

Data Visualization Assignment

ECON 370: Data Science for Economic Analysis

Department of Economics

Williams College

Fall 2024

For this project, you will use the provisions data to explore a specific research question, and then present your conclusions in the form of one or two data visualizations. You will present your results in the form of three or four slides (or pdf pages):

- A title page with the title of your project and the names of your group members
- A page explaining the research question that you explored in your analysis
- Your visualizations, complete with titles and notes explaining them

Given the number of groups and the amount of time we have available, you'll be presenting your results in a poster session. Each group will affix the four pdf pages (which you will submit before midnight on Thursday September 26, and I will print out at 6:00 AM on Friday September 27) to a piece of poster board which we will mount on the wall. At least one group member needs to stand next to the poster at all times to answer questions from me and your fellow students.

Potential Research Questions:

1. How have sales and revenues been evolving over time? What are the factors that predict unusually high-volume days? Is the composition of sales different on those days?
2. How does the weather impact sales? Are sales higher/lower on rainy days? When it is particularly warm or cold? For this question, you will need to merge the data with information on local weather conditions.
3. How does purchasing behavior differ by gender? The transactions data set contains information on the first names of some customers (though they may not be a representative sample?). What proportion of customers are women? Do they purchase different types of products than men? Or is their behavior as customers different in some other way? Do men tend to buy beer? Do women tend to buy wine?
4. How can Peter (the owner of Provisions) identify products that aren't selling? Which types of products are the most/least reliable in terms of sales? Are there specific items that should be cut from the catalog?
5. What happens when a new product is introduced? How long does it take for a new product to start selling? How do sales evolve over time? How can we determine whether a new product is a keeper?

6. What share of sales/revenue comes from returning or regular customers (and how would we define them?), and what share comes from one-time customers? How do their purchasing patterns differ?
7. How long would it take Provisions to sell their current inventory? Which products might they be holding too much of, and which products (or categories) tend to sell out? How might they adjust their inventory practices to increase sales or profits?
8. Another research question of your choosing.

What We Are / I am Looking For:

1. How interesting is your question? Do we learn something about patterns of consumer demand that is of interest to economists? Do we learn something that will help Provisions increase profits or provide better service to customers? (The answer to at least one of those questions should be yes.)
2. How much have you done to answer your research question? Have you attempted visualizations that are challenging, and that go beyond what we've done in the labs? Do your visualizations provide new and interesting insights?
3. How do your visualizations look? Have you followed the “rules” for good data visualizations (or violated them intentionally to achieve some specific end)?
4. Does your code process the data and generate your visualizations?