CANADIAN COLLEGE OF TECHNOLOGY AND BUSINESS

2023

Canadian College of Technology and Business

Invest in Yourself

Descriptive Statistics

CANADIAN COLLEGE
OF TECHNOLOGY
AND BUSINESS

Descriptive Statistics

Overview

- Descriptive statistics is a **simple, intuitive way of summarizing and describing** the main features of a dataset to have a quick overview
- Often the first step in data analysis
  - Provide a preliminary understanding of the data
  - Help to identify issues needing to be addressed, like outliers, skewness, or missing values

Descriptive Statistics

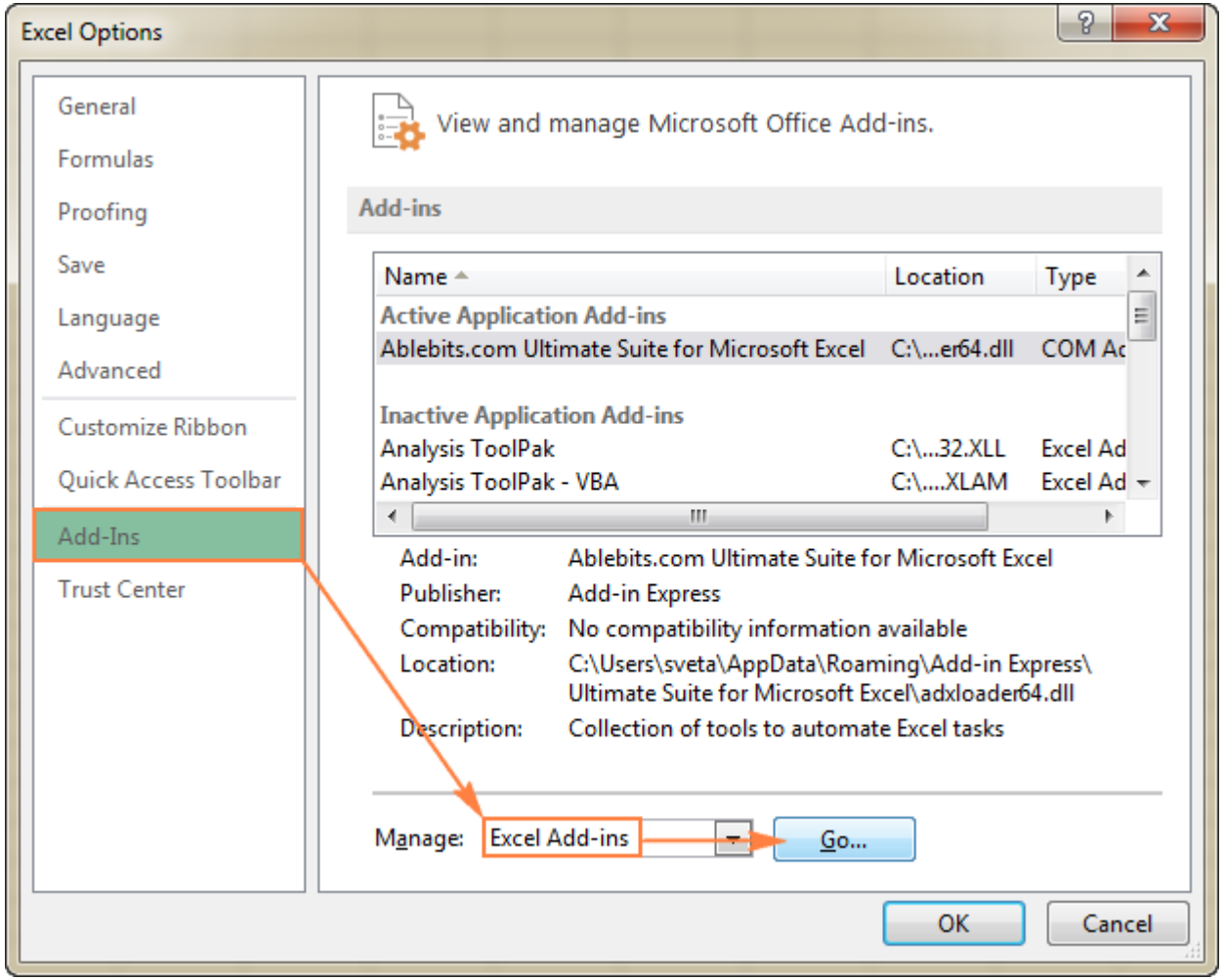
Main Techniques Used in Descriptive Statistics

