Conceptual difficulties when interpreting histograms: A review

Lonneke Boels\textsuperscript{a,b,*}, Arthur Bakker\textsuperscript{a}, Wim Van Dooren\textsuperscript{c}, Paul Drijvers\textsuperscript{a}

\textsuperscript{a}Utrecht University, the Netherlands
\textsuperscript{b}Christelijk Lyceum Delft, the Netherlands
\textsuperscript{c}KU Leuven, Belgium

ARTICLE INFO

Keywords:
- Big ideas of statistics
- Misconception
- Statistical concepts
- Statistics education
- Statistical knowledge for teaching (SKT)
- Histogram

ABSTRACT

Histograms are widely used and appear easy to understand. Research nevertheless indicates that students, teachers and researchers often misinterpret these graphical representations. Hence, the research question addressed in this paper is: What are the conceptual difficulties that become manifest in the common misinterpretations people have when constructing or interpreting histograms? To identify these conceptual difficulties, we conducted a narrative systematic literature review and identified 86 publications reporting or containing misinterpretations. The misinterpretations were clustered and—through abduction—connected to difficulties with statistical concepts. The analysis revealed that most of these conceptual difficulties relate to two big ideas in statistics: data (e.g., number of variables and measurement level) and distribution (shape, centre and variability or spread). These big ideas are depicted differently in histograms compared to, for example, case-value plots. Our overview can help teachers and researchers to address common misinterpretations more generally instead of remediating them each individually.

1. Introduction

Statistical literacy is a core competence for citizenship and therefore an important goal of statistics education for students of all ages (Ben-Zvi, Makar, & Garfield, 2017). It includes the ability to interpret graphical representations of statistical data (Ben-Zvi & Garfield, 2004; Garfield & Ben-Zvi, 2007). Graphical representations of statistical data can be found in newspapers, schoolbooks, research articles, government policy reports, television news bulletins and other common sources of information. “Graphical representations serve as useful tools to communicate aspects of a distribution as they facilitate a focus on aspects of the data that may be missed with the use of descriptive statistics alone” (Leavy, 2006, p. 90); see also Pastore, Lionetti and Altoé (2017). Different representations reveal different aspects of the data. Many real-life examples show that lives can literally be saved if people master the ability to switch between different representations of data to reveal different aspects. One example is from Nightingale who saved many lives with her famous polar graph (Martineau, 1859) which showed that more soldiers died from preventable diseases—caused by bad hygienic circumstances in the hospitals—than from the war wounds caused by the Crimean War.

A graphical representation widely used to represent the distribution of univariate scale data is the histogram. What researchers consider a histogram is rarely defined. In addition, some researchers (e.g., Stevens & Palocsay, 2012; Wong, 2009), teachers and citizens use—often implicitly—a definition of a histogram that deviates from what statisticians refer to as a histogram (e.g., Cooper & Shore, 2010; Friel, Curcio, & Bright, 2001). In the statistics literature (e.g., Bruno & Espinell, 2009; Cooper & Shore, 2010; Pearson, 1895; Shaughnessy, 2007), a regular histogram is defined as a graph with bars that meets the following criteria (see Fig. 1 for an example and a non-example):
The data of only one statistical variable are presented on the horizontal axis;
The data are measured at interval or ratio measurement level;
The variable is preferably continuous;
The vertical axis typically displays the class density, or—when bin widths or class intervals are equal—relative frequency or frequency.

Connected bars are neither a defining nor a distinctive feature to distinguish histograms from other graphs with bars (e.g., Ioannidis, 2003; Rufilanchas, 2017).

At first sight, histograms may appear easy to understand, but research indicates otherwise (e.g., Lem, Onghena, Verschaffel, & Van Dooren, 2014). In fact, many errors, misconceptions and mistakes in interpreting histograms have been documented in the literature (e.g., Bruno & Espinel, 2009; Derouet & Parzysz, 2016; Friel et al., 2001; Kaplan, Gabrosek, Curtiss, & Malone, 2014; Lem, Onghena, Verschaffel, & Van Dooren, 2013b). However, a systematic overview of these misinterpretations—a term we use as an umbrella for the ways in which people interpret histograms incorrectly—has not yet been compiled. Research repeatedly showed the persistence of the misinterpretations, despite various attempts to improve statistics education (e.g., Ben-Zvi & Garfield, 2004; Ben-Zvi et al., 2017; Chance, del Mas, & Garfield, 2004; Cohen, 1996; Garfield & Ben-Zvi, 2007; Kaplan et al., 2014; Shaughnessy, 2007). Hence there is a need to reflect on what conceptual difficulties may lie at the basis of these persistent misinterpretations. The aim of this review is therefore to make an inventory of the misinterpretations that occur when people use histograms, as well as to categorise these misinterpretations along the conceptual difficulties that become manifest in these misinterpretations. In this review, the word people refers to students of all levels, as well as to teachers, researchers, teacher educators and others. The question guiding this review is: What are the conceptual difficulties that become manifest in the common misinterpretations people have when constructing or interpreting histograms?

Organizing misinterpretations by conceptual difficulties that may lead to them seems to have several advantages. First, it provides a better understanding of the misinterpretations (in terms of types or common difficulties). Secondly, once the conceptual difficulties that become manifest in the most common misinterpretations are made plausible, researchers and educators can address these more generally instead of treating or remediating misinterpretations one by one. Such a pedagogical route would be in line with the current view in statistics education, which aims to ensure that students develop an understanding of the big ideas of statistics in relation to each other. In the statistics education literature, the term big ideas is often used instead of the term key concept. We use key concept and big ideas as interchangeable terms. Up to now, research usually focuses on a specific misinterpretation (e.g., of the standard deviation) instead of multiple misinterpretations that together are a manifestation of a conceptual difficulty with a big idea or key concept (e.g., of the distribution). Thirdly, this overview is useful for all researchers in education and not only for mathematics education research, because many education researchers either use statistics to analyse the results of their research or teach statistics in—for example—Psychology or Education. For researchers using statistics, graphing is a logical first step in analysing quantitative data—for instance when looking at the distribution of univariate data—and is often advised to do before calculating a measure (Ben-Zvi & Garfield, 2004; Pastore, Lionetti, & Altoè, 2017). Fourthly, research indicates that we need subject-topic-specific information for certain aspects of teaching and learning of this topic (Leinhardt, Zaslavsky, & Stein, 1990; Pareja Roblin, Schunn, & McKenney, 2018).
2. Theoretical background

2.1. Graphical representations

Statistical graphs often serve the analysis of data or enquiry—as Gal (2002) phrases it—and communication of results. This requires graph comprehension (Curcio, 1981, 1987; Friel et al., 2001). Difficulties with graphical representations have been extensively studied (e.g., Arcavi, 2003; Carpenter & Shah, 1998; Larkin & Simon, 1987; Leinhardt et al., 1990; Tufte, 1983/2001; Tversky, 1997). Statistical graphs represent not only data but also statistical concepts—especially graphs that represent data in an aggregated form (e.g., boxplots and histograms). In turn, statistical concepts are inextricably represented in some form—sometimes numerically, sometimes graphically, or both. For example, for most people the concept of the normal distribution is inextricably connected to the bell shape as a graphical representation (Bakker & Hoffmann, 2005).

In this review we therefore focus on the relation between the graphical representation and big ideas in statistics (statistical key concepts). Although some misinterpretations might be unique to the graphical representation itself, we anticipated that most misinterpretations would be a manifestation of a conceptual difficulty. Some conceptual difficulties may appear with other graphical representations too, for example with a boxplot, which is also a graphical representation of univariate data measured at interval or ratio measurement level (e.g., Bakker, Biehler, & Konold, 2005; Lem et al., 2013b).

The literature on statistical graphs revealed that experts and novices analyse graphs in different ways. Experts tend to view a graph globally, while novices seem to focus on local features of the graph (Khalil, 2005). Konold, Higgins, Russell, and Khalil (2015) showed that some elementary school students regard the data as a pointer to the context or situation, which is in line with the findings from other researchers that students see the graph as a picture (e.g., Friel et al., 2001; Leinhardt et al., 1990). According to Konold et al., other students have their focus on individual cases in the graph, for example the shortest person, or where a specific person can be found in the graph. Yet other students see the data in a graph as classifiers—for example for the mode or “the winning outcome” (p. 314). Elementary school students rarely see the data as aggregates, meaning that their focus is mostly not on the entire distribution. Which perspective is useful depends on the question posed to the data.

