Data & Analytics For Good •

No Internet = No High School Diploma? Data to Identify Counties at-risk of Increased Dropout Rates during the COVID-19 Pandemic (I1.A5)

Lincoln H. Groves, Linsey Dyson, Lauren Homel, Atul Kachare, Hiwot Tesfaye, and

Published on: Sep 02, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Structured Abstract

Source Data Article Overview

These secondary data originate from a study on the pre-pandemic association between high school dropout rates and internet access in the home, while controlling for several other demographic, economic, and social factors affecting the high school dropout rate. As the coronavirus (COVID-19) forced many students into remote learning, our work sought to identify counties across the United States where dropout rates were already high – and could be exacerbated as vulnerable students lack a critical tool to help them complete their diplomas in an online world: reliable internet access.

In this source data article, we seek to foster a community of like-minded researchers interested in the links between high school dropout rates, local economic conditions, and socioeconomic factors. To help establish this community within the *Data and Analytics for Good Journal*, we provide data and source code, and share underlying methodologies in the hope that other researchers join our quest to better understand how the COVID-19 pandemic will impact vulnerable students across the United States. Internet access in the home will be just one of the factors affecting student performance, and we hope that others will provide data which can help us better understand our potential looming educational crisis.

Data Value

This project strives to start an empirical conversation about vulnerable populations and equal access to education with the *Data and Analytics for Good Journal*. These data support *U.N. Sustainability (UNSDG) Goal #4 – Quality Education* and *UNSDG #10 Reduce Inequality*. The goal of this secondary data compilation is to help academics, decision makers, and citizen data scientists to think more critically about where remote learning stemming from the COVID-19 pandemic could exacerbate achievement gaps in education. Moreover, these data can be used to identify geographic regions that may need additional resources to help remediate youth returning to school.

Data Description

Our data provide demographic, economic, and socioeconomic data for 3,133 counties all across the United States. This secondary data compilation primarily uses several

public resources produced by the U.S. government, for all counties reporting data. Most of the variables in the data are derived from individual survey response data from the 2019 American Community Survey, which is collected by the U.S. Census Bureau. We also use economic data from the Local Area Unemployment Statistics (LAUS) from the U.S. Bureau of Labor Statistics, Small Area Income and Poverty Estimates (SAIPE) data from the U.S. Census Bureau, and the Social Vulnerability Index from the U.S. Center for Disease Control and Prevention.

Data Application

The final data set contains a host of county-level factors related to educational attainment, socioeconomic status, access to technology, and social vulnerability. Data can be used for simple descriptive analyses – including geospatial analysis – as well as more complex modeling such as decision trees and regressions. Possible research questions include, but are not limited, to:

- What is the relationship between high school dropout rates and internet access in the home – before COVID19?
- How do local employment opportunities – as proxied by Census occupation classifications¹ – affect high school dropout rates?
 - And do these opportunities differ by race/ethnicity and gender?
- How do dropout rates by race and ethnicity vary across counties in the United States?
 - How do they differ by race/ethnicity and gender?
- How are socioeconomic variables, such as median household income and poverty rates, related to high school dropout rates?

Indexing Table

Supported UN SDGs	Goal 4: Quality Education; Goal 10: Reduce Inequality
Type of Data/Article	Archival Data
Class of Analytics	Descriptive
Data Tables	Six tables ACS Data; LAUS Data; SAIPE Data; SVI Data; County + PUMA Pop (2010); PUMA to County Crosswalk

Key words	High school dropout rates; home internet access; local economic factors; socioeconomic status
------------------	---

Introduction

High school dropouts face increased risks for a number of socially undesirable outcomes, including less stable family environments, greater financial hardships, and higher rates of incarceration (Apel and Sweeten, 2010; Rumberger and Lamb, 2003; Carlson et al., 2004; Cherlin, 2010; Western and Wildeman, 2009; Blanchflower and Freeman, 2000; Pettit and Western, 2004; Campbell, 2015). High school dropouts also experience less stability in the labor market and lower earnings over their lifetime (Western et al., 2001; Heckman and Lafontaine, 2010). Reducing the high school dropout rates should be a priority for policymakers interested in increasing the long-term human capital accumulation of their citizens, which will make individuals more self-sufficient and less reliant upon public assistance programs throughout their lifecycle (Bane and Ellwood, 1996; Hamiul-Luker, 2006).

