Business Analytics Executive Overview

Module 1 - Analytics Beyond the Spreadsheet

Doug Laney

Lesson 1.1

Overview of Analytics

Questions Analytics Can Answer

What happened?

How or why did it happen?

What's happening now?

What is likely to happen next?

Benefits of Analytic Analysis

Discover and take advantage of trends.

Predict and forecast occurrences or amounts.

Find hidden patterns.

Identify risks and threats.

Benefits of Analytic Analysis

Identify opportunities, esp. for new products, services, etc.

Optimize the performance or quality of processes, people, and machines.

Improve customer experience.

Digitalize offerings.

Benefits of Analytic Analysis

New business models.

Monetize data.

Improve safety.

Perform “analytics for good”.

Lesson 1.2

Beyond Basic Business Intelligence

Reaching Beyond Business Intelligence

Evolving from traditional business intelligence (BI) has been slow in many organizations.

Reaching Beyond Business Intelligence

Analytics (business analytics, data analytics, business intelligence) is a key way to package and deliver information, to make it more usable and valuable to people and processes.

Basic BI Implementations Are Everywhere

Examples are:

- Personal spreadsheets
- Financial and production reports
- Executive dashboards

Business Intelligence vs Advanced Analytics

BI informs managers of performance indicators.

Advanced analytics can provide farther reaching benefits.

Reaching Beyond Business Intelligence

Analytics is a key way to package and deliver information to make it more usable and valuable to people and processes.

Lesson 1.3

The Analytics Continuum

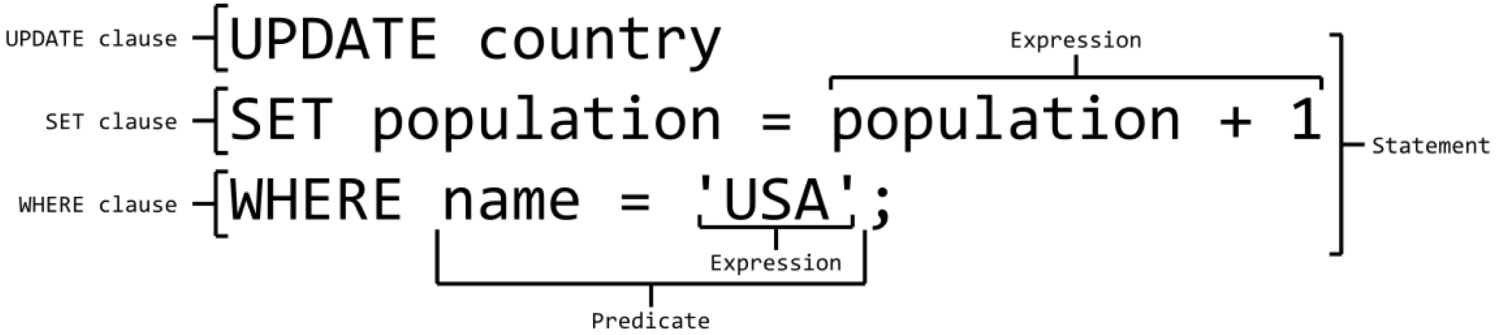
Lesson 2

Analytic Output

Ad-hoc Query

Either programmed using a query language, like SQL, or the code is generated by the application you're using.

SQL Structure Example



Benefits of Ad-hoc Query

It is empowering.

It enables self-service.

Fosters curiosity and creativity.

You can identify.

Easily modifiable.

Challenges of Ad-hoc Query

Non-standardized.

Time-consuming.

Requires some technical ability.

May discourage collaboration.

May increase system resources.

Scheduled Reporting

A report produced on a periodic basis, often weekly, monthly, quarterly, or more often.

Scheduled Reporting Examples

Quarterly financial statements.

Monthly blog hits.

Weekly safety reports.

Daily add and drop rates for a MOOC.

Scheduled Reporting Benefits

Standardized and accepted throughout the organization.

Expected and anticipated.

Easy to understand.

Easily shared with others.

Scheduled Reporting Challenges

Just a single page.

Difficult to change or modify.

Not often very interactive.

Typically hindsight oriented.

May not be available between report production periods.