2.2. Misinterpretations and conceptual difficulties

In this review, we distinguish between conceptual difficulties and misinterpretations. In line with other research (Lem et al., 2013b), we use the term misinterpretation to denote a repeatable and explicit mistake or error that occurs in different people (Leinhardt et al., 1990) and that relates to the conclusion being drawn from a given graph. The term conceptual difficulty is widely used in the literature on physics and chemistry education when people have an incorrect, naïve, or incomplete idea of a concept (e.g., Battaglia, DiPaola, & Fazio, 2017; Hammer, 1996; Garnett & Treagust, 1992). As a clear definition was not found in this literature, we define a conceptual difficulty as having not fully grasped or understood the key concept or big idea at hand. People who have fully grasped the key concept are not expected to show misinterpretations when drawing conclusions from graphs. When we identify a misinterpretation, we can therefore conclude that it is a manifestation of a conceptual difficulty.

An example may further clarify the distinction between a misinterpretation and a conceptual difficulty. When college statistics teachers state that a graph has more variability because the graph is bumpier (meaning: more difference in heights of the bars; e.g., Dabos, 2014) they assess the variability of the frequency bars in the histogram instead of the variability of the variable at hand. We infer from this behaviour (i.e., showing a misinterpretation) that these teachers have difficulties with the statistical concept of variability which is part of the big idea of distribution. This is further clarified through the examples in the next section.

2.3. The big ideas of data and distribution

The research on the teaching and learning of statistics identified several big ideas of statistics—key concepts that underlie statistical investigations (Garfield & Gal, 1999)—as the core goals of statistics education. These big ideas encompass several statistical concepts such as trend, model, sample and graphical representation (e.g., Bakker, 2004; Ben-Zvi et al., 2017; Gal, Garfield, & Gal, 1997; Pfannkuch & Ben-Zvi, 2011). The statistical concepts are intricately connected (Bakker & Derry, 2011). Which statistical concepts are at stake depends on the particular context and research question posed to the data. Fig. 2 summarises how the various statistical concepts fit together when it comes to solving a statistical problem involving univariate data that can be represented in a histogram.

During the analysis of the groups of misinterpretations (the axial codes, see section 3.2) it became clear that the usual theoretical framework of big ideas in statistics—a collection of big ideas and the statistical concepts they are related to—lacked a specification of the relationships between these statistical concepts. We therefore propose a network of statistical concepts based on the theoretical framework of big ideas found in the literature (see Fig. 2). As it is unlikely that there is a generic relationship between these statistical concepts, we focused on those relevant for solving statistical problems that may involve the representation of univariate data in a histogram. Our contribution consists of three parts. First, we added connections between the statistical concepts, which led to a coherent network that, from our analysis1, turned out to be relevant. In this network, we outlined how these connections can be

---

1 Ioannidis (2003) uses the word “bucket” instead of “bin.”
2 In some languages the word frequency refers to relative frequency only and the word count is used to address absolute numbers. An example is
understood. Secondly, we linked this network to the statistical investigation by assigning a specific statistical concept to a specific part of the statistical investigation—such as posing a question or collecting data (Wild & Pfannkuch, 1999). Assigning the concepts to the statistical investigation clarifies the consequence of misinterpretations for statistical investigations and inferential reasoning in education and research. Thirdly, we added measurement level and number of variables as separate statistical concepts, as from the grouping of our data it became clear that these concepts were lacking in the existing theoretical framework of big ideas. In addition, we added some statistical concepts that, beforehand, we did not expect to find in this review (e.g., correlation and covariance), because these statistical concepts do not make sense for histograms. For example, correlation is only possible with at least two variables, whereas a histogram depicts only one variable. From the coding (axial codes), it nevertheless became clear that misinterpretations related to this statistical concept sometimes played a role, so we added this to the network.

We now discuss the two key concepts or big ideas that turned out to be most relevant during the analyses phase. The descriptions of these big ideas are taken from Garfield and Ben-Zvi (2004, p. 400):

- **Data**—[...] data represent characteristics or values in the real world [...]  
- **Distribution**—a representation of quantitative data that can be examined and described in terms of shape, centre and spread [variability], as well as unique features such as gaps, clusters, outliers and so on.

Because we know from the literature in this review that the big ideas data and distribution are hard to grasp for most people, we synthesise the main characteristics in two examples.

### 2.3.1. The big idea of data

The big idea of data includes how many variables are depicted in the graph (see the letter F in Fig. 2) as well as the measurement level (nominal, ordinal, interval or ratio) of its attributes (see K in Fig. 2). In Fig. 3, the big idea of data is explained through the example of babies born in a hospital in Queensland, Australia (Dunn, 1999). For our explanation, only two variables of this data set are used: a number referring to each baby girl that was born (instead of her name) and her weight in grams. To visualise these data, a so-called case-value plot or value bar chart is used, which is a special type of bar graph that shows a value (birth weight) for every case (baby girl; see Fig. 3a).

From the case-value plot with two statistical variables (see Fig. 3a)—thus a bivariate distribution—a histogram can be constructed through six intermediate steps. These intermediate steps are needed to tackle one of the most common misinterpretations related to the big idea of data. This misinterpretation—existing not only among students but also among researchers and mathematics teachers—is that the number of axes determines the number of statistical variables measured—thus defining whether the distribution is univariate or bivariate (see Fig. 3) (e.g., Cohen, 1996). In the first step, one variable is removed from the graph. The resulting series of graphs is therefore univariate, including the histogram (see Fig. 3b–g).

During three of the six steps described here, and during a seventh step outside Fig. 3, information reduction⁵ occurs (Gal et al., 1997). The first information reduction is the removal of the names of the baby girls (here anonymised; see Fig. 3b), possibly inducing—for example—the misinterpretation that bars in a histogram can be reordered (e.g., Bruno & Espinel, 2009). The second information reduction occurs when the dots are stacked (see Fig. 3f), possibly inducing—for example—the misinterpretation that only the middle value of the bar is observed (e.g., Biehler, 1997). The third information reduction occurs when the dots are removed from the bars, making it necessary to use a second axis for the height of the bars (density or frequency), possibly inducing the misinterpretation that two statistical variables are depicted instead of one (see Fig. 3g, e.g., Baker, Corbett, & Koedinger, 2002; Dabos, 2014). When bin widths are unequal, another step is needed. A fourth step in information reduction is therefore using frequency density instead of frequency (not shown in Fig. 3) possibly inducing—for example—wrong labelling of the vertical axis (e.g., Derouet & Parzysz, 2016).

---

(footnote continued) found in French textbooks where the word *effectifs* is used for absolute frequency and the word *fréquence* is used for relative frequency (e.g., Derouet & Parzysz, 2016). In English and in our manuscript, the word frequency means absolute frequency.

³ For example, concepts related to hypothesis testing are not included in this network as these did not emerge from our analysis.

⁴ Most of these steps are available in educational software such as VUstat, Tinkerplots, Tableau and InZight. The R-code for making these graphs is available on the website of the first author.

⁵ We prefer information reduction over the term data reduction, as the original data themselves are not reduced—only aggregated—making other aspects, such as patterns in the data, more visible.
A six-step cartoon on how to convert a case-value plot into a histogram; all graphs are plots of the birth weight of baby girls that were born in a hospital in Australia (Dunn, 2018)

**Fig. 3.** Example of the steps needed to convert a bivariate distribution—using the variables birth weight and name—into a histogram, thus a univariate distribution of the variable birth weight Dunn, 2018.
2.3.2. The big idea of distribution

The big idea of distribution encompasses shape, centre and variability (see L–V and part of W in Fig. 2). The distribution depends on the type of data (see D–K in Fig. 2). In line with Cooper and Shore (2010) we argue in this section that shape (part of W in Fig. 2), centre (see M and T in Fig. 2) and variability (see L and S in Fig. 2) are assessed differently, depending on the type of graph at stake (see W in Fig. 2). Identifying the mean and variation in a histogram—a univariate distribution—can be done by drawing a vertical line for the mean and examining the horizontal spread of the bars, meanwhile taking the heights of the bars into account, see Fig. 4, left. Identifying the mean and variation or spread in a case-value plot—a bivariate distribution—can be done by drawing a horizontal line for the mean and examining the variation of the heights of the bars around this line, see Fig. 4, right. Note that in a histogram less variation in the heights of the bars often indicates more variability of the variable represented on the horizontal axis (here: weight), whereas more variation in the heights of the bars in a case-value plot always indicates more variability of the variable represented on the vertical axis (here: weight). Although the graphical representations in Fig. 4 look quite different, the underlying distribution of the variable at hand (weight) is the same. This big idea of distribution is often misunderstood as people tend to think of a distribution as the shape of the graph and not as an abstract statistical concept leading to, for example, not recognising different graphical representations of the same data (e.g., delMas, Garfield, Ooms, & Chance, 2007).

