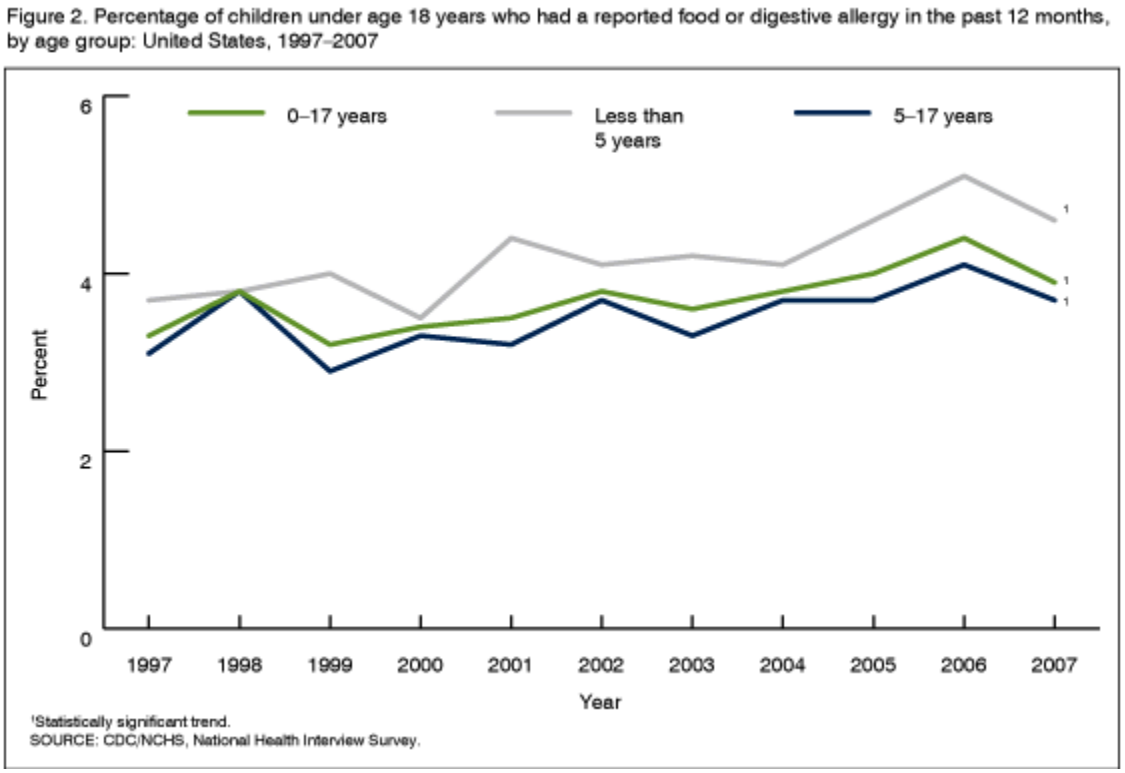
Isaac Zhong, William Cai, Adith Sivakumar
COMPSCI 571 Project Milestone Process Book

