Text, Tables, Graphs, and Maps:

Communicating with Data Visualization

by

Leslie Raine Carman

A Thesis
Submitted in partial fulfillment
of the requirements for the degree
Master of Environmental Studies
The Evergreen State College
October 2019

This Thesis for the Master of Environmental Studies Degree

by

Leslie Raine Carman

has been approved for

The Evergreen State College

by

____

Kathleen Saul, Ph.D.
Member of the Faculty

____

Date

ABSTRACT

Text, Tables, Graphs, and Maps:
Communicating with Data Visualization

Leslie Raine Carman

Scientists, researchers, and analysts visualize data that is then often shared with the public. Knowing the audience for whom the visualization is intended for is important; however, it can be difficult to predict how the audience will interpret these visualizations. Elements of a visualization such as wording, font, size, color, theme, style, and the type of visualization can affect what the audience interprets. To explore these elements of a visualization, I created my own maps using GIS to depict an impactful topic, Cancer. I surveyed an audience on the information they took away from those maps. I used this survey to test these methods and to find out what my audience saw in my maps. My survey results suggest that individual experiences affect which depictions people prefer and how they interpret information, but despite these differences in experience, common themes are pulled out by the audience.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my thesis reader, Dr. Kathleen Saul, for being an amazing mentor, an enthusiastic supporter and motivator, and a great friend. Kathleen has a level of kindness and understanding that cannot be matched. She provided the support I needed when in doubt and is always an encourager of unconventional ideas that allow her students to grow and produce amazing work. I always looked forward to our weekly meetings and appreciate all that Kathleen has done to make this thesis possible.

I thank my family for sharing their stories and love. My mom for exploring the country with me on road trips and teaching me to write and communicate, my dad for telling me stories and making me laugh, and my brother for living close to home to keep them busy and for inspiring me with his own brand of weirdness.

I would like to thank my friends, Nicole and Averi, for sending me funny memes, going out on adventures with me, and being there during life's challenges. Thank you for all the trips to the beach, trivia nights, pumpkin patches, sundaes on Sundays, and volleyball leagues. 5 out of 7, would do again. Thank you Nicole for bringing me baby Purr-sula and thank you Averi for coming with me to the shelter to bring home Hercules.

I would also like to thank Danny for cooking and cleaning during this last month of writing, going on adventures, laughing at my memes, and spoiling my cats. Thank you for your IT support and all your work to improve your cooking and tea brewing skills.

Hercules and Purr-sula enjoying the holidays.

Introduction

Croaking frogs, clicking cicadas, clucking chickens, bellowing cows, and the ballad of mourning doves woke me every summer morning in Illinois. I ran through prairies and cornfields, watched monarch butterflies transform from caterpillar to chrysalis, bottle fed goats, caught fireflies, and looked on as storms swirled the landscape. I never thought I would leave such an enchanting place, to rip my roots from the soil and migrate like the monarchs. These were the years before we had internet on our farm, before social media and instant news.

As summer ended, we canned our tomatoes, put straw out for the chickens, and put our gardens to bed. We would go back to school and decorate for Halloween. Living in the middle of nowhere makes for boring trick-or-treating, so we would go to Aunt Carol’s house every year as she lived in town. We dressed up as vampires, cats, Jiffy-Pop, and batman, ate our yearly Halloween dinner at McDonald’s, and went to Aunt Carol’s for candy. At least we did, until Aunt Carol was sick, and then gone. She had breast cancer, something old people get, I thought.

We got older and Halloween gave way to Thanksgiving, which meant a visit to my grandparents in Ohio where my dad had once been a kid. I looked forward to hearing my grandma’s sweet West Virginian accent and my grandpa’s quirky Pittsburgh accent. We baked pies and cookies with grandma, ate our Thanksgiving dinner, and went out bowling with my grandpa. We carried out this tradition every year until my grandpa couldn’t bowl anymore. The next year, we spent Thanksgiving at his funeral, cancer had spread through his whole body. We barely aged and Thanksgiving turned to Christmas, the time when we would spend the holidays with my mom's family back in Illinois. We opened presents, played games with my grandma, and watched NASA shows with my grandpa, Richard. His last Christmas was his last day at home. Glioblastoma had come suddenly and unexpectedly, and he was gone too soon. During his last few months I became familiar with the process of chemotherapy and the frequent trips to the cancer center. Richard inspired my interest in science and space, and later GIS and this thesis.

I left Illinois hoping to escape cancer. I continue to watch as my friends from Illinois are touched by cancer. I see their updates on social media and see that they experience the effects of cancer more often than my friends from other states. This thesis is dedicated to my family and friends whose lives have been touched by cancer. To those who have fought, are still fighting, have lost their fight, and have lost their loved ones.

Richard and me one Christmas Eve.

This thesis blends the science behind data, the art of visuals, and an understanding of people. We create visuals to understand science and to convey that understanding to others. Scientists, researchers, analysts, and data visualizers can influence what their clients and the public understand about the data they are sharing depending on how they craft these visualizations. Fonts, colors, and the order of the visuals are a few of the important elements that can help or hinder how well an audience can interpret visualizations. This thesis explores those elements and tests them using an online survey. An anonymous survey allows for the discovery of what information is pulled from the visualizations, what common themes and conclusions the audience comes to, and what visualizations the audience finds most informative and appealing. This thesis builds on the survey data with information from multiple disciplines to suggest best practices for creating a visualization that is accessible, interesting, and informative.

Literature Review

Data can be presented in many ways, but how do we decide what the best way to present that data is? We must think about who our audience is, how our data visualization will be used and interpreted, and what mediums are best for our audience to understand and interact with information. If we do not understand our audience, we cannot effectively communicate the stories we aim to tell with the data. Elements in a visualization such as color, font, themes, and styles all aid data visualizers in communicating information and can help an audience interpret and engage in that visualization. Each of these elements are unique to each visualization, but there are some best practices that can be followed, as explained in the sections that follow.

Poop Maps

The visualizations below (Figures 1 - 5) show how a dataset can be visualized in five different ways, two maps, two bar graphs, and a line graph. The first three visualizations are all appropriately–or inappropriately–colored in brown. These data visualizations tell the story of human feces left behind in the San Francisco area effectively using color themes and scale to alert the viewer to the enormity of the issue, although each in its own way. The first map produced by "Open the Books" (2019) displays all incidences from 2011 to 2019 at the same time, showing the viewer that uncollected waste is a huge and widespread problem. However, there are so many points overlapping that location information, such as road names, cannot be read and the distinction of land and water is not easy to see. Regardless, this visualization grabs the viewer’s attention and elicits an emotional response. This is an effective method of visualization to elicit this emotional response, but it does not tell the whole story.

While this first map shows the reported incidences all at once, creating a large brown blob, the bar graph in Figure 2 breaks down the information through time, telling viewers a different story. By including time, the bar graph sheds light on another dimension of the issue: how incidences have increased from 2011 to 2018. The viewer would not have been able to gather this information from the map visualization, but by having an accompanying bar graph, the viewer can gather two different types of information and two different stories from the same dataset.

The map from “RealtyHop” (2018) (Figure 3) takes this same dataset and cleans it up (pun intended) to give us another dimension of understanding to the dataset, which is an aggregate by location. By aggregating the points into polygons representing neighborhoods, viewers can understand which neighborhoods are the messiest and should be addressed first when tackling this issue. The “RealtyHop” map also provides additional line and bar graphs that are multi-color instead of brown. These visualizations do not follow color themes, but the line graph depicting incidents by year and state needs these additional colors so that the viewer can separate the years from one another. The bar graph visualized by “RealtyHop” displaying days of the week also displays multiple colors. The different colors on the bar graph are not necessary for understanding the data, and could be confusing to a viewer. The colors in the bar graph are the same as those in the line graph but do not correspond to the same time frame of information. The brown bar graph from “Open the Books” is arguably a better bar graph visualization than that from “Realty Hop” due to this color issue.

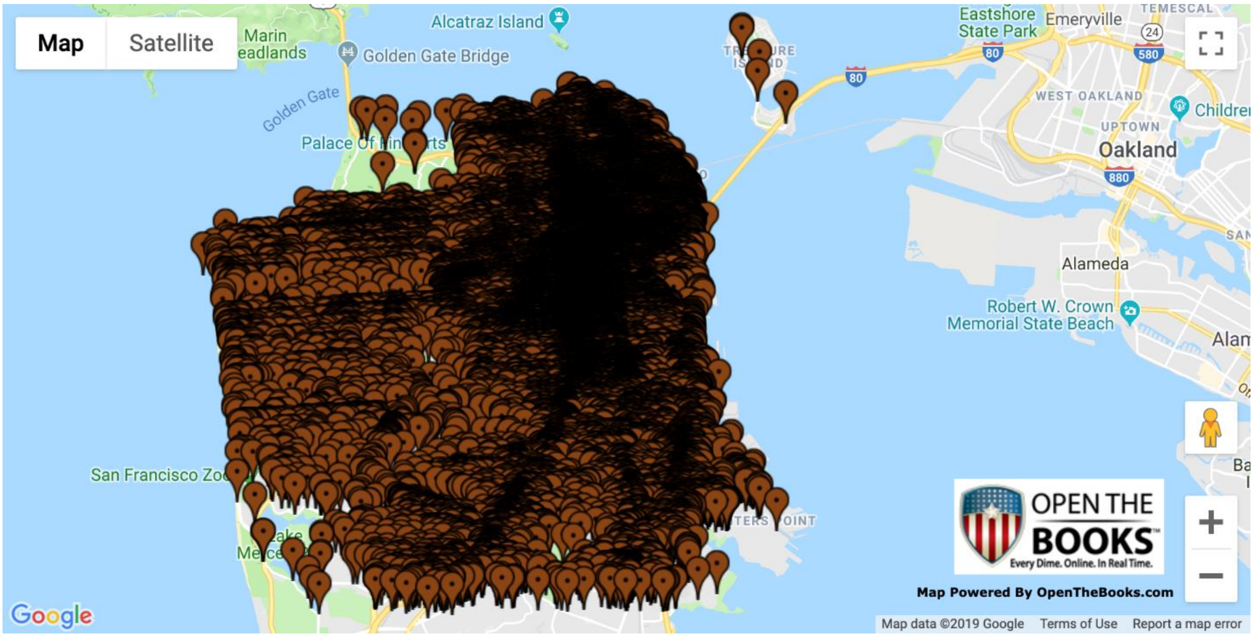


Figure 1. Open the Books map showing all incidents from 2011-2019.

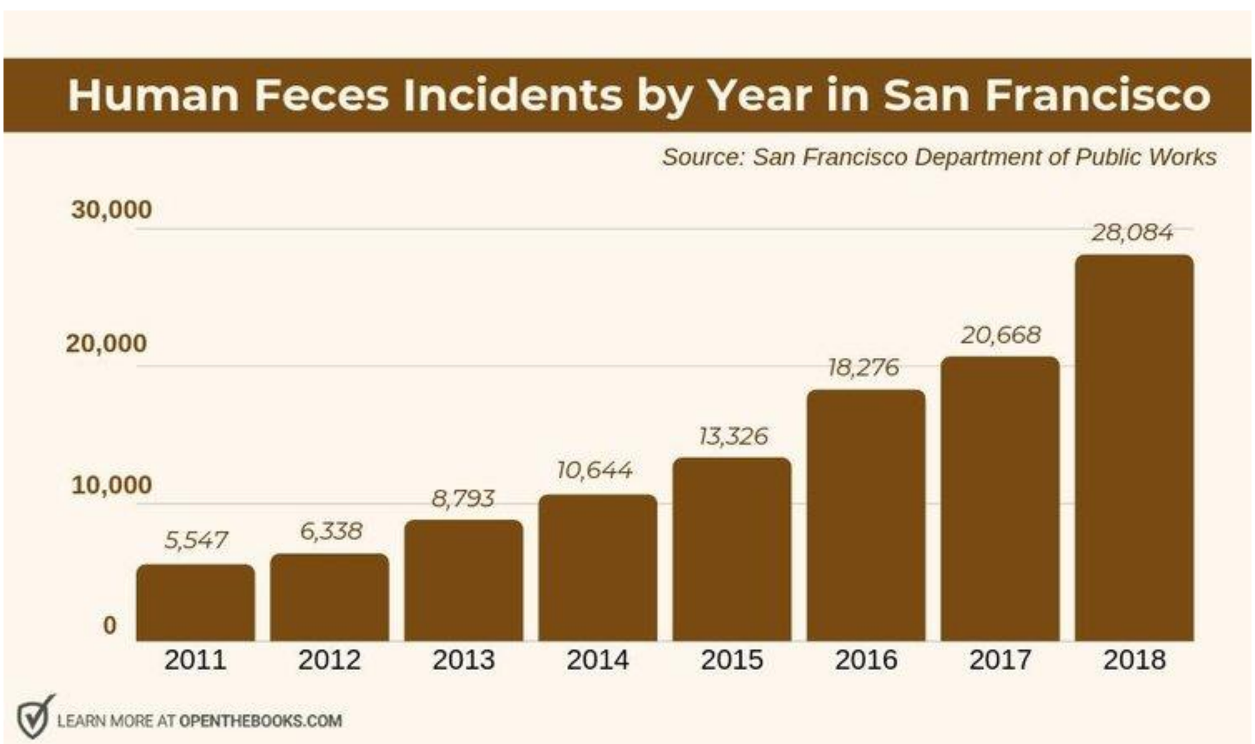


Figure 2. Open the Books bar graph separating the number of incidences by year by 2011-2018.

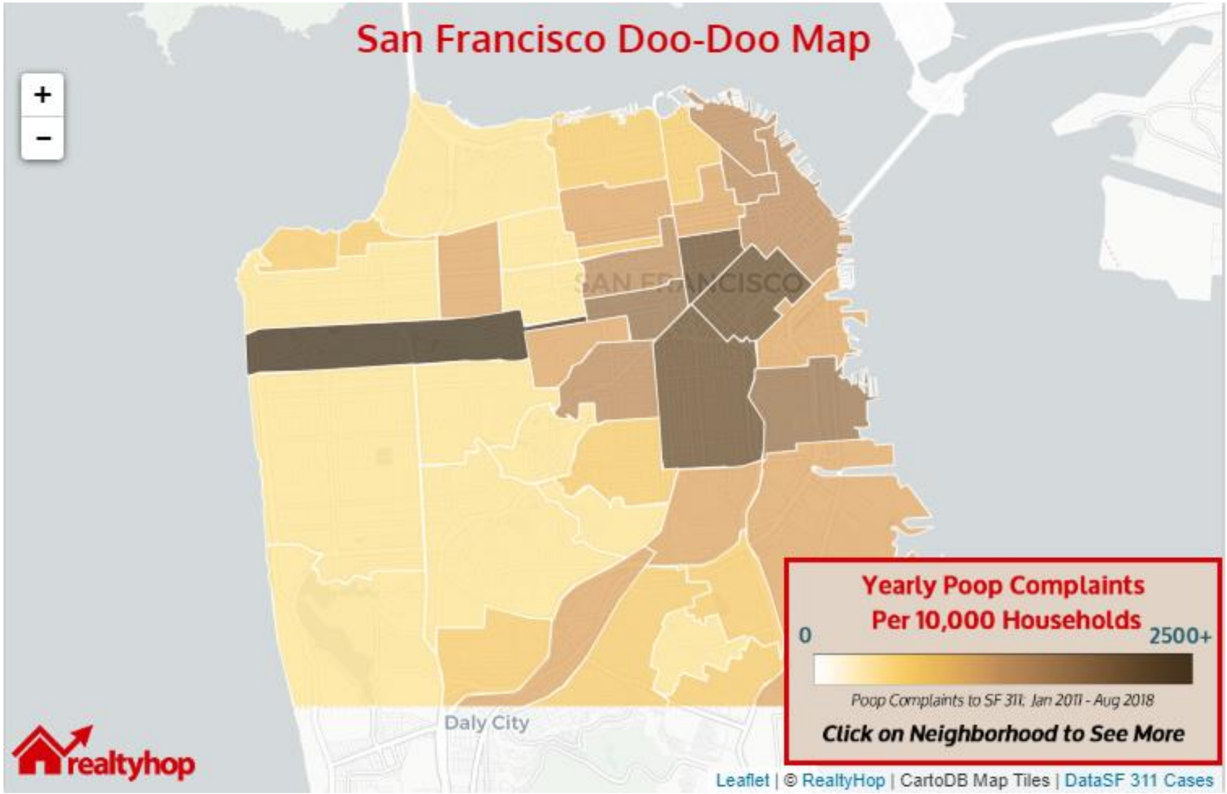


Figure 3. RealtyHop map showing incidents grouped by neighborhood in 2017.

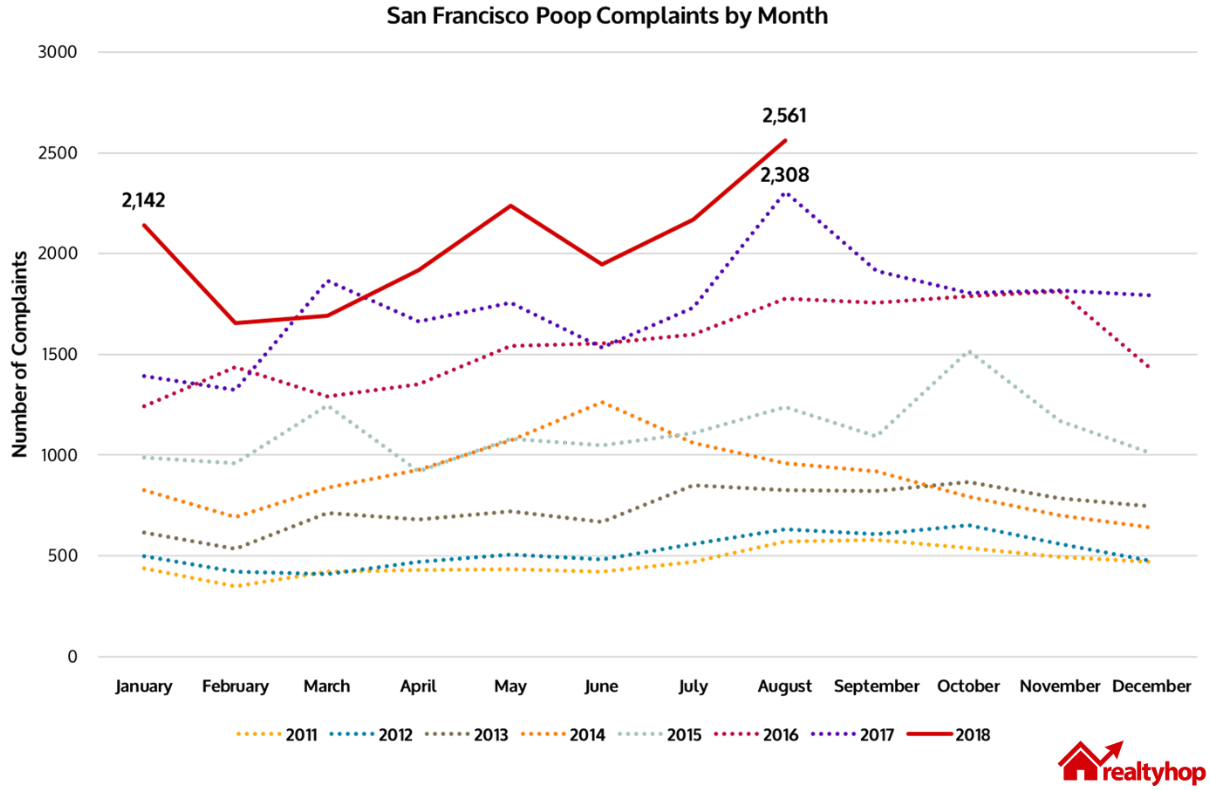


Figure 4. Realty Hop line graph show incidents by year and month from 2011-2018.

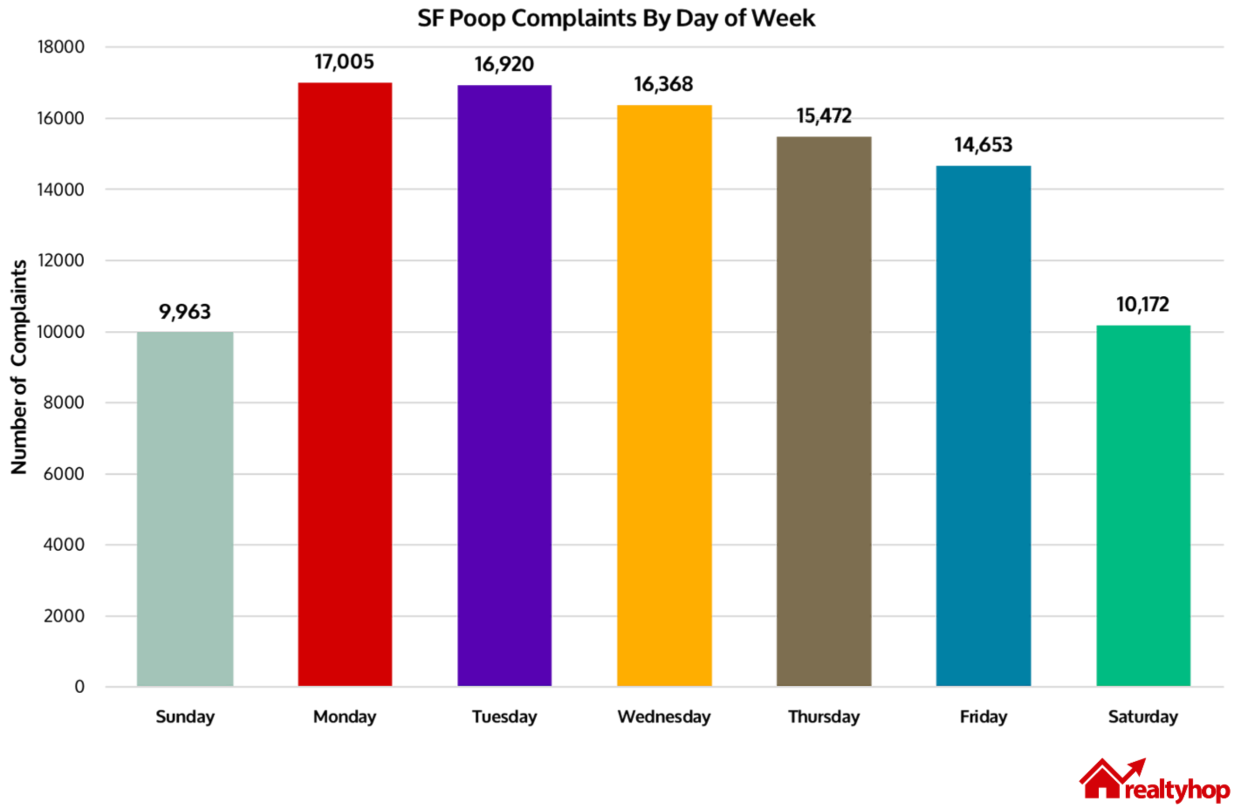


Figure 5. Realty Hop bar graph showing incidents by the day of the week.

A map from "(Human) Wasteland" depicts incidents in a different time period and displays the points as a poop emoji shown below in Figure 6. This map reduces some of the clutter by only showing incidents from January 2015 (Blum 2017). This map may not elicit the same emotional response as the "Open the Books" map but does allow the viewer to better identify areas of the reported incidents. The map in Figure 6 also follows the brown theme used by "Open the Books" and the "RealtyHop" map but takes the color one step further by using poop emojis to display these incidents. It clearly cements the theme of the map in the mind of the viewer. This map arguably displays the best use of theme of all the visualizations.

The original author of this map, Jennifer Wong, also does something the previous visualizations do not. She addresses the issues of other visualizations of this data and how others have interpreted or used these visualizations (Wong, 2018). For example, she has a marquee across the website addressing the misuse of her maps. Wong has not only stated on her website, but in interviews as well, that these visualizations only depict incidents reported through dialing 311, which are not necessarily reports of human waste. The collected waste is not tested to determine if it is from people or animals (Eskenazi, 2018). On her website, she also provides additional resources that address why there is an increasing number of human waste incidents in San Francisco including the lack of access to public toilets in the city (Swan, 2014). By providing these extra resources and alerting first time viewers to her website about the issues of the data and its representation, Jennifer Wong helps educate her viewers. Wong also uses text as an element of her visualization. By doing so, she arguably provides the most information within the map, and provides this information in an honest way. The effectiveness of these visualizations can be explored further by looking into the elements behind visualizations.

The map cannot be displayed from the (Human) Wasteland website as of October 2019.

Figure 6. The (Human) Wasteland map showing incidents in January 2015 (retrieved from Blum 2017)

Color and Color Vision Deficiency

The way something appears to us is an important factor in the human experience whether we are choosing a meal or viewing a representation of data. Color is an important element of appearance and our vision as well as an important element of creating a visualization. We must not only choose colors that are culturally and thematically appropriate, but we must also consider how the colors we choose appear to someone that views colors differently. Because color vision deficient individuals see the world, and the visualizations they are presented, in a different way, we must create visualizations with these individuals in mind.

For our audience members without color vision deficiencies, we can look to fields of study such as consumer preferences and color theory when choosing a color palette for data visualizations. For example, when choosing a salad, people tend to prefer salads with colors that are saturated, have depth, and have high contrast such as reds and greens together. Not only are high contrast colors preferred, but they also communicate higher complexity to the viewer. (Paakki, 2019). These color preferences may have root in biology (Palmer, 2010) and may be explained by color theory.

Color theory was originally proposed by Goethe who wrote the *Theory of Color*, which organized color in orders and proposed the color wheel (Margo, 2018). Color theory also suggests that when choosing color to represent a category or to flood a poster, we must also know our audience. Factors such as culture (Cotgreave, 2019) or the colors of an individual's school or favorite football team can affect the colors the viewer prefers (Palmer, 2010). Likewise, some colors may elicit unintended responses from your audience (Field, 2018) and may be offensive (Brewer, 2016).

We must also match the data represented to the types of palettes: sequential (ranging from light to dark), diverging (a midpoint color with extremes on either side), qualitative (different hues), and bivariate (two sequential palettes) (Field, 2018). For example, it would be more appropriate to use a diverging color palette over a sequential color palette to visualize a political map that displays preference over two parties.

In addition, we need to consider the needs of individuals with color vision deficiencies. Color vision deficiencies affect how hue is perceived by an individual which means color palettes will be viewed differently by those individuals. By providing sets of colors that can be discerned by those with color vision deficiencies, such as red and blue, we can hope to reach a larger audience and allow our visualizations to be as accessible as possible (Fields, 2018). If we are stuck with using palettes that are not color vision deficient friendly due to organization constraints, such as red/yellow/green, we can mix another hue such as blue into green, to make our visualizations more readable (Cotgreave, 2019). Luckily, there are many tools data visualizers can access to create color vision deficient friendly visuals such as Colorbrewer2 (2002) and tools to test those visualizations such as Colblindor (2006). Textures (lines, dots, cross hatching) can also be added in visualizations when color cannot (Wu 2018). Textures may also help those without vision feel visualizations if the visualization is created in 3D using tools such as 3D printers (Field, 2018).

Other Visual Considerations

Although color is an important element in creating a visual, there are other visual considerations that are just as important such as eye tracking, size, and font. Eye tracking takes into account the way a viewer's eyes move about a page, or in this case, a data visualization. Elements such as size, location, and images affect where the eyes and attention are drawn. Larger fonts will grab attention as well as images. The types of images we select can also determine which images people tend to focus on. For example, when viewing a visualization on workers compensation, viewers tended to focus most on an image of the human body that appeared nude over images of maps and graphs (Alberts, 2019).

