Diverting Droughts with Data: Designing a Visualization Tool to Drive Water Conservation Efforts Across Massachusetts

Jethro R. Lee Michelle Wang Yuting Zheng

ABSTRACT

Our visualizations serve to support three domain tasks. We aim to help users determine the relationship between precipitation values and drought levels among region(s) of interest. For instance, if users anticipate the possibility of a drought given the precipitation levels reported in their area, they can take early action in conserving water to prevent being harmed by an increase in drought level severity, having precipitation totals lower than normal. Our second task is to help users determine which regions tend to have more/less precipitation than other region(s) during one or multiple months across one or multiple years. This ability will hopefully encourage users to prioritize conservation efforts during certain months in places where the precipitation is lower than others during the same month. Our third task is to aid users in determining if SPI levels correspond with annual precipitation for each region in a given year. This task allows users to be better informed about how changes in annual precipitation and SPI drought levels are related and take appropriate action when they hear about a possible decrease in the annual precipitation total for their region that reflects the annual precipitation of a previous year with many monthly precipitation totals lower than normal, for instance.

1 INTRODUCTION

The first task this visualization tool supports is allowing users to determine the relationship between rainfall values and drought levels among region(s). They can accomplish this task by using a scatter chart that plots a region's precipitation and corresponding SPI value over certain month(s) and year(s). SPI, or the number of standard precipitation totals from normal precipitation amounts, is marked across the x-axis, whereas precipitation is marked along the y-axis in terms of inches. The data on the scatter plot can be filtered by month and year. The user can first identify regions they want to be represented on the scatter plot with the help of a geographical map that shows the positions of cities in Massachusetts. Markers are spread across the map to indicate a county that a group of cities belongs to. The color of each marker corresponds to a drought region in Massachusetts. Thus, the map enables users a convenient way of identifying the drought region a city belongs to. From there, the user can filter the scatter plot by the regions that interest them. The user can also see more specific data about the precipitation and SPI value that a region experienced at a certain time by hovering over points that interest them. The user should be able to see that SPI and precipitation are positively correlated.

Therefore, this task is important to support because our plot confirms that forecasting precipitation patterns is a justifiable way to prevent drought occurrences across Massachusetts. If users anticipate the possibility of a drought given the precipitation levels reported in their area, they can take early action in conserving water to prevent being harmed by an increase in drought level severity. Precipitation predictions are very accessible to the public, so strengthening people's knowledge of the relationship between precipitation levels and drought severity can help raise individuals' awareness of whether a region they live in is prone to having droughts, requiring water to be conserved more thoroughly.

The second task supported by this visualization tool is allowing users to determine which regions tend to have more or less precipitation than other regions at certain times. This task is supported by two of our visualizations. The segmented line chart allows users to compare precipitation totals for Massachusetts regions across all months in one given year. The scatter plot allows users to compare precipitation totals for a region(s) across as many months and as many years (2017-2019) as they wish. The regions can be filtered based on a user's preferences, which can derive from the same strategies a user employs when deciding which regions to be symbolized on the previously mentioned scatter plot for domain task one. The line chart will emphasize whether a region consistently experiences lower precipitation amounts than other regions at certain months. For instance, if a point is noticeably a lot lower compared to points representing other regions where x = June during multiple years, then people should prioritize conservation efforts during June in that region. People should generally focus their conservation efforts during certain months in places where the precipitation is lower than others during the same month.

Furthermore, the scatter plot can help users identify which precipitation totals for a month tend to be considered lesser than normal (when SPI is negative). Suppose across some regions and some years, a user sees that regions experience lower precipitation levels and lower SPI levels (higher drought severity) consistently in July. Then, users might therefore focus on taking action in July in the future because our visualization suggests that across many years and months, many regions have low precipitation and negative SPI levels in July.

This task is also important to support since many people are unsure of when they should be more conscious of the amount of water they are using. Since that instance is the case, people just generally use water to their own liking all the time. If users were notified of the specific times when their region is at high risk of low precipitation amounts, then they can be better prepared to live through periods of high water scarcity and take appropriate measures beforehand. Perhaps regions in Massachusetts that experience higher precipitation totals during a certain month can also assist nearby regions with lower precipitation totals at that time.

The third task this visualization supports is enabling users to determine if a region having a higher or lower SPI corresponds with that region having higher or lower annual precipitation totals. They can explore this relationship with a linked bar plot and scatter plot. The bar plots display the total annual precipitation for regions in a certain year. The data can be filtered by the regions that interest the user as well as the year represented (2017, 2018, or 2019). The scatter plot linked to this bar chart is essentially a replica of the previously mentioned scatter plot with its associated features (e.g., tooltip). However, this chart also contains a dashed line where SPI = 0 to help users determine if there is a relationship between the distribution in the number of points to the right and left of where SPI = 0 and annual precipitation. If so, the plots further justify how precipitation is a good indicator of drought severity.

More importantly, the users can see whether a region has been experiencing more monthly precipitation totals lower than normal over time through the linked charts. For instance, from 2017 to 2018 to 2019, if the number of points to the left of where SPI = 0 (negative SPI values) increases for a certain region, then this finding should indicate to the user that the region has been experiencing abnormally low precipitation values for an extended period of time. The user can also take note of which annual precipitation totals have historically been correlated with more negative SPI values per month.

Therefore, the task is critical to support as it enables users to caution people in that region about the potential drought severity they could experience in the future and encourage the implementation of water conservation efforts in that region. The general public likely lacks the knowledge necessary to infer whether where they live is at risk of having a drought in the near future. Even though they can passively listen to meteorologists, our visualizations provide concrete evidence that certain regions are prone to experiencing droughts through digestible charts. Therefore, one can present the tool's elements to others to help them see why water conservation is critical in Massachusetts at certain times. By providing people with a way to understand the importance of saving water in their region in an intuitive manner, more people are likely to conserve water as they are not presented with confusing and intimidating information that could detract from their motivation.

To reiterate, the end user of these visualizations are part of the general public of Massachusetts. The data that is painted through these visualizations are precipitation and SPI data for all the regions in Massachusetts across time. Looking at both precipitation and SPI data together allows the user to draw important conclusions about the relationship between the two to ensure where they live is prepared for the next drought.

2 Related Work

2.1 Related Work #1

"Visualization of High-Resolution Weather Model Data" discusses a weather prediction system, WRF, that is optimized for meteorological simulations [2]. The researchers attempted to create a model with high-resolution severe weather simulations and visualizations for the area close to the O'Hare International Airport. Their model encompasses an area of around 224 kilometers and a height of at least 21 kilometers. They obtained their meteorological data for their visualization from the National Centers for Environmental Prediction. They used TACC's Stampede supercomputer to create the most high-resolution simulation possible with efficient data processing and visualization work. The developers were only interested in modeling a particular set of two-dimensional and three-dimensional attributes, such as radar reflectivity, cloud water content, cloud ice content, snow content, and graupel content. They developed a custom reader to limit the data read from disk and used a tool called ParaView to incorporate their visualizations with a lot of detail via isosurfaces, clipping, slicing, and streamlines. Regarding the data itself, they modeled rainwater mixing ratios as volumes, used colors to represent different levels of rain values, symbolized wind velocity as streamlines and differentiated magnitude by color, and more. The researchers were inspired by other studies and experiments that used WRF simulations extreme in resolution or scaled on distinct platforms. The researchers were able to generate a model that projected data from specific places and time frames infused with high resolution. They believe the model they made exceeds the typical time and space constraints of common weather simulations. Due to the immense amount of detail they can project in their model, they believe they can help meteorologists notice even the slightest changes in weather in local areas.

