

Lecture 4.10.18 : AMS 7

This time: Variable types ; Histograms

next time: Center & Spread ; normal curve (bell curve)

\* HW #1 due date extended to Friday 4.20 @ 11:59 pm

1.2 • Variable types (Data types cont.)

	Variables	Possible values	Ratio (unique true zero)
Plant size	seedling height (cm)	7.42 cm, ... (quant.) *continuous numbers	quantitative numerical (have unique spot on the number line) (quant.) *can be continuous numbers (w/right sig. figures) or whole numbers (discrete)
	# of leaves (adult)	46, 111, ... (quant.) *pure # counts & discrete #s	
	Hair color in hamsters	brown, black (qual.) (Nominal)	
	Success at maze running	Very slow, slow, moderate & fast (ordered categorical) (qual.) aka Ordinal	

\* qualitative  
 ↳ categorical  
 • can be ordinal  
 → don't have a unique spot on the number line  
 ↳ nominal: have no order to them

ex: hair color

Can take the mean of data sets of quant. variables

Quant. → ratios & ratio scales

a unique true zero on the measurement scale  
 with a direct physical meaning (ex: seedling height)

Interval scale

has a constant size interval but lacks a true zero

(ex: temperature)  $^{\circ}F = \frac{9}{5}^{\circ}C + 32^{\circ}$

\* age - qualitative & discrete

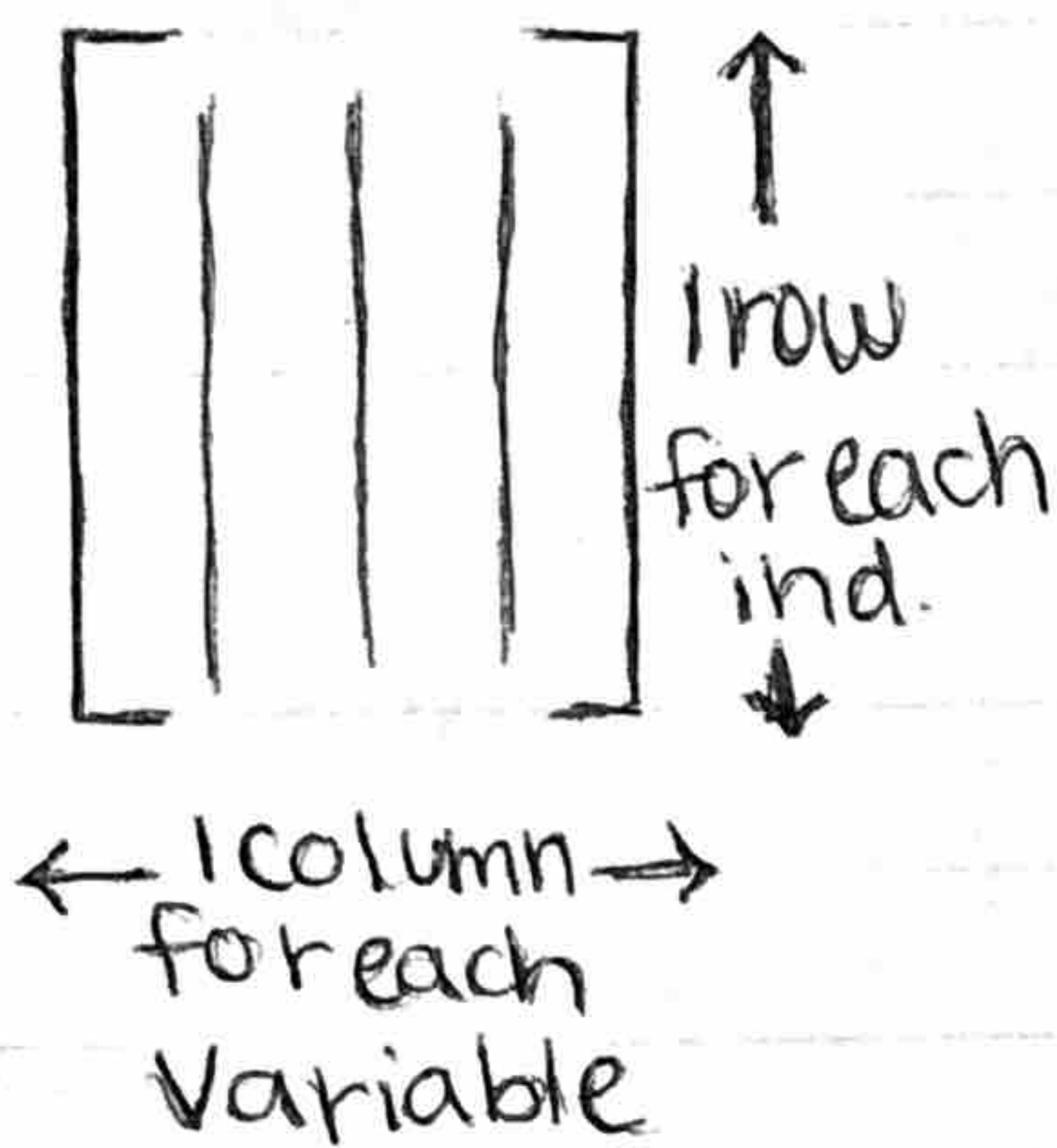
↳ can take average

\* animal ID = even though variables are numerical, this is QUALITATIVE (no meaning to mean)



2

\* all of data sets have property of one row for each individual & one column for each of the individuals



