INFO 3300 Project 2 Final Report

Group Members:
Xinyi Zhou (xz255) | Warren Hua (wsh48) | Minghan Gao (mg2328)

---

1. Data Description

This project utilizes multiple datasets to explore regional economic and demographic trends in China. The data includes geographical boundaries (china.json), per capita consumption expenditure (consumption-expenditure.csv), per capita disposable income (disposable-income.csv), gross regional product (GRP.csv), local government revenue (local-government-revenue.csv), and population (population.csv), spanning various years and provinces.

china.json

- The topoJson file for the Chinese province is provided by Matt Bertrand on his github.
- Data source:https://gist.github.com/mbertrand/5530456c2816c7cad94b

consumption-expenditure.csv

- This dataset presents the average amount of money spent per person in different years on goods and services within different regions in China.
- Data source: https://data.stats.gov.cn/english/easyquery.htm?cn=C01
- Variables:
    - Region: Chinese Provinces
    - Year: columns from 2005 to 2023 represents the years
    - Per Capita Consumption Expenditure:The per capita annual consumption expenditure for the respective region and year.
- Reformat:
    - Change the Per Capita Consumption Expenditure from string to int format.

disposable-income.csv

- This dataset presents the average disposable income people have in different years within different regions in China.
- Data source: https://data.stats.gov.cn/english/index.htm
- Variables:
    - Region: Chinese Provinces
    - Year: Columns from 2005 to 2023 represents the years
    - Per Capita Disposable Income:Per capita disposable income for each region and year.
- Reformat:
    - Change the Per capita disposable income from string to int format.

GRP.csv

- This dataset presents the gross regional product in different years within different regions in China.
- Data source: https://data.stats.gov.cn/english/index.htm
- Variables:
    - Region: Chinese Provinces
    - Year: Columns from 2004 to 2023 represents the years
    - GRP:The economic productivity of each region annually.

- Reformat:
    - Change the GRP from string to int format.

local-government-revenue.csv

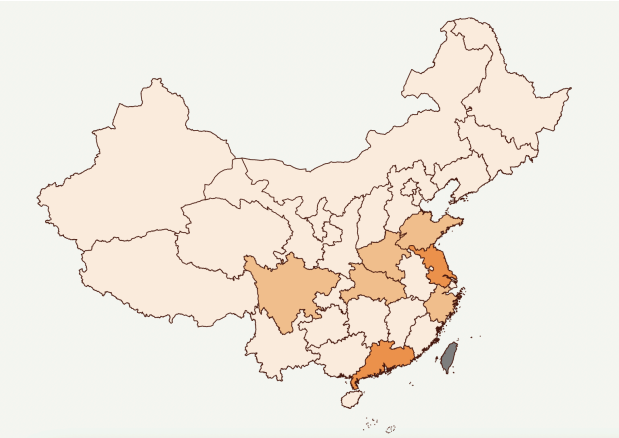
- This dataset presents the local government revenue in different years within different regions in China.
- Data source: https://data.stats.gov.cn/english/index.htm
- Variables:
    - Region: Chinese Provinces
    - Year: Columns from 2004 to 2023 represents the years
    - Local government revenue: Local government revenue of each region annually.
- Reformat:
    - Change the local government revenue from string to int format.

population.csv

- This dataset presents the population in different years within different regions in China.
- Data source: https://population.un.org/wpp/
- Variables:
    - Region: First column presents Chinese Provinces
    - Year: Other Columns represents the years
    - Population: Population size of each region annually.
- Reformat:
    - Change the population size from string type to int format.

