Iowa County Population and Demographic Information
Final Report
Carson Kirby

1. Introduction

According to a PEW Research[1] study from 2018, nearly half of the rural counties in the United States had a lower population at that time than they did at the time of the 2000 Census. What does the overall movement of the population in the state of Iowa indicate about other factors that describe the population? Or to frame it in another way, how do other demographic factors like income and unemployment rate impact the growth or decline in a county's population? Are there certain key factors that can be used as indicators to predict how a county's population will change when counted in future censuses? I am particularly interested in if there is any relationship between a county's rural/urban classification and its population growth. Do national trends hold true in a state like Iowa that has an abundance of rural areas?

In this project, I used Iowa county population data from an Iowa State University[2] and other demographic information like unemployment rate and education from the National Institutes of Health[3]. Additionally, I got information about the rural/urban status of counties from the Iowa Data Center[4] I joined the data I obtained from these websites into a single dataframe.

2. Data

This project uses two main sources of data: Census data from the Iowa State University and county demographic data from the National Health Institute.

2.1 County Census Information

I collected data from the Iowa State University website, specifically the page dedicated to the Iowa Community Indicators Program. The site contained a table that displayed official population data from the U.S. Census Bureau. I chose to use web scraping to gather the data

[1] https://www.pewresearch.org/social-trends/2018/05/22/demographic-and-economic-trends-in-urban-suburban-and-rural-communities/
[2] https://www.icip.iastate.edu/tables/population/census-counties

[3]https://hdpulse.nimhd.nih.gov/data-portal/social/map?age=001&age_options=ageall_1&demo=00010&demo_options=income_3&race=00&race_options=race_7&sex=0&sex_options=sexboth_1&socialtopic=030&socialtopic_options=social_6&statefips=19&statefips_options=area_states
[4] https://www.iowadatacenter.org/index.php/data-by-source/decennial-census/urban-and-rural-population

from the website for the purposes of this project. I scraped the web table on the Iowa State website that listed the population data received at each of the decennial censuses from 1980 thru 2020. I saved this web scraping data as a single csv file titled 'county_data.csv'. I had initially found this data at another site called the Iowa Data Center. I was having trouble conducting web scraping on that data, so I switched to the Iowa State University website. Both websites contained official population data from the U.S. Census Bureau.

In the end, I created a pandas dataframe using the web scraping script I wrote. The dataframe contained the county name and the county population recorded in the years 1980, 1990, 2000, 2010, and 2020. Included in the dataframe was also a FIPS code for each county. This is a Federal Information Processing Standard code that uniquely identifies counties. Due to the focus on the state of Iowa, each county's given name is a unique identifier, so the FIPS code was not needed.

2.2 Population Demographic Information

The demographic information I used in this project was downloaded from the National Institute of Health website. It was specifically in the subsection of the website, National Institute on Minority Health and Health Disparities. This is an official branch of the U.S. Department of Health & Human Services.

I downloaded each of the datasets used for the demographic information as separate csv files ('unemployment.csv' & 'income_csv') that I imported into Jupyter Notebooks as pandas dataframes. This included demographic information for median household income and unemployment rate. These datasets simply the county name and demographic I was focused on.

The other demographic factor I found was the rural/urban indicator. I incorporated this information into my project by downloading it for the Iowa Data Center website as a csv ('rual_urban.csv'). This was the site I used throughout this project for reference. Similar to the Iowa State website, the Iowa Data Center site used information directly from the U.S. Census Bureau. This ensured that the data was consistent across the board.

The data I downloaded as csv files directly from the internet did not need much cleaning at all. The main form of cleaning I did was rename columns and delete unneeded columns. This prepared the datasets for merging.

2.3 Combining Census & Demographic Information

The task of combining the data frame containing the census data with the multiple data frames containing demographic information was straightforward. Each of the data frames had the same amount of records because I ensured all of the data I used was focused on information about Iowa counties. For this reason, it was pretty obvious that I would be joining all of the data frames on the 'County' column. To make this process as smooth as possible, I renamed all of the columns containing county names to 'County'.

I started creating my integrated dataframe by merging the 'population_change' dataframe, containing the census data, with the 'income' dataframe, containing the median household income data. I then proceeded to conduct another inner join on the unemployment and rural_urban dataframes. This resulted in the 'integrated_counties2' that I saved as a csv file. This data cleaning process was all conducted within the '001_cdkirby_final_project' notebook in Jupyter. The remainder of the project, including additional cleaning and the entirety of the analysis and visualization, was conducted in the '002_cdkirby_final_project' Jupyter notebook file.