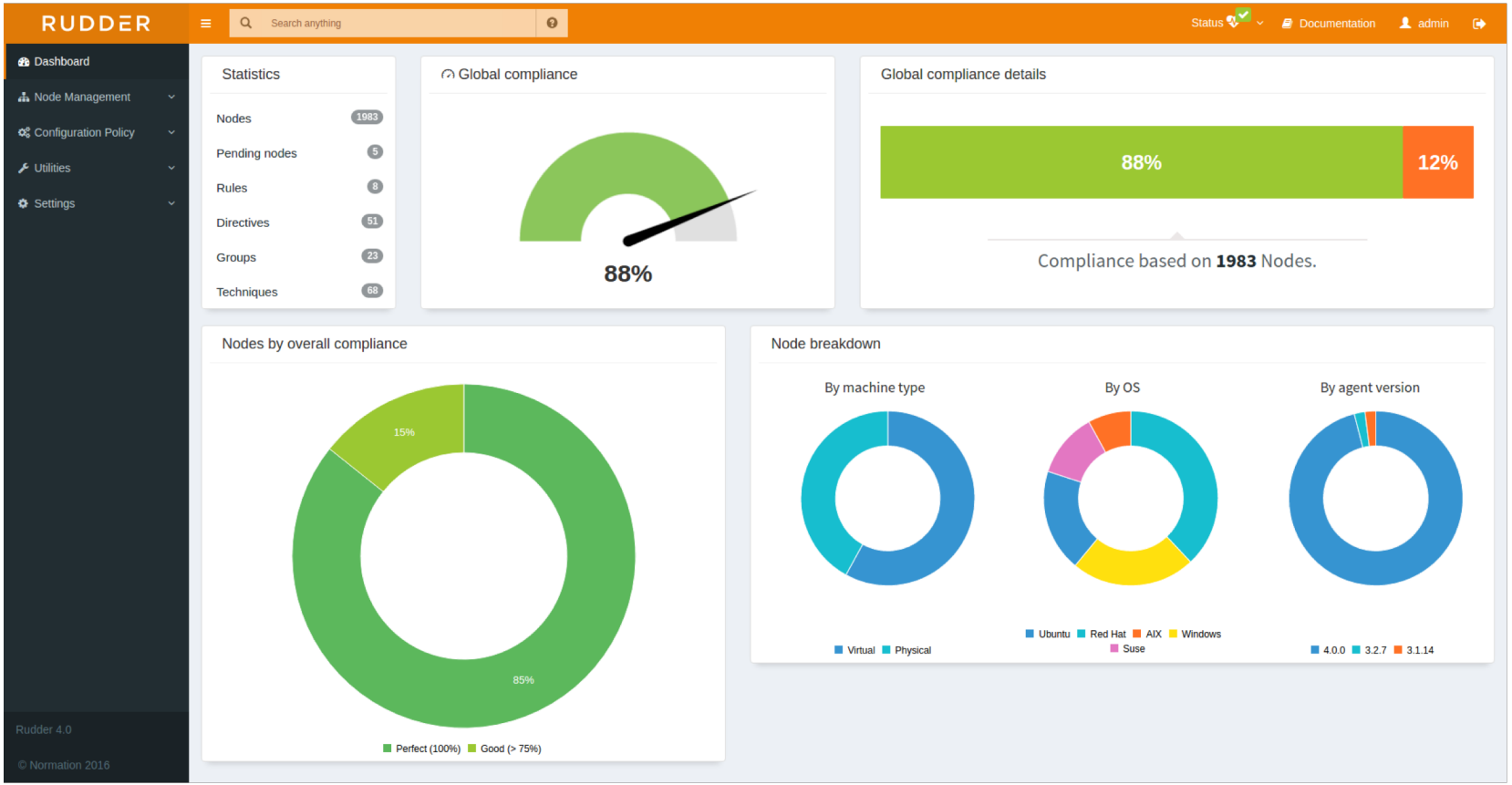
Increases reporting sprawl .

Dashboards and Scorecards

Simple interfaces or displays that provide an at-a-glance view of data.

Often a summary of Key Performance Indicators (KPIs).

Dashboards and Scorecards



(Moostik, 2017)

Dashboards and Scorecards Benefits

Easy to understand.

Can tell a story.

Great for non-technical users.

Updated continuously.

Standardized.

Good visuals.

Can be interactive and even customizable.

Dashboards and Scorecards Challenges

Often become cluttered

May display data you do not need.

Often show just historical or current data.

Conflicting performance indicators.

Often don’t show changing conditions very well.

Alerts and Notifications

A way of giving people or applications a heads-up about something important or urgent.

Alerts and Notifications

A way of giving people or applications a heads-up about something important or urgent.

Alerts and Notifications Benefits

Getting information or insights as/when needed.

Little need to design complex displays or interfaces.

Hands-free.

Can use existing messaging channels.

Don’t need to be visual – they can elicit a sound or vibration.

Can be integrated into dashboards.

Alerts and Notifications Challenges

Overloading users with alerts.

The alert itself cannot instantly convey much information, but it can carry additional information.

Embedded Analytics

Analytics for machines and applications.

Embedded Analytics Benefits

No need for any displaying data.

Faster reaction time than humans.

Instantaneous feedback (closed loop analytics).

Continuous operation.

Labor reduction.

Embedded Analytics Challenges

May not be able to respond to ambiguous context or unforeseen conditions.

Limited ability to apply human experience.

Loss of institutional (tacit) knowledge.

Lesson 3

Basic Analytic Techniques

Condensing

Basic calculations like sums, counts, averages, and ranges like standard deviations, or functions like lookups, filtering, searching, and sorting.

Condensing

Simple functions to gain an aggregate understanding of the data–as a whole or a selection of the data.

Correlation

Discovers the strength of (or lack of) relationship or association between two or more things.

Correlation

Methods include simple linear regression using the least-squares method or more sophisticated Bayesian approaches.

Pattern Matching

Identifying recurring or known sequences in data or determining to what degree a selection of data approximates a known entity.

Classification/Segmentation

Entities are grouped into established segments according to particular characteristics.

Clustering

Creating new artificial or derived segments.

Anomaly Detection

The identification of outliers or rare observations, items, or events in data that contrast with the norm.

