

Lesson 2: Data Grouped into Classes, Histograms, and  $\bar{x}$ , Md, Mo. May 1st 2023

How to Construct a Table of Condensed Data Grouped into Classes  $[a, b[$   $[b, c[$   $[c, d[$

for a given distribution of continuous quantitative values or many different discrete values.

Pg 32

A hospital has recorded the weight of 25 recent newborns. Here is the data that was collected expressed in grams:

$x = \{ 3155, 3185, 2150, 2255, 3010, 3030, 2770, 2890, 2895, 3035, 3560, 3740, 2895, 3220, 3230, 3275, 3405, 3885, 3105, 3130, 2315, 2580, 2590, 2745, 3530 \}$

$x =$ Weight (g)	Tally	# of babies Frequency	% of babies Frequency (%)
[2000, 2400[		3	12
[2400, 2800[		4	16
[2800, 3200[	 	10	40
[3200, 3600[	 	6	24
[3600, 4000[		2	8
total		25	100% ✓

Should be more than max.

nota bene:

$3200 \notin [2800, 3200[$   
does not belong to the interval

$3200 \in [3200, 3600[$

Including bracket

excluding bracket } the classes are mutually exclusive

$$R.F. \% = \frac{\text{part}}{\text{total}} \times 100\%$$

$$R.F. \% = \frac{\text{freq}}{\text{total}} \times 100\%$$

$$R.F. \% = \frac{3}{25} \times 100\%$$

$$R.F. \% = 12\%$$

$$R.F. \% = \frac{4}{25} \times 100\% = 16\%$$

$$R.F. \% = \frac{10}{25} \times 100\% = 40\%$$

$$R.F. \% = \frac{6}{25} \times 100\% = 24\%$$

$$R.F. \% = \frac{2}{25} \times 100\% = 8\%$$

8% of babies weighed [3600, 4000[ grams.

interval  $[a, b[$   $[b, c[$   $[c, d[$   
min  $\rightarrow$  max  
lower endpoint  $\rightarrow$  upper endpoint  
width =  $d - c$

step i: Pick # of classes from 5 - 10.

ex. 5 classes

step ii: Identify the min, max, range of dist.

$$\text{min} = 2150 \text{ g}$$

$$\text{max} = 3885 \text{ g}$$

$$\text{Range} = \text{max} - \text{min}$$

$$\text{Range} = 3885 - 2150$$

$$\text{Range} = 1735 \text{ g}$$

step iii: Determine the widths of the classes.

Pick a nice rounded # larger than:

$$\frac{\text{Range}}{\# \text{ of classes}}$$

$$\frac{1735}{5}$$

$$347$$

400 (easier than 350)

step iv: Determine the 1st lower endpoint of class.

Pick a nice rounded # smaller than min = 2150

ex 2000

step v: Construct table + put row titles on memory aid

You do

**3.2 Data Tables Grouped into Classes**

We surveyed a class of 25 History students and asked for their final grade in the course. Here are the results:

Class	Tally	Frequency	Rel. Frequency
Total			

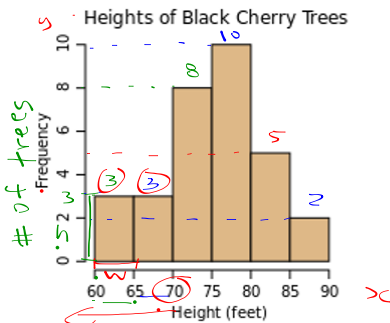
47 91 69 70 80  
 81 57 99 78 74  
 63 69 69 70 80  
 55 77 50 53 85  
 77 81 48 65 88

6 classes

It makes more sense to group this data into classes (i.e. a range of numbers or intervals). Let's construct the grouped table and determine which percentage of students passed the course.

## Reading Histograms (no spaces between bars)

Example: Students surveyed the heights of black cherry trees at a local orchard and presented their results in the following histogram:



Read the graph and answer the following questions:

Question 1: How many trees had a height between 60 to 65 feet?