Beyond historical factors contributing to high school dropouts, the coronavirus upended education in the United States, forcing students to rapidly adapt to an online learning environment. While many school districts provided internet hotspots and computers for youth to continue learning at home in a virtual world, other districts lacked the resources and infrastructure to provide these vital services. To better understand how the COVID pandemic could impact existing trends in high school dropout rates, we created a data set to explore the pre-COVID links between internet access in the home and high school dropout rates, while controlling for other factors. In this secondary data compilation, we provide data that can be used to identify areas of the country that could be particularly vulnerable to the transition to remote learning – because they were already experiencing low graduation rates and low access to technology outside of the classroom – and that may need additional remedial resources after primary and secondary education returns to normal after the COVID pandemic subsides.

Given that numerous factors influence the dropout rates (Heckman and Lafontaine, 2010; Murnane, 2013), the true impact of the potential learning time “lost” to the COVID pandemic will take many years to unravel. The goal of this source data is to start a broader discussion of how the COVID pandemic will affect high school dropout rates in both the short- and long-run. This question is not yet answerable – and it will be difficult to disentangle the causal mechanisms once a sufficient time has passed for

data collection. Moreover, we know that access to technology is just one of the many challenges affecting learning in the United States. That stated, data set creation is the first important step in exploring these relationships.

This project aligns with *UNSDG Goal #4 - Quality Education* and *UNSDG #10 - Reduce Inequality*. In this source data article, our intent is to provide the initial framework for others to build off this work and add to our body of knowledge. As more data become available – both in terms of a post-COVID evaluation period and other explanatory variables that help us disentangle the endogenous relationship between internet access and high school dropout rate, we hope to contribute to the broader conversation on whether COVID-19 impacted students at-risk of dropping out from high school. Towards that goal, we provide the source SAS code to our secondary data compilation in a GitHub repository so that other researchers can more easily extend our work.²

U.N. Goal ID

Primary: *Quality Education: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all; Goal 4*

Secondary: *Reduce inequality within and among countries; Goal 10*

Education is traditionally seen as a great equalizer in the United States (Holmes and Zajacova, 2014; Torche, 2011). Regardless of a child's circumstances at birth, the notion is that if they work hard, invest in education, then they can prosper in the long run.

We know that the story is much more complex than this, as a host of factors influence a child's educational trajectory (Holmes and Zajacova, 2014; Torche, 2011; Heckman and Lafontaine, 2010; Murnane, 2013). However, as we strive to reduce inequality in the United States (UNSGD #10), quality education (UNSDG #4) is a critical component of those efforts.

Value of Data

Our data provides demographic, economic, and socioeconomic data for 3,133 counties in the United States. Excluding the Social Vulnerability Index variables (2018), which we will define in the next section, data are for 2019 – the year before the pandemic swept across the United States. This allows us to establish a pre-period baseline, where we could examine relationships between high school dropout rates and

technology access under “good” times. It is reasonable to assume that counties struggling in the pre-pandemic period would perform even worse given the challenges of learning in a COVID environment.

Variables in the final data set includes the following categories shown in Figure 1:

Measures of High School Dropout Rates	Internet Access	Geographic Data	County Demographics	Poverty + Income Measures	Local Labor Force Characteristics
<ul style="list-style-type: none"> 18- to 20-year old dropout rates, all residents 18- to 20-year old dropout rates, by gender 18- to 20-year old dropout rates, by race-ethnic group 	<ul style="list-style-type: none"> % of household in county with internet access in the home % of household in county with high speed internet access in the home 	<ul style="list-style-type: none"> State + County identifying information, including FIPS Codes 	<ul style="list-style-type: none"> County population, residents County population, daytime population Sex ratio (men to women) Social Vulnerability Index, from CDC 	<ul style="list-style-type: none"> Estimates of all people in poverty, raw count and % of population Estimates of all people in poverty, ages 0-17 Estimates of children in poverty, ages 5-17 in families 	<ul style="list-style-type: none"> Raw count of labor force size, employment, and unemployment Unemployment rate Employment rate, all residents and by gender Labor force participation rate, all residents and by gender Not in labor force, aged 25-54 Percent of local labor force engaged by 7 occupational sectors Percent of local labor force engaged by 7 occupational sectors, by gender

