

Multivariate Graphiken & ggplot2

Vorbereitungen:

Laden Sie den Datensatz `wdi` sowie die zugehörige Lookup-Tabelle `wdi_info` von der Homepage und lesen Sie diese in R ein.

Aufgabe 1: ggplot2

`ggplot2` von Hadley Wickham ist ein Paket, welches die Erstellung von Graphiken durch eine logische Syntax im Vergleich zu Standard-Graphiken in R teilweise deutlich vereinfacht. Verschaffen Sie sich einen Überblick über die Graphikerstellung mit `ggplot2`, z.B. anhand des offiziellen Buches oder des Cheat Sheets.

- a) Was ist die *Grammatik* von `ggplot2`-Graphiken? Was sind Ebenen (layers) und woraus setzen sich diese zusammen?
- b) Machen Sie sich mit den verschiedenen Layer-Attributen genauer vertraut:
 - i) *aesthetic mapping*: Plotten Sie die Arbeitslosigkeit aller westeuropäischen Länder in einem Scatterplot (`geom_point()`) gegen das Jahr. Geben Sie den Ländern unterschiedliche Farben.
 - ii) *geom*: Verwenden Sie statt `geom_point()` nun `geom_line(aes(group = Land))` für ein Liniendiagramm. Kombinieren Sie Punkte und Linien in einer Graphik. Erstellen Sie außerdem einen Boxplot der Arbeitslosigkeit über alle Jahre sowie unterteilt nach Jahren.
 - iii) *stat*: Welche Funktion zur Datentransformation wird bei der Erstellung eines Scatterplots angewandt? Welche bei einem Balkendiagramm? Welche bei einem Histogramm?
 - iv) *position*: Plotten Sie die Anzahl der Länder pro Kontinent für 2013 in einem Balkendiagramm. Unterteilen Sie die Balken dabei farblich nach der kategorisierten Arbeitslosigkeit. Welche Darstellungsmöglichkeiten bietet Ihnen hierbei das `position`-Argument in `geom_bar()`? Stellen Sie eines der Balkendiagramme horizontal dar, um sich mit dem einfachen Umgang mit Koordinatentransformationen in `ggplot2` vertraut zu machen.

- c) Weitere Anpassungsmöglichkeiten:

- i) **theme:** Machen Sie sich mit Themes in `ggplot2` vertraut (`?theme`, `?theme_minimal`, `?theme_update`). Wie können Sie das Standard-Theme dauerhaft umstellen?
- ii) **Plotbeschriftung:** Wie verändern Sie den Titel, die Achsenbeschriftungen sowie die Labels und Ticks der x- und y-Achse? Wie unterdrücken Sie diese? Setzen Sie den angezeigten Bereich der x-Achse auf `[0,100]`. Unterdrücken Sie im Standard-Theme von `ggplot2` das einzeichnen der Gitter im Hintergrund.
- iii) **Umgang mit Legenden:** Plotten Sie die Kindersterblichkeit gegen den Zugang zu Sanitäranlagen. Geben Sie den Kontinenten dabei verschiedene Farben, um eine automatische Legende zu erzeugen. Wie können Sie folgende Parameter der Legende verändern: Titel, Position, verwendete Farben, verwendete Symbole?
- iv) **Faceting:** Benutzen Sie `facet_grid()`, um den Scatterplot der Kindersterblichkeit gegen den Zugang zu Sanitäranlagen im Jahr 2013 getrennt nach Kontinenten anzuzeigen. Unterteilen Sie den Plot anschließend zusätzlich nach der kategorisierten Bevölkerungsanzahl.

Aufgabe 2: Scatterplots

Verwenden Sie für diese Aufgaben nur die Daten aus dem Jahr 2013.