Table 1: Data Dictionary

Column	Type	Source	Description
County	Object	Iowa State University	The name of each of Iowa's 99 counties
1980_pop	Integer	Iowa State University	The population of each Iowa county as of the 1980 U.S. Census
1990_pop	Integer	Iowa State University	The population of each Iowa county as of the 1980 U.S. Census
2000_pop	Integer	Iowa State University	The population of each Iowa county as of the 1980 U.S. Census
2010_pop	Integer	Iowa State University	The population of each Iowa county as of the 1980 U.S. Census
2020_pop	Integer	Iowa State University	The population of each Iowa county as of the 1980 U.S. Census
Median_Household_Income	Integer	National Institute of Health	The median household income in each Iowa county as of research from 2018-2022

Unemployment_rate	Float	National Institute of Health	Unemployment rate recorded in each county as of research from 2018-2022
Urban_Percent	Object	Iowa Data Center	Percent of a county classified as urban as of the 2020 Census
Rural_Percent	Object	Iowa Data Center	Percent of a county classified as rural as of the 2020 Census
40_year_growth	Float	Derived from 1980_pop and 2020_pop	Calculated variable representing the percentage of growth a county experienced from the 1980 Census to the 2020 Census
30_year_growth	Float	Derived from 1990_pop and 2020_pop	Calculated variable representing the percentage of growth a county experienced from the 1990 Census to the 2020 Census
20_year_growth	Float	Derived from 2000_pop and 2020_pop	Calculated variable representing the percentage of growth a county experienced from the 2000 Census to the 2020 Census
10_year_growth	Float	Derived from 2010_pop and 2020_pop	Calculated variable representing the percentage of growth a county experienced from the 2010 Census to the 2020 Census
Growth_Rank	Integer	Derived from the 10_year_growth column	Calculated variable that ranks each county on its 10-year population growth, with 1 being the largest growth percentage
Urban_Rank	Integer	Derived from the Urban_Percent column	Calculated variable that ranks each county based what percentage of it is classified as urban, with 1 being the most urban

3. Analysis

3.1 Urban Percentage vs Population Growth

I wanted to start by analyzing if there was any relationship between the population growth that occurred between the 2010 and 2020 censuses and the urban/rural classification of Iowa counties. This was intriguing to me because it would confirm to me if Iowa counties followed the national trend of urban counties growing at higher rates than rural counties.

I started my analysis by grouping the Counties dataframe on the 'Urban_Percent' column. I then displayed the county, its urban percentage, 10-year growth, and its growth ranking. The 'Growth_Rank' is a column I created that ranks the counties based on their rate of population growth. This indicator helps contextualize the population growth rate. I specified that the Urban_Percent column should not be ascending, allowing the highest percentage of urban areas to be listed first. Based on the results from this conditional coding, it was clear to see that there was certainly a trend occurring within the data. Four of the five fastest growing counties in the state were in the top ten for urban percentage.

	County	Urban_Percent	10_year_growth	Growth_Rank
76	Polk	95.50%	14.341678	3
6	Black Hawk	87.00%	0.041193	29
81	Scott	86.70%	5.716482	10
56	Linn	86.30%	9.029665	6
96	Woodbury	83.90%	3.688878	15
51	Johnson	83.30%	16.787641	2
16	Cerro Gordo	79.70%	-2.319313	53
24	Dallas	78.70%	50.718984	1
84	Story	74.20%	10.045565	5
29	Dickinson	73.20%	6.215876	8

Figure 1: Top 10 Urban Counties Table

To see a different perspective to answer this question, I grouped the results on the 'Rural_Percent'. To make it clear, a county has both an 'Urban_Percent' column and a 'Rural_Percent' column. These columns add up to 100% for each county respectively. I once again displayed the top 10 most rural counties in the state. The results from this grouping were not as conclusive. The 'Growth_Rank' results from the rural counties on this list did not indicate that they were declining in growth at a faster rate than some of their more urban counterparts. In fact, three of the ten most rural counties in the state experienced some level of population growth between 2010 and 2020.

To visualize that relationship between a county's urban percentage and population growth, I created a scatterplot with a trendline. I used the 'Urban_Rank' column as the x-variable. This was a column I created that indicates how each county ranks compared to other counties when looking at urban percentage. Any county ranked 63rd indicates that they are 100% rural. There were nearly 40 Iowa counties classified as completely rural. The trendline that was fit to this data represents the inverse relationship between a county's urban ranking and its population growth. In other words, as a county is classified as less urban, its level of population growth between 2010 and 2020 declined.

