Understanding Your Objective

## In the Beginning…

Though taboo to openly debate with a children book writer, we think @Seuss1990 may have gotten it wrong. To be clear, we have no doubt that learning to code and conduct analytics will take you to many exciting places that are difficult to foresee. We also think that exploring novel data sets to see what you discover is a stimulating way to spend a casual afternoon. However, rather than simply leaving open the places you will go, any good research or applied project (read: likely to be efficiently successful) starts with a clear objective and well-defined deliverable. That is, you should know where you are going.

This chapter is about three things related to “knowing where you are going” with analytic tasks and projects. The first involves figuring out exactly what needs to be done. Oftentimes you might hear seemingly straightforward questions in need of a data-based answer such as “Which patients do we do better with?”, “Which staff have clients who show better progress?”, or “What is the biggest driver of canceled appointments?” Just like operational definitions (e.g., *CITATION*), data definitions can be nuanced, varied, and change over time. Different ways in which we define different data components of a question will lead to different answers. Thus, the first critical analytic step is to pin down exactly what the analytic deliverable will be.

The second item this chapter is about is clearly specifying who will be involved. This comes in two forms. First, who will be involved in actually pulling off the analytic project so that the deliverable can be, well, delivered. Second, who will use the answer once known and how will they use it. As a question, what is the function of these data within the end-users workflow? To meet this function well, those involved in analytic projects need to clearly identify the roles of each team member in delivering the deliverable and how the deliverable will be used once delivered.

The third item this chapter is about is scoping your work to fulfill the identified function. Stories have been told of mythical humans who luxuriously meander through project ideation and framing, who can complete projects on their own schedule and at their own pace, and who can continuously tinker until everything is exactingly perfect. For the rest of us, deadlines are real. Dissertations need to be done. Projects need to be presented. And, deliverables need to be delivered. All of which need to be on time and to budget. Once an analytic deliverable is clearly outlined, people want to know when it will be delivered. Thus, the final section of this chapter reviews how to take the information gathered about deliverables and roles to determine how long it will take you to do the voodoo that you do so well.

## Specifying the Deliverable

Specifying the deliverable is an act of translation. Specifically, translating someone’s question about the world into two components: (1) the data that can best answer the question, and (2) the analytic method that can best answer the question.

### Getting the Best Data.

Identifying the data that can best answer an analytic question might be a new skill for some behavior analysts. Behavior analysts *are* trained in how to create operational definitions for behaviors of interest and intervention conditions (e.g., baseline vs. DRA). Given something we want to track or know about, behavior analysts are trained to define exactly what the behavior or environmental event looks like. In so doing, relatively precise data can be collected (e.g., frequency, duration, latency, treatment fidelity) specific to the occurrence of a phenomenon (e.g., self-injurious behavior; implmenting treatment). Note, however, this is a future-oriented task. I have something I am interested in. I develop a definition. I then collect data to answer that specific question (e.g., “What is the function of a behavior?”; “What is the client’s current manding repertoire?”). This “data-defined-as-what-I-will-do” is the common approach in clinical and research settings.

As a data analyst or data scientist, you also will answer analytic questions in a past-oriented and future-oriented way. That is, given a question in need of a data-based answer, we have to look at what data we already have and that was likely defined previously and not by us. We also have to consider what data would be ideal to answer the question perfectly. For example, if I am interested in knowing which patients our clinic tends to do better with, I would look at what data I have now around each patient’s overall clinical presentation, what data I have now around measuring progress consistently across all patients, and what the ideal data might be for each of these areas. Assessing your data in this way lets you know: (1) what is possible with the data you have right now; (2) what are the limitations to our answer with the data we have right now; (3) what might it take to generate the ideal data sets in terms of time, effort, and cost; and (4) whether collecting those ideal data sets will be worth it.

