

Unit 4: Organization of Data for Analysis

NOV. 5/2018

Data Concepts & Graphical Summaries

Statistics is the gathering, organization, analysis, and presentation of data.

Types of Data:

(1) Numerical (quantitative) data are measurements.

(a) discrete data have specific values.

e.g., # correct answers, # of blue socks

(b) continuous data can have any value in a range.

e.g., weight, time, volume, temperature

Nov 4-5:42 PM

Types of Data:

(2) Categorical (qualitative) data can be counted, but the data items themselves are not measurements.

(a) ordinal data can be ranked (ordered), and may have numbers assigned to them

e.g., rating scale 1-10, poor/fair/good/excellent

(b) nominal data cannot be ranked

e.g., eye colour, vehicle make/model, favourite...

Displaying Data (see handout)

Nov 4-5:42 PM

Frequency Table (Tally Chart)

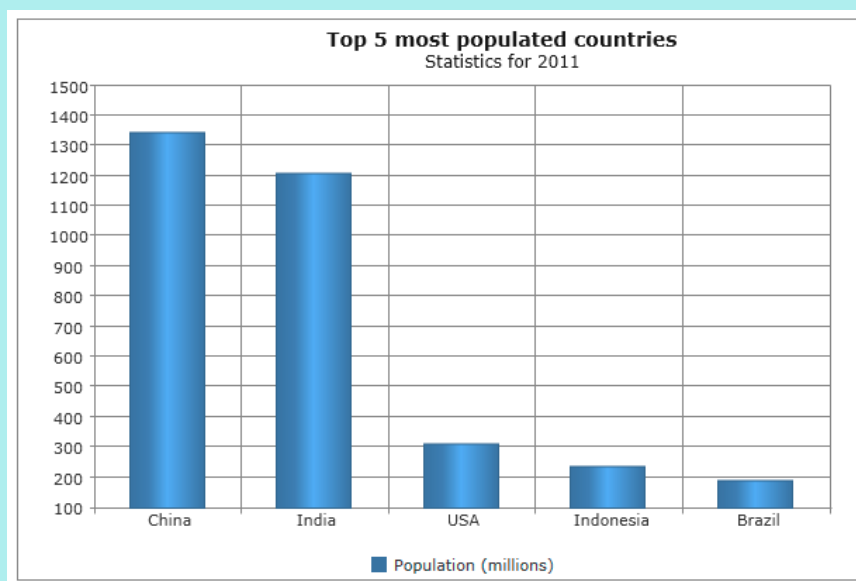
Marks	Tally	Frequency
1	///	3
2	///	3
3	//	2
4	//	2
5	//	2
6	#####	5
7	////	4
8	#####	5
9	//	2
10	//	2
Total		30

- counts occurrences of each criteria being measured
- can use for all data types

Nov 4-5:54 PM

Bar or Column Chart

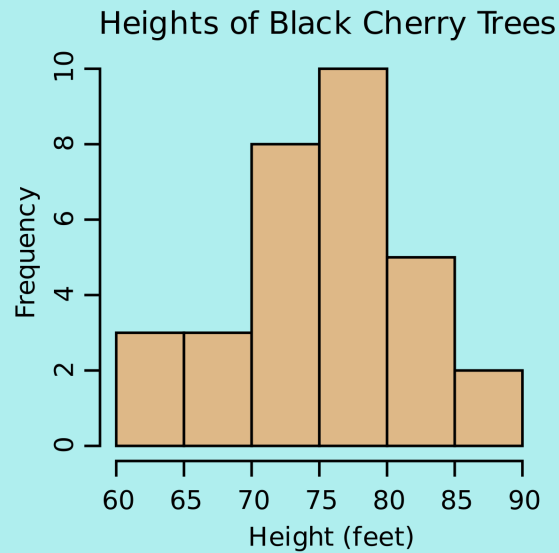
- good for discrete data or categorical data
- bars do not touch



