KU LEUVEN

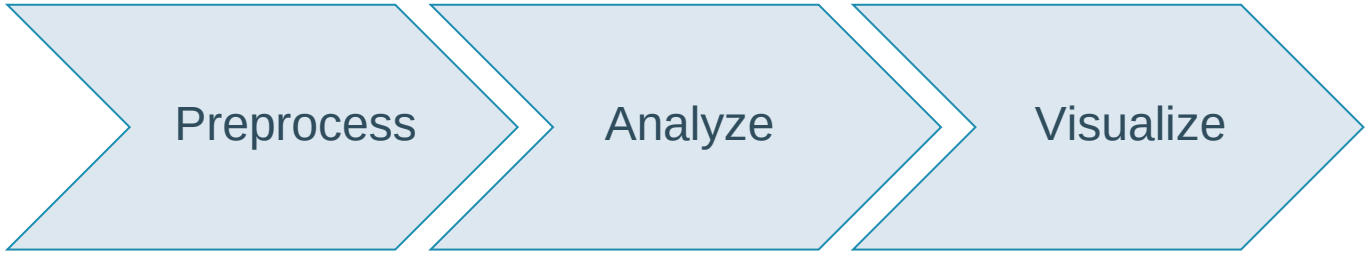
Modern Data Analytics: Presentation Group Sweden

Kinga Nowak, Sara Rutten, Emma Kessenich, Ilias Willems, Fan Huang & Peter Day

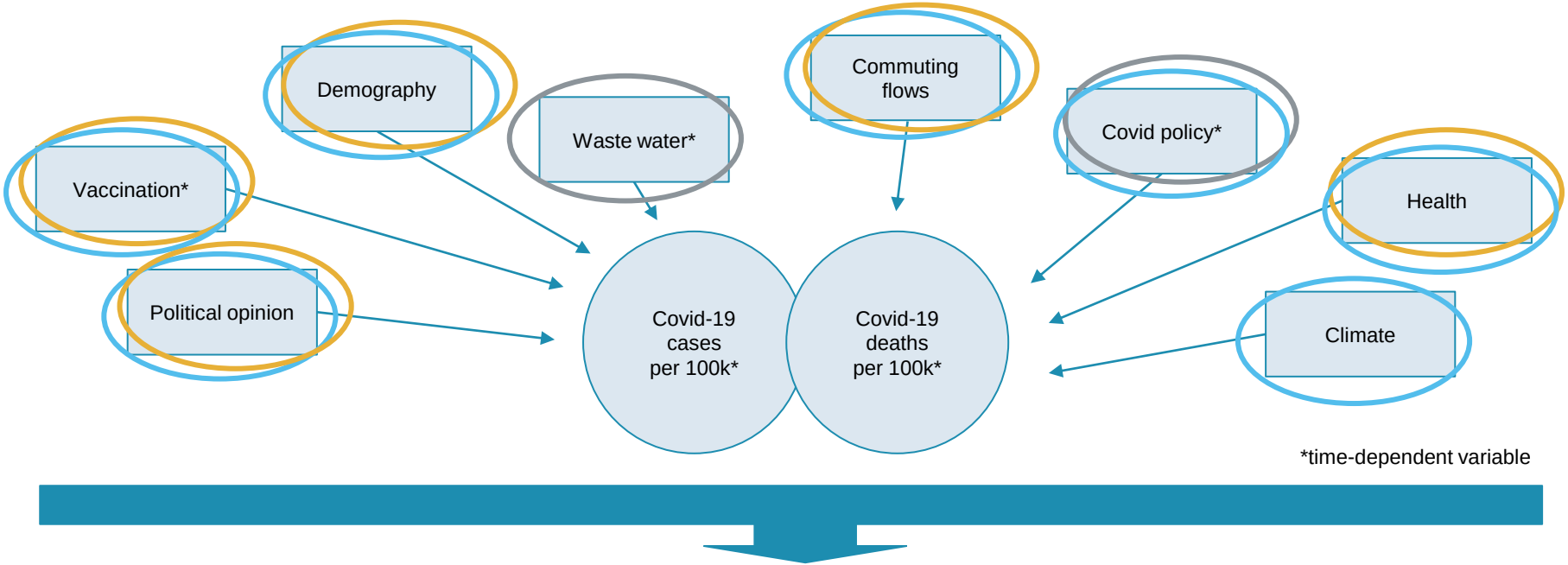
Our project: Covid-19 in the United States

RQ: How did Covid-19 evolve in the US between 21-01-2020 and 20-04-2022?

