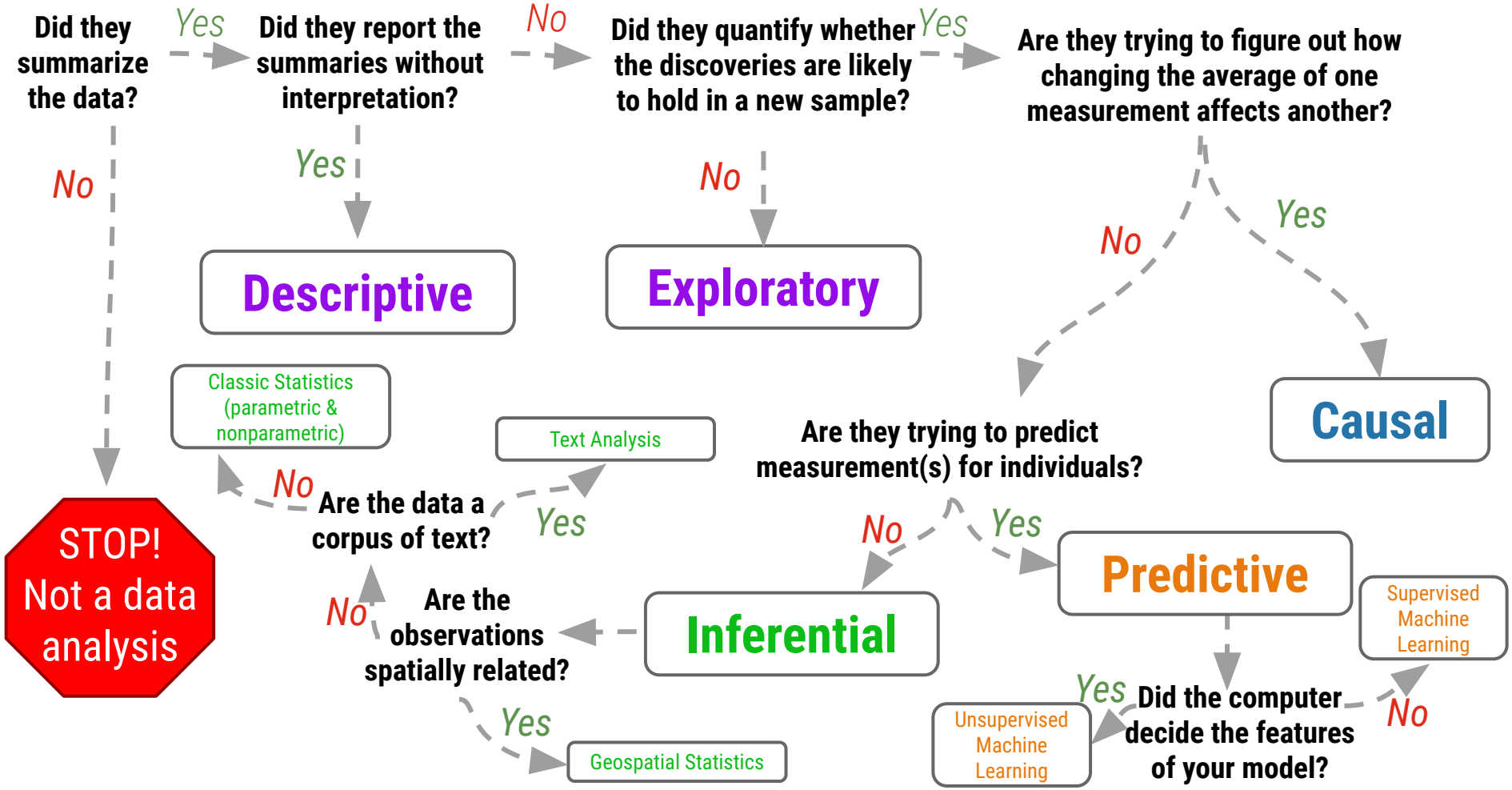
Plan:

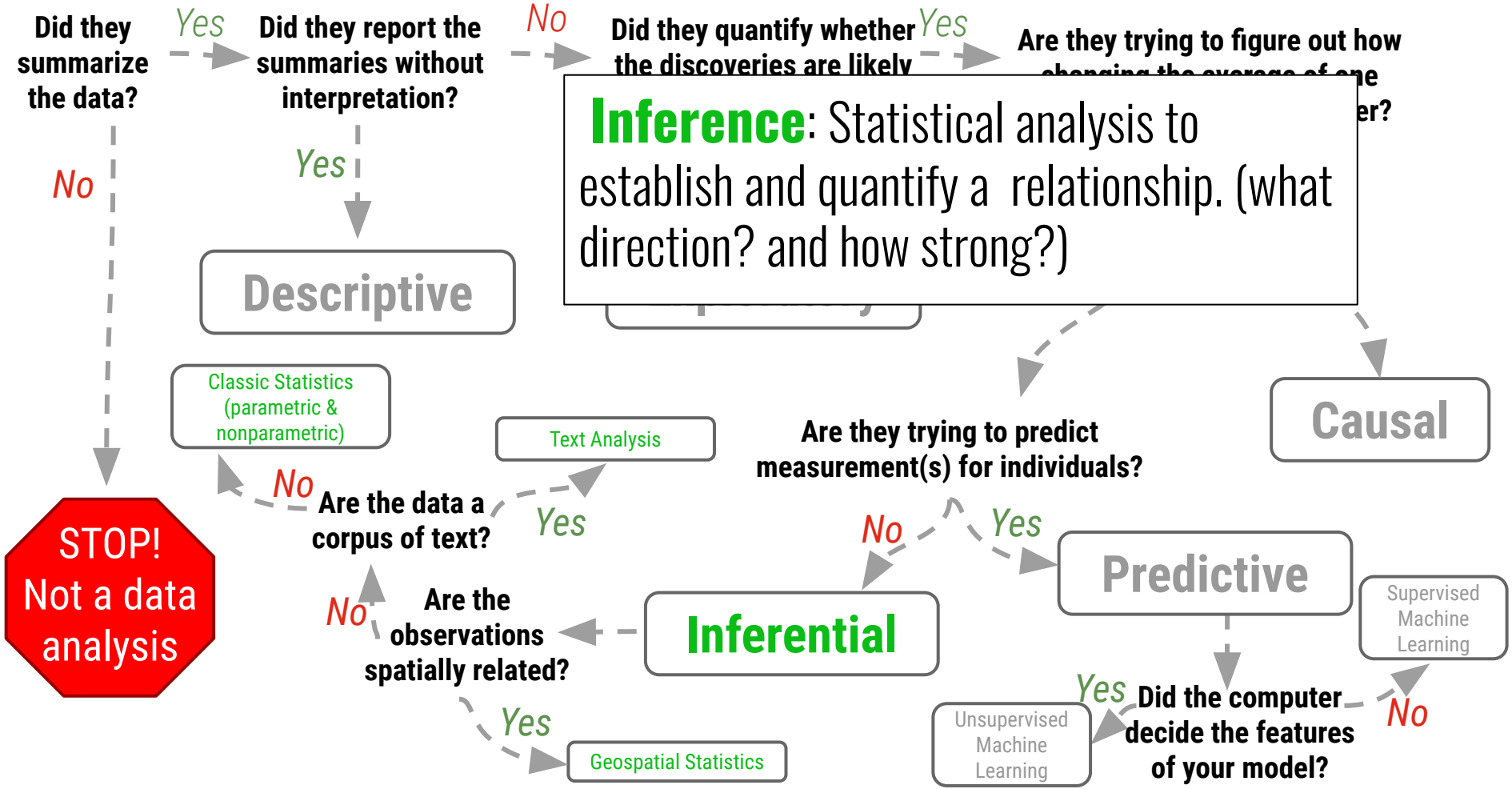
1. Introduce Inferential analysis
2. Discuss randoming sampling

