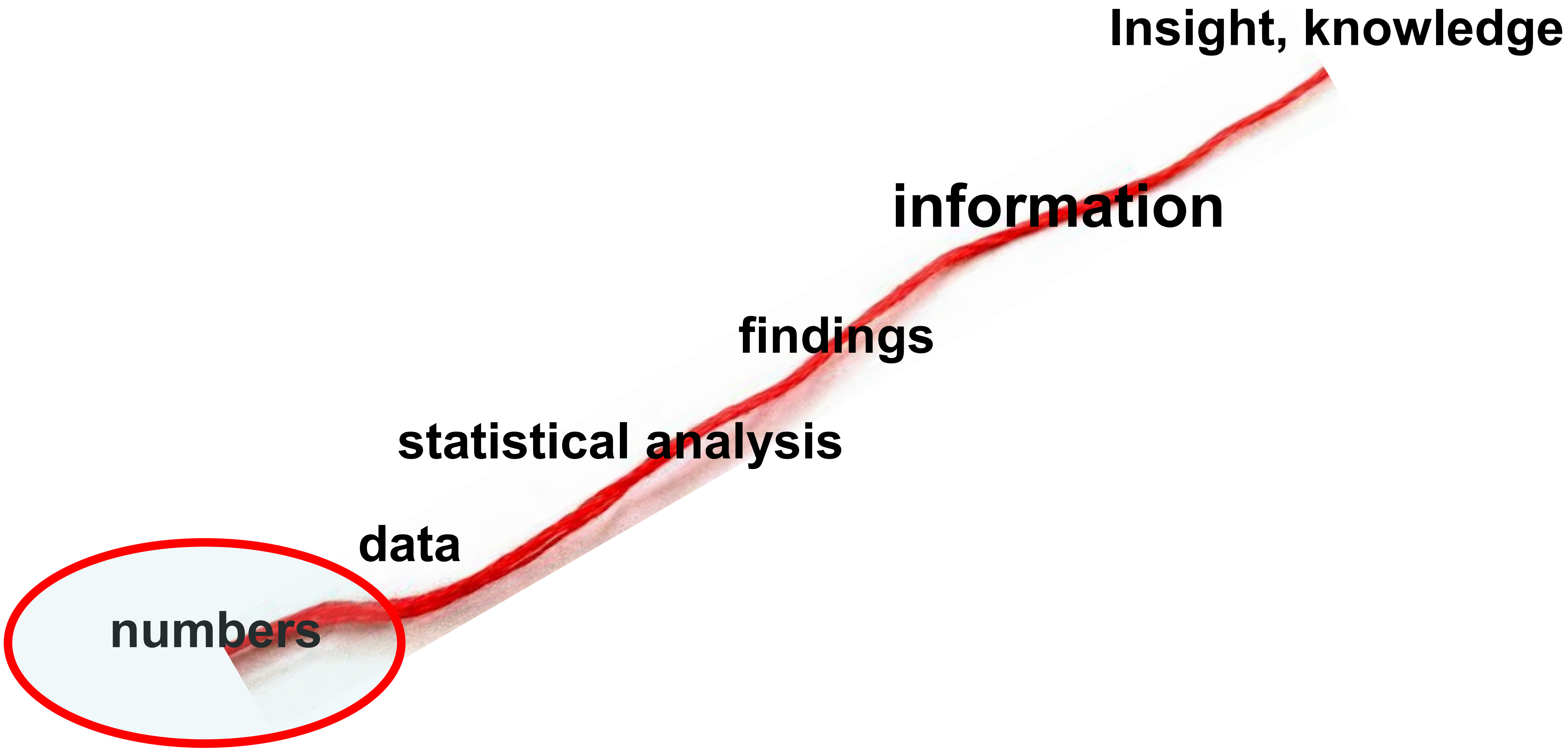
A Biomed Data Analyst Training Program

Data types and data integration

Professor Ron S. Kenett


Insight, knowledge
information
findings
statistical analysis
data
numbers

http://moovitapp.com

moovit

https://moodle.com

moodle

www.netflix.com

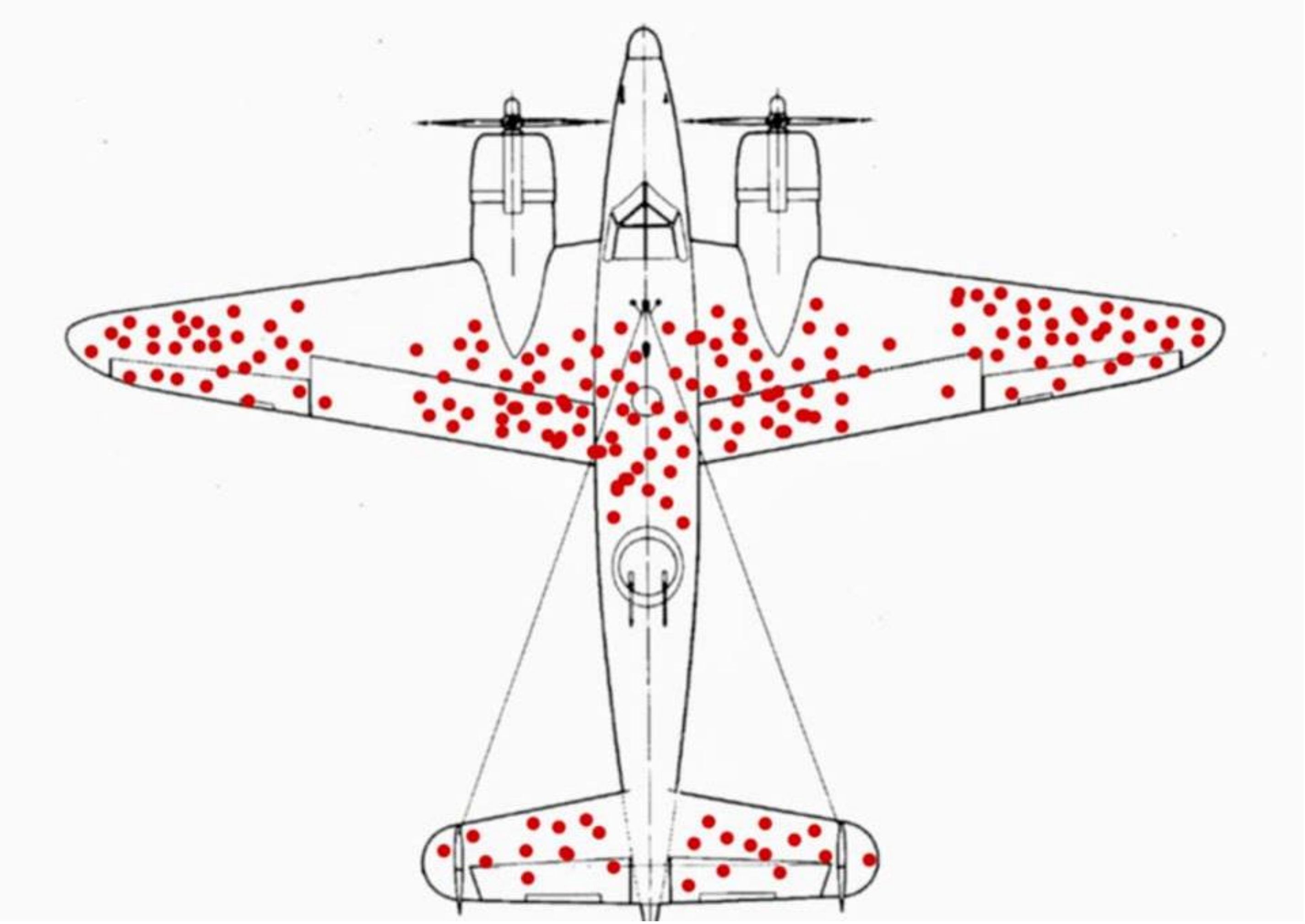
NETFLIX

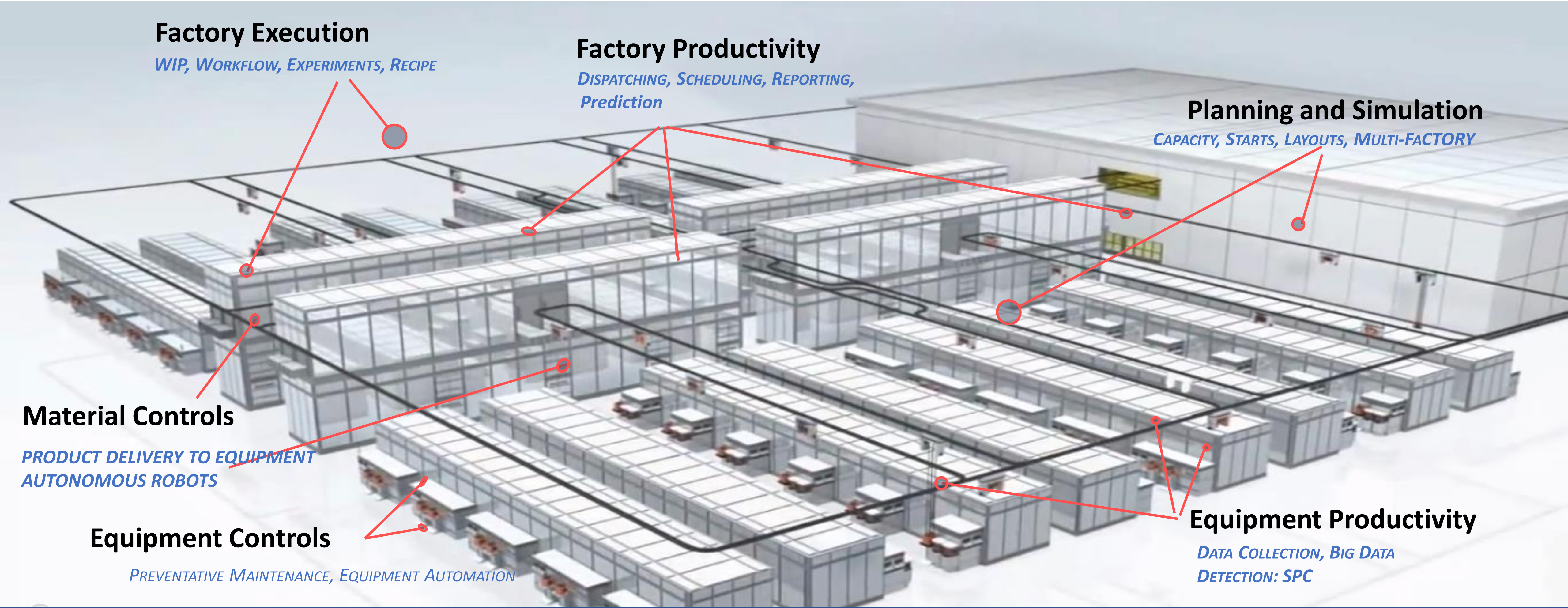
http://waze.com

www.google.com/maps

https://tinder.com/

tinder





Factory Execution

Factory Productivity

Planning and Simulation

Material Controls

Equipment Productivity

Equipment Controls

Data and the Fourth Industrial Revolution

Ron S. Kenett and **Shirley Y. Coleman** outline the roles played by data and statistics in "Industry 4.0", from monitoring manufacturing processes to the building of "digital twins"

The word "manufacturing" conjures images of galleries of machines running day and night, maybe with rows of workers adjusting or sifting and sorting. What is missing from these mental images, though, are the sensors embedded in each of those machines, collecting data continuously on different aspects of production, transmitting that data to analytics computer packages, and – at the end of it all – a statistician monitoring the outputs in an effort to understand what is going on and to make sure things are working at their very best.

There is a whole world of data analytics based on statistics and artificial intelligence going on behind the scenes in manufacturing plants all over the world. In this article we aim to give some insight into what goes on and why it is important.

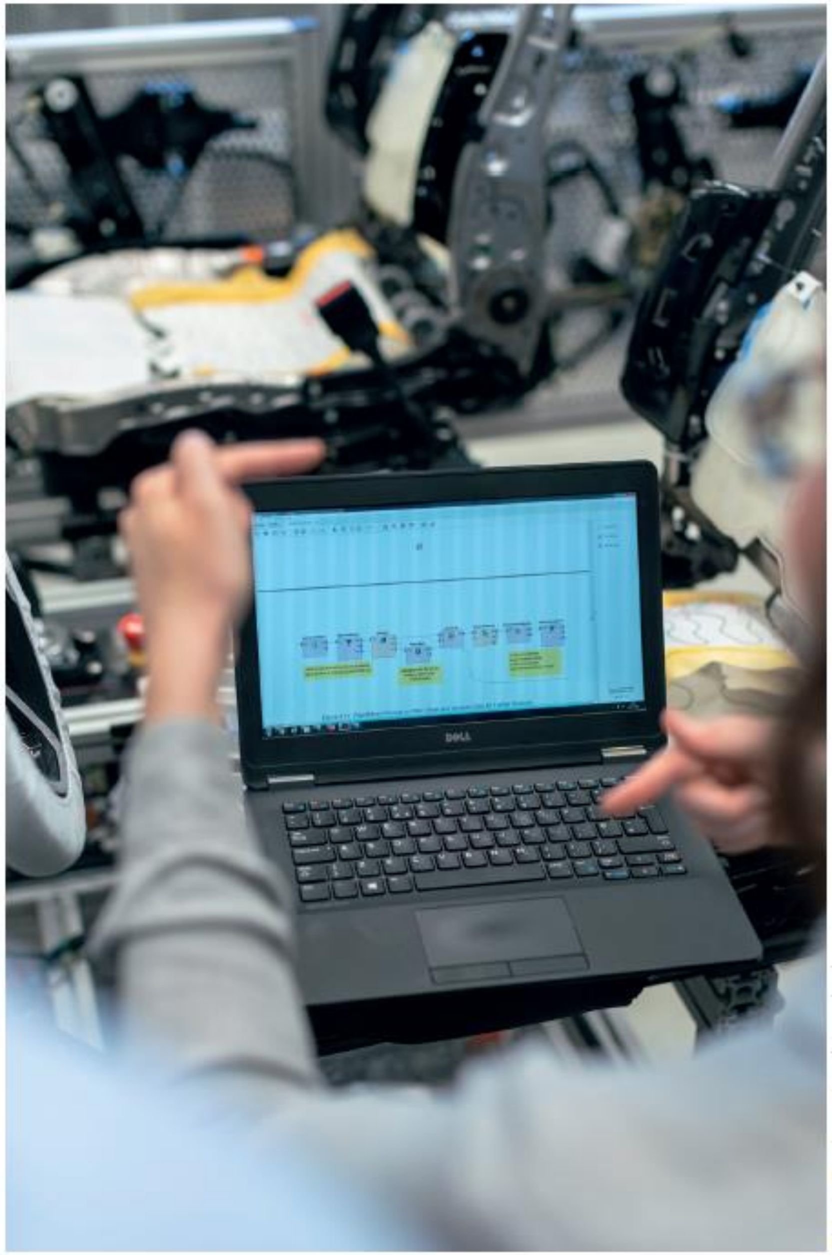
Monitoring and adjusting

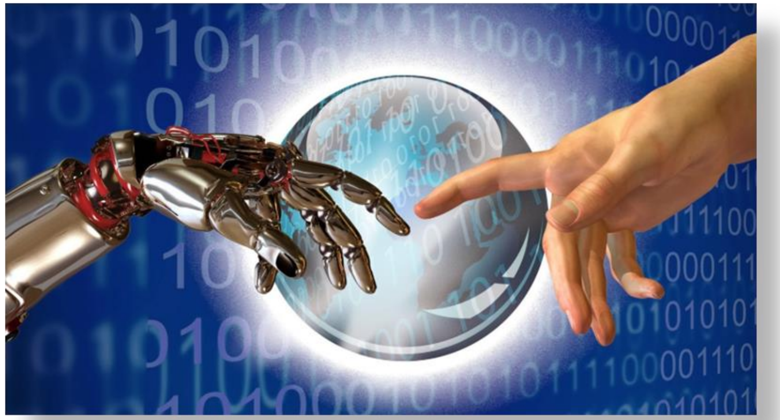
During the last decade, companies

continuous measurements such as temperature, flow rate, colour and purity between different parts of the production process.

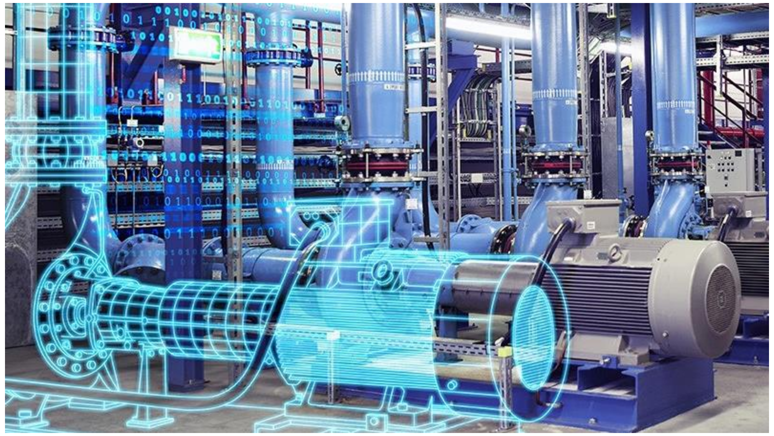
2. *Flexible manufacturing* capabilities – such as 3D printing – that can efficiently produce batches of products to order.
3. *Data analytics*, including statistical analysis, machine learning and artificial intelligence that powers industry with the capability to control and optimise processes.

Consider a hypothetical scenario. You are the operations manager of a company manufacturing medical devices. The Covid-19 pandemic creates a worldwide shortage of ventilators and your company has recently transformed itself to meet Industry 4.0 standards. As a result, you can predict operating



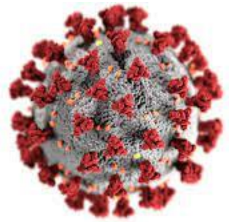


Digital Twin



BMC Part of Springer Nature https://academic.oup.com/pnasnexus/article/1/3/pgac125/6673789

Israel Journal of Health Policy Research



Home About Articles Submission Guidelines

https://academic.oup.com/pnasnexus/article/1/3/pgac125/6673789

Perspective | Open Access | Published: 20 April 2022

The role of statisticians in the response to COVID-19 in Israel: a holistic point of view

Itai Dattner, Reuven Gal, Yair Goldberg, Inbal Goldshtein, Amit Huppert, Ron S. Kenett, Orly Manor, Danny Pfeffermann, Edna Schechtman, Clelia di Serio & David M. Steinberg

https://rdcu.be/cLMKZ

https://ijhpr.biomedcentral.com/articles/10.1186/s13584-022-00531-y

Qualitative Data

- Measurements that do not exist on any naturally occurring numerical scale; they are classified into categories.

