

# **DATA ANALYSIS: TOOLS FOR ENVIRONMENTAL SCIENTISTS**

Descriptive Statistics and  
Exploratory Data Analysis

**IDA RAHU**

# VARIABLES & THEIR TYPES

VARIABLE = RECORDED INFO



precipitation (mm)  
rainy days per month      snow cover present (yes/no)  
soil type      number of trees  
age      species      drought severity  
temperature (°C)      weight      wind speed (m/s)  
fire danger level      nitrate concentration      land cover class

# VARIABLES & THEIR TYPES



VARIABLE = RECORDED INFO

CATEGORICAL or QUALITATIVE

snow cover present (yes/no)

soil type

species

drought severity

land cover class

fire danger level

NUMERIC or QUANTITATIVE

precipitation (mm)

rainy days per month

number of trees

age

wind speed (m/s)

temperature (°C)

weight

nitrate concentration

# VARIABLES & THEIR TYPES



VARIABLE = RECORDED INFO

## CATEGORICAL or QUALITATIVE

### NOMINAL

no order based on magnitude

snow cover present (yes/no)

soil type

species

land cover class

### ORDINAL

rank/order

drought severity

fire danger level

## NUMERIC or QUANTITATIVE

### DISCRETE

integers only, countable

rainy days per month

number of trees

### CONTINUOUS

measured on continuous scale

nitrate concentration  
precipitation (mm)

age  
temperature (°C)

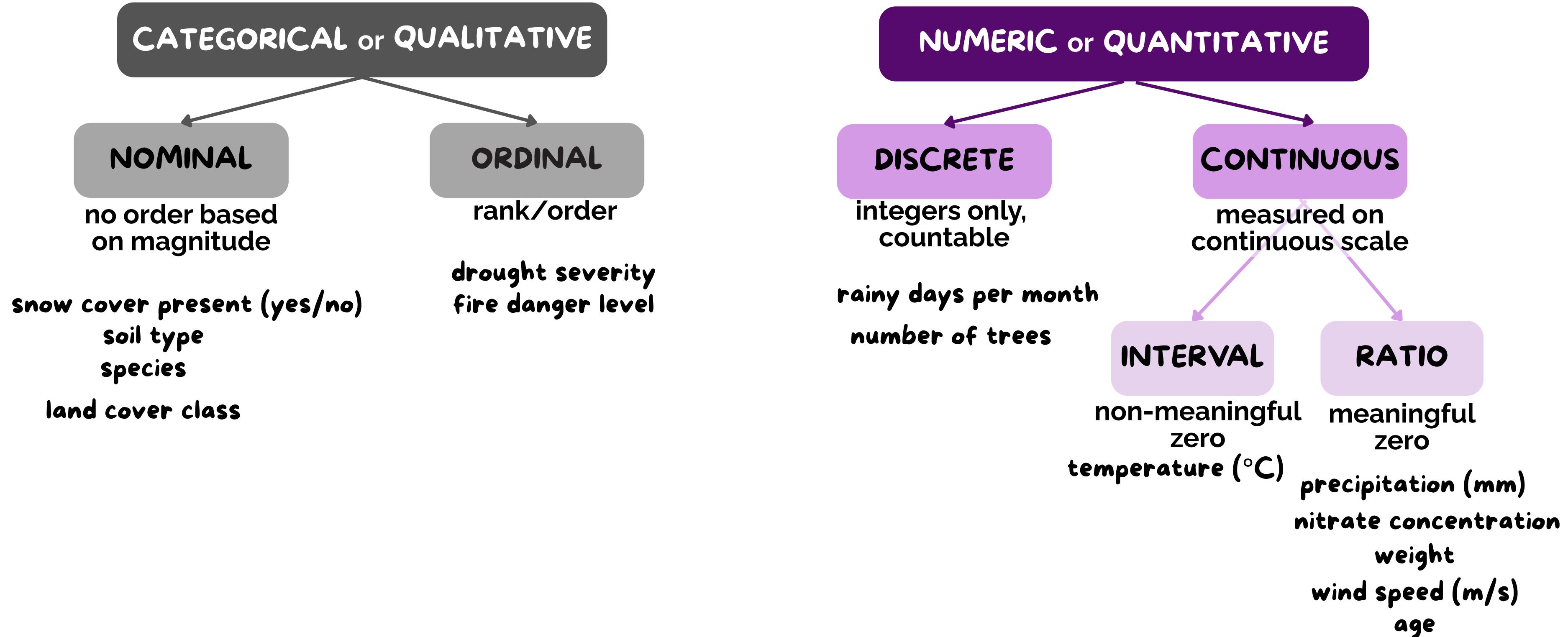
weight

wind speed (m/s)

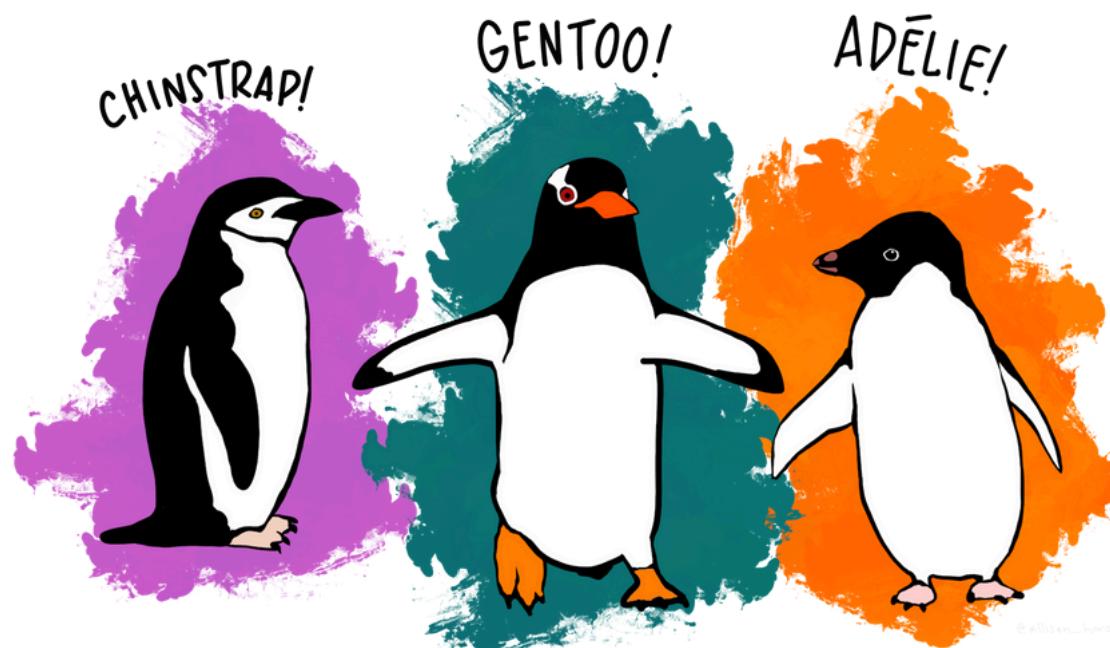
# VARIABLES & THEIR TYPES



VARIABLE = RECORDED INFO



# CATEGORICAL VARIABLES



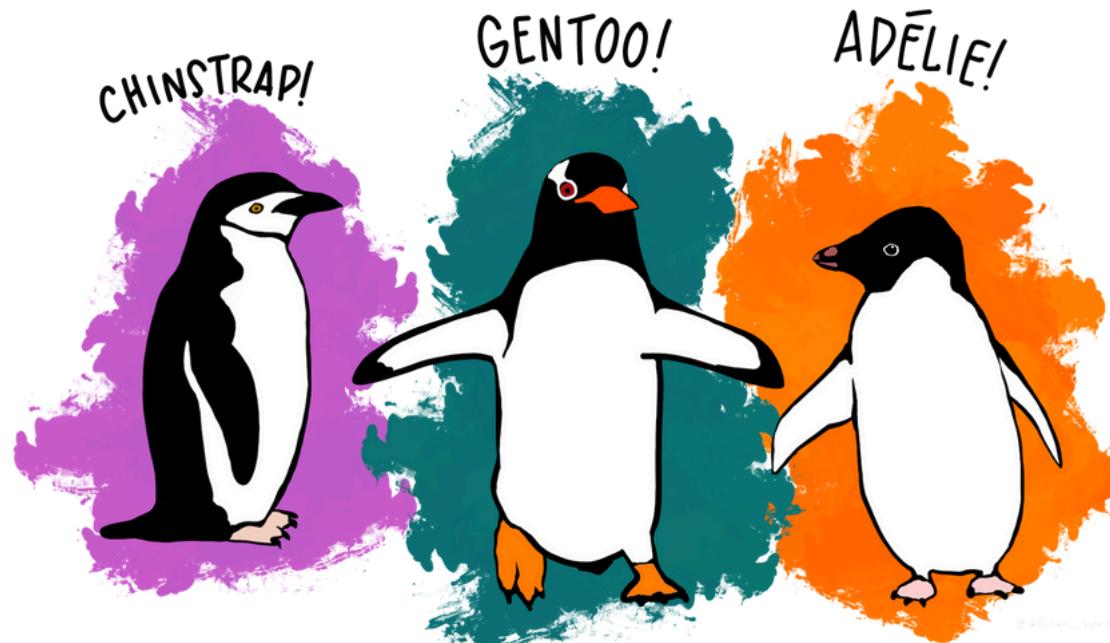
The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

8 variables (n = 344 penguins)

<b>Species</b>	Penguin species (Adélie, Chinstrap, Gentoo)
<b>Island</b>	Island in the Palmer Archipelago where observed
<b>Bill length (mm)</b>	Length of the penguin's bill (mm)
<b>Bill depth (mm)</b>	Depth (thickness) of the penguin's bill (mm)
<b>Flipper length (mm)</b>	Length of the penguin's flipper (mm)
<b>Body mass (g)</b>	Body mass of the penguin (g)
<b>Sex</b>	Male or female (some values missing)
<b>Year</b>	Year of observation (2007–2009)

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# CATEGORICAL VARIABLES



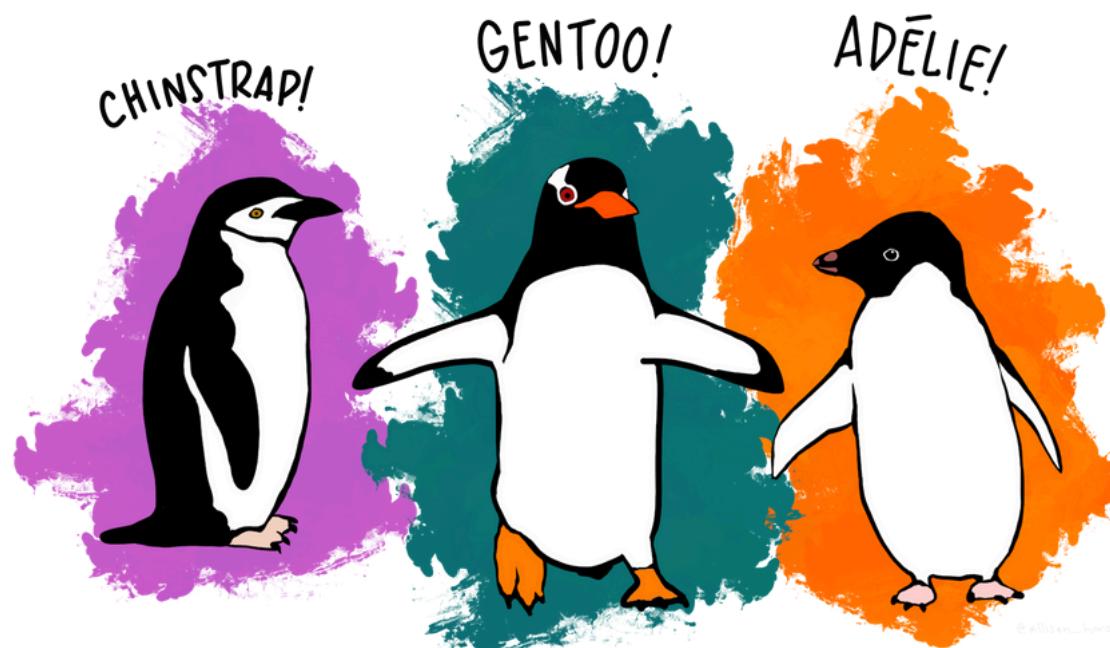
8 variables (n = 344 penguins)

The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

<b>Species</b>	Penguin species (Adélie, Chinstrap, Gentoo)
<b>Island</b>	Island in the Palmer Archipelago where observed
<b>Bill length (mm)</b>	Length of the penguin's bill (mm)
<b>Bill depth (mm)</b>	Depth (thickness) of the penguin's bill (mm)
<b>Flipper length (mm)</b>	Length of the penguin's flipper (mm)
<b>Body mass (g)</b>	Body mass of the penguin (g)
<b>Sex</b>	Male or female (some values missing)
<b>Year</b>	Year of observation (2007–2009)

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# CATEGORICAL VARIABLES



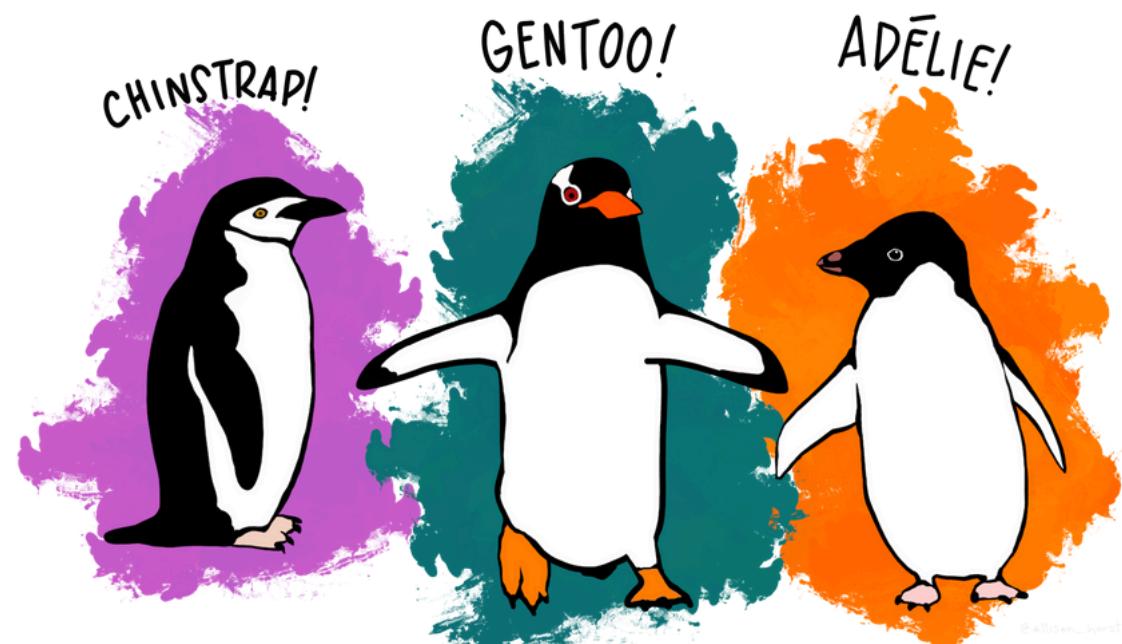
8 variables (n = 344 penguins)

<b>Species</b>	Penguin species
<b>Island</b>	Island in the Palmer
<b>Bill</b>	Length of the
<b>Bill</b>	Depth (thickness) of
<b>Flipper</b>	Length of the
<b>Body</b>	Body mass of the
<b>Sex</b>	Male or female (some)
<b>Year</b>	Year of observation

<b>Island</b>	<b>Frequency</b>	<b>Proportion</b>
Biscoe	168	
Dream	124	
Torgersen	52	
<b>TOTAL</b>	<b>344</b>	

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# CATEGORICAL VARIABLES



8 variables (n = 344 penguins)

<b>Species</b>	Penguin species
<b>Island</b>	Island in the Palmer
<b>Bill</b>	Length of the
<b>Bill</b>	Depth (thickness) of
<b>Flipper</b>	Length of the
<b>Body</b>	Body mass of the
<b>Sex</b>	Male or female (some)
<b>Year</b>	Year of observation

Island	Frequency	Proportion
Biscoe	168	0.489
Dream	124	0.36
Torgersen	52	0.151
<b>TOTAL</b>	<b>344</b>	<b>1</b>

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.



## PIE CHART

Torgersen

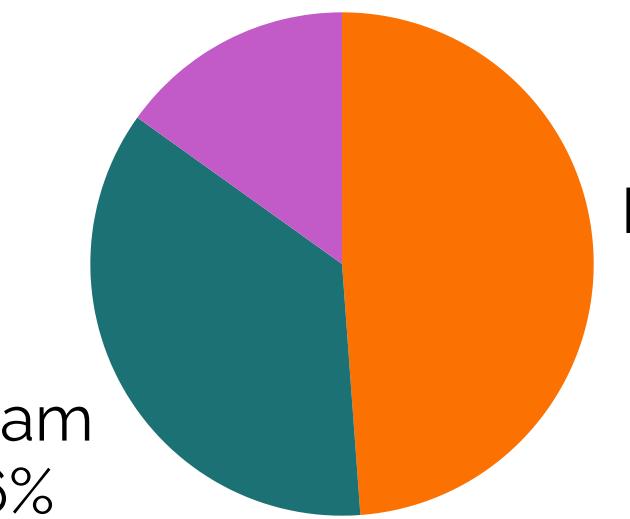
15.1%

Dream

36%

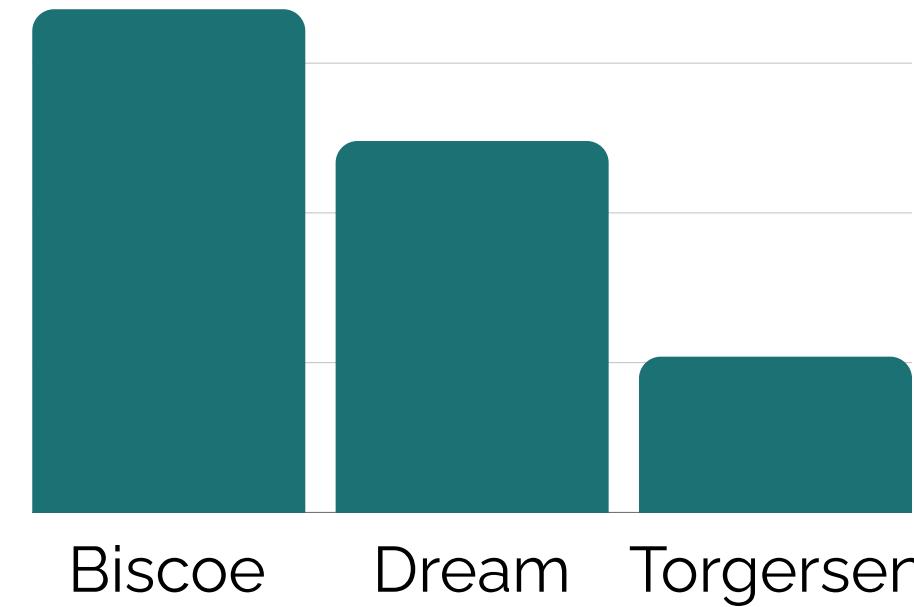
Biscoe

48.8%

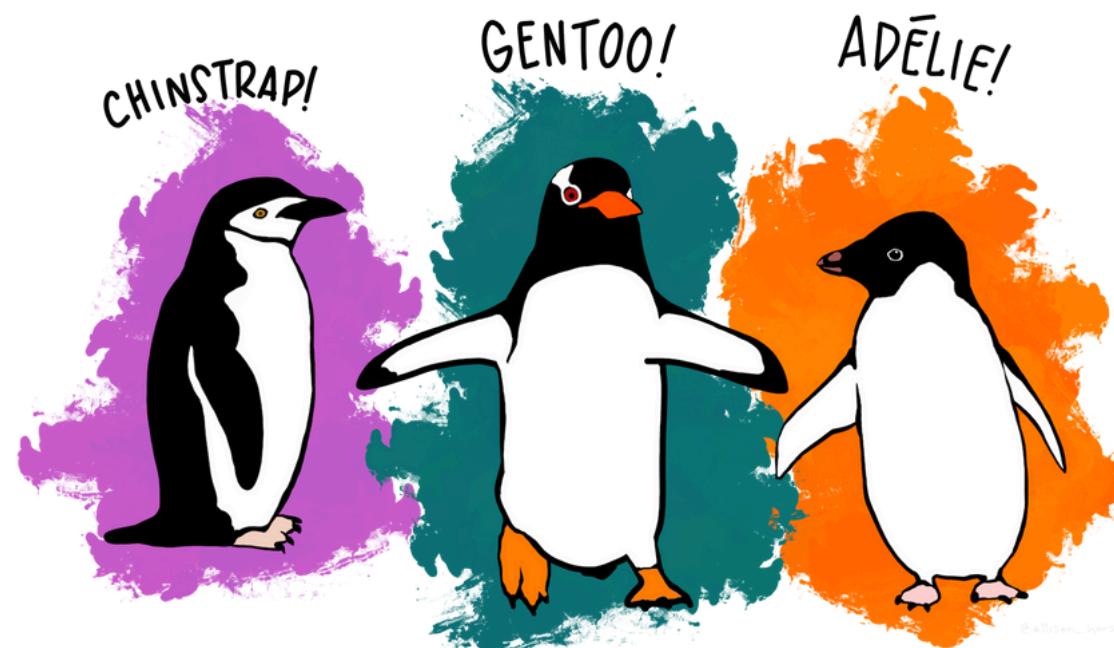


## BAR CHART

200  
150  
100  
50  
0



# CATEGORICAL VARIABLES



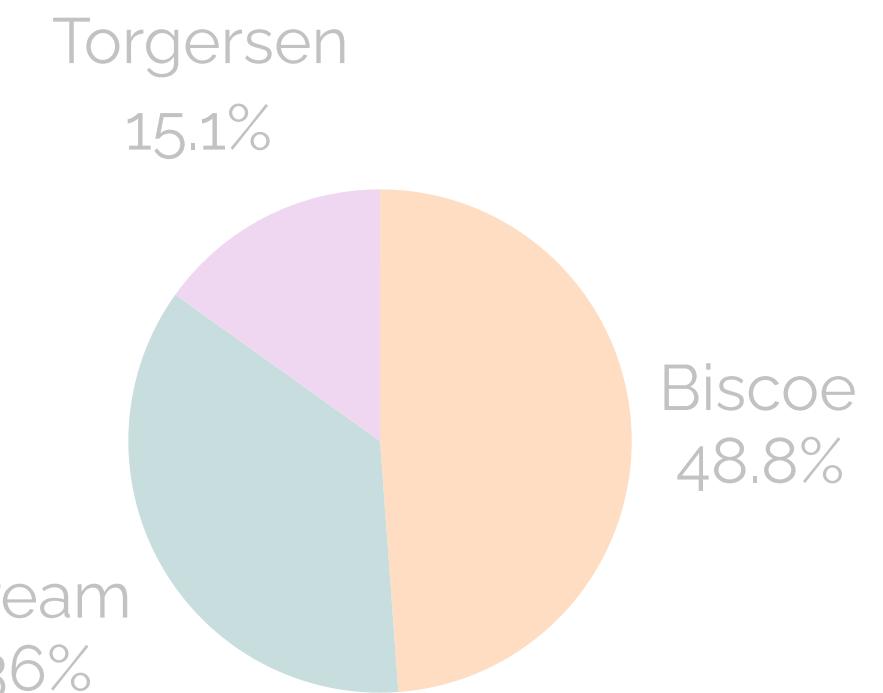
8 variables (n = 344 penguins)

