Sebastian Kjær
Curating Data
Aarhus University

25-9-25
Characters: 4.876

Assignment 1: Collecting objects and building dataset

My collection of objects is based on my collection of 20 golf clubs, that I both described and datafied. Each club was measured and registered with information like name, brand, loft in degrees, length in centimeters, average distance when hitting a golf ball in meters and a subjective assessment of the condition. Length of the club were measured physically, and loft was taken from the brands website, while the average distance was estimated from several shots on the driving range, with the help of a shot tracker that measured all the shots precisely. All the data was then put into a spreadsheet. The data attributes that I chose had to balance the clubs' physical properties and functional dimensions. Distance became a central data attribute, because it allowed the clubs to be compared directly. The dataset can then easily be analyzed with the connection between length and stroke length of the clubs. But makes other stories invisible such as other dimensions like precision, ball control or user experience, and the fact that range balls don't fly as long as quality branded golf balls.

The decisions about variables greatly shaped how the data set could be interpreted. I could see a coherence between the club length and distance, the longer the golf club was, the longer the golf shot became. While a steeper loft angle generated a shorter golf shot. This shows how the length and loft of the club each affect the result, but also how they impact each other. A high loft angle is most often found on shorter clubs, which amplifies the effect and makes the ball travel both shorter and higher. On the other hand, clubs with lower loft angle provides lower ball flight and more distance on the golf shot. It's therefore obvious for this dataset, that the single variables can't be understood on their own but must be seen in relation to each other to explain the variations in distance. The variable about condition is something that can't be measured precisely but builds on a subjective assessment. The clubs' condition is categorized with terms such as "Very good", "Used", "A bit used" and "Very used". But these assessments is an estimate rather than an objective measurement. Therefore, is the dataset not only a neutral registration but also formed by my own choices and interpretations.

At the same time the digitalization process shaped the dataset. Moving from a physical observation to a spreadsheet required an adjustment to rows and columns. This structure also highlighted certain patterns but hid other dimensions of the golf clubs, such as perception and

shot shape. The digitalization is therefore not a neutral transfer, but an active shaping of what can be seen and analyzed.

Here it is relevant to draw on Rob Kitchin's distinction between data and capta. Kitchin points that data isn't "raw" or given but is something that is taken from the real world through selective decisions and technological frameworks (Kitchin, 2022). In relation to my dataset this means that "stroke distance" and "condition" is capta, selected aspects of the clubs while other possible dimensions were never captured. This underlines that the dataset is a product of selection and not a neutral understanding.

Jacqueline Wernimont describes in her analysis of quantification, that the assignment of numbers is never just a description but a performative and creative practices that create certain understandings of the world (Wernimont, 2021). By choosing an average stroke distance as a central unit of measurement, I am making distance the dominating perspective on the golfclubs, while precision and feel in the golf shot disappears. The choice of what is made measurable is therefore also a choice of which narratives can be produced.

Finaly in Rowley's paper about the DIKW hierarchy gives a framework to understand how my dataset moves from raw measurements to knowledge. According to Rowley is the transformation from data to information and knowledge nonlinear but depends on context and interpretation (Rowley, 2007) In this dataset, the measurements only become information when I analyze relationships such as loft and length in relation to distance and then turned into knowledge when I translate it into practical advice on choosing a club when playing.

The dataset shoes overall that curation is not just about colleting and storing information, but also about designing future possibilities. The variables I chose highlighted certain stories, especially in relation to loft, length and distance, while other possible dimensions weren't stated. Data is situated and produced, never "raw" or neutral. Currenting data is therefore an active process where choices determine what is measured and how it is organized, and that creates patterns and relations which can later be interpreted as knowledge.

Sebastian Kjær
Curating Data
Aarhus University

References:

- Kitchin, R. (2022). *The data revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures*. Sage Publications Limited.
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706
- Wernimont, J. (2021). Quantification. In *The MIT Press eBooks* (pp. 427–432). https://doi.org/10.7551/mitpress/12236.003.0047