PROCESS BOOK — Final Submission

Team Members: Cooper Hartzog, Ayden Sabol, Matthew Myers

(Superstore Sales Dashboard)

1. Overview and Motivation

Our project explores patterns in retail performance using the Superstore dataset. This dataset contains detailed information on sales, profit, customers, products, and geographic regions across the United States.

The motivation behind this project is to build an interactive data visualization dashboard that helps users quickly understand which product categories generate the most revenue, how sales relate to profit, how performance changes over time, and how different states and regions contribute to total sales.

The dashboard allows business users to identify trends, outliers, and regional differences in an intuitive visual format through coordinated, interactive views.

2. Related Work

Several dashboards and visual studies of the Superstore dataset exist online, especially in Tableau Public. These examples inspired our design, including Tableau's "Superstore Sales Dashboard" focusing on category breakdowns, D3-based scatterplots from ObservableHQ, and geographic sales heatmaps commonly used in BI tools like Power BI.

In class, examples such as bar charts, scatterplots, and maps helped us understand visual encodings and interactions. We incorporated principles such as color consistency, reducing chartjunk, choosing appropriate marks for different data types, and—critically—space allocation based on information density.

3. Questions

Our project focuses on the following questions:

- Which product categories drive the most sales?
- Is there a relationship between sales and profit at the transaction level?
- What trends exist in sales over time?
- How do different states contribute to overall sales? Which regions perform best?

During early exploration, we also considered profit by sub-category, quantity vs discount effects, and customer-level performance. However, we narrowed our scope to four core visualizations to keep the dashboard focused and clean.

4. Data

The dataset used is Superstore.csv, provided in earlier modules. It contains approximately 10,000 rows spanning 2011–2014 and includes order details, customer information, product categories, profit and sales metrics, and state/region/city for geographic mapping.

Data Preparation

Steps we took:

- Converted numeric fields (Sales, Profit, Quantity, Discount)
- Parsed Order Date into JavaScript Date objects with validation to filter invalid dates
- Aggregated sales for categories, months, and states using D3 rollup functions
- Cleaned inconsistent state values

No major missing values were found.

5. Exploratory Data Analysis

We performed initial exploration using Python and spreadsheets:

- A pivot table showed Technology and Furniture dominate total sales
- Scatterplots showed profit has strong variance, including many negative values
- Time-series plots showed seasonal fluctuations with late-year spikes
- Summaries by region showed the West and East regions dominate performance

These insights helped determine the four visualizations in our dashboard and informed our decision to emphasize the scatterplot, which reveals the most complex relationships in the data.

6. Design Evolution

Early Sketches (Milestone 3)

Our first sketches included a category summary bar chart, a dual-axis line chart, and a bubble map in a 2×2 equal-space grid layout.

Design Iteration Based on Feedback

Professor feedback highlighted that the equal space allocation was not justified—the bar chart and line chart are synthetic views that don't need much space, while the scatterplot is information-dense and was cramped. Based on this feedback, we made significant layout and design changes.

Final Design Principles Applied

- Asymmetric layout: Bar and line charts are compact; scatterplot gets full width
- Bar chart for categorical comparison (Technology, Furniture, Office Supplies)

- Scatterplot for correlation analysis with enhanced readability features
- Line chart for temporal trend (monthly sales 2011–2014)
- Choropleth map for geographic context (replaced original bubble map)
- Dark theme for better visual contrast and hierarchy
- Coordinated filtering across all views via clicking/hovering
- Clear titles, consistent spacing, and tooltips throughout

7. Implementation

The dashboard was implemented using HTML/CSS for layout, D3.js v7 for all visualizations, TopoJSON/US Atlas for the US map (loaded from CDN), and GitHub Pages for hosting.

Visualizations

- Sales by Category bar chart — compact, with rotated labels and hover effects
- Monthly Sales Trend line chart — shows 2011–2014 with proper date parsing
- Sales by State choropleth map — clickable states with color gradient
- Sales vs Profit scatterplot — full-width with enhanced features

Scatterplot Enhancements

Based on professor feedback to improve readability, the scatterplot received significant enhancements:

- Better scales with expanded domains and padding to spread points
- 60% opacity for overlapping point visibility
- Grid lines for easier value reading
- Red dashed zero-profit reference line
- Color-coded by category with legend
- Detailed tooltips showing category, sales, profit, and margin
- Hover effects (circles grow and become fully opaque)

Map Redesign

The original bubble map had visibility issues, so we converted it to a choropleth:

- States filled with sequential color scale (white to red) based on sales
- Uses US Atlas TopoJSON from CDN for reliable loading
- Square root scale for better differentiation at lower values
- Click any state to select/filter other views
- Yellow border highlights selected state

Coordinated Views & Interactions

All four visualizations are now connected through coordinated filtering:

- Hovering over a category bar filters the scatterplot to show only that category's transactions
- Clicking a state on the map filters the bar chart (shows overlay), line chart (shows state trend), and scatterplot
- Click again to deselect and return to the full dataset view
- All charts include tooltips on hover for detailed data exploration

Dark Theme

We implemented a dark theme (navy/purple background with white text) to improve contrast. This was especially important for the choropleth map, where lighter-colored states were hard to see against a white background.

8. Evaluation

Key Insights Discovered

- Technology is the highest-grossing category
- Many transactions show high revenue but low or negative profit—the scatterplot reveals this clearly
- Late-year months (Q4) show consistently higher sales, likely holiday-driven
- California, New York, and Texas dominate total sales

Design Improvements Made

- Tooltips added to all charts
- Coordinated filtering by category and state
- Map redesigned from bubbles to choropleth for cleaner appearance
- Space reallocation based on information density
- Dark theme for improved contrast
- Fixed label cutoff issues on bar chart and line chart axes

Conclusion

The final dashboard successfully answers our core questions and demonstrates the importance of design iteration. The key takeaway from the data is that profitability varies dramatically even within high-revenue categories—some of the biggest transactions actually lose money. The coordinated filtering allows users to drill down from any angle: by category, by state, or by exploring individual transactions in the scatterplot.