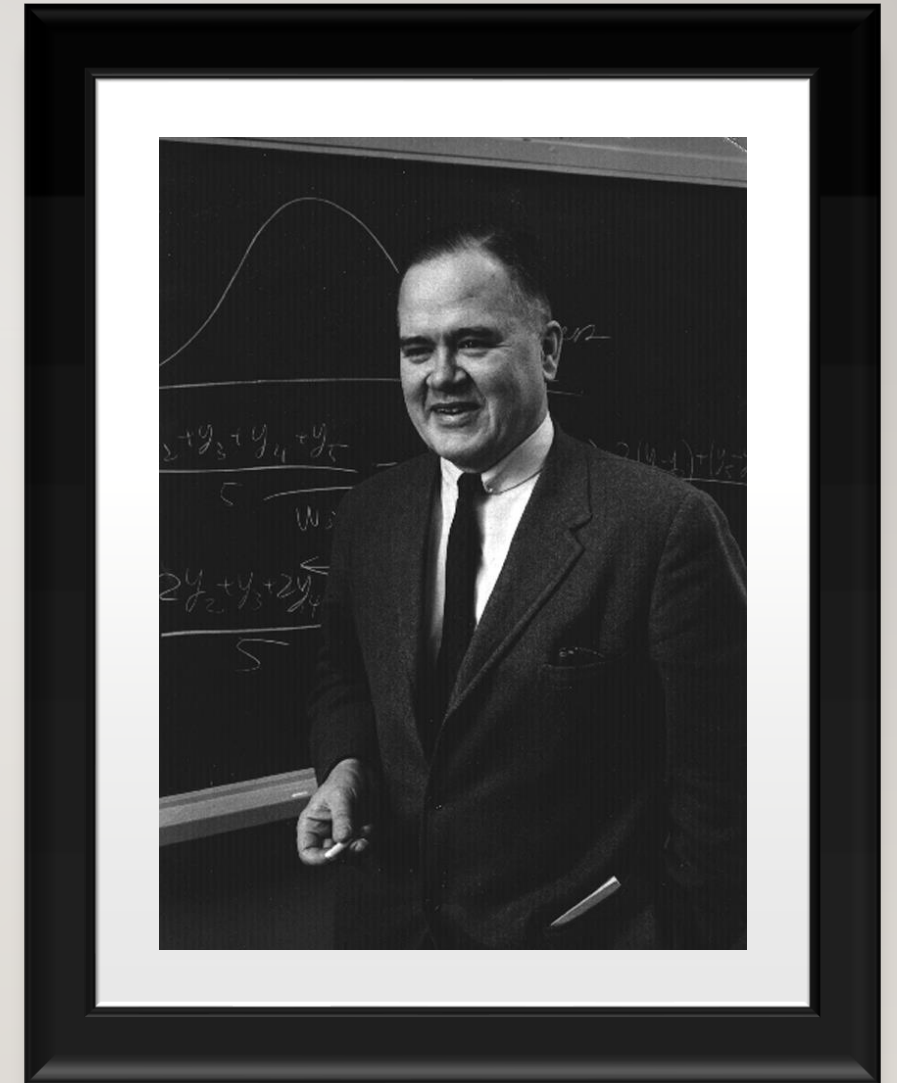
---



# JOHN TUKEY (1915-2000)

---

It is far better an **approximate** answer to the right question, which is often vague, than an exact answer to the **wrong** question, which can always be made precise

