

ggplot2 - Theoretische Einführung

Dirk Seidensticker/Clemens Schmid

7. Juli 2017

Geschichte, Einordnung und Philosophie

“Grammatik” und Elemente

Quellen und Empfehlungen

Geschichte, Einordnung und Philosophie



Figure 1: Hadley Wickham

- **Hadley Wickham**

- Chief Scientist von **RStudio** und Adjunct Professor of Statistics an der **Rice University** (Houston, Texas)
- Entwickler etlicher zentraler Pakete für R (dplyr, tidyr, stringr, readr, devtools, DBI, ...)
- <http://had.co.nz/>

- Kontinuierliche Entwicklung **seit 2005**
- **github-Repository** (3400 commits bis zum 04.02.16)
- Kontext
 - Vergangenheit
 - -> **base graphics** (Ross Ihaka in den 90ern)
 - —> “grid”-graphics (Paul Murrell 2000)
 - —> lattice (Deepayan Sarkar)
 - Gegenwart
 - —> **ggplot2** (Hadley Wickham 2005)
 - —> **htmlwidgets**
 - Zukunft
 - —> ggvis (Hadley Wickham 2014) + htmlwidgets
 - ?

- Philosophie

- Einfachere, schnellere und bessere Grafiken
- Didaktisches Konzept:
 - Leichter Einstieg durch Anwendung
 - Verbesserung der eigenen Fähigkeiten durch Begreifen der Theorie
 - Erweitern des Systems durch Einbringen eigener Konstrukte
- Standardisierte “Grammatik” zur Konstruktion von Abbildungen

*ggplot2 is a plotting system for R, based on the **grammar of graphics**, which tries to take the good parts of base and lattice graphics and avoid bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics.*

- <https://github.com/hadley/ggplot2>

- “The Grammar of Graphics” von Leland Wilkinson 2005: Elemente, die jeder statistischen Grafik zugrunde liegen

“Grammatik” und Elemente

data: Daten, die visualisiert werden sollen

```
R <- data.frame(  
  V1 = c(1,2,3,4),  
  V2 = c(3.5,4.3,5.6,6.5),  
  V3 = c("A","A","B","A")  
)
```

R

```
##   V1  V2 V3  
## 1  1 3.5  A  
## 2  2 4.3  A  
## 3  3 5.6  B  
## 4  4 6.5  A
```

noch nicht lauffähig

```
ggplot(data = R)
```


aesthetic mappings: Zuweisung, wie einzelne Variablen aus dem Datenbestand in der Grafik veranschaulicht werden sollen - z.B. durch Farbe, Form, Größe etc.

```
R
```

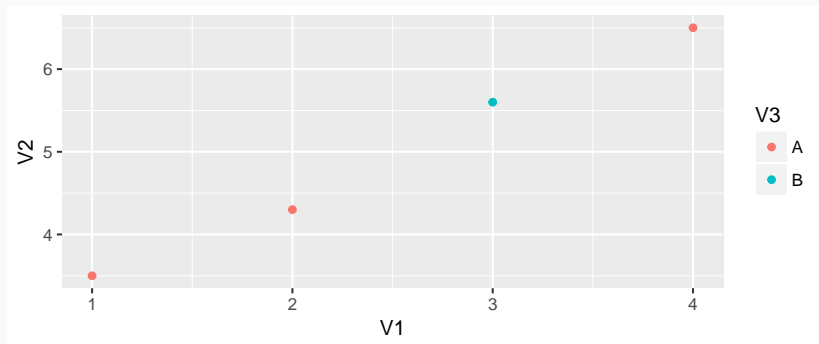
```
##   V1  V2 V3  
## 1   1 3.5 A  
## 2   2 4.3 A  
## 3   3 5.6 B  
## 4   4 6.5 A
```

```
# immer noch nicht lauffähig
```

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3))
```

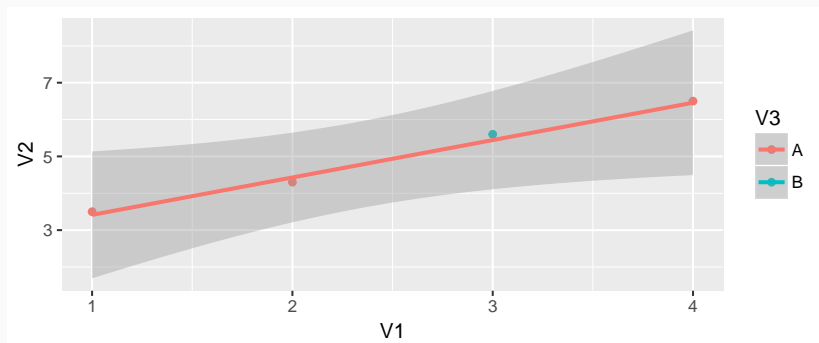
geometric objects: Elemente, aus denen sich der Plot zusammen setzt - z.B. Punkte, Balken, Linien etc. (definiert den Plottyp)

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3)) +  
  geom_point()
```



