Evaluating Social Disparities in Autism Spectrum Disorder Diagnosis and Access to Care in Virginia

The Blue Backpackers
Amelia Doyle, Shane Hill, and Jorge Suárez-Marill

Executive Summary

We were tasked with uncovering any disparities in Autism Spectrum Disorder (ASD) coverage between urban and rural populations in the state of Virginia. Our sponsors wanted concrete data on if coverage in rural areas was lacking so that they could apply for a grant to address such an issue. We used drive-time analysis to calculate how long it would take to drive from a school that reported students with ASD, and then performed a regression analysis to see if there was a correlation between the drive time and whether an area was rural or urban. Our current conclusion is that there is no disparity, but there is future work that can be done to see if this is truly the case and why.

sponsored by
the Virginia Tech Center for Autism Research
December 13, 2017

1 Problem Statement

We have been tasked with discovering whether or not there are any disparities in Autism Spectrum Disorder (ASD) diagnosis and resource coverage between rural and urban areas in Virginia, particularly in the ease of access for an ASD patient to reach an Autism care center. If we conclude that such a disparity exists, then appropriate responses can be taken to alleviate issues. This is an important problem to investigate and address this issue as the Autism service cost of lifelong care can be reduced by two-thirds from $236-262 billion annually with early diagnosis and intervention [2].

Our sponsor, the Virginia Tech Center for Autism Research (VTCAR), sought us out because they asked the U.S. government to allocate more resources to serve those affected by ASD in rural Virginia, but they were turned down due to a lack of concrete evidence that such funding required. Our project will present this missing data and provide some justifications towards or against whether extra resources are needed in rural areas from which VTCAR plans to inform policy and apply for grants.

2 Ethical Considerations

One ethical concern of ours involves the model's validity. Our model could possibly be lacking in representation of the true ASD population. First of all, our sponsors asked us to focus on children with ASD, which comes at the expense of the adult ASD population. We may have missed out on a significant proportion of the ASD population as a result of this narrow focus, which would be unfortunate for the adults who need just as much help as the children. Another issue about the model's validity is that we can only use data that was a result of an official ASD diagnosis. Some parents may seek a diagnosis for their children if their children show any traits of ASD, whether it is because of a lack of knowledge on ASD, or it is a fear of what would happen if the child actually receives a diagnosis. We could have missed some ASD patients in our model as we can only work with official diagnosis data. Hopefully, our model includes enough people to be considered valid and that the missing people are not the majority of the ASD population.

The other ethical concern of ours is in regards to any consequences of policy decisions based on our model. Our sponsors asked us to collect data on ASD patients because they want a grant in order to improve services in rural areas. This means that our model will be shown to U.S. government officials as proof that there is a great disparity in ASD coverage between urban and rural areas. If our model is not satisfactory in the government's eyes or shows that no disparity exists, then our sponsors will be denied the grant. This could be disastrous for ASD patients who do have serious problems with reaching a service center as this grant denial can prevent new centers that would be closer from being built, preventing these patients from getting the intervention and help they need. While a good model that shows no disparity is not a terrible thing, a poor model that does not display a disparity when one exists can be catastrophic for the ASD population.

3 Literature Review

Several different organizations have done research in regards to ASD. The CDC has reported that the number of U.S. children with ASD has increased by 119.4% from the year

2000 to 2010 and that one in 68 children are diagnosed with ASD. The CDC also reported that Autism services cost U.S. citizens $236-262 billion annually, but they believe that the cost of lifelong care can be reduced by two-thirds with early diagnosis and intervention [2]. Our own sponsors have also done research on ASD, more specifically on the disparity problem between rural and urban populations in Virginia. They concluded that there was a disparity issue between two due to a number of problems such as the difficulty for rural families to reach centers and the lack of reliance on health care professionals [1]. Unfortunately, when they asked the U.S. government for a grant to address the issue, the government told our sponsors that their findings lacked concrete data and were too inconclusive for a grant to be issued.

4 Project Criteria

We divided our project into two components in order to organize our criteria better: data collection and data analytics. For data collection, our solution criteria is listed as follows in order from most important to least important:

1. *Content*: the topics and details the data provides
2. *Access*: how easily accessible the data actually is
3. *Data integration*: how well can the data work with other sets of data
4. *Credibility of data sources*: whether or not the data is trustworthy
5. *Data cleanliness*: how organized the data already is

We felt that the content of the data was the priority in regards to whether or not the data was worth using in our solution. Data is the backbone of our project, and a solution would be worthless if it does not include enough. We specifically focused on data that attached a geographical location to each school that reported ASD patients and ASD service centers in the dataset. We felt that access was also very important, as we needed to make sure the data is actually obtainable, as health care data is not easily accessible to the common person.

Similarly, for data analytics, we based a solution's value on the following criteria organized from most important to least important:

1. *Methods*: how efficiently the solution can analyze and determine social disparities of ASD between the two populations
2. *Software*: what software (Python, R, etc.) the solution can be used in
3. *Visualization*: how well the results can be displayed in a visual medium of some sorts
4. *Sophistication and scalability*: how well the solution will scale when the amount of data increases, and whether or not the data's integrity would be maintained

We deemed the actual methodology of the solution to be its most important aspect. We wanted to make sure that the solution actually helps answer the question of whether or not a disparity actually exists. The software the solution can be used in was second in priority as we wanted to make sure use something that can help us communicate our findings. Visualization was third in importance as a visual can help get findings across much easier than text would. Sophistication and scalability was last as we felt that most solutions we could have used would maintain the data's integrity.

5 Selected Solutions

As no datasets were readily available, web scraping was automatically identified as the primary method for data collection. Fortunately, Autism Speaks, an Autism advocacy group that sponsors Autism research and spreads awareness aimed at influencing public health policy, provides a list of Autism services and their locations by state. In order to gather this information into a desirable format for analysis, the Python package `Beautiful Soup` was used to scrape the locations of the services in Virginia from the Autism Speaks website.

Considering the spatial nature of our project objective, we immediately searched for a geographical solution. Originally, our plan was to evaluate service coverage for each ASD facility by creating Thiessen polygons, which determine geographical relations between points by defining areas around predetermined points such that everything located within that area is closest in proximity to that point. However, our strategy quickly shifted to a drive-time analysis approach as this would utilize road networks rather than elementary Euclidean distance, providing a much more realistic measure of access. The `Find Nearest` proximity tool was implemented in Esri's ArcGIS Online to compute the drive times to ASD early intervention centers and local Autism organizations from various rural and urban starting points. Furthermore, the `Create Drive Time Areas` proximity tool in ArcGIS Online was utilized to create drive-time areas around each service facility in order to visualize the areas within Virginia which can reach that facility within a specified driving time.

Regression analysis was then used to investigate whether the urban or rural classification of a point is a significant factor in its computed drive time. This was to find if there was a correlation between the time it took to reach a center and whether the population in question belonged to a rural area or an urban one. With the actual drive times and this regression analysis, we can deduce whether or not there is a disparity in ASD coverage between the two populations.

6 Obstacles

The largest obstacle we faced was the lack of readily available data. First of all, there was little data on ASD patients and even less for Virginia. This limited the number of resources we could collect data from. Another issue is that a lot of this data is considered health care data, which has many privacy protections in place to prevent the common person from accessing data on other people. Special care for this type of data had to be taken into consideration before explaining the results of our findings.

