

## Explorative Clusteranalyse von geo-codierten Daten

Für Forscher kann es eine Herausforderung sein, Muster von ähnlichen Objekten innerhalb eines Datensatzes zu entdecken und zu analysieren. Zudem ist ein fundiertes Wissen erforderlich, um aussagekräftige Cluster zu bilden und Schlussfolgerungen ziehen zu können. Daten zu analysieren, die nicht korrekt transformiert oder standardisiert sind, kann zu schlechteren Ergebnissen führen. Die Auswahl der Methoden und Abstandsmasse zur Generierung von Clustern sind von den ausgewählten Variablen abhängig. Auch die Anwendung von Clustering-Algorithmen und die Visualisierung der erzeugten Ergebnisse können eine Herausforderung darstellen, da genügende Programmierkenntnisse erforderlich sind. Das Fehlen dieser Fähigkeiten führt dazu, dass Forscher schlechte Datenanalyse betreiben. Die hohe Komplexität für die Auswertung von geclusterten Daten ist ein Problem, mit dem viele Forscher konfrontiert sind. Vor allem unter Einbeziehung von Kompositionsdaten, die ungewohnte Techniken erfordern und Einschränkungen unterliegen, welche eingehalten werden müssen. Um dieses Problem zu lösen, wurde eine R-basierte Anwendung entwickelt. Das Programm stellt alle notwendigen Schritte bereit, die Forscher bei der korrekten Datenauswertung unterstützen. Eingelesene Datensätze werden auf ihre Struktur und ihr Format überprüft, um die Funktionalität der Applikation sicherzustellen. Die Anwendung wird in Übereinstimmung mit einigen uns bekannten Datensätzen entwickelt, um Einblicke in mögliche Fehlerquellen zu geben, die bei der Ausführung der Funktionen auftreten könnten. Eine breite Palette von Open-Source-Funktionen sind in die Anwendung eingebettet. Diese ermöglichen es, Clustering-Ergebnisse vollständig zu generieren, einschliesslich der Visualisierung der projizierten Daten auf Karten und Diagrammen. Für die Projektion von Datenpunkten auf dynamische und stabile Karten sind Datensätze erforderlich, die geografische Referenzen zur Bestimmung eines Beobachtungsortes liefern. In Übereinstimmung mit den generierten visuellen Ausgaben werden Ergebnisse ausgegeben, welche dem Benutzer wichtige Informationen zur Qualität der generierten Cluster liefern. Es wurde festgestellt, dass die Verwendung der ClustToolApp die Forscher bei der Analyse der Daten unterstützen kann, da sie über eine breite Palette eingebetteter Funktionen verfügt und vielfältige Ausgaben erzeugt. Sie hat sich bei der Erfassung, Analyse und Darstellung von Daten für die explorative Forschung als erfolgreich erwiesen.



### Diplomierende

Kerim Alqadi  
Kevin Steffen

### Dozent

Matthias Templ



Projektion von Geodaten auf eine von ClustToolApp bereitgestellte dynamische Karte mit spezifizierter Beschriftung für jede einzelne Beobachtung.



Der Silhouettenplot dient zur Beurteilung der Qualität jedes erzeugten Clusters. Es wird grafisch dargestellt, wie gut jede einzelne Observation klassifiziert wurde.