Work Examples:

Data Visualization: An extension of the dissertation with support from the UNT Willis Library Dean's Innovation Grant is the newer dashboard with updated datasets and public library information to evaluate the state of digital connectivity across the U.S.
https://public.tableau.com/app/profile/cary.k.jim/viz/ALAConferenceJune2022/Map1

Programming with R or Python for Data Analysis:
I've completed and presented an analytics project as part of the participation of the IES sponsored training of using PIAAC dataset. This work is co-presented with another PhD student at the 2020 AERA Satellite Conference at Stanford University – Conference on Educational Data Science.

I developed a Python and a R version of this analysis. They can be viewed in the following repository.

- Exploratory Analysis of PIAAC in R: https://rpubs.com/caryjim/653859
- Exploratory Analysis of PIAAC in Python https://github.com/caryjim/PIAAC-Exploratory-Cluster-Analysis/blob/master/PIAAC_Cluster_Analysis_Final.ipynb

Students Digital Opportunity: Conceptualization, Analysis, and Implications

Candidate for

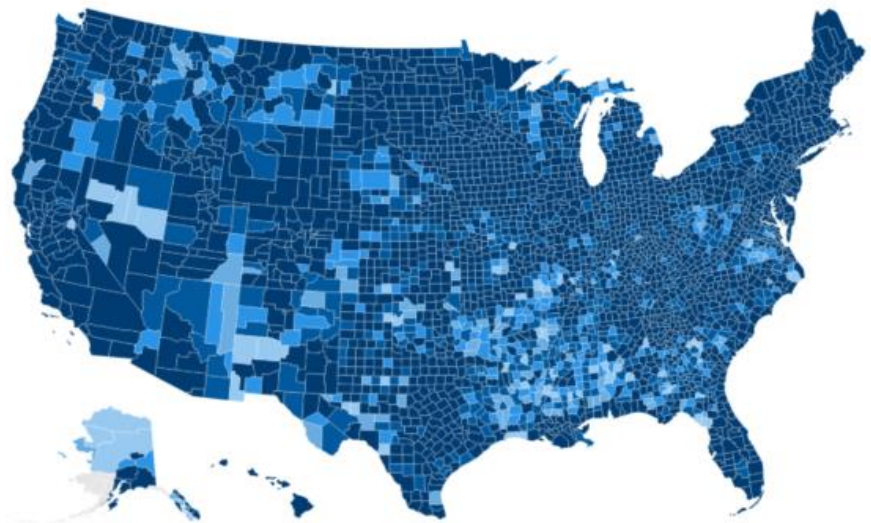
Senior Research Analyst

Cary K. Jim

Note: This work was initiated at the Global XPRIZE Education Open Data Challenge by Microsoft and the Open Data Institute (UK).

Maps showing FCC fixed broadband availability and broadband usage based on Microsoft data updated as of October 2020

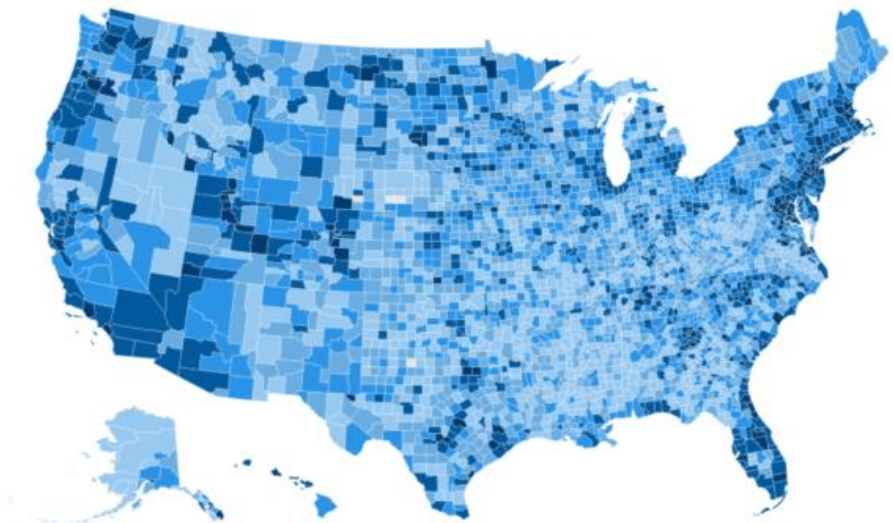
FCC indicates broadband is not available to ~14.5M people



FCC broadband availability*: 0% | >0 to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100%

* FCC Broadband has or "could" provide greater than or equal to 25 Mbps / 3 Mbps

Microsoft data indicates ~120.4M people do not use the internet at broadband speeds



Broadband usage**: 0% | >0 to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100%

** Broadband speeds greater than or equal to 25 Mbps

Select a View

- FCC broadband availability
- FCC and Microsoft
- Congressional districts
- Broadband subscriptions

