

Data analytics for psychology and business

Dirk Wulff, Rui Mata

February 21, 2020

Take 5 minutes to fill out this survey

bit.ly/cds_dataanalytics

**The
Economist**

MAY 6TH-12TH 2017

Crunch time in France

Ten years on: banking after the crisis

South Korea's unfinished revolution

Biology, but without the cells

The world's most valuable resource

google

amazon

Microsoft

Google

USER

f

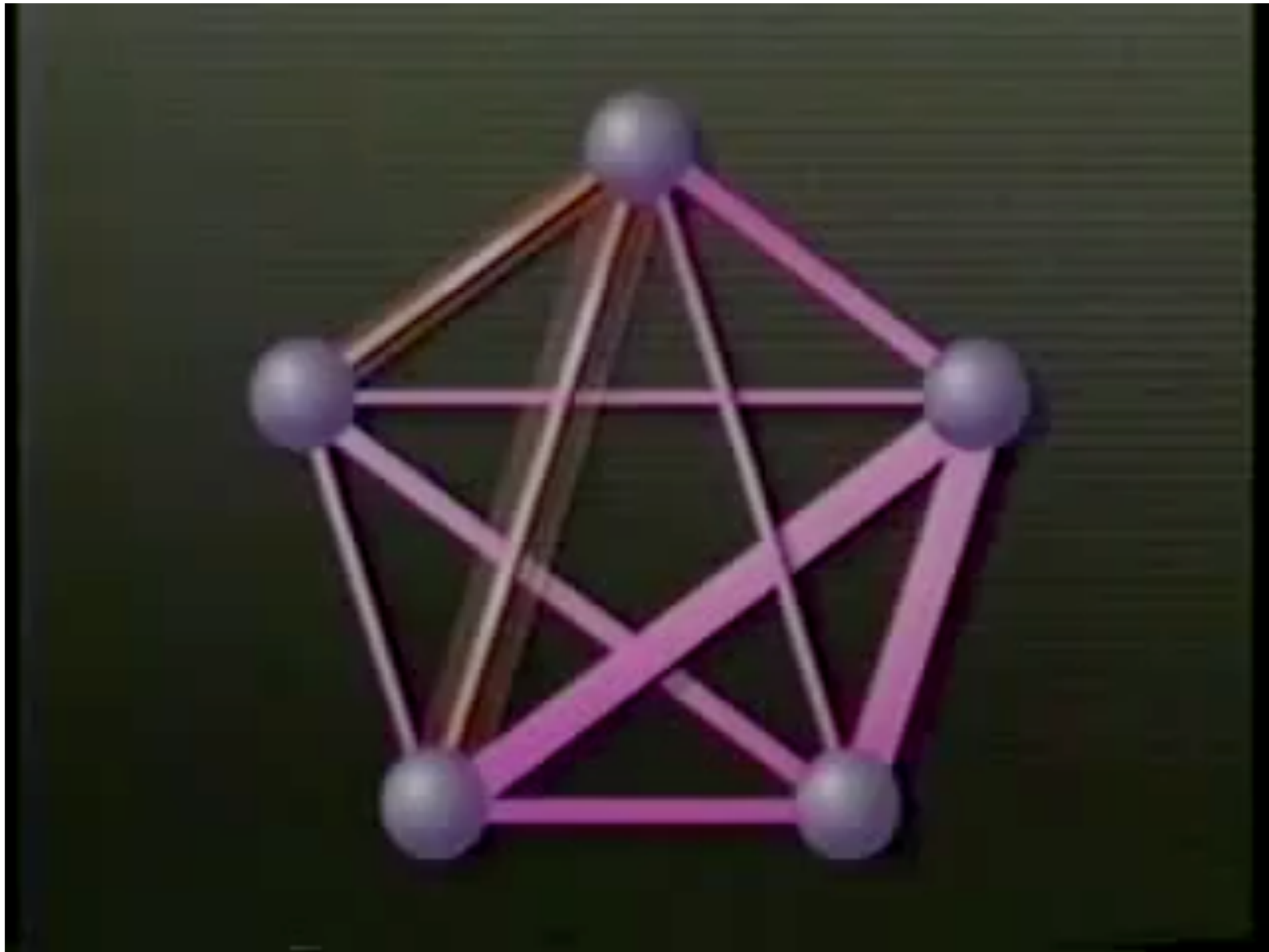
TESLA

f

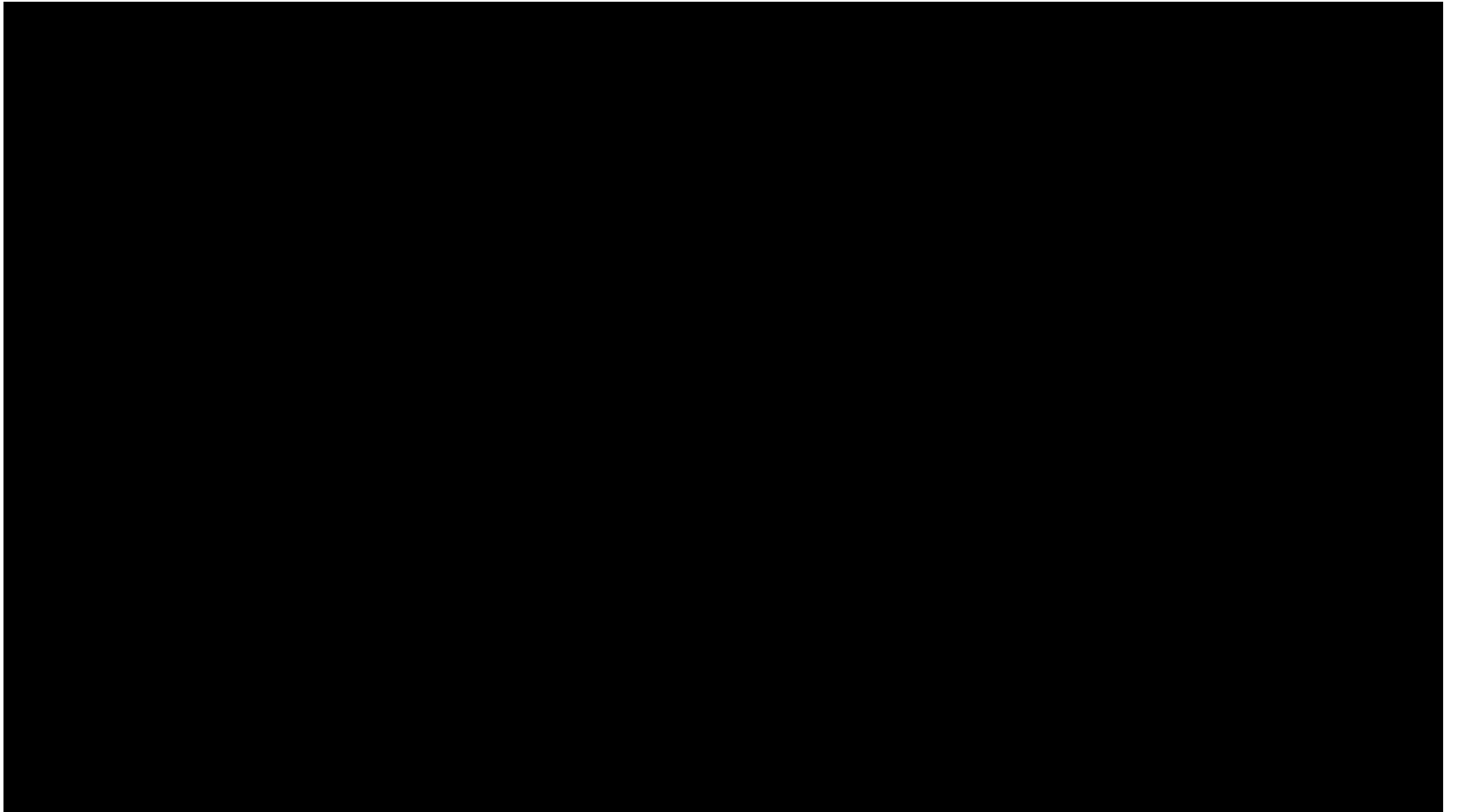
**Data and the new rules
of competition**



