

Project Work Division Plan: Tracing Federal Dollars (VA, CA, MD)

Robel Endashaw, Jonathan Hale, Daniel Berhane, Benjamin Tanowitz

April 23, 2025

Project Overview

This project aims to visualize U.S. federal contracting data from USASpending.gov for Fiscal Years 2019-2024, focusing on recipients in Virginia, California, and Maryland. The goal is to explore geographic distribution, agency funding flows, industry sector breakdowns, and temporal trends using various visualization techniques outlined in the planning document.

Team Roles and Responsibilities

Robel Endashaw: Data Acquisition and Initial Processing

- **Data Download:** Use USASpending.gov's Custom Award Data tool to download relevant contract data.
 - Filter by Award Type: Contracts.
 - Filter by Date Range: FY 2019-2024 (using 'action_date'). Handle potential multi-year downloads and merging.
 - Filter by Recipient Location: VA, CA, MD.
 - Output Format: CSV.
- **Initial Data Cleaning & Structuring:**
 - Load only necessary columns into the chosen analysis environment (). Refer to Table 1 in federal contracting pdf for essential fields.
 - Optimize data types for memory efficiency (e.g., categorical, smaller numeric types).
 - Assess data quality, particularly for location fields ('recipient_location_zip_code', 'recipient_location_county_code', 'place_of_performance_zip_code', etc.) required for geographic mapping. Document findings on usability.
 - Differentiate between recipient location and place of performance, preparing data based on the primary analysis choice ('recipient_location_state_code') while retaining the other for context.
- **Preliminary Aggregation:** Calculate basic aggregate metrics needed for other team members (e.g., total obligation per state/year/agency/NAICS).
- **Deliverable:** Cleaned, documented datasets ready for visualization tasks.

Jonathan Hale: Geographic Distribution and Agency Flow Visualizations

- **Visualization Strategy A (Geographic):**

- Implement county-level choropleth maps for VA, CA, MD showing total ‘federal_action_obligation’ or contract count. Requires obtaining relevant shapefiles/GeoJSON.
- Implement interactive bubble maps using recipient ZIP codes, with bubble size representing total contract value. Focus on interactivity (zoom, pan, tooltips) to manage density.
- Base implementation on data quality assessment provided by Robel.

- **Visualization Strategy B (Agency-to-State Flow):**

- Implement a Sankey Diagram visualizing ‘federal_action_obligation’ flow from top N awarding agencies to the three states (VA, CA, MD).
- Determine appropriate value for N (e.g., top 10-15 overall, or top 5-7 per state) to ensure clarity.

- **Tooling:** Collaborate on selecting and utilizing appropriate visualization libraries/tools (JavaScript: D3.js/).

- **Deliverable:** Code and resulting visualizations for Geographic Distribution and Agency Flows.

Daniel Berhane: Sector/Industry and Temporal Trend Visualizations

- **Visualization Strategy C (Sector/Industry - NAICS):**

- Aggregate contract data to the 2-digit NAICS sector level.
- Implement Faceted Grouped Bar Charts comparing total ‘federal_action_obligation’ for VA, CA, MD within the top N NAICS sectors.
- Optionally, implement supplementary 100% stacked bar charts (faceted by state) to show relative sector proportions within each state.

- **Visualization Strategy D (Temporal Trends):**

- Implement a multi-line chart showing total ‘federal_action_obligation’ over time (FY 2019-2024, using ‘action_date’ for aggregation) for VA, CA, and MD.
- Implement supplementary Faceted Stacked Area Charts (one per state) showing the composition of total obligation over time by top 3-5 awarding agencies or NAICS sectors.
- Consider adding annotations for significant events (e.g., COVID-19 spending) if DEFC data allows.

- **Tooling:** Collaborate on selecting and utilizing appropriate visualization libraries/tools.

- **Deliverable:** Code and resulting visualizations for Sector Breakdowns and Temporal Trends.

Benjamin Tanowitz: State Comparison, Integration, and Documentation

- **Visualization Strategy E (State Comparison):**

- Implement Grouped Bar Charts to compare key aggregate metrics (e.g., total obligation, average obligation, key sector totals) across VA, CA, MD.
- Implement Faceted Histograms or Box Plots to compare the distribution of individual ‘federal_action_obligation’ values across the three states.
- Create a summary table with key comparative numerical values.

- **Integration and Finalization:**

- Integrate all visualizations developed by the team into a cohesive final product (e.g., a report, a dashboard, a web page).

- Ensure consistent styling, labeling, and interactivity across all visualizations.
- Review visualizations for clarity, accuracy, and adherence to project goals.
- **Documentation and Reporting:**
 - Ensure code is clean, well-documented, and potentially publishable (, via GitHub).
 - Contextualize findings with external benchmarks where appropriate.
- **Tooling:** Coordinate final tool choices and integration strategy.
- **Deliverable:** Integrated final visualization product, final report, and documented codebase.

Collaboration Notes

- Regular team checkin via groupme to discuss progress, address challenges, and ensure consistency.
- A shared code repository (e.g., GitHub) will be used for version control and collaboration.
- The final selection of visualization tools (e.g., libraries, , D3.js) will be confirmed collectively based on project needs and team expertise.