Analysis of the INC.com Top 5000 Companies and Comparison with Census Data

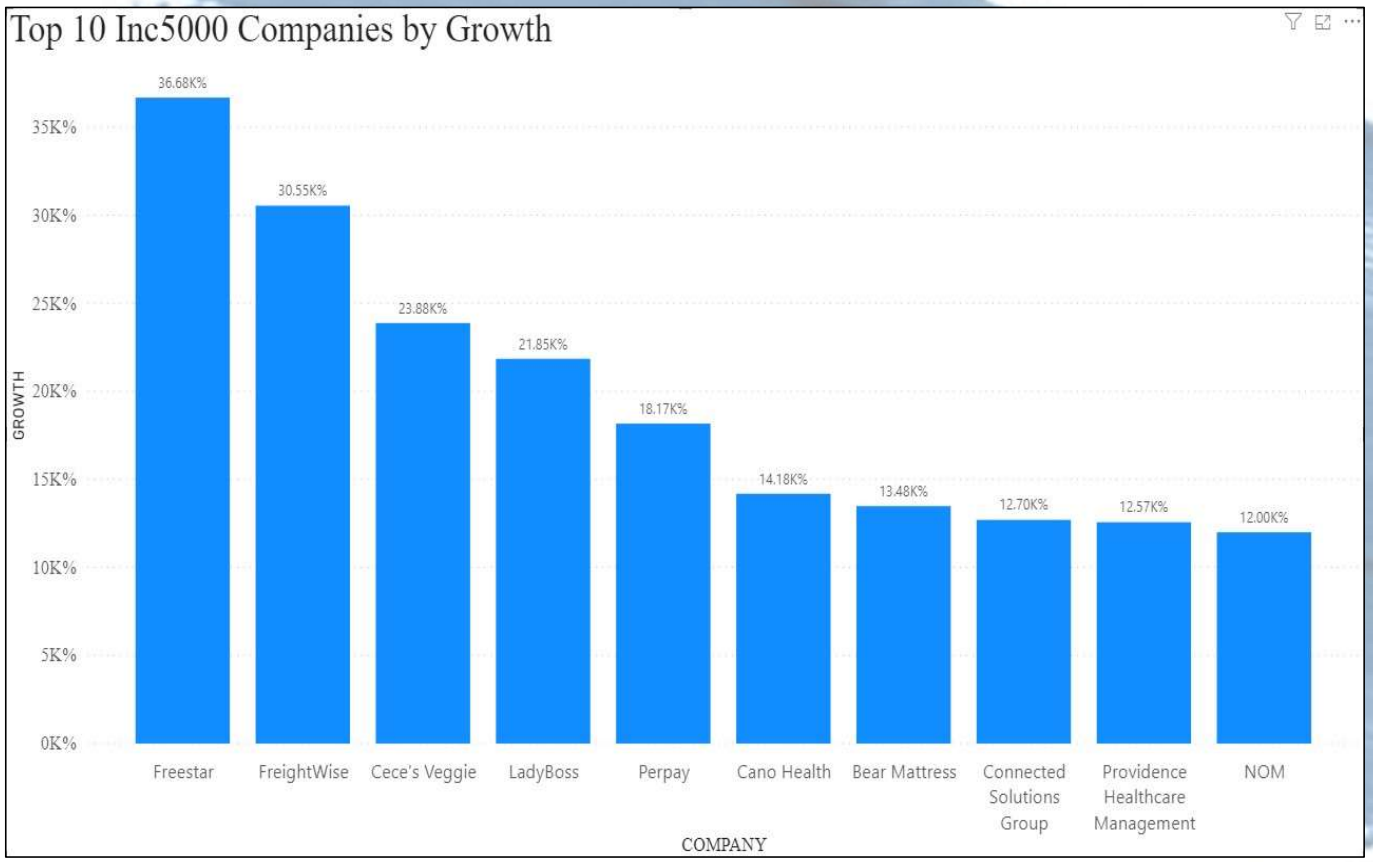
By Dan Amrine

November 1,2020

Agenda:

- Introduction
- Objectives
- Research Hypothesis
- Methodology
- Findings
- Conclusion and Recommendations

Introduction:



INC.com list of top 5000 growing companies from 2018-2019

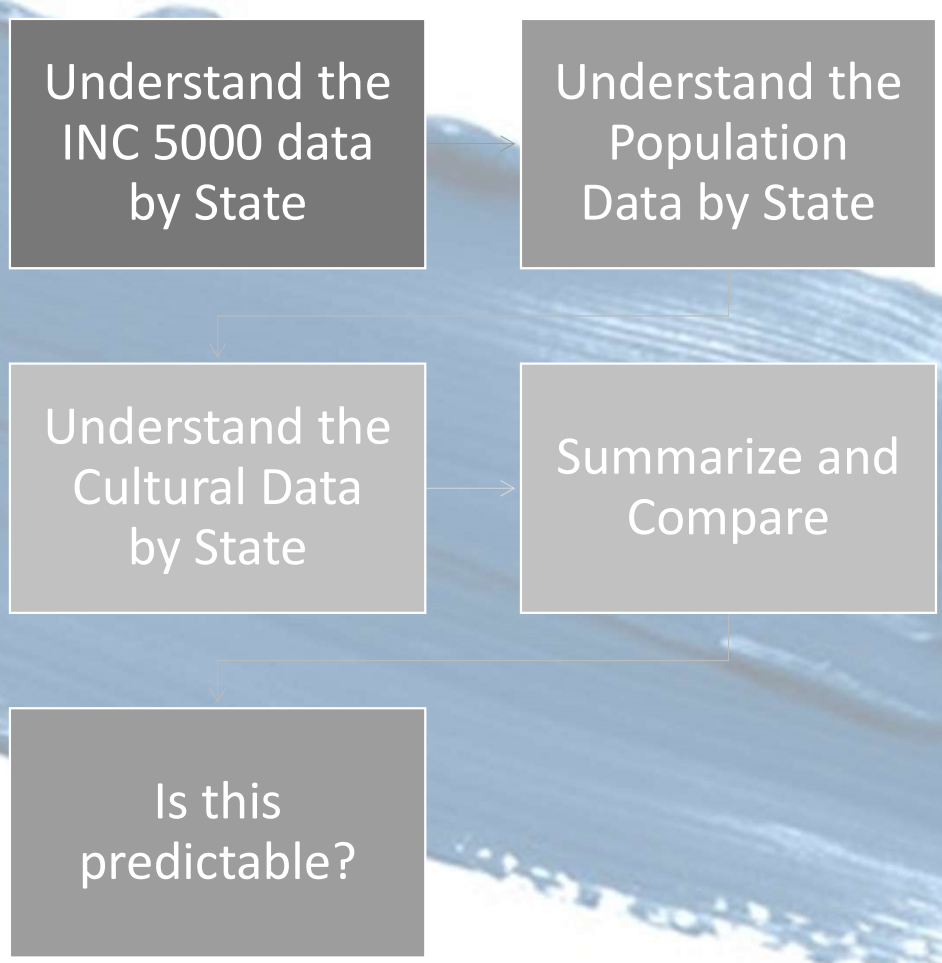
What drives this growth?

This project compares the growth of these companies with Census data collected by the United States.

The main goal is to see if there is a relationship with the growth and these attributes:

- Population
- Poverty
- Education
- Income
- Urban Density

Objectives:


Understand the INC 5000 data by State
Understand the Population Data by State
Understand the Cultural Data by State
Summarize and Compare
Is this predictable?

Hypothesis:

- **Does each variable have normal distribution**?
  - Ho: It does not have normal distribution
  - Ha: It does have normal distribution

- **Is there a strong correlation between the state population, growth, urban density, and Business Growth**?
  - Ho: There is no correlation
  - Ha: There are strong or weak positive or negative correlations.

- **Is there a strong correlation between the population's education, earnings, and poverty with the growth of these companies**?
  - Ho: There is no correlation
  - Ha: There are strong or weak positive or negative correlations.

- **Is it possible to identify a combination of these attributes that would either enhance or discourage company growth**?
  - Ho: It is not possible or there is no definitive combination
  - Ha: It is possible to find a specific combination

Literature and Limitations:

Literature

- Difficult to find examples of this analysis
- However, it is clear small business growth is an essential part of economy
- Maps have been used to find correlation
- The following is the criterion to included in this list

