

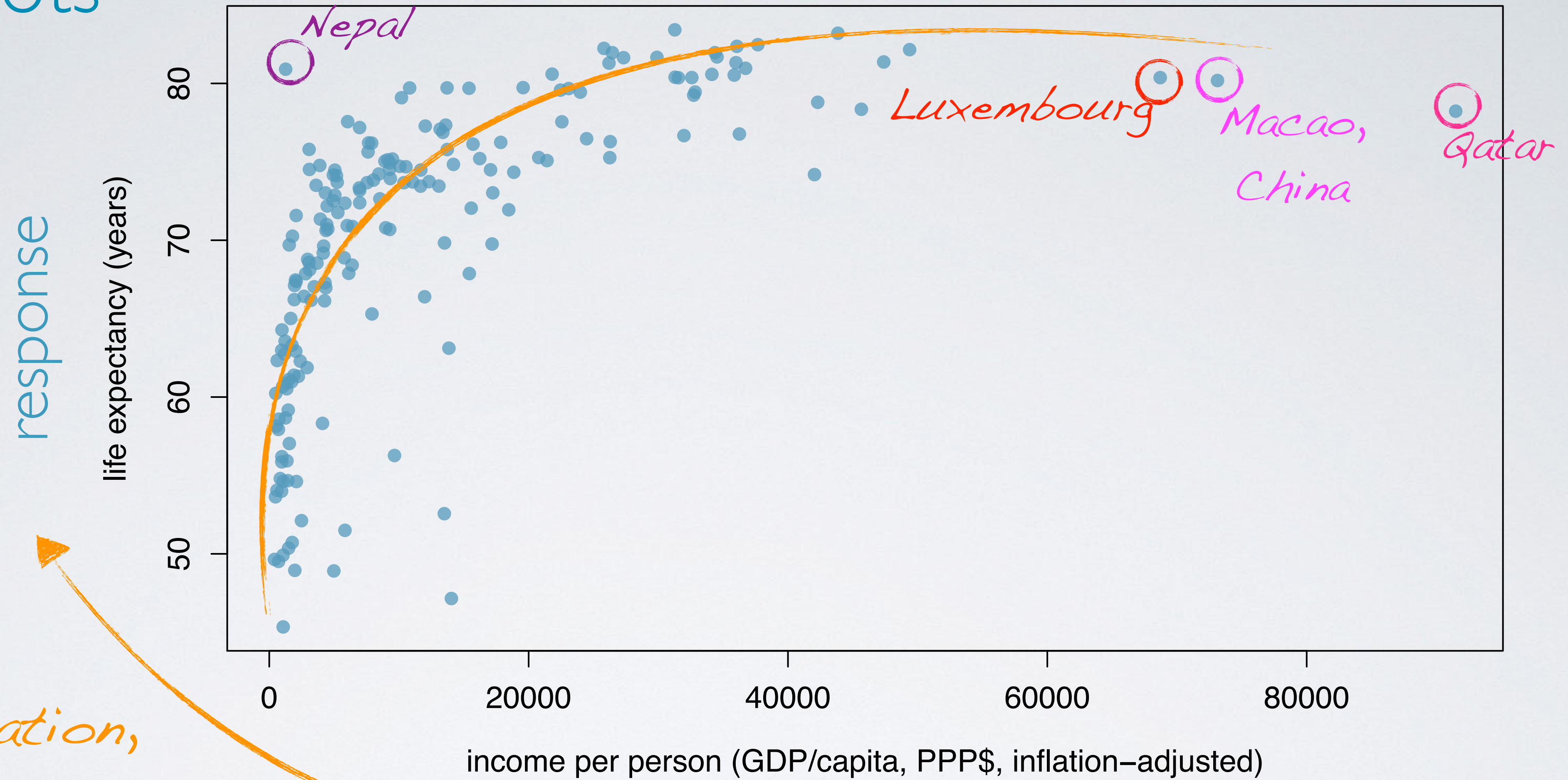
visualizing numerical data

- ▶ scatterplots for paired data
- ▶ other visualizations for describing distributions of numerical variables

data	income per person (\$, 2012)	life expectancy (years, 2012)
Afghanistan	1359.7	60.254
Albania	6969.3	77.185
Algeria	6419.1	70.874
...
Zimbabwe	545.3	58.142

