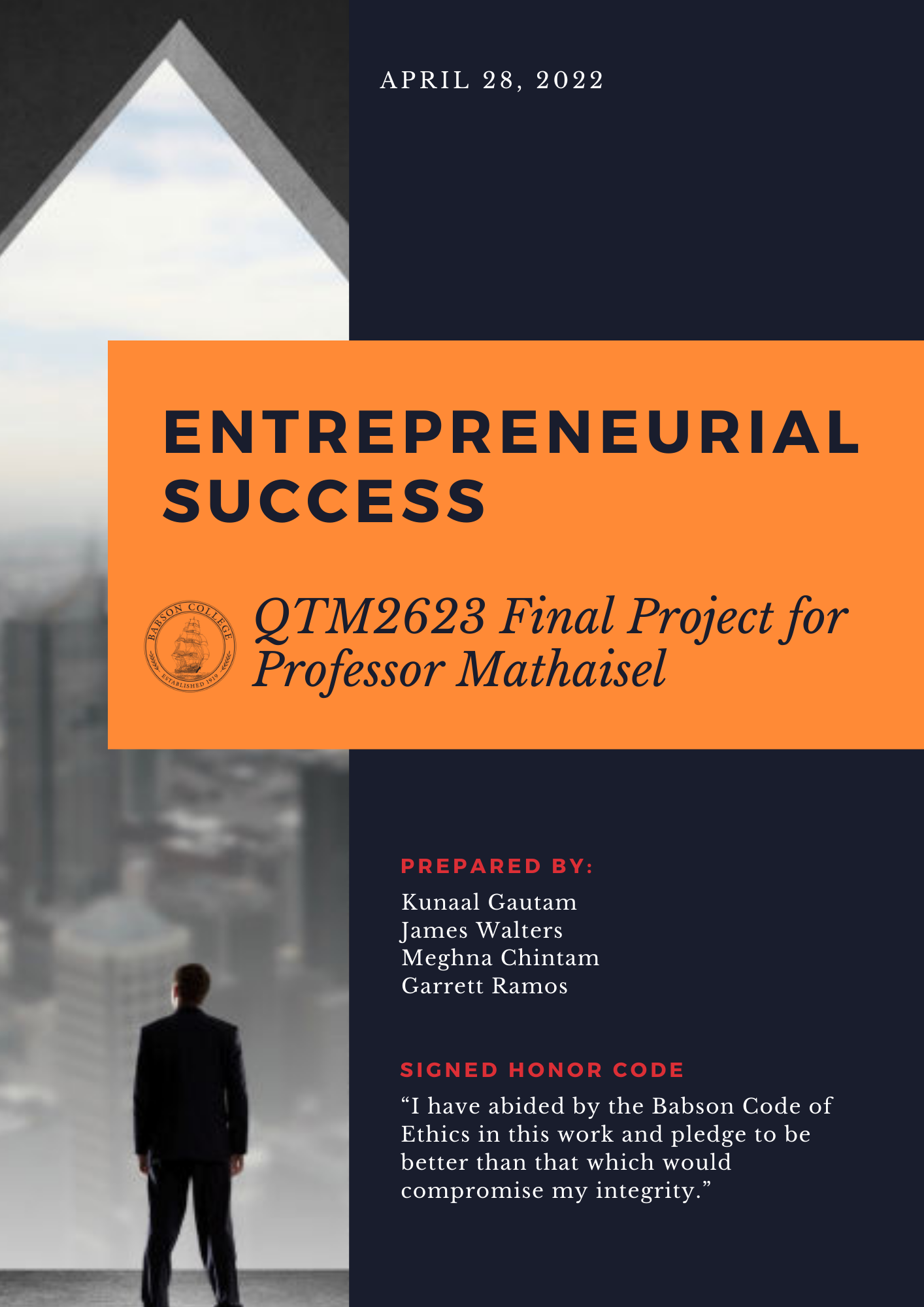
****

**1** **Introduction**

This report analyzes the several economic and social census attributes and their relationship with an entrepreneurship indicator. Two of the datasets were taken from the United States Census Bureau’s website on the American Community Survey (<https://www.census.gov/programs-surveys/acs>). The datasets, which are named “ACS 2019 Social Characteristics Data - for States Transposed.xlsx” and “ACS 2019 Economic Data - for States Transposed.xlsx” (note: the datasets were renamed in R-Studio to be Social and Economic respectively) include data about the population activity by state. It is important to note that although all datasets include information by the 50 US States, each dataset has a different number of observations, with Economic having 52 observations and 126 variables, and Social having 51 observations and 172 variables. Additionally, there are almost 300 variables but only a few variables will be selectively chosen from Economic and Social to carry out the regression models and visualizations. The other dataframe was taken from a dataset named ‘“Early-Stage Entrepreneurship” on the Kauffman Indicators of Entrepreneurship website (note: dataset is available electronically at<https://indicators.kauffman.org/data-downloads>). The dataset was taken directly from the Kauffman Indicators of Entrepreneurship link. The website offers two datasets, the Early-Stage Entrepreneurship dataset and the New Employer Business dataset. The Early-Stage Entrepreneurship, which is a set of measures that represents new business creation in the United States, integrating several high-quality, timely sources of information on early-stage entrepreneurship, will be the dataset that is used for this report and analysis. The dataset, which is named “2021\_kese\_download\_final.csv”, has data about entrepreneurship indices for each of the 50 US states, spanning from 1996 to 2021. The dataset contains variables that include Rate of New Entrepreneurs, Opportunity Share of New Entrepreneurs, Startup Early Job Creation, Startup Early Survival Rate, and the Kauffman Early-Stage Entrepreneurship Index. Each of these variables measures a different metric in relation to early stage entrepreneurship and business owners. It is important to note that the data frame, which was renamed ‘KESE’, has 1846 observations and 10 variables.

From a cursory glance, we also find that the datasets have missing data. For the Social and Economic datasets, we find that there are completely empty columns in order to segregate the dataset. The missing values are denoted by spaces in Excel, but as an ‘NA’ in R-Studio. The same applies for the KESE dataset, making these values unreadable in R-Studio, and there are 4530 missing values. Overall, the datasets are very insightful, and the results of this analysis could offer information to help understand what influences entrepreneurial behavior around early job creation and the various social and economic factors that influence this.

The objective of this analysis is to load the 3 datasets into R-Studio to analyze its data structure, convert it to a user-friendly format to further analyze for insights, create regression models, and use advanced and interactive graphics in R to try to ascertain what economic and social attributes are the most significant attributes of startup early job creation. The packages used for this analysis include “dplyr”, “formattable”, “shiny”, “ggplot2”, “plotly”, “readxl”, “forecast”, and “pROC” which were installed. Through the manipulation of the structure, configuration of the data in the datasets, and creation of regression models and visualizations insights will be generated about the relationship between startup job creation and various social and economic attributes. The analysis aims to make the data usable, and creation of visualizations, regression models, and visualization analysis will reveal any trends in the three datasets.

# **2 Data Dictionary**

The data dictionaries describe the structure of each dataset, list the variables, and provide a description that explains each variable. The data dictionaries summarize the datasets succinctly, and will serve as a frame of reference for each analysis conducted later on. The data dictionaries for these dataset include hundreds of variables, and for the sake of brevity, are not included in the main report. Listed in the Appendix are the details about each individual variable for each dataset.

# **3 Analysis**

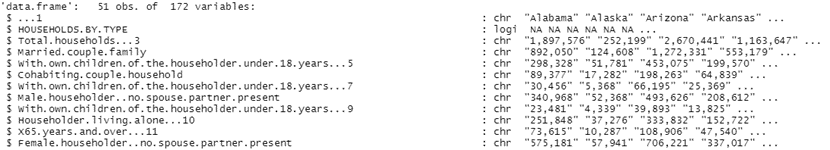
**3.1 Data Structure Analysis**

We begin the analysis by checking if each data structure is a data frame. Through R-Studio, the data can be viewed in tabular form, with the first few rows being displayed. From this initial internal structure analysis, we find that each dataset is a data frame, with each column containing values of one variable and each row containing one set of values from each column. Additionally, there are column names, the data stored in the data frame falls under numeric, factor, or character type, and each column contains the same number of data items. We use a function in R-Studio in order to verify that each of the datasets (Economic, Social, and KESE) is a data frame, which returns as true.

We continue the analysis with a variable analysis. The goal of this analysis is to create a data dictionary for the Economic, KESE, and Social datasets (displayed in the Appendix) that include the variable names, variable attribute type, units and a description of each variable. First, we examined the variable name and variable attribute type. Since the data came from external websites like the US Census and the Kauffman website, the variable names and descriptions were based on information found there and other online articles. In order to find out the variable attribute type, an analysis was conducted on R-Studio to look into the structure of each variable as numeric, integer, etc. A snippet of the output of this analysis, which was done separately for each dataset, is displayed in Exhibit 1. It is important to note that the original Social and Economic datasets include number values as character attribute type because of the commas separating the numbers, which is the incorrect format for visualizations. These variables will have the commas removed and be converted into ‘num’ type, which was a process that was repeated for the KESE dataset. Lastly, the units were discerned by the variable attribute type and description. For example, the variable named ‘Fiscal Policy’ in Economic is numeric, and described as “the overall fiscal policy score as an index with freedom”, and based on this information we discern that the unit is index.