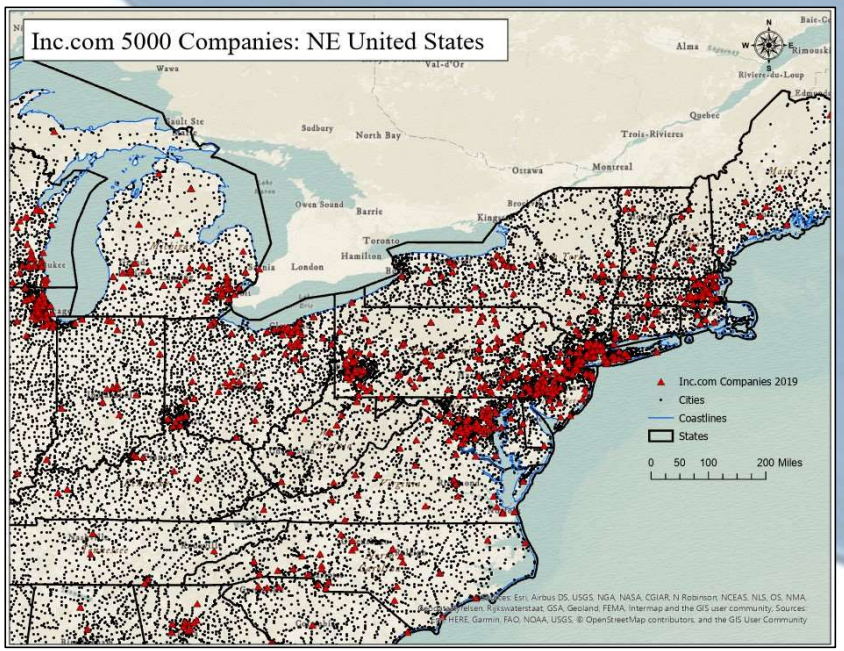
1. Be Privately-owned, based in the United States, and independent
2. Have started earning revenue by March 31, 2016
3. Had revenue no less than $100,000 in 2016
4. Had revenue no less than $2,000,000 in 2019
5. Revenue in 2019 exceeds revenue in 2016

Limitations

- Missing Data such as number of employees, initial capital, total costs, and other proprietary attributes.
- There is a problem with the population in Puerto Rico this is not a serious limitation as we are focusing on the lower 48 states.
- This led to a very thorough review of all the attributes used to find and mitigate any other issues.

Methodology:

Collecting and Mapping



Discussion:

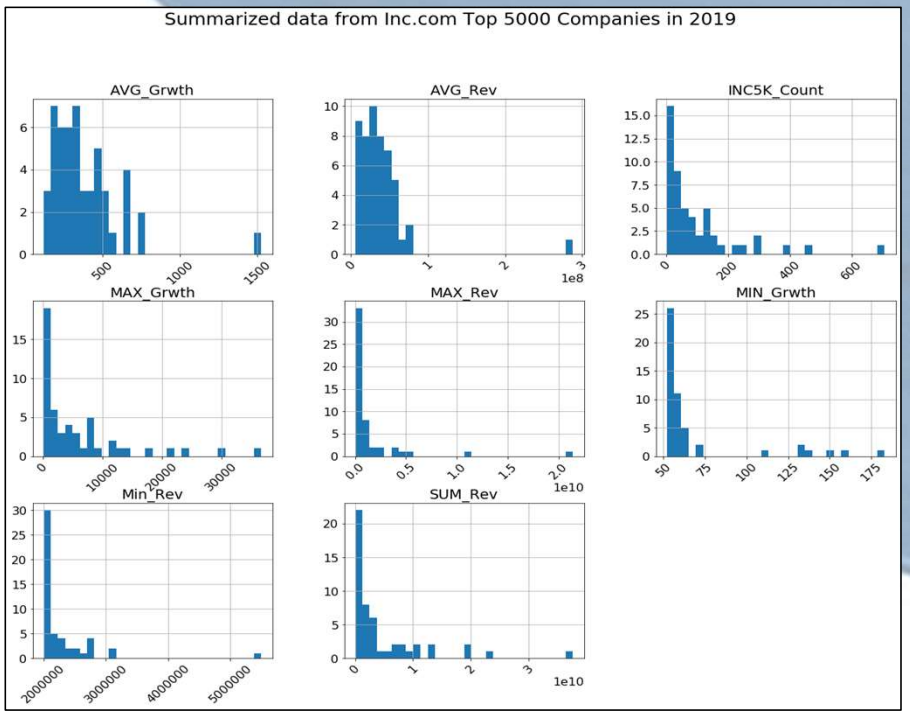
- This process involved scraping the 5000 records from the public Inc.com website into an excel sheet.
- To make the map the names of the cities had to be corrected and when a company used a regional or common name it had to be linked to the closest census designation.
- This data is at the city level with corresponding state and county.

