Effective Use of AI Tools in Map-Making Workflows for Visualizing Epidemiological Data

Tre Hellwig[1]; Melanie A. Sutton[1], PhD; Teresa Lynne Caples[1], MS, D.Bioethics, CIP; Anthony Okafor[2], PhD; Justice Mbizo[1], DrPH
[1]Department of Public Health, Usha Kundu, MD College of Health, University of West Florida; [2]Hal Marcus College of Science and Engineering, University of West Florida

UNIVERSITY of WEST FLORIDA

ABSTRACT

BACKGROUND:
In 2024, Google rolled out AI Overviews, placing an AI chatbot in front of every user during Google searches. Given the public's awareness of these AI tools, public health professionals must incorporate them into their data visualization workflows to create effective visualizations for public-facing communications.

OBJECTIVE:
- Compare methods for effective priming of AI tools to generate improvement suggestions for created maps.
- Demonstrate tips and tricks to attain the full potential of AI tools in the map-making process for epidemiological data.

METHODS:
The Department of Public Health formed a multidisciplinary faculty/student team to explore the capabilities of a suite of AI tools in designing and evaluating epidemiological maps derived from data coded at country, state, city, and census tract-level levels. Representative datasets included variables for crime, disease occurrence, and medical testing facility locations.

RESULTS:
Strategies for effectively using free versions of chatbots like Chat GPT 3.5/4 and Google Gemini 1.0/1.5 Flash, were developed based on several map types created by faculty and students using Google Geo Charts or Excel 3D Maps. Tips for priming AI tools in the process of designing and interpreting public-facing epidemiological maps to ensure equity and accessibility were created, alongside guidance on recognizing AI hallucinations.

CONCLUSION:
Epidemiological data is commonly visualized with publicly distributed maps, and utilizing AI tools in the map-making process will ensure thorough, well-designed displays. Incorporating AI tools into data visualization workflows provides students, faculty, and public health professionals with additional foundational knowledge to ensure that generated maps contribute to building trust in public health.

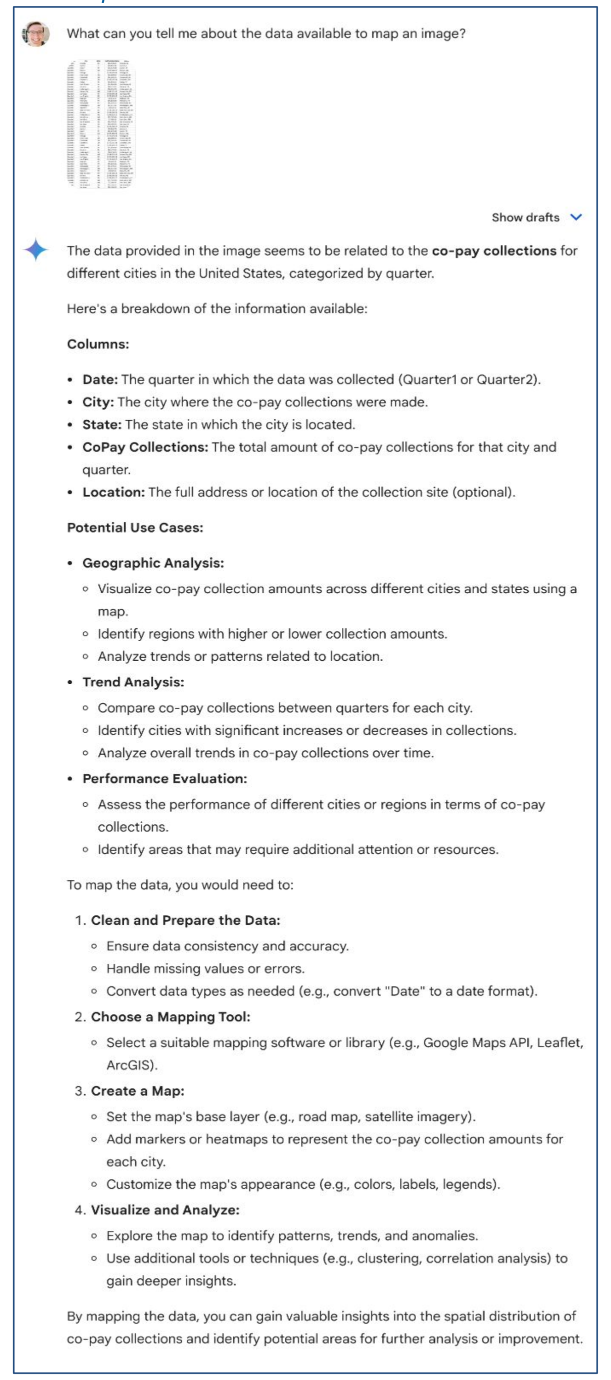
KEY POINTS WHEN MAP-MAKING WITH AI CHATBOTS
- Identify desired output (AI-facilitated learning versus AI-augmented workflow)
- Recognize relevant AI hallucinations in map-making
- Leverage AI chatbots to enhance and develop critical thinking in map design
- Expand learning outcomes through reiteration/exploration with AI chatbots
- Approach a solution to Bloom's 2 Sigma Problem (AI's ability to act as a tutor)

