Network of Childhood Resources in the

St. Louis Area

By Mandy Klein

Program Name: Masters of GIS, Department of Geography, The Pennsylvania State University

Date: Dec 2020

Advisor: Stephen A. Matthews

**Abstract:**

Tableau Public was used to create two dashboards visualizing data about several childhood indicators and location/attribution information for facilities parents may need in the St Louis, MO, area. The purpose of this project was to create a dynamic visualization new parents could use in finding resources to care for their young child and also to aid non-profits with directing parents to resources they may need. This paper walks through each of the data sources and provides information about the datasets, any known limitations, how each dataset was used, and, where applicable, lists the variables used from that source. The process of preparation, formatting, and manipulating the data required significant work to make the data useable for visualization. Once the data were prepared, the visualizations were created in Tableau public by first creating the individual visualizations, each of the four dashboards were created (two data driven dashboards and two project information dashboards), then the story for the project was completed before finally publishing the story. The future of this project, potential data sources, dashboard improvements, promotion of use, etc. are discussed as well as the potential rewards and challenges each improvement could bring to the project.

**Keywords:**

Tableau public, non-profit, childhood indicators, public data, data preparation, data visualization, Notepad ++

Contents

[1. Introduction 4](#_Toc58261199)

[2. Background 5](#_Toc58261200)

[3. Data Harvesting & Methodology 7](#_Toc58261201)

[a. Data Sources 7](#_Toc58261202)

[b. Processing 9](#_Toc58261203)

[4. Tableau Dashboard/Story Creation 23](#_Toc58261204)

[a. Background Data Dashboard 24](#_Toc58261205)

[1. Load Background Data 24](#_Toc58261206)

[2. Create Visualizations & Filters for Background Data 25](#_Toc58261207)

[3. Create Background Data Dashboard 30](#_Toc58261208)

[b. Facilities Dashboard 31](#_Toc58261209)

[1. Load Facilities Data 31](#_Toc58261210)

[2. Create Visualizations & Filters for Facilities Data 32](#_Toc58261211)

[3. Create Facilities Dashboard 34](#_Toc58261212)

[c. Create Introduction & User Guide Dashboards 35](#_Toc58261213)

[d. Create Story 37](#_Toc58261214)

[e. Publish Story 38](#_Toc58261215)

[5. Future Improvements 38](#_Toc58261216)

[a. Data Availability, Coverage, and Linkage 39](#_Toc58261217)

[b. Search Tools/Filters 45](#_Toc58261218)

[c. Data Visualization 46](#_Toc58261219)

[d. Promotion of use 48](#_Toc58261220)

[6. Conclusion 49](#_Toc58261221)

[7. Bibliography 52](#_Toc58261222)

[Appendix A - Tableau User Guide 54](#_Toc58261223)

[e. Background Data Dashboard 55](#_Toc58261224)

[f. Facilities Dashboard 57](#_Toc58261225)

[1. Sliders: 59](#_Toc58261226)

[2. Tools and Functionality: 60](#_Toc58261227)

# Introduction

In the St Louis, Missouri metropolitan area there is a registered non-profit called Nurses for Newborns (*Nurses for Newborns*, 2020) (Figure 1). I learned of this organization through an article on a local news website late one night when my own son was an infant (Figure 2) and was struggling to sleep. Having grown up in poverty and understanding the daily struggles for basic necessities as well as having my own newborn to care for without a local support network, I was inspired to focus my capstone project on building a resource that could help other parents find diverse information they need.

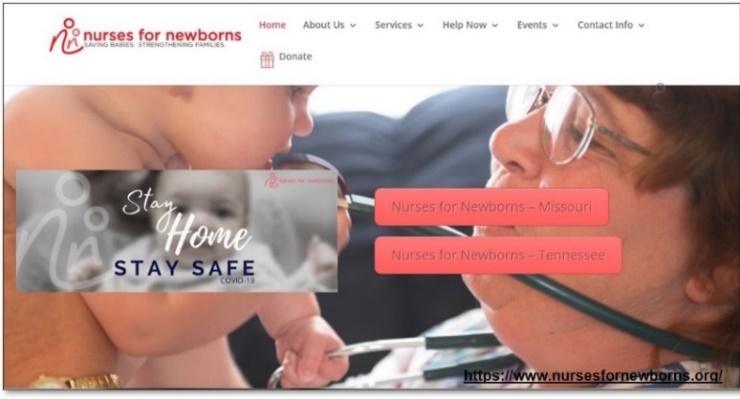
****

Figure . Main page of the Nurses for Newborn website.



Figure . My son, Milo, the motivation for creating a network of resources for parents.

Using Nurses for Newborns as a starting point to build my project idea, I planned my project to create a dashboard that could be used by non-profit organizations as well as parents within the St Louis County and City. This dashboard is intended to be a resource to help identify childcare facilities that may best meet their needs or find medical help. Instead of a parent calling a long list of childcare facilities, for example, they would be able to go to the dashboard and narrow their search to a much smaller and manageable list that provides contact information. For non-profit organizations in the St Louis area, this dashboard can visualize basic childhood indicators as data points that can help in directing resources. The tool could also be a self-service option that might reduce the workload for the local non-profits.

# Background

One of the leading commercial software options for visualization is called Tableau. For this project, I used the free version of Tableau, called Tableau Public (*Tableau*, 2020). Often, publicly available data requires reformatting and cleanup before it can be used in Tableau visualizations. The Tableau visualizations created for this project are made up of two data centric dashboards, one shows a variety of graphs to be investigated through filters, and the other is a geospatial (map) interface that gives context to results of similar filters and queries. Using data provided by different parts of the State of Missouri (specifically, the Department of Health & Senior Services and the Department of Social Services) enabled me to use publicly available government furnished and vetted data. This capstone project is intended to make information about critical childhood resources discoverable to new parents as well as non-profit and government organizations that serve parents and children in the St Louis County and City communities.

A dashboard is comprised of individual visualizations and filters (Figure 3). This can be as simple as one visualization with one or two filters or it can have many individual visualizations with many filters.

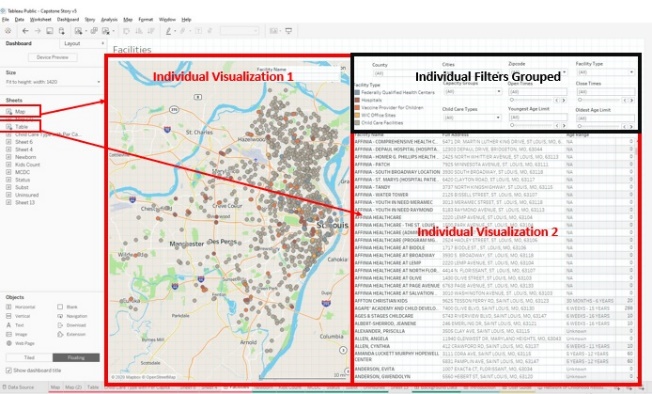


Figure . Overview of Tableau configuration with multiple visualizations and filters.

Tableau allows a person to create custom dashboards to fit the data being used or the purpose of the dashboard/the question being asked. A person creating visualizations can create a single dashboard to be published as a standalone project or they can produce a story which is multiple dashboards published in one project. For this project, two data-based dashboards (Facilities and Background Data) comprised of multiple types of visualizations and filters were created as well as two dashboards (Introduction and User Guide) with project specific information (Figure 4).

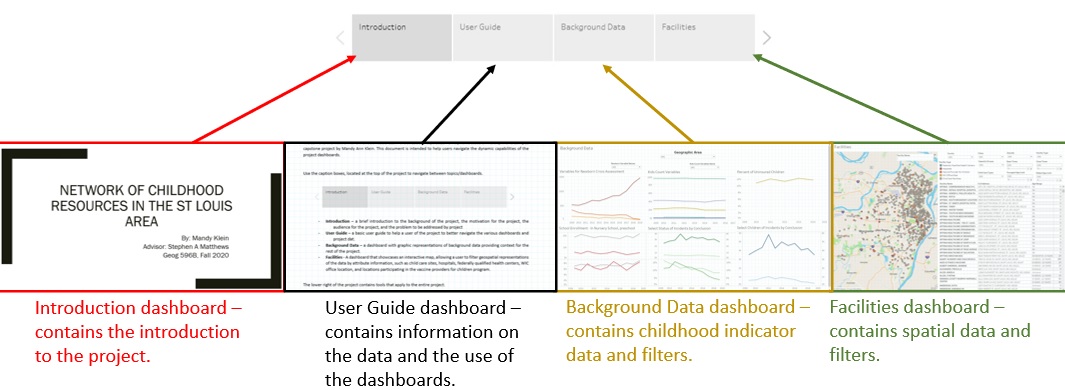


Figure . View of the project story as it relates to each of the individual dashboards in Tableau Public.

This paper is divided into 5 major subsequent sections with additional subheadings within each section. In section 3-part a, details regarding the data sources used in creating this project are detailed including where the data was obtained from, any known limitations, and how the data was used for the project. In part b, how the data was cleaned, formatted, and processed for use in the project is detailed for each data source. In Section 4, details on how the Tableau dashboards/story have been created are explained. This section is broken into five main parts following the workflow from loading the data through to publishing the story. Section 5, details future improvements and breaks down the types of improvements by focus area. Section 6 is the conclusion of the paper where the purpose of the project, lessons learned, and the limitations of the project as well as the data are all summarized. Section 7, contains the bibliography of the sources used to create the final project. Finally, the last section is Appendix A which contains the Tableau User Guide created to aid new users with understanding the data and effective manipulation of the dashboards.

# Data Harvesting & Methodology

For this project I gathered data from local and state governments, a non-profit, and a GIS data retrieval website.

## Data Sources

All of the data sources were located online from local/state government sources or a not-for-profit organization. Online data is good because it is assumed to be most current due to the government entities monitoring and reporting on information related to childcare sites, Vaccines for Children sites, as well as the annual reports on children’s services. However, it can also be problematic because it often contains inconsistent formatting and errors that likely result from free text entry, unbounded data entry, or from bulk data loads within the system hosting the data. Even with the problems associated with cleaning the data, it is still the best available source for the information to be displayed within this project and there is an expectation the data has been through some validation prior to being made available to the public. The dashboard is built using 6 primary sources and 1 service used to geocode addresses:

1. Missouri Department of Health and Senior Services (MDHSS) publishes information on all childcare facilities in the state and can be queried by county as well as other limited criteria such as facility name (*MDHSS - Child Care Facility Search*, 2020). Information contained in the MDHSS database includes facility name, phone number, address, hours of operation, age range, and capacity. 86% of the addresses geocoded without error and 14% did not contain enough information to place. The site itself does not provide any information as to the frequency of updates, however, looking at several of the facilities showed data updates through November 2020. This dataset is used in the spatial dashboard.
2. Missouri Department of Health and Senior Services (MDHSS) publishes information on the Vaccines for Children program locations (*Vaccines for Children Providers*, 2020). Information contained in the database is only partial address information. Google Maps was used to supply ZIP code information to aid in geocoding. 78% of the addresses geocoded without error and 22% did not contain enough information to place. A second search of Google Maps was performed to see if a better or more accurate address could be obtained based a review of OpenStreetMap and imagery. This search did not yield any additional geocoded results. This dataset is used in the spatial dashboard.
3. Missouri Spatial Data Information Service Open Data Site from the University of Missouri - Geography Department (MSDIS) archives and retrieves geospatial datasets that were used in multiple phases of this project (*MSDIS*, 2020). Hospitals, federally qualified health centers, and WIC Office point shapefiles were downloaded for display in the spatial dashboard.
   1. County boundary data were obtained from this source and used in the geoprocessing phase of the project to clip other datasets to isolate only the county and city as well as for data verification. This dataset is not included in the final project.
4. Missouri Department of Social Services – Missouri Children’s Division publishes an annual report of county level statistics regarding children services (*Children’s Division Annual Report*, 2020). Reports from 2009 through 2019 were used in this project. Several variables including Newborn Crisis Assessment and data on Substantiated, Unsubstantiated, and Unsubstantiated – Preventative Services Indicated were included in this project. This data is used in the background data dashboard.
5. The Annie E Casey Foundation’s KIDS COUNT Data Center publishes a dataset containing data on children within predefined geographic areas (for this project, county) (*KIDS COUNT*, 2020). Several variables were extracted from this site to provide additional variables on the background data dashboard.
6. The Missouri Census Data Center (MCDC) provides US Census data among other datasets both nationwide as well as Missouri specific datasets (*MCDC*, 2020). For this project, enrollment in nursery school, preschool data was obtained to integrate into the background data dashboard. This dataset contains both counts of children enrolled in nursery school, preschool as well as the percentage of enrolled children. The data cover the decade between 2009 to 2019.
7. The US Census Bureau geocoding service uses their Tiger Line road network data to geocode addresses from their website (*Census Bureau Geocoder Tool*, 2020). The datasets obtained from the Missouri Department of Health & Senior Services for the spatial dashboard was spatially located via the Census geocoding service using the Public\_AR\_Current Benchmark (Address Range) on the service. The output file contained XY coordinates which were then converted to a shapefile. Without this service, geocoding the address data from MDHSS would have been much more challenging as one service only allowed a finite amount of records to be geocoded at a time and other services required payment for use.

## Processing

Each source presents unique processing and formatting challenges. Data manipulation and data cleaning efforts can be performed in a range of software packages, ranging in cost, and complexity. In this project, MS Excel (*Microsoft Excel*, 2020) is used extensively for data manipulation in tandem with Notepad ++. Notepad ++ (*Notepad ++ Home Page*, 2020) is a free source code editing software for Windows. In addition to supporting a wide range of programming and scripting languages, Notepad++ provides powerful text editing tools. The macro recording functionality is leveraged in this project for performing complex data parsing tasks required to ensure formatting is compatible with geocoding engines and Tableau.

1. The Missouri Department of Health and Senior Services provides a Child Care Search page on their website to extract information on childcare facilities (*MDHSS - Child Care Facility Search*, 2020). Using the Child Care Search page (Figure 5) from the Missouri Department of Health and Senior Services, tabular data on childcare locations as well as attribution information were utilized in this project. Several searches were performed using the Child Care Search page within both St Louis County and St Louis City so I could determine the best data available from the search tool as well as the structure of the data.

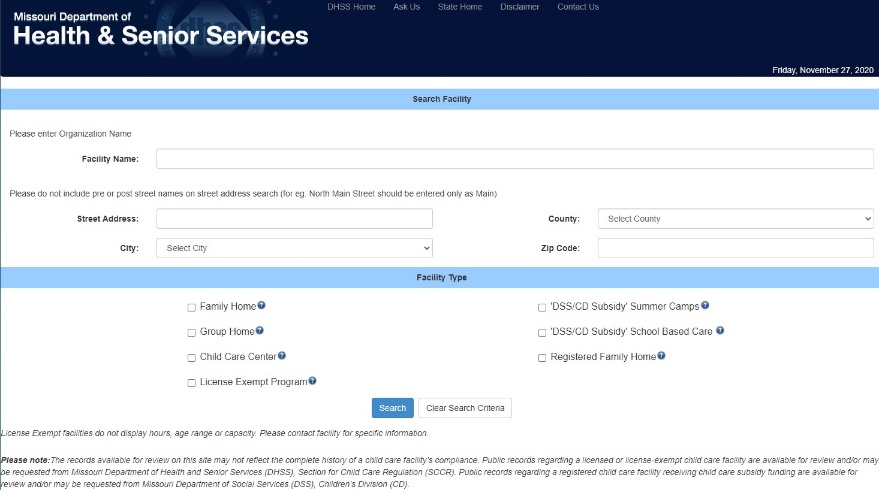


Figure . Missouri Department of Health & Senior Services Child Care Search page.

I focused on St Louis County and St Louis City for the project and copied and pasted all of the records listed under five of the Facility Types into MS Excel spreadsheets to be manipulated into a format readable by a geocoding service and Tableau (Figure 6). Metadata information on the different facility types and site information were also collected and input into the spreadsheet. Those five facility types are: Family Home, Group Home, Child Care Center, License Exempt Program, and Registered Family Home (defined in the User Guide section).

