

# Seminar: Introduction to statistical analysis of areal data

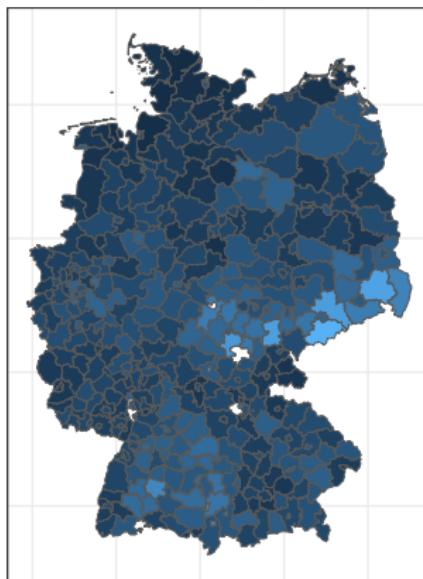
## Application Presentation

Sven Maurice Morlock

2023-07-13

# Motivation I

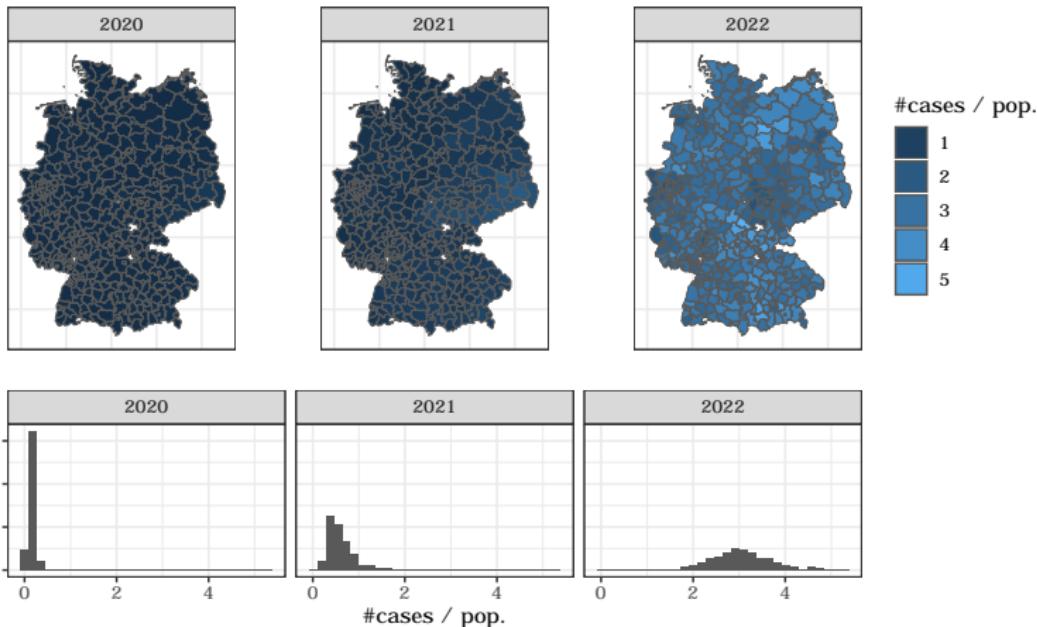
Example data Germany



- ▶ Corona cases in Germany
- ▶ We can see some regional patterns
- ▶ Can some districts be clustered?

# Motivation II

Aggregated corona cases by year (adjusted by population)