<b>Species</b>	Penguin species
<b>Island</b>	Island in the Palmer
<b>Bill</b>	Length of the
<b>Bill</b>	Depth (thickness) of
<b>Flipper</b>	Length of the
<b>Body</b>	Body mass of the
<b>Sex</b>	Male or female (some)
<b>Year</b>	Year of observation

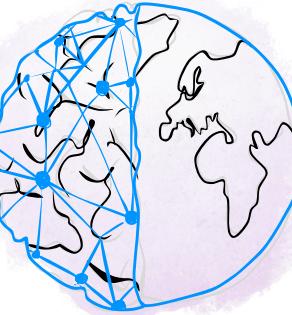
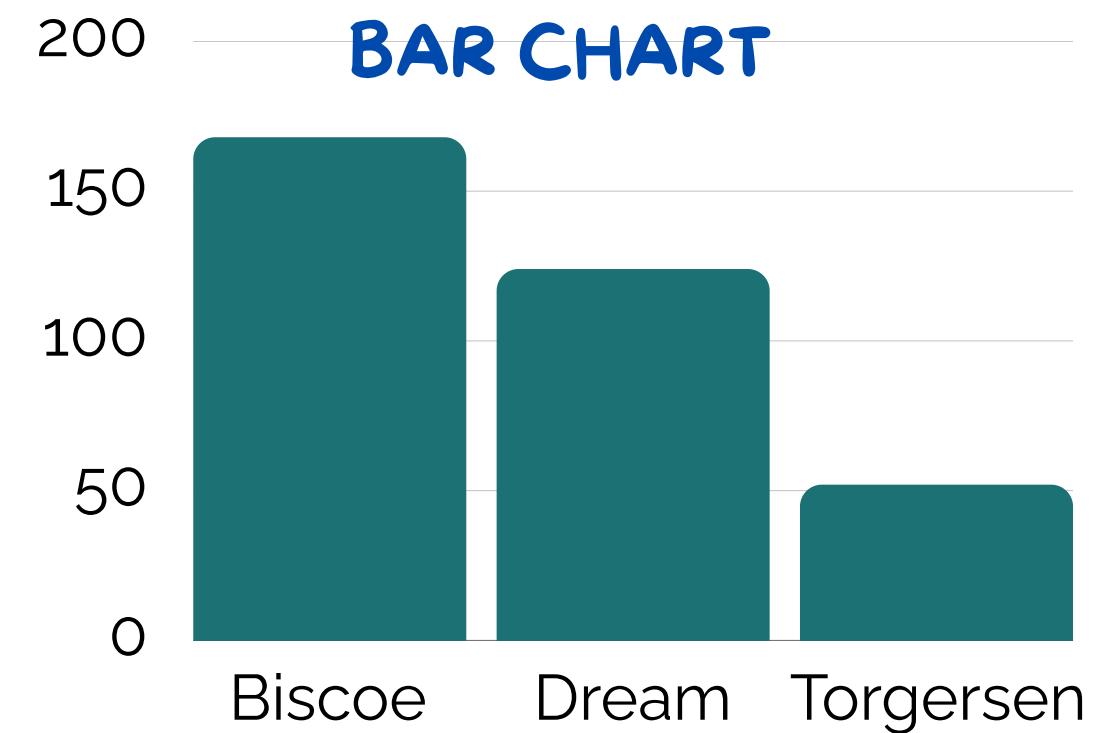
Island	Frequency	Proportion
Biscoe	168	0.489
Dream	124	0.36
Torgersen	52	0.151
<b>TOTAL</b>	<b>344</b>	<b>1</b>

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

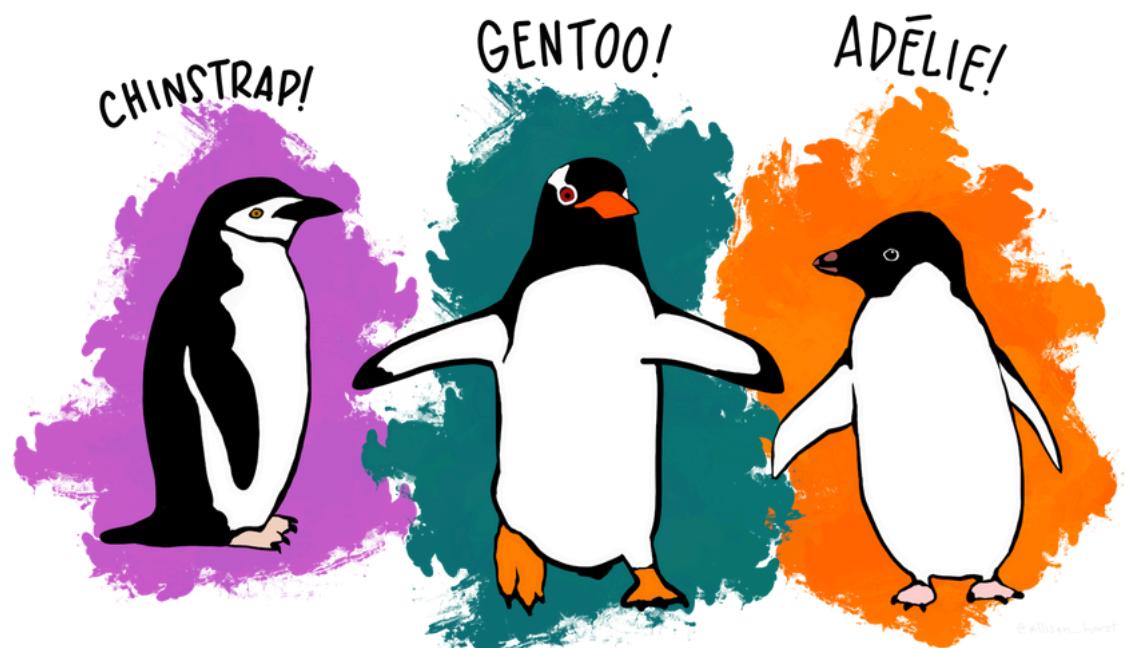
## PIE CHART



## BAR CHART



# NUMERIC VARIABLES



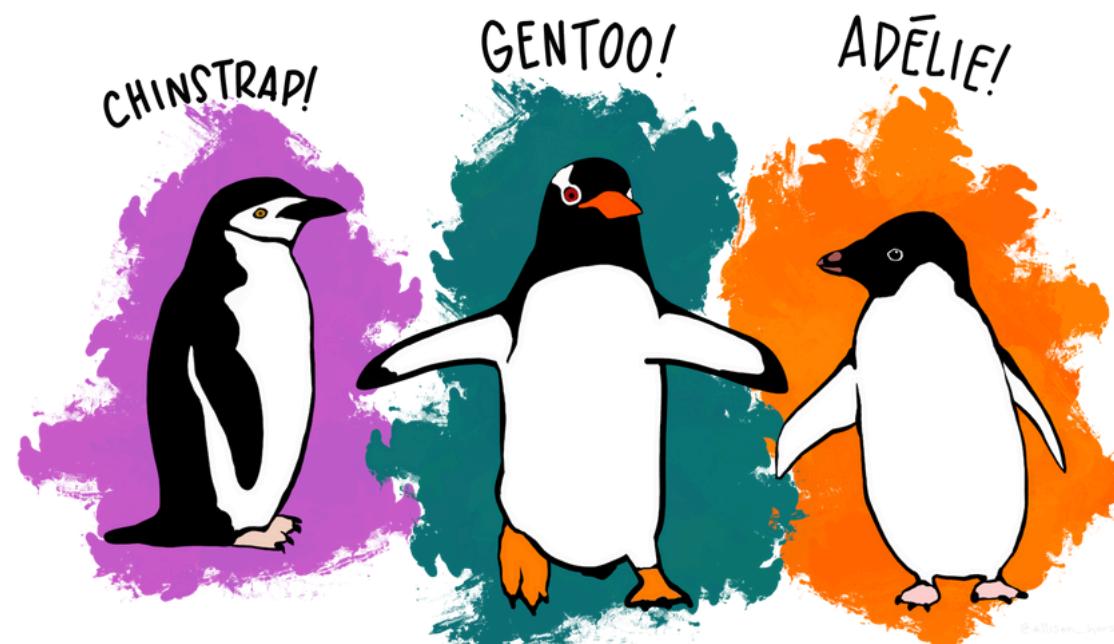
The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

8 variables (n = 344 penguins)

<b>Species</b>	Penguin species (Adélie, Chinstrap, Gentoo)
<b>Island</b>	Island in the Palmer Archipelago where observed
<b>Bill length (mm)</b>	Length of the penguin's bill (mm)
<b>Bill depth (mm)</b>	Depth (thickness) of the penguin's bill (mm)
<b>Flipper length (mm)</b>	Length of the penguin's flipper (mm)
<b>Body mass (g)</b>	Body mass of the penguin (g)
<b>Sex</b>	Male or female (some values missing)
<b>Year</b>	Year of observation (2007–2009)

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# NUMERIC VARIABLES



8 variables (n = 344 penguins)

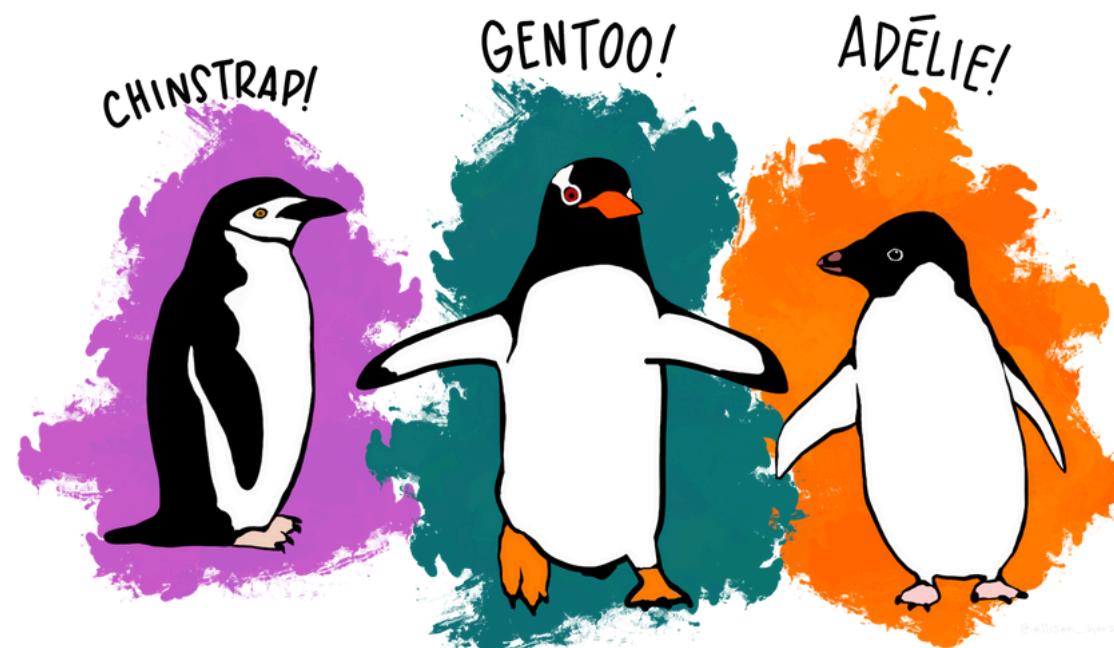
The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

73 male Adelie penguins  
 min body mass = 3325 g  
 max body mass = 4775 g

<b>Species</b>	Penguin species (Adélie, Chinstrap, Gentoo)
<b>Island</b>	Island in the Palmer Archipelago where observed
<b>Bill length (mm)</b>	Length of the penguin's bill (mm)
<b>Bill depth (mm)</b>	Depth (thickness) of the penguin's bill (mm)
<b>Flipper length (mm)</b>	Length of the penguin's flipper (mm)
<b>Body mass (g)</b>	Body mass of the penguin (g)
<b>Sex</b>	Male or female (some values missing)
<b>Year</b>	Year of observation (2007–2009)

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# NUMERIC VARIABLES



8 variables (n = 344 penguins)

Speci	Penguin
Island	Island in the Palmer Archipelago
Bill	Length of the bill
Bill depth	Depth of the bill
Flippe	Length of the flipper
Body	Body mass of the penguin
Sex	Male or female
Year	Year of collection

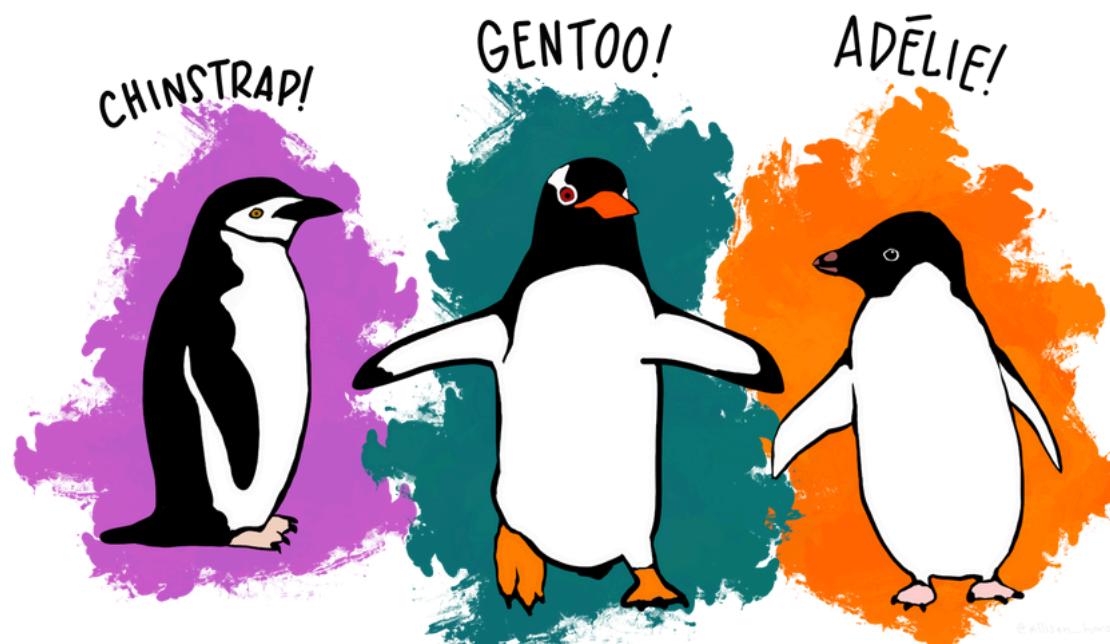
The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

73 male Adelie penguins  
 min body mass = 3325 g  
 max body mass = 4775 g

	Body mass (g)	Frequency	Proportion
	[3325...3615)	9	
	[3615...3905)	20	
	[3905..4195)	18	
	[4195...4485)	18	
	[4485...4775)	8	
	<b>TOTAL</b>	<b>73</b>	

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# NUMERIC VARIABLES



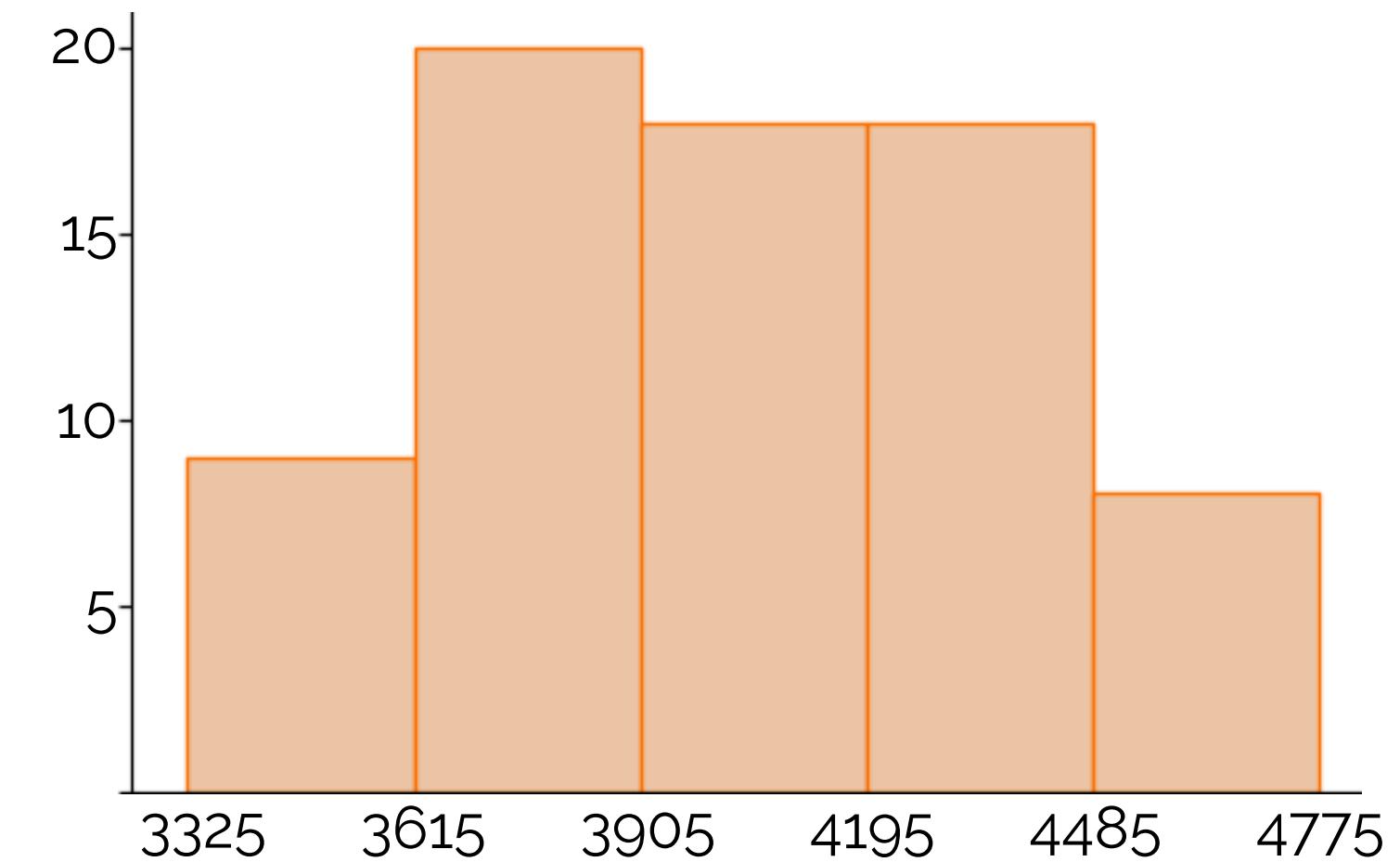
8 variables (n = 344 penguins)

Speci	Penguin
Island	Island in the Palmer Archipelago
Bill	Length of the penguin's bill
Bill depth	Depth of the penguin's bill
Flippe	Length of the penguin's flipper
Body	Body mass of the penguin
Sex	Male or female
Year	Year of observation

Body mass (g)	Frequency	Proportion
[3325...3615)	9	0.123
[3615...3905)	20	0.274
[3905..4195)	18	0.247
[4195...4485)	18	0.247
[4485...4775)	8	0.11
<b>TOTAL</b>	<b>73</b>	<b>1</b>

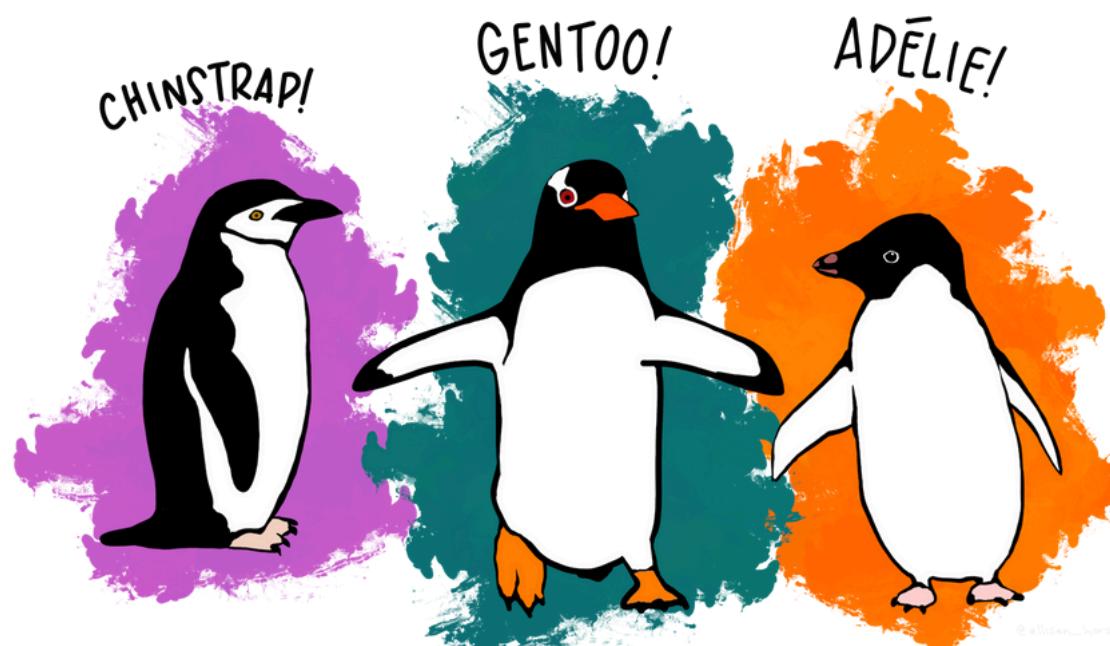
The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

## HISTOGRAM



These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# NUMERIC VARIABLES



8 variables (n = 344 penguins)

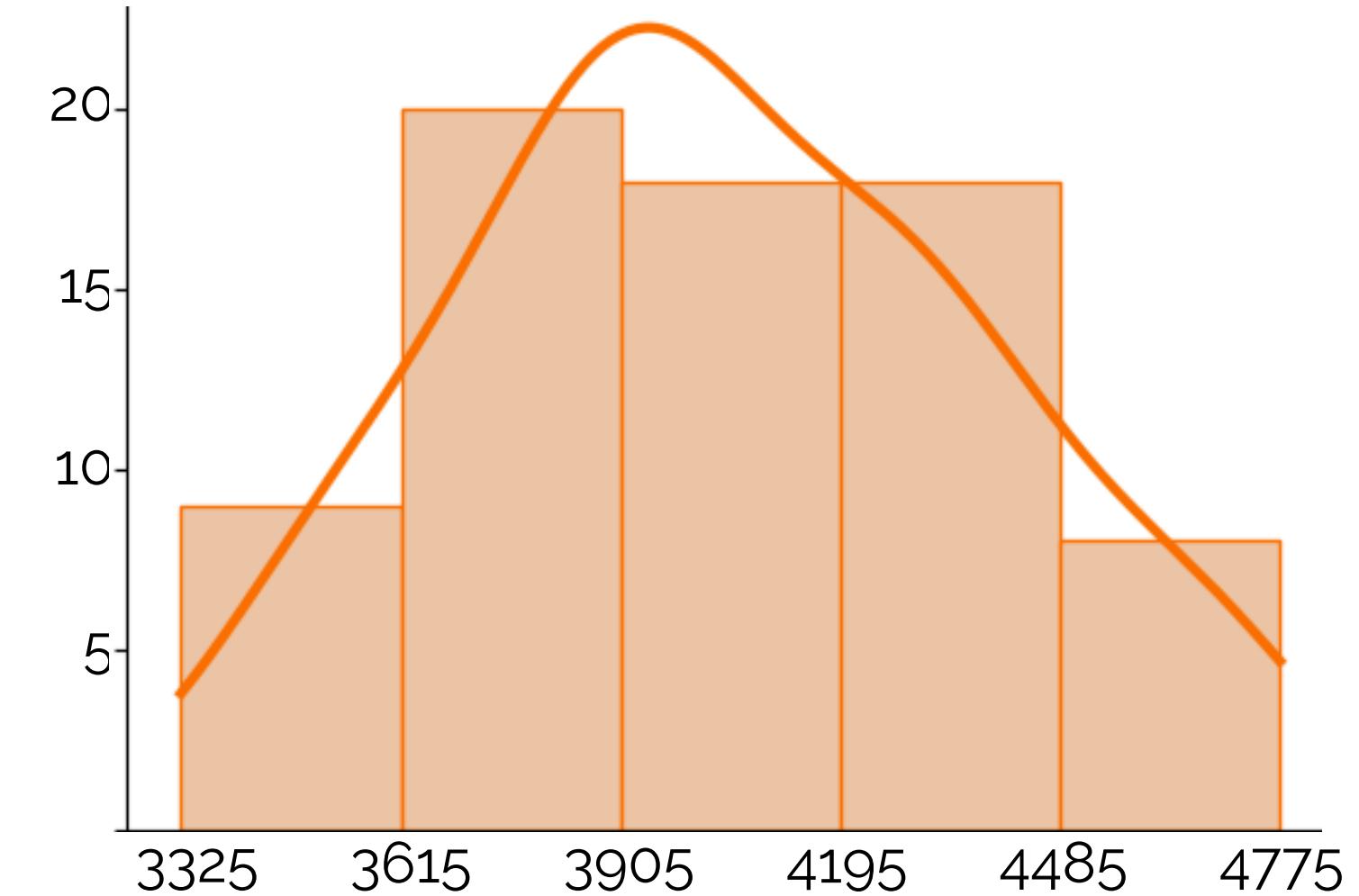
Speci	Penguin
Island	Island in the Palmer Archipelago
Bill	Length of the penguin's bill
Bill depth	Depth of the penguin's bill
Flippe	Length of the penguin's flipper
Body	Body mass of the penguin
Sex	Male or female
Year	Year of collection

Body mass (g)	Frequency	Proportion
[3325...3615)	9	0.123
[3615...3905)	20	0.274
[3905..4195)	18	0.247
[4195...4485)	18	0.247
[4485...4775)	8	0.11
<b>TOTAL</b>	<b>73</b>	<b>1</b>

The **palmerpenguins** data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.

