Data driven Leadership: Harnessing the Power of Analytics

Introduction

Does Will Smith know more than what Jim Carrey or any other knows?

- Year 2007 and 2008 were terrible years https://www.boxofficemojo.com/year/?sort=year&sortDir=asc&grossesOption=totalGrosses&ref_=bo_yl_resort#table
- **"I am legend"** earned $77Mn
- In 2008, $625Mn earning with poor and critical review to **"Hancock"**

Target Figured Out A Teen Girl Was Pregnant Before Her Father Did!

An angry man went into a Target outside of Minneapolis, demanding to talk to a manager:

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#190024096668

How Target did it?

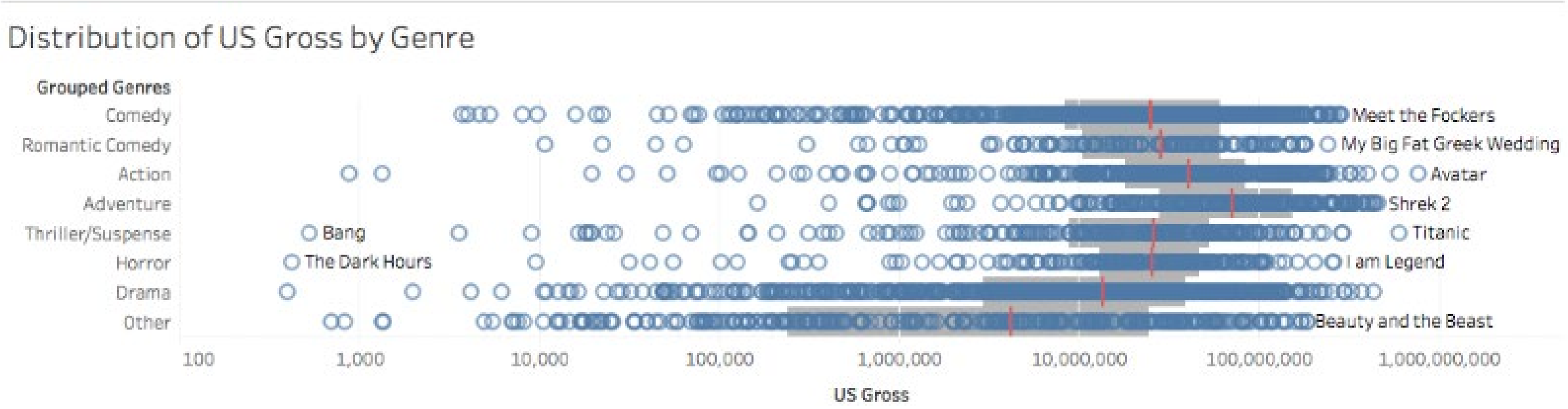
- Target assigns every customer a Guest ID number, tied to their credit card, name, or email address that becomes a bucket that stores a history of everything they've bought and any demographic information
- Historical buying data for all the ladies who had signed up for Target baby registries in the past.
- Ran test after test, analyzing the data, and before long some useful patterns emerged.
- Women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester.
- In the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium and zinc.
- Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signals they could be getting close to their delivery date.

Decisions

PREDICTION

OPTIMALITY

Data Science: finding useful pattern in data



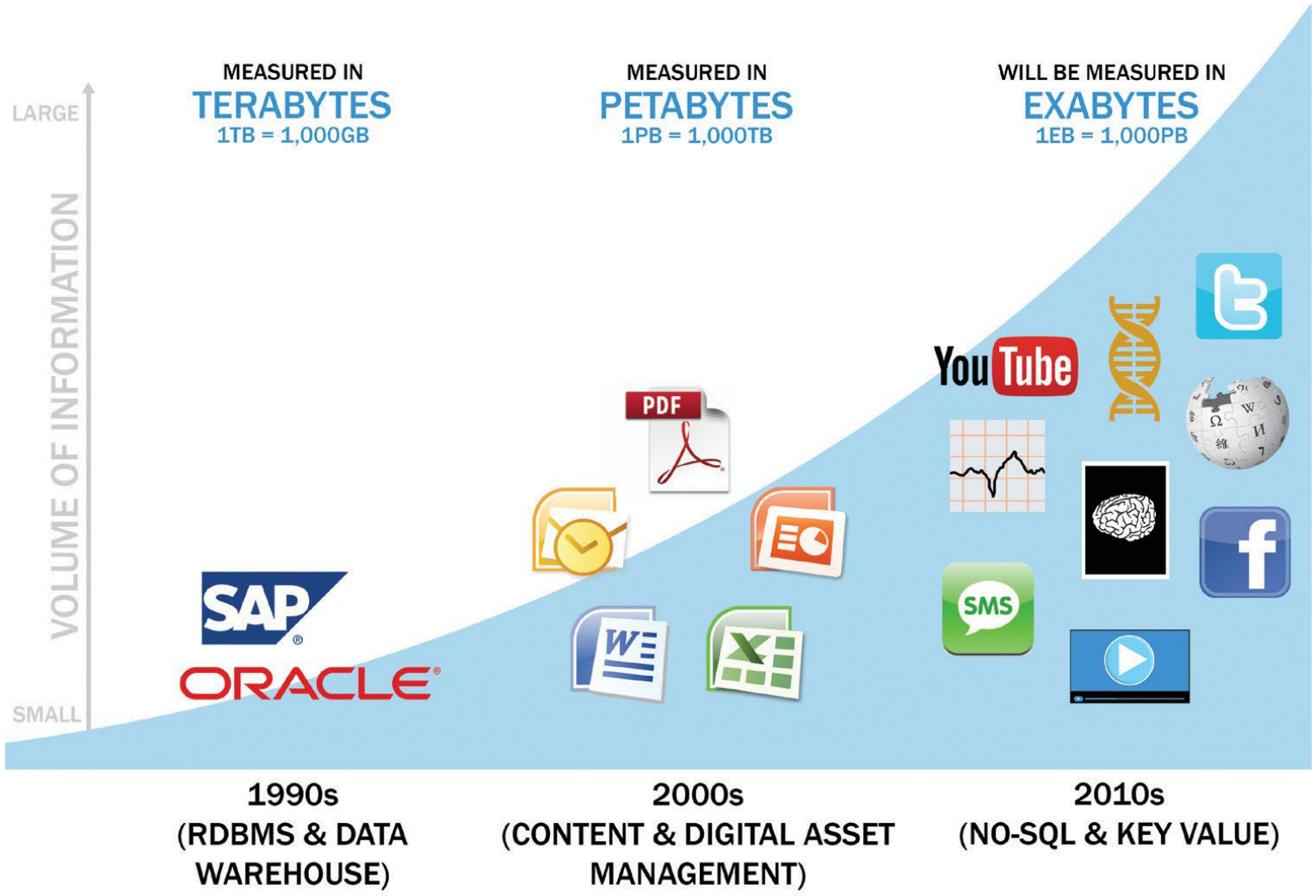
Data and Big Data

source: Data Science & Big Data Analytics, Wiley

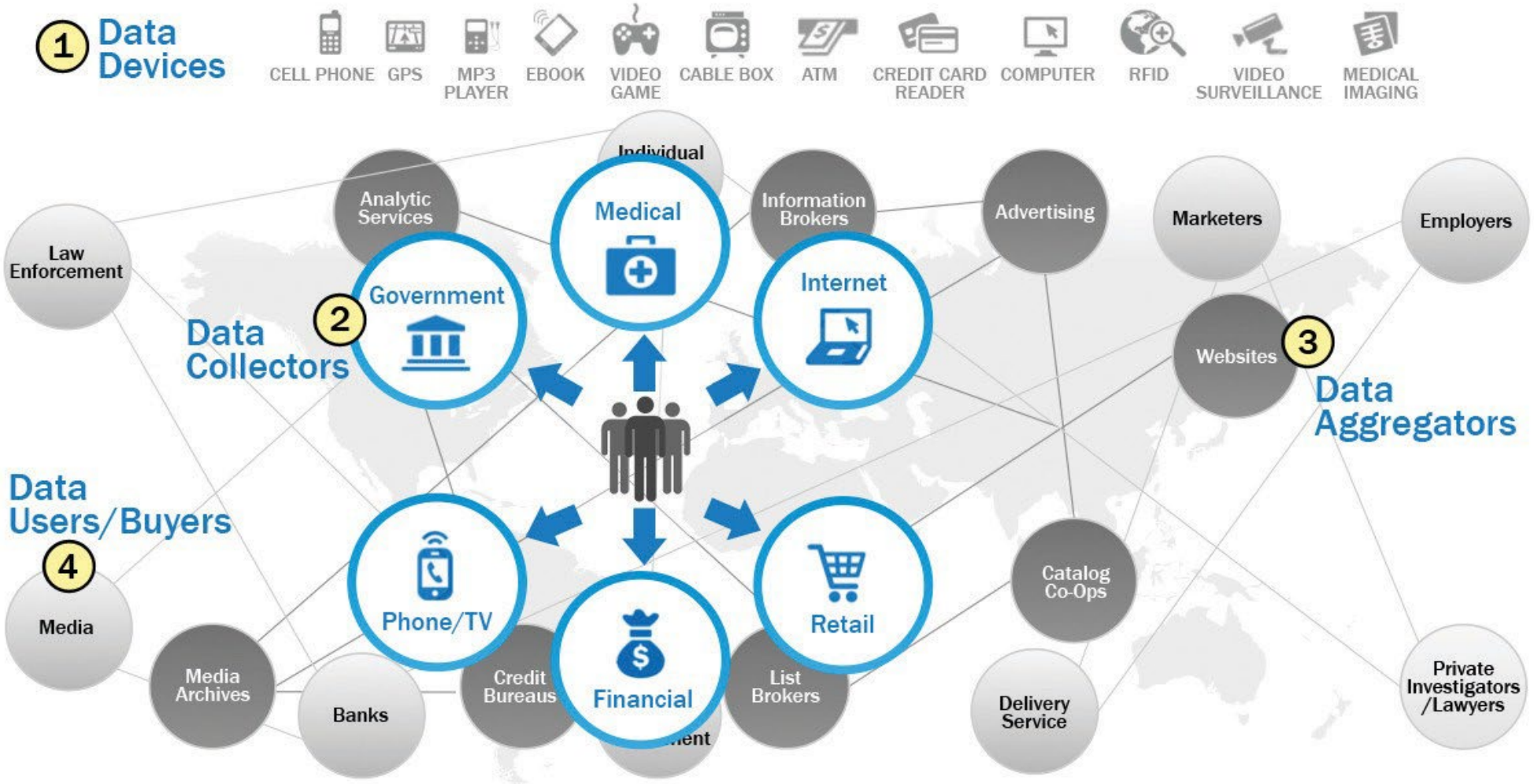
What's Driving Data Deluge?

