

# Cluster Analysis and Visualization of Administrative Data from the Catholic Church

Claudia Schiavetti & Manuel Soffici



1 Introduction

2 Our data

3 Grouping the data

4 Seeing the data

5 Summary

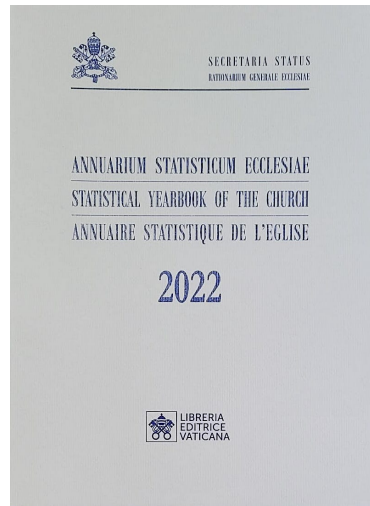
- **Partners:** Prof. Dr. Jean Olivier Nke Ongono (Chair of Global Church Leadership, LMU), Prof. Dr. Andreas Wollbold (Chair of Pastoral Theology, LMU).
- **Students:** Claudia Schiavetti, Manuel Soffici.
- **Supervisor:** Dr. Anna-Carolina Haensch (Statistics and Data Science in Social Sciences and Humanities, LMU).

- Discussion of **data** source, collection method, structure, and quality.
- Introduction to **cluster analysis**, and illustration of our procedure and results.
- Live presentation of the main **web-app** with open testing.

# Outline

- 1 Introduction
- 2 Our data**
- 3 Grouping the data
- 4 Seeing the data
- 5 Summary

- The ***Annuarium Statisticum Ecclesiae*** is compiled yearly by the Central Office of Church Statistics. The publication is based on internal survey data.
- It provides numerous tables of data, mostly at country-level and sometimes with values for past years, covering various Church activities.
- It is structured around seven sections. Our data source is the **2022 edition** of the *Annuarium*, published by Libreria Editrice Vaticana in 2024.



- Ch. 1 - **Area and population**

- ▶ Catholics, inhabitants, parishes, and pastoral centers.
- ▶ Ecclesiastical territories, mission stations, area of diocesan territories.
- ▶ *24 tables, mostly continent-level time series.*

- Ch. 2 - **Workforce for the Apostolate**

- ▶ Priests (diocesan and religious), bishops, and deacons.
- ▶ Priestly ordinations, deaths, and defections by year.
- ▶ *35 tables, mostly time series.*

- Ch. 3 - **Training centres and educational institutions**

- ▶ Religious schools, priestly vocations, seminary dropouts
- ▶ *38 tables, mostly continent-level time series.*

- Ch. 4 - **Practice of religion**

- ▶ Baptisms, first communions, confirmations, and marriages
- ▶ *6 tables, mostly country-level.*

- Ch. 5 - **Welfare institutions**

- ▶ Hospitals, orphanages, nurseries, etc., run by the Church
- ▶ *1 country-level snapshot table.*



- Ch. 6 - **Papal-jurisdiction Institutes of Consecrated Life and Societies of Apostolic Life**

- ▶ Houses, members, novices, and associated laypeople
- ▶ Seminararians, priests, deacons, and religious women
- ▶ *33 tables, mostly global-level time series.*

- Ch. 7 - **Diocesan and religious tribunals**

- ▶ Marriage annulment sentences and appeals
- ▶ *9 country-level snapshot tables.*

## ● Limitations

- ▶ The data is recorded on paper only: there are no official digital records.
- ▶ The employed country scheme does not always follow international recognition.
- ▶ Negligibly small non-zero values are recorded as double dots ".".
- ▶ Some variables have two or more different names across the Yearbook sections.