- RQ 1: Which clusters of states/counties evolved similarly and which features influence these clusters?
- RQ 2: How can we predict the evolution of Covid-19 cases?
- RQ 3: Which counties in the US can be classified as hotspots and which features explain this classification?



Preprocessing: Data sources

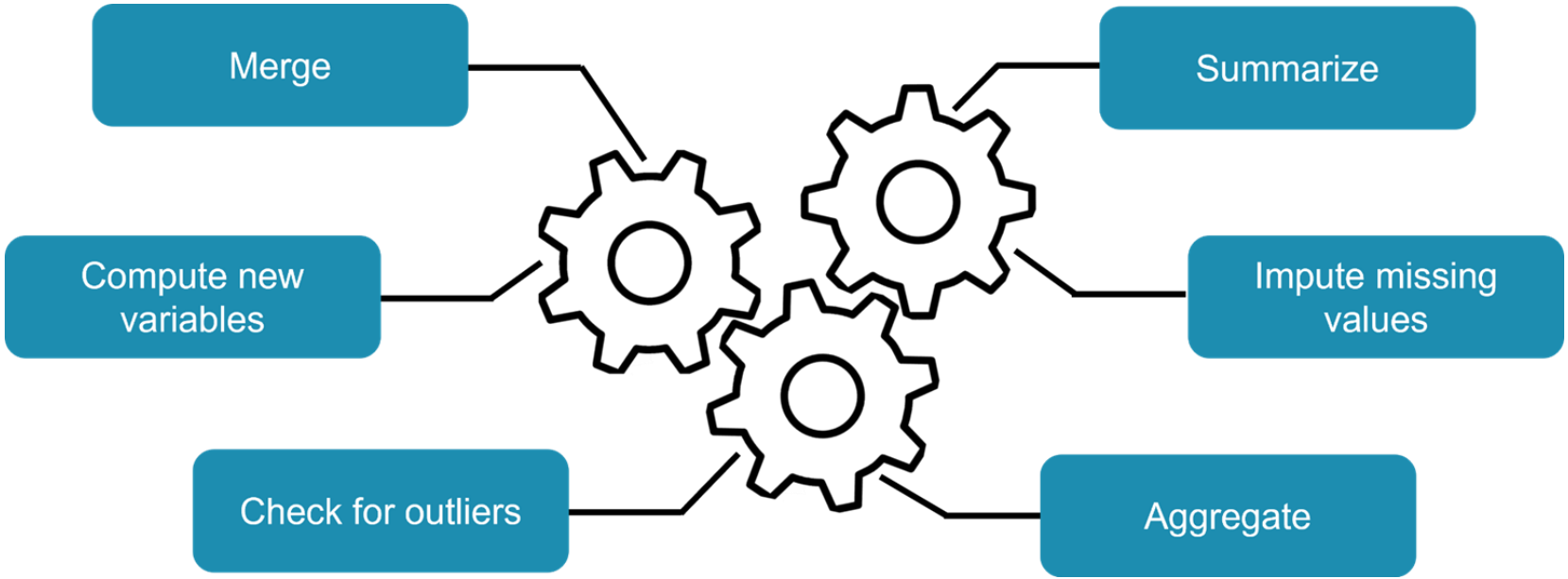


RQ 1: Communities
measured per state & period (LPA)
measured per county (Kmeans)

RQ 2: Prediction
measured per day (linear model)
measured per state & week (SVM)

RQ 3: Hotspots
measured per county & period (random forest)