Lesson 4

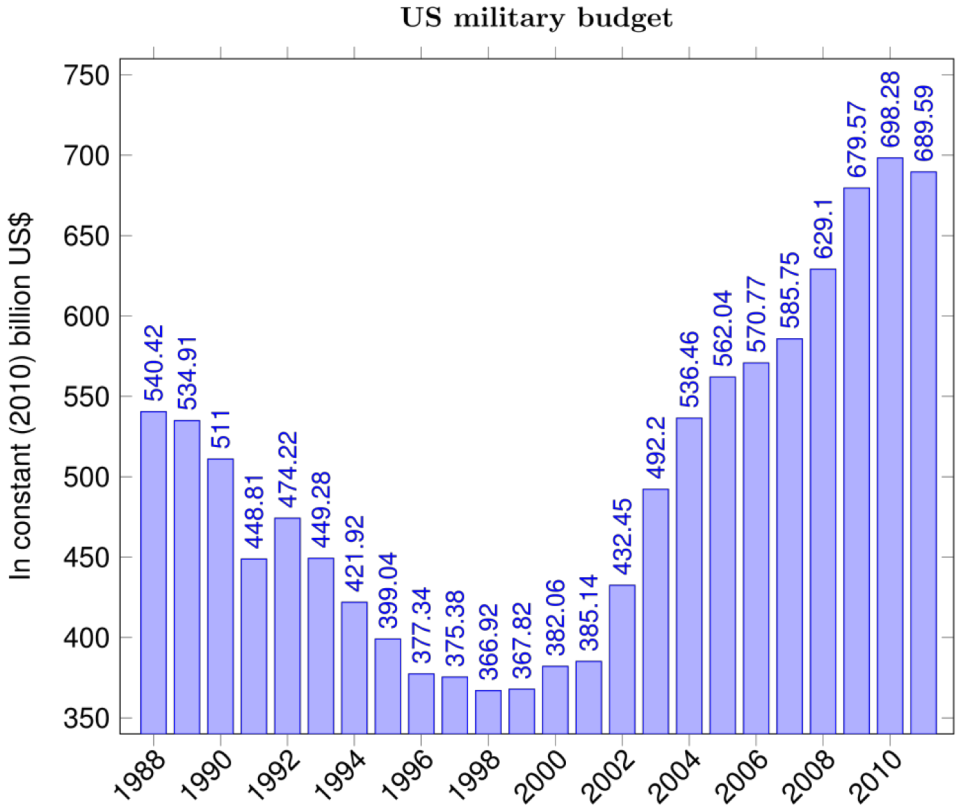
Analytic Graphical Representation (Visualization)

Table

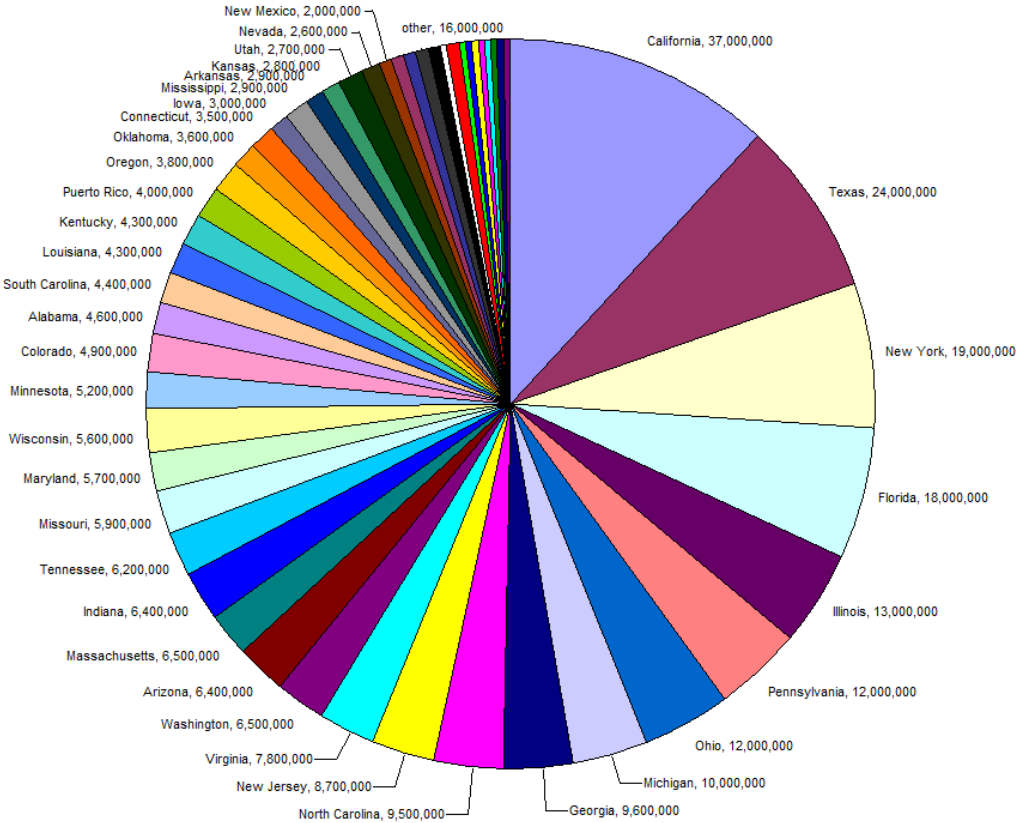
A vs. B	Residual Value A	Residual Value B	Higher	Higher as % of Lower	Intensity
Accord vs. Accord Hybrid	52%	46%	A	113%	3
Accord vs. Pilot	52%	44%	A	118%	4
Accord vs. CR-V	52%	55%	B	106%	2
Accord vs. Element	52%	48%	A	108%	2
Accord vs. Odyssey	52%	48%	A	108%	2
Accord Hybrid vs. Pilot	46%	44%	A	105%	2
Accord Hybrid vs. CR-V	46%	55%	B	120%	5
Accord Hybrid vs. Element	46%	48%	B	104%	1
Accord Hybrid vs. Odyssey	46%	48%	B	104%	1
Pilot vs. CR-V	44%	55%	B	125%	6
Pilot vs. Element	44%	48%	B	109%	2
Pilot vs. Odyssey	44%	48%	B	109%	2
CR-V vs. Element	55%	48%	A	115%	4
CR-V vs. Odyssey	55%	48%	A	115%	4
Element vs. Odyssey	48%	48%	A	100%	1

