

Operations

Prof. Dr. Nicolas Meseth

Data is the new oil, according to the mathematician [Clive Humby](#):

“Data is the new oil. Like oil, data is valuable, but if unrefined, it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity. So, must data be broken down, analysed for it to have value.”

If we take this analogy seriously, the data, like oil, needs to be refined to turn it into something of value. Two important tools for refining data into a valuable output are *data transformation* and *data visualization*, both of which are the main focus of this book. In this part of the book, we first need to learn how to transform data so that we can apply visualization later on.

To learn how to transform data, we need to learn how to to the following operations:

- Remove any variables we don't currently need (or specify those we **do** need)
- Remove any records we don't currently need (or specify those we **do** need)
- Add new variables that don't exist yet
- Summarize many records into one or a few numbers
- Change the order of the records

The goal of the following chapters is to introduce means to perform theses five operations with R.