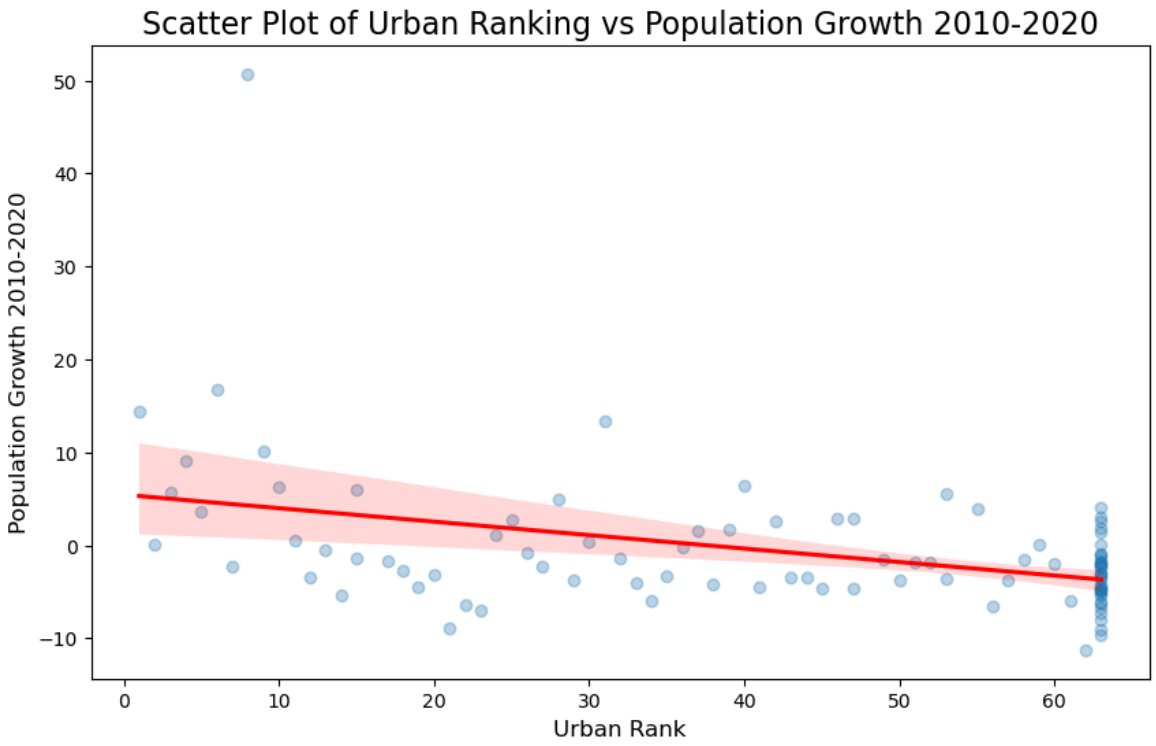

Figure 2: Scatter Plot of Urban Ranking vs Population Growth

3.2 Highest and Lowest Population Growth

The objective of my next question is very straightforward: what counties are growing the fastest? I ultimately asked this question to draw conclusions on the results based on my previous knowledge of Iowa counties. For example, of the counties with the highest growth on the list below I already associate counties with the city located there. I know Iowa City is in Johnson County, Cedar Rapids is in Linn County, and Des Moines is in Polk County. With this simple knowledge I can make the estimate that a lot of the areas with Iowa's biggest cities are the counties that are experiencing the most growth.

Based on the data frame I created, grouped by the '10_year_growth' column', it is clear that counties with the highest growth were more likely to be urban counties. All of the counties in the top ten in population growth had an urban ranking within the top half of all counties.

	County	10_year_growth	Growth_Rank	Urban_Rank
24	Dallas	50.718984	1	8
51	Johnson	16.787641	2	6
76	Polk	14.341678	3	1
90	Warren	13.365062	4	31
84	Story	10.045565	5	9
56	Linn	9.029665	6	4
83	Sioux	6.432471	7	40
29	Dickinson	6.215876	8	10
30	Dubuque	5.993401	9	15
81	Scott	5.716482	10	3

Figure 3: Top 10 Highest Population Growth Table

I then grouped the counties dataframe by 'Growth_Rank', this time organized descending to see the counties with the highest levels of population decline. In this list, I had less prior knowledge to inform me of what can be drawn from this list. Coming from a rural county myself, I recognize several names on this list as rural counties themselves. The inclusion of the 'Urban_Rank' column shows that 4 of the 10 counties with the least amount of growth were considered 100% rural.

