UPPSALA
UNIVERSITET

Clustering as a tool for identifying drought-prone regions

A Swedish example

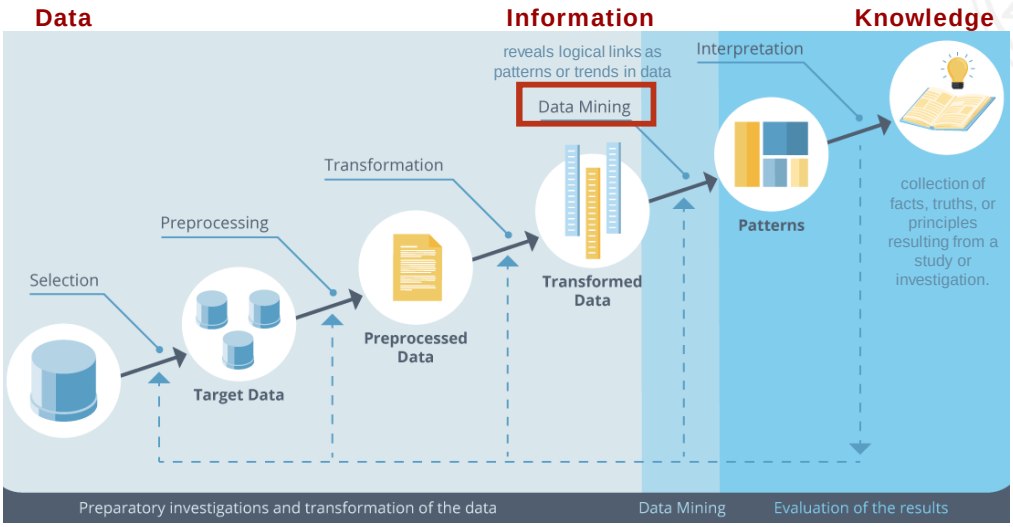
Claudia Teutschbein
Andrijana Todorović and Thomas Grabs

UPPSALA
UNIVERSITET

INTRODUCTION

- **information age**: we collect data faster than it can be processed
- data acquisition has grown exponentially since the introduction of mainframe computers
- improved methodologies for extracting information from data in almost every aspect of life

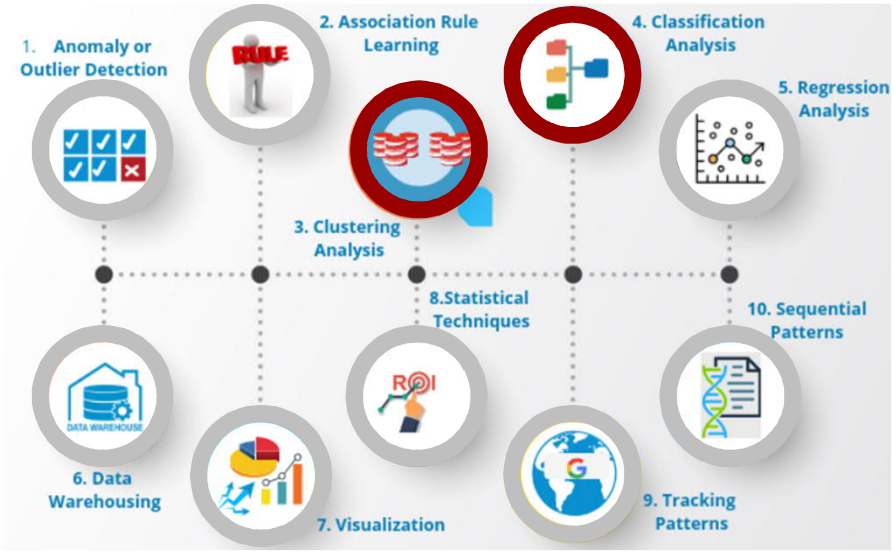
INTRODUCTION



Sources: https://www.industry-analytics.de/datenanalyse-und-data-mining-verfahren-im-ueberblick/ and https://job-wizards.com/en/wp-content/uploads/2019/08/JW_graphic_fs_DATA_MINING_KW29_EN.png

UPPSALA UNIVERSITET

INTRODUCTION

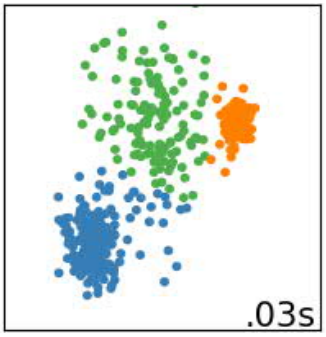


Sources: https://www.bizprospex.com/top-10-data-mining-techniques-for-business-success/

UPPSALA UNIVERSITET

UNDERSTANDING CLUSTERING

- Clustering is a data mining method that involves grouping of data points
- Each data point is classified into a specific group, called a cluster.



