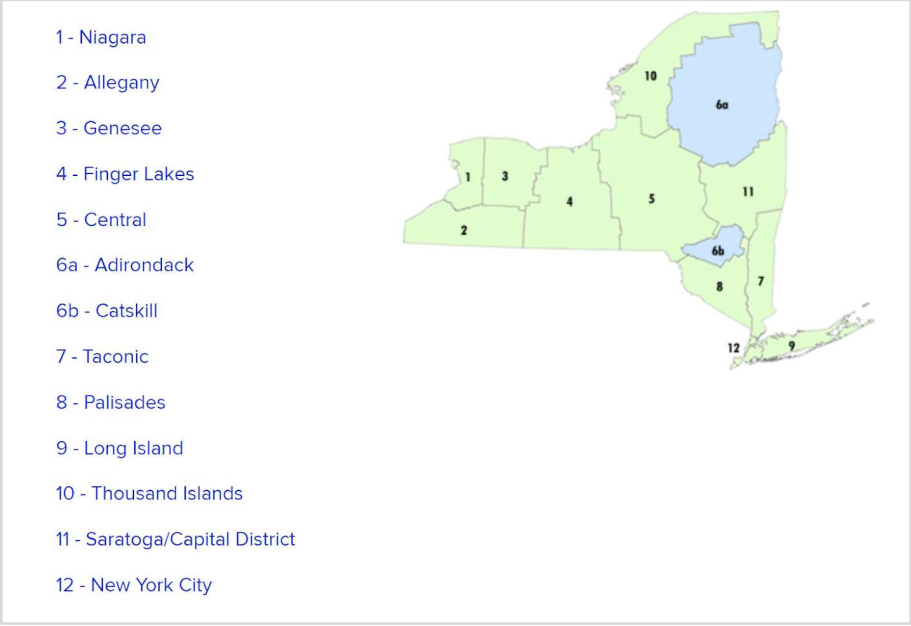
NYS State Park Annual Attendance Dataset Visualizations Using Tableau

Sita Robinson

7/13/18

Introduction

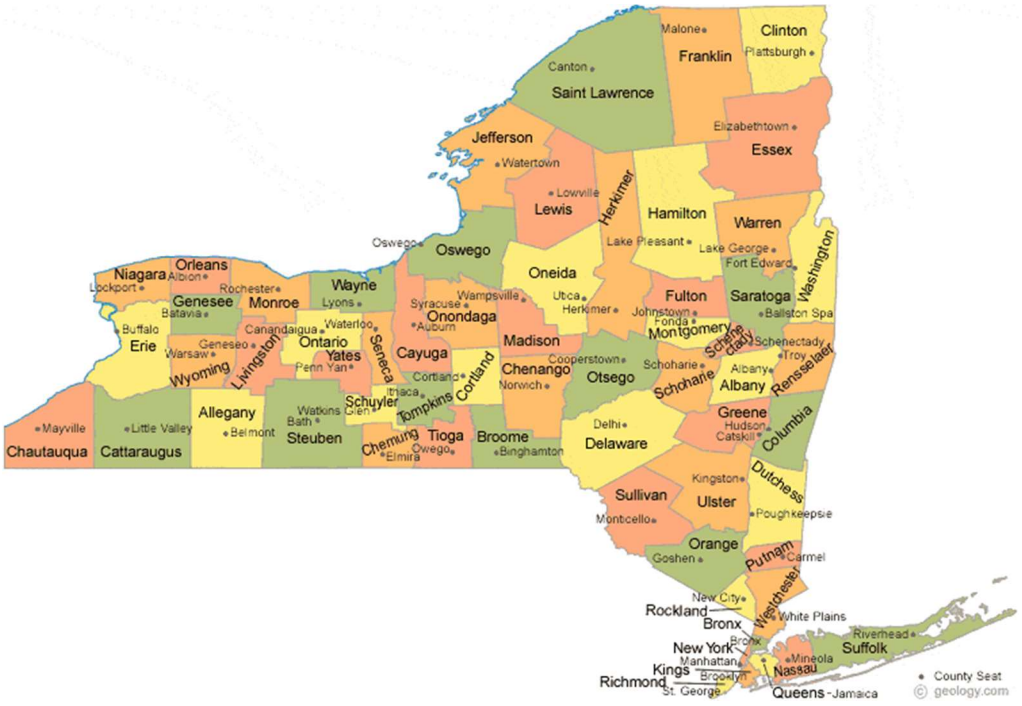
The dataset chosen for this assignment was the NYS State Park Annual Attendance Figures dataset from Kaggle. The data is provided by the State of New York and provides yearly attendance for facilities operated by the New York State Office for Parks, Recreation and Historic Preservation (OPRHP). There are 66 unique counties in this dataset but according to the New York State government there are 62 counties. This is because some of the records contain multiple counties. There are 3296 columns not including the description columns. There are five columns in this dataset which are Year, OPRHP Region, County, Facility, and Attendance respectively. Year is a numerical variable, OPRHP Region is textual, County is geographical, Facility is textual, and Attendance is numerical. The following figures exclude the missing data. The OPRHP Regions are separated into eleven sections and is shown in the image below from the New York State Parks, Recreation, and Historic Preservation (OPRHP) website:



The Adirondack and the Catskill regions (shown in blue) are not part of the OPRHP and are overseen by the Department of Environmental Conservation. The dataset shows some parts of all eleven of these unique OPRHP Regions. The image below shows the results after a UNIQUE statement in Google Sheets:

OPRHP Region
Allegany
Central
Finger Lakes
Genesee
Long Island
New York City
Niagara
Palisades
Saratoga
Taconic
Thousand Islands

The counties in New York are shown in the image below.

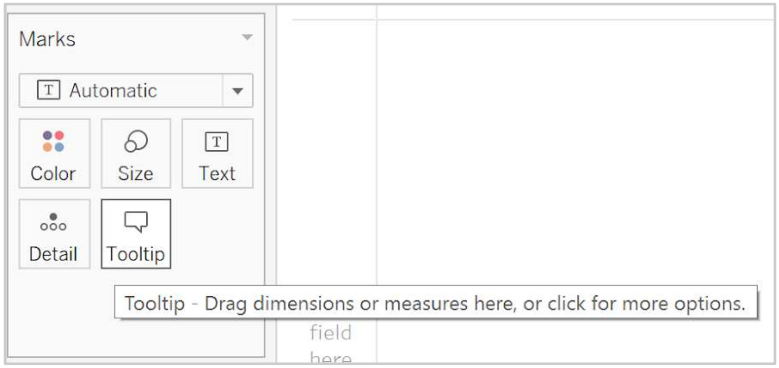


The counties in this dataset are shown here:

County
Cattaraugus
Chautauqua
Oswego
Otsego
Chenango
Broome
Madison
Onondaga
Oneida
Herkimer
Onondaga/Madison/Oneida
Broome/Delaware
Tompkins
Wayne
Seneca
Cayuga
Ontario
Steuben
Yates
Chemung
Ontario/Seneca
Schuyler
Monroe
Livingston
Genesee
Allegany/Cattaraugus/Livingston/Monroe
Orleans
Livingston/Wyoming
Wyoming
Suffolk
Nassau/Suffolk
Nassau
Queens
Richmond
Kings
New York
Bronx
Niagara
Erie
Rockland
Dutchess/Orange/Putnam/Rockland
Orange/Rockland
Ulster
Orange
Sullivan
Greene
Rensselaer
Essex
Saratoga
Columbia
Albany
Delaware
Fulton
Schoharie
Saratoga/Warren
Montgomery
Putnam
Dutchess
Westchester
Dutchess/Putnam
Columbia/Dutchess
Dutchess/Ulster
St Lawrence
Jefferson
Clinton
Lewis

As you can see some of the counties are shown together. Tableau's "Marks" section allows six different options which are Color, Size, Label, Detail, Tooltip, and Path. These can be chosen by dragging the particular