→ this info should help with HW#1, problem #1

L-15

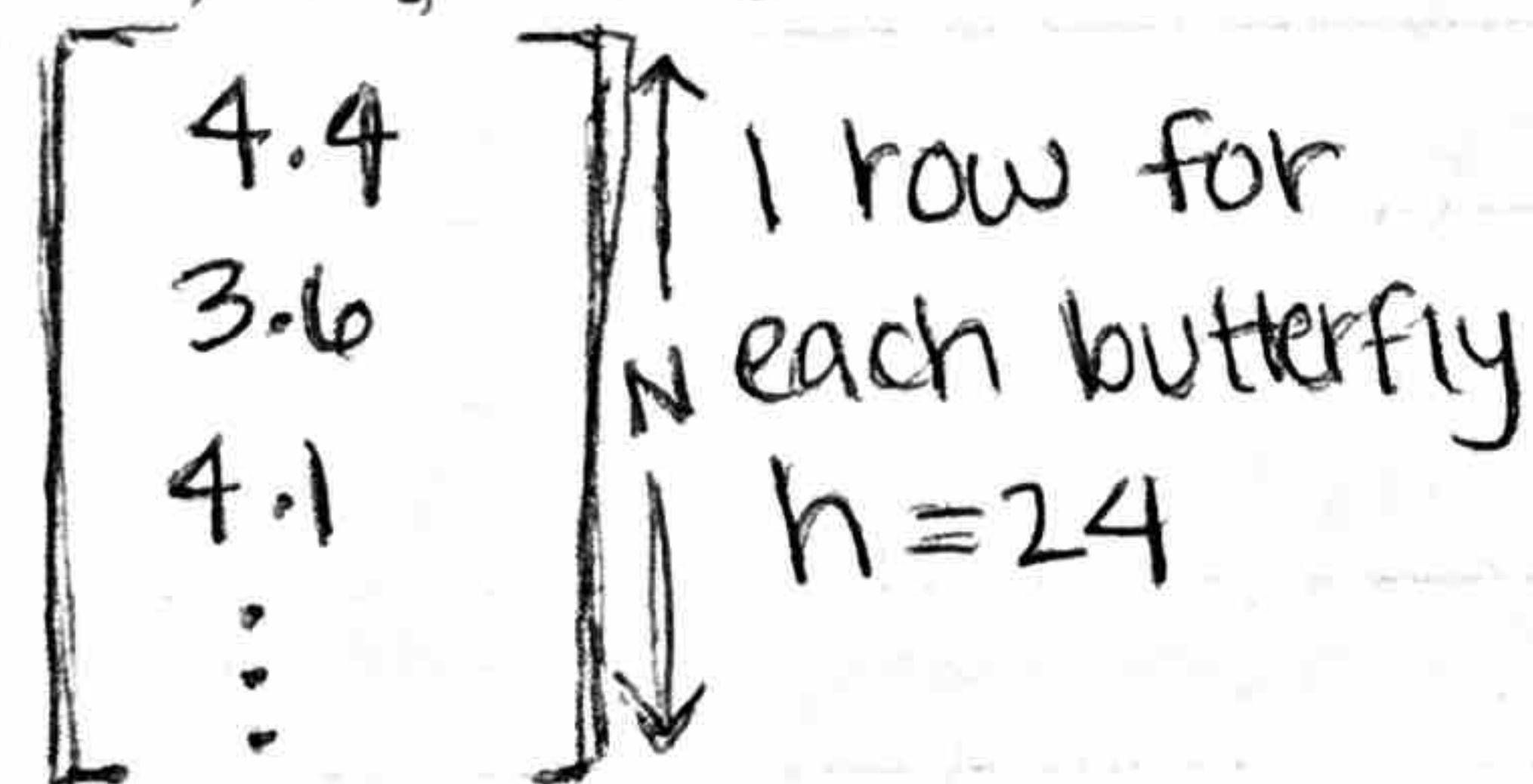
1.3.1 Graphical descriptive methods

Example: butterfly wing lengths

Sample  $n=24$

Variable of interest = Wing length (cm)

Wing length (cm)



\* Step 1

What was the wing span of big vs small butterfly?

→ Sort data from smallest to largest

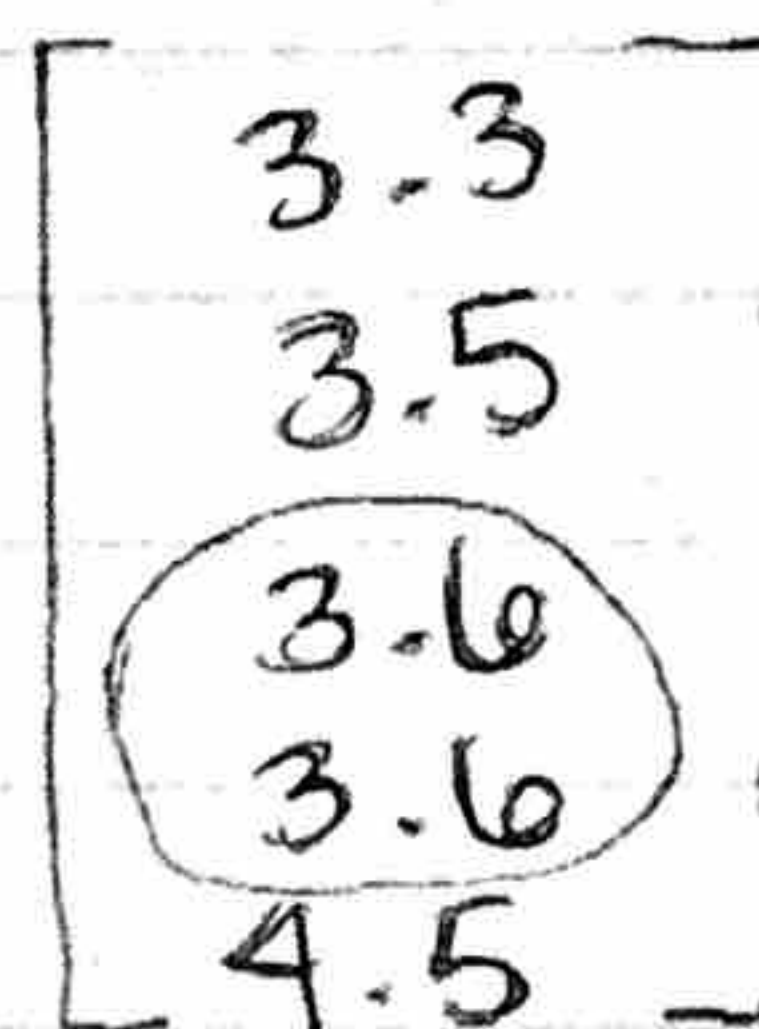
\* Step 2

Raw Freq. Distribution

count

wing length value (cm)	raw frequency
3.3	1
3.4	0
3.5	1
3.6	2
⋮	
4.5	
<b>Total :</b>	<b>24 = n</b>

0 b/c it does not exist in data set



Sorting will lose the order → not relevant in this ex.

