

# 05

## Datenvisualisierung mit ggplot2

Dominic Schmitz & Janina Esser

# Datenvisualisierung

- Datenvisualisierung kennt zwei Hauptmotivationen

## 1. Darstellung

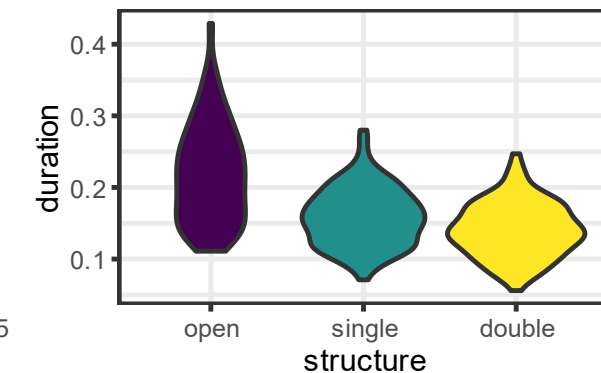
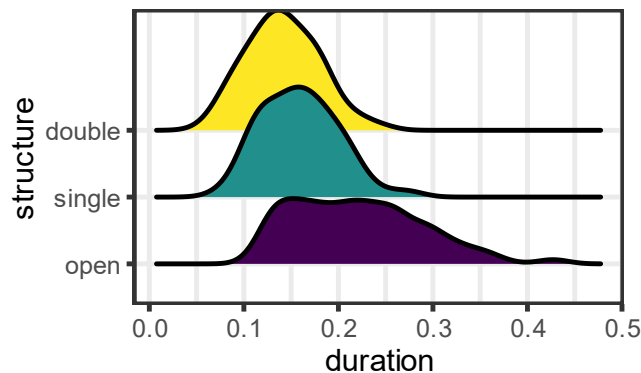
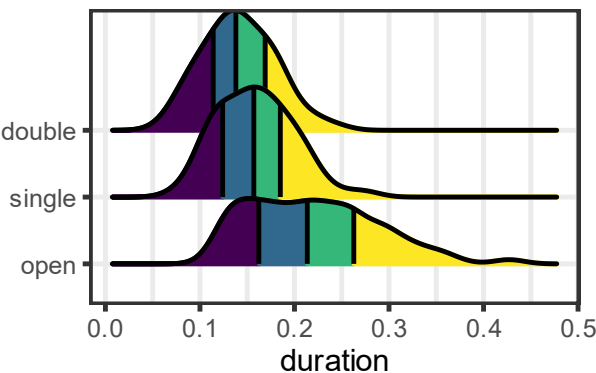
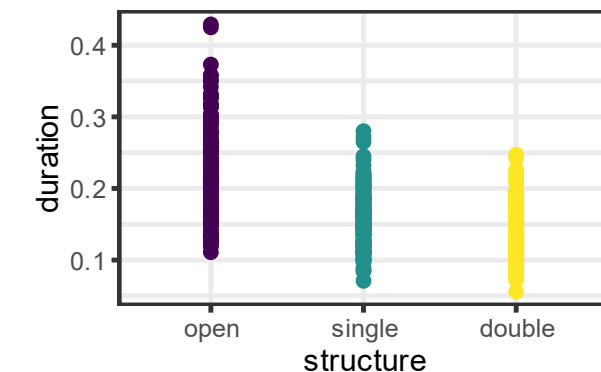
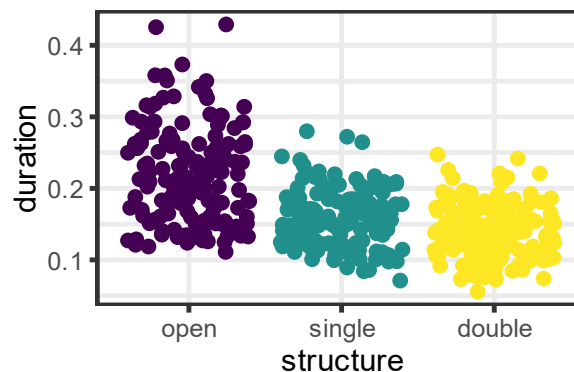
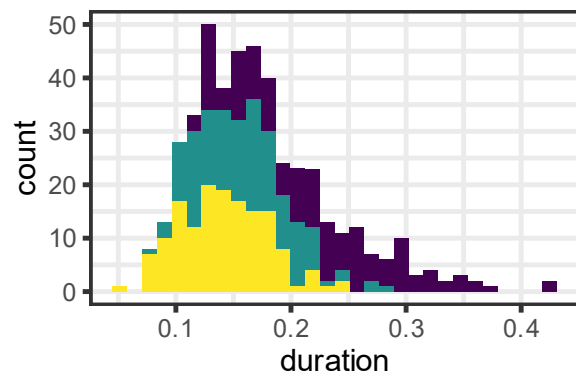
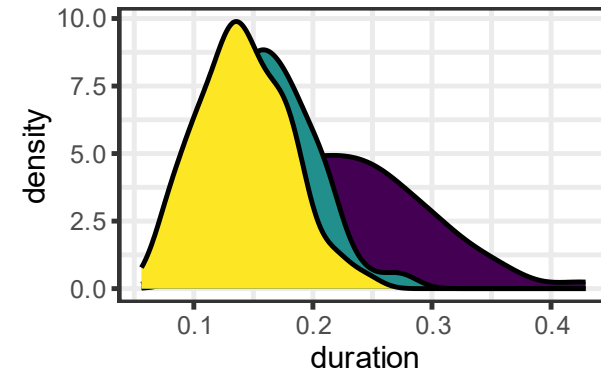
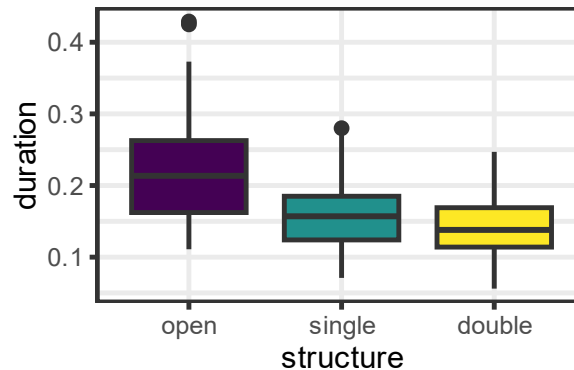
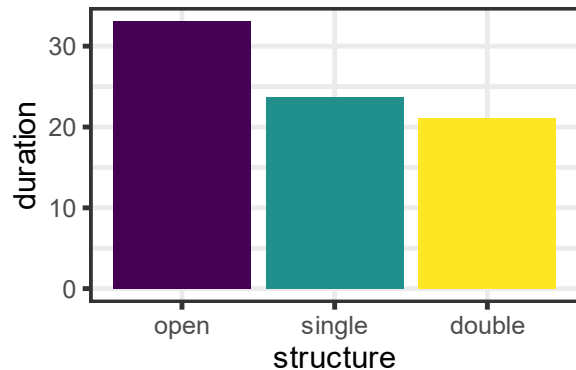
Daten sollen für eine bestimmte Zielgruppe aufbereitet dargestellt werden, um eine bestimmte Botschaft mitzuteilen

## 2. Analyse

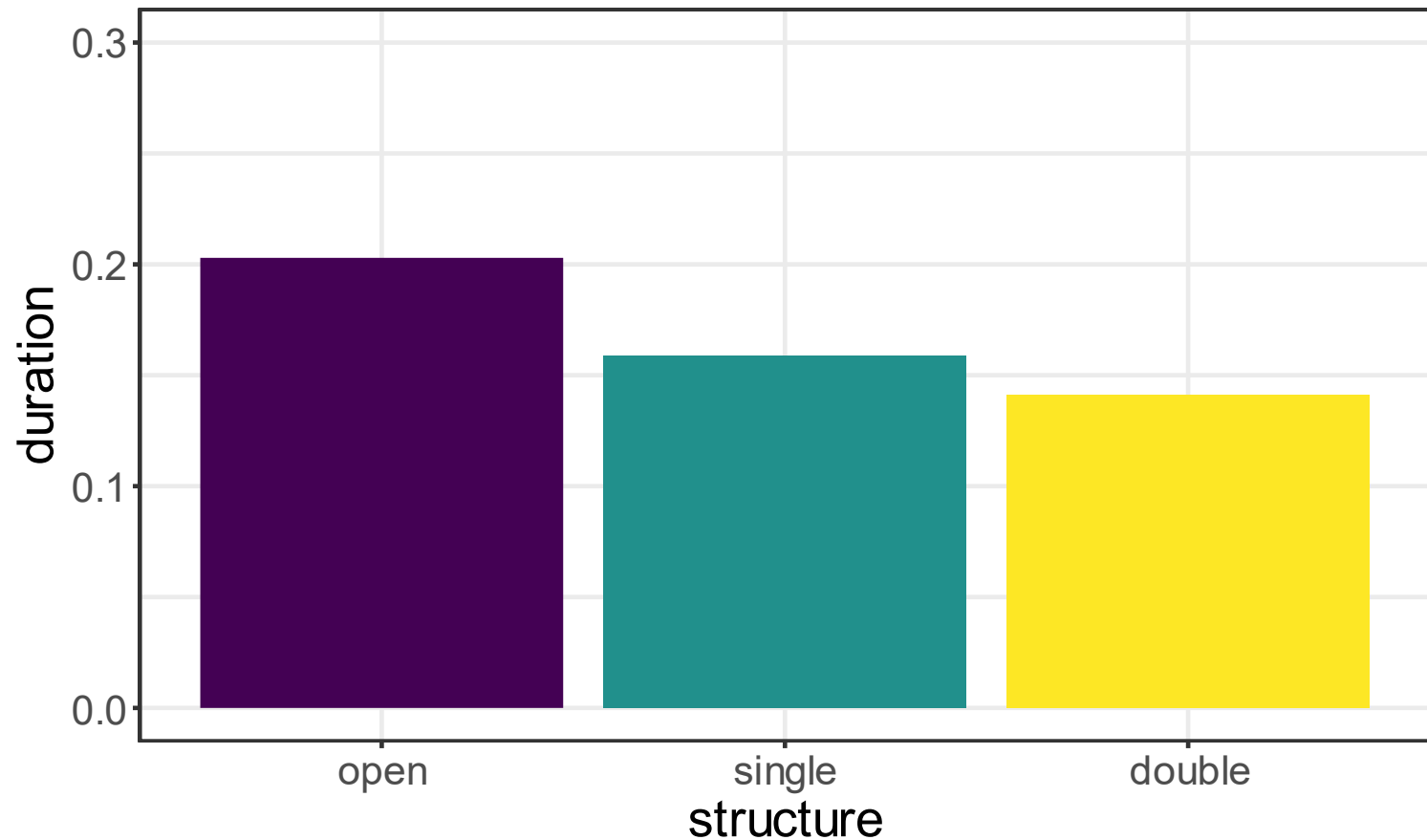
Zur Exploration der Daten können visuelle Darstellungen manchmal hilfreicher sein als bloße Zahlen