## ● Advantages

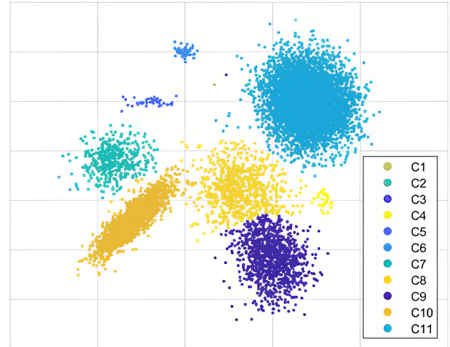
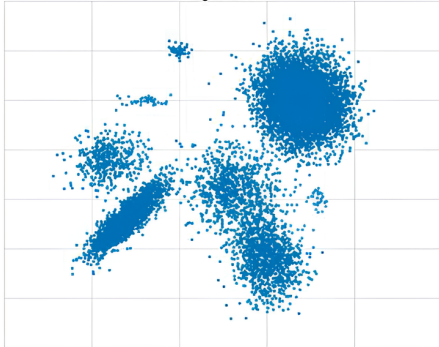
- ▶ No recording errors are (to the best of our knowledge) present.
- ▶ Many items are measured, providing a rich picture of the Church activities.

# Outline

- 1 Introduction
- 2 Our data
- 3 Grouping the data**
- 4 Seeing the data
- 5 Summary

# Cluster analysis basics

Original data





- *How shall we measure the **distance between points** with respect to a given variable?* For instance, we may compute the absolute difference between the values taken by the data points.
- *How shall we define the **distance between clusters**?* We may, for example, define it as the distance (as previously established) between the median points of the clusters.
- *Where shall we "cut the tree"?* I.e., when should the clustering process be stopped? This **model selection** problem is solved with a mix of qualitative and quantitative criteria. Given the distance metrics, there is no absolutely optimal number of clusters.

- **Goal:** Identify groups of countries with similar Catholic and pastoral profiles.
- **Variables Used:** (from Part I of the Annuario)
  - ▶ Population & Territorial Structure
    - ① *Share of Catholics*
    - ② *Catholics per km<sup>2</sup>*
    - ③ *Catholics per pastoral centre*
  - ▶ Pastoral Workforce Composition
    - ① *Share of diocesan pastors*
    - ② *Non-vacant parishes administered by non-pastor priests per Catholic*
    - ③ *Share of non-vacant parishes entrusted to religious women or laypeople*
  - ▶ Pastoral Capacity & Coverage
    - ① *Parishes entirely vacant per Catholic*
    - ② *Catholics per priest*

# Design 1: The Territorial Division of the Church (2)

Design 1: Correlation Heatmap (Population & Territory)

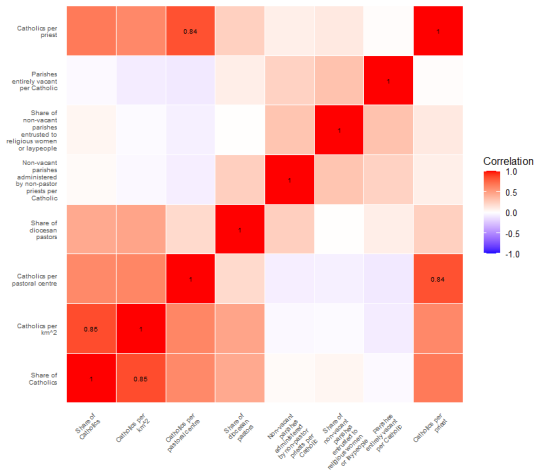


Figure: Correlation Heatmap (1)



- **Goal:** Identify groups of countries with similar clergy dynamics and sacramental practice.
- **Variables Used:** (from Part II, III and IV of the Annuario)
  - ▶ Clergy dynamics
    - ① *Yearly ordinations, deaths, defections, and net balance of diocesan clergy (shares of priests at Jan 1).*
  - ▶ Vocation pipeline
    - ① *Vocation rates per 100k inhabitants and per 100k Catholics*
    - ② *Candidates (philosophy+theology) per 100 priests*
    - ③ *Candidates for diocesan and religious clergy per Catholic*
  - ▶ Sacramental practice
    - ① *Baptisms: infant/adult per Catholic; baptisms per inhabitant*
    - ② *Marriages: share of mixed marriages; mixed marriages per inhabitant; marriages between Catholics per Catholic*
    - ③ *Initiation: Confirmations per Catholic, First Communions per Catholic*

# Design 2: Clergy and Ecclesiastical Practice (2)

Design 2: Correlation Heatmap (Clergy & Sacraments)