source: Data Science & Big Data Analytics, Wiley

Drivers of Data

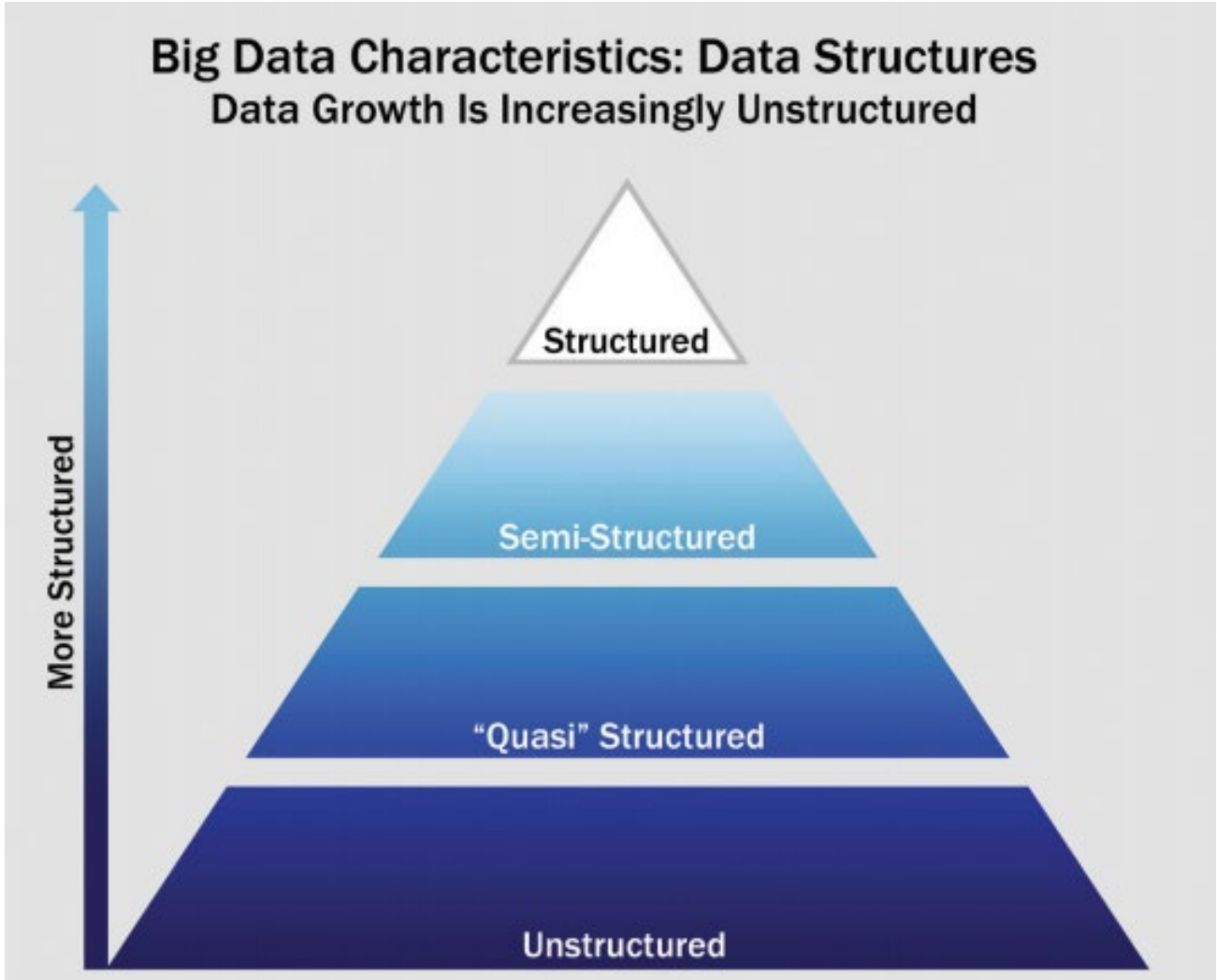


Data eco-system



source: Data Science & Big Data Analytics, Wiley

source: Data Science & Big Data Analytics, Wiley

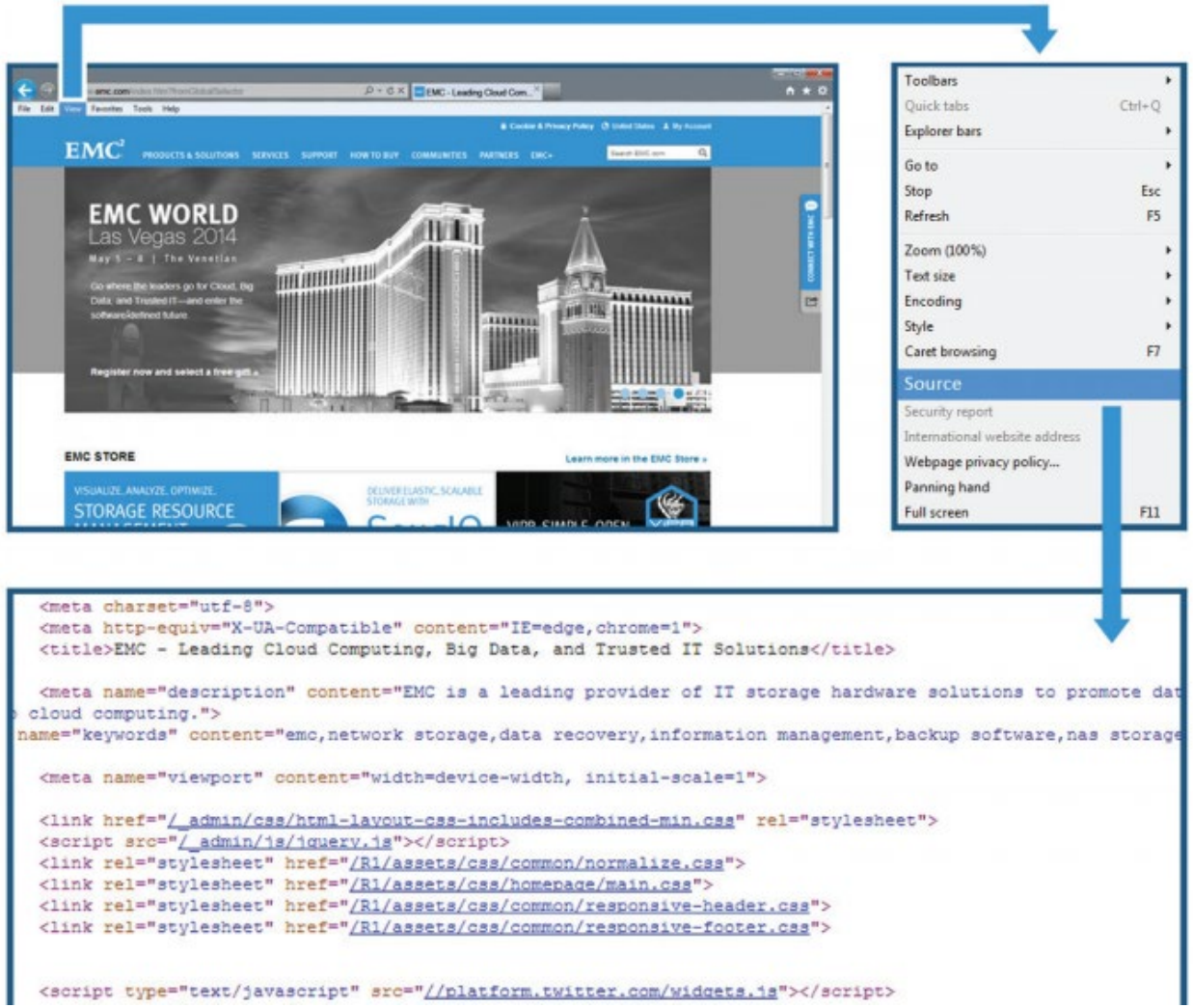


source: Data Science & Big Data Analytics, Wiley

Structured data

SUMMER FOOD SERVICE PROGRAM 1]				
(Data as of August 01, 2011)				
Fiscal Year	Number of Sites	Peak (July) Participation	Meals Served	Total Federal Expenditures 2]
	----------Thousands----------		--Mil.--	---Million $---
1969	1.2	99	2.2	0.3
1970	1.9	227	8.2	1.8
1971	3.2	569	29.0	8.2
1972	6.5	1,080	73.5	21.9
1973	11.2	1,437	65.4	26.6
1974	10.6	1,403	63.6	33.6
1975	12.0	1,785	84.3	50.3
1976	16.0	2,453	104.8	73.4
TQ 3]	22.4	3,455	198.0	88.9
1977	23.7	2,791	170.4	114.4
1978	22.4	2,333	120.3	100.3
1979	23.0	2,126	121.8	108.6
1980	21.6	1,922	108.2	110.1
1981	20.6	1,726	90.3	105.9
1982	14.4	1,397	68.2	87.1
1983	14.9	1,401	71.3	93.4
1984	15.1	1,422	73.8	96.2
1985	16.0	1,462	77.2	111.5
1986	16.1	1,509	77.1	114.7
1987	16.9	1,560	79.9	129.3
1988	17.2	1,577	80.3	133.3
1989	18.5	1,652	86.0	143.8
1990	19.2	1,692	91.2	163.3

