

# Chapter 6. Seeing Patterns of Distribution

In this chapter, you will look for patterns in how your verbal data are distributed across the categories of your coding scheme. Using a frequency table, you will create and interpret distribution graphs of their patterns. Techniques for graphing are introduced.

## Introduction to Patterns of Distribution

In this chapter we take the first step in seeing patterns in the data you have just coded. In particular, we ask, “How did the way I assigned data to my coding categories vary with my built-in contrasts?” The answer to this question is called its distribution and visual representations of this pattern are called distribution graphs.

### A Note on Counting

Before we go into the details of how to create and understand patterns of distribution like those shown in Figure 6.1, we need to address the issue of counting. Behind every pattern visualized in a distribution graph is an act of counting. Judgments about the frequency with which something occurs—judgments of *more* or *less*, *similar* or *different*, are based on counting. In the case of distribution, we are basing our judgments on counting the number of segments that have been assigned to the coding categories in a coding scheme.

Counting has often been identified as the dividing line between qualitative and quantitative approaches to analysis. Simplistically, quantitative researchers use numbers; qualitative researchers don't. But, as Maxwell (2010) has pointed out, many prominent qualitative researchers, including literacy researchers Shirley Brice Heath and Brian Street (2008), do use numbers.

Making meaning of patterns is more than a matter of counting. Saldaña (2016) repeatedly warns us that, "Frequency of occurrence is not necessarily an indicator of significance." But as Miles, Huberman, and Saldaña (2014) acknowledge, numbers are often important to meaning making. In particular, these researchers identify three reasons for counting: First, they help us see overarching patterns in our data that we might otherwise not notice when focused on individual pieces of data. Second, they can help to verify (or question) a hypothesis that we bring to the data. Finally, they can serve as a check on our intuitions, a test for our biases.

Our focus in this and the following chapters is on using numbers to identify patterns, using what Sandelowski, Voils, and Knafl (2009) have called "quantitizing" (p. 210), the direct counting of qualitative data. Their discussion of quantitizing is useful because it describes the judgment underlying two important moves we make in coding verbal data: identifying clear and non-overlapping boundaries, the essential move we made in segmentation as described in Chapter 3, and sorting into categories, the essential move we made in coding as described in Chapter 4.

Maxwell (2010) notes the importance of the use of numbers to support the "internal generalizability" (p. 478) of a researcher's claims. We have already seen the power of numbers at work in Chapter 5 on reliability. There, counting the number of agreements and disagreements helped us to refine our intuitions and identify our biases as we worked to achieve reliability. In this chapter, we see another example of the power of numbers at work, this time helping us to provide evidence that the claims we want to make are actually characteristic of the data set as a whole, not just the result of selective and biased focus.

## ■ Basics of Distribution

Distribution refers to the way that your data is spread among the categories

in your coding scheme. It is the topology of the data. The patterns shown in Figure 6.1, for example, are distribution graphs of speaker data across the categories of Ed, John, and Cheryl for Meetings 1 and 2.

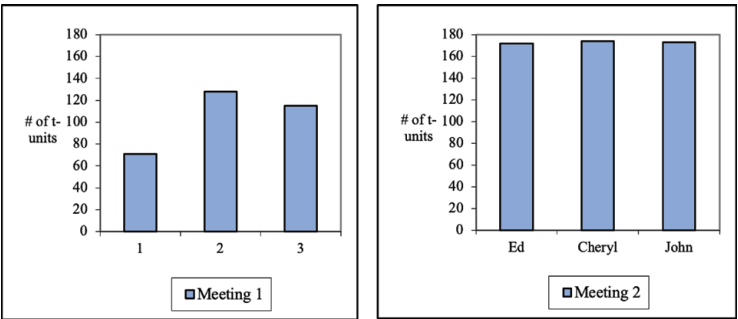


Figure 6.1: Distribution of speaker frequency in two meetings.

The most obvious use for distribution graphs comes when you want to know whether the distribution of the data varies according to the built-in contrasts that you created in your original analytic design. With that design, you articulated a plan for selecting data that you expected to show contrasting coding patterns in the phenomenon of interest. With these distribution graphs, you can begin to see whether your expectations were met: Does the distribution of your data over the categories in your coding scheme vary by the source of the data?

If, for example, we had chosen to look at data from Meeting 1 and Meeting 2 because we expected a built-in contrast, we could use the distribution shown in Figure 6.1 to see how these two meetings differed along the dimension of speaker participation. When the frequencies are looked at visually, we see two things immediately: that Meeting 2 was a lot more talkative and that Meeting 2 was a lot more equitable.

## ■ Absolute versus Relative Frequency

While the graphs in Figure 6.1 showed us distribution in terms of absolute frequency, those in Figure 6.2 show it in terms of relative frequency. That is, we have graphed the percentages of time that each category occurred. This data does not show us how talkative one meeting was compared to the other. For

that we would have to go back to the graphs in Figure 6.1. But these graphs tell us what proportion of the talk was contributed by each speaker and how those proportions varied between the two meetings.

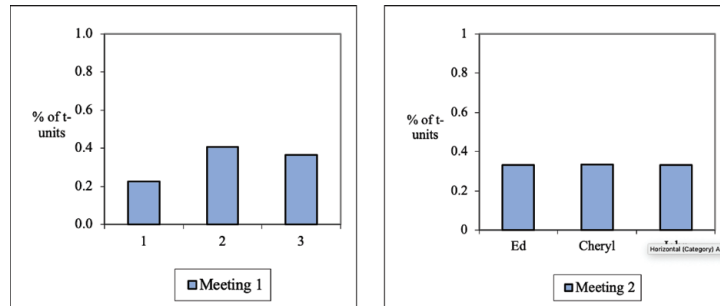


Figure 6.2: Relative frequencies of speaker distributions in two meetings.

If the total number of segments in one set of data is quite different from the total number in a comparison set, then comparing the absolute frequencies can be quite deceptive. In Figure 6.1, for example, we see that the frequencies for John are 115 in Meeting 1 and 173 in Meeting 2, two figures that seem quite different. But when we look at them as relative frequencies, as in Figure 6.2, we see that in Meeting 1, John spoke 37% of the time while in Meeting 2, he spoke 33% of the time, two figures that seem more similar. The difference here lies in the differences in the total number of segments in the two meetings. In Meeting 1 there were a total of 314 segments; in Meeting 2, there were 519 segments, two-thirds again as much. So while John spoke more often in absolute terms in Meeting 2, in relative terms, he spoke about as often.

Which of these two ways of looking at distribution is better, using the absolute frequency as we have done in Figure 6.1 or using relative frequency as we have done in Figure 6.2? The answer is both. Relative frequencies like those shown in Figure 6.2 allow us to compare across data sets. But if we looked at relative frequencies alone we might overlook important differences among the data sets in terms of overall size like the difference in meeting length shown in Figure 6.1.

It is also possible that a given category occurs with the same absolute frequency across data sets, rather than in terms of relative frequency. If we looked, for example, at the frequency of greetings in speeches, we would see about the

same absolute number of greetings across speeches, even though the speeches themselves might vary widely in length. That is, no matter how long the speech, speakers may tend to give the same number of greetings in their opening remarks. This pattern of absolute size would not show up if we just looked only at relative frequency. The point is that both measures of frequency tell us important information about the way our data is distributed over the categories making up our coding scheme. As a consequence, you should always examine both.