Source: gapminder.com

scatterplots

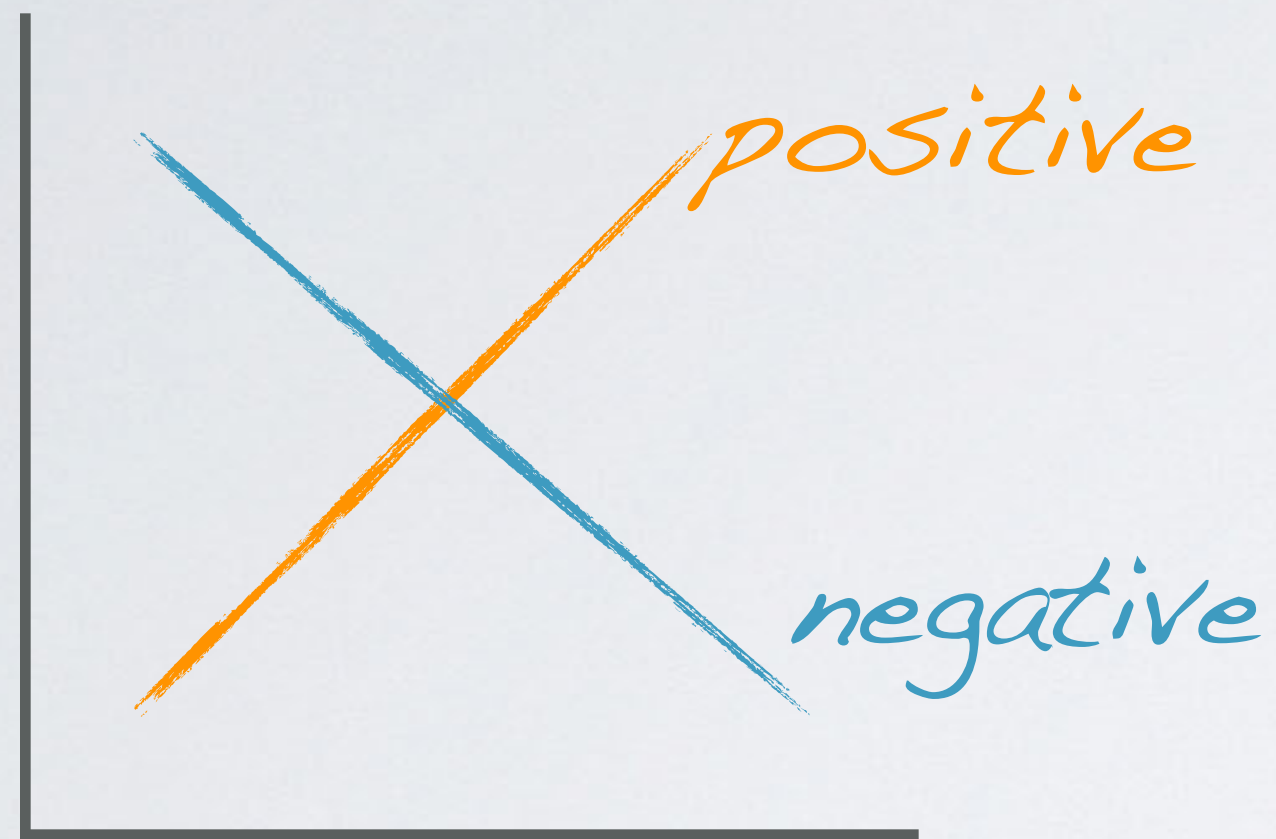


correlation,
not causation

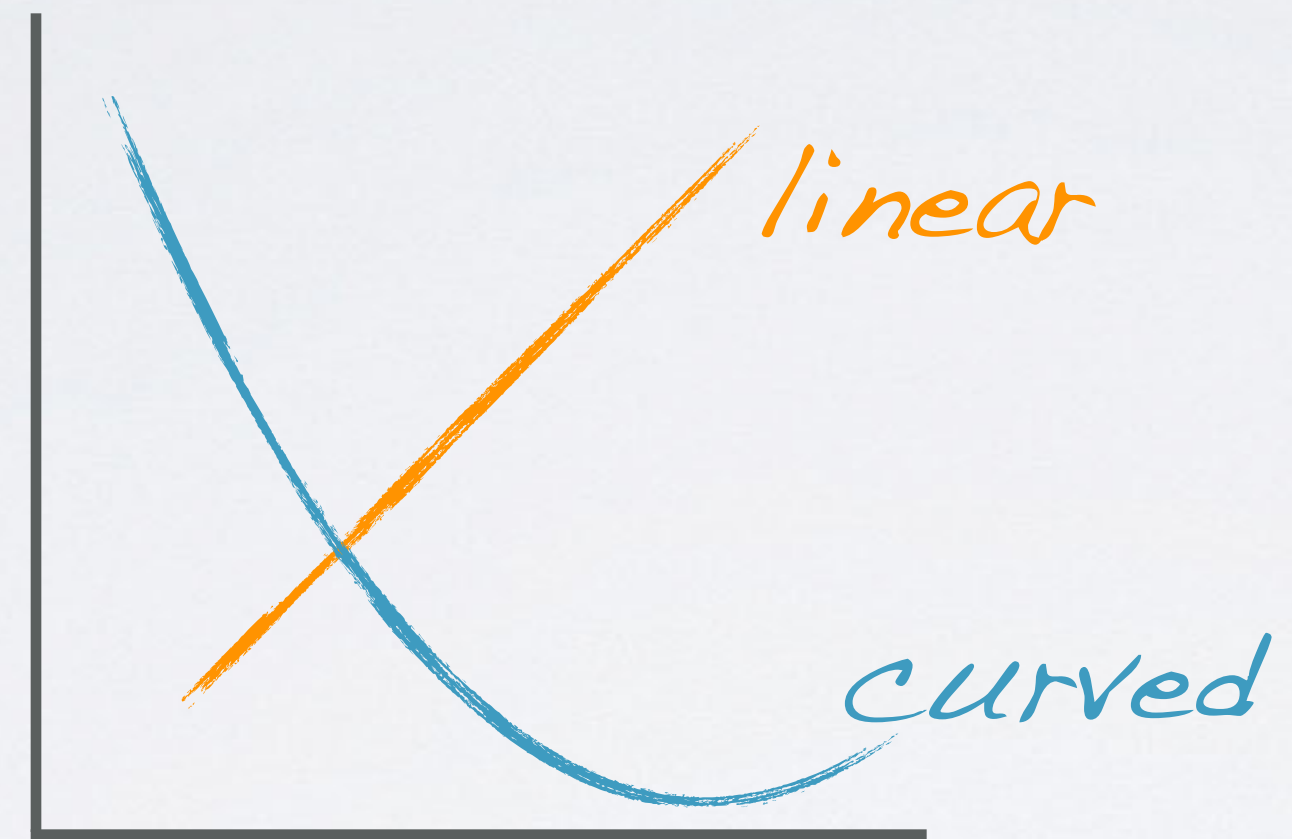
explanatory

evaluating the relationship

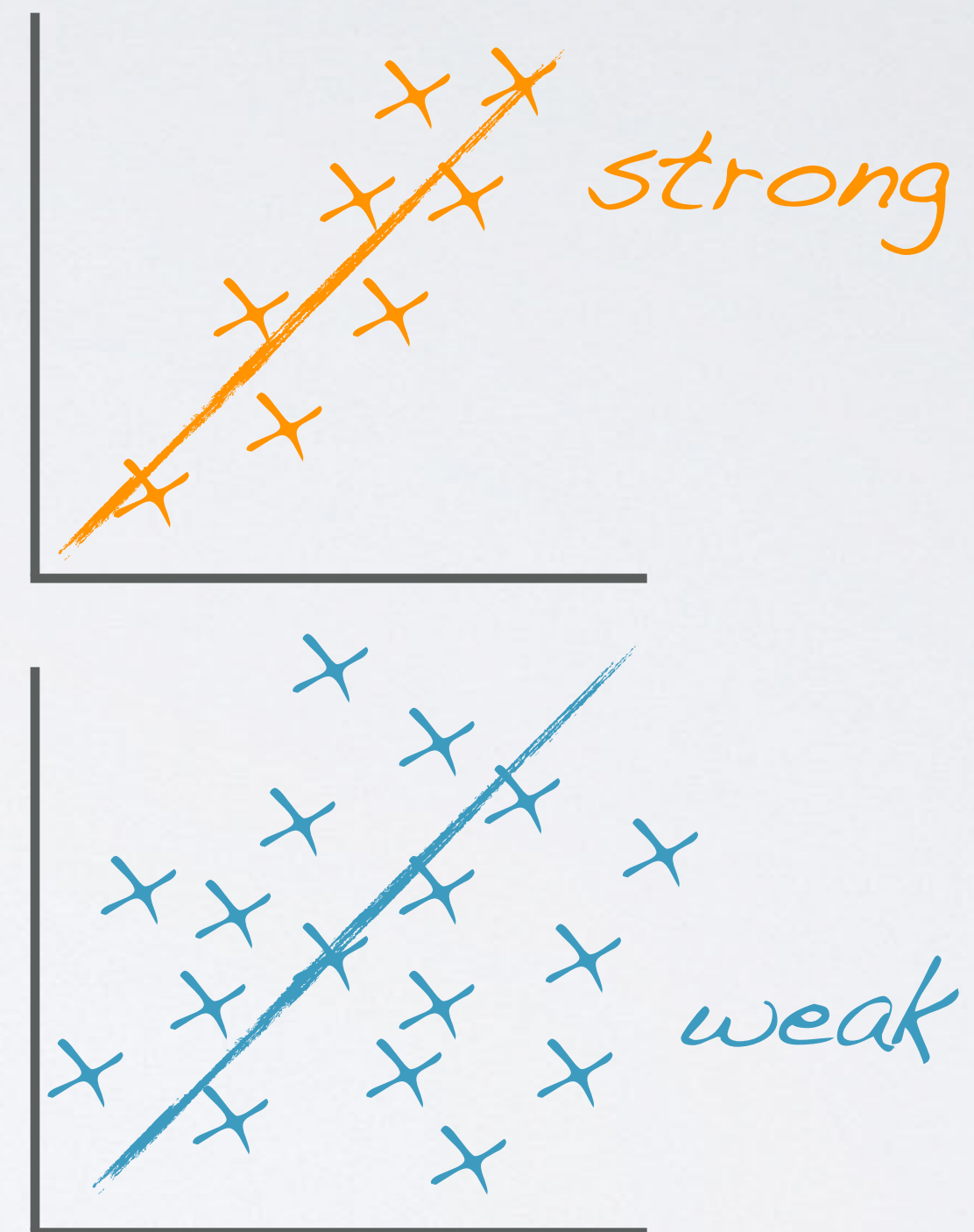
direction



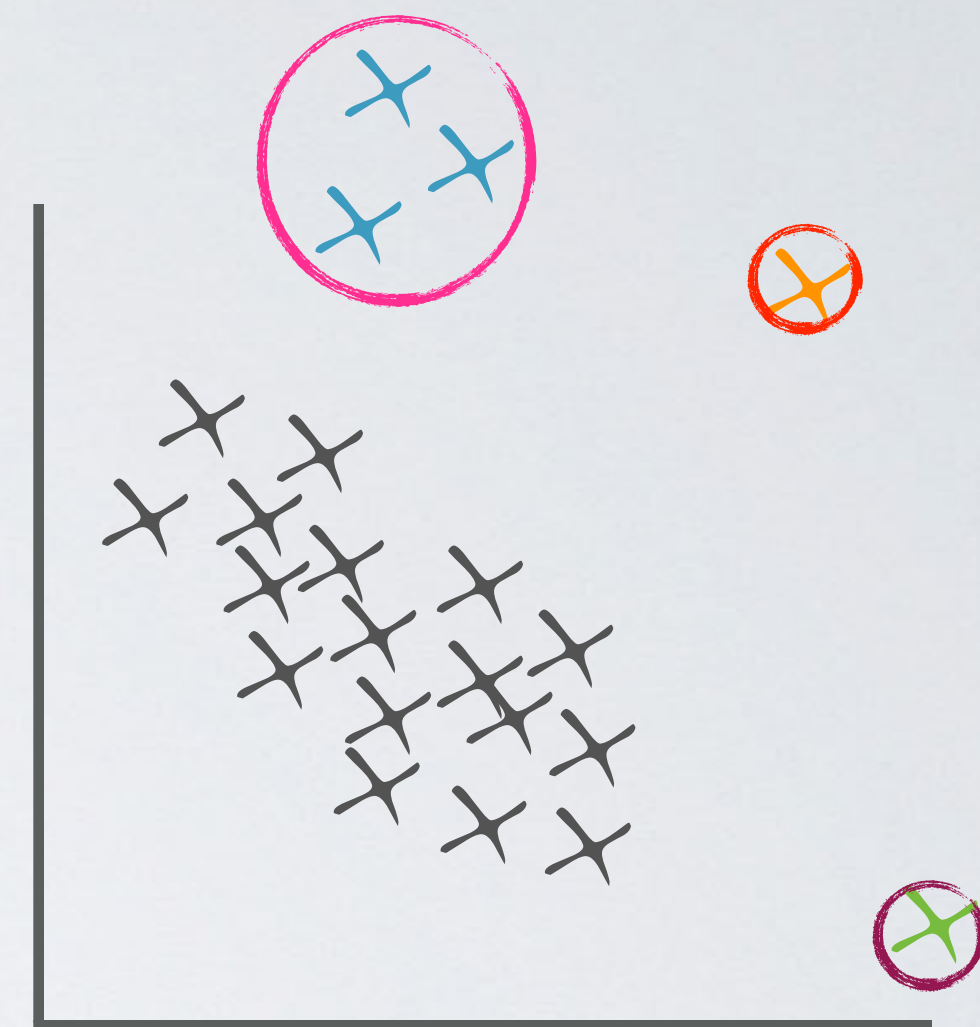
shape



