Sampling

EDP 619 Week 3

Dr. Abhik Roy

Two Approaches to Sampling

Nonprobability

each person in your target population *does not* have an equal chance of being selected

Probability

each unit in your target population *must* have an equal chance of being selected

Nonprobability Sampling

- Probability is usually unknown
- Does not rely on numerical data
- Inability to generalize to any populous
- Used when you want to say something about a discrete phenomena, a few select cases (people, places, objects, etc)

Characteristics

- Easier than probability based methods
- Nonrandom selection
- Sampling bias is present
- Samples are not considered representative of the populations from which they were drawn

Primary Types

Convenience

Purposive

Quota

Snowball

Convenience Sampling[1]

- Samples are selected based on
  > their availability to the researcher
- Good for
  > administering a pilot study
  >
  > generating a hypothesis
  >
  > gaining an initial sense of attitudes or opinions

Example

Crowdsourcing survey participants from a platform[2]

[1] aka **haphazard** or **accidental** sampling

[2] like Amazon Mechanical Turk (MTurk)

Purposive Sampling

- Samples are selected based on
  - elective criteria that define a unique group
  - targeting knowledgeable individuals[1]
- Good for
  - focusing on the depth of relatively small samples
  - identifying cases, individuals, or communities best suited for a study

Example

Choosing skilled candidates for a job vacancy

[1] aka **key informants**

Quota Sampling

- Samples are selected based on
  > defined subgroups that exhibit certain characteristics of interest
- Good for
  > gaining insight about a characteristic of a particular subgroup
  >
  > investigating relationships between different subgroups

Example

Assessing the the differences in the career goals among university freshman, sophomores, juniors, and seniors

Snowball Sampling[1]

- Samples are selected based on

  > individuals recruited by other individuals

- Good for

  > researching people with specific traits who might otherwise be difficult to identify and/or gain access to
  >
  > keeping costs low

Example

Studying the current living status of ex-convicts

[1] aka **chain** or **network** sampling

Why should I even care?

Because:

> Any choice will limit the type of utilizable quantitative study
>
> Not everything can be explained quantitatively
>
> Some studies even mandate a mixed methods design

Probability Sampling

- Based solely on the idea that a population can be represented by a subset of it given some error: **Random selection**

  Example: 45% ± 3% agree with...

- Ability to generalize to a certain populous

  Inability to describe individual phenomena at any great depth

Characteristics

- Greater difficulty than non-probability based methods
- Random selection
- Sampling bias is minimal, and samples are considered representative of the populations from which they were drawn
- Samples are representative of the populations from which they were drawn

Primary Types

Census

Simple Random Sample (SRS)

Systematic

Stratified

Cluster

Census

- Samples are selected based on

  > an official count or survey of a population, typically recording various details of individuals

