



Lisbon School
of Economics
& Management
Universidade de Lisboa



Carlos J. Costa

BUSINESS INTELLIGENCE

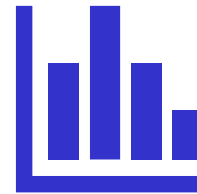
What Is Business Intelligence?



Business Intelligence



Goal is to deliver accurate real-time information to decision makers



Main analytic functionalities of BI systems

Production reports

Parameterized reports

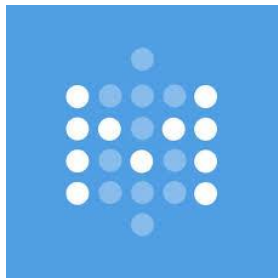
Dashboards/scorecards

Ad hoc query/search/report creation

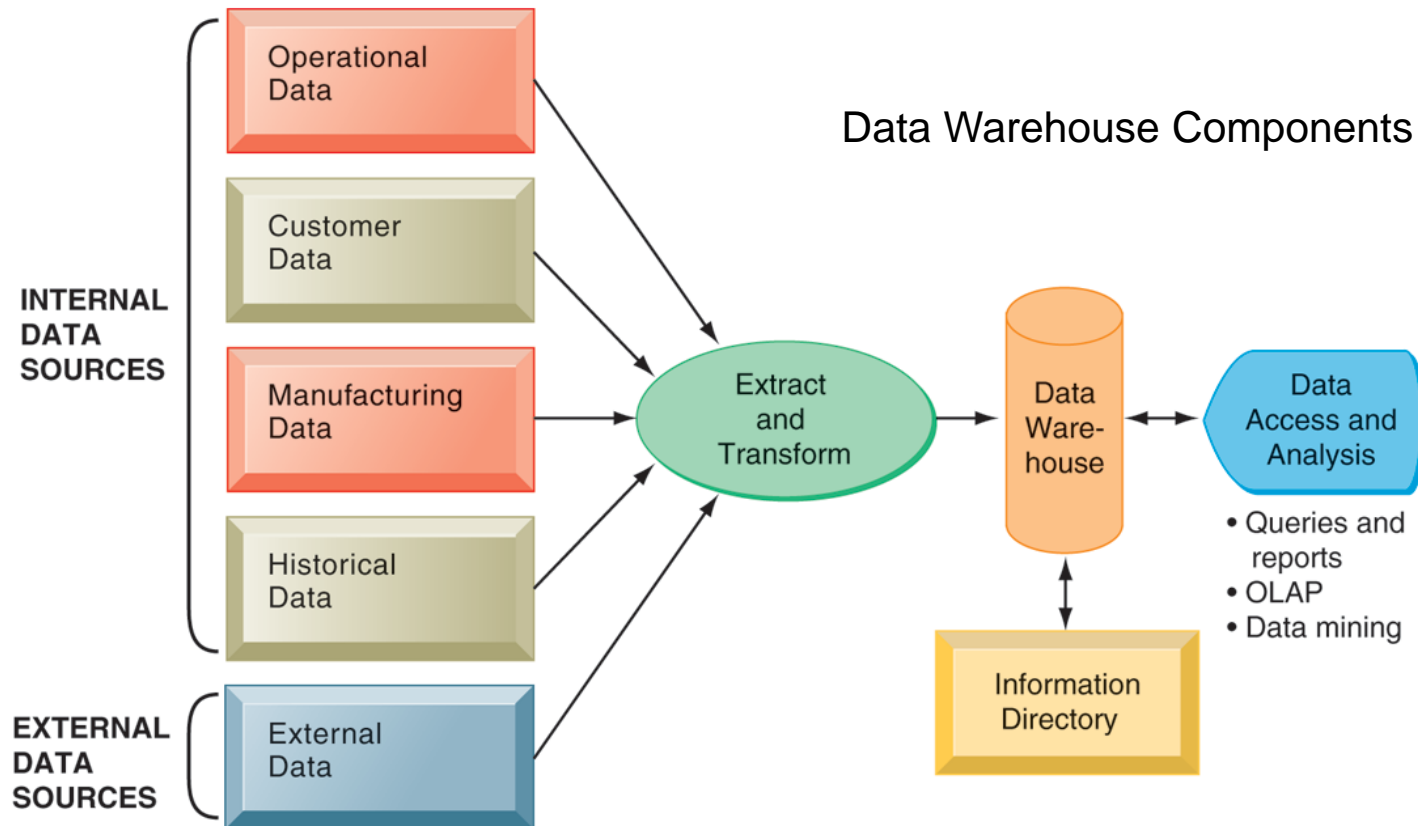
Drill down

Forecasts, scenarios, models

Business intelligence vendors

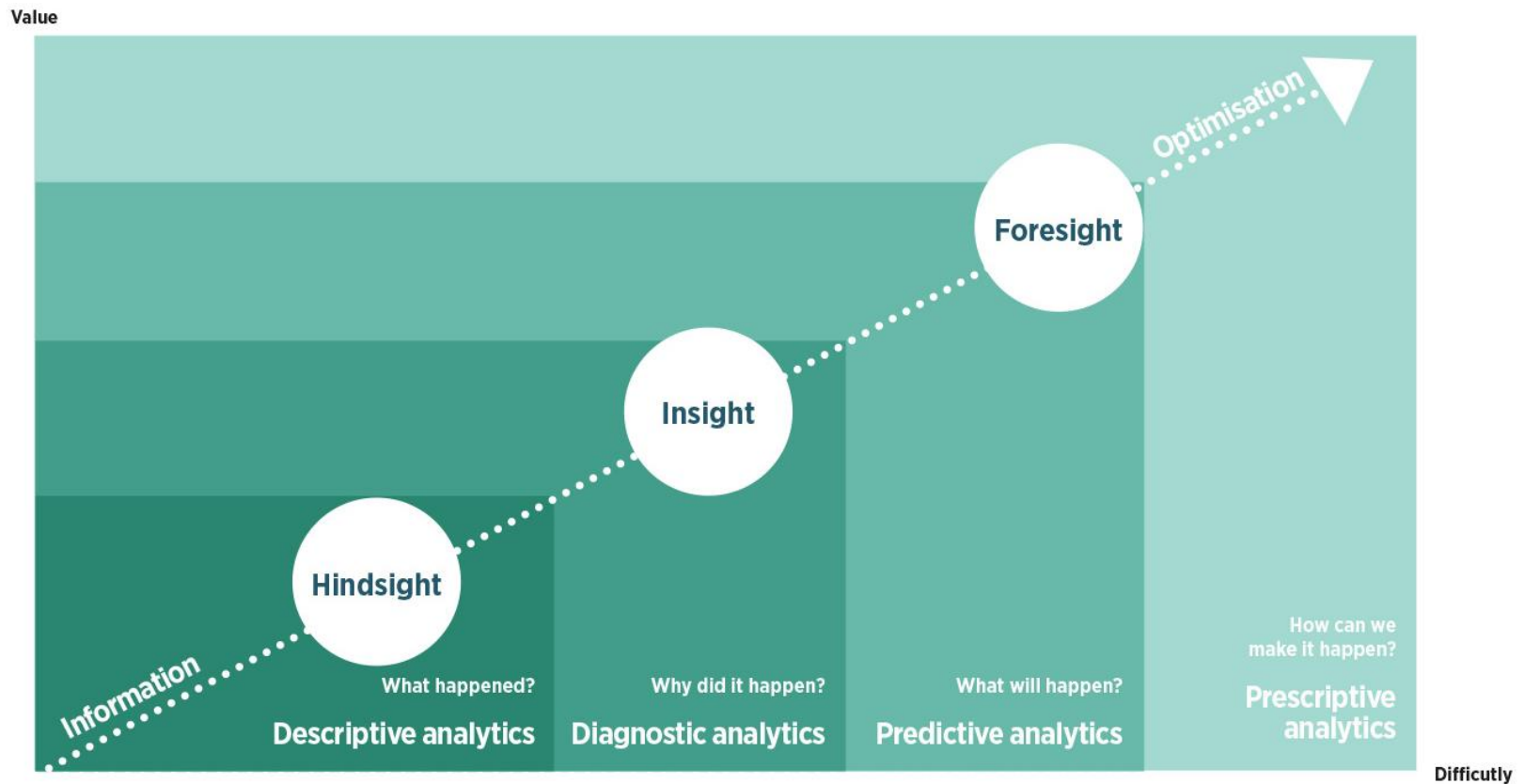


Data Warehouse



Laudon & Laudon (2012)