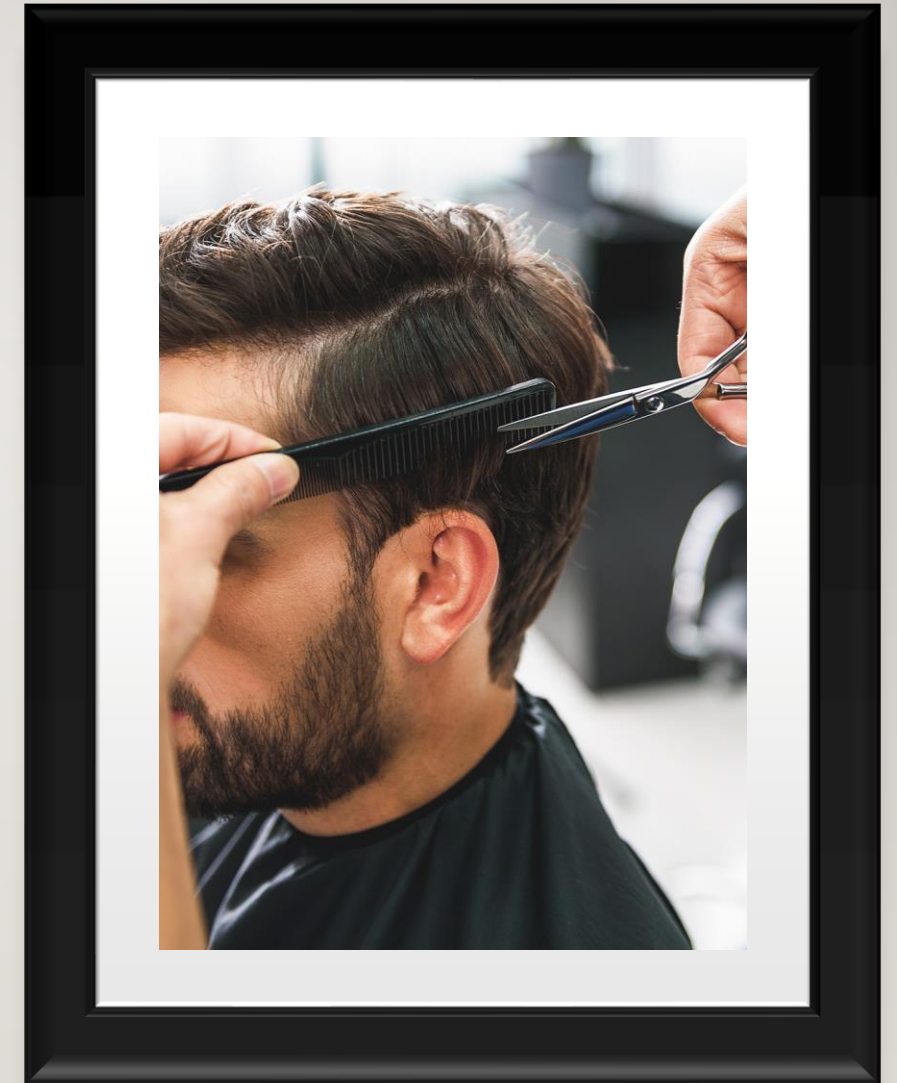




# WILL BARBERS OR DATA SCIENTISTS BE AUTOMATED?

---

- Identify real-world problems, choose analysis plan
- Design data collection, assess quality, correct labels, reject data
- Incorporate domain knowledge, e.g., via feature engineering