Inferential Analysis: Sampling

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu


Did they summarize the data?
Yes
Did they report the summaries without interpretation?
No
Did they quantify whether the discoveries are likely to hold in a new sample?
Yes
Are they trying to figure out how changing the average of one measurement affects another?
No
Yes
Causal
No
Descriptive
Yes
No
Exploratory
Yes
STOP! Not a data analysis
No
Classic Statistics (parametric & nonparametric)
Are the data a corpus of text?
Yes
Text Analysis
Are they trying to predict measurement(s) for individuals?
No
Yes
Predictive
Supervised Machine Learning
No
Inferential
Are the observations spatially related?
No
Yes
Geospatial Statistics
Yes
Unsupervised Machine Learning
Did the computer decide the features of your model?
No


Did they summarize the data?
Yes
Did they report the summaries without interpretation?
No
Did they quantify whether the discoveries are likely
Yes
Are they trying to figure out how
No
Yes
Inference: Statistical analysis to establish and quantify a relationship. (what direction? and how strong?)
Descriptive
Classic Statistics (parametric & nonparametric)
Causal
Text Analysis
Are they trying to predict measurement(s) for individuals?
No
Are the data a corpus of text?
Yes
No
Yes
STOP! Not a data analysis
Predictive
Supervised Machine Learning
No
Are the observations spatially related?
Inferential
Yes
Unsupervised Machine Learning
Yes
Did the computer decide the features of your model?
No
Yes
Geospatial Statistics

- **Problem:** Does Sesame Street affect kids brain development?
- **Data science question:** What is the relationship between watching Sesame Street and test scores among children?
- **Type of analysis:** Inferential analysis

Sesame Street viewership

??

Test scores

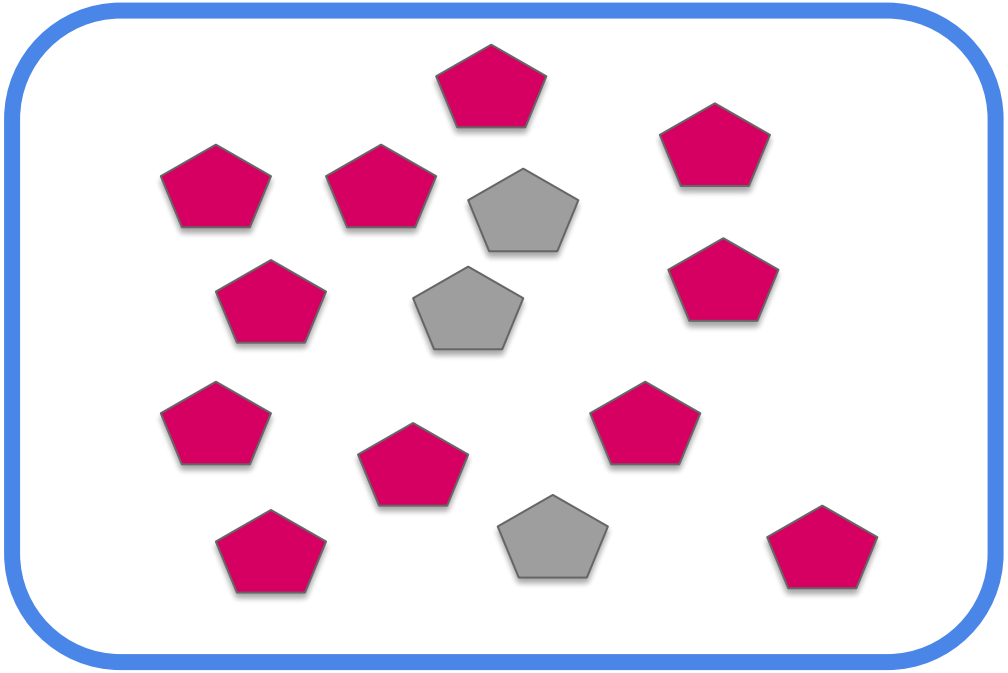
Establishing & Stating Your Null and Alternative Hypotheses Helps Guide Your Analysis

Null Hypothesis:

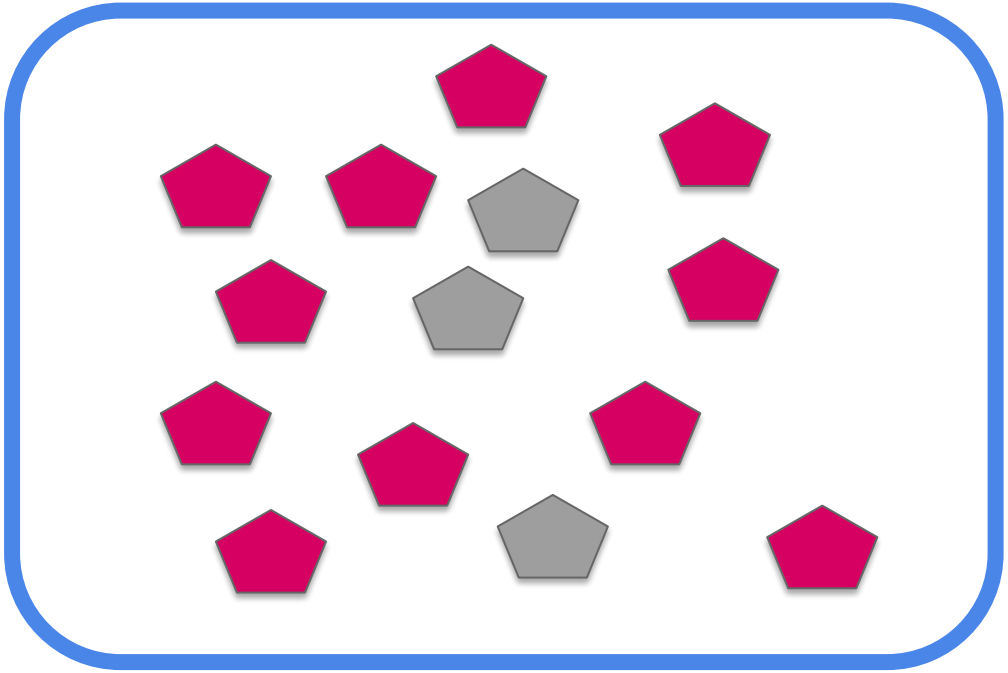
H_0: Sesame Street has *no effect* on kids brain development

