Data Mining

DAPT 631

Vishal Patel

Spring 2022

DERIVE
www.derive.io

- **Vishal**
- Founder of DERIVE, LLC
- MS in Computer Science (IIT, Chicago), and
  MS in Decision Sciences (VCU, Richmond)
- Mining data since 2003

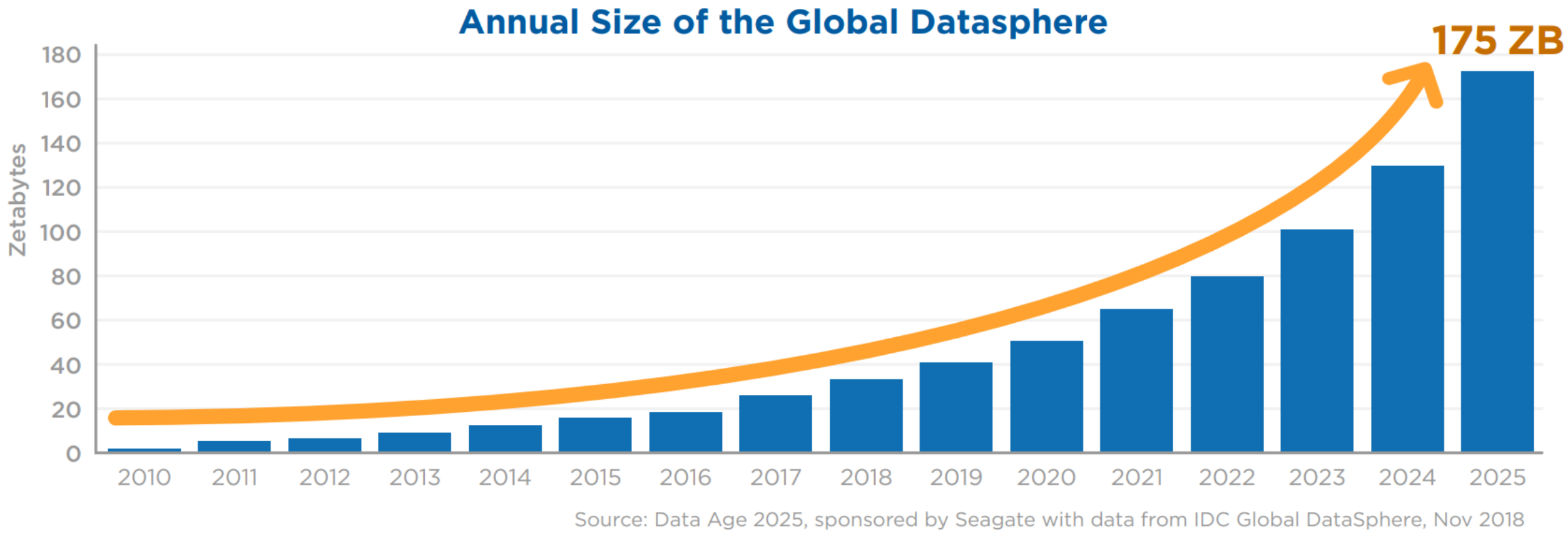
Synchronoss
NORTH HIGHLAND
Michaels
ActiveCampaign
MOEN
BOSS HUGO BOSS
Verdata
The Container Store®
AMERICA'S TEST KITCHEN
Neutrogena®
PERFETTI van Melle
SoFi
BABYLON harvested here.
biogen idec
PETRO HOME SERVICES
Virgin mobile
McDonald's
KEURIG™ GREEN MOUNTAIN
Pizza Hut®
BLOCKBUSTER®
IBM
SEARS
JAGUAR
Sprint
AARP®
Humana®
sanofi aventis
Michelob ULTRA
NEXTEL®

- Introduction
- History
- Course Structure

Introduction

- History
- Course Structure


Cambrian Era



One zettabyte = One trillion gigabytes = One sextillion (10^{21}) bytes

For comparison, the universe is 4×10^{17} seconds old.

INFRASTRUCTURE

STORAGE
amazon S3, Azure Storage, Google Cloud Storage, IBM Storage, PURESTORAGE, wasabi, ALLUXIO, panasas, nimblestorage, COHESITY, Qumulo, VAST, CLUMIO

HADOOP
CLOUDERA, amazon EMR, Google Cloud Dataproc, Azure HDInsight, HPE Ezmeral Data Fabric, Pivotal, jethro, SAP Cloud Platform, IBM InfoSphere BigInsights, IBM InfoSphere, arm TREASURE DATA

DATA LAKES
Azure Data Lake Storage, aws Lake Formation, IBM Power Systems, Yellowbrick, Qubole, dremio, CAZENA

DATA WAREHOUSES
amazon REDSHIFT, Google BigQuery, Google Cloud Spanner, ORACLE Autonomous Data Warehouse, Pivotal, Infoworks, snowflake, FIREBOLT

STREAMING / IN-MEMORY
amazon Kinesis, Google Cloud Dataflow, databricks, SAP Cloud Platform, ORACLE GoldenGate, confluent, GIGASPACES, GridGain, hazelcast, kx, Materialize, Wellaroo.io, FASTDATA, memsql

NoSQL DATABASES
Google Cloud Bigtable, amazon DynamoDB, amazon DocumentDB, Azure CosmosDB, ORACLE, mongoDB, MarkLogic, Couchbase, DATASTAX, redislabs, AEROSPIKE, SCYLLA, ArangoDB

NewSQL DATABASES
SAP HANA, Amazon Aurora, nuodb, Cloud Spanner, Pivotal, MariaDB, imply, Clustrix, MEMSQL, influxdata, Cockroach Labs, splice machine, VOLTDB, TiDB, TIMESCALE, Trafodion, paradigm4

GRAPH DBs
neo4j, Amazon Neptune, IBM graph, ORACLE, OrientDB, STARDOG, InfiniteGraph, Objectivity, TigerGraph

MPP DBs
TERADATA, VERTICA, IBM Data Warehouse Systems, actian, Kognitio, Exasol

SERVER-LESS
Microsoft Azure, Azure Functions, Google Cloud Functions, nuclio, Huawei Function Service

CLUSTER SVCS
Amazon ECS, IBM Power Systems, IBM, Amazon EKS, Google Cloud Dataproc, MESOSPHERE, packet, PENGUIN COMPUTING, CYCLECLOUD, Bright Computing, Hewlett Packard Enterprise, anyscale

ETL / DATA TRANSFORMATION
talend, pentaho, alteryx, TRIFACTA, amazon Glue, amazon Athena, Azure Data Factory, Paxata, StreamSets, INFIFI, tamr, FIGHTBYTE ANALYTICS, dataform

DATA INTEGRATION
SAP Data Services, Segment, Informatica, MuleSoft, TEALIUM, snapLogic, Stitch, Qlik Data Catalyst, tray.io, Fivetran, ATTUNITY, ZALONI, import.io, MATILLION, Infoworks, SNOWPLOW, Narrator, Census, SEEKWELL

DATA GOVERNANCE
Informatica, SailPoint, IBM, Alation, collibra, McAfee Skyhigh Security Cloud, dremio, MANTA, IMMUTA, OKERA, Waterline Data, data.world, PRIVAGERA

DATA QUALITY
talend, TORO, SODA, datafold, Databand, precisely, great_expectations, MONTE CARLO

MGMT / MONITORING
amazon Cloudwatch, New Relic, DATADOG, splunk>, APPDYNAMICS, rubrik, dynatrace, solarwinds, druva, WAVEFRONT by VMWARE, actifio, Numerify, Moogsoft, SCALYR, chronosphere, unravel, pagerduty, zenoss, ScienceLogic, MAGNITUDE, veeam, Grafana Labs, OpsRamp

DATA GENERATION & LABELLING
amazon mechanical turk, HIVE, scale, Upwork, appen, unity, LIONBRIDGE, Labelbox, AI.REVERIE