A smaller problem we had was figuring out the best way to analyze how difficult it is for someone to reach a service center. As discussed earlier, we were originally going to use a solution approach called Thiessen Polygons that only took geographic distance into account, but later switched to a solution that used drive-time analysis instead. This way, our model would have a better basis as the latter took the actual geography and road networks of Virginia into account.

7 Results

In this section, we will provide a detailed account of our project's results. Two analytical methods have been implemented with the task of identifying the existence of disparities in ASD access to care between rural and urban populations - geographic information systems (GIS) to compute drive times and linear regression to determine the significance of urban and rural classifications. For each approach, we will explain the significant outcomes we have obtained, what those outcomes mean in terms of our project objective, and any limitations to our results. The use of clear and captivating maps and plots will assist in visualizing the scope and impact of our project.

7.1 Spatial Analysis

The United States Census Bureau defines three classifications of geographical areas:

- *Urbanized Areas*: Areas with a population of 50,000 or more
- *Urban Clusters*: Areas with a population between 2,500 and 50,000
- *Rural*: All populations, housing, and territory not included within an urban area, where an urban area refers to either an urbanized area or urban cluster [6].

In compliance with the U.S. Census Bureau's definition of urban and rural, the Center for Rural Virginia has expanded these geographical classifications, placing all localities in Virginia into one of the following four categories:

- *Urban*: The locality's population density is at least 500 people per square mile, 90% of the locality population lives in urban areas, and the locality's population in urbanized areas is at least 50,000 or 90% of the population.
- *Rural*: The locality's population density is less than 500 people per square mile and 90% of the county population is in rural areas, or the county has no urban areas with a population of 10,000 or more.
- *Mixed Rural*: The locality meets neither urban nor rural criteria, and its population density is less than 320 people per square mile.
- *Mixed Urban*: The locality meets neither the urban nor rural criteria, and its population density is at least 320 people per square mile. [3]

Figure 1 shows that rural and mixed rural localities account for approximately 85% of the land area in the state, but only support one-third of the population [3]. As a result of this wide dispersal of people, rural communities are often disadvantaged in terms of geographic proximity to care and often face challenges such as poverty, inadequate transportation, social and demographic transitions, especially the out-migration of younger citizens, and unequal access to information technology.

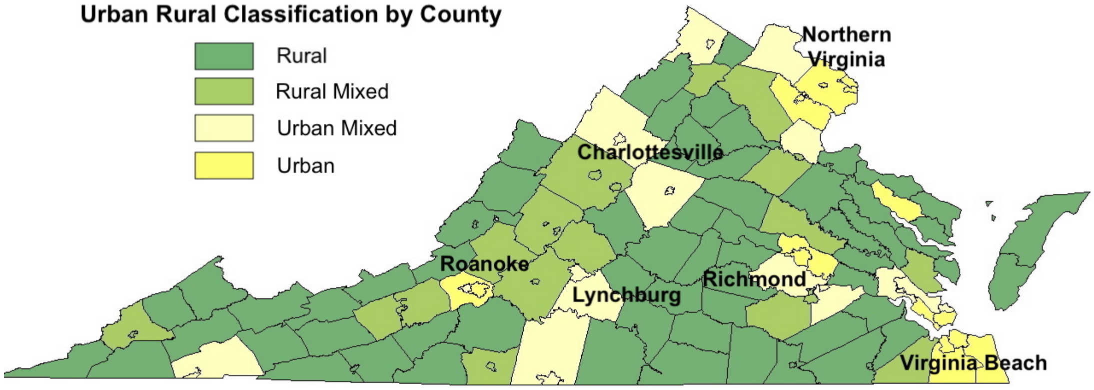

Figure 1: Urban and Rural Counties in Virginia (2010) [5]

According to the Centers for Disease Control and Prevention's (CDC) Autism and Developmental Disabilities Monitoring Network, about 1 in 68 children are diagnosed with Autism Spectrum Disorder [4]. It is useful to map the prevalence of ASD in Virginia compared to this national average. Because of data availability and confidentiality, our team's evaluations of ASD prevalence and access to care are strictly based on and limited to the public school system. We feel analysis at this scope will still yield appropriate findings as Autism is most commonly recognized and diagnosed in school-aged children.

The Virginia Department of Education's Fall Membership and Special Education Child Count reports contain data on the total number of students attending public schools in every county of Virginia. Furthermore, a count of children with various learning disabilities attending public schools is recorded for each county. From the 2016 data, our team has been able to gather the most recent county totals of children diagnosed specifically with ASD as well as calculate the percentage of students with ASD living in every county of Virginia. Overall, we were able to gather 778 observations of schools with students diagnosed with ASD. We calculated this percentage using the following formula:

$$\text{Percent ASD} = \frac{\text{Number of students with ASD}}{\text{Total number of students}}.$$

Figure 2 maps these percentages in relation to the national average (1 in 68). Visualizing the prevalence of ASD throughout Virginia will help us understand where ASD diagnoses are distributed and assess the regions that are in most need of relief-providing resources and governmental assistance. Of all the localities with a percentage of Autistic children above or equal to the national average, about 55% are classified as rural or mixed rural. This statistic already validates the importance of studying access to quality care in rural communities as there is an apparent need for such care.

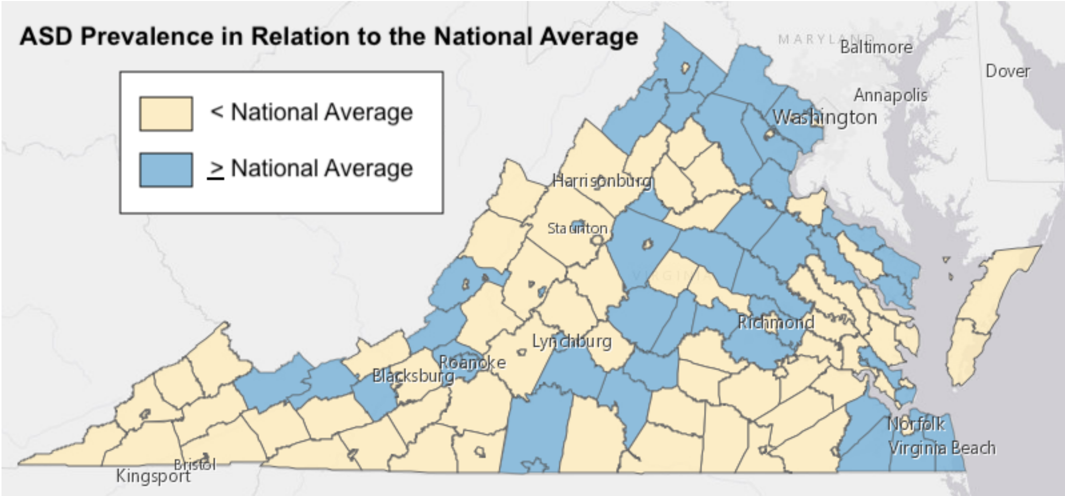

Figure 2: ASD Prevalence in Virginia (2016)

Our team decided that the most robust way to determine a population's access to a given service is through travel time. The `Find Nearest` proximity tool offered by Esri's ArcGIS online platform can be utilized to calculate drive times from a source to a destination point. For the purpose of this project, drive-time analysis will be useful in determining whether or not rural populations across Virginia affected by ASD are underserved in terms of their access to medical resources, particularly specialty care and local support organizations. If this is the case, we can expect rural localities to be associated with longer drive times. Determining the significance of an urban or rural classification in the drive time calculation will be assessed in the next section covering regression analysis.

