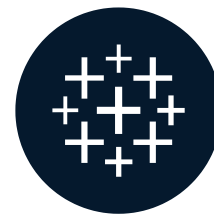# Welcome to the course!

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

Exploratory Data Analysis (EDA)

- Main characteristics of your data

- Spot extreme values

- Suggest hypotheses

- Assess assumptions

General goal: get an idea of the overall structure of your data

Univariate EDA

- Summary table

- Bar plot

- Histogram

- Box plot

Tables & bar plots

Visualize the distribution of a single, categorical variable
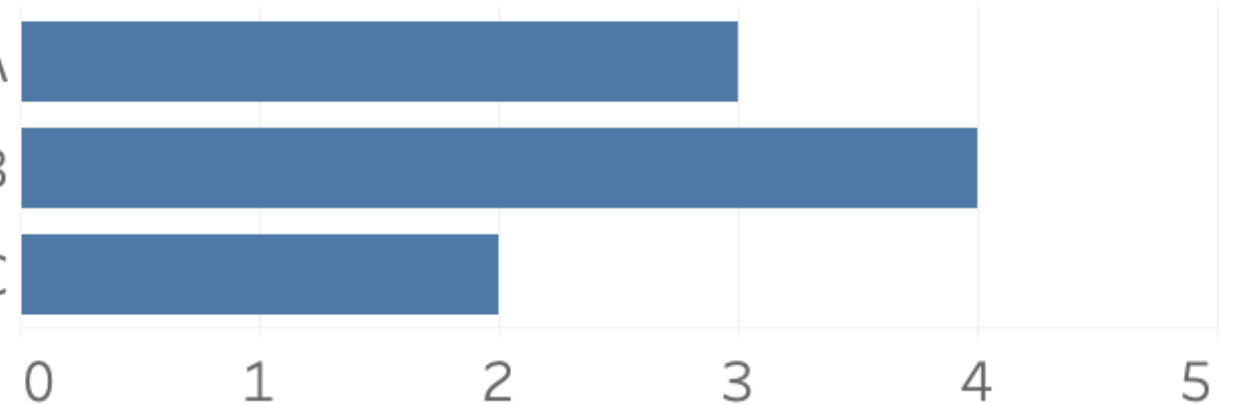
Category

