UNIVERSITY OF TEXAS ARLINGTON
COLLEGE OF SCIENCE

DATA 1301
Introduction to Data Science

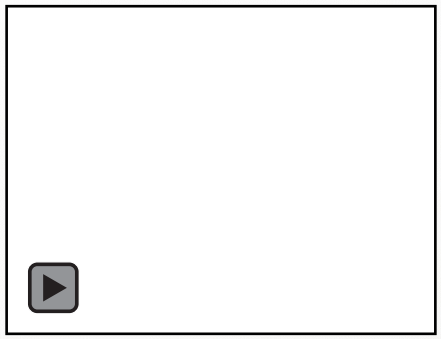
Student-Faculty connection day
The Pillars of Science

Amir Shahmoradi

Department of Physics / College of Science
Data Science Program / College of Science
The University of Texas
Arlington, Texas

The Computational and Data Science Lab @ Department of Physics / College of Science / UTA

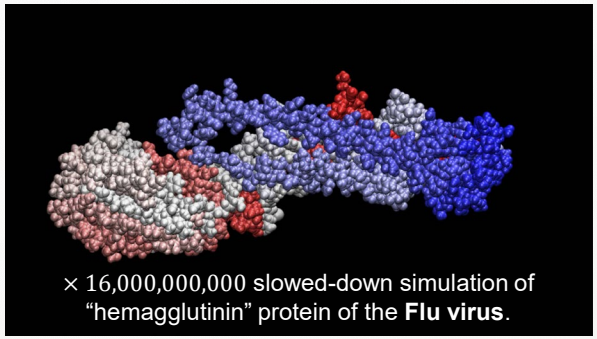
Join us @ **cdslab.org**



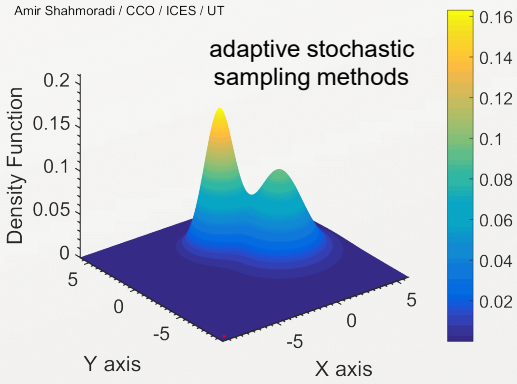
Physics of Gamma-Ray Bursts
The most powerful explosions in the universe

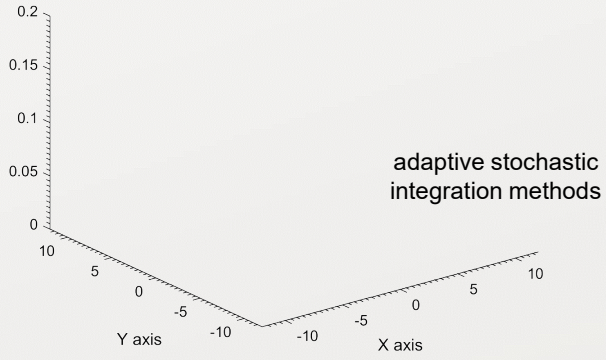
Open-source software development:
Machine Learning and Monte Carlo Methods

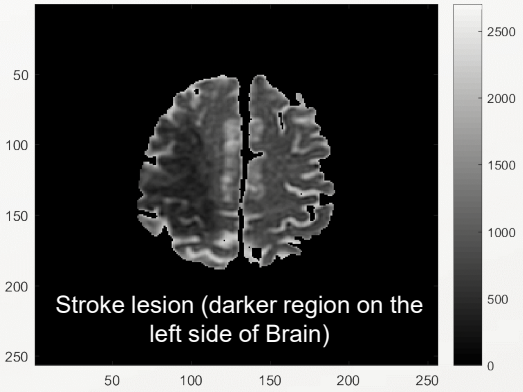
cdslab.org/pm

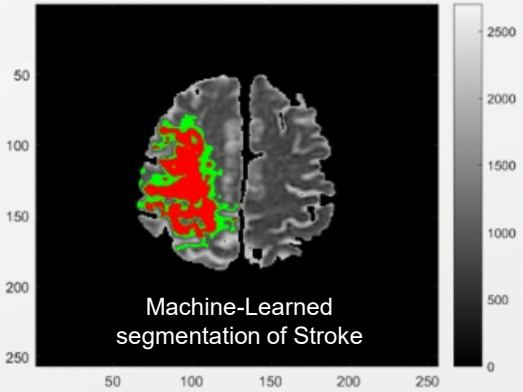


Bioinformatics / Biophysics









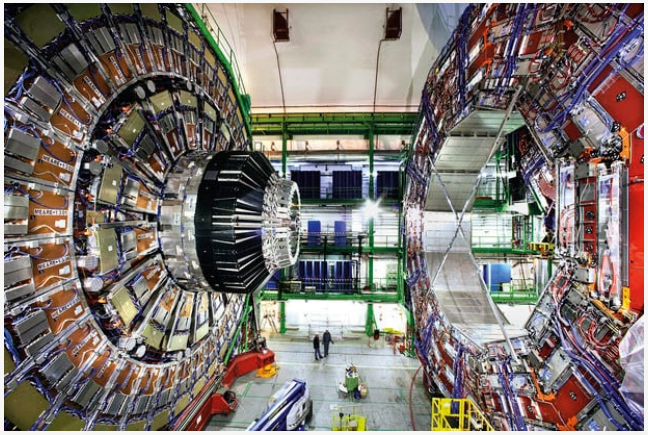
Biomedical Data Science

Traffic Engineering

Physicists' contributions to the foundations of Data Science

Particle Physicists

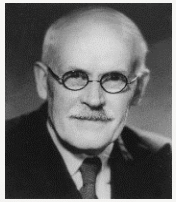
Pioneers in high-throughput Big Data and the Internet technology



Physicists / Astronomers

The founding fathers and strong advocates of
Bayesian Inference and
stochastic optimization / integration methods

Pierre Laplace
(1749 – 1827)
Astronomer / Mathematician



Harold Jeffreys
(1891 – 1989)
Astronomer / Geophysicist

Richard Cox
(1898 – 1991)
Physicist

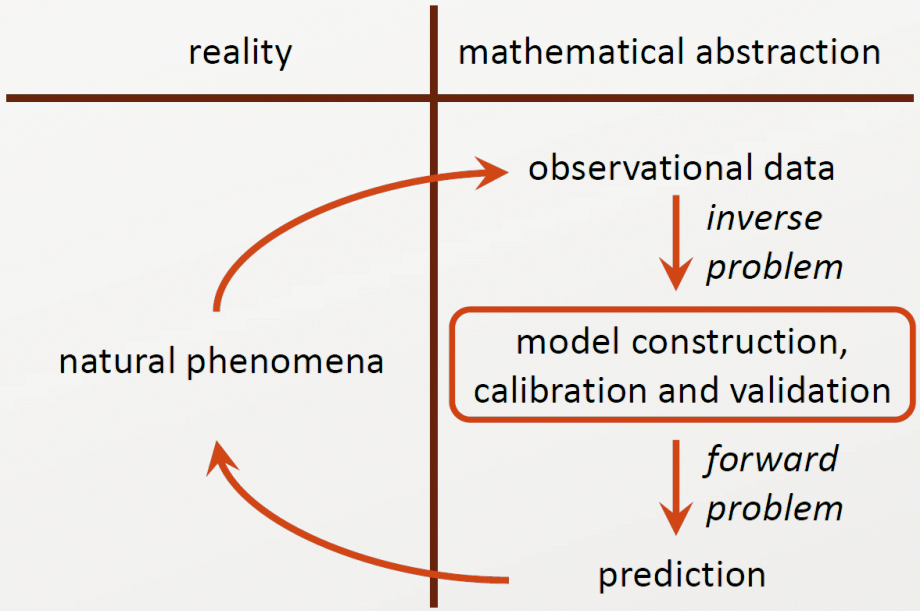
Edwin Jaynes
(1922 – 1998)
Physicist

Prior knowledge has a fundamental role
in inference along with Data
(via the Bayes' Rule)
e.g., hyperparameter tuning

The two classical pillars of science: Experiment and Theory

How do we make a scientific inference?