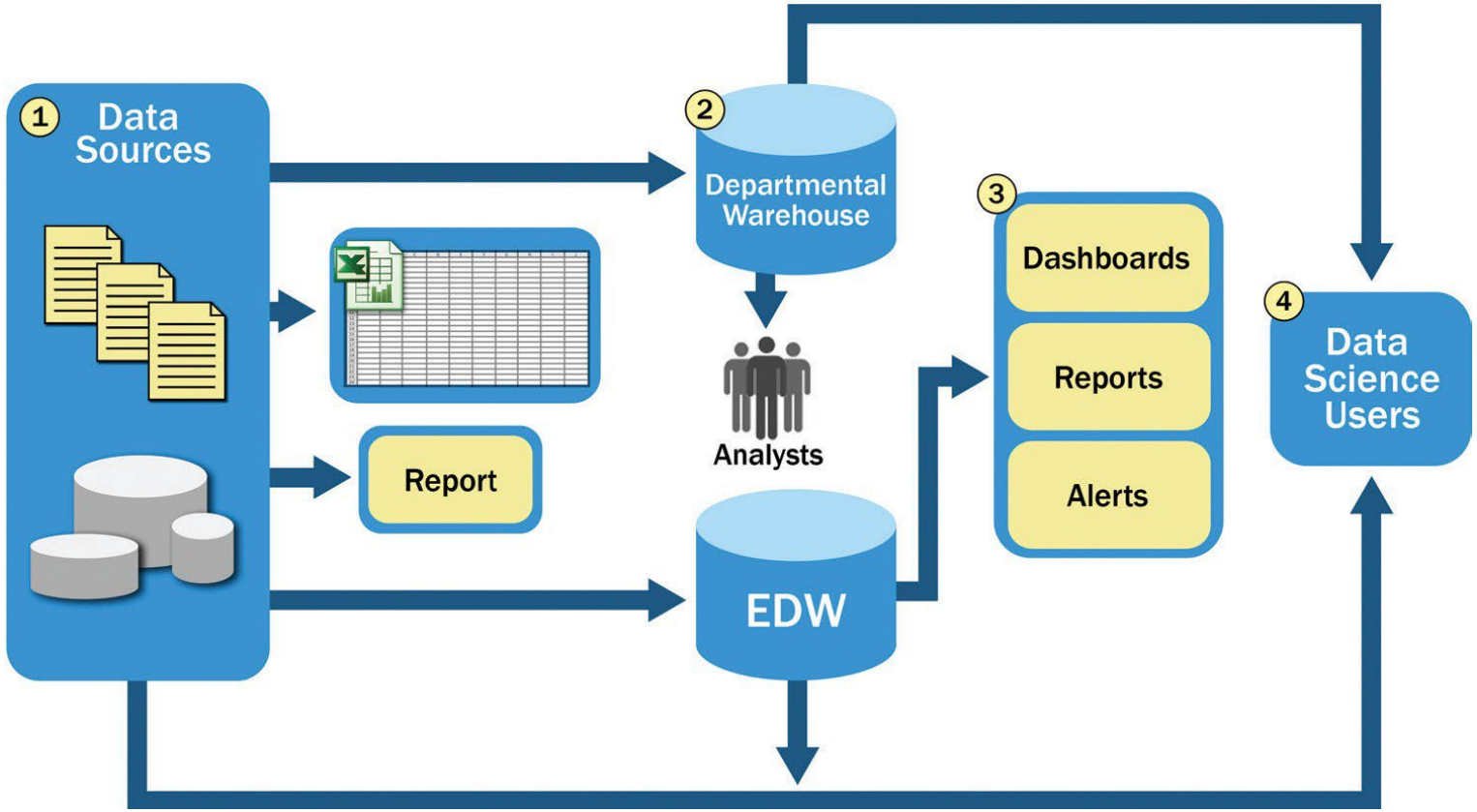
Semi-Structured data



source: Data Science & Big Data Analytics, Wiley

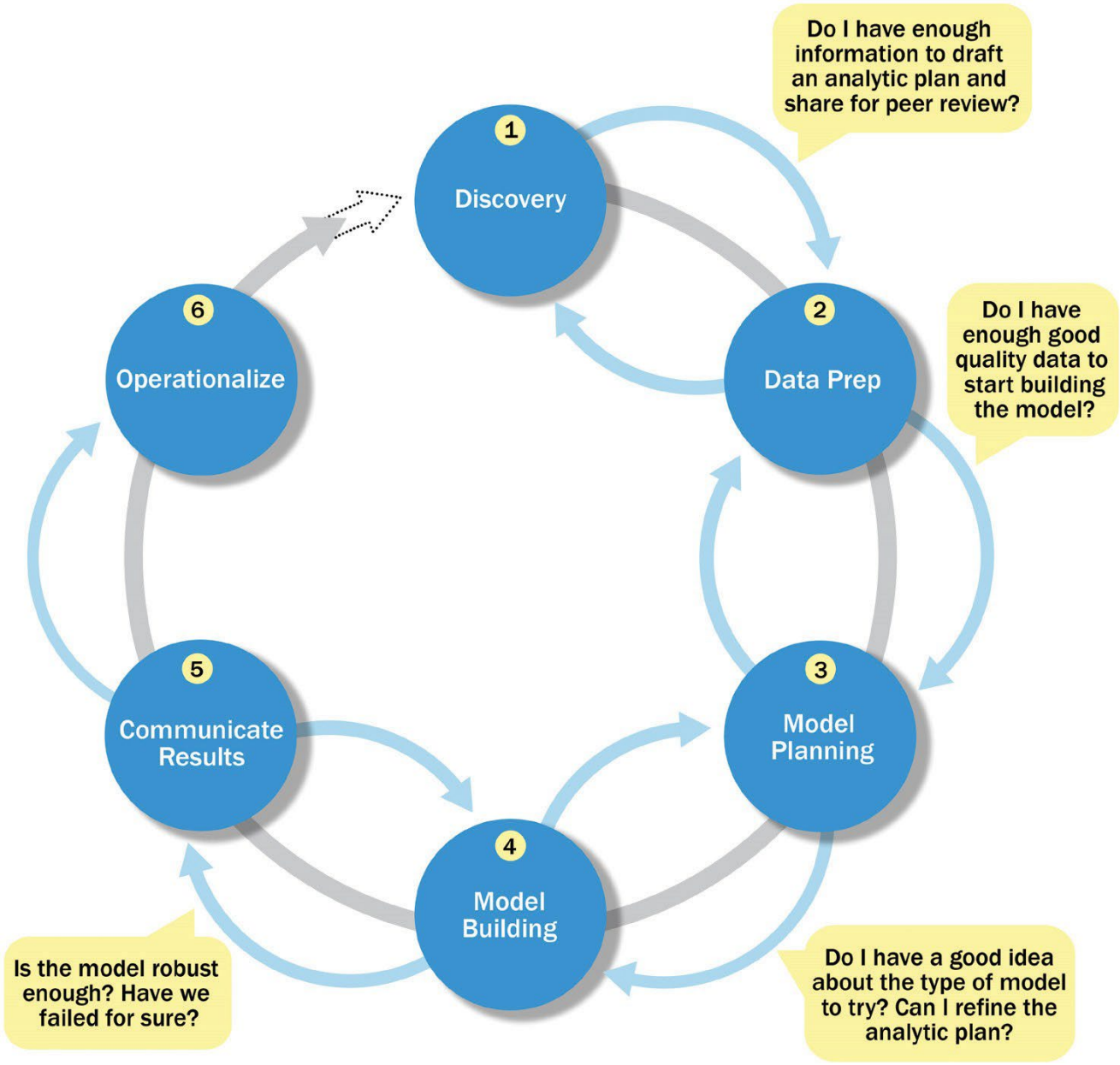
source: Data Science & Big Data Analytics, Wiley

Analytical Structure



source: Data Science & Big Data Analytics, Wiley

Data Analytics Lifecycle



Analytics

- Business Analytics
  - Wayne Winston: “using data for better decision making.”
- Major fields:
  1. Descriptive analytics
  2. Predictive analytics
  3. Prescriptive analytics

Examples of Descriptive Analytics in Business

- **Sales and revenue analysis:** to see what months or days had the highest sales and adjust your **marketing strategy** accordingly.
- **Customer behaviour analysis:** Descriptive analytics can give insight into customer behaviour, such as which products they buy the most, how frequently they purchase, and which promotions they respond to best.
- **Market share analysis:** You can see how your brand stacks up against your competitors by analysing market share data.
- **Inventory Analysis:** to track inventory levels, identify trends in demand, and optimise their supply chain.

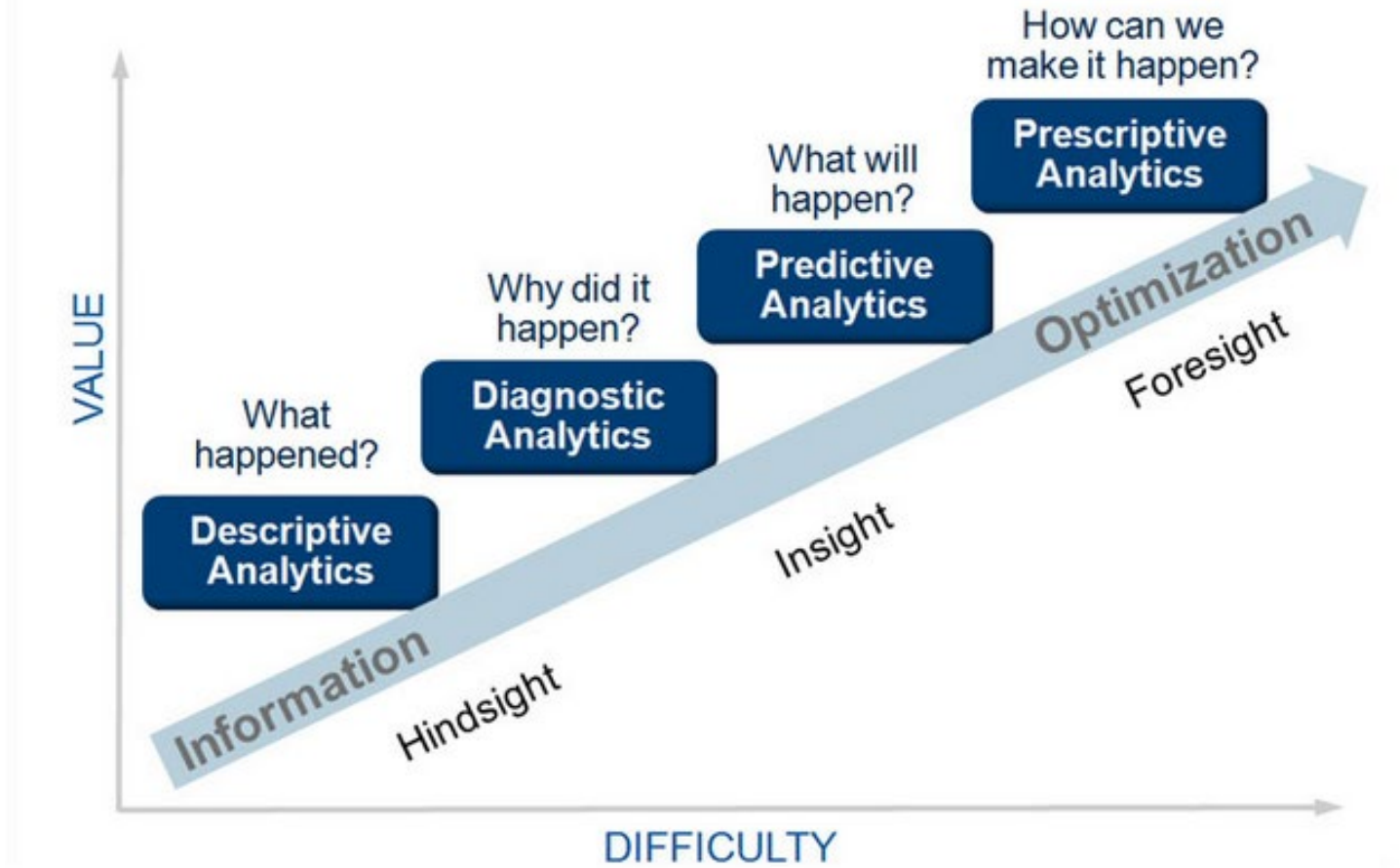
Examples of Predictive Analytics in Action

- **Retail Industry:** Analyse customer data and **forecast future sales**. This information is used to optimize inventory levels and improve supply chain management.
- **Banking and Financial Services:** To **detect fraud**, **assess credit risk**, and identify **potential investment opportunities**.
- **Healthcare Industry:** To forecast **future demand** for medical services, identify **at-risk patients**, and improve patient outcomes through **personalized care plans**.
- **Manufacturing:** To predict when equipment is **likely to fail**, allowing them to schedule maintenance and **prevent unplanned downtime**.

Business Applications of Prescriptive Analytics

- **Supply Chain Optimization:** helps companies make informed decisions on inventory levels, production schedules, and transportation routes.
- **Fraud Detection:** By analysing historical data, prescriptive analytics can help financial institutions identify patterns and anomalies in transactions that may indicate fraud.
- **Customer Segmentation:** Prescriptive analytics helps companies better understand their customers by segmenting them based on demographic, geographic, or behavioural characteristics.

Analytic Value Escalator



Use Cases for Cognitive Analytics

- **Healthcare:** Cognitive analytics can analyse medical images, such as x-rays or MRIs, to help diagnose diseases or identify potential health risks.
- **Customer service:** customer feedback and sentiment, such as comments on social media or support tickets, to identify trends and patterns that can help improve customer experiences.
- **Retail:** analyses customer purchasing patterns, such as the items they buy and when they buy them, to help retailers optimise their inventory and improve the customer experience.
- **Market Research:** Cognitive analytics can analyse consumer data to gain insights into consumer behaviour and preferences.