The Virginia public schools were assigned as the source points in the drive-time analysis, the locations at which the drive time calculation starts. Only the public schools in which Autism-diagnosed children are currently enrolled were considered. Data was collected from the 2016 Virginia Department of Education's December 1 Special Education Child Count report. Private school data was not included.

To avoid limiting the results of our project to a single service type, we decided to run two separate drive-time analyses - one for early intervention centers and another for local Autism organizations. Considering facility expertise allows us to evaluate and compare access to different kinds of services in hope that the results reveal patterns or trends that might otherwise have remained hidden. Perhaps rural areas are lacking access to early intervention care but not to local support organizations, for example.

Early intervention services focus solely on the initial diagnosis and testing for ASD and primarily treat young children ages two to five. In comparison, local Autism organizations can serve Autism patients and their families at different age ranges and stages of treatment and provide a variety of professional support services, such as applied behavioral analysis (ABA) therapy, assessment-based educational treatment plans, research-based behavior intervention plans, and Individualized Education Program (IEP) reviews. The locations of the early intervention centers and local Autism organizations in Virginia were scraped from the Autism Speaks website using the `Beautiful Soup` package in Python.

The built-in geocoder in ArcGIS allowed us to batch upload the addresses of the schools and service centers and map them in relation to each other. Running the `Find Nearest` proximity tool produces drive times, drive distances, and the corresponding driving routes using established road networks from each public school to its nearest early intervention center (see Figure 3) and local Autism organization (see Figure 4). Traffic and time of day spent traveling was not considered in the drive time calculations.

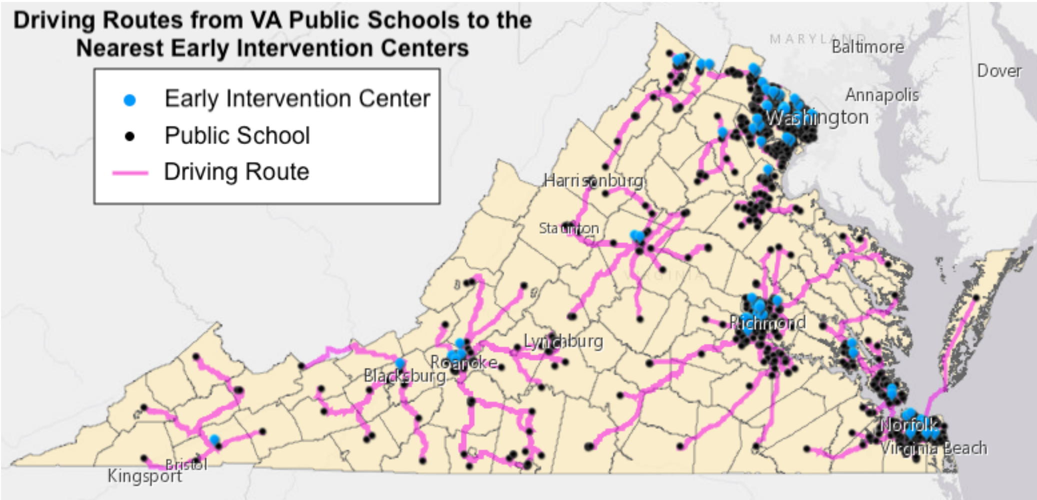

Figure 3: Nearest Early Intervention Centers to Schools

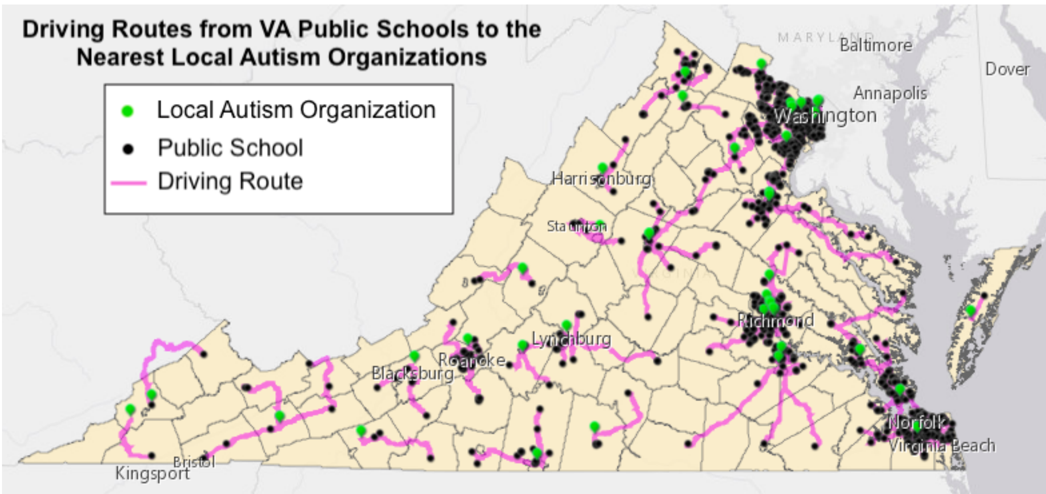

Figure 4: Nearest Local Autism Organizations to Schools

From these maps, we can already see that ASD early intervention centers, in particular, and local Autism organizations are clustered in large, urban cities. In fact, Table 1 summarizes the average time it takes to drive from a school to the nearest service center based on if the school is in an urban or rural locality.

Table 1: Average Drive Times (minutes) by Service Type

	Service Type	
	Early Intervention Center	Local Autism Organization
Urban	14.69	16.54
Rural	33.76	26.24

Splitting the spatial analysis results into two categories - urban and rural - has uncovered evidence to support a potentially significant pattern. For both early intervention centers and local Autism organizations, the driving time from a public school to its closest ASD care center was, on average, longer for schools located in rural areas.

In order to gain a more complete visualization of the drive-time analysis results, we implemented the `Create Drive Time Areas` proximity tool in ArcGIS Online, which creates drive time buffers around a centralized point to describe areas that can be reached within a specified driving time. With these drive time areas, we can identify from which regions in Virginia a person would have to drive the longest to reach a service center.

For example, in Figure 5, we can conclude that the southern border of Virginia has poor access to Autism early intervention services as the travel time from those counties is greater than two hours. Thinking back to our project's overall objective, it is important to point out that the majority of those counties are rural (see Figure 1). By simply allocating resources to add a new early intervention center in the heart of this "red zone", quality care would become much more accessible throughout all of Virginia and not just the urban cities and their outskirts. However, recall that data was only collected for Virginia, so these results may be limited as they do not consider access to services that may sit on the other side of the state borders.