Fonts can also be important in a visualization and improve readability if used correctly. Lowercase letters tend to be easier to read while uppercase letters provide distinction. Choosing a font with more space between characters may also be easier to read, but may not always be possible to use due to space constraints (Field, 2018).

Depending on your audience, you may also want to choose a font that is easier for individuals with dyslexia to read, such as Dyslexie font shown below in Figure 7. This font helps those with dyslexia by using techniques such as letter angle, large letter openings, and longer sticks (i.e. a lowercase h) to make each letter unique and harder to mistake for another letter (Dyslexie Font, 2017). Fonts like these can make visualizations more accessible to a wider audience but may not be accessible to all data visualizers due to cost and restrictions an organization must follow when presenting papers, posters, or other documents.

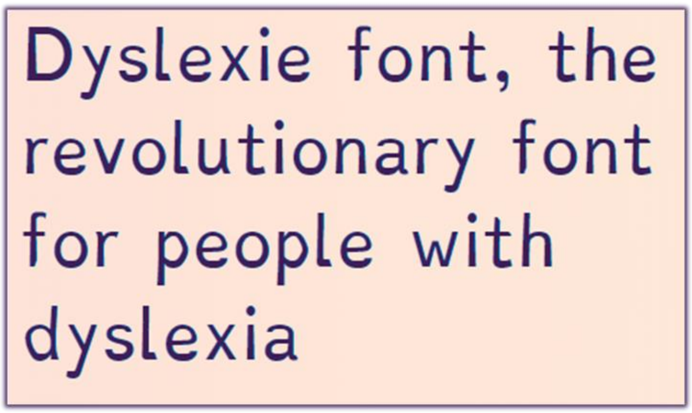


Figure 7. An example of Dyslexie font: slanted letters, larger character spaces, and longer sticks give each letter unique characteristics so that they cannot be mistaken for other letters as easily.

White space and the amount of information included are also important in a visualization. Too much information can overwhelm your audience and they may not view all aspects of your visualization (Alberts, 2019). Removing unnecessary information can reduce clutter and the appearance of a busy visualization, allowing your audience to gather important information without being overwhelmed or put-off. (Field, 2018). Designing to a grid (a framework of crossing horizontal and vertical lines) can also help organize information and reduce clutter (Cotgreave, 2019).

"So me think, why waste time say lot word when few word do trick?" -*Kevin, The Office*

Themes and Styles

Themes should also be employed when creating a visualization. Themes and styles may be specific to an organization being represented or to the specific content (Cotgreave, 2019). Aesthetic elements in a visualization such as font and color come together to create a style that fits into the theme of the data you are visualizing. These styles and themes can help elicit certain feelings from an audience (Field, 2018) and can help your audience engage with your visualization (Nelson, 2018). For example, if you were to map filming locations for Game of Thrones, then you may want to take styles from the series (Houghton, 2019) as shown below in Figure 8.

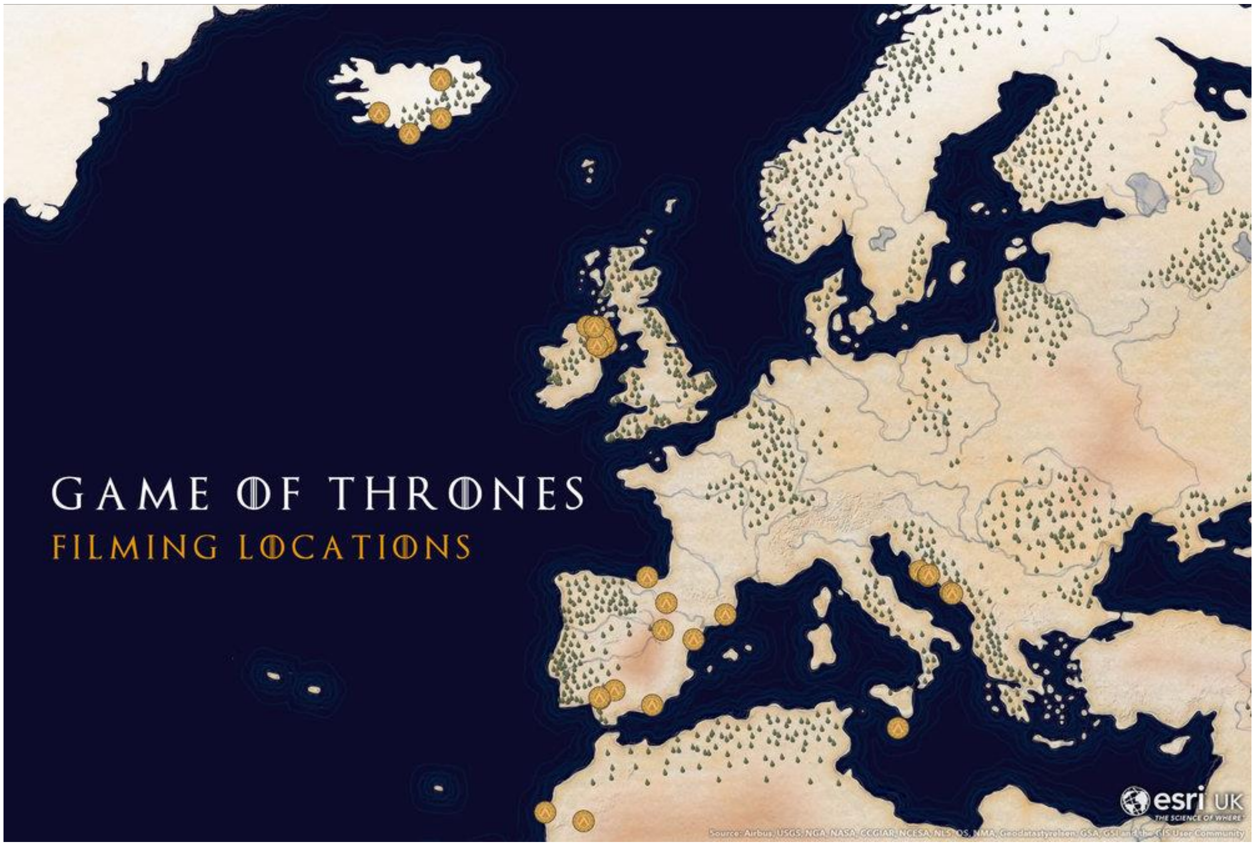


Figure 8. A map depicting Game of Thrones filming locations using styles to create a theme.

Conclusion

Color, placement of data, themes, the use of type, and white space all came together in the visualization of the San Francisco poop problem presented at the start of this section. While elements such as style help engage the viewer, other elements such as font and color allow the viewer to read your visualizations. By improving the accessibility of a visualization, you can allow your visualization to reach a wider audience. Best practices that you are familiar should be used and researching your audience and new techniques are important in creating better visualizations for each project.

Methodology

Throughout my research, I was interested in what an audience understands from a visualization. I wanted to know what my audience may see in a visualization that I don't and if there is any information that they are pulling from my visualization I did not intend–similar to the (Human) Waste Project visualization that was mis-interpreted and mis-used. To conduct this study, I created my own visualizations using GIS and chose a topic to visualize that would be important to myself and others–cancer incident rates by zip code in Illinois. To test the elements of this visualization, I provided the same information four different ways and with two different color palettes. I wanted to test a color vision deficient friendly color palette against a standard non-friendly palette. I was interested not only in color vision deficient individuals' preference, but also if non color vision deficient individuals would prefer the more accessible palette. I created my visualizations by following best practices such as creating to a grid, reducing clutter, and following a theme. I created a survey to test these visualizations and first tested my maps at the 2019 Esri User Conference. I used feedback and survey results from the conference to improve my survey and send to my final test audience by using social media.

Creating the Maps

I created maps using ArcGIS 10.6.1 to visualize cancer incidence by zip code in Illinois using data from the Illinois Department of Health. To select a color palette that was appropriate for the dataset and color vision deficient friendly I used ColorBrewer2, which provided many options. To select one of these options, I used a color palette closest to the colors of stained cancer tissue when examined under a microscope. I continued to use this color palette throughout my thesis including a poster and the surveys I created in order to follow a theme. I created four maps using the same data, but displayed by different classes and class groupings. I created a 15-class natural break (jenks) map, an 8-class quantile map, a 5-class standard deviation map, and a 2-class standard deviation map. I used the pink/purple color palette selected from ColorBrewer2 for the 15-class and 8-class maps shown below in Figure 9 and Figure 10, respectively. For the 5-class standard deviation map, I used a diverging pink/green color palette selected from ColorBrewer2 shown in Figure 11 and selected my own color palette of purple/off-white for the 2-class standard deviation map shown in Figure 12. I also re-colored the 15-class and 8-class maps with a standard Esri red/yellow/green palette to use in my survey shown in Figure 13 and Figure 14, respectively. Step-by-step instructions on how to create these maps are shown in Appendix 1.

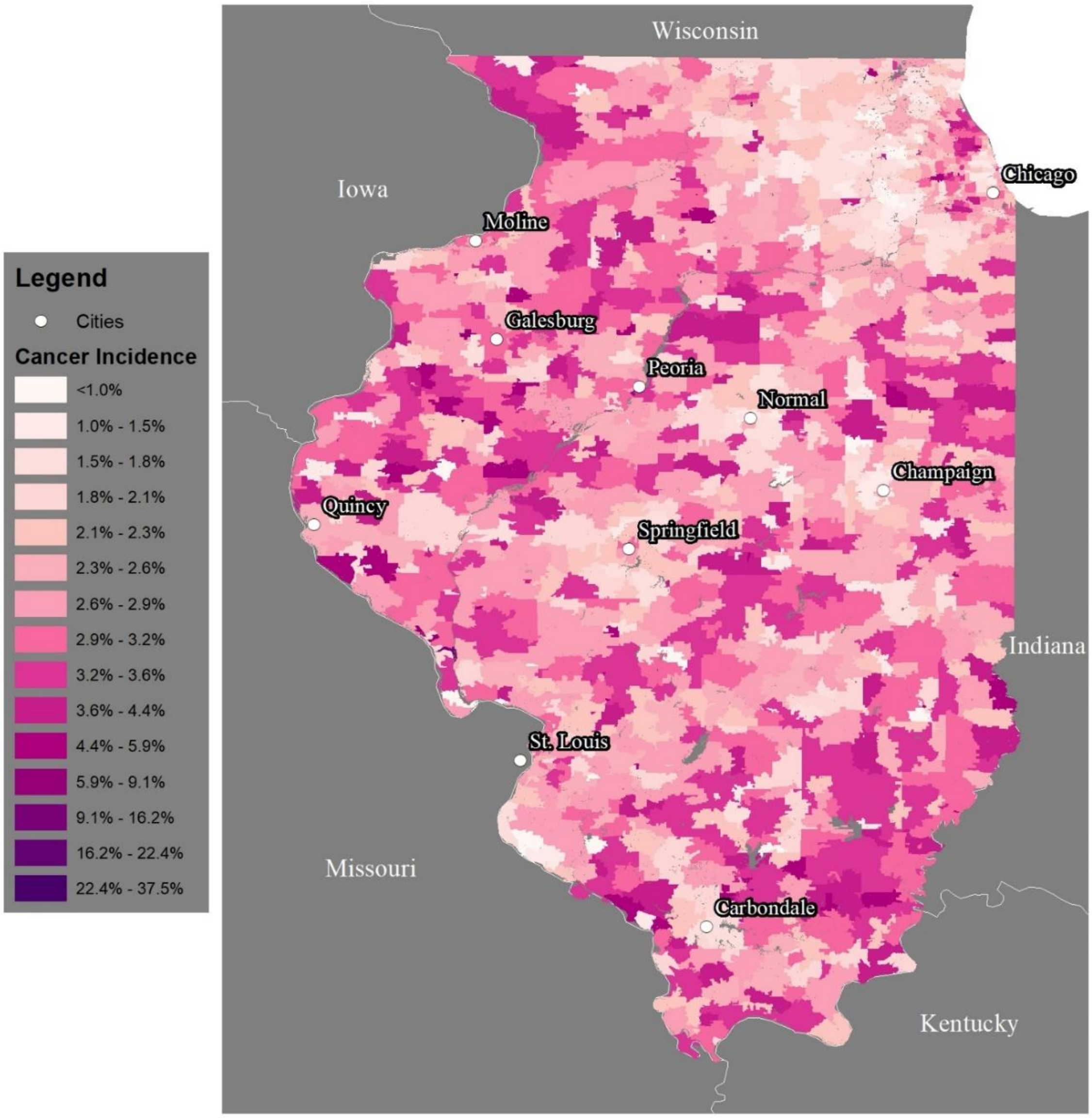


Figure 9. The 15-class natural break (jenks) pink/purple map I created using ArcMap 10.6.1.

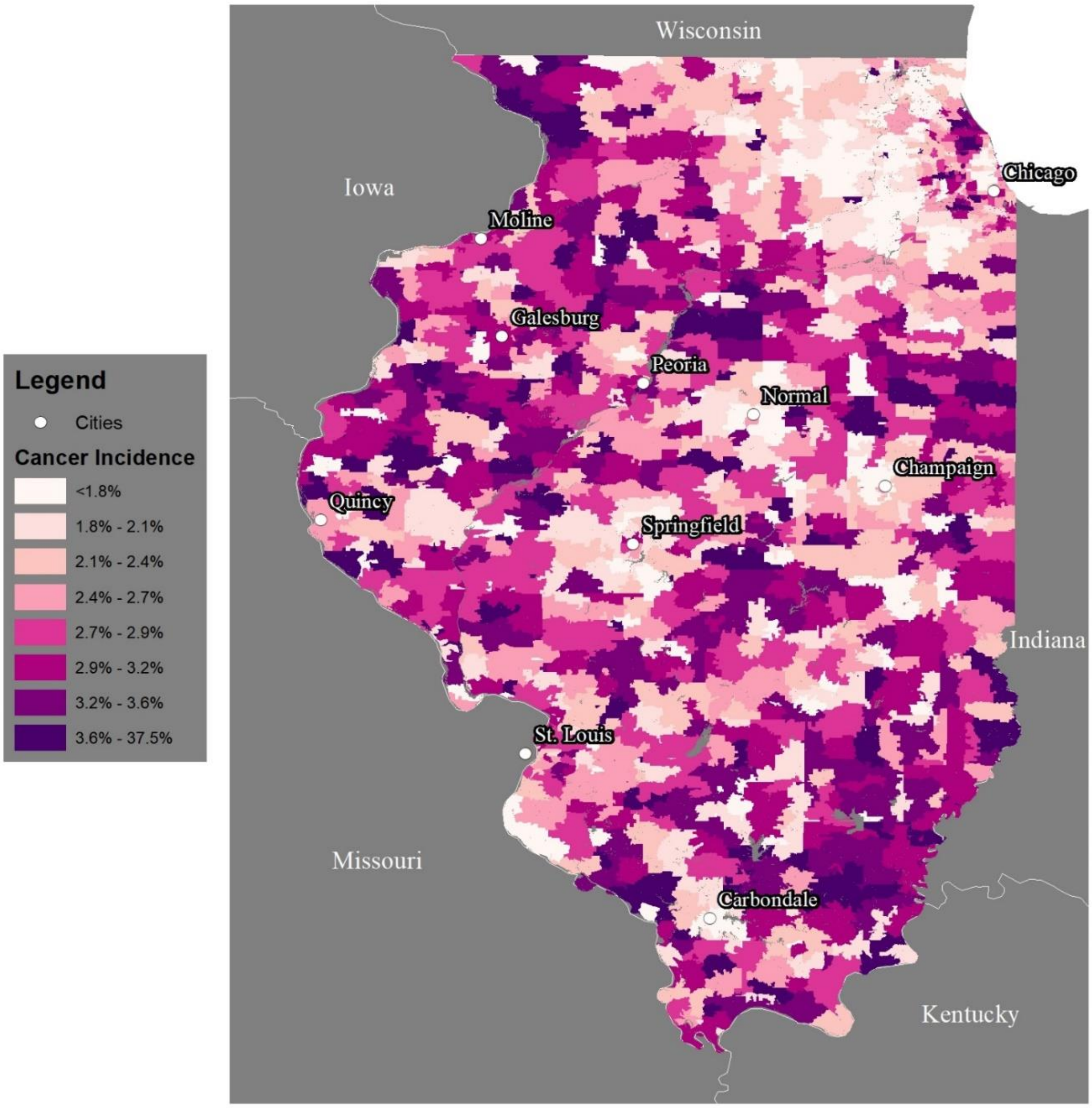


Figure 10. The 8-class quantile pink/purple map I created using ArcMap 10.6.1.

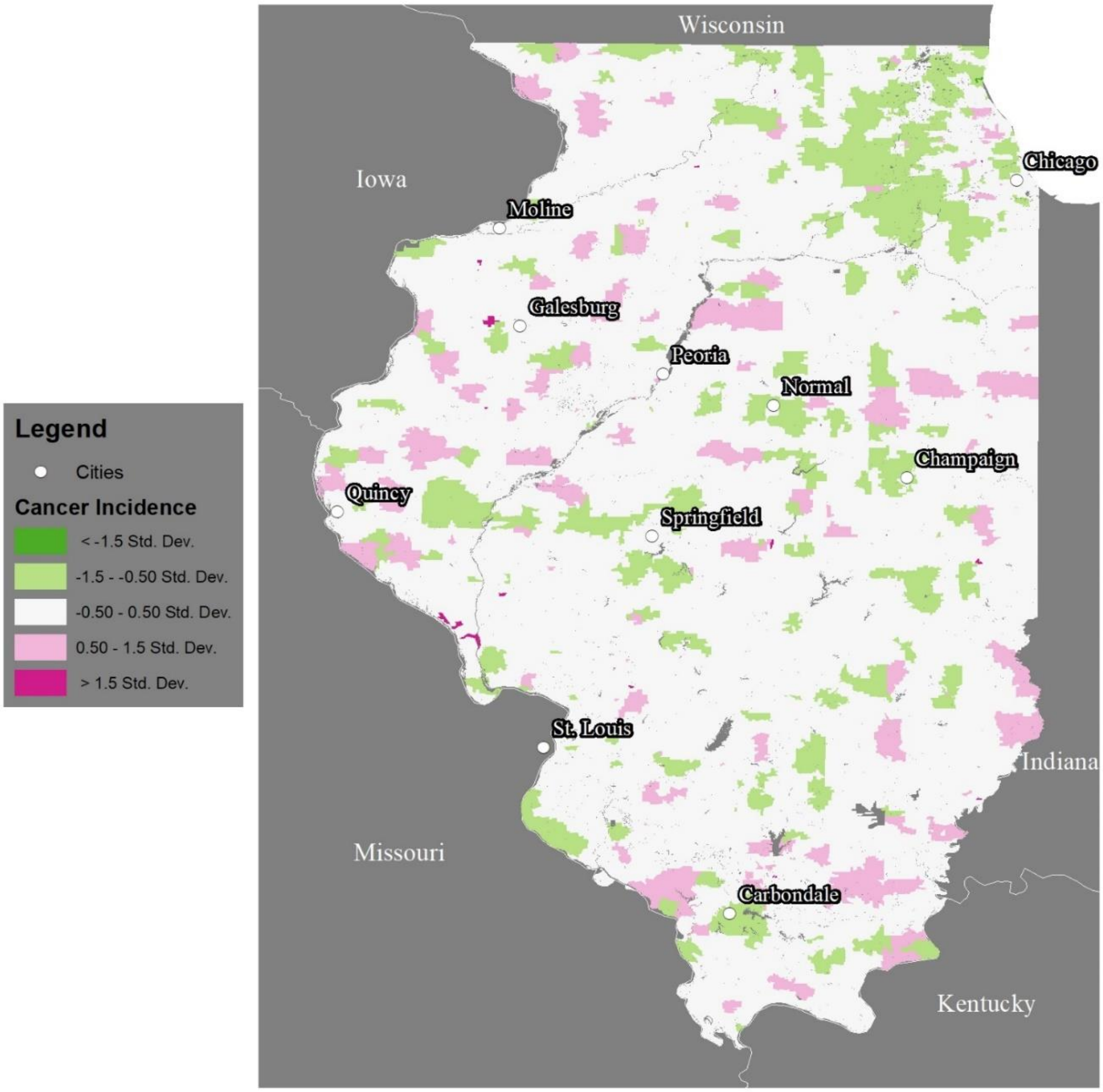


Figure 11. The 5-class standard deviation map I created using ArcMap 10.6.1.

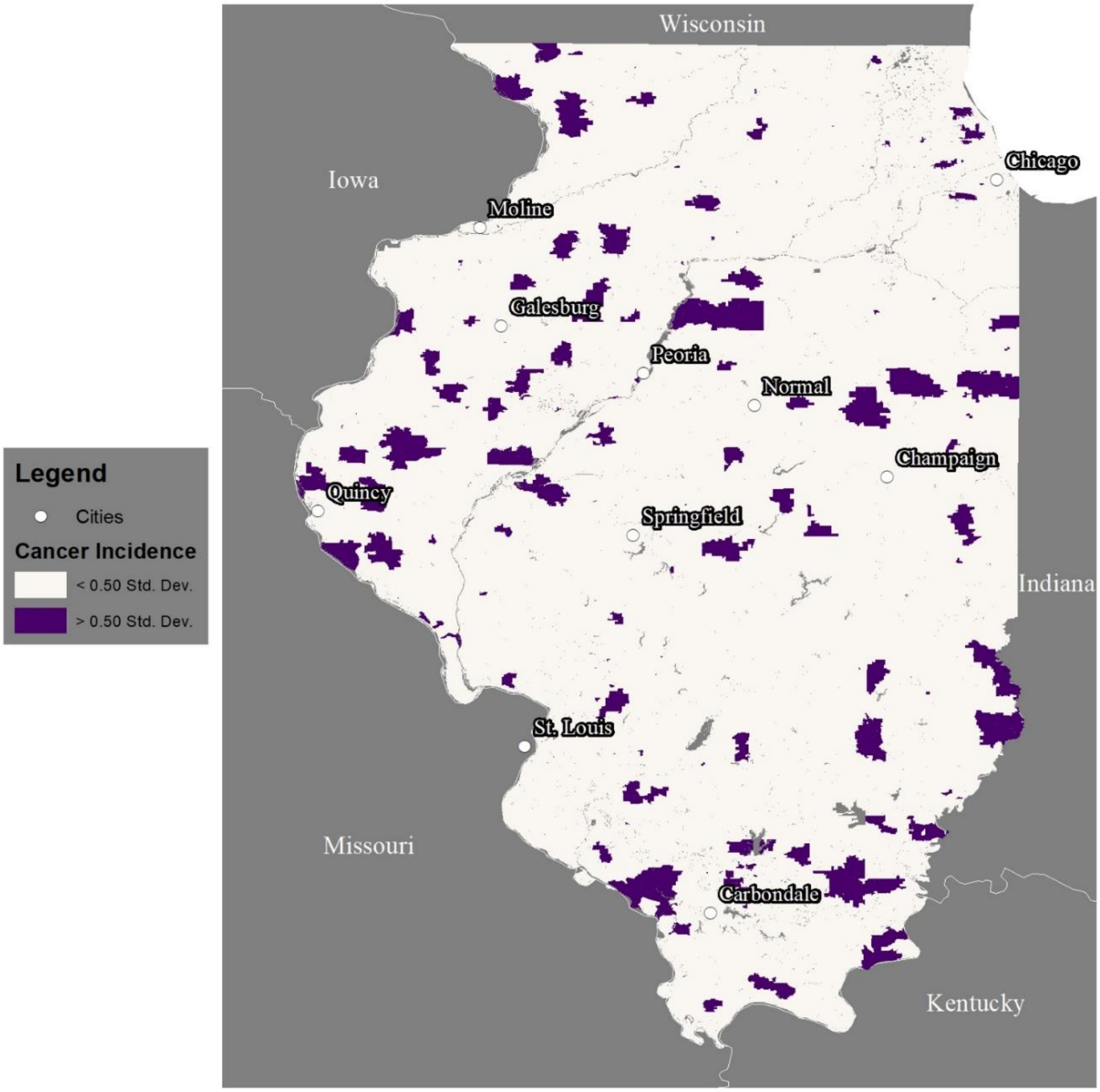


Figure 12. The 2-class standard deviation map I created using ArcMap 10.6.1.

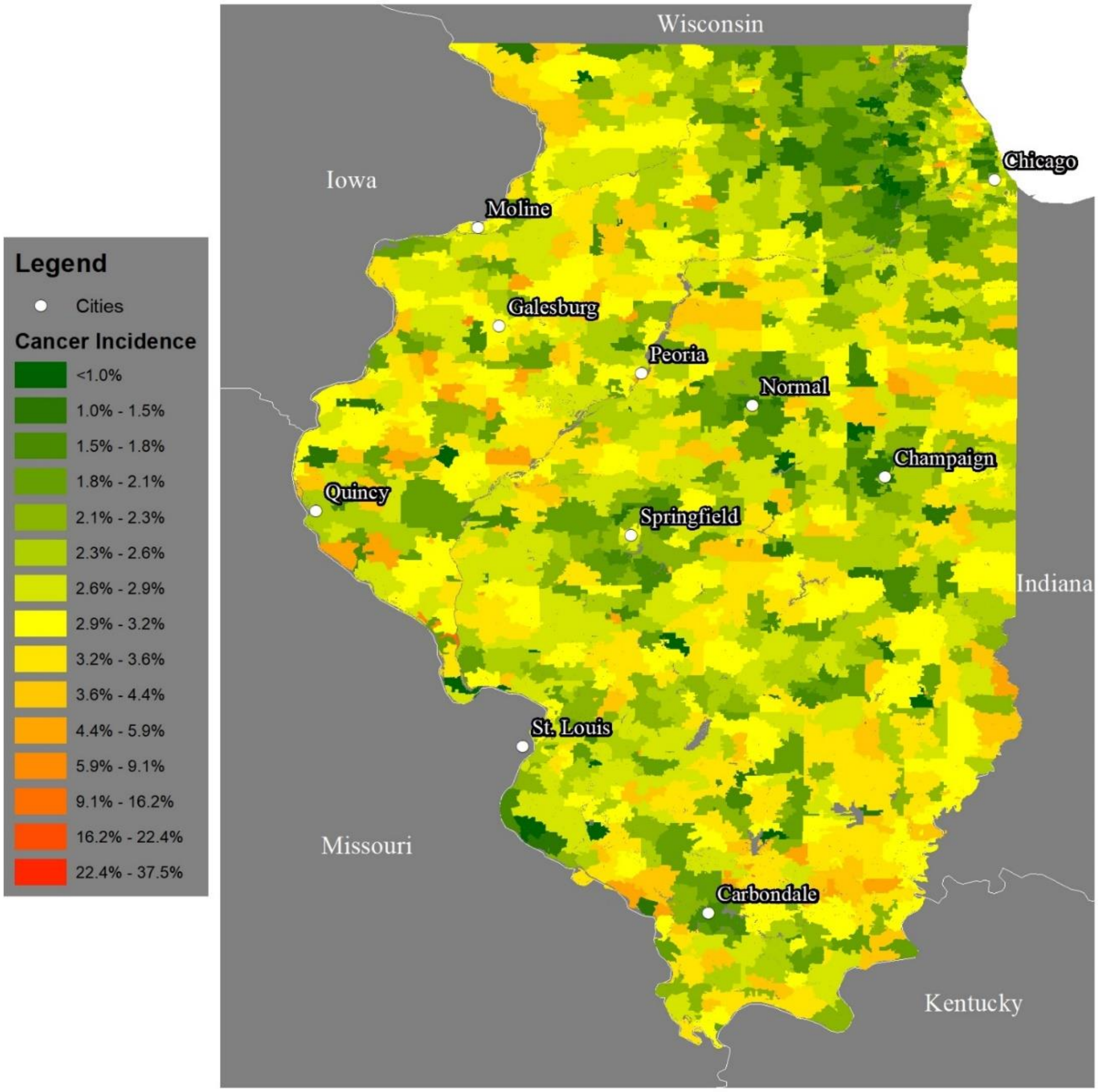


Figure 13. The 15-class natural break (jenks) red/yellow/green map I created using ArcMap 10.6.1.