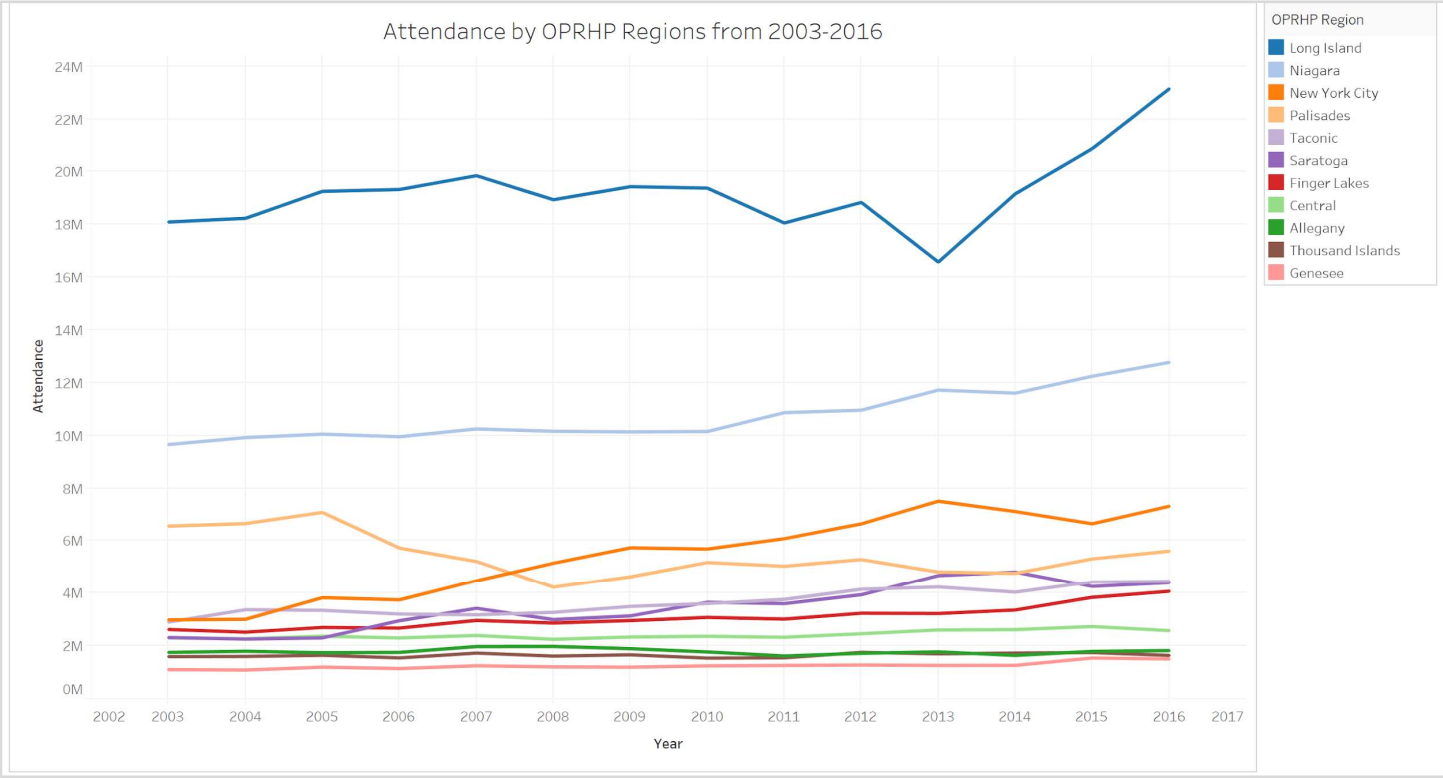
dimension or measures to the specific box. If you hover over each mark option, a description pops up explaining each of these options:



There are twenty-four different visualization options and the ones that can be used are clickable after the marks, columns, and/or rows are chosen.

