অংকুরণ

Descriptive vs Inferential Statistics

Descriptive and inferential statistics are two main branches of statistics that are used to analyze and interpret data.

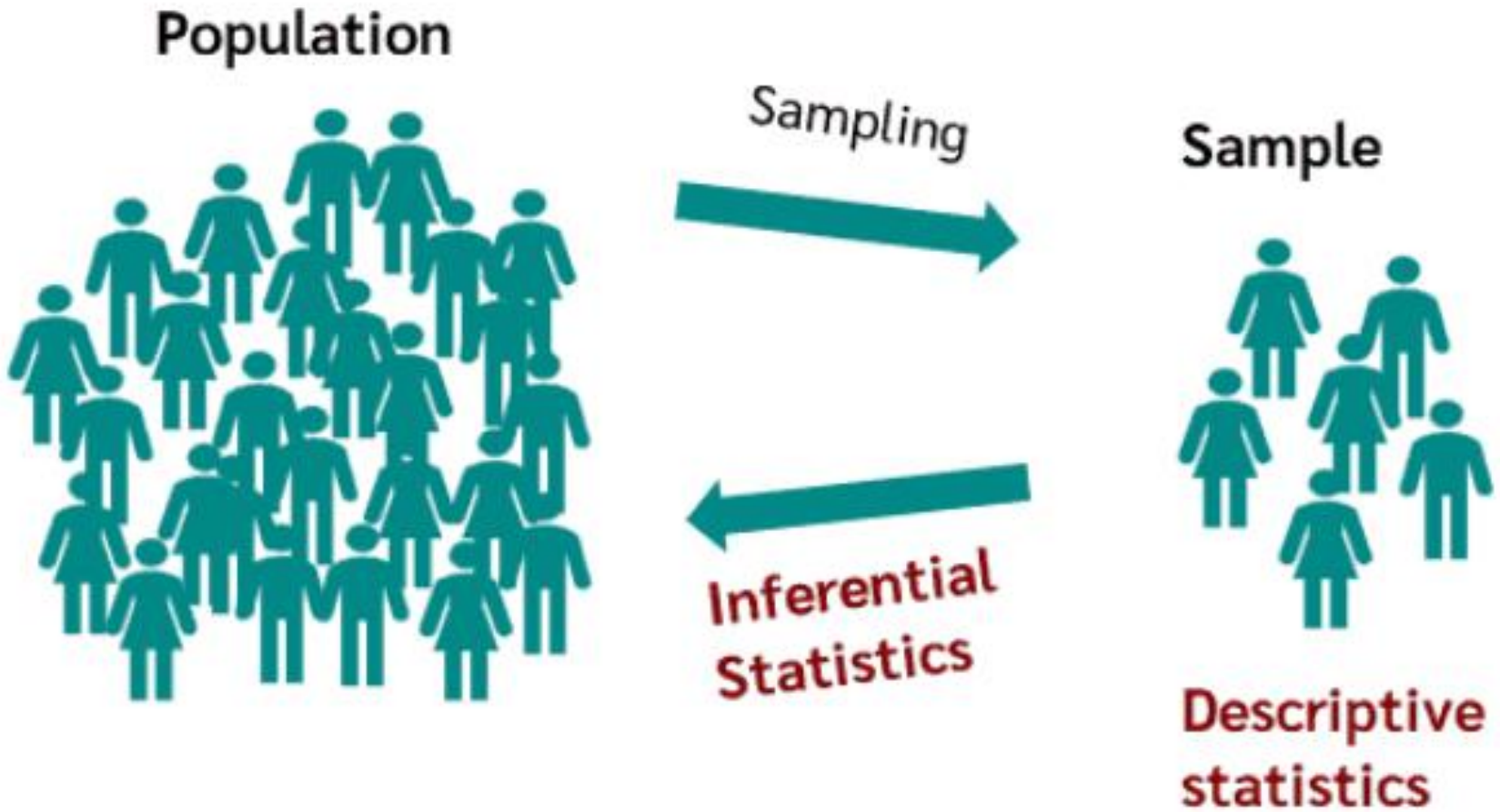
DESCRIPTIVE

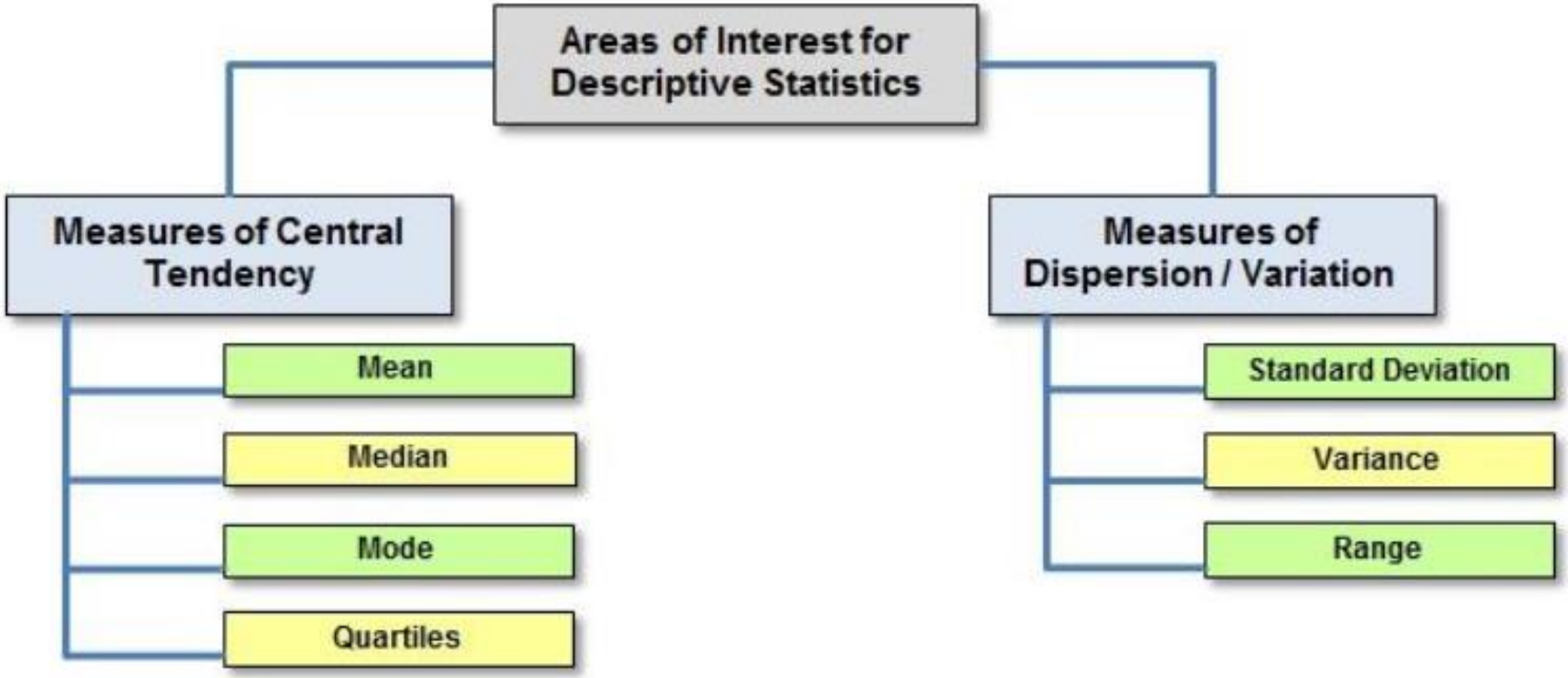
Descriptive statistics is a branch of statistics used to summarize and describe the characteristics of a dataset. Descriptive statistics involves calculating summary measures, such as the mean, median, mode, range, standard deviation, variance.