AI OPS
ALGORITHMIA, HPE Ezmeral ML Ops, Weights & Biases, comet, Sagemaker, Determined AI, Verta.ai, Dataiku, SPELL, fiddler, datatron, datmo, EVIDENTLY AI

GPU DBs & CLOUD
kinetica, Paperspace, SCREAM, omnisci, FAUNA, brytlyt, PG-Strom, blazingdb, FLOYDHUB

AI HARDWARE
Google TPU, arm, intel AI, NVIDIA, 3DFX Power Systems, GRAPHCORE, intel Nervana, WAVE COMPUTING, covariant, CEREBO, HAILO, MYTHIC, Movidius, Hewlett Packard Enterprise, vicarious, CERNAN, Cerebras

ANALYTICS & MACHINE INTELLIGENCE

BI PLATFORMS
looker, Google Cloud Looker analytics, amazon QuickSight, DOMO, ThoughtSpot, SISENSE, ATSCALE, Qlik, GoodData, Information Builders, birst, MicroStrategy, Keen IO

VISUALIZATION
tableau, Microsoft Power BI, SAP, Google Cloud Data Studio, Periscope Data, zepl, YOMDATA, plotly, CHARTIO, TOUCAN TOCO, Observable, GRID

DATA ANALYST PLATFORMS
Microsoft, pentaho, alteryx, sisu, MODE, Dataweer, Starburst, anodot, outlier, Digital Reasoning, ENDOR, Microsoft Power BI, ATTIVIO, incorta, interana, GUAVUS, ASCEND.IO, switchboard

DATA SCIENCE NOTEBOOKS
jupyterhub, binder, colab, SaturnCloud, Google Cloud Dataproc, Azure Notebooks, Amazon SageMaker, Deepnote, COCALC, Nextjournal, count

DATA SCIENCE PLATFORMS
IBM Watson Studio, databricks, data iku, ORACLE Data Science Platform, rapidminer, sas, ANACONDA, Altair Knowledge Works, TIBCO, KNIME, DOMINO, MathWorks

MACHINE LEARNING
Azure Machine Learning, Amazon SageMaker, Google Cloud AutoML Vision, H2O.ai, DataRobot, gamalon, VISENZE, ELEMENT AI, deepsense.ai, OctoML

COMPUTER VISION
Face++ 旷视, 商汤, Microsoft Azure, AWS Rekognition, Cloud Vision API, clarifai, EVERAI, deepomatic, neurala, twentybn, UBIQUITY6, ADEE, 依图 YITU, trax, ELEMENT AI, synthesia, DataGrid, onfido

HORIZONTAL AI
IBM Watson, Cortana, NEURAL NETWORKS, Voyager Labs, Affectiva, PROPHESEE, CognitiveScale, Numenta, PETUUM, SCAPE TECHNOLOGIES, narologics, OSARO, CURIOUS AI, BLUE VISION, Paradigm, Pelorian

SPEECH & NLP
Google Natural Language API, twilio, amazon alexa, amazon Translate, amazon Polly, amazon Transcribe, amazon Comprehend, Azure Cognitive Services, Wellsaid Labs, semantic machines, Mobvoi, Liulishuo, SoundHound Inc., PRIMER, Intelligent Voice, SMARTLING, Unbabel, Babylon Health, NUANCE, voicera, MonkeyLearn, Grid.speech, cogito, PolyAI, Verbalytics, Lexical Analytics, Lexalytics, cortical.io

SEARCH
elasticsearch, amazon Kendra, ORACLE ENDECA, amazon CloudSearch, Azure Cognitive Services, algolia, coveo, SINEQUA, Lucidworks, ATTIVIO, swiftype, EXALEAD, alphasense, omni:us, MAANA

LOG ANALYTICS
splunk>, Google Cloud Logging, sumologic, solarwinds loggly, kibana, TIMBER, logz.io, logdna

SOCIAL ANALYTICS
Hootsuite, sprinklr, NETBASE, synthesio, simplereach, bitly, tracx, SimilarWeb

WEB / MOBILE / COMMERCE ANALYTICS
Google Analytics, mixpanel, SIGOPT, Airtable, RESCI, granify, Amplitude

APPLICATIONS – ENTERPRISE

SALES
einstein, BANG, CHORUS, conversica, aviso, peopleai, ClearBit, tact.ai, fuse, machines

MARKETING - B2B
App Annie, Lattice, sense, Reflektion, tubular, ENGAGIO

MARKETING - B2C
TEALIUM, ACTIONIQ, Segment, Simon, attentive, Active Campaign, SendGrid, CONTENTSQUARE, zeta, mparticle, Amperity, bloomreach, Bluecore, INVOCA, braze, [PERSADO]

CUSTOMER EXPERIENCE / SERVICE
qualtrics, SurveyMonkey, UserTesting, CLARABRIDGE, MEDALLIA, zendesk, Kustomer, freshdesk, Gainsight, pendo, HEAP, Amplitude, Watson Assistant, Dialogflow, LIVEPERSON, INTERCOM, Drift, ASAPP, ada, afiniti, nextiva, CallDesk, aircall, Eureka, talla, frame.ai

HUMAN CAPITAL
HireVue, pymetrics, textio, PI, mya, Allyo, Wade&Wendy, ere

LEGAL
RAVEL, Seal, DISCO, ContractPodai, kira, ROSS

PARTNERSHIPS
OneTrust, GRIDSMART, DATA REPUBLIC, INFOSUM

REGTECH & COMPLIANCE
Bigid, text IQ, TESSIAN, Comply Advantage, ETHYCA, securiti.ai

FINANCE
Anaplan, zuora, SAP S/4 HANA, TRADESHIFT, SCALEFACTOR, botkeeper, pilot, DIGITS, appzen

AUTOMATION & RPA
UiPath, HyperScience, ZINIER, blueprism, VIDADO, WorkFusion, workato, Catalytic, KRYON, ARTWORK, ALKYMI, Fortress IQ, Nightingale, Electric, INSTABASE, Moveworks, celonis

SECURITY
TANIUM, BlackBerry Spark, zscaler, StackPath, Illumio, FORTER, feedzai, DATAVISOR, pindrop, sift science, riskrecon, ANOMALI, Recorded Future, SHIELD AI, sparkcognition, cybereason, DARKTRACE, ThreatMetrix, VECTRA, SOCURE, CLOUDFLARE, exabeam, SENTINELONE, BLUEHEXAGON, Kaspersky, GRIDGUARD, Armorblox, CyberAngel, PRIVITAR, netskope, risk.ified

APPLICATIONS – INDUSTRY

ADVERTISING
AppNexus, MediaMath, criteo, IAS, ORACLE MOAT, salesforce datorama, albert, gumgum, Appier, theTradeDesk, TAPAD

EDUCATION
Liulishuo, 猿辅导, KNEWTON, declara, 松鼠AI Squirrel AI Learning, KORBIT

REAL ESTATE
REDFIN, VTS, Opendoor, Orchard, reonomy, SKYLINE, SPACEMAKER, GEOPHY

GOV'T & INTELLIGENCE
Palantir, OPENGOV, Dataminr, MARK43, ANDURIL, FiscalNote, Quid, PRIMER

COMMERCE
FAIRE, STITCH FIX, HowGood, STANDARD

FINANCE - INVESTING
Symphony AYASDI, KENSHO, ADDEPAR, NUMERAI

FINANCE - LENDING
affirm, Monedo, TALA, ZEST, upgrade, LEAD BANC, aura, Upstart, 100Credit, Active.AI

INSURANCE
ROOT, Metromile, Lemonade, GUIDEWIRE, Shift Technology, ARTURO, CAPE ANALYTICS, TRUTHFINDER, EvolutionIQ, ZELROS, zesty.ai