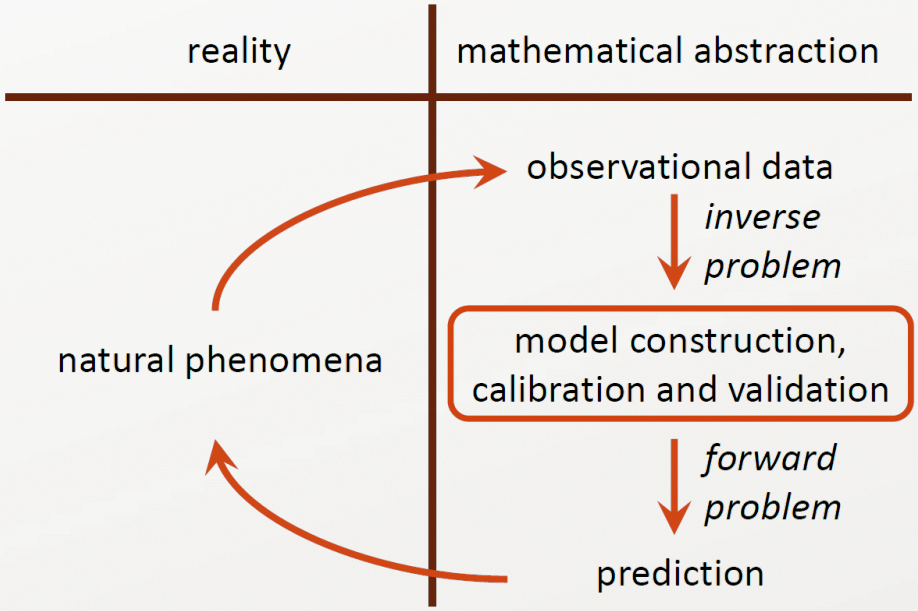
A very elementary depiction of the scientific method



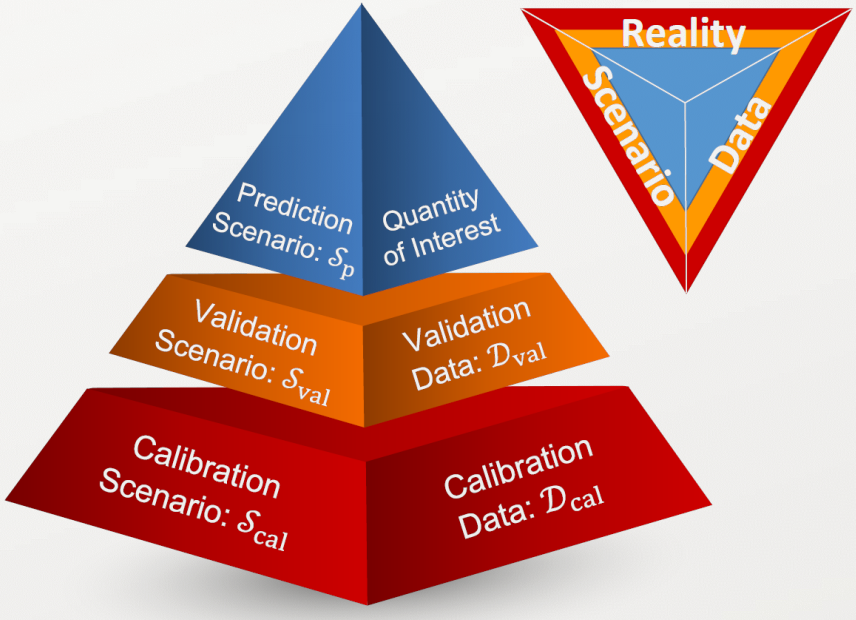
The two classical pillars of science: Experiment and Theory

How do we make a scientific inference?

A very elementary depiction of the scientific method



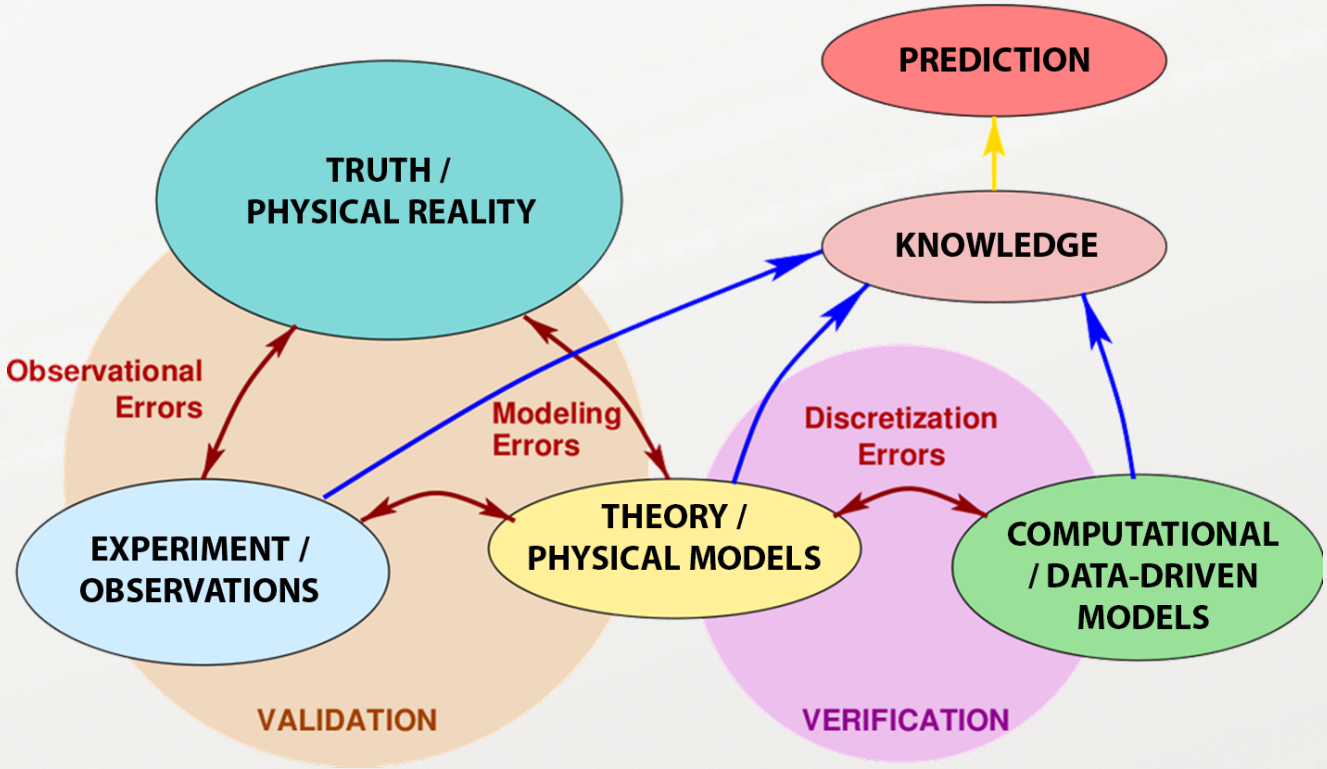
The prediction pyramid



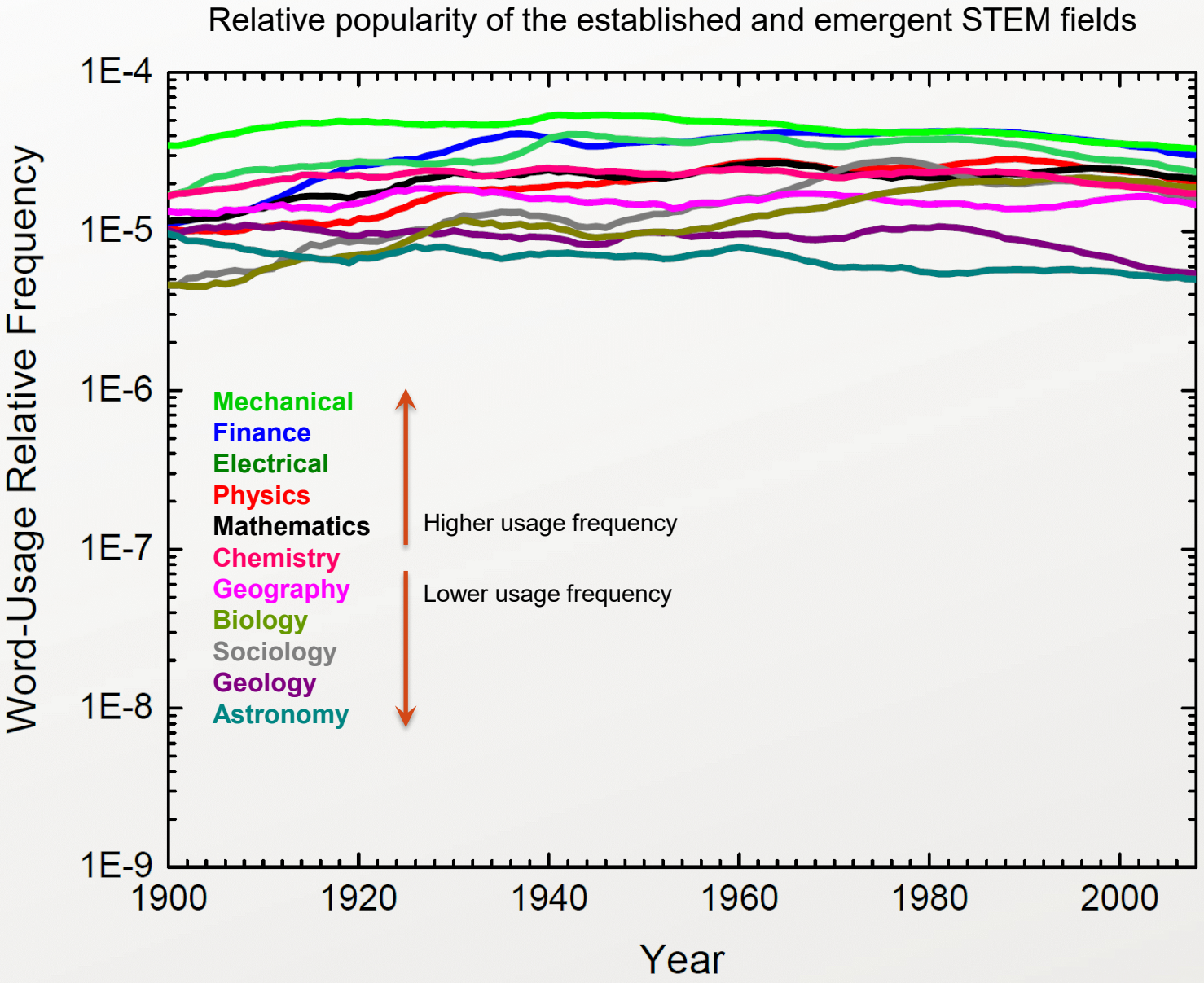
The three pillars of science: Experiment, Theory, Computation+Data