- Nominal

Nominal data is categorical data that has no order or ranking.
Examples include: eye color, gender, country of origin, etc.
Nominal data can be represented by numbers, but the numbers do not have any meaning beyond labeling categories.

- Ordinal

Ordinal data is categorical data that has a natural order or ranking.
Examples include: education level, income level, rating scales (e.g., 1-5), etc.
Ordinal data can be represented by numbers, and the numbers have meaning in terms of their rank, but the intervals between them are not necessarily equal

Interval and Ratio Data

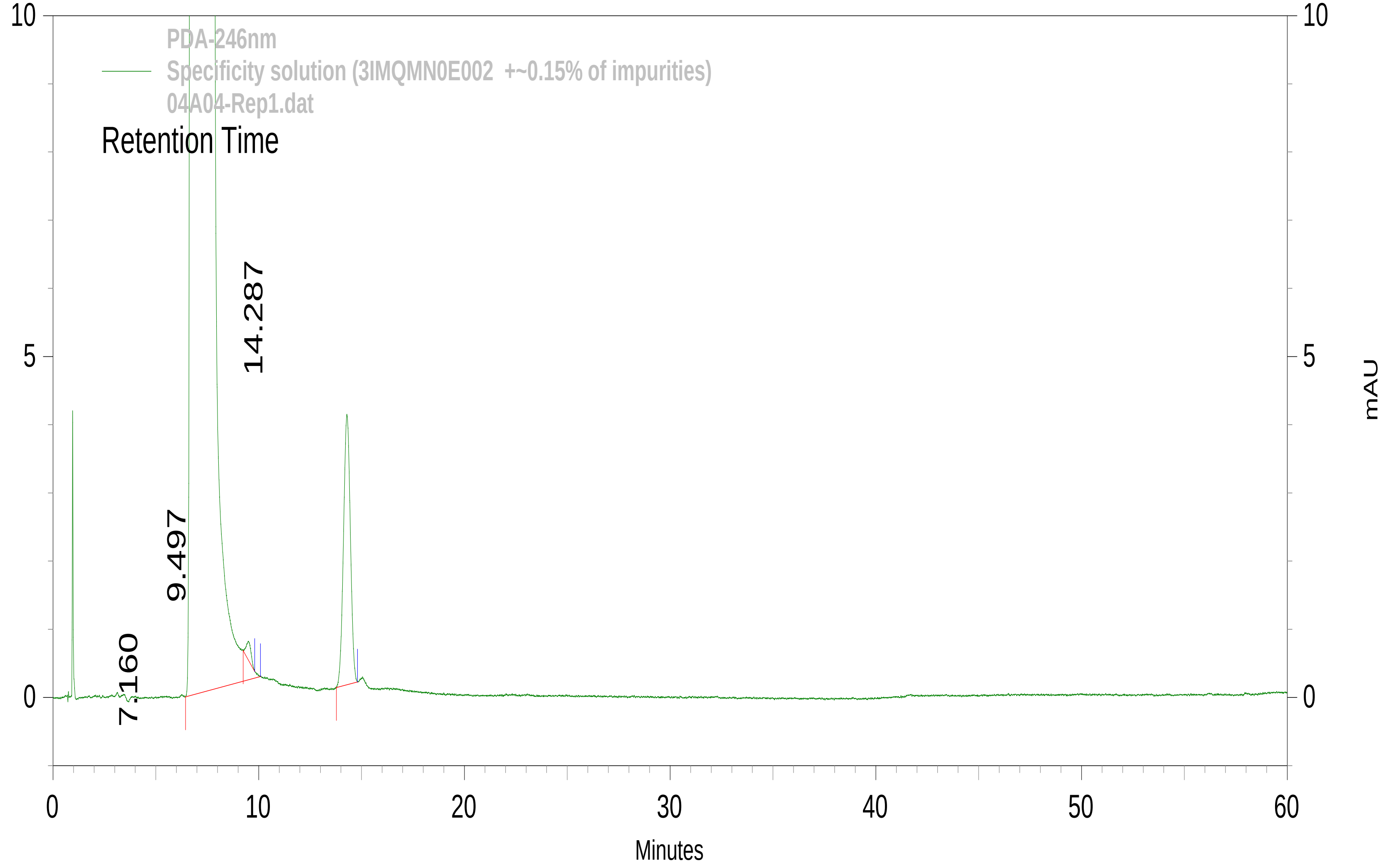
Interval data is numerical data that has equal intervals between values, but no true zero point. Examples include: temperature (in Celsius or Fahrenheit), years (e.g., 1950, 2000, etc.), etc.

Interval data can be added, subtracted, and averaged, but it doesn't make sense to multiply or divide them.

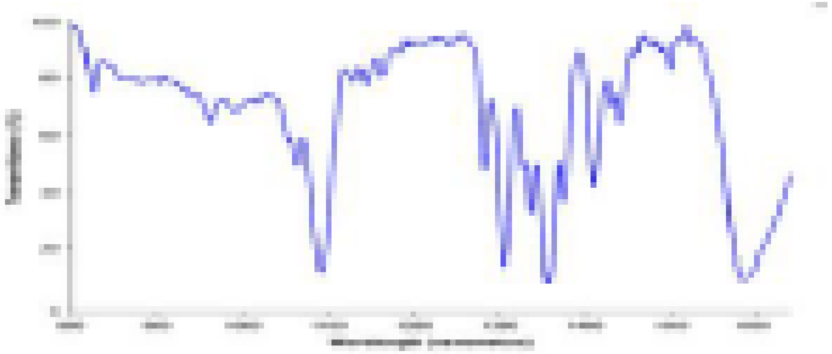
Ratio data is numerical data that has equal intervals between values and a true zero point. Examples include: height, weight, distance, age, etc.

Ratio data can be added, subtracted, multiplied, and divided

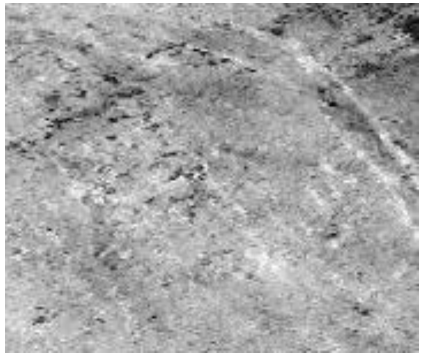
Quantitative data



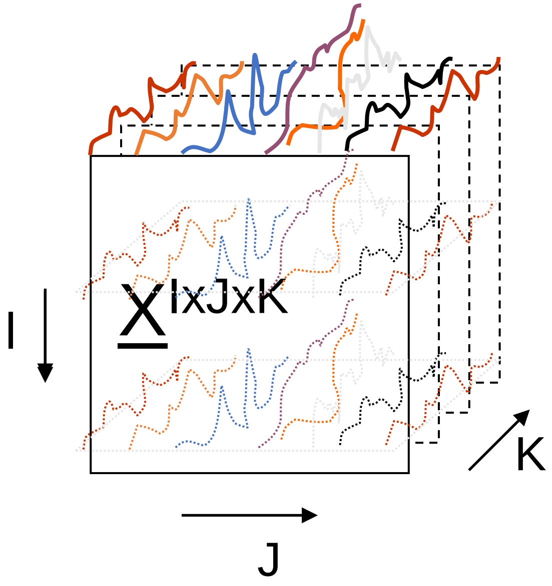
1D



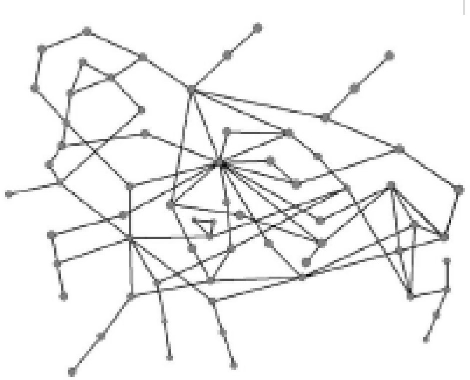
2D

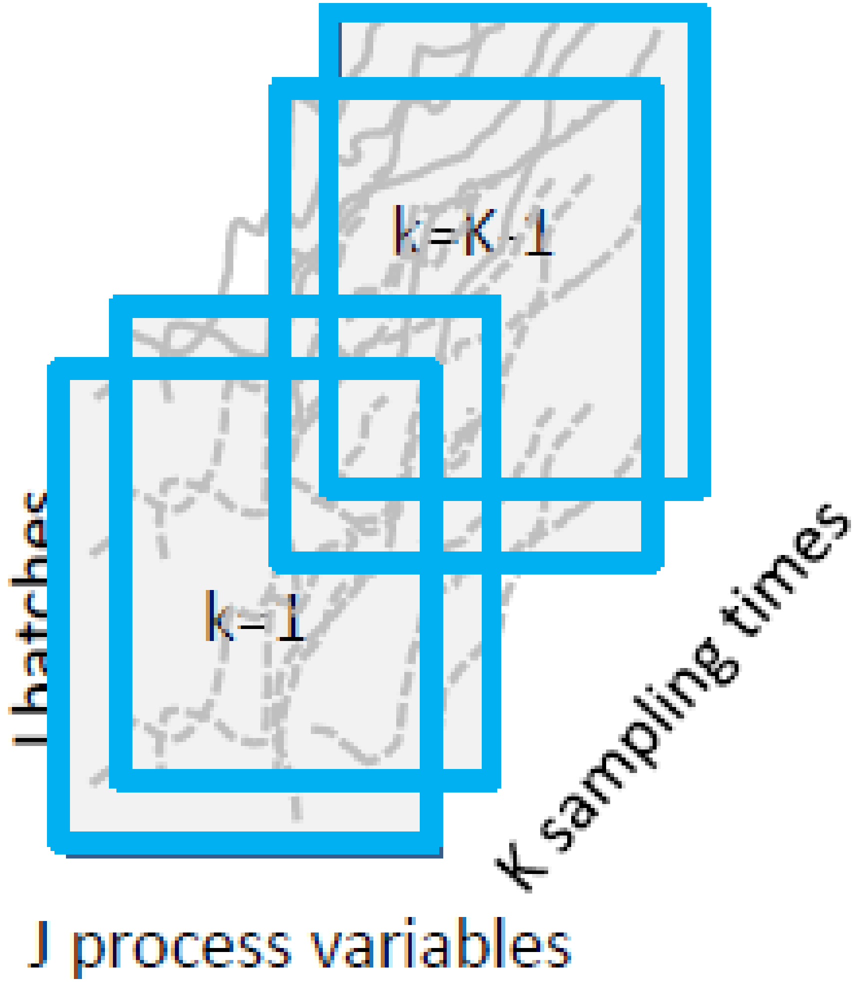


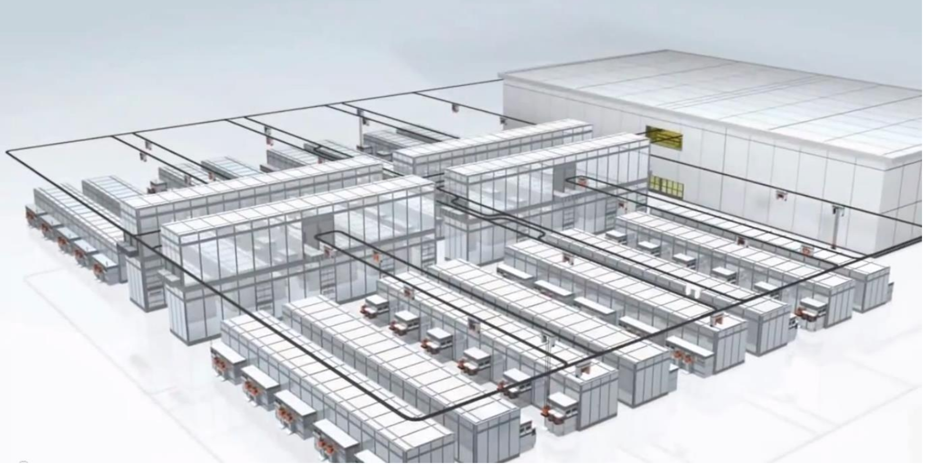
3D



Unstructured







- Maintenance schedule, etc…
  - Day 1
    - Oper8002
    - Oper9050
    - Oper9070
  - Day 2
    - Oper8002
    - Oper9050
    - Oper9070

The data: v7022, v7023, v7024

Variables

Operations

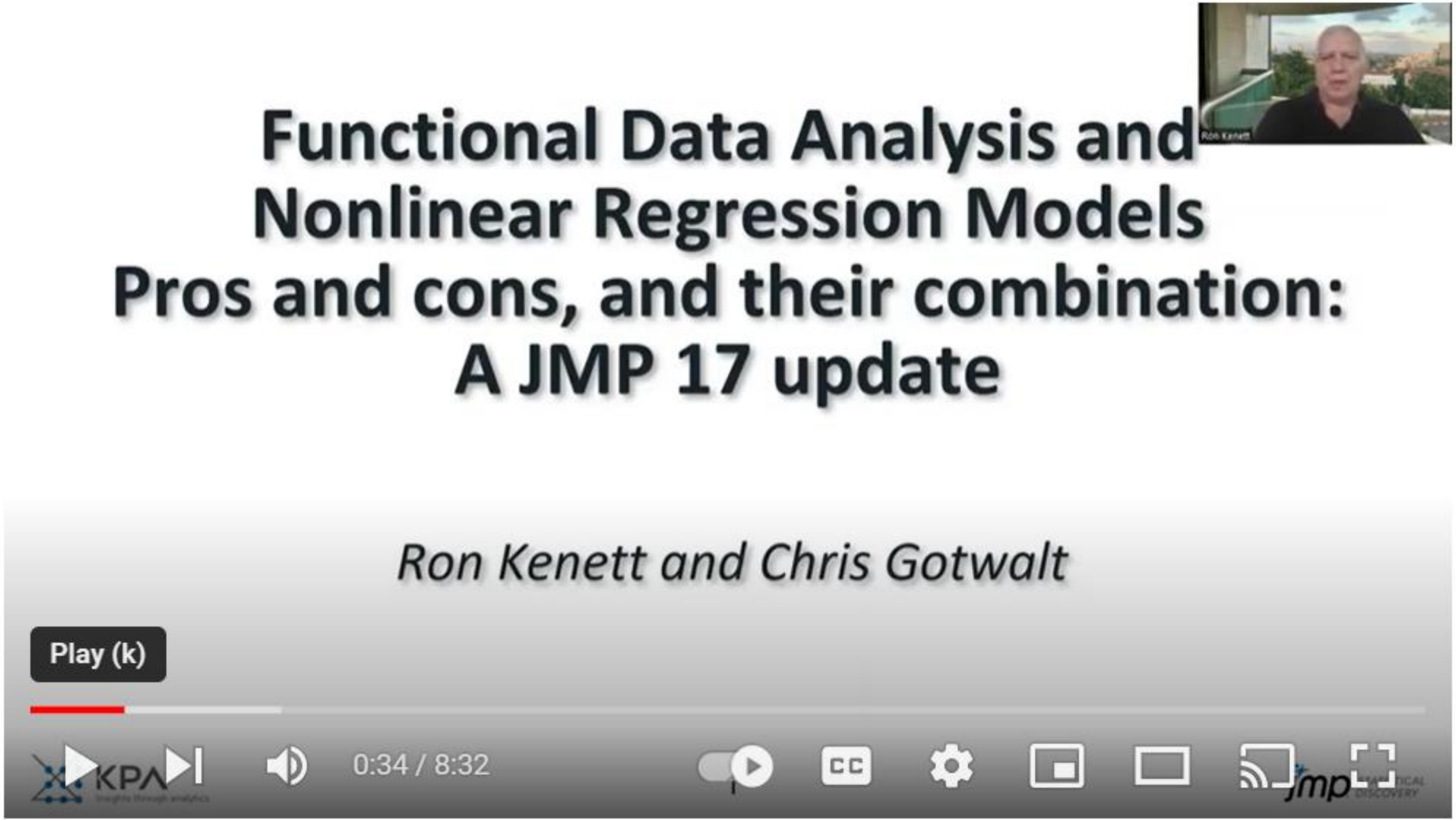
Timestamps

Obs	Lot_Unit	v7022	v7023	v7024	Oper8002	Oper9050	Oper9070
1	Lot01_U1	0.81	-0.85		2016-06-15-12:43	2016-06-17-06:42	2016-06-21-15:13
2	Lot01_U2	0.84	-0.87		2016-06-15-14:22	2016-06-17-04:59	2016-06-21-12:01
3	Lot01_U3	0.78	-0.83		2016-06-15-16:16	2016-06-17-05:29	
4	Lot01_U4	0.76	-0.81	-3.21	2016-06-15-17:43	2016-06-17-04:53	
5	Lot02_U1			-3.23	2016-06-15-14:36		
6	Lot02_U2			-3.11			
7	Lot02_U3			-3.05			
8	Lot02_U4			-3.17	2016-06-15-21:09		
9	Lot03_U1			-3.03	2016-06-16-03:15	2016-06-18-18:26	2016-06-21-16:53
10	Lot03_U2			-3.19	2016-06-16-03:29	2016-06-18-19:38	2016-06-21-17:17
11	Lot03_U3			-3.06	2016-06-16-04:37	2016-06-18-21:13	
12	Lot03_U4				2016-06-16-05:27		