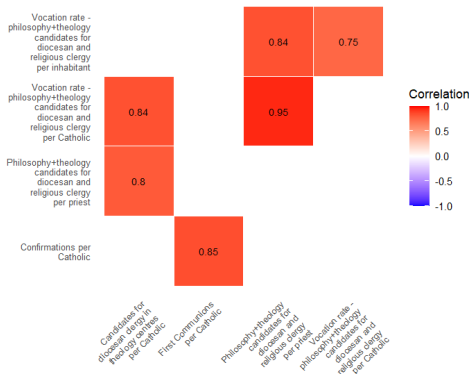


Figure: Correlation Heatmap (2)

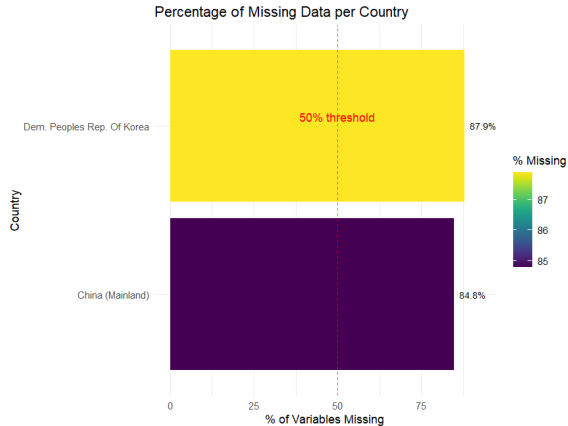


Figure: Missing Values per Country

# Clustering Procedure

- *Distance*: Euclidean on standardized variables
- *Algorithm*: Hierarchical clustering (Ward.D2)
- *k selection*: Calinski–Harabasz (CH) index + informal criteria

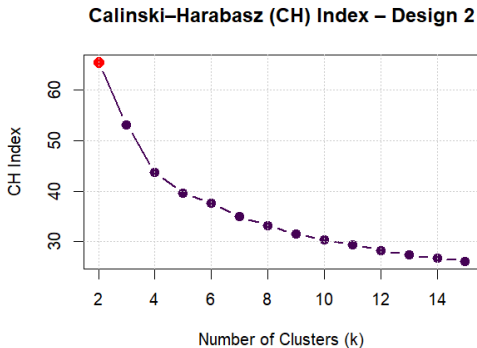
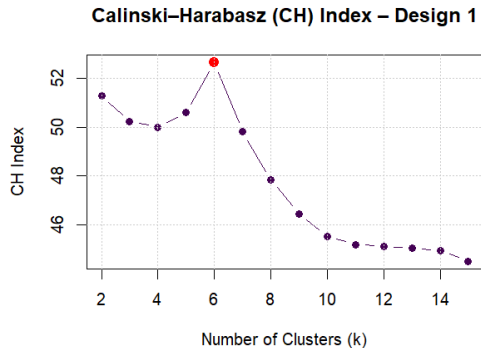


Figure: Calinski–Harabasz index for the two clustering designs.

Table: Medoid: **Zimbabwe**

Indicator	Value
Catholics as % of population	12.7
Catholics per km <sup>2</sup>	5.52
Catholics per pastoral centre	907
Share of diocesan pastors (%)	77.9
Non-vacant parishes by non-pastor priests (per 100k Catholics)	0.42
Share of non-vacant parishes entrusted to religious women or laypeople (%)	0.0
Parishes entirely vacant (per 100k Catholics)	0.00
Catholics per priest	3,431

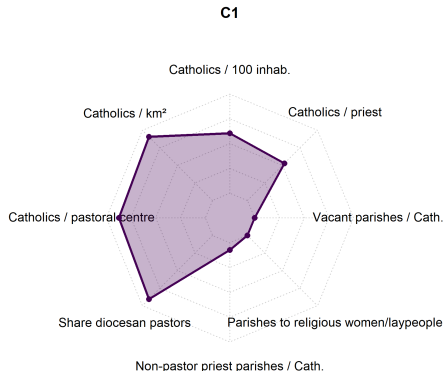


Figure: Cluster 1 profile (Design 1).

