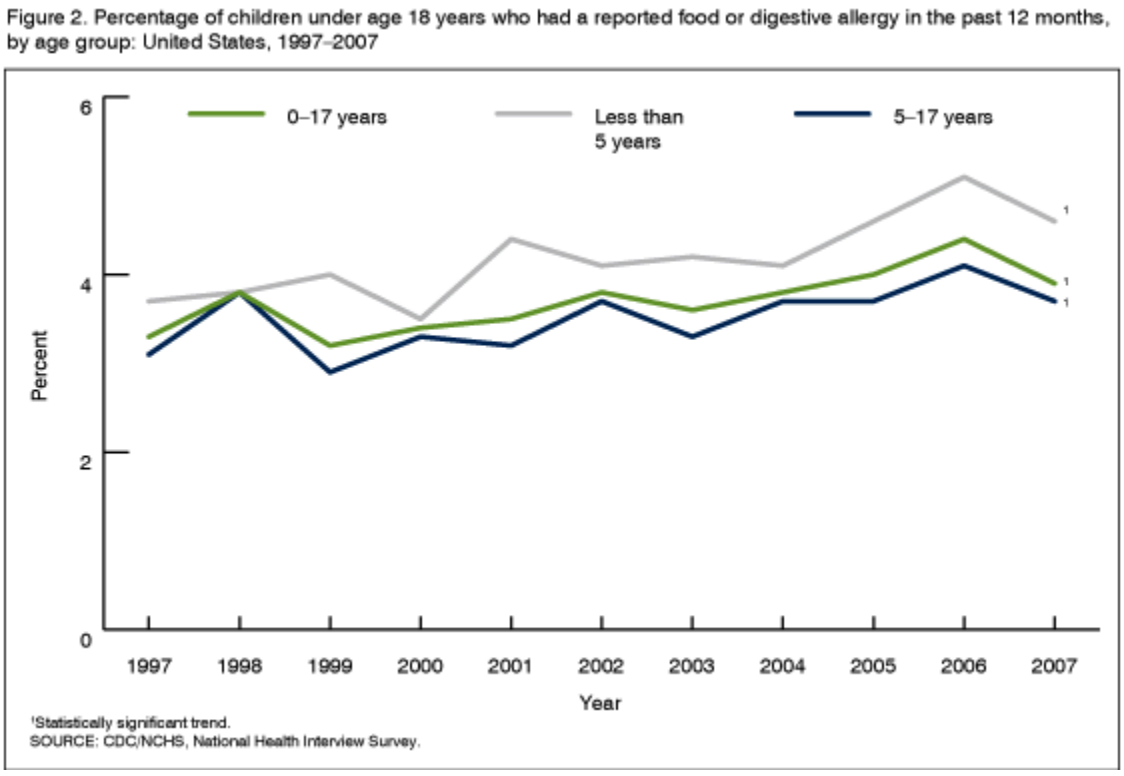
Isaac Zhong, William Cai, Adith Sivakumar
COMPSCI 571 Final Project Process Book

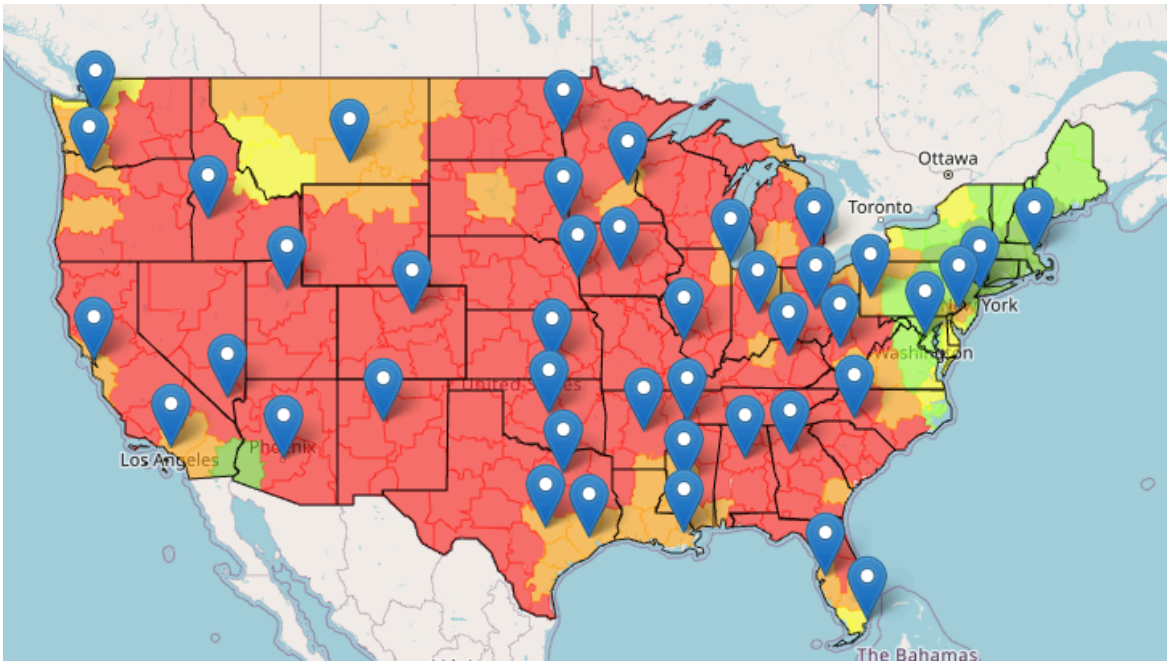
1. **Overview and Motivation: Provide an overview of the project goals and their motivation. Consider that this will be read by people who did not see your project proposal.**

   Many individuals suffer from allergies, which can significantly impact daily life, from simple activities to emergencies. Our project is designed to address this challenge by creating clear, interactive visualizations of allergy data. These visualizations aim to help the public, including patients, caregivers, healthcare professionals, and policymakers, grasp the scope and specifics of allergy-related issues quickly and effectively.

