UNIVERSITY OF AMSTERDAM

A paradigm shift in auditing

From IT auditor to data scientist?

Prof. Dr. Edo Roos Lindgreen, UvA / KPMG

May 18, 2017

Paradigm shift

Paradigm: a distinct set of concepts or thought patterns, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field (Wikipedia).

“A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.“ – Max Planck

Kuhn, T. (1963). The Structure of Scientific Revolutions.

Paradigm shift in (IT) auditing

- Current approach based on obsolete paradigm
  - Defined, static and controlled environment
  - Sample testing and test of controls
  - Application controls / general IT controls

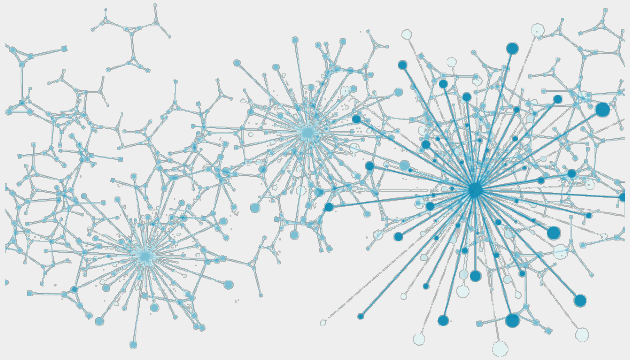
- New approach should be based on new paradigm
  - Fluid, dynamic, uncontrolled environment
  - 100% substantive testing, also for controls
  - Data analytics / machine learning


DARK MATERIA

Old wine in a new bottle?

- "Nothing new here. Auditors have used data for at least 100 years and digital data for at least 50 years."

- True. And the next 10 years, they will start to use:
  - enormous amounts of
  - structured and unstructured,
  - internal and external data
  - from many different sources
  - using algorithms, artificial intelligence, machine learning.



Effect of computerisation on audit

- Frey & Osborne, 2013
- Probability computerisation will lead to job losses within the next two decades

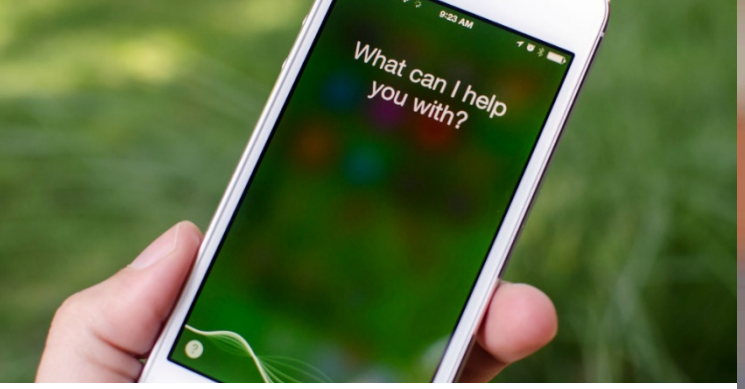
Bring on the personal trainers

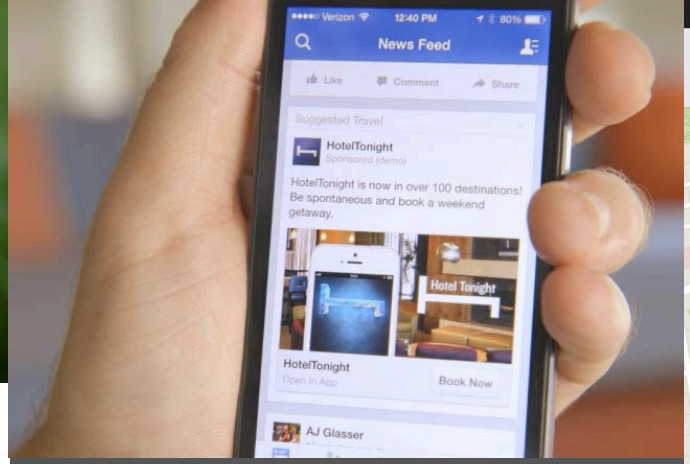
Probability that computerisation will lead to job losses within the next two decades, 2013
(1=certain)