- Für alle Fälle gibt es **nahezu endlos viele Möglichkeiten** zur Darstellung; die Wahl des richtigen Formats ist oftmals entscheidend

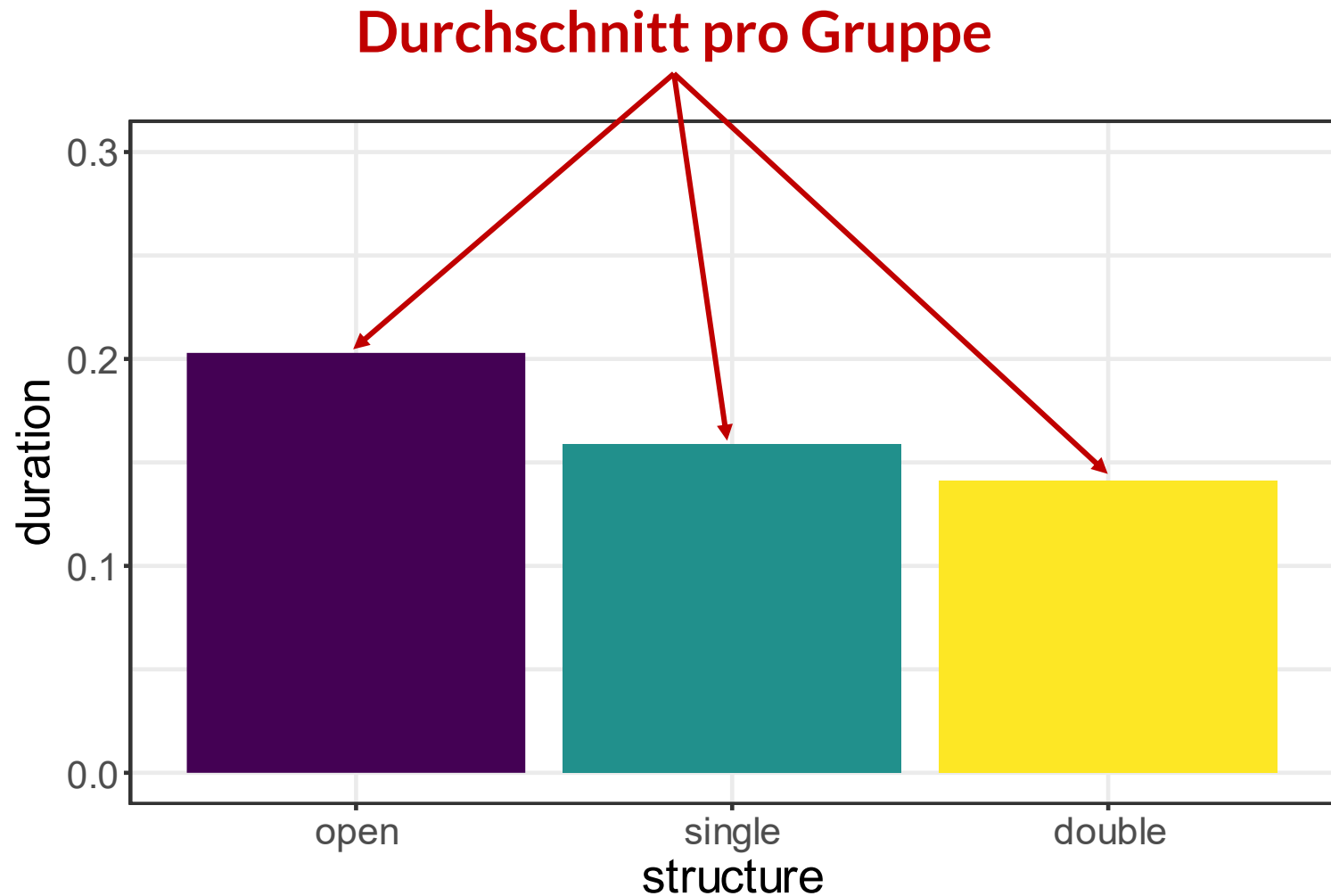
# 2 Variablen, $\infty$ Plots



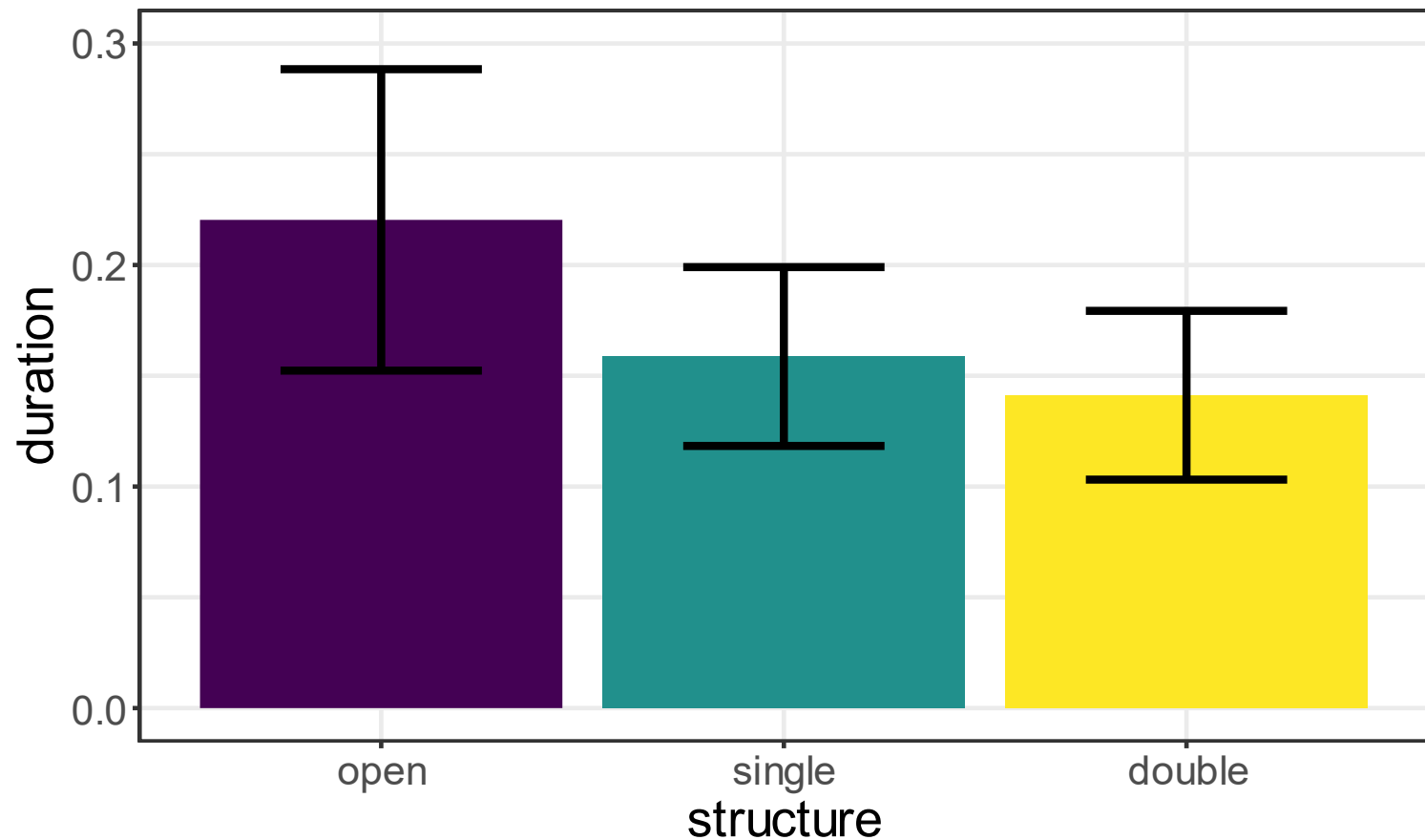
