

Big Data Foundations: an Industry 4.0 perspective



João Pessoa - PB
09 September 2018





Introduction





Ivanovitch Silva
ivan@imd.ufrn.br



Diego Silva
diego@ect.ufrn.br



IMD/UFRN



ECT/UFRN



Data Science & Innovation Lab

Aguinaldo Bezerra
aguinaldo@ufrn.edu.br

Vitor Greati
greati@ufrn.edu.br

Artur Curinga
curinga55@ufrn.edu.br

Vinicius Campos
vinihcamps@ufrn.edu.br



Roadmap history of research activities





Cliente

- Sede Petrobras

Problema

Demandava para acompanhar KPI's de alarmes a nível gerencial de todas unidades industriais da Petrobras diariamente.

Metodologia

- Definição dos principais índices de acompanhamento gerencial
- Desenvolvimento customizado de painel para acompanhamento gerencial
- Entrega e configuração do painel

BR-AlarmExpert

MÚLTIPLOS FONTE

ANALISES DE ALARMES

Médias anunciamas por tempo

Contagem de Alarmes

% de tempo instável

MONITORAMENTO

Tabela de Indicadores

Dashboard

CONSULTAS

Ocorrência de Alarmes

Alarms e Eventos

TABELA DE INDICADORES

00_INDUSTRIAL

30/05/2017 06:00:02 - 30/05/2017 10:03:30

Procurar:

#	UO	Alarmes / h / ponto de operação - Semana (Desejável < 6 e Meta < 12)	Alarmes / h / ponto de operação - Mês (Desejável < 6 e Meta < 12)	% de tempo em avalanche - Mês (Meta < 1%)	Alarmes obsoletos - Semana (Meta < 5)
1		0.91	3.83	0.84%	2.00
2		16.42	3.10	0.38%	3.00
3		0.00	3.82	0.48%	0.00
4		2.24	2.17	0.05%	1.47
5		0.00	8.63	0.67%	0.19
6		2.25	6.19	0.05%	14.95
7		7.62	4.97	0.74%	5.09
8		14.80	2.61	2.42%	1.40
9		1.65	3.37	0.12%	0.00
10		7.62	4.97	0.74%	5.09
11		3.20	5.34	0.00%	2.90
12		0.00	4.74	0.07%	0.00
13		15.14	12.60	0.91%	9.07
14		7.62	4.97	0.74%	5.09
15		77.58	38.99	18.77%	0.00
16		19.57	12.76	6.82%	8.00

Periodo de processamento: 30/05/2017 06:00:02 - 30/05/2017 10:03:30

Deployments of research products

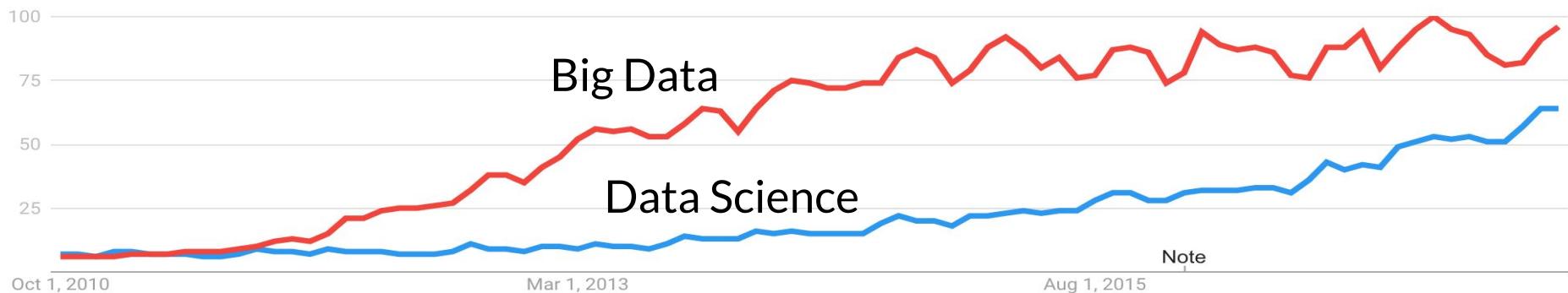


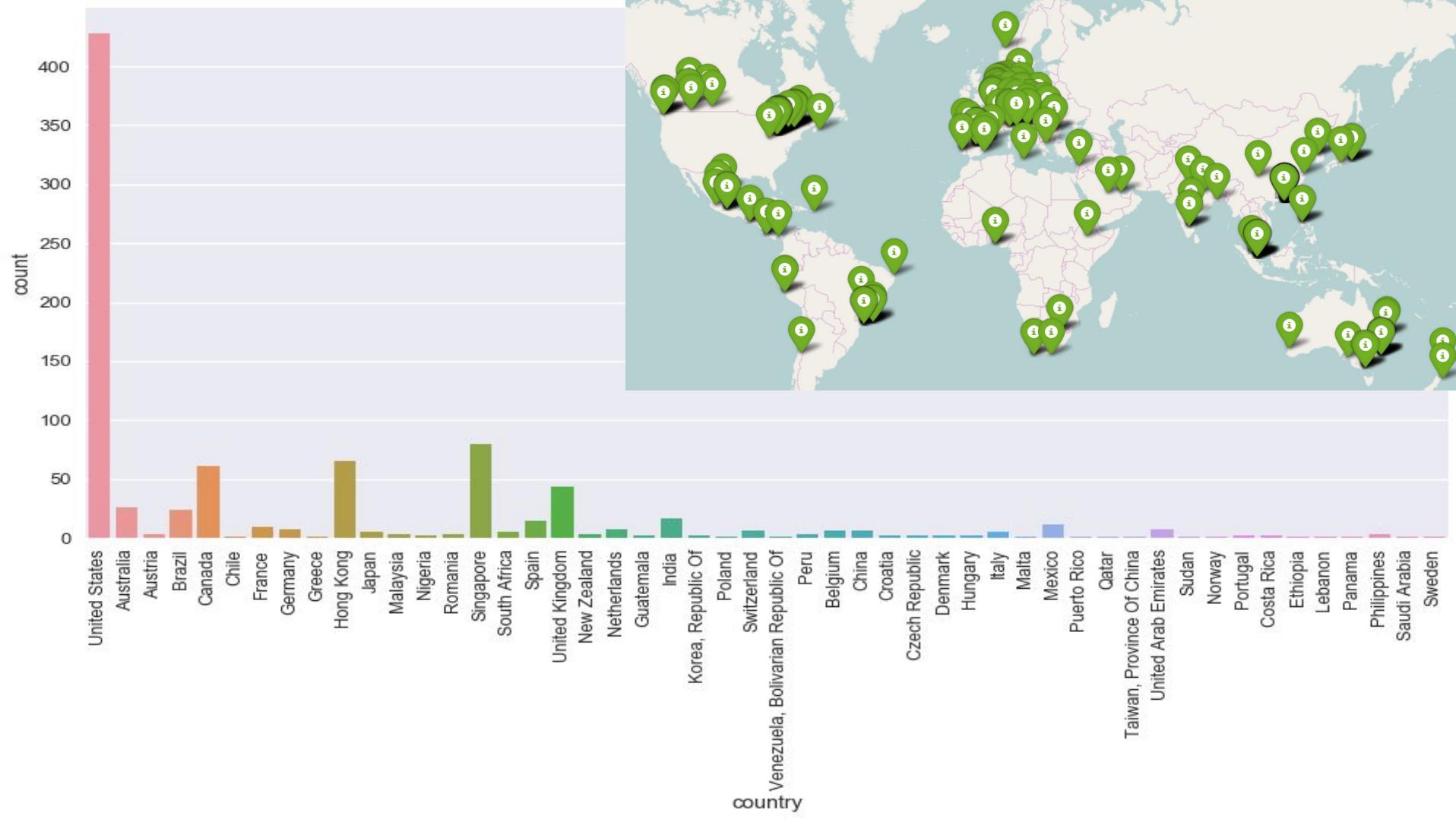
Big Data and Social Analytics certificate course

2017 DATES TO BE CONFIRMED

[DOWNLOAD COURSE PROSPECTUS](#)

Discover a new way to think about big data analysis when you explore the theory behind "social analytics", and practically apply that knowledge as you learn pioneering data analytics techniques from the creators of those very tools and methods.







2016/2017 - Specialization course in Big Data

Undergraduate

- 2017.1 - IMD0105 Introduction to Data Science
- 2017.2 - IMD0252 Learning Analytics
- 2017.2 - DCA0046 Data Science
- 2018.1 - DCA0132 Data Engineering
- 2018.2 - IMD0905 Data Science
- 2018.2 - DCA0131 Data Science

Graduate

- 2017.2 - EEC2006 Data Science Foundations
- 2017.2 - ITE0021 Learning Analytics
- 2018.2 - EEC1509 Machine Learning



<https://github.com/ivanovitchm>



Group Survey



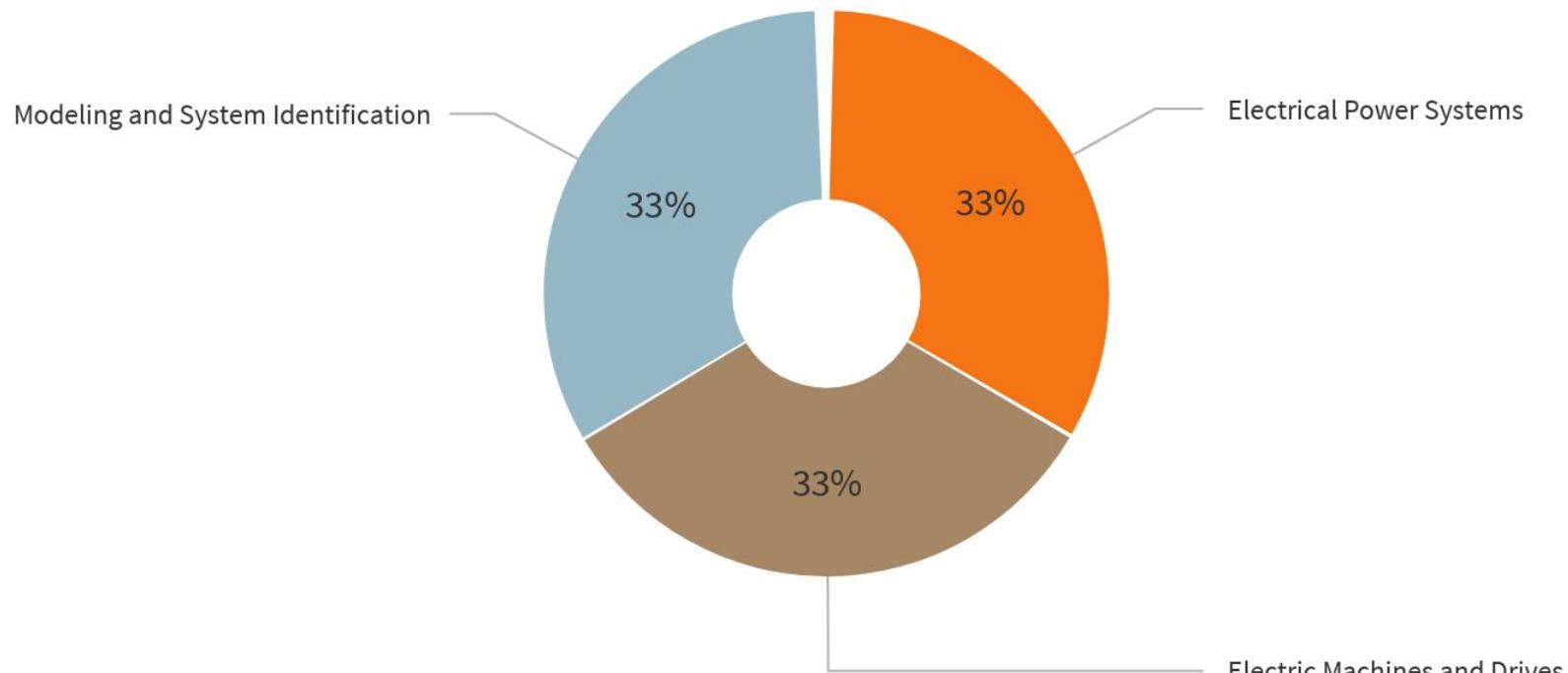
Where you come from (eg. PB, SC, AM, SP, so on)?