A

B

A

B

B

C

A

B

C

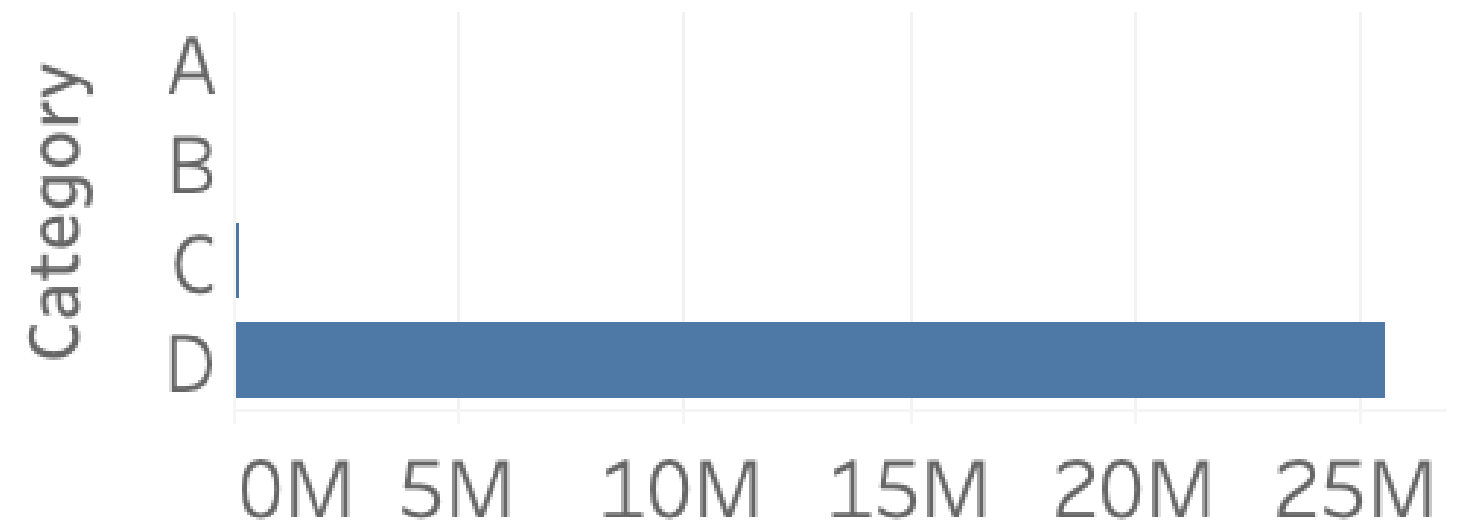| Category	
A	3
B	4
C	2

When to use a table vs. a plot

- Focus is on individual values (snapshot) and not on trends

- Dataset contains few values

- Small difference between values is crucial

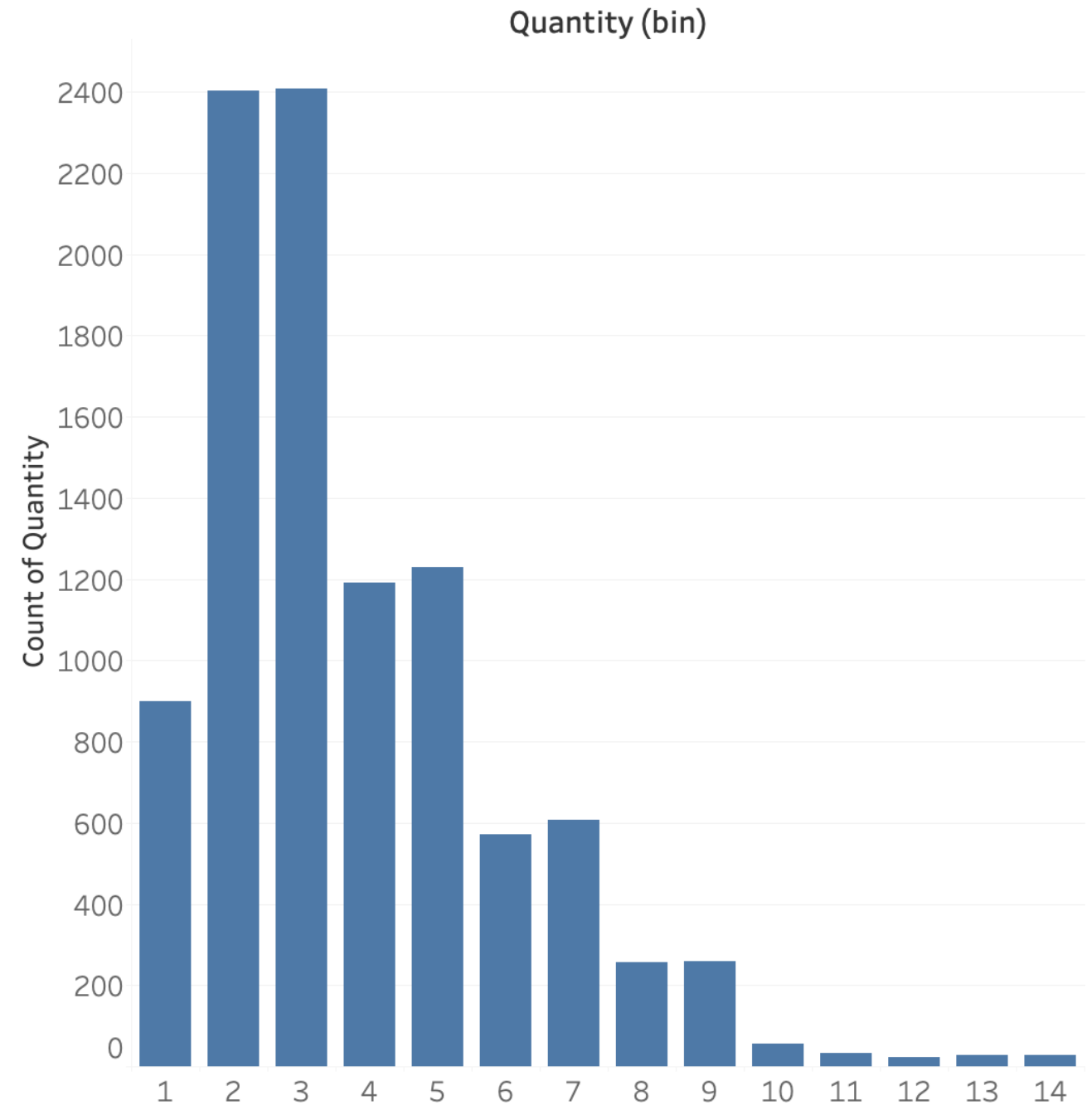- Data is presented in a non-interactive way

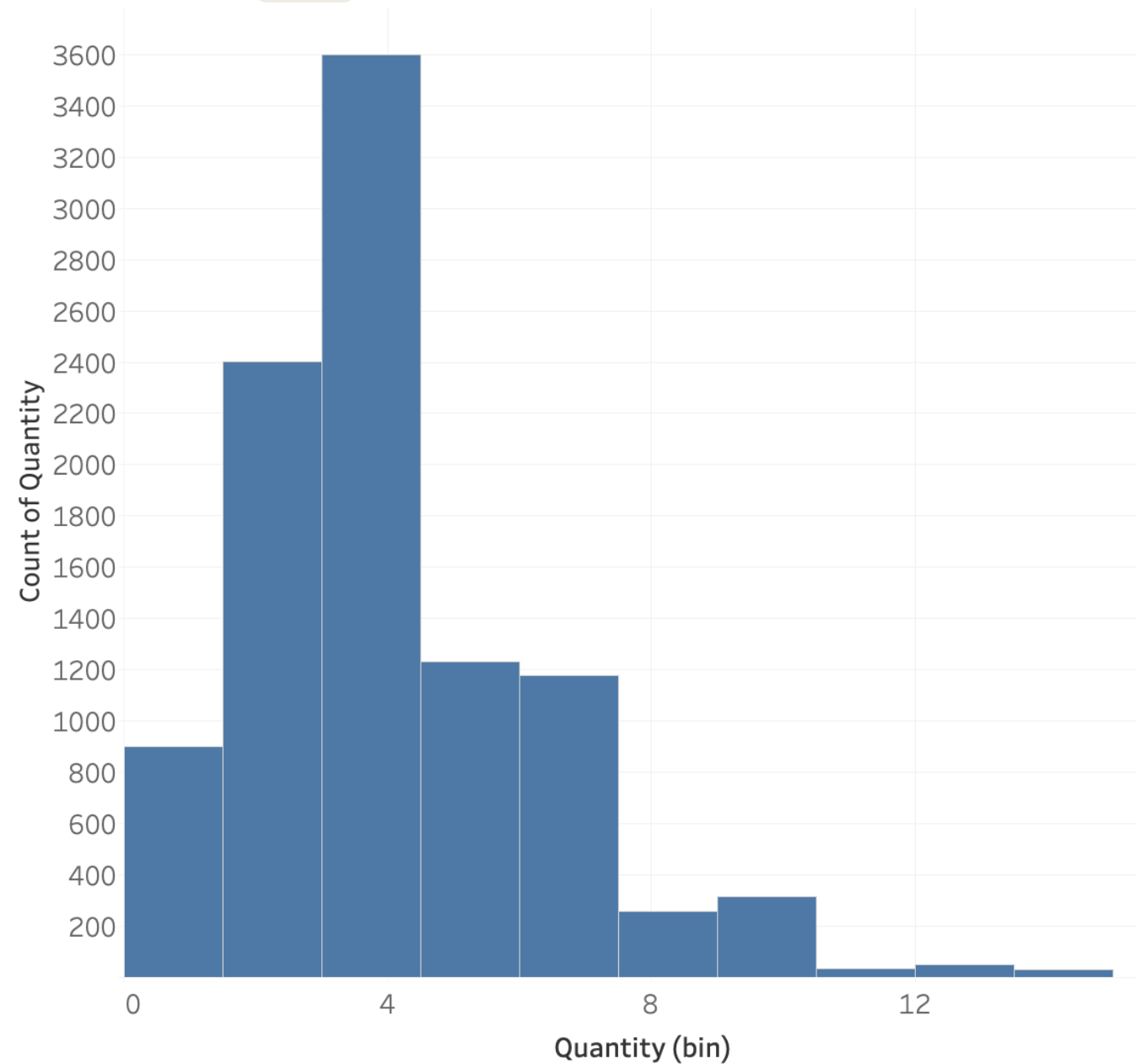| Category	
A	20
B	400
C	160.000
D	25.600.000