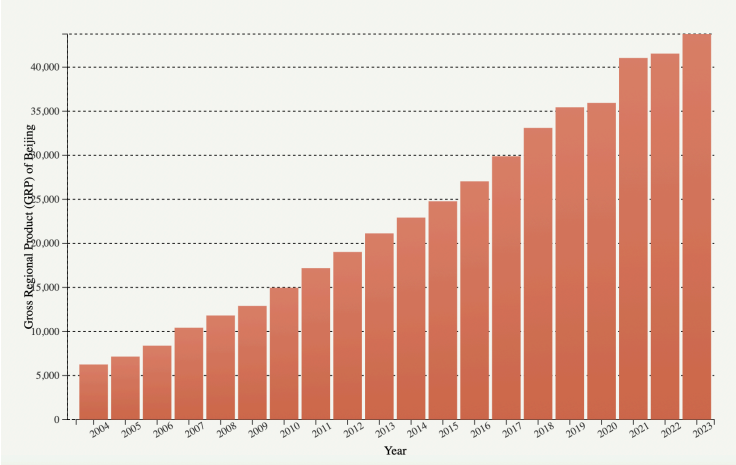
2. Overview of Visual Design Rationale

- **Mapping Data to Visual Elements**
    - **Choropleth Map of China:**
        - Data: Provincial-level economic metrics such as Gross Regional Product (GRP), Population, and Disposable Income, mapped to geographical regions.
        - Marks: Each province is represented as a geographic shape on the map.
        - Channels:
            - Color: A sequential color scale represents the magnitude of the selected data metric. Higher values are mapped to more saturated shades, while lower values are less saturated shades. We used a quantile-based color scale to ensure the effective representation of data distribution across all provinces.
            - Position: The geographic layout ensures users can easily associate data with real-world locations.

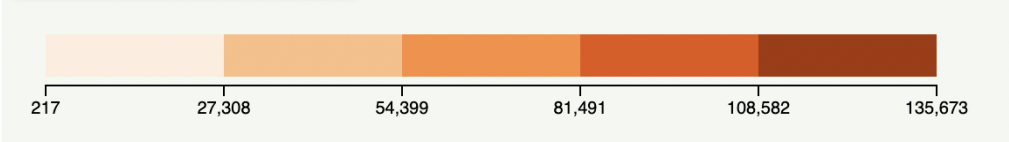


- Justification: The choropleth design allows users to quickly compare the economic performance of provinces spatially, leveraging the user's familiarity with geographic layouts.

- **Histogram for Temporal Data:**
  - Data: Yearly values of the selected metric for the chosen province.
  - Marks: Vertical bars represent data values for each year.
  - Channels:
    - Height: Encodes the magnitude of the data value (e.g., GRP) for each year.
    - Color: A gradient color fill was used for aesthetic appeal and to align with the project's theme.
    - Position: Bars are positioned sequentially along the x-axis to represent time.
  - Justification: The histogram provides a clear view of temporal trends, helping users identify changes in economic performance over the years.

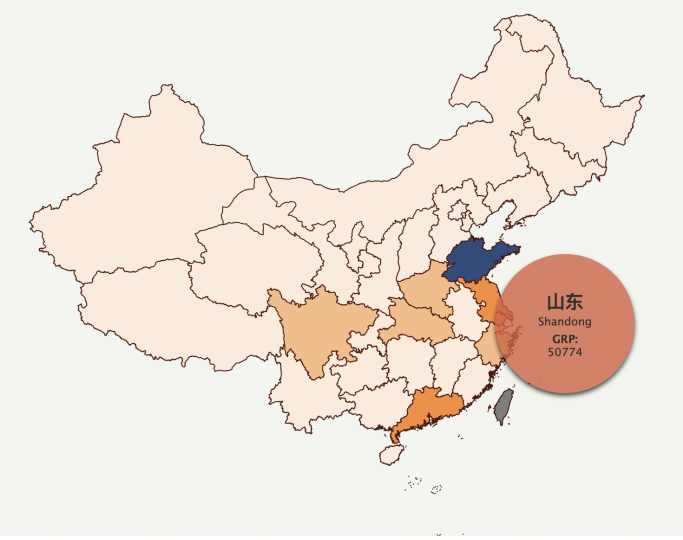


- **Legend and Tooltip:**
  - Legend:
    - Explains the mapping of data values to colors, using discrete steps.
    - Positioned right of the map for easy reference without cluttering the main visualization.



  - Tooltip:
    - Displays precise values for the selected province and metric when users hover over it.