Predictive Analytics



Difficulty

Source: Gartner

40 ZETTABYTES

[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

6 BILLION PEOPLE have cell phones



WORLD POPULATION: 7 BILLION

Volume SCALE OF DATA



It's estimated that **2.5 QUINTILLION BYTES**

[2.3 TRILLION GIGABYTES]
of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT are shared on Facebook every month



Variety DIFFERENT FORMS OF DATA



By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month



400 MILLION TWEETS are sent per day by about 200 million monthly active users

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION during each trading session



Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

- almost 2.5 connections per person on earth



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure



1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions



in one survey were unsure of how much of their data was inaccurate

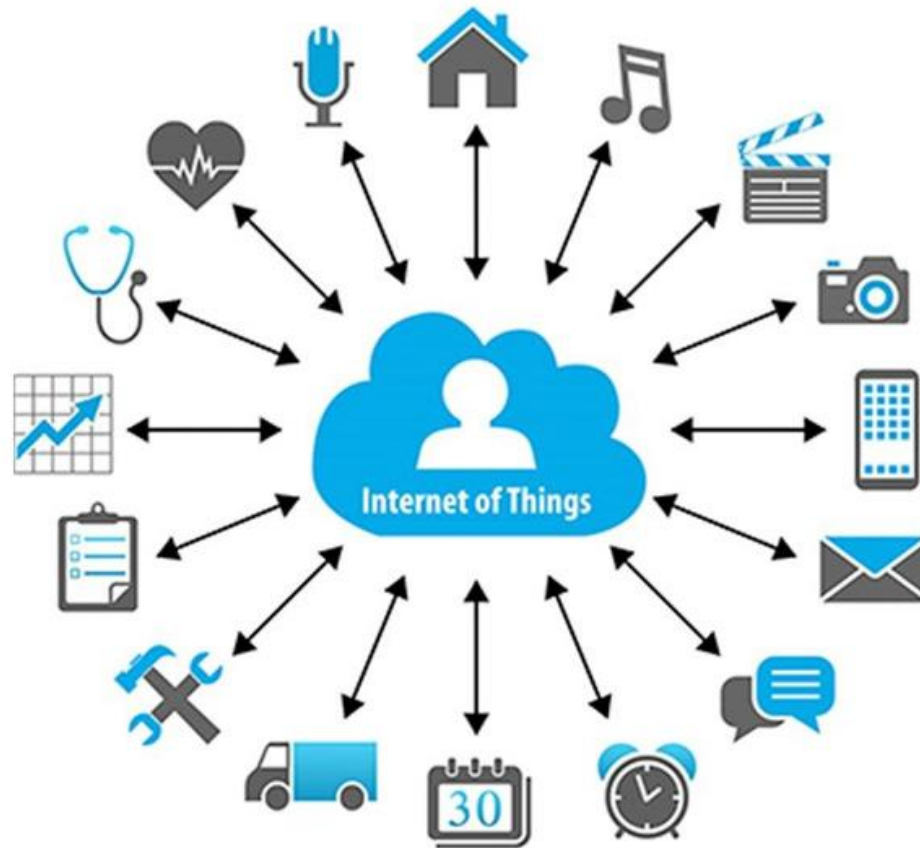


Veracity UNCERTAINTY OF DATA

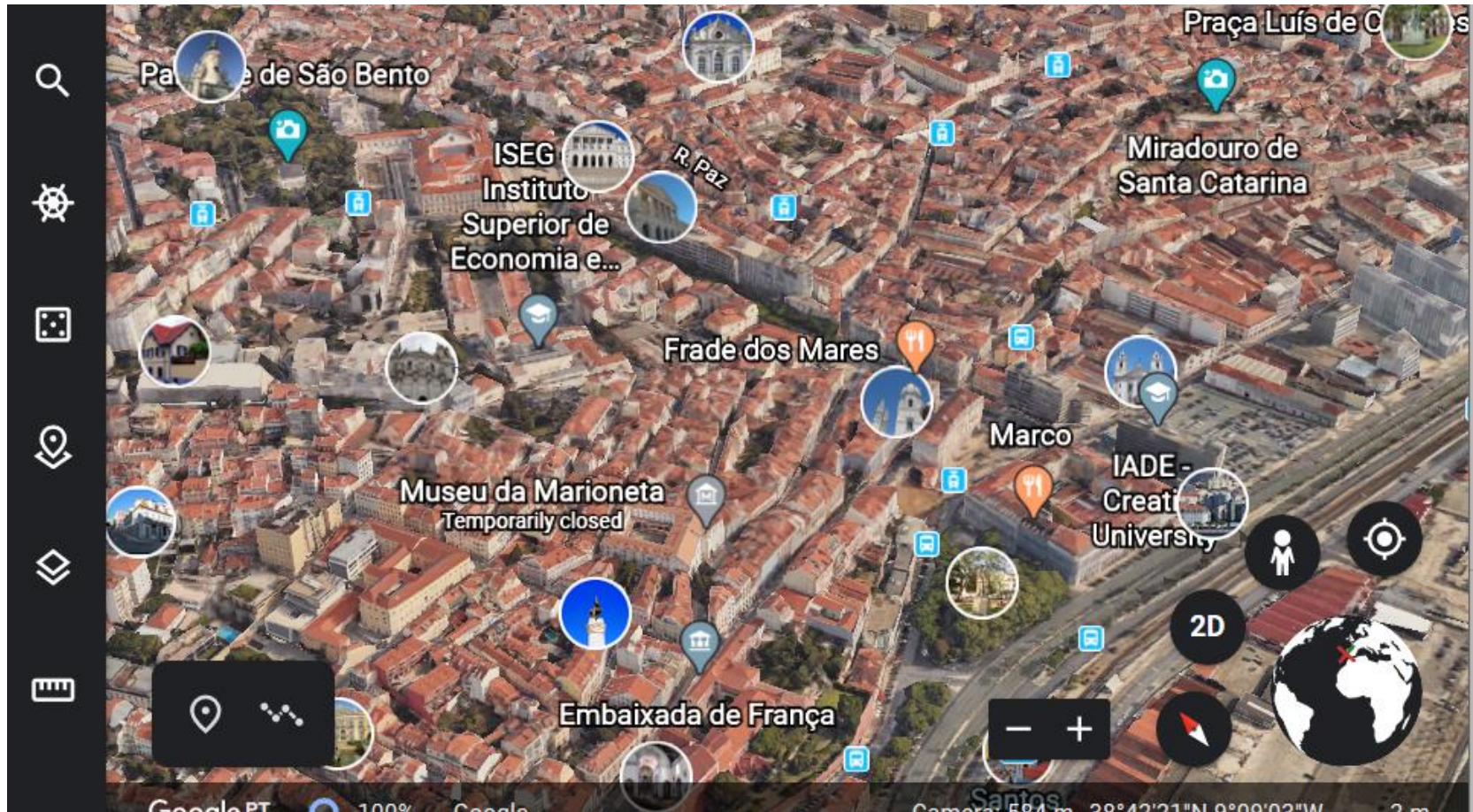
Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