### Exercise 6.1 Test Your Understanding

The two graphs in Figure 6.3 show the absolute and relative frequency of men and women 13 or more years past dissertation who have been promoted to the rank of full professor in the five colleges at a university. What do the absolute frequencies tell you? What do the relative frequencies tell you? What are the differences between them?

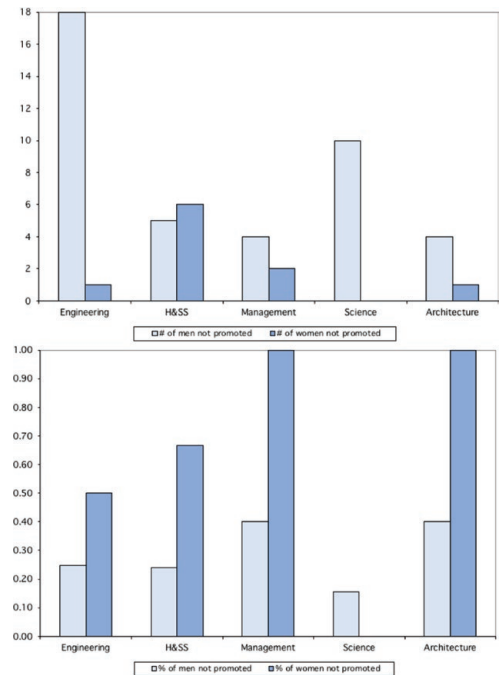


Figure 6.3: The absolute and relative frequency of promotion to the rank of full professor.

**For Discussion:** Which graph would you use to call attention to the problem of promotion among women faculty? Why?

## ■ Building a Frequency Table

Graphs like those shown in Figure 6.1 and 6.2 are based on frequency tables that summarize the data from a data worksheet. A frequency table shows the number of times data segments were coded in each of the categories of a coding scheme (see Excel Procedures 6.1 and 6.2 and MAXQDA Procedure 6.1). Along the side are the names of the coding categories, along the top are the names of the data samples. In Figure 6.4, for example, we have set up a frequency table for the speaker data for Meeting 1. On the top, we have labeled the source of the data (Meeting 1). Along the side, we have listed the categories (Ed, Cheryl, and John).

	Meeting 1
Ed	71
Chery	128
John	115

*Figure 6.4. A frequency table.*

### ■ Memo 6.1: Building Frequency Tables

Build a frequency table for each of your data sets, making sure to add marginals. What are your thoughts about how they compare?

---

## Excel Procedure 6.1: Naming Data in Excel

<https://goo.gl/LUK52D>

In Excel, we begin building a frequency table by naming the appropriate data range in the data worksheet.

1. In your data worksheet, select the column for the dimension you want to analyze for frequency.
2. Select the **Define Name** command under the **Insert** menu.
3. In the dialogue box as shown in Figure 6.5, give the data range a name that combines the data source and the dimension name.

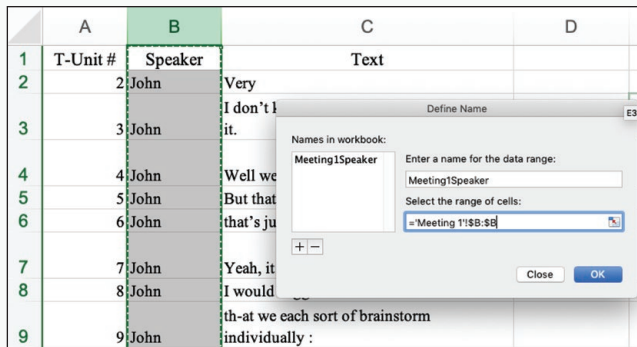


Figure 6.5: Naming a Data Range.

For example, we could name a data sheet Meeting1Speaker.

4. Click **OK** to complete the naming process.

## Excel Procedure 6.2: Making a Frequency Table in Excel

<https://goo.gl/LUK52D>

1. Create a frequency table with data source at the top and coding categories down the side.
2. Click in the first blank cell of your table.
3. Type in the following formula:  
=countifs(
4. Choose your first data range by clicking **Insert > Name > Paste** as shown in Figure 6.6.

Continued ...

 **Excel Procedure 6.2: Making a Frequency Table in Excel (continued)**

<https://goo.gl/LUK52D>

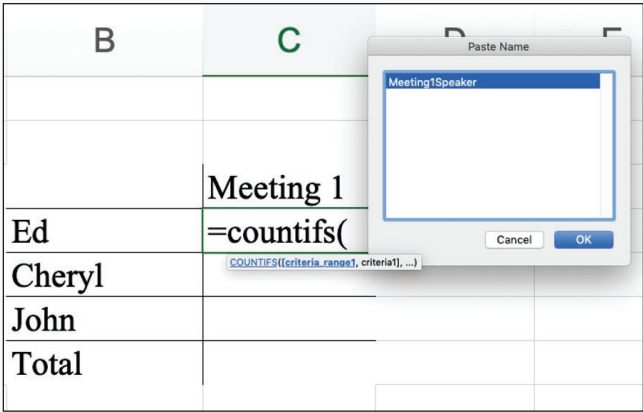


Figure 6.6: Filling a Frequency Table with a countifs formula in Excel.

- Click **OK**.
- Type a comma and then click on the cell corresponding to first coding category.

Note that the spelling and case must be an exact match.

- Type ) to close the formula and hit enter.

The number of times the first coding category was used in your first data set now appears in the cell.

To use this formula to fill the rest of the frequency table, edit and drag the formula using the following procedure:

- Click into the cell with the countifs formula.



Figure 6.7: Editing the column name in the countifs formula in Excel.

- Edit the countifs formula by placing a \$ before the column name category as shown in Figure 6.7.
- Select and drag this edited formula down the column.



## MAXQDA Procedure 6.1: Making a Frequency Table in MAXQDA

<https://goo.gl/LUK52D>

1. In the **Document System** window, activate the document for which you want to build a frequency table.
2. In the **Code System** window, activate the codes for your chosen dimension.
3. Select the **Overview of Codes** command from the **Codes** menu.
4. Click on the **Statistic** icon at the top of the **Overview of Codes** window.
5. In the **Statistic of Document Variables** window, select **Coded segments of activated documents** in the drop-down menu at the top.

MAXQDA will display a frequency table like that shown in Figure 6.8.

	Frequency	Percentage	Percentage (valid)
Cheryl	128	40.9	40.9
John	115	36.7	36.7
Ed	70	22.4	22.4
Total (Valid)	313	100.0	100.0
Missing	0	0.0	
Total	313	100.0	

Figure 6.8: Frequency table displayed in the Code Matrix Browser in MAXQDA.

## ■ Graphing Distribution

One of the best ways to see patterns in verbal data is through graphs—the spatial array of data points on a two- and sometimes three-dimensional coordinate system. Almost all of us are better able to see and interpret patterns through the spatial arrays shown in Figures 6.1 than the numeric frequency tables like the one in Figure 6.4. It is relatively easy to move from frequency table to distribution graph (see Excel Procedure 6.3 and MAXQDA Procedures 6.2 and 6.3).