(Tschneidr, 2017)

Bar Chart


US military budget
In constant (2010) billion US$
750
700
650
600
550
500
450
400
350
540.42
534.91
511
448.81
474.22
449.28
421.92
399.04
377.34
375.38
366.92
367.82
382.06
385.14
432.45
492.2
536.46
562.04
570.77
585.75
629.1
679.57
698.28
689.59
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010

Pie Chart

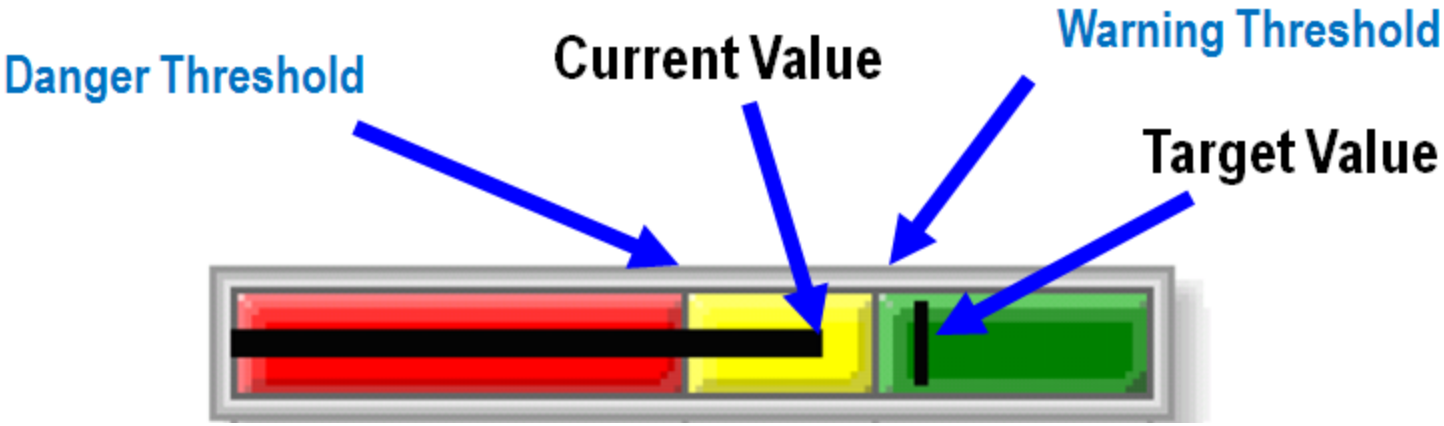

California, 37,000,000
Texas, 24,000,000
New York, 19,000,000
Florida, 18,000,000
Illinois, 13,000,000
Pennsylvania, 12,000,000
Ohio, 12,000,000
Michigan, 10,000,000
Georgia, 9,600,000
North Carolina, 9,500,000
New Jersey, 8,700,000
Virginia, 7,800,000
Washington, 6,500,000
Arizona, 6,400,000
Massachusetts, 6,500,000
Indiana, 6,400,000
Tennessee, 6,200,000
Missouri, 5,900,000
Maryland, 5,700,000
Wisconsin, 5,600,000
Minnesota, 5,200,000
Colorado, 4,900,000
Alabama, 4,600,000
South Carolina, 4,400,000
Louisiana, 4,300,000
Kentucky, 4,300,000
Puerto Rico, 4,000,000
Oregon, 3,800,000
Oklahoma, 3,600,000
Connecticut, 3,500,000
Iowa, 3,000,000
Mississippi, 2,900,000
Arkansas, 2,900,000
Kansas, 2,800,000
Utah, 2,700,000
Nevada, 2,600,000
New Mexico, 2,000,000
other, 16,000,000

Line Charts

(XOLE2129, 2018)