3 trees

Question 2: How many trees had a height between 65 to 70 feet?

3 trees

Question 3: How many trees had a height less than 70 feet?

6 trees

Question 4: What else is the graph communicating?

there are 31 trees w/ a height in between 60-90 feet

sample size { 70 feet

Question 5: Which axis are the survey responses/results represented on?

x-axis

Question 6: Which axis are the frequencies (or relative frequencies) represented on?

y-axis

Question 7: What do the frequencies represent in this example?

# of trees

Question 8: Do the vertical bars always have the same width?

yes

Question 9: Are there any spaces between the bars?

No

Question 10: What type of variable is represented on the x-axis (discrete quantitative, continuous quantitative, or qualitative)?

continuous (no spaces)

discrete (spaces)

Read titles

Think - Pair - Share

→ Answer the question

→ Share answers w/ partner.

## Constructing Histograms of a Distribution of Raw Data

A hospital has recorded the weight of 25 recent newborns. Here is the data that was collected expressed in grams:

3155, 3185, 2150, 2255, 3010, 3030, 2770, 2890, 2895, 3035, 3560, 3740, 2895, 3220, 3230, 3275, 3405, 3885, 3105, 3130, 2315, 2580, 2590, 2745, 3530.

$x$ Weight (g)	Tally	(general) Frequency	Relative Frequency (%)
[2000, 2400[ <small>min</small>		3	12
[2400, 2800[		4	16
[2800, 3200[	 	10	40
[3200, 3600[	 	6	24
[3600, 4000[		2	8

step i construct a table of data grouped into classes (same steps as before)

step ii Label the survey responses as  $x$  and the freq. or rel. freq. as  $y$ .

step iii Find for  $y$ -variable

$\min_y = 0$

$\max_y = 40$

$\text{Range}_y = \max - \min = 40 - 0$

$R_y = 32$

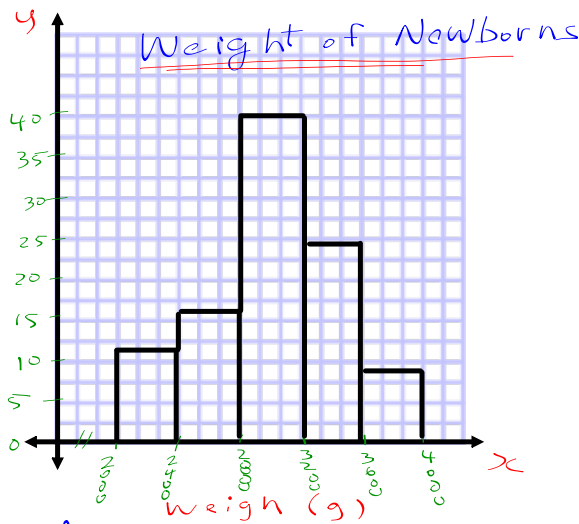
$\left[ \begin{matrix} \text{go} \\ \text{up} \\ \text{by} \end{matrix} \right] = \frac{\text{Range}}{5} = \frac{32}{5} = 6.4$   
any # between 5-10

$\therefore \left[ \begin{matrix} \text{go} \\ \text{up} \\ \text{by} \end{matrix} \right] = 5$

step iv: Graduate the  $x$  /  $y$  axis (Break axis if min is far from 0)

step v. Draw bars w same width, no spaces, w heights corresponding to (relative) freq.

step vi For full mark, write the 3 titles



. Relative Freq.  
. % of babies

You do

pg 35 of textbook # 1.13 (check answer)