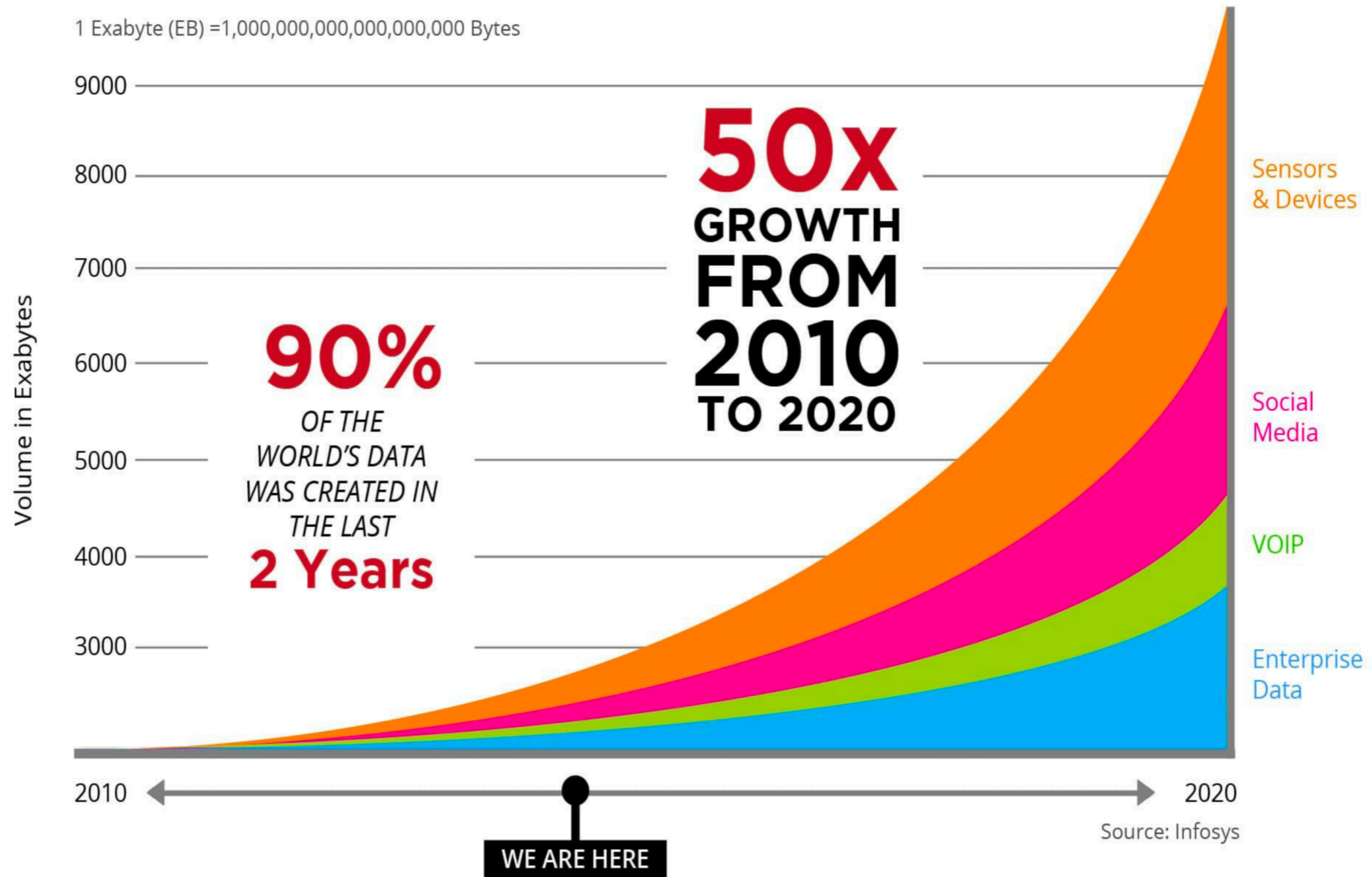
AI in the 50s & 60s



AI today



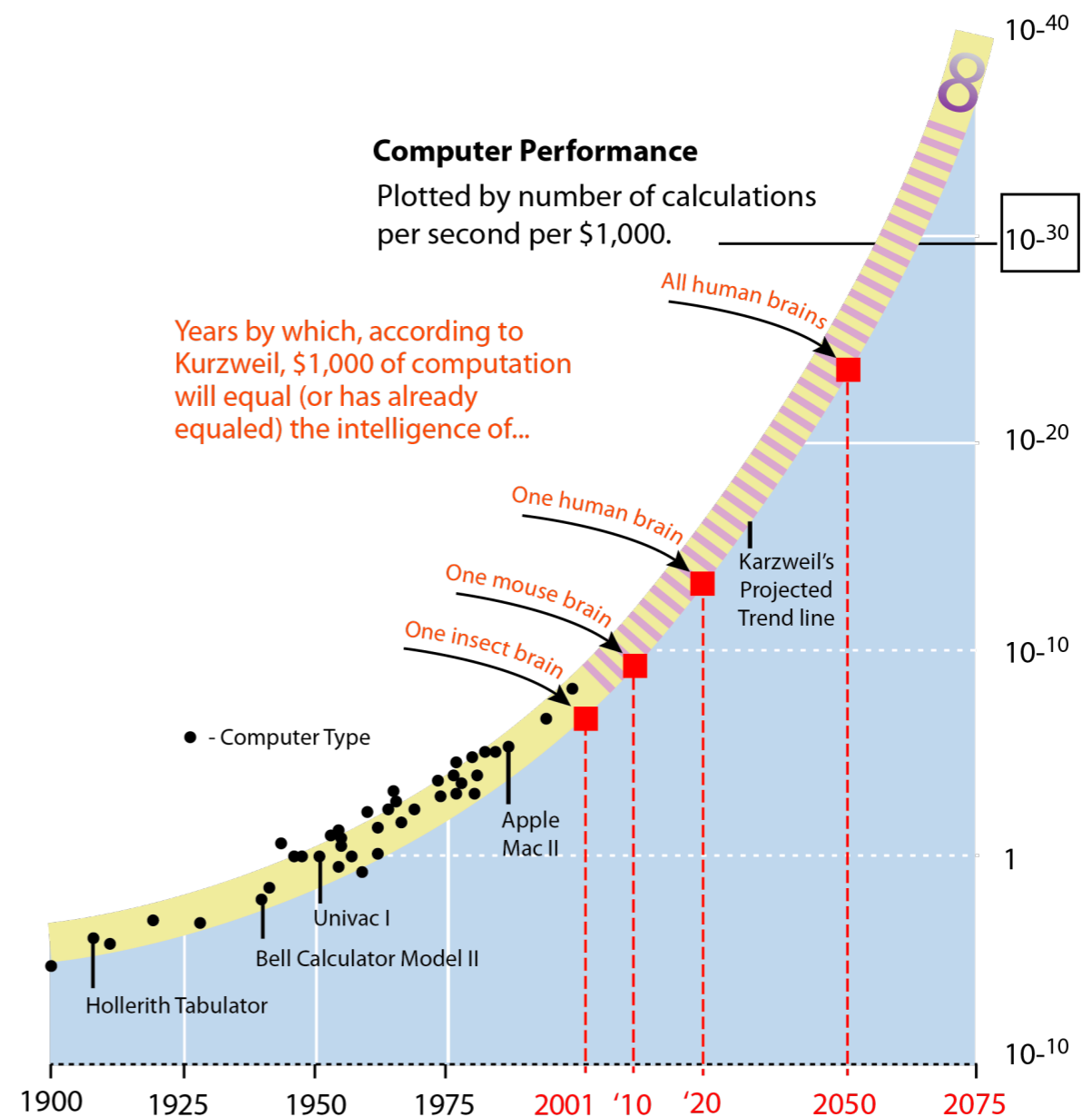
Exponential growth of digital data



Exponential growth of computing efficiency



