

CANADIAN COLLEGE  
OF TECHNOLOGY  
AND BUSINESS

2023



**Canadian College of  
Technology and Business**  
Invest in Yourself



# Introduction to Statistics



CANADIAN COLLEGE  
OF TECHNOLOGY  
AND BUSINESS



# Introduction to Statistics



## Overview

- Data is **everywhere**
- Statistics is a science dealing with **collection, analysis, interpretation, and presentation of data**
  - It is important that the statistical reports are accurate and unbiased

# Introduction to Statistics



## Application of Statistics

- **Advertising**
  - Evaluate post-campaign to determine the effectiveness of advertising strategies
- **Operations**
  - Investigate existing and emerging trends
- **Biology and medicine**
  - Study collected data to uncover reasons for the occurrence of diseases
- **Finance and economics**
  - Track a firm's performance and assess economic growth

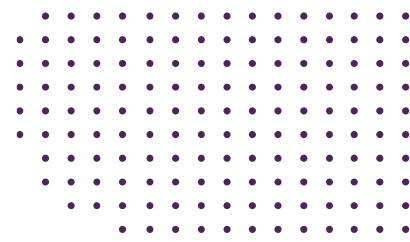
# Introduction to Statistics



## Application of Statistics

- **Agriculture**
  - Forecast agricultural output and identify variables that influence crop growth
- **Marketing research**
  - Conduct and analyze consumer surveys to determine needs and wants
- **Weather**
  - Forecast changes and fluctuations in climate conditions

# Introduction to Statistics



## Categories of Statistics

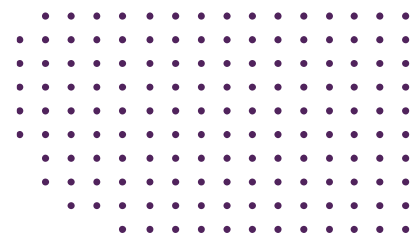
- **Descriptive statistics**

- Deal with **summarizing and describing** the main features of a set of data
- Involve using **measures**, such as **mean**, **median**, **mode**, to describe the **distribution of data** and its central tendencies

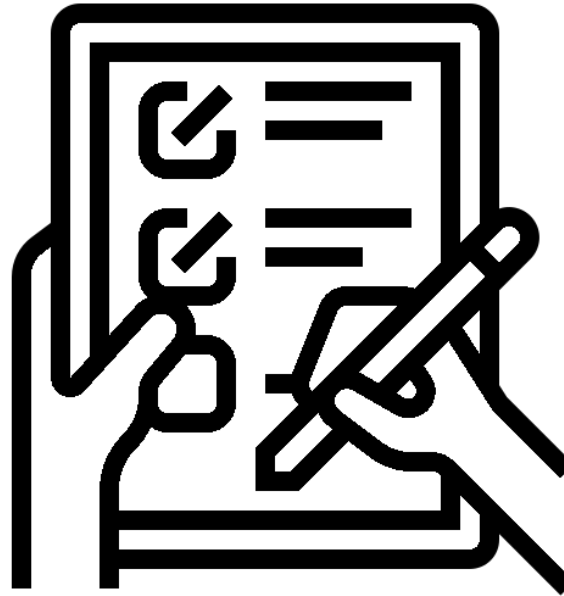
- **Inferential statistics**

- Use sample data to **make inferences** about a larger population
- Involve using **statistical techniques to make predictions**, **estimate population parameters**, and **test hypotheses about relationships between variables**

