Statistics for Linguistics

Session 03

Data Visualisation

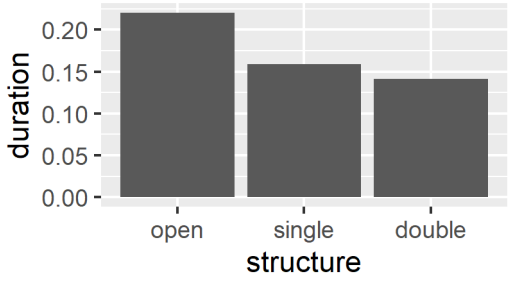
Data Visualisation

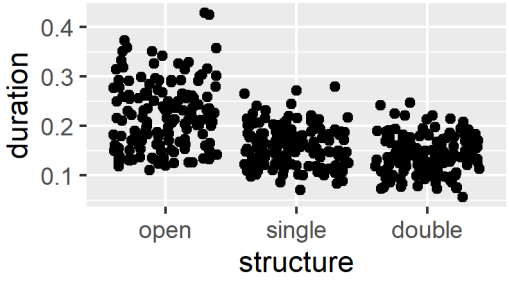
Two Reasons for Data Visualisation

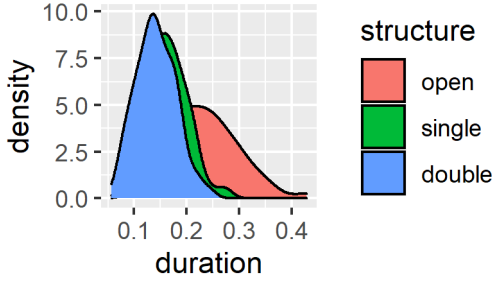
1. We wish to present data in a more visual form, i.e. as plots, to convey a message
2. We wish to take a closer look at our data, i.e. to find patterns, check distributions, etc.

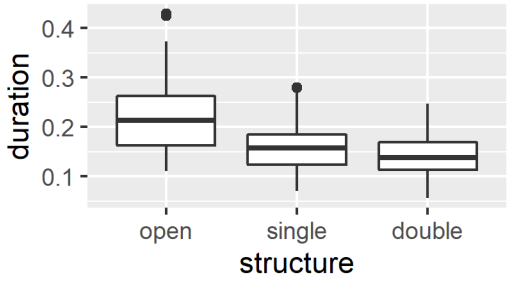
- In any case, there are many different types of plots, but choosing the right one is often crucial to transport our message