73 male Adelie penguins  
min body mass = 3325 g  
max body mass = 4775 g

## KERNEL DENSITY (PLOT)



These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# ESTIMATES OF LOCATION MEASURE OF CENTRAL TENDENCY

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS



# ESTIMATES OF LOCATION MEASURE OF CENTRAL TENDENCY

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
 $\&$   
IQR

SKEWNESS

MODE

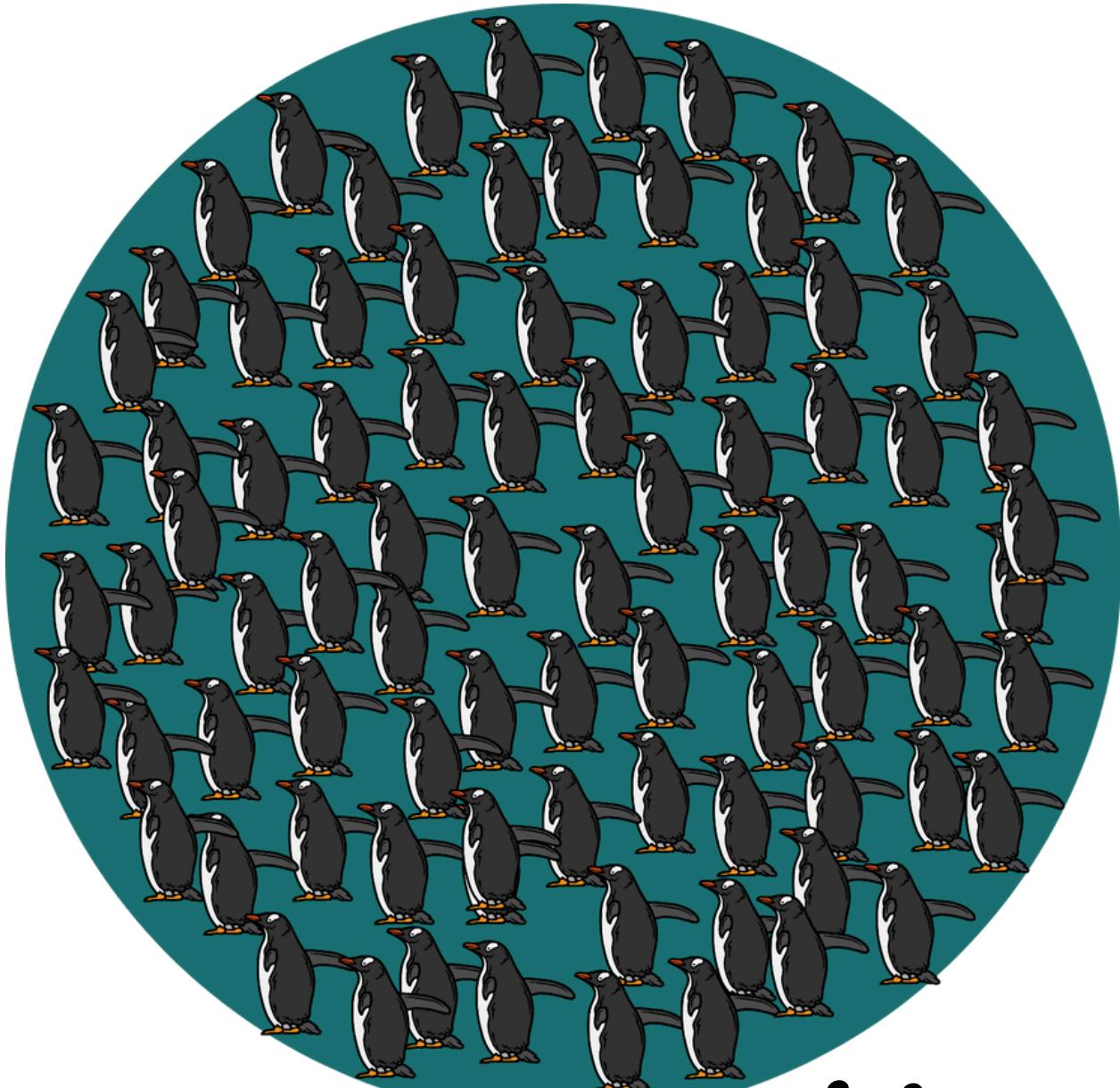
VARIANCE  
 $\&$   
STD DEV

KURTOSIS



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



POPULATION  $\mu$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQR

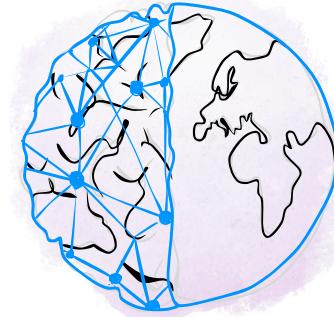
VARIANCE  
&  
STD DEV

COEFFICIENT  
OF  
VARIATION

SKEWNESS

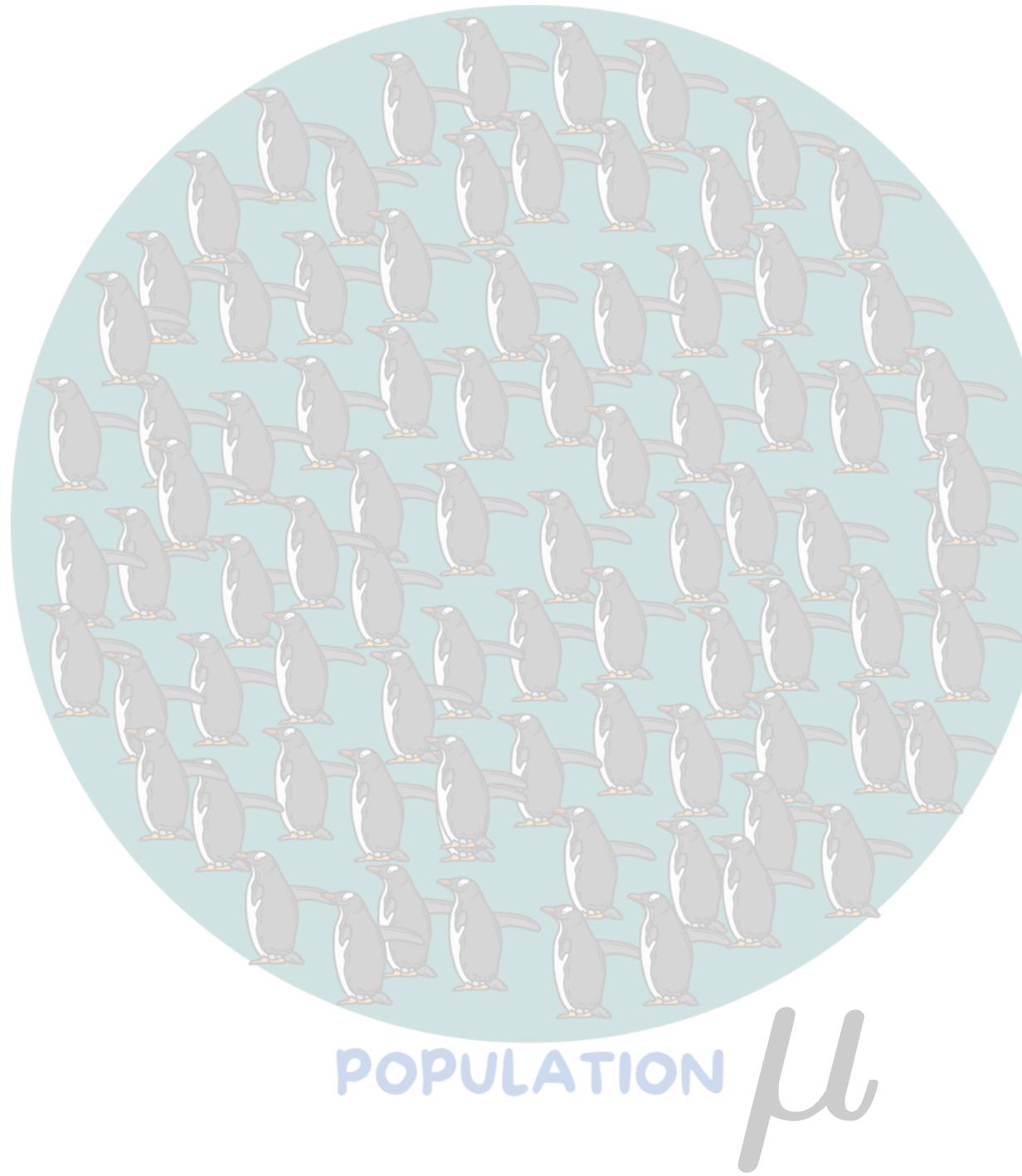
KURTOSIS

$$\mu = \frac{\sum x}{N}$$



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

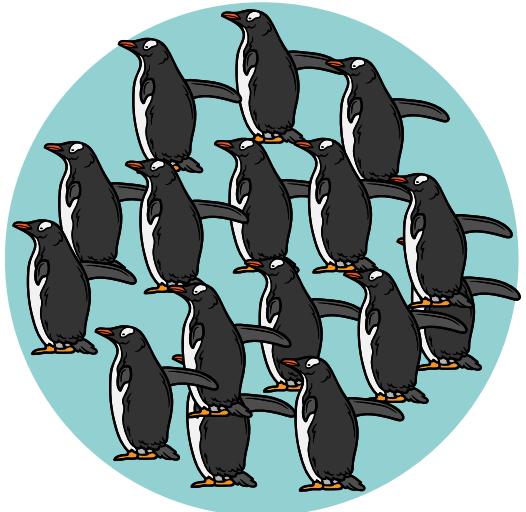
KURTOSIS

$$\mu = \frac{\sum x}{N}$$



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

SKEWNESS

MODE

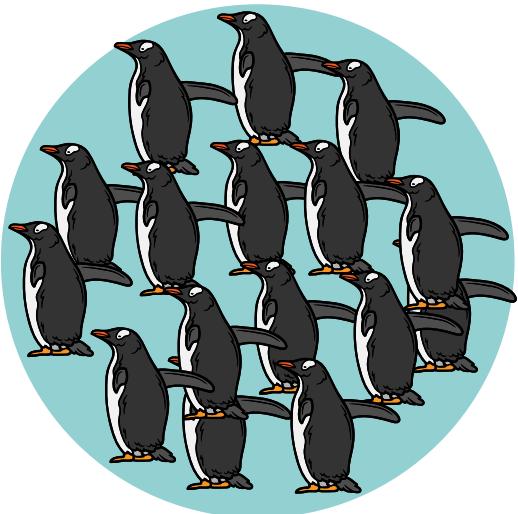
VARIANCE  
&  
STD DEV

KURTOSIS



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

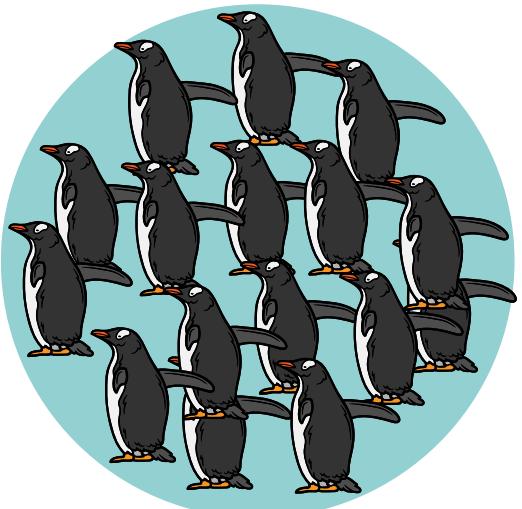


[10, 28, 28, 33, 54]

$$\bar{x} = \frac{10 + 28 + 28 + 33 + 54}{5} = 30.6$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

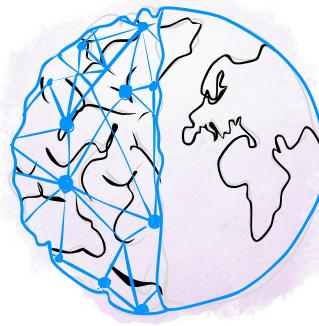
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



### WEIGHTED MEAN

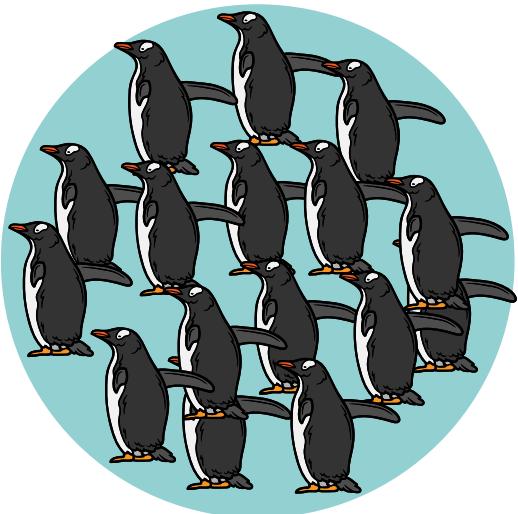
[10, 28, 28, 33, 54]

$$\bar{x} = \frac{10 + 28 + 28 + 33 + 54}{5}$$

x	Frequency
10	1
28	2
33	1
54	1

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



### WEIGHTED MEAN

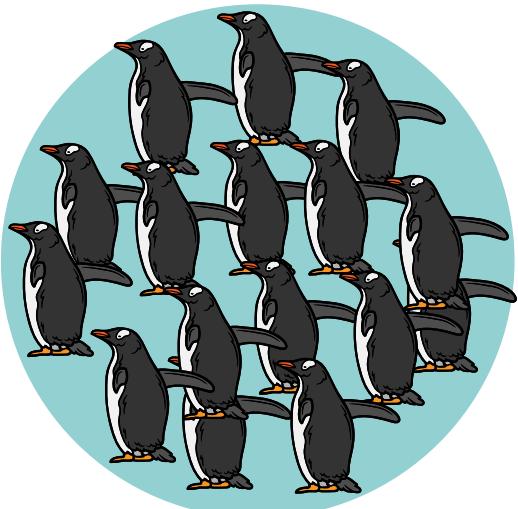
[10, 28, 28, 33, 54]

$$\bar{x} = \frac{10 + 2 \times 28 + 33 + 54}{5}$$

x	Frequency
10	1
28	2
33	1
54	1

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

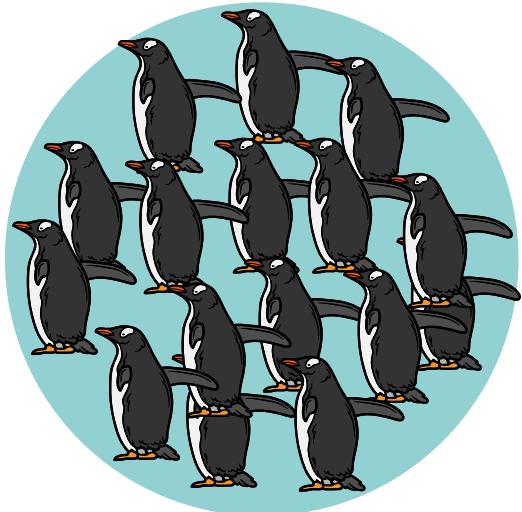
WEIGHTED  
MEAN

[10, 28, 28, 33, 54]

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

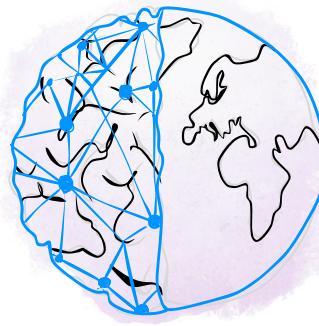
RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS

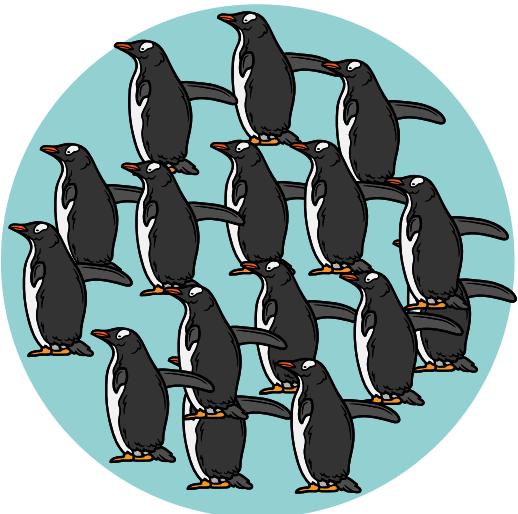


### MEAN OF CATEGORICAL VARIABLE?

[Male, Female, Female, Male, Female, Female]

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



### MEAN OF CATEGORICAL VARIABLE?

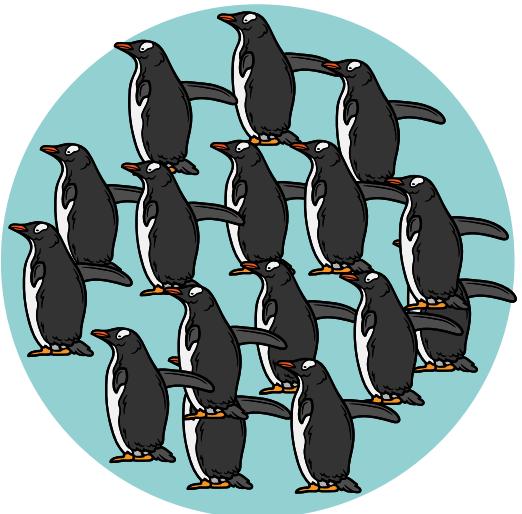
[Male, Female, Female, Male, Female, Female]

[0, 1, 1, 0, 1, 1]

$$\bar{x} = \frac{0 + 1 + 1 + 0 + 1 + 1}{6} = 0.66\ldots$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

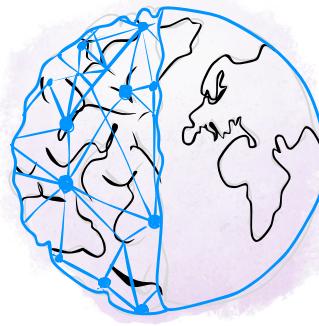
RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS



### MEAN OF BINARY VARIABLE > PROPORTION

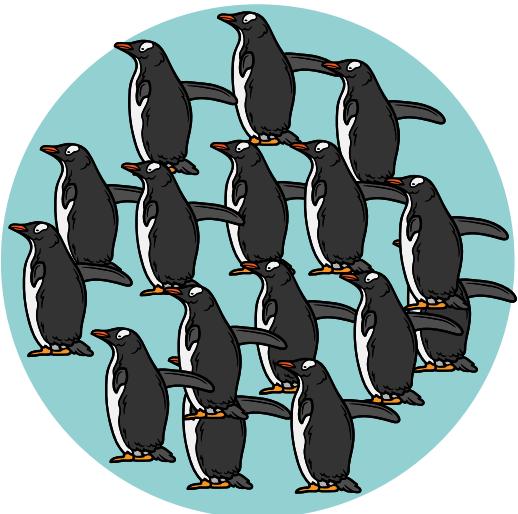
[Male, Female, Female, Male, Female, Female]

[0, 1, 1, 0, 1, 1]

$$\bar{x} = \frac{0 + 1 + 1 + 0 + 1 + 1}{6} = 0.66\ldots$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS

OTHER MEANS

TRIMMED MEAN

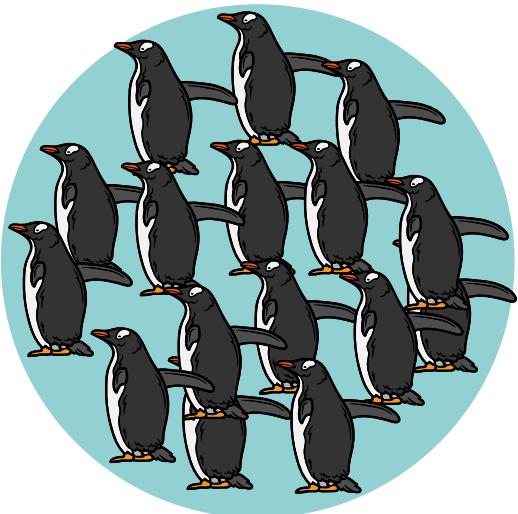
[3, 5, 6, 6, 7.5, 8, 8.5, 9, 37]

$$\bar{x} = 10$$



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS

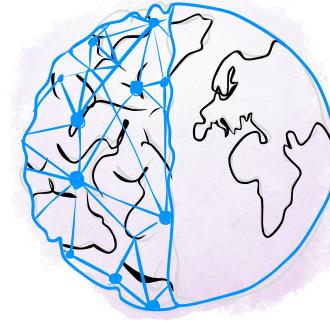
OTHER MEANS

TRIMMED MEAN

[3, 5, 6, 6, 7.5, 8, 8.5, 9, 37]