- **Measures of central tendency:** Include mean, median, and mode
  - Provide a summary of "typical" value in a dataset
- **Measures of variability:** Include the range, variance, and standard deviation
  - Describe how spread out the values in a dataset are
- **Graphical representations:** Use graphs such as histograms, bar charts, and scatter plots
  - Visually display the data to help identify patterns, trends, and relationships
- **Tabular summaries:** Use tables such as frequency tables and contingency tables
  - Summarize the data and to help identify patterns and relationships in the data

Excel's Analysis ToolPak

How to Enable

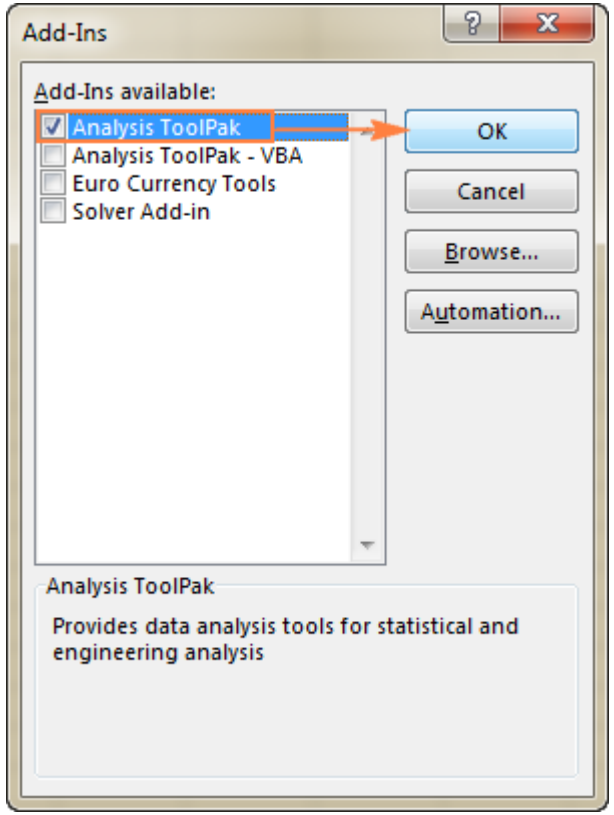
1. Click File > Options. In the Excel Options dialog, click Add-Ins on the left sidebar, select Excel Add-ins in the Manage box, and click the Go button



Excel's Analysis ToolPak

How to Enable

2. In the Add-Ins dialog box, check the Analysis ToolPak box, and click OK to close the dialog



Displaying Descriptive Statistics

CANADIAN COLLEGE
OF TECHNOLOGY
AND BUSINESS

Displaying Qualitative Data

Qualitative Data

- Recall that qualitative data are values that are **categorical**
  - Can be either **nominal** or **ordinal** measurement levels
  - Describe a characteristic of the data, such as gender or level of education

Displaying Qualitative Data

Techniques to Display Qualitative Data

- The following can be used
  - Frequency Tables
  - Bar Charts
  - Pie Charts

Displaying Qualitative Data

Frequency Tables

- A tabular summary of the **number of occurrences** (frequency) of each unique value in a dataset
- Frequency tables can also show the **relative frequencies** of each category
    - The number of times a category occurs divided by the total number of outcomes

Displaying Qualitative Data

Frequency Tables

- An example of a dataset with 16 samples (colors of cars sold per day) for a frequency table

Dataset	
Red	Red
Blue	Green
Red	Blue
Green	Red
Blue	Green
Red	Red
Green	Blue
Blue	Green

Frequency Table

Color	Frequency	Relative Frequency
Red	6	$^{6}/_{16} = 0.3750$
Blue	5	$^{5}/_{16} = 0.3125$
Green	5	$^{5}/_{16} = 0.3125$
Total	**16**	