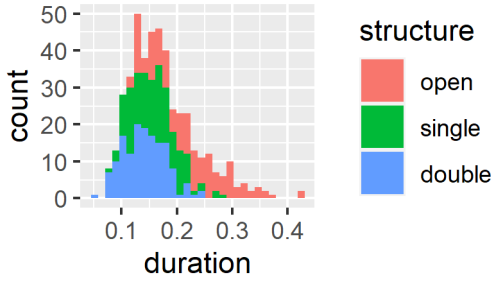
2 Variables, ∞ Plots

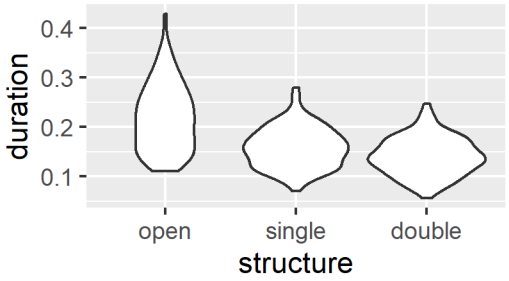


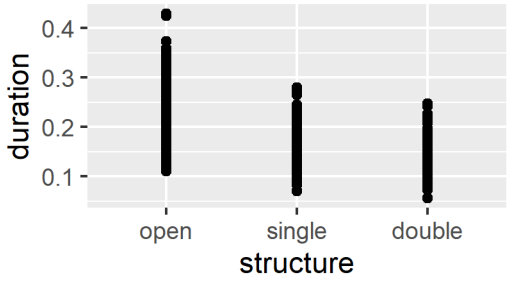


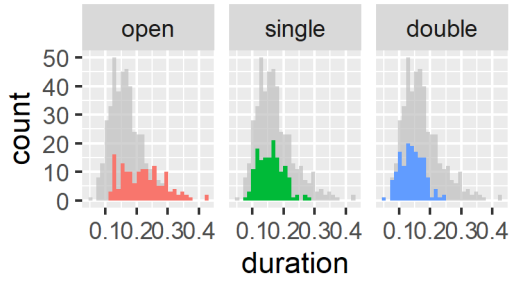


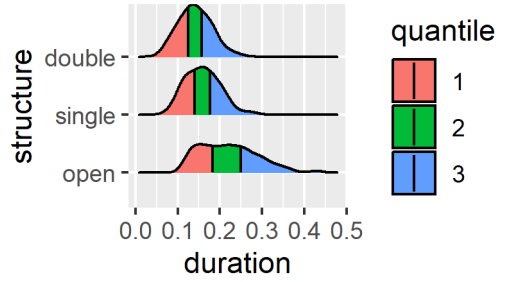