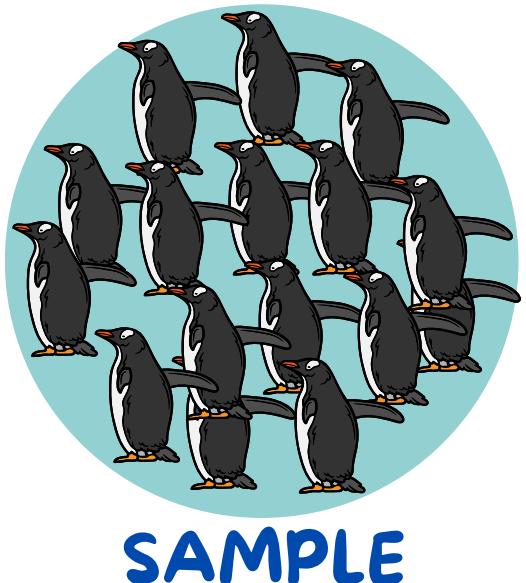
$$\bar{x} = 10$$

$$\bar{x}_{\text{trimmed}} = 7.14$$



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

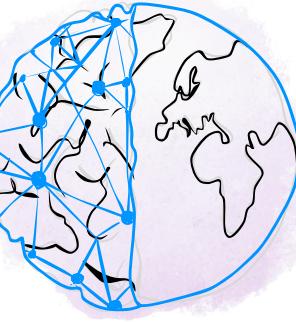
SKEWNESS

MODE

VARIANCE  
&  
STD DEV

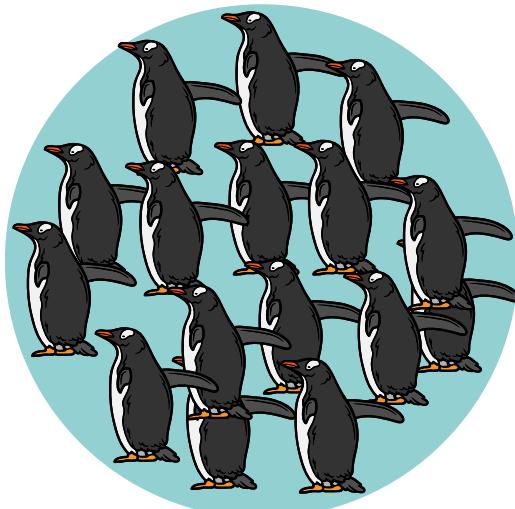
KURTOSIS

OTHER MEANS  
GEOMETRIC MEAN



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

RANGE  
&  
IQR

SKEWNESS

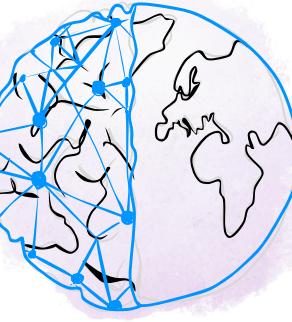
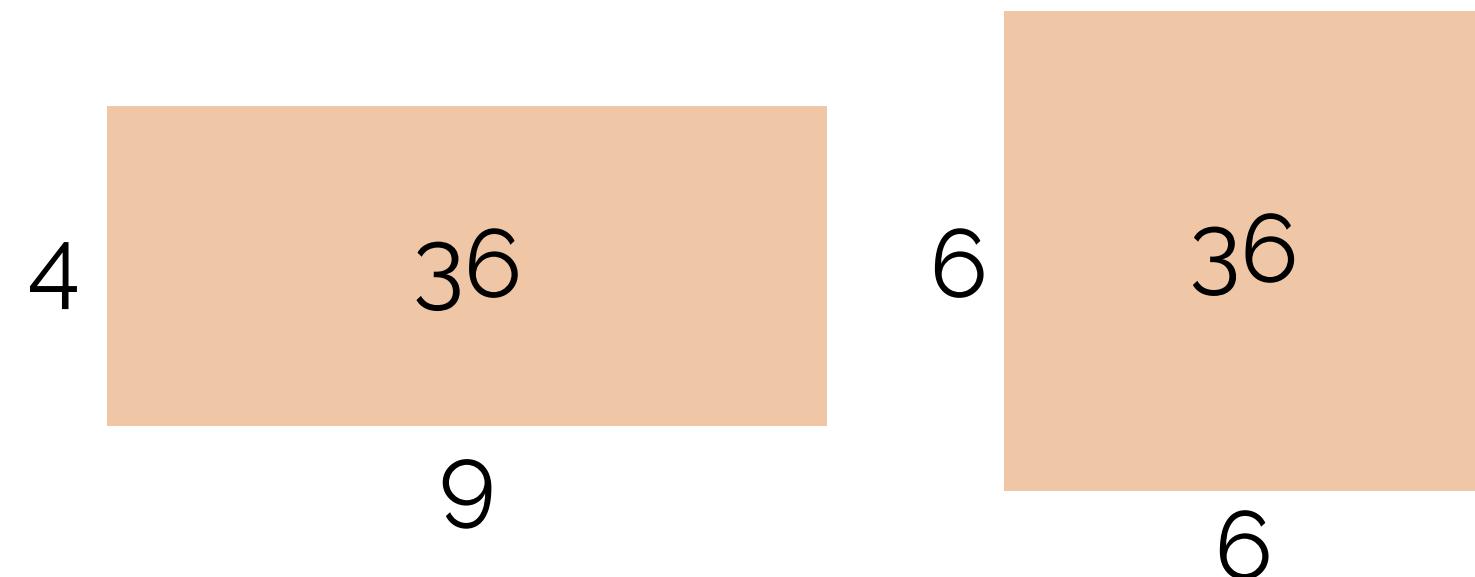
MODE

VARIANCE  
&  
STD DEV

KURTOSIS

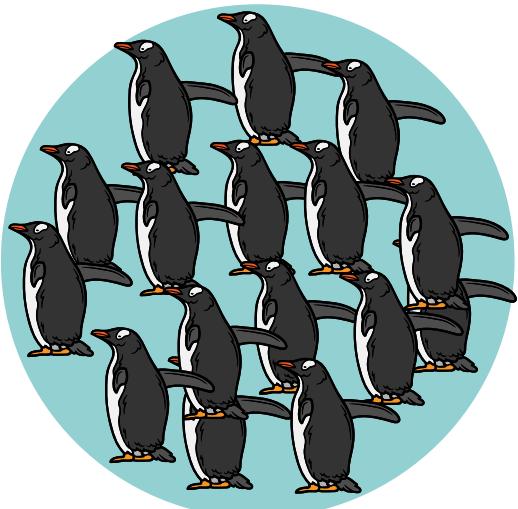
OTHER MEANS  
GEOMETRIC MEAN

$$gm = \sqrt[n]{\prod_{i=1}^n x_i}$$



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

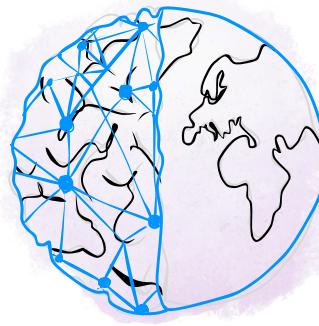
RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS



Day	E. coli in water (cfu/mL)
1	5
2	50
3	500

OTHER MEANS  
GEOMETRIC MEAN

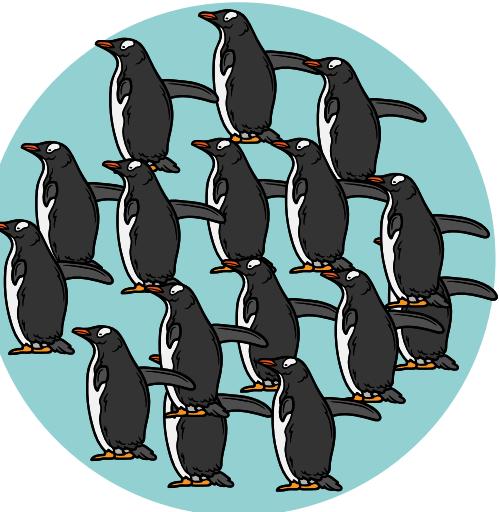
Calculating the average of ratios

$$gm = \sqrt[3]{5 \times 50 \times 500} \approx 50$$

$$gm = \sqrt[n]{\prod_{i=1}^n x_i}$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

MEDIAN

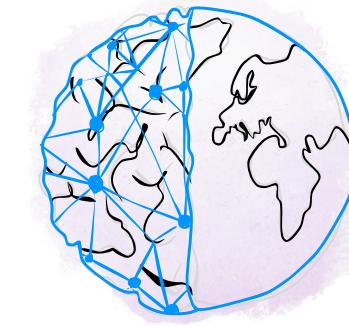
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



Day	E. coli in water (cfu/mL)
1	5
2	50
3	500

$$gm = \sqrt[3]{5 \times 50 \times 500} \approx 50$$

OTHER MEANS  
GEOMETRIC MEAN

Calculating the average of ratios

$$\log(gm) = \frac{1}{n} \sum_{i=1}^n \log x_i \Rightarrow gm = \exp \left( \frac{1}{n} \sum_{i=1}^n \log x_i \right)$$

$$gm = (x_1 \times x_2 \times \dots \times x_n)^{\frac{1}{n}}$$

$$\frac{\log 5 + \log 50 + \log 500}{3} \approx 3.91$$

$$\exp(3.91) \approx 50$$

$$gm = \sqrt[n]{\prod_{i=1}^n x_i}$$

The geometric mean (gm) is equivalent to log-transforming your data, then calculating the arithmetic mean, and transforming the result back (with the antilog exponent.)

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

SAMPLE

MEAN

QUANTILES

COEFFICIENT  
OF  
VARIATION

MEDIAN

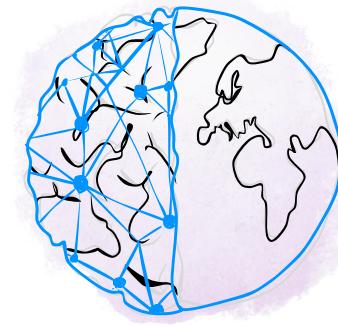
RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

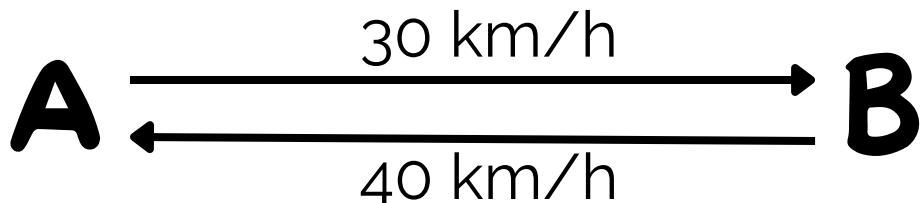
KURTOSIS



OTHER MEANS  
HARMONIC MEAN

Appropriate for averaging rates

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



$$\bar{x} = \frac{\sum x}{n}$$

ARITHMETIC  
MEAN

MEAN

MEDIAN

MODE

QUANTILES

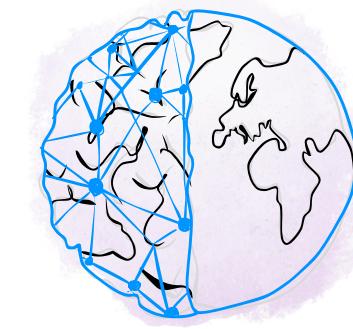
RANGE  
&  
IQR

VARIANCE  
&  
STD DEV

COEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

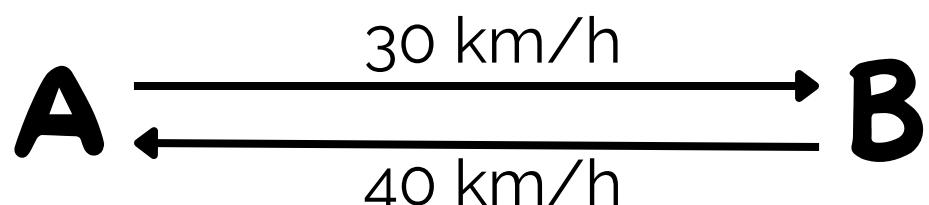


OTHER MEANS

HARMONIC MEAN

Appropriate for averaging rates

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

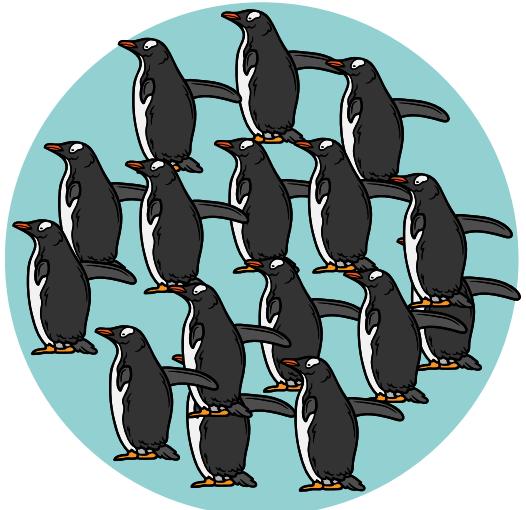


$$H = \frac{2}{\frac{1}{30} + \frac{1}{40}} \approx 34.3$$

$$H = \left( \frac{\sum_{i=1}^n \frac{1}{x_i}}{n} \right)^{-1}$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
¶  
IQRVARIANCE  
¶  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

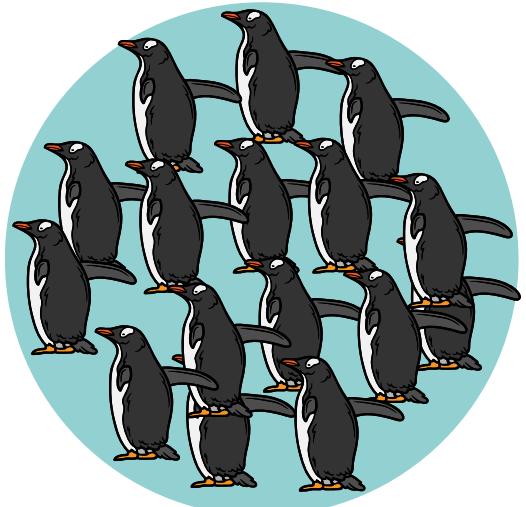
KURTOSIS



[10, 28, 28, 33, 54]

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

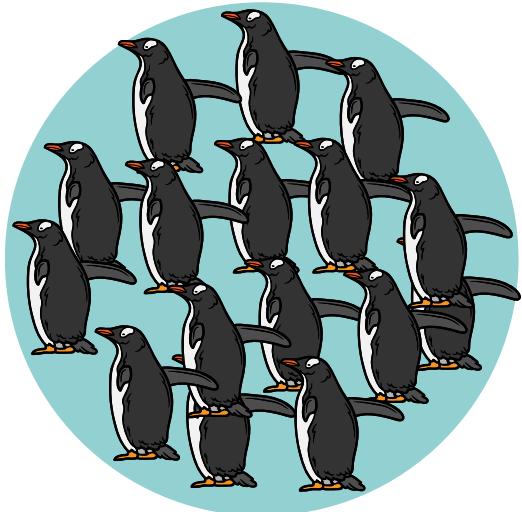


[10, 28, 28, 33, 54]

 $Med = 28$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

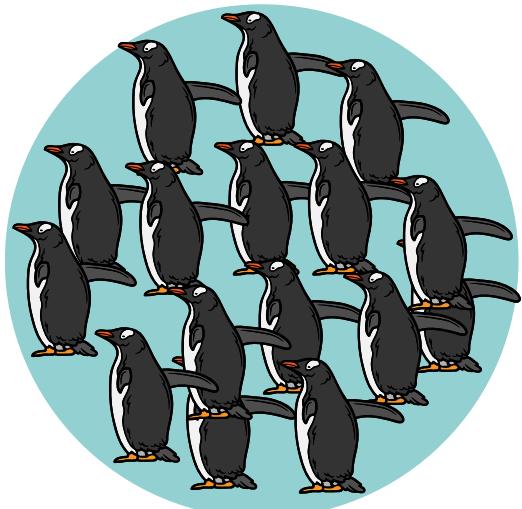
KURTOSIS



[10, 28, 28, 33, 54, 59]

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



Middle value of ordered observations (50% below, 50% above)

[10, 28, 28, 33, 54, 59]

$$Med = \frac{28 + 33}{2} = 14.5$$

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

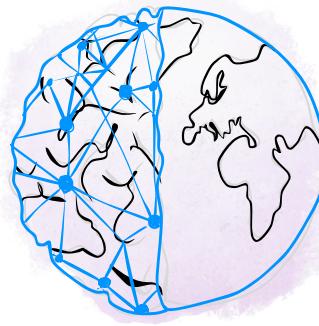
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



Mean:

- sensitive to outliers
- “balance” point

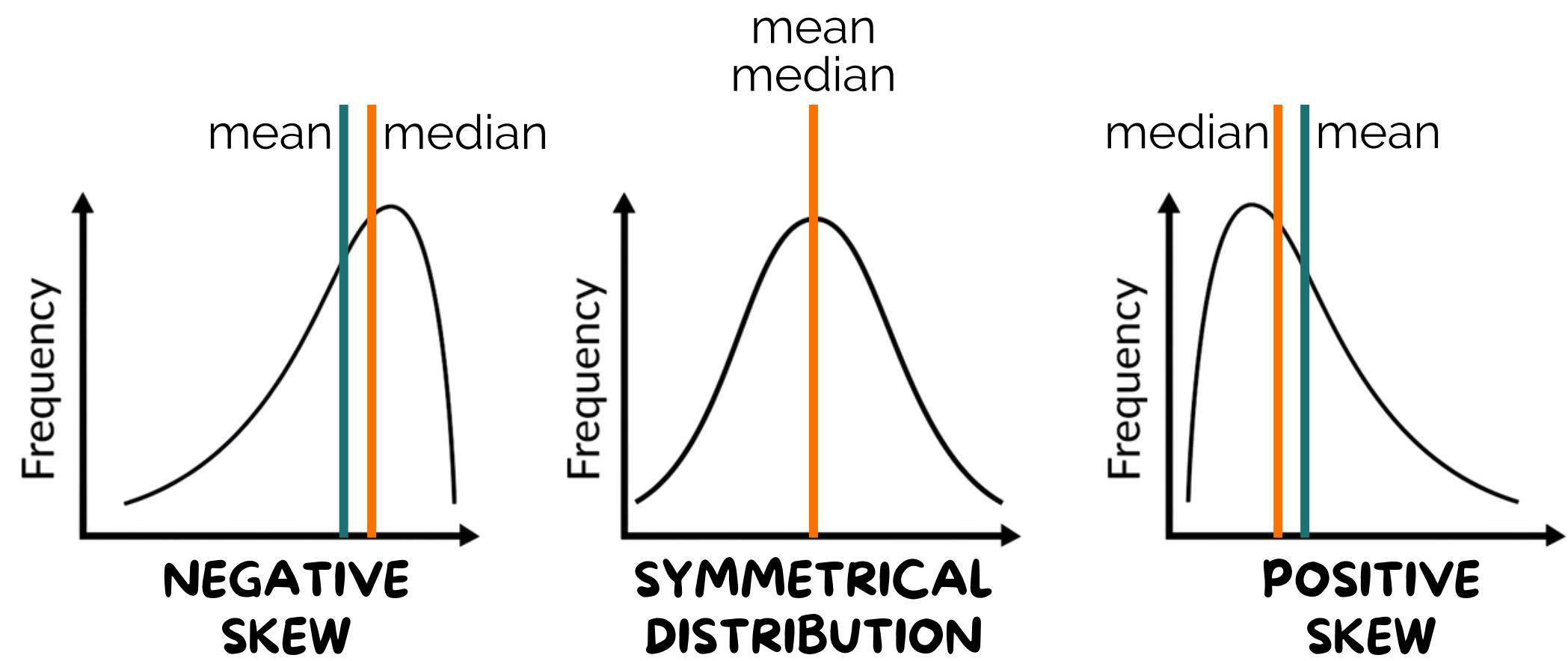
Median:

- not sensitive to outliers
- cuts data in half

[10, 28, 28, 33, 54]

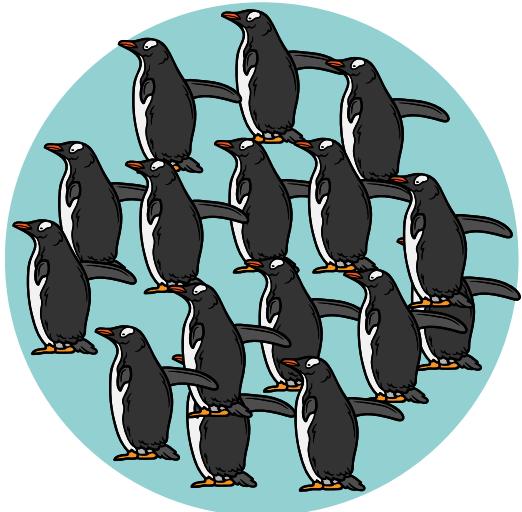
$$\bar{x} = \frac{10 + 28 + 28 + 33 + 54}{5} = 30.6$$

Med = 28



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

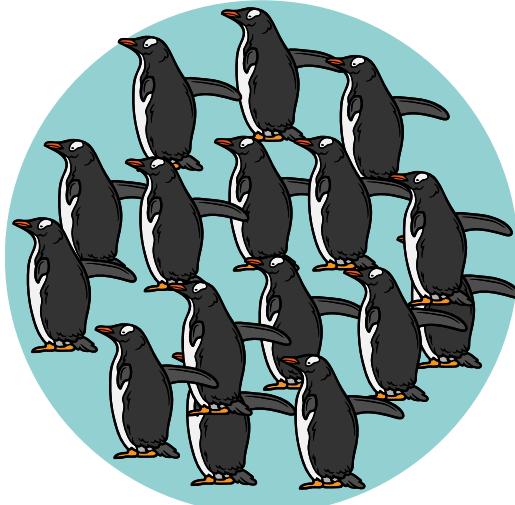


The observation with the highest frequency

[10, 28, 28, 33, 54]

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



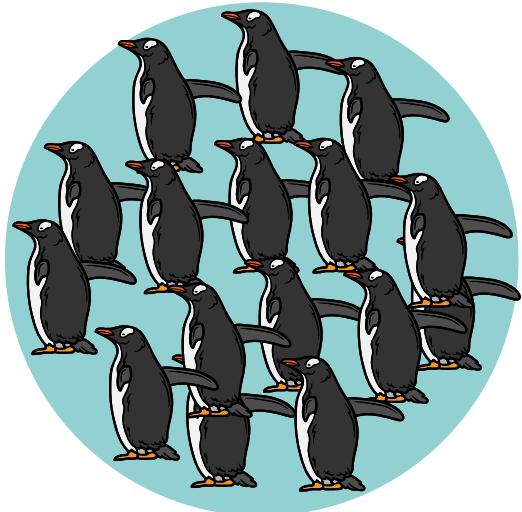
The observation with the highest frequency

[10, 28, 28, 33, 54]

*Mode* = 28

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



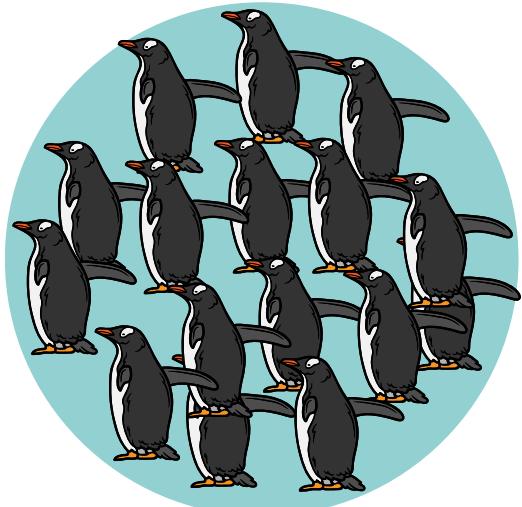
The observation with the highest frequency

[10, 27, 28, 33, 54]

*Mode* = ?

