

# How to provide light to COVID data by means of FCA

## RealDataFCA Workshop

D. López-Rodríguez<sup>1</sup>, P. Cordero<sup>1</sup>, M. Enciso<sup>2</sup>, Á. Mora<sup>1</sup>

<sup>1</sup>Dep. de Matemática Aplicada

<sup>2</sup>Dep. de Lenguajes y Ciencias de la Computación

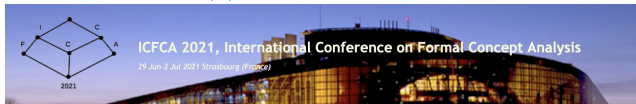


UNIVERSIDAD  
DE MÁLAGA



ICFCA 2021, International Conference on Formal Concept Analysis

29 Jun-2 Jul 2021 Strasbourg (France)



# Table of Contents

Motivation

Data sources

Data curation and preprocessing

Use of FCA

Some results

Conclusions and future works

# Motivation

- A global pandemic, the first in a century, that has led to:
  - Different restriction measures taken in the EU
  - Different evolution of the disease
  - Governments and people worried about health AND economy
- We aim at studying the relationship between all these parameters
- So, data has to come from many perspectives: health, economy and government.
- Formal Concept Analysis is the perfect framework to study these heterogeneous data.
- It's a challenging task from the point of view of data analysis.

- **European Centre for Disease Prevention and Control (ECDC)**, an agency of the European Union<sup>1</sup>. Provides data on:
  - Notification of new cases, deaths, hospitalizations, excess mortality, ICU occupancy as parameters defining the evolution of the pandemic at country-level
  - The restriction measures taken in each country, and for which period of time
- **Eurostat** data about the Economy. We have used the repository that this institution has mainly created to measure the impact of the crisis in Europe<sup>2</sup>:
  - Changes in Unemployment rates
  - Changes in the GDP of each country

---

<sup>1</sup><https://www.ecdc.europa.eu/en>

<sup>2</sup><https://ec.europa.eu/eurostat/web/COVID-19/data>

# Data curation and preprocessing

Restrictions data are presented in the following format:

Country	Response_measure	date_start	date_end
Austria	AdaptationOfWorkplace	2020-03-16	2020-04-13
Austria	AdaptationOfWorkplace	2021-04-01	NA
Austria	BanOnAllEvents	2020-11-03	NA
Austria	ClosDaycare	2020-03-16	2020-05-04
Austria	ClosDaycarePartial	2020-11-17	2020-12-07
Austria	ClosHigh	2020-03-16	2020-09-30

We have aggregated data according to the quarter the restrictions were active, and built binary variables to indicate if a given restriction has been or hasn't been active any time in the given quarter (Q2 or Q3).

For each quarter (Q2 and Q3), we have discretized the COVID data (deaths and cumulative cases in percent with respect to the country's population), ICU and hospitalization data, and excess mortality data:

- First, at country-level, the figures were normalized by dividing by the country's total population.
- Second, we have performed interval scaling to determine if the value of a variable in the country was lower than, equivalent, or greater than Europe's average. For instance, if for a country the variable `COV_Q2_ExcessMortality_low` is `TRUE`, this will indicate that the percentual mortality excess in the country is lower than the mean percentual mortality excess in Europe.

For economical data, we have used the relative changes produced from Q1 to Q2 and Q2 to Q3, indicating if the variable has increased or decreased in a given quarter, or it was stable.

Recall that this is applied to the harmonised unemployment rate and the gross domestic product.

- The previous data preprocessing has led us to a binary dataset, that can be considered as a formal context  $\mathbb{K} = (G, M, I)$ , with  $G$  the set of countries,  $M$  the set of measures and variables, and  $I$  the incidence relationship.
- We aim to extract knowledge from this dataset using FCA tools.

Particularly, we want to:

- Find measures that were taken by all or none of the EU countries.
- Find equivalences between variables or measures (that is, attributes  $a$  and  $b$  such that  $I(\cdot, a) = I(\cdot, b)$ )  $\longrightarrow$  Context clarification.
- Represent and inspect the knowledge in the context in the form of the Duquenne-Guigues basis of implications, allowing for symbolic manipulation of the knowledge.
- Find the co-occurrence of measures and variables with interesting attributes  $\longrightarrow$  Closure computation.



# Some results

This formal context has 175 attributes and 25 objects. The dataset has been imported and processed in R using the `fcaR` package<sup>3</sup>.

## Common variables found

- In the 3rd quarter, no country imposed a ban on all events or restrictions on private gatherings or closed restaurants, hotels and entertainment venues
- The GDP in the second quarter decreased in all countries whereas it increased in the third quarter (summer), irrespectively of the restriction measures taken.
- The unemployment rate wasn't stable in both quarters.
- The strong lockdown measures were not needed in the Q3 period.