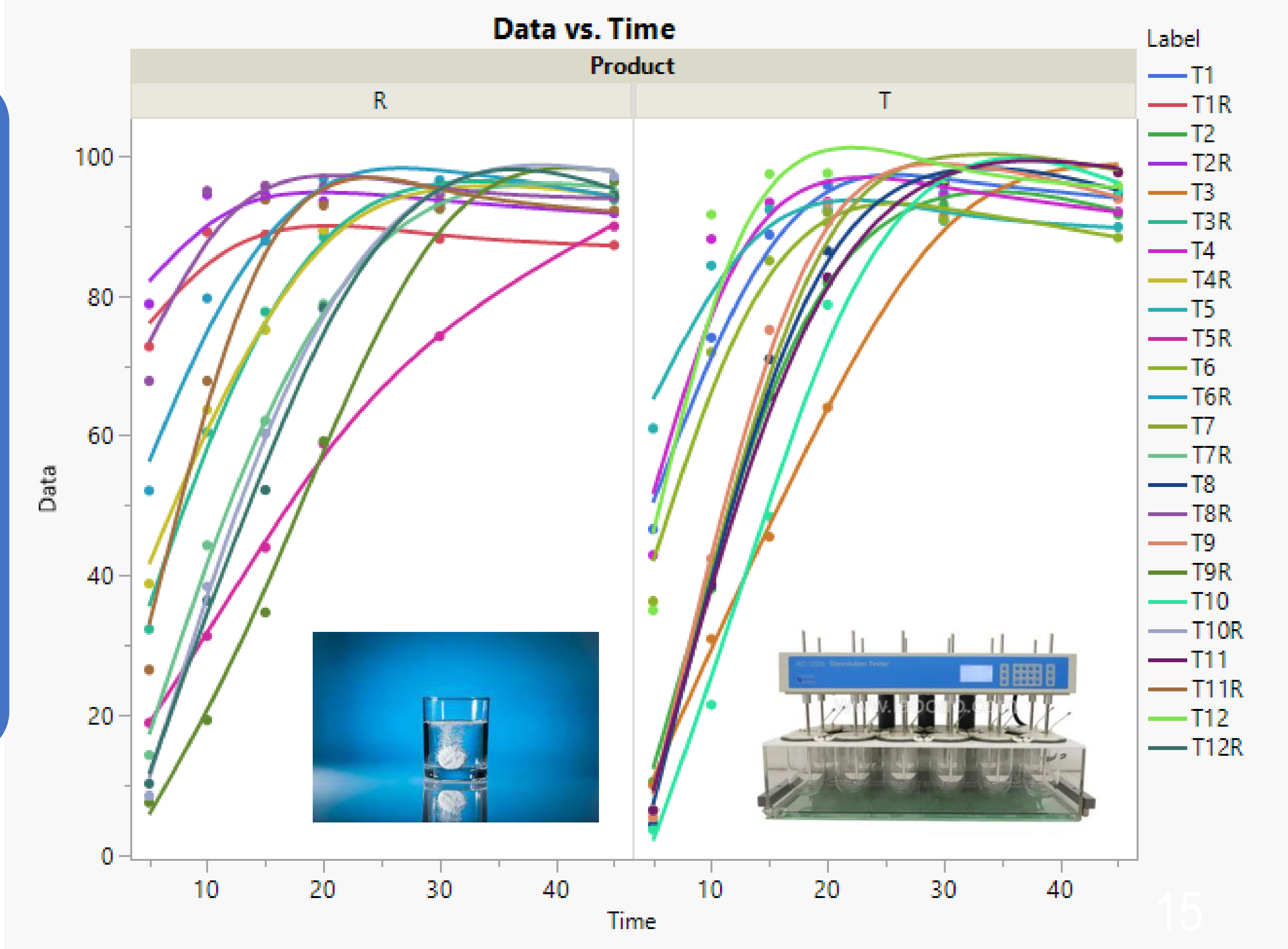
YouTube IL

Search

https://www.youtube.com/watch?v=g4gxLG2IQeo

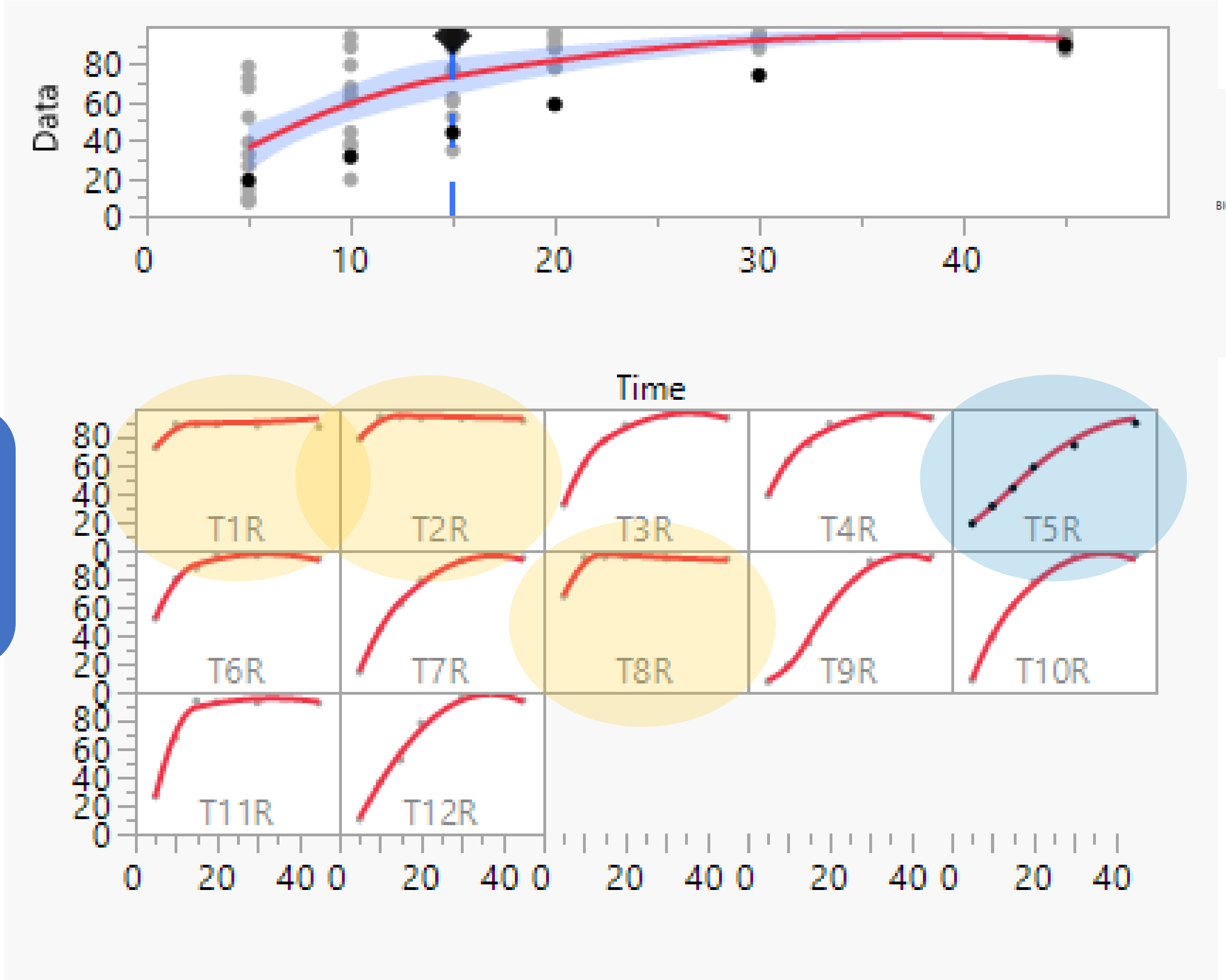


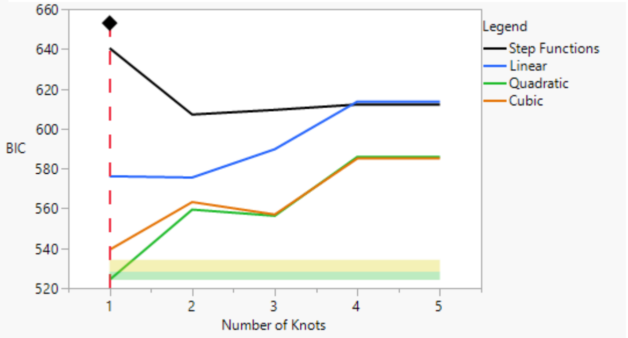
Dissolution Curves of 12 tablets as Test and Reference



Test T5R

Is different

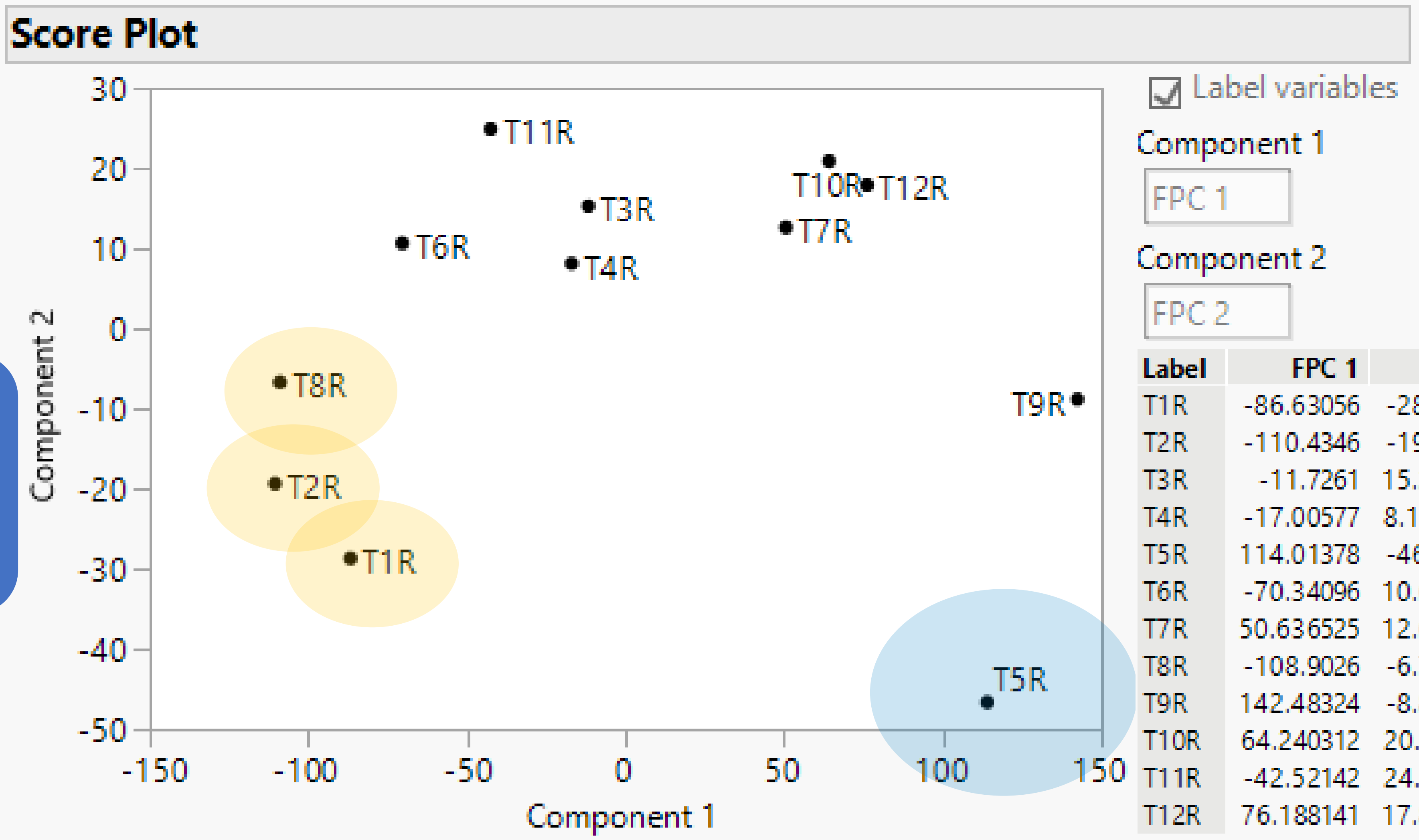




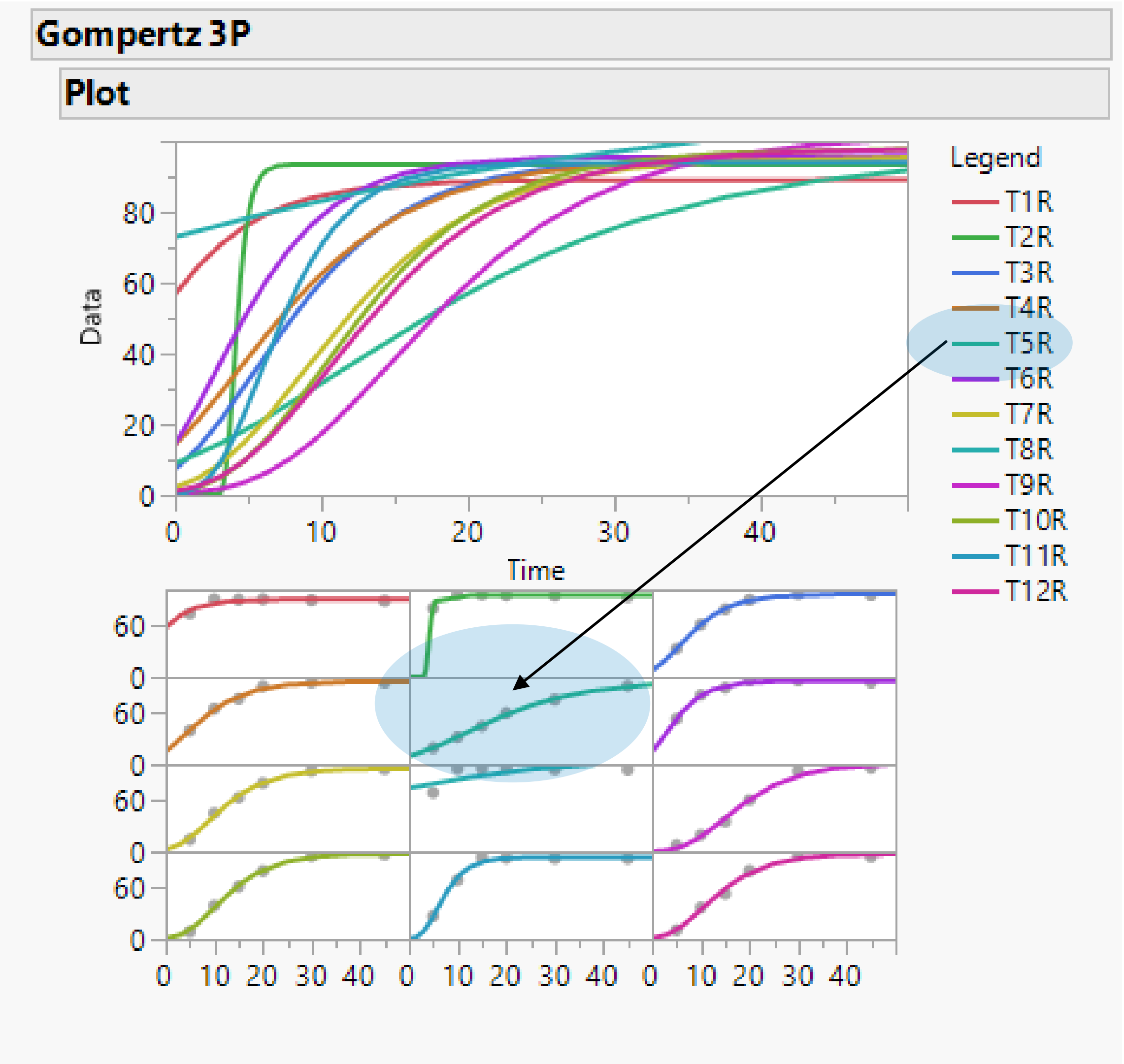
Quadratic B-spline with 1 knot

Test T5R

Is different



Label	FPC 1	FPC 2	FPC 3
T1R	-86.63056	-28.70704	0.996566
T2R	-110.4346	-19.38626	7.7488408
T3R	-11.7261	15.242747	0.0425637
T4R	-17.00577	8.1173121	3.7777815
T5R	114.01378	-46.64264	-12.92808
T6R	-70.34096	10.648925	3.9284304
T7R	50.636525	12.627046	-3.211241
T8R	-108.9026	-6.710816	0.0497409
T9R	142.48324	-8.840222	11.976841
T10R	64.240312	20.878974	-0.66174
T11R	-42.52142	24.890182	-18.14564
T12R	76.188141	17.881793	6.4259358



Prediction Model

$$a \cdot \mathrm{Exp}\left(-\mathrm{Exp}\left(-b \cdot \left(\mathrm{Time} - c\right)\right)\right)$$

a = Asymptote

b = Growth Rate

c = Inflection Point

Nonlinear model

	Label	Asymptote	Growth Rate	Inflection Point
T1R 1	T1R	89.072244404	0.2185624809	-3.625806907
2	T2R	93.480399791	1.76758908	4.0002987548
3	T3R	95.117117858	0.1732544061	5.4556204689
4	T4R	95.393703545	0.1508168903	4.2794474042
T5R 5	T5R	97.047132531	0.0750862269	11.579937352
6	T6R	95.886344295	0.2282099484	2.8239229453
7	T7R	95.608682945	0.1500953986	8.8540204547
8	T8R	113.26922091	0.0355126872	-23.11022674
9	T9R	102.16502758	0.1201635618	14.766362121
10	T10R	97.965019617	0.1562304451	10.087517474
11	T11R	94.032980681	0.3037771891	5.8648755174
12	T12R	97.966870258	0.1439240958	10.549169714