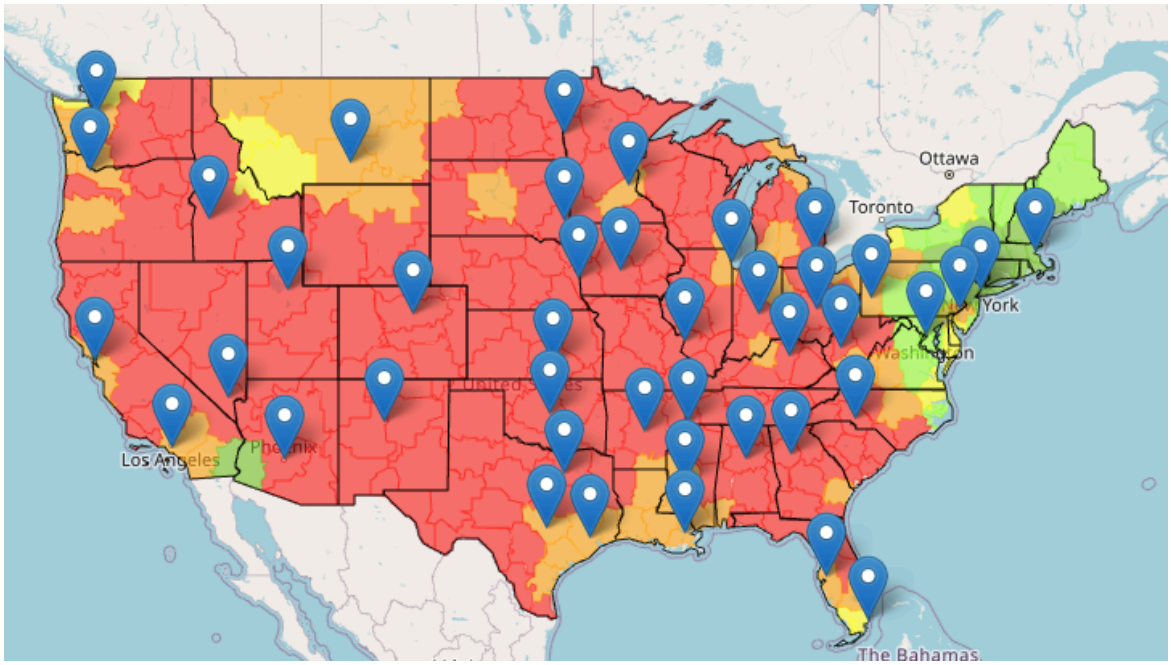
1. **Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.**

   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 Allergy 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. This is also a reason 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 under 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 brought in 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 to include 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. Secondly, we also use the pollen allergy called "Allergy Capital" 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 that have 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 use to initially 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. For food allergies, we used bar charts to compare allergy types across different states. One interesting insight was that coastal states tend to report higher rates of seafood allergies compared to inland states, suggesting a possible link between regional diets or exposure and allergy prevalence. We also explored pollen data using 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 trend of food allergy increasing. 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 the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?**
   We considered several types of visualizations to best represent our data 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 at a glance. We chose a sequential color scale to ensure that users could perceive differences clearly without misinterpreting the data. We also consider pie charts to represent the allergy by age groups within each state. The pie chart could see the data clearly. For pollen scores, we considered both heatmaps and ranked city lists. We ultimately used a color gradient for these cities to represent the worst pollen score and better pollen score, which gives a better view of the pollen score.
   We did make some deviations from our original proposal. Initially, we planned to use pie

charts for allergy by age group, but based on our team discussions and what we learned about perceptual accuracy, we replaced them with showing different portions for each state for clearer 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 first map gives the user an insight into the allergies that cause Anaphylaxis by each state. Users can click each state to see the top five foods that cause anaphylactic reactions and the percentage of an anaphylaxis diagnoses. Also, users can see the top 100 cities that have the highest total pollen score when they hover the small circle on the map. The top 100 cities are classified by color, with red representing the worst and green representing better.

   The second map allows users to see the food allergy prevalence (in percentages) in the US by three different age groups: ages 0-13, ages 14-40, and ages 41 and older. When a user hovers over a state, they can see these statistics for that state. They can also click on a button to show or hide all states and prevalence for each age group in list form.

The third map gives the user an insight into the percent change in allergy-related medical claims by each U.S. state. The map uses a red-colored choropleth scale to indicate the severity of change, where darker red shades represent states with a higher percent increase in claims, up to 300%.

Users can interact with the map in multiple ways:

Region Filter: Allows users to highlight a specific U.S. region while dimming all other states.

State Filter: Isolates a selected state by disabling interaction with all other states.

Focus Toggle (≥200%): Highlights only states with ≥200% increase in claims, hiding all others.

