Mouktika Kolluru & Elizabeth Dods
CPSC 4030

Health Inequality Project Process Book

Overview and Motivation:

The focus of this project is to investigate the way that income inequality impacts people's health. A main motivator for choosing this project was our interest in the medical field. I (Elizabeth) am a Bioengineering major and specialize in the design of medical devices. Because of this, I have prior knowledge about the healthcare industry, but I was interested in how income impacts life expectancy and other medical outcomes. In addition, I (Mouktika) am from India and have seen so many people there who live in unlivable conditions due to the fact that they can't afford it, and as a consequence, their health is almost always in bad shape. During my visits to India, I've observed an increasing number of people who struggle with low income and its effects. Because of this, I have a deep empathy for them and wanted to explore and study more as to what exactly is affected, and this dataset diving into the United States of America is a good starting point. Our combined interests made it extremely easy to decide on this project, and we are looking forward to the discoveries we will make within this study!

Project Goals: We wanted to investigate how income impacts life expectancy in the United States. We wanted to determine if there were substantial differences between lower-income and higher-income life expectancies. We also wanted to see if there were geographic trends in this data.

Questions

The original questions we wanted to answer were:

- Which commuting zones in the US have the highest and lowest life expectancies?
- What are the local changes in life expectancy in the top 100 commuting zones in the US?
- What is the relationship between age at death and average household income? How does this change with gender?
- What factors influence life expectancy in these commuting zones?
    - We had data such as % of people in each county with health insurance, etc., that could affect life expectancy.

Because of limits in our dataset, we shifted our questions:

- Firstly, what is the relationship between income and life expectancy?
    - Is there a correlation between life expectancy and income?
    - Do life expectancies vary based on gender (M or F)?
- Secondly, how do life expectancies change annually based on income?
    - Is there a difference in the change in life expectancy between high and low-income people (ie is health inequality getting worse)?
    - Is there a difference in the annual change depending on location?
- Thirdly, is there a geographic trend for life expectancy?

    - Is life expectancy worse in certain regions of the country for low-income people than in others?

We originally considered using commuting zone data. Because our commuting zone data was limited to the top 100 most populated commuting zones in the country, we decided this would limit our visualizations and shifted our questions to work on the state level. In the future, it would still be worthwhile to visualize county or commuting zone-level data if it becomes available. It would also be interesting to visualize the factors related to income that affect life expectancy, like access to doctors, having health insurance, etc.

September 21st, 2023- Dataset Selection Re-do

We started off with a different dataset for this project. The dataset we originally had holds multiple opinionated facebook accounts and analyzes how much misinformation is actually posted and spread.

Concerns about original dataset:

- Not enough attributes for potential visualizations
- No clear distinction between different items (news sources or posts)
- Formatting of document was confusing

We redid the first milestone and decided to go with a different dataset that showcases the relationship between life expectancy and income in different areas of the U.S.

https://healthinequality.org/data/

Data Source: The Health Inequality Project- Data collected by a team of researchers from Stanford, MIT, and Harvard

Pros of new dataset:

- Has multiple attributes to work with
- Multiple sheets of data to utilize
- Easier to come up with different visualizations that actually showcase different aspects of data instead of being repetitive

In our milestone 1 report, we talked about all the different items and visualizations we considered when planning out what exactly we want to do for our project. We then

October 1st, 2023- Project Proposal (Milestone 3)

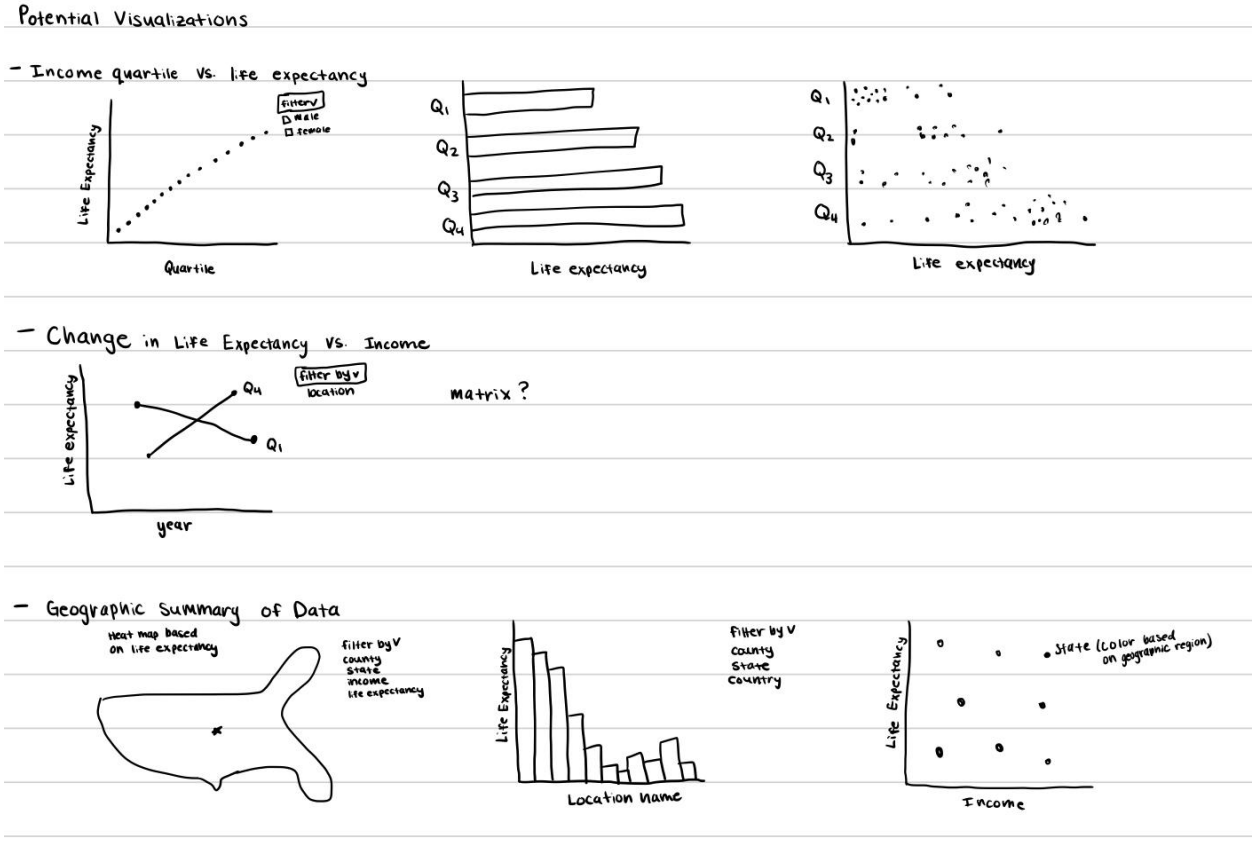
The three questions we wanted to primarily answer with visualizations are as follows:

1) What is the relationship between life expectancy and income?
2) How do life expectancies change annually based on income?
3) Is there a geographic trend for life expectancy?

First we had to process some data to make it easier for us to do the visualizations. The first thing we did was consolidate our data into fewer Excel sheets as opposed to the 16 we were given.

Another thing we had to do was do calculations in Excel to get the specific information we wanted that depended on the already existing information. We were originally given data that was separated for males and females, and by quartile. We averaged this data so that we had aggregated data that we could filter later through interactions. With these two steps, it was much easier to work with the data to create our visualizations.

Initial Brainstorming:



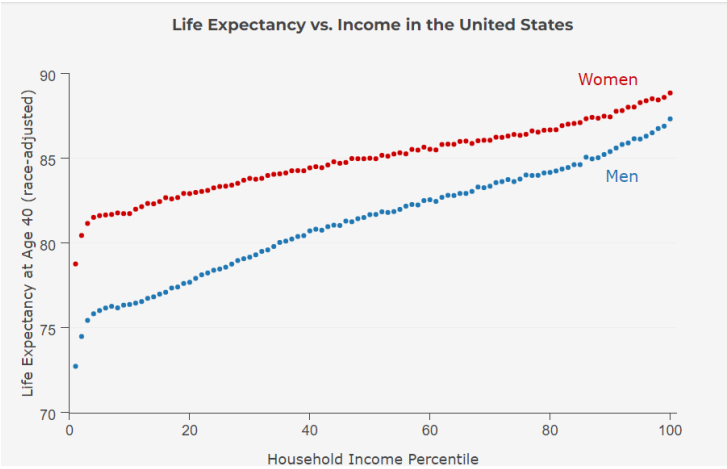
(Elizabeth's original ideas for the three different visualizations)

After brainstorming, we narrowed down our top 3 visualizations for each question and created them in Tableau:

3 visualizations for the first question:

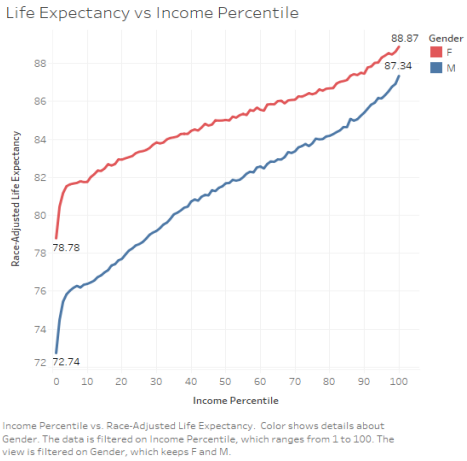
Idea 1: This design shows life expectancy vs income percentile for males and females.

Inspiration:



This is a visualization from the original study. We liked how easy it was to see the trend of increasing life expectancy as income percentile increases. We wanted to make ours a line chart however so that users could select a segment of the line relating to quartile, rather than an individual point.