Internet of Things (IoT)



Location Analytics and Geographic Information Systems



BI Users

**Power Users:
Producers
(20% of employees)**

Capabilities

**Casual Users:
Consumers
(80% of employees)**

IT developers

Production Reports

Customers/Suppliers
Operational employees

Super users

Parameterized Reports

Senior managers

Business analysts

Dashboards/Scorecards

Managers/Staff

Analytical modelers

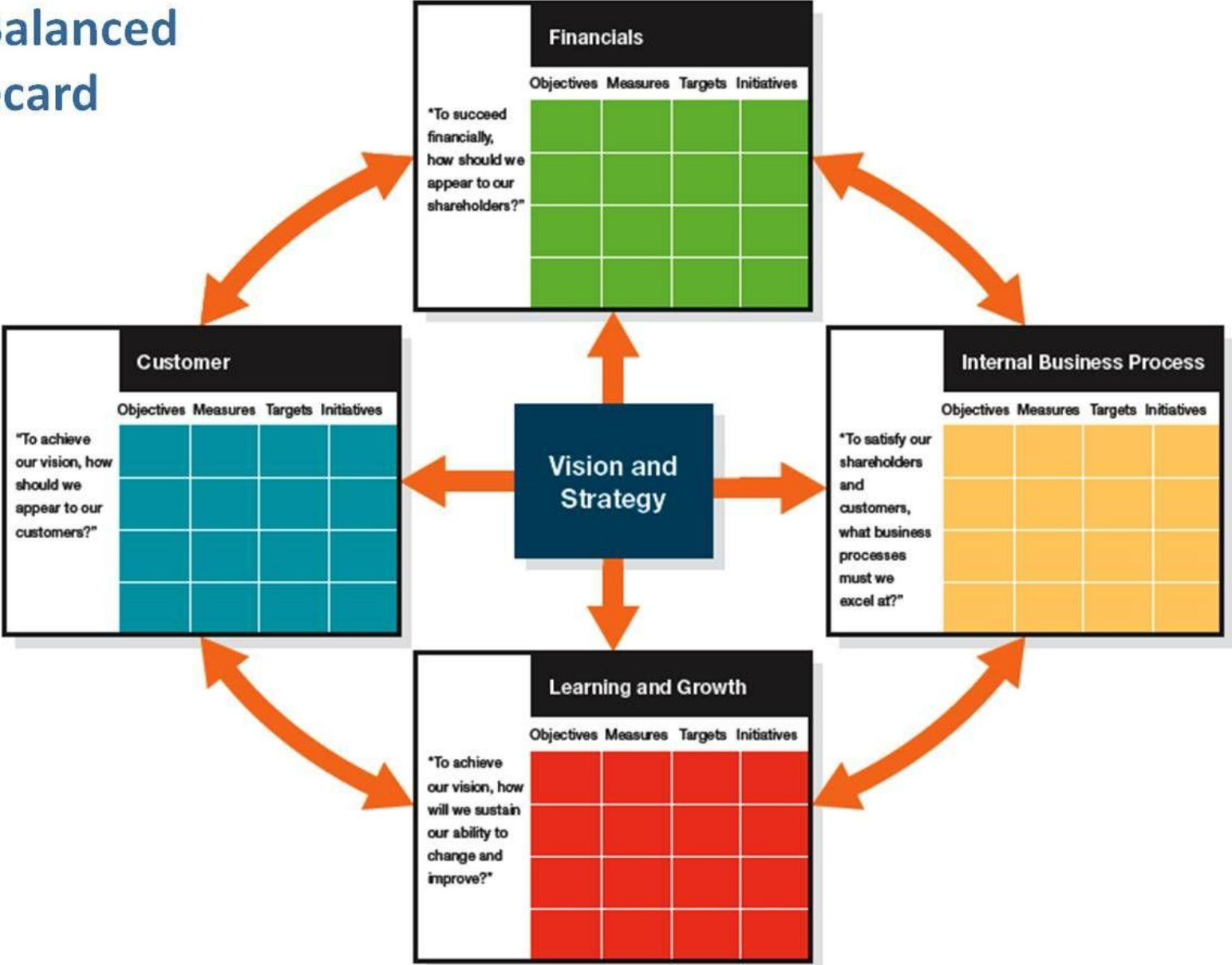
Ad hoc queries; Drill down
Search/OLAP

Business analysts

Forecasts; What if
Analysis; statistical models

Laudon & Laudon (2012)

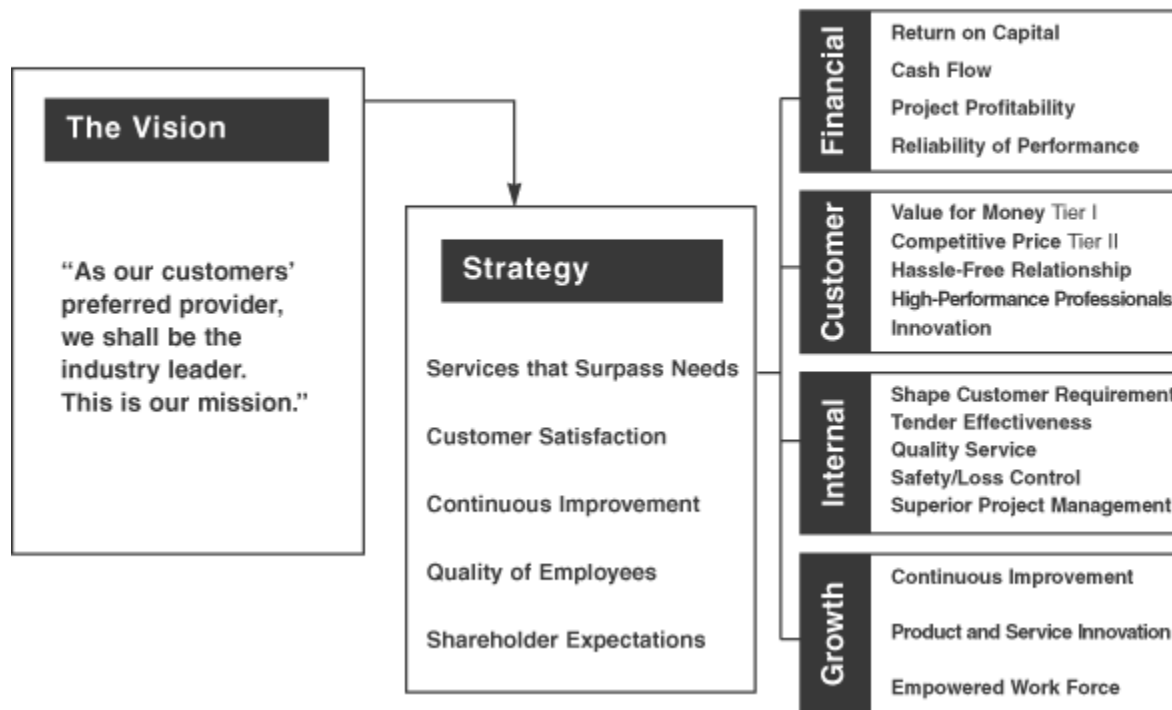
The Balanced Scorecard



Decision Support for Senior Management

The Balanced Scorecard Framework

Rockwater's Strategic Objectives



Kaplan & Norton (1993)

Decision Support for Senior Management



Decision Support Systems



Allows support for semi structured decisions



Use mathematical or analytical models



Allow a variety of analysis:

- “What-if” analysis
- Sensitivity analysis
- Backward sensitivity analysis
- Multidimensional analysis / OLAP (ex: pivot tables)

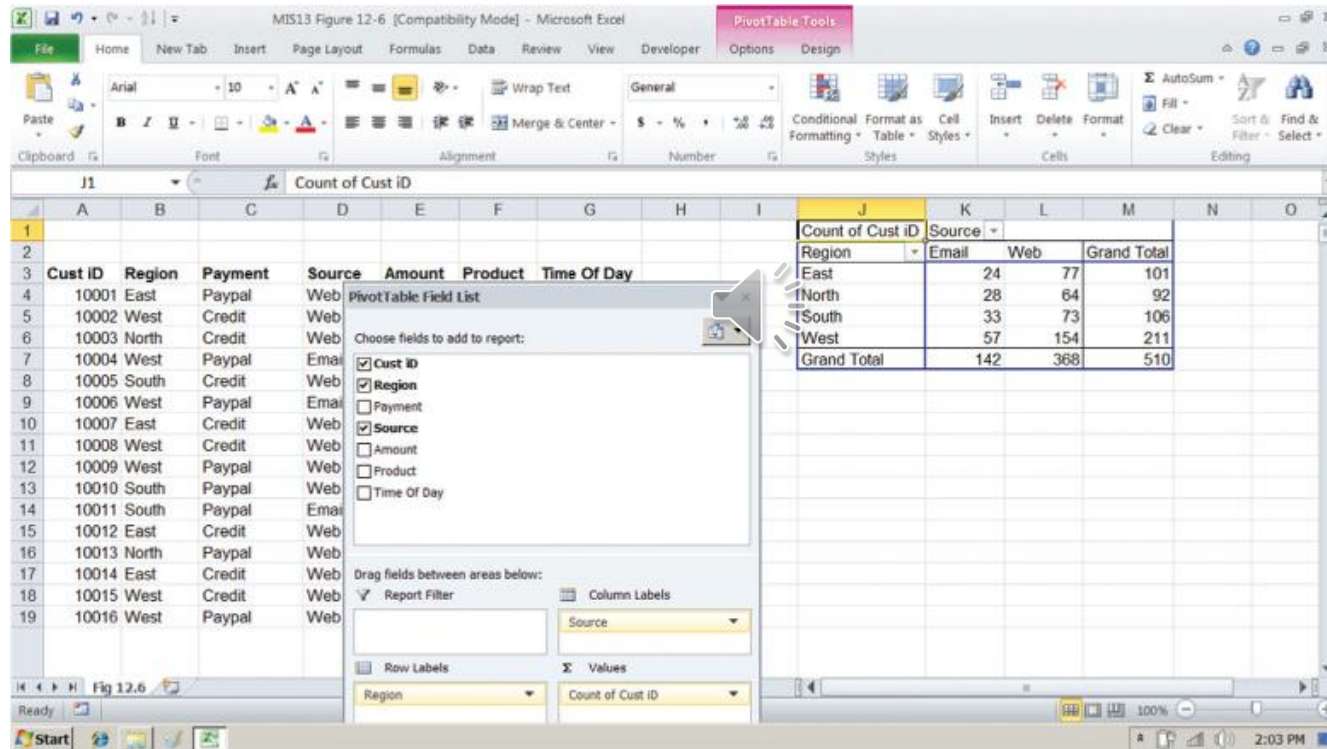
Sensitivity Analysis

Total fixed costs	19000					
Variable cost per unit	3					
Average sales price	17					
Contribution margin	14					
Break-even point	1357					
		Variable Cost per Unit				
Sales	1357	2	3	4	5	6
Price	14	1583	1727	1900	2111	2375
	15	1462	1583	1727	1900	2111
	16	1357	1462	1583	1727	1900
	17	1267	1357	1462	1583	1727
	18	1188	1267	1357	1462	1583

Laudon & Laudon (2012)

A Pivot Table

Examining Customer Regional Distribution and Advertising



Laudon & Laudon (2012)

Group Decision-Support Systems (GDSS)

- Voting
- Brainstorming
- Other techniques





Lisbon School
of Economics
& Management
Universidade de Lisboa



Carlos J. Costa

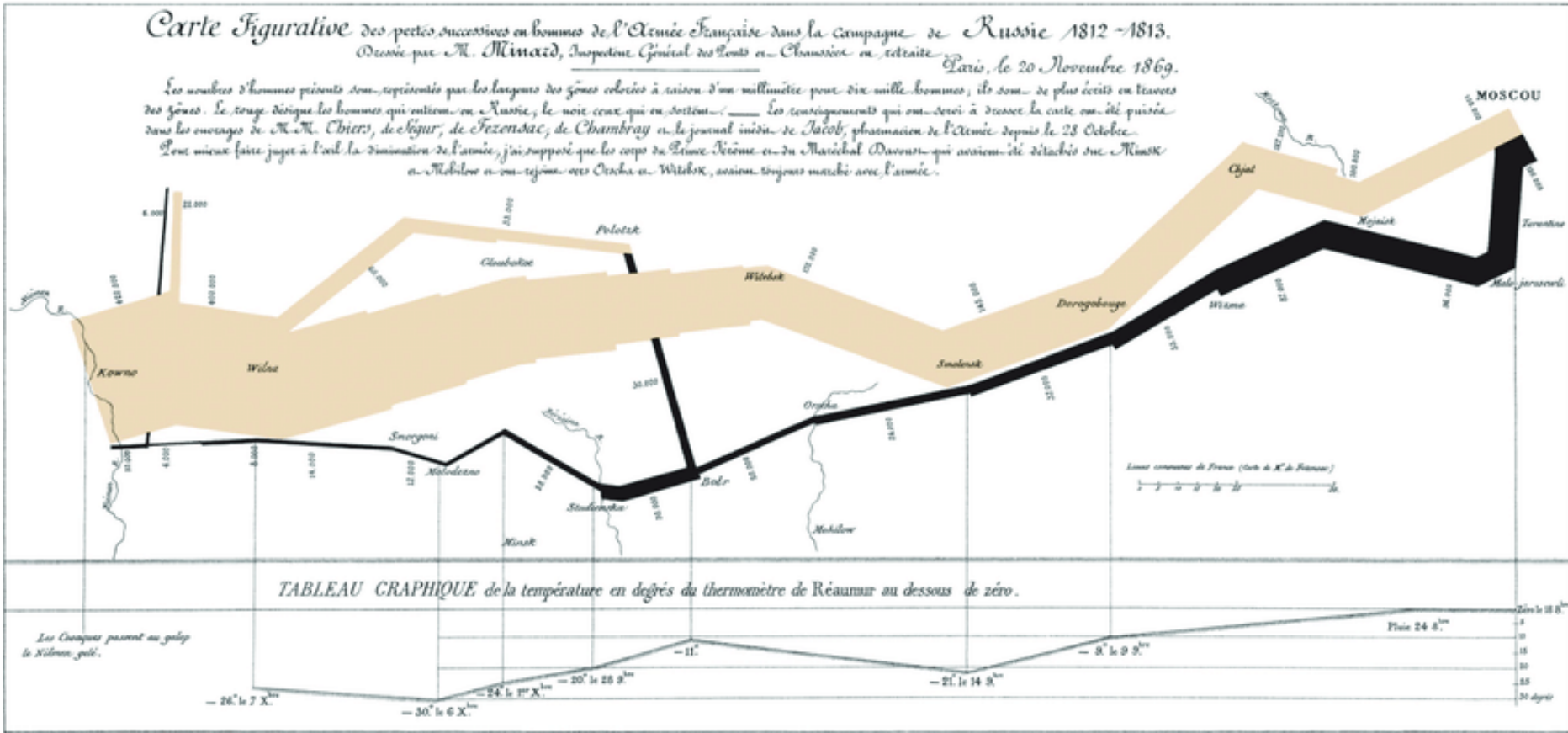
DATA VISUALIZATION

Data visualization

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes perdus sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Legur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que le corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilew n'en rejoignent pas l'armée.



