

# Session 05: Datenvisualisierung mit ggplot2

Dominic Schmitz & Janina Esser

Verein für Diversität in der Linguistik

# Datenvisualisierung

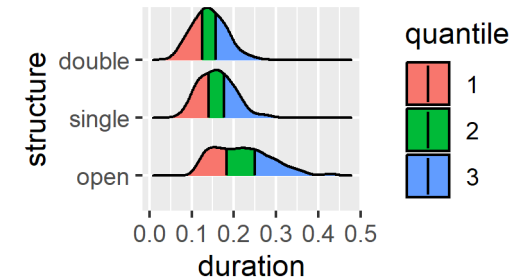
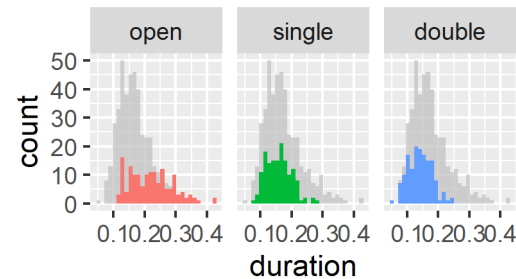
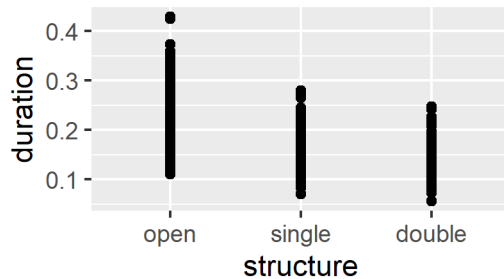
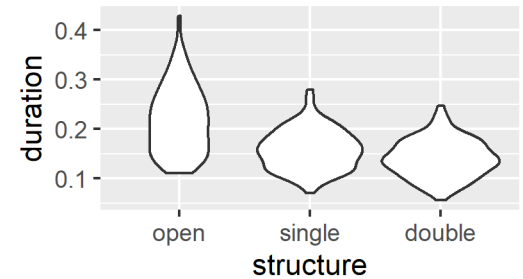
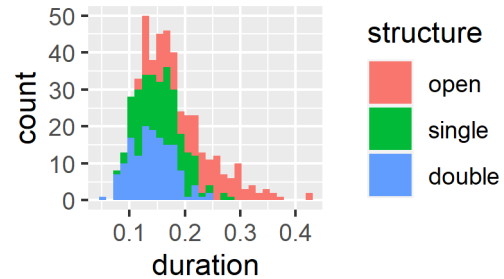
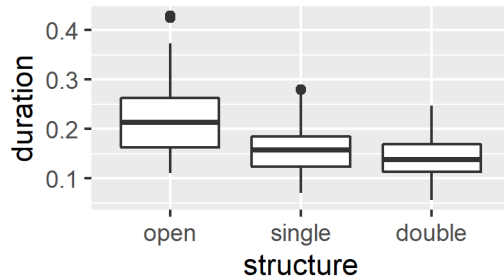
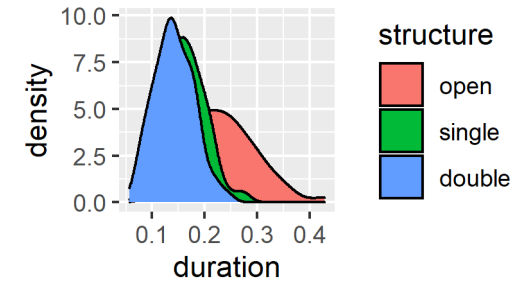
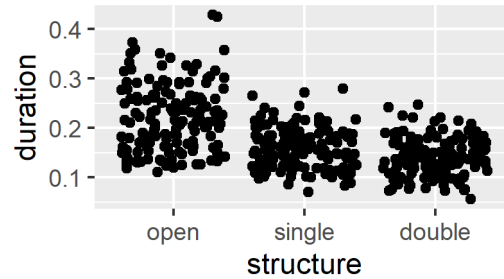
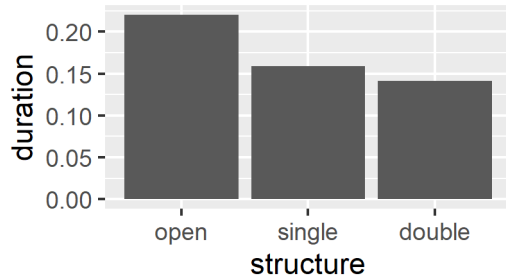


- Datenvisualisierung kennt zwei Hauptmotivationen
  1. Darstellung:

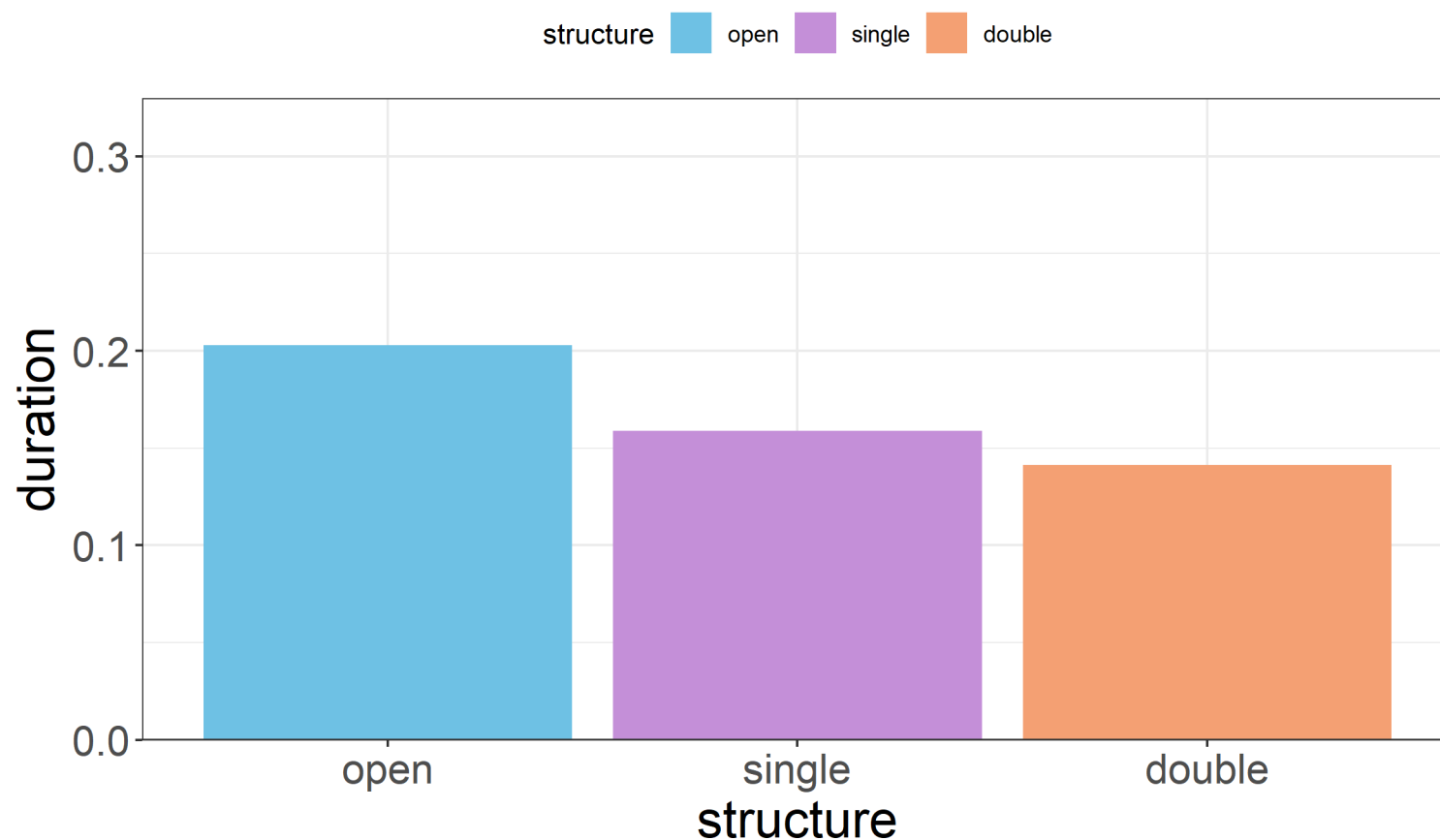
Daten sollen für eine bestimmte Zielgruppe aufbereitet dargestellt werden, um eine bestimmte Botschaft mitzuteilen
  2. Analyse:

Zur Exploration der Daten können visuelle Darstellungen manchmal hilfreicher sein als bloße Zahlen
- Für alle Fälle gibt es nahezu endlos viele Möglichkeiten zur Darstellung; die Wahl des richtigen Formats ist oftmals entscheidend

# 2 Variablen, $\infty$ Plots



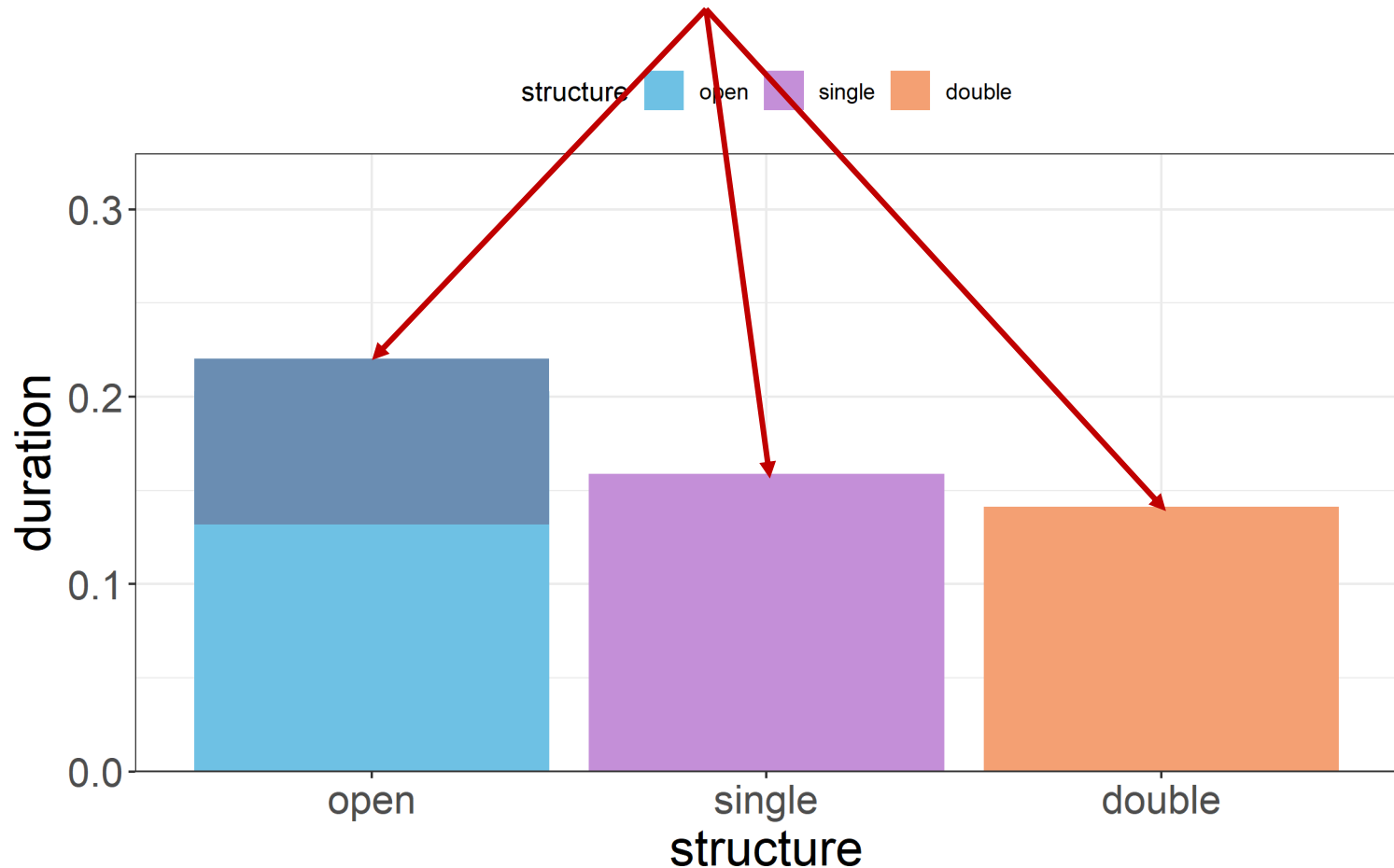
# Balkendiagramm / Bar Plot / Bar Chart / Bar Graph