HEALTHCARE
flatiron, KYRUUS, METABIOTA, babylon, 3DMed, PathAI, TEMPUS, patientslikeme, AICure, LinkDoc, swipe health, Olive, imagen, komodo health, PAIGE, health, innovaccer, lumiata, springhealth, THE HUMAN DIAGNOSIS PROJECT, Vida, enlitic, ARTERYS, ezra, Caption Health, IMAGEN, zebra

LIFE SCIENCES
Benevolent AI, color, Genome Medical, DNAnexus, verily, Genome Biology, Z, genuity, inato, freenome, Atomwise, Phenomic AI, SOPHIA, OWKIN, insitro, Exscientia, RECURSION, ConcertAI, healx, Verana Health

TRANSPORTATION
UBER, TESLA, WAYMO, ZOOX, cruise, nuro, APTIV, CAMBRIDGE MOBILE TELEMATICS, Aurora, nauto, AUROTIVE, AIRO, PILOT.AI, NIO, OPTIMUS, G7, nexar, Kodiak, moovit, blaze, GHOST, PERCEPTIVE AUTOMATA

AGRICULTURE
FARMERS' BUSINESS NETWORK, Granular, JOHN DEERE, BLUE RIVER, TARANIS, prospera, semios

INDUSTRIAL
AVEVA, SIEMENS, Sightmachine, UPTAKE, SHIFT ENERGY, PREDIX, KONUX, SCORTEX, TACHYUS, ALICE, BUILT ROBOTICS, BEYOND LIMITS

OTHER
stem, Amper, ByteDance, STATESPACE, VERDIGRIS, Second Spectrum, duetto, Lightricks

OPEN SOURCE

FRAMEWORKS
TensorFlow, Spark, Hadoop, Flink, YARN, TEZ, kubernetes, MESOS, docker, CDAP, Red Hat, HighAvailability, HELIX

QUERY / DATA FLOW
Spark SQL, presto, HIVE, DRILL, SLAMDATA, GraphQL, Flink, HAWQ, Trafodion

DATA ACCESS & DATABASES
mongoDB, redis, Elasticsearch, cassandra, Cockroach Labs, riak, Amundsen, druid, Uber Databook, Dataportal, Cloud Spanner, SciDB, OpenTSDB, ACCUMULO, EDGEDB, HBASE

ORCHESTRATION & PIPELINES
Apache Airflow, talend, Apache Zookeeper, Luigi, Apache Ambari, MESOS, etcd, PREFECT, Kedro, DAGSTER, dbt, MARQUEZ, RAY, Flyte, DASK, argo, TEKTON

STREAMING & MESSAGING
Spark Streaming, kafka, beam, PULSAR, Flink, STORM, nifi, Apache RocketMQ

STAT TOOLS & LANGUAGES
python, R, Scala, NumPy, pandas, SciPy, Studio, julia

AI OPS & INFRA
mlflow, Kubeflow, DVC, SELDON, Polyaxon, BENTOML

AI / MACHINE LEARNING / DEEP LEARNING
TensorFlow, Keras, PyTorch, fastai, MXNet, Caffe, Microsoft Cognitive Toolkit, DMTK, OpenAI, theano, PaddlePaddle, Apache SINGA, DIMSUM, FeatureFu, VELES, neon, Chainer, Michelangelo, ONNX, WEKA, LOGIQ, PyTorch, DSSTNE, mlib, DL4J, MAHOUT, Aerosolve, fastai, mlr, OpenML, mindsdb

SEARCH
elasticsearch, Solr, Lucene, Sphinx, Sonic, Toshi Search

LOGGING & MONITORING
elasticsearch, kibana, SENTRY, logstash, Prometheus, fluentbit, Grafana, VECTOR

VISUALIZATION
Superset, matplotlib, Metabase, redash, TensorBoard, seaborn, bokeh

COLLABORATION
Beaker, jupyter, Zeppelin, ANACONDA

SECURITY
Apache Ranger, KNOX, Sentry, ACCUMULO, snyk

DATA SOURCES & APIs

DATA MARKETPLACES & DISCOVERY
aws Data Exchange, snowflake Data Exchange, DAWEX, EXPLORIUM, data.world, narrative

FINANCIAL & ECONOMIC DATA
Bloomberg, THOMSON REUTERS, DOW JONES, Quandl, S&P CAPITAL IQ, CB INSIGHTS, PLAID, SECOND MEASURE, ENVESTNET YODLEE, THE WORLD BANK, Xignite, StockTwits, Thinknum, earnest, predata, MX, tink, SmartNews

AIR / SPACE / SEA
Orbital Insight, AIROBOTICS, spire, Descartes Labs, WINDWARD, MarineTraffic, LOFT ORBITAL, RS Metics, ExoLabs, Qianxun SI, Synspective

PEOPLE / ENTITIES
zoominfo, acxiom, experian, EPSILON, InsideView, Brandwatch, Quantcast, BASIS TECHNOLOGY, SAFEGRAPH, Demyst, melissa

LOCATION INTELLIGENCE
FOURSQUARE, mapbox, sense360, pitney bowes, PlaceIQ, esri, CARTO, Radar, Mapillary, cuebiq, OpenStreetMap

OTHER
DATA.GOV, DEFENSE INNOVATION UNIT, IMAGENET, LabelMe, Berkeley DeepDrive, COCO, APOLLOSCAPE, CRUX

DATA RESOURCES

DATA SERVICES
QUANTUMBLACK, Booz | Allen | Hamilton, kaggle, ElectrifAI, fractal, EXL, DataKind, INNOPLEXUS

INCUBATORS & SCHOOLS
PLURALSIGHT, GENERAL ASSEMBLY, DataCamp, DataElite, galvanize, METIS, INSIGHT, The Data Incubator

RESEARCH
OpenAI, facebook research, MILA, MIRI, VECTOR INSTITUTE, CSAIL, DFKI, AI2 ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE

Robert Cohn, Peter Wang | Keynote: How Open Data Science Opens the World of Innovation

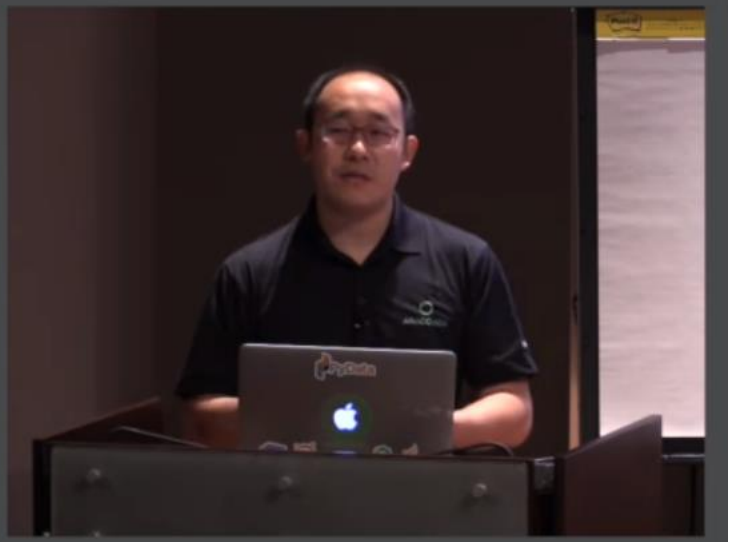
Watch later Share

CONTINUUM ANALYTICS

Era of Data Literacy

- Data exploration and analysis are going to be a new kind of **literacy** that will be required to do great work in any field.
- Language is a human instinct and is a natural path to insight. We see this in our interaction with Python/PyData users, whose passion chiefly stems from this *expressiveness* and *agility*.
- An analytical language is "**thoughtware**", not "software".

ANACONDA



PyData DC 2016

29:26 / 41:23

YouTube

Data mining is the process of
discovering patterns in large data sets
involving methods at the intersection of
machine learning, statistics, and database systems.

[Wikipedia]

Data mining is the extraction of
implicit, previously-unknown,
and potentially-useful information from data.

