Adding Variables and Quantifying Codes 10

Even in strictly qualitative research, a set of standardized (quantitative) data will usually be collected, e.g., to record sociodemographic characteristics such as age, gender, education, number of children, religious affiliation, and much more. This data can be used very well to form groups and compare them with one another. What do women say about a certain topic, for example, and what do men say? Quantitative data, referred to as "variables" in MAXQDA, is of course also used in mixed methods projects, where qualitative and quantitative methods, data, and analyses are combined and interlinked. A third area that involves numbers also directly relates to qualitative analysis. Each time you assign a code to a segment of data, you are essentially classifying it, which in turn generates information about the frequency of code assignments per code and per case. You can then use this information in your analysis, for example, to find out who spoke about which topic and how often.

In This Chapter
- Understanding the meaning and purpose of quantitative data (variables)
- Getting to know MAXQDA's variable functions: the "List of Variables" and "Data Editor"
- Entering and editing variable values
- Transforming code frequencies into variables
- Statistically analyzing and visually presenting quantitative data

Benefits of Using Variables for Quantitative Data?

There are many examples of how standardized, quantitative information can be integrated into qualitative research, even beyond mixed methods strategies. Anyone who conducts an interview study will have information about the people they have

interviewed, be it just the location the interview was conducted. Questionnaires are often used in parallel to interviews to alleviate the need to cover standardized information in the latter and to allow more time for interactive questions instead (Witzel & Reiter, 2012; Kuckartz, 2014). If you conducted a qualitative study in which you interviewed the educators, leaders, and financiers of a kindergarten in a socially deprived area, for example, you would automatically have access to information about grouping criteria, in this case the professional positions of those surveyed. As in this example, it is clear that qualitative studies often rely on standardized characteristics in their selection of qualitative samples, whether it is through a conscious selection process, quota system, or a strategy based on theoretical sampling (Corbin & Strauss, 2015). When carrying out a focus group study, sometimes homogeneous, other times heterogeneous compositions of groups are chosen. In other words, some additional information about the individual participants must also be available in these cases. When importing Twitter feeds into MAXQDA, standardized information about the authors of the tweets and the tweets themselves is automatically available in addition to their 280-character qualitative texts. This standardized data, such as the number of followers the author has, the language of the tweet, or whether it is a retweet, opens up extensive filtering and contrasting possibilities with respect to the qualitative data. The same applies to online surveys with closed and open questions: for each case, there is a set of standardized information available along with the answers to the open questions.

MAXQDA uses the term "variable" for standardized, quantitative data; you could also use the terms "attributes" or "characteristics," which are available as supplementary and descriptive information to the individual cases. This is because, in contrast to quantitative research, where the primary objective is the aggregation and consolidation of data using mean, standard deviation, and other statistics, qualitative data analysis focuses more on individual cases. In view of the examples mentioned above, quantitative data in the form of variables in MAXQDA can offer significant benefits for qualitative data analysis, including:

- They can be used for individual case analyses as supplementary information to help classify, explain, and interpret the available data and can also be integrated as descriptive features when creating case summaries.
- They can be used to form groups and allow you to contrast and compare cases.
- In addition to group formation, they can also be used for group descriptions, especially if the groups are derived from the qualitative data itself. This would be the case, for example, if respondents in an interview study are divided into three groups on the basis of their qualitatively analyzed responses to questions concerning their self-esteem—i.e., those with high, medium, or low self-esteem. Subsequently, these groups can be compared according to sociodemographic variables, such as the average age or the proportion of men.
- They can support the formulation of hypotheses about a given case that need to be tested.
- They can be used for sorting and filtering—especially when it comes to large volumes of data.

- They serve as a central link between qualitative and quantitative data in mixed methods analyses. When working with MAXQDA Stats, the add-on module for descriptive and inferential statistics, document variables provide this link between qualitative and quantitative data and enable further mixed methods analyses.
- They can be used to record quantitative information about a case, e.g., how often a topic has been coded for that case.