3. Method

A narrative systematic review of the literature with a configurative synthesis was conducted (Gough, Oliver, & Thomas, 2017), with a query-based search strategy in the following databases: PsycINFO, Web of Science, Scopus, ERIC and Google Scholar, see Fig. 5 for the flowchart. These five databases are commonly used for scientific literature in mathematics and statistics education.

3.1. Search strategy

This paper includes publications that describe or contain misinterpretations when constructing or interpreting histograms by people (students, teachers, researchers and others). A publication was excluded when histograms were only used as a graphical representation of the research data of the study. Statistics textbooks were also excluded. Searching for the keyword histogram not including patents gave more than 260,000 hits (12 May 2016). Most of these hits were not relevant, as many publications only use histograms for presenting the results. Searching in the title of articles only would exclude many relevant publications. Several rounds of checking of keywords for inclusion and exclusion were conducted to arrive at a workable query. The keywords for exclusion were chosen from non-relevant publications and checked in a quick separate search before adding the keyword to the list, to ensure that no important publications were missed. The 300 publications identified during this iterative process were kept. It proved necessary to formulate many keywords for exclusion to avoid a large number of non-relevant publications, see Table 1. We also excluded patents and citations. This query resulted in another 299 publications to be checked, resulting in a total of 599 publications identified in Google Scholar. As every database has its own search engine with different options, we had to slightly adapt the search strategy. Table 2 provides examples of how this was done for PsycINFO. For the other databases, a similar procedure was used. Where possible, we searched in title, abstract, keywords and topic or category. The search in all those databases resulted in more than 1000 publications to be checked. The procedure depicted in Fig. 5 resulted in 86 publications that are enclosed in this review. As we searched for the most common misinterpretations, in addition to backward snowballing, we applied a checking procedure at the end of the search until saturation occurred, to make sure that we did not miss any key publications or common misinterpretations.

3.2. Data analysis

For every publication included in this review, we collected the misinterpretations that were either reported or detected in the publication. To identify the conceptual difficulties that become manifest in the most common misinterpretations, we grouped these misinterpretations into axial codes. Using the big ideas and the statistical concepts that they encompass as a lens, we inferred through abduction (Peirce, 1994) that misinterpretations stem from a lack or misunderstanding of these concepts or big ideas. Abduction is the process of generating explanatory hypotheses. Hoffmann (2011) states that we can stop this process “when an abductive insight has been achieved” which he defines as “the experience that what someone created in abductive reasoning is plausible and gives an acceptable argument for the phenomenon” (p. 572). As explained in section 2.2, the following holds. People who have fully grasped a key concept or big idea are not expected to show misinterpretations when drawing conclusions from graphs. When we identify a misinterpretation, we can therefore conclude that it is a manifestation of a conceptual difficulty.

How the network of statistical concepts was used, is now explained with two examples. The first example is the misinterpretation of students who used two statistical variables when asked to draw a histogram (Baker et al., 2002). This misinterpretation is categorised as indicating a problem with understanding the big idea of data (see box F in Fig. 2: 1, 2 or more variables and attributes), as it indicates that these students do not differentiate between a histogram—which represents a univariate distribution of one variable—and a bivariate distribution of two statistical variables (the latter often being depicted in a scatterplot). A second example is students who do not understand that a distribution that looks unimodal in a histogram can turn out to be bimodal if the bin width is made smaller (Karagiannakis, 2013). This misinterpretation is categorised as indicating a problem with understanding the big idea of

---

6 Our unit of analysis is publications, not studies.
distribution, as it indicates that these students do not understand the influence of grouping on the graphical representation, displayed by the arrow from grouping (see box R in Fig. 2: group or ungroup) to graphical representation (see box W in Fig. 2: graphical representation: graph with bars, histogram). As further explained in the codebook (see online extra materials for the full version), the selective code grouping was assigned here.

We used open, axial and selective coding (Corbin & Strauss, 1990) to cluster the identified misinterpretations exhaustively and mutually exclusively into three categories: (1) data-related conceptual difficulties, (2) distribution-related conceptual difficulties and (3) miscellaneous. Three examples of axial codes (a group of misinterpretations) are: ‘larger frequency thus larger mean’, ‘bell-
shaped = histogram’ and ‘bumpier = higher variability’. From these axial codes, the selective codes were created through abduction from the network of statistical concepts (see Fig. 2). Provided with the codebook and the open codes (description of what was reported or found in the publication) and axial codes (the first grouping of the misinterpretations), an external coder was asked to assign one of eleven selective codes to the description of the misinterpretations. Of the more than 300 descriptions of misinterpretations (open codings), 73 were coded by the first author and an external coder. The interrater reliability—Cohen’s kappa—was 0.84, suggesting a reliable coding procedure with “almost perfect” agreement (Landis & Koch, 1977, p. 165). A summary of the codebook is given in Table 3; an extended version can be found in Table A1 in the Appendix of this article. The full version is available as online material (Extras).

The selective codes in the codebook categorise the misinterpretations at the level of a specific concept that were then merged into three categories of conceptual difficulties. At this final level, categories summarise whether the conceptual difficulties that become manifest in the misinterpretations are related to the data represented, or related to the distribution represented, or neither of these two (miscellaneous). The level of selective codes identifies subcategories of specific concepts that are misinterpreted. These subcategories are characterised briefly in the last column of the codebook and are illustrated with types of misinterpretations listed. The characterisation ends with a note when not to assign this code, so as to make the second coder aware of the boundaries of a particular code (subcategory) (Boyatzis, 1998).

Some misinterpretations are possibly caused by the translation into English. In English, different words are created to distinguish histograms (one variable; numerical measurement level, see Fig. 1, left) and distribution bar graphs (one variable; categorical measurement level, see Fig. 1, right) on the one hand from case-value plots (see Fig. 3a; two variables) and time-plots (also two variables) on the other. Other languages may lack such different words. Several researchers refer to a graph with bars as a histogram while it is not. If this misinterpretation was held by researchers from non-English-speaking countries, it might be due to translation only. Therefore, these specific misinterpretations were excluded from the results (Kramarski, 1999; Mevarech & Kramarsky, 1997).

4. Results

The results show that the conceptual difficulties that become manifest in the most frequently reported misinterpretations fall into three different categories: data-related, distribution-related and other. The misinterpretations that are a manifestation of difficulties with the concept of data include: not understanding how many statistical variables are depicted in a histogram (only one) and not understanding that a histogram is suitable for numeric variables only (see Fig. 2 F and K). The misinterpretations that are a manifestation of difficulties with the concept of distribution include: not knowing how shape (part of W, Fig. 2), centre (see Fig. 2 M and T) and variability (see Fig. 2 L and S) are depicted in a histogram and not understanding the effect of grouping into bins in a histogram (see Fig. 2 R). In line with Bakker and Hoffmann (2005) our research shows that these two conceptual difficulties cannot be isolated from their sign, the histogram. The third category of miscellaneous conceptual difficulties is more loosely related to the sign—the histogram—and entails difficulties that occur due to the software used, to confusion on whether the sample or the population is depicted in the histogram and to the context. The most common misinterpretations resulting from these conceptual difficulties are elaborated further in the next sections. Table 4 and 5 give an overview of the publications included in this review. The full details of

| Table 1 |
| Keywords for inclusion | Keywords for exclusion |
|------------------------|------------------------|
| Histogram              | Automated              |
| Mistake                | DNA                    |
| Education              | Neural                 |
| Literacy               | Radiology              |
| Bank                   | Driver                 |
| Bayesian               | Forensic               |
| Compression            | Genetic                |
| Disability             | MRI                    |
|                       |                       |