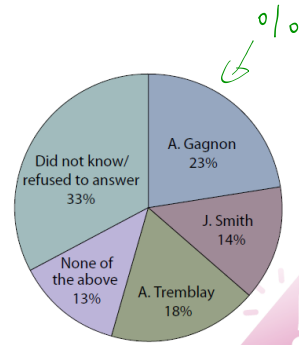
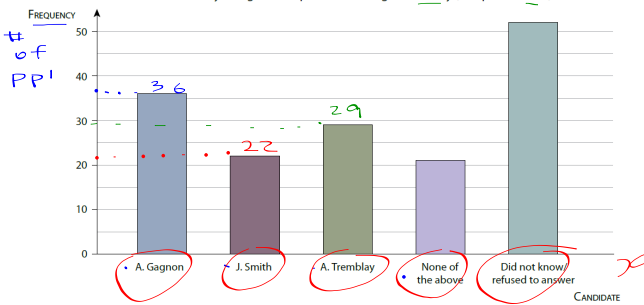
less than 15 mins (@ 11h12)

# Reading Bar and Circle Graphs

→ Read titles

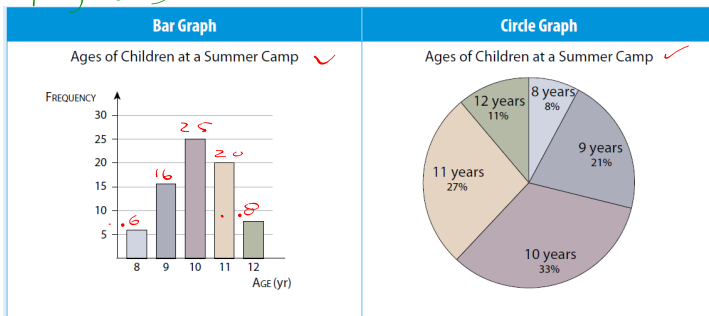
Turn page 22 in textbook

✓ Mayor  
**Graph 1.2**  
 Candidate Most Likely to Fight Corruption According to a Survey (sample size: 160) → people



What is the variable of survey response?  
 → qualitative ✓

pg 23



variable: quantitative  
 discrete

pg 50 of textbook  
 for a summary

- READ
- put on memory aid

## Data Collection

→ do surveys

→ look @ results. How? Look at:

→ distribution of results [raw + long]

→ data tables/graph [organized + long]

→ 3 measures of central tendency of distribution of results ( $x$ )

- #1 Mean (Average) - a data value that summarizes the central trend/result of dist. ( $\bar{x}$ )  
(fig. most important)
- #2 Median (Middle) - the (literal) centre data value of an ordered distribution. ( $M_d$ )
- #3 Mode (Frequent) - the data value that appears the most / the most frequent. → has highest frequency. ( $M_o$ )

## Determining the Measures of Central Tendency

e.x. given the following grades,  
Find the mean, median, and mode

95, 80, 52  
60, 83, 82  
60

#1.

$$\text{mean} = \frac{\text{sum of data values}}{\# \text{ of data values}}$$

$$\text{mean} = \frac{(95 + 80 + 52 + 60 + 83 + 82 + 60)}{7}$$

$$\text{mean} = 73.1\%$$

#2 For median:

step i. order the values (min to max)

52, 60, 60, 80, 82, 83, 95  
 1<sup>st</sup> 2<sup>nd</sup> 3<sup>rd</sup> 4<sup>th</sup>  
 3 values      " md      3 data values

step ii Find median's position  $\bar{w}$  equation

$$P_{Md} = \frac{n+1}{2} \quad \text{where } n = \# \text{ of data values}$$

$$P = \frac{7+1}{2}$$

$$P = 4^{\text{th}} \text{ position}$$

$$\therefore Md = 80\%$$

step iii check that median divides distribution into 2 equal parts (sub-distribution)

#3

$$Mo = 60\%$$

You do Ex 1.2

→ 5 min.  
moving on  
@ 12:40

If finished read the additional notes on Choosing Between Mean, Md + Mo.

check double count

Going Further  $\bar{w}$  Mean, Md, Mo.

e.x. consider the grades

$$x = \{ \overset{x_1}{70}, \overset{x_2}{72}, \overset{x_3}{72}, \overset{x_4}{84}, \overset{x_5}{84}, \overset{x_6}{85} \}$$

# 1.

mean =  $\frac{\text{add up all values}}{\# \text{ of values}}$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_6}{6}$$

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

start  $\rightarrow$   $\bar{x} = 77.8\%$

$n = \#$  of data values.