Gauges and Meters

Bullet Charts

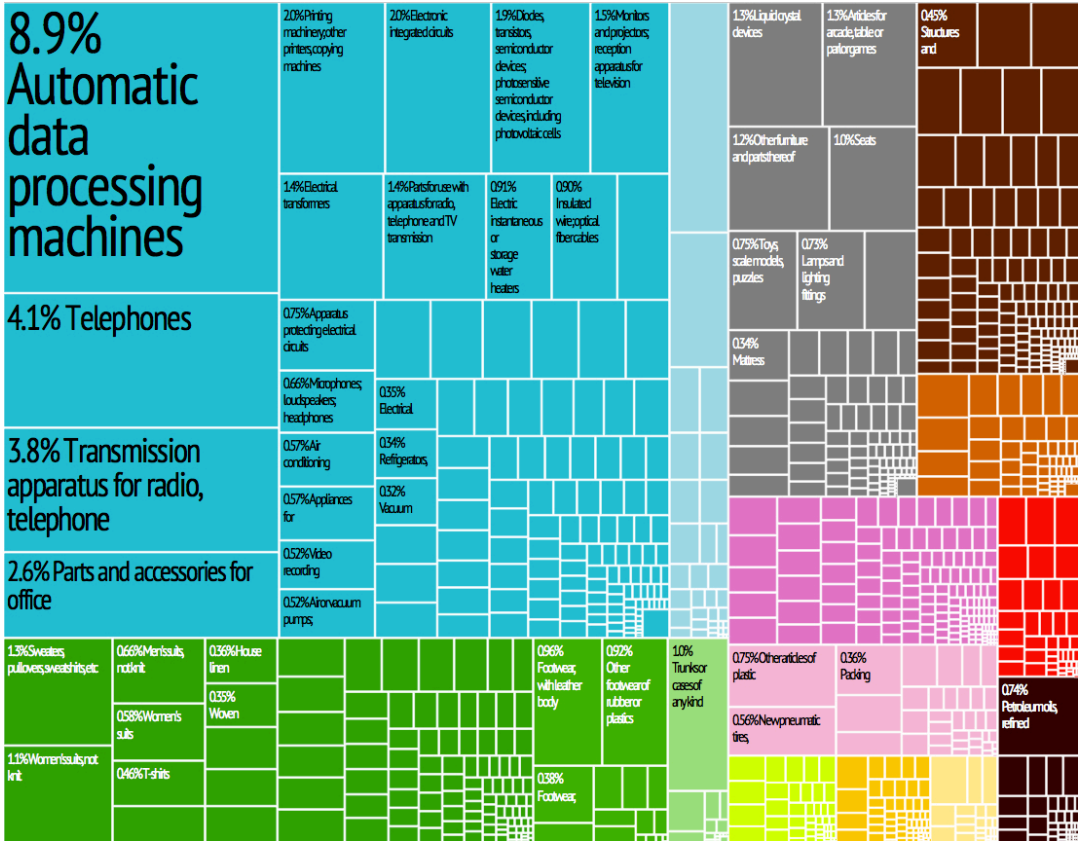


(Dmccreary, 2010)

Eye Tracking Heatmap

(Tschneidr, 2017)