Displaying Qualitative Data

Frequency Tables

- When qualitative data has an **ordinal scale**, the frequency table can be extended with **cumulative frequency and cumulative relative frequency**
    - Cumulative frequency and cumulative relative frequency are used to determine the frequency and relative frequency of observations that are less than or equal to a specific value in an ordinal scale
    - Calculated by adding the frequencies of each value **up to the desired value**

Displaying Qualitative Data

Frequency Tables

- Example: Consider the dataset which contains the scores of students in a Statistics class

Score	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
A	2	2	$^{2}/_{16} = 0.1250$	0.1250
B	5	2 + 5 = 7	$^{5}/_{16} = 0.3125$	0.4375
C	4	7 + 4 = 11	$^{4}/_{16} = 0.2500$	0.6875
F	5	11 + 5 = 16	$^{5}/_{16} = 0.3125$	1
Total	**16**			

Displaying Qualitative Data

Bar Charts

- Bar chart is a common way to **represent categorical data**
  - Displays the frequencies or relative frequencies of different categories in the form of bars
- Constructing a bar chart
  - The order of the categories is **arbitrary towards nominal scale**
  - For the **ordinal scale**, the categories should be **listed in their natural order**
    - Example: Rating of customer satisfaction (Very Poor, Poor, Fair, Good, Excellent)

Displaying Qualitative Data

Vertical Bar Charts

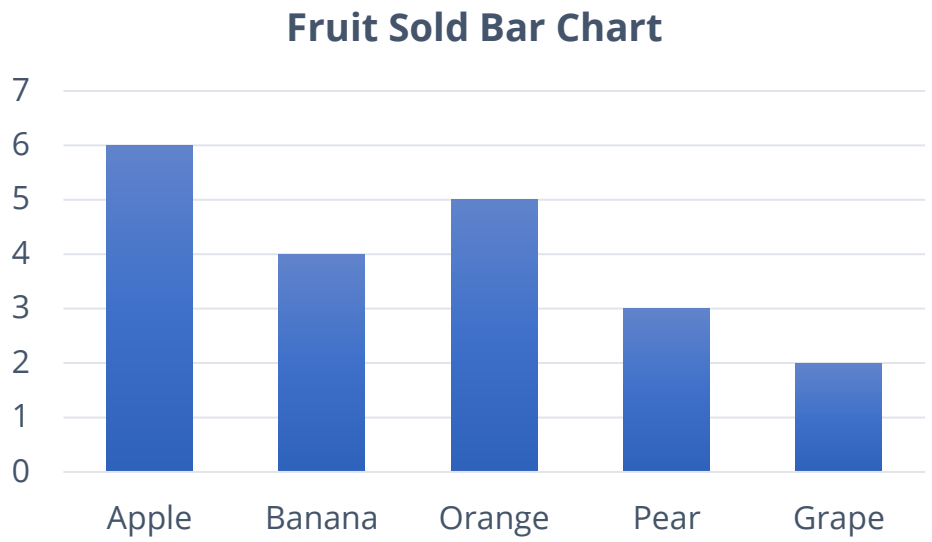
- The x-axis lists the categories
- The y-axis shows the frequency of each category
    - The y-axis can also be the relative frequency for each category
        - The chart will still look the same

Displaying Qualitative Data

Vertical Bar Charts

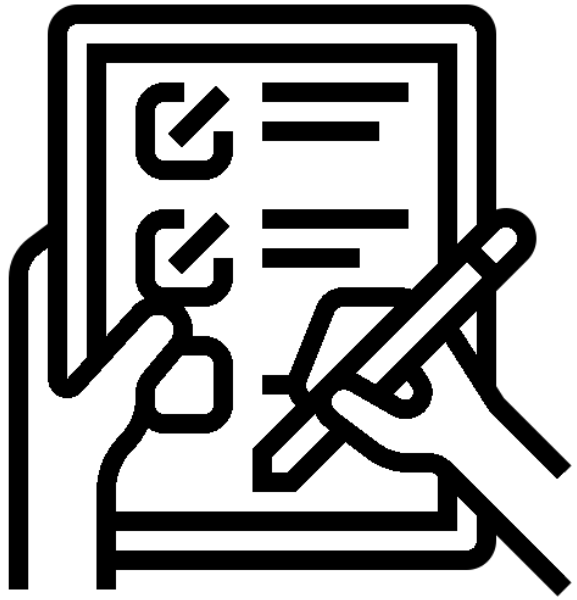
- Vertical bar chart for fruits sold in a supermarket per hour