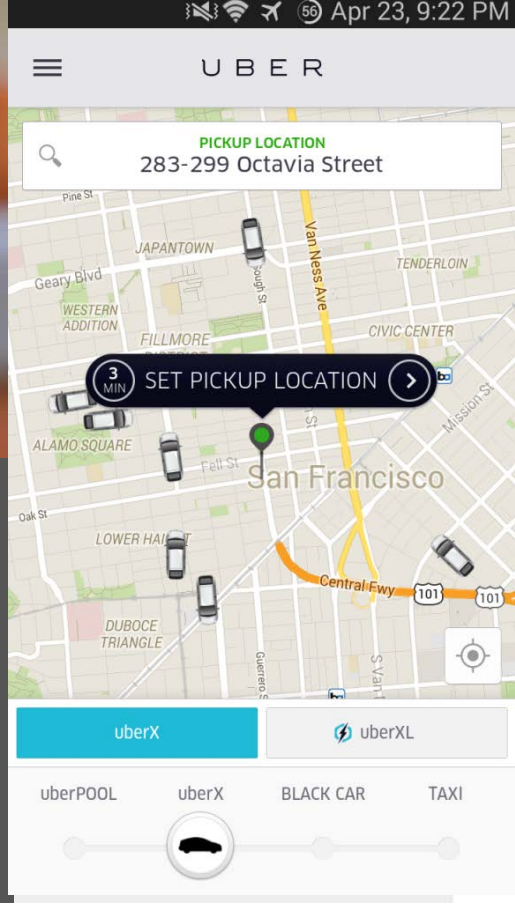
Job	Probability
Recreational therapists	0.003
Dentists	0.004
Athletic trainers	0.007
Clergy	0.008
Chemical engineers	0.02
Editors	0.06
Firefighters	0.17
Actors	0.37
Health technologists	0.40
Economists	0.43
Commercial pilots	0.55
Machinists	0.65
Word processors and typists	0.81
Accountants and auditors	0.94
Telemarketers	0.99

Source: "The Future of Employment: How Susceptible are Jobs to Computerisation?" by C.Frey and M.Osborne (2013)

Accountants and auditors 0.94


What can I help you with?


News Feed
HotelTonight
HotelTonight is now in over 100 destinations! Be spontaneous and book a weekend getaway.
Hotel Tonight
Book Now
AJ Glasser


Apr 23, 9:22 PM
UBER
PICKUP LOCATION
283-299 Octavia Street
3 MIN
SET PICKUP LOCATION
San Francisco
uberX
uberXL
uberPOOL
uberX
BLACK CAR
TAXI


nest
IN 25 MIN
72


amazon


Miele


ABSA

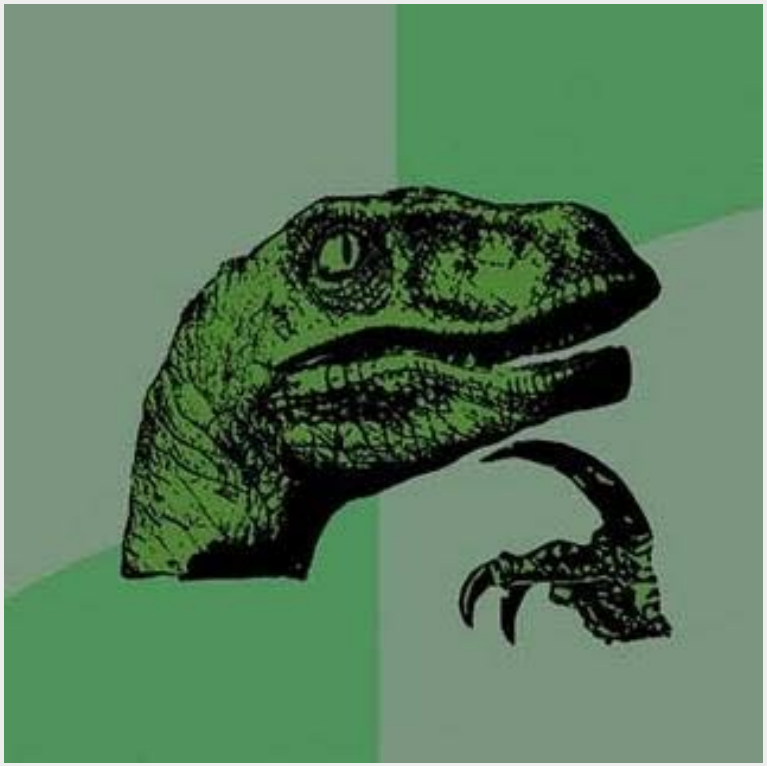
Use of analytics lower than expected

- **Wang, T. and Cuthbertson, R. (2014). Eight Issues on Audit Data Analytics We Would Like Researched. Journal of Information Systems. Vol. 29. No. 1. pp. 155-162.**
- Acceptance and utilization of traditional and nontraditional computer-assisted audit techniques (CAATs) or, more specifically, data analytics for an audit, is lower than expected.
- Reasons:
  - Lack of confidence in own abilities
  - Organizational pressure and technical infrastructure
  - Performance expectancy and facilitating conditions

