"Innovation Hubs”: A Story of Cities and Patents

Nathan Caplan, Tingyu Chang, Rohun Iyer, Sarah Sachs

Abstract

In recent years there has been great interest around cities and their new product output, where they are usually labeled as ‘innovation hubs’ or ‘tech hubs’. This capstone project will investigate the factors that facilitate innovation growth within a city using publicly available patent data. To understand how this process develops, we will analyze patent data in the United States from 2001-2012. Our regression analysis will explore many features that influence the growth of innovation. Upon running multiple analyses across the years, we find that there are certain features that have higher influence on patent output amongst the top cities. We also find these features are missing among cities with less patent output. What this experiment would recommend to cities desiring greater patent output is that they should invest in higher education, in earning Small Business Innovation Research (SBIR) grants, and looking into becoming empowerment zones.

1. Introduction

1a. Importance of Study

Cities can measure success in many ways, one being economic output. Cities grow and shrink based on their output, whether it is measured through the number of jobs, employers, GDP, innovation, and so on[9] It becomes difficult for newer or smaller cities to increase their own output comparable to that of larger cities[1]. For instance startups like to form in regions where they have the best chance for success, such as incubators in Silicon Valley. But what got Silicon Valley to where it is today? And can other cities replicate their success? The goal of this paper to identify how cities can invest in their economic future. This is accomplished by developing an innovation score. This will provide an understanding of the urban structure that facilitates the growth of innovation in the form of patent production.

1b. Patents as Innovation

Ever since the first Patent Act of the U.S. Congress in 1790, the patent has been a key representation of innovation and progress in the United States[22]. Keeping that in mind, this project looks into the ecosystem of innovation in the United States using the registry of patents as its foundational aspect.

There is an economic incentive for a city to provide a structure conducive for technological growth and information. When a city attracts a robust network of inventors, investors and collaborative creators -- a foundation for future progress is galvanized. Supported by research that used patents as a metric for innovation, this type of network is exhibited in places such as Silicon Valley[2].

Using patent information is a well-established strategy for understanding the development of technology, and spread of information as it relates to economic growth. It has been explained that there has been use of patent data to explore topics such as inventive activity, the scale of manufacturing in cities and the factors which drive technological advances in a region[2,10,15,19].

1c. Factors That Lead to Innovation

Using previous studies, we have identified three main determinants of a city's classification as an innovative 'innovation hub': regulatory, socioeconomic, and spatial. Different types of regulations include federal programs for research and development (R&D) funding, tax benefits, and government subsidies. In addition to the regulatory environment, literature also suggests that city growth can be measured by changes in demographic and socioeconomic factors[19]. Lastly, these cities are being analyzed over time and in space to identify long-term trends and factors inherent to their geographic properties.

1c1. Regulatory Factors:

In 1997, the National Research Council mentions R&D investment as one indicator in measuring input of innovation[25]. The Information Technology and Innovation Foundation also mentions “The Innovation Success Triangle” in one of its reports with one leg being a strong innovation policy system. It elaborates that such a system includes investment in innovation infrastructure, funding technology and industry research as well as active tax incentives to spur innovation[18]. Therefore, supportive regulations by government play a significant role for cities to thrive in an increasingly competitive global economy. Our approach will take into account multiple regulations and policies to understand the extent to which government regulations influence patent production. Based on this previous research, we believe that cities receiving supportive regulations are more likely to have significant growth in innovative development.

1c2. City Diversity Factors:

Our second set of features contains socioeconomic data since 1997 that help explain the foundation upon which a city’s innovative culture is built. Census data has been used to investigate how changes in population and income would support city growth in terms of innovation[12]. Jonathan Quigley expands upon this research and finds that there is a relationship between nativity, racial, and occupational diversity and economic growth[9].

Urbanist Richard Florida continued to support these claims, saying that cities which achieve diversity in population are the result of greater acceptance in these cities, contributing to greater economic output. Florida continues to state that the 'creative class' -- individuals who are "fully engaged in the creative process" in STEM and the arts -- and other post-higher-education occupations, is a driving force of economic development[16]. We expect that changes in demographic diversity, such as increases in creative class population, the number of creative class establishments, number of foreign born residents, and education diversity contribute to innovation growth.

2. Data and Methods

2a. Patent Data

The patent data obtained is from Patentsview.org. It contains detailed information on every patent assigned to a United States based organization from 1976-2014[23]. Patents and their associated citations have shown to indicate the level of a firm's innovative capacity, and aggregating this data by city, we can scale this model to evaluate a that capacity[6,17].

For the analyses, the patents were aggregated by city using a couple significant features of patents in the United States. Every patent has a list of assignees -- those who own the rights to the patent -- and a list of inventors -- those who contributed to the innovation itself -- each with an associated company and location. Additionally, every patent contains a number of citations -- the amount of new innovations built off of this patent. By aggregating the number of patents assigned and patents invented with their associated citations, two scores were generated that indicate innovative development: Patents Assigned and Patents Invented.

2b. Regulatory Data

A federal award can be defined as money the federal government has promised to pay to companies, organizations, government entities or individuals. This is done by contracts, grants, loans or direct payments. Federal awards data are available from 2001 to 2018 with each year having millions of awards. Each award has 260 features ranging from funding agency, federal obligation, to recipient, and performance center, and location.

Each year's data was aggregated to average amount of federal obligation and total number of awards based on recipient city and primary place of performance. The average amount was used instead of the total amount of federal obligation to account for the large variation in the sizes of cities being analyzed.

Empowerment Zones and the Small Business Innovation Research (SBIR) program were investigated as well. The SBIR program is a federal funding program that enables small businesses to get financial awards from federal agencies' R&D budgets helping thousands of small businesses with over $100 million awarded every year since 1982[5]. For the SBIR program, data are available from 1983 to 2019 with an average of six thousand awards each year and each award containing general information of each business including its location, the amount of award the business receives, its funding agency and topic/field for each awarded project. The average amount of funding per business received and number of businesses awarded were calculated for each city in our model. The goal is to measure if this program has encouraged innovation within cities.

The Empowerment Zone Initiative is a tax incentive and public funding program initiated by the US Department of Housing and Urban Development in 1994 that intended to revitalize economically distressed areas. Empowerment Zone data are available at city level and included in our model to see if these zones witnessed transformations into innovation hubs. For these empowerment zones, a binary variable was put in place to indicate if the city federal assistance.

2c. City Diversity Data

Demographic and household data can be collected decennially going back to 1970 from the Census' IPUMS National Historic GIS at the place levels for the entire U.S. Features of interest include total population, median household income, education, and nativity[4].

In order to determine how many people fit Richard Florida's creative class, we collected US Economic 5 Year Data from the Census API. Years available were 1997, 2002, 2007, and 2012. Data collected includes the number of employers and employees per each job title as described by North American Industry Classification System (NAICS) per Census designated place. Richard Florida describes the creative class as those in academia, arts, and other professions requiring an advanced degree. We mapped the job titles to create, or not, and summed the number of creative and non-creative employees per city in order to determine the size of each city's creative class[16].

Processing the US Economic Census data resulted in five features including the number of 'creative' and 'regular' employees and employers, and a unique city identifier.

2d. Data Aggregation:

A unique code per each city was designed as such, city_state (ex. sanjose_ca), among all datasets in order to join them. We then performed a left join on of our collected and processed data onto the patent data in order to keep as many of the original 1000 top patent producing cities. All final features and their explanations can be found within the appendix.

2e. Model Selection:

Our analysis consists of creating a classification and using coefficients to understand each feature's influence on the classes. To do this we applied a logistic regression to our given data, splitting our data into two classes. We defined the split at the 75th percentile of patent production and can tentatively say the top 25% of cities for each score constitute innovative centers. What this experiment hopes to do is to identify features, and subsequently policy decisions, the bottom 75% of cities might hope to enact for greater patent output. Multiple model scoring metrics were used to determine the model performance including log loss, area under curve (AUC), confusion matrix and precision-recall. After calculating the scores per feature per year, scores were formatted into a time series across our years of interest to highlight feature influence changes over time.

A random forest regression was used to confirm and add a measure of robustness to the feature importance results from the logistic regression. Outside of the impurity scoring feature importance given by sci-kit learn, we calculated the model score using a method that permuted the features and dropped a feature over multiple iterations. As a result, we are able to determine each feature's added value to the model.

2f. Limitations

A major limitation faced by this experiment was the decline in number of municipalities through various stages of this study. With the original top 1,000 patent producing cities, the retention rate after all the joins with other datasets was found to be between 55%-80%. One reason this occurred was due to PatentsView irregular use of geographies within the same location feature. For instance, New York City, NY and Woodlawn, IL -- a neighborhood within Chicago -- were both listed within the top 1,000 patent producing cities. Finding neighborhood level data across various data sources was not possible and, as a result, neighborhoods, towns, and small municipalities were mostly dropped. This amount of data reduction allowed for limited model selection, with the random forest and logistic regressions as the best options.

External data sets not only limited the number of cities, but also the years of analysis. PatentsView data went back to the 1970s, however, between all these external data sources, the window of analysis was shorted to 2001 through 2012. This window allows for the experiment to determine what features are significant to city patent production, but not how these cities became patent producing 'hubs'.