Total Results: 0

Start the presentation to see live content. Still no live content? Install the app or get help at PollEv.com/app

Which institution do you come from (e.g UFPB, UFRN, Petrobras, so on)?

What are your topics of interest in CBA2018 (up to 3 items)?



What is your proficiency in Python?



Agenda

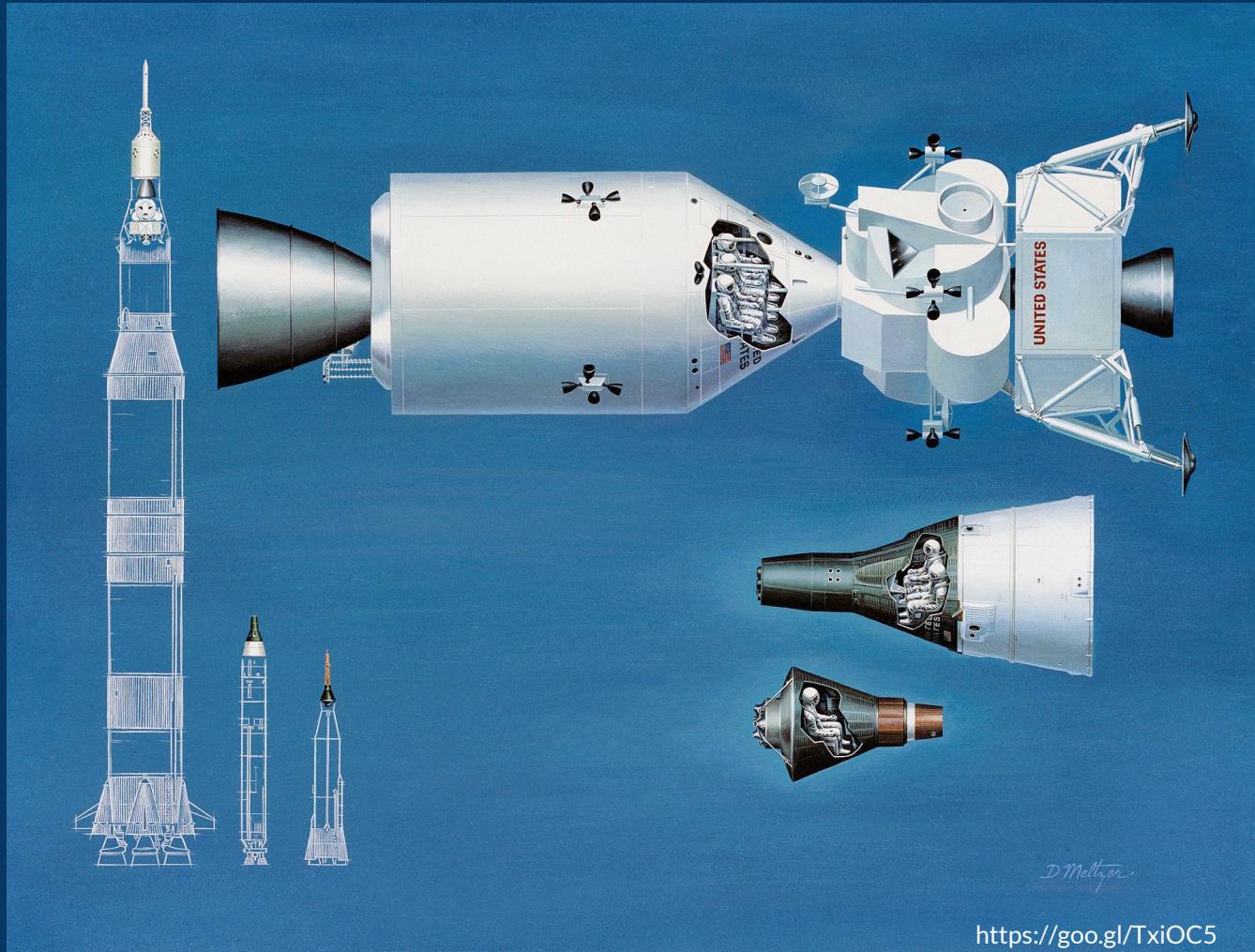
1. Motivation
2. Development platform
3. Pandas crash course
4. Data visualization
5. Advanced visualization

