



Holy Deviations!- Stories, Superheroes, and Data Science Education

- Charlotte Bolch, Midwestern University
- Daniel Maxwell, University of Florida
- Justine Andersen, Sequential Artist
- Elli Rhodes, Sequential Artist

Big data from multiple sources is a common situation that students will face...



...but their technical foundation often looks like this.

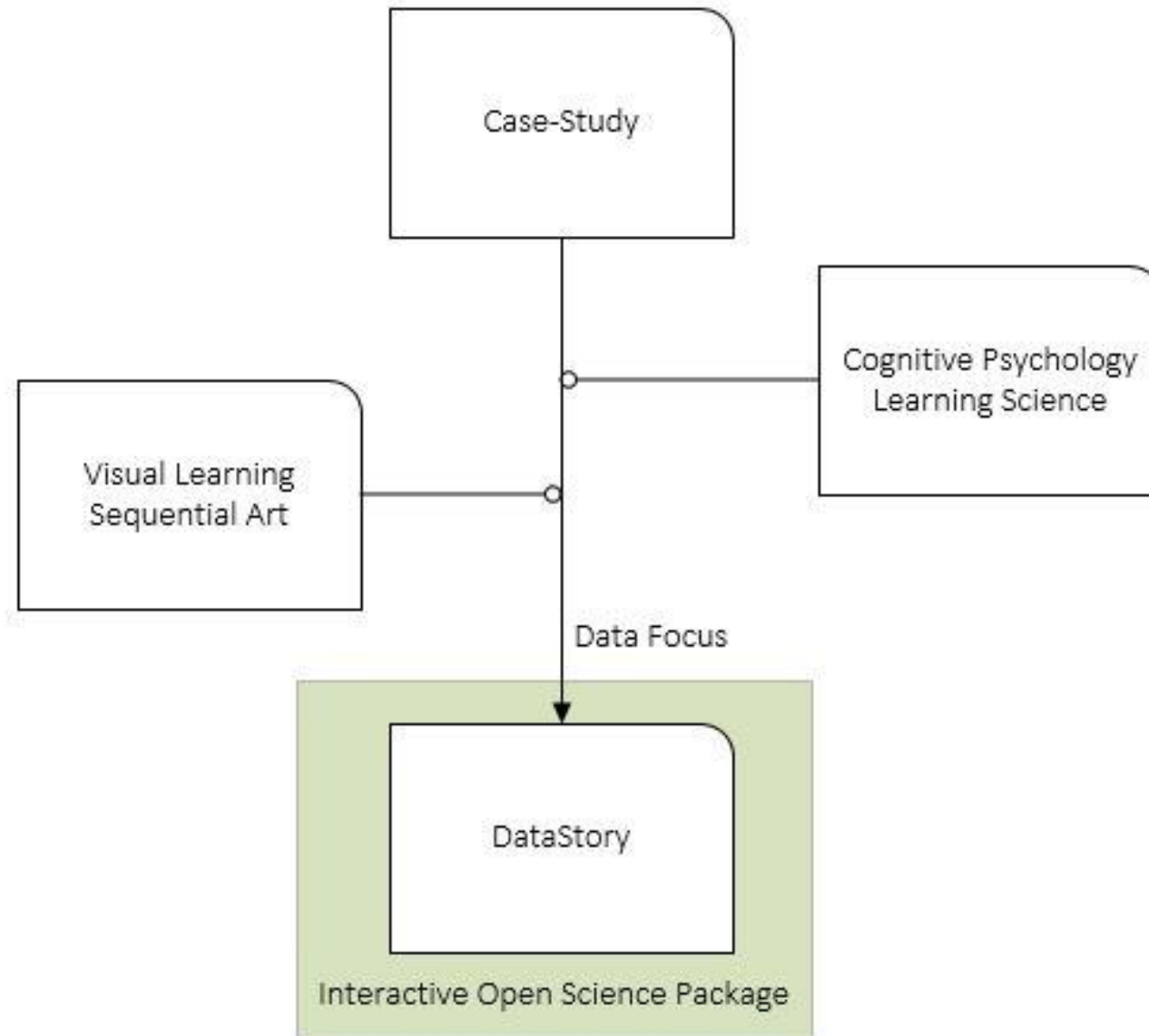


...and technical training can be overwhelming and boring.



Solution: DataStory

- DataStory is a combination of sequential art (comics), a programming language and data!
- The primary objective is to educate students about foundational data science topics while engaging them in a journey through a narrative arc.
- Focus on integration of data, conflict, action, character development, while being mindful of the density of information.