# Bar Plot / Bar Chart / Bar Graph



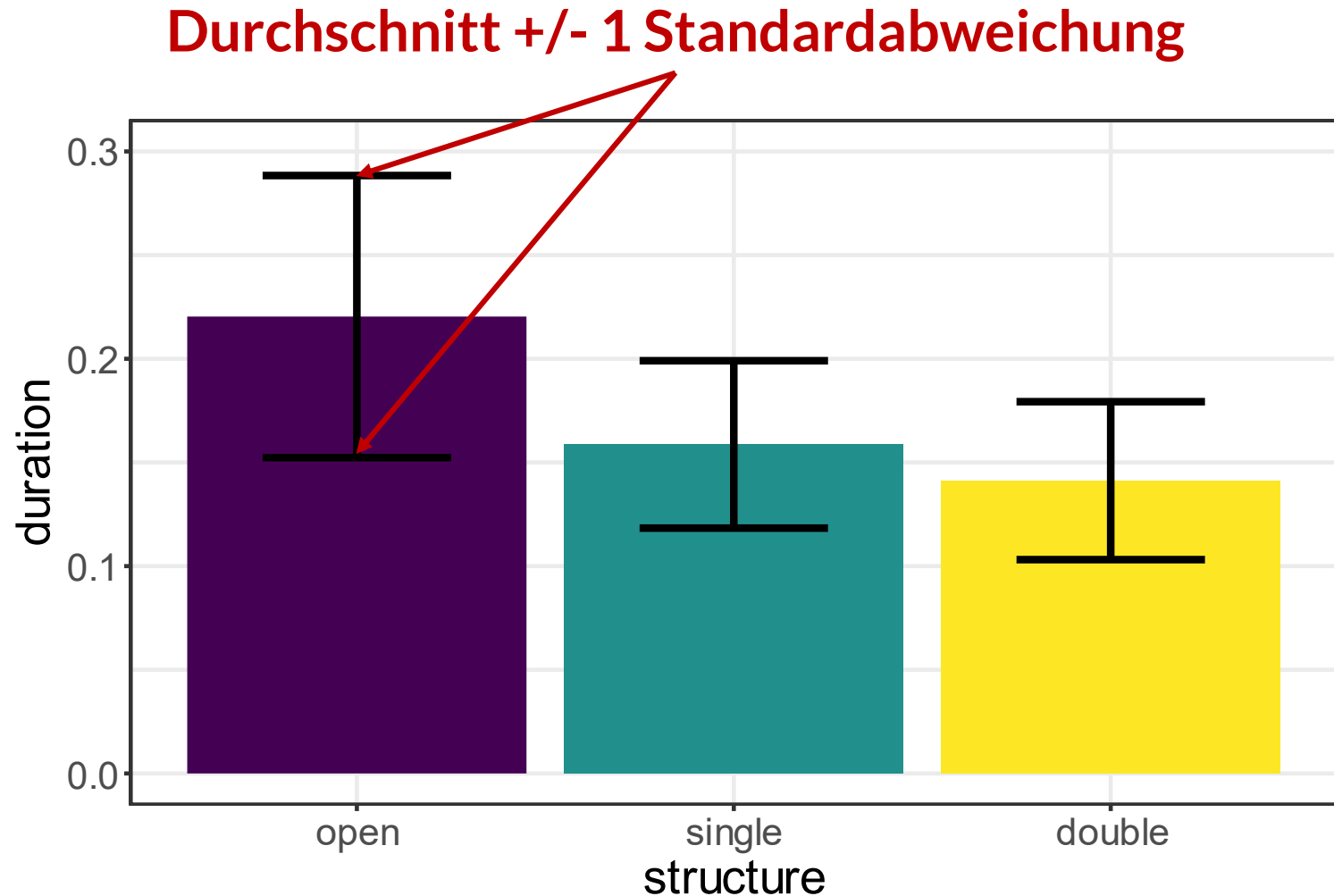
# Bar Plot / Bar Chart / Bar Graph



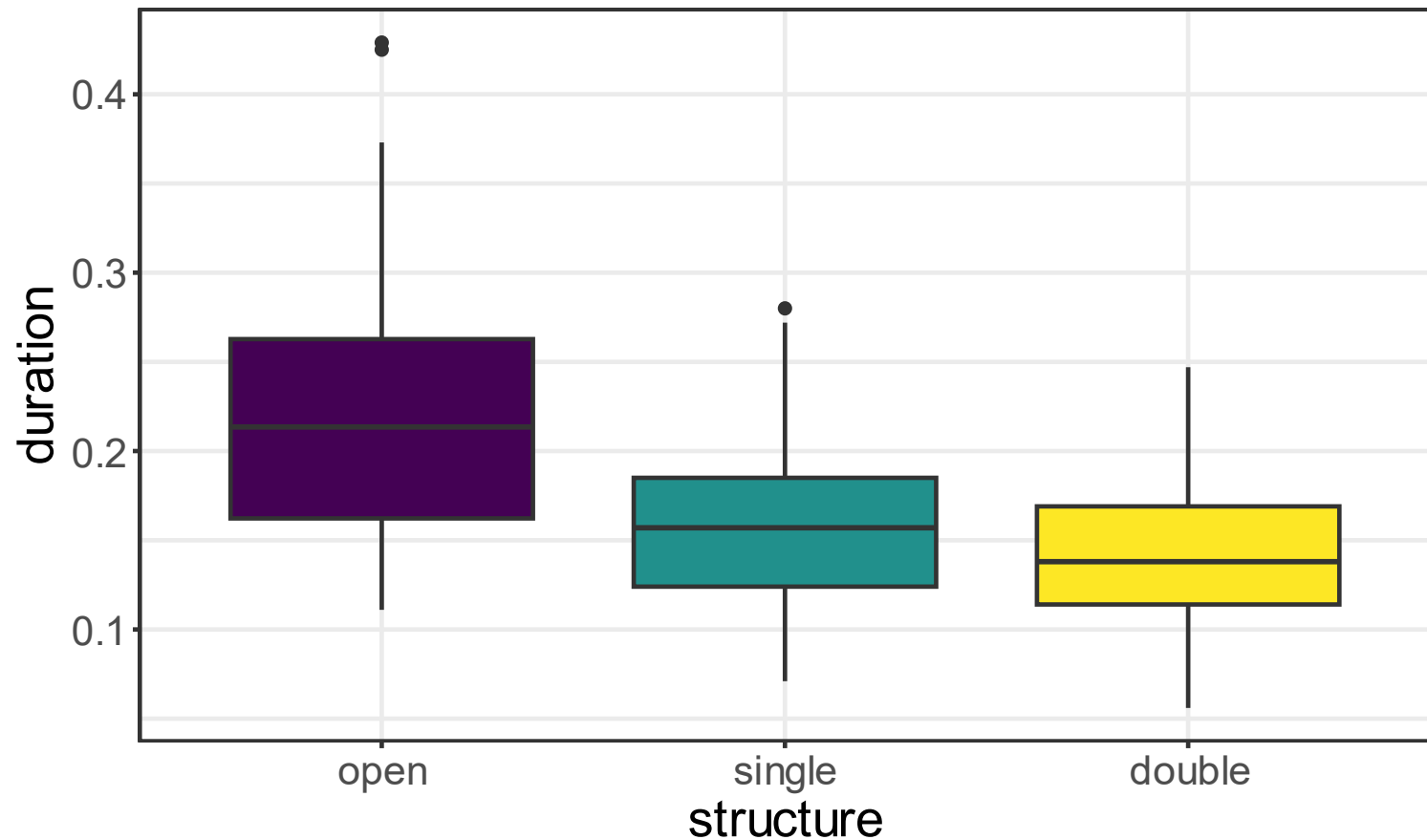
# Bar Plot / Bar Chart / Bar Graph



# Bar Plot / Bar Chart / Bar Graph

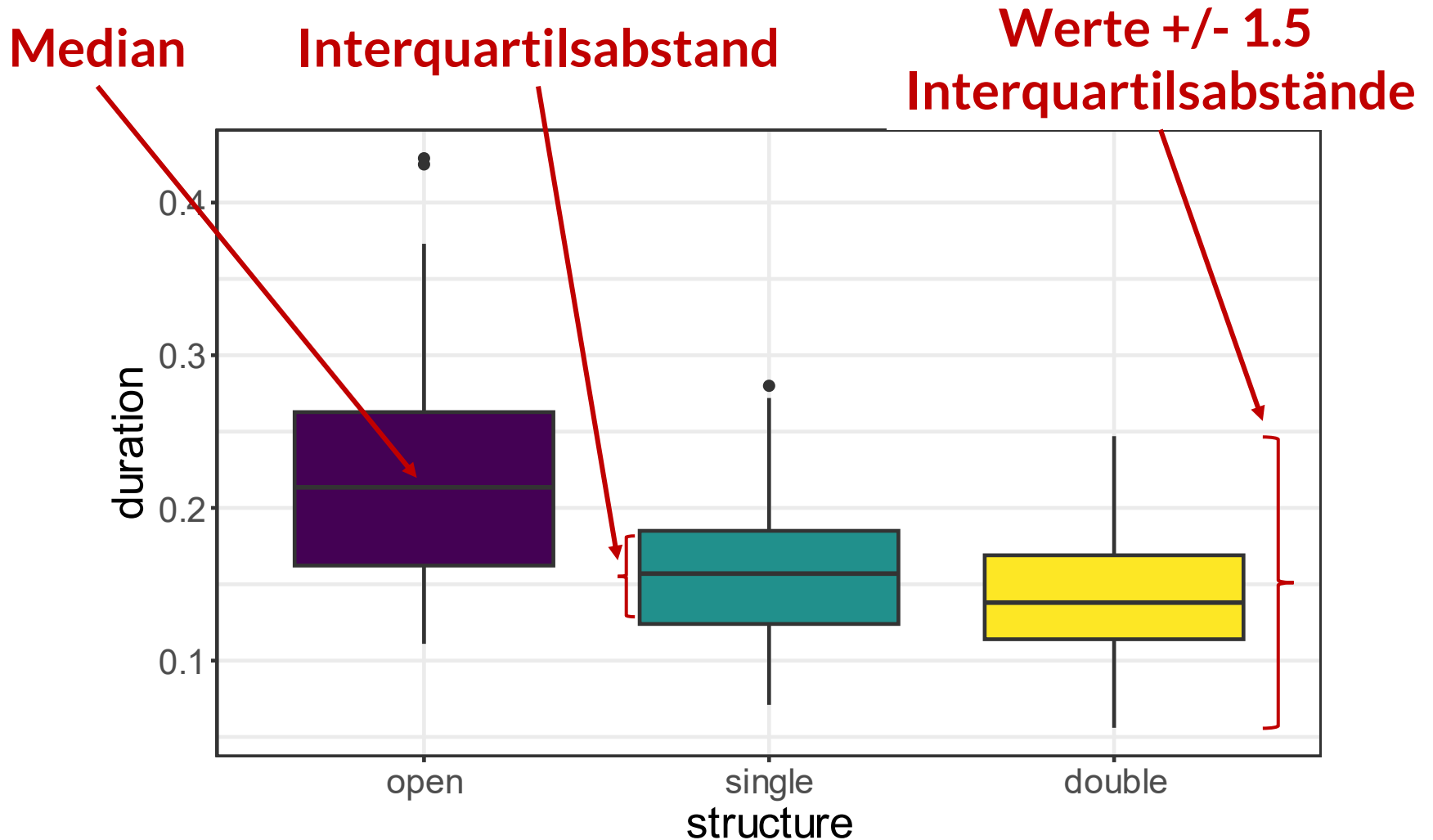