Note: Using four key words for inclusion and none for exclusion led to zero publications identified, so the search strategy had to be slightly adapted by using fewer key words.

| Table 2 |
| Example of searching strategies, keywords and number of identified and relevant publications in database PsycINFO. |
| Database | Search in | Key words for inclusion | Number of publications identified | New relevant publications |
|----------|-----------|--------------------------|-----------------------------------|--------------------------|
| PsycINFO | Abstract  | Histogram and mistake and education and literacy | 0 | 0 |
|          |           | Histogram and education and literacy | Almost 600, so more key words were used | 8 | 1 |
|          | Abstracts of predefined category 2240 statistics & mathematics | Histogram and education and literacy | Over 200, so more key words were used | 40 | 0 |
|          | Title     | Histogram | 43 | 0 |
|          | Total     | | 91 | 1 |

Note: Using four keywords for inclusion and none for exclusion led to zero publications identified, so the search strategy had to be slightly adapted by using fewer key words.<ref>100291</ref>
all misinterpretations can be found in the data paper (Boels, Bakker, & Drijvers, in preparation); in the online extras some more summaries of the findings are given. The misinterpretations described or detected in the publications, including almost 16,000 students, teachers and researchers, are incorporated in this review. This includes slightly over 400 primary school students, almost 7,000 secondary school students and approximately 8000 college and university students. The remainder includes college statistics teachers, mathematics teachers and researchers. Most participants are from the USA (see Appendix A).

4.1. Misinterpretations related to difficulties with the concept of data

4.1.1. Identifying the measured variable only

As explained in the theoretical background, by definition a histogram displays the distribution of one statistical variable. Seven publications reported or showed misinterpretations regarding the measured variable. A widespread misinterpretation is that a histogram could display the data of two variables, which was reported or found in nine sources (e.g., Cohen, 1996; Gilmarin & Rux, 2000; Meletiou & Lee, 2002; Meletiou-Mavrotheris, 2000; Stevens & Palocay, 2012; Zaidan et al., 2012) and which is related to the misinterpretation that the number of bars is seen as the number of cases (Dabos, 2014; Ismail & Chan, 2015; Sorto, 2004). Other often found misinterpretation is that the frequency is seen as the measured value (Baker, 2004; Chance et al., 2004; delMas & Liu, 2005; Friel & Bright, 1996; Kaplan et al., 2014; Lem et al., 2013b) and that the horizontal axis is seen as a timescale when it is not (Dabos, 2014; Kaplan et al., 2014; Meletiou & Lee, 2002; Meletiou-Mavrotheris, 2000; Zaidan et al., 2012). This confusion is aggravated as frequency and number (count) are commonly interchangeable terms. The definition of a histogram nevertheless implies that the vertical axis depicts the frequencies or number counts of the measured values that are depicted on the horizontal axis. Consequently, a time-plot—with, for example, years on the horizontal axis—is not a histogram, as it is nonsensical to count how often a year occurs in a year (see Fig. 6). Furthermore, it is often stated that the bars of a time-plot must be connected when intervals are consecutive, but this is only true for histograms.

### Table 3

| Phase | Conceptual difficulty | Selective codes |
|-------|-----------------------|-----------------|
| Orientation on histogram | Data-related | Number of variables (F, H) or measurement level (K) or both (K, F, H). |
| Interpreting histogram | Distribution-related | Variability (L, S), centre (M, T), shape (W) and grouping (R). |
| In review but not included in results | Miscellaneous | Context (A), Population (B), ICT or unknown. |

* ICT is found along the arrows from population to a sample. Only where relevant for this review ICT is indicated.

### Table 4

| Misinterpretations related to difficulties with the concept of data | Misinterpretations related to difficulties with the concept of distribution |
|---------------------------------------------------------------|----------------------------------------------------------------------------|
| Abrahamson & Wilensky, 2007; Agro, 1977; Baker, Corbett, & Koedinger, 2001, 2002; Bakker, 2004; Bruno & Espinel, 2009; Capraro, Kulm, & Capraro, 2005; Chace et al., 2004; Clayden & Croft, 1990; Cohen, 1996; Cooper & Shore, 2008; Corredor, 2008; Dabos, 2014; delMas, Garfield, & Ooms, 2005; delMas et al., 2007; Derouet & Parzysz, 2016; Enders, 2013; Estab & Schwartz, 2002; Friel & Bright, 1996; Gilmarin & Rux, 2000; Hawkins, 1997; Humphrey, Taylor, & Mitig, 2014; Ismail & Chan, 2015; Kaplan et al., 2014; Kramarski, 2004; Kulm, Dager, Wilson, & Kitchen, 2005; C. Lee & Meletiou-Mavrotheris, 2003; Lem et al., 2013b; McKinney, 2015; Meletiou & Lee, 2002; Meletiou-Mavrotheris, 2000; Redfern, 2011; Ruiz-Primo, Li, Yu, & Shavelson, 1999; Sorto, 2004; Stevens & Palocay, 2012; Stone, 2006; Strasser, 2007; Tiefenbruck, 2007; Watts et al., 2016; Whitaker & Jacobbe, 2017; Wong, 2009; Yun, Ko, & Yoo, 2016; Zaidan, Ismail, Yusof, & Kashefi, 2012. | Baker et al., 2001; Batanero, Tauber, & Sánchez, 2004; Biehler, 1997; Bruno & Espinel, 2009; Capraro et al., 2005; Chan & Ismail, 2013; Chace et al., 2004; Cohen, 1996; Cooper & Shore, 2008; Cooper & Shore, 2010; Corredor, 2008; Dabos, 2014; delMas & Liu, 2005; delMas et al., 2005; delMas et al., 2007; Dabos, 2014; DelMas & Liu, 2005; delMas et al., 2007; delMas et al., 2007; Derouet & Parzysz, 2016; Friel & Bright, 1995, 1996; Gilmarin & Rux, 2000; Gonzalez, 2014; Higg, 2015; Ismail & Chan, 2015; Kaplan et al., 2014; Kaplan, Fisher, & Rognes, 2009; Karaiskakis, 2013; Kelly, Sloane, & Whitaker, 1997; Kondol, Pollatsek, Well, & Gagnon, 1997; Kuklinsky, 2016; Kulm, Dager Wilson, & Kitchen, 2005; C. Lee & Meletiou-Mavrotheris, 2003; T. Lee, 1999, Lem, Ongena, Verschaffelt, & Van Dooren, 2011, 2013; Lem et al., 2013b, 2014, Madden, 2008; Martin, 2003, McGatha, Cobb, & McClain, 2002; McKinney, 2015; Meletiou & Lee, 2002; Meletiou-Mavrotheris, 2006; Meletiou-Mavrotheris & Lee, 2005; Mervarech & Kramarsky, 1997; Oulide, 2014; Roth, 2005; Rumsey, 2002; Sorto, 2004; Stevens & Palocay, 2012; Stone, 2006; Tiefenbruck, 2007; Tureg & Rieger, 2011; Vernet & Guttuso, 2014; Whitaker & Jacobbe, 2017; Whitaker, Foti, & Jacobbe, 2015; Wong, 2009. |

Some statistics educators prefer the more general term of “attribute” (W. Finzer, personal communication, July 12th, 2018). As other people may think that attribute only refers to a nominal measurement level, we avoided this term.

See also our footnote in the introduction on the influence of language on the interpretation of the term frequency.

And some researchers also use separate bars in histograms, e.g., Ioannidis (2003).
4.1.2. Identifying the measurement level only

Eighteen publications reported or contained misinterpretations regarding the measurement level. Five of these publications reported people referring to a normal distribution—which is only possible for numerical data—while the measurement level of the data was nominal or ordinal (delMas et al., 2007; Humphrey et al., 2014; Kaplan et al., 2014; Redfern, 2011; Whitaker & Jacobbe, 2017). Nine publications reported or contained ‘histograms’ with nominal or ordinal measurement level (e.g., Stone, 2006; Tiefenbruck, 2007; Watts et al., 2016; Wong, 2009). People showing this misinterpretation may consider the blood type graph (see Fig. 1) as ‘right skewed’ or ‘not normally distributed’. These people overlook that the measurement level is nominal and therefore the bars are not in scale order and that the theoretical model of a normal distribution is therefore not applicable.

