Communication is fundamental to good data analysis. What we aim to address in this chapter is the role of routine communication in the process of doing your data analysis and in disseminating your final results in a more formal setting, often to an external, larger audience. There are lots of good books that address the "how-to" of giving formal presentations, either in the form of a talk or a written piece, such as a white paper or scientific paper. In this chapter, though, we will focus on:

1. How to use routine communication as one of the tools needed to perform a good data analysis; and
2. How to convey the key points of your data analysis when communicating informally and formally.

Communication is both one of the tools of data analysis, and also the final product of data analysis: there is no point in doing a data analysis if you're not going to communicate your process and results to an audience. A good data analyst communicates informally multiple times during the data analysis process and also gives careful thought to communicating the final results so that the analysis is as useful and informative as possible to the wider audience it was intended for.

The main purpose of routine communication is to gather data, which is part of the epicyclic process for each core activity. You gather data by communicating your results and the responses you receive from your audience should inform the next steps in your data analysis. The types of responses you receive include not only answers to specific questions, but also commentary and questions your audience has in response to your report (either written or oral). The form that your routine communication takes depends on what the goal of the communication is. If your goal, for example, is to get clarity on how a variable is coded because when you explore the dataset it appears to be an ordinal variable, but you had understood that it was a continous variable, your communication is brief and to the point.

If, on the other hand, some results from your exploratory data analysis are not what you expected, your communication may take the form of a small, informal meeting that includes displaying tables and/or figures pertinent to your issue. A third type of informal communication is one in which you may not have specific questions to ask of your audience, but instead are seeking feedback on the data analysis process and/or results to help you refine the process and/or to inform your next steps.

In sum, there are three main types of informal communication and they are classified based on the objectives you have for the communication: (1) to answer a very focused question, which is often a technical question or a question aimed at gathering a fact, (2) to help you work through some results that are puzzling or not quite what you expected, and (3) to get general impressions and feedback as a means of identifying issues that had not occurred to you so that you can refine your data analysis.

Focusing on a few core concepts will help you achieve your objectives when planning routine communication. These concepts are:

1. **Audience**: Know your audience and when you have control over who the audience is, select the right audience for the kind of feedback you are looking for.
2. **Content**: Be focused and concise, but provide sufficient information for the audience to understand the information you are presenting and question(s) you are asking.
3. **Style**: Avoid jargon. Unless you are communicating about a focused highly technical issue to a highly technical audience, it is best to use language and figures and tables that can be understood by a more general audience.
4. **Attitude**: Have an open, collaborative attitude so that you are ready to fully engage in a dialogue and so that your audience gets the message that your goal is not to "defend" your question or work, but rather to get their input so that you can do your best work.

The Audience

For many types of routine communication, you will have the ability to select your audience, but in some cases, such as when you are delivering an interim report to your boss or your team, the audience may be pre-determined. Your audience may be composed of other data analysts, the individual(s) who initiated the question, your boss and/or other managers or executive team members, non-data analysts who are content experts, and/or someone representing the general public.

For the first type of routine communication, in which you are primarily seeking factual knowledge or clarification about the dataset or related information, selecting a person (or people) who have the factual knowledge to answer the question and are responsive to queries is most appropriate. For a question about how the data for a variable in the dataset were collected, you might approach a person who collected the data or a person who has worked with the dataset before or was responsible for compiling the data. If the question is about the command to use in a statistical programming language in order to run a certain type of statistical test, this information is often easily found by an internet search. But if this fails, querying a person who uses the particular programming language would be appropriate.

For the second type of routine communication, in which you have some results and you are either unsure whether they are what you'd expect, or they are not what you expected, you'll likely be most helped if you engage more than one person and they represent a range of perspectives. The most productive and helpful meetings typically include people with data analysis and content area expertise. As a rule of thumb, the more types of stakeholders you communicate with while you are doing your data anlysis project, the better your final product will be. For example, if you only communicate with other data analysts, you may overlook some important aspects of your data analysis that would have been discovered had you communicated with your boss, content experts, or other people.

For the third type of routine communication, which typically occurs when you have come to a natural place for pausing your data analysis. Although when and where in your data analysis these pauses occur are dictated by the specific analysis you are doing, one very common place to pause and take stock is after completing at least some exploratory data analysis. It's important to pause and ask for feedback at this point as this exercise will often identify additional exploratory analyses that are important for informing next steps, such as model building, and therefore prevent you from sinking time and effort into pursuing models that are not relevant, not appropriate, or both. This sort of communication is most effective when it takes the form of a face-to-face meeting, but video conferencing and phone conversations can also be effective. When selecting your audience, think about who among the people available to you give the most helpful feedback and which perspectives will be important for informing the next steps of your analysis. At a minimum, you should have both data analysis and content expertise represented, but in this type of meeting it may also be helpful to hear from people who share, or at least understand, the perspective of the larger target audience for the formal communication of the results of your data analysis.