- Anticipate risks, manage them
- Manage biases, ethical issues, impact of project in society
- Analyze and critique performance
- Explain insights to human stakeholders, convince them



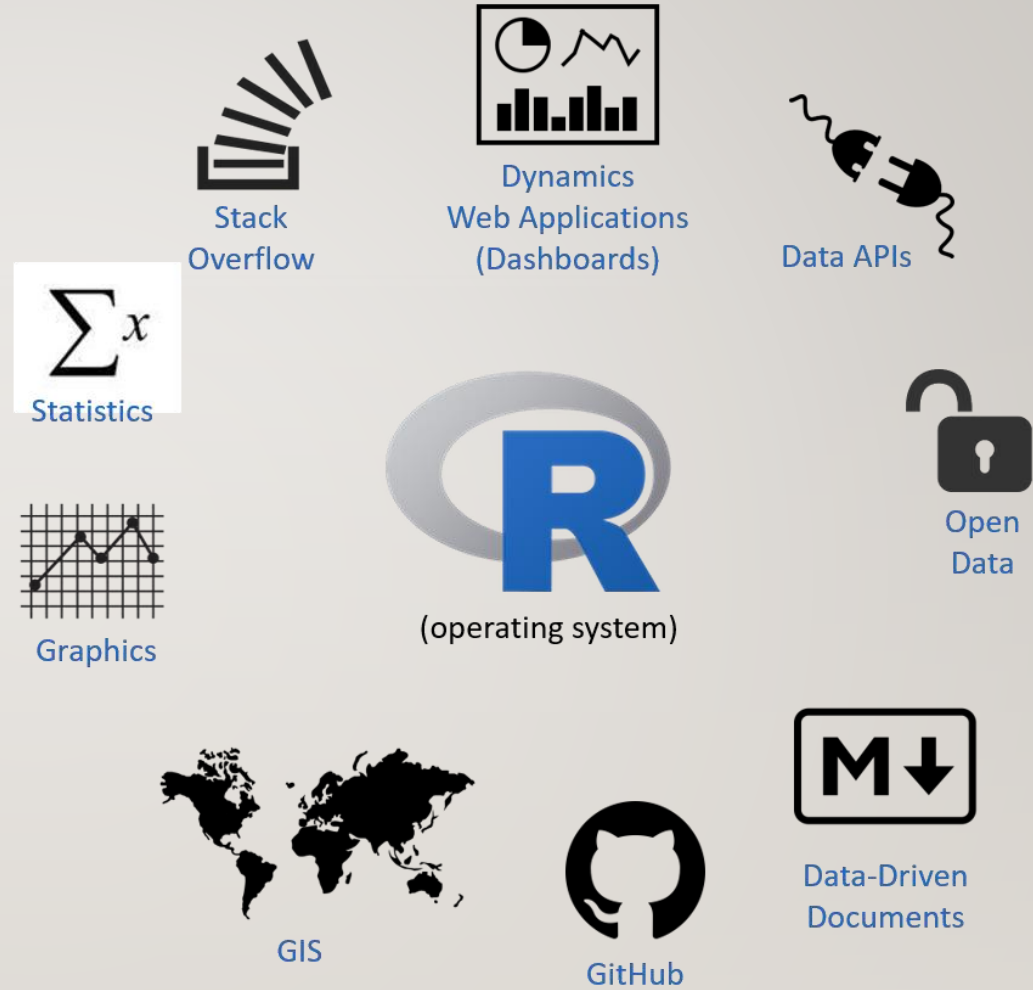
# HOW ABOUT R?