Ray Kurzweil
Director of Engineering @ Google



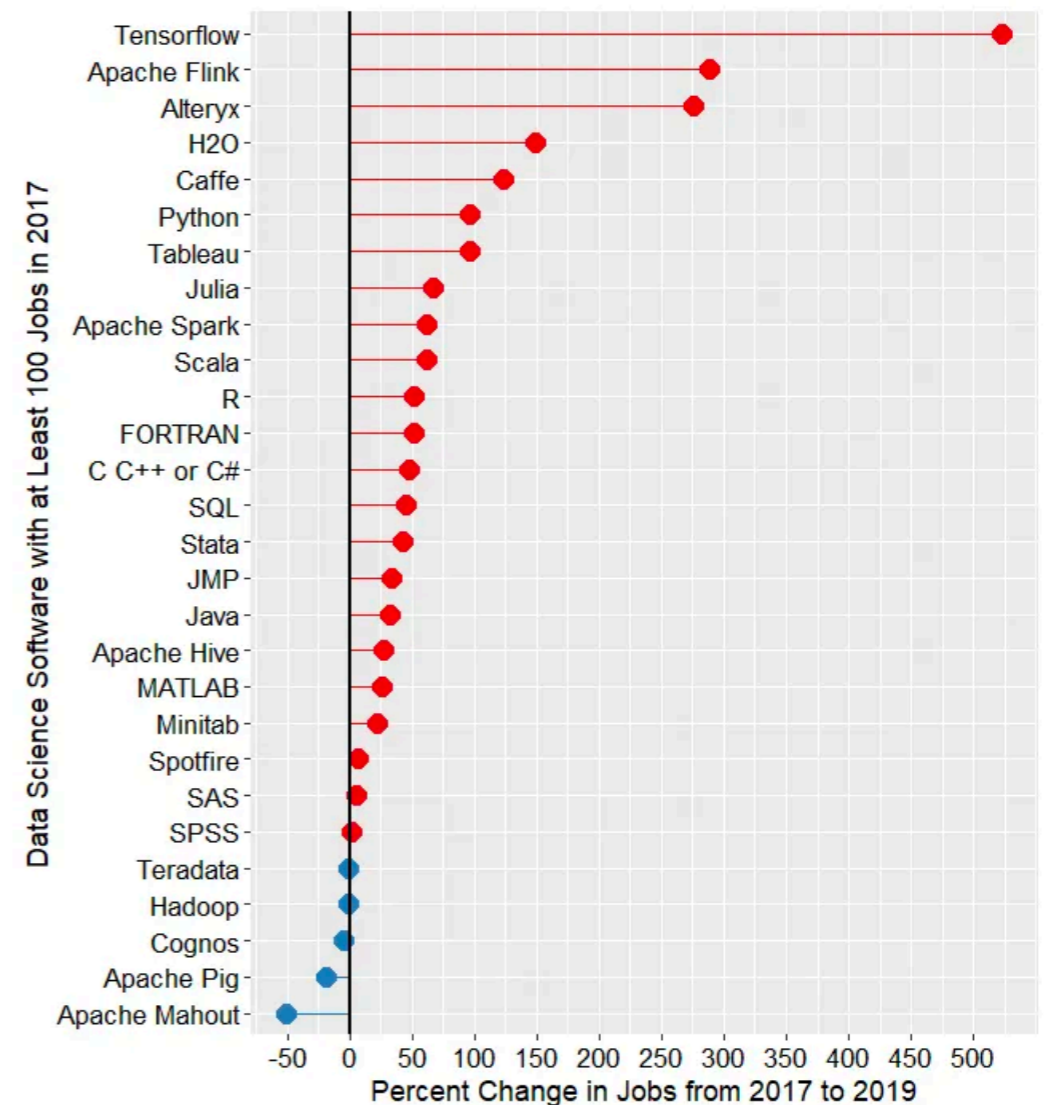
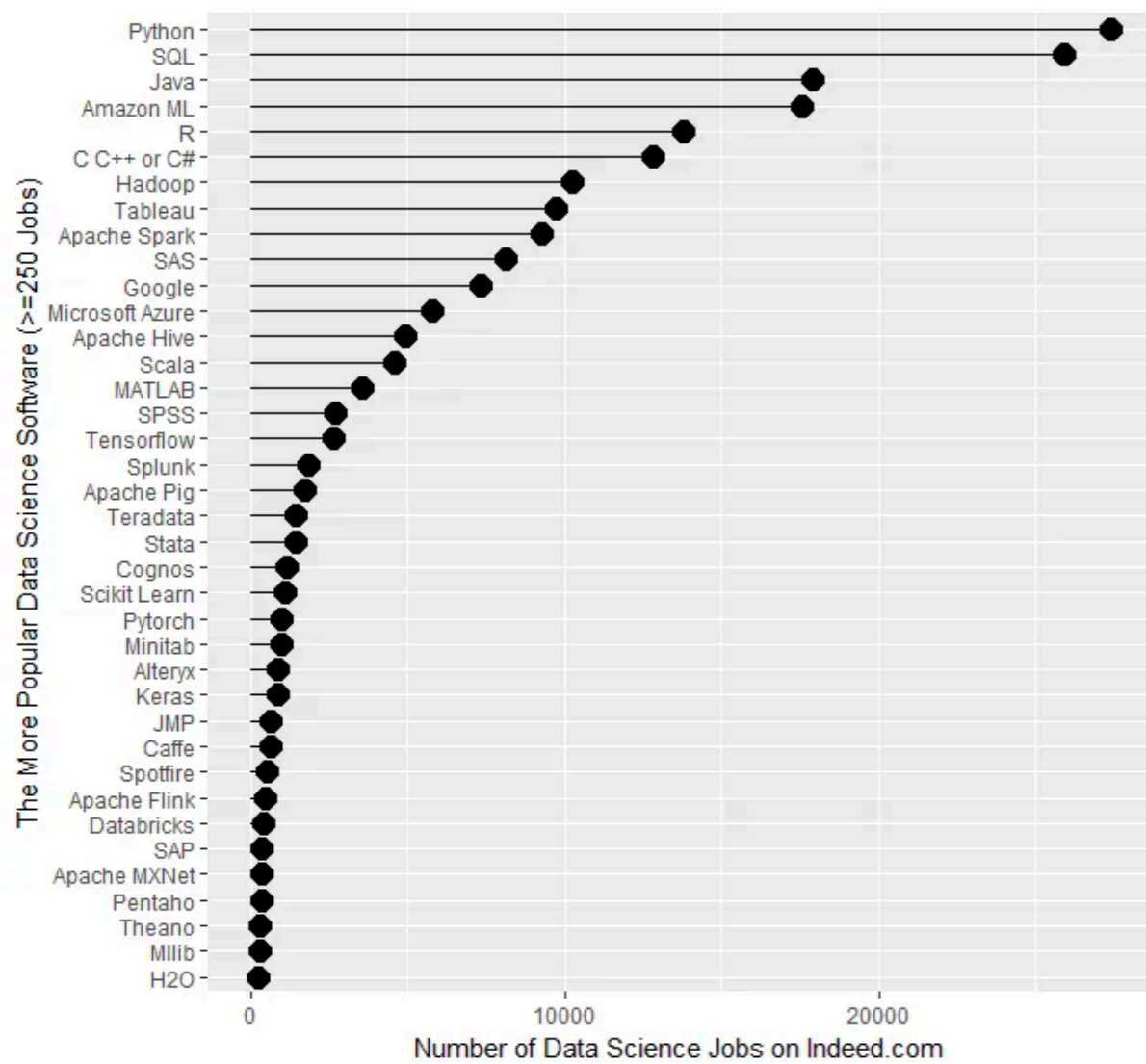
(Exponential) growth of data science tools