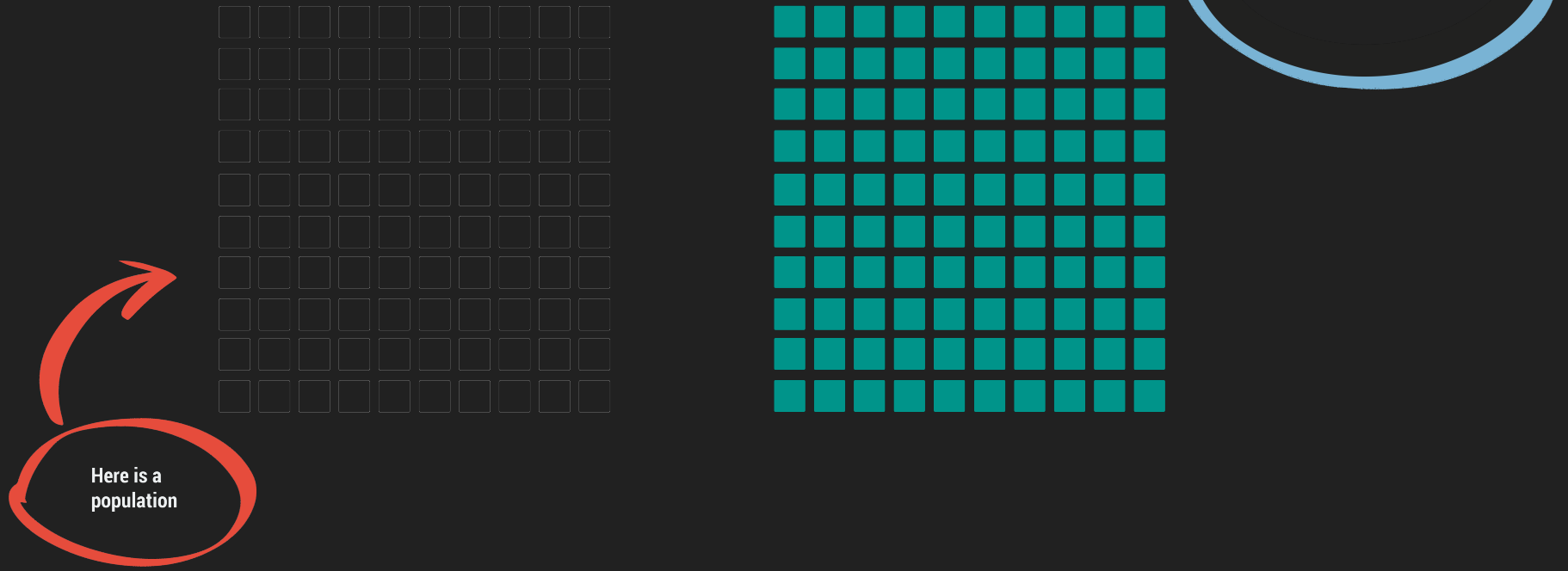
- Good for

  > ease of administration
  >
  > generalizing to an overall populous
  >
  > simple data analysis
  >
  > small samples

Example

The United States Census

General Idea



Characteristics

Benefits

> a lack of an error associated with a result
>
> self-weighting

Drawbacks

> extremely expensive
>
> time consuming
>
> typically infeasible

Simple Random Sample (SRS)

- Samples are selected based on
  - an equal probability of being picked
- Good for
  - ease of administration
  - generalizing to an overall populous
  - simple data analysis
  - situations where not a lot is known about a population
  - large samples

