1. Introduction – each DataStory begins like this, with a title card and learning objectives. StatCat and DataDog are the primary characters in each story. StatCat is a 1970’s throwback, with a ridiculous academic tone. He specializes in all things statistical. DataDog, on the other hand, is the computer whiz / programmer. She plays off of StatCat’s inflated sense of himself. Character definition is key as we want both of our characters to act in consistent ways, readily identified by the reader.
2. Classic cartoon format – pays homage to the great cartoons of the mid-50’s and 60’s. This story pays homage to a famous cartoon directed by Chuck Jones in 1953 – Duck Dodgers in the 24th and a half century. The depletion of illudium phosdex (the shaving cream molecule) forces Daffy Duck to search for it on planet X. However, he faces Marvin the Martian when he arrives on its surface. This story references the phosdex molecule. But in this case, it’s the key ingredient in an elixir StatCat believes will grant eternal life.
3. Technical note – this is a ShinyApp, running on an RStudio Shiny server. We use the learnr package to deliver code exercises, provide hints and solutions, and deliver multiple-choice quizzes. In this case, the code is R, packaged in a Rmarkdown document. We also plan to develop Python DataStories delivered in a Jupyter Notebook format. Of course, Jupyter Notebooks will change the look and feel just a bit and they don’t support the same level of interactivity.
4. Chapter 1 (Advertisement) – the DataStory learning experience features “advertisements” in certain key points. This was a prominent part of Saturday morning cartoons, what we’re calling Staturday morning cartoons. Ads serve two roles – they provide a break for the student, and they also allow us to revisit and/or review other data science concepts. In learning science, the concept of interleaving one’s practice is well established. The idea is that you practice one activity and then periodically switch to a different activity. These ads are one way of doing that, creating mini-review sessions.
5. Chapter 2 (Coding Assignment) – the student is being asked to assist DataDog in loading a dataset. Coding assignments are mediated by DataDog as she’s the programming whiz. Our pedagogical goal is to create a friendly bond between the student and DataDog.
6. Chapter 3 (Content Delivery) – once the student engages with the story, we then deliver some foundational content in a lecture format. Because we want to avoid a textbook look-and-feel, these are typically short, featuring a rather pompous StatCat.
7. Chapter 3 (Plotting) – our platform supports plotting, as pictured here. The assignment below this plot features a Solution button, in the event the student gets stuck. It also features a Submit button for assignment submission.
8. Chapter 4 (Statistics) – in this DataStory, we present some foundational statistical ideas, the concept of positive, negative, and zero correlation. Eventually, we want to provide a link here to the StatCave where students can get in-depth information on this topic as well as other statistical concepts. The StatCave is StatCat’s lair (think BatCave from Batman and Robin) where we feature traditional statistical textbook content. Likewise, the DataDen is DataDog’s home base with in-depth programming and technical information.
9. Chapter 5 (Data Ethics) – In this story, students learn about positive, negative, and zero correlation. Naturally, StatCat has a vested interest in the data showing a strong positive correlation between the liberal use of Illudium Phosdex elixir and the number of lives a cat lives. More elixir equals more lives. The problem is that the data set StatCat relies on to prove his case was create by his uncle Carbuncle years ago (Chapter 3, Panel 7). If you look closely at this image, you will notice something interesting. Do you see it? Easter egg surprises are part of the DataStory learning experience. In the trash can, we see a printout of an official Census data set. Why isn’t that data used? This raises the ethical question data provenance – how was the data gathered or created? Who oversaw the process? Did they have an agenda, as is possibly the case here? Later in this story, DataDog realizes that this is indeed the case (Chapter 5, Panel 2) when he says, “Uncle Carbuncle was a snake oil salesman…” And then he follows up by saying that he will need to dig through the census data Uncle Carbuncle discarded. DataDog spends the night working up the census data and then he confronts StatCat with the truth (Chapter 6, Panel 3). This is the defining moment in the story, the climax. Will StatCat grow in his understanding of his ethical and moral obligations? Or will he dodge them? In this case, he formulates a new hypothesis (Chapter 6, Panel 6). He engages in what is called “data dredging.” Because StatCat is so intent on ‘proving’ that the frequent and liberal use of Illudium Phosdex elixir leads to immortality, he crosses an ethical line at this point. Rather than examine the data objectively, he seeks to impose his will on it. Thus, we close the story with a question mark, thereby providing an opening for the instructor to have an open-ended ethical conversation. Was StatCat acting in an ethical manner?
10. What we have discovered over the past year is that a sequential art format allows us talk about more than just technology. It allows us to display emotions and motivations, creating opportunities for data science educators to talk about ethics, positive and negative ways of conducting data analysis.
11. Reference – This chapter summarizes the functions used in the story.

Design Principles

1. Humor – Humor is an important part of each DataStory. After all, these are cartoons. Although sequential art is often derided as superficial, an inappropriate medium for delivering serious content, we believe that it can be a powerful pedagogical tool. Fear shuts down the learning process while humor opens it up. It’s impossible to be fearful when you’re laughing.
2. Dual Coding – Because sequential art combines a text-based storyline with a visual depiction of it, dual coding theory serves as an ideal theoretical framework in that it concerns itself with the impact of text in combination with visuals on the learning process. Indeed, dual coding theory posits that the human mind processes incoming information in two channels, one for visuals, a second for verbal input. Hence, learning experiences which combine the two should be more effective, provided their design reflects best-practices.