## Setup I

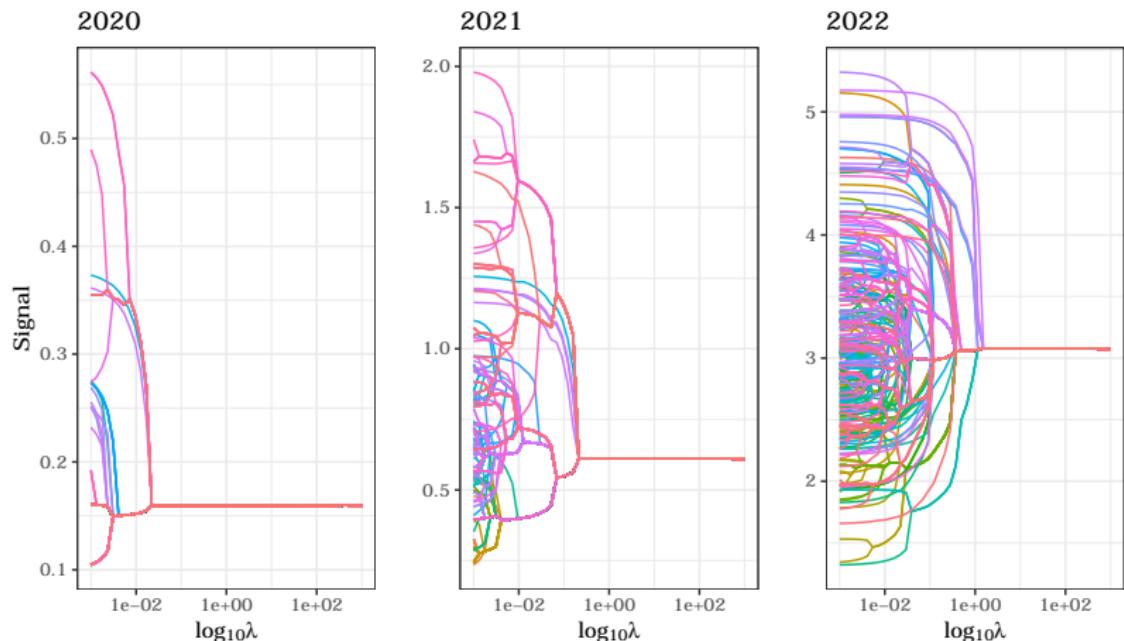
- ▶ data on corona cases in Germany for years 2020 to 2022
- ▶ 396 districts (2 non-adjacent)
- ▶ interpret each district as vertex/node
- ▶ (initial) edge weight 1/0 if districts adjacent or not
- ▶ Spatial segmentation/clustering based on Goepp and Kassteele (2022) via adaptive ridge
  - ▶ only one 'spatial signal' no other variables are taken into account
  - ▶ spatial signal  $\hat{=}$  
$$\frac{\text{cummulative cases in period}}{\text{avg. population in period}}$$

## Setup II (adaptive ridge)

- ▶ in each iteration  $I$ :
  - ▶ calculate new weights  $v_k$
  - ▶ calculate new signal  $\theta_k$
  - ▶ minimize  $I^{pen}(\boldsymbol{\theta}, \mathcal{V}^I) = I(\boldsymbol{\theta}) + \frac{\lambda}{2} + \sum_{j \sim k} v_{j,k}^{(I)} (\theta_j - \theta_i)^2$
- ▶ after successful iteration we will end up with new signals for each district
- ▶ but due to regularization some districts will have the same signal (we are penalizing difference in signals)
- ▶ districts having same signal assigned to same cluster
- ▶ this is done for some schedule of  $\lambda$  values
  - ▶  $\rightsquigarrow$  assessment after each iteration cycle by AIC/BIC/GCV