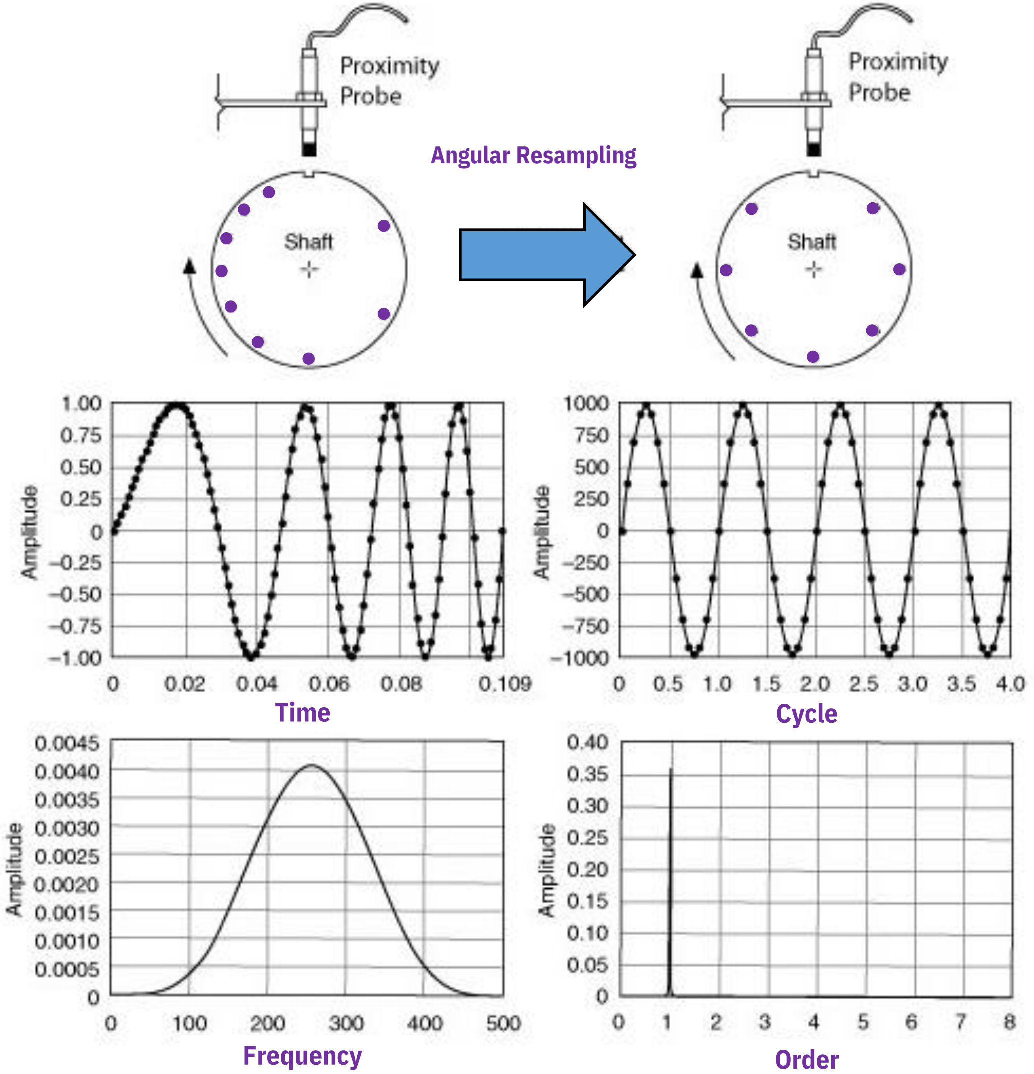
Prediction Model

$$a \cdot \mathrm{Exp}\left(-\mathrm{Exp}\left(-b \cdot \left(\mathrm{Time} - c\right)\right)\right)$$

a = Asymptote
b = Growth Rate
c = Inflection Point

Nonlinear model

Angular Resampling of Ball Bearing Engine



Text Data – Medical Device Inspections

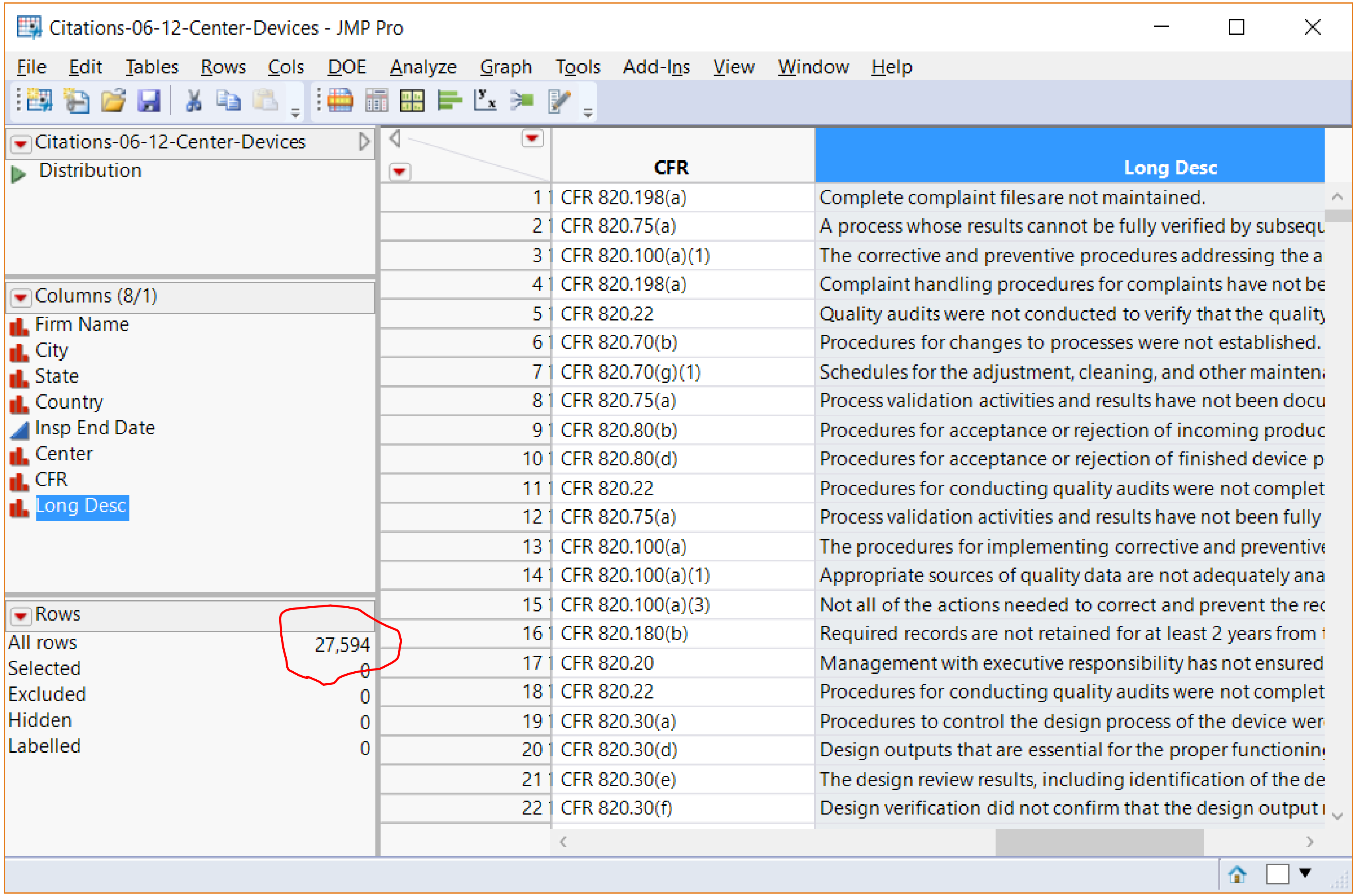
- **Data**: 27,594 inspection observations from fda.gov.

- **Objective**: Determine the most frequent themes in inspection observations for medical devices.

.

The Data

CFR - Code of Federal Regulations



Citations-06-12-Center-Devices - Text Explorer of Long Desc - JMP Pro

Text Explorer for Long Desc

Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-empty Cases	Portion Non-empty per Case
1062	27594	431455	15.6358	27594	1.0000

Term and Phrase Lists

Term	Count
procedures	12817
established	8404
device	6406
quality	5694
adequately	5049
design	4010
documented	3931
implemented	3353
product	3280
corrective	3156
preventive	2913
process	2818
ensure	2538
system	2396
complete	2377
activities	2291
control	2277
action	2189
complaints	2179
maintained	2071
requirements	2039
record	1974
records	1895
complaint	1808
results	1791
defined	1778
history	1703

Phrase	Count	N
adequately established	3204	2
corrective and preventive	2913	3
quality system	2163	2
corrective and preventive action	1882	4
preventive action	1882	2
nonconforming product	1328	2
quality audits	1247	2
device history	1080	2
device master record	1053	3
device master	1053	2
master record	1053	2
evaluating complaints	1037	2
quality problems	987	2
procedures for receiving	981	3
reviewing and evaluating complaints	977	4
reviewing and evaluating	977	3
written mdr	977	2
corrective and preventive actions	935	4
preventive actions	935	2
procedures for receiving reviewing	934	4
receiving reviewing and evaluating	934	4
receiving reviewing	934	2
complaint handling procedures	929	3
complaint handling	929	2
handling procedures	929	2
preventive action activities	911	3
action activities	911	2

Topic Words

Topic1	
Term	**Score**
serious	0.23412
injury	0.23397
death	0.23252
marketed	0.23166
suggests	0.23065
30	0.23041
becoming	0.23011
aware	0.22950
reasonably	0.22545
days	0.21984
within	0.21638
recur	0.18705
malfunctioned	0.18705
malfunction	0.18705
likely	0.18705
contribute	0.18705
submitted	0.18482

Topic2	
Term	**Score**
problems	0.28969
nonconforming	0.23010
causes	0.22648
actions	0.21586
recurrence	0.20936
corrective	0.20896
correct	0.20680
needed	0.19748
sources	0.19630
preventive	0.19380
prevent	0.18391
parties	0.17575
dissemination	0.17337
data	0.16424
including	0.15170

Topic3	
Term	**Score**
disclosure	0.26797
financial	0.25905
certification	0.24059
statement	0.21569
participating	0.20885
allow	0.20645
submit	0.20242
obtained	0.20230
commitment	0.19676
includes	0.19389
year	0.19187
signed	0.18538
investigator	0.16801
sponsor	0.16683
agreement	0.16625
accurate	0.16524

Topic4	
Term	**Score**
trustworthy	0.33022
generally	0.33022
paper	0.33022
reliable	0.32189
equivalent	0.29324
electronic	0.28572
limitation	0.25209
trail	0.25208
access	0.21608
audit	0.19973
check	0.19182

Topic5	
Term	**Score**
system	0.28155
fulfilling	0.26882
objectives	0.25451
effective	0.25435
quality	0.24894
verify	0.24870
regular	0.24778
prescribed	0.24778
sufficient	0.21578
intervals	0.21549
internal	0.20033
audits	0.19772
conducted	0.16045

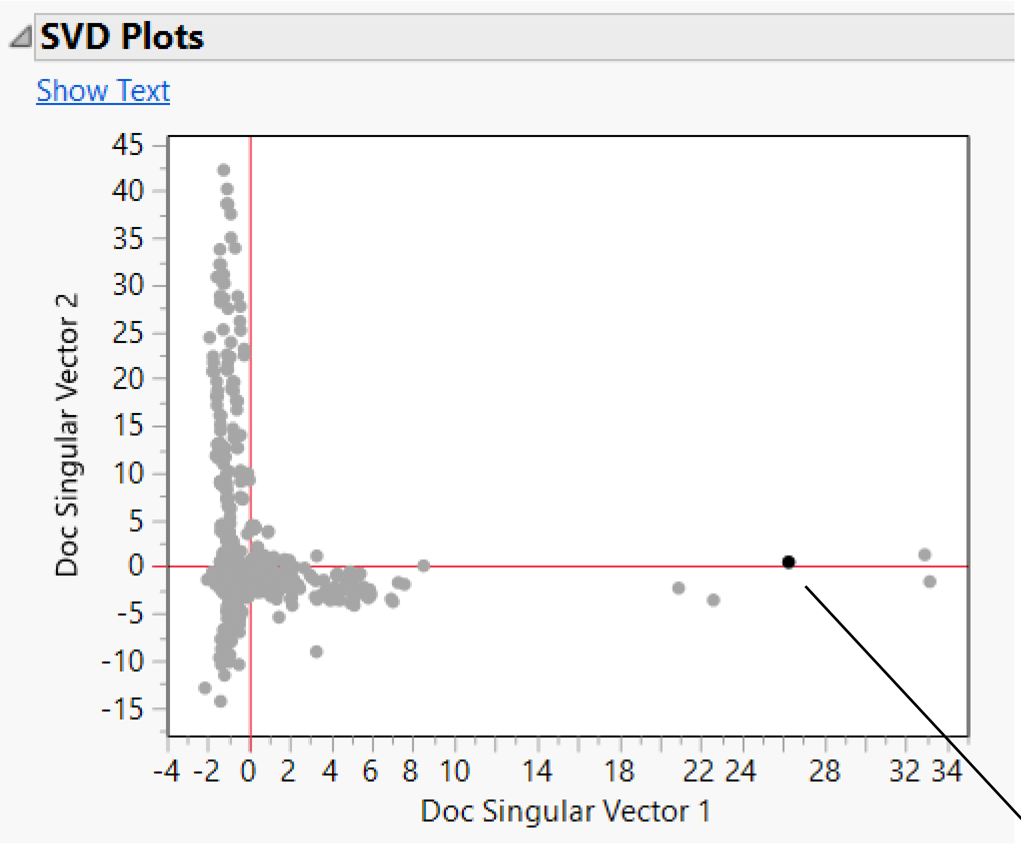
Topic6	
Term	**Score**
subsequent	0.32026
whose	0.32026
inspection	0.31169
test	0.30422
verified	0.30219
according	0.28521
validated	0.27202
fully	0.24942
results	0.22145
approved	0.18579
process	0.15686

Topic7	
Term	**Score**
health	0.28786
reported	0.27381
correction	0.26683
removal	0.26448
fda	0.24325
writing	0.23956
reduce	0.23956
posed	0.23956
risk	0.19181
fd&c	0.17168
act	0.16817
violation	0.16723
present	0.16723
might	0.16723

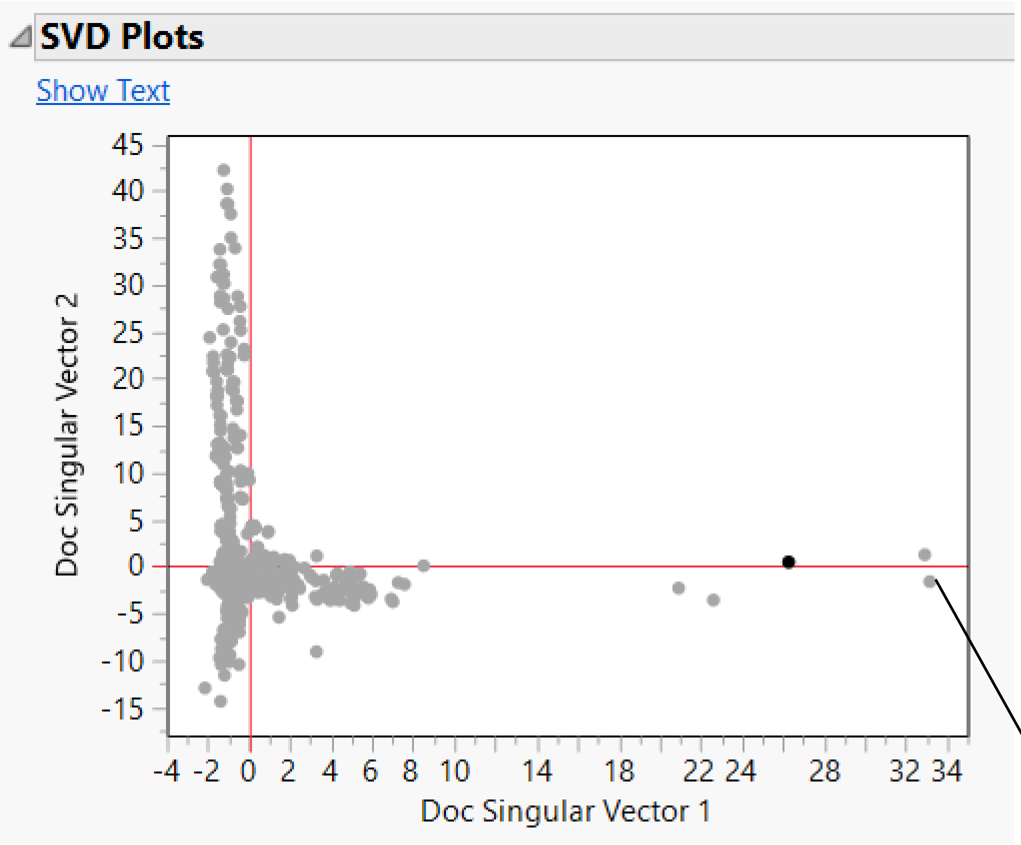
Topic8	
Term	**Score**
batch	0.24329
manufactured	0.23190
demonstrate	0.23184
device	0.21879
history	0.20094
record	0.19734
accordance	0.19553
master	0.19455
regulation	0.19144
lot	0.18818
records	0.16649
code	0.15537
mark	0.14889
systems	0.14876
number	0.14709

Topic9	
Term	**Score**
provides	0.26933
procedure	0.26067
reporting	0.25708
event	0.21770
internal	0.20418
standardized	0.19515
determining	0.19303
medical	0.18473
events	0.16167
written	0.15374
communication	0.14876
include	0.14163
criteria	0.14143
timely	0.14027

Topic10	
Term	**Score**
effect	0.25470
adverse	0.25123
learning	0.24373
prepared	0.24040
unanticipated	0.23452
working	0.23179
first	0.22702
10	0.20946
effects	0.15673
storage	0.14708
damage	0.14302
rooms	0.13304
stock	0.13244



An MDR report was not submitted within 30 days of receiving or otherwise becoming aware of information that reasonably suggests that a marketed device may have caused or contributed to a death or serious injury. [127]

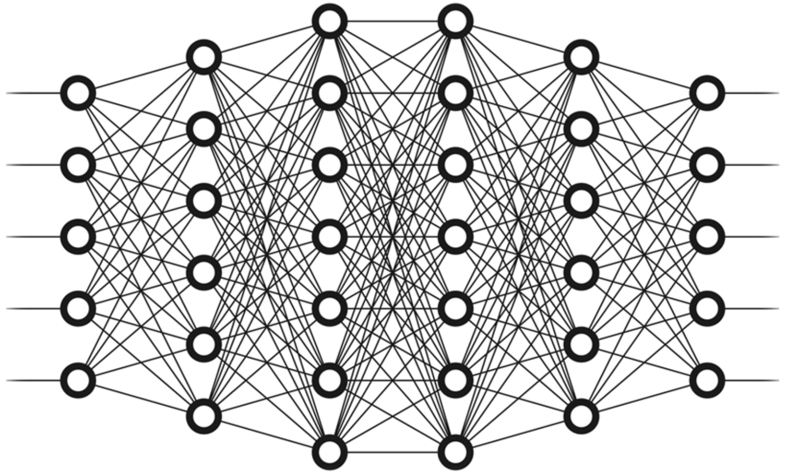


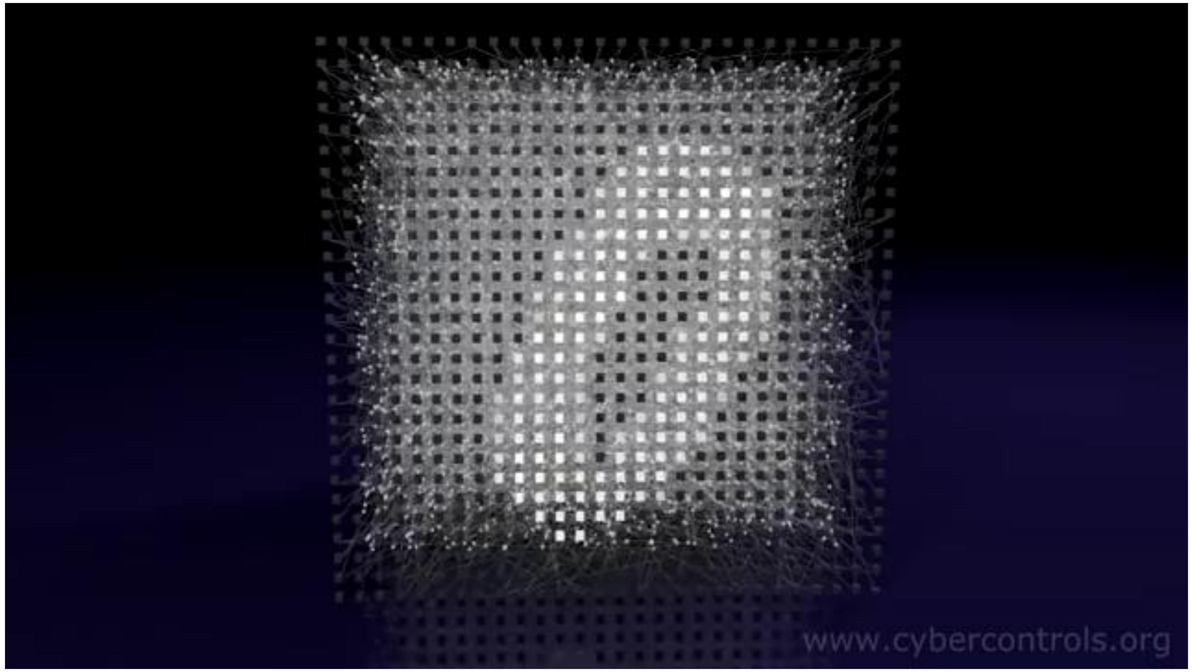
An MDR report was not submitted within 30 days of receiving or otherwise becoming aware of information that reasonably suggests that a marketed device has malfunctioned and would be likely to cause or contribute to a death or serious injury if the malfunction were to recur. [271]