3. Results

The models were built using the two patent scores discussed earlier as the dependent variable and the various independent variables (Appendix Section 4). There was a fixed effect applied to all demographic and creative class features as they are five-year aggregates.

The following results are from our logistic regressions across all features for all years. The figures below details them for two of the features in the logistic model.

In order for cities in their current state to see how they could achieve patent output of the top 25% of patent producing cities, we looked at 2012, the final year in our analysis. The major differences found between our top 25% and bottom 75% of cities among both scores is the significant influence SBIR award mean, the percentage of graduate students, and the Empowerment Zone status of a city. Performance mean had significant influence on the number of invented patents and if the city was an empowerment zone had significant influence on the number of assigned patents for the top 25% of cities.

In order to evaluate the robustness of our logistic regressions, the regression scores were compared for each feature across all cities, the top 25% and the bottom 75% of patent producers for each dependent variable across all years. Between assigned and invented patents, the trend in feature importance matched for eight out of the ten features for all cities, seven for the top 25%, and seven for the bottom 75%. On average our model showed consistent trends across both scores with a 73.33% match across all the features for all city breakdowns, indicating high robustness and confirming general trends.

Figure 1: Changes in Feature Influence of Percent Graduate Students Over Time

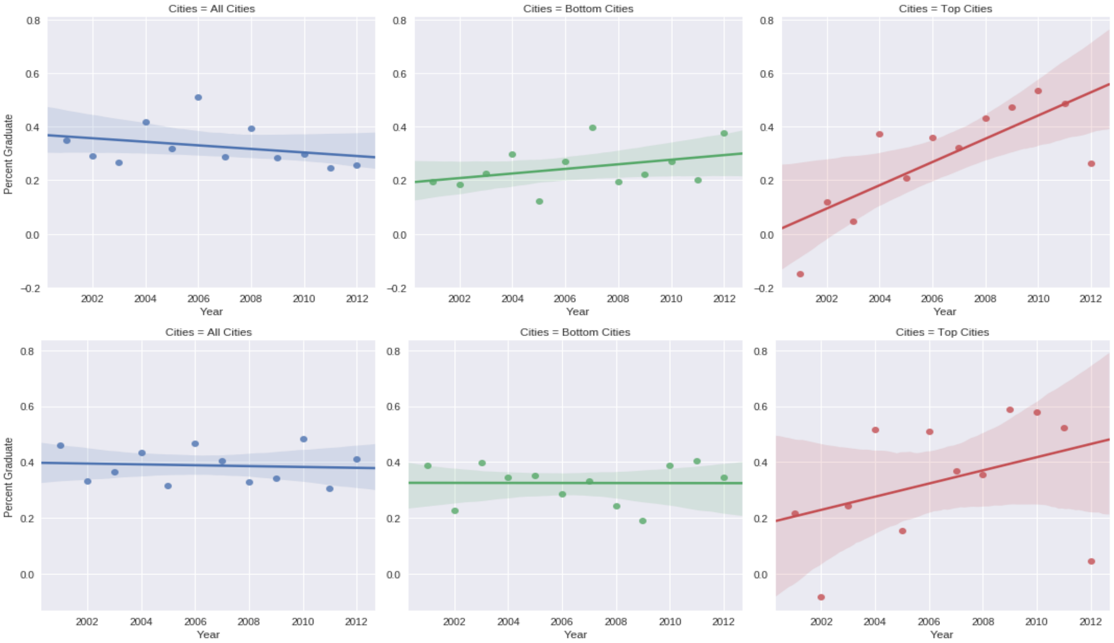


Figure 1 details the similar trend in feature significance for percent of graduate degrees earned for both patent scoring methods. Where assigned patent score is on top and invented is on bottom.

Using an AUC scoring metric, our random forest regression score was found to be between 0.71 and 0.86 though our years of study, sufficient to explore the feature importance. The out of sample was in between 0.26 and 0.40 across those years; a decent indicator within the social sciences of a meaningful relationship. We see both upwards and downwards trends across multiple features, leading us to infer that features change in importance towards innovative growth. Seeing that cities today may want to know what features are currently important in innovative growth, Figure 2 shows the feature importance for all of our features from the random forest regression for assigned patents.

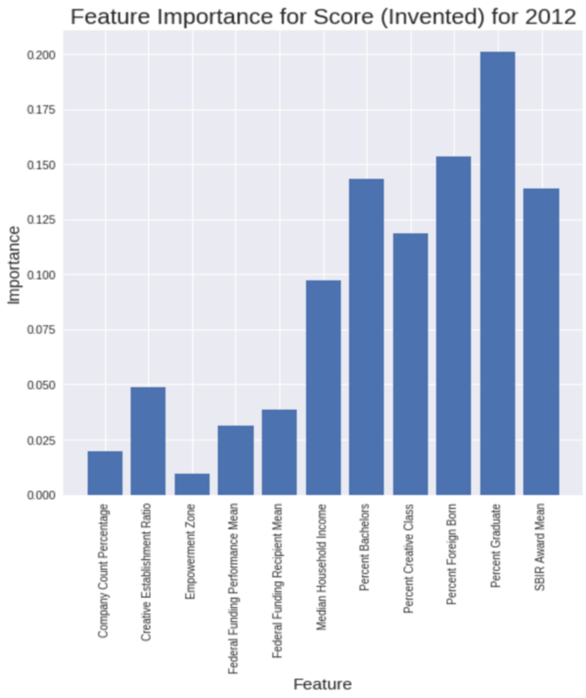


Figure 2 details the results from the random forest regression for 2012. It shows the high influence in SBIR Awards, percent graduates and bachelor's, and percent foreign born.

Across both scoring methods, SBIR Award Mean, percentage of the population with graduate degrees, the percentage with bachelor's degrees, and the percentage of foreign born residents had the highest feature importance. The only feature of those four to decrease over time is the percentage of the population with bachelor's degrees, suggesting that graduate degrees are a more telling feature for innovative output. Unlike the logistic regression, explained below, the model was applied on all cities. Inconsistent results were yielded when applied to the top 25% and bottom 75% of patent outputting cities. This was due to the split yielding too few items to perform the random forest and likely overfitting.

Comparing the feature importance for each feature across all the years, the trends matched for eight of the ten features for all cities for both assigned and invented patent scores, consistent with. That means there is consistent trends across all features for all cities, indicating high robustness between the random forest model and the two scores.

Lastly, a visualization was produced, as seen below in Figure 3, that displays the spatial element (All remaining regression plots can be found in the appendix for each regression.)

Figure 3: Exploratory Spatial Visualization of Patent Output in Cities Across America

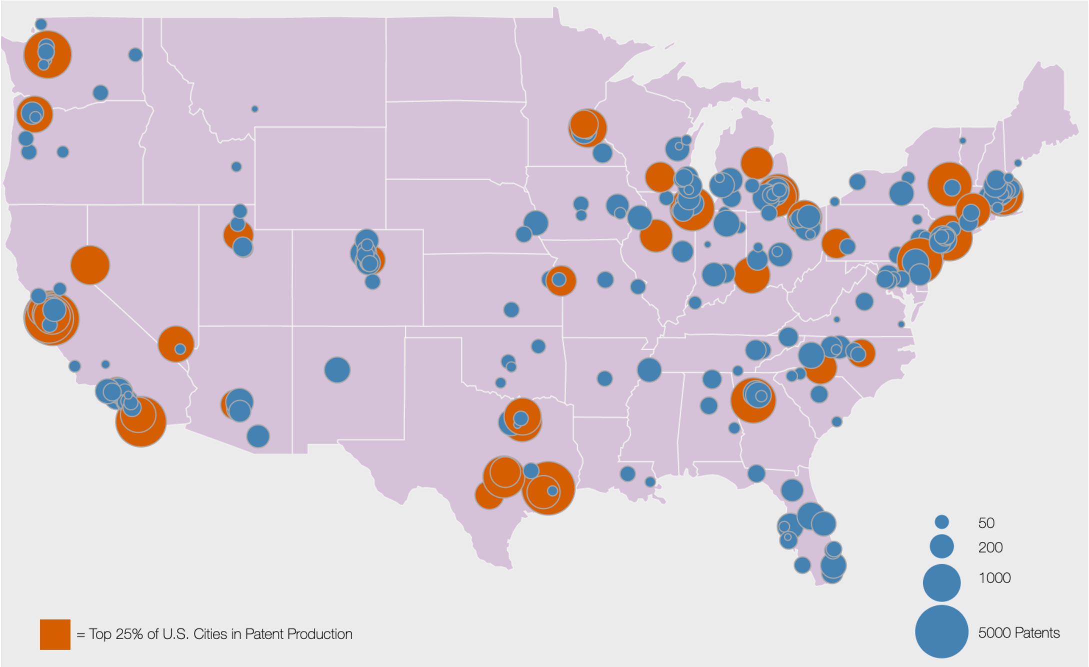


Figure 3 shows the initial spatial visualization of patents in United States. Orange cities represent the top 25% of cities from the analyses above.