The two paragraphs above highlight that conducting analytics in practice may require behavior analysts to think differently about how they ask and answer questions with data. The questions likely asked of a data analyst or data scientist rarely will be things where the data can be collected in days or weeks for subsequent analysis. It can take months or years to build the ideal data set for a specific question. But, this doesn’t mean you have to sit on your hands and do nothing while you wait. As noted above, people want answers and information that can help them make better decisions today—even if those answers are not perfect.

The main takeaway of this section is that we rarely have access to data that perfectly matches a question in need of a data-based answer. This makes identifying the “best” data set a relative term. Whatever data we have to answer a data-based question is unlikely to be perfectly ideal and that data set may never arrive. But, by looking at what is available, we can identify what data most closely matches the question at hand, what questions are possible to answer with the data we have, and how we might want to change our data collection systems moving forward.[[1]](#footnote-21) In short, matching your data to a question is likely to shape exactly what deliverable(s) is (are) possible.

So how do you know you have the best data for the task at hand? In short, it involves becoming an expert in the data collection systems and databases at your company. Early in your, this will likely involve asking someone who knows the databases and data collection systems better than you.[[2]](#footnote-22). Over time, you will likely become less reliant on others as you learn about the unique databases, tables, fields, meta-data, and (gasp!) Excel docs floating around your company and how to access them.

### Choosing the Best Method.

Assuming you have identified what data you will use to complete some kind of analytic project, the next step surrounds choosing the best method to complete the project. Later in the book we’ll chat about how to implement and choose among the many different types of analytic behaviors you will engage in. Here, however, we are talking about: (1) types of deliverables; (2) the function of the deliverable in the end-users’ daily workflows; and, (3) the frequency that the analytic task needs to occur moving forward.

Like coffee and breakfast bagels, different people likely prefer their analytics presented differently. Table 1 shows common deliverables you might be asked to produce. For example, some people may prefer a fully functioning dashboard with 6-8 unique graphs, dynamic filters, trend lines, statistical significance clearly indicated, and the ability to export the underlying data if they choose. For others, a simple data extract as an .xlsx or .csv file will do. Still others will want a 5-7 page report of your methods and the insight generated. And, still others will want a quantitative or computational model that can be deployed to make predictions about something in the future. Knowing which of these many options you are being asked to produce will be important for scoping the time it will take to produce the deliverable.

| Type of Data Deliverable | Definition |
| --- | --- |
| Dashboard |  |
| Data Extract |  |
| Full report of methods and findings |  |
| Recommended action for a decision |  |
| Quantitative model for future inference |  |
| < Others > |  |
| < Others > |  |
| < Others > |  |
| < Others > |  |

[Identifying the primary end-user. Who is the audience? What function does the deliverable serve in their daily workflow? How precise does it need to be? Or is directional / order-of-magnitude sufficient.]

[Mocking up deliverable to ensure alignment among stakeholders.]

[Understanding repeat-ability and refreshment requirements (e.g., one-off requests, ongoing maintenance, iteratively improving projects]

## Knowing Your Role

* What is your role as a data analyst?
  + Assume outside of academia
  + Who else is on your team? SMEs
  + Quality validation
  + Setting expectations
* Go from business requests to formulating deliverables, timeline, scope

## Chapter Checklist

* Have some kind of checklist or tool - something tangible to reference
  + Maybe action item list at the end of each section

x

[1] 1 2 3 4

Add a little bit about knowing your role, which includes being a content area expert. Also know what questions to ask and who to ask. You don’t have to have all of the answers. Not a linear end point, it’s a cyclical/iterative process.

1. Before you go lobbying for data definitions and systems of collection to change, it is important to identify what other clinical or operational processes rely on those data. It is possible your current analytic project is of lower priority than other areas and how well the available data fits your need is unlikely to change anytime soon. [↑](#footnote-ref-21)
2. “Hey, Steve—I am trying to find data on patient progress and how it relates to hours per week of ABA. Do you know where this data lives?” [↑](#footnote-ref-22)