Methodology (cont.):

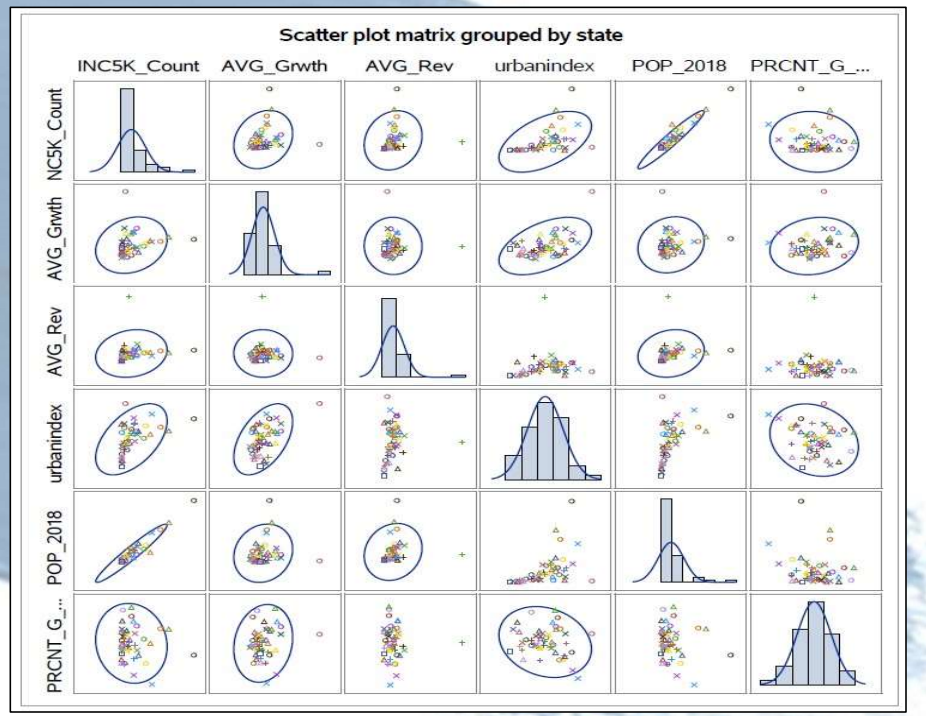
Software and Languages	Tools and Analysis
➢ Python: Used to test the data distribution ➢ SAS Studio used to explore the relationships between the data ➢ Microsoft PowerBI for advanced visualization and live functionality ➢ Microsoft Excel for the initial analysis	➢ Shapiro – Wilks Test of Gaussian (Normal) Distribution ➢ Pearson Correlation Coefficient ➢ Boxplot visualization which allows the comparison between three variables

Results:

Distribution: Not Normal



Data Exploration: There are Correlations



Results: SAS, Excel, and Correlation

6 With Variables:	Under_30962 Prcnt_UND_30962 POV30962_pr_100K TOTAL_ERNGS MALE_ERNGS FEMALE_ERNGS
8 Variables:	INC5K_Count AVG_Grwth MAX_Grwth MIN_Grwth SUM_Rev MAX_Rev Min_Rev AVG_Rev

Pearson Correlation Coefficients, N = 49

	INC5K_Count	AVG_Grwth	MAX_Grwth	MIN_Grwth	SUM_Rev	MAX_Rev	Min_Rev	AVG_Rev
Under_30962	0.94020	0.09627	0.51815	-0.30828	0.83969	0.39520	-0.29659	0.12884
Prcnt_UND_30962	0.02805	-0.04189	0.00304	-0.04205	-0.03781	-0.16544	-0.07795	-0.17295
POV30962_pr_100K	0.02805	-0.04189	0.00304	-0.04205	-0.03781	-0.16544	-0.07795	-0.17295
TOTAL_ERNGS	0.11679	0.49925	0.47403	-0.11217	0.09390	0.12133	-0.01604	0.09695
MALE_ERNGS	0.06331	0.41901	0.40276	-0.03933	0.05438	0.10197	0.05450	0.09461
FEMALE_ERNGS	0.14327	0.52038	0.48738	-0.10911	0.11198	0.12260	-0.03499	0.07689