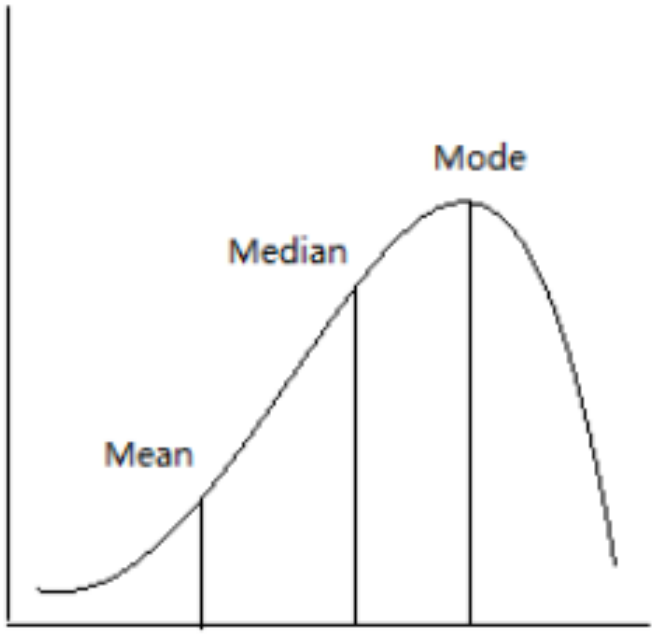
VS

INFERENTIAL

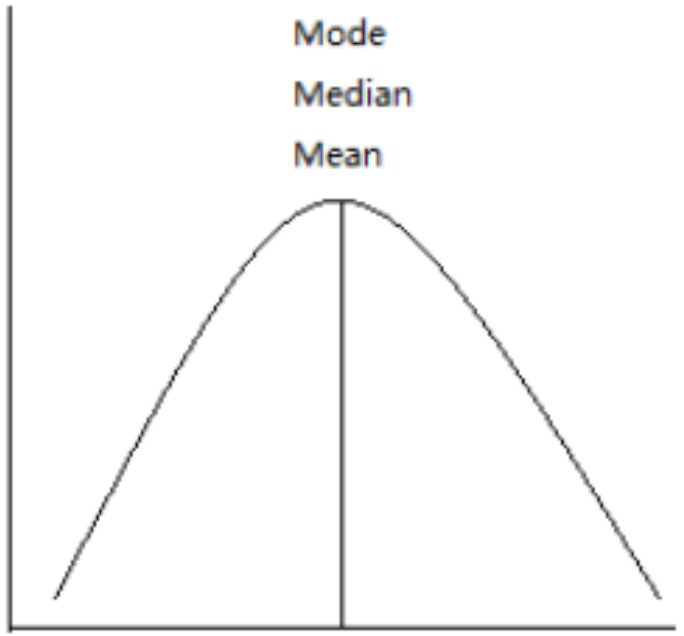
Inferential statistics is a branch of statistics used to make inferences or predictions about a population based on a sample of data. Inferential statistics involves using statistical tests, such as hypothesis tests and regression analysis.


Population
Sampling
Sample
Inferential Statistics
Descriptive statistics

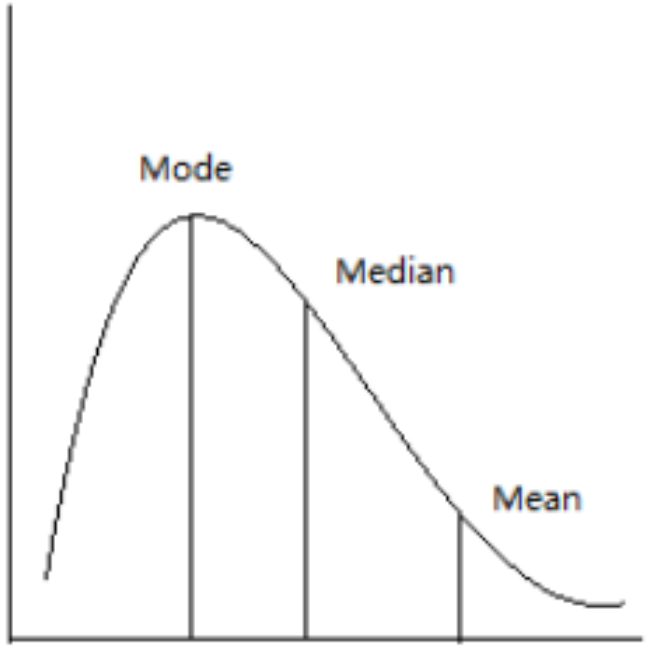

Areas of Interest for Descriptive Statistics
Measures of Central Tendency
Mean
Median
Mode
Quartiles
Measures of Dispersion / Variation
Standard Deviation
Variance
Range