Challenges with Business Analytics

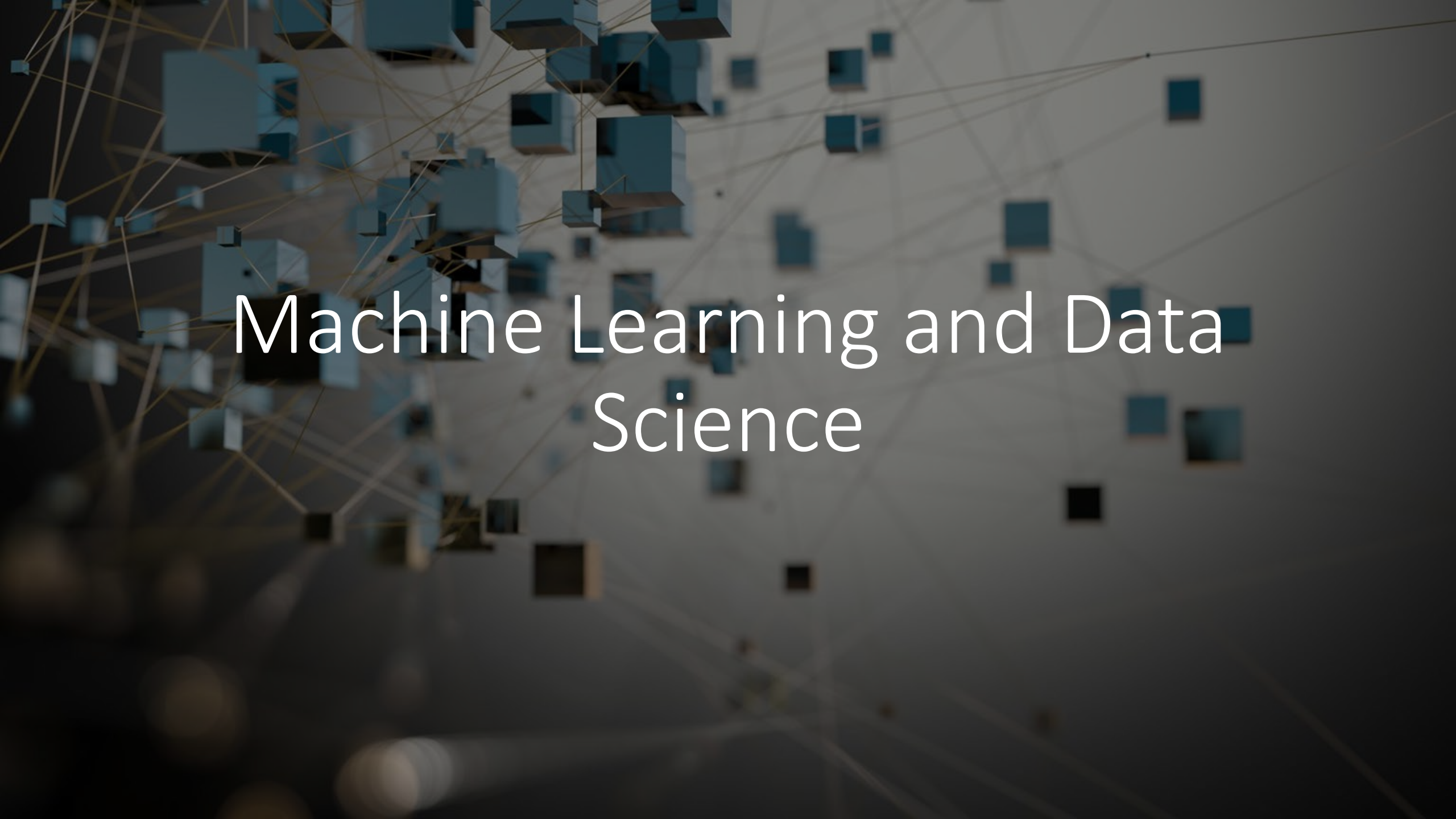
- Lack of Management Science Experts
  - Spreadsheet modeling
    - Simple formulation
    - Seek practical solutions
    - But limited in the amount of data they can store
- Analytics Bring Change in the Decision-Making Process
  - Information based decision can upset traditional power relationship

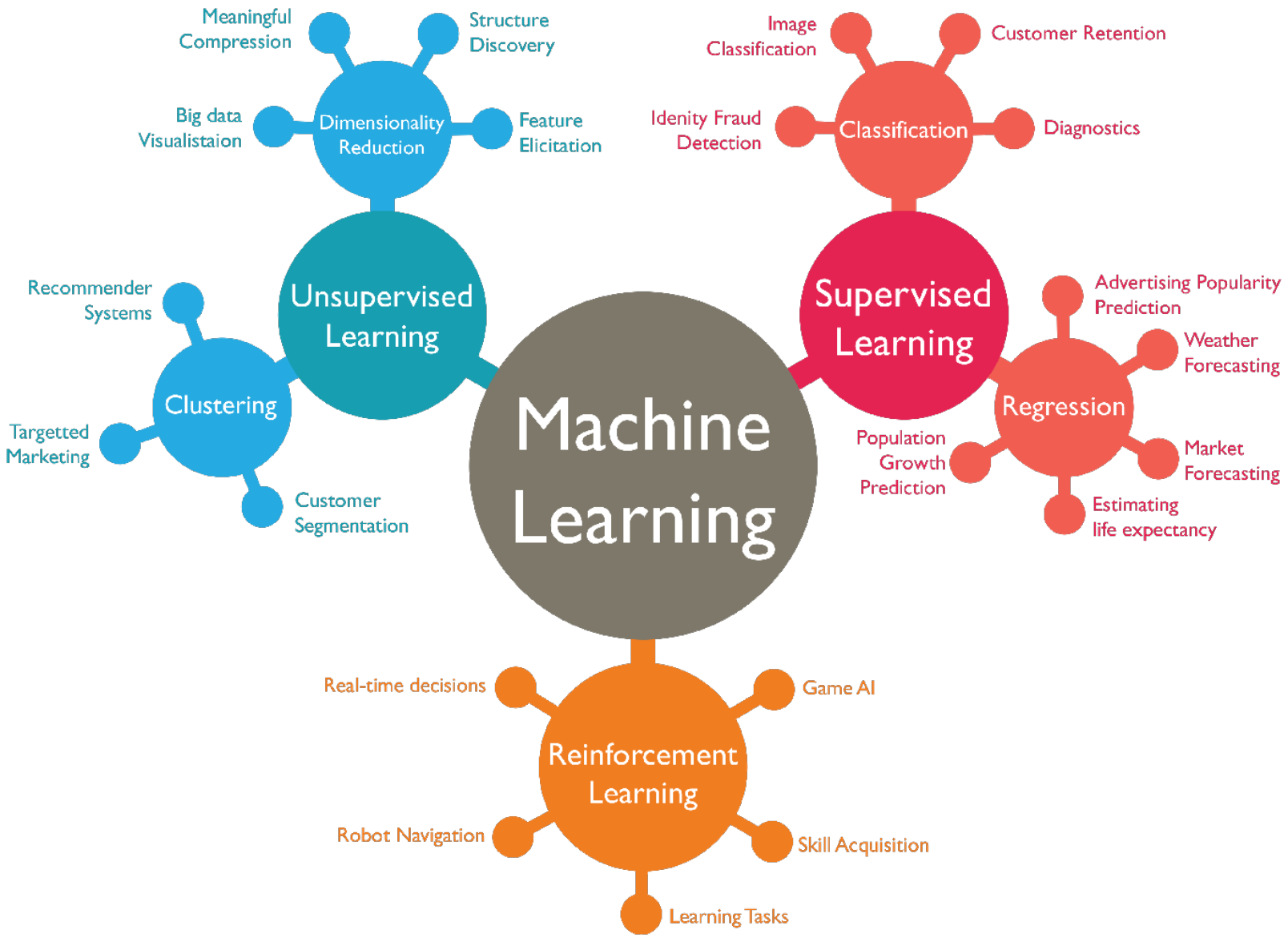
Business Driver

Optimize business operations	**Sales, pricing, profitability, efficiency**
Identify business risk	**Customer churn, fraud, default**
Predict new business opportunities	**Upsell, cross-sell, best new customer prospects**
Comply with laws or regulatory requirements	**Anti-Money Laundering, Fair Lending**

(Prediction and Recommendation)

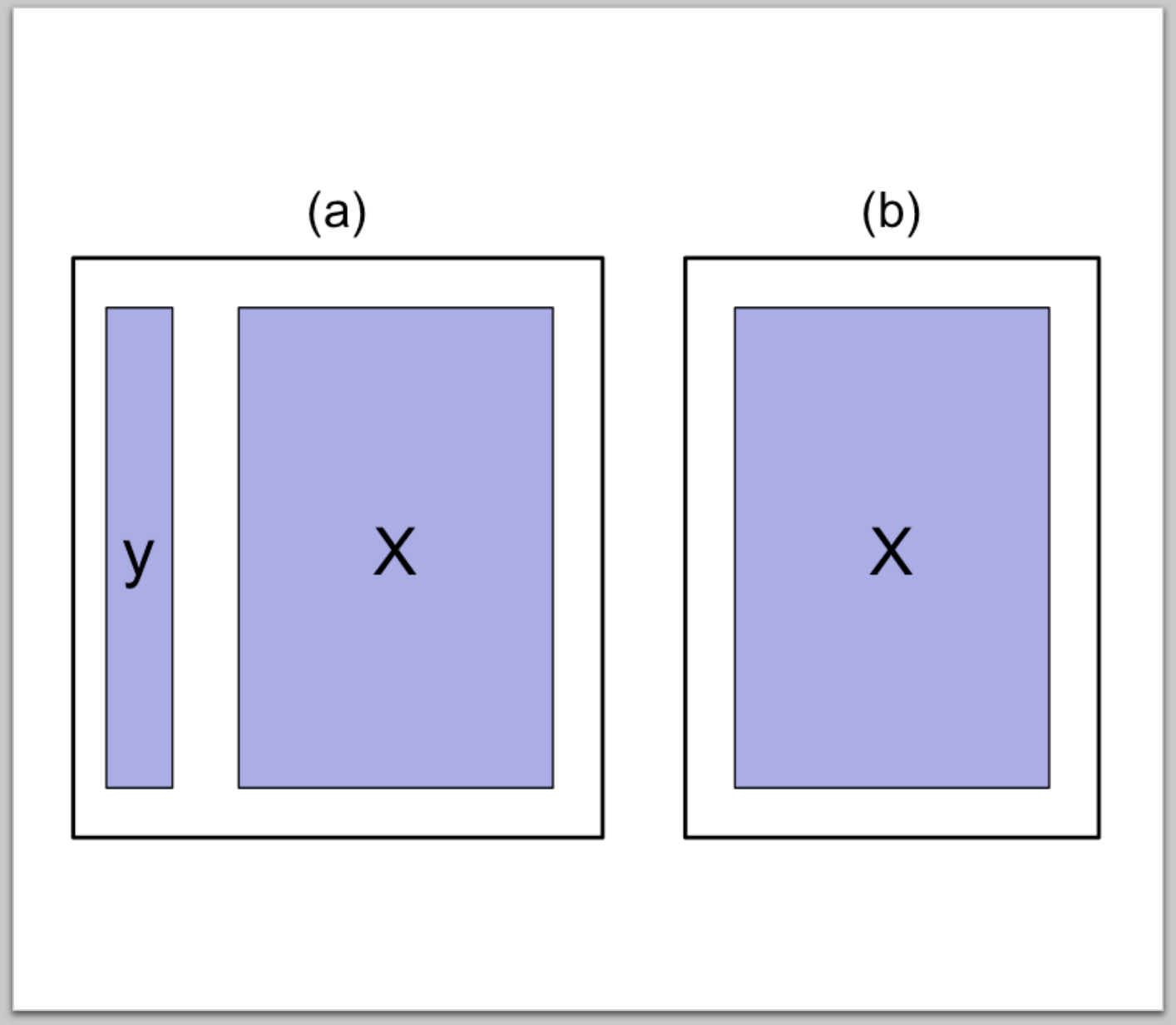
source: Data Science & Big Data Analytics, Wiley


Machine Learning and Data Science


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Supervised Learning vs Unsupervised Learning

- Supervised Learning: both X and Y are known
- Unsupervised Learning: only X



Supervised Learning: Techniques

- Linear Regression
- Logistic Regression
- Decision Trees
- Artificial Neural Network
- Support Vector Machines

