Tobias Kyhn Mikkelsen
Curating Data
AU

25/09/2025
Characters: 4888

Assignment 1 - Collecting

When I started collecting golf balls for this dataset, it looked like a very simple task. But during the process I realised how many choices were involved in turning physical objects into data. I ended up with 20 golf balls, and for each of them I wrote down five variables: *Mærke, Kvalitet, Farve, Tilstand,* and *Model*. At first this felt obvious, but when I worked on it I saw how these categories came from my own decisions. The dataset looks neat in a spreadsheet, but it hides the work and the judgement that made it.

I collected the balls from what I already had at home. It was not a finished collection like stamps or postcards. Just the act of saying that golf balls are "worth" collecting was already a decision. I could have chosen other objects, but golf balls had enough similarities and differences to make sense as a dataset. For digitising, I did not use pictures or scans, but instead typed them into a table. This was easier and made them comparable, but it also removed their physical form. A ball is something you can touch and play with, but in my dataset it is just a line with words. The physical scratches, weight in the hand, or memories of using the ball are all lost in this translation.

Choosing the variables was the most important step. I picked *Mærke* because brands like Titleist or Callaway are well known and carry a certain status. *Kvalitet* was harder, because I had to decide if a ball was "god," "medium," or "dårlig." That is clearly a personal judgment. *Farve* and *Tilstand* were more clear, but still I had to choose when a ball was "ny" or "brugt." *Model* gave extra detail, but sometimes I was not sure if I recognised it correctly. At the same time I left out many other things: weight, year of production, or personal stories of use. So the dataset reflects both what I thought was important and what I could easily measure. Someone else might have built a very different dataset with the same balls, showing how subjective these steps really are.

Rob Kitchin writes that data are not simply "given," but "taken" from reality. He uses the word *capta* to show that we always select from many possible details (Kitchin, 2022). This fits my dataset well. Each row looks neutral, but it is only one version of the ball, shaped by what I decided to include. I could have focused on different aspects, and then the dataset would tell a very different story. What seems like "raw data" is really the result of my own choices. Even a dataset that looks very small and simple shows this selective nature. Kitchin also points out that such selections have consequences: they decide what kind of questions can be asked later and which answers are possible. My golf ball data can be used to ask about brands or conditions, but not about history, weight, or who played with them.

Jacqueline Wernimont also helps to understand the process. She says quantification is not only description, but also an act of making meaning. By writing that a ball is "god" or "dårlig," I create a hierarchy. A Callaway Chrome Soft marked as "god" looks more valuable

than a Top Flite marked as “dårlig.” These numbers and words are not universal truths. They are my way of organising the world, just like Wernimont’s example where a dinner party can be counted as “four people” or “two couples.” Both are correct, but they give different views. In the same way my dataset creates one specific way of seeing the balls, not the only possible one. Quantification therefore does not only describe reality, it shapes it.

Seen like this, my dataset is simple and complex at the same time. On one side, the table looks clean: 20 rows, 5 categories. On the other side, it hides all the interpretation and labour behind it. Kitchin reminds us that what looks raw is really selective. Wernimont shows that quantification is not neutral, but always a choice that makes some things visible and hides others. The clarity of the spreadsheet is therefore a bit misleading. It looks like pure information, but it is also a cultural product with my own values inside it. If another person had made the dataset, it would look different, because their judgments and priorities would not be the same as mine.

In the end, my small golf ball collection shows the bigger point of the assignment. Data are never just data. They are shaped by decisions, categories, and judgments. My spreadsheet is not only about golf balls, but also about the way I chose to describe them. Even something as ordinary as this becomes a good example of how theories about *capta* and quantification matter in practice. The process made me aware that whenever we make datasets, we also make choices about what counts, what does not, and how objects will be understood in the future. In that way the assignment was useful, because it forced me to see the hidden work behind something that first looked easy.

Literature review:

Kitchin, R. (2022). Introducing data. In *The data revolution: Big data, open data, data infrastructures and their consequences* (pp. 1–19). SAGE Publications Ltd.