α 400 900

1 2
—— I
2

500 → 50

1:2

1:2

True ref^n
!
Same as
popn
→ Daily

Equal Ref^n
Same # of
obs
1:1
Plus

400
450
500 Pum   900:450

$$P = \frac{1x}{ALL}$$ 0.99
0.97
0.91

Both (Cluster Sampling | Stratified Sampling) → we take a group of

* Organize the population (groups of rows)   obs. from the population at once

divide

Clusters   Stratas
(Heterogenous)   (Homogenous)

1400 rows

50
50 clusters 600   1200
1000

≥ 1
→ multiple obs. at once

1 Apple

1F 2M
1:2

900 male   600 female

100 1:1

Supermarket

Fruits

10 Apple

Fixed Ratio

Equal distribution

50 M

50 ml
50 F M

1:1:1:1
1:2

1M 1F

9:5

→ sample

1M 1F

1M 1F

1M 1F

Cluster Seen

Whole Basket

← Fruit Basket

Fruit stalls / crates →

1:4
1:1
1:10
5:1

1A 16B

check

different types of fruit in one basket

CLUSTER SAMPLING

- **Cluster sampling:** In a clustered sample, we use the subgroups of the population as the sampling unit rather than individuals. The population is divided into subgroups, known as clusters, and a whole cluster is randomly selected to be included in the study:
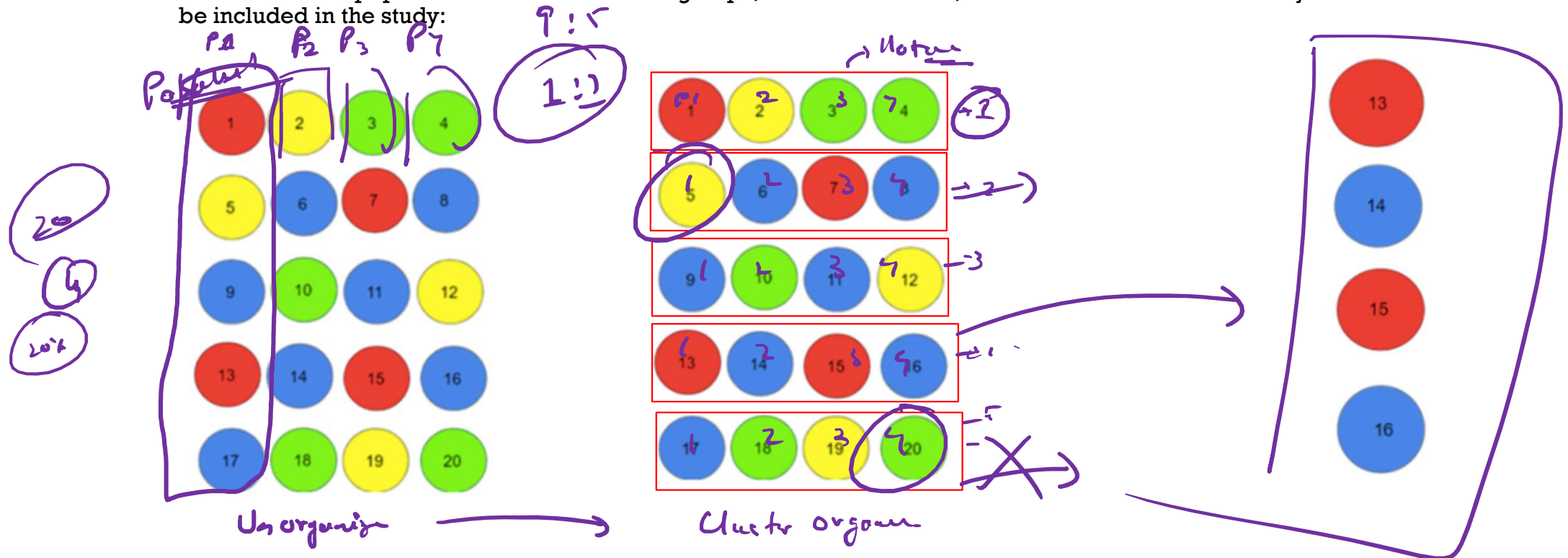

- In the above example, we have divided our population into 5 clusters. Each cluster consists of 4 individuals and we have taken the 4th cluster in our sample. We can include more clusters as per our sample size.