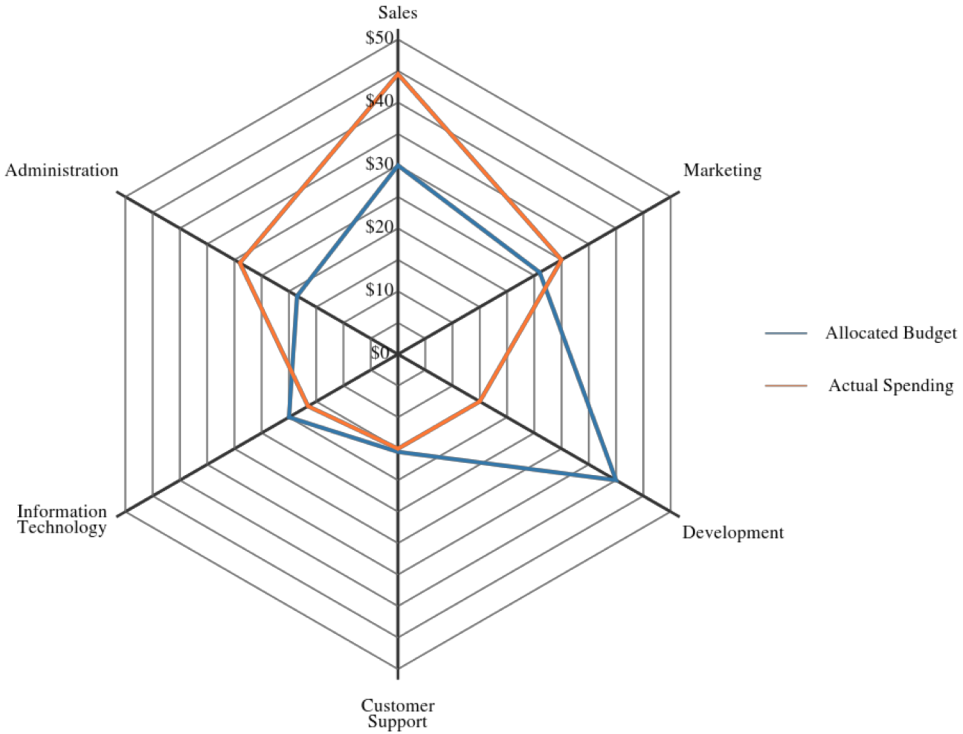
Treemaps



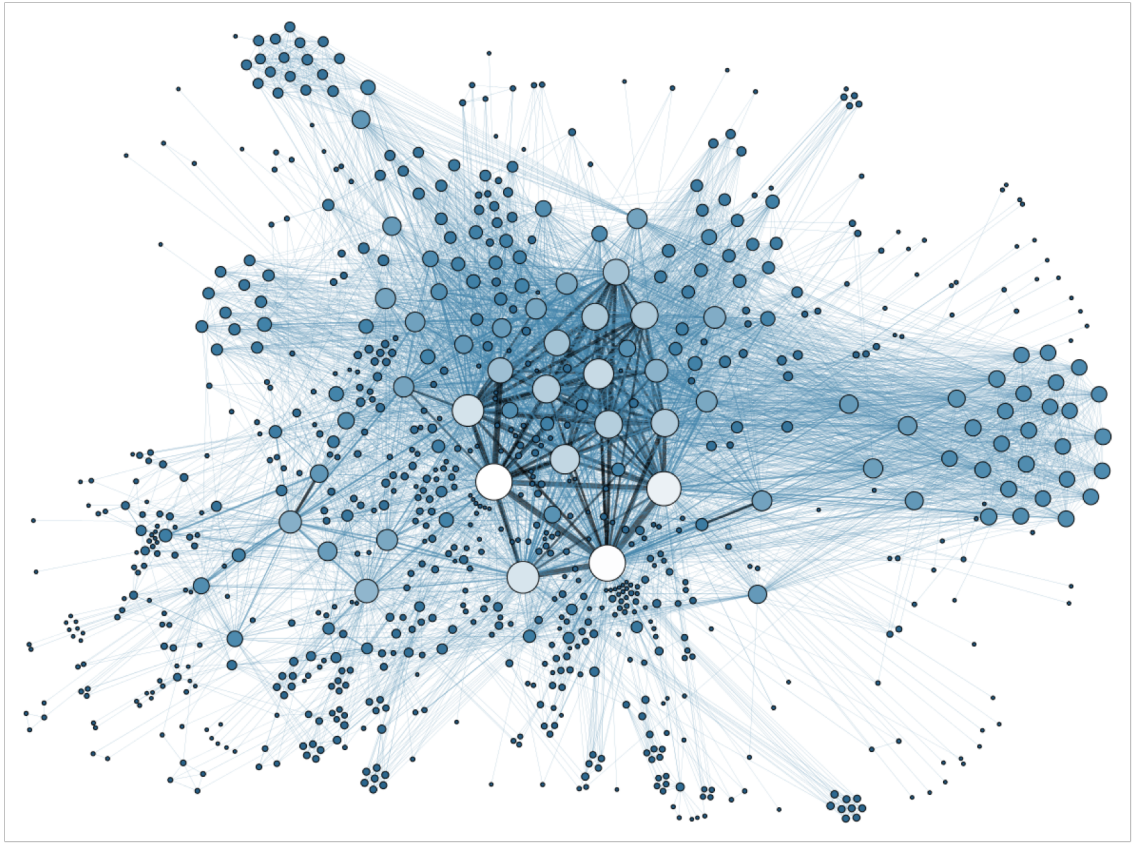
(Haussman, Cesar Hidalgo, et. al., 2012)