Fruit	Frequency
Apple	6
Banana	4
Orange	5
Pear	3
Grape	2



Displaying Qualitative Data

Activity 1



Displaying Qualitative Data

Horizontal Bar Charts

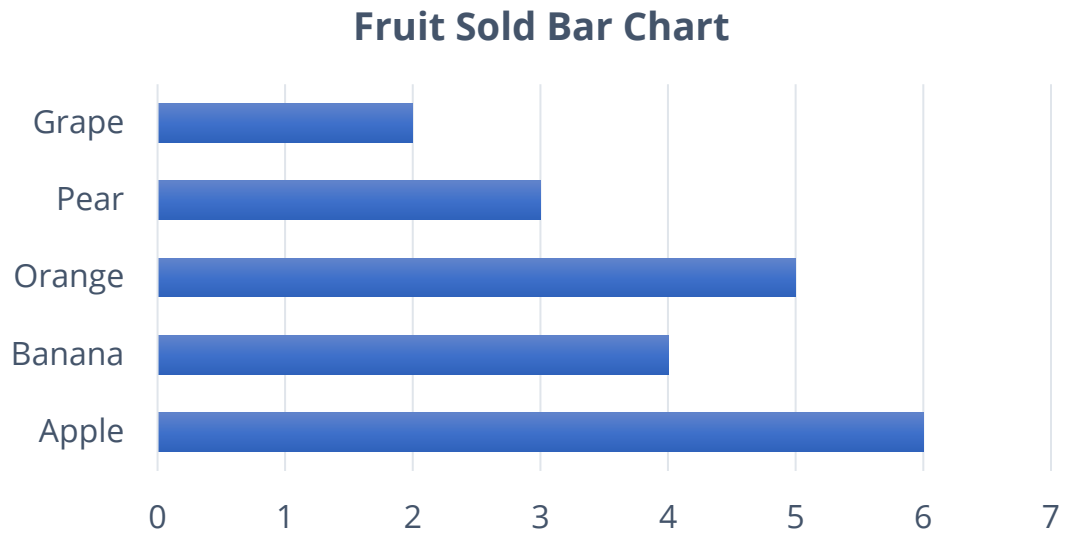
- We could flip the x- and y-axis to generate a horizontal bar chart
  - The information would still be the same, just the orientation of the chart has changed
- The horizontal bar chart is useful when
  - The categories have **long labels**
  - Want to **emphasize** the magnitude of the **frequencies** over the categories themselves

Displaying Qualitative Data

Horizontal Bar Charts

- Horizontal bar chart for fruits sold in a supermarket per hour

Fruit	Frequency
Apple	6
Banana	4
Orange	5
Pear	3
Grape	2



Displaying Qualitative Data

Pie Charts

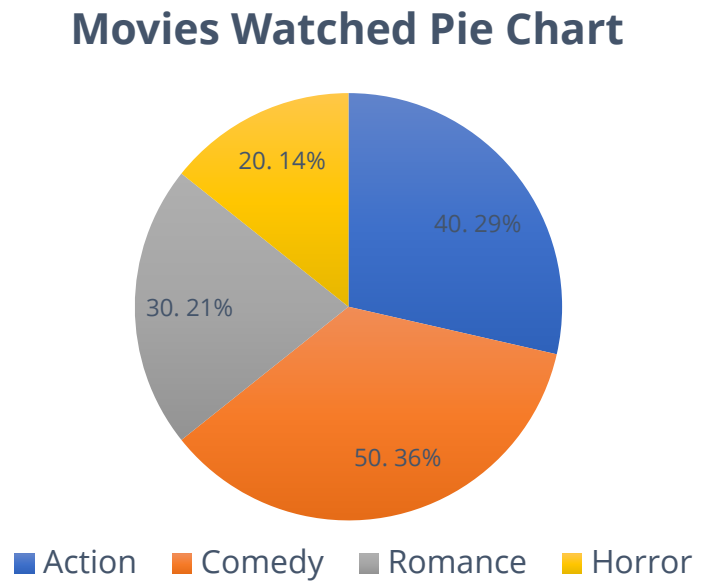
- A pie chart is a circular statistical graphic divided into slices to **represent the proportion** of the whole set of data
    - Each slice in the pie chart represents a category
    - The size of the slice is proportional to the quantity of that category
        - The slices can also have the percentages presented with them