Content

The most important guiding principle is to tailor the information you deliver to the objective of the communication. For a targeted question aimed at getting clarification about the coding of a variable, the recipient of your communication does not need to know the overall objective of your analysis, what you have done up to this point,

or see any figures or tables. A specific, pointed question along the lines of "I'm analyzing the crime dataset that you sent me last week and am looking at the variable "education" and see that it is coded 0, 1, and 2, but I don't see any labels for those codes. Do you know what these codes for the "education" variable stand for?"

For the second type of communication, in which you are seeking feedback because of a puzzling or unexpected issue with your analysis, more background information will be needed, but complete background information for the overall project may not be. To illustrate this concept, let's assume that you have been examining the relationship between height and lung function and you construct a scatterplot, which suggests that the relationship is non-linear as there appears to be curvature to the relationship. Although you have some ideas about approaches for handling non-linear relationships, you appropriately seek input from others. After giving some thought to your objectives for the communication, you settle on two primary objectives: (1) To understand if there is a best approach for handling the non-linearity of the relationship, and if so, how to determine which approach is best, and (2) To understand more about the non-linear relationship you observe, including whether this is expected and/or known and whether the non-linearity is important to capture in your analyses.

To achieve your objectives, you will need to provide your audience with some context and background, but providing a comprehensive background for the data analysis project and review of all of the steps you've taken so far is unnecessary and likely to absorb time and effort that would be better devoted to your specific objectives. In this example, appropriate context and background might include the following: (1) the overall objective of the data analysis, (2) how height and lung function fit into the overall objective of the data analysis, for example, height may be a potential confounder, or the major predictor of interest, and (3) what you have done so far with respect to height and lung function and what you've learned. This final step should include some visual display of data, such as the aforementioned scatterplot. The final content of your presentation, then, would include a statement of the objectives for the discussion, a brief overview of the data analysis project, how the specific issue you are facing fits into the overall data analysis project, and then finally, pertinent findings from your analysis related to height and lung function.

If you were developing a slide presentation, fewer slides should be devoted to the background and context than the presentation of the data analysis findings for height and lung function. One slide should be sufficient for the data analysis overview, and 1-2 slides should be sufficient for explaining the context of the height-lung function issue within the larger data analysis project. The meat of the presentation shouldn't require more than 5-8 slides, so that the total presentation time should be no more than 10-15 minutes. Although slides are certainly not necessary, a visual tool for presenting this information is very helpful and should not imply that the presentation should be "formal." Instead, the idea is to provide the group sufficient information to generate discussion that is focused on your objectives, which is best achieved by an informal presentation.

These same principles apply to the third type of communication, except that you may not have focused objectives and instead you may be seeking general feedback on your data analysis project from your audience. If this is the case, this more general objective should be stated and the remainder of the content should include a statement of the question you are seeking to answer with the analysis, the objective(s) of the data analysis, a summary of the characteristics of the data set (source of the data, number of observations, etc.), a summary of your exploratory analyses, a summary of your model building, your interpretation of your results, and conclusions. By providing key points from your entire data analysis, your audience will be able to provide feedback about the overall project as well as each of the steps of data analysis. A well planned discussion yields helpful, thoughtful feedback and should be considered a success if you are left armed with additional refinments to make to your data analysis and thoughtful perspective about what should be included in the more formal presentation of your final results to an external audience.

Style

Although the style of communication increases in formality from the first to the third type of routine communication, all of these communications should largely be informal and, except for perhaps the focused commmunication about a small technical issue, jargon should be avoided. Because the primary purpose of routine communication is to get feedback, your communication style should encourage discussion. Some approaches to encourage discussion include stating up front that you would like the bulk of the meeting to include active discussion and that you welcome questions during your presentation rather than asking the audience to hold them until the end of your presentation. If an audience member provides commentary, asking what others in the audience think will also promote discussion. In essence, to get the best feedback you want to hear what your audience members are thinking, and this is most likely accomplished by setting an informal tone and actively encouraging discussion.

Attitude

A defensive or off-putting attitude can sabotage all the work you've put into carefully selecting the audience, thoughtfully identifying your objectives and preparing your content, and stating that you are seeking discussion. Your audience will be reluctant to offer constructive feedback if they sense that their feedback will not be well received and you will leave the meeting without achieving your objectives, and ill prepared to make any refinements or additions to your data analysis. And when it comes time to deliver a formal presentation to an external audience, you will not be well prepared and won't be able to present your best work. To avoid this pitfall, deliberately cultivate a receptive and positive attitude prior to communicating by putting your ego and insecurities aside. If you can do this successfully, it will serve you well. In fact, we both know people who have had highly successful careers based largely on their positive and welcoming attitude towards feedback, including constructive criticism.

NOTE: Part of this reading was taken from *The Art of Data Science* by Peng & Matsui.