---

### Exercise 6.2 Try It Out

In the exercise available at <https://wac.colostate.edu/books/practice/coding-streams/>, create a graph of the distribution of speaker for Meeting 2 in the same style as the graph for Meeting 1. Do this by creating a style template for the Meeting 1 graph and then applying it to the Meeting 2 graph.

### ■ Memo 6.2: Distribution Graphs

Create a distribution graph for each of your data sets. Make sure they have a common scale on the vertical axis.

What patterns do you now see with these visual representation?

---

## Excel Procedure 6.3: Creating a Distribution Graph for Excel Data

<https://goo.gl/LUK52D>

1. As shown in Figure 6.9, select the cells of the frequency table representing your code totals and code names. Do not select column totals.

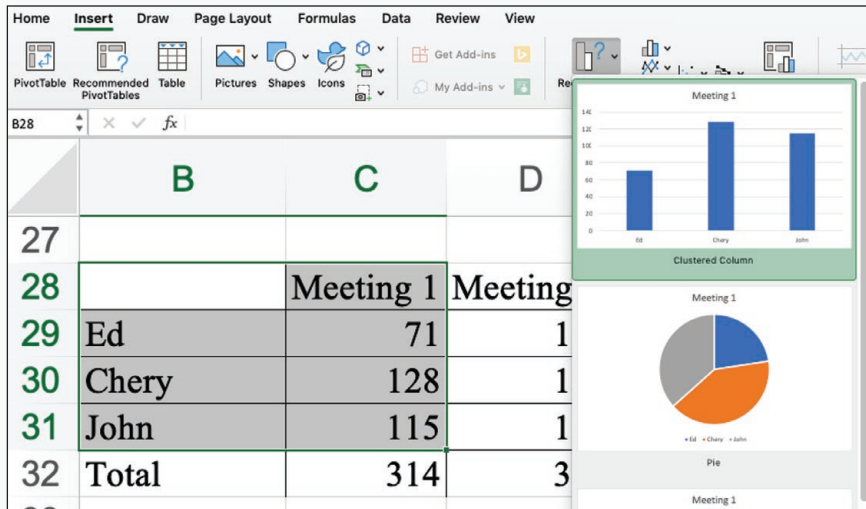


Figure 6.9: Inserting a clustered column chart in Excel.

2. On the **Insert** menu, select **Column** from the **Chart** options.
3. Add a label to your primary vertical axis by choosing **Vertical Axis** under **Axis Titles** using the **Add a Chart Element** command on the **Chart Design** tab.

Excel will now display a distribution graph.

To insure that distribution graphs are on the same scale:

4. Double click on the vertical axis.
5. In the **Format Axis** pane, under **Axis Options**, set the **Minimum and Maximum** bounds as desired.

To transfer these features to other distribution graphs:

6. Click on the chart with the desired features.
7. On the **Chart Design** tab, under **Change Chart Type**, select the **Save as Template** command.



## MAXQDA Procedure 6.2: Creating a Distribution Graph in MAXQDA

<https://goo.gl/LUK52D>

1. Select the **Overview of Codes** command from the **Codes** menu.
2. Click on the **Statistic** icon at the top of the **Overview of Codes** window.
3. In the **Statistic of Document Variables** window, select **Coded segments of activated documents** in the drop-down menu at the top.
4. Click on the **Chart View** icon at the top of the window.

MAXQDA will now display a distribution graph for the activated document.



## MAXQDA Procedure 6.3: Creating Distribution Graphs with a Common Scale for MAXQDA Data

<https://goo.gl/LUK52D>

MAXQDA does not allow you to adjust the scale of the axis in its charts. If you want to use a common scale for a set of distribution charts, you may want to export the frequency table to Excel and then produce your distribution graphs as follows:

1. Select the **Overview of Codes** command from the **Codes** menu.
2. Click on the **Statistic** icon at the top of the **Overview of Codes** window.
3. In the **Statistic of Document Variables** window, select **Coded Segments** of activated documents in the drop-down menu at the top.
4. Click on the **Export** icon to save the frequency table to Excel.
5. As shown in Figure 6.9, select the cells of your frequency table representing your code totals and code names. Do not select the column totals.
6. On the **Insert** menu, select **Column** from the **Chart** options.
7. Add a label to your primary vertical axis by choosing **Vertical Axis** under **Axis Titles** using the **Add a Chart Element** command on the **Chart Design** tab.
8. Excel will now display a labeled distribution graph.

Continued . . .

## ■ Interpreting Patterns of Distribution

Once you have constructed distribution graphs of the absolute and relative frequencies of your coding categories for each sample of your data, you are in a position to begin the process of interpreting the patterns they show. In other words, with the distribution graphs you have just created, you can begin to explore the nature of your built-in contrasts. To aid you in this endeavor, we review in the following sections the more common distribution patterns as well as ways you can look for patterns across streams of data and across categories.

### ■ Direct and Inverse proportions

When is the distribution of coding categories across one sample of your data similar to another sample? The simple cases are distributions where the frequencies of each coding category are identical. In the case of Meeting 1 and Meeting 2, for example, we would certainly say that the distributions are the similar if Cheryl, Ed, and John talked for exactly the same amount of time in each the two meetings.



#### MAXQDA Procedure 6.3: Creating Distribution Graphs with a Common Scale for MAXQDA Data (continued)

<https://goo.gl/LUK52D>

To insure that both distribution graphs are on the same scale:

9. Double click on the vertical axis.
10. In the **Format Axis** pane, under **Axis Options**, set the **Minimum and Maximum bounds** as desired.

To transfer these features to other distribution graphs:

11. Click on the chart with the desired features.
12. On the **Chart Design** tab, under **Change Chart Type**, select the **Save as Template** command.
13. Give the chart an appropriate name such as Frequency and click **OK**.
14. To use the frequency template, select it from the templates among the **Chart** options under the **Insert** menu.

Our understanding of similarity goes beyond identical frequencies, however. Suppose that Meeting 2 was twice as long as Meeting 1, but that the relative frequencies with which Cheryl, Ed, and John talked were the same. In this case, we usually say that the distributions are similar—that Cheryl, Ed, and John talked about the same amount across the two meetings—despite the overall difference in the length of the meetings.

The underlying pattern for this larger sense of similarity is called direct proportionality: With direct proportionality, as the absolute frequency of one category changes, the absolute frequency of other categories changes proportionately and in the same direction; if one goes up, the others go up. If one goes down, the others go down.

The two graphs in the top half of Figure 6.10 show distributions that are directly proportional to one another. Despite the fact that Meeting 1 is a lot longer than Meeting 2 (275 t-units versus 134 t-units), the relative frequencies of the coding categories are nearly identical between them: 64% for Ed, 24% for Cheryl, and 12% for John in Meeting 1, and 64% for Ed, 23% for Cheryl, and 13% for John in Meeting 2. Sometimes frequency distributions are inversely rather than directly proportional. In inverse proportion, as the absolute frequency of one category changes, the absolute frequencies of other categories change proportionately but in the opposite direction. If one goes up; the others go down. If one goes down, the others go up. This is the case for the pair of distribution graphs shown in the bottom half of Figure 6.10. In Meeting 1, Cheryl talked for 34% of the time and John for 66%; in Meeting 2, the relative frequencies are reversed with Cheryl talking for 66% of the time and John for 34%. This is a clear case of inverse proportion.