The image displays a vast collection of logos for data science tools, organized into several main categories:

- INFRASTRUCTURE:** Includes HADOOP ON-PREMISE (Cloudera, Hortonworks, MPR, Pivotal, IBM InfoSphere, jethro), HADOOP IN THE CLOUD (AWS, Google Cloud, SAP Cloud Platform, IBM InfoSphere BigInsights, ARM, Double, CAZENA), and STREAMING / IN-MEMORY (Amazon Kinesis, Databricks, SAP Cloud Platform, Oracle, Confluent, Stream, Hazelcast, GridGain, GIGASPACE, Wallaroo, PASTDATA, Ix).
- ANALYTICS & MACHINE INTELLIGENCE:**
 - DATA ANALYST PLATFORMS:** Microsoft, Pentaho, Alteryx, Digital Reasoning, GUAVUS, AYASDI, ATTIVO, Datameer, incorta, interana, MODE, ENDOR, sisu, switchboard, Starburst.
 - DATA SCIENCE PLATFORMS:** IBM, Databricks, data iku, DOMINO, rapidminer, TIBCO, ANACONDA, SAS, MathWorks, KNIME.
 - BI PLATFORMS:** Looker, Amazon Analytics, AWS, Domo, Arcadia Data, ThoughtSpot, SAP, Qlik, AtScale, Information Builders, Birst, MicroStrategy, Keen IO.
 - VISUALIZATION:** Tableau, Power BI, SAP, Google Cloud, Celonis, Zepi, Chartist, PowerView.
 - MACHINE LEARNING:** Azure ML, Google Cloud AutoML, H2O, DataRobot, Gamalon, Visonze, Element AI, DeepSense AI.
- APPLICATIONS - ENTERPRISE:**
 - SALES:** CHORUS, INSIDESALES.COM, peopleAI, conversica, clari, avso, tact.ai, fuses/machines, clearbit.
 - MARKETING - B2B:** RADIUS, App Annie, EVERSTRING, Lattice, MINTIGO, bsense, tubular, ENGAGIO, KNOTCH, mrp.
 - MARKETING - B2C:** Zeta, Bloomreach, SendGrid, braze, ACTIONIQ, BLUECORE, CONTENT SQUARE, TEALIUM, Amperio, amperity, QUANTIFIND, Simon, Lifica, PERSADO, refresh.
 - CUSTOMER EXPERIENCE / SERVICE:** qualtrics, MEDALLIA, SurveyMonkey, UserTesting, CLARABRIDGE, zendesk, Customer, freshdesk, INTERCON, Drift, LIVEPERSON, Gainsight, pendo, HEAP, Amplitude, Watson Assistant, Dialogflow, DigitalGenius, A5APP, ada, AUTOMAT, ahniti, CallDesk, netomi, frame.ai.
 - ENTERPRISE PRODUCTIVITY:** slack, ORACLE, GURU, Lumina, DIFFBOT, clara, talla, Kasisto.
- APPLICATIONS - INDUSTRY:**
 - ADVERTISING:** AppNexus, Oracle, MOAT, theTradeDesk, TAPAD, dataxu, gumgum, Appier, Criteo, xAd, OpenX, distillery, Liventel, Tapad, dataxu, gumgum, Appier.
 - EDUCATION:** Knewton, Clever, kidaptive, PANORAMA, Knowre, gradscope.
 - REAL ESTATE:** Redfin, Opencor, VTS, CREDIFI, GEOPHY, COMPSTAK, STREETSHIDATA, Opencor.
 - GOVT:** OPENGOV, mark43, LiveStories, Passport, SmartProcure, STREETSHIDATA, Opencor.
 - INTELLIGENCE:** Palantir, Dataminr, Quid, PRIMER, FORGE.
 - FINANCE - INVESTING:** KENSHC, Quantopian, ADDEPAR, NUMERA, ISENTIUM, ALGORZ, Fluoropack, PAGAYA.
 - FINANCE - LENDING:** ondeck, Affirm, JIANPUAI, Kreditech, AVANT, TALA, Upstart, Curo, LEARNBANC, upgrade, 100Credit, WeLab, Weosha, TrueAccord, MoneyLion, ActiveAI, aire, cignifi.
 - INSURANCE:** Metromile, Lemonade, Hippo, Shift Technology, ROOT, zesty.ai, TRACTABLE, CAPE.
- HEALTHCARE:** Flatiron, Clover, AYRUS, HealthTap, METABIOTA, Gingerio, Glow, babyon, 3DMed, zebra, PARTAI, ovia, TEMPUS, patientslikeme, AICure, insitro, notable, citizen, prognos, enitic, imago, Blackthorn, BAYLABS, Qventus, ARTERYS, IMAGEN, innovaccor, PAIGE, DATAVANT, DWKIN.
- LIFE SCIENCES:** color, BenevolentAI, verily, Clear Labs, freemote, PHOSPHORUS, DNAnexus, CITRINE, twoAR, Atomize, GENOMICS, SODIUS, DWKIN.
- TRANSPORTATION:** UBER, TESLA, CLEARPATH, CRUISE, NURO, ARGO, drive.ai, SAMBRIDGE, AURORA, nauto, AUTOMOTIVE, PILOTAI, NIO, OPTIMUS, moovit, Ibe, nexar, Kodiak, comma.ai, netradyne, German Autotech, Civil Maps, cognata, chind, INRIX.
- AGRICULTURE:** FARMERS, Granular, JOHN DEERE, BLUE RIVER, FarmersEdge, AgriStar, FarmLogs, TARANIS, GAMAYA, Terraviva, prospera.
- COMMERCE:** instacart, FAIRE, STITCH FIX, Du & Co, RetailNext, HowGood, heuritech, eharmony, stem, Amper, ByteDance, hoppin, celest, SOJERN, B@XEVER, VERIDICIS, duetto, Jukedeck, Second Spectrum, Electric, ZINIER, Spoke, Moveworks.
- INDUSTRIAL:** AVEVA, SIEMENS, PREDIX, UPTAKE, SCORTEX, KODIAX, TACHYUS.