– Witten and Frank

Dictionary

data mining

data mining

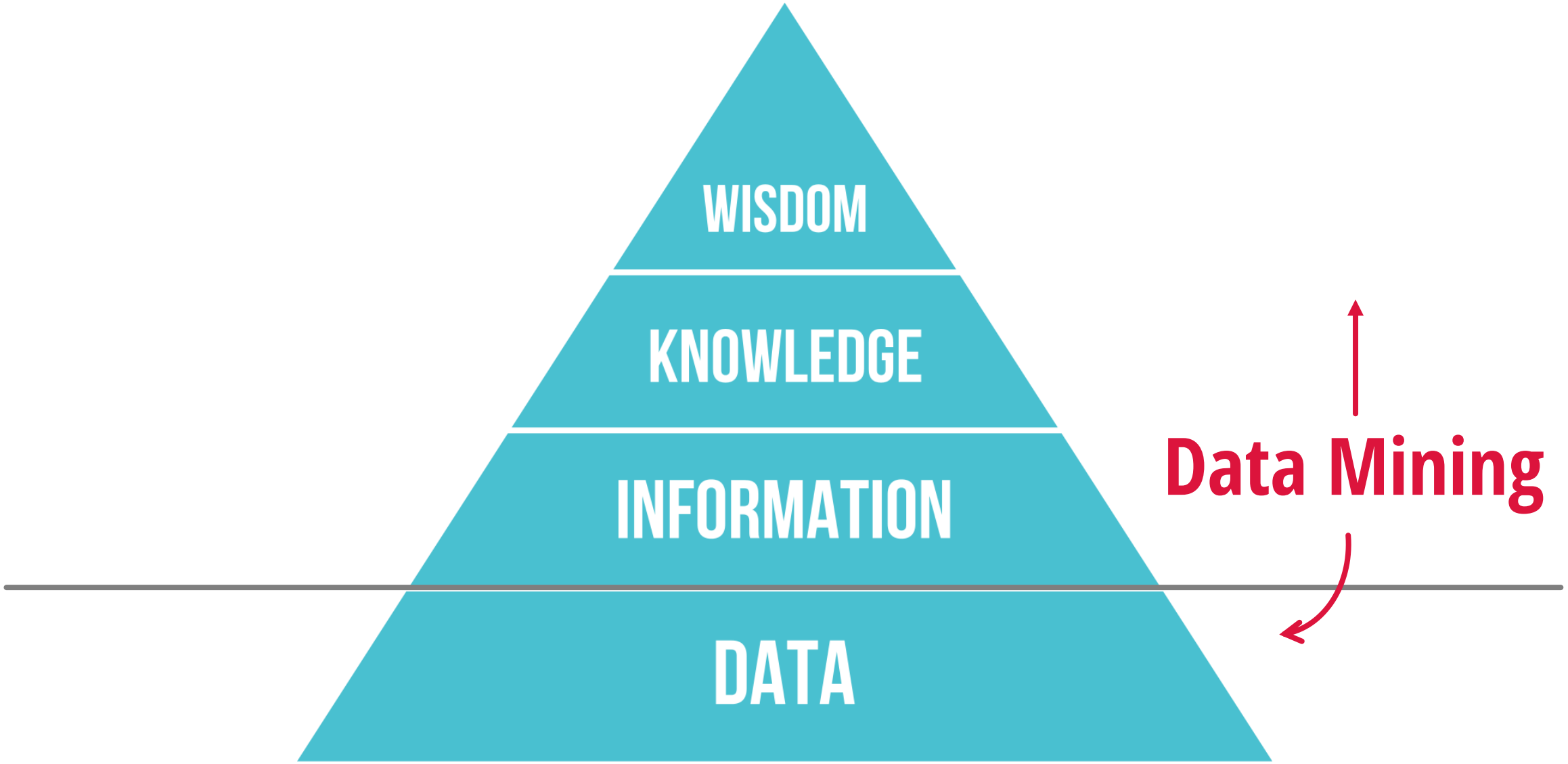
noun COMPUTING

noun: **data mining**; noun: **datamining**

the practice of examining large pre-existing databases in order to generate new information.

Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting
through large amounts of data stored in repositories,
using pattern recognition technologies as well as statistical and mathematical techniques.

– Gartner


WISDOM
KNOWLEDGE
INFORMATION
DATA
Data Mining


Data Mining
Data
Decision
Past
Future

Data Mining Tasks

- Description
- Estimation
- Prediction
- Classification
- Clustering

Applications of Data Mining

- Direct Marketing
- Fraud Detection
- Loan Approval
- Medical Diagnosis
- Music Reccos.
- Energy Forecasting
- Image Classification
- Market Basket Analysis
- Terrorism Surveillance
- Document Classification

- Introduction
- **History**
- Course Structure

Statistics

Census

Mortality tables

Accounting

From Latin: *status* state

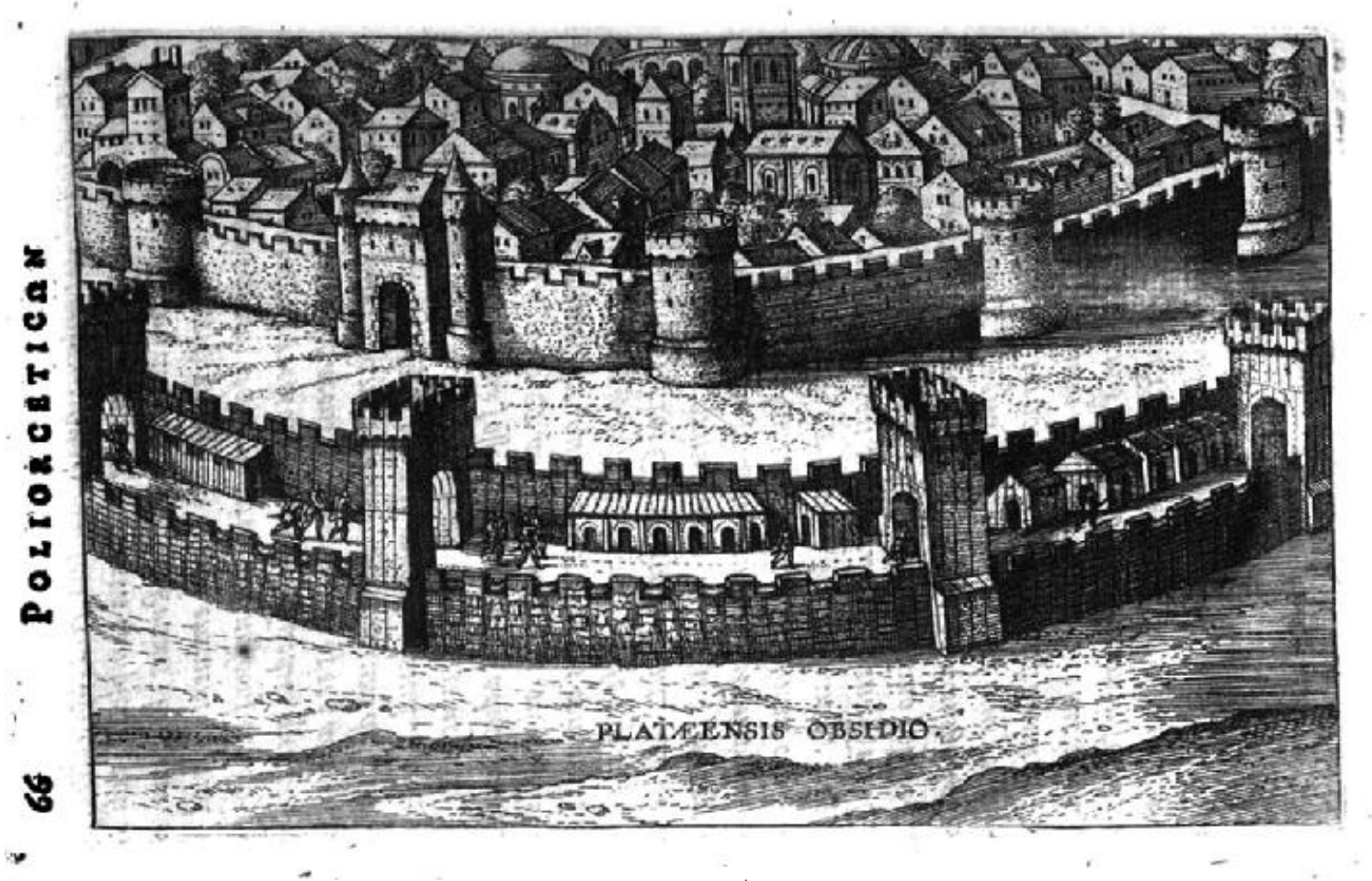
... teaches us what is the political arrangement of all modern states of the world.