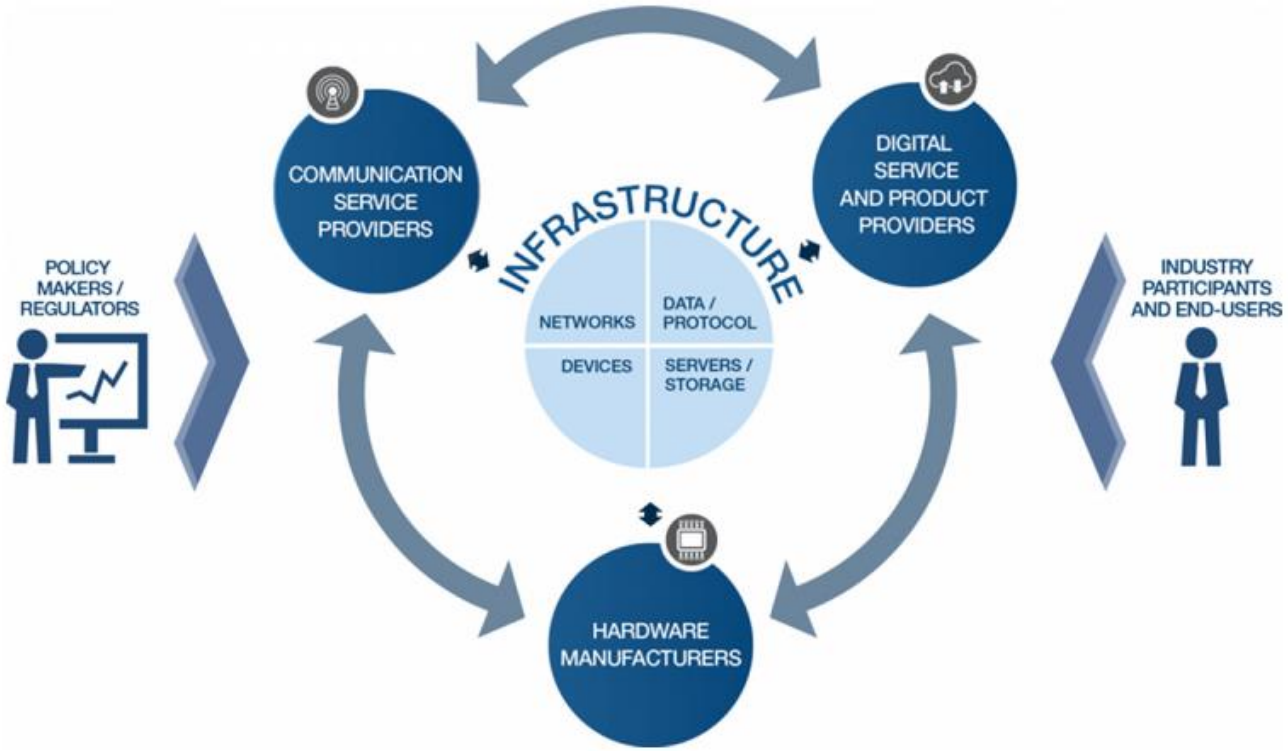
Select a State

All

Sources: FCC Fourteenth Broadband report based on form 477 data from December 2019 and Microsoft data from October 2020
To assist with additional broadband mapping analysis data has been made downloadable here. Learn more in this GitHub repository.

What is going on and how to address it?

World Economic Forum



- Unequal internet penetration and adoption
- Affordability
- Unequal skills among users
- Lack of motivation to participate online

Students Digital Opportunity (SDO)

Broadband Availability

Broadband Usage

Speed Quality (Download & Upload)

