Describing Data

Data → Qualitative (categorical data)

EX.
- eye color
- favorite animal for a pet
- least favorite school lunch
- zip codes
- phone #

Data → Quantitative

- How many pets do you have?
- How many pens do you have in your backpack?

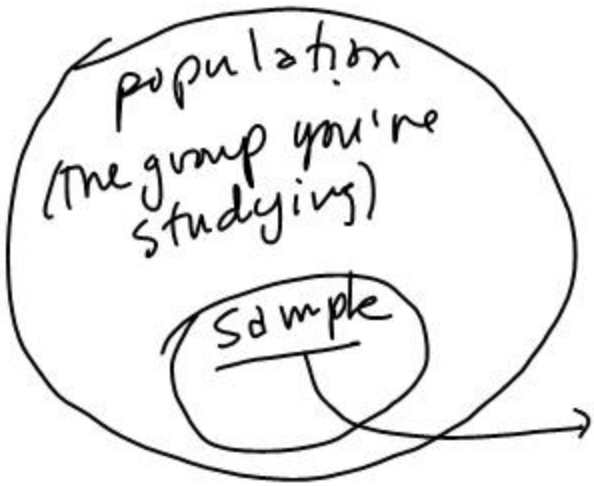
→ discrete data

- height
- weight
- IQ
- Test score

→ continuous data

---

populations + samples



A good sample needs to be random.

→ The members of the population that you measure.

Presenting Data

Frequency tables + bar charts

These are ways to present discrete data.

Number of books read last summer

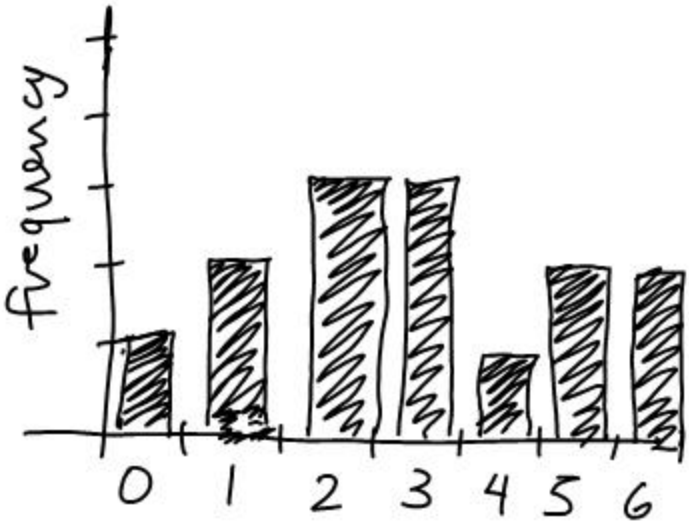
2, 5, 1, 2, 3, 3, 1, 0, 4, 5, 2, 6, 3, 6

Frequency table

Books	Tally	Frequency
0	I	1
1	II	2
2	III	3
3	III	3
4	I	1
5	II	2
6	II	2

HH

bar chart



Note: bars are not touching
(It's discrete data)

Continuous data

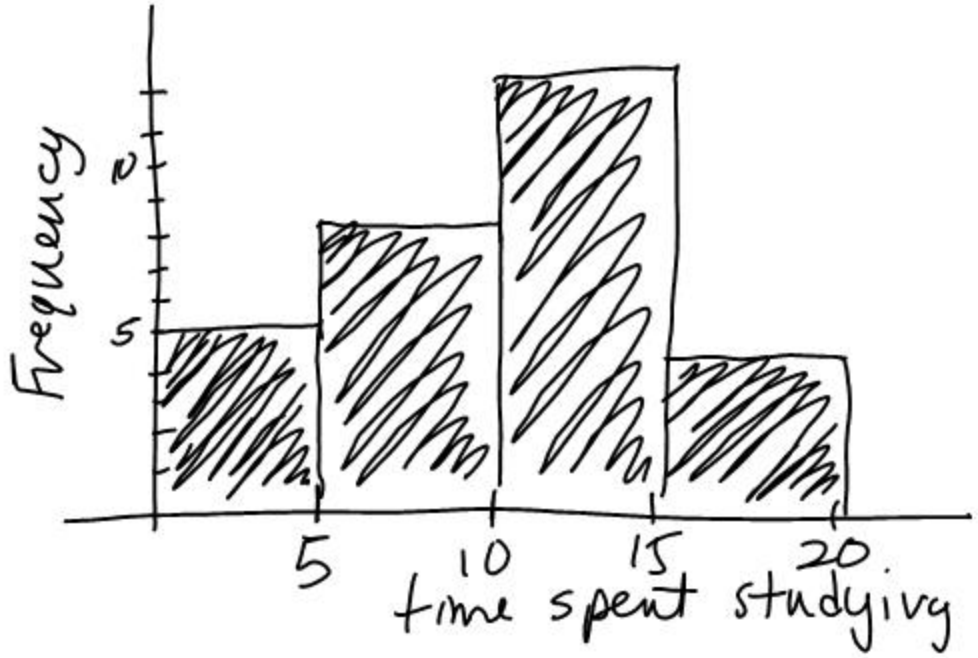
Grouped frequency tables and histograms

Minutes spent on math HW

grouped frequency table

time studying	$0 \le t < 5$	$5 \le t < 10$	$10 \le t < 15$	$15 \le t < 20$
number of students	5	8	12	4

histogram (no gaps between bars)



SB #3 (a) continuous

(b) grouped freq. table

weight	$1 \le w < 2$	$2 \le w < 3$	$3 \le w < 4$	$4 \le w < 5$
number of students	8	24	50	14

(c) 96

Two Important ways to describe a large data set

- Measures of central tendency
  - Mean
  - Median
  - Mode
- Measure of Dispersion (spread)
  - 5 number summary
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation

Ex. Find the 5-number summary for

$[0, 1, 1, 2, 3, 3, 3, 4, 5, 7, 9, 9, 10, 11, 11, 12, 13]$

median $\frac{17+1}{2} = 9$

2.5 is the first quartile (Q_1) (25^{th} percentile)

median or 2nd quartile (Q_2) (50^{th} percentile)

10.5 is the 3rd quartile (Q_3) (75^{th} percentile)

5 number summary: $0, 2.5, 5, 10.5, 13$

min, Q_1, Q_2, Q_3, max

Range: Maximum − Minimum

$13 - 0 = \underline{\underline{13}}$

Interquartile Range: $Q_3 - Q_1$

$10.5 - 2.5 = \underline{\underline{8}}$

Variance & Standard Deviation

Ex. 5, 6, 8, 9 Find variance

$\mu = 7$ ← deviation from the mean

x	$x-\mu$	$(x-\mu)^2$
5	-2	4
6	-1	1
8	1	1
9	2	4
		10 ← sum of the squared deviations

average squared deviation: $\frac{10}{4} = \underline{2.5} = \sigma^2$

variance

Standard deviation: $\sigma = \sqrt{2.5} = \underline{1.58}$

HW 8B #2ab, #4

8C #1

* Find the variance of

10, 14, 15, 21