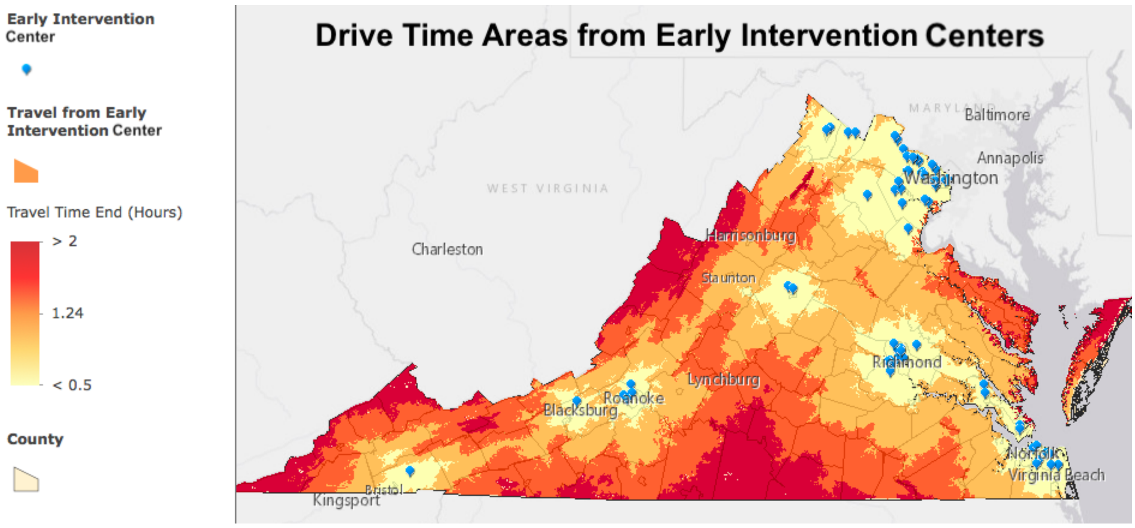

Figure 5: Drive Time Areas from Early Intervention Centers

Figure 6 tells us that access to local Autism organizations is much more distributed across Virginia. Despite the increase in accessibility, there are still apparent "red zones" where travel time is expected to take over one hour and a half. Again, these "red zones" correspond to the rural areas of Virginia. However, to reiterate the limitation expressed above, this analysis did not take into account possible Autism service center locations in the neighboring states. Additionally, it may also be important to note that, although similar spatial analysis techniques could be applied to various other disabilities or diseases, the analysis described in this report is only valid for evaluating ASD and should not be used as justification in other studies of any other medical condition.

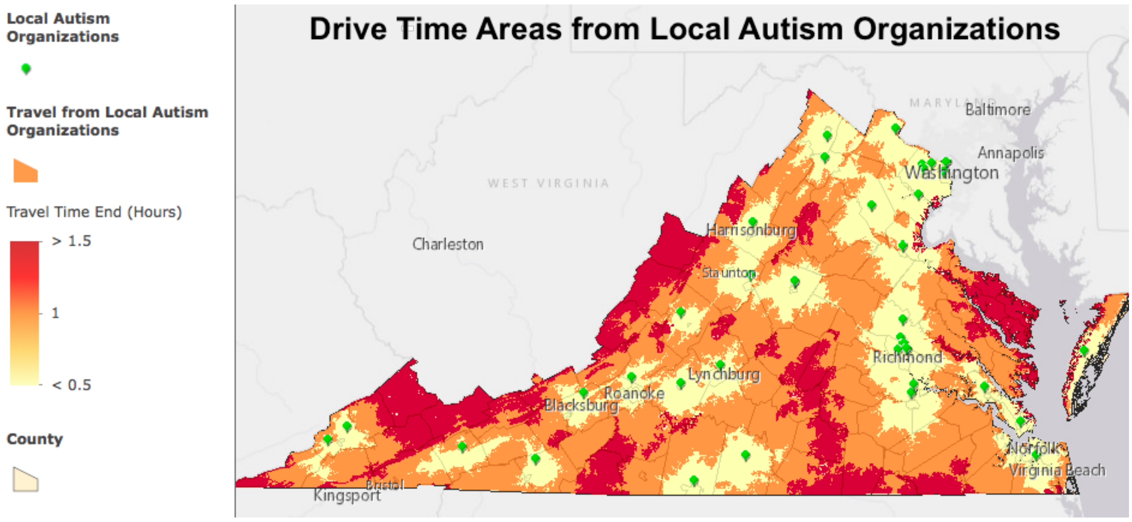

Figure 6: Drive Time Areas from Local Autism Organizations

By performing robust techniques for spatial analysis on the ASD data provided by the Virginia Department of Education and Autism Speaks, we can walk away from this semester having answered several questions our project set out to find. (1) There is evidence that quality care for those affected by Autism is needed as areas throughout Virginia have a prevalence at or above the national average. (2) Access to specialty care for early intervention is particularly scarce. (3) Regions associated with longer travel times tend to be located in rural areas. From this initial investigation, we suspect urbanity and rurality will be a significant factor in evaluating how much access a particular community or population has to a desired service. To test this hypothesis, our team has conducted a regression analysis using the drive-times calculated in ArcGIS, the definitions of urban and rural, and the aforementioned ASD data.

7.2 Regression Analysis

When originally brainstorming our preferred solutions, our team had decided that a regression analysis would most likely be irrelevant as a solution to our problem. However, after obtaining useful data sets, we saw that there could in fact be a significance between drive time and other useful factors represented in these data sets. Since our client is particularly interested in adolescents diagnosed with ASD, we decided it would be best to build two separate regression models with one relating to Early Intervention Centers and another for Local Autism Organizations.

The data set we constructed included the factors:

- `School.Name`: The name of the school where students diagnosed with ASD were observed.
- `VDOE.School.Code`: A number designated as an identifier for a school by the Virginia Department of Education.
- `Division.Name`: The division in which the school is located.
- `VDOE.Division.Number`: A number designated as an identifier for a division by the Virginia Department of Education.

- `Address`: The street address representing the school's geographical location.
- `City`: The city in which the school is located.
- `State`: The state in which the school is located. (*Note: All observations for this study are located in Virginia. We included this factor in hopes of expanding the analysis to a nationwide level.*)
- `Zip`: The postal zip code corresponding to where the school is located.
- `Level`: Designates which level of education the school provides (Elementary, Middle, or High).
- `Census.Tract`: The United States Census Tract number in which the school is located.
- `Rural`: A binary representation of whether the school is located in a Rural area.

  `1`: School is located in a rural area.

  `0`: School is not located in a rural area; e.g. Urban, Suburban, or Urban Cluster.
- `ASD.Count`: The number of students in the school diagnosed with ASD as of December 2016.
- `Num.Students`: The total number students enrolled in the school as of December 2016.
- `ASD.Prevalence`: The proportion of students diagnosed with ASD in relation to the total number of students.
- `Min.Travel.Time.EIC`: Drive time, in minutes, from the observed school to the nearest Early Intervention Center using road networks.
- `Travel.Dist.EIC`: Distance, in miles, from the observed school to the nearest Early Intervention Center using road networks.
- `Min.Travel.Time.LAO`: Drive time, in minutes, from the observed school to the nearest Local Autism Organization using road networks.
- `Travel.Dist.LAO`: Distance, in miles, from the observed school to the nearest Local Autism Organization using road networks.

Before running the analysis, we wanted to visually see if rurality had an overall impact on drive time to the nearest service center.