5. Conclusions

In order for a city to improve their innovative output, major policy and economic decisions can be made. The contents of this paper yield results that demonstrate that there are certain socioeconomic and policy decisions that could be made for a city with limited patent output to achieve greater levels comparable to larger patent producing cities. Based on the random forest regression, a city's mayor or key investors should push for investments into higher education, in earning SBIR grants, and incentives for members of immigrant communities to move to their cities. As for the logistic regression, push for investments to higher education, earning SBIR grants, and looking into becoming empowerment zones. Feature importance differences can be attributed to the differences in the underlying scoring structure of the two models used in our analysis. However, seeing SBIR Award Mean, percent graduate degrees earned, and percent foreign born residents were consistently influential across all years and both scoring methods, we are confident in our recommendations and findings listed above.

References

1. Berry, B. J. (1961, July). City Size Distributions and Economic Development. Economic Development and Cultural Change, 9(4), 573-588Breandán Ó hUallacháin, Timothy F. Leslie. Spatial Convergence and Spillovers in American Invention. *Annals of the Association of American Geographers* 95, 866–886 Informa UK Limited, 2005. Link
2. Carlino, Gerald A. and Carr, Jake and Hunt, Robert M. and Smith, Tony E., The Agglomeration of R&D Labs. FRB of Philadelphia Working Paper vol. 11, no.42. Sept. 2011
3. "Census Data API:" *api.census.gov*, 28-Nov-2011. [Online]. Available: https://api.census.gov/data.html.
4. "Dashboard," *Dashboard*. [Online]. Available: https://www.sbir.gov/awards/annual-reports. [Accessed: 30-Apr-2019].
5. David Audretsch Zoltan Acs. Patents as a Measure of Innovative Activity. *Kyklos*. 1989.
6. Falcone, J.A., 2015, U.S. conterminous wall-to-wall anthropogenic land use trends (NWALT), 1974–2012: U.S. Geological Survey Data Series 948, 33 p. plus appendixes 3–6 as separate files, http://dx.doi.org/10.3133/ds948.
7. *International Patent Classification (IPC)*. [Online]. Available: https://www.wipo.int/classifications/ipc/en/. [Accessed: 30-Mar-2019].
8. J. M. Quigley, "Urban Diversity and Economic Growth," *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 127–138, Apr. 1998.
9. Johnson, D. K. N. and Brown, A. How the West has won; Regional and industrial inversion in U.S. patent activity. Economic Geography, 80: 241–60. 2004.
10. Knut Blind, The influence of regulations on innovation: A quantitative assessment for OECD countries. October 2011.
11. L. M. Bettencourt, J. Lobo, C. Kuhnert, and G. West, "Growth, innovation, scaling, and the pace of life in cities," *PNAS*, vol. 104, no. 17, pp. 7301–7306, Nov. 2006.
12. M. Batty and P. Longley, *Fractal cities: a geometry of form and function*. London: Academic Press, 1994.
13. Mario Pianta Daniele Archibugi. Measuring Technological Through Patents and Innovation Surveys. *Institute for Studies on Scientific Research*. 1996.
14. Pred, A. R.. The spatial dynamics of U.S. urban-industrial growth, 1800–191, Cambridge, MA: MIT Press.1966.
15. R. M. Florida and G. Gates, "Technology and Tolerance: The Importance of Diversity to High-Technology Growth," *Center on Urban & Metropolitan Policy*, pp. 1–12, Jun. 2001.
16. Riitta Katila. Using Patent Data to Measure Innovation Performance. *International Journal of Business Performance Management*. 2000.
17. Robert D. Atkinson, Understanding the U.S. National Innovation System, The Information Technology and Innovation Foundation. June 2014. [Online]. Available: Link
18. Sokoloff, K. L. Inventive activity in early industrial America: Evidence from patent records 1790–1846. The Journal of Economic History, 48: 813–50. 1988.

19. S. Nagaoka, K. Motohashi, and A. Goto, “Patent Statistics as an Innovation Indicator,” *Handbook of the Economics of Innovation, Volume 2 Handbook of the Economics of Innovation*, pp. 1083–1127, 2010.
20. Tinguely, X.The Nature of the Innovation Process and the New Geography of Innovation. In *The New Geography of Innovation*. Palgrave Macmillan. 2013.
21. “U.S. PATENT ACT -- OVERVIEW,” *Legal Information Institute*. [Online]. Available: https://www.law.cornell.edu/patent/patent.overview.html.
22. “Why Explore Patent Data?,” *PatentsView*. [Online]. Available: http://www.patentsview.org/api/doc.html.
23. “World Topographic Map,” *arcgis.com*. [Online]. Available: https://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f. [Accessed: 30-Apr-2019].
24. "5 Improving Information on Industrial R&D." National Research Council. 1997. *Industrial Research and Innovation Indicators: Report of a Workshop*. Washington, DC: The National Academies Press. doi: 10.17226/5976.

Appendix

1. Calculating Innovative Scores

Sadao Nagoaka[2] cites multiple patent-related values that can be used to measure innovation including raw patent numbers, patent citation numbers and patent citations controlled for the diversity of patent classifications. For this preliminary analysis, we focused on the first two measures: patent numbers and patent citations. The calculated scores are as follows:

Score 1	Inventor Patent Citations
Score 2	Assigned Patent Citations
Score 5	Inventor Patents Awarded
Score 6	Assigned Patents Awarded
Score 7	Combined Patents Awarded

**Scores 3 and 4 have been omitted for this analysis because of issues with the underlying data*

As expected, the patents and their citations per city had an incredibly large spread and were very skewed. They were standardized as described in Appendix, Section 7.

2. Standardization Process

Each feature and patent scoring metric was standardized in the following way:

$$\text{value} = \text{value}^{(1/\log(\max(\text{current\_feature}))}$$

Using this standardization, we were able to account for the various different skewed features with one standard function.

To split our data into two classes, we settled on using the 75th percentile as our cut-off score. As a result, we had a total of 182 cities marked as innovative versus 547 as not innovative. This was used in our logistic regression and in any stratified regressions.

3. Sci-kit learn’s Regression Scoring Metrics Used

a. Log Loss

“Log loss, also called logistic regression loss or cross-entropy loss, is defined on probability estimates. It is commonly used in (multinomial) logistic regression and neural networks, as well as in some variants of expectation-maximization, and can be used to evaluate the probability outputs (predict_proba) of a classifier instead of its discrete predictions.”[https://scikit-learn.org/stable/modules/model_evaluation.html]

A log loss scoring metric is being used to evaluate the accuracy of our model outside of pure true-positive, true-negative measures and taking into account the confidence of the models we have created.

b. Area Under Curve (AUC)

“The roc_auc_score function computes the area under the receiver operating characteristic (ROC) curve, which is also denoted by AUC or AUROC. By computing the area under the roc curve, the curve information is summarized in one number.”[scikit-learn.org/stable/modules/model_evaluation.html]

The AUC scoring metric was used to evaluate the ability of our logistic regression model to choose true-positive results. For evaluation, we want the AUC score to be as close to 1 as possible. Anything below 0.5 is worse than randomizing our model.

c. Confusion Matrix

“The confusion_matrix function evaluates classification accuracy by computing the confusion matrix with each row corresponding to the true class”[https://scikit-learn.org/stable/modules/model_evaluation.html#confusion-matrix]

The added benefit of using the confusion matrix allows to properly identify the number of false-positive and false-negative classifications we have and where our model could possibly improve.

d. Precision and Recall Scores

“The precision is the ratio tp / (tp + fp) where tp is the number of true positives and fp the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative. The best value is 1 and the worst value is 0.” [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html#sklearn.

metrics.precision_score]

“The recall is the ratio tp / (tp + fn) where tp is the number of true positives and fn the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples. The best value is 1 and the worst value is 0.” [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html#sklearn.metrics.recall_score]

The added benefit of using precision and recall scores quantifies the values we are seeing in the confusion matrix and shows possible areas we are lacking.