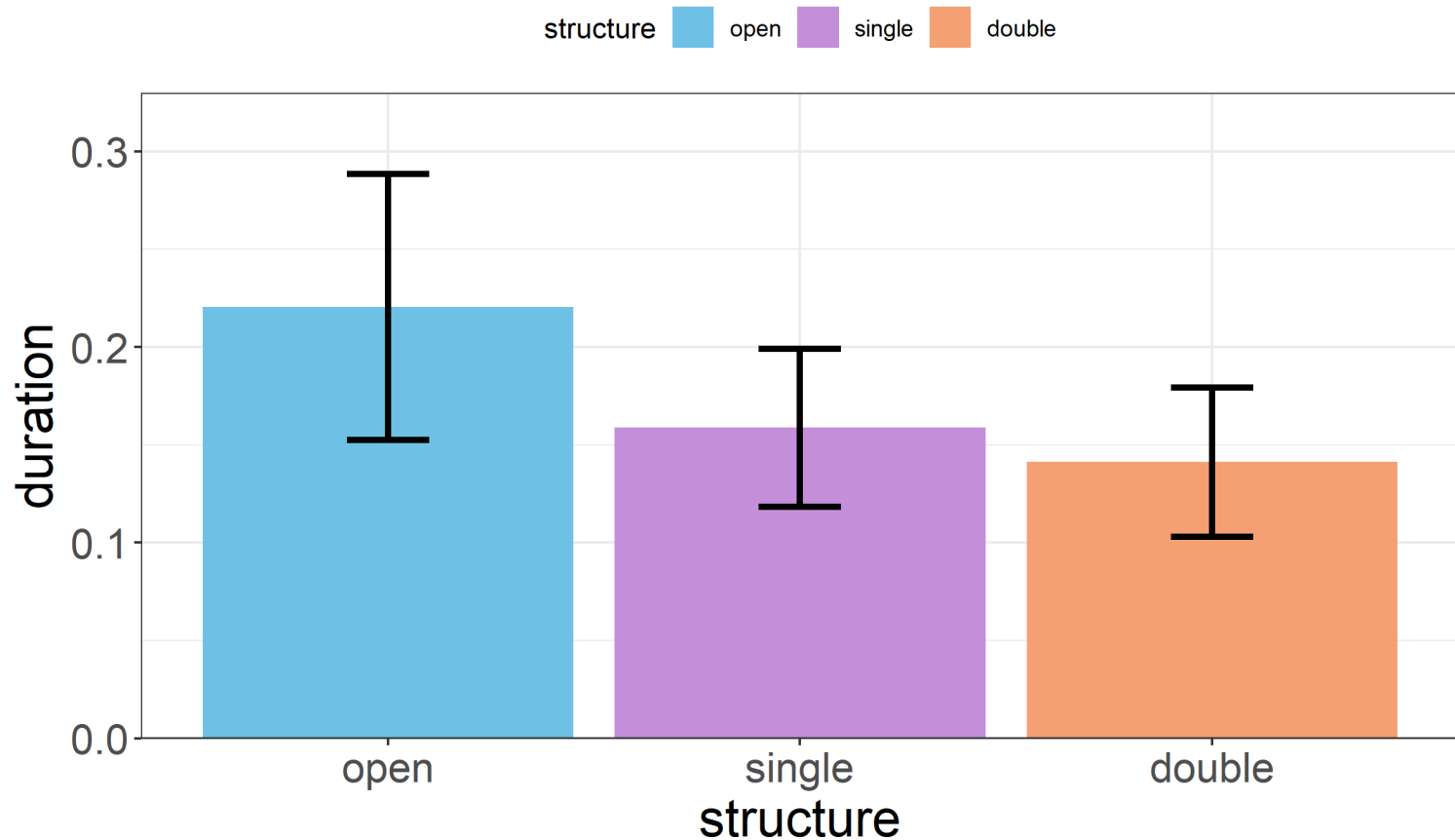
# Balkendiagramm / Bar Plot / Bar Chart / Bar Graph



## Durchschnitt pro Gruppe



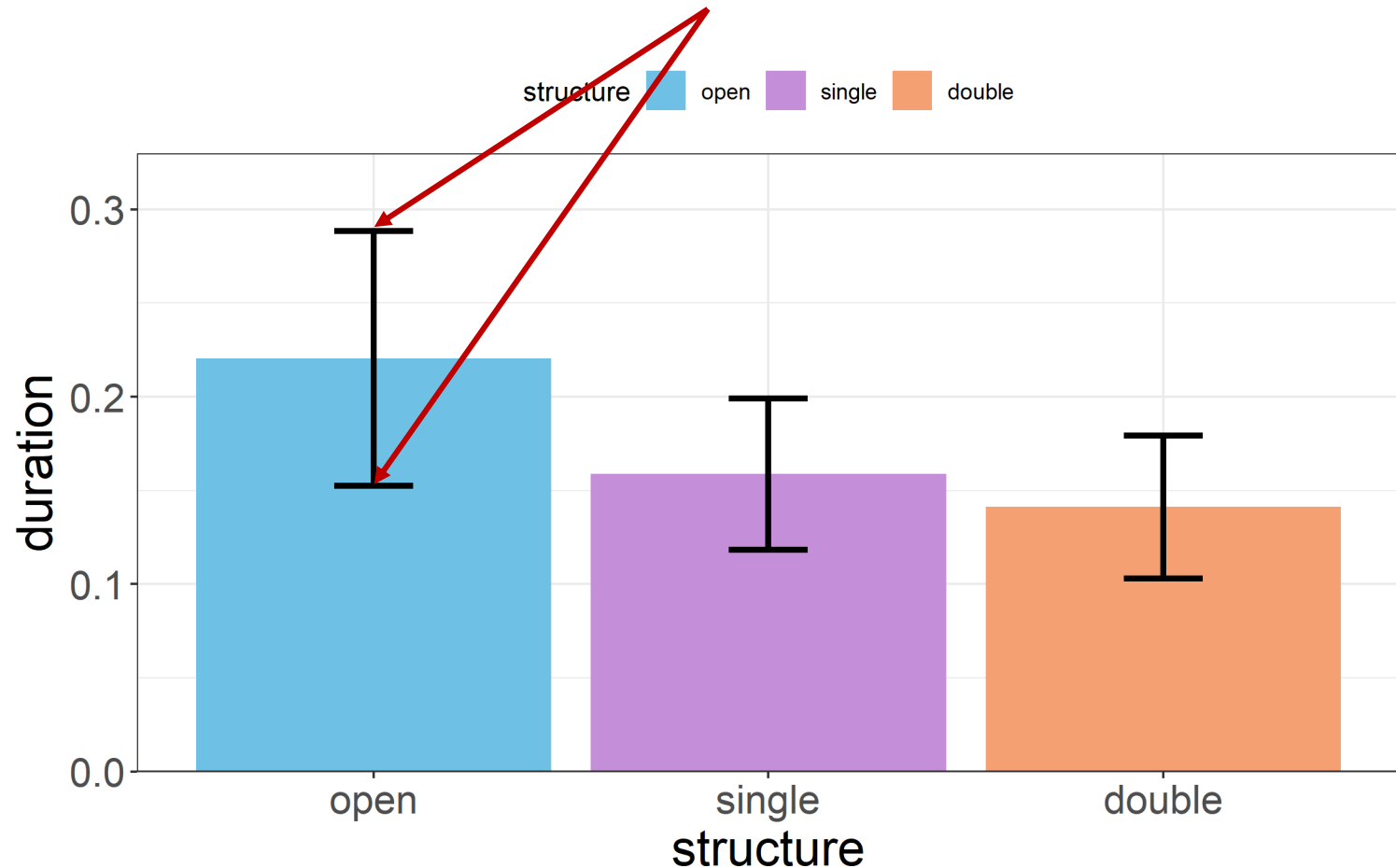
# Balkendiagramm / Bar Plot / Bar Chart / Bar Graph



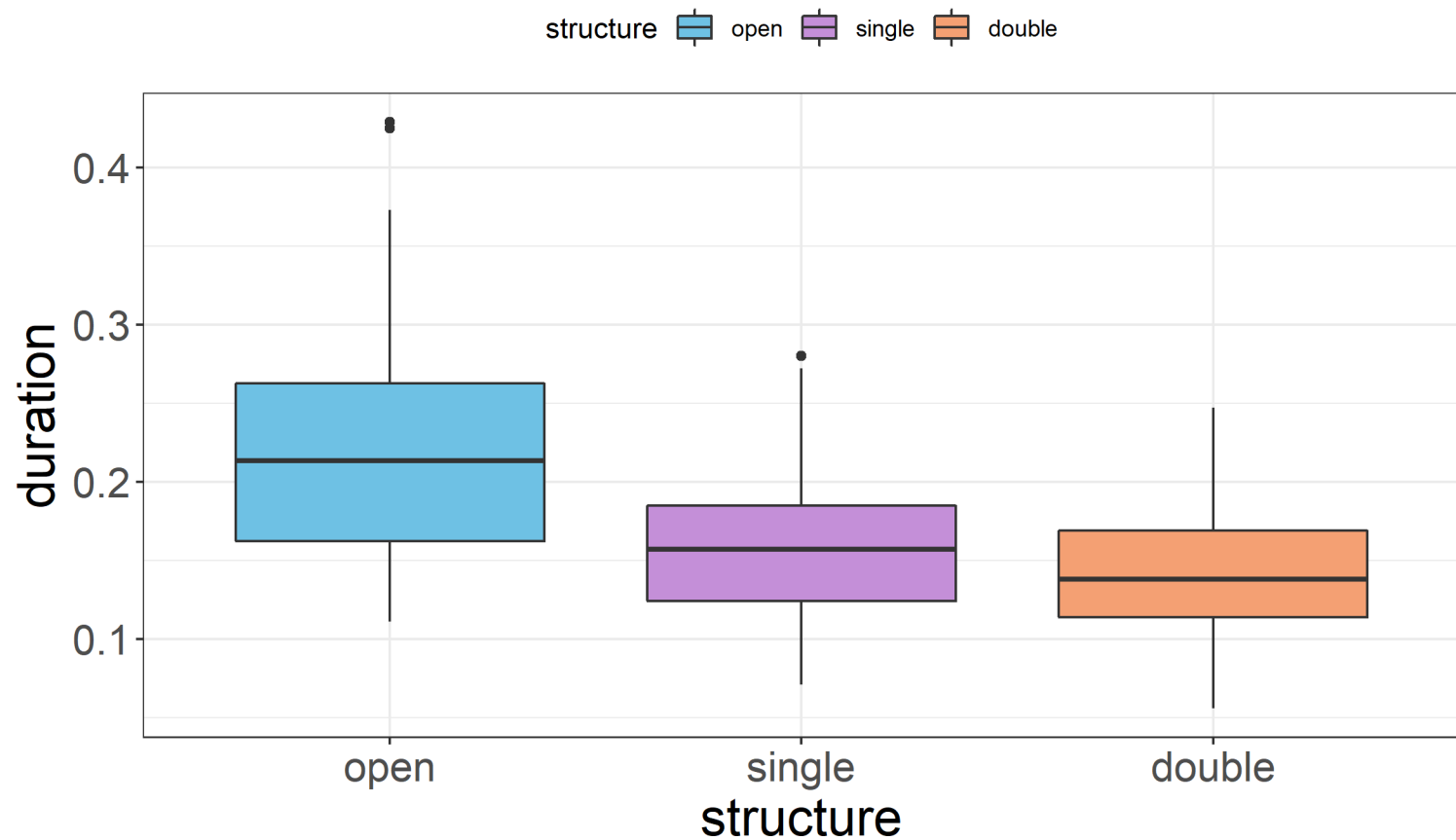
# Balkendiagramm / Bar Plot / Bar Chart / Bar Graph