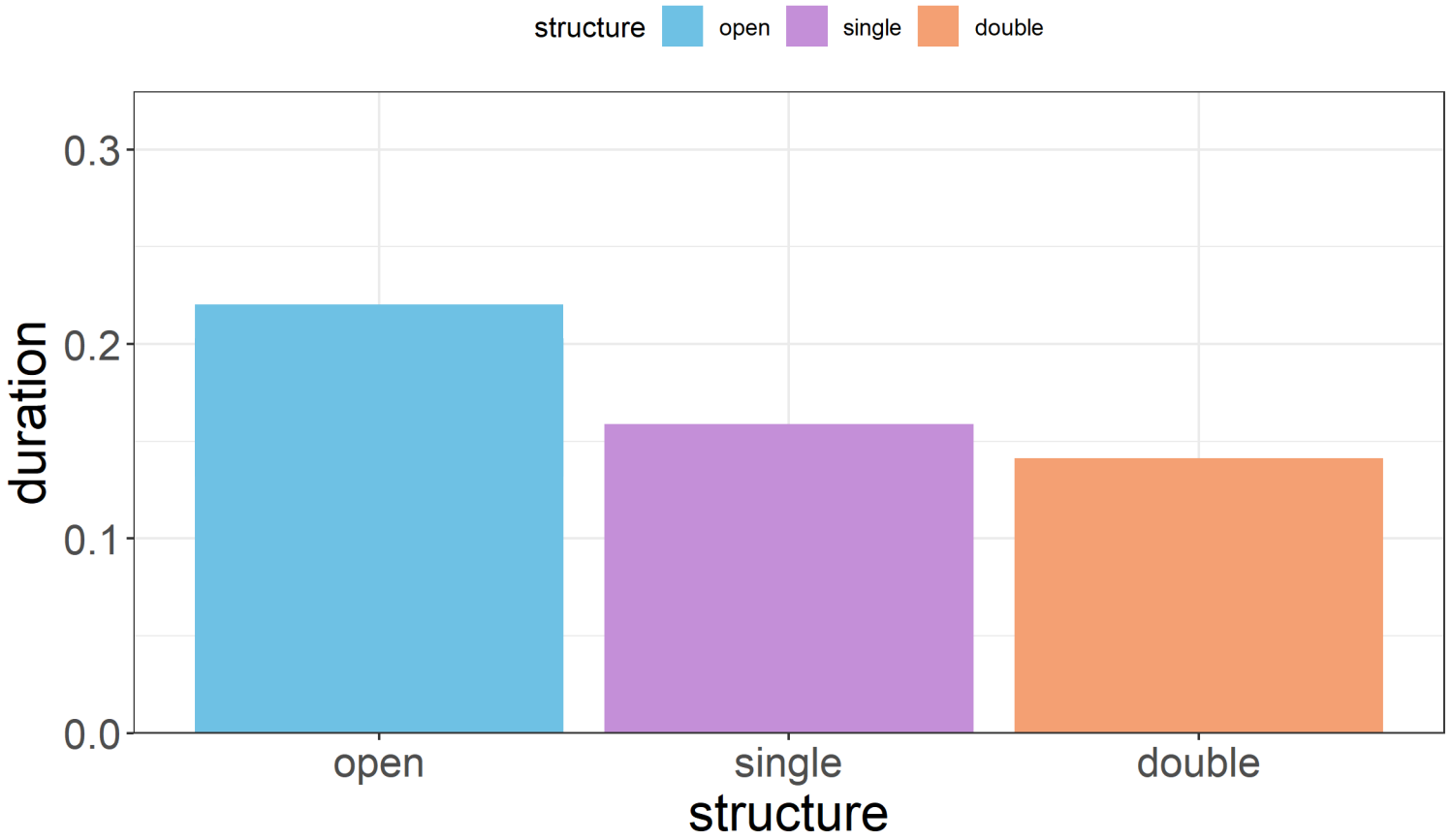




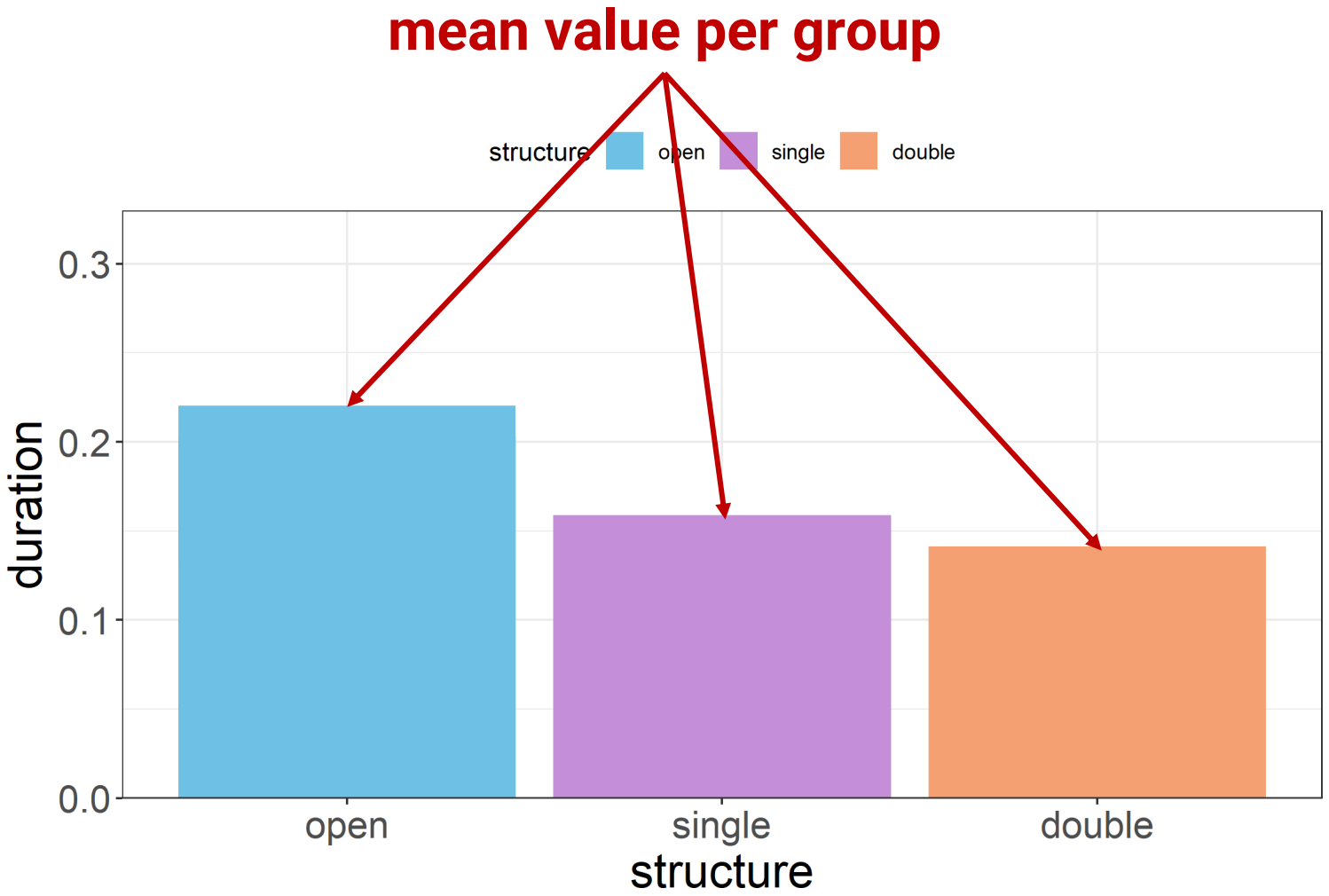




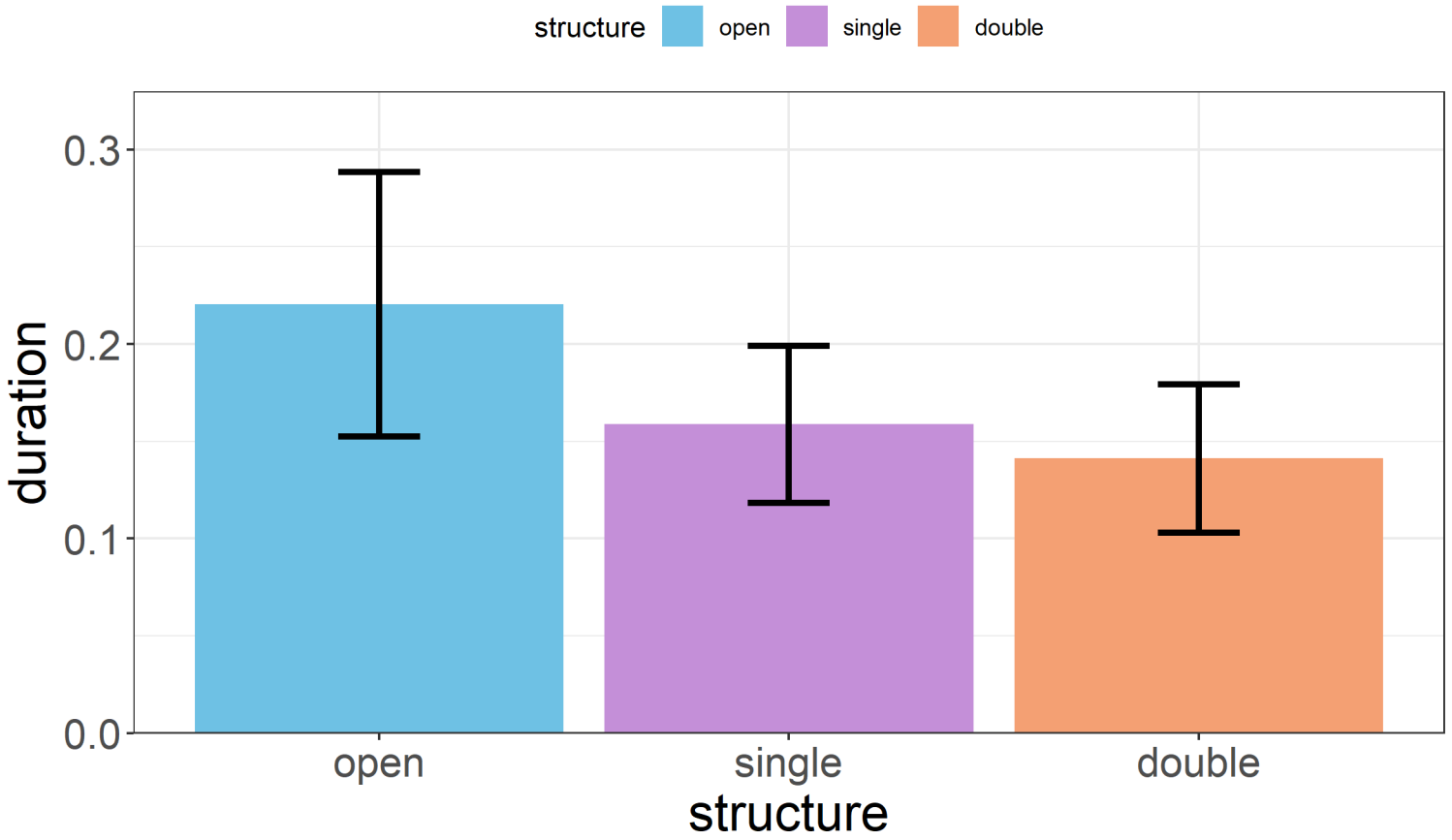
Bar Plot / Bar Chart / Bar Graph



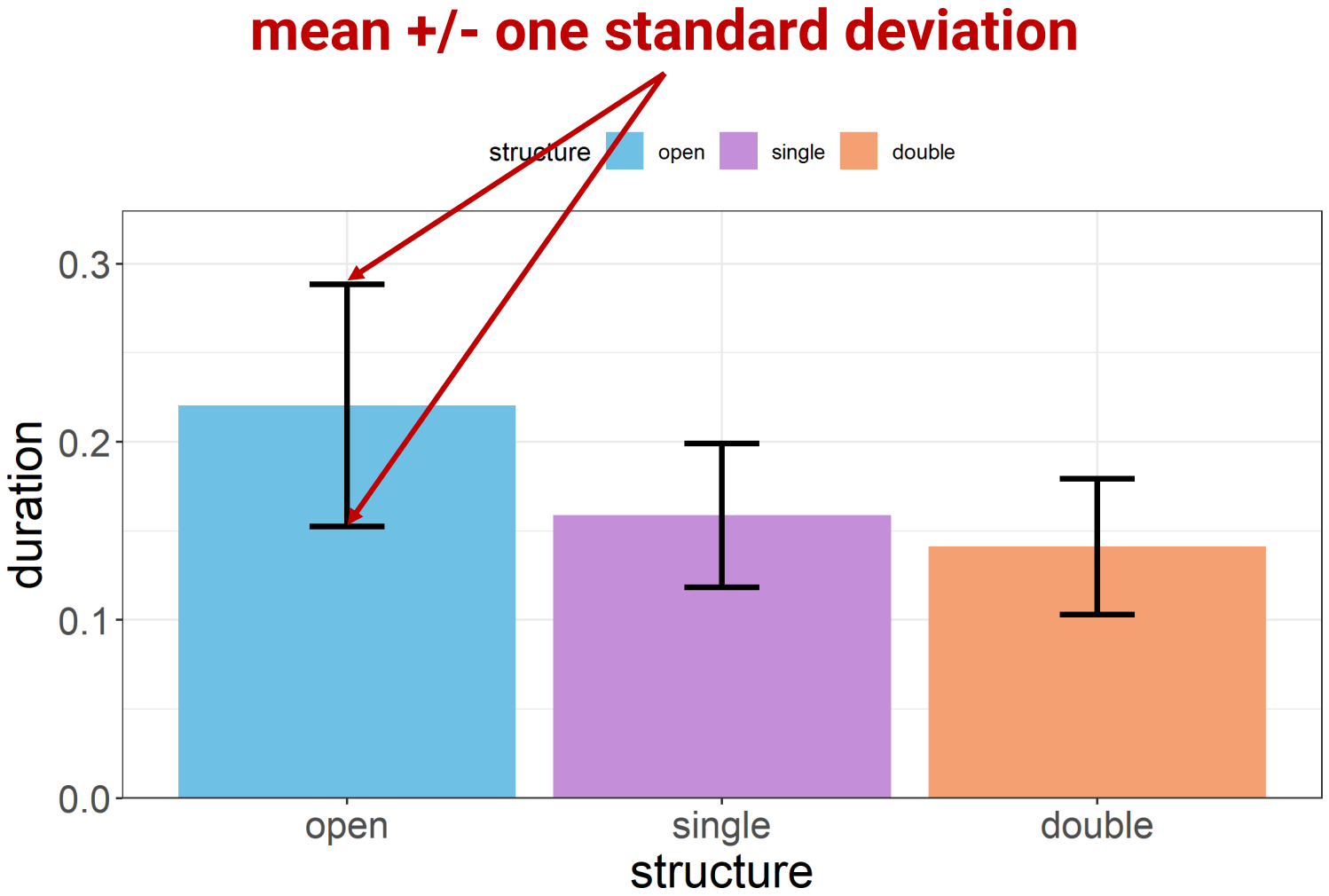
Bar Plot / Bar Chart / Bar Graph



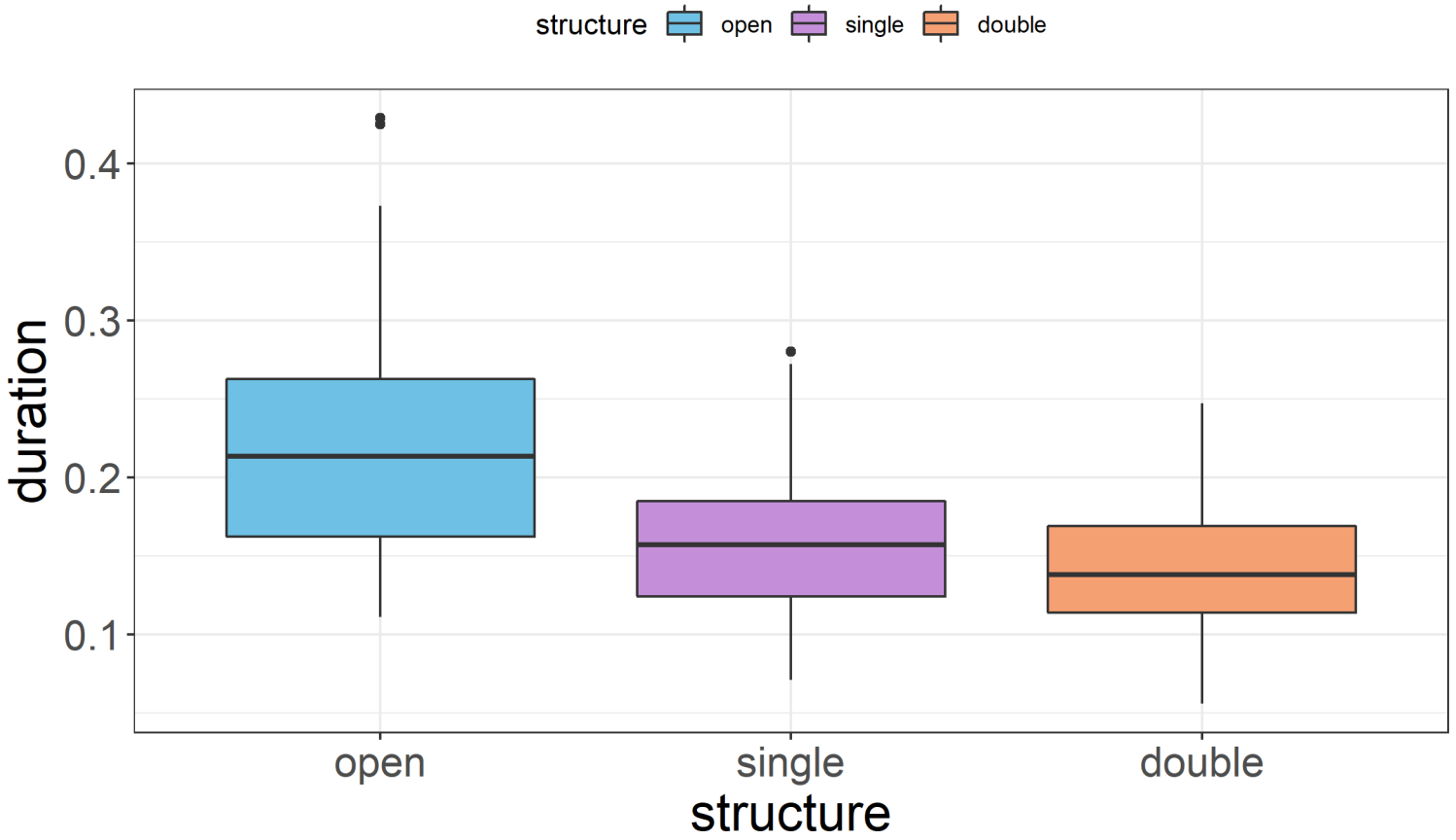