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



The observation with the highest frequency

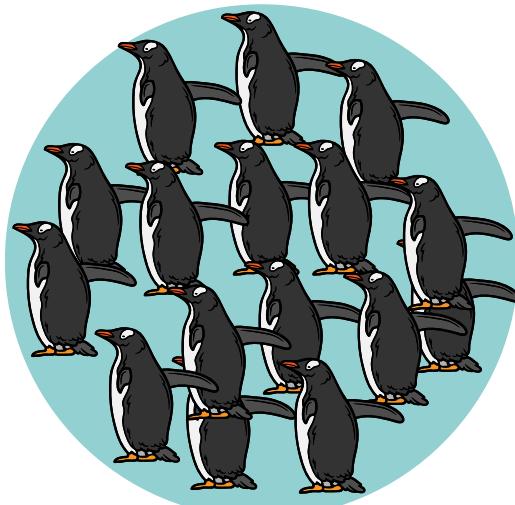
[10, 27, 28, 54, 54]

*Mode* = 54

Not suitable for small datasets

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

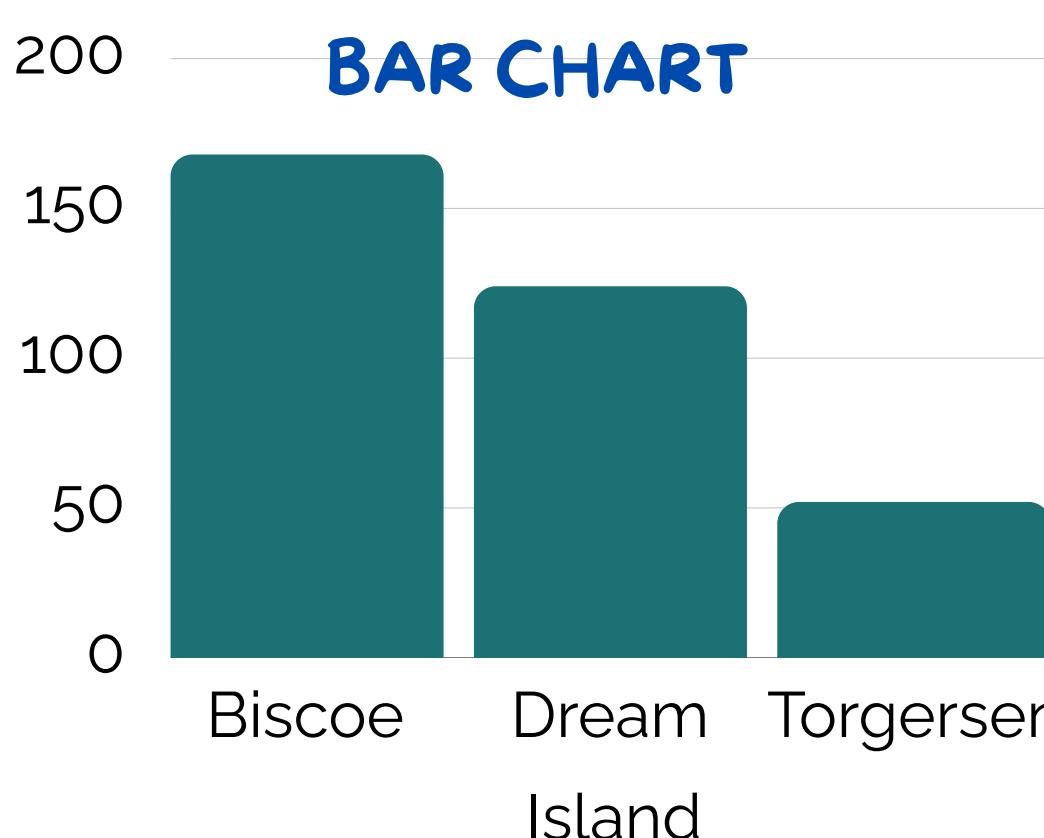
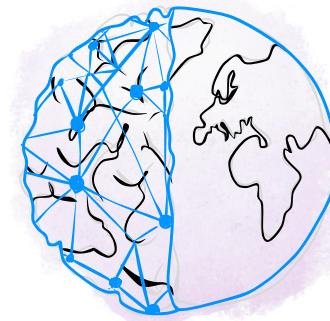
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



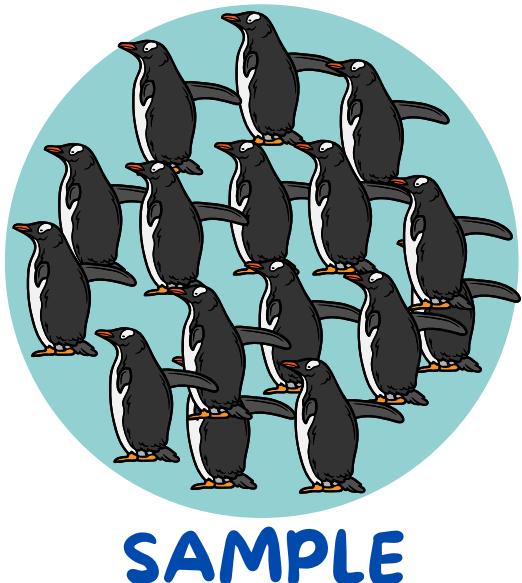
The observation with the highest frequency

*Mode = Biscoe*

These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



$$\bar{x} = \frac{\sum x}{n}$$

MEAN

MEDIAN

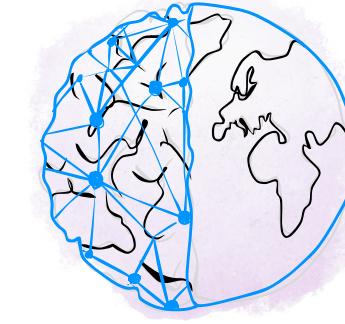
MODE

QUANTILES

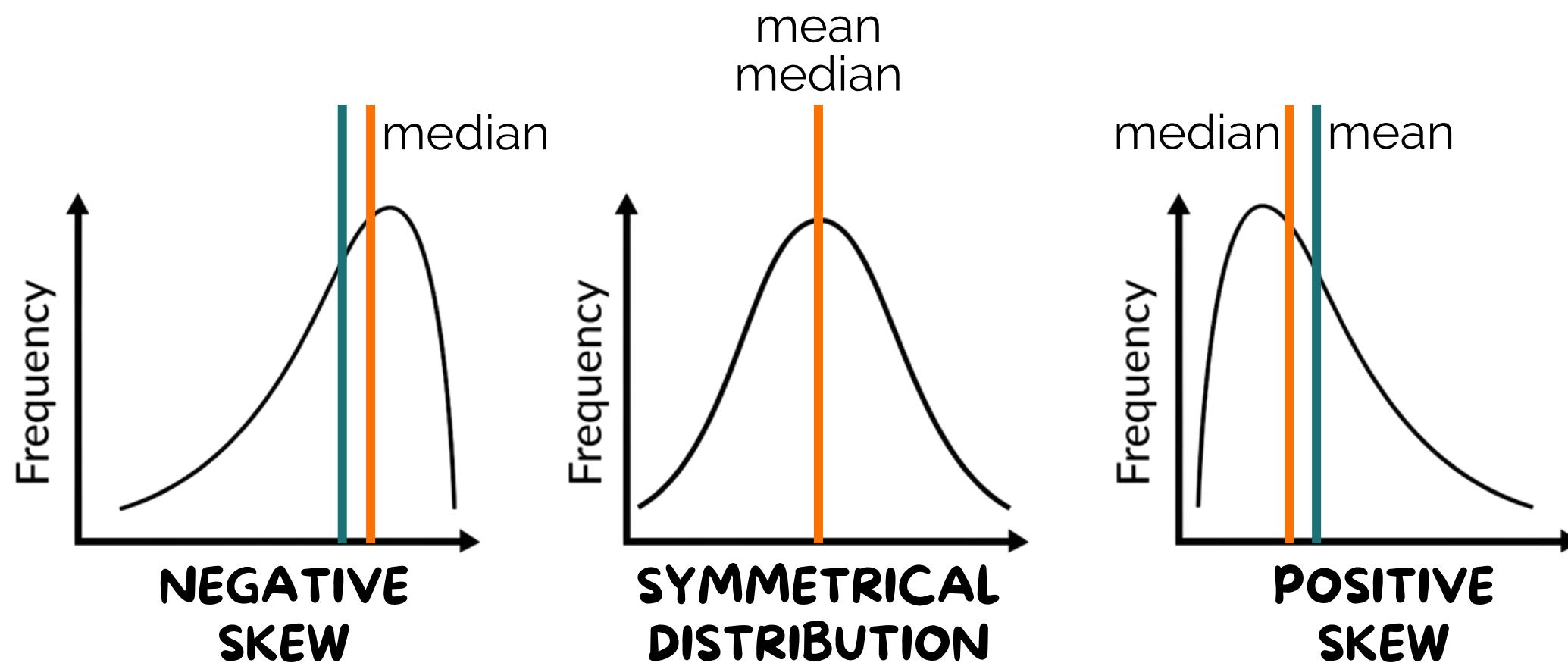
RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



Mode: The observation with the highest frequency



Mean:

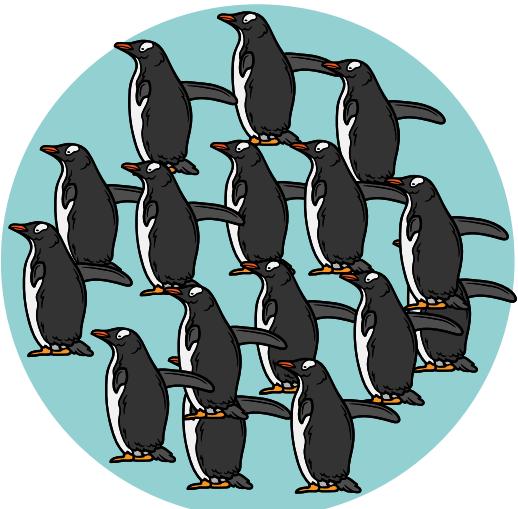
- sensitive to outliers
- “balance” point

Median:

- not sensitive to outliers
- cuts data in half

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

Mean:

- sensitive to outliers
- “balance” point

Median:

- not sensitive to outliers
- cuts data in half

Mode:

The observation with the highest frequency

MEAN

MEDIAN

MODE

QUANTILES

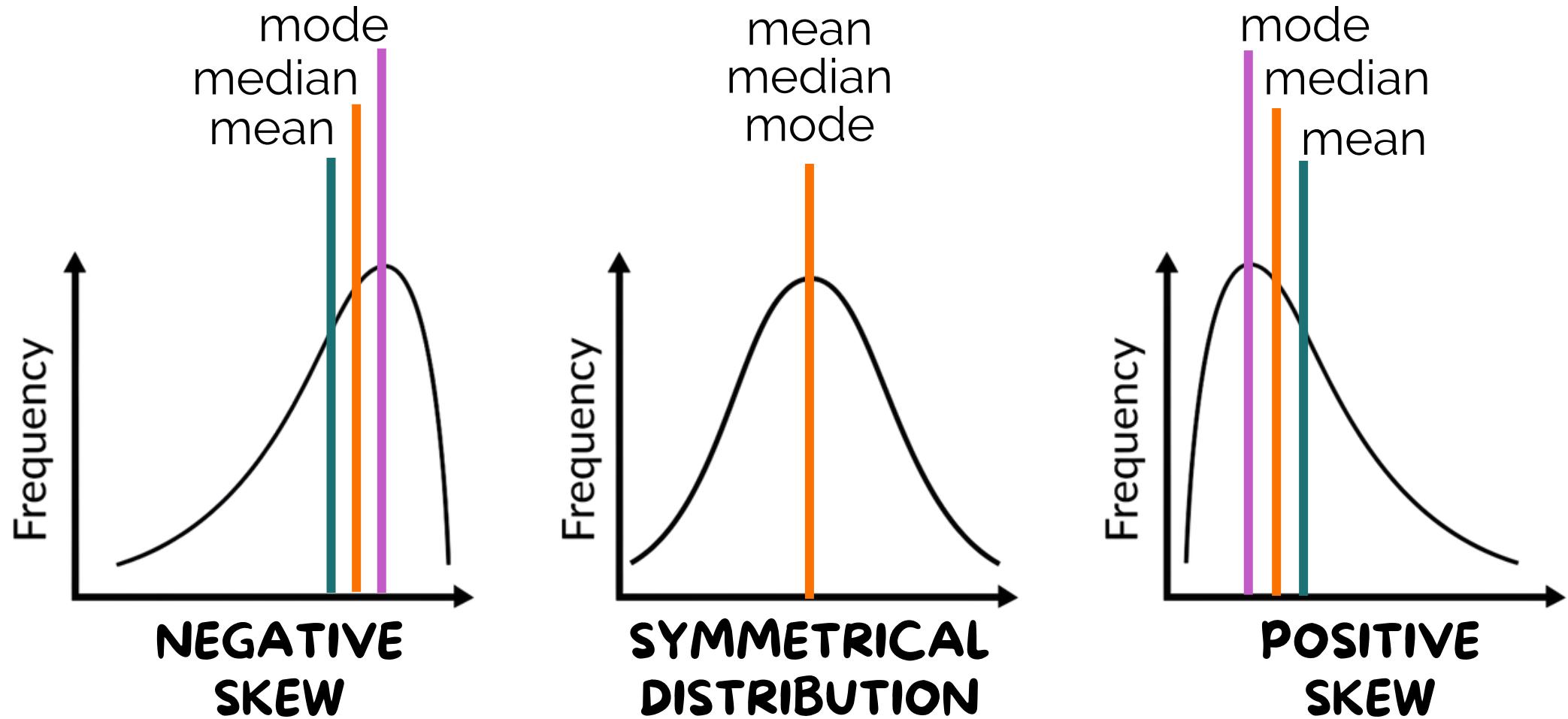
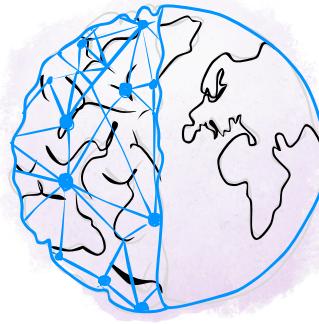
RANGE &amp; IQR

VARIANCE &amp; STD DEV

COEFFICIENT OF VARIATION

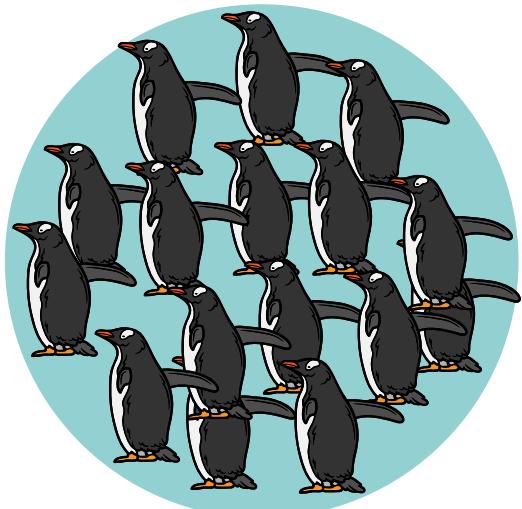
SKEWNESS

KURTOSIS



# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

Mean:

- sensitive to outliers
- “balance” point

Median:

- not sensitive to outliers
- cuts data in half

Mode:

The observation with the highest frequency

MEAN

MEDIAN

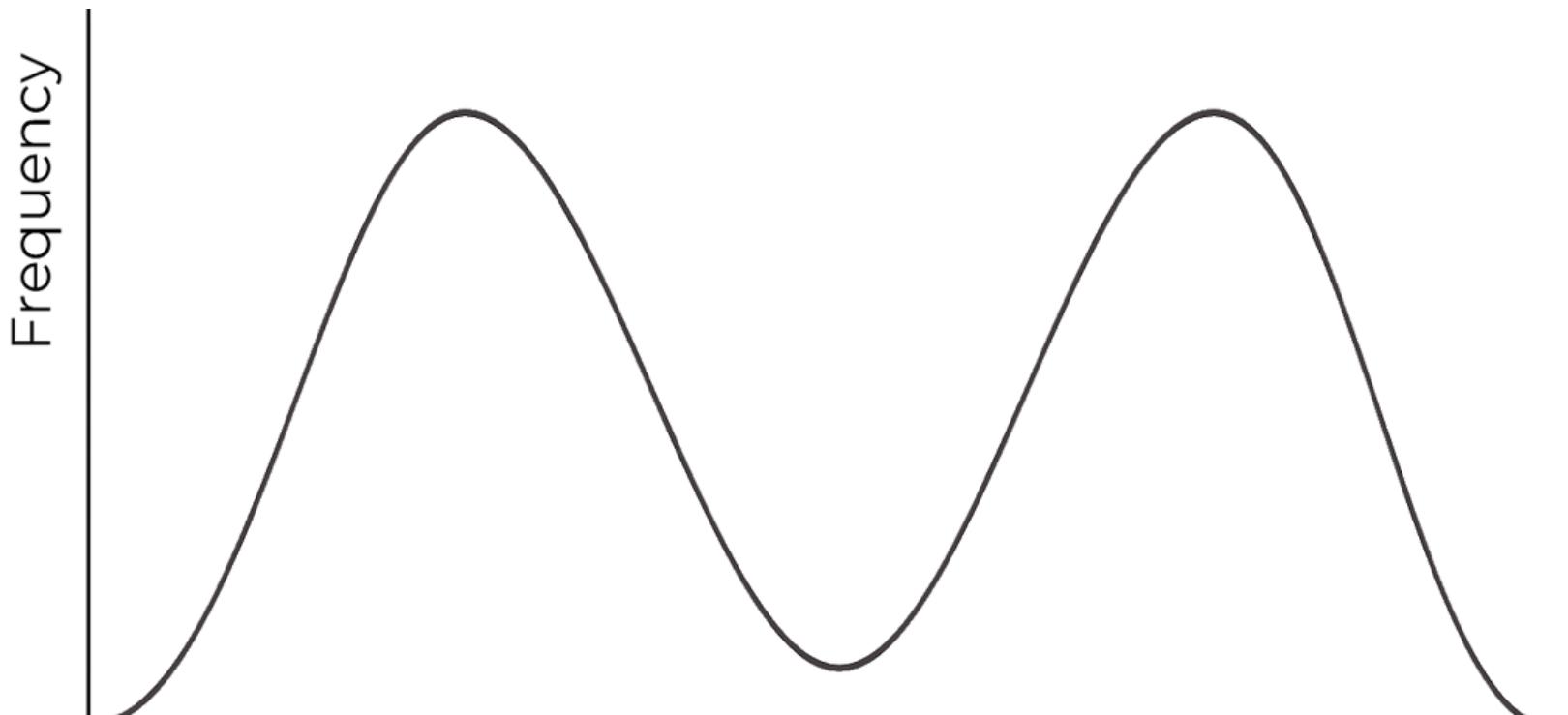
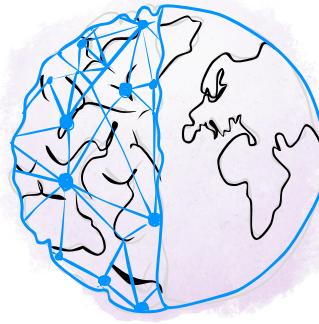
MODE

QUANTILES

RANGE  
¶  
IQRVARIANCE  
¶  
STD DEVCOEFFICIENT  
OF  
VARIATION

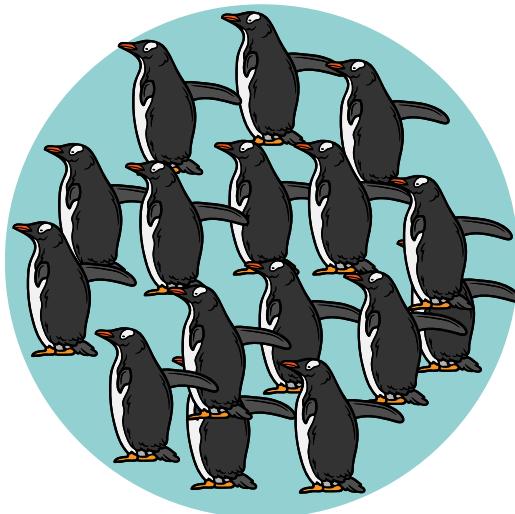
SKEWNESS

KURTOSIS

SYMMETRICAL  
DISTRIBUTION

# ESTIMATES OF LOCATION

## MEASURE OF CENTRAL TENDENCY



SAMPLE

$$\bar{x} = \frac{\sum x}{n}$$

Mean:

- sensitive to outliers
- “balance” point

Median:

- not sensitive to outliers
- cuts data in half

Mode:

The observation with the highest frequency

MEAN

MEDIAN

MODE

QUANTILES

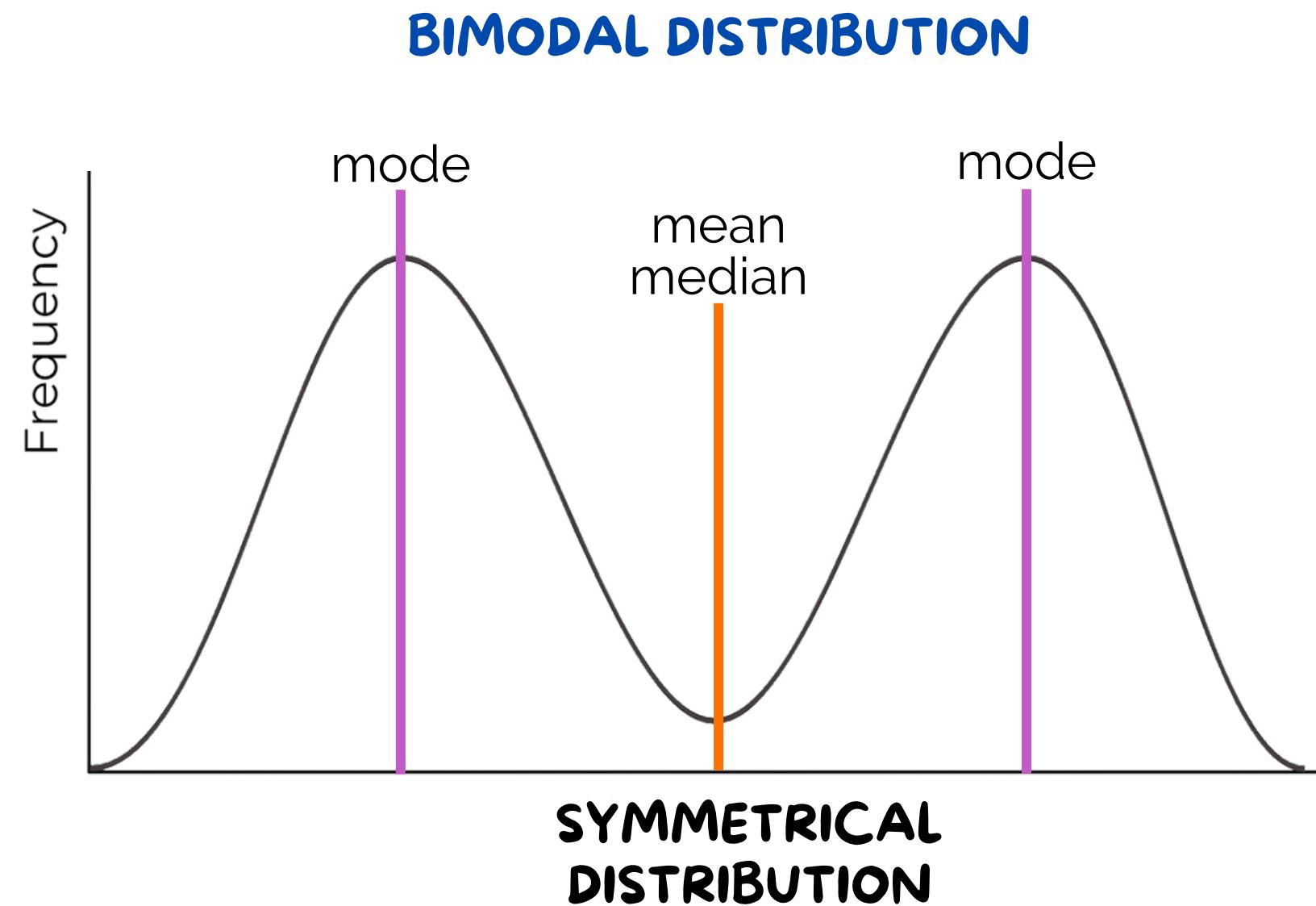
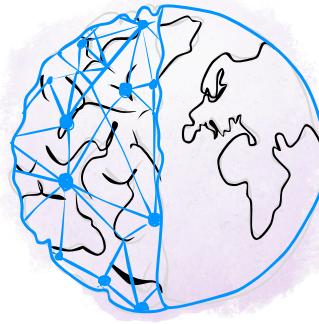
RANGE &amp; IQR