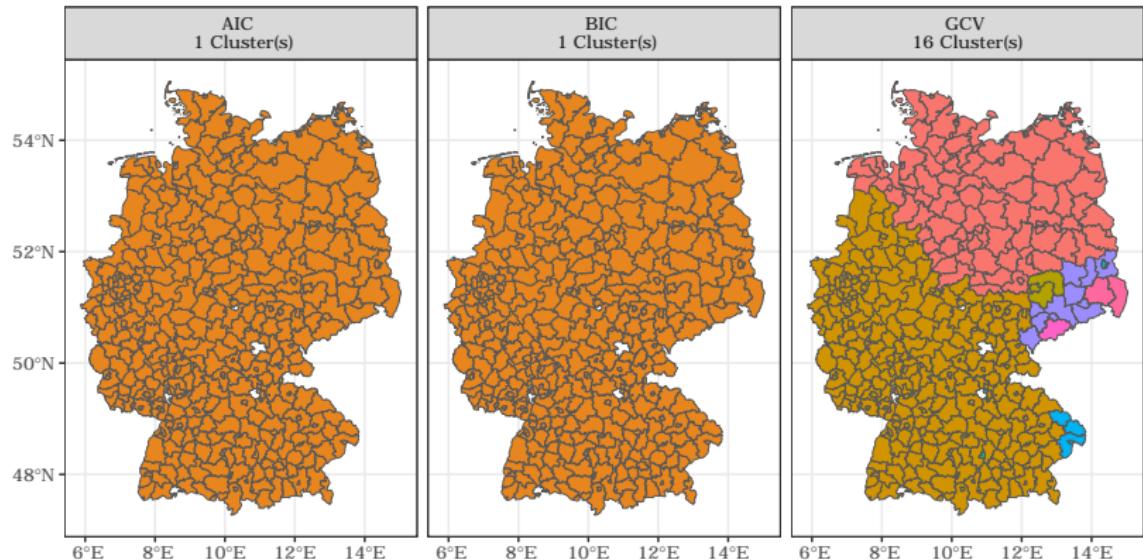
# Results I

Regularization paths for each year



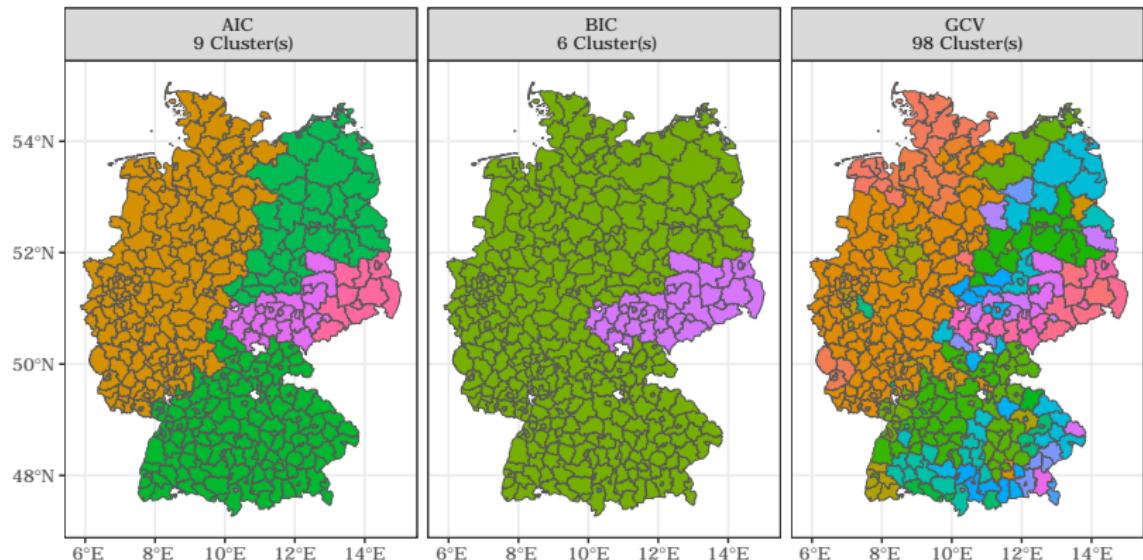
# Results II (2020)

Piecewise constant signal 2020



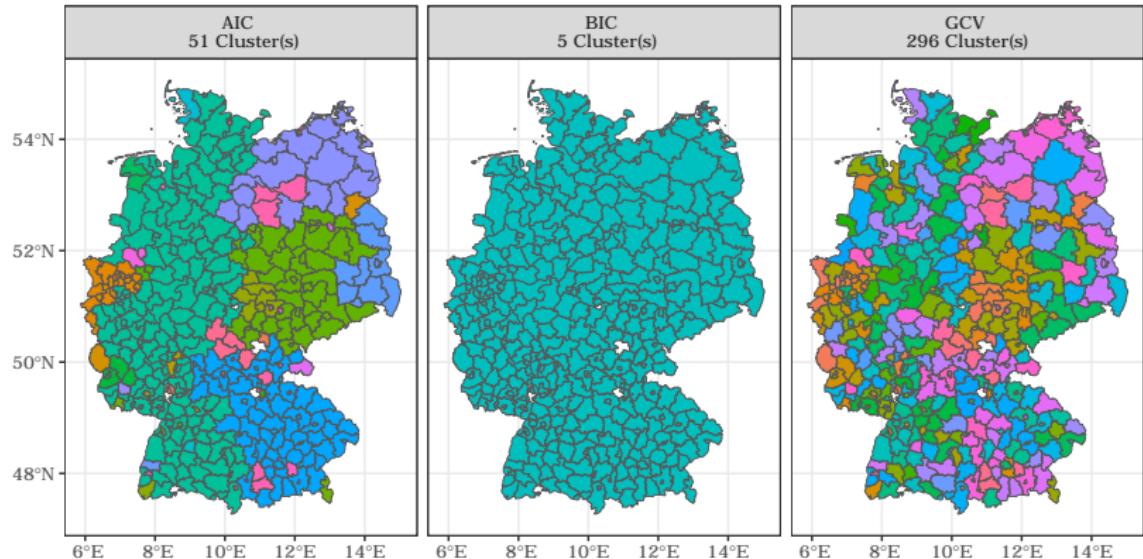
# Results III (2021)