Alternative Hypothesis:

H_a: Watching Sesame Street *has an effect* on kids' brain development

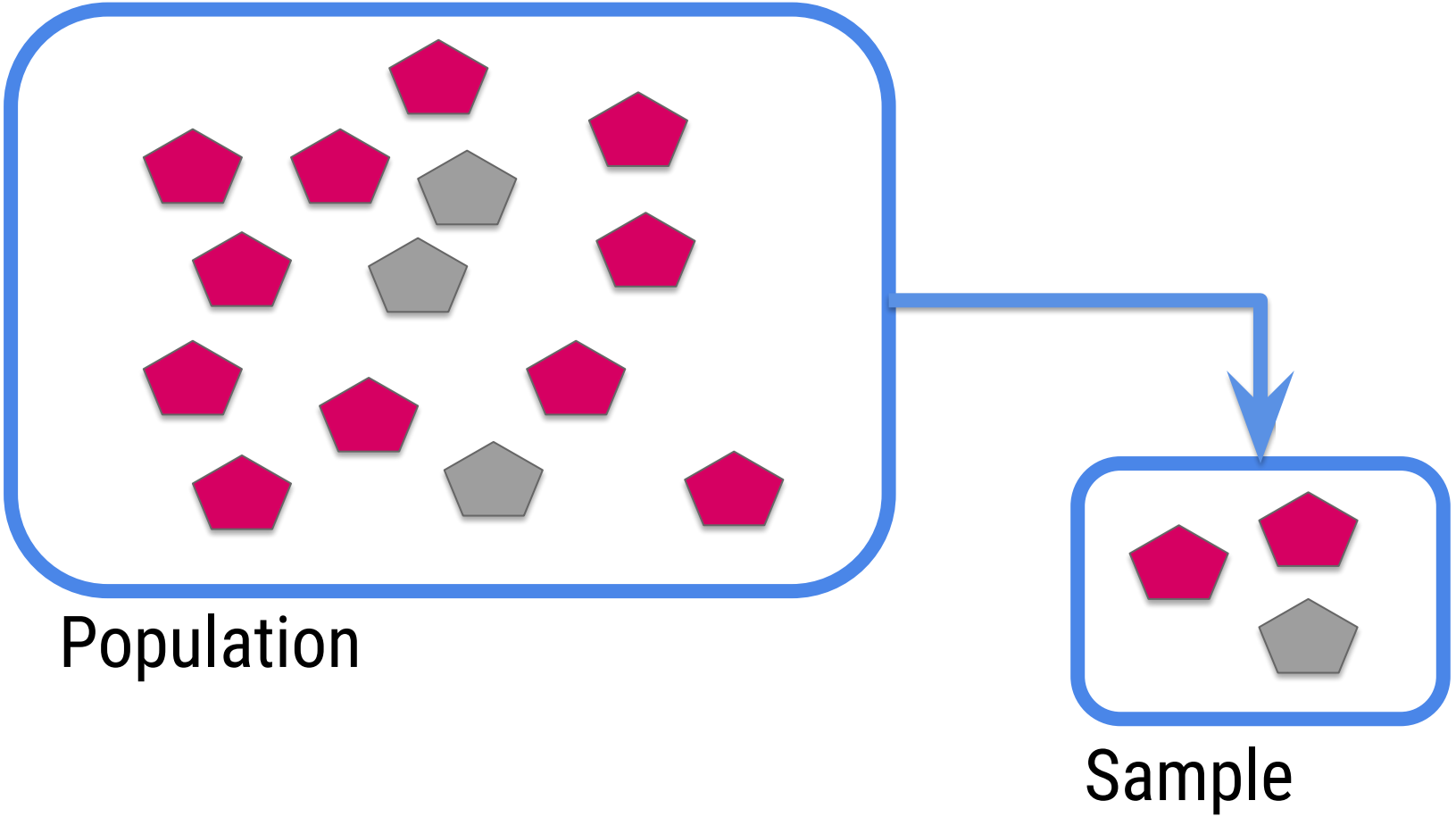


Population



Population

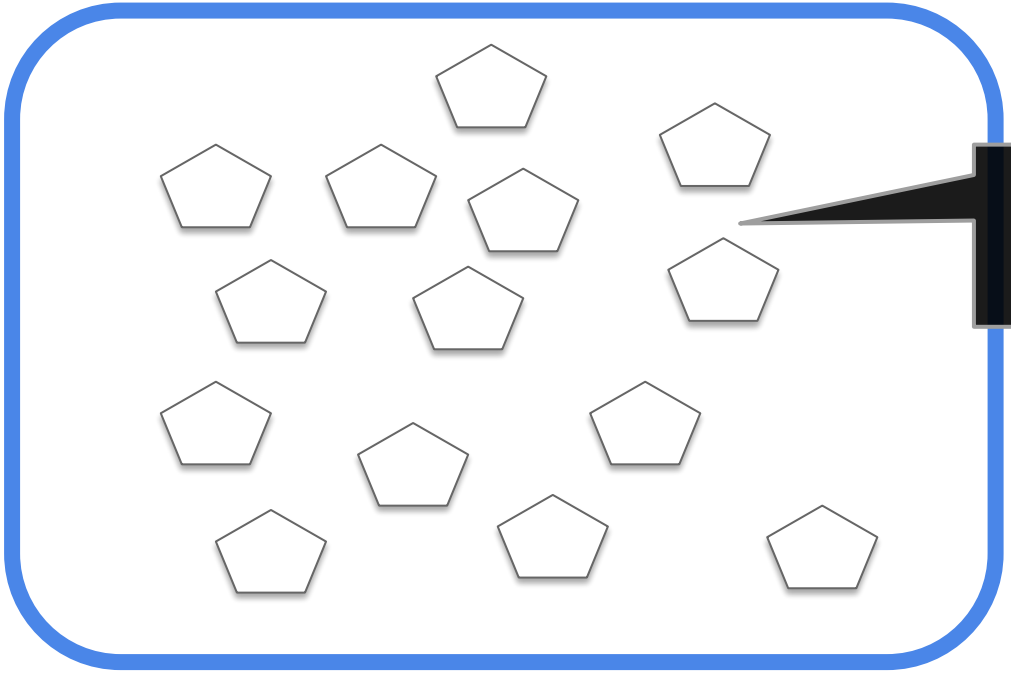
In our Sesame street example, the population would be all children