W Hooper, 1770

DATA COLLECTIONS + ANALYSIS + DECISION MAKING

Statistics

EXAMPLE #1: UNCERTAINTY



Siege of Plataea (5th Century BCE)

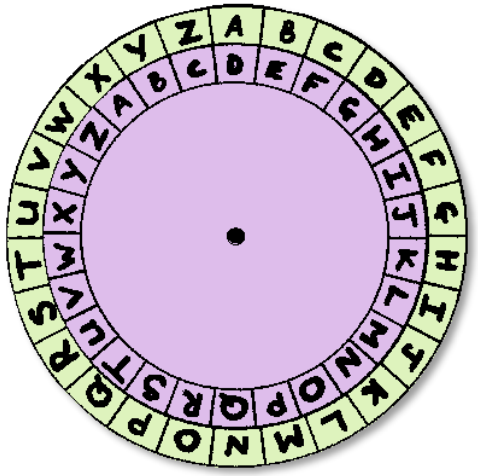
Poll 1

What do you think they used as the best estimate for the height of the wall?

A. Mean
B. Median
C. Mode
D. Max

Statistics

EXAMPLE #2 FREQUENCY ANALYSIS, CRYPTOANALYSIS

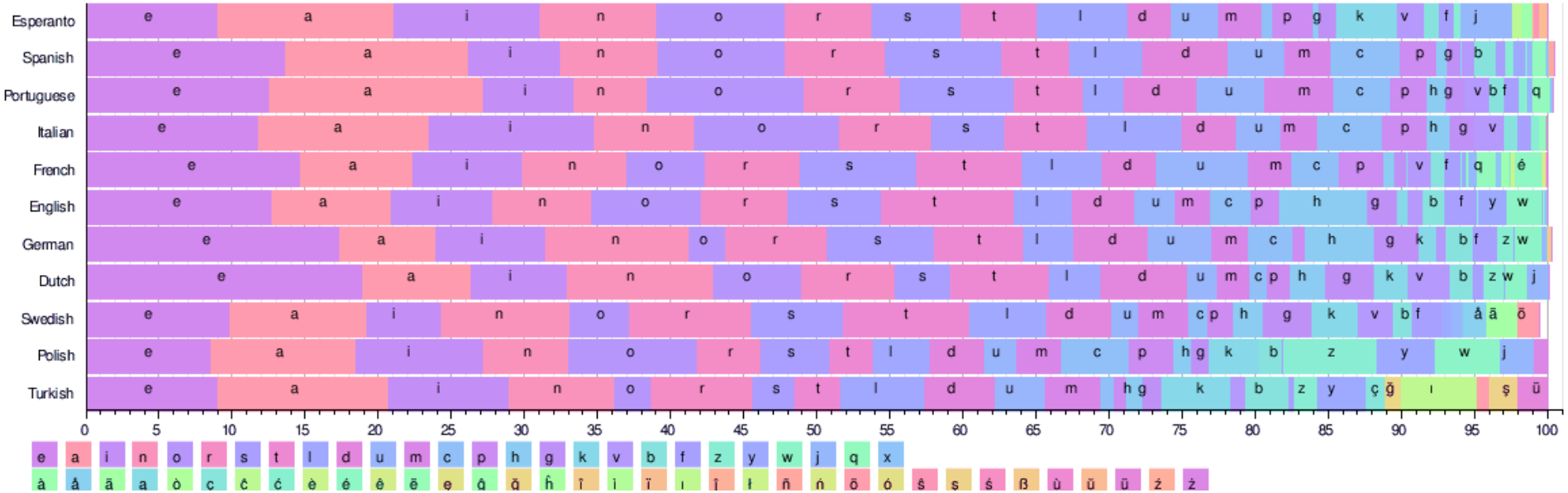


Caesar Cipher

Original message: Et tu, Brute?

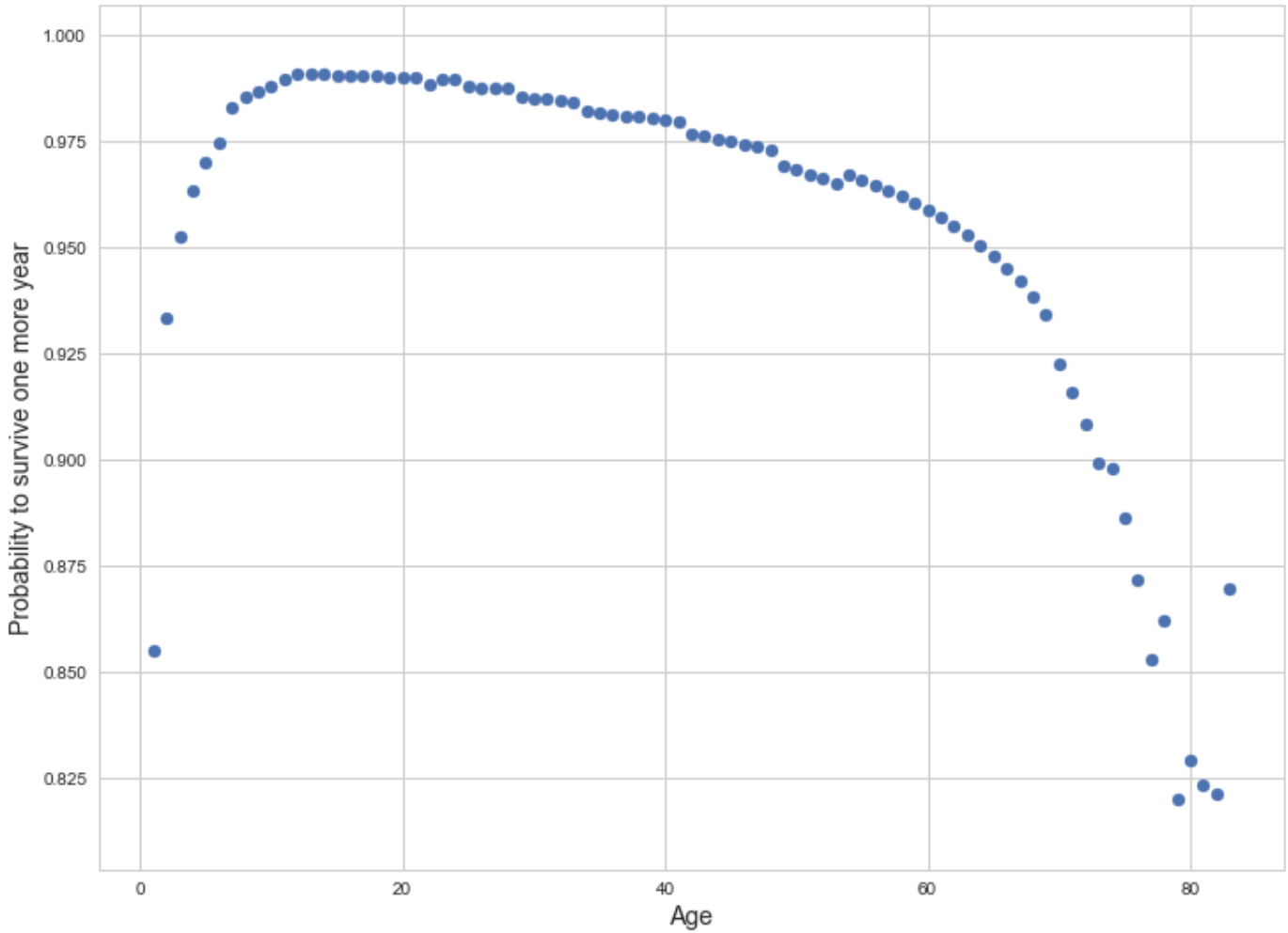
Encrypted message: Hw wx, Euxwh?