---



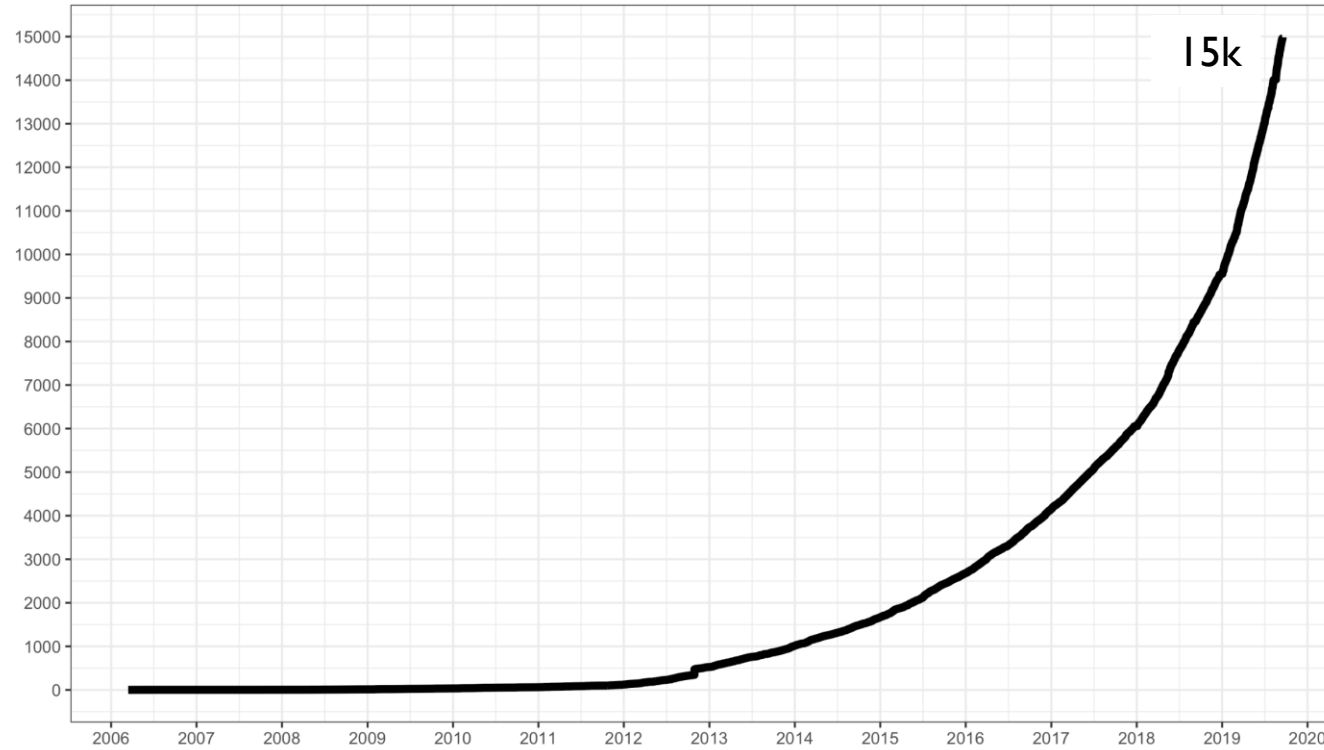
# ECOSYSTEM

---





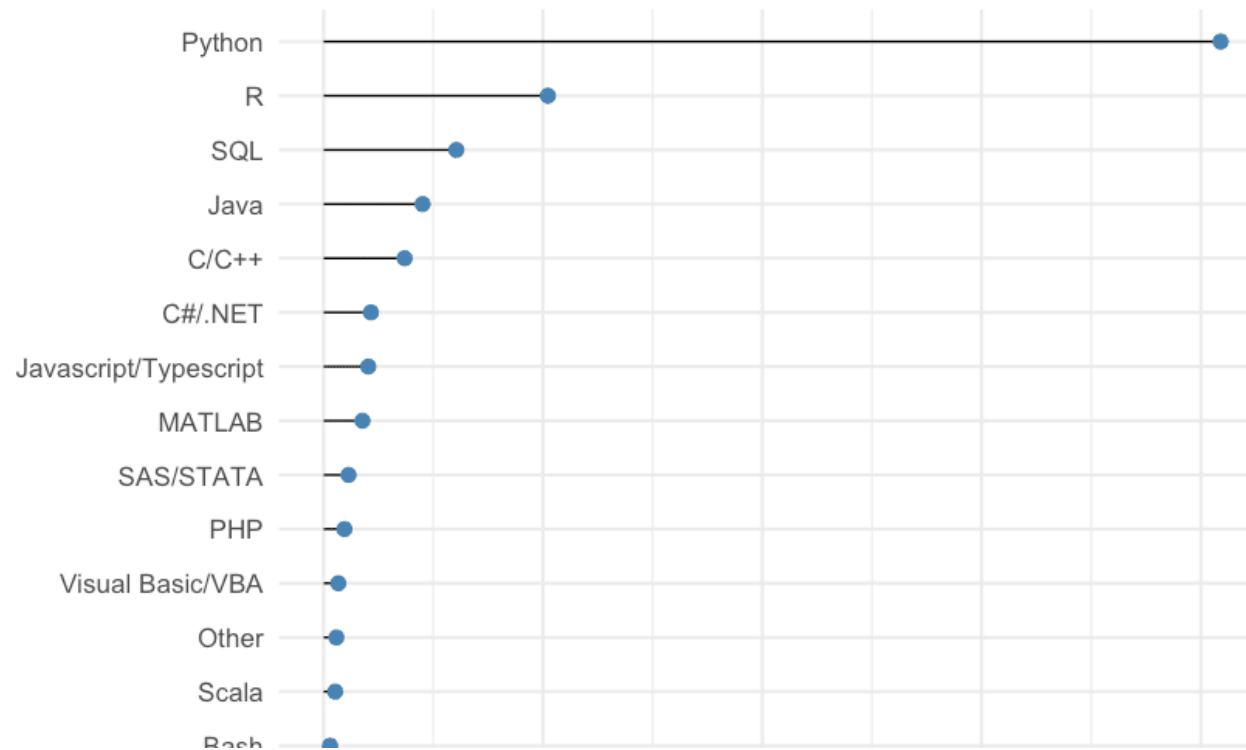
Number of R packages ever published on CRAN