	County	10_year_growth	Growth_Rank	Urban_Rank
35	Fremont	-11.235049	99	62
26	Decatur	-9.601514	98	63
79	Ringgold	-9.121029	97	63
31	Emmet	-8.872064	96	21
1	Adams	-8.066518	95	63
4	Audubon	-7.272430	94	63
50	Jefferson	-7.005878	93	23
86	Taylor	-6.664556	92	63
32	Fayette	-6.566092	91	56
55	Lee	-6.432993	90	22

Figure 4: Top 10 Highest Population Decline Table

3.3 Median Household Income vs Population Growth

Next, I wanted to navigate the relationship between median household income and population growth in a county. Intuitively, I predicted that households with a higher family income were likely in counties that were experiencing higher levels of population growth. I can imagine that people with higher incomes can afford to buy homes in new housing developments that contribute to population growth in those counties. Also, areas that are growing likely attract more employers to the county, therefore contributing to good-paying jobs and increased median household incomes in the county.

To put my prediction to the test, I started by grouping the county dataframe on the 'Median_Household_Income' column. I displayed the county, income, growth rank, and 10-year growth percentage. What struck me immediately about the result was the evident relationship between income and growth. Number 1 and 2 in the 'Growth_Rank' respectively, Dallas and Johnson County, are also 1 and 2 in median household income. In fact, the top 6 counties in terms of population growth all fell within the top ten counties for median household income. Of the other counties on the list, Mills County, was the only one I would classify as an outlier. It was ranked 6th in median household income, but 69th in terms of population growth.

	County	Median_Household_Income	Growth_Rank	10_year_growth
24	Dallas	130231	1	50.718984
51	Johnson	112575	2	16.787641
84	Story	109432	5	10.045565
90	Warren	107434	4	13.365062
60	Madison	105706	11	5.542445
64	Mills	100809	69	-3.818315
76	Polk	99161	3	14.341678
8	Bremer	98823	17	2.932938
7	Boone	97936	24	1.554778
56	Linn	96369	6	9.029665

Figure 5: Top 10 Counties with Highest Annual Median Household Income

Next, I altered my code slightly to produce results showing the lowest median household incomes from across the state. The growth rankings of this set were not as conclusive as the counties with the top ten highest median household incomes. Despite this, it is still clear to see that the lowest earning counties did not experience growth like the top earning counties. Of this

list, 8 of the 10 counties with lowest incomes in the state were in the bottom half of the state in population growth.

	County	Median_Household_Income	Growth_Rank	10_year_growth
3	Appanoose	60169	73	-4.423062
26	Decatur	66754	98	-9.601514
89	Wapello	67016	33	-0.527719
50	Jefferson	68298	93	-7.005878
19	Clarke	70580	12	4.975232
88	Van Buren	70893	81	-4.848085
32	Fayette	71025	91	-6.566092
34	Franklin	72089	89	-6.189139
72	Page	72199	75	-4.525483
98	Wright	72665	50	-2.161917