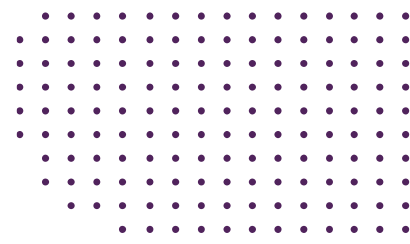
# Introduction to Statistics



## Activity 1

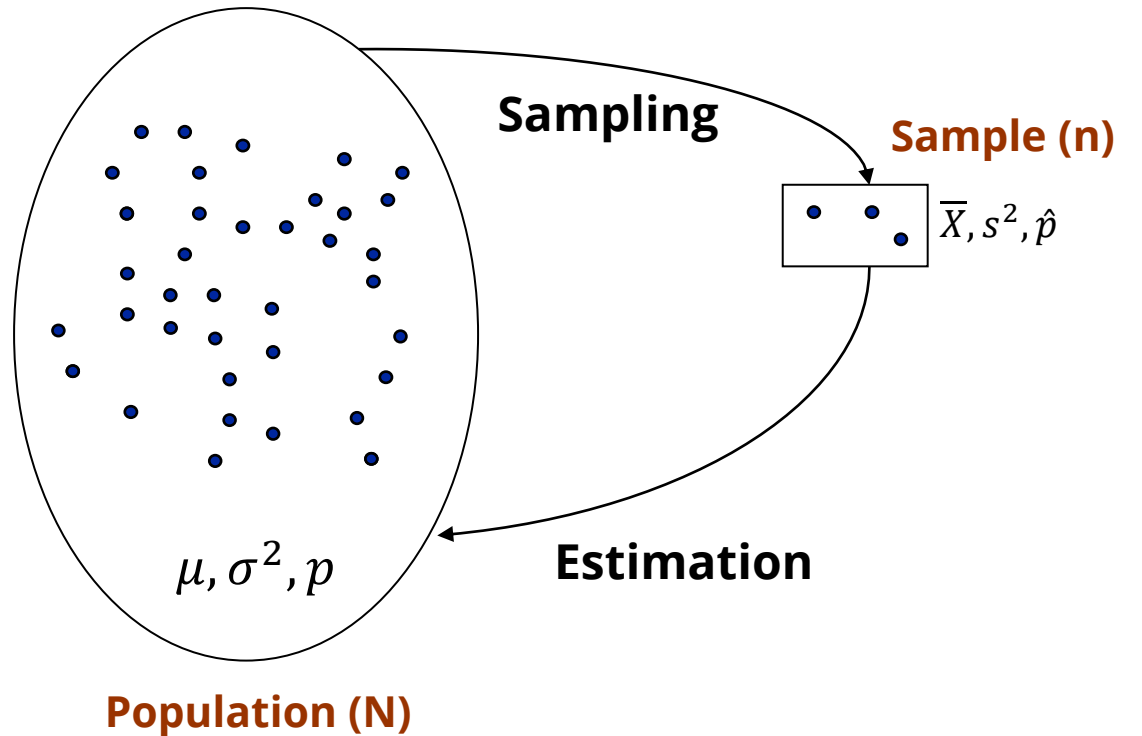


# Introduction to Statistics



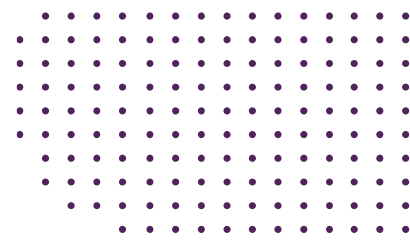
## Population vs. Sample

- The entire set of data being studied is referred to as the **population**
- A portion of this data, chosen for analysis, is referred to as a **sample**





# Introduction to Statistics

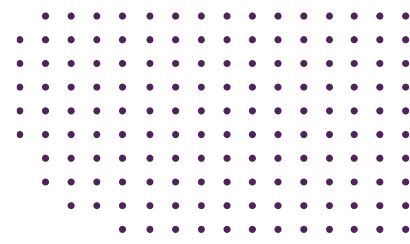


## Population vs. Sample

- If the population is **small enough** and it is **possible to gather data on every member** of the population, this method of data collection is referred to as a **census**
  - **Why Sample?**
    - Study a population may be
      - Impossible
      - Impractical
      - Too costly
- Sample may be used instead

