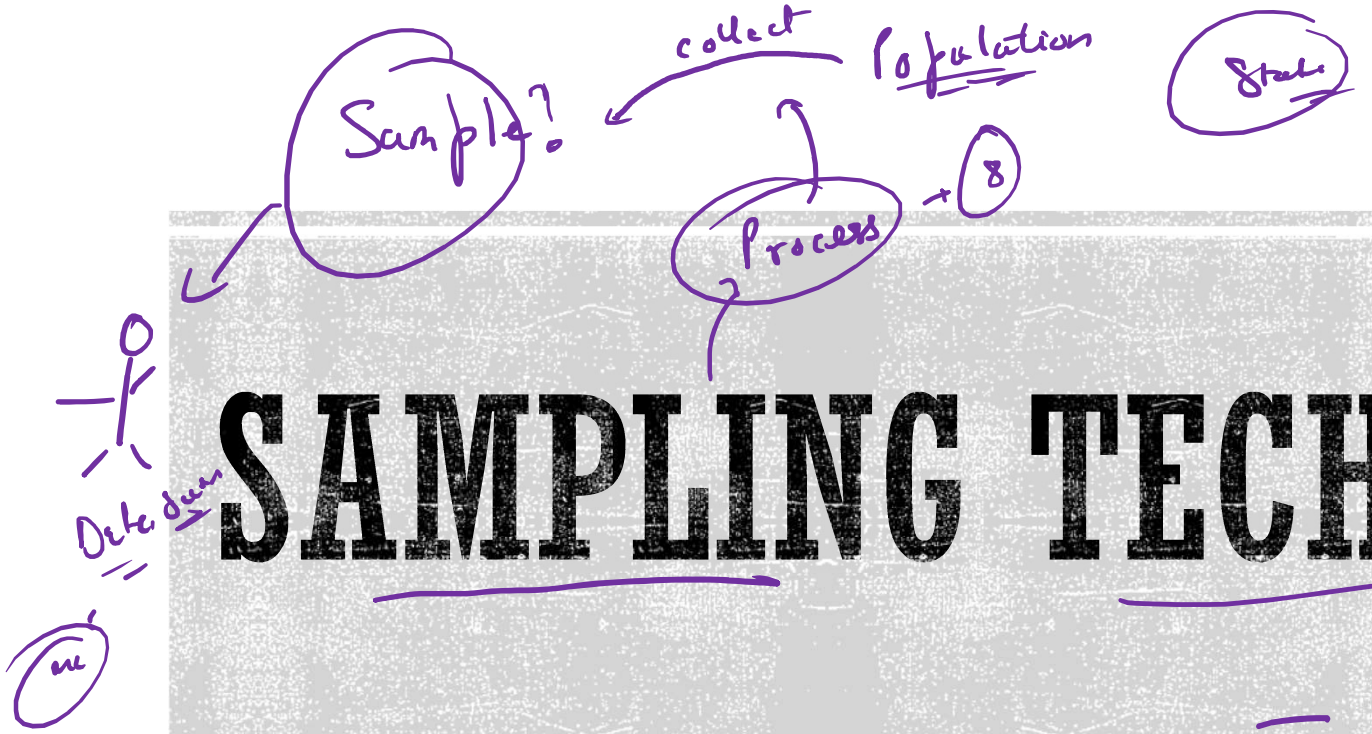
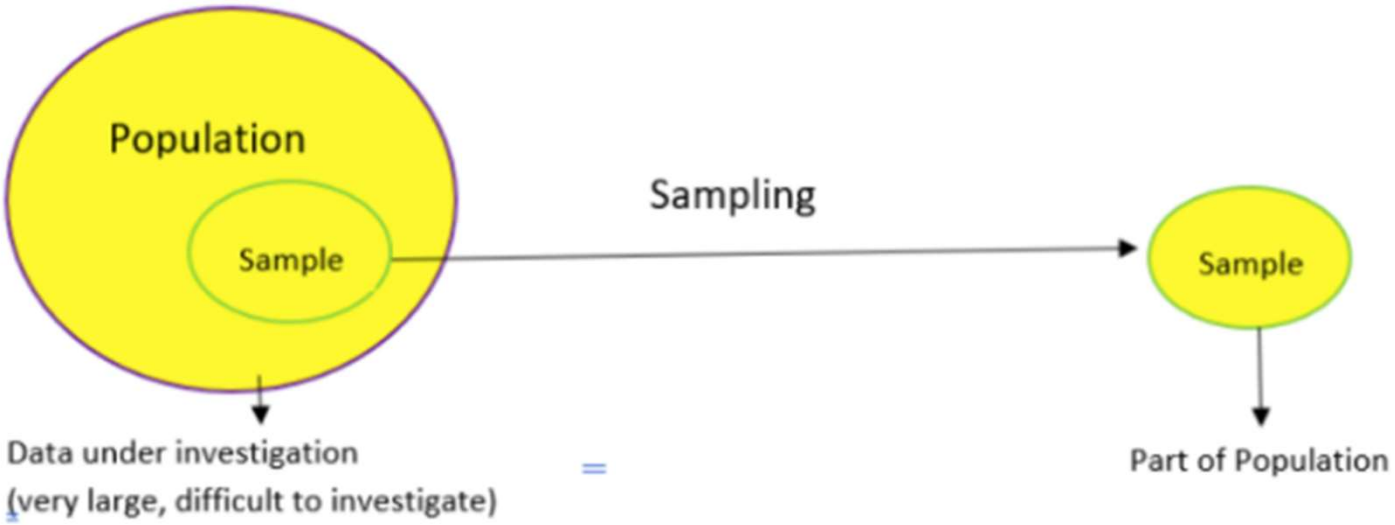