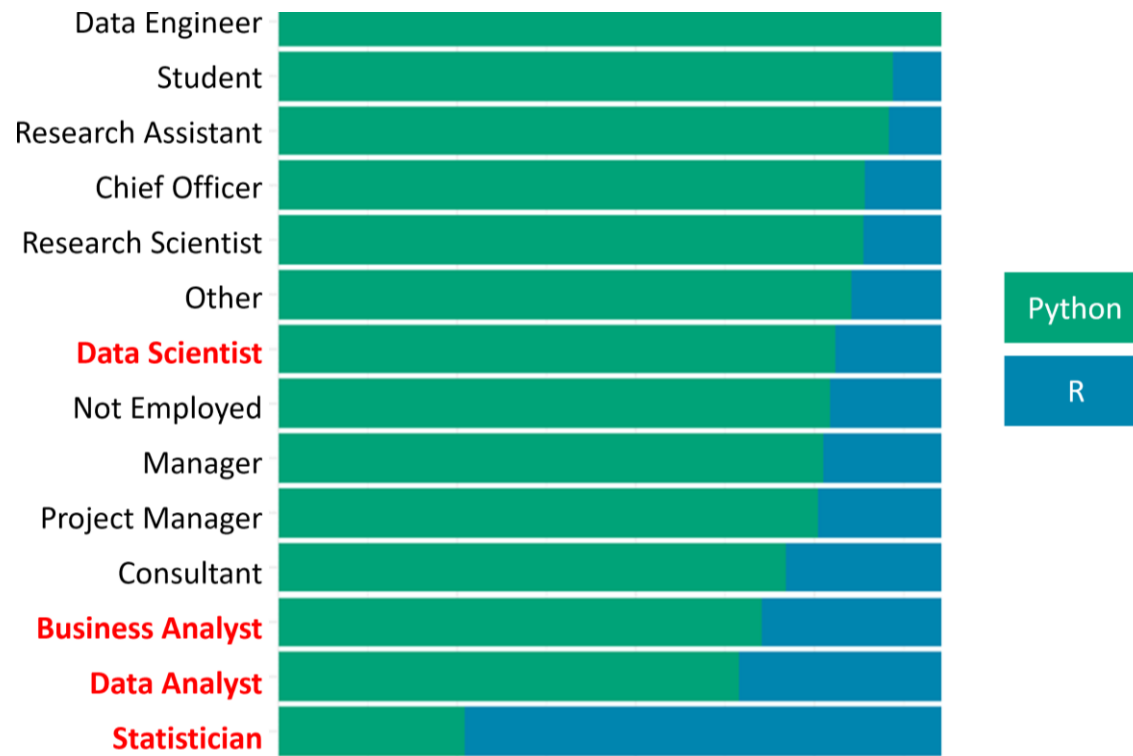


CRAN  
PACKAGES

# R VS PYTHON

---





Source: <https://www.r-bloggers.com/r-or-python/>

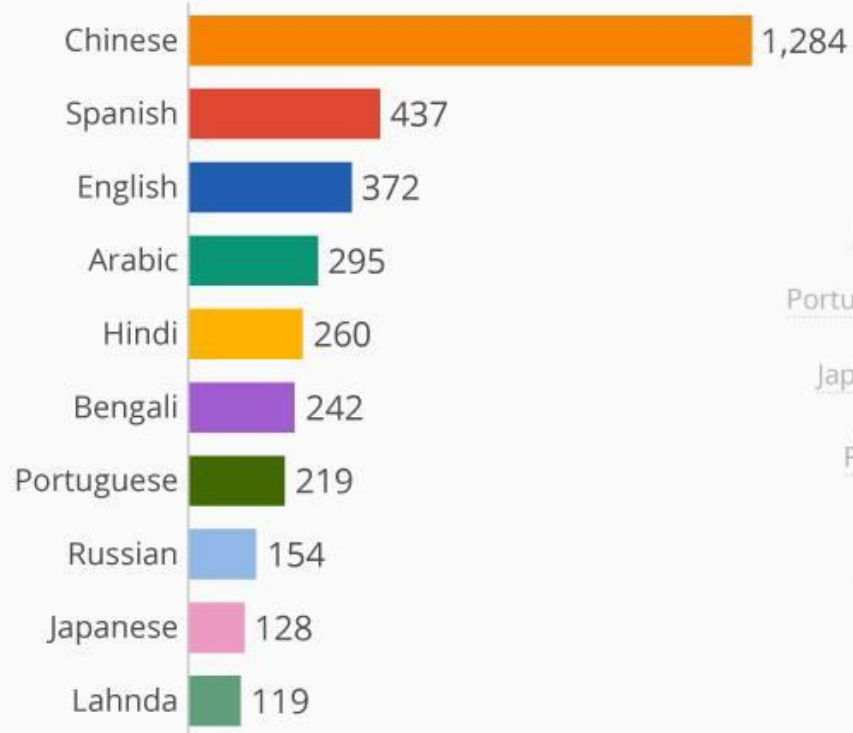
# R VS PYTHON

---



## Two Worlds: Languages IRL and Online

Estimated number of first-language speakers worldwide in 2017 (millions)\*



Percentage of websites using various content languages\*



(c) 2019 Data Science Consulting Ltd.

