Visualizing Data as a Tool for Gaining Insight into the Choice DJS Program

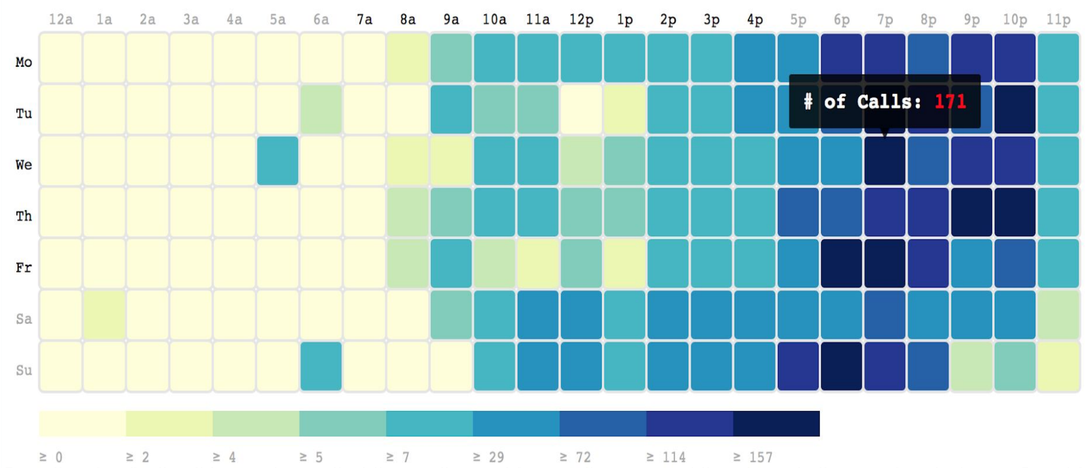
Sam Mendimasa, Danila Frolov, Celestine Wong, Cathy Poore, and Achuachua Tesoh-Snowsel

University of Maryland, Baltimore County

Overview and Goals

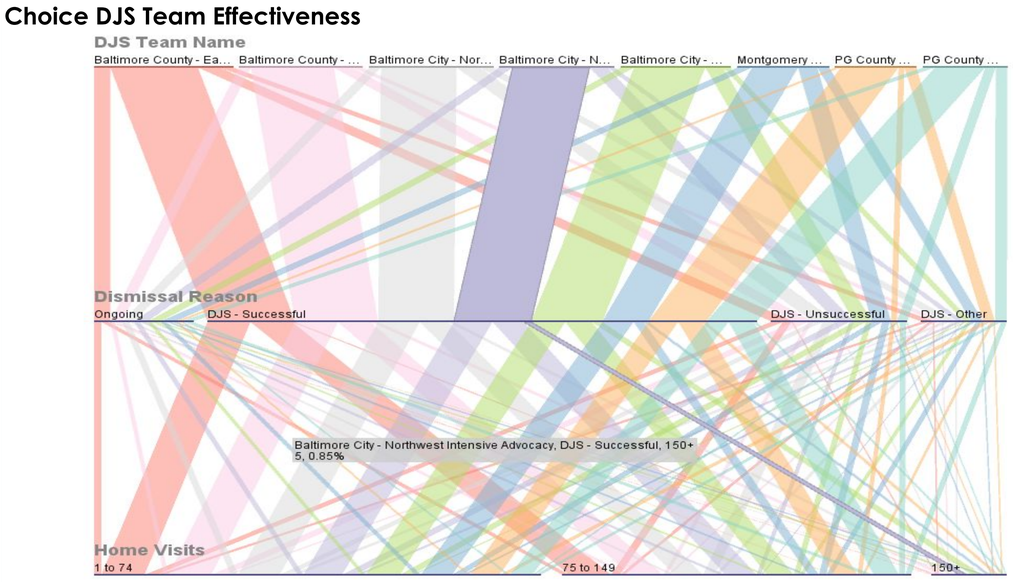
The context of the project derives from the initiative set by the University of Maryland, Baltimore County (UMBC)'s Choice Program. The Choice Program provides 50 AmeriCorps members the resources to provide different services to advocate for youth living in Maryland's high-risk communities. One of these services includes the Choice DJS Intensive Advocacy Program, a community-based alternative to incarceration. The Choice DJS Program works to reduce recidivism with daily face-to-face contact, increasing family engagement, and supporting youth with probation/legal requirements. Since DJS members advocate a large number of young people in several regions of the State of Maryland and visit dozens of people during a day, there is a lot of data such as demographics or daily reports that had to be collected and stored. Our client, Ms. Shirey Baig, is interested in finding a way to visually present this data in a way that will help to identify patterns and correlations of interest for her or the audience she wants to present it for.
Thereby, the main goal of our team is to design the visualizations that will extract the knowledge from generated visualizations in order to answer the following data-related questions of our clients: how the young participants of the Choice Program are distributed within regions served by DJS, how effective is the service of each of the DJS teams and finally, what is the best time for DJS members to reach the youth during the day. We are interested in understanding how effective these visualizations are for users, both as observers of a presentation and as analysts.