Histograms

Visualize the distribution of a single, continuous variable

- Lowest/highest value
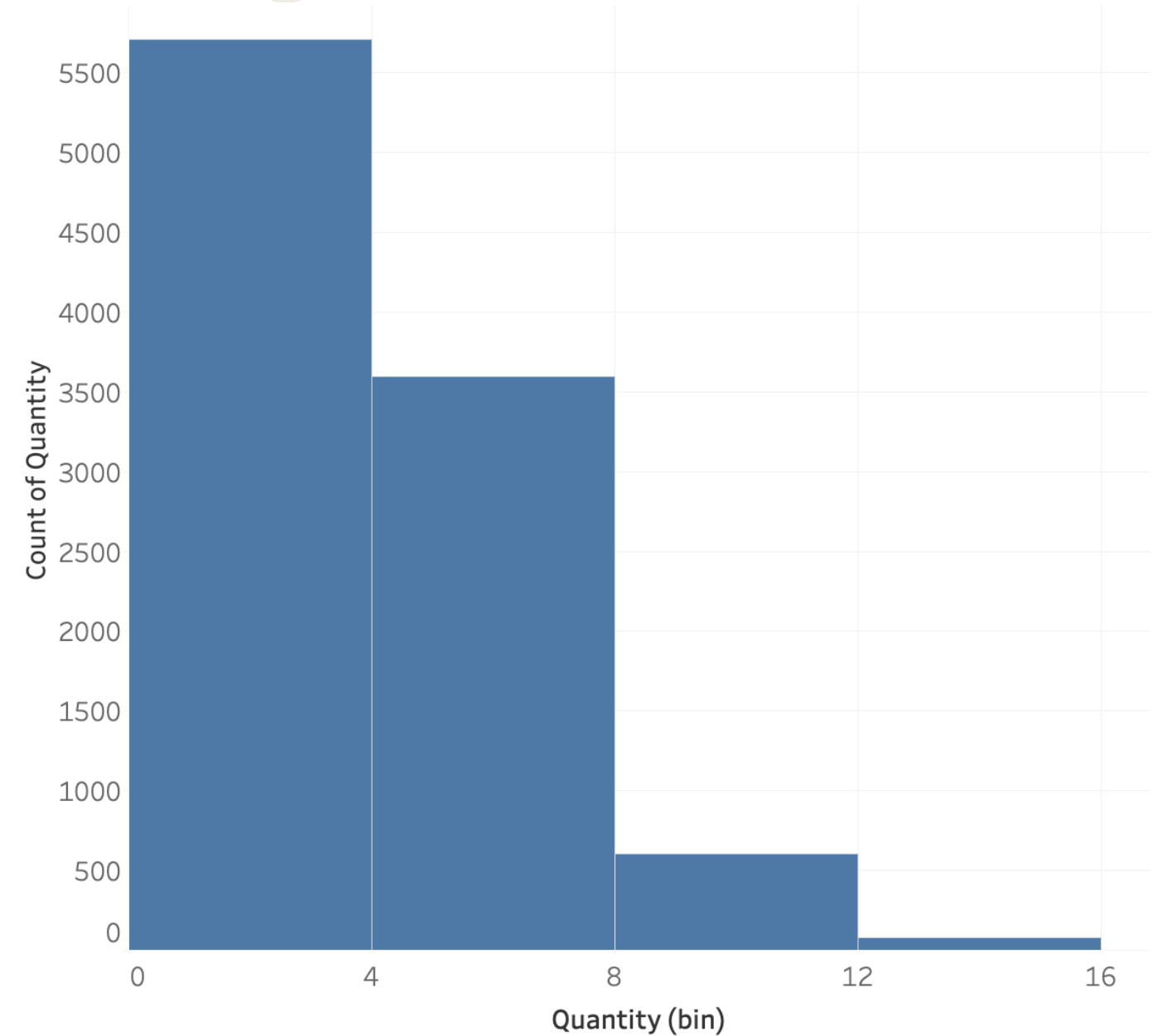
- Most common value(s)

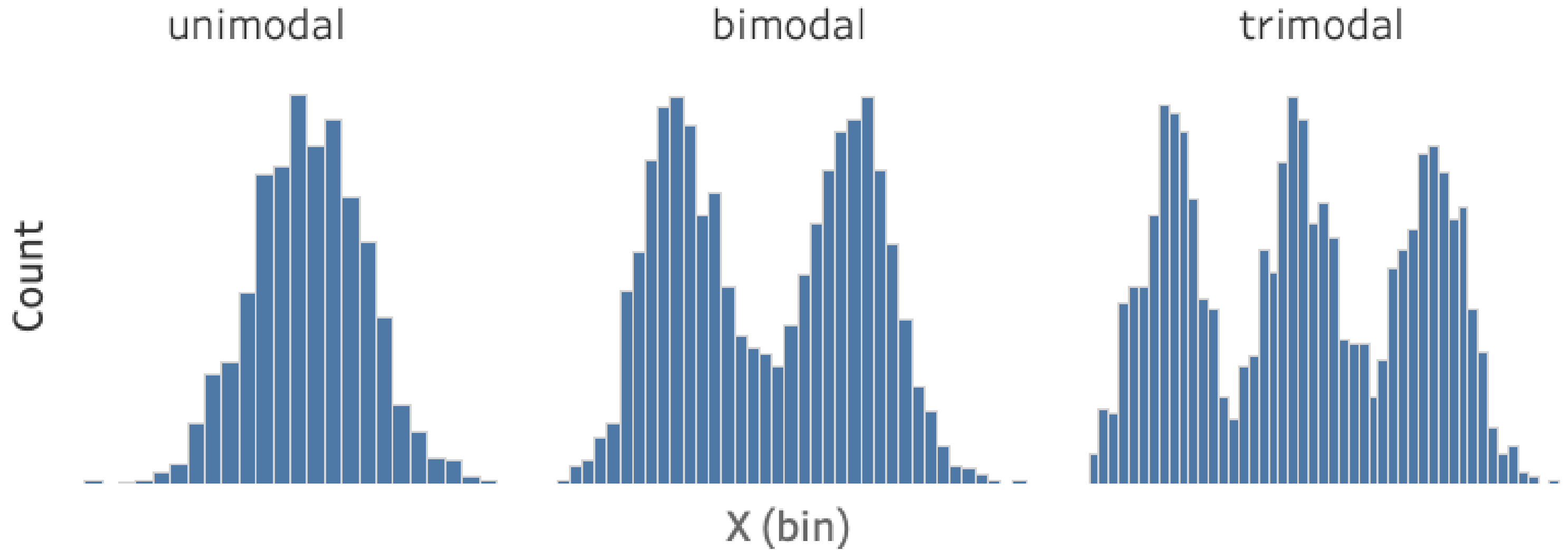- Splitting variable in bins

Size of bins

Binwidth = `1.5`

Binwidth = `4`

Modality

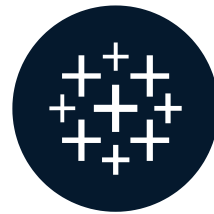unimodal　　　　　　　　bimodal　　　　　　　　trimodal

Count

X (bin)

Mode: most occurring value

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU

EDA in Tableau: tables and bar plots

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

Tableau Time!

