# Sampling Methods

# Learning Outcomes

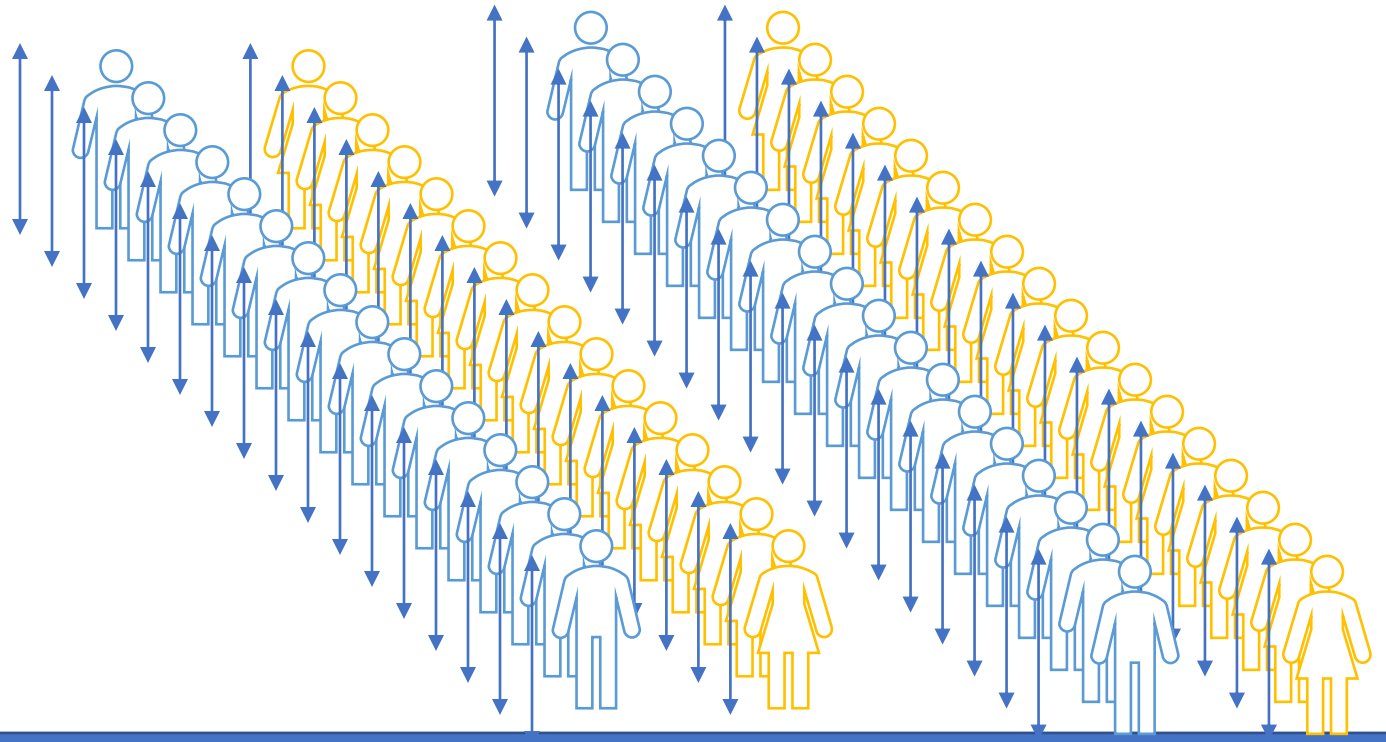
- By the end of this lecture, students will:
  - Value the concept of sampling and understand its rational.
  - Understand simple random, stratified, clustering, systematic random, and convenience sampling.
  - Be able to choose a suitable sampling method for a given problem.
  - Understand the difference between sampling and non-sampling errors.

# Outline

- Introduction
- Sampling Techniques
  - Simple Random Sampling
  - Convenience Sampling
  - Systematic Sampling
  - Cluster Sampling
  - Stratified Sampling
- Sources of Bias
- Summary