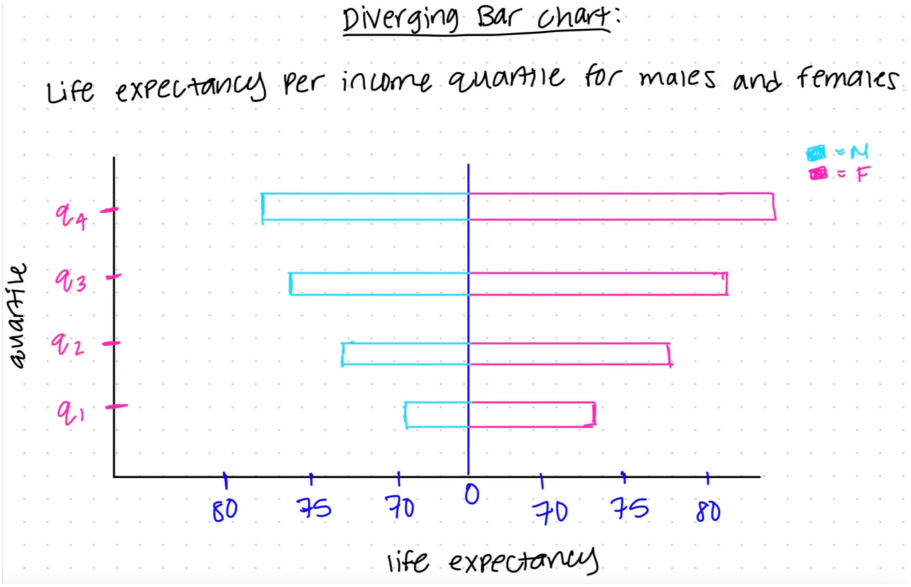
Our design:



Pros: The line chart very clearly shows that, as income percentile increases, so does life expectancy. It also clearly shows the gap in life expectancies based on gender because of the two different colors used. The line naturally draws the user's eyes up, so they can very easily tell that there is an increase.

Cons: The line chart is a very simple design so it may not be visually interesting to the user. Additionally, we would need to find a way determine where the user is clicking on the line and determine the income quartile based off of this coordinate for future interactions.

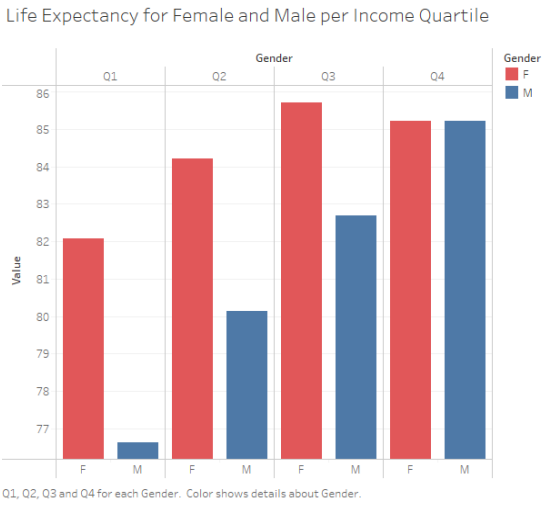
Idea 2: This design shows the life expectancy per income quartile for males and females through a diverging bar chart.



Pros: With this diverging bar chart, it is easy to compare the life expectancies for males and females and also easy to see that life expectancy is higher with each income quartile. We chose to break this data down into quartiles rather than the exact percentiles so that the bar chart did not get too crowded. The colors help to visualize male and female data.

Cons: Our concerns with this design were that the male life expectancies could initially be perceived as negative values because the bars are going towards the left. This design also has less detail than the previous chart, as it only shows trends per quartile. Because we had the exact percentile data available to us, we did not want to prefilter this data for the user by breaking it down into quartiles.

Idea 3: This design shows the life expectancy per income quartile for males and females through a grouped bar chart.

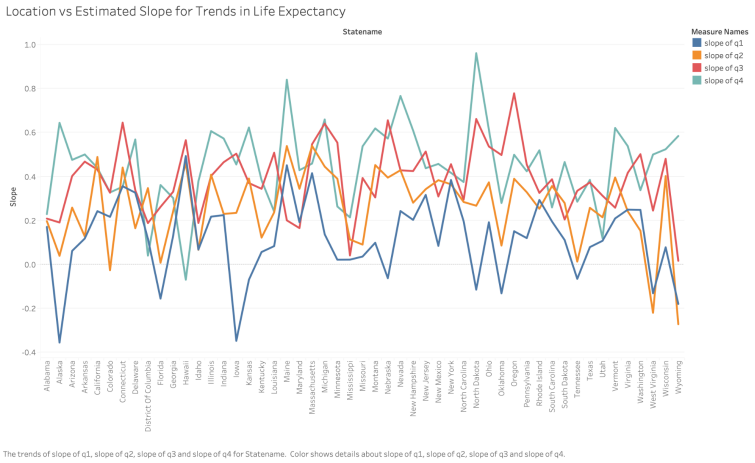


Pros: Grouped bar charts allow for easy comparison between two groups, in this case male and female life expectancies. It is also easy to compare the heights of different bar charts based on their height.

Cons: We did not end up choosing this design because, once again, it only shows quartile data rather than exact percentile values. We also felt that, because we wanted to show the increase in life expectancy as income increases, it is harder to visualize a clear increase with this data.

3 visualizations for the second question:

Idea 1: Our first prototype displays trends in life expectancy per state. Negative values represent a life expectancy that is decreasing, while positive values represent life expectancies that are increasing. Each line represents the life expectancy trend for each quartile.

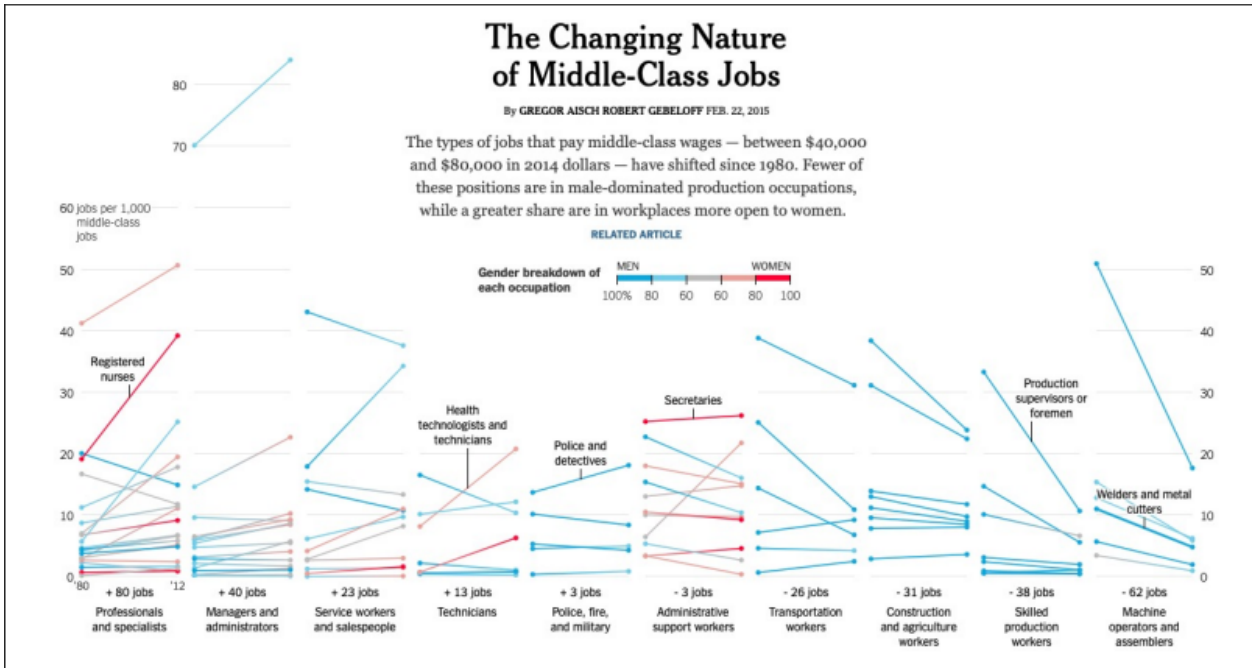


Pros: This graph allows for comparison between quartiles in each state because, looking at the column for each state, it can be seen that the lines closer to the bottom have lower changes in life expectancy than those above them. The colors also help distinguish between quartiles and could be filtered to only show one quartile.

Cons: We felt that the graph was misleading. By using a line graph, we are implying that there is a connection between the change in life expectancy in one state and the change in another state, when, in actuality, there could be no correlation at all.

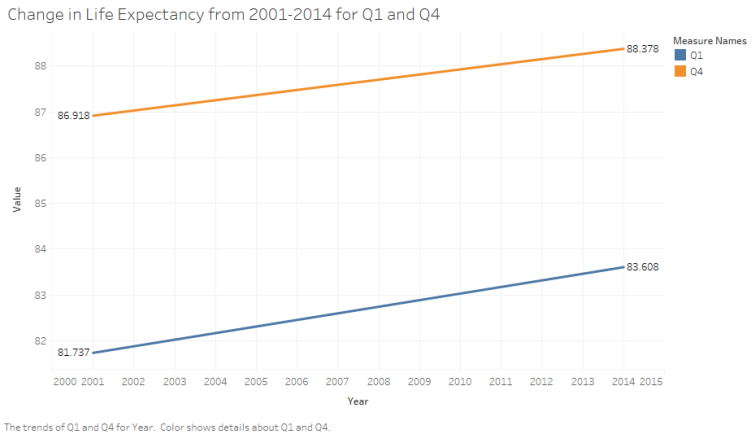
Idea 2: This design shows the change in life expectancy from 2001-2014 for one state at a time. The blue line shows the change for Q1, and the orange line shows the change for Q4.

Inspiration:



We felt this visualization did a good job of visualizing trend data over a period of time.