Focus Group Study

- Invited participants to participate in a 1.5 hour Zoom call
- Research Questions:
 - What are the participants' reactions to the DataStory (characters, storyline, etc.)?
 - What are the participants' feedback about the DataStory in terms of the flow, content (learning objectives), and interactivity (engaging students in the data learning events)?
- Five Subjects Participated in Initial Testing of the Sardine DataStory
 - 4 PhD Students, 1 Faculty
 - Varying levels of prior knowledge about statistics, data science, and programming language

A Sardine Tale

Title Page

Introduction

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Ending

Exercises

Start Over



Learning Objectives:

- 1 – Importing and Cleaning New Datasets in R
- 2 – Visualizing Continuous Variables in R
- 3 – Conducting Smoothing Techniques in R: LOESS Line
- 4 – Calculating and Interpreting Pearsons Correlation Coefficient in R
- 5 – Conducting Regression Techniques in R: Simple Linear Regression



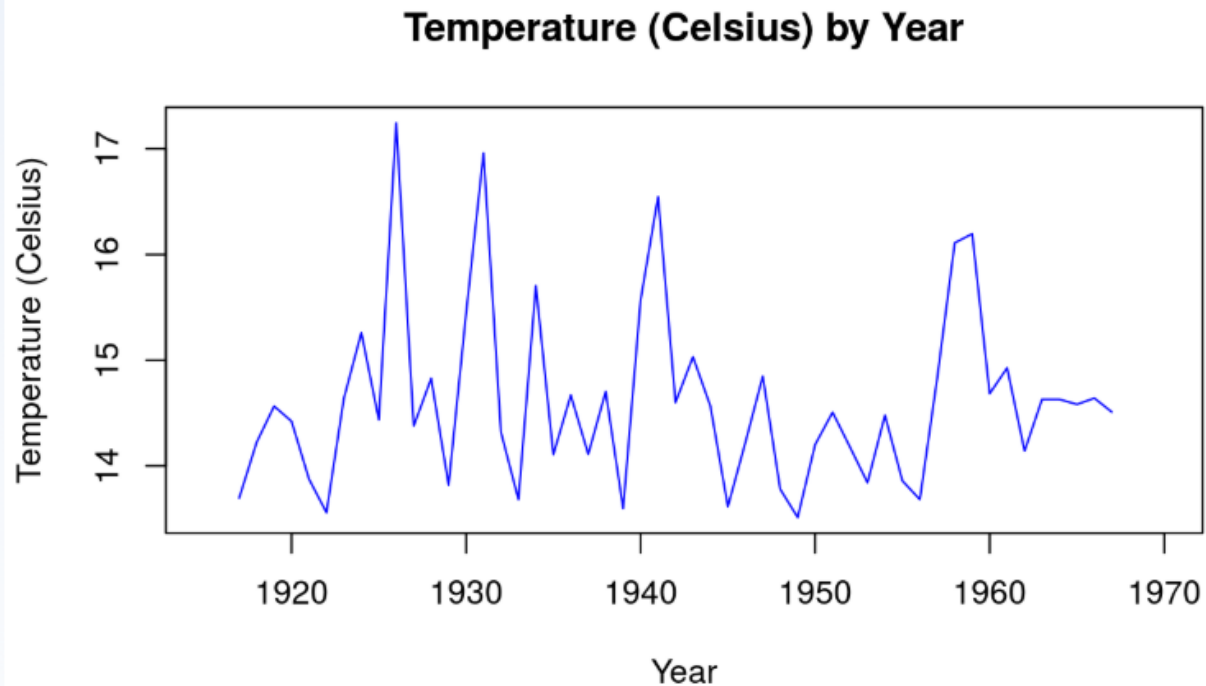
Now that our code is perfect and we have a dataset that contains average ocean temperatures by year, we need to create a graph to see what is going on in the data. Let's create a graph with the year on the x-axis and the average ocean temperature on the y-axis.