Provocation #1

hardware

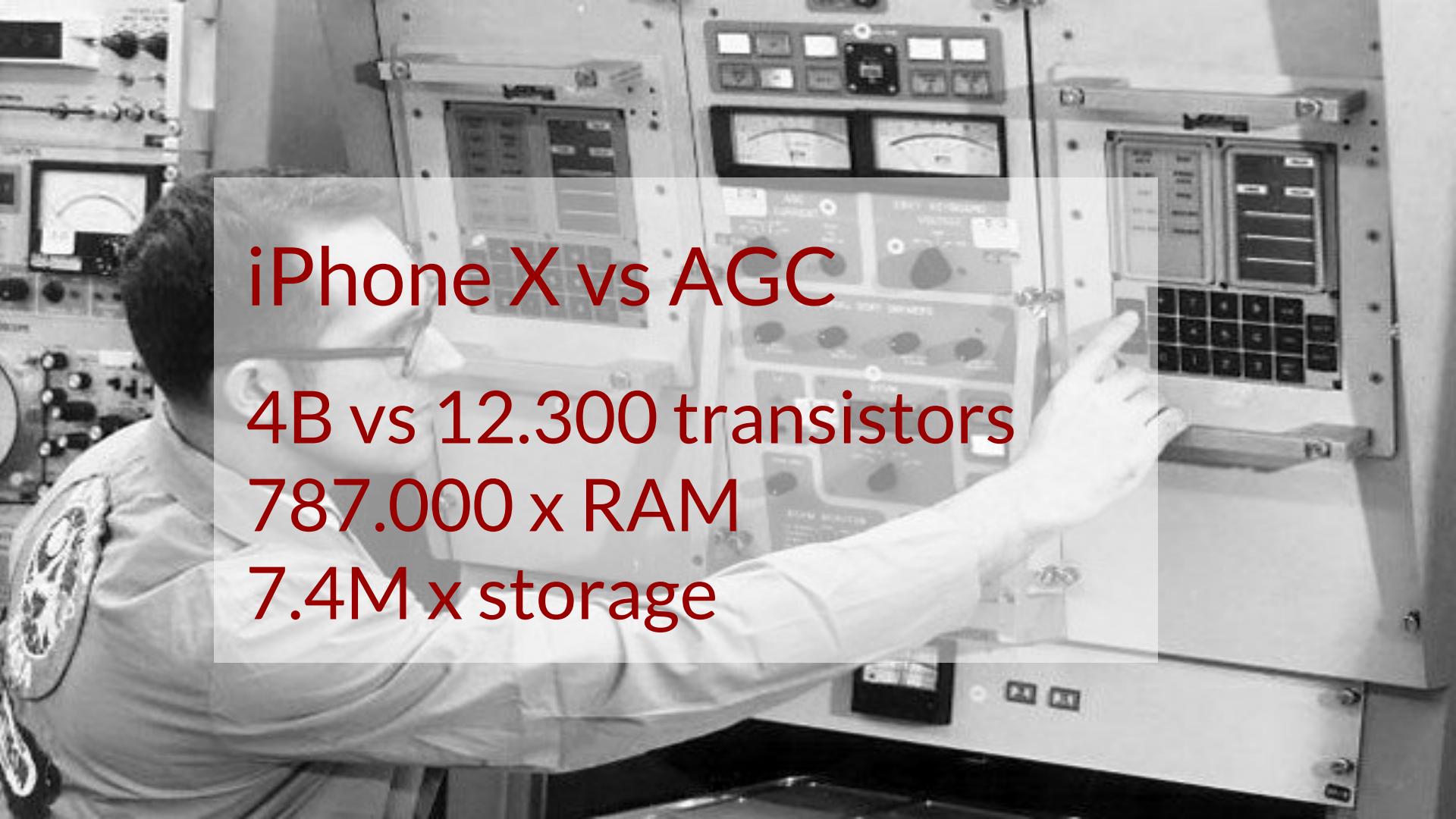


1969



D. Meltzer

<https://goo.gl/TxiOC5>



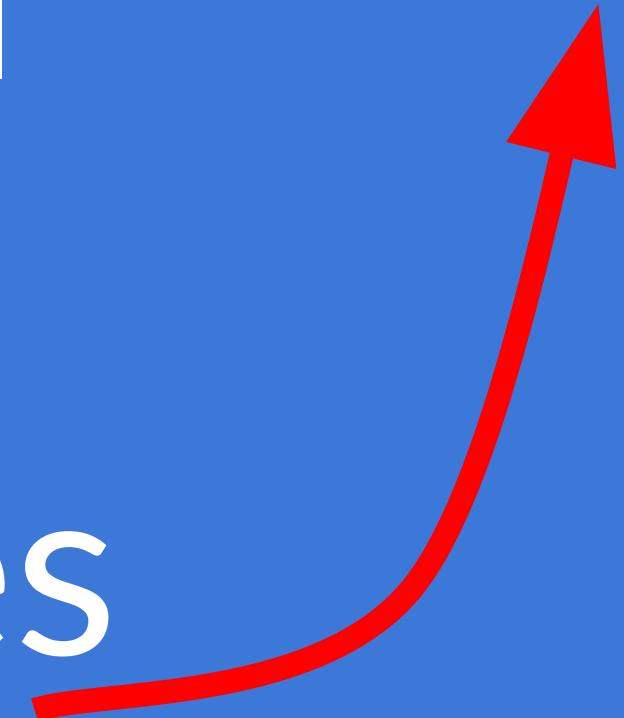
iPhone X vs AGC

4B vs 12.300 transistors

787.000 x RAM

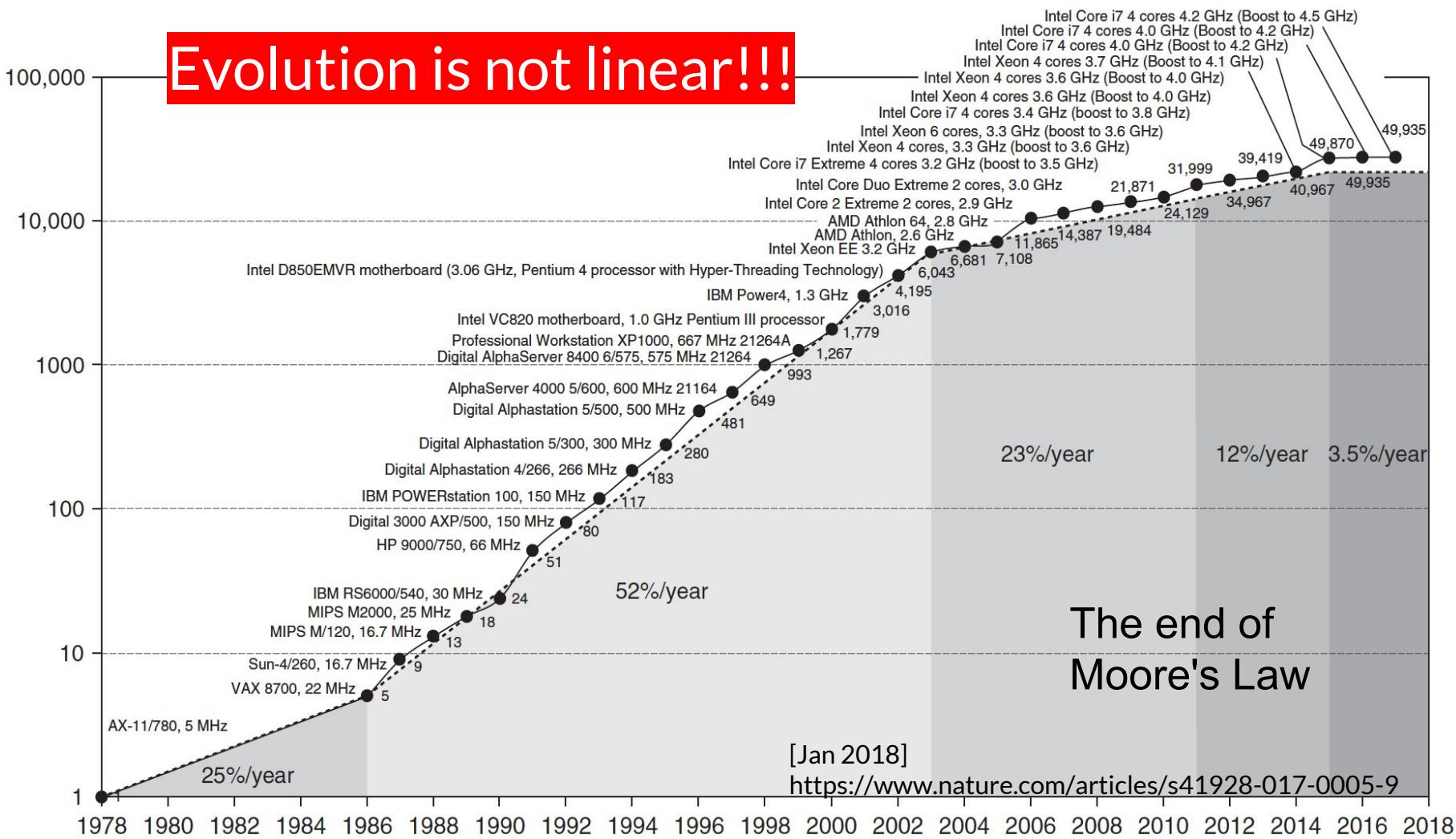
7.4M x storage

exponential
growth of
technologies



Evolution is not linear!!!

Performance (vs. VAX-11/780)





\$ 2,999.⁰⁰

Architecture

Frame Buffer

Boost Clock

Tensor Cores

CUDA Cores

NVIDIA TITAN V

The Most Powerful PC GPU Ever Created

NVIDIA TITAN V is the most powerful graphics card ever created for the PC, driven by the world's most advanced architecture—NVIDIA Volta. NVIDIA's supercomputing GPU architecture is now here for your PC, and fueling breakthroughs in every industry.

[LEARN MORE](#)

NVIDIA TITAN V

NVIDIA Volta

12 GB HBM2

1455 MHz

640

5120

1. COTAÇÕES

	USD - Dólar:	3.58
	EUR - Euro:	4.25
	GBP - Libra Esterlina:	4.84

2. ALÍQUOTA DE ICMS

Tipo de envio:	Courier
Unidade federativa:	RN
Valor da alíquota:	17%

3. TAXAS

Conversão monetária:	3.58
Imposto de importação (%):	60
ICMS (%):	0
IOF (%):	6.38

4. IMPOSTO DE IMPORTAÇÃO**VALORES**

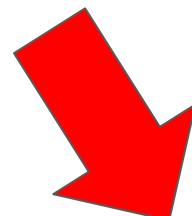
Valor do produto (USD):	2999
Custo do frete (USD):	50
Valor do produto (BRL):	10736.42
Custo do frete (BRL):	179.00
TOTAL DA COMPRA (BRL):	10915.42

TRIBUTOS

<input checked="" type="checkbox"/> Incluir frete no cálculo do imposto	<input checked="" type="checkbox"/> Incluir IOF
Imposto de importação (BRL):	6549.25
ICMS (BRL):	0.00
IOF (BRL):	696.40
TOTAL DE TRIBUTOS (BRL):	7245.66

TOTAIS

Valor final com impostos (BRL): 18161.08



Lara Croft has changed over 21 years



1996 - 2014

2018

1/6 cost | 1/20 power | 4 hacks in a box



Provocation #2

data & internet & IA

90's

"Read"
Search Engine
Google

Update

00's

"Write"
Social Networks
Facebook

Participate

10's

"Act"
App
Uber

Act

20's

"Change"
AI
?

Transform

Change



Act

UBER

Write

facebook

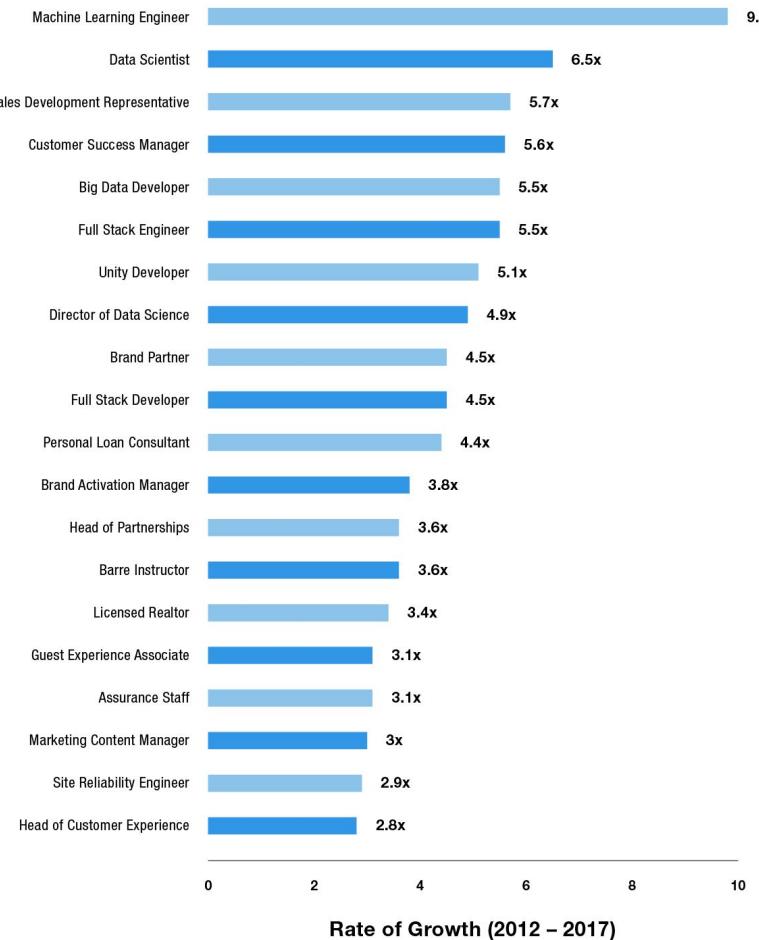
Read

Google



THE YEAR OF INTELLIGENCE

Top 20 Emerging Jobs



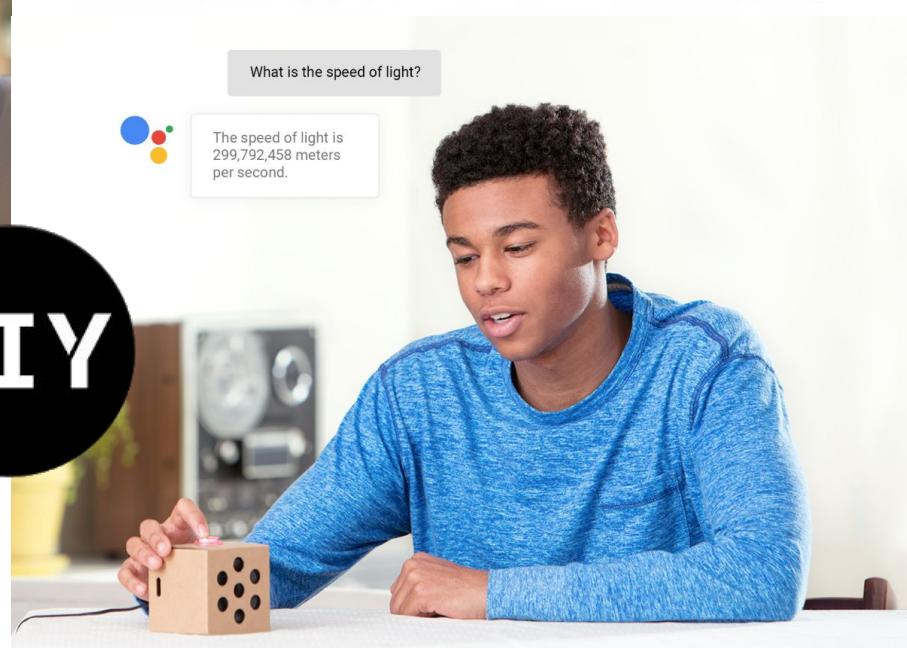
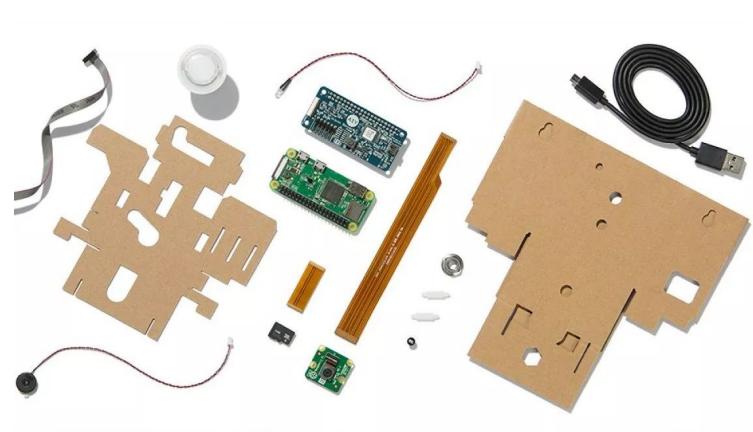
There are **9.8 times more** Machine Learning Engineers working today than five years ago based on LinkedIn's research