Choice DJS Weekly Phone Call Heat Map

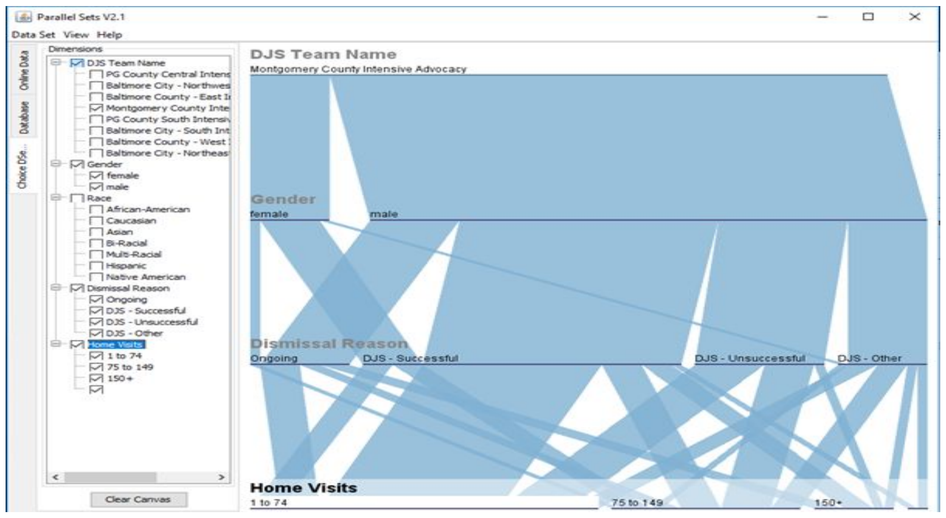


From the visualization it is easy to see that the youths most frequently answered the phone in the evenings between 5 and 10 pm. Very few phone calls were answered in the early mornings between midnight and 9 am, with a few exceptions as youth or AmeriCorps members are expected to be asleep. The day and time labels that are shown in black (Mo-Fr and 7a-4p) indicate the time frame in which the youths are typically in school, and as expected, the colors are lighter within that block indicating that less youth contact was made during those times.

Choice DJS Team Effectiveness



This image analyzes the effectiveness of the eight teams by exploring home visits per team and youth's outcome. Each team is represented by a different colored bar along the top with the width of the bar corresponding to the number of youths that the team served during the year. From the visualization, it is clear that about 60% of youths successfully completed the program during the year of the given data. It is easy to follow the percentage of youths that completed the program and to see the corresponding amount of home visits they received. About half of the youths received 75 or more home visits and it is clear from the visualization that the majority of those youths had successful outcomes. Hovering over each line in the visualization will highlight its overall connection as well as give its percentage of the dimension above and below. Through interaction with this visualization we can see that the Montgomery County Intensive Advocacy team had a higher percentage of unsuccessful youths and made less home visits than the more successful PG County South.



This image analyses the effectiveness of the Montgomery County team based on gender, outcomes, and home visits. Hovering over any of the lines give you the percentages of youths within this team that correspond to said category. An interesting trend which can be explored from this visualization is that 12% of this team are females, and despite their number of visits, none were unsuccessful. On the other hand, 88% of this team were male and of the 88% percent, 20% were unsuccessful. Of the 20% that were unsuccessful, 54% received 1 - 74 visits, 38% received 75 to 149 visits, and the remaining 8% received 150+ visits. On the other hand, although only 46% the Montgomery County team was successful, of those that were successful, 70% received 1 - 74 visits. Hence, one can conclude that having a lower amount of visits does evaluate to a decrease in effectiveness.