VARIANCE &amp; STD DEV

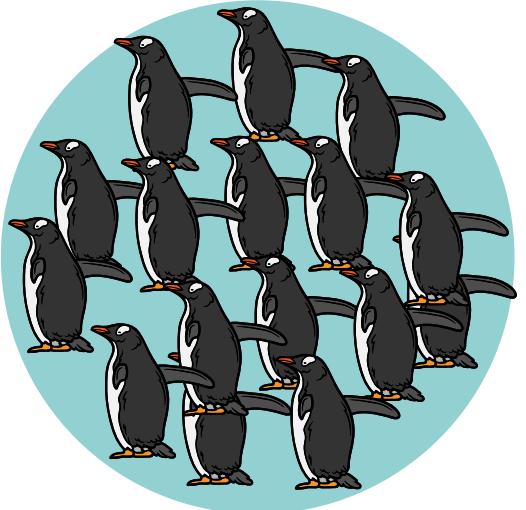
COEFFICIENT OF VARIATION

SKEWNESS

KURTOSIS



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

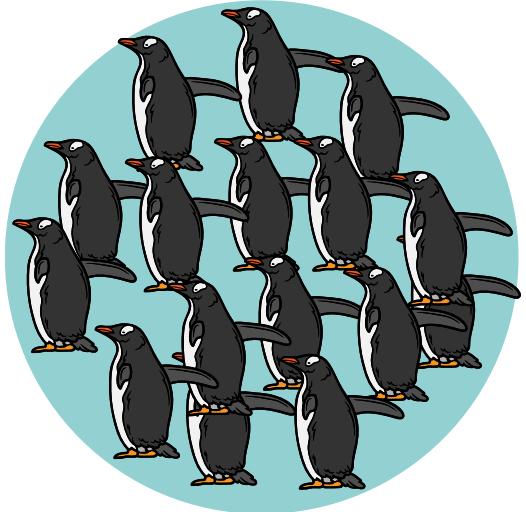
SKEWNESS

KURTOSIS

[4, 7, 8, ..., 178, 180, 189]

 $n = 23$ 

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

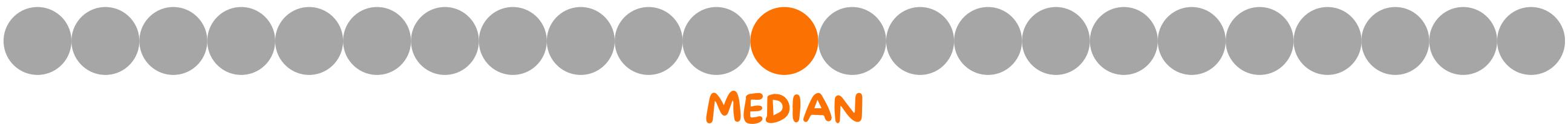
MODE

QUANTILES

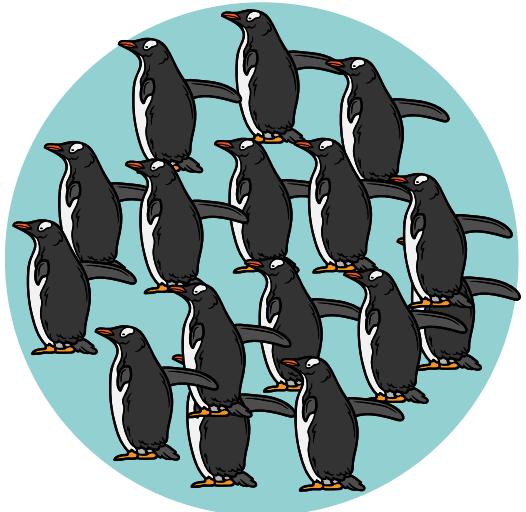
RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

 $[4, 7, 8, \dots, 178, 180, 189]$  $n = 23$ 

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



# SAMPLE

# MEAN

# MEDIAN

# MODE

# QUANTILES

# COEFFICIENT OF VARIATION

# RANGE & IQR

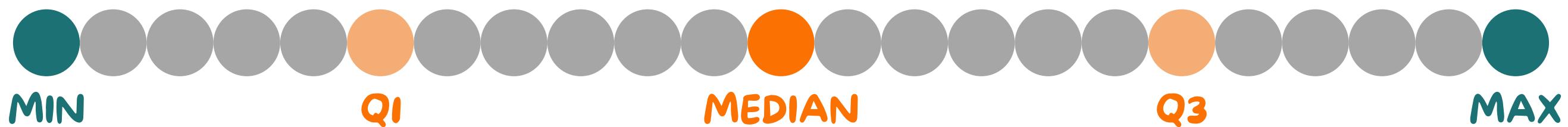
# SKEWNESS

# VARIANCE $\xi$ STD DEV

# KURTOSIS

[4, 7, 8, ..., 178, 180, 189]

$$n = 23$$

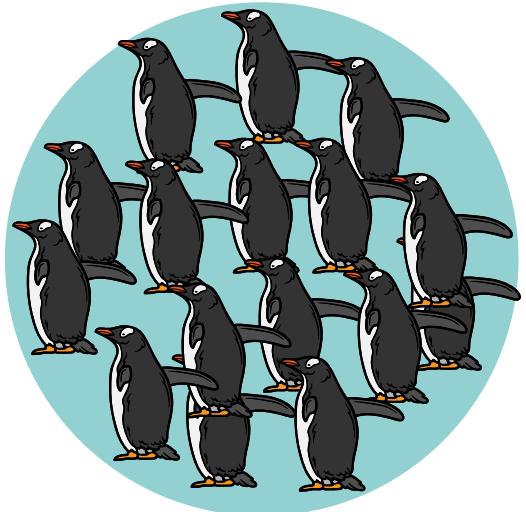


First quartile (Q1) : 25% (1/4)  
of observations below

Third quartile (Q3) : 75% (3/4) of observations below



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



# SAMPLE

# MEAN

# MEDIAN

# MODE

# QUANTILES

# COEFFICIENT OF VARIATION

# RANGE & IQR

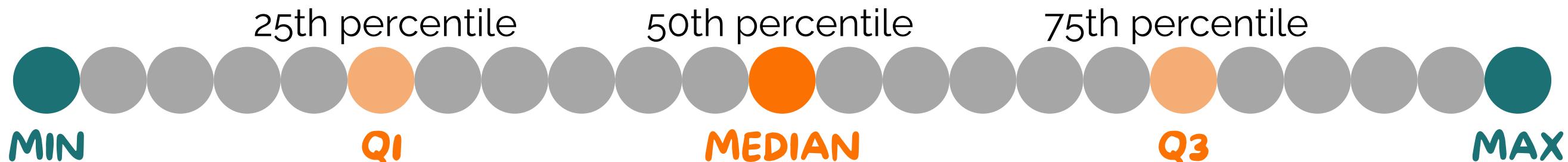
# VARIANCE $\&$ STD DEV

# SKEWNESS

# KURTOSIS

[4, 7, 8, ..., 178, 180, 189]

$$n = 23$$

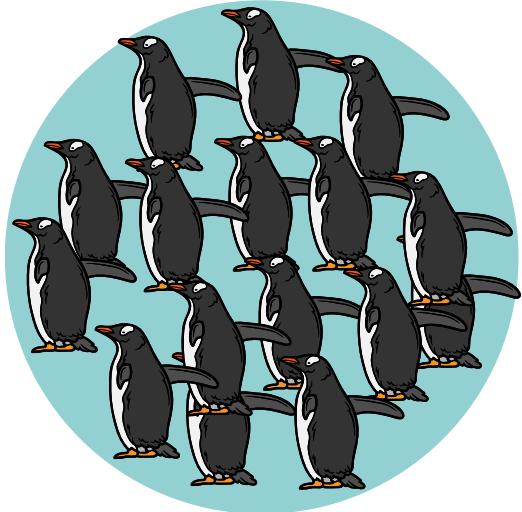


First quartile (Q1) : 25% (1/4)  
of observations below

Third quartile (Q3) : 75% (3/4) of observations below



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

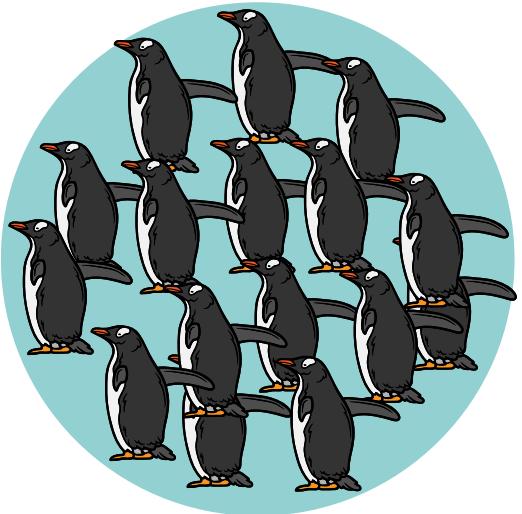
KURTOSIS

[2, 2, 5, 6, 9, 10, 13]

2 2 5 6 9 10 13



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

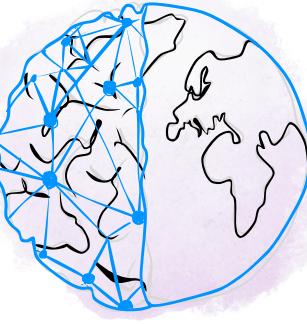
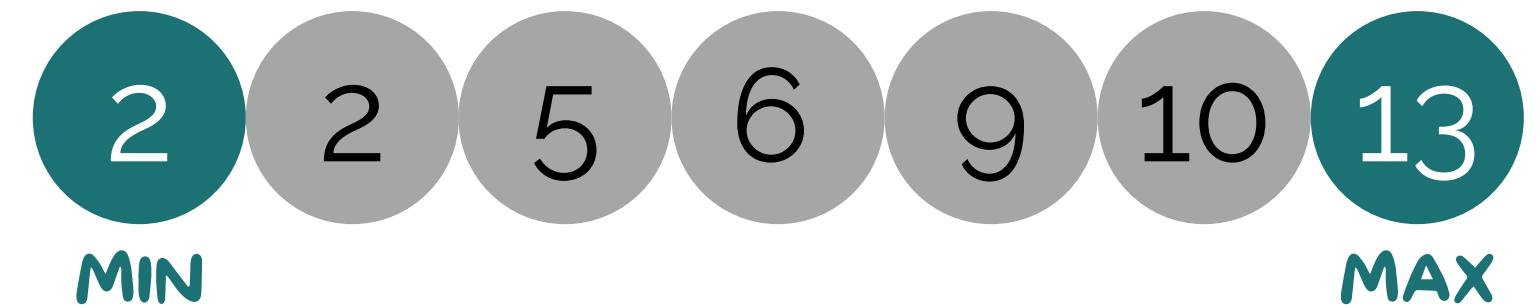
MODE

QUANTILES

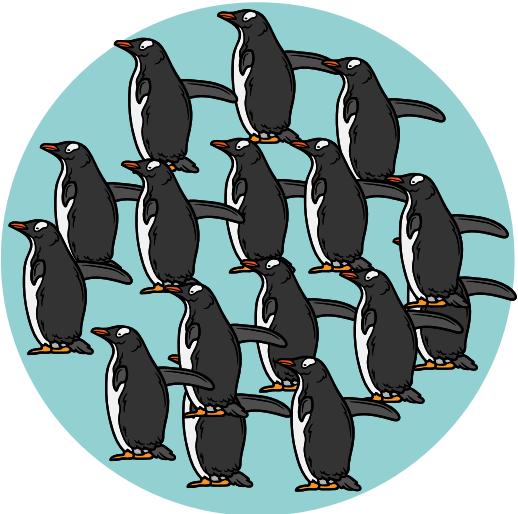
RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

 $[2, 2, 5, 6, 9, 10, 13]$ Range:  $\text{MAX} - \text{MIN} = 13 - 2 = 11$ 

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

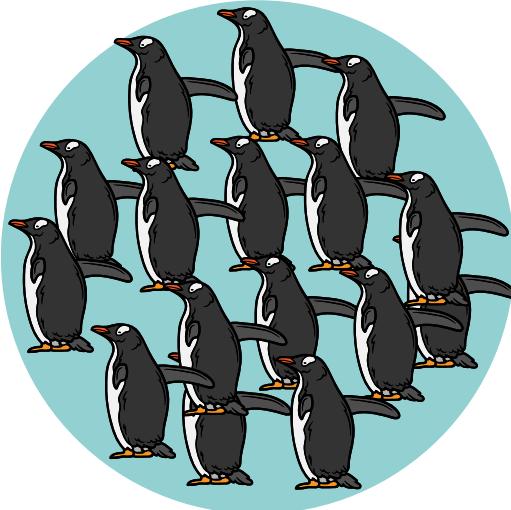
[2, 2, 5, 6, 9, 10, 13]

Range: MAX - MIN = 13 - 2 = 11

IQR: Q3 - Q1 = 10 - 2 = 8



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

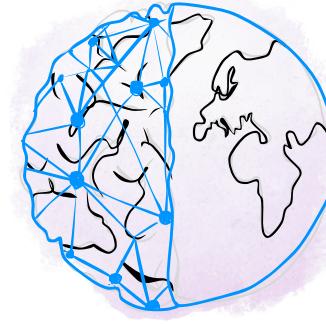
SKEWNESS

KURTOSIS

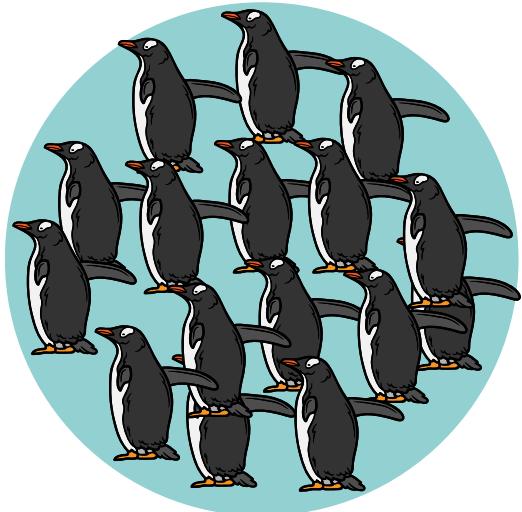
[2, 2, 5, 6, 9, 10, 58]

Range:  $\text{MAX} - \text{MIN} = 58 - 2 = 56$  Range is sensitive to outliers

IQR:  $\text{Q}_3 - \text{Q}_1 = 10 - 2 = 8$  The middle 50% of the ordered data  
Not sensitive to outliers



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

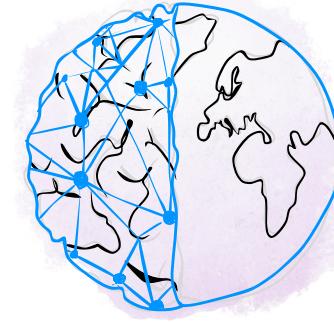
MODE

QUANTILES

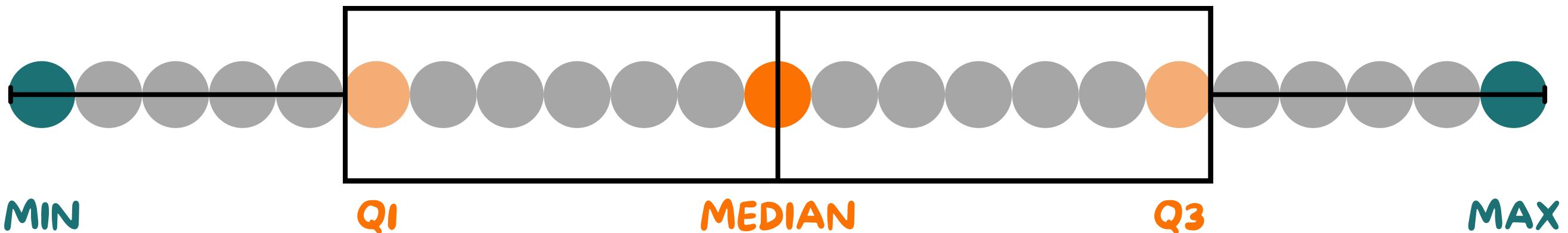
RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

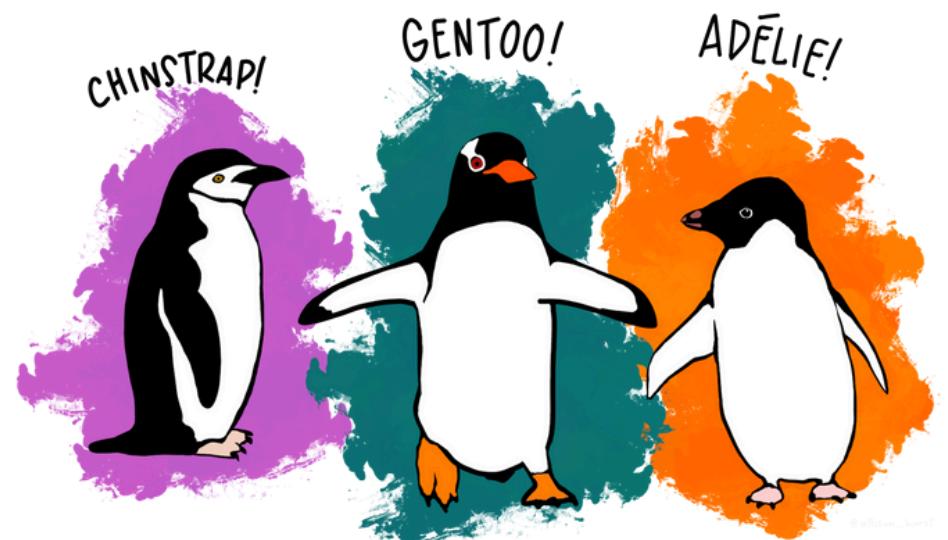
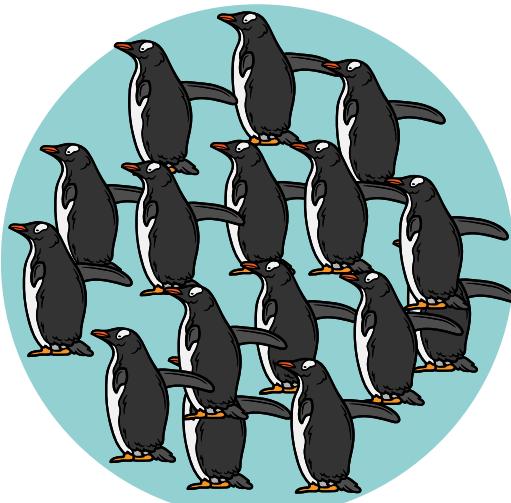
 $n = 23$ 

## BOXPLOT



The box-and-whisker plot was first introduced in 1970 by John Tukey

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

MODE

QUANTILES

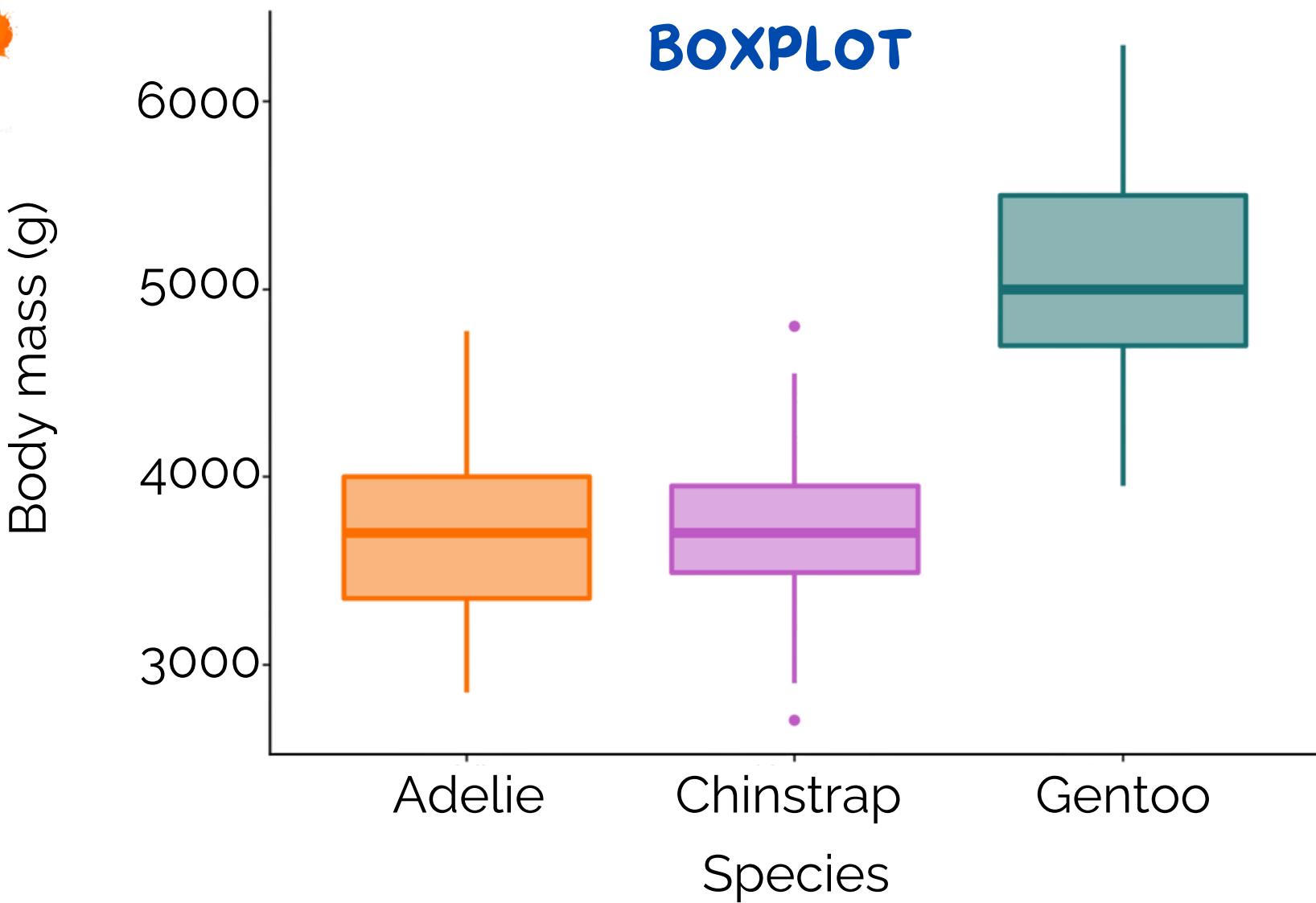
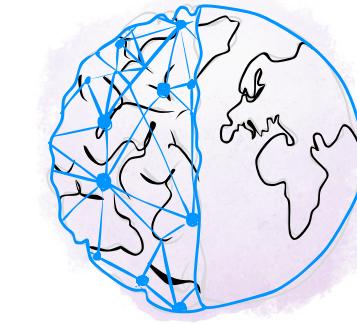
RANGE  
&  
IQR