- CROSS-INFRASTRUCTURE/ANALYTICS:** AWS, Google Cloud, Microsoft, IBM, SAP, Hewlett Packard Enterprise, SAS, IOIODATA, VMware, TIBCO, TERADATA, ORACLE, NetApp, syncsort, MPR, cloudera.
- OPEN SOURCE:**
- FRAMEWORKS:** Spark, Flink, YARN, TEZ, MESOS, doctar, CDAP, HELIX.
- QUERY / DATA FLOW:** Spark, presto, SLAMDATA, GraphQL, Flink.
- DATA ACCESS & DATABASES:** cassandra, mongoDB, redis, Cockroach LAAS, druid, COUCHDB, CLOUDSPANNER, SCIDB, triak, HBASE, OCCUTULO.
- ORCHESTRATION & MGMT:** talend, Apache Ambari, etcd, Kang.
- STREAMING & MESSAGING:** Spark, nifi, Flink, beam, kafka, STORM, Apache RocketMQ.
- STAT TOOLS & LANGUAGES:** python, Scala, Studio, SciPy, julia, SELDON, Polysol.
- AI OPS & INFRA:** miflow, KubeFlow, DVC, Seldon.
- AI / MACHINE LEARNING / DEEP LEARNING:** TensorFlow, Keras, PyTorch, OpenAI, DM, theano, Apache SINGA, DIMSUM, FeatureFu, mxnet, VELES, ONNX, PyTorch, neon, DSSSTNE, mlflow, MAHOUT, Aerosolve, fast.ai, mlr, openml.
- SEARCH:** elasticsearch, Solr.
- LOGGING & MONITORING:** elasticsearch, kibana, SENTRY, logstash, Prometheus, fluentbit, fluentd, Grafana, Vector.
- VISUALIZATION:** matplotlib, TensorBoard, seaborn, Bokeh.
- COLLABORATION:** BeakerX, jupyter, ZenML, ANACONDA.
- SECURITY:** Apache Ranger, KNOX, SENTRY, ACCURITO.
- DATA SOURCES & APIs:**
- HEALTH:** Apple, VALIDIC, practice fusion, fitbit, GARMIN, HUMAN AP, kinsa, MINIC.
- IOT:** GE Digital, UPTAKE, thingworx, helium, samsara, estimate.
- FINANCIAL & ECONOMIC DATA:** Bloomberg, THOMSON REUTERS, DOW JONES, S&P CAPITAL IQ, CB INGHITS, PLAID, SECOND MEASURE, ENVESTNET YODLEE, THE WORLD BANK, @estimate, PREMISE, Quandl, Eagle Alpha, Stocktwits, xignite, Thinknum, earnest, predata.
- AIR / SPACE / SEA:** Orbital Insight, planet, SKYCAT, AIRBOTICS, spire, kespri, FREEDOMOV, DRONESTORY, telluslabs, WINDWARD, DroneDeploy, MarineTraffic, LOG ENVI, REACT.
- PEOPLE / ENTITIES:** acxiom, experian, EPSILON, InsideView, Crimson Hexagon, BASIS, Quantcast, SAFEGRAPH.
- LOCATION INTELLIGENCE:** FOURSQUARE, MapAnything, mapbox, sense360, plney, hexagon, PlaceIQ, esri, factual, CARTA, Mapillary, Streetline, cuebq, Radar, OpenStreetMap, Lgraffiti.io.
- OTHER:** DATA.GOV, IMAGENET, wki, Wikidata, ApolloScale, ApolloScale, CRUX, Lgraffiti.io.
- DATA RESOURCES:**
- DATA SERVICES:** OPERA, LIQ, fractalx, kaggle, DataKind, INNOPLXUS.
- INCUBATORS & SCHOOLS:** PLURALSIGHT, GA, galvanize, DataCamp, DataElite, INSIGHT, The Data Incubator, METIS.
- RESEARCH:** facebook research, OpenAI, MIRI, VECTOR INSTITUTE, ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE, AIZ.

R (and Python) dominate the data science world



Data scientists missing

emerging
roles,
global
change
by 2022

133
Million

Top 10 Emerging

1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. General and Operations Managers
4. Software and Applications Developers and Analysts
5. Sales and Marketing Professionals
6. Big Data Specialists
7. Digital Transformation Specialists
8. New Technology Specialists
9. Organisational Development Specialists
10. Information Technology Services