2. **Related Work: Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.**



   The source of inspiration was the CDC's statistics about Food Allergies Among U.S. Children: Trends in Prevalence and Hospitalizations, which provides comprehensive allergy data for children. This data shows that allergies increased from 1997 to 2007, which is also why we want to do the allergy trend and prevalence in the USA.



Additionally, environmental allergens also contribute to allergies. We looked at the Allergens and Pollen section on the CDC's website. Furthermore, we looked at the visual design of interactive dashboards on sites like Pollen.com, which present pollen trends over time in a clean, user-friendly format, as well as a report from FAIR Health about food allergies by each state.

3. **Questions: What questions are you trying to answer with your visualization(s)? How did these questions evolve/change throughout the project? What new questions did you consider as the project progressed?**
   At the beginning of our project, our main question was: "How does the prevalence of allergies vary across different states in the U.S.?" We wanted to understand whether certain regions experience more allergy cases and whether environmental or demographic factors might play a role. As we gathered and visualized the data, new questions emerged. Like, how does allergy prevalence differ by age group? This led us to explore age-specific patterns, such as whether children or older adults report more allergies. And "Are allergies increasing or decreasing over time?" This created a temporal dimension and prompted us to visualize trends across years. Lastly, not only about food allergies, we want to see what pollen allergies affect US people/ cities. Thus, we came up with the question, "Which cities experience the highest pollen levels, and how might this relate to allergy

data?" This question expanded our focus beyond reported allergies, including environmental factors like pollen scores.

4. **Data: Include information about the source, how you collected it (e.g., web scraping), cleaning methods, etc.**
   We have food allergy data by state collected by FAIR Health, a national, independent, nonprofit organization that collects and analyzes data on healthcare costs. We also use the "Allergy Capital" pollen allergy from the Asthma and Allergy Foundation of America. It provides information about tree, grass, and weed pollen scores throughout the year. Also, it ranks the first 100 cities with the highest total pollen score.
   Our data is from the web. It was initially presented in text form, which we manually converted into a CSV file and incorporated into our project.

5. **Exploratory Data Analysis: What visualizations did you initially use to look at your data? What insights did you gain? How did these insights inform your design?**

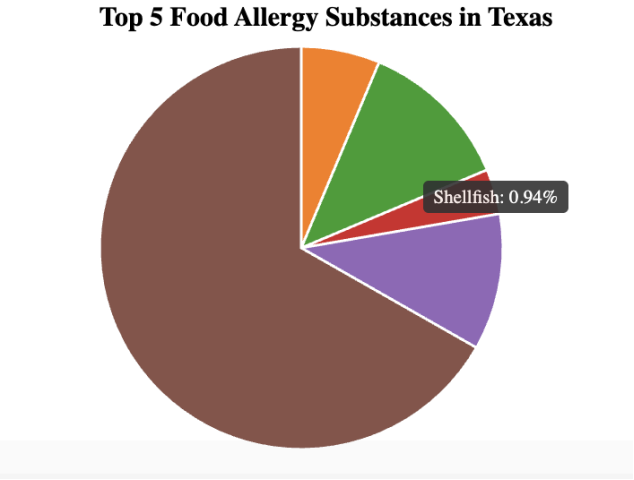
   We started with basic visualizations such as bar charts and pie charts. These helped us identify patterns and outliers early on. We used bar charts to compare allergy types across different states for food allergies. One fascinating insight was that coastal states report higher rates of seafood allergies than inland states, suggesting a possible link between regional diets or exposure and allergy prevalence. We also explored pollen data by plotting ranked cities onto the map. Through this, we discovered that Wichita had the highest total pollen score, and warmer, more humid states—especially in the South—consistently showed higher pollen levels. This supported the idea that climate plays a significant role in environmental allergies. These early findings directly informed our design decisions. We chose to emphasize regional differences with interactive maps, highlight age-group comparisons, present pollen score ranking with color gradients, and show the increasing food allergy trend. This structure made our insights more intuitive and accessible for users exploring allergy trends.

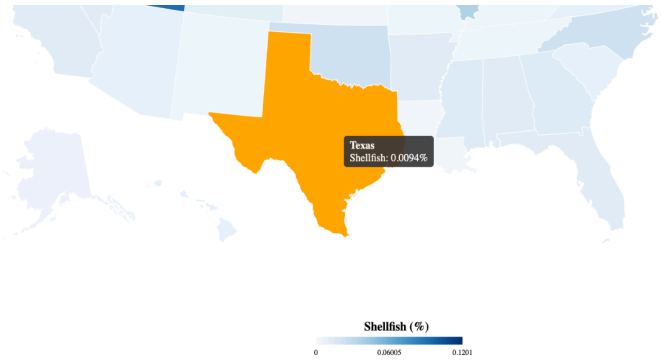
6. **Design Evolution: What are the different visualizations you considered? Justify your design decisions using the perceptual and design principles you learned in the course. Did you deviate from your proposal?**
   We considered several types of visualizations to represent our data best and answer our key questions. Choropleth maps were used to show allergy prevalence by state. This design leverages spatial positioning and color encoding to make geographic patterns easy to compare. We chose a sequential color scale to

ensure that users could perceive differences clearly without misinterpreting the data. We also consider pie charts representing the allergy by age groups within each state. The pie chart shows the data. For pollen scores, we considered both heatmaps and ranked city lists. We used a color gradient for these cities to represent the worst and best pollen scores, giving a better view of the pollen score.