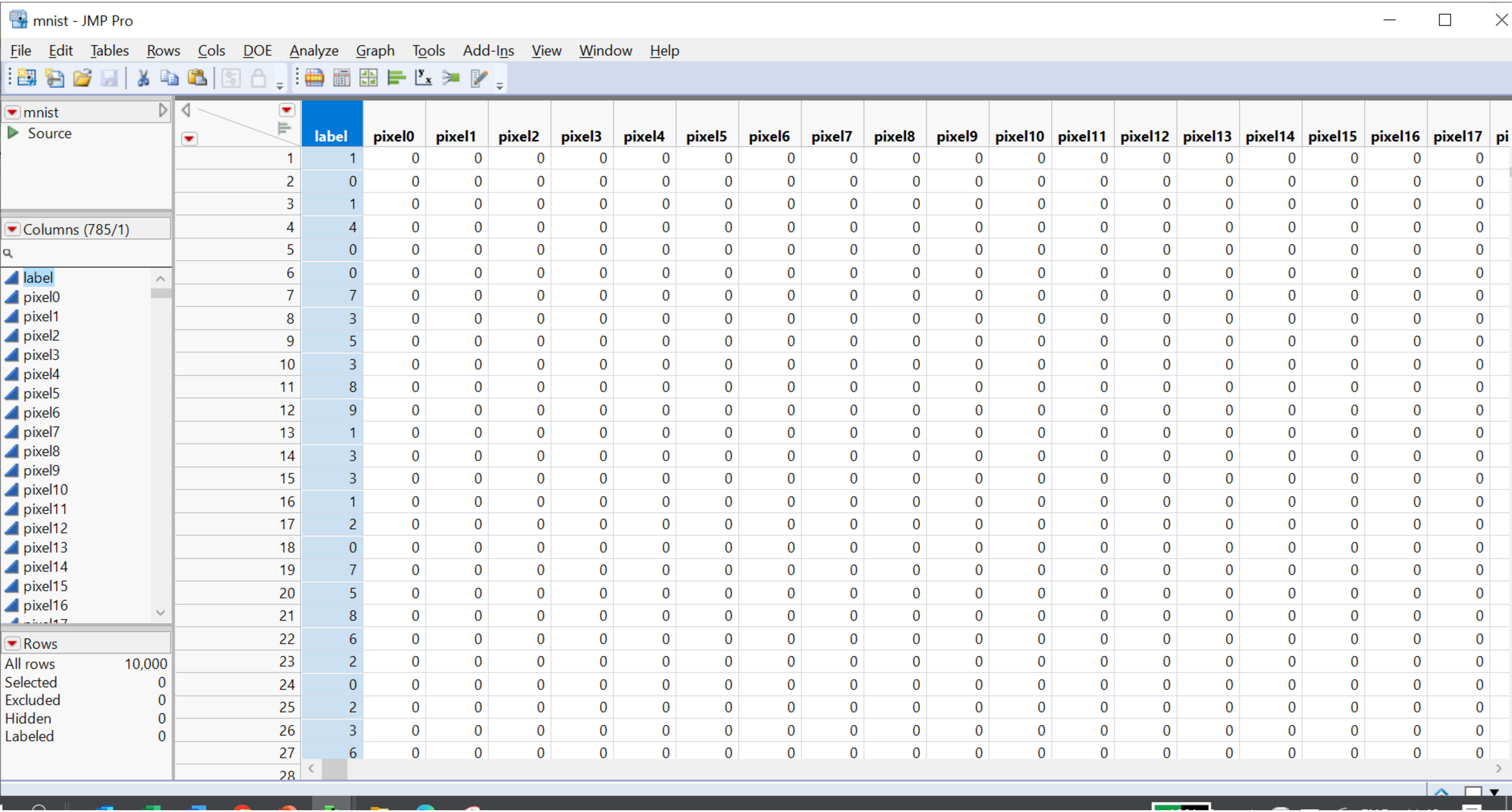
Deep learning

MNIST
Modified **N**ational **I**nstitute of **S**tandards and **T**echnology

Contains 60,000 training images and 10,000 testing images

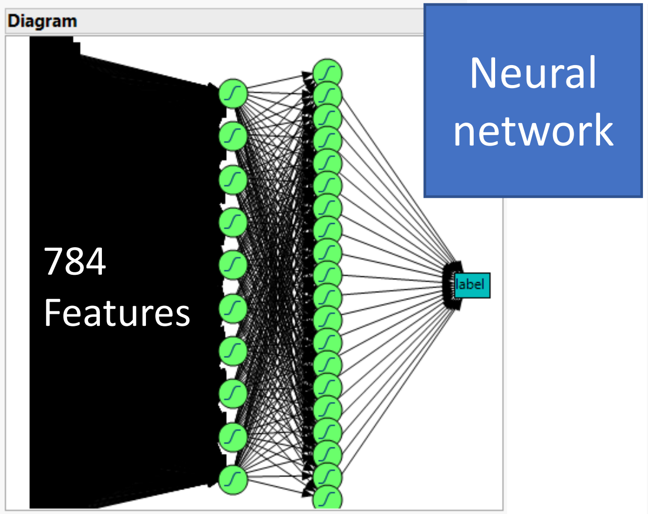


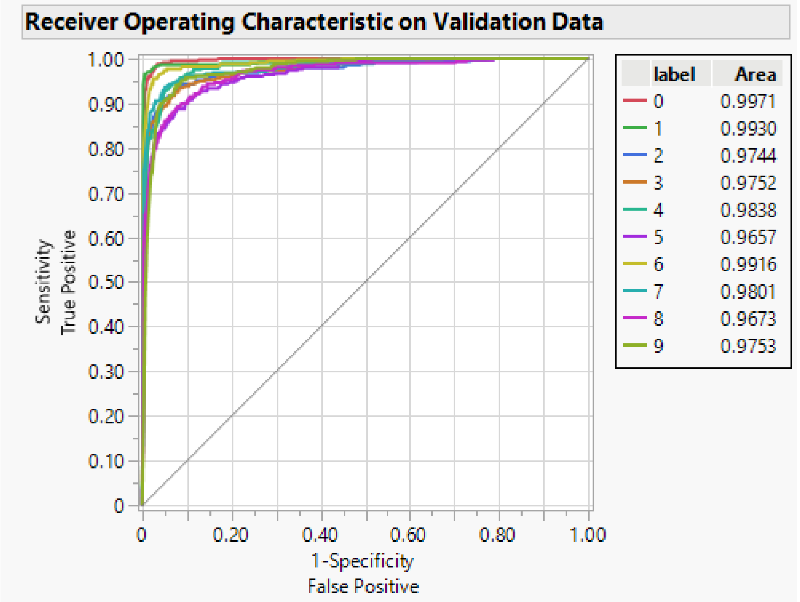




	label	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	pixel9	pixel10	pixel11	pixel12	pixel13	pixel14	pixel15	pixel16	pixel17	pi
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28																				

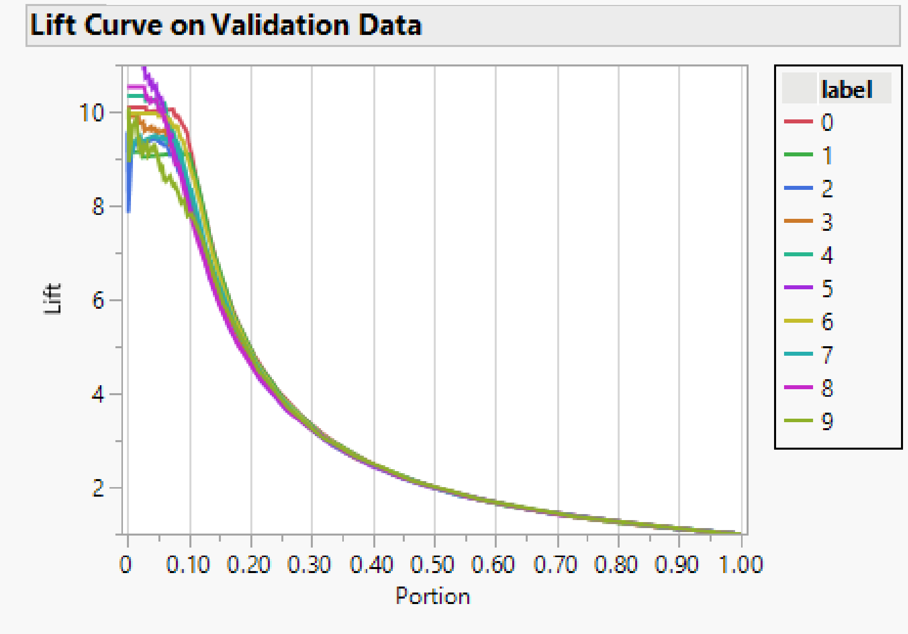
Measures	Value
Generalized RSquare	0.9791839
Entropy RSquare	0.7573229
RASE	0.3655983
Mean Abs Dev	0.1904833
Misclassification Rate	0.1509886
-LogLikelihood	1864.1996
Sum Freq	3338



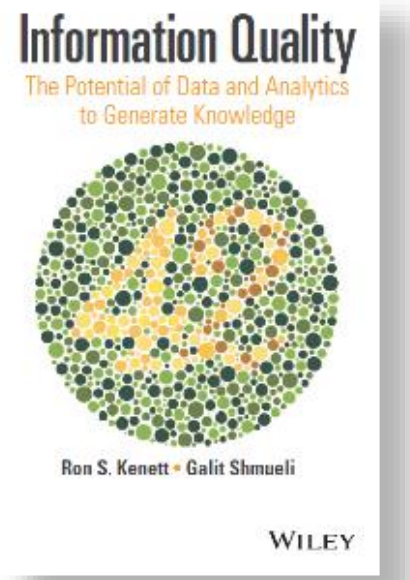


Confusion Matrix

Actual	Predicted Count									
label	0	1	2	3	4	5	6	7	8	9
0	315	0	0	1	0	6	6	1	2	0
1	0	352	2	0	3	0	1	3	4	0
2	9	1	290	16	6	2	7	1	15	2
3	6	4	11	278	1	19	2	5	6	5
4	1	0	5	0	264	1	7	2	3	40
5	9	4	7	16	5	228	4	2	20	7
6	6	2	8	0	7	6	298	4	4	0
7	2	5	4	4	3	1	2	292	7	27
8	2	8	15	11	5	17	2	3	246	8
9	2	0	3	2	23	6	2	16	7	271

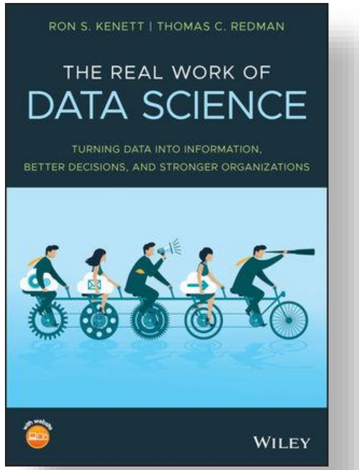


$$InfoQ(f,X,g) = U(f(X|g))$$

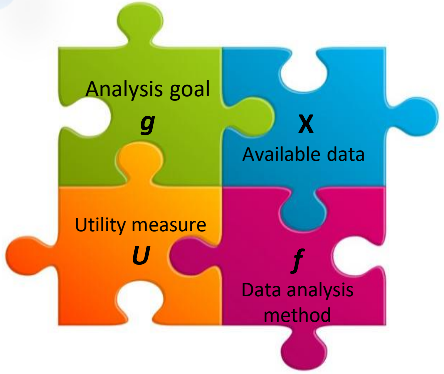


InfoQ

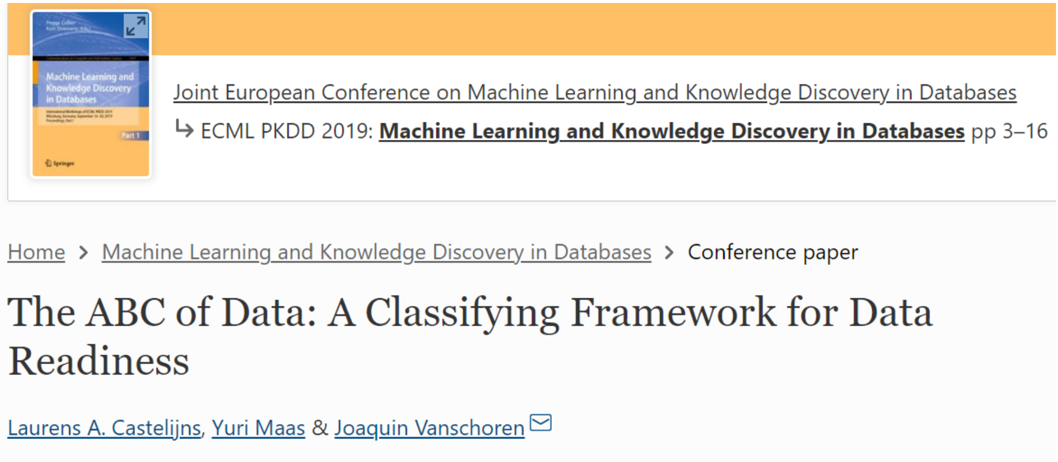
1. Data resolution
2. Data structure
3. Data integration
4. Temporal relevance
5. Chronology of data and goal
6. Generalizability
7. Operationalization
8. Communication



How



https://link.springer.com/chapter/10.1007/978-3-030-43823-4_1


Joint European Conference on Machine Learning and Knowledge Discovery in Databases
↳ ECML PKDD 2019: **Machine Learning and Knowledge Discovery in Databases** pp 3–16

Home > Machine Learning and Knowledge Discovery in Databases > Conference paper

The ABC of Data: A Classifying Framework for Data Readiness

Laurens A. Castelijns, Yuri Maas & Joaquin Vanschoren

Information Quality

The potential of a particular dataset to achieve a particular goal using a given empirical analysis method

g	**A specific analysis goal**
X	**The available dataset**
f	**An empirical analysis method**
U	**A utility measure**

https://github.com/pywash/pywash

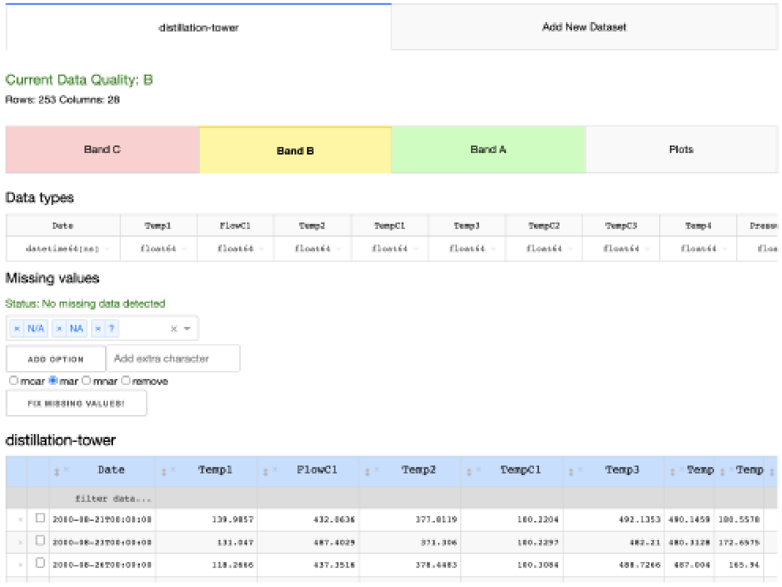
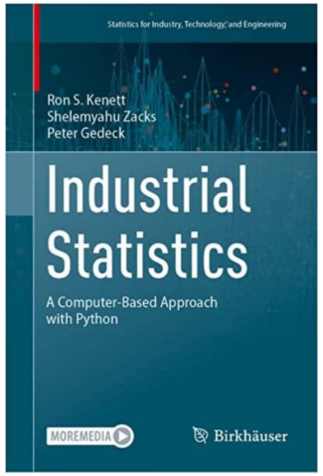


Fig. 8.5: Screenshot of PYWASH software

Band C (Conceive)
Band B(Believe)
Band A (Analyze)
Band AA (Allow Analysis)
Band AAA (Full Readiness)



$$InfoQ(f,X,g) = U(\, f(X|g)\,)$$

https://github.com/gedeck/mistat

Kenett, R.S. and Shmueli , G. (2014) On Information Quality , *Journal of the Royal Statistical Society, Series A* (with discussion), Vol. 177, No. 1, pp. 3-38. http://ssrn.com/abstract=1464444.

Assessing Information Quality

Assess dimensions versus goal

InfoQ dimensions

1. Data resolution
2. Data structure
3. Data integration
4. Temporal relevance
5. Chronology of data and goal
6. Generalizability
7. Operationalization
8. Communication

Assess properties

"Quality of Statistical Data"
(Eurostat, OECD, NCSES,…)

- Relevance
- Accuracy
- Timeliness and punctuality
- Accessibility
- Interpretability
- Coherence
- Credibility

http://www.nsf.gov/statistics/information-quality.cfm
http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf
http://www.oecd.org/std/qualityframeworkforoecdstatisticalactivities.htm

An Italian Case Study

$$InfoQ(f,X,g) = U(f(X|g))$$

f: Use **Bayesian networks** to model the dependence structure of the variables in the data set and to calculate the conditional rank correlations

Kenett, Ron S., Applications of Bayesian Networks (2021). http://dx.doi.org/10.2139/ssrn.2172713

g: Understand the influence on sales of several variables, such as number of employees, to make predictions and derive diagnostics.