Our design:



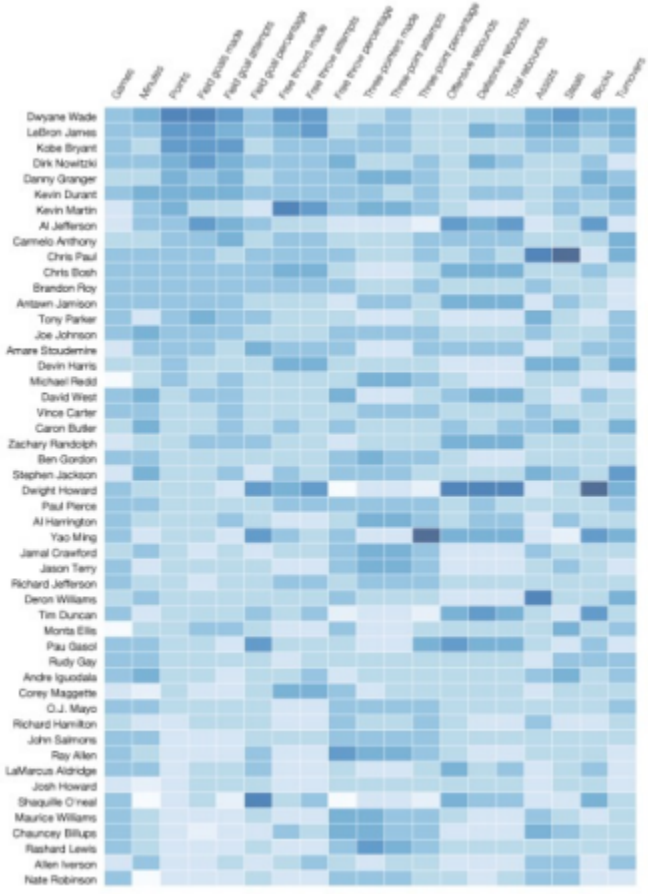
Pros: By using a line, it is very clear to the user that life expectancies for both quartiles are increasing over the time of the study. This visualization also gives the exact life expectancies at the start and end of the study, which could be information the user wants to see.

Cons: We felt this graph lacked impact because it does not show how this data compares to data in other locations. Once again, this graph is also limited to showing only Q1 and Q4, instead of all four income quartiles. Additionally, it is hard to gauge, based on the incline of the line, which

quartile has a higher change in life expectancy, which is a point we want to stress. There is also not a lot of potential filtering or interactions available with this visualization.

Idea 3: Our final visualization is a matrix. On the vertical axis, there is the state name. On the horizontal axis, there is the quartile name (Q1-Q4). The trends in life expectancy per state per quartile are shown in squares.

Inspiration:



We felt this matrix made it easy to see outliers (really dark and light colors), which we wanted to emulate.

Our design:

State-level Trends in Life Expectancy per Income Quartile

Statename	slope of q1	slope of q2	slope of q3	slope of q4
Alabama	0.171	0.198	0.209	0.230
Alaska	-0.354	0.040	0.192	0.645
Arizona	0.063	0.259	0.403	0.476
Arkansas	0.119	0.128	0.468	0.501
California	0.243	0.490	0.432	0.440
Colorado	0.217	-0.025	0.327	0.328
Connecticut	0.357	0.441	0.646	0.354
Delaware	0.327	0.166	0.339	0.569
District Of Columbia	0.118	0.348	0.189	0.041
Florida	-0.154	0.009	0.263	0.362
Georgia	0.132	0.244	0.331	0.303
Hawaii	0.494	0.463	0.566	-0.069
Idaho	0.069	0.067	0.190	0.379
Illinois	0.218	0.409	0.403	0.607
Indiana	0.225	0.231	0.466	0.574
Iowa	-0.347	0.235	0.505	0.456
Kansas	-0.069	0.392	0.371	0.624
Kentucky	0.057	0.123	0.345	0.384
Louisiana	0.084	0.238	0.509	0.241
Maine	0.452	0.539	0.201	0.841
Maryland	0.193	0.345	0.166	0.430
Massachusetts	0.416	0.540	0.548	0.460
Michigan	0.138	0.446	0.641	0.660
Minnesota	0.022	0.391	0.555	0.264
Mississippi	0.023	0.114	0.042	0.215
Missouri	0.037	0.091	0.394	0.538
Montana	0.100	0.453	0.306	0.619
Nebraska	-0.062	0.396	0.656	0.573
Nevada	0.244	0.429	0.429	0.767
New Hampshire	0.204	0.282	0.425	0.610
New Jersey	0.317	0.343	0.514	0.438
New Mexico	0.085	0.384	0.310	0.458
New York	0.385	0.365	0.457	0.417
North Carolina	0.198	0.285	0.295	0.376
North Dakota	-0.114	0.268	0.662	0.962
Ohio	0.193	0.374	0.537	0.614
Oklahoma	-0.131	0.087	0.498	0.280
Oregon	0.152	0.390	0.779	0.500
Pennsylvania	0.121	0.329	0.455	0.425
Rhode Island	0.294	0.254	0.326	0.520
South Carolina	0.196	0.358	0.388	0.260
South Dakota	0.112	0.280	0.205	0.466
Tennessee	-0.065	0.014	0.336	0.286
Texas	0.080	0.258	0.374	0.385
Utah	0.108	0.215	0.313	0.120
Vermont	0.210	0.397	0.259	0.621
Virginia	0.250	0.242	0.418	0.539
Washington	0.249	0.154	0.502	0.338
West Virginia	-0.130	-0.219	0.246	0.501
Wisconsin	0.079	0.403	0.481	0.525
Wyoming	-0.179	-0.271	0.018	0.585

Measure Values
-0.354 0.962

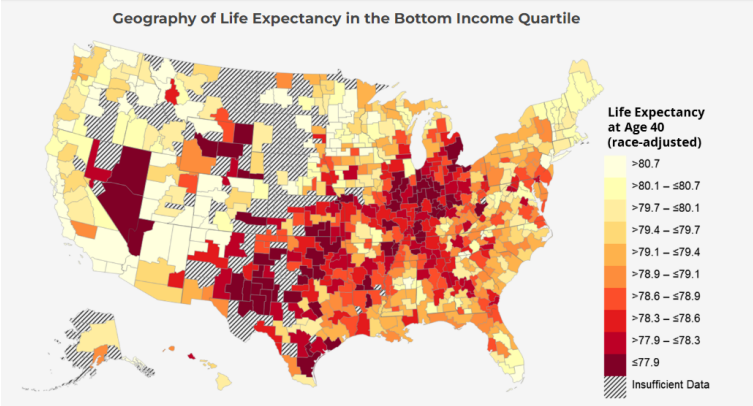
Slope of q1, slope of q2, slope of q3 and slope of q4 (color) broken down by Statename.

Pros: The diverging color map of the matrix makes it very clear which quartiles in each state have decreasing life expectancies (red) and which are increasing (green). The opacity of these colors also increases as the values get more extreme, so the user's eyes are drawn to the outliers of the data. We also included the exact data values in case the user wanted more information.

Cons: The matrix would take up a lot of space on our final dashboard. Additionally, using spatial arrangements of shapes may make it easier to view increasing or decreasing trends rather than just using color.

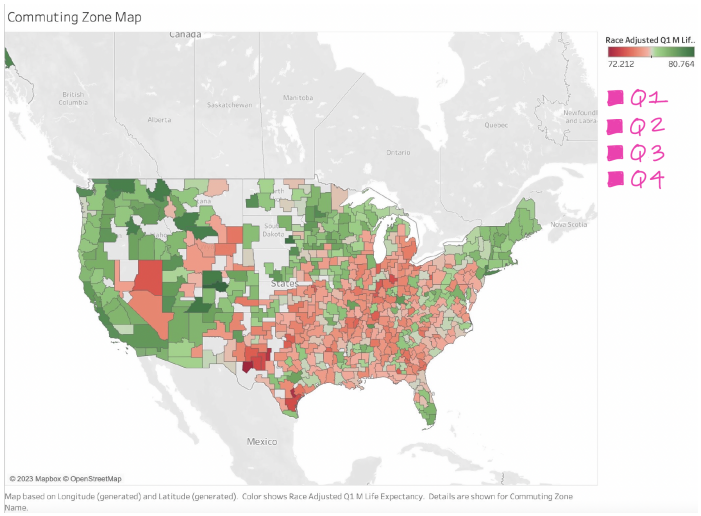
3 visualizations for the third question:
Idea 1: This design uses a map geometry to convey which areas of the country have higher or lower life expectancies. A color gradient is used to show life expectancies. The lower the life expectancy, the more red the area on the graph. The higher the life expectancy, the more green the area is on the map.

Inspiration:



We liked this visualization of life expectancy categorized by commuting zones because you could see geographic trends easily.

Our design:

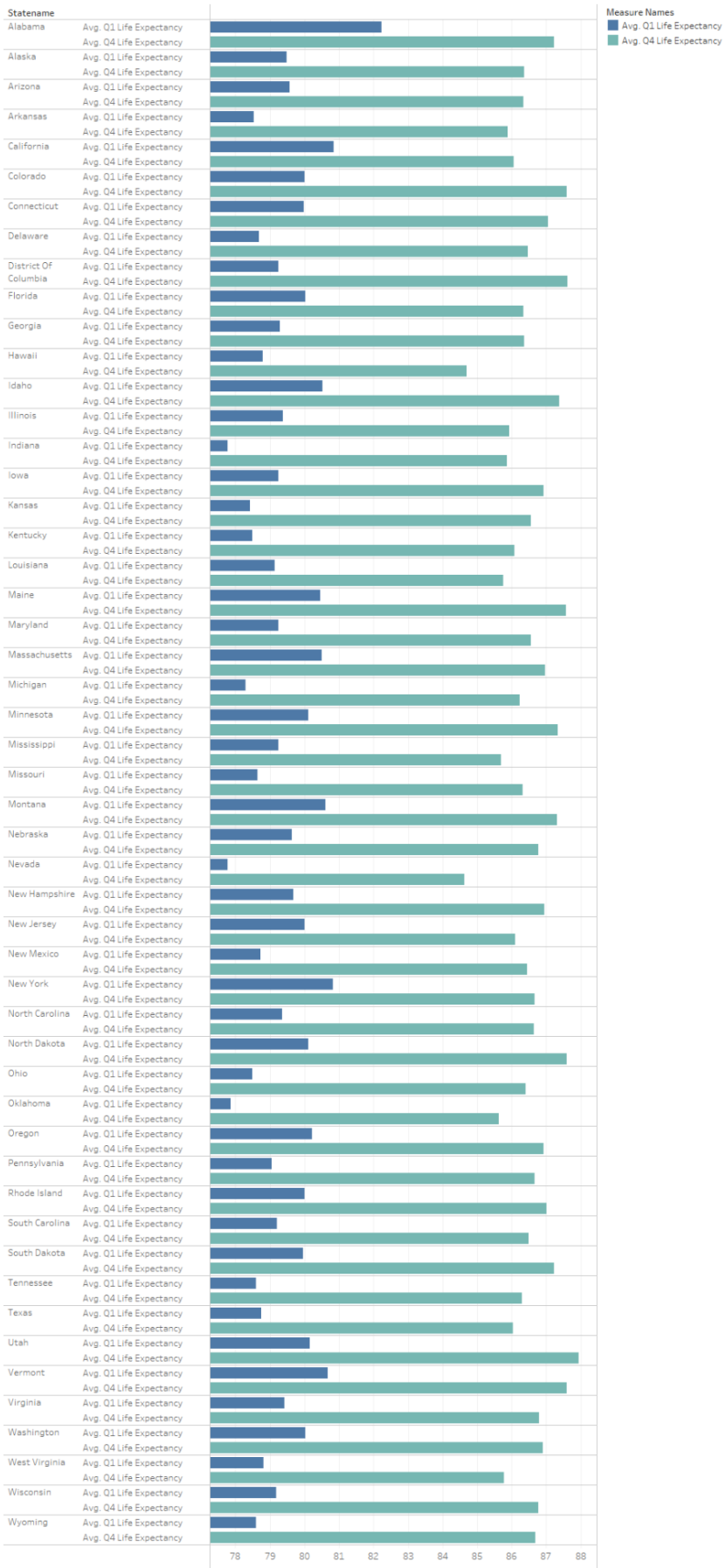


Pros: The map allows users to notice geographic trends, for example, if there is a lower life expectancy in southern states. Additionally, this graph would allow for a lot of potential filtering, like looking at aggregate data or at data broken down by gender and quartile.