Exhibit 1: Structure Analysis

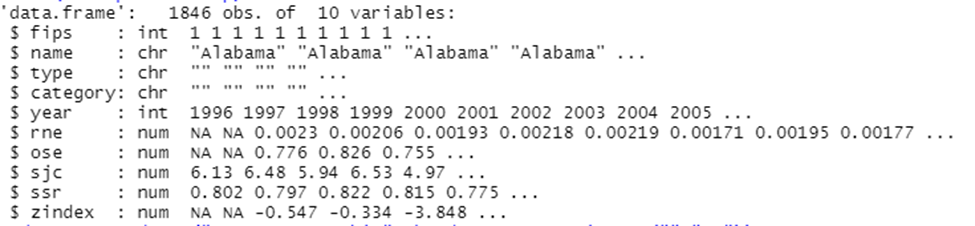
*Social dataset*

**

*Economic dataset*

**

*KESE dataset*

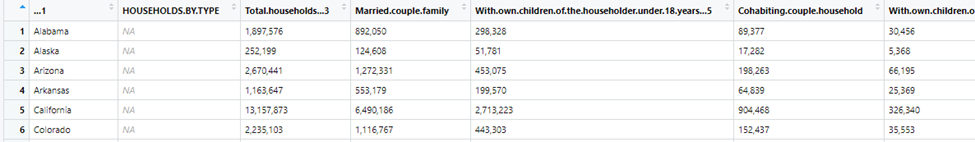
**

As mentioned earlier, it is important to note the missing values because of the empty columns. Social has 918 missing values, Economic has 416 total missing values (including the 52 missing values under ‘Unemployment Rate’ after further conversion), and KESE has 4530 missing values total, and the majority of them come from the ‘type’ and ‘category’ variables. The Economic and Social datasets include entire variable columns of missing values that were originally used to segregate the different variables into broad categories like fertility and employment, as does the KESE dataset, which can be seen in Exhibit 2. These columns will not be removed even though they carry completely empty values.

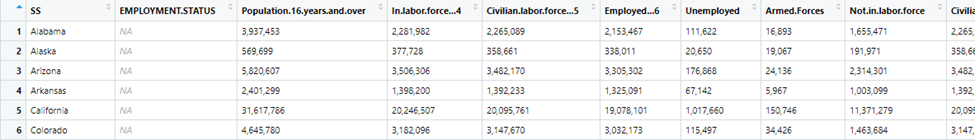
Lastly, in order to do further analysis, it is important to categorize the variables as attributes or as a variable. Considering the information from the data dictionaries, we can conclude that almost all variables in all datasets are attributes because they provide important information that could generate valuable insights. However, in order to be usable for the analysis the chosen variables need to be converted to a numeric type (since some variables were originally classified as character type because of commas). Additionally, some variables, like ‘HOUSEHOLDS BY TYPE’ from the Social dataset can be considered a variable because it is unusable for analysis in its current format. The justification of a variable as an attribute is subjective, and in the next few sections, specific variables will be justifiably chosen for the regression models and visualization analysis.

Exhibit 2: Missing Values

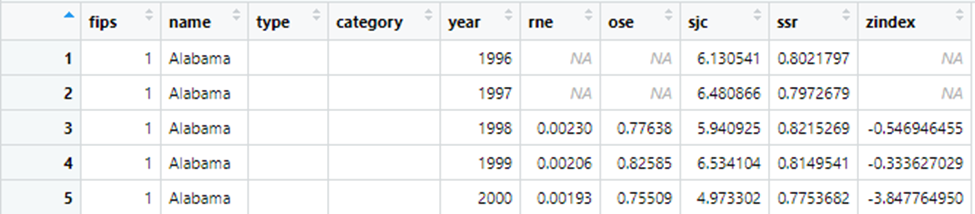
*Social dataset*

**

*Economic dataset*

**

*KESE dataset*

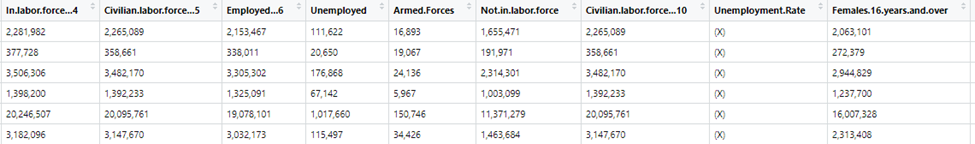
**

**3.2 Missing Value Analysis**

In order to be able to analyze the dataset and generate insights from the analysis, the missing values need to be converted or removed. This conversion and removal was done through R-Studio separately for each dataset. The variable ‘Unemployment rate’ from the Economic dataset had missing values denoted by ‘(X)’, which R-Studio does not recognize as a missing value. The values were changed to ‘NA’ in order to complete this analysis, which can be seen in Exhibit 3. As stated earlier, it is also important to note that some variables, like ‘HOUSEHOLDS BY TYPE’ from the Economic dataset and ‘COMPUTERS AND INTERNET USE’ from the Social dataset are variables with only missing value observations, and others have missing values as observations. All of the missing values from all 3 datasets were removed in R-Studio.

Exhibit 3: Missing Value Conversion

*Original dataset*

**

*After Conversion*

**

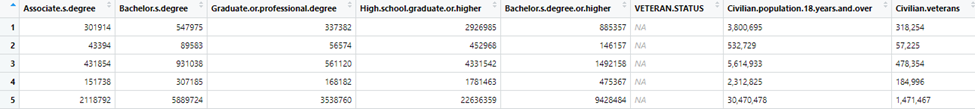
**3.3 Variable Selection and Conversion**

In order to complete the analysis, the attributes being analyzed must be chosen. ‘Wholesale Trade’ (industry), ‘Information’ (industry), ‘Computer use’, ‘Veterans’, ‘Less than 9th grade’, ‘Some college’, ‘Bachelors’, ‘Language other than english’, ‘Never married’, ‘Average family size’ were chosen from the Social dataset, and ‘Self-employed’, ‘Government workers’ and ‘Median-income’ were chosen from the Economic dataset as the independent variables, and ‘sjc’ from the KESE dataset as the dependent variable for the regression models. These variables will be analyzed through the regression models and visualizations. Each variable was chosen to analyze its effect on startup job creation, with specific variables like ‘Self employed; being chosen because they represent the number of entrepreneurs in each state.

The data dictionary shows that the variables in the Economic dataset and the Social dataset are in number units, but recognized by R as character type, which makes these three variables in each dataset not usable or easily readable for an analysis. Using R-Studio, the commas were first removed from the variables that were used to create visualizations, and then each variable was converted to numeric form in the same variable.

These new variables have the same values as the previous one, just without the ‘,’ sign and in a format that is readable by R-Studio and appropriate for the upcoming analysis. For reference, the first few values of the modified Social dataset for the education level variables are displayed in Exhibit 4. This process was repeated for the variables in the Economic dataset and the KESE dataset. As expected, the values under each variable stayed the same, with the only obvious difference being the ‘,’ sign removal.

Exhibit 4: Variable Conversion for Economic dataset



**3.4 Data Preparation and Manipulation**

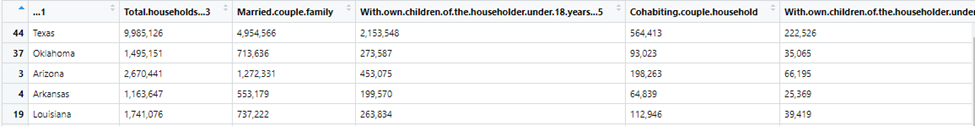
In order to begin this analysis, the KESE, Economic and Social datasets need to be altered. The states need to be grouped into 4 regions for this analysis. The US Census website was used for the region classification and division (note: can be found electronically here:<https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf>). This was done using a for loop in R that looped through a list classifying each state under West, Midwest, South, and North categories based on the link above. These new dataframes, called Total\_West, Total\_Midwest, Total\_South, and Total\_North separate the KESE, Social and Economic datasets by state, but have all the hundreds of combined variables from the original datasets. It is important to note that although they have a different number of observations from each other (West has 11, North has 8, Midwest has 10, South has 18), these represent the number of states in each region. All region data frames will have the same number of variables, however, from combining the KESE, Social and Economic datasets. Additionally, it should be noted that a few states, like Puerto Rico were not included in any of the region classifications and will be ignored in this analysis. A snippet of each of the region data frames are displayed in Exhibit 5.