X: combined survey data and individual company performance with data reported to the stock exchange.

U: Sales prediction error in employment policy economic programs

#1 Data Resolution

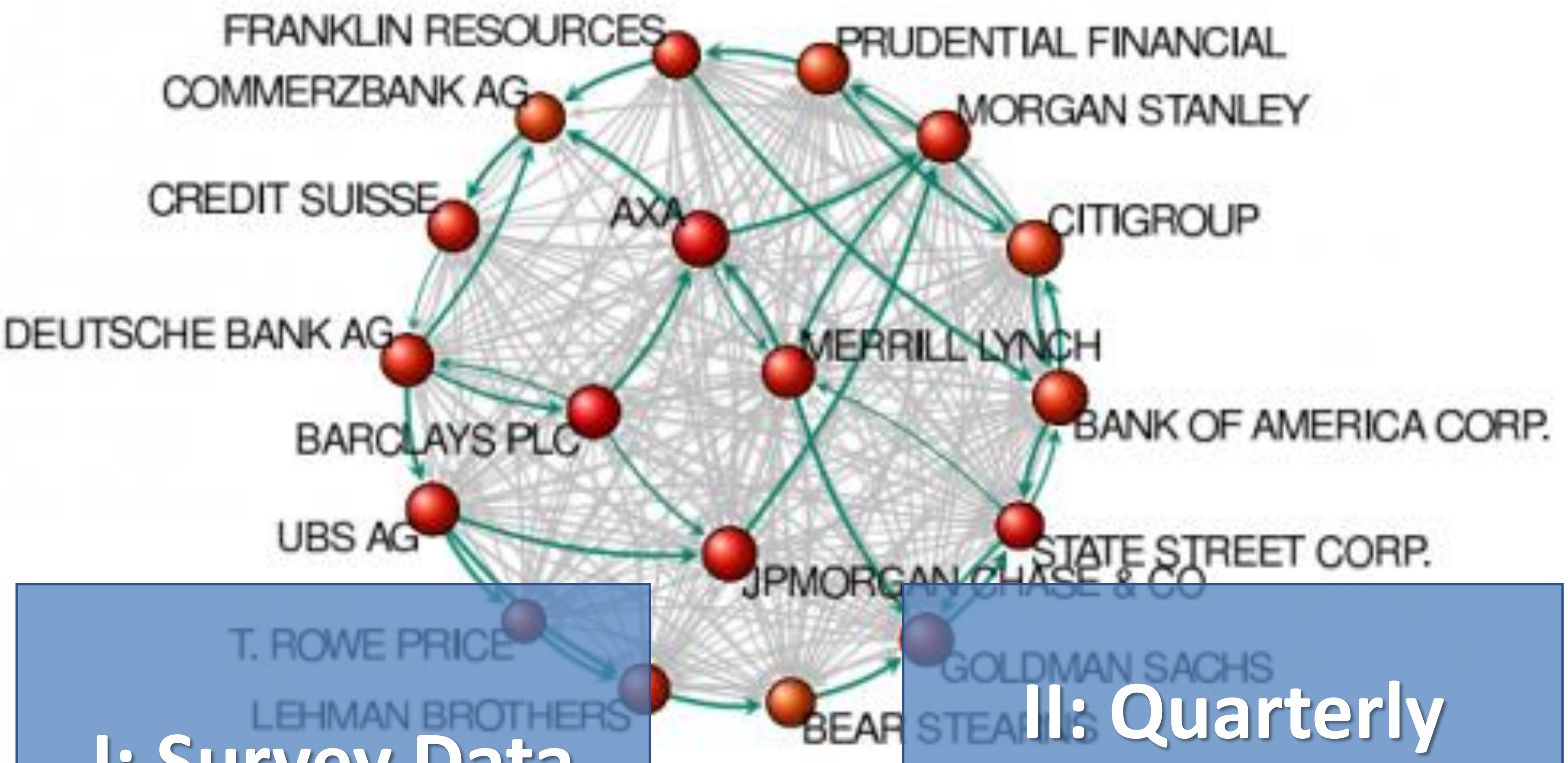
Data collected at the company level.

I: Periodic survey waves of self reports

II: Quarterly stock exchange reports

Goal: Predict sales using # employees in the context of a regional development plan

#2 Data Structure



I: Survey Data

II: Quarterly Reports

I: The Assolombarda Data

- Assolombarda is an Italian association of about 5,000 firms located in the province of Milan and in other provinces of the north of Italy, and represents manufacturing and service companies.
- The associated firms employ about 300,000 workers locally and several hundred thousands in the whole country.
- Assolombarda periodically collects data through questionnaires sent to the associated firms, in order to gather information about the economic climate, firms' activity and production, and the number and types of employees.
- The data analyzed contains information collected through one of the association surveys in 2007, and it is about 167 firms located in the provinces of Milan and Lodi.

I: The Assolombarda Data

The variables in the dataset are:

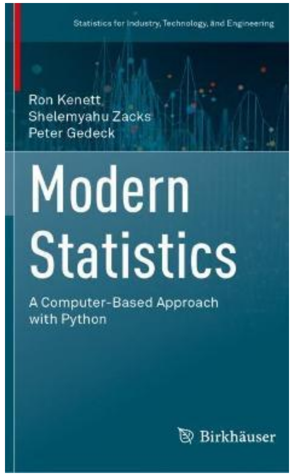
- *sales*: firm annual turnover;
- *emp*: average number of employees;
- *rise*: number of managers receiving wage rise;
- *rise2*: number of managers that will receive wage rise in the following year;
- *prom*: number of employees gaining a promotion;
- *horiz* : number of employees involved in horizontal movements;
- *ext*: number of people employed in the external market;
- *grad*: number of newly-graduated employees;
- *qual*: number of newly-qualified employees.

X

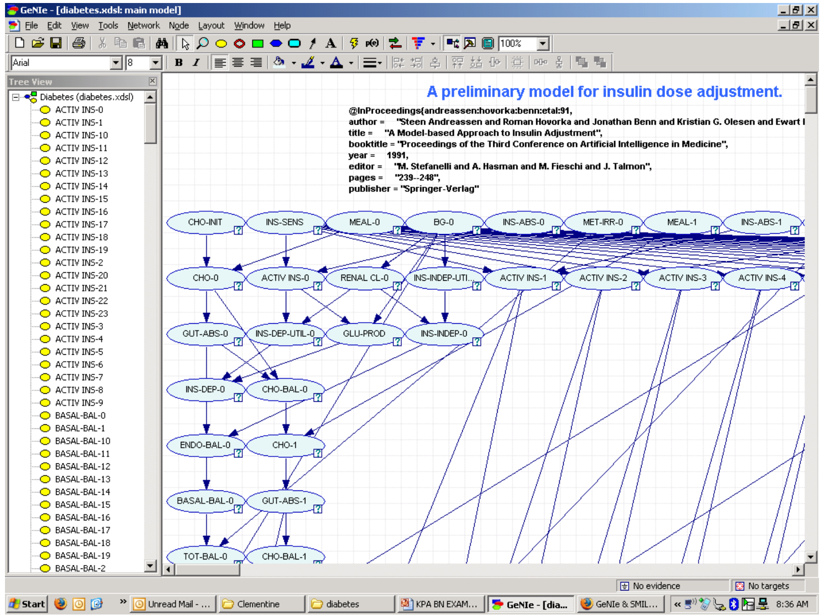
8.3 Bayesian Networks

Bayesian networks (BNs) were introduced in Sect. 2.1.6. They implement a graphical model structure known as a directed acyclic graph (DAG) that is popular in statistics, machine learning, and artificial intelligence. BNs enable an effective representation and computation of the joint probability distribution over a set of random variables (Pearl 1985). The structure of a DAG is defined by two sets: the set of nodes and the set of directed arcs; arcs are often also called edges. The nodes represent random variables and are drawn as circles labeled by the variable names. The arcs represent links among the variables and are represented by arrows between nodes. In particular, an arc from node X_i to node X_j represents a relation between the corresponding variables. Thus, an arrow indicates that a value taken by variable X_j depends on the value taken by variable X_i. This property is used to reduce the number of parameters that are required to characterize the joint probability distribution (JPD) of the variables. This reduction provides an efficient way to compute the posterior probabilities given the evidence present in the data

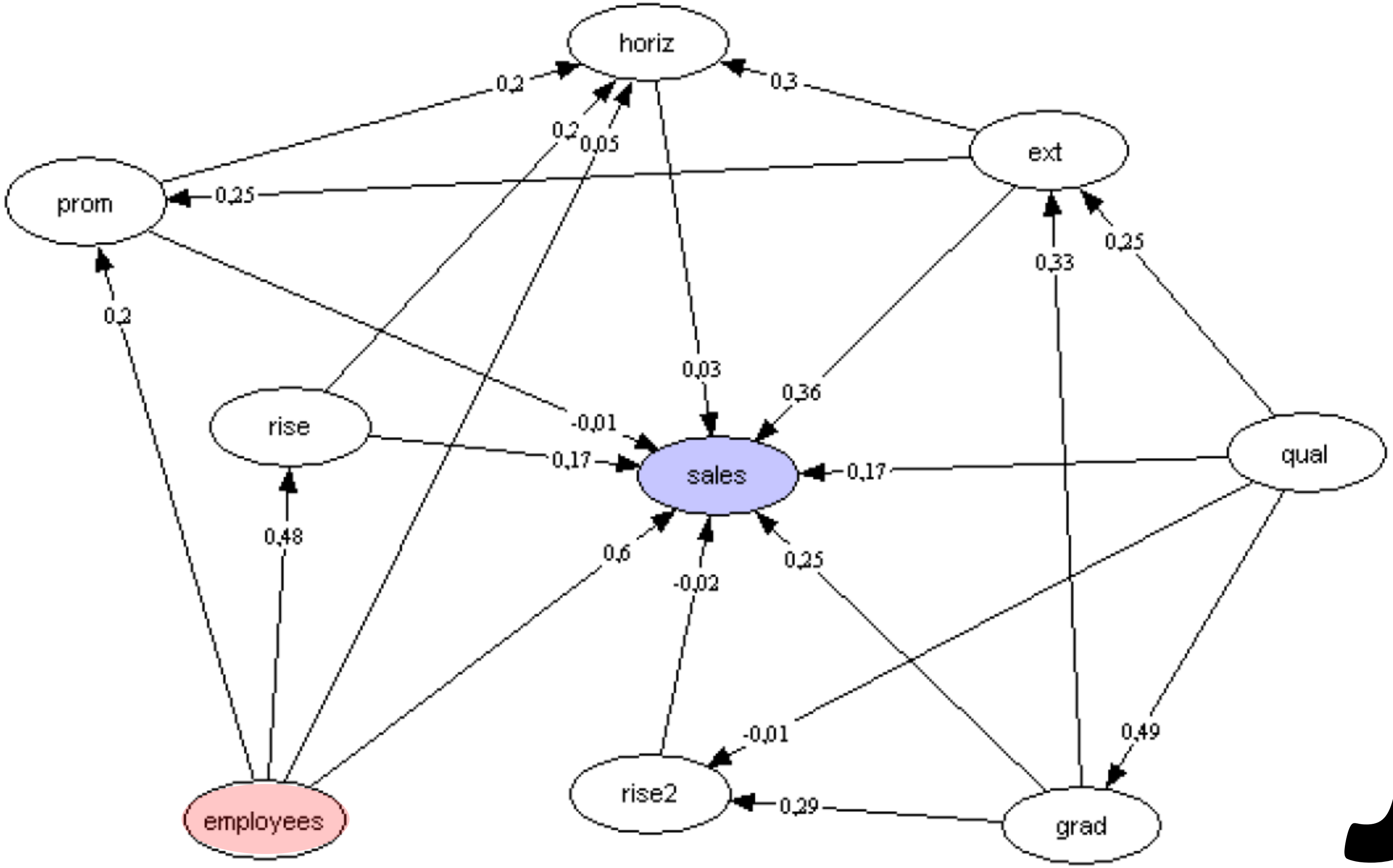
Kenett, Ron S., Applications of Bayesian Networks (2021). http://dx.doi.org/10.2139/ssrn.2172713

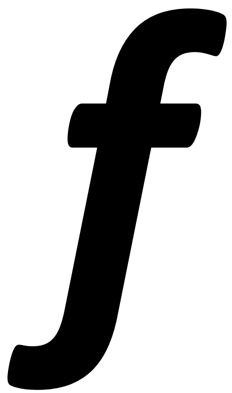


GeNIe 2.1



I: The Assolombarda Data





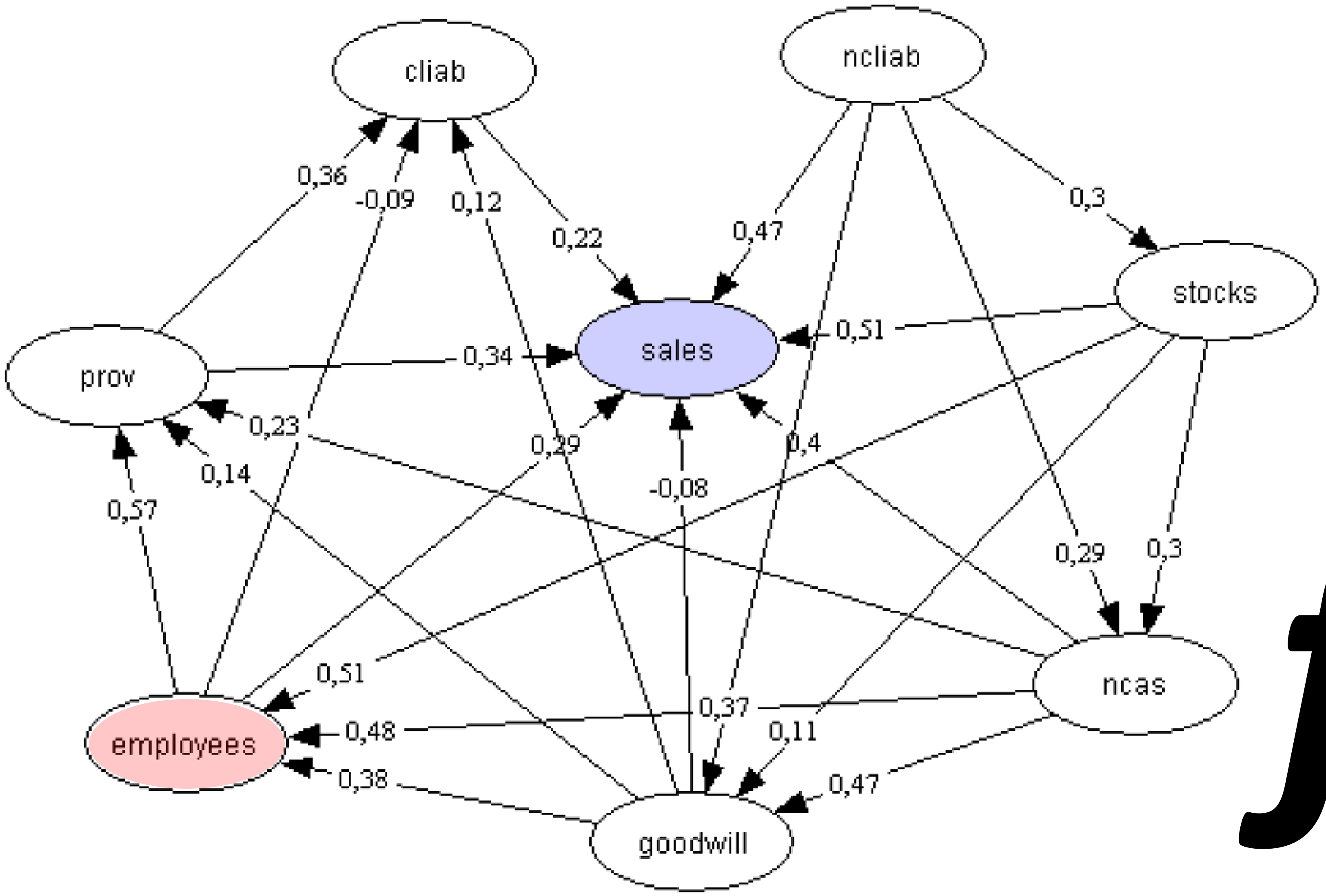
II: The FTSE-MTB Data

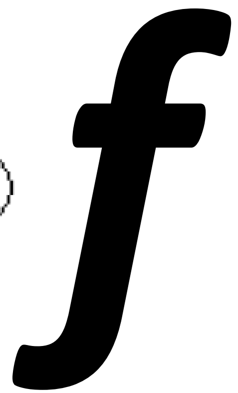
The FTSE-MIB is the benchmark stock market index for the Italian national stock exchange and consists of the 40 most-traded stock classes on the exchange. The dataset analyzed here contains information from the balance sheets of the 40 largest Italian firms belonging to the Italian stock market. The variables used in the analysis are:

- *sales*: firm annual turnover;
- *emp*: average number of employees;
- *goodwill*: difference between the balance sheet assets and the sum of intangible assets and equipment at market value;
- *ncas*: non-current financial assets;
- *stocks*: stocks and work in progress;
- *prov*: provisions for liabilities and non-recurring expenses;
- *ncliab*: non-current liabilities;
- *cliab*: current liabilities.