**Durchschnitt  $\pm$  1 Standardabweichung**



# Box-Plot / Box-and-Whisker Diagram





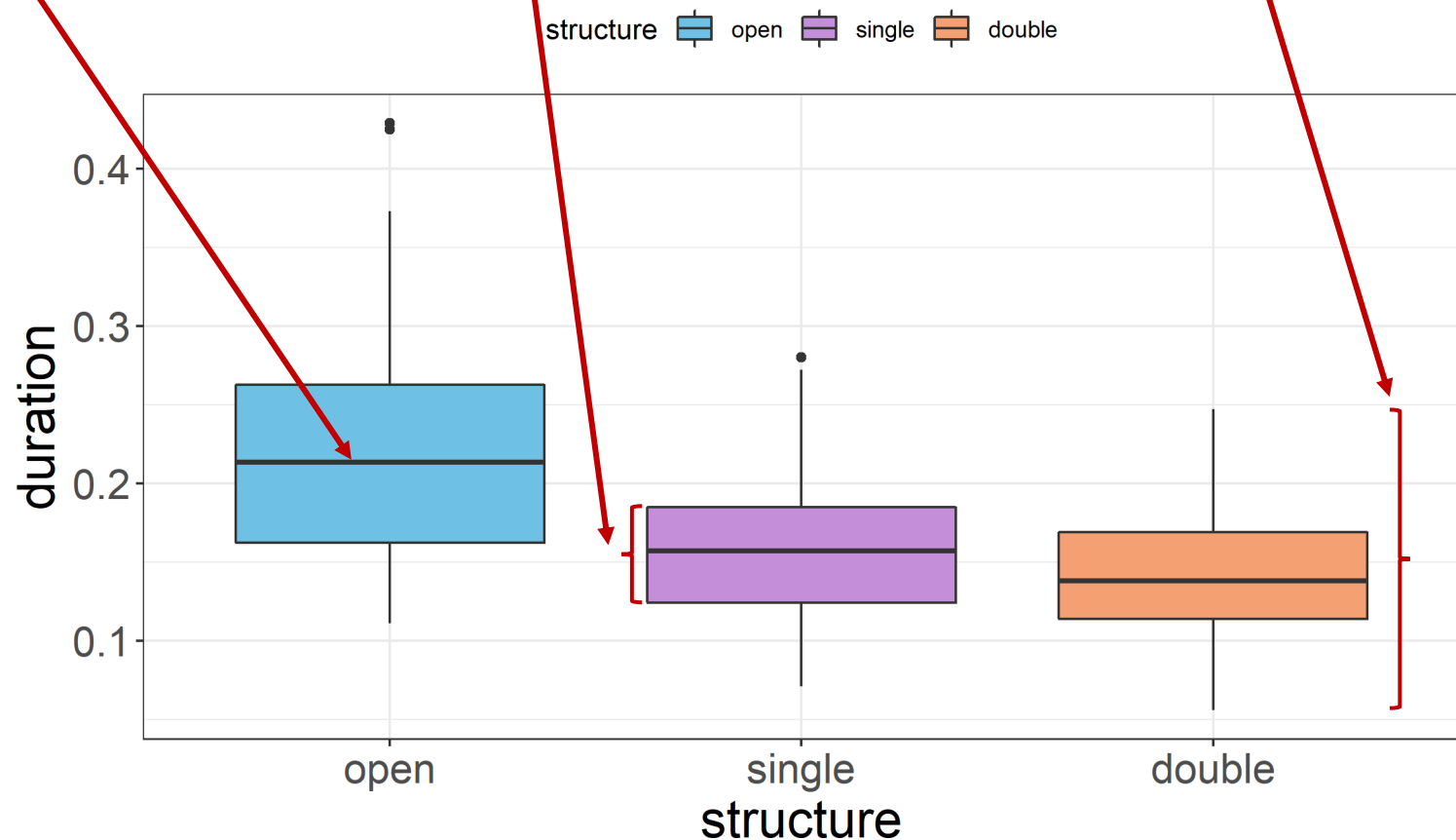
# Box-Plot / Box-and-Whisker Diagram



**Median**

**Interquartilsabstand**

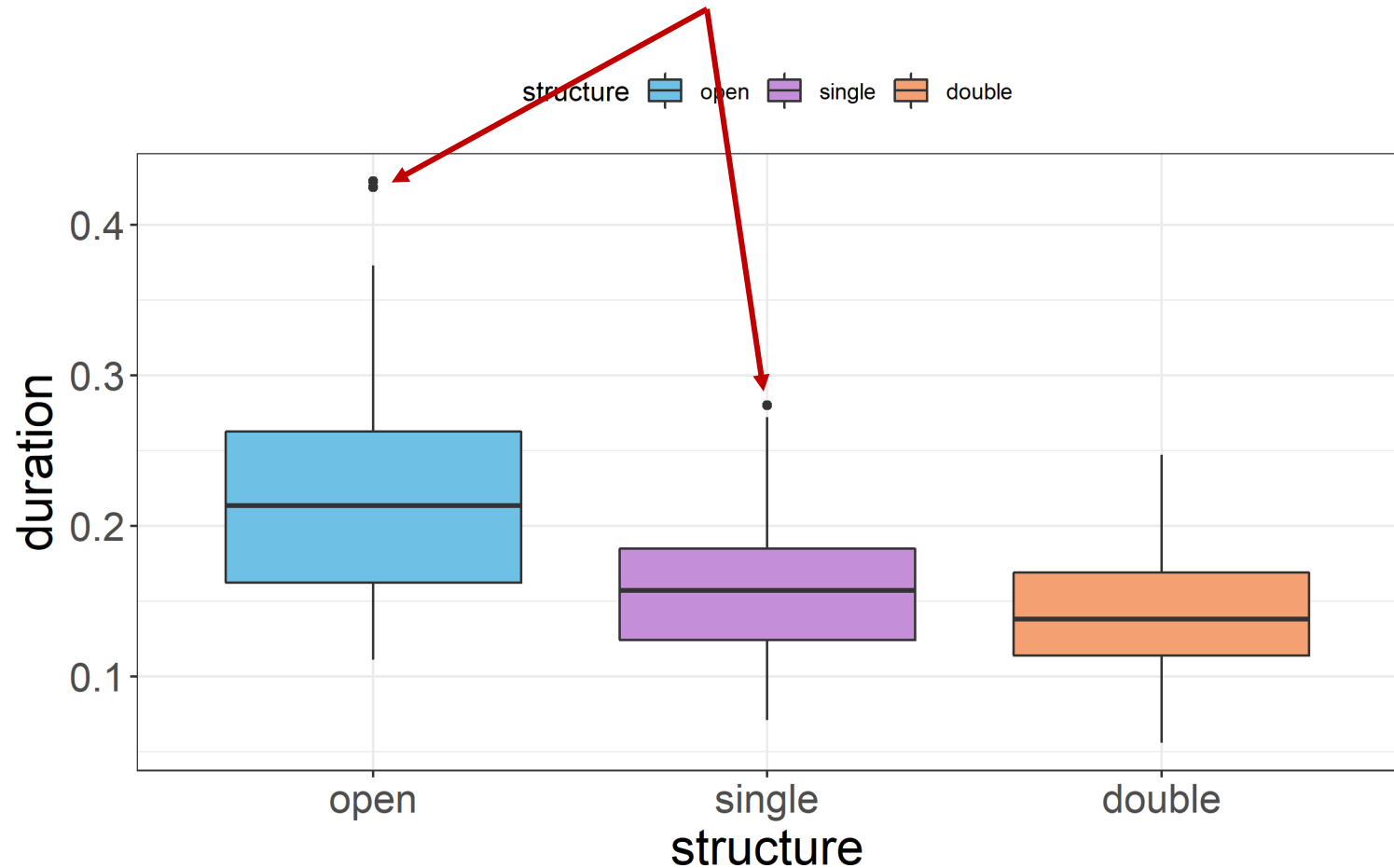
**Werte  $\pm 1.5$   
Interquartilsabständen**