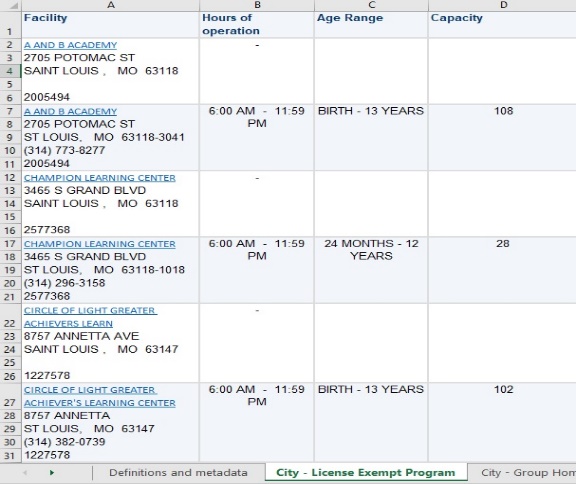


Figure . Unformatted tabular data copied from the Missouri Department of Health and Senior Services Child Care Search Page.

Each of the ten individual data listings obtained from MDHSS utilized in this project required slightly different modifications to correct formatting problems. Some of the listings could have universal modifications, such as parsing text into columns, while others did not have consistent enough formatting to support universal functions. For example, the county level License Exempt Program data contains relevant facility information in rows, versus the needed columns (Figure 7). All facility information for each site except for address is in the first line of each record and does not contain characters or spaces that can be used to parse into columns. Additionally, facility information is not consistently complete, with some records missing key information, such as Hours of Operation or Age Ranges. Additionally, the addresses do not all use the same number of lines due to missing information.

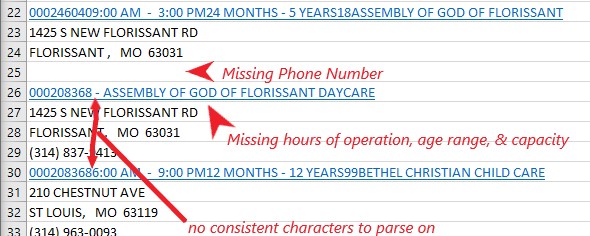


Figure . Example of raw MDHSS county level License Exempt Program childcare data in MS Excel.

The remedy for the above issues involves a combination of manual and semi-automated modifications to each of the ten data listings. Each sheet is exported as comma separated text files (\*.csv). Because commas appear in some of the facility names, the field delimiter specified upon export is a semi-colon (;). Each sheet is opened in Notepad++ individually (Figure 8).



Figure . Example of unprocessed MDHSS childcare data opened in Notepad ++.

For the first step, each record must be brought onto the same line, with the proper delimiters in place (in this case, semi-colons). Because the modifications that have to be made aren’t consistent, the record & play macro functionality cannot be utilized for the entire file clean up. To use the macro, click on the end of the top row and press the record button. Press delete, go to the end of the line, press delete again, and repeat until all of that record is contained on one line. Then arrow down to the next record and hit the end button to get to the end of the record. After these steps have been taken, press the playback button and if the records are consistent, use the ‘Run Macro Multiple Times…’ button and choose the number of times or run until the end of the file. If there isn’t a pattern to allow for a more automated processing, click on the end of the next record’s first line to end. Once all of the records have been processed the file has all of the records on one line (Figure 9).

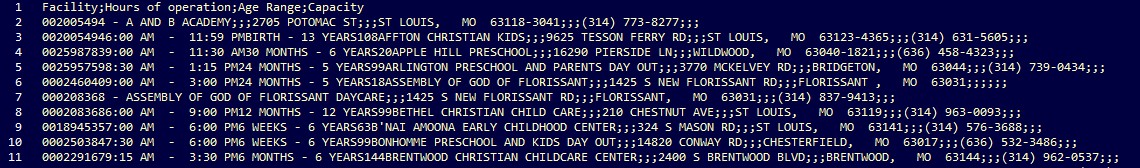


Figure . Childcare data on one line in Notepad ++.

Facility IDs are unique 9-character numbers for each facility and can be parsed using the multi-line cursor function in Notepad ++. Hold the *alt* button, click the end of the first Facility ID and drag straight down to the end of the file, keeping the cursor after the 9th character. A single vertical line starts to blink like a regular cursor. Next add a semi-colon as a delimiter so when the file is opened in MS Excel the Facility ID will be in its own field. Save and close the file. Then open the file in MS Excel and select a delimiter of a semi-colon. Select the Facility IDs in the first column, cut the records then scroll to the last used field, and paste. Remove empty fields from the data. Select the *Hours of Operation* field (first field) and use the sort command to sort it. Make sure to expand the selection when requested. This puts the hours of operation more or less in a group. Save this file as a csv again but with a different name and using semi-colons as delimiters.

Open this newly created file in Notepad ++ again (Figure 10). This step has to be semi-automated due to data entry inconsistencies. For the lines that only list facility names (versus the times), use the *alt+click+drag down* technique to select the first position on each line and then add a semi-colon (Figure 11).

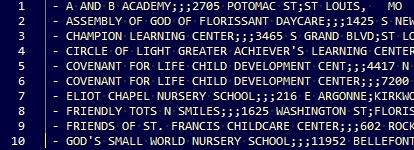


Figure . CSV data prior to semi-colon added to the first position.

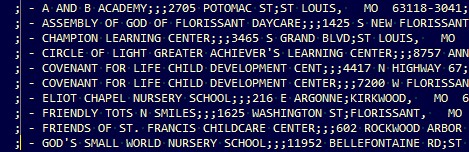


Figure . CSV data after adding semi-colon to the first position.

Next use the *alt+click+drag down* technique to select places between hours of operation and ages and enter semi-colons. Save the csv and open it up in MS Excel. Cut/paste the hours of operation column to move it to the far right and delete the empty column on the left. Save the file, close MS Excel, and open again in Notepad++. Repeat the last steps, entering a semi-colon in the far-left position for all records that are missing age ranges. The simplest way to get the next required parse in is to do a find/replace one record at a time. Press *ctrl+H* go open the *replace* dialog, type in “YEARS” and “YEARS;” (Figure 12). Proceed one record at a time, skipping when needed by hitting *find next* instead of *replace* (if the range is x years – y years, the semi-colon only goes after the second years) (Figure 13). Save the file and close it.

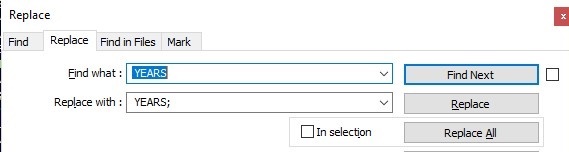


Figure . Find & Replace to parse the age range data.

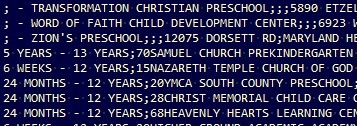


Figure . Data after the Find & Replace for the age range data has been completed.

Reopen the file in MS Excel, cut and paste the age range field in the same manner as previous fields then delete the empty field afterwards. Next, save and close the file then reopen it in Notepad ++. Using the same methodology as the Age Range field, repeat the process for the Capacity field (between the capacity and the facility name). Remember to add the semi-colon to the beginning of the lines that do not have capacity data as described for previous fields. Finally, save and close the file then reopen it in MS Excel. Reorder and remove fields, edit the field names as needed, and then save the files as a MS Excel spreadsheet.

Once all 10 of the worksheets in Excel were formatted, all of the information was combined into a single worksheet for geocoding. I then began to sort the data in various ways (by county, by facility type, by address, etc.) to locate duplicate records. This was present between the Family Homes and Registered Family Homes facility types due to the location being placed into both types. When this situation was present, I would normally keep the registered family home due to it having the most attribution available. Another part of cleaning the data up was to add “Unknown” or “NA” (Not Available) in fields were the attribution wasn’t supplied by the state (Figure 14).



Figure . The cleaned childcare dataset in MS Excel.

Then I focused on cleaning up the inconsistent formatting for the attribution fields, specifically the Age Range and Hours of Operation fields. The Hours of Operation field was also split into two additional fields, one for Opening Times and one for Closing Times to allow a user of the dashboard more flexibility in searching for childcare options. The same process was used for the Age Range field which was split into Youngest Age and Oldest Age in addition to the original field, for more options to filter on the dashboard. I used the US Census Bureau’s Geocoding service (Figure 15) to geocode the childcare facilities addresses without any additional attribution information included in the file (*Census Bureau Geocoder Tool*, 2020). After getting the results of the geocoding service, I combined the geocoded file with the associated attribution and verified the records were correct. There were a number of records that did not geocode in the childcare facilities data and they were removed from the final dataset. Due to the geocoding issues, this dataset is known to be incomplete. The records from the Family Home and Registered Family Home facility types would be next to impossible to locate without a correct physical address due to it being a home in a residential area and indistinguishable from the homes around it in most cases. Additionally, the originating source does not provide a currency statement for the data.

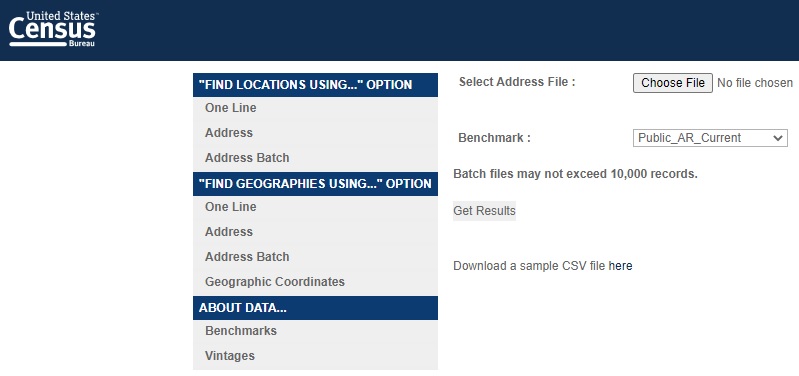


Figure . The US Census Bureau geocoding service page.

1. Then I collected tabular data from the Vaccines for Children program (Figure 16) and placed the data into a MS Excel spreadsheet (*Vaccines for Children Providers*, 2020). The data was prepared for geocoding by placing all of the components of the address into separate fields. This file required examination of individual records and research to locate ZIP code information not provided by the listing from the original data source. The spreadsheet with only the address information and a unique identifier were entered into the US Census Bureau’s geocoding service (*Census Bureau Geocoder Tool*, 2020). I verified the results afterwards in Tableau to ensure they would display and in the correct area. There were several records that did not geocode due to issues with the address, I attempted to manually find addresses for those locations using Google Maps and web searches but was unable to find reliable location information. The next step was to combine this data for the Vaccines for Children program with the childcare facilities geocoded spreadsheet. After combining the two datasets, I filled in attribution not native to the Vaccines for Children program data to help complete the data for visualization in Tableau Public.

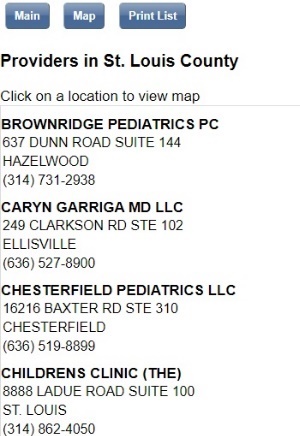


Figure . Vaccines for Children program search results.

1. Geospatial data was obtained from the Missouri Spatial Data Information Service Open Data Site from the University of Missouri Geography Department website (*MSDIS*, 2020). The four datasets I downloaded from the MSDIS were county boundaries (used only for reference, verification, and clipping to the project area of interest), hospitals, federally qualified health centers, and WIC (Women, Infants, and Children) Offices. I loaded all four into ArcGIS Desktop 10.8 provided by The Pennsylvania State University WebApps page (*ESRI*, 2020).

All of the entities (hospitals, federally qualified health centers, and WIC Offices) were clipped to the project county boundaries. The clipped \*.dbf files were moved from the Penn State virtual machine to my local folder on my laptop. I opened each of the \*.dbf files in Excel and copied the necessary attribute and location information from those files into the spreadsheet containing childcare and Vaccines for Children program information. Attribution for fields where information is unknown or not applicable were updated in the newly added data. Additionally, a unique identifier was added for all of these records.

With all of the geospatial data combined into one file with complete attribution, it was converted to a \*.txt file and then brought back into the Penn State virtual environment. I opened the county boundaries shapefile and then loaded the text file in ArcGIS Desktop 10.8 (*ESRI*, 2020). The original facilities shapefiles were added for spatial verification purposes. I chose to Display XY for the text file. After spot checking several random facilities to ensure the spatial accuracy was maintained, I exported the displayed XY to a shapefile (Figure 17). I moved the shapefile of all facilities to my local project folder. Then I loaded the file into Tableau on its own to verify it would visualize and be manipulated by the software properly.

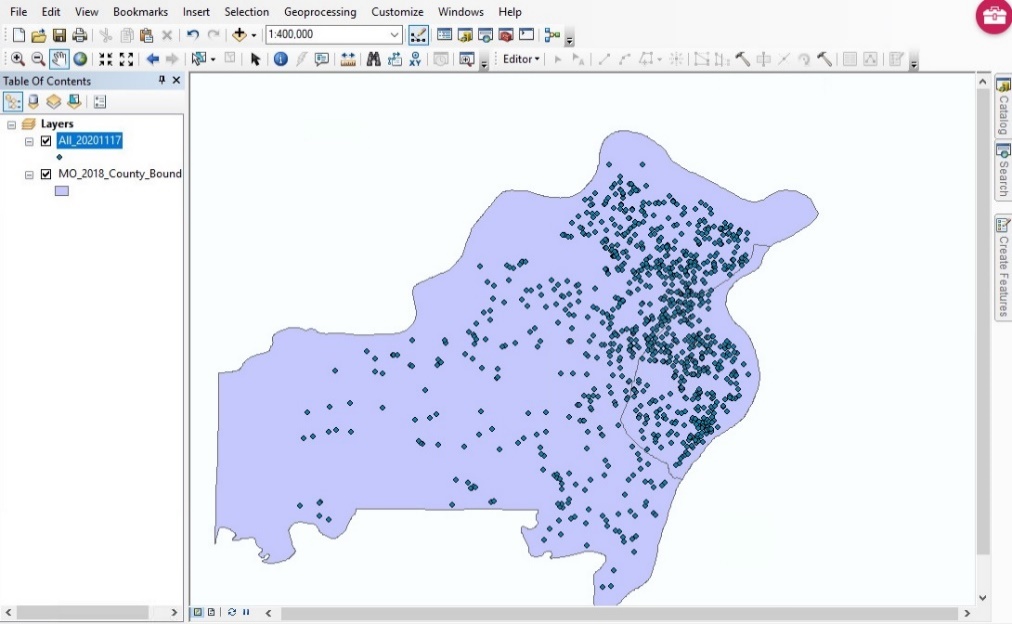


Figure . All GIS data combined together and clipped to the area of interest boundaries.

1. I downloaded the Missouri Department of Social Services’ Missouri Children’s Division Annual Reports for each year between 2009 and 2019 (the most recent year available)(*Children’s Division Annual Report*, 2020) (Figure 18) to my local project folder.



Figure . Missouri Department of Social Services Missouri Children’s Division Annual Report Page.

Then I opened the report for each of the project years and entered information about key indicators that would be visualized on a Tableau dashboard into a MS Excel spreadsheet with metadata on the variables and key terms available from the reports. There were six variables in all used from the Children’s Division Annual Report including three for Newborn Crisis Assessment (Figure 19), Percent Uninsured, Status of Incidents by Conclusion, and Status of Children.

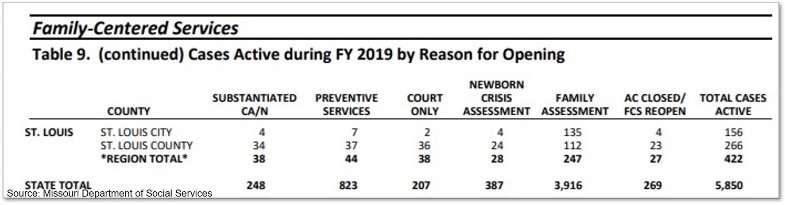


Figure . One of Missouri Children’s Division Annual Report’s Newborn Crisis Assessment indicators.

The field structure is Year, Variable Full Name, County, and then several fields for the type of data (Number, Rate, Percent). After those fields from the data source, I added a field for the data source site and the website address to assist with keeping track of each data source as the project progressed. Some of the variable names changed over the project time frame. Both variable names have been preserved in the spreadsheet. For example, within the Newborn Crisis Assessment heading the variable for CA/N (Child Abuse/Neglect) Incidents and Referrals During FY (XXXX) changed to CA/N Hotline Reports and Referrals during FY (XXXX) in 2016 (Figure 20). In Tableau this will be corrected by grouping the records with both names and entering one name to display for the variable.