# Introduction to Statistics

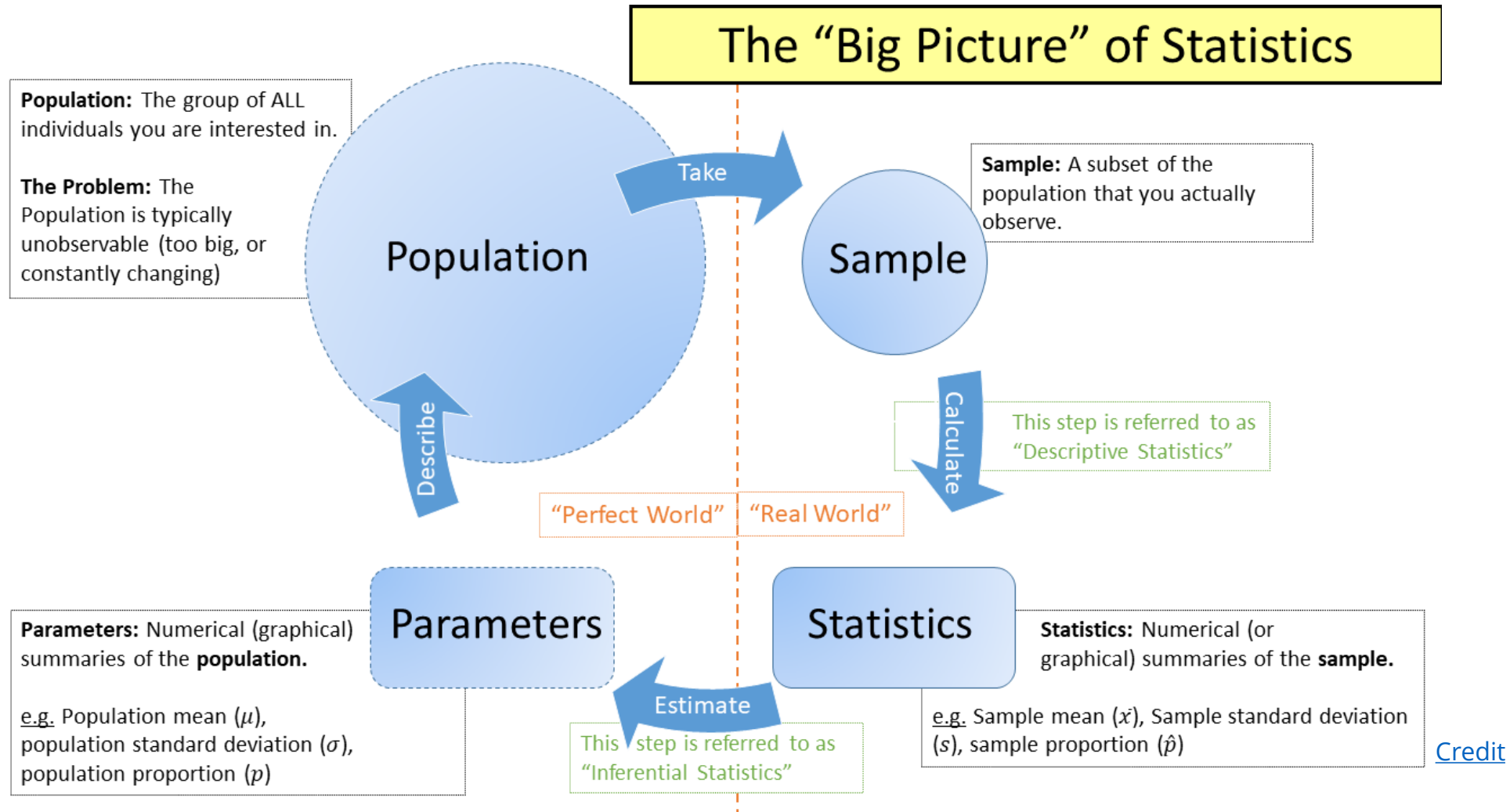
---



## Parameter vs. Statistic

- A **parameter** is a numerical value that describes a **population** characteristic
  - A fixed value that is true for the entire population
- A **statistic** is a numerical value that summarizes a characteristic of a **sample**
  - An estimate of a population parameter based on a sample of data
  - Since statistics are based on a sample, they will vary from sample to sample, and are subject to sampling error

# Introduction to Statistics



# Sampling



## Simple Random Sample

- A subset of a population that **each member** of the population has an **equal chance to be selected**
  - Help to minimize bias
  - Ensure the sample represents the population – more accurate inferences and predictions

# Sampling



## Biased Sample

- A type of sample that does **not accurately** represent the population from which it was drawn
- This occurs when the selection process is **not random and systematic**
  - Some members of the population have a **higher/lower chance of being selected** than others
- A biased sample can lead to **inaccurate conclusions and inferences** about the population