Three major roles of computational models in contemporary science

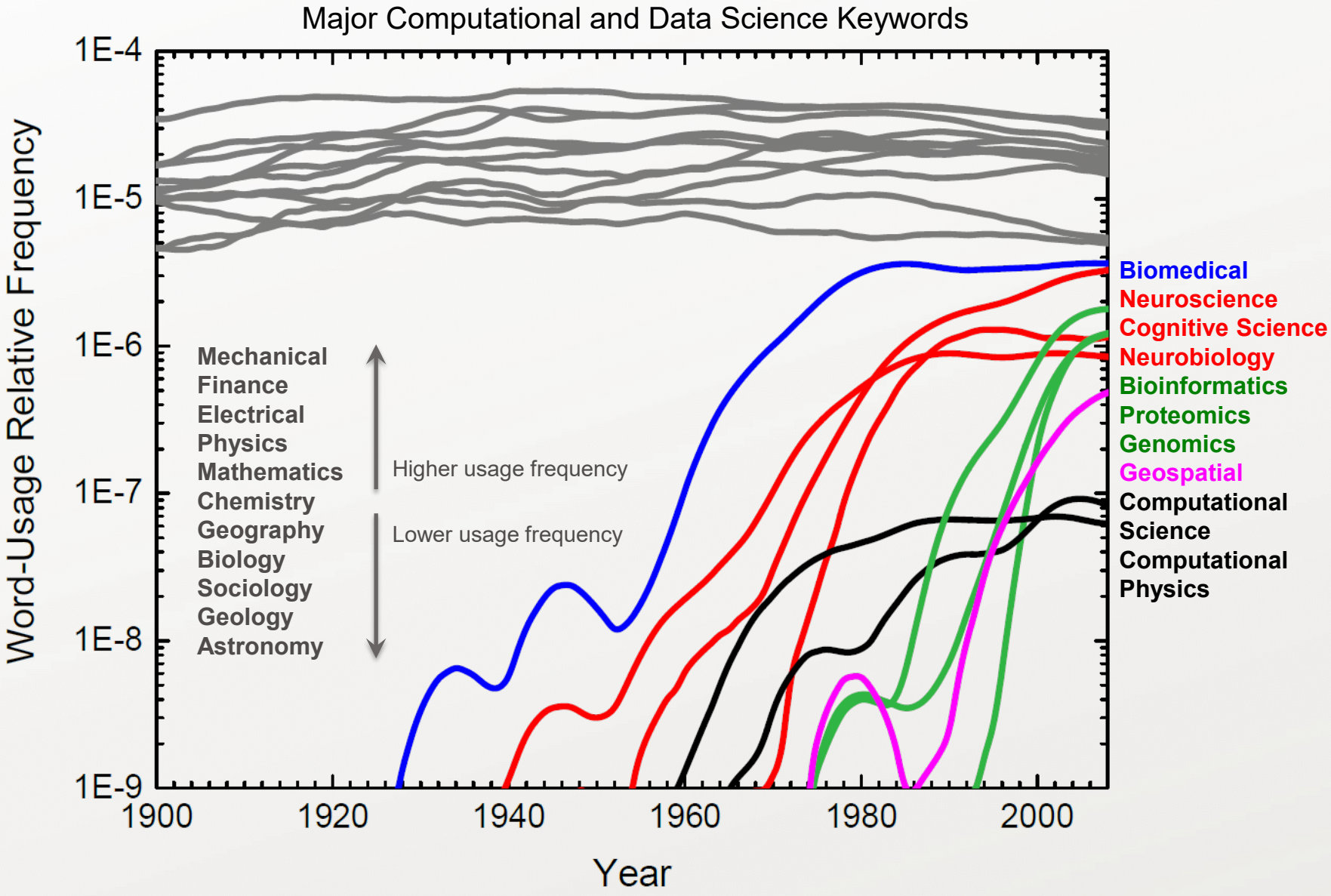
1. a workflow bridging data, hypothesis/theory, and prediction (**predictive computing**).
2. a substitute for experiment and observational data, where it is not available (**numerical simulation**).
3. a substitute for theory, where it is not available (**data-driven discovery via machine learning**, **deep learning, …**).



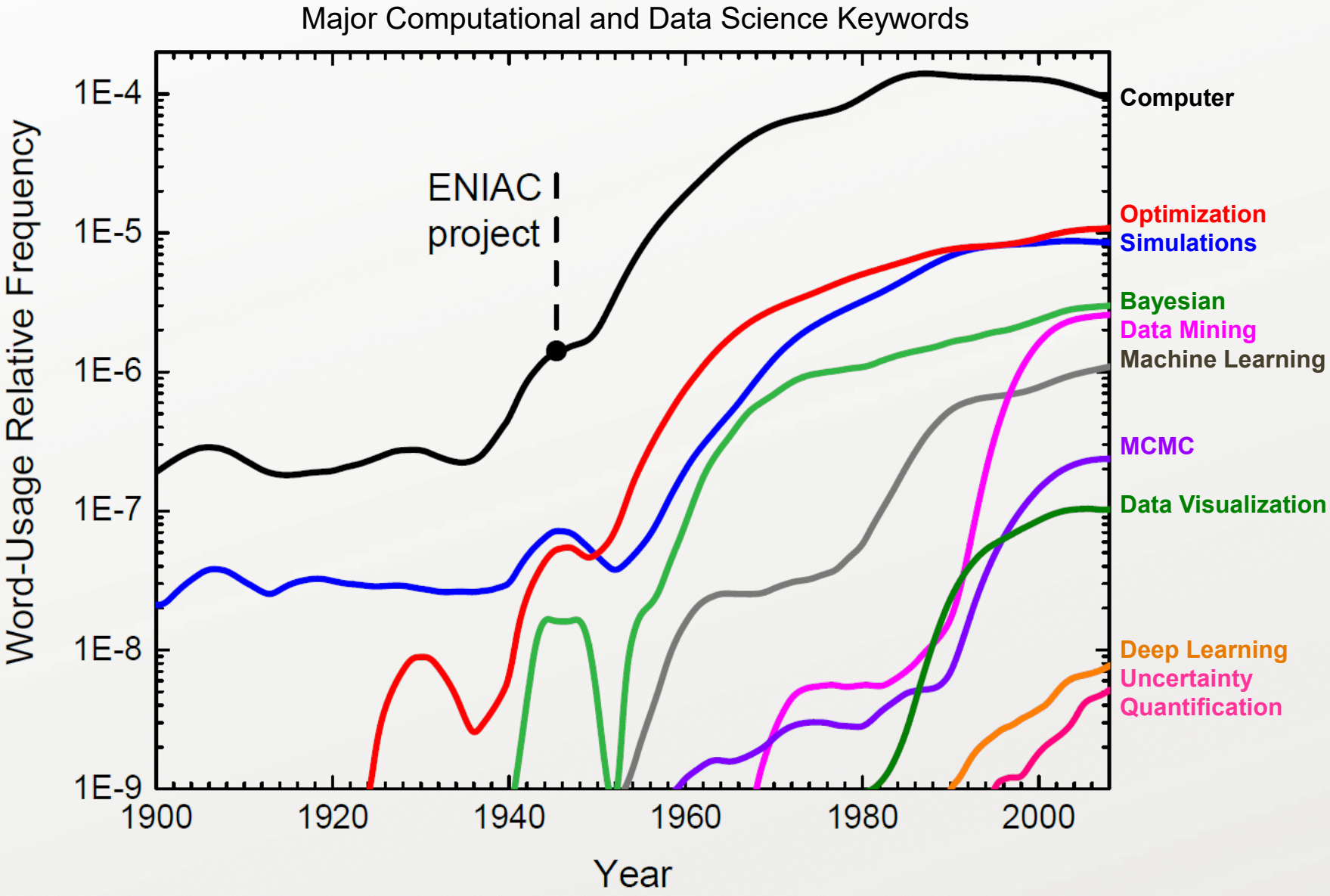
The emerging third pillar of science: Computational and Data Sciences