- **This type of sampling is used when we focus on a region or area.**

STRATIFIED SAMPLING

- In this type of sampling, we divide the population into subgroups (called strata) based on different traits like gender, category, etc. And then we select the sample(s) from these subgroups:
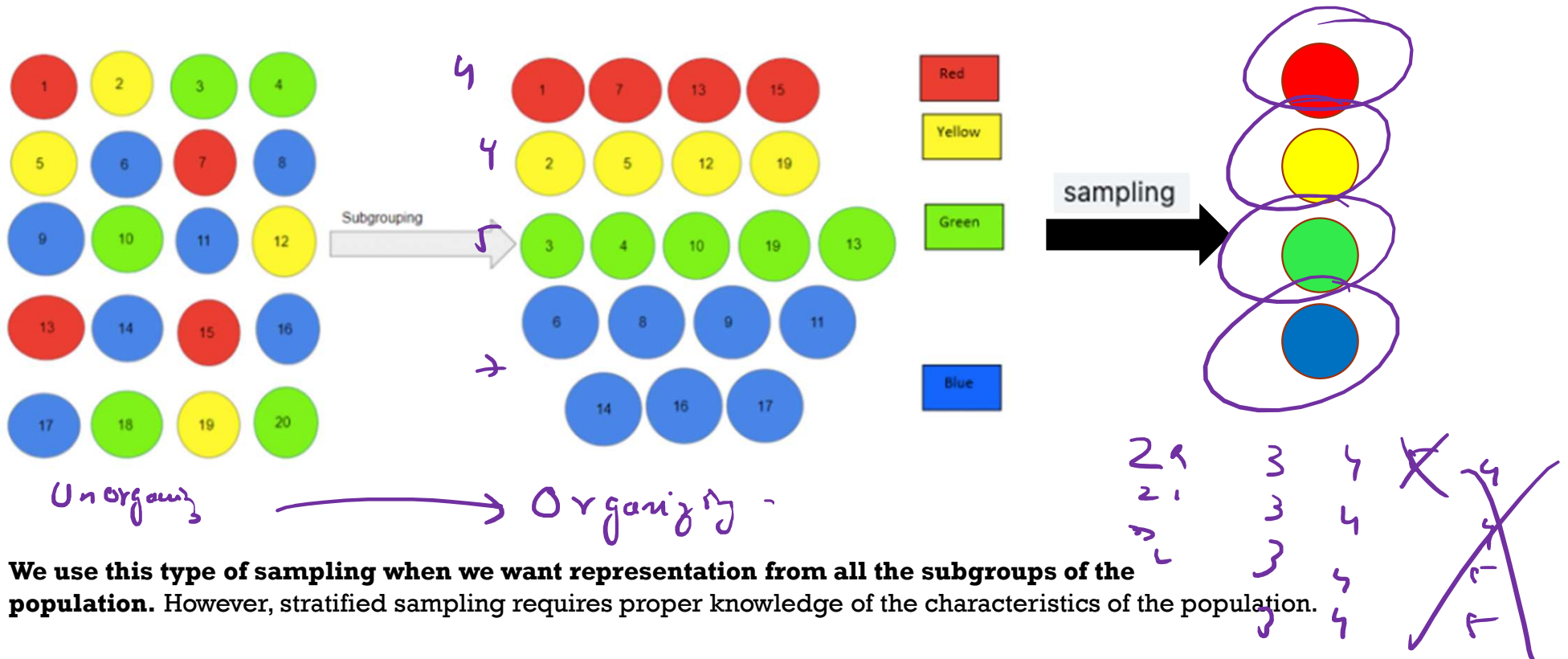
We use this type of sampling when we want representation from all the subgroups of the population. However, stratified sampling requires proper knowledge of the characteristics of the population.

Random → Equal

Systematic → Equal

1000

$\frac{1}{1000}$

100

$(10^{th} \; 1^{st} \; 1^{st})$

Clus → Each cluster
has eq Prob

State → each strata is giving Eq obs.

↳ Population does not Exist
↑
In the Imaginary Population
↓
unequal chances
of

Stochastic
non-prob.

BIASED
ALWAYS!!!

biased

Always n → No other options

selection
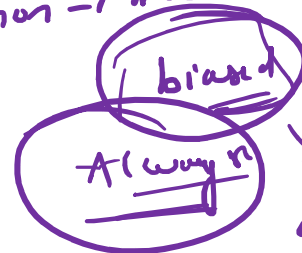
NON PROBABILITY SAMPLING
↳ mostly used

- **Convenience sampling**: This is perhaps the easiest method of sampling because individuals are selected based on their availability and willingness to take part.