Al-Kindi
(801–873 AD)



Statistics

EXAMPLE #3 MORTALITY TABLES, DEMOGRAPHY



Data from Edmond Halley's *An Estimate of the Degrees of Mortality of Mankind (1693)*, table p.600.

The graph shows the probability of surviving one of more year(s) at a certain age.

Modern Statistics

Normal distribution

t distribution

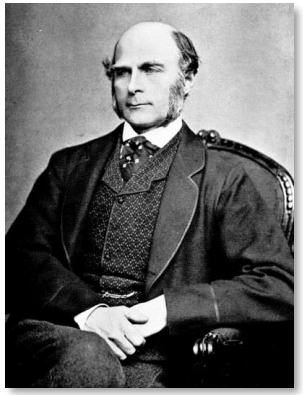
Random sampling

Design of-experiments

Bayesian Statistics

A rigorous mathematical discipline

for analysis, decision making, and inference



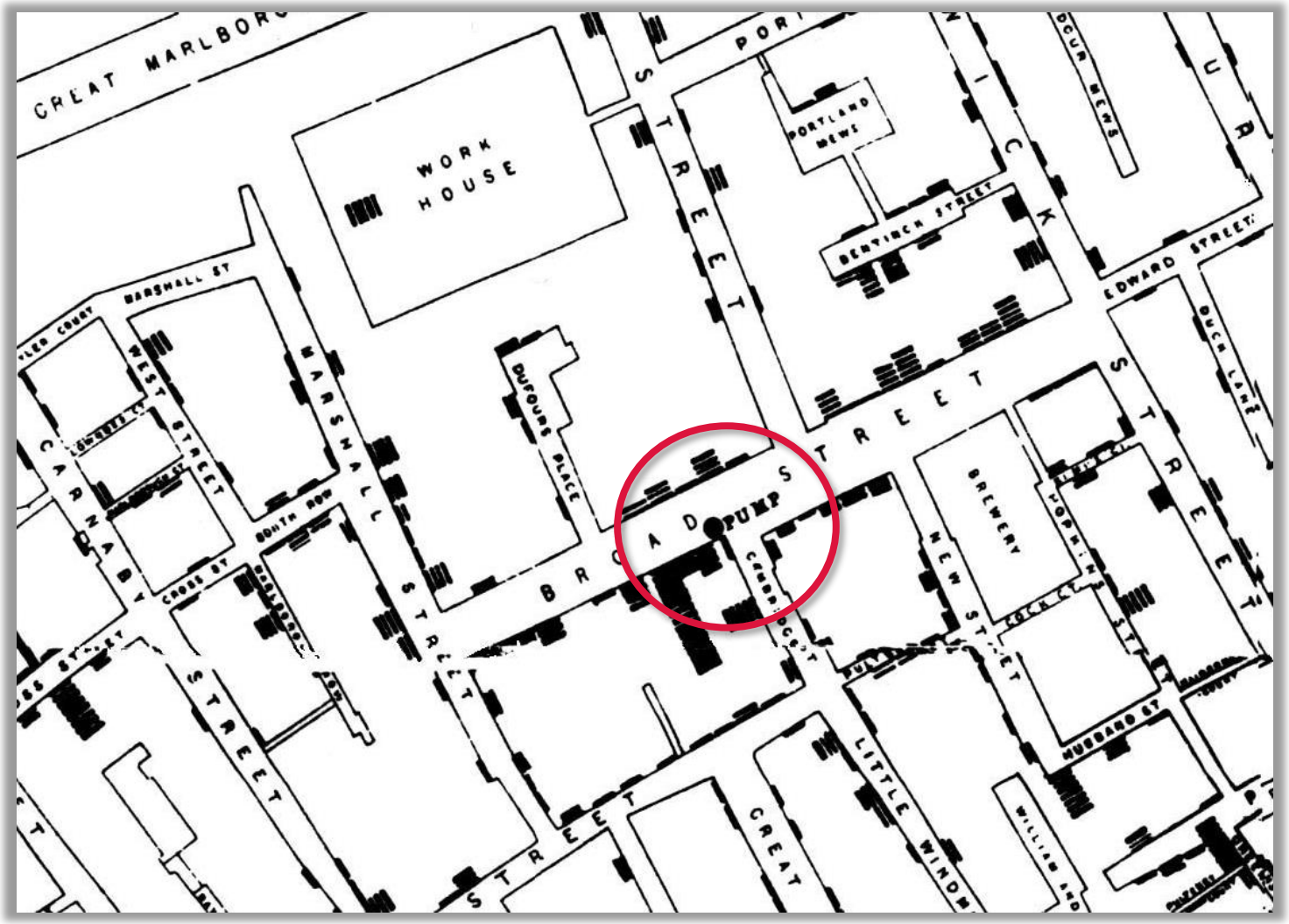
Sir Francis Galton
(1822–1911)
Correlation, regression

Carl Pearson
(1857–1936)
Founder of mathematical statistics

R A Fisher
(1890–1962)
ANOVA, Maximum Likelihood, DOE

Modern Statistics

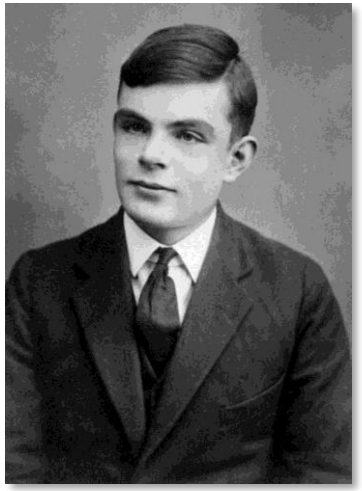
EXAMPLE: DATA VISUALIZATION



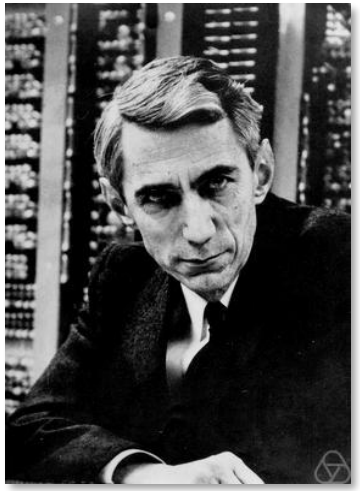
Original map by **John Snow** showing the clusters of cholera cases in the **London epidemic of 1854** [Source]

Data Mining

Algorithms &
Computation
Computer Science
Neural Networks
Decision Trees
Genetic Algorithms
Relational Databases