Wing length (cm)  
 \* quant.  
 \* continuous  
 \* ratio

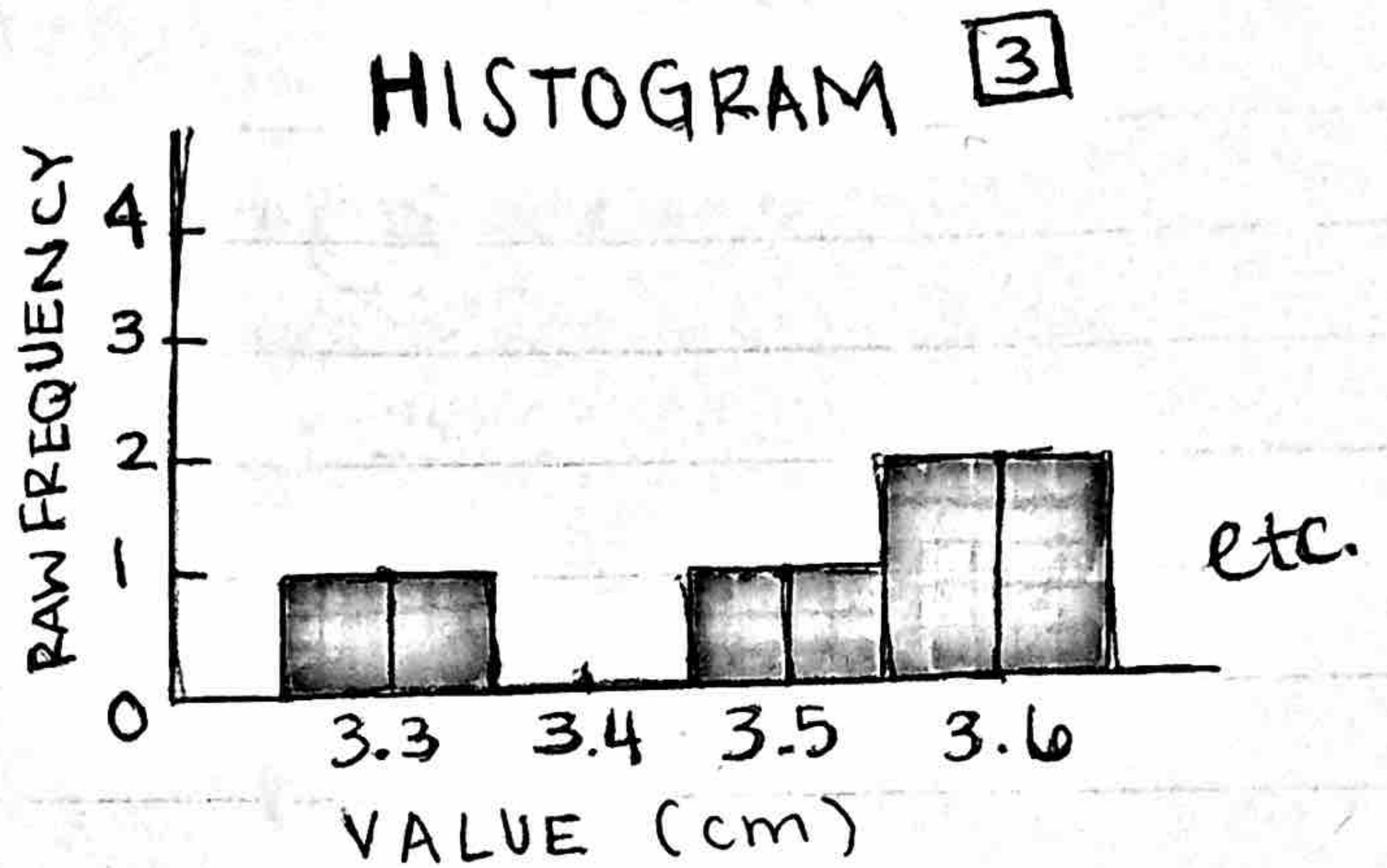


Graph  $\left\{ \begin{array}{l} \text{Value} \rightarrow \text{Horizontal Scale} \\ \text{Freq.} \rightarrow \text{Vertical Scale} \end{array} \right.$   
 \* can be switched b/c it is arbitrary

Histogram

a special case of a bar graph; a plot with locations identified along the horizontal axis corresponding to values a variable takes on and bars over those locations with heights given by the (raw) frequencies of those values

- \* Can be drawn as a summary of any quantitative (nominal or ordinal) variable
- \* qualitative data set can't have histogram but can have bar graph (arbitrary)