4. Features Table

Feature	Data Group	Explanation
SBIR Award Mean	Regulatory Data	Average amount of award per company has received of the City
Percent Creative Class	Socio-Economic Data	Percent of the population whose NAICS occupation was considered as a part of the creative class.
Creative Establishment Ratio	Socio-Economic Data	Ratio of the employers whose NAICS title was considered as a part of the creative class over the total establishments
Percentage of Establishments Receiving SBIR Funding	Regulatory Data	Percentage of the establishment of the City that has received SBIR funding
Percent Population Earning Bachelor Degrees	Socio-Economic Data	Percentage of the population that has earned a bachelor degree
Percent Population Earning Graduate Degrees	Socio-Economic Data	Percentage of the population that has earned a graduate degree
Percent Foreign Born	Socio-Economic Data	Percentage of the population that was born outside the US
Median Household Income	Socio-Economic Data	Median Household Income of the City
Federal Funding Recipient Mean	Regulatory Data	Average amount of funding per recipient has received whose company has legally registered

		in the City
Federal Funding Performance Mean	Regulatory Data	Average amount of funding per recipient has received whose primary work of award has performed in the City
Empowerment Zones	Regulatory Data	A binary variable indicating whether the city has ever been included as an empowerment zone

5. NAICS titles deemed to be a part of the Creative Class:

'Heavy and civil engineering construction'

'Veneer, plywood, and engineered wood product manufacturing'

'Pharmaceutical and medicine manufacturing'

'Electronic computer manufacturing'

'Computer terminal and other computer peripheral equipment manufacturing'

'Navigational, measuring, electromedical, and control instruments manufacturing'

'Electromedical and electrotherapeutic apparatus manufacturing'

'Surgical and medical instrument manufacturing'

'Computer and computer peripheral equipment and software merchant wholesalers'

'Computer and computer peripheral equipment merchant wholesalers'

'Surgical, medical, and hospital supplies merchant wholesalers'

'Direct life, health, and medical insurance carriers'

'Direct health and medical insurance carriers'

'Direct insurance (except life, health, and medical) carriers'

'Offices of lawyers'

'Legal aid societies and similar legal services'

'Other legal services'

'All other legal services'

'Accounting, tax preparation, bookkeeping, and payroll services'

'Other accounting services'

'Architectural, engineering, and related services'

'Custom computer programming services'

'Other computer related services'

'Scientific research and development services'

'Research and development in the physical, engineering, and life sciences'

'Research and development in the physical, engineering, and life sciences'

'Research and development in biotechnology'

'Research and development in the physical, engineering, and life sciences (except biotechnology)'

'Research and development in the physical and engineering sciences'

'Research and development in other life sciences'

'Research and development in the social sciences and humanities'

'Marketing research and public opinion polling'

'Economic or industrial planning or development organization'

'Business schools and computer and management training'

'Professional and management development training'

'Art, drama, and music schools'

'Family planning centers'

'HMO medical centers'

'Freestanding ambulatory surgical and emergency centers'

'General medical and surgical hospitals'

'General medical and surgical hospitals, government'

'General medical and surgical hospitals, except government'

'Residential intellectual and developmental disability, mental health, and substance abuse facilities'

'Residential intellectual and developmental disability facilities'

'Musical groups and artists'

'Symphony orchestras and chamber music organizations'

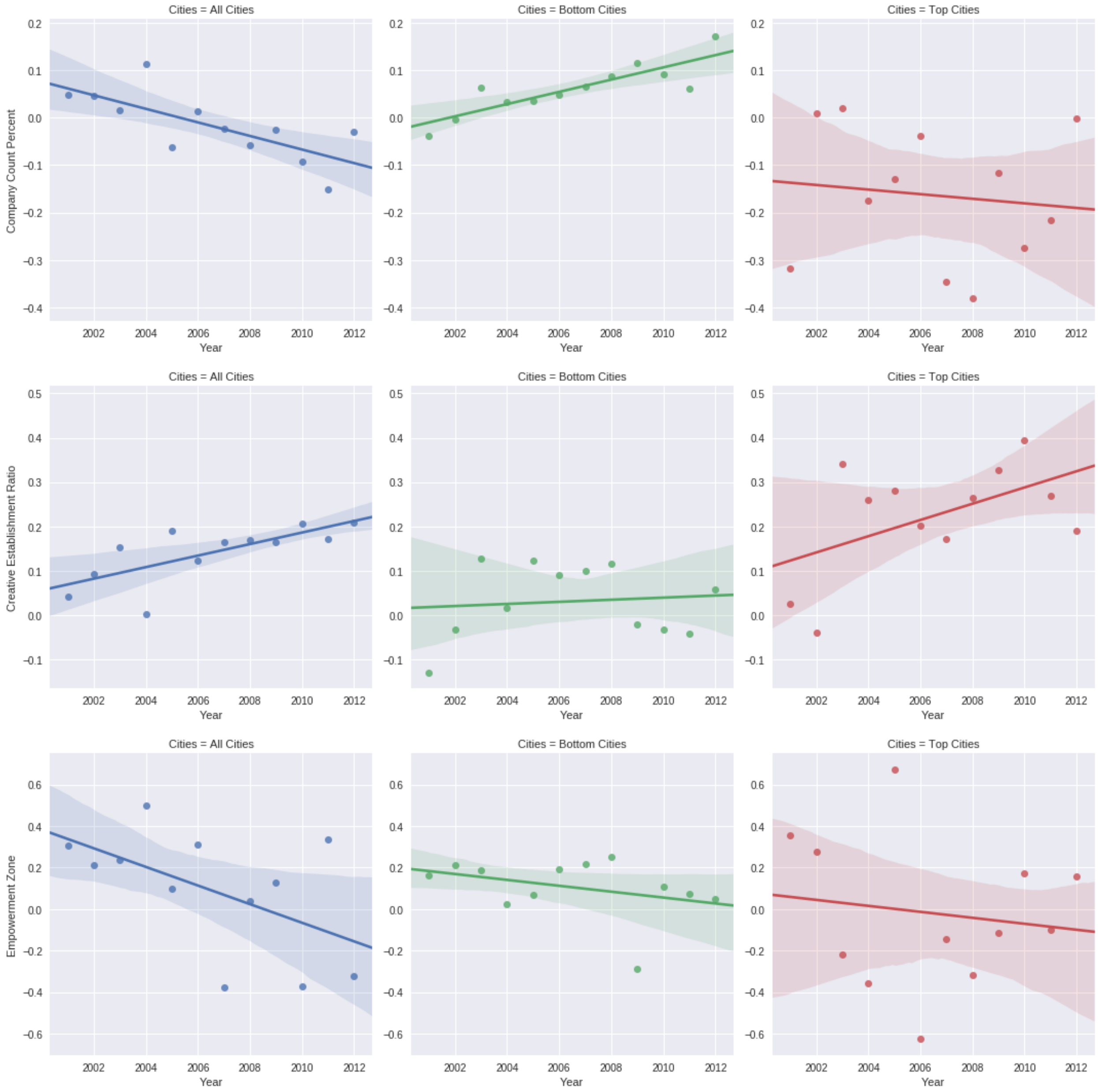
'Other music groups and artists'

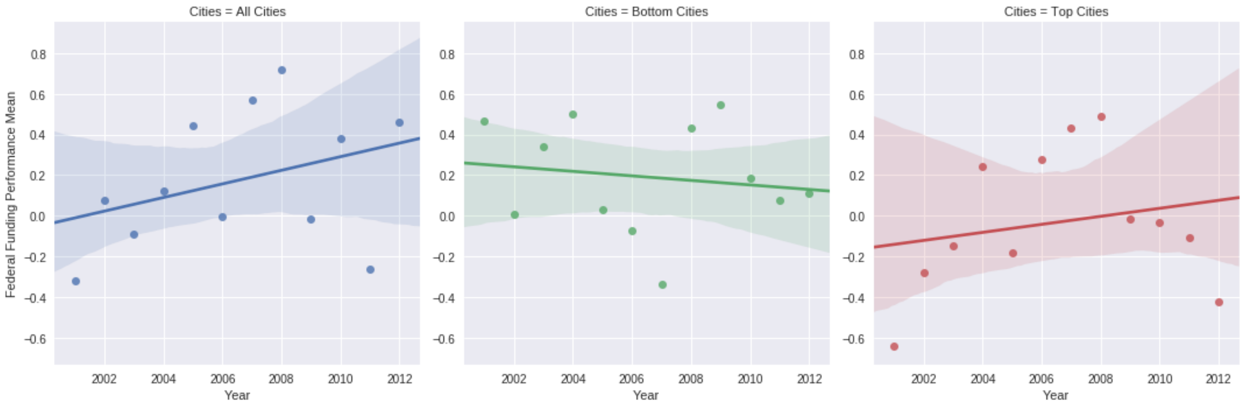
'Agents and managers for artists, athletes, entertainers, and other public figures'

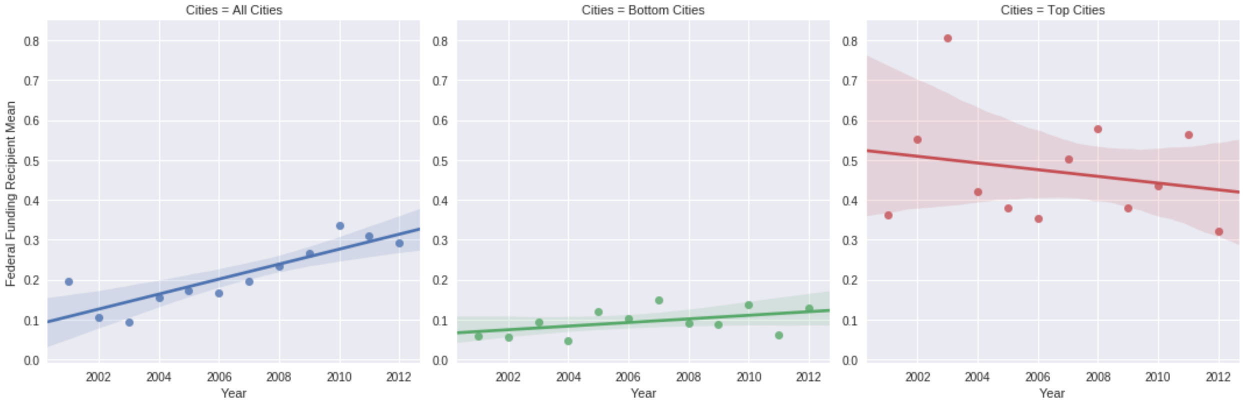
'Independent artists, writers, and performers'