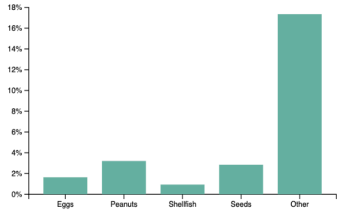
We did make some deviations from our original proposal. Initially, we planned to use pie charts for allergies by age group. Still, based on our team discussions and what we learned about perceptual accuracy, we replaced them with showing different portions for each state for more precise interpretation. Our design decisions were to minimize cognitive load and maximize clarity, ensuring that users can easily extract insights without being overwhelmed.

7. **Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.**

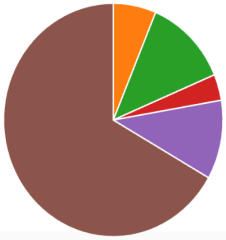




The precentage of food allergy group with anaphylaxis in Texas

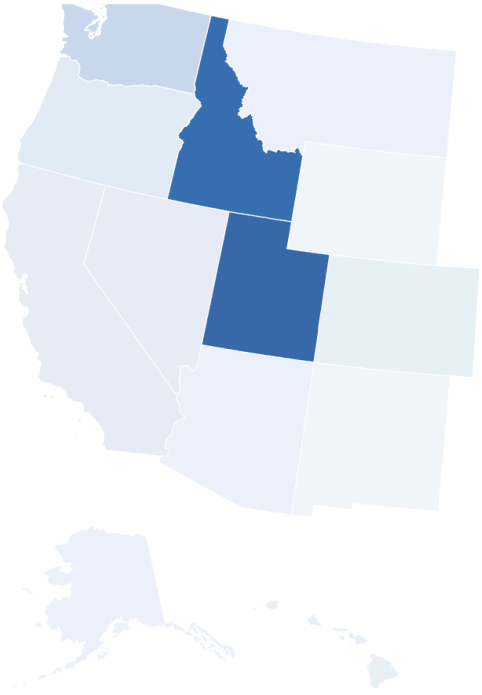


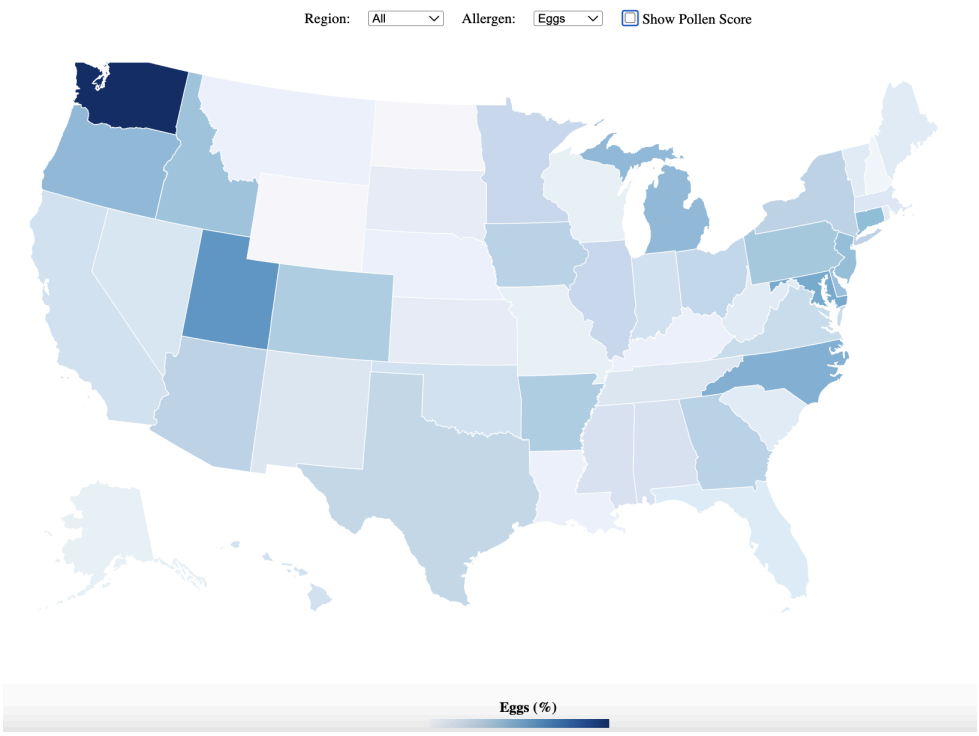
Top 5 Food Allergy Substances in Texas



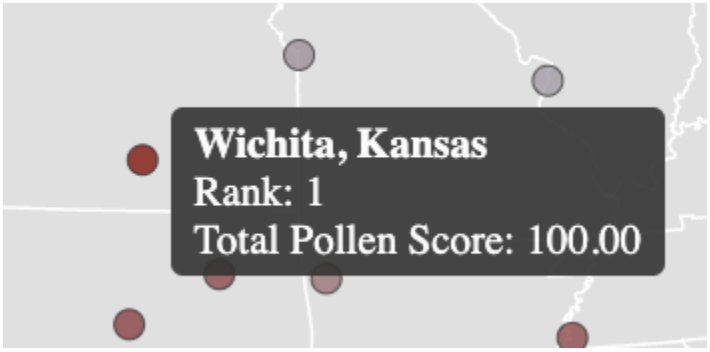
Allergies That Cause Anaphylactic Reactions