Figure 6: Bottom 10 Counties Based on Median Household Income

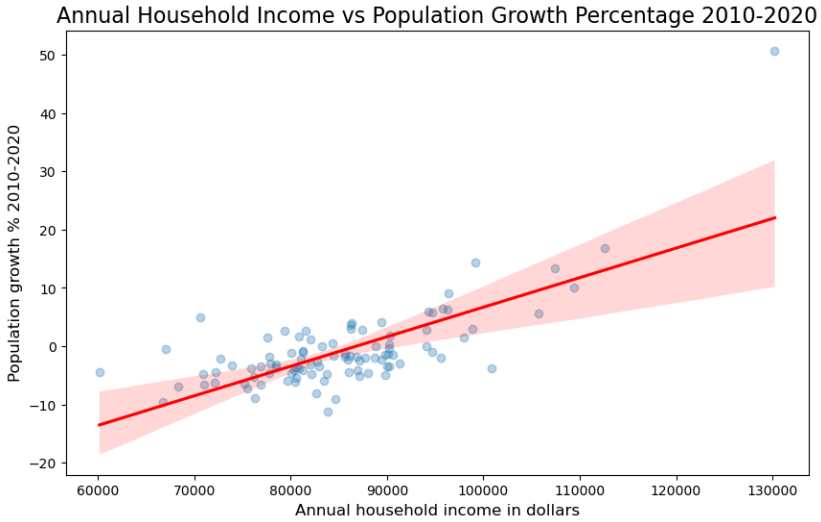

Figure 7: Scatterplot of Household Income vs Population Growth

To further analyze the relationship between annual household income and population growth, I decided to plot the two variables on a scatter plot to visualize the relationship. The annual median household income is on the x-axis and the population growth percentage between 2010 and 2020 is on the y-axis. Based on the graph and the trendline I plotted, it is clear to see there is a positive relationship between annual household income and the population growth between 2010 and 2020.

3.4 Unemployment vs Population Growth

The last comparison I wanted to make was the relationship between unemployment rates and the population growth a county experienced. I grouped the dataframe on 'Unemployment_Rate'. I produced two new data frames, one with the unemployment rate sorted ascending and the other sorted descending. I also included the Urban_Percent and Urban_Rank to draw additional conclusions from these results.

	County	Unemployment_Rate	Growth_Rank	10_year_growth	Urban_Percent	Urban_Rank
4	Audubon	0.7	94	-7.272430	0.00%	63
65	Mitchell	0.9	48	-1.958055	0.00%	63
7	Boone	1.3	24	1.554778	46.80%	37
14	Cass	1.4	86	-5.940097	50.30%	34
44	Howard	1.5	36	-1.014008	0.00%	63
92	Wayne	1.6	25	1.468062	0.00%	63
79	Ringgold	1.7	97	-9.121029	0.00%	63
35	Fremont	1.7	99	-11.235049	0.10%	62
13	Carroll	1.9	32	-0.269024	48.90%	36
25	Davis	2.0	13	4.078602	0.00%	63

Figure 8: Top 10 Counties by Lowest Unemployment Rate

	County	Unemployment_Rate	Growth_Rank	10_year_growth	Urban_Percent	Urban_Rank
72	Page	6.6	75	-4.525483	66.20%	19
75	Pocahontas	6.3	59	-3.173735	0.00%	63
3	Appanoose	6.1	73	-4.423062	42.80%	41
87	Union	5.7	58	-3.159406	61.80%	20
63	Marshall	5.6	39	-1.335859	68.30%	15
84	Story	5.6	5	10.045565	74.20%	9
50	Jefferson	5.2	93	-7.005878	58.80%	23
6	Black Hawk	5.2	29	0.041193	87.00%	2
10	Buena Vista	5.0	19	2.778875	57.00%	25
88	Van Buren	4.9	81	-4.848085	0.00%	63

Figure 9: Top 10 Counties by Highest Unemployment Rate

The comparison between unemployment rate and growth ranking was very inconclusive. While I might have expected areas with low unemployment rates to have higher growth rates, this was not necessarily true. In fact, 3 of the counties that were in the top ten for population decline, as seen in Figure 8, were also in the top ten for low unemployment. When looking at the counties with the top 10 highest rates of unemployment, there was a wide variety of growth rankings associated.

Taking a look at these derived data frames through the lens of urban percentage told another story. Only one county (Cass County) within the top ten counties for unemployment had an urban percentage of over 50%. In fact, 6 of the top 10 counties for unemployment were classified as 100% rural. When looking at the 'Urban_Rank', due to nearly 40 counties having 0% urban population, 63rd is the rank that all of those tied counties carry. So this shows me that some of the most rural counties in the state had the lowest levels of unemployment.