Device Ownership & Internet Subscription at Home

Data Sources

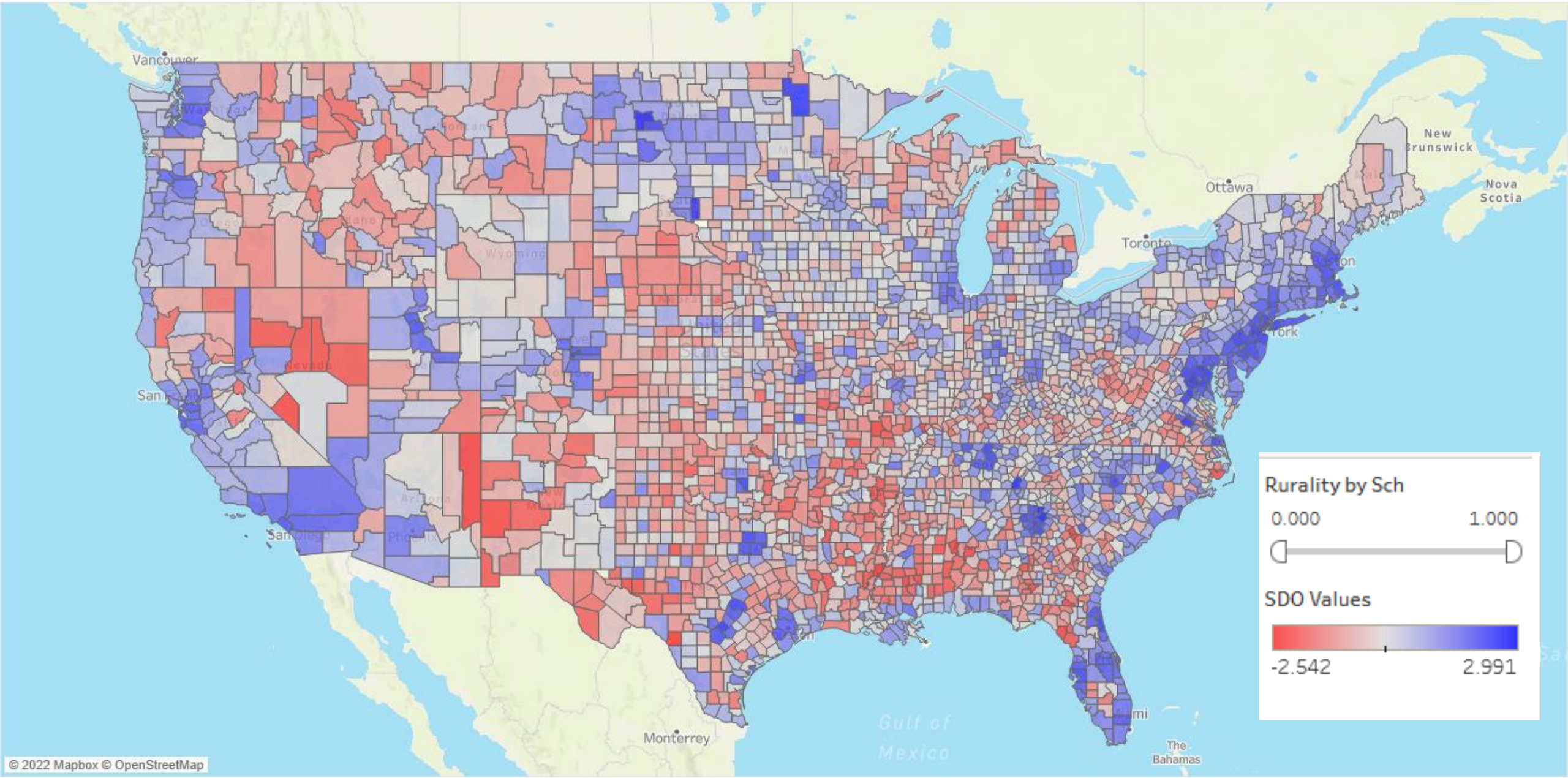
NCES Elementary/Secondary Information System (50 states + District of Columbia)
2019 - 2020 K-12 public school's administrative data

Microsoft Airband Initiatives
2020 Broadband availability and usage data

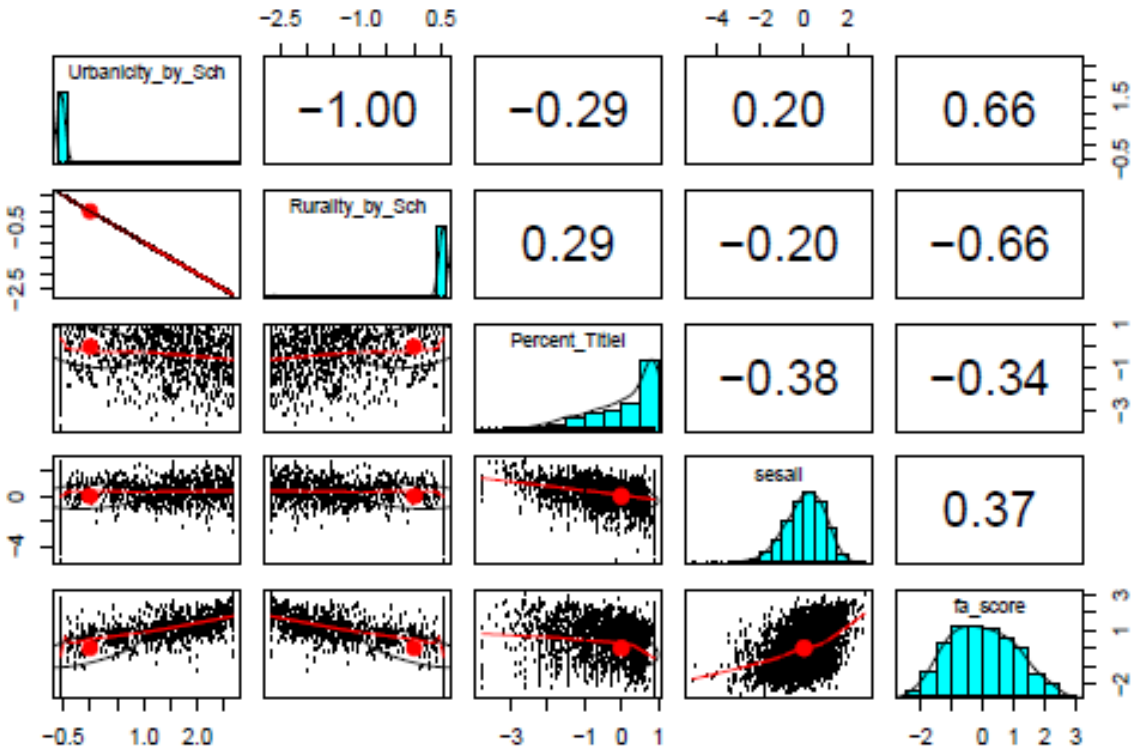
U.S. Census ACS Survey 2019 5-years estimates
Computer and types of internet subscription in household (with children under 18 yrs. old)