# Box-Plot / Box-and-Whisker Diagram



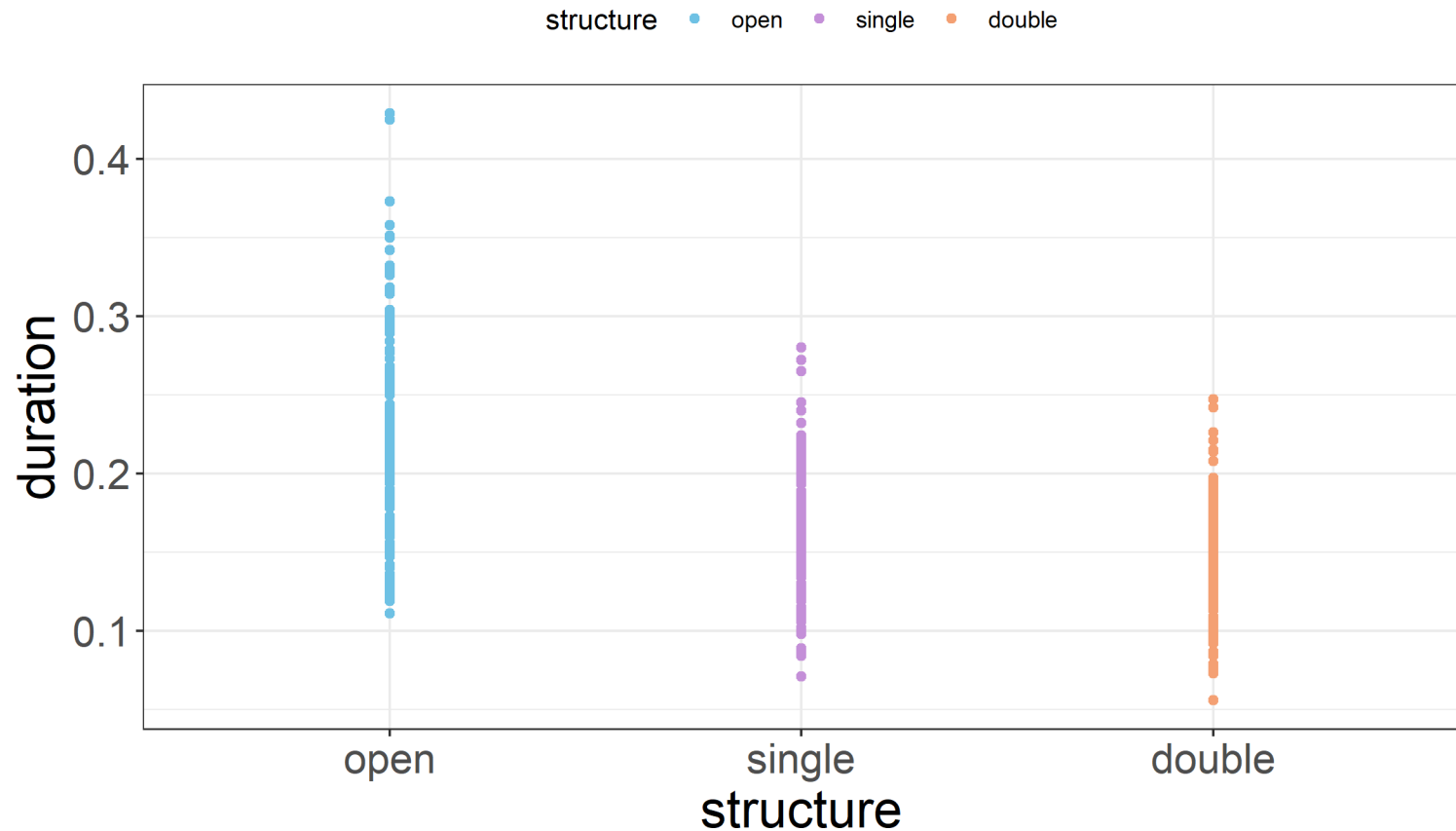
**Outliers / Ausreißer**



# Punktdiagramm / Point Plot / Dot Plot / Dot Chart



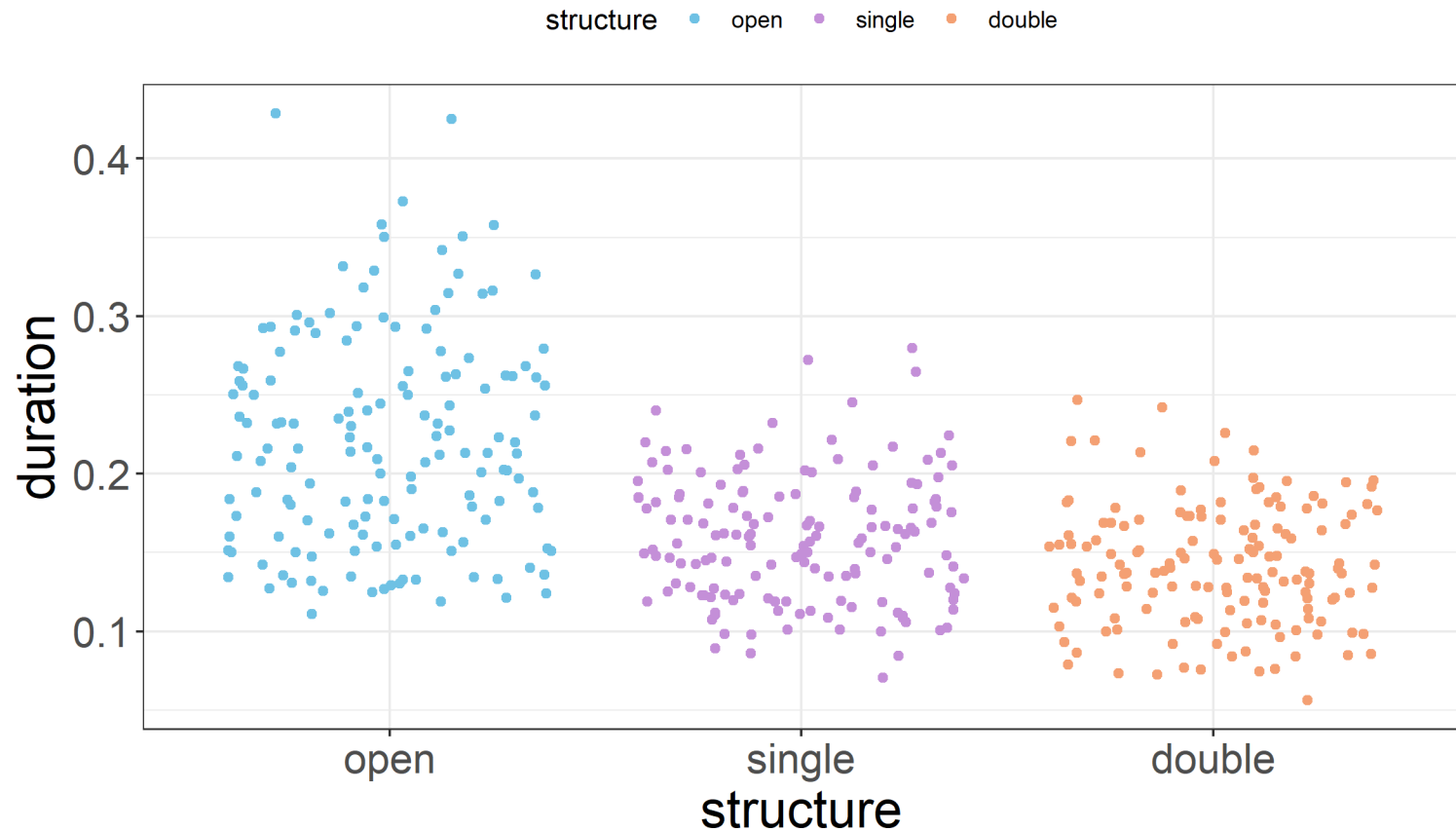
**1 Punkt = 1 Datenpunkt**



# Jitter Plot / Scatter Plot / Scattergram



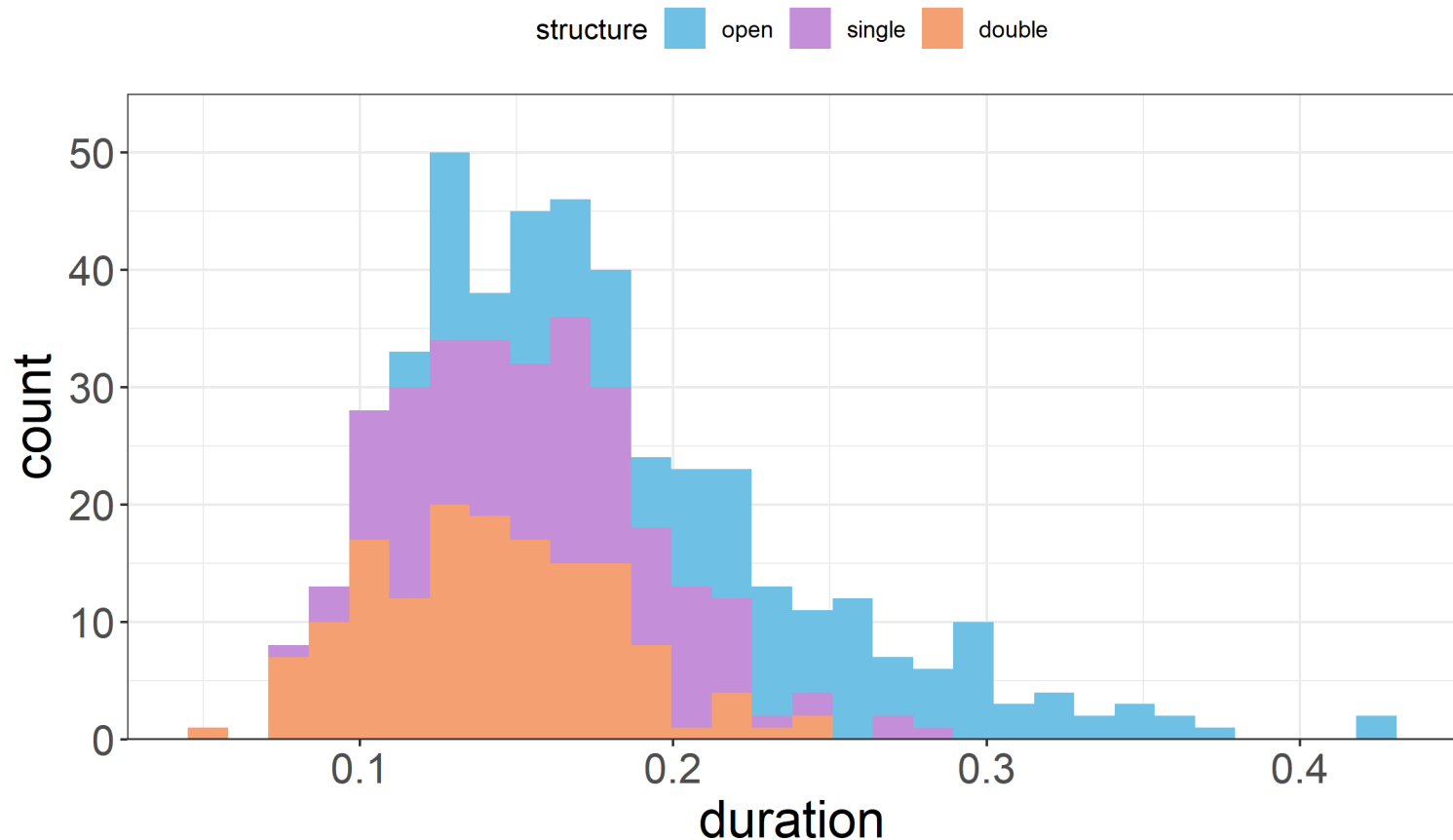
**1 Punkt = 1 Datenpunkt**



# Histogramm



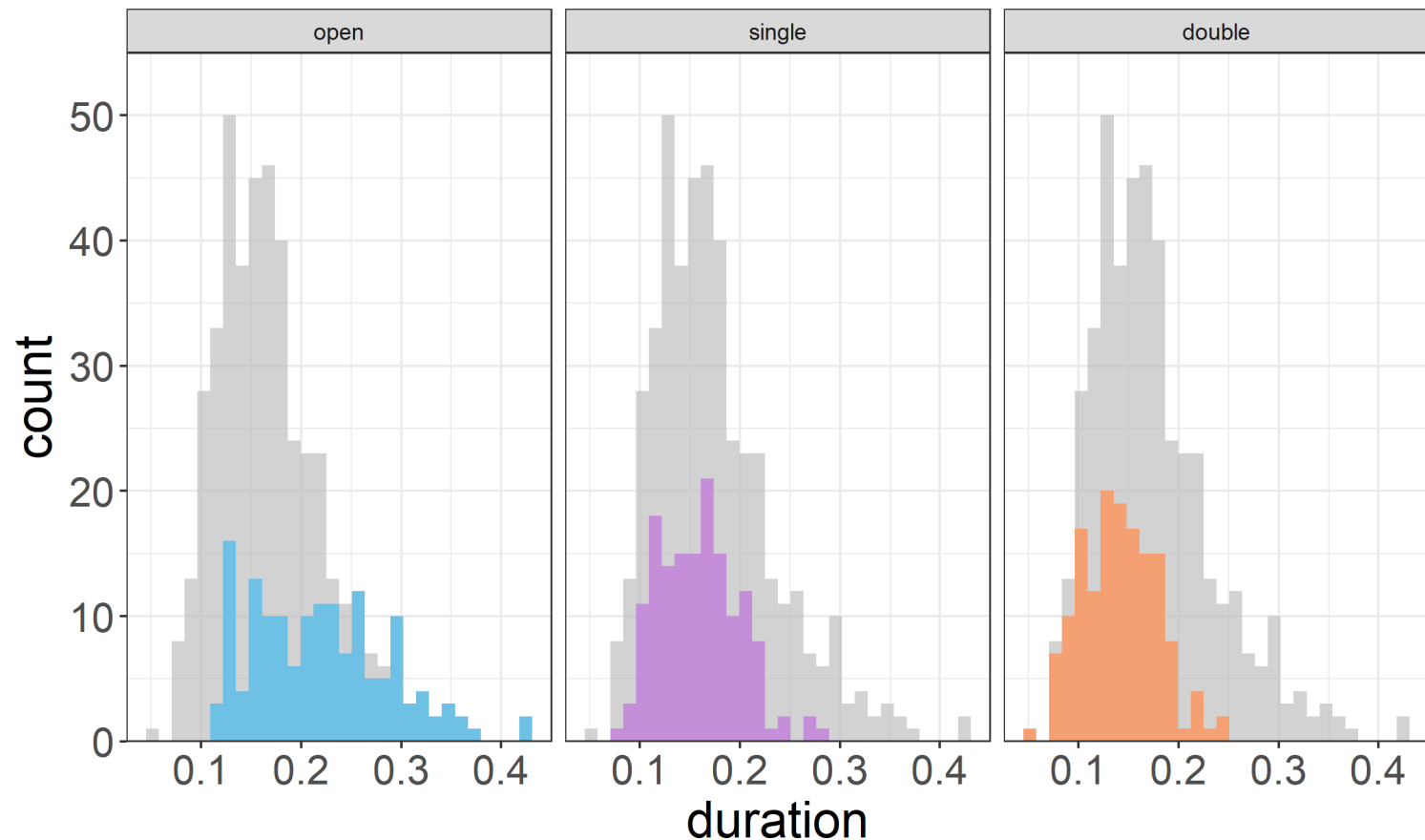
nützlich zum Check der Verteilung – Lesbarkeit?



# Histogramm



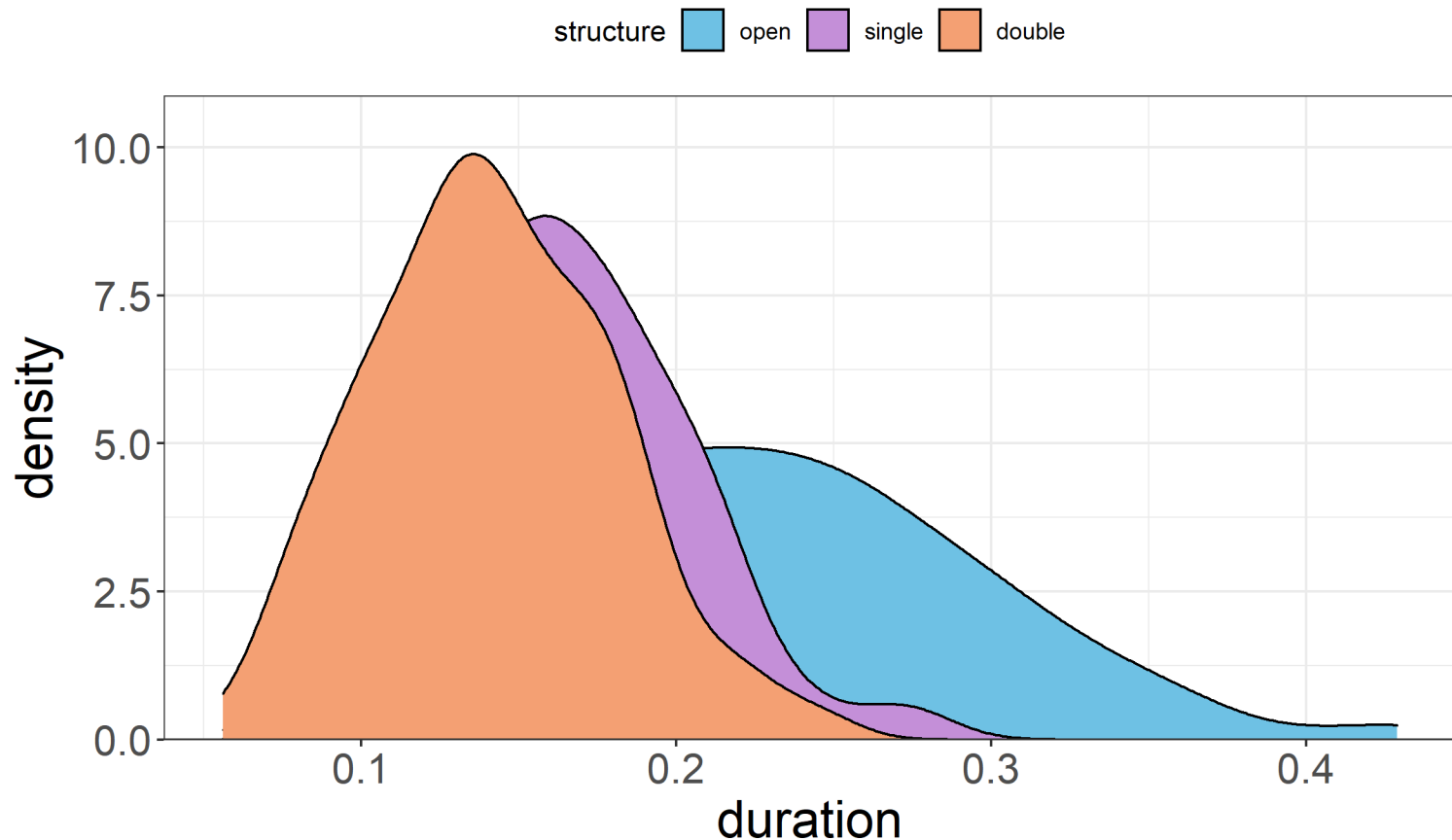
ein Plot pro Level; Hintergrund = alle Daten



# Dichte-Diagramm / Density Plot



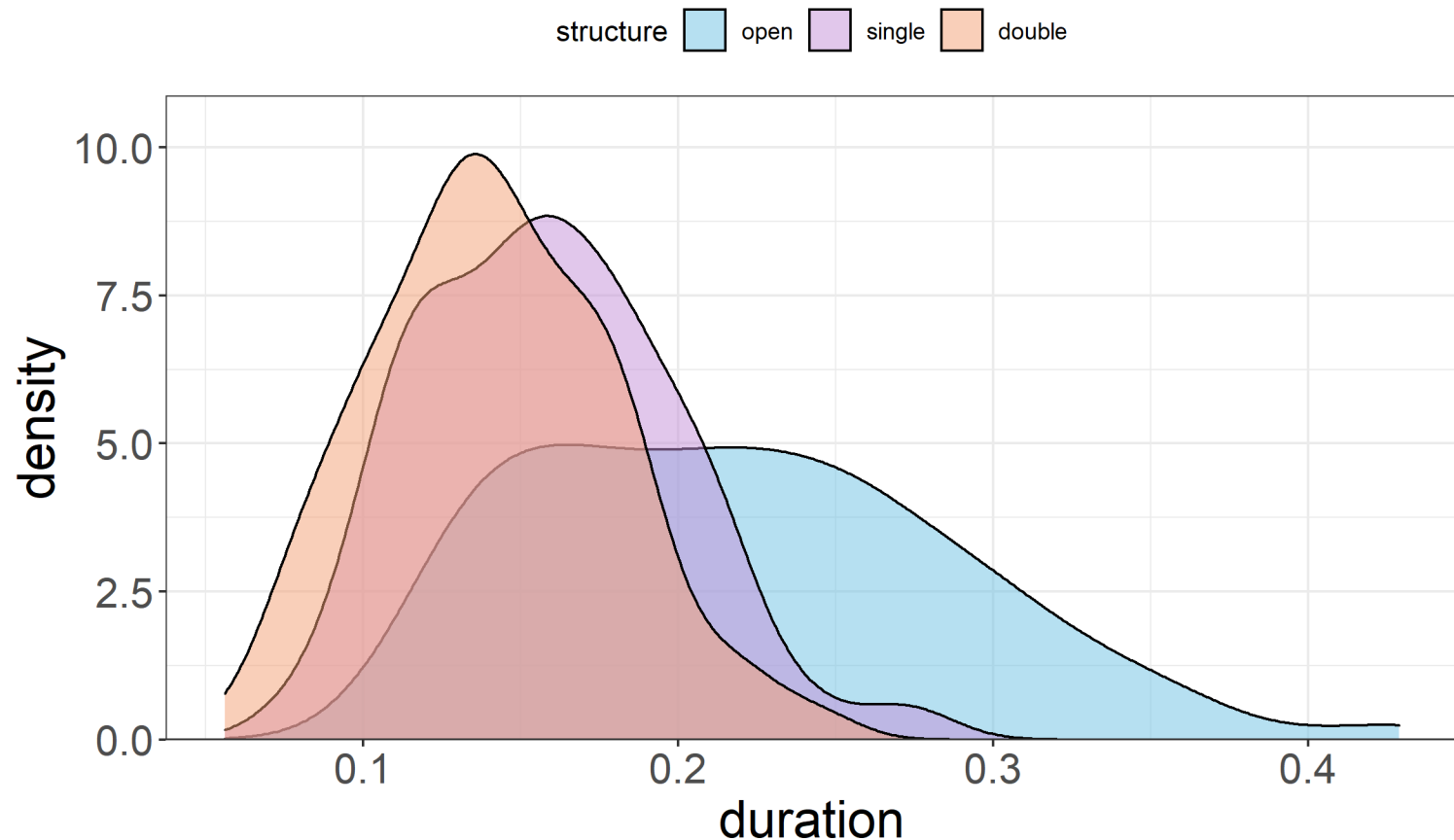
**nützlich zum Check der Verteilung – Lesbarkeit okay**