STATISTICAL TECHNIQUES IN TABLEAU

EDA in Tableau: histograms

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

datacamp

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU

Box plots and distribution characteristics

STATISTICAL TECHNIQUES IN TABLEAU
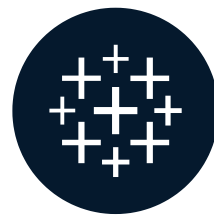
Maarten Van den Broeck
Content Developer at DataCamp

Box plots

Visualize the distribution of a single, continuous variable
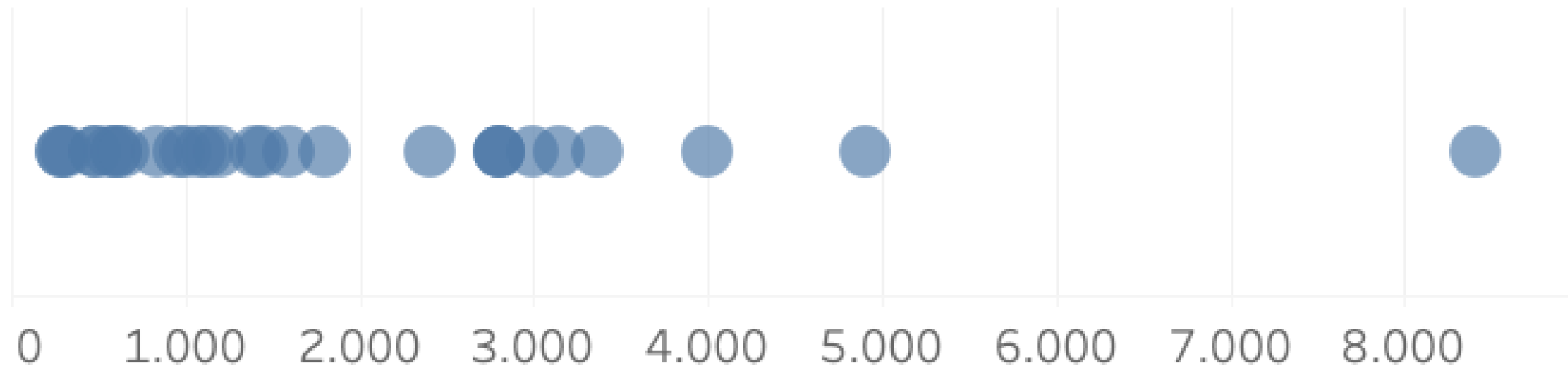
Box plots

Visualize the distribution of a single, continuous variable
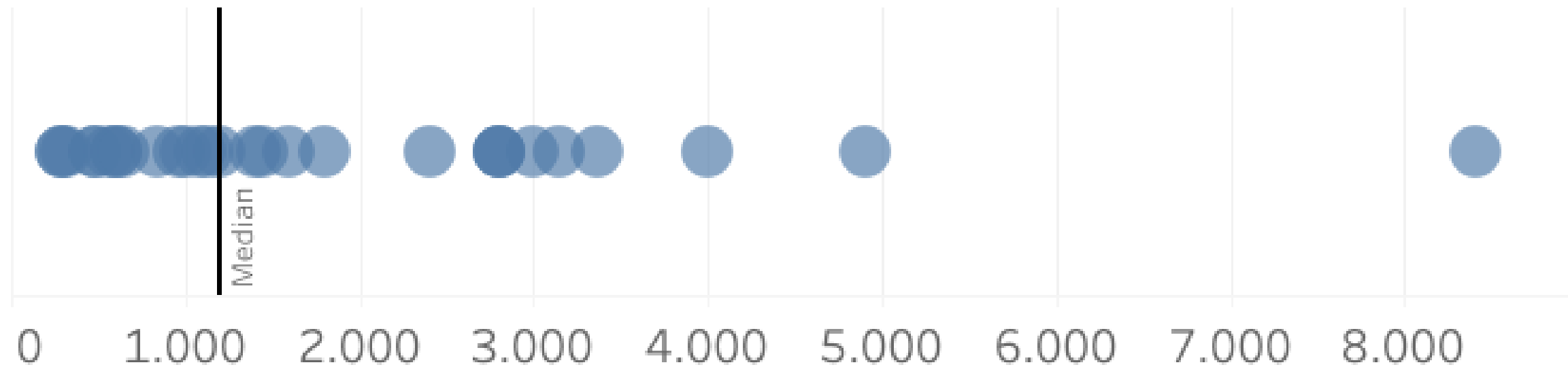
Box plots

Visualize the distribution of a single, continuous variable
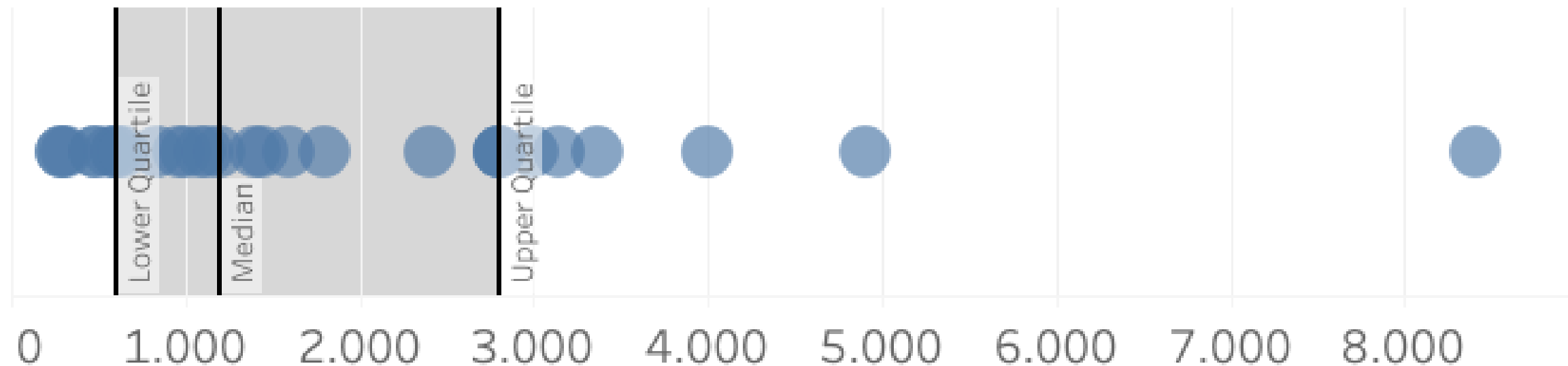
- Distance between lower quartile and upper quartile is the interquartile range (IQR)