Nov 4-6:01 PM

Histogram

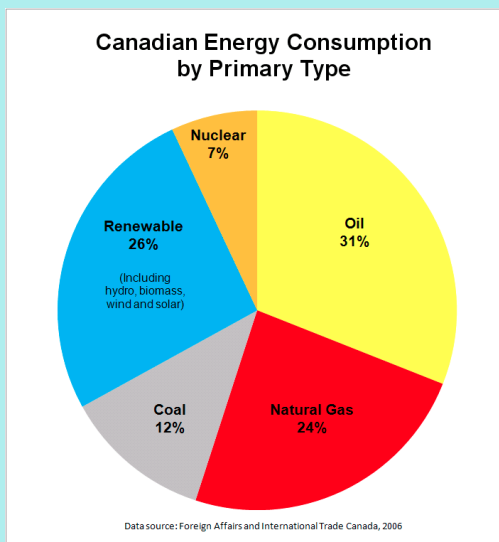
- use with continuous data
- displays data using intervals of fixed width
- bars touch to show continuity of data across intervals



Nov 4-6:03 PM

Circle Graph (Pie Chart)

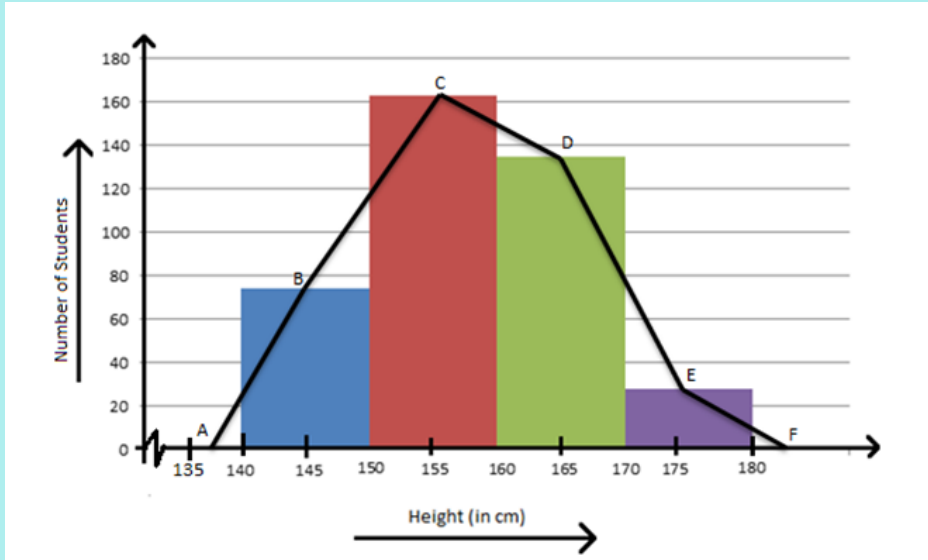
- used for categorical data
- good visual representation of part to whole



Nov 4-6:05 PM

Frequency Polygon

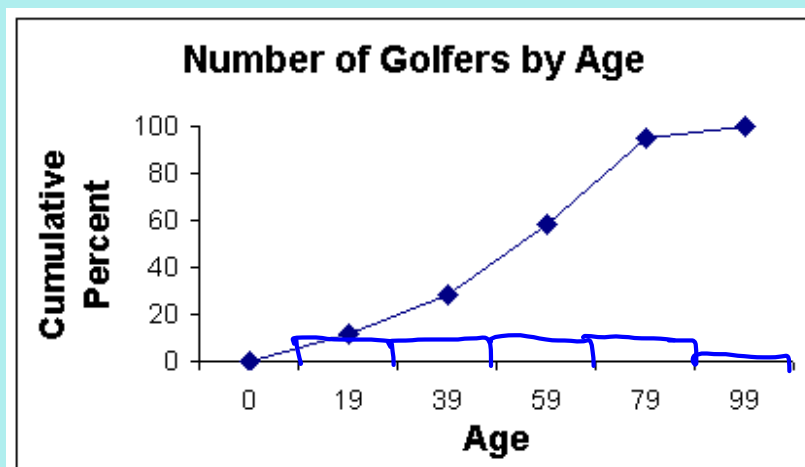
- connects midpoints on bar graph or histogram with a line
- relative frequency polygon shows same graph shape, but scaled to percentages (100% total).