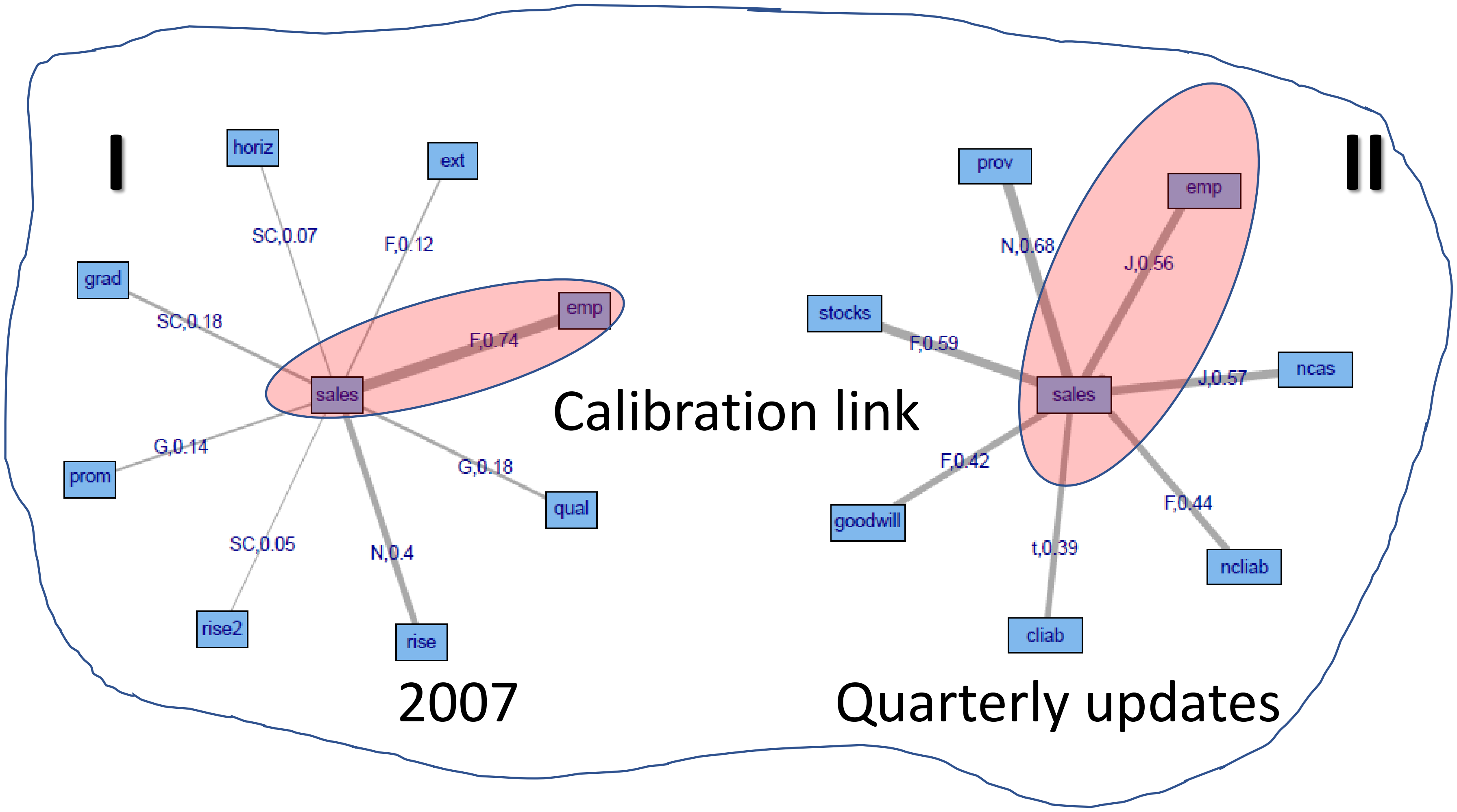
X

II: The FTSE-MTB Data





#3 Data Integration

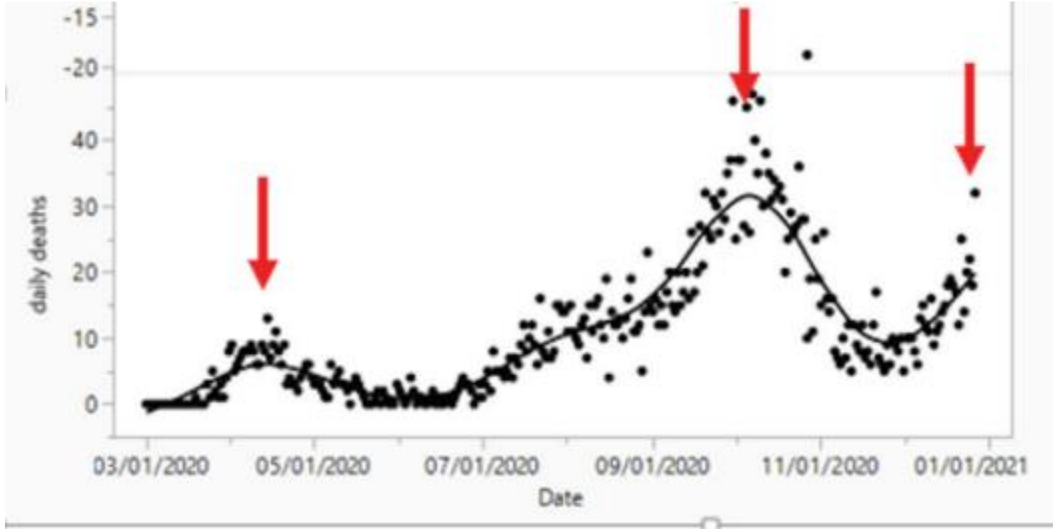


https://libri.unimi.it/index.php/milanoup/catalog/download/73/155/653-1


Citizen Behavior & Health Indicators in Israel During COVID-19: A Systematic Analysis of Data Over Time

Ron S. Kenett[1] *and Carmit Rapaport*[2]

[1]Samuel Neaman Institute, Technion, and KPA, Israel and University of Turin, Italy – ron@kpa-group.com
[2]Institute for Regulation of Emergency and Disaster, College of Law and Business and Department of Geography and Environmental Studies, University of Haifa, Israel – carmit.rapaport@gmail.com



International Journal of
Environmental Research and Public Health

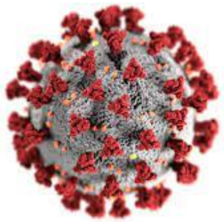
MDPI

Article

Integrated Analysis of Behavioural and Health COVID-19 Data Combining Bayesian Networks and Structural Equation Models

Ron S. Kenett [1], Giancarlo Manzi [2], Carmit Rapaport [3,4] and Silvia Salini [2,*]

1 KPA Group and Samuel Neaman Institute, Raanana 43100, Israel; ron@kpa-group.com
2 Data Science Research Centre, Department of Economics, Management and Quantitative Methods, University of Milan, 20122 Milan, Italy; giancarlo.manzi@unimi.it
3 Department of Geography and Environmental Studies, University of Haifa, Haifa 3498838, Israel; carmit.rapaport@gmail.com
4 NIRED—National Institute for Regulation of Emergency and Disaster, College of Law and Business, Ramat Gan 5110801, Israel
* Correspondence: silvia.salini@unimi.it



https://www.mdpi.com/1660-4601/19/8/4859/pdf

Goals of Research

- To assess the impact of pandemic management and mitigation policies on pandemic spread and population activity.

- To examines the effect of mobility restriction measures in Italy and Israel and compares the association between health and population mobility data.

- To provide decision makers a way to conduct scenario analysis to help support pandemic management

Methodology

1. **Collect** data on health and population behavior from ministries of health and google mobility

2. **Integrate** the data using Bayesian networks and determined proper lags using arc strength indicators.

3. **Assess** the derived network structure using confirmatory SEM.

4. **Discretize** the data accounting for local thresholds and use the resulting BNs to assess alternative scenarios. For example: what would be the impact of closing airports?

5. **Calibrate** the data from Italy and Israel using "wave" time windows and using country-based thresholds.

6. **Compare and contrast**: The fact that we did this analysis in two countries proved very effective from a methodology viewpoint.

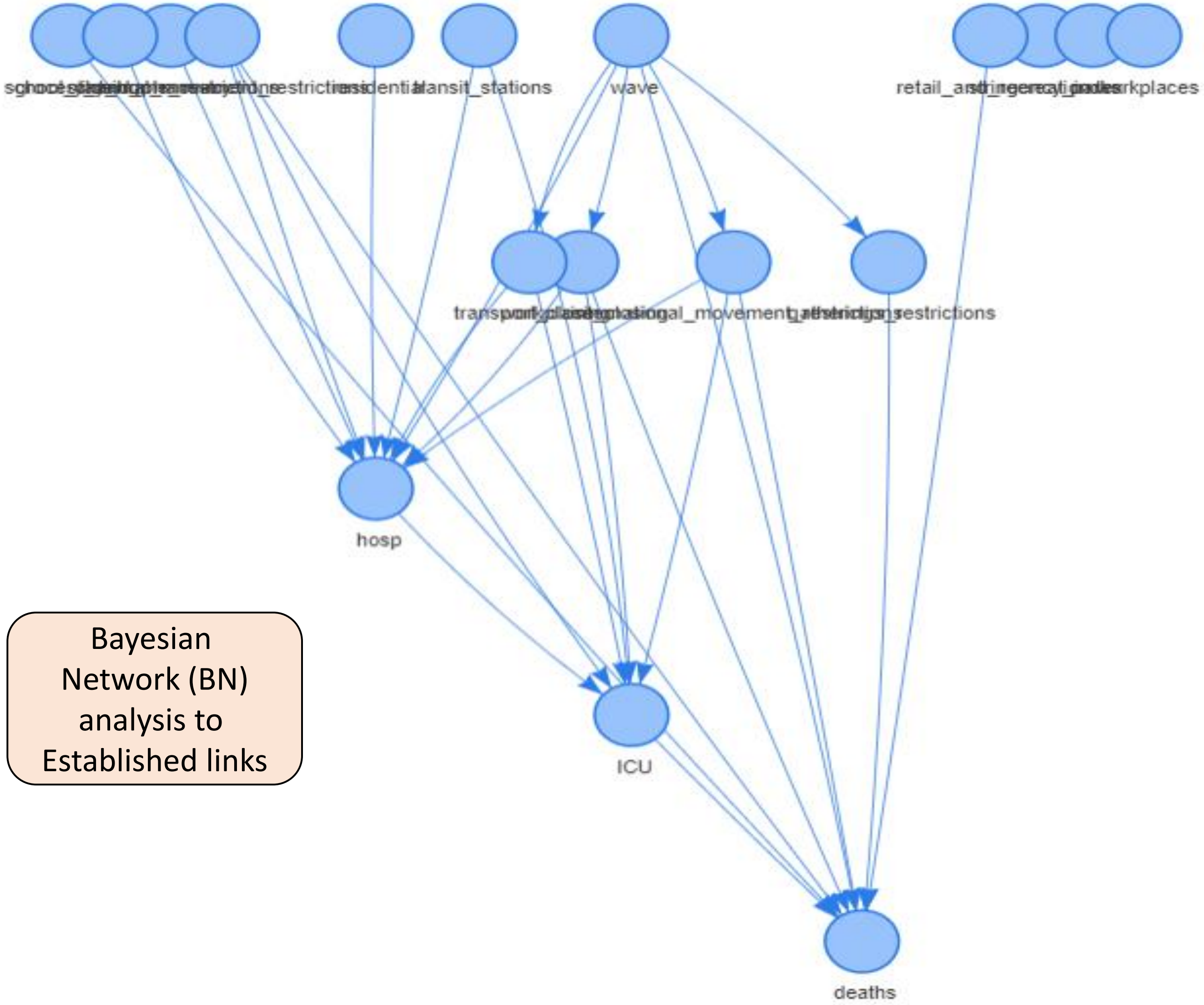
Covid19 Israel

Monitoring of emergency

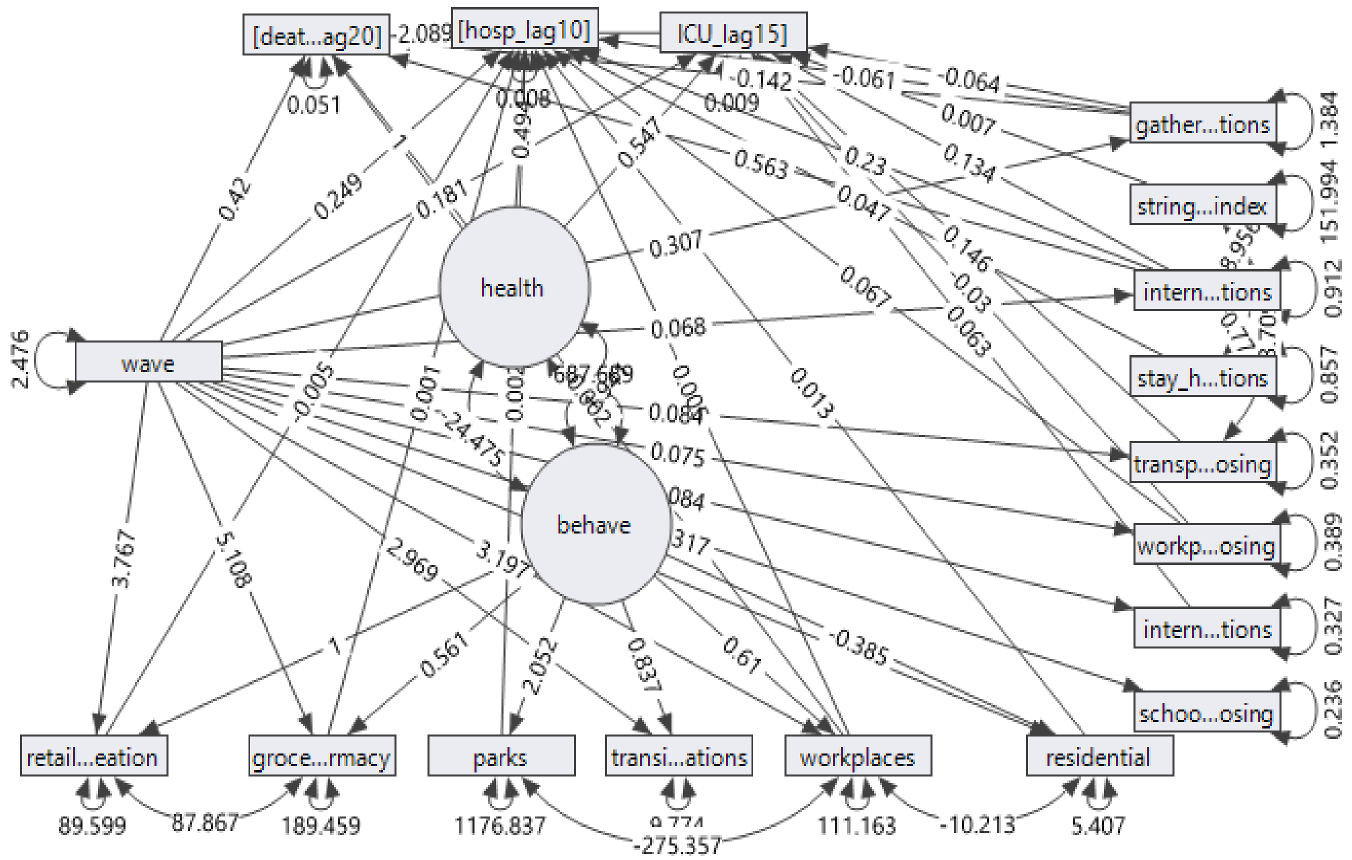
Bayesian Network (BN) analysis to establish links