A visualization technique employed by the researchers was focusing on specific areas in the target area of interest. The target area of interest formed a cylindrical area around the O'Hare International Airport with the dimensions of 1345 x 1345 x 234 grid cells. Despite this expansive area, the meteorologists emphasized the importance of getting high-resolution data. Since the researchers were able to closely observe weather data on the visualization's geographical map, they could incorporate detailed figures in their paper showing scenes of rainwater mixing ratios as well as wind and reflectivity. The paper discussing WRF emphasizes the importance of highlighting subtle changes in weather patterns to make the best meteorological conclusions possible [2]. Hence, perhaps people with advanced experience in analyzing precipitation data could conceive compelling insights with our visualization if they were allowed to inspect the depicted drought patterns very closely. For this reason, we believe it will be beneficial for our map of Massachusetts to be a focus and context visualization. So, we will allow users to zoom across our map of Massachusetts to better understand what region they live in. With this in mind, users can use our other two visualizations to better understand more information about precipitation and drought levels in just specific areas of interest (such as the region they live in) instead of broad information about Massachusetts in general.

2.2 Related Work #2

"An Exploration Framework to Identify and Track Movement of Cloud Systems" discusses a framework that enables users to observe and track the movement of cloud systems via computational topology and computer vision [1]. A user of the framework can investigate cloud movement at different scales in time and space. The researchers claim their work is friendly to any users with any expertise in cloud clusters, unlike other similar frameworks. They also wanted to address how methods that monitor the movement of cloud clusters that already exist don't capture the local movement that occurs within certain systems. Furthermore, those existing methods tend to require users to specify additional parameters, yet the absence or presence of parameters provided can affect the movement projected and the consistency of the results. The researchers also wanted to make sure their tool could track cloud movements at various thresholds since each threshold presents different findings about what is being studied. To identify clouds, the researchers relied on two scalar functions: infrared brightness temperature and precipitation. Different temperatures are associated with different clouds: colder temperatures are linked with tall/Cumulonimbus clouds, while higher temperatures are correlated with short/Stratus clouds. As for precipitation, higher values likely indicate tall clouds since Cumulonimbus clouds create more rain than Stratus clouds. The movement of the clouds themselves is monitored with the optical flow between pairs of IR brightness temperature images. The researchers used an implementation of an algorithm that allowed them to depict intra-cloud system movements that, as mentioned above, existing cloud tracking systems cannot capture. The algorithm also allowed the developers to capture the smooth local motion of clouds for users to query and see interesting trends regarding the way the clouds move. The researchers concluded based on their visualization that the majority of tropical convective processes are multi-scale occurrences integrated with smaller-scale cloud systems. Their cloud motion visualization is currently only used to support queries, but the researchers hope to add extensibility by expanding it to be also used as a user interface that interacts with the visualization.

The visualization discussed in the paper also centers around weather data and stresses the importance of tracking the weather over a particular region over a period of time to draw compelling meteorological conclusions [1]. Particularly, the cloud movement tool shows the smooth local motion of clouds through a velocity field. The velocity field is derived from the output of the algorithm mentioned above to construct a cloud motion graph. This graph maps clouds from one time step to clouds in the next time step, showing how clouds move and interact over time. Basically, the graph shows how clouds are present when they are first formed, followed by when they combine and break apart from other clouds, and finally when they dissipate. Accordingly, we want to emphasize how weather data changes over time within our visualization, particularly in our map of Massachusetts. We will use year/month checkboxes as well as year drop-downs. This will allow users to see a visual geographic depiction of how drought levels have changed over time in Massachusetts.

3 USE CASE

Given the current Level 2 Drought in some areas of Massachusetts, the visualization is intended to make the public and citizens of Massachusetts more aware of this drought issue. Clearly, the state is already keeping watch on it and staying aware of it through the Massachusetts Drought Management Plan. However, the citizens should also be informed and aware so they can be encouraged to focus on water conservation.

The visualization will not be too complex as the public is not familiar with using lots of tooltips and other visualization features. We want the people in Massachusetts to conserve more water. They can use our visualizations to become more motivated as to exactly what state of water shortage specific regions in Massachusetts are in. For instance, suppose anyone in the general public wants to encourage people in Massachusetts to conserve more water. The person can hold a meeting with individuals from Boston to push more water conservation efforts. The person can give a presentation during this meeting and use our visualization tool to strengthen their claims. One of the visualizations can be a line chart easily digestible to the public presenting the average amount of rainfall each region in Massachusetts experiences in a certain month over several years. The person can take advantage of filters to highlight certain subsets of the rainfall data, such as data from a particular region(s) and month(s). The user can then bring attention to, for instance, certain regions that have relatively less rainfall than other regions to assert that they need to be prioritized with water conservation efforts.

4 DATA

4.1 Data Selection

The precipitation database data is made available to the public on the Mass.gov website, which is owned by the Commonwealth of Massachusetts [3]. The database that holds the data in the file has been monitored by the Massachusetts Department of Conservation and Recreation's (DCR) Office of Water Resources. The website states that the precipitation data itself derives from observers and cooperative agencies.

The standardized precipitation index (SPI) data is on the same website as the precipitation database data [4]. The standardized precipitation index data is calculated monthly from values in the precipitation database data. Hence, it relies on data collected from the same sources as the precipitation database data.

4.2 Biases and Ethical Considerations

The precipitation database data, especially since it's derived from a government agency, overall seems reliable. However, there are some minor possible issues with how the data was collected. The Mass.gov website says about 60 precipitation observation stations collected the precipitation data, but there are no details on the identity and legitimacy of the observers themselves. Hence, it's unclear whether the stations are reliable and provide accurate data. Furthermore, each station across the Massachusetts cities may be using different measurement tools with varying levels of precision and accuracy. Therefore, we cannot conclude whether the data is standardized. The cooperative agencies that are also referred to on the Mass.gov website have unknown identities. Thus, like the ambiguous observers, the precision, accuracy, and standardization of the agencies' precipitation data cannot be verified.

Furthermore, the precipitation measurements are rounded to two decimal places on the public dataset. There is a possibility that this level of precision may not allow us to make the most effective visualizations possible. The rounding restricts us from presenting the most detailed variation in precipitation patterns across Massachusetts (for example, a value of 2.883 inches of precipitation would be visualized the same way as 2.881 inches).