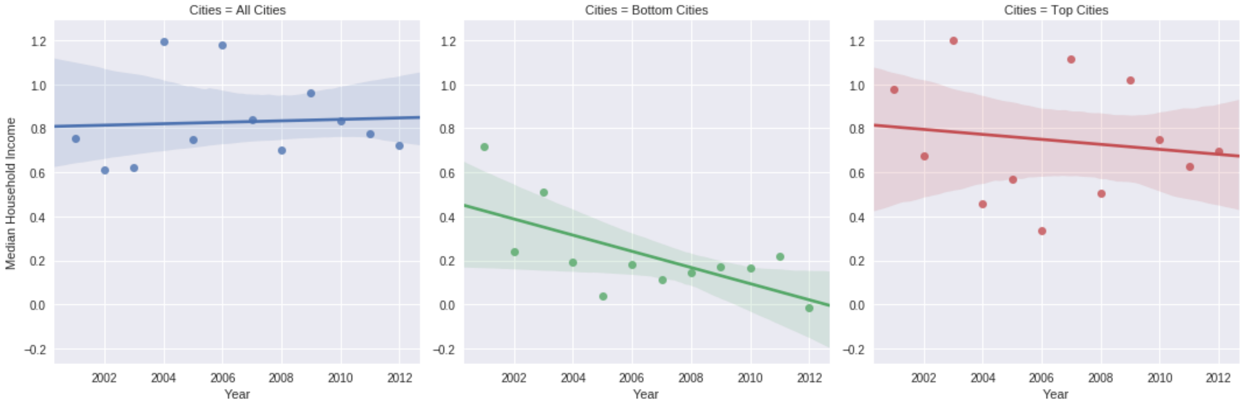
6. Logistic Regression Feature Influence Graphs

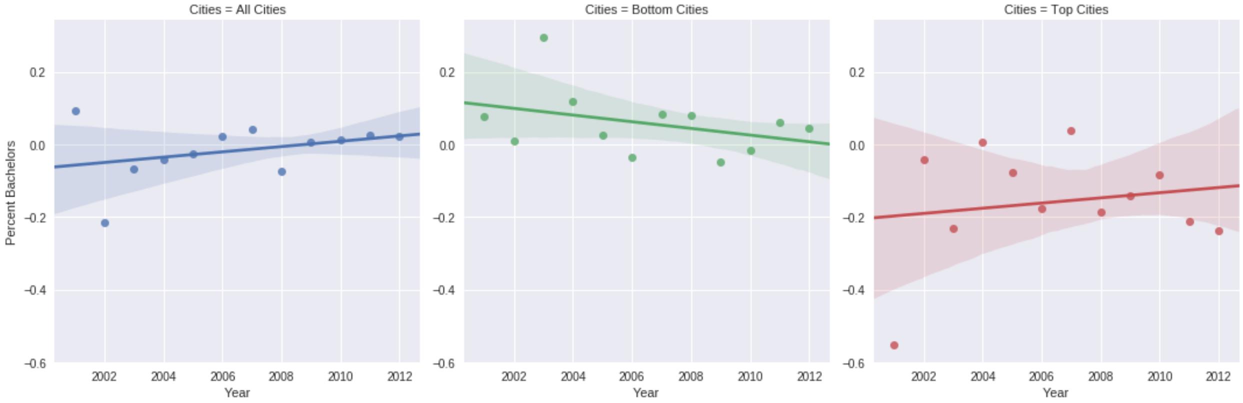
a. 6.1 Assigned Patent Index

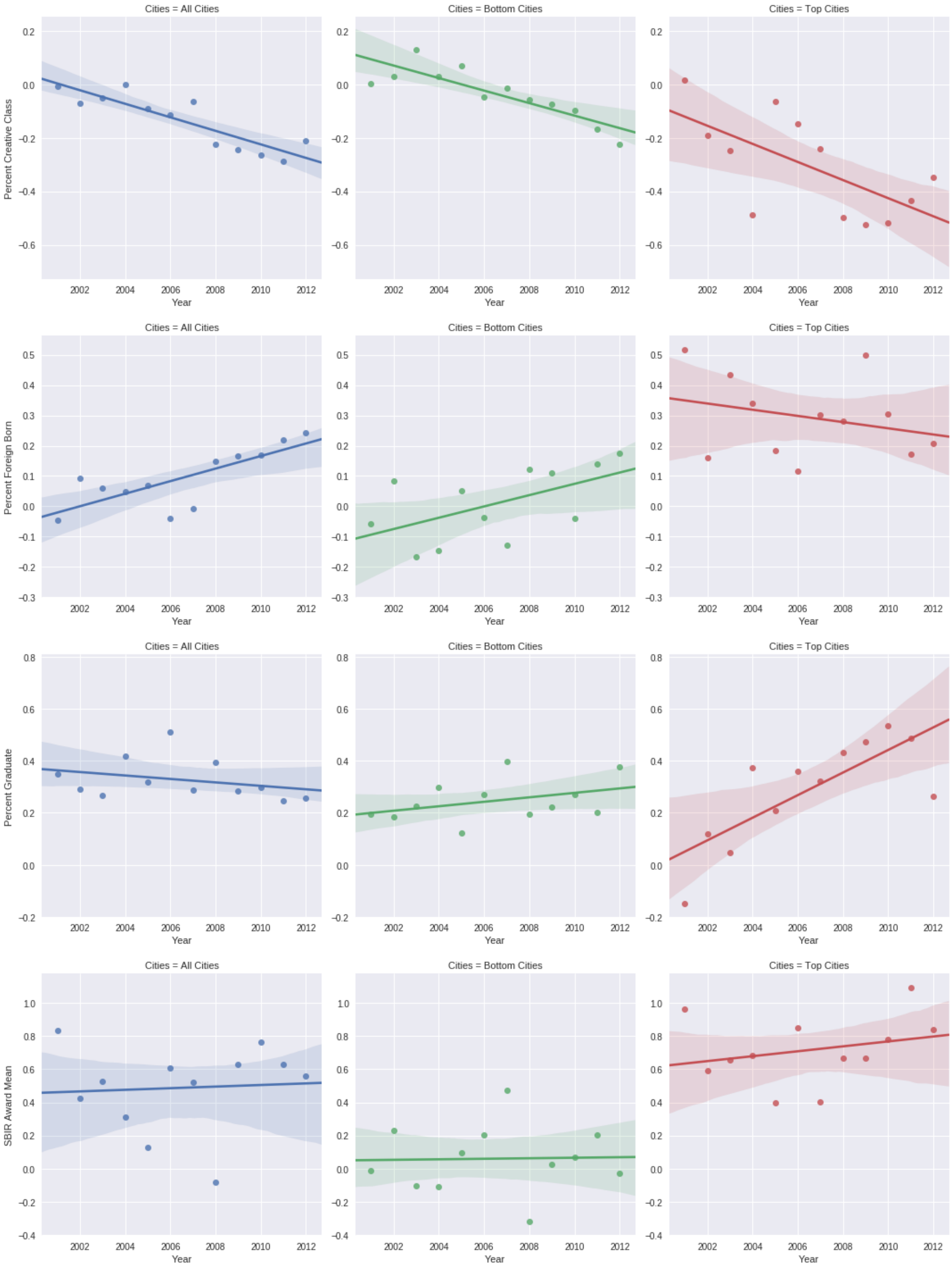



Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Federal Funding Performance Mean
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
2002
2004
2006
2008
2010
2012
Year