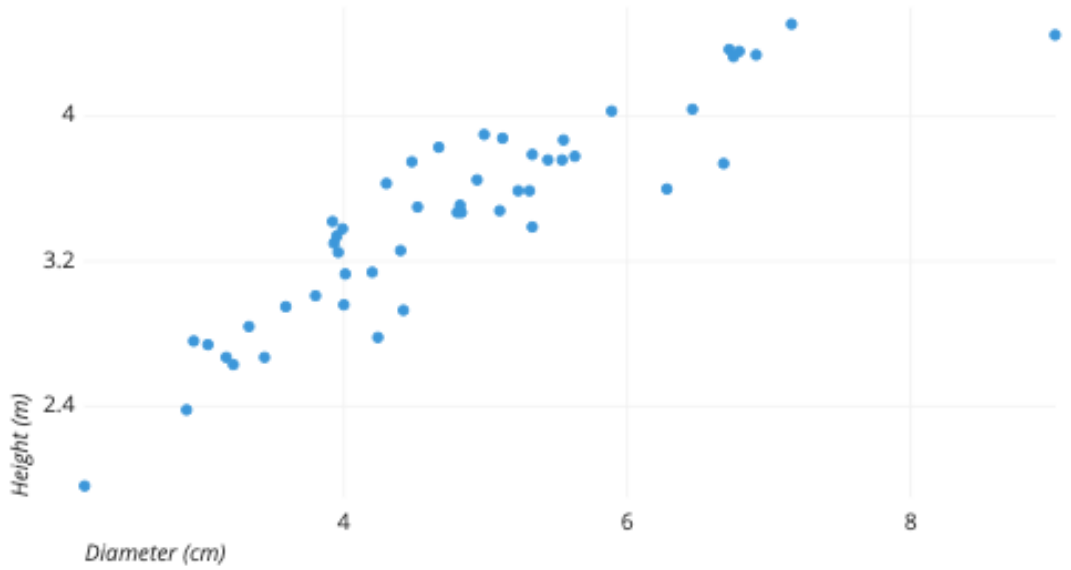
Unsupervised Learning: Techniques

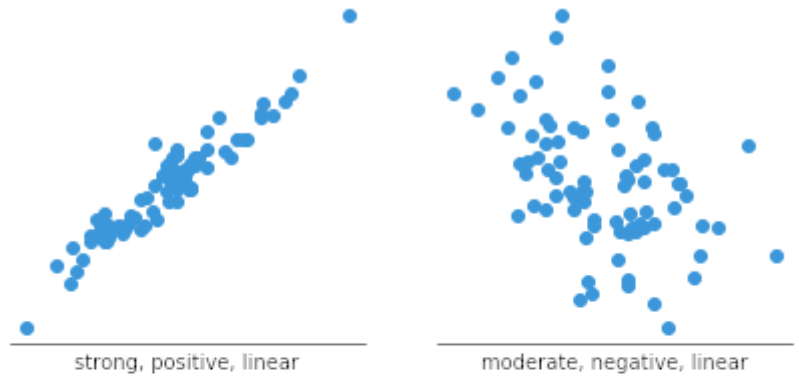
1. Principal Component Analysis: a tool used for data visualization or data pre-processing before supervised techniques are applied.
2. Clustering: a broad class of methods for discovering unknown subgroups in data.

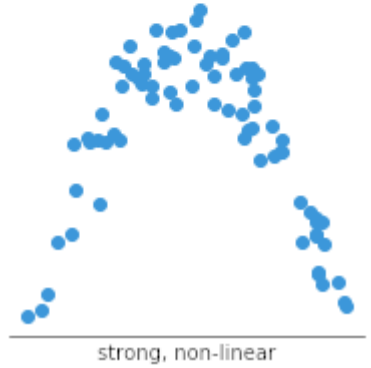
Tools: R, Python and KNIME

Visualization of data

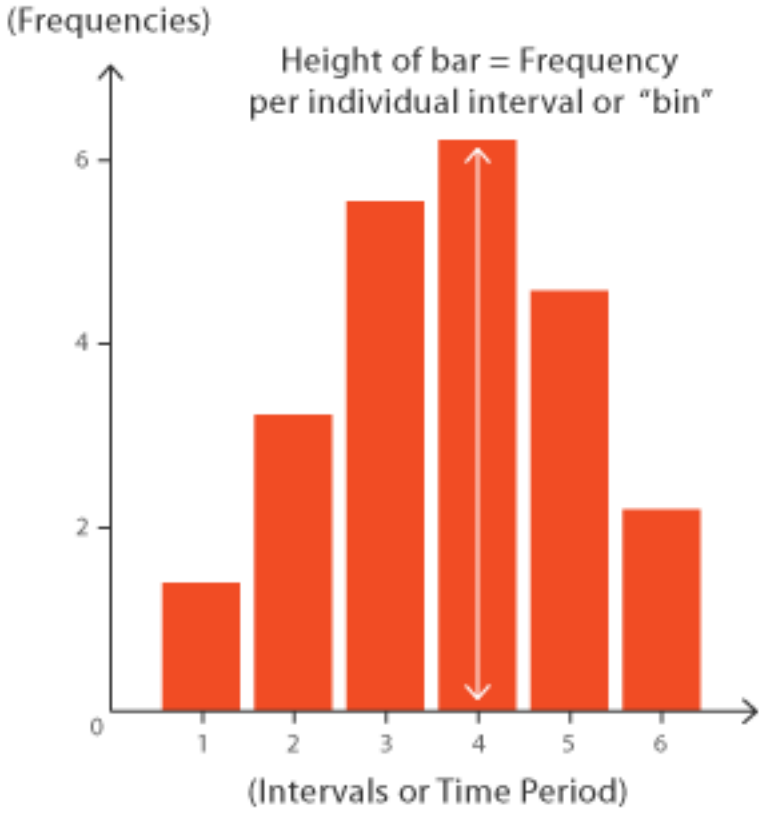
Scatter plot

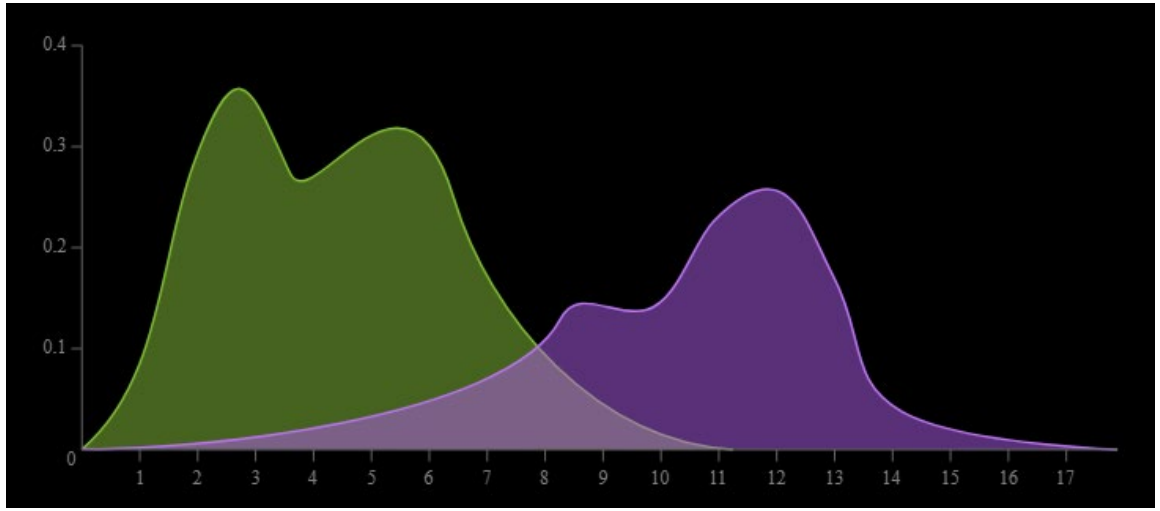


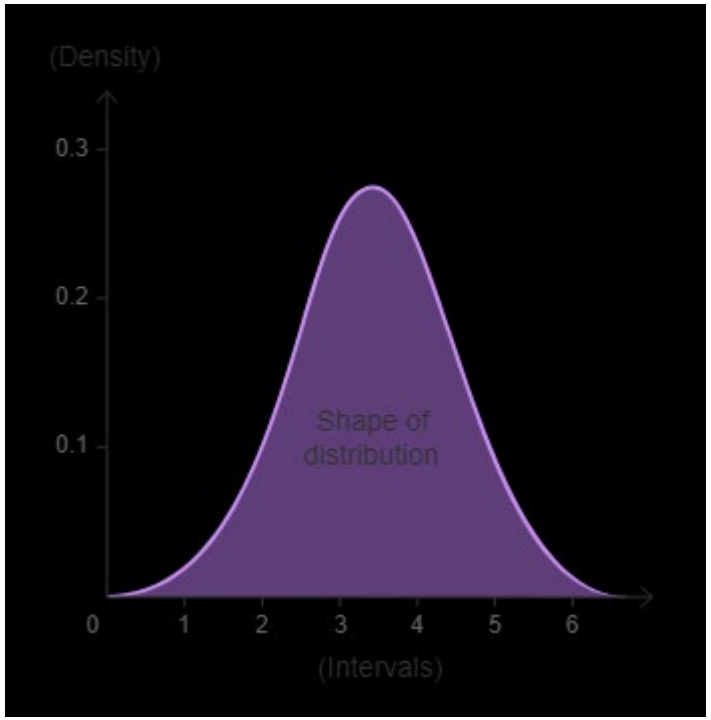




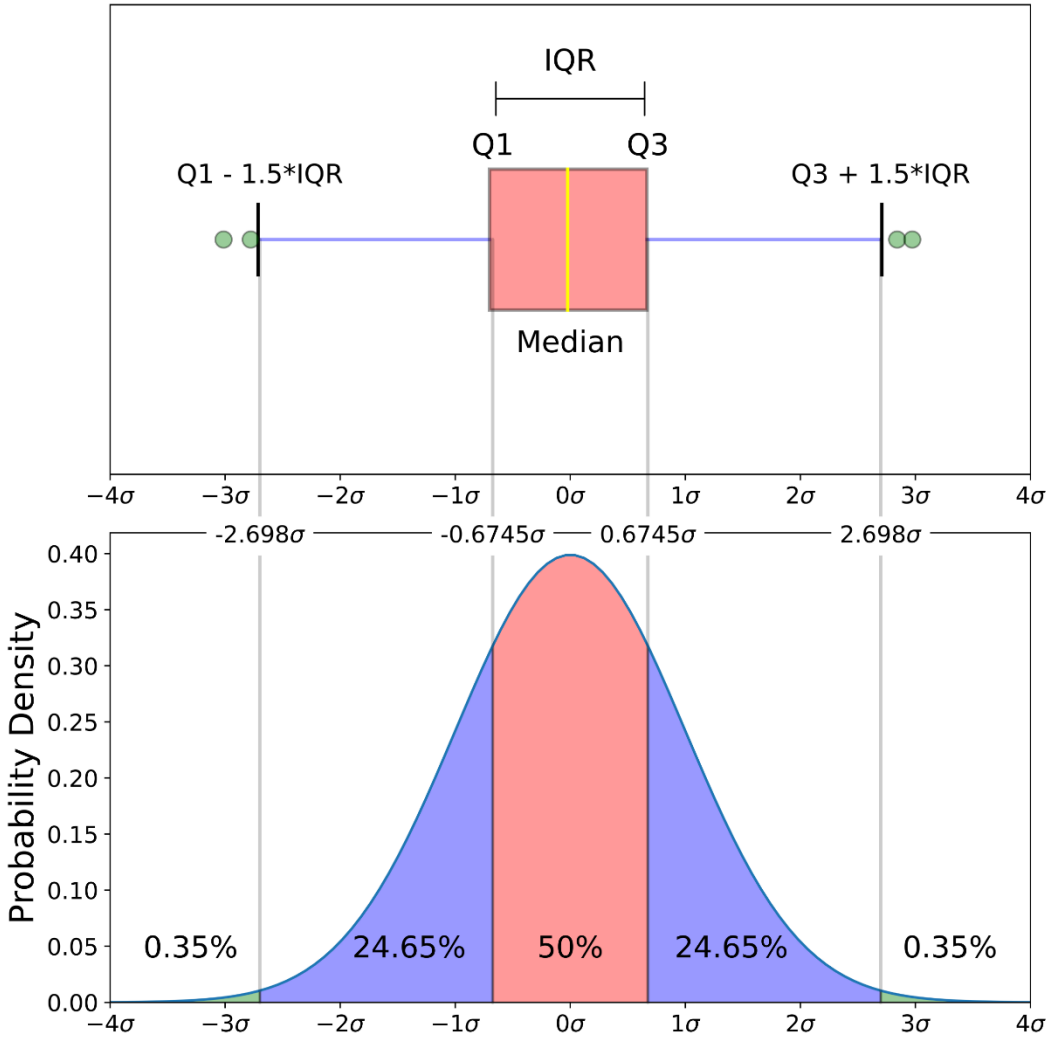
histogram


(Frequencies)
Height of bar = Frequency
per individual interval or "bin"
6
4
2
0
1
2
3
4
5
6
(Intervals or Time Period)