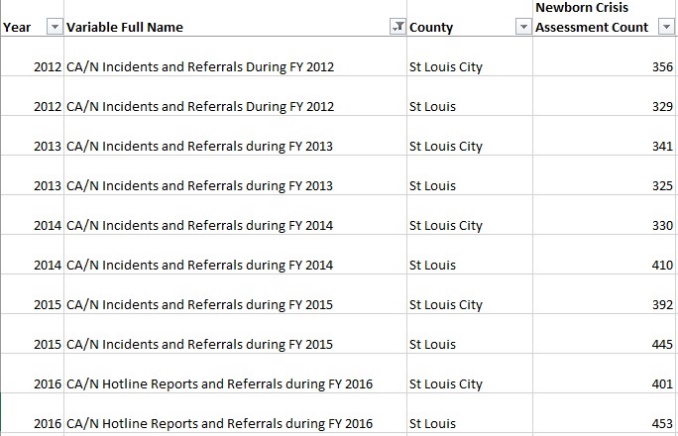


Figure . Example of a variable name change.

1. Next, I gathered additional variable data from the Annie E Casey Foundation KIDS COUNT Data Center’s page to use on the Background Data dashboard (*KIDS COUNT*, 2020). Six variables were gathered for the project area of interest within the 2009 to 2019 timeframe. The variables did not span the entire time frame of the project with five stopping in 2016 and the remaining variable only spanning 2010 to 2014. Each variable was extracted from the Kids Count Data Center Page (Figure 21 and Figure 22) in a text format.

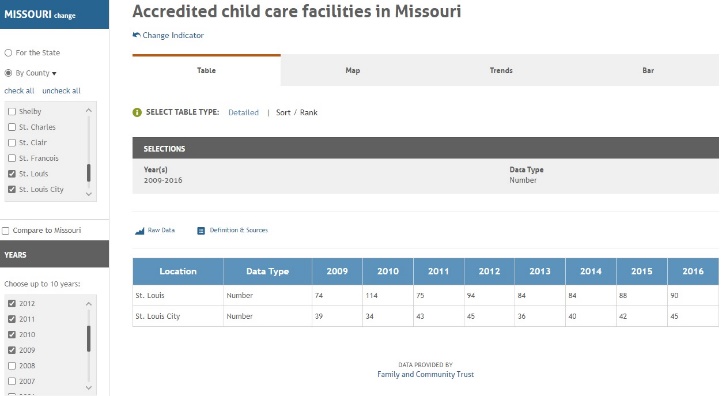


Figure . ‘Accredited child care facilities in Missouri’ data from the KIDS COUNT Data Center page.



Figure . ‘Preventable hospitalizations for children in Missouri’ data from the KIDS COUNT Data Center page.

All of the data from the KIDS COUNT Data Center page was gathered and entered into a spreadsheet with metadata about the variables. The variable information for this data source was formatted in the same manner as the data from the Missouri Children’s Division Annual Reports so they would be better visualized within the Tableau software and to function in a similar way to the each other. The fields created within MS Excel were Year, Variable, County, and fields for the type of data (Number, Rate, and Percent) (Figure 23). After the data was structured and entered in, fields for the data source site and the website address were filled out to assist with keeping track of each data source. Then this data from the KIDS Count Data Center page was added to another worksheet within the MS Excel spreadsheet containing the Missouri Children’s Division Annual Report (Figure 24).

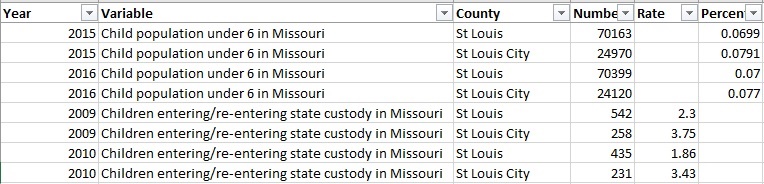


Figure . Example of KIDS COUNT data formatted to match the Children’s Division data.



Figure . Final KIDS COUNT data variable tabs.

1. The last variable gathered for the project was collected from the MCDC (Missouri Census Data Center) website (*MCDC*, 2020). After exploring the data contained on this website and looking at what was relevant to the project and available during the project timeframe, I decided to include information about School Enrollment – In Nursery School, Preschool from ACS (American Community Survey) at the county level (Figure 25). This data source has several scales available including national, census tract, and metropolitan area among others.

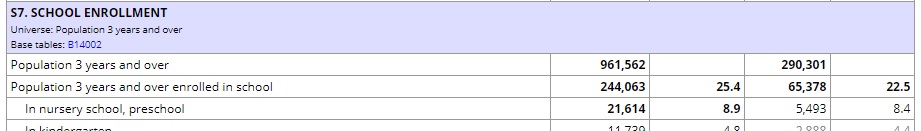


Figure . Sample of School Enrollment data available on the Missouri Census Data Center (MCDC) website.

After gathering the data from 2009 to 2019 for both St Louis County and St Louis City, I entered all of the data into a MS Excel spreadsheet with fields for Year, Variable Name Full, County, Number, and Rate. Then I created fields for the source site name as well as the website address. After completing any additional clean up on the file, I copied the data and pasted it into a separate worksheet within the KIDS COUNT and Children’s Division data file.

Once all of the workbooks were combined into a single workbook with multiple worksheets (Figure 26), each worksheet was reviewed to ensure no issues were created while copying the data from one file to the final file. The ‘All SocialServices’ worksheet contains all of the variables from the Department of Social Services as well as the four black tabs to the right. This constitutes the final workbook data and worksheets used in the Background Data dashboard.



Figure . All Background Data workbook tabs added and completed.

1. The US Census Bureau geocoding service is used in this project to supply latitude/longitude coordinates for each entity by geocoding addresses to the US Census Bureau’s transportation network (*Census Bureau Geocoder Tool*, 2020). This service uses Tiger Line road network data to geocode addresses to the centerline of the correct block.

Gathering all of the variables into one Excel spreadsheet makes dashboard creation much simpler as they all have the County field in common and all of the variables are structured in a similar manner. After moving all of the tabular datasets into one MS Excel spreadsheet file, I loaded the file into Tableau Public on its own to ensure it would visualize within the software as expected. During this initial test of the variable data, I discovered one of the KIDS COUNT variables had to be moved to its own worksheet because the percent data was formatted differently from the source and wouldn’t visualize. After this variable was moved to a separate sheet, there were no additional issues with the KIDS COUNT data.

# Tableau Dashboard/Story Creation

Tableau Public 2020.3 is used to create and publish the dashboards for public use (*Tableau*, 2020). As the name implies, the Tableau Public software and hosting site is available for free for anyone to use. The published dashboards are in the public domain along with any associated data as well as the Tableau Workbook. The basic workflow (Figure 27) of the dashboard/story creation was loading the data, creating the individual visualizations and filters, dashboard creation, story creation, and publishing the story.

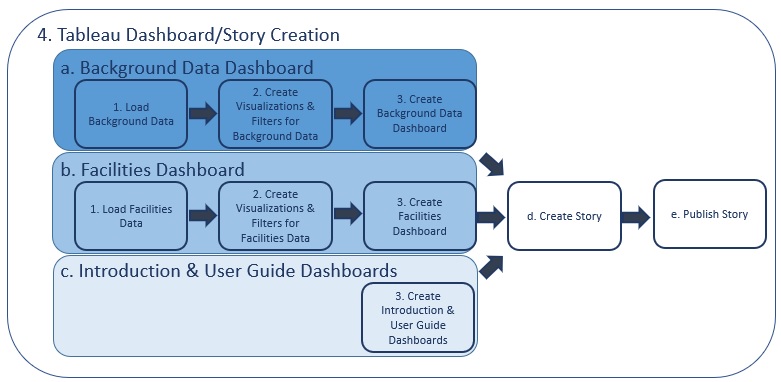


Figure 27. Workflow of the Tableau story creation process.

All of the data must be loaded into Tableau. As Tableau Public has limited data manipulation capabilities, if there are issues with the data (IE, formatting, data type, geographic role) it is best to fix the data in an external application and re-load it into Tableau if errors or issues are found in the dataset.

## Background Data Dashboard

### Load Background Data

Once the tabular data has been ‘cleaned and formatted,’ it can be loaded into Tableau Public (Figure 28). For this project, each dataset is a worksheet and all worksheets are contained in one Excel Workbook. A many-to-many relationship can be defined on the County field that all of the sheets have in common. For this project, the Newborn Crisis dataset is used as the primary dataset.

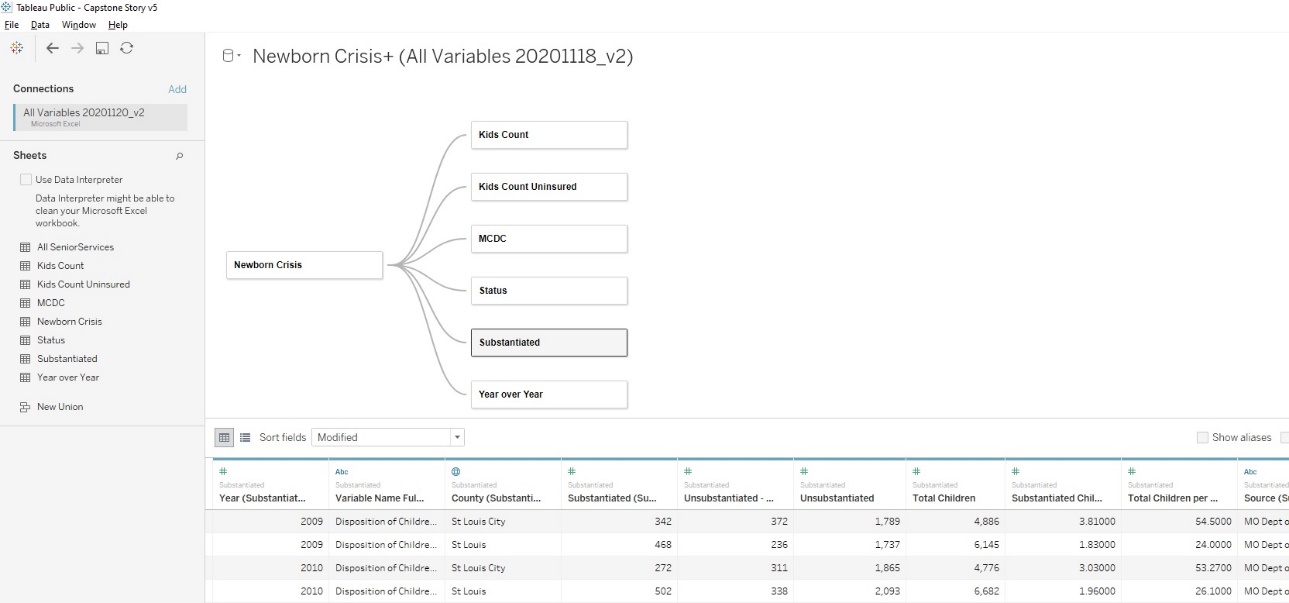


Figure . Data Source tab in Tableau Public showing the data connections for the tabular data.

### Create Visualizations & Filters for Background Data

Each visualization is a unique representation of the data behind it. The axes, units, and filters can differ between each visualization based on what the data can support and the information being communicated to the user. Additional information about the variables can be entered into the ‘Tooltip’ boxes that appear to a user when hovering over an item within the visualization.

The Variables for Newborn Crisis Assessment visualization (Figure 29) has two filters, one for County and another to filter on the variable name. There is also a legend for the colors used in the visualization which has been deleted from the final dashboard to keep it uncluttered. This was customized by removing the word ‘Year’ from the bottom axis and the words ‘Newborn Crisis Assessment Count’ from the other. The ‘Year’ information is self-explanatory to a user and the ‘Newborn Crisis Assessment Count’ is explained via the hover box. All three of the variables were visualized together in a line graph over time to show if each variable was going up or down.

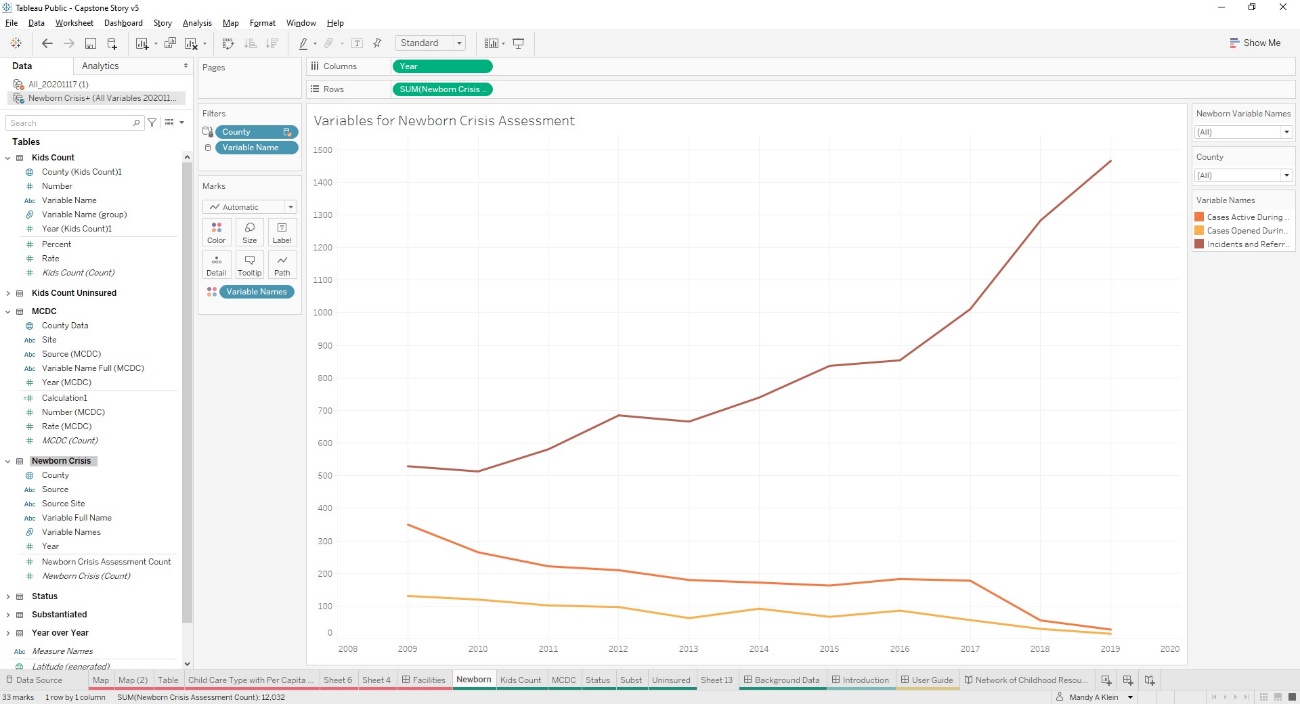


Figure . The visualization of Variables for Newborn Crisis Assessment within Tableau Public.

The next visualization created was for six of the variables obtained from the KIDS COUNT Data Center page (Figure 30). This visualization was created using number instead of the percent or rate data provided by KIDS COUNT. All six of the variables were visualized together in a line graph over time to show if each variable was going up or down. Then the axes were cleaned to remove ‘Year’ and ‘Number’ from their respective axes. The colors of the graph were adjusted to keep them unique from the other visualizations going onto the Background Data dashboard. Two filters were created, one dropdown for Geographic Area (County) and one for the variables themselves to allow a user to see all, one, or any combination of the individual variables.