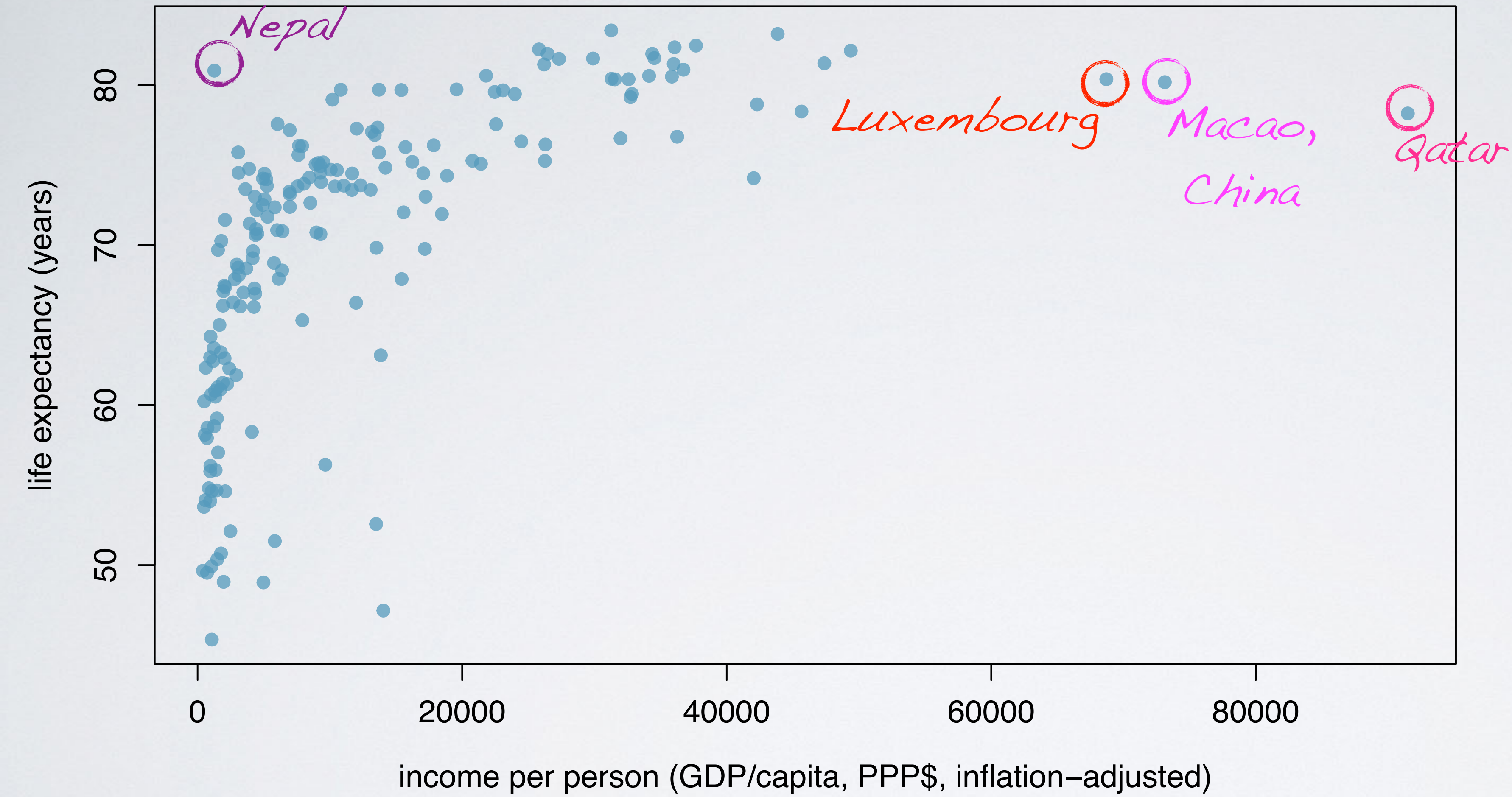
strength



outliers

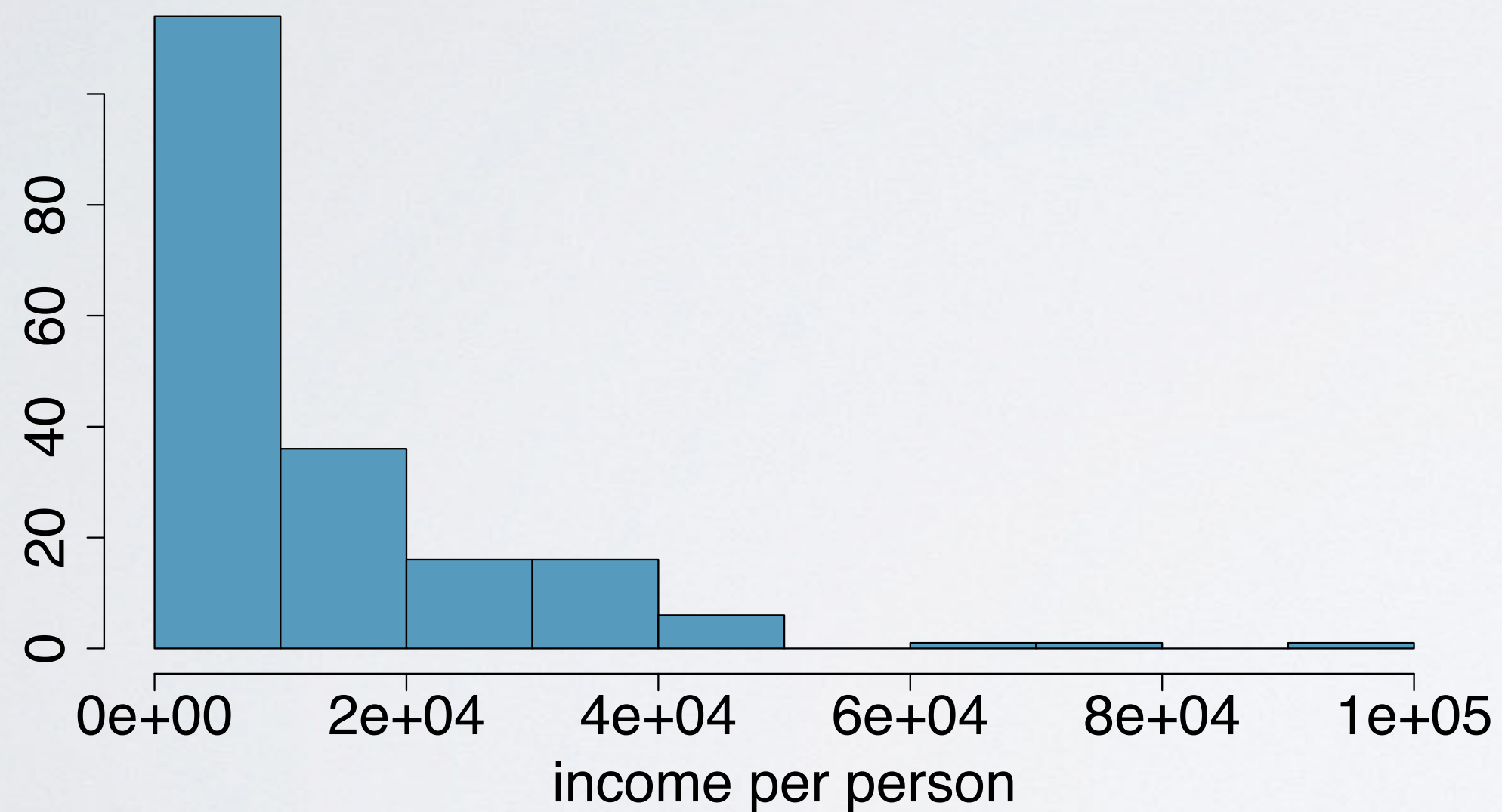
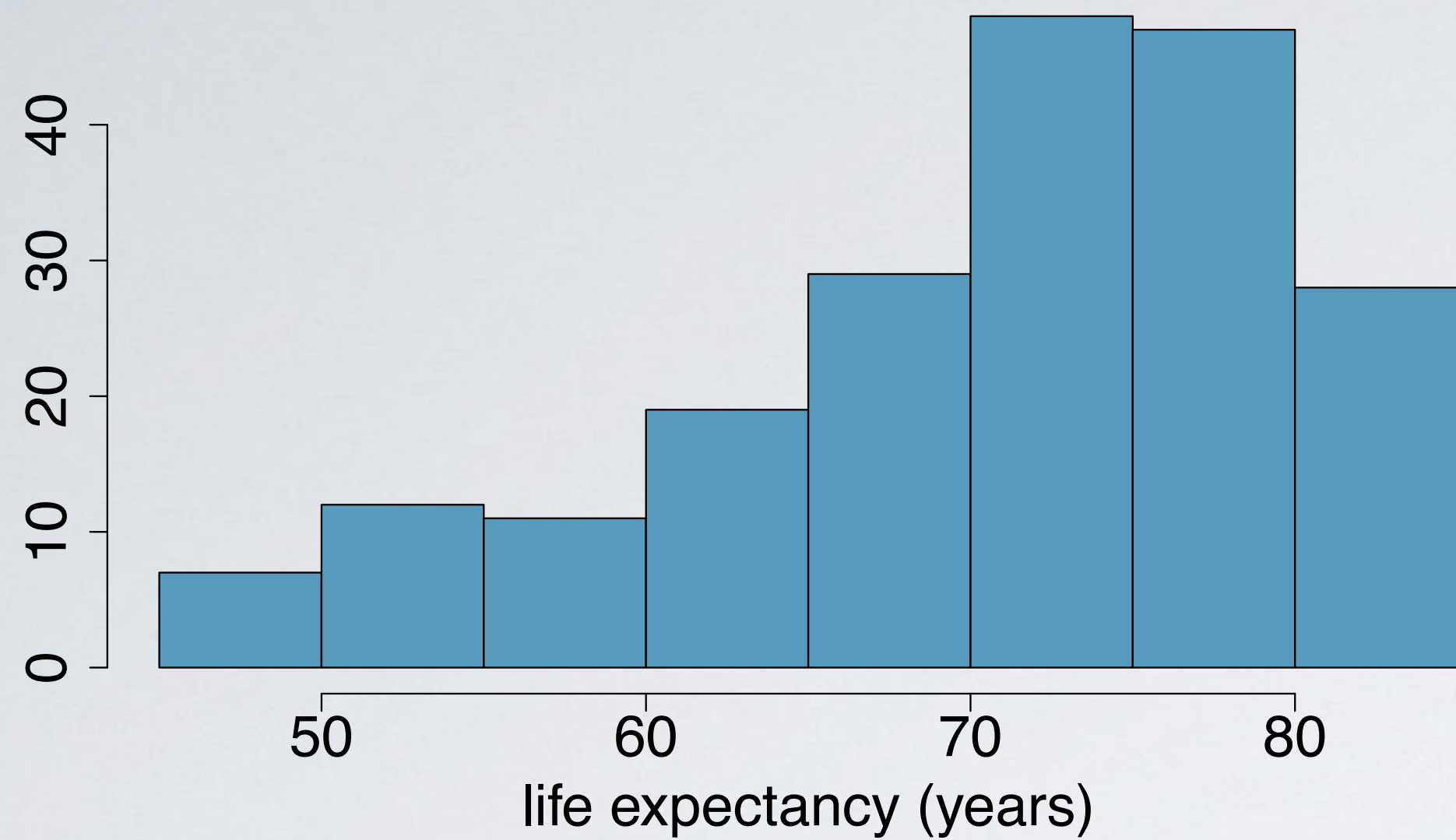


[revisit]



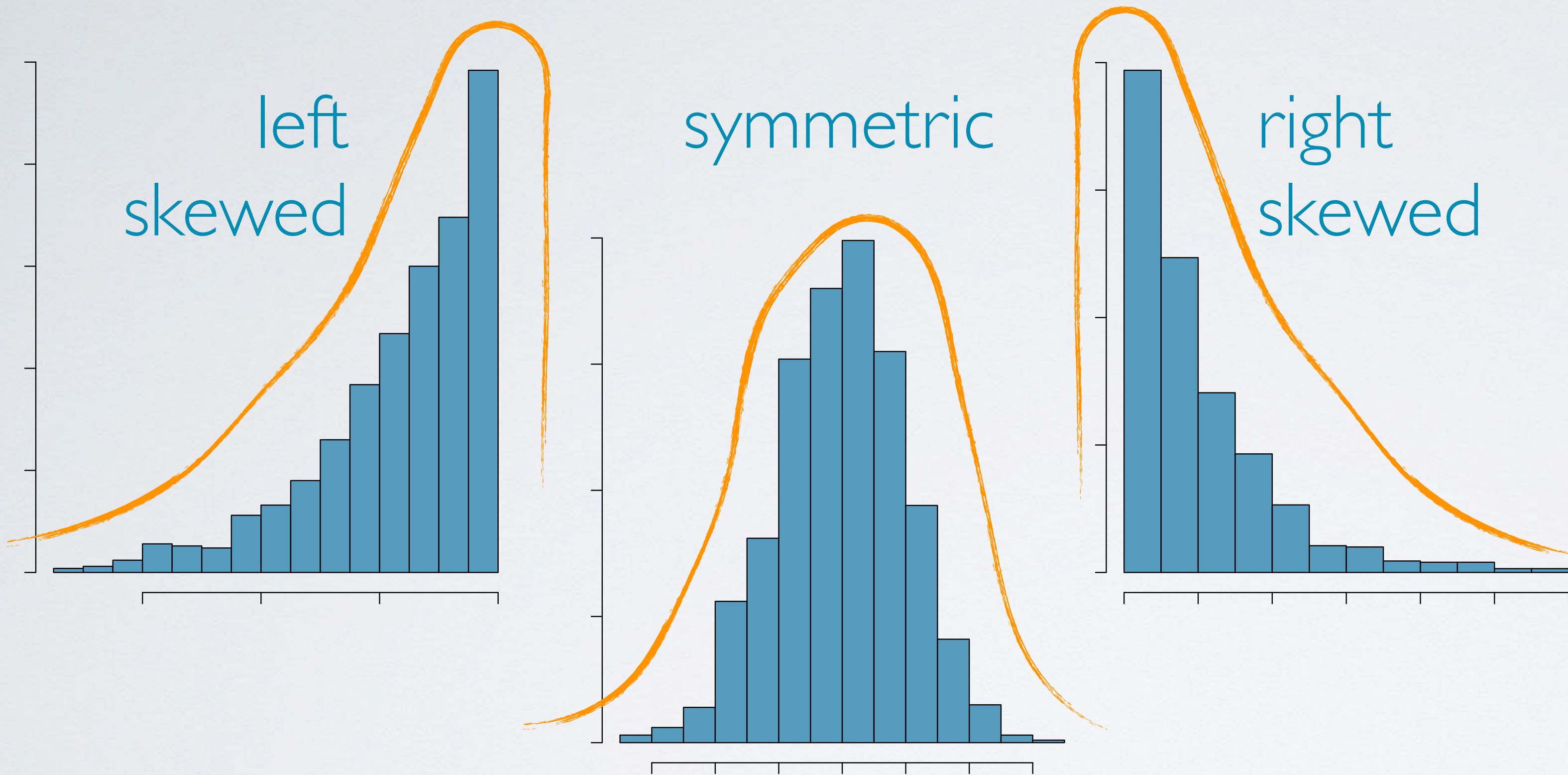
histogram

- ▶ provides a view of the data density
- ▶ especially useful for describing the shape of the distribution