Amis par Bignon, à Paris le 17 Mars 1819.

Imp. Leth. Bignon et Bouché.

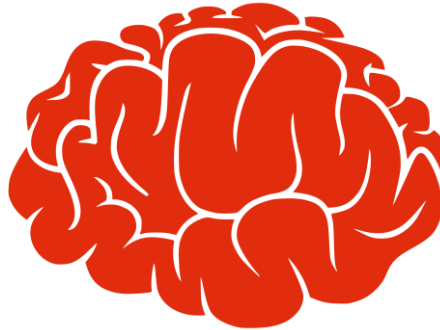
Problems with data visualization



Problems with data visualization



Memory

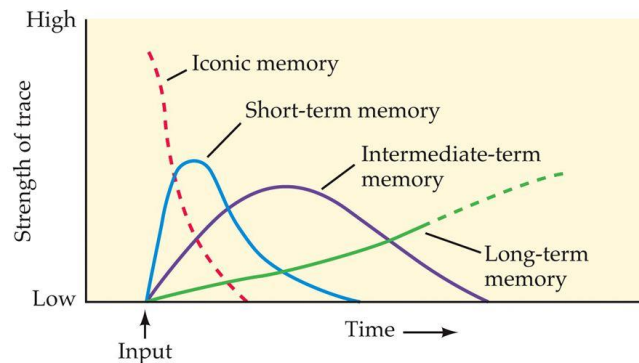


Iconic
memory

Short Term
memory

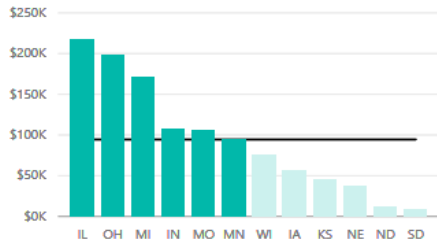
Long Term
memory

Stages of Memory Formation



Gestalt's principles

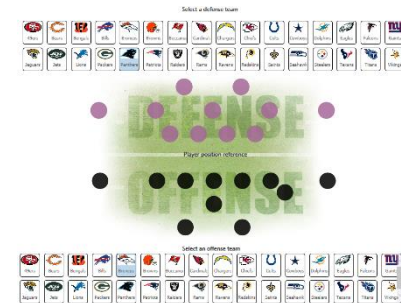
Proximity and Similarity



Continuity



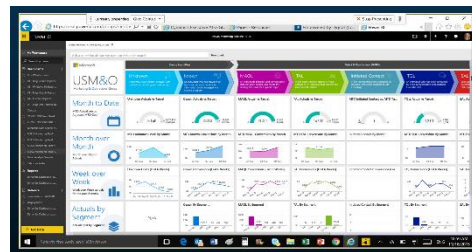
Figure-Ground



Closure and Symmetry



Symmetry and continuity

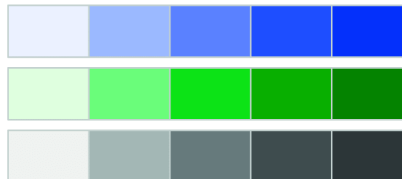


Colour

(A) Qualitative (isoluminant)



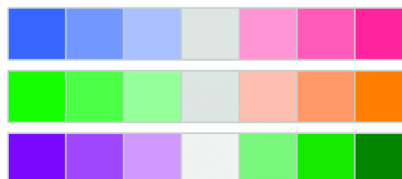
(B) Sequential (single hue)



(C) Sequential (multi hue)



(D) Diverging



Character Fonts

Font Family

serif

Sans serif

Display

Handwriting

Monospace

Text Indent
Line height
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi id magna a lorem sollicitudin fermentum. Pellentesque suscipit ante lorem, bibendum luctus enim imperdiet id. Phasellus finibus nisi lectus, at pharetra libero cursus a. Nulla fringilla elit eu lacus molestie volutpat.

Font Size
18px
T

Letter spacing
Word Spacing
Typography is good for design

Normal

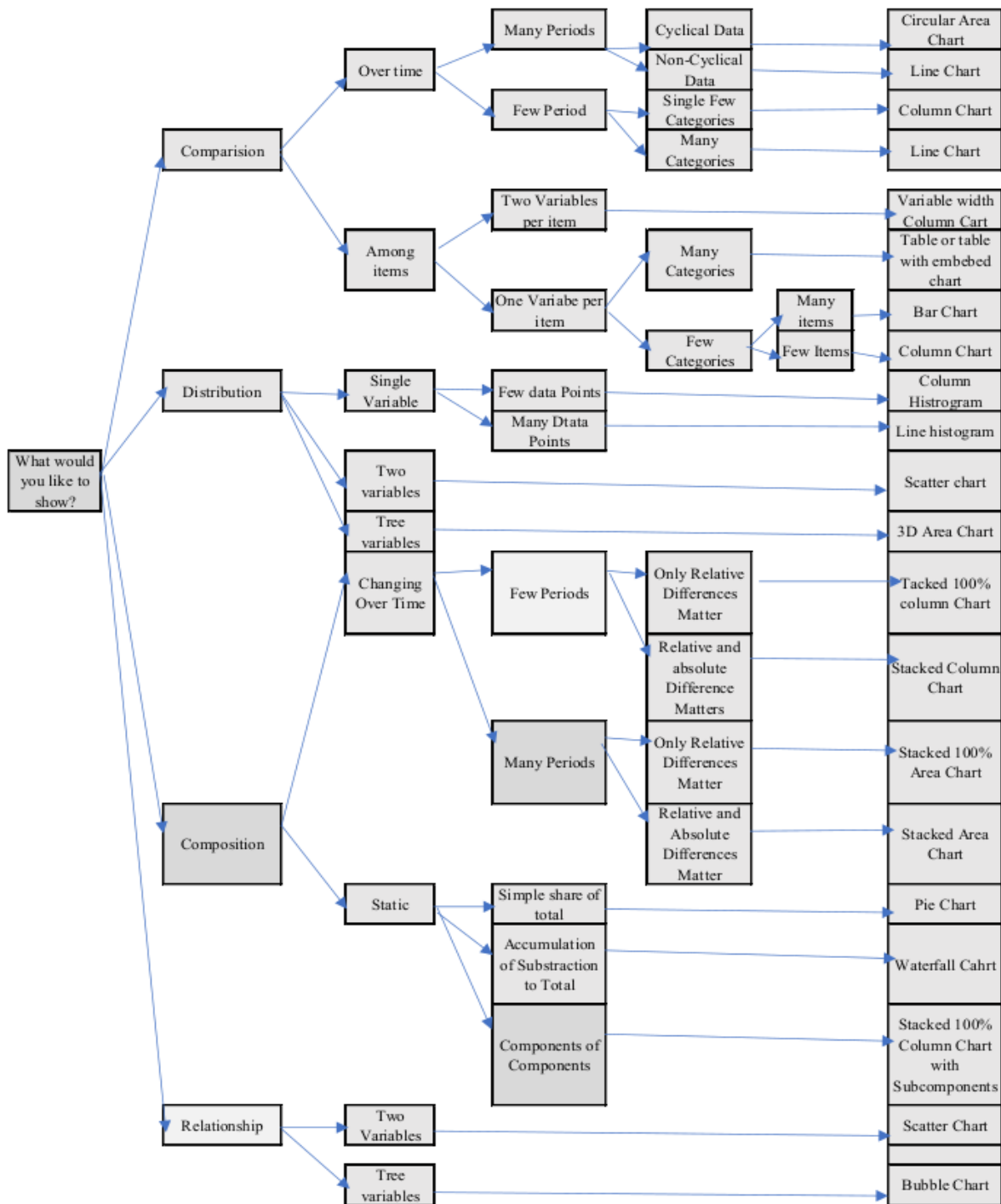
Italic

Oblique

Black Bold Regular Light Thin

Good Dashboard

sqlbi.		The 15 Rules to Design a Perfect Dashboard			http://sql.bi/dashboard				
1	 Design for a target	2	 Keep everything at a glance	3	 Keep it simple	4	 Align elements	5	 Be consistent
6	 Highlight the most relevant information	7	 Be clear	8	 Start from zero	9	1M Shorten the numbers	10	 Show the context
11	 Choose the right colors	12	 Design dashboards, not reports	13	 Show variations	14	 Leave the noise off	15	 Pick the right charts



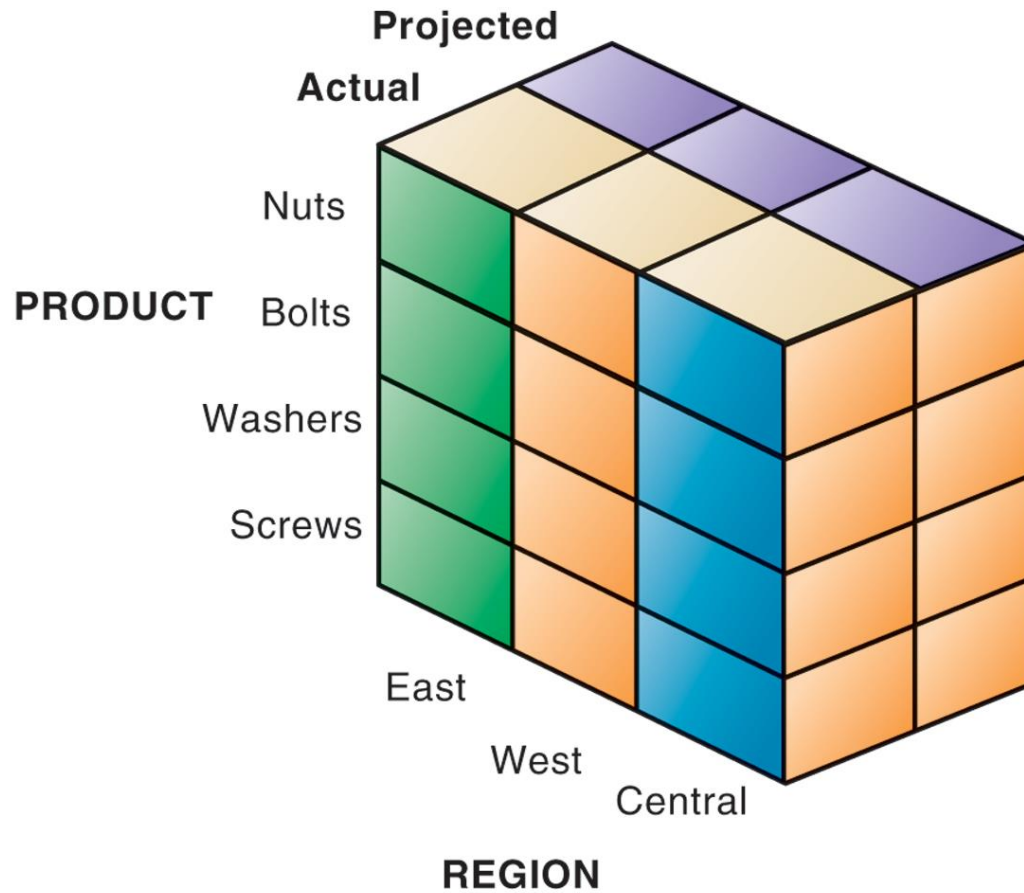


Lisbon School
of Economics
& Management
Universidade de Lisboa



MULTIDIMENSIONAL STRUCTURE

OLAP



Multidimensional Data Model

- a variation of the relational model that uses multidimensional structures to organize data and express the relationships between data

O'Brien & Marakas (2009).

Dimensional Schema



Primary keys



Foreign keys



Fact tables



Dimension Tables



Star schemas



Snowflake schemas

What is a Primary



Foreign Key

Customer

FirstName	LastName	CustID
Elaine	Stevens	101
Mary	Dittman	102
Skip	Stevenson	103
Drew	Lakeman	104
Eva	Plummer	105