Direct and inverse proportions are relatively common in verbal data. For example, suppose you have two conversations with the same person. In the first conversation, she is rather quiet; in the second conversation, she is more loquacious. If we were to look at the distribution of talk between speakers in these two conversations, we would probably find they were directly proportional. That is, as your interlocutor spoke more (or less), you probably also spoke more (or less) in return. As a result, the speaker distributions in the two conversations, which vary in overall length, are nevertheless similar.

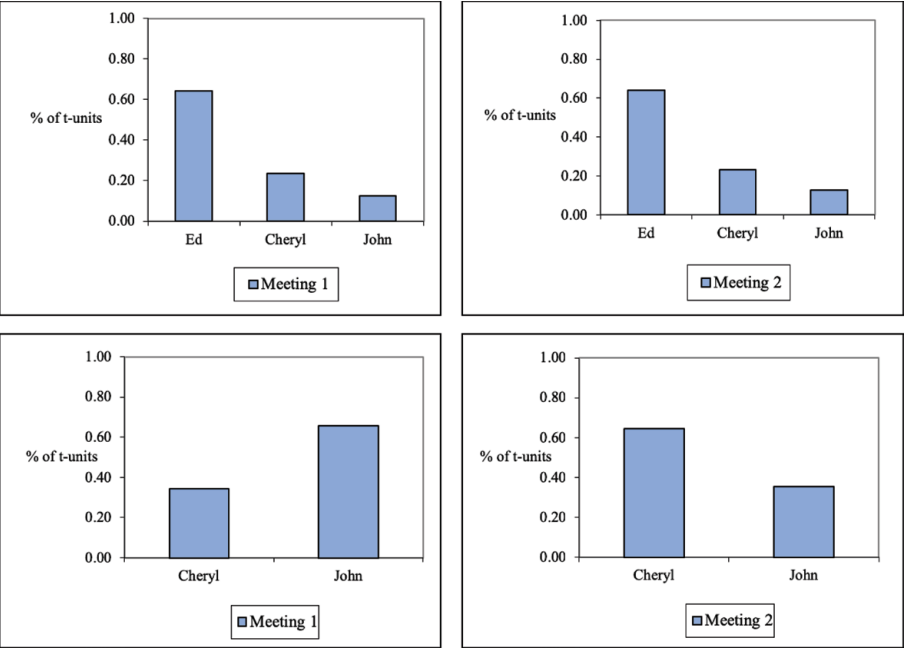


Figure 6.10: Patterns of direct (left) and inverse (right) proportion.

Suppose, however, that your conversations with your interlocutor occur within a formal one-hour meeting. Now, when she speaks more, she leaves less time for you to speak. As a result, the speaker distributions become inversely proportional. On days she speaks more, you speak less; one days she speaks less, you speak more. As this example suggests, inverse proportion is fairly common in cases of fixed resources—such as the total time available for a meeting.

## ■ Constants, Absolute and Proportional

Direct and inverse proportionality are patterns that reflect systematic differences between distributions. By contrast, some relationships between distributions can be characterized more by what they have in common than by how they differ. In the top of Figure 6.11, for example, the two distributions show a common value for Cheryl's contribution: Cheryl spoke 94 t-units in both

Meeting 1 and Meeting 2 even though Ed and John's contributions differed across the two meetings.

When one category has the same absolute frequency in two or more samples, we see in them a pattern involving an absolute constant. In our example, the absolute frequency of Cheryl's contributions—94 t-units—is an absolute constant across the two meetings.

A related pattern, shown in the bottom of Figure 6.11 involves a proportional constant. Here, the relative rather than absolute frequency remains constant across samples. In both meetings graphed in the bottom graphs in Figure 6.11, for example, Cheryl speaks about 33% of the time in both meetings.

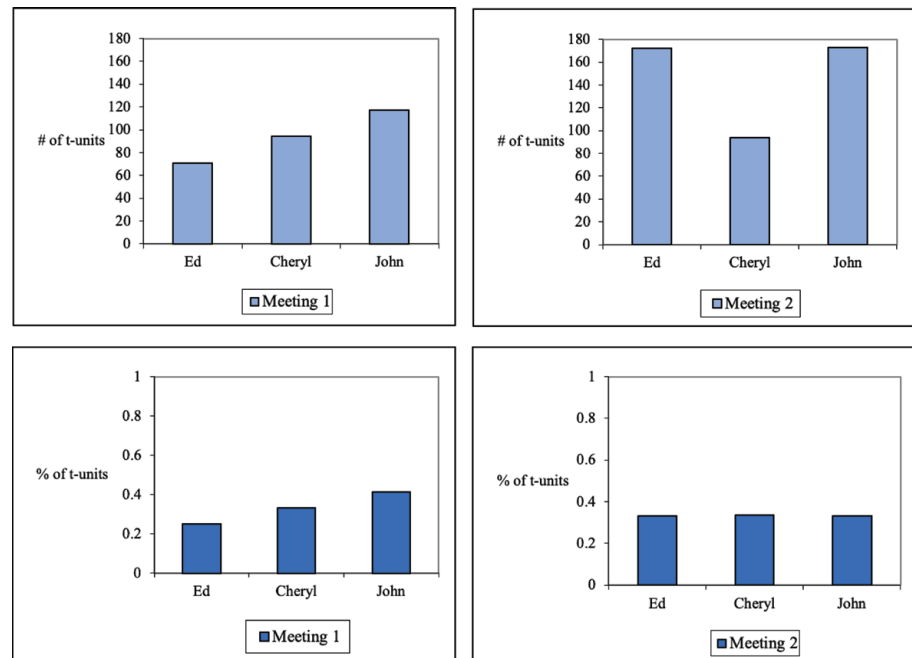


Figure 6.11: Patterns of absolute (top) and proportional (bottom) constants.

Absolute and proportional constants are not unusual in verbal data. Suppose, for example, a weekly meeting is always opened with a five-minute recap of last week's meeting. All verbal phenomenon associated with that recap will

tend to be fixed in quantity regardless of how long the rest of the meeting takes. In this case, meetings from two different weeks will tend to exhibit a pattern of absolute constant.

Proportional constants are even more common. Suppose, for example, that the number of questions asked by the prosecution and defense in a trial was a fixed percentage of the trial proceedings. Then the contribution of attorneys' questions will be a proportional constant, increasing as the length of the trial increases, but remaining more or less at a constant proportion of the proceedings. Witness contributions, on the other hand, may be quite variable, with some witnesses responding extensively while others have relatively little to say. In this case, in contrast to attorneys' contributions, the length of witness contributions might have no relationship to the length of the trial.

## ■ Equity and Dominance

The distribution patterns shown in Figure 6.12 are last but not least in the analysis of verbal data. Here, in Meeting 1, we see a situation in which one category, speaker John, dominates. In Meeting 2, by contrast, speaker contribution is equitably distributed across speaker categories.

Patterns of equity and dominance are significant in the analysis of verbal data. In most classroom conversations, for example, the teacher's contributions tend to dominate regardless of other participation. Equitable distributions like those shown to the right in Figure 6.12 are relatively rare.