Figure 1: Type of Variables in Final Data Set

Data Sources

This submission relies solely on data produced by the United States government. A brief description of each resource is found below – and more details can be found online by accessing the links provided.

[American Community Survey \(ACS\) - U.S. Census Bureau](#)

The American Community Survey is an ongoing survey conducted by the U.S. Census Bureau. By collecting monthly data on jobs and occupations, educational attainment, household composition, and other factors, the ACS helps determine how more than \$675 billion in federal and state funds are distributed each year.³ The ACS household survey was created to replace the “long form” of the decennial census program, and Census Bureau representatives target 3.5 million household addresses annually to gather this important information.⁴

Data used in this compilation comes from IPUMS USA.⁵ Originally “the Integrated Public Use Microdata Series”, the IPUMS databases are maintained by the Minnesota Population Center (MPC) at the University of Minnesota. Within the IPUMS USA project, the MPC harmonizes data across sixteen federal censuses, as well as all the American Community Surveys.⁶

In this secondary data compilation, we used microsample data from the 2019 ACS 5-year sample. Statistics were first aggregated to the lowest public use microdata area (PUMA) in the data – and then mapped to the county levels.⁷ Please see the *Data Compilation* section for more details.

Local Area Unemployment Statistics (LAUS) - U.S. Bureau of Labor Statistics

To better understand current economic conditions, Local Area Unemployment Statistics are collected by the U.S. Bureau of Labor statistics each month.⁸ In this compilation, we use county-level measures of local economic conditions, including the total number in the labor force, the number of employed and unemployed individuals, and the unemployment rate – and use the full-year 2019 LAUS data to establish a pre-pandemic baseline.

Small Area Income and Poverty Estimates (SAIPE) - U.S. Census Bureau

The primary goal of the U.S. Census Bureau’s Small Area Income and Poverty Estimates program is to produce annual estimates of income and poverty for all U.S. states and counties.⁹ In this source data article, we use SAIPE estimates for 2019 – to provide a pre-pandemic baseline – and these data provide us with a range of poverty and median household income estimates.

Social Vulnerability Index (SVI) - Center for Disease Control and Prevention

Finally, the CDC Social Vulnerability Index seeks to measure how well communities may respond to events such as natural disasters or infectious disease outbreaks. By focusing on factors such as socioeconomic status, household composition, minority status, or housing type and transportation, this index strives to help identify which areas are most vulnerable when external stressors are placed on the community.¹⁰

The SVI index uses 15 variables from the U.S. Census to help local officials identify communities that may need support before, during, or after disasters.¹¹ Data in the SVI are from 2018 – the last reporting year before the pandemic. The index is bound between 0 and 1, where a higher number indicates greater vulnerability. Finally, the individual components of the index – called SVI Themes – are included in our compilation. The four SVI themes are socioeconomic status, household composition, race/ethnicity/language, and housing/transportation.

Data Compilation

In this section, we outline how we compiled the data from various federal resources to create a county-level data set. This section is organized by the corresponding SAS programs used to wrangle the data.

[00 - Employment Data Import - 05.06.2021.sas](#)

This SAS program imports the Local Area Unemployment Statistics data from the U.S. Bureau of Labor Statistics. Original data are in the .CSV format. Select labor force variables are retained in a SQL statement - and relabeled to provide clarity.