eye color

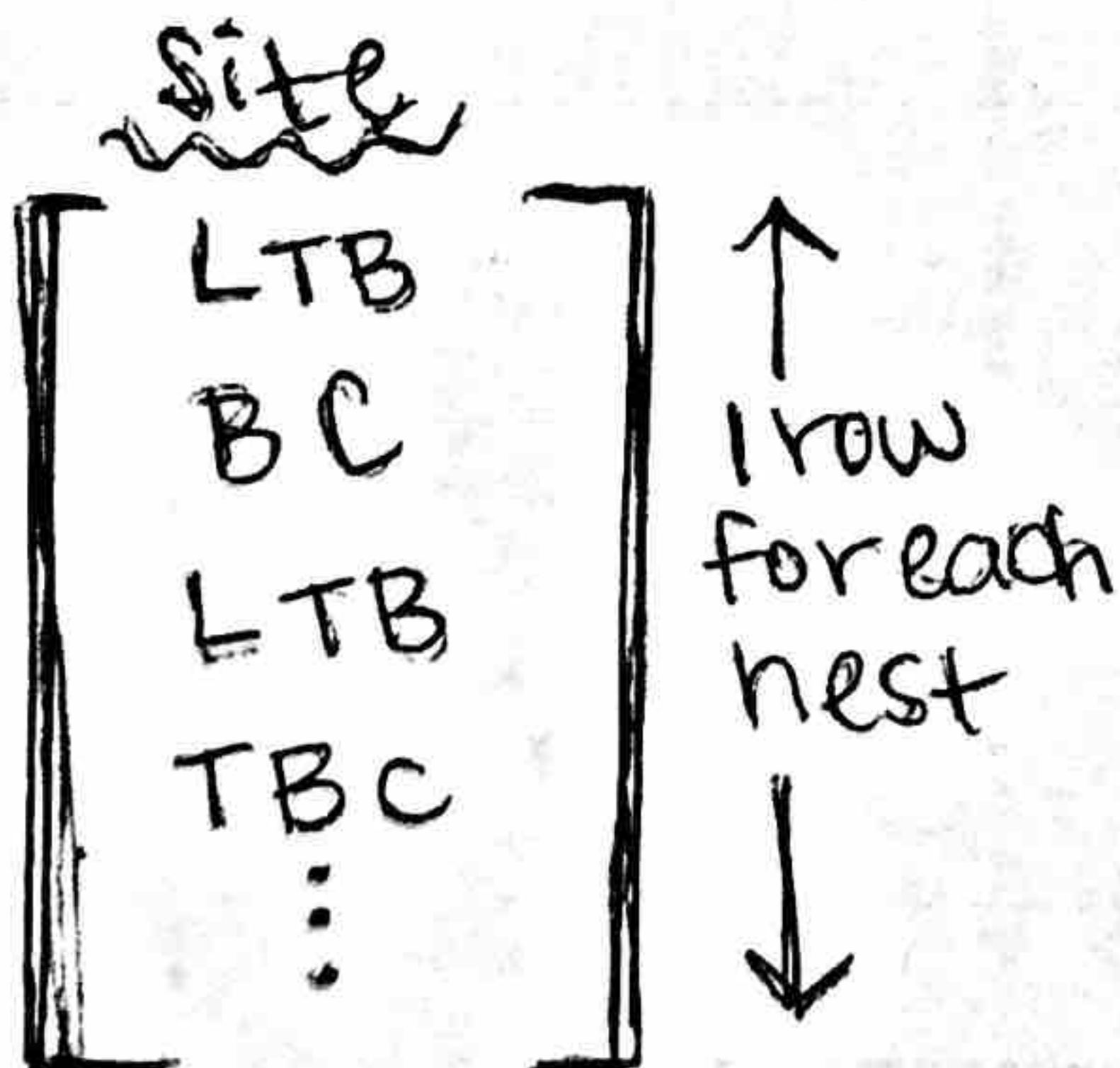


↳ has no unique spot on the number line

• more graphical examples

\* the location of sparrow nests - Raw Freq. Distribution

<u>Nest site</u>	<u>(Number of Nests observed)</u>	
Vines	56	- qual.
BC - Building caves	60	- nominal
LTB - low tree branches	46	- not
TBC - Tree & building cavities	49	dichotomous