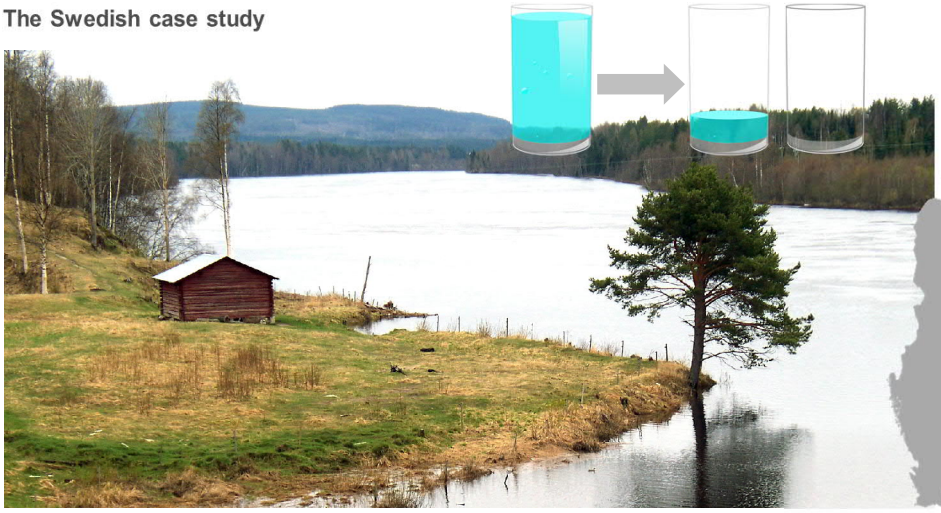
- Each cluster has distinct characteristics.
- Data points within the same group should be as similar as possible
- Data points in different groups should have highly dissimilar properties and/or features.
- Various clustering approaches can be used, all coming with some advantages or disadvantages.

UPPSALA UNIVERSITET

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html

CLUSTERING IN PRACTICE

The Swedish case study



UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

The Swedish case study

1996

Grundvattnet rekordlågt

Göteborgs-Posten | 21 mars 1996 | 483 ord

2003

Torka påverkar dricksvattnet

Hudiksvalls Tidning | 24 juli 2003 09:02 | 271 ord

2018

Många söker stöd för skadad skörd

Land Lantbruk | 28 sep. 2018 | sida 10 | 54 ord

2019

Risk of water shortages in 2019

Too little rain has led to low groundwater levels in parts of the country. There is a risk that it will be a difficult summer with water shortages in several counties, according to the Geological Survey of Sweden, SGU.

2020

FÖRETAGANDE UPPDATERAD: 19 MAJ 2020

Rapsen skadad av frost och torka – men hoppet är inte ute

2021

Lantmännens skördeprognos: Värme och torka i juni sänker skörden i Halland, Skåne och Blekinge

TIS, JUL 13, 2021 07:00 CET

Report this

Rekordsolig torka **2022**

UPPDATERAD 5 SEPTEMBER 2022 PUBLICERAD 1 APRIL 2022

UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

The Swedish case study

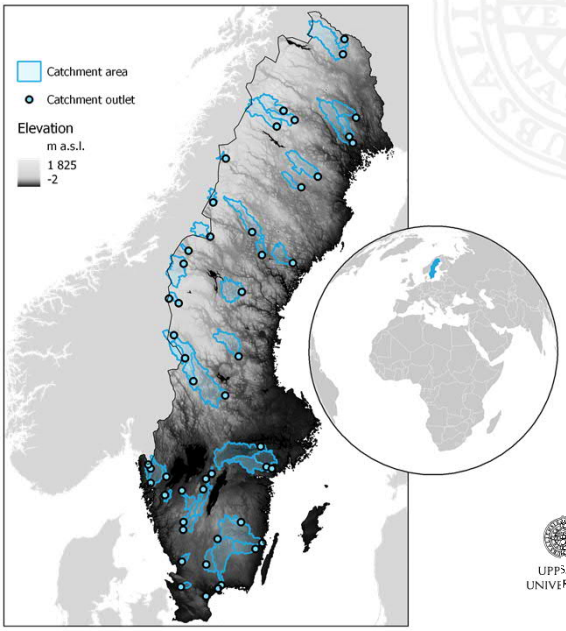
- Have streamflow droughts become more common in Sweden?
- Are all regions affected to the same extent?

UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

The Swedish data set

- Streamflow measured over 60 years at 50 stations across Sweden
- Different catchment properties (e.g., area, topography, landuse)
- Different climate conditions



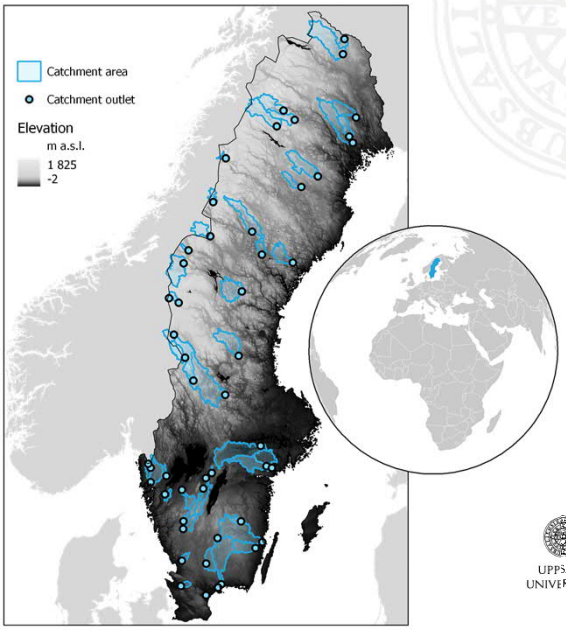
UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Problem

It can be challenging to derive meaningful insights when dealing with a diverse range of catchments.

How can we move beyond analyzing individual catchments and instead identify common patterns that can be generalized?

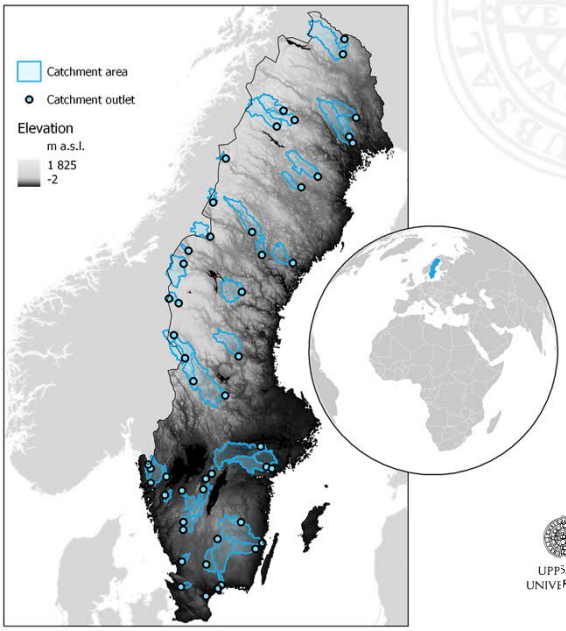


UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Motivation

We let the data "speak" to see which catchments behave similarly and to evaluate emerging spatial patterns.



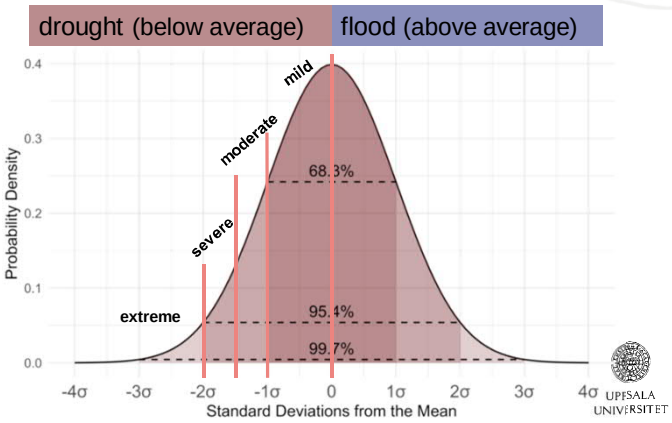
UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Clustering based on drought behaviour

Standardize streamflow index (SSI)

- Represents anomalies from a normal situation in a standardized way
- Calculated based on monthly data
- Transforms data into a normal distribution
- For each month, it is assessed how much the value deviates from the mean (how many stds)

