A data-driven approach to early childhood intitiatives

Elba Gomez, Patricia Angkiriwang, Patrick Laflamme, Shenyi Pan
Data Science for Social Good 2017
The University of British Columbia, Canada

Summary: Avenues of Change - Guildford West is an initiative within Surrey whose goal is to increase the readiness of children to enter Grade 1. Child readiness is measured at the population level using the Early Development Instrument (EDI) [2]. The neighbourhood of focus for the initiative, Guildford West, was selected for two reasons: (1) the measured EDI score in that region was notably higher than other regions; and (2) it had an existing network of service providers working to achieve the same goal. Our approach aimed to probe both of these pieces in an effort to make data-driven recommendations on how to improve the initiative's efficacy. We partitioned our project into two parts, each addressing different dimensions of the Avenues of Change initiative.

In the first part, we examined two sets of relationships: first, the relationship between neighborhood environment (measured by library account creation and community program registration) and EDI and second, the relationship between community resource engagement (measured by service requests and crime data) and EDI. We concluded that none of these measures were significantly correlated with EDI. This suggests that even if Avenues of Change did have some impact on the neighbourhood of Guildford West, this impact may not necessarily be measurable by their key indicator, EDI. This highlights the importance of measuring other indicators of impact in the community.

In a second part, we examined the network of service providers using a network analysis. This was done with the internal Avenues of Change network, as well as with the network of early years service providers in Surrey as a whole. We visualise existing relationships and highlight some important areas in which the network could be optimized. In addition, we introduce an analytical tool to suggest new collaborations that could improve communication and information flow within the network. We provide a few examples in order to demonstrate how to use the tool. By better understanding the structure of the network of service providers, the initiative can make more informed decisions about changes to its structure.

Contents

1 INTRODUCTION

In this section, we introduce the background of the our project. This includes the context of the project, and the history of the initiative with whom we collaborated. Critical information that is relevant the present work is introduced for clarity. We then outline the areas of focus within this project, and why they are important. Finally, we detail how we intend to achieve our project goals.

1.1 BACKGROUND

Throughout the globe, governments are launching large-scale initiatives to ensure that their cities are sustainable, livable and equitable for their citizens. At the City of Surrey, this concept has been embraced through a multidisciplinary effort to improve their infrastructure, economy, and human capital. As part of their efforts, they are committed to improving health and development outcomes for their youngest citizens, specifically children aged 0-5.

One of the key early childhood development metrics adopted by the BC province and City of Surrey is the Early Development Instrument (EDI). EDI is a population-level measure of the overall developmental health and school readiness of children aged 0-5 within a particular region [5]. Given the underlying complexity of measuring childhood development, the measure incorporates a wide range of skills involved in school success - from communication skills and language, to physical strength and fine motor skills. EDI data is collected through surveys filled by preschool teachers, which are aggregated and presented as population scores within five distinct EDI categories [1]. If a child is labeled vulnerable in any of these categories, they are considered vulnerable in the aggregated EDI score. Thus, a higher regional EDI score means more children are vulnerable in one or more EDI categories. A more detailed discussion of the EDI measure can be found in section 4.1: ”Additional Details”.

In Surrey, the proportion of vulnerable children has remained relatively constant, and is comparable to the provincial average, both hovering around 30% [2]. Yet when we look at the 24 neighborhoods within Surrey, shown in Figure 1, we can observe very drastic differences in vulnerability rates. For example, in 2013, the lowest vulnerability score in Surrey was 22%, in Newton South-East, while the highest was 50%, in Newton East [6]. A complete table of vulnerability scores by neighbourhood over the 5 waves of data collection is shown in Table 1

In 2012, the City of Surrey tasked a group of IBM researchers, as part of a Smarter Cities challenge grant, to investigate how the City and its partners can best invest in early childhood development and share information and insights across service providers, funders, and sup-

[1] The 5 EDI Categories are ’Physical Health and Well-Being’, ’Social Competence’, ’Emotional Maturity’, ’Language and Cognitive Development’ and ’Communication Skills and General Knowledge’

[2] In the most recent (2016) EDI wave, BC scored 32.2% while Surrey scored 34%. Link: `http://earlylearning.ubc.ca/media/edi_w6_communityprofiles/edi_w6_communityprofile_sd_36.pdf` page 9

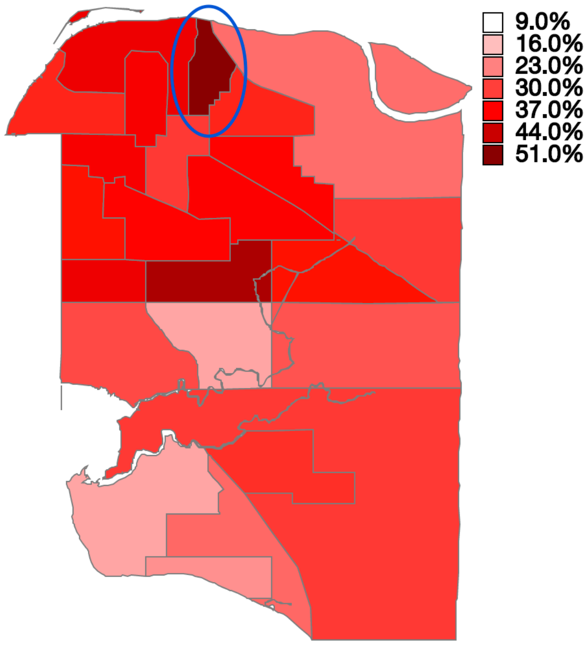

FIGURE 1: Overall EDI scores within Surrey for 2015. Note that Guildford West (Circled in blue) has the highest vulnerability in the region. Also note the drastic variability in EDI scores between neighbourhoods.

porters [1]. Their report recommended establishing a task force on early childhood development that would share information as well as implement pilot programs [1]. Soon after, the City of Surrey partnered with United Way to implement IBM's recommendations in some of the more vulnerable neighbourhoods in the city. This task force became known as Avenues of Change (AoC) - Guildford West, a network of local agencies and public partners focused on early childhood development. The neighbourhood Guildford West was selected for two main reasons: first, its consistently high EDI vulnerability rates, and second, Guildford West's already existing community infrastructure and the large amount of initiatives already present. The latter was used to distinguish Guildford West from other equally vulnerable neighbourhoods, such as Newton East, as it allows for a more immediate implementation of change within the neighbourhood.

By introducing Avenues of Change, City of Surrey and United Way hoped to observe an increase in collaboration between the multitude of agencies addressing children's development within Surrey. Instead of having agencies working on their immediate mandates, the network allowed for an all-encompassing long-term goal and a consistent stream of interactions. The network is broken into two parts: a Joint Leadership Team (JLT) and working groups. The JLT would provide guidance and aim to increase the network of partner organizations, building a larger support system for the community. Meanwhile, the working groups would work on specific mandates, such as child mental health, or park safety.

In terms of the community itself, the neighbourhood Guildford West was selected for two main reasons: first, its consistently high EDI vulnerability rates, and second, Guildford West's

Neighborhood	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Surrey City Centre	42	37	40	31	39
Whiterock	27	30	25	27	21
Newton Southeast	25	14	18	22	19
Semiahmoo	14	11	17	23	19
South Surrey East	27	21	22	35	31
Cloverdale South	26	22	30	27	28
Cloverdale North	20	22	18	34	35
Fleetwood Southwest	31	25	33	31	38
Fleetwood Northeast	24	24	34	35	36
Guildford East	21	14	38	28	25
Guildford Center	26	30	33	36	33
Whalley North	31	46	43	37	40
Whalley West	33	24	34	30	33
Newton West	38	44	39	39	40
Newton Northwest	31	27	38	41	35
Newton East	35	36	46	50	47
Newton Northeast	40	31	33	42	36
Whalley Southwest	40	40	38	45	39
Whalley Southeast	25	24	32	29	31
Clayton	24	29	17	26	31
Rosemary	9	10	15	32	32
South Surrey West	22	21	20	30	26
Guildford West	38	40	45	43	51
Newton Southwest	33	27	29	37	29

TABLE 1: Vulnerability scores by neighbourhood in Surrey over the 5 waves of data collection.

already existing community infrastructure and the large amount of initiatives already present. The latter was used to distinguish Guildford West from other equally vulnerable neighbourhoods, such as Newton East, as it allows for a more immediate implementation of change within the neighbourhood.