Displaying Qualitative Data

Pie Charts

- Suppose we want to represent the types of movies watched by a group of people in a pie chart

Movie Type	Number of Movies Watched	Proportion of Total Movies Watched
Action	40	$^{40}/_{140}$ = 28.57%
Comedy	50	$^{50}/_{140}$ = 35.71%
Romance	30	$^{30}/_{140}$ = 21.43%
Horror	20	$^{20}/_{140}$ = 14.29%
Total	**140**	



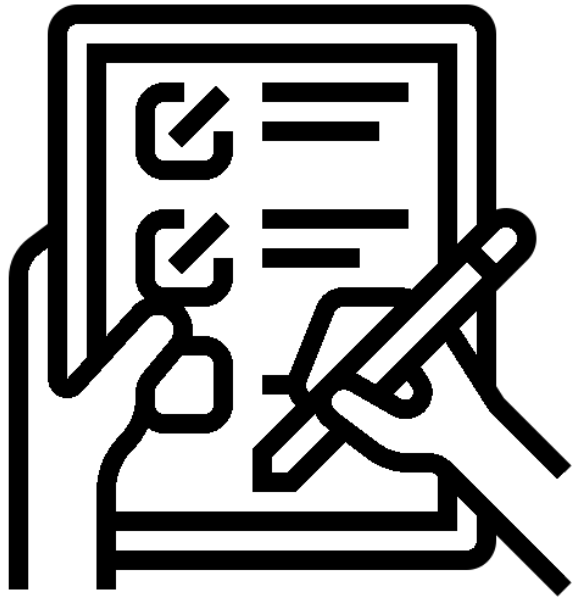
Displaying Quantitative Data

Quantitative Data

- Quantitative data refers to numerical data that can be measured and compared, which can be
  - **Discrete:** Have a **finite or countable** number of values
    - Represent something that has been counted
    - Example: The number of siblings a person or the number of siblings a person
  - **Continuous:** Have an **infinite number of values** within a given range
    - Usually measured rather than counted
    - Example: The height of a person or the temperature of a room

Displaying Quantitative Data

Activity 2



Displaying Quantitative Data

Techniques to Display Quantitative Data

- The following can be used
  - Frequency Tables
  - Histogram
  - Stem-and-Leaf Plots

Displaying Quantitative Data

Frequency Tables

- Like qualitative data, quantitative data can be displayed using a frequency table
- Suppose we have a set of data that represents the number of pets owned by 10 people

2	0	3	1	2	1	0	3	2	1

Number of Pets	Frequency	Relative Frequency	Cumulative Relative Frequency
0	2	0.2	0.2
1	3	0.3	0.5
2	3	0.3	0.8
3	2	0.2	1

Displaying Quantitative Data

Constructing a Frequency Distribution Using Grouped Quantitative Data

- To construct a frequency distribution for grouped quantitative data, we need to **first group the data into intervals or classes**
    - For certain data sets, especially those with continuous values, it is necessary to group several values into a single class to be summarized and represented effectively
- This **grouping process helps to avoid having too many classes** in the frequency distribution, which can make it challenging to identify patterns and trends in the data

Displaying Quantitative Data

Steps to Construct a Frequency Distribution Using Grouped Quantitative Data

1. Determine the range of the data
    - Find the **minimum and maximum values** of the dataset
    - Subtract the minimum value from the maximum value to determine the range
2. Determine the number of classes
    - A method for determining the number of classes in a frequency distribution is the rule

$$2^k \geq n$$

where: k = Number of classes

n = Number of data points

    - Example: Given n = 60, we have $2^6 = 64 > 60 \Rightarrow$ k = 6 classes

Displaying Quantitative Data

Steps to Construct a Frequency Distribution Using Grouped Quantitative Data

3. Find the class width

- The class width refers to the **range of values assigned to each class**

$$\textbf{class width} = \frac{\textbf{Data range}}{\textbf{k}}$$

- The width should be rounded to a whole number to make frequency distribution readable
- There is no single, definitive answer for determining the class width