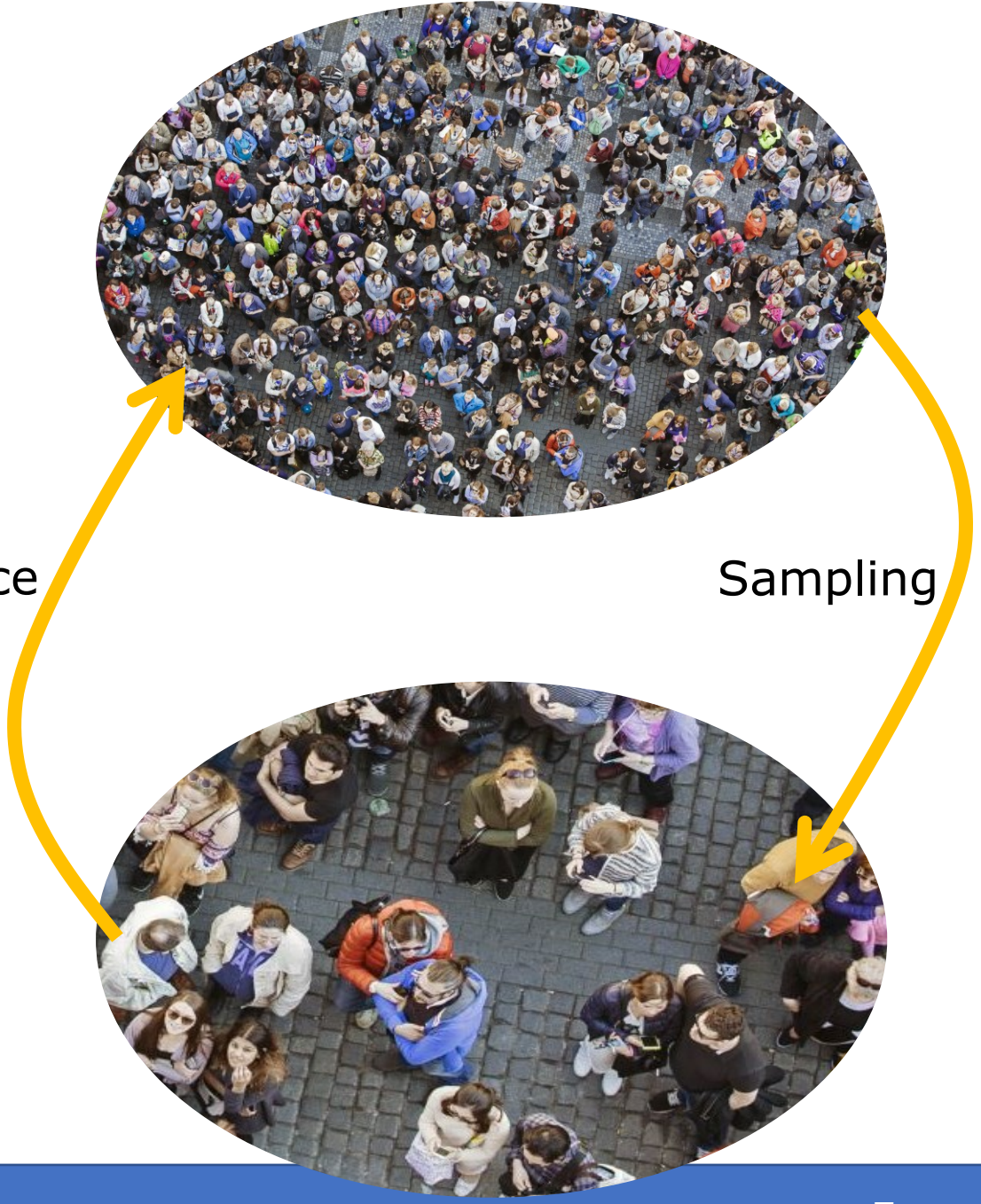
# Introduction

- Why to sample?



# Introduction

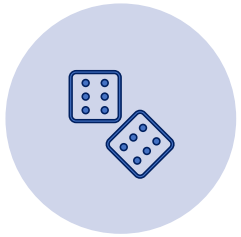
- Why to sample?
- What is a sample?
- How to sample?



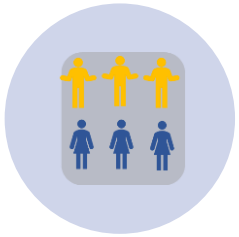
# Choosing a Sample

- **Assumption:** a sample is a good representation of the population.
- However, there is uncertainty as to how representative the sample is.
  - The sample is not perfect.
  - Not all samples are the same.
- **Goal:** maximize precision and minimize cost.
- Therefore, the way the sample is taken matters.