Cons: Because we are visualizing commuting-zone level data, it can be hard to tell where states start and end. Additionally, because some commuting zones have no data, there are parts of the map a user couldn't interact with.

Idea 2: Our second design utilizes grouped bar charts to show how average life expectancies in different states differ between quartile 1 and quartile 4.

Life Expectancy of Q1 vs Q4 by State

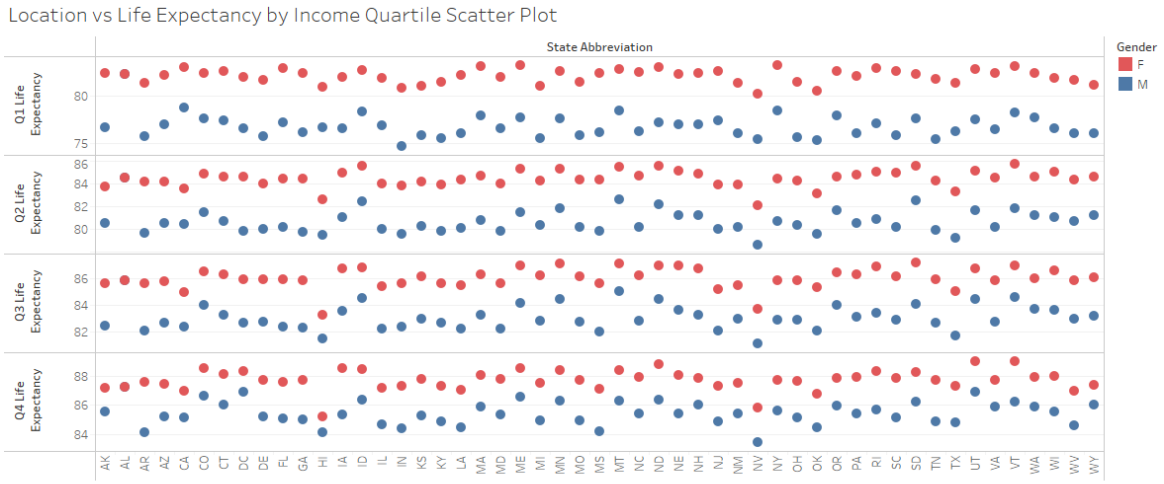


Avg. Q1 Life Expectancy and Avg. Q4 Life Expectancy for each Statename. Color shows details about Avg. Q1 Life Expectancy and Avg. Q4 Life Expectancy.

Pros: Bar charts are very scalable, so they can handle a lot of data. It is easy for a user to compare the heights of the bars to determine what states have the highest and lowest life expectancies and which have the biggest gaps between life expectancies.

Cons: This visualization takes up a lot of room on the final dashboard. It also only looks at Q1 and Q4 data, which is limiting to the user. It is also more difficult to see geographic trends when compared to the map.

Idea 3: Our final visualization is a scatter plot. The x-axis shows the 50 different states and the y-axis displays the life expectancy. Red circles represent female life expectancies, while blue circles represent male. There are 4 graphs, each representing the life expectancy in each state for a different income quartile.



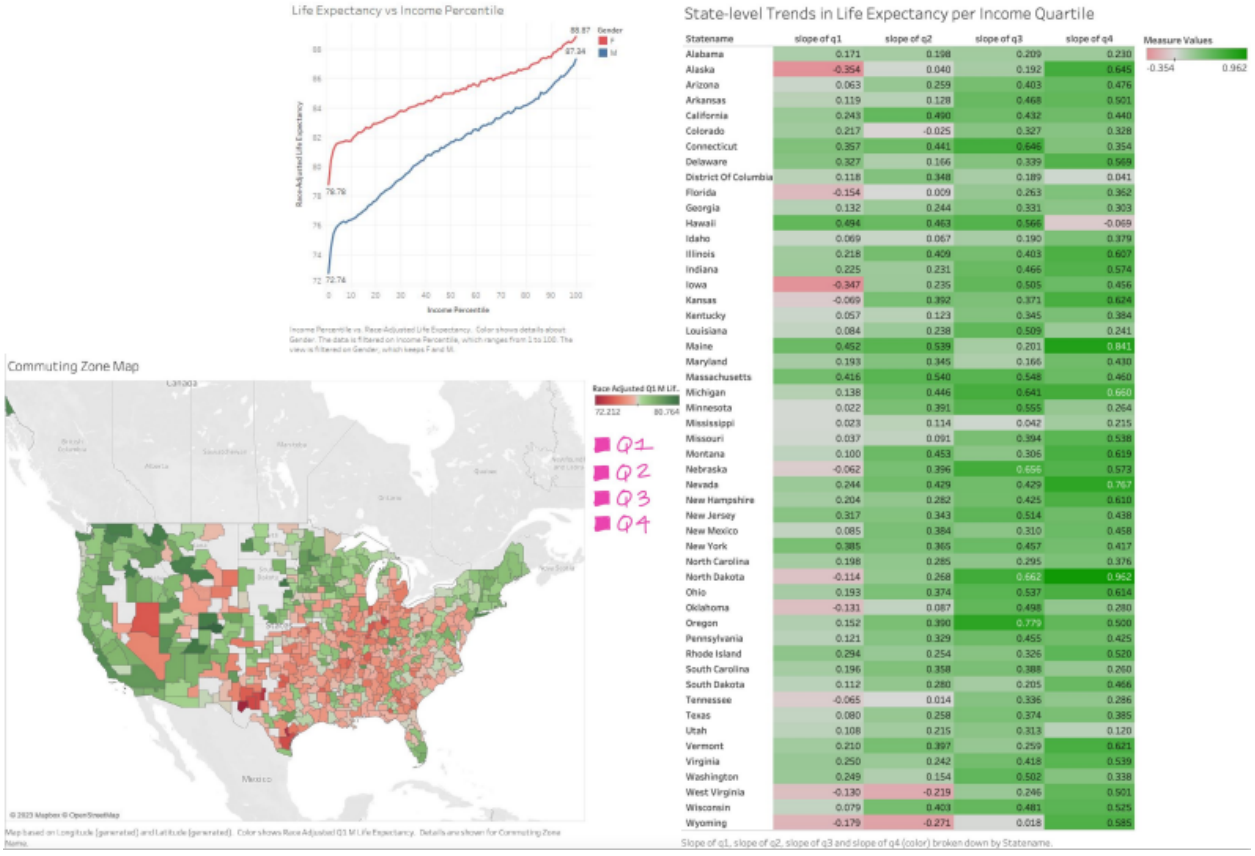
Pros: It is easy to see trends in life expectancy, for example, female life expectancy is typically higher than male. It is also easy to compare the life expectancies between states based on the height of the dots.

Cons: We decided that this visualization was too cluttered, and it can be hard to find the correct circle for the state you are looking at. Additionally, it is hard to read the exact life expectancy because of the scaling of the y-axis. Because the quartiles are separated, it can be hard to compare data between quartiles. This visualization also makes it difficult to determine geographic trends.

Final dashboard:

- Visualization #1: Line graph showing life expectancy vs income quartile
    - Clearly shows that life expectancy increases as income increases
    - Clearly shows the difference between male and female life expectancies
    - Visually, the user's eyes follow the lines of a line graph and are able to perceive the difference between different points along that graph

    - Able to convey a level of detail the other two prototypes could not (only showed quartile data, not exact percentile data)
- Visualization #2: Matrix showing trends in life expectancy for each income quartile in all 50 states
    - Good for data that has two keys (in this case, state or quartile)
    - Color allows for the user to easily notice trends and outliers (used diverging color map so users could quickly detect which values were less than or greater than zero)
    - Summarizes the data neatly, while the other prototypes were too crowded or lacked detail
- Visualization #3: Map of the United States showing geographic trends in life expectancy for each income quartile
    - Using a map allows the user to detect geographic trends in data (ex. there are lower life expectancies in southern states), which was not conveyed in charts that organized state data alphabetically
    - Diverging color map helps show the user which areas have higher or lower life expectancies



Feedback based on Milestone 3:

- Try not to aggregate data too much- the lines of the first visualization are poor and do not allow for much interaction

- The matrix is not the best design- it takes up too much space but doesn't make extracting data any easier, do you really need to see all the quartiles to understand the trends?
- Maybe try to have a visualization with individual data
- Diverging color map for the map may not be the best scale to use because life expectancy is lower, not opposite (like not positive vs negative situation)
- Try to be consistent with colors so as not to confuse the user

October 26th, 2023- Meeting with Dr. Iuricich to Discuss Feedback from Milestone 3

Based on the feedback we received from Milestone 3, we chose to set up a meeting with Dr. Iuricich to explain our dataset in more detail, as some visualization suggestions he had were not possible due to the data available to us. In the meeting, we went through the visualizations and discussed how they could interact with each other as well.