1.2 PROBLEM STATEMENT

With the release of the most recent EDI collection wave came great disappointment for Avenues of Change. Childhood vulnerability in Guildford West had gotten worse, increasing from 43% vulnerable in 2013 to 51% vulnerable three years later [6]). Additionally, the change indicators developed at the beginning of the initiative had not yet been collected and there was minimal data available to speak of their impact. There were two possibilities for this: (1) Avenues of Change - Guildford West was not having a positive impact on the community as intended, or (2) the impact was not measured by the current indicators.

In order to identify the impact that Avenues of Change is having on Guildford West, the City of Surrey provided us with a number of data sets that are transactional in nature. Repurposing procedural interaction data into something meaningful to early development adds an additional layer of complexity to our project. In addition, we were tasked with increasing the amount of data-driven decision making within the initiative. To unify these two tasks, we aim to devise a project that identifies points of impact within the community and highlighted how data-driven decision making can help magnify that impact. For this reason, we split the project into two parts.

In the first part, we explore the relationship between the various transactional datasets provided to us and the neighbourhood EDI scores. EDI is the main metric used for assessing progress within AoC. Yet EDI is affected by a multitude of factors, which makes it challenging to separate the impact of different AoC initiatives only by looking at this aggregated population metric. Several AoC network members, who provided various anecdotal explanations regarding the significant increase in EDI between 2013 and 2016, echoed this sentiment. Thus, this part aims to address some of these anecdotes within the context of the data available. In addition, we investigate the impact, if any, of the Avenues of Change project on the variables within transactional data collected in Guildford West.

In the second part, we aim to increase our understanding of how the network of partners in the Avenues of Change initiative is structured. This network, a direct consequence of recommendations made in the IBM Smarter Cities report to improve early childhood outcomes, needs to rely on effective communication to achieve its main goal. An upcoming restructuring of the AoC network is an opportune time to evaluate the ways in which the present network structure succeeded, and where it could use improvement. To date, there has only been anecdotal evidence describing the network dynamics within AoC, so we supplement this qualitative understanding with concrete metrics and visualisations to allow for a better understanding and assessment of the network. The goal of this part is to describe the network in its current form and to provide an analytical framework for the network structure to be efficiently analyzed as changes are made in the future.

Finally, we aim to visualize all of the data provided to us by the City of Surrey and Avenues of Change. In order to cater to members that are starting the transition towards more data driven analysis, we develop a dashboard to contain the disparate data sources and present them in a unified way. We hope that this dashboard tool, in addition to acting as an interactive summary piece of our project, inspires network partners to interface with the data they have available when making decisions.

2 RESULTS AND CONCLUSIONS

In this section, we discuss the main results and conclusions we obtain from our project. We begin with the analyses regarding the prediction of EDI. Then we introduce the network analysis results on Avenues of Change. Finally, we give an overview of the dashboard tool that we implement for data visualization and exploration.

2.1 PREDICTION OF EDI

Since EDI is currently used as the main measure to evaluate the effectiveness of AoC program, it is of interest to further investigate whether library account creations, community program registrations, as well as other available variables, are associated with the change of EDI. However, EDI has only been measured five times within the city of Surrey: in 2005, 2008, 2010, 2012, and 2015. The sparse number of EDI observations limit our ability to fit advanced models to the available data sets. As a result, in our project we mainly adopt linear regression as well as logistic regression models to predict EDI based on various variables.

First, we start the analysis by predicting EDI with library account creations and community program registrations. For these two data sets, we have access to the monthly account creation counts and program registration counts starting from 2001. Taking into account that EDI is only available in the form of five waves, we aggregate library account creations and community program registrations by year and neighborhood, then correct the yearly count data by the population size of each neighborhood to obtain the yearly library account creation counts and program registration counts per capita and use them as the predictors for EDI. However, by computing the correlation coefficient between these two variables and EDI, we find that neither are meaningfully correlated with EDI. For the library account creation counts, we find that there is simply no relationship with EDI, with an observed correlation of 0.03. For the program registration counts, we find that there is a weak relationship with EDI, with correlation coefficient of 0.33. However, we find that if we perform the same analysis while statistically holding Social Economic Status (SES) constant, we find that this relationship disappears, meaning that program registration counts correlate with EDI only because it contains information about SES. Furthermore, fitting a linear regression model with these two variables as well as their interaction term barely has any prediction power on EDI, yielding coefficient of determination equal to 0.0955. The coefficient of determination can vary between 0 (no predictive power) and 1 (perfect predictive power), so the observed value indicates nearly no relationship. Both of these results suggest that early library account creation counts and program registration counts per capita are not strong predictors for EDI.

Despite their poor predictive capabilities on EDI, it is worth noting that there is some reason to believe that library account creations and program registrations have increased since the beginning of Avenues of Change. Figure 2 shows that the number of library account creations per capita in 2015 and 2016 are up considerably. Figure 3 shows that while program registration counts have decreased since 2010, they have shown a marked recovery since 2015. This may suggest that Avenues of Change is indeed having a positive impact in the neighborhood, but that this impact is simply not captured by such a big-picture measure such as EDI.

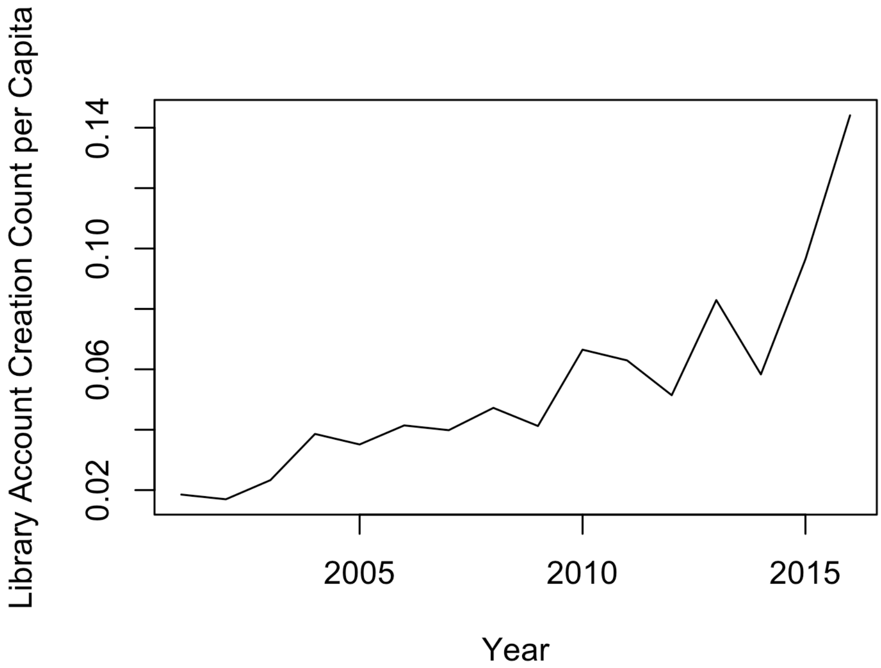

FIGURE 2: Library account creations per capita over time within the Guildford West neighbourhood. It is clear that the library account creations have seen a considerable increase since the beginning of the Avenues of Change program.

Next, we use population density of each neighborhood to predict EDI. The neighborhood-wide population data are available every five years through the Canadian census. We apply linear interpolation to the census population data to get the estimated population data for every year. Although linear interpolation cannot may not be able to capture the nonlinearity in the population change, it is still a reasonable approximation for the population data since the population change between two census waves is usually small. Linear interpolation was used since, upon visual inspection, most of the population changes across the four censuses appeared roughly linear in nature. It is worth noting, however, that considerable error is added to our final variables of interest due to this interpolation step. Unfortunately, the years in which EDI data was collected seldom overlap with census years, nor do we have access to data that would give us more reliable yearly estimates from 2006-2016. As a result, interpolation is the only viable option available.

