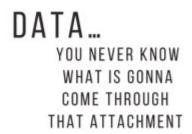To that end, I'd like to present some of my favorite quotes from *Advancing into Analytics*, along with a bit of context and relevant passages from the book.
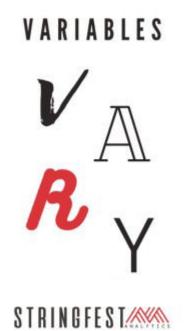

"Data… you never know what's gonna come through that attachment"

This quote is a paraphrase from the opening of the TV show *Pawn Stars*, where Rick Harrison says: "You never know *what* is gonna come through that door." It opens Chapter 1, Foundations of Exploratory Data Analysis. As data analysts, we often peddle in data as varied as Corey and Chumlee do antiques:

… confronted with a new dataset, you never know what you are going to find. This chapter is about exploring and describing a dataset so that we know what questions to ask of it. The process is referred to as exploratory data analysis, or EDA.

Fortunately, exploratory data analysis (EDA) gives us a structured process to size up whatever dataset makes its way to your workstation. Through a series of descriptive and visual techniques covered in this chapter, you'll be poised to test relationships in the data later in the book, using Excel, R and Python.
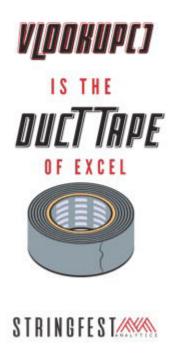
VARIABLES

VARY

STRINGFEST

"Variables vary"

Well, I guess this isn't so much a quote as a tautology, but it's an important one. Have you ever considered why variables have that name? I explain it in Chapter 1 of the book:

We call them variables because their values may vary across observations. If every observation we recorded returned the same measurements, there wouldn't be much to analyze.

In fact, researchers have traditionally struggled to analyze data where over 90% of observations take on the same value. Between larger datasets and increased computing power, this has become less of a constraint, but the *ability to vary* is still what data's all about.

"VLOOKUP() is the duct tape of Excel"

Chapter 5 of *Advancing into Analytics* is titled "The Data Analytics Stack." While the book only covers Excel, R and Python in depth, I felt it was important to situate them in the wider analyst toolkit. In explaining how relational databases work, I like to start off with an example that every seasoned Excel analyst knows well: the mighty `VLOOKUP()`. (Yes, I hear you, `INDEX()/MATCH()` and now `XLOOKUP()` fans. The analogy works for you, too.)

As I explain in the book, "I like to call `VLOOKUP()` the *duct tape of Excel* because of its ability to connect
datasets together." Now, duct tape is a helpful tool. With skilled hands, it can accomplish a lot. But, you may not want to trust duct tape for every job. In data, we have relational database joins to make more stable, efficient connections between data. As I go on to write, "If `VLOOKUP()` is like duct tape, then relational database joins are welders."

"There's a package for that!"

This quote comes from Chapter 6, First Steps with R for Excel Users. In this chapter, readers install the R code base. I draw the following analogy to explain the difference between this code base and packages:

Imagine if you weren't able to download applications on your smartphone. You could make phone calls, browse the internet, and jot notes to yourself—still pretty handy. But the real power of a smartphone comes from its applications, or apps. R ships much like a "factory-default" smartphone: it's still quite useful, and you could accomplish nearly anything necessary with it if you were forced to. But it's often more efficient to do the R equivalent of installing an app: installing a package.

Both R and Python feature staggering collections of packages for all sorts of tasks. To paraphrase the famous slogan: "There's a package for that!" This analogy carries conceptual but also practical value: to use an app, you download it once, but *open* it each time you need it. The same goes in R and Python… so you'll be using `library()` and `import`, respectively, a lot.

"It's your world… the data's only living in it"