Also, the measurement tools, observation stations, and cooperative agencies may have evolved, along with the precision and accuracy of the precipitation data. For instance, the cooperative agencies could have relied on tools in 1900 that were not as competent at collecting precipitation data as those used in 2015. Therefore, there's a chance that precipitation data from older years such as 1900 may not be as accurate as that from more recent years such as 2015.

Finally, each region in Massachusetts does not have the same amount of precipitation measurements in the data set. Some have more measurements than others, so our visualization could have more precise information about the precipitation pattern of regions in Massachusetts represented by more rows in the data set. There is a substantial amount of data for each region, but this consideration is still important to keep in mind.

These same biases and ethical considerations apply to the SPI data since it derives from the precipitation database data. We will be relying on 2017, 2018, and 2019 SPI data since they are the only ones available that most directly correlate with the precipitation database data we are using, which was generated in October 2019. The 2017-2019 SPI data might not contain the most representative values indicating the drought severity of a Massachusetts region. There might have been some weather anomalies throughout 2017-2019 that may have caused a region in Massachusetts to experience more or less precipitation during those years than usual. However, we are assuming the tools used to collect values for the precipitation database have improved over time. Hence, we think the 2017-2019 data is reliable since it was gathered only 4-6 years ago.

4.3 Data Cleaning

Our visualizations containing data only on precipitation rely on these attributes from the original precipitation database data set: "YEAR" (the year when the corresponding precipitation data was collected), "JAN," "FEB," "MAR," "APR," "MAY," "JUN," "JUL," "AUG," "SEP," "OCT," "NOV," "DEC," (each of which indicates the amount of precipitation a region experienced in the month corresponding to the column label), and "Region" (the region in Massachusetts where a city is located in). Before doing any reformatting of the data, we dealt with missing values. We assumed that the missing values indicated data that could not be collected and not that there was no precipitation during a month. We first inspected the number of null values for each column as well as for each row. We realized that for all rows, there was less than half of the data missing. Therefore, we performed imputation for all the month columns, which represent precipitation for a certain month, on the remaining missing cells. We replaced any missing values with the median of all the values per column, taking into account any potential outliers.

We went on to reformat the data set into a new one called precipitation_cleaned.csv so that it would be easier to manipulate when we made the visualizations using D3. We kept the "YEAR" column and the "Region" column. However, we made a new "Month" column and "Precipitation" column, which uses the values in "JAN," "FEB," "MAR," "APR," "MAY," "JUN," "JUL," "AUG," "SEP," "OCT," "NOV," and "DEC" from the original dataset. The "Month" column indicates the month a precipitation value corresponds to and the "Precipitation" column details the quantitative precipitation value (in inches) corresponding to the month, year, and region that share the same row.

Also, we want to make a note about the meaning of our "Precipitation" column in our final data sets. We decided to calculate the average amount of precipitation across all the cities in each region per month by using the groupby function in Python and inserted those values in a new data set. There are 159 unique cities in the original precipitation database data set. Our visualization would have been too hectic if we tried to present the annual precipitation data for each city. Hence, we decided to focus on the precipitation data per region rather than per city.

Next, we created a data set that combines information from both the original precipitation database as well as the SPI data (combined_prep_spi.csv). The SPI data on the website we retrieved it from were in tables in PDF format. There were tables available for each month (January-December) and the years 2017 to 2019. We read them in as data frames with Python and merged them into separate CSV files for each Massachusetts drought region. Then, we merged the files for each region into one huge data frame and removed the "number of sites" column since it is not important to us in our analysis. At that point, we had a data frame with the following columns: "Year," "Drought Region," "Month," "3-month SPI," "6-month SPI," and "12-month SPI." From there, we decided to remove the "6-month SPI" and "12-month SPI" columns since we initially just wanted to show 3-month SPI values to the users (however, in our future work, we explain our intent to expand this). We used the remaining 4 columns in combination with two other columns derived from the precipitation database data ("Precipitation" and "Annual Precipitation") to create the new data set. The "Precipitation" column was generated by using groupby, as discussed in the previous paragraph. For each value in the "Precipitation" column, the value for "Year" and "Month" in the same row are the same values for "YEAR" and "Month" in the row it originally came from in "precipitation_cleaned.csv." The Annual Precipitation column is also a derived attribute. First, the data in combined_prep_spi.csv was copied and grouped by drought region and year. Then, the average precipitation values corresponding to each distinct year and drought region combination was calculated. Finally, those values were manually copied and pasted into combined_prep_spi.csv. The values were pasted in rows that included the year and drought region combination they represented.

5 Data Abstraction

1. Precipitation data (precipitation_cleaned.csv):

   - Each item of the data set indicates a year, region, month, and average amount of precipitation the region has across all its cities during that month and year.

     YEAR: Ordered (Sequential, Quantitative) (indicates the year when a row of precipitation data was collected)

     Region: Categorical (indicates the region represented by a row of precipitation data)

     Month: Ordered (Ordinal, Cyclical) (indicates the month when a row of precipitation data was collected)

     Precipitation: Ordered (Sequential, Quantitative) (indicates the average amount of precipitation, in inches, collected across a certain region's cities)

2. Drought and Precipitation data (combined_prep_spi.csv):

   - Each item of this data set indicates a year, drought region, month, corresponding 3-month look-back of the number of standard deviations from normal precipitation totals (3-month SPI) the region has that month and year, precipitation value of that region during the corresponding year and month, and annual precipitation of the region during that year.

     Year: Ordered (Sequential, Quantitative) (indicates the year when a row of SPI data was collected)

     Drought Region: Categorical (indicates the region in which the 3-month SPI, precipitation, and annual precipitation were monitored)

     Month: Ordered (Ordinal, Cyclical) (indicates the month when data was collected)

     x: Ordered (Divergent, Quantitative) (indicates a 3-month SPI value, the number of standard deviations from normal precipitation totals from across 3 months; corresponds with the region, month, and year in the same row)

     Precipitation: Ordered (Sequential, Quantitative) (indicates the average amount of precipitation, in inches, collected across a certain region's cities)

     Annual Precipitation: Ordered (Sequential, Quantitative) (indicates the average amount of precipitation, in inches, collected in a year for a certain region)

6 Task Abstraction

In this section, we will use Munzner's visualization task taxonomy.

6.1 Domain Task #1

The first domain task is to determine the relationship between rainfall values and drought levels among region(s). The visualization task for this action would be to compare rainfall values across regions in Massachusetts to the respective drought levels. The user can do this task by observing a map to determine which region(s) of Massachusetts they want to scrutinize and then look at the relationship between precipitation and drought severity on a scatter plot for that region(s).

The high-level tasks would be to consume and discover. The user is not generating any data. Using the information presented on the scatter plot and the map, users can make compelling findings with the rainfall pattern data and its relationship to the drought data. They also would likely not know (but can predict) if there is a relationship between drought intensity and rainfall patterns since this tool is targeted toward the general public. They would discover and make conclusions about this relationship on their own based on their interpretations of our visualization tool.

