Equitable Equations: *Introduction to statistical sampling*

Problem 1

Identify the sampling technique used in each of the following. If applicable, identify possible sources of sampling bias.

(a) Following an earthquake in San Francisco, a relief agency selects forty city blocks at random and then sends agents door-to-door to assess damage.

This uses the cluster sampling to collect data. It vulnerable to non-response bias.

(b) A corn field is divided into 25 one-acre plots and a random sample is taken from each one to estimate crop yield.

This is stratified sampling. There are no obvious sources of bias.

(c) From an alphabetical list of customers, a car dealership sends satisfaction surveys to every tenth person.

This uses systematic sampling but the alphabetical list of customers makes it vulnerable to sampling bias. Nonresponse bias as well.

(d) A student gathers data on satisfaction with a school's cafeteria by surveying all of their friends and carefully recording answers in a numerical spreadsheet.

This is convenience sampling which is not a great method and very vulnerable to sampling bias

(e) At a basketball game, ten seats are chosen at random and the occupants given prizes.

This is simple random sampling with no obvious sources of bias. Empty seats - nonresponse(?)

Problem 2

The following problem is taken from *OpenIntro Statistics*, Fourth Edition, by David Diez, Mine Çetinkaya-Rundel, and Christopher Barr. Pay what you want or download for free at `https://www.openintro.org/book/os/`

1.23 Evaluate sampling methods. A university wants to determine what fraction of its undergraduate student body support a new $25 annual fee to improve the student union. For each proposed method below, indicate whether the method is reasonable or not.

(a) Survey a simple random sample of 500 students.

(b) Stratify students by their field of study, then sample 10% of students from each stratum.

(c) Cluster students by their ages (e.g. 18 years old in one cluster, 19 years old in one cluster, etc.), then randomly sample three clusters and survey all students in those clusters.

A) It is reasonable but it is subject to non-response bias.

B) It is reasonable.

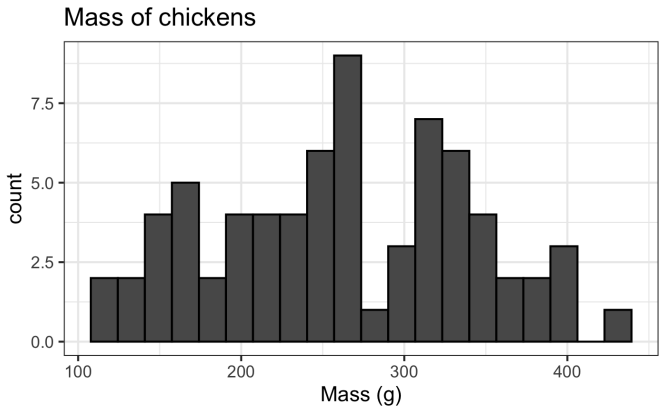
C) It is reasonable but the groups would likely have to be age groups since colleges can have a large range of student ages.

Equitable Equations: *Describing data qualitatively*

Problem 1

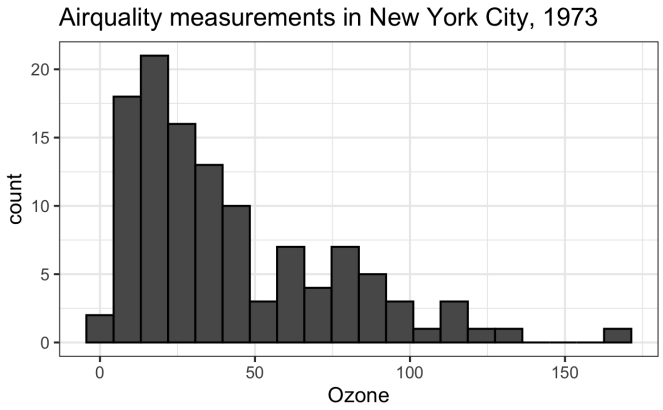
Qualitatively describe the following plots using the vocabulary developed in class.

(a)



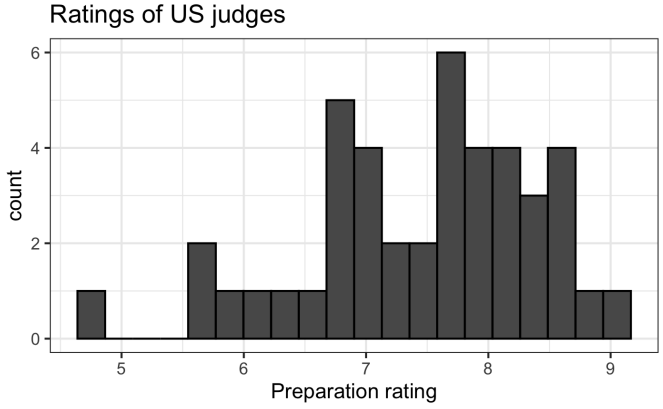
This histogram is relatively symmetrical.

(b)



This histogram is right-skewed with a possible outlier on the far right.

(c)



This histogram is left-skewed.