Sparklines

Spider (Radar) Charts

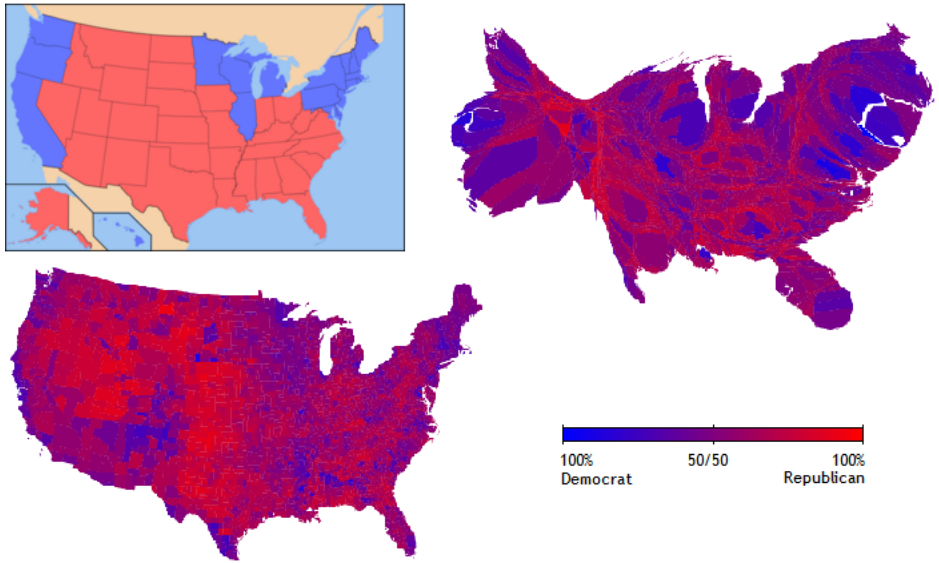


Network Diagrams



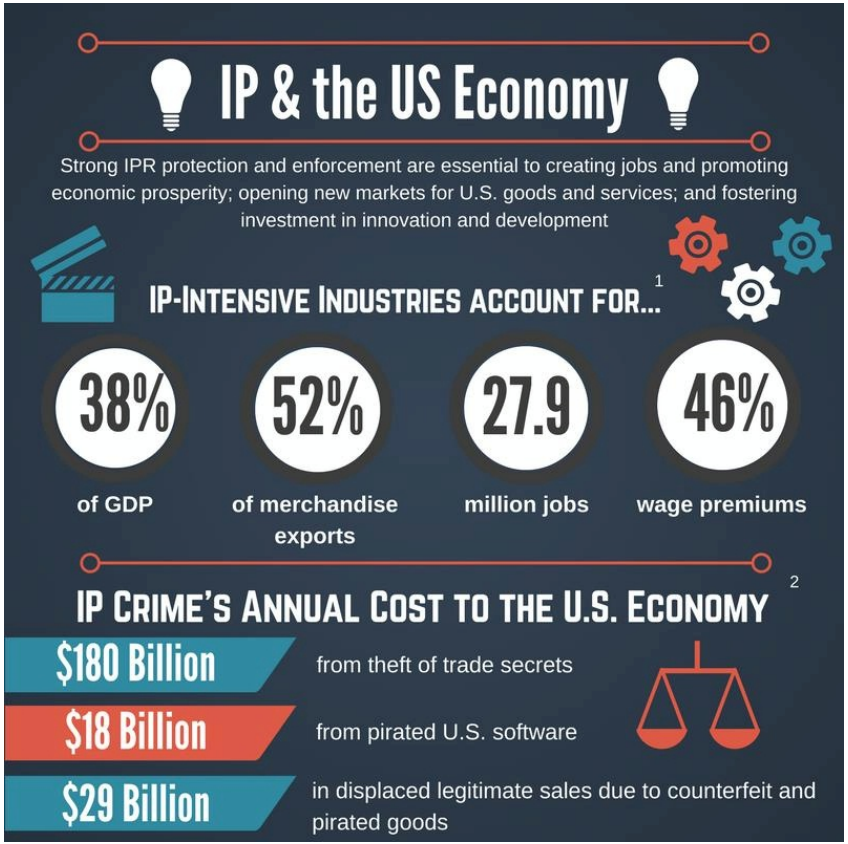
(Grandjean, 2013)

Maps or Cartograms



(Zscout370, Newman, Gastner, Shalizi, 2005)

Infographics



Lesson 5.0

Key Analytics Concepts to Know (Data Literacy)

Lesson 5.1

Data, What Data? (Sources of Data)

Enterprise (Operational Data)

Includes customer data, transaction data, process and production data, inventory and sales data, data about employees, partners, suppliers, and so forth.

Dark Data

Emails, documents, reports, and system log files that track the operation and performance of devices, computers, software, or other equipment.

Open (Public) Data

Data made freely available by government or other organizations. It also includes public financial reports.

Social Media Content

Data from social media sites is available for economic benefits, but also to identify trends.

Syndicated Data (Data brokers and marketplaces)

Data collected and sold by data aggregators or data brokers like AC Nielsen, Experian, Neustar, Dun & Bradstreet, Bloomberg, Reuters, Standard & Poors, etc

Web Content

Websites are a goldmine of information, and technologies have emerged that can crawl websites and harvest content from competitor product pricing to executives that have joined or left a company.

Partner Data

Data on suppliers and suppliers of suppliers.

Reference Data

Master data and meta data.

Internet of Things (IOT)

Data from and about equipment, machines, cars, planes, manufacturing, mining equipment, traffic signals, etc.

Alternative Data

External or exogenous data that can provide some unique or supplemental competitive advantage over data you already generate or collect.

Lesson 5.2

Where to Put the Data and How to Get It There?

Database/DBMS

A software application for storing, organizing, indexing, processing, and querying data.

Dataset

A somewhat generic term for a collection of related records.

Extract

A dataset that is copied from a database.

Data Warehouse

A special data structure (or schema or data model) architected to integrate data from multiple source systems for high performance data querying and analysis.

Data Warehouse

- Subject oriented
- Historical
- Time variant
- Non-volatile

Data Mart

A kind of mini-data warehouse that is specific to a particular subject.

Operational Data Store

A data warehouse that integrates data from multiple systems on a more frequent, real-time, or near real-time basis.

Data Lake

An analytic data structure that allows data to be structured as-is, with structures being designed then applied at the time data is queried.

Logical and Virtual Data Warehouses

Logical– a hybrid form of data structures.

Virtual– a view (a special schema just for querying) is created that can make it look to a data analyst that that he or she is querying a single datawarehouse, when in fact the query is being distributed across multiple source systems.

Workspace or sandbox

An area designated for data scientists or advanced business analysts to play with extracts of data.

Archive

A slower, cheaper form of a database where seldom used data is stored.

Backup

An exact copy of a database, usually updated continuously.

Schemas

Special ways to organize data for ease of navigation and querying.

Multidimensional databases

Databases that automatically create or impose multidimensional schemas.

Content, Content/Document Management Systems

Systems for storing, searching, and managing documents.

Blockchain

A distributed, decentralized, public leger of transactions or other records.

Data Integration (ETL)

Special software applications that one programs to extract, transform, and load (ETL) data into a data warehouse or other analytic structure.

API

An application programming interface (API) is a defined method provided by a business application to access functionality or its data without having to query the underlying database directly.

Data Preparation or self-service data prep

A class of simple-to-use technology for business users rather than programmers.

Lesson 5.3

Analytic Methods, Techniques, and Concepts

Business Intelligence

The discipline and tools for analyzing and reporting on data.

SQL

Short for structured query language, SQL is the predominant way to query databases.

NoSQL

A type of non-relational database in which data is modeled and queried using its own more efficient language.

OLAP/MOLAP/ROLAP

General terms for performing analytic functions like searching, sorting, filtering, and calculating data

Drill-down/across

Performing additional queries to access increasingly specific data.

Drill-down/across

Performing additional queries to access increasingly specific data.

Algorithm

A calculation involving a number of discrete steps or processes.

Enterprise Reporting

Defining, producing, and publishing periodic standardized analysis.

Corporate Performance Management (CPM)

The methodologies, metrics, processes, and systems used to monitor and manage the business performance of an organization.

Stream Analytics

Analyzing data in motion, in real-time, or near real-time almost immediately after it’s created.

Visualization

Presenting data and analyses in easily interpreted and compelling graphical ways.

Write-back

Calculations or other output that is stored back in the database for applications to access for automated decision-making.

Graph Analytics

A technique for determining the strength and direction of the relationship between multiple objects.

Natural Language Query (NLQ)

Interfaces or applications that convert English (or other language) questions into SQL.

Spatial Analysis

Understanding the relationships between objects and geographic positions.

A-B Testing

A standard method for analytic experimentation.

Sentiment Analysis

The relative affinity of individuals or groups for something.

Text (Voice, Video) Analytics

A special class of methods and tools for finding patterns or otherwise understanding various kinds of unstructured content.

Self-Service Analytics

Analytics performed by business people who are not analytics professionals.

Augmented Analytics

Analytic applications with features to help users automatically identify interesting or important data.

AI and Machine learning

Technologies for emulating how humans learn and apply knowledge or perform tasks.

Lesson 5.4

Other Data Concepts

More Important Data Concepts

Big data

Data science

Data mining

Sampling

Structured vs unstructured data

Data model/ modelling

Data architecture/ architect

More Important Data Concepts

Data lifecycle

Data strategy

Derived data

Data as a service

Data quality/ cleansing

More Important Data Concepts

Data governance

Anonymization, encryption, masking

Stored procedures

Natural language generation

Lesson 6.0

Purpose-Built Analytics Solutions

Lesson 6.1

Data-Specific Analytic Solutions

Corporate Performance Management (CPM)

The methodologies, metrics, processes, and systems used to monitor and manage the business performance of an enterprise.

Web and Mobile Analytics

The core process of web analytics is collecting, monitoring, and analyzing customers' behavior on desktop and mobile sites.

Logfile Analytic

Analytics purpose-built for understanding what is happing under the covers of a machine, device, or application.

Social Media Analytics

The process of understanding what is being discussed or shared over social media channels like FB, Twitter, LinkedIn, Instagram, WeChat, blogs, and so forth.

Business Process Analysis

Tools that are primarily intended for use by business end users looking to document, analyze, and streamline complex processes.

Lesson 6.2

Business Function-Specific Analytics

Sales and Marketing

- Understand customer preferences.
- Identify trends.
- Find new prospects or customers (enlarge pipeline).
- Identify new markets.
- Streamline sales process.

Sales and Marketing - Westpac

By analyzing data by customer, they were able to increase target offerings from just 1% of customers to more than 25%, yielding $25M is just the first nine months.

Customer Experience

- Understand customer experience
- Measure customer satisfaction.
- Identify instances of potential customer dissatisfaction.
- Improve information available to customer service agents.
- Streamline customer service process.

Customer Experience- Burberry

Integrates and analyzes data from 800K followers and 15M fans on Instagram and Facebook, customers' Twitter posts, purchase history, and surveys, and data pulled from RFID tags on clothing.

New Products or Services

- Understand unanticipated product/service usage
- Identify market gaps (white space)
- Identify new types of materials
- Improve information available to designers and engineers
- Streamline the new product introduction (NPI) process

New Products or Services - AMBiotech

Analyzing tens of billions of short DNA sequences in the creation of customized x-aptamers.

Financial

- Understand impact of financials on the business.
- Identify instances of fraud (closer to when/where they occur).
- Identify new leading financial indicators (improved forecasting).
- Improve information available to accountants, controllers, and investors.
- Streamline the financial reporting process.

Financial - Ahold

Detected that vendor was charging for expensive product while delivering cheaper product.
Recovered value ~$160K.

Human Resources

- Understand the motivations and habits of people.
- Identify opportunities to improve people's capabilities.
- Identify opportunities to automate repetitive tasks.
- Identify workplace issues.
- Improve information available to employees.
- Streamline the talent acquisition, evaluation, and other HR processes.

Human Resources - Lockheed Martin

Identified a concise set of attributes that are predictors of program performance.

Increased program foresight by 3x. Enables earlier program assessments.

Save\ds potential losses of hundreds of millions of dollars from program delays.

Supply Chain

- Understand all the moving parts and processes of the supply chain.
- Model the supply network down to multiple levels of suppliers.
- Analyze and prevent systemic or ad-hoc supply chain issues.
- Identify opportunities to improve, eliminate, or outsource processes.
- Improve information availability to supply chain processes.

Supply Chain - NCR

$110M contribution to revenues.

5% increase in sales efficiency.

Information Technology (IT) and Operational Technology (OT)

- Understand and enable flexible, dynamic capacity and performance.
- Enable the cooperation of disparate technologies.
- Analyze and limit system downtime.
- Protect systems from hacking.
- Identify opportunities to improve, eliminate, or outsource technologies (environmental concerns, too).
- Correlate technology metrics and business metrics.

IT and OT - Tesco

200,000 refrigeration-related data points across all stores, including compressor pressures, liquid levels, temperatures, energy use, currents, etc.

70 million daily from readings every 3 minutes.