Exhibit 5: Combining Social and Economic datasets by Region

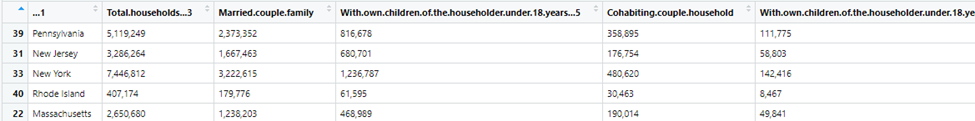
*myData\_West*

**

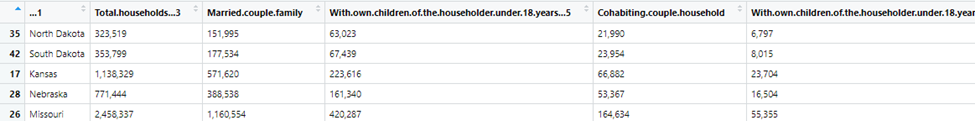
*myData\_South*

**

*myData\_North*

**

*myData\_Midwest*

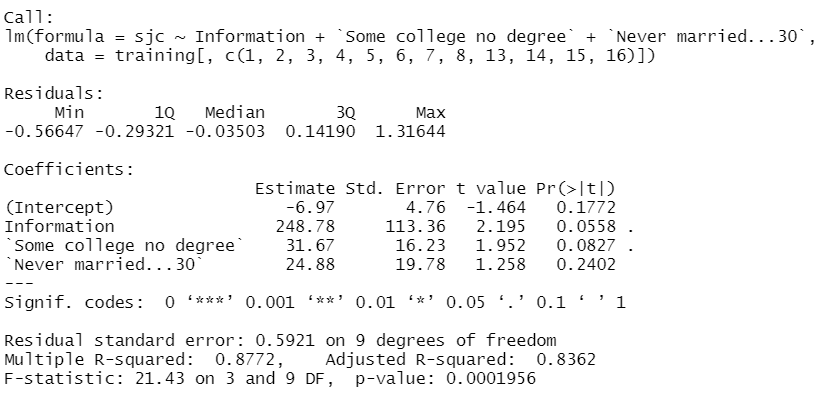
**

It is important to note that some Region categories are much larger than others. Additionally, each state has a different population, and this can skew the data and observations. In order to combat that, the per capita data for all variables was calculated by dividing the variable by the variable ‘Population 16 years and over’. These new per capita variables by region can be found in the new dataframes named ‘myData\_South\_per\_capita’, ‘myData\_Nortth\_per\_capita’, ‘myData\_Midwest\_per\_capita’, and ‘myData\_West\_per\_capita’. These new variables, which are percentages of the total population, take into account the differences in population and will allow for more accurate insights.

**3.5 Linear Regression Models**

In order to determine the relationship between the Social and Economic variables with the Startup Job Creation variable (sjc) from the KESE dataset, linear regression models were created for each of the 4 region categories in R-Studio. The models were created using the variables ‘Information’ (industry), ‘Computer use’, ‘Veterans’, ‘Less than 9th grade’, ‘Some college’, ‘Bachelors’, ‘Language other than english’, ‘Never married’, ‘Average family size’ from the Social dataset, and ‘Self-employed’, ‘Government workers’ and ‘Median-income’ from the Economic dataset as the independent variables, and ‘sjc’ from the KESE dataset as the dependent variable. It is important to note that each Region category has less than 20 observations, making the dataset very small. In order to address and prevent any overfitting, stepwise variable selection was used to find the subset of variables in the data set resulting in the best performing model, and to lower the prediction error.

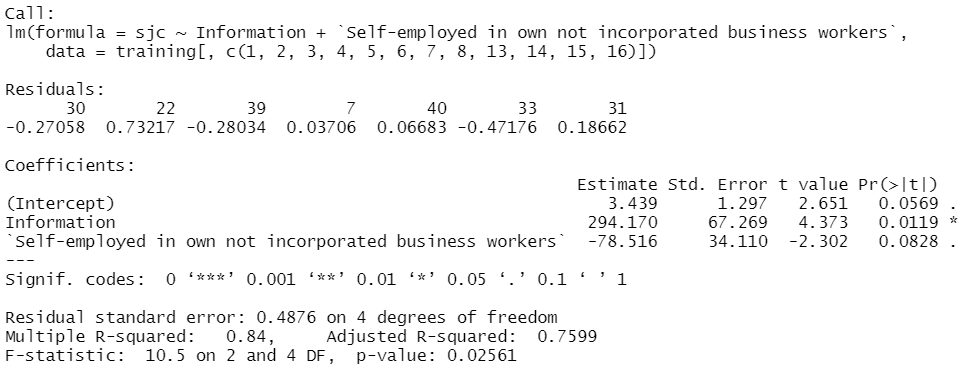
***Model for South Region:***

**

**

The first linear regression model is for the South Region. Stepwise variable selection shows that the variables most relevant for this model are ‘Information’, ‘Some College’, and ‘Never married’. Although the p-value aren’t isn’t statistically significant, the values are relatively low, especially for the ‘Information’ variable. All of the chosen variables have positive values, indicating a positive relationship between them and the sjc variable. The results of the linear regression reveals that for every 1% increase in Information sector jobs, there will be a 243.78% increase in startup job creation. Similarly, a 1% increase in population with some college and a 1% increase in population that was never married results in a 31.67% and 24.88% increase in sjc, respectively. The MAPE of this regression model is 14.66%, meaning this model has relatively good forecast accuracy. The Adjusted R-squared value is high at 0.8362, and a higher R-squared indicates the model is a good fit. Overall, this would mean that an increase in Information jobs, population that has some college education, and population that was never married will result in an increase in startup job creation.

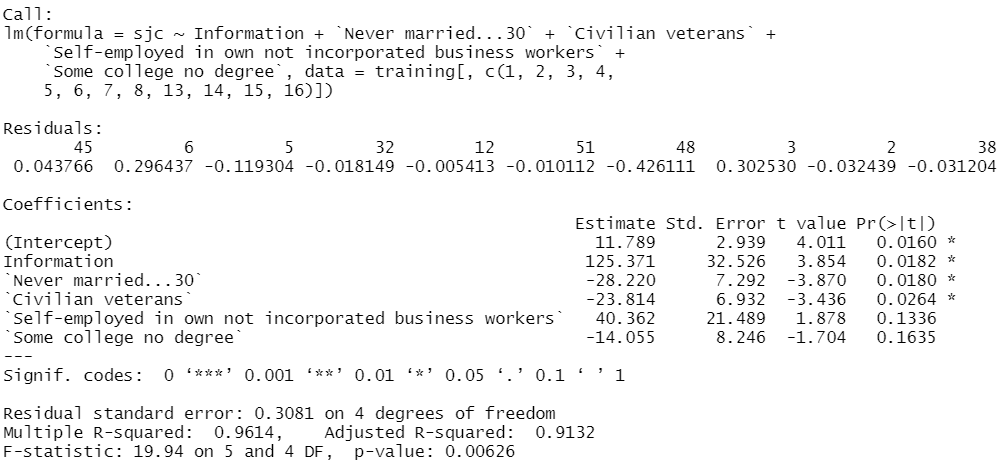
***Model for North Region:***

**

**

The next regression model is for the North Region. Stepwise variable selection shows that the variables most relevant for this model are ‘Information’ and ‘Self employed’. Although the p-value for the ‘Self employed’ variable is not statistically significant, the ‘Information’ variable is. The ‘Information’ variable has a positive coefficient, and the ‘Self employed’ variable has a negative coefficient, meaning that they have a positive and negative relationship with startup job creation, respectively. The results of the linear regression reveals that for every 1% increase in Information jobs, there will be a 294.17% increase in startup job creation. Similarly, for every 1% increase in the number of self-employed people, there will be a 78.52% decrease in startup job creation. The MAPE of this model is relatively higher at 38.51%, meaning that this model does not have great forecast accuracy. The Adjusted R-squared value is relatively high at 0.7599, and a higher R-squared indicates the model is a good fit. Overall, this would mean that an increase in Information jobs would result in more startup job creation, and an increase in self-employment would result in a decrease.