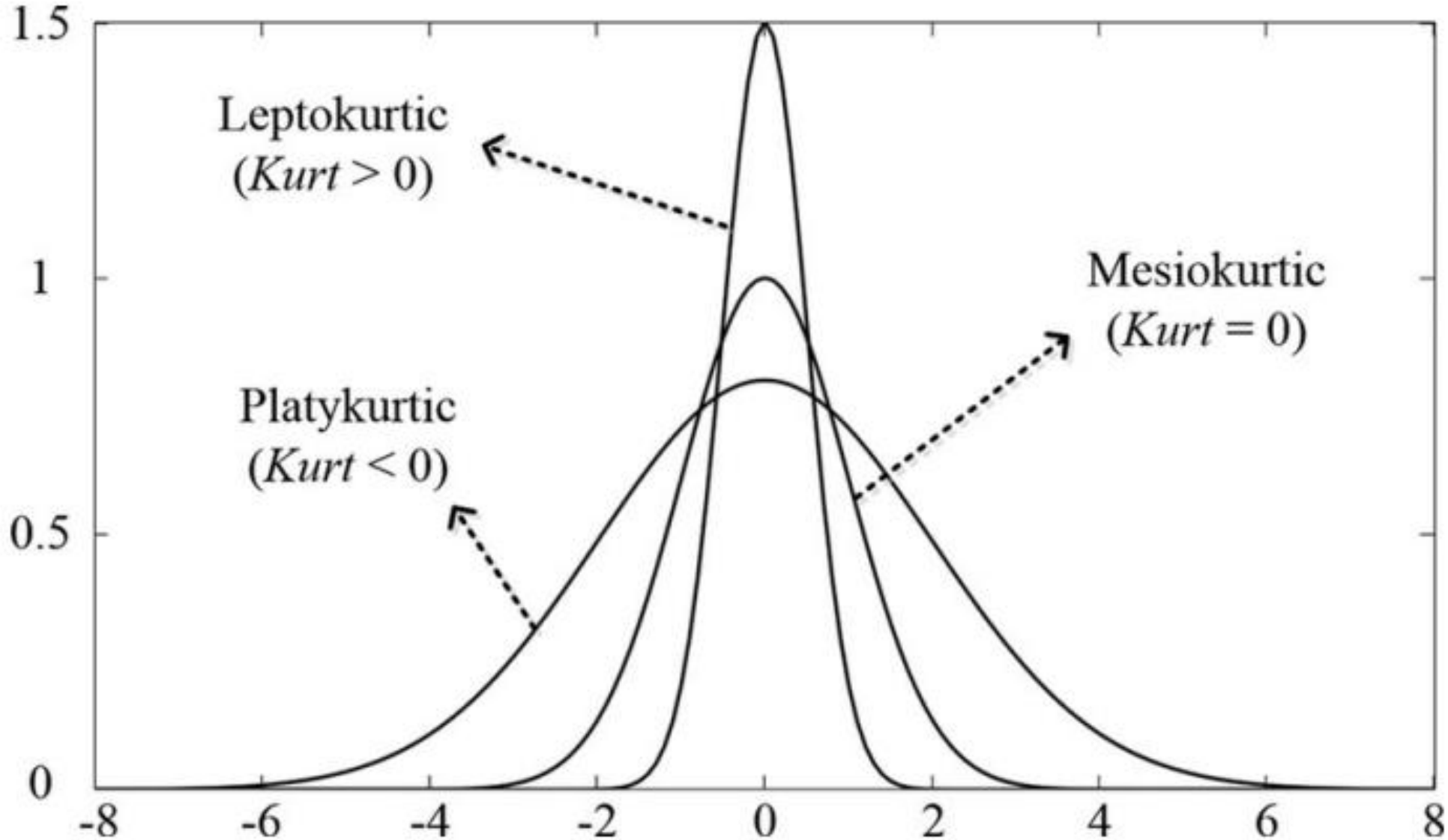
Left skew



Normal Distribution

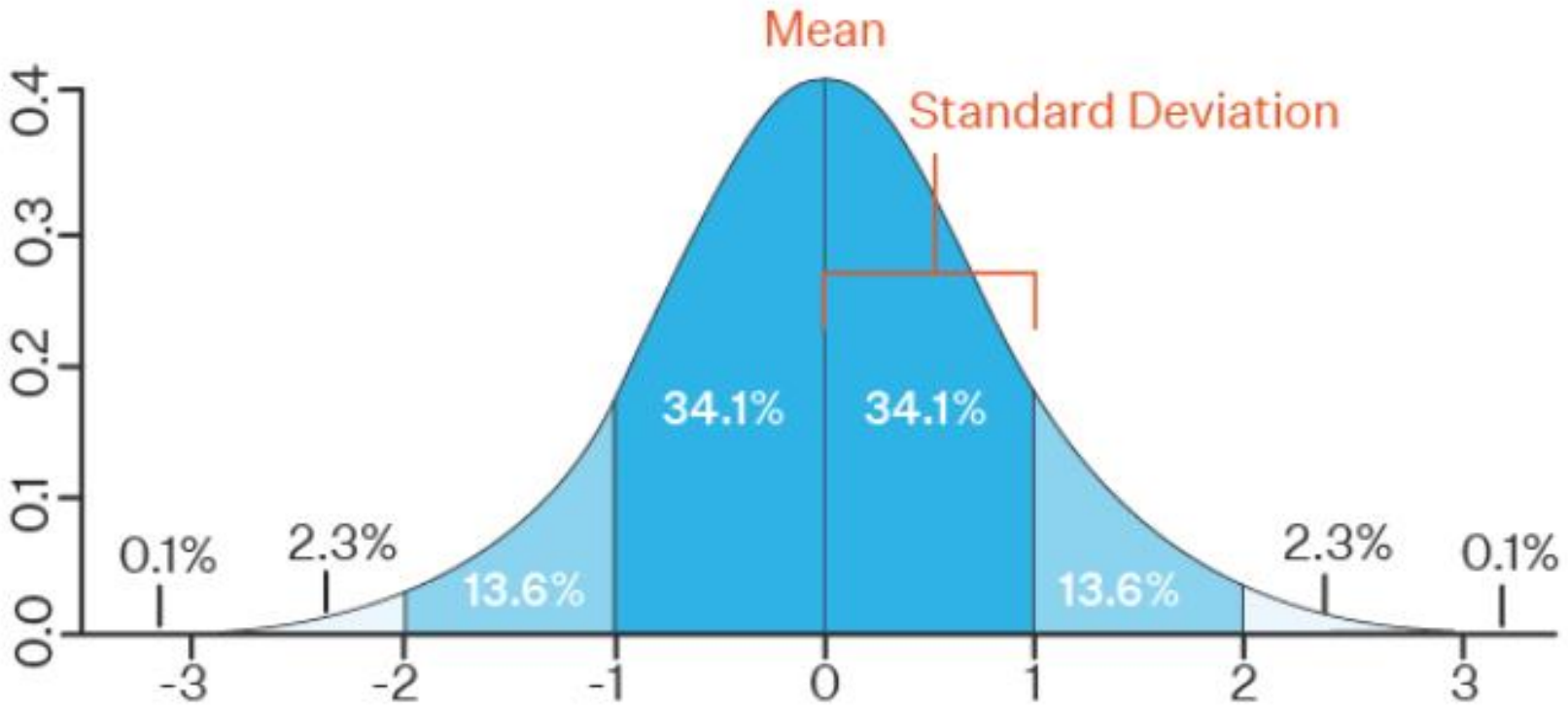


Right skew


1.5
1
0.5
0
-8
-6
-4
-2
0
2
4
6
8
Leptokurtic
(Kurt > 0)