The emerging third pillar of science: Computational and Data Sciences



The emerging third pillar of science: Computational and Data Sciences



Data Science job market has grown by 700% over a half-decade

Skill	2013-2018 Growth
Machine Learning	809%
R	298%
Data Analysis	86%
Data Management	78%
SQL	45%

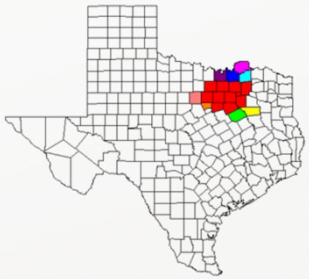
Role	2013-2018 Growth
Data Scientist	663%
Marketing Data Analyst	194%
Bioinformatician	75%
Financial Quantitative Analyst	57%

Data: Burning Glass Technologies 2019

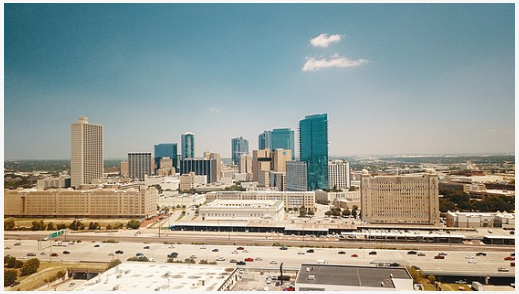
The Data Science Program of The University of Texas Arlington

UTA

- Established 1895
- Located at DFW metroplex
  - 8,000,000 population
  - 3rd-largest concentration of Fortune 500
  - the largest economic engine of Texas
- >42,000 students
- Second-largest institution within the University of Texas System
- 5th most-diverse campus in the US (2018)
- The UTA Data Science program
  - established by the College of Science
  - Undergraduate Major and Minor
  - Final approval in April 2020



Dallas-Fort-Worth Metroplex



Fort-Worth

Dallas

Arlington

UTA College of Science

- Biology
- Chemistry
- Environmental Science
- Mathematics
- Physics
- Psychology

Science & Engineering Innovation & Research Building

The UTA Data Science faculty

The UTA Data Science program is unique in the nation for its focus on Natural Sciences

Data Science Codes of the **Texas higher education** coordinating board:

- **30.7001 Data Science, General.**

A program that focuses on the **analysis of large-scale data sources** from the interdisciplinary perspectives of **applied statistics**, **computer science**, **data storage**, **data representation**, data modeling, mathematics, and statistics. Includes instruction in **computer algorithms**, **computer programming**, **data management**, data mining, information policy, information retrieval, mathematical modeling, quantitative analysis, statistics, trend spotting, and visual analytics.

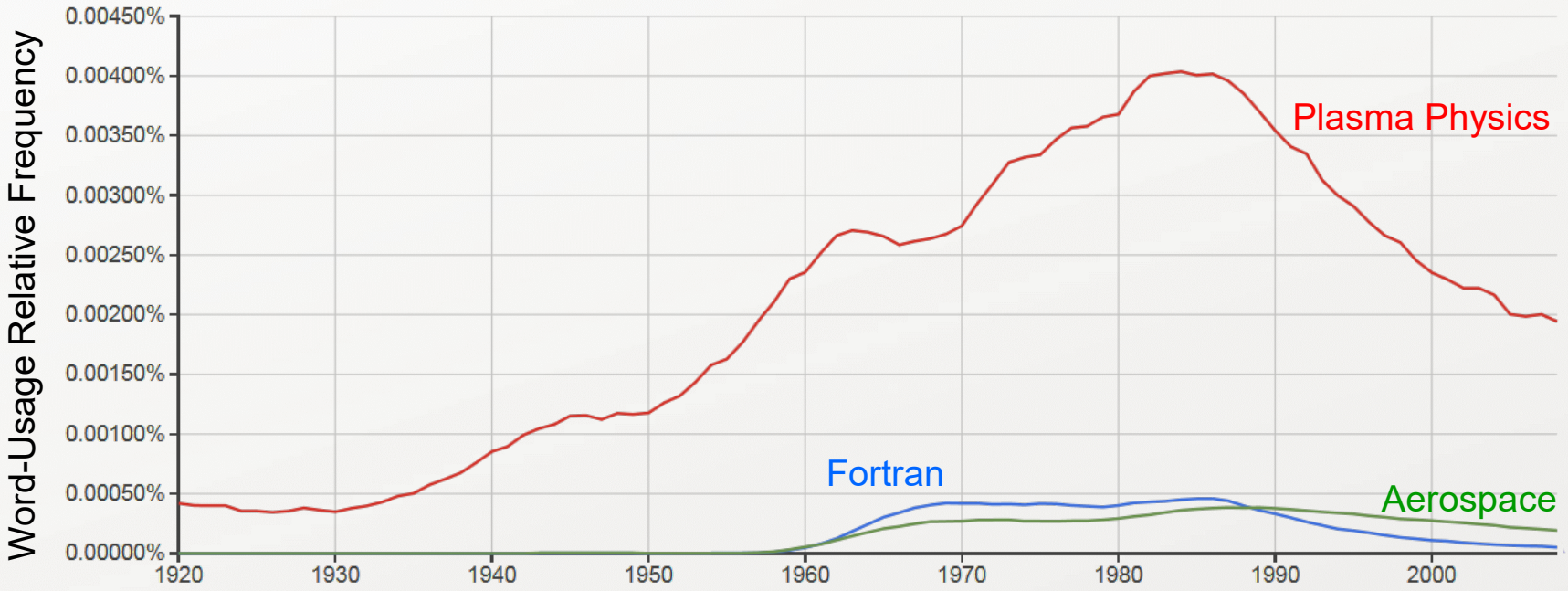
- **30.7101 Data Analytics, General.**

A program that prepares individuals to **apply data science** to generate insights from data and identify and predict trends. Includes **instruction in computer databases**, **computer programming**, inference, machine learning, optimization, probability and stochastic models, **statistics**, strategy, uncertainty quantification, and visual analytics.

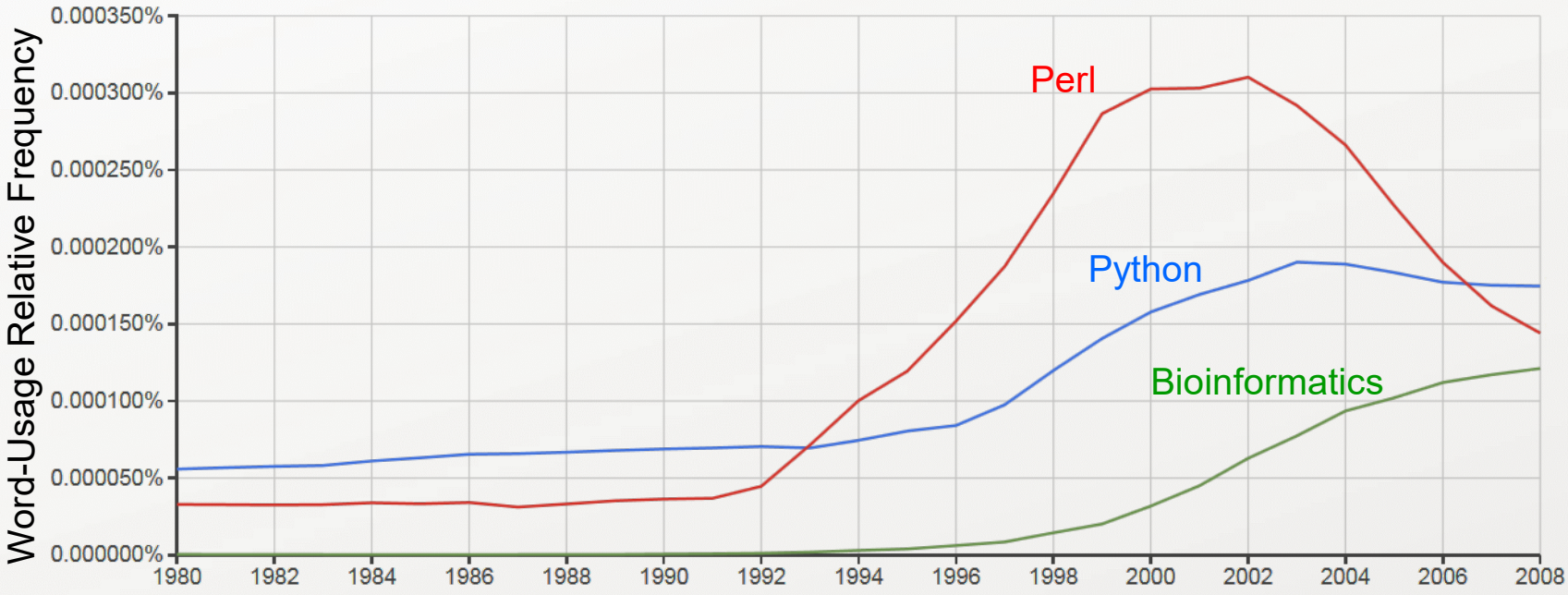
- **30.7099 Data Science, Other.**

Any instructional program in data science not listed above.

