Project 2 Report

Team: Amber Su, Curtix Xu, Chi Zhang

a) A description of the data. Report where you got the data.
Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. Editing is important! You are not required to use every part of the dataset.
Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.

We utilized two primary datasets for this project:

1. **Commercial Buildings Energy Consumption Survey (CBECS)**
2. **Residential Energy Consumption Survey (RECS)**

Both datasets are open-source and provided by the U.S. Energy Information Administration (EIA). Additionally, we used **U.S. Census Divisions** data sourced from the U.S. Census Bureau to provide geographic context for visualizations.

Variables and Data Selection

The original CBECS and RECS datasets include a large number of variables. To improve usability and focus, we selected and cleaned a subset of key variables:

- **Construction Year**: The year the building was constructed.
- **Building Area**: The total area of each building.
- **Census Division**: The geographic division where the building is located.
- **Normalized Energy Consumption**: Calculated values for gas, electricity, heating, and cooling consumption normalized by building area.

Data Preprocessing

To prepare the data for analysis and visualization:

1. **Variable Selection:**
    - The original files has been processed to keep only the relevant information mentioned above
2. **Variable Alignment**:
    - Construction years and census divisions from both datasets were standardized into shared scale buckets to ensure compatibility.
3. **Outlier Removal**:

    - Buildings with abnormally high energy consumption values were identified as outliers and excluded to prevent skewed results.
4. **Shapefile Conversion**:
    - The U.S. Census Division map data, originally in shapefile format, was preprocessed using Python to convert it into TopoJSON format for compatibility with D3.js visualizations.

Integration of Data

The CBECS and RECS datasets were combined based on matching census divisions and construction year buckets, ensuring consistent geographic and temporal alignment. The TopoJSON map data was then used to provide spatial visualization for energy consumption patterns.

By focusing on this refined subset of variables and applying preprocessing steps, we aimed to balance comprehensiveness and efficiency, ensuring the dataset is clean, uniform, and ready for meaningful visualizations.

b) An overview of your visual design rationale. A good rule of thumb to follow is "every pixel must be justified." Instead of a 100,000-element breakdown, give us an overview of the design decisions you made and the trade-offs inherent in how you displayed the data. This part ought to include a description of the mapping from data to visual elements. Describe marks and channels you employ such as position, color, or shape. Mention any transformations you performed, such as log scales.

Each set of interactive visualizations for residential/commercial energy consumption includes a map of energy consumption based on geographic division, a bar chart of building distribution based on energy consumption, and a scatter plot showing the relationship of two user-defined inputs.
They are arranged side by side sharing the same canvas height so that every visualization shares the same visual hierarchy. User-input panel, which controls the update of graphs based on change in y-axis, x-axis, "Area Range" and "Year Range", is located on the left side. Its height is set to align with other graphs, ensuring a seamless viewing experience without scrolling while observing updates to the graphs.

Marks and channels:
Map of Commercial/Residential Buildings by Energy Consumption:
Design justification: This map provides a clear overview of energy consumption, comparing residential and commercial buildings across different geographic locations.

- Mark: U.S. division shape

- Visual Channels: gradient from light to dark color, color hue, geographically aligned division shapes

Distribution of Commercial Buildings by Energy Consumption
Design justification: The bar chart is used to represent the distribution of buildings by energy consumption due to its simplicity and ease of comparison across discrete ranges based on user inputs

- Mark: rectangles
- Visual Channels: horizontal and vertical aligned position, color hue(changes with user-input updates)

Scatter Plot of Commercial Buildings by Energy Consumption
Design justification: scatterplot is very efficient at displaying patterns among data, which is great for user-driven exploration, especially when the user would like to find the relationships between two inputs. Therefore, we give users the flexibility to define both the x and y inputs, so they could explore the data relationships in the dynamic visualizations.

- Marks: circles
- Visual Channels: horizontal and vertical aligned position, color hue(changes with user-input updates)

c) An overview of your interactive elements and their design rationale. Give us an outline of the design decisions that went into the interaction affordances you added to your visualization. What process did you use to choose the interactions you developed? How did you make them discoverable, usable, and interesting?

The interactive interface is designed to enable users to explore energy consumption data through multiple interaction points, both inside and outside the graphs.

We incorporated two primary methods for user interaction:

1. **Toolbar Controls**
   Users can control the visualization through interface elements like radio buttons, dropdown menus, and sliders located in the toolbar. These allow users to customize the x- and y-axes of the graphs based on their preferences. Additionally, sliders enable users to filter the dataset by specifying a range for parameters such as building area or year built. These features make it easier for users to focus on subsets of the data that are most relevant to their interests.
2. **Graph Interactions**
   Users can interact directly with the graphs for a more immersive experience. Clicking on the map allows users to select a specific region of the United States, which then filters the dataset displayed in both the histogram and scatter plot. This graph-based interaction

complements the toolbar controls by providing a visual and intuitive way to refine data exploration. Like the sliders, this feature helps users efficiently navigate and analyze large datasets. Additionally, when hovering over specific elements, detailed information will be displayed to allow users to inspect the actual number and other relevant information.

These interaction methods were chosen to balance usability and discoverability. Toolbar controls are straightforward and familiar, offering clear functionality, while direct graph interactions make the experience engaging and visually intuitive. Together, these design choices empower users to explore the data efficiently and effectively while keeping the process interesting and user-friendly.

d) The story. What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

This visualization is based on the Commercial Buildings Energy Consumption Survey (CBECS) and Residential Energy Consumption Survey (RECS) data. It analyzes energy usage patterns in residential and commercial buildings across construction years, building areas, and regions in the United States, offering insights into gas, electricity, heating, and cooling energy consumption.
To explore the implications of building characteristics—such as usage type, climate region, construction year, and building area—this project presents a series of visualizations and interactive filters.
It is surprising to observe that, while residential buildings consume less energy on gas, electricity, and cooling compared to commercial buildings, their energy consumption for heating is remarkably similar.
Key insights reveal that smaller buildings tend to have lower energy consumption per square meter, while newer buildings show higher electricity consumption, likely driven by the increased use of electronic devices and modern systems. Additionally, heating and cooling demands vary significantly across regions, reflecting the influence of climate, whereas average gas and electricity consumption exhibit relatively minor regional differences.

3. At the end of your PDF file, include an outline of team contributions to the project. Identify how work was broken down in the group and explain each group member's contributions to the project. Give a rough breakdown of how much time you spent developing and which parts of the project took the most time.

Amber Su:

1. Design and implement the basic interaction logic
2. In charge of the distribution plot
3. Data preprocessing and cleaning

Curtis Xu:

1. In charge of the Scatter Plot design and Implementation
2. Design and implement the website layout
3. Contribute to the graph interaction and connection functionality

Chi Zhang:

1. Made Mapping Commercial Buildings by Energy Consumption