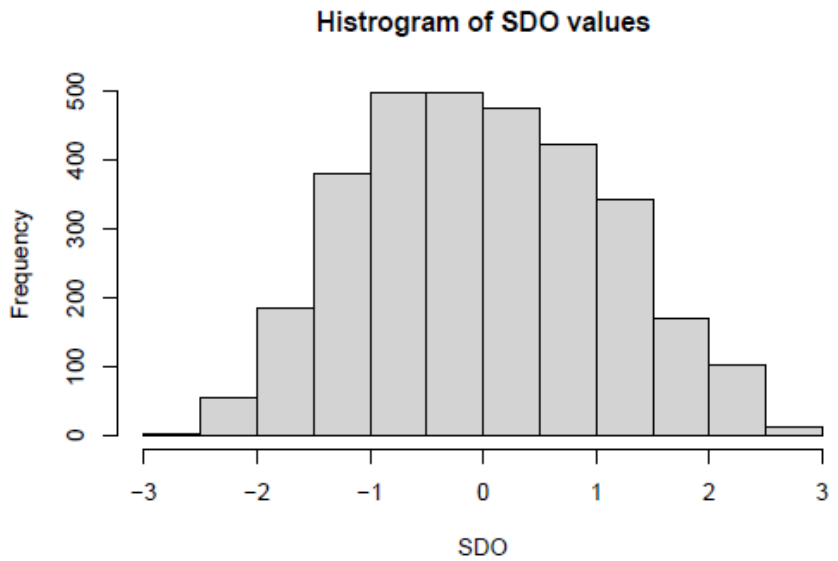
Ookla SpeedTest
Download Speed, Upload Speed, and Latency

Stanford Education Data Archive (v 4.1)
Performance scores, socio-economic measure, poverty level


Vancouver
New Brunswick
Nova Scotia
Ottawa
Toronto
Monterrey
Gulf of Mexico
The Bahamas
Rurality by Sch
0.000
1.000
SDO Values
-2.542
2.991
© 2022 Mapbox © OpenStreetMap

SDO	BELOW AVERAGE (< 0)	ABOVE AVERAGE (> 0)
Number of County	1781 (56.8%)	1357 (43.2%)
Total Student Enrollment	5,777,902 (11.5%)	44,488,214 (88.5%)
Number of K-12 Schools	17,555 (18.6%)	77,064 (81.4%)





Spatial Analysis: Spatial Autocorrelation

Dataset + SDO + 2019 Census Shapefiles (County Level)

Global Moran's I Statistics (Analytic Method)
+ Monte Carlo simulation (for Comparison)

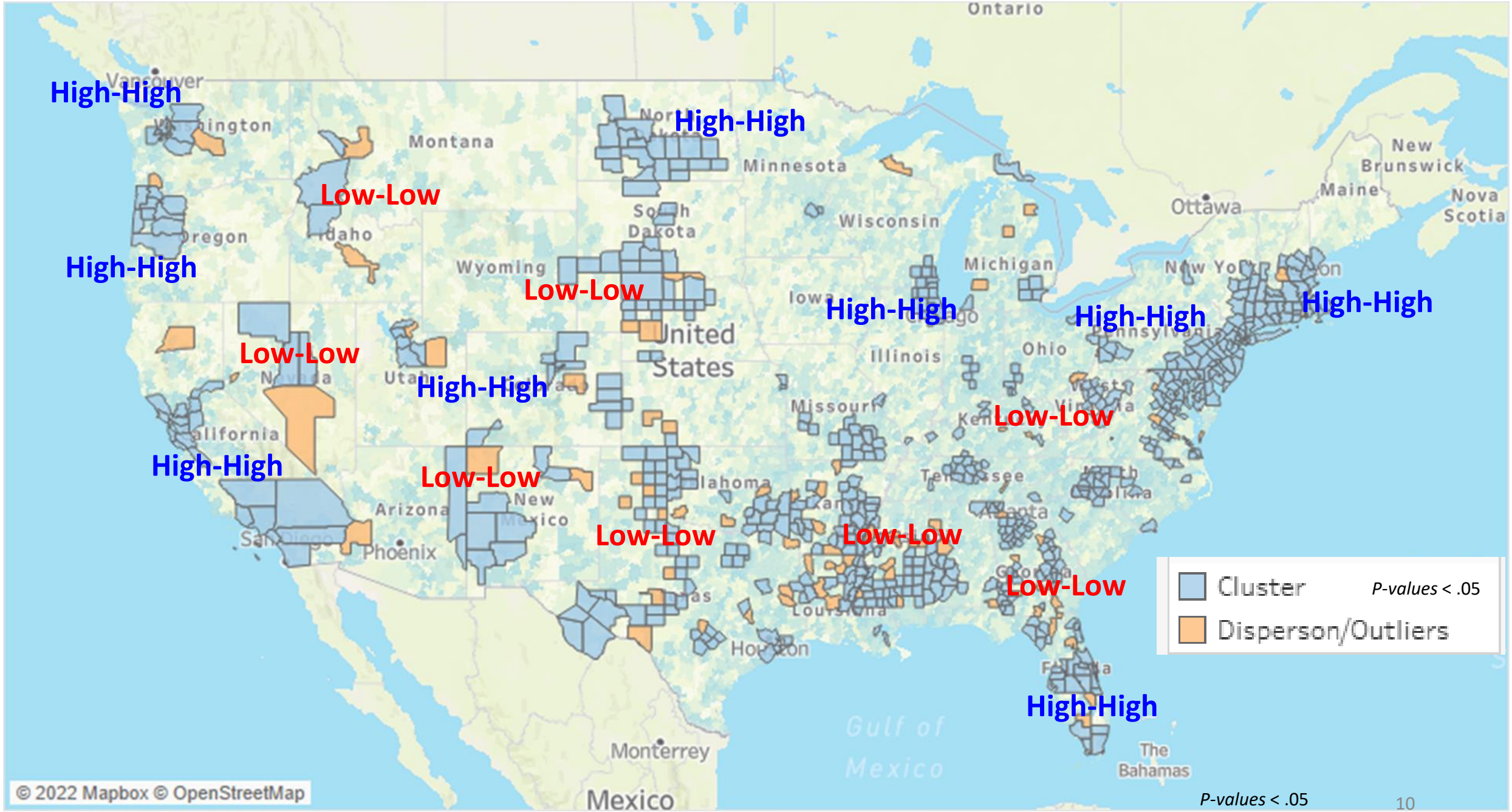
Estimate Local Moran's I Statistics & *p-values*