Box plots

Visualize the distribution of a single, continuous variable
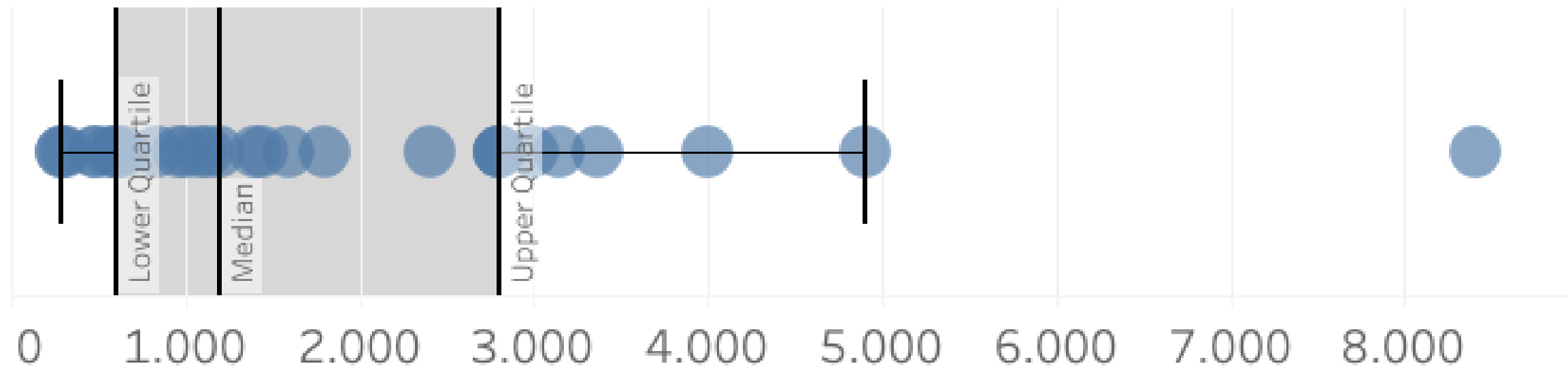
- Distance between lower quartile and upper quartile is the interquartile range (IQR)

- Whiskers: length of 1.5 x IQR

- Outlier: extreme value outside whiskers

When to use a box plot
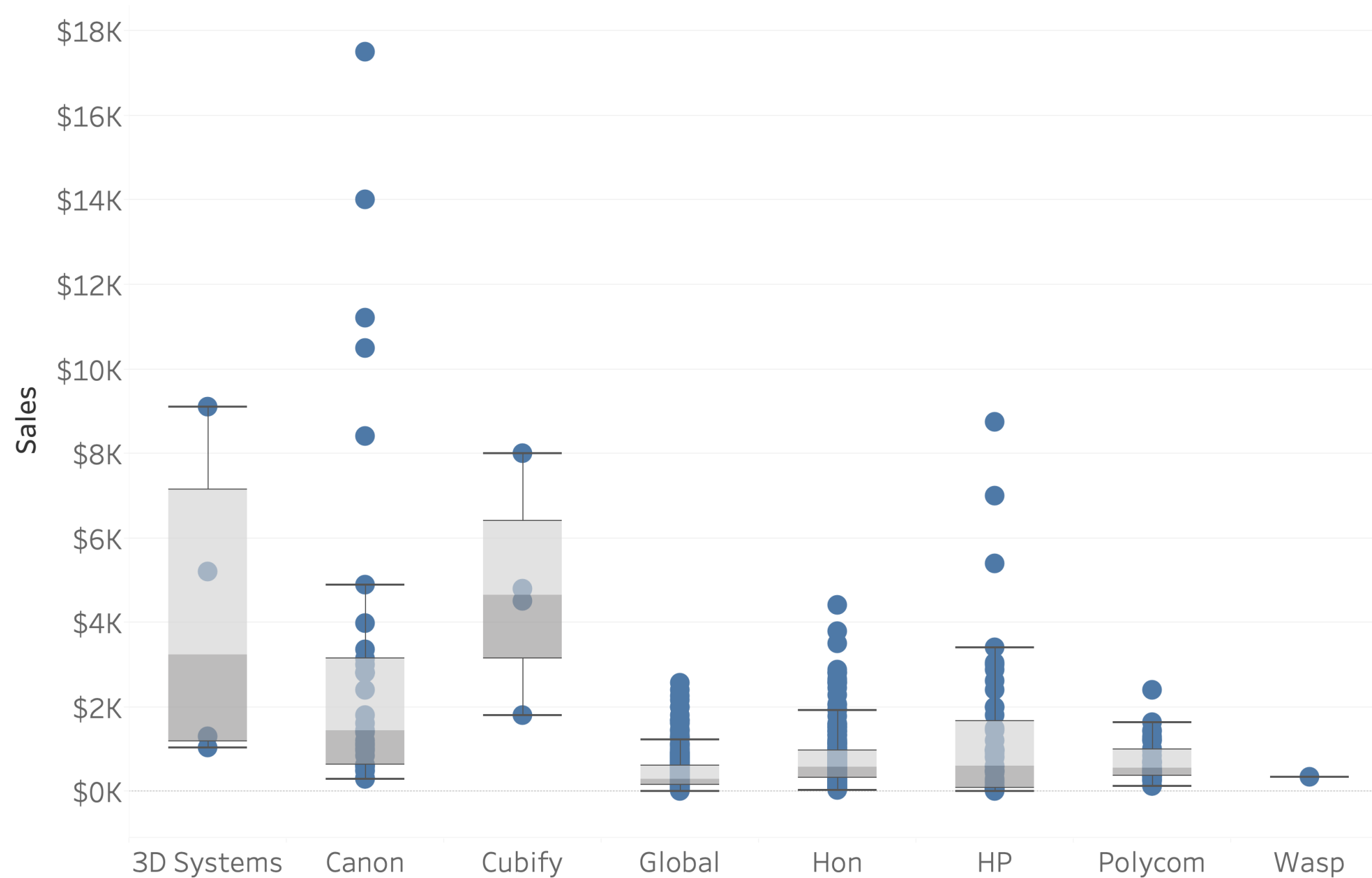
When to use a box plot

- Compare distributions among multiple categories

- Spot trends and differences between categories

What about the mean?