skewness

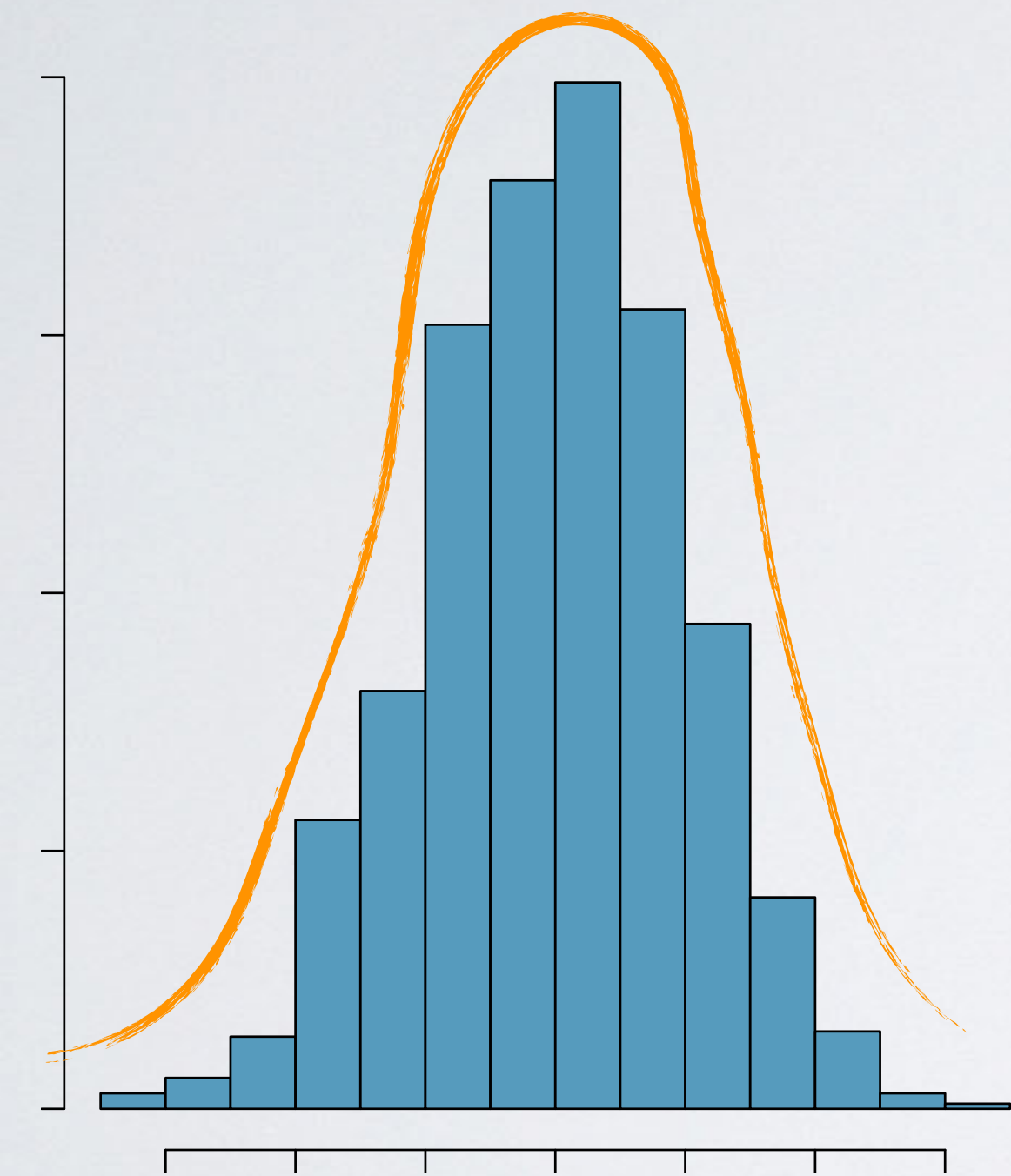
distributions are skewed to the side of the long tail



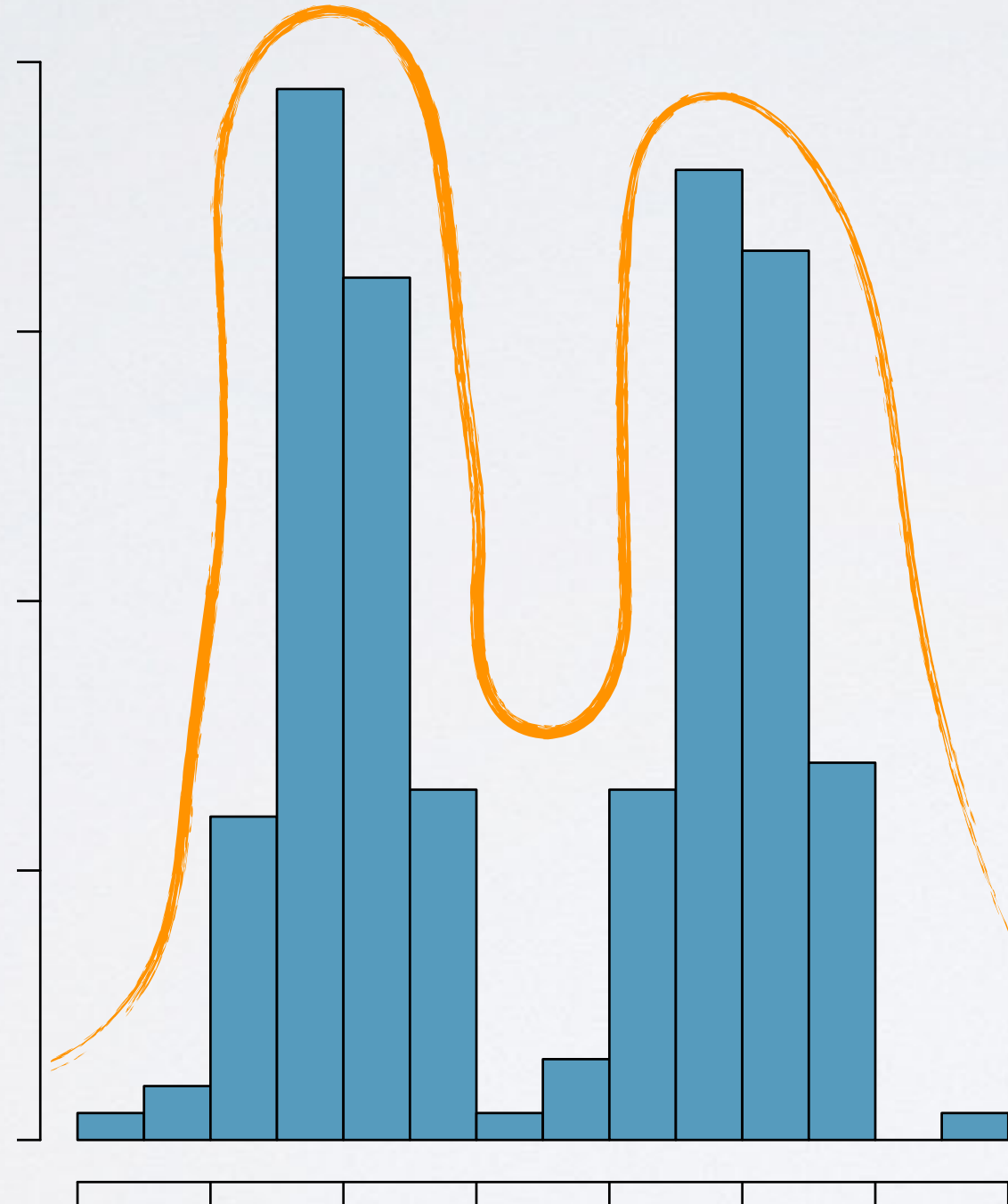
modality

distributions are skewed to the side of the long tail

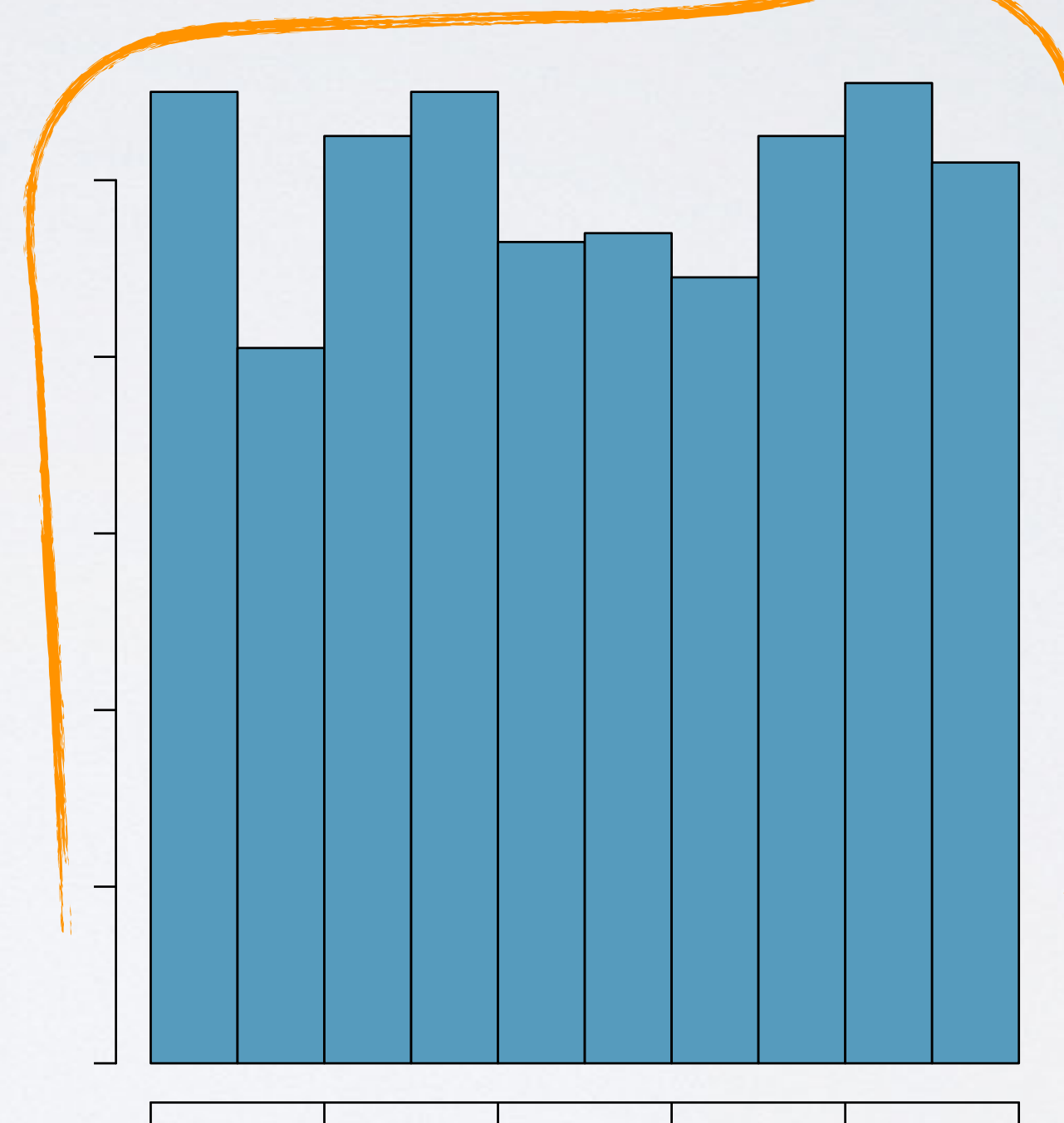
unimodal



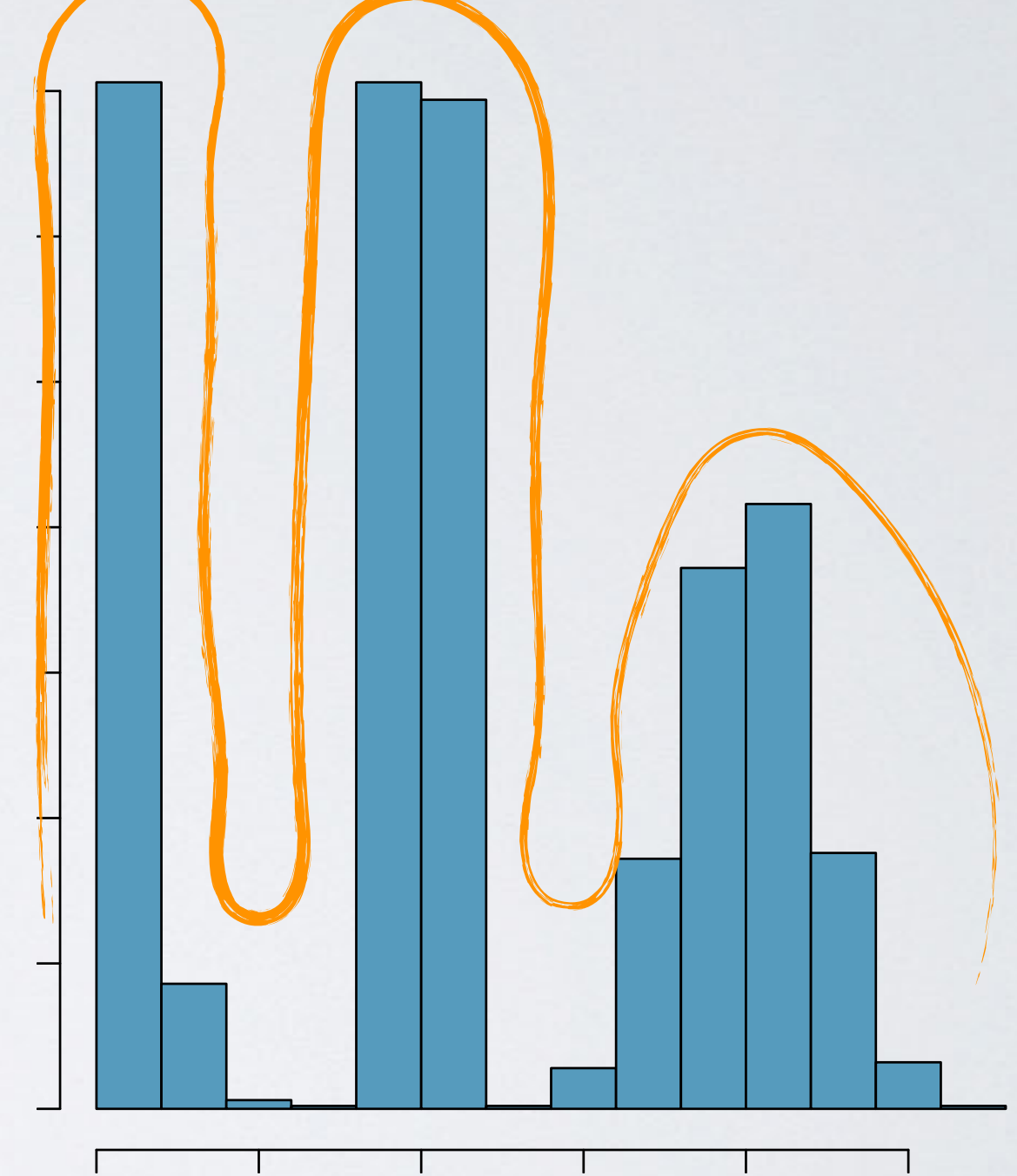
bimodal



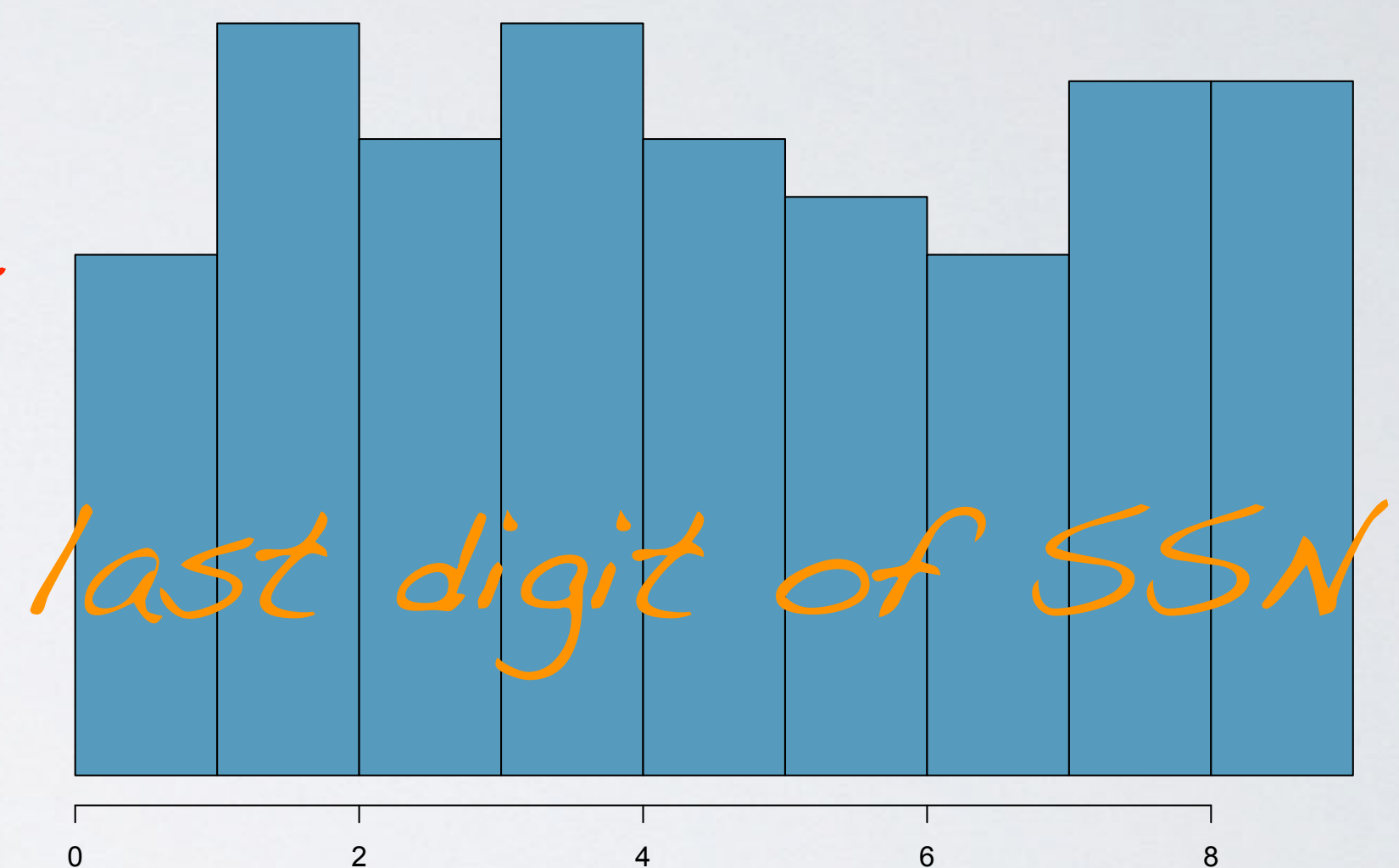
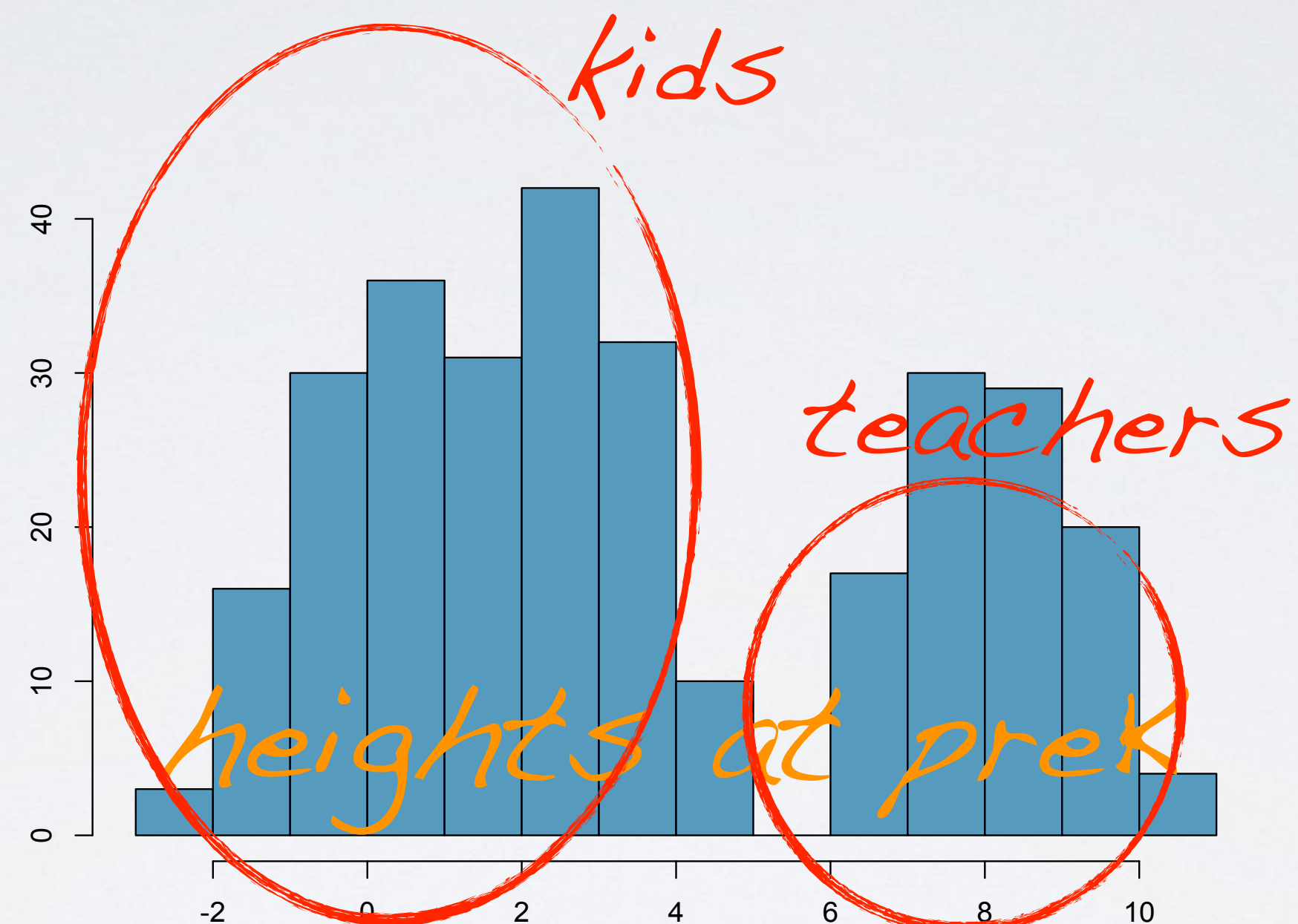
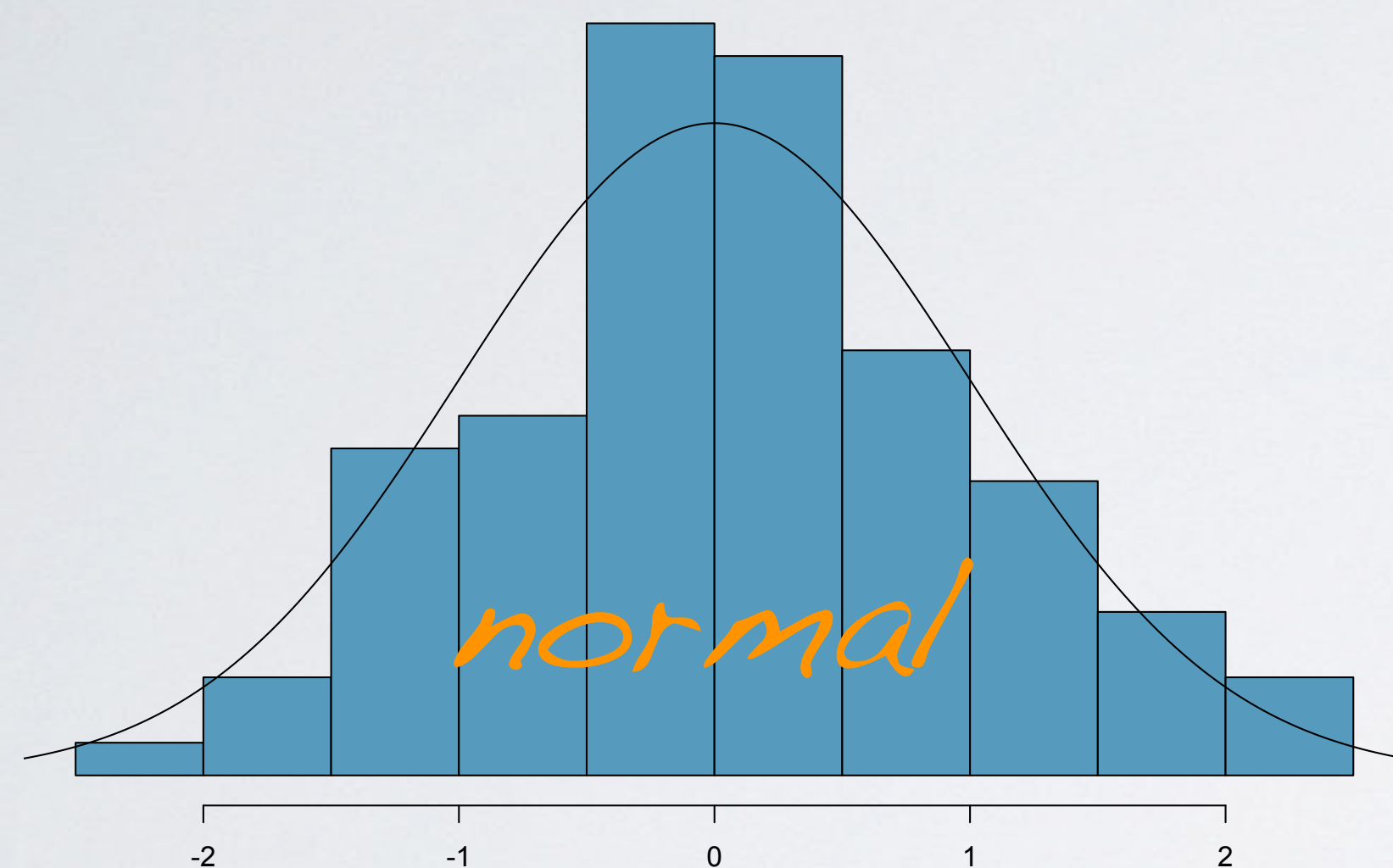
uniform



