First Name: _____ Last Name: _____ Student ID: _____

Chapter 5: Data Visualization, Sampling Techniques, Bias in Surveys

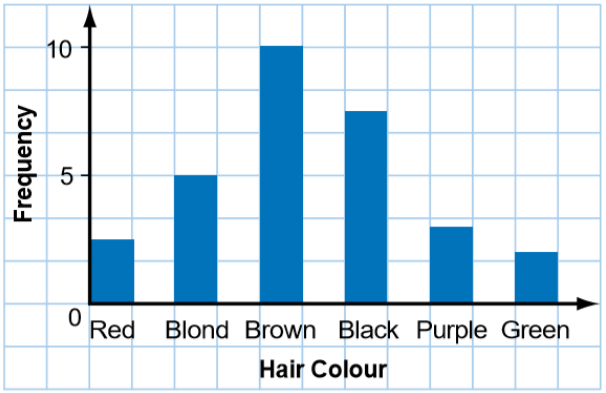
➢ Definitions

To begin this chapter, we first look at some statistical language that we need to understand before we continue.

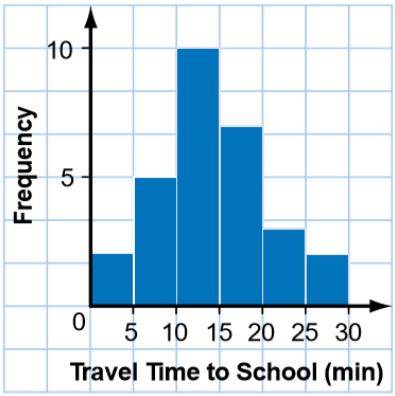
- Statistics
- Raw Data
- Continuous Variable
- Discrete Variable
- Categorical Data
- Range