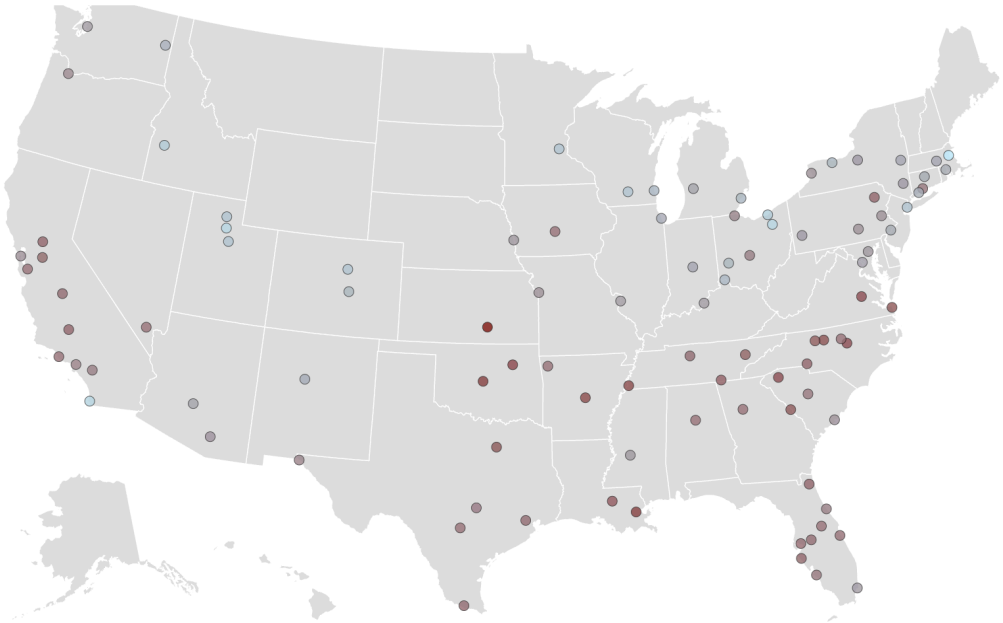
Region: West Allergen: Shellfish Show Pollen Score





The first map gives the user an insight into the allergies that cause Anaphylaxis in each state. Users can click each state to see the top five foods that cause anaphylactic reactions and the percentage of anaphylaxis diagnoses. Also, users can see the trend by selecting different allergens. In addition, users can focus on other regions by choosing a region like the northeast, the west, etc. Lastly, users can see the top 100 cities with the highest total pollen score by checking the pollen score box. After the map is shown, users can hover over the small circle on the map; it will show the city's name and state they are in, the rank, and the total pollen score. The top 100 cities are classified by color, with dark brown representing the worst and light blue representing better.



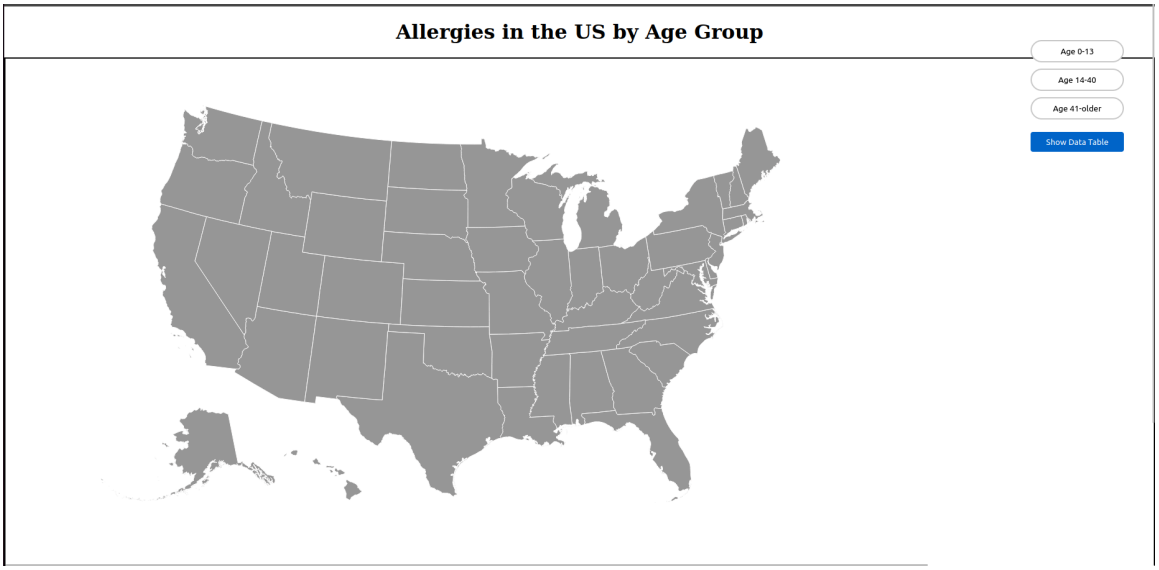


For the total pollen score, these southern states have higher scores for pollen score rankings than the northern states. The warm weather and humid environment might allow more plants to blossom. The user can see the pollen score because we use the color gradient to show the difference in pollen score. The redder the color of the city, the worse the pollen score. For the food allergy part, we can see that seafood allergy has higher incidents in states having coastlines; on the other hand, inland states have a lower seafood allergy rate.