Start date

2019-02-01

End date

2019-12-10

Group by

- ☒ date
- ☐ week
- ☐ month
- ☐ total

Job titles

Channels

- ☒ Technical Copywriter
- ☒ Frontend Engineer
- ☒ Graphic Designer
- ☒ Head of Growth
- ☒ Head of Sales
- ☒ Project Leader
- ☒ Data Analyst
- ☒ PM
- ☒ Data Scientist
- ☒ unknown
- ☒ Software Engineer

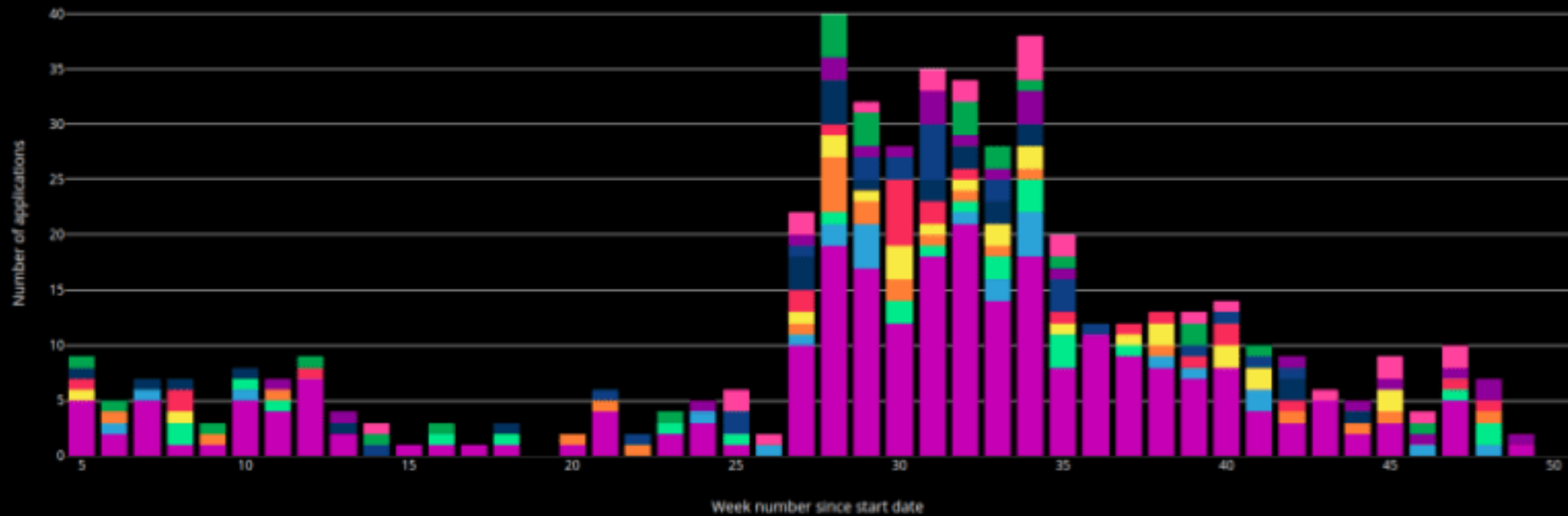
Select All

Clear All

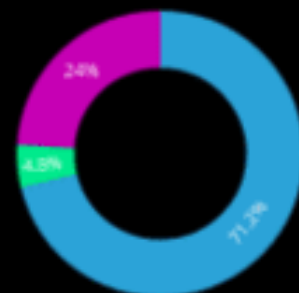
Theme

Dark

## Timeline - how many candidates we had for each job position?



## Candidates gender



## Candidates nationality



**"Open app store"**

tap to edit

**It doesn't look like you have  
an app named 'app store'. If  
you'd like, I can help you  
look for it on the App Store.**

**App Store**





IN TERMS OF GENERAL INTELLIGENCE, WE'RE NOT EVEN CLOSE  
TO A RAT.

---

YANN LECUN  
FACEBOOK AI DIRECTOR



**AI**

- sounds sexy
- gets us money from VCs
- what we all hope is the future

**Machine Learning**

- the only real “AI”
- traditionally an academic discipline
- not concerned with real-world software

**Data Science**

- applies machine learning to create actual products
- deals with real-world complexity

## SUMMARY

- Artificial Intelligence is a **statistical modeling**
- Machine Learning is **statistical modeling**
- Useful Analytics **is still mostly descriptive**
- Data Science is **carpentry**



# THANK YOU

---

DAN@DAT-SCI.COM