Population
Sample

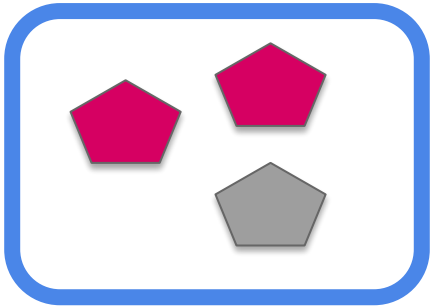
Population

In our Sesame street example, the sample would be the children included in the study

Sample

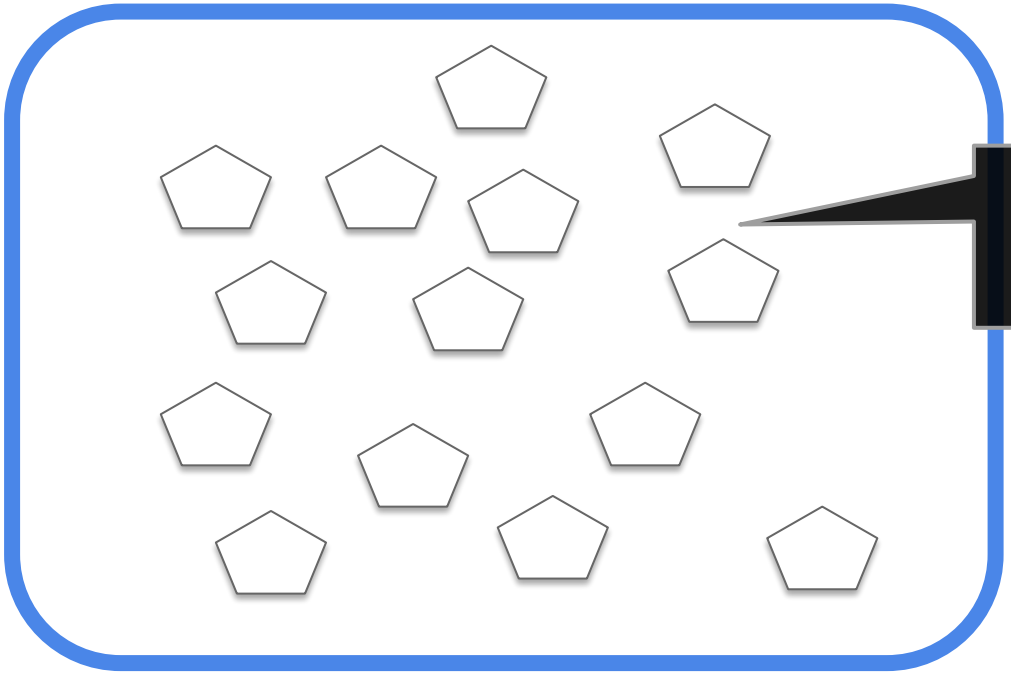


Population



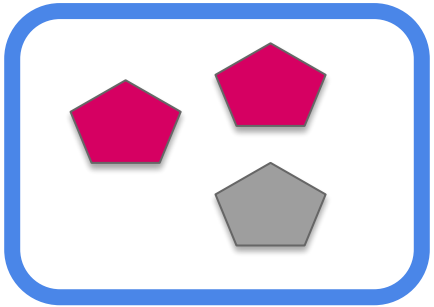
Sample

We don't know how much Sesame street was watched by or the tests scores of <u>all kids</u>