[01 - ACS Data - 05.06.2021.sas](#)

This program is the main file in our compilation. In this program, we accomplish several tasks. The first set of tasks focused on preparing the individual level data, where we:

1. Import data from the American Community Survey (ACS) using code provided by IPUMS USA.
2. Limited data to individuals aged 16-64, not enrolled in school.
3. Prepared data formats and labels to make the data more useful in subsequent analyses
4. Used occupational codes from 2010 to aggregate up into occupational code grouping, including "No Occupation", "White Collar", "Community, Health and Service", "Office Sales and Support", "Service Occupations", "Blue Collar Laborers", and "Protective Services".
5. Identified high school dropouts in the data. Dropouts were defined as individuals aged 18 to 20 that were no longer enrolled in high school but did not have a high school diploma. This is one the methodologies used in the academic literature, including Cohodes et al. (2016) and Groves (2020).
6. Prepared select demographic variables for subsequent calculations, such as race and ethnic group.

Once the individual-level data was imported and prepared, the next step was to address a challenge with the ACS data: the smallest level of geography published is a public use microdata area (PUMA). A PUMA covers no fewer than 100,000 residents - and is utilized to protect the identity of survey respondents. There are a few hurdles with this definition. To start, several small counties can fall into a single PUMA in rural areas. At the opposite end of the population spectrum, several PUMAs can constitute the same county in a dense urban area (for example, Los Angeles County in California

comprises 35 PUMAs). Consequently, we needed a methodology to aggregate PUMAs up to larger geographic areas (i.e., counties) when necessary and use the estimate for a single PUMA across multiple counties when appropriate. Moreover, PUMAs can overlap counties and counties have overlapping PUMAs – all of which complicates the estimates derived from the ACS.

For the PUMA-to-county challenge, we used the methodology advocated by the University of Michigan Population Studies Center (MPSC).¹² We started with the MPSC’s PUMA-to-county crosswalk and saved this file as *PUMA to County Crosswalk*. Using data from the Missouri Census Data Center,¹³ we then appended 2010 populations to this crosswalk so that we could use the PUMA populations as statistical weights for the cases where we needed to average county-level statistics across multiple PUMAs. Without weights we would produce a simple weighted average, meaning that the less populated PUMA would receive a disproportionate impact on a calculated statistic for that county.

With the PUMA, county, and weight dataset collected, we used the Census person-weights for each individual respondent in the ACS sample to estimate a series of variables for educational attainment, internet access in the home, and labor force activity. We estimated these statistics at the three geographic levels recorded in the ACS: (1) the public use microdata area (PUMA), (2) the consistent PUMA (CPUMA), 2000-2010, and the county Federal Information Processing Standards (FIPS) code. Only one of these geographic levels was provided for each ACS survey respondent.

The final step in our data preparation process was to merge the ACS estimates with the PUMA-to-county crosswalk and then aggregate data to the county-level. After this work, we then had a series of estimates for the 3,133 counties in our compilation.

02 - Create Vulnerable Population Data Set - 05.06.2021.sas

The third, and final, SAS program was used to load the SAIPE and Social Vulnerability data into SAS and to combine the four main pieces of our data set: (1) the LAUS data, (2) the ACS data, (3) the SAIPE data, and (4) the SVI index. We also did a bit of variable clean-up, including relabeling, to make the data more accessible.

Figure 2 shows the Entity Relationship Diagram for the resources used to compile our data:

Figure 2: Entity Relationship Diagram

As shown, underlying data sets in our compilation were linkable by the FIPS code for that state and county.

Data Summary

Table 1 contains simple descriptive statistics for all variables included in our data set:

Visit the web version of this article to view interactive content.

Table 1: Descriptive Statistics

For the main outcome of interest in our secondary data compilation – high school dropout rates – county estimates of the population aged 18-20 that are no longer enrolled in school and do not have a high school diploma range from 0% to 52%, with

an average dropout rate of 18% across all counties.¹⁴ For our primary input of interest, we see that high-speed internet access in the home estimates range from 44% of the county population to 96% of the population, at the upper end. 75% of all county residents have access to high-speed internet in the home. Figure 3 and Figure 4 provide an overview of how internet access in the home is distributed across counties in the United States:

Figure 3: % of Population in County with Home Internet Access

Figure 4: % of Population in County with High-Speed Internet Access in the Home

Note that while the individual county estimates use the ACS sample weights, the *Mean* column in Table 1 is a simple average.