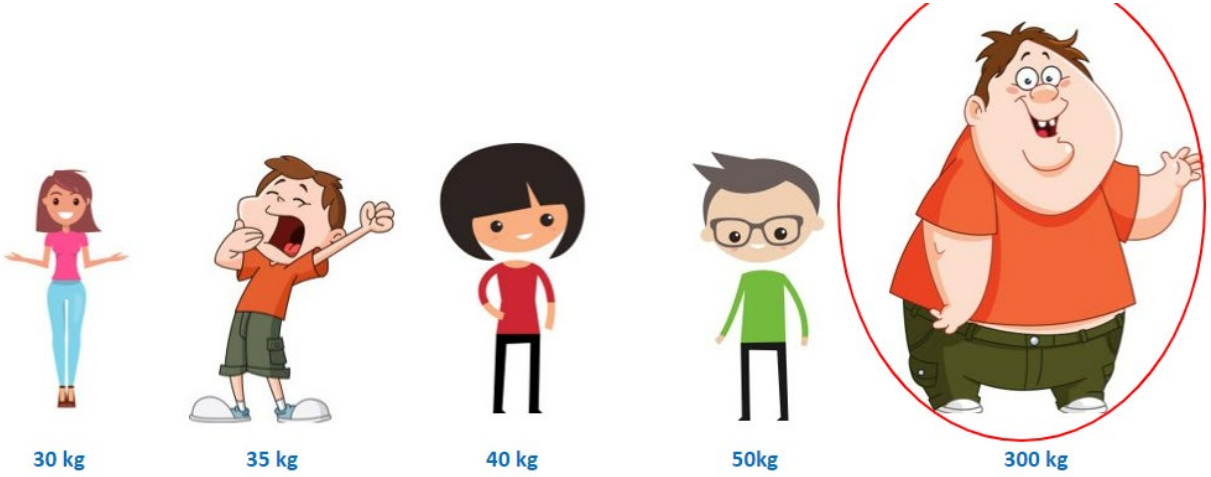

0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17


(Density)
0.3
0.2
0.1
Shape of
distribution
0
1
2
3
4
5
6
(Intervals)

Box Plot: Measure of central tendency

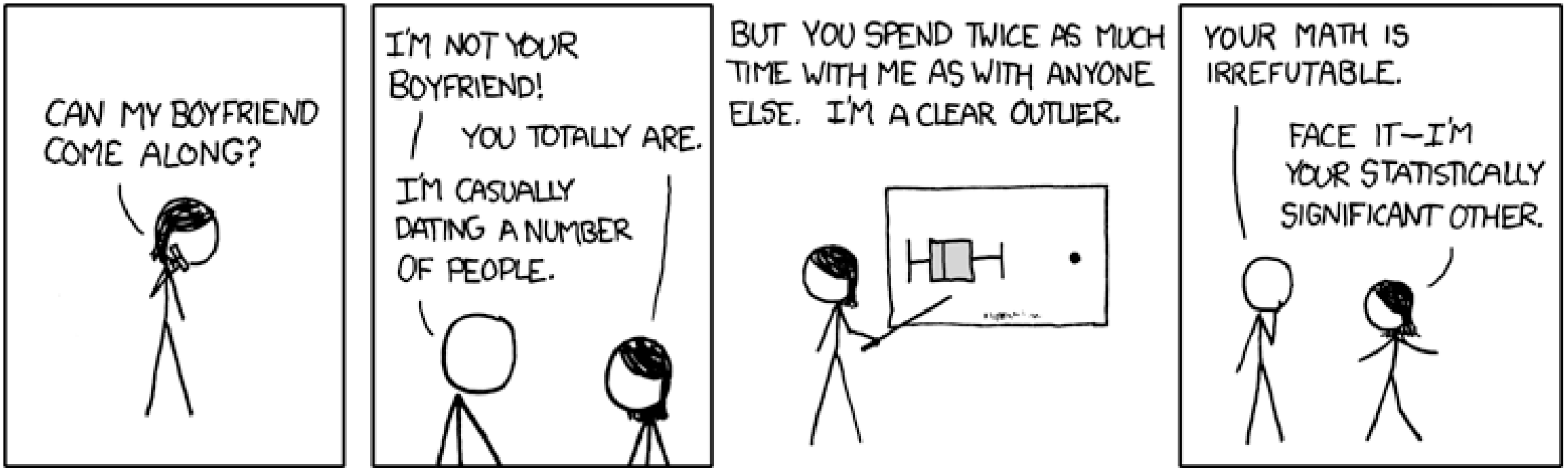


Outlier detection



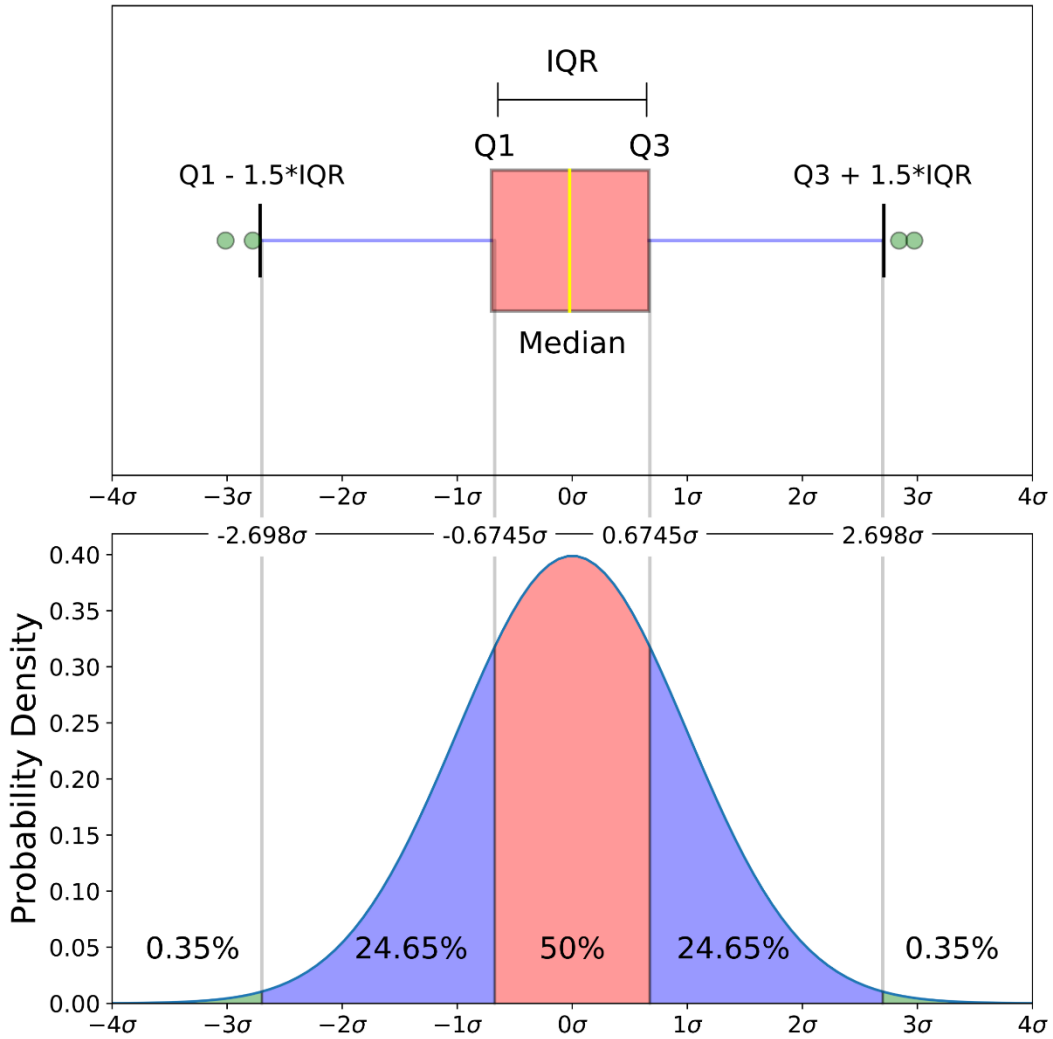
Outliers in a collection of data are the **values which are far away from most other points**.
A boxplot is usually used to visualize a dataset for spotting unusual data points. However, is an outlier abnormal or normal? It needs to be decided by data analysts.
The boxplot displays five descriptive values which
are minimum, Q1, median, Q3 and maximum.

Analytics

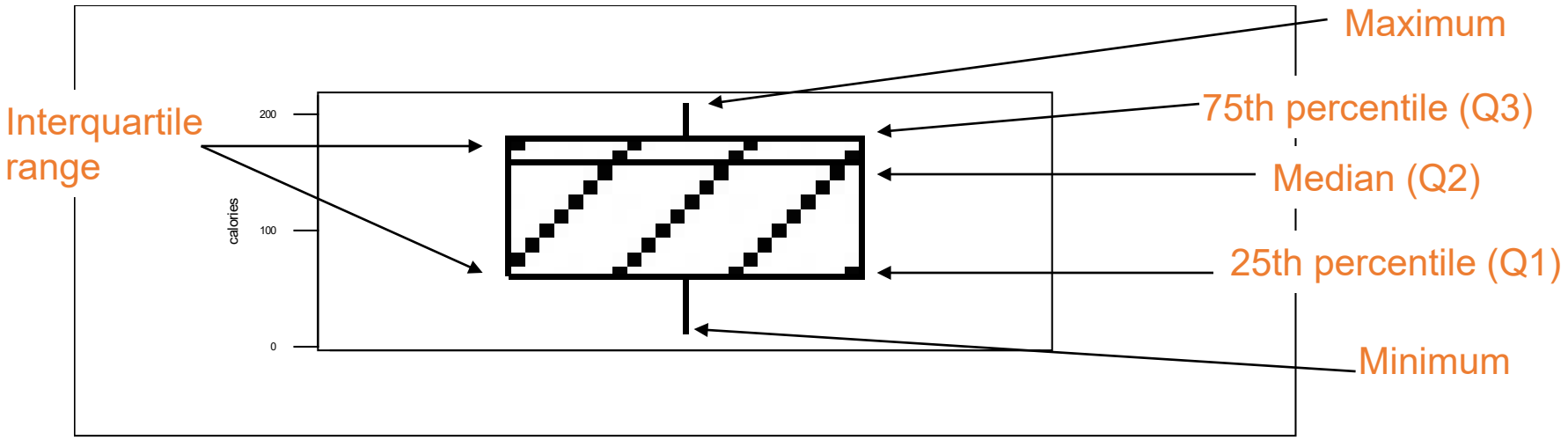

CAN MY BOYFRIEND COME ALONG?
I'M NOT YOUR BOYFRIEND!
YOU TOTALLY ARE.
I'M CASUALLY DATING A NUMBER OF PEOPLE.
BUT YOU SPEND TWICE AS MUCH TIME WITH ME AS WITH ANYONE ELSE. I'M A CLEAR OUTLIER.
YOUR MATH IS IRREFUTABLE.
FACE IT—I'M YOUR STATISTICALLY SIGNIFICANT OTHER.