BEST PRACTICES

ENGINEERING PROMPTS FOR PRECISE OUTPUTS
To begin, what is your desired output and outcome? We recommend using AI chatbots for brainstorming to generate ideas, for expanding understanding through exploration, or as a "new set of eyes" to provide the final map adjustments.

EXPLORING DATASETS
- Upload data files
- Provide a specific task for the chatbot to complete
- Identify variables and potential uses
- Start considering map types

Image 1
Data Exploration

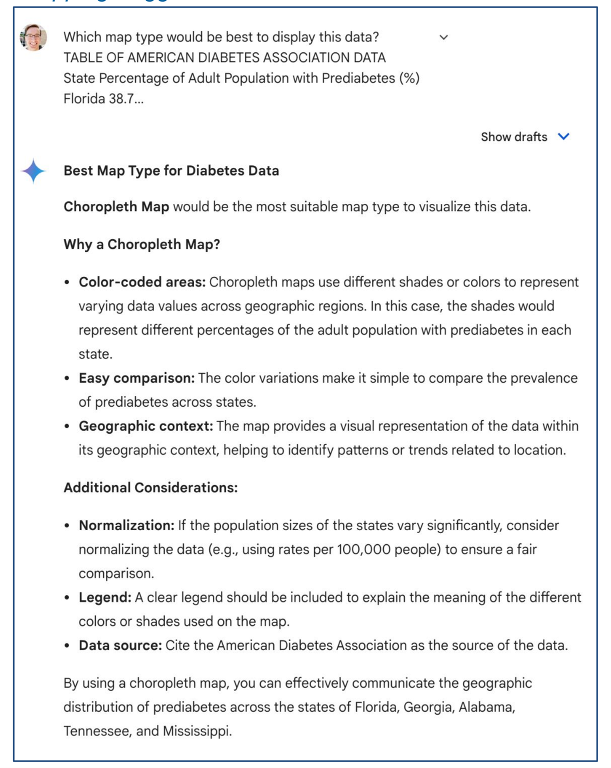

What can you tell me about the data available to map an image?
Show drafts
The data provided in the image seems to be related to the **co-pay collections** for different cities in the United States, categorized by quarter.
Here's a breakdown of the information available:
Columns:
Date: The quarter in which the data was collected (Quarter1 or Quarter2).
City: The city where the co-pay collections were made.
State: The state in which the city is located.
CoPay Collections: The total amount of co-pay collections for that city and quarter.
Location: The full address or location of the collection site (optional).
Potential Use Cases:
Geographic Analysis:
Visualize co-pay collection amounts across different cities and states using a map.
Identify regions with higher or lower collection amounts.
Analyze trends or patterns related to location.
Trend Analysis:
Compare co-pay collections between quarters for each city.
Identify cities with significant increases or decreases in collections.
Analyze overall trends in co-pay collections over time.
Performance Evaluation:
Assess the performance of different cities or regions in terms of co-pay collections.
Identify areas that may require additional attention or resources.
To map the data, you would need to:
1. Clean and Prepare the Data:
Ensure data consistency and accuracy.
Handle missing values or errors.
Convert data types as needed (e.g., convert "Date" to a date format).
2. Choose a Mapping Tool:
Select a suitable mapping software or library (e.g., Google Maps API, Leaflet, ArcGIS).
3. Create a Map:
Set the map's base layer (e.g., road map, satellite imagery).
Add markers or heatmaps to represent the co-pay collection amounts for each city.
Customize the map's appearance (e.g., colors, labels, legends).
4. Visualize and Analyze:
Explore the map to identify patterns, trends, and anomalies.
Use additional tools or techniques (e.g., clustering, correlation analysis) to gain deeper insights.
By mapping the data, you can gain valuable insights into the spatial distribution of co-pay collections and identify potential areas for further analysis or improvement.