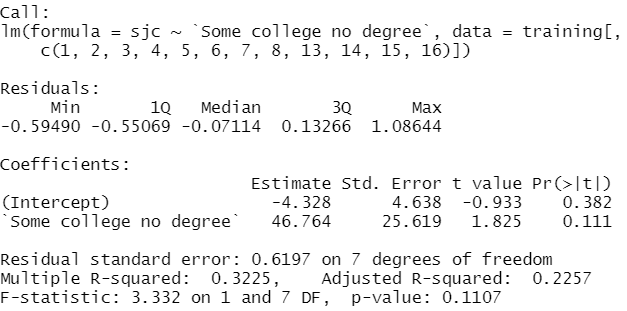
***Model for West Region:***

**

**

The next regression model is for the West Region. Stepwise variable selection shows that the variables most relevant for this model are ‘Information’, ‘Never married’, ‘Self employed’, ‘Some college no degree’ and ‘Civilian veterans’. Although the p-values for the ‘Self employed’ and ‘Some college no degree’ variables are not statistically significant, the p-values for the ‘Information’, ‘Never married’, and ‘Civilian veterans’ variables are. ‘Never married’, ‘Civilian veterans’, and ‘Some college no degree’ have negative coefficients, meaning that they have a negative relationship with the ‘sjc’ variable. On the other hand, ‘Information’ and ‘Self employed’ have positive coefficients, indicating a positive relationship with the ‘sjc’ variable. The results of the linear regression reveals that a 1% increase in Information jobs or self-employed people will cause a 125.37% and 40.36% increase in startup job creation, respectively. Similarly, a 1% increase in people who were never married, civilian veterans, or people with some college education but no degree will cause a 28.22%, 23.81%, and 14.06% decrease in startup job creation, respectively. The MAPE of this model is 22.32%, indicating a good forecast accuracy. The Adjusted R-squared value is very high at 0.9132, and a higher R-squared indicates the model is a good fit.Overall, this would mean that an increase in Information jobs or people who are self employed will result in an increase in startup job creation, and an increase in people who were never married, civilian veterans, or people with some college education but no degree will result in a decrease in startup job creation.

***Model for Midwest Region:***

**

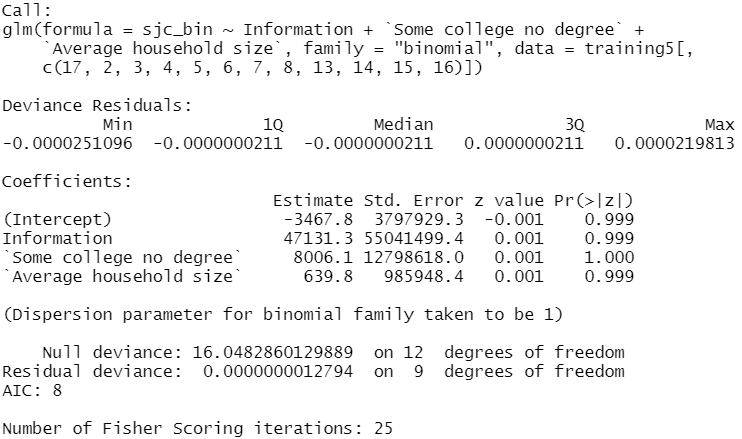
**

The last regression model is for the Midwest Region. Stepwise variable selection shows that the variable most relevant for this model is ‘Some college no degree’. Although the p-value for this variable is not statistically significant, it is relatively low. The variable has a positive coefficient, indicating a positive relationship with the ‘sjc’ variable. The results of the linear regression reveals that a 1% increase in the number of people with some college education but no degree will result in a 46.76% increase in startup job creation. The MAPE of this model is very low at 8.62%, indicating great forecast accuracy. Overall, this would mean that an increase in the number of people with some college education but no degree will result in an increase in startup job creation.

**3.6 Logistic Regression Models**

Similar to the linear regression models addressed above, stepwise selection was used for the purpose of variable selection, to ensure the best predictors were being used for each model. Our dependent variable, early startup job creation, was continuous for the purpose of linear regression. However, for the purpose of logistic regression, we decided to reclassify our dependent variable into a binary, factor variable. The early startup job creation was equal to 1 if the original, continuous value was greater than the regional average for early startup job creation, and a 0, if the value was less than the regional average. Other than this manipulation of the dependent variable, we used the same variables of interest during the modeling process, and a similar partition of the data, separating each region's observations into training and validation sets.

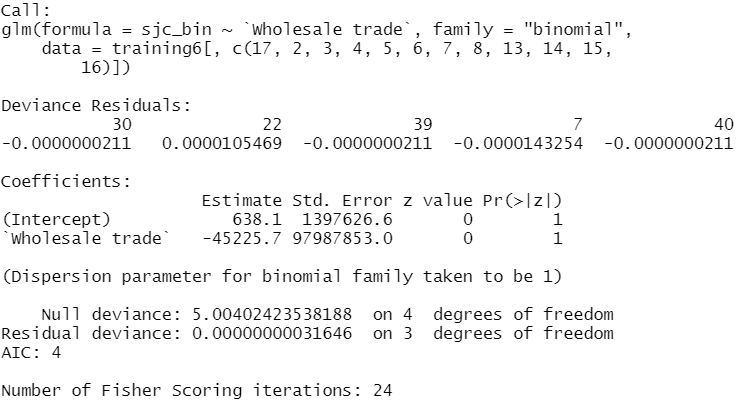
**Model for South Region:**



**AUC: 0.75**

In the case of the logistic regression conducted for the South region, after running stepwise variable selection, our final model found that ‘Information’, ‘Some college, no degree’, and ‘Average household size’ were the best predictors in combination with one another. In terms of the Area Under the Curve (AUC) for this particular model, the final model had an AUC of 0.75.

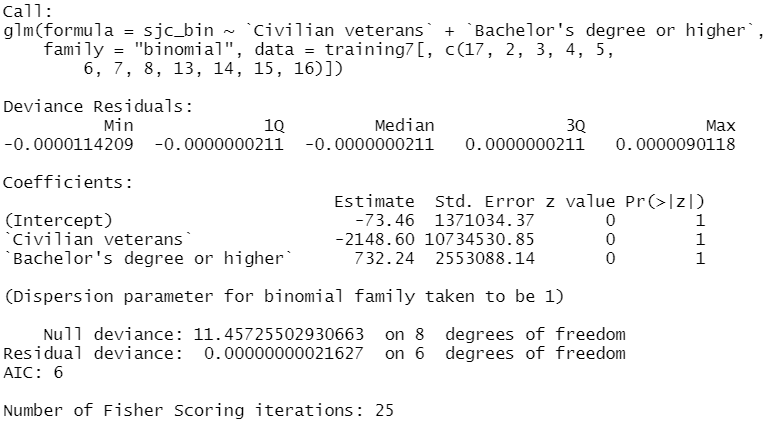
**Model for Northeast Region:**



**AUC: 1**

In the case of the logistic regression conducted for the Northeast region, after running stepwise variable selection, our final model found that just ‘Wholesale trade’ was required for the best predictive outcome. In terms of the Area Under the Curve (AUC) for this particular model, the final model had an AUC of 1.

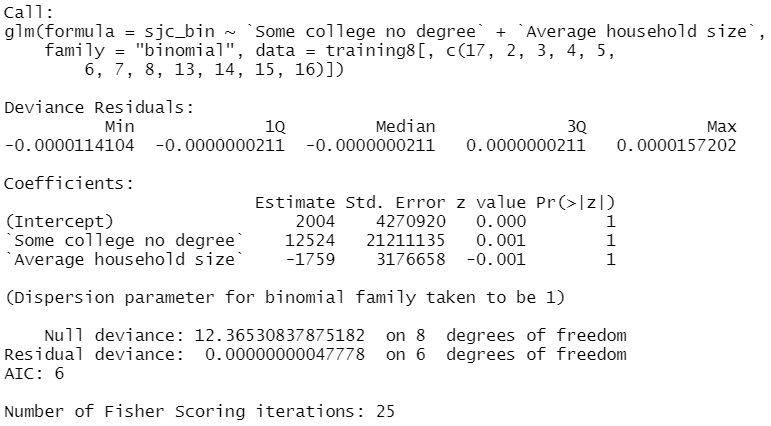
**Model for the West Region:**



**AUC: 1**

In the case of the logistic regression conducted for the West region, after running stepwise variable selection, our final model found that Civilian veterans, and ‘Bachelor’s degree or higher’ were the best predictors in combination with one another. In terms of the Area Under the Curve (AUC) for this particular model, the final model had an AUC of 1.