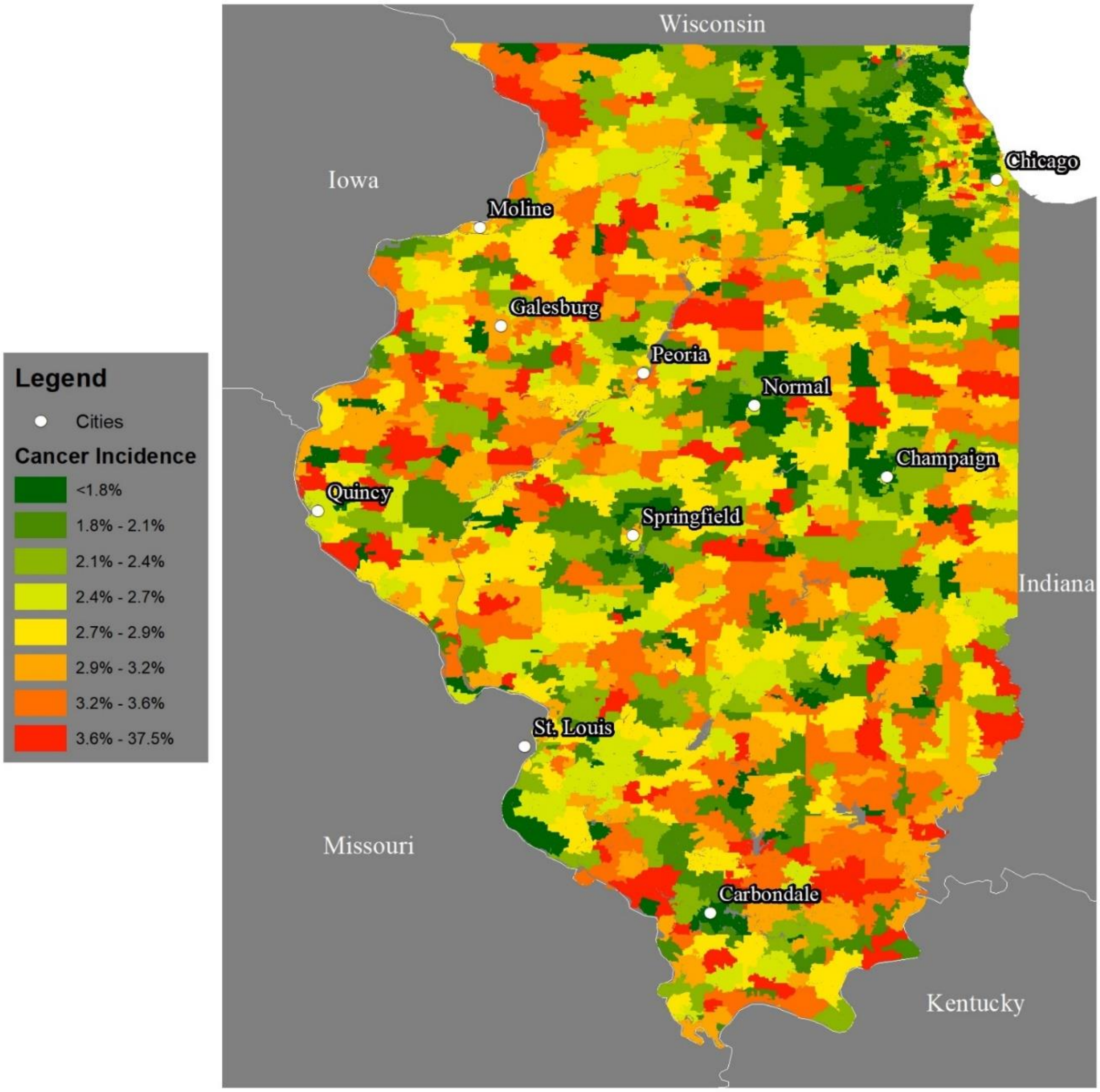


Figure 14. The 8-class quantile red/yellow/green map I created using ArcMap 10.6.1.

Survey Methods-PreTest Phase

After completing my maps, I created a poster to display at the 2019 Esri User Conference in San Diego, California. To keep my thesis theme, I used the image of invasive lobular carcinoma from John Hopkins Pathology as the background image and used pinks and purples for text, boxes, and a custom QR code. I created the custom QR code using QR Code Generator to select for a rounded style with a purple frame and pink edges. I kept the poster simple with little text and kept the focus of the poster on the maps and the QR code which linked to my survey. I included the 15-class natural break (jenks) map, 8-class quantile map, 5-class standard deviation map, and 2-class standard deviation map in the poster. I also included a call out text to highlight one of the zip codes with a cancer rate above the standard deviation shown below in Figure 15.

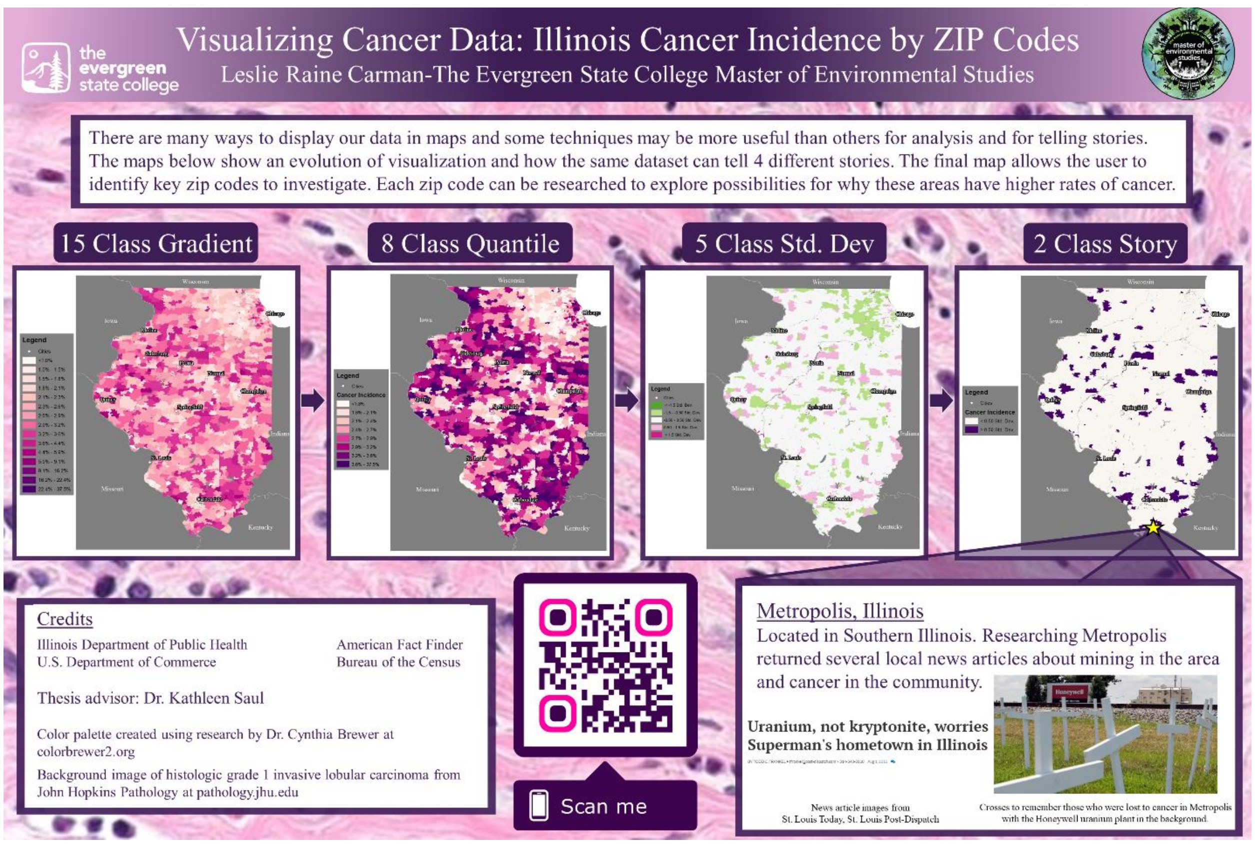


Figure 15. The poster I created to display at the 2019 Esri User Conference in San Diego, California

I had created business cards with the black QR code, shown in Figure 16 below, to distribute at the conference to get more survey responses from respondents at other conference events who may not have attended the poster session. The business cards also made it easier for people to scan the code and complete the survey between or after conference events.

Figure 16. Business cards I created with a custom QR code generated by QR Code Generator to elicit survey responses from conference attendees.

I provided business cards in a pencil bag as well as paper versions of the test survey in file folders next to my poster. The paper version of the survey was provided to elicit survey responses from conference attendees who may not have or preferred not to use mobile devices, shown in Figure 17 below. I also tested my poster’s readability for color vision deficient individuals using Colblindor. The poster viewed through different types of color vision deficiency are shown in Appendix 2.

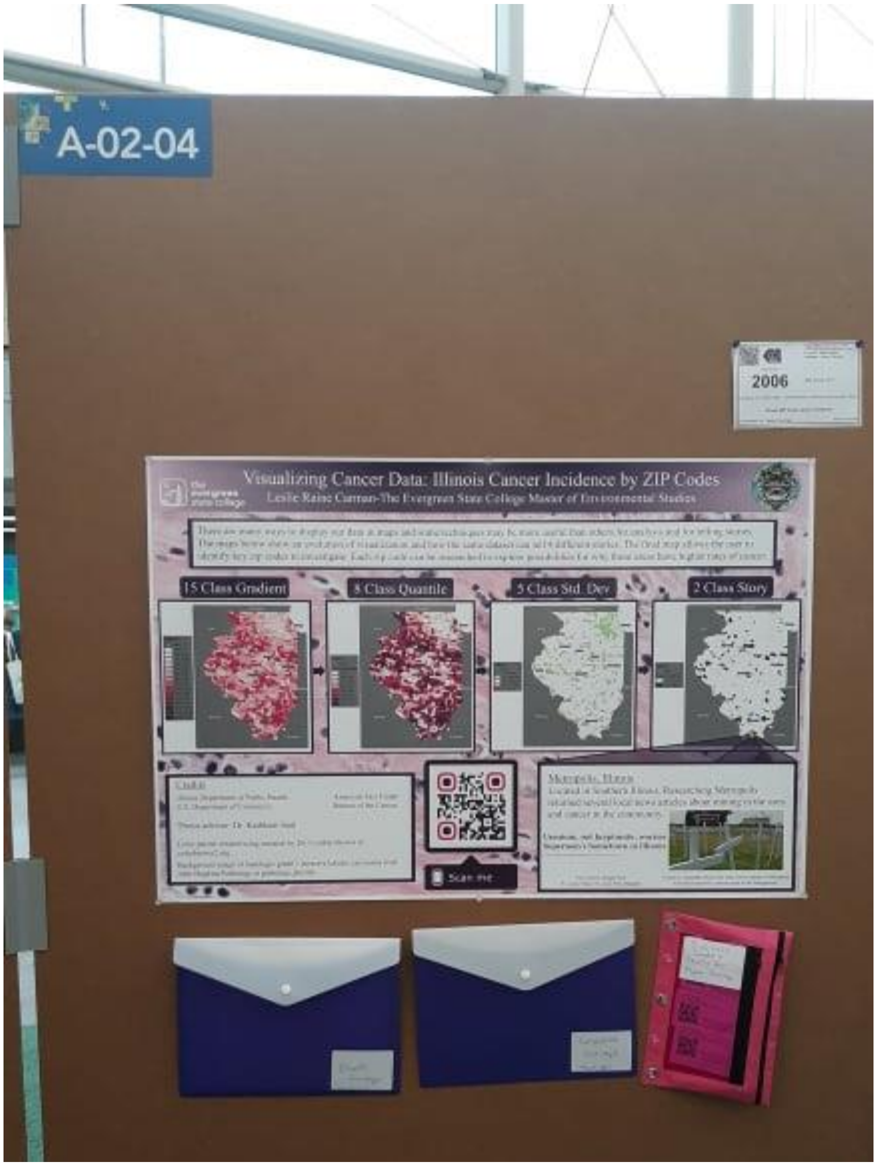


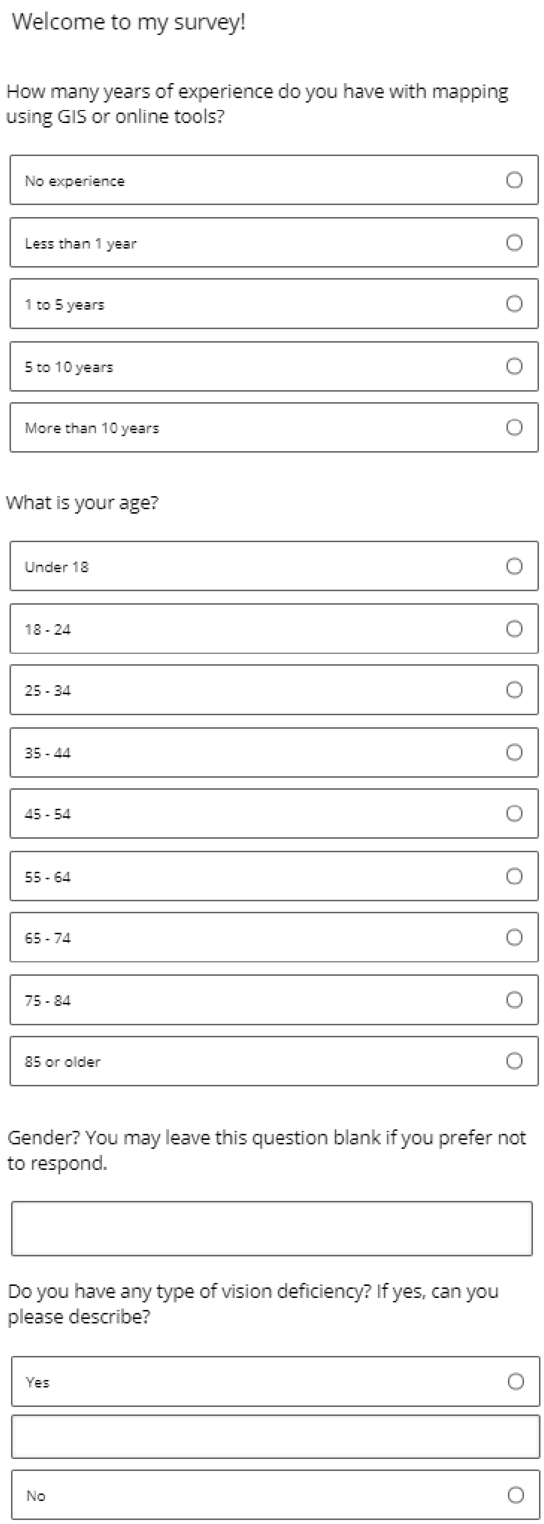
Figure 17. My poster displayed at the 2019 Esri User Conference with paper copies of the survey in the left file folder, an empty file folder to the right for completed surveys, and a pencil bag with pencils to use for the paper surveys. The pencil bag also contained extra business cards for respondents to take with to complete the survey when convenient.

To create both the test and final surveys, I used Qualtrics. I started the test survey with demographic questions, then asked respondents about vision deficiencies, and their history with and knowledge of Illinois. I then moved into the map-related questions: I asked respondents what they learned from the 15-class natural break (jenks) map, the 8-

class quantile map, the 5-class standard deviation map, the 2-class standard deviation map, the 15-class natural break (jenks) red/yellow/green map, and the 8-class quantile red/yellow/green map. I was able to improve my survey in real-time based on the survey responses and face-to-face conversations and completed a final survey improvement after the conference. The final and improved survey questions were then saved as a new survey in Qualtrics.

The Final Survey

To elicit surveys for the improved and final survey, I created public posts on Facebook and LinkedIn asking respondents to take my survey and share the survey link with others. The questions shown in Figures 18-27 below show all the questions that final survey respondents could have received. Some questions had display logic and would only display for respondents that selected certain answers. For example, the questions, "Can you read this map? Please provide any details you would like to share." would only display if the respondent selected "Yes" to a previous question, "Do you have any type of vision deficiency? If yes, can you please describe?"



Welcome to my survey!

How many years of experience do you have with mapping using GIS or online tools?

- No experience
- Less than 1 year
- 1 to 5 years
- 5 to 10 years
- More than 10 years

What is your age?

- Under 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 - 84
- 85 or older

Gender? You may leave this question blank if you prefer not to respond.

Do you have any type of vision deficiency? If yes, can you please describe?

- Yes
- No

Figure 18. Page 1 of the sent survey. By leaving the gender question as a text box answer, the question is more inclusive. The final question, "Do you have any type of vision deficiency?", has display logic so that vision related questions are displayed to a "Yes" answer later in the survey.

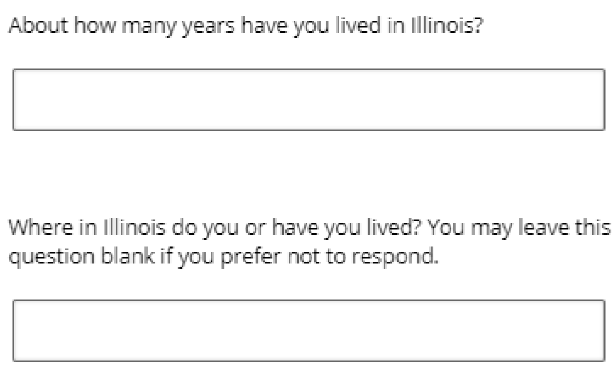

How much would you say you know about the state of Illinois?

- Nothing at all
- Very little
- Somewhat
- Quite a bit
- A great deal

Do you currently or have you ever lived in Illinois?

- Yes
- No

Figure 19. Page 2 of the sent survey. The final question, "Do you currently or have you ever lived in Illinois?", has display logic so that Illinois related questions on the next page of the survey are displayed to a "Yes" answer. Respondents that selected "No" did not see the next page of questions.

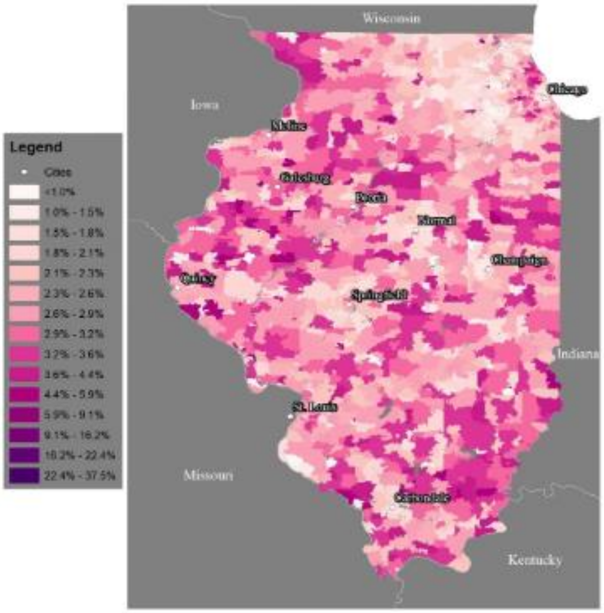

About how many years have you lived in Illinois?

Where in Illinois do you or have you lived? You may leave this question blank if you prefer not to respond.

Figure 20. Page 3 of the sent survey. This page only displayed to respondents that selected "Yes" to the previous question, "Do you currently or have you ever lived in Illinois?"

This is the fun part of the survey! You will see 6 maps followed with 1 to 3 questions about each map.

This map shows the percentage of cancer incidence by zip code.



What conclusions do you draw from this map?

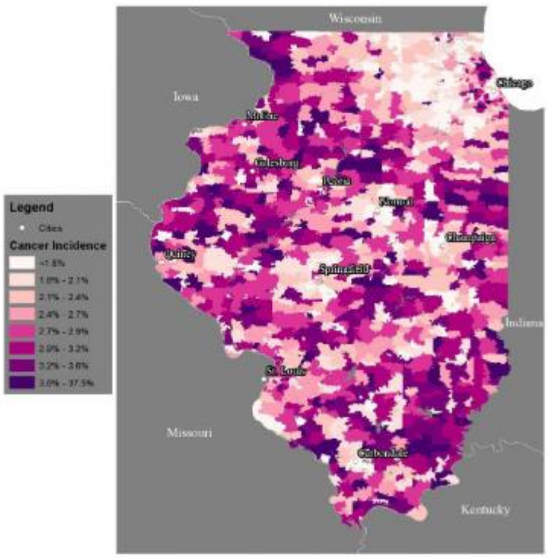
Did anything surprise you? If yes, please describe.

- Yes
- No

Can you read this map? Please provide any details you would like to share.

Figure 21. Page 4 of the sent survey asking respondents about the 15-class natural break (jenks) pink/purple map. The final question, "Can you read this map?", only displayed to respondents that selected "Yes" to the previous question "Do you have any type of vision deficiency?"

This map shows the percentage of cancer incidence by zip code.



What conclusions do you draw from this map?

Did anything surprise you? If yes, please describe.

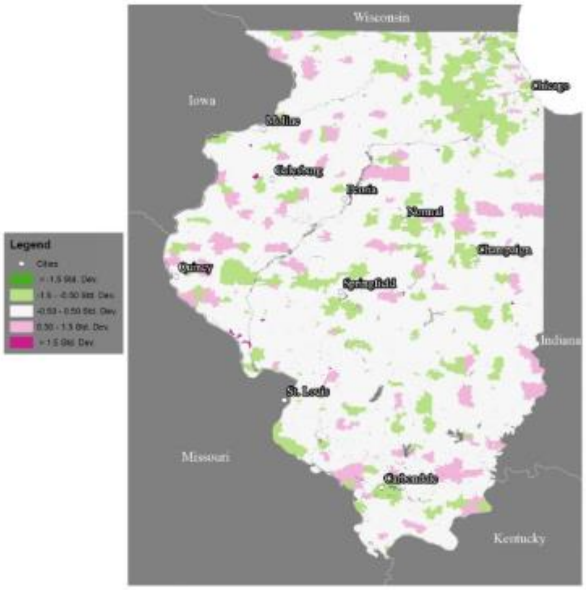
Yes

No

Can you read this map? Please provide any details you would like to share.

Figure 22. Page 5 of the sent survey asking about the 8-class quantile pink/purple map. The final question, "Can you read this map?", only displayed to respondents that selected "Yes" to the previous question, "Do you have any type of vision deficiency?"

This map shows cancer incidence by zip code using standard deviation. Green indicates zip codes below the standard deviation and pink indicates zip codes above the standard deviation.



What conclusions do you draw from this map?

Did anything surprise you? If yes, please describe.

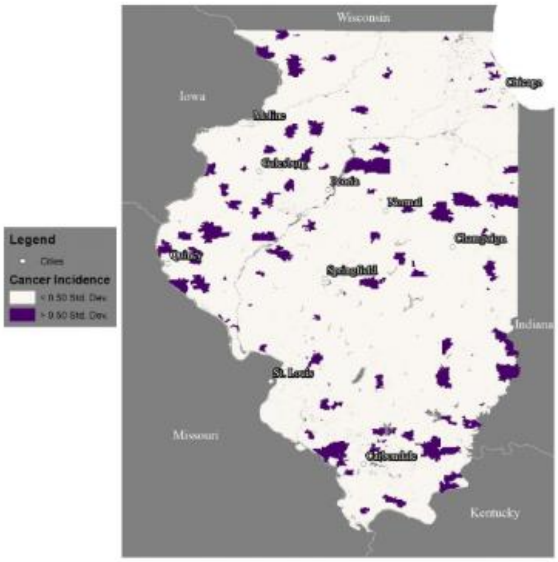
Yes

No

Can you read this map? Please provide any details you would like to share.

Figure 23. Page 6 of the sent survey asking respondents about the 5-class standard deviation map. The final question, "Can you read this map?", only displayed to respondents that selected "Yes" to the previous question, "Do you have any type of vision deficiency?"

This map shows cancer incidence by zip code using standard deviation using two categories instead of five. Purple areas indicate zip codes with cancer rates above the standard deviation.



What conclusions do you draw from this map?

Did anything surprise you? If yes, please describe.

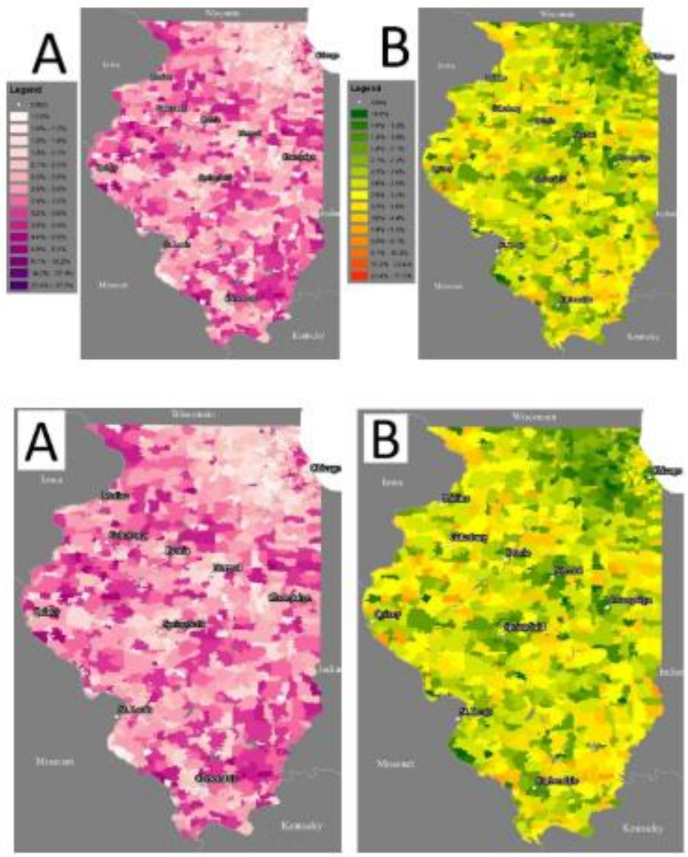
Yes

No

Can you read this map? Please provide any details you would like to share.

Figure 24. Page 7 of the sent survey asking respondents about the 2-class standard deviation map. The final question, "Can you read this map?", only displayed to respondents that selected "Yes" to the previous question, "Do you have any type of vision deficiency?"