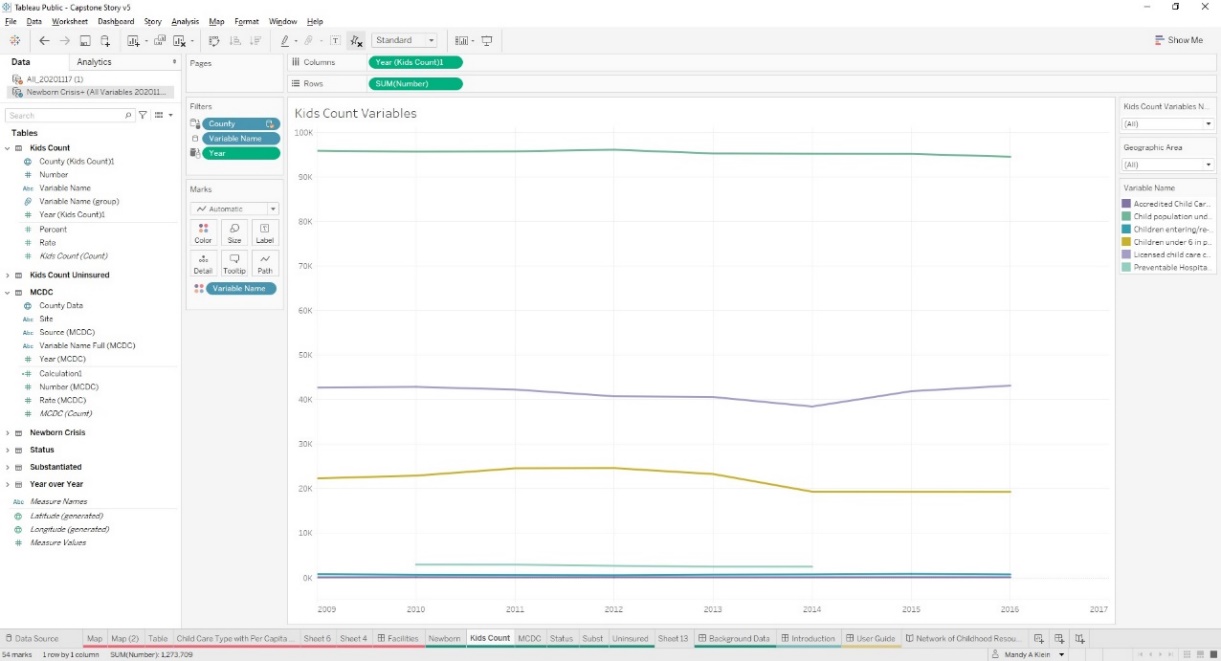


Figure . KIDS COUNT variables visualization within Tableau Public.

The Missouri Census Data Center data was visualized using both the number and percent data by county (Figure 31). Both axes were cleaned up to remove ‘Year’, ‘Percent’, and ‘Number’. The percentage data was adjusted to display as a percentage so a user could quickly understand the difference between the top and bottom portions of the visualization. Additionally, the colors were adjusted to make the visualization distinct from the others on the Background data dashboard and a filter for County was created.

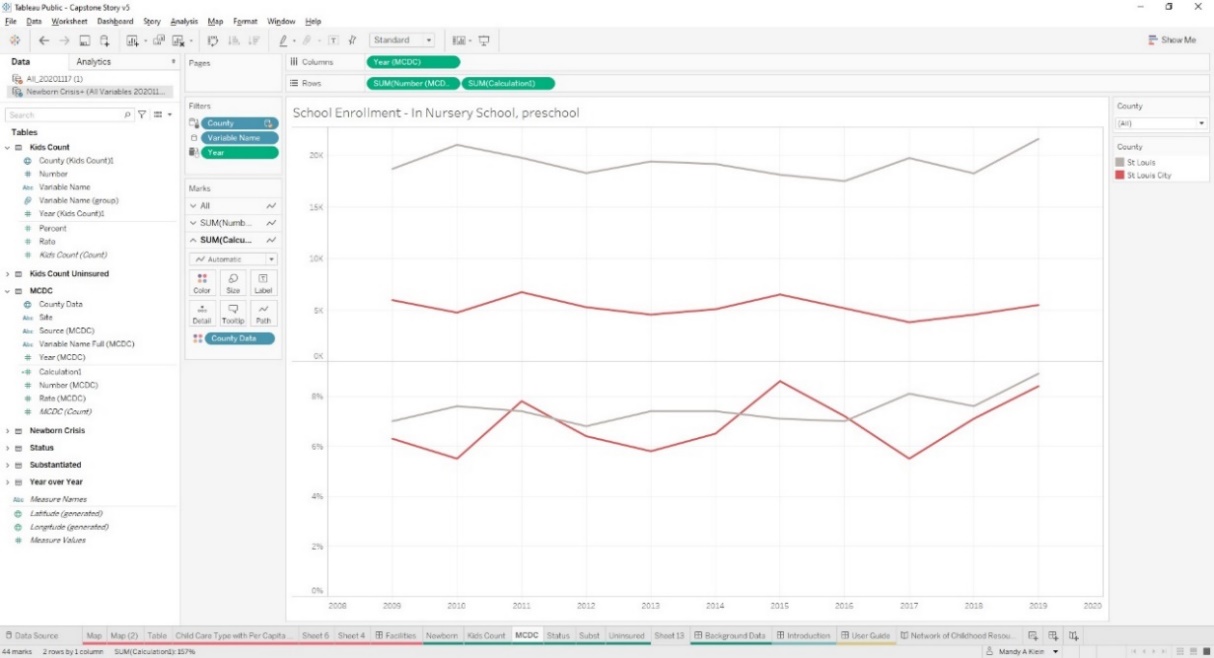


Figure . The Preschool Enrollment variable from the Missouri Census Data Center visualized in Tableau Public.

The Select Status of Incidents by Conclusion visualization (Figure 32) was created by using a synced dual axis line graph over time displaying both available numbers and percentages of Substantiated versus Unsubstantiated incidents based on the Missouri Children’s Division Annual Report data. The axes were cleaned up to remove ‘Year’, ‘Number’, and ‘Percent’ and the axis displaying percentage information was modified to display percentages to aid users. A filter was created and displayed for county data to test the visualization. The colors were adjusted to keep this visualization unique from the others going onto the Background Data dashboard.

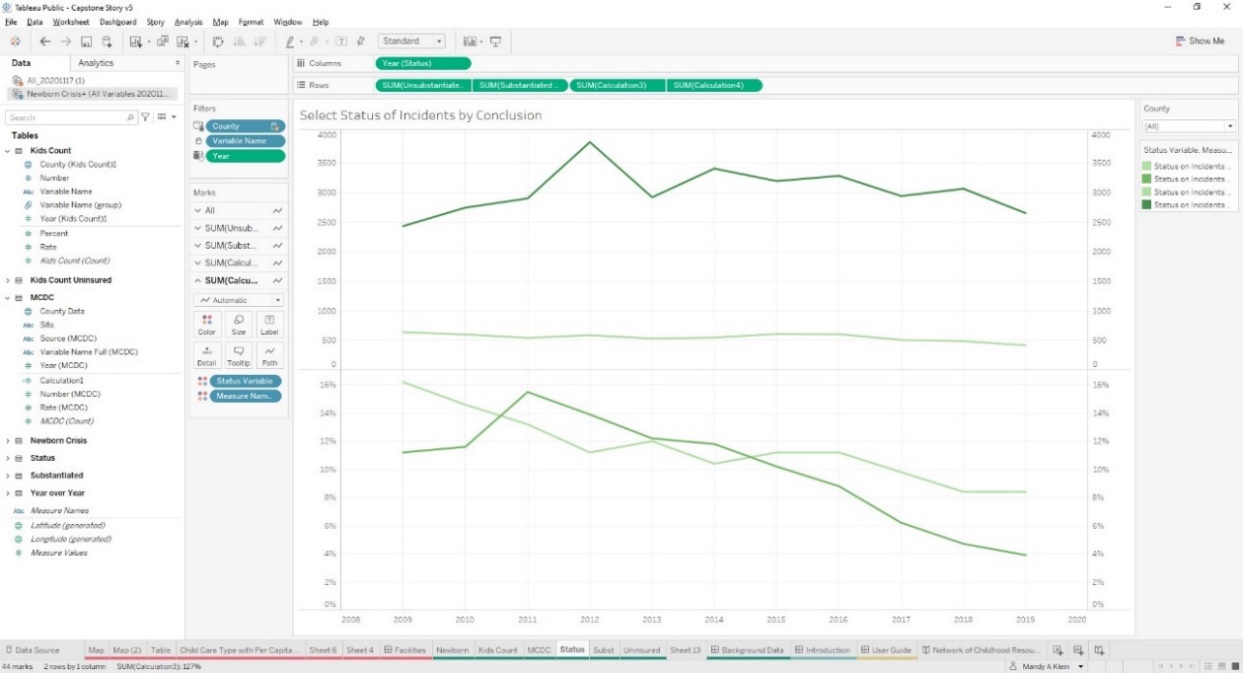


Figure . Select Status of Incidents by Conclusion data visualization from the Missouri Children’s Division Annual Report.

Then I created the Select Children of Incidents by Conclusion visualization with data obtained from the Missouri Children’s Division Annual Report (Figure 33). This is a synchronized dual axis line graph over time displayed based on Unsubstantiated and Unsubstantiated – Preventative Services Indicated. The graph was created and then cleaned up to remove ‘Year’, ‘Unsubstantiated’, and ‘Unsubstantiated – PSI’ from the axes. The colors were modified to ensure this visualization would be distinguishable from the others going on the Background Data dashboard. A filter was created and displayed for county to check the data and ensure it was filtering properly prior to moving onto the final visualization.

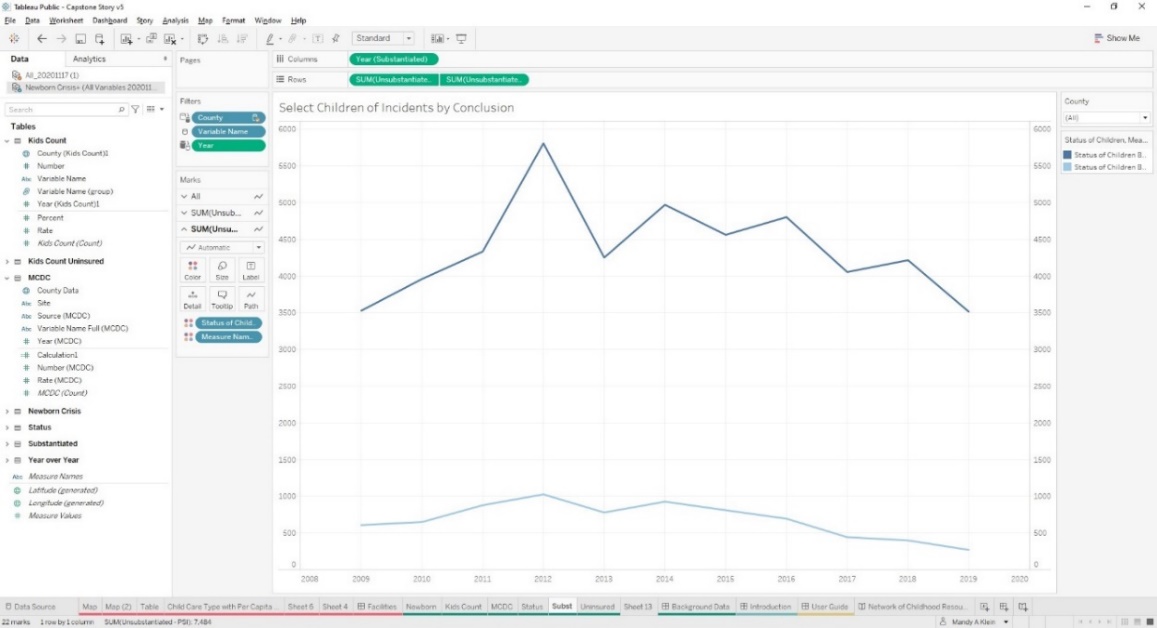


Figure . Select Children of Incidents by Conclusion data from the Missouri Children’s Division Annual Report visualized in Tableau Public.

The last visualization for the Background Data dashboard is the Percent of Uninsured Children (Figure 34) visualization based on data from the KIDS COUNT Data Center. This visualization was created separately from the other KIDS COUNT Visualization due to it having only percentage information and formatting issues as well as I felt it was important enough to display on its own. This data was visualized as a line graph over time based on percentages. The visualization was cleaned up to remove ‘Year’ and ‘Percent’ from the axes and display was modified to show percentages properly. The color of the visualization was modified to use the same color palette as the other KIDS COUNT data to tie them together. A filter was created and displayed based on county to test to ensure the data was filtering as expected.

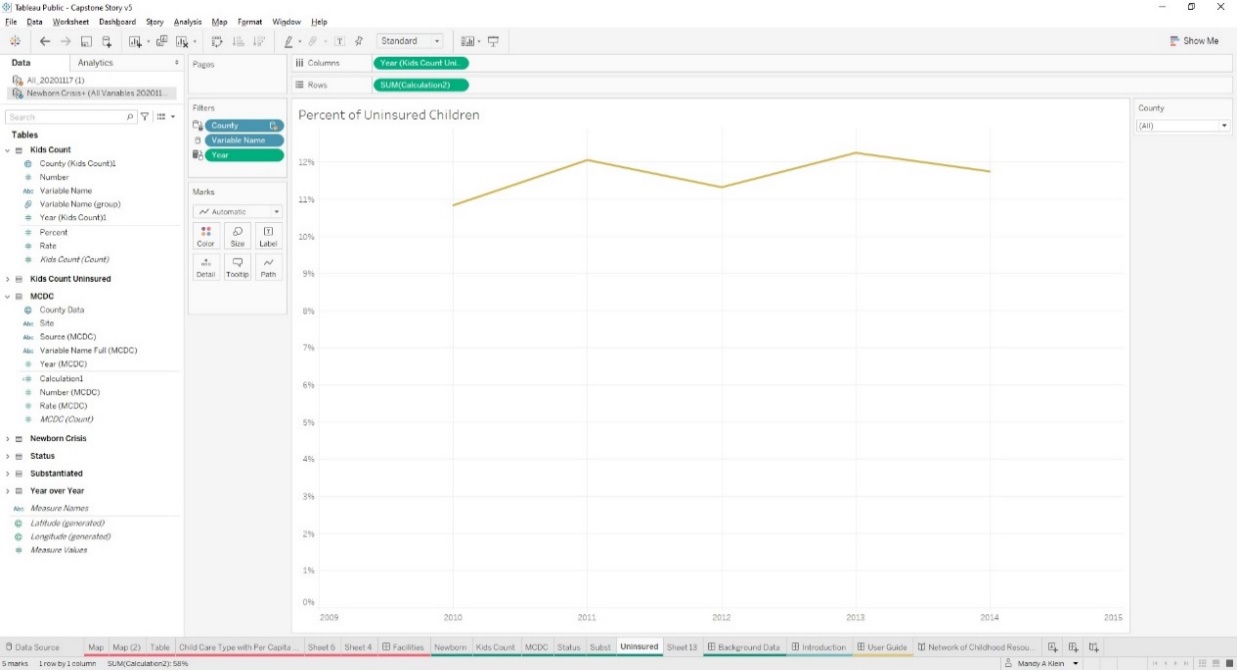


Figure . Percent of Uninsured Children data from KIDS COUNT visualized in Tableau Public.

### Create Background Data Dashboard

After all of the visualizations were created, the Background Data dashboard was created in Tableau Public and all of the visualizations were added one by one. All of the symbology from each of the visualizations was removed to keep the dashboard as simple and uncluttered as possible. Several variations of the layout were experimented with and visualized before the final product (Figure 35) was reached. There are only three filters present on the dashboard; Geographic Area which applies to all of the visualizations, Newborn Variable Name which applies only to the Variables for Newborn Crisis Assessment visualization, and the Kids Count Variables Name which applies only to the Kids Count Variables visualization.