# Box Plot / Box-and-Whisker Diagram

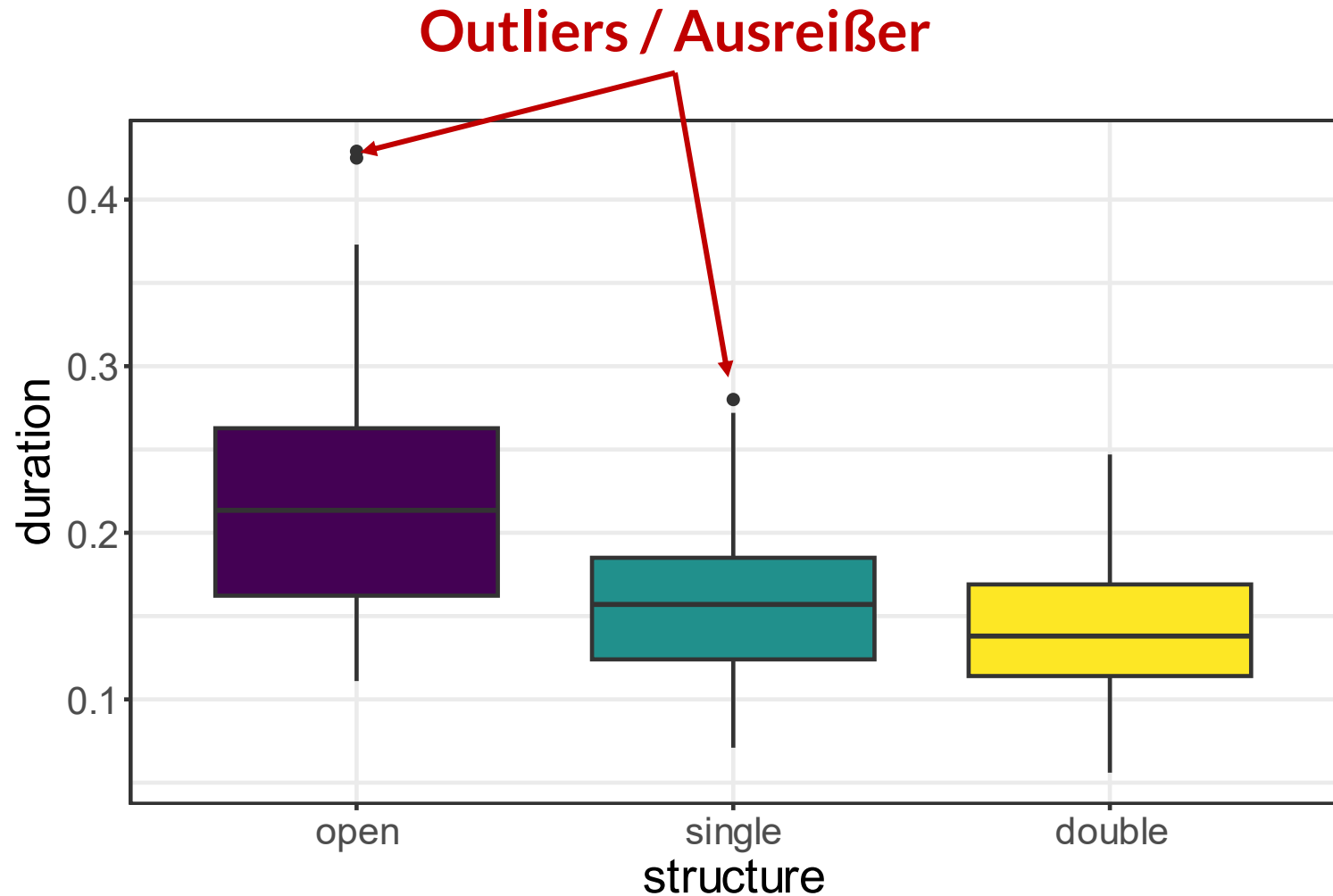




# Box Plot / Box-and-Whisker Diagram

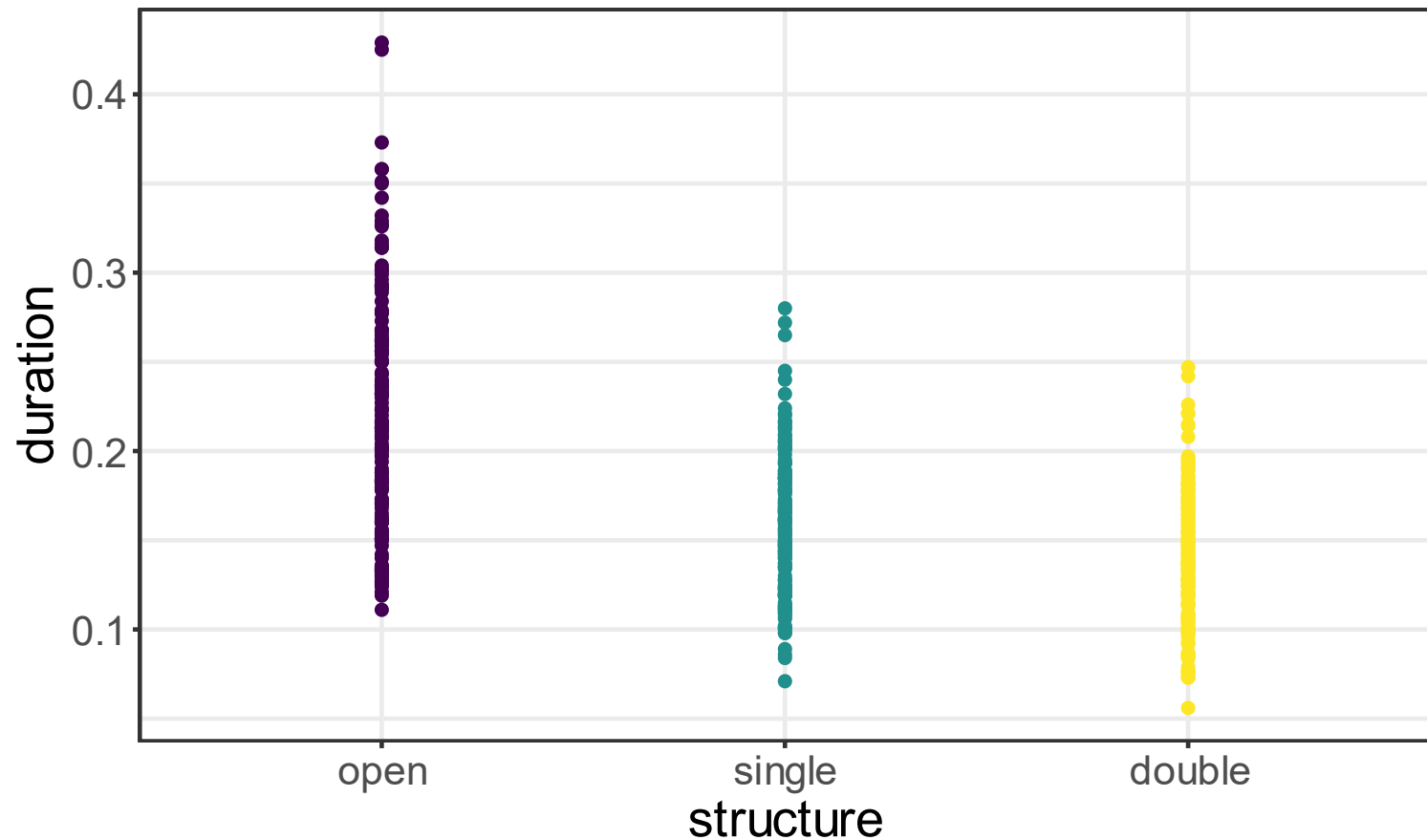


# Box Plot / Box-and-Whisker Diagram



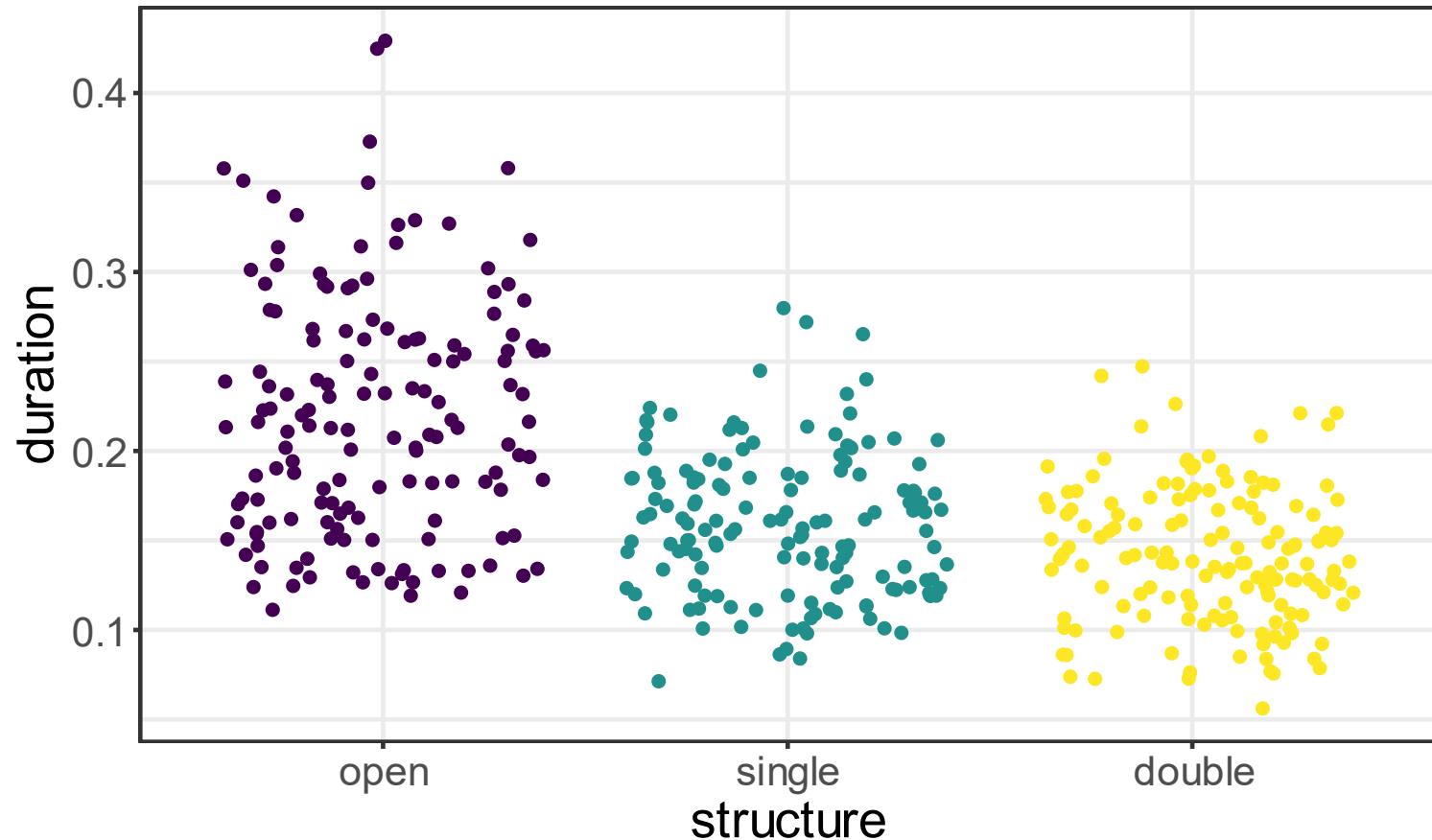