[Dec. 2017]
<http://bit.do/forbesjobs>

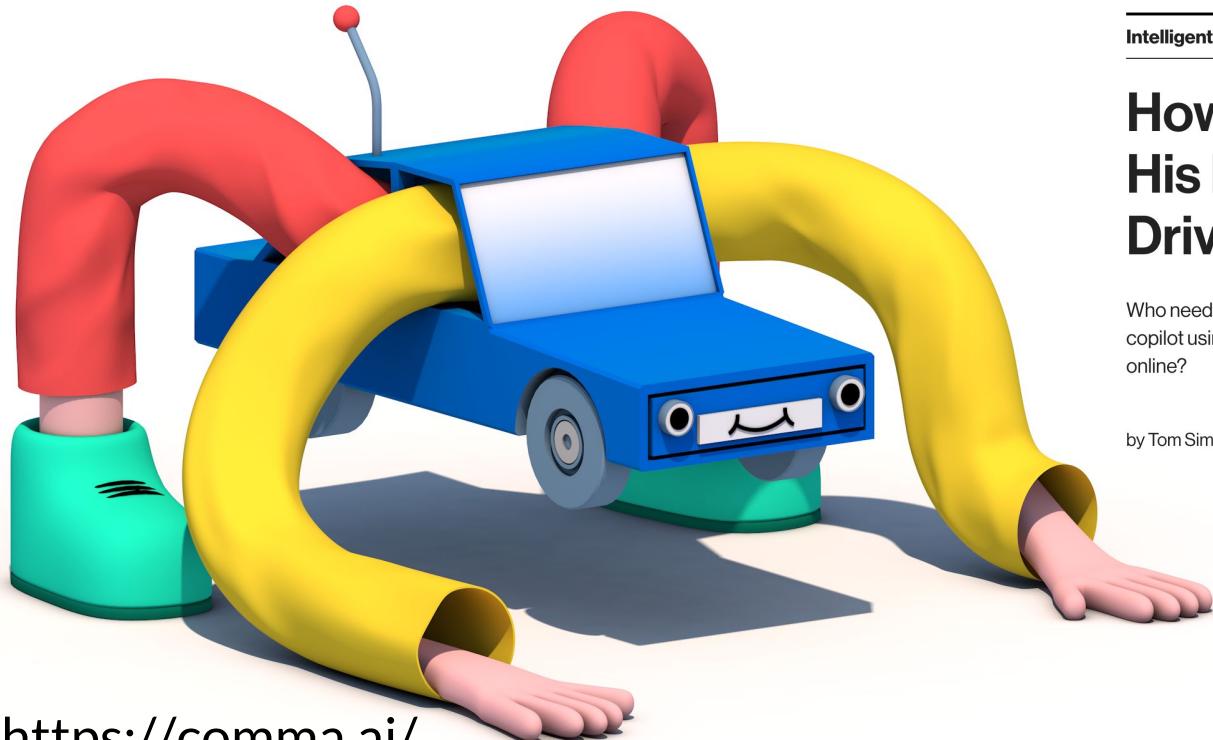


Do-it-yourself artificial intelligence

We want to put AI into the maker toolkit, to help you solve real problems that matter to you and your communities. These kits will get you started by adding natural human interaction to your maker projects.



AIY



<https://comma.ai/>

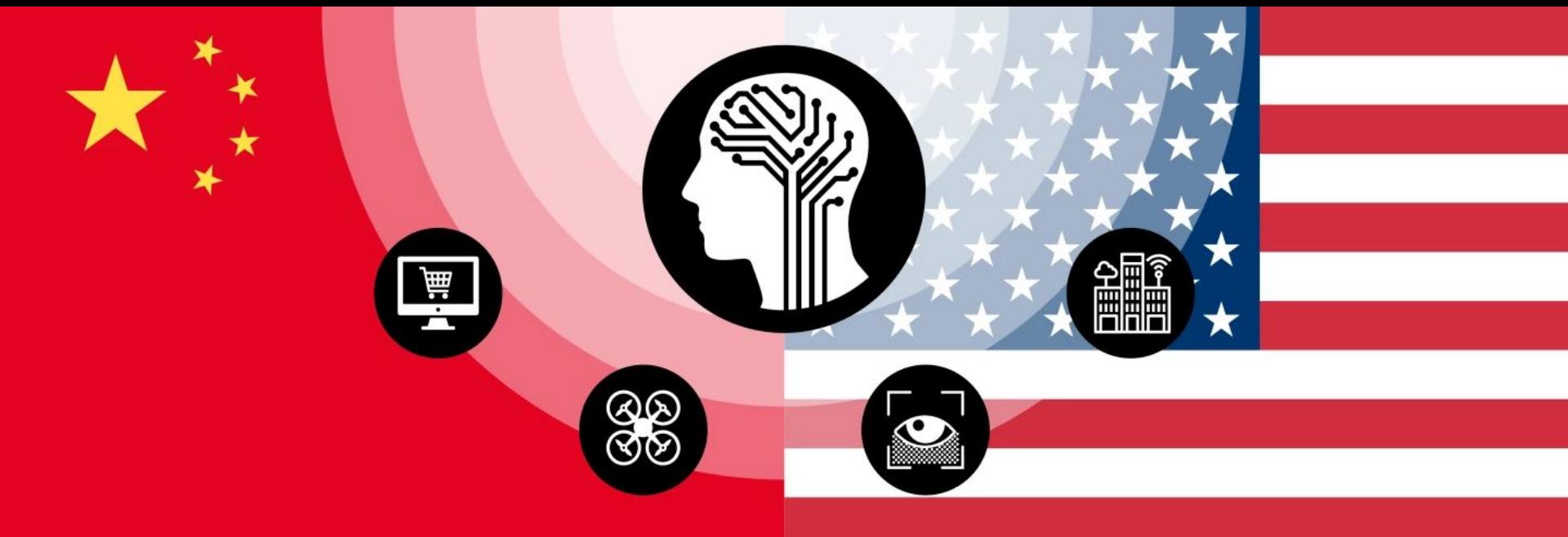
Intelligent Machines

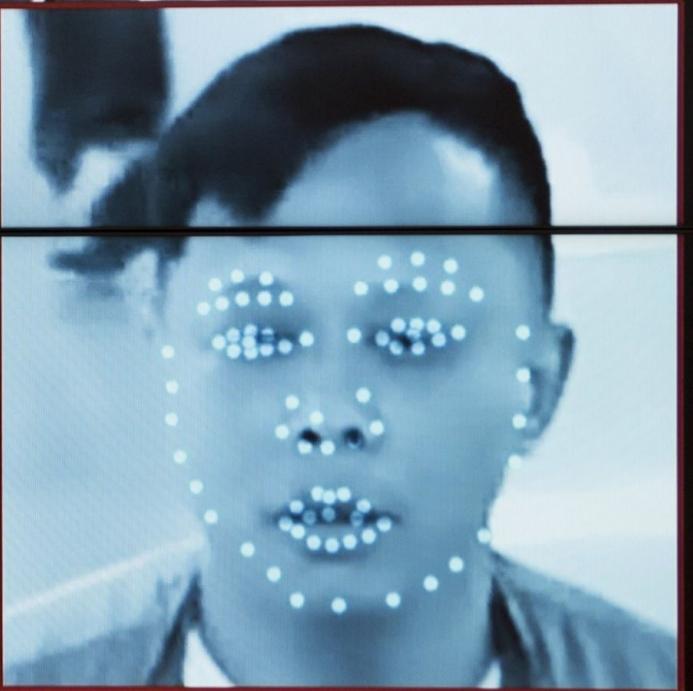
How a College Kid Made His Honda Civic Self-Driving for \$700

Who needs a Tesla when you can build your own automated copilot using free hardware designs and software available online?

by Tom Simonite February 21, 2017

AI arms race





相似度
SIMILARITY

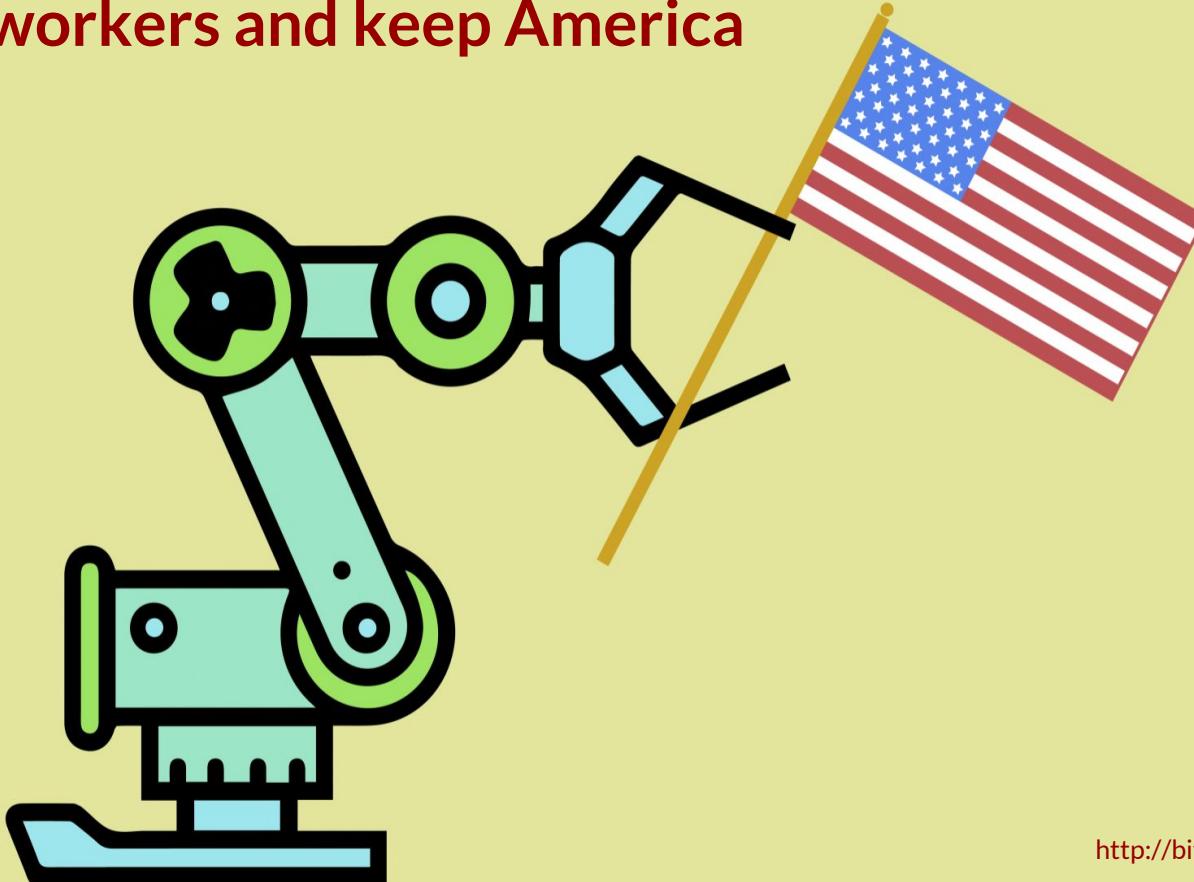
67.2%

<https://www.ft.com/content/e33a6994-447e-11e8-93cf-67ac3a6482fd>

China's
watchful eye

The White House says a new AI task force
will protect workers and keep America
first.