**Model for Midwest Region:**

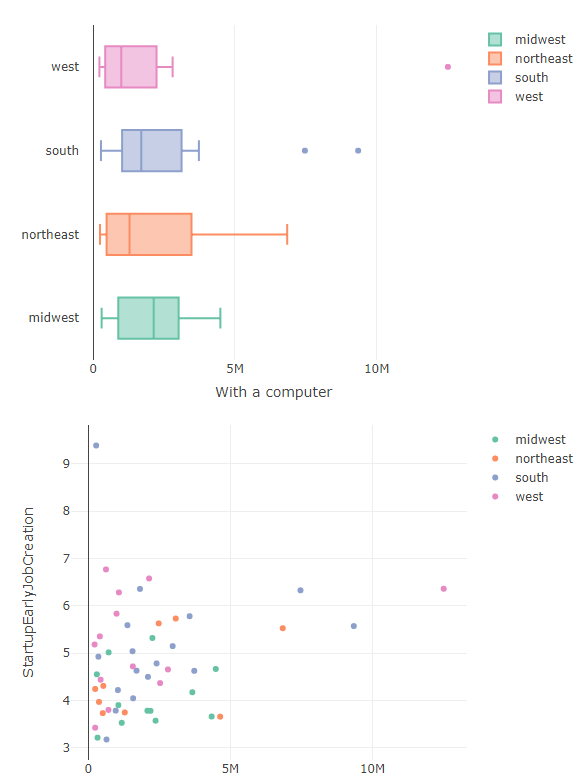


**AUC: 0.5**

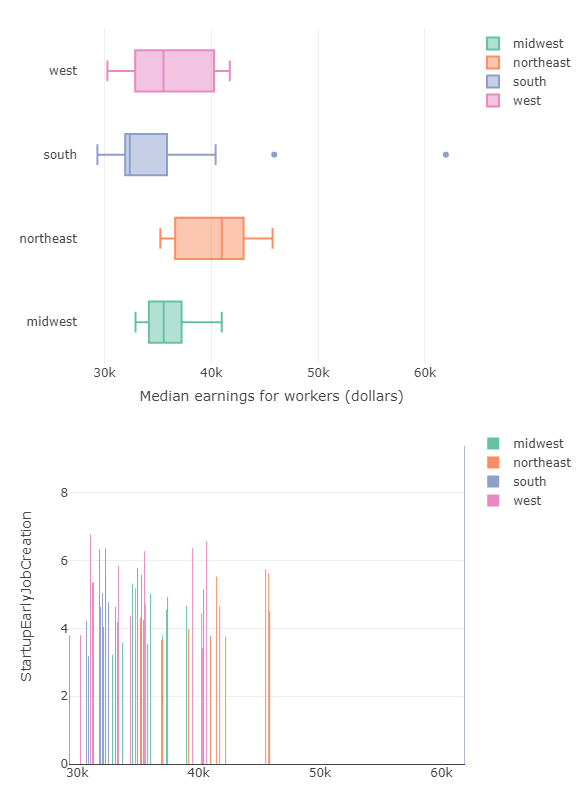
In the case of the logistic regression conducted for the Midwest region, after running stepwise variable selection, our final model found that ‘Some college, no degree’, and ‘Average household size’ were the best predictors in combination with one another. In terms of the Area Under the Curve (AUC) for this particular model, the final model had an AUC of 0.5.

**3.7 Shiny App and Visualizations**

The next step in this report was to visualize some of the variables contained within the regression models. This was done to better understand the data and see if it conflicted with any of the models. Data visualization was done using plotly and several different plots were utilized. These plots include a box plot, a scatter plot, a bar chart, and a histogram. Two graphics were presented at a single time to enable easy comparison. Before showing the visualizations, a choice was made to keep the box plot consistent, and not allow that visualization to be changed to a different type of plot to allow for easy comparison between the regional boxplot and the by state visualizations. Additionally, the variable Startup Early Job creation was always the Y axis on the secondary plot. Each plot shares the same x axis, so it is only denoted once. The variables eligible for selection were limited to the variables used in the regressions.

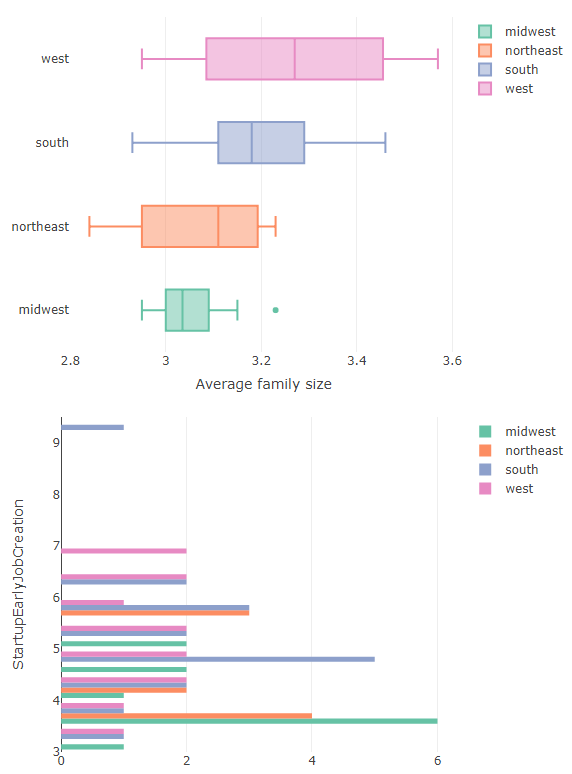
*First Visualization: Computer usage vs. Startup Early Job Creation by Region & State*

The first graphic in this visualization is a boxplot plotting the average number of individuals with computers in a given region. The second graphic is a scatter plot charting Startup Early Job Creation against the number of individuals with computer access in a given state, colored by region. This visualization shows that while the midwestern region has the most computer users on average, it does not boast significantly high startup job creation, seeing as the vast majority of points on the scatter plot fall below 4 jobs created. This tells us that while the midwest has significant access to computers, it is not effectively utilizing computer usage to create jobs within startups, compared to other regions. Another interesting trend presents itself when observing the western region, the home of Silicon Valley. Intuition suggests this region would have high amounts of computer access, however it is the lowest on average, barring the exception of California. This state is represented by the furthest right point on the second graphic, indicating that this state has a large amount of computer access, and also effectively utilizes computer usage to generate startup jobs.

*Second Visualization: Median earnings vs Startup Early Job Creation by Region & State*

This second visualization explores the relationship between median worker income and startup early job creation. The first graphic is a boxplot charting the average median income for each region and the second visualization is a bar chart plotting median income on the x axis followed by startup early job creation on the y axis. This visualization shows an interesting outlier in Washing DC having the highest income, as well as the highest job creation rate, which is interesting considering it is in the southern region, which is ranked last for average income according to the first graphic. Even though it is such an outlier, it does relatively little to bring the average of the south up, and accordingly, is kept in the dataset to allow for additional analysis. This visualization is important because it shows how even though a state may appear to be an outlier, its overall effect on the analysis of this report is unaffected by its inclusion. With or without DC, the south would still remain in its low position relative to the rest of the regions. This helps this report's analysis by allowing us to not worry about having to deal with outliers in the regression analysis, and further corroborates earlier findings.

*Final Visualization: Average Family Size vs Startup Early Job Creation by Region & State*

This last visualization explores the relationship between family size and startup early job creation using a histogram. Please note, the x axis for the histogram (the second graphic) is not average family size, rather a count of states in a given region that share a similar SJC score. These visualizations are important in understanding the relationship between family size and entrepreneurial success in a given state. Intuition would suggest that a larger family would mean a larger support network of mentors and inspirations, so it is reassuring to see that the region with the largest average family size, the west, is also fairly entrepreneurially successful. The western region occupies two out of the top three SJC score bins, meaning that the western region is experiencing a high level of job creation while having the largest number of family members on average. While this variable was not the most valuable in the regression analysis, it still provides useful in drawing generalizations about the family life of a successful entrepreneurial endeavor. 

*Final Visualization Notes:*

Having presented three visualizations out of the possible numerous combinations, it wasn’t easy to narrow down the choice of which visualization to present. However, the shiny app developed provides an easy way to visualize all combinations of variables used in the regression analysis. The app allows for an interactive visualization of 11 variables in 4 different plot types.

**4 Conclusion**

Overall, the KESE, Economic and Social datasets show information about the relationship between startup job creation and various social and economic census data. The creation of visualizations and regression models of the dataset reveals the relationship between these variables and the sjc variable.

In order to create the logistic and linear regression models, we needed to select variables using stepwise selection. There were common variables across all of the linear and logistic regression models including “Some college, no degree” and “Self-Employed”. The average adjusted r-squared values across all linear regression models was 0.68 which tells us that there is a fair correlation between the variables we selected and startup early job creation. Moreover, the average AUC for all logistic regression models was 0.81. This is a clear indication that the logistic regression models are fairly able to distinguish between positive and negative classes. It’s imperative to note that it was difficult to build these models with very few observations as we only had 50 states to work with. Also, there were signs of statistical significance (if we used a 95% confidence level) in some variables within particular models. The visualizations also offered unique insights. The first visualization revealed that states that have a large amount of computer access, can potentially utilize computer usage to generate startup jobs. The last visualization revealed that western region is experiencing a high level of job creation while having the largest number of family members on average. The shiny app provides an easy way to visualize all combinations of variables used in the regression analysis, offering a multitude of visualizations and variable plot types.