Three publications identified the misinterpretation that the interval is a ‘label’ with for example students and authors of schoolbooks treating this label as a nominal measurement level (see Fig. 7), neglecting the numerical scale (Bruno & Espinel, 2009; Derouet & Parzyzs, 2016; Humphrey et al., 2014).

Another misinterpretation is the use of histograms for Likert scales when words combined with numbers are used. An example of how seriously this can go wrong when used by non-statisticians can be found in McKinney (2015) where the following strange attribution for a 5-point Likert scale is used: none at all (1), very little (2), strong degree (3), quite a bit (4) and a great deal (5) for a Self-Efficacy Scale for Teaching Mathematics Instrument (SETMI). This SETMI was developed by McGee (2012) and is used in several other studies (e.g., McCampbell, 2014).

4.1.3. Identifying the measured variable and the measurement level

Seventeen publications reported or contained misinterpretations regarding both the number of variables and the measurement level. The most often reported or found misinterpretation (10 publications) is that people think that there is no difference between a histogram and a bar graph, or that the only difference is that bars are connected in a histogram, neglecting the required measurement level (Capraro et al., 2005; Clayden & Croft, 1990; Eshach & Schwartz, 2002; Gilmartin & Rex, 2000; Humphrey et al., 2014; Kramarski, 2004; Kulm et al., 2005; Sorto, 2004; Stevens & Paloccsay, 2012; Tiefenbruck, 2007). Six publications contained or reported the misinterpretation that a histogram could be used for nominal or ordinal data and two variables (Baker et al., 2002, 2001; Dabos, 2014; delMas & Liu, 2005; Eshach & Schwartz, 2002; Ruiz-Primo et al., 1999). Four publications reported the misinterpretation that bars could be rearranged in a histogram, for example from highest to lowest bar (Dabos, 2014; Humphrey et al., 2014; Kaplan et al., 2014; Whitaker et al., 2017).

4.2. Misinterpretations related to difficulties with the concept of distribution

As explained in section 2, theoretical background, the number of measured variables as well as the measurement level define the type of graphical representation, which in turn influences the interpretation of the distribution: shape, centre and variability. For example, variability can be seen as weighted deviation from the arithmetical mean (Cooper & Shore, 2010). In a case-value plot with nominal data on the horizontal and numerical data on the vertical (two measured variables), the relevant measured value is on the vertical axis and variability can be seen as variation in the heights of the bars. In a histogram, the only measured value is on the horizontal axis and therefore the horizontal spread of these measurements must be taken into account—in combination with the heights of the bars. Several studies report that students and teachers confuse variation in the frequencies in a histogram—the heights of the bars—with variation in the measured value—hence the variability in a histogram (e.g., Lem, Onghena, Verschaffel, & Van Dooren, 2013a). In this section four groups of misinterpretations are reported, regarding: variability, centre, shape and data reduction through grouping (Fig. 7).

4.2.1. Variability

Twenty-six publications reported on misinterpretations regarding the statistical concept variability or regarding the variability combined with the statistical concepts centre and/or shape. Eleven publications reported the misinterpretation that a higher difference in the heights of the bars only implies more variation in the data (Cooper & Shore, 2008; delMas et al., 2007; Kaplan et al.,...
Eight publications reported an overgeneralisation of the idea that a certain shape (normal, uniform or symmetrical distribution) has the highest or lowest variability (Cooper & Shore, 2010; Dabos, 2014; González, 2014; Kaplan et al., 2014; Meletiou-Mavrotheris & Lee, 2005; Turegun & Reeder, 2011; Vermette & Gattuso, 2014; Whitaker & Jacobbe, 2017). Range can be regarded as a simple or preliminary measure of variability, especially for secondary school students. Seven publications reported misinterpretations about the variability in the data when range was used (Cooper & Shore, 2008; Dabos, 2014; Kaplan et al., 2014; Lem et al., 2013b; Madden, 2008; Meletiou-Mavrotheris & Lee, 2005; Olande, 2014) and two misinterpretations about variability and centre, when range was used (Kukliansky, 2016; Lem et al., 2013a). Various misinterpretations regarding the standard deviation in a histogram are reported, including that a certain shape, or ordering of the bars (e.g., ascending or descending heights) leads to the largest or smallest standard deviation, that a larger mean implies a larger standard deviation and that gaps between bars (frequency zero) do not influence the standard deviation (delMas & Liu, 2005).
the misinterpretation that standard deviation and mean in a histogram are the same (Chan & Ismail, 2013) or that once the means in both histograms are the same, the standard deviation is the same as well (Kukliansky, 2016). Misinterpretations regarding variability are also found among teachers (e.g., González, 2014). Variability is the variation of the data, for example around the mean, see Fig. 8. As the mean is depicted differently in a histogram than in a case-value plot, the variability also has to be assessed differently. In a case-value plot, the variability is the variation in the heights of the bars. In a histogram, the variability is the weighted horizontal spread of the bars.

4.2.2. Centre

Thirteen publications reported on misinterpretations regarding the statistical concept of centre. Four publications reported a misinterpretation where the mean of the frequencies (vertical axis) was used instead of the mean of the measured values of the variable (horizontal axis, see Fig. 8 for more explanation) (Cooper & Shore, 2008; Lem et al., 2013a,b, 2014). Five publications reported a similar misinterpretation regarding the median (Cooper & Shore, 2008; Ismail & Chan, 2015; Kaplan et al., 2014; Lem et al., 2013a), the mode (Huck, 2016; Ismail & Chan, 2015; Kaplan et al., 2014), or both (Kaplan et al., 2014). All these misinterpretations are related to the type of graphical representation, as it depends on the type of graph whether the frequency is a statistical variable or not. For example, in a time plot the frequency is the measured value (see Fig. 6). Other misinterpretations include that the median is seen as the middle class (Stevens & Palocsay, 2012), that it is seen as the midpoint of the scale on the horizontal axis, or as the midrange (Cooper & Shore, 2008).

In many Introductory Statistics Courses, rules of thumb are taught for the position of mean and median in relation to the skewness of the distribution (thus the shape in the histogram). One such rule of thumb is that the mean is typically lower than the median in left or negative skewed distributions. Although this holds true in many situations, Huck (2016) states that this was helpful when people lacked strong computers, but nowadays these kinds of rules are no longer needed, as they can also mislead us when analysing results. Huck claims that: “Unfortunately, the application of those rules can make one think data are skewed left when they are skewed right (or vice versa)” (p. 26). Therefore, we carefully need to reconsider questions that test, for example, if students know the rule of thumb that the mean is bigger than the median in right-skewed distributions (Cooper & Shore, 2008; delMas et al., 2007; Karagiannakis, 2013; C.; Lee & Meletiou-Mavrotheris, 2003; Whitaker & Jacobbe, 2017).

4.2.3. Information reduction through grouping

People have difficulties with the information reduction (Gal et al., 1997) present in histograms. As explained in the theoretical background, one step in information reduction is that several values are grouped into one bin. Bakker (2004) already pointed out that this grouping is difficult for students in grades 7 and 8. Fifteen publications reported or contained misinterpretations regarding the grouping in bins. Misinterpretations include not using or mentioning density for unequal bin width (Derouet & Parzysz, 2016; Gilmartin & Rex, 2000; Kelly et al., 1997; McGatha et al., 2002) and choosing a wrong bin width or wrong boundaries for the bins (Bruno & Espinel, 2009; delMas et al., 2005; Martin, 2003; McKinney, 2015; Whitaker & Jacobbe, 2017). Three publications reported misinterpretations regarding the measured values, either that all possible values in a bin are measured (Lem et al., 2013b; Meletiou-Mavrotheris, 2000) or that only the middle value of a bar is measured (Biehler, 1997).

Fig. 8. Centre and thus variability is assessed differently in a histogram (left) compared to a case-value plot (right).
4.2.4. Shape

Twenty-eight publications reported or contained misinterpretations concerning the graphical representation of a histogram itself. Six reported that students cannot link a histogram to a corresponding boxplot (Corredor, 2008; delMas et al., 2005; delMas et al., 2007; Karagiannakis, 2013; Lem et al., 2011, 2015). Ten reported or contained misinterpretations regarding graph conventions (Baker et al., 2001; Batanero et al., 2004; Bruno & Espinel, 2009; Lem et al., 2013b; Martin, 2003; McGatha et al., 2002; Mevarech & Kramarsky, 1997; Roth, 2005), for example that connected bars are for easier comparison (Capraro et al., 2005; Kulm et al., 2005). Some authors state that histograms are not suitable for discrete variables (Batanero et al., 2004; Cohen, 1996; Friel & Bright, 1995, 1996; Tiefenbruck, 2007). However, even though variables can be either discrete or continuous in theory, in practice data sets that represent them are always discrete, due to the limited accuracy of the measurement instrument. Also, even discrete variables can meet some interval and ratio measurement criteria. For example, the discrete variable "the number of siblings in your family", the value six indeed is twice as large as the value three. Therefore, we decided not to exclude discrete variables. Students using graphs with poles instead of bars can be found in McGatha et al. (2002).