← SAS Results

Total Correlation Results

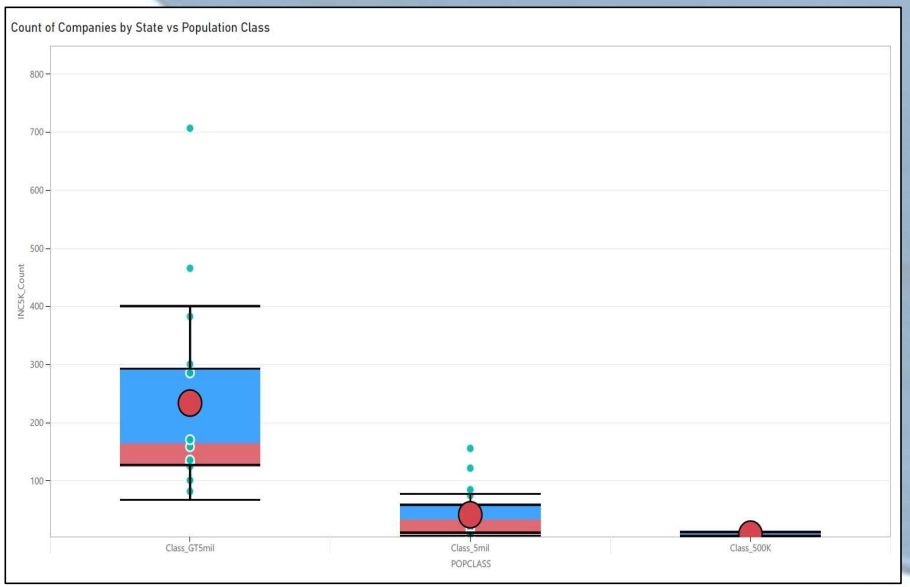
TYPE	_NAME_	INC5K_Count	AVG_Grwth	MAX_Grwth	MIN_Grwth	SUM_Rev	MAX_Rev	Min_Rev	AVG_Rev
CORR	EDCTN_Total	0.95837443	0.117512879	0.533639868	-0.334520798	0.866925197	0.436893354	-0.32382892	0.168162622
CORR	NoSch_pr100k	0.718270471	0.114037509	0.424655165	-0.293470236	0.620284559	0.311313339	-0.31450816	0.097395715
CORR	HighSch_pr100k	-0.344818138	-0.43273001	-0.45530853	0.183269009	-0.268042256	-0.167224479	0.007789392	-0.060061869
CORR	Assoc_pr100k	-0.158593613	-0.33443077	-0.32595045	0.183546239	-0.034713529	0.136962119	0.377398242	0.171607238
CORR	Bach_pr100k	0.202860211	0.403517323	0.389443433	-0.154253725	0.225400726	0.257170692	-0.014613706	0.212212756
CORR	Master_pr100k	0.134449132	0.503673458	0.492273075	-0.102802087	0.066652499	0.034616351	-0.164042979	0.007250042
CORR	Proff_pr100k	0.099278298	0.691999682	0.588892256	-0.095619547	0.060373737	0.04157868	-0.012800041	-0.003313791
CORR	Doct_pr100k	0.058468016	0.520529414	0.446226856	0.01076076	0.018307242	0.01283099	-0.012794489	-0.034872433
CORR	No_schooling	0.929680516	0.128021085	0.516713438	-0.231773538	0.841011145	0.433261571	-0.22345575	0.104917074
CORR	high_school_dipl	0.903790763	0.080400166	0.486599084	-0.356375436	0.82597632	0.408737106	-0.348295696	0.187129101
CORR	Associate_degree	0.937441255	0.107042025	0.50574243	-0.33882388	0.865813659	0.465529043	-0.320559653	0.203164266
CORR	Bachelor_degree	0.967577922	0.147423554	0.556496777	-0.334322006	0.879299703	0.462130546	-0.321344059	0.183157867
CORR	Master_degree	0.946807419	0.142144369	0.552997624	-0.33894313	0.833189201	0.420250626	-0.335264821	0.158913279
CORR	Professional_degree	0.94655521	0.158539752	0.53478203	-0.313415643	0.846626739	0.442820354	-0.303181356	0.154134541
CORR	Doctorate	0.95134486	0.137038568	0.524643805	-0.317811195	0.857109878	0.451708281	-0.314010385	0.157415485

Scale	
1	Strong Positive
0.5	Weak Positive
0	None
-0.5	Weak Negative
-1	Strong Negative

