

A Shiny app for bivariate Choropleths

Jenny Kloster

Matrikelnummer: 01/942243

Einleitung

Die Grundfunktion unserer Shiny App war das Erstellen bivariater Choroplethenkarten. Damit sollen unsere App Benutzer die Möglichkeit haben, in denen von ihnen interessierten Bereichen Datenexploration zu betreiben. Auf diese Weise sollen sie einen visuell attraktiven Eindruck von der aktuellen Datenlage bekommen. Die Datenvisualisierung und Exploration wird auch im Bereich der psychologischen Forschung immer relevanter. Mit dieser App können potenzielle Zusammenhänge zwischen zwei Variablen über verschiedene geographische Gebiete hinweg erforscht oder die Befunde solcher Zusammenhänge visualisiert werden. Die App kann je nach Daten auf jedes beliebige Thema angewandt werden. So kann sich zum Beispiel ein Gesundheitspsychologe über gesundheitsbezogene Daten informieren (z.B. die selbst wahrgenommene Gesundheit).

Da allgemein der Umgang mit R und speziell das Erstellen bivariater Choroplethenkarten für R-Neulinge kompliziert und zeitintensiv sein kann, soll diese App die Datenexploration und die Erstellung bivariater Choroplethenkarten vereinfachen. Diese App wurde mithilfe von Shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2018) und der R Version 3.5.2 (2018-12-20) kreiert.

Natürlich muss hinzugefügt werden, dass es sich bei der App ausschließlich um ein Datenexplorationsinstrument handelt, was keinerlei statistisches Hypothesentesten inkludiert.

Bivariate Choroplethenkarten

Die bivariaten Choroplethenkarten in dieser App wurden nach der Anleitung von Art Steinmetz (2017) erstellt. Die Variablen, welche in der Abbildung miteinander verglichen werden, werden auf die Bevölkerungsgröße normalisiert. Ansonsten würden sich Unterschiede in der Populationsgröße der geographischen Gebiete in den Choroplethenkarten widerspiegeln (Joshua Stevens, 2015). Beide Variablen werden in drei Gruppen kategorisiert. Daraufhin wird eine einzelne Variable erstellt, welche die klassifizierten Variablen vereint. Diese Variable beinhaltet neun (3×3) Gruppen, wobei jeder eine bestimmte Farbe zugeordnet wird. Die geographischen Formen, welche zu den

geographischen Gebieten aus dem Datenset gehören, werden dann nach dieser Variablen gefärbt. Alle bivariaten Choroplethenkarten werden dann mithilfe des Paketes tmap (Tennekes, 2018) erstellt und dann auf der Weltkarte mithilfe von leaflet (Cheng, Karambelkar, & Xie, 2018) eingezeichnet.

Das eigene Datenset

Der Benutzer hat nun zwei Möglichkeiten, Daten bei der App hochzuladen. Die erste Möglichkeit ist das Hochladen eines eigenen Datensatzes. Aktuell können nur Excel Datensätze (.xls oder .xlsx) hochgeladen werden. Zudem muss der Datensatz im Voraus gereinigt werden und im wide Format vorliegen.

Nachdem der Datensatz hochgeladen wurde, wird die Tabelle mit den Daten dargestellt, wobei die Anzahl an dargestellten Zeilen manuell justiert werden kann. Dies ermöglicht dem Nutzer verschiedene Variablen in der Seitenbox auszuwählen. Hierbei kann der Nutzer zwei Variablen von Interesse und eine Variable mit der geographischen Information (entweder die deutschen Bundesländer oder Europa) auswählen. Zusätzlich kann bei Bedarf eine Variable mit Zeitinformationen (z.B. Jahre) ausgesucht werden, um diese zu spezifizieren. In diesem Fall kann jede Abstufung der Variablen ausgesucht werden. Des Weiteren muss der Benutzer die Karte auswählen, auf der die Daten später dargestellt werden. Bisher sind Karten von Europa oder der deutschen Bundesländer wählbar, der Benutzer kann selbst keine Karten hochladen. Die Abstufungen der geographischen Variablen des Datensatzes müssen mit der Karte kongruieren.

Wenn eine Zeitvariable und irgendeine Abstufung der Variablen ausgesucht wurden, wird der Datensatz nach der Abstufung dieser Variablen gefiltert (z.B. nur Daten des Jahres 2016 werden fortan verwendet). Die Variablen, die nun bei der bivariaten Choroplethenkarte verglichen werden, werden daraufhin anhand der Variablen, die Informationen bezüglich der Populationsgröße von jedem geographischen Gebiet enthält, normalisiert.

Wird schließlich der Button „Create plot“ gedrückt, wird die bivariate Choroplethenkarte und eine Legende dargestellt. Nun kann eine Datei mit dieser Karte heruntergeladen werden. Dies funktioniert jedoch nur, wenn die App im Brower geöffnet wurde.