- Includes both the province's name and the data for the chosen topic and year.



- Justification: These elements enhance accessibility by providing detailed data without overwhelming the main visualization.

- **Design Decisions and Trade-Offs**
  - Sequential Color Scales:
    We opted for a perceptually uniform sequential color scale to represent data magnitudes.
  - **Trade-Off**: While effective for representing magnitude, the color scale may obscure small differences in data values when the range is wide.

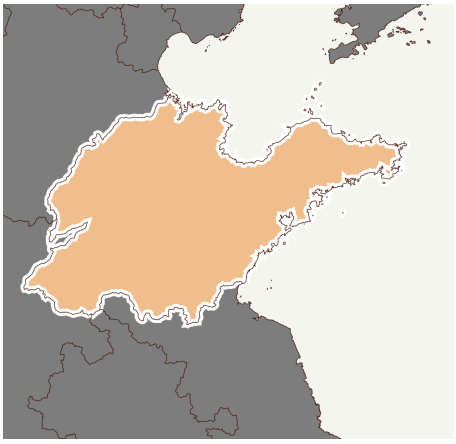
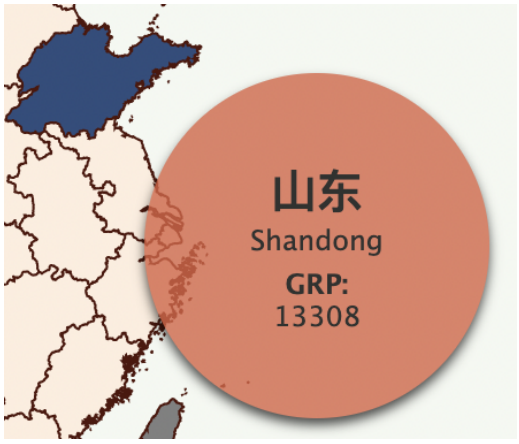
- **Logarithmic vs. Linear Scales**
  - We chose linear scales for simplicity, given the audience's likely familiarity with this representation.
  - **Trade-Off**: Linear scales can obscure smaller variations in data values compared to logarithmic scales.

- **Axis and Label Design**
  - X and Y axes for the histogram are clearly labeled with a rotated x-axis to avoid label overlap.
  - Y-axis values are formatted with commas for readability.
  - **Trade-Off**: Rotated x-axis labels may be slightly harder to read but are necessary to prevent overlap for long time-series data.

3. Overview of Interactive Elements & Design Rational

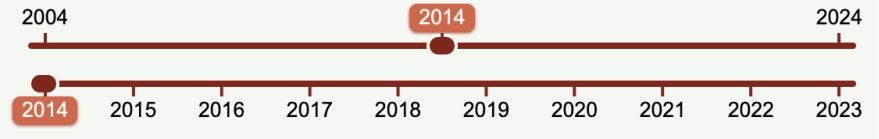
- **Choropleth Map Interaction**
  - **Interactions**:
    - Clicking on a province highlights it, grays out all other provinces, and updates the histogram for the selected province. At the same time, all animation effects pause to minimize distractions when users focus on a specific province. The animations resume once users zoom in or out, signaling their intent to continue exploring other regions.
    - Hovering over a province displays a tooltip with detailed information including name of the region, data topic, and data value.
    - The map supports zooming and panning for better exploration of individual regions.

- **Design Rationale**:
  - The click interaction focuses user attention on one province, allowing for detailed exploration without distractions.
  - The hover tooltip provides immediate feedback, making the data accessible without additional clicks.
  - Zoom and pan interactions enable detailed inspection of smaller regions.
- **Process**:
  - We implemented these interactions to balance focus and exploration. Clicking simplifies the display by graying out unselected provinces, while hovering provides quick, discoverable details