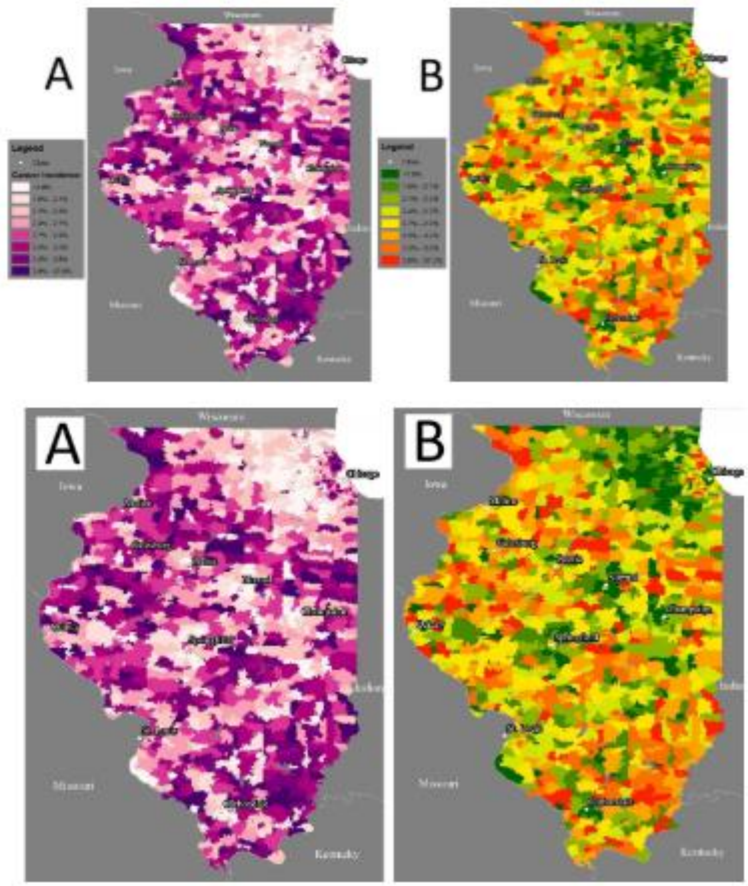
These maps show the percentage of cancer incidence by zip code. The top set shows the legend, and the bottom set is a larger version for easier viewing.



Which of these maps do you prefer and why?

Figure 25. Page 8 of the sent survey asking respondents to choose between the 15-class natural break (jenks) pink/purple map and the 15-class natural break (jenks) red/yellow/green map. I provided a smaller version with legends and a larger version without legends for easier viewing in the Qualtrics survey format.

These maps show the percentage of cancer incidence by zip code. The top set shows the legend, and the bottom set is a larger version for easier viewing.



Which of these maps do you prefer and why?

Figure 26. Page 9 of the sent survey asking respondents to chose between the 8-class quantile pink/purple map and the 8-class quantile red/yellow/green map. I provided a smaller version with legends and a larger version without legends for easier viewing in the Qualtrics survey format.

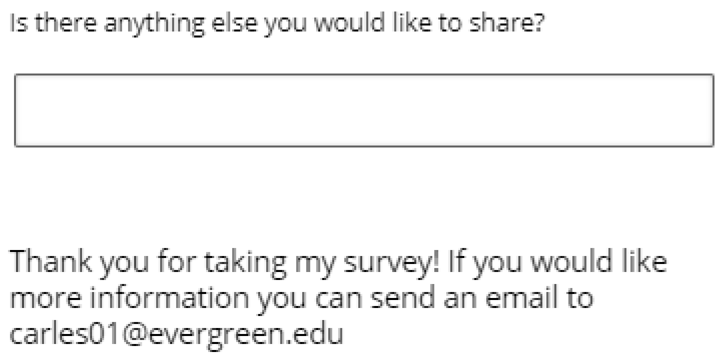


Figure 27. Page 10 of the sent survey allowing the respondents to share any information that was not covered with the survey questions.

I visualized and analyzed my survey results using Tableau which required downloading the table from Qualtrics and preparing the table before uploading in Tableau. Because Qualtrics is a more complex survey tool and may not widely be available due to cost, I provided step-by-step instructions in Appendix 3 on how to clean the exported Qualtrics survey results in Excel. Cleaning the exported results makes it easier to use data visualization tools such as Tableau. Some institutions have a Qualtrics/Tableau bridge in order to avoid this step, but this feature is not widely available due to cost.

Results and Discussion

The following results are from the final sent survey sent to respondents through Facebook and LinkedIn. Results from the survey used at the Esri conference are not included. The results from the Esri survey were used to improve this final survey.

Demographics

Out of 61 responses, half of the survey respondents had no experience with mapping using GIS or online tools. The respondents ranged in age from under 18 to the 65-74 age bracket with most respondents in the 25-34 age bracket (32.7%). All age brackets up to age 74 were represented in the survey with no respondents aged over 74. The question asking respondents about gender was a text box so that all spectrums of gender would be captured. Of the respondents that reported gender, only female and male and variations (such as f or m) were reported. Of these respondents, 79.6% were female and 20.4% were male. These demographics were visualized using Tableau 2019.2 as shown in Figure 28 below.

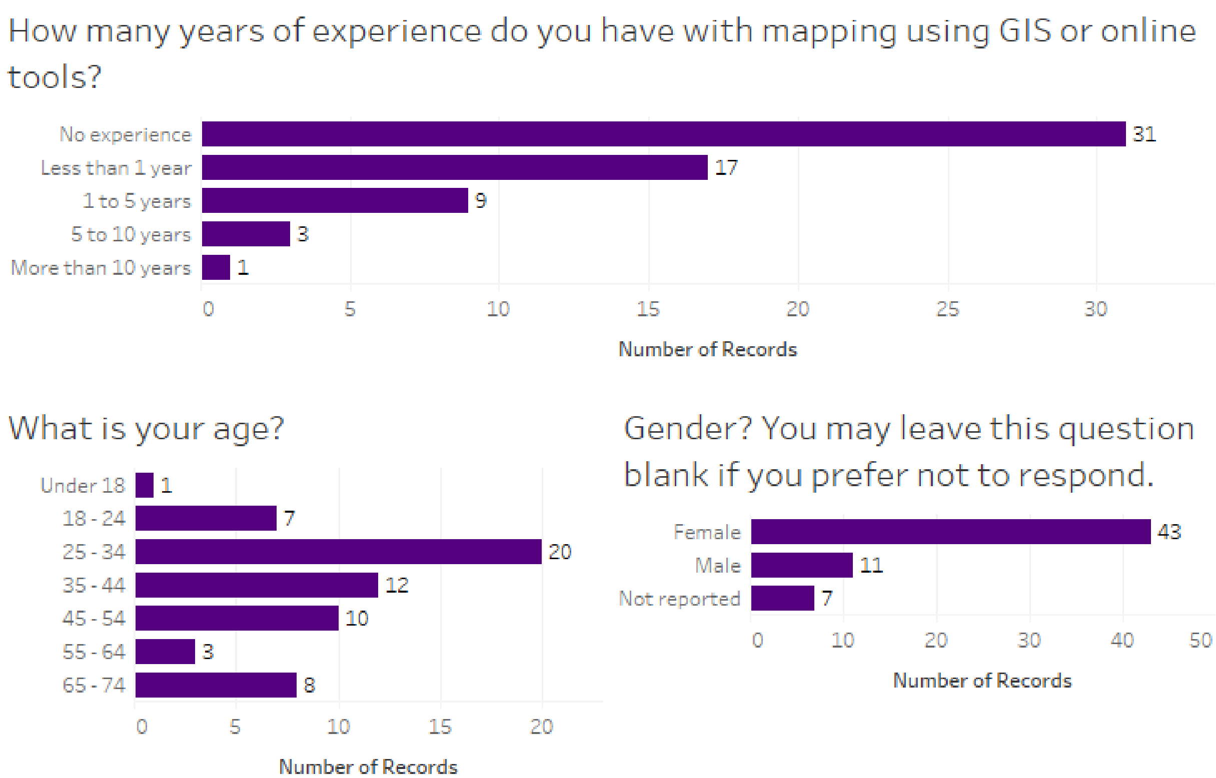
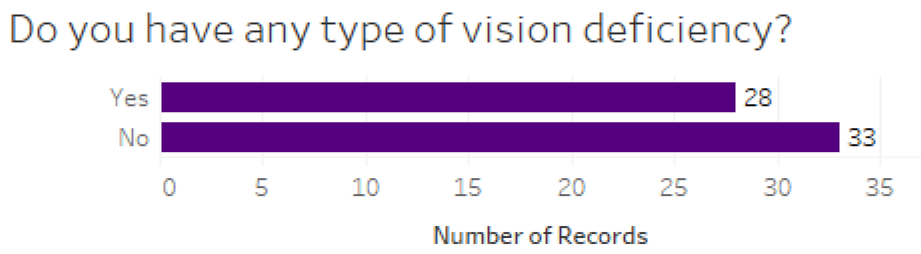


Figure 28. Survey respondent demographics displayed using Tableau 2019.2. Half of respondents had no experience with GIS, most respondents were aged between 25 and 34, and most respondents were female.

Vision Deficiency

I asked respondents about vision deficiencies and asked respondents to describe their vision deficiency. 45.9% of respondents reported a vision deficiency. Of the 28 respondents that reported a vision deficiency, 23 described their vision deficiency, which included glasses and contacts, astigmatism, near-sightedness, far-sightedness, one respondent with lazy eye, and one respondent with cataracts. I asked this vision question hoping to have respondents with color vision deficiency so that I could see if there was a difference between which map respondents preferred based on whether they had color vision deficiency or not. The red/yellow/green maps would not be as easy to read to those with color vision deficiency. However, no respondents that took the survey reported color vision deficiency. It is likely that the ratio of female to male respondents contributed to the lack of color vision deficient individuals that participated in the survey as men are more likely to have color vision deficiencies than women. Color vision deficiency also only affects 4% of the population (Field, 2018) and a larger group of respondents would have increased the likelihood of some respondents with color vision deficiency. These results are shown below in Figure 29.



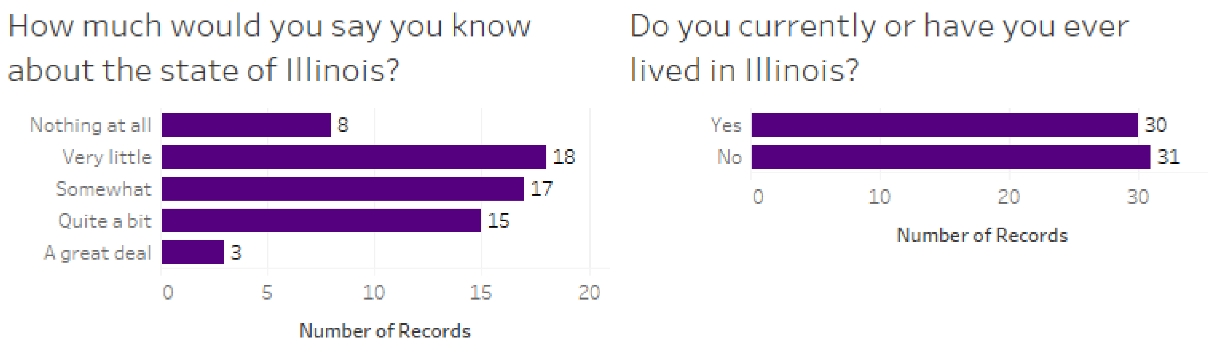
If yes, can you please describe?

Astigmatism
Astigmatism, lazy eye
Bifocals
Cataracts, near-sighted, astigmatism
constant wear of glasses
Contact lens wearer
Far sighted
far sighted, astigmatism
Farsided
Glasses
Glasses/contacts/astigmatism
Near sighted
Near sighted w/astymatism
Near sighted, astigmatism
Near sightedness
Nearsighted
Nearsightedness
Need to wear glasses
Needs glasses/contacts
Short-sighted
Slightly far sighted-reading/computer glasses needed
Use glasses
Wear glasses

Figure 29. Survey respondents' vision deficiencies displayed using Tableau 2019.2. 28 respondents reported a vision deficiency and 23 of those respondents described their deficiency. No respondents reported color vision deficiency.

Illinois Knowledge and Residency

Because the maps were of Illinois, I asked respondents about their knowledge of Illinois and if they have ever lived in Illinois. Most respondents (86.9%) had at least some knowledge of Illinois. Of the respondents that had not lived in Illinois, most had at least some knowledge of Illinois with 58.1% having very little knowledge of Illinois. Of the respondents that had lived in Illinois, most reported somewhat and quite a bit of knowledge of Illinois, 43.3% and 46.7% respectively. Most respondents that had lived in Illinois, had been there for 30 to 40 years (30% of respondents). These results are shown below in Figure 30.



Knowledge and Residency

Illinois Knowledge	Illinois Residency Yes	No
Nothing at all		8
Very little		18
Somewhat	13	4
Quite a bit	14	1
A great deal	3	

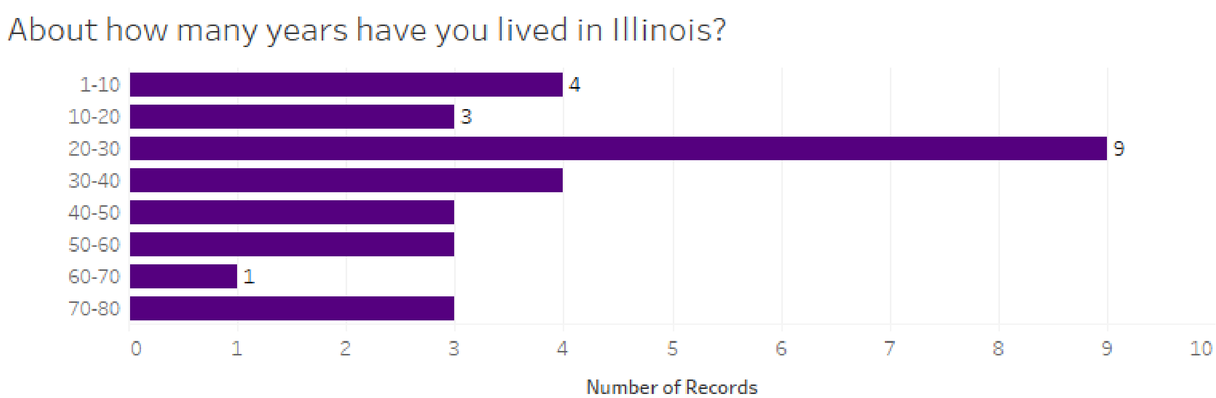


Figure 30. Survey respondents' knowledge and residency in Illinois displayed using Tableau 2019.2. Most respondents had at least some knowledge of Illinois with half of the respondents residing in Illinois at some points in their lives.

I also asked respondents where in Illinois they had lived or currently were living. The survey respondents were spread throughout Illinois. Many respondents gave multiple answers with some respondents reporting city and others reporting county. Some respondents gave vague answers such as "northwest". These vague responses were not visualized in Figure 31 below. In future studies, it may be beneficial to ask for specific answers.

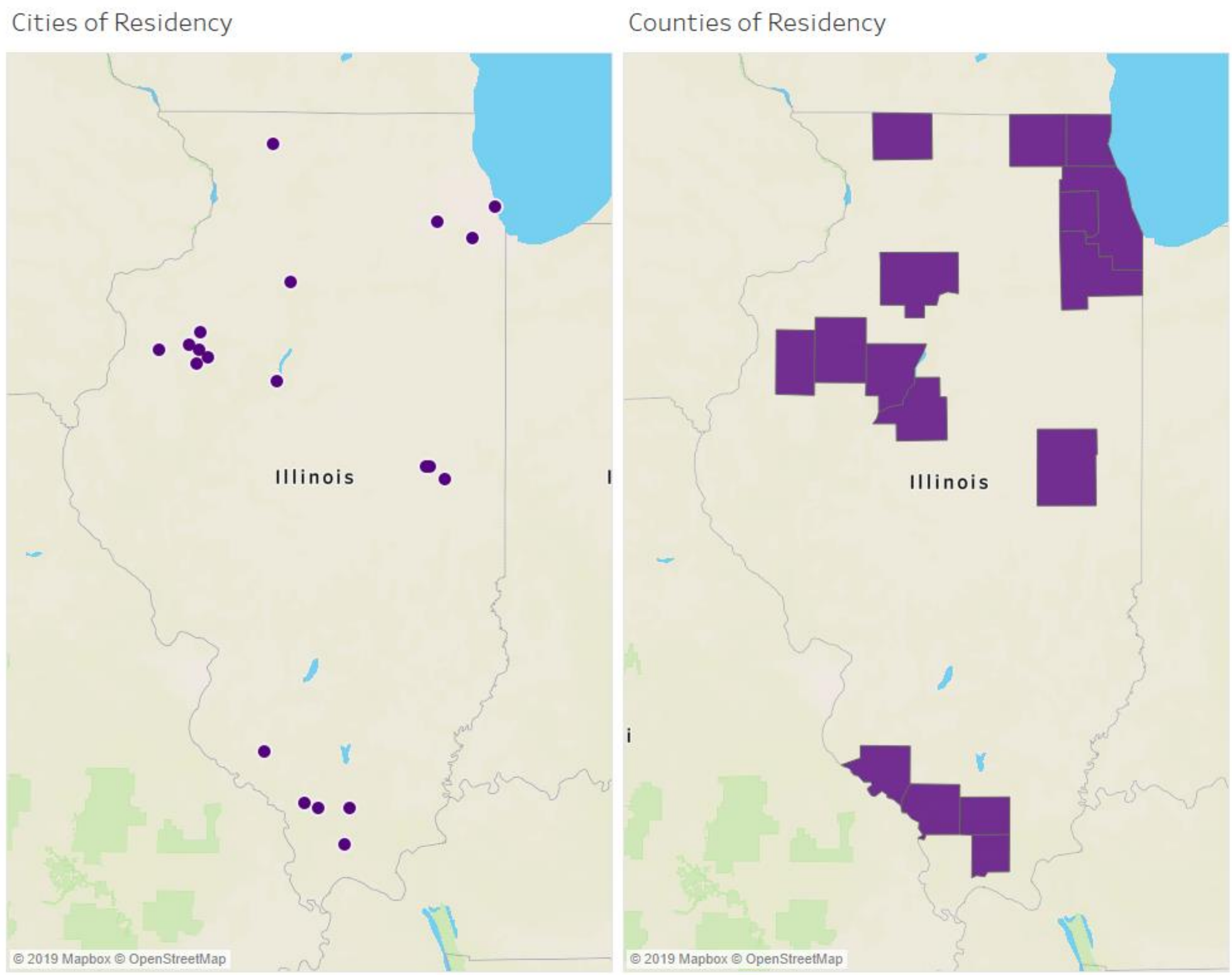


Figure 31. Two maps of where survey respondents currently live or had lived in Illinois using Tableau 2019.2. Some respondents reported city while others reported county. The county map also includes the city responses by displaying the county of the city that respondents reported.

Maps

I asked respondents the same two to three questions about four different maps. Respondents who selected "Yes" to the question, "Do you have any type of vision deficiency?”, were prompted to answer three questions per map while the other respondents replied to two.

1. What conclusions do you draw from this map?
2. Did anything surprise you? If yes, please describe.
3. (For those with a vision deficiency) Can you read this map? Please provide any details you would like to share.

For the first map, the 15-class natural breaks (jenks) pink/purple map, 50 respondents answered the first question, “What conclusions do you draw from this map?” Answers to this question were similar, with a general sentiment that there were high rates of cancer, yet some respondents saw patterns in the data while other saw no pattern at all. One respondent commented that the map had a camouflage-like pattern. Three respondents indicated that the patterns they saw indicated farming, rivers, and agrochemicals might have caused higher incidences of cancer, resulting in their interest in buying organic produce. It was interesting to see some respondents come to conclusions as to why there may be higher rates of cancer. Examining the zip codes with the highest rates of cancer indicated that rivers may be a factor as well as industry such as mining. Some respondents sought more information on groupings such as income, population, rural vs. urban, or age. Population had been accounted for in the analysis, but this information was not shared with the respondents. I intentionally included as little information as possible as to my methods to see what respondents saw in the information. Responses such as this one indicate that when sharing information with the public it is best to assume that some individuals will not understand all methods, and some text may be needed to accompany a visualization in order to explain similar information. Many respondents indicated that they noticed that urban areas had lower rates of cancer and some indicated surprise at this observation. As the person creating these visualizations, I was also initially surprised by this observation. Others commented that the size of the zip code area might have affected the visual representation. This is another area that the survey indicated text may benefit a visualization if it is to be shared with the public. Health data such as cancer incidence is sensitive information and is often difficult to get access to. When researching this topic, it was difficult to find data more granular than county. Zip code data was the most granular data that I could get access to, and I could only find zip code data that was publicly available for two states, Illinois and New York. One respondent with no GIS experience even stated that it seemed that cancer rates seemed to be tied to zip code. This response also indicates that additional text would be beneficial as this was a map that examined cancer incidence by the reported zip code. Overall, survey response attention focused on the darker purple areas with higher rates of cancer.

Twenty-six respondents answered the second question, “Did anything surprise you?”, for this map. The answers were similar to the first question, but mostly focused on the urban vs. rural divide, with respondents stating that they saw lower incidence in urban areas and higher incidence in rural areas. Fourteen respondents commented on the low rates in the Chicago suburbs. I was also surprised by the low rates in the Chicago suburbs when creating these maps.

Eight respondents answer the third question, “Can you read this map?”. Respondents indicated that they could read the map with other respondents indicating that they thought they could read it, but that the question made them doubt themselves. Those who doubted themselves had no experience with GIS. This question was not intended to make respondents doubt themselves but was intended for respondents with color vision deficiencies. For future studies, it may be beneficial to have a check box of vision deficiencies with color vision deficiency as an option so that this question only displays for those individuals. For reference, this map is shown in Figure 32 below.

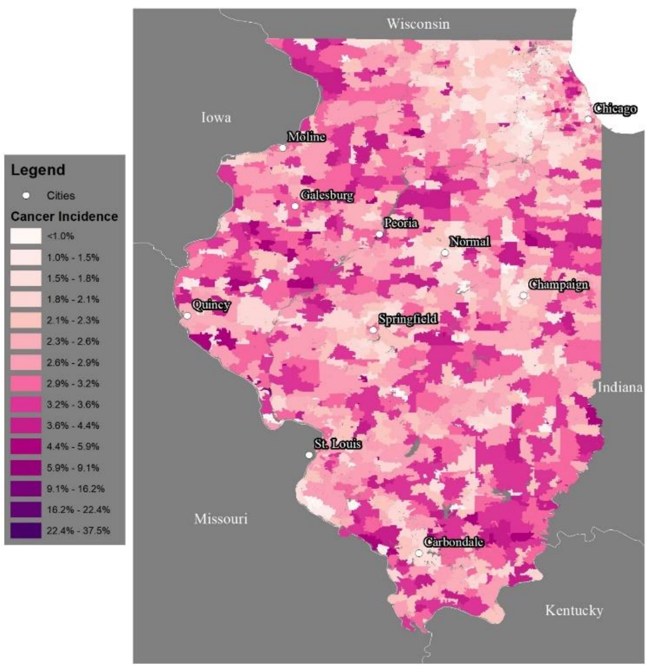


Figure 32. 15-class natural break (jenks) pink/purple map

For the second map, the 8-class quantile pink/purple map, 40 respondents answered the first question, “What conclusions do you draw from this map?” The answers were similar to the answers of the first map, with more respondents commenting on the map pattern. A few of the respondents commented on the reduced number of classes and the change in color. One respondent asked why the maps looked different when they were said to show the same information, this respondent had no experience with GIS. Another respondent with no GIS experience commented on specific cities and areas in Illinois and guessed that wealthier, non-agricultural areas had no agricultural pesticides and therefore lower rates of cancer. This survey respondent also guessed that the darker areas may have chemical companies. When researching these dark areas, chemical companies were present. This was the first time a chemical company was discussed; for the first map, only agricultural practices had been mentioned.

Nineteen respondents answered the second question, “Did anything surprise you?”, for this map. The answers to this question were similar to the questions for the previous map although one respondent indicated that this 8-class map was easier to read. This is not surprising as this map had fewer classes which is recommended in cartography and was a quantile map, which creates a more visually interesting map. Another respondent felt this map was trying to trick them with the change in number and size of classes. Interestingly, this respondent had 1-5 years of experience with GIS.

Four respondents answered the third question, “Can you read this map?”. Two respondents said yes, one was interested in land use and more information on agricultural areas, and one wanted the population to be included in the map. Again, population information had been used in the analysis, but this was not shared with the respondents. For reference, this map is shown in Figure 33 below.

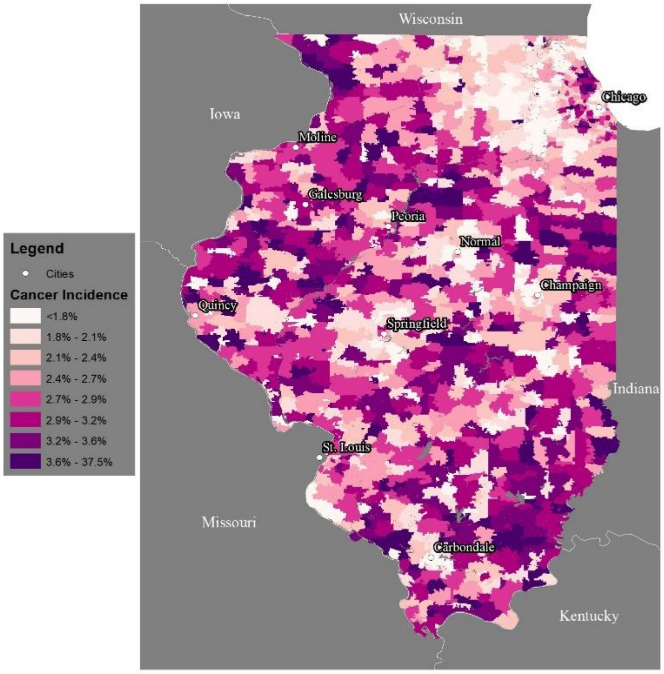


Figure 33. 8-class quantile pink/purple map.

For the third map, the 5-class standard deviation pink/purple map, 35 respondents answered the first question, "What conclusions do you draw from this map?" The answers to this question were similar to previous answers with some respondents indicating that they did not understand the map. These respondents had no experience and just 1-5 years of experience using GIS. I was surprised that respondents with some GIS experience did not understand this map. I believe that the confusion from this map stemmed from the standard deviation where some respondents may not have an understanding of statistics or the language used in statistics. I intentionally did not share text information about statistics for this map but these responses indicate that if maps like these are to be shared with the public, then text information about the statistics would be helpful. One respondent stated that statistics are not facts, this respondent had no experience with GIS.

Eleven respondents answered the second question, “Did anything surprise you?”, for this map. The answers to this question were similar to the previous questions in the survey. One respondent stated that the cancer didn’t seem as bad with this map compared to the other maps; this respondent had 1-5 years of experience using GIS. This response is not surprising as this map has more white space than the previous maps.

Four respondents answered the third question, “Can you read this map?”, for this map. Three respondents said yes; however, the last respondent felt that all of the maps were only good for conversation and research starters. This respondent had no experience with GIS. This respondent is not wrong, these maps are intended as a tool to research areas with higher cancer rates. For reference, this map is shown in Figure 34 below.

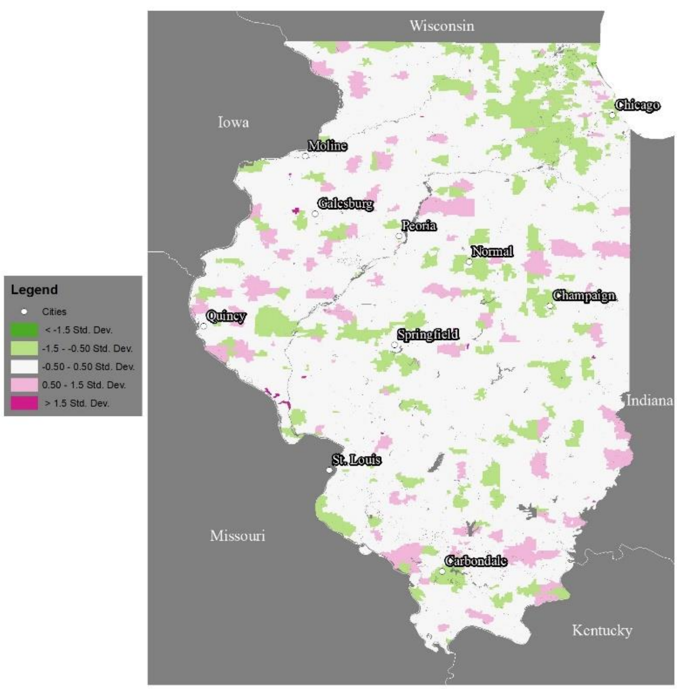


Figure 34. 5-class standard deviation map.