MAXQDA allows you to define variables not only for documents but also for codes in your code system. Code variables were originally introduced in MAXQDA 10 to store standardized information for the participants in group interviews. MAXQDA 12 introduced especially designed focus group variables for this purpose, which are described in detail in Chap. 15. Code variables have continued to be available in versions since MAXQDA 12 and can be used, for example, to trace the origins of codes (concept-driven vs. data-driven), or their creation date, and this information can in turn be useful for your analysis. The procedure for working with variables is the same for all variable types, so in this chapter we will limit ourselves to describing the most frequently used document variables.

Managing Variables in the "List of Variables"

By selecting *Variables* > *List of Document Variables* you can open a list of the available document variables in MAXQDA, as shown in Fig. 10.1. Here you can create new variables and edit or delete existing ones.

When you open the "List of Document Variables" for the first time, it will already contain six variables that MAXQDA automatically creates in each project and which cannot be deleted. You will recognize them by the red square in the first column. MAXQDA stores important information about your imported documents in these: the document group, document name, date of import, current number of coded segments and memos within the document, as well as the name of the "author,"
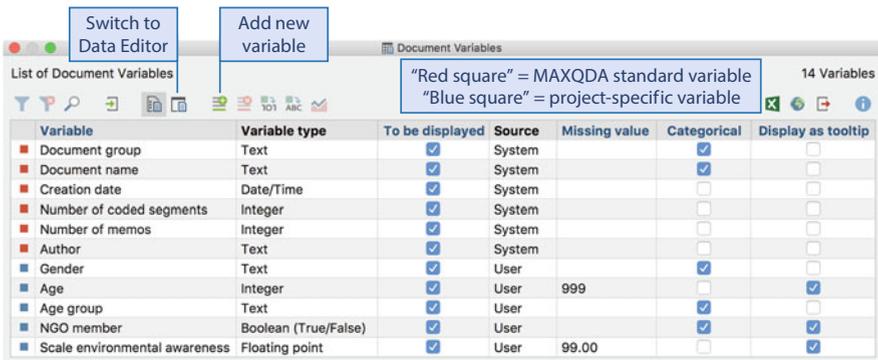

Variable	Variable type	To be displayed	Source	Missing value	Categorical	Display as tooltip
Document group	Text	✓	System		✓	
Document name	Text	✓	System		✓	
Creation date	Date/Time	✓	System			
Number of coded segments	Integer	✓	System			
Number of memos	Integer	✓	System			
Author	Text	✓	System			
Gender	Text	✓	User		✓	
Age	Integer	✓	User	999		✓
Age group	Text	✓	User		✓	
NGO member	Boolean (True/False)	✓	User		✓	✓
Scale environmental awareness	Floating point	✓	User	99.00		✓

Fig. 10.1 The "List of Document Variables"

i.e., the person who imported the respective document. In Fig. 10.1, five project-specific variables have already been defined in addition to the so-called "system variables." The values of the additional variables can be changed and are therefore marked with a blue square as "user variables" in the list.

A new variable is added to a project using the icon of the same name at the top of the window. A dialog box will then open in which you can enter a descriptive variable name of up to 63 characters and specify the variable type. While you can change the name later, the variable type can only be changed to a limited extent, which is why it is important to consider this choice of type carefully before adding a new variable. The following types are available:

- Text—This variable type allows you to enter any text of up to 63 characters as a value for each document. While in statistics programs standardized information is usually represented with numbers (e.g., 1, female; 2, male), this can prove very impractical in MAXQDA, because then you would always need a table of correspondence at hand to interpret the data. Instead, it is usually more helpful to work with plain text labels for the variable values in MAXQDA—which is why, in the example in Fig. 10.1, the gender has been defined as a text variable—so that you can enter "female" and "male" or "w" and "m" as immediately identifiable text.
- Integer—This variable type lets you specify (positive and negative) integer values. A classic example of this "integer" type is age. Negative values are less common than positive values but do occur, for example, when assessing the level of a student's seminar on a scale from "too low (-3)" to "too high (+3)," where negative values can be chosen.
- Floating point—Whenever you want to enter numbers with decimal places as additional information, you need this variable type. For example, this type can be used "for mean values of attitude scales or other test values." MAXQDA allows the input of two decimal places and always displays entered values with two decimal places.
- Date/time—This type is suitable for all situations in which you want to record the date and time for a case. For each case, you can enter both a date with a time and a date without a time. It is not possible to enter only a time. MAXQDA automatically recognizes most input formats and displays the values entered according to the current operating system settings. It makes no difference whether you enter the year as two or four digits; MAXQDA saves this date in a uniform, universal format and displays it as such.
- Boolean (true/false)—This variable type only has two values, namely, the logic values "true" and "false," which in a concrete case can also mean "yes" vs. "no" or "applies" vs. "does not apply." This variable type is not used as often, since it also comes with some restrictions, e.g., missing values cannot be defined using the Boolean variable type. Usually, the types "Integer" and "Text" offer more flexibility, as you can work with "0 vs. 1" or with "yes vs. no" and achieve the same results.

What Are "Missing Values"?
Once you have created a new variable, you can define further settings in the "List of Variables." In the "missing value" column, you can enter a variable value that will not be taken into account in subsequent case selections and other analyses based on variable values. Let us assume that in an interview study, there is no information on age for two cases. You can then assign the value "999" (or "−99") to these cases and enter this value in the column "missing value." If you then use MAXQDA to select people over 40 years of age, the cases with the values 999 will automatically be ignored. In addition to the values defined as missing, empty cells are also considered missing by MAXQDA. Empty cells can only occur for text and date variables. For Boolean variables, it makes no sense to define missing values, because here MAXQDA only lets you set a check mark denoting "yes" or "applies" or not; other values cannot be entered.

► **Please Note** For the variable types "integer" and "floating point," there can be no empty cells in MAXQDA. Hence, when a new variable is created, the value 0 is entered as the initial value in all cells. You should always take this into account if the value "0" may also occur as a real value and you do not know whether an existing 0 was deliberately entered for a case or whether it already existed as an initial value. For example, if you were to define the number of a person's children as a variable, you should define the missing value as "−99" to be able to distinguish people with no children from those for whom the number of children is unknown.

What Does the Variable Property "Categorical" Mean?
In general, different measurement levels are used for quantitative variables (Kuckartz, Rädiker, Ebert, & Schehl, 2013, pp. 16–20):

• With *nominal-scaled variables*, individual variable values cannot be sorted; each value is of equal importance, for example, "Gender: female, male."
• With *ordinal-scaled variables*, individual values can be ranked, for example, "Education levels: low, medium, high."
• With *interval-scaled variables*, individual values can be ranked and the intervals between values are always identical, for example, "Age: 10, 11, 12, 13 years." (In fact, ages in years are ratio-scaled, since they have an absolute zero point, but ratio-scaled variables in the social sciences are often only treated as interval-scaled.)

Nominally and ordinally scaled variables are also referred to as "categorical" variables; their categories, i.e., their individual variable values, are central to them, and you cannot calculate statistical measures, like mean values, for them as you could for interval-scaled variables. When you set the "categorical" property for a variable, you are instructing MAXQDA not to interpret its values as interval-scaled.

This would be necessary, for example, if you wanted to create the variable "education level" with the values 1, low; 2, medium; and 3, high. These values are ordinally scaled, and it does not make sense to calculate averages for them. By setting the property "categorical," certain functions in MAXQDA will calculate percentages for cases—for example, the percentage of people with a high level of education—instead of calculating their mean values.

Setting the "categorical" property is only useful in the case of "integer" or "floating point" variables, since MAXQDA sets the other variable types as categorical anyway. In other words, only in the case of numerical variables does the question arise as to whether they should be interpreted as categorical or interval-scaled.