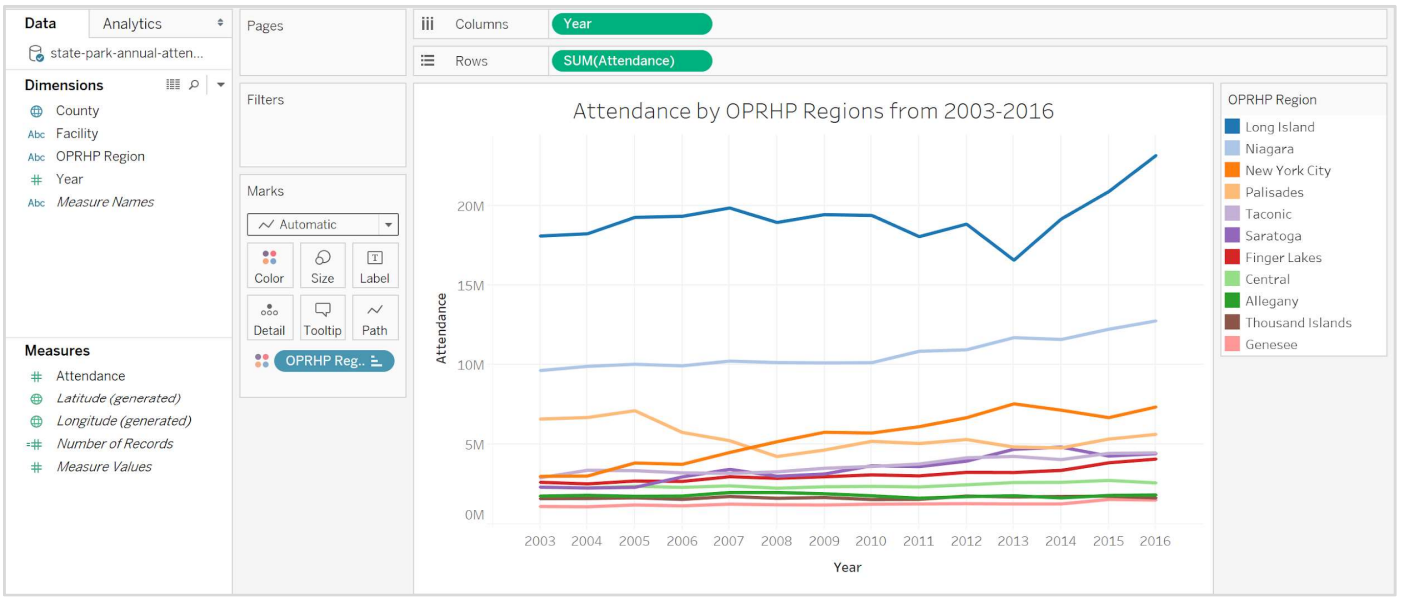
Visualizations

Figure 1: Attendance by OPRHP Regions from 2003-2016



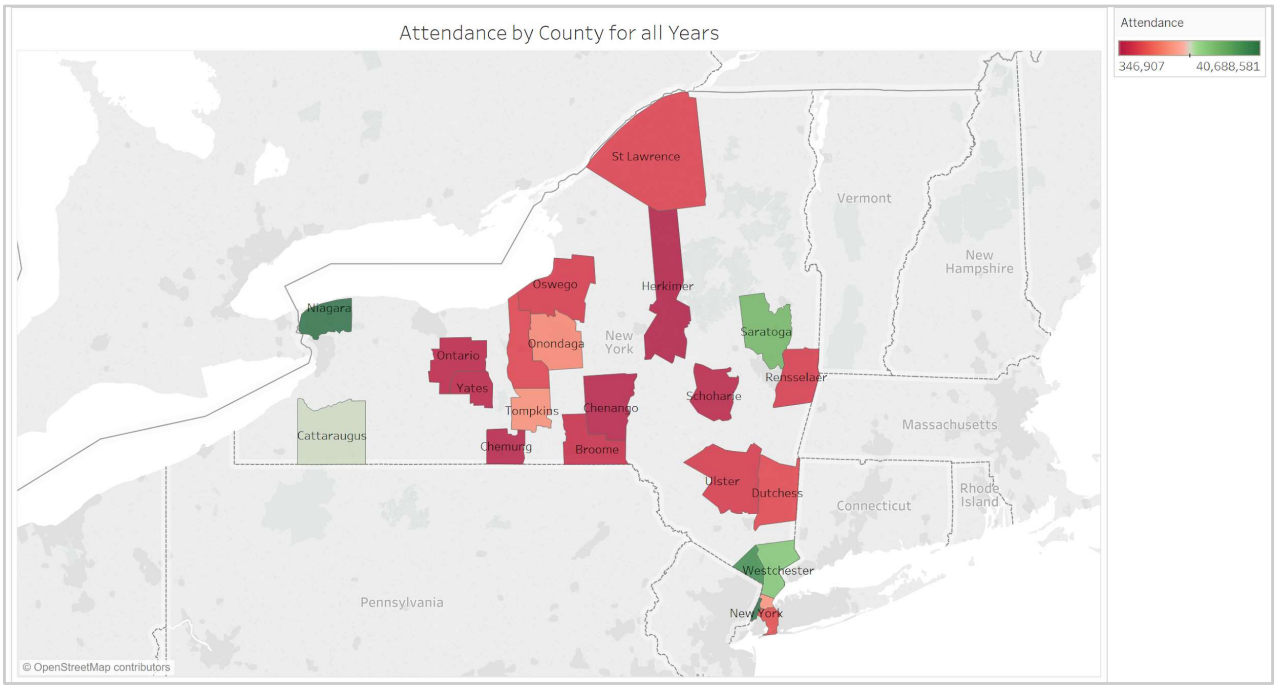
This line visualization shows the attendance for different OPRHP Regions over time from 2003 to 2016. From this we can see from this data is that the Long Island has always had the highest attendance through all years followed by Niagara. Palisades is the third highest from 2003 to 2007 and then New York City has the most attendance from 2008 to 2016. The Genesee region has had the least amount of attendance from 2003 to 2016. This is one of the more useful maps in this document since it separates attendance out by year.