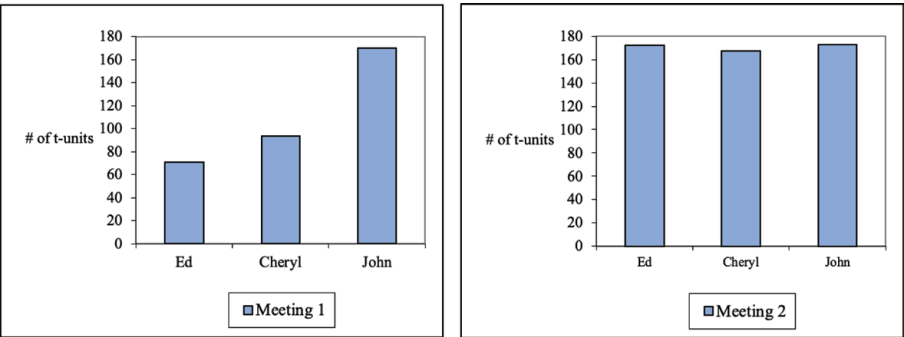


Figure 6.12: Patterns of equity and dominance.

## ■ Lack of Pattern

Finally, the last, and certainly not least important pattern to be on the lookout for in the analysis of distributions is the pattern of “no pattern.” All verbal data is characterized by natural variations due to innumerable factors. Sometimes these variations add up to variation across distributions that cannot be described as falling into any particular pattern. In Figure 6.13, for example, no discernible pattern exists linking Meeting 1 and Meeting 2. Cheryl’s contributions go way down; John’s stay about the same; Ed’s go up. While we can develop narrative descriptions of these relationships, there does not appear to be an overall pattern as was in the cases described in the previous sections.

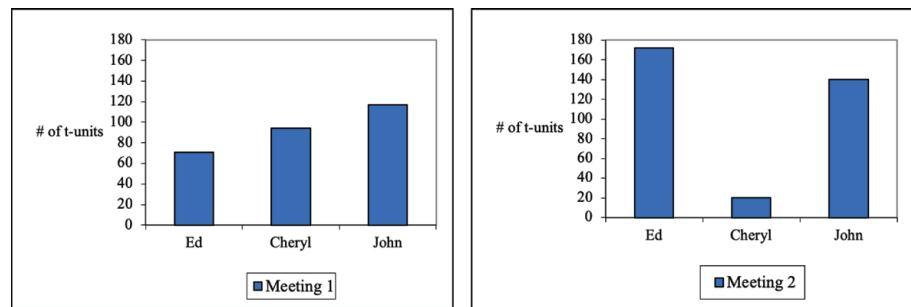


Figure 6.13: Distributions without pattern.

If, given the literature or general beliefs, you expected differences between data samples, the lack of pattern can be quite interesting. If, on the other hand, you are looking for ways to characterize the relationship between two data samples, the lack of pattern for the dimension analyzed should be a message to send you looking at other dimensions or to other phenomena altogether.

The patterns just reviewed are only some of those that you may encounter as you compare the distributions across the built-in contrasts of your design. They are tools to help you to notice and characterize the contours of your data samples. If they indicate that something is “going on” in your data, you may want to refine your analysis to produce a better characterization of the phenomenon. In the remainder of this chapter, we deal with two such refinements—collapsing streams and combining categories—as well as some of the interactions that can make such refinements overly simple and thus problematic.

### Exercise 6.3 Try It Out

The graphs in Figure 6.14 show the absolute frequency and relative frequency with which texts that were private (intended for oneself as reader) and texts that were public (intended for other readers) were used during a working session that involved five different applications: desktop software for a personal digital assist (PDA), email, web browser, word processing, and spreadsheet. How would you describe the differences between the applications in terms of these distributions?

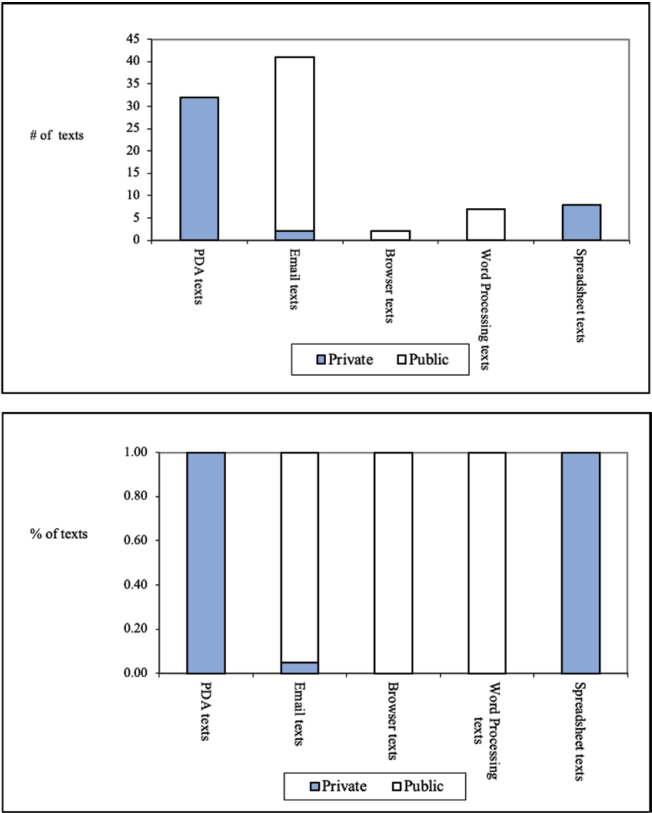


Figure 6.14: Distribution of public and private texts.

**For Discussion:** Compare your descriptions with those of your classmates.

## ■ Memo 6.3: Patterns

Review your distribution graphs across your built-in contrast. Do the distributions on one side of your contrast have more in common than they do with the distributions on the other side?

How would you best compare them?

---

## ■ Refining Patterns across the Built-In Contrast

### ■ Collapsing Streams

Thus far, we have described the distributions you may want to analyze as if they consist of just one stream of language on each side of your built-in contrast. Yet good analytic designs include multiple streams across a built-in contrast as we described in Chapter 2. Suppose, for example, we had designed our analysis to contrast two kinds of meetings—design meetings and managerial meetings—and we had transcripts from two meetings of each kind—four meetings in total.

Our analysis would begin, of course, with the distributions of each of the four streams as shown in Figure 6.15. In this case, we find that there is a proportional similarity between Meetings 1 and 2, the two design meetings, in which Ed and John talk with about the same relative frequency and Cheryl talks less frequently. The two managerial meetings, Meetings 3 and 4, are likewise similar to each other, though different from the design meetings. Here, Ed talks most frequently followed by Cheryl and with John left far behind.

When data streams have more in common with each other than they do with the streams across a built-in contrast, we may want to collapse data within the contrast in order to explore the pattern (see Excel Procedure 6.4 and MAXQDA Procedure 6.4). In Figure 6.16, for example, we have collapsed the data from the two design meetings and the two managerial meetings to give a portrait of how, overall, the design meetings and managerial meetings differ.