# Sampling

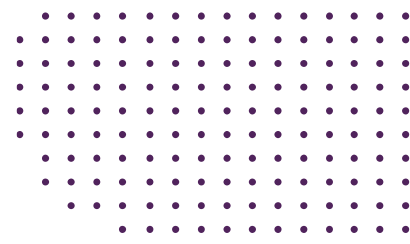


## Reasons of Biased Sample

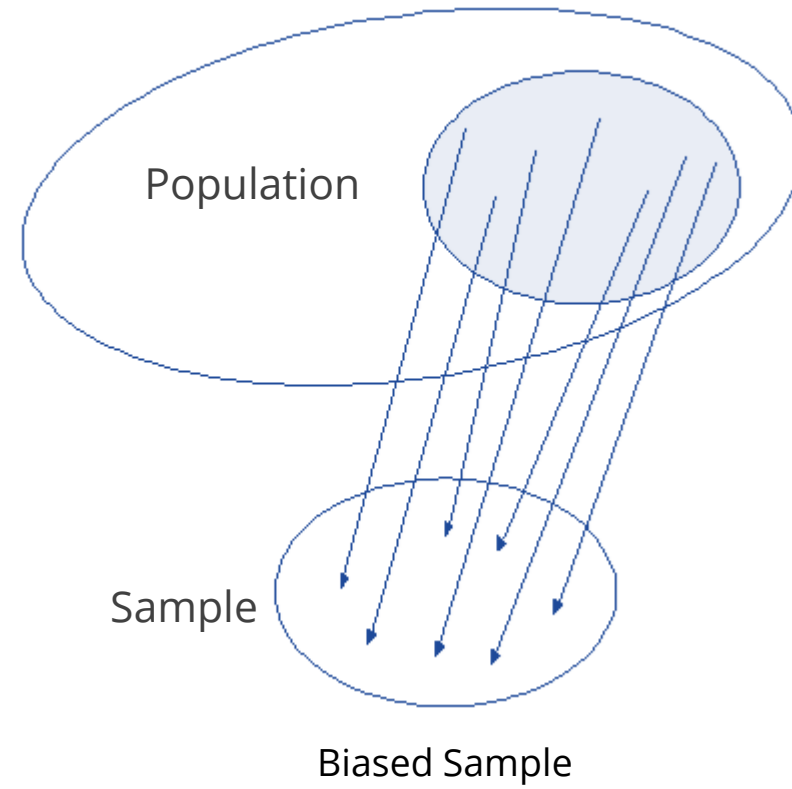
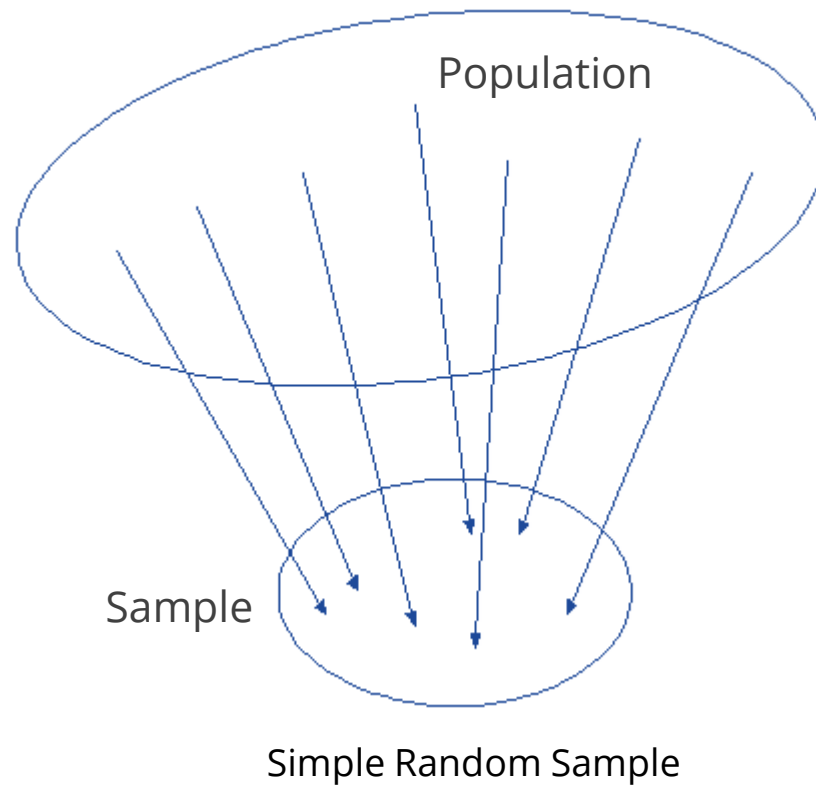
- **Convenience sampling**
  - Select individuals who are readily available or easy to access
- **Self-selection sampling**
  - Allow individuals to volunteer to be part of the sample
- **Voluntary response sampling**
  - Ask individuals to respond to a survey voluntarily
- **Non-response bias**
  - When certain individuals do not respond to the survey