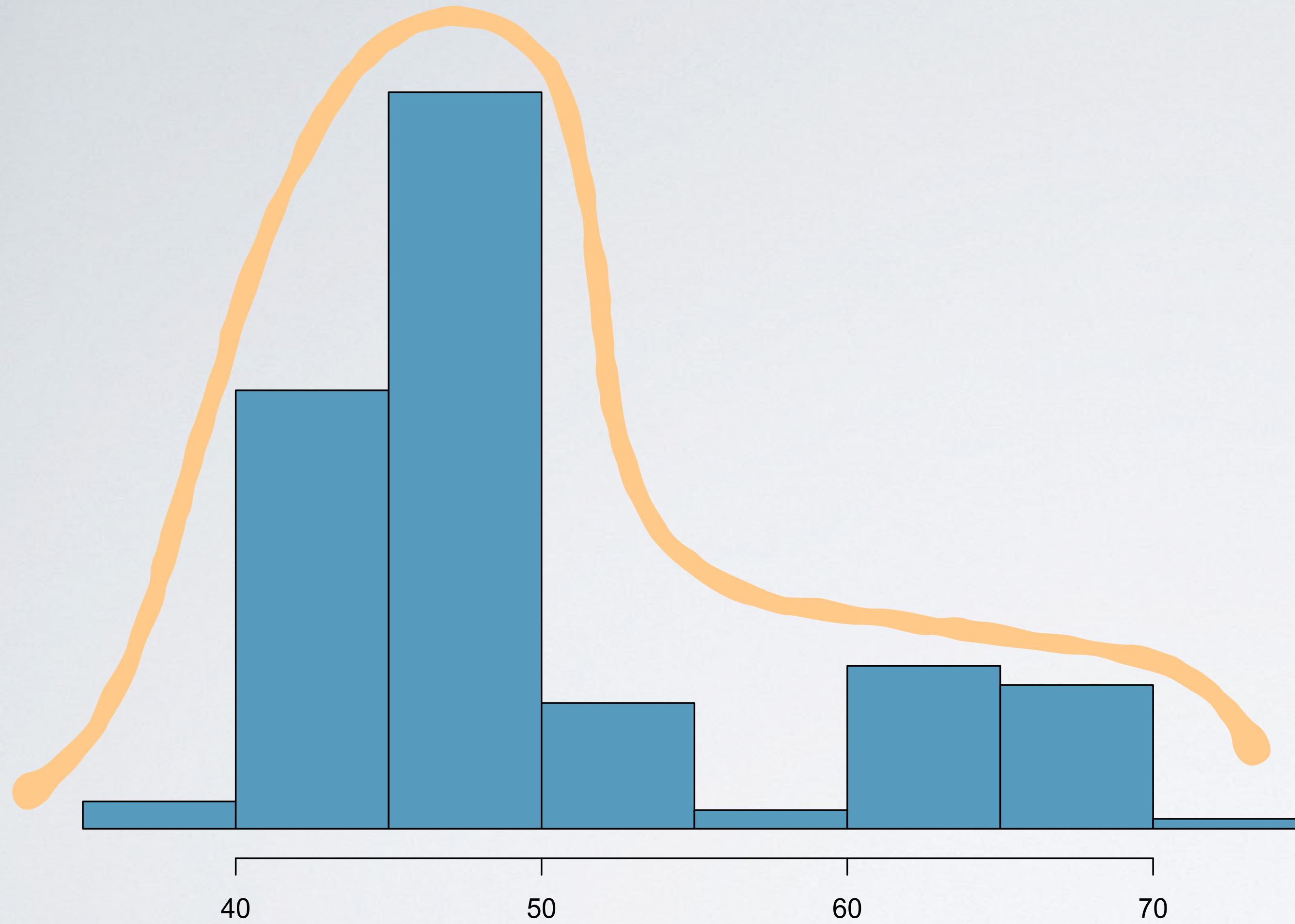
multimodal



modality (cont.)

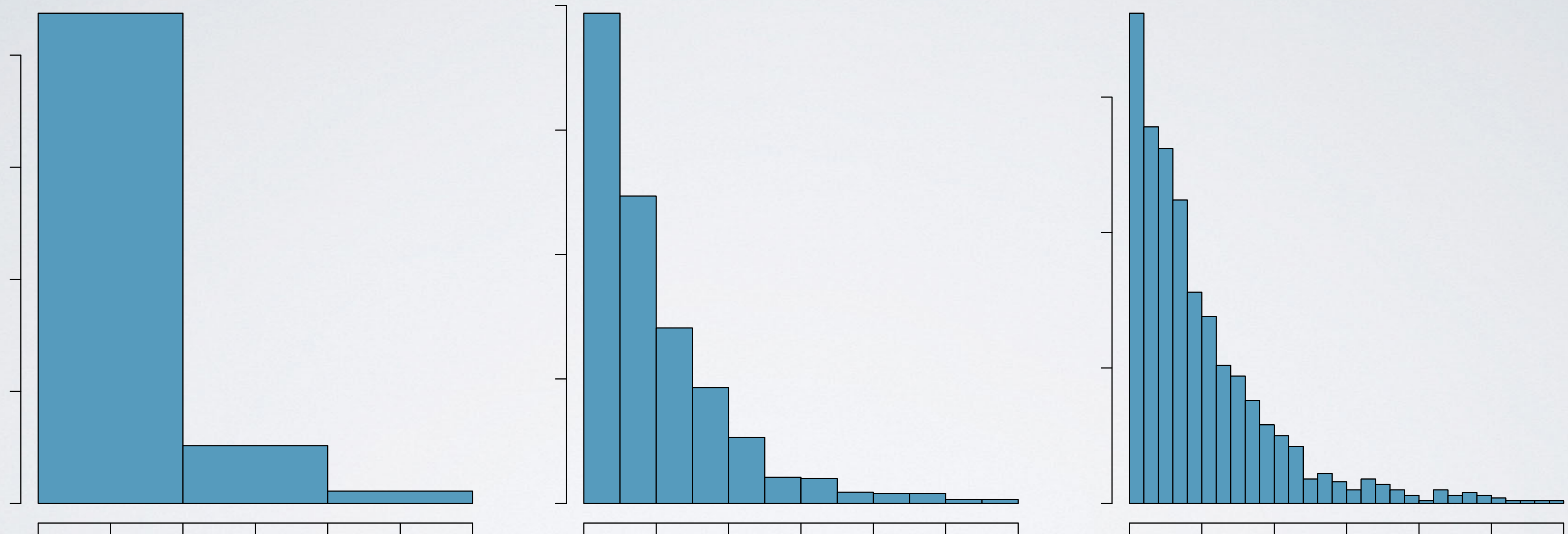


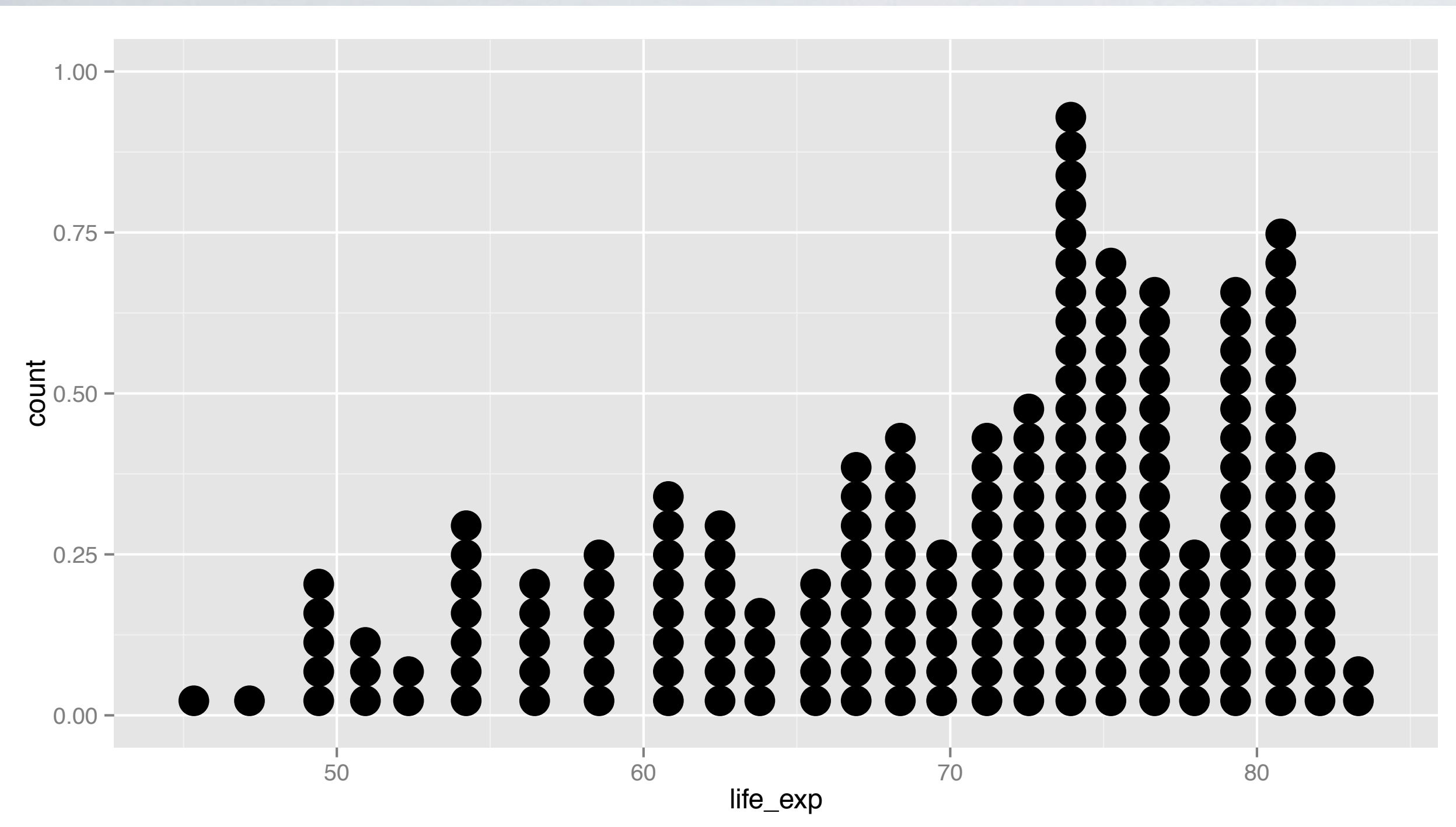
modality (cont.)