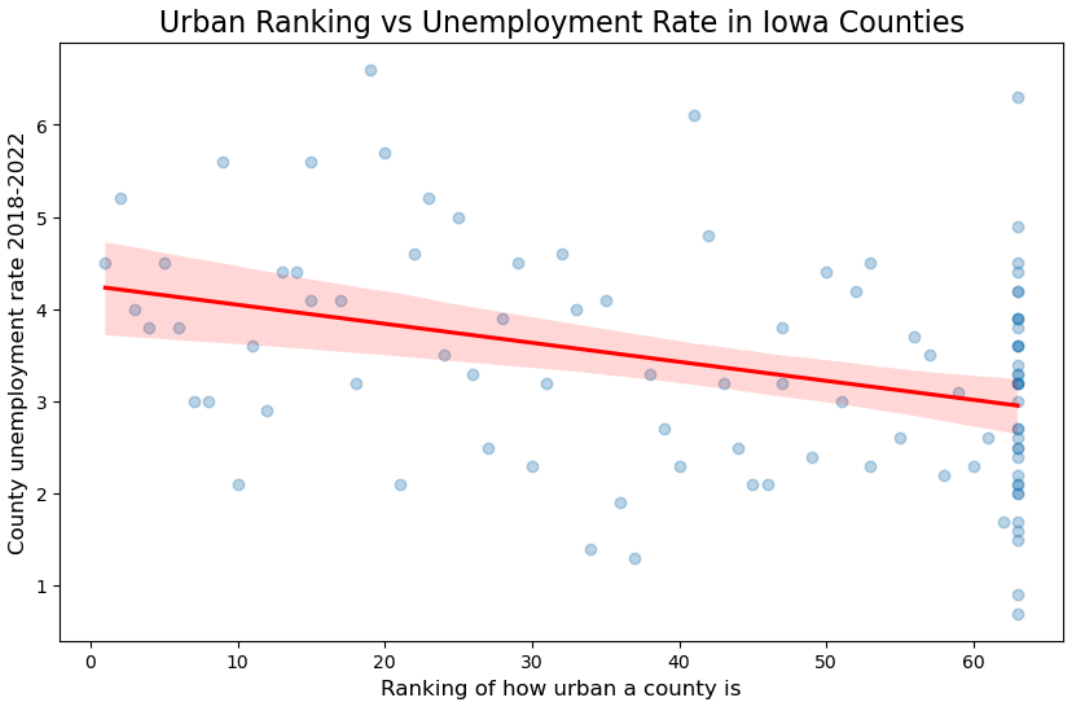

Figure 10: Scatterplot Depicting Urban Ranking vs Unemployment Rate

4. Conclusion

In this project, I analyzed several different demographic areas for each of Iowa's 99 counties. The different topics I looked into were population, rural/urban makeup, unemployment rate, and household income. When making these conclusions I will try to avoid extrapolating or making generalizations about the data. Here are the conclusions I drew from the questions I posed in my project proposal:

1. *Does a county's ratio of urban to rural residents have any impact on the growth rate of the county?*

As depicted in Figure 2, the results indicated that there is an inverse relationship between a county's urban classification and the level of population growth it experienced between 2010 and 2020.

2. *Which counties in Iowa saw the highest growth in population between the 2010 and 2020 Census? Similarly, which counties saw the highest percentage drop in population?*

Figure 3 and 4 show the top 10 highest and lowest population growth percentages, respectively. The results show that Dallas, Johnson, and Story county were the county's with the largest percentage of population growth between the 2010 and 2020 censuses. Fremont, Decator, and Ringgold county were the counties with the highest level of population decline in the same period. It is notable that the highest growth occurred in counties within the top 10 for urban percentage ratings. The three highest counties with population decline occurred in counties classified as nearly or totally rural.

3. *Does the average household income of a county have any effect on the county's growth?*

Figure 7 shows a positive relationship between the median household income of a county and the county's population growth between 2010 and 2020. In other words, as the median annual household income in a county increased, there was generally also an increase in the population growth of the county.

4. *Do the unemployment levels of a county have an impact on the county's population?*

Figure 10 depicts a weak inverse relationship between a county's unemployment and the county's urban classification. It is important to note that I used rankings for the x axis (urban ranking), which means a county's point on the scatterplot is relative to where it is in relation to the other counties, not its actual urban percentage.

With all of the relationships I found throughout this project, I want to reiterate that I did not classify any of them as causal. For the scope of this project, I was able to identify positive or inverse relationships, but I did not look into causality. If I were to expand upon this project, that is a factor I would consider looking into.

When reflecting on this project, I can see it has several limitations. While I had a special interest in understanding the population trends in Iowa, I can understand that the results of this project may not be applicable to many. Focusing my project on just the state of Iowa limited that amount of rows to only 99, which is quite small. Another drawback of my project is the fact that I didn't use some of the data in my conclusions that I included in my data dictionary. This was the population and growth rates from the 1980, 1990, and 2000 censuses. When getting into my project, I had trouble working that data into the project in a way that didn't overcomplicate my questions.

In the future, I would expand this project to include more demographic factors like age, race, and household income. I have also considered including school district enrollment information into the dataset to see its relationship with a county's population growth. This project could also be expanded to include counties from all the states in the United States. I could also look at dividing the population into different geographical groupings like city or state.