

Fakultät für
**Kultur- und
Sozialwissen-
schaften**

Datenvisualisierung mit ggplot2 | Teil 1

mit Sebastian Kuhn



Themen

Was ist ggplot2?

„Grammatik“ der Visualisierung

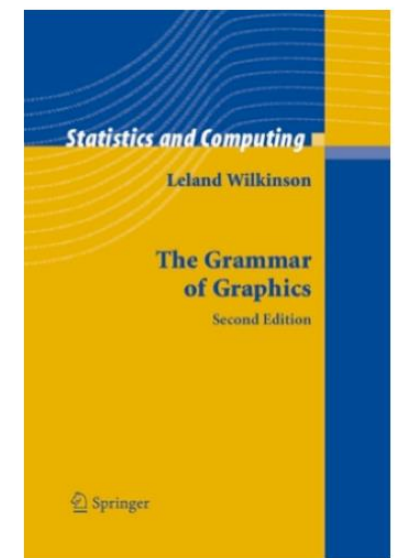
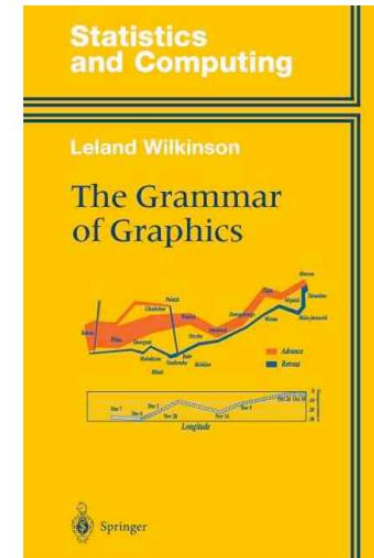
Grundsätzliches zu ggplot2

Was ist ggplot2?

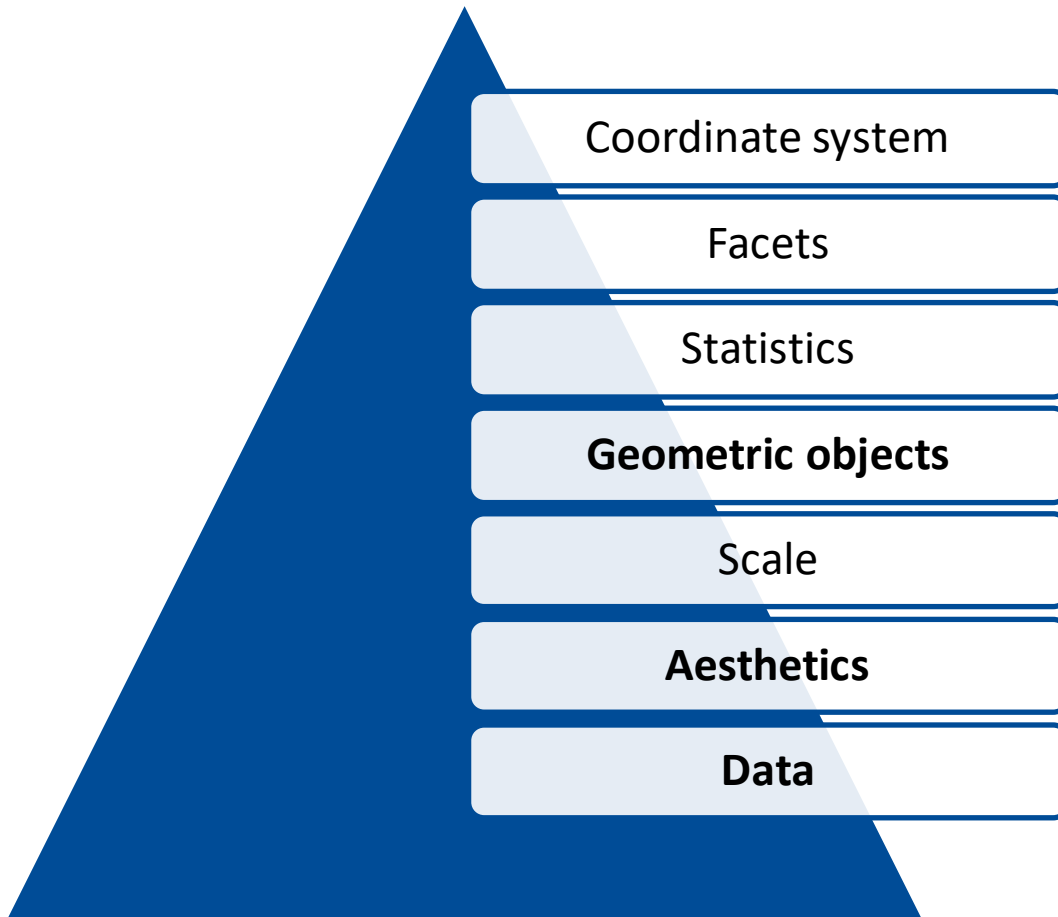
- ggplot2 ist ein (open-source) Paket zur Datenvisualisierung und Teil der tidyverse-Familie (<https://www.tidyverse.org/>)
- ggplot2 wurde 2005 von Hadley Wickham entwickelt und stellt im Prinzip eine open-source-Implementation der „Grammar of Graphics“ von Leland Wilkinson für R dar.
- Die Grundidee der „Grammar of Graphics“ bestand (und besteht) darin, eine grammatikalische Grundstruktur für das Erstellen Datendarstellungen herzuleiten und „Grammatikregeln“ für alle statistischen Grafiken zu beschreiben.

examples that explore different components and their interactions, in more detail. The article concludes by discussing some perceptual issues, and thinking about how we can build on the grammar to learn how to create graphical “poems.”