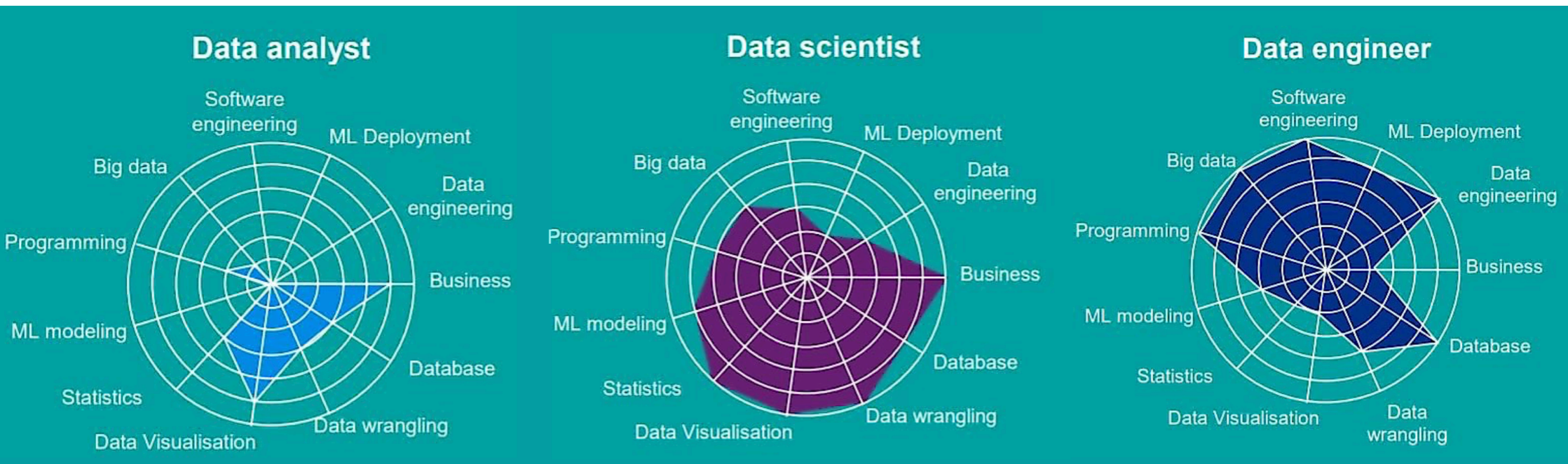
declining
roles,
global
change
by 2022

75
Million

Top 10 Declining

1. Data Entry Clerks
2. Accounting, Bookkeeping and Payroll Clerks
3. Administrative and Executive Secretaries
4. Assembly and Factory Workers
5. Client Information and Customer Service Workers
6. Business Services and Administration Managers
7. Accountants and Auditors
8. Material-Recording and Stock-Keeping Clerks
9. General and Operations Managers
10. Postal Service Clerks

Data science is not (only) machine learning and AI



Data science's misunderstood hero

"Each of the three data science disciplines has its own excellence. Statisticians bring rigor, ML engineers bring performance, and analysts bring speed."

"Your analyst is the sprinter; their ability to quickly help you see and summarize what-is-here is a superpower for your process."

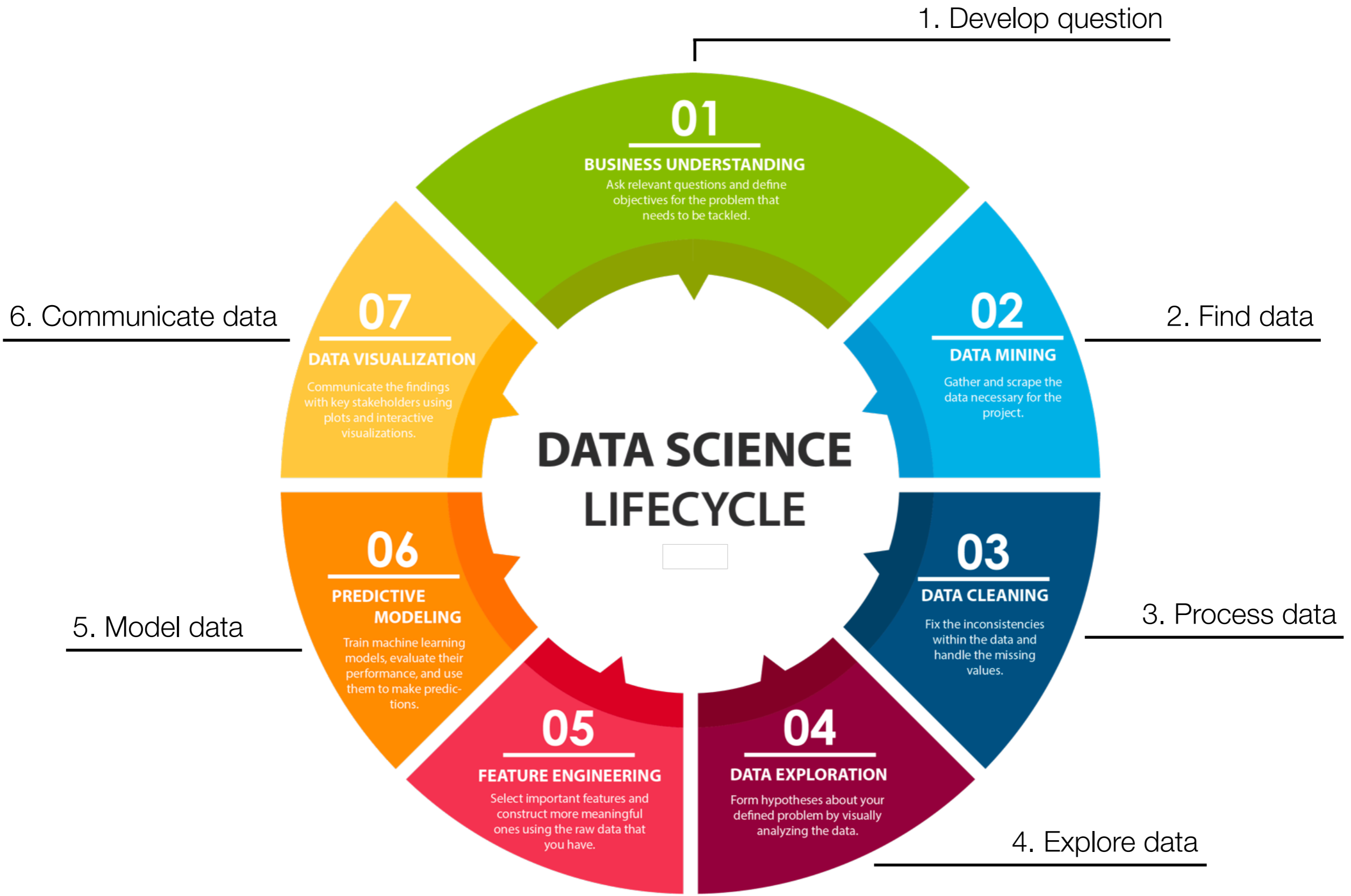
"The only roles every business needs are decision-makers and analysts. If you lose your analysts, who will help you figure out which problems are worth solving?"



Cassie Kozyrkov

Chief Decision Scientist @ Google





Who am I

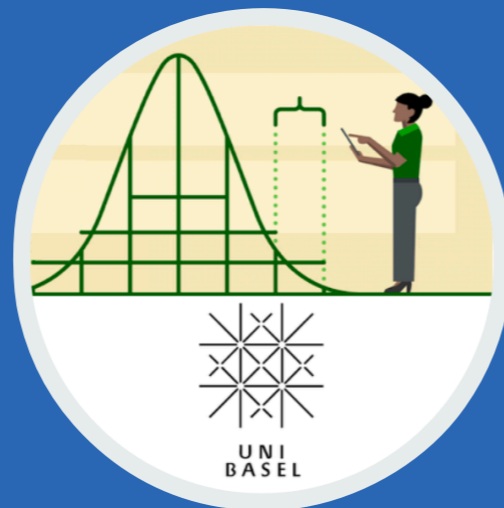


WE TEACHER

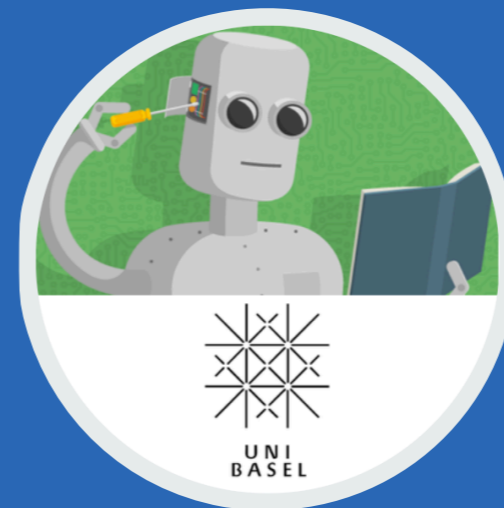
WWW.THERBOOTCAMP.COM



Explorative
Datenanalyse mit R



Statistik mit R



Machine learning mit R



Reporting mit R

Who am I



Cognitive and Decision Sciences

Am Zentrum für Cognitive and Decision Sciences (CDS) untersuchen wir, wie Menschen unterschiedlichen Alters Entscheidungen treffen. Wir verwenden komputationale Modelle, bildgebende Verfahren sowie Verhaltensexperimente um zu verstehen, wie kognitive und affektive Prozesse Entscheidungen beeinflussen, und somit Menschen zu besseren Entscheidungen im Alltag zu verhelfen.