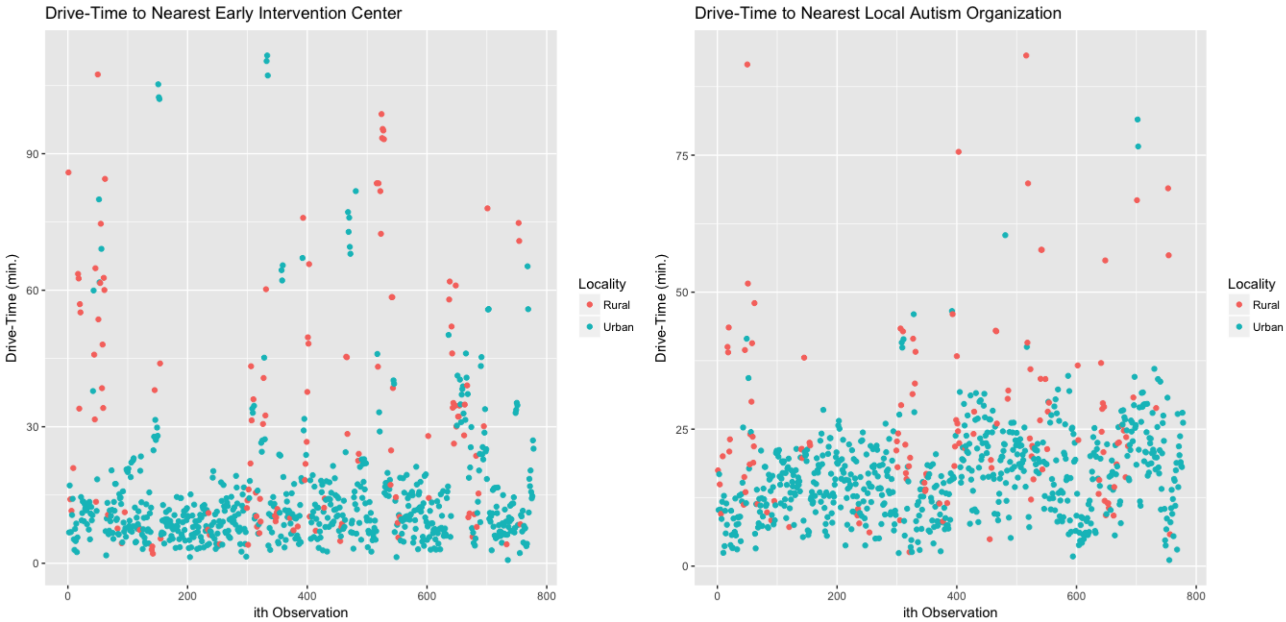

Figure 7: Plot of Drive Time for Early Intervention Centers and Local Autism Organizations

Plotting the drive-times and denoting rural and urban areas, it appears that more rural areas (represented as red points) are associated with longer drive times than urban areas (represented as blue points). That is, we see that when drive time increases, more points are located in rural areas. In both graphs, we see that the majority of urban locations have less than a 30 minute drive time to the nearest service center. Understanding the results of these graphs helps to show that rural areas have a disadvantage when it comes to access to care.

7.2.1 Early Intervention Centers

To conduct a regression analysis, our team sought to identify whether drive time to the nearest Early Intervention Center was significantly influenced by rurality, ASD prevalence, total number of students, and the distance to the respective Early Intervention center. Thus, our regression model was represented as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

where:

y: `Min.Travel.Time.EIC`

x_1: `Rural`

x_2: `ASD.Prevalence`

x_3: `Num.Students`

x_4: `Travel.Dist.EIC`

Before relating the results of our model to our problem, our team checked the model assumptions that residuals of the proposed model (1) are independent and normally distributed and (2) have constant variance and a mean of 0. These assumptions can be best expressed as $\varepsilon_i \overset{IID}{\sim} N(0, \sigma^2)$.

To evaluate if the residuals are independent and normally distributed, we look to the Normal Quantile-Quantile (Q-Q) Plot and the Histogram for the Studentized Residuals of the model.

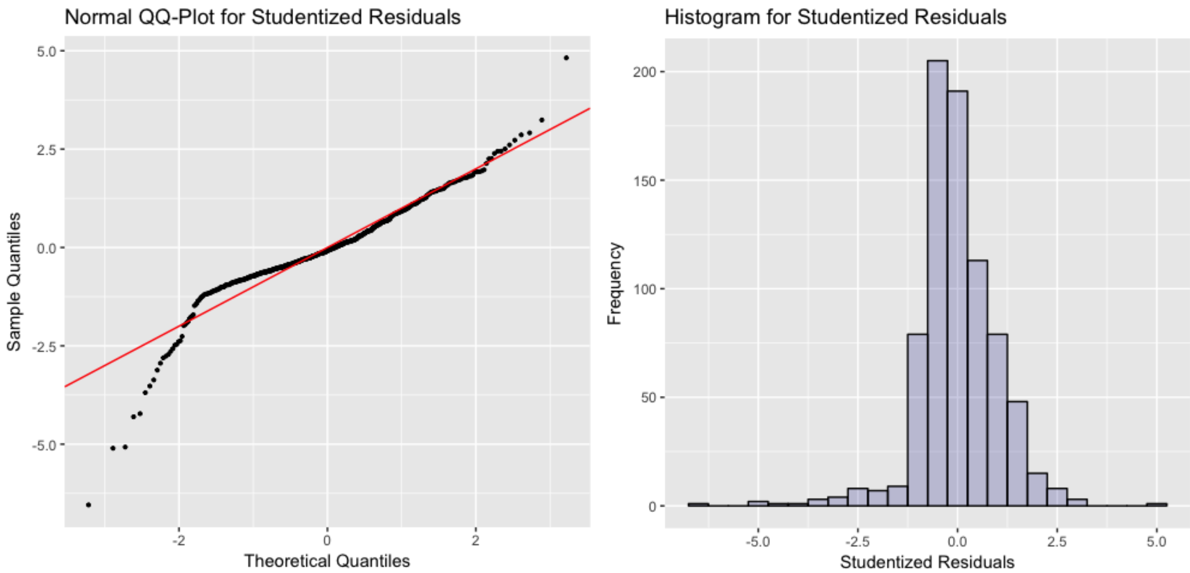

Figure 8: Normal Q-Q Plot and Histogram for Early Intervention Centers Model

As we see, the residuals are approximately normal as they lie relatively close to the line in the Normal Q-Q Plot. The "tails" of the line are most likely attributed to outliers in our data. Taking into consideration the context of our problem that rural areas have a lack of access to care to service centers, the concern for these outliers is low due to rural areas having longer drive times to the nearest service center than urban areas. From the Histogram, we see a relatively normal approximation that follows a bell curve. Overall, there is reason to believe that the residuals for this model are independent and normally distributed.

To determine if residuals have constant variance and a mean of zero, we examine the Residuals vs. Fit plot.

