

# Räumliche Aspekte der Fallzahlenentwicklung in Deutschland

Höser, Christoph

*Anschrift Korrespondierende/r Autor\*in:  
Universitätsklinikum Bonn, Gebäude 64 (Thofern Haus)  
Venusberg-Campus , D-53127 Bonn*

*Email Korrespondierende/r Autor\*in: choeser@ukbonn.de*

## Motivation

Die Entwicklung der Fallzahlen zu COVID19 in Deutschland bedarf nicht nur einer epidemiologischen Betrachtung insgesamt, sondern verdient eine räumlich differenzierte Betrachtung. Dabei geht es nicht nur darum, die möglicherweise kausalen Ursachen für das lokale Ausbreitungsgeschehen besser zu verstehen, sondern auch nicht-kausale Zusammenhänge aufzuzeigen, etwa um das Ausmaß der Betroffenheit aus verschiedenen Blickwinkeln zu beleuchten. Weiterhin sollte der Charakter z.B. eines Landkreises aus statistischer Sicht in allen Facetten bekannt sein, bevor versucht wird die lokalen Ergebnisse einer Schwerpunktstudie, wie z.B. in Heinsberg oder Kupferzell, zu verallgemeinern. Auch zur Untersuchung der Wirksamkeit lokaler Maßnahmen und zum adäquaten Einsatz von Beschränkungen ist das Hintergrundwissen zum Ort sicher notwendig.

## Konzept

Das Konzept zur Lösung dieser Aufgabenstellung sieht vor, dass die Meldedaten zum Infektionsgeschehen aggregiert werden, jedoch nicht wie üblich auf Ebene der Bundesländer, sondern nach charakteristischen Kreistypen unterschiedlichster Themen. Damit kann nun untersucht werden, inwieweit ein ausgewähltes Thema ein differenziertes Outcome enthält. Dabei muss das Thema nicht notwendigerweise als kausal betrachtet werden.

Die Umsetzung erfolgt mit sql und r, das interface ist mittels shiny(1) realisiert. Das Werkzeug ist daher auf verschiedenen Betriebssystemen verwendbar.

## Daten

Das Grundgerüst der Auswertelogik ist entsprechend dem Konzept flexibel gehalten und kann in sehr unterschiedlicher räumlicher Granularität eingesetzt werden – soweit einerseits Daten zum Infektionsgeschehen und andererseits Informationen zu statistischen Hintergrunddaten in der gleichen Auflösung zur Verfügung stehen. Das Werkzeug wurde exemplarisch für die Ebene der Kreise in Deutschland realisiert.

## Datengrundlage

Für das Infektionsgeschehen werden die tagesscharfen Meldedaten des RKI auf Kreisebene verwendet.(2) Der komplette Datensatz wird täglich in eine eigene SQL Datenbank überführt und enthält daher alle Nachtragungen und umfasst Fallzahlen, Verstorbene und Genesene.

Der Abruf aus der eigenen Datenbank erfolgt direkt aus r mittels stored procedures. Die Qualität der Meldedaten wird an dieser Stelle nicht diskutiert. (Bei Verwendung der jüngst zurückliegenden Tage erfolgt im Interface eine Warnung, da diese Daten i.d.R. besonders instabil sind.)

Für die räumliche Statistik werden einerseits bereits ausgearbeitete Klassifikationen der Kreise aus dem BBSR verwendet (Raumbeobachtung)(3) und andererseits statistische Maßzahlen auf Kreisebene aus der INKAR(4) Datenbank. Da hier keinerlei API zur Verfügung stand, wurden mehr als 400 Themen ausgewählt und inklusive der Metadaten zur Methodik etc in eine eigene Datenbank überführt. Dies ist problemlos erweiterbar.

Für die exemplarische Umsetzung ist die Datenqualität der statistischen Hintergrunddaten aus BBSR und INKAR unstrittig.

### Optionen der Selektion

Die Auswertung beschränkt sich auf einen interaktiv festzulegenden Zeitraum, so dass beliebige Abschnitte aus dem Verlauf des Infektionsgeschehens betrachtet werden können. Als Zwischenergebnis wird eine Karte dargestellt.

Eine weitere Option erlaubt die Auswahl von Bundesländern, deren Kreise berücksichtigt werden sollen. So können bei Bedarf auch Teilbereiche Deutschlands untersucht werden.

### Datenaufbereitung

Die Daten des Infektionsgeschehens werden aggregiert für Kreise, wobei die prinzipiell verfügbaren Differenzierungen zu Geschlecht und Altersklasse zunächst nicht weiter berücksichtigt werden. Die Stadtbezirke Berlins werden zusammengefasst, da seitens des BBSR und INKAR in den Daten auf Kreisebene keine Differenzierung vorliegt.