Box plot: Measure of Central tendency

- **median (Q2/50th Percentile)**: the middle value of the dataset.
- **first quartile (Q1/25th Percentile)**: the middle number between the smallest number (not the “minimum”) and the median of the dataset.
- **third quartile (Q3/75th Percentile)**: the middle value between the median and the highest value (not the “maximum”) of the dataset.
- **interquartile range (IQR)**: 25th to the 75th percentile.
- **outliers (shown as green circles)**
- **“maximum”**: Q3 + 1.5*IQR
- **“minimum”**: Q1 -1.5*IQR



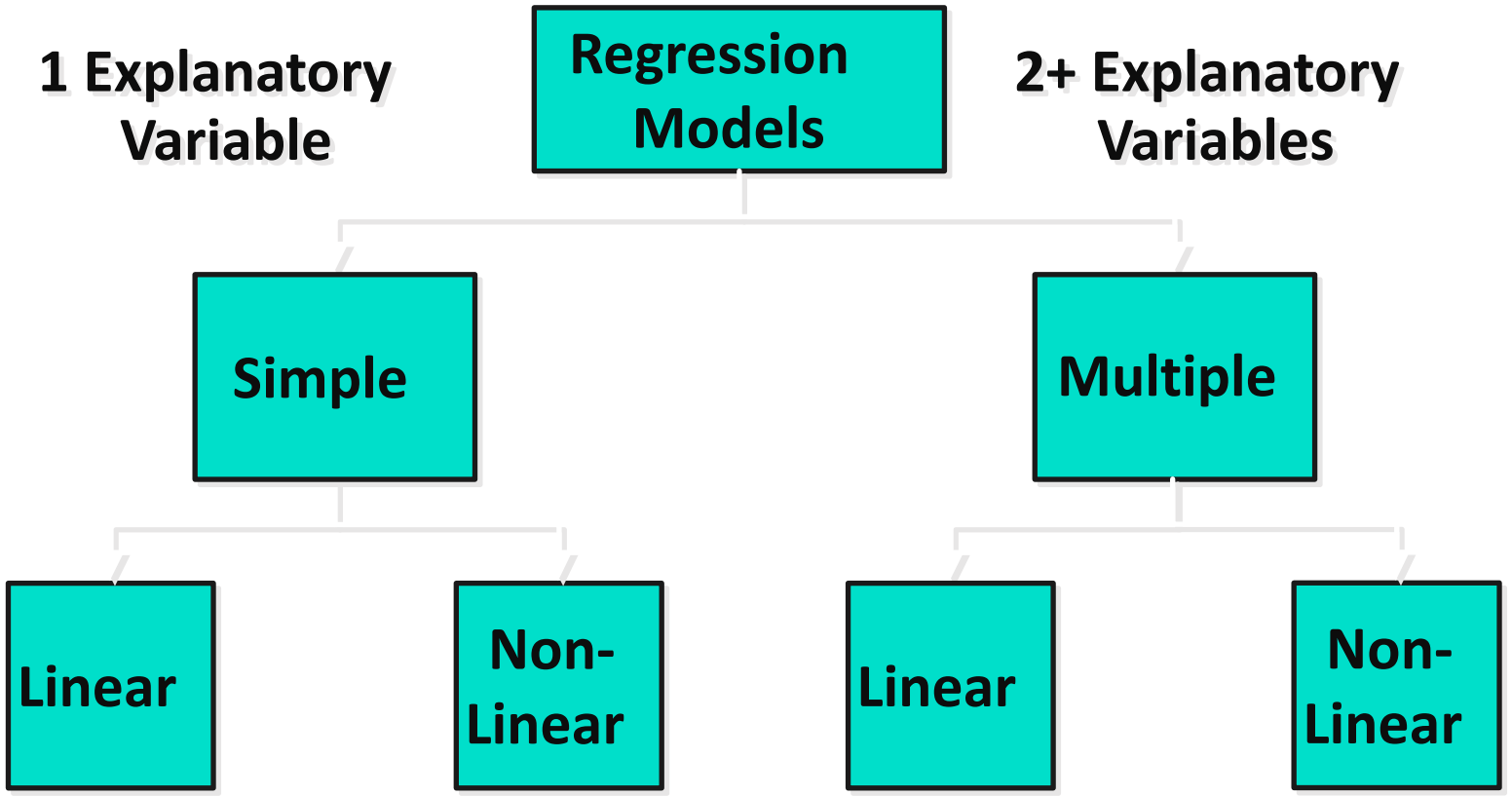
A visualization of most of the basic statistics.



Box Plot

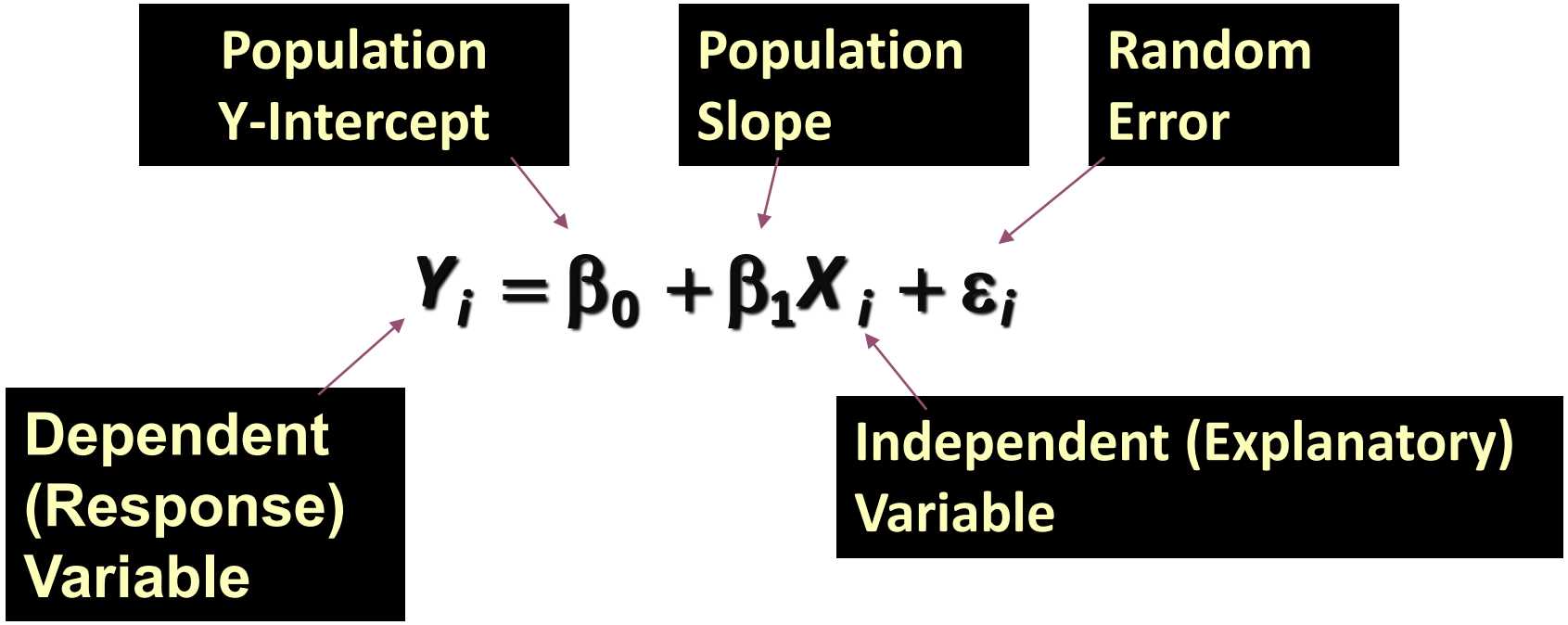
Regression Models

Types of Regression Models

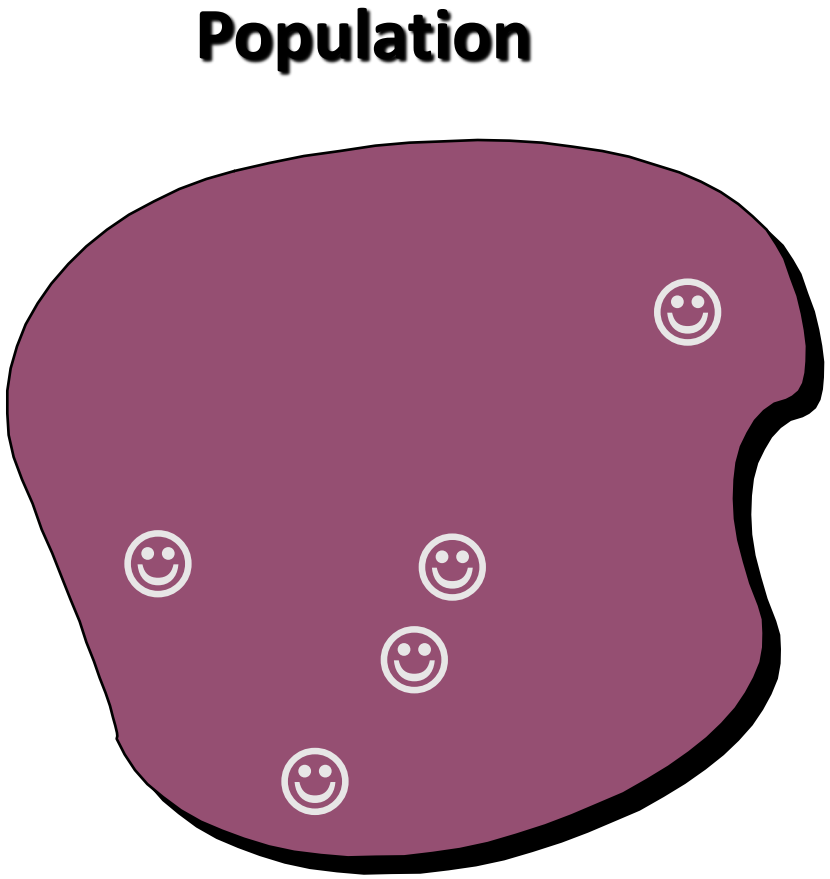


Linear Regression Model

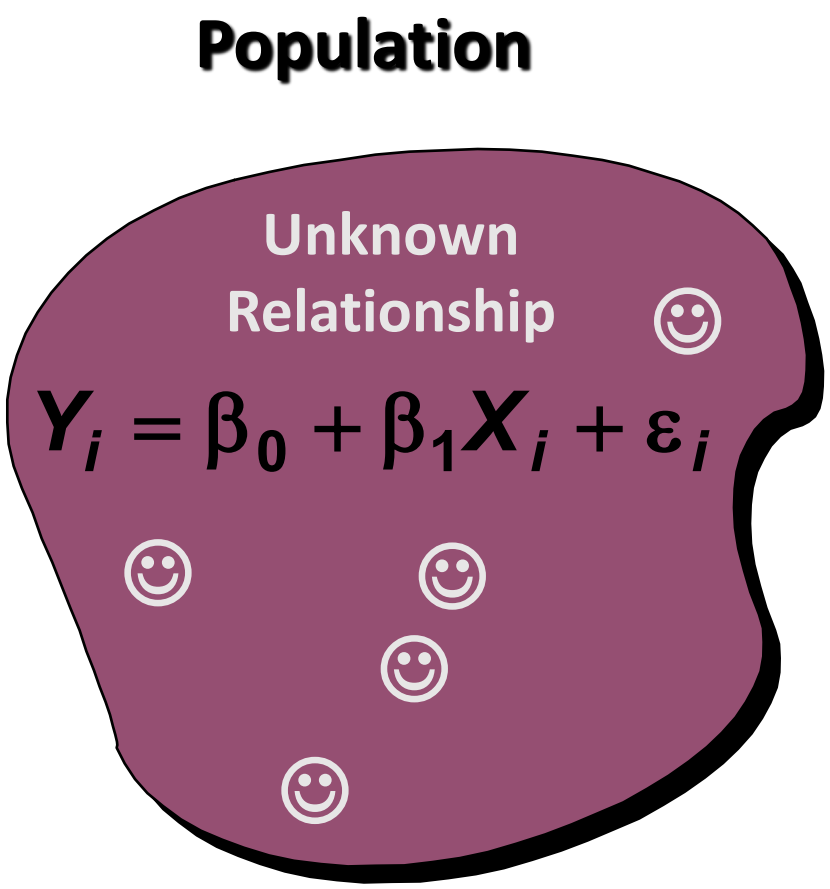
- 1. Relationship Between Variables Is a Linear Function