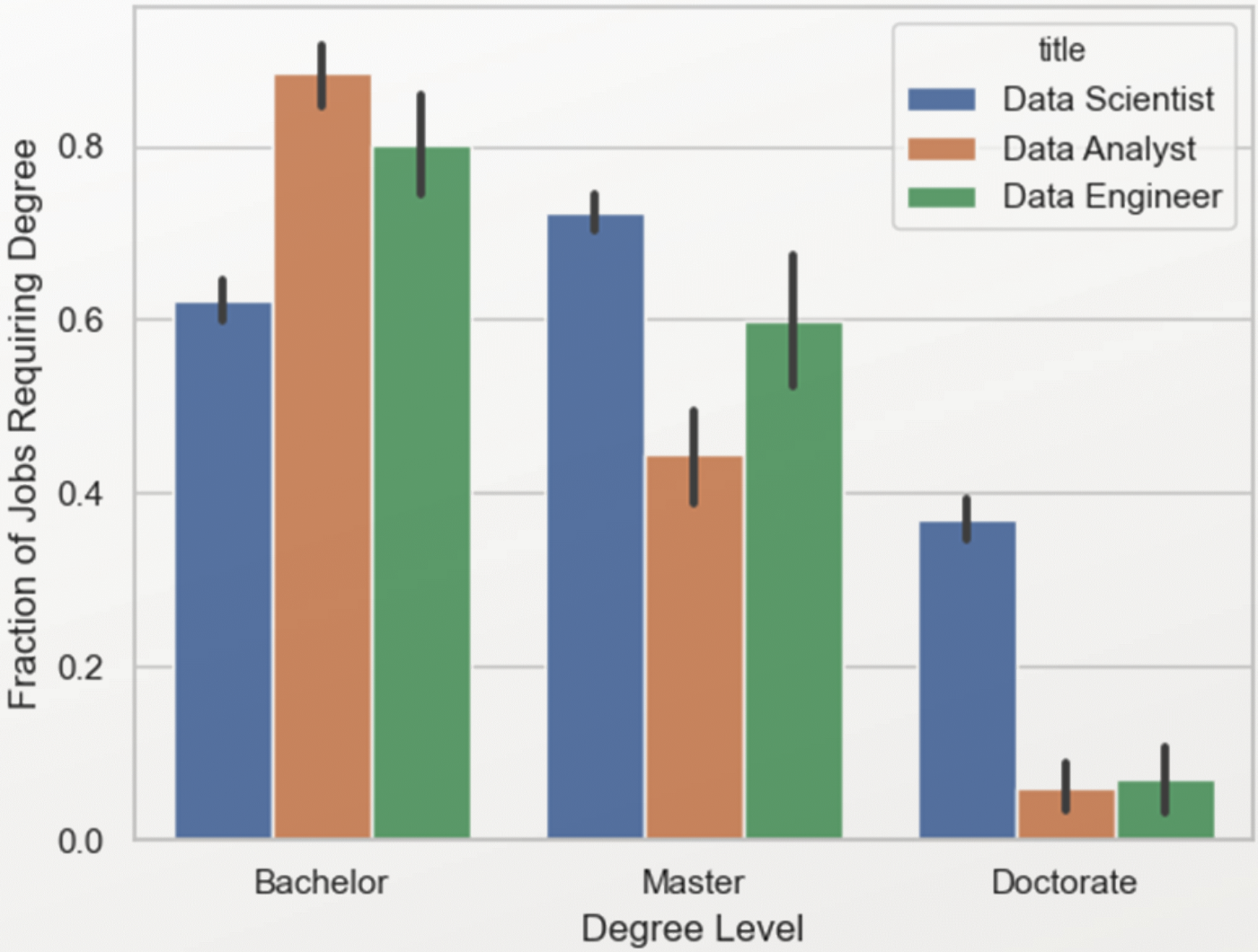
The landscape of scientific tools is highly dynamic and domain-specific



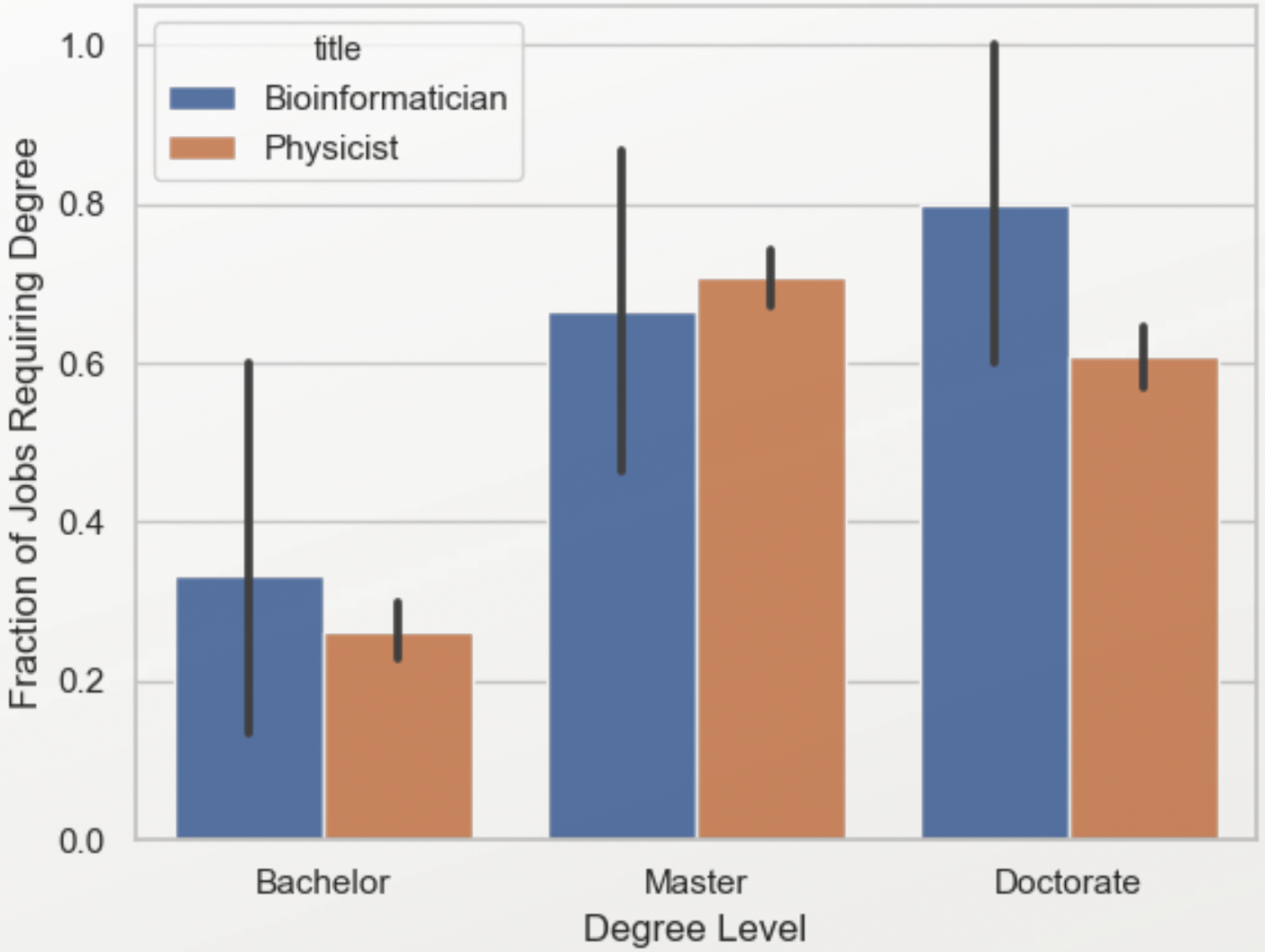
The landscape of scientific tools is highly dynamic and domain-specific