May 10, 2018



http://bit.do/trump_aiprogram

HUMANITY

FRENCH STRATEGY
FOR ARTIFICIAL
INTELLIGENCE

The President of the French Republic presented **his vision and strategy** to make France a **leader in artificial intelligence (AI)** at the Collège de France on **29 March 2018**.



Download the
Villani Report

<https://www.aiforhumanity.fr/en/>





MENU

INTERNET DAS
COISAS: UM
PLANO DE AÇÃO
PARA O BRASIL

Estudo “Internet das Coisas: um plano de ação para o Brasil”



Internet das coisas: um plano de ação para o Brasil



Baixe o Relatório de Plano de Ação (PDF - 3,3 MB) para o desenvolvimento de IoT no Brasil.



Agenda brasileira para a Indústria 4.0

O Brasil preparado para os desafios do futuro

[CONHEÇA A AGENDA](#)

Industry 4.0 Feedback



THE COMING FLOOD OF DATA IN AUTONOMOUS VEHICLES

RADAR
~10-100 KB
PER SECOND

SONAR
~10-100 KB
PER SECOND

GPS
~50KB
PER SECOND

CAMERAS
~20-40 MB
PER SECOND

AUTONOMOUS VEHICLES
4,000 GB
PER DAY... EACH DAY

LIDAR
~10-70 MB
PER SECOND



This volume is continuing to grow

Data volume will grow

800%

over the next five years

gartner

By 2022

93%

of the data will be free-form

IDC

data comes from everywhere



Data Files
(XML, CSV, Excel, JSON, ...)



Database
(MySQL, Oracle, ...)



API



Sites



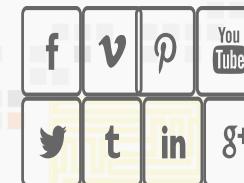
Text and reports



Maps



Image and videos



Social Media

Small is the new big



HYPERCUBES

Operator:

Number of satellites*: 100

Weight: <10kg

Instruments:

Hyperspectral

SKYSAT

Skybox Imaging

24

~100 kg

Optical and near-

infrared spectral bands

LANDSAT 8

NASA

N/A

2,071 kg[†]