Global Moran's I : 0.445
p-value <0.001

Obtain local Moran's I values and Quadrant's information

Quadrants:
High – High Clusters
Low – Low Clusters

Dispersion/Outliers:
High – Low Clusters
Low – High Clusters


High-High
High-High
Low-Low
High-High
Low-Low
Low-Low
High-High
High-High
High-High
High-High
Low-Low
High-High
Low-Low
Low-Low
Low-Low
Low-Low
High-High
Cluster
Disperson/Outliers
P-values < .05
P-values < .05
© 2022 Mapbox © OpenStreetMap
Ontario
Montana
Minnesota
Wisconsin
Michigan
Ottawa
New Brunswick
Maine
Nova Scotia
Wyoming
Iowa
Illinois
Ohio
Missouri
Utah
Arizona
Phoenix
Monterrey
Mexico
Gulf of Mexico
The Bahamas
United States

Intermediate Needs and Supports

Internet Connectivity:
Reliable broadband technology and affordable cost for the public

Encourage adoption & usage:
Multi-mode options for learning and access of information

Flexible content for equipment types, bandwidth, speed options

Long-Term Needs and Supports

Training and Knowledge Sharing:
Training for children beyond classroom
Upskilling for adults
Information and knowledge sharing

Maintenance
Infrastructure and technology
Human capital

Utilize SDO measure for progress monitoring, evaluation, and improvement

Leverage on existing social structure

Understand how culture, attitudes, beliefs & mindsets play a role

Thank you!

Questions and Feedbacks

Supplemental Materials

Three-Level of Digital Divide

1st Level: Physical Access

- Lack of access to broadband internet or technology (Servon, 2002; Grubesci & Murray, 2002; Riddlesden & Singleton, 2014)
- Affordability (Chao & Park, 2020; Cotten et al., 2011; Gonzales, 2016)

2nd Level: Usage and Skills

- Adoption and use (DiMaggio & Hargittai, 2001)
- Unequal skills in relation to usage (Warschauer, 2003)

3rd Level: Tangible Benefits (Social and Economic Terms)

- Employability
- Economic Outcomes
- Social Mobility
- Quality of Life

How does students' digital opportunity vary across the U.S. at the county level?

Students Digital Opportunity (SDO)

- Level 1: Physical Access + Level 2: Uses of the Internet

Spatial Analysis

- Level 3: Social and Tangible Benefits

Data Cleaning and Processing

Checking for missing data and list the rows with NA

```
usage_county_2020[!complete.cases(usage_county_2020), ]
```

```
##      ST COUNTY.ID                COUNTY.NAME BROADBAND.AVAILABILITY.PER.FCC   BROADBAND.USAGE
## 68   AK      2013     Aleutians East Borough                              NA             0.066
## 69   AK      2016 Aleutians West Census Area                              NA             0.023
## 71   AK      2050         Bethel Census Area                              NA             0.050
## 72   AK      2060        Bristol Bay Borough                              NA             0.054
## 74   AK      2070
## 82   AK      2164
## 93   AK      2158
## 95   AK      2282
## 1657 NE     31005
## 1658 NE     31007
## 1713 NE     31117
## 2243 OR     41069
## 2658 TX     48269
## 2922 VA     51580
## 2924 VA     51595
## 2925 VA     51600
## 2933 VA     51678
## 2936 VA     51685
## 2937 VA     51690
## 2952 VA     51830
```

```
library(missRanger)
# Since the dataframe has been reduced, there are 3138 records with 18 columns(variables)
# The random forest imputation is fairly quick with a small dataframe (file size 552.5 KB)
data_imputed <-missRanger(final_data, pmm.k = 50, num.trees = 1000)

##
## Missing value imputation by random forests
##
##   Variables to impute:       Percent_TitleI, BROADBAND.AVAILABILITY.PER.FCC, BROADBAND.USAGE,
download_Mbps, upload_Mbps, avg_latency, sesall
##   Variables used to impute:  COUNTY.ID, ST, County.Name, Total_STU, Urbanicity_by_Sch, Rurality_by_Sch,
Percent_TitleI, Total_Household, Total_household_w_children, Percent_Home_w_Stu, Percent_Stu_home_w_PC,
Percent_Stu_PC_and_Broadband, BROADBAND.AVAILABILITY.PER.FCC, BROADBAND.USAGE, download_Mbps, upload_Mbps,
avg_latency, sesall
## iter 1:  .......
## iter 2:  .......
## iter 3:  .......
## iter 4:  .......
```