What Is the Benefit of Setting the "Tooltip" Option for a Variable?
In computer programs, the "tooltip" is the information that appears when you place your cursor over an object. In MAXQDA, there are tooltips in several places: when you hover your cursor over a memo, a preview of the memo will appear; for a document link, a preview of the link target will appear; and for a document in the "Document System," information about who imported the data and how many memos are assigned to it is displayed. If you select the option "Display as tooltip" for a variable in the "List of Variables" (Fig. 10.1), the selected variable name and the variable value entered for the document will additionally appear in this document's tooltip (Fig. 10.2). This is particularly practical for quickly exploring cases, because instead of needing to read the variable information from a table when writing a case summary, for instance, you can simply hover your cursor over the document name to display the relevant context information about the case. The "Tooltip variable" setting also proves useful in other places in MAXQDA. Coded segments from the "Retrieved Segments" window or the Overview of Coded Segments can be exported together with their tooltip variables as additional information. For interviews, for example, you can include important case information on every retrieved or exported coded text passage, such as the age of the person interviewed.
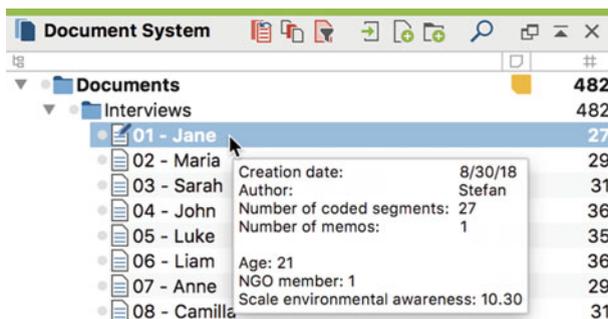

Fig. 10.2 Tooltip variable information in the "Document System"

Entering, Changing, and Viewing Variable Values Within the "Data Editor"

In MAXQDA's "List of Variables," you can organize your variables, while in the "Data Editor," you can edit and view the data for all or only selected cases. The "Data Editor" is opened either via the *Variables* ribbon tab or by clicking on the icon 🗐 with the same name at the top of the "List of Variables." The icon 🗐 switches back to the "List of Variables", so that you can switch between the variable and data views at any time. The structure of the "Data Editor" corresponds to a typical rectangular data matrix: cases (documents) form the rows, while variables form the columns (Fig. 10.3). As in all MAXQDA table overviews, columns whose values can be changed by the user have a blue header, while columns that cannot be changed and are set by MAXQDA have a black header.

If you only have a few variables and cases, you can usually enter the data quite quickly and conveniently by hand once you have created the relevant variables. To enter the data, click in a cell with the mouse, either a single or double-click will do, and enter a value. If you are entering this data column by column, it is best to confirm the entry with the *Enter* key (↵) as MAXQDA will then automatically select the next row. If you plan to enter the data on a case by case basis, you should press the *Tab* key (↦) after the entry; then MAXQDA will jump to the next column in the "Data Editor."

By default, the "autocomplete" function is switched on for all variables, with the exception of the "Boolean" type. When you start to enter a value, the entered letters will be supplemented with suggested variable values that already exist in the column, just like in Excel. You can see this, for example, in the column "Gender" in Fig. 10.3. All you need to do is type an "m" and MAXQDA will automatically complete this value to "male," so that you only need to press the enter or tab key to accept the value.
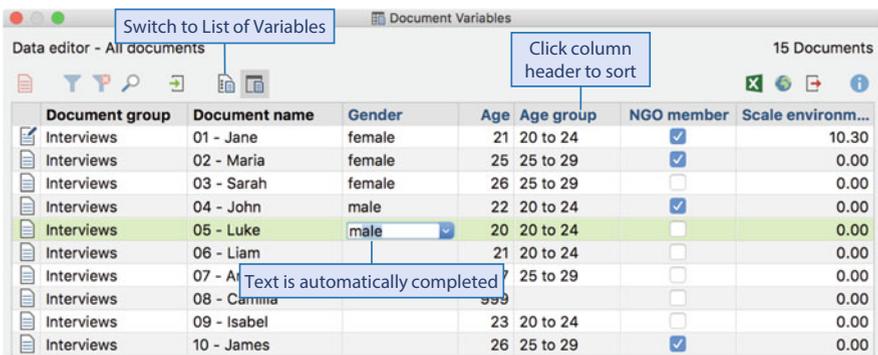