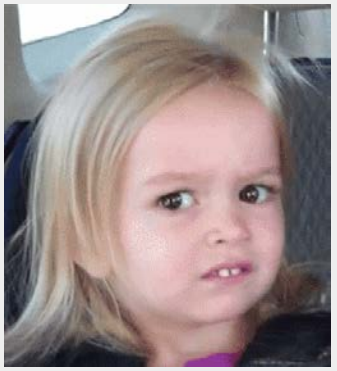
Big data analytics in financial audit

- **Cao, M., Chychyla, R. and Stewart, T. (2015). Big Data Analytics in Financial Statement Audits. Accounting Horizons. Vol. 29. No. 2. pp. 423–429.**
- Big data analytics is the process of inspecting, cleaning, transforming, and modeling big data to discover and communicate useful information and patterns, suggest conclusions, and support decision making.
- Big data has been used for advanced analytics in many domains but hardly, if at all, by auditors.
- Hypothesis: Big data analytics can improve the efficiency and effectiveness of financial statement audits.

Dealing with the paradigm shift



Basic statistics



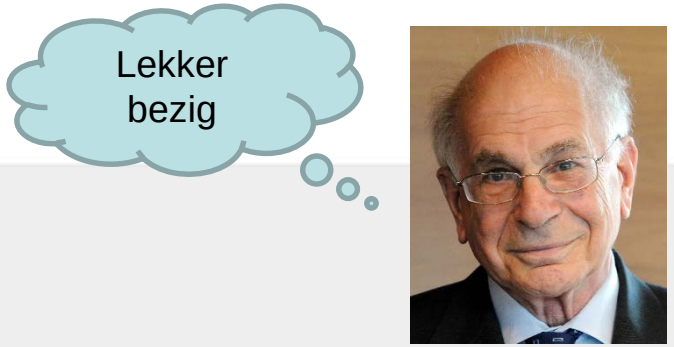
- Deviation = $X - \mu$
  - Difference from mean μ for single value
- Variance $\sigma^2 = E[(X - \mu)^2]$
  - Average of squared deviation for series of values
- Standard deviation $\sigma_X = \sqrt{} E[(X - \mu)^2]$
  - Square root of variance
- Covariance $COV(X,Y) = E[(X - E[X]) * (Y - E[Y])]$
  - Average of products of deviations for x and y
  - Measure of joint variability of x and y
- Correlation $\rho_{X,Y} = COV(X,Y) / SD(X) * SD(Y)$
  - Covariance divided by product of standard deviation for x and y
  - Square root of b from regression analysis

R

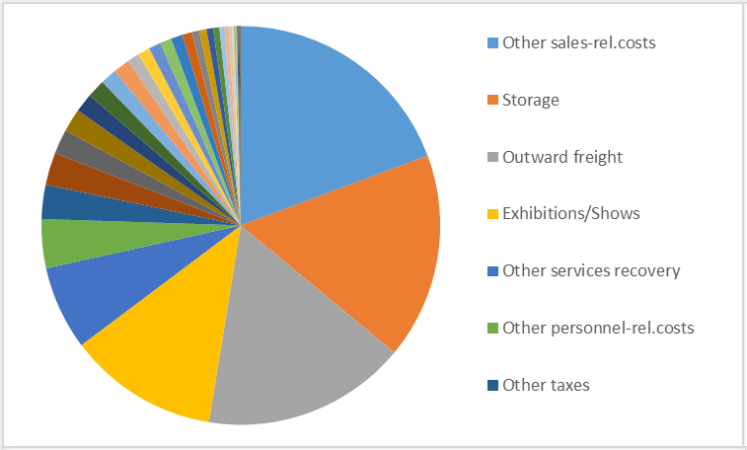
- So many tools. Which one to choose?
- Focus on one language: R
- R most widely used and rising
- Widely used by companies and universities
- Very effective in data manipulation / ETL
- Strong visualisation
- Open source, thousands of packages available
- After a year, learn Python too