Meeting notes:

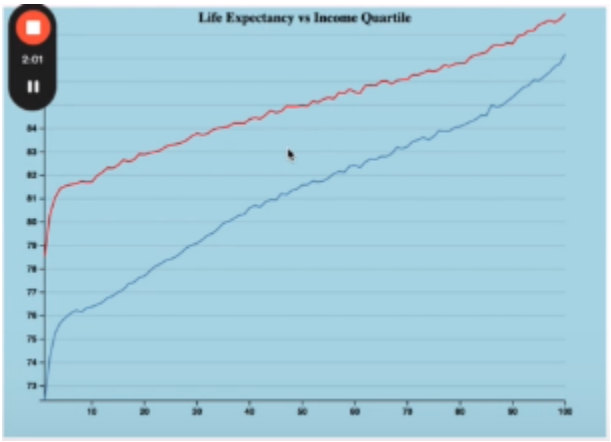
- Start making the process book like a log book, then clean up based on the final dashboard. Have a paragraph explaining how we got to the final visualization (Final summary showing specific decisions).
- Counties are the smallest item. The line chart will have the max level of aggregation, then want visualization at the state level and visualization at the county level
- Use state for map visualization and click on it to get data for the county; county-level data can be messy so don't strive for clarity in that visualization initially
- The third visualization is there to only show relevant data when the state is clicked on (or can change color to stand out)
- The map shows data for the entire data set (compiled data that's not broken up by quartile)
- Expect to see static visualization for Milestone 4

Our main takeaways from this meeting: we plan to keep the aggregated line chart as our first visualization, make the map our second visualization but use a state map rather than commuting zones or counties, and find a new third visualization to show trends in life expectancy for each county in a state, because we did not have data on the individual level. To interact with the elements, the user will click either the male or female line, which will cause male or female data to show in the map. Additionally, where the user clicks on the line will determine which quartile's data will show on the map (because we don't have location data for every percentile for each state). Finally, our visualization showing trends in life expectancy should be at the county level so that, by clicking on a state in the map, only data pertaining to counties in that state would appear.

October 30th, 2023- Work on Visualization 1

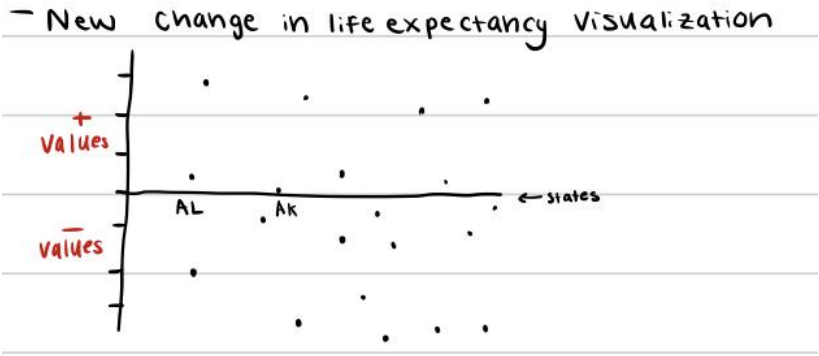
After our meeting, we met up to begin work on coding the line chart in d3. We struggled with appending our visualizations to the website we had previously created, so we worked on each visualization separately and then combined them into one file after they had all been created. We

also found that when we tried to create two separate lines in d3 for the male and female data, an additional third line appeared connecting the two. To fix this error, we had to do additional data processing in the d3 files so that the male and female data were in separate columns and were getting read in separately. This allowed us to append to paths to our visualization, and allowed us to make the color for the female line red and the color for the male line blue.



<u>October 31st, 2023- Work on Visualization 2</u>

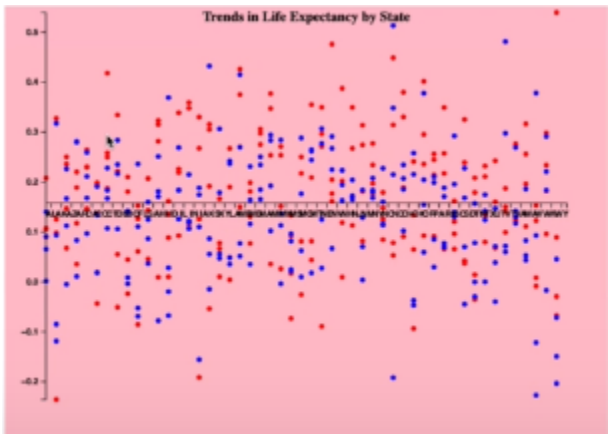
After analyzing our dataset again, we realized that we only had trend in life expectancy data at the state and commuting-zone level. The commuting-zone data was filtered to the top 100 most populated commuting zones, and we felt using this data would limit our visualization. Instead we decided to continue working at the state level so our data was less limited. We also decided on a new visualization idea, instead of using the original matrix idea we had proposed:



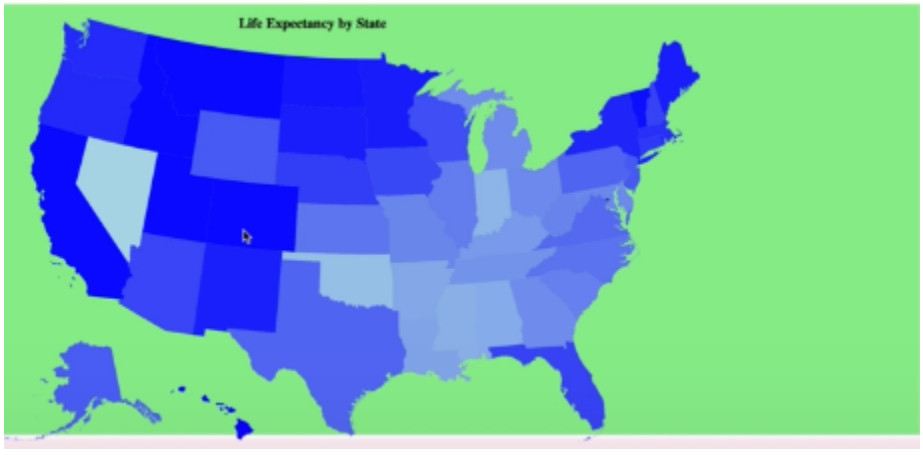
The new visualization would be a scatterplot with state names on the x-axis and the values of the change in life expectancies on the y-axis. Points on the graph would represent quartile data for a specific gender for each state. For example, for one state there would be eight dots, male quartile 1-4 dots and female quartile 1-4 dots.

Pros: We felt that using a scatterplot helped users easily detect where life expectancies were increasing and decreasing. All decreasing values are below the x-axis and all increasing values are above the y-axis. Additionally, we liked that all the quartile data was aligned with the state on

the x-axis so users could visually see which demographic in each state had the lowest and highest changes in life expectancy. This graph also made it easy to view outliers.
Cons: Because we did not have the county or commuting-zone level data for this visualization, we felt that the visualization would be too simple after we filtered the data. Ideally, we would have had male and female quartile 1-4 points for every county or commuting-zone in the country, but at the state level we have fewer points.



November 1st, 2023-November 2nd, 2023- Work on Visualization 3
Based on feedback from Milestone 3, we chose to make changes to our map visualization. We chose to use a US state map, rather than a county or commuting zone map, so elements would be easier for the user to view and click on. Additionally, rather than using a diverging color map we switched to a sequential color map that uses a single hue, just changing the saturation of the color. Darker colors on the map represent areas with higher life expectancies (and vice versa).



November 5th, 2023- Milestone 4
[insert pic of dashboard]

Feedback from Milestone 4:

- Choose the graph placement on the webpage based on how much data it is trying to convey
    - Try switching the map and scatterplot positions so the scatterplot is less cramped, as the map visualization is showing less data
- Make sure the visualizations fit on the display- shouldn't make users scroll down the page to view all the visualizations
- Make sure that colors stay consistent throughout visualization: if the map is blue, that represents male data and should use a different color when the map is showing aggregate data
- Overall, visualizations make sense for the data and work in relation to each other

November 16th, 2023- In-class Feedback on Visualization

Because we switched from viewing county/commuting-zone level data in the scatterplot to viewing state-level data, we were concerned that filtering would make our scatterplot too boring to look at. We were considering switching to a different kind of graph when the scatterplot was completely filtered. For example, the user would choose the gender and quartile they wanted to view using the line chart. This would leave only the points on the scatterplot corresponding to that gender and quartile. Then using the map, the user would choose what state they would like to view. This would leave only one dot remaining on the scatterplot. At that point, we were considering a transition to a line chart that showed this dot's value in a different way. Because the dots represent the change in life expectancy over a period of time, we were considering moving to a line chart with years on the x-axis and the life expectancy on the y-axis. The slope of the line would be the same value as the dot on the scatterplot. We went to the in-class feedback sessions to get input on this concern/idea:

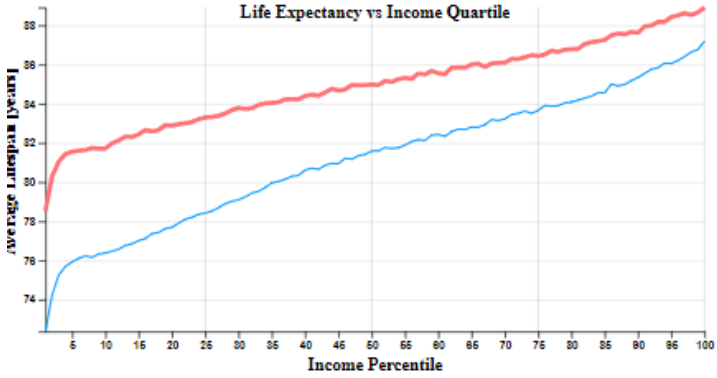
- Instructor feedback:
    - Transitioning to a new visualization is not necessary
    - Don't have to completely remove points from the scatterplot when they're filtered out
    - Can decrease the opacity of points that weren't selected or increase the size of the points that were so that they stand out more
    - Can also try a different way of sorting data in scatterplot other than alphabetical (ex. Highest to lowest changes, etc.)
    - For final visualization, background should be either black or white
- Class feedback
    - Need legend to indicate whether higher or lower saturation on the graph represents higher or lower life expectancy
    - Add outlines around the states so that they're more visible
    - Maybe increase the outline around a state after it has been clicked to maintain a reference for the user