In conclusion, this dataset can be used to show the relationship between social and economic variables with the startup job creation of an area. For further analysis, more information can be collected from performing further regression analysis to show specific relationships between certain variables, and other different types of visualizations could be used to create interesting insight

**5 Appendix**

We created a Data Dictionary because it is helpful to catalog and communicate the structure and content of the data to the viewer. It provides descriptive details about each variable found in the dataset. The table below is the Data Dictionary for all three datasets:

**Kauffman Early Stage Entrepreneurship Data Dictionary**

| **Variable Name** | **Variable Data-Type\*** | **Units\*\*** | **Description** |
| --- | --- | --- | --- |
| fips | Integer | Number | Number pertaining to the “name” variable. |
| name | Character | Text | Name of the region/state. |
| type | Character | Text | Type of category pertaining to U.S. as a whole (e.g. Race, Sex, Total, Age) |
| category | Character | Text | Subcategory of “type” variable (e.g. Ages 35-40, Black) |
| year | Integer | Unit of Time | Year data was collected in (each state has data collected from 1996 to 2021). |
| rne | Numeric | Percentage | **RATE OF NEW ENTREPRENEURS**  Percent of population that starts a new business |
| ose | Numeric | Percentage | **OPPORTUNITY SHARE OF NEW ENTREPRENEURS**  Percent of new entrepreneurs who created a business by choice instead of necessity |
| sjc | Numeric | Number of Jobs | **STARTUP EARLY JOB CREATION**  Average number of jobs created by startups in their first year (normalized by population) |
| ssr | Numeric | Percentage | **STARTUP EARLY SURVIVAL RATE**  Percent of startups that are still active after one year |
| zindex | Numeric | Ratio (index normally does not have a unit) | **KAUFFMAN EARLY-STAGE ENTREPRENEURSHIP (KESE) INDEX**  Summary index that reflects entrepreneurial activity in the United States |

*\*Data type is written as how R reads them*

*\*\*Percentages are “rates per 100”, the units are generally what the percentage entails, and an index normally does not have a unit as it can be a ratio*

**ACS 2019 Economic Data - for States Transposed.xlsx**

| **Variable Name** | **Attribute Type** | **Units** | **Description** |
| --- | --- | --- | --- |
| N/A (Blank heading but it is States) | Character | Text (referring to location) | State located in the USA |
| Population 16 years and over | Character | Number of People | Total Number of people that are 16 years of age and older |
| In Labor Force | Character | Number of People | Total number of people that are in the labor force |
| Civilian Labor Force | Character | Number of people | Employed or unemployed individuals, who are not active-duty military personnel, institutionalized individuals, agricultural workers, and federal government employees. Retirees, handicapped and discouraged workers are also not part of the civilian labor force. |
| Employed | Character | Number of people | Number of people that have a paid job |
| Unemployed | Character | Number of people | Number of People who are jobless, looking for a job, and available for work. |
| Armed Forces | Character | Number of people | Number of people who are classified as being in the armed forces |
| Not in Labor Force | Character | Number of people | Number of people who are neither employed, or unemployed. |
| Civilian labor force | Character | Number of people | Employed or unemployed individuals, who are not active-duty military personnel, institutionalized individuals, agricultural workers, and federal government employees. Retirees, handicapped and discouraged workers are also not part of the civilian labor force. |
| Unemployment Rate | Character | Rate (percentage, subset of people / set of people) | Dividing the number of unemployed persons by the number of persons in the labor force (employed or unemployed) and multiplying that figure by 100. |
| Females 16 years and over | Character | Number of people | Total number of females who are 16 years of age and older |
| In labor force | Character | Number of people | Number of women that are in the labor force |
| Employed | Character | Number of people | Number of women employed |
| Own children of the householder under 6 years | Character | Number of people | Number of households that have children that are under 6 years of age |
| All parents in family in Labor Force | Character | Number of people | Number of parents in family that are in the labor force with children under 6 years of age |
| Own children of the householder 6 to 17 years | Character | Number of people | Number of households that own children from 6 to 17 years of age |
| All parents in family in labor force | Character | Number of people | Number of parents in a family that are in the labor force with children from 6 to 17 years of age |
| Workers 16 years and over | Character | Number of people | Number of workers that are 16 years of age and over |
| Car, truck, van - drove alone | Character | Number of people | Number of people that used a vehicle to drive alone to commute to work |
| Car, truck,van - carpooled | Character | Number of people | Number of people that used a vehicle to carpool to commute to work |
| Public transportation (excluding taxicabs) | Character | Number of people | Number of people that used public transportation to commute to work |

| Walked | Character | Number of people | Number of people that walked to commute to work |
| --- | --- | --- | --- |
| Other means | Character | Number of people | Number of people that used other means to commute to work |
| Worked from Home | Character | Number of people | Number of people that worked from home (probably did not need to use a mode of transportation or walk to work) |
| Mean travel time to work (minutes) | Character | Minutes | The average travel time it took for a person to get to work |
| Civilian employed population 16 years and over | Character | Number of people | The number of civilians that are employed and are 16 years of age and older |
| Management, business, science, and arts occupations | Character | Number of people | The number of people that are in management, business, science and arts occupations |
| Service occupations | Character | Number of people | The number of people that are in the service occupation |
| Sales and office occupations | Character | Number of people | The number of people that are in a sales and office occupation |
| Natural resources, construction, and maintenance occupations | Character | Number of people | The number of people that are in natural resources, construction, and maintenance occupations |
| Production, transportation, and material moving occupations | Character | Number of people | The number of people that are in production, transportation, and material moving occupations |
| Civilian employed population 16 years and over | Character | Number of people | The number of people that are civilians and are employed and are 16 years of age and over |
| Agriculture, forestry, fishing and hunting, and mining | Character | Number of people | The number of people that are in the agriculture, forestry, fishing and hunting, and mining industry |
| Construction | Character | Number of people | The number of people that are in the construction industry |
| Manufacturing | Character | Number of people | The number of people that are in the manufacturing industry |
| Wholesale trade | Character | Number of people | The number of people that are in the wholesale trade industry |
| Retail trade | Character | Number of people | The number of people that are in the retail trade industry |
| Transportation and warehousing, and utilities | Character | Number of people | The number of people that are in transportation and warehousing, and utilities industry |
| Information | Character | Number of people | The number of people that are in the information industry |
| Finance and insurance, and real estate and rental and leasing | Character | Number of people | The number of people that are in finance and insurance, and real estate and rental and leasing industry |
| Professional, scientific, and management, and administrative and waste management services | Character | Number of people | The number of people that are professional, scientific, and management, and administrative and waste management services industry. |

| Educational services, and health care and social assistance | Character | Number of people | The number of people that are in educational services, and health care and social assistance industry |
| --- | --- | --- | --- |
| Arts, entertainment, and recreation, and accommodation and food services | Character | Number of people | The number of people that are in arts, entertainment, and recreation, and accommodation and food services industry |
| Other services, except public administration | Character | Number of people | The number of people that are in other services, except public administration industry |
| Public administration | Character | Number of people | The number of people that are in public administration industry |
| Civilian employed population 16 years and over | Character | Number of people | The number of people that are in the civilian employed population and are 16 years of age and over |
| Private wage and salary workers | Character | Number of people | The total number of private wage and salary workers |
| Government workers | Character | Number of people | The total number of Government workers |
| Self-employed in own not incorporated business workers | Character | Number of people | The total number of Self-employed in own not incorporated business workers |
| Unpaid family workers | Character | Number of people | The total number of Unpaid family workers |
| Total households | Character | Number of households | The total number of households |
| Less than $10,000... | Character | Number of people | The total number of people with income and benefits totalling less than $10,000 |
| $10,000 to $14,999... | Character | Number of people | The total number of people with income and benefits that is between $10,000 and $14,999 |
| $15,000 to $24,999 | Character | Number of people | The total number of people with income and benefits that is between $15,000 and $24,999 |
| $25,000 to $34,999 | Character | Number of people | The total number of people with income and benefits that is between $25,000 and $34,999 |
| $35,000 to $49,999 | Character | Number of people | The total number of people with income and benefits that is between $35,000 and $49,999 |
| $50,000 to $74,999 | Character | Number of people | The total number of people with income and benefits that is between $50,000 and $74,999 |
| $75,000 to $99,999 | Character | Number of people | The total number of people with income and benefits that is between $75,000 to $99,999 |
| $100,000 to $149,999 | Character | Number of people | The total number of people with income and benefits that is between $100,000 to $149,999 |
| $150,000 to $199,999 | Character | Number of people | The total number of people with income and benefits that is between $150,000 to $199,999 |
| $200,000 or more | Character | Number of people | The total number of people with income and benefits that is between $100,000 to $149,999 |

| Median household income (dollars) | Character | Dollars $ | The median household income in dollars |
| --- | --- | --- | --- |
| Mean household income (dollars) | Character | Dollars $ | The mean household income in dollars |
| With earnings | Character | Number of people | The number of people that have income with earnings. |
| Mean earnings (dollars) | Character | Dollars $ | The average earnings in dollars |
| With Social Security | Character | Number of people | The number of people with social security |
| Mean Social Security income (dollars) | Character | Dollars $ | The average social security income in dollars |
| With retirement income | Character | Number of people | The number of people that have retirement income |
| Mean retirement income (dollars) | Character | Dollars $ | The average retirement income in dollars |
| With Supplemental Security Income | Character | Number of people | The number of people with supplemental security income |
| Mean Supplemental Security Income (dollars) | Character | Dollars $ | The average supplemental security income in dollars |
| With cash public assistance income | Character | Number of people | The number of people with cash public assistance income |
| Mean cash public assistance income (dollars) | Character | Dollars $ | The average cash public assistance income in dollars |
| With Food Stamp/SNAP benefits in the past 12 months | Character | Number of people | The number of people with food stamp/SNAP benefits in the past twelve months |
| Families | Character | Number of families | The total number of families |
| Less than $10,000 | Character | Number of people | Families with income less than $10,000 |
| $10,000 to $14,999 | Character | Number of people | Families with income between $10,000 to $14,999 |
| $15,000 to $24,999 | Character | Number of people | Families with income between $15,000 to $24,999 |
| $25,000 to $34,999 | Character | Number of people | Families with income between $25,000 to $34,999 |
| $35,000 to $49,999 | Character | Number of people | Families with income between $35,000 to $49,999 |
| $50,000 to $74,999 | Character | Number of people | Families with income between $50,000 to $74,999 |