Bar Plot / Bar Chart / Bar Graph



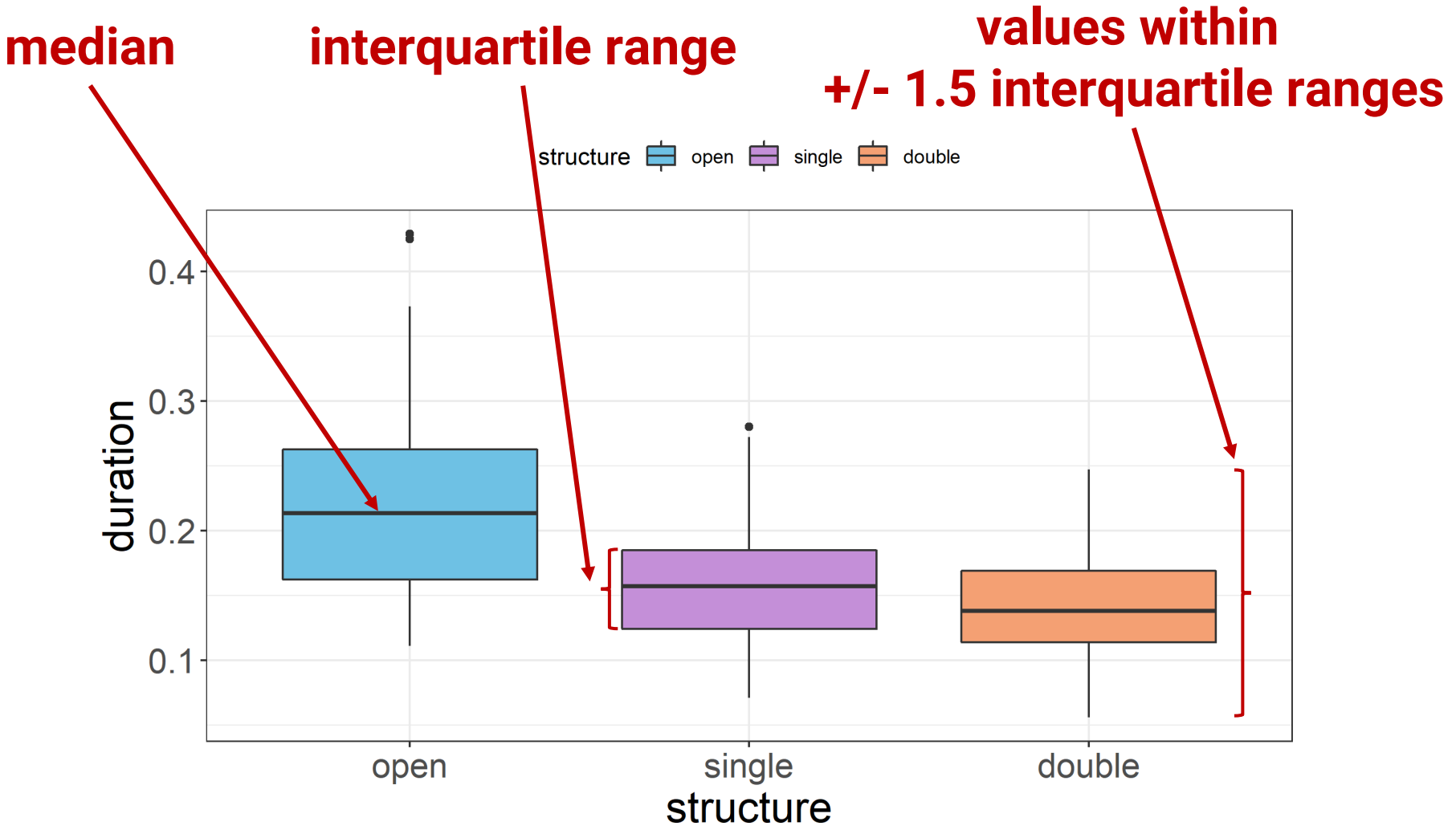
Bar Plot / Bar Chart / Bar Graph



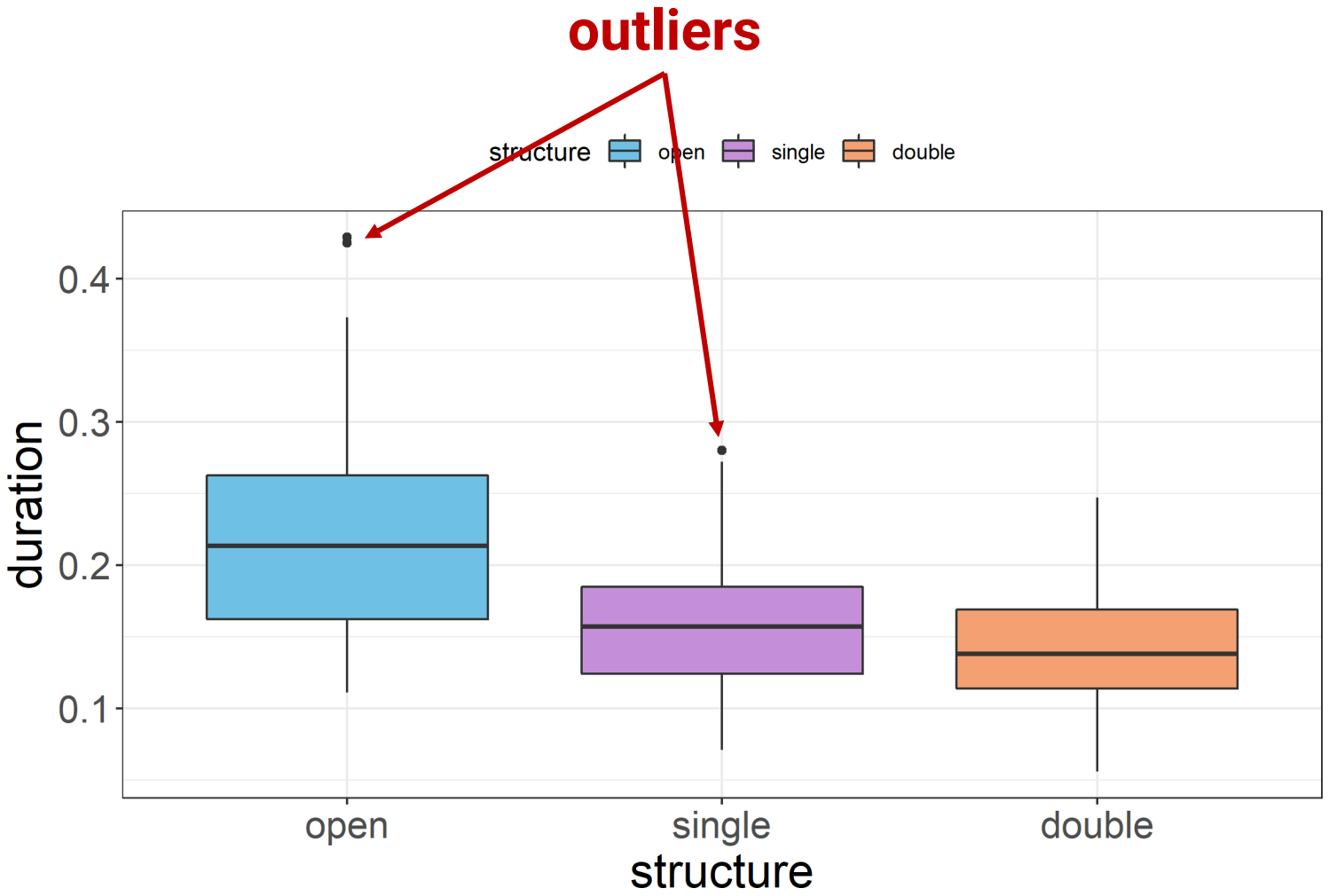
Box Plot / Box-and-Whisker Diagram



Box Plot / Box-and-Whisker Diagram

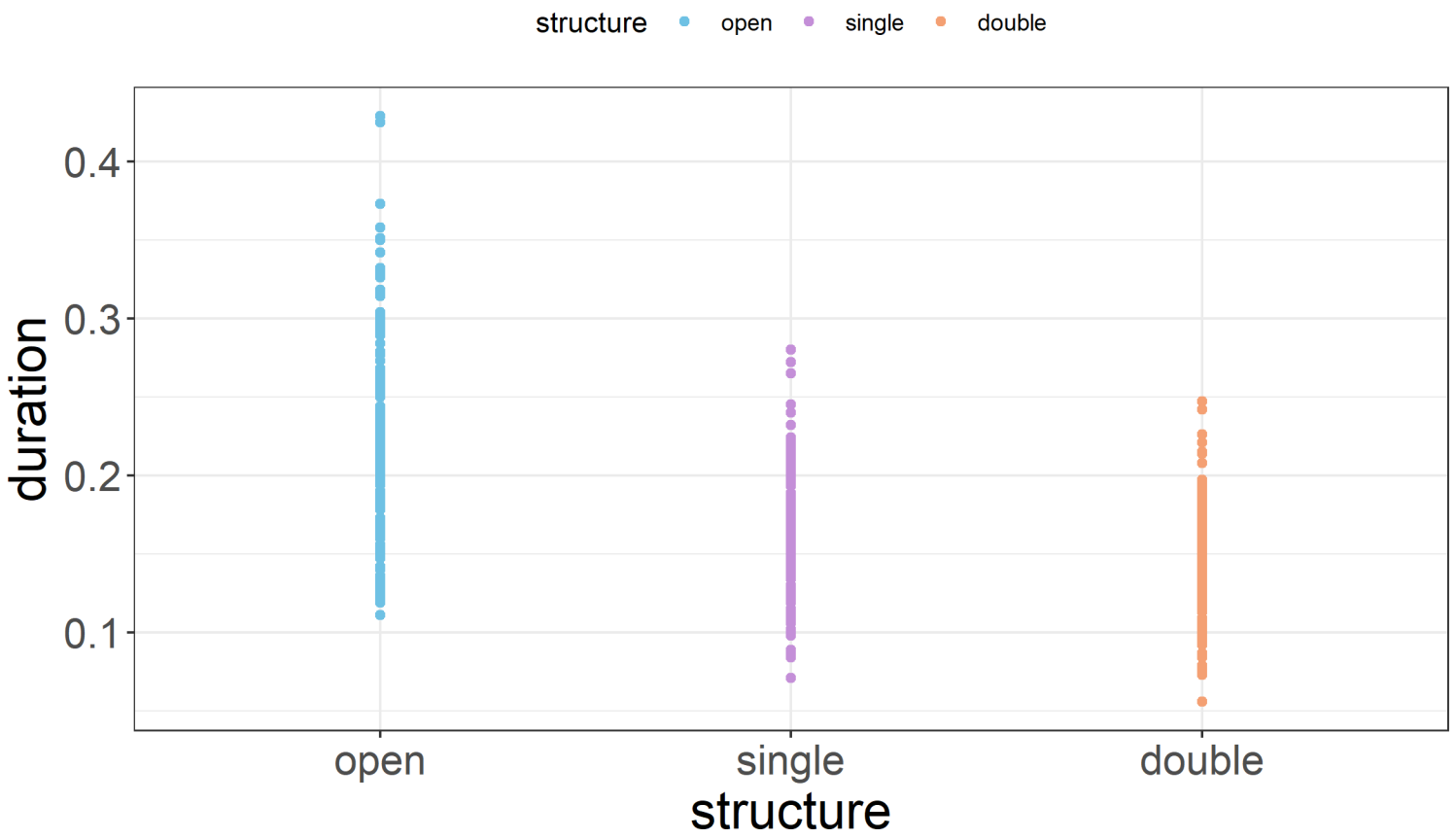


Box Plot / Box-and-Whisker Diagram



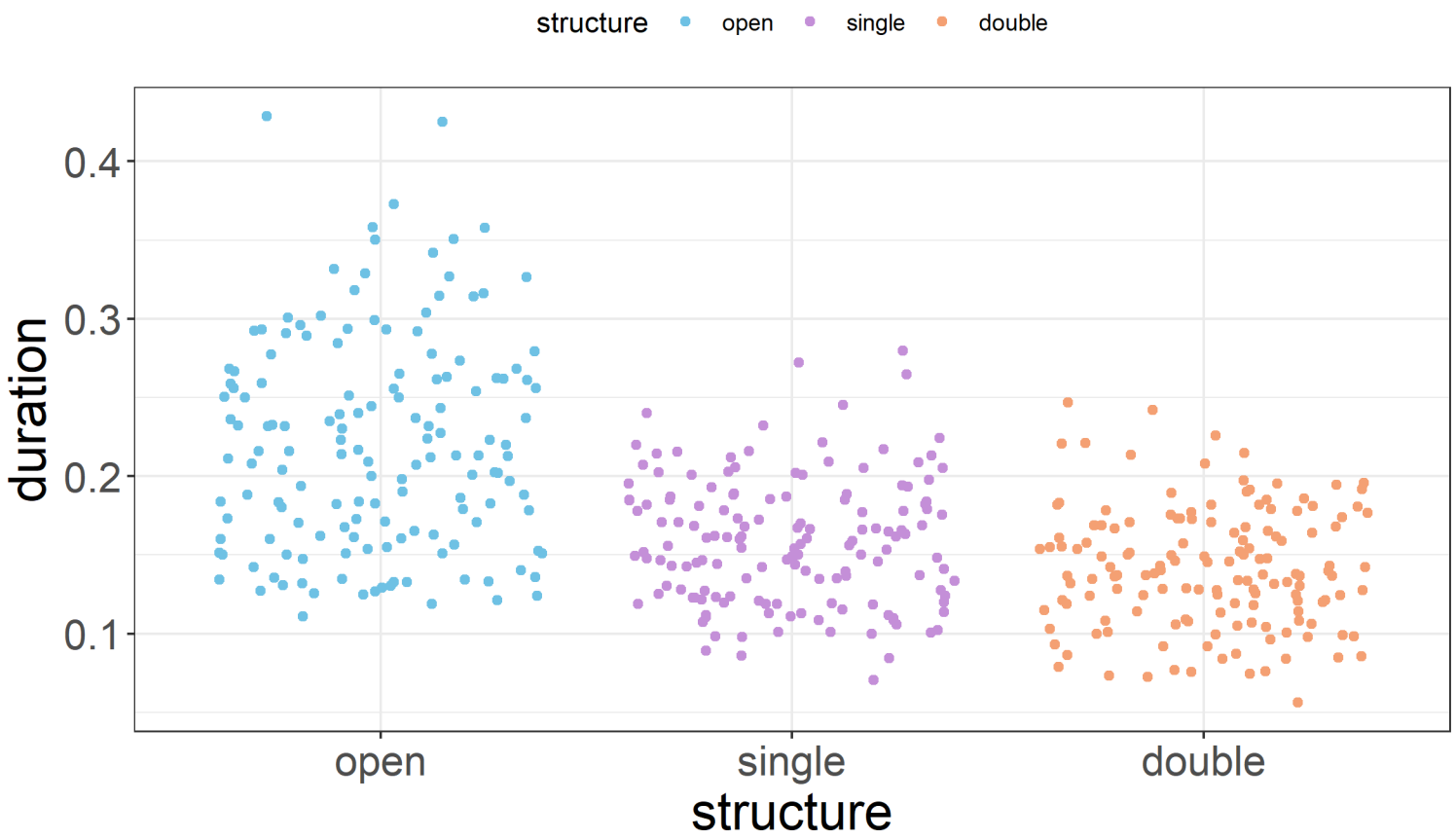
Point Plot / Dot Plot / Dot Chart

one point per value



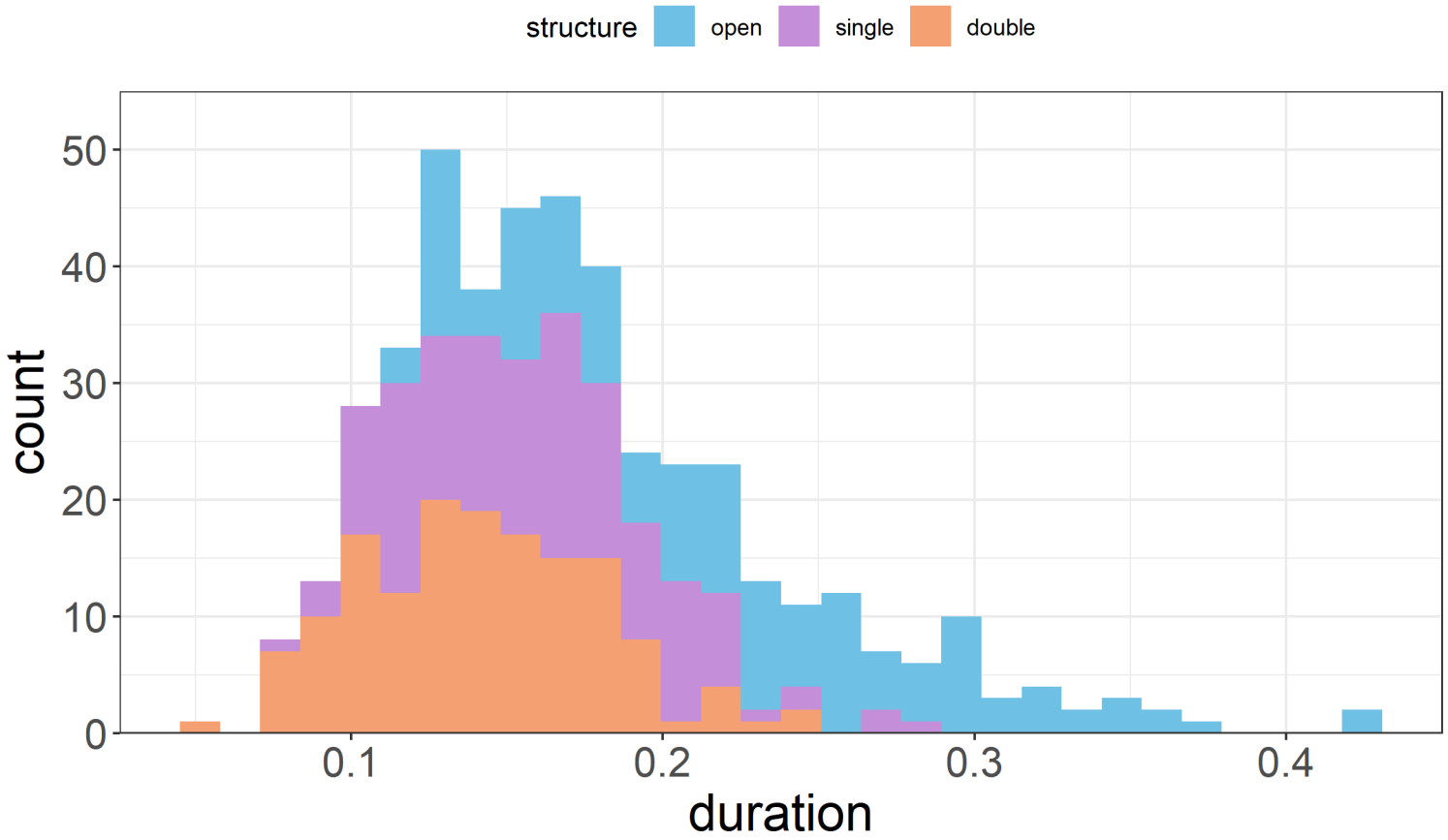
Jitter Plot / Scatter Plot / Scattergram