- The dots on the scatterplot are not aligned with the x-axis labels

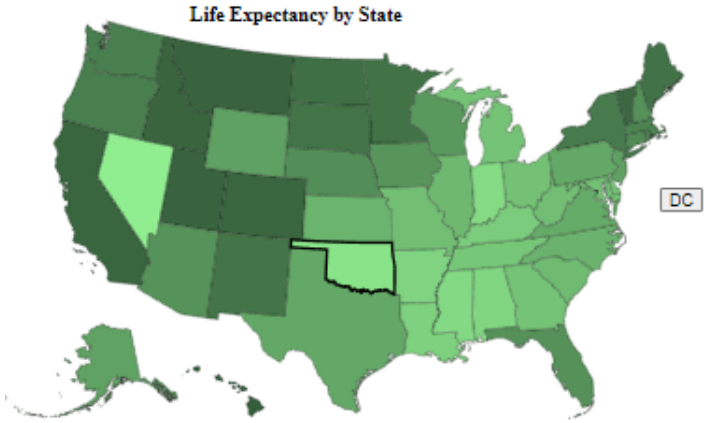
November 20th, 2023- Hover interactions

We wanted to implement hover features to indicated to the user that an element could be clicked on. These hover features would be available for both lines on the line chart, each of the states on the map, and each point on the scatterplot

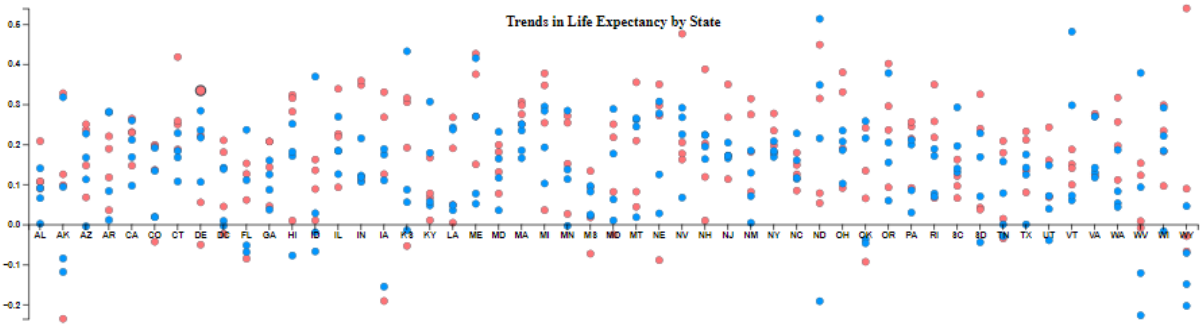
- Line chart- the line width of the male or female line increases when it is being hovered over



- State map- the outline of the state being hovered over increases in thickness



- Scatterplot- each dot increases in radius when it is hovered over. The stroke-width of the dot also becomes bolder



These changes disappear when an element is no longer being moused over.

November 28th, 2023- Meeting with Dr. Iuricich about Implementing Interactions

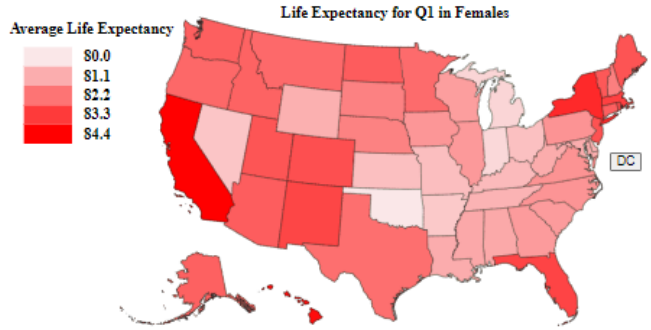
Concerns: We have no idea how to make visualizations interact with each other
Meeting Notes:

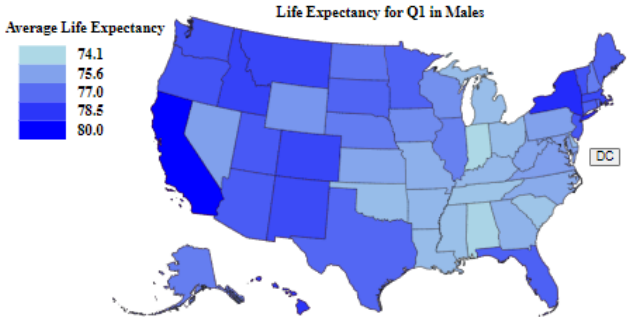
- Use d3.select() and d3.selectAll()
- Have hook, id, or class to select elements that you want to change
- Add attribute to elements when you draw them using .attr(“id”, “point”)
- Then, use d3.selectall(“#point”) to filter them later
- Might need boolean to determine what has already been selected/clicked
- Can also use .attr(“class”, “...”) for class of items
- Can use g to group elements together
- Create .on(‘click’, function(d){}) and write the name of function, then define the function outside

November 29th, 2023- Line chart filtering map interaction

Goal: To have the user click on either the male or female line and then have map data switch to being data for that gender. The x-coordinate of where the user clicks on the line will determine the quartile that is shown on the map.

- Made color scale of map change to red if female data was chosen or blue if male data was chosen. These colors correspond to the colors of the line in the line chart.
- Color scale changes for each quartile and gender chosen. This enables comparison of state life expectancies within a quartile so users can determine which states have worse health inequality (worse health inequality = lower life expectancy for lower incomes).

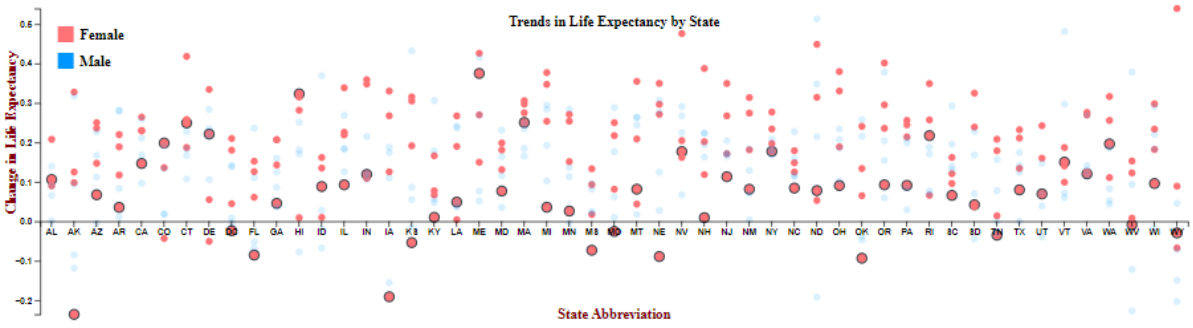




November 30th, 2023- Line chart filtering scatterplot interaction

Goal: To have the user click on either the male or female line and then have the scatterplot data filter depending on the gender chosen. The x-coordinate of where the user clicks on the line determines the quartile that is shown on the scatterplot.

- To draw the user's eyes to the points that were selected, we increased the radius and stroke width of the points. This design choice enables users to easily see which states have increasing or decreasing life expectancies for the specified gender and quartile.
- We chose to decrease the opacity of points that were not chosen instead of deleting them off the chart- this allows the user to still see points in other quartiles and enable comparisons between the chosen quartile and the other quartiles and gender.



December 1st, 2023- Meeting with Dr. Iuricich and independent work

After working on our interactions further, we had additional questions about the correct way to visualize certain data after filtering.

- How should we fix overlapping points in the scatterplot?
    - Concerns: certain points on the scatterplot overlap, which makes them hard to read
    - Feedback:
        - Not concerned about overlapping points, just go stronger on the opacity of other points that were not selected to make the plot more readable when someone clicks on it
        - Should also keep selected circles bigger so they stand out
- How do we correct hovering bugs in the scatterplot?
    - Concerns: Every time we hover over a point in the scatterplot that has been selected through interactions with the line graph, the point becomes unselected when we mouse away due to mouseover features we implemented
    - Feedback:
        - Best way is to create a global boolean variable to keep track of clicks
        - Start as always false but turns to true when clicked
        - Could also do a d3.select(item) and read the attributes of the item to see if the circle has a radius of 6, then decide not to apply hovering changes based on that (may be more difficult)
- How do we revert data back to normal if it is no longer selected?
    - Concerns: having trouble using selectAll to revert previously selected points back to normal after a new dataset is chosen
    - Feedback:
        - Way have to create separate selectAll()s for each attribute id we are trying to change
- What color scale should we use for the map?
    - Concerns: not sure if our current aggregate color scale using the color green has enough contrast, also not sure if the color scale should change for each quartile selected or if we should use a universal scale that stays the same regardless of what quartile is shown, also not sure if we should include buttons on visualization to show aggregate data for male and females
    - Feedback:
        - The green color scale is good for aggregate data, has enough contrast
        - Need a chart legend to show the color steps of the chart
        - Could use a universal range, the only problem is because life expectancies in the quartiles have similar values, the map becomes all one color step when a specific quartile is chosen-doesn't really show how location impacts life expectancy

        - Keep separate ranges for each quartile- point of visualization is to show how location impacts life expectancy within a quartile (ie does a certain location have a worse life expectancy for Q1 but a really high life expectancy for Q4?), easier to achieve with a scale that changes based on quartile
        - Not necessary to have aggregate male and female buttons that don't filter by income quartile because point of project is to specifically show the impact of income, not gender, on life expectancy
- How can we click on states in the map to create filtering in the scatterplot?
    - Concerns: haven't been able to click on specific states to filter scatterplot data
    - Feedback:
        - Use a reference id for specific states
        - Add the attr() after the append path so it's called once per state
        - Could read in the name of the state or the number of the state, which could then be related to the position of states in the scatterplot data's array
- General comments:
    - Look into z order function to move graphical elements forward when they are selected
    - Add a text or title to the map that changes based on the gender and quartile that was selected
    - Could add vertical lines on the line chart to indicate where each quartile begins and ends (make very light grey, like horizontal lines)

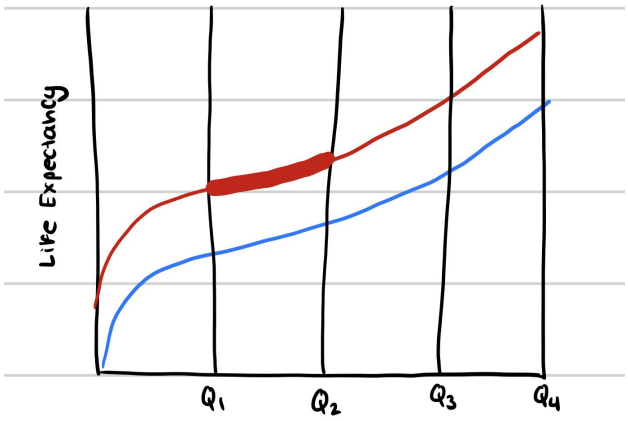
December 3rd, 2023- Map filtering scatterplot interactions

Goal: When a user clicks on a state on the map, the data not associated with this state will be filtered out or hidden/decrease in opacity so that the data for the selected state stands out. Ideally, the user would be able to click on multiple states to be able to compare their data to each other.