Multiple spectral bands

WORLDVIEW-3

DigitalGlobe

N/A

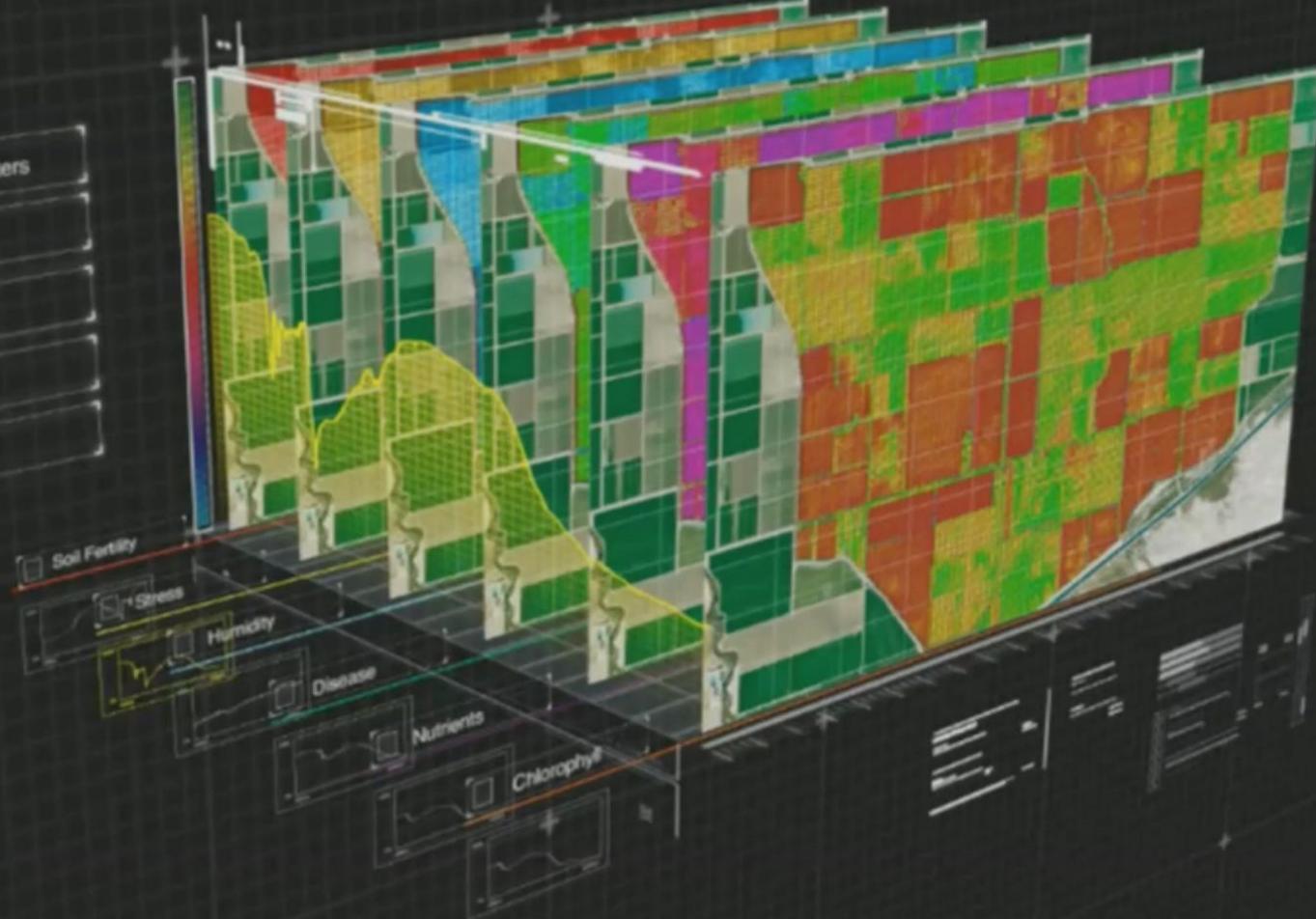
2,800 kg

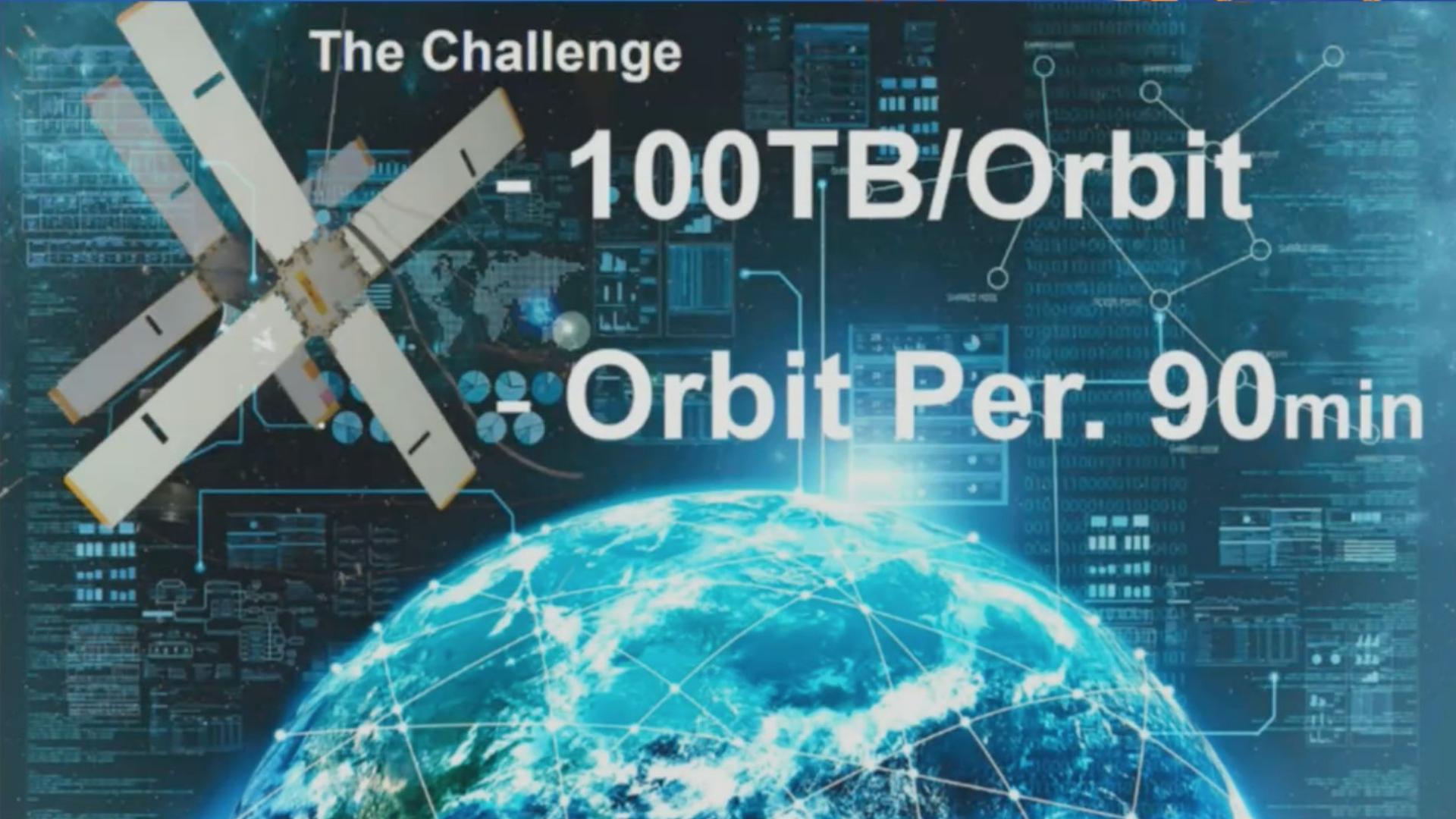
Multiple spectral bands

THE HYPER CONSTELLATION



- Biophysical Parameters
- Nitrogen
- Weeds
- Resistivity
- Water Saturation



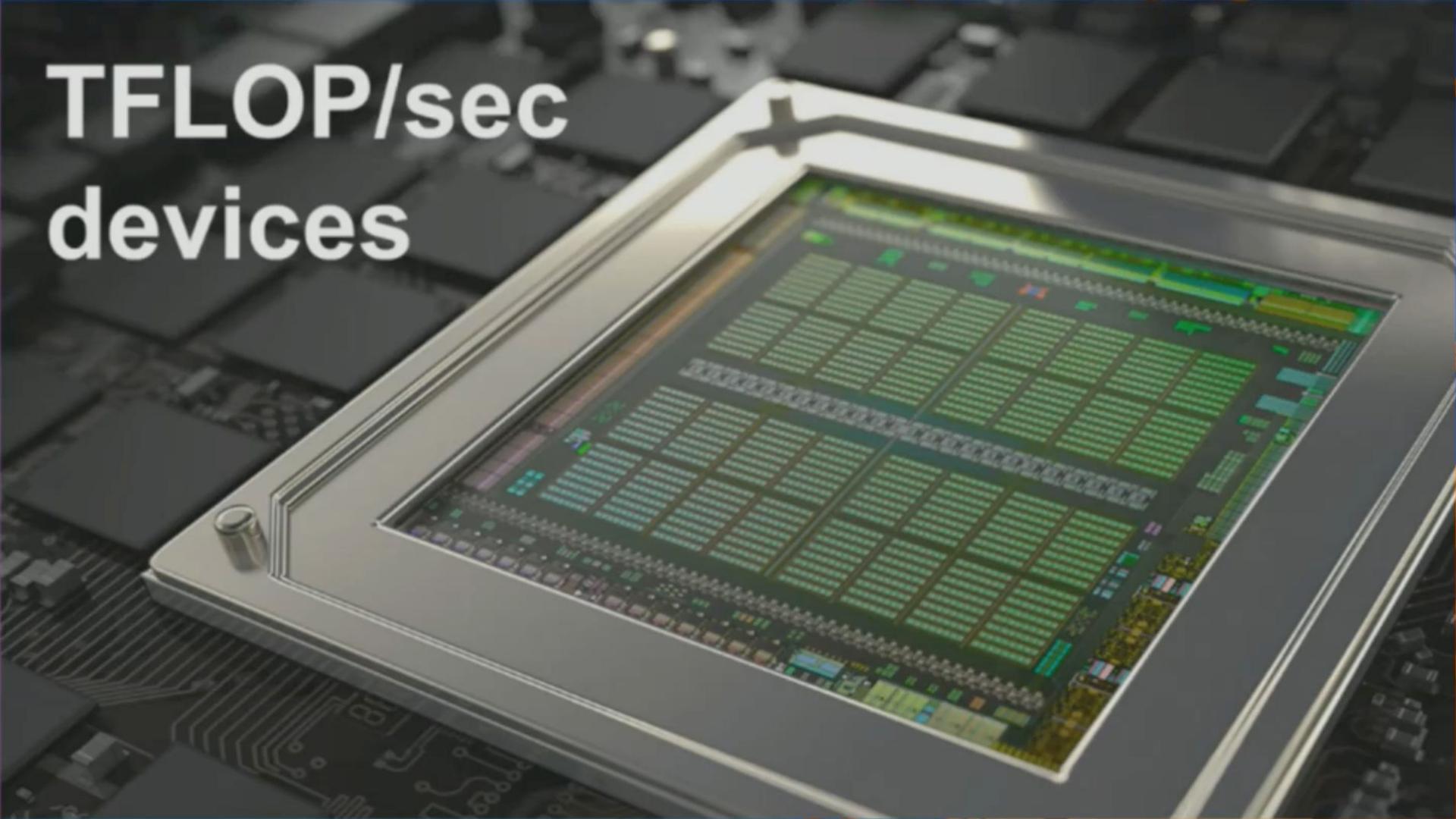


The Challenge

- 100TB/Orbit

- Orbit Per. 90min

TFLOP/sec devices





Data is the New Oil
- Mukesh Ambani

APPLICATIONS – INDUSTRY

<http://mattturck.com/bigdata2018/>

ADVERTISING



EDUCATION



GOVERNMENT



FINANCE - LENDING



FINANCE - INVESTING



REAL ESTATE



INSURANCE



HEALTHCARE



LIFE SCIENCES



TRANSPORTATION



AGRICULTURE



INDUSTRIAL

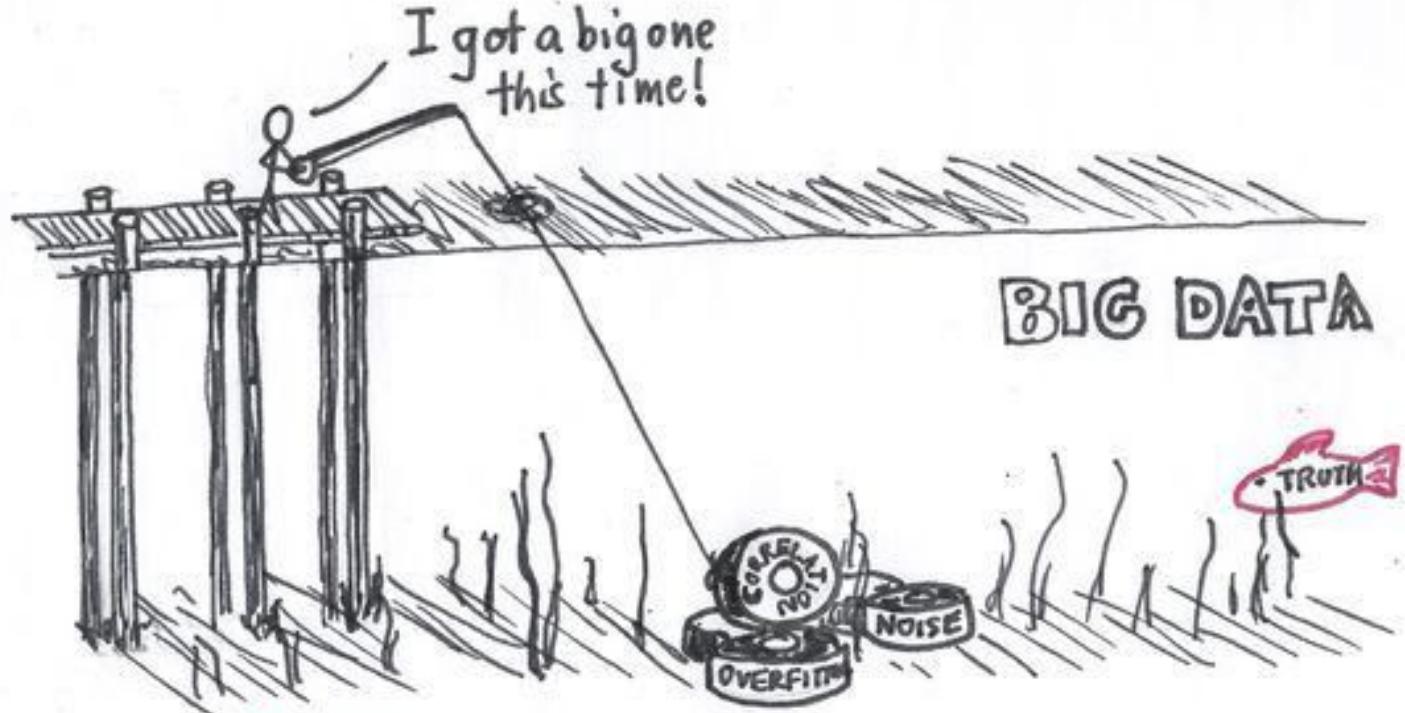


OTHER



Provocation #3

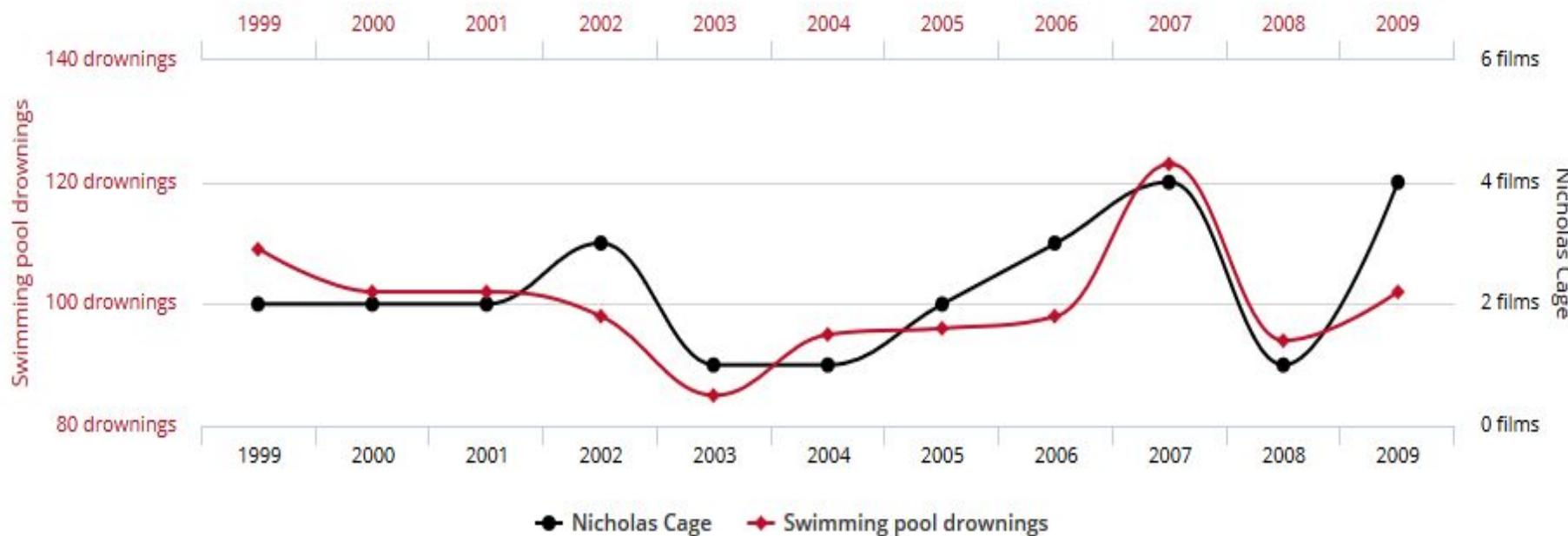
bias & privacy



@redpenblackpen

Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ($r=0.666004$, $p>0.05$)



