

Quadrilaterale Scientific Conference
Enterprises in hardship: economics, managerial and juridical perspectives
Messina, Italy (September 24-26th, 2014.)

***OUTLIER DETECTION
IN FUNCTION OF
QUALITY IMPROVEMENT
OF BUSINESS DECISIONS***

Authors:

Milan Stamenković

Faculty of Economics, University of Kragujevac, Serbia

Marina Milanović,

Faculty of Economics, University of Kragujevac, Serbia

INTRODUCTION

Intensive development
and use of ICT

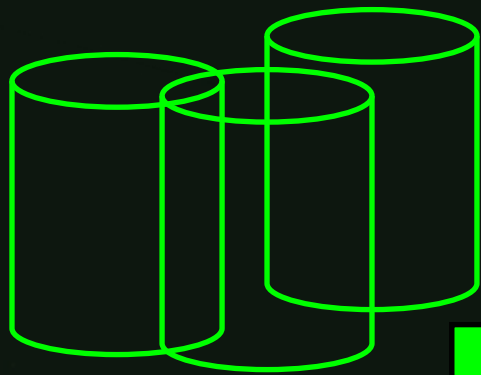
Dynamism and
unpredictability

Knowledge-based economy



**good
information**

**Quality
data**

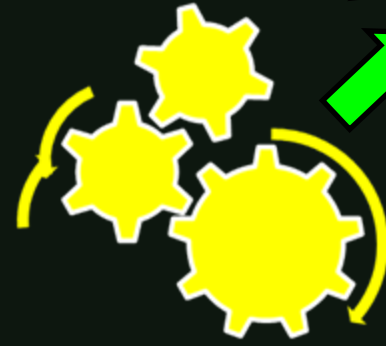


Different sources
of data



Organization's databases

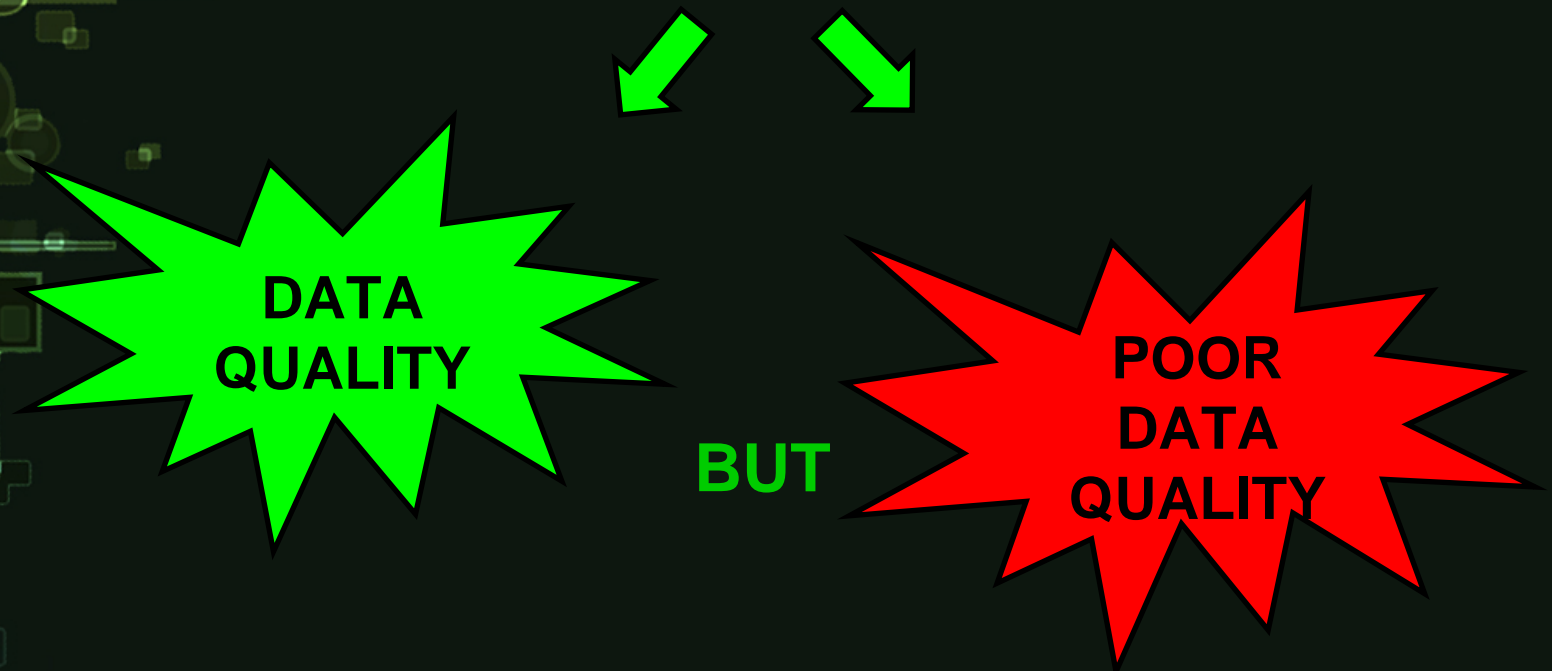
Valuable information



Sophisticated data
analysis methods

WE LIVE IN A DATA AGE!

- Data are everywhere;
- Strategic asset of every organization;