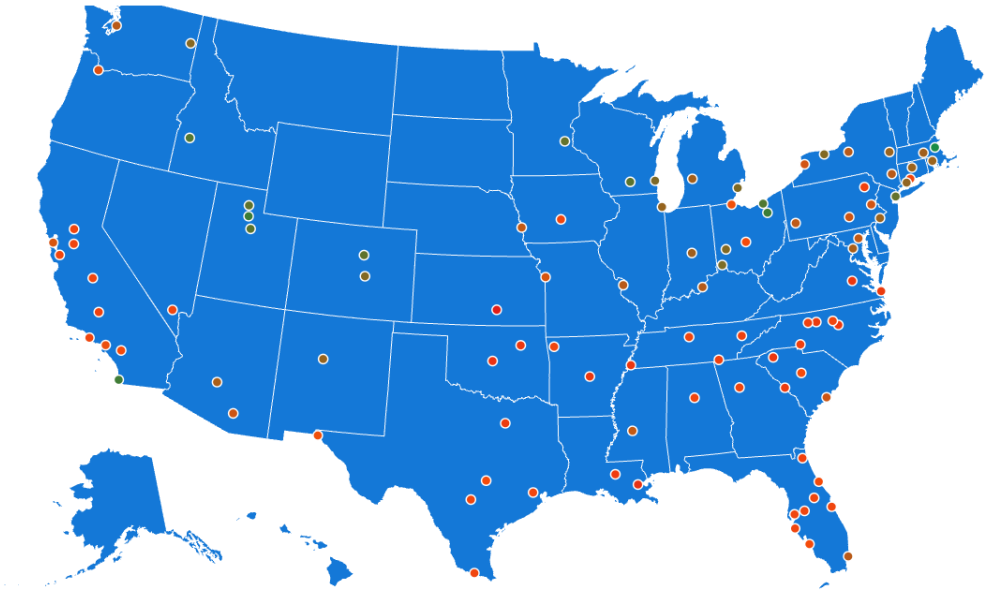
Tooltips: Displays the state name and percent change on hover for active states.
Germ Icon Overlay: Adds a scaled bacteria SVG icon to states with ≥200% increase when focus mode is on.

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 further improve it? (Be honest here. Limitations are a part of any project, and they will be noticeable during the grading process. Acknowledging them in this section indicates thoughtfulness in your design process, and, as such, will only help your grade.)**

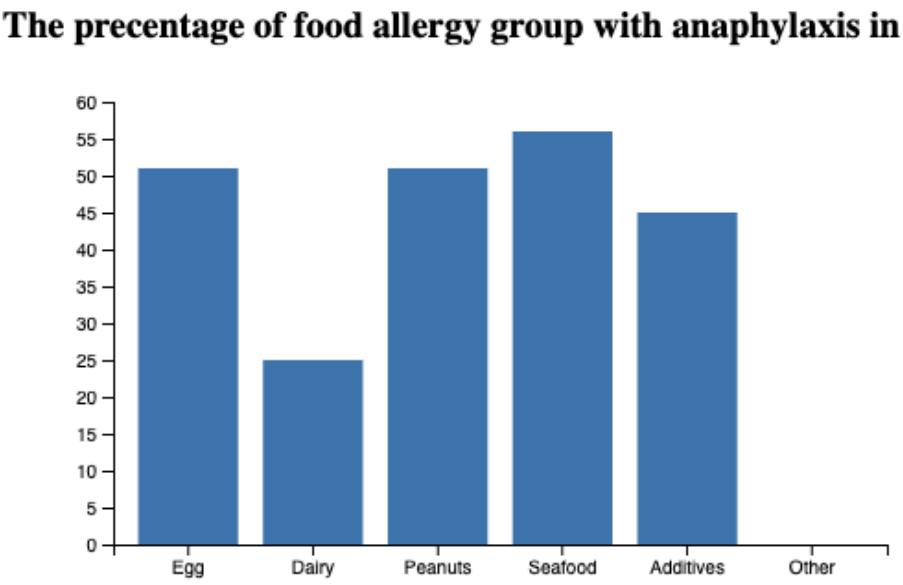
   By creating our visualizations, we were able to gain insights into the prevalences of food allergies in the US. We get a glimpse into allergies causing anaphylaxis in various cities.