Structural Equation Models (SEM) to confirm links

Bayesian Network (BN) analysis to Established links

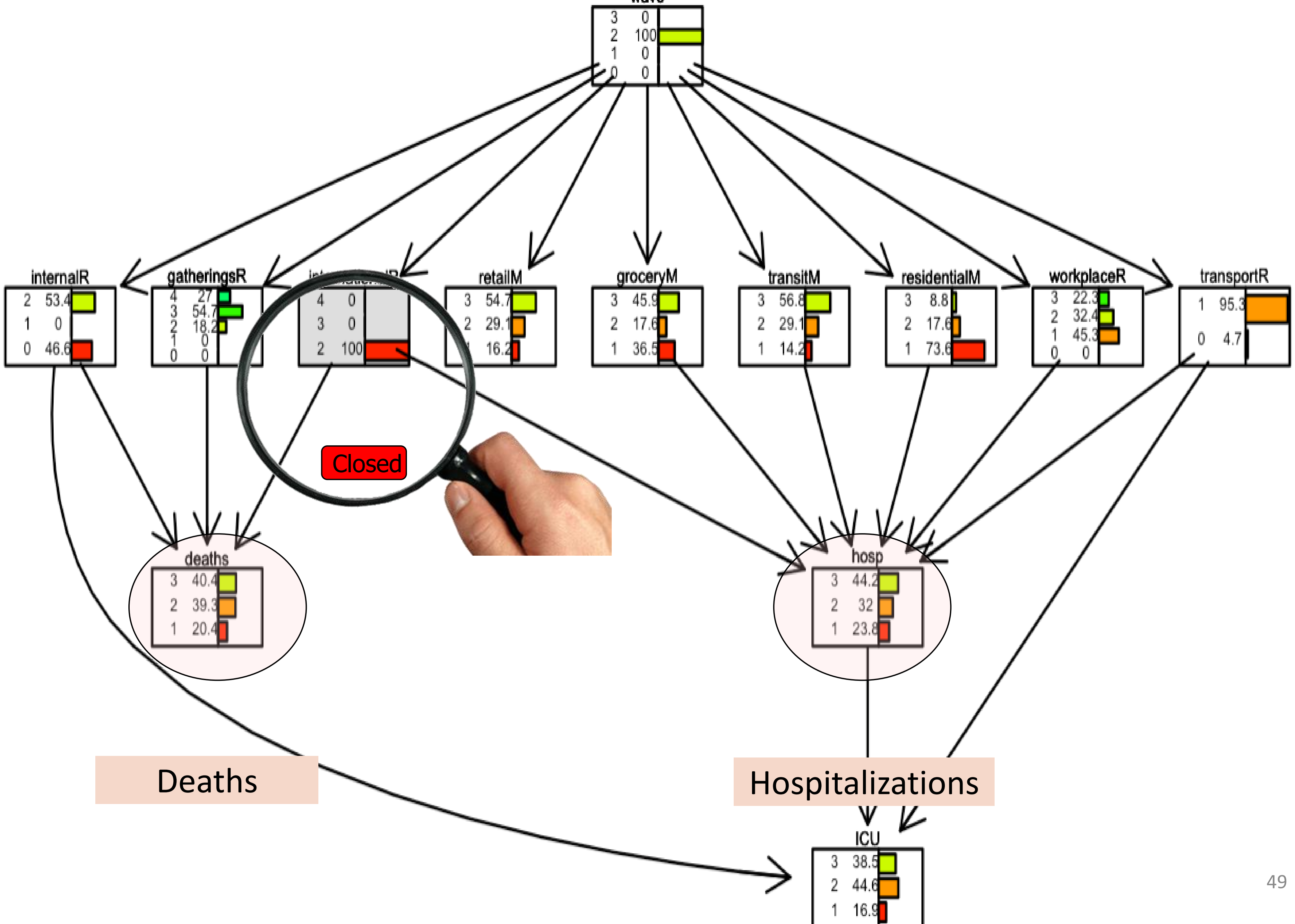


Hosp Lag 10, ICU Lag 15, Deaths Lag 20

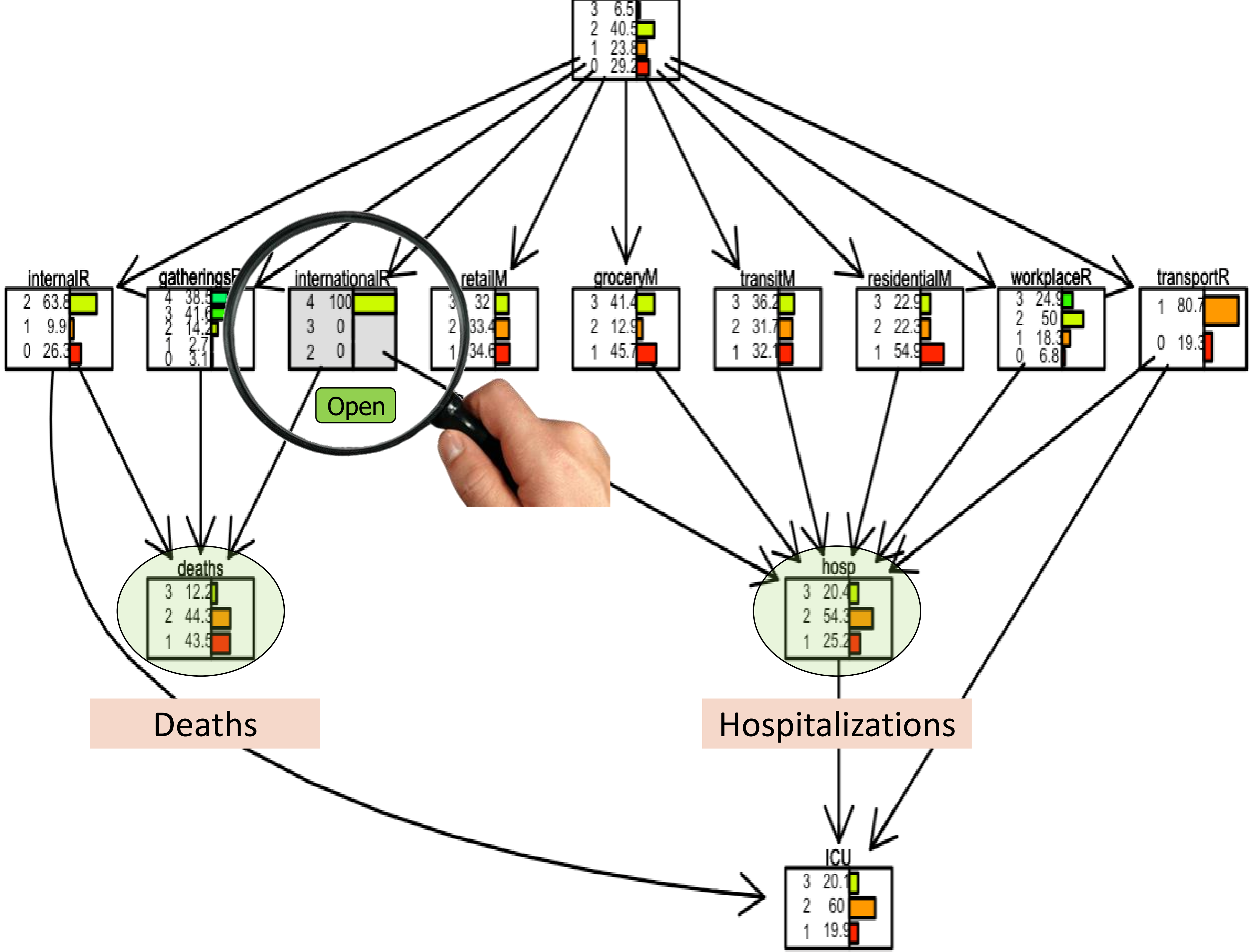


Regressions	Estimate	SE	Prob>|Z|
workplaces → [hosp_lag10]	0.0051733	0.0013351	0.0001*
workplaces → [death_lag20]	-0.001852	0.0012394	0.1350
workplace_closing → [icu_lag15]	-0.030367	0.0278552	0.2756
workplace_closing → [hosp_lag10]	0.0667816	0.0270191	0.0134*
wave → workplaces	3.19686	0.5874759	<.0001*
wave → workplace_closing	0.0754524	0.0217294	0.0005*
wave → transport_closing	0.0837236	0.0126287	<.0001*
wave → transit_stations	2.9688032	0.4265175	<.0001*
wave → [icu_lag15]	0.1807456	0.0274846	<.0001*
wave → [hosp_lag10]	0.2486583	0.0245819	<.0001*
wave → [death_lag20]	0.4204896	0.1221507	0.0006*
wave → school_closing	-0.316598	0.0169194	<.0001*
wave → retail_and_recreation	3.7666174	0.59323	<.0001*
wave → residential	-1.129711	0.2063731	<.0001*
wave → international_movement_restrictions	0.0842087	0.0199013	<.0001*
wave → internal_movement_restrictions	0.0676107	0.0158063	<.0001*
wave → grocery_and_pharmacy	5.10812	0.5543045	<.0001*
wave → gatherings_restrictions	0.3068871	0.0409729	<.0001*
transport_closing → [icu_lag15]	0.1455176	0.0273859	<.0001*
stringency_index → [icu_lag15]	0.0067435	0.0038163	0.0772
stay_home_restrictions → [hosp_lag10]	0.0472529	0.022111	0.0326*
[icu_lag15] → [death_lag20]	1.1659765	0.2939227	<.0001*
[hosp_lag10] → [death_lag20]	-2.089182	0.410195	<.0001*
retail_and_recreation → [hosp_lag10]	-0.005147	0.0014207	0.0003*
residential → [hosp_lag10]	0.013011	0.004218	0.0020*
parks → [hosp_lag10]	0.0019993	0.0004296	<.0001*
international_movement_restrictions → [icu_lag15]	0.0630446	0.018092	0.0005*
internal_movement_restrictions → [icu_lag15]	0.1338373	0.0667171	0.0449*
internal_movement_restrictions → [hosp_lag10]	0.230179	0.0574105	<.0001*
internal_movement_restrictions → [death_lag20]	0.5626452	0.1869919	0.0026*
grocery_and_pharmacy → [hosp_lag10]	0.000544	0.0008651	0.5294
gatherings_restrictions → [icu_lag15]	-0.063505	0.0407235	0.1189
gatherings_restrictions → [hosp_lag10]	-0.061278	0.035522	0.0845
gatherings_restrictions → [death_lag20]	-0.141549	0.0740964	0.0561
Covariances	**Estimate**	**SE**	**Prob>|Z|**
behave ↔ health	-24.47497	8.5668423	0.0043*
grocery_and_pharmacy ↔ retail_and_recreation	87.867134	9.2979598	<.0001*
residential ↔ workplaces	-10.21344	1.4937324	<.0001*
stay_home_restrictions ↔ internal_movement_restrictions	0.7703725	0.0643261	<.0001*
stringency_index ↔ internal_movement_restrictions	8.9560929	0.7747634	<.0001*
stringency_index ↔ stay_home_restrictions	8.8102202	0.7537571	<.0001*
stringency_index ↔ transport_closing	3.708656	0.3479073	<.0001*
workplaces ↔ parks	-275.3571	24.166104	<.0001*

Structural Equation Models (SEM) to confirm links

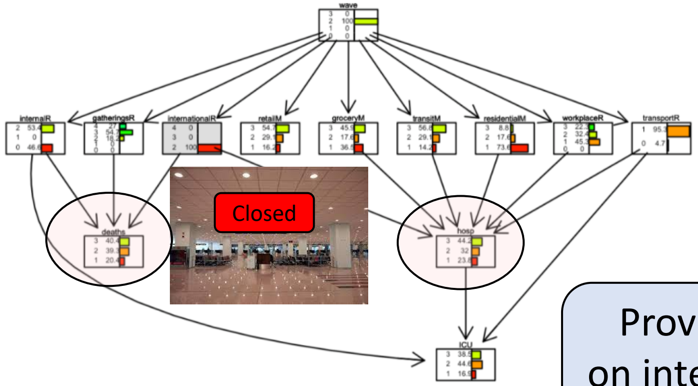


3 0
2 100
1 0
0 0
internalR
2 53.4
1 0
0 46.6
gatheringsR
4 27
3 54.7
2 18.2
1 0
0 0
4 0
3 0
2 100
Closed
retailM
3 54.7
2 29.1
16.2
groceryM
3 45.9
2 17.6
1 36.5
transitM
3 56.8
2 29.1
1 14.2
residentialM
3 8.8
2 17.6
1 73.6
workplaceR
3 22.3
2 32.4
1 45.3
0 0
transportR
1 95.3
0 4.7
deaths
3 40.4
2 39.3
1 20.4
hosp
3 44.2
2 32
1 23.8
Deaths
Hospitalizations
ICU
3 38.5
2 44.6
1 16.9

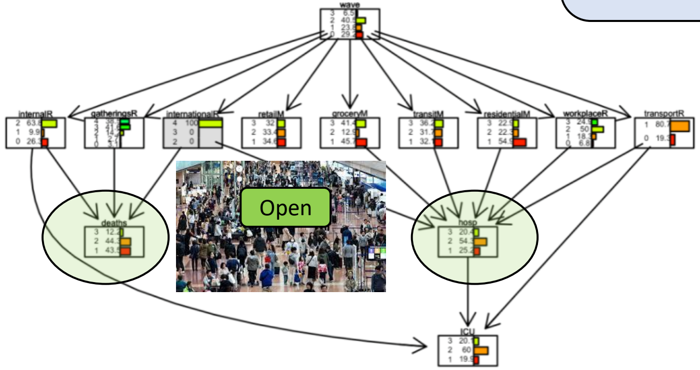

3 6.5
2 40.5
1 23.8
0 29.2
internalR
2 63.8
1 9.9
0 26.3
internationalR
4 100
3 0
2 0
Open
retailM
groceryM
3 41.4
2 12.9
1 45.7
transitM
3 36.2
2 31.7
1 32.1
residentialM
3 22.9
2 22.3
1 54.9
workplaceR
3 24.9
2 50
1 18.3
0 6.8
transportR
1 80.7
0 19.3
deaths
3 12.2
2 44.3
1 43.5
hosp
3 20.4
2 54.3
1 25.2
Deaths
Hospitalizations
ICU
3 20.1
2 60
1 19.9

Validation of method
by applying models to two
Different countries

	ITALY			ISRAEL	
hosp	<5000	>15000	>30000	<250	>1100
icu	<200	>2000		<40	>150
deaths	<100	>500		<5	>20



Provides "what if" scenario
on international transportations
(airports)



Pentaho Download - Lumada Dat ✕ | Pentaho Data Integration: Evaluat ✕ | +

https://www.hitachivantara.com/en-us/products/dataops-software/data-integration-analytics/download-pentaho.html?ecid=ps_eme...

Email link | RonLinks | Facebook | LinkedIn | News Ynet | הארץ | בנק הפועלים – לקו... | Demo: Industry 4.0 | קורונה - לוח בקרה | Wiley StatsRef: Stati... | » | Other bookmarks

HITACHI
Inspire the Next

United States | Community | Get Support | Search

Hitachi Vantara

Solutions | Products | Services | Newsroom | Partners | Company | Contact Us

Home / Products / DataOps Software / Data Integration and Data Analytics / Pentaho Download - Lumada Data Integration and Analytics

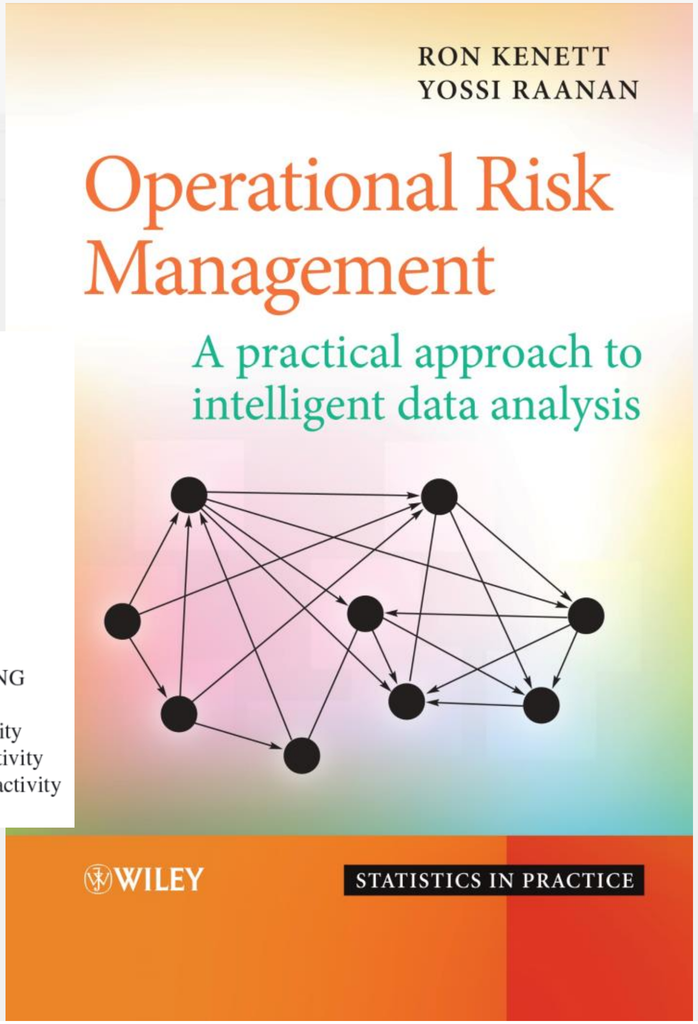
Extract – Transform – Load (ETL)

Lumada Data Integration and Analytics powered by Pentaho

Download Pentaho – Turn Data into Actionable Insights.

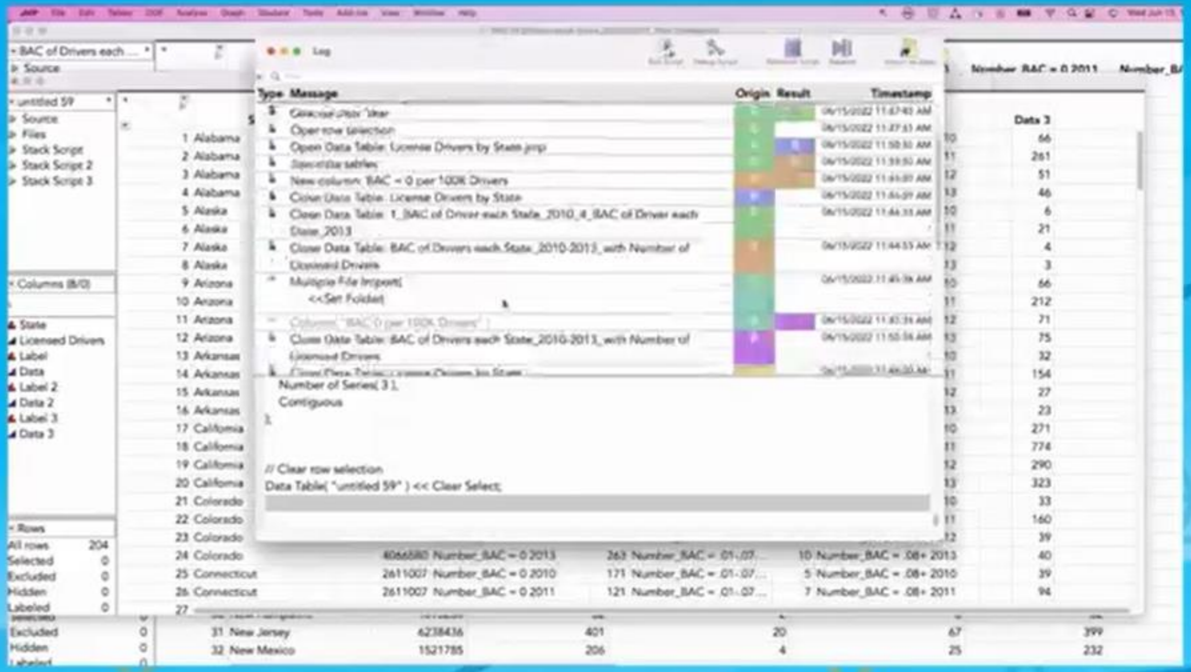
Pentaho Data Integration: Evaluation (PDI1028S)

(60 mins) Self-paced, interactive online training with virtual lab environment for hands-on practice



Mastering JMP®

Automated Analytics: Automating ETL Made Easy with Workflow Builder



jmp STATISTICAL DISCOVERY

Development > **Software Engineering** > **ETL**

Writing production-ready ETL pipelines in Python / Pandas

Learn how to write professional ETL pipelines using best practices in Python and Data Engineering.

4.2 ★★★★☆ (555 ratings) 4,381 students

Created by Jan Schwarzlose

Last updated 7/2022 English English [Auto]

₪59.90 ~~₪279.90~~ 79% off

Add to cart

Buy now

30-Day Money-Back Guarantee

What you'll learn

https://link.springer.com/article/10.1007/s10845-021-01817-9

https://www.taylorfrancis.com/books/mono/10.1201/9780429292835/pragmatic-programmer-machine-learning-marco-scutari-mauro-malvestio?context=ubx&refId=4c1ab861-6c68-40d7-be0f-6cf3717fe7e6

Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-021-01817-9

Challenges of modeling and analysis in cybermanufacturing: a review from a machine learning and computation perspective

SungKu Kang[1] · Ran Jin[1] · Xinwei Deng[1] · Ron S. Kenett[2]

Received: 31 December 2020 / Accepted: 19 July 2021

Abstract

In Industry 4.0, smart manufacturing is facing its next stage, cybermanufacturing, founded upon advanced communication, computation, and control infrastructure. Cybermanufacturing will unleash the potential of multi-modal manufacturing data, and provide a new perspective called computation service, as a part of service-oriented architecture (SOA), where on-demand computation requests throughout manufacturing operations are seamlessly satisfied by data analytics and machine learning. However, the complexity of information technology infrastructure leads to fundamental challenges in modeling and analysis under cybermanufacturing, ranging from information-poor datasets to a lack of reproducibility of analytical studies. Nevertheless, existing reviews have focused on the overall architecture of cybermanufacturing/SOA or its technical components (e.g., communication protocol), rather than the potential bottleneck of computation service with respect to modeling and analysis. In this paper, we review the fundamental challenges with respect to modeling and analysis in cybermanufacturing. Then, we introduce the existing efforts in computation pipeline recommendation, which aims at identifying an optimal sequence of method options for data analytics/machine learning without time-consuming trial-and-error. We envision computation pipeline recommendation as a promising research field to address the fundamental challenges in cybermanufacturing. We also expect that computation pipeline recommendation can be a driving force to flexible and resilient manufacturing operations in the post-COVID-19 industry.

Keywords Computation pipelines · Cybermanufacturing · Industry 4.0 · Machine learning · Manufacturing modeling and analysis