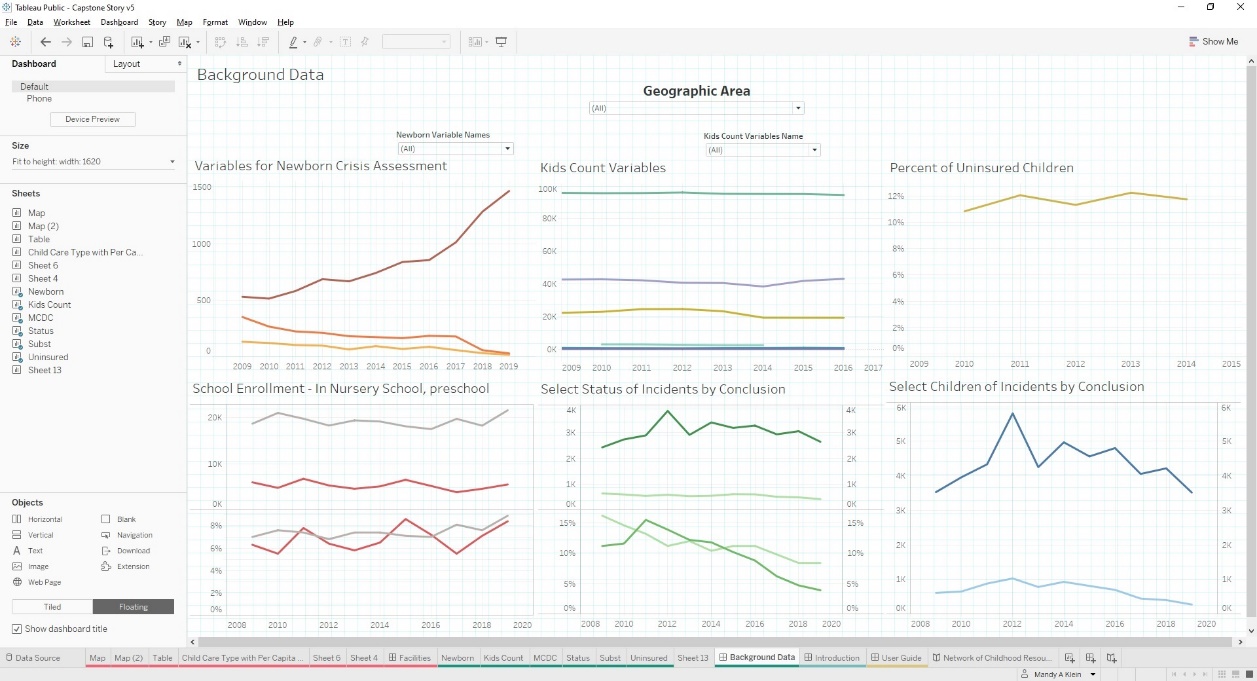


Figure . The final layout of the Background Data dashboard.

## Facilities Dashboard

### Load Facilities Data

To begin I added the GIS data into Tableau using the out of the box Tableau spatial capabilities. Once loaded, the data can be reviewed in a table format in the ‘Data Source’ screen (Figure 36).

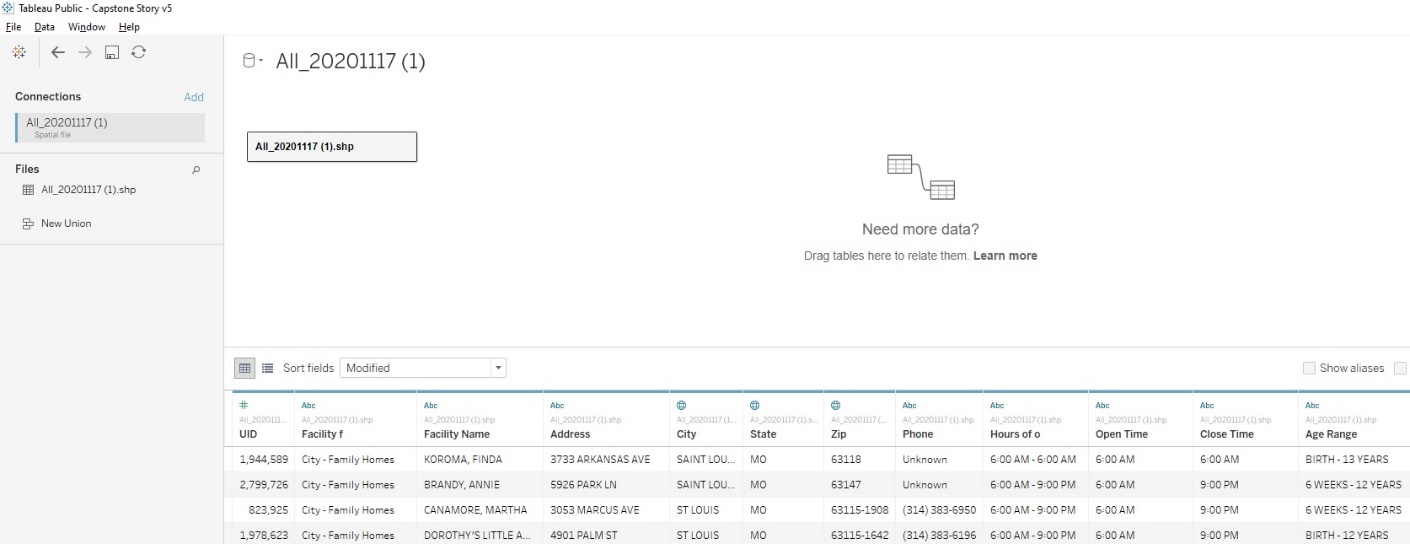


Figure . Data Source tab in Tableau Public showing the fields of the GIS data to be used in the dashboard.

A verification step was taken by moving into one of the visualization sheets and checking to ensure there were two data sources listed (Figure 37) at the top and the associated fields would appear when each data source was selected. This step was to double check both datasets were loaded correctly and were being displayed correctly within the software so the individual visualizations could be made and then incorporated into a dashboard.

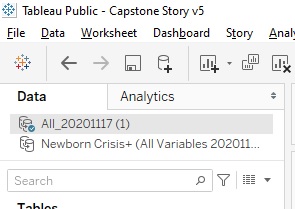


Figure . Verification of two data sources after both were loaded into Tableau Public.

### Create Visualizations & Filters for Facilities Data

The facilities visualization (Figure 38) is an interactive map of the childcare facilities driven by the geocoded childcare facilities and Vaccines for Children data in addition to GIS based files for hospitals, federally qualified health centers, and WIC Offices. To create the map, I had to assign the X, Y fields a Geographic Role within the software so it understood X = Longitude and Y = Latitude. After assigning the proper roles, I added the fields into the display and adjusted their settings so they would display as a map. The background map was then changed to the OpenStreetMap basemap that comes standard with Tableau Public. I chose this basemap to allow users to see where they are in an area, cross roads around the area, and any landmarks around them. Next, I worked on the symbology of the points. I chose to have the childcare facilities as tan so the other types of facilities, which are fewer in number, would stand out to a user. While working on symbolizing, I also darkened the circle around the points to help to differentiate them from the background map. After the data was visualized effectively, I customized the hover box to display important information for each facility (Name, Address, Phone Number, Hours of Operation, and Age Range).

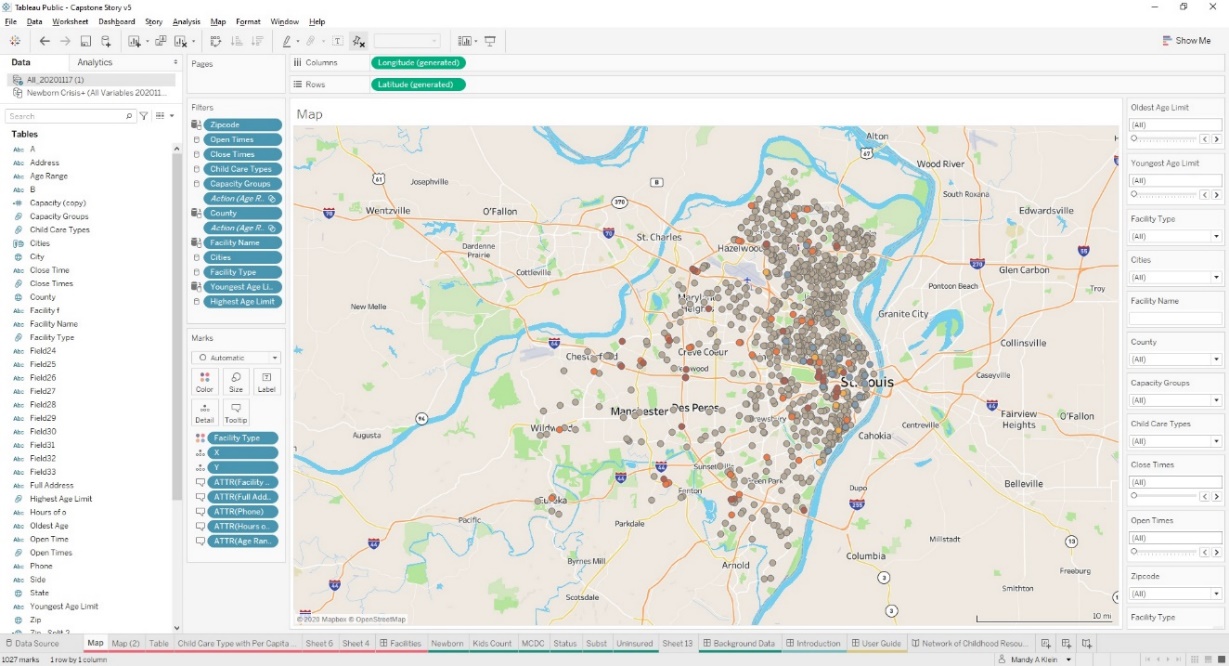


Figure . The facilities map for St Louis County and City symbolized on Facility Type.

Each filter (Figure 39) was created individually, then displayed, and lastly customized (filter name and type) to best filter the data it would be used for. There are dropdown filters as well as slider filters and a search box. The search box allows for free text search on the Facility Name field.

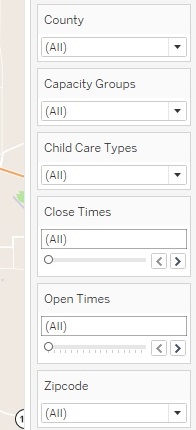
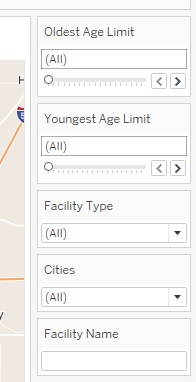


Figure . Filters to allow a user to better refine their search criteria.

Once the filters are complete, the results table (Figure 40) can be created. The results table displays the results of user specified filters and map selections. This table shows the Facility Name, Full Address, Age Range, and Capacity for every facility in the geocoded file. The hover box information was edited for clarity and formatting.

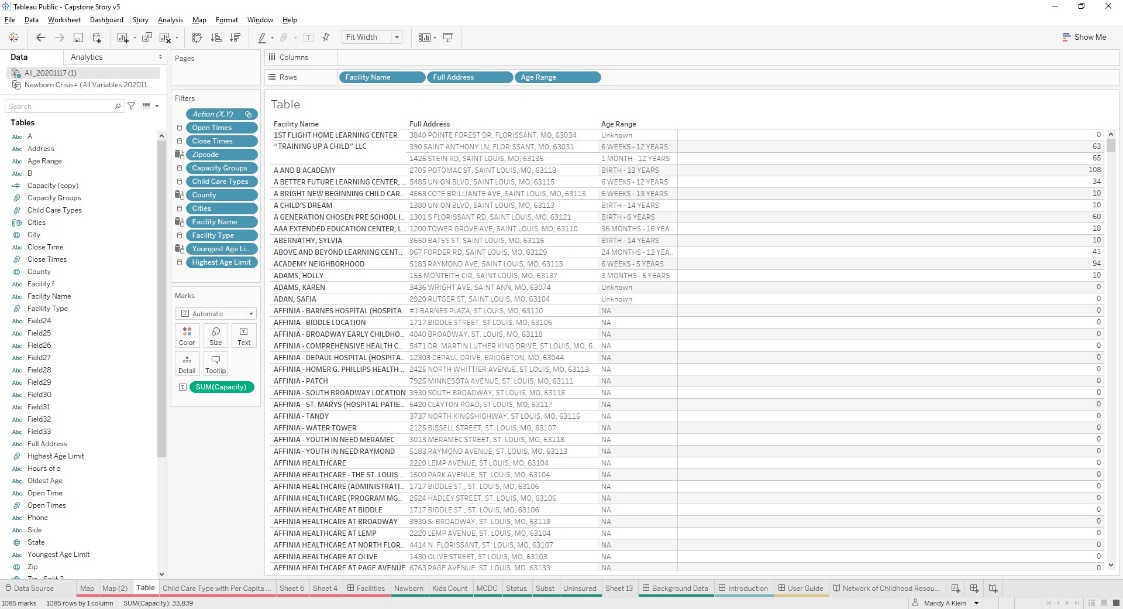


Figure . The table visualization set up in Tableau Public.

### Create Facilities Dashboard

Once complete, the map and table visualizations were added to a new “Facilities” dashboard (Figure 41). Having both visualizations and the filters in the dashboard allows for tweaking and adjusting filters and layout. The Facility Name search box is located within the map to distinguish it from the other filters since it is the only search box. It was placed in the upper right so there would not be a conflict with the “Search Map” feature that is standard within Tableau map visualizations. The filters were rearranged several times to better group them by similar functions; by geographic search, by times, by ages, etc.

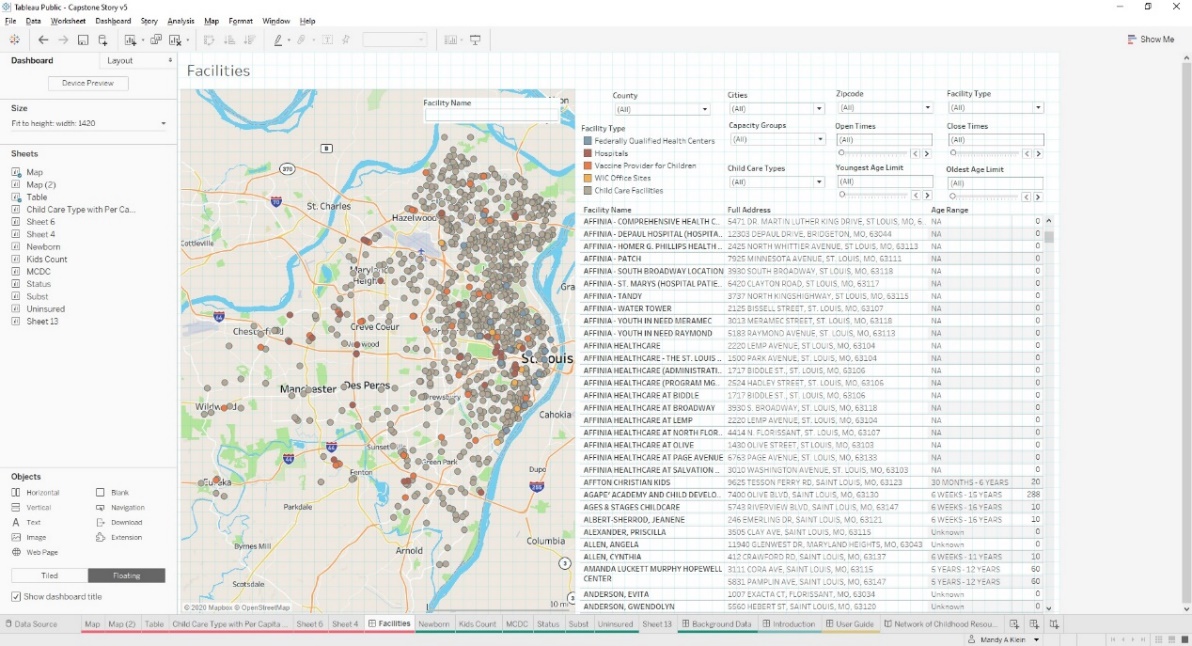


Figure . Facilities dashboard final layout within the Tableau Public software.

## Create Introduction & User Guide Dashboards

For the creation of the Introduction and User Guide dashboards, I created all of the materials outside of Tableau first, saved the materials as screen grabs, and then added them into the appropriate dashboard as images. This was completed after the Background Data and Facilities dashboard were set up and working.

On the Introduction dashboard (Figure 42) in Tableau, I used the PowerPoint I created in Geog 596A as a template for some of the information I wanted to share on this dashboard. I modified the first four slides to reflect the work I had completed and the changes to the project since the end of Geog 596A the previous semester. After completion of the edits and a few reworks of the slides, I saved screengrabs of each of the four slides I planned to use after cropping them to remove any unneeded information. In Tableau, I added each image one at a time and sized it to fit the display. This was repeated for each of the remaining three slides. As I continued to edit the originating slide deck and gathered feedback, I iterated through this process several more times.

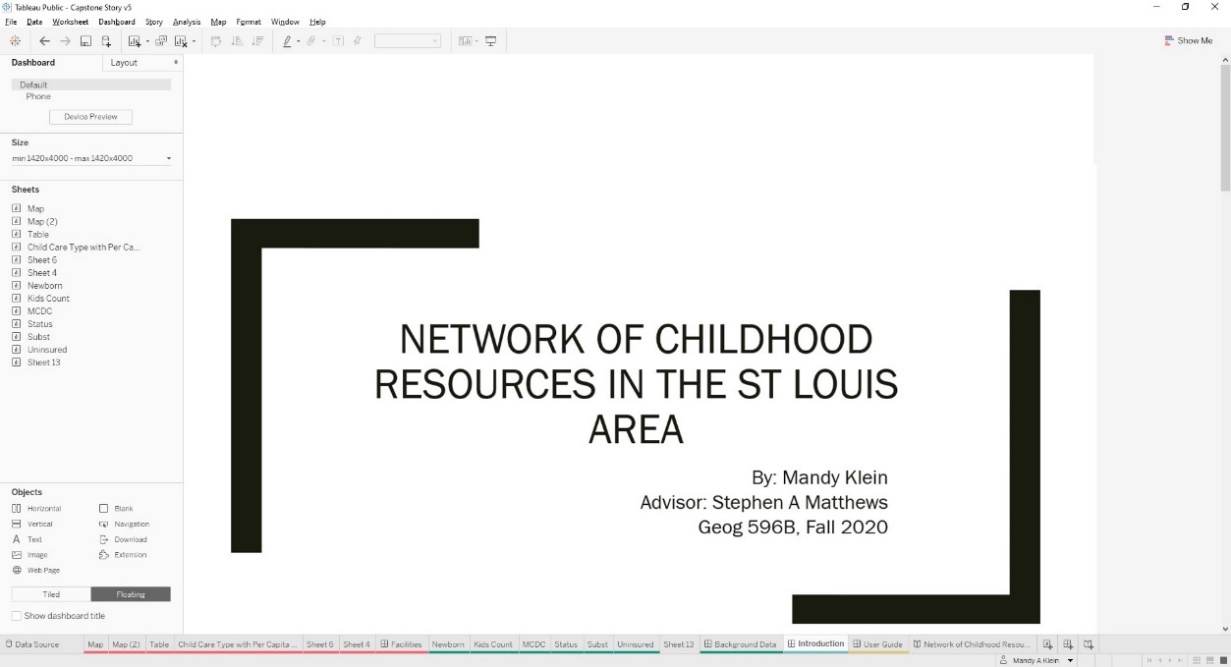


Figure . The Introduction dashboard visualized within Tableau Public.

The final step in the dashboard creation process was to create and integrate a User Guide dashboard (Figure 43) for those who would use the dashboard but not have familiarity with the technology. This was originally created as a Word document with explanations of the Background Data and Facilities dashboards. The user guide not only explains various ways to interact and manipulate a dynamic dashboard, it also provides definitions for the different components of the story as a whole. The guide also explains some information about the different datasets and terms from within the various datasets. At the end of the User Guide is a listing of the sources used throughout the project as reference materials in addition to collecting and editing the data.



Figure . The User Guide displayed within Tableau Public.

## Create Story

After all of the dashboards were set up, a story was set up within Tableau Public (Figure 44). This process was relatively quick and simple since there are not many options to customize a story. I created a blank story then added three more caption boxes to toggle between with blank pages associated to be populated with the different dashboards I created. Next, I named each one to match the different components of the project: Introduction, User Guide, Background Data, and Facilities. The final step in the story creation was to add each of the four dashboards to the correct part of the story and toggle through to ensure each was in the correct location.

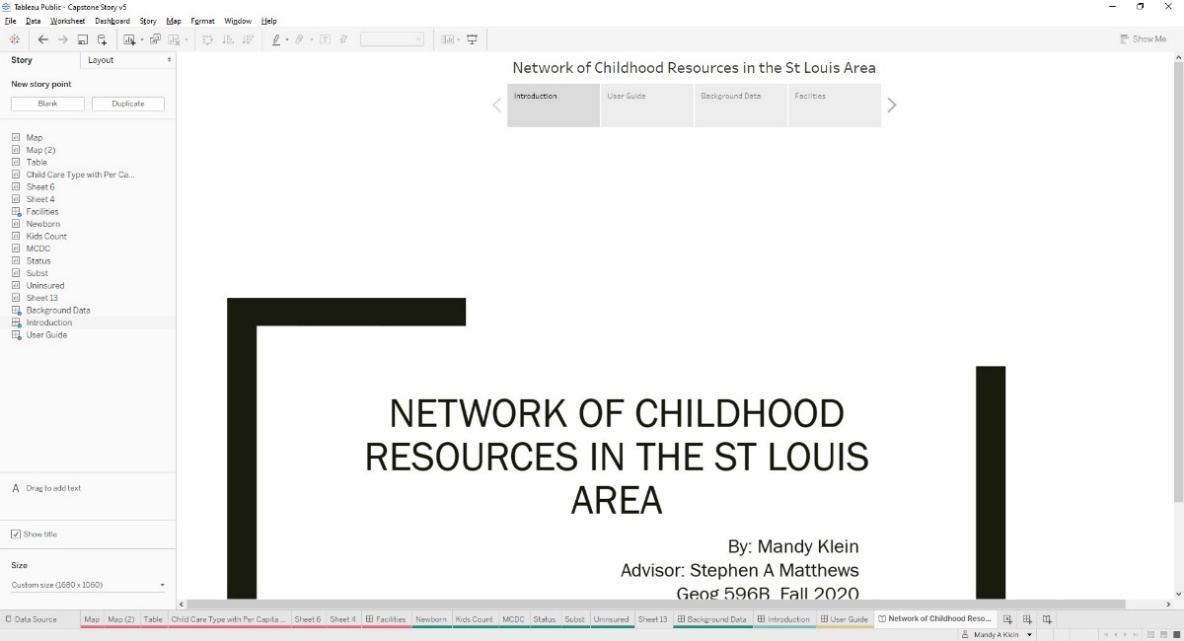


Figure . Network of Childhood Resources in the St Louis Area story within Tableau Public.

## Publish Story

Once the story was set up, it was published to the Tableau Public website so I could share the link with others and get feedback on issues, edits needed, and questions. There were several rounds of edits and adjustments based on feedback from friends, family, colleagues, and my project advisor until I reached a stage where the project was displaying the information in a meaningful and clear manner for users to understand and manipulate the data.

Published link: <https://public.tableau.com/profile/mandy.a.klein#!/vizhome/CapstoneStoryv4/NetworkofChildhoodResourcesintheStLouisArea>

# Future Improvements

There are four main areas of focus for future improvements to the project dashboards I have focused on for the purposes of this paper. Each individual area contains elements that would greatly advance the data contained within the dashboards with either new data or enhancements to current data, use of other tools or potential tools in Tableau, new visualizations, and the promotion of the project to potential stakeholder organizations. Any of the future improvements contained in this section would improve the user experience with the dashboard along with the information obtained.

## Data Availability, Coverage, and Linkage

Below is an outline of the potential future data enrichments, data additions, and project expansion. Current Facility Types Enrichment

* + Vaccines for Children program
  + Hospitals
  + Federally Qualified Health Centers
  + Childcare Facilities
* Head Start Data
* Non-Profits
* Family Support Division Resource Centers
* Urgent Care Facilities
* Local Public Health Agencies
* Parking lots
  + On street parking
* Expanding the project area of interest
  + Rural Health Clinics

In the future, I would start by enriching the attribution information for all of the current facility types (Hospitals, Federally Qualified Health Centers, WIC Offices, Childcare Facilities) available to provide more complete and detailed information that would be helpful to all potential users of the dashboard.

The **Vaccines for Children program’s information**, for example, was very limited: no name, no ZIP code, and no hours of operation attribution. The listing I found that allowed a search by county only provided the street address and city name. I would also plan to track down the locations that could not be located either by geocoding or a manual search of the address to make more of the facilities available on the dashboard. In subsequent searches, I located a listing of Vaccine’s for Children locations with names as well as ZIP codes. With that information there is the ability to also get phone numbers, county of the facility, and the Provider Pin.

For the **Hospitals** dataset, I would like to include information on any hours or restrictions on hours. Then I would incorporate attribution on capacity if available. I would also gather and incorporate information on services, especially information on pediatric services as well as call out information about NICU (Neonatal Intensive Care Unit) facilities and the level of the NICU. Within the data I would distinguish between children’s hospitals within the project area of information and hospitals that also treat pediatric cases. This information is located on many different websites including each hospital’s website.

The data lacks hours of operation for **Federally Qualified Health Centers** and that would be key information for a person looking for medical care. Another enhancement would be information on services, specifically if they are equipped to provide pediatric care, provided at each location would be very helpful to users of the site, capacity, and if medications can be filled onsite so parents don’t have to travel to another location to fill a prescription.

Additional attribution that would be extremely useful in the **childcare facilities** visualization would be price range. This is typically not published, but it would be very useful for narrowing down searches. Also, working to get more of the records to geocode from the MDHSS data by investigating the address information and fixing the issues so the record(s) will geocode successfully. This could be due to spelling errors, formatting, or incomplete address information. To remedy this, the data could be tracked down, either through internet research or through other government data sources. Another future item would be to investigate other ways of representing the Hours of Operation (& potentially the Age Ranges data) within the MS Excel spreadsheet. A potential method would be to create a field for each 30-minute increment of a 24-hour day (for example, 0900 to 0930) and represent the data by: if the facility is open during that half hour mark enter a 1, if it is not open it would get a value of zero. This methodology may allow for a more inclusive visualization within the Tableau software.

In addition to enriching the childcare facilities data already on the dashboard, I would like to provide **Head Start** location and attribution information. According to the US Department of Health & Human Services’ Head Start | Early Childhood Learning & Knowledge Center page, there are more than 70 Head Start and Early Head Start locations within the St Louis Metropolitan Area. From the listing on their site, I can obtain the facility name, address information, phone number, and who to register through for a location and their phone number. I would want to enrich the data to include capacity at each of the locations, their hours of operation, if there are age ranges or restrictions at any sites, additional services provided at the location, and information on the type of Head Start each location is (Head Start, Early Head Start, Head Start and Early Head Start, Migrant and Seasonal Head Start, and American Indian and Alaska Native). This information will assist users of the dashboard to locate their nearest Head Start facility as well as information on how to contact the entity in charge of that program to learn more information. The challenge with getting all of this information is that it is spread across several webpages and would take time to gather all of the data, clean, and format for Tableau.

In the future I would also like to add in locations of **non-profit organizations** like Nurses for Newborns, Crisis Nursery locations, United Way, food banks/pantries, and shelter locations. All of these locations can be of assistance to parents in various situations they may encounter from being homeless to needing resources to care for a newborn to a temporary crisis within a family. By providing all of this information on a map with Name, Address, Phone Number, Hours of Operation, Capacity, and Available Services, it will enable parents to more easily locate a resource.

Data is available from the State of Missouri Data Portal for **Family Support Division Resource Centers** to be downloaded as a CSV file or several other formats. These locations can provide an array of resources to families who need support and help navigating the process of getting help. The information contains the facility name, address information for both mailing and physical, hours of operation (all currently listed as closed though some are reportedly taking appointments), and phone numbers for a 24-hour information line and a family support division information center. This information would then be enriched with information on services provided at each location. This enriched information does not appear to be available online at either the Data Portal or the Family Support Division site.

Additionally, I would add the location of **urgent care facilities** around the project area as they can provide medical services when doctor’s offices are closed but the emergency room is not needed. This information would be very helpful to parents trying to find additional resources or when figuring out to who contact for help. I would want to include the Facility Name, Address, Phone Number, Hours of Operation, services offered, if they equipped to handle pediatric care, and whether or not prescriptions can be filled onsite. The challenge to gathering this data would be the disparate nature of the data. This data would be spread out over many corporate and, in the case of hospital associates urgent care, hospital network websites and not all of the desired attribution would be publicly available for download or copying. The information that can be found will need a significant amount of cleaning and processing to be incorporated into the facilities dataset.

To provide more information and resources, I would also include the locations of **Local Public Health Agencies** within the project area of interest. This information is already available from the Missouri Spatial Data Information Service Open Data Site for download as a GIS file. The file already contains location information, facility name, address information, and county information. I would add additional information on phone numbers, hours of operation, services provided, and any information specific to pediatric services. While the starting information would be easily obtained, enriching that data would require going to the website for each location and gathering the data manually. This would benefit a user of the dashboard to have the contact and location information for the local public health agency if they have questions or need to go there for a service provided, such as vaccines.

Another useful dataset to potentially be included would be **parking lot** (paid or free) locations from within the current project area of interest but also a potentially expanded project area. This data would enable users of the dashboard to see information about parking lots in the vicinity of a childcare facility or another facility type on the dashboard. Potentially using the Parkopedia website as a data source, address information, phone information, days open, hours of availability (if not 24hrs), cost, payment options, total lot parking spots, information on restrictions and handicapped parking attribute information are available (*Parkopedia Main Page*, 2020). This dataset does not appear to have a download function or an easy way to copy the data and would more than likely require a lot of manual entry to create. The inclusion of this potential dataset would help users to better prepare for a trip with as much information on parking as the data supports.

Another potential data need would be the availability of **on street parking**, whether paid or free. The Parkopedia website does provide limited locations and information regarding on street parking. This information would be extremely challenging to keep up to date and maintain as cars are potentially coming and going throughout the day. If a dataset containing parking meters within the project area of interest could be located, this could provide potential information to distinguish on street parking from residential parking along a street.

Another opportunity for improvement is to **expand the area of interest** from St Louis County and City to the entire metropolitan area of St Louis. This would include the Metro East areas in Illinois. This would expand the project area significantly, from 2 counties to approximately 15 counties. This would create a more complete picture of the entire St Louis Metro area for both the Facilities and Background Data dashboards. For the Facilities, it would provide people living in Illinois with resources in their state instead of always having to travel into Missouri for resources or assistance. For Background Data, expanding the area of interest would provide non-profits with a better understanding of the entire metropolitan area rather than a subsection of the area. This will help to see how resources can be redirected to better serve all of the communities. For the local, county, and state governments, there will be expanded areas to see how the variables are performing over time rather than two counties only. One of the challenges may be how Illinois presents their data on the variables and facilities to be featured on the map and visualization. It is unknown if Illinois makes all of the information Missouri has available, open to the public. Then after obtaining the data, cleaning, formatting, attributing, and integrating the data may present other challenges depending on how the data is received.

Once the project area is expanded, **Rural Health Clinics** could be downloaded from the Missouri Spatial Data Information Service Open Data Site. Contained in the file when it is downloaded is spatial information, facility name, address information, facility phone number, a facility ID, and the county. I would enrich this attribution information to include hours of operation, capacity, services provided, specifically services for pediatric patients, and information if medications can be filled on site. This information would be in addition to hospital, federally qualified health centers I have already included in the current project but for the potentially expanded project area, the Rural Health Clinics dataset is also available as another source of medical treatment, vaccines, and other services. The enrichment of the data would be challenging due to the information not being contained on one site but likely multiple sites for each of the individual locations within the dataset.

## Search Tools/Filters

* Multiple geospatial data layers
* Hours of Operations and Age Range Sliders
* Page function

With future improvements to Tableau’s mapping capabilities, the public transit system could be visualized in the Facilities map visualization, giving a user a great deal more context to their query and visual results (*St Louis Metro Developer Resources*, 2020). Being able to see that a childcare facility is within a block or two of a bus stops might be valuable information to have. Currently, Tableau is limited to displaying one geometry type at a time but a recent press release provided a preview into **multiple geospatial data layers** being visualized at once (*Tableau 2020.4 Release Notes*, 2020). This limitation currently also applies to custom geospatial files, such as the facility points. While trying to display several GIS files of the same geometry type, the software cannot accommodate multiple files and will not display properly.

There are several improvements to the dashboard capabilities I would like to see within Tableau Public, external tools to augment Tableau Public (user created tools available to the public), or programmed custom solutions. The **Hours of Operations and Age Range sliders** on the current dashboard allow the user only to query exact matches to a set of times or ages. Allowing a user to define an age range or range of hours of operation will be far more useful. This flexibility will allow users to find more facilities that might meet their needs, instead of only supplying matches to their exact criteria.