For the fourth map, the 2-class standard deviation pink/purple map, 33 respondents answered the first question, “What conclusions do you draw from this map?”

The answers to this question were mostly similar to the previous questions in the survey. More respondents were confused by this map as with the previous standard deviation map and many respondents noted that the cancer rates seemed less severe. One respondent suggested people reduce smoking and eat healthier foods. This respondent had 1-5 years of experience using GIS. This map did not provide any data on health habits of Illinois residents, but this topic could be explored further if researching why these areas have higher cancer rates. Another respondent remarked that data could be manipulated based on personal agendas. This respondent had no experience with GIS. While the data in these maps was not changed, it can be visualized in different ways to tell different stories, which was the purpose of the thesis. One respondent noted that there was less detail and color in this map while others indicated that they did not like this reduction in detail while others were able to pull out areas with cancer rates above the standard deviation. This map is intended to identify the areas with the highest cancer rates in order to research why these zip codes have higher rates of cancer. It is not the most appealing map but is a great research tool.

Five respondents answered the second question, “Did anything surprise you?”, for this map. The answers to this question were similar to the previous questions in the survey.

Two respondents answered the third question, “Can you read this map?”, for this map. Both said yes with one commenting that this map was their least favorite. This respondent had no experience with GIS and likely would not be familiar with using GIS as a tool for research. However, this map was not created to be the most appealing. The 8-class quantile map was intended for this purpose, so this response is not surprising. It is also not surprising that fewer respondents answered the last questions. This indicates that respondents were experiencing survey fatigue and shorter surveys may yield different results. For reference, this map is shown in Figure 35 below.

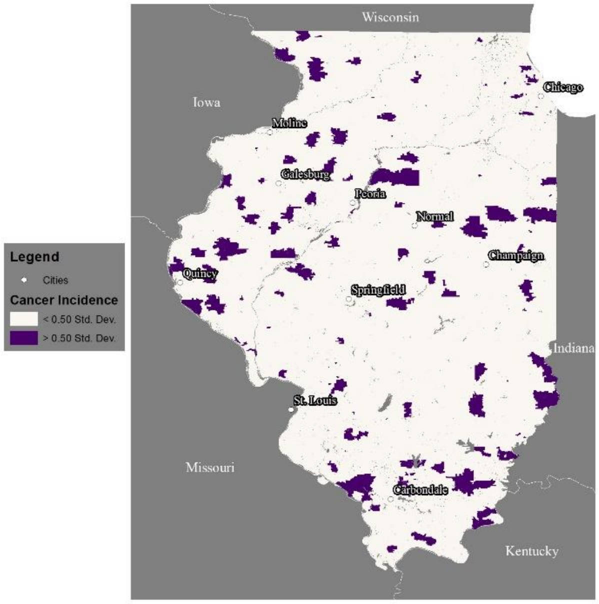
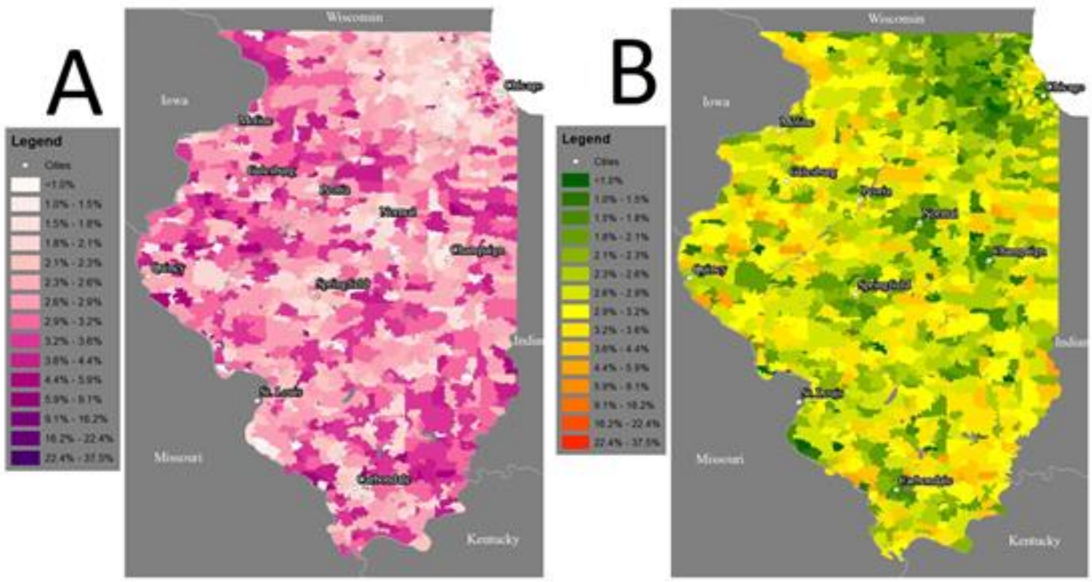
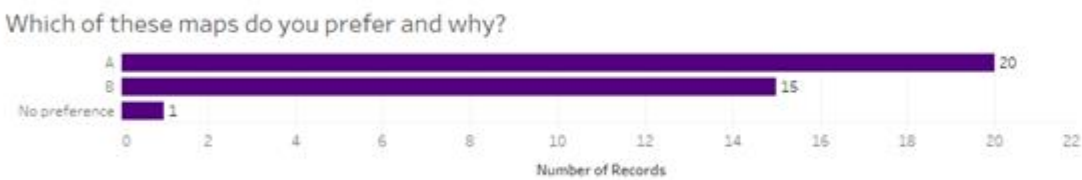


Figure 35. 2-class standard deviation map.

Preferred Map

I asked respondents to choose between similar maps symbolized with different color palettes. Thirty-six respondents answered this question for the first set of maps, shown in Figure 36: the 15-class natural break (jenks) pink/purple and red/yellow/green maps. Most respondents preferred the pink/purple map (55.6%) with under half of the respondents preferring the red/yellow/green map (41.7%). Most respondents with no GIS experience preferred the pink/purple map (58.8%), all respondents that had 5 to 10 and more than 10 years of experience using GIS also preferred the pink/purple map. More respondents with less than one year of experience using GIS preferred the red/yellow/green map (54.5%) and respondents with 1 to 5 years of experience using GIS were divided evenly between the pink/purple and red/yellow/green maps. This was surprising initially as I expected respondents with no GIS experience to prefer the red/yellow/green maps. The red/yellow/green palette is not an appropriate color palette for sequential data, but those with a few years of GIS experience may prefer this palette because they have not learned this yet and the red/yellow/green palette is standard when learning GIS. These responses indicate that the red/yellow/green color palette preferences are due to the teaching style of tools such as GIS and not due to a natural preference for these colors. Younger respondents were more likely to prefer the red/yellow/green map while older respondents were more likely to prefer the pink/purple map. Respondents that were 35-44 largely preferred the red/yellow/green map over the pink/purple map (85.7%). It is unclear why there may be this divide in age and color preference, but shared events in time, such as childhood, may affect this preference. Both male and female respondents preferred the pink/purple map over the red/ yellow/ green map with 54.2% of female respondents and 85.7% of male respondents. This was surprising as I expected male respondents may reject a pink color palette. These results were visualized using Tableau 2019.2 shown in Figure 36 below.





15-Class Natural Break Preferred Map and GIS Experience

	Natural Break Preferred Map		
GIS Experience	A	B	No preference
No experience	10	6	1
Less than 1 year	5	6	
1 to 5 years	3	3	
5 to 10 years	1		
More than 10 years	1		

15-Class Natural Break Preferred Map and Age

	Natural Break Preferred Map		
Age	A	B	No preference
Under 18		1	
18 - 24	4	1	
25 - 34	7	5	
35 - 44	1	6	
45 - 54	4	2	
55 - 64	1		
65 - 74	3		1

15-Class Natural Break Preferred Map and Gender

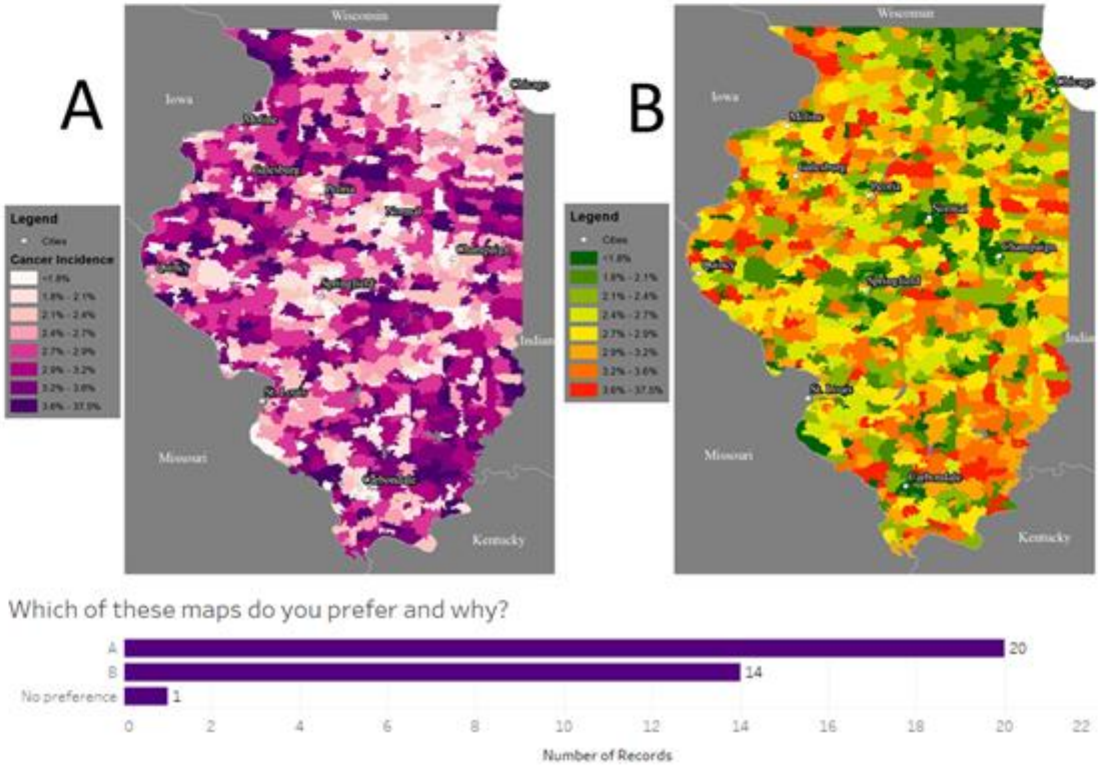
	Natural Break Preferred Map		
Gender	A	B	No preference
Female	13	11	
Male	6	1	1
Not reported	1	3	

Figure 36. Survey respondents' color palette preference displayed using Tableau 2019.2. Most respondents preferred the pink/purple map.

Thirty-five respondents answered this question for the second set of maps, those with the 8-class natural break (jenks) pink/purple and red/yellow/green maps. Most respondents preferred the pink/purple map (57.1%) with under half of the respondents preferring the red/yellow/green map (40%).

Half of the respondents with no GIS experience preferred the pink/purple map while the only respondent with more than 10 years of experience using GIS preferring the red/yellow/green map. I expect that this preference may be due to using red/yellow/green palettes as a default in GIS. The other respondents with less than 1 year, 1 to 5 years, and 5 to 10 years of experience using GIS preferred the pink/purple map (60%, 66.7%, and 100%, respectively).

The only respondent that was under 18 preferred the red/yellow/green map and most respondents in the 35-44 age bracket preferred the red/yellow/green map as well (71.4%). Respondents in the 45-54 and 65-74 age brackets were divided on their preference of maps. All other ages preferred the pink/purple map with 100% of those aged 18-24, 58.3% aged 25-34, the only respondent aged 55-64. Both male and female respondents preferred the pink/purple map over the red/yellow/green map with 54.2% of female respondents and 57.1% of male respondents. One respondent had a strong reaction to the red/yellow/green map stating that the map looked like “Digiorno Supreme Pizza Barf”. I agreed with this assessment. These results were visualized using Tableau 2019.2 shown in Figure 37 below.



8-Class Quantile Preferred Map and GIS Experience

	Quantile Preferred Map		
GIS Experience	A	B	No preference
No experience	8	7	1
Less than 1 year	6	4	
1 to 5 years	4	2	
5 to 10 years	1		
More than 10 years		1	

8-Class Quantile Preferred Map and Age

	Quantile Preferred Map		
Age	A	B	No preference
Under 18		1	
18 - 24	5		
25 - 34	7	5	
35 - 44	2	5	
45 - 54	2	2	
55 - 64	1		
65 - 74	2	1	1

8-Class Quantile Preferred Map and Gender

	Quantile Preferred Map		
Gender	A	B	No preference
Female	13	11	
Male	4	2	1
Not reported	2	1	

Figure 37. Survey respondents' color palette preference displayed using Tableau 2019.2. Most respondents preferred the pink/purple map.

Six respondents changed which map they preferred between the purple/pink palette and the red/yellow/green palettes when choosing a preference in the 15-class then

8-class maps. Three respondents preferred the pink/purple color palette (map A) for the 15-class natural break (jenks) map but preferred the red/yellow/green palette (map B) for the 8-class quantile map. Three respondents preferred the red/yellow/green color palette (map B) for the 15-class natural break (jenks) map but preferred the pink/purple color palette (map A) for the 8-class quantile map. This result was surprising as I expected color preference to be the same for both maps.

Conclusion

Future Study

If I could conduct this study again, I would host a focus group to gain a more in-depth understanding and detail from participants on what they understood from the maps, what they liked or didn't, and why. The focus group would remove the survey fatigue element from the study and allow participants to build their responses from each other. This could reduce confusion that some participants experienced in answering some of the questions. For example, some participants misunderstood the question in my survey, "Which of these maps do you prefer?" and stated that they preferred the maps with legends. While these types of answers highlight areas where the survey could be cleaner, they did not provide additional input as to which map color palette was preferable.

A future study could also provide an additional story map that draws out the stories behind every zip code with cancer rates above the standard deviation, similar to the callout I created in my poster (Figure 15). Many respondents did not like the 2-class map but presenting this map as a story map–it's intended purpose–may yield different results. Respondents are also likely to have a different set of reactions to a story map than those reactions elicited from the static maps.

Final Conclusions

An audience has a wide variety of knowledge and backgrounds but there are tools and methods that we can employ to make visuals more understandable. While the first step in the process is generally to know your audience, we must also understand that we can never truly know everything about our audience. There are so many factors and experiences that each individual participant or audience member brings to the table that we cannot always account for. We ourselves are also people with biases and experiences that shape our opinions and views of our audience and influences how we visualize the data we share. These opinions we hold will never fully be true and so we must do our best to create visuals to the best of our abilities. We must employ techniques that we are familiar with and work diligently to continue to learn and grow as researchers and visualizers to best serve our clients and the public.

The test of a truly exceptional data visualizer is not only determined by the knowledge they have, but also by the knowledge they continue to grow. An excellent data visualizer is open to new ideas, strives to learn, and unafraid to test these new ideas and methods. As data visualizers, we must believe that perfection can never be reached, but strive to come as close as possible with every visualization we create.

Bibliography

Alberts, A., & Cotgreave, A. (2019). Eye-tracking: what it teaches us about dashboard design. Retrieved from https://www.tableau.com/learn/webinars/eye-tracking-what-it-teaches-us-about-dashboard-design-1.

Blum, S. (2017, January 11). This Map Shows San Francisco Is Covered in Human Poop. Retrieved from https://www.thrillist.com/news/nation/human-wasteland-map-plots-all-of-san-franciscos-poop.

Brewer, C., Harrower, M., & The Pennsylvania State University. (2002). COLORBREWER 2.0. Retrieved from http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3.

Brewer, C. A. (2016). *Designing better maps: a guide for Gis users*. Redlands, CA: Esri Press.

Colblindor. (2006). Coblis - Color Blindness Simulator. Retrieved from https://www.color-blindness.com/coblis-color-blindness-simulator/.

Cotgreave, A., Shaffer, J., & Wexler, S. (2019). Dashboards for insight and impact. Retrieved from https://www.tableau.com/learn/webinars/dashboards-insight-and-impact-0.

Data Access and Dissemination Systems (DADS). (2010, October 5). American FactFinder. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t.

Dyslexie Font. (2017). Dyslexie font, the revolutionary font for people with dyslexia. Retrieved from https://www.dyslexiefont.com/en/home/.

Eskenazi, J. (2018, October 15). 'Poop map' creator has a message: Stop crapping on San Francisco. Retrieved from https://missionlocal.org/2018/10/poop-map-creator-has-a-message-stop-crapping-on-san-francisco/.

Field, K. (2018). *Cartography*. Redlands, CA: Esri Press.

Harvard Library. (2018). Harvard Geospatial Library. Retrieved from https://library.harvard.edu/services-tools/harvard-geospatial-library.

Houghton, J. (2019, April 12). On the Map: Game of Thrones filming locations. Retrieved from https://communityhub.esriuk.com/geoxchange/2019/4/12/game-of-thrones-filming-locations.

Illinois Department of Public Health. (2019). Illinois State cancer registry. Retrieved from http://www.idph.state.il.us/iscrstats/ZP/Show-ZP-Table.aspx.

Margo, C. E., & Harman, L. E. (2018). Helmholtz's critique of Goethe's Theory of Color: more than meets the eye. *Survey of Ophthalmology*, *64*(2), 241–247. doi: 10.1016/j.survophthal.2018.10.004

Nelson, J., & Esri Press. (2018). *Mapping with style* (Vol. 1). Retrieved from https://www.esri.com/content/dam/esrisites/en-us/media/pdf/Mapping With Style, Volume 1.pdf

Open the Books. (2019). 011-2019 San Francisco Human Waste Reportings . Retrieved from https://www.openthebooks.com/map/?Map=32504&MapType=Pin.

Paakki, M., Sandell, M., & Hopia, A. (2019). Visual attractiveness depends on colorfulness and color contrasts in mixed salads. *Food Quality and Preference*, *76*, 81–90. doi: 10.1016/j.foodqual.2019.04.004

Roshal, A. (2002). RARLAB. Retrieved from https://www.rarlab.com/download.htm.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877–8882. doi: 10.1073/pnas.0906172107

QR Code Generator. (2019). Create your QR Code for free. Retrieved from https://www.qr-code-generator.com/.

RealtyHop. (2018, September 29). Doo-Doo, the New Urban Crisis. Retrieved from https://www.realtyhop.com/blog/doo-doo-the-new-urban-crisis/#figure1.

Swan, R. (2014, March 19). Occupied: San Francisco: Understanding the City Through its Toilets. Yes, Really. Retrieved from https://archives.sfweekly.com/sanfrancisco/occupied-san-francisco-understanding-the-city-through-its-toilets-yes-really/Content?oid=2948873&showFullText=true.

Wong, J. (2018). (Human) Wasteland. Retrieved from http://mochimachine.org/wasteland/#.

Wu, F.-G., Tseng, C.-Y., & Cheng, C.-M. (2018). The composition of visual texture design on surface for color vision deficiency (CVD). *Computers in Human Behavior*, *91*, 84–96. doi: 10.1016/j.chb.2018.02.028

Appendices

Appendix 1

Creating the maps

1. I went to the Harvard Geospatial Library and selected "Launch HGL": https://library.harvard.edu/services-tools/harvard-geospatial-library

2. I then used the zoom tool and zoomed to Illinois to find the zip code tabulation areas from 2000 as shown in Figure 38.

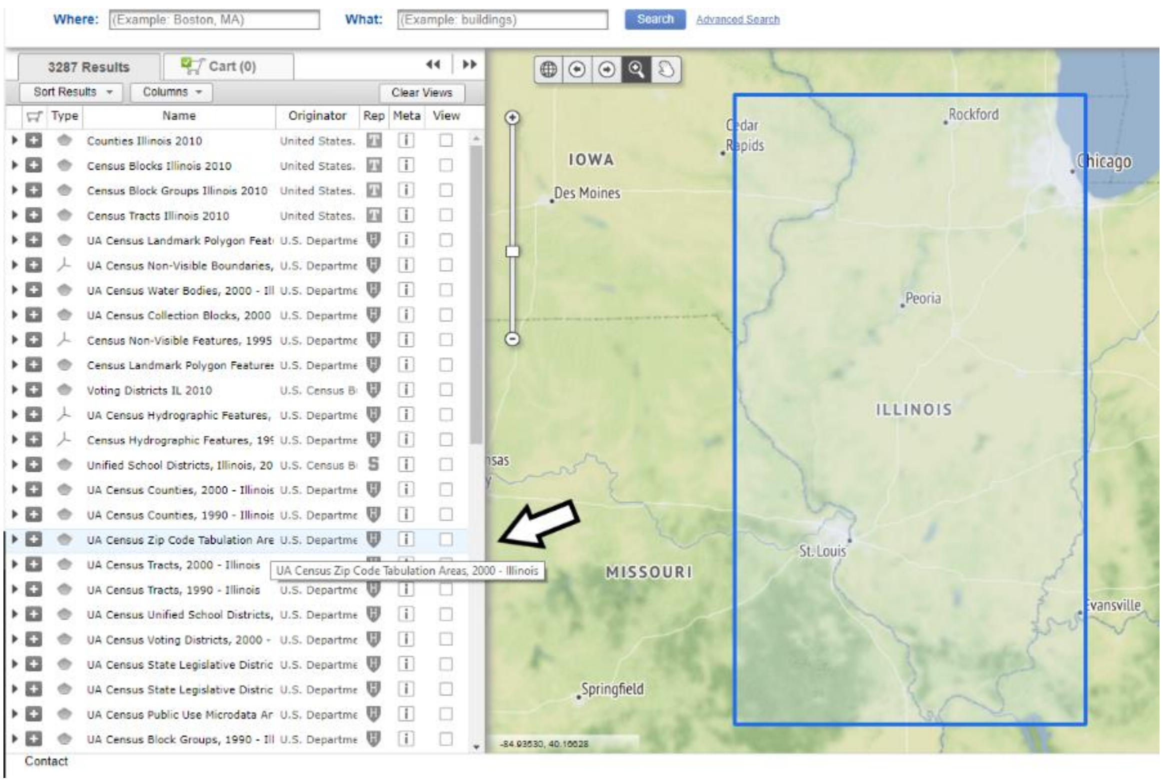


Figure 38. Zoom in to Illinois and scroll through the available data to find the data titled "UA Census Zip Code Tabulation Areas, 2000 – Illinois" from the U.S. Department of Commerce, Bureau of the Census, Geography Division.

3. I then added the UA Census Zip Code Tabulation Areas, 2000-Illinois from the Harvard Geospatial Library data-set to my cart to download.

4. I downloaded the data-set, unzipped the file folder, and added the shapefile to ArcMap 10.6.1.

5. I accessed the Illinois Department of Public Health website to download the zip code data for cancer incidence from 1996 to 2000: http://www.idph.state.il.us/about/epi/intropds.htm

   The cancer incidence data is coded as shown in Figure 39 below and the description of the data is as follows:

   "The files include incidence data for invasive cancers only with the exception of cancer of the bladder. Carcinoma in situ of the breast is provided in a separate category. Non-melanoma skin cancers, cases reported with unknown or "other" sex, and cases with an unknown age are omitted." (Illinois Department of Public Health, 2019).

FILE LAYOUT for ZIP code files

ZPCD9600.DAT

Record Format
(all fields are numeric)

Data Field	Positions	Length
ZIP code	1-5	5
cancer site groups	6-7	2
age group code	8-8	1
sex code	9-9	1
diagnosis year group	10-13	4

CODES FOR DATA FIELDS

ZIP code
Valid Illinois ZIP codes

cancer site group
1 oral cavity
2 colorectal
3 lung
4 breast - invasive (females)
5 cervix - invasive
6 prostate
7 urinary system
8 central nervous system
9 leukemias and lymphomas
10 all other cancers
11 breast in situ (females)

age group code
1 0-14
2 15-44
3 45-64
4 65 +

sex code
1 male
2 female

diagnosis year group

9600 diagnosed between 1996 and 2000

Figure 39. The file layout for Illinois cancer data by zip code tabulation areas.

6. To download the data, I right clicked on the zip code download link "ZPCD9600.EXE (871KB)", selected "save link as", and saved the .EXE file to my hard drive.

7. The file could not be unzipped and extracted as usual, so I downloaded WinRAR to extract the data. For my system, I downloaded "WinRAR x64 (64 bit) 5.80 beta 1": https://www.rarlab.com/download.htm

8. I then right clicked on the .EXE file and selected "Extract Here".

9. I then right clicked the ZPCD9600.DAT file that was extracted from the .EXE file and selected "Open with > Notepad".

10. I then right clicked in Notepad and selected "Select All" then copied and pasted the text into an Excel workbook.

11. Each data entry displayed as a string of numbers in one column with the ZIP code in positions 1-5, cancer site groups in positions 6-7, age group code in position 8, sex code in position 9, and diagnosis year group in positions 10-13.

12. I used the following formulas in Excel to produce the results also shown in Table 1 below:

    ZIPCode: =LEFT(A2, 5)
    CancerSiteCode: =MID(A2, 6, 2)
    AgeGroupCode: =MID(A2, 8, 1)
    SexGroupCode: =MID(A2, 9, 1)
    DiagnosisYearCode: =MID(A2, 10, 4)

    These formulas select certain characters in the string. The first formula for ZIPCode selects the first 5 characters in the string and the other formulas select the indicated number of characters in the string starting with the indicated position number. For example, the second formula for CancerSiteCode selects two characters starting at position 6 in the string.

Table 1. The first column labeled "Coded" shows an example of how the data displayed after being extracted from the Illinois Department of Public Health website into Excel. The remaining columns show the data after the Excel formulas were applied to each column using the string in the first column.

Coded	ZIPCode	CancerSiteGroupCode	AgeGroupCode	SexGroupCode	DiganosisYearCode
60084 2419600	60084	2	4	1	9600

13. I then removed the "Coded" column and added the table in ArcMap.

14. I opened the table in ArcMap and used the ZIPCode column to create a new table that shows the number of cancer incidences per ZIP code.

15. I joined the new table of cancer incidences per ZIP code to the Illinois ZIP code tabulation area shapefile.

16. Because the dataset shows the total number of cancer incidences per ZIP code, I downloaded Illinois 2000 population by ZIP code tabulation area from American Fact Finder by filtering for topic > year > 2000 and then filtering for geographies > 5-digit ZIP code tabulation area > Illinois > all 5-digit ZIP code tabulation areas fully within/partially within Illinois: https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

17. I used the "Profile of General Demographic Characteristics: 2000 Census 2000 Summary File 1 (SF 1) 100-Percent Data" table, the table ID is "DP-1".

18. I removed extra population data from the table, removed characters incompatible with ArcMap, and added the edited table in ArcMap.

19. I then combined the uploaded population table with the previously joined ZIP code cancer incidence table in ArcMap.

20. I saved my shapefile with cancer incidences and population to my file geodatabase, uploaded the new shapefile into my map, and removed the previous shapefile that did not have this new information attached to it.

21. To account for cancer incidences per population in the ZIP code areas, I created a new field in the attribute table of the shapefile and opened the editor in ArcMap.