---

<sup>3</sup><https://malaga-fca-group.github.io/fcaR>

## Equivalent attributes

- Every country that did not close entertainment venues in the 2nd quarter (Q2) of 2020, in the 3rd quarter (Q3), they had to issue a recommendation to stay home for risk groups (older adults, disabled people or people with other pathologies).
- Also, having a high number of deaths in Q2 was equivalent to a high number of deaths in Q3. This is interesting since no matter the measures adopted if the pandemic has led to a high number of deaths during the first stage, in the summer, the number of deaths remains at a high level, and it was also greater than Europe's average.
- A high hospitalization occupancy ratio is shown to be equivalent to the ICU occupation ratio during Q2.

## Use of implications

- The Duquenne-Guigues basis of this formal context has 4086 implications.
- The average size of its LHS and RHS is 11.3 and 1.6 attributes, respectively.
- After using the simplification logic to remove redundancies, the resulting implication set has 4.47 and 1.19 attributes in the average size in LHS and RHS.
- The decrease of the implication size eases its reading and interpretation.

## Influence of Q2 measures in Q2 COVID cases or deaths (short-term effects in health)

NOT\_Q2\_MassGather50  $\Rightarrow$  Q2\_cases are high

Q2\_AdaptationOfWorkplace  $\Rightarrow$  Q2\_cases are medium

Q2\_RestaurantsCafes, Q2\_MasksMandatoryAllSpaces  $\Rightarrow$  Q2\_deaths are medium

## Influence of Q2 measures in Q2 economy (short-term effects in economy)

NOT\_Q2\_WorkplaceClosures  $\Rightarrow$  Q2\_UnemploymentChange is increased

NOT\_Q2\_StayHomeOrder  $\Rightarrow$  Q2\_UnemploymentChange is increased

## Influence of Q2 measures in Q3 COVID (mid-term effects in health)

NOT\_Q2\_StayHomeRiskG, NOT\_Q3\_MasksVoluntaryAllSpaces  
⇒ Q3\_ChangeInHospital is increased

## Influence of Q2 measures in Q3 economy (mid-term effects in economy)

NOT_Q2_BanOnAllEvents	⇒	Q3_UnemploymentChange is decreased
NOT_Q2_StayHomeOrder	⇒	Q3_UnemploymentChange is decreased
Q2_PrivateGatheringRestrictions	⇒	Q3_UnemploymentChange is decreased

## Influence of Q3 measures in Q3 COVID

NOT\_Q2\_StayHomeRiskG, NOT\_Q3\_MasksVoluntaryAllSpaces

⇒ Q3\_ChangeInHospital is increased

NOT\_Q2\_StayHomeRiskG, NOT\_Q3\_MasksMandatoryAllSpaces

⇒ Q3\_ChangeInHospital is increased

Q3\_StayHomeRiskG

⇒ Q3\_inICU is medium, Q3\_ChangeInHospital is decreased,

Q3\_ExcessMortality is equivalent

## Closure of *interesting* attributes

We want to answer the question ‘*What had in common the countries with a low and a high hospital occupancy rate in the second quarter?*’

→ Compute the closures of the corresponding attributes (`Q2_inHospital` is low and `Q2_inHospital` is high).

- The closure of Q2\_inHospital is low is Q2\_ClosHigh, Q2\_ClosPrim, Q2\_MassGather50, Q2\_MassGatherAll, Q2\_OutdoorOver1000, Q2\_EntertainmentVenues, NOT\_Q2\_ClosureOfPublicTransport, NOT\_Q3\_MasksMandatoryAllSpaces, NOT\_Q3\_IndoorOver50, NOT\_Q3\_OutdoorOver50, NOT\_Q3\_Teleworking.
- The high occupancy closure provides the following set of attributes: Q2\_ClosSec, NOT\_Q2\_HotelsOtherAccommodation, NOT\_Q2\_ClosureOfPublicTransport, NOT\_Q3\_Teleworking, Q2\_ExcessMortality is more, Q3\_ExcessMortalityChange is decreased.



# Conclusions and future works

- We have used FCA to address the problem to illustrate the relationships among the government measures, the economic impact and the health situation in the Pandemic.
- We have combined real public data from all countries in the EU stored in different data sets, establishing a data pre-processing task.
- We have made a knowledge representation through the implications using the fcaR package.
- Reducing the original Duquenne-Guigues basis using the Simplification logic eases a further examination of the set of implications, allowing to deduce valuable insights about the COVID disease, its impact and the measures that can be used to deal with the virus.

- We intend to consider all the countries worldwide and collect more attributes for the evolution of the data in 2021. A more comprehensive conceptual scaling should be performed to manage such large dataset.
- We will also build the concept lattice that provides other helpful information and a graphical representation of the information, essential for further navigation among the closure set of attributes.
- We will also address the negation issue (the relation between contrary attributes) by means of mixed (negative and positive) attributes.

# How to provide light to COVID data by means of FCA

## RealDataFCA Workshop

D. López-Rodríguez<sup>1</sup>, P. Cordero<sup>1</sup>, M. Enciso<sup>2</sup>, Á. Mora<sup>1</sup>

<sup>1</sup>Dep. de Matemática Aplicada

<sup>2</sup>Dep. de Lenguajes y Ciencias de la Computación



UNIVERSIDAD  
DE MÁLAGA