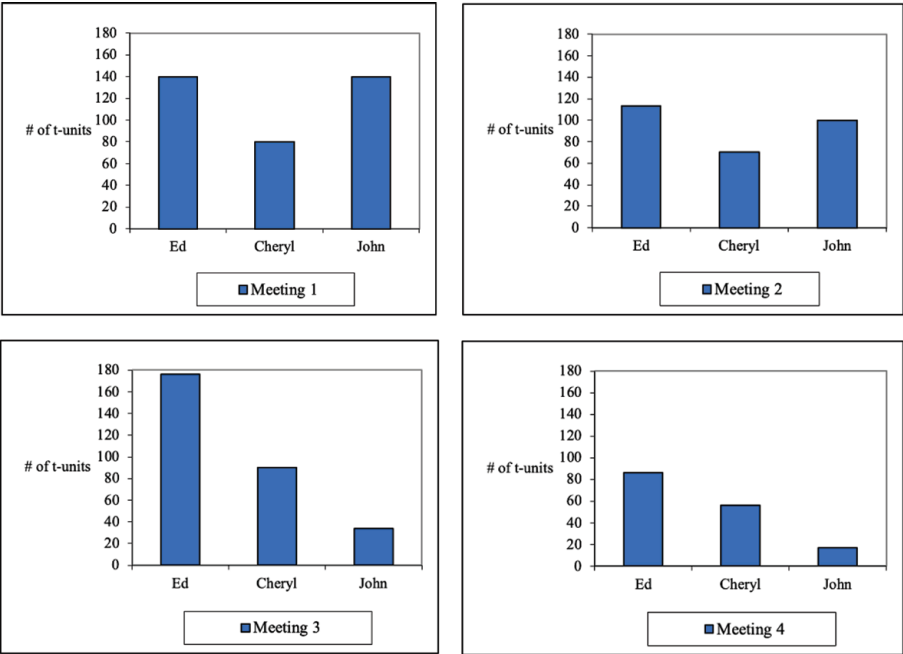


Figure 6.15: Distributions of multiple streams.

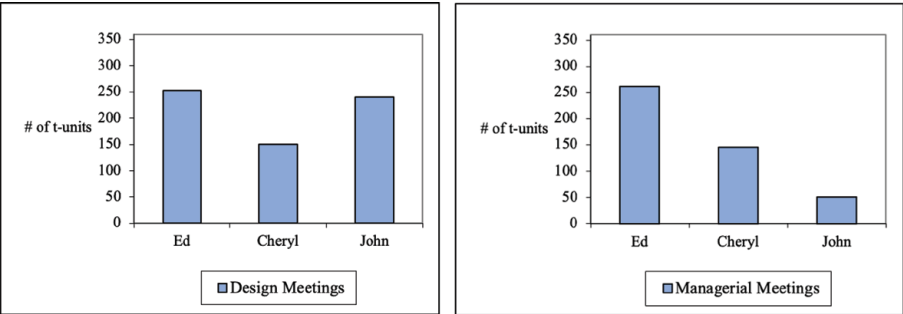


Figure 6.16: Collapsing streams.



## Excel Procedure 6.4: Collapsing across Streams in Excel

<https://goo.gl/LUK52D>

To collapse data streams, add together the frequencies from the data sets on one side of the contrast as follows:

1. Create an empty summary frequency table like the one shown in the bottom of Figure 6.17. Label it to reflect one side of your built-in contrast.
2. Click in the first empty cell of this table and begin typing the formula with  
=  
3. Click in the corresponding cell of the frequency table for the first data stream on one side of your contrast.
4. Type  
+  
5. Click in the corresponding cell of the frequency table for the next data stream on the same side of your contrast.
6. Type  
)  
7. Hit enter to add the values together.

As shown in Figure 6.17, this procedure will enter the following formula in the first cell of your summary table:

=B6+B12

To add additional data streams to your summary, repeat steps 4 and 5, adding to your formula a + followed by the corresponding cell for each additional frequency table.

Fill the rest of the summary table by dragging the formulas down the column.

	Meeting 1
Ed	140
Cheryl	80
John	140
	Meeting 2
Ed	113
Cheryl	70
John	100
	Design Meetings
Ed	=C6+C11
Cheryl	
John	

Figure 6.17: Adding frequencies to collapse streams.



## MAXQDA Procedure 6.4: Collapsing across Streams in MAXQDA

<https://goo.gl/LUK52D>

To collapse across data streams in MAXQDA, you create a document variable and use it to classify each of your documents according to your built-in contrast; then you collapse across documents.

1. From the Variables menu, select the **List of Document Variables** command.
2. In the pop-up **Document Variables** window, click on the **New Variable** icon.
3. Type in a name for the contrasting variable as shown in Figure 6.18 and click **OK**.

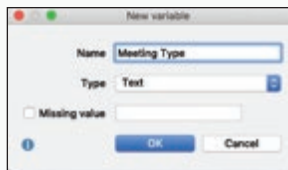


Figure 6.18: Adding a document variable in MAXQDA.

4. From the **Variables** menu, select the **Data Editor** command.
6. In the **Document** window, click in the first row under the new variable and type in the name for one side of your contrast as shown in Figure 6.19 and hit enter.

Document g...	Document n...	Creation date	Number of c...	Number of ...	Author	Meeting Type
	Meeting 4	.../18 4:25 AM	159	0	Cheryl Geisler	Management
	Meeting 3	.../18 4:25 AM	301	0	Cheryl Geisler	Management
	Meeting 1	.../18 12:38 PM	313	0	Cheryl Geisler	Design
	Meeting 2	.../18 12:37 PM	314	0	Cheryl Geisler	Design

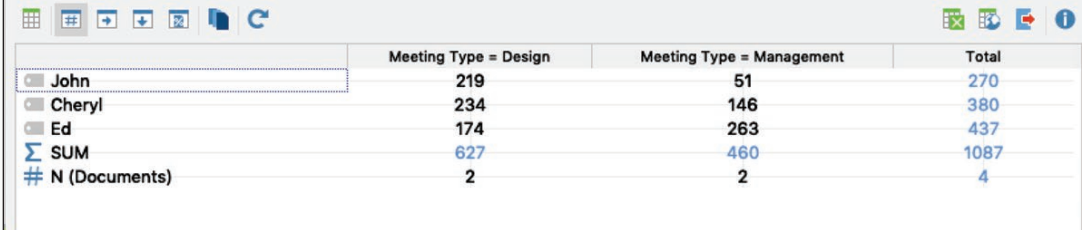
Figure 6.19: Classifying documents by variable in MAXQDA.

Continued...



## MAXQDA Procedure 6.4: Collapsing across Streams in MAXQDA (continued)

<https://goo.gl/LUK52D>



	Meeting Type = Design	Meeting Type = Management	Total
John	219	51	270
Cheryl	234	146	380
Ed	174	263	437
<b>SUM</b>	<b>627</b>	<b>460</b>	<b>1087</b>
<b># N (Documents)</b>	<b>2</b>	<b>2</b>	<b>4</b>

Figure 6.20: Frequency table for collapsed streams in MAXQDA.

7. In the next row, classify the next document, either by typing in the name or selecting from the drop-down menu and hit enter.
8. Repeat step 7 to complete the classification for all your documents.
9. From the **Mixed Methods** menu, select the **Crosstab** command.
10. In the **Crosstab** window, select your variable of contrast and click on the right arrow to move it into the second column.
11. In the third column, make sure that the = option is selected and use the drop-down menu to choose a value for your built-in contrast.
12. Repeat for the remaining variables of contrast.
13. Deselect the options for **Only activated documents** and **Only activated codes**, and then click **OK**.