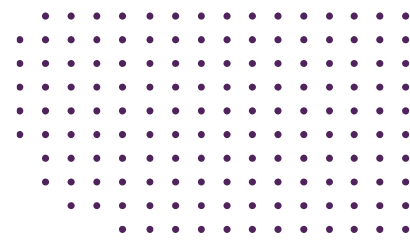
# Sampling



## Simple Random Sample vs. Biased Sample



# Data



## What is Data?

- Data refers to a set of measurements collected from a population to investigate some certain characteristics, also known as observations
  - Database is a collection of data points organized into rows(records) and columns(fields)

First Name	Last Name	Date Of Birth	Department	Email	Salary
John	Rodriguez	1980-06-15	Marketing	john123@gmail.com	\$50,000
Jane	Kim	1990-12-23	HR	janekim456@yahoo.com	\$75,000
Johnson	Patel	1985-03-07	IT	pateljohnson789@hotmail.com	\$45,000

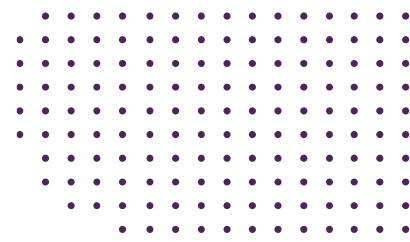
# The Sources of Data



## Primary Data

- Collected **directly** from original sources for a specific purpose
  - Often collected through methods such as surveys, interviews, or experiments
  - The most accurate and relevant type of data for a specific research or analysis
- Disadvantages
  - Can be expensive and time consuming to gather

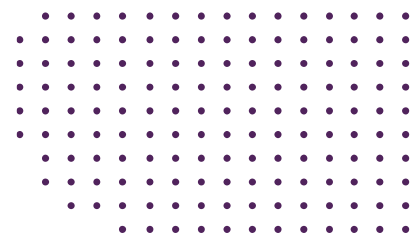
# The Sources of Data



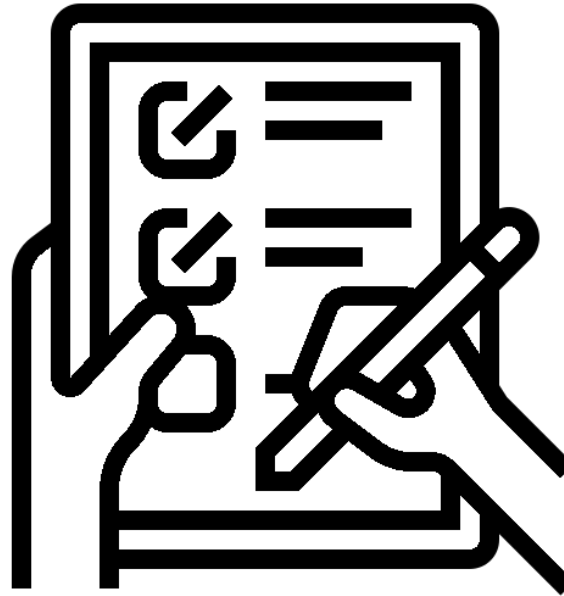
## Secondary Data

- **Previously collected by someone else** and is used for another purpose
  - Readily available, usually less expensive and time-consuming to obtain than primary data
  - Can come from sources like government reports, industry publications, and online databases
  - May require further verification or analysis to ensure its accuracy and relevance
- Disadvantages
  - Data quality may suffer if not gathered and recorded properly
  - No control over the data collecting process

# The Sources of Data

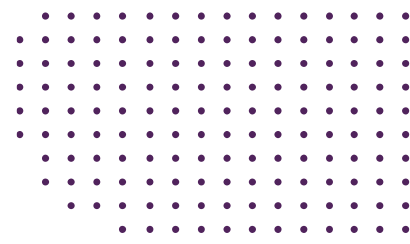


## Activity 2



# Main Types of Data

---



## Qualitative Data

- Use descriptive terms to **classify and label** something of interest
- Examples
  - Gender (male, female, etc.)
  - Marital status (married, single, divorced, etc.)
  - Education level (high school, bachelor's, master's, etc.)
  - Occupation (teacher, doctor, engineer, etc.)
- Mathematical operations, like addition, subtraction, multiplication, and division, cannot be applied to qualitative data