22. In the new field, I calculated cancer incidence divided by population and multiplied by 100 to show the percentage of cancer per zip code using the following formula:

([Cnt_ZIPCode] / [HC01VC01]) *100

23. I created four maps classified four different ways. The first map was classified by natural breaks (jenks) with 15 classes. The second map was classified by quantile with eight classes.

24. The third map was classified by standard deviation with five classes and the fourth map was also classified by standard deviation, but with two classes. The two-class map was divided into cancer rates greater than 0.5 standard deviation and those below 0.5 standard deviation.

25. To create a color palette that would be visible to those with color vision deficiency, I opened www.colorbrewer2.org and selected sequential data, colorblind safe, changed the number of data classes to the maximum of 9, and then selected the pink/purple multi-hue. I chose the pink/purple multi-hue color palette because it was colorblind safe, was a more visually interesting sequential color palette, and was a color palette that is similar to cancer cell samples that have been stained to study under the microscope as shown in Figure 40 below. By choosing this color palette, I was able to create and follow a theme throughout my thesis.

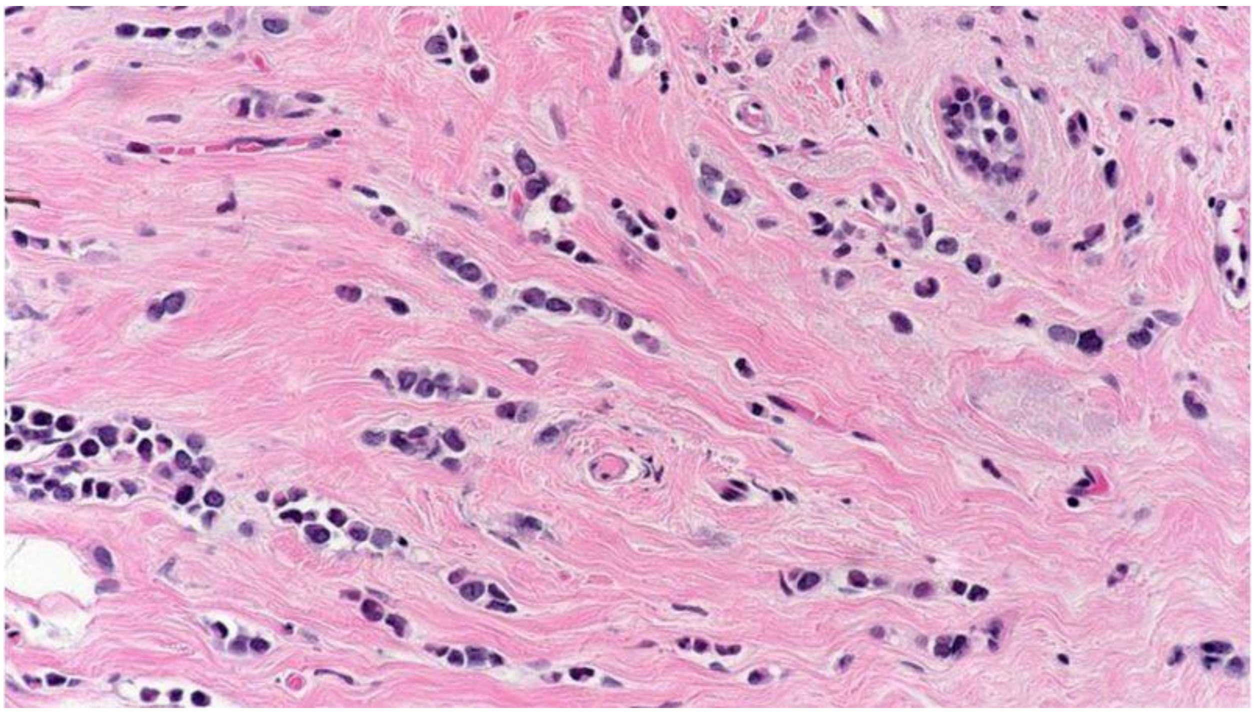


Figure 40. Histologic grade 1 invasive lobular carcinoma from John Hopkins Pathology at https://pathology.jhu.edu/breast/types-of-breast-cancer/. This type of breast cancer is the second most common type of breast cancer according to Johns Hopkins Medicine.

26. Since my first map had 15 classes, I needed to add additional colors to the color palette generated by Colorbrewer2, which only had 9 classes. Colorbrewer2 defaults to display the hex code number so I selected RGB to display the RGB values instead as shown below in Figure 41. To create the additional colors, I selected new colors evenly spread between the existing colors in the color palette provided by Colorbrewer2 and selected RGB values between the existing colors as shown below in Table 2. For example, I created the color for values between 1.0% and 1.5%. The R value for that color, 254, is between the R value of the previous R value, 255, and the next R value, 253.

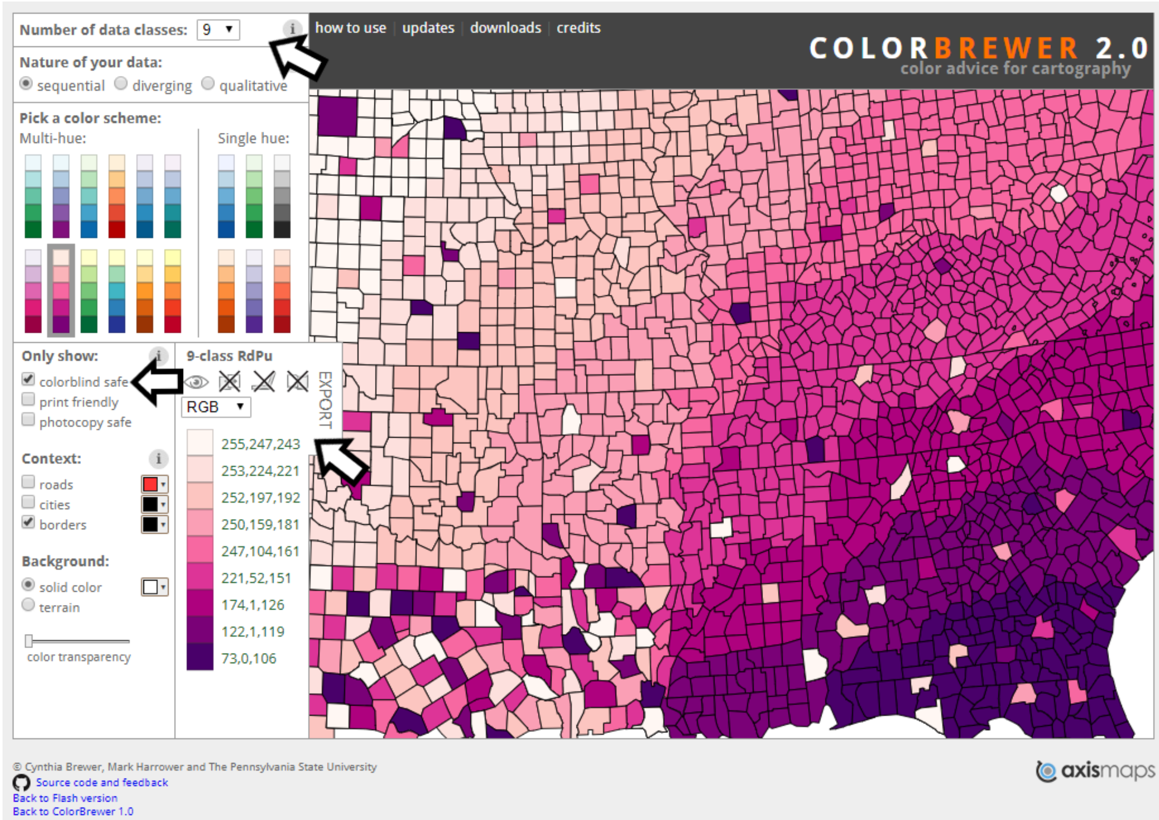


Figure 41. Colorbrewer2 with arrows pointing to changing the number of data classes, type of data, colorblind safe selection, and the RGB values. The color palette displayed is the color palette I selected to create my 15-class natural break and 8-class quantile maps.

Table 2. This table shows the RGB values used to visualize the 15-class natural break (jenks) map with the resulting color shown below each value. The values highlighted in pink were colors I created to add between the colors generated by Colorbrewer2.

15-Class Natural Break Map

	<1.0%	**1.0%-1.5%**	**1.5%-1.8%**	**1.8%-2.1%**	**2.1%-2.3%**
R	255	254	253	252	252
G	247	235	224	215	197
B	243	233	221	213	192

	2.3%-2.6%	**2.6%-2.9%**	**2.9%-3.2%**	**3.2%-3.6%**	**3.6%-4.4%**
R	251	250	247	221	200
G	175	159	104	52	30
B	187	181	161	151	140

	4.4%-5.9%	**5.9%-9.1%**	**9.1%-16.2%**	**16.2%-22.4%**	**22.4%-37.5%**
R	174	150	122	100	73
G	1	1	1	1	0
B	126	120	119	115	106

27. To create a visually interesting 8-class quantile map, I used the 9-class sequential colors generated by Colorbrewer2 after removing the middle color. The RGB values used in this map are shown below in Table 3.

Table 3. This table shows the RGB values used to visualize the 8-class quantile map with the resulting color shown below each value. This palette was created using the 9-class sequential values generated by Colorbrewer2 with the middle color removed from the palette so that the resulting map would be more visually interesting than the 8-class sequential palette generated by Colorbrewer2.

8-Class Quantile Map

	<1.8%	**1.8%-2.1%**	**2.1%-2.4%**	**2.4%-2.7%**	**2.7%-2.9%**	**2.9%-3.2%**	**3.2%-3.6%**	**3.6%-37.5%**
	255	253	252	250	221	174	122	73
G	247	224	197	159	52	1	1	0
B	243	221	192	181	151	126	119	106

28. I used a different color palette generated from Colorbrewer2 for the 5-class standard deviation map. To generate this color palette, I selected diverging data, colorblind safe, five classes, and selected the pink/green diverging palette as shown in Figure 42 below.

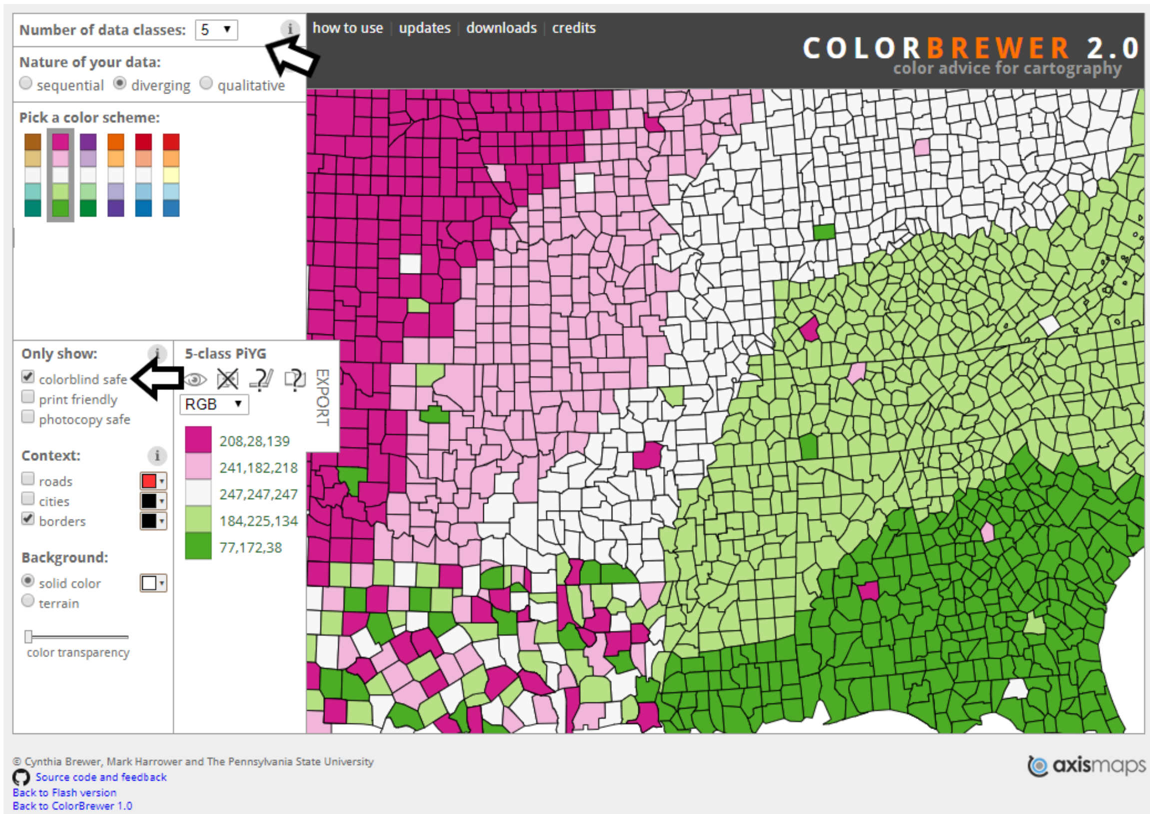


Figure 42. Colorbrewer2 with the pink/green diverging color palette. Arrows point to changing the number of data classes, type of data, and the colorblind safe selection. The color palette displayed is the color palette I selected to create my 5-class standard deviation map.

29. For the 5-class standard deviation map, I used the same color palette generated by Colorbrewer2 without any changes as shown in Figure 42 above and Table 4 below.

Table 4. This table shows the RGB values used to visualize the 5-class standard deviation map with the resulting color shown below each value. This palette was generated by Colorbrewer2.

5-Class Standard Deviation Map

	< -1.5 Std. Dev.	**-1.5 – -0.50 Std. Dev.**	**-0.50 – 0.50 Std. Dev.**	**0.50 – 1.5 Std. Dev.**	**> 1.5 Std. Dev.**
R	208	241	247	184	77
G	28	182	247	225	172
B	139	218	247	134	38

30. To create the color palette for the 2-class standard deviation map I used the darkest purple from the 15-class natural break (jenks) map and created a purple-toned off-white color as shown in Table 5 below.

Table 5. This table shows the RGB values used to visualize the 2-class standard deviation map with the resulting color shown below each value. I used the darkest purple generated by Colorbrewer2 from the 15-class natural break (jenks) map for values above the standard deviation and create a purple-toned off-white color for values below the standard deviation.

2-Class Standard Deviation Map

	< 0.50 Std. Dev.	**>0.50 Std. Dev.**
R	250	73
G	247	0
B	243	106

31. I also created two additional maps with a standard red/yellow/green palette. I used the same 15-class natural break (jenks) map and 8-class quantile maps as before but recolored them using the red/yellow/green palette generated by ArcMap 10.6.1 as shown in Figure 43 below.

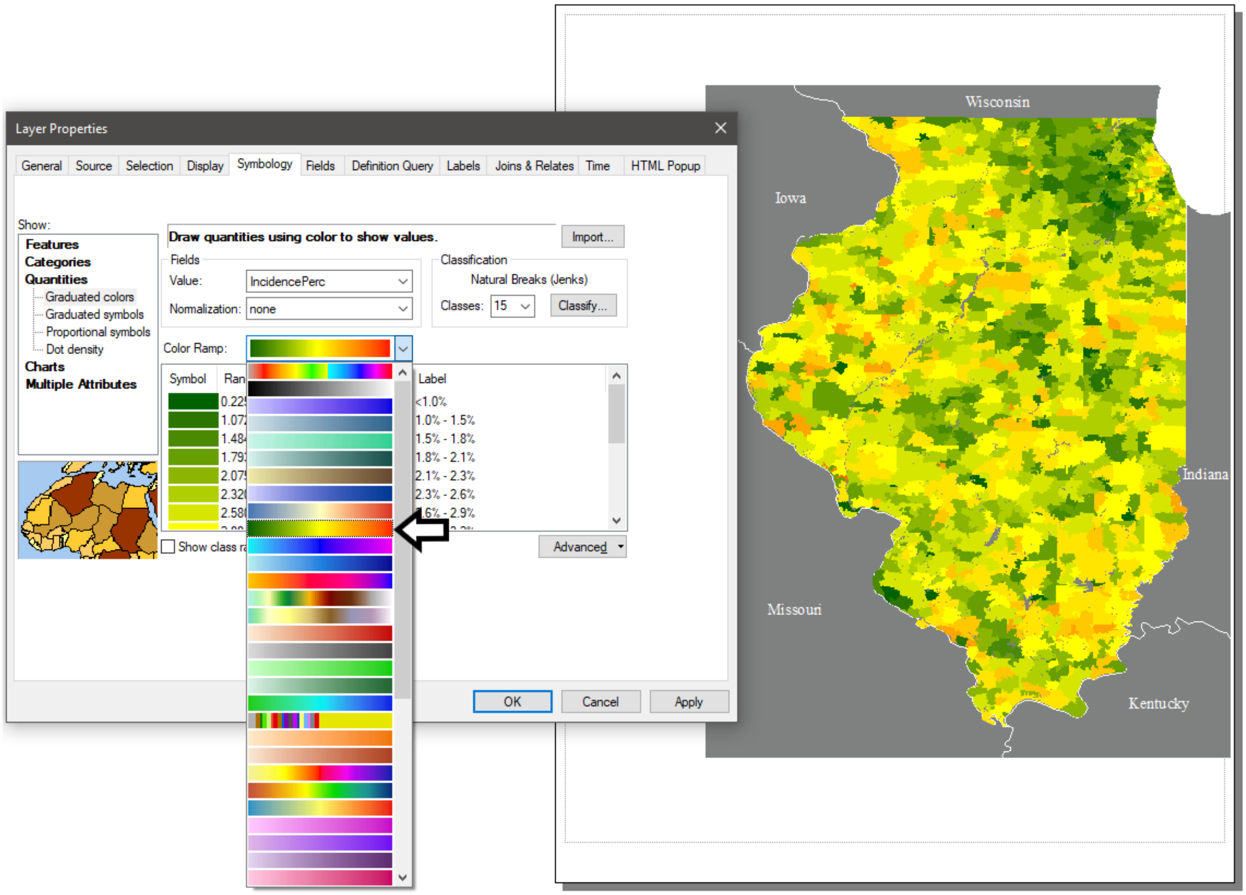


Figure 43. The 15-class natural break (jenks) map recolored with a standard red/yellow/green palette generated by ArcMap 10.6.1 with an arrow pointing to the selected color palette.

32. To make the maps look clean and polished, I added a USA states layer from Esri using the “Add Data” > “Add Data From ArcGIS Online…” tool and using the search term “States” as shown in Figure 44 below.

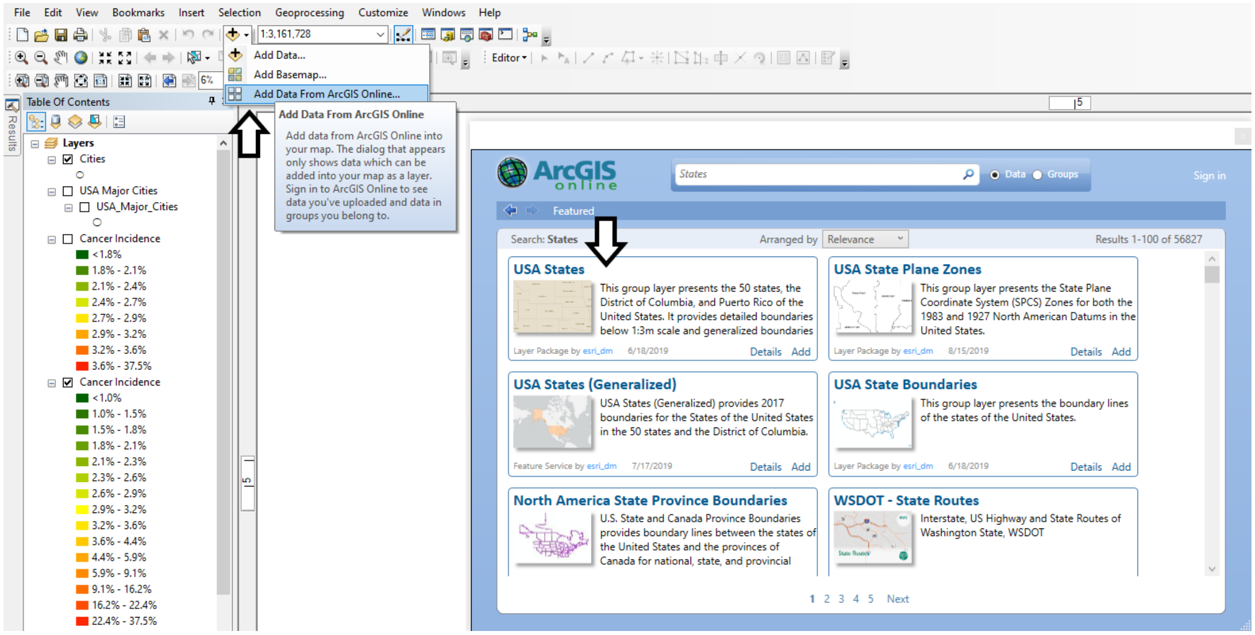


Figure 44. ArcMap 10.6.1 with arrows pointing to the "Add Data" > "Add Data From ArcGIS Online" tool and the selected USA States layer package managed by Esri.

33. I symbolized the USA States > USA States (over 1:3m) using black (RGB: 0, 0, 0) – which displayed as a dark gray on the map – with a white outline (RGB: 255, 255, 255).

34. I then added state name labels symbolized with white text (RGB: 255, 255, 255).

35. Next, I added a USA major cities layer from Esri using the "Add Data" > "Add Data From ArcGIS Online…" tool and using the search term "USA Major Cities" as shown in Figure 45 below.

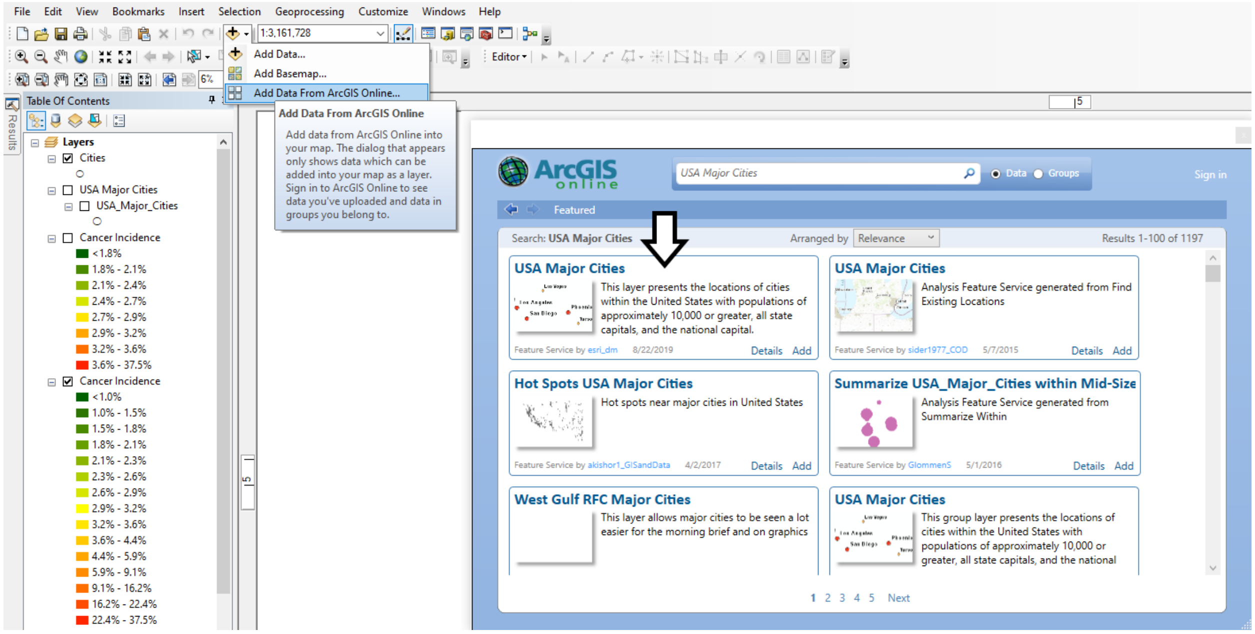


Figure 45. ArcMap 10.6.1 with an arrow pointing to the selected USA Major Cities layer package managed by Esri.

36. I then changed the symbology from the default population symbology to a single symbol of a white circle size 8.00 with a black outline.

37. I then used the query builder under the definition query tab in layer properties to only symbolize larger cities in and near Illinois so that map viewers could have a reference to cities throughout the state. I symbolized Moline, Chicago, Galesburg, Peoria, Normal, Champaign, Quincy, Springfield, and Carbondale, Illinois; and St. Louis, Missouri, by using the following definition query which is also shown in Figure 46 below:

    "NAME" IN ('Moline' , 'Chicago' , 'Galesburg' , 'Peoria' , 'Normal' , 'Champaign' , 'Quincy' , 'Springfield' , 'St. Louis' , 'Carbondale').

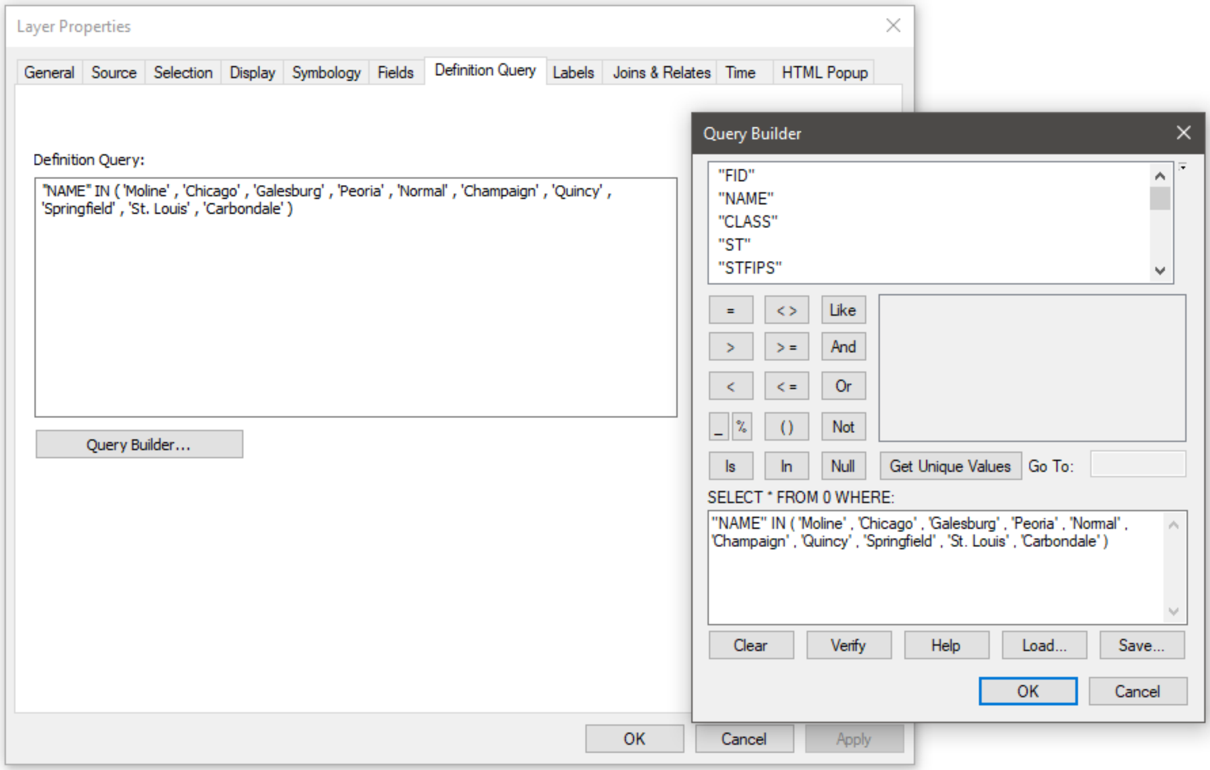


Figure 46. The definition query used in the query builder to select for cities in and near Illinois.