➢ Common Types of Data Visualization

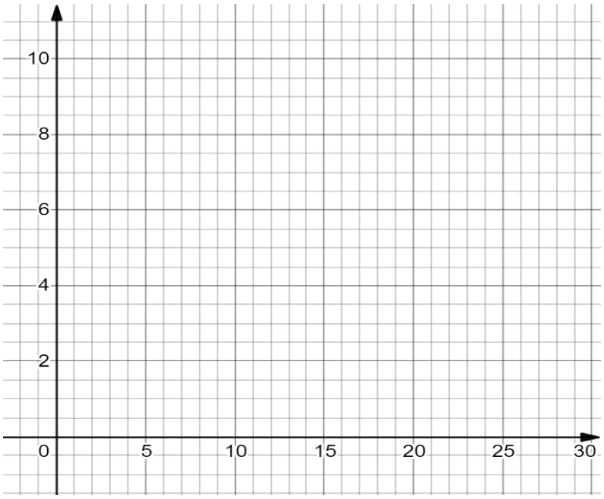
Bar Graph



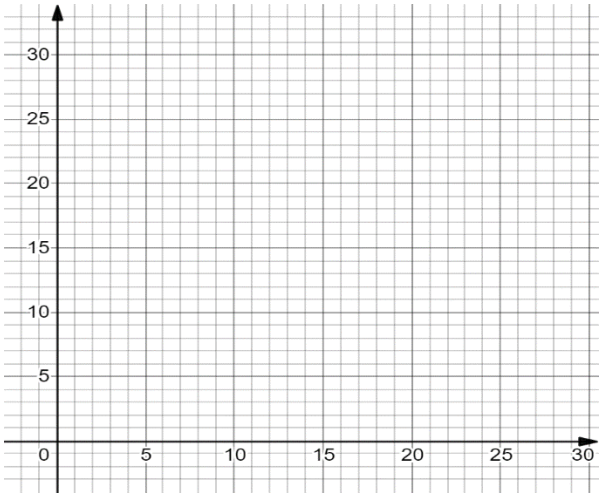
Histogram



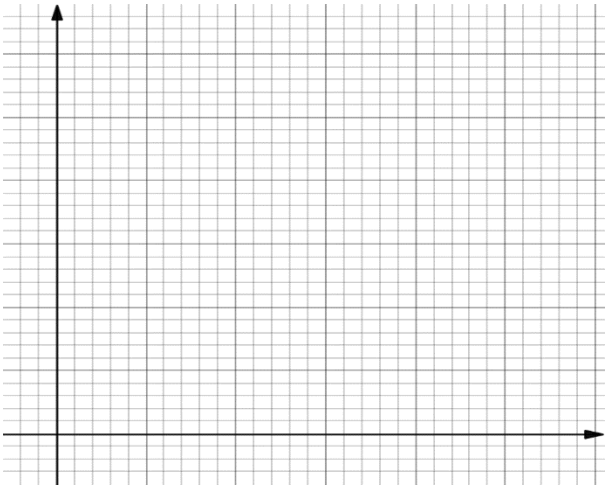
Frequency Polygon



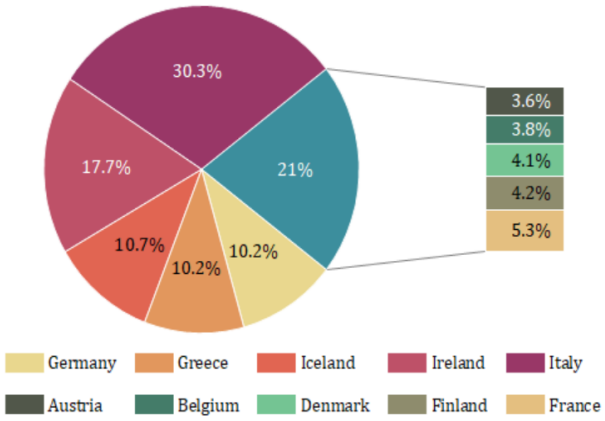
Cumulative-frequency-Polygon



Relative Frequency Histogram/Polygon



Pie Chart



➢ Examining Data and Constructing Graphs

When presented with a set of data, there are a few things we should consider before proceeding with data visualization or analysis.

- The nature of the data: is it continuous or discrete?
- Is it numerical or categorical?
- Should we use a grouped or ungrouped format to display the data?

Ex. 1: Consider the following list of daily high temperatures in July for Waterloo, ON.

July	1	2	3	4	5	6	7	8	9	10
Temp (°C)	27	25	24	30	32	31	29	24	22	19

July	11	12	13	14	15	16	17	18	19	20
Temp (°C)	21	25	26	31	33	33	30	29	27	28

July	21	22	23	24	25	26	27	28	29	30	31
Temp (°C)	26	27	22	18	20	25	26	29	32	31	28

a) Is temperature a discrete or continuous variable? Is it numerical or categorical? Explain.

b) Should this data be presented in a *grouped* or *ungrouped* format?

c) Determine a range and suitable interval width for the frequency table.

d) Complete the following frequency table.

Temperature Intervals	**Midpoint**	**Tally**	**Frequency**	**Cumulative Frequency**	**Relative Frequency**

e) Construct a frequency histogram, a relative frequency histogram and a cumulative frequency polygon.

f) On how many days was the maximum temperature 25°C or less? On how many days was the temperature greater than 30°C?

- **Sampling Techniques**

When you conduct research about a group of people, it is rarely possible to collect data from every person in that group. Instead, you select a sample. To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. There are two types of sampling methods:

- **Probability sampling**:

- **Non-probability sampling**:

First, we need to understand the difference between a **population** and a **sample**, and identify the target population of your research.

- **Population:**

- **Sample:**

If the population is very large, demographically mixed, and geographically dispersed, it might be difficult to gain access to a representative sample.

- **Sampling frame:**
- **Sample Size:**

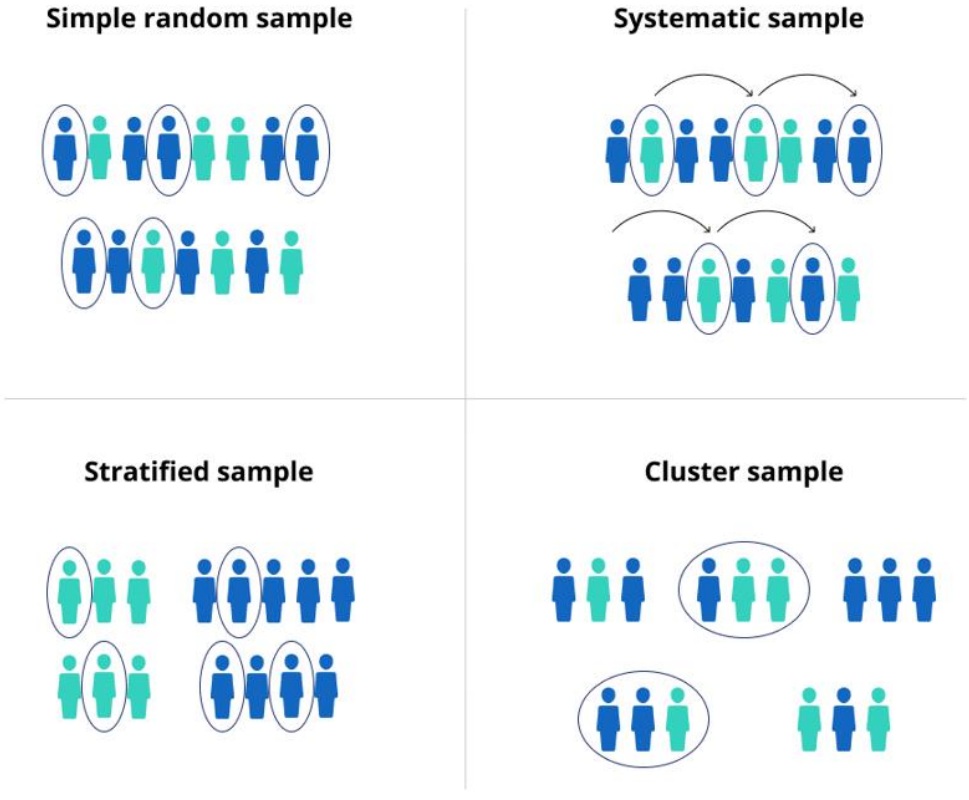
Ex.2: You are doing research on working conditions at Company X with 1000 employees.