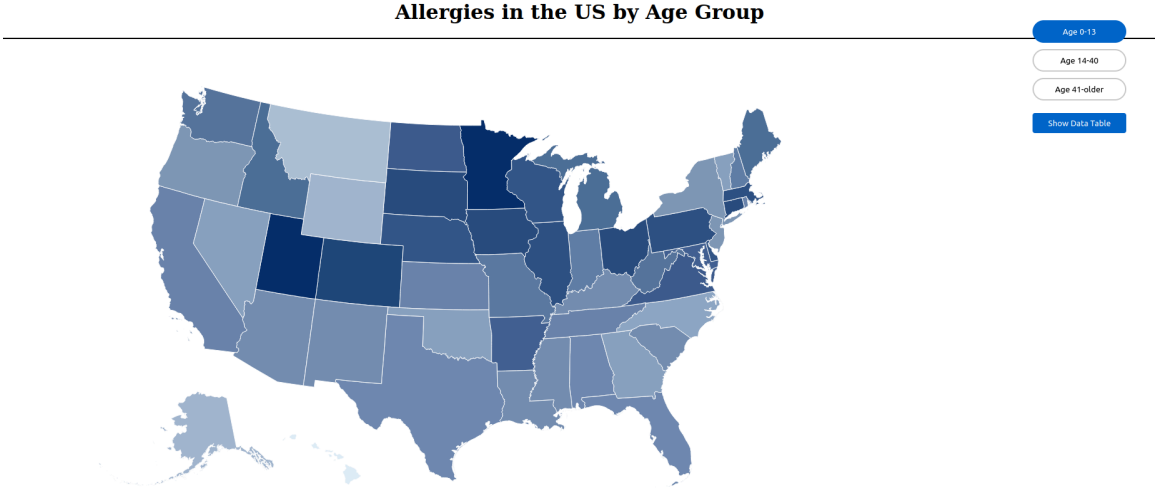
The second map offers insights into the prevalence of food allergies across different age groups in the US. When the map is first opened up, all of the states are gray. Users can hover over each state to see these prevalence rates in three groups: Age 0-13, Age 14-40, and Age 41-older.

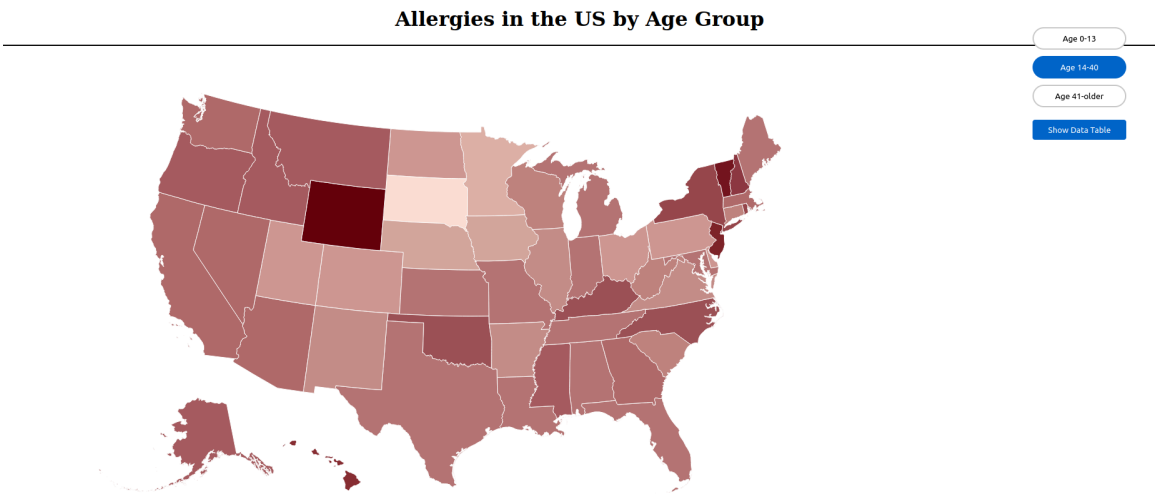
To the right of the map are buttons for these three age groups. When a user clicks on one of these buttons, the states will become colored, with darker hues representing higher percentages and lighter hues representing lower percentages for that group. The map becomes blue for Age 0-13, red for 14-40, and green for 41-older.