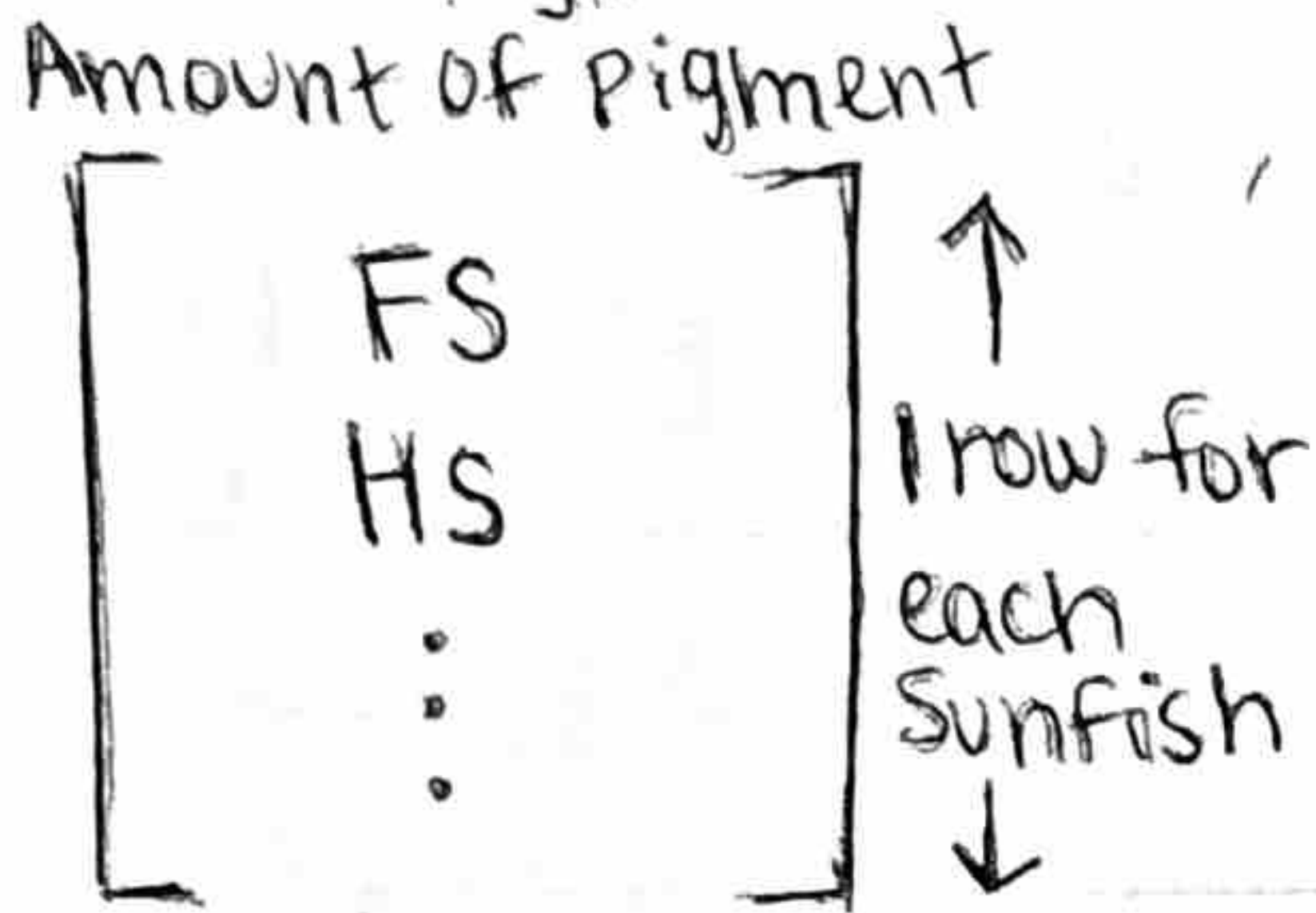


- no histogram  
 - can have bar graph



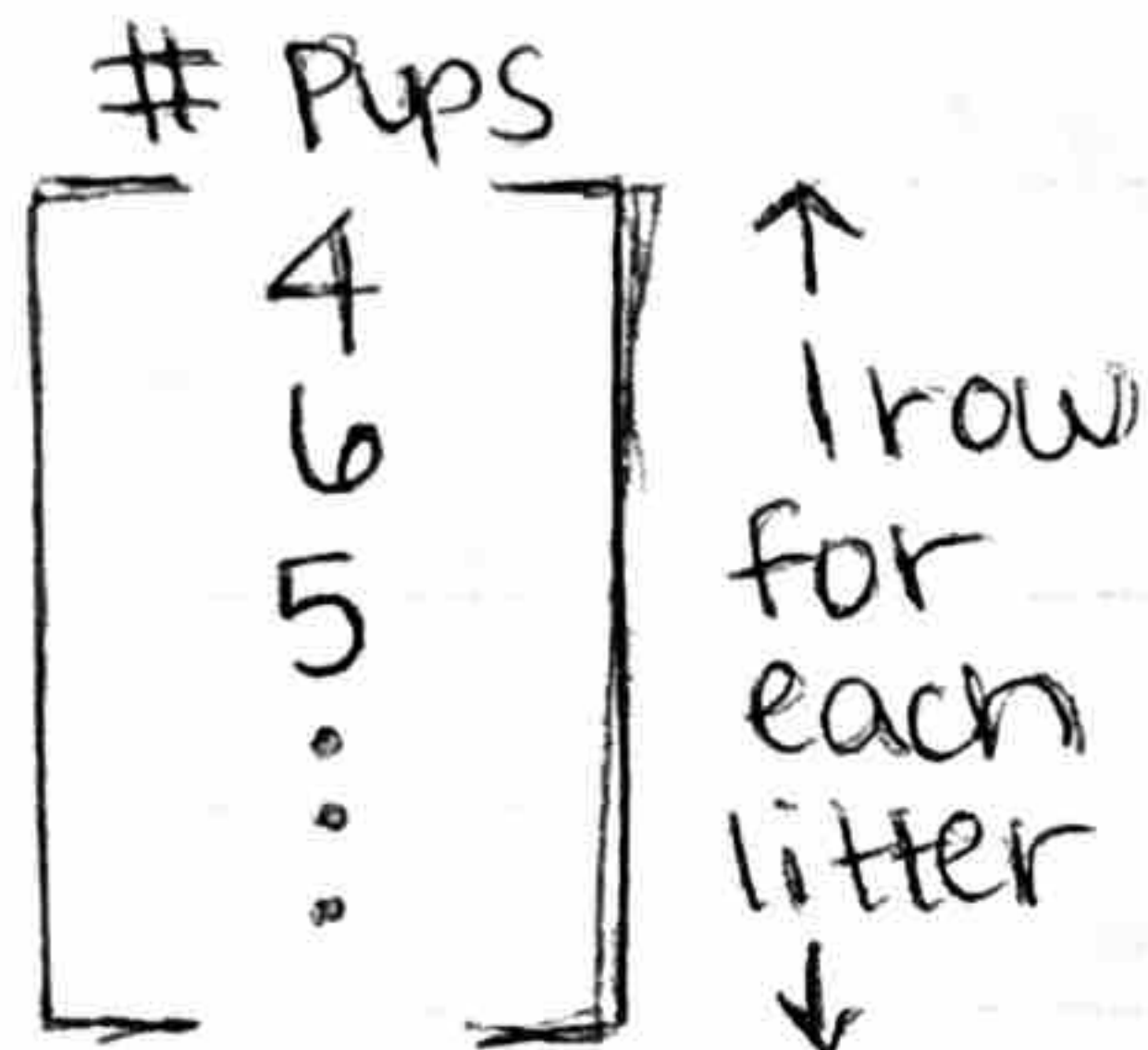
4

\* numbers of sunfish, tabulated according to amount of pigmentation - a freq. table of ordinal data



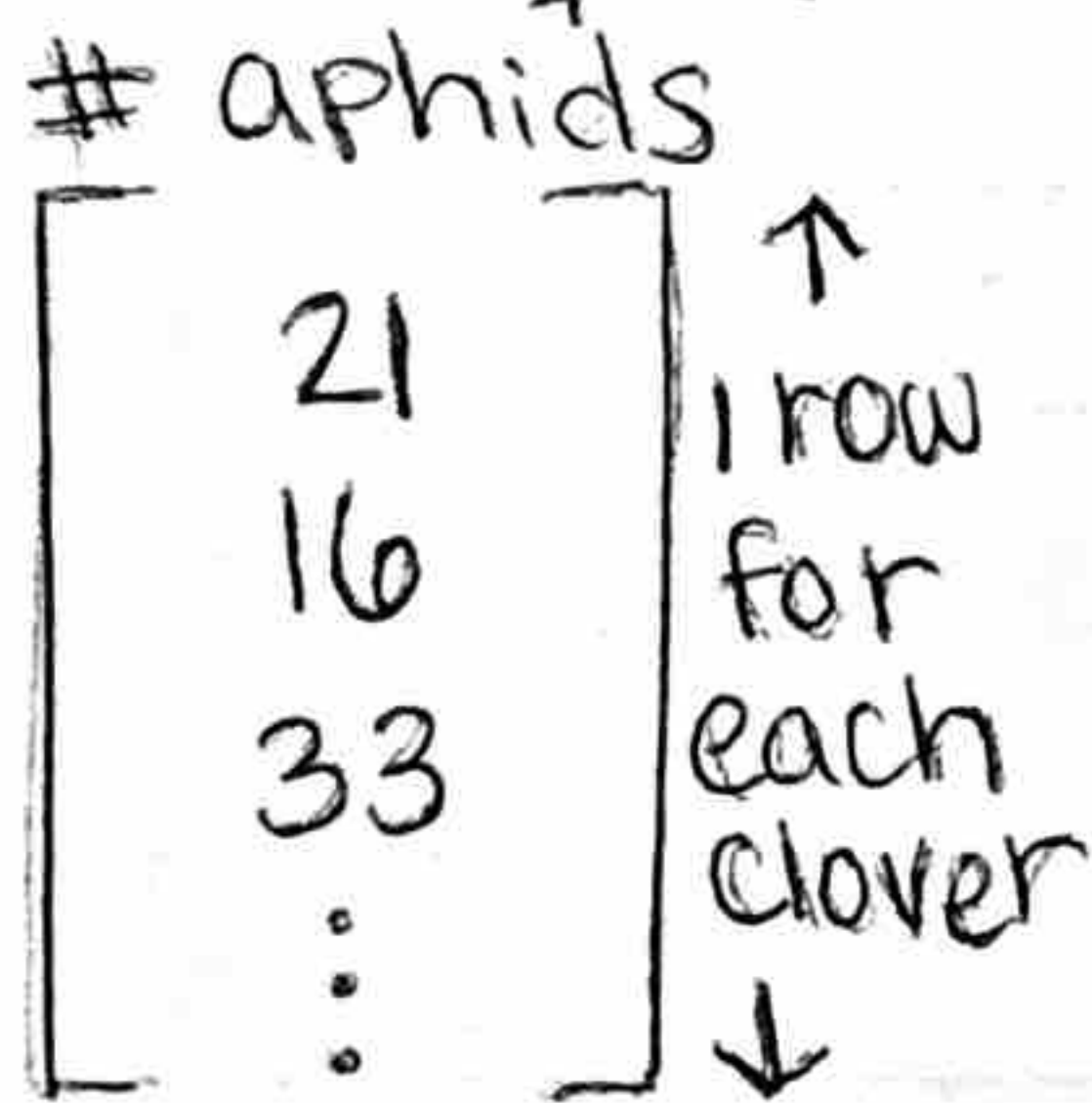
- qual.
- ordinal
- not dichot.
- no histogram
- has bar graph