Platykurtic
(Kurt < 0)
Mesiokurtic
(Kurt = 0)

Standard Deviation

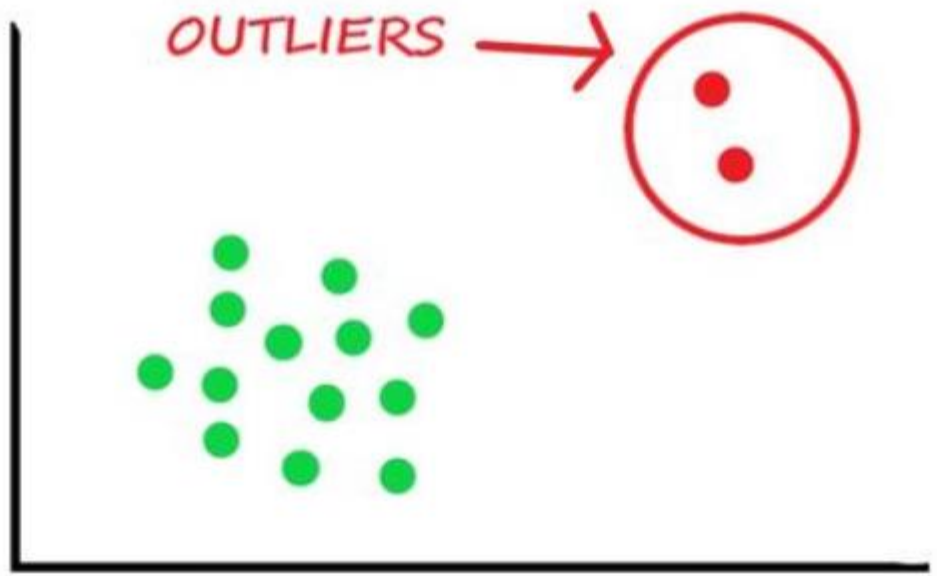
- a measure of how spread out the values in a dataset are from the mean.

