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” from time to time. These serve two roles – they provide a break for the student and they also provide a way for 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 in this learning environment.
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) – Because StatCat is so intent on “proving” that the frequent and liberal use of his uncle’s Illudium Phosdex potient leads to immortality, he crosses an ethical line. Rather than examine the data objectively, he seeks to impose his will on it. In this chapter, DataDog wises up to StatCat’s shenanegans and discovers that uncle Carbuncle’s data is totally fabricated – the data used to “prove” the relationship between Phosdex concentration and cat lives lived. He finds a more reputable dataset from the Census bureau that disproves everything StatCat believes. A story format allows us to 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.