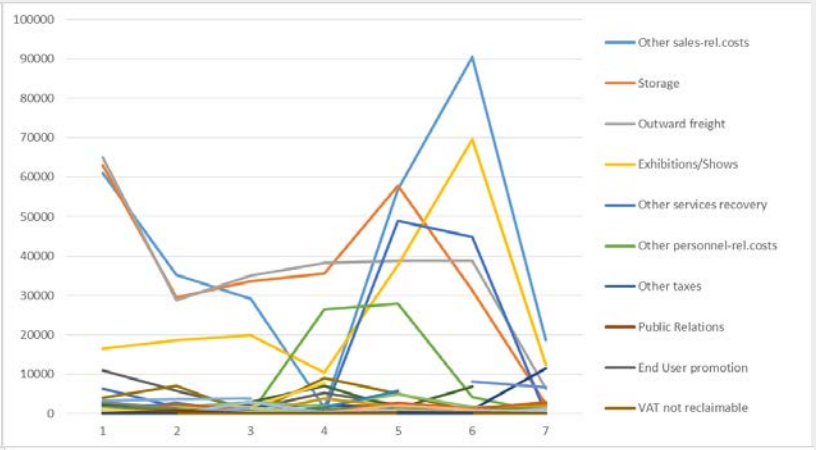
Asking the right question

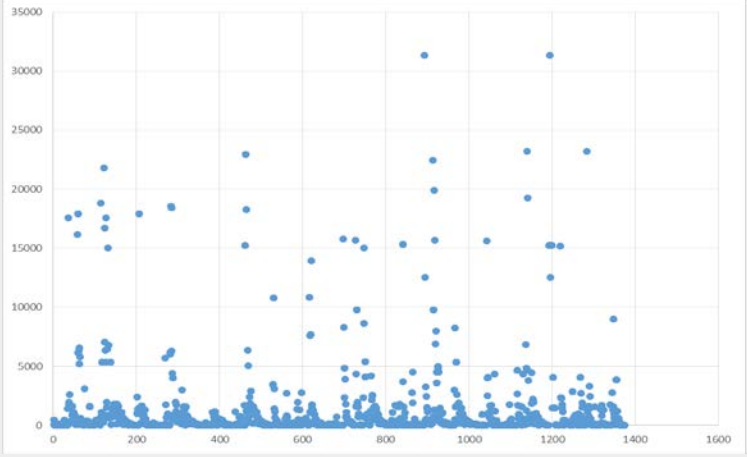
Descriptive	Just describe the data – “how many debtors”
Exploratory	Explore the data to find patterns that support a hypothesis – “can we find a relation between debtor characteristics and risk of default”
Inferential	Test the hypothesis on representative sample – “does our hypothesis hold for a representative sample of debtors across Europe”
Predictive	**Determine predictors** **“Predict risk of default based on debtor location”**
Causal	Find out **why** there is a relation between variables E.g. “Why do debtors from northern countries have a higher risk?”
Mechanistic	Find out the exact mechanism that causes a specific phenomenon