Data Application and Conclusion

The COVID-19 pandemic could have a profound impact upon educational outcomes in the United States, especially for students lacking the essential resources to thrive in a remote learning environment. This source data article stems from a project conducted during the SAS Institute's Second Annual Social Innovations Summit in the fall of 2020, in which we focused on one of the essential ingredients to learning during the quarantine: access to high-speed internet in the home. However, the authors recognize that a myriad of other factors will affect both the short- and long-term graduation rates in the post-COVID environment (Heckman and Lafontaine, 2010; Murnane, 2013). Thus, our primary intent is to start a conversation and promote evolving databases within the Data and Analytics for Good Journal ecosystem. More specifically, we hope that others can help us build upon these data with either:

- (1) a longer time series (i.e., more years of data – particularly in the recent period),
- (2) more explanatory variables, and, perhaps,

(3) data at another unit of analysis (i.e., district-level or individual student-level data).

Education is a critical ingredient in reducing inequality in America. By highlighting vulnerable communities that could be even more at-risk of experiencing short- and long-term increases in the high school dropout rates resulting from technological barriers stemming from new methods of classroom delivery due to COVID-19, we hope that our collaborative work can contribute to a path forward to help students at risk of not completing high school get the remedial support they require.

Coders Appendix

The three programs used to compile the final *Vulnerable_Populations_DataSet* can be found in the GitHub repository found here:

<https://github.com/lincolngroves/Data-and-Analytics-for-Good-Journal-v1>

For additional questions on code and methodology, please contact the corresponding author.

Acknowledgements

The genesis of this work was the SAS Institute's Second Annual Social Innovations Summit in the fall of 2020. While in quarantine as Team 6, the authors focused on vulnerable populations and sought to highlight the "COVID slide", which was gaining prominent media attention. As part of the Summit, Sterlina Smith at SAS provided invaluable administrative support, guidance, and encouragement.

The authors would also like to thank the editors of the Data and Analytics for Good Journal for their timely feedback and input, which helped transform the idea from Social Innovations Summit project to journal article.

Conflict of Interest

None.

References

Apel, Robert, and Gary Sweeten. 2010. "The Impact of Incarceration on Employment during the Transition to Adulthood." *Social Problems* 57 (3): 448–79.
<https://doi.org/10.1525/sp.2010.57.3.448>.

Bane, Mary Jo, and David T Ellwood. 1996. *Welfare Realities*. Harvard University Press. <https://www.hup.harvard.edu/catalog.php?isbn=9780674949133>.

Blanchflower, David G., and Richard B. Freeman. 2000. "The Declining Economic Status of Young Workers in OECD Countries." In *Youth Employment and Joblessness in Advanced Countries*. University of Chicago Press.

Campbell, Colin. 2015. "High School Dropouts After They Exit School: Challenges and Directions for Sociological Research." *Sociology Compass* 9 (7): 619-29. <https://doi.org/10.1111/soc4.12279>.

Carlson, Marcia, Sara McLanahan, and Paula England. 2004. "Union Formation in Fragile Families." *Demography* 41 (2): 237-61. <https://doi.org/10.1353/dem.2004.0012>.

Cherlin, Andrew J. 2010. "Demographic Trends in the United States: A Review of Research in the 2000s." *Journal of Marriage and Family* 72 (3): 403-19. <https://doi.org/10.1111/j.1741-3737.2010.00710.x>.

Cohodes, Sarah R., Daniel S. Grossman, Samuel A. Kleiner, and Michael F. Lovenheim. 2016. "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions." *Journal of Human Resources* 51 (3): 727-59. <https://doi.org/10.3368/jhr.51.3.1014-6688R1>.

Groves, Lincoln H. 2020. "Still 'Saving Babies'? The Impact of Child Medicaid Expansions on High School Completion Rates." *Contemporary Economic Policy* 38 (1): 109-26. <https://doi.org/10.1111/coep.12431>.