# Sampling Methods



SIMPLE RANDOM



STRATIFIED

CLUSTER

SYSTEMATIC

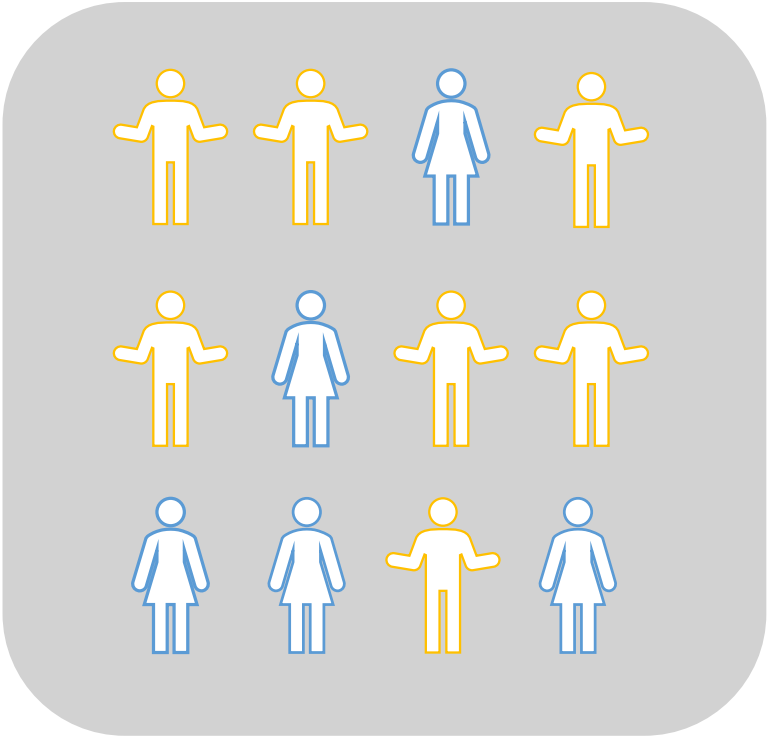
CONVENIENCE

# Simple Random Sample

- Selected in a manner such that every sample of size n from the population has an equal chance of being selected
- Unbiased
- Representative
- Practical for manufacturing or production populations
- Can be difficult and expensive when dealing with people
- Needs a sampling frame

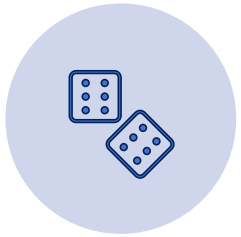
# How to choose a random sample?

- Number all members of the population sequentially.
- Select random values from the numbers assigned to the population members.
- Create the sample by using population members with numbers corresponding to those randomly selected.

