Lucas Simonsen 25/9/2025

Assignment 1: Collecting objects and building datasets

The process of creating a spreadsheet dataset began with collecting data from my own kitchen. Twenty four analogue objects were selected, ranging from condiments, oils, spices, bread and protein. The kitchen was chosen due to the wide variety of items it offered, allowing for a sampling rule of one entry per unique object. All selected objects were consumable and gathered from either the pantry or fridge. Initially, only items not requiring refrigeration were considered, however this did not yield enough entries, so items from the fridge were included. Although the sampling rule stated, only allowed one entry per unique object, the decision of including variations of the same product type with differentiating forms or use was made to allow for enough consumable items. For example, coffee entered into the dataset three times, as grounded, capsuled and instant. This would of course increase the share of one type of object entries.

The process of digitising included creating a spreadsheet dataset in Microsoft Excel. One row indicates one object within the dataset while the first row represents the titles of the columns, describing variables within the dataset. By structuring the dataset with rows as objects and columns as variables, information could easily be easily sorted and compared. Each object was entered into the spreadsheet using nine variables: object, color, origin, category, container material, quantity, unit, flavor profile and physical state. The choice of having nine variables within the dataset, ensured that the information about the objects were detailed, meanwhile ensuring manageability. The variables chosen to represent information about the objects, were based on the decision to ensure that the most relevant qualities of the objects were represented in a consistent and comparable way. Other possible variables were excluded such as brand, expiry date and price to reduce complexity of the dataset. This would of course limit certain aspects of possible analyses. Consistent spelling formats were used throughout this process, with the first letter in all cells within the spreadsheet being capitalised, except the column indicating the unit variable. The entries within this column were spelled with lowercase letters, such as ml and g, which enhanced clarity of which unit representation the object contained, as it matches measurement standards. Mixing unit measurement in a column with g and ml makes the dataset easier to read but harder to compare quantitatively. Consistent capitalization of all other cells substantially improved readability. The titles of the columns were colored with a lighter green and every other row in the dataset was colored with a light gray. This reduced visual strain and helped track rows. To further ensure readability of the dataset, columns were marked with edges, ensuring explicit grid structure. As Dourish (2017) argues, spreadsheets are not just neutral

storage tools, they also structure how information is organized and acted upon. The formatting and coloring of this dataset supports this idea, as it reflects how the spreadsheet shapes what is visible.

Controlled vocabularies were established throughout the dataset for variables to avoid inconsistencies from entries. At the same time, ambiguity was avoided within the flavor profile to make the dataset more machine-readable. The process of digitising helped transferring analogue objects into a structured digital format and the next process of datafying, helped translate the kitchen objects' complex qualities into simplified variables. This process involved more subject decision making as only one entry had to be selected for each variable to describe the qualities of the objects. To describe the flavor profile of an object multiple taste notes had to be shortened into only one label. The flavor profile of olive oil was for example described with the label, fruity even though the taste notes also include some bitterness. The same was true for coffee which was labeled with a flavor profile of bitter, even though it also has taste notes of earthiness. A subjective decision was made to label the primary flavor profile of the objects. This made the dataset tidy, but biased towards the descriptor that was chosen. As Kitchin (2022) emphasizes, data is not neutral but produced through social and technical decisions.

To indicate the origin of objects, the packaging of the objects were studied. The origin variable within the dataset indicates where the finished product originates from. A suspicion arose due to the fact that not all original origins of the objects were listed. To most adequately indicate the original origin of these objects, google searches were conducted to research where they originally came from. The coffee capsules for example were listed as distributed from Denmark. A quick google search indicated that they were in fact from Switzerland. By recording the country of origin for the finished product, instead of where it for example was grown, might have limited the number of countries indicated within the dataset.

Bibliography:

Kitchin, Rob. 2022. “Introducing Data.” In The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, 1–19. SAGE Publications Ltd.

Dourish, Paul. 2017. “Spreadsheets and Spreadsheet Events in Organizational Life.” In The Stuff of Bits: An Essay on the Materialities of Information, 81–104. Cambridge, [Massachusetts]: The MIT Press.