FOUNDATIONAL STUDENT OUTCOMES
- Communicate the strengths and weaknesses of using an AI tool to complete an assignment
- Identify AI hallucinations and ways to confirm the output's correctness
- Utilize critical thinking when designing a prompt to generate the desired output
- Describe their target audience and the accessibility needs
- Identify gaps in their knowledge to target areas of improvement
- Reiterate key ideas and elaborate to strengthen their understanding
- Recognize biases in the outputs

WEIGHING MAPPING OPTIONS
- Explore common options for a mapping display
- Review strengths and weaknesses of available options
- Factor considerations to reach your target audience
- Prepare for map creation

QR Code to Project Folder in Google Drive

Image 2
Mapping Suggestions


Which map type would be best to display this data?
TABLE OF AMERICAN DIABETES ASSOCIATION DATA
State Percentage of Adult Population with Prediabetes (%)
Florida 38.7...
Show drafts
Best Map Type for Diabetes Data
Choropleth Map would be the most suitable map type to visualize this data.
Why a Choropleth Map?
Color-coded areas: Choropleth maps use different shades or colors to represent varying data values across geographic regions. In this case, the shades would represent different percentages of the adult population with prediabetes in each state.
Easy comparison: The color variations make it simple to compare the prevalence of prediabetes across states.
Geographic context: The map provides a visual representation of the data within its geographic context, helping to identify patterns or trends related to location.
Additional Considerations:
Normalization: If the population sizes of the states vary significantly, consider normalizing the data (e.g., using rates per 100,000 people) to ensure a fair comparison.
Legend: A clear legend should be included to explain the meaning of the different colors or shades used on the map.
Data source: Cite the American Diabetes Association as the source of the data.
By using a choropleth map, you can effectively communicate the geographic distribution of prediabetes across the states of Florida, Georgia, Alabama, Tennessee, and Mississippi.

THE TEACHER IS NOT THE ADVERSARY

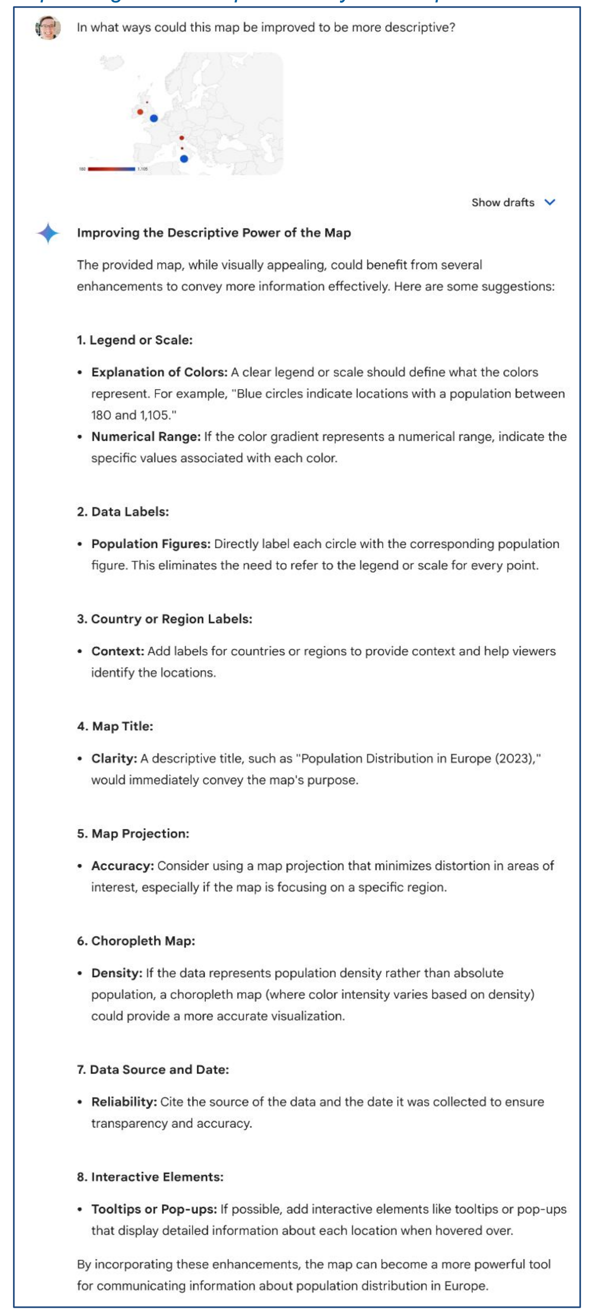
MOVING PASSED THE RUBRIC
We've seen our students adhere to the grading rubric and show little interest in expanding their skill level. After adding an exploration step using an AI chatbot, the students' submissions improved in ways that exceeded our expectations.