Fig. 10.3 The "Data Editor" for editing the variable values per individual case, here: documents

Explore Variables for Individual Documents, Document Groups, or Sets
You can easily set constraints over the rows shown in the "Data Editor." To explore only the variable values of documents in a document group, right-click on a document group in the "Document System," and select the *Overview of Variables*. The "Data Editor" which is displayed in this case will only contain the documents from the selected document group. You can also restrict the view to a single document or the documents in a set. Alternatively, you can restrict the displayed rows via activations in the "Document System" and the left-hand icon **Only activated documents** in the Data Editor toolbar.

It is also important to note that the "Data Editor" and "Document System" are interactively linked. If you select a document in the "Document System," the corresponding row is selected in the open "Data Editor" and vice versa. You can use this feature to explore documents if you prefer not to work with tooltip variables, which might be very useful when working with data from online surveys.

Import an Existing Data Matrix from Excel or SPSS
In studies, the standardized data are often already available in a data matrix of "cases x variables," and it is not necessary to reenter such data by hand. MAXQDA lets you import this data, as long as it is available in Excel or in the SPSS format, which is often used by statistical software. To make sure the data is assigned to the correct cases when importing it, MAXQDA requires the imported matrix to contain two columns for the assignment of cases, one for the document group and one for the document name. For imports from Excel, these columns should be titled as such; for SPSS imports, these columns can be selected as desired.

The import function is available via the *Import Document Variables* option in the *Variables* ribbon tab or via the corresponding icon in the "Data Editor" or "List of Variables." A file dialog box will then open in which you can select an Excel or SPSS file. MAXQDA will open the file, list all the variables contained in it, and allow you to specify for each new variable which variable type should be used.

If a given variable name already exists and has the same variable type, its values will be updated with the values of the imported variable—and empty values also overwrite already existing values! When importing SPSS files, you can select which variable column contains the document group and which the document name. In addition, you can specify that the variable labels should be imported, instead of the often-abbreviated variable names, and the value labels rather than the numerical codes—both of these options are generally to be recommended.

► **Please Note** Document variables are always defined globally for the entire project, i.e., you cannot define different variables for different document groups. But this is seldom necessary, because you can create any number of variables and only assign variable values to the documents for which they are relevant. For other documents, these cells will simply remain empty or the defined "missing value" will be displayed.

Transforming Code Frequencies into Document Variables

In the course of the coding process with MAXQDA, quantitative information about how often a code was assigned to a case is generated for each document. MAXQDA allows you to create document variables that reflect the frequency distribution of a code across individual documents. These variables are dynamic and are always automatically adapted to the current status of the project.

Let us assume that all text passages in which interviewees describe their environmentally conscious behavior were coded with the category of "personal behavior." By right-clicking on this code in the "Code System" and selecting the option ***Transform into Document Variable***, you can create a new document variable whose values reflect how often that code has been assigned in individual documents (Fig. 10.4). The code name is used as the variable name, which can be changed at any time in the "List of Document Variables." While the name of this option in the context menu may suggest otherwise, the code is completely preserved by this action and will not be changed in any way.

► **Please Note** The values of a transformed code are updated automatically by MAXQDA whenever the code is assigned to a new segment or deleted from an existing one and so always reflect the current status of the project.