A frequency table across your contrast will appear in the crosstab window as shown in Figure 6.20.

## ■ Calculating Averages

The data graphed in Figure 6.16 represent the total frequency of speaker categories in two design meetings and two managerial meetings. Another, perhaps more meaningful statistic for this data is the average frequency of speaker contribution per meeting. Averages can tell you the average frequency of a given code on one side of your built-in contrast. Ed, for example, contributed an average of 126.5 t-units per meeting in design meetings and 131.0 t-units per meeting in managerial meeting. Such averages give a better sense of the average magnitude of a phenomenon (like Ed's speaking) and are thus sometimes more useful in communicating with your readers. To learn more, see Excel Procedure 6.5 and MAXQDA Procedure 6.5.



### Excel Procedure 6.5: Calculating Averages in Excel

<https://goo.gl/LUK52D>

1. Click in the cell where you want the average to appear.
2. Type  
=average(
3. Select the range of values you want to average together.
4. Type  
)
5. Hit enter to complete the formula.

Excel will place a formula like the following in the cell:

```
=AVERAGE(C6:C7)
```



## MAXQDA Procedure 6.5: Calculating Averages in MAXQDA

<https://goo.gl/LUK52D>

1. Right click on a code in the **Code System** window as shown in Figure 6.21 and select **Transform into a Document Variable**.
2. Close the **Document Variables** window that pops up.
3. Repeat steps 1-2 for each of your codes.
4. From the **Mixed Methods** menu, select the **Typology Table** command.
5. If necessary, select the names of your codes and click **Continue**.
6. In the **Typology Table** pop-up window, select your variable of contrast and click on the right arrow to move it into the second column.
7. In the third column, make sure that the = option is selected and use the drop-down menu to choose a value for your built-in contrast.
8. Repeat for the other value(s) of your built-in contrast.
9. Deselect the options for **Only Activated Documents** and then click **OK**.

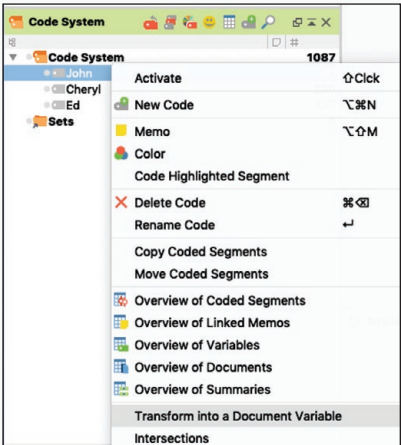


Figure 6.21: Transforming a code into a document variable in MAXQDA.

A typology table across your contrast will appear in the **Typology Table** window as shown in Figure 6.22. It shows the average (or mean) number of segments assigned to each code across your built-in contrast.

	Meeting Type = Design (N=2)	...anagement (N=2)
John, Mean (SD)	109.5 (5.5)	25.5 (8.5)
Cheryl, Mean (SD)	117.0 (11.0)	73.0 (17.0)
Ed, Mean (SD)	87.0 (17.0)	131.5 (45.5)
N (Documents)	2 (50.0%)	2 (50.0%)

Figure 6.22: A typology table showing the averages (means) for each code across a built-in contrast in MAXQDA.

## ■ Checking for Interactions

Whenever you collapse data streams, you make the implicit claim that the streams so collapsed follow the same basic pattern. Such need not be the case. Look at the data in Figure 6.23. As the frequency tables at the bottom of the figure suggest, the collapsed frequency for design and managerial meetings is the same here as it was in the data presented in Figure 6.16. In this case, however, this collapse presents a very misleading portrait of the individual streams.

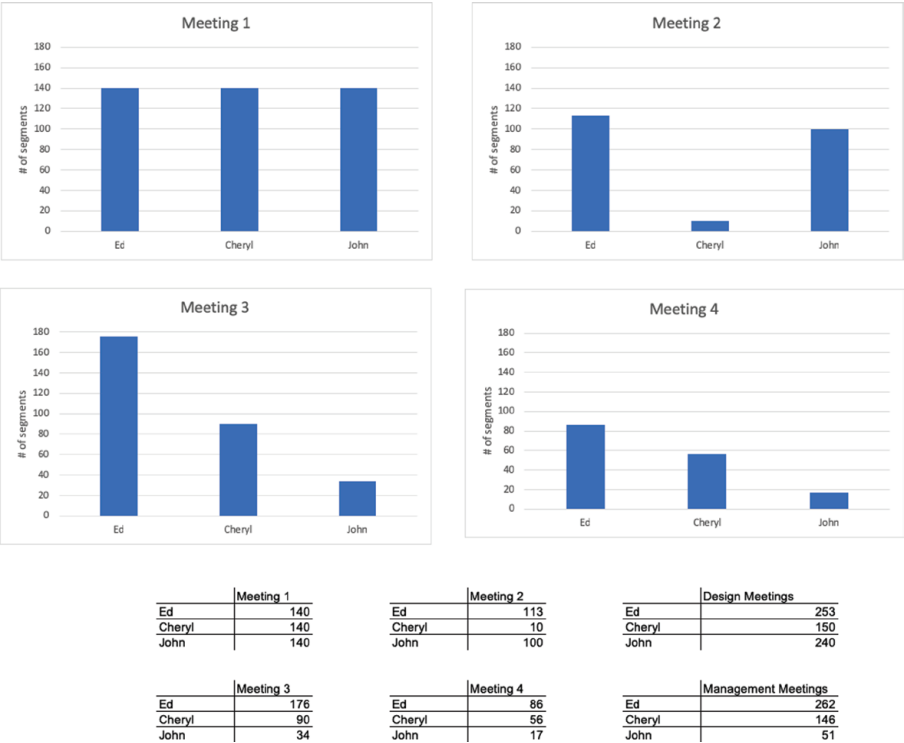


Figure 6.23: Interaction within streams.

To see how this is, compare the kinds of descriptions we might generate with the collapsed data with descriptions based on the individual streams. If we were to say, for instance, that Cheryl contributed, on average 75 t-units per

meeting in the design meetings, this characterization would hardly be a good description of her contribution in Meeting 1 (where she spoke for 140 t-units) or her contribution in Meeting 2 (where she spoke for 10 t-units).

Nor are comparisons between her total contribution and the total contributions of Ed and John valid. In the collapsed data as shown in Figure 6.24, Cheryl looks as if she speaks less than either Ed or John. But the data for the individual streams in Figure 6.25 makes clear that in Meeting 1 she spoke with about the same frequency as Ed and John, but in Meeting 2, she spoke far less.