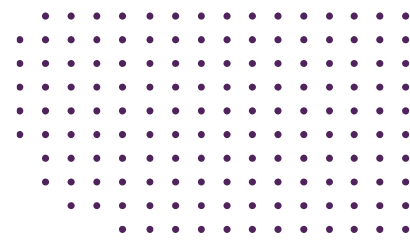
# Main Types of Data



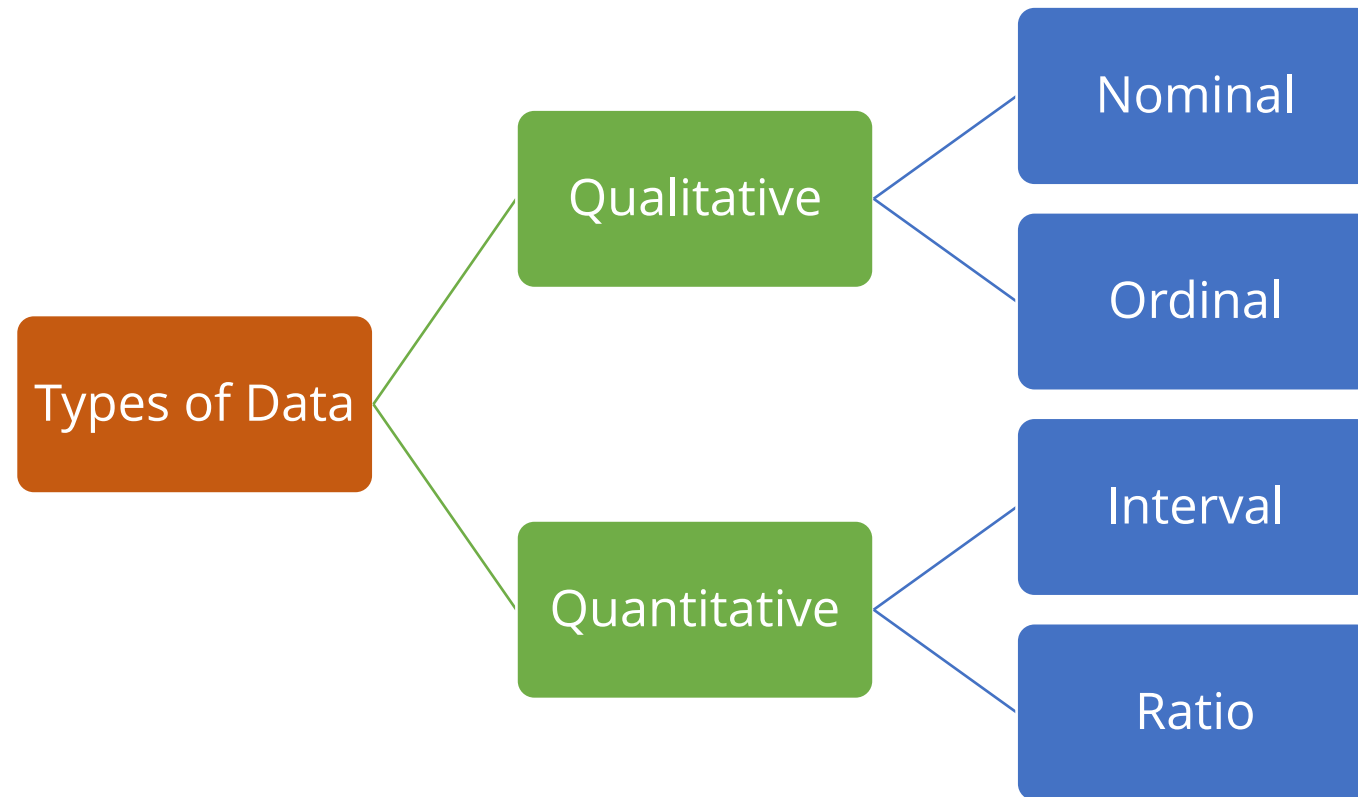
## Quantitative Data

- Use numerical values obtained by a **measurement or count** to describe something of interest
- Examples
  - Age
  - Height
  - Weight
- Note that quantitative data can either be **continuous** (value within a range, such as height) or **discrete** (only take specific values, such as number of siblings)