Parent Table

Primary Key

One to Many Relationship

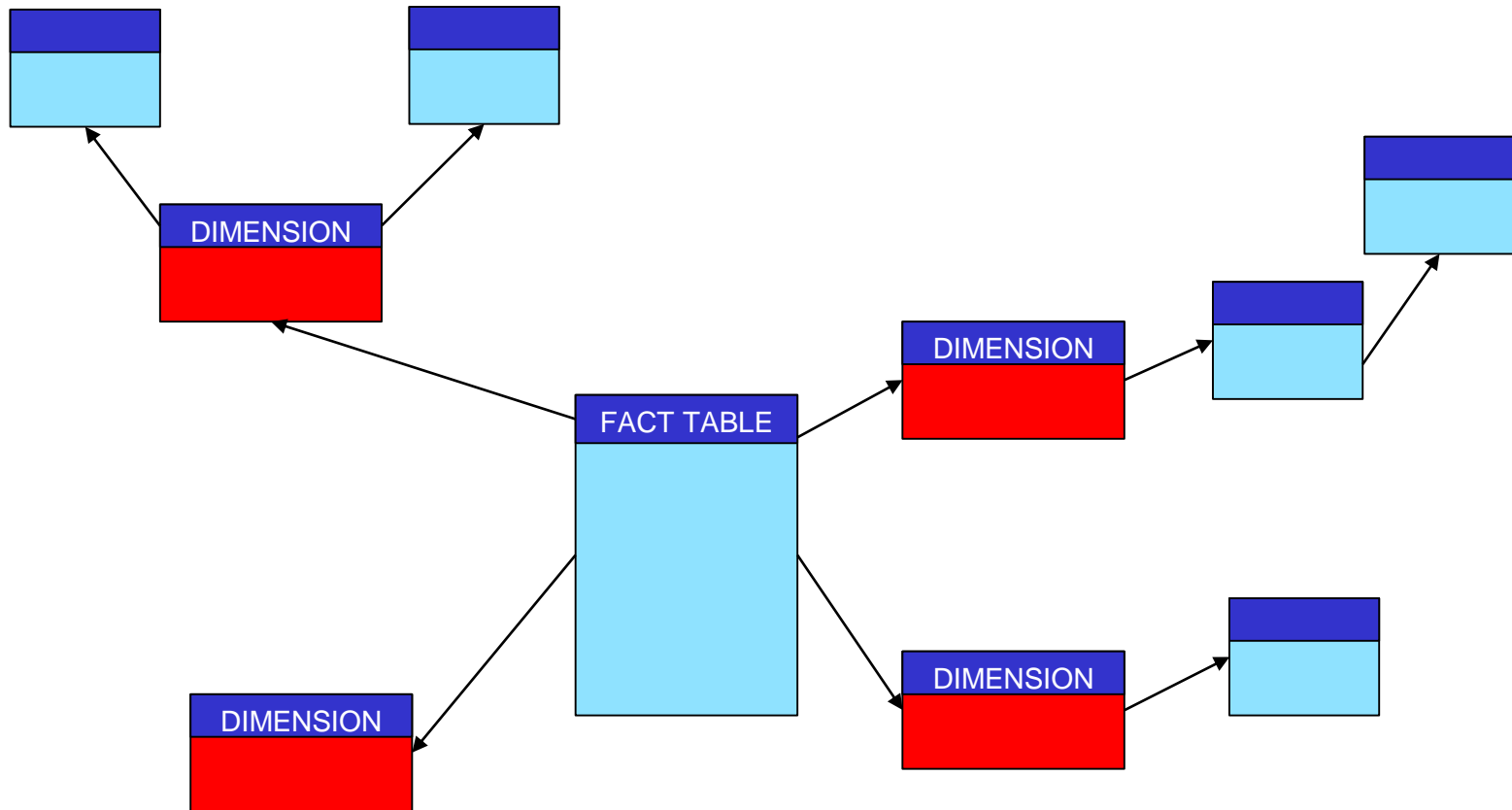
Contact

CustID	ContactInformation	ContactType
101	555-2653	Work
101	555-0057	Cell
102	555-8816	Work
104	555-0949	Work
103	555-0650	Work
101	555-8855	Home
105	Plummer@akcomms.com	Email
101	Stevens@akcomms.com	Email
101	555-5787	Fax
103	Stevenson@akcomms.com	Email
105	555-5675	Work
102	Dittman@akcomms.com	Email