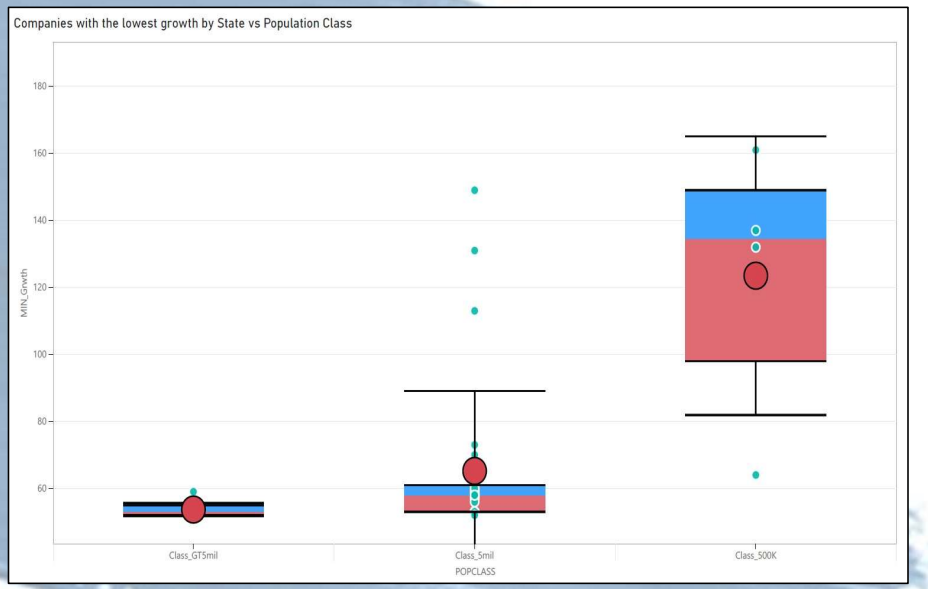
TYPE	_NAME_	Intercept	INC5K_Count	AVG_Grwth	MAX_Grwth	MIN_Grwth	SUM_Rev	MAX_Rev	Min_Rev	AVG_Rev
CORR	urbanindex		0.52	0.50	0.60	-0.40	0.42	0.20	-0.37	0.13
CORR	POP_2018		0.96	0.14	0.54	-0.32	0.86	0.43	-0.30	0.14
CORR	PRCNT_G_1718		-0.09	0.15	-0.03	-0.14	-0.08	-0.05	0.09	-0.01

Results: Box Plots

Box Plot: Population vs Total Count of Inc 5000 Companies



Box Plot: Minimum growth per state vs population



Conclusion:

- Of the original 4 hypotheses this analysis rejected the null on the last three but not the first. The data does not have normal distribution but there are strong correlations, and it is possible to combine them to build a case
- There is a negative correlation with low growth and population. The strongest correlation showed the more population the more companies on the list you have
- This analysis has shown the even though lower population decreases the count, the states with a population less than 500,000 improve the companies with a lower growth rate
- This is the initial pass on the data and because of this workflow many more cases will be discovered
- With the established workflow the data can be refined, and more relationships investigated
- For questions please contact Grand River Analytics: dan.amrine@grandanalytics.net
- For the documents and python code please go to github: https://github.com/DCCP80/Analysis-of-Inc5000-2019-Companies-and-Census-Data

References:

- Bonsu, S., & Kuofie, M. (2019). Small business survival. *Journal of Marketing & Management*, *10*(1), 51–63.
- Brownlee, J. (2018, August 14). 17 Statistical hypothesis tests in Python (cheat sheet). Machine learning mastery. https://machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/
- US Census Bureau. (2020, January 22). Poverty thresholds: the united states census bureau. https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html
- Fairlie, R. W., Miranda, J., & Zolas, N. (2019). Measuring job creation, growth, and survival among the universe of start-ups in the United States using a combined start-up panel data set. *ILR Review*, 72(5), 1262–1277. https://doi.org/10.1177/0019793919862764
- Fivethirtyeight/data. (n.d.). GitHub. https://github.com/fivethirtyeight/data/tree/master/urbanization-index
- Inc. 5000 2020: An exclusive guide to America's most inspiring entrepreneurs. (n.d.). Inc.com. https://www.inc.com/inc5000
- Rosenblum, D., Castrillo, F. M., Bourgois, P., Mars, S., Karandinos, G., Unick, G. J., & Ciccarone, D. (2014). Urban segregation and the US heroin market: A quantitative model of anthropological hypotheses from an inner-city drug market. *International Journal of Drug Policy*, 25(3), 543-555. https://doi.org/10.1016/j.drugpo.2013.12.008
- Rupasingha, A., & Wang, K. (2017). Access to capital and small business growth: Evidence from CRA loans data. *Annals of Regional Science*, 59(1), 15–41. https://doi.org/10.1007/s00168-017-0814-9
- Wach, K. (2020). A typology of small business growth modelling: A critical literature review. *Entrepreneurial Business & Economics Review*, *8*(1), 159–184. https://doi.org/10.15678/EBER.2020.080109