Example

Drawing names from a hat

Characteristics

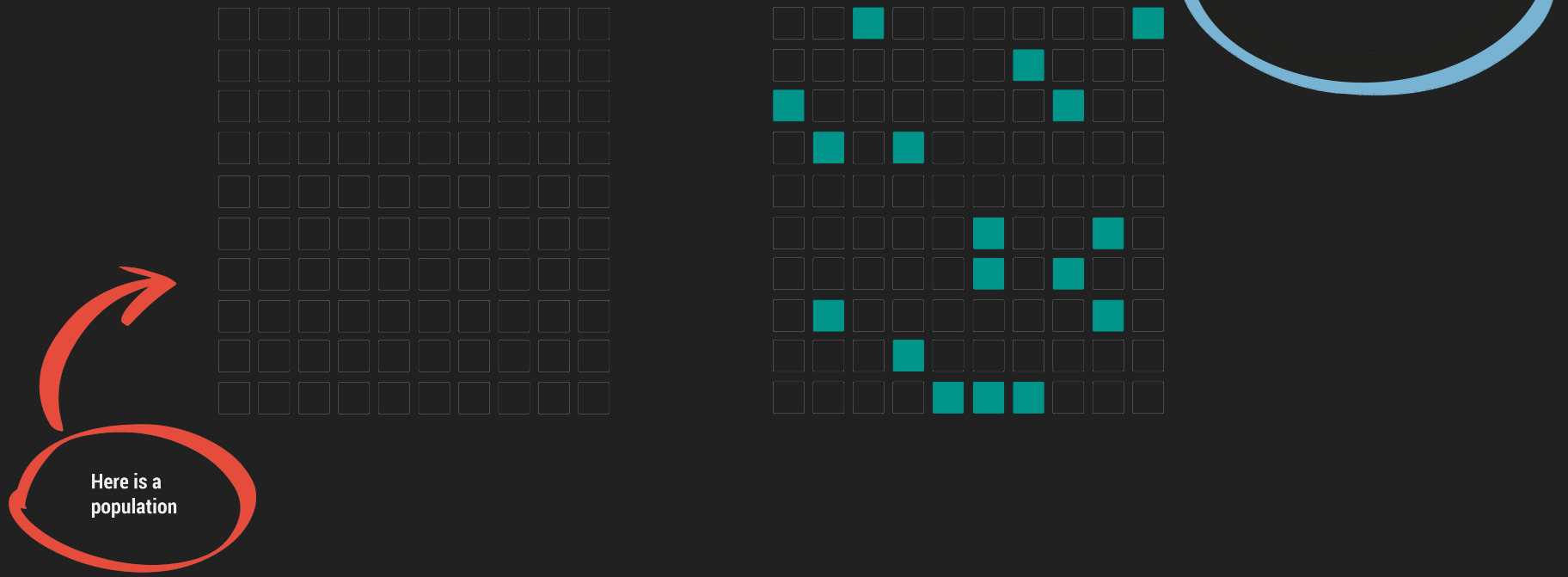
Benefits

- data collection can be efficiently performed on randomly distributed items
- simple error calculation
- self-weighting

Drawbacks

- expensive
- likely impractical
- possible underrepresentation of subgroups
- tedious
- time consuming
- vulnerable to sampling errors

General Idea



Systematic Sample

- Samples are selected based on

  > arranging of a population according to some ordering pattern and then the selection of elements at regular intervals from that that ordered list

- Good for

  > ease of administration
  >
  > automation of selection process[1]
  >
  > providing more information about a population than an SRS

Example

Picking every third house on a block to poll

[1] after selecting the first unit

Characteristics

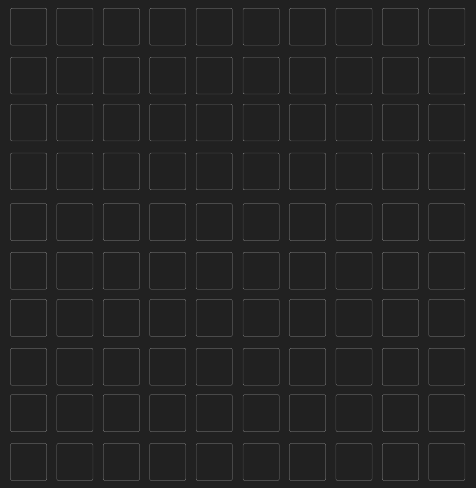
Benefits

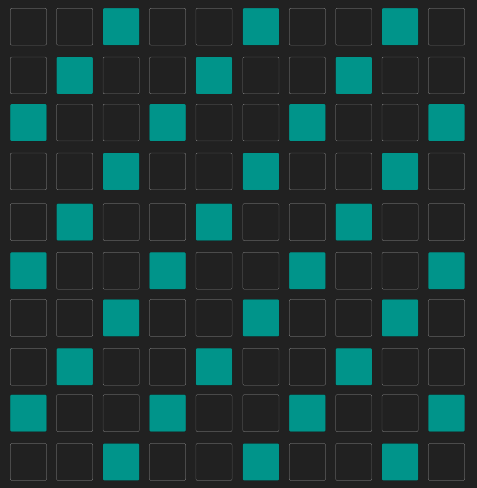
- most likely will provide a more robust information set per unit cost than an SRS
- less subjective to selection error than SRS
- simple selection process

Drawbacks

- dependence on a previous and next unit
- vulnerable to periodicities

General Idea





A sample is found by
1. lining everyone up
2. determining an initial person/interval
3. choosing indivduals based on their position

Here is a population

Stratified Random Sampling

- Samples are selected based on

  > a population being divided and subdivided into distinct groups[1] followed by a simple random or systematic sample in each

- Good for

  > ease of administration
  >
  > automation of selection process[1]
  >
  > providing more information about a population than an SRS