Aus: Wickham 2010: A Layered Grammar of Graphics



Hauptbestandteile der „Grammar of Graphics“



Daten: Welche Daten (und welche Dimensionen) sollen visualisiert werden?

Ästhetik: Wie sollen die Variablen aus dem Datensatz mit bestimmten visuellen Eigenschaften (Position, Größe, Form, Farbe, etc.) verknüpft werden?

Skalierung: Müssen wir die potenziellen Werte skalieren, eine bestimmte Skala verwenden, um mehrere Werte oder einen Bereich darzustellen?

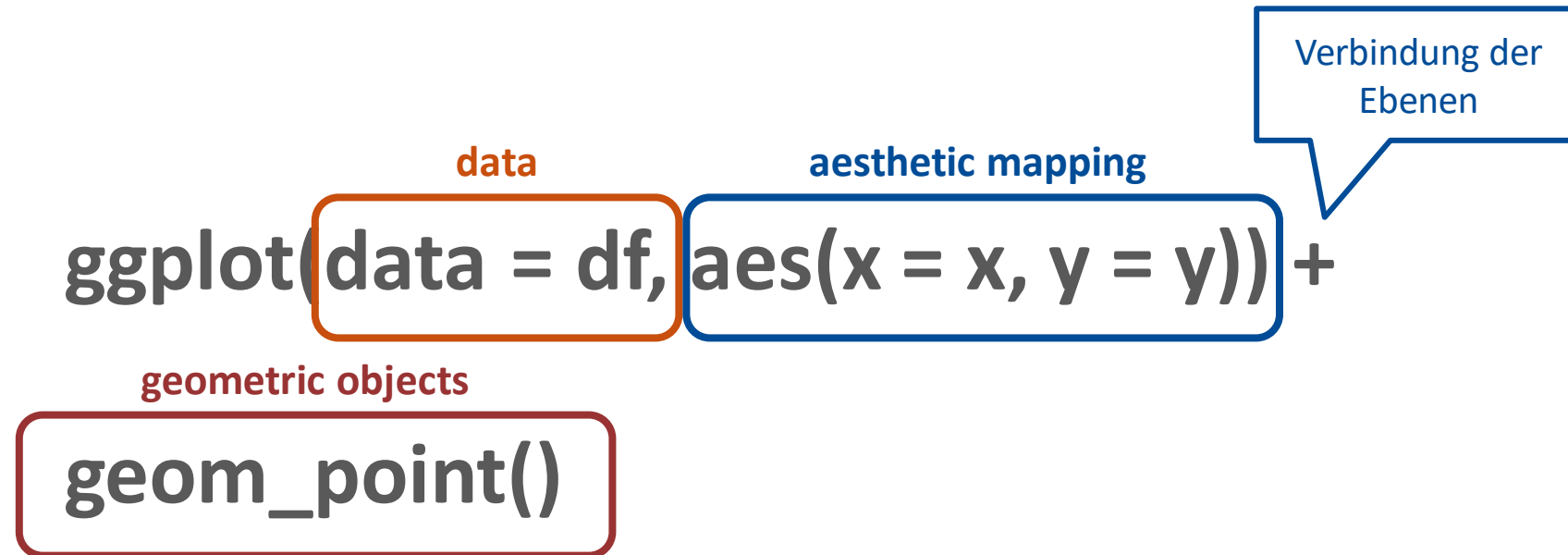
Geometrische Objekte: In welcher Form sollen die Datenpunkte in der Visualisierung dargestellt werden (Punkte, Balken, Linien etc.)?

Statistik: Sollen statistische Kennzahlen (z.B. Kennzahlen für zentrale Tendenz, Streuung oder Konfidenzintervalle) dargestellt werden?

Facetten: Sollen die Darstellungen getrennt für bestimmte Datendimensionen (z.B. einzelne Jahre oder Untersuchungseinheiten) dargestellt werden?