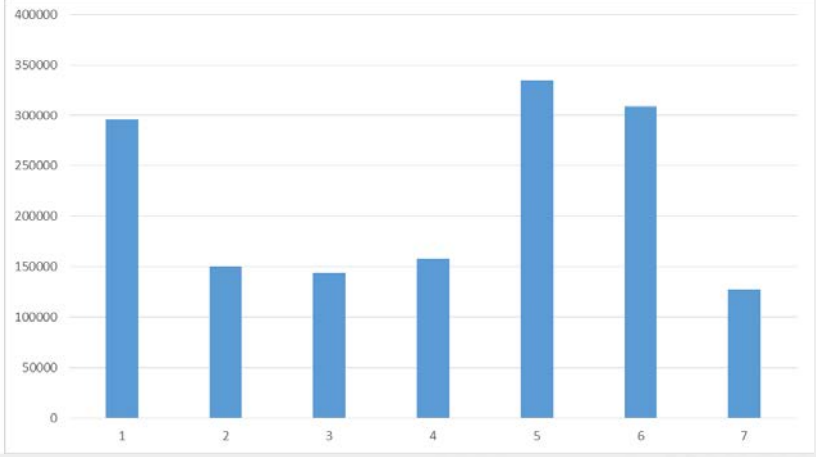


Visualisation

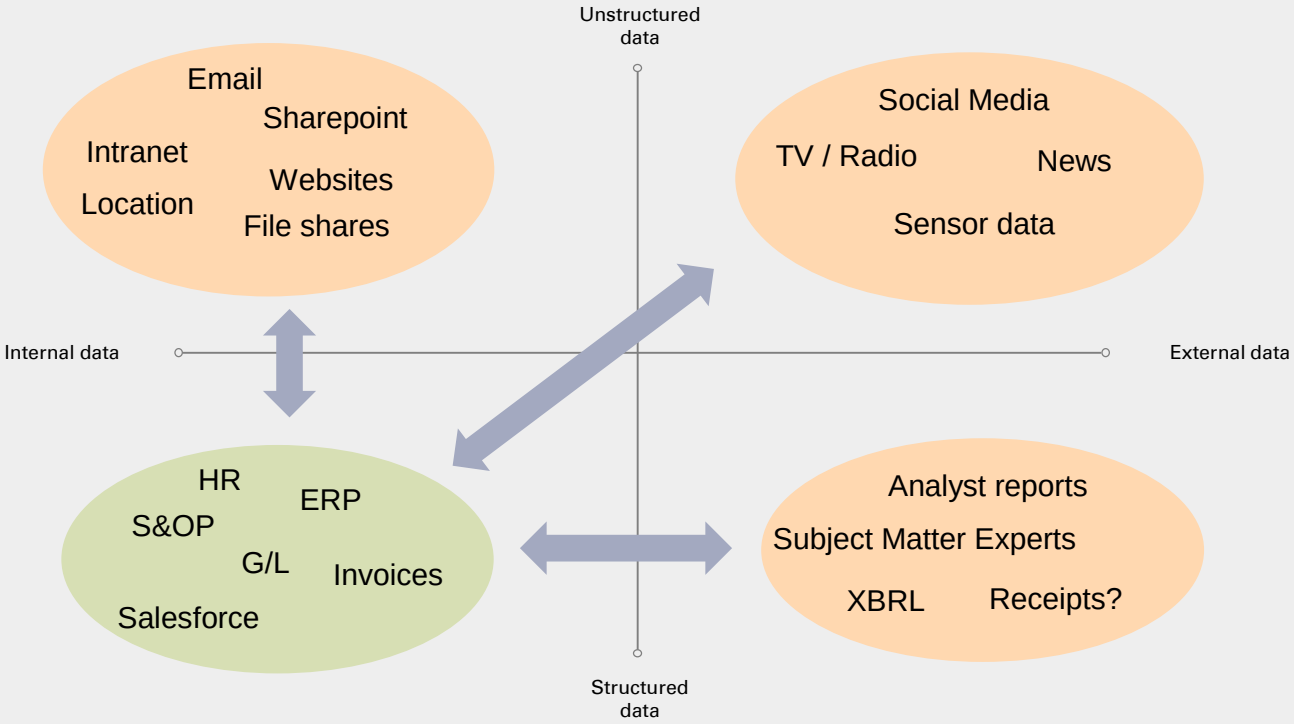






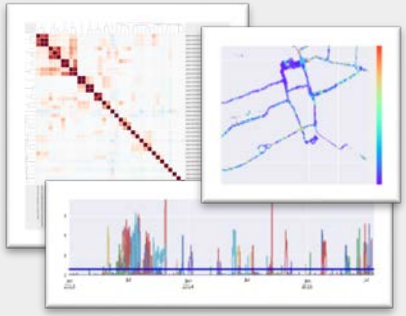


Expanding the data space: Big Data



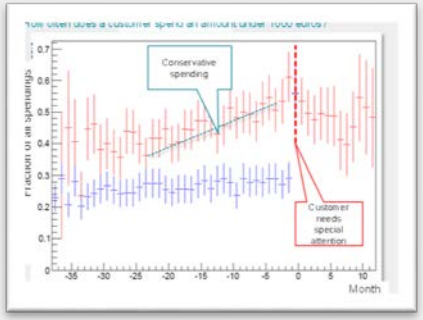
Examples of (big) data analytics

Predicting maintenance

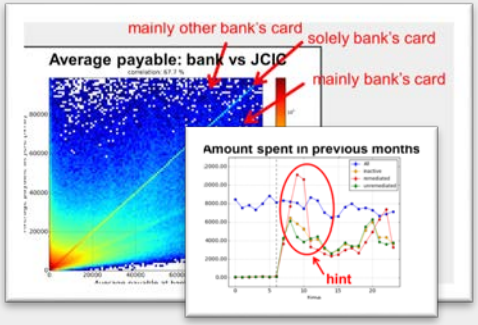


Predicting railroad switch failures

Reducing loan defaults

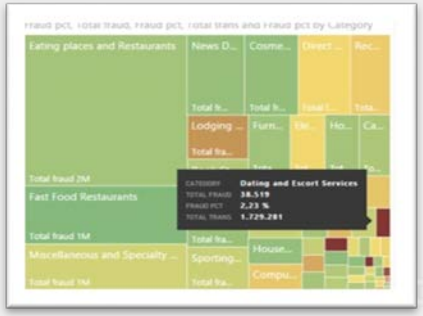


Reducing customer churn



Identifying cross-sell opportunities

Identifying fraud



	Conventional data	Big data
What?	- 'Traditional' data volumes (< 1 TB) - Data structure determined in advance	- Large data volumes (>1TB) - Conventional techniques no longer fit, due to volume, velocity or complexity
Technology	- Data Warehouse - **Excel**, R, SPSS, Python, **SQL**	- Data Lake - Hadoop, MapReduce, Python, Spark
Profiles	- Data engineers, architects - DBA, BI-specialists	- Data engineers, architects - Data scientists
Examples	- SAP entries on projects - Reports - Data about storage goods	- Transaction data - Sensor data - Public transport tap-in tap-out data - Internet of Things