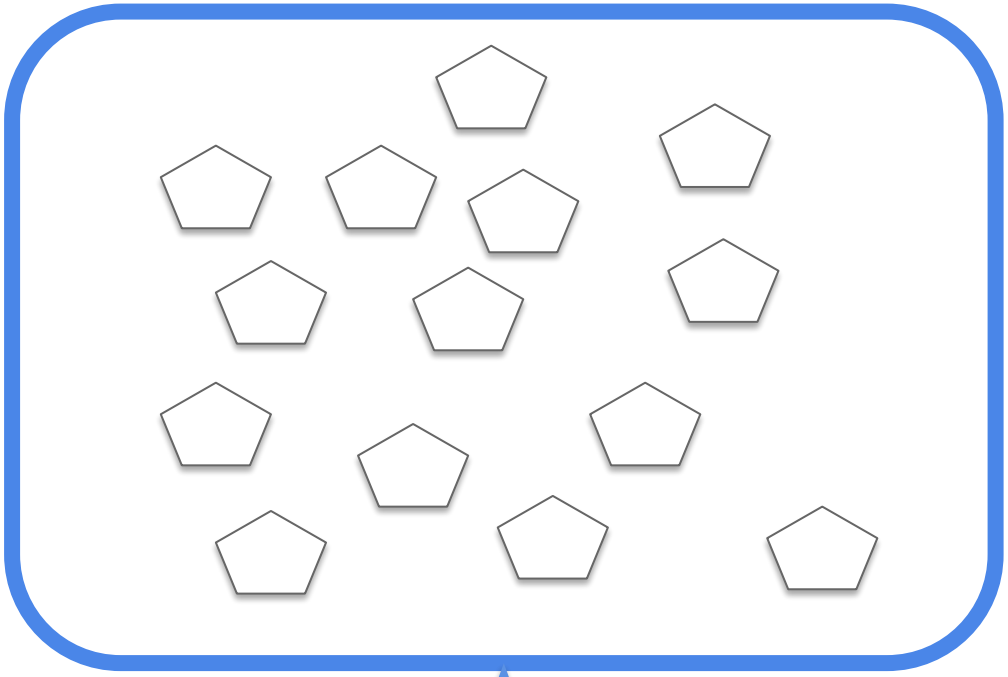


¯_(ツ)_/¯

Population

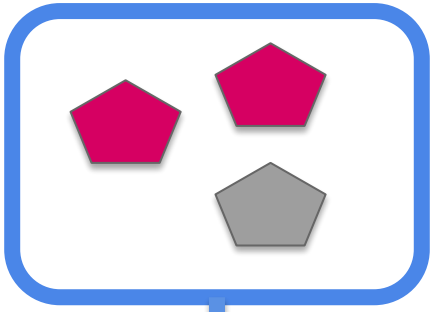


Sample



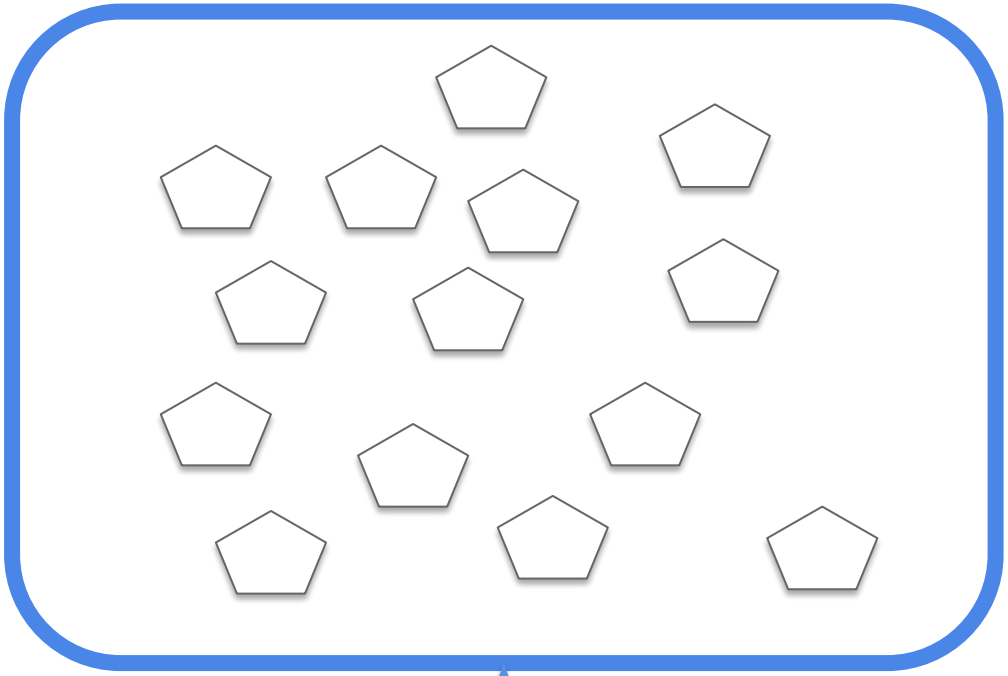
Population

Based on the relationship we see in our sample, we can infer the answer to our question in our population



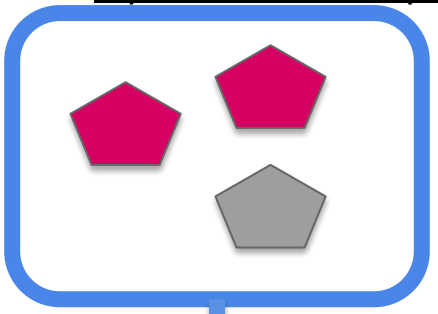
Sample

Inference!



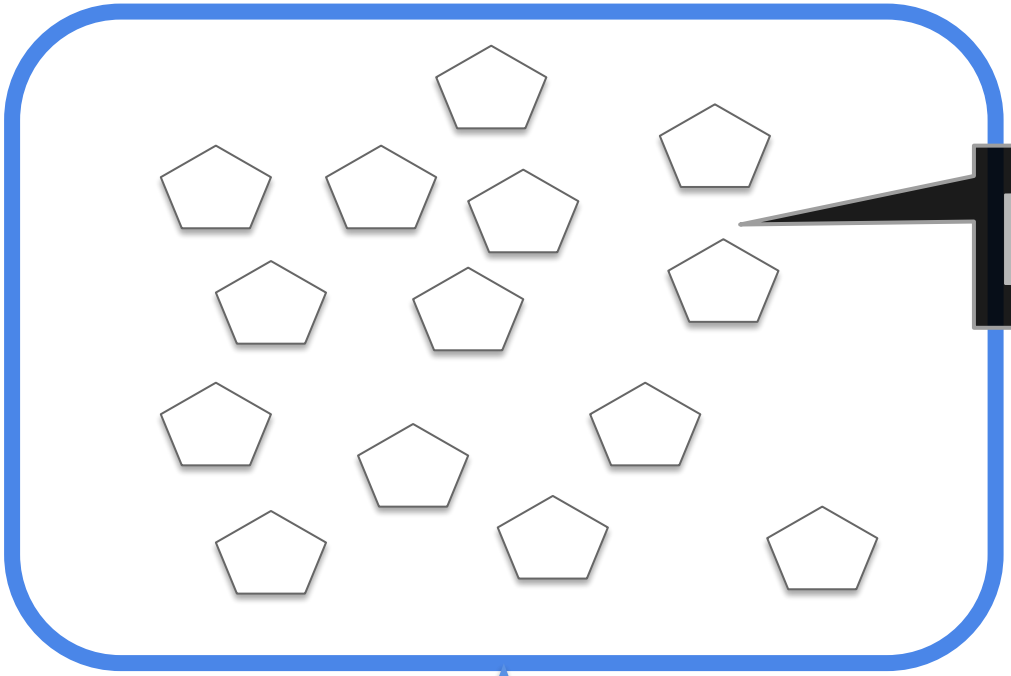
Population

So we look at Sesame street viewing and test scores in a representative sample of kids



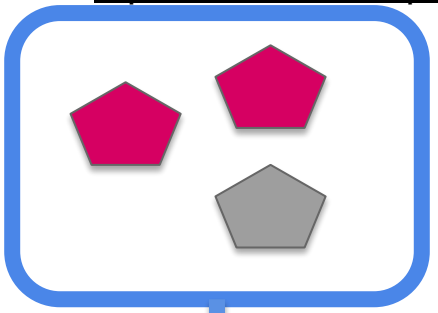
Sample

Inference!