4.3. Misinterpretation related to miscellaneous concepts

In addition to the two aforementioned categories, there are less frequent miscellaneous difficulties that can be summarised as: not understanding the histogram in relation to the given context, not understanding the difference between a histogram of a sample and a histogram of a population, and the influence of ICT (ICT often does not differentiate between histograms and other types of graphs with bars). Some descriptions in publications do not provide enough details for specifying the type of misinterpretation and are classified as unknown (Baker et al., 2001; Behrens, 1997; Biehler, 1997; Carrión & Espinel, 2006; Chance et al., 2004; Konold et al., 1997; Shaughnessy, 2007; Yun & Yoo, 2011).

4.3.1. Context

Nine publications reported misinterpretations due to the context. One example of a misleading context is height (Whitaker & Jacobbe, 2017), as students in this specific context more easily interpret the height of the bars in a histogram as the measured height leading to the confusion of a case-value plot with a histogram. The misinterpretation of a time scale on the horizontal axis can sometimes also stem from the context and is described in subsection 4.1.1 Identifying the measured variable only. Furthermore, students and teachers occasionally use context knowledge or personal experience instead of the data (Friel et al., 2001; Madden, 2008; Shaughnessy, 2007). The opposite equally occurs: students having trouble linking the histogram to the original data collection or context (delMas et al., 2005; Yun & Yoo, 2011). This is in line with research from Kaplan, Lyford, and Jennings (2018) who showed that students’ descriptions of histograms systematically differ, depending on the specific wording of the question (including the word distribution or variable or both in the question) as well as the context (income or hours of sleep).

4.3.2. Sample or population?

Seven publications reported misinterpretations regarding the population. Five of these report the misinterpretation that the histogram of a sample and the histogram of a population have the same properties—for example the same shape or distribution (Chance et al., 2004; Hawkins, 1997; Slauson, 2008; Stone, 2006; Whitaker & Jacobbe, 2017). Not distinguishing sample and population might also lead to ignoring the effect of random noise (Biehler, 1997; Nuhfer et al., 2016).

4.3.3. Influence of ICT

Although ICT can be a helpful tool to understand statistics, it can also introduce new misinterpretations. The most common misinterpretation is embedded in the software where no distinction is made between a histogram and a bar graph (Hawkins, 1997), often leading to histograms with strange or even wrong boundaries of the bins (Abrahamson, 2006, 2008, 2009; Abrahamson & Cendak, 2006; Abrahamson & Wilensky, 2007; McKinney, 2015; Prodromou & Pratt, 2006). Two publications reported the misinterpretation that the number of classes is fixed, possibly due to a fixed number of classes in the software (Cohen, 1996; Yun et al., 2016).

5. Conclusion and discussion

In this review, the aim was to make a systematic inventory of the misinterpretations that occur when people use histograms, as well as to categorise these misinterpretations along the conceptual difficulties that become manifest in them. It turned out that the most common conceptual difficulties could be related to two big ideas in statistics: data and distribution. The category misinterpretations that are related to the difficulties with the key concept of data includes misinterpretations about the number of variables depicted in a histogram and the measurement level of the data, including the wrong application of theoretical models. The category of misinterpretations that are related to difficulties with the key concept of distribution includes misinterpretations about variability, centre, shape and information reduction through grouping. The third and more diverse category of misinterpretations is related to other conceptual difficulties and includes having trouble to link the context to the histogram, not understanding the difference between a histogram of a sample and of a population, and the influence of ICT. The analysis of the publications in our
review also led to the identification of a network of statistical concepts specific to interpreting histograms, see the theoretical background section. From our analysis, it furthermore became clear that two statistical concepts needed to be added to the big idea of data: number of variables and measurement level. These two concepts were not yet explicitly part of the collection of big ideas in statistics.

Furthermore, our review study reveals that most publications investigate students’ or teachers’ notions of shape and variability, which is an important topic for college and university students. Hence, these publications focus on misinterpretations that are related to difficulties with the big idea of distribution. Although misinterpretations regarding identifying the number of variables and the measurement level of its attributes are more often observed, research specifically addressing these misinterpretations is scarce. The latter two sub-categories of misinterpretations are related to difficulties with the big idea of data. The data-related conceptual difficulties may be underlying the distribution-related conceptual difficulties, as the data (number of statistical variables and measurement level) define the type of graph, and in turn how variability and centre are depicted in the graphical representation (e.g., Cooper & Shore, 2010). We speculate that the persistence of people’s misinterpretations of histograms is partly due to overlooking the impact of data-related conceptual difficulties. This might also result in underreporting of misinterpretations regarding data-related conceptual difficulties, as well as misinterpretations regarding shape and centre.

Our findings are in line with findings about mathematical graphs from Leinhardt et al. (1990), such as the tendency to overgeneralise. An example of overgeneralisation is the idea that the number of axes is the number of measured variables (true for a case-value plot, but false for a histogram; see: subsection 4.1.1. identifying the measured variable only). Another example is the overgeneralisation of the effect of shape (e.g., uniform distribution) on variability (see: subsection 4.2.1 variability) and of theoretical models (normal distribution, see: subsection 4.1.2 identifying the measurement level only). Leinhardt et al. also found an interference with the context or daily life observations (see: subsection 4.3.1 context).

According to Friet al. (2001), the basic level of reading the data is often not very difficult for students for most graphs. This may be true for reading off a particular value, but our review shows that many misinterpretations are related to the data depicted in a histogram, hence to reading the data (thus the big idea of data). In addition, during the application of the theoretical framework of big ideas it became clear that not only the big ideas and statistical concepts are important, but also the connections between them, as, for example, the grouping in bins influences the shape of the distribution, thus the graphical representation of the data. We therefore proposed a coherent network of statistical concepts relevant for research questions that may involve the interpretation of histograms (see section 3.2 data analysis).

Systematic reviews of the literature have limitations. A geographical selection bias seems to exist. A large proportion of the studies in this literature review was carried out in the United States, followed by European countries (see the Appendix). The English-speaking countries generally pay more attention to statistics in their curriculum than other countries (e.g., Franklin, 2019). This suggests that the problem may be bigger than what was found here. We do not want to suggest representativeness, as we were mainly interested in the types of conceptual difficulties that become manifest when people (students, teachers, researchers and others) interpret histograms.

Furthermore, we speculate that the misinterpretations identified in this literature review also hold for Asian, African and South American countries, as well as for Australia. The reasons for this speculation are that in some countries statistics is not yet or only recently part of the curriculum, for example in Thailand, (e.g., Burrill & Ben-Zvi, 2019; Franklin, 2019; González & Chitmun, 2019) and there are some, although not yet many, studies of Asian countries indicating misinterpretations when interpreting histograms (Ismail & Chan, 2015; Yun et al., 2016; Yun & Yoo, 2011).

Several implications for future research and education arise from this review. The first is that data-related conceptual difficulties seem to be understudied, and therefore would require more explicit attention from researchers. Ignoring the difficulties with the concept of data may possibly explain the persistence of misinterpreting histograms. Researchers, teachers and teacher trainers are encouraged to be more aware of the differences regarding data (number of variables, measurement level) and distribution between a case-value plot, a distribution bar graph and a histogram and the consequences of these differences for shape as well as assessing variability and centre. Furthermore, for languages that lack distinct words for case-value plots, time-plots, distribution bar graphs and histograms, our advice is to create and introduce those words in research and implement them in the statistics education curriculum from primary school level up to university level, as this will support the awareness of the differences. In addition, this literature review adds to the framework of big ideas in statistics education that there is a hierarchy in those big ideas. The big idea data (number of variables and measurement level) is fundamental for a deep understanding of the big idea distribution as shape, centre and variability are depicted differently in different types of graphical representations.

