

rstatsZH - Data Science mit R

Datenvisualisierung verschiedener Datentypen

Lars Schöbitz

2021-11-08

Rückblick - Woche 2

Daten erkunden

- Über mapping visueller Eigenschaften
 - `aes()`
- Über Faceting
 - `facet_grid()` - 2 dimensional
 - `facet_wrap()` - 1 dimensional
- Mit Funktionen
 - `str()`
 - `nrow()`
 - `ncol()`
 - `View()`



Ziele für diese Woche

Am Ende dieser Woche könnt ihr:

- die Farb- und Achsenskalierung eines Plots anpassen
- den passenden Plot für eure Daten finden
- einzelne Funktionen des R Package `{dplyr}` anwenden
- beschreiben wie das Symbol `%>%` genutzt wird



Hausaufgabe - Hallo OGD

Übung 3 - Beispiel

Plot

Code

Übung 4 - Beispiel

Plot

Code

Praktikum 3 - {ggplot2}

1. **RStudio Cloud**: Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
2. **RStudio Cloud / Projects**: Öffne erneut das Praktikum 03
3. Folgt wieder auf dem Bildschirm

Eine Skala ersetzen - Farbskalen (color)

```
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, colour = species)) +  
  geom_point() +  
  scale_color_viridis_d()
```

Eine Skala ersetzen - Farbskalen (fill)

```
ggplot(penguins, aes(x = species, fill = island)) +  
  geom_bar() +  
  scale_fill_viridis_d()
```

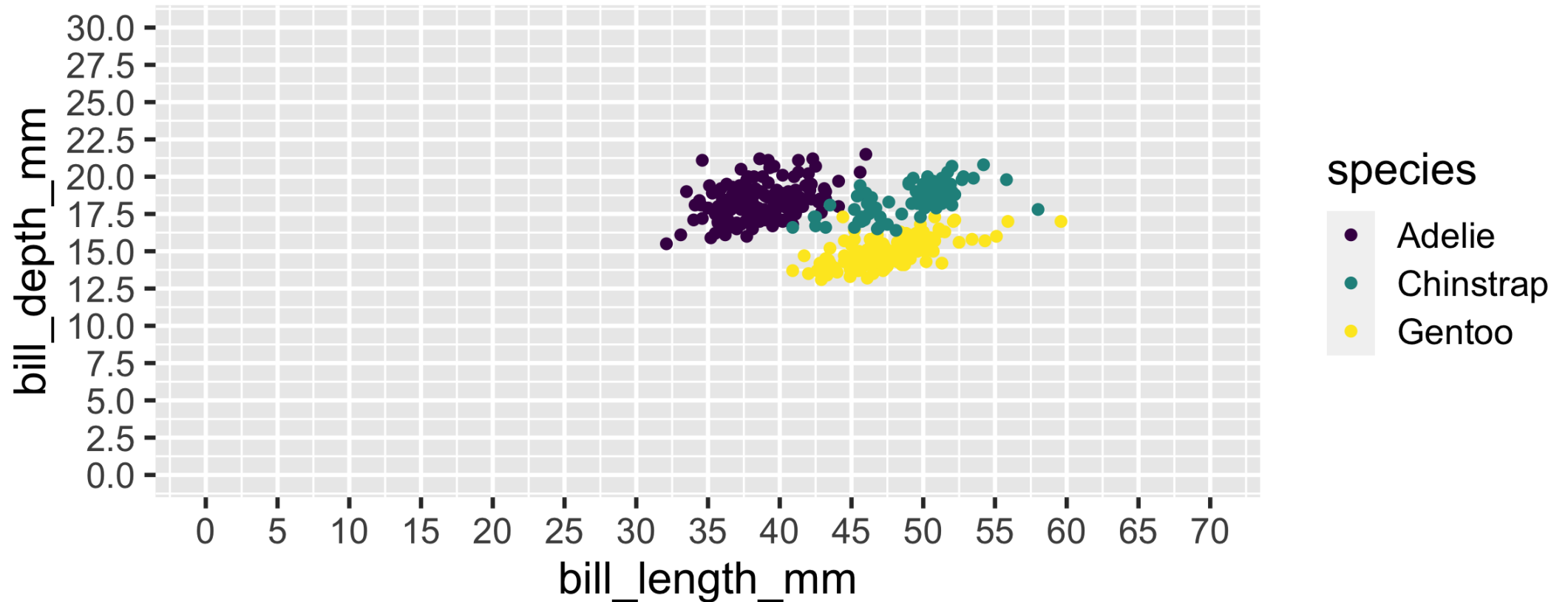
Eine Skala ersetzen - Farbskalen (colorbrewer2.org)

```
ggplot(penguins, aes(x = species, fill = island)) +  
  geom_bar() +  
  scale_fill_brewer(type = "qual", palette = "Set2")
```