ELABORATION IN ACTION
- Use the output to expand topic areas
- Explore alternate methods of labeling and coloring
- Complement the map with tables and text

BEWARE OF HALLUCINATIONS
ChatGPT defines and describes AI hallucinations as:

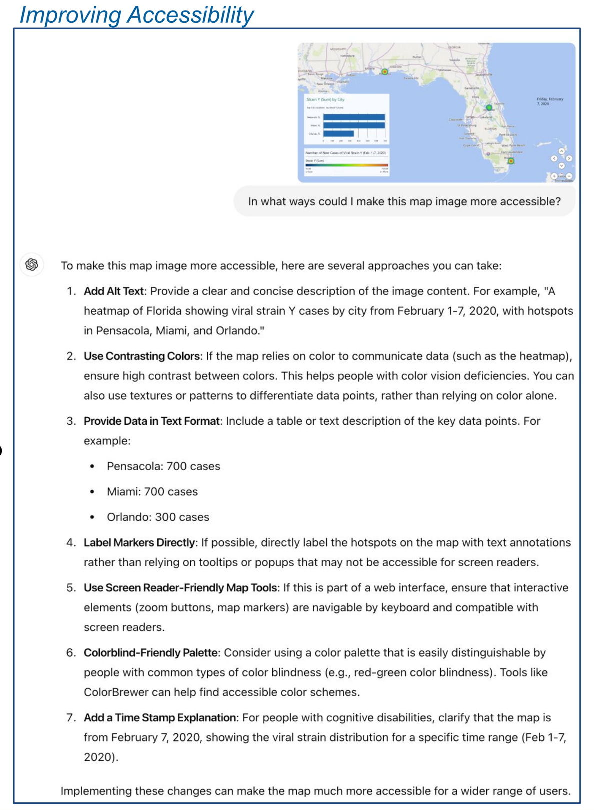
> "Artificial hallucination refers to the phenomenon of a machine, such as a chatbot, generating seemingly realistic sensory experiences that do not correspond to any real-world input. This can include visual, auditory, or other types of hallucinations. Artificial hallucination is not common in chatbots, as they are typically designed to respond based on pre-programmed rules and data sets rather than generating new information. However, there have been instances where advanced AI systems, such as generative models, have been found to produce hallucinations, particularly when trained on large amounts of unsupervised data" (Alkaissi & McFarlane, 2023).

Image 3
Expanding the Descriptive Ability of a Map


In what ways could this map be improved to be more descriptive?
Show drafts
Improving the Descriptive Power of the Map
The provided map, while visually appealing, could benefit from several enhancements to convey more information effectively. Here are some suggestions:
1. Legend or Scale:
Explanation of Colors: A clear legend or scale should define what the colors represent. For example, "Blue circles indicate locations with a population between 180 and 1,105."
Numerical Range: If the color gradient represents a numerical range, indicate the specific values associated with each color.
2. Data Labels:
Population Figures: Directly label each circle with the corresponding population figure. This eliminates the need to refer to the legend or scale for every point.
3. Country or Region Labels:
Context: Add labels for countries or regions to provide context and help viewers identify the locations.
4. Map Title:
Clarity: A descriptive title, such as "Population Distribution in Europe (2023)," would immediately convey the map's purpose.
5. Map Projection:
Accuracy: Consider using a map projection that minimizes distortion in areas of interest, especially if the map is focusing on a specific region.
6. Choropleth Map:
Density: If the data represents population density rather than absolute population, a choropleth map (where color intensity varies based on density) could provide a more accurate visualization.
7. Data Source and Date:
Reliability: Cite the source of the data and the date it was collected to ensure transparency and accuracy.
8. Interactive Elements:
Tooltips or Pop-ups: If possible, add interactive elements like tooltips or pop-ups that display detailed information about each location when hovered over.
By incorporating these enhancements, the map can become a more powerful tool for communicating information about population distribution in Europe.