# Point Plot / Dot Plot / Dot Chart

**1 Punkt = 1 Datenpunkt**



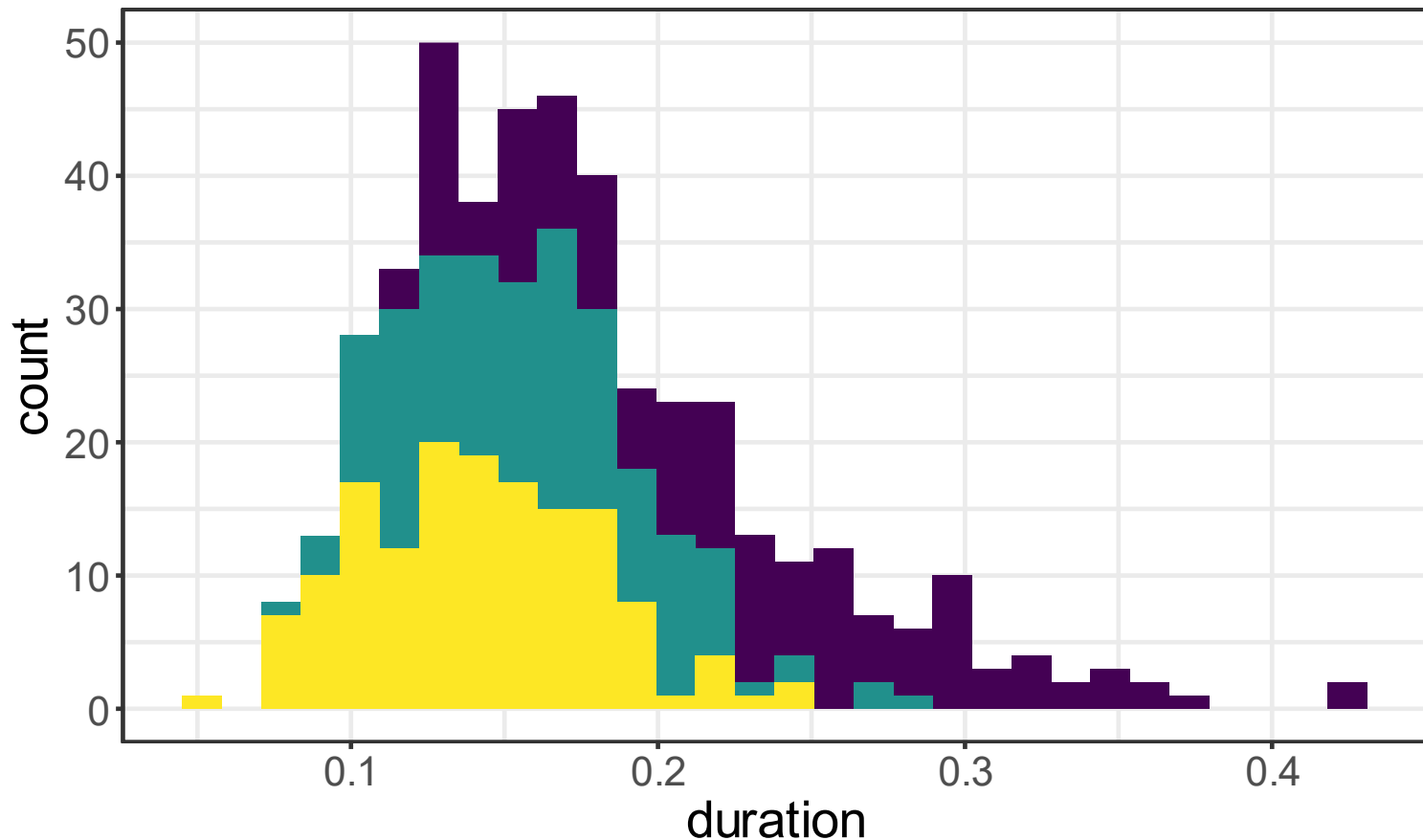
# Jitter Plot / Scatter Plot / Scattergram

**1 Punkt = 1 Datenpunkt**



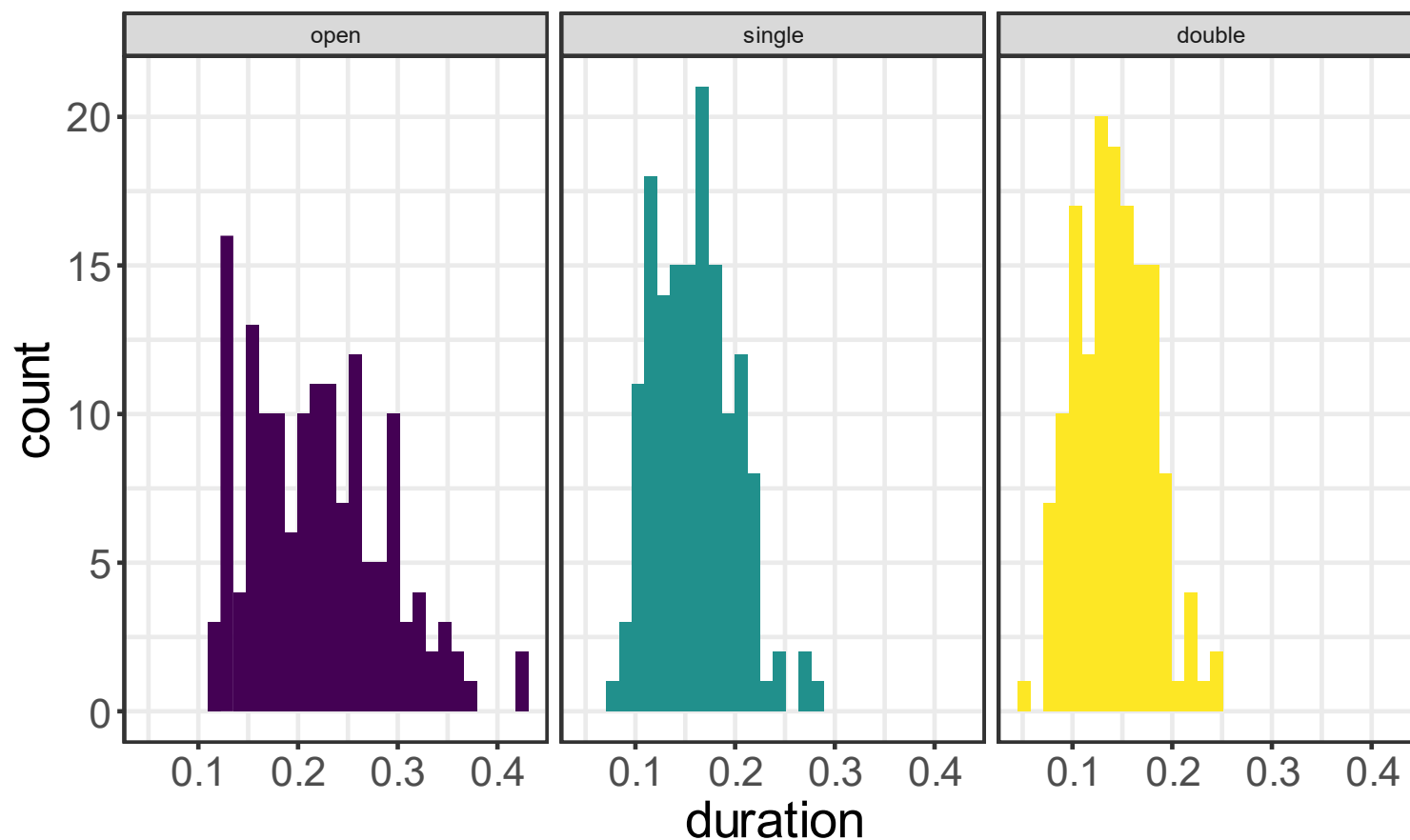
# Histogram

nützlich zum Check der Distribution – Lesbarkeit?



