PROCESS BOOK — Draft for Milestone 3

(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
- 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.

---

2. Related Work

Several dashboards and visual studies of the Superstore dataset exist online, especially in Tableau Public. These examples inspired our design:

- Tableau's "Superstore Sales Dashboard" focusing on category breakdowns
- D3-based scatterplots from ObservableHQ
- 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, and choosing appropriate marks for different data types.

---

3. Questions

Our project focuses on the following questions:

1. **Which product categories drive the most sales?**
2. **Is there a relationship between sales and profit at the transaction level?**
3. **What trends exist in sales over time?**
4. **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
- Customer-level performance

But 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 ~10k rows and includes:

- Order details
- Customer information
- Product categories

- Profit and sales metrics
- State, region, and city for geographic mapping

Data Preparation

Steps we took:

- Converted numeric fields (Sales, Profit, Quantity, Discount)
- Parsed Order Date into JavaScript Date objects
- Aggregated sales for categories, months, and states
- 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 Furniture + Technology dominate total sales
- Scatterplots showed profit has strong variance, including negative values
- Time-series plots showed seasonal fluctuations
- Summaries by region showed the West and East regions dominate performance

These insights helped determine the four visualizations in our dashboard.

---

6. Design Evolution

Early Sketches

Our first sketches included:

- A category summary bar chart
- A dual-axis line chart
- A bubble map

We iterated to improve clarity and decided on a **2×2 dashboard grid** layout.

Design Principles Applied

- **Bar chart** for categorical comparison
- **Scatterplot** for correlation (sales vs profit)
- **Line chart** for temporal trend
- **Choropleth map** for geographic context
- Color choices balanced between distinct palettes
- Limited gridlines and reduced noise
- Clear titles and consistent spacing throughout

We originally planned interactive filtering but postponed it for Milestone 4/5.

---

7. Implementation

The dashboard was implemented using:

- **HTML/CSS** for layout
- **D3.js** for all visualizations
- **GeoJSON** for US maps
- **GitHub Pages** for hosting

Figures included in the dashboard:

- Sales by Category bar chart
- Sales vs Profit scatterplot
- Monthly Sales Trend line chart
- Sales by State choropleth map

Each chart is defined in the `script.js` file with modular functions, clean axes, tooltips, and color scales.

---

8. Evaluation

What we learned:

- Technology is the highest-grossing category
- Many sales transactions show high revenue but low or negative profit
- Late-year months show consistently higher sales
- California, New York, and Texas dominate total sales

Future Improvements:

- Add tooltips to all charts
- Add filtering by region or category
- Add transitions between states in the scatterplot
- Add legends to the map
- Improve color accessibility

Overall, the prototype successfully answers our core questions and forms a solid foundation for the final dashboard.