- Basis for a good information;
- And therefore for effective decision making;

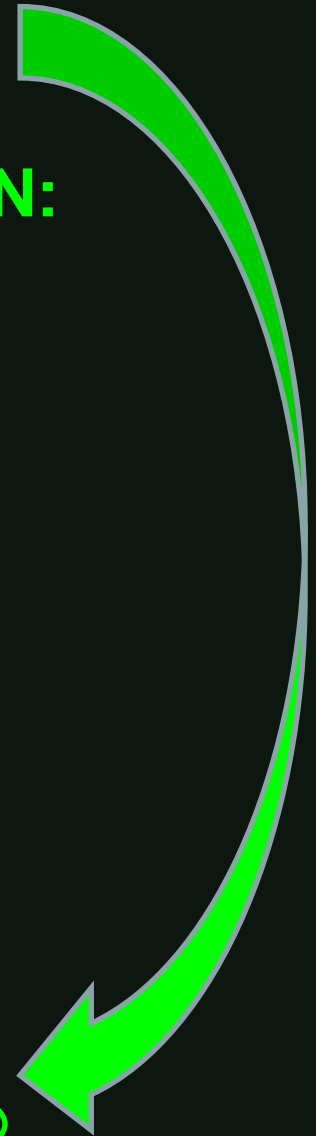


POOR DATA QUALITY

COSTS FOR AN ORGANIZATION:

- **Direct financial costs;**
- **Reduced productivity;**
- **Missed market opportunities;**
- **unsatisfied customers;**
- **Poor strategic planning;**
- **Organization's credibility;**
- **Etc.**

POOR DECISION MAKING



“Ultimately, poor data quality is like dirt on the windshield. You may be able to drive for a long time with slowly degrading vision, but at some point, you either have to stop and clear the windshield or risk everything.” Ken Orr

**WE NEED TO CLEAR OUR WINDSHIELD
WE NEED TO CLEAR OUR DATA**



Preprocessing activities

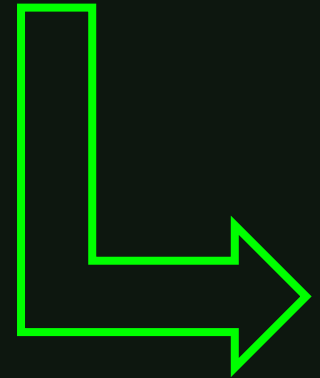
- Data integration;
- Data exploration;
- Data transformation;
- Data reduction;
- DATA CLEANING:
- Missing values;
- OUTLIER DETECTION