one point per data point



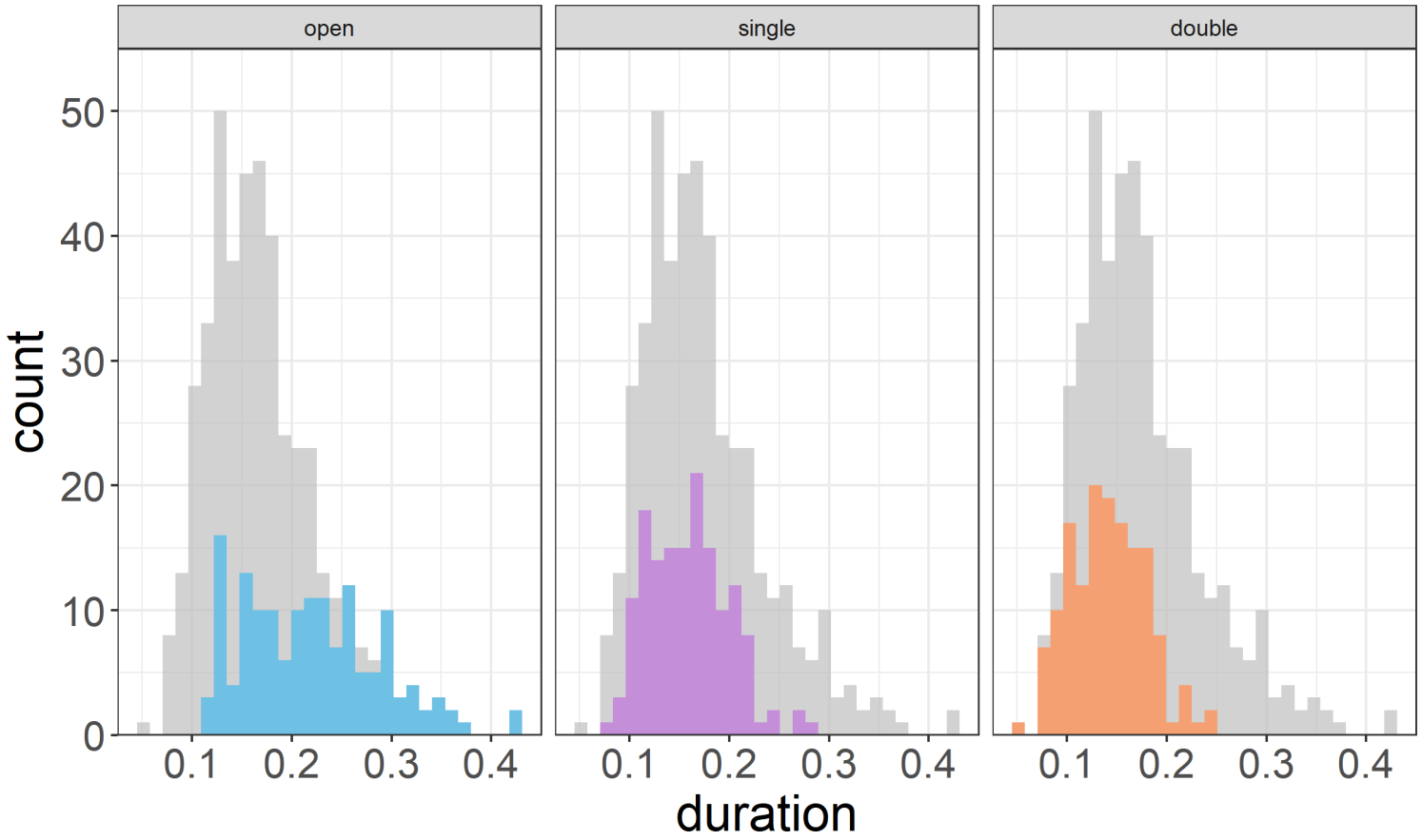
Histogram

useful for distribution checks – readability though?



