

Elementare Methoden

Abgabe über die NextCloud bis 23:59 Uhr des o.g. Datums.

Aufgabe 1 Attribute Mapping

(8 Punkte)

In `spread_data.csv` befindet sich eine Tabelle mit Einträgen zur Ausbreitung einer infektiösen Krankheit (Spalten), gegliedert nach Bundesland (Zeilen). In dieser Aufgabe sollen verschiedene Eigenschaften dieser Daten auf einer Karte dargestellt werden. Die Datei wurde bereits eingelesen.

a) (1 Punkt)

Verschaffen Sie sich einen Überblick über die Daten mittels `print(data_frame)`. Die Einträge `x_pos` und `y_pos` verweisen auf eine Pixelposition in der Datei `germany.png`, welche eine Deutschlandkarte zeigt. Lesen Sie dieses Bild ein. Es sollen zwei Darstellungen verglichen werden. Erstellen Sie dafür nebeneinander *zwei* Plots der Karte mittels `imshow`.

b) (2 Punkte)

Erstellen Sie pro Bundesland einen Kreis, dessen Fläche die Anzahl der Infizierten der letzten Woche zeigt. Mappen Sie die Daten auf der ersten Karte direkt auf die Kreisfläche. Skalieren Sie die Flächen linear mit dem Faktor 0.05, damit die Punkte nicht zu groß werden.

Hinweise:

- Das Auslesen einer Spalte aus einem Pandas *Dataframe* kann mit `data_frame["columnName"]` erfolgen.
- Der Scatterplot-Funktion kann im Parameter `s` die Größe der Punkte als *Fläche* übergeben werden (siehe Cheatsheet *Scatterplots*).

c) (3 Punkte)

Verwenden Sie nun *Steven's area judgement scale*, um die Kreisflächen zu bestimmen. Es soll wieder die 7-Tage-Anzahl behandelt werden. Stellen Sie das Ergebnis zum Vergleich auf der zweiten Karte dar. Verwenden Sie zur Skalierung $\alpha = 300$.

Hinweise:

- Ein guter Startpunkt ist die Beziehung $I = \alpha r^{1.4}$. In dieser Formel ist I der Attributwert (also hier die Anzahl der Infizierten).

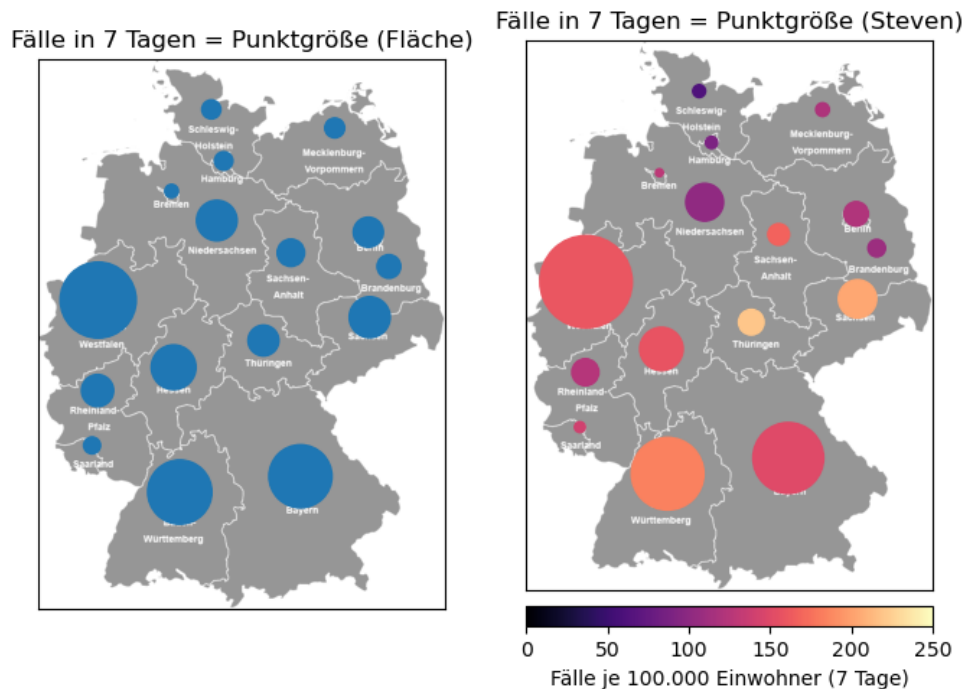
d) (2 Punkte)

Zeigen Sie nun die Werte der relativen Infektionen in 7 Tagen pro 100.000 Einwohner durch eine Colormap. Die Einwohner pro Bundesland stehen ebenfalls in der Tabelle. Die Colormap soll zwischen 0 und 250 skaliert sein. Zeigen Sie diese Skalierung mit einer `colorbar`.