Implementation

Our product consists of three distinct visualizations, where each of them is designed to comply with our client requests.

1. Choice DJS Location Distribution:

One of the requests of our client was to provide a user the ability to see how the youth advocated by DJS teams are distributed within different regions of Maryland. Our solution to this problem was to create a dot distribution map where each dot represents a particular youth. The color of the dot corresponds to the team supporting that particular region. The key in the bottom of the image displays the names and corresponding colors of each of the eight teams. The visualization follows Shneiderman's mantra of "overview first, zoom and filter, details on demand". Glancing first at this visualization the viewer sees a geographical map with the group of colored dots representing all the youth in the database. Filtering this data is achieved via an extremely simple mechanism. By clicking on the name of the team displayed in the bottom a user can see the distribution of dots (youth) depending on the team selected. The search box on the right serves to filter participants by ID number, dismissal reason, date of birth, or race. Finally, scrolling over a particular dot allows the user to view demographic information for that participant, including number of home visits by team members and program dismissal reason. This visualization was created using Java and Tableau, where Java was used to parse the data and convert the addresses of the youth into the geographical coordinates of latitude and longitude, and then the modified dataset was uploaded to Tableau in order to generate a dot distribution map where each dot shows an accurate location of each youth participating in Choice Program.

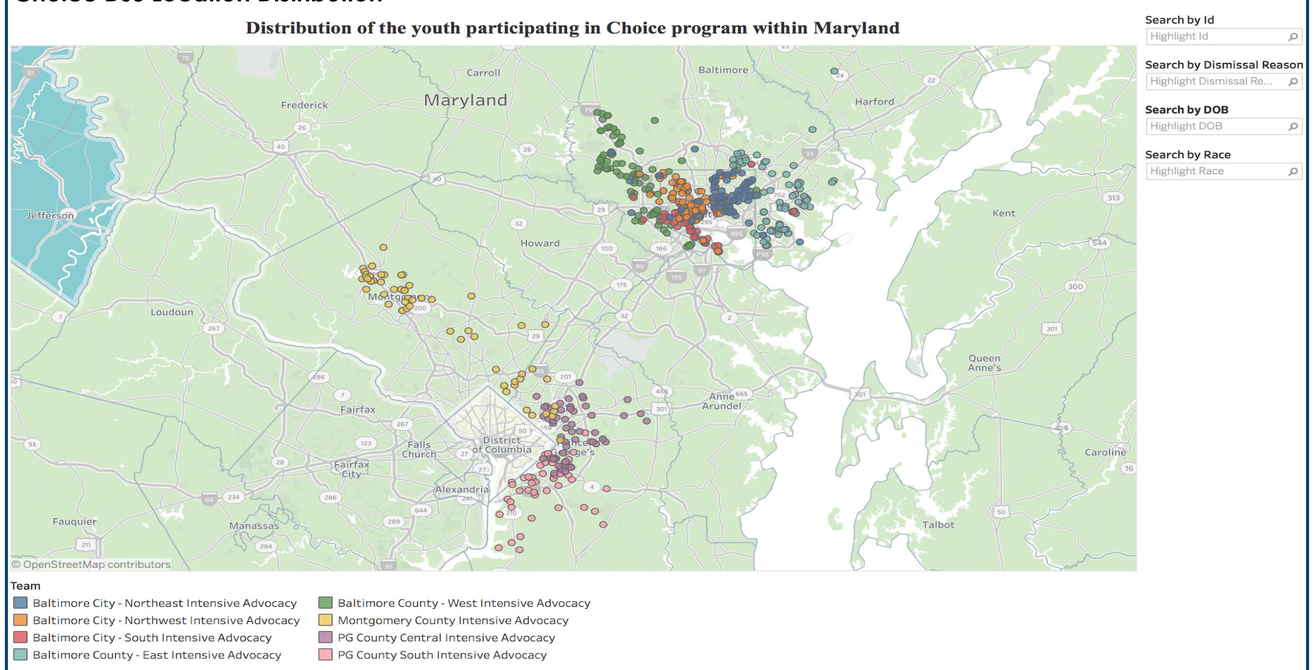
2. Choice DJS Weekly Phone Call Heat Map:

The second visualization is a heat map that addresses the best time to contact youth by phone during the week. This visualization required a time scale, and some means to show frequency of successful calls. The grid of the heat map is made from each day of the week and each hour in the day. The hour of the day is listed horizontally along the top of the grid, while the day of the week is listed vertically along the left of the grid. The colored squares represent the total number of calls in which youths were successfully contacted during that time of day of that day in the week across the entire year. Lighter colors represent fewer successful calls and darker blue squares represent higher numbers of successful calls. This heat map was made using the D3 and d3-tip library. The data is formatted to day-hour-value in a .tsv file, and then given to the D3 Day/Hour heatmap code. The code then creates ranges based on these values and groups them to a color. The color uses a sequential color scale with darker color meaning more successful calls. The heat map is responsive when hovering over each square. A tooltip for each square will appear and display the number of calls made throughout the year on that day, at that hour.

3. Choice DJS Team Effectiveness:

The third visualization is a Parallel Set that allows exploration of various dataset points (categories) that can be used to explore the effectiveness of the eight teams. Home visits are included to resolve our client's need and additional categories such as outcomes from the program, and race and gender are also included. This visualization was created using Parallel Sets V2.1, a Java based open source visualization application for categorical data. Parallel Sets V2.1 is limited in the sense that it only accepts categorical data in .csv files and it only works with Java jre 1.6. We have modified it such that it now works for jre 1.7 and above, and have included additional functionality, particularly with the colors to allow other trends to have a more shadowy and grayscale look when something is selected. Additionally, since some of our data were continuous, we used Java for parsing, removing errors, and converting all continuous data into to categorical data as a way of limiting our data points for our visualization to be more effective. Hence, continuous data for home visits were converted into 3 ranges: 1 - 74, 75 - 149, and 150+ visits. Also, categorical data such as youths outcomes that initially had fourteen outcomes were reduced to four in order to ensure that this visualization is less cluttered. This product can only be run locally, which is a positive as Choice Data contains personal identifiable information, and our project has been designed to only be used by our client and other members of Choice.

Choice DJS Location Distribution



This visualization allows the user to explore demographics of the youths and shows their location on a map. Users can look for clusters of youths that have additional characteristics in common, such as DJS team, race, or dismissal reason. Since the teams are each represented by a different color, it is easy to see the amount of youth served by the corresponding team. With each dot representing an individual participating in the program, there is clearly a larger distribution of youths within Baltimore City, around its edges, and clustered closer to the edge of Washington, D.C. than the outer areas of the different counties. This indicates that a lot of the youth served live in or near large cities. Almost each region has outliers, i.e. youth moved outside of the area that was initially attached to one of the DJS advocacies, however this team continues its service with that particular youth.

Conclusion and Future Direction

The three visualizations presented aim to answer many questions revolving around Choice's DJS Program. These visualizations give information that is more accessible than looking at the raw text data. For future improvements to looking at how number of visits may impact youth success, other components of the DJS program can be further examined. Such factors to consider can include location of youth, duration of visits, or involvement of parents to see if these may influence the number of successful cases. The heat map could also be expanded to add additional breakdowns by month or team and allow the user to toggle though those views for potential differences. An implementable feature that can be added to the heat map is the ability to upload multiple data from different years for comparison. The hope is that these visualizations can help the Choice program in future services and interactions with their youth.

Acknowledgements

The authors wish to thank Dr. Rheingans for providing valuable suggestions and guidance while creating this product, and Ms. Shirey Baig for evaluating our work to ensure our product would help and improve the Choice program.

Select References

[1] Stephen Few, "Introduction to Geographical Data Visualization", *Perceptual Edge Visual Business Intelligence Newsletter* , March/April 2009
[2] Jarke van Wijk and Edward van Selow, Cluster and Calendar based Visualization of Time Series Data. Proceedings of Information Visualization 99, pp., 1999
[3] Fabian Bendix, Robert Kosara, Helwig Hauser , "Parallel Sets: Visual Analysis of Categorical Data", Proceedings of IEEE Information Visualization 2005, pp. 133-140, 2005