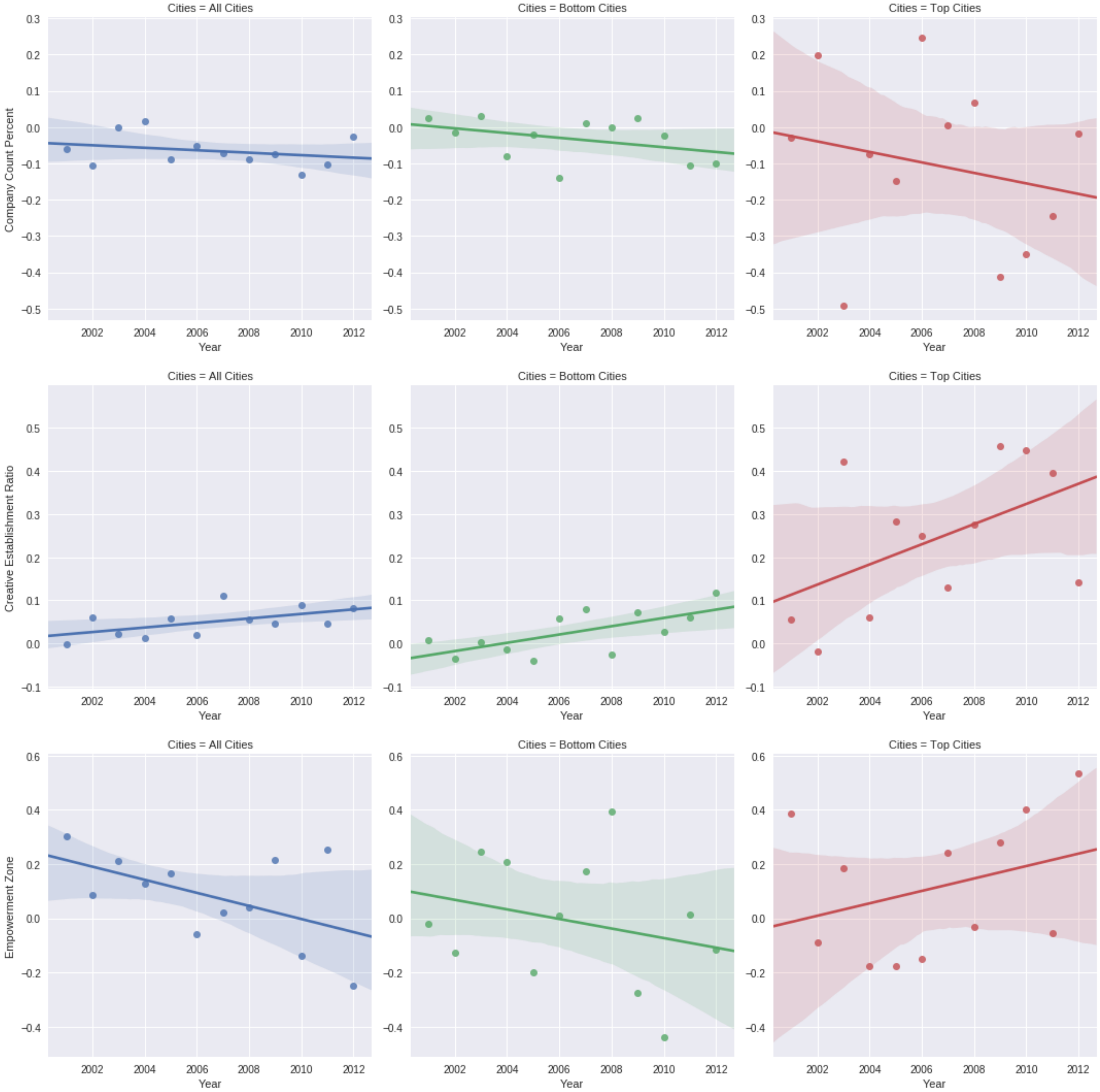

Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Federal Funding Recipient Mean
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2002
2004
2006
2008
2010
2012
Year

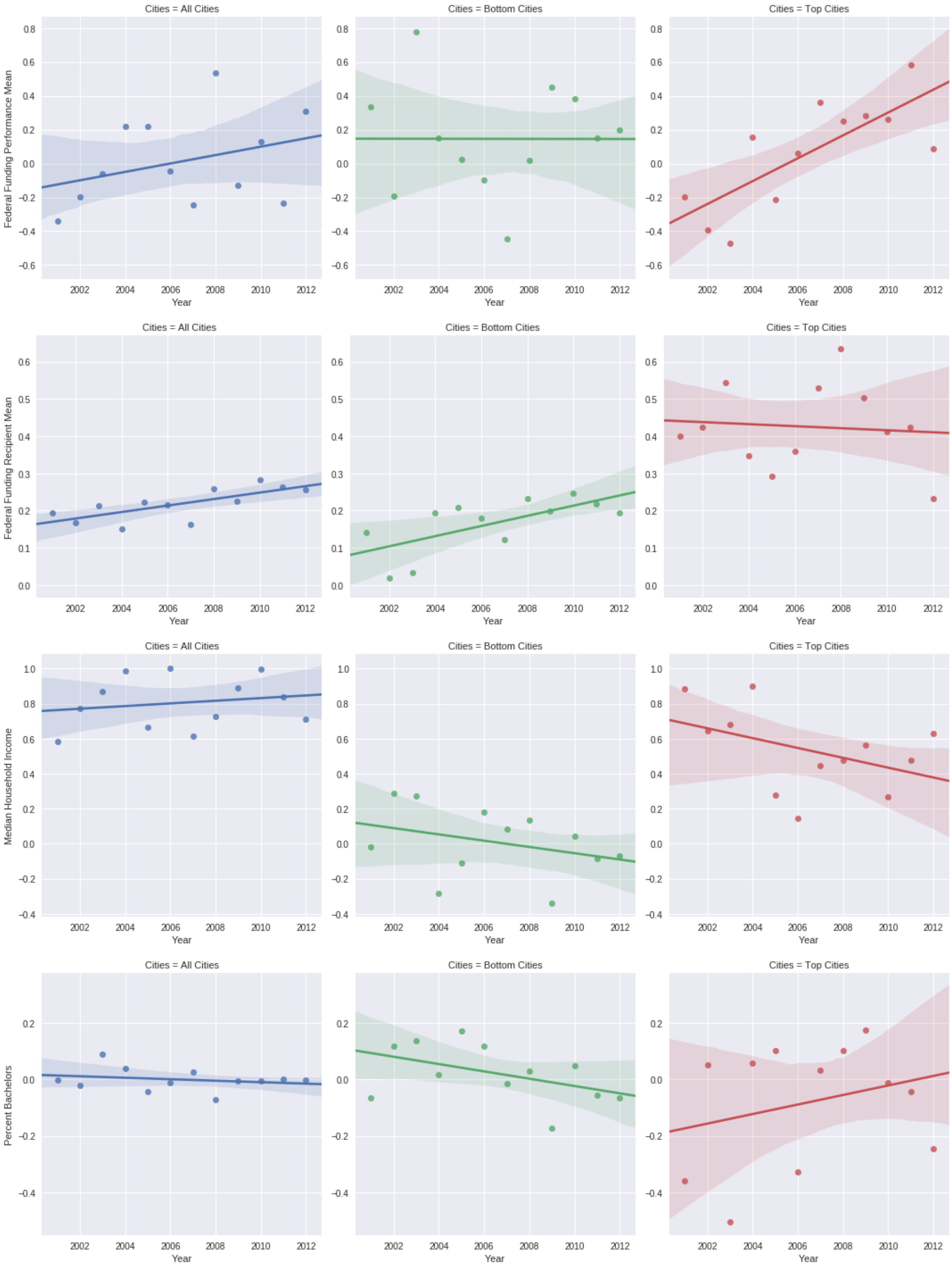

Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Median Household Income
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
2002
2004
2006
2008
2010
2012
Year