SAMPLING TECHNIQUES

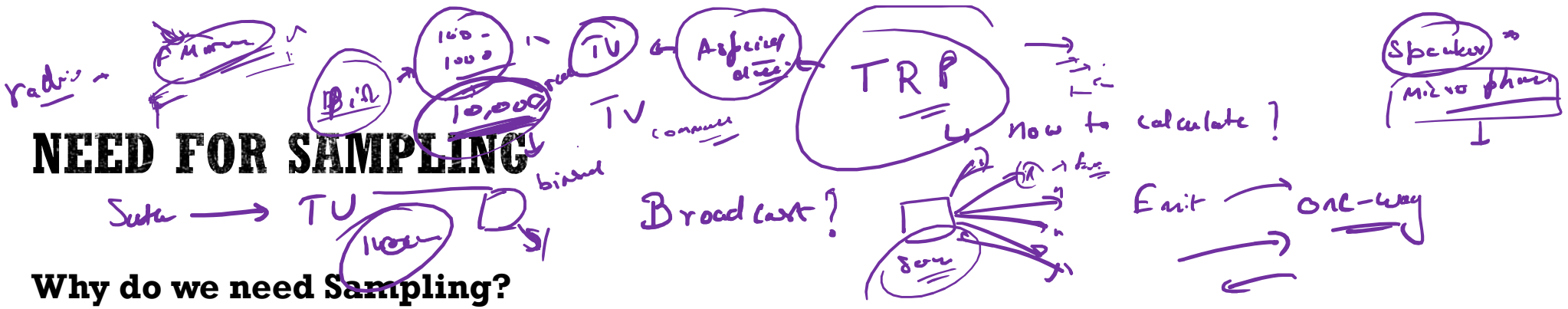
* As a data scientist, you will NOT be using it MUCH

SAMPLING

- *Sampling is a method that allows us to get information about the population based on the statistics from a subset of the population (sample), without having to investigate every individual.*