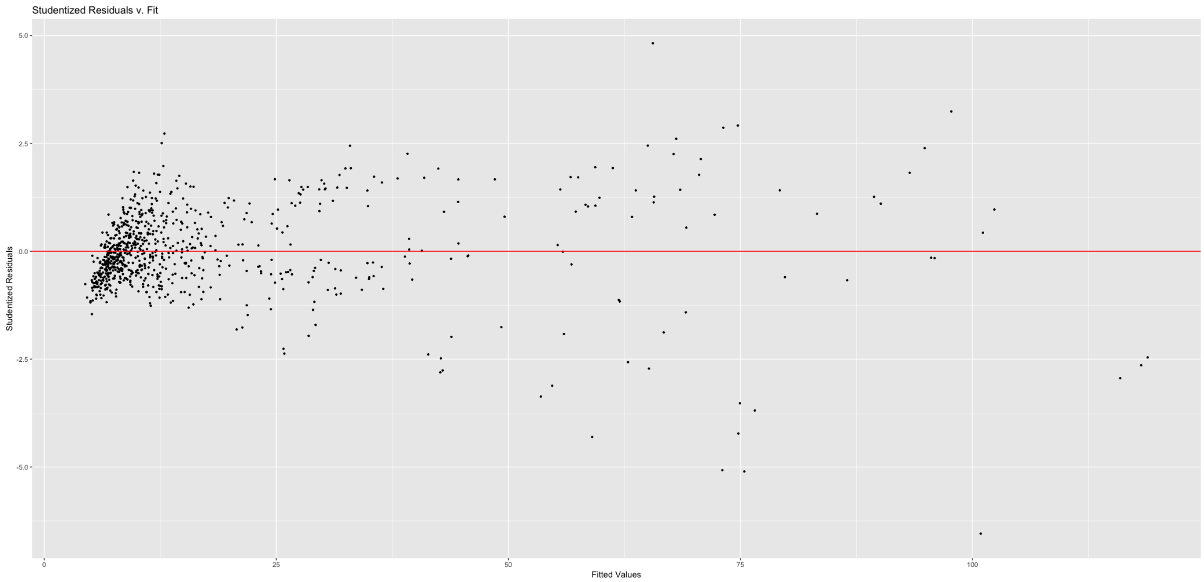

Figure 9: Residuals v. Fit for Early Intervention Center Model

Using software, we calculated the mean of the Studentized Residuals as -0.0005943622, which is approximately equal to zero. From the Residuals v. Fit plot, there are key points to note. Overall, the residuals tend to be centered around the mean. However, there seems to be a bunching of residuals towards lower fitted values. This would suggest that a transformation of the response would help to scatter the residuals into a more favorable random pattern. Another key observation is a "trumpeting effect" of the residuals; in other words, as fitted values increase, residuals increase further from the mean. This effect would suggest that a transformation of one or more of the predictors would help to create the favorable constant width across the residuals.

Following these suggestions, transforming our response variable could ultimately change the whole interpretation of our model. We strayed away from this proposal as we wanted to keep a linear model. We did, however, attempt to transform the predictors to obtain a better Residuals v. Fit. Log-transformations, interactions, and including new predictors failed to provide a significant change in the Residuals v. Fit.

Similarly to observing the Normal Q-Q Plot and Histogram, we took into account the potential for the effect of outliers on our model. When detecting outliers, a good rule-of-thumb when interpreting the Residuals v. Fit plot is to look for observations where $|r_i| > 2$. Noting this, there are 33 observations that match this criteria. These 33 observations relate to longer drive times in rural areas. Though deleting these points would essentially give us a more desirable plot, we decided to keep them given the context of our overall problem. Thus, we attributed that the long drive times for rural areas are most likely the cause for the skewed Residuals v. Fit plot and continued with our analysis.

After reviewing and verifying the model assumptions, we were able to summarize the results of our model. Using R, we obtained the following summary from running our regression model.

Listing 1: Summary of Early Intervention Center Model

```
Call:
lm(formula = Min.Travel.Time.EIC ~ Rural + ASD.Prevalence + Num.Students +
    Travel.Dist.EIC, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-19.1078  -1.6104  -0.2589   1.7308  14.3963

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.4766874  0.3903149  11.469   <2e-16 ***
Rural           -0.1177554  0.3133282  -0.376    0.707
ASD.Prevalence  -5.4408493  5.1153874  -1.064    0.288
Num.Students     0.0002138  0.0002368   0.903    0.367
Travel.Dist.EIC  1.1392326  0.0072496 157.144   <2e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Residual standard error: 3.05 on 773 degrees of freedom
Multiple R-squared:  0.9744,    Adjusted R-squared:  0.9743
F-statistic:  7356 on 4 and 773 DF,  p-value: < 2.2e-16
```

Conducting an overall F-Test helped us to understand whether the predictors used in this model are significant in predicting drive-time to the nearest Early Intervention Center. Following the format for an overall F-Test we formulated the hypotheses:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_a: at least one $\beta_j \neq 0$ for some j

assuming $\varepsilon_i \overset{IID}{\sim} N(0, \sigma^2)$ and $\alpha = 0.05$

From the output we saw an F-statistic = 7356, and an associated *p-value* of $< 2.2e-16$. Comparing this to the proposed $\alpha = 0.05$, we rejected H_0, since the *p-value* is significantly less than α. Thus, we concluded that `Rural`, `ASD.Prevalence`, `Num.Students`, and `Travel.Dist.EIC` are useful for predicting `Min.Travel.Time.EIC`.

Another key observation is `Multiple R-Squared = 0.9744`. This statistic tells us that 97.44% percent of variability in `Min.Travel.Time.EIC` is explained by the predictors present in the model. Continuing noting key observations, we also determined whether each variable is individually significant to the model. Following the format for Individual t-Tests on the predictors we have for each predictor of the hypotheses:

H_0: $\beta_k = 0$ vs. $H_a : \beta_k \neq 0$

assuming $\varepsilon_i \overset{IID}{\sim} N(0, \sigma^2)$ and $\alpha = 0.05$

For clarity, it's best to represent each test on the predictors in a table:

Variable	t-value	p-value	Conclusion
`Rural`	`-0.376`	`0.707`	Fail to reject H_0 since $p > \alpha$. There is not enough sufficient evidence to suggest that `Rural` significantly contributes to `Min.Travel.Time.EIC`.
`ASD.Prevalence`	`-1.064`	`0.288`	Fail to reject H_0 since $p > \alpha$. There is not enough sufficient evidence to suggest that `ASD.Prevalence` significantly contributes to `Min.Travel.Time.EIC`.
`Num.Students`	`0.903`	`0.367`	Fail to reject H_0 since $p > \alpha$. There is not enough sufficient evidence to suggest that `Num.Students` significantly contributes to `Min.Travel.Time.EIC`.
`Travel.Dist.EIC`	`157.144`	`<2e-16`	Reject H_0, since $p < \alpha$. `Travel.Dist.EIC` significantly contributes to `Min.Travel.Time.EIC` after we have accounted for `Rural`, `ASD.Prevalence`, and `Num.Students`.

Table 2: Individual t-Tests for Early Intervention Center Model

From these tests, we saw that only one predictor, `Travel.Dist.EIC`, is significant. Looking back at the output, we can see that for every one mile driven, drive time increases by approximately 1.139 minutes. However, we also saw that `Rural` - the key predictor we wished to relate - was not only not significant, but also had a negative coefficient on its estimate. That is, if a school is in a rural location, drive time *decreases* by approximately 0.118 minutes, or approximately seven seconds. This revelation failed to match up with all of our previously noted conclusions that a rural location should increase drive time. That is, rural locations lack appropriate service centers and people in these locations have to drive further to receive care.

7.2.2 Local Autism Organizations

Following the same format as the Early Intervention Center model, our team sought to identify whether drive time to the nearest Local Autism Organization was significantly influenced by rurality, ASD prevalence, total number of students, and the distance to the respective Local Autism Organization. Thus, our regression model was represented as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

where:

y: `Min.Travel.Time.LAO`

x_1: `Rural`

x_2: `ASD.Prevalence`

x_3: `Num.Students`

x_4: `Travel.Dist.LAO`

Similarly, we checked the assumptions that the residuals for our model are independent, normally distributed, have a mean of zero, and have constant variance. Checking independence and normality, we, again, utilized the Normal Q-Q Plot and Histogram for the studentized residuals.