Appendix A

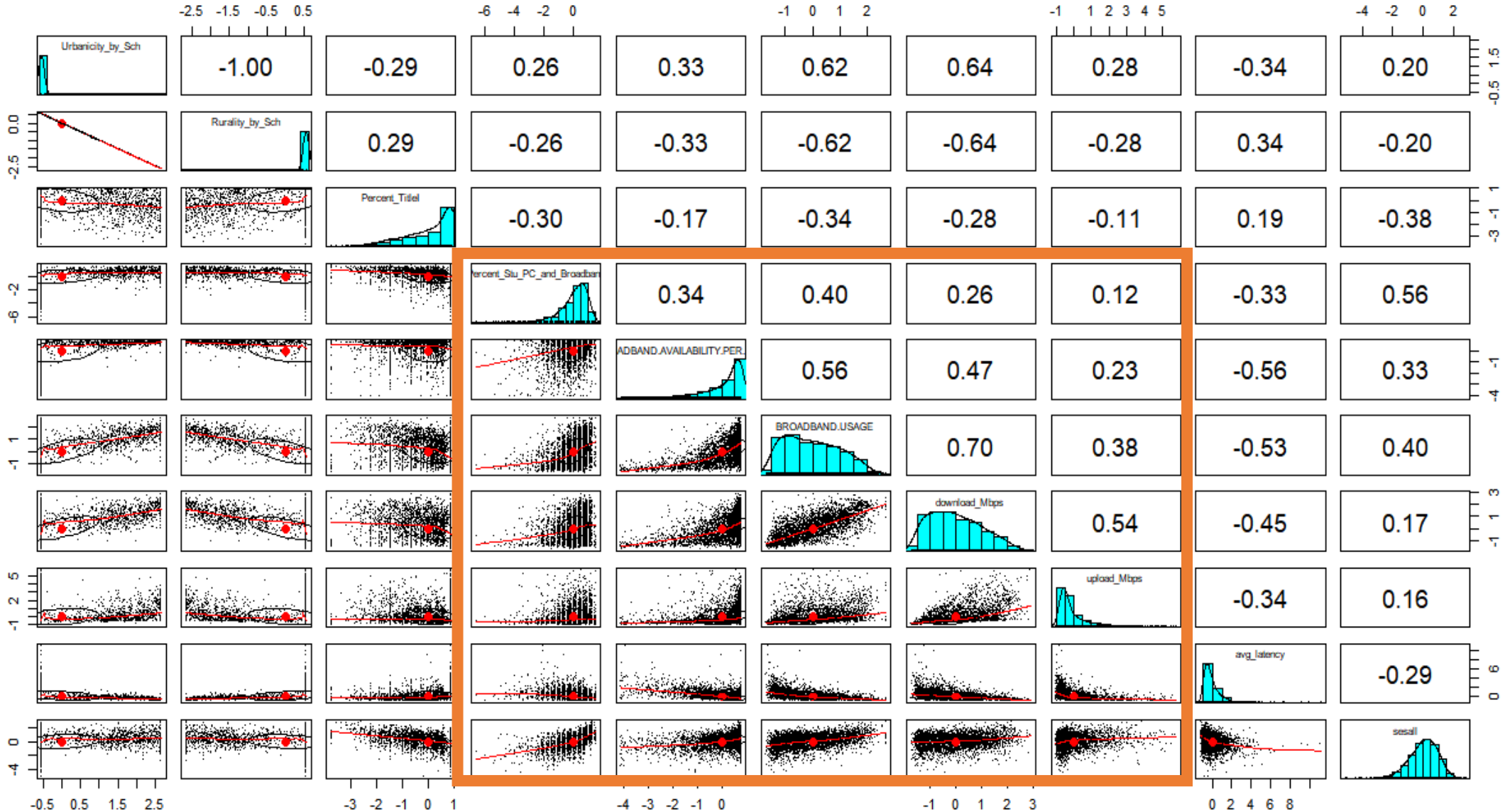
Data Elements, Description, and Sources of Each Variables in the Final Dataset

Data Elements	Data Description & Processing	Data Source
COUNTY.ID	County code (4-5 digit)	Kahan & Ferres (2021)
ST	State (2 Letter abbreviation)	Kahan & Ferres (2021)
County.Name	County name	NCES (2019-2020)
Total_STU	Total number of K-12 student enrollment excluding adult education (Enrollment numbers were sum to the county level)	NCES (2019-2020)
Urbanicity_by_Sch	Urban-centric Locale Categories in NCES: City: Large (11), Mid-size(12), Small(13); Suburb: Large(21), Mid-size(22), Small(23). (The sub-levels were aggregated into the two main categories for city and suburb locale and then calculated a percent of schools for this category at the county level)	NCES (2019-2020)
Rurality_by_Sch	Urban-centric Locale Categories in NCES: Town: Fringe(31), Distant(32), Remote(33); Rural: Fringe(41), Distant(42), Remote(43). (The sub-levels were aggregated into the two main categories for town and rural locale and then calculated a percent of schools for this category at the county level)	NCES (2019-2020)
Percent_TitleI	Title I School Status in NCES: 1 - School is eligible for Title 1 Targeted Assistance (TAS) but provides no program 2 - School is eligible for Title 1 Targeted Assistance (TAS) and provides TAS program 3 - School is eligible for Title 1 Schoolwide program (SWP) and provides TAS program 4 - School is eligible for Title 1 Schoolwide program (SWP) and provides no program 5 - School is eligible for Title 1 Schoolwide program (SWP) and provides SWP program 6 - School is not eligible for either Title 1 Targeted Assistance or Schoolwide Program (SWP) 7 – Unknown status added by author for missing values (The levels were first combined in 3 main types based on eligibility regardless of service provided in the reported data: No program or non-response for level 6 and 7, Title I Targeted Assistance (TAS) including both eligible and identified schools for level 1 and 2, Title I School Wide Program (SWP) for all eligible and identified schools for level 3, 4, 5. A percent of Title I was calculated based on the total number of schools with TAS and SWP over all schools at the county level)	NCES (2019-2020)