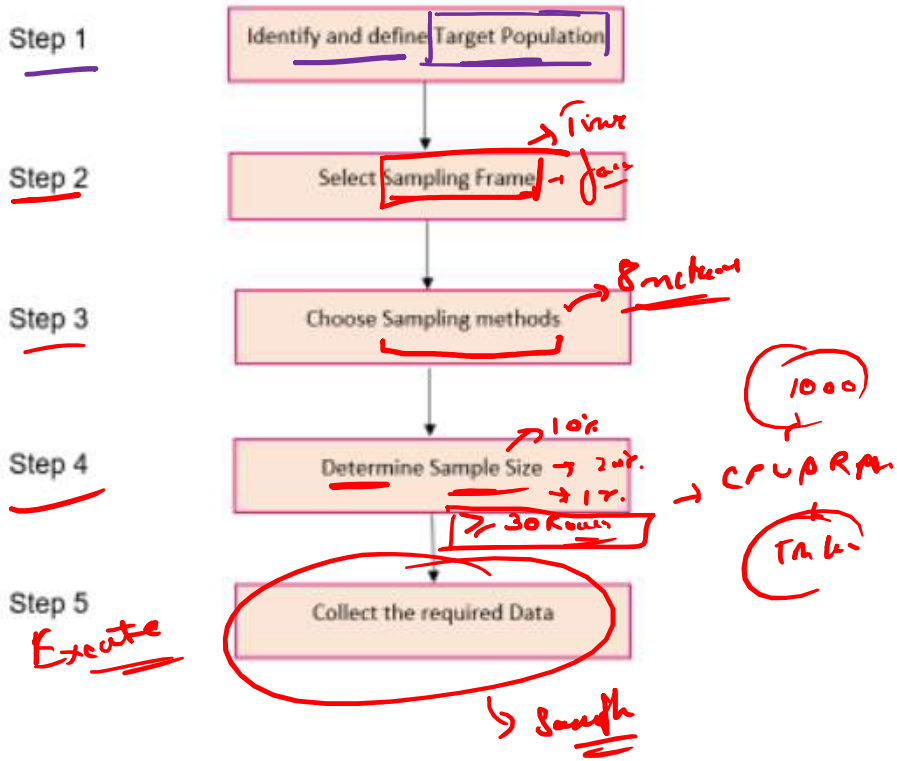
NEED FOR SAMPLING



Why do we need Sampling?

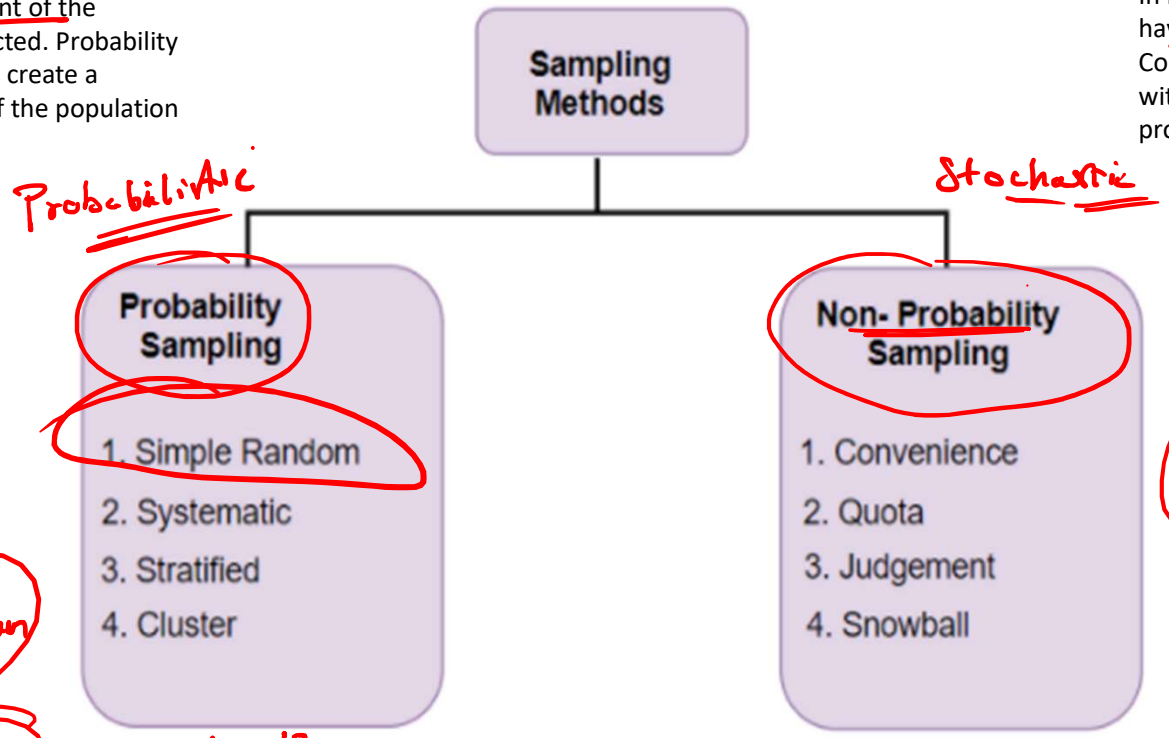
- Sampling is done to draw conclusions about populations from samples, and it enables us to determine a population's characteristics by directly observing only a portion (or sample) of the population.
- Selecting a sample requires less time than selecting every item in a population
- Sample selection is a cost-efficient method
- Analysis of the sample is less cumbersome and more practical than an analysis of the entire population