The second implication is that the role of information reduction seems to be understudied (see: section 2. theoretical background). The literature on information reduction is very scarce. Bakker (2004) is one of the few examples indicating the difficulty of the idea of grouping. Nevertheless, indications for this difficulty are also found in other research (Ismail & Chan, 2015; Lem et al., 2013b; Meletiou-Mavrotheris, 2000; Sorto, 2004). Researchers, teachers and teacher trainers are advised to be aware that information reduction plays an important role in the following four stages when turning a case-value plot into a histogram. The first stage is when one of the measured variables is removed (resulting in, for example, a dotplot). Students who see the data in a graph as a pointer to the situation (Konold et al., 2015) and students who consider a histogram as a case-value plot might not have understood the case information removal phase. The second phase is when the dots in a dotplot are stacked into classes with a certain bin width. People who think that only the middle value of a bar is observed might not have understood this grouping phase. The third phase is when the
dots are omitted from the bar, making it compulsory to use a second axis. People who regard a histogram as a bivariate distribution, might have problems with this third phase. The fourth phase, which is hardly studied, is when the frequency is turned into frequency density. This phase is of key importance for the transition to continuous probability distributions. The other two understudied areas are: the difference between the histogram of a sample and of a population and the influence of the context. These two areas are only loosely related to histograms.

The third implication is that future research is needed in those countries that are not yet included in our review to substantiate our claim that the identified conceptual difficulties can be found all over the world and are not due to a specific way of teaching or an educational system, as a geographical gap seems to exist in the research literature. Also, active promotion of publishing in English journals of work published earlier in other languages is needed to make this literature available for many more researchers, as well as translation of English literature into other languages.

As an implication for task design in research and education, this review makes it clear that items containing graphs with bars without context or labels cannot be identified with regard to the type of graph and must be avoided in schoolbooks as well as assessments and research items.

An implication for education—now that the conceptual difficulties that become manifest in the most common misinterpretations are made plausible—is that researchers and educators can address these more generally instead of treating or remediating misinterpretations one by one. Such a pedagogical route would be in line with the current view in statistics education which aims to ensure that students develop understanding of the big ideas (key concepts) of statistics in relation to each other. Our overview opens up the possibility of systematically dealing with these misinterpretations first in research and eventually in primary and secondary schools and statistics introductory courses, as well as developing and testing materials especially designed to tackle these misinterpretations. Teachers and teacher trainers now have access to an overview of all the common misinterpretations identified in the publications. This adds to their Statistical Knowledge for Teaching (SKT, see Groth, 2007). According to Pareja Roblin et al. (2018), an overview is very important as “positive student outcomes were associated with curriculum materials […] that provide teachers with information about students’ ideas” (p. 260).

One might conclude that histograms are too difficult to use and teach. Can we do without them in education and research? Our answer is no. First, histograms reveal aspects of the distribution that other graphs do not (e.g., Pastore et al., 2017). Secondly, histograms are omnipresent in research and education and should therefore be learned. Thirdly, the alternatives entail some of the same disadvantages such as the height misinterpretation in dotplots (Lyford, 2017), as well as other disadvantages such as an irregular shape (dotplots) or an even more advanced step in information reduction (boxplots). Fourthly, it is the key concepts underlying a histogram that are hard to grasp (the big ideas of data and distribution). Unfortunately, we cannot learn those big ideas without signs (e.g., histogram), as the representation of the data as well as how the distribution manifests itself (through its shape) strongly depends on the specific type of graph with bars, as we explained in subsection 2.3 of the theoretical background section. It is when interpreting histograms that these underlying conceptual difficulties become manifest, making histograms a good diagnostic instrument for teachers and researchers as well.

Acknowledgement

The research is funded with a Doctoral Grant for Teachers from The Dutch Research Council (NWO), number 023.007.023 awarded to Lonneke Boels. The authors thank research assistants Floris Kooij and Rian Ligthart for their contribution to digitalizing the classification in big ideas, the screening of several doctoral theses, the addition of the number of students and countries and the backward snowballing procedure. In addition we thank PhD-student Marianne van Dijke-Droogers for her contribution as a second coder. We also thank Nathalie Kuipers for checking the document on APA style and English. Furthermore, the authors thank ICT developer Kees van Eijden for writing some of the R-code that we used for making the graphs, especially the non-stacked dot plot and the histodot.

Appendix A

Appendix A.1. Shorted version of the codebook

The full version of the codebook is provided in the online extra materials. The codebook is used for classification of the misinterpretations. If misinterpretations regarding distribution related difficulties can be assigned to two subcategories (thus selective codes, see Table A1), the following hierarchy applies: grouping takes priority over all other subcategories, then variability, then measures of centre and finally shape. We made this hierarchy with in mind that, for example, variability is often regarded as a more difficult concept to grasp than centre.
### Table A1

| Category                        | Selective code | Assign this code if                                                                 |
|--------------------------------|----------------|-------------------------------------------------------------------------------------|
| Data-related (DA)               | Identifying the measured variable only (VO) | The misinterpretation is related to identifying what was measured (which variable), but not to the measurement level of this variable. This includes the following misinterpretations:  
  - frequency is regarded as a measured value  
  - it is stated that frequency or number count on the vertical axis implies a histogram  
  - a histogram is chosen to depict two variables,  
  - the graph is called a histogram but has time-scale on the horizontal axis (e.g., the percentage of unemployed in 2018, 2019, ..., or the number of mortalities at certain time periods of a day).  
  - people use the number of bars as the number of measured values  
  Do not assign this code if:  
  - there is also an issue with the measurement level (see the code identifying the measured variable & measurement level) |
| Measurement level only (MO)     |                | The misinterpretation is related to only the measurement level of this variable, including:  
  - what is called a 'histogram' is actually a distribution bar graph (nominal or ordinal data on horizontal, (relative) frequency on the vertical)  
  - a statement is made about continuous distributions (e.g., normal distribution) in a graph with nominal or ordinal data  
  - a bar graph (nominal or ordinal data) is chosen with a 'bell-shape' instead of a histogram  
  Do not assign this code if:  
  - a statement about the normal distribution is related to variability in a histogram (e.g., a normal distribution has the lowest standard deviation; see the code variability) |
| Identifying the measured variable & measurement level (VM) |                | Both previous misinterpretations are at stake, for example if:  
  - frequency is regarded as a measured value in combination with a wrong measurement level (nominal or ordinal measured values)  
  - a statement is made that bars can be rearranged  
  - bar graph and histogram are used as synonyms  
  Do not assign this code if:  
  - the authors who made the mistake are not native English speakers nor statistics teachers (see the language code)  
  - the software does not distinguish between bar graphs and histograms (see the code ICT) |
| Distribution-related (DI)       | Variability (VY) | (A measure of) Variability is wrongly used in relation to the shape of the distribution in a histogram, including the following misinterpretations:  
  - variability is assessed as variation in the heights of the bars (thus variation in the frequency) instead of variation around the mean (e.g., range of measured value, IQR). This includes using words like ‘bumpy’.  
  - measures of variability, such as standard deviation, are wrongly used in a histogram  
  Do not assign this code if:  
  - a statement is made about continuous distributions (e.g., normal distribution) in a graph with nominal or ordinal data (see the code measurement level only) |
| Centre (CE)                     |                | (A measure of) Centre is wrongly used in relation to the shape of the distribution in a histogram, including the following misinterpretations:  
  - mean, mode or median are assessed of the frequency or heights of the bars  
  Do not assign this code if:  
  - people use the number of bars as the number of measured values (see the code identifying the measured variable only) |
| Shape (SH)                      |                | The misinterpretation is related to how the distribution of the data is depicted in a histogram, including graph conventions. The following are examples of misinterpretations:  
  - a histogram is wrongly matched to or compared with another type of graph (e.g., boxplot of the same data)  
  - area or density is wrongly used  
  Do not assign this code if:  
  - a statement is made about continuous distributions (e.g., normal distribution) in a graph with nominal or ordinal data (see the code measurement level only) |