Preprocessing: General steps



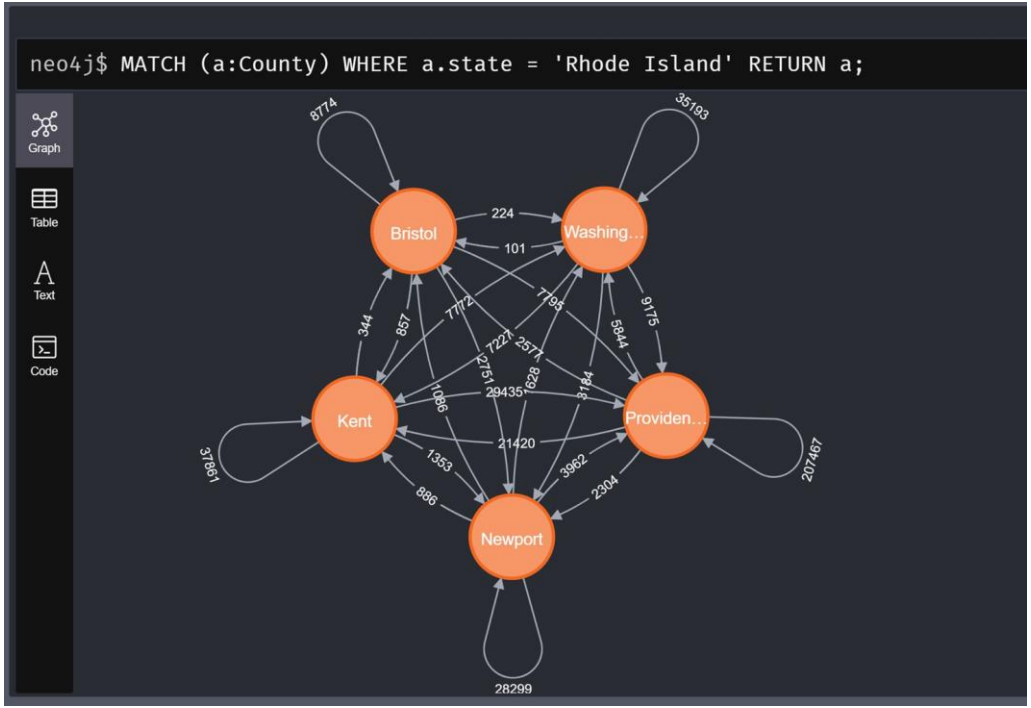
Preprocessing: Commuting flows

County level

- Closeness centrality
- Betweenness centrality
- Pagerank score

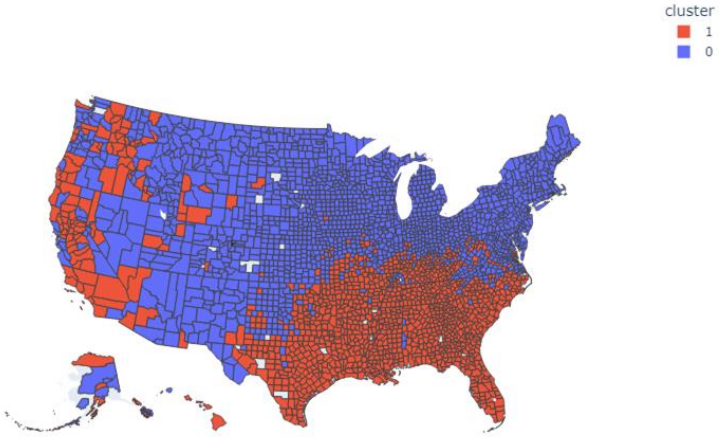
State level

- Degree centrality



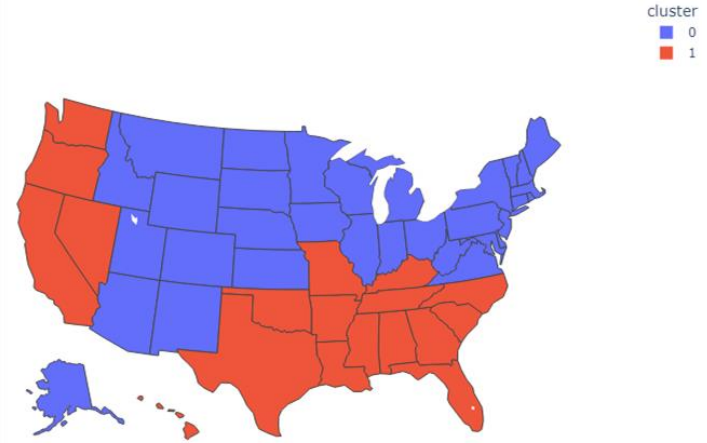