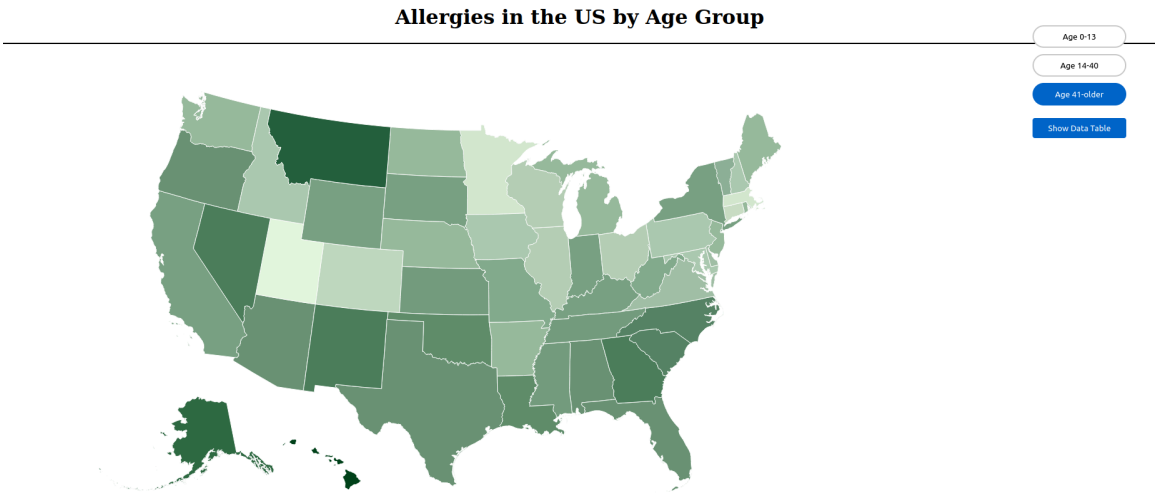
Allergies in the US by Age Group



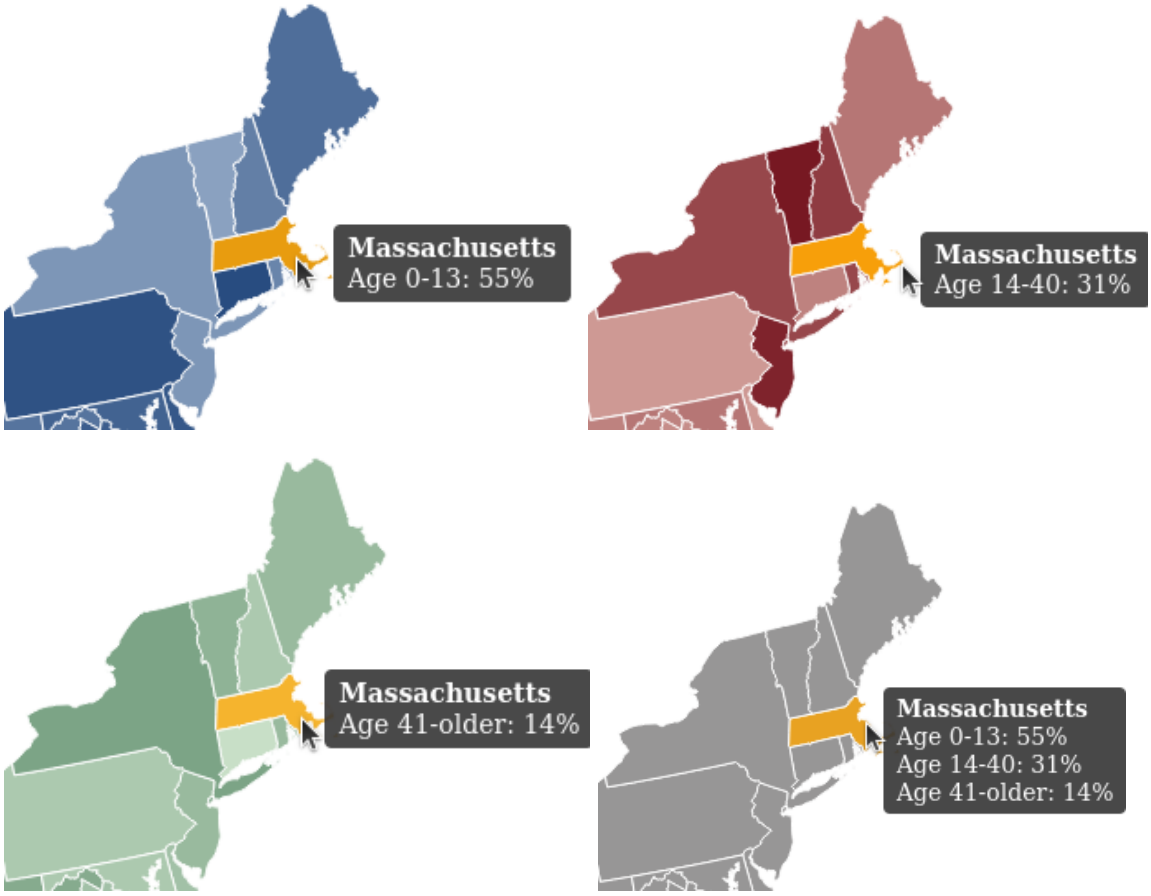
Allergies in the US by Age Group



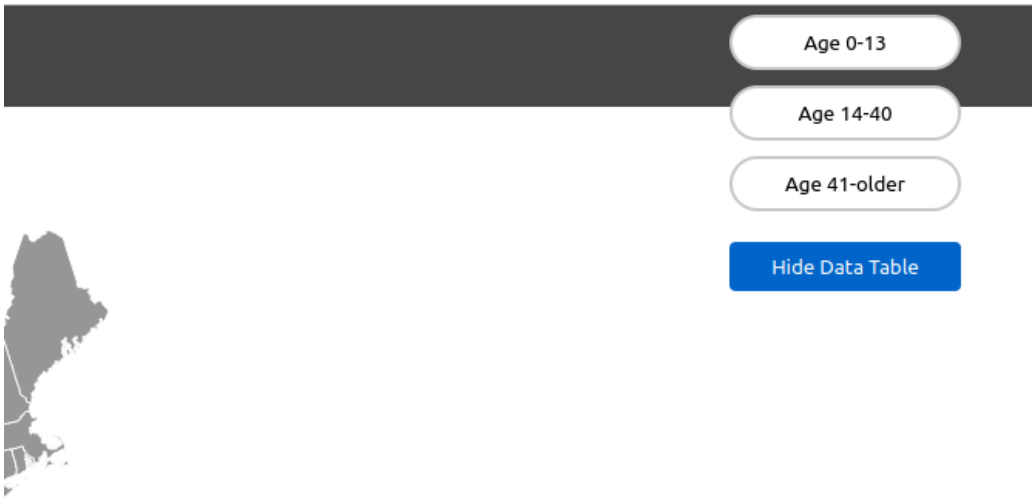




In many states, especially Midwestern states, people aged 0-13 make up the biggest percentage of allergy prevalences. This could be because younger people tend to have less developed immune systems compared to older people. The prevalence of allergies in older adults aged 41 and up tend to be either the lowest or the same as those for those aged 14-40.