The image below shows the options chosen for the Marks section and for the Rows and Column Filter.



Each line is colored by the OPRHP Region. Year is in the Columns filter and the sum of the attendance is in the Rows filter so therefore Year is along the x-axis and attendance is around the y-axis.

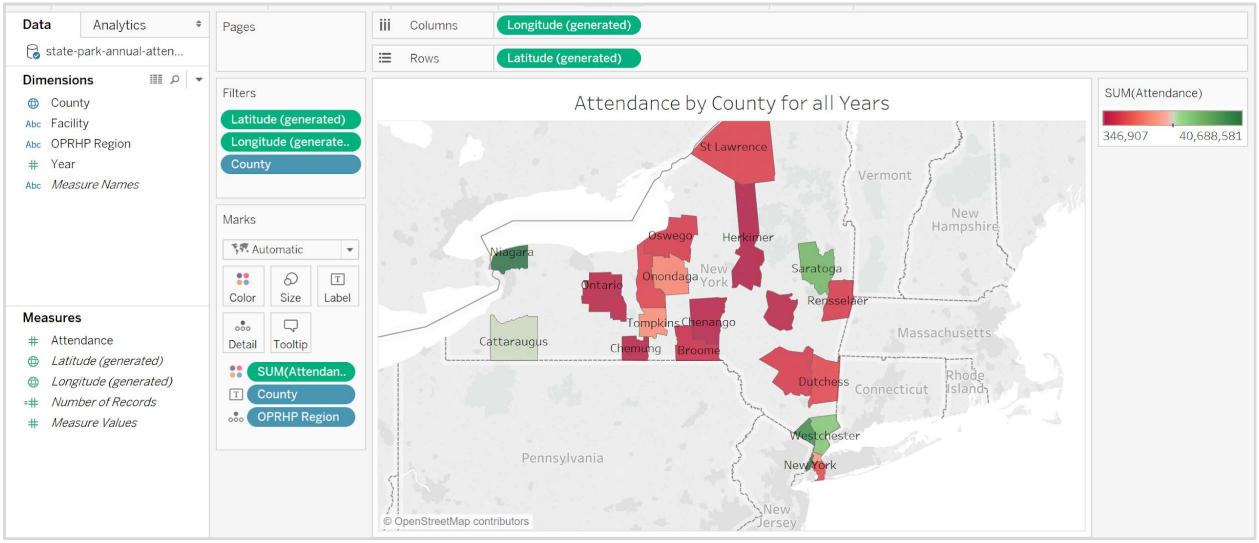
Figure 2: Attendance by County for all Years