After obtaining the yearly population data, we then divide the interpolated population data by the area of each neighborhood. It is found that there is a moderately high correlation coefficient between population density and EDI (correlation coefficient of 0.6065). The scatter

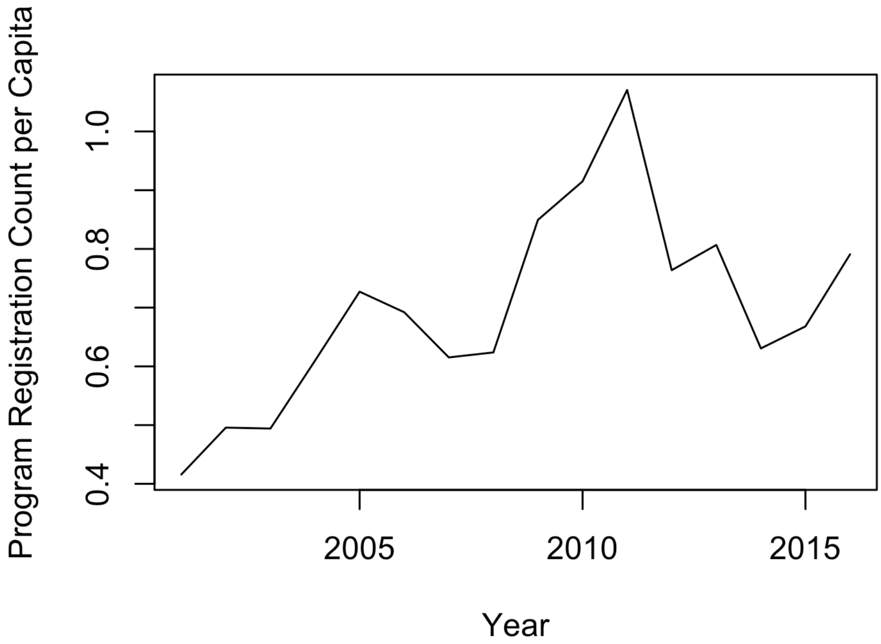

FIGURE 3: Community program registrations per capita over time within the Guildford West neighbourhood. While community registrations seem to have peaked around 2010, there has been a marked recovery in community registrations since 2015.

plot of EDI against population density can be found in Figure 4. If we fit a linear regression model with population density as the only predictor, the coefficient of determination of the fitted model is 0.3132, implying that around 31% of the total variance in EDI can be explained by population density, which makes population density alone a reasonably good predictor for EDI. However, previous research from UBC HELP shows that Social Economic Status (SES), which is a linear combination of several variables measured in census, is highly negatively correlated with EDI (our analysis shows that the correlation coefficient is -0.8342) [7]; the scatter plot of EDI against SES can be found in Figure 5. Further analysis shows that the correlation between SES and population density is -0.7258 (the scatter plot of SES against population density can be found in Figure 6), while the partial correlation between population density and EDI conditioning on SES is close to 0. This suggests that the correlation between population density and EDI can be mostly attributed to their high correlations with SES. Nonetheless, one limitation of SES is that it is extracted from census data, which is only available every five years. Given that SES is a good predictor of EDI, we can use yearly population density data to approximate SES, and then use this approximation as an estimate yearly SES levels within the neighbourhoods.

Finally, we analyze the service request and crime report data sets to check their association with EDI. Note that service request and crime report data are only available starting from 2011. There are only two corresponding EDI waves, which further reduces the number of observations. Therefore, instead of predicting the EDI values, we decide to predict whether the

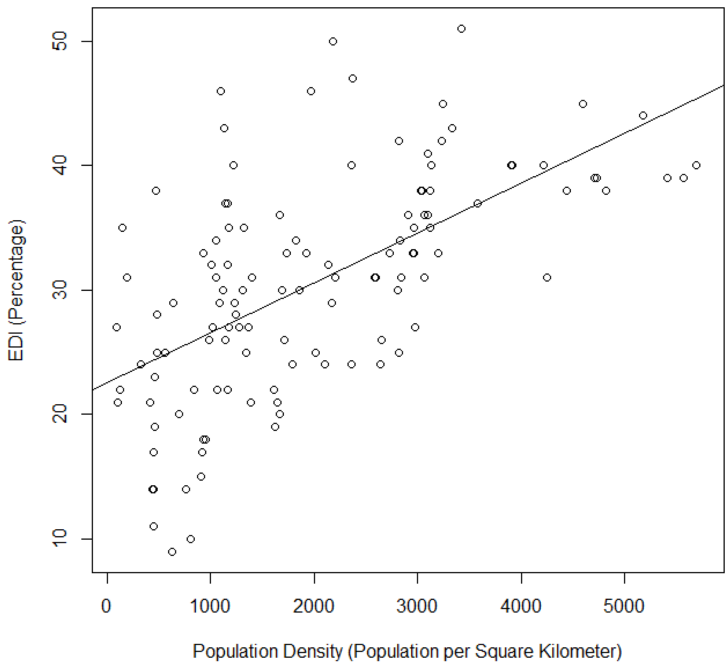

FIGURE 4: Scatter plot of EDI against population density in 2006, 2009, 2011, 2013 and 2015.

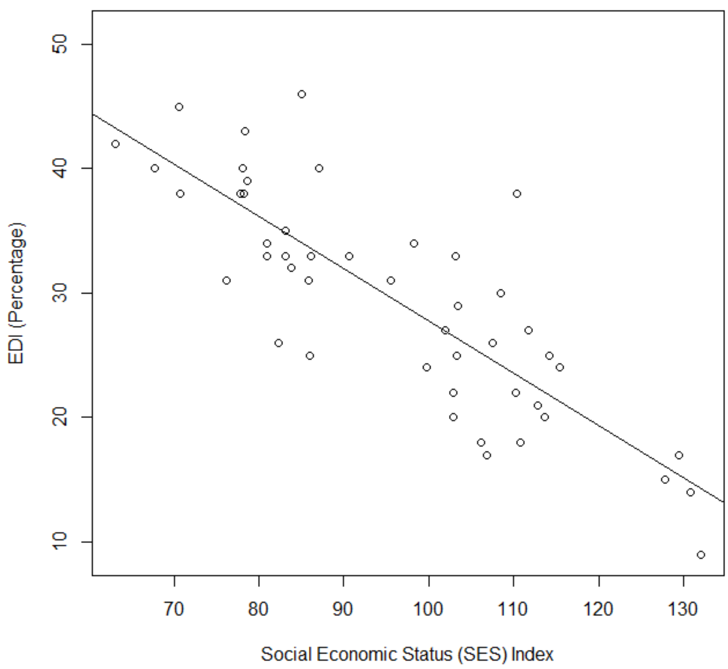

FIGURE 5: Scatter plot of EDI against SES in 2006 and 2011.