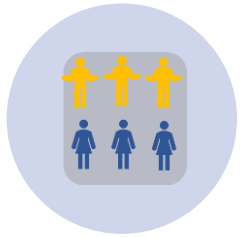


1 3 7 12

# Sampling Methods



SIMPLE RANDOM



STRATIFIED

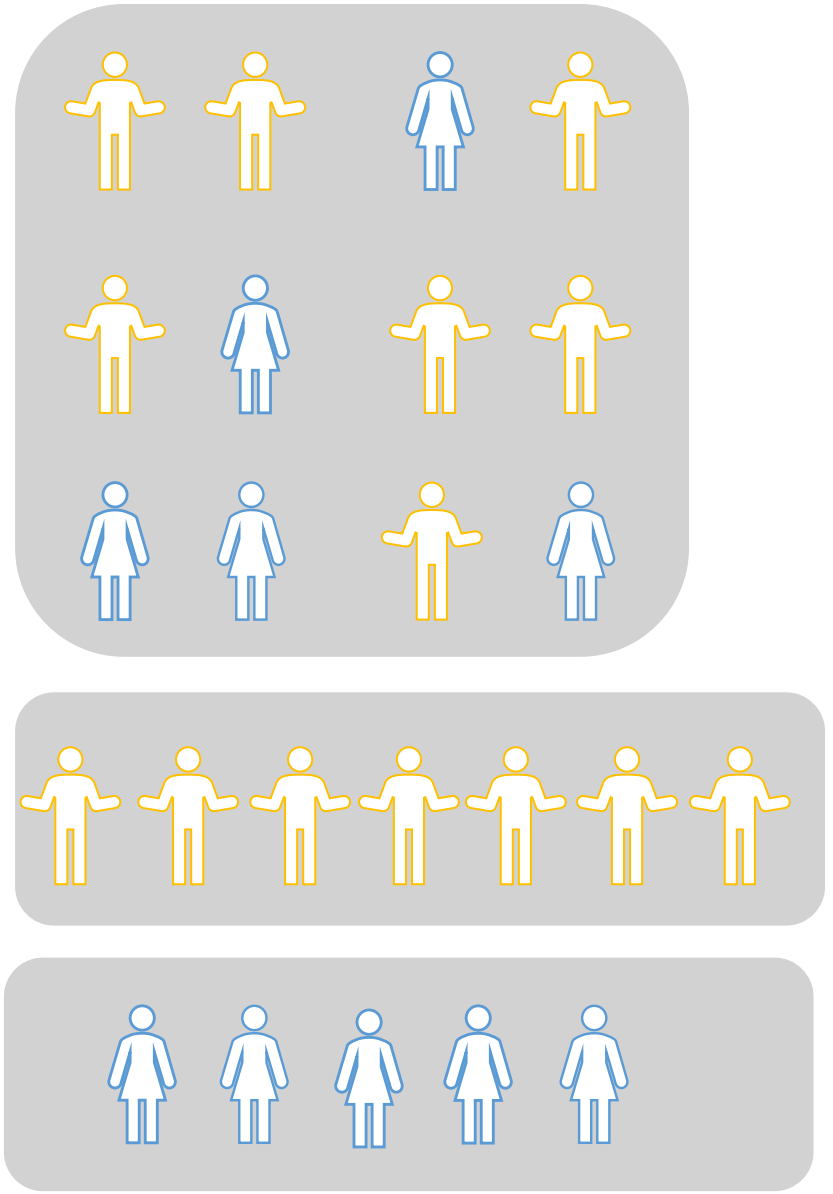
CLUSTER

SYSTEMATIC

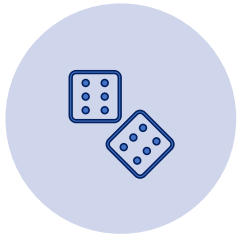
CONVENIENCE

# Stratified Sampling

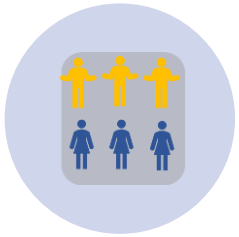
- Divide the entire population into distinct subgroups called strata. Then, draw a random sample from each stratum.
- Strata is chosen specifically to represent different characteristics within the population.
- Can be a very good random representation.
- Can be hard to implement.
- Sampling frame with considerable information about the population is required.
- Ensures the diversity of the sample.



# Sampling Methods



SIMPLE RANDOM



STRATIFIED

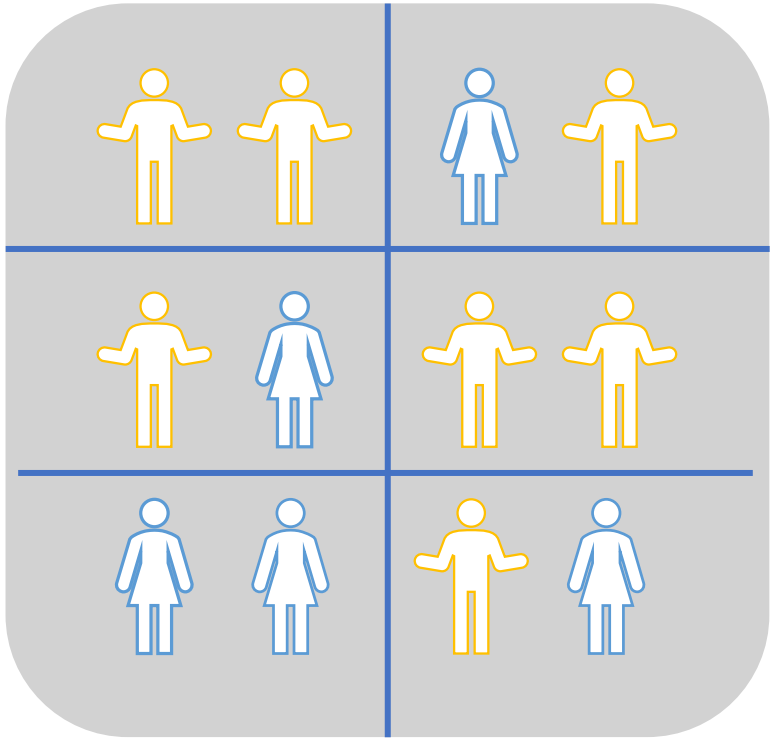
CLUSTER

SYSTEMATIC

CONVENIENCE

# Cluster Sampling

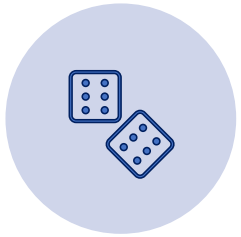
- Divide the entire population into existing segments of clusters. Make a random selection of clusters. Include every member of each selected cluster in the sample.
- More convenient and practical than simple random sample
- Can lead to biased sample or a sample that is not representative
- Less cost