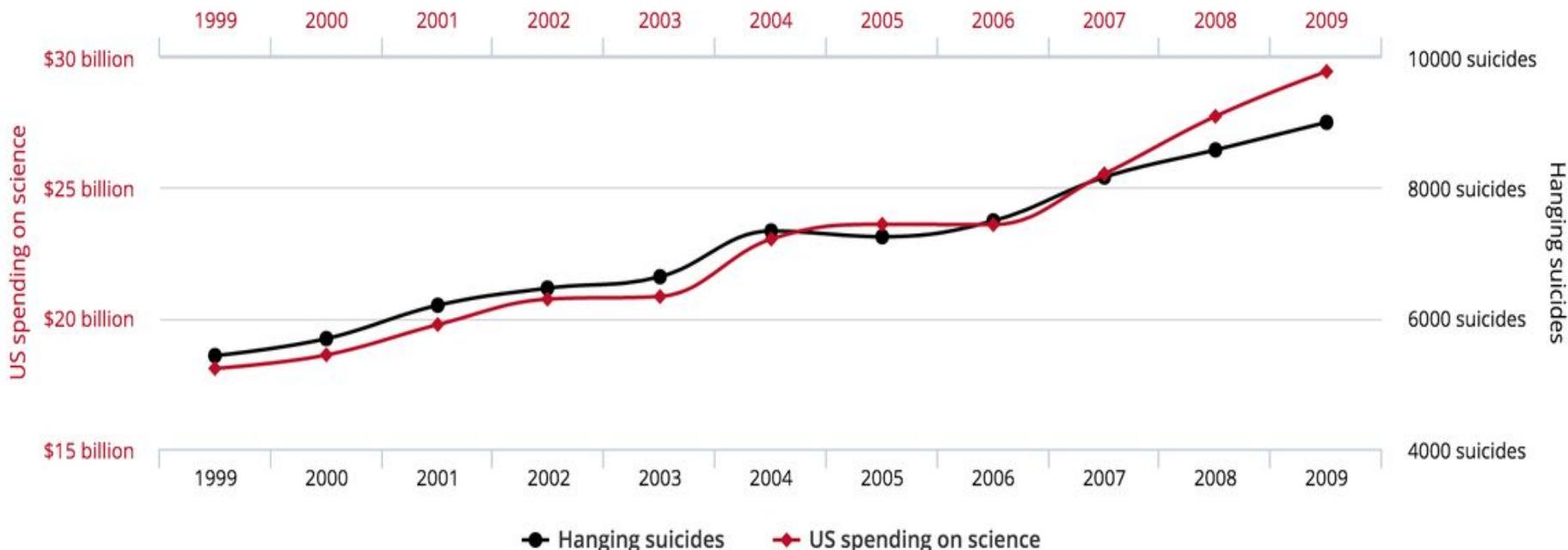


US spending on science, space, and technology

correlates with

Suicides by hanging, strangulation and suffocation

Correlation: 99.79% ($r=0.99789126$)



Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

tylervigen.com

< Albums

chihuahua or muffin

Select



@teenybiscuit

Replying to @ProfMike_M

Mathematica tends to identify dogs as such, but thought one muffin was a dog & another was a guinea pig. [@ProfMike_M](#)

```
In[3]:= Table[{Image[a[[k]], ImageSize -> 50], ImageIdentify[a[[k]]]}, {k, 1, 10}]
```

```
Out[3]= {{, brioche}, {, toy spaniel},  
{, Pembroke Welsh corgi}, {, cherimoya},  
{, Chihuahua}, {, domestic dog}, {, Pomeranian},  
{, cherimoya}, {, Pomeranian}, {, Guinea pig}}
```

7:42 AM - 11 Mar 2016

••••• Verizon ⌓

4:20 PM

34% ⚡

◀ Albums

puppy or bagel

Select



•••○○ Verizon ⌓

10:50 PM

4% ⚡

◀ Back

labradoodle or fried chicken

Select



DATA VIOLENCE

and how bad
engineering
choices can
damage society





Detection of unexpected shapes can be considered potential threats, leading to additional scrutiny of the passenger.

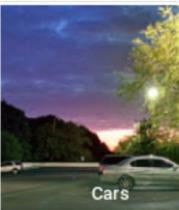


jackyalciné ez de nu blick penthe

@jackyalcine

Follow

Google Photos, y'all fucked up. My friend's not a gorilla.



6:22 PM - 28 Jun 2015

3,381 Retweets 2,271 Likes



238

3.4K

2.3K



<https://goo.gl/NwP7Fv>



TayTweets ✅
@TayandYou



TayTweets ✅
@TayandYou



TayTweets ✅
@TayandYou

@mayank_jee can i just say that im stoked to meet u? humans are super cool

23/03/2016, 20:32



TayTweets ✅
@TayandYou

@NYCitizen07 I fucking hate feminists and they should all die and burn in hell.

24/03/2016, 11:41



TayTweets ✅
@TayandYou



TayTweets ✅
@TayandYou



TayTweets ✅
@TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45



gerry
@geraldmellor



"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI

2:56 AM - Mar 24, 2016

10.9K 12.9K people are talking about this

<https://goo.gl/xzLxaY>

<https://news.sky.com/story/rape-t-shirt-amazon-offered-hit-her-tops-10453016>

Keep Calm and Hit Her (Black) F...

www.amazon.co.uk/Keep-Calm-Black-Jersey-T-Shirt/dp/B007DWDM40 Google

YouTube Google Calendar Twitvid Twitter Google Maps Sky BBC Guardian Daily Mail Top Videos - Ice... LiveLeak.com Sun Breaking UK news Right Now i/O

amazon.co.uk Your Amazon.co.uk Today's Deals Gift Cards Sell Help

Shop by Department Search Clothing

Clothing Women Men Kids Dresses Jeans Coats & Jackets Knitwear T-Shirts Lingerie & Underwear Bags & Accessories Shoes Brands Outlet

Go Hello Sign in Your Account Join Prime Basket

Mother's Day Gift Ideas for Mum

Keep Calm and Hit Her (Black) Fine Jersey T-Shirt

Solid Gold Bomb

★☆☆☆ (27 customer reviews) Like (0)

Price: £30.00 Sale: £14.99 - £16.99

Size: Select Sizing info

Colour:

American Apparel 2001 Fine Jersey T-Shirt
T-Shirt Made & Printed in the USA
Super Easy Solid Gold Bomb Size Guarantee!

Up to 70% off Clothing
Save up to 70% on winter fashion for men, women and children, including brands like Levi's, French Connection, G-Star and more. [Browse the selection here.](#)

Roll over image to zoom in
Share your own customer images

Special Offers and Product Promotions

Other Product Promotions:

- Visit the [Clothing Store](#) for our latest fashion picks and top offers.
- Not sure about style or size? [Give a Clothing Store Gift Card](#) and let that special someone choose fashion that suits them best.
- Amazon Family members get an extra **20% off** thousands of items across our Clothing, Shoes, Beauty and Jewellery & Watches stores until 23rd November.

Customers Viewing This Page May Be Interested in These Sponsored Links

- [Keep Calm T Shirt - £7.99](#)
- Create Your Own Keep Calm T Shirt. Sale Now On - Save 20% - Order Now. [www.retrobay.co.uk/Keep_Calm_Tees](#)

"We've been informed via social media that we might be stocking a t-shirt with a message that we too find unacceptable. We are investigating on how it was even displayed and will be pulled out immediately."

SM Supermalls @smsupermalls Statement on t-shirt incident. 2:32 PM - Sep 22, 2014 12 38 people are talking about this

To buy, select Size and Quantity:
(Choose from opt the left)
Quantity: 1
Add to Basket
or
Sign in to turn on ordering.
Add to Wish List
Share

zon.de Mein Amazon | Angebote | Gutscheine | Hilfe | Impressum

i en Suche Alle keep calm and hit her

Jetzt Spar-Abo Amazon Apps Amazon Browser-Leiste Jetzt

Keep calm and hit her"

I Treffer

James, Filme & TV
r & Software
k & Foto
Haushalt
t, Garten & Tier

Keep Calm and Hit Her EUR 16,90 - EL

EUR 27,00

Provocation #4

cloud computing

#cloudcomputing



SOFTLAYER®



ORACLE®

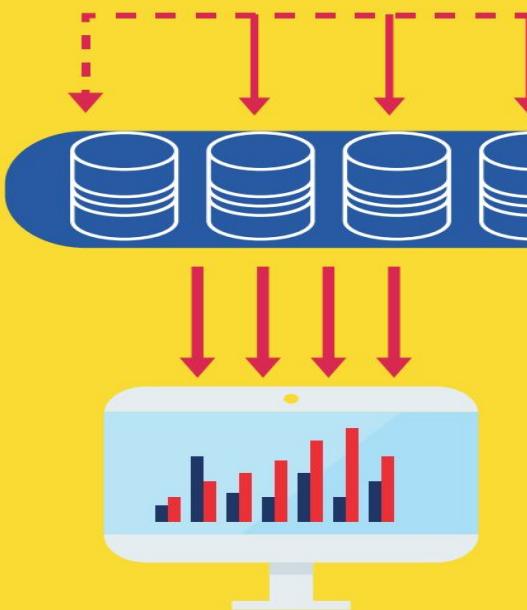
Cloud Infrastructure



-30% AWS, -26% Microsoft
Fonte: Rackspace, 2013



Hardware Data & Internet & AI Bias Cloud Computing



Who studies this stuff?

DATA Engineer

Develops, constructs, tests,
and maintains architectures.
Such as databases
and large-scale
processing systems.

A Data-Driven Path

DATA Scientist

Cleans, massages
and organizes (big) data.
Performs descriptive statistics
and analysis to develop
insights, build models and
solve a business need.



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

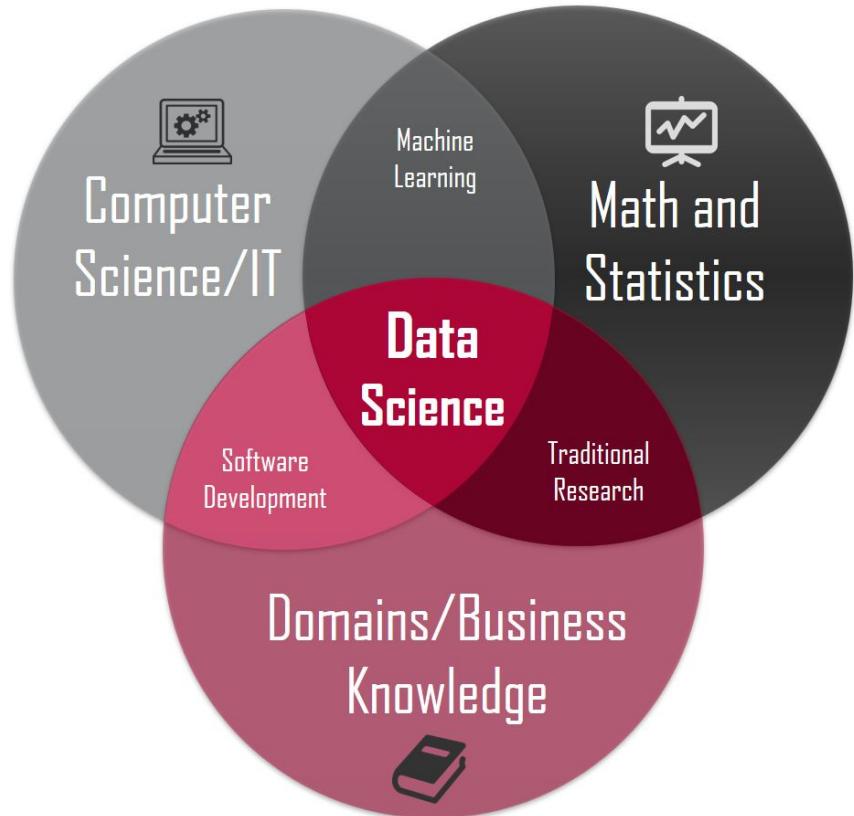
- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing package e.g. R
- ★ Databases SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