- Here, let's say individuals numbered 4, 7, 12, 15 and 20 want to be part of our sample, and hence, we will include them in the sample.
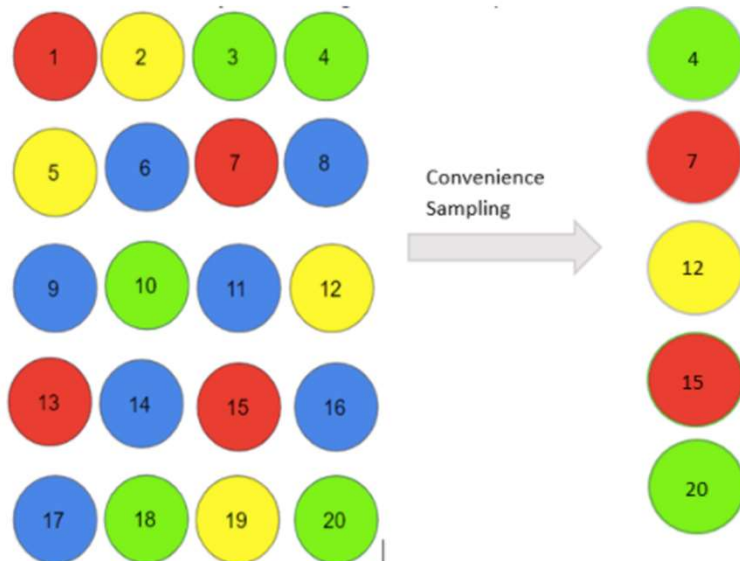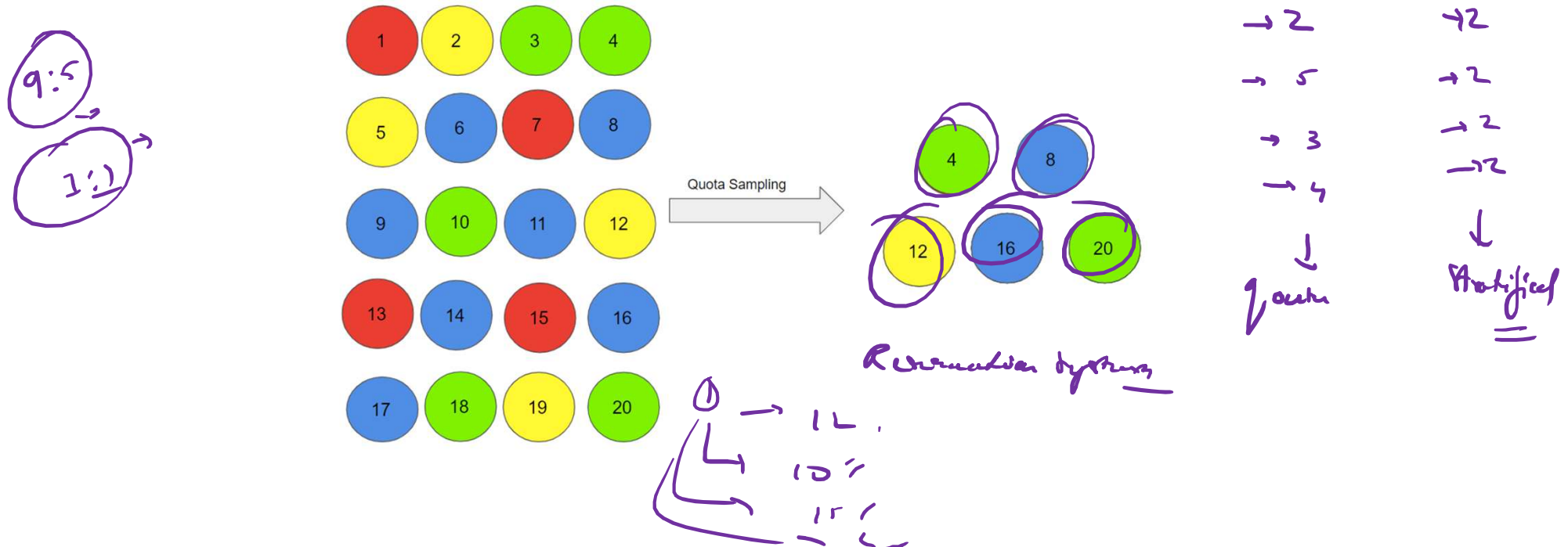

- For example, start-ups and NGOs usually conduct convenience sampling at a mall to distribute leaflets of upcoming events or promotion of a cause – they do that by standing at the mall entrance and giving out pamphlets randomly.

- Convenience sampling is prone to significant bias, because the sample may not be the representation of the specific characteristics such as religion or, say the gender, of the population.

QUOTA SAMPLING

Stratified sampling but with unequal distr'n

- **Quota sampling:** In this type of sampling, we choose items based on predetermined characteristics of the population. For example you could divide the population into strata and then select from each strata based on Quota. Consider that we have to select individuals having a number in multiples of four for our sample:

Quota Sampling

$9:5$

$1:1$

20 L
(3 or

→ 2   →2
→ 5   →2
→ 3   →2
→ 4   →2
↓   ↓
youth   Stratified

Reservation system

→ 12
→ 10%
→ 15%

- Therefore, the individuals numbered 4, 8, 12, 16, and 20 are already reserved for our sample.
- In quota sampling, the chosen sample might not be the best representation of the characteristics of the population that weren't considered.
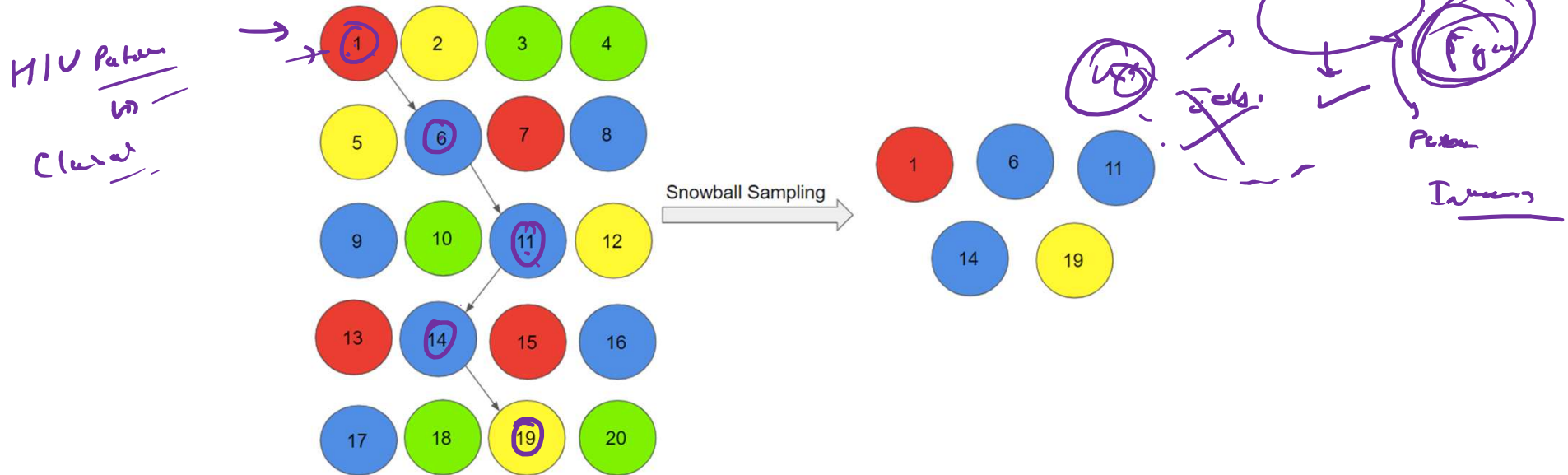
SNOWBALL SAMPLING

(handwritten: Snowball Effect)

- In this technique **Existing people are asked to nominate further people known to them so that the sample increases in size like a rolling snowball.** This method of sampling is effective when a sampling frame is difficult to identify.

(handwritten: HIV Patient in Cluster)

Snowball Sampling

- Here, we had randomly chosen person 1 for our sample, and then he/she recommended person 6, and person 6 recommended person 11, and so on.

1->6->11->14->19

- There is a significant risk of selection bias in snowball sampling, as the referenced individuals will share common traits with the person who recommends them.

JUDGMENT SAMPLING

- It is also known as selective sampling. It depends on the judgment of the experts when choosing whom to ask to participate.
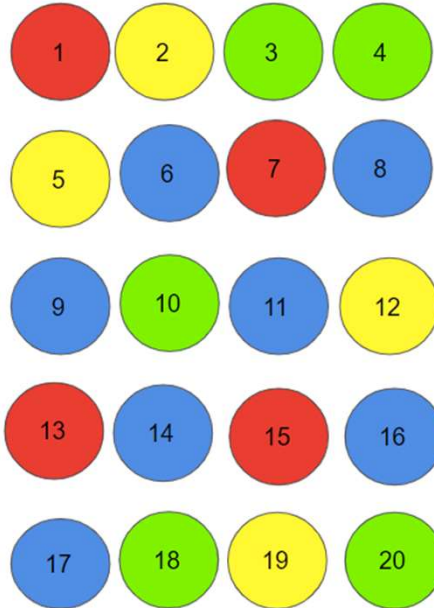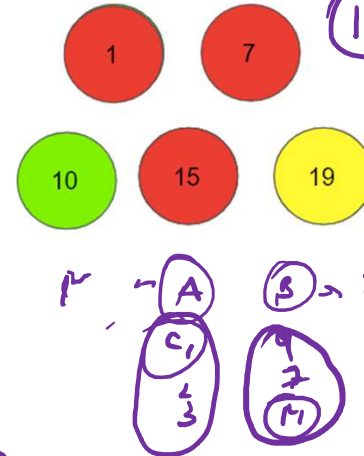

- Suppose, our experts believe that people numbered 1, 7, 10, 15, and 19 should be considered for our sample as they may help us to infer the population in a better way. As you can imagine, quota sampling is also prone to bias by the experts and may not necessarily be representative.