ENHANCING CRITICAL THINKING
- What was my initial mapping goal?
- What does my map currently say?
- Does my map tell a story?
- Does my map have the necessary labeling?
- Am I using the appropriate basemap?
- Is my map accessible to my target audience?
- Are there groups that will be unable to see my map?

Image 4
Improving Accessibility


In what ways could I make this map image more accessible?
To make this map image more accessible, here are several approaches you can take:
1. Add Alt Text: Provide a clear and concise description of the image content. For example, "A heatmap of Florida showing viral strain Y cases by city from February 1-7, 2020, with hotspots in Pensacola, Miami, and Orlando."
2. Use Contrasting Colors: If the map relies on color to communicate data (such as the heatmap), ensure high contrast between colors. This helps people with color vision deficiencies. You can also use textures or patterns to differentiate data points, rather than relying on color alone.
3. Provide Data in Text Format: Include a table or text description of the key data points. For example:
Pensacola: 700 cases
Miami: 700 cases
Orlando: 300 cases
4. Label Markers Directly: If possible, directly label the hotspots on the map with text annotations rather than relying on tooltips or popups that may not be accessible for screen readers.
5. Use Screen Reader-Friendly Map Tools: If this is part of a web interface, ensure that interactive elements (zoom buttons, map markers) are navigable by keyboard and compatible with screen readers.
6. Colorblind-Friendly Palette: Consider using a color palette that is easily distinguishable by people with common types of color blindness (e.g., red-green color blindness). Tools like ColorBrewer can help find accessible color schemes.
7. Add a Time Stamp Explanation: For people with cognitive disabilities, clarify that the map is from February 7, 2020, showing the viral strain distribution for a specific time range (Feb 1-7, 2020).
Implementing these changes can make the map much more accessible for a wider range of users.

REFERENCES
Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus, 15(2), e35179. https://doi.org/10.7759/cureus.35179
Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher, 13(6), 4–16. https://doi.org/10.2307/1175554
Gao, A. (2023) Prompt engineering for large language models. SSRN Electr J. https://doi.org/10.2139/ssrn.4504303

AI CHATBOTS RAISE THE CEILING ON PERFORMANCE

STUDENT-LED EXPLORATION AND IMPROVEMENT
To enhance student learning and help students develop skills for the future, faculty are encouraged to allow students the space to explore and discover. Integrating these AI tools into the coursework, we have added a final preparatory step to help the students reflect, elaborate, validate, and re-affirm their knowledge.

By uploading their maps into an AI chatbot and receiving feedback from their custom prompts, the students have one last opportunity to consider the assignment rubric and ways to take their submission to the next level. This step raised the ceiling on student performance, surpassing submissions from prior semesters.

Image 5
Student Submission: Labeling

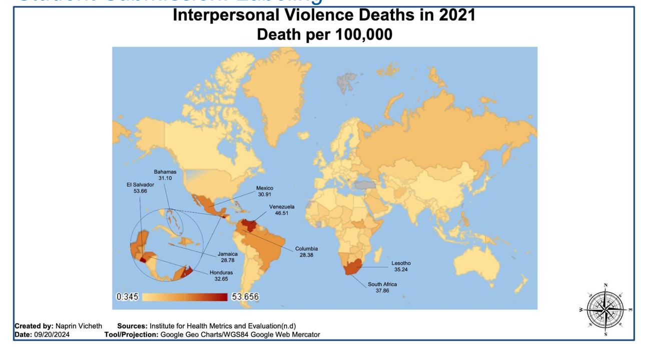
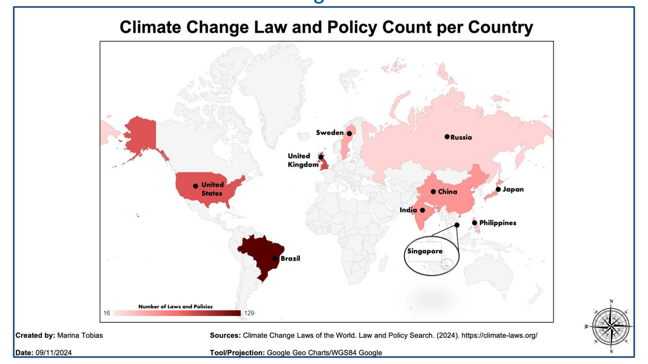