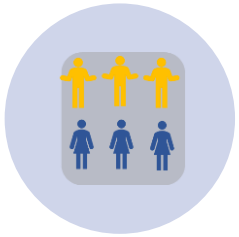
# Think-Pair-Share (10 minutes)

- When are estimates based on stratified random sampling likely to be more precise than estimates based on simple random samples of comparable total size?
- When are estimates based on cluster sampling is likely to be most precise?

# Sampling Methods



SIMPLE RANDOM



STRATIFIED

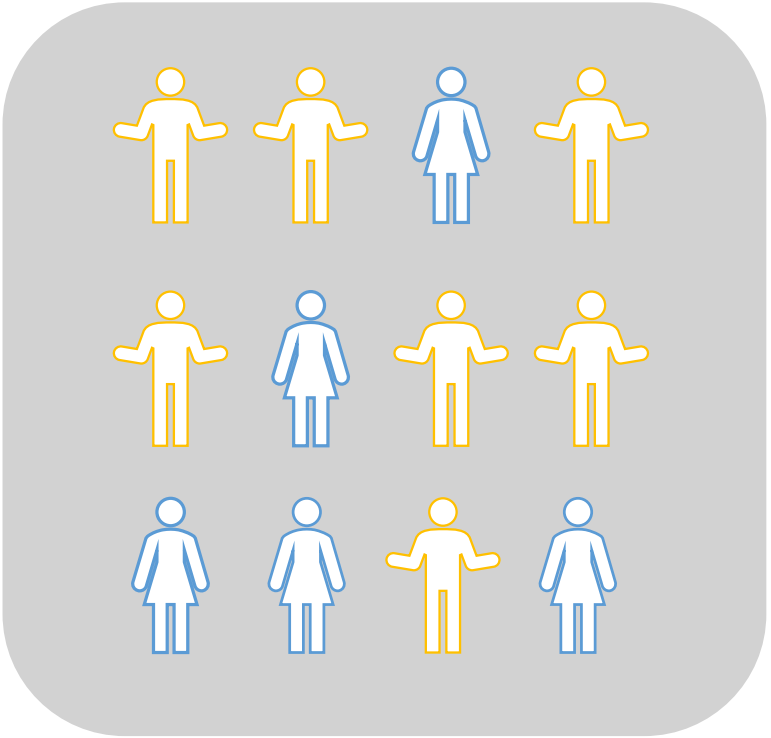
CLUSTER

SYSTEMATIC

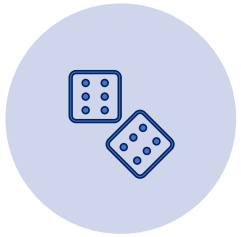
CONVENIENCE

# Systematic Sampling

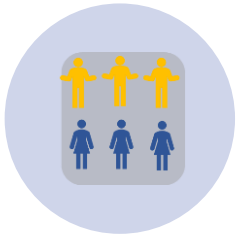
- Number all the members of the population sequentially. From a starting point selected at random, include every $Kth$ member of the population.
- Easier to implement than simple random samples
- Good approximation of a random sample
- May result in a bias if there is a certain pattern in the population.