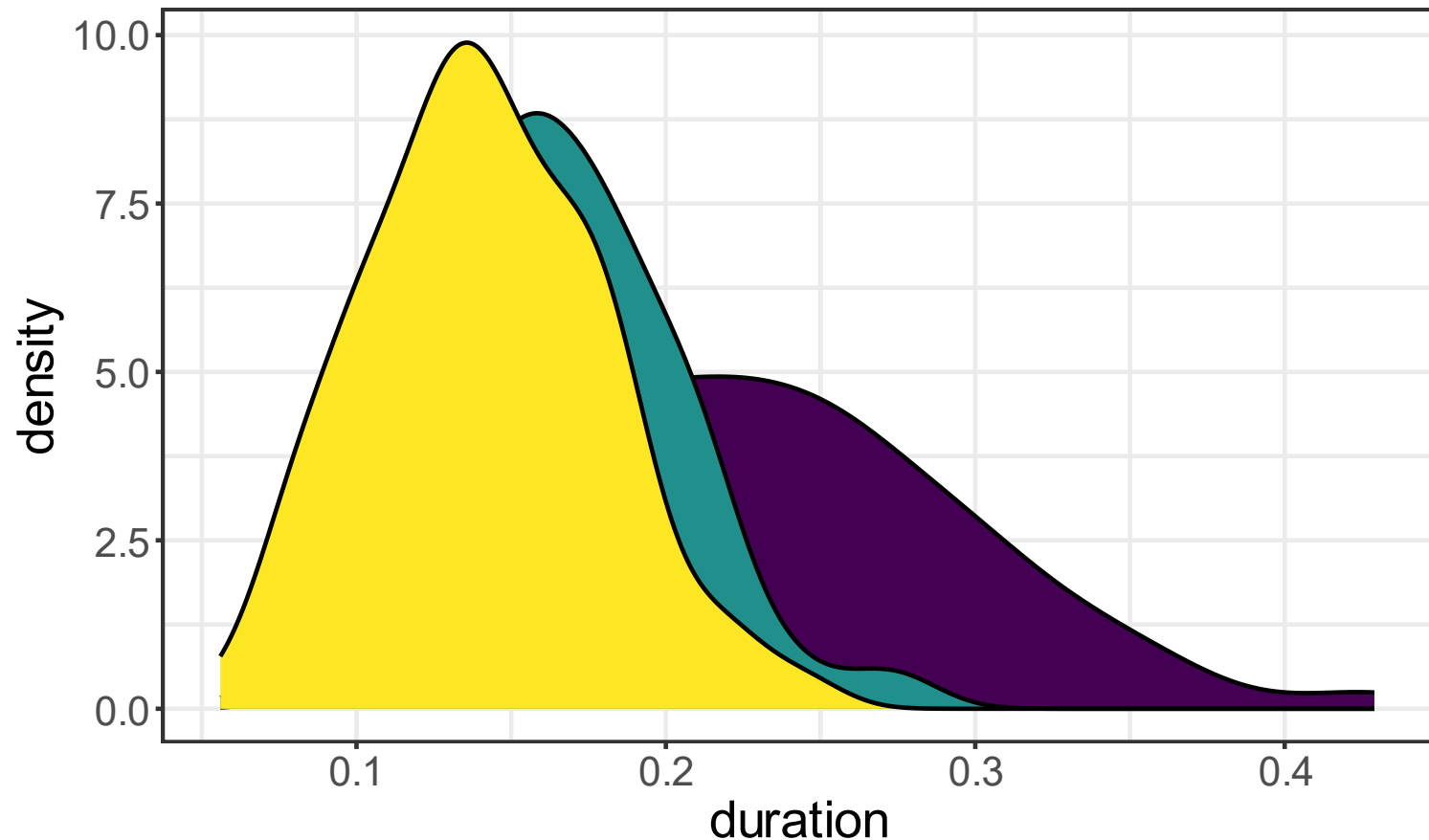
# Histogram

ein Plot pro Level; Hintergrund = alle Daten



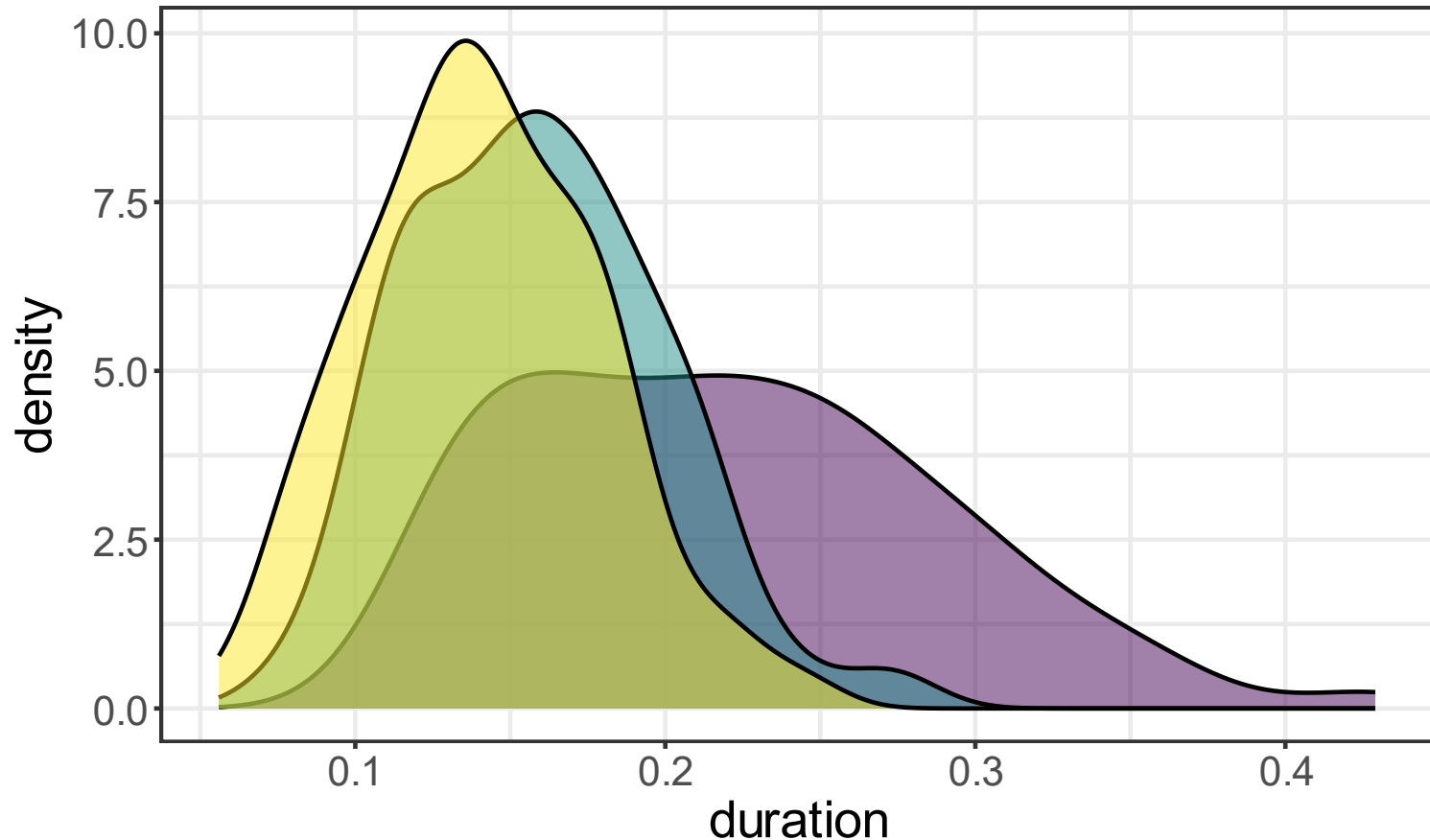
# Density Plot

nützlich zum Check der Distribution – Lesbarkeit okay



# Density Plot

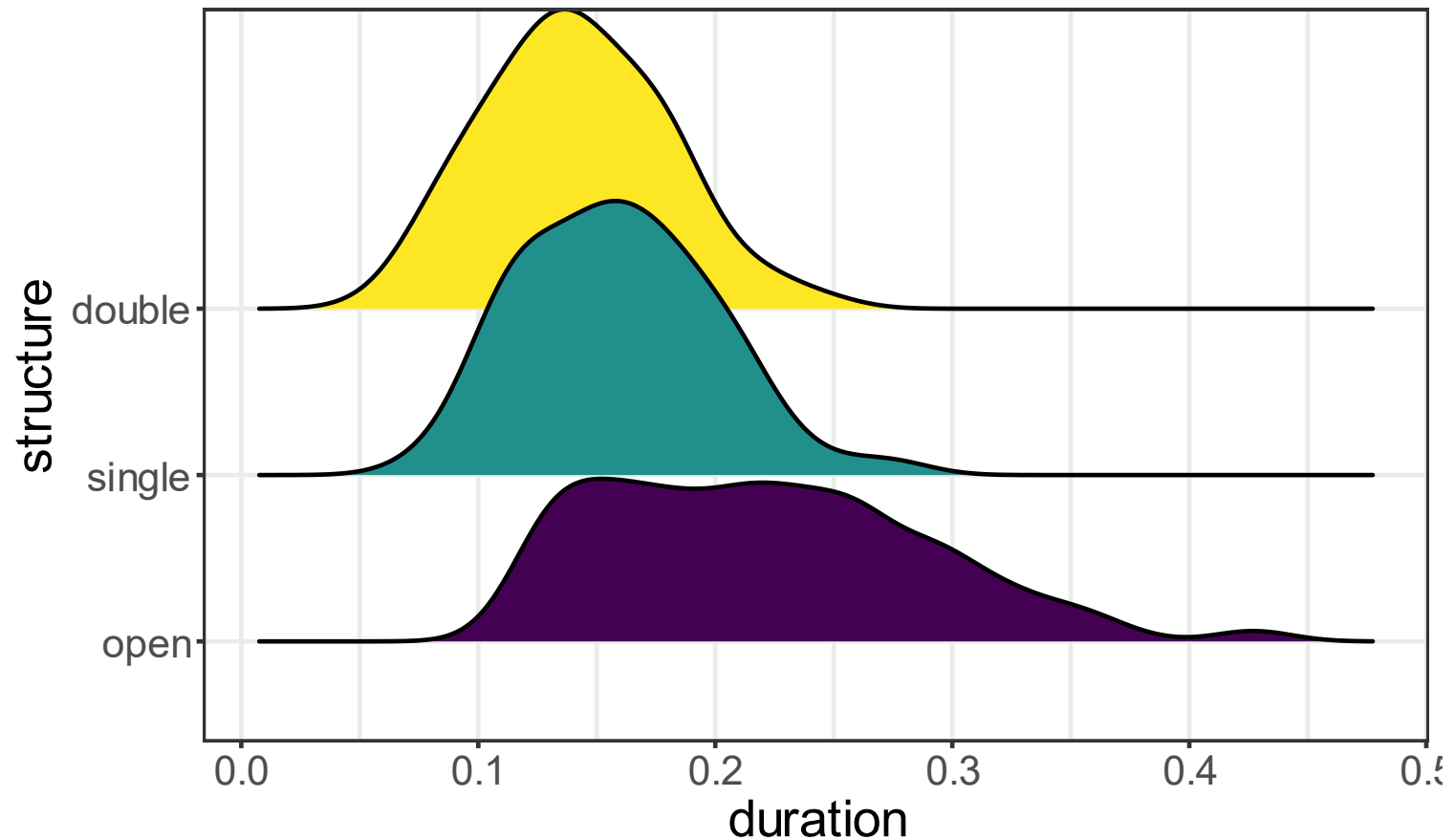
**nützlich zum Check der Distribution – Lesbarkeit gut**