Interpersonal Violence Deaths in 2021
Death per 100,000
0.245
53.656
Created by: Naprin Vicheth
Date: 09/20/2024
Sources: Institute for Health Metrics and Evaluation(n.d)
Tool/Projection: Google Geo Charts/WGS84 Google Web Mercator

Image 6
Student Submission: Labeling and Color Gradient


Climate Change Law and Policy Count per Country
Sweden
Russia
United Kingdom
United States
China
Japan
India
Philippines
Singapore
Brazil
Number of Laws and Policies
Created by: Marina Tobias
Date: 09/11/2024
Sources: Climate Change Laws of the World. Law and Policy Search. (2024). https://climate-laws.org/
Tool/Projection: Google Geo Charts/WGS84 Google

STUDENT COMMENTS:
- "The AI output brought to my attention that countries that are not included or visible in the choropleth map need to be stated. The AI advice given also gave me confirmation that displaying numeric values is appropriate and helpful in these types of charts; I was initially unsure if it was allowed to detail the chart further. However, this also brings light to how little detail is in these maps by default."
- "I made adjustments to include street labels, with a satellite view, and a red-yellow color palette all suggested by Gemini."

Image 7
Student Submission: Base Map and Accessibility

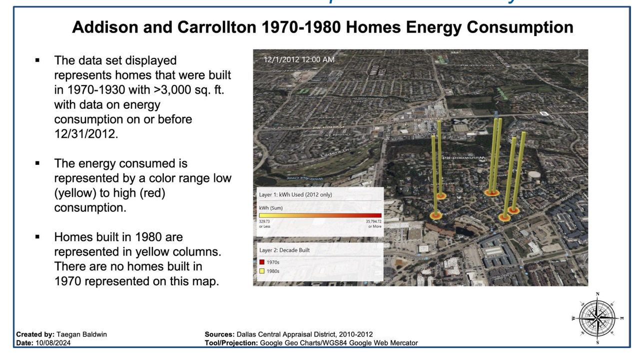

Addison and Carrollton 1970-1980 Homes Energy Consumption
The data set displayed represents homes that were built in 1970-1930 with >3,000 sq. ft. with data on energy consumption on or before 12/31/2012.
The energy consumed is represented by a color range low (yellow) to high (red) consumption.
Homes built in 1980 are represented in yellow columns. There are no homes built in 1970 represented on this map.
Created by: Taegan Baldwin
Date: 10/08/2024
Sources: Dallas Central Appraisal District, 2010-2012
Tool/Projection: Google Geo Charts/WGS84 Google Web Mercator

Image 8
Considering the Target Audience

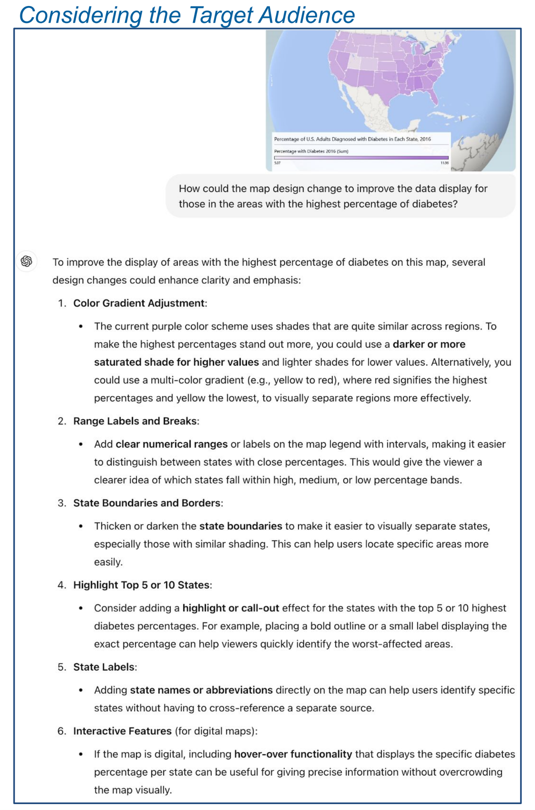
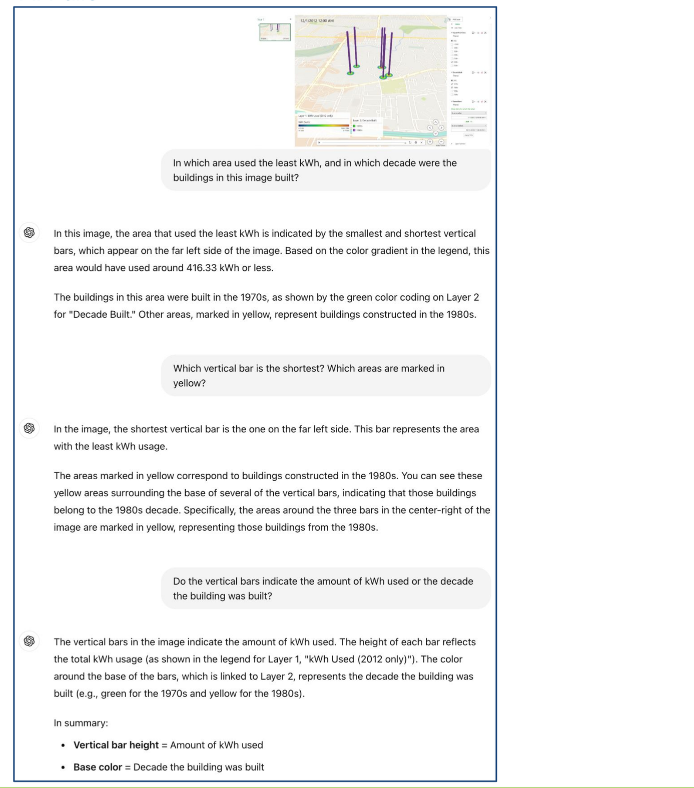