Example

Administering a survey to random units of all apartment complexes in a town

[1] aka **strata**

Characteristics

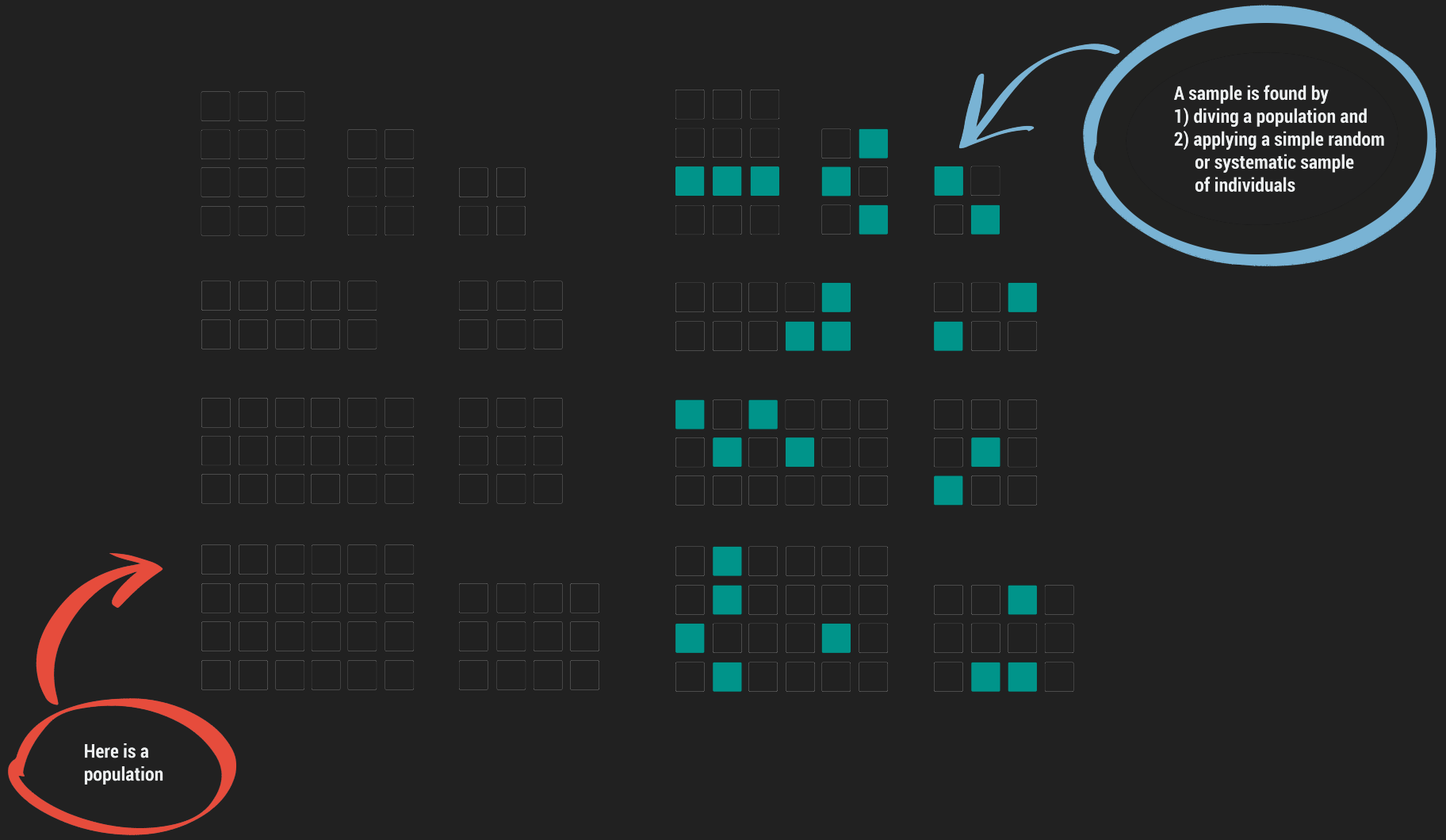
Benefits

- less variability than an SRS
- reduced sampling error
- reduced reported error and increases precision compared to an SRS

Drawbacks

- may be expensive
- strata must be implicitly defined

General Idea



Cluster Random Sampling

- Samples are selected based on

  > a population being divided and subdivided into distinct groups[1] followed by a random sample of those units with census in each