EDI measurement increases or decreases since the last wave to transform the regression problem to a binary classification problem. In addition, since there are a relatively large amount of variables in the crime report and service request data sets, we decide to perform principal com-

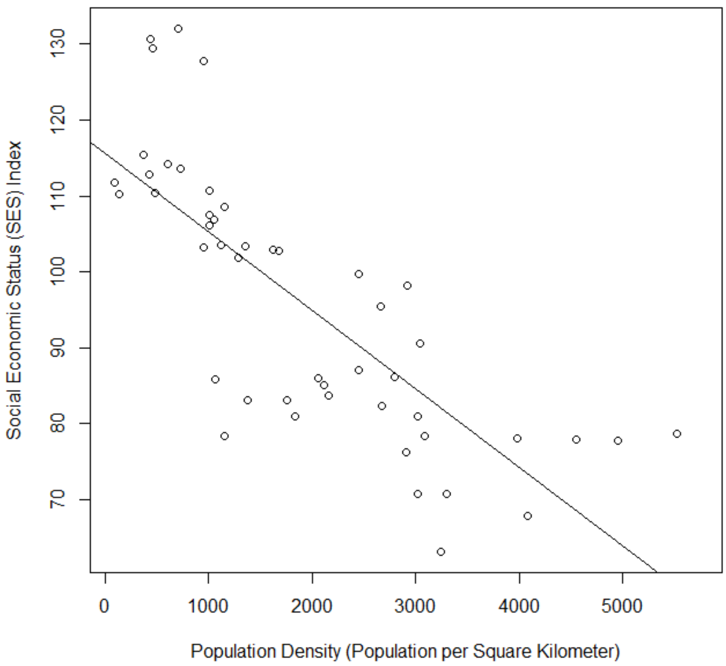

FIGURE 6: Scatter plot of SES against population density in 2006 and 2011.

ponent analysis (PCA) on those variables. Then we use the meaningful principal components as the predictors instead of the original variables to achieve dimension reduction. In this case, our analysis yielded 4 meaningful principal components. However, fitting a logistic regression model with these principal components has high error rate on predicting EDI increase or decrease. For example, the fitted logistic regression model with the principal components obtained from the current year and the previous year gives prediction error rate 45.83%, which is not much lower than the 50% error rate of random guessing. Furthermore, we also apply logistic regression to predict whether there is a significant increase in EDI or not [3] using the same set of principal components. Again the fitted logistic regression model with the principal components obtained from the current year and the previous year gives a high error rate of 37.5% for predicting whether there is a significant increase in EDI. Both analyses suggest that service request and crime report data are not strong predictors for EDI.

[3] the technical details about EDI significant change can be found on http://earlylearning.ubc.ca/supporting-research/critical-difference/

2.2 NETWORK ANALYSIS

To better understand the present structure of the network of agencies involved in early childhood development initiatives, we construct network maps and perform social network analysis on the AoC network. We repeat this process on the larger network of early years service providers in Surrey to understand how the AoC network fits into the broader context of Surrey-wide early years initiatives. In each of the network maps, we refer to each entity as “nodes” and the links between them as “edges”.[4]

For the AoC network, a graph was constructed based on meeting co-attendances from the JLT and working group meetings conducted from the beginning of 2016 to early 2017, shown in Figure 7. For privacy purposes, most agency names are not disclosed in this publication and are instead replaced with an anonymous number IDs. In cases where more than one individual from a single organisation attends meetings, their attendance is only considered once. In Figure 7, we see that the members who are a part of the JLT are very tight-knit, while the working-group only members remain on the periphery. This may simply be an artefact of the data, since meeting minutes for the JLT meetings were much more plentiful. However, it may also be a sign that the JLT is relatively isolated from other members. Beyond the biases introduced by the imbalance in meeting frequencies, constructing a network using meeting minutes does not capture the level of engagement with other network members. It is possible that some members of the network engage in proceedings to differing degrees.

To supplement the information provided by the meeting minutes data, and to address our concerns regarding the biases within it, we conducted a survey of AoC network members to ask two simple questions: first, which (up to 3) agencies they have strong collaborations with, and second, to which partners they would go to in order to share new ideas. We obtained responses from 13 individuals from 9 out of the 10 organisations currently on AoC’s members list. We plot the results from the survey in the map is shown in Figure 8. It is worth noting that nodes 4, 8, 11, and 13 are very central. All four of those nodes were named as collaborators or innovators by many people and, in network analysis terms, have a high in-degree centrality.

For these networks, we also build an analytical tool to suggest new connections that could improve a network by various metrics, such as an increase in efficiency of information transfer. To do this, the tool loops through edges that do not currently exist, computing the average shortest path length for the graph with each possible new edge. Average shortest path length is a measure of the efficiency of information flow, and we ignore any directionality that the graphs might have when computing this measure. Of all the possible edges that decrease av-

[4] As a caveat, we make no attempt to distinguish between attendance of various members of the same organisation. The result is that while an organisation in our network may appear to be connected to 3 other organisations, it is possible that each of these organisations interface with a different individual. If communication is less than effective between individuals of a large organisation, this adds significant error to our analyses of information flow.

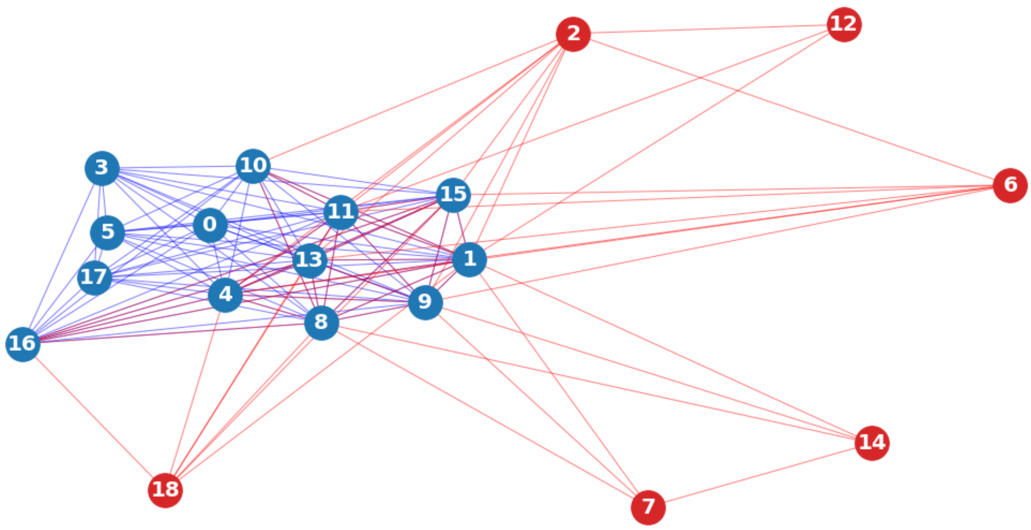

FIGURE 7: Network graph built from Avenues of Change Guildford West (AoC) meeting co-attendance data from 2016 to early 2017. Red edges: co-attendance at AoC working group meetings. Blue edges: co-attendance at AoC Joint Leadership Team (JLT) meetings. Purple edges: co-attendance in both types of meetings. Red nodes: agencies who have only attended working group meetings. Blue nodes: agencies who have attended at least one JLT meeting. The closeness of the nodes related to the number of meetings the agencies co-attend. In the case where an agency is represented by more than one individual at a meeting, their attendance is counted only once.