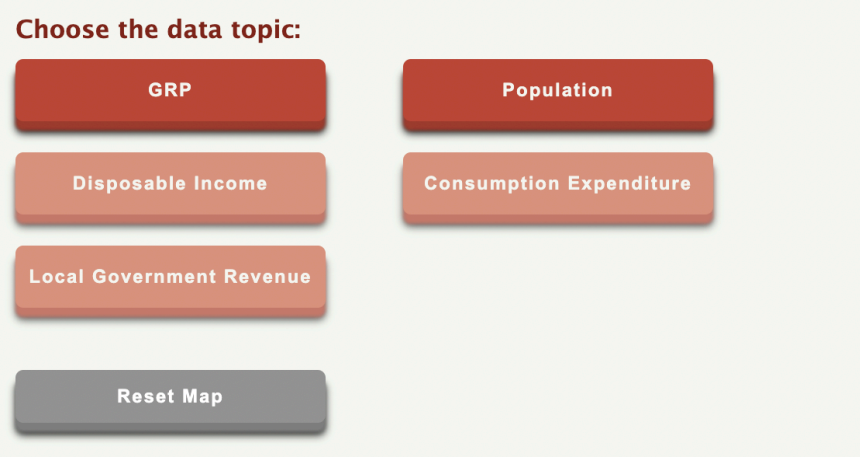



- **Discoverability**:
  - When a specific region is selected, the animation pulse will be accompanied by corresponding text, guiding users on how to resume the animation after deselecting the region.
  - The hover tooltip is intuitive as it appears immediately when the mouse is over a province.
  - Clicking and zooming are discoverable through visible changes in the map and accompanying histogram.

- **Topic and Year Selection**
  - **Interactions**:
    - Users can select a year by clicking on the year button or by scrolling the range bar. Both methods are synchronized and will change the map.



        - Topic buttons allow users to switch between economic topics (e.g., GRP, Population) and years, dynamically updating the map and histogram.



    - **Design Rationale**:
        - Topic buttons simplify navigation between datasets, ensuring users can focus on one dimension at a time.
        - The year slider and buttons provide flexibility, allowing users to explore both specific years and broader ranges.
    - **Process:**
        - We prioritized simplicity for topic selection by using large, clearly labeled buttons.
    - **Discoverability**:
        - Buttons are labeled and visually distinct, making their function clear.
        - The year slider is accompanied by numeric labels, ensuring users understand its purpose.

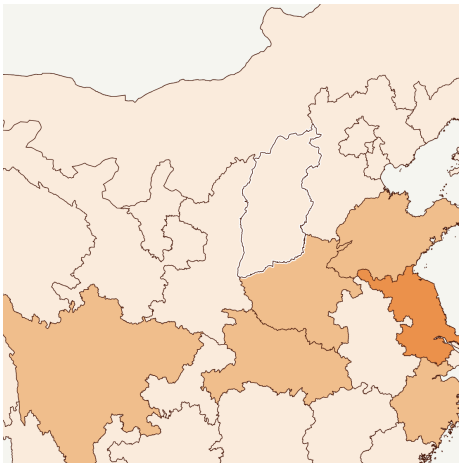
- **Histogram Interaction**
    - **Interactions**:
        - The histogram dynamically updates to show yearly data for selected province and topic.
    - **Design Rationale**:
        - The histogram complements the choropleth map by providing a detailed temporal view of the data.
    - **Process**:
        - We chose the bar chart for its effectiveness in communicating time-series data.
    - **Discoverability**:
        - The histogram automatically updates when a province is clicked, creating a direct connection between map interaction and the temporal data.

- **Reset Button**
    - **Interactions**:
        - Clicking the reset button resets the map to its default zoom level and restores the colors of all provinces.

    - **Design Rationale**:
        - This feature ensures that users can easily return to the default view after exploring specific regions.
    - **Process**:

        - We added this interaction to improve usability, especially for users unfamiliar with zoom and pan controls.
    - **Discoverability**:
        - The button is labeled "Reset Map" and positioned prominently next to the map controls.

