

sampling & sources of bias

- ▶ census vs. sample
- ▶ sources of bias
- ▶ sampling methods



Dr. Mine Çetinkaya-Rundel
Duke University

census

Wouldn't it be better to just include everyone and "sample" the entire population, i.e. conduct a [census](#)?

- ▶ Some individuals are hard to locate or measure, and these people be different from the rest of the population.
- ▶ Populations rarely stand still.

Illegal Immigrants Reluctant To Fill Out Census Form

by PETER O'DOWD

March 31, 2010 4:00 AM



There is an effort underway to make sure Hispanics are accurately counted in the 2010 Census. Phoenix has some of the country's "hardest-to-count" districts. Some Latinos, especially illegal residents, fear that participating in the count will expose them to immigration raids or government harassment.



inference

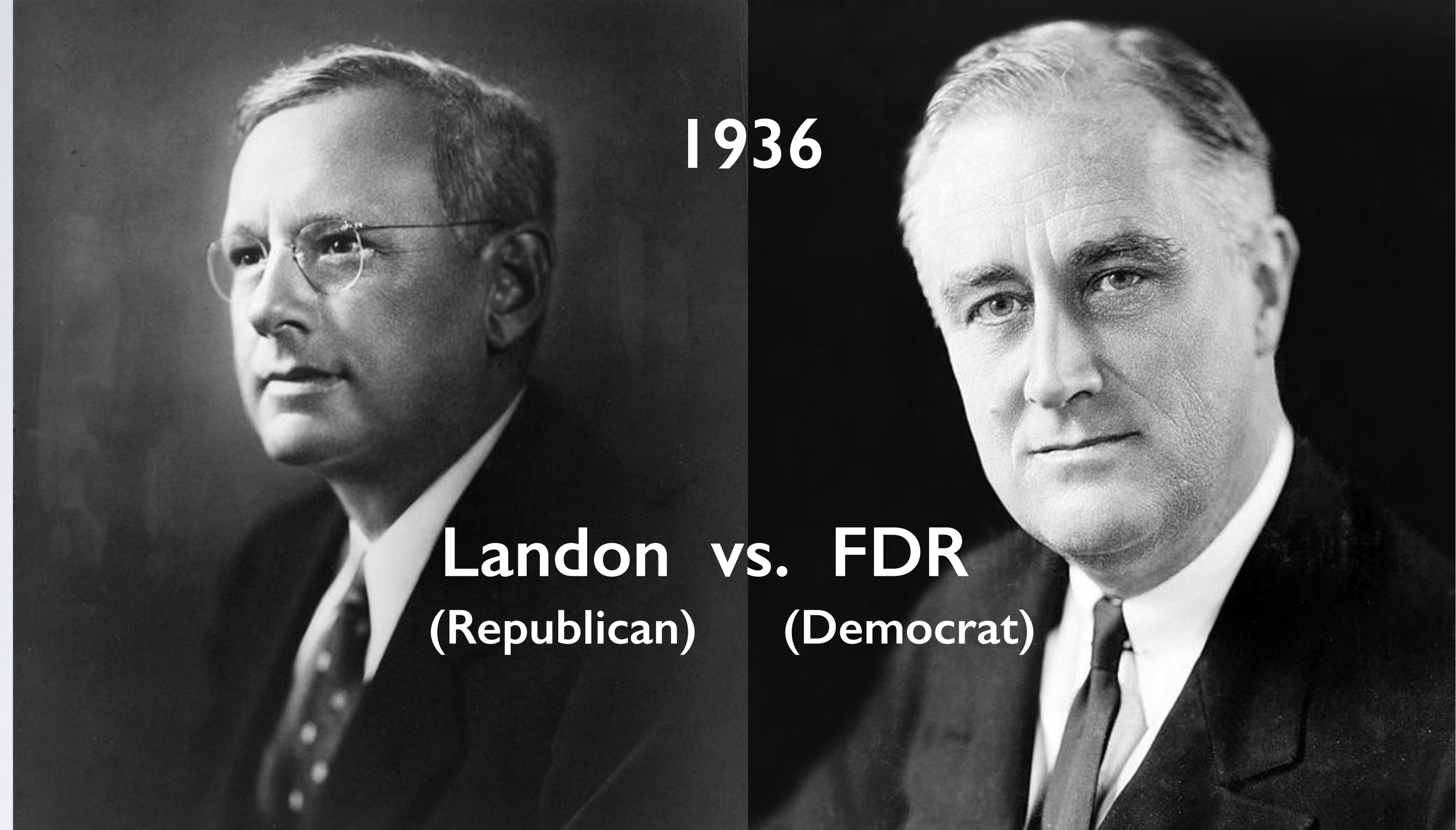
representative
sample

exploratory
analysis

a few sources of sampling bias

- ▶ **Convenience sample:** Individuals who are easily accessible are more likely to be included in the sample
- ▶ **Non-response:** If only a (non-random) fraction of the randomly sampled people respond to a survey such that the sample is no longer representative of the population
- ▶ **Voluntary response:** Occurs when the sample consists of people who volunteer to respond because they have strong opinions on the issue





The Literary Digest

Election results

Lose with 43% of the votes

Win with 62% of the votes



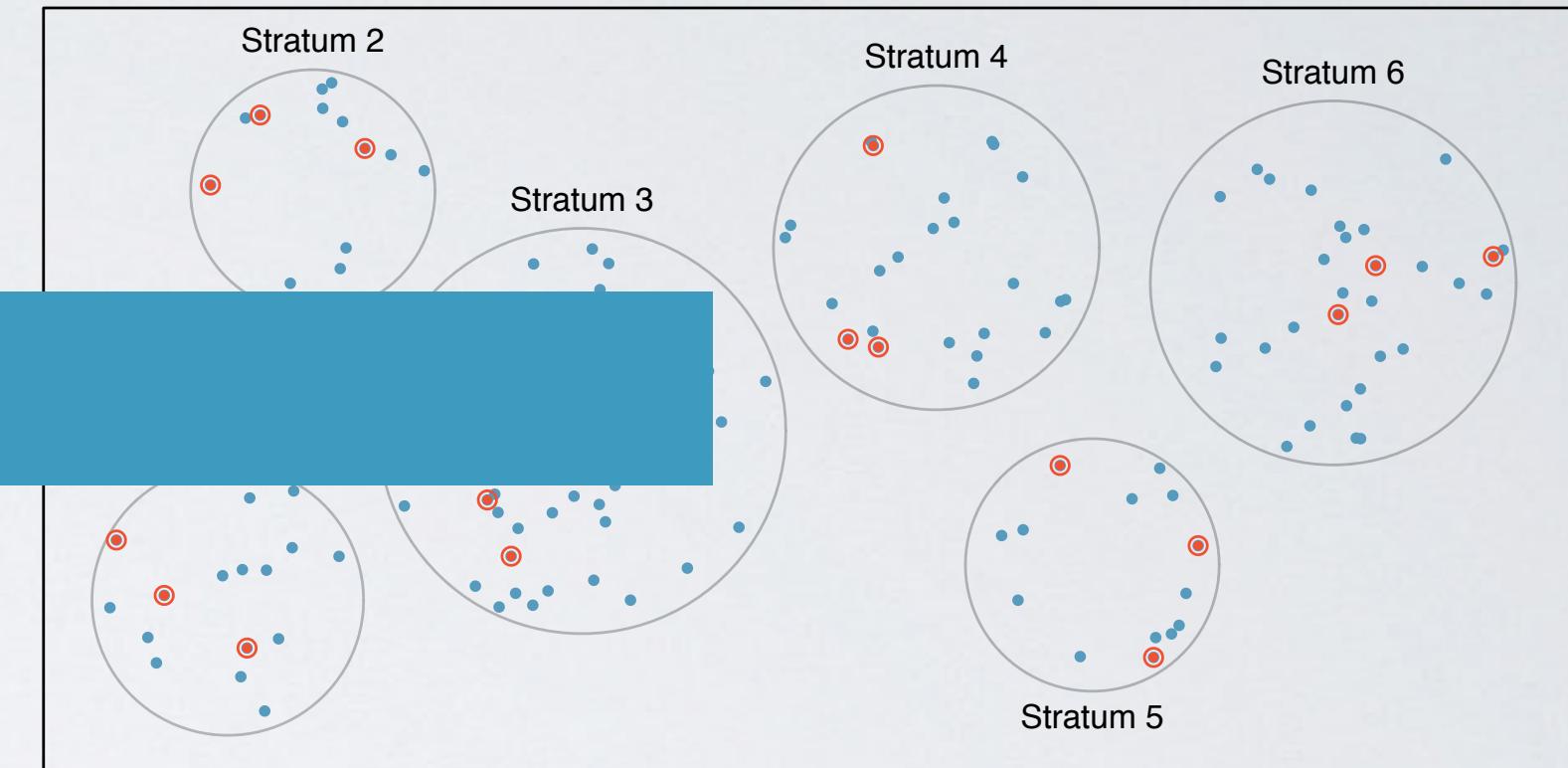
Image credit: Wonderlane CC BY 2.0 <http://www.flickr.com/photos/wonderlane/6231888661>

sampling methods

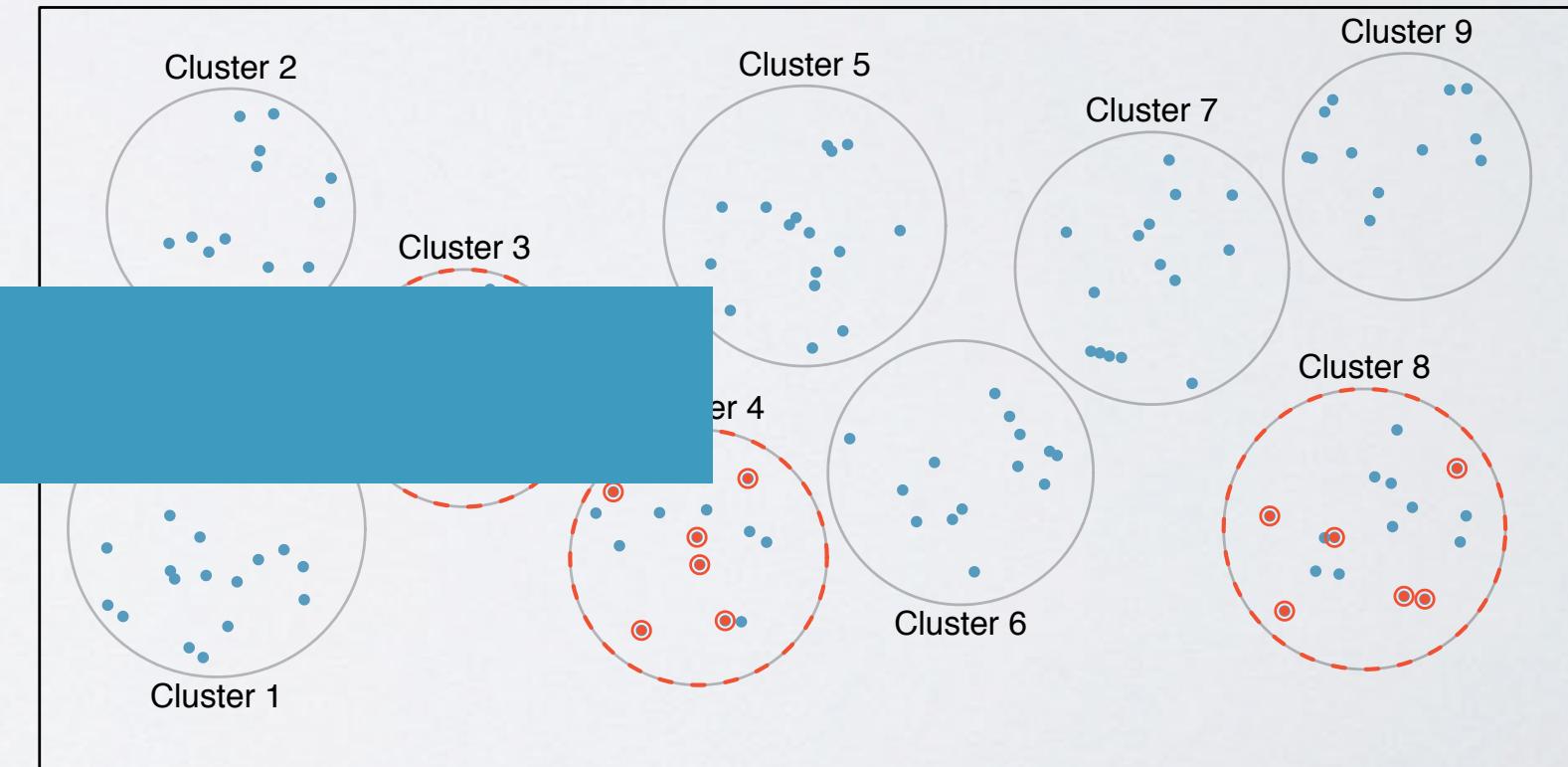
simple random sample (SRS)



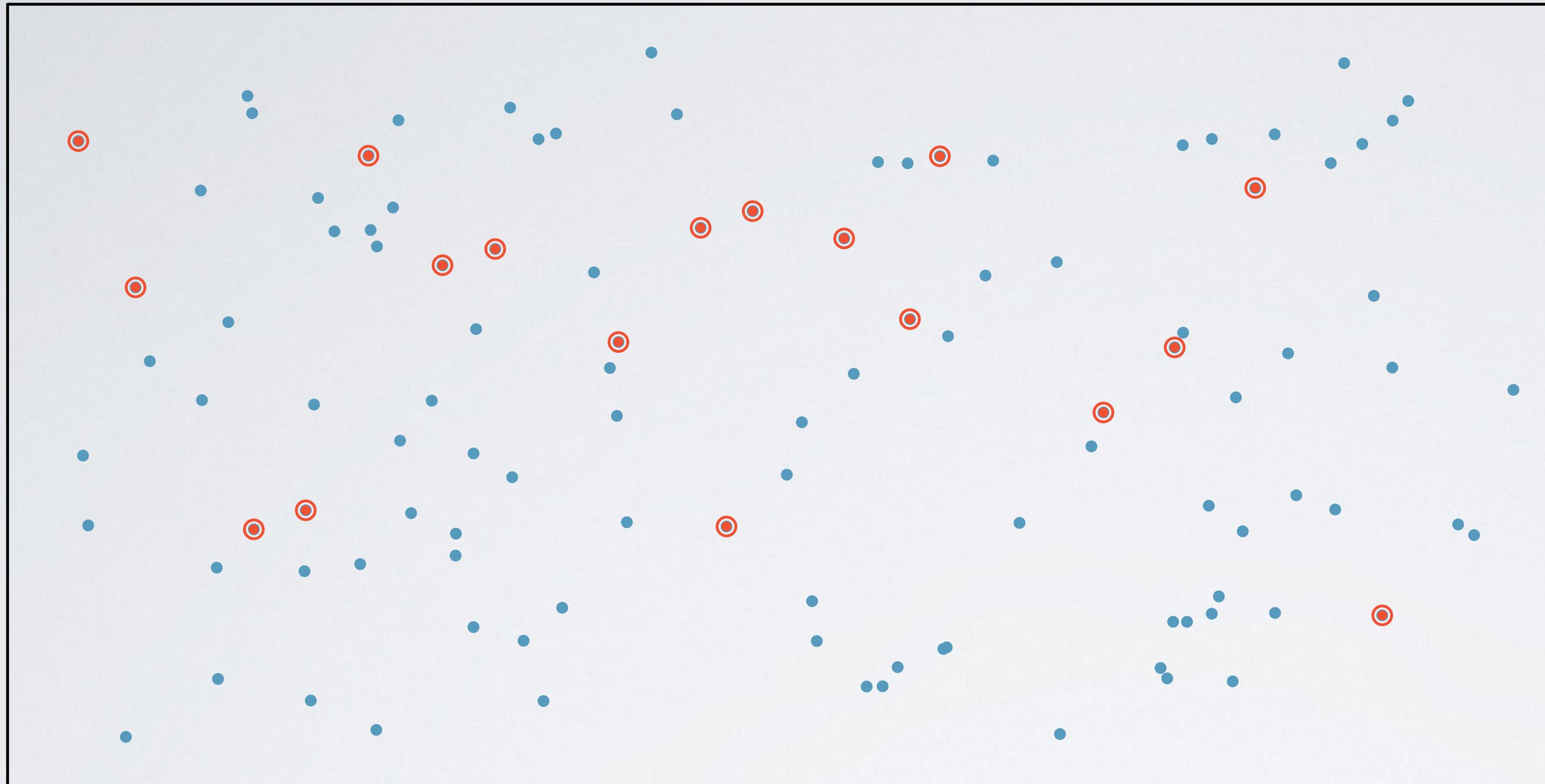
stratified sample



cluster sample

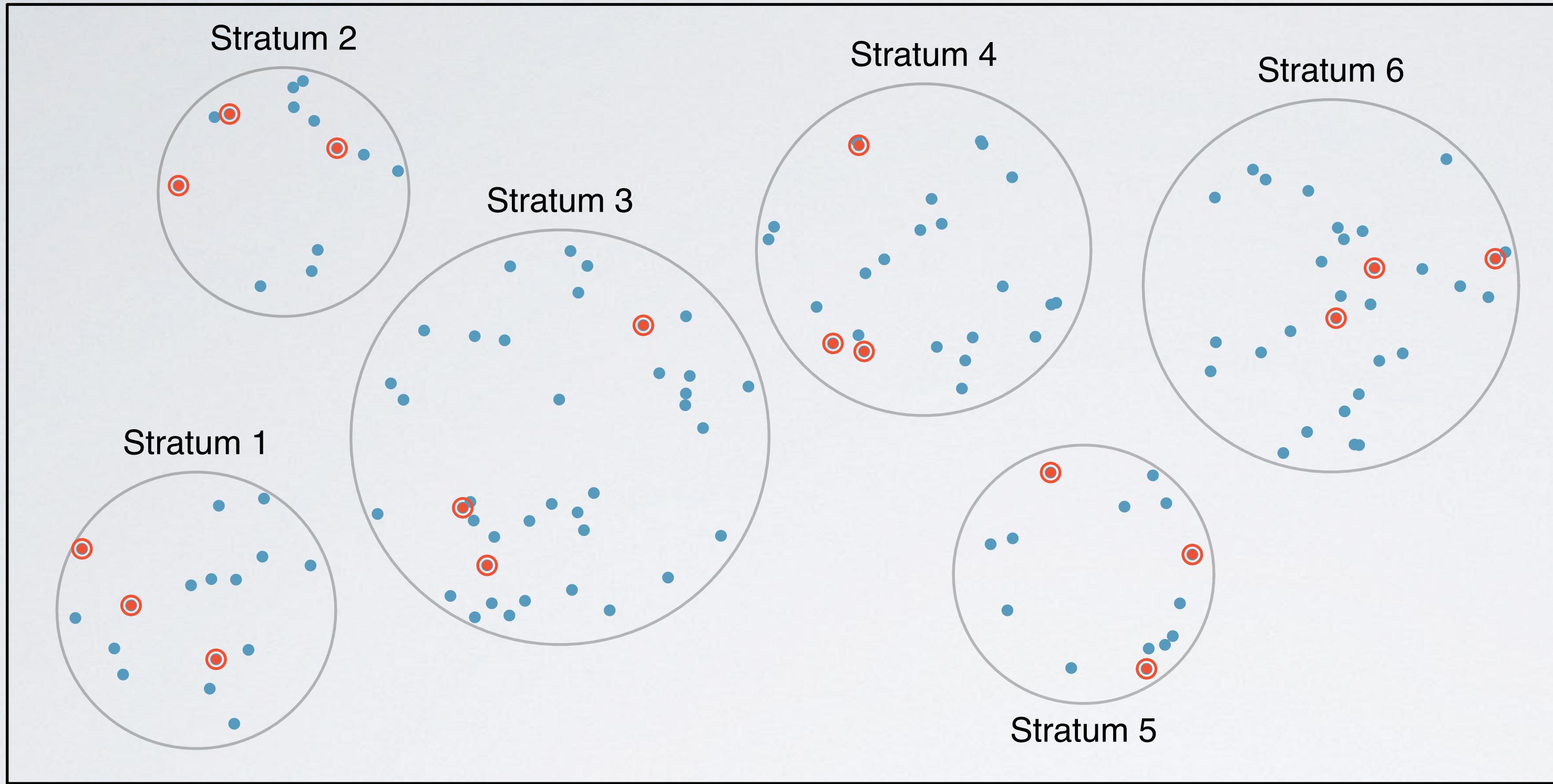


simple random sample (SRS)



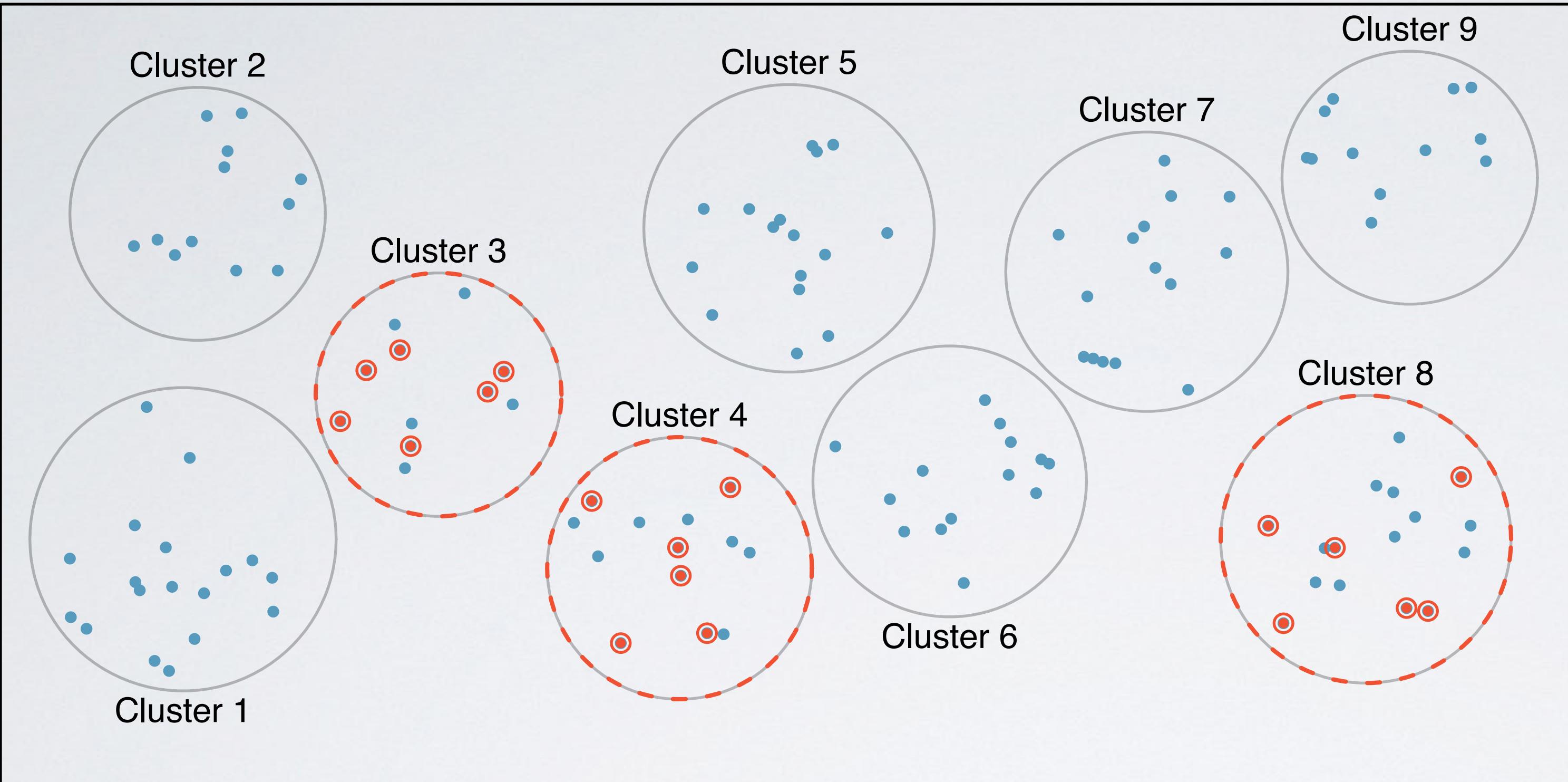
each case is equally likely to be selected

stratified sample



divide the population into homogenous **strata**, then
randomly sample from within each stratum

cluster sample



divide the population **clusters**, randomly sample a few clusters, then randomly sample from within these clusters