The medium-level task would be to locate. A user would want to see whether a set of SPI values and corresponding precipitation measurements prove that there is a relationship between the two. The target finding is clear: any sign of correlation between precipitation and SPI, whether it is positive, negative, logarithmic, exponential, or another type of relationship. Users are just mainly searching for any signs of correlation between the two variables. They would not know which months and years to observe the data from, though. In fact, the user can only sufficiently accomplish this task by observing data from the scatter plot across multiple years and months. Therefore, the target is known but the location for this task is unknown.

The low-level task would be to compare, as the user is comparing the rainfall trends of a region with the corresponding drought levels to see if there is a relationship between them. More specifically, they can observe those findings on the scatter plot. From there, the user can conclude whether declining rainfall patterns should be used as a predictor for drought in specific regions.

6.2 Domain Task #2

The second domain task is allowing users to determine which regions tend to have more or less precipitation than other region(s) during one or multiple months across one or multiple years. The visualization task for this action would be to identify regions that have lower precipitation levels than other regions given the same year and month as well as given different years and the same month.

The high-level tasks would be to consume and discover. The user is not generating data and using the information presented in our visualizations to make compelling findings. They also would likely not know when to help a region of Massachusetts with more water resources since this tool is targeted toward the general public and would thus discover that critical information on their own.

The medium-level task would be to locate. A user would want to find the specific times (months/years) when a region(s) in Massachusetts is most prone to droughts. Thus, the target finding is clear: marks on the line chart that are consistently lower compared to others during the same month across one year and/or marks on the scatter plot that are lower compared to other regions during the same month across multiple years. However, they would not know which months and years to observe the data from to access the target's location and would therefore need to locate the appropriate data via filtering and testing out different regions, months, and years to be shown on the plots.

The low-level task would be to identify, as users would be observing precipitation levels in Massachusetts across regions over time to see which months a region(s) has precipitation totals lower than others given a one-year time frame and/or multiple-year time frame.

6.3 Domain Task #3

The third domain task is enabling users to determine if a region having a higher or lower SPI value corresponds with that region having higher or lower annual precipitation totals. From there, the visualization task for this action would be to see if higher or lower annual precipitation values correspond with more negative or positive SPI data points for all the regions. This task allows users to be better informed about how changes in annual precipitation and SPI drought levels are related and take appropriate action when they hear about a possible decrease in the annual precipitation total for their region that reflects the annual precipitation of a previous year with many monthly precipitation totals lower than normal.

The high-level tasks would be to consume and discover. The user is not generating data and using the information presented in our visualizations to make compelling findings. They also would likely not know the relationship between annual precipitation and SPI, whether a region has actually been experiencing a greater amount of negative SPI data points over time, or which annual precipitation totals correspond with more monthly negative SPI data points for a region during the same year. This assumption derives from how this tool is targeted toward the general public. The user would make these conclusions after experimenting with our visualization.

The medium-level task would be to locate. A user would want to see whether the number of months a region has experienced negative SPI totals has increased or decreased and then identify the annual precipitation totals that align with more negative SPI totals. The target finding is clear since users are just looking for any sign of a relationship between time, negative SPI occurrences during a month, and annual precipitation for a region, even if the relationship itself is not known. Also, users know to look for annual precipitation totals that correspond with more negative SPI totals on the scatter plot (i.e. values on the bar chart that are linked to a greater amount of points to the left of where SPI = 0 on the x-axis). Yet, they would not know which months and years to observe the data from to make proper conclusions. In fact, the user can only sufficiently accomplish this task by observing data from the linked bar and scatter plots across multiple years and months. Therefore, the target for this task is known while the location is unknown.

The low-level tasks would be to compare and identify, as the user is comparing the number of months a region experiences negative SPI values across multiple years, each of which is associated with unique annual precipitation totals that get identified by the user. If that number has increased, then the region has been experiencing increasingly lower precipitation totals than normal per month, which should raise concern for that region regarding its water supply. Also, if a region is projected to have an annual precipitation total that has historically been connected with more monthly precipitation totals lower than normal, then people should be concerned. If the number has stayed around the same or decreased, then that region should be prioritized with fewer conversation efforts compared to others that are clearly experiencing declining SPI values per month.

7 Design Process

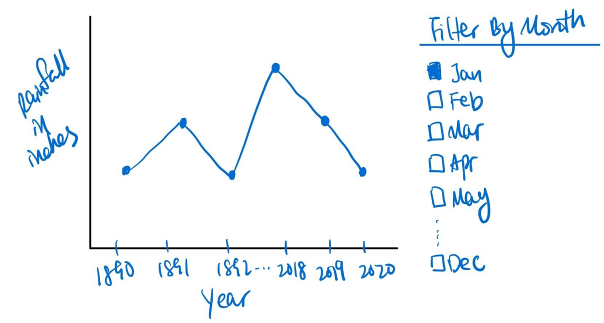

Figure 1: Visualization 1 - Rough Sketch: Shows a line chart showing the relationship between drought severity (using SPI) and total precipitation in inches. The visualization can be filtered by month There is a line for each region and the regions are color coded.

For the rough sketch in visualization 1, we thought of a tool that only allows users to select one month to filter by when looking at the average amount of precipitation a region in Massachusetts yields across its cities over time (with years being on the x-axis). For the rough sketch in visualization 2, we show how the tool would also include a line chart presenting precipitation over time, where the x-axis in this sketch is in terms of months rather than years and the y-axis represents the average amount of precipitation in inches across cities in each region of Massachusetts. We also allow for filtering by region so that unwanted data isn't cluttered in the graph. We used our ideas from the first two rough sketches (visualization 1 and 2) to determine our revisions for the final visual encoding. We will allow users to filter the precipitation data by multiple months if they want to see the average amount of precipitation each region experiences across each of its cities from the 1800s to 2015, as the first sketch implies. We will also allow users to filter by year (and not show all the years on the x-axis). Accordingly, on the line chart's x-axis, instead of years, both years and months would be shown. For instance, say a user is interested in 2015 and 2016 data from January and March. The x-axis would then contain 1/2015, 3/2015, 1/2016, and 3/2016. These revisions are reflected in the final visualization.

For the rough sketch in visualization 3, we designed a tool that allows users to filter data about the standard precipitation indices across each region by year as well as month. In the final visualization, this map will probably use the same region, year, and month filters as the other visualization showing precipitation distributions in inches over time. Therefore, we probably won't be using the slider on this sketch to filter the months, but rather just the month checkboxes used in the other visualization. Additionally, we will make a note to the user in our final visualization that this map will default to 2019 data as we only have 2017, 2018, and 2019 drought data. Additionally, the coloring of each region will be based on the earliest year and month indicated by the checkboxes. In other words, if the years 2017 and 2018 were chosen along with January and February, then the coloring of the regions would be based on the 3, 6, or 12-month SPI values of January 2017 for each of the regions.