# Dichte-Diagramm / Density Plot



**nützlich zum Check der Verteilung– Lesbarkeit gut**

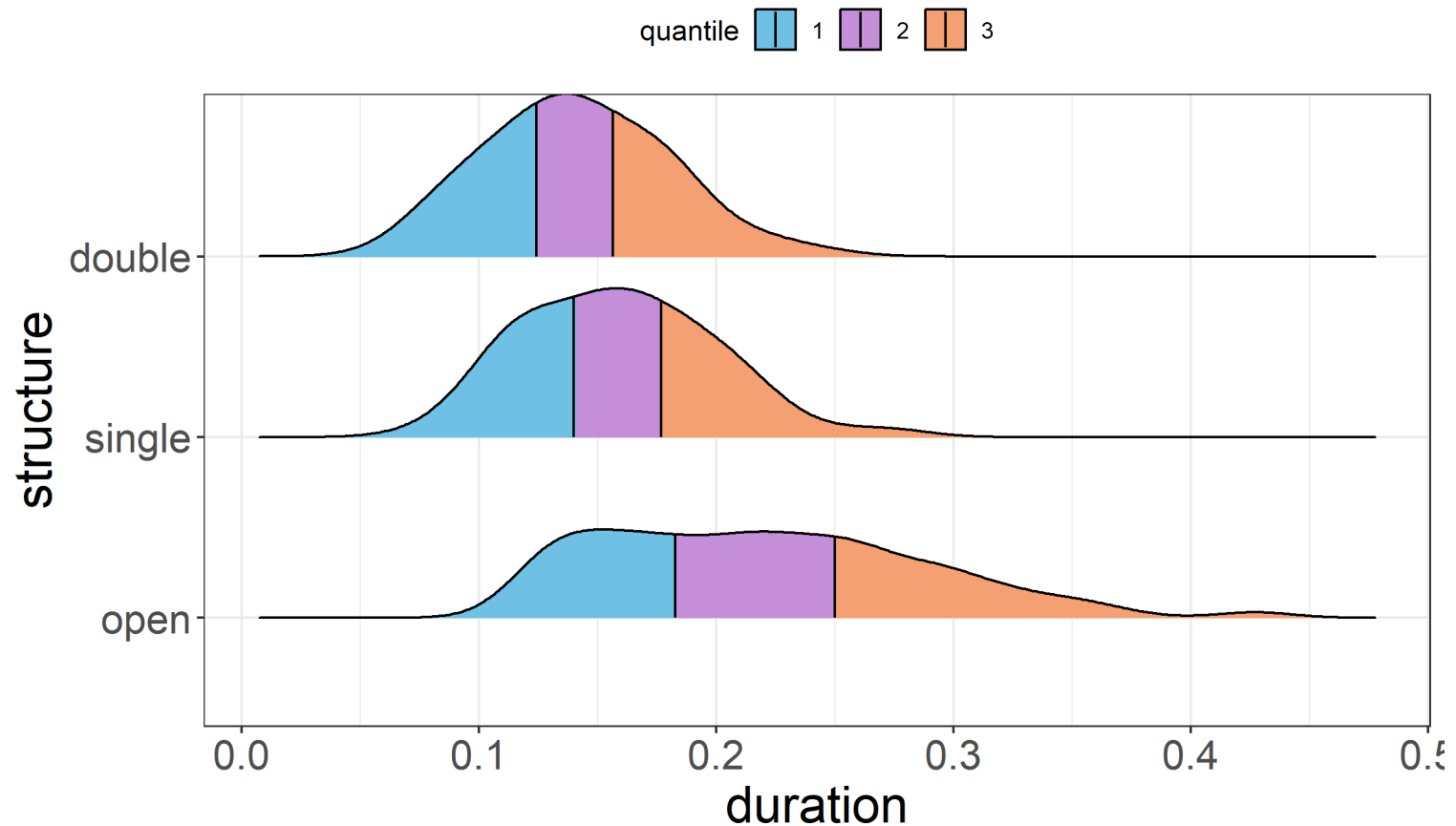




# Dichte-Diagramm / Density Plot



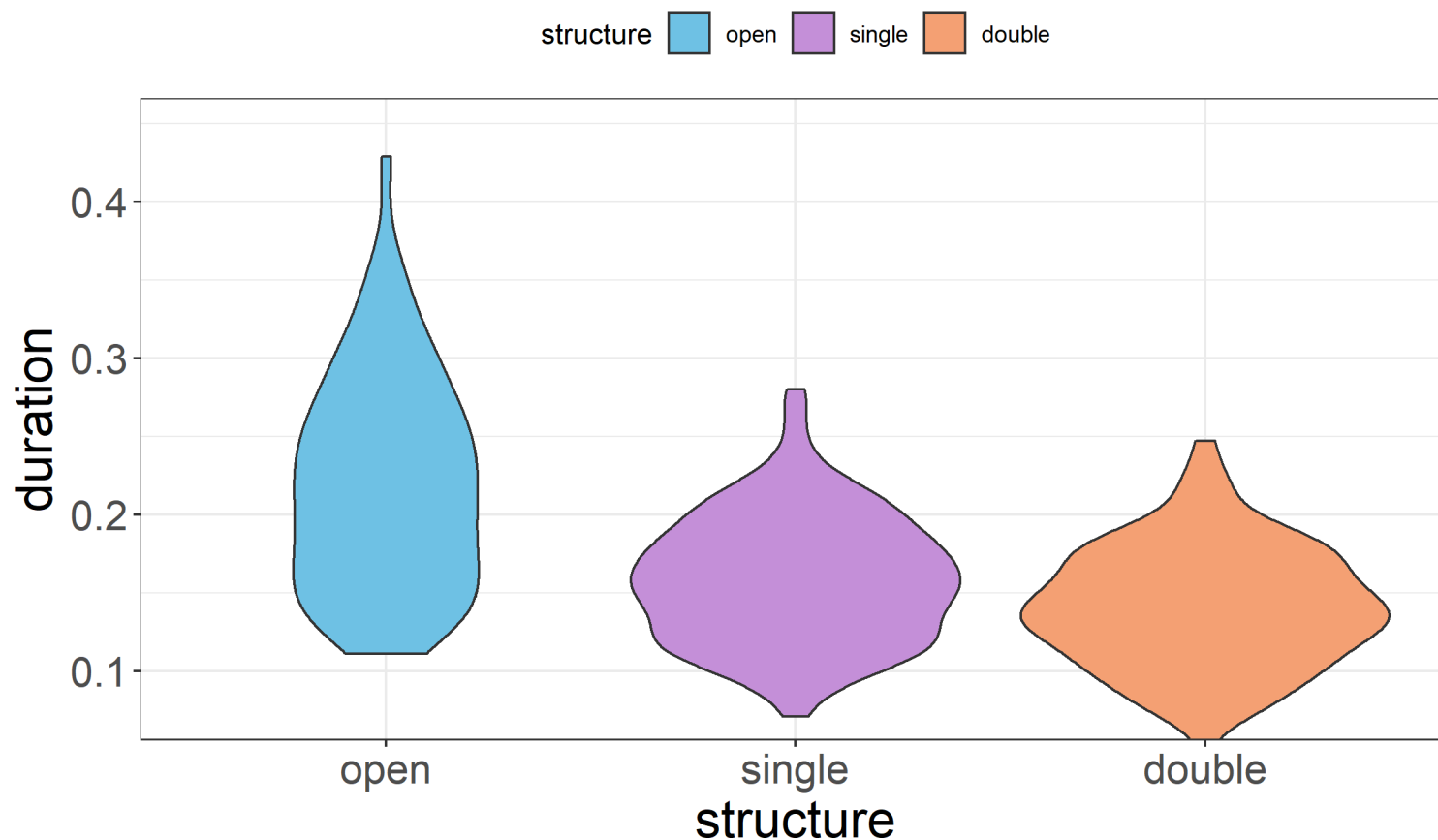
nützlich zum Check der Verteilung



# Violinen-Diagramm / Violin Plot



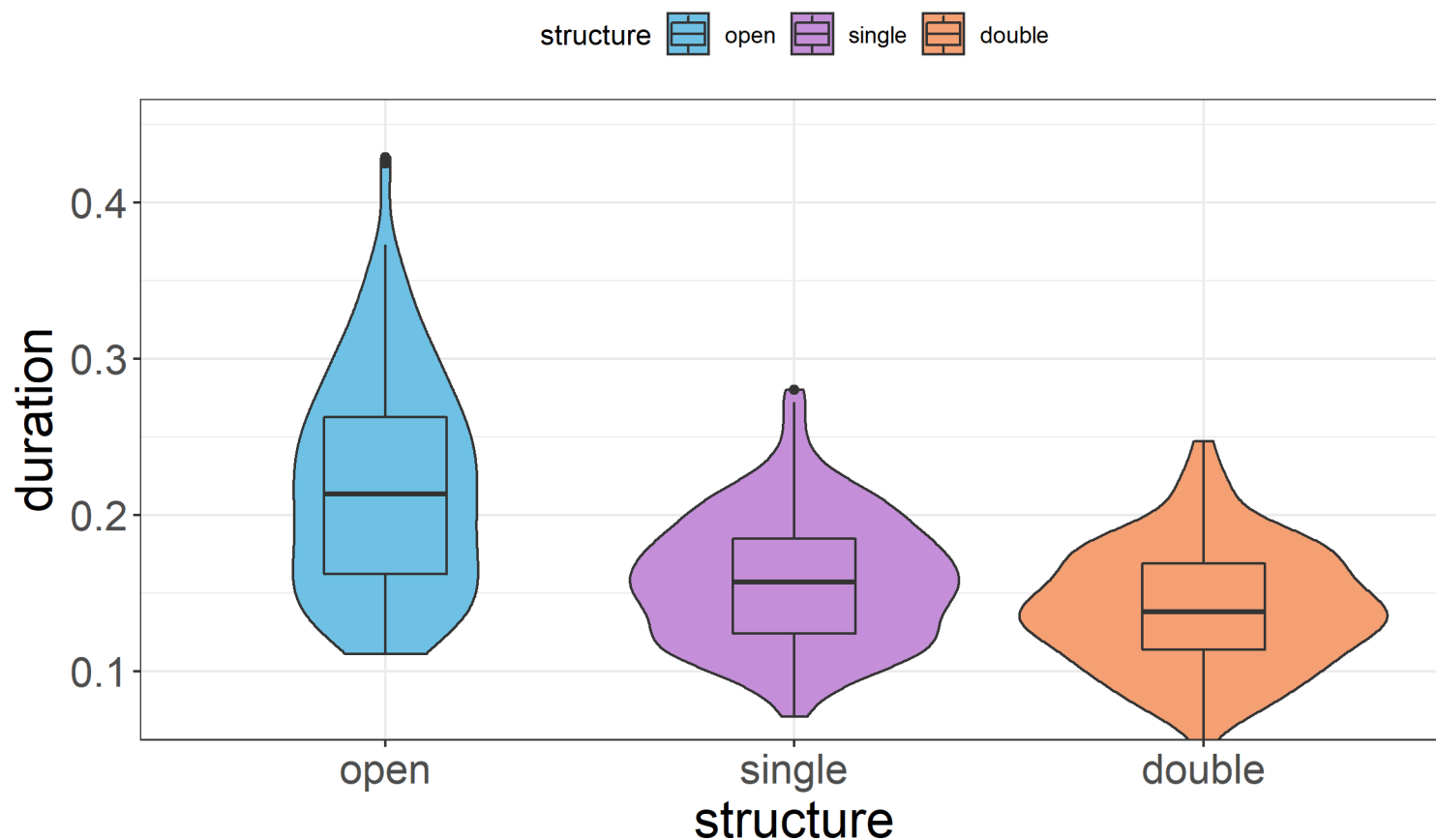
**nützlich zum Check der Verteilung**



# Violinen-Diagramm / Violin Plot



**oft mit Box-Plots kombiniert**



# The Grammar of Graphics

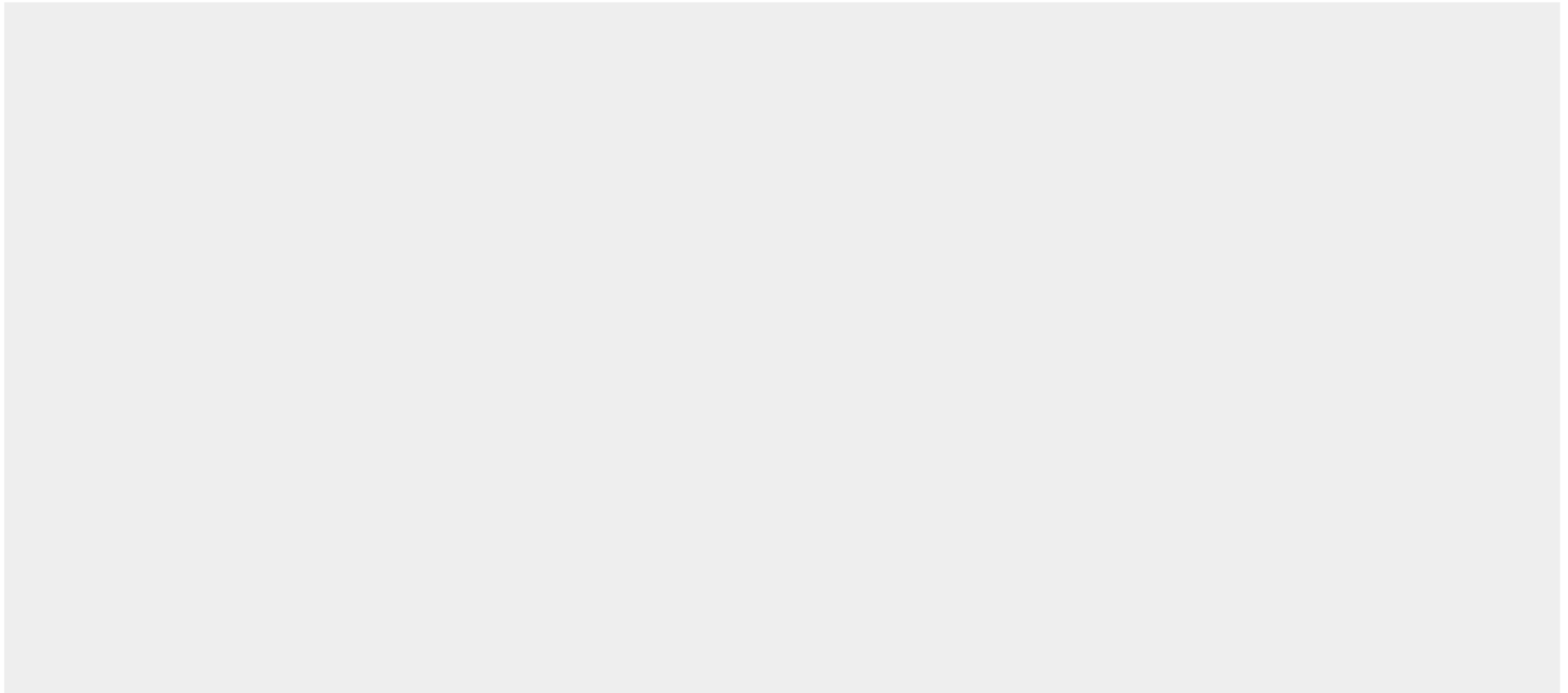


- Die meisten (und besten) Plots in R werden mit dem `ggplot2` Package erstellt
