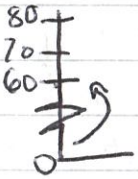
∴ = therefore ; thus

BASIC GRAPHS 9/22/2023

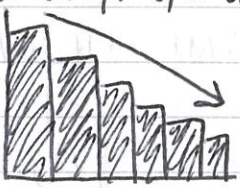
visual representations of data
- pi chart
- line graph } basic basic
- bar graph

- people will only look at it w/ abt 7 seconds
- starts with TITLE
- ends with Analysis
  - highs / lows / trends AND why did you get what you got

BAR GRAPHS
- categorical data
- can be | or —
- TITLE and label axis
- keep increments and width the same



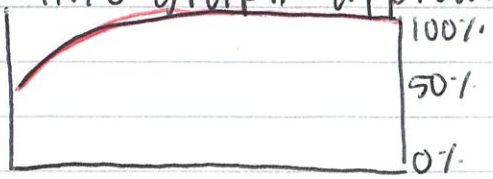
- "squiggle" - one time mulligan; can let you jump
- pareto charts - vertical bar graphs arranged tall to small (always go down)

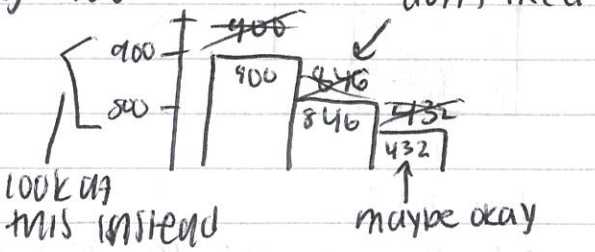


put number 10 —

NOT 10 —

- Ogive - line graph approaching 100

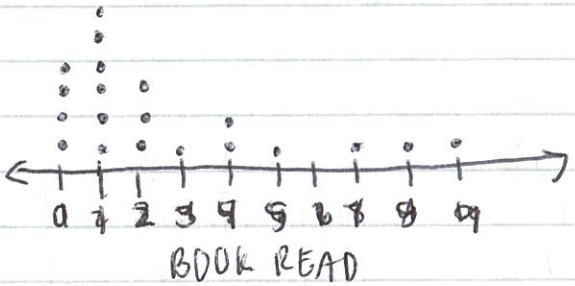




PIE CHARTS
- categorial data
- good for %
- circle or disc divided
- put # or % inside sector

Example:

Books read last summer



ANALYSIS:

center: 1.5 (median)

shape: skewed right

spread: ch. 3

outliers: none

HISTOGRAMS

- Quantitative data
- width of bar has meaning
- large amounts of data
- shapes helps us recognize patterns
- deviations can be hidden
  - specific data is hidden
- $w = \frac{\text{maximum} - \text{minimum}}{\#\text{ classes}}$ ★ always round up
- usually don't start with 0; use lowest data then step by width
- use class limits ~ limit bars/ classes

$\frac{\text{hi} - \text{lo}}{\#} \quad \frac{44-14}{10} = 30$

hungry hippo the last class to grab the remainder

class limits	Tally	Freq.	Relative Freq
14-16	IIII I	6	6/150
17-19			
20-22			
23-25			
26-28			
29-31			
32-			
...			

3 ↺ 3 ↺

2.3 STEM & LEAF DISPLAYS

1. Find max and min
2. Draw vertical line to split stem and leaf
3. Data flow
4. Clean it up and put in order

★ Title, key, analysis (CCSO) include units!

C: 7-11 = 9 (median)

S: skewed right (high) 2|3 = 23 seconds away

S: Ch3

O: 37

0	0 3 8
1	2 7 4 5
2	2 3 6

Pg "Off"

stem	leaf
0	1 2 2 3 3 4 7
1	1 3 5 7
2	1 3
3	7

1 min ex:

min: 1

max: 37

stem		leaf
(0-4) (5-9)	0	7 2 4 2 1 2 3
(10-14)	1	5 3 1 7
	2	1 3
	3	7
	4	

Key: 2|3 = 23 seconds off

Analysis (M)

Center:

