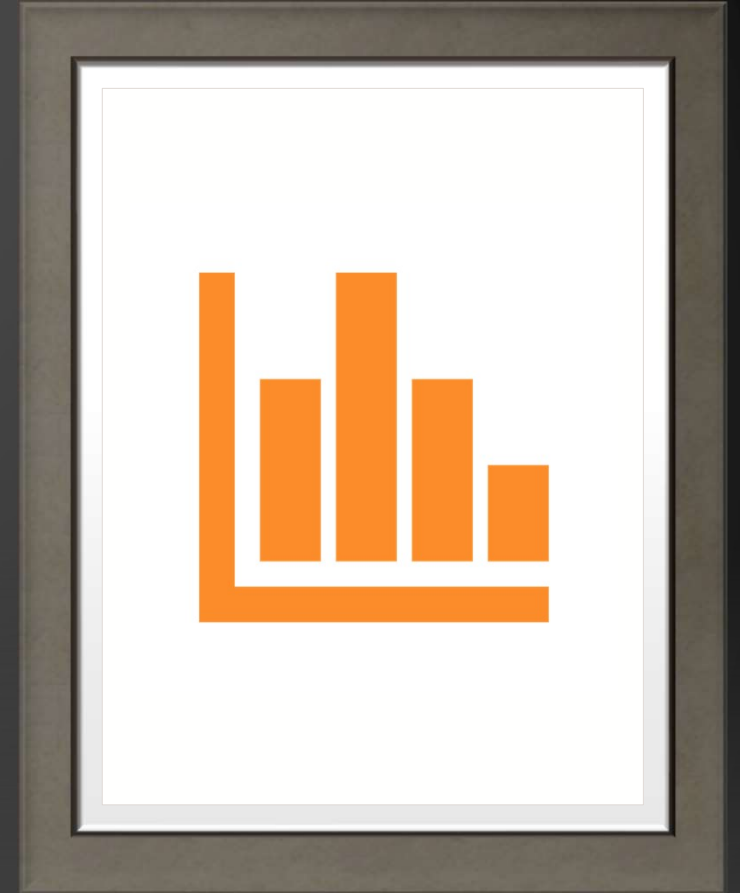
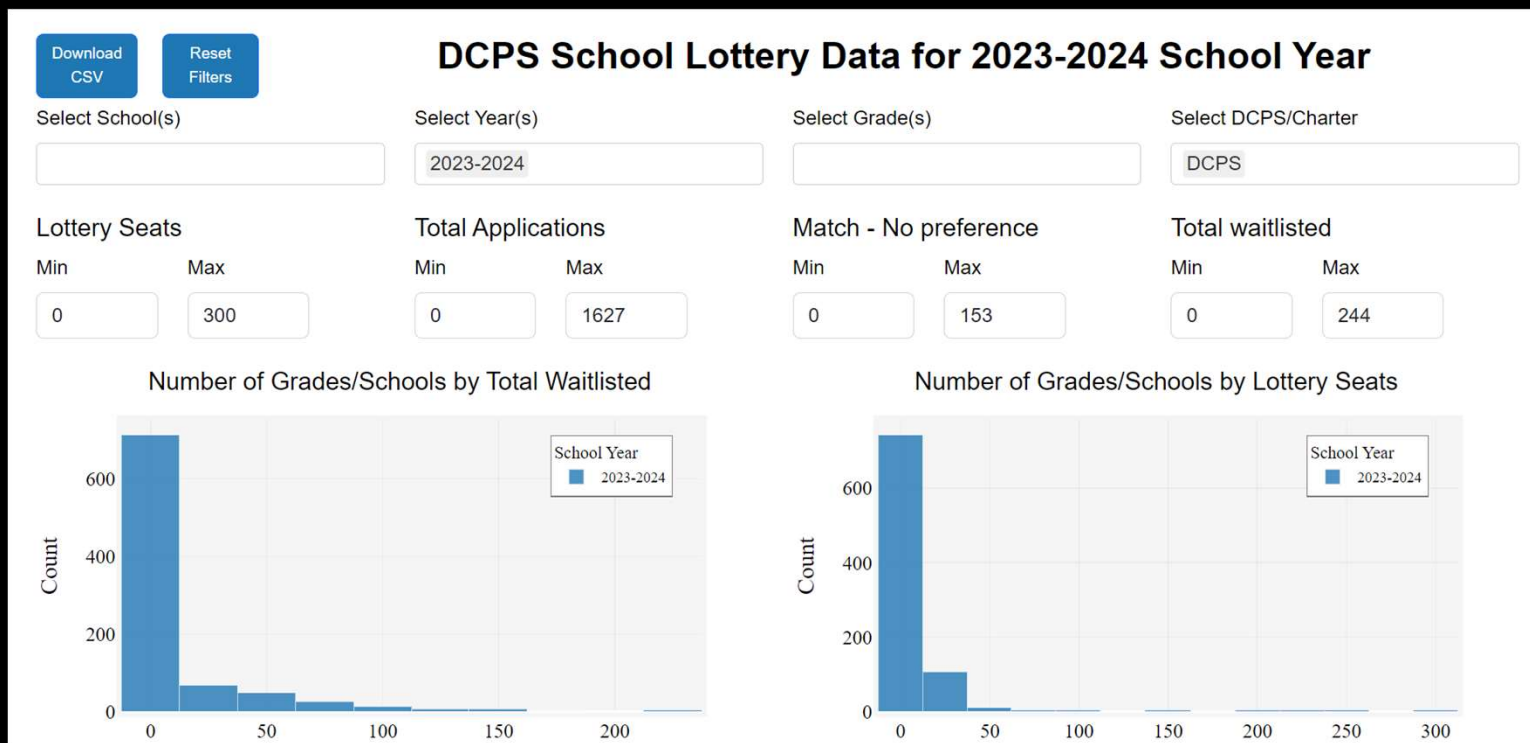