Alan Turing
(1912 –1954)
Theoretical Computer Science



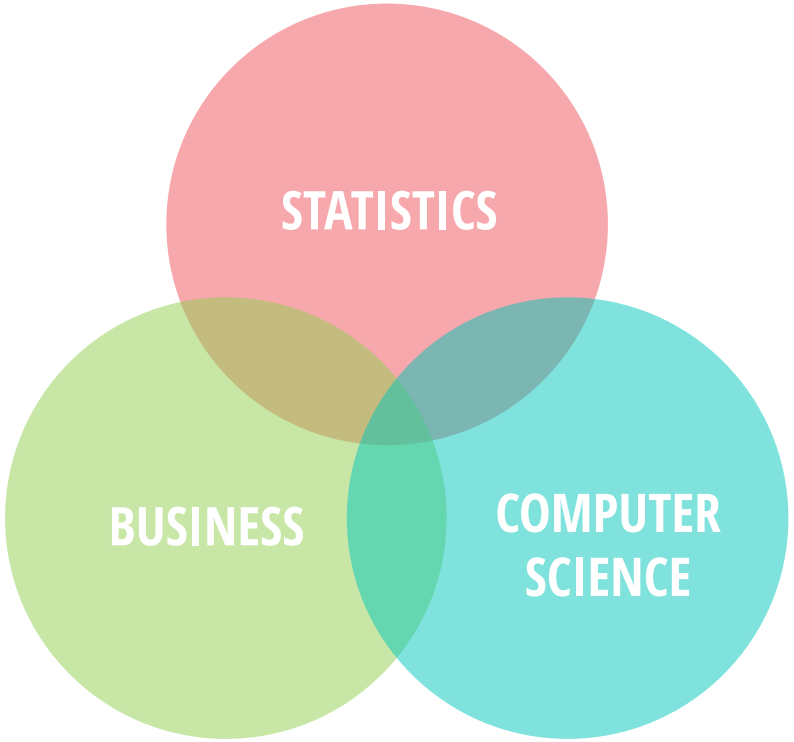
Claude Shannon
(1916 –2001)
Information Theory

- Warren McCulloch and Walter Pitts created a computational model for neural networks. (1943)
- John Holland introduced Genetic Algorithm based on the concept of Darwin's theory of evolution. (1960)
- E. F. Codd published an important paper to propose the use of a relational database model. (1970)

Data Science

Gradient Boosting
Random Forests
Support Vector-Machines
Recommender-systems
Unstructured data
Open source
Big Data

Data science is an interdisciplinary field
that uses scientific methods, processes, algorithms and systems
to extract knowledge and insights
from data in various forms,
both structured and unstructured,
similar to data mining.[†]