# Sampling Methods



SIMPLE RANDOM



STRATIFIED

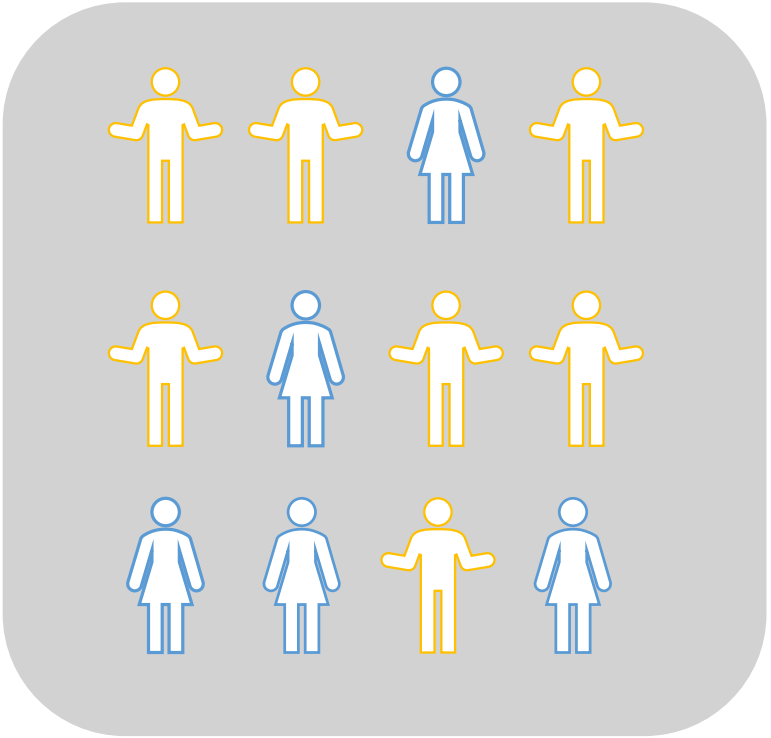
CLUSTER

SYSTEMATIC

CONVENIENCE

# Convenience Sampling

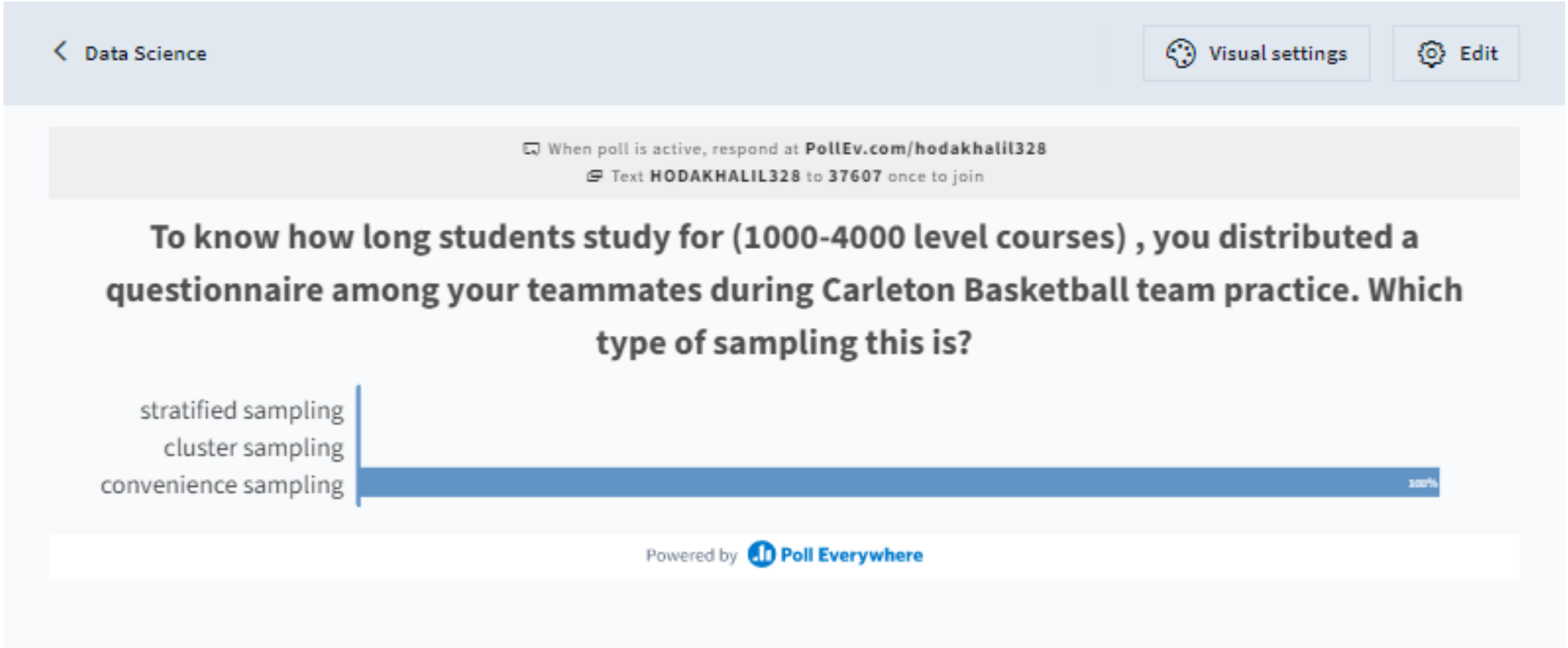
- Use data from the population members that are readily available
- Biased
- Quick and cheap