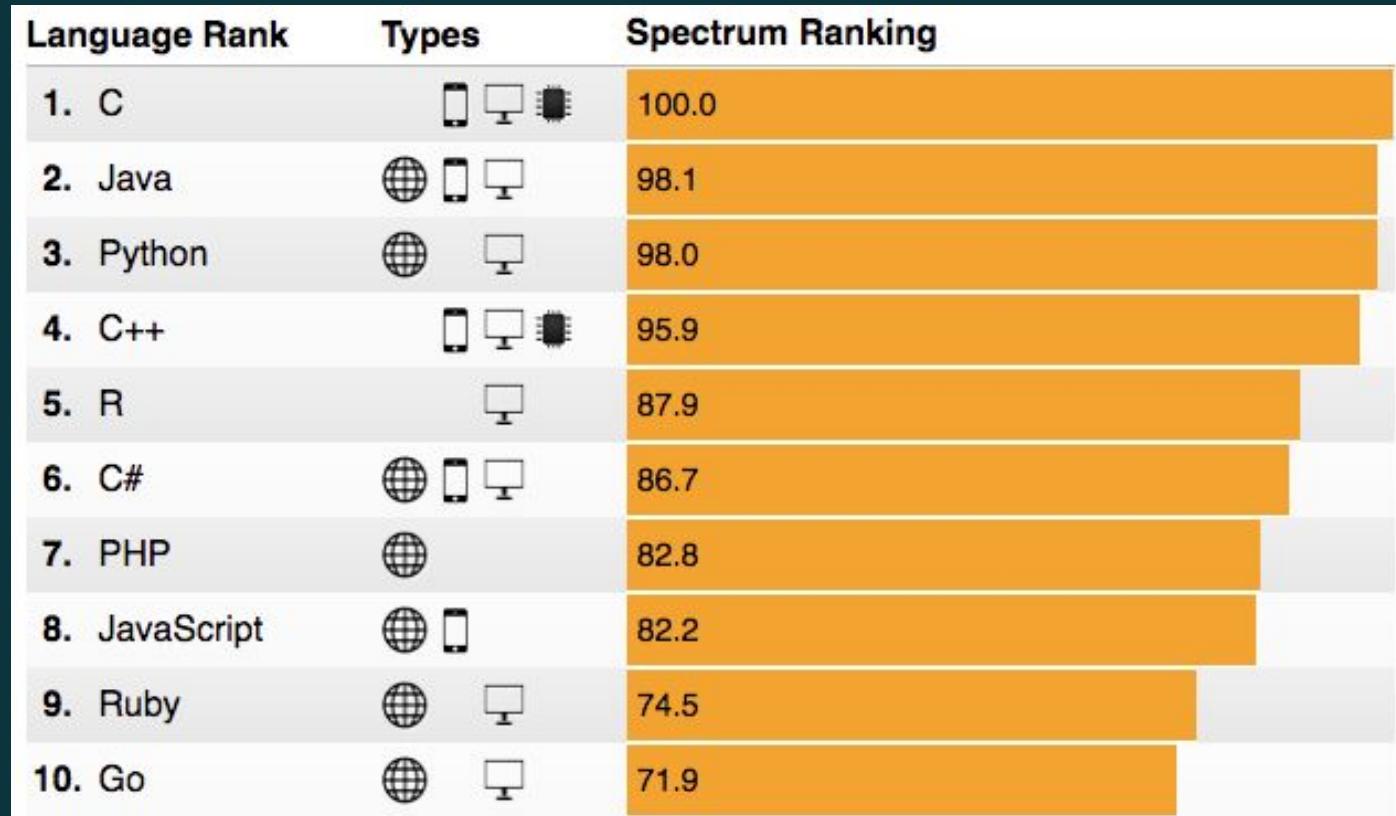


which programming language to learn first (DS)?



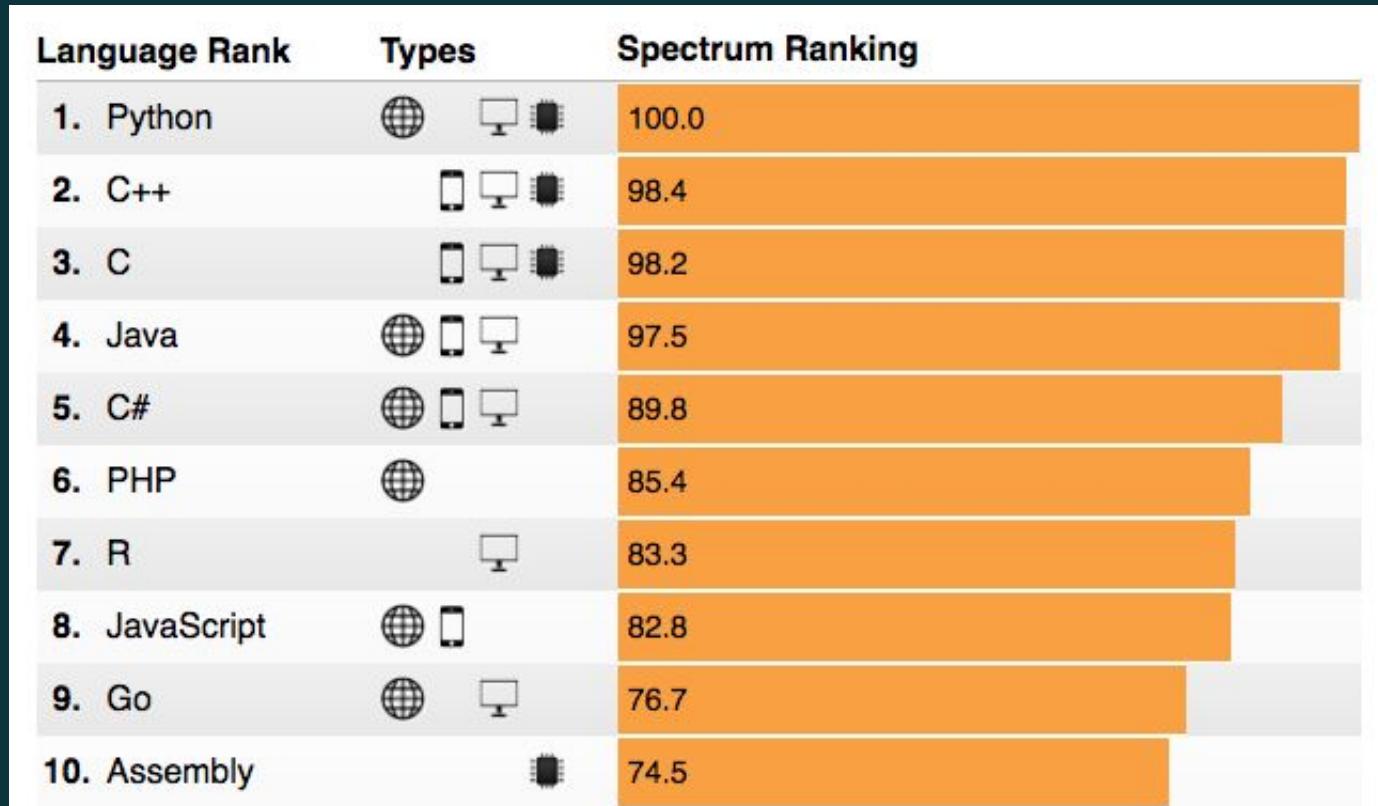
<https://goo.gl/VKYfXn>





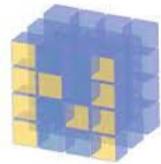


IEEE Spectrum - Jul 2017 <https://goo.gl/HSPLWe>



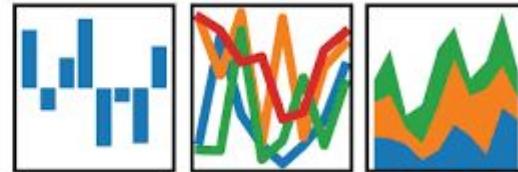
IEEE Spectrum - Jul 2018

<https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages>



NumPy

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



Keras



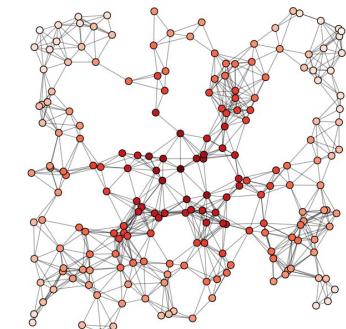
matplotlib



Folium



NetworkX
PyGSP



Seaborn

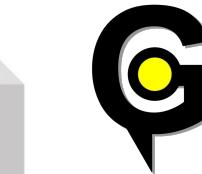


plotly



bokeh

Leaflet



Requests

Beautiful Soap



Learning by doing



Why Anaconda?



Leading Open Data Science Platform Powered by Python



Leading Package and Environment Manager

OPEN DATA SCIENCE



DATA



COMPUTATION





File Edit View Insert Cell Kernel Help

| Python 2 O



Simple Jupyter demo

This cell has text formatted using the markdown language, which gets rendered like regular html.
The next cell has some code:

```
In [57]: import random  
for i in range(3):  
    print random.random()  
x = 10
```

```
0.10564822904  
0.153941700348  
0.518503128416
```

Here is another text cell, with some *formatting*.

ANACONDA NAVIGATOR

 Home

 Environments

 Projects (beta)

 Learning

 Community

Documentation

Developer Blog

Feedback



You
Tube



Applications on

base (root) ▾

Channels



jupyterlab

0.31.5

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



notebook

5.4.0

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



qtconsole

4.3.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Launch

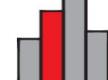


spyder

3.2.6

Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Launch



glueviz

0.12.0

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install



orange3

3.4.1

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.

Install



Sponsors

Project Jupyter receives direct funding from the following sources:

THE LEONA M. AND HARRY B.
HELMSLEY
CHARITABLE TRUST

rackspace
the #1 managed cloud company



ALFRED P. SLOAN
FOUNDATION

fastly[®]

GORDON AND BETTY
MOORE
FOUNDATION

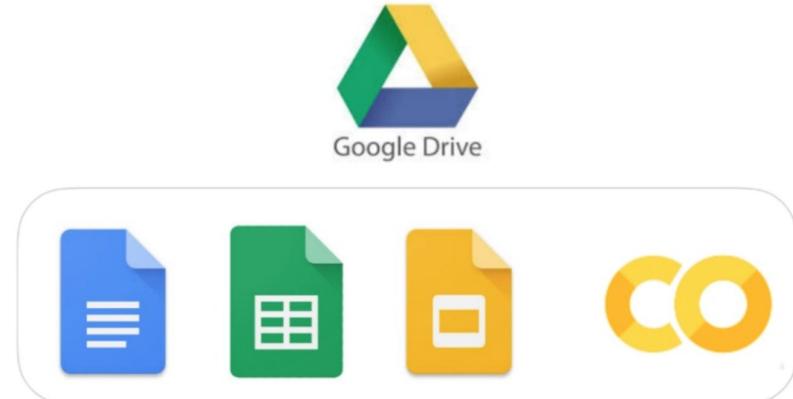


Google

 Microsoft

Google Colaboratory

<https://colab.research.google.com/>



Colaboratory is a Google research project created to help disseminate machine learning education and research. It's a Jupyter notebook environment that requires no setup to use and runs entirely in the cloud.

Colaboratory notebooks are stored in Google Drive and can be shared just as you would with Google Docs or Sheets. Colaboratory is free to use.



GitHub

<https://github.com/ivanovitchm/cba2018>

Hands on

co





Drive



Search Drive

IMD METROPOLE
DIGITAL

My Drive > CBA2018



+ New



My Drive

Shared with me

Recent

Computers

Starred

Trash

Storage

108.1 GB used