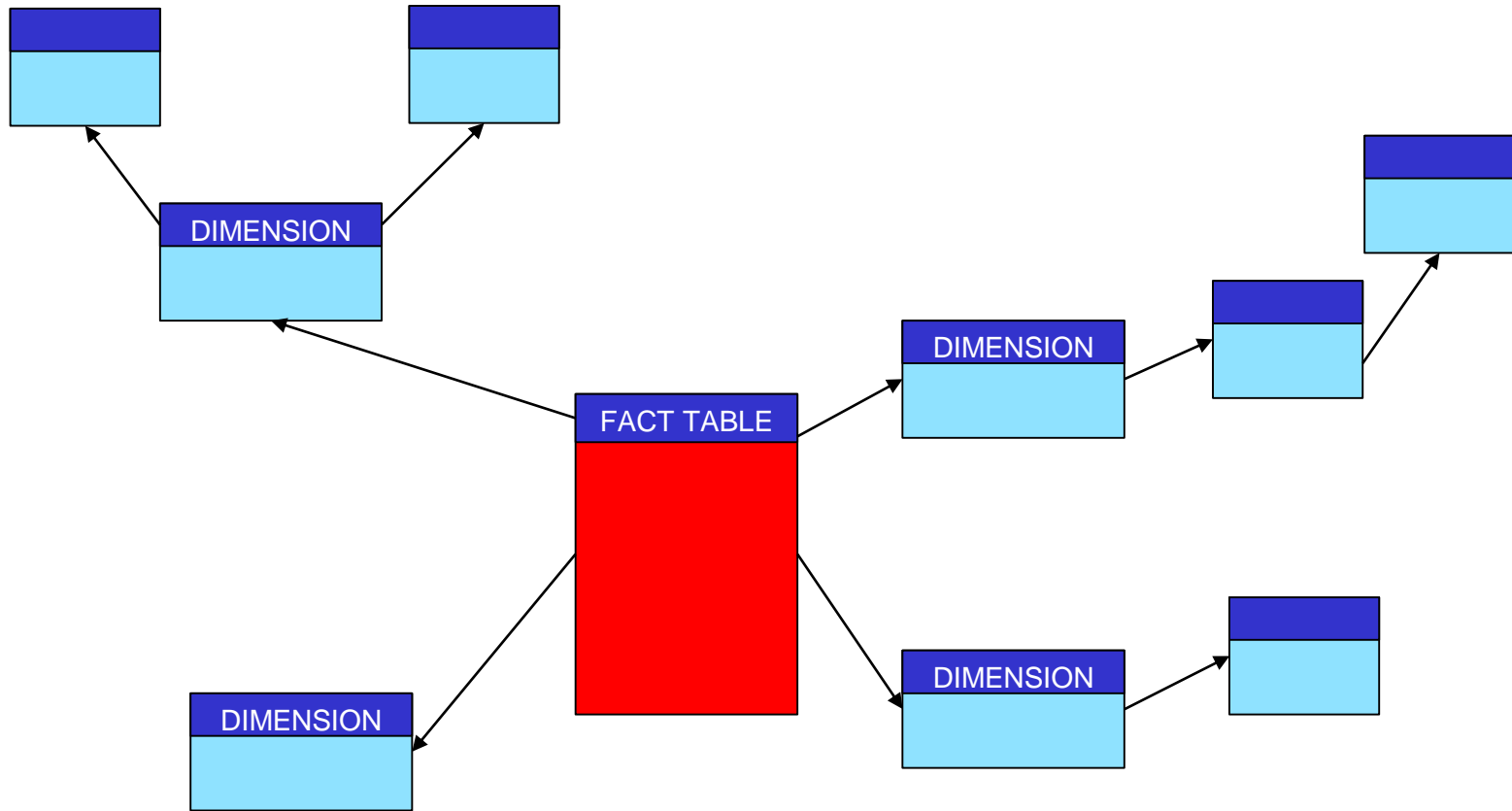
Foreign Key

Child Table

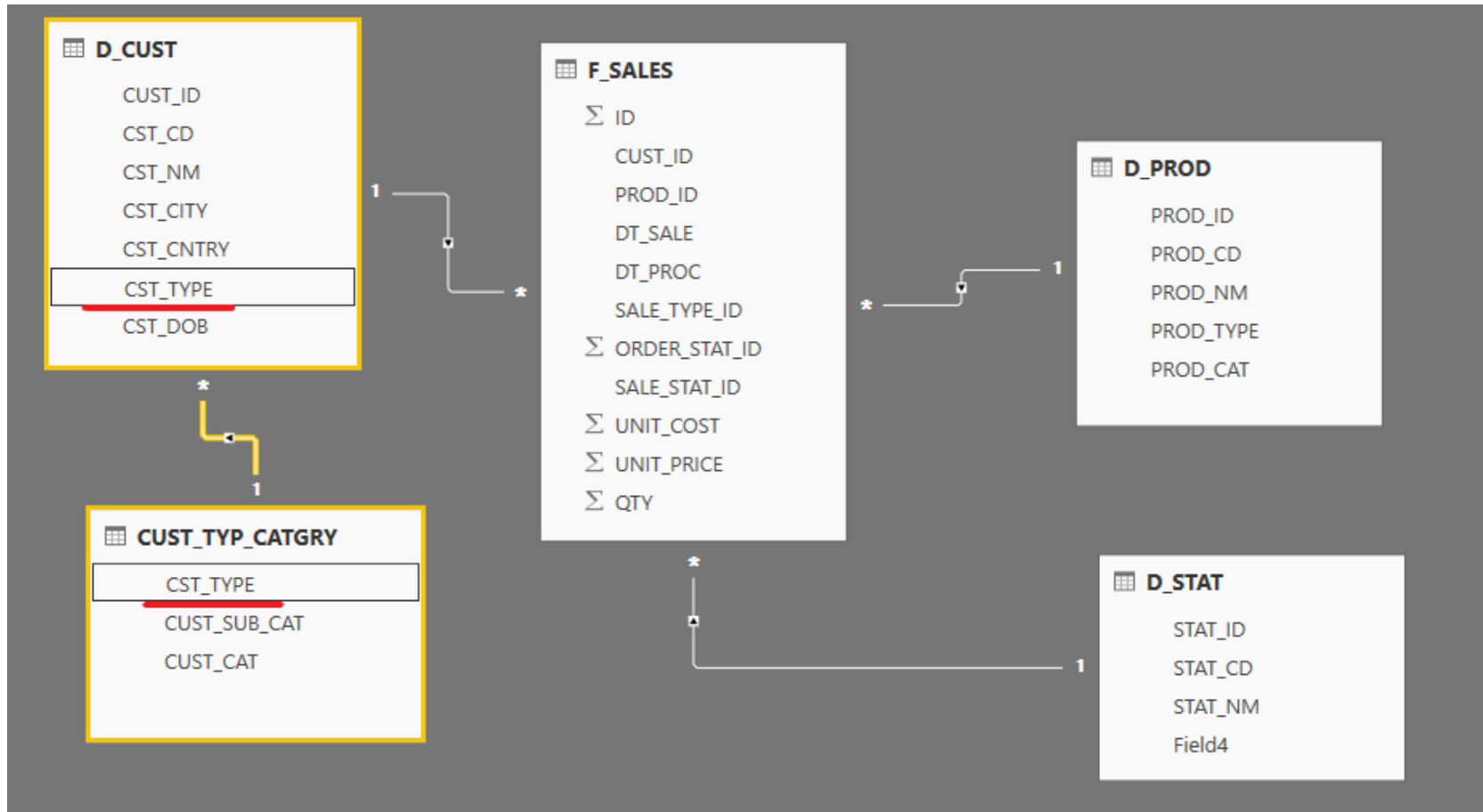
Dimension tables



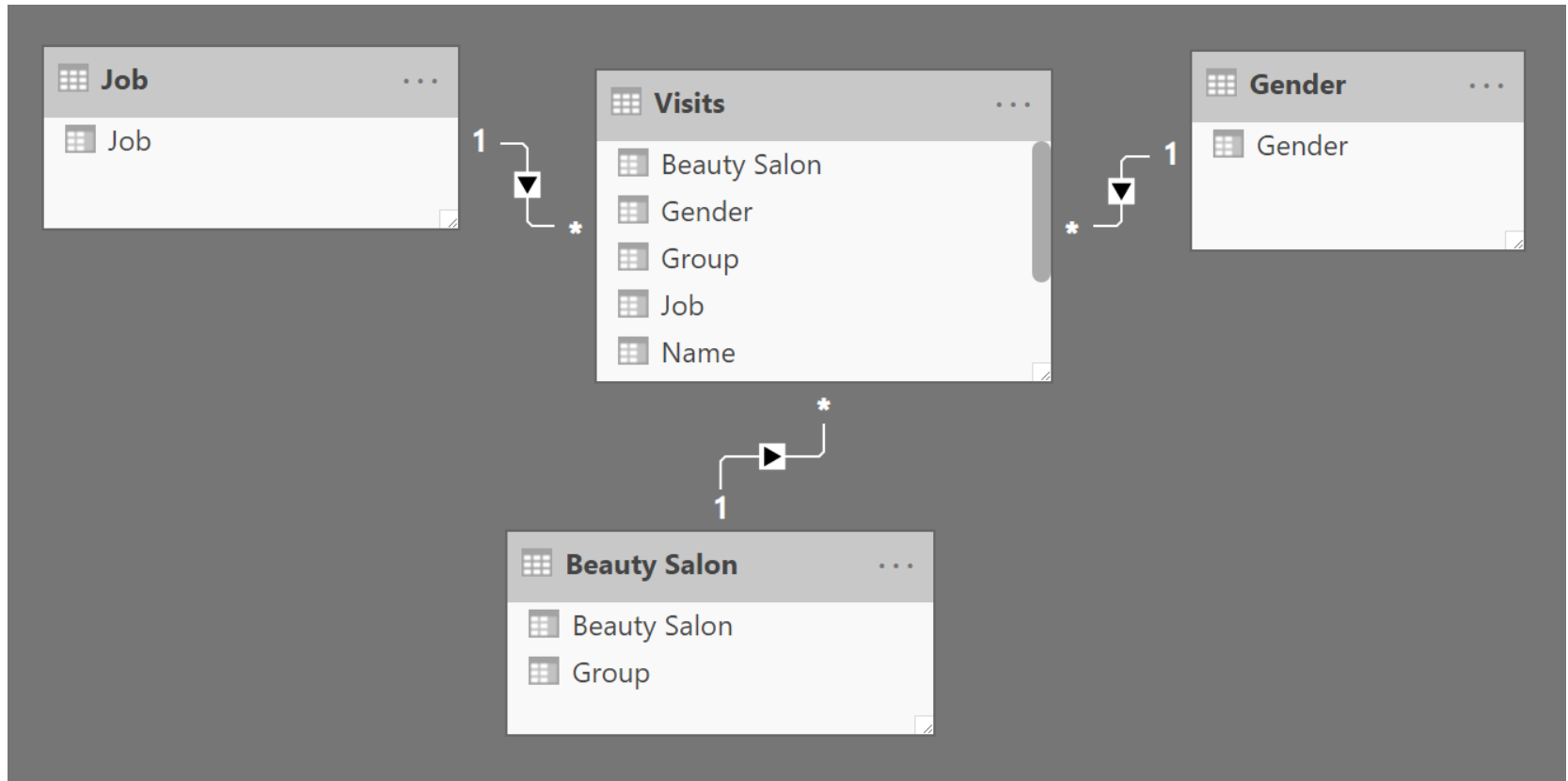
Fact Table

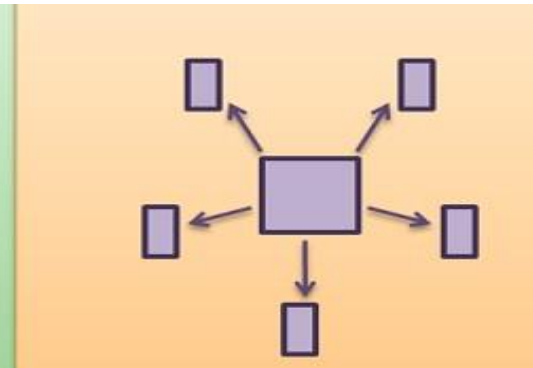
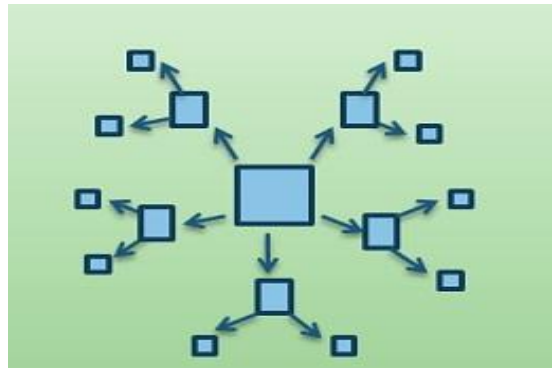


Snowflake schemas



Star schemas





Snowflake Schema

Star Schema

Ease of maintenance / change	No redundancy and hence more easy to maintain and change	Has redundant data and hence less easy to maintain/change
Ease of Use	More complex queries and hence less easy to understand	Less complex queries and easy to understand
Query Performance	More foreign keys-and hence more query execution time	Less no. of foreign keys and hence lesser query execution time
Type of Datawarehouse	Good to use for datawarehouse core to simplify complex relationships (many:many)	Good for datamarts with simple relationships (1:1 or 1:many)
Joins	Higher number of Joins	Fewer Joins
Dimension table	It may have more than one dimension table for each dimension	Contains only single dimension table for each dimension
When to use	When dimension table is relatively big in size, snowflaking is better as it reduces space.	When dimension table contains less number of rows, we can go for Star schema.
Normalization/ De-Normalization	Dimension Tables are in Normalized form but Fact Table is still in De-Normalized form	Both Dimension and Fact Tables are in De-Normalized form
Data model	Bottom up approach	Top down approach