The variable can serve as an indicator of the extent to which people behave in an environmentally conscious manner, but it is immediately clear that these purely quantitative results should be interpreted with a degree of caution, since the frequency with which a topic is mentioned is not solely dependent on how important it is considered to be. The interview situation, the extroversion of an interviewee and, above all, the coding rules on how to deal with repeated occurrence of identical statements also have an effect on the number of times a code is assigned.

Fig. 10.4 Code transformed into a document variable in the "Data Editor"
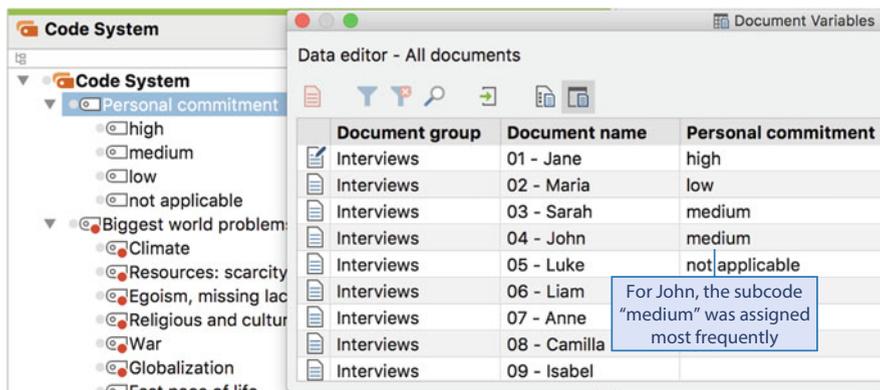
Fig. 10.5 Code converted to a categorical document variable (left) in the Data Editor (right)

The transformation of the frequencies of qualitative codes into variables with numerical values is also described in mixed methods literature as "quantitizing" (Kuckartz, 2017). Accordingly, the ***Mixed Methods*** ribbon tab also contains an icon bearing the same name, which can be used to transform several codes into document variables at once.

Transforming Codes into Categorical Document Variables
To continue our example, where the interviewees' personal engagement in environmental protection has been coded using scaled subcategories such as "low," "medium," "high," and "unknown," MAXQDA offers you the option of transforming the parent code "Personal engagement" into a document variable. The values of this variable will then be set to the subcategory which was assigned most frequently in each document (Fig. 10.5). To create such a variable, click on the parent code and choose ***Transform into Categorical Document Variable***. Transforming the parent code in this way provides an overview of the levels of personal engagement per document, which can be used for evaluative content analyses (Kuckartz, 2014; Mayring, 2014; Schreier, 2012) or to group documents together and contrast them, for example, the less committed vs. the highly committed.

Creating Frequency Tables and Charts for Document Variables

For the first analysis of the variables, you can use frequency tables and charts in which the individual variable values are listed and counted. They allow you to get an idea of the distribution of variable values and can be integrated into reports describing the sample and results. You can create tables and charts by going to
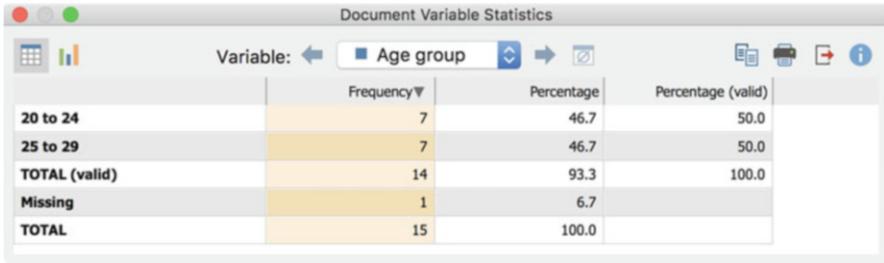
	Frequency▼	Percentage	Percentage (valid)
20 to 24	7	46.7	50.0
25 to 29	7	46.7	50.0
TOTAL (valid)	14	93.3	100.0
Missing	1	6.7	
TOTAL	15	100.0	

Fig. 10.6 Frequency table for a document variable

Variables > Document Variable Statistics. In the dialog box that opens, simply select several or all the available variables. The result is a window with a frequency table as shown in Fig. 10.6.