The **Pages functionality** in Tableau Public would make it possible to have fewer visualizations on the background data dashboard at one time. There would be a filter a user could use to ‘page’ through the variables along with a drop-down menu to select an individual variable to be viewed. This would create a more integrated visualization experience for the dashboard.

## Data Visualization

* Dynamic zoom
* Toggle between basemaps
* Census Block (or Tract) level indicators
* More recent or complete indicators

In the Facilities dashboard, the map cannot **dynamically zoom** based on the user spatial filters the user selects. When a user selects a county, city or ZIP code, the map stays at the current zoom. For someone unfamiliar the current project area, this could be very frustrating. If the project area were to be expanded in the future, the map not zooming to geographic filters or facility selection will create a larger area for a user to scan. An improvement I would like to have in the future is for the map to automatically zoom to an area based on the filter(s) selected on the dashboard. This will help users to see where the facilities are easier but also to narrow the viewable map on the dashboard. This capability is present within Tableau when the dashboard is published on its own, outside of a story but it doesn’t work once the dashboard is incorporated into a story. It is unclear what the solution to keep this functionality working in a story environment and if it will be fixed in a future release of Tableau Public. While Tableau has recently made several improvements in their spatial capabilities over when I began using the program (the ability to view more than point files without a well-known hack, heat maps, integrated data layers native to Tableau), it is unclear if this zoom capability in a story is something the company may have in development for future releases.

Currently, a user of the Facilities dashboard can only view the Streets basemap. An improvement would be to create a way to toggle between the Streets basemap and the Satellite basemap. Both are currently available within the Tableau Public software and there is an option to bring in a basemap from another source if one were to be removed in future versions of the software. There is no way within Tableau Public to **toggle between basemaps** so this would have to be developed by the company. The benefit for a user to be able to toggle between the Streets basemap and the Satellite basemap would be once they find a facility on the Streets basemap they could switch over to the Satellite basemap to see if there is a parking lot or garage in close proximity or if the facility is within a larger complex, within a strip mall setting, set along a busy road, or easy to enter or exit via car. A temporary solution for this would be to duplicate the Facilities dashboard and change the basemap from Streets to Imagery. It is unknown if the company has this in a backlog of features for a future release.

I would like to create another dashboard similar to the Background Data one that would contain **Census Block (or Tract) level data** concerned with poverty and most specifically children in poverty over the period of 2009 to 2019 when the data is available. This information can help non-profits better direct resources into those communities or outreach efforts. Additionally, this data along with the facility information may potentially show gaps in a service provider network in some communities that need those resources the most. This information could also be used for local governments to better direct their resources, public education plans, and efforts to bring different programs and facilities into an area. The challenge would be to gather the data and then format it in a way Tableau can understand to effectively visualize and filter data to be useful for those interacting with the dashboard.

For the variables already included in the Background Data dashboard, I would like to find datasets with more recent data and data that is available for a larger area. **More recent and complete data** will help those using this dashboard to direct resources have the most up to date information while making the decisions. Having information from 2009 to 2019 will also provide a decade worth of data on a given variable to see if it is improving, getting worse, or maintaining. The challenge to finding newer data is that a lot of the time newer data isn’t publicly available and would require contacting the government organizations (IE County Health Departments) or non-profits that warehouse the data to gain access to the data.

## Promotion of use

Another future plan would be to contact non-profit organizations and government entities that would be interested in this type of dashboard to gather their feedback on what would be most useful to their clients. There are a number of early childhood and children in general non-profits within the St Louis Metropolitan area I would reach out to for feedback on the current dashboard as well as suggestions on things to include, change, remove, and potential sources of data. I would reach out to different organizations individually and work with them on their needs understanding each entity might have different needs or wants for a visualization like this and one size may not work for all of the them. The reward of reaching out to the different entities would be they could potentially adopt and use this dashboard during their outreach efforts or even link to it via their websites. The challenge would be to gather the potential requirements from the different organizations and implement them on one or more dashboards for their use. This could result in a lot of time and effort on my part to get to a desired end state for the organization.

# Conclusion

The purpose of this project is to create a Tableau dashboard for parents and non-profits to help find different types of resources for families of young children in St Louis County and City, Missouri. There is a lot of information and data available in the public domain from government sources as well as non-profit organizations to put into a project like this. A lot of the data is inconsistently structured, missing key attribution, and required extensive work to clean and format for use. This presented the major challenge in setting up and visualizing the data. If the data would come from the original sources with rigid structuring and attribution, it would be easier to use the data. Tableau overcame some of the structure and formatting issues with basic data preparation tools available off the shelf. The software proved to be very useful in cleaning several issues in the data and visualizing the required components in an easily understood display. In the end, the final product is a practical resource a user can access for information on childhood resources.

This project not only proved the concept of an interactive dashboard focused on childhood resources but it is also accessible to everyone who would need or want to use it today. Multiple data sources were gathered and combined to create a data rich dashboard allowing parents, caregivers, non-profits, and government organizations to access useable information in a timely, consumable, and user-friendly format. The MDHSS website, where the childcare and Vaccines for Children data were obtained from, provides only a listing with limited search capabilities. Parents can use the Facilities dashboard to locate childcare and/or vaccine options in both a table and a map with many filtering options based on their criteria and needs. Additional, GIS based data for other useful facilities was also incorporated into the facilities dashboard to assist users. The result is a self-service dashboard that is both free and available to the public.

There are several limitations to the data and software used for this project. One of the limitations of the project is the data used. The Child Care Search and the Vaccines for Children data had to be geocoded in order to display on a map. There were a number of facilities in both datasets the geocoding service could not match due to address format errors. The facilities that failed to geocode are missing from the dashboard and are a known limitation of the data. Another limitation was the currency of the data in general. The MDHSS data didn’t provide any currency statement for the childcare or Vaccines for Children datasets. The data used for several of the visualizations on the Background Data dashboard are several years old. The lack of current data, limits the value of the information as well as potential insights gained regarding trends. Tableau also has some limitations in how it visualizes and filters data. There are limitations on Tableau’s ability to visualize geospatial data. At this time, it can only visualize one geometry type at a time and only one spatial file at a time. To visualize points data, a user must merge them all into one file prior to using Tableau. Being able to visualize lines in addition to the facility point data would have allowed the use of public transit spatial files. Another limitation of the software was how some of the filters worked. For example, the Open and Close times, a user can only filter for those facilities with those exact times instead of those that fall within that time range. This would allow for more facilities to remain in the results.

Over the course of this project, I learned there is a lot of valuable public information available to be used for this kind of project. The childcare and Vaccines for Children datasets required a lot of cleaning prior to visualization. I learned to utilize different software packages (Notepad ++) to clean the data in a more automated and consistent process. This produced more reliable results to work with and visualize. While researching geocoding services, I learned about the US Census Bureau’s geocoding service that is free to use and provided reasonably accurate results in a short period of time. The service took some trial and error to have the datasets set up to process correctly. This service was very valuable to spatially visualize data in this project. I learned a great deal about organizations, government and non-profit, that are trying to help children and families in poverty get the resources they need in order to survive, grow, and be successful. A lot of the organizations provide data on indicators to see if things are improving or not. I learned some of the data out there is limited in the time frame as well as the type of data (number, rate, percent).

I learned additional features, tools, and methods while using Tableau. Having used Tableau professionally for several years prepared me for this project, some of the challenges presented learning opportunities. The data itself required me to use the data preparation tools to normalize and clean the data sources. In the past, I have used some of the data preparation tools (changing field types, grouping) but this project required using several field calculations to clean percent data into a format Tableau was able to display properly, splitting several fields to create additional filter data, and creating many-to-many relationships with their current user interface. Some of the filters required me to customize parts in ways that required research on how best to structure the data and then set up the filter for use. Also, the version of the software I use for work is a number of years older and has a few of the capabilities removed (various basemaps, Background Map Layers, and newer features such as visualizations within hover boxes). I was unable to use the Background Map Layers and visualizations within hover boxes in the end product, but I did experiment with the capabilities while setting up different test visualizations. The knowledge of this much newer version of Tableau can easily be applied to some of the projects I am tasked with in my current job.

# Bibliography

*Census Bureau Geocoder Tool*. (2020, November 28). [Tool]. United States Census Bureau. https://geocoding.geo.census.gov/geocoder/locations/addressbatch?form

*Child Care Facility Search*. (2020, November 22). [Facility Search]. Missouri Department of Health & Senior Services. https://healthapps.dhss.mo.gov/childcaresearch/

*Child Wellbeing Indicators & Data | KIDS COUNT Data Center*. (2020, November 20). https://datacenter.kidscount.org/data/

*Coming Soon*. (2020, December 6). Tableau. https://www.tableau.com/products/coming-soon

*ESRI ArcGIS Desktop*. (2020, November 30). https://desktop.arcgis.com/

*Find Parking Lots, Parking Garages, Parking Meters, Private Garages—Reserve Parking*. (2020, December 6). https://en.parkopedia.com/

*Medically Underserved Areas/Populations Search Page*. (2020, November 22). Health Resources & Services Administration. https://data.hrsa.gov/tools/shortage-area/mua-find

*Microsoft Excel Online, Spreadsheet Software*. (2020, December 2). Microsoft Excel. https://www.microsoft.com/en-us/microsoft-365/excel

*Missouri Census Data Center*. (2020, November 18). https://mcdc.missouri.edu/

*Missouri Children’s Division Annual Report*. (2020, November 15). Missouri Department of Social Services. https://dss.mo.gov/re/csar.htm

*Missouri Spatial Data Information Service Open Data Site from the University of Missouri Geography Department*. (2020, November 18). https://msdis.missouri.edu/

*Notepad++*. (2020, December 2). Notepad ++. https://notepad-plus-plus.org/

*Nurses for Newborns Homepage*. (2020, November 28). Nurses For Newborns. https://www.nursesfornewborns.org/

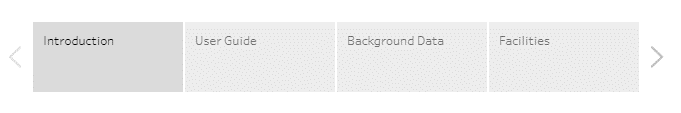
*St Louis Metro Developer Resources*. (2020, November 29). Metrostlouis.Org Site. https://www.metrostlouis.org/developer-resources/

*Tableau Public*. (2020, November 30). Tableau Public. https://public.tableau.com/en-us/s/

*Vaccines for Children Providers*. (2020, November 17). [Search page]. Missouri Department of Health & Senior Services. https://ogi.oa.mo.gov/DHSS/VFC/index.html

# Appendix A - Tableau User Guide

Use the caption boxes, located at the top of the project to navigate between topics/dashboards.



* **Introduction** – a brief introduction to the background of the project, the motivation for the project, the audience for the project, and the problem to be addressed by project
* **User Guide** – a basic user guide to help a user of the project to better navigate the various dashboards and project dat.
* **Background Data** – a dashboard with graphic representations of background data providing context for the rest of the project.
* **Facilities** - A dashboard that showcases an interactive map, allowing a user to filter geospatial representations of the data by attribute information, such as childcare sites, hospitals, federally qualified health centers, WIC office location, and locations participating in the vaccine providers for children program.

The lower right of the project contains tools that apply to the entire project.



 = Undo – undo action(s) performed on the project.

 = Redo – redo action(s) performed on the project.

 = Reset – reset the project to the initial default.

 = Share – a pop up appears with embed code and a link of either the Current View or Original View to share.

 = Download – a pop up appears with options of what to download.

 = Full Screen – project will become full screen.

## Background Data Dashboard

Click the “Background Data” dashboard in the topic panel to view the background indicators for the project. Figure 45 is the default view. Each of the six graphs on the dashboard examine different indicators related to select childhood concerns. Hover over individual graphs to see more information on the topic (Figure 46).

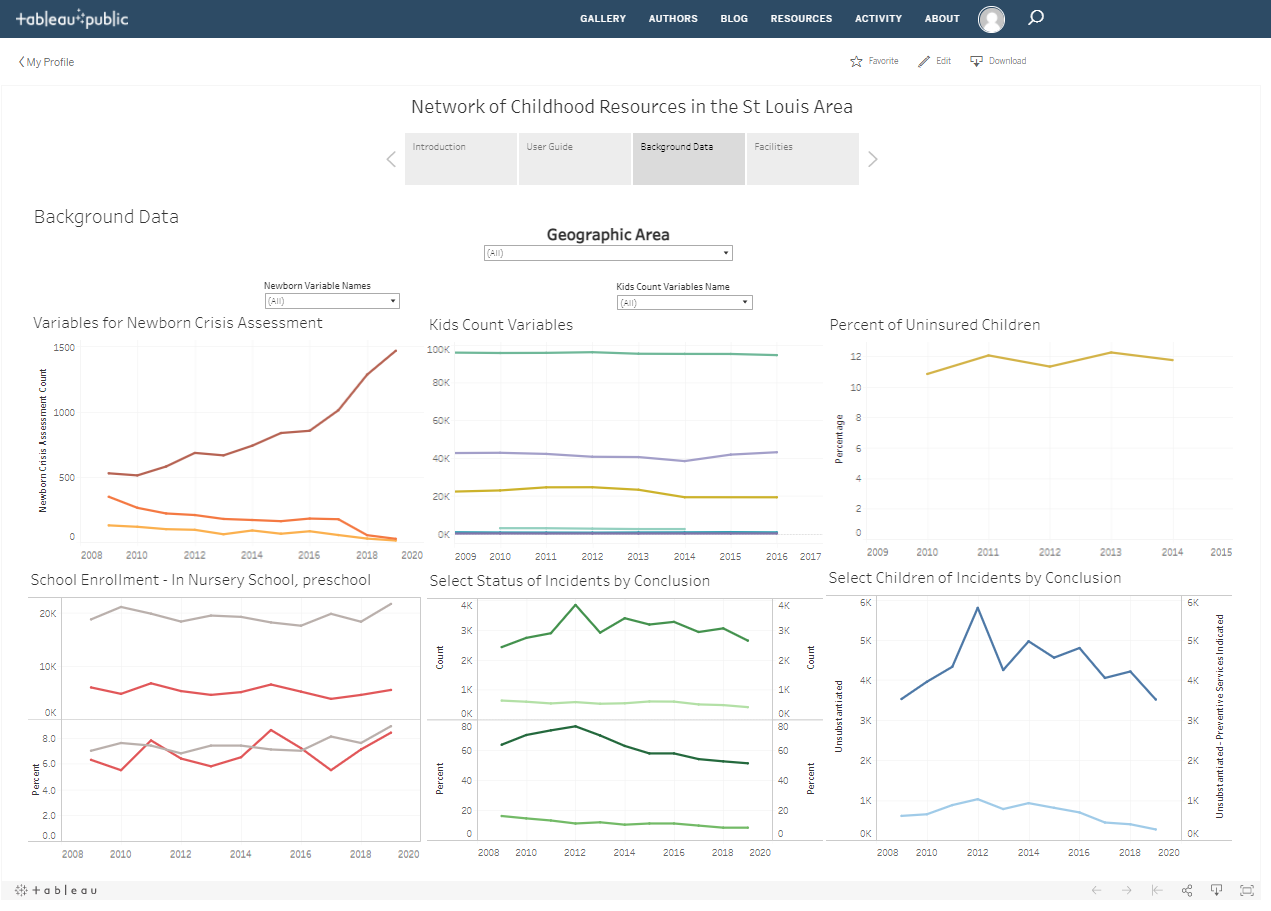


Figure . Default view of the Background Data dashboard upon initial load.

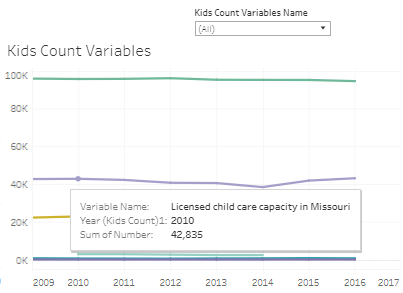


Figure . Pop up window with additional information on the variable.

All of the indicator graphs are filtered based on the Geographic Area filter defined at the top center of the dashboard (Figure 47).

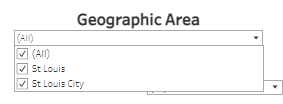


Figure . Geographic Area filter.

Above the Variables for Newborn Crisis Assessment graph is a filter for only that graph to view the three variables individually or a combination of the three (Figure 48).

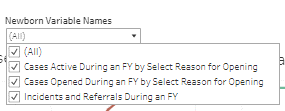


Figure . Newborn Variable filter.