- Good for

  > when lacking a sampling frame
  >
  > cost efficiency is needed

Example

Picking every third house on a block to poll

[1] aka **clusters**

Characteristics

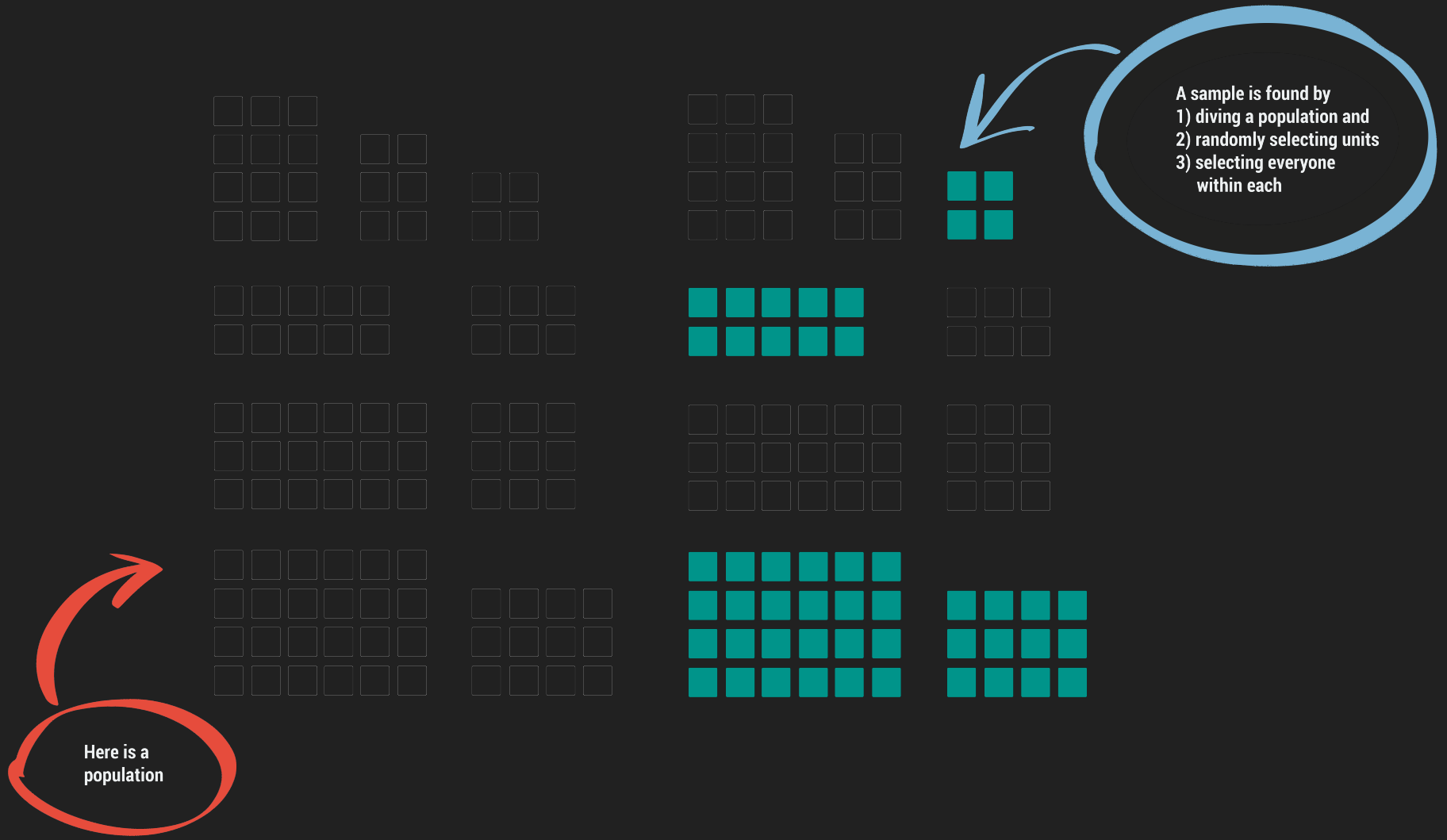
Benefits

- clusters can be stratified if necessary which results in increased precision
- less subjective to selection error than SRS
- simple selection process

Drawbacks

- may not represent diversity within a populous
- prone to high sampling errors
- requires a larger sample size than SRS

General Idea



That's it!

If you have any questions, please reach out