Data Elements	Data Description & Processing	Data Source
Total_Household	Total number of household estimated at the county level	U.S. Census Bureau (2019)
Total_household_w_children	Total number of household with population under 18 years estimated at the county level	U.S. Census Bureau (2019)
Percent_Home_with_Stu	A percent of household with under 18 years was calculated at the county level	Author
Percent_Stu_home_w_PC	Total number of household with population under 18 years and has a computer estimated at the county level. (A percent was calculated for the number of household with children who owns a computer at the county level)	U.S. Census Bureau (2019) Table B28005
Percent_Stu_PC_and Broadband	Total number of household with population under 18 years, has a computer with a broadband Internet subscription. (A percent was calculated for the number of household with children who owns a computer and broadband internet at the county level)	U.S. Census Bureau (2019) Table B28005
BROADBAND.AVAILABILITY.PER.FCC	Percent of population per county with access to fixed terrestrial broadband at speeds of 25 Mbps/3 Mbps as of the end of 2019 in FCC report	Kahan & Ferres (2021)
BROADBAND.USAGE	Percent of population per county that use the internet with at least 25 Mbps download speed as of Oct 2020	Kahan & Ferres (2021)
download_Mbps	Average download speed in megabits per second (Mbps) over 4 quarters in Year 2020	Ookla Speed Test (2020)
upload_Mbps	Average upload speed in megabits per second (Mbps) over 4 quarters in Year 2020	Ookla Speed Test (2020)
avg_latency	Average delay milliseconds (ms) over 4 quarters in Year 2020	Ookla Speed Test (2020)
sesall	SES Composite based on U.S. Census ACS median family income, proportion of adults with a bachelor's degree or higher, proportion of adults that are unemployed, the household poverty rate, the proportion of households receiving SNAP benefits, and the proportion of households with children that are headed by a single mother. (See SEDA 4.1 Technical Documentation for more details)	SEDA 4.1 (Reardon et al., 2021)

Note. Stanford Education Data Archive Technical Documentation Version 4.1 (June 2021) can be accessed through https://stacks.stanford.edu/file/druid:db586ns4974/seda_documentation_4.1.pdf

Exploratory Data Analysis



Correlation of the Standardized Variables

Variables	1	2	3	4	5	6	7	8	9	10
1. Urbanicity_by_School	--									
2. Rurality_by_School	- 1.00	--								
3. Percent_TitleI_School	- .29	.29	--							
4. Percent_with _PC_and_Broadband	.26	- .26	- .30	--						
5. Broadband Access/ Availability (FCC)	.33	- .33	- .17	.34	--					
6. Broadband Usage	.62	- .62	- .34	.40	.56	--				
7. Download_Speed_Mbps	.64	- .64	- .28	.26	.47	.70	--			
8. Upload_Speed_Mbps	.28	- .28	- .11	.12	.23	.38	.54	--		
9. Average_latency_ms	- .34	.34	.19	- .33	- .56	- .53	- .45	- .34	--	
10. Socioeconomic Status	.20	- .20	- .38	.56	.33	.40	.17	.16	.29	--