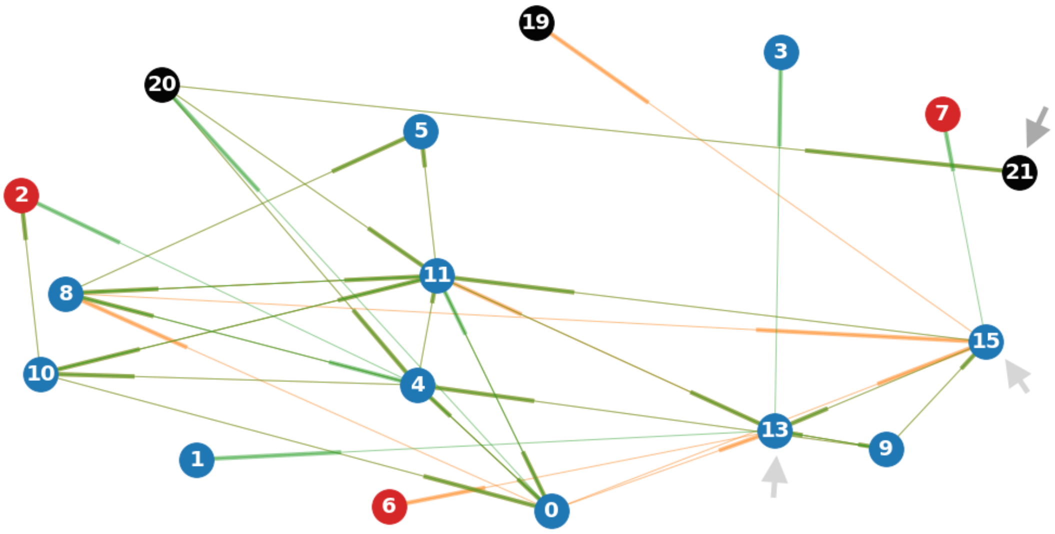

FIGURE 8: Network graph built from survey data on collaboration and innovation. Agencies on the Avenues of Change Guildford West (AoC) members list were asked to cite the top 3 agencies for each of the following criteria: (1) have had strong collaborations with in the past two years, and (2) they would talk to discuss a new or innovative idea. Orange edges: cited collaboration with given agency. Light green edges: cited innovation with given agency. Olive green edges: both collaboration and innovation. The thicker portion of the line represents an "arrowhead", indicating the directionality of the relationship. Red nodes: agencies who have only attended AoC working group meetings. Blue nodes: agencies who have attended at least one AoC Joint Leadership Team (JLT) meeting. Black nodes: agencies who did not attend AoC meetings in 2016 to early 2017. Our analyses suggest that connecting node 21 with node 13 or 15 (indicated by grey arrows) would improve efficiency of information transfer.

erage shortest path length (i.e. increase efficiency), the tool selects those above the 99.9th percentile and labels them as the suggested new connections that would most dramatically increase information flow.

In the network built from our survey data on collaboration, our tool finds that a new connection between node 21 and node 13 or 15 would increase efficiency of information flow the most. From the network map in Figure 8, we can see that this would connect a relatively isolated node with extremely connected nodes in the network (see arrows in Fig. 8). In the case that a specific organisation, say the City of Surrey (node 4), were choosing agencies to collaborate with, our tool suggests that collaborating with nodes 1, 3, 6, or 13 would most improve information flow in the network.

It is informative to compare the two perspectives on the same AoC network, meeting co-attendance and survey responses in Figures 7 and 8 respectively, by mapping their overlaps and non-overlaps. In Figure 9 below, we present the same graph derived from AoC meeting co-attendances shown in Figure 7, but we highlight the co-attendances between two agencies that also involve a survey response on collaboration or innovation. It is clear that many agencies have co-attended the same meetings but do not cite the other agency in their survey responses, or were not surveyed because they were not on the most recent AoC members list. All but three of the highlighted edges between agencies (involving both co-attendance and collaboration) connect two agencies that meet at a JLT meeting, not just a working group meeting.

In Figure 10, we present the same graph derived from survey responses shown in Figure 8, but we highlight the survey responses (collaboration and/or innovation) that also involve meeting co-attendances at a JLT meeting, working group meeting, or both. Again, many of the cited collaborations were between two agencies who have co-attended JLT meetings, or both JLT meetings and working group meetings– rarely just working group meetings. In fact, it is just as likely to see a cited collaboration between two agencies who do not meet at all as part of AoC as it is to see one between two that have met only at working group meetings.

Finally, to characterise the broader Surrey network, we use official membership data for each working group, table, or task force focused on early years initiatives in Surrey, linking agencies to each group they belong to. This shown in Figure 11, and suggests there are many redundancies between groups, with all pairs of groups sharing 3 to 10 of the same member agencies. Though having multiple groups containing the same agencies is sometimes necessary, too many redundancies could eventually lead to time lost in unproductive meetings.

Applying the edge analysis tool to the group membership graph, one can obtain recommendations about which members should be added to a given working table in order to increase information flow across the network. For example, if the network wanted to improve the efficiency of information travels from one agency or group to another, they should encourage table 26 to add nodes 29, 41, or 42 to their membership list (see arrows in Fig. 11). It is worth noting, however, that these suggestions do not take into account the mandate of the working table, and so should be interpreted critically by the person applying the tool. For example, it

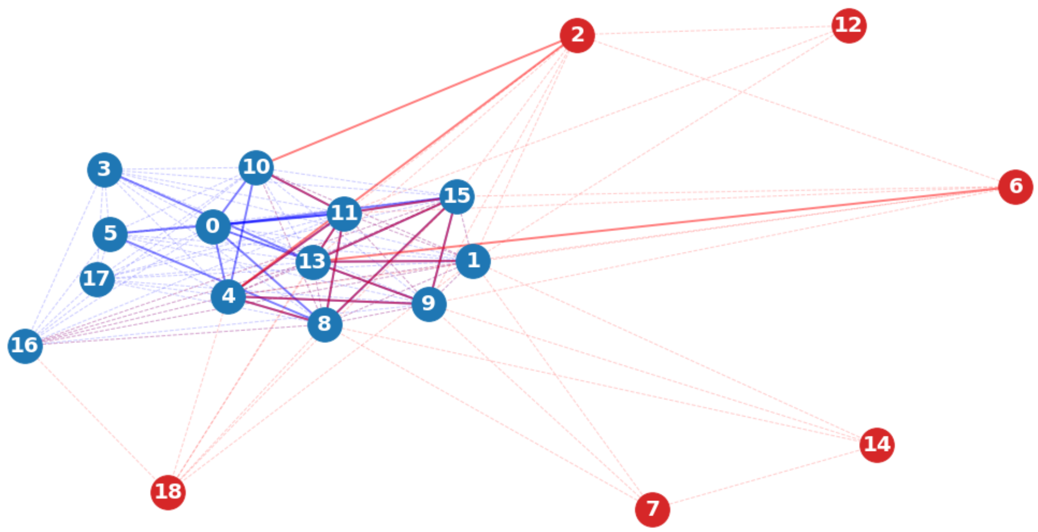

FIGURE 9: Network graph built from Avenues of Change Guildford West (AoC) meeting co-attendance data from 2016 to early 2017, with added information from survey data on collaboration and innovation. Here, solid lines represent edges connecting two agencies that both co-attend meetings and have a cited collaboration, while faded dashed lines represent co-attendances without cited collaborations. Red edges: co-attendance at AoC working group meetings. Blue edges: co-attendance at AoC Joint Leadership Team (JLT) meetings. Purple edges: co-attendance in both types of meetings. Red nodes: agencies who have only attended working group meetings. Blue nodes: agencies who have attended at least one JLT meeting. The closeness of the nodes related to the number of meetings the agencies co-attend. In the case where an agency is represented by more than one individual at a meeting, their attendance is counted only once.