$\sum$  greek letter "sigma"

in math « take the sum of »

$x_i$  - individual data value

#2 Median of dist.  $\bar{w}$  even  $\#$  of data values

e.x. consider the grades

$$x = \{ \overset{x_1}{70}, \overset{x_2}{72}, \overset{x_3}{72}, \overset{x_4}{84}, \overset{x_5}{84}, \overset{x_6}{85} \}$$

$\leftarrow$   $\overset{78}{\text{md}}$   $\rightarrow$

$$P_{md} = \frac{n+1}{2}$$

$$P_{md} = \frac{6+1}{2}$$

$$P_{md} = 3.5$$

3<sup>rd</sup> / 4<sup>th</sup>

Median is the middle/average of those values

$$md = \frac{72 + 84}{2}$$

$$md = 78\%$$

average of 3<sup>rd</sup> + 4<sup>th</sup>

check if median divides dist. into 2 equal parts

#3. Mode = 72% and 84%  
bimodal

- You do #1.2.1
- if done continue handout
- if start homework handout #4.

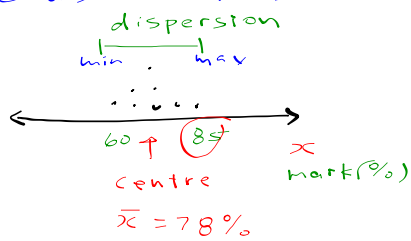


width / length

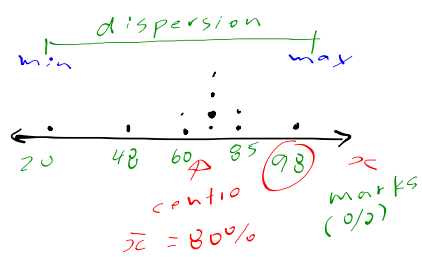
## Understanding Measures of Dispersion

Ex : Range - tells us how dispersed/spread out the data values are.

Ms. Short's class marks



Mr. Wong's class marks



S	W
+++	

nota bene → the less dispersed a dist. is, the more reliable the measures of central tendency are.

## Weighted Mean:

Takisha is evaluated in her mathematics class according to the following scheme:

Breakdown of grades

Assignments	10%
Quizzes	30%
Math Portfolio	15%
Tests	45%

i.e. Assignments are worth 10% of final mark.

Takisha received 83% on assignments, 73% on quizzes, 89% on her math portfolio and 80% on tests. What is Takisha's final grade? **The final grade is a weighted mean!**

	Weighting (%)	Takisha's results
A	10%	83%
Q	30%	73%
MP	15%	89%
T	45%	80%

$$\bar{x} = \% \text{ of } x_1 + \% \text{ of } x_2 + \dots + \% \text{ of } x_n \quad \begin{matrix} \text{per cent} \\ \div 100 \end{matrix}$$

$$\bar{x} = \% \times x_1 + \% \times x_2 + \dots + \% \times x_n \quad \text{where } \% \text{ expressed as decimals.}$$

$$\bar{x} = 10\% \times 83 + 30\% \times 73 + 15\% \times 89 + 45\% \times 80$$

$$\bar{x} = \frac{10}{100} \times 83$$

$$\bar{x} = 0.1 \times 83 + 0.3 \times 73 + 0.15 \times 89 + 0.45 \times 80$$

$$\bar{x} = 79.55\% \quad \text{You do } 3.1 + 3.2$$

- HMWK:
- pg 36 #1.14
  - pg 41 #1.20
  - pg 40 #1.19
  - pg 62 table 2.5 ...
  - pg 65 #2.10
  - pg 7 #1.5