- `ggplot2` folgt den Ideen aus *The Grammar of Graphics*, einem vielfach zitierten Werk zu Datenvisualisierung von Leland Wilkinson (doi: 10.1007/0-387-28695-0)
- Jeder Plot besteht bei `ggplot2` aus 3 Elementen:
  1. data
  2. aesthetics
  3. geometric object

# ggplot2 – data



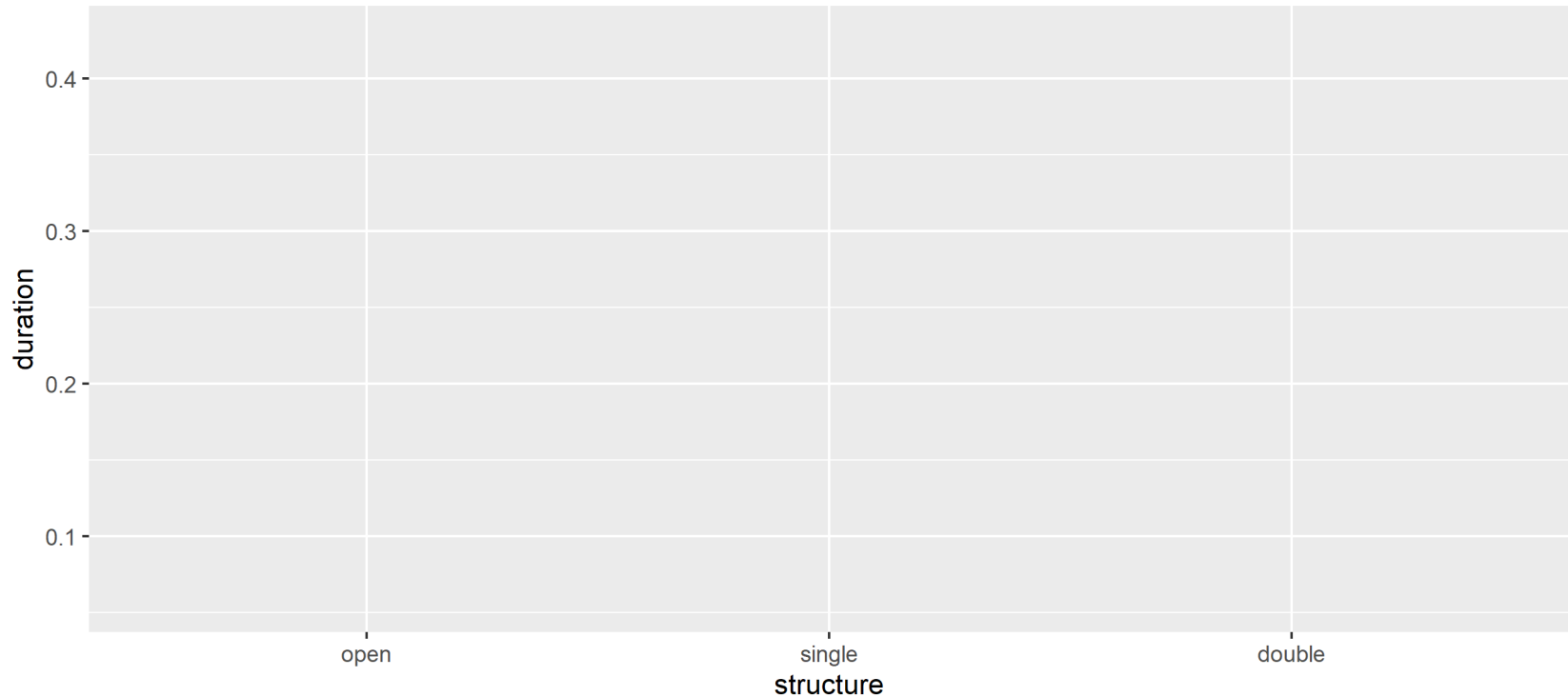
```
> ggplot(data)
```



# ggplot2 – aesthetics



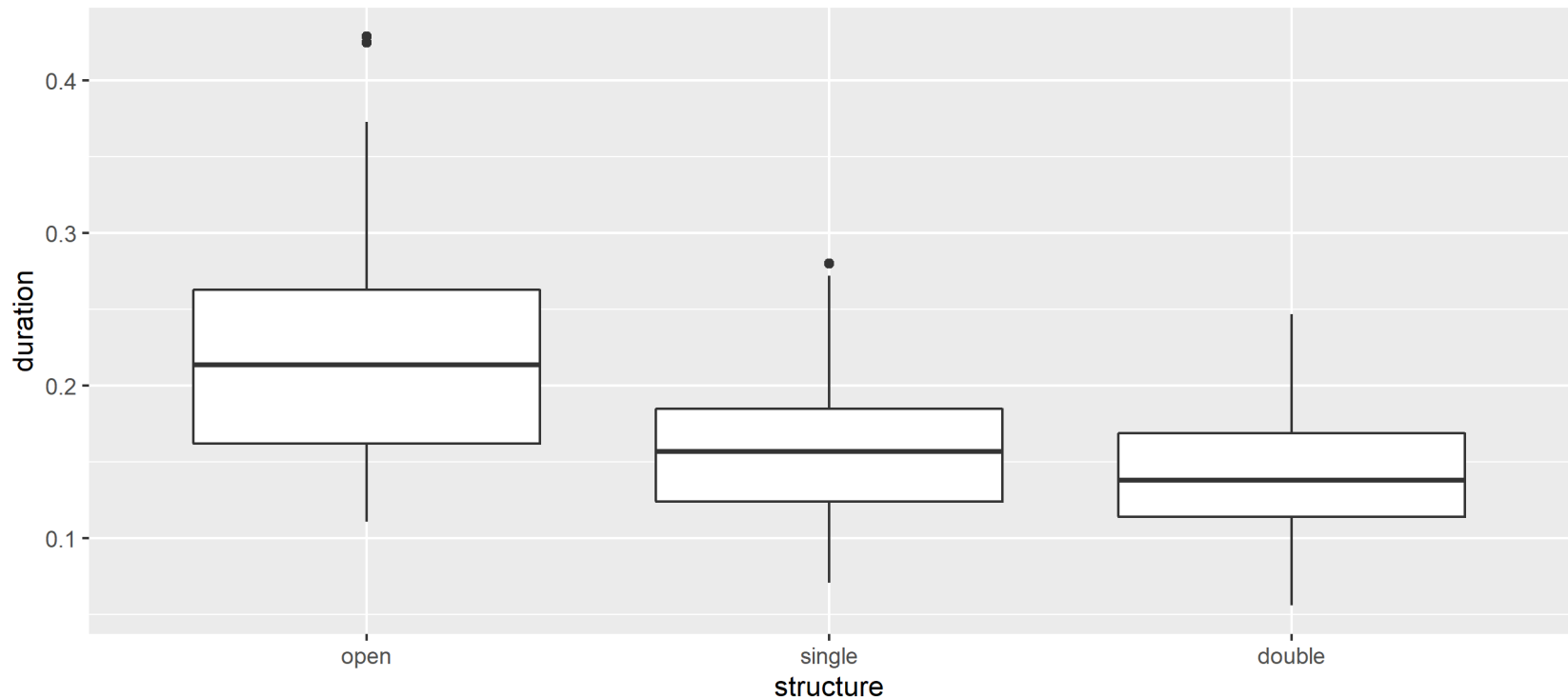
```
> ggplot(data, aes(x=structure, y=duration))
```



# ggplot2 – geometric object



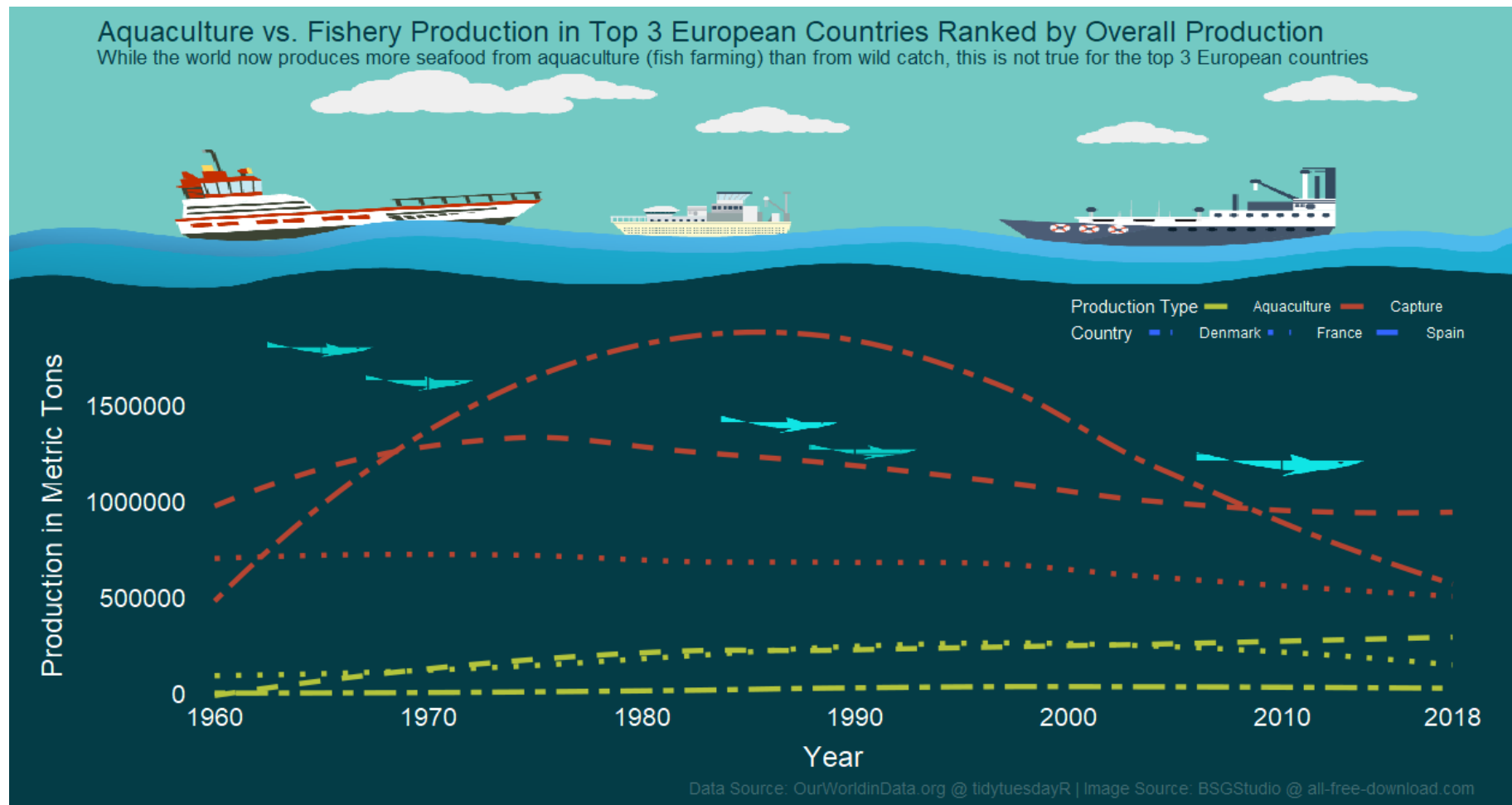
```
> ggplot(data, aes(x=structure, y=duration)) +  
  geom_boxplot()
```



# ggplot2 – Endlose Möglichkeiten



- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.