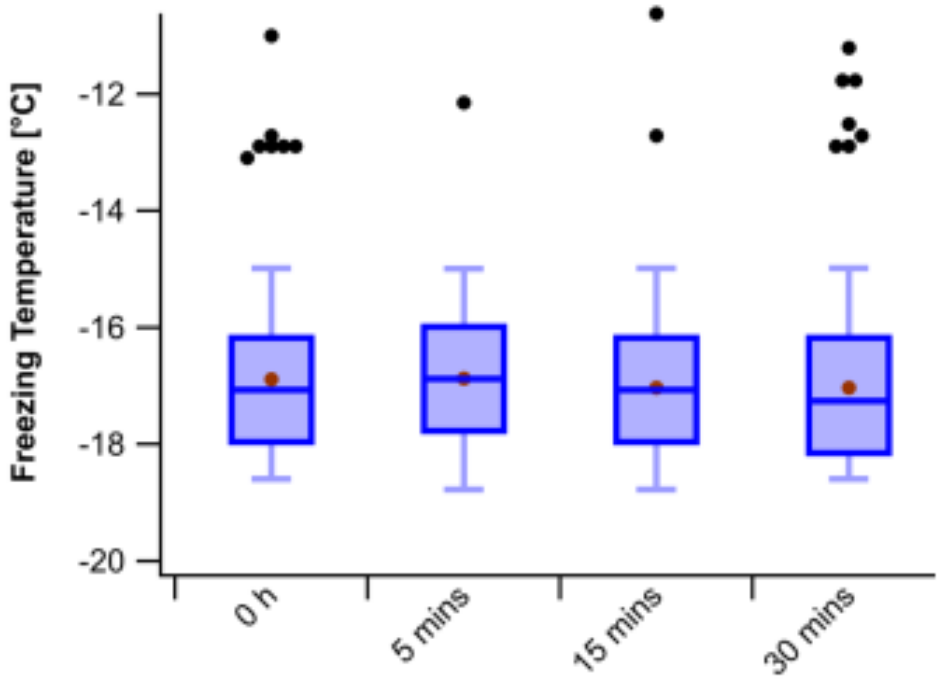


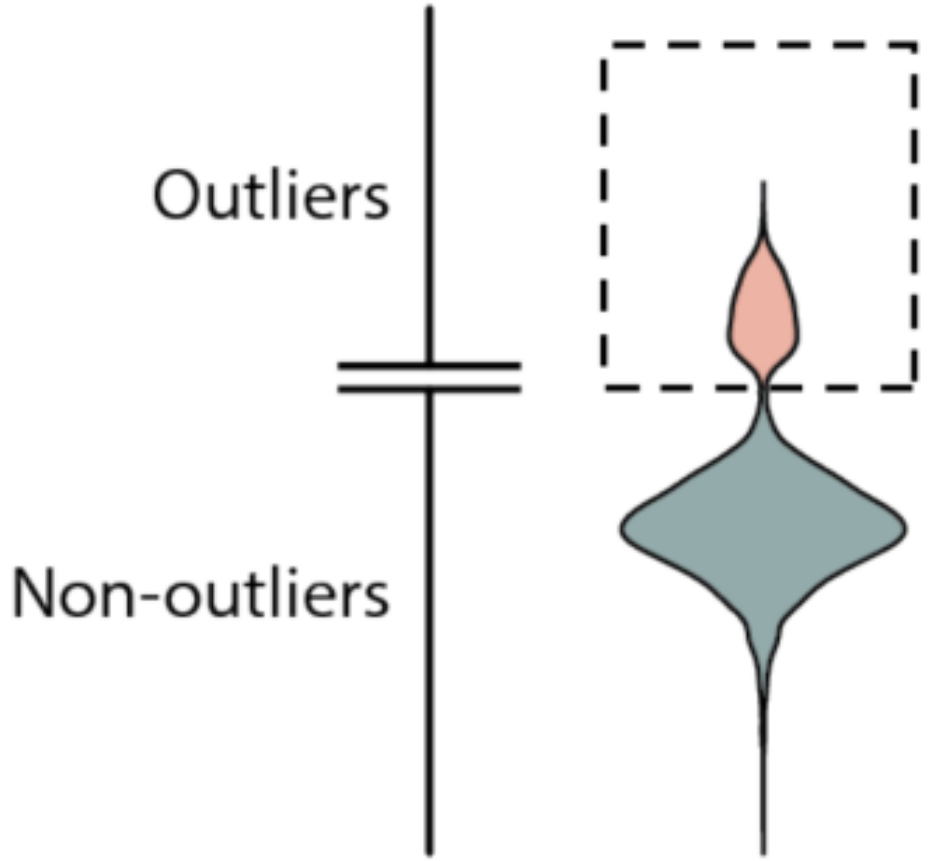
Empirical Rule (68-95-99.7 Rule)

This rule describes how data is distributed in a normal distribution:

- **68% of the data** falls within **1 standard deviation (σ)** of the mean (between -1σ and +1σ).
- **95% of the data** falls within **2 standard deviations (σ)** of the mean (between -2σ and +2σ).
- **99.7% of the data** falls within **3 standard deviations (σ)** of the mean (between -3σ and +3σ).