Hamil-Luker, Jenifer. 2005. "Trajectories of Public Assistance Receipt among Female High School Dropouts." *Population Research and Policy Review* 24 (6): 673-94. <https://doi.org/10.1007/s11113-005-5751-0>.

Heckman, James J, and Paul A LaFontaine. 2010. "The American High School Graduation Rate: Trends and Levels." *Review of Economics and Statistics* 92 (2): 244-62. <https://doi.org/10.1162/rest.2010.12366>.

Holmes, Christopher J., and Anna Zajacova. 2014. "Education as 'the Great Equalizer': Health Benefits for Black and White Adults." *Social Science Quarterly* 95 (4): 1064-85. <https://doi.org/10.1111/ssqu.12092>.

Murnane, Richard J. 2013. "U.S High School Graduation Rates: Patterns and Explanations." Working Paper 18701. National Bureau of Economic Research.

<http://www.nber.org/papers/w18701>.

Pettit, Becky, and Bruce Western. 2004. "Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration." *American Sociological Review* 69 (2): 151-69.

Torche, Florencia. 2011. "Is a College Degree Still the Great Equalizer? Intergenerational Mobility across Levels of Schooling in the United States." *American Journal of Sociology* 117 (3): 763-807. <https://doi.org/10.1086/661904>.

Western, Bruce, Jeffrey R. Kling, and David F. Weiman. 2001. "The Labor Market Consequences of Incarceration." *Crime & Delinquency* 47 (3): 410-27. <https://doi.org/10.1177/0011128701047003007>.

Western, Bruce, and Christopher Wildeman. 2009. "The Black Family and Mass Incarceration." *The ANNALS of the American Academy of Political and Social Science* 621 (1): 221-42. <https://doi.org/10.1177/0002716208324850>.

End Notes

[1] For more detail on Industry and Occupation Classifications, see <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/industry-and-occupation-classification.html>

[2] <https://github.com/lincolngroves/Data-and-Analytics-for-Good-Journal-v1>

[3] <https://www.census.gov/programs-surveys/acs/about.html>

[4] <https://www.bls.gov/lau/acsqa.htm#Q01>

[5] Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021. <https://doi.org/10.18128/D010.V11.0>

[6] <https://usa.ipums.org/usa/about.shtml>

[7] To protect the identity of respondents, the smallest unit of geography in the ACS is a public use microdata area (PUMA). From the Census Bureau, PUMAs are "Public Use Microdata Areas (PUMAs) are non-overlapping, statistical geographic areas that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each."

[8] <https://www.bls.gov/lau/>, <https://www.bls.gov/lau/#cntyaa>

[9] <https://www.census.gov/programs-surveys/saipe.html>

[10] <https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>

[11] <https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>

[12] See: <https://www.psc.isr.umich.edu/dis/census/Features/puma2cnty/>

[13] See: <https://mcdc.missouri.edu/applications/geocorr2018.html>

[14] When N is less than 3133 for an estimate, it means that insufficient data exists to provide a reliable estimate. In particular, ethnic minorities are less likely to live in rural areas, and this is reflected in the estimates.

Appendix A | Entity Data Dictionary

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
25	COUNTY	Char	25	\$25.	
4	CountyFIPS	Num	8		FIPS county code (0 for US or state level records)
38	Dropouts_All	Num	8	PERCENT9.2	Dropouts - All
40	Dropouts_Black	Num	8	PERCENT9.2	Dropouts - Blacks
41	Dropouts_Hispanic	Num	8	PERCENT9.2	Dropouts - Hispanic
43	Dropouts_Men	Num	8	PERCENT9.2	Dropouts - Men

42	Dropouts_Other	Num	8	PERCENT9.2	Dropouts - Others
39	Dropouts_White	Num	8	PERCENT9.2	Dropouts - Whites
44	Dropouts_Women	Num	8	PERCENT9.2	Dropouts - Women
28	E_TOTPOP	Num	8	BEST12.	Estimate of Daytime Population in County
35	Employed	Num	8	COMMA11.	Employed 2019
47	Employment_Rate_Men	Num	8	PERCENT9.1	Employment Rate Men
50	Employment_Rate_Women	Num	8	PERCENT9.1	Employment Rate
26	FIPS	Num	8	BEST12.	
46	High_Speed_Internet	Num	8	PERCENT9.2	High Speed Internet in the Home
30	Household_Composition_Disability	Num	8	BEST12.	Household Composition & Disability
32	Housing_Type_Transportation	Num	8	BEST12.	Housing Type & Transportation
45	Internet_Access	Num	8	PERCENT9.2	Internet Access in the Home
27	LOCATION	Char	44	\$44.	