Code

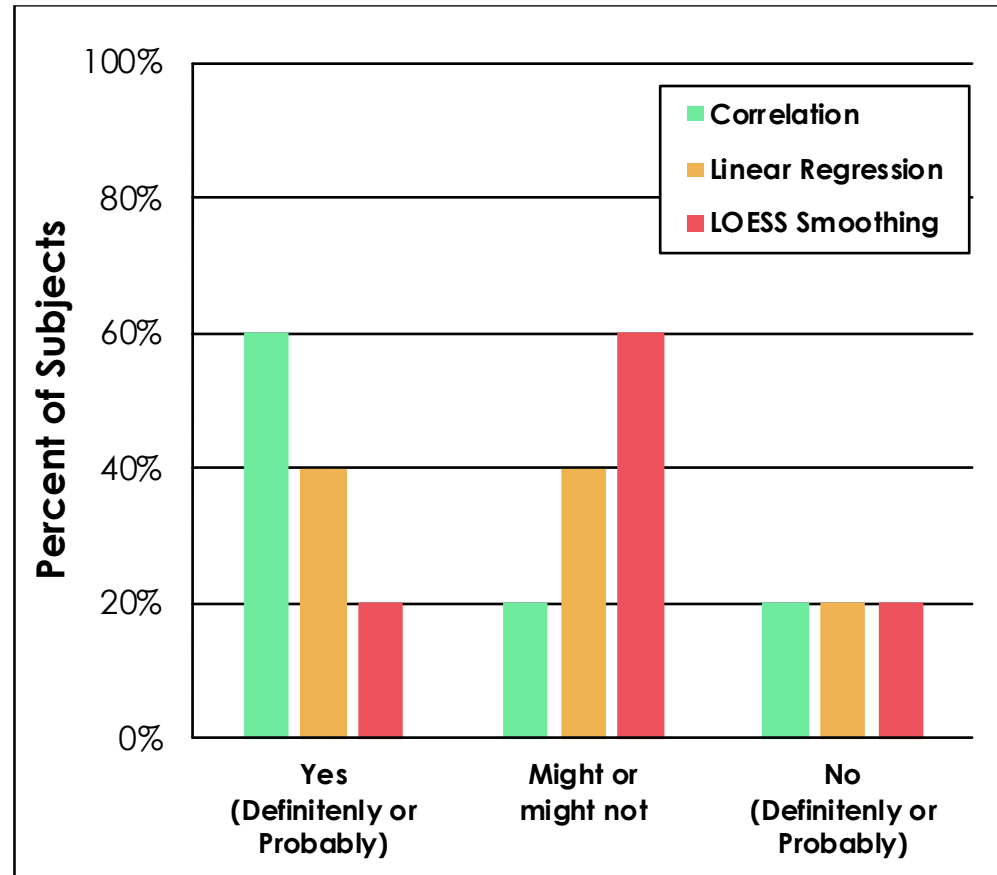
[Start Over](#)

[Run Code](#)

```
1 plot(avg_temps_df$year, avg_temps_df$avg_temp, # x-axis, y-axis.  
2     type = "l", # Type of graph is a line graph ("l" for lines).  
3     col = "blue", # Color of the lines.  
4     main = "Temperature (Celsius) by Year", # Overall title for the plot.  
5     xlim = c(1915, 1970), # Range of values for the x-axis.  
6     xlab = "Year", # Label for the x-axis.  
7     ylab = "Temperature (Celsius)") # Label for the y-axis.
```



After completing the DataStory, participants' level of agreement with the following learning objectives:



Categories for Improvement

Visuals

- Visual of characters
- Visual of learning content

Flow or Progression

- Introduction length
- Referencing past section

Education Delivery

- Technicality
- Vocabulary in the story

Interactivity

- Coding in Story
- Exercises

Narrative

- Character traits
- Story-line (dramatic)

Usability

- Confusion about role
- Sound issues

DataStory version 2

The Carbuncle Data Debacle

Title Page

Chapter 1

Chapter 2

Chapter 3

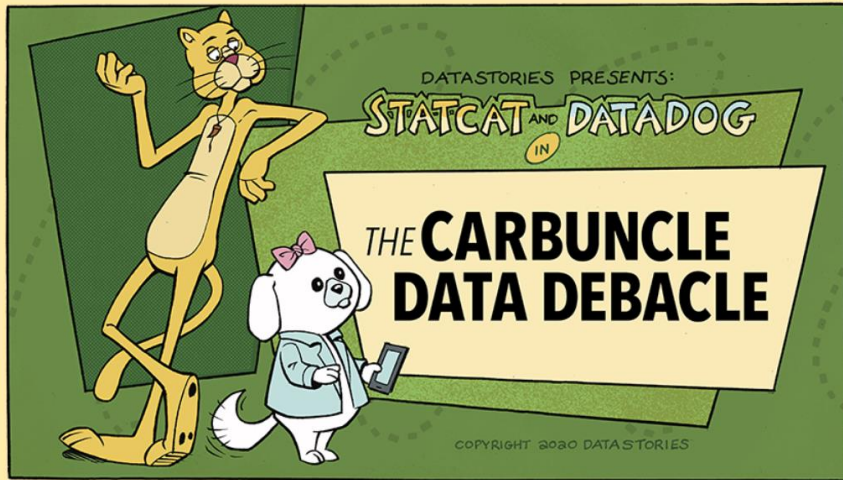
Chapter 4

Chapter 5

Chapter 6

Reference

Start Over



- Learning Objectives:
 - Importing and Understanding Dataframes in R
 - Developing an Analysis Plan
 - Joining/Merging Dataframes in R
 - Calculating and Interpreting Pearson's Correlation Coefficient in R
- Characters:
 - DataDog
 - StatCat

Poster Session Q&A: June 29th 3-3:45pm EST

Please play around with the DataStory!

DataStory Link for The Carbuncle Data Debacle:

https://mydatastory.shinyapps.io/phosdex_story_main_color/