\* Freq. of occurrence of various litter sizes in foxes  
a table of discrete ratio-scale data



- quant.
- discrete
- ratio
- Has histogram

\* number of aphids observed per clover plant  
A freq. table of discrete, ratio-scale data

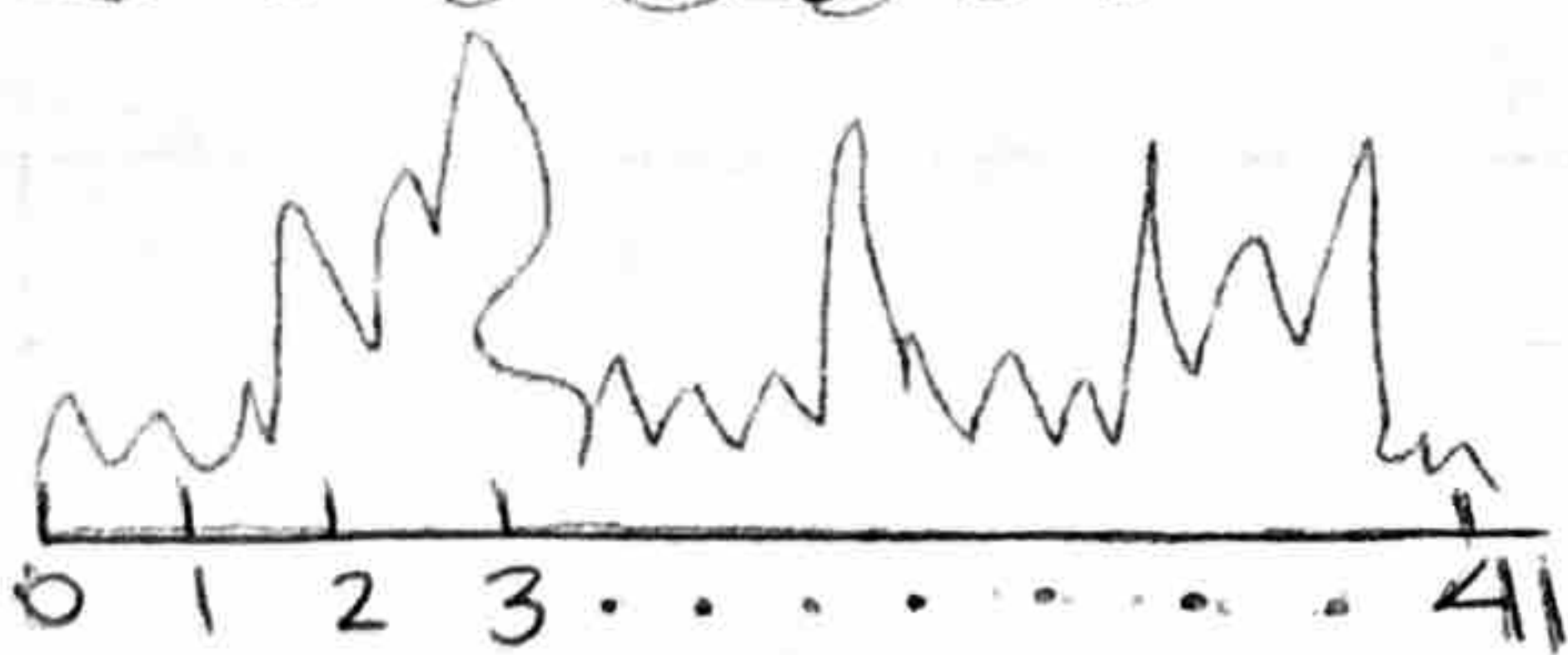


- quant.
- discrete
- ratio
- nominal/ordinal? NA
- has histogram

N = 41

↳ big number, can make diff types of histograms

Bad Histograms



too many bars  
too jumpy  
42 bars



too few bars  
all shape info is lost  
only 1 bar.

Goldilocks moment



- How to draw a good histogram of this example

# of aphids on plants	# of plants observed	# of aphids on plants	# of plants observed
0	3	0-3	6
1	1	4-7	17
2	1	8-11	40
3	1	12-15	54
4	2	⋮	⋮
⋮	⋮	40-43	1
41	1		



HW#1  
Problem#2

↳ this data will create a better histogram  
will look like the bell curve

\* Determination of the amount of phosphorus in leaves  
A freq. data table of continuous data

Phosph. conc.