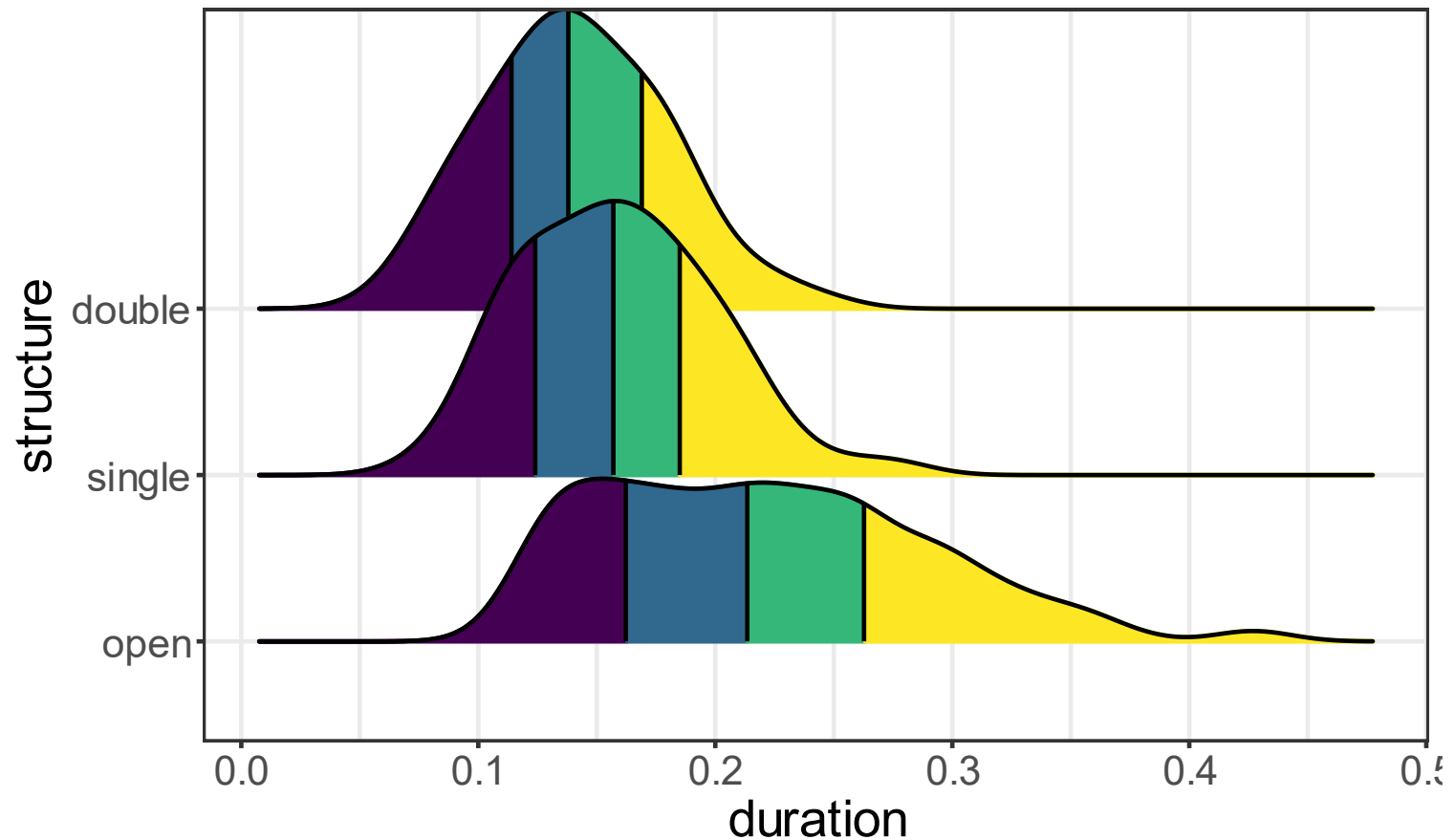
# Density Plot

nützlich zum Check der Distribution



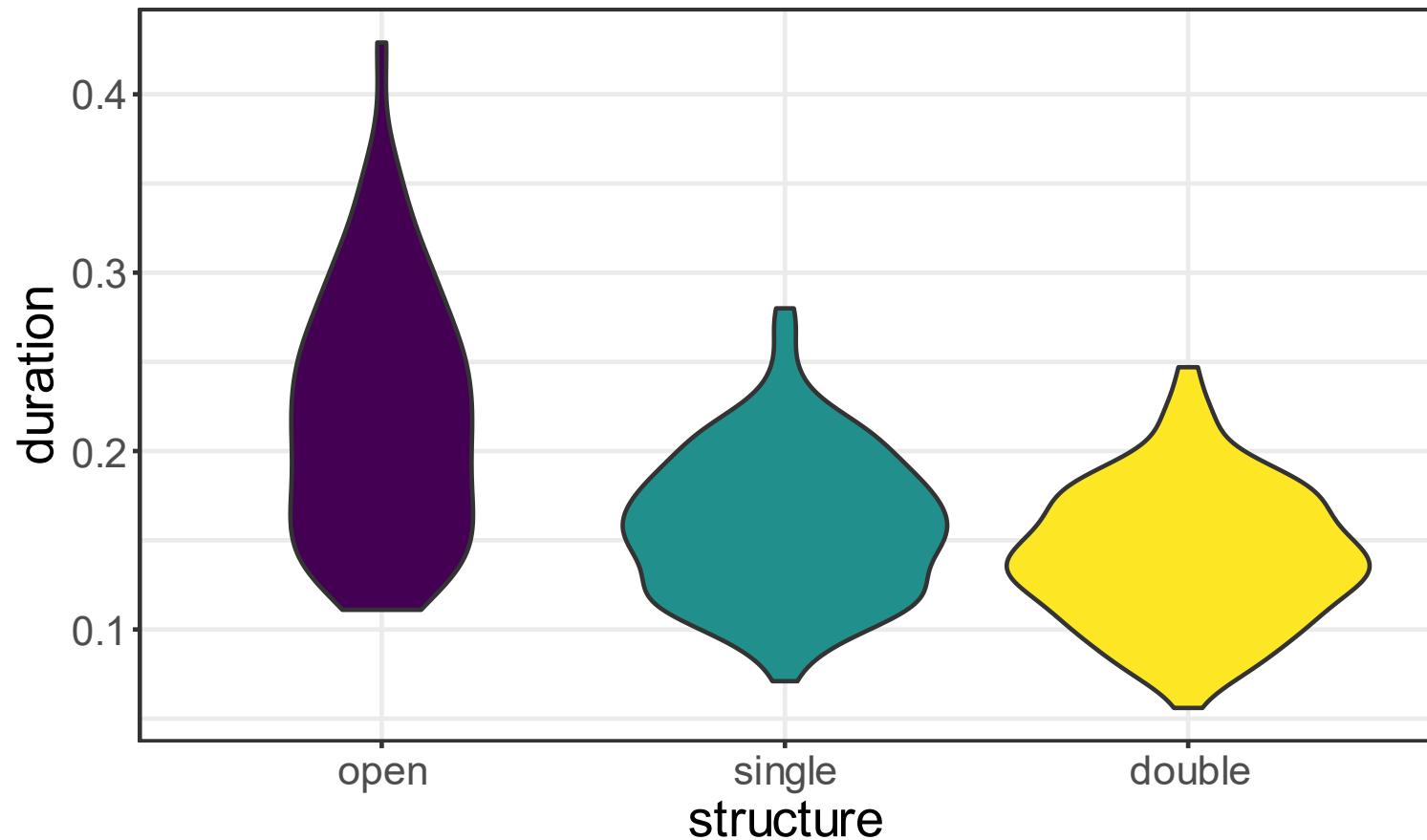
# Density Plot

nützlich zum Check der Distribution



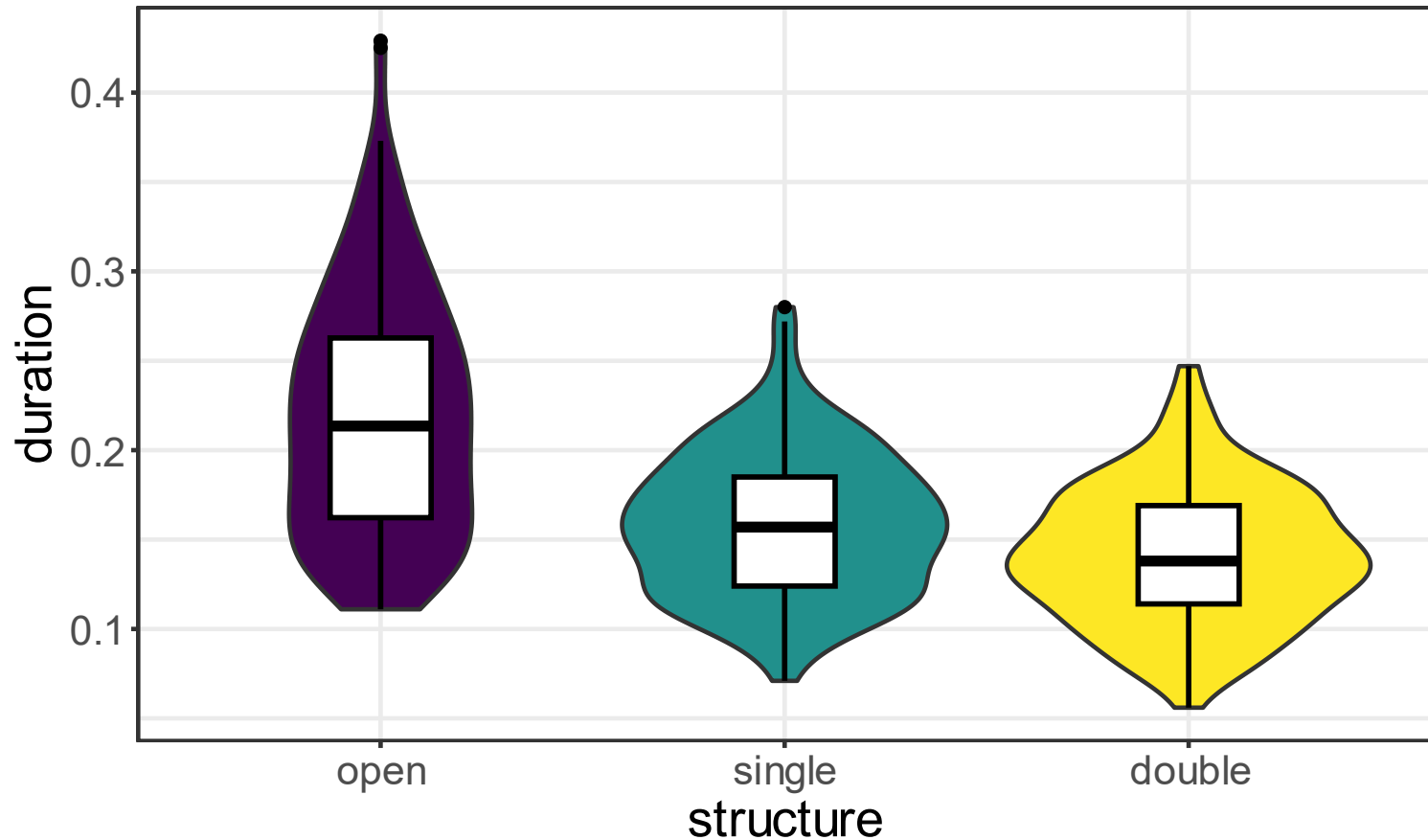
# Violin Plot

nützlich zum Check der Distribution



# Violin Plot

oft mit Box Plots kombiniert

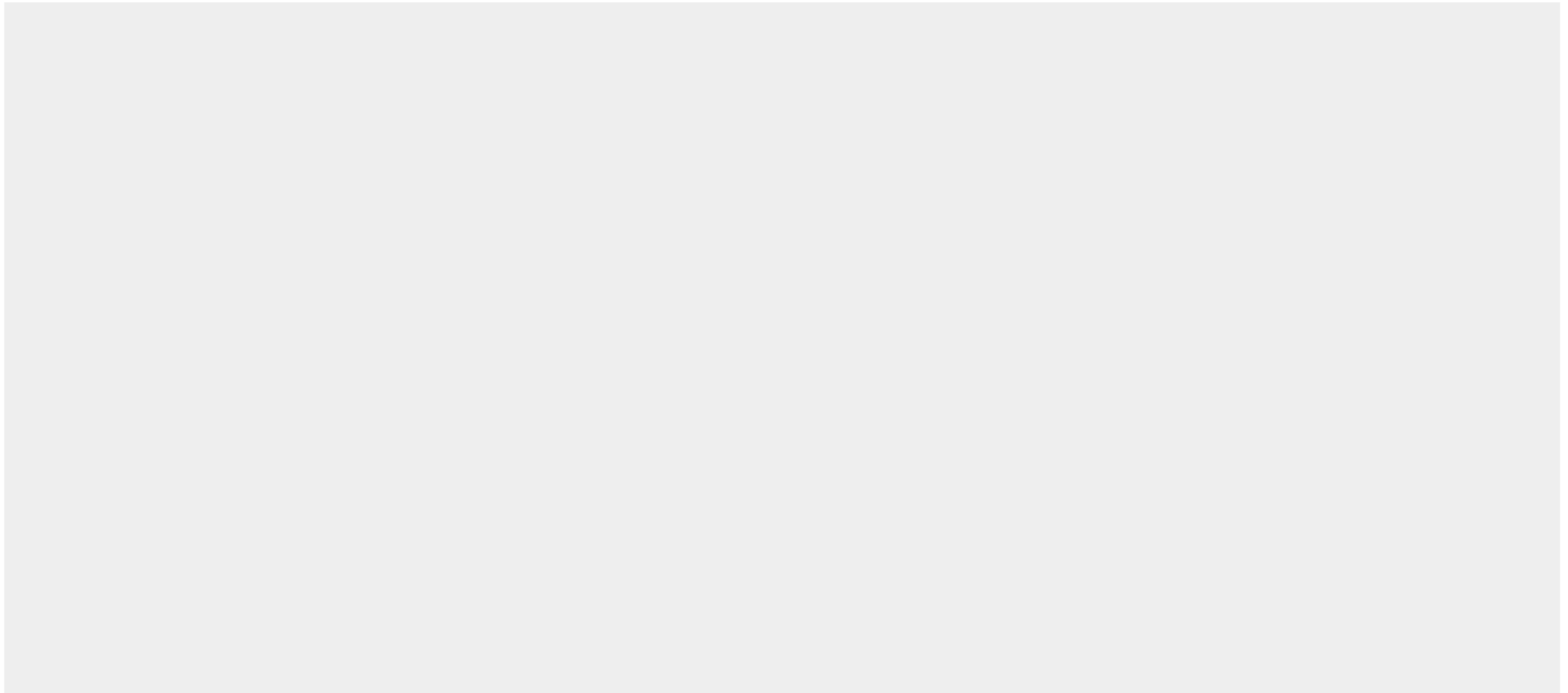


# The Grammar of Graphics

- Die meisten (und wohl auch 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

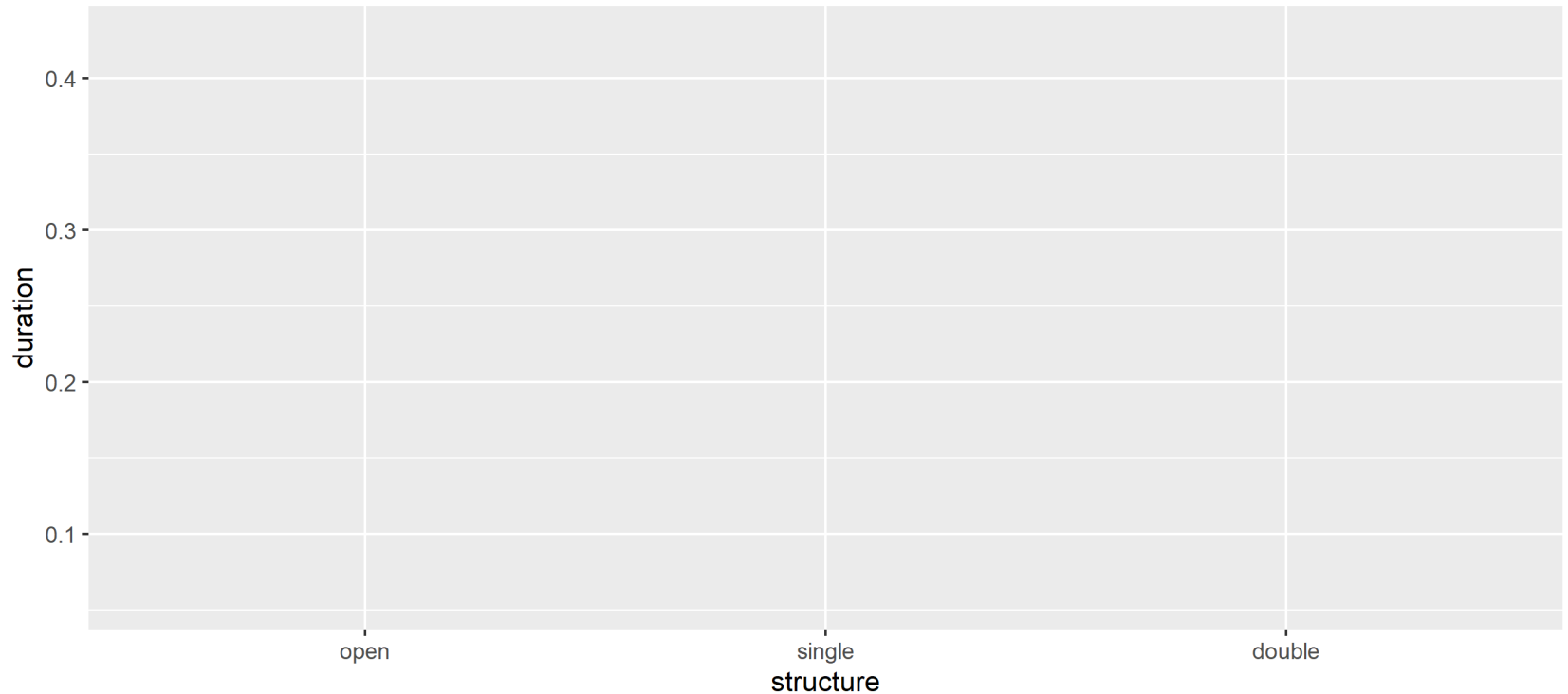
# ggplot2 – data

```
> ggplot(data)