- Average = arithmetic mean

- $\dfrac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$

- Average and mean are often used interchangeably

Skewness

right (positive) skewed

Average
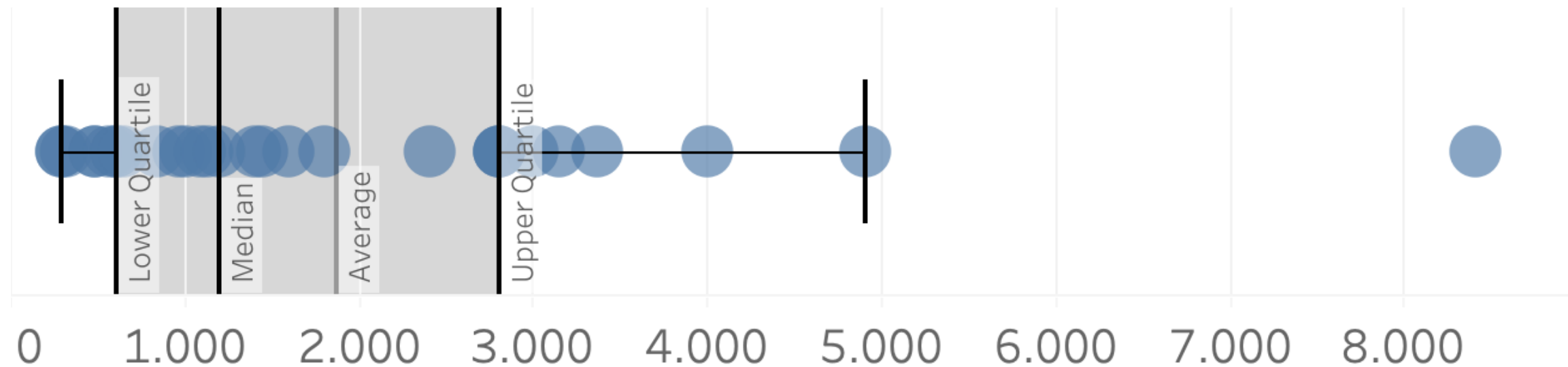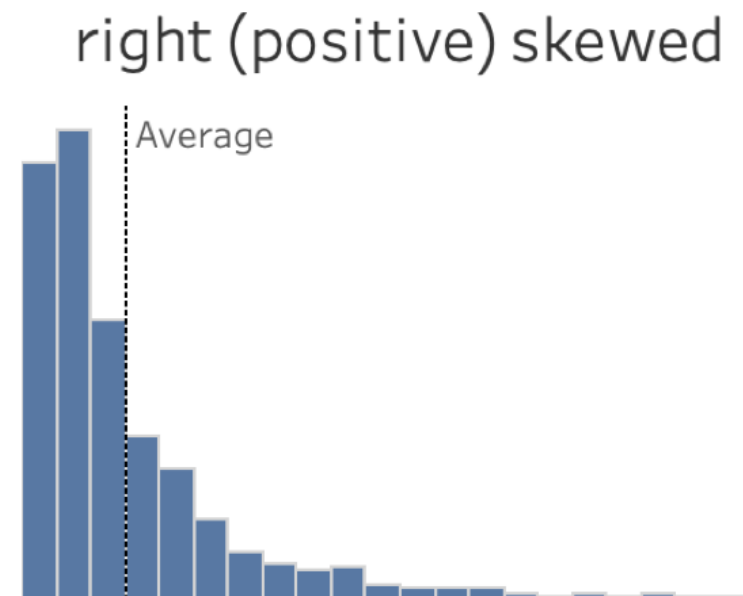
Skewness