December 4th, 2023- Scatterplot interacting with line chart and map

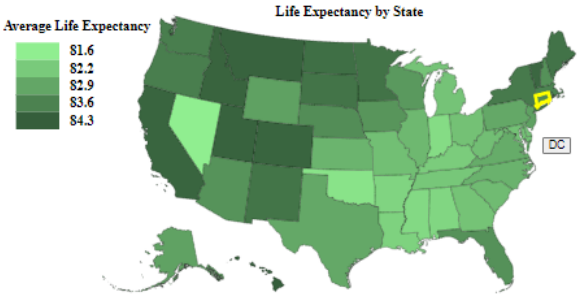
Goal: We wanted the scatterplot to interact with the other two visualizations instead of just being filtered by them. We wanted the user to be able to click on a point in the scatterplot and see what line and state it was associated with.

- When the user clicks a point, we want just the section of the male or female line corresponding with that point to become bolder (increase its stroke-width). For example, if a Q1 Male point was chosen, only the section of the male line that counts as Q1 would become bolder.



What would ideally happen if a quartile 2 female point was chosen-the quartile 2 section of the female line would become bolder

- When the user clicks a point, we want the corresponding state to be highlighted on the map to help the user see the geographic location of the state.

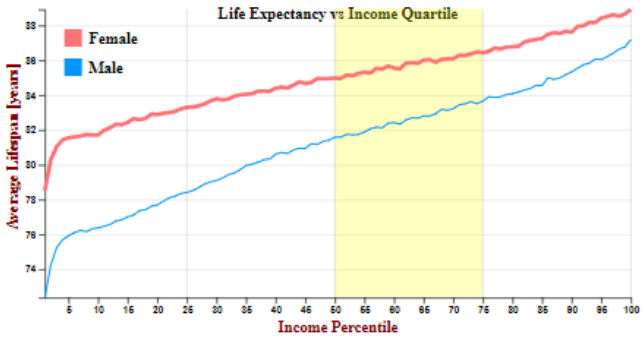


Result of clicking on a scatterplot point associated with Connecticut: This interaction is currently only working for the aggregated data map, not the filtered data map but ideally, the same thing would occur after filtering.

<u>December 5th, 2023- Oral presentation</u>

After giving our presentation in class, we received the following feedback:

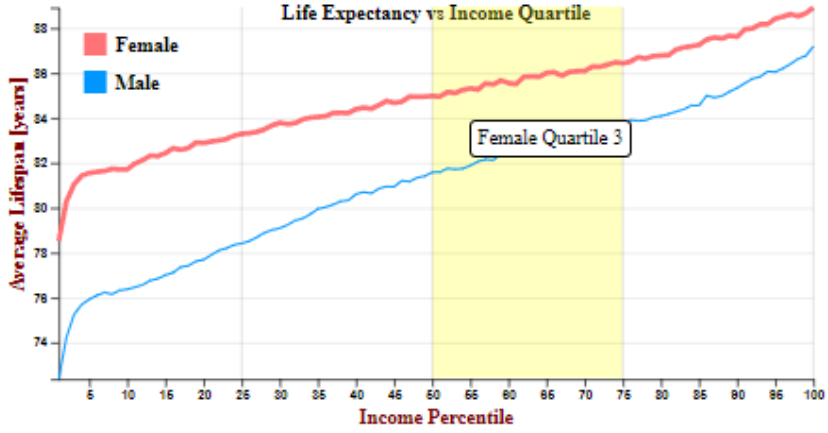
- For the final website walkthrough, put more focus on the patterns that our visualizations reveal- for example, does one state always have the highest life expectancy?
- Also, point out outliers specifically in the data.
- To highlight one section of the lines on the line chart, we can try to just highlight the background rather than the actual line



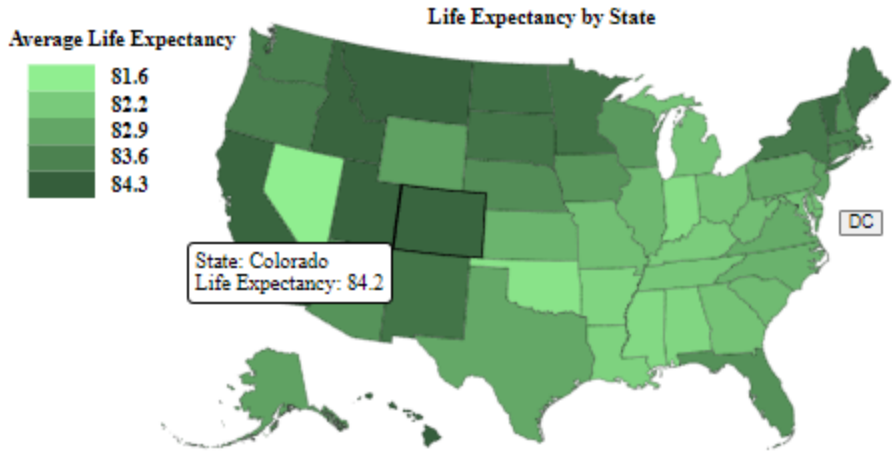
December 6th, 2023- Tooltips

Goal: To have tooltips on the line chart, map, and scatterplot that appear and give the user more information when they hover over a certain element.

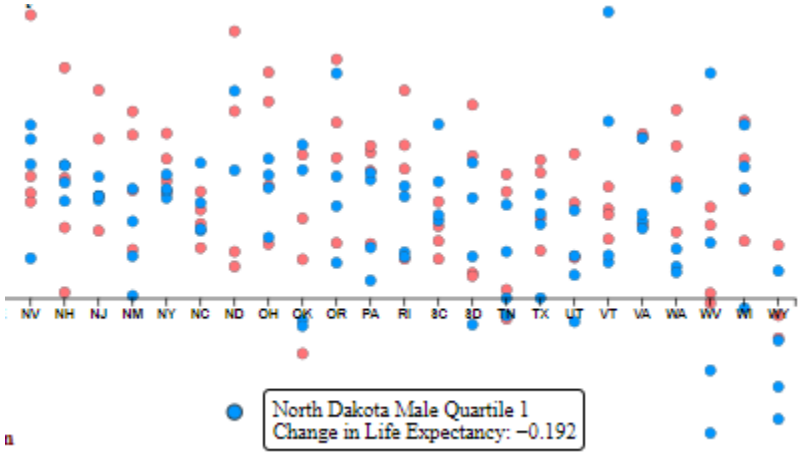
- Line chart: When the user tracks their mouse up the line, a tooltip will appear showing the life expectancy for each income percentile.
    - This didn't end up working because the method we used to track the mouse along the line took away the ability for the user to click on the line, which we felt was a more important interaction than the tooltip. Instead, we chose to just have a tooltip that displayed the gender and quartile associated with each section of the lines.



- Map: When the user hovers over a state, the state name and average life expectancy for that state will be displayed. This should work before and after filtering the data.

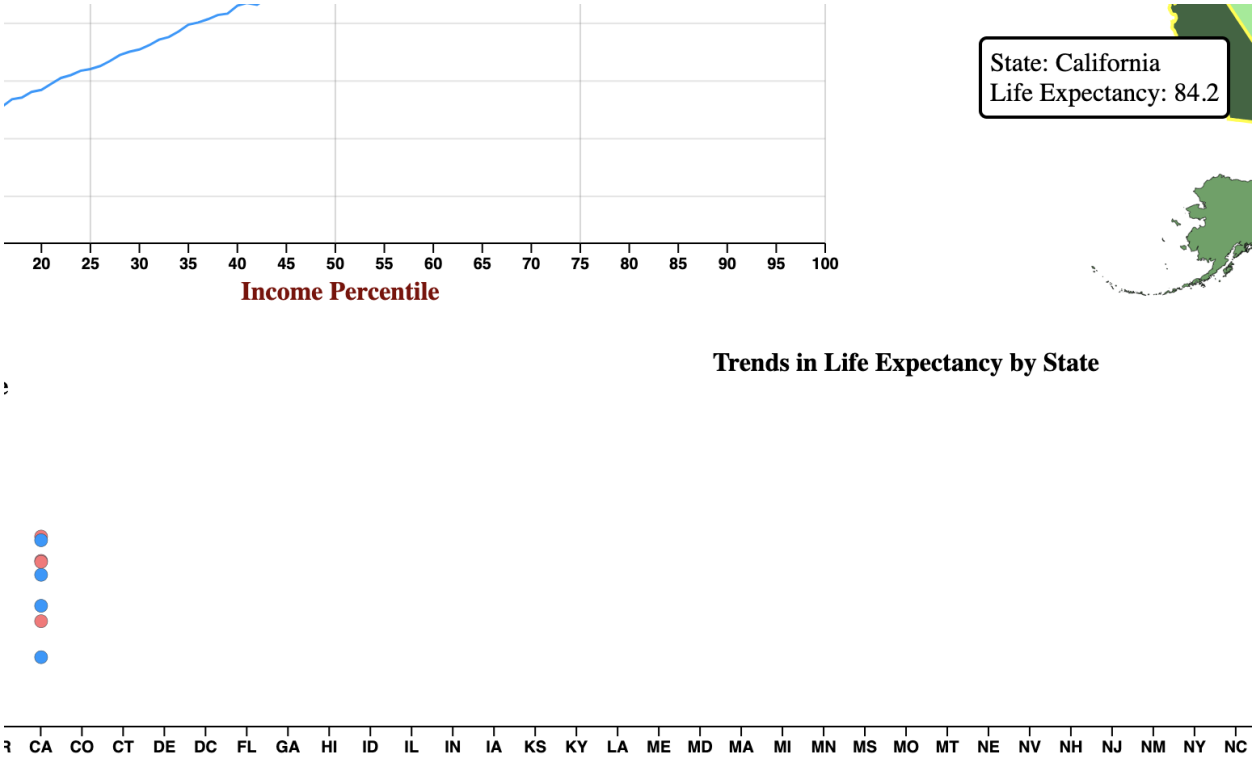


- Scatterplot: When the user hovers over a point, the gender, quartile, and change in life expectancy will be displayed.