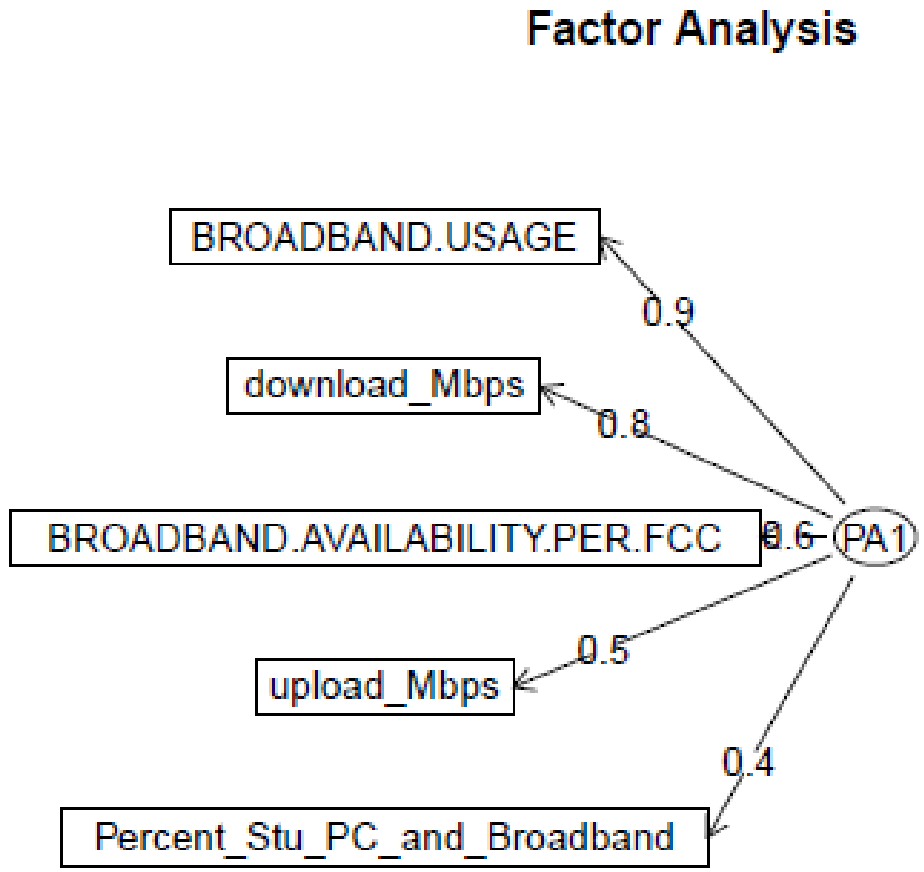
(Correlation *p-values* for all variables are < .001)

Factor Analysis: Principal Axis Factoring Model

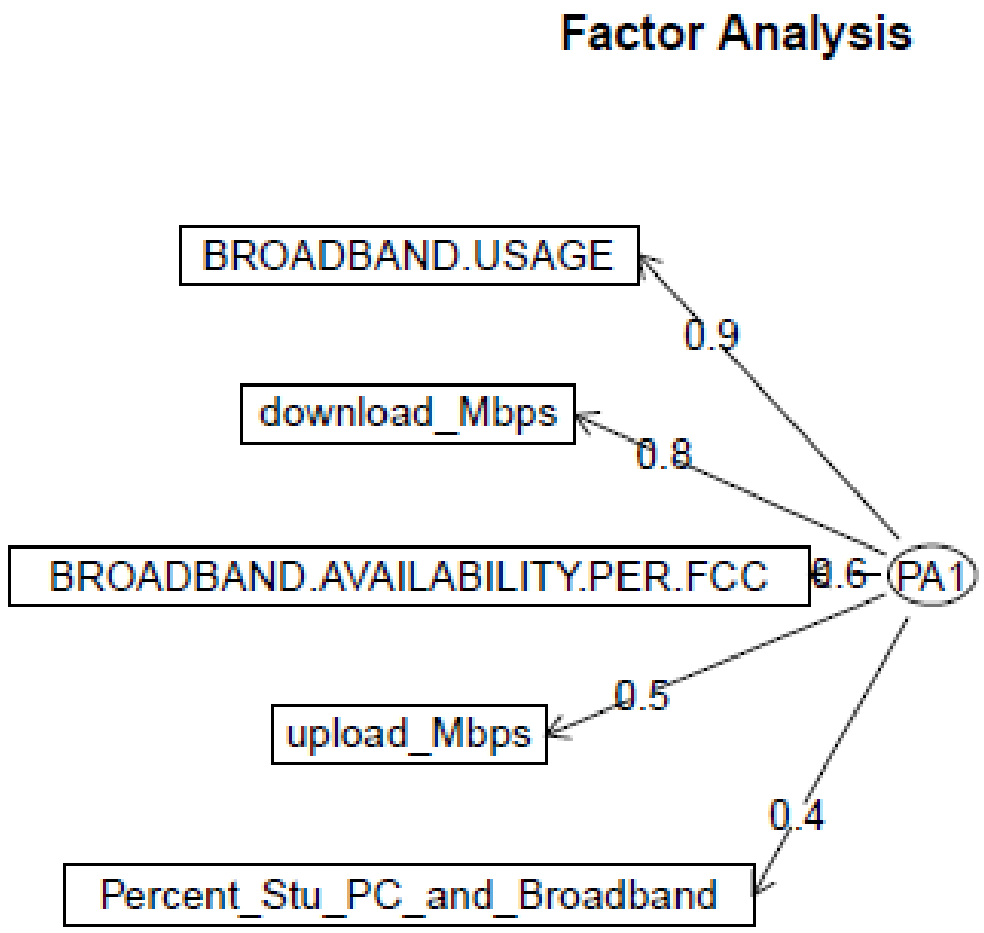
- Multivariate statistics
- Highlight relationships
- Latent (underlying) structure

- Correlation matrix
- Assume error variance

- Standardized scores
  (Bartlett method)
- Reliability assessed Cronbach's Alpha



Factor Analysis – Principal Axis Factoring



Variables	Factor Loadings	Communality (h^2)	Unique Variance (u^2)	Cronbach Coefficient alpha
PC Ownership and Broadband Subscription	.41	.17	.83	.79
Broadband Access/Availability	.61	.37	.63	.73
Broadband Usage	.87	.76	.24	.66
Speed: Download	.83	.69	.31	.67
Speed: Upload	.49	.24	.76	.77