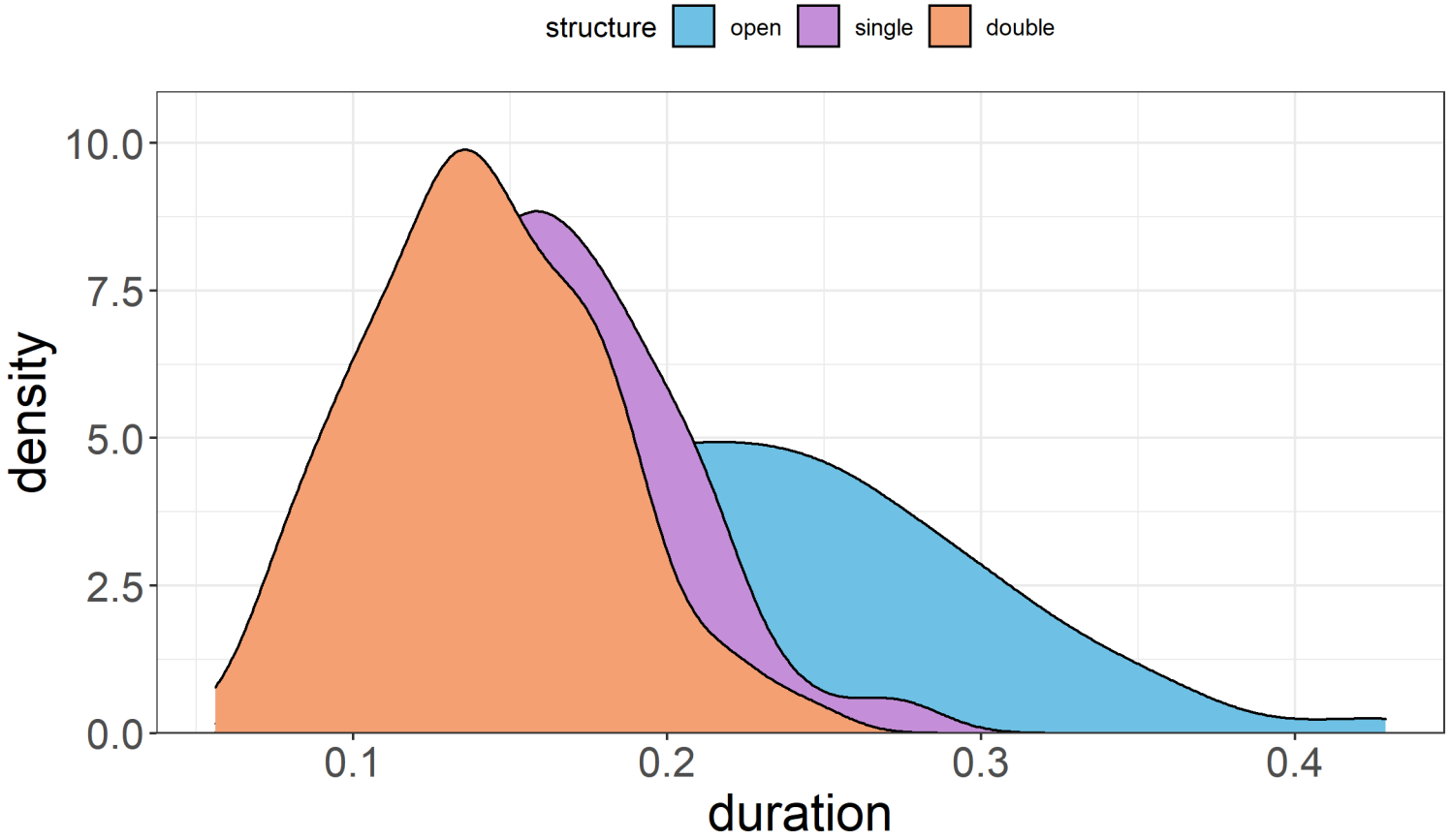
Histogram

one plot per level; background = all levels



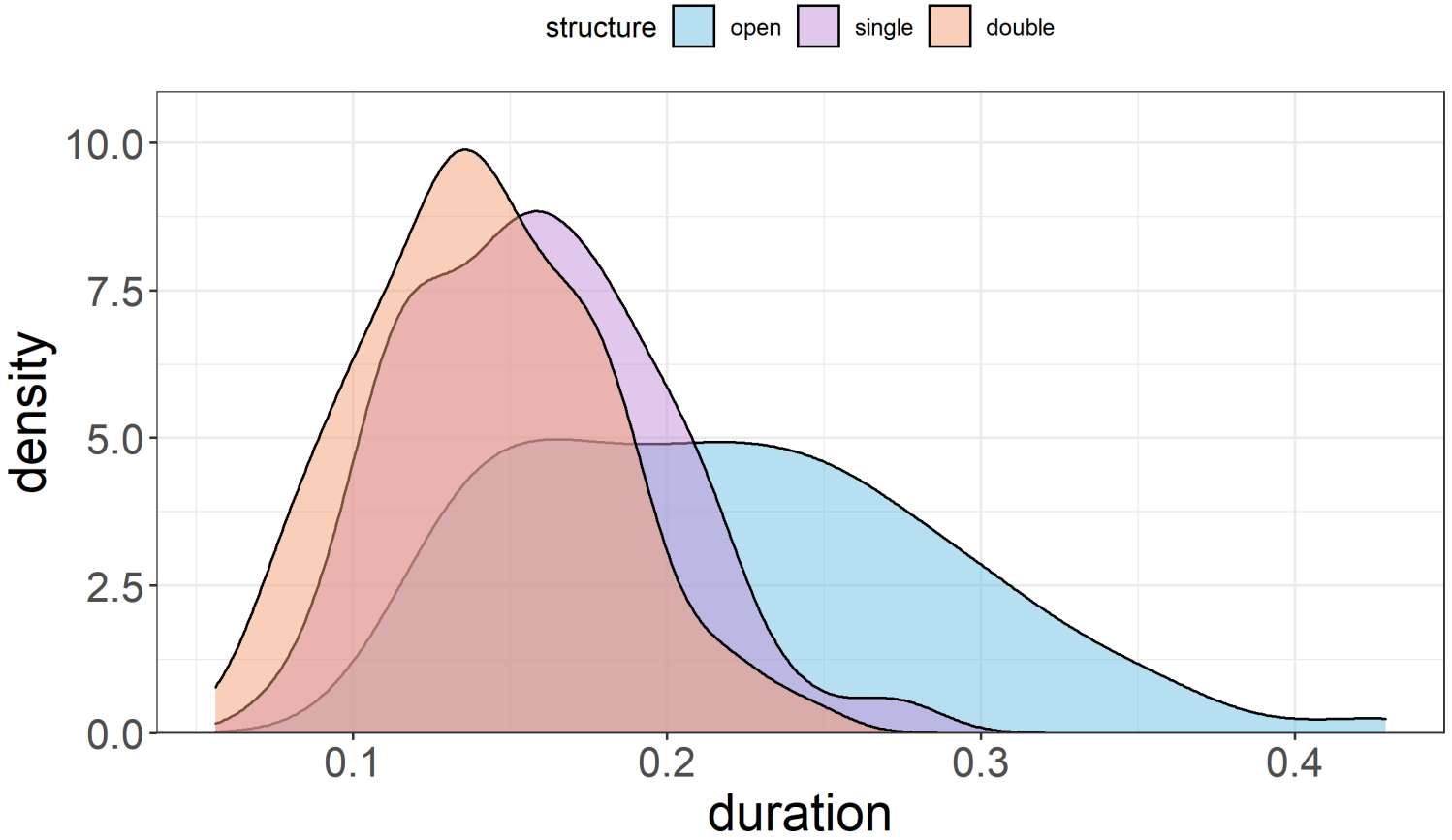
Density Plot

useful for distribution checks – okay readability



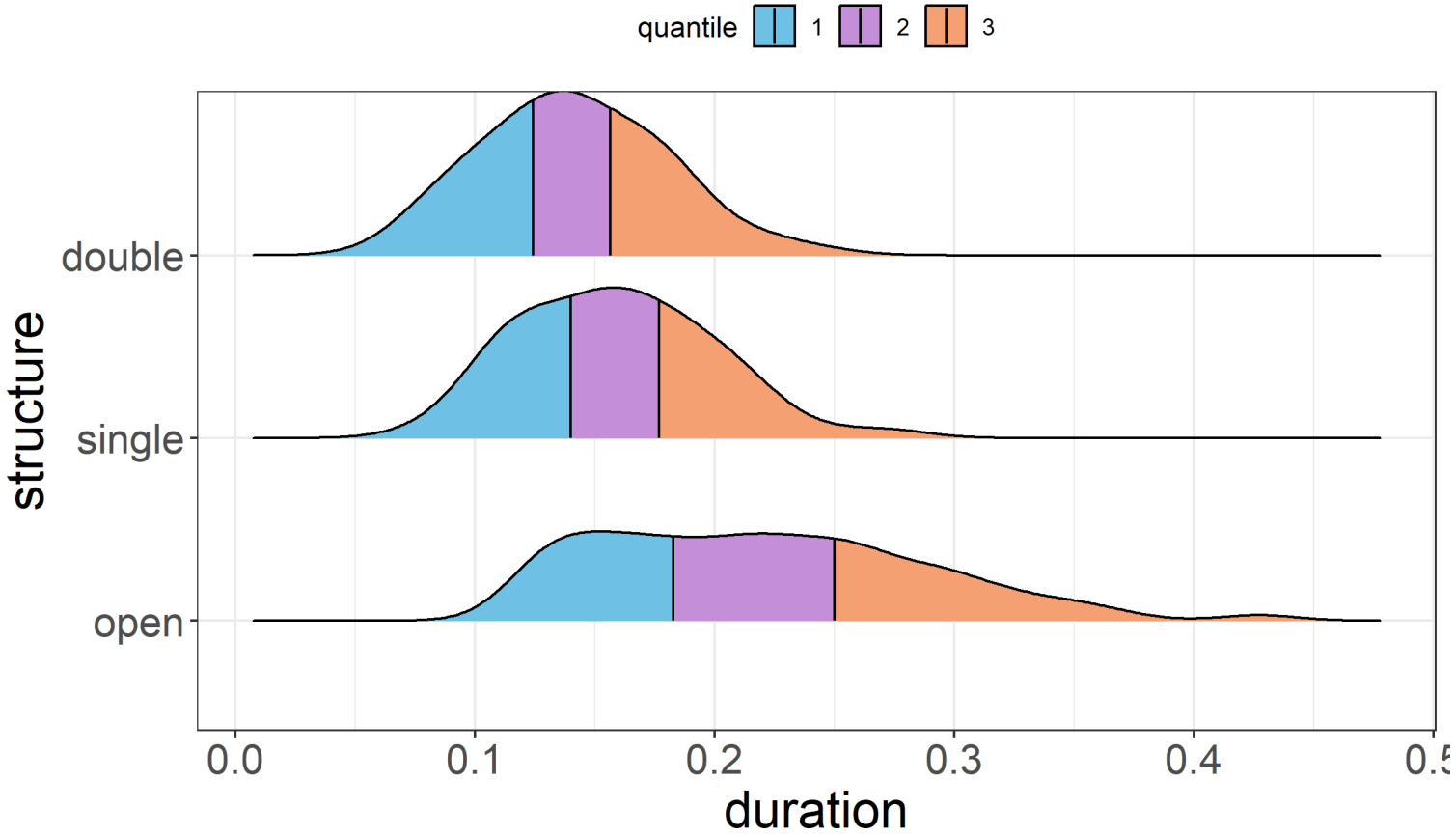
Density Plot

useful for distribution checks – good readability



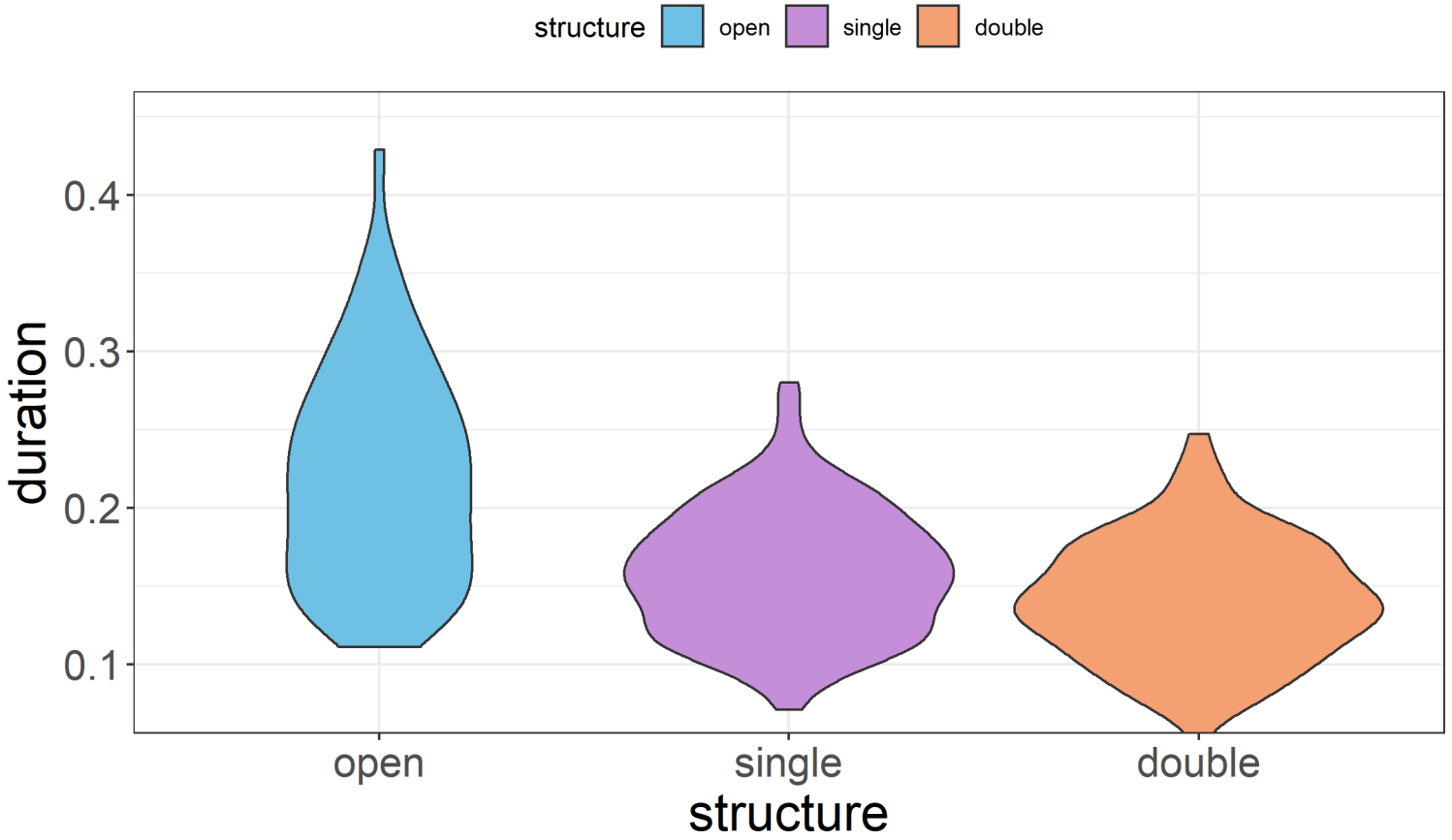
Density Plot

useful for distribution checks



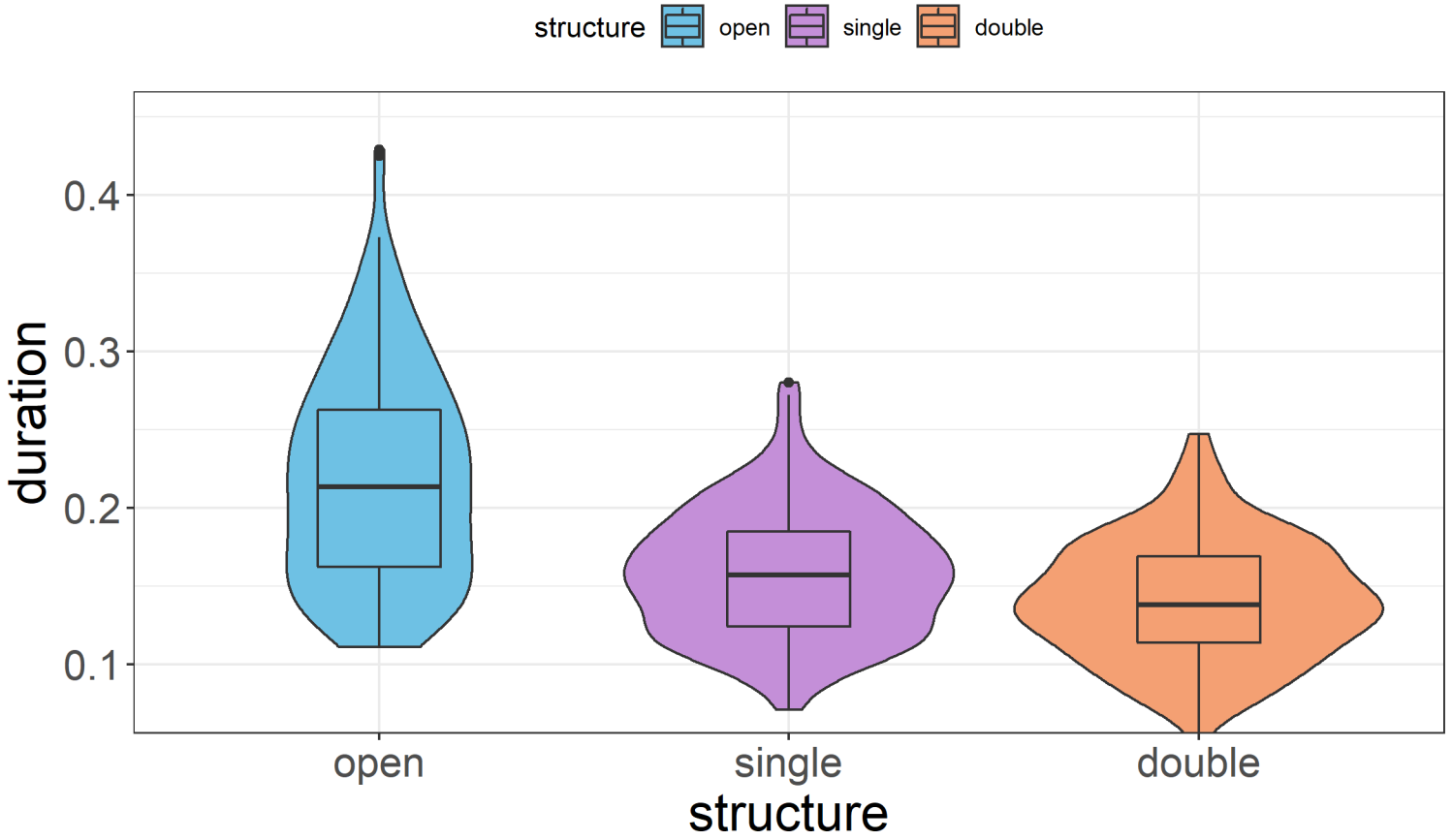
Violin Plot

useful for distribution checks