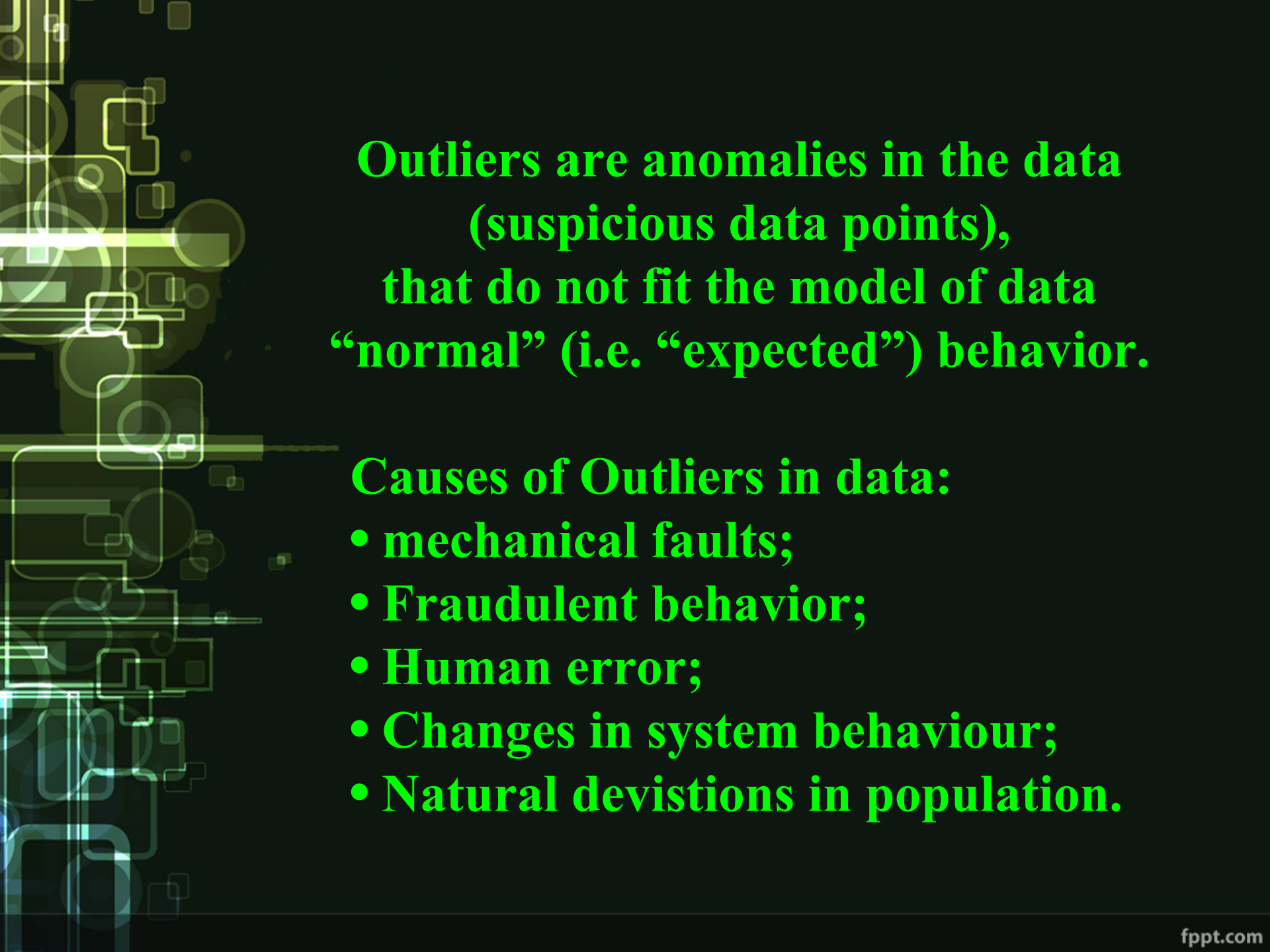


“Dirty” data



“Clear” data
Good data quality



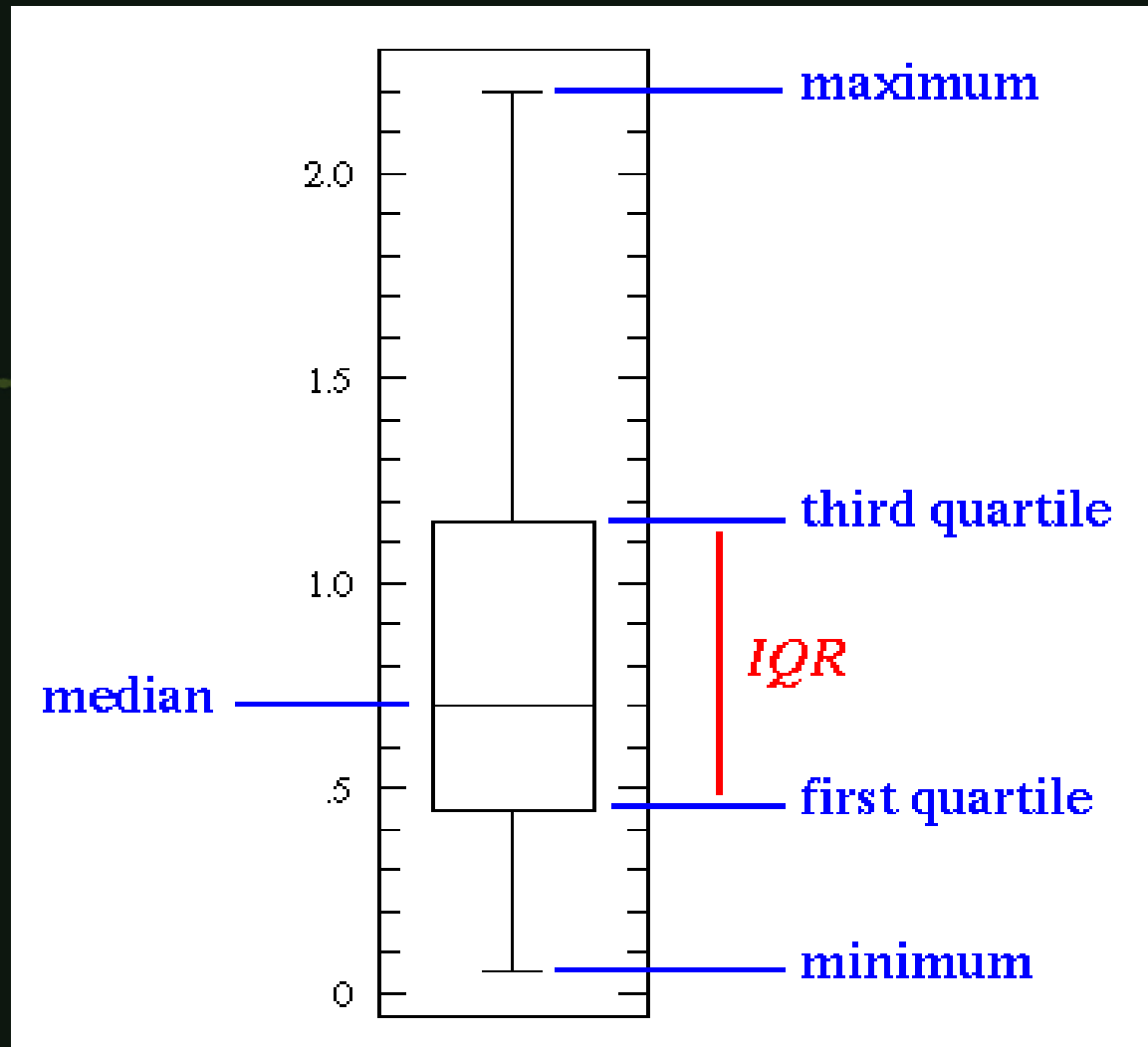


**Outliers are anomalies in the data
(suspicious data points),
that do not fit the model of data
“normal” (i.e. “expected”) behavior.**