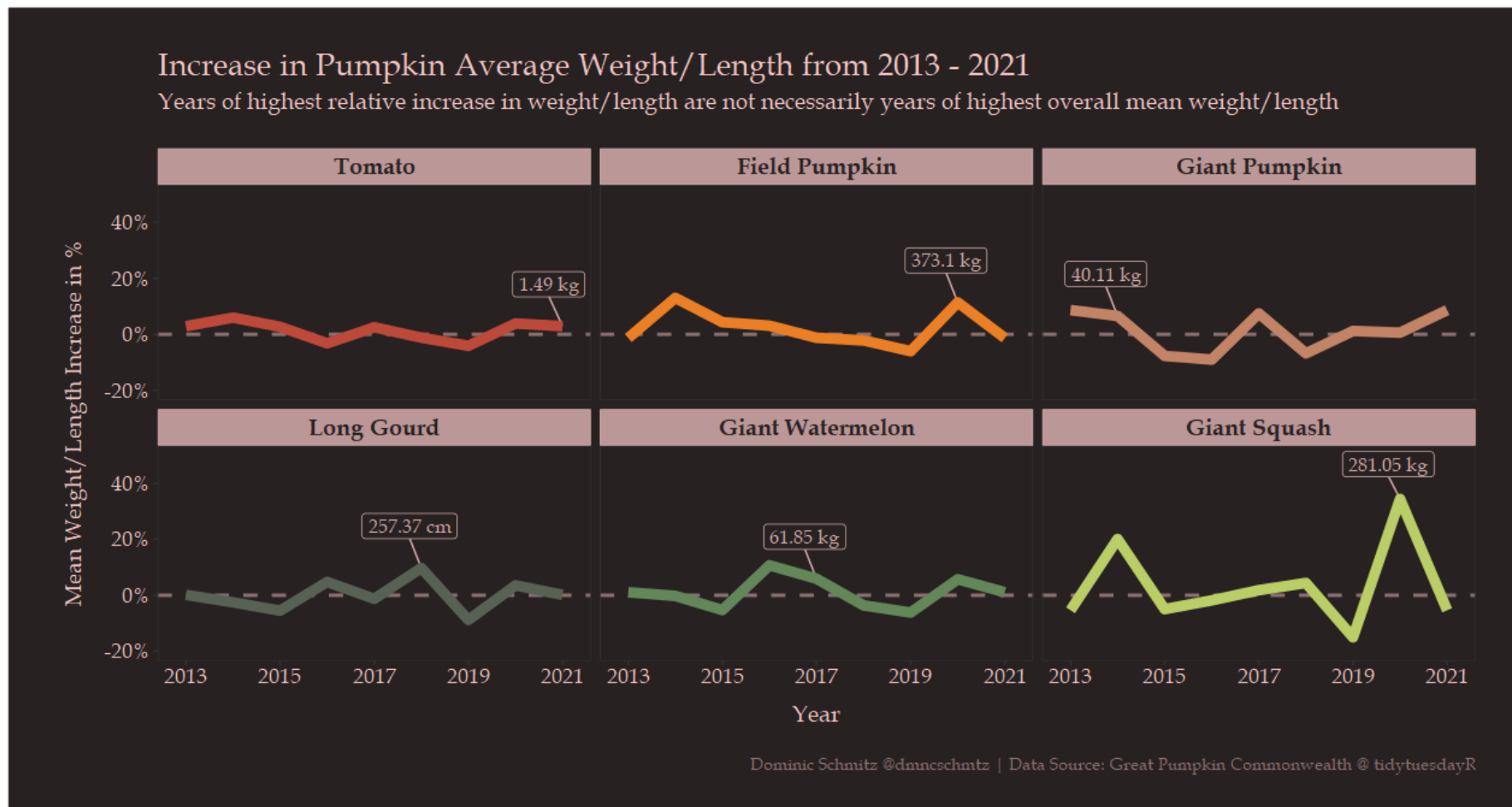




# ggplot2 – Endlose Möglichkeiten



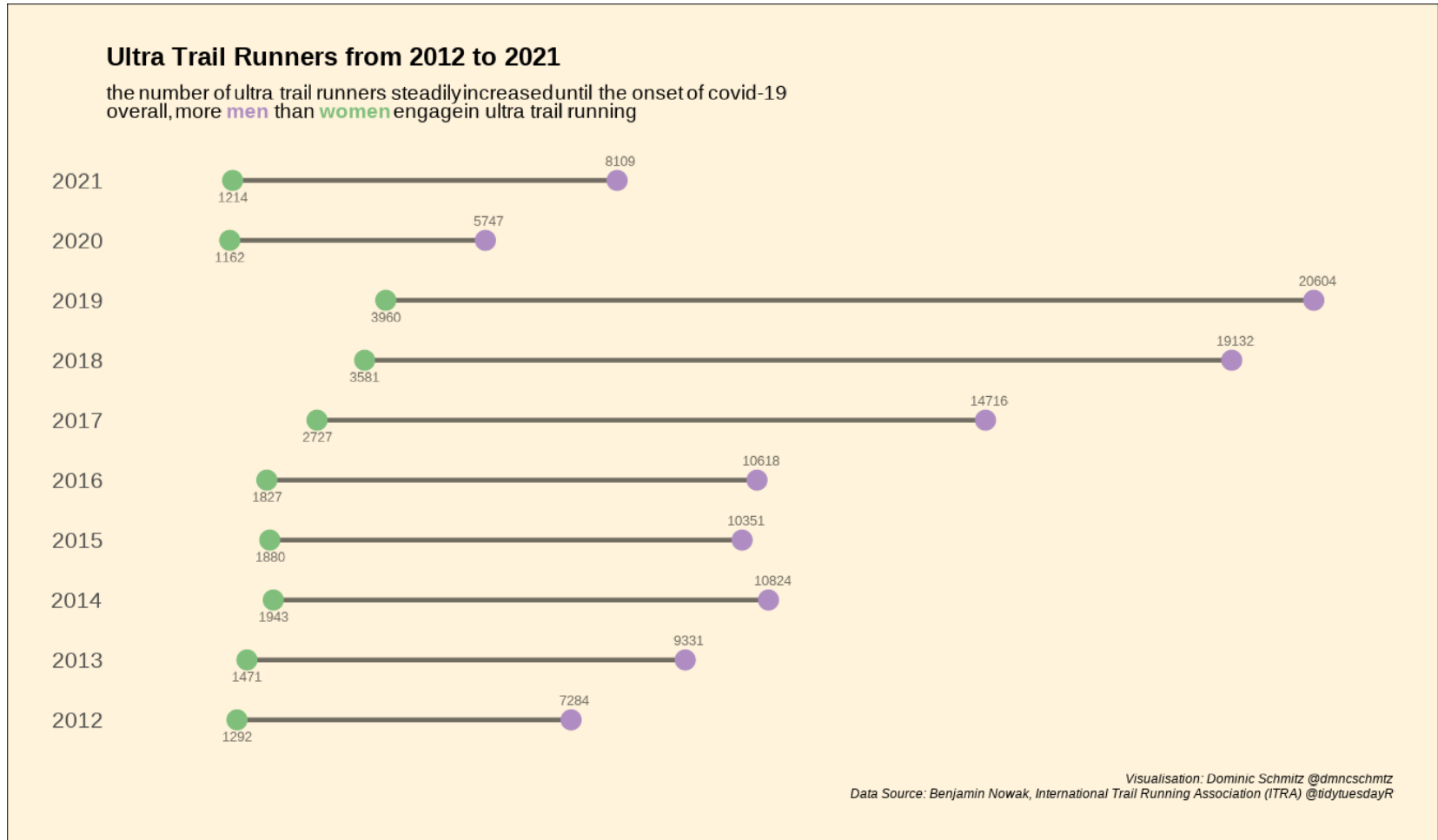
- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.



# ggplot2 – Endlose Möglichkeiten



- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.

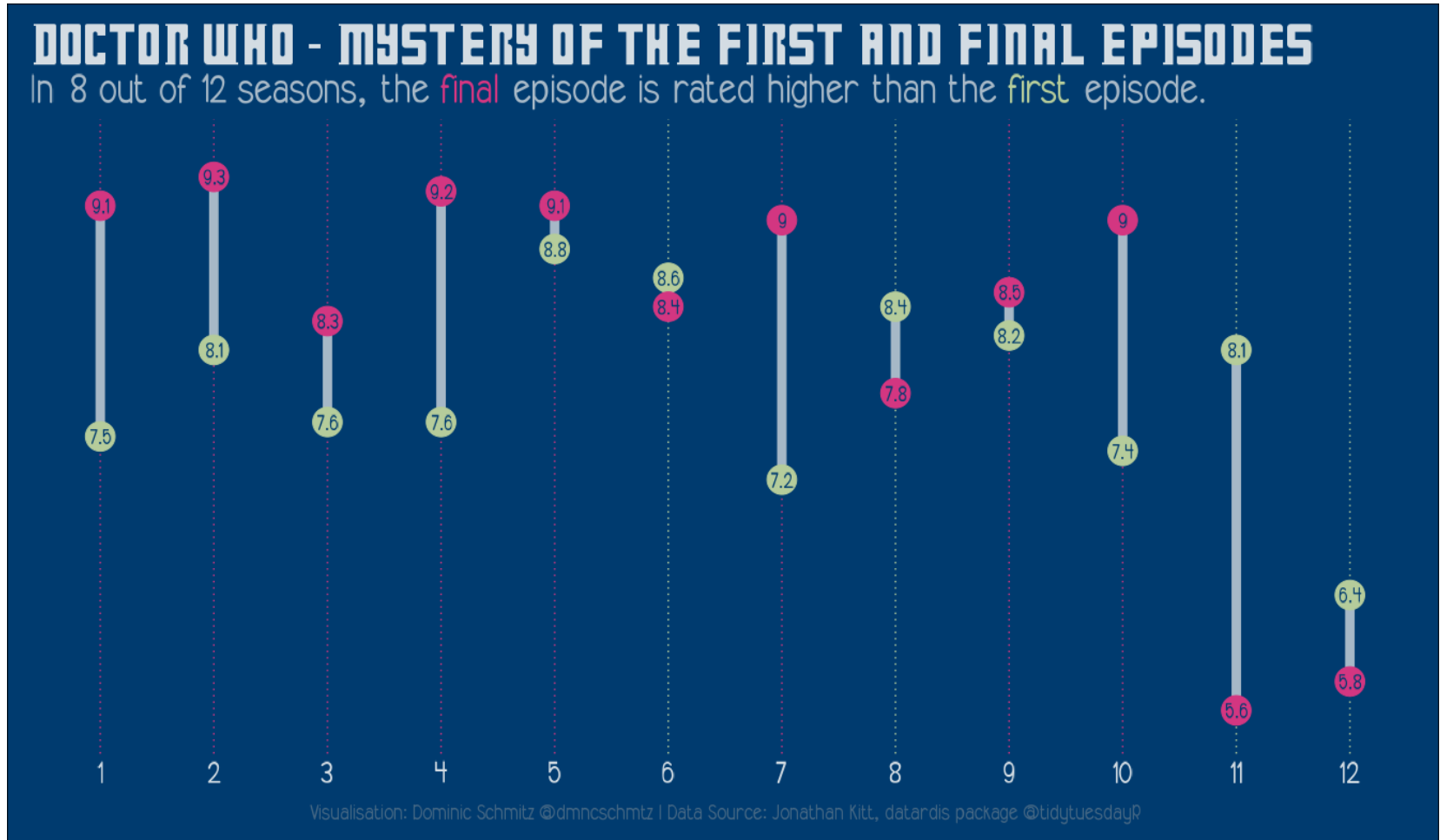


- [illegible]

# ggplot2 – Endlose Möglichkeiten



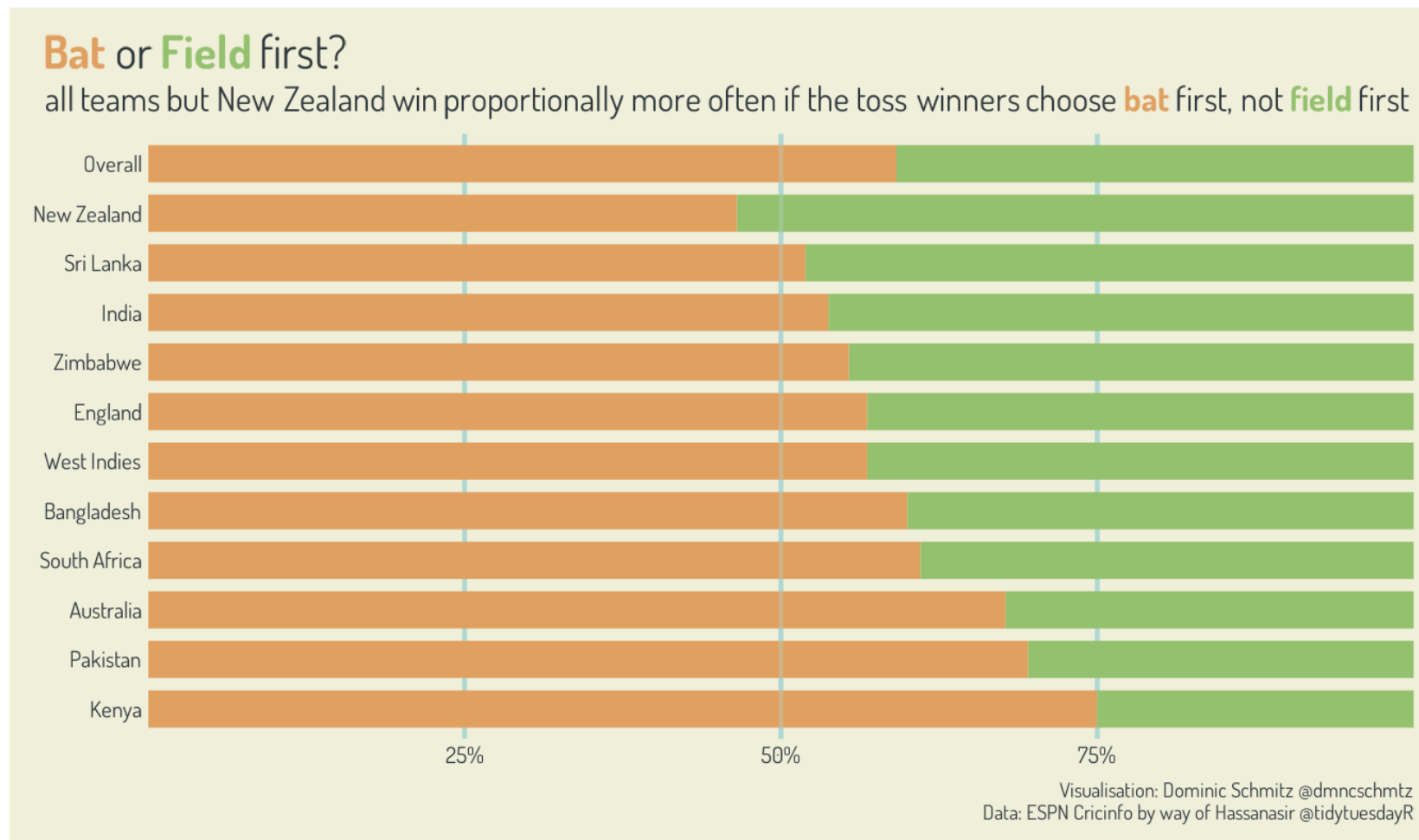
- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.