34	Labor_Force	Num	8	COMMA11.	Labor Force 2019
21	Median_HH_Income_Lower	Num	8		90% confidence interval lower bound of estimate of median household income
22	Median_HH_Income_Upper	Num	8		90% confidence interval upper bound of estimate of median household income
20	Median_Household_Income	Num	8		Estimate of median household income
49	Men_Not_LF_25_54	Num	8	PERCENT9.	Not in LF Men 25-54
31	Minority_Language	Num	8	BEST12.	Minority Status and Language
52	Not_in_LF_25P	Num	8	PERCENT9.	Not in Labor Force 25+
48	Not_in_LF_Men	Num	8	PERCENT9.	Not in Labor Force Men
51	Not_in_LF_Women	Num	8	PERCENT9.	Not in Labor Force Women

54	Occup_Group_0	Num	8	PERCENT9.1	No Occupation
55	Occup_Group_1	Num	8	PERCENT9.1	White Collar
56	Occup_Group_2	Num	8	PERCENT9.1	Community, Health and Service
57	Occup_Group_3	Num	8	PERCENT9.1	Office Sales and Support
58	Occup_Group_4	Num	8	PERCENT9.1	Service Occupations
59	Occup_Group_5	Num	8	PERCENT9.1	Blue Collar Laborers
60	Occup_Group_6	Num	8	PERCENT9.1	Protective Services
61	Occup_Group_0_Men	Num	8	PERCENT9.1	No Occupation Men
68	Occup_Group_0_Women	Num	8	PERCENT9.1	No Occupation Women
62	Occup_Group_1_Men	Num	8	PERCENT9.1	White Collar Men
69	Occup_Group_1_Women	Num	8	PERCENT9.1	White Collar Women
63	Occup_Group_2_Men	Num	8	PERCENT9.1	Community, Health and Service Men
70	Occup_Group_2_Women	Num	8	PERCENT9.1	Community, Health and Service Women

64	Occup_Group_3_Men	Num	8	PERCENT9.1	Office Sales and Support Men
71	Occup_Group_3_Women	Num	8	PERCENT9.1	Office Sales and Support Women
65	Occup_Group_4_Men	Num	8	PERCENT9.1	Service Occupations Men
72	Occup_Group_4_Women	Num	8	PERCENT9.1	Service Occupations Women
66	Occup_Group_5_Men	Num	8	PERCENT9.1	Blue Collar Laborers Men
73	Occup_Group_5_Women	Num	8	PERCENT9.1	Blue Collar Laborers Women
67	Occup_Group_6_Men	Num	8	PERCENT9.1	Protective Services Men
74	Occup_Group_6_Women	Num	8	PERCENT9.1	Protective Services Women
33	Overall_SVI_Index	Num	8	BEST12.	Overall Social Vulnerability Index
5	Poverty	Num	8		Estimate of people of all ages in poverty
11	Poverty_0_17	Num	8		Estimate of people age 0-17 in poverty

12	Poverty_0_17_Lower	Num	8		90% confidence interval lower bound of estimate of people age 0-17 in poverty
13	Poverty_0_17_Upper	Num	8		90% confidence interval upper bound of estimate of people age 0-17 in poverty
17	Poverty_5_17_Families	Num	8		Estimated percent of related children age 5-17 in families in poverty
18	Poverty_5_17_Families_Lower	Num	8		90% confidence interval lower bound of estimate of percent of related children age 5-17 in families in poverty