STEP FOR SAMPLING



TYPES OF SAMPLING

In probability sampling, every element of the population has chance of being selected. Probability sampling gives us the best chance to create a sample that is truly representative of the population

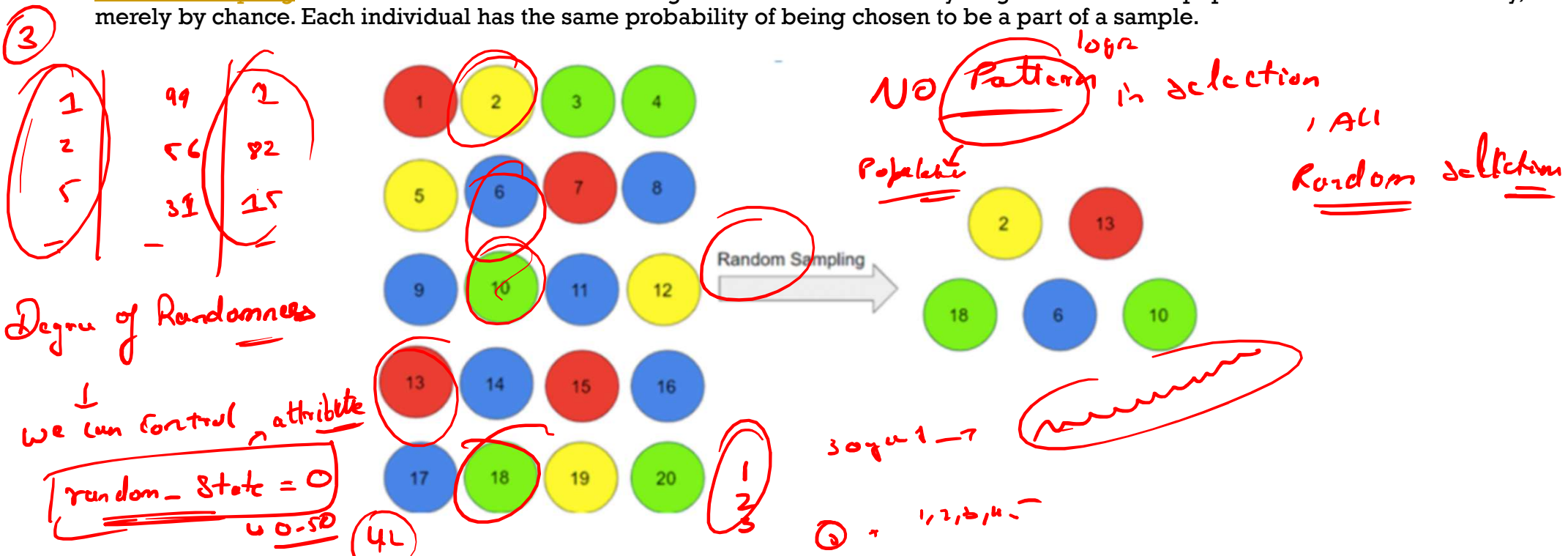


In non-probability sampling, all elements do not have an equal chance of being selected. Consequently, there is a significant risk of ending up with a non-representative sample which does not produce generalizable results