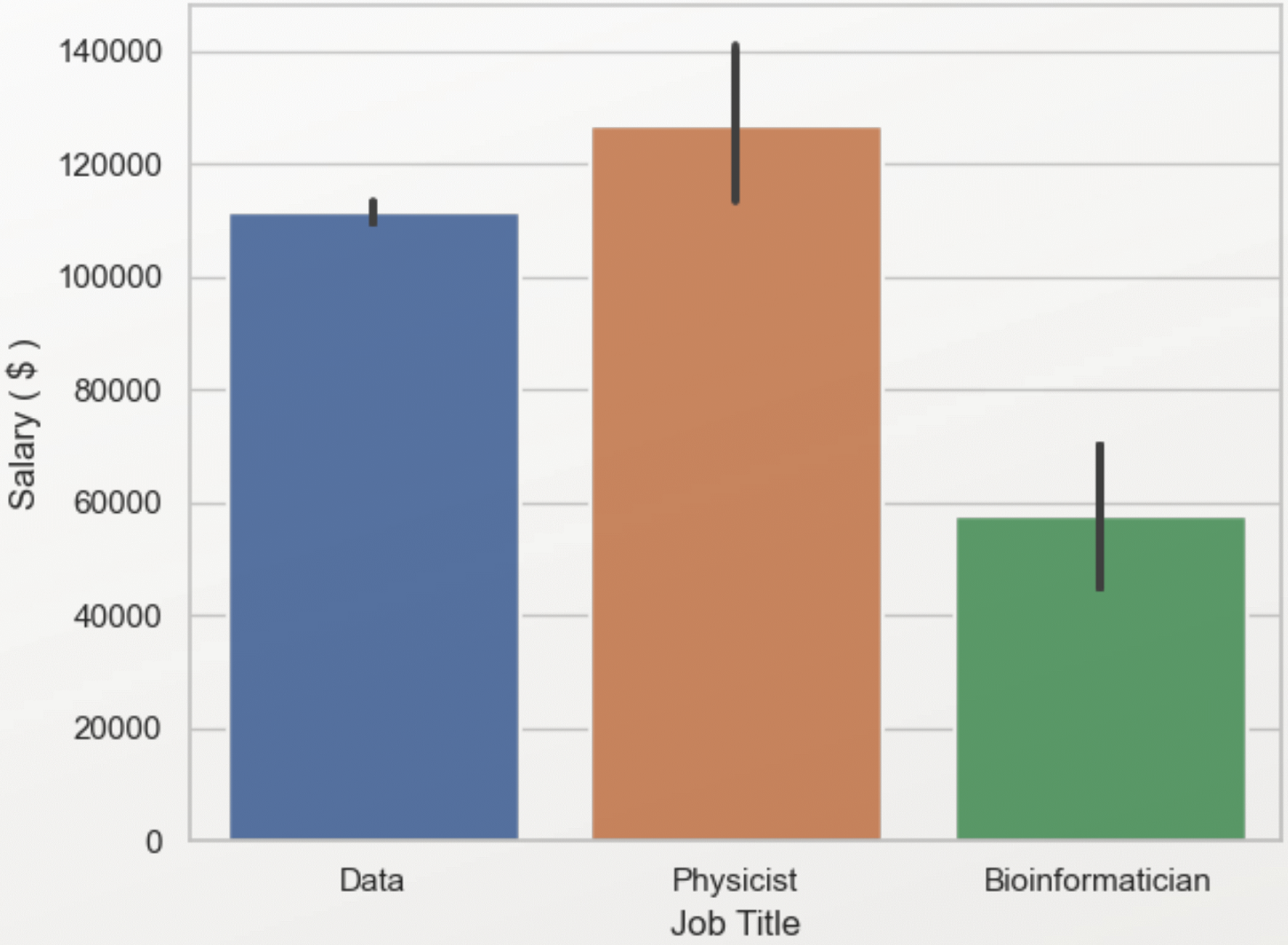
Data-related jobs generally require less advanced degrees than STEM



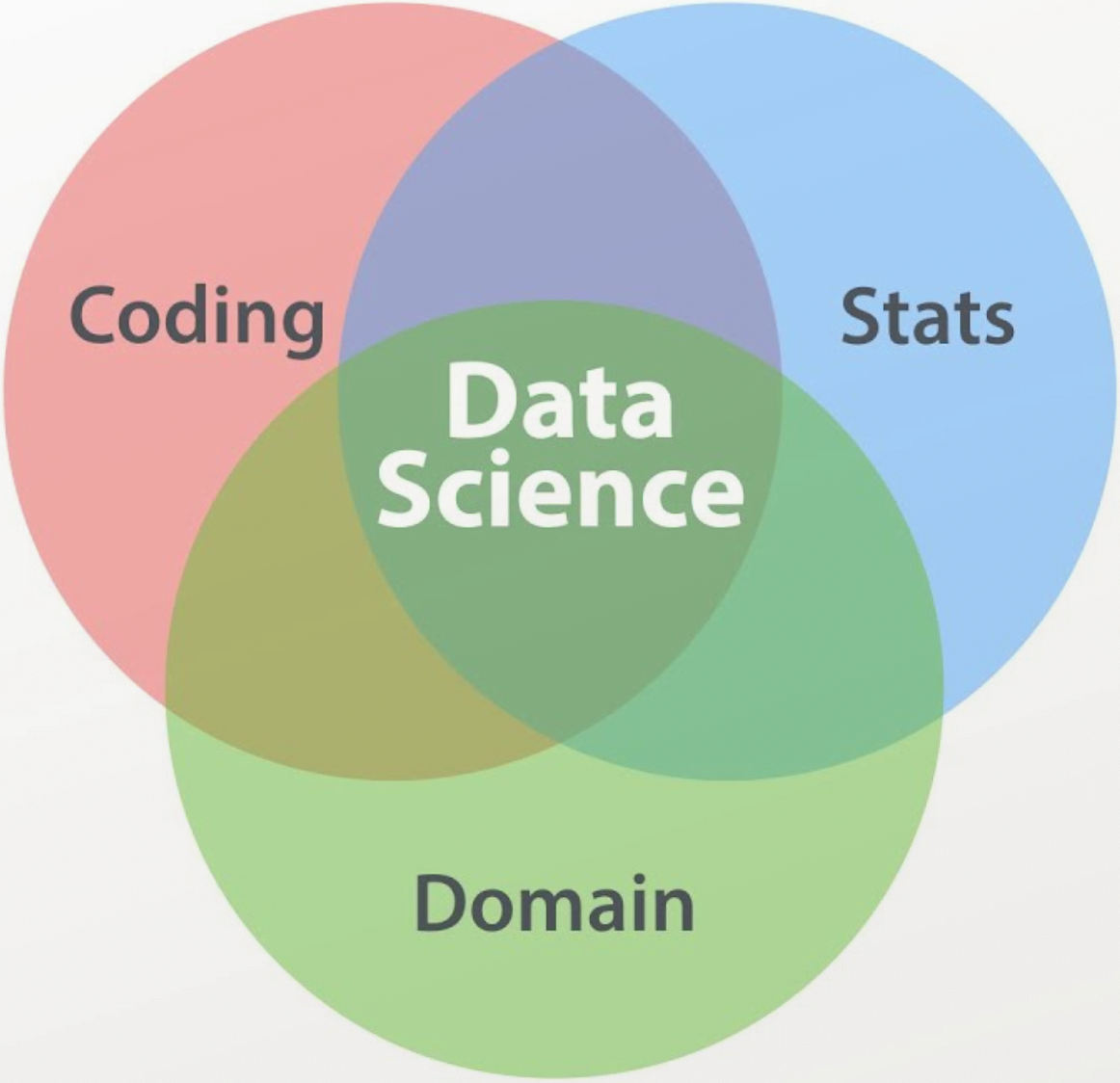
Data-related jobs generally require less advanced degrees than STEM