Nov 4-6:06 PM

Cumulative Frequency Polygon

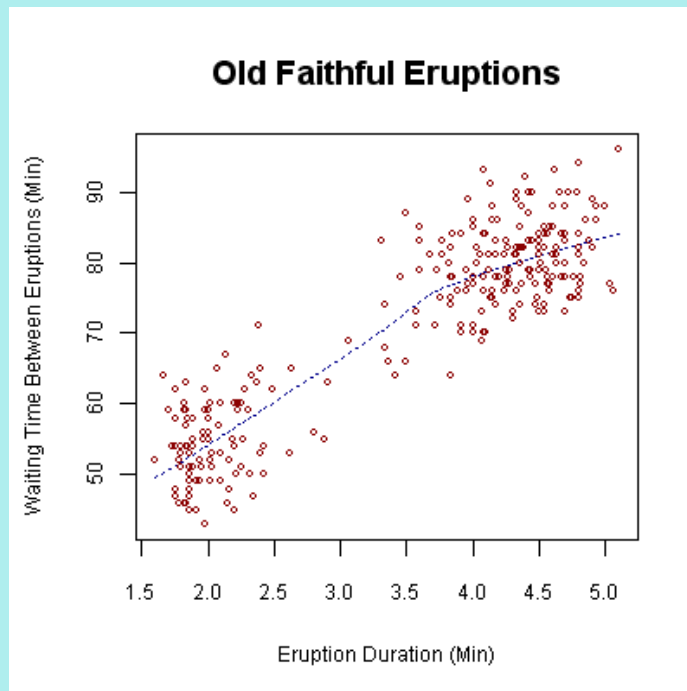
- shows rate of change in frequency from one interval to the next



Nov 4-6:07 PM

Scatter Plots

- shows relationships between two variables
- can be used to determine correlations and trends in data



Nov 4-6:11 PM

When drawing a histogram (continuous, numerical data), you may need to limit the number of intervals used to describe or graph the data.

1. Determine the number of intervals you want to present. Usually between 5 to 20, with around 8 best.

2. Make all intervals equal in width.

$$\text{bin width} = \frac{\text{max} - \text{min}}{\# \text{ of intervals}}$$

3. Use interval notation to define your intervals and make sure there are no gaps between them.

Nov 4-7:12 PM

Ex. Create a frequency table and simple histogram using five intervals for the following data.

13,7,5,7,9,10,5,11,8,7,9,10,10,11,14,10,6,12,6,9,7,12,9,6

① intervals

→ interval width?

→ intervals to use?

$$\max = 14 \quad \min = 5$$

$$\text{width} = \frac{14-5}{5} \leftarrow \text{my choice, no "best" answer}$$

$$= 1.8$$

$$\approx 2$$

interval	tally	frequency
[5, 7)		5
[7, 9)		5
[9, 11)		8
[11, 13)		4
[13, 15)		2

does not include 7
 $[5, 7)$
 ↑ includes 5

Nov 4-7:19 PM

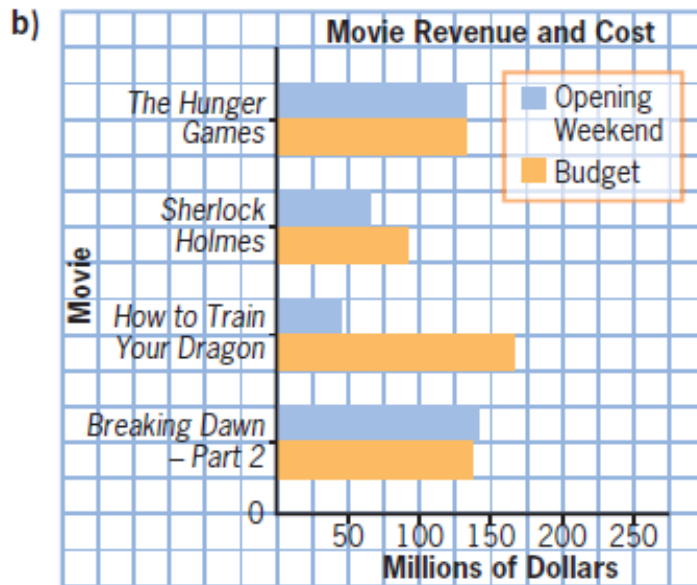
Assigned Work:

p.194 #1 - 3 **1b**

p.203 #1, 3, 4, **8**, 9

c

Nov 4-7:21 PM



Nov 6-2:03 PM

1. Use examples to describe similarities and/or differences between the two types of data.
 - a) continuous versus discrete
 - b) ordinal versus nominal
 - c) numerical versus categorical

Nov 6-2:04 PM

Extend

8. Your teacher will direct you to a website that shows the popularity of various names.
 - a) Search for your name and describe how its popularity has changed over the years.
 - b) Enter a common name at your school. Is it just as popular on this site? Explain.

Nov 6-2:05 PM