

rstatsZH - Data Science mit R

Datenvisualisierung verschiedener Datentypen

Lars Schöbitz

2020-03-22

Rückblick - Woche 2

Daten erkunden

- Über mapping visueller Eigenschaften
 - `aes()`
- Über Faceting
 - `facet_grid()` - 2 dimensional
 - `facet_wrap()` - 1 dimensional
- Mit Funktionen
 - `glimpse()`
 - `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

1. **RStudio Cloud:** Öffnet den Arbeitsbereich für den Kurs
2. **RStudio Cloud - Projects:** Öffnet das Projekt für Hausaufgabe 2
3. **File-Manager:** Öffnet eure R Markdown Datei (ha-02.Rmd) für die Hausaufgabe 2
4. Strickt das Dokument

Wer möchte das Ergebnisse der Übung 3 präsentieren?

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()
```

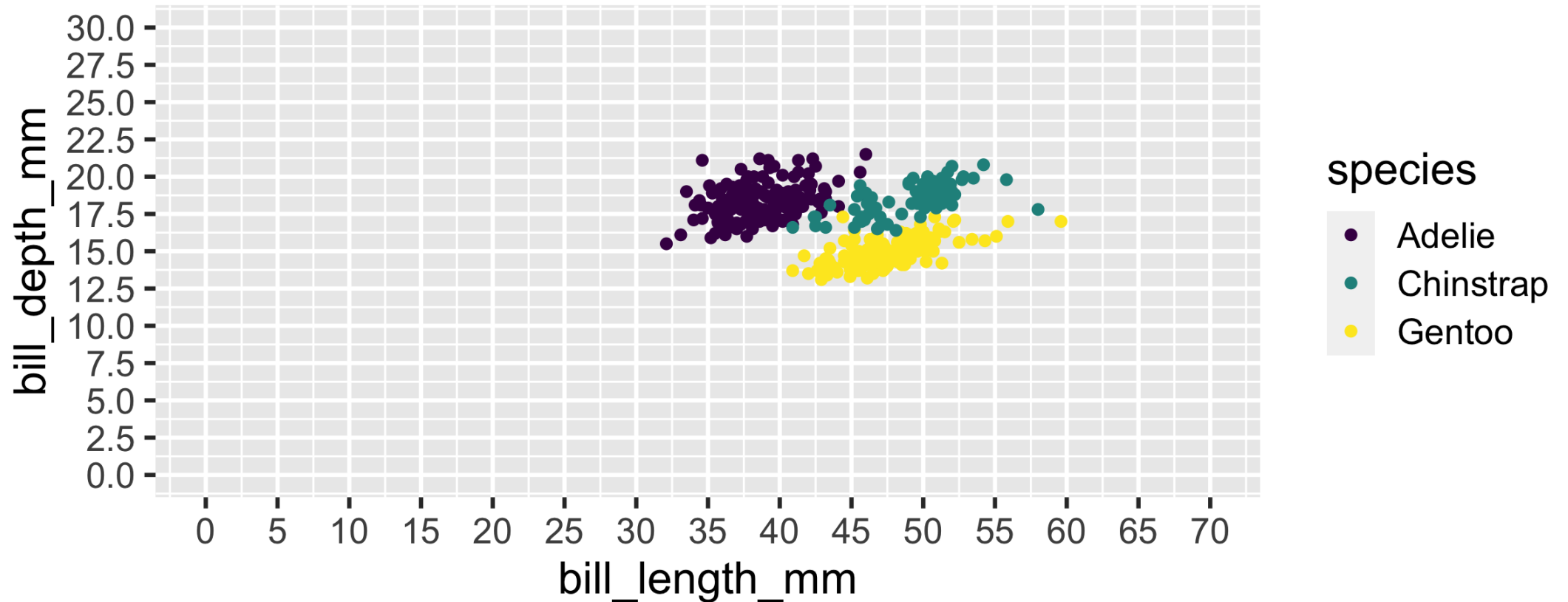
Eine Skala ersetzen - Farbskalen (fill)

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

Eine Skala ersetzen - 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)

Übung

Arbeitet in 2er Teams.

1. **RStudio Cloud**: Öffnet euren Arbeitsbereich für den Kurs
2. **RStudio Cloud / Projects**: Öffnet Praktikum 3
3. Unter dem Kapitel "Daten kennenlernen"
 - erkundet den Datensatz `penguins` mit den Funktionen `glimpse()` und `str()`
 - öffnet die Hilfedatei für den Datensatz `penguins`
 - beschreibt den Variablen Typen für jede Variable im Datensatz

Variable	Variablen Typ 1	Variablen Typ 2
Var 1	numerisch/kategorisch	diskret/stetig (kontinuierlich), oder ordinal/nicht ordinal

10:00

Lösung

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.