After a user selects an age group from one of the three buttons, they can hover over a state to see the percentage for that group only in that state. When that age group is deselected, the states become gray again.

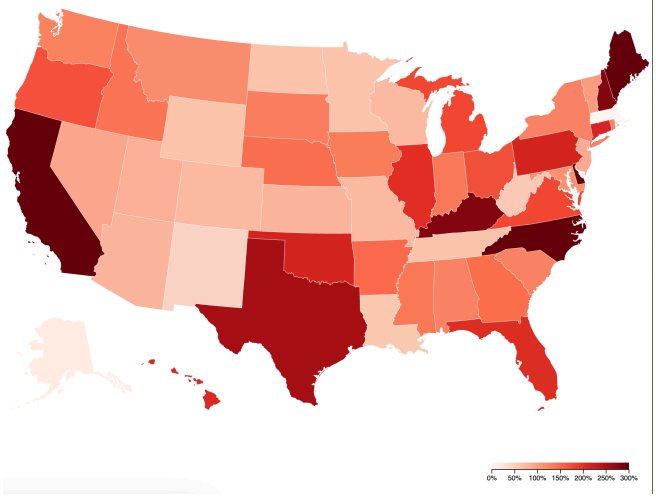


Age Group Data

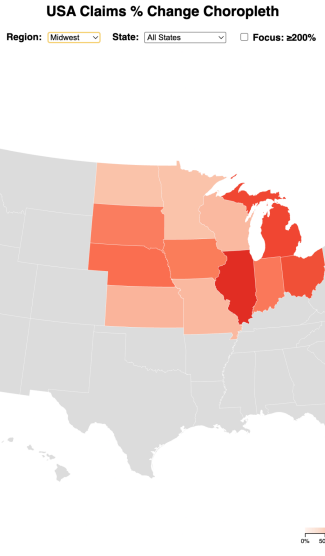
State	Age 0-13	Age 14-40	Age 41-older
Alabama	45%	30%	25%
Alaska	37%	32%	31%
Arizona	44%	31%	25%
Arkansas	52%	28%	20%
California	46%	31%	23%
Colorado	57%	27%	16%
Connecticut	56%	29%	15%
Delaware	54%	27%	19%
District of Columbia	48%	34%	18%

Additionally, there is a button to show or hide a table showing the age group data in tabular form. This table lists all of the states in one column and the percentages of allergies for each age group are in the next three columns. By consolidating the data into one table, users can easily see the prevalences for multiple states at once.

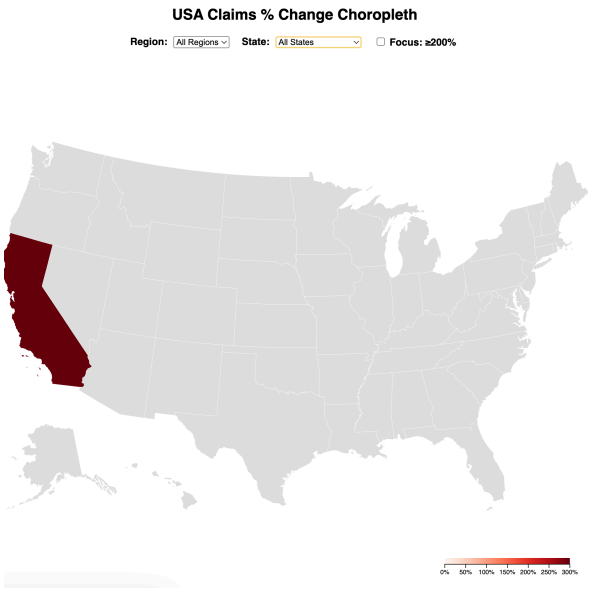
Our interactive choropleth map visualizes the **percent change in allergy-related medical claims** across U.S. states, using a red scale where darker shades indicate higher increases. The visualization allows users to explore national, regional, and state patterns through several dynamic features. **Hovering over any active (non-grayed) state** reveals a tooltip displaying the **state name** and its corresponding **percent change in allergy claims**. This feature provides immediate, on-demand detail without requiring clicks or additional screens.



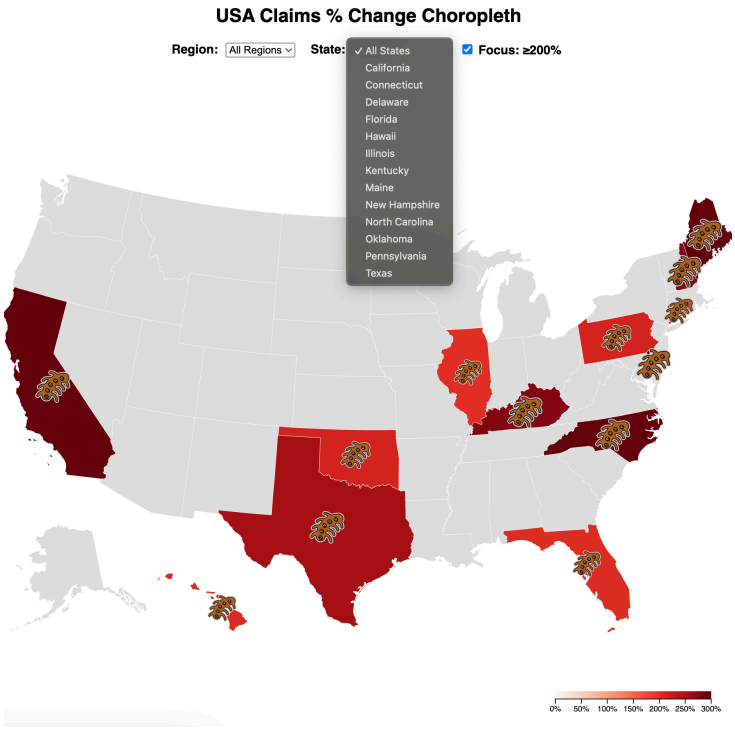
The map initially displays all U.S. states, colored based on the percent change in allergy claims. This view provides a **broad overview** of trends nationwide, helping users quickly identify states with notably high increases.