Mehr Informationen zu unserer Arbeit finden Sie auf [Forschung](#) und [Blog](#).



The role of data analytics in psych

Four Phases and Major Events in Psychotherapy Research

Phase I (1900s-1940s) The Birth of Psychotherapy Research (Outcome and Process Research)	Phase II (1950s-1960s) The Search for Scientific Rigor	Phase III (1970-1980s) The Birth of Meta-Analysis and Institutionalisation of Randomized Control Trials (RCT)	Phase IV (1990-present) Institutionalization of Evidence-based Practice in Medicine and Psychology
<p>1899 Freud publishes <i>The Interpretation of Dreams</i></p> <p>1936 Publication of <i>Some Implicit Common Methods in Diverse Forms of Psychotherapy</i> by Saul Rosenzweig</p> <p>1940 Carl Rogers records and transcribes therapy sessions</p> <p>1946 Publication of the <i>International Classification of Diseases</i> manual (ICD-6) that first included a section on mental disorders</p> <p>1948 Use of placebo control design by Medical Research Council</p>	<p>1951 Publication of <i>Client-centered Therapy</i> by Carl Rogers</p> <p>1952 Publication of the first <i>Diagnostic and Statistical Manual of Mental Disorders</i> (DSM)</p> <p>1952 Hans Eysenck suggests that effects of psychotherapy are equal or less than spontaneous recovery</p> <p>1956 Placebo control in psychotherapy research advocated by Rosenthal & Frank</p> <p>1958 Publication of <i>Psychotherapy by Reciprocal Inhibition</i> by Joseph Wolpe</p> <p>1958 APA and NIMH organize conferences on Psychotherapy</p>	<p>1975 Publication of <i>Cognitive Therapy</i> by Beck, Mahoney, & Meichenbaum</p> <p>1976 Gene Glass coins the term meta-analysis</p> <p>1977 Meta-analysis shows that psychotherapy is effective by Smith & Glass</p> <p>1979 Systematic analysis of language in psychotherapy by Russell & Stiles</p> <p>1980 FDA requires double-blind placebo design</p> <p>1982 Publication of <i>Converging Themes in Psychotherapy</i>, a book promoting common factors theory</p>	<p>1993 Cochrane Collaboration is founded to systematise (and communicate) evidence in medicine (in 1999 sister organisation for the social sciences, Campbell, is founded)</p> <p>1995 Empirically supported treatments (EST) designated by Div. 12 (Clinical Psychology) APA on the basis of RCTs</p> <p>2001 Institute of Medicine adopts evidence-based practice in medicine</p> <p>2006 APA adopts evidence-based practice in psychology</p>

Russell, R. L., & Orlinsky, D. E. (1996). Psychotherapy research in historical perspective: Implications for mental health care policy. *Archives of General Psychiatry*, 53(8), 708–715.

Wampold, B. E. & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge: New York.

The role of data analytics in psych

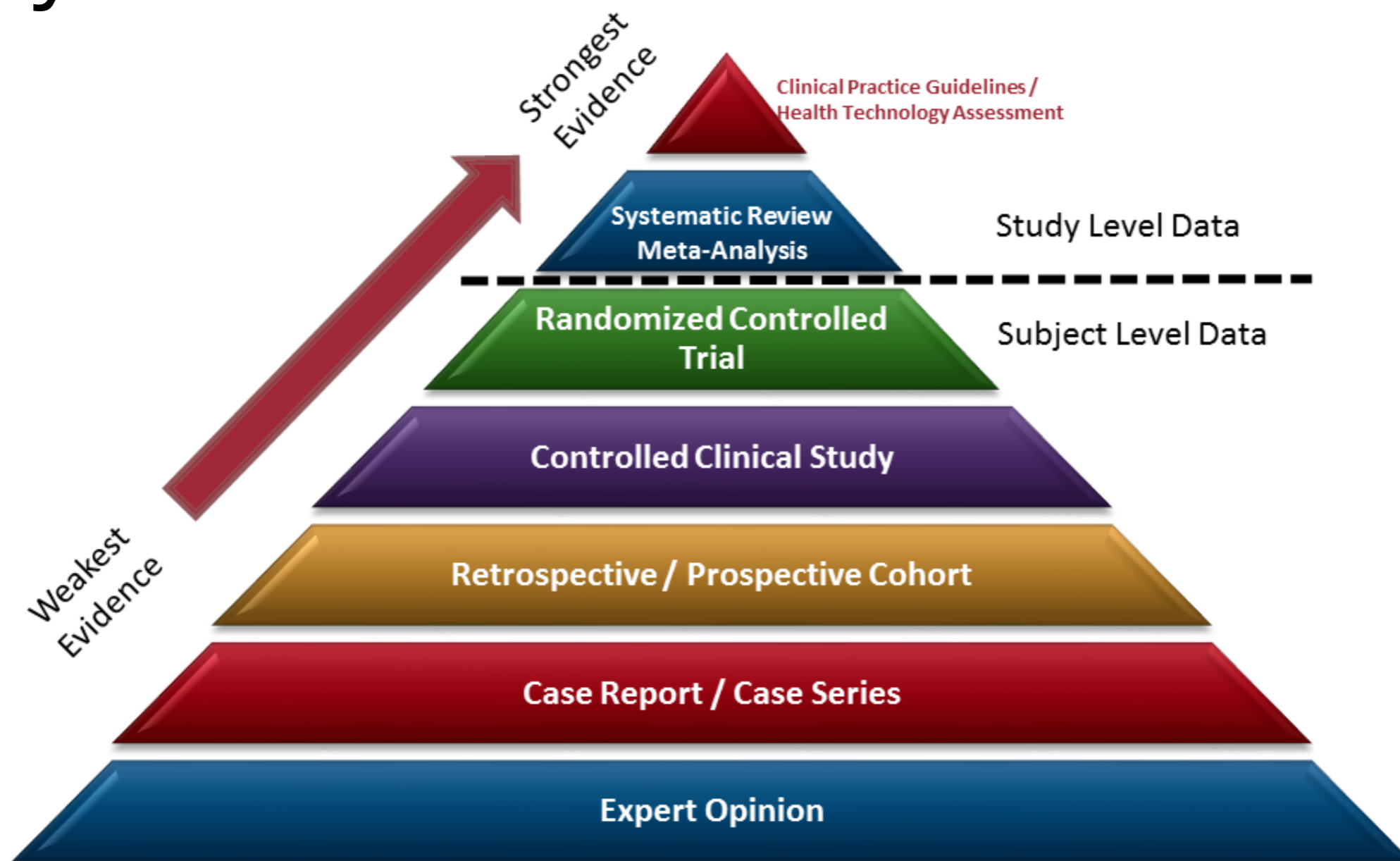
Policy Statement on Evidence-Based Practice in Psychology

The following statement was approved as policy of the American Psychological Association (APA) by the APA Council of Representatives during its August, 2005 meeting.