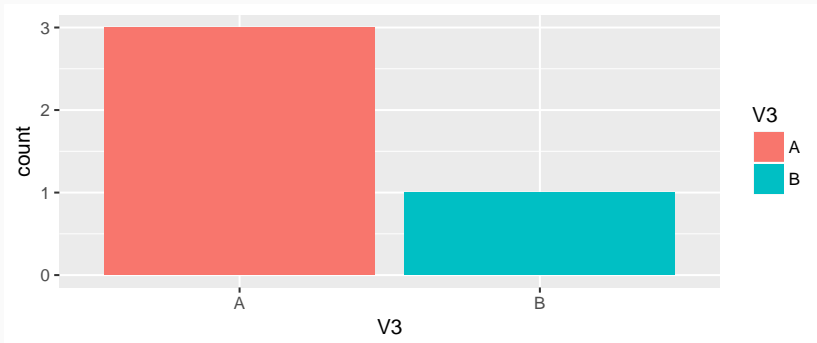
statistical transformations: Stats fassen Daten je nach Bedarf zusammen - z.B. kann man mit ihnen einen Spline einpassen oder die Häufigkeit des Auftretens für ein Histogramm zählen

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3)) +  
  geom_point() +  
  stat_smooth(method = "lm")
```



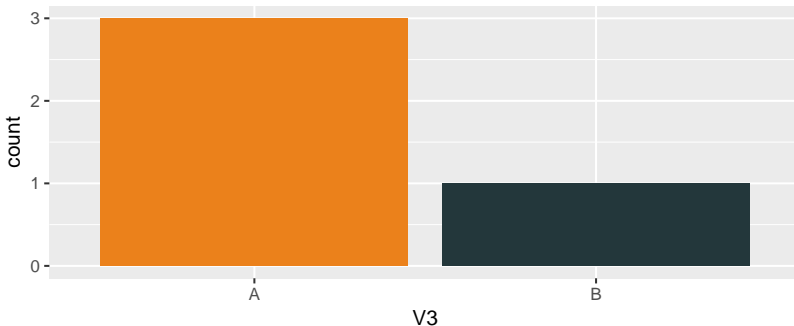
statistical transformations: Stats fassen Daten je nach Bedarf zusammen - z.B. kann man mit ihnen einen Spline einpassen oder die Häufigkeit des Auftretens für ein Histogramm zählen

```
ggplot(data = R, aes(x = V3, fill = V3)) +  
  geom_bar(stat = "count")
```



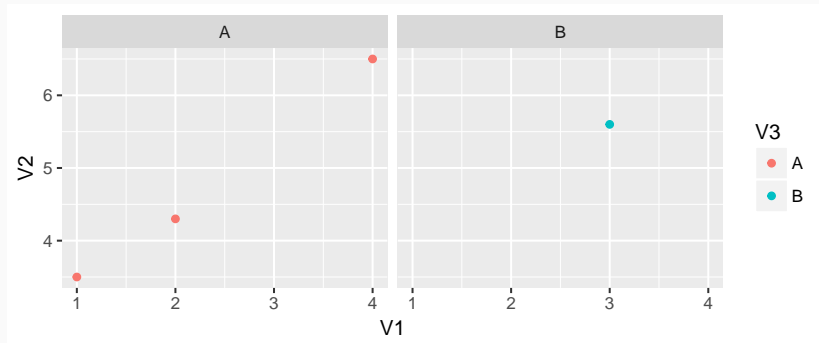
scales: Spezifizierung der Zuordnung von Daten zu grafischer Veranschaulichung (data -> aes). Scales steuert auch Legenden und Achsenskalierung

```
ggplot(data = R, aes(x = V3, fill = V3)) +  
  geom_bar(stat = "count") +  
  scale_fill_manual(  
    values = c('A' = '#EB811B', 'B' = '#23373B'),  
    guide = FALSE  
  )
```



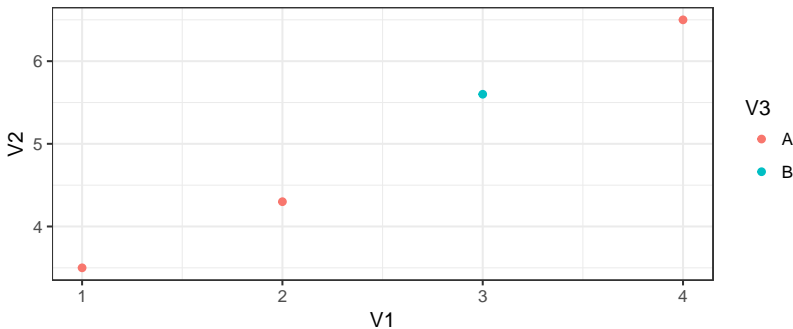
faceting: Aufteilen von Daten nach Kategorien (subsets) und getrennte Visualisierung der Kategorien

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3)) +  
  geom_point() +  
  facet_wrap(~V3)
```