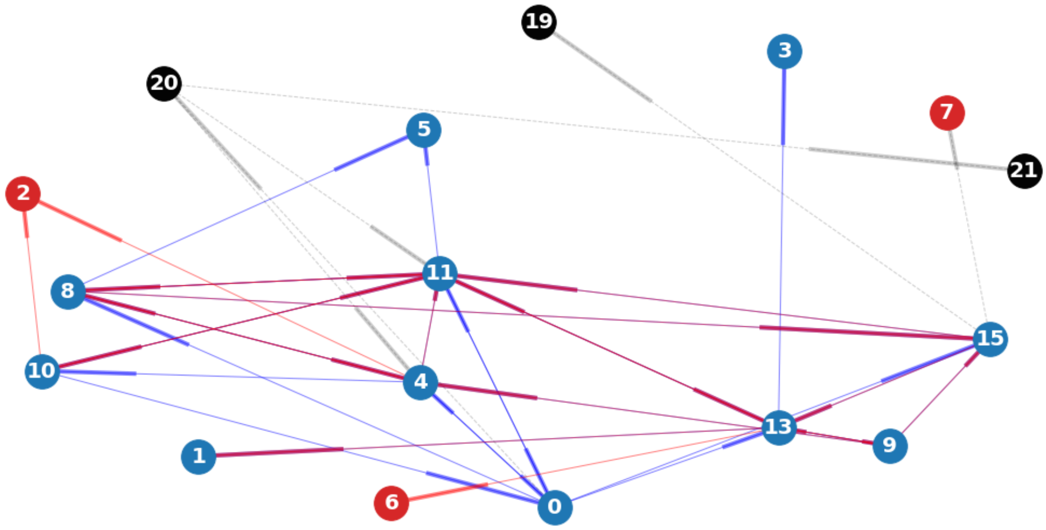

FIGURE 10: Network graph built from survey data on collaboration and innovation, with added information from Avenues of Change Guildford West meeting co-attendances. Edges connect two agencies where at least one cites the other as one of the top 3 agencies they either (1) have had strong collaborations with in the past two years, or (2) they would talk to discuss a new or innovative idea. The colour of the edges represents the type of meetings at which the two agencies co-attend, if at all. Red edges: co-attendance at AoC working group meetings. Blue edges: co-attendance at AoC Joint Leadership Team (JLT) meetings. Purple edges: co-attendance in both types of meetings. Grey edges: no co-attendance at AoC meetings. The thicker portion of the line represents an "arrowhead", indicating the directionality of the relationship. Red nodes: agencies who have only attended AoC working group meetings. Blue nodes: agencies who have attended at least one AoC Joint Leadership Team (JLT) meeting. Black nodes: agencies who did not attend AoC meetings in 2016 to early 2017.

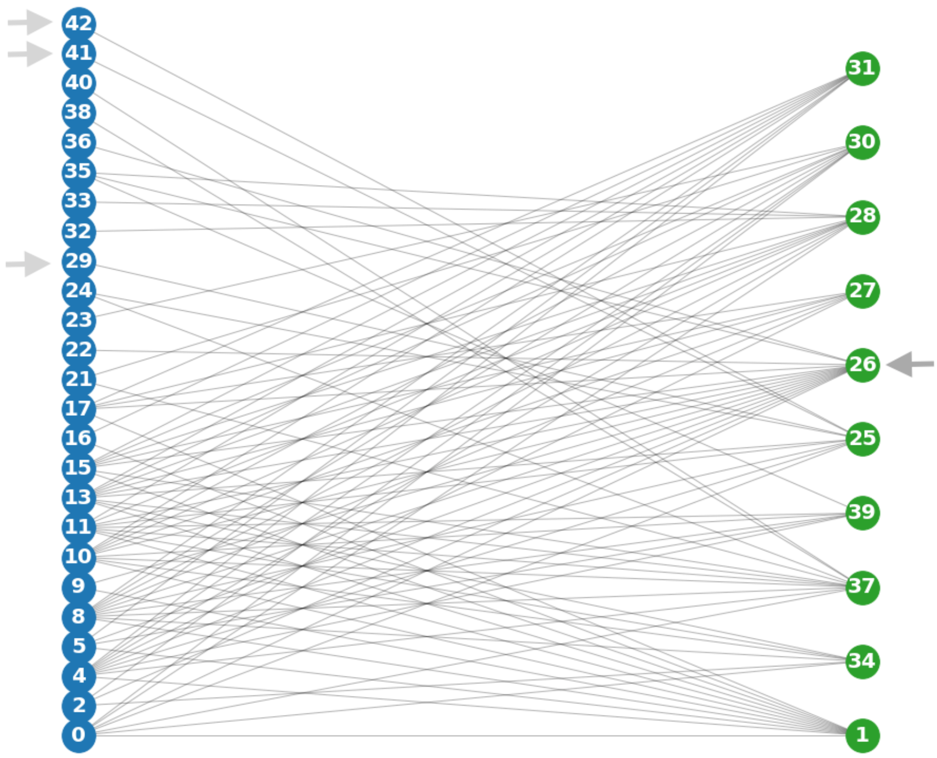

FIGURE 11: Two-part network graph built from group membership data for early childhood-related working tables and task forces in Surrey. Blue nodes: Early Years agencies. Green nodes: Early Years groups. Our analyses suggest that connecting node 26 with nodes 29, 41 or 42 (indicated by grey arrows) would improve efficiency of information transfer.

may not make sense to add a group from Guildford West to a working table that works on issues in Newton East. Nonetheless, this tool may be helpful in inspiring new relationships within the network.

An alternative way to visualise this group membership information is to plot only the agencies on the graph and to link them based on whether they are both members to the same group, as shown in Figure 12. Here, the thickness of a line connecting any two agencies is proportional to the number of groups the two agencies both belong to. With this representation of the graph, we observe a central cluster of agencies that tend to meet in the context of multiple task forces or working groups. This is consistent with the earlier observations of redundancy between groups, with multiple groups containing similar sets of agencies.

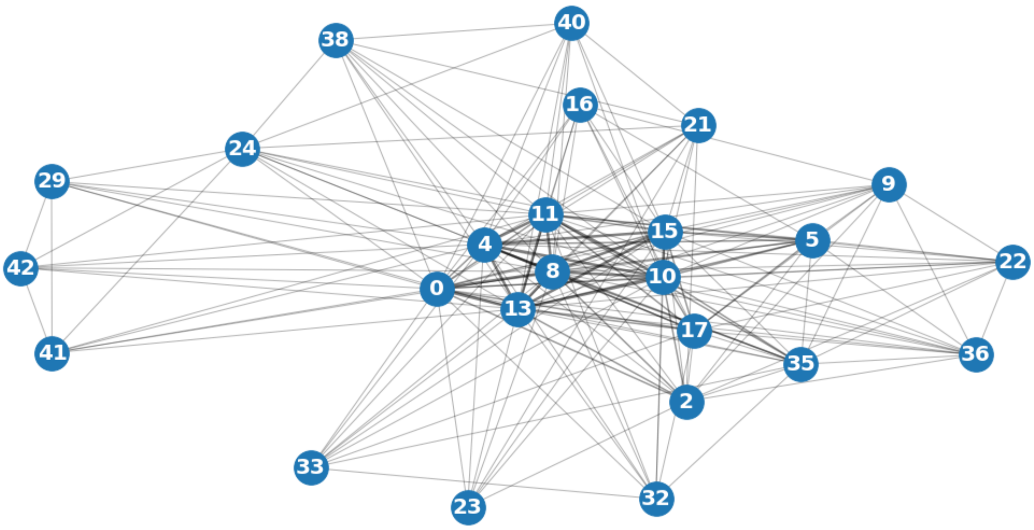

FIGURE 12: Network graph representing the relationships resulting from early childhood-related working tables and task forces in Surrey. Each of the blue nodes represents an agency on an early childhood-related working table or task force in Surrey. They are connected by edges to agencies that belong to at least one common group. The closeness of the nodes related to the number groups the agencies share.

2.3 DASHBOARD

The last goal for the present project was to make the data we worked with accessible to a wider audience. Many of the datasets we worked with, such as the Canadian Census and community program registrations, required extensive cleaning, organisation, and manipulation to prepare them for use and interpretation. Unfortunately, not everyone has to the skills to perform such steps. Nonetheless, using data to drive decision making can greatly aid the impact of initiative such as Avenues of Change. To address this discrepancy, we developed a dashboard that presents the cleaned data in a way that is both easily understood and queried.

Two driving principles shaped the nature of the dashboard. First, it must be easy and intuitive to navigate. To address this, we built the dashboard as a website. In this way, users will already be comfortable with how to navigate from one visualization to another. Second, it must be easy to interact with the visualizations, and ask questions of the data presented. We addressed this concern by using simple sliders and drop-down menus as the primary method of interaction with our dashboard. In addition, we ordered our visualizations such that they guide the user through the data in a coherent, structured manner.

Through the course of building the dashboard, a few notable observations surfaced. First and foremost, visualizing the population within Surrey reveals a very high concentration of early children within the southern portion Guildford West, starting in 2006 in Figure. The high concentration of young children in Guildford West has become increasingly drastic in the past decade. In conjunction with our findings that population density is related to SES and EDI, it may explain the sudden up-tick in vulnerability rates within the neighbourhood in the past 3 waves. On the other hand, another notable finding was that Guildford West was not the most vulnerable in all categories. In Figure 13, with respect to emotional vulnerability, the neighbourhood has been far from the most vulnerable region, falling within the regional norm.