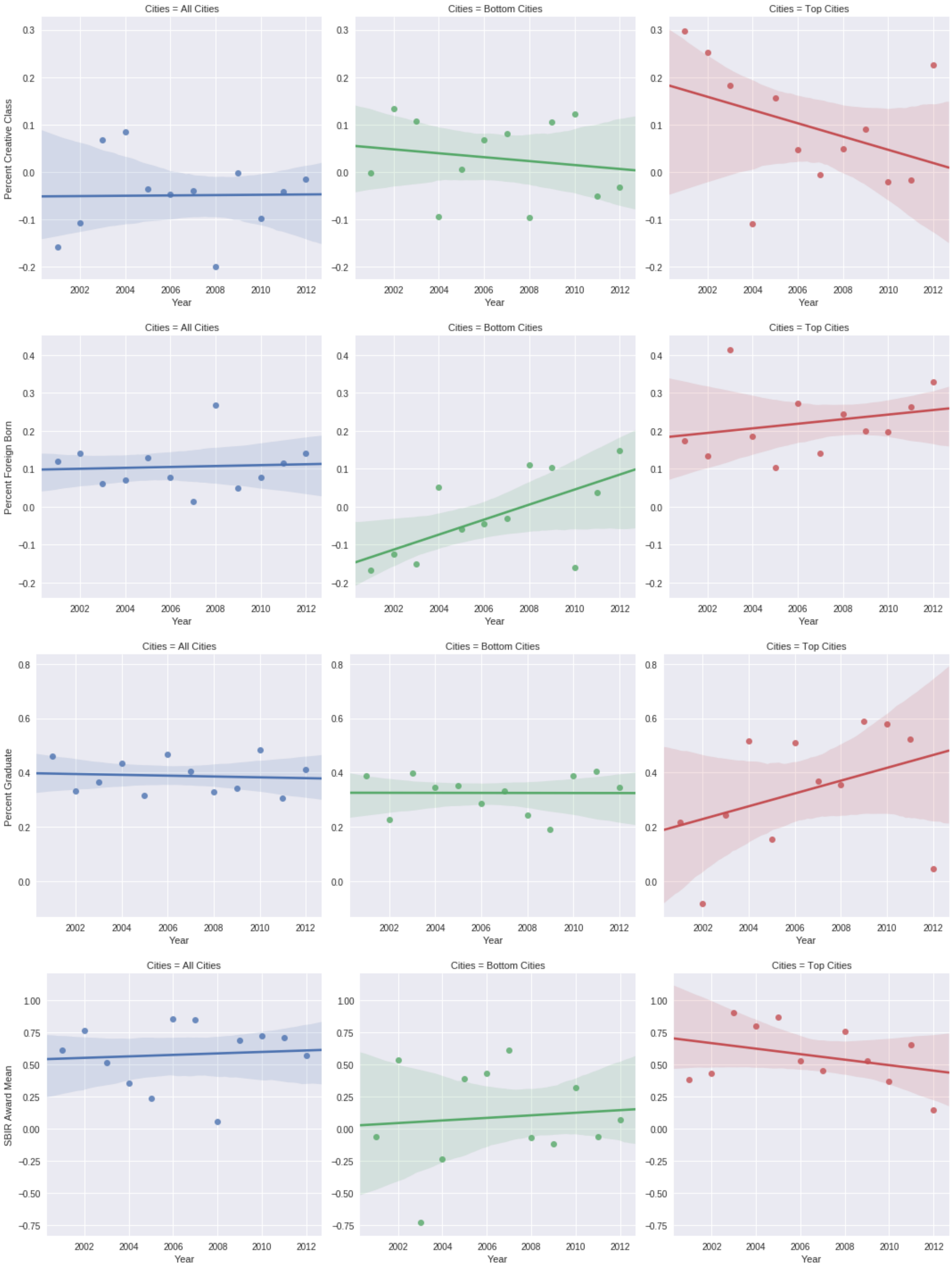

Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Percent Bachelors
0.2
0.0
−0.2
−0.4
−0.6
2002
2004
2006
2008
2010
2012
Year


Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Percent Creative Class
0.2
0.0
-0.2
-0.4
-0.6
Percent Foreign Born
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Percent Graduate
0.8
0.6
0.4
0.2
0.0
-0.2
SBIR Award Mean
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
2002
2004
2006
2008
2010
2012
Year

b. 6.2 Invented Patent Index

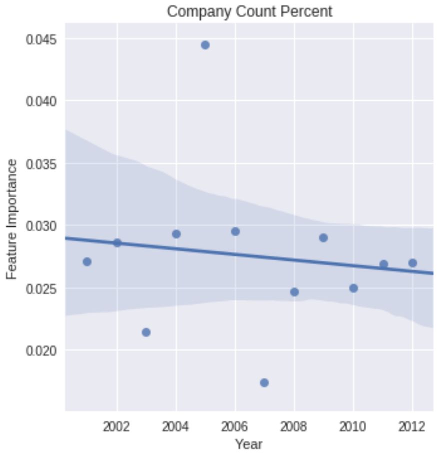


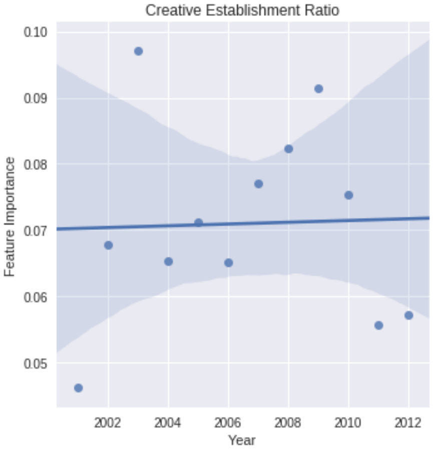

Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Federal Funding Performance Mean
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
2002
2004
2006
2008
2010
2012
Year
Federal Funding Recipient Mean
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Median Household Income
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
Percent Bachelors
0.2
0.0
-0.2
-0.4

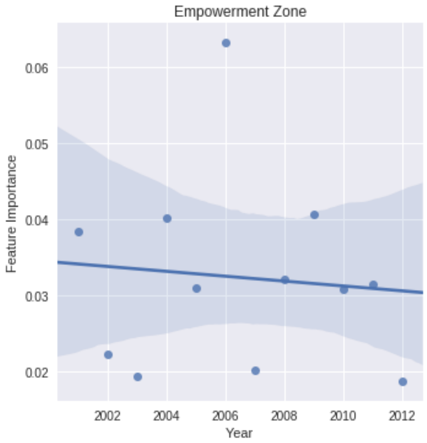

Cities = All Cities
Cities = Bottom Cities
Cities = Top Cities
Percent Creative Class
0.3
0.2
0.1
0.0
-0.1
-0.2
Year
2002
2004
2006
2008
2010
2012
Percent Foreign Born
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
Percent Graduate
0.8
0.6
0.4
0.2
0.0
SBIR Award Mean
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75

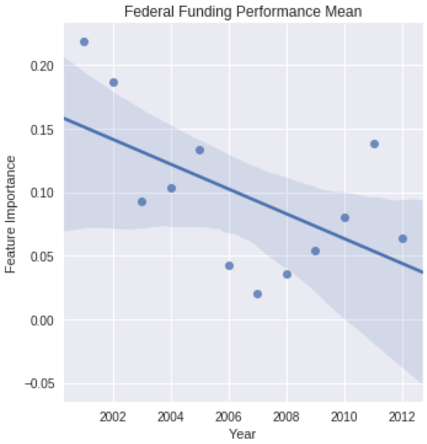
7. Random Forest Regression Importance Graphs