Eurostat Daten

Die zweite Möglichkeit, Daten bei der App hochzuladen, besteht darin, einen Eurostat Datensatz auszuwählen. Die App erlaubt es nach Eurostat Datensätzen zu suchen, indem man einen Suchbegriff in eine Suchleiste eingibt. Sobald der Nutzer einen für ihn interessanten Datensatz gefunden hat, kann er oder sie den zugehörigen Code in ein zweites Eingabefeld entweder händisch oder bequem per copy & paste einfügen. Der Datensatz wird dann automatisch heruntergeladen. Zugriff zu den Eurostat Datensätzen bekommt die App durch das Paket „eurostat“ (Lahti, Houvari, Kainu, & Biecek, 2017).

Nachdem das Datenset heruntergeladen wurde, wird die Tabelle mit den Daten wieder dargestellt. Hier kann wieder die Anzahl an dargestellten Zeilen manuell eingestellt werden. Zuerst muss der Nutzer die Variablen von Interesse aussuchen, dabei können es entweder zwei verschiedene Variablen oder zweimal dieselbe Variable sein (um zwei Abstufungen derselben Variablen miteinander vergleichen zu können).

Als nächstes wird eine Tabelle mit Spaltenfiltern angezeigt. Der Nutzer muss hier alle Variablen filtern, welche nicht blockiert sind. Hierbei kann der Nutzer nur eine Ausprägung pro Spalte auswählen und ein Jahr zwischen 1990 und 2016. Falls diese Bedingungen nicht zutreffen, wird der Filter nicht angewendet und es passiert einfach nichts. Ebenso sind die Spalten mit den vorher ausgewählten Variablen blockiert, sowie die Spalte mit den geographischen Informationen, die Spalte mit den Werten und die Spalte der Variablen „unit“, da diese nur eine Ausprägung hat.

Je nach Auswahl der Filter, wird die gewünschte Information aus dem Eurostat Datensatz extrahiert und vom long ins wide Format konvertiert. Daraufhin werden die Abstufungen der vorher ausgewählten Variablen angezeigt und müssen ausgesucht werden.

Nach dem Drücken des „Create plot“ Buttons, wird eine weitere Variable mit Informationen über die Bevölkerungsdichte zum Datensatz hinzugefügt, damit die ausgewählten Abstufungen normalisiert werden können. Im Moment wird die Information bezüglich der Bevölkerungsdichte vom Eurostat Datenset „demo_r_d3dens“ erhalten. Eine Limitation dieser Datenquelle ist, dass diese Informationen lediglich für die Jahre 1990 bis 2016 vorliegen. Ebenfalls nennenswert ist, dass alle Zeilen der Datensätze, bei welchen die geographischen Informationen nicht mit einem Gebiet in dem „demo_r_d3dens“ Datenset übereinstimmen, entfernt werden.

Schließlich wird eine bivariate Choroplethenkarte und eine Legende dargestellt. Hier kann erneut eine Datei mit der Karte heruntergeladen werden. Zusätzlich kann das endgültige Datenset ebenfalls heruntergeladen und später bei Bedarf wieder bei der App hochgeladen werden, um wieder dieselbe Abbildung zu erzeugen.

Aufgabenteilung

Diese App wurde zusammen mit Julia E. Koller erstellt. Das Konzept der App haben wir zusammen erarbeitet. Während sich Julia hauptsächlich mit der praktischen Implementierung unseres Konzeptes in der Shiny App beschäftigt hat und hierbei besonders mit der Möglichkeit eigene Datensätze hochzuladen, um dann Choroplethenkarten zu erstellen, habe ich das Eurostat Paket entdeckt und mich mit dem Datenmanagement beschäftigt, damit man die Eurostat Datensätze benutzen kann.

Referenzen

Art Steinmetz. (2017). Exploring the distribution of NYC Pre-K seats vs. neighborhood income.

Retrieved 19.02.2019, from <https://rpubs.com/apsteinmetz/prek>

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2018). shiny: Web Application Framework for R. Retrieved from <https://CRAN.R-project.org/package=shiny>

Cheng, J., Karambelkar, B., & Xie, Y. (2018). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. Retrieved from <https://CRAN.R-project.org/package=leaflet>

Joshua Stevens. (2015). Bivariate Choropleth Maps: A How-to Guide. Retrieved 19.02.2019, from <http://www.joshuastevens.net/cartography/make-a-bivariate-choropleth-map/>

Lahti, L., Huovari, J., Kainu, M., & Biecek, P. (2017). eurostat R package. , 9(1), 385–392. Retrieved from <https://journal.r-project.org/archive/2017/RJ-2017-019/index.html>

Tennekes, M. (2018). tmap: Thematic Maps in R. Journal of Statistical Software, 84(6), 1–39.