Skewness

Skewness
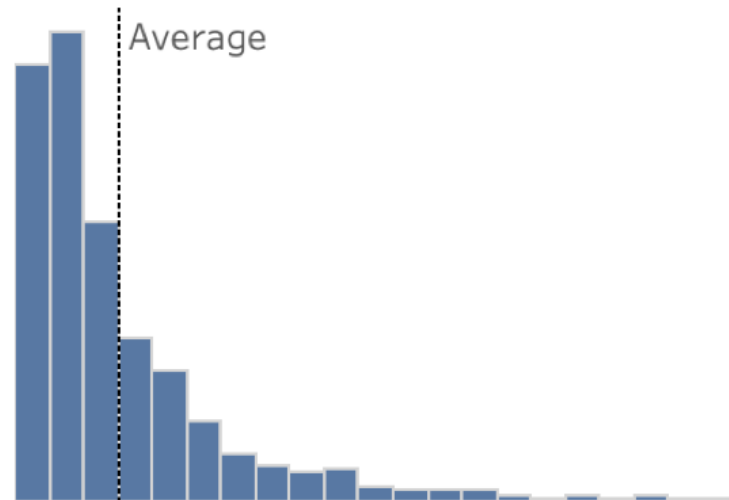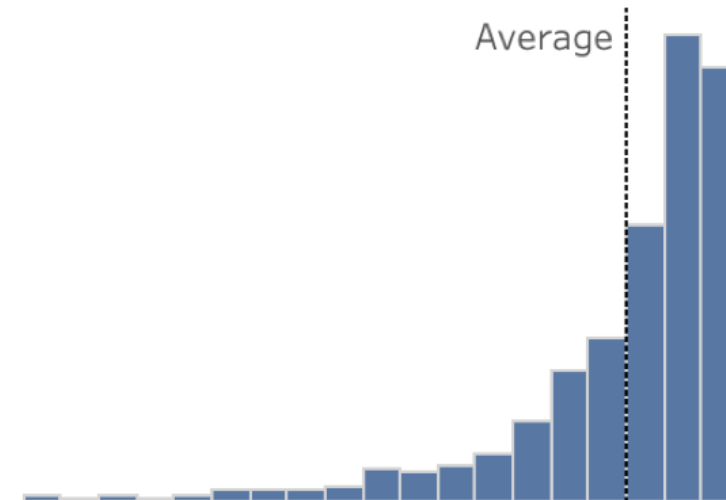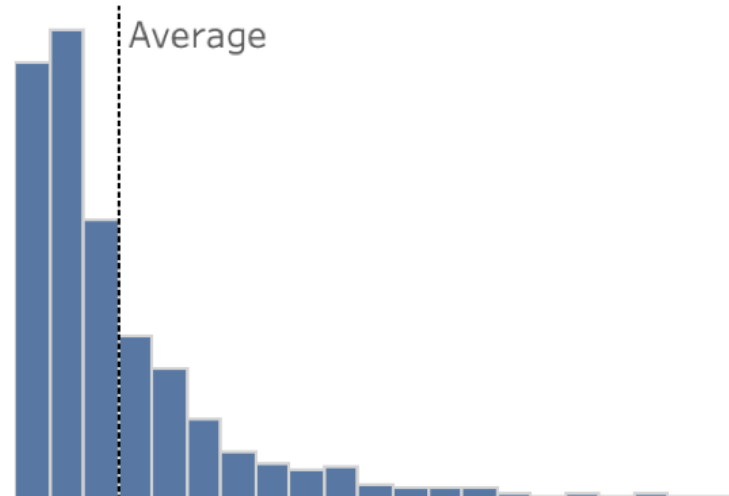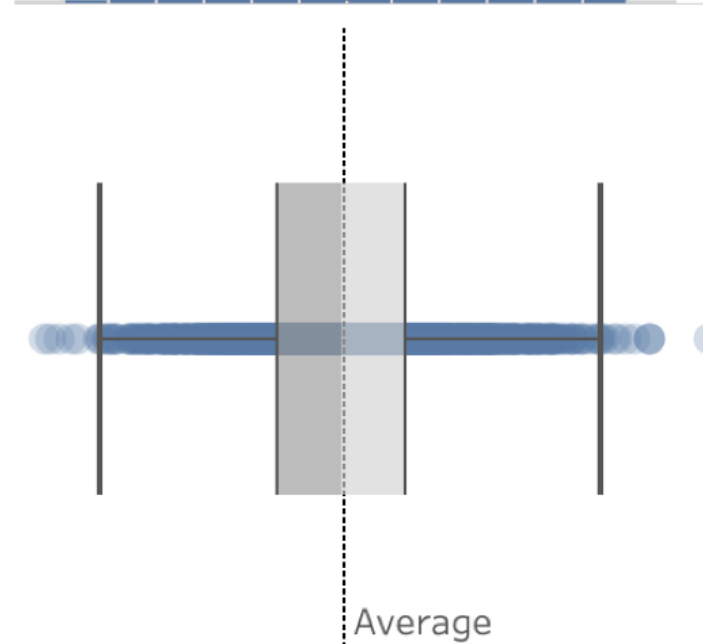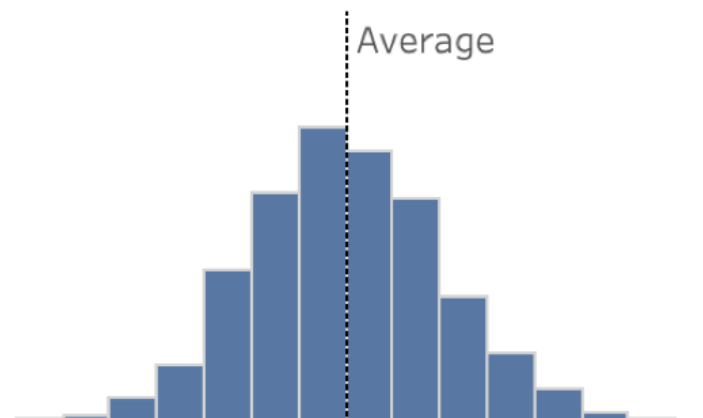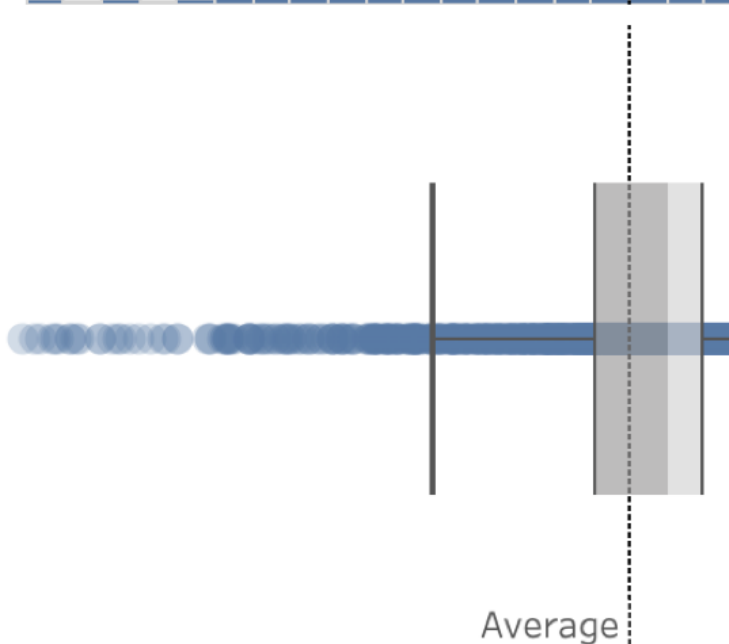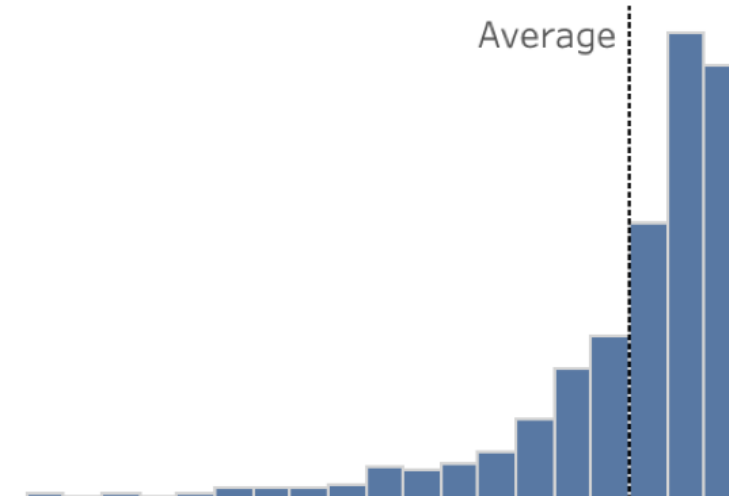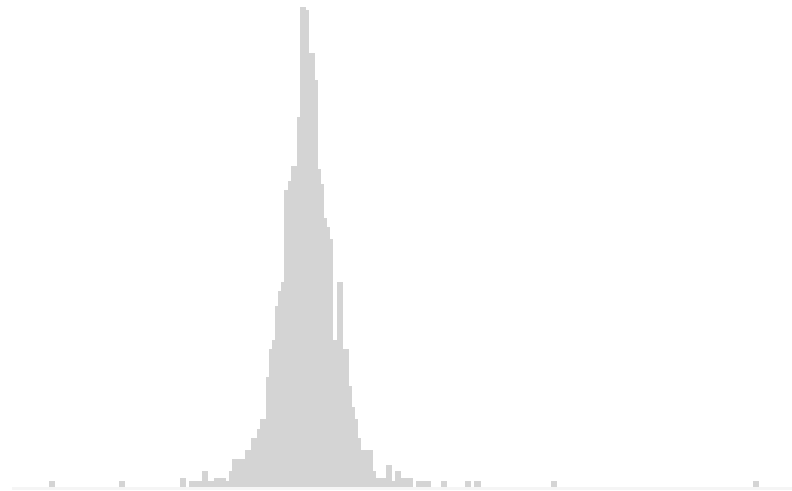