histogram & bin width

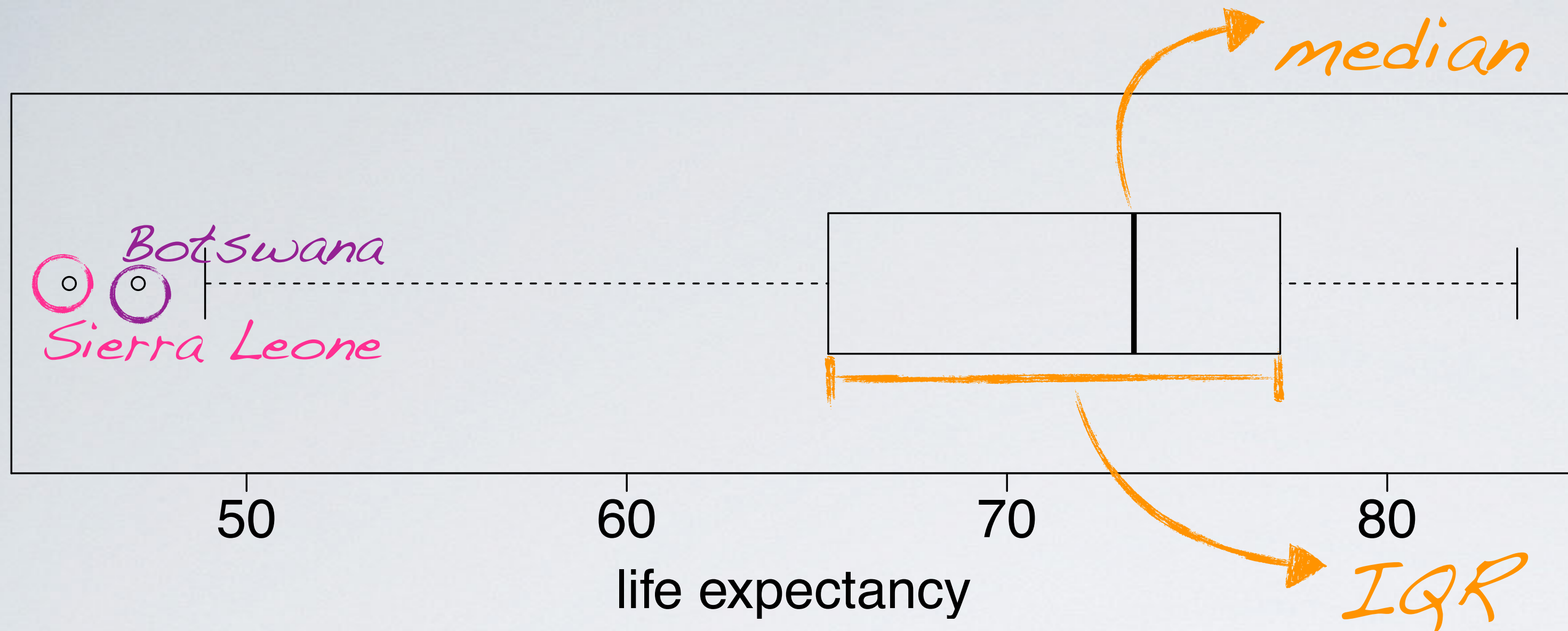
The chosen **bin width** can alter the story the histogram is telling.





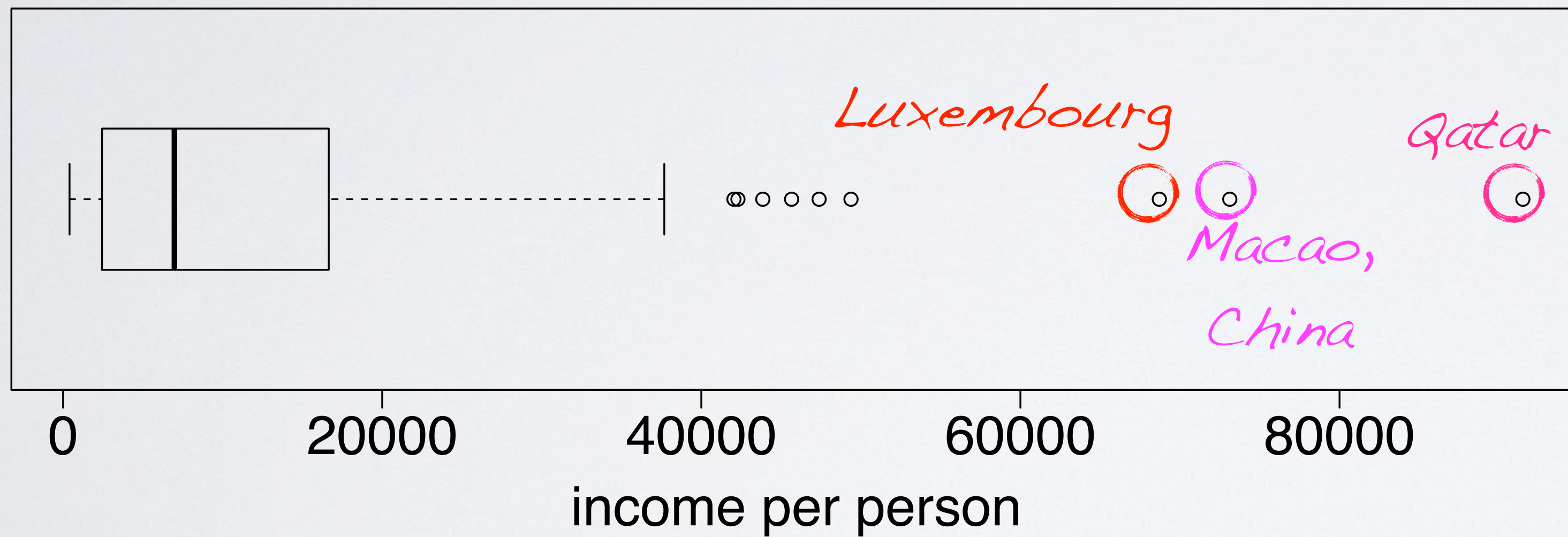
dotplot

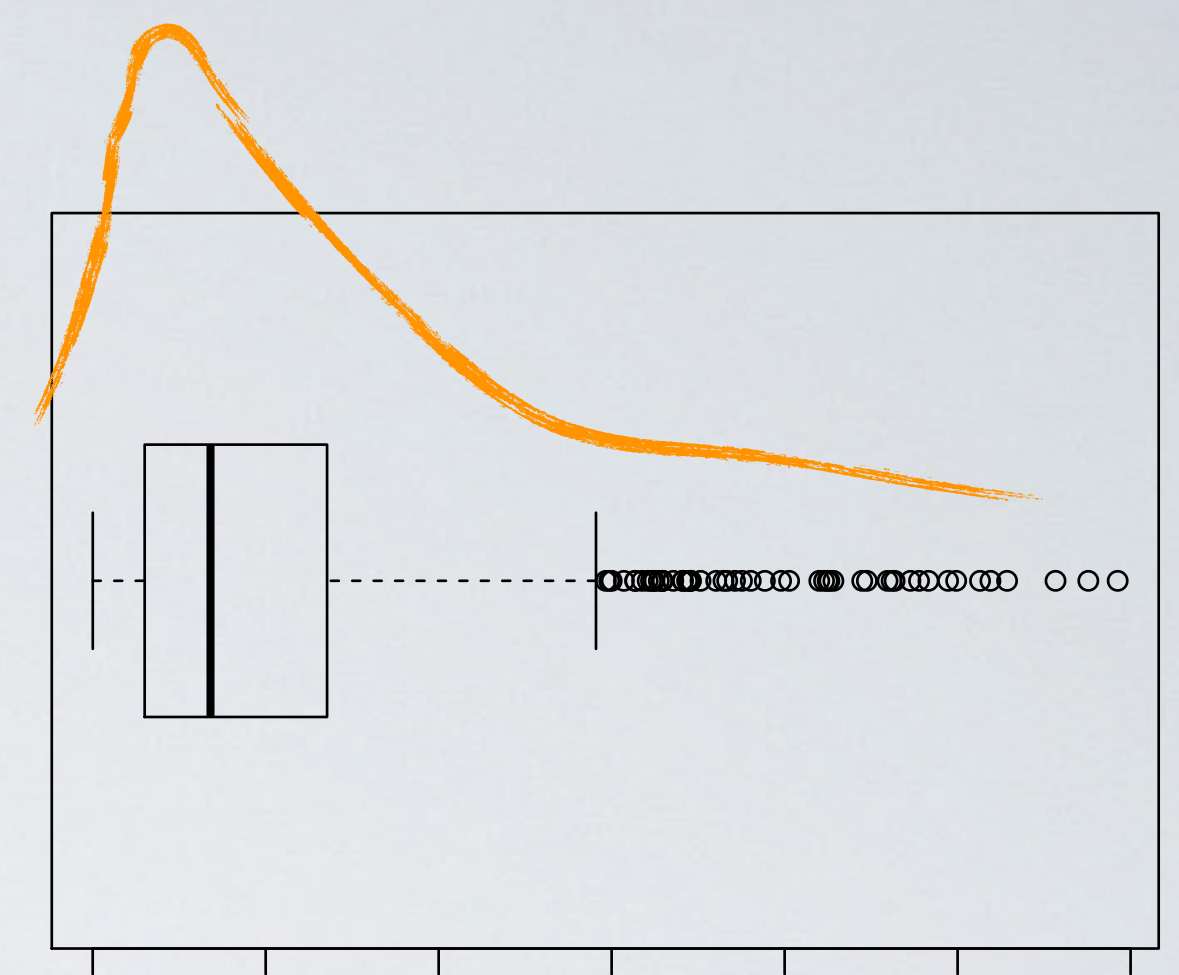
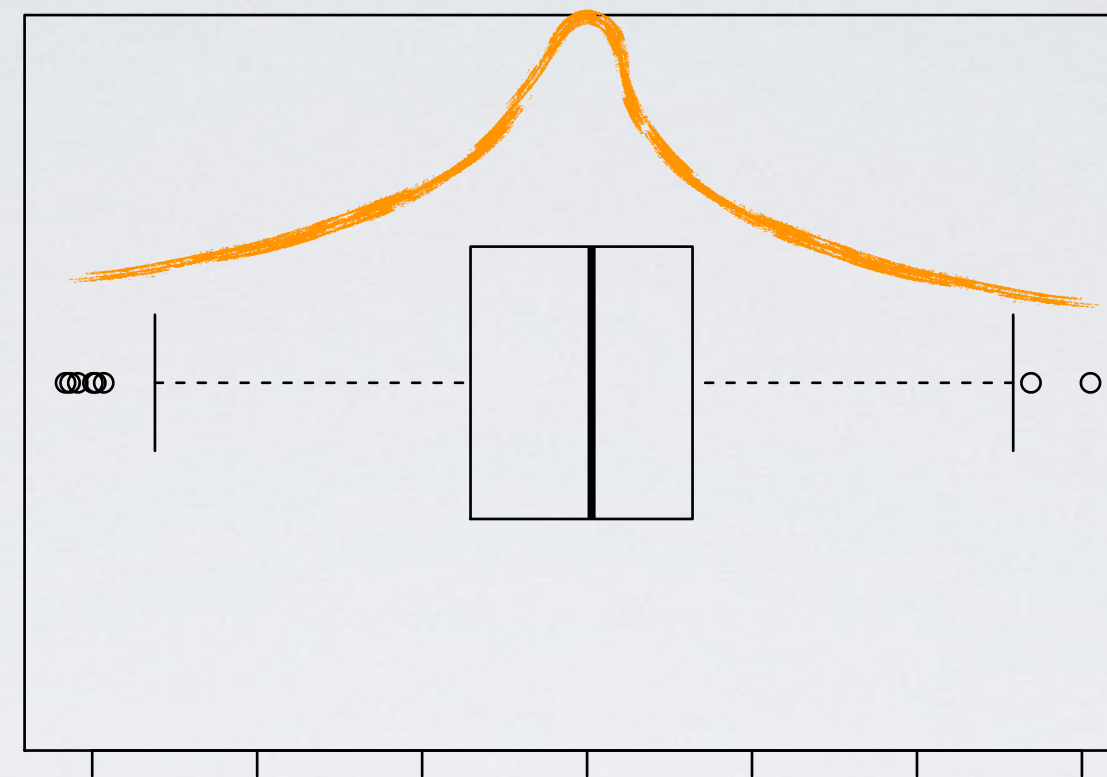
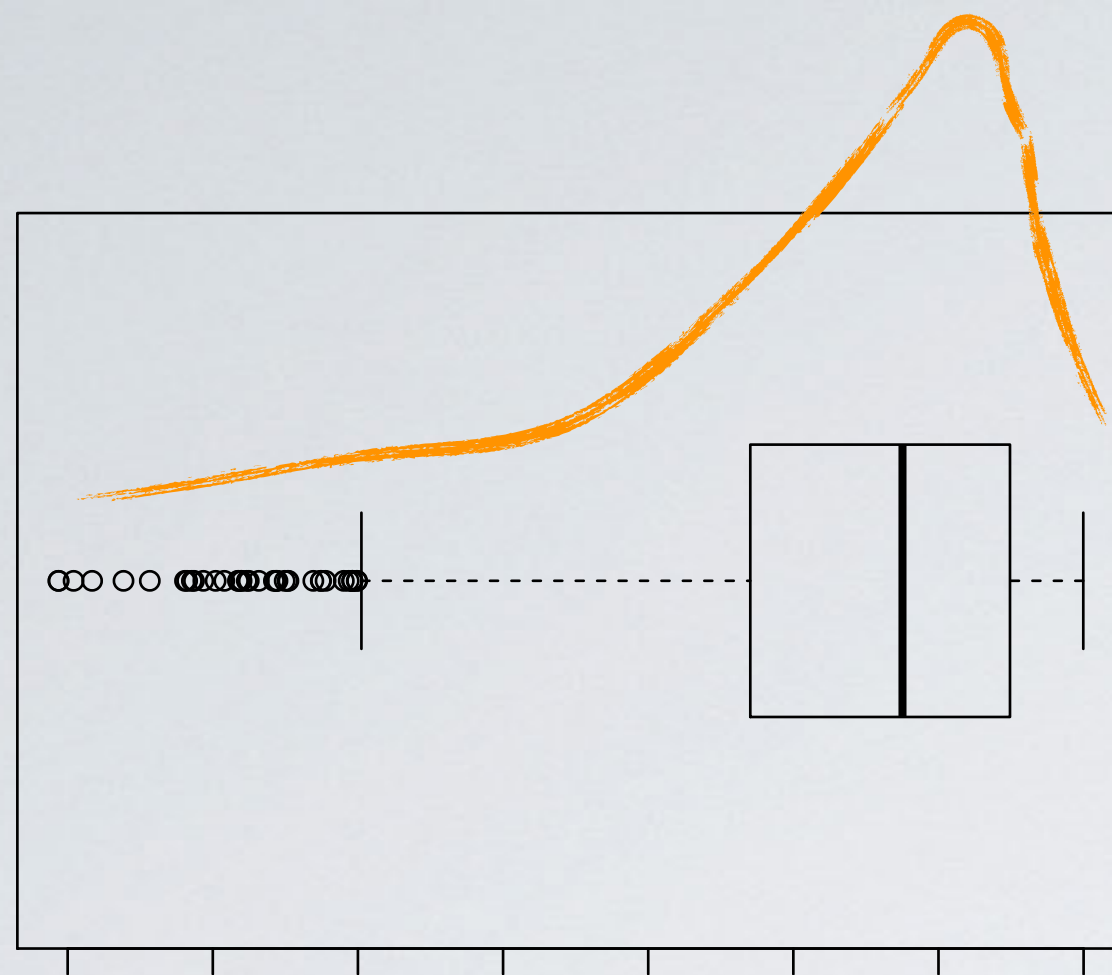
- ▶ useful when individual values are of interest
- ▶ can get busy as the sample size increases