- a) Plotten Sie die Kindersterblichkeit gegen den Zugang zu Sanitäranlagen. Stellen Sie die Randverteilungen einmal durch einfache Markierungen an den Achsen dar (`geom_rug()`) und einmal als eigenständige Plots in einem Plot-Gitter (`grid.arrange()` aus Paket `gridExtra`). In welchen Datensituationen ist die erste Darstellungsart problematisch? Wie können Sie diese Problematik umgehen?
- b) Eine weitere Möglichkeit zur Visualisierung von Verteilungen sind Contour plots. Erzeugen Sie eine solche Grafik, indem Sie einen Kerndichteschätzer über obigen Scatterplot legen (`geom_density2d()`).
- c) Plotten Sie erneut die Kindersterblichkeit gegen den Zugang zu Sanitäranlagen. Welche Möglichkeiten haben Sie, weitere Variablen innerhalb des 2D-Scatterplots zu visualisieren? Stellen Sie auf diese Weise zusätzlich den Zugang zu Elektrizität, die Bevölkerungsanzahl sowie den Kontinent dar.
- d) Vergleichen Sie die Kontinente, indem Sie in obigen Scatterplot per `geom_smooth()` LOESS-Glättungskurven einzeichnen. Vergleichen Sie zur besseren Übersichtlichkeit dabei nur Afrika, Amerika und Europa. Unterteilen Sie den Plot schließlich auch per `facet_grid()` in mehrere Fenster.
- e) Welche Probleme können bei Scatterplots für hochdimensionale Daten auftreten? Welche Strategien gibt es, um diese zu umgehen? Simulieren Sie sich per `rnorm()` 5000 Datenpunkte für einen Scatterplot und wenden Sie alle Strategien an.

- f) Stellen Sie die Kindersterblichkeit in einem 3D-Scatterplot gegen den Zugang zu Sanitäranlagen und zu Elektrizität dar. Verwenden Sie hierzu die Funktion `scatterplot3d` aus dem gleichnamigen Paket). Bewerten Sie die Übersichtlichkeit der Graphik.

Aufgabe 3: Spezielle Graphiken

- a) Erster Überblick über Datensatz:

Verwenden Sie `ggpairs()` aus dem Paket `GGally`, um auf einen Blick einen Überblick über alle numerischen Variablen sowie den Kontinent zu bekommen.

Sehen Sie sich zusätzlich die Korrelationsstruktur aller metrischen Variablen mit Hilfe von `corrplot()` aus dem gleichnamigen Paket an. Welche unterschiedlichen Visualisierungsmöglichkeiten von Korrelationen bietet die Funktion?

- b) Vergleich einzelner Beobachtungen / Cluster:

Vergleichen Sie in einem Sterndiagramm fünf Länder Ihrer Wahl bzgl. ihrer Ausprägungen in den metrischen Variablen. Benutzen Sie dafür die (nicht auf `ggplot2` basierende!) Funktion `radarchart()` aus dem Paket `fmsb`.

Machen Sie sich mit der Visualisierung der Chernoff-Gesichter vertraut (ohne diese in R anzuwenden). Für welche Situationen kann diese Darstellungsart sinnvoll sein?

- c) Karten:

Laden Sie sich die Datei `GDP.csv` von der Homepage. Plotten Sie das logarithmierte BIP aller Länder auf einer Weltkarte. Verwenden Sie `map_data()` zum Extrahieren der Polygon-Daten aller Länder und `geom_polygon()` zum Plotten der Weltkarte. Verwenden Sie testweise auch eine alternative Kartendarstellung mit `coord_map('ortho')`.

Aufgabe 4: Interaktive Graphiken

- a) `manipulate`:

Stellen Sie erneut die Kindersterblichkeit gegen den Zugang zu Sanitäranlagen in einem Scatterplot dar. Verwenden Sie die Funktion `manipulate()` aus dem gleichnamigen Paket, um das Jahr, für welches die Daten geplottet werden, interaktiv auswählbar zu machen.

Legen Sie auf den Scatterplot nun zusätzlich eine Kerndichteschätzung und machen Sie beide Bandweiten (in x- und y-Richtung) interaktiv einstellbar.

- b) `shiny`:

`shiny` ist ein R-Paket, welches Funktionen aus R interaktiv zugänglich macht, indem es eine HTML-Oberfläche erzeugt, welche mit jedem Browser angezeigt werden kann. Sehen Sie sich hier Beispielpprogramme an, welche mit shiny erstellt wurden. Verschaffen Sie sich mit diesem Tutorial einen Überblick über die Grundstruktur eines Programms in shiny. Setzen Sie für die Graphik aus Aufgabe a) ein interaktives shiny-Programm auf.