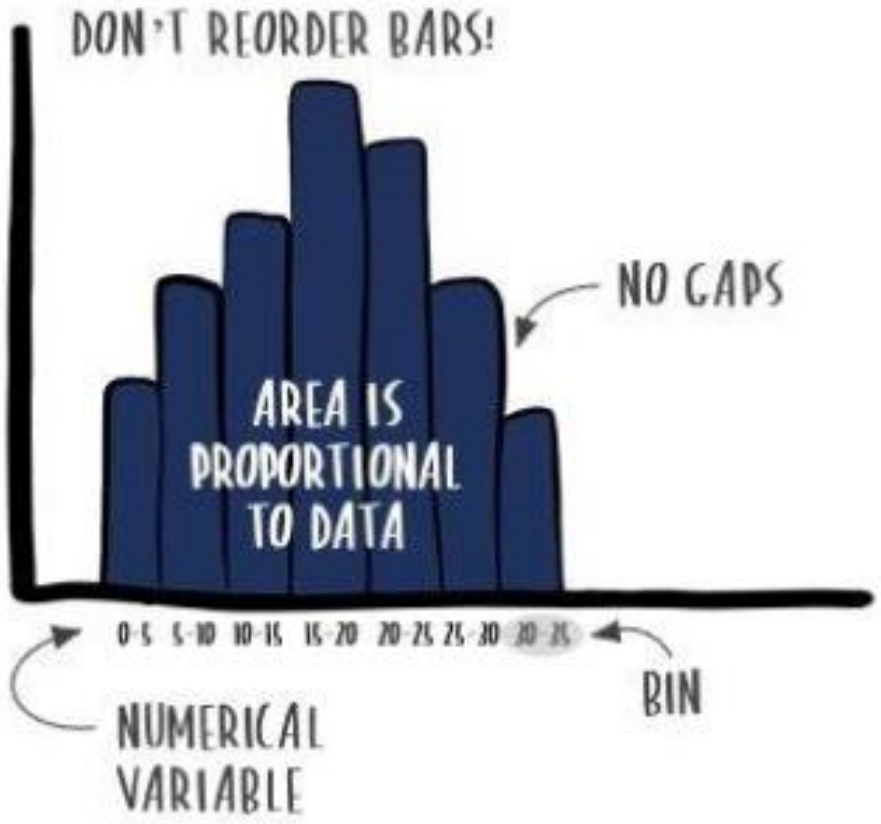

OUTLIERS


Freezing Temperature [°C]
-12
-14
-16
-18
-20
0 h
5 mins
15 mins
30 mins

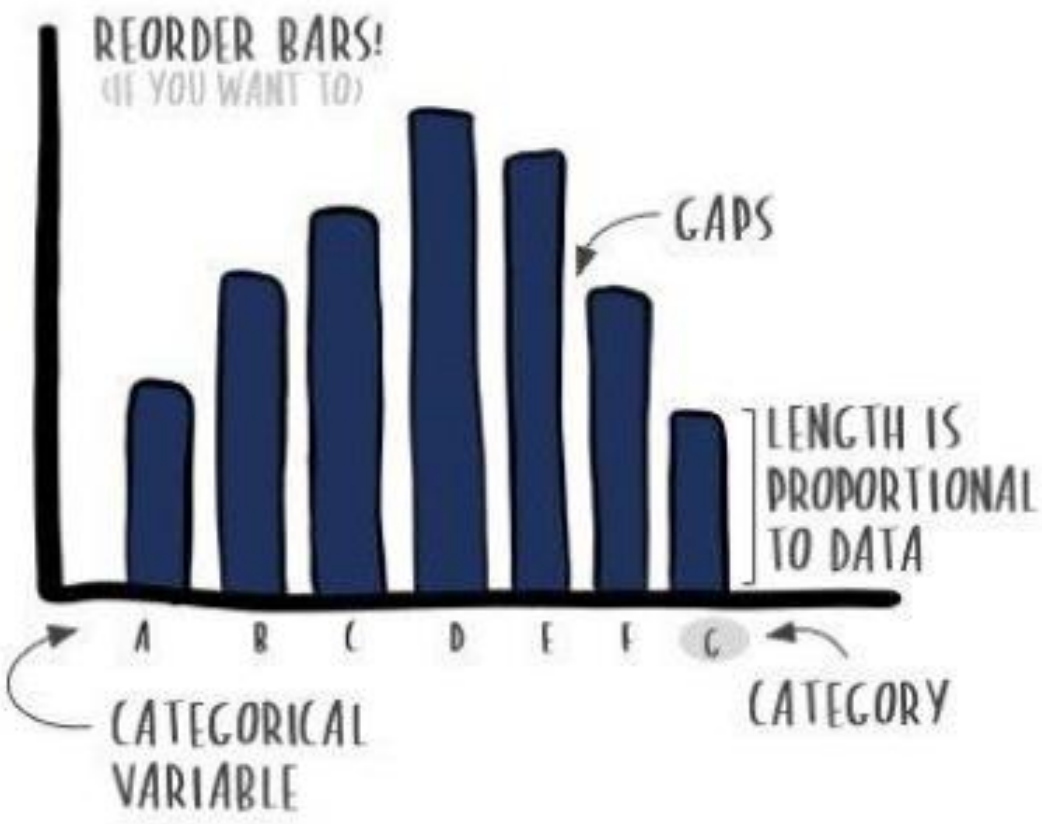

Outliers
Non-outliers

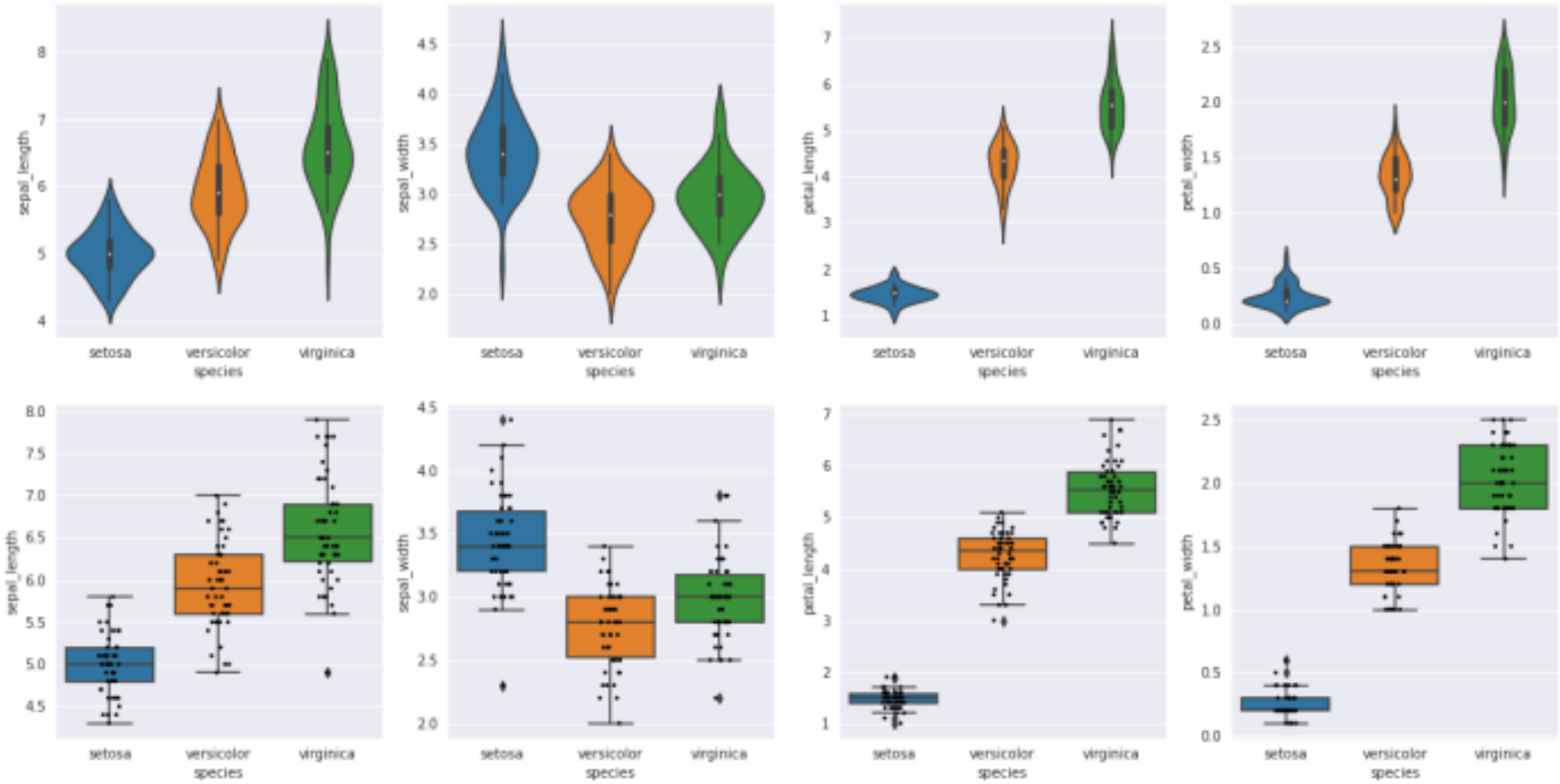

Don't let variability, outliers &
skewness hide: ban bar graphs
?
?
?