For the rough sketch in visualization 4, we design a tool that would allow one to look at the relationship between the total average amount of precipitation in inches each region in Massachusetts

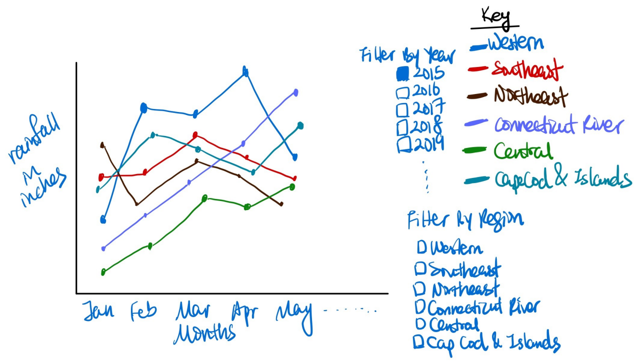

Figure 2: Visualization 2 - Rough Sketch: Line charts presenting the precipitation over time. Users can filter by year. Users can filter by region (only showing lines for a set of chosen regions rather than all six). Hovering over a line indicating a certain region would highlight parts on visualization 3 that correspond to the same region.

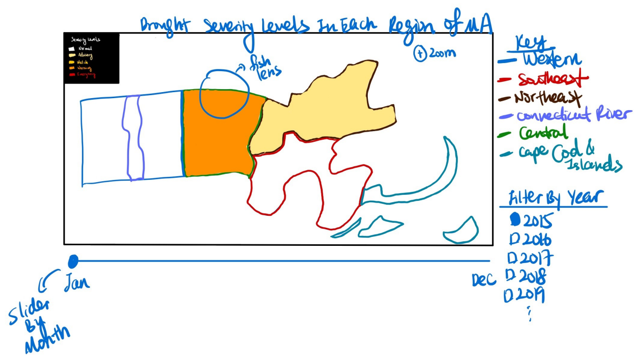

Figure 3: Visualization 3 - Rough Sketch: Map of Massachusetts indicating drought severity levels per month and year Filter for the year (2017 SPI data vs. 2018 SPI data, for instance) - this filter can be the same one used in visualization 2. Slider to change which month(s) you are trying to look back on. Hover over each region to see the 3, 6, and 12-month SPI. Color map for different levels of drought intensity. Pan and zoom/fisheye lens.

experienced across its cities per month in one year and drought severity (which is measured by the 3, 6, or 12-month SPI values for that given year and month). We tried to incorporate a year filter. This component will be useful in allowing users to select the year(s) from which data is observed. We decided that we should also incorporate the previously used month filter to allow users to select the desired months represented. Lastly, we decided to incorporate a region filter allowing users to select 1 or many regions to be represented so this visualization doesn't look cluttered. These three filters would be the same ones used for the other two encodings of the final visualization. We also realized that using a scatter plot would make more sense with regards to showing a relationship between the two variables as opposed to connecting the points and making a line graph.

Our final sketch includes three different visual encodings. It shows how we planned to have a line chart showing the average amount of rainfall that cities across regions in Massachusetts experienced in inches over time. The users would have been able to filter the data by year(s), month(s), and region(s). We also planned

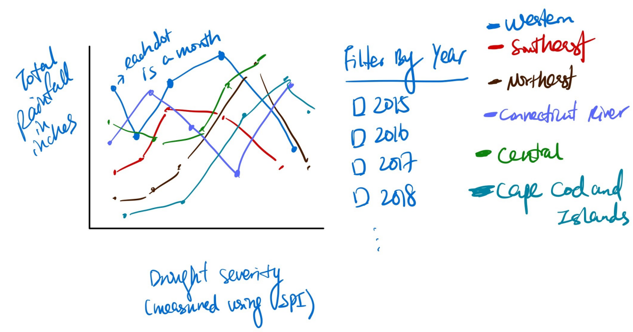

Figure 4: Visualization 4 - Rough Sketch: Shows a line chart showing the relationship between precipitation and year. The y-axis is the total precipitation in inches. The x-axis shows the years The line chart is filtered using months. The user in this case can only choose one month. There will be a line for each of the regions: Western, Southeast, Northeast, Central, Cape Cod and Islands, and Connecticut River. The rough sketch below only shows one of many lines that there can be (representing different regions).

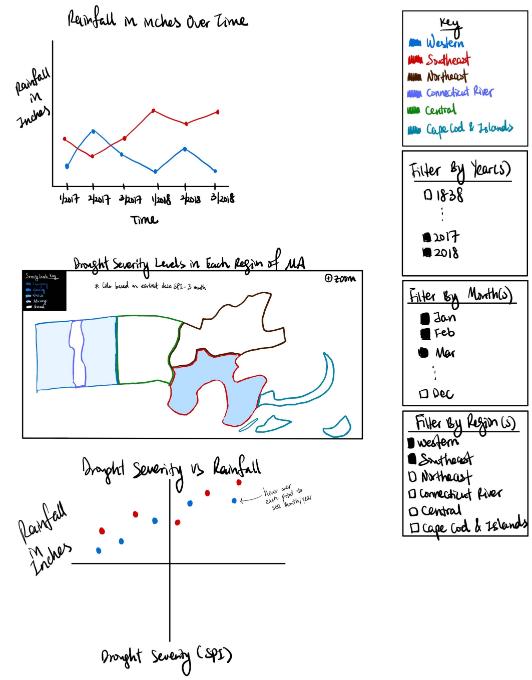

Figure 5: Visualization 5 - Final polished design with all three visual encodings together in one. Region, year, and month filters will be shared among all three visual encodings, therefore linking all three visualizations together. Whenever a filter is changed, all three visualizations will change accordingly. Additionally, when hovering is done on the lines for the line chart showing average precipitation distributions across cities in regions over time, the corresponding region on the map will be highlighted along with the corresponding point(s) on the scatterplot. This same hovering function will work whenever hovering is done on the other two visualizations as well. (Note: the interactive and coordinated elements of this visualization are described in further detail below.)

to have a map of Massachusetts showing drought severity levels across regions in Massachusetts that could also get filtered by year, month, and region. We also considered adding a filter to adjust the "look-back" periods represented by the colors on the map (3, 6, or 12 months). Because we only have drought data for the years 2017-2019, we would have defaulted our colors based on the 2019 information. Lastly, we planned to have a scatterplot showing the relationship between the average amount of rainfall a region experienced across each of its cities over time and drought severity (SPI), which the user could influence by filters for the region(s), year(s), and month(s) represented.

With regards to the line chart displaying the average rainfall across cities in regions in Massachusetts in inches over time, we included a tooltip function where the user can get exact rainfall measurements, the specific date (month and year), and region represented by a point when it gets hovered over. There is also a key/legend on the side detailing the region each line represents. In terms of marks used in the visualization, lines and points are being used. For channels, color hue is used to represent the different regions, and horizontal and vertical positions are used to designate the specific month and rainfall in inches that a point represents during the year picked in the drop-down menu, respectively.