Best guess

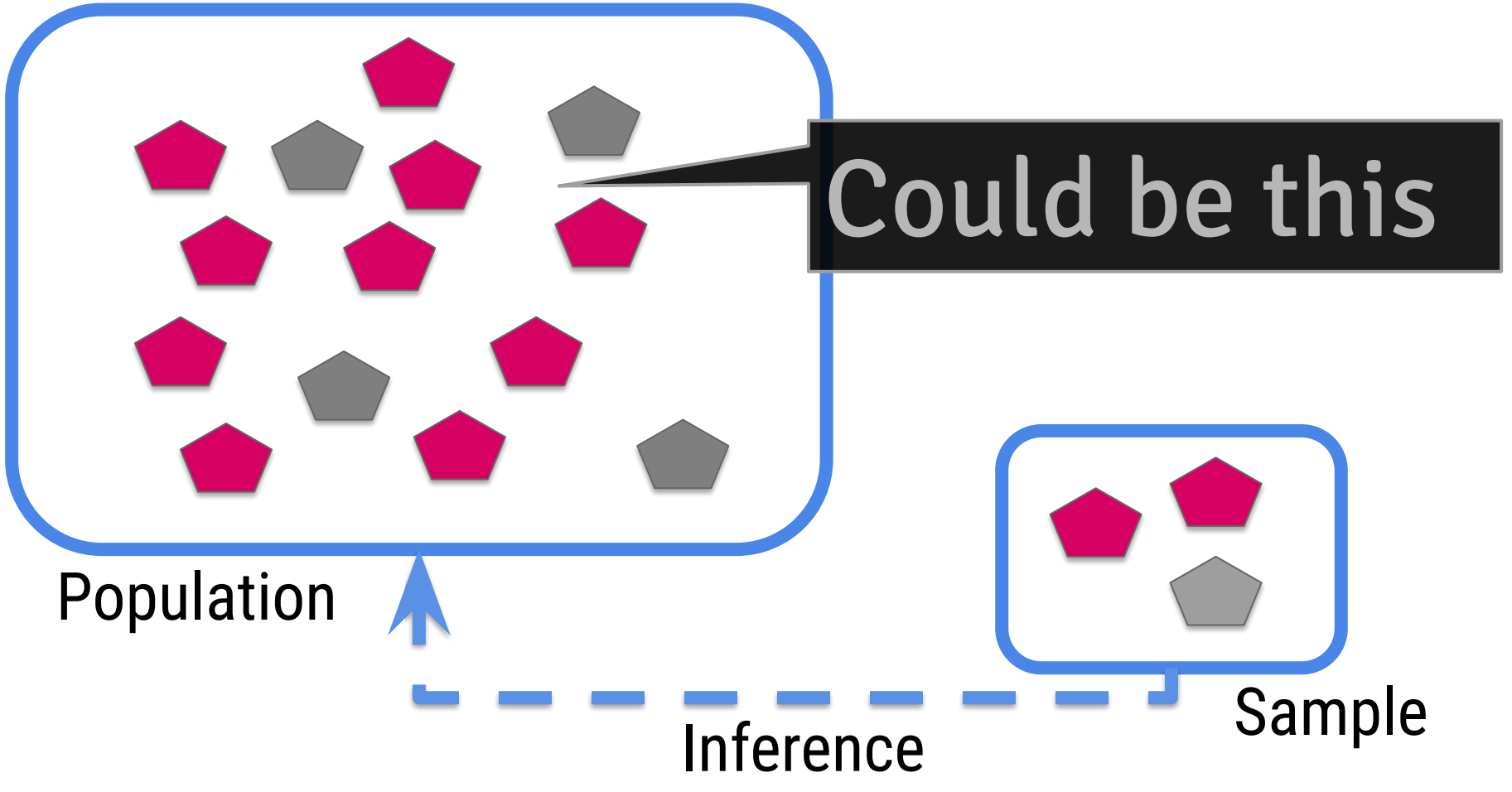
So we look at Sesame street viewing and test scores in a representative sample of kids