Sub-question 1: Identify communities

K-means cases



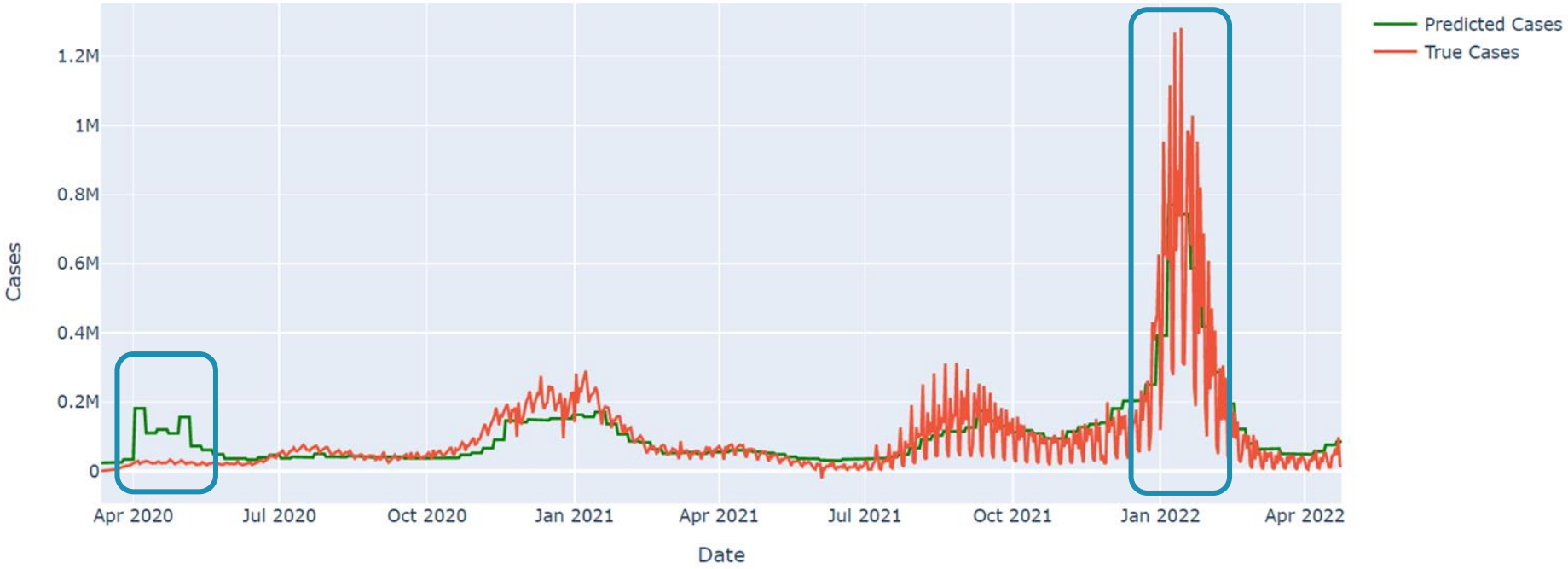
Vaccination rate **
Poverty rate ***
Political votes ***
Airports ***
Life expectancy ***
Uninsured ***

LPA cases (Jun-Dec 2021)

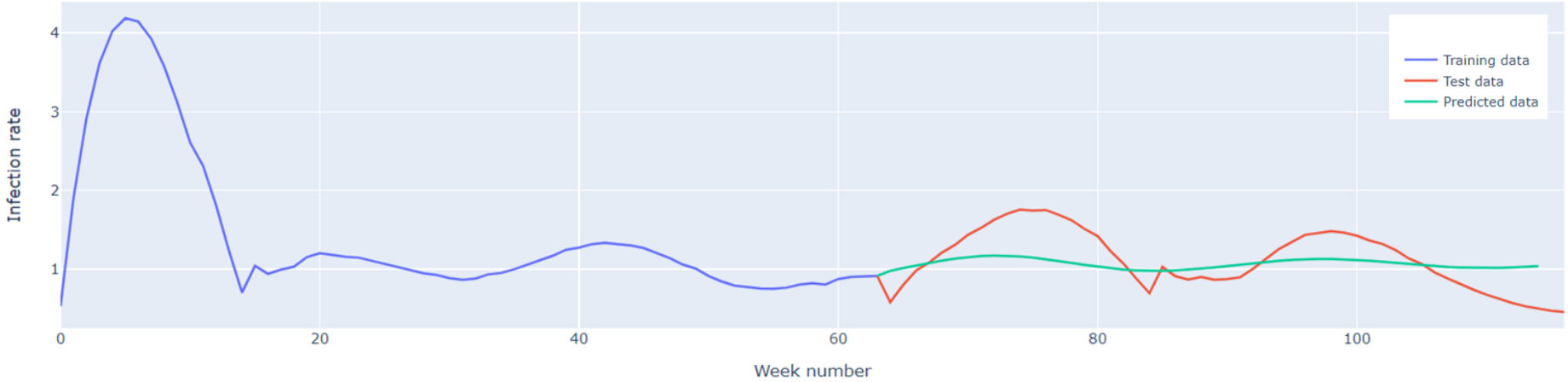


Uninsured *
Poverty rate *
(Vaccination rate *)