Evidence-based practice in psychology (EBPP) is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences. This definition of EBPP closely parallels the definition of evidence-based practice adopted by the Institute of Medicine (2001, p. 147) as adapted from Sackett and colleagues (2000): "**Evidence-based practice is the integration of best research evidence with clinical expertise and patient values.**"

Best research evidence refers to scientific results related to intervention strategies, assessment, clinical problems, and patient populations in laboratory and field settings as well as to clinically relevant results of basic research in psychology and related fields. A sizeable body of evidence drawn from a variety of research designs and methodologies attests to the effectiveness of psychological practices. Generally, evidence derived from clinically relevant research on psychological practices should be based on systematic reviews, reasonable effect sizes, statistical and clinical significance, and a body of supporting evidence. The validity of conclusions from research on interventions is based on a general progression from clinical observation through systematic reviews of randomized clinical trials, while also recognizing gaps and limitations in the existing literature and its applicability to the specific case at hand (APA, 2002). Health policy and practice are also informed by research using a variety of methods in such areas as public health, epidemiology, human development, social relations, and neuroscience.

The role of data analytics in psych



https://en.wikipedia.org/wiki/Levels_of_evidence

The role of data analytics in business and beyond

DATA SCIENCE AND ITS RELATIONSHIP TO BIG DATA AND DATA-DRIVEN DECISION MAKING

Foster Provost¹ and Tom Fawcett²

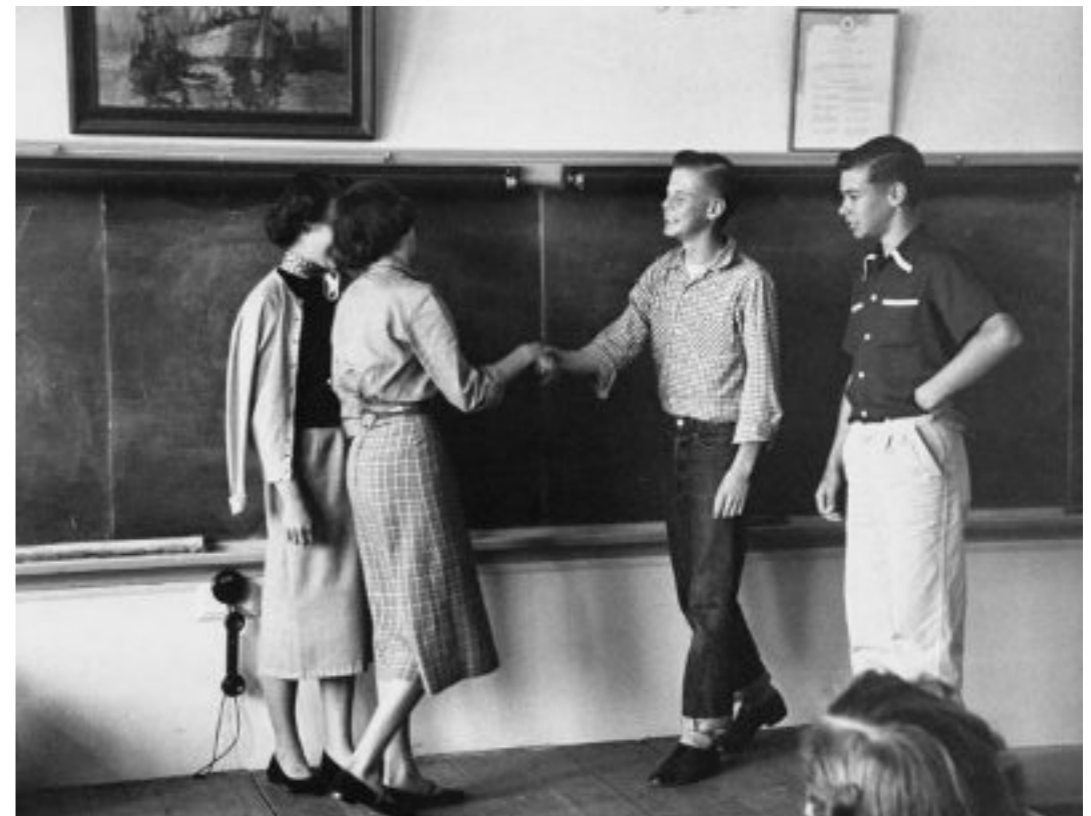


Abstract

Companies have realized they need to hire data scientists, academic institutions are scrambling to put together data-science programs, and publications are touting data science as a hot—even “sexy”—career choice. However, there is confusion about what exactly data science is, and this confusion could lead to disillusionment as the concept diffuses into meaningless buzz. In this article, we argue that there are good reasons why it has been hard to pin down exactly what is data science. One reason is that data science is intricately intertwined with other important concepts also of growing importance, such as big data and data-driven decision making. Another reason is the natural tendency to associate what a practitioner does with the definition of the practitioner’s field; this can result in overlooking the fundamentals of the field. We believe that trying to define the boundaries of data science precisely is not of the utmost importance. We can debate the boundaries of the field in an academic setting, but in order for data science to serve business effectively, it is important (i) to understand its relationships to other important related concepts, and (ii) to begin to identify the fundamental principles underlying data science. Once we embrace (ii), we can much better understand and explain exactly what data science has to offer. Furthermore, only once we embrace (ii) should we be comfortable calling it data science. In this article, we present a perspective that addresses all these concepts. We close by offering, as examples, a partial list of fundamental principles underlying data science.

Introduction round

1. What is your name?
2. Why did I choose the seminar?
3. What is my experience with data analysis?
4. What do you want to learn?
5. What do you want to learn it for or what are your plans for the future?



cdsbase1.github.io/dataanalytics

Data Analytics for Psychology and Business

[Syllabus](#) [Readings](#) [Materials](#) [Projects](#) [Installation](#)



Welcome



This is the home of the course *Data Analytics for Psychology and Business* in FS2020 at the Faculty of Psychology, University of Basel.

Syllabus

cdsbase1.github.io/dataanalytics

Today

Session 1

Introduction; project groups; explorative data analysis

March 13

Session 2

Statistical model and causal inference; machine learning; blitz talks

April 24

Session 3

In-class project work

May 15

Session 4

Project presentations; meet-the-expert

Project time line

cdsbase1.github.io/dataanalytics

When

What

21.02.

Building groups

21.02. - 04.03.

Find 1-3 project ideas and corresponding datasets

04.03. - **noon**

Email ideas to **Rui**

04.03. - **evening**

Receive feedback on project ideas

04.03. - 13.03.

Prepare project proposal and pitch

13.03.

Pitch project proposal

13.03. - 24.04.

Project work

24.04.

In-class project support

24.04. - 15.05.

Finalize project and prepare presentation

15.05.

Deliver final presentation