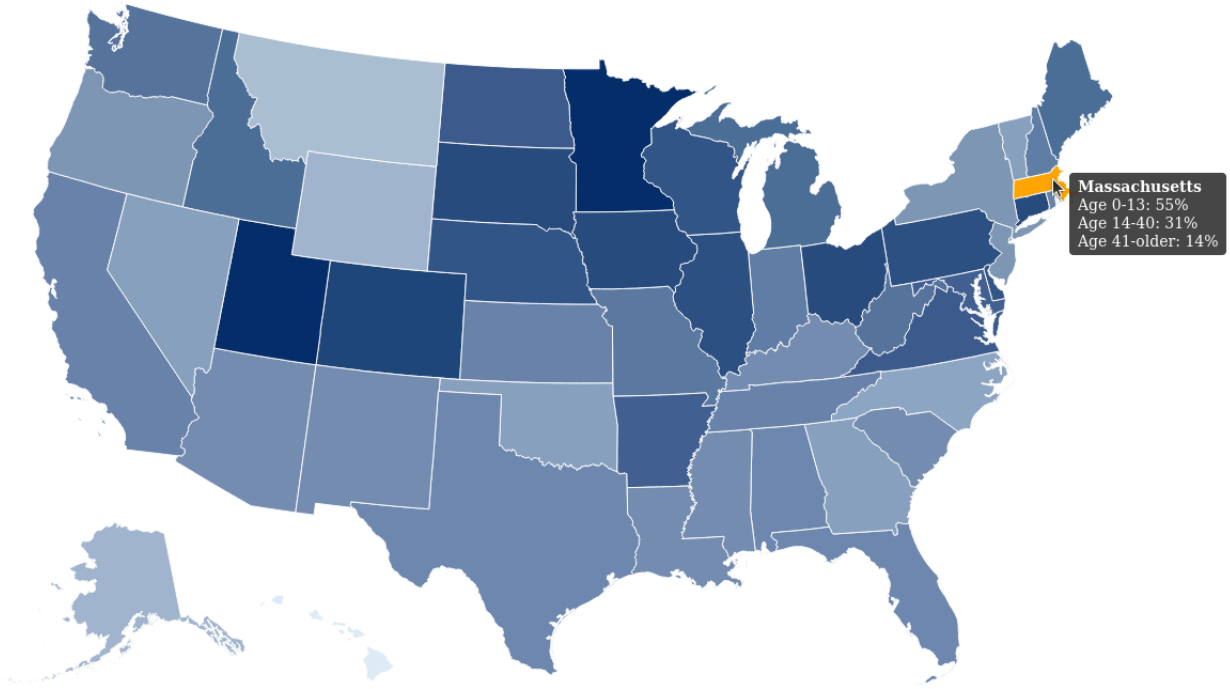
   Based on the allergy prevalence by age group in each state, we found that in almost all states, a large percentage occurs in people aged 0-13.

   In map 1, we can notice that these southern states have higher scores for pollen score rankings than those northern states. This might be the warm weather and humid environment that allow more plants to blossom. The



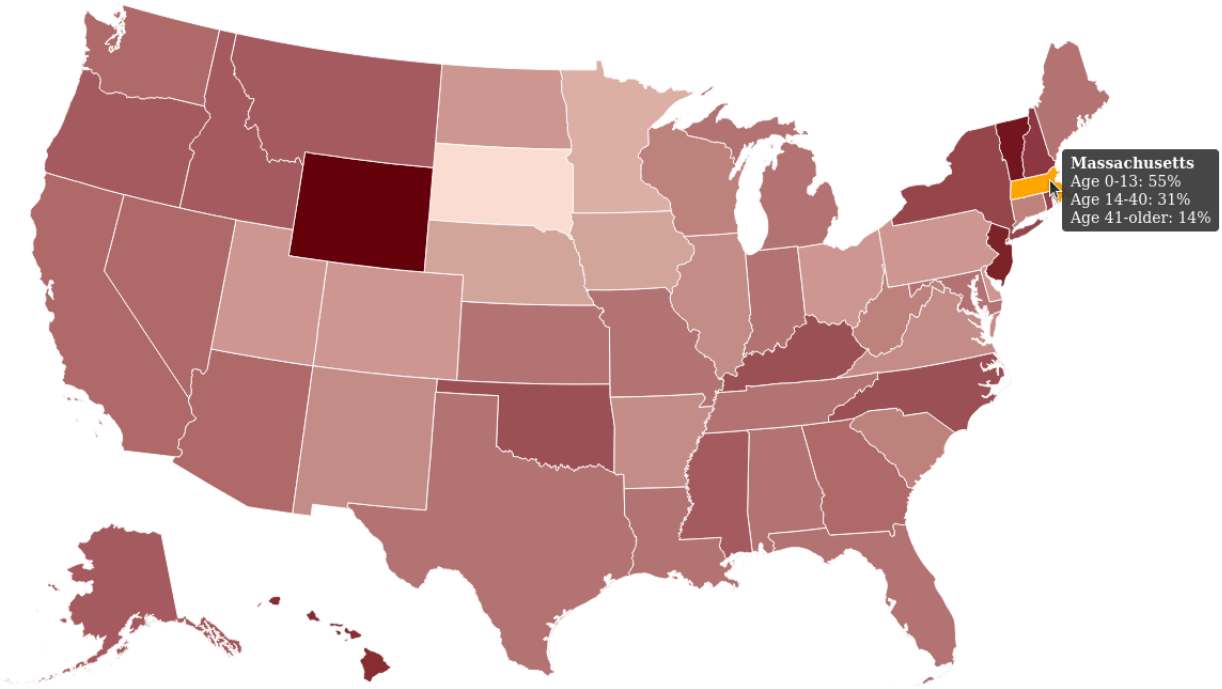
   pollen score can be clearly seen by the user because we use the color gradient to show the difference in pollen score. The redder the color of the city, the worse the pollen score. For the food allergy part, we can see that seafood allergy has higher incidents in states having coastlines; on the other hand, inland states have a lower seafood allergy rate.