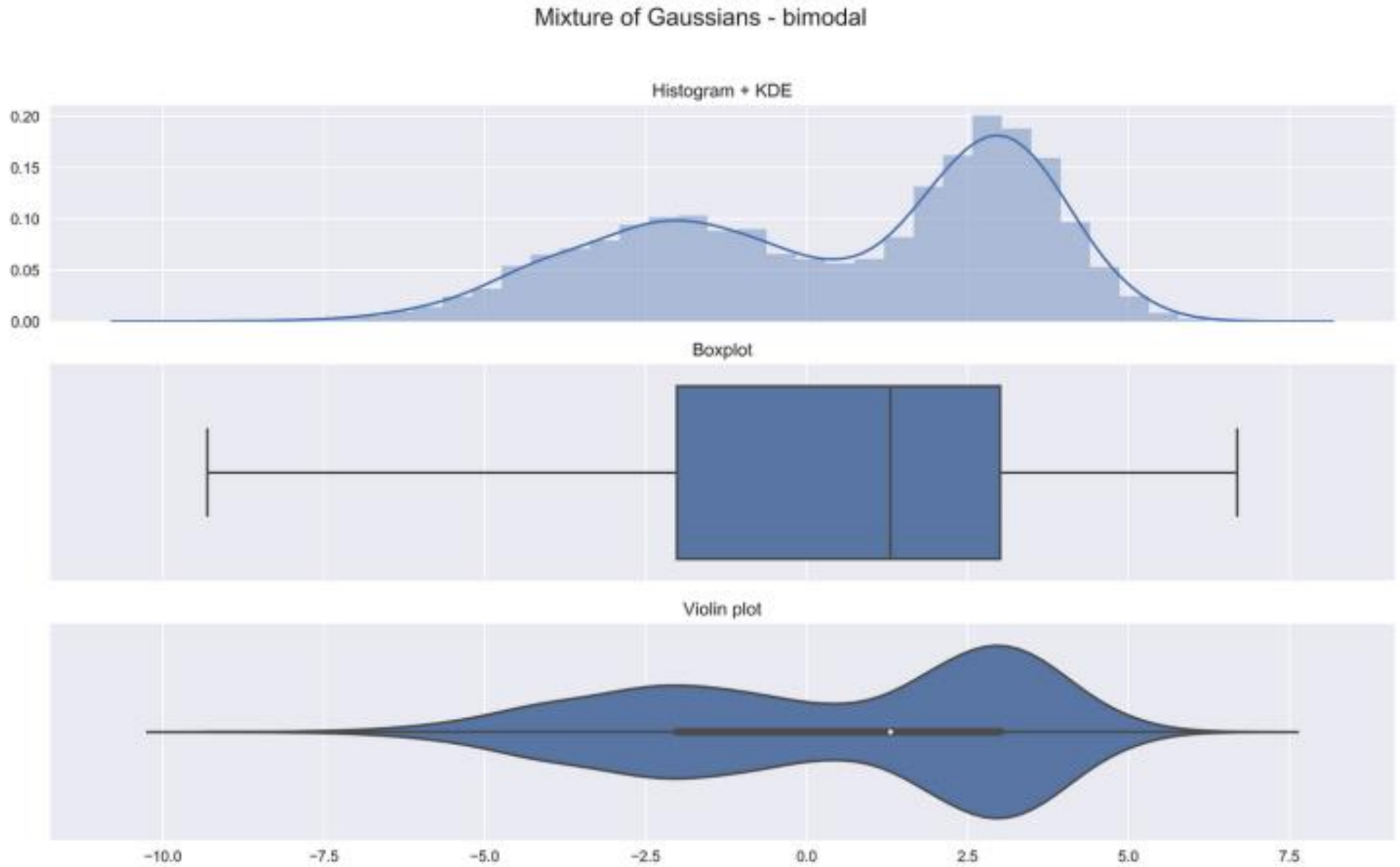
This is a **histogram**...