Table: Medoid: **Chile**

Indicator	Value
Catholics as % of population	74.0
Catholics per km <sup>2</sup>	19.40
Catholics per pastoral centre	3,699
Share of diocesan pastors (%)	77.9
Non-vacant parishes by non-pastor priests (per 100k Catholics)	0.87
Share of non-vacant parishes entrusted to religious women or laypeople (%)	0.6
Parishes entirely vacant (per 100k Catholics)	0.02
Catholics per priest	8,215

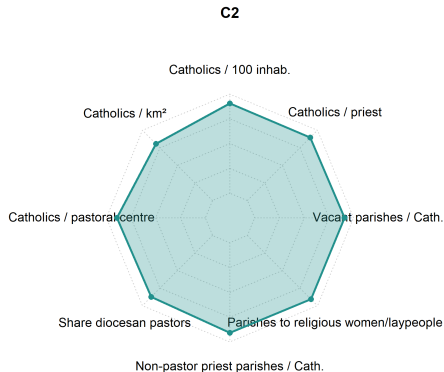


Figure: Cluster 2 profile (Design 1).

Table: Medoid: **Western Sahara**

Indicator	Value
Catholics as % of population	0.03
Catholics per km <sup>2</sup>	0.00
Catholics per pastoral centre	115
Share of diocesan pastors (%)	0.0
Non-vacant parishes by non-pastor priests (per 100k Catholics)	0.00
Share of non-vacant parishes entrusted to religious women or laypeople (%)	0.0
Parishes entirely vacant (per 100k Catholics)	0.00
Catholics per priest	172

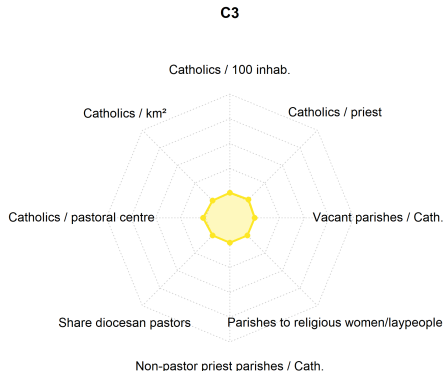


Figure: Cluster 3 profile (Design 1).

C1

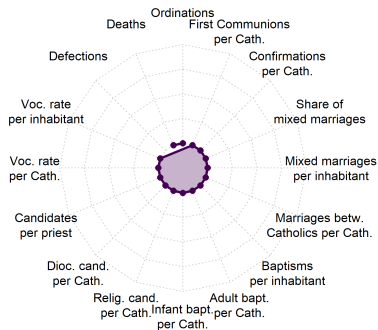


Table: Medoid: **Niue**

Indicator	Value
Ordinations (share, %)	0.0
Deaths (share, %)	0.0
Defections (share, %)	0.0
Vocation rate (per 100k inh.)	0.00
Baptisms (per inhabitant)	0.0000
Marriages (per Catholic)	0.0000
Share of mixed marriages (%)	0.0
Confirmations (per Catholic)	0.0000
First Communions (per Catholic)	0.0000



C2

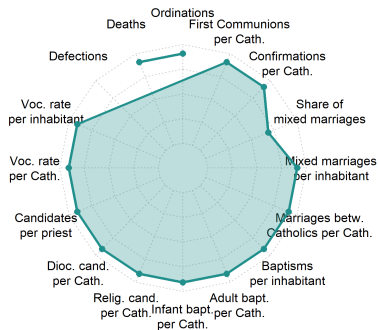


Table: Medoid: **Costa Rica**

Indicator	Value
Ordinations (share, %)	2.5
Deaths (share, %)	0.8
Defections (share, %)	0.2
Vocation rate (per 100k inh.)	3.15
Vocation rate (per 100k Catholics)	3.94
Candidates for clergy per 100 priests	19.88
Infant baptisms (per Catholic)	0.00538
Adult baptisms (per Catholic)	0.00057
Share of mixed marriages (%)	7.7
First Communions (per Catholic)	0.00669

C3

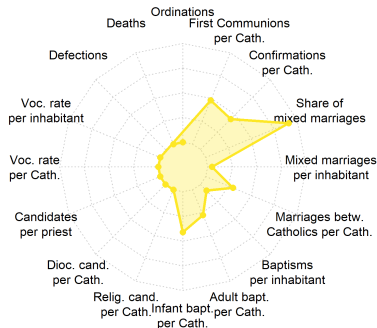


Table: Medoid: **Cayman Islands**

Indicator	Value
Ordinations (share, %)	0.0
Deaths (share, %)	0.0
Defections (share, %)	0.0
Vocation rate (per 100k inh.)	0.00
Infant baptisms (per Catholic)	0.00622
Adult baptisms (per Catholic)	0.00044
Baptisms (per inhabitant)	0.00087
Marriages (per Catholic)	0.00100
Share of mixed marriages (%)	18.2
First Communions (per Catholic)	0.00513