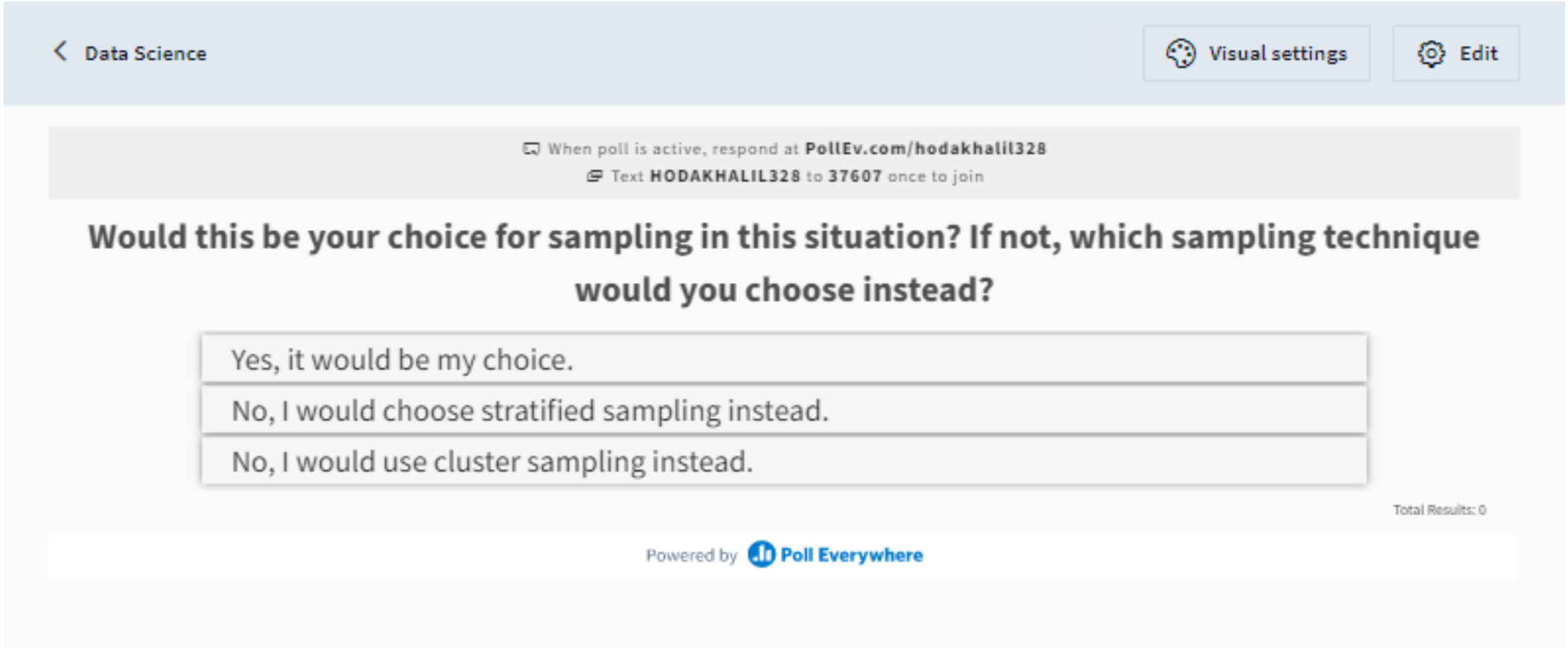
# Questions to ask when choosing a sampling method

- What is the nature of the population?
- What are the available resources?
- Will this sampling produce a biased or unbiased sample?