right (positive) skewed
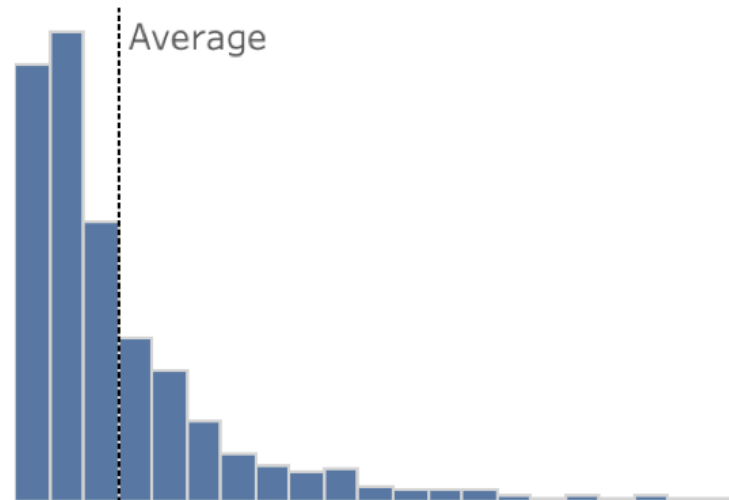
symmetric
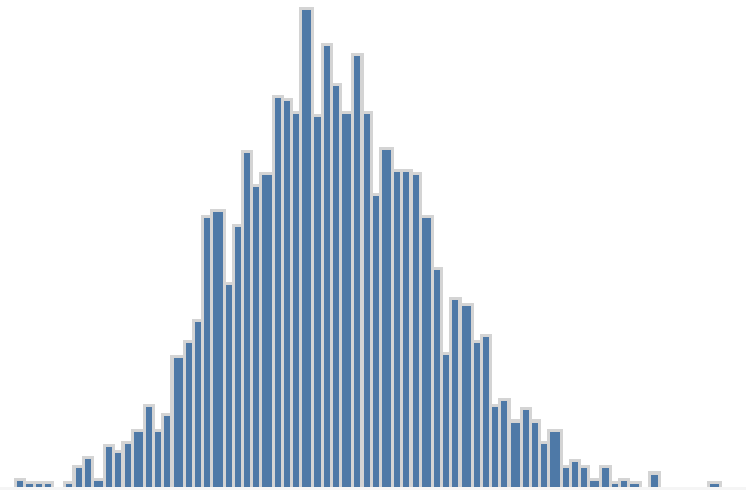(zero skewness)

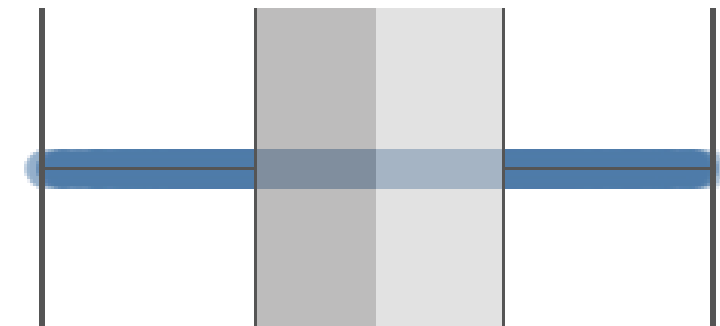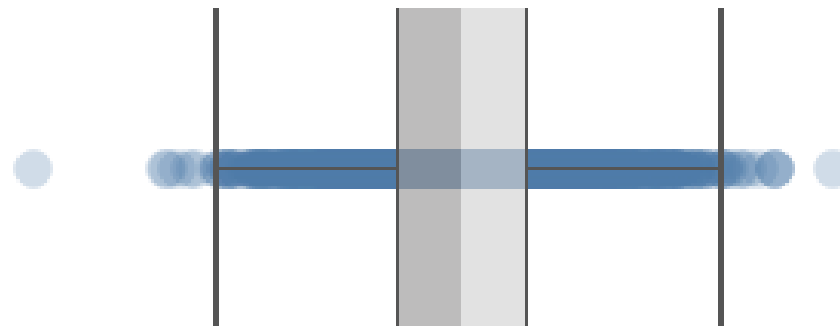left (negative) skewed

Excess kurtosis
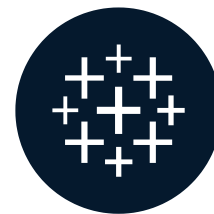
Let's practice!
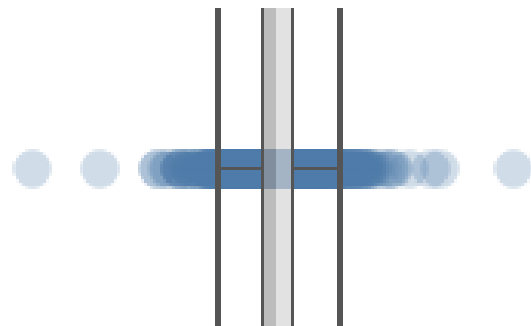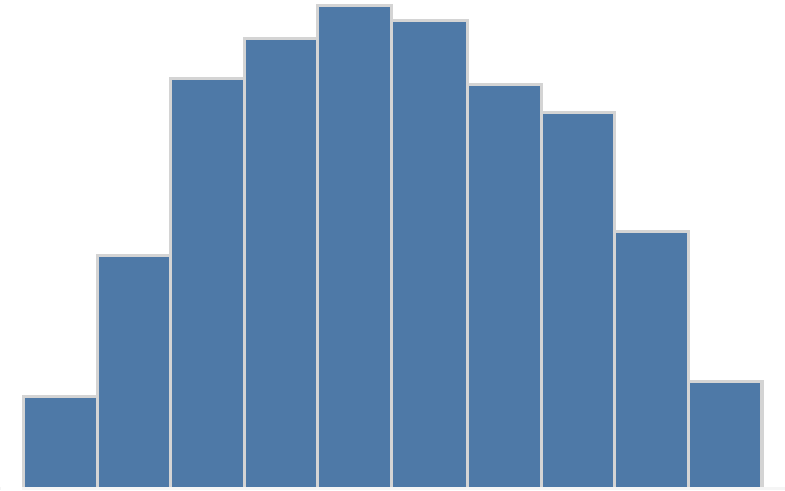
STATISTICAL TECHNIQUES IN TABLEAU

datacamp

EDA in Tableau: box plots

STATISTICAL TECHNIQUES IN TABLEAU

Maarten Van den Broeck
Content Developer at DataCamp

Let's practice!

STATISTICAL TECHNIQUES IN TABLEAU