# Outline

- 1 Introduction
- 2 Our data
- 3 Grouping the data
- 4 Seeing the data**
- 5 Summary

**<https://tinyurl.com/CatholicAnnuario>**

# Outline

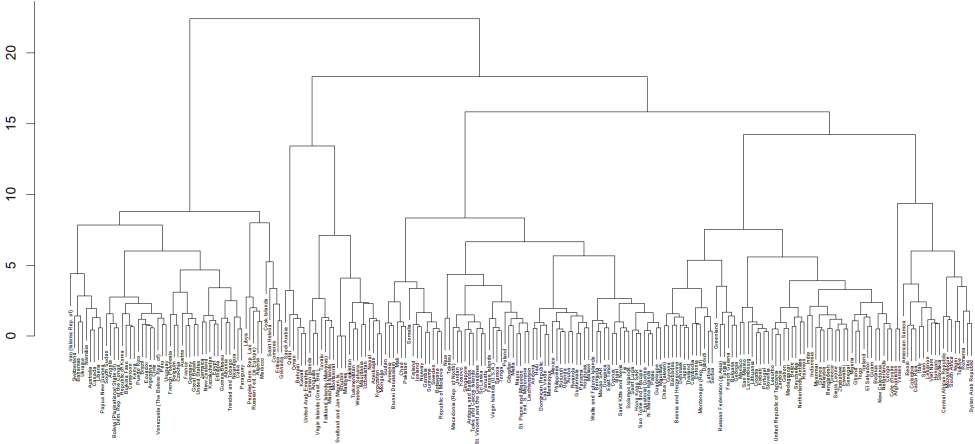
- 1 Introduction
- 2 Our data
- 3 Grouping the data
- 4 Seeing the data
- 5 Summary**

- Our primary achievement is the development and deployment of the three web-apps built with Shiny.
- An exploratory cluster analysis on key metrics of interest was also carried out.
- The preliminary data cleaning (in small part), merging, and harmonization were, of course, necessary for both.

**Thank you for being here!**

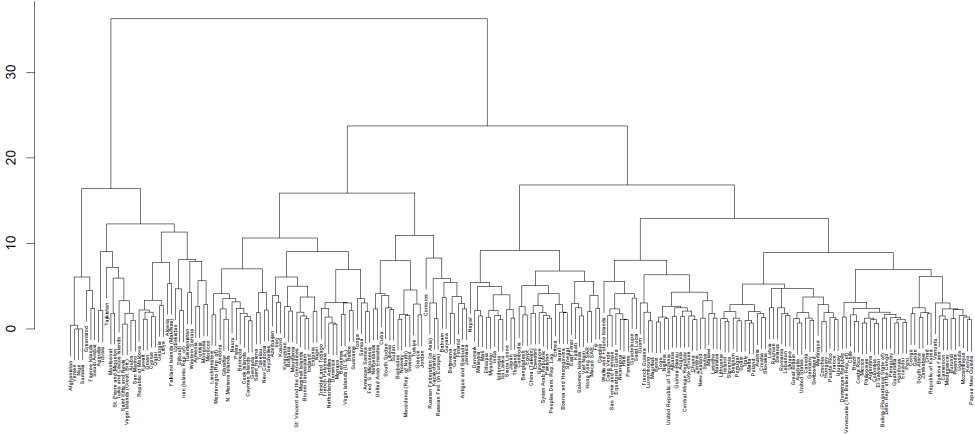
And special thanks to Prof. Dr. Ongono and Prof. Dr. Wollbold.

HC dendrogram (Euclidean, Ward.D2)





HC dendrogram (Euclidean, Ward.D2)



**[github.com/ClaudiaSchiavetti/Consulting\\_Catholic\\_Church](https://github.com/ClaudiaSchiavetti/Consulting_Catholic_Church)**

**[github.com/manuki-stats/church-data-world-map](https://github.com/manuki-stats/church-data-world-map)**