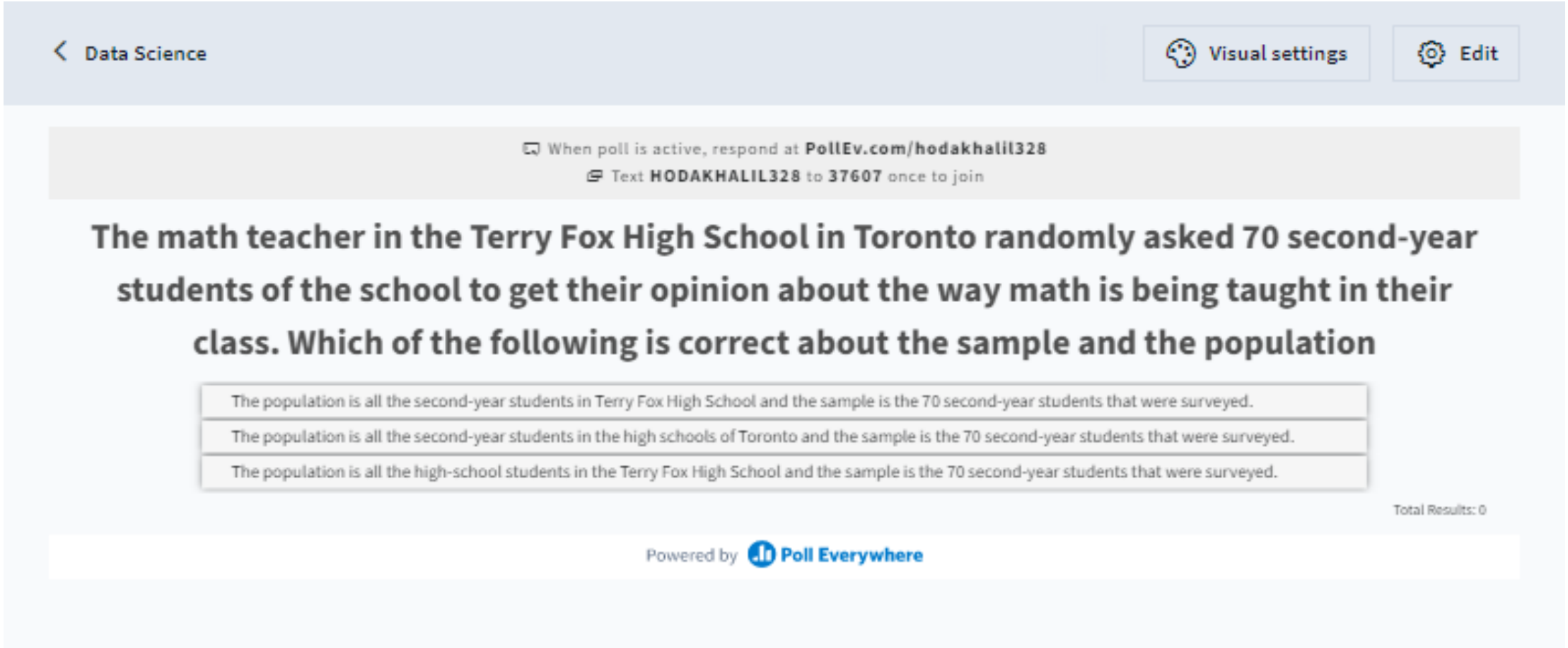
# In-class exercise (10 minutes) 1/4



# In-class exercise (10 minutes) 2/4



# In-class exercise (10 minutes) 3/4



# In-class exercise (10 minutes) 4/4

# Variation

- Real or natural variation
- Explainable/Confounding variation
- Sampling error
- Non sampling error (e.g., measurement errors)

# Sources of Bias in a Sample

## Sampling Error

- The difference between measurements from sample and corresponding measurements from the perspective population.
- Caused by the fact that the sample does not perfectly represent the population.
- Will be always there, but can be reduced by increasing the sample size.

## Nonsampling error

- The result of poor sampling design, inaccurate data collection, faulty measuring instruments, bias in questionnaires, and so on.
- A mistake has occurred.
- Can be reduced by fixing the design or methodology.
- Selection bias and nonresponse bias.

# Summary

- When an analyst is interested in studying characteristics of a population, they may draw a sample from that population if they don’t have access to the entire population, the means to study the population are not available, or to minimize errors.
- Inferential statistics allows us to allow us to make conclusions beyond the data we have through inferring conclusions about the population from the sample data.
- The basic ideas behind inference are that the sample is likely to be representative, there is uncertainty as to how representative the sample is, and the way the sample is taken matters.
- Some sampling methods are simple random, cluster, convenience, and systematic sampling. Deciding which sampling method to use depends on the nature of the population and the resources available.
- No matter how the sampling was done, there will be sampling error due to the difference between measurements from the sample and corresponding measurements from the perspective population. Sampling errors can be reduced by increasing the population size. On the contrary, non-sampling errors cannot be reduced except by fixing the design or the methodology.

# Further Optional Readings

- Introduction to Statistics and Data Analysis with Exercises and Solutions in R, by Christian Heumann; Michael S. Shalabh. DOI 10.1007/978-3-319-46162-5DOI 10.1007/978-3-319-46162-5
- Statistical Analysis and Data Display, by Richard M. Heiberger and Burt Holland. DOI 10.1007/978-1-4939-2122-5