Above the Kids Count Variables graph is a filter based on a combination of one or more of six variables (Figure 49).

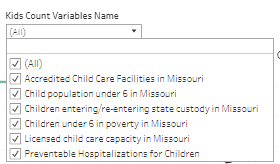


Figure . Kids Count Variable filter.

## Facilities Dashboard

Click on “Facilities” dashboard in the topic panel to view the information for the childcare types, hospitals, WIC (Women, Infants, and Children) Offices, federally qualified health centers, and locations participating in the vaccine providers for children program (Figure 50).

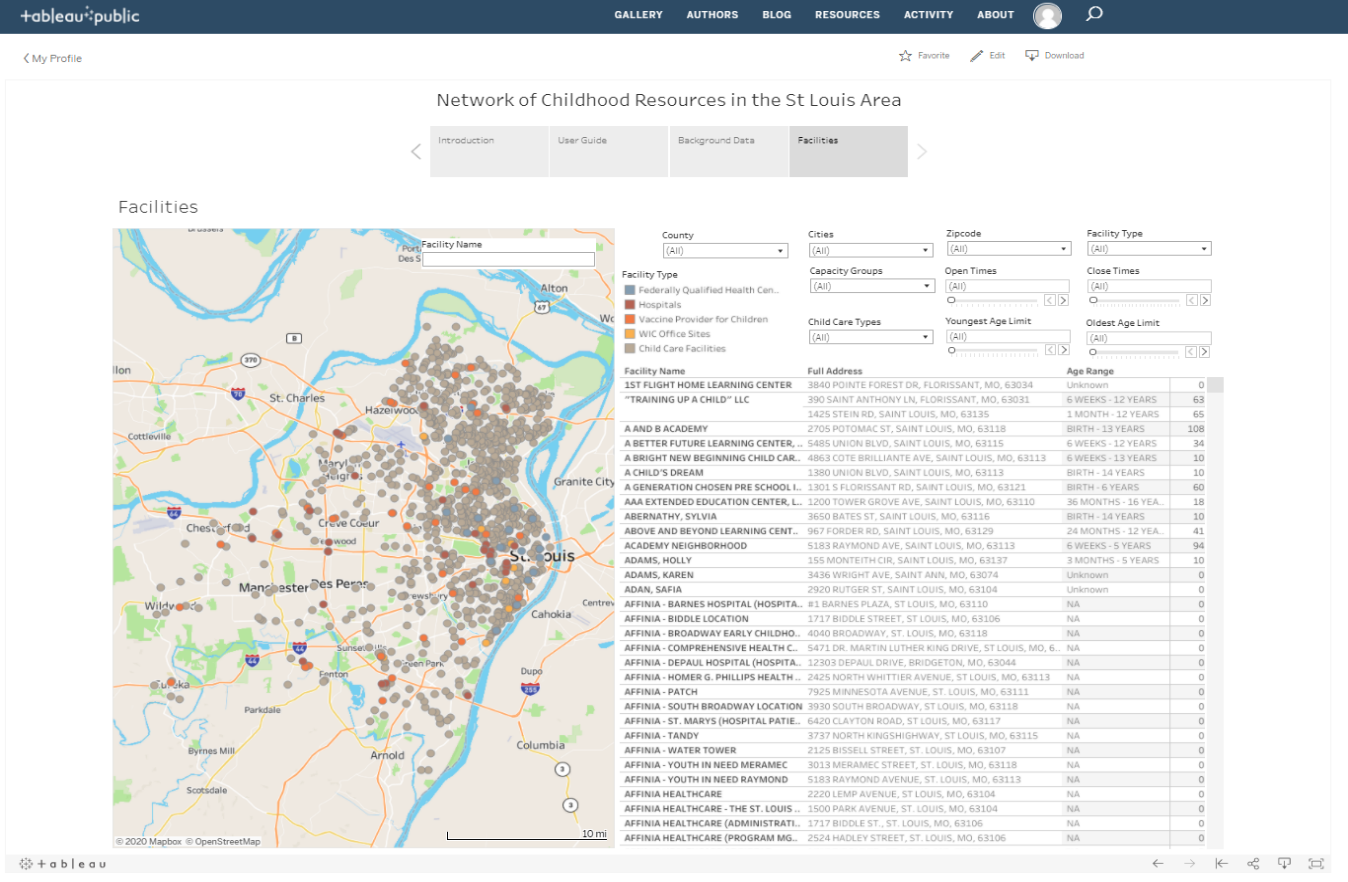


Figure . Default view of the Facilities Dashboard upon initial load.

Use the drop-down menus and slider bars at the top of the dashboard (Figure 51) to filter data based on attribution. **\***

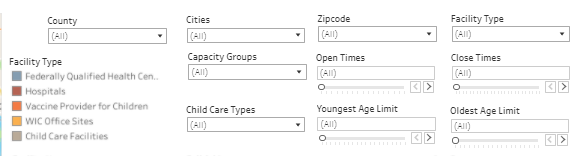


Figure . Overview of the filters on the “Facilities” dashboard.

**\***note: de-selecting all in any of the categories removes all points from the map and data from the table (Figure 52). To visualize points again, select at least one item from each category drop down.

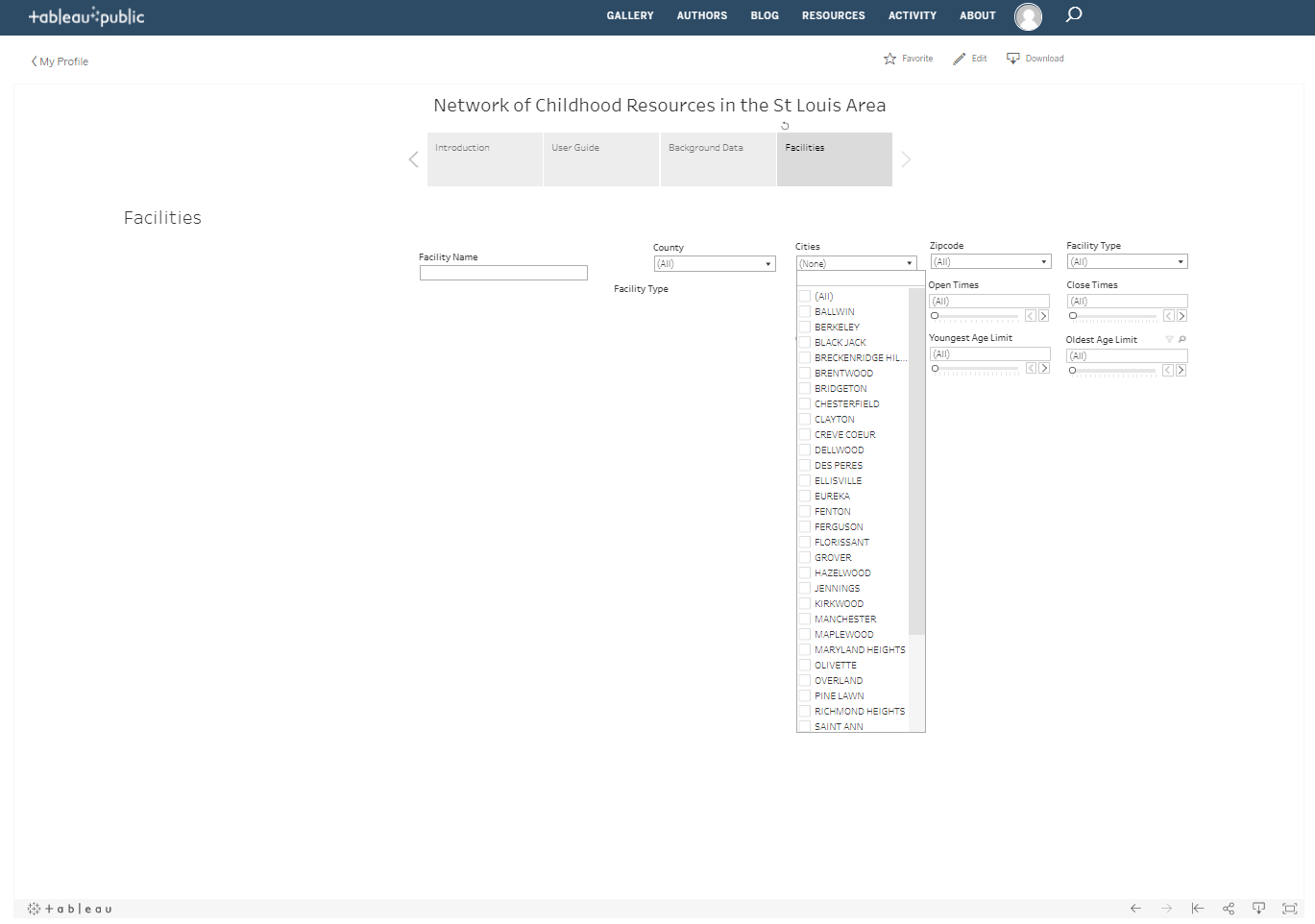


Figure . De-Selecting all in any of the categories results in the map and table disappearing.

### Sliders:

Sliders are used to define ranges as additional filtering mechanisms. Define the ranges by moving the slider bars or by clicking the left or right arrows next to them.

A limitation of this dashboard is that it only filters on exact ranges. For example, enter an Open Times of 6AM and a Close Times of 6PM (Figure 53), it will only find facilities with those times, not ones that open at 7AM and close at 5PM as an example. To find the broadest amount of facilities matching criteria, it would be best to select either Open OR Close Times and leave the other slider set to (All).



Figure . Sliders for time range.

This limitation is also true for the age range sliders (Figure 54). For example, enter 6 WEEKs for the Youngest Age Limit as well as 9 Years for Oldest Age Limit, only results showing facilities with those exact criteria.



Figure . Sliders for age range.

At the upper right of the map is a free text search for facility name (Figure 55). Enter part of a name for the tool to search for results matching the entry.



Figure . Free text search for a facility name.

To view information about a site on the map, hover the mouse over the point (Figure 56).

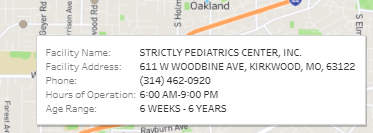


Figure . Example of a hover box containing information on a childcare facility.

On upper left of the map are a few helpful tools, hover over the map for them to appear (Figure 57).



Figure . Overview of the tools available on the map.

### Tools and Functionality:

 = Search Map Tool – This tool allows searches at the ZIP code or city level. There is no functionality to search an address.

 = Zoom In/Zoom Out – This tool allows zooming in and out in the map by clicking on the respective button (functionality can also be achieved by using the scroll bar on a mouse).

 = Zoom Home – Zooms the map to the default view from when the page originally loaded.

 = Clicking this button reveals additional functionality for the map described below:

 = Zoom Area – This tool allows a click and drag to create a box over an area to zoom in.

 = Pan – This tool allows panning around the map.

 = Rectangle – This tool allows click and drag a box over an area and select/filter facilities on the map to see them in the table.

 = Radial – This tool allows click at a point and drag a radial (circle) out a desired distance to select/filter facilities (Figure 58).

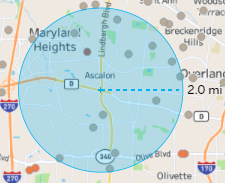


Figure . Example of the radial selection tool.

 = Lasso – This tool allows click and draw a freehand shape to select/filter facilities (Figure 59).

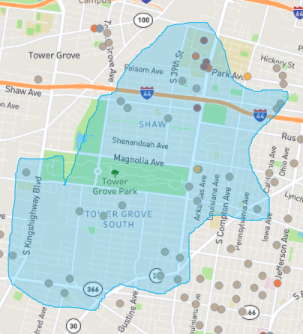


Figure . Example of the lasso freehand shape tool.

Use the Facility Type drop-down filter (Figure 60) to choose the type(s) of facilities to display on the map and in the table.

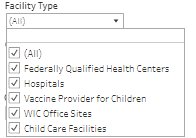


Figure . Facility Type filter with all options visible.

**Federally Qualified Health Centers** - This dataset published by the Missouri Department of Health and Senior Services from data provided by the Missouri Primary Care Association. An FQHC provides services to medically underserved populations**\***. An FQHC may be in either an urban or rural area. October 2020 update.

**\*** for the purposes of this project, medically underserved populations are defined as are areas or populations designated by HRSA (Health Resources & Services Administration) as having too few primary care providers, high infant mortality, high poverty or a high elderly population. Health Professional Shortage Areas (HPSAs) are designated by HRSA as having shortages of primary medical care, dental or mental health providers and may be geographic (a county or service area), population (e.g. low income or Medicaid eligible) or facilities (e.g. federally qualified health center or other state or federal prisons)(*MUA Find Page*, 2020).

**Hospitals** - This dataset published by Missouri Department of Health and Senior Services. Missouri hospitals, including VA hospitals, rehabilitation centers and psychiatric hospitals October 2020 update.

**Vaccine Provider for Children** – This dataset published by the Missouri Department of Health and Senior Service – Vaccines for Children (VFC) Program and are locations where children fitting certain criteria can get vaccinated for free (thought there maybe still be a fee for the office visit and the shot).

**WIC Office Sites** - This dataset published by the Missouri Department of Health and Senior Services. Locations where people can sign up and receive services from the Women, Infant, and Children (WIC) program. Satellite sites are associated with offices and are likely to be open for a short period of time only one or two days a week. WIC provides supplemental nutrition services to pregnant women, new mothers, infants and children up to their 5th birthday. June 2020 update.

**Child Care Facilities** - This dataset published from the Missouri Department of Health and Senior Services – Child Care Search Site. Details of the types of childcare facilities used in this dashboard are detailed below in the Child Care Types filter section.

Use the Child Care Type drop-down filter (Figure 61) to choose the type(s) of childcare facilities to display on the map and in the table.

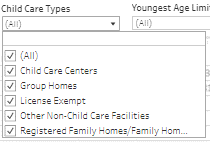


Figure . Child Care Types filter with all options visible.

.

**Child Care Centers** – A program licensed by DHSS/SCCR (Department of Health and Senior Services/ Section for Child Care Regulation) where care is given for any number of children dependent on the director’s qualifications, available staffing, amount of usable indoor floor space, amount of outdoor play space and materials and equipment. A childcare center may be in a location other than the provider’s permanent residence or a separate from the provider’s living quarters. This includes before and after school care programs located in a school building. This dataset is from the Missouri Department of Health and Senior Services – Child Care Search Site.

**Group Home** - A program licensed by DHSS/SCCR (Department of Health and Senior Services/ Section for Child Care Regulation) where care is given by a person licensed as a group childcare home provider for 11-20 children. A group childcare home may be a location other than the provider’s permanent residence or separate from the provider’s living quarters. Group childcare homes must meet the same requirements for staff/child ratios as childcare centers. This dataset is from the Missouri Department of Health and Senior Services – Child Care Search Site.

**License Exempt** - Nursery School:  A program inspected by DHSS/SCCR (Department of Health and Senior Services/ Section for Child Care Regulation) for preschool children that is operated for no more than four hours per child per day. Child Care Programs Operated by a Religious Organization: A program inspected by DHSS/SCCR (Department of Health and Senior Services/ Section for Child Care Regulation) that is under the exclusive control of a religious organization caring for children up to 17 years of age. This dataset is from the Missouri Department of Health and Senior Services – Child Care Search Site.

**Other Non-Child Care Facilities**- This selection will display only Federally Qualitied Health Centers, Hospitals, Vaccine Provider for Children Sites, and WIC Office Sites.

**Registered Family Homes/Family Home** – Family Home: A program licensed by DHSS/SCCR (Department of Health and Senior Services/ Section for Child Care Regulation) where care is given by a person licensed as a family childcare home provider. If there is one adult childcare provider, family childcare homes may be licensed for up to six children, including a maximum of three children under age two and/or for up to ten children including a maximum of two children under age two. If only four children are present, all the children may be under age two. If the provider has an assistant present, the home may be licensed for up to ten children, including a maximum of four children under age two, or for up to eight children who may all be under age two. Registered Family Home: A program registered with DSS/CD (Department of Social Services/Children’s Division) to receive subsidy funds in which a non-licensed provider provides care in their own home or the child’s home. The provider is limited to no more than six children, with a maximum of three children under the age of two. This dataset is from the Missouri Department of Health and Senior Services – Child Care Search Site.