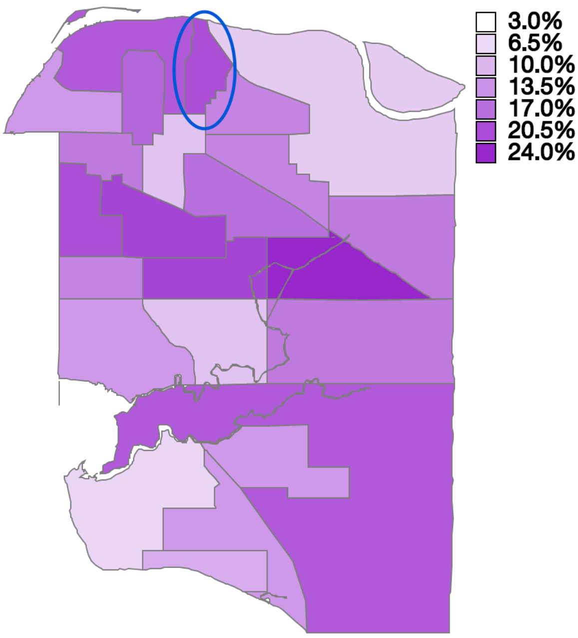

FIGURE 13: Emotional vulnerability rates with Surrey in 2015. Darker colours are associated with higher vulnerability rates. Note here that Guildford West (circled in blue) is not outside the regional norm.

3 DISCUSSION

Our findings suggest that none of the four elements we explored (library account creations, program registrations, service requests, and crime rates) are strongly correlated with EDI. This means that even if Avenues of Change did have some impact on the neighbourhood of Guildford West, this impact may not necessarily be measurable by their key indicator, EDI. These findings highlight the importance of choosing other suitable performance indicators and putting in place the infrastructure needed to collect that data in the community.

In addition, we may also glean meaning from the fact that the factors we explored do not seem to be related to EDI. For example, from our research, we can observe that our crime dataset does not meaningfully predict EDI in Surrey neighbourhoods. This is an interesting result since AoC identified safety as a key pathway to ensure long-term success. In their Theory of Change framework, the network identifies the importance of ensuring that the built environment is free from harm and that recreational areas are safe. Yet, our results show that even if crime diminishes, EDI would not necessarily see a proportional reduction. This may suggest that safety is not as important to early childhood development as initially perceived. However, it is much more likely that our crime dataset may not a perfect measure of safety within neighbourhoods. Rather, crime rates are the only measure that closely estimates this idea, and supplementary metrics are needed to quantify neighbourhood safety.

Besides collecting additional performance indicators, it is also critical to examine the current data in order to best direct efforts to reduce childhood vulnerability. The differing vulnerability scores across the 5 major sections may yield useful insights into where the most energy is required in order to decrease childhood vulnerability within a region. For example, as observed in Figure 13, emotional vulnerability scores in Guildford West are not outside of the regional norm. As such, it may not be a priority to address factors that relate to emotional vulnerability. On the other hand, given that language vulnerability is much higher in Guildford West compared to the rest of the region, this may be an excellent point of focus.

In order for such decisions to be made, effective network communication must take place. Previously, the Avenues of Change network had not yet been described as a graph, nor had it been quantitatively examined in order to understand its connectivity. Our analysis of the network achieved these goals, and highlighted some important pieces that may improve communication among members, as well as increase member buy-in. Many observations reflect anecdotal sentiments within the network, such as the observation on group redundancies at Early Years working groups and tables in Surrey. It should be noted, however, it is not always the same individual representing the same organisations at these various groups, but rather different people within the organisation that may or may not communicate with each other.

3.1 IMPACT FOR SOCIAL GOOD

Ultimately, the overarching goal of this project was to provide some benefit to society through the means of data. Of course, similar to many social good initiatives, large-scale impact is hardly ever straightforwardly measured. Our project is no different, despite our being partnered with an organization whose core mandate is a social good initiative in and of itself. Nonetheless, we believe that through our work we have identified some key areas of opportunity that will help push their mandate forward.

For example, our data collection suggestions will hopefully help catalyze a more data driven approach in the near future. In turn, this will help both Avenues of Change and the individual community members track meaningful changes within their society. Data collection is a difficult endeavour, which we experienced first-hand when trying to collect basic locational measures for some of our datasets, yet projects like these help set precedent for future data collection collaborations and contribute to a shift in culture regarding data within the organization. Additionally, through our research, we shed light to the nuance in large-scale initiatives that aim to drive systemic change. We believe, impact can be measured, but asking one metric, such as EDI, to account for all change will not work at the service of evaluating whether an initiative is working or not.

Finally, despite the various limitations in our analysis, we believe our findings will help understand complex social issues, such as interactions between partner organizations, in a more concrete way. We hope that in turn, this can allow for more cohesive efforts in reducing early childhood vulnerability in Guildford West and beyond.

3.2 LIMITATIONS

3.2.1 PREDICTION OF EDI

Limitations within our EDI analysis mostly revolved around the sparse number of observations we had, both spatially (24 neighbourhoods in Surrey) and temporally (at most 5 EDI observations across time). The lack of observations limited the number of variables that we could put into our potential predictive approaches. As a result, we discarded methods such as mixed effects model and random forest, which could have included more variables and better accounted for the complex relationship between the predictors and response variables. Instead, we opted for more straightforward models such as linear regression and logistic regression. Nevertheless, fitting these models leads to the poor predictive performance. Hence, we did not pursue the idea of analyzing the statistical significance of the parameters within our fitted model.

Additionally, although EDI is a robust, widely adopted measure of child development, there is a level of measurement error associated with it, especially relevant in our case when trying to calculate ‘critical differences’ in EDI over time. The measurement errors associated with identifying meaningful changes in vulnerability, especially within smaller neighbourhoods, is widely documented in the EDI literature [4], mostly related to uncertainty regarding the biases in children sampled, the way in which teachers interpret the survey, and neighbourhood size. Therefore, we must account for a degree of measurement error within our response variable.

While we were able to show that population density is currently a viable proxy for SES within Surrey, it is important to note that this may change with time. Currently, regions of high population density are associated with regions of lower SES, meaning that low SES neighbourhoods tend to contain more high-rises and apartments than higher SES neighbourhoods. However, periods on significant gentrification has the potential to eliminate the relationship between SES and population density. For this reason, we strongly advise that great caution be taken when using population density as a proxy measurement of SES. In the very least, we recommend that this relationship be re-examined with the release of each Canadian census (ie. every 5 years).

3.2.2 NETWORK ANALYSIS

With the strengths of our network analysis and tool come some important limitations, foremost of which is the completeness of the data. With respect to the graph built from the Surrey-Wide group membership data, it is critical to recognize that the data is official membership lists. The true engagement of individual agencies may vary, and the depth of involvement is unlikely to be equal across all members. Nonetheless, we have considered all connections to be of equal magnitude due to a lack of information. Nuances such as this, as well as whether or not information is passed on between different individuals who represent a single organisation, is not captured in much of our analyses. Gaining access to measures that quantify the depth of involved of each member of a table would greatly increase the graph's ability to represent the network structure of table membership.

With the analyses of the network built from survey data, the network built from our survey contains the responses of most, but not all, of the Avenues of Change members. As such, we cannot say with complete confidence that the analyses we have performed truly reflect the Avenues of Change network as a whole. In addition, there are many other agencies in Surrey working on early childhood development who were not surveyed because they were not a part of Avenues of Change Guildford West.

3.2.3 DASHBOARD

One limitation from our dashboard is the general interpretability of some of our visualizations. For example, our featured heatmaps allow for excellent comparison across neighborhoodsi for a particular year, but given the way the heatmap is rendered, comparison across years is not recommended, as the scales change to adjust to the data of a particular year.