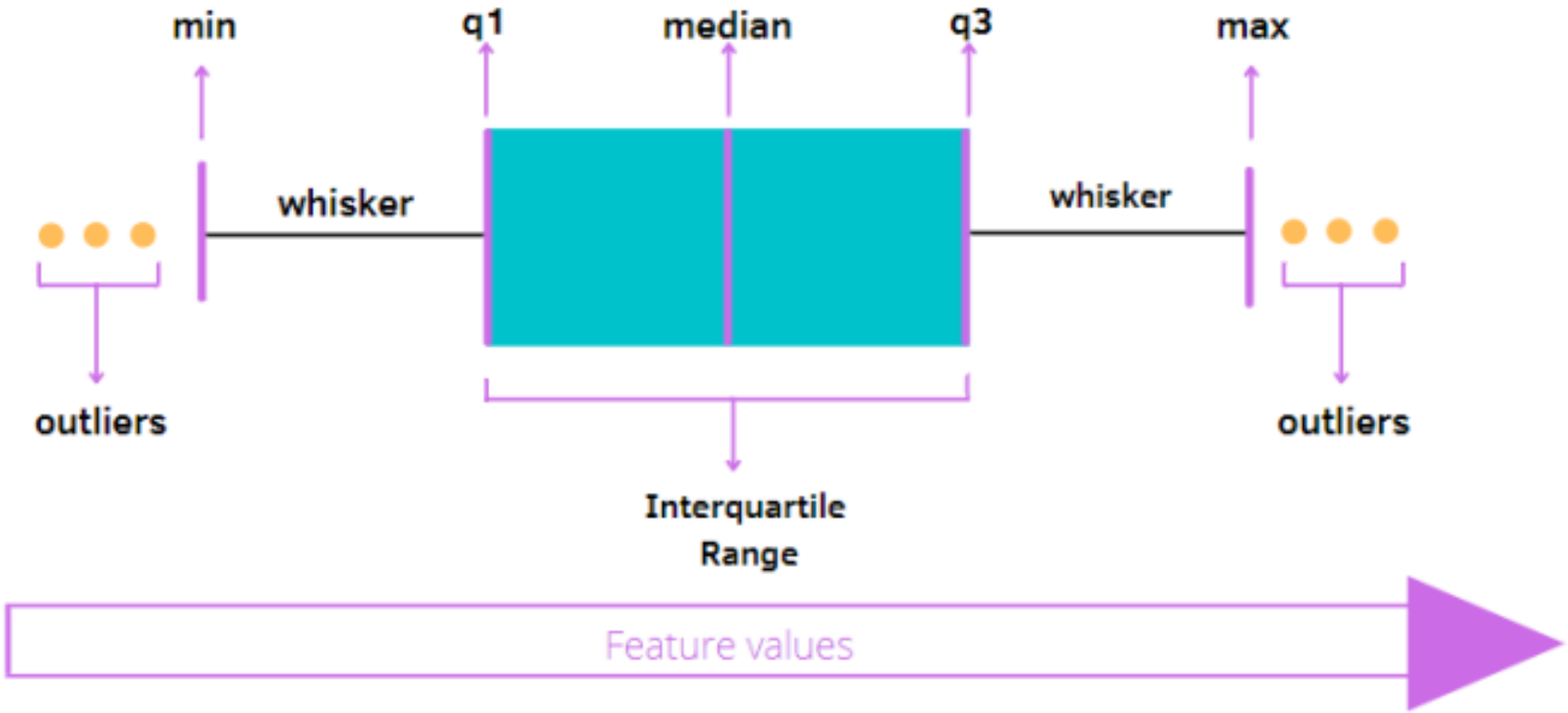


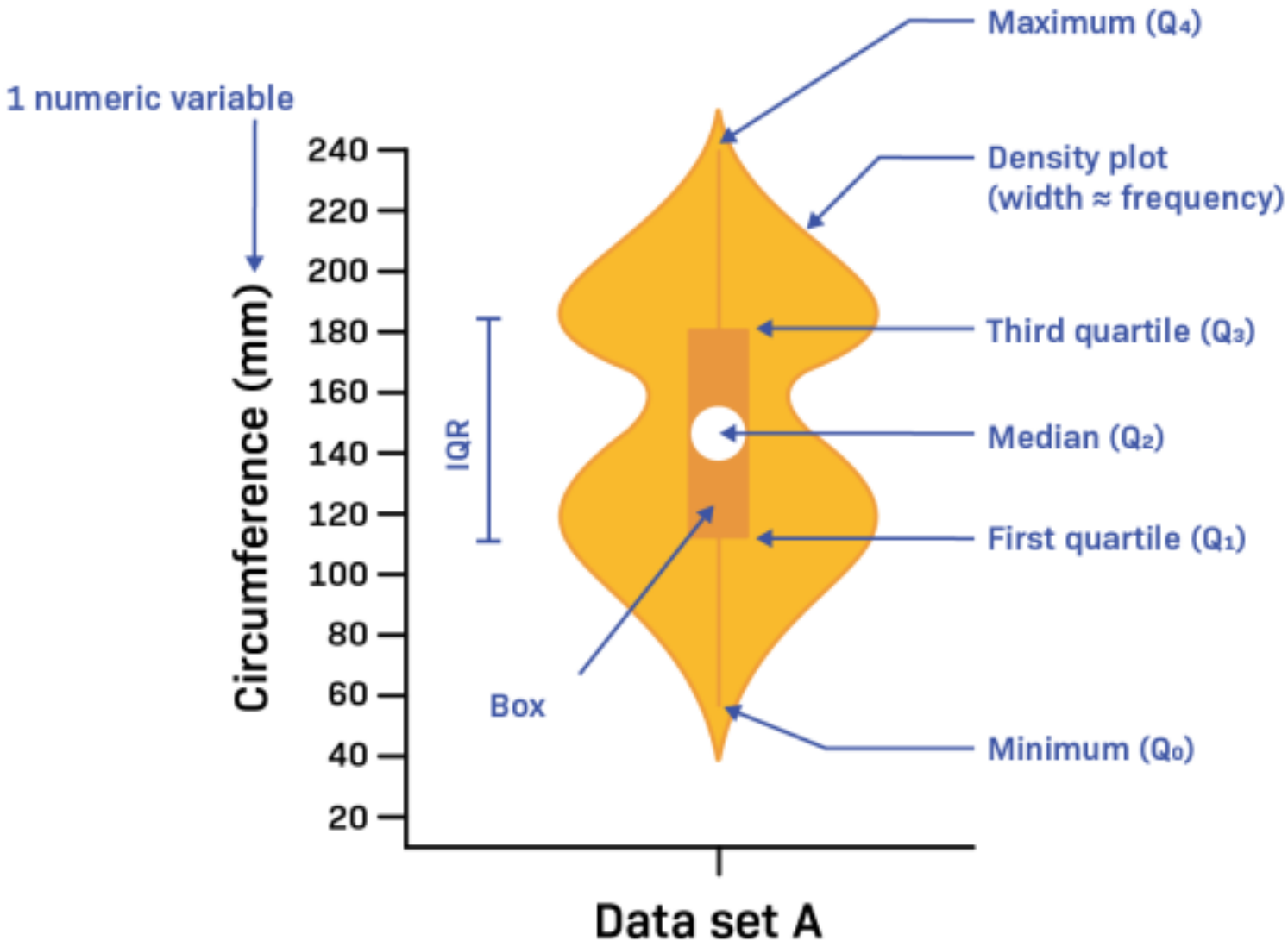
This is a **bar chart**...




sepal_length
sepal_width
petal_length
petal_width
setosa
versicolor
virginica
species


Mixture of Gaussians - bimodal
Histogram + KDE
0.20
0.15
0.10
0.05
0.00
Boxplot
Violin plot
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5


min
q1
median
q3
max
whisker
whisker
outliers
outliers
Interquartile
Range
Feature values


1 numeric variable
Circumference (mm)
240
220
200
180
160
140
120
100
80
60
40
20
IQR
Box
Maximum (Q4)
Density plot
(width ≈ frequency)
Third quartile (Q3)
Median (Q2)
First quartile (Q1)
Minimum (Q0)
Data set A