With regards to the scatterplot showing the relationship between rainfall and drought severity, there is a tooltip function that allows users to see the exact rainfall and SPI values a point represents when it's hovered over. They will also be presented with the month and year combination as well as the region a point represents. In this visualization, the point mark is used to represent each of the drought severity and rainfall data values. For channels, the color hue channel is used to represent the region a point represents. Horizontal and vertical positioning is also used to symbolize the average amount of rainfall a region has across its cities and the drought severity level (based on standard precipitation indices) each data point indicates, respectively.

With regards to the map showing drought severity levels in each region of Massachusetts, we scrapped the idea of allowing a user to see the SPI values on a map in which the color of each region is determined by SPI thresholds. We instead focused on making a map that allows a user to see which region they live in (through markers, not borders). The markers on the map are surrounded by cities that a user may live in. Users can then click on each marker to see the county associated with each nearby city, helping the user to determine which county they live in. The colors of each marker correspond with a certain drought region, helping the user to also realize which drought region they live in. From there, they can observe data on the region they live in across our encodings if they wish. Our original map idea would have taken too much time to implement. It also would have provided data repetitive with our SPI scatter plot, which also shows the SPI values each region experienced over time. In terms of marks used in the visualization, area marks are used to show the different regions, and points are dotted across the map to indicate the county that cities surrounding the point are a part of. In terms of channels, area (spatial position) is used to represent each region in Massachusetts distinctly. Additionally, color hue is used to distinguish the region each marker represents. The spatial position of the points is also scaled based on the exact latitude and longitude measurements of counties in Massachusetts.

We also added the linked bar and scatter plots (see "Final Design"). The bar charts present the annual precipitation totals each Massachusetts region accumulated. The chart can be filtered by region and by the years 2017, 2018, or 2019. The scatter plot is similar to the one we already planned to include (showing the relationship between SPI and precipitation) except that it also has a dotted line where SPI = 0 on the horizontal axis. As mentioned, we originally wanted the user to link between the line and the original scatter plot via region, but we realized that there would be no clear purpose for this feature. The bar and scatter plots do as they enable a user to determine the relationship between the distribution in the number of negative SPI values a region experienced in a year and the region's annual precipitation for that year. They do so by comparing the number of points to the right and left of the line that marks where SPI = 0 with that region's annual precipitation. From there, the user can see whether a region has been experiencing more monthly precipitation totals lower than normal over time. They can perform this operation by changing the years represented on both charts and comparing the annual precipitation values as well as the corresponding number of scatter points that represent negative SPI values across each year. If the annual precipitation values tend to increase as the number of points to the right of the vertical line on the scatterplot where SPI = 0 increases, then a positive relationship between annual precipitation and SPI can be implied.

The user has many tools to aid them when navigating the linked charts. They can hover over the bars and points to see a tooltip presenting the information each mark represents (region, annual precipitation/SPI, month for just the scatterplot, and year). They can click on a bar to highlight points that represent the same region on the scatter plot. Finally, they can brush over points on the scatter plot to highlight corresponding bars that represent the same region.

For these visualizations, the line mark is used to represent annual precipitation values on the bar chart while the point mark is used to represent each of the drought severity and rainfall data values on the scatter plot. For channels, the color hue channel is used to represent the region that a point or bar represents. For the bar chart, the region a bar symbolizes is emphasized via horizontal position, and that region's annual precipitation value during a certain year is portrayed through vertical position. For the scatter plot, horizontal positioning is used to symbolize the average amount of rainfall a region has across its cities while vertical positioning is used to symbolize the drought severity level (based on standard precipitation indices) each data point indicates.

As mentioned before, we planned to include region, year, and month filters that would have been shared among all three original encodings, therefore linking all three visualizations together. When designing, we decided to no longer allow a user to filter by month for the line chart and just focused on allowing people to filter by year. A month filter would have taken too much time to implement. Besides, allowing users to remove data from certain months from their view could make them prone to making incorrect conclusions about the precipitation trends of regions in a certain year. Furthermore, regarding the checkboxes, years and month checkboxes are no longer shared among all the visualizations (while a region one still is). The precipitation data, which is used to primarily build the line and bar charts, spans from 1838 to 2019. Conversely, the SPI data, which is primarily used to build both the linked and unlinked scatter plots, only spans from 2017 to 2019. Hence, it wouldn't make sense to use one year filter for all the plots. Instead, users can now filter the line chart data by year using a drop-down menu, while the users can filter the years presented on the unlinked scatter plot with checkboxes that enable them to see data across multiple years. They can filter the year represented on the linked charts with a drop-down menu with the years 2017, 2018, and 2019 since the linked scatter plot only contains data for those years. We also included buttons for the charts so that data only appear after all the necessary filters have been selected by a user. In this way, the plot gets a chance to reset before showing the data that aligns with the user's choice of filters.

Some smaller changes include changing the colors of the region key from the ones in our sketch since we were restricted by the default colors of the library we used to build our map (Leaflet). We still kept in mind choosing colors that create a categorical color map as we are showing distinct regions. We also dismissed our original idea of having 4 quadrants for the unlinked scatter plot since negative precipitation values don't exist. Finally, we decided to group our visualization tool into tabs to make it more organized.

7.1 Usability Testing

After deciding on these changes, we performed usability testing with the visualization tool we currently had. For the line chart, we instructed subjects to pick a year and region(s) with the filters. We then instructed them to look at the line plot and locate the region with the highest and lowest precipitation value for the month of May. We had them think out loud while performing this so we can better understand their thinking process while using the visualization. Our visualization does not present data initially until a user selects a year AND regions to be shown on the chart (none of the region checkboxes are originally checked). Thus, some subjects were confused as to why data wasn't initially present on the chart. We did not have enough time to implement the necessary code that would allow lines to initially appear on the page, so we saved the idea as a future consideration. We also changed the colors of the instructions for using our visualization into one that popped out (royal blue) since some subjects were confused about how to get data to appear. Some subjects also suggested that the scale of the y-axis be changed each time they click a button to present the line chart since the minimum and maximum precipitation values used for the axis don't necessarily align with those for the year and region(s) it represents. The scaling is inappropriate for data during certain years and regions. We didn't have enough time to implement this feature but it would be a future consideration. Overall, subjects found this task on a Likert scale ("Very Easy," "Easy," "Normal," "Hard," "Very Hard") to be "Easy."

We also asked the subjects to filter the chart to only show data for Central, Northeast, and Western Massachusetts in 2006. We then asked them to identify how much precipitation Central Massachusetts received in August. On average, subjects took 19 seconds and were mainly confused with how the tooltip that contained the appropriate information to complete the task was the same color as one used in the chart. Subjects also suggested reformatting the information on the tooltip so that the most important information was at the top row (e.g., region). We accepted these criticisms and adjusted the tooltip for the line chart appropriately. In general, subjects rated the task as "Normal" on a Likert scale.