[†] From Wikipedia

Artificial Intelligence

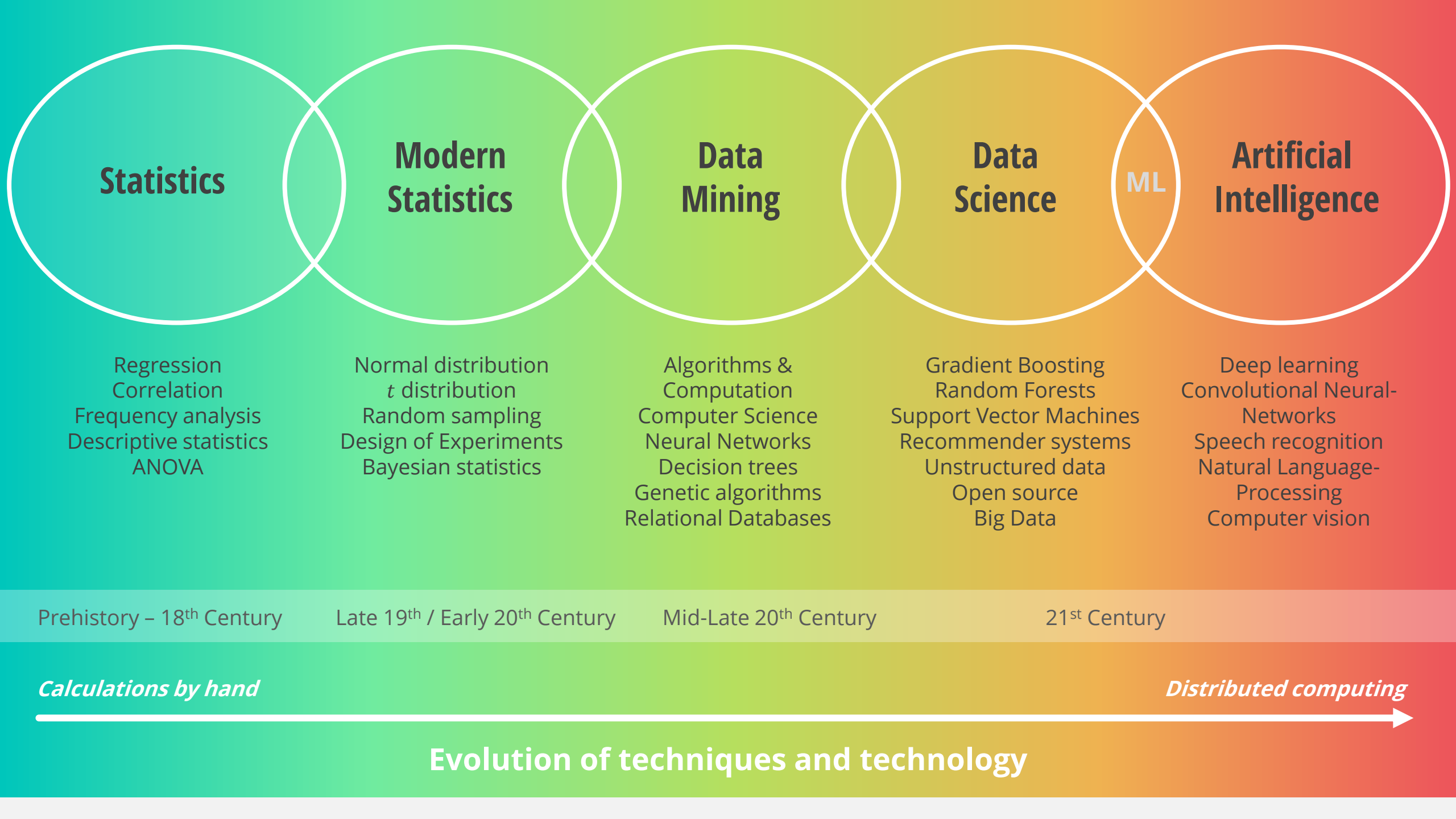
Deep learning

Convolutional-Neural Networks

Speech recognition

Natural Language-Processing

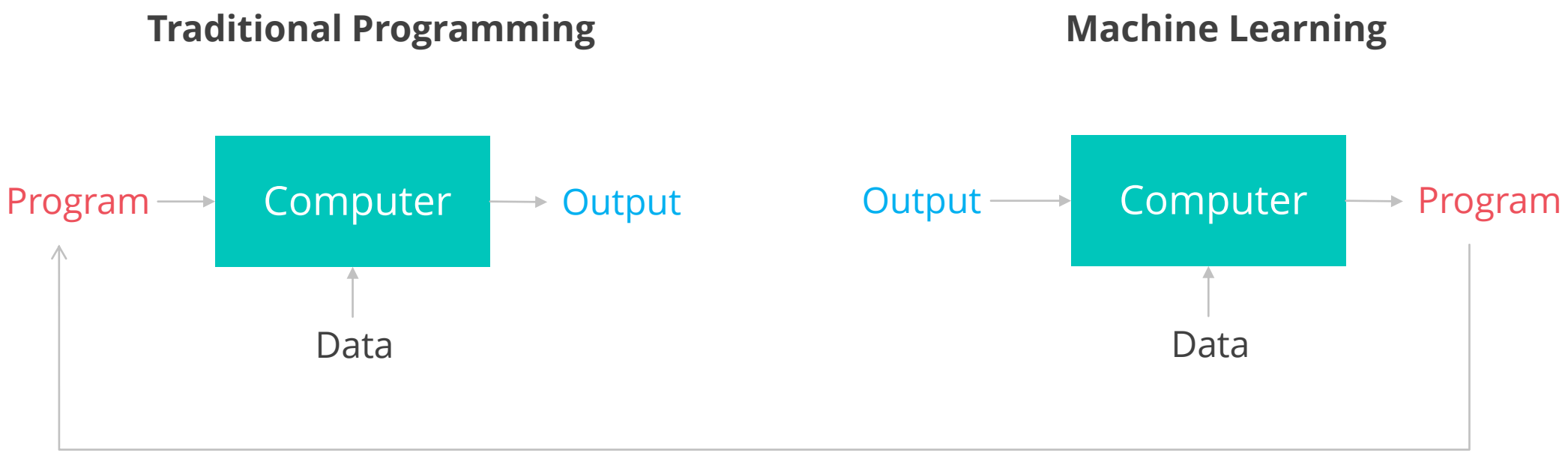
Computer vision


Statistics
Modern Statistics
Data Mining
Data Science
ML
Artificial Intelligence
Regression
Correlation
Frequency analysis
Descriptive statistics
ANOVA
Normal distribution
t distribution
Random sampling
Design of Experiments
Bayesian statistics
Algorithms & Computation
Computer Science
Neural Networks
Decision trees
Genetic algorithms
Relational Databases
Gradient Boosting
Random Forests
Support Vector Machines
Recommender systems
Unstructured data
Open source
Big Data
Deep learning
Convolutional Neural-Networks
Speech recognition
Natural Language-Processing
Computer vision
Prehistory – 18th Century
Late 19th / Early 20th Century
Mid-Late 20th Century
21st Century
Calculations by hand
Distributed computing
Evolution of techniques and technology

Machine Learning

Field of study that gives computers the ability to learn without being explicitly programmed.

Artur Samuel,1959



- Introduction
- History
- **Course Structure**

Data Science ≈ Data Mining

- The specific definitions and boundaries between these disciplines remain fuzzy.
- For the purpose of this class, I will use the terms ‘Data Science’ and ‘Data Mining’ interchangeably (with a preference to the former).
- We will cover several Data Science techniques in this class, e.g., Gradient Boosting.

Two Cultures

Statistics	Data Science
THEORETICAL	PRACTICAL
INFERENCE	PREDICTION
ASSUMPTIONS	EMPIRICAL
MANUAL	AUTOMATION

Course Outline

1. Introduction
2. The Data Science Process
3. Supervised Learning
4. Unsupervised Learning
5. Wrap Up

Class Structure

1. Ask questions at any time!
2. Collaboration is encouraged.
3. All content (course material) will be available on Blackboard (and on a git repository).
4. Data Mining + Python
5. Homework assignments in Python

My Objectives

1. Provide a practical knowledge of data mining algorithms.
2. Give a broader perspective to help understand what role data mining plays in the decision-making process.
3. Help you develop an appreciation for the beauty of the theoretical foundations underlying data mining.
4. Help you think more like a Data Scientist.
5. (For myself) Continue learning.

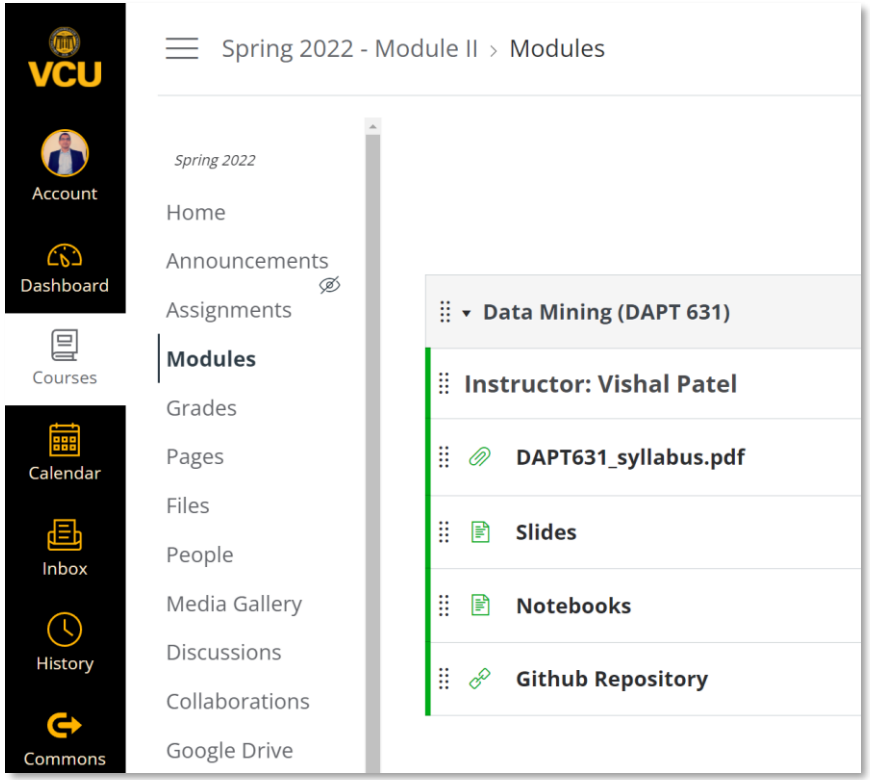
Data Mining + Python

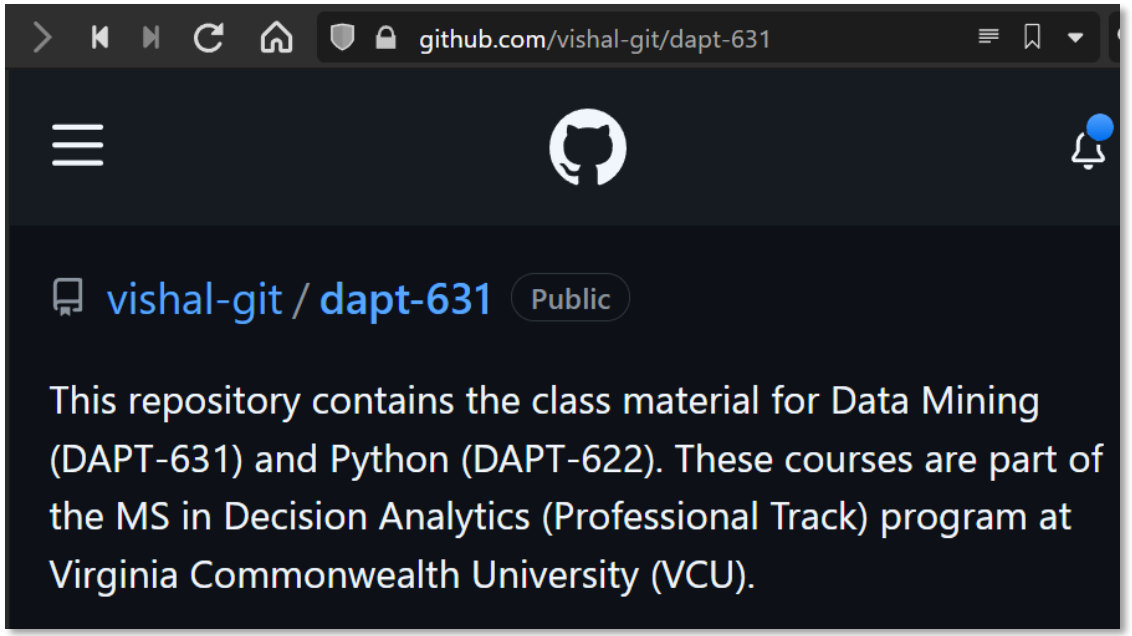
Learn Data Mining Concepts

→

Implement them in Python

Course Material





			HyFlex		HyFlex		HyFlex		HyFlex	
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9
	Saturday	8-Jan	22-Jan	5-Feb	19-Feb	5-Mar	19-Mar	2-Apr	16-Apr	30-Apr
	7:30 - 8:00	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast
Session 4	8:00 - 9:45	Data Mining	Python	Data Mining	Data Mining	Python	Data Mining	Data Mining	Python	Data Mining
Session 5	10:00 - 11:45	Data Mining	Python	Data Mining	Data Mining	Python	Data Mining	Data Mining	Python	Data Mining
	11:45 - 12:30	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch

vishal@derive.io

www.linkedIn.com/in/VishalJP

@derive_io