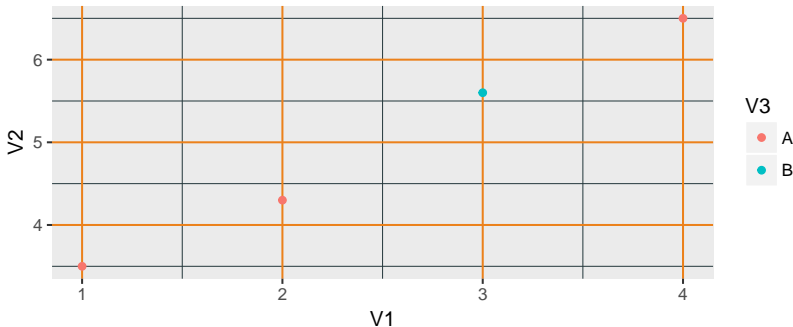
theme: Steuerung der nicht oder nicht unmittelbar datenbezogenen, visuellen Qualitäten eines Plots - z.B. Hintergrundfarbe, Gridlines, Titel, Beschriftungen

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3)) +  
  geom_point() +  
  theme_bw()
```



theme: Steuerung der nicht oder nicht unmittelbar datenbezogenen, visuellen Qualitäten eines Plots - z.B. Hintergrundfarbe, Gridlines, Titel, Beschriftungen

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3)) +  
  geom_point() +  
  theme(  
    panel.grid.minor = element_line(colour = '#23373B'),  
    panel.grid.major = element_line(colour = '#EB811B')  
  )
```



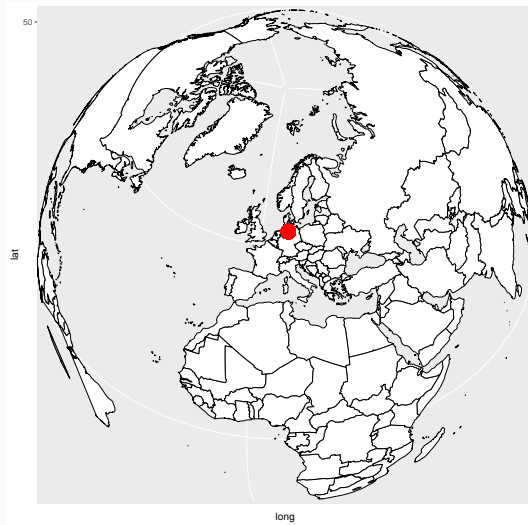
coordinate system: Beschreibt, wie Koordinaten auf die Plotfläche projiziert werden und stellt Achsen und Gridlines bereit. Neben dem normalen Kartesischen Koordinatensystem stehen z.B. auch geographische Koordinatensysteme zur Verfügung

```
library(maps)
```

```
world <- map_data("world")
```

```
worldmap <- ggplot(world, aes(x = long, y = lat, group = group)) +  
  geom_polygon(fill = "white", colour = "black") +  
  coord_map("ortho", orientation = c(48, 9, 0))
```

```
worldmap +  
  geom_point(aes(x = 10, y = 53.33), colour = "red", size = 7)
```



data: Ausgangsdaten

aes: Zuweisung Variablen -> Visualisierung. Spezifizierung mit **scales**

geom: Plottyp/Elemente

stats: Zusammenfassungenoperationen

faceting: Mehrere Teilplots erstellen

theme: Plotlayout/Design

coord: Koordinatensystem

- `ggplot(data, aes()) +`
 - `geom_...()` +
 - `geom_...(aes(), stat)` +
 - `scale_..._...()` +
 - `facet_...(~)` +
 - `theme_...()` +
 - `coord_...()`

Speichern der erstellten Plots als Vektor- oder Rastergrafiken.

```
ggplot(data = R, aes(x = V1, y = V2, colour = V3)) +  
  geom_point()  
ggsave("output/filename.pdf", width = 8, height = 6)  
ggsave("output/filename.png", width = 8, height = 6)
```

Quellen und Empfehlungen

Web

- <http://ggplot2.org/> - ggplot-Homepage
- <https://github.com/hadley/ggplot2> - Github-Repository zu ggplot2
- <http://docs.ggplot2.org/current/index.html> - Online Dokumentation und Hilfe
- <http://had.co.nz/> - Hadley Wickhams Homepage

Literatur

- W. Chang, R graphics cookbook: practical recipes for visualizing data (Sebastopol 2013)
- H. Wickham, ggplot2: elegant graphics for data analysis (New York 2009) und neuere Versionen des Buchs (<https://github.com/hadley/ggplot2-book>)
- L. Wilkinson/G. Wills, The grammar of graphics (New York 2005)