# ggplot2 – Endlose Möglichkeiten



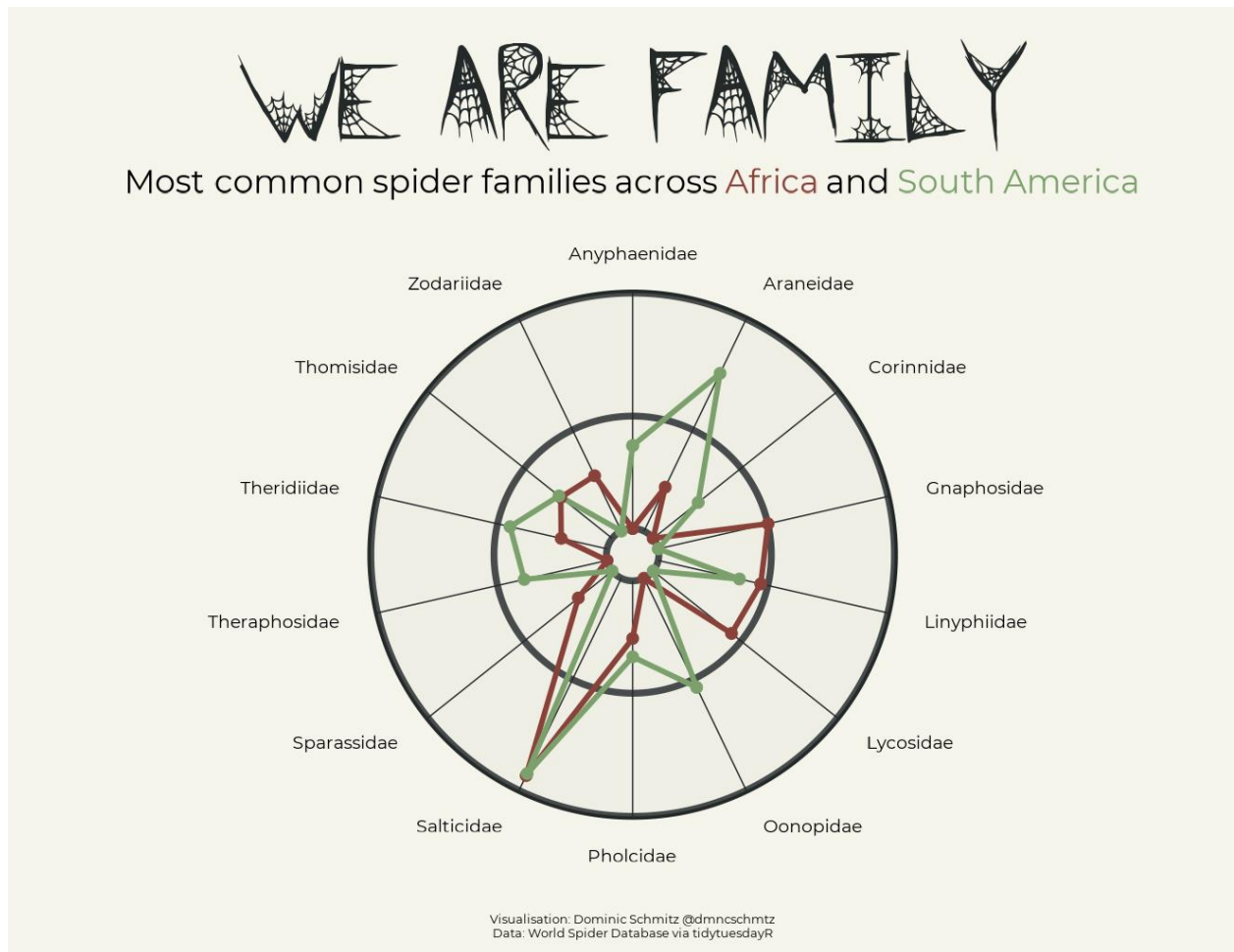
- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.



# ggplot2 – Endlose Möglichkeiten



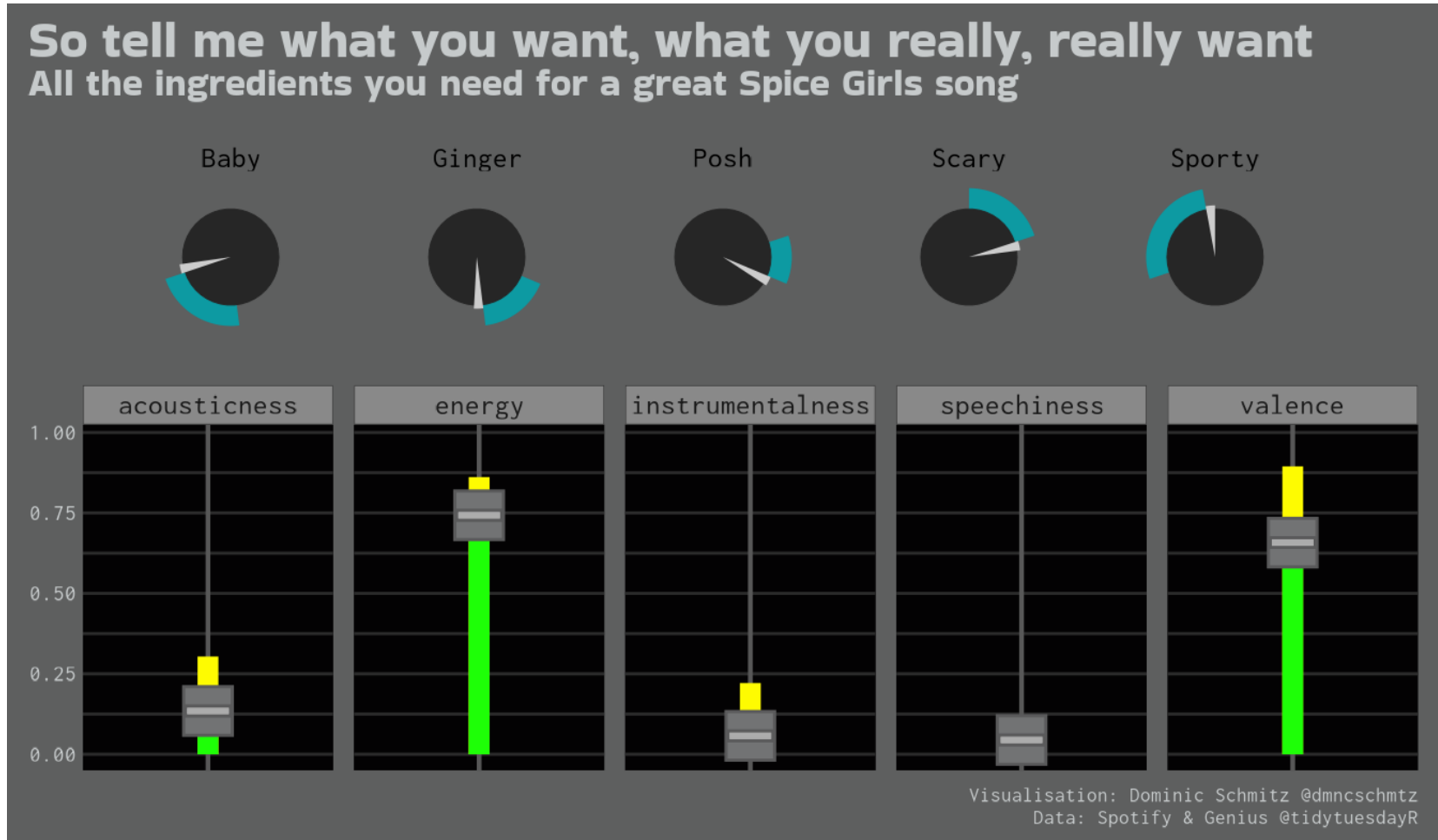
- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.



# ggplot2 – Endlose Möglichkeiten



- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.



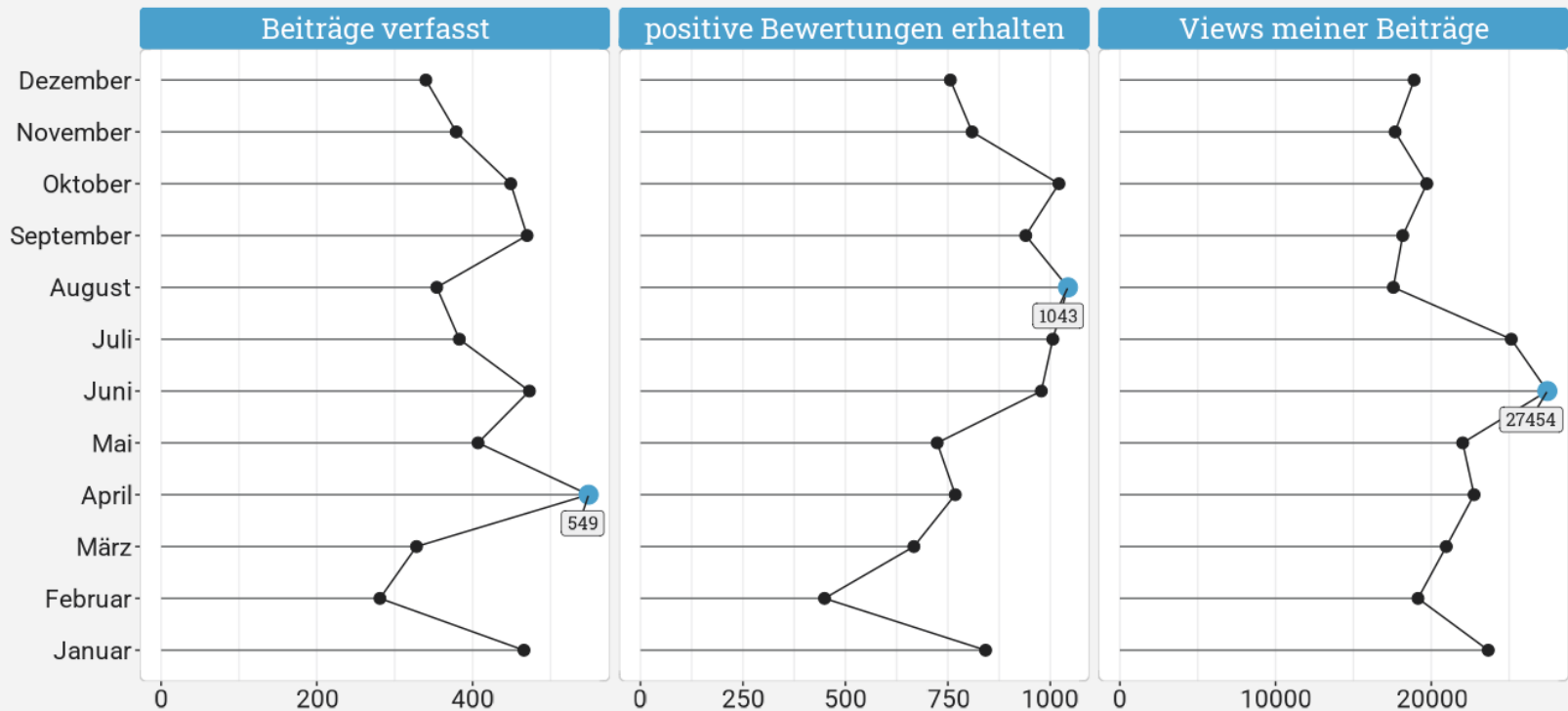
# ggplot2 – Endlose Möglichkeiten



- Hat man ggplot2 gemeistert, hat man endlos viele Möglichkeiten, z.B.

## Mein Jahr 2021 auf gutefrage.net

Daten aus meinem monatlichen Report als »Community-Experte« für Sprache, Gendern, Gender, LGBT & LGBTQ



Visualisation: Dominic Schmitz @dmnnschmtz  
Data Source: gutefrage.net @gutefrage\_net