| $75,000 to $99,999 | Character | Number of people | Families with income between $75,000 to $99,999 |
| --- | --- | --- | --- |
| $100,000 to $149,999 | Character | Number of people | Families with income between $100,000 to $149,999 |
| $150,000 to $199,999 | Character | Number of people | Families with income between $150,000 to $199,999 |
| $200,000 or more | Character | Number of people | Families with income $200,000 or more |
| Median family income (dollars) | Character | Dollars $ | The median family income in dollars |
| Mean family income (dollars) | Character | Dollars $ | The average family income in dollars |
| Per capita income (dollars) | Character | Dollars $ | A measure of the amount of money earned per person |
| Nonfamily households | Character | Number of households | Number of households that are non-family classified |
| Median nonfamily income (dollars) | Character | Dollars $ | The median non-family income in dollars |
| Mean nonfamily income (dollars) | Character | Dollars $ | The average non-family income in dollars |
| Median earnings for workers (dollars) | Character | Dollars $ | The median earnings for workers in dollars |
| Median earnings for male full-time, year-round workers (dollars) | Character | Dollars $ | The median earnings for male full-time, year-round workers (dollars) |
| Median earnings for female full-time, year-round workers (dollars) | Character | Dollars $ | The Median earnings for female full-time, year-round workers (dollars) |
| Civilian noninstitutionalized population | Character | Number of people | People 16 years of age and older residing in the 50 States and the District of Columbia who are not inmates of institutions, and who are not on active duty in the Armed Forces. |
| With health insurance coverage | Character | Number of people | The number of people that have health insurance coverage |
| With private health insurance | Character | Number of people | The number of people with private health insurance |
| With public coverage | Character | Number of people | The number of people with public coverage |
| With no health insurance coverage | Character | Number of people | The number of people with no health insurance coverage |
| Civilian noninstitutionalized population under 19 years | Character | Number of people | The number of civilian noninstitutionalized population under 19 years of age |
| No health insurance coverage | Character | Number of people | The number of civilian noninstitutionalized population under 19 years of age with no health insurance coverage |

| Civilian noninstitutionalized population 19 to 64 years | Character | Number of people | The number of civilian noninstitutionalized population 19 to 64 years |
| --- | --- | --- | --- |
| In labor force | Character | Number of people | The number of people in the labor force |
| Employed: | Character | Number of people | The number of people that are employed |
| With health insurance coverage | Character | Number of people | The number of people with health insurance coverage |
| With private health insurance | Character | Number of people | The number of people with private health insurance |
| With public coverage | Character | Number of people | The number of people with public coverage |
| No health insurance coverage | Character | Number of people | The number of people with no health insurance coverage |
| Unemployed | Character | Number of people | The number of people that are unemployed |
| With health insurance coverage | Character | Number of people | The number of people with health insurance coverage |
| With health insurance coverage | Character | Number of people | The number of people with health insurance coverage |
| With public coverage | Character | Number of people | The number of people with public coverage |
| No health insurance coverage | Character | Number of people | The number of people with no health insurance coverage |
| Not in labor force: | Character | Number of people | The number of people not in the labor force |
| With health insurance coverage | Character | Number of people | The number of people with health insurance coverage |
| With private health insurance | Character | Number of people | The number of people with private health insurance |
| With public coverage | Character | Number of people | The number of people with public coverage |
| No health insurance coverage | Character | Number of people | The number of people with no health insurance coverage |

**ACS 2019 Social Characteristics Data - for States Transposed.xlsx**

| **Variable Name** | **Attribute Type** | **Units** | **Description** |
| --- | --- | --- | --- |
| State | Character | Text (referring to location) | State located in the USA |
| Total households. | Character | Number of households | Number of total households in the US |
| Married-couple family | Character | Number of married-couple families | The total number of married-couple families |
| With own children of the householder under 18 years | Character | Number of households | Number of households with own children of the householder that is under 18 years of age (belonging to a married-couple family) |
| Cohabiting couple household | Character | Number of households | The number of cohabiting couple households |
| With own children of the householder under 18 years | Character | Number of households | Number of households with own children of the householder that's under 18 years (belonging to a cohabiting couple) |
| Male householder, no spouse/partner present | Character | Number of males | The total number of male householders with no spouse or a partner present (single) |
| With own children of the householder under 18 years | Character | Number of households | The total number of single male householders with no spouse or a partner present that own children under 18 years of age |
| Householder living alone | Character | Number of people | The total number of lonely householders |
| 65 years and over | Character | Number of people | The total number of householders that are 65 years of age and older |
| Female householder, no spouse/partner present | Character | Number of female | The total number of single female householders with no spouse and partner present |
| With own children of the householder under 18 years | Character | Number of households | Number of single females with own children of the householder that are under 18 years of age |
| Householder living alone. | Character | Number of people | Number of female householders that live alone |
| 65 years and over. | Character | Number of people | Number of female householders that are 65 years of age and older |
| Households with one or more people under 18 years | Character | Number of households | The total number of households with one or more people under 18 years of age |
| Households with one or more people 65 years and over | Character | Number of households | The total number of households with one or more people 65 years of age and over |
| Average household size | Character | Number of people | The average number of people in an average household |
| Average family size | Character | Number of people | The average number of people in a family |
| Population in households | Character | Number of people | The total number of people that live in households |

| Householder | Character | Number of people | The total number of people that are householders |
| --- | --- | --- | --- |
| Spouse | Character | Number of people | The total number of people that are spouses |
| Unmarried partner | Character | Number of people | The total number of people that have an unmarried partner |
| Child | Character | Number of people | The total number of people that have a child |
| Other relatives | Character | Number of people | The total number that have other relatives |
| Other nonrelatives | Character | Number of people | The total number of people that have other nonrelatives |
| Males 15 years and over | Character | Number of males | The total number of males that are 15 years and over |
| Never married | Character | Number of males | The total number of males that were never married |
| Now married, except separated | Character | Number of males | The total number of people that are now married, except for those that are separated from each other |
| Separated | Character | Number of males | The total number of males that are separated |
| Widowed | Character | Number of males | The total number of males that are widowed |
| Divorced | Character | Number of males | The total number of males that are divorced |
| Females 15 years and over | Character | Number of females | The total number of females that are 15 years and older |
| Never married | Character | Number of females | The total number of females that were never married |
| Now married, except separated | Character | Number of females | The total number of females that were now married except separated |
| Separated | Character | Number of females | The total number of females that are separated |
| Widowed | Character | Number of females | The total number of females that are widowed |
| Divorced | Character | Number of females | The total number of females that are divorced |
| Number of women 15 to 50 years old who had a birth in the past 12 months | Character | Number of females | The total number of women 15 to 50 years old who had a birth in the past 12 months |
| Unmarried women (widowed, divorced, and never married) | Character | Number of females | The total number of unmarried women who are widowed, divorced, and never married. |

| Per 1,000 unmarried women | Character | Number of people per 1000 | The number of women per 1000 that are unmarried |
| --- | --- | --- | --- |
| Per 1,000 women 15 to 50 years old | Character | Number of people per 1000 | The number of women per 1000 who are 15 to 50 years old |
| Per 1,000 women 15 to 19 years old | Character | Number of people per 1000 | The number of women per 1000 who are 15 to 19 years old |
| Per 1,000 women 20 to 34 years old | Character | Number of people per 1000 | The number of women per 1000 who are 20 to 34 years old |
| Per 1,000 women 35 to 50 years old | Character | Number of people per 1000 | The number of women per 1000 who are 35 to 50 years old |
| Number of grandparents living with own grandchildren under 18 years | Character | Number of grandparents (people) | The number of grandparents living with their own grandchildren who are under 18 years old |
| Grandparents responsible for grandchildren | Character | Number of people | The total number of grandparents who are responsible for their grandchildren |
| Years responsible for grandchildren \*\* | Character | Years (unit of time) | The number of years in which grandparents have been responsible for their grandchildren |
| Less than 1 year | Character | Number of grandparents | Number of grandparents who have been responsible for their grandchildren for less than a year |
| 1 or 2 years | Character | Number of grandparents | Number of grandparents who have been responsible for their grandchildren for 1 to 2 years |
| 3 or 4 years | Character | Number of grandparents | Number of grandparents who have been responsible for their grandchildren for 3 to 4 years |
| 5 or more years | Character | Number of grandparents | Number of grandparents who have been responsible for their grandchildren for over 5 years |
| Number of grandparents responsible for own grandchildren under 18 years | Character | Number of grandparents | The total number of grandparents who are responsible for their grandchildren under 18 years |
| Who are female | Character | Number of grandparents | The total number of grandparents who are female |
| Who are married | Character | Number of grandparents | The total number of grandparents who are married |
| Population 3 years and over enrolled in school | Character | Number of enrolled | Number of people enrolled in school who are over 3 years old |
| Nursery school, preschool | Character | Number of enrolled | The number of people enrolled in nursery school, or preschool |
| Kindergarten | Character | Number of enrolled | The number of people enrolled in kindergarten |
| Elementary school (grades 1-8) | Character | Number of enrolled | The number of people enrolled in elementary school (grades 1 through 8) |
| High school (grades 9-12) | Character | Number of enrolled | Number of people enrolled in high school (grades 9 through 12) |