8.63	↑ 1 row for each leaf ↓
9.14	
⋮	

- quant.
- cont.
- ratio
- has histogram  
↳ Bell Curve

Total freq. = 130 → the sample size

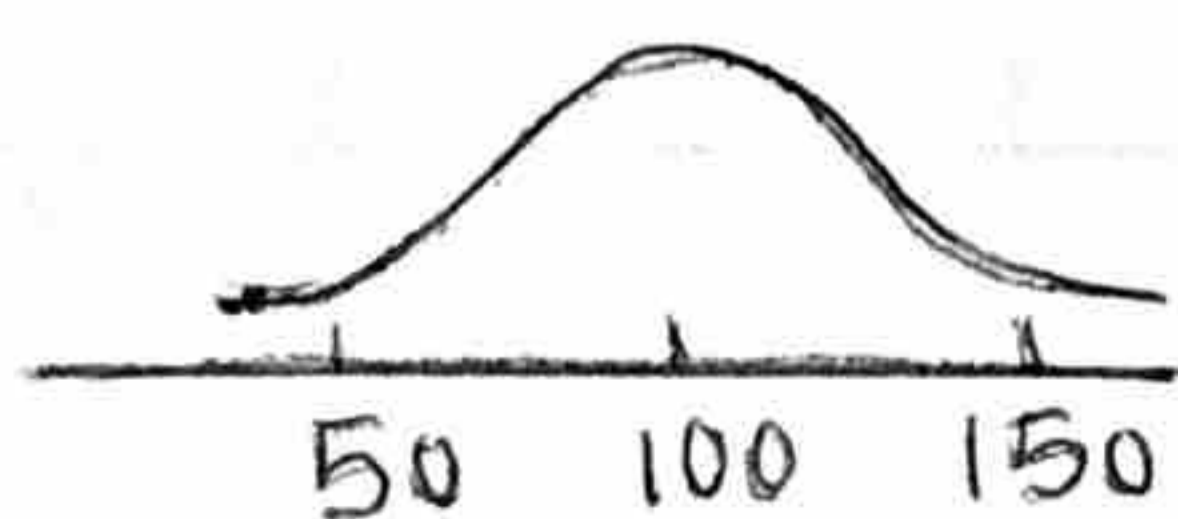
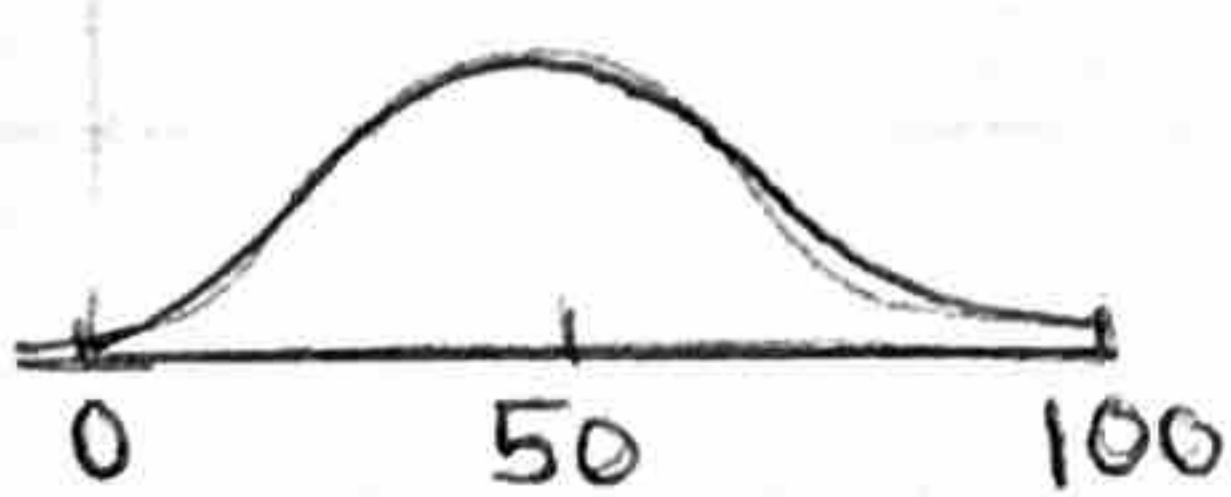
### 1.3.2 ▷ Numerical Descriptive Methods

#### → Measures of center

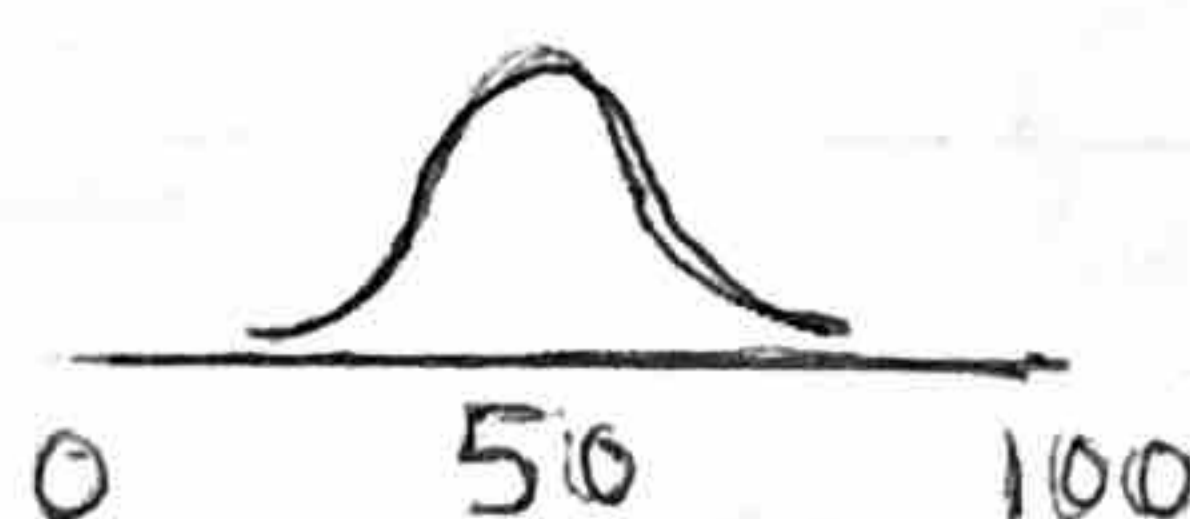
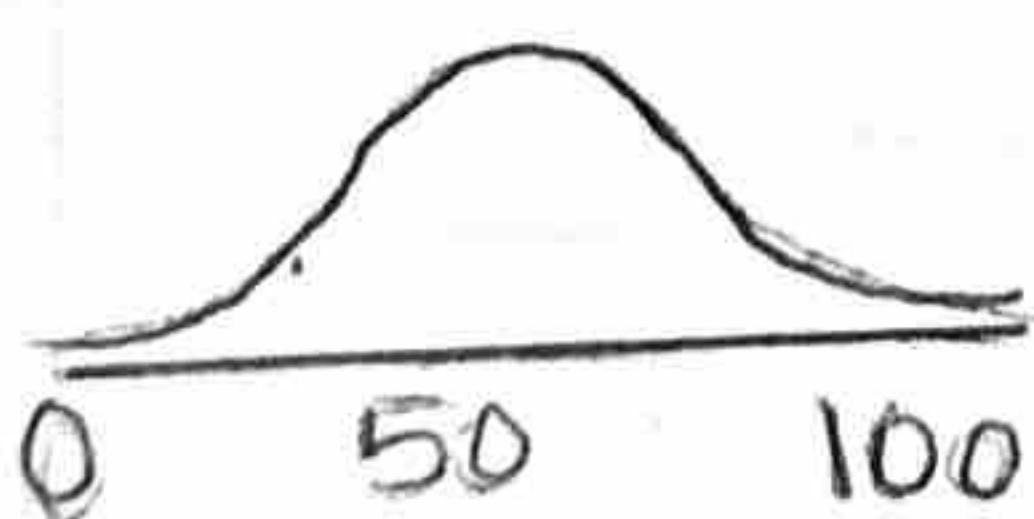
a type of numerical summary to capture most of the info a variable contains in fewer than n numbers.