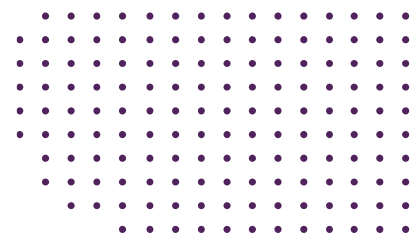
# Main Types of Data



## Scales of Measurement



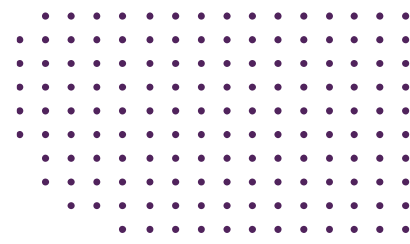
# Main Types of Data



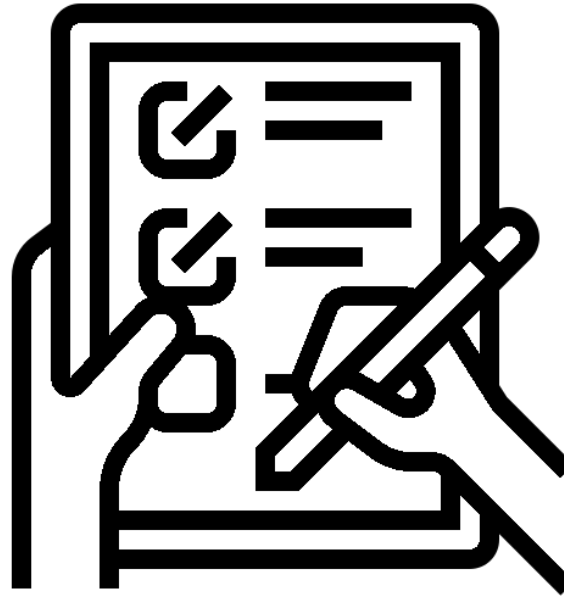
## Scales of Measurement

Data Type	Level	Description	Example
Qualitative	Nominal	<ul style="list-style-type: none"><li>• Lowest level of measurement</li><li>• Categorize data into distinct groups</li><li>• Do not have any inherent order or structure</li></ul>	<ul style="list-style-type: none"><li>• Gender (male, female)</li><li>• Eye color (brown, blue, green)</li></ul>
	Ordinal	<ul style="list-style-type: none"><li>• Builds upon nominal by adding a level of order/ranking</li><li>• Can be ranked or ordered<ul style="list-style-type: none"><li>• The difference between categories is not meaningful</li></ul></li></ul>	<ul style="list-style-type: none"><li>• School grades (A, B, C)</li><li>• Income brackets (low, middle, high)</li></ul>
Quantitative	Interval	<ul style="list-style-type: none"><li>• There is order and difference between two values is meaningful</li><li>• Defined zero point<ul style="list-style-type: none"><li>• Zero point doesn't represent the absence of the quantity measured</li></ul></li></ul>	<ul style="list-style-type: none"><li>• Calendar year (2018, 2019)</li></ul>
	Ratio	<ul style="list-style-type: none"><li>• Includes all the characteristics of interval measurement, but also has a meaningful zero point</li><li>• The difference between two values can be expressed as a ratio</li></ul>	<ul style="list-style-type: none"><li>• Weight (5kg, 15kg)</li><li>• Distance (20km, 50km)</li></ul>

# Main Types of Data



## Activity 3



# Time Series vs. Cross-Sectional Data

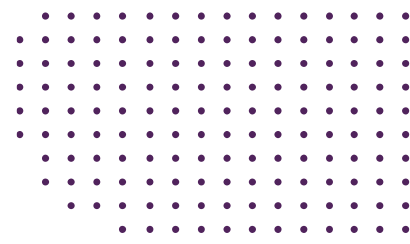


## Time Series Data

- Set of data collected **over time**
- Used to observe trends, patterns, and changes in the data
- Often used in fields such as economics, finance, and environmental studies to track changes in variables such as GDP, stock prices, and temperature over time

# Time Series vs. Cross-Sectional Data

---

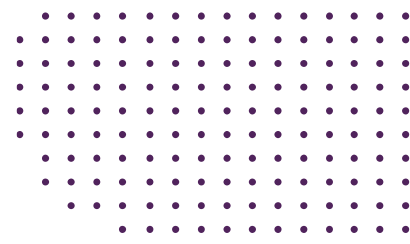


## Cross-Sectional Data

- Collected at a **single point in time** from a sample of individuals, firms, or other units
- Used to make comparisons across units at a specific point in time
- Often used in fields such as sociology, psychology, and marketing to compare characteristics such as income, education, or brand loyalty across individuals or groups