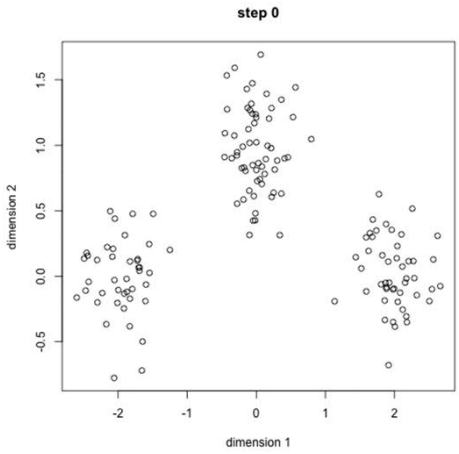


Lloyd-Hughes & Saunders 2002. A drought climatology for Europe. https://doi.org/10.1002/joc.846

CLUSTERING IN PRACTICE

Clustering with K-Means

- Probably most well-known algorithm
- Pretty fast
- Requires pre-selection of number of groups

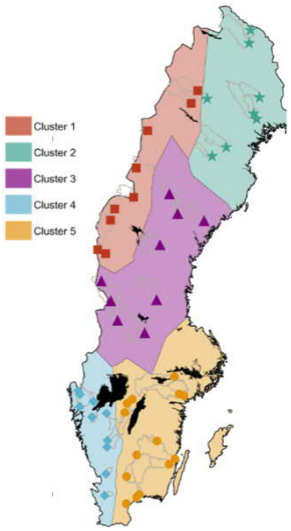


https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68

CLUSTERING IN PRACTICE

Resulting Clusters

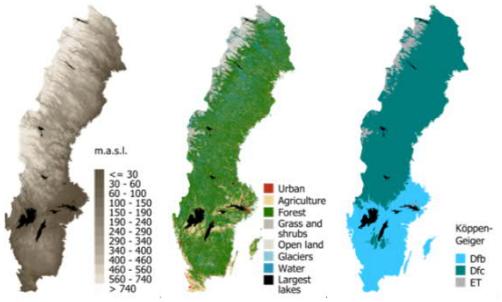
- 5 geographically distinct regions emerged from the clustering
- These clusters varied also in geographic, hydroclimatic and land-cover characteristics

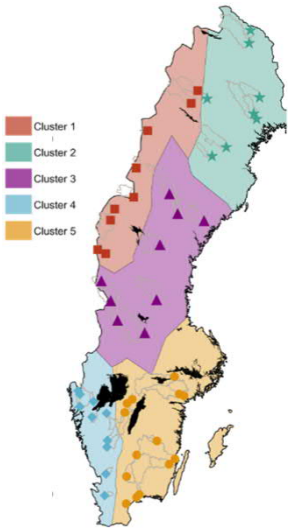


UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Resulting Clusters





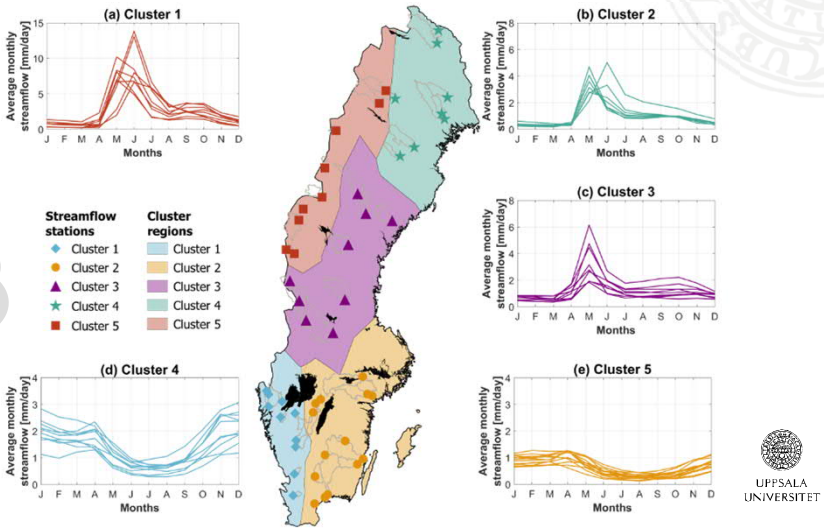
UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Resulting Clusters

- These clusters varied also in their hydrological regimes.