How could the map design change to improve the data display for those in the areas with the highest percentage of diabetes?
To improve the display of areas with the highest percentage of diabetes on this map, several design changes could enhance clarity and emphasis:
1. Color Gradient Adjustment:
The current purple color scheme uses shades that are quite similar across regions. To make the highest percentages stand out more, you could use a darker or more saturated shade for higher values and lighter shades for lower values. Alternatively, you could use a multi-color gradient (e.g., yellow to red), where red signifies the highest percentages and yellow the lowest, to visually separate regions more effectively.
2. Range Labels and Breaks:
Add clear numerical ranges or labels on the map legend with intervals, making it easier to distinguish between states with close percentages. This would give the viewer a clearer idea of which states fall within high, medium, or low percentage bands.
3. State Boundaries and Borders:
Thicken or darken the state boundaries to make it easier to visually separate states, especially those with similar shading. This can help users locate specific areas more easily.
4. Highlight Top 5 or 10 States:
Consider adding a highlight or call-out effect for the states with the top 5 or 10 highest diabetes percentages. For example, placing a bold outline or a small label displaying the exact percentage can help viewers quickly identify the worst-affected areas.
5. State Labels:
Adding state names or abbreviations directly on the map can help users identify specific states without having to cross-reference a separate source.
6. Interactive Features (for digital maps):
If the map is digital, including hover-over functionality that displays the specific diabetes percentage per state can be useful for giving precise information without overcrowding the map visually.

Image 9
AI Fails


In which area used the least kWh, and in which decade were the buildings in this image built?
In this image, the area that used the least kWh is indicated by the smallest and shortest vertical bars, which appear on the far left side of the image. Based on the color gradient in the legend, this area would have used around 416.33 kWh or less.
The buildings in this area were built in the 1970s, as shown by the green color coding on Layer 2 for "Decade Built." Other areas, marked in yellow, represent buildings constructed in the 1980s.
Which vertical bar is the shortest? Which areas are marked in yellow?
In the image, the shortest vertical bar is the one on the far left side. This bar represents the area with the least kWh usage.
The areas marked in yellow correspond to buildings constructed in the 1980s. You can see these yellow areas surrounding the base of several of the vertical bars, indicating that those buildings belong to the 1980s decade. Specifically, the areas around the three bars in the center-right of the image are marked in yellow, representing those buildings from the 1980s.
Do the vertical bars indicate the amount of kWh used or the decade the building was built?
The vertical bars in the image indicate the amount of kWh used. The height of each bar reflects the total kWh usage (as shown in the legend for Layer 1, "kWh Used (2012 only)"). The color around the base of the bars, which is linked to Layer 2, represents the decade the building was built (e.g., green for the 1970s and yellow for the 1980s).
In summary:
Vertical bar height = Amount of kWh used
Base color = Decade the building was built

UNIVERSITY of WEST FLORIDA