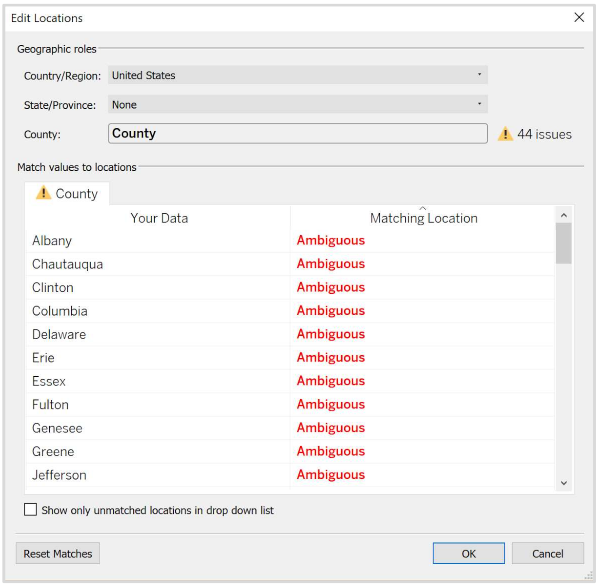
A lot of New York is not colored which is probably because of the number of missing values, Tableau does not recognize the latitude and longitude of certain counties, and maybe also because it is an incomplete sample. So this method of visualization in this situation is probably not the most truthful. Each "block" shows a different county. The county with the most amount of attendance overall from 2008 to 2016 is the Niagara county (Niagra OPRHP Region) with

140,628,246 followed by New York (New York City OPRHP Region) with 42,904,894, and Rockland (Palisades Region) with 35,714,082. Herkimer county has had the least attendance with 346,907 over the years. This can be figured out by looking at the map but the tooltip tool is the most useful here but unfortunately can't be shown in a text document. Also, many of the rows with multiple counties together are not able to be shown in this map.

The image below shows the options chosen for the Marks section and for the Rows and Column Filter.



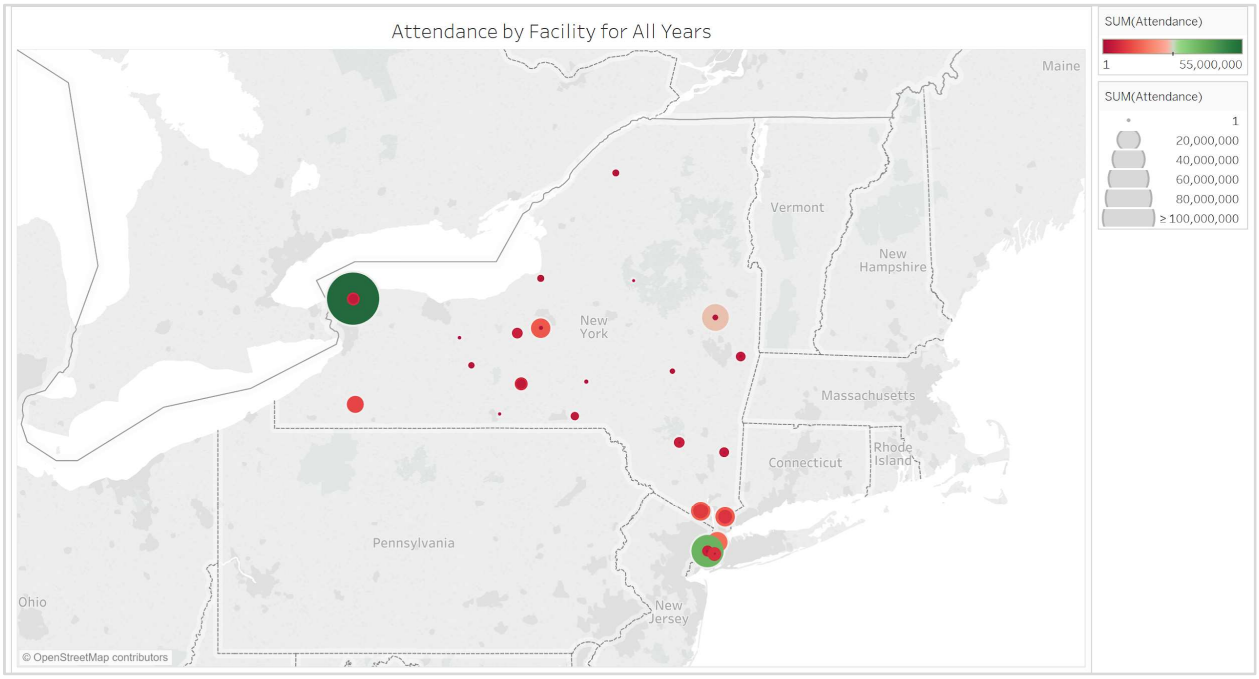
The counties are mostly labeled (unless there is not enough room) and colored by the sum of attendance over all years for all facilities in that county. The OPRHP Region is also included in the tooltip which can be seen if you hover or click on the county.