| College or graduate school | Character | Number of enrolled | Number of people enrolled in college or graduate school |
| --- | --- | --- | --- |
| Population 25 years and over | Character | Number of people | Number of people who is 25 years and older |
| Less than 9th grade | Character | Number of people with educational attainment | Number of people who have completed school less than ninth grade |
| 9th to 12th grade, no diploma | Character | Number of people with educational attainment | Number of people who are in high school but had no diploma |
| High school graduate (includes equivalency) | Character | Number of people with educational attainment | Number of people who are high school graduates and includes equivalency |
| Some college, no degree | Character | Number of people with educational attainment | Number of people who went to some college but got no degree |
| Associate's degree | Character | Number of people with educational attainment | Number of people who have an associates degree |
| Bachelor's degree | Character | Number of people with educational attainment | Number of people who got a bachelor’s degree |
| Graduate or professional degree | Character | Number of people with educational attainment | Number of people who got a graduate or professional degree |
| High school graduate or higher | Character | Number of people with educational attainment | Number of people who are high school graduates or higher |
| Bachelor's degree or higher | Character | Number of people with educational attainment | Number of people who have a bachelor’s degree or higher |
| Civilian population 18 years and over | Character | Number of people | The total number of civilians who are 18 years or older |
| Civilian veterans | Character | Number of veterans | Number of civilians who are veterans |
| Total Civilian Noninstitutionalized Population | Character | Number of people | Number of people 16 years of age and older residing in the 50 States and the District of Columbia who are not inmates of institutions, and who are not on active duty in the Armed Forces. |
| With a disability | Character | Number of people | Number of people with a disability who are in the civilian noninstitutionalized population |
| Under 18 years | Character | Number of people | Total number of people under 18 years of age |
| With a disability. | Character | Number of people | Total number of people with a disability under 18 years of age |
| 18 to 64 years | Character | Number of people | Total number of people who are 18 to 64 years old |
| With a disability. | Character | Number of people | Total number of 18 to 64 year olds with a disability |
| 65 years and over | Character | Number of people | Total number of people who are 65 years and over |

| With a disability | Character | Number of people | Total number of people who are 65 years and over with a disability |
| --- | --- | --- | --- |
| Population 1 year and over | Character | Number of people | The total number of people who had a residence 1 year and over ago |
| Same house | Character | Number of people | The total number of people who had the same house 1 year and over ago |
| Different house in the U.S. | Character | Number of people | The total number of people had a different house in the US 1 year and over ago |
| Same county | Character | Number of people | The total number of people that had a house in the same county 1 year and over ago |
| Different county | Character | Number of people | The total number of people that had a house in a different county 1 year and over ago |
| Same state | Character | Number of people | The total number of people who had a house in the same state 1 year and over ago |
| Different state | Character | Number of people | The total number of people who had a house in the different state 1 year and over ago |
| Abroad | Character | Number of people | The total number of people who had a house abroad 1 year and over ago |
| Total Population | Character | Number of people | The total number of people born |
| Native | Character | Number of people | The total number of native born |
| Born in US | Character | Number of people | The total number of people born in the US |
| State of Residence | Character | Number of people | The total number of people born in their current state of residence |
| Different State | Character | Number of people | The total number of people born in a different state than their current residence |
| Born in Puerto Rico, U.S. Island areas, or born abroad to American parent(s) | Character | Number of people | The total number of people born in puerto rico, US island areas, or born abroad to American parent(s) |
| Foreign Born | Character | Number of people | The total number of foreign born people |
| Foreign-born population | Character | Number of people | The total number of foreign born people |
| Naturalized U.S. citizen | Character | Number of people | The total number of naturalized US Citizens |
| Not US Citizen | Character | Number of people | The total number of Non-US citizens |
| Population born outside United States | Character | Number of people | The number of people born outside US |

| Native | Character | Number of people | The number of natives |
| --- | --- | --- | --- |
| Entered 2010 or Later | Character | Number of people | The number of natives entering US 2010 or Later |
| Entered before 2010 | Character | Number of people | The number of natives entered before 2010 |
| Foreign Born | Character | Number of people | The number of foreign born people |
| Entered 2010 or Later | Character | Number of people | The number of foreign born people who entered 2010 or later in US |
| Foreign-born population, excluding population born at sea | Character | Number of people | The number of foreign born population, excluding the population born at sea |
| Europe | Character | Number of people | The number of people born in Europe |
| Asia | Character | Number of people | The number of people born in Asia |
| Africa | Character | Number of people | The number of people born in Africa |
| Oceania | Character | Number of people | The number of people born in Oceania |
| Latin America | Character | Number of people | The number of people born in Latin America |
| Northern America | Character | Number of people | The number of people born in Northern America |
| Population 5 years and over | Character | Number of people | The number of people who are 5 years and older |
| English only | Character | Number of people | The number of people whose language is english only |
| Language other than English | Character | Number of people | The number of people who speak a language other than english |
| Speak English less than "very well" | Character | Number of people | The number of people who speak a language other than language and can speak english less than very well |
| Spanish | Character | Number of people | The number of people who speak spanish |
| Speak English less than "very well" | Character | Number of people | The number of people who speak spanish and can speak english less than very well |
| Other Indo-European languages | Character | Number of people | The number of people who speak other indo-european languages |
| Speak English less than "very well" | Character | Number of people | The number of people who speak other indo-european languages and can english less than very well |

| Asian and Pacific Islander languages | Character | Number of people | The total number of people who speak asian and pacific islander languages |
| --- | --- | --- | --- |
| Speak English less than "very well" | Character | Number of people | The total number of people who speak asian and pacific islander languages and speak english less than very well |
| Other languages | Character | Number of people | The total number of people who speak other languages |
| Speak English less than "very well" | Character | Number of people | The total number of people who speak other languages and can speak english less than very well |
| Total population | Character | Number of people | The total population of people in US |
| American | Character | Number of people | The total number of people with american ancestry |
| Arab | Character | Number of people | The total number of people with arab ancestry |
| Czech | Character | Number of people | The total number of people with czech ancestry |
| Danish | Character | Number of people | The total number of people with danish ancestry |
| Dutch | Character | Number of people | The total number of people with dutch ancestry |
| English | Character | Number of people | The total number with english ancestry |
| French (except Basque) | Character | Number of people | The total number of french (except basque) ancestry |
| French Canadian | Character | Number of people | The total number of people with french canadian ancestry |
| German | Character | Number of people | The total number of people with german ancestry |
| Greek | Character | Number of people | The total number of people with Greek ancestry |
| Hungarian | Character | Number of people | The total number of people with Hungarian ancestry |
| Irish | Character | Number of people | The total number of people with Irish ancestry |
| Italian | Character | Number of people | The total number of people with Italian ancestry |
| Lithuanian | Character | Number of people | The total number of people with Lithuanian ancestry |
| Norwegian | Character | Number of people | The total number of people with Norwegian ancestry |

| Polish | Character | Number of people | The total number of people with Polish ancestry |
| --- | --- | --- | --- |
| Portuguese | Character | Number of people | The total number of people with Portuguese ancestry |
| Russian | Character | Number of people | The total number people of Russian ancestry |
| Scotch Irish | Character | Number of people | The total number of people with scotch irish ancestry |
| Scottish | Character | Number of people | The total number of people with scottish ancestry |
| Slovak | Character | Number of people | The total number of people with slovak ancestry |
| Subsaharan African | Character | Number of people | The total number of people with sub saharan african ancestry |
| Swedish | Character | Number of people | The total number of people with swedish ancestry |
| Swiss | Character | Number of people | The total number of people with swiss ancestry |
| Ukrainian | Character | Number of people | The total number of people with ukrainian ancestry |
| West Indian (excluding Hispanic origin groups) | Character | Number of people | The total number of people with west indian excluding hispanic origin groups ancestry |
| Total households | Character | Number of people | The total number of households |
| With a computer | Character | Number of people | The total number of households with a computer |
| With a broadband Internet subscription | Character | Number of people | The total number of households with a broadband Internet subscription |