a) Identify the following:

Population:

Sampling frame:

There are four main types of **probability sampling**:



Ex.2 cont'd:

b) Suppose you will select a sample of 100 employees of Company X for your research. Describe how you can select the sample using each of the four probability sampling methods.

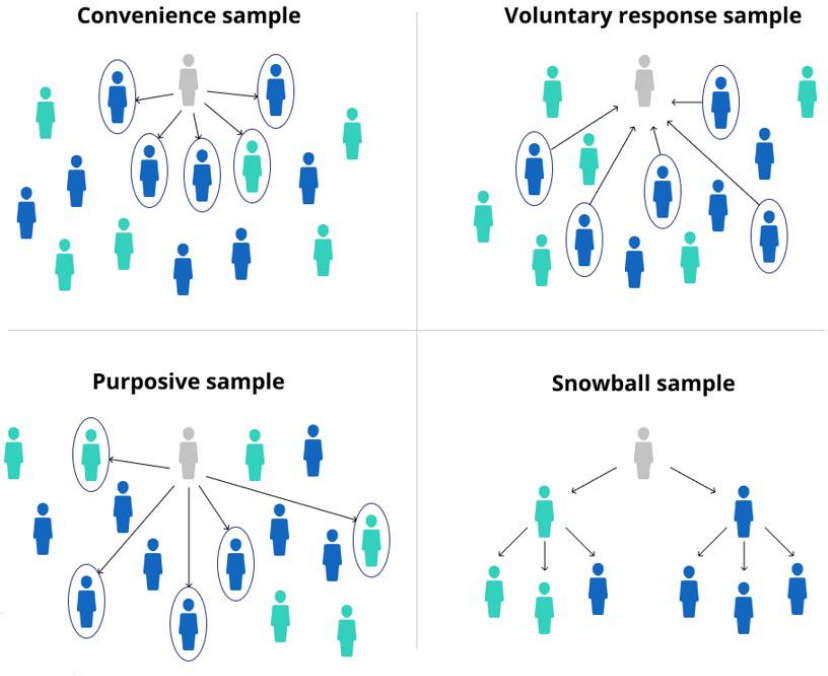
i) simple random sampling

ii) systematic sampling

iii) stratified sampling

iv) cluster sampling

Non-probability sampling techniques are often appropriate for **exploratory** and **qualitative** research. In these types of research, the aim is **not** to test a hypothesis or make inferences about a broad population, but to develop an initial understanding of a small or under-researched population.



- **Bias in Survey**

Statistical Bias is any factor that favours certain outcomes on responses, hence systematically skewing the survey results. Often bias is unintentional, but there are times when some people deliberately bias surveys in order to get the results they want.

Sampling Bias: This type of bias is why we need to carefully choose a sampling technique! If you are looking to draw conclusions about a certain population, you'll need to ensure that everyone in that group has an equal chance of receiving an invite to respond to your survey.
Ex.3: A telecommunication company is looking to find out how satisfied are their customers with its service. The company shares a link to the survey via social media.

Non-Response Bias: Regardless of how careful you are about inviting everyone to respond, there will be some people who are either unwilling or unable to respond. These people are often systematically different than those who do respond.
Ex.4: A science class asks every fifth student entering the cafeteria to answer a survey on environmental issues. Less than half agree to complete the questionnaire. The completed questionnaires show that a high proportion of the respondents are concerned about the environment and are well informed about environmental issues.

Measurement Bias: This type of bias occurs when the data-collection method consistently either under- or overestimate a characteristic of the population.
Ex.3: A highway engineer suggests that an economical way to survey traffic speeds on an expressway would be to have police officers patrol the highway and record the speed of the traffic around them every half hour. How is this sample biased?

Response Bias: Humans are complex creatures. Getting people to respond to your survey is hard enough, but you also need to think about how to get them to respond accurately. Both subconscious and conscious cognitive factors can result in less-than-truthful responses.
Ex.4: A high school wanted to know what percent of its students smoke cigarettes. During the week when students visited the counselors to schedule classes, they asked every student in person if they smoked cigarettes or not. The data showed that 5% of students smoked cigarettes.

- Leading questions:

- Loaded questions: