Sampling

Math 122 - Introduction to Statistics

Recall

- What is statistics?
- Rare event rule
- Observational studies vs. Experiments
- Categorical vs. Quantitative data
- Continuous vs. Discrete data
- Population vs. Sample
- Statistic vs. Parameter

With proper sampling,
a statistic from a sample
may be a good approximation
of the corresponding population parameter.

Sampling

Simple Random Sample

- A SIMPLE RANDOM SAMPLE is a sample which is collected in such a way that every possible sample of the same size has the same chance of being selected.
- This is the IDEAL.
- Most statistical tools require a simple random sample. But...

- SIMPLE RANDOM SAMPLE - all samples equally likely

- To select a simple random sample of 50 students at CU, I could randomly select names from a directory.

Simple Random Sample

Suppose that I find a sample of Concordia students by selecting 10 Freshmen, 10 Sophomores, 10 Juniors, 10 Seniors, and 10 Super Seniors.

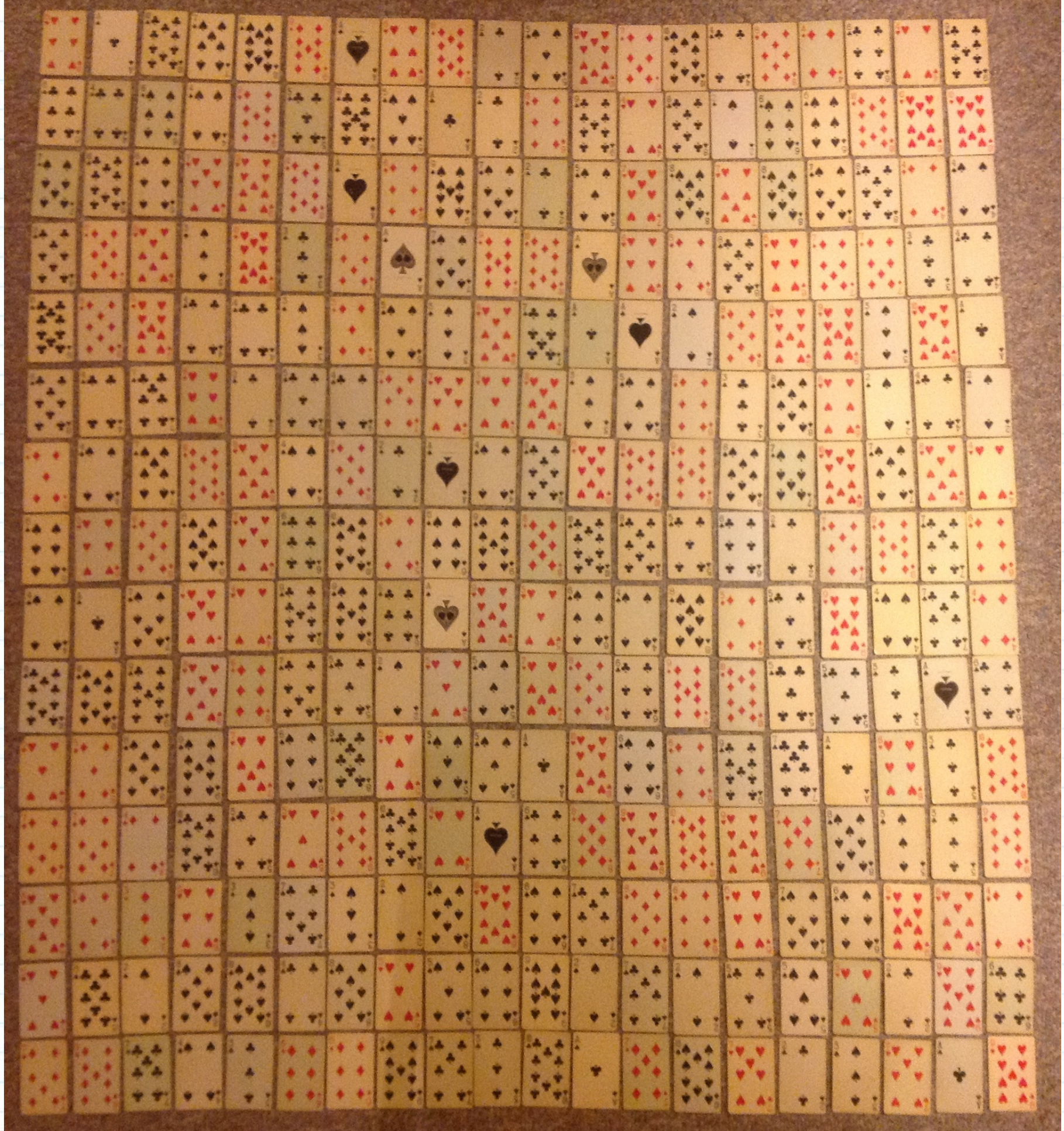
Why is this NOT a simple random sample of size 50?

Simple Random Sample

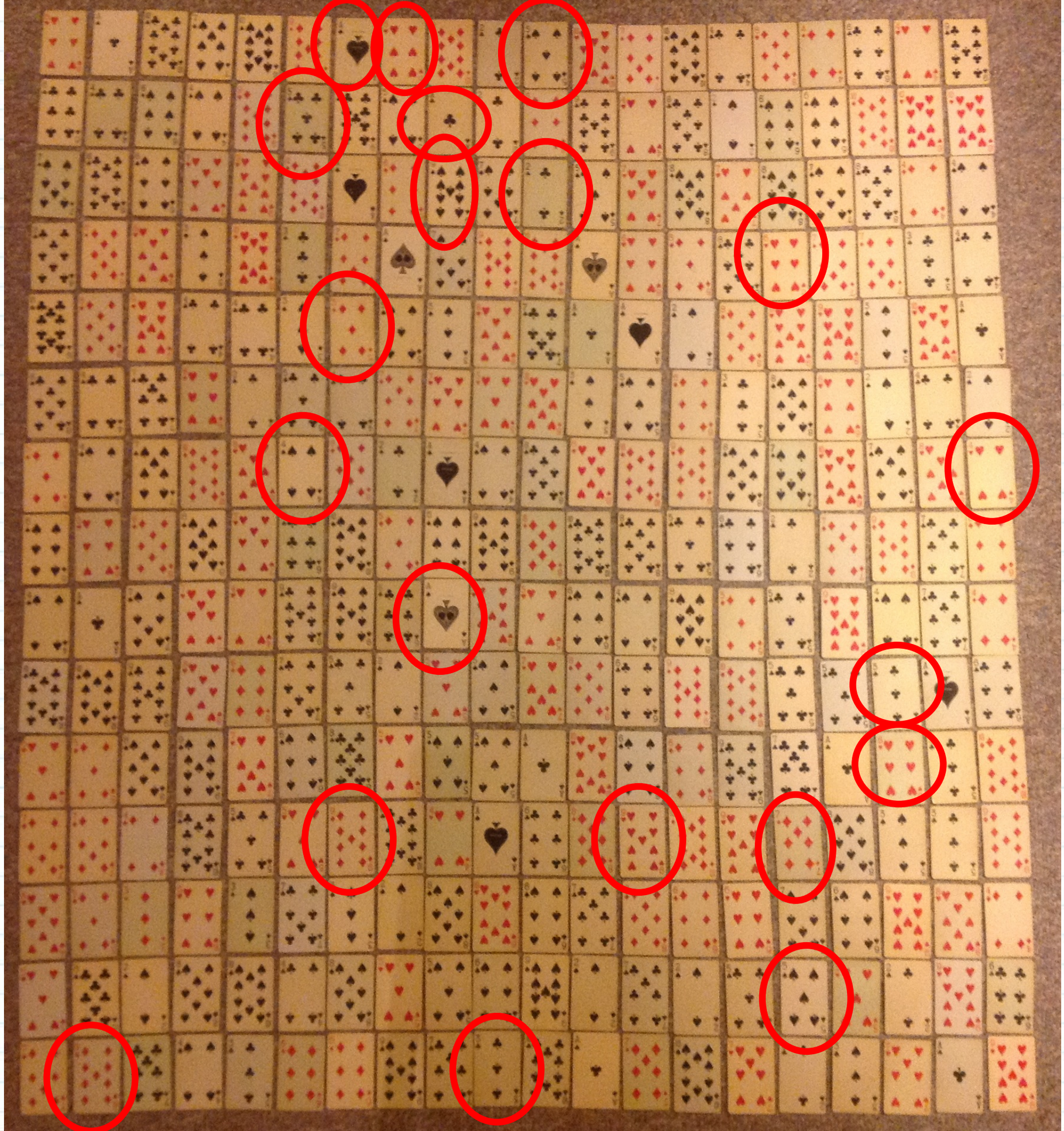
According to Gallop,

- 43% of the US identifies with the Democratic Party
- 39% of the US identifies with the Republican Party
- 18% of the US does not identify with either of these parties.

Suppose that I collect 43 Democrats, 39 Republicans, and 18 others. Why is this NOT a simple random sample of size 100?



I generated random row and column numbers from Excel and circled the corresponding cards.



Random Sample

- A RANDOM SAMPLE is a sample which is selected in such a way that every individual member of the population has an equal chance of being selected.

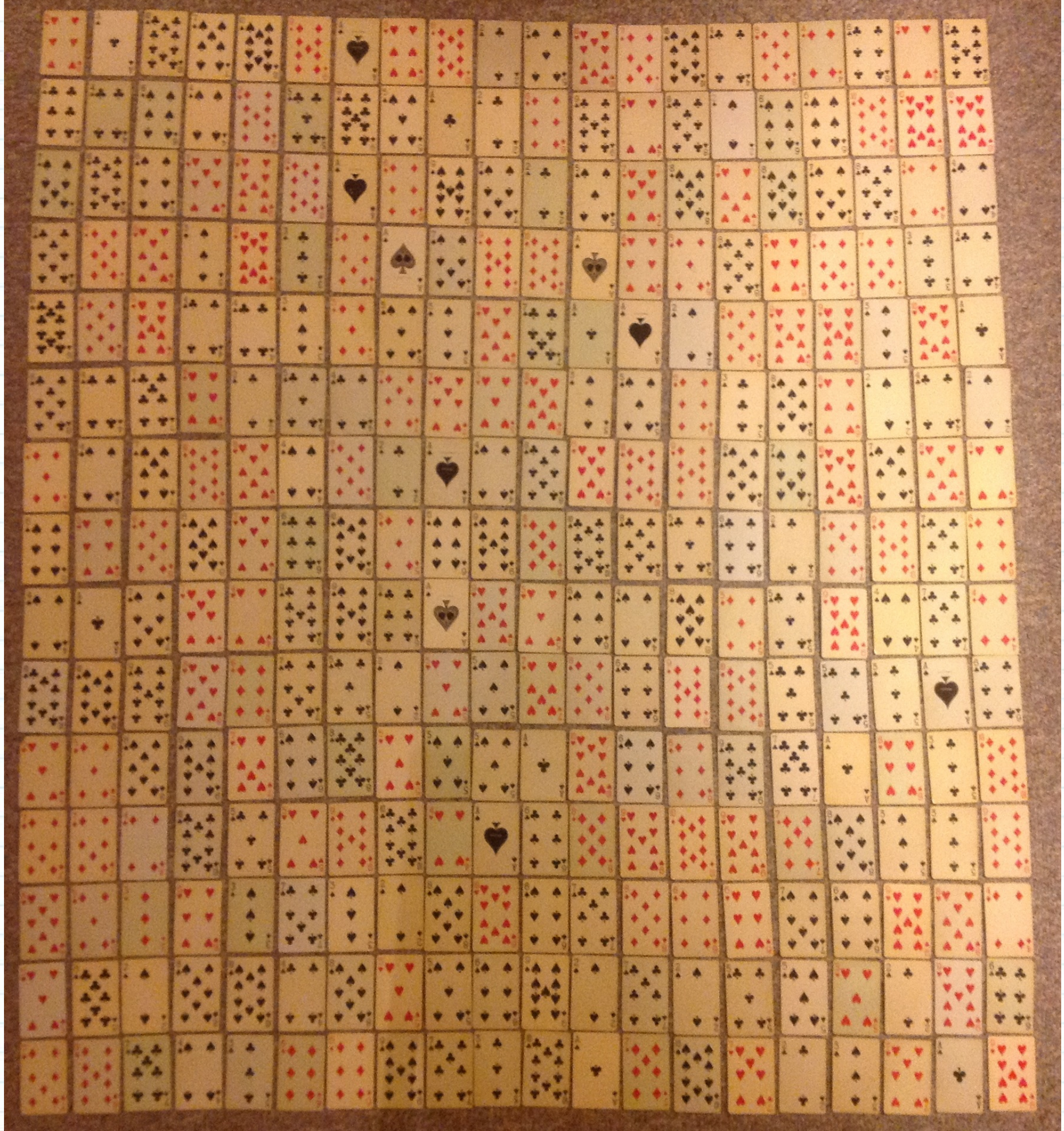
- RANDOM SAMPLE - all individuals equally likely
- To select a random sample (not simple) of 50 students at CU, I could randomly select 5 dorm floors and then randomly select 10 people from each floor.
- Everyone may be equally likely to be selected, but some combinations cannot happen.

Some types of sampling

- Convenience sampling
- Systematic sampling
- Stratified sampling
- Cluster sampling

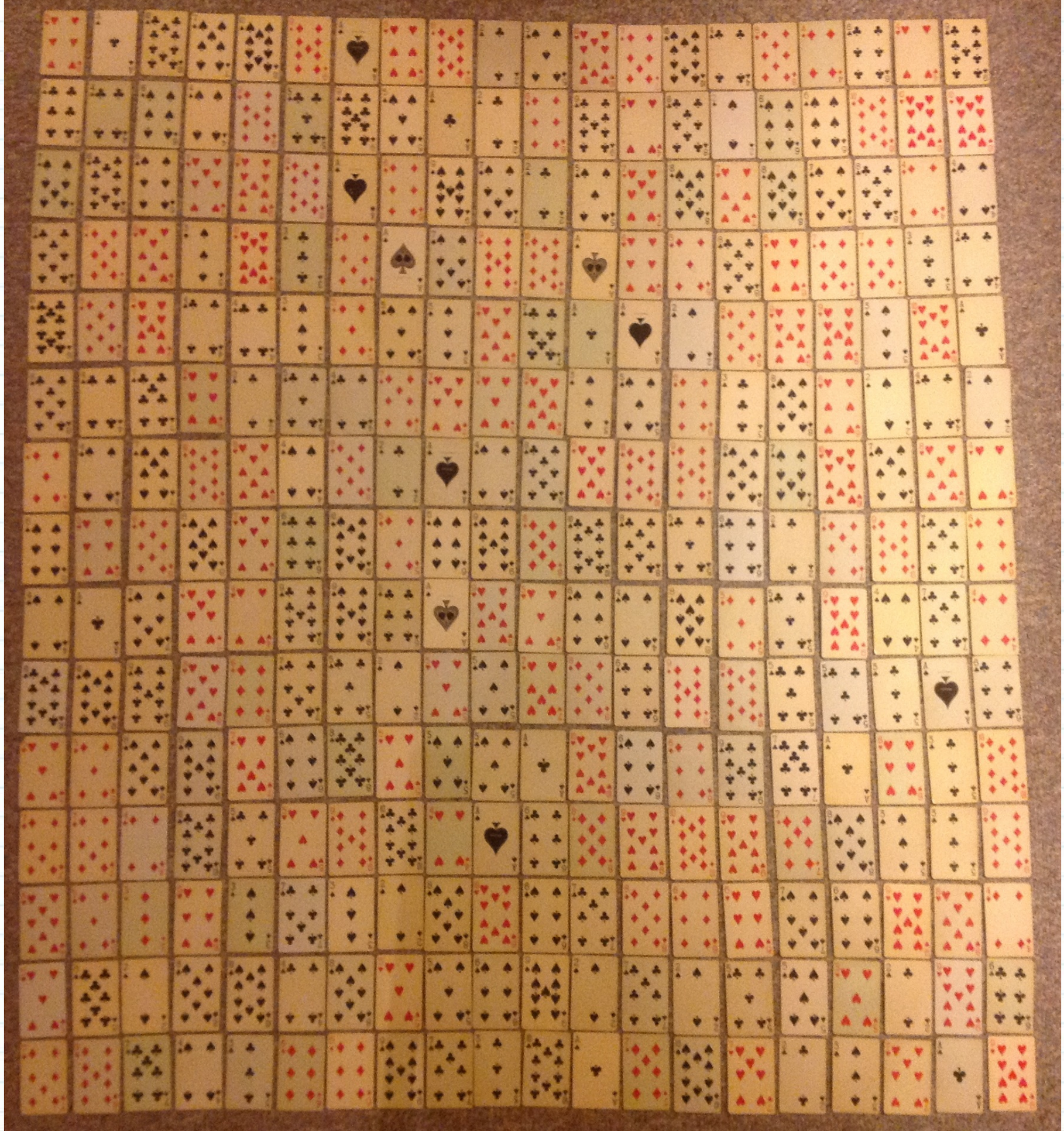
Convenience Sampling

Use data that is easily accessible



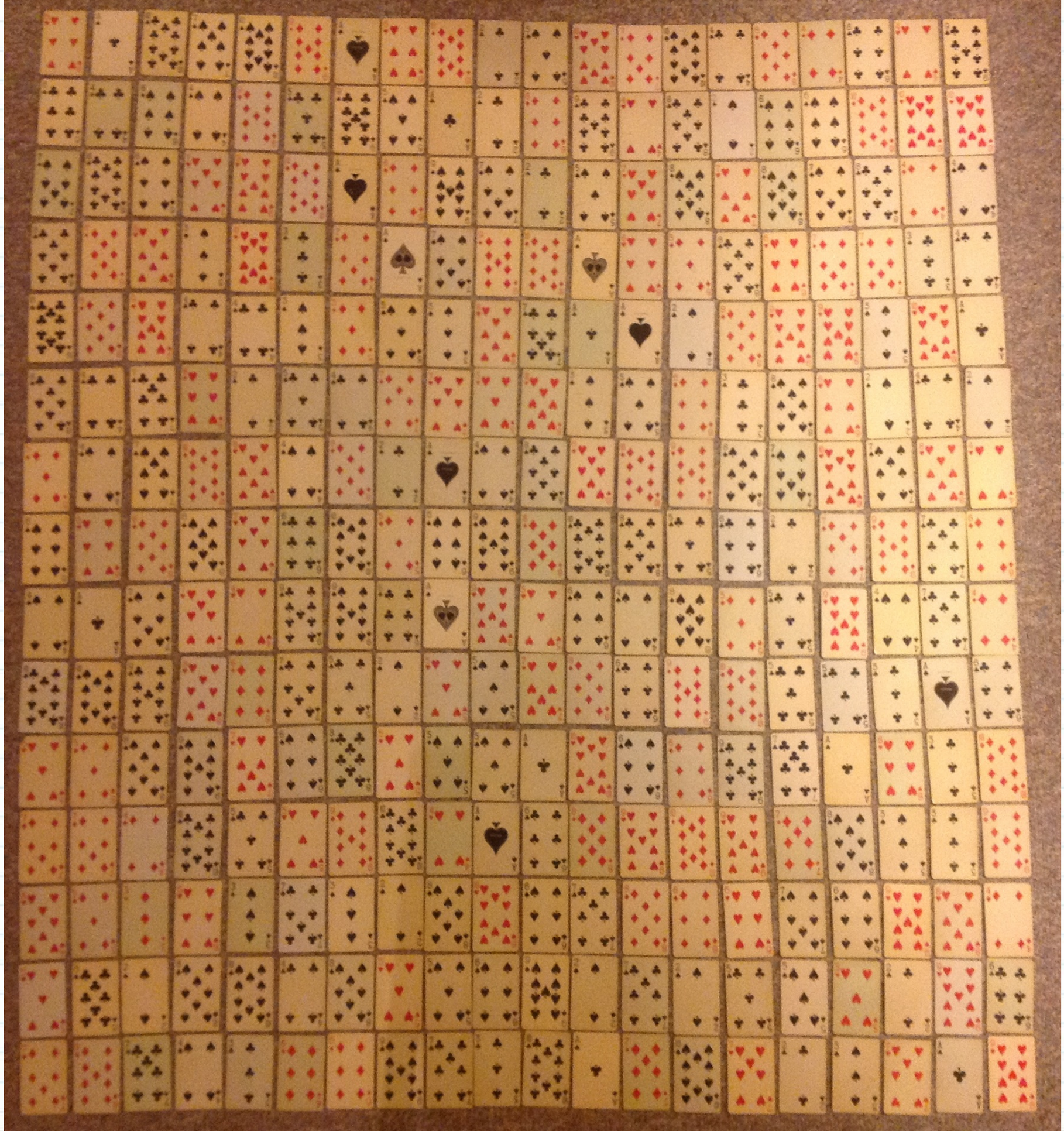
Systematic Sampling

- Arrange the population in a list and select subjects at regular intervals.



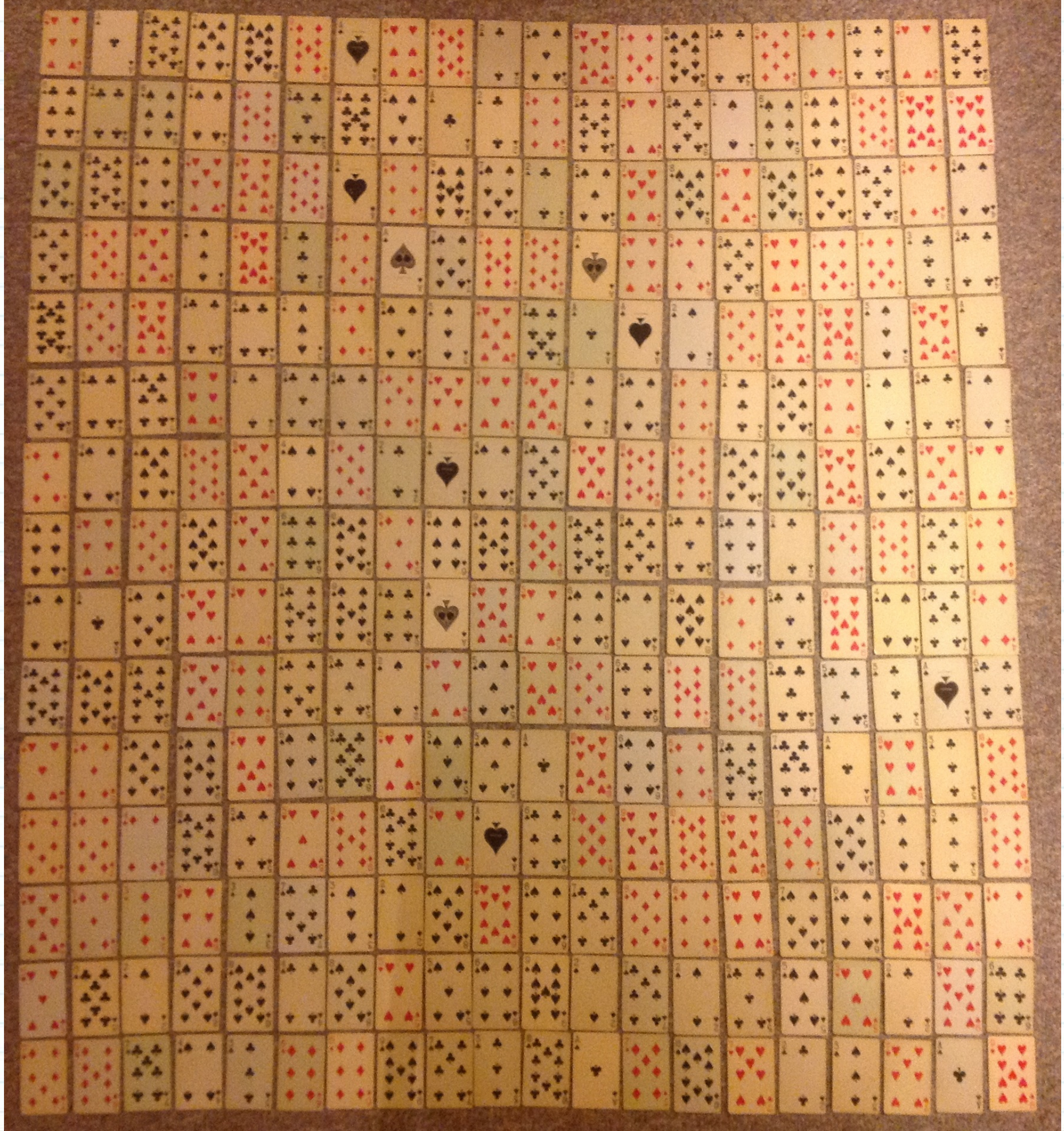
Stratified Sampling

- Divide the population into levels or strata
- Select subjects randomly from each level



Cluster Sampling

- Divide the population into levels or clusters
- Select an entire level or cluster



How might I find a sample of Concordia students using…

- Convenience Sampling?
- Stratified Sampling?
- Cluster Sampling?
- Systematic Sampling?

Some things to consider

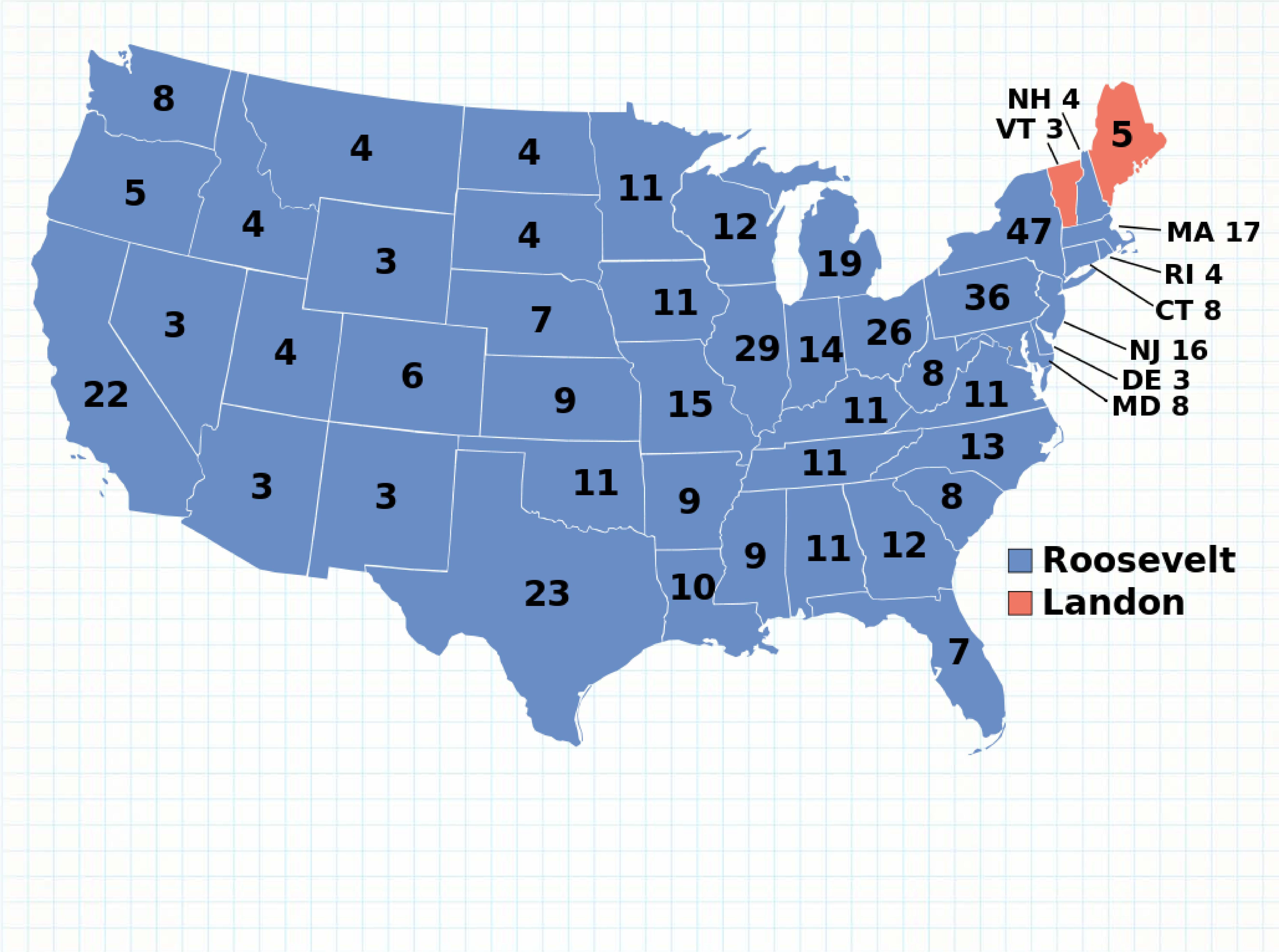
- Sampling
- Source
- Context
- Conclusions
- Practicality