We also tested the subjects' capabilities to handle our scatter plot. We first asked them to pick a year on the drop-down menu, check all the month checkboxes, and then pick a region(s) to be represented on the plot via checkboxes. We then asked them to explain the overall trend of SPI vs. precipitation for the selected regions. We also asked them to explain what this trend means in a broader sense using the meaning of SPI defined on the webpage. In particular, a subject chose to filter the data by June 2018 and accurately stated that precipitation and SPI increase somewhat linearly and positively. All subjects were able to accurately state that there is a positive correlation between precipitation and SPI. We still thought that we could have made the definition of SPI clearer on our webpage (e.g., by presenting the definition in an eye-catching color) since the people using our tool most likely don't have a background in meteorology. We also used the results of this experiment to further motivate us to change the colors of the instructions to a more standout color. We also decided to change the number of ticks on the x-axis as some values (e.g., -1) were placed on multiple ticks. Overall, subjects rated this task as "Easy" on a Likert scale.

For the same plot, we asked subjects to filter the chart to only represent the Cape Cod and Islands, Connecticut River, and Southeast regions of Massachusetts. We also asked them to filter the plot to only show data from May 2018. We then asked them to name the 3-month SPI value for each of those regions in May 2018. On average, subjects accurately named the correct values in 19 seconds and had no additional suggestions, claiming the task was "Easy" on a Likert scale.

We then evaluated the subjects' capabilities to use our map as intended. We first asked whether they could tell which part of the world our map focused on (Massachusetts). Since Massachusetts is highlighted in the map in blue in contrast to the rest of the map, on average, subjects got the right answer in 6 seconds. Overall, subjects found the task to be "Easy" on a Likert scale.

We then asked subjects to identify how many regions of Massachusetts the map focuses on. In general, subjects actually found this task to be "Hard" on a Likert scale and took, on average, 13 seconds to complete it. They initially thought that each marker corresponded to a region of interest, not realizing that each marker symbolized a different county that belongs to a specific region. All the markers were originally blue, so we then decided to color code each marker to indicate the region a county is a part of. The subjects also suggested adding layers to our map to aid in this task, which we would address as a future consideration.

Finally, we tested how comfortable our subjects felt using our linked charts. We first asked them to filter the data to only show values from 2019 using the checkboxes. We then asked them to use the brushing and linking features to name the annual precipitation value that Northeastern Massachusetts received in that year as well as one of its monthly SPI values for that year. On average, subjects completed the task in 28 seconds and found it "Easy" to accomplish on a Likert scale. We again affirmed on making the instructions for navigating the encoding a different color. There were also moments in which the tooltip wouldn't appear when subjects hovered over bars and points, so we had to fix the offset of the tooltip to address that issue.

Our last task for the subjects with handling our visualization tool was having them freely play around with the brushing and linking for 30 seconds. While experimenting, we wanted them to notice if there was any relationship between the distribution in the number of points on the scatter plot to the left and right of the x = 0 line versus the annual precipitation on the bar chart for each region. In general, subjects correctly identified that higher precipitation values correlate with a greater number of points to the right of the line, or positive SPIs, and that lower precipitation values correspond with a greater number of points to the left of the line, which represent negative SPIs. Overall, subjects found this task "Easy" on a Likert scale and offered no further suggestions.

8 Final Design

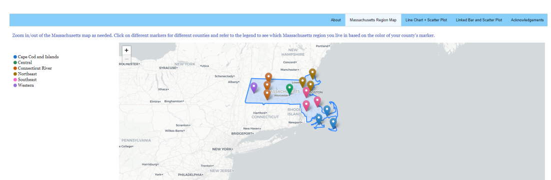

Figure 6: Visualization 1 - Final visualization of the Massachusetts region map)

With regards to this visualization's design, the above figure presents a map of Massachusetts because the visualizations center around that state. With regard to the colors, we used a universal categorical color map that presents the region a marker represents. There are also simple directions at the top of the visualization to clearly tell the user how to use the map.

The Massachusetts region map allows users to zoom in and out and utilize the tooltip to understand better which region they live in based on the counties targeted on the map. The user can utilize this map to easily locate the region they are located in and therefore, better understand which regions of interest they would want to focus on for the other visualizations to address any domain tasks/problems (domain tasks one and two in particular). First, the user would locate which city they live in on the map. Then, they would click on the marker closest to that city. By clicking over the markers, the user would know which county of Massachusetts they likely live in. They also would use the legend to determine which drought region they live in based on the marker's color.

Figure 7: Visualization 2 - Final visualization of the precipitation over user-selected times and regions line chart and a projection of whether lower precipitation values indicate a higher risk of intense droughts scatter plot)

With regard to the visualization design in Figure 7, we used the same categorical color map that the map uses to indicate a region that lines and points represent. For the line chart, the independent variable is region while the dependent is precipitation in inches across months in a year. For the scatter plot, the independent variable is SPI over certain months across one or more years while the dependent is precipitation in inches over certain months across one or more years. Above each of the plots, we also included filters specific to each visualization see "Design Process"). Filters shared between the two plots (i.e. the region filter) are to the left of the screen. There are a lot of directions and instructions as to how to interpret the variables as well as how to use the visualization (e.g., knowing which filters to use) in order to make the visualization easy to navigate for the user.

The visualization above allows users to filter both plots by region on the left-hand side using the provided checkboxes. The user can then use the year filter for the line chart to select one year. And for the scatter plot, the user can filter by one or multiple years using the provided checkboxes as well as filter by one or multiple months. After making all these necessary selections, the user can press the button for the respective data to be shown on the plots. By hovering over any of the points on either the line chart or the scatter plot, the user can take advantage of the tooltip that gives more specific, detailed data on the specific point hovered.

The line chart supports domain task two (as mentioned in "Task Abstraction") and the scatter plot supports domain tasks one and two (as mentioned in the "Task Abstraction").

When working with the line chart, the user filters the data by region using the checkboxes on the left-hand side, filters it by year using the dropdown above the line chart visualization, then presses the "Display Line Chart" button. Looking at the resulting chart can help users prioritize conservation efforts during certain months in places where the precipitation is lower than others during the same month.

When working with the scatter plot, the user filters it by region using the checkboxes on the left-hand side, filters it by one or multiple years using the checkboxes above the scatter plot visualization, filters by one or multiple months with checkboxes above the plot, and then presses the "Display Scatter Plot" button. Looking at the displayed plot informs users of the relationship between precipitation and drought levels (how does one change as the other changes) and therefore use this understanding to anticipate the possibility of droughts in the future given information about the precipitation values in a certain region. Additionally, the scatter plot allows users to better understand regions that need greater conservation efforts given the same months across various years.

With regard to the design of the visualization in Figure 8, we used

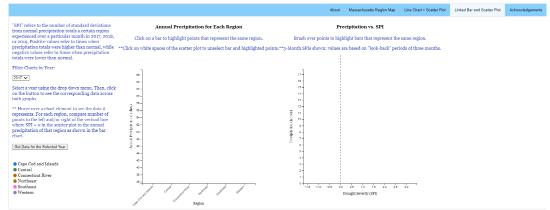

Figure 8: Visualization 3 - Final visualization of a bar chart showing the annual precipitation for each region linked with a scatter plot of Precipitation vs. SPI)

the same categorical color map used for the other visualizations to indicate the region that each bar and point symbolizes. The independent variables of the bar plot are regions while the dependent is annual precipitation. For the scatter plot, the independent variable is SPI over certain months across one or more years while the dependent is precipitation during a certain month and year in inches. A filter used to adjust the year represented on each plot is on the left of the screen. There are a lot of directions and instructions as to how to interpret the variables as well as how to use the brushing and linking for each of the plots on the left side as well as above the charts.