Selecting a U.S. region (e.g., *Midwest*) from the Region dropdown highlights only the states within that region. All other states are **grayed out and non-interactive**, enabling users to focus on localized trends and compare states within a specific geographic group.



Choosing a specific state (e.g., *California*) from the State dropdown zooms in on that state by **dimming all others**. This supports a **clean, distraction-free analysis** of individual state-level data.



When the **"Focus: ≥200%"** checkbox is enabled, the map dynamically filters to highlight only those states experiencing a **≥200% increase** in allergy-related medical claims. All non-qualifying states are grayed out and turned off to eliminate visual distractions.

- To draw further attention, a **bacteria-shaped SVG icon** is overlaid on each high-risk state. The size of the icon scales with the severity of the percent change, making extreme outliers visually prominent.

- The **State filter dropdown** is also refined in focus mode, showing **only the eligible states** (i.e., those with ≥200%

increase), allowing users to drill down into individual high-risk states.

- As with all other modes, **tooltips remain active** on hovered states, providing the **exact % change** and **state name**.

This interaction mode helps users **quickly identify and analyze** states with unusually high allergy trends, supporting targeted storytelling and public health insight.

8. **Evaluation: What did you learn about the data by using your visualizations? How did you answer your research questions? How well does your visualization work, and how could you improve it? (Be honest here. Limitations are part of any project and will be noticeable during grading. Acknowledging them in this section indicates thoughtfulness in your design process, and, as such, will only help your grade.)**

   By creating our visualizations, we gained insights into the prevalence of food allergies in the US and a glimpse into allergies causing anaphylaxis in states and age groups.

   To answer our research question, we found that allergies aren't evenly distributed or follow a specific trend. Instead, several factors play a role in allergic reactions, such as geographic location, environmental factors, climate, and lifestyle.

   The user will see how allergic reaction claims happened with different allergens. Furthermore, users can also see the pollen score distribution, which shows that southern states have higher total pollen scores than northern states.

   Improvement: For map 1, we can visualize each allergen by a different color. This will help users better understand when they select different allergens.

By interacting with the visualizations, we have found that allergies tend to be prevalent across people of all ages, but these prevalences vary across each state. The filters for the different age groups give us insights into how much these groups are affected by allergies in their respective states. Distinct colors (blue, red, and green) and hues are able to very clearly differentiate the age groups.

We find that in many states, especially Midwestern states, people aged 0-13 see the biggest percentages of food allergies in those states. The percentage in Minnesota is 61%, tied with Utah for being the highest in age 0-13. Hawaii, meanwhile, has 28% which is

the lowest in age 0-13, likely due to climate and environmental factors which particularly stand out in Hawaii. For age 14-40, the percentages tend to be rather high in the Northeast and the lowest in the Midwest. For age 41 and older, the percentages are overall the highest in the Southeast and lowest in the Northeast and Midwest.

When a user clicks on a button for a specific age group, they can hover over a state to only see the percentage for that age group, and not be distracted by other values.

The table consolidates all of the data presented in the map, allowing users to more easily see the percentages for multiple states at once.

One improvement would be to have the highest and lowest percentages in the table stand out by bolding or a different color, such that users can easily find these values without having to scroll through the whole table.

By interacting with our choropleth map visualizations, I learned that food allergy-related medical claims have increased significantly in specific states, with a few, like California, Texas, and Maine, showing extreme spikes above 200%. The visualizations made it easy to spot these outliers due to the red intensity and the added SVG bacteria icons in focus mode.

The filters helped me answer our core research questions: *Which regions see the highest allergy claim increases?* A*re there geographic clusters of high activity?* With the region filter, I could isolate specific clusters (e.g., the Midwest). With the **state filter**, I could drill down into states like California or Illinois to view their trends. The **focus mode** further streamlined this process by dynamically surfacing only the states with extremely high claim increases and reducing distractions from irrelevant data.

The visualization works well in terms of clarity, engagement, and storytelling. The interactivity is smooth, the color scale is intuitive, and the tooltips and icon overlays make the insights feel immediate and accessible.

That said, there are still areas for improvement. One limitation is that the focus mode completely turns off region and state filters unless handled carefully—we addressed this by filtering the dropdown dynamically. However, it could still be refined for even better usability. Another challenge was placing the SVG icons precisely on each state; for smaller states, they sometimes overlap or obscure the shape. Lastly, this map only reflects percentage change, not population size or absolute claim counts, which might skew interpretation if not contextualized.

Overall, the tool fulfills its purpose and offers a compelling, user-driven way to explore allergy claim trends across the U.S. while leaving room for future enhancements like absolute counts, time filters, or mobile responsiveness.