# TRANSITIONING TO CODING FOR VISUALIZATION VIA GPT

ABIGAIL HADDAD



# MY SECOND DASHBOARD, IN SHINY FOR PYTHON



## THE TWO-MINUTE VERSION

---

Think of GPT like your autodidact intern; give small, detailed assignments, expect some oddness

---

Coding can let you do more, faster, and more transparently – GPT can help with each step

---



# HELLO

Public Policy Ph.D.

Research/stats/data science for the Department of Defense

Lead Data Scientist

You can find this

here: <https://github.com/abigailhaddad/slides/>

You can also find me here:

<https://presentofcoding.substack.com>

# PROMPTING FOR CODE: BE REALLY SPECIFIC

---



**INPUTS:** What does your data look like? It is a table? What does each row represent? What does each column represent? What are your column names?



**OUTPUTS:** Do you want a map which shows US states? Which has roads on it? A stacked bar chart? A regular bar chart? Is something aggregated? Cleaned in some other way



**GENERAL GUIDANCE:** R? Python? Go slowly, explain each step

# SPECIFIC PROMPTS FROM TODAY THAT GOT WORKED-ON- THE-FIRST-TRY CODE



I'm trying to show trade offs between false negatives and false positives. I have three variables I vary in my model: small vs. large, count (1, 100, or 1000) and tolerance (.4, .5, .6, .7, .8, .9). I have two tables: one with rate of misidentification among true positives, one with rate of misidentification among true negatives, by my varying variables. I also have line charts. How can I show in a chart this trade off between false positives and negatives?



I'm interested in something that looks like a ppf or utility curve? I think we could do that for each of the lines that represents a permutation of the three variables



I think I want to try this. Can you write me Python code? I have two tables as inputs, df\_pos and df\_neg. Columns for both are small\_1, small\_10, small\_100, large\_1, large\_10, large\_100. Index is .4, .5, .6, .7, .8, .9. I like plotly



Can you show me how to add the index info from the original dfs to the hovertext?



No, sorry, I want to show the curve that's the lowest point ? Like for each rate of false positive, the point that's the lowest rate of false negatives

## OTHER INFO TO GIVE GPT (ESPECIALLY FOR NEW LIBRARIES)

---



DOCUMENTATION



EXISTING CODE





## WHY TRANSITION TO CODING FOR DATA VISUALIZATION?

- **Reusability**
- **Version Control**
- **Transparency/Documentation**
- **Separation of Code and Data**
- **Ownership/Reduced Vendor Lock-In:**
- **Customizability**



## GPT CAN HELP YOU WITH THOSE, TOO

---

“How do I get started on GitHub?”

“I’m getting this error message from git. What should I do?”

“This is my code. Can you help me make it more modular?”

“This is my code. Can you write me a readme file for my git repo?”

“Can you write me function-level documentation for all of my functions?”

## WHAT IF YOU LIKE CODING?

- Learn some of the basics
- There are resources at every level of price and structure

## A FEW MORE NOTES AND QUESTIONS

- **There's going to be a Data Science DC talk on transformers, which are the technology actually behind LLMs. (as in, the T in generative pre-trained transformer, or GPT. John Kaufhold is the speaker.**
- GPT 4 > GPT 3.5. I didn't like Bard. I haven't used Copilot yet. I hear that trained versions of the leaked Facebook model (LLAMA) are getting as good as GPT-4; I used a smaller version that I'm running locally, and it's not there yet

slido



Questions?

① Start presenting to display the poll results on this slide.