Displaying Quantitative Data

Steps to Construct a Frequency Distribution Using Grouped Quantitative Data

4. Determine the class limits
    - The class limits define the **beginning and end of each class interval**
    - The lower-class and upper-class limit are the smallest and largest values in the interval
5. Count the number of observations that fall within each class interval

Displaying Quantitative Data

Constructing a Frequency Distribution Using Grouped Quantitative Data

- Suppose we have a set of data representing the heights (in inches) of 60 people

63	68	66	67	69	61	61	63	67	68
71	65	71	64	60	71	63	65	60	61
61	69	64	68	61	62	68	64	71	69
62	71	68	63	68	68	67	65	63	69
61	68	71	65	67	68	60	71	65	67
62	68	69	63	68	65	63	67	68	68

Displaying Quantitative Data

Constructing a Frequency Distribution Using Grouped Quantitative Data

- Step 1: Data range = 71 – 60 = 11
- Step 2: Given n = 60, we have $2^6 = 64 > 60 \Rightarrow$ k = 6 classes
- Step 3: Class width = $\frac{\text{Data range}}{k} = \frac{11}{6} \approx 2$
- Step 4: Class limits
    - Class 1: 60 to less than 62
    - Class 2: 62 to less than 64
    - Class 3: 64 to less than 66
    - Class 4: 66 to less than 68
    - Class 5: 68 to less than 70
    - Class 6: 70 to less than 72

Displaying Quantitative Data

Constructing a Frequency Distribution Using Grouped Quantitative Data

- Step 5: Construct the frequency table

Class Label	Frequency	Relative Frequency
60 to less than 62	9	0.15
62 to less than 64	10	0.17
64 to less than 66	9	0.15
66 to less than 68	7	0.12
68 to less than 70	18	0.30
70 to less than 72	7	0.12

Displaying Quantitative Data

Rules for Classes for Grouped Data

1. **Equal class widths:** Classes should have the same range of values
    - Easier to compare their frequencies
2. **Non-overlapping classes:** Classes should not overlap with each other
3. **Include all data values:** All data points should be accounted for in the frequency distribution
4. **Avoid empty classes:** It's not ideal to have classes with no data points
5. **Avoid open-ended classes:** Classes with no defined upper or lower boundary
    - Violate the rule of equal class widths
6. **Choose an appropriate number of classes:** The number of classes should be chosen carefully to ensure that the frequency distribution is easy to read and effectively represents the data

Displaying Quantitative Data

Histogram

- A graph that **displays the distribution of a dataset** by dividing the data into a set of "bins" and counting the number of observations that fall into each bin
  - The resulting histogram shows the shape of the data by displaying the frequency of values within each bin as a bar
  - The x-axis represents the values in the data set
  - The y-axis represents the frequency of occurrence of those values

Displaying Quantitative Data

Histogram

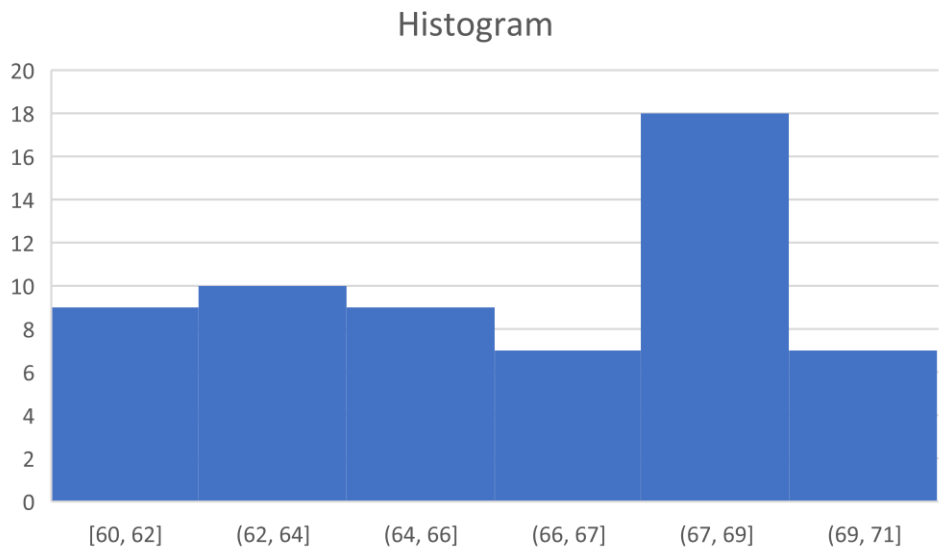
- A histogram resembles a bar chart, but while a **bar chart is used to display categorical data**, a **histogram is used to display numerical data**
    - The bars in a histogram are only separated by gaps if a class in the data set has no values