All variable values that occur are listed in the first column of the result table—in the example in Fig. 10.6, the two age groups "20–24 years" and "25–29 years." The second column indicates how many documents contain these variable values—in the example, there are seven interviewees (documents) per age group. The "Missing" row indicates how many documents contain either an empty value or a value defined as missing. In the example, this is the case for an interviewee for whom there is no age information available. The "Percentage" column contains the relative frequencies of the variable values with respect to all the values, including the missing ones. Usually, however, you will ignore the missing values in your analysis and therefore refer to the values in the "Percentage (valid)" column, which only takes into account the number of documents with valid values.

If the first column only contains numbers, which would be the case with a numeric variable, for example, the "number of children" variable, you can click the *Descriptive statistics* icon ☑ to request the mean value, standard deviation, quartile values, and other statistical measures to describe the variable's distribution in addition to the frequency table.

The *Chart view* icon ▮▮ switches the display of this data from a frequency table to a chart (Fig. 10.7). Once you have changed to this view, you can choose between a vertical or horizontal bar chart and a pie chart. The display can be adjusted using the icons at the top of the window; for example, you can toggle between displaying absolute values and relative frequencies as percentages, and you can also hide or display any missing values. The order of the columns, bars, and circle segments is always dependent on the order of the rows in the frequency table—there you can sort the variable values by frequency or by alphabetical order by clicking on the column headings. Both charts and frequency tables can be exported in numerous formats via the usual icon at the top right-hand corner of the window.
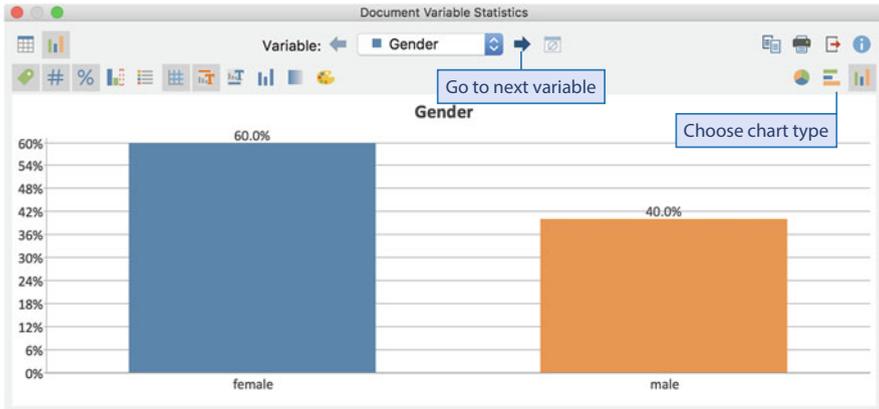
Fig. 10.7 Bar chart for the document variable "Gender"

▶   **Tip** You can also generate a frequency table or a chart directly from the "List of Variables" by clicking on the Statistics icon after selecting one or more variables with your mouse (see Fig. 10.1).

References

Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA: SAGE.

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice & using software*. Thousand Oaks, CA: SAGE.

Kuckartz, U. (2017). Datenanalyse in der Mixed-Methods-Forschung: Strategien der Integration von qualitativen und quantitativen Daten und Ergebnissen. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 69*(S2), 157–183. https://doi.org/10.1007/S11577-017-0456-Z.

Kuckartz, U., Rädiker, S., Ebert, T., & Schehl, J. (2013). *Statistik: Eine verständliche Einführung* (2nd ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. Klagenfurt. Retrieved from http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: SAGE.

Witzel, A., & Reiter, H. (2012). *The problem-centred interview: Principles and practice*. Thousand Oaks, CA: SAGE.