

# Exercise Sheet: Foundations of Data Collection

Data, Data Storage, Data Collection

## Exercise 1 - Quantitative vs Qualitative Data

For each of the following, indicate whether the data is quantitative or qualitative:

- (1) Hours of sleep students get the night before the exam.
- (2) Transcripts of customer service calls.
- (3) Number of steps recorded daily by a fitness tracker.
- (4) Photos taken by wildlife cameras in a forest.
- (5) Results of a multiple-choice survey about favorite foods.
- (6) Scores from standardized math tests.
- (7) Zip codes of shipment addresses.
- (8) Ratings (1-5 stars) given to movies on a review site.

## Exercise 2 - Scope of Data

The California Environmental Protection Agency (CalEPA), the Office of Environmental Health Hazard Assessment (OEHHA) developed the CalEnviroScreen project to study how environmental hazards relate to individual health in California.

The project studies connections between population health and environmental pollution using:

- Demographic summaries from the US Census
- Health statistics from the California Department of Health Care Access and Information
- Pollution measurements from air monitoring stations maintained by the California Air Resources Board

Data are available at the level of census tracts (aggregation over geographic areas), not individuals. For example, one can examine rates of asthma hospitalizations vs. air quality across tracts, but not the impact on any specific individual.

- (1) What is the target population for the original research question?
- (2) What is the access frame used in this study?
- (3) What is the sample in this context?
- (4) Why does aggregation at the census tract level limit the conclusions that can be drawn?