Sampling

- Bad sampling techniques render statistics useless.
  - Self-selected/Voluntary Response
  - Small sample
  - Non-representative sample
  - Missing or corrupted data

1936 Presidential Election

- *Literary Digest* performed a self-selected survey with 2.3 million responses that predicted Alf Landon would win.
- Gallup polled 50,000 people and concluded FDR would win.


8
4
4
11
NH 4
VT 3
5
5
4
12
3
4
19
47
MA 17
RI 4
3
7
11
36
CT 8
4
6
29
14
26
NJ 16
DE 3
MD 8
22
9
15
8
11
11
13
3
3
11
9
11
8
9
11
12
23
10
7
Roosevelt
Landon

Small Sample

- The Children's Defense Fund published *Children out of School in America* in which it reported that 67% of the children in one region had been suspended from school at least 3 times.
- Their data was based on a sample of size 3.

Reported Results

- It is always better to measure yourself than to ask subjects to report results.
  - In a recent poll of 1000 eligible voters, 70% claim to have voted in the last election.
  - Actually, only 61% voted.

Source

- Is the source biased?
- What is the motivation?
- Has the data been corrupted?

Benefits of Caffeine

- Increases memory
- Stimulates hair growth
- Relieves post-workout muscle pain
- Wards off Alzheimer's
- Lowers risk of suicide
- Improves reaction time and logical reasoning
- Reduces asthma symptoms
- May prevent weight gain

Who do you think pays for these studies?

Scuffed Shoes

"According to a nationwide survey of 250 hiring professionals, scuffed shoes was the most common reason for a male job seeker's failure to make a good first impression."

Financed by…Kiwi Brands, who make shoe polish.

Context

- What does the data represent?
- What type of data is it?
- Determines what calculations and tests make sense.

Conclusions

- Confusing correlation and causation
- Precise numbers do not mean that the study was precise.
- Some conclusions are simply impossible to support with statistics.

Causation vs. Correlation

A recent study concluded that people who die young typically slept more as children.

In response, some doctors directed parents not to let their children sleep too much.

Causation vs. Correlation

My brother's pediatrician discouraged him from letting my nephew participate in gymnastics because male gymnasts are typically shorter than average.

Practicality

- Results can be statistically significant without being practically significant.

Weight Watchers

In a study printed in the *JAMA*, 40 subjects were on the Weight Watchers diet for a year. They lost an average of 3lb with a standard deviation of 4.9lb.

The weight loss is *statistically significant* but…