Die Themen des BBSR und aus INKAR wurden nach Sachgebieten hierarchisch in 4 Stufen gegliedert, um eine komfortable Auswahl zu ermöglichen. Für die Raumbeobachtung sind die Klassifikationen bereits vorgegeben. Für die statistischen Maßzahlen aus INKAR ist dies nicht der Fall, hier muss eine Klassifikation noch erfolgen. Die Klassifikation erfolgt bei diesen Themen interaktiv nach dem Verfahren natural breaks (Fisher).(5) Die Klassenbreite und Klassenbesetzung ist daher unterschiedlich, aber die Klassen bilden die gegebene Verteilung der eingehenden Einzelwerte optimal ab, der Godness of Fit (GOF) wird ausgewiesen. Bei der Berechnung der Klassen wird versucht, den interaktiv festgelegten GOF mit möglichst wenigen Klassen zu erreichen. Das Verfahren wird beendet, wenn entweder der GOF oder die maximal gewünschte Klassenzahl erreicht ist. Die erreichte Klassifikation wird als Zwischenergebnis in einer Karte dargestellt.

## Auswertung

### Verfahren

Die Zahlen zum Infektionsgeschehen werden nun, entsprechend der zuvor gesetzten räumlichen und temporalen Filter, nach dem Thema und der erzeugten Klassifikation je Klasse aggregiert. Hierzu werden die korrekten Inzidenzen per 100k Einwohner für die Summe aller beteiligten Kreise jeder Klasse neu gebildet (kein gewichteter Durchschnitt, dies wäre falsch!).

Eine Darstellung in Form eines Scatterplots verbietet sich aus dem Grund, da die reporting areas extrem unterschiedliche Einwohnerzahlen als Basis verwenden und daher nicht vergleichbare Sensitivität für die Erfassung eines Risikos besitzen. Daher ist das Konzept der Klassenbildung unerlässlich.

## Ergebnis und Export

Im Ergebnis erfolgt eine graphische Visualisierung der Klassen in aufsteigender Reihenfolge. In der Graphik werden die einzelnen Kreise im Hintergrund als farbige dots dargestellt, um einen Eindruck der Streuung zu erhalten. Wesentlich ist jedoch die rote Markierung je Klasse. Die gepunktete Linie stellt den Wert für die Summe der eingegangenen Werte dar. Die Visualisierung ist mittels der Library plotly(6) erstellt und erlaubt vielfältige interaktive Manipulationen und Exportoptionen.

Zusätzlich zur graphischen Darstellung ist eine vollständige Dokumentation der Datenquellen und der interaktiv gewählten Filter abrufbar, der im Detail die Grundlagen erläutert, die zu diesem Ergebnis geführt haben. Dies umfasst auch alle Angaben zur Klassifikation sowie die Datentabelle selbst.

## Screenshot:



## Ausblick

Eine weitere Graphik wird den zeitlichen Verlauf je Klasse darstellen, denn gleichartige Gesamtergebnisse zweier Klassen können sich innerhalb des gewählten Zeitraums unterschiedlich entwickelt haben. Hierbei wird ein 7tägiges gleitendes Mittel verwendet werden, um die wochentägliche Schwankung der Meldedaten auszugleichen. Eine Auswahloption wird selektierte Kreise in der Graphik hervorheben.

## Fazit

Die Kombinationsmöglichkeiten aus mehr als 400 Themen, den Optionen zur Klassenbildung, den räumlichen Filtern und den temporalen Filtern sind nahezu unendlich. Gleichzeitig ist ein sehr verantwortungsvoller Umgang mit den Themen unerlässlich, um die Ergebnisse richtig

zu interpretieren. Daher ist das entstandene Werkzeug als Expertensystem zu betrachten, der Zugang steht über den Autor zur Verfügung.

## Referenzen

Für die Unterstützung bei der Aufbereitung der Metadaten zu den INKAR-Daten danke ich Frau Nadine **Blätgen**, **BBSR**.

## Literaturverzeichnis

1. Shiny [Internet]. [zitiert 3. September 2020]. Verfügbar unter: <https://shiny.rstudio.com/>
2. RKI COVID19 [Internet]. [zitiert 3. September 2020]. Verfügbar unter: [https://npgeo-corona-mpgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6\\_0](https://npgeo-corona-mpgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0)
3. BBSR - Raumbewachtung [Internet]. [zitiert 3. September 2020]. Verfügbar unter: [https://www.bbsr.bund.de/BBSR/DE/forschung/raumbewachtung/\\_node.html](https://www.bbsr.bund.de/BBSR/DE/forschung/raumbewachtung/_node.html)
4. Start - INKAR online [Internet]. [zitiert 3. September 2020]. Verfügbar unter: <https://www.inkar.de/>
5. Fisher WD. On Grouping for Maximum Homogeneity. J Am Stat Assoc. Dezember 1958;53(284):789–98.
6. Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny. CRC Press; 2020. 278 S.