38. To add the city names to the map, I went to the labels tab in layer properties by checking the “Label features in this layer” box, selected “NAME” for the label field, and symbolized with white test (Times New Roman size 12) with a black mask sized 1.50 shown in Figure 47 below.

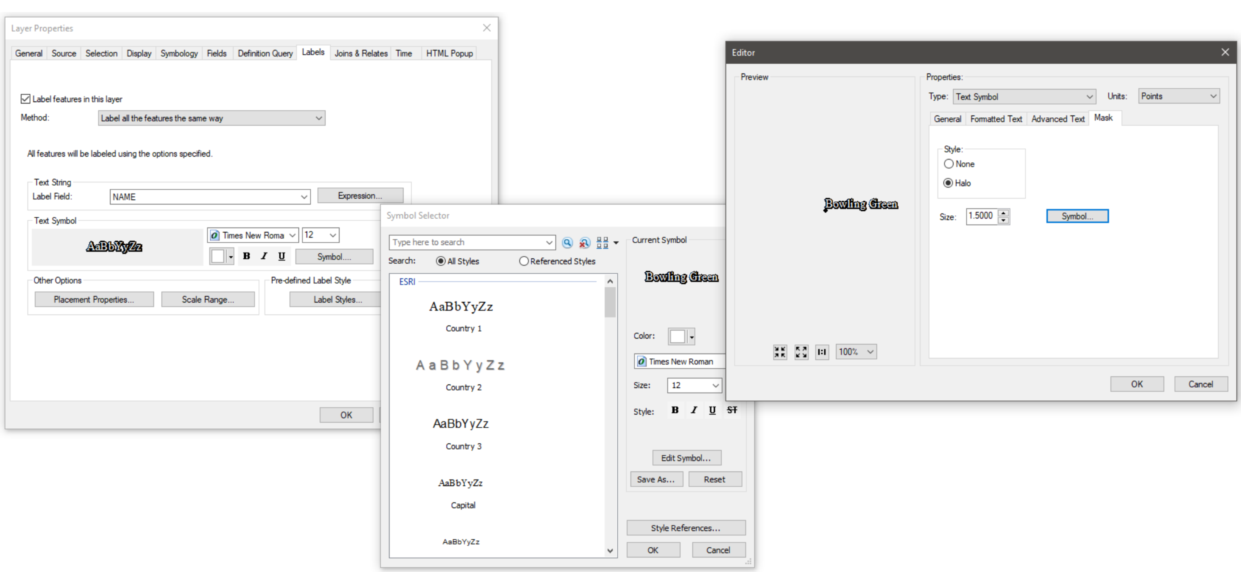


Figure 47. The symbology selected in ArcMap 10.6.1 to label cities.

39. The final maps that I created are shown below in Figures 48-53.

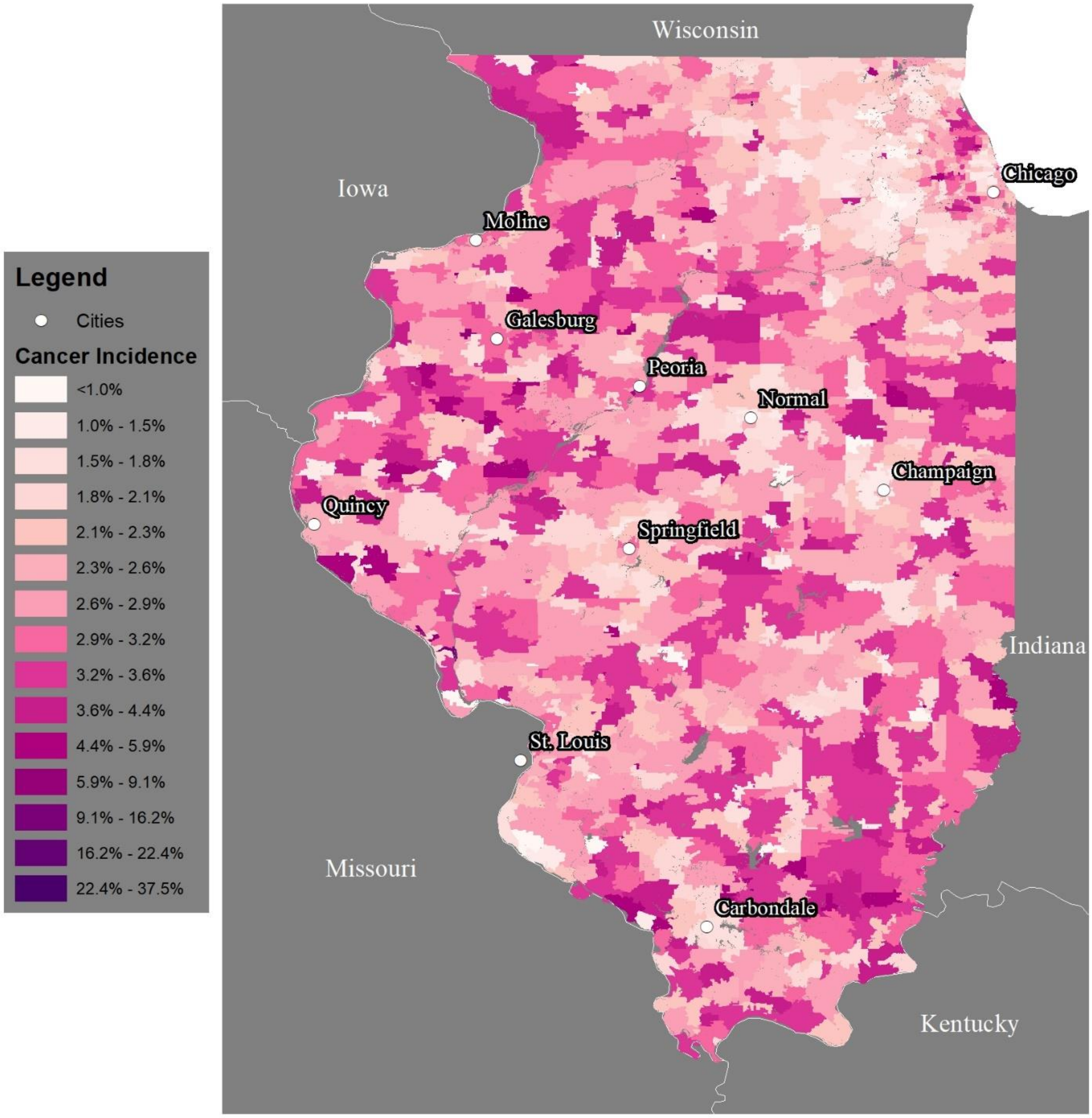


Figure 48. The 15-class natural break (jenks) pink/purple map I created using ArcMap 10.6.1.

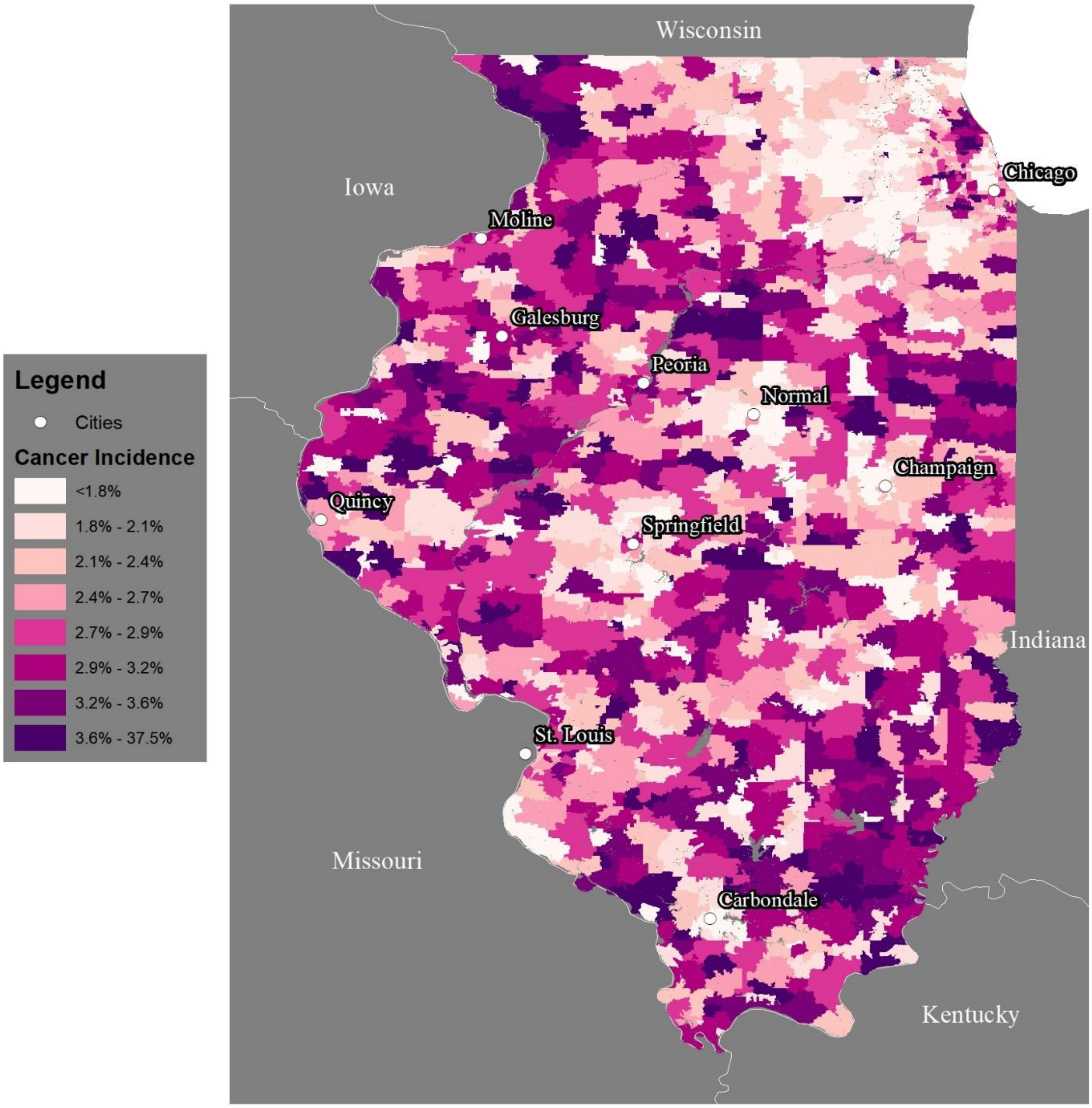


Figure 49. The 8-class quantile pink/purple map I created using ArcMap 10.6.1.

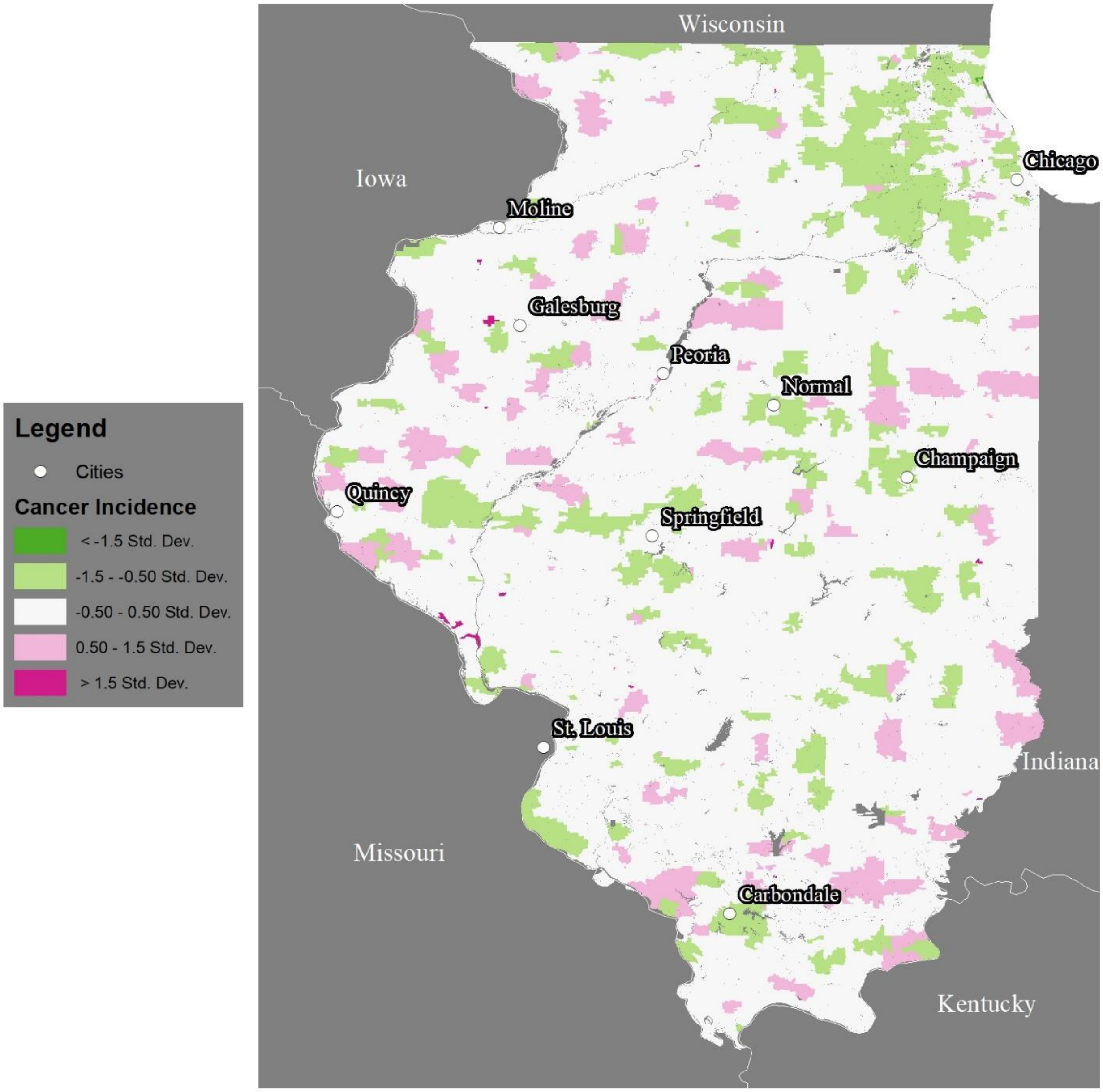


Figure 50. The 5-class standard deviation map I created using ArcMap 10.6.1.

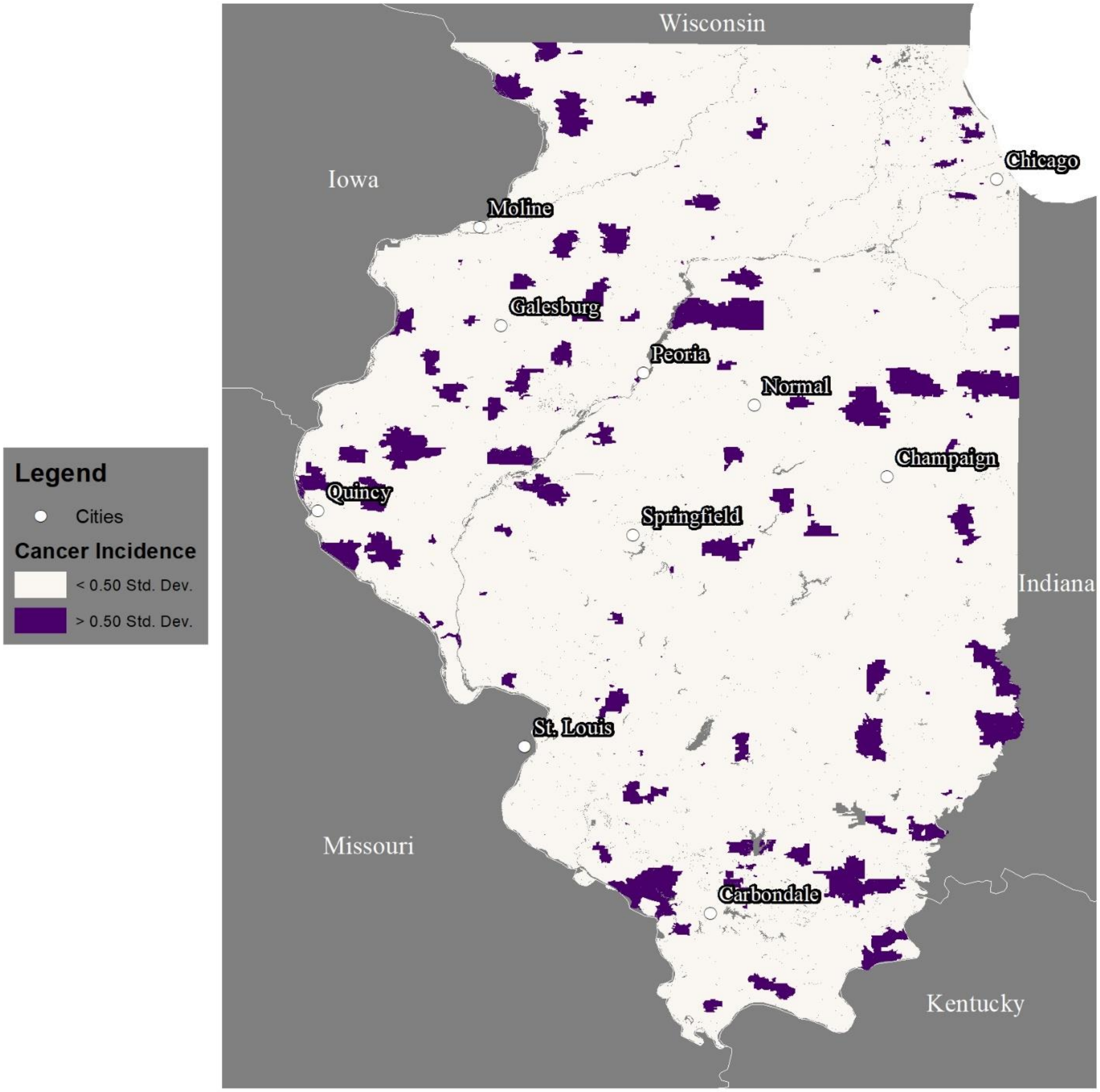


Figure 51. The 2-class standard deviation map I created using ArcMap 10.6.1.

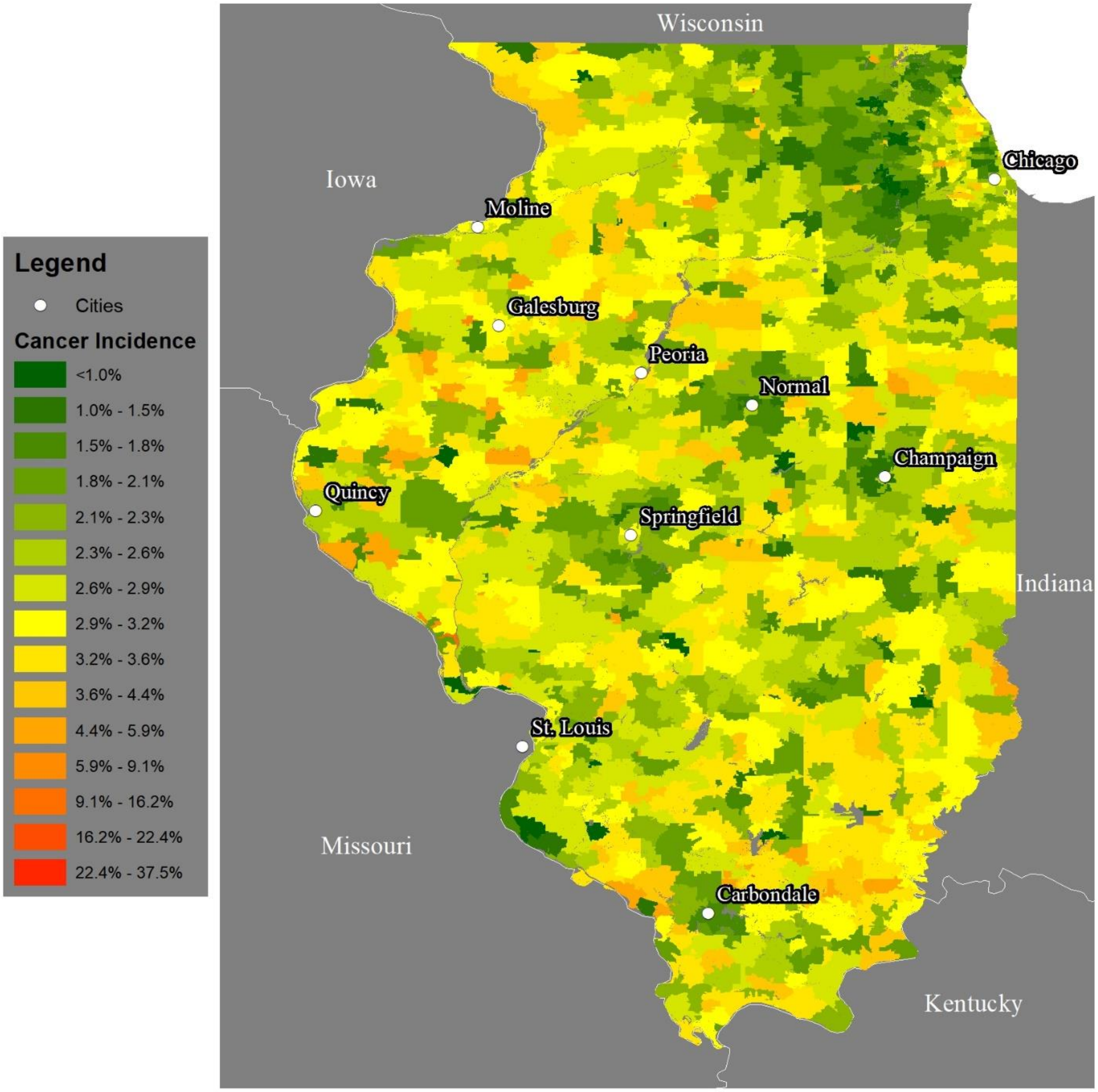


Figure 52. The 15-class natural break (jenks) red/yellow/green map I created using ArcMap 10.6.1.

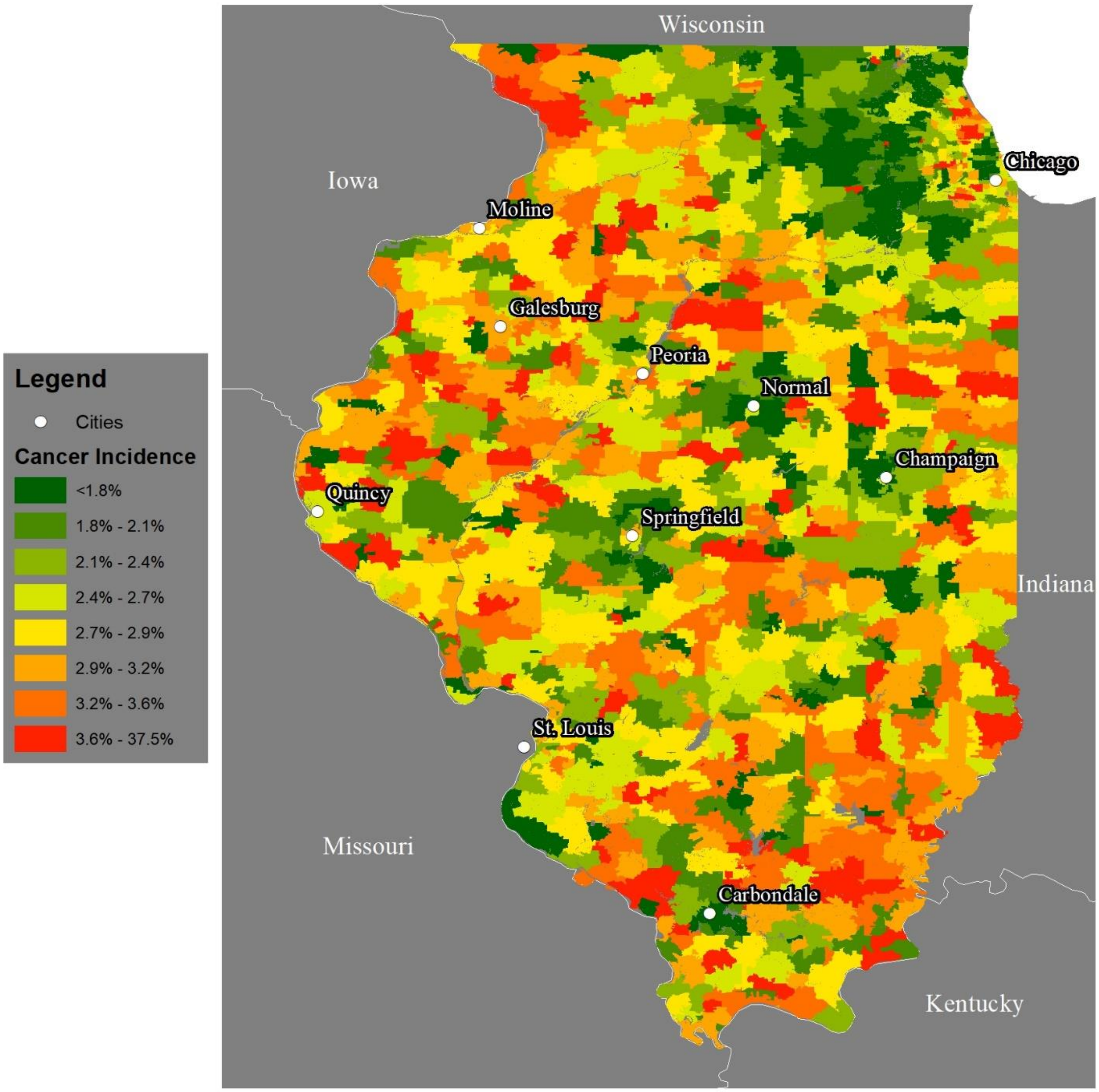


Figure 53. The 8-class quantile red/yellow/green map I created using ArcMap 10.6.1.

Appendix 2

Qualtrics Survey

1. I exported my data from Qualtrics as a .CSV file and opened the file in Excel.

2. To make the Qualtrics download useable in Tableau, I edited the Excel sheet in a separate tab and saved the file as an .XLSX file.

3. I deleted the Import ID row generated by Qualtrics (row 3) the first 14 Qualtrics generated columns (A-Q). These columns contain information such as distribution channel and time of survey completion.

4. I then changed the column headers with the question number and an abbreviated version of the question. For example, question 1 "How many years of experience do you have with mapping using GIS or online tools?" has the header "1 Years of experience". Some questions generate two columns, for example question 4 "Do you have any type of vision deficiency? If yes, can you please describe?" allows the survey respondent to select yes or no and provides a text area to the "yes" selection. For these types of questions, both columns will have a header that starts with the question number. For question 4, the columns are titled "4 Vision deficiency" and "4 Text".

5. I then removed the last column header row generated by Qualtrics (row 2) and removed 4 survey responses that had not been completed.

6. Since question 4 was a text box question, I created a new column next to question 4 and entered the answers in the same format so that answers such as "f" and "female" were all reported as "female".

7. I also created a new column next to question 7 and entered the information as a whole number since some answers were reported as text answers and reported non-residence status as "0".