3.3 FUTURE DIRECTIONS

One of the most important results stemming from our project is the pressing need to collect data for the purpose of analysis and evaluation. Therefore, we hope that future directions include a large data collection initiative encompassing two main categories: (1) measures for tracking impact in the community and (2) measures for understanding the evolving nature of the network structure. We provide a list of data collection suggestions in our additional information section. Once data is systematically collected, the opportunities for analysis will be significantly broadened.

Additionally, we hope that as data becomes available, it will be incorporated into our data analysis frameworks, such as PCA, and regression. Working with data collection and data analysis in tandem, rather than doing one after the other, will allow the data collectors to discard metrics that are not meaningful early on and realize if the measure that are relevant need further granularity.

In terms of the analysis we conducted, we also hope that the CLASS dataset, which was an extremely rich source of data, will be further analyzed and expanded on in order to answer questions related to City of Surrey's list of courses, where they are provided, and even predicting when a particular individual will stop taking courses. These questions were beyond the scope of our project, but we still believe they are important and can be partially answered with the data at hand.

A promising next step for our network analysis would be to expand the survey conducted to include all early years (and possibly also early-middle years) agencies in Surrey, not just members of Avenues of Change Guildford West. With a complete set of responses, we could be considerably more confident that the graph built from the survey responses quantifies the network structure of agencies within Surrey.

4 ADDITIONAL DETAILS

4.1 THE EARLY DEVELOPMENT INDICATOR (EDI)

Before the Avenues of Change initiative began, a measure of childhood readiness was necessary in order to ensure that efforts are focused where they are needed most. This measure needed to be broad enough to encapsulate the wide range of skills involved in school success - from communication skills and language, to physical strength and fine motor skills. This measure needed to be broad enough to encapsulate the wide range of skills involved in school success - from communication skills and language, to physical strength and fine motor skills. Such a measure was developed in the late 2000s by a group at McMaster University [2]. Named the Early Development Instrument (EDI), it is a broad, population-level measure of childhood readiness for school, when entering grade 1 at the age of 6 . The term population-level means that individual scores for each child are not particularly meaningful. Instead, looking at the aggregate scores for children within an area can give insight into the overall readiness of children for the start of their school lives. The EDI consists of 104 questions. Kindergarten teachers are asked each of the questions for each child in their class, and are asked to give children a score of 0, 5, or 10 for each questions. A "0" means that the skill is not present, a "5" means that the skill is developing, and a "10" means that the skill is well developed. The 104 questions touch on 5 major areas: Communication skills, Social skills, Physical ability, Language skills, and Emotional regulation. The total scores in each of these 5 major areas are calculating by adding together the a child's score for all questions that fit into each category. The total scores for the 5 areas, measured in childhood, have been shown to strongly predict health, education and social outcomes as adolescents and adults [3].

At a regional level, the EDI is reported as the proportion of children who are considered "vulnerable" within that region. In order to determine the score at which a child is considered vulnerable, an initial wave of data collection was performed from 2004-2006. In this wave, the 10th percentile of each major area was set to be the cutoff for vulnerability - If a child scores below that cutoff, they would be considered vulnerable. The proportions of vulnerable children are reported separately for each major area of measurement. In addition to each major area, an "overall" vulnerability is also reported, which gives the proportion of children who are vulnerable in one or more areas. Since the initial wave in 2004-2006, 4 additional waves of data collection have occurred, happening over each 3 year period.

4.2 WISHLIST / DATA COLLECTION SUGGESTIONS

- Avenues of Change
  - Collaborator Survey
    * **Implementation Strategy:** Distribute email survey with questions intending to measure the strength of the connections between collaborators (broader network members).
    * **Benefit:** Supplements anecdotal evidence and allows for concrete visualization of interconnectedness between a relatively complex set of individuals/groups.
    * **Example:** AoC Network survey and visualization.
  - Event Attendance
    * **Implementation Strategy:** In its simplest form, do a headcount of individuals attending an event/workshop at a particular date, especially if the nature of the workshop is ongoing. Given the opportunity, it would be beneficial to write additional details, such as: age, sex and other attendee characteristics that may be relevant.
    * **Benefit:** Will provide an initial benchmark to compare the workshop/event's improvement at a later date.
    * **Example:** Course attendance for CLASS data.
  - Parent Questionnaire
    * **Implementation Strategy:** Repeat the parent questionnaire frequently in order to keep a finger on the pulse of the community. Since it has already be written and prepared, the main task would be running a collection initiative on a regular basis. In fact, it may be beneficial to be continually collecting responses from the questionnaire.
    * **Benefit:** This will allow Avenues of Change to keep track of their impact in a continuous form. This way, they know as soon as possible when something is or isn't working.

- City of Surrey
  - Library Circulation Data
    * **Implementation Strategy:** It came to our attention during one of the AoC meetings that Surrey libraries have the ability to collect library circulation data but so far haven't done so. It would be ideal to encourage libraries in Surrey to systematically collect this information.
    * **Benefit:** Usage and circulation data would provide more meaning to library creation accounts as they would highlight (1) how much an account is actually being

used and (2) what kind of usage it's been given (e.g. educational v recreational purposes, children's or adult's books, etc.).

* **Example:** Library account creation visualization is a starting point.

– Health Data

* **Implementation Strategy:** There are two open source health databases from Fraser Health that can be explored: (1) Community Health Atlas (2) My Health My Community Atlas (MHMC atlas). Community Health Atlas has quantitative health variables but at the municipal level (not neighborhood level). The MHMC atlas has data at the neighborhood level but has different neighbourhood shapes and is conducted via a survey in 2013-14, but more geared towards the community.
* **Benefit:** Health data at a neighborhood level through time would presumably have a strong link with EDI, thus it can highlight areas of opportunity for service providers to contribute towards. Furthermore, it can help with the particular goal outlined in AoC's preliminary map of change, namely: 'Mothers are healthy and give birth to healthy infants who remain healthy.'
* **Example:** Health data can be incorporated into the PCA and regression analysis.

– Neighbourhood Environment and Greenspace

* **Implementation Strategy:** LIDAR data from the City of Surrey is already publicly available in their Surrey Open Data Catalogue, as well as various datasets on parks, park catch basins, trails, natural areas, outdoor recreation facilities, sport fields, trees, and more. Extensive data cleaning and integration would need to be performed on these various datasets.
* **Benefit:** It is possible that greenspace may be able to estimate social capital to some degree. In addition, greenspace may be associated with opportunities for children to play outside, which could be predictive of physical vulnerable scores within the EDI.
* **Example:** Greenspace data can be incorporated into the dashboard visualizations, as well as into the PCA and regression analysis.

5 References

[1] IBM's Smarter Cities Challenge. Ibm's smarter cities challenge: Surrey report. Technical report, IBM, July 2012.

[2] Magdalena Janus, Sally Brinkman, Eric Duku, Clyde Hertzman, Robert Santos, Mary Sayers, Joanne Schroeder, and Cindy Walsh. The early development instrument: a population-based measure for communities. *Offord Centre for Child Studies, McMaster University*, 2007.

[3] Paul Kershaw, Lynell Anderson, Bill Warburton, and Clyde Hertzman. 15 by 15: A comprehensive policy framework for early human capital investment in bc. Technical report, Human Early Learning Partnership, University of British Columbia, August 2009.

[4] Human Early Learning Partnership. Understanding critical difference in edi results. Technical report, The University of British Columbia, 2012.

[5] Human Early Learning Partnership. Edi [early years development instrument] british columbia provincial report. Technical report, University of British Columbia, School of Population and Public Health, October 2016.

[6] Human Early Learning Partnership. Edi (early years development instrument) report. wave 6 community profile, 2016. surrey (sd36). Technical report, University of British Columbia, School of Population and Public Health, October 2016.

[7] Simon Webb, Magdalena Janus, Eric Duku, Rob Raos, Marni Brownell, Barry Forer, Martin Guhn, and Nazeem Muhajarine. Neighbourhood socioeconomic status indices and early childhood development. *SSM - Population Health*, 3:48 – 56, 2017.