Data-related jobs generally pay much higher salaries than other STEM degrees

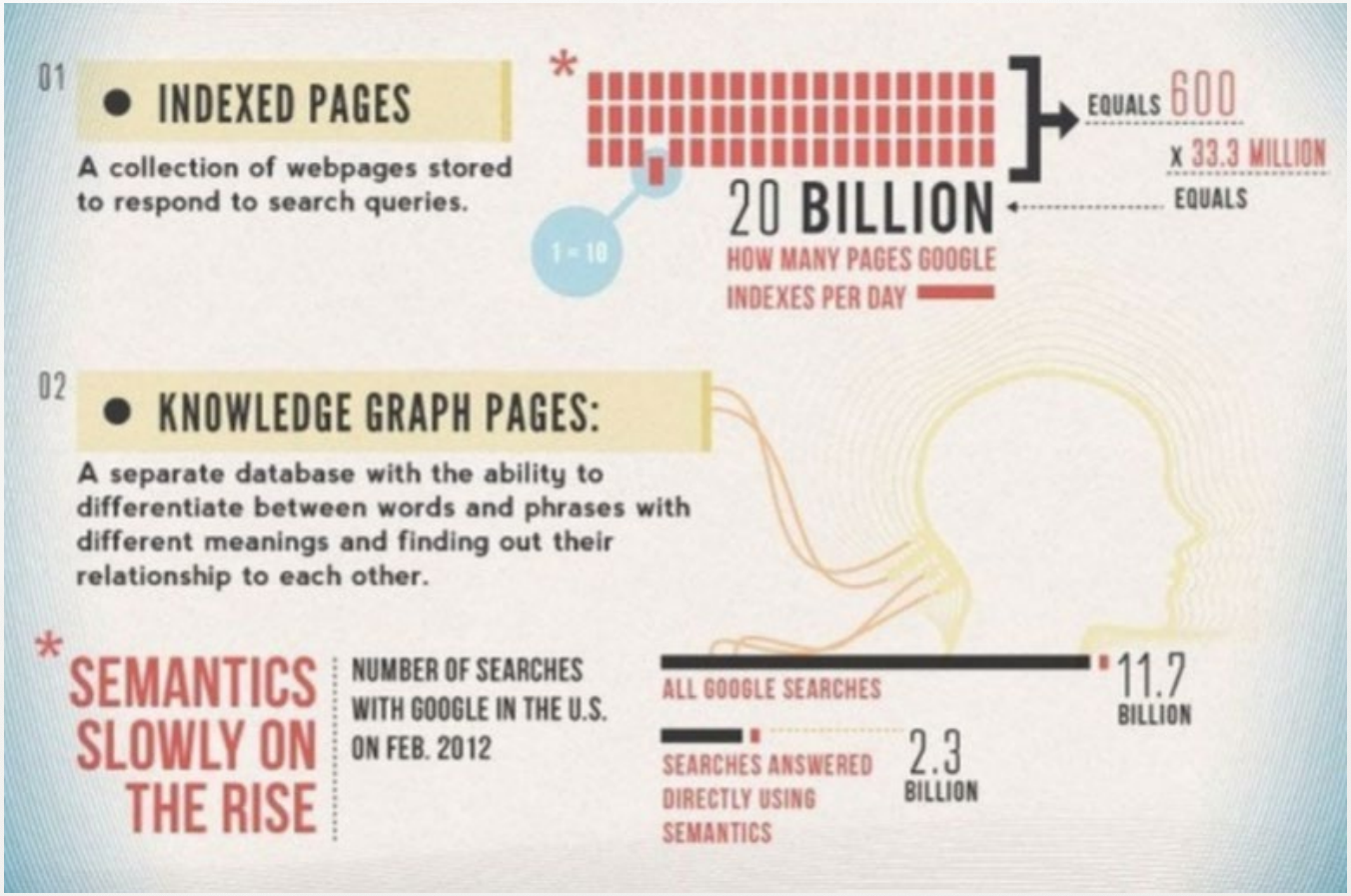


Learning Data Science competencies via Word Embeddings



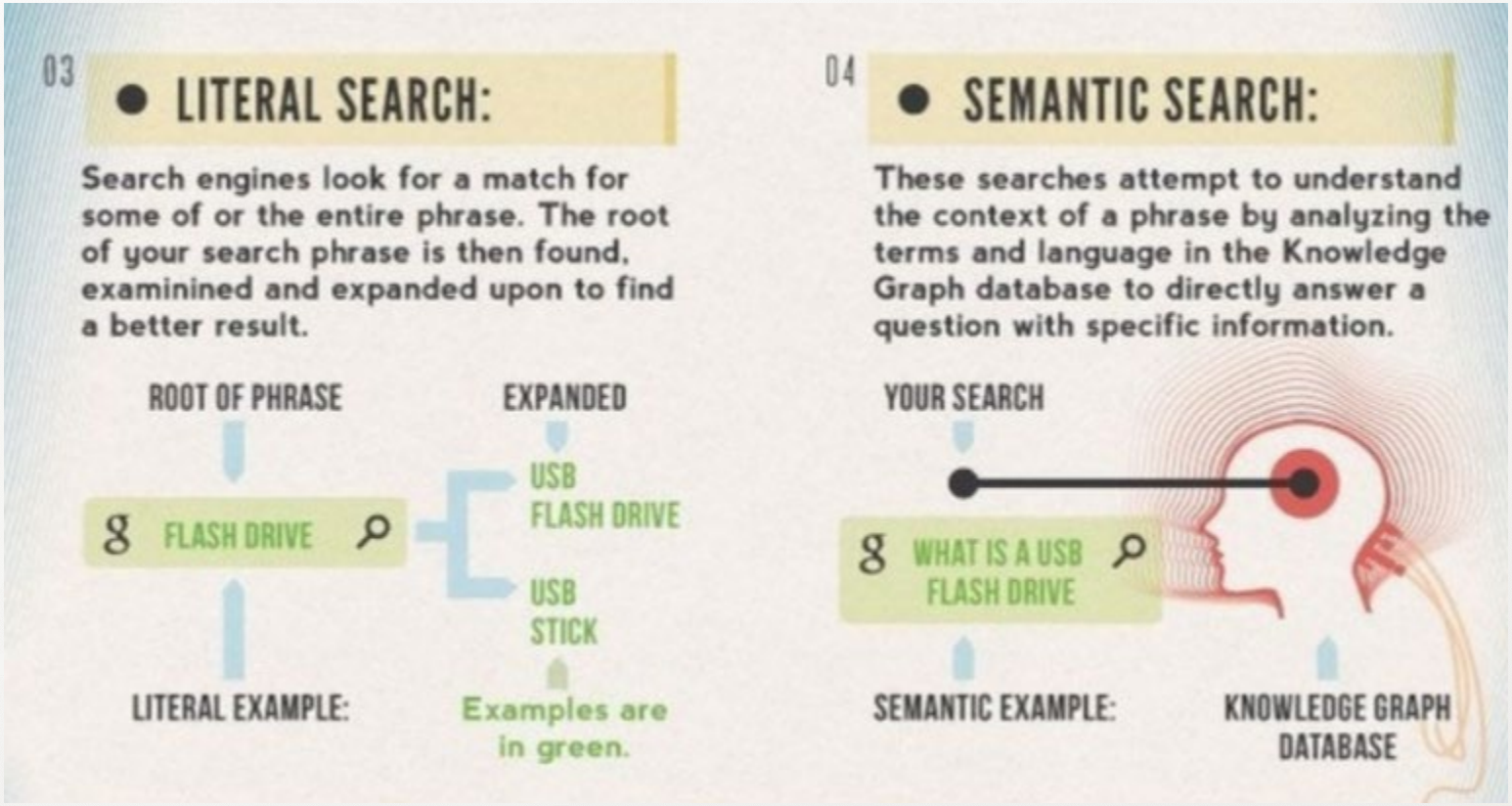
BIG DATA: How does a search engine work

The search results come from indexed pages and knowledge graphs.



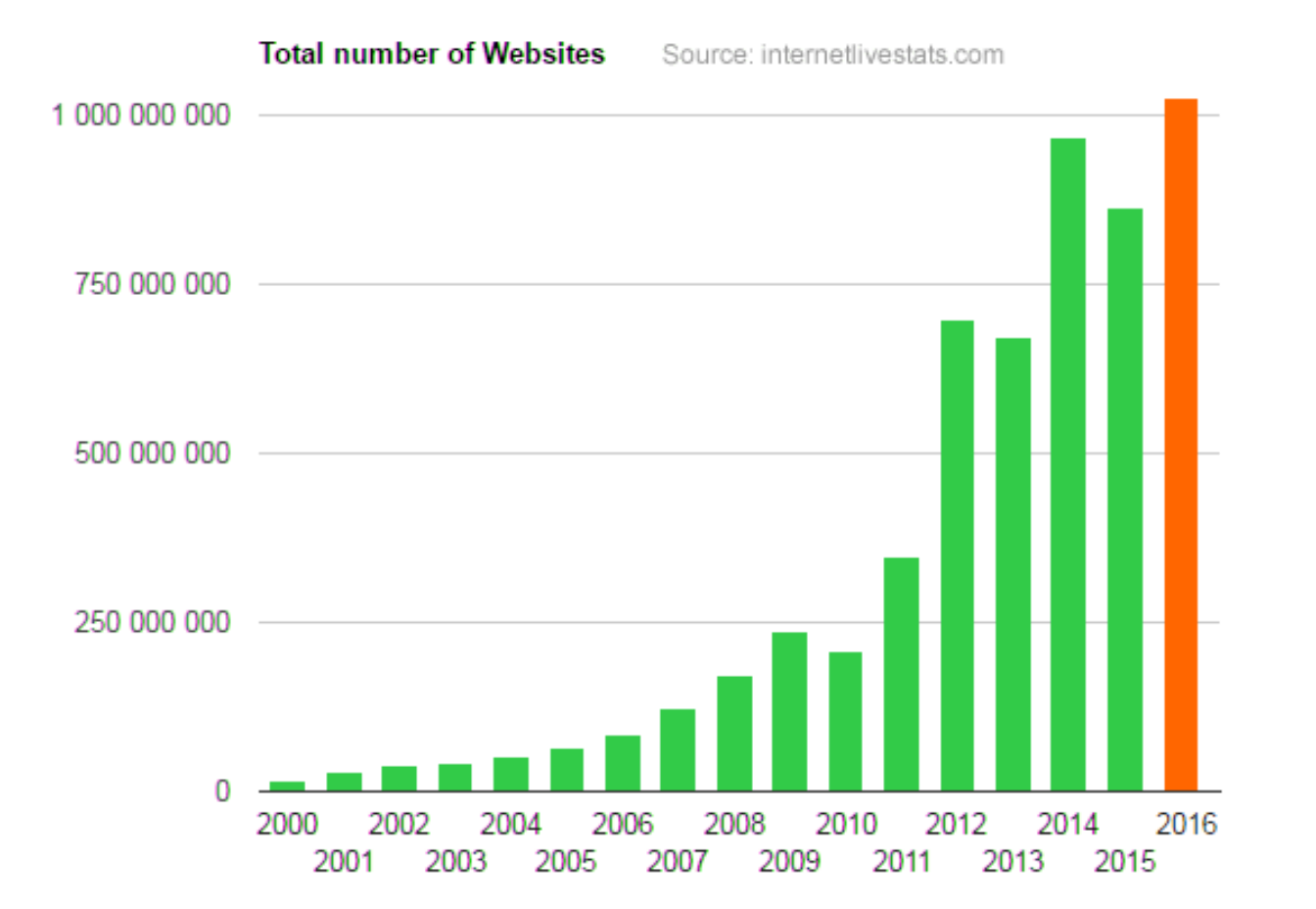
How does a search engine work

When you type a phase into the search bar, Google analyzes the literal and semantics of the query and searches for both