VARIANCE  
&  
STD DEV

COEFFICIENT  
OF  
VARIATION

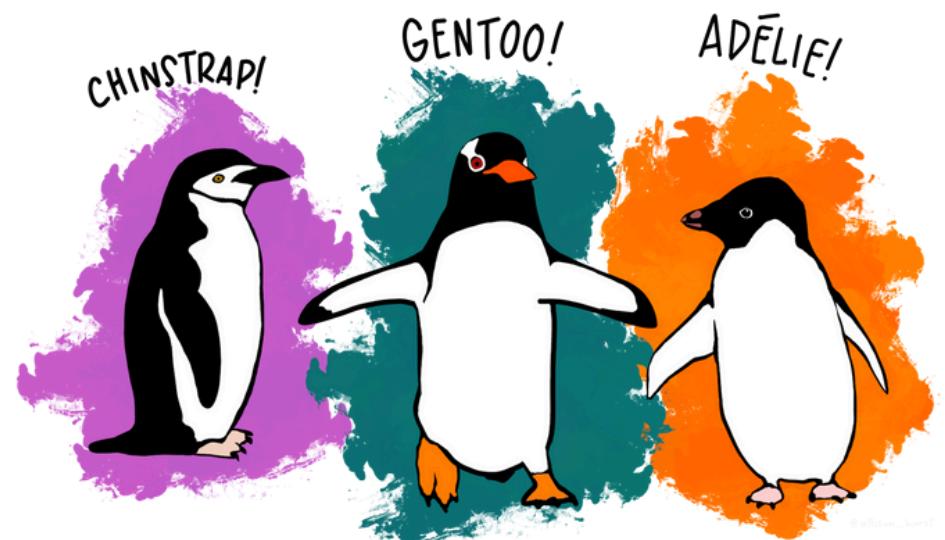
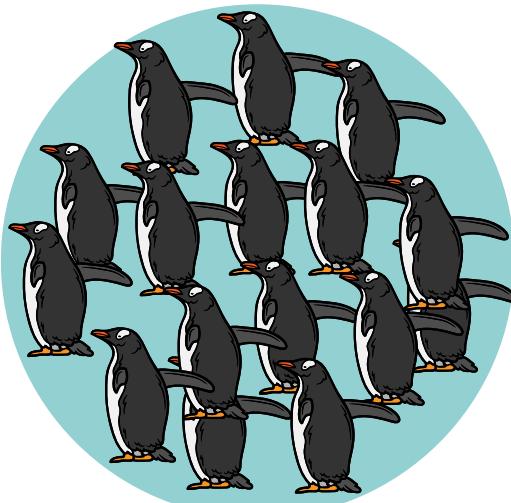
SKEWNESS

KURTOSIS



These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

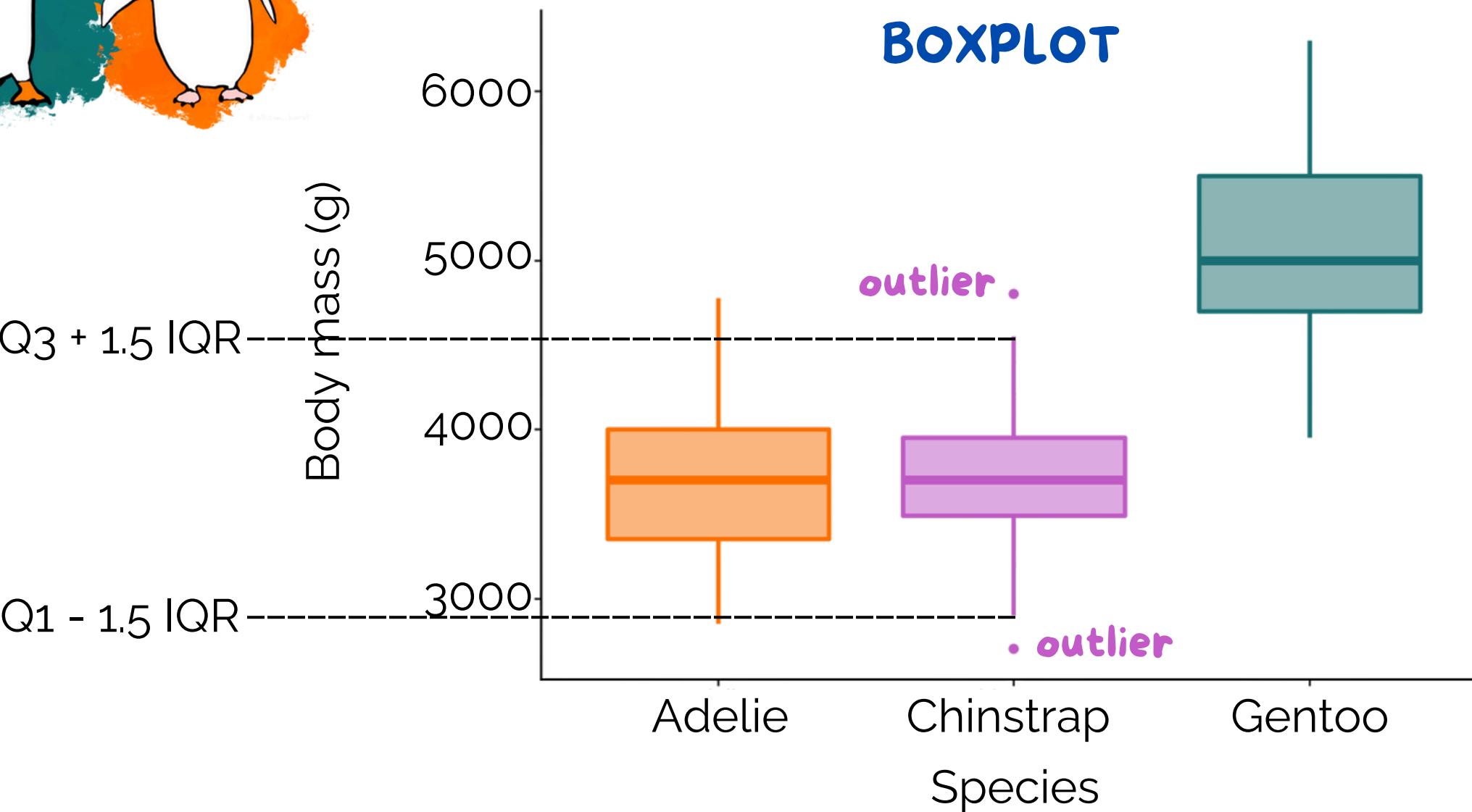
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

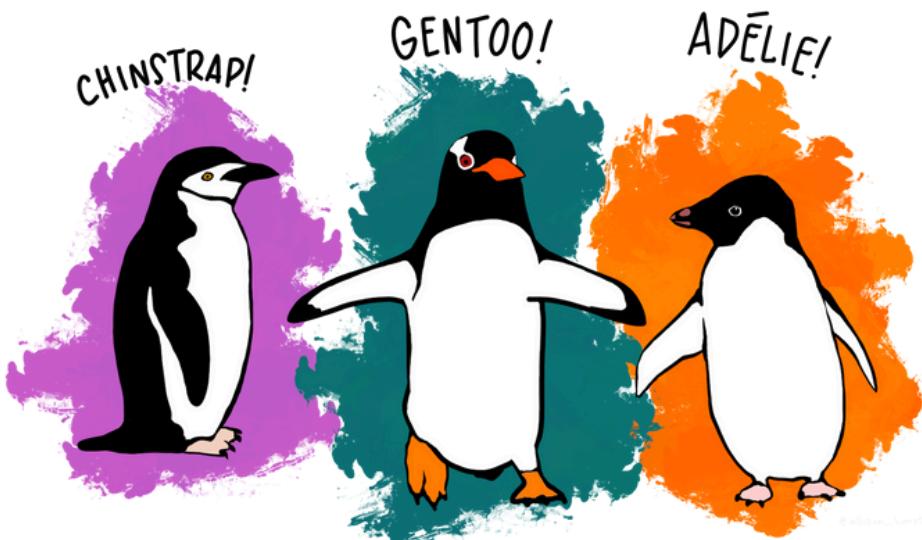
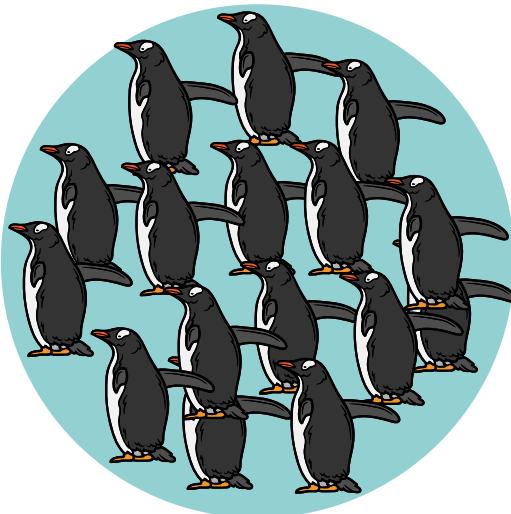


These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.



## PART I

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



MEAN

MEDIAN

MODE

QUANTILES

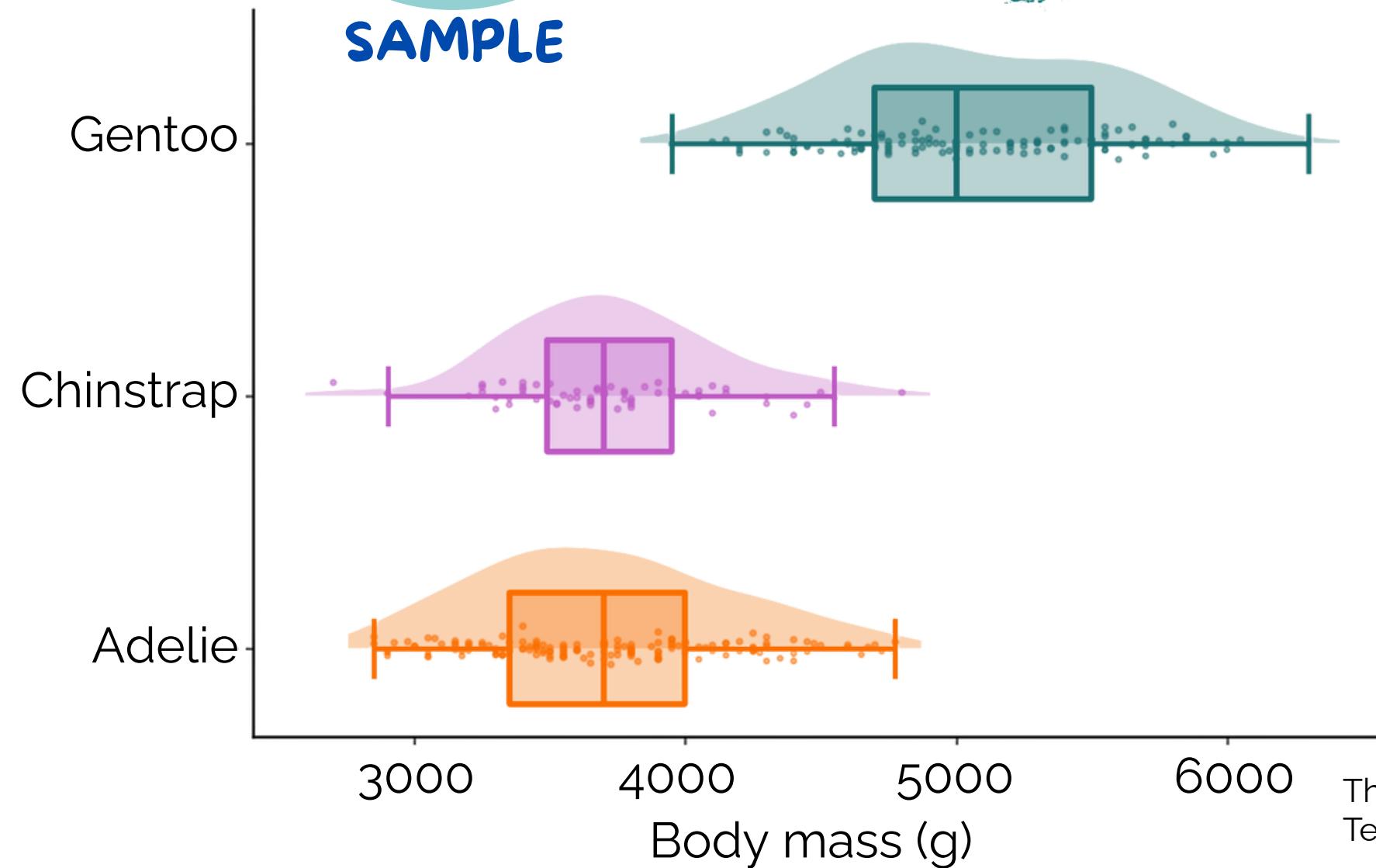
RANGE  
&  
IQR

VARIANCE  
&  
STD DEV

COEFFICIENT  
OF  
VARIATION

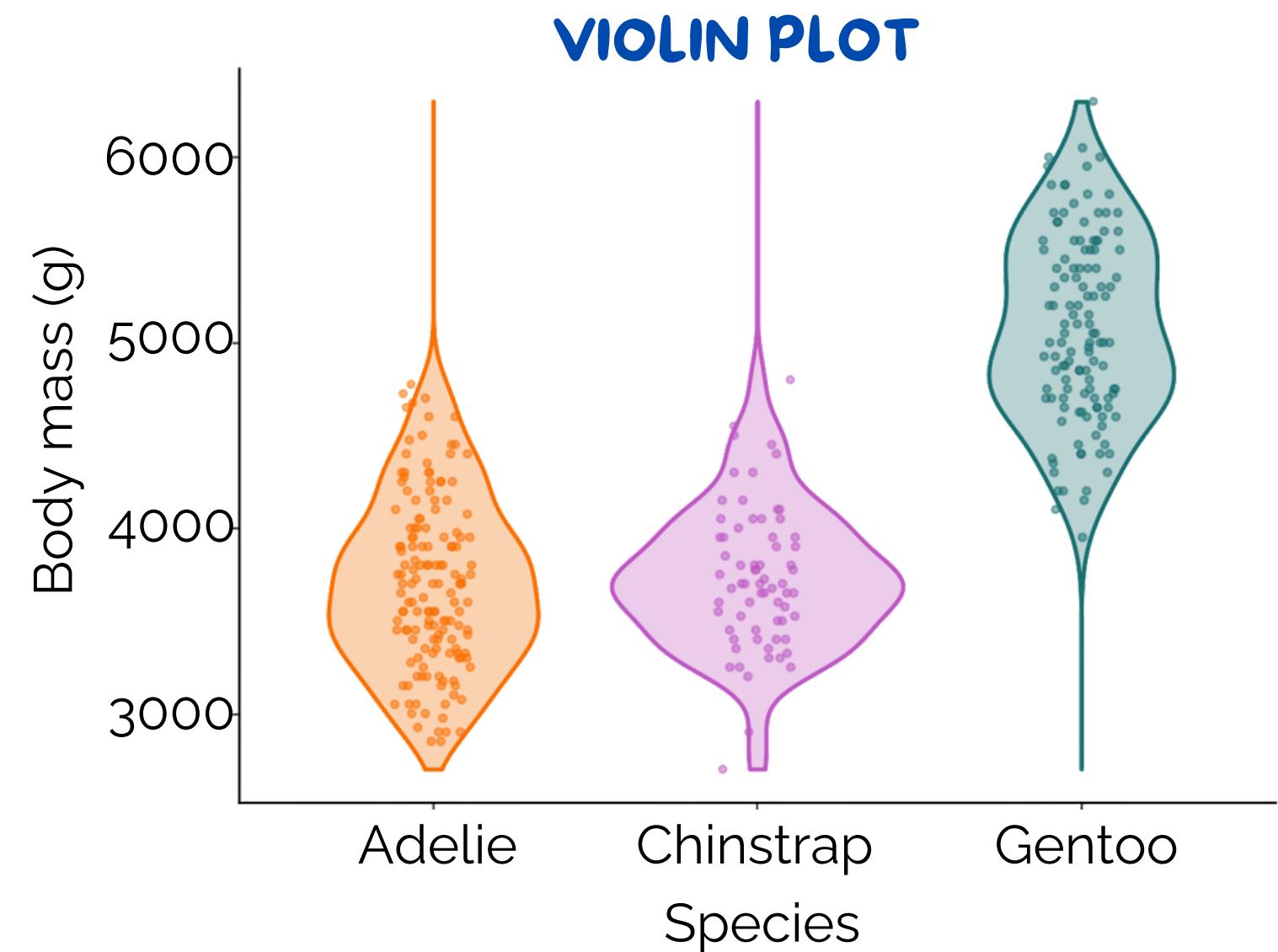
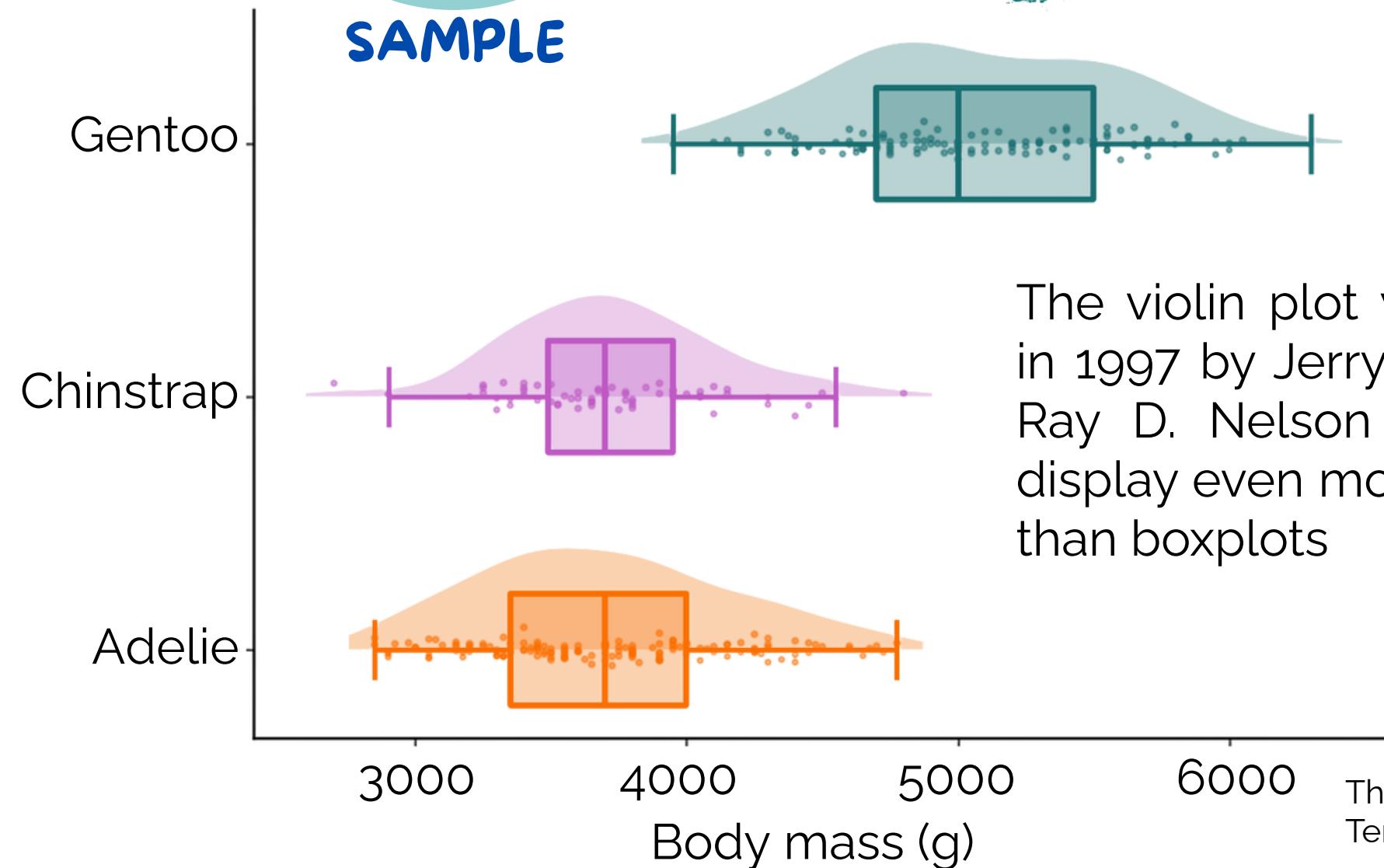
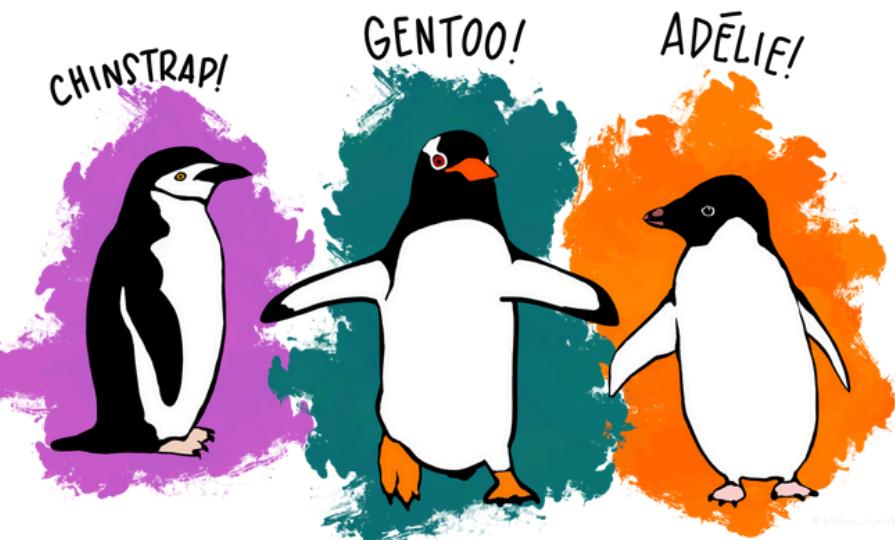
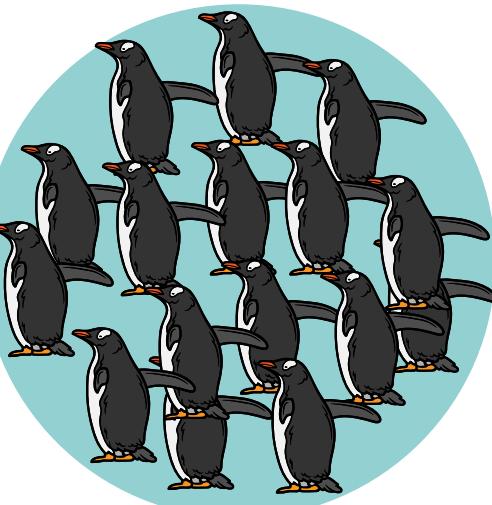
SKEWNESS

KURTOSIS



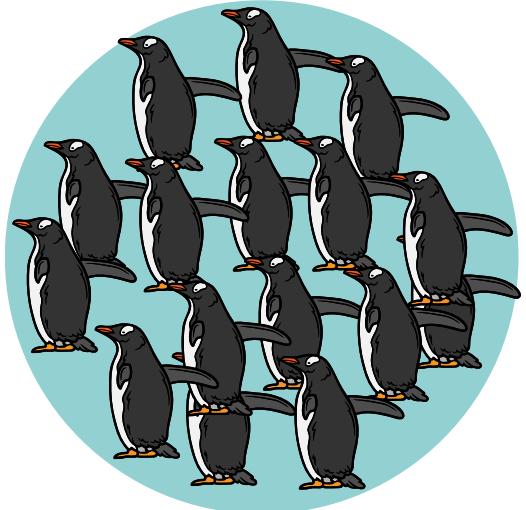
These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

MEDIAN

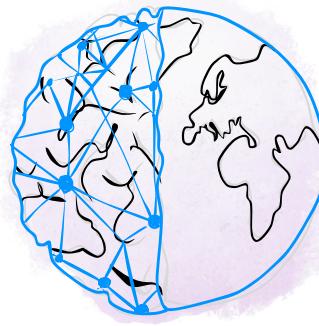
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



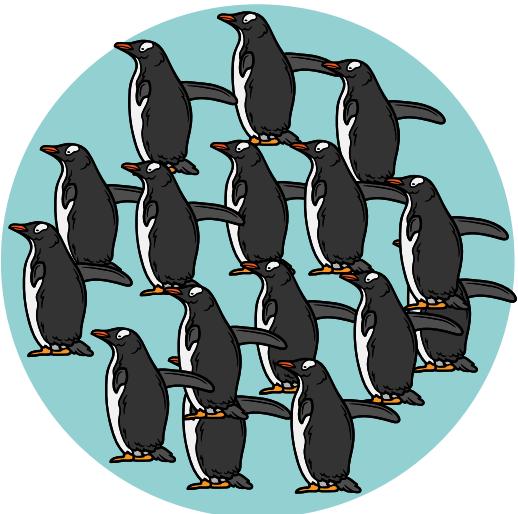
Deviation measures the difference between an observed value and the mean value of a data set.