Eine Skala ersetzen - Achsen (kontinuierliche Variable)

Plot

Code



Praktikum 3 - Lösungen

- **GitHub Organisation:** rstatsZH
 - github.com/rstatsZH
- **Repo:** prak-03-vis-ggplot2
 - <https://github.com/rstatsZH/prak-03-vis-ggplot2>
- **R Markdown Datei:** prak-03-solutions.Rmd
 - <https://github.com/rstatsZH/prak-03-vis-ggplot2/blob/main/prak-03-solutions.Rmd>

Variablen Typen

Variablen Typen

Numerisch

Diskrete Variablen

- nicht negative
- zählbare
- ganze Zahlen
- z.B. Anzahl Schüler, Würfelwurf

Stetige (kontinuierliche) Variablen

- unendliche Anzahl von Werten
- zwischen zwei Werten
- auch Datums/Uhrzeitwerte
- z.B. Länge, Gewicht, Grösse

Nicht numerisch

Kategoriale Variablen

- endliche Anzahl von Werten
- eindeutige Gruppen (z.B. EU Länder)
- **ordinal**, wenn diese eine logische Reihenfolge/Rangordnung aufweisen (z.B. Wochentage, Schulnoten)

Variablen Typen - Penguins

```
penguins
```

```
# A tibble: 344 × 8
  species island    bill_length_mm bill_depth_mm flipper_length_...
  <fct>    <fct>          <dbl>          <dbl>          <int>
1 Adelie  Torgersen        39.1           18.7           181
2 Adelie  Torgersen        39.5           17.4           186
3 Adelie  Torgersen        40.3            18           195
4 Adelie  Torgersen         NA            NA            NA
5 Adelie  Torgersen        36.7           19.3           193
6 Adelie  Torgersen        39.3           20.6           190
# ... with 338 more rows, and 3 more variables: body_mass_g <int>,
#   sex <fct>, year <int>
```

Variablen Typen - Penguins

```
str(penguins)
```

```
tibble [344 × 8] (S3: tbl_df/tbl/data.frame)
 $ species      : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1
 $ island       : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3
 $ bill_length_mm : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 .
 $ bill_depth_mm : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 .
 $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
 $ body_mass_g   : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250
 $ sex          : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA
 $ year         : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
```

Variablen Typen - Penguins

Variable	Variablen Typ 1	Variablen Typ 2
species	kategorisch	nicht ordinal
island	kategorisch	nicht ordinal
bill_length_mm	numerisch	kontinuierlich
bill_depth_mm	numerisch	kontinuierlich
flipper_length_mm	numerisch	kontinuierlich
body_mass_g	numerisch	kontinuierlich
sex	kategorisch	nicht ordinal
year	numerisch	diskret

from Data to Viz

EXPLORE

STORY

ALL

CAVEATS

POSTER

ABOUT

CONTACT



from Data to Viz

Zwei Bücher für eure weitere Reise

- `{ggplot2}`: Elegant Graphics for Data Analysis <https://ggplot2-book.org/index.html>
- Fundamentals of Data Visualization: <https://clauswilke.com/dataviz/>



Für die Aufmerksamkeit!

Für die R packages `{xaringan}` und `{xaringanthemer}` mit welchen die Folien geschrieben wurden.

Eine PDF Version der Folien kann hier heruntergeladen werden:

https://github.com/rstatsZH/website/raw/master/slides/e1_d03-data-vis-num/e1_d03-data-vis-num.pdf

Für [Data Science in a Box](#) und [Remaster the Tidyverse](#), von welchen ich Materialien für diesen Kurs nutze und welche genau wie diese Folien mit [Creative Commons Attribution Share Alike 4.0 International](#) lizenziert sind.