Violin Plot

often combined with box plots



The Grammar of Graphics

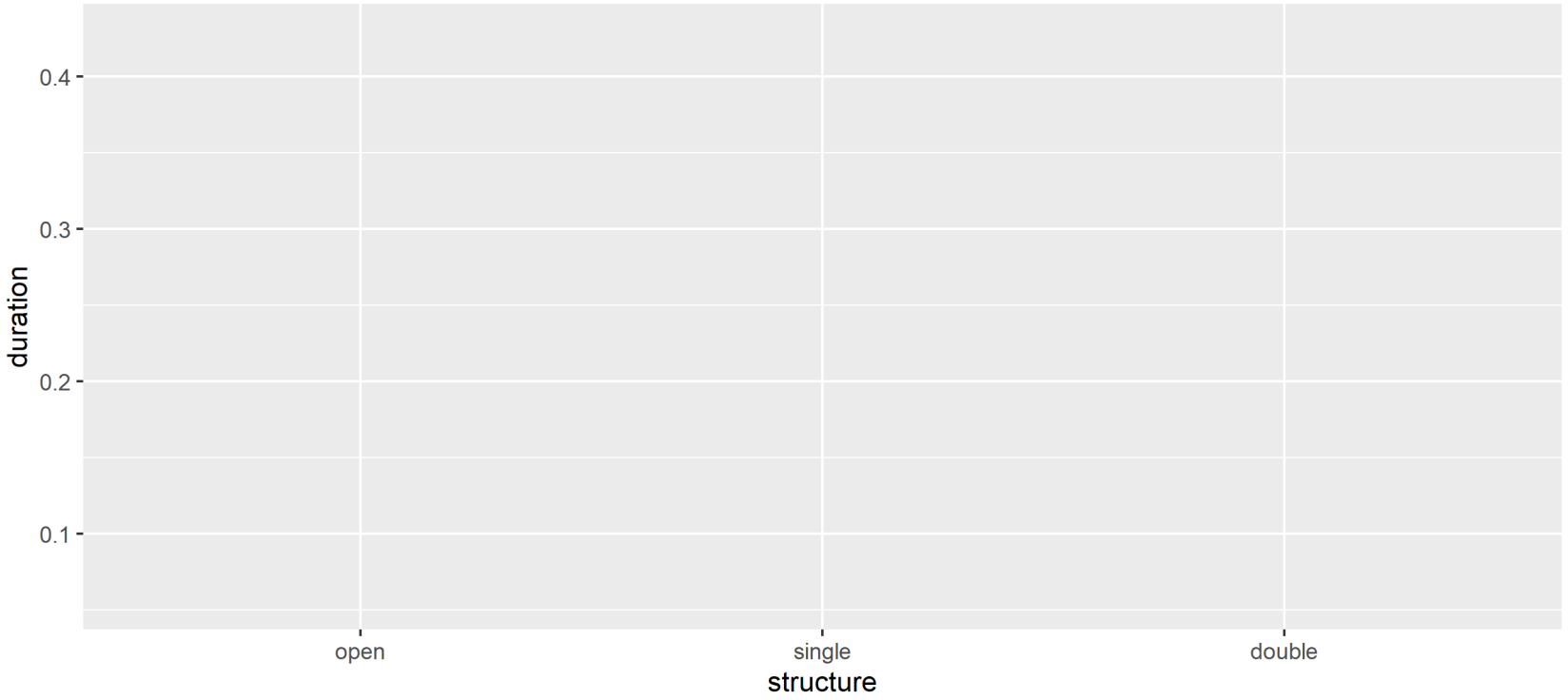
- In R, most plots are created using the `ggplot2` package
- `ggplot2` follows the ideas of *The Grammar of Graphics*, a seminal work on data visualisation by Leland Wilkinson (doi: 10.1007/0-387-28695-0)
- In `ggplot2`, each plot consists of 3 elements:
  1. data
  2. aesthetics
  3. geometric object

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()
```