TYPES OF PROBABILITY SAMPLING - SAMPLING METHODS

SIMPLE RANDOM SAMPLING

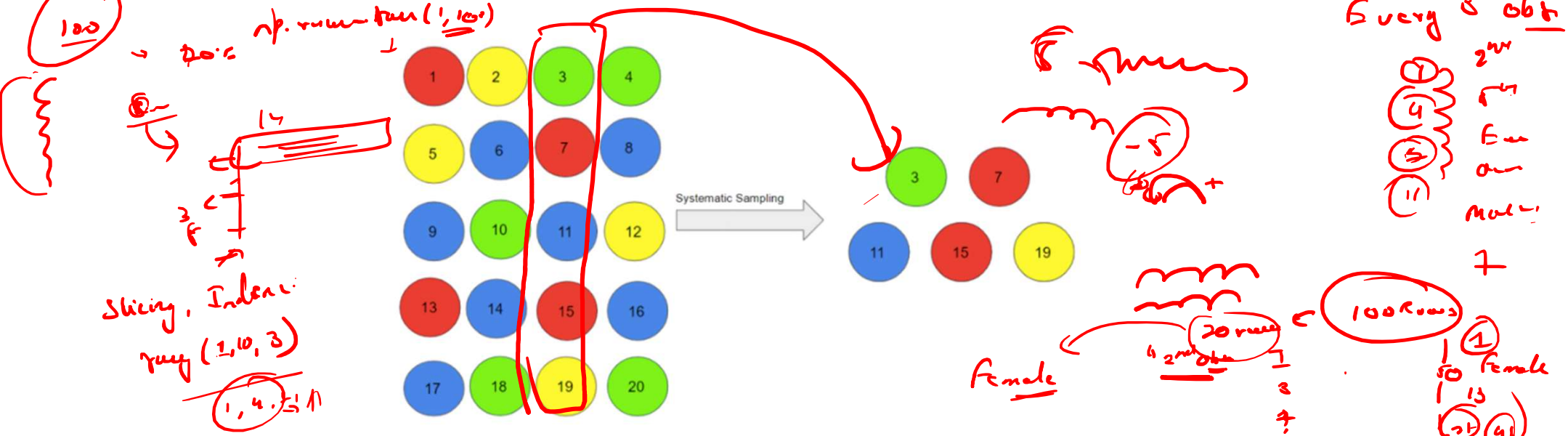
- **Simple random sampling:** One of the best probability sampling techniques that helps in saving time and resources, is the Simple Random Sampling method. It is a method of obtaining information where every single member of a population is chosen randomly, merely by chance. Each individual has the same probability of being chosen to be a part of a sample.



For example, in an organization of 500 employees, if the HR team decides on conducting team building activities, it is highly likely that they would prefer picking chits out of a bowl. In this case, each of the 500 employees has an equal opportunity of being selected.

SYSTEMATIC SAMPLING

- Researchers use the method to choose the sample members of a population at regular intervals. It requires the selection of a starting point for the sample and sample size that can be repeated at regular intervals. This type of sampling method has a predefined range, and hence this sampling technique is the least time-consuming.



- Rule: Say our population size is x and we have to select a sample size of n. Then, the next individual that we will select would be x/nth intervals away from the first individual. We can select the rest in the same way.
- For example, a researcher intends to collect a systematic sample of 500 people in a population of 5000. He/she numbers each element of the population from 1-5000 and will choose every 10th individual to be a part of the sample (Total population/ Sample Size = 5000/500 = 10).
- However, it might also lead to bias if there is an underlying pattern in which we are selecting items from the population. There should be no hidden pattern in the order.