19	Poverty_5_17_Families_Upper	Num	8		90% confidence interval upper bound of estimate of percent of related children age 5-17 in families in poverty
6	Poverty_Lower	Num	8		90% confidence interval lower bound of estimate of people of all ages in poverty
8	Poverty_Pct	Num	8		Estimated percent of people of all ages in poverty
14	Poverty_Pct_0_17	Num	8		Estimated percent of people age 0-17 in poverty
15	Poverty_Pct_0_17_Lower	Num	8		90% confidence interval lower bound of estimate of percent of people age 0-17 in poverty

16	Poverty_Pct_0_17_Upper	Num	8		90% confidence interval upper bound of estimate of percent of people age 0-17 in poverty
9	Poverty_Pct_Lower	Num	8		90% confidence interval lower bound of estimate of percent of people of all ages in poverty
10	Poverty_Pct_Upper	Num	8		90% confidence interval upper bound of estimate of percent of people of all ages in poverty
7	Poverty_Upper	Num	8		90% confidence interval upper bound of estimate of people of all ages in poverty
23	STATE	Char	20	\$20.	
24	ST_ABBR	Char	2	\$2.	
53	Sex_Ratio	Num	8	6.2	Sex Ratio - Men to Women

29	Socioeconomic	Num	8	BEST12.	Socioeconomic Status
3	StateFIP	Num	8		FIPS state code
1	State_Name	Char	45		State or county name
2	State_Postal	Char	2		Two-letter Postal State abbreviation
36	Unemployed	Num	8	COMMA11.	Unemployed 2019
37	Unemployment_Rate	Num	8	PERCENT9.1	Unemployment Rate 2019

[High School Dropout Rates and Internet Access.pdf](#)

2 MB

Footnotes

1. For more detail on Industry and Occupation Classifications, see <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/industry-and-occupation-classification.html> ↗
2. <https://github.com/lincolngroves/Data-and-Analytics-for-Good-Journal-v1> ↗
3. <https://www.census.gov/programs-surveys/acs/about.html> ↗
4. <https://www.bls.gov/lau/acsqa.htm#Q01> ↗
5. Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021. <https://doi.org/10.18128/D010.V11.0> ↗
6. <https://usa.ipums.org/usa/about.shtml> ↗

7. To protect the identity of respondents, the smallest unit of geography in the ACS is a public use microdata area (PUMA). From the Census Bureau, PUMAs are “Public Use Microdata Areas (PUMAs) are non-overlapping, statistical geographic areas that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each.” [↵](#)

8. <https://www.bls.gov/lau/>, <https://www.bls.gov/lau/#cntyaa> [↵](#)

9. <https://www.census.gov/programs-surveys/saipe.html> [↵](#)

10. <https://www.atsdr.cdc.gov/placeandhealth/svi/index.html> [↵](#)

11. <https://www.atsdr.cdc.gov/placeandhealth/svi/index.html> [↵](#)

12. See: <https://www.psc.isr.umich.edu/dis/census/Features/puma2cnty/> [↵](#)

13. See: <https://mcdc.missouri.edu/applications/geocorr2018.html> [↵](#)

14.

When N is less than 3133 for an estimate, it means that insufficient data exists to provide a reliable estimate. In particular, ethnic minorities are less likely to live in rural areas, and this is reflected in the estimates.

Appendix A | Entity Data Dictionary

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
25	COUNTY	Char	25	\$25.	
4	CountyFIPS	Num	8		FIPS county code (0 for US or state level records)
38	Dropouts_All	Num	8	PERCENT9.2	Dropouts - All
40	Dropouts_Black	Num	8	PERCENT9.2	Dropouts - Blacks
41	Dropouts_Hispanic	Num	8	PERCENT9.2	Dropouts - Hispanic
43	Dropouts_Men	Num	8	PERCENT9.2	Dropouts - Men
42	Dropouts_Other	Num	8	PERCENT9.2	Dropouts - Others
39	Dropouts_White	Num	8	PERCENT9.2	Dropouts - Whites
44	Dropouts_Women	Num	8	PERCENT9.2	Dropouts - Women