Piecewise constant signal 2021



# Results IV (2022)

Piecewise constant signal 2022



## Discussion

- ▶ We can see different results depending on chosen IC and time period
- ▶ By design only 1 variable (graph signal) is considered
  - ▶ no covariates included
  - ▶ but also no notion of dependent/independent variable as in regression
- ▶ limited interpretation

# Backup

## Backup

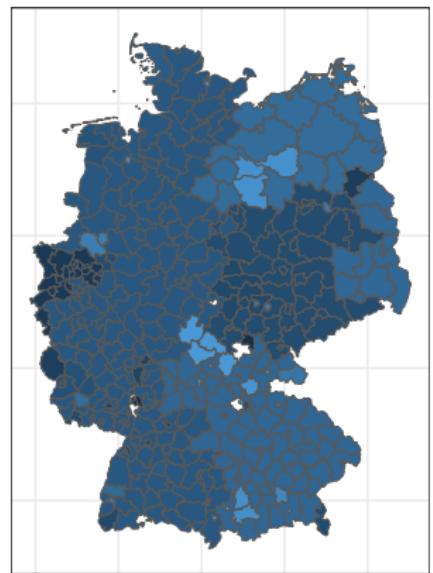
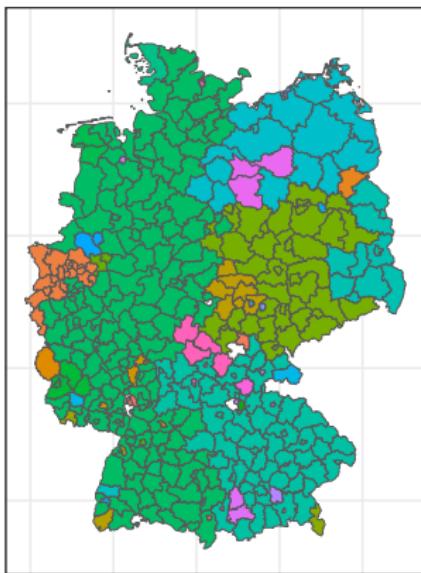
- ▶ Adaptive ridge iteratively optimizes:

$$\boldsymbol{\theta}^{(l)} = (\boldsymbol{\Sigma}^{-1} + \lambda \boldsymbol{K}^{(l-1)})^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{x} \quad (1)$$

$$v_{j,k}^{(l)} = \frac{1}{(\theta_j^{(l)} - \theta_k^{(l)})^2 + c} \quad (2)$$

# Backup

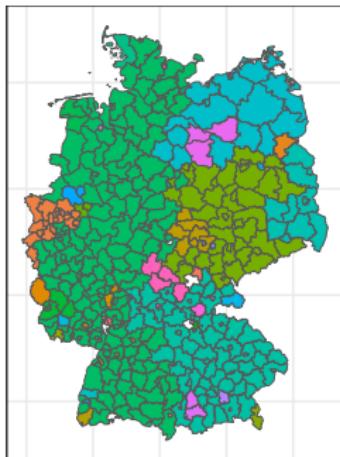
Result for year 2022 (AIC)  
51Cluster(s)



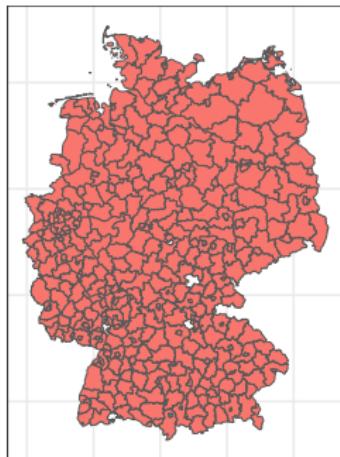
Signal  
2  
3  
4  
5

# Backup

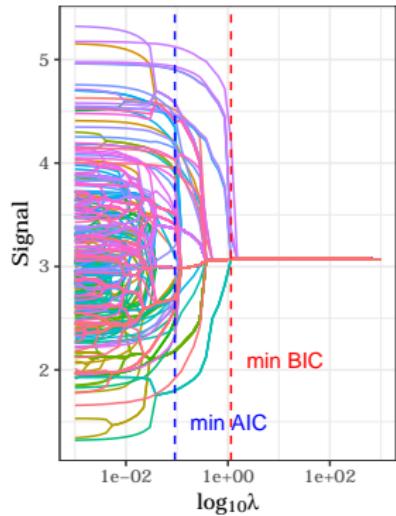
Segmentation 2022 (AIC)  
51 Cluster(s)