Clustering made sense from a physical & process-based perspective

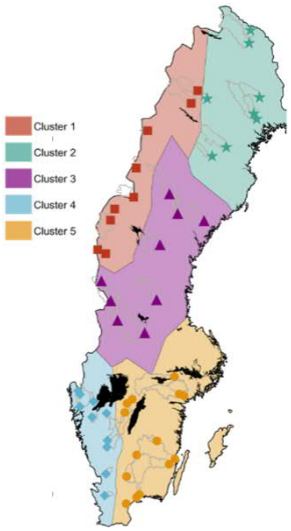


UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Cluster-wise computations

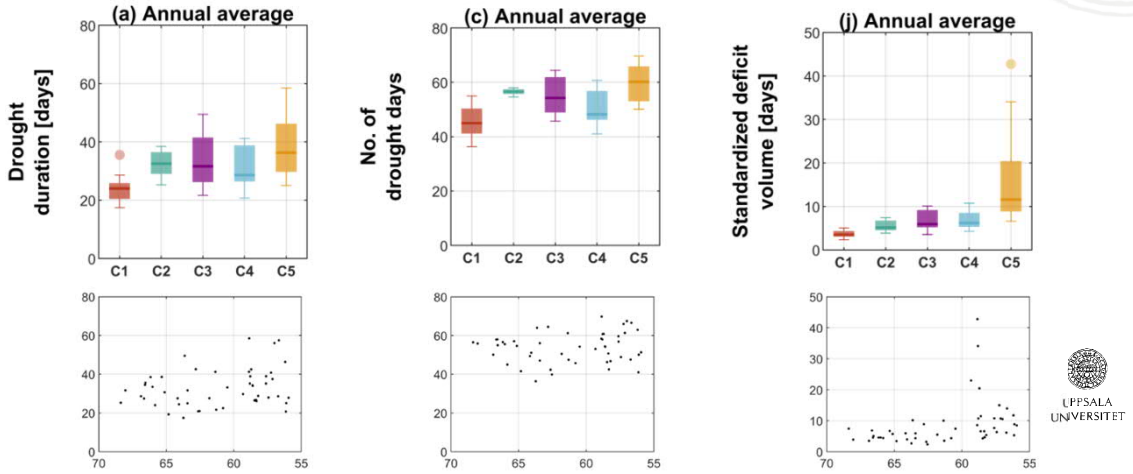
- 17 drought statistics
- Mann-Kendall's test and Sen's slope estimates were computed to detect potential significant trends



UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Drought Characteristics per cluster

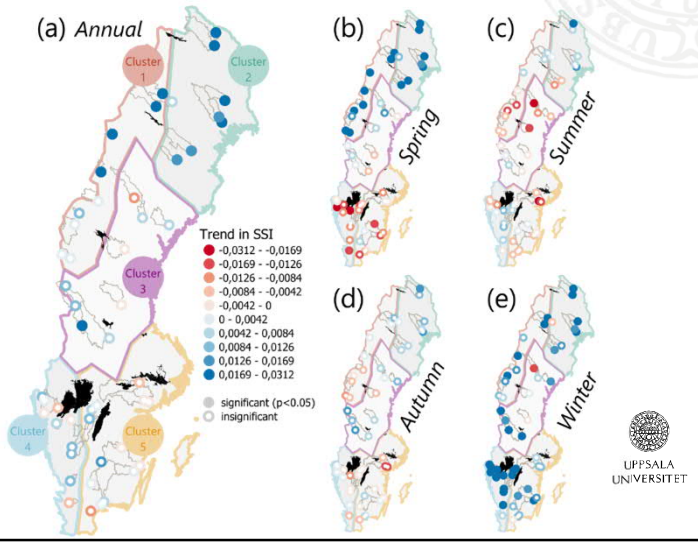


UPPSALA UNIVERSITET

CLUSTERING IN PRACTICE

Trends per cluster

- Can be difficult to detect patterns
- Plotting by region revealed considerable differences



UPPSALA UNIVERSITET

CONCLUSION

- **Knowledge is power:** clustering can give you more knowledge!
- **Clustering can** be used to
  - Simplify your data and make it easier to analyze
  - Identify hydrological similarities among catchments
  - Identify areas at risk of floods or droughts and inform planning strategies
  - Uncover the underlying hydrological processes that influence streamflow patterns
- **Careful:** clustering can help efficiently evaluate past streamflow drought behavior and identify drought-prone areas, but it's important to verify and interpret the patterns with process-based knowledge.

UPPSALA UNIVERSITET

ELSEVIER

Journal of Hydrology: Regional Studies
Volume 42, August 2022, 101171

Streamflow droughts in Sweden: Spatiotemporal patterns emerging from six decades of observations

Claudia Teutschbein [a], Beatriz Quesada Montano [a], Andrijana Todorović [b], Thomas Grabs [a]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ejrh.2022.101171

Get rights and content

SCAN ME

UPPSALA UNIVERSITET

Link: https://doi.org/10.1016/j.ejrh.2022.101171