Sub-question 2: Prediction – Waste water (linear)

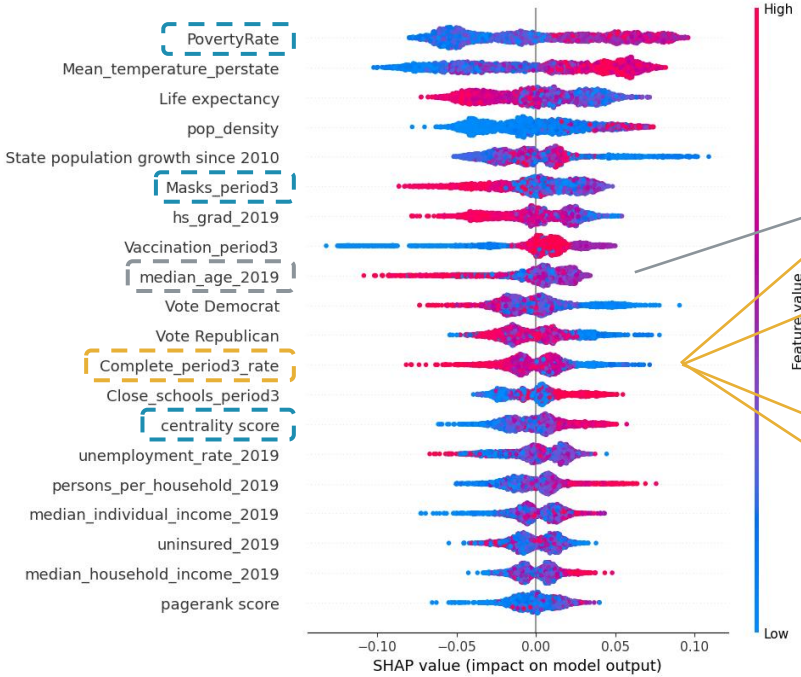


Sub-question 2: Prediction – Covid policy (SVM)

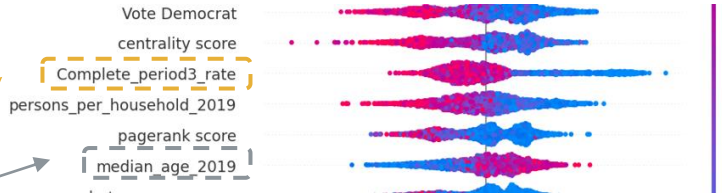


Sub-question 3: Explain hotspots (random forest)

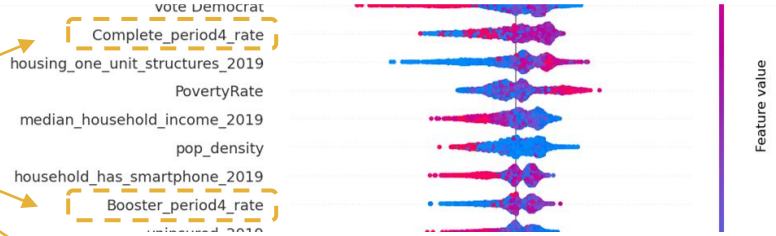
Hotspot in terms of Covid cases (Dec-Jun 21)



Hotspot in terms of Covid deaths (Dec-Jun 21)



Hotspot in terms of Covid cases (Jun-Dec 21)



Hotspot in terms of Covid deaths (Jun-Dec 21)

Our application

Modern Data Analytics project: Covid data

Made by Team Sweden

1. Visualization of the new cases per county on a weekly basis.

In this section, we display the newly reported cases per county and per week. The slider can be used to choose the week for which the data must be shown.

In order to get more insight in how Covid-19 evolves throughout the United States, we can try to cluster counties together based on how they evolve in terms of cases per week. To this end, we applied a K-means and spectral clustering algorithm on the data set. If you'd like to know more about the clusters found by these algorithms and how they were obtained, please select "clusters" in the dropdown box below (it may take a while to load the map).

Covid-19 cases for week 20. (08/Jun/2020 to 14/Jun/2020)

Conclusion

RQ 1: Which clusters of states/counties evolved similarly and which features influence these clusters?

- different clusters in different time periods
- vaccination rate and uninsurance

RQ 2: How can we predict the evolution of Covid-19 cases?

- Wastewater
- SVM: no useful results

RQ 3: Which counties in the US can be classified as hotspots and which features explain this classification?

- different hotspots in different time periods
- different explanatory factors in different time periods

→ vaccination influence on death hotspots