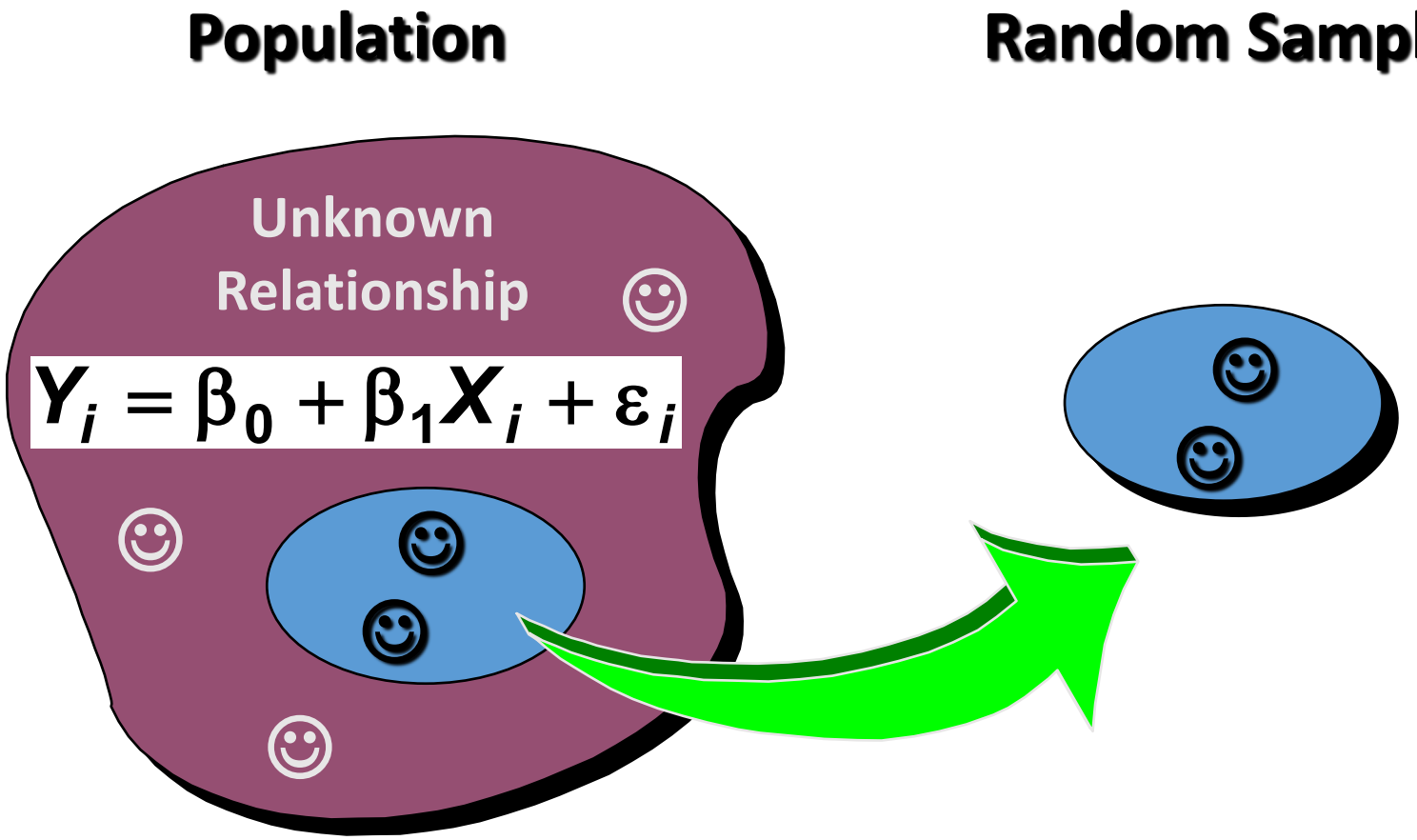
Population & Sample Regression Models



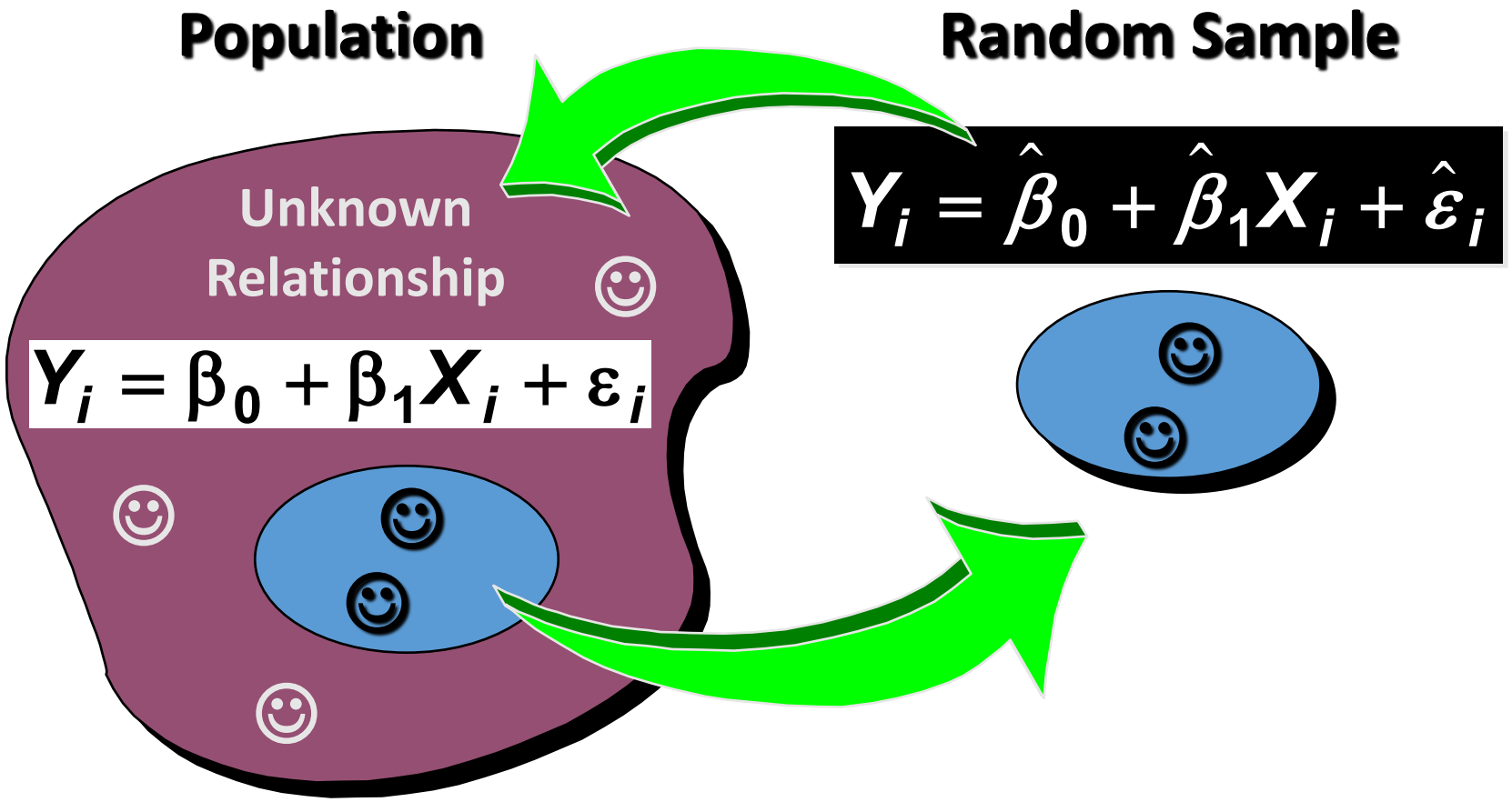
Population & Sample Regression Models



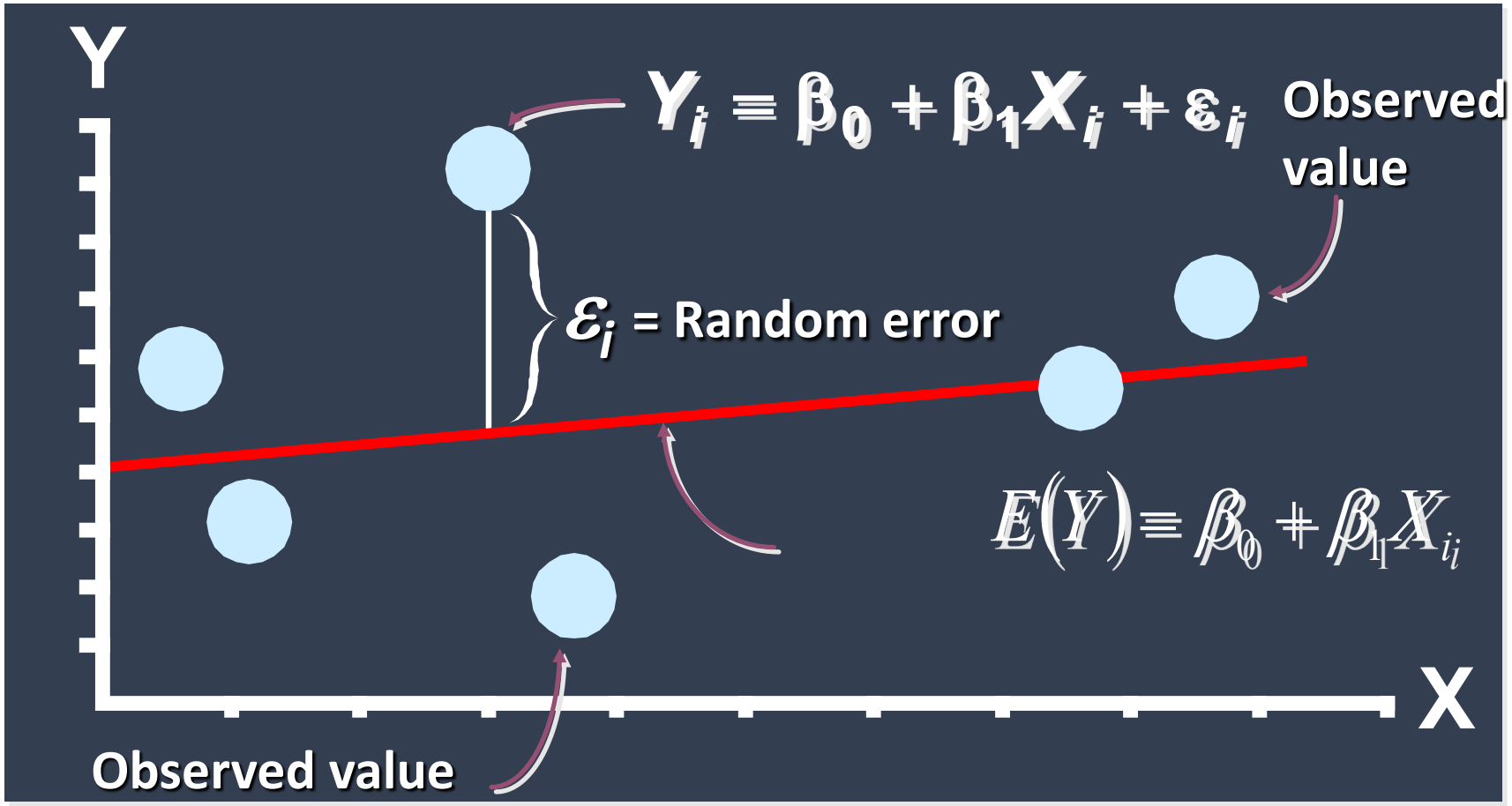
Population & Sample Regression Models



Population & Sample Regression Models



Population Linear Regression Model



Sample Linear Regression Model