(continued on next page)
### Table A1 (continued)

| Category     | Selective code                                                                                                                                                                                                 | Assign this code if                                                                                           |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| **Grouping (GR)** | The misinterpretation is a misunderstanding of the process of data reduction encompassed in a histogram leading to a possibly different shape or modality (e.g., bimodal, depending on the bin width), including the following misinterpretations:  
  - a statement is made that all values in a bar or only the midpoint are/is measured  
  - a wrong bin width is chosen (e.g., different bin widths without using density on the vertical axis), or not enough (e.g., two) or too many  
Do not assign this code if:  
  - the wrong bin width is generated by the software (see the code ICT) | |
| **Miscellaneous (MI)** | Context (CO)  
  The misinterpretation is due to the context or the research question, including the following:  
  - a wrong description of the distribution in a histogram is given in relation to context, or personal knowledge is used instead of the data in the histogram  
Do not assign this code if:  
  - time is mentioned (see the code identifying the measured variable only).  
| **Population (PO)** | The misinterpretation of the histogram is related to the distinction between sample and population, including the following:  
  - a statement is made indicating that sample and population in a histogram are the same (e.g., distribution, shape) | |
| **ICT (IT)** | The misinterpretation is embedded in the software by the software designers, including the following:  
  - the software does not distinguish between bar graphs and histograms or produces histogram with wrong boundaries | |
| **Unknown (U)** | A misinterpretation is mentioned by the authors of the publication but is not specific enough to be coded. | |
| **Translation (T)** | The misinterpretation may be caused by translation, including the following:  
  - if authors use the word bar graph for a histogram—or as synonym—and are not from a native English-speaking country. | |

### Appendix A.2. Provenance of participants of studies included in this review

The participants of studies included in the publications in this review come from the United States of America (12,959) followed by Europe (1572). Asia (1298) and South America (84) are relatively underrepresented (see Fig. A1); no African studies were found during the search. If the misinterpretation was in the publication itself or if the number of participants were not given, count was set to zero.
Fig. A1. Spread of the participants included in the studies in this review.
Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2019.100291.

References

Abrahamson, D. (2006). The shape of things to come: The computational pictograph as a bridge from a combinatory space to outcome distribution. *International Journal of Computers for Mathematical Learning, 11*(1), 137–146. https://doi.org/10.1007/s10758-006-9102-y.

Abrahamson, D. (2008). Bridging theory: Activities designed to support the grounding of outcome-based combinatory analysis in event-based intuitive judgment—A case study. *Proceedings of the international congress on mathematical education (ICME 11)* Monterrey, Mexico: ICME. Retrieved from https://iase-web.org/documents/papers/icme11/ICM11_TSG13_01P_abrahamson.pdf.

Abrahamson, D. (2009). Embodied design: Constructing means for constructing meaning. *Educational Studies in Mathematics, 70*(1), 27–47. https://doi.org/10.1007/s10649-008-9137-1.

Abrahamson, D., & Cendak, R. M. (2006). The odds of understanding the law of large numbers: A design for grounding intuitive probability in combinatory analysis. *Proceedings of the thirtieth conference of the international group for the psychology of mathematics education: Vol 2*, (pp. 1–8).

Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools in a technology-based design for statistics. *International Journal of Computers for Mathematical Learning, 12*(1), 23–55. https://doi.org/10.1080/10691898.2010.1188953.

Agro, S. (1977). Graphing. *USMES Intermediate “how to” set*. Columbus, OH, USA: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics, 52*(3), 215–241. https://doi.org/10.1023/A:1023213210781.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001). Toward a model of learning data representations. *Proceedings of the 23rd annual conference of the cognitive science society (pp. 45–50).*

Bakker, R. S., Corbett, A. T., & Koedinger, K. R. (2002). The resilience of overgeneralization of knowledge about data representations. New Orleans, LA *Annual meeting of the American educational research association* (pp. 1–10). Retrieved from https://files.eric.ed.gov/fulltext/ED465530.pdf.

Abrahamson, D. (2006). Design research in statistics education: On symbolizing and computer tools. Doctoral dissertation. Retrieved from https://dspace.library.uu.nl/handle/1874/893.

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about boxplots? In G. Burrill, & M. Camden (Eds.). *Curricular development in statistics education: International association for statistical education 2004 roundtable* (pp. 163–173). Voorburg, the Netherlands: International Statistical Institute.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1–2), 5–26. https://doi.org/10.1080/10986065.2011.538293.

Bakker, A., & Hoffmann, M. H. (2005). Diagrammatic reasoning as the basis for developing concepts: A semiotic analysis of students’ learning about statistical distribution. *Educational Studies in Mathematics, 60*(3), 333–358. https://doi.org/10.1007/s10649-005-5536-8.

Batanero, C., Tauber, L. M., & Sánchez, V. (2004). Students’ reasoning about the normal distribution. In D. Ben-Zvi, & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 257–276). (1st ed.). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/1-4020-2278-6.

Battaglia, O. R., DiPaola, B., & Fazio, C. (2017). A quantitative analysis of educational data through the comparison between hierarchical and not-hierarchical clustering. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(8), 4491–4512. https://doi.org/10.2973/eurasia.2017.00943a.

Behrens, J. T. (1997). Toward a theory and practice of using interactive graphics in statistics education. In J. B. Garfield, & G. Burrill (Eds.). *Research on the role of technology in teaching and learning statistics. Proceedings of the 1996 IASE Round Table Conference* (pp. 111–121). Voorburg, the Netherlands: ISI.

Ben-Zvi, D., & Garfield, J. (Eds.). (2004). The challenge of developing statistical literacy, reasoning, and thinking(1st ed.). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-3-030-03472-6.

Bennet, S. W., & Ismail, Z. (2013). Assessing misconceptions in reasoning about variability among high school students. *Procedia-Social and Behavioral Sciences, 93*, 1478–1483. https://doi.org/10.1016/j.sbspro.2013.06.067.

Biehler, R. (1997). Students’ difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. B. Garfield, & G. Burrill (Eds.). *Research on the role of technology in teaching and learning statistics* (pp. 169–190). Voorburg, the Netherlands: ISI.

Boels, L., Bakker, A., & Drijvers, P. (in preparation). Datapaper.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Cleveland, OH, USA: Sage.

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science & Technology, 40*(4), 473–493. https://doi.org/10.1080/0020729902759584.

Burrill, G., & Ben-Zvi (2019). Introduction. In G. Burrill, & D. Ben-Zvi (Eds.). *Topics and trends in current statistics education research.* International perspectives (pp. v–vi).

Capraro, M. M., Kulm, G., & Capraro, R. M. (2005). Middle grades: Misconceptions in statistical thinking. *School Science & Mathematics, 105*(4), 165–174. https://doi.org/10.1111/j.1949-8594.2005.tb18156.x.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 75–100. https://doi.org/10.1037/1076-8594.4.2.75.

Carrión, J., & Espinel, M. (2006). An investigation about translation and interpretation of statistical graphs and tables by students of primary education. *The seventh international conference on teaching statistics*Salvador, Bahia, Brazil: International Statistical Institute and International Association for Statistical Education. Retrieved from https://iase-web.org/documents/papers/ics7/C332.pdf.

Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi, & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–325). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/1-4020-2278-6_13.

Chan, S. W., & Ismail, Z. (2013). Assessing misconceptions in reasoning about variability among high school students. *Procedia-Social and Behavioral Sciences, 93*, 1478–1483. https://doi.org/10.1016/j.sbspro.2013.06.067.

Clayden, A., & Croft, M. (1990). Statistical consultation—who’s the expert? *Annals of Mathematics and Artificial Intelligence, 2*, 65–75. https://doi.org/10.1007/BF01530997.

Cohen, S. (1996). Identifying impediments to learning probability and statistics from an assessment of instructional software. *Journal of Educational and Behavioral Statistics, 21*(1), 35–54. https://doi.org/10.2307/1165254.

Cooper, L. L., & Shore, F. S. (2008). Students’ misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education, 16*(2). https://doi.org/10.1080/10691898.2010.11899487.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift für Soziologie, 19*(6), 418–427. https://doi.org/10.1515/zsoz.1990.0602.

Corredor, J. A. (2008). Learning statistical inference through computer-supported simulation and data analysisDoctoral dissertation. Retrieved from http://d-scholarship.pitt.edu/6683.

Curcio, F. R. (2011). The effect of prior knowledge, reading and mathematics achievement, and sex on comprehending mathematical relationships expressed in graphsDoctoral dissertation. Retrieved from https://files.eric.ed.gov/fulltext/ED210185s.pdf.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382–393. https://doi.org/10.1080/10691898.2010.11899559.