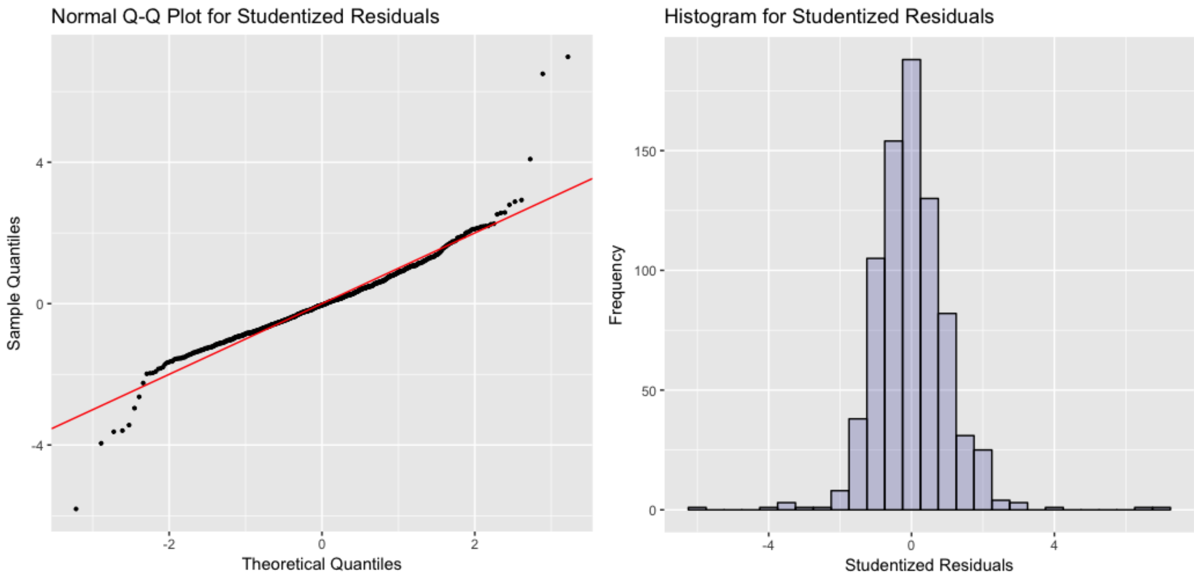

Figure 10: Normal Q-Q Plot and Histogram for Local Autism Organization Model

The Normal Q-Q Plot for this model follows more of a normal distribution than the Early Intervention Center Model. While we still have deviating “tails” of the line, we know that outliers due to longer drive times in rural areas are the leading cause to these deviations. As previously mentioned, we sought to keep these outliers since we know that people in rural areas often have to drive longer distances for care. From the Histogram, we see that the the peak is centered exactly on zero and follow a desirable bell curve. These two plots help to verify that the residuals are, in fact, independent and normally distributed for the Local Autism Organization model.

Next, we checked the Residuals v. Fit plot to verify that the residuals for the model have constant variance and are centered around a mean equal to zero.

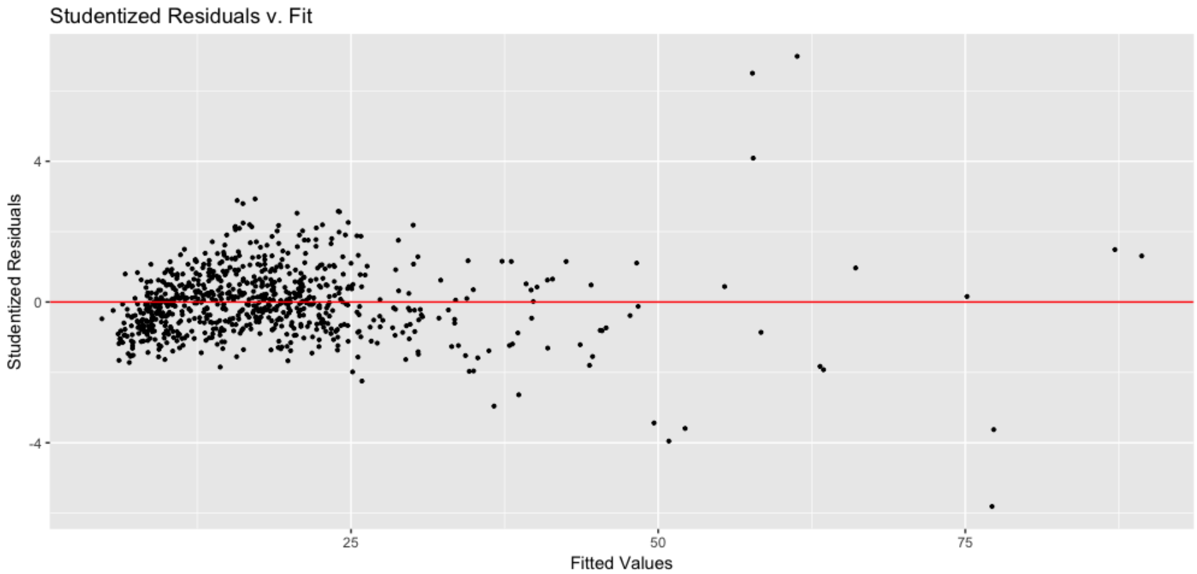

Figure 11: Residuals v. Fit for Local Autism Organization Model

Again, using software, we were able to calculate that the mean of the residuals is equal to 0.000528853, which is approximately equal to zero. While the residuals in the Residuals v. Fit plot are still bunched near lower fitted values, we see more of a dispersal of residuals across the fitted values than the Early Intervention Center model. More so, we fail to see the "trumpeting" effect that the Early Intervention Center model displayed. However, there could be a concern for potential outliers as there are 31 observations where $|r_i| > 2$. In the context of our problem, these observations are most likely the long drive time from rural areas where service centers are not locally present. Similarly to the Early Intervention Model, we kept these observations in our data set as they are relevant to the overall problem of lack of access to care in rural areas. As mentioned before, we attempted to transform predictor variables, but no transformation significantly made a difference in the residuals. Therefore, we kept the original model parameters with no transformations.

Overall, the assumptions for this model seem to be better verified than the Early Intervention Center model. Given more time on this project, we could have completed more analyses to determine why the two models have varying residual analyses.

After reviewing and verifying the model assumptions, we were able to summarize the results of our Local Autism Organization model. Using R, we obtained the following summary from running our regression model.

Listing 2: Summary of Local Autism Organization Model

```
Call:
lm(formula = Min.Travel.Time.LAO ~ Rural + ASD.Prevalence + Num.Students +
    Travel.Dist.LAO, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-16.7748  -1.8851  -0.1199   1.6316  20.1929

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.783e+00  3.988e-01  14.501  < 2e-16 ***
Rural           -1.304e+00  3.121e-01  -4.179 3.26e-05 ***
ASD.Prevalence  -1.496e+01  5.074e+00  -2.949  0.00329 **
Num.Students     4.346e-04  2.341e-04   1.857  0.06371 .
Travel.Dist.LAO  1.142e+00  1.263e-02  90.484  < 2e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Residual standard error: 3.026 on 773 degrees of freedom
Multiple R-squared:  0.9235,    Adjusted R-squared:  0.9232
F-statistic:  2334 on 4 and 773 DF,  p-value: < 2.2e-16
```

As in the Early Intervention Center model, we conducted an overall F-test for this model, as well, with the hypotheses:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_a: at least one $\beta_j \neq 0$ for some j

assuming $\varepsilon_i \overset{IID}{\sim} N(0, \sigma^2)$ and $\alpha = 0.05$

Using the output, we see an F-statistic $= 2334$, and an associated *p-value* of $< 2.2e-16$. Comparing this to the proposed $\alpha = 0.05$, we rejected H_0, since the *p-value* is significantly less than α. Thus, we concluded that `Rural`, `ASD.Prevalence`, `Num.Students`, and `Travel.Dist.LAO` are useful for predicting `Min.Travel.Time.LAO`.