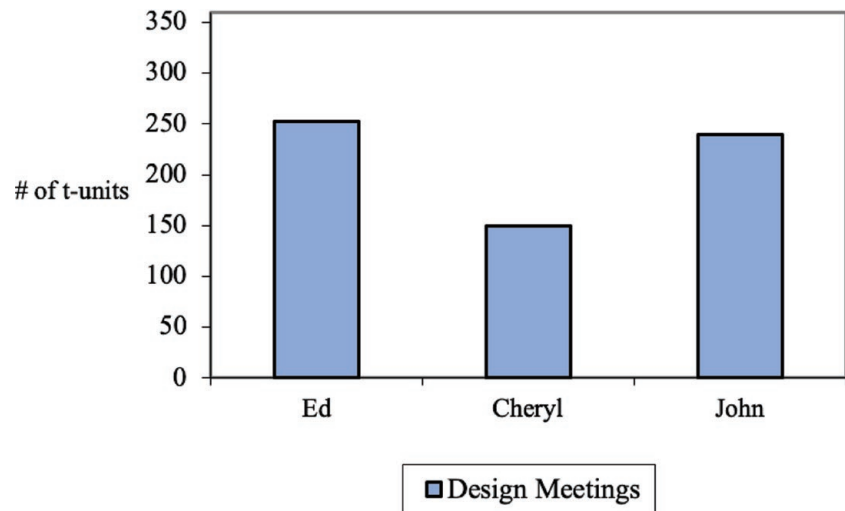


Figure 6.24: Misleading collapsed streams.

The situation we have been describing—in which the collapsing of streams provides a misleading portrait of the individual streams—is called an interaction. In an interaction, collapsing data across streams obscures major variations within the streams that makes the pooled statistics—either the collapsed frequencies or the averages—poor descriptions of the streams themselves. These are called interactions because the choice of stream interacts with pattern of distribution.

So, for example, if we want to answer the question, “how did the distribution of speaker contribution differ between the design meetings and the

managerial meetings?” for the data presented in Figure 6.23, we would have to answer, “It depends . . .” The first design meeting, Meeting 1, looks a lot more equitable than the managerial meetings, Meetings 3 and 4. But the second design meeting, Meeting 2, does not look at all equitable. In this data set, as in all interactions, there simply is no way to generate a general characterization that fits all the individual streams. Seeing these interactions may be telling you that the coding scheme you’ve chosen does not allow you to get at what makes the contrasts contrastive.

The overall lesson to be learned here is that while collapsing streams can be a very useful technique for exploring patterns, it can also be misleading if it obscures real variation within streams. Thus, it is important to always look at the distribution of each stream individually before taking the step of collapsing them together across your built-in contrasts.

### ■ Memo 6.4: Collapses across Contrast

Collapse your streams across contrast and create distributions graphs to show the sides of your built-in contrast with this collapsed data. Does the collapsed data match your intuitions that you developed earlier using single stream distribution graphs? Is the collapsed data a fair representation of the individual streams or do you have significant interactions?

Write one to two sentences describing each of your individual stream graphs as well as the collapsed graphs.

---

## ■ Refining Patterns Across Codes

A second technique for refining distribution patterns involves combining codes within your coding scheme (see Excel procedure 6.6 and MAXQDA Procedure 6.6). In Figure 6.25, for example, we show meetings with four speakers combined into two categories, managers and subordinates. This 100% stacked column graph neatly suggests that in both meetings managers spoke about twice as often as subordinates, a very suggestive finding.

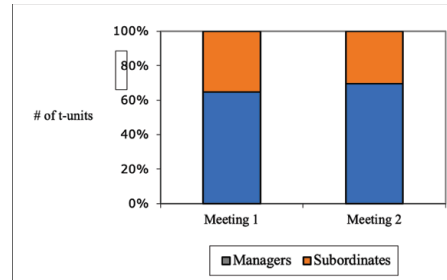


Figure 6.25: Combining categories.



### Excel Procedure 6.6: Combining Codes in Excel

<https://goo.gl/LUK52D>

To add together the frequencies for two codes you want to combine:

1. Create an empty summary frequency table with cells for the frequencies of the new combined code.
2. Click in the first empty cell for the new combined code and begin typing the formula with =
3. Click in the cell with the frequency of the first code you want to combine.
4. Type +
5. Click in the cell with the frequency of the second code you want to combine.
6. Type ) and hit enter to add the values together.
7. Fill in the remaining frequencies in the same way for the combined codes.



### MAXQDA Procedure 6.6: Combining Codes in MAXQDA

<https://goo.gl/LUK52D>

1. Create a new higher-level code representing the combination.
2. Move the original codes under the new code to become its subcodes.
3. To create a frequency table for the higher-level codes:
4. From the **Codes** menu, select the **Overview of Codes** command.
5. In the **Overview of Codes** pop-up window, click on the icon for **Aggregate on 1st level**.

MAXQDA will now display the frequencies for just the top-level codes.

## ■ Limitations on Combining Codes

Two limitations need to be placed on the use of combining codes as a way of refining distribution patterns. To begin with, codes can only be combined in ways that made sense with respect to the phenomena. Combine apples with apples, not with oranges (unless you are interested in fruit!).

Second, you must always check the distributions of the individual codes in advance of combining them to make sure that no interactions exist. Look at the distributions in Figure 6.26, for example. These represent the individual streams combined to produce the patterns shown in Figure 6.25. Ed and Roger are the managers; Cheryl and John are the subordinates. Combining the speaker codes as we have done makes the implicit claim that Ed and Roger have more in common with each other than they do with Cheryl and John. But the individual streams in Figure 6.26 give lie to this claim. Here we see that Roger, although a manager, looks more like Cheryl and John in his participation than he does like Ed. Combining his data with Ed's data to produce an overall picture of managers' contribution is, therefore misleading. There is an interaction: how managers act depends upon which manager: Ed or Roger. In such cases as this one, you should avoid combining categories.

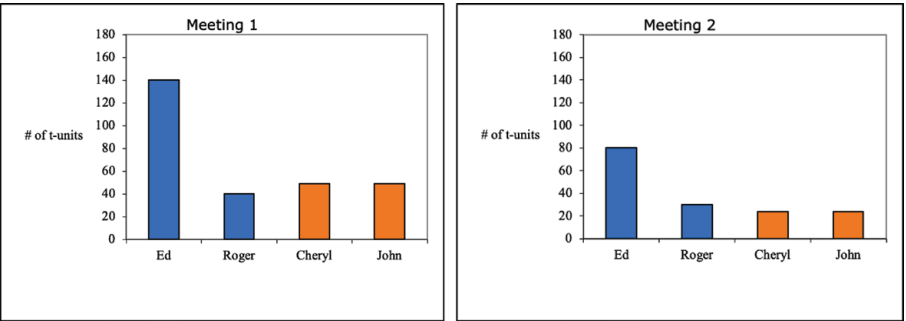


Figure 6.26: Interactions within categories.

## ■ Memo 6.5: Combined Codes

Are there codes that you might consider combining to better show the overall patterns in your data? Try to combine them and look again at the distribution of the combined codes.

Does the combined data fairly represent the individual codes?

---

## ■ Selected Studies Using Frequency Distributions

Campbell, K. S., & Naidoo, J. S. (2017). Rhetorical move structure in high-tech marketing white papers. *Journal of Business and Technical Communication*, 31, 94-118.

## ■ For Further Reading

Heath, S. B., & Street, B. V. (2008). *Ethnography: Approaches to language and literacy research*. New York: Teachers College Press.

Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry* 16(6), 475-482.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage.

Saldaña, J. (2016). *The coding manual for qualitative researchers*. London: Sage.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Method Research* 3(3), 208-222.