There are 44 counties that were either ambiguous or unrecognized.

Appendix

Figure 3: Attendance by Facility for All Years



This map shows the attendance by facility. The size and color is based on the amount of attendance for that specific facility over all years. Once again, the tooltip is the most useful tool. The Niagara Reservation is the facility with the most attendance over the years, the second most popular is Riverbank State Park followed by Saratoga Springs. The tooltip on this visualization shows the County, Facility, OPRHP Region, and the Attendance. Once again there is the issue of Tableau not understanding the location. There are 44 counties that were either ambiguous or unrecognized.

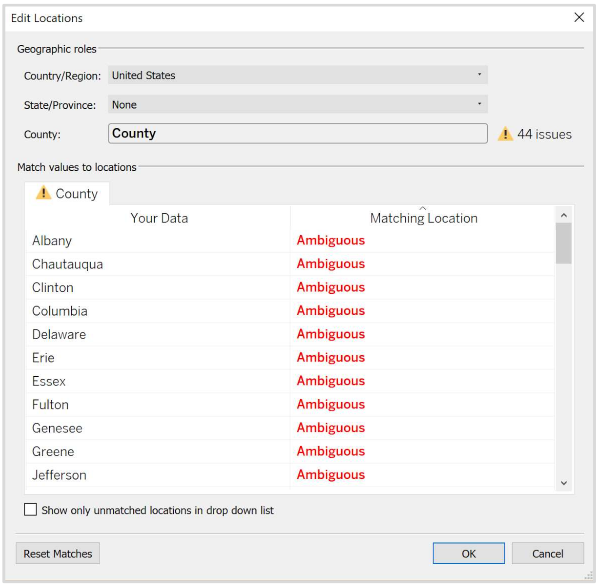


Tableau gives the option to enter a longitude and latitude in this case.

User Tasks in Terms of Shneiderman's Taxonomy for *Figures 1-3*

The article *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations* explains how users should look at visualizations. Shneiderman explains seven data types for visualizations "1-, 2-, 3-dimensional data, temporal and multi-dimensional data, and tree and network data." *Figure 1* which is the line graph, goes under the temporal data type taxonomy. *Figure 2* and *3* goes under the 2-dimensional data type taxonomy since it is showing the information geographically. There are also seven main user tasks which are overview, zoom, filter, details on demand, relate, history, extract.

The three main steps in Shneiderman's Taxonomy are "Overview first, zoom and filter, then details-on-demand." First the user can look at the general visualization to see what it is showing. They might need to find additional information about this data which can be found on the Kaggle website to understand more. Once they get the general overview they can look at each of the individual lines, facilities, or counties and see how much the attendance has been. After that, the user can get more details by clicking or hovering directly over certain points in the line.

Conclusion

The best Tableau visualizations for this dataset seems to be line graphs and maps. Bar graphs, circle views, and treemaps are possible but they either take up too much space or look very complicated. The colors can be easily changed by picking a different palette.

References

Shneiderman, B. (1996) The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. of IEEE Symposium on Visual Languages, Los Alamos, pp. 336-343.
https://www.kaggle.com/new-york-state/nys-state-park-annual-attendance-figures/home
https://parks.ny.gov/regions/
https://onlinehelp.tableau.com/current/pro/desktop/en-us/concepts.html
https://geology.com/county-map/new-york.shtml