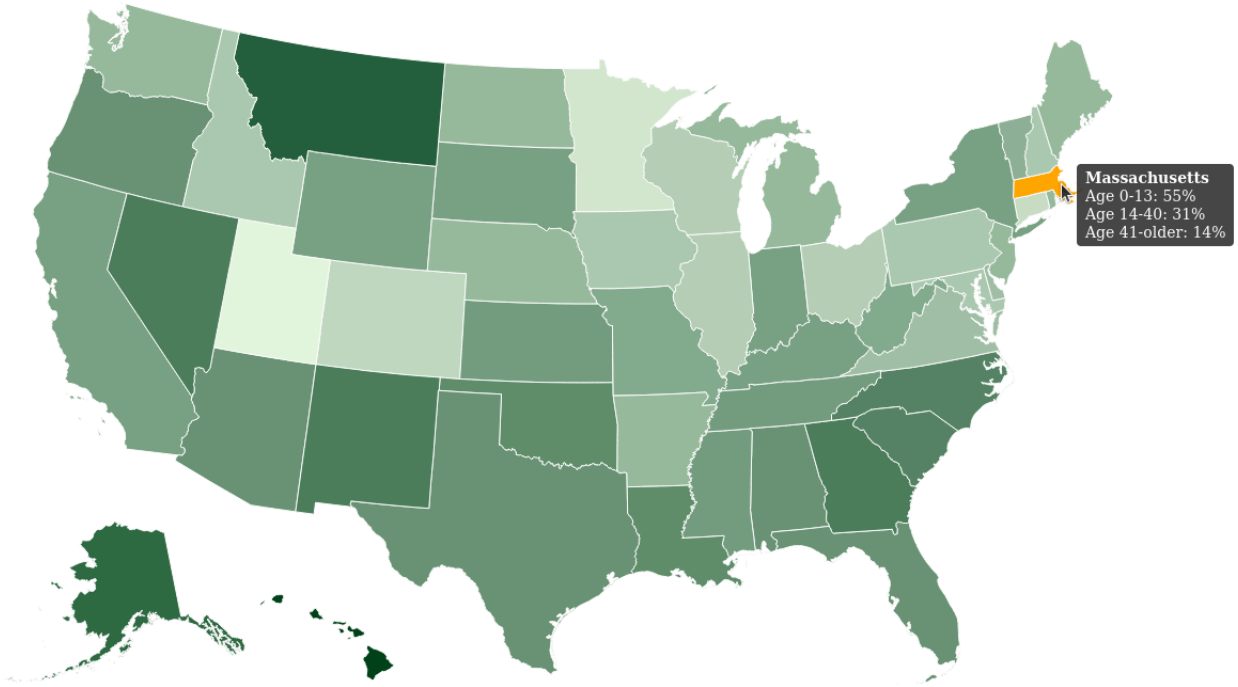


In map 2, we can see the prevalences of food allergies by age groups in each state. Here, the blue states show these trends among people aged 0-13, with darker shades of blue representing higher percentages. These percentages tend to be among the highest in the Midwestern states (Minnesota has the highest at 61%, tied with Utah) and roughly the lowest in the southeast.