[10, 28, 28, 33, 54]

$$\bar{x} = \frac{10 + 28 + 28 + 33 + 54}{5} = 30.6$$

mean absolute deviation =  $\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{|10-30.6| + \dots + |54 - 30.6|}{5} = 10.32$

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

MEDIAN

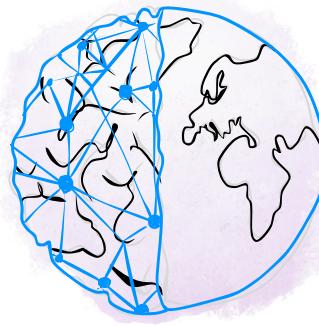
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



mean absolute deviation =  $\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$

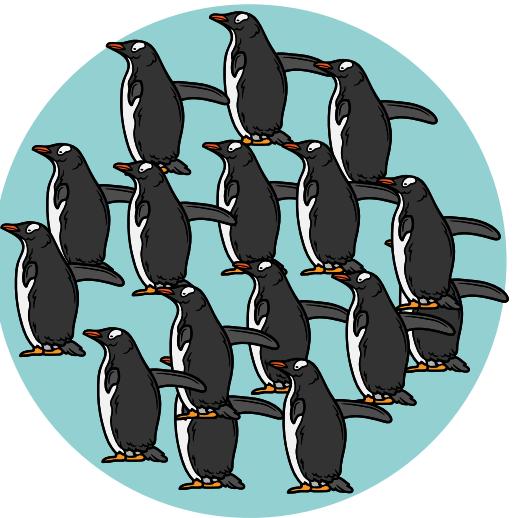
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

STANDARD DEVIATION

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



POPULATION

$$\text{mean absolute deviation} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE

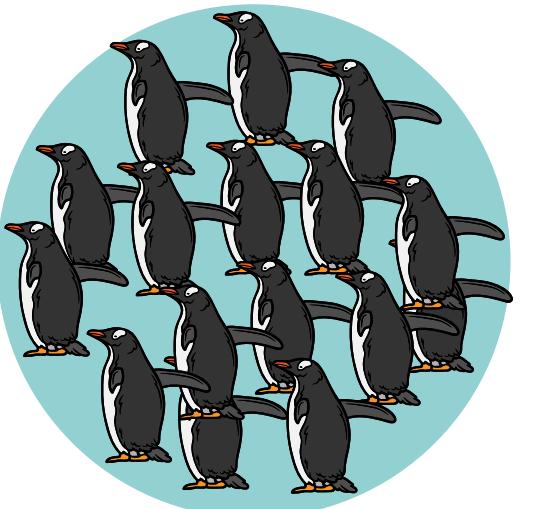
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

$$\sigma = \sqrt{\sigma^2}$$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

STANDARD DEVIATION

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

**BESSEL'S CORRECTION**

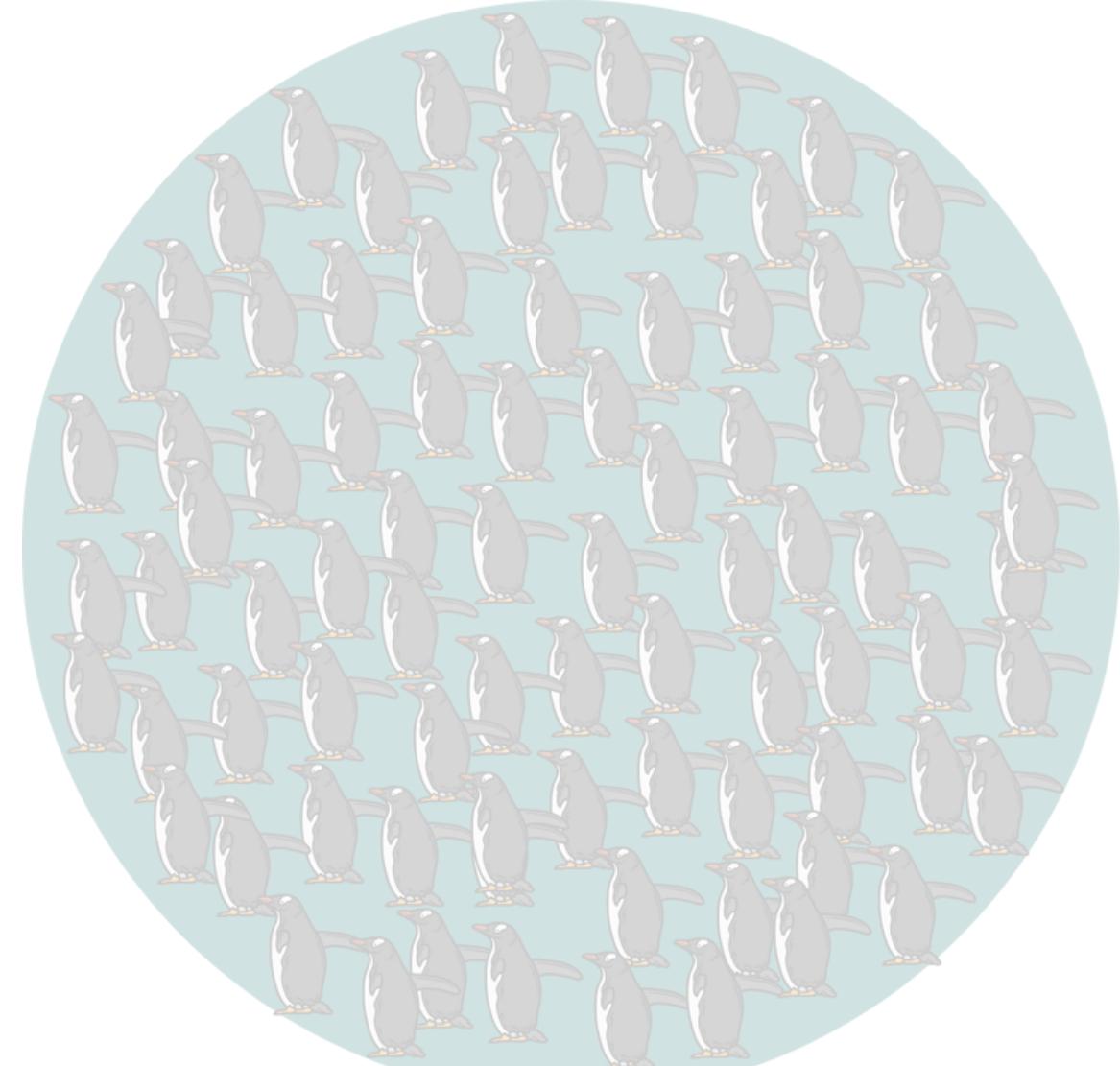
mean absolute deviation =  $\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

$$\sigma = \sqrt{\sigma^2}$$

The distance  $(x_i - \bar{x})$  is, on average, a little smaller than  $(x_i - \mu)$

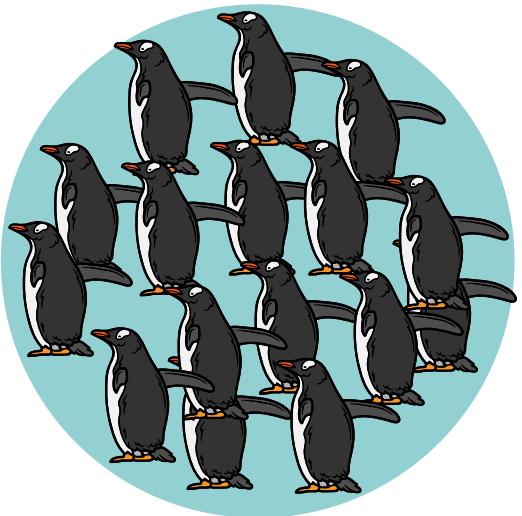
There are  $n - 1$  degrees of freedom since there is one constraint:  
the standard deviation depends on calculating the sample mean.



POPULATION



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS

mean absolute deviation =  $\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$

$$\bar{x} = \frac{10 + 28 + 28 + 33 + 54}{5} = 30.6$$

[10, 28, 28, 33, 54]

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{(10 - 30.6)^2 + (28 - 30.6)^2 + (28 - 30.6)^2 + (33 - 30.6)^2 + (54 - 30.6)^2}{5-1} = 247.8$$

VARIANCE

$$s = \sqrt{s^2} \approx 15.74$$

STANDARD DEVIATION

WHAT WOULD THE SAMPLE  
STANDARD DEVIATION BE?

WHAT HAPPENS WHEN THE  
SAMPLE SIZE IS LARGE?



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

MEDIAN

MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



$$A = [1, 2, 3]$$

$$\bar{A} = 2 \quad s_A = 1$$

$$B = [101, 102, 103]$$

$$\bar{B} = 102 \quad s_B = 1$$

# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

MEAN

QUANTILES

**COEFFICIENT  
OF  
VARIATION**

MEDIAN

RANGE  
&  
IQR

SKEWNESS

MODE

VARIANCE  
&  
STD DEV

KURTOSIS

$$A = [1, 2, 3]$$

$$\bar{A} = 2 \quad s_A = 1$$

$$CV(A) = \frac{1}{2} = 0.5$$

Proportional to  
the sample mean

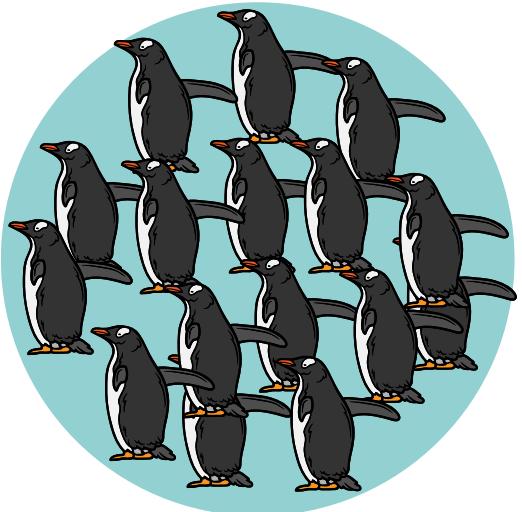
$$B = [101, 102, 103]$$

$$\bar{B} = 102 \quad s_B = 1$$

$$CV(B) = \frac{1}{102} \approx 0.0098$$



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

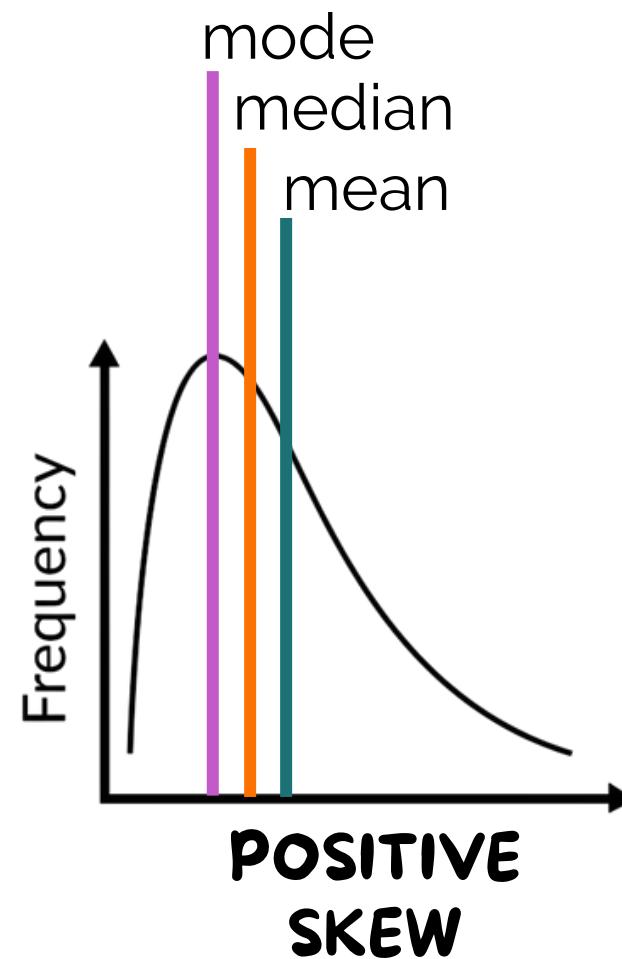
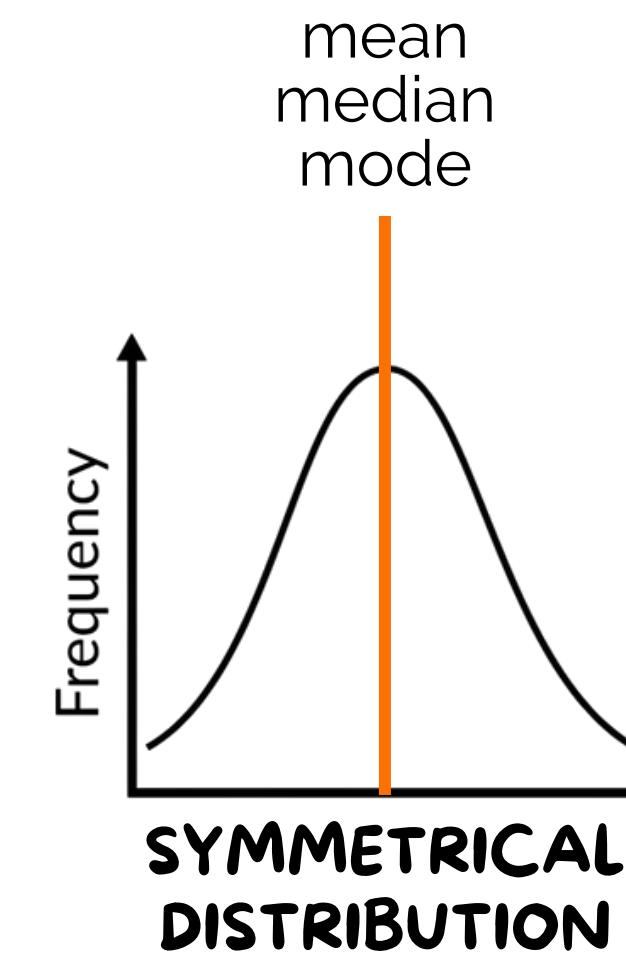
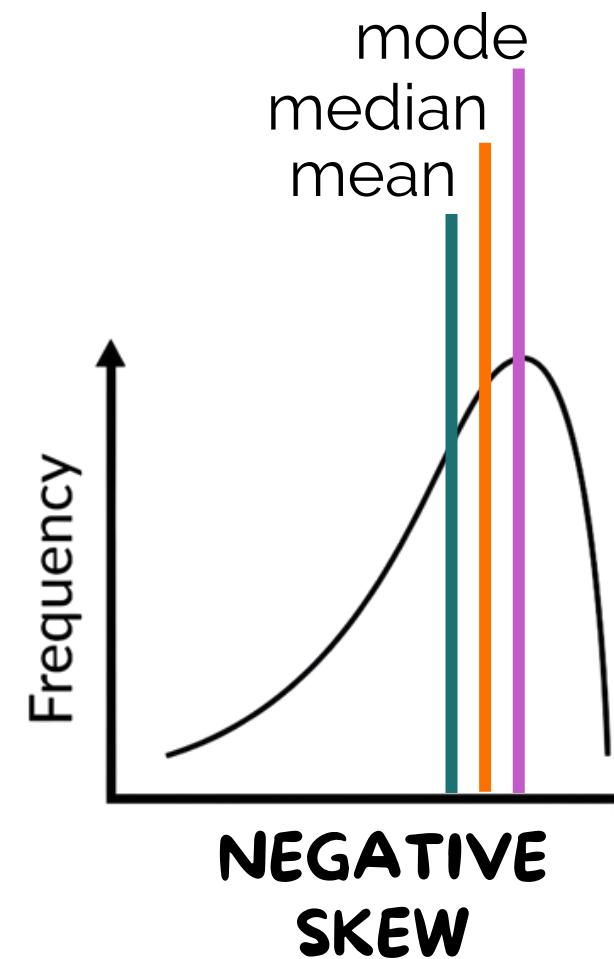
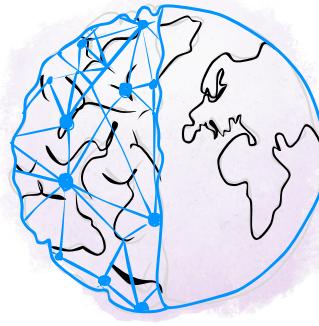
MODE

QUANTILES

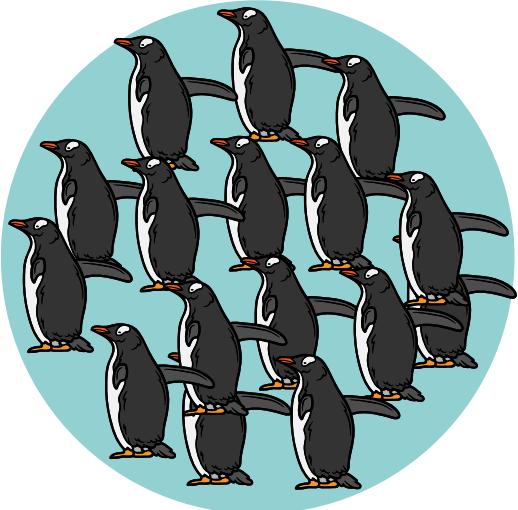
RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

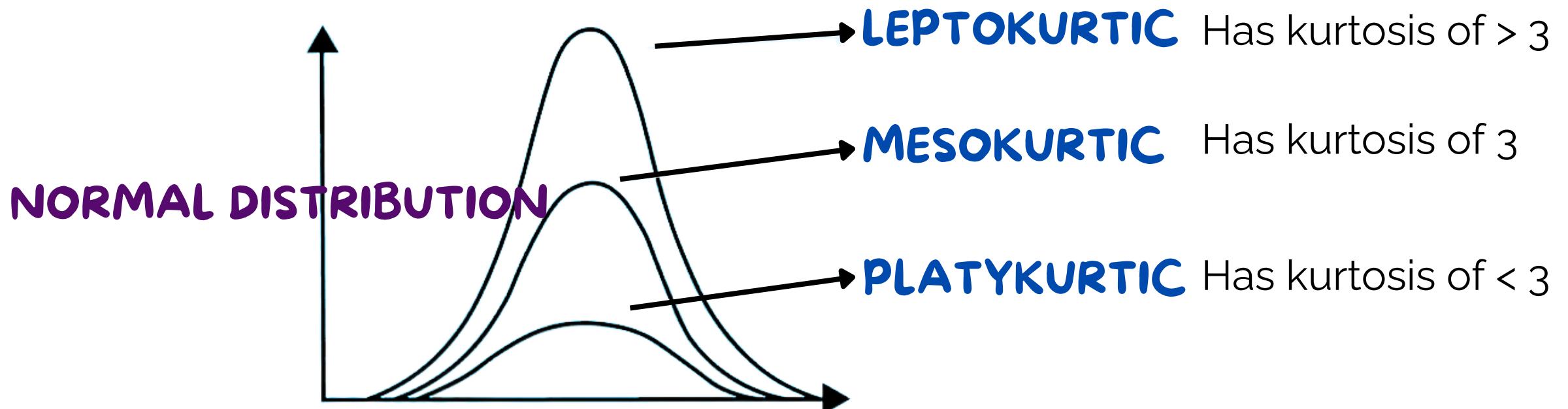
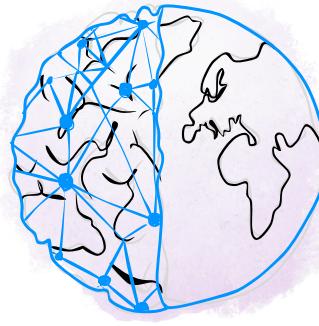
MODE

QUANTILES

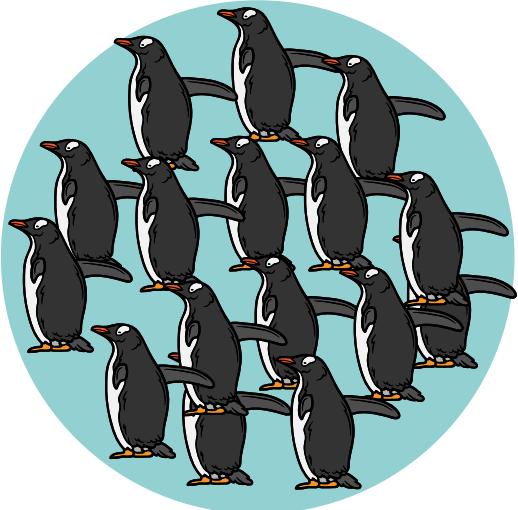
RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



# ESTIMATES OF LOCATION DESCRIBING THE DISTRIBUTION



SAMPLE

MEAN

MEDIAN

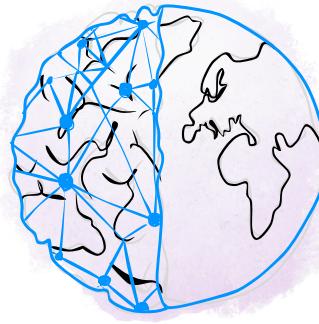
MODE

QUANTILES

RANGE  
&  
IQRVARIANCE  
&  
STD DEVCOEFFICIENT  
OF  
VARIATION

SKEWNESS

KURTOSIS



Excess kurtosis = kurtosis - 3