8. I added 24 columns next to question 8 with the headers "8 City 1", "8 County 1", "8 State 1", etc. to enter the location by city, county, and state for each response so that they could be mapped in Tableau. Answers that were vague, such as "West central Illinois", did not receive a city or county location in the newly added columns.

9. I added two additional columns after questions 21 and 22 and entered "A", "B", "No preference", or "Null" so that the answers were all in the same format, as most responses contained descriptive text.

Appendix 3

Color Vision Deficiency Test

To test the readability of the poster to different types of color vision deficiency, I uploaded an image of the poster to the Colblindor Color Blindness Simulator and selected for each type of color vision deficiency as shown in Figures 54-60 below: https://www.color-blindness.com/coblis-color-blindness-simulator/

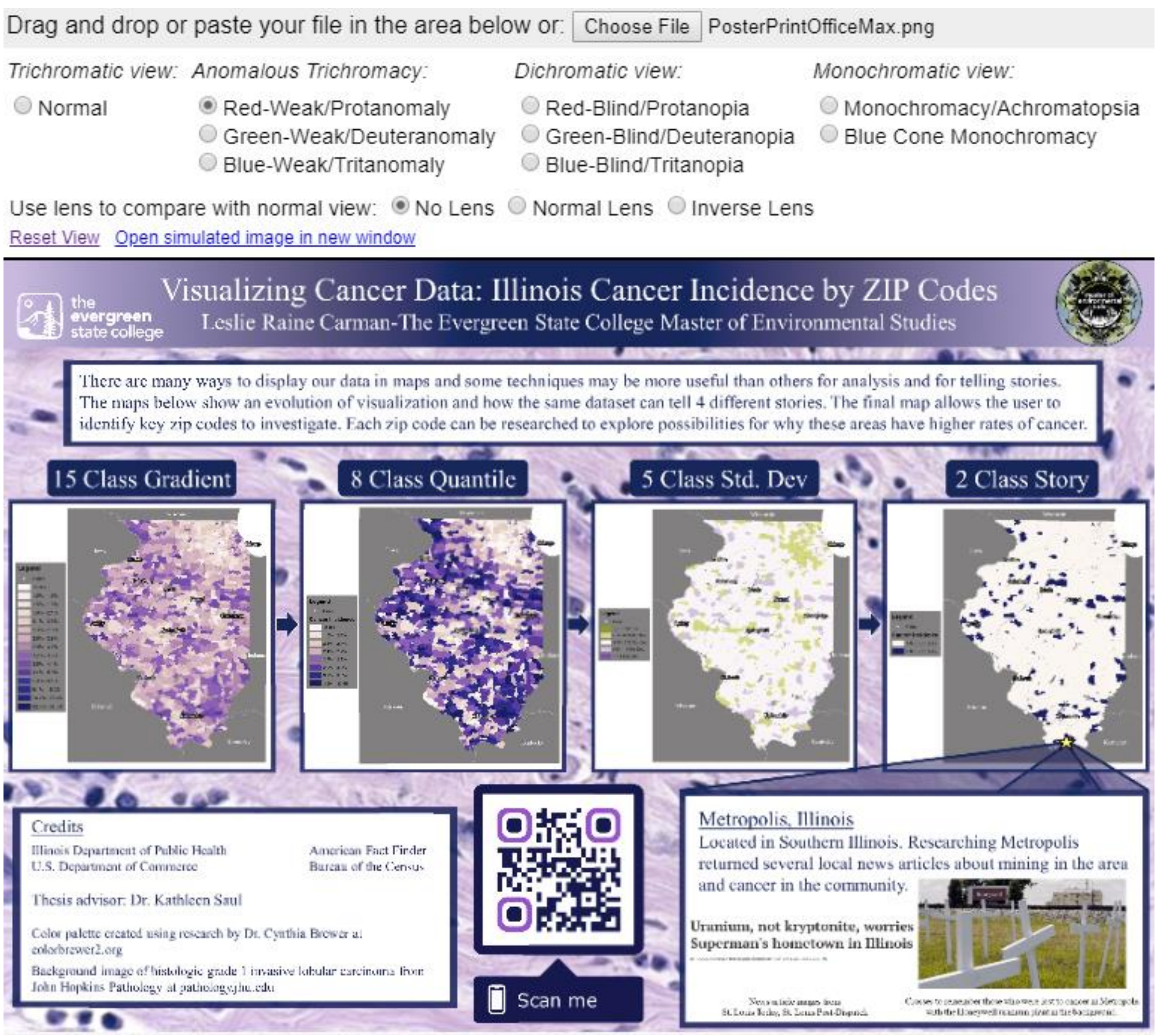


Figure 54. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with red-weak color blindness (protanomaly).

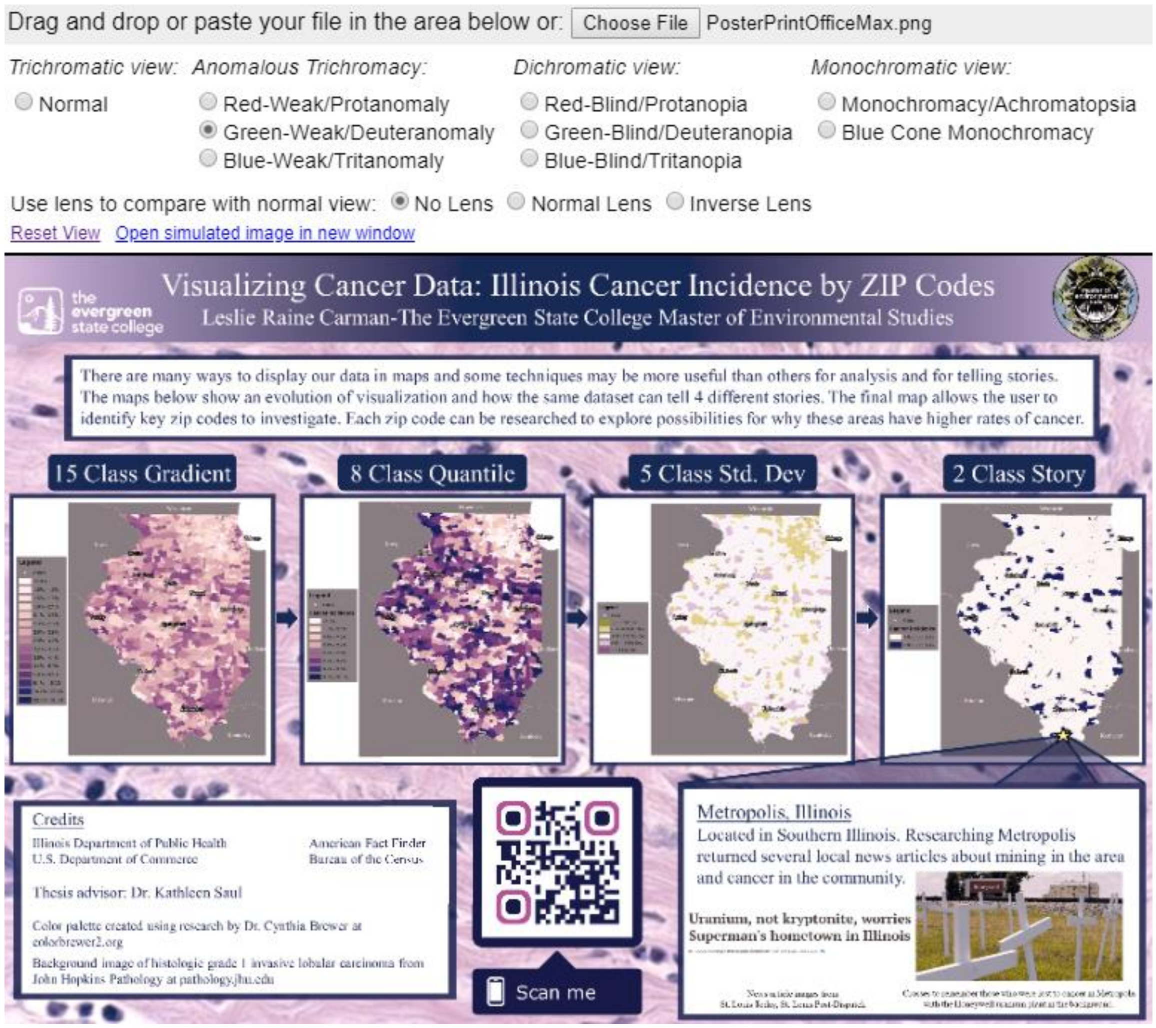


Figure 55. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with green-weak color blindness (deuteranomaly).

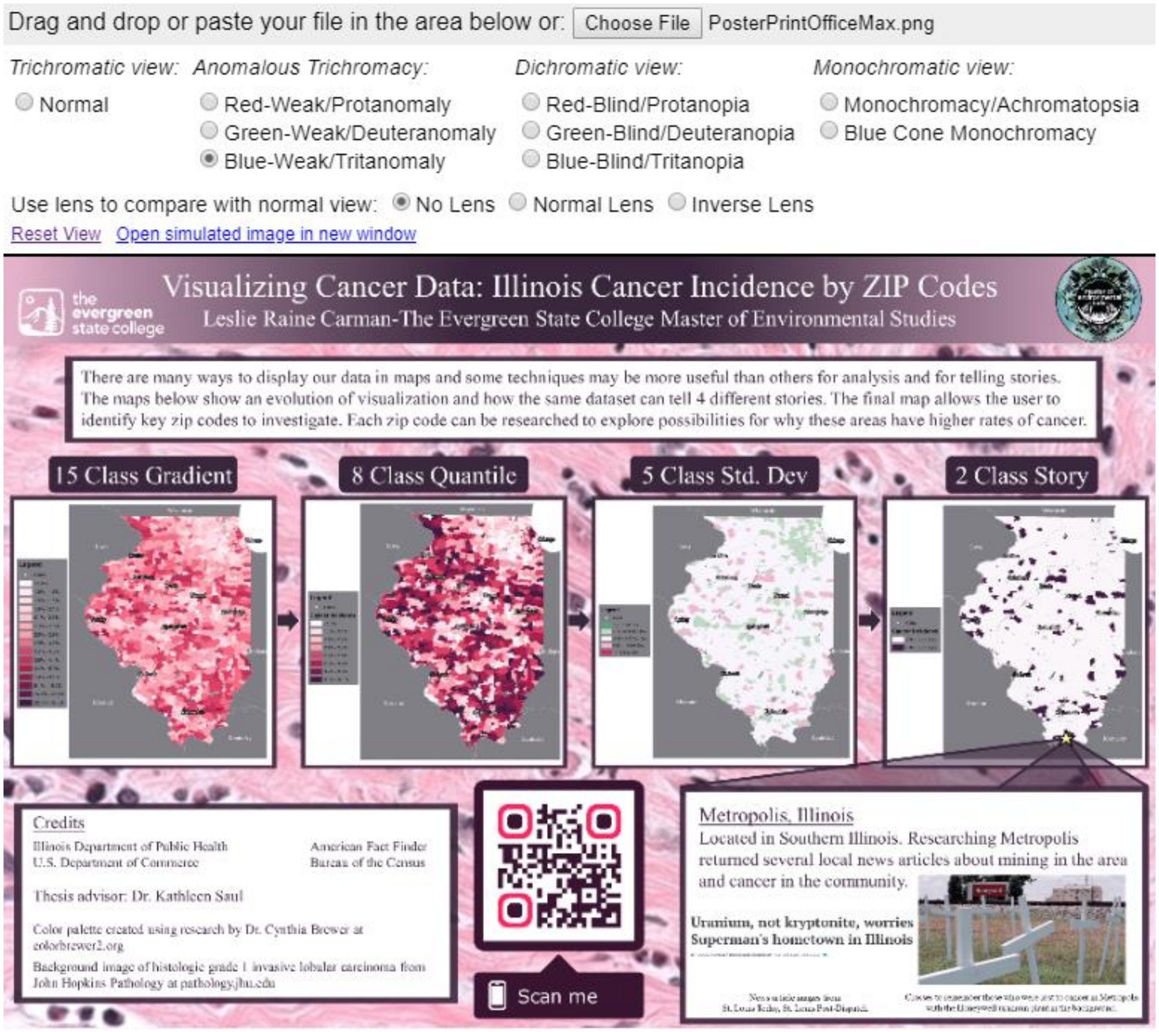


Figure 56. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with blue-weak color blindness (tritanomaly).

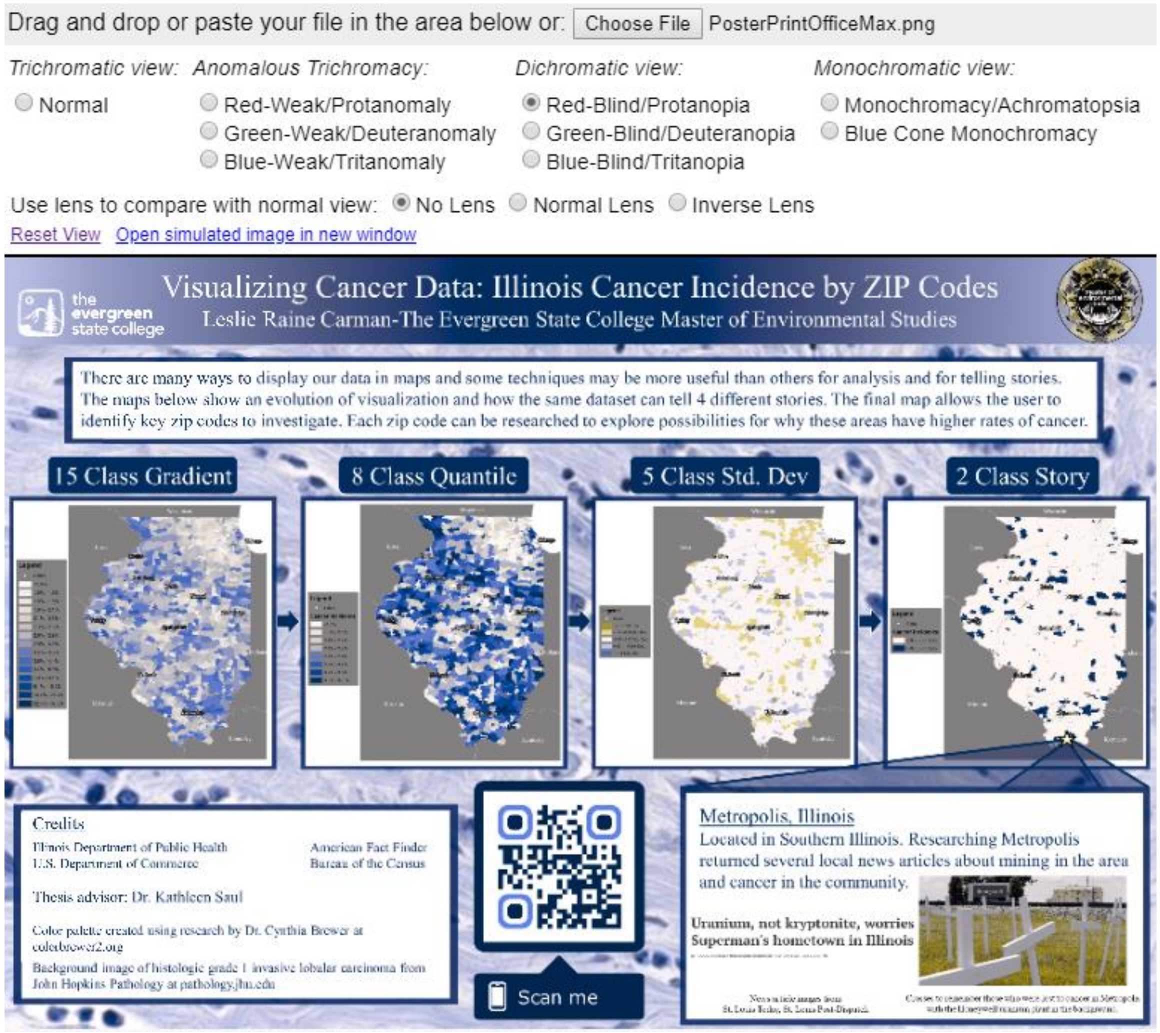


Figure 57. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with red-blind color blindness (protanopia).

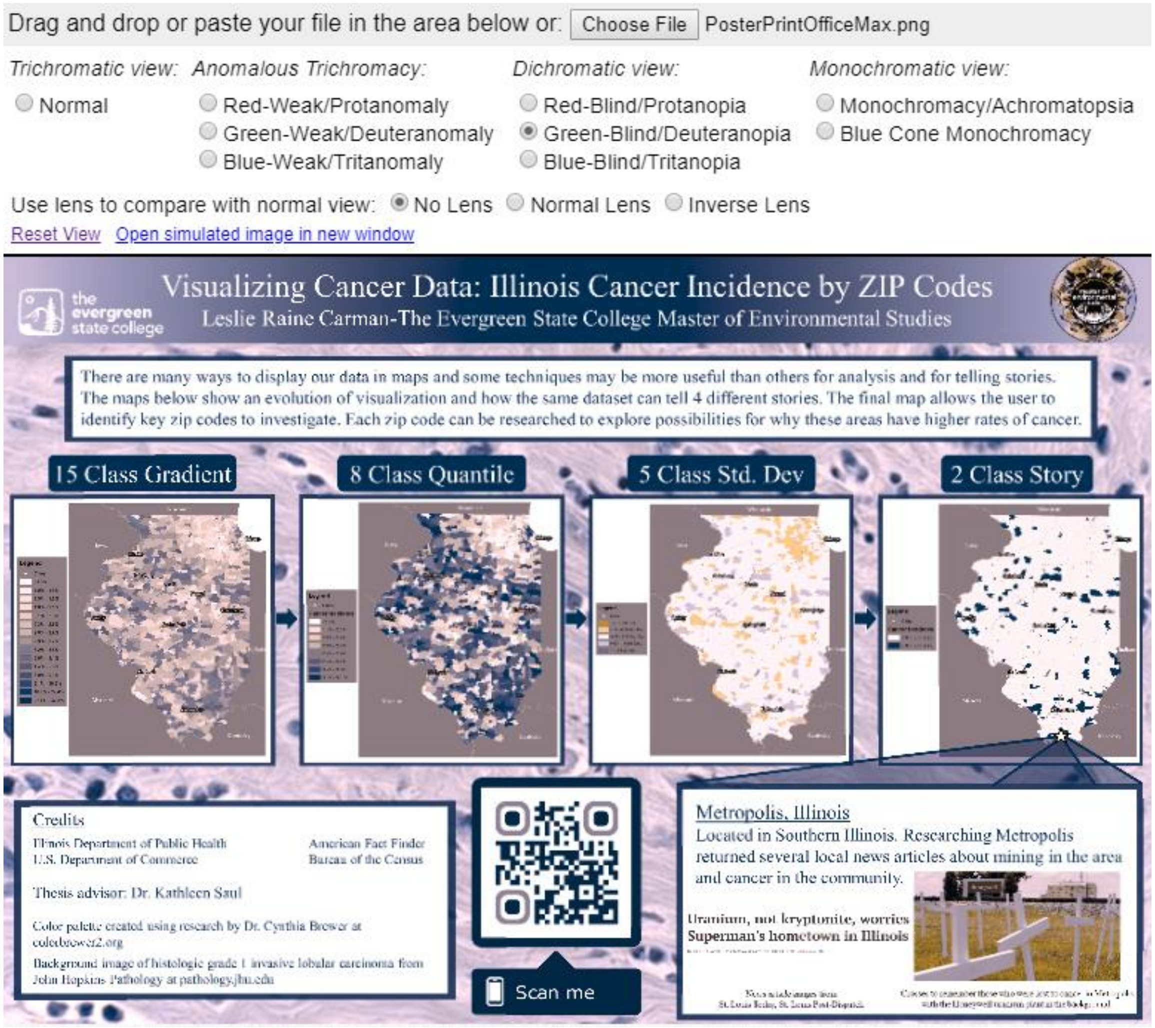


Figure 58. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with green-blind color blindness (deuteranopia).

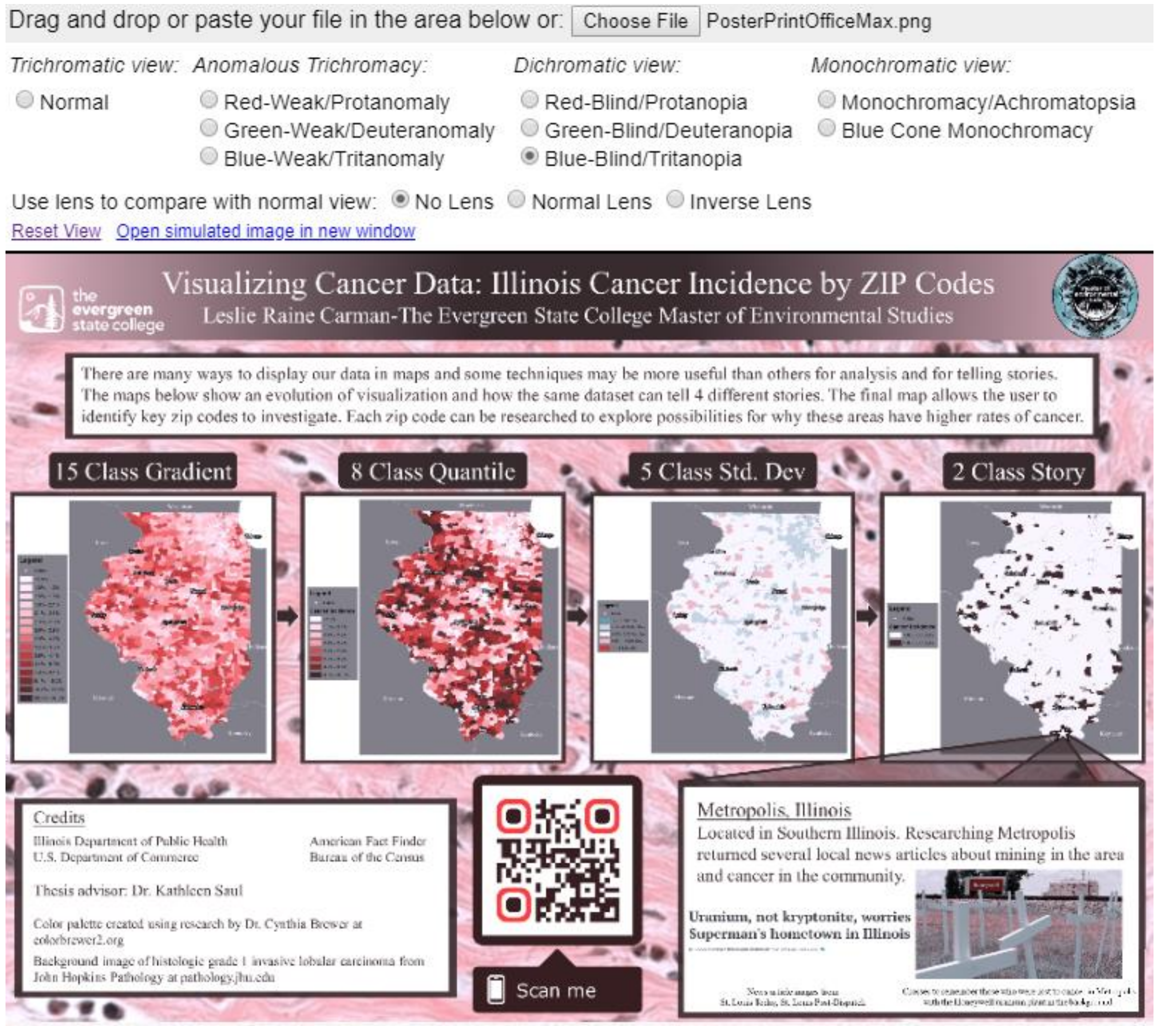


Figure 59. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with blue-blind color blindness (tritanopia).

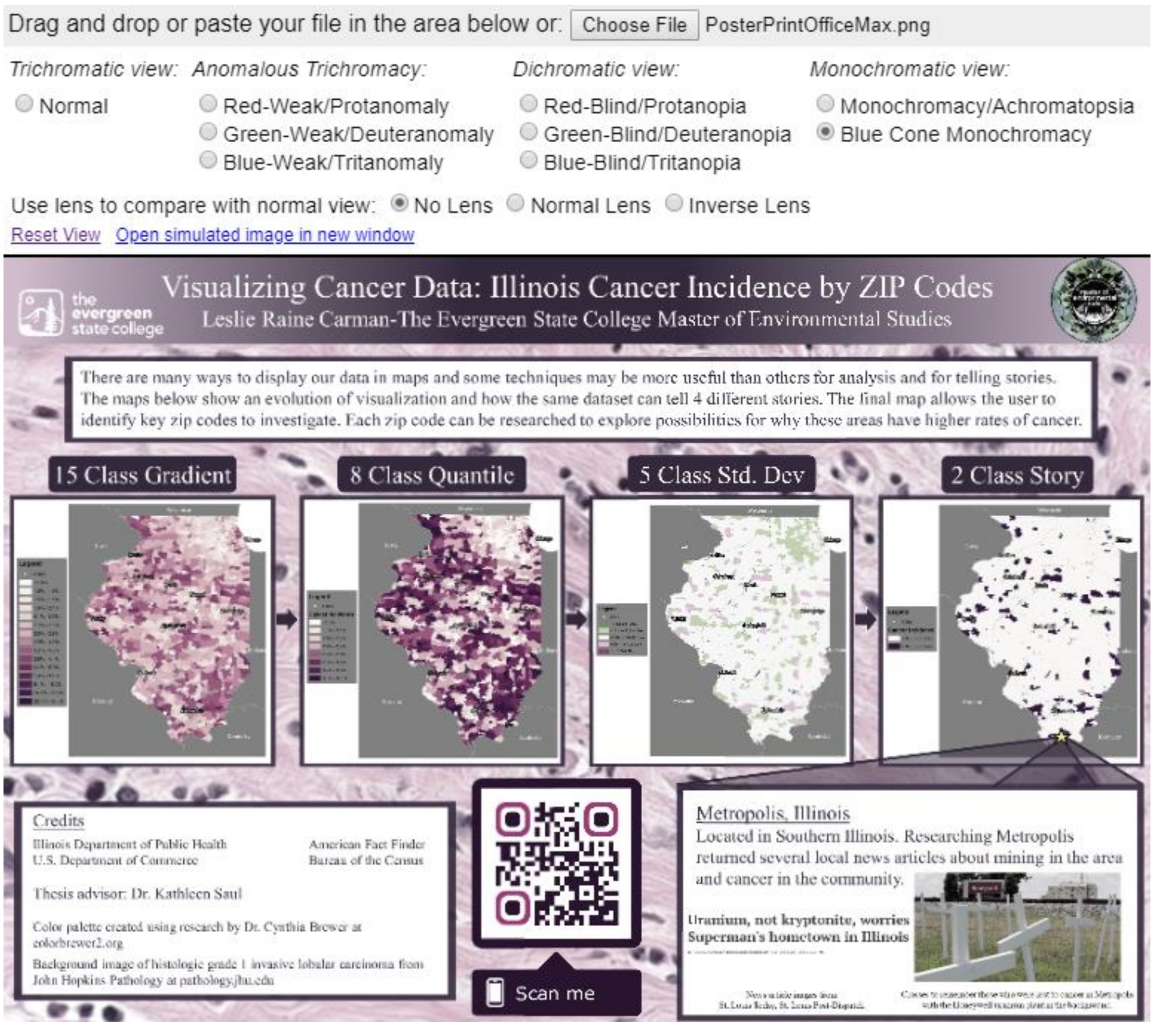


Figure 60. My poster viewed in the Colblindor Color Blindness Simulator as it would be seen by someone with blue cone monochromacy.