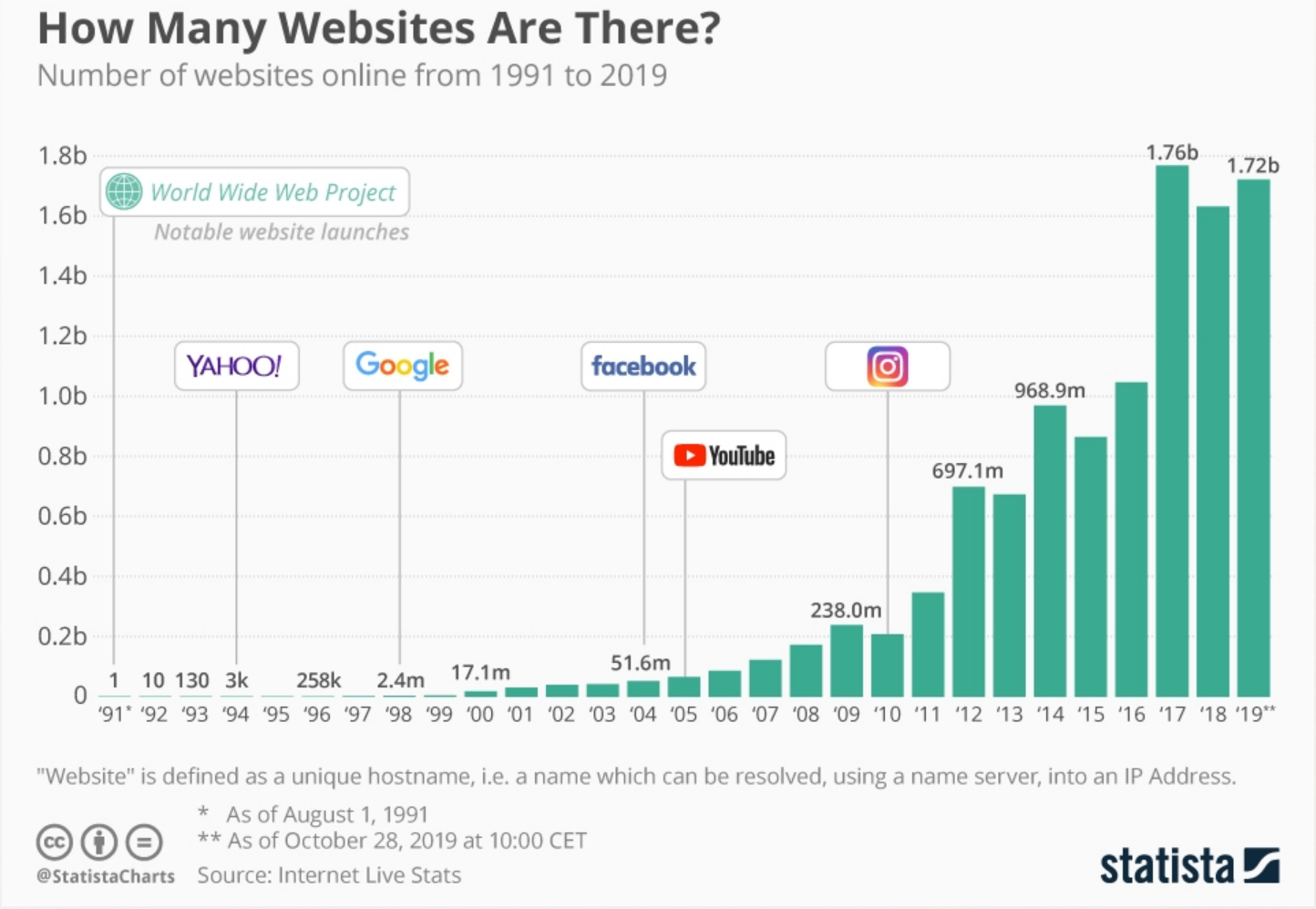
How does a search engine work

When you type a phase into the search bar, Google analyzes the literal and semantics of the query and searches for both



How does a search engine work

When you type a phase into the search bar, Google analyzes the literal and semantics of the query and searches for both



How does a search engine work

When you type a phase into the search bar, Google analyzes the literal and semantics of the query and searches for both

accept your query

parse your query

figure out the word order

look up the information in its database

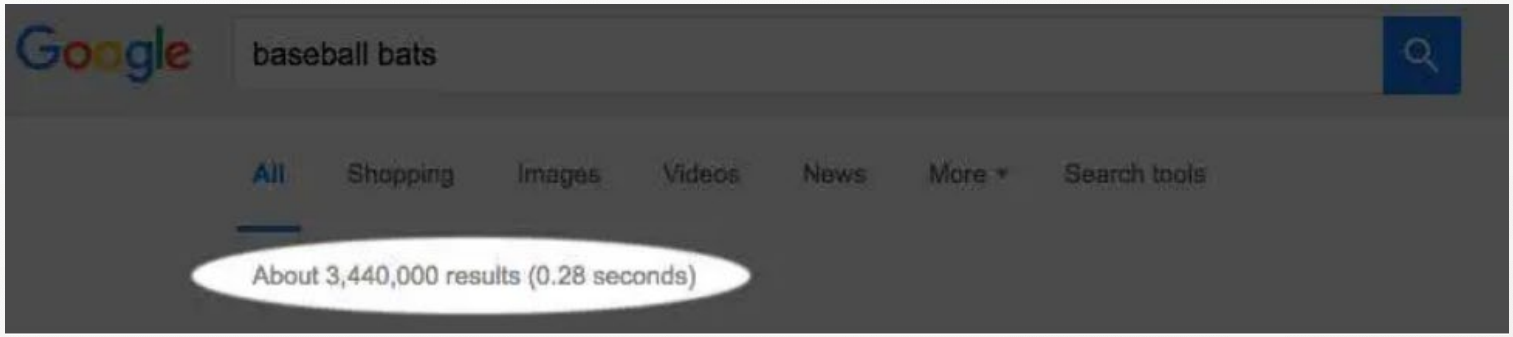
personalize your results by taking into account everything it knows about you

rank the results

send the results to your browser

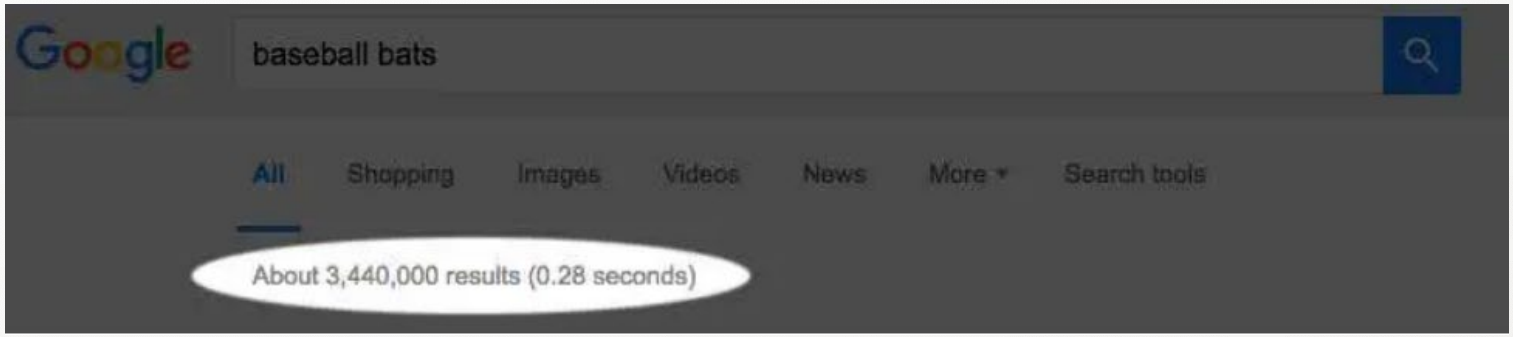
How does a search engine work

When you type a phase into the search bar, Google analyzes the literal and semantics of the query and searches for both



How does a search engine work

When you type a phase into the search bar, Google analyzes the literal and semantics of the query and searches for both



Multiple Datacenters with a Worldwide Load Balancing Network

Hundreds of Computers in Each Datacenter Using Distributed Lookups

Custom File System and Custom Software

Caching, Index Stored in RAM, Pre-fetching Results