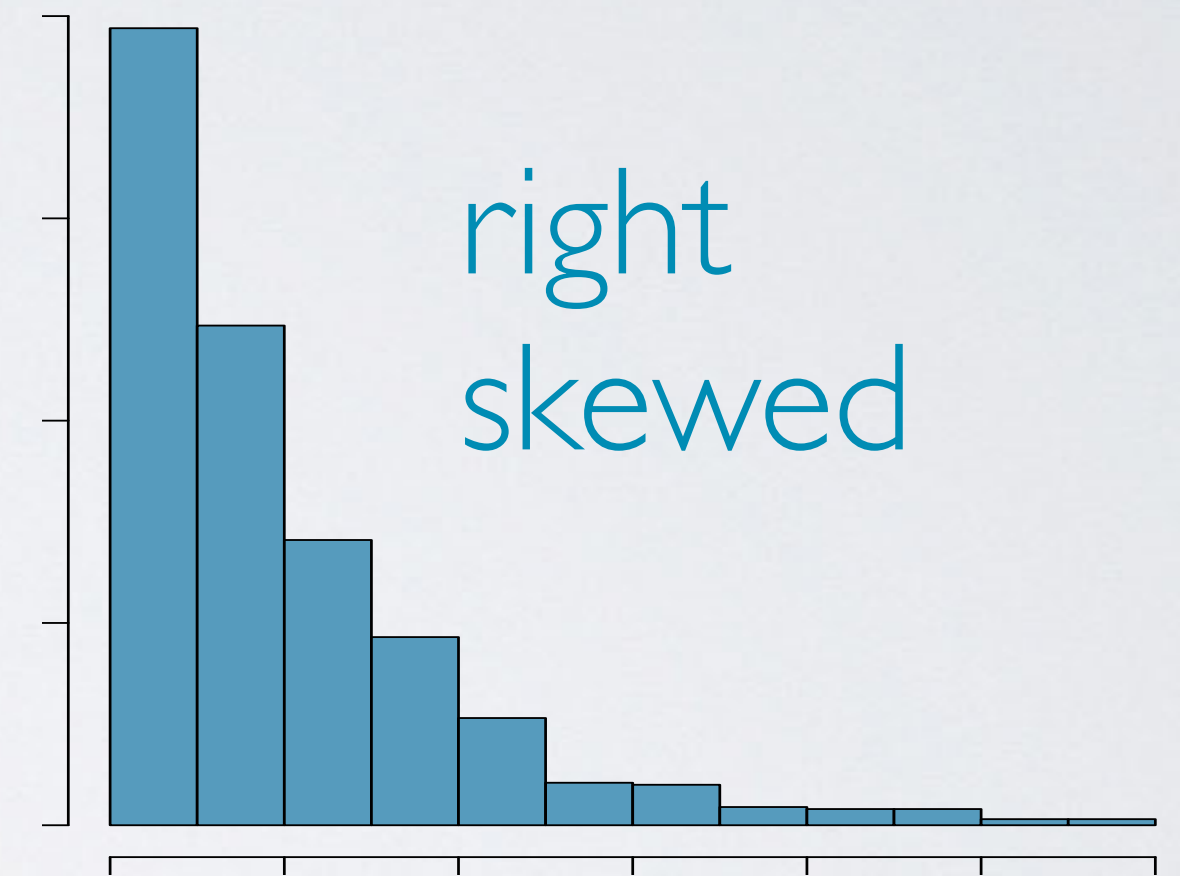
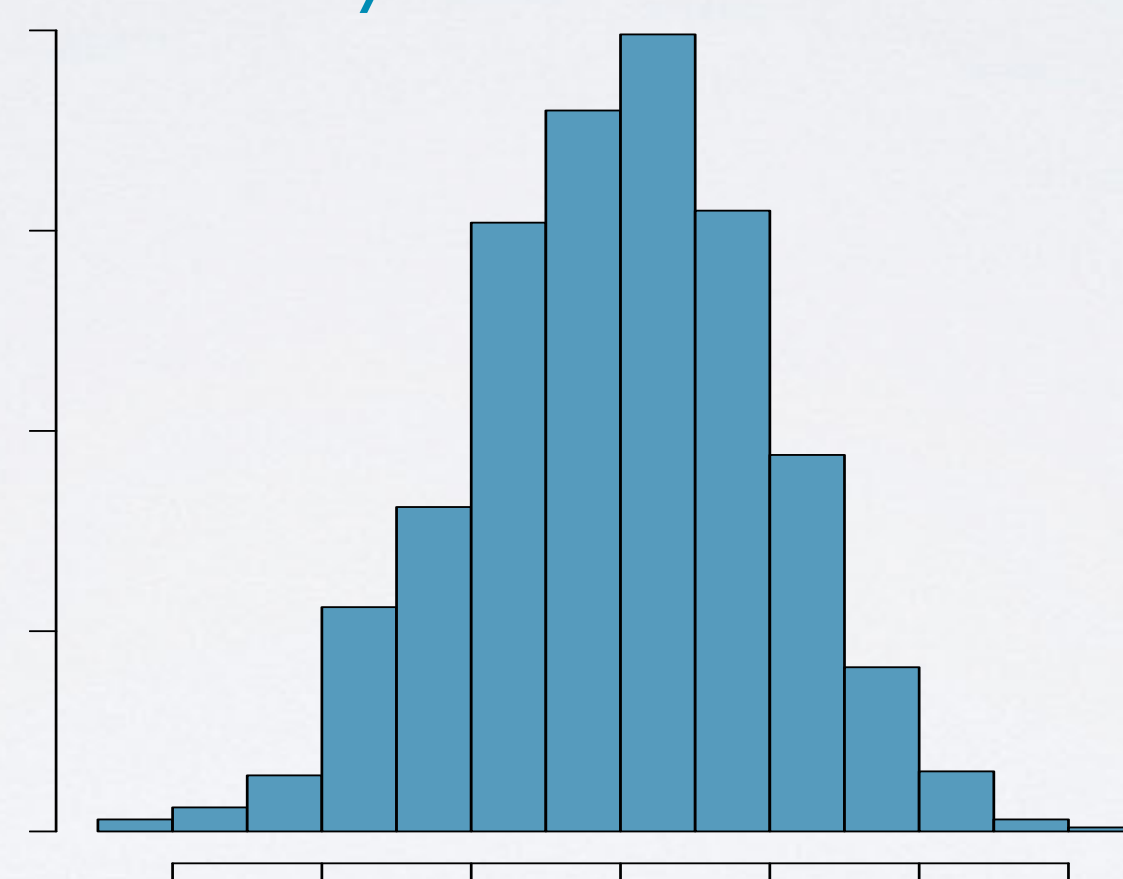
box plot

useful for highlighting outliers,
median, IQR





symmetric



intensity map

- Useful for highlighting the spatial distribution.