```



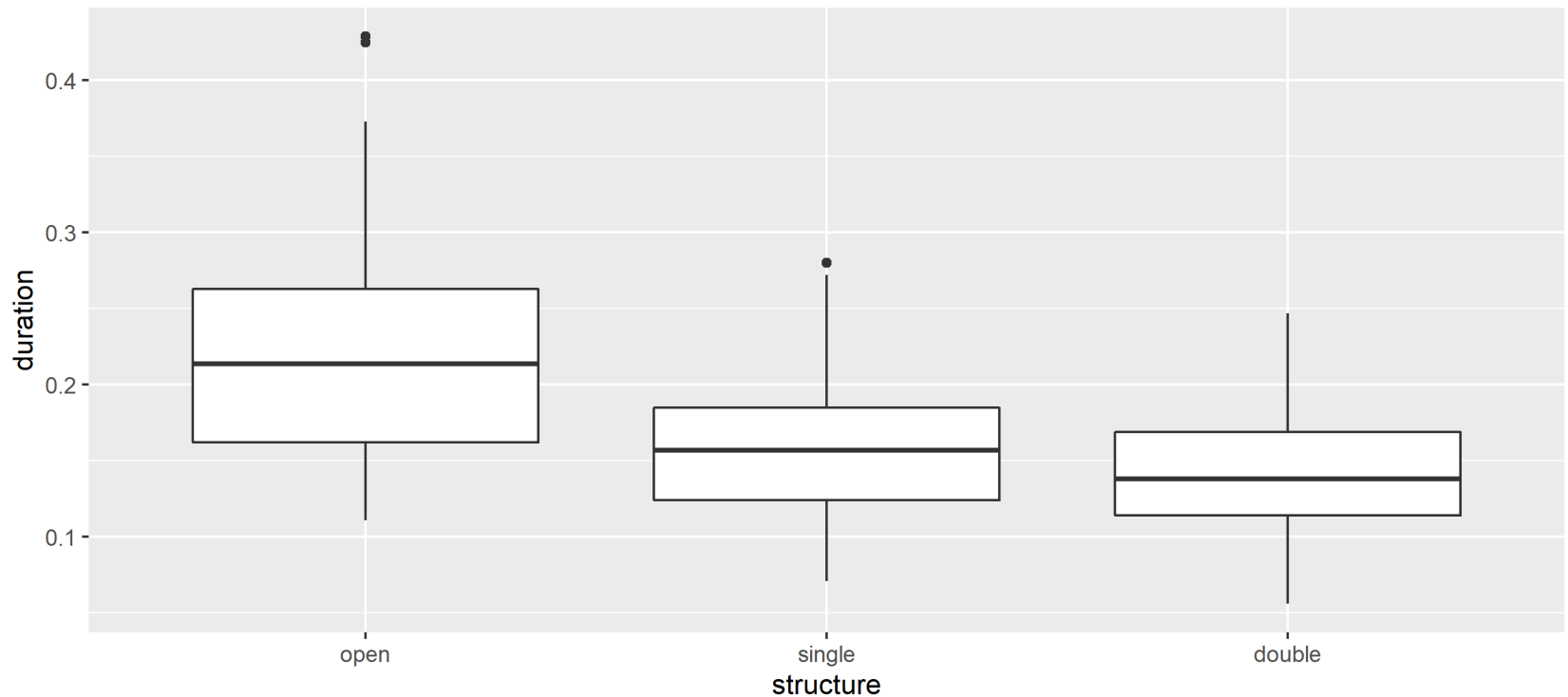
# ggplot2 – aesthetics

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



# ggplot2 – geometric object

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





# Datenvisualisierung

## Problem

Woher weiß ich, welchen Plot ich für meine Daten nutzen kann bzw. soll?

## Lösungen

1. (Un-)systematisches Ausprobieren...
2. Was nutzen andere Wissenschaftler\*innen für ähnliche Daten?
3. From Data to Viz & The R Graph Gallery nutzen

<https://www.data-to-viz.com/>

<https://r-graph-gallery.com/index.html>