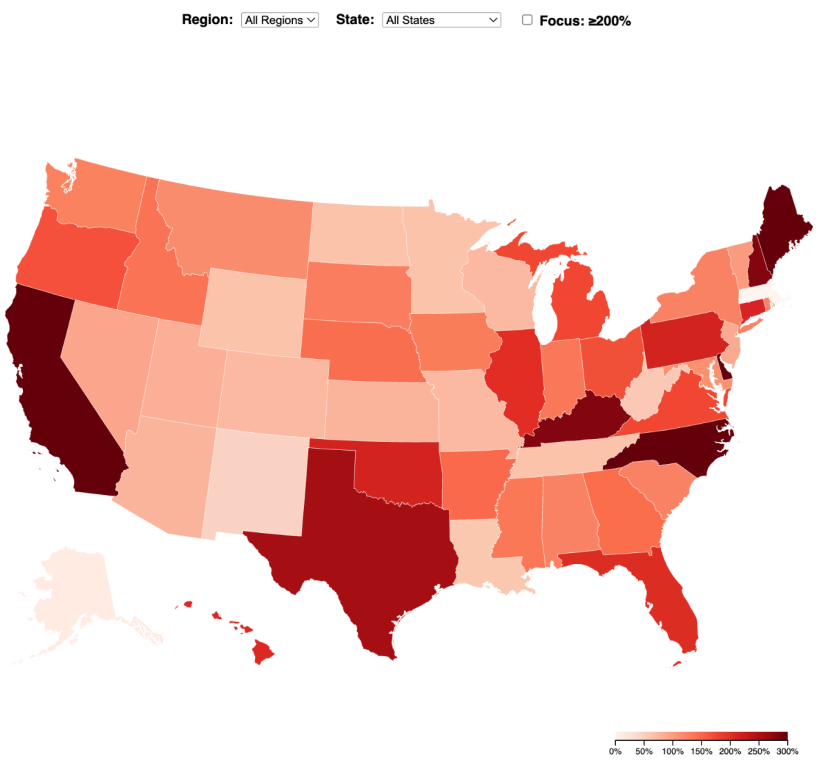
Overall, food allergies are higher in ages 0-13 than older people probably because younger people have immune systems that aren't as developed as those of older people.



Here, the red states show these trends among people aged 14-40 with darker shades of red representing higher percentages. These percentages tend to be rather high in the Northeast and the lowest in the Midwest.



Here, the green states show these trends among people aged 41 and older with darker shades of green representing higher percentages. These percentages tend to be low in the Midwest and Northeast, and higher in the southeast. While allergies can still be prevalent among older people, they tend to have more developed immune systems, so these percentages for each state tend to be the smallest among the age groups.



By using the choropleth map, we observed clear geographical trends in the percent change of food allergy-related claims across U.S. states. States like California, Texas, New York, and Virginia show the highest increases, with ≥200% changes, indicating potential regional surges in either allergy prevalence, awareness, or reporting. The interactive features — such as filtering by region or state and focusing on extreme cases — allowed us to explore patterns in a more targeted way and answer our main research questions regarding regional disparities and high-risk zones.

The visualization works well for identifying outliers, regional clusters, and engaging interaction, especially through the filter and focus toggles. It also enables quick comparisons by state and reveals that certain states in the West and Southeast exhibit disproportionately high increases, which could warrant further health policy attention or environmental study.

However, there are limitations. The map shows percent change only, not absolute values, which might mislead interpretation — a state with few initial cases may show high % growth that isn't as impactful in real terms. Also, we assume consistent data quality and reporting across states, which might not be true. Lastly, the color scale could be fine-tuned to better separate mid-range changes (e.g., 100–200%).

Improvement: For map 1, we can visualize each category to see how it affects each state. This will help users better understand why coastal states have higher seafood allergy rates than inland states.

The visualization for map 2 could be improved by adding a feature for the list form of allergy prevalence such that one can see any group ranked in increasing (or decreasing) order by %. This allows an easier reading of which states have the most food allergies within one age group.

Going forward, this visualization for MAP 3 could be improved by:

- Adding tooltips with both absolute numbers and % change
- Integrating time sliders to show trend evolution
- Overlaying external data (e.g., pollution, population density) to enrich insight