Segmentation 2022 (BIC)  
5 Cluster(s)

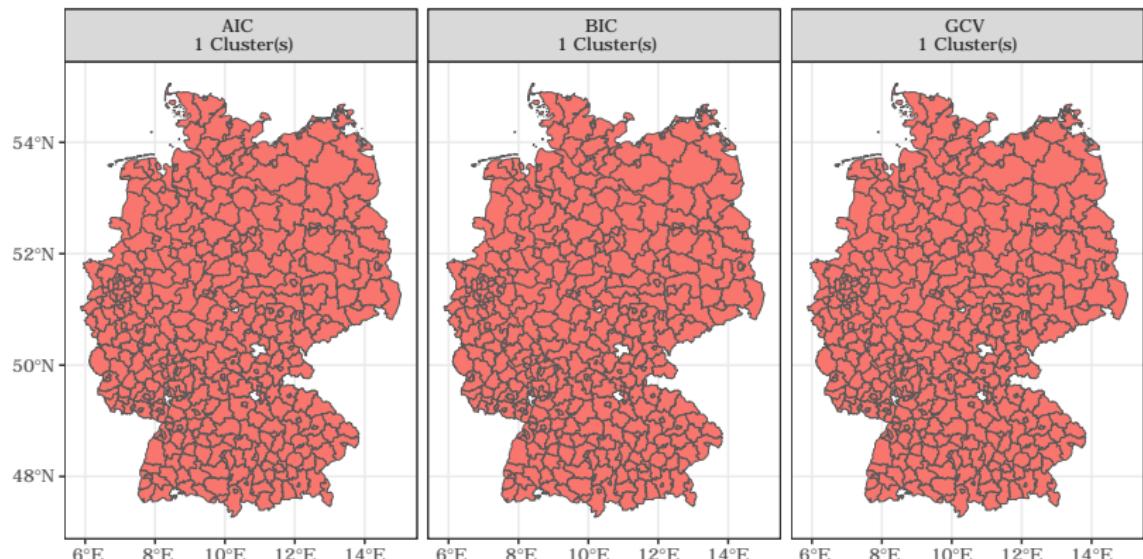


Regularization path 2022



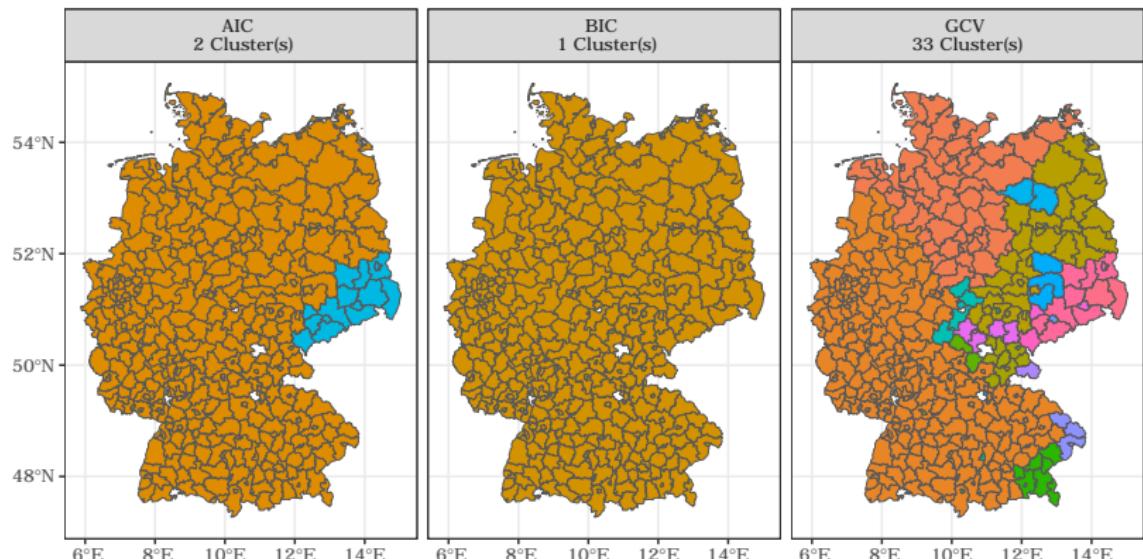
# Backup

Piecewise constant signal  
First wave (10/2020 – 20/2020)



# Backup

Piecewise constant signal  
Second wave (40/2020 – 8/2021)



## References

Goepp, Vivien, and Jan Kassteele. 2022. "Graph-Based Spatial Segmentation of Health-Related Areal Data," May.