In this model, we saw a `Multiple R-Squared = 0.9235`. Thus, 92.35% percent of variability in `Min.Travel.Time.LAO` is explained by the predictors present in the model.

Continuing the analysis, we tested each predictor using Individual t-Tests to determine if each is significant for predicting `Min.Travel.Time.LAO`, with the hypotheses:

H_0: $\beta_k = 0$ vs. $H_a : \beta_k \neq 0$

assuming $\varepsilon_i \overset{IID}{\sim} N(0, \sigma^2)$ and $\alpha = 0.05$

Again, one can better interpret the results of these tests as a table:

Variable	t-value	p-value	Conclusion
`Rural`	`-4.179`	`3.26e-05`	Reject H_0 since $p << \alpha$. `Rural` significantly contributes to `Min.Travel.Time.LAO` after accounting for `ASD.Prevalence`, `Num.Students`, and `Travel.Dist.LAO`.
`ASD.Prevalence`	`-2.949`	`0.00329`	Reject H_0 since $p < \alpha$. `ASD.Prevalence` significantly contributes to `Min.Travel.Time.LAO` after accounting for all other predictors.
`Num.Students`	`1.857`	`0.06371`	Fail to reject H_0 since $p > \alpha$. There is not enough sufficient evidence to suggest that `Num.Students` significantly contributes to `Min.Travel.Time.LAO`.
`Travel.Dist.LAO`	`90.484`	`<2e-16`	Reject H_0, since $p << \alpha$. `Travel.Dist.LAO` significantly contributes to `Min.Travel.Time.LAO` after accounting for all other predictors.

Table 3: Individual t-Tests for Local Autism Organization Model

Contrary to the previous model, three out of the four predictors are significant in predicting `Min.Travel.Time.LAO`: `Rural`, `ASD.Prevalence`, and `Travel.Dist.LAO`. From the summary output, `Rural` has an estimate of -1.304. In terms of this model, if an observation is located in a rural area, drive team will *decrease* by 1.304 minutes. Again, we did not see an affirmative conclusion in this model either that rurality would increase drive time like originally proposed. More so, we see that `ASD.Prevalence` has an estimate of -1.496e+01. For every 1% increase in ASD prevalence, drive time decreases by approximately 14.96 minutes. This observation makes sense in terms of our problem as we would normally see a higher prevalence of ASD diagnosis in urban areas. Prevalence is likely to increase due to more service locations in an area. Lastly, `Travel.Dist.LAO` has an estimate of 1.142. This observation makes sense, as well, that for every one mile driven, drive time increases by 1.142 minutes.

7.2.3 Regression Analysis Conclusion

Overall, these two regression models have provided some key insights into drive time to service centers from rural and urban areas. The Early Intervention Model had rurality as insignificant, while the Local Autism Organization model had rurality as very significant. Regardless, rurality was kept in both models as it is one of the key indicators we were looking to examine. Moreover, we have seen in both models that rurality ends up *decreasing* drive time; we had originally concluded based on other observations that it should increase drive time. Given more time on this project, our team could have conducted more tests and collected more data to understand why rurality in these models seem to decrease drive-time.

8 Division of Labor

Amelia took a prominent role in the spatial component of the project. Given her previous experience working with geographic information systems (GIS), she was responsible for gathering and cleaning the information pertaining to the locations of the ASD early intervention centers, local Autism organizations, and Virginia public schools and for conducting the drive-time analysis of the data in Esri's ArcGIS Online suite.

Shane took lead in conducting all of the Regression Analysis for the project. Given four differing data sets, he helped to clean and solely consolidated all useful information from each into one final universal data set on which Regression Analysis was run.

Jorge was the lead in communicating ideas to other parties and research into the topic and possible solutions. He also helped with cleaning so that the data could be inserted into Amelia's GIS solution.

9 Future Work

Due to data limitations and time constraints, the scope of our project was strictly limited to those Virginia census tracts containing the locations of public schools that currently have enrollment of students diagnosed with ASD. If this project were to be expanded, the next steps would be to perform similar spatial and regression analyses on every census tract in Virginia in order to gain a more complete insight into the current situation of the state. Similarly, additional analyses could be conducted to determine what factors cause rurality to decrease drive time in our study. Would combining all the types of service locations into one data set result in a model where rurality increases drive time as originally proposed? Questions such as this are important to understanding the lack of access to care that is explicitly prevalent in rural Virginia.

Additionally, given more time, the following questions that VTCAR were also interested in answering could be addressed.

- How does socioeconomic status impact ASD service access and delivery across rural and urban areas?
- Do rates of governmental assistance differ for ASD across rural and urban areas, and also compared to other developmental disabilities?
- Does the level of parental educational attainment play a role in ASD diagnosis or under-diagnosis of school-aged children?

One could even build on our research to study the impacts of ASD beyond Virginia's state borders. The prevalence of ASD and access to various care specializations, early intervention and other service centers in rural and urban localities could be analyzed and compared throughout the entire eastern region of the United States. Improving human health is one of the most challenging policy goals facing the U.S. today. Despite expenditures in the trillions of dollars, U.S. health outcomes continue to fall behind those in other developed countries. Access to care remains a significant problem for millions of citizens, including children and the elderly. Moreover, access is only one of many determinants of health. Indeed, a broad range of social, economic, and environmental factors shape individuals' and communities' opportunities and barriers to health. These "social determinants of health" must be addressed if the U.S. is to achieve greater health equity by reducing the persistent disparities in health outcomes that adversely affect particular groups of people.

References

[1] Antezana, L., Scarpa, A., Valdespino, A., Albright, J., & Richey, J. A. 2017 "Rural Trends in Diagnosis and Services for Autism Spectrum Disorder". *Frontiers in Psychology*, 8, 590. http://doi.org/10.3389/fpsyg.2017.00590

[2] Autism Society. August 2015. *Facts and Statistics.* http://www.autism-society.org/what-is/facts-and-statistics/

[3] Center for Rural Virginia. *Rural Virginia Prosperity Commission Data Update.* Richmond, Virginia, 2010.

[4] Centers for Disease Control and Prevention. *Autism Spectrum Disorder (ASD) Data & Statistics.* https://www.cdc.gov/ncbddd/autism/data.html

[5] Copenhaver, B., Elmer, M., Gaffney, R., Shaw, C., Alwang, J. *The Twin-Transfer Squeeze on Rural Households and Governments in Virginia.* Agricultural and Applied Economics Association, Choices Magazine, 3rd Q. Vol. 27(3), 2012.

[6] United States Census Bureau. *Urban and Rural.* https://www.census.gov/geo/reference/urban-rural.html