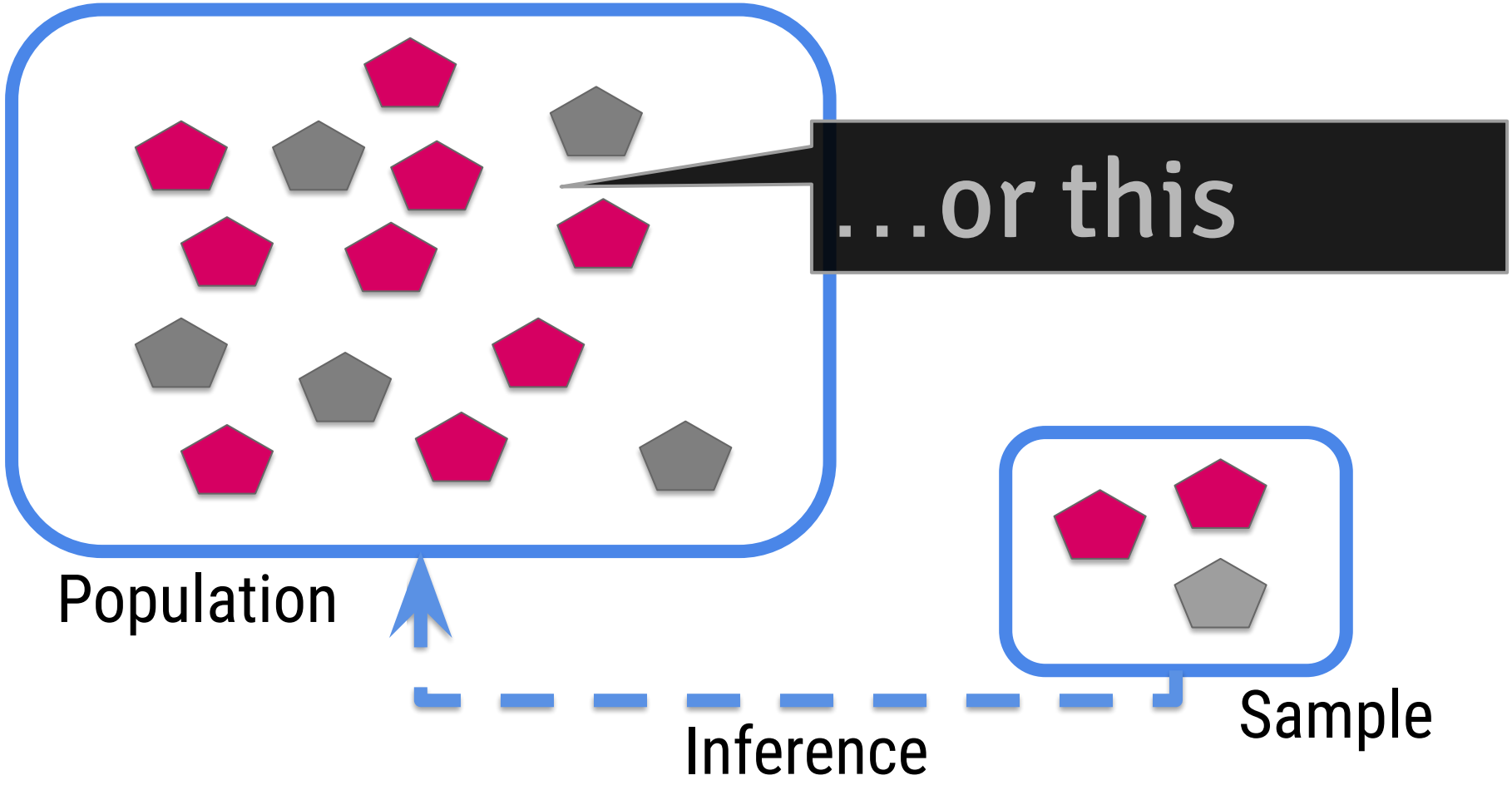


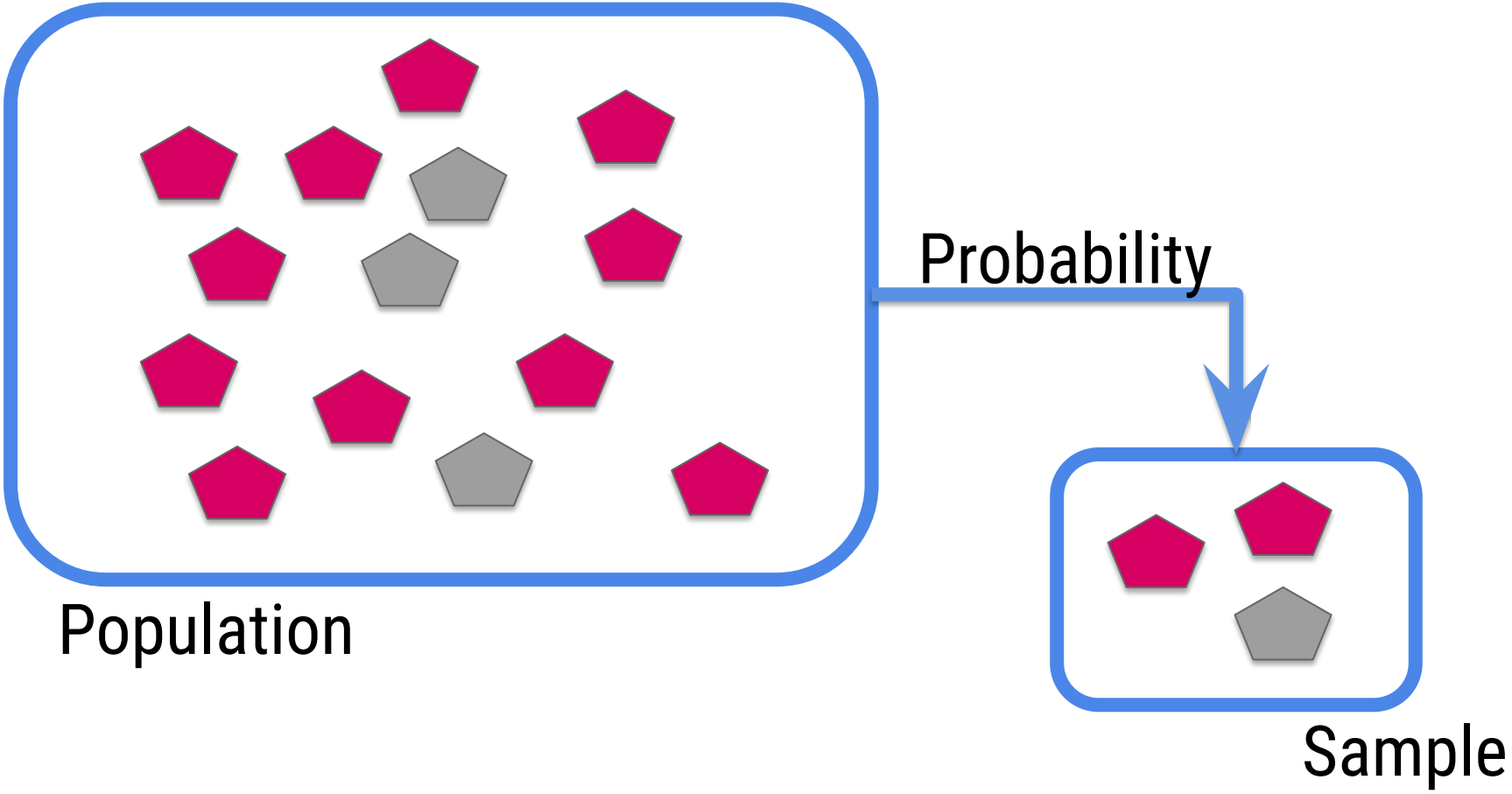
Population

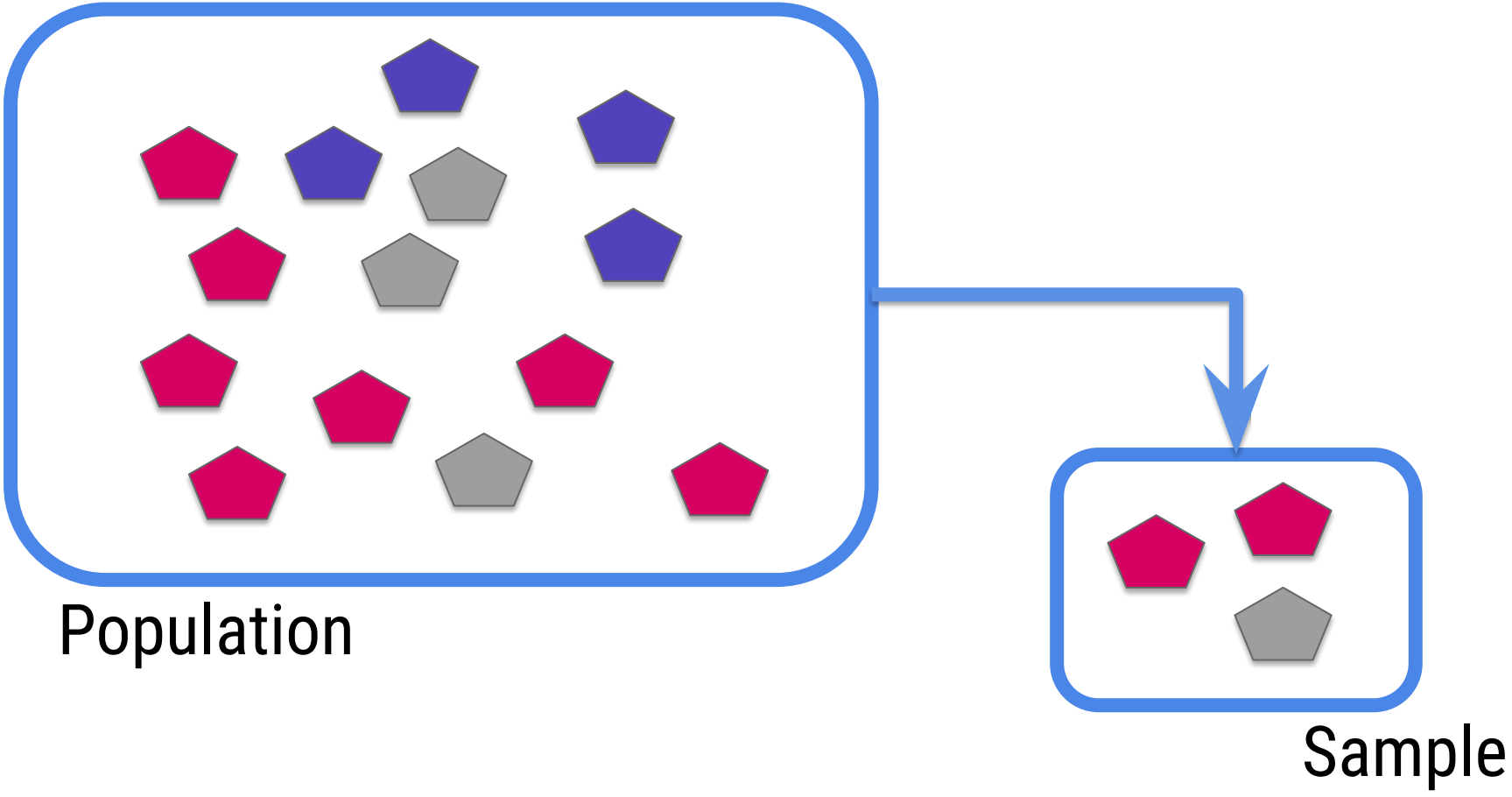
Inference!

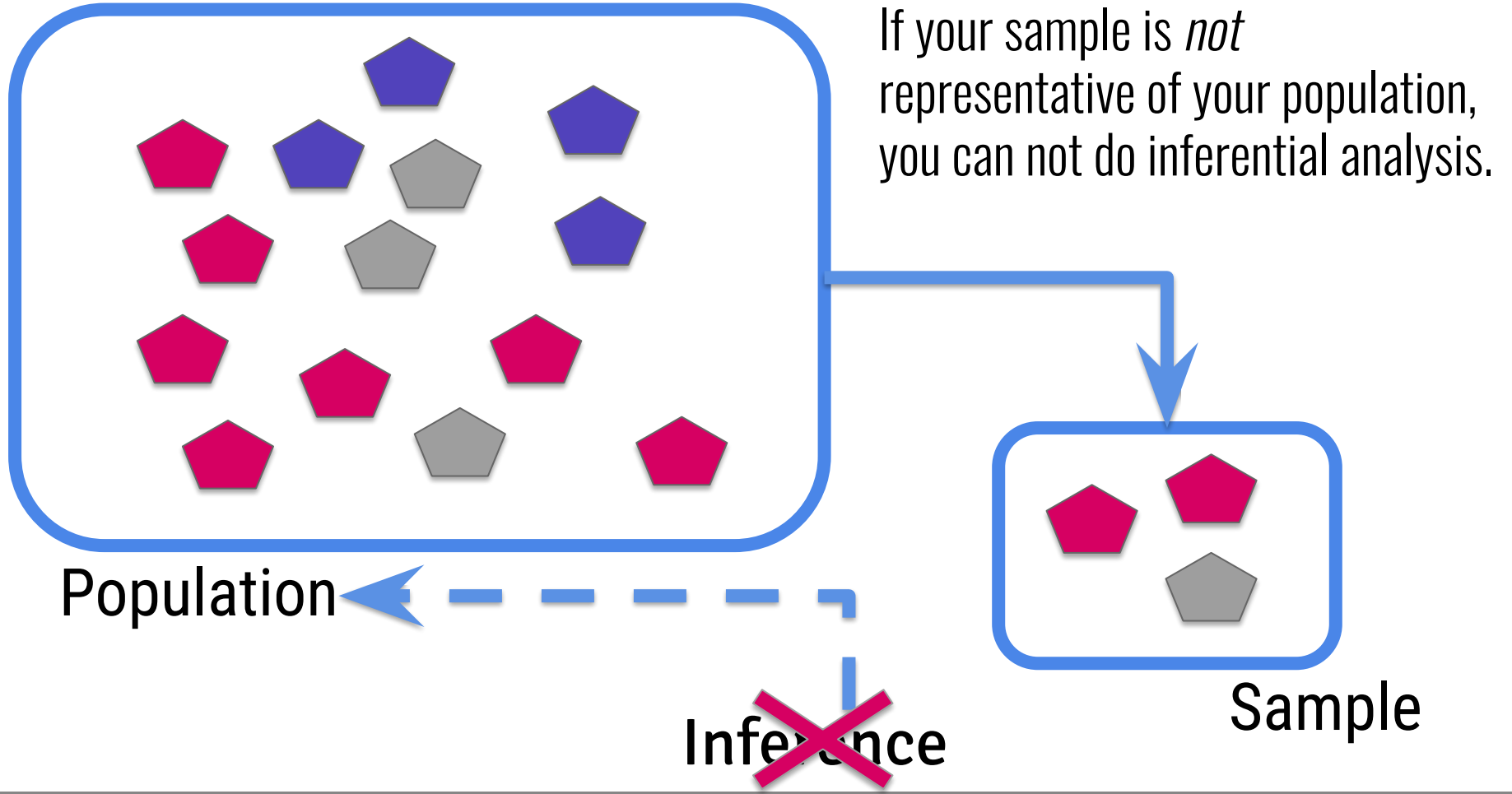
Sample


Could be this
Population
Inference
Sample


...or this
Population
Sample
Inference


Probability
Population
Sample


Population
Sample


If your sample is *not* representative of your population, you can not do inferential analysis.
Population
Sample
Inference