The visualization in Figure 8 allows users to filter both plots based on a year between 2017 and 2019. Then, the user can press the button "Get Data for the Selected Year" to display the corresponding information on both the bar chart and scatter plot to the right. This visualization helps to address the third domain task, as discussed in "Task Abstraction." By selecting a year of choice, the user has two ways of better understanding how SPI levels correspond with annual precipitation for all regions of a given year. One option is to click on a bar on the bar chart. This action will then help users to see how many corresponding points are to the right and left of the SPI = 0 vertical line since the corresponding points that represent the same region would get highlighted. Users can identify if there is a relationship between the number of points left versus the right of the SPI = 0 line in relation to the annual precipitation of the region (relative to other regions) to see which annual precipitation totals for a region tend to correspond with more negative SPI values per month. The second option for users to address the third domain task is to brush on the scatter plot, which will then highlight the bars on the bar chart that represent the same region(s) indicated by the points that are brushed. This action again allows the user to analyze the previously mentioned relationship between annual precipitation and SPI for a region.

9 Discussion

Our final visualization tool addresses the domain problems we set out to solve, but not to the fullest extent. As we built our tool, we had to change some of the domain tasks to better align with the visualizations we were able to create during the process of implementation. We were limited by the data available on the data set as well as our knowledge of creating visualizations. For instance, 3-month SPI values may not be the most representative metrics for drought severity, impeding a user's ability to determine a region's susceptibility to droughts with the best precision possible. Therefore, our visualization doesn't provide the most intricate insight into a region's precipitation and drought patterns. The visualization tool also helps to emphasize the rise of droughts across Massachusetts, but it fails to address the exact measures users can take to resolve the issue.

One limitation of the tool is that it only shows 3-month SPI data. If we could make improvements in the future, we would show 3-month, 6-month, as well as 12-month SPI drought level data. Allowing users to filter the look-back periods for the SPI data would be useful in providing a greater understanding of how drought levels relate to precipitation. By incorporating more look-back periods in the visualization, the conclusions made about a region's future drought severity can be strengthened or weakened depending on whether SPI consistently increases or decreases for that region over time, regardless of the choice of look-back periods. Therefore, we would include checkboxes for both scatter plots to allow users to filter the SPI data shown by the look-back period.

Another limitation is that our axes in the line, bar, and scatter plots are static. In the future, it would be helpful to implement dynamic axes that change according to the range of data that is being plotted. Because the axes are static in the visualizations, it is sometimes hard to make quick distinctions among y-variable values that are close to each other. The axes currently depend on the minimum and maximum precipitation, annual precipitation, or SPI values in the imported data sets as a whole and don't depend on values from user-specified filters. Therefore, if, for example, a user filters the line chart to only present data from 2000 for each region, if the range of the 2000 monthly precipitation values for those regions is much smaller than the range of precipitation values in the precipitation data set in its entirety, then the lines on the chart would appear squished, and the graph would contain a lot of unnecessary white space.

Our visualization also has no initial data marks presented on the web page, which can seem misleading to the user as this feature might cause users to think that our web page doesn't function. This impediment would draw users away from making potentially important conclusions about a region's drought and precipitation patterns, decreasing the chance a region receives the conservation efforts it needs. Therefore, in the future, we hope to have some marks and lines on each chart of the web page by default.

Our map visualization also has very limited capabilities as it only has markers for a specific set of counties. In the future, we hope to add layers to the map to make it more visually appealing as well as so users can link the map to other visualizations to show more meaningful information. For instance, if a user hovers over a line on the line chart, the part of the map that represents the same region as that line could get highlighted. The map should also be altered so that borders distinguish the drought regions on the Massachusetts map. While markers are fine, the map would look too messy if markers for all of Massachusetts's counties to indicate a region a city is located in were implemented. Dividing the map of Massachusetts by region with borders would offer a clean way to see all the cities that belong to a region. The borders would be filled in with a transparent color based on the region represented.

Lastly, the visualization should allow users a way to see how certain efforts can change precipitation/SPI levels in the future. This task is way too complex for us with our current skills and the data we had and seems to require the need for machine learning. For instance, perhaps a multiple regression model can be shown on the tool to predict a region's annual precipitation patterns if a region reduces its deforestation by a certain amount. Or, a random forest regressor can train on data showing how deforestation, pollution, and temperature affect precipitation and SPI levels to predict values for a region's annual precipitation and SPI based on inputted features. By embedding machine learning in our tool, users will have a more specific idea of what kinds of actions they can take to conserve as much water in a region as possible while minimizing trial and error.

10 Conclusion

Our visualizations hope to serve our target domain of the general public of Massachusetts. The ultimate goal is to increase awareness of the precipitation patterns and drought levels throughout Massachusetts to encourage the general public to focus on water conservation moving forward. The domain tasks supported by our three sets of visualizations will hopefully help them achieve that goal. The first task is allowing users to determine the relationship between rainfall values and drought levels among region(s). The second task is to help users determine which regions tend to have more/less precipitation than other region(s) during one or multiple months across one or multiple years. The third task is enabling users to determine if a region having a higher or lower SPI corresponds with that region having higher or lower annual precipitation totals, as well as the specific annual precipitation totals linked with more negative monthly SPI values. The visualizations included to support these tasks include a map, unlinked scatter plot, unlinked line chart, and a set of a linked bar chart and scatter plot.

10.1 Contributions

We equally contributed to each of the project milestone assignments for the project. Michelle worked on the abstract, introduction, data cleaning, data abstraction, final design, discussion, and conclusion. Jethro worked on related works, use cases, data selection, biases and ethical considerations, final design, and conclusion. Jane worked on the abstract, introduction, domain tasks, and design process. We all contributed to code debugging and formatting the website, working through any barriers we had along the way.

References

[1] V. N. H. Doraiswamy and R. S. Nanjundiah. An exploration framework to identify and track movement of cloud systems. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2896–2905, Dec. 2013. doi: 10.1109/TVCG.2013.131

[2] A. G. Liu S., Foss G. and B. A. Visualization of high-resolution weather model data. *EGPGV@ EuroVis*, pp. 63–66, May 2015. doi: xmlui/bitstream/handle/10.2312/pgv.20151156.063-066/063-066.pdf?sequence=1isAllowed=y

[3] C. of Massachusetts. Precipitation database data. `https://www.mass.gov/info-details/precipitation-data`, 2023.

[4] C. of Massachusetts. Standardized precipitation index (spi). `https://www.mass.gov/info-details/precipitation-data`, 2023.