Causes of Outliers in data:

- **mechanical faults;**
- **Fraudulent behavior;**
- **Human error;**
- **Changes in system behaviour;**
- **Natural deviations in population.**

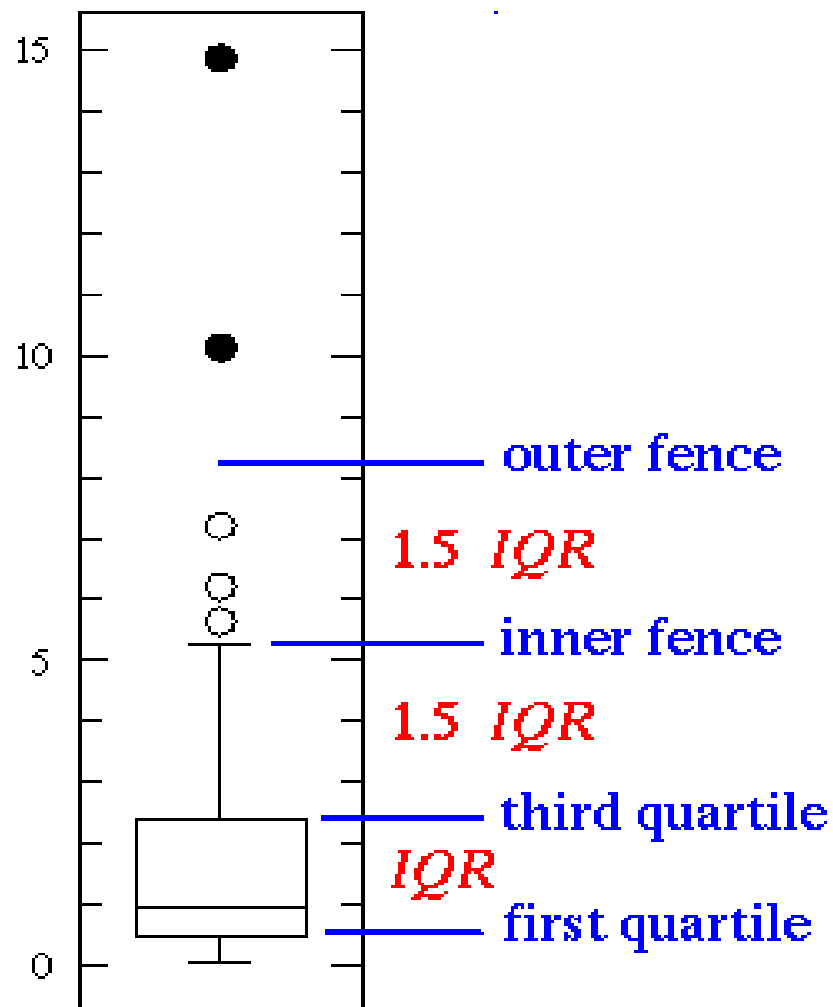
Box-plot with whiskers from min to max



Box-plot and outlier detection

outliers

suspected
outliers



Grubbs' test for outlier detection

H_0 : There are no outliers in data set

H_1 : There is at least one outlier in data set

**Grubbs'
test statistic**

$$G = \frac{\max |x_i - \bar{x}|}{s}$$

\bar{x} - sample mean,

s - standard deviation of all data, and

x_i - suspected single outlier

Grubbs' critical values table

N	G_{critical} (95%)**	G_{critical} (97.5%)**	G_{critical} (99%)**
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82



business application fields
of outlier detection analysis:

- Structural defect detection;
- Loan application processing;
- Fraud detection;
- Intrusion detection;
- Activity monitoring;
- Detection of unexpected entries in databases;
- etc.

The background features a complex, abstract digital graphic on the left side, consisting of glowing yellow and blue lines, squares, and circles. The rest of the background is a solid dark color. The text is centered in a bold, yellow, serif font.

**THANK YOU
FOR
YOUR
ATTENTION**