Hinweise:

- Verwenden Sie `cmap='magma'` beim Erstellen des Scatterplots und übergeben Sie die richtigen Werte an den Parameter `c` (color).
- Im Beispiel von letzter Woche sehen Sie den Umgang mit der `colorbar`. Eine gute Darstellung erhält man mit den zusätzlichen Parametern `orientation='horizontal'` und `pad=0.02`.

Das Ergebnis sieht so aus:



Aufgabe 2 Theorie

(7 Punkte)

Geben Sie die Antworten auf die Theorieaufgaben in der Multiple-Choice-Datei MC03.txt an. Es ist immer genau eine Auswahlmöglichkeit richtig. Bitte keine anderen Anmerkungen in diese Datei schreiben und den Dateinamen nicht verändern.

a) (1 Punkt)

Für die Verwendung von Flächen als visuellen Repräsentationskanal von Daten gilt:

- (a) Verhältnismäßig große Flächen werden i.d.R. genauso groß wahrgenommen wie sie sind.
- (b) Eine große Fläche wird eher kleiner wahrgenommen als sie ist.
- (c) Eine große Fläche wird meist größer wahrgenommen als sie ist.

b) (1 Punkt)

Was ist das entscheidende Merkmal, warum Balkendiagramme (bar charts) oft besser geeignet sind als Kreisdigramme (pie charts), um Daten zu vergleichen?

- (a) Sie nutzen Position und Länge als Codierung.
- (b) Sie benötigen keine Farbe.
- (c) Sie erlauben mehr Kategorien.

c) (1 Punkt)

Welcher Diagrammtyp sollte am ehesten zum Vergleich vieler Zeitserien gewählt werden, wenn das Ziel ist, schnell große und kleine Wertebereiche zu finden?

- (a) *Line graph.*
- (b) *Horizon graph.*
- (c) *Stacked graph.*
- (d) *Small multiples.*

d) (1 Punkt)

Welcher Diagrammtyp sollte am ehesten für eine kleine Anzahl von Zeitserien gewählt werden, wenn das Ziel ist, die Serien direkt zu vergleichen?

- (a) *Line graph.*
- (b) *Horizon graph.*
- (c) *Stacked graph.*
- (d) *Small multiples.*

e) (1 Punkt)

Berechnen Sie für die folgenden Daten das untere Quartil q_1 , den Median q_2 und das obere Quartil q_3 : 1, 5, 8, 7, 5, 2, 6, 8, 4, 7, 2, 1, 4. Antwortmöglichkeiten:

- (a) $q_1 = 1, \quad q_2 = 5, \quad q_3 = 7$
- (b) $q_1 = 2, \quad q_2 = 4, \quad q_3 = 7$
- (c) $q_1 = 2, \quad q_2 = 5, \quad q_3 = 6$
- (d) $q_1 = 3, \quad q_2 = 5, \quad q_3 = 7.5$
- (e) $q_1 = 2, \quad q_2 = 5, \quad q_3 = 7$

f) (1 Punkt)

Was zeigen *parallele Koordinaten*?

- (a) Die horizontalen Verbindungen zeigen die Korrelation zwischen Datenproben.
- (b) Jede horizontale Verbindung zeigt die Veränderung einer Variable über die Zeit.
- (c) Jede horizontale Verbindung zeigt die Ausprägung einer Datenprobe bzgl. der dargestellten Attribute.

g) (1 Punkt)

Welche Aussage gilt *nicht* bei der Verwendung von Glyphen zur Attributcodierung?

- (a) Komplexe Glyphen sollten in der Visualisierungsdomäne eher dünn verteilt sein.
- (b) Glyphen sollten da auffälliger sein, wo wichtige Attribute zu finden sind.
- (c) Ellipsoide werden in 3D Darstellungen genutzt, da Sie blickrichtungsunabhängig sind.
- (d) Die Anzahl von Attributen, die durch eine Glyphe kommuniziert werden können, ist begrenzt durch menschliche Wahrnehmungslimits.