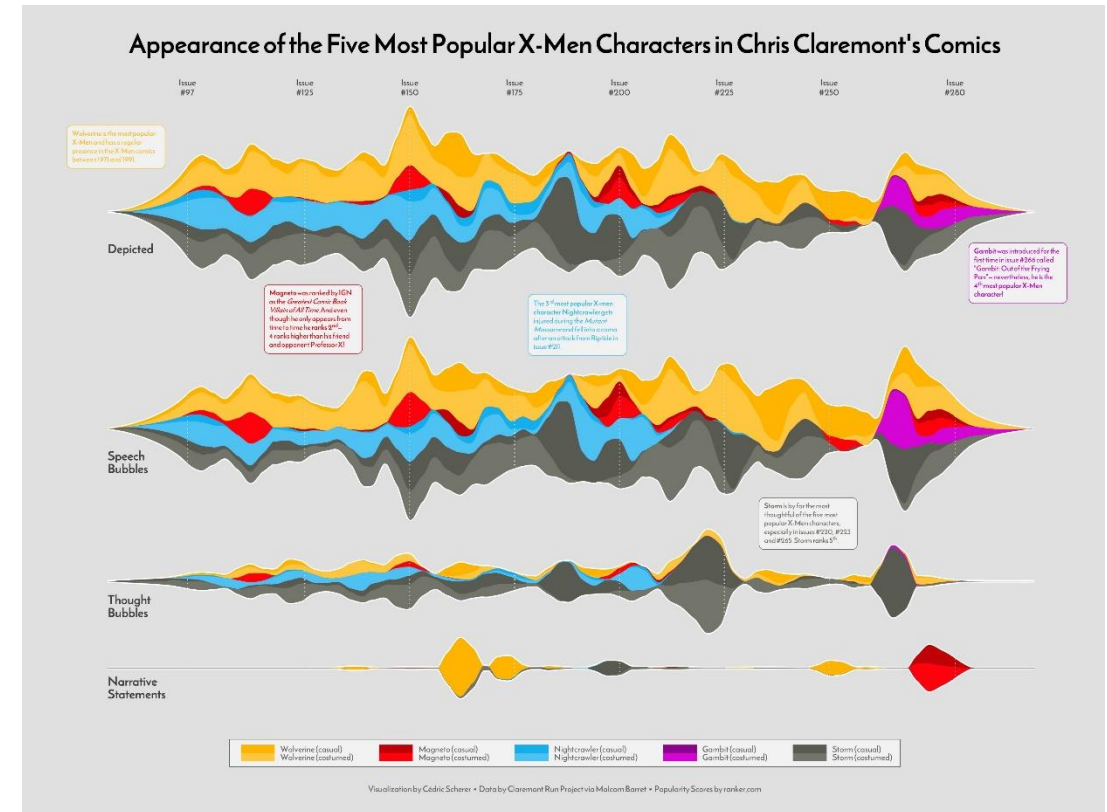
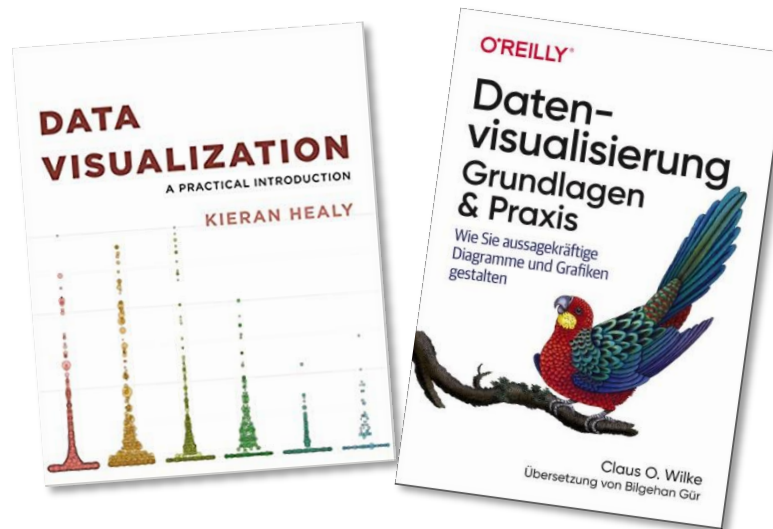
Koordinatensystem: Auf was für einem Welches Koordinatensystem soll für die Visualisierung genutzt werden (z.B. kartesisch oder polar)?

...das ist alles ziemlich viel Theorie!



Seien Sie kreativ!

- Nutzen Sie Blogs und Tutorials wie z.B.:
<https://www.r-graph-gallery.com/>
https://uc-r.github.io/ggplot_intro
<https://www.cedricscherer.com/>
- Die Skripte für viele aktuelle Bücher finden sich häufig auf den Homepages (github) der Autoren!



Quelle: <https://www.cedricscherer.com/2021/01/01/review-2020-personal-dataviz-highlights/>

Hilfe!

- Eine sehr umfangreiche Dokumentation von ggplot2 findet sich unter <https://ggplot2.tidyverse.org/> (häufiger ist eine Suche hier schneller und effizienter als in der Dokumentation in R-Studio).
- Hilfreich sind auch sogenannte „Cheat Sheets“, die Sie (unter anderem) auch unter <https://ggplot2.tidyverse.org/> finden.
- Bei etwas „kniffligeren“ Fällen lohnt sich ein Blick in <https://stackoverflow.com/> - einer Internetplattform, die sich primär an Softwareentwickler richtet.