Shape: skewed small

Spread: Ch3

Outliers: 35 min

tens	ones
0	1 2 2 2 3 4
1	7
2	1 3
3	5 7
.	1 3
.	
.	
3	7

↑ spacing matters

How long is my trek to school (min)

① max and min

35 3

② M F

6 7 3	0	5 5 5 5 7 5 7 8 8 5
0 5 0	1	0 3 5
5 7 0 0	2	5 5 5 8
5	3	5

③

KEY:

F 3|5 = 35 min commute

M 0|2 = 20 min commute

④

7 6 3	0	5 5 5 5 5 5 7 7 8 8
5 0 0	1	0 3 5
7 5 0 0	2	5 5 5 8
5	3	5

Analysis (F):

Center:

Shape: skewed high

Spread: Ch3

Outliers: 35 min

1.4 - 1.6
AP Classroom ☰ Notes ✍ 10/01/2023

Representing a categorical variable with graphs:

Making bar charts:
- label axis (x-axis = variable name y-axis = freq/relative freq.
- scale axis
- draw bars
  - equal width lengths
  - space between bars

Pie charts:
- include legend/key
- relative frequencies are better when comparing tables with different proportions

Frequency ~~charts~~ tables
- label axis
- scale axis
- draw bars
- add key

1.5 representing a quantitative variable with graphs
categorical variable — variable that takes values with names or labels
quantitative variable — variable with numerical value for a measured or counted quantity
- discrete — variable that can take on a countable # of values (w/ gaps) e.g. # of sibs
- continuous ~ can take infinitely values, but can't be counted (no gaps) e.g. height

3.1 Stats Recording Notes 10/05

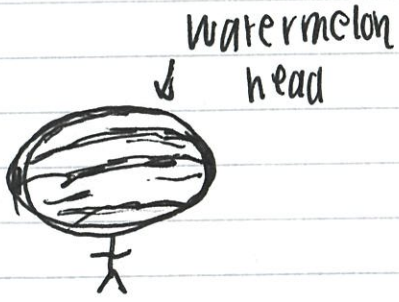
HOW Many Siblings?
0 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 5

The Averages
- Mean
- Median
- Mode
- Trimmed Mean

The Arithmetic Mean
- $\frac{\Sigma x}{n}$ (sum numeric values ÷ # of values)
- $\bar{x}$ bar for sample $\bar{X}$ $\quad \frac{367}{5} = 73.40$
- μ for a population
- size matters ~ numeric value

0. _____
1. _____
everything else = _____] rounding rules

- sean med 440 sibs affects the mean $\quad \frac{447}{28} = 15.96$
- Disadvantage – easily affected by outliers



The median = 1
- middle value when #s are in order

The Mode = 1
- most often occuring value
- outliers don't affect mode
- can be bimodal and no mode

Trimmed means
- resists extreme
- EX: 10% trimmed mean

$\frac{\Sigma x}{n} = \frac{40}{23} = 1.7391$

10% of 29 (cut 3 highest and 3 lowest values)
0.1 × 29 → 3
2.9

Measures of Variation 10/09/23

Measures of Variation

study guide: pg 102-110
- one # may not represent an entire set of #s well so, a cross reference is measuring the spread of data
- name 3 measures of variation
  - ☐ range
  - ☐ variance
  - ☐ standard deviation or σ
- provide advantages and disadvantages of the range
  - ☐ doesn't tell how much other values vary from each other or from the mean ✗ doesn't tell much
  - ☐ tells the spread of the data ✓
- the measurement that helps us see how the data is different from the mean is standard deviation
- why divide by n-1 sometimes and by N other times
  - ↳ sample standard deviation
  - ↳ population standard deviation
- obtuse Ollie believes he has discovered an easier method for ex 6 (Big Blossom Greenhouse) $\Sigma(8-\underline{9.5})^2/7$
  - ↳ median not mean
- compare and contrast sample and pop. formulas for the mean and standard deviation

$\frac{\Sigma}{n}$ we deleted during

STANDARD DEVIATION: $s=?$ $s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

X: 2, 4, 6, 8, 10

$\bar{X}$: $30/5 = 6$

$X-\bar{X}$	$(X-\bar{X})^2$
-4	16
-2	4
0	0
2	4
4	16

$40 \div 4 = 10$ $\sqrt{s^2}$ $n-1$

$3.16 = s$

Tab

2.10 1.95 2.60 2.00 1.85 2.25 2.15 2.25

$\mu = \$2.14$

$\sigma = 0.22$

$CV = \frac{\sigma}{\mu} = \frac{0.22}{2.14} = 0.1028 = 10.28\%$

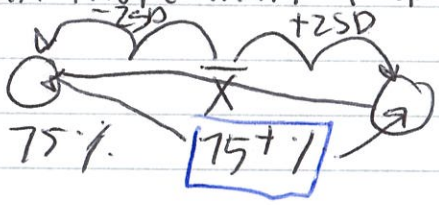
Chebyshev's Theorem

$1 - 1/k^2$ at least more than 1 SD

2sd 3sd 4sd

ex: 2 = k

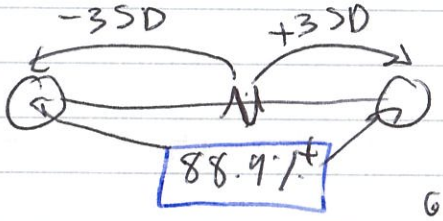
$1 - \frac{1}{4} = 3/4 = 75\%$



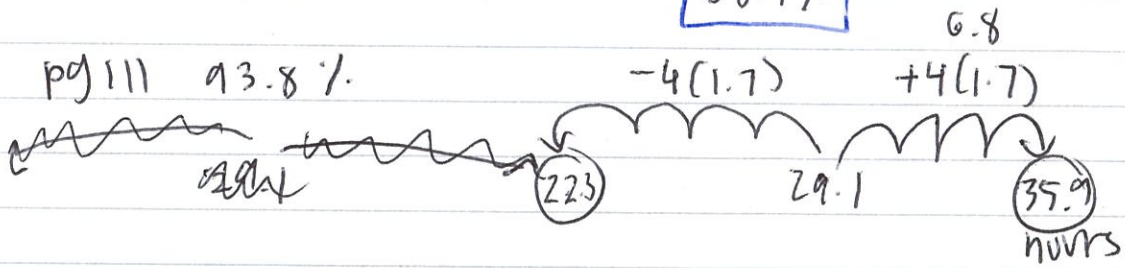
ex: k=3

$1 - \frac{1}{9}$

$8/9 = 88.9\%$



8 pg 111 93.8 %.



at least 93.8% of students would fit into the group that varies from 22.3 to 29.1 hours each semester

Number Summary & Box and Whisker Plot 10/11

P_{99} = 99%

Q_1 Q_2 Q_3

↑

P_{50} aka Median

At what age would you panic at?

MIN

(1) 7 10 14 18 20 21 | 21 22 22 22 23 23 24 | 24 | 25 25 25 25 25

25 | 26 26 27 27 28 28 (30)

MAX

Q_2 (P_{50}) = median

Q_1 (P_{25}) = 21 LQ

IQR = $Q_3 - Q_1$

= 25.5 − 21 = 4.5

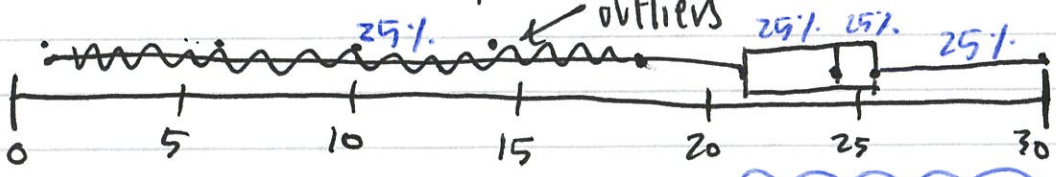
Q_3 (P_{75}) = 25.5 UQ

- trims the low and high
- resistant to extremes

5 # summary: Max, Min, Q_1, Q_2, Q_3

Box and whisker plot

P4 Dreamy Ages



AGES

outliers

$Q_3 \pm 1.5(IQR)$

ANA:

- Center – 24 (median)
- Shape – skewed low / ~~symmetri~~ uniform
- Spread – 4.5 IQR "banded" right / clumped
- Outliers – 1, ~~5~~, 7, 10, 14

25.5 + 1.5 (4.5)

25.5 + 6.75

32.25

no outliers

25.5 − 1.5(4.5)

18.75

outlier → 14.25

CHAPTER 2 = Midterm ch 1, 2, 3 ★ 10/17/2023

chapter 2 study guide 10/17

Key concepts: Frequency distributions, histograms, etc.

~~FREQUENCY TABLES~~ BASIC GRAPHS

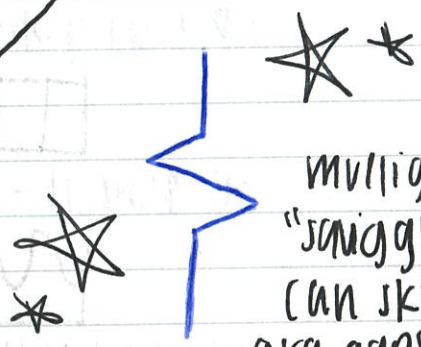
Bar, Pie, Line, Picto

NECESSITIES / INFO:
- represents data visually, quickly, and understandably
- people only look at it for 7 seconds
- TITLE and (maybe) x/y axis titles
- Analysis!

ANALYSIS:
High
Low
Trends
Why?

BAR GRAPHS
- can compare multiple subjects
- categorical data
- equal increments (steps)
- same bar width
- x-axis = variable
- y-axis = frequency

mulligan: "squiggle" can skip big ~~gra~~ gaps once



→ Pareto charts ~ bar graph from tallest to smallest

CIRCLE (PIE) GRAPHS
- compare parts of a whole /100%
- categorical data
- #/% inside slices OR use a key

↖ ignore

LINE GRAPHS
- equal increments
- connected dots
- only use squiggle at the start
- can have multiple lines to compare

~~Arrogance~~

✗ CONJECTURE - cause/why smt is happening

→ Time plots - type of line graph w/ x-axis as time!

PICTO GRAPHS
- cute symbols
- equal sized symbols that are simple
- full/partial symbols in a KEY

♡ = 50 smiles
♡ = 25 smiles

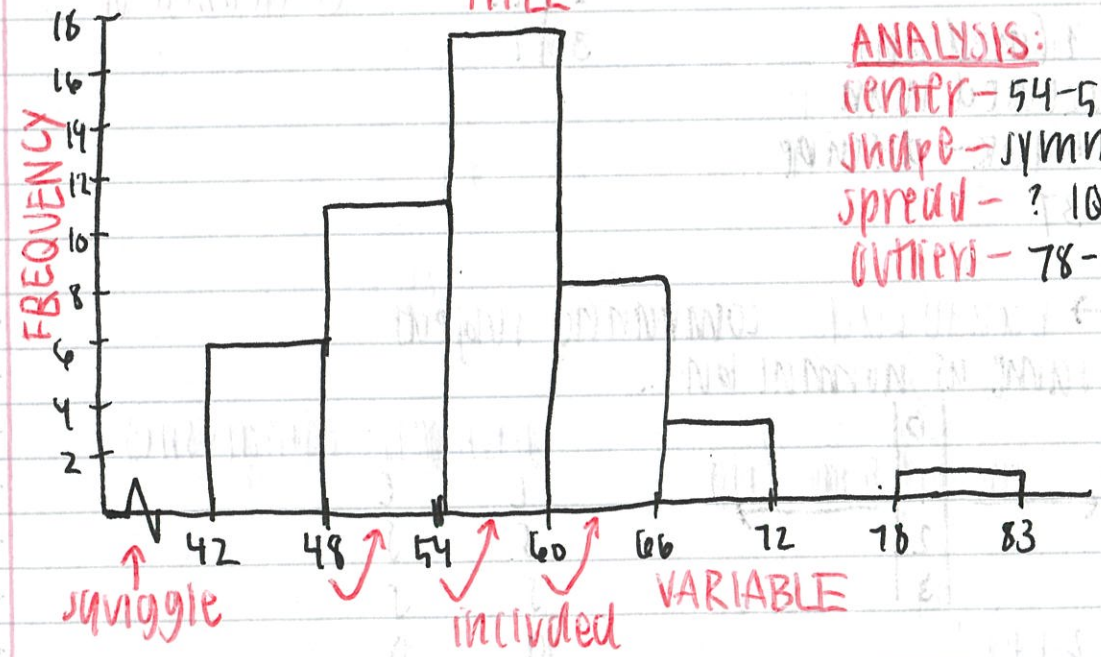
→ Example #2: presidents problem

class width = $\frac{\text{max} - \text{min}}{\text{classes}} = \frac{78-42}{7} = 5.14 \rightarrow \boxed{6}$

Frequency Table:

	class	tally	freq.	relative freq.
1	42 - 47	𝍸 I	6	6/46 = 0.13
2	48 - 53	𝍸 𝍸 I	11	11/46 = 0.24
3	54 - 59	𝍸 𝍸 𝍸 II	17	17/46 = 0.37
4	60 - 65	𝍸 III	8	8/46 = 0.17
5	66 - 71	III	3	3/46 = 0.07
6	72 - 77		0	0/46 = 0
7	78 - 83	I	1	1/46 = 0.02

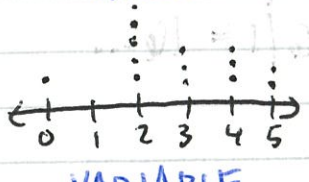
Histogram:



state

ANALYSIS:
center - 54-59 (median)
shape - symmetric / normal
spread - ? IQR - *describe*
outliers - 78-83 → state

DOTPLOTS:
- good for quantitative data
- not good for high spread
- shows original, specific data!



ANALYSIS:
C
S
S
O

CHAPTER 3:

Chapter 3 study guide 10/17/2023

WHAT DOES "MEAN" MEAN???

measures of central tendency – arithmetic mean, median, mode, trimmed mean

THE AVERAGES (pretty average ngl :/)

MEAN (ARITHMETIC):
- $\bar{x}$ = mean of a sample
  μ = mean of a population
- numerical values → size matters
- disadvantages: pulled/affected by outliers

$\frac{\Sigma x}{n}$ = $\frac{\text{sum of values}}{\#\text{ of values}}$

MEDIAN:
- middle value when values are put in order small → large
- advantages: ignores outliers unlike arithmetic mean

MODE:
- categorical OR quantitative
- most often occuring value
- Advantages: ignores outliers

TRIMMED MEAN:
- mean the eliminates extremely low/high values
- Ex: 10% trimmed mean
  ↳ 0.1 X n
  0.1 X 29 = 2.9 → 3 = trim 3 highest and lowest values

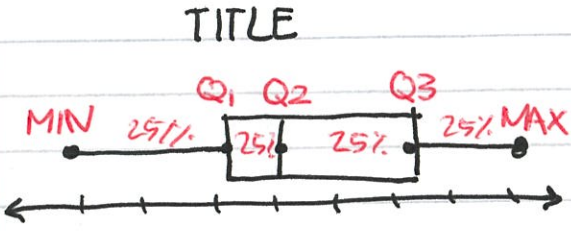
ROUNDING RULES!!
0. _ _ _ _ _
1. _ _ _ _
anything else. _ _

CENTER!

MEASURES OF VARIATION

OVERVIEW:
- r = range → high - low (useless)
- CV = coefficient of variation
- IQR = inner quartile range
- s = sample standard deviation / σ = population standard deviation
- s^2 = sample variance / σ^2 = population variance
- Chebychev's theorem
- empirical rule

BOX-AND-WHISKER PLOT:



5# SUMMARY:
- MAX
- MIN
- Q1 P_{25}
- Q2 median P_{50}
- Q3 P_{75}

VARIABLE

ANALYSIS:

- spread before OR after outliers — C, S
- cut out outliers (if you want) — S, O

$Q_3 \pm 1.5(IQR)$

GOOD LUCK!