# Time Series vs. Cross-Sectional Data



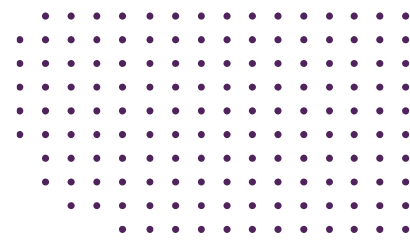
- In summary, time series data tracks changes in a single variable over time, while cross-sectional data compares variables across units at a specific point in time

	Unemployment Rate		
Year	Ontario	Quebec	British Columbia
2015	7.0%	6.5%	6.0%
2016	6.6%	6.2%	5.7%
2017	6.2%	6.0%	5.5%
2018	5.9%	5.8%	5.3%
2019	5.7%	5.6%	5.1%

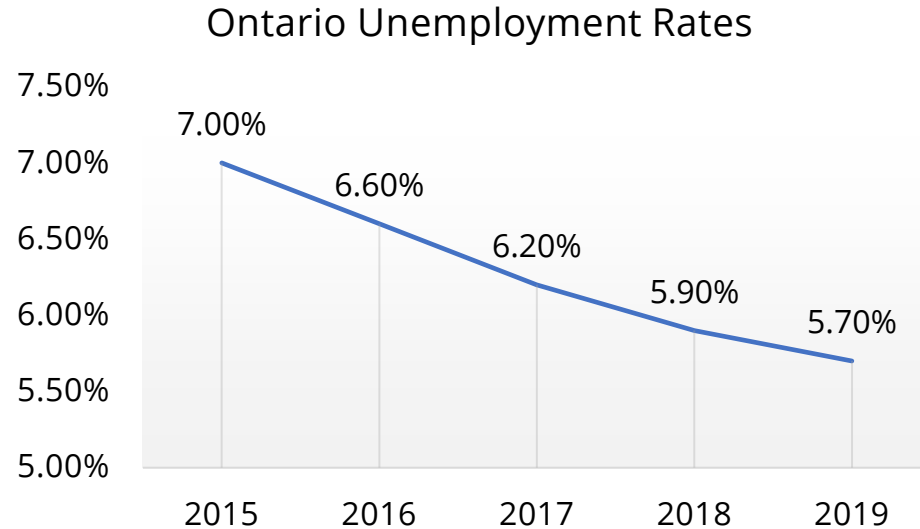
Cross-Sectional Data

Time Series Data

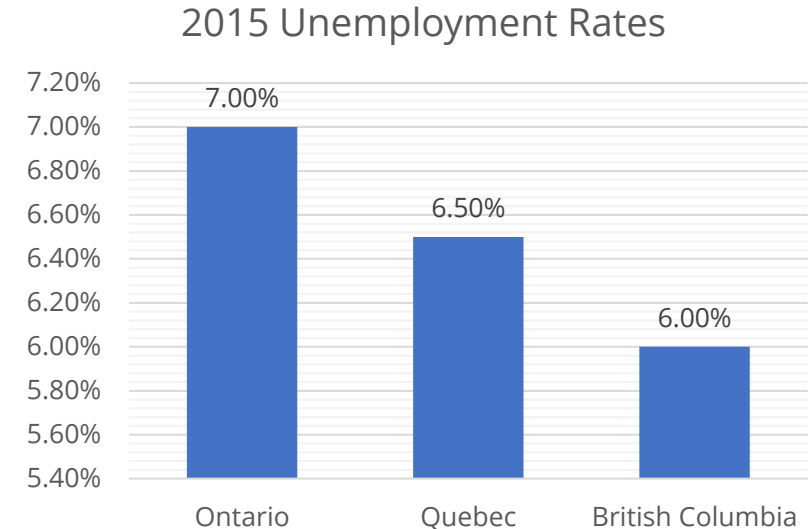
# Time Series vs. Cross-Sectional Data



## Visualization



Time Series Data



Cross-Sectional Data

# Any Questions?

---



CANADIAN COLLEGE  
OF TECHNOLOGY  
AND BUSINESS

*Thank  
You*



**CANADIAN COLLEGE  
OF TECHNOLOGY  
AND BUSINESS**

**Website**

[www.canadianctb.ca](http://www.canadianctb.ca)

**Email**

[info@canadianctb.ca](mailto: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