Displaying Quantitative Data

Histogram

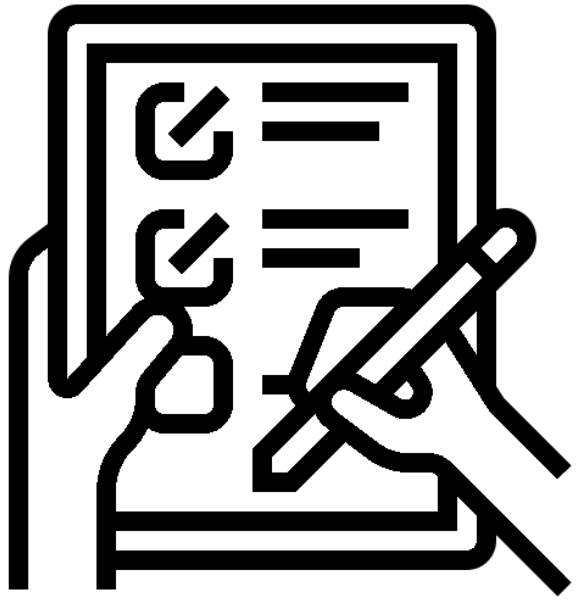
- Consider the set of data representing the heights (in inches) of 60 people

Class Label	Frequency	Relative Frequency
60 to less than 62	9	0.15
62 to less than 64	10	0.17
64 to less than 66	9	0.15
66 to less than 68	7	0.12
68 to less than 70	18	0.30
70 to less than 72	7	0.12



Displaying Quantitative Data

Activity 3



Displaying Quantitative Data

Stem-and-Leaf Plots

- A type of data visualization to **display the distribution of a dataset**
- It consists of two columns
  - One for the "stems" – The **leading digit or digits** of each value in the data set
  - One for the "leaves" – The **trailing digits** of each value in the data set

Displaying Quantitative Data

Stem-and-Leaf Plots

- To create a stem-and-leaf plot for this data, we would:
  - Sort the data from lowest to highest
  - Divide each value in the dataset into "stem" (the first digit) and "leaf" (the remaining digits)
  - Arrange the stems in order and list the corresponding leaves under each stem

Displaying Quantitative Data

Stem-and-Leaf Plots

- Suppose we have a data set of exam scores for a class of 20 students ordered

  78, 80, 81, 82, 83, 84, 85, 86, 87, 88, 88, 89, 89, 90, 90, 91, 91, 92, 93, 94

- The resulting stem-and-leaf plot

Stem	Leaf
7	8
8	0 1 2 3 4 5 6 7 8 8 9 9
9	0 0 1 1 2 3 4

- This stem-and-leaf plot shows that most of the exam scores in the class fall between 80 and 90, with a peak in frequency at 88-90

Any Questions?

CANADIAN COLLEGE OF TECHNOLOGY AND BUSINESS

Thank You

CANADIAN COLLEGE OF TECHNOLOGY AND BUSINESS

Website www.canadianctb.ca

Email info@canadianctb.ca

Telephone +1 604-515-7880

Address 626 West Pender Street - Suite 600
Vancouver, British Columbia, V6B 1V9, Canada

Connect with CCTB

@CanadianCTB

DLI O134304821852