\* Includes the Mean, the Median, & the Mode

#### • Histogram Shapes Comparisons



\* same shape & spread, different centers

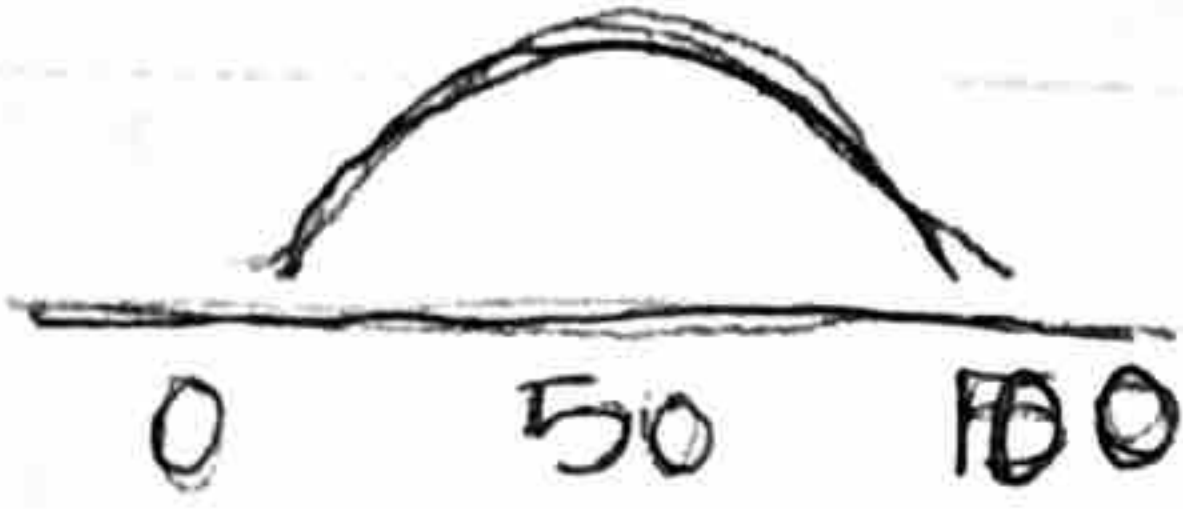


\* same shape & center, different spread



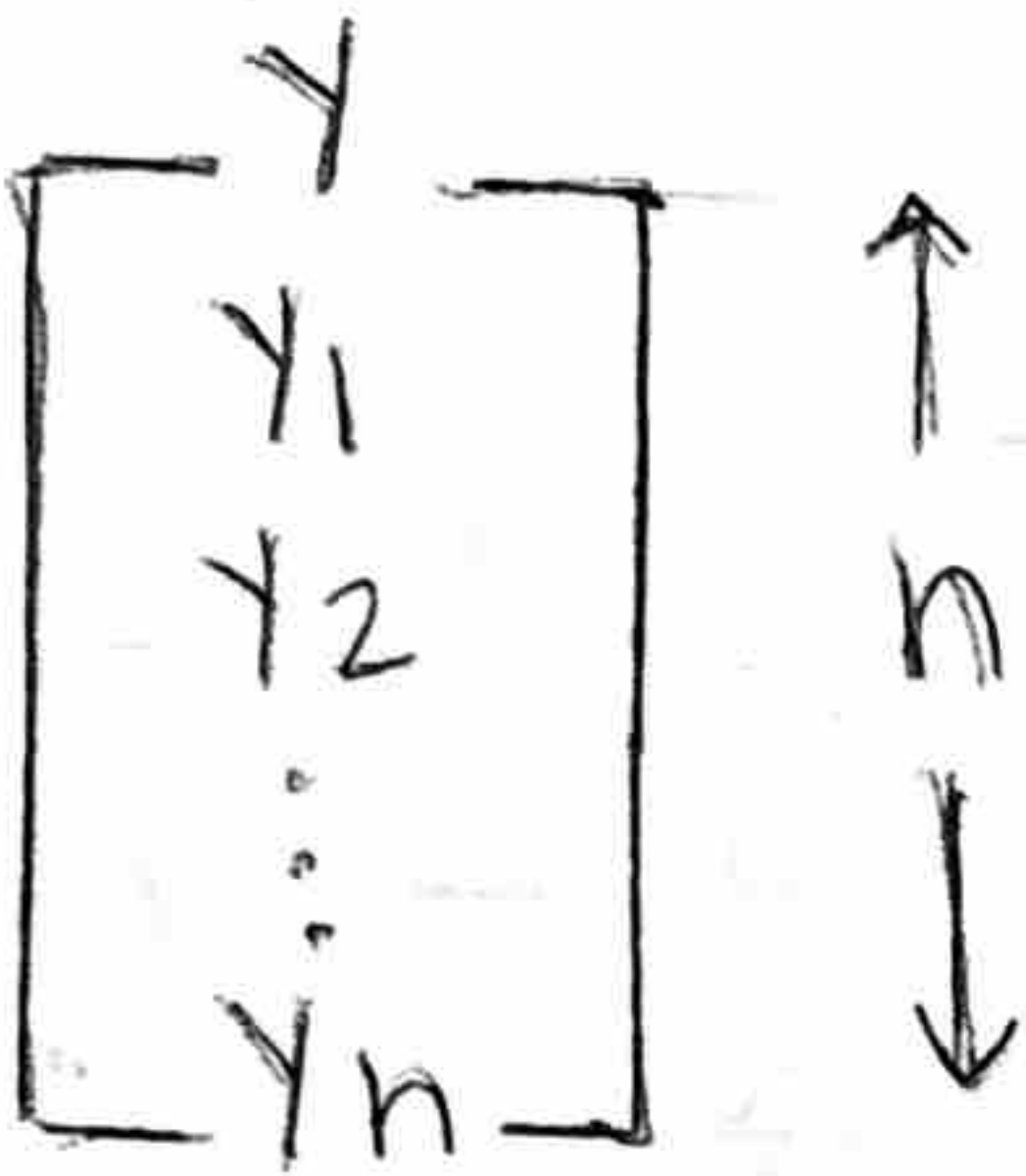
4.10.18

6



\* Same center  
& spread  
diff shape

• The Mean



average = (arithmetic mean)

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} (y_1 + \dots + y_n)$$

*y bar*

$$= \frac{1}{n} \sum_{i=1}^n y_i$$

*summation notation*

$$= \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$

• The Mode

Highest Point of a histogram