- **Zoom and Pan**
    - **Interactions**:
        - Users can zoom in and out using the mouse wheel and pan across the map by dragging.
    - **Design Rationale**:
        - These interactions allow users to explore smaller provinces or specific regions in greater detail.
    - **Process**:
        - We included zoom and pan to complement the map's geographic data, enabling closer inspection while maintaining spatial context.
    - **Discoverability**:
        - Zoom and pan are standard interactions for map visualizations, making them intuitive for most users.



- **Discoverability, Usability, and Interest**
    1. **Discoverability**:
        - Interactions like hovering, clicking, and zooming are standard and intuitive.
        - Buttons and tooltips are visually distinct, making their purpose immediately clear.
    2. **Usability**:
        - Consistent interaction patterns (e.g., clicking for details, hovering for tooltips) ensure ease of use.
        - The reset button offers a safety net for users exploring unfamiliar regions.
    3. **Interest**:
        - Interactive elements encourage exploration, allowing users to uncover insights through dynamic updates and personalized exploration paths.

4. The Story

- **Key Narratives**
    1. **Economic Disparities Across Provinces**:
        - **Observation**: The visualization reveals significant disparities in economic performance across provinces. For instance, coastal provinces like Guangdong and Jiangsu consistently outperform inland provinces like Gansu and Qinghai in terms of GRP and Disposable Income.
        - **Insight**: This highlights the uneven distribution of economic development in China, driven by factors such as industrialization, foreign investment, and geographic advantages.
    2. **Economic Growth**:
        - **Observation**: The visualization reveals the majority of the region face economic growth throughout the years, with increasing GRP, disposable income, consumption expenditure, and local government revenue.

    - **Insight**: This highlights China is facing steady economic progress, people's living standard are becoming better and better.
3. **Population and Resource Utilization**:
    - **Observation**: Population data shows that provinces with larger populations (e.g., Henan, Sichuan) don't always align with the highest GRP or Disposable Income values.
    - **Insight**: This indicates that resource utilization and industrial development are more critical than sheer population size in driving economic performance.
4. **Government Revenue Patterns**:
    - **Observation**: Local Government Revenue often correlates strongly with GRP, but there are outliers where revenue collection is unusually high or low relative to the province's economic size.
    - **Insight**: This may reflect varying taxation policies, economic structure (e.g., reliance on state-owned enterprises), or administrative efficiency.

- **Surprising Findings**
    1. **Persistent Underperformance in Certain Provinces**:
        - Provinces like Tibet and Qinghai consistently rank at the bottom for economic indicators, even as the national economy grows. This stark contrast highlights long-standing challenges in developing remote and resource-scarce regions.
    2. **Impact of Major Cities**:
        - Municipalities like Shanghai and Beijing show extraordinary economic performance despite their small geographic size, emphasizing the outsized role of urban centers in China's economy.

- **What We Want Viewers to Take Away**
    1. **Regional Complexity**: China's economic landscape is incredibly diverse, with each province telling its own story of development, challenges, and opportunities.
    2. **Growth and Inequality**: While China's overall economy has grown exponentially, this growth is not evenly distributed, and understanding this inequality is crucial for crafting future policies.
    3. **Interconnected Factors**: Economic performance is influenced by a web of factors, including geography, population, and government policies. Our visualization provides a tool to explore these interconnections in depth.

5. Outline of Team Contributions

- **Broken down of work**
    - Xinyi Zhou: responsible for the first part: Filter labels to display population, GDP, and consumption data in geographic maps. Also responsible for writing the corresponding section of the final report.
    - Minghan Gao: responsible for the second part: implementing the emphasis effect when hovering over, and the displaying of corresponding data for specific provinces.
    - Warren Hua: responsible for the navigation: including interaction between the user and data visualization. Also responsible for writing the corresponding section of the final report.

- **Time spent & most time consuming part**
    - Xinyi Zhou: Spent roughly 12h to implement the map, implement the interaction of the year and theme buttons, the two range bars, the mouseover effect, and the styling of the pages and buttons.

- Minghan Gao: Spent roughly 12h on implementing color interactions, zoom in/out, as well as how different part of function coordinate with each other
- Warren Hua: Spent roughly 12h on implementing color legend, fixing bugs as well as writing final reports.