December 7th, 2023- State filtering in scatterplot

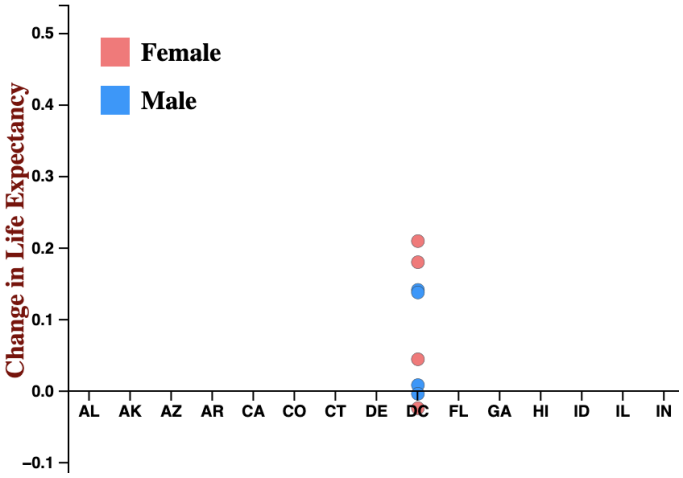
One issue we were having throughout the project is filtering the scatterplot when clicking on a state. One problem we had for a while is that whenever we click on a state, it would work fine, however if you click on another state, all the dots would disappear. I think this is because initially, we were using the methods .exit() and .remove(). We couldn't figure this out for a while, so we messaged our professor and he told us another way we can do it was to change opacity to 0 and then when something else is clicked, change the opacity back to 1. We finally got it to work that way.



In addition, we also messaged our professor to get some final feedback for our project. We went ahead and recorded a video explaining all our interactions, what we were having trouble with, and if there was anything he thought we needed to add. Once we receive the feedback, we want to incorporate it into our project.

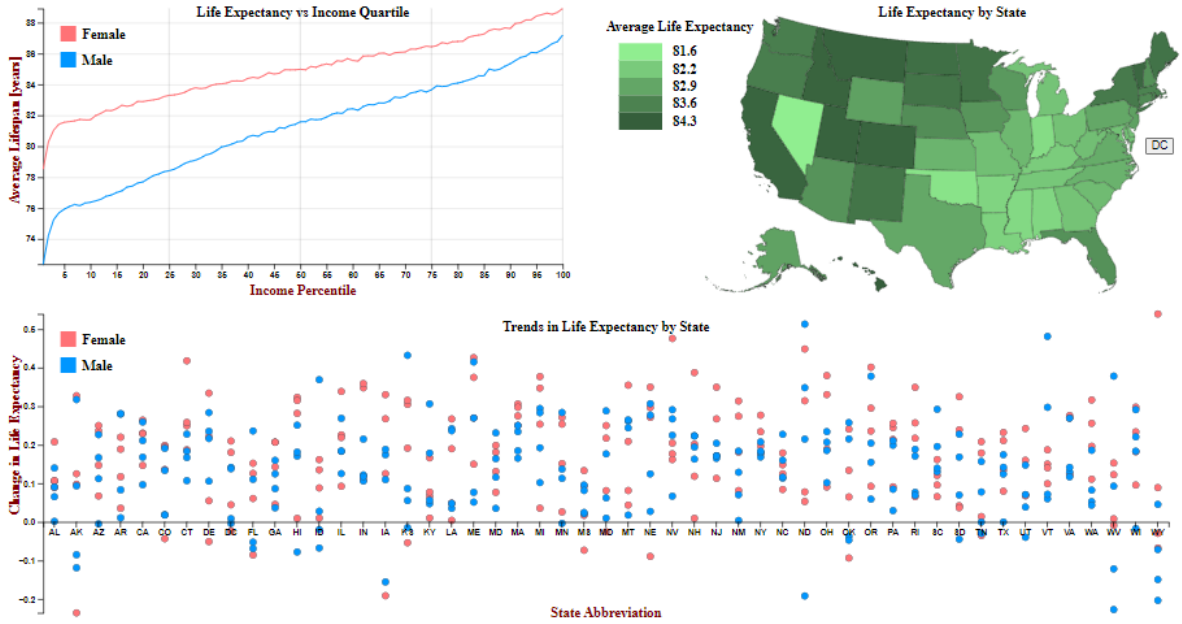
December 8th and 9th, 2023- Final touches

In these two days, we basically worked on making the final touches for our projects and making any last minute updates if needed. We are still waiting for feedback from our professor, so if he has any suggestions we want to use them and make more updates to our project. Some of the final changes we made is including the DC button and making sure that works, and working out minor bugs. An issue we are having trouble with is highlighting the associated state when clicking on a line and then clicking on a scatterplot point. The cause behind this issue is that we are generating a new map every time, so the highlighting isn't properly working because the map sits on top of the highlight (we know this because if a border state is highlighted, we can see a bit of the highlight popping out of the map). *pictured below is the filtering for DC*



December 10th, 2023- Final Project Screen-Cast

Our goal today was to record our final 5 minute presentation and make final touches and submit the project! For our final touches, we finally figured out our highlight issue. We just had to read in the actual map instead of the filtered map, so that the highlight will show up on the map. We then zoomed and rehearsed our presentation 2 times, recorded the video, and submitted it.



Evaluation/What we learned:

By working with our data, we were able to answer our original three questions:

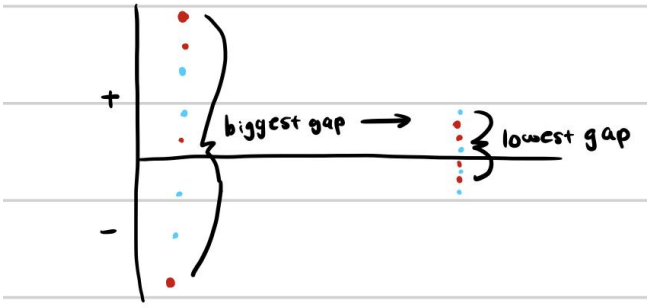
- As income percentile increases, so does life expectancy.

- In general, female life expectancies are higher than male life expectancies for every income quartile. As income increases, the gap between gender life expectancies decreases.
- States in the south and midwest generally had lower life expectancies for people of every income quartile. States in the northeast and west tended to have higher life expectancies. Nevada was an outlier, as it consistently had one of the lowest life expectancies for every quartile.
- When viewing lower quartiles, the difference between life expectancies for different locations in the US ranged by about 4 years for women and 6 years for men. As income increased, this gap decreased to about 2 years, so as income quartile increased location played less of a role in determining life expectancy.
- In most states, life expectancies are increasing for people of all income quartiles. If a state did have a decreasing life expectancy for a demographic, it was generally for people belonging to quartiles 1 and 2. An exception to this trend was Hawaii, where quartile 4 men had the biggest annual decrease in life expectancy and quartile 1 women had the biggest increase in life expectancy.

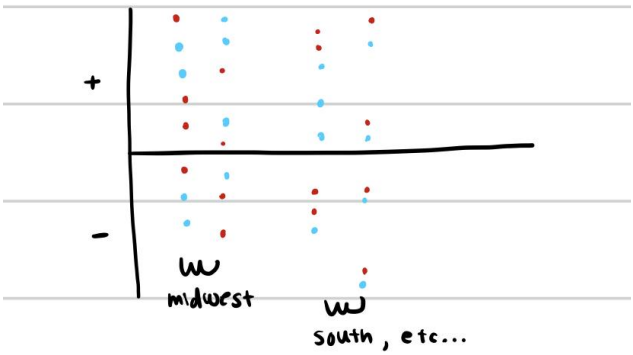
Further improvements:

Because neither of us had used d3 before, our visualizations were limited by our coding abilities. We have several additional features we would like to add in the future after we become more familiar with d3:

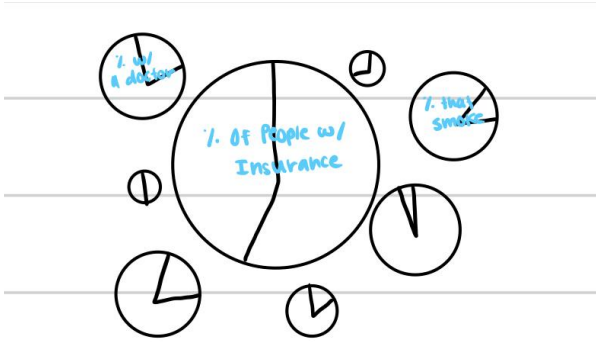
- Implement a sorting mechanism besides alphabetical for the scatterplot. Potential sorting mechanism could include:
    - Sorting data in ascending or descending order from states with the lowest differences in the change in life expectancy between quartiles to states with the highest differences in the changes in life expectancy between quartiles- this would help show which states have more income inequality than others



- Grouping states in the chart based on geographic location (like southern, midwestern, etc.)- this would help users notice geographic trends in the changes in life expectancy



- Add multiple states to the scatterplot after filtering. Currently, when you click on a state on the map, only data for that state remains in the scatterplot. Then, if you click another state, the data from the original state disappears, and the new state's data appears. This helps to filter the data, but we thought maybe in the future, if a user clicked on multiple states, multiple states' data would appear on the scatterplot. This could help users compare changes in life expectancy for a few states they selected.
- Including revert buttons that, when pressed, would return a visualization back to its original state. We did not include this feature originally because it could be achieved with the refresh button but including a button to refresh one visualization at a time (for example, just the map or scatterplot) would be useful.
- In our early design phase, we also considered having a visualization that showed how certain factors that impact life expectancy (like access to hospitals, having health insurance, etc.) change with income. We still think this would be an interesting set of data to visualize and we could visualize a selected group of attributes for each state and the percentage of people in those states that have that attribute (like % with access to a doctor). This visualization would change depending on what state was clicked on, and the size of the circles would change based on which bubble the user hovered over.


% of People w/ Insurance