Machine learning

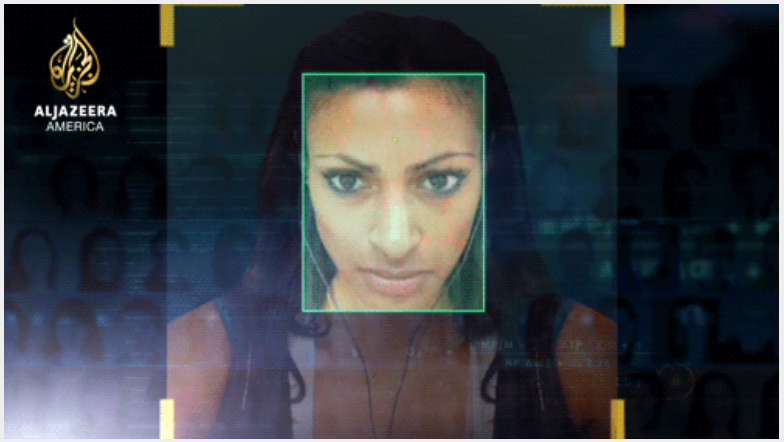
- Arthur Samuel (1959): Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed

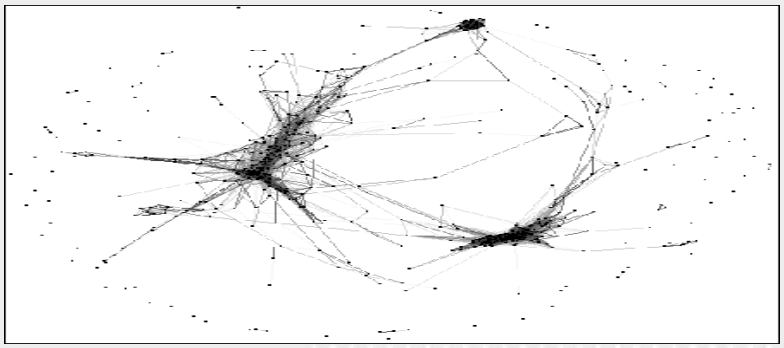
- Main types of machine learning:
  - Supervised – “train” algorithm with question/answer pairs
  - Unsupervised – algorithm finds its own way
  - Reinforcement – algorithm gets feedback as it navigates through problem space

- Machine learning: solving known problems
- Data mining: discovering unknown patterns

Machine learning techniques

- Regression
- Decision trees
- Neural networks
- Bayesian networks
- Nearest neighbors
- Cluster analysis
- Anomaly detection
- Deep learning





Nice vids

IBM Watson:
https://www.youtube.com/watch?v=_Xcmh1LQB9I

Google Machine Learning:
https://www.youtube.com/watch?v=l95h4alXfAA
https://www.youtube.com/watch?v=_rdINNHLYaQ

Call for action

How can we deal with the paradigm shift?

Dust off statistics

Master the tools of the trade

Become a leader in data munging and wrangling

Invest in yourself!