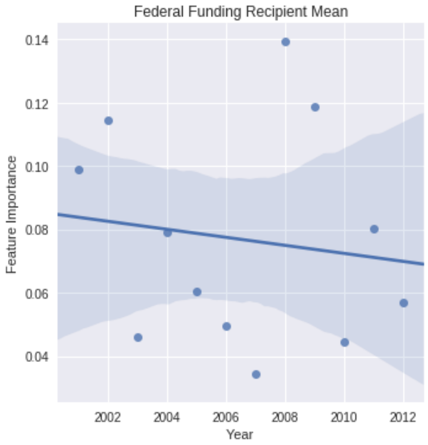
a. 7.1 Assigned Patent Index

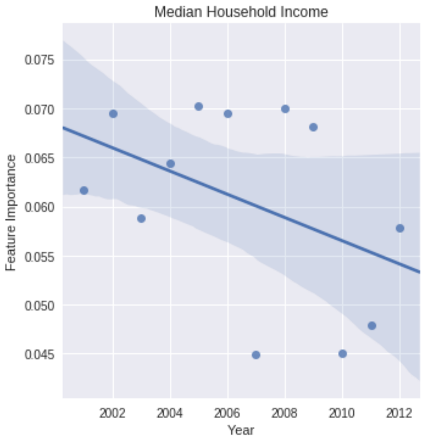




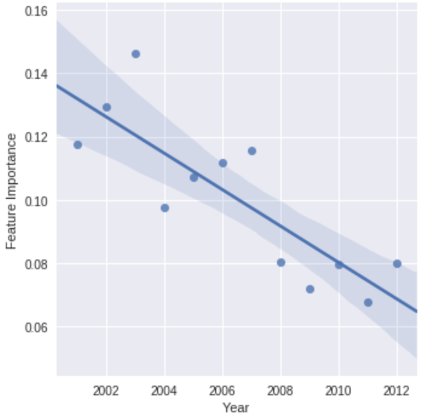




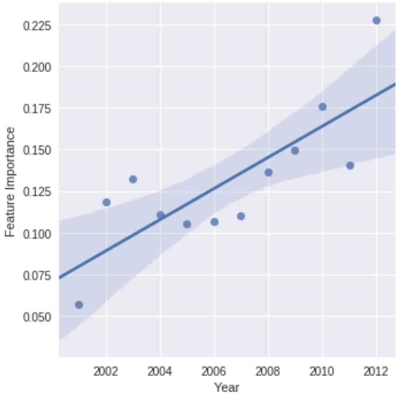




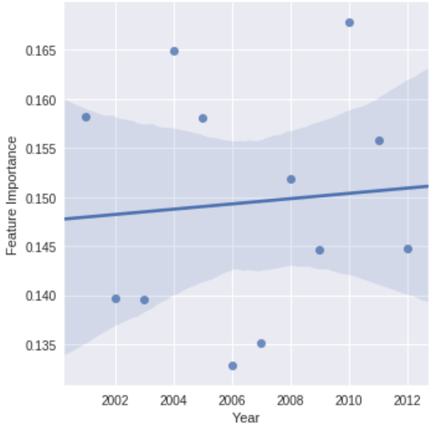
Percent Bachelors



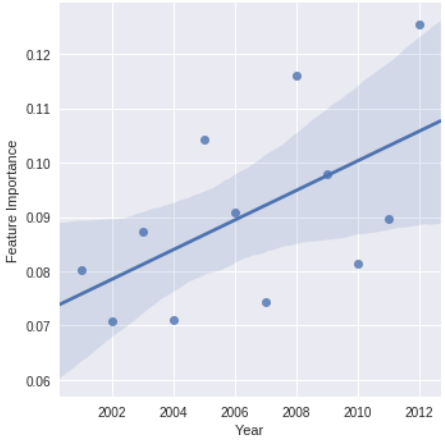
Percent Creative Class



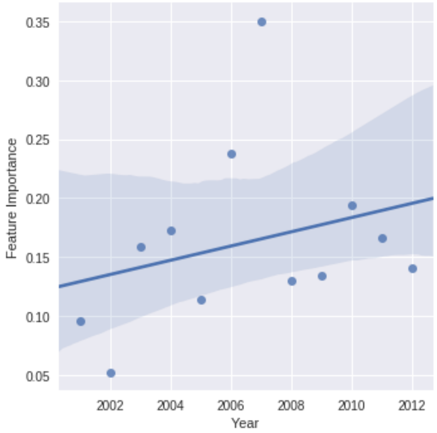
Percent Foreign Born



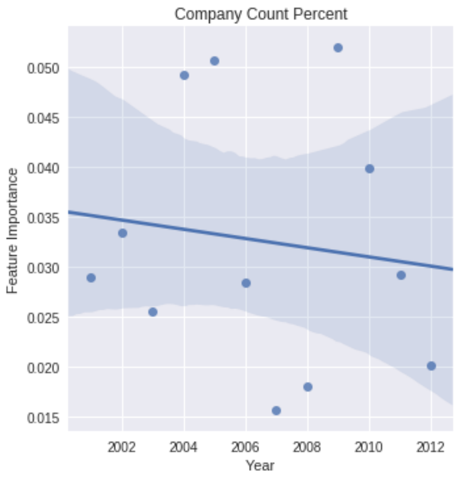
Percent Graduate

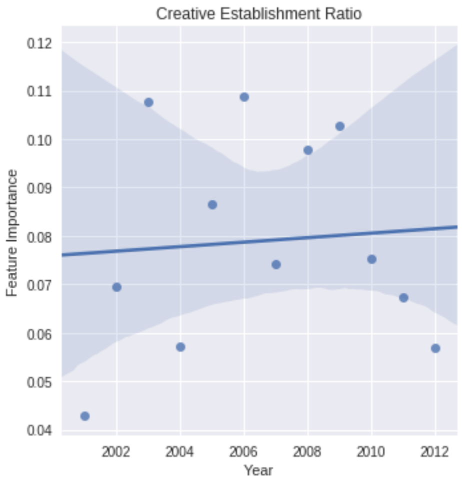


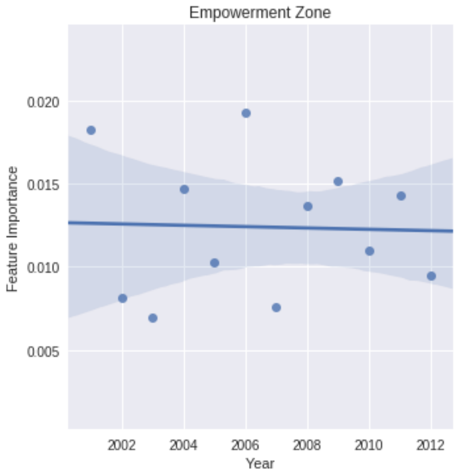
SBIR Award Mean

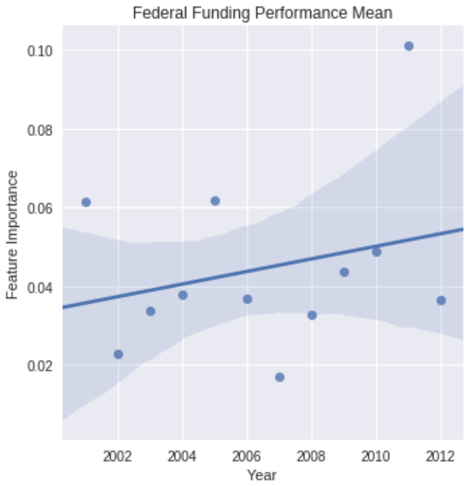


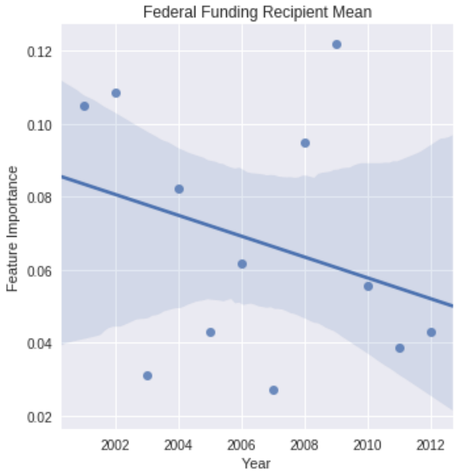
b. 7.2 Invented Patent Index

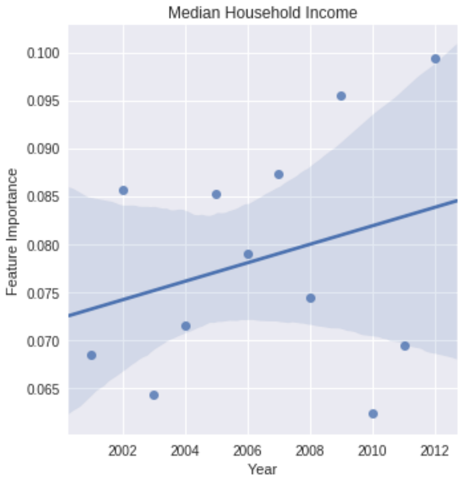


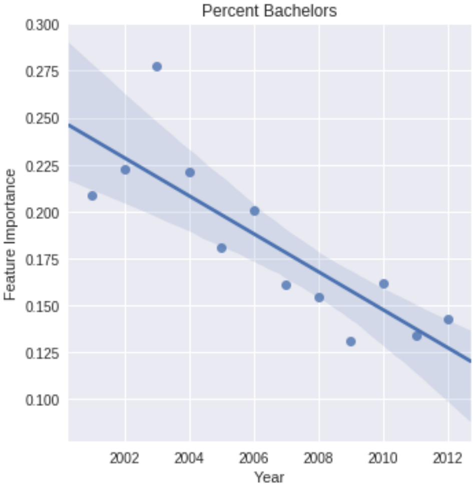


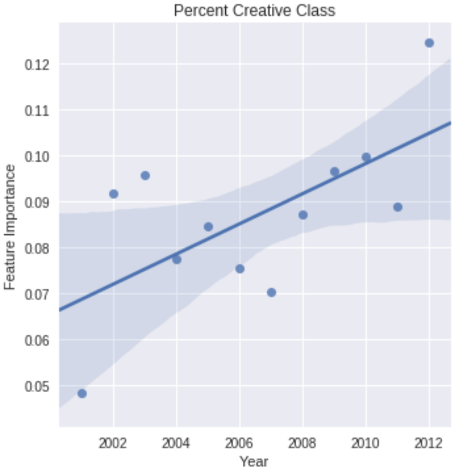


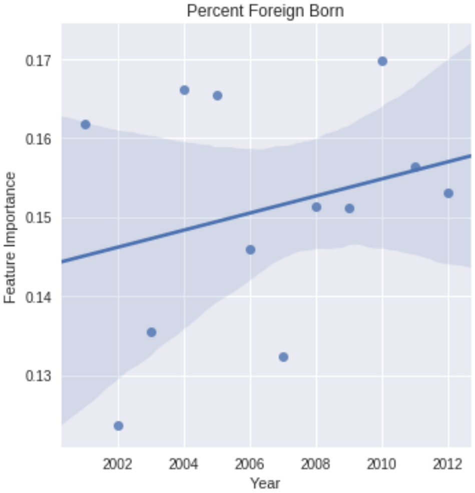


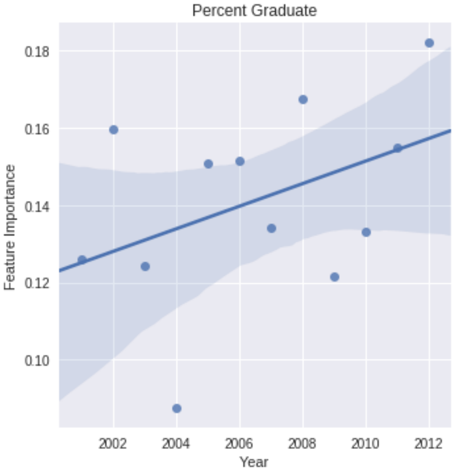


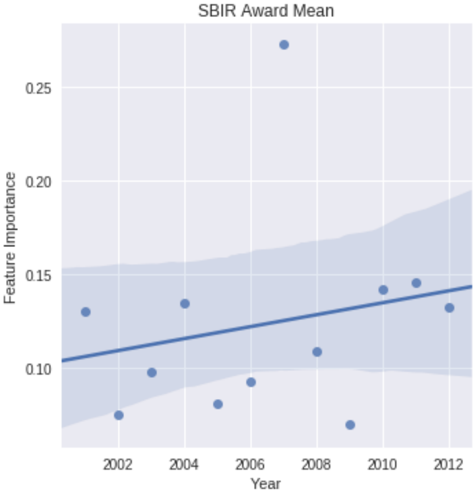



Percent Bachelors
Feature Importance
0.300
0.275
0.250
0.225
0.200
0.175
0.150
0.125
0.100
2002
2004
2006
2008
2010
2012
Year


Percent Creative Class
Feature Importance
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
2002
2004
2006
2008
2010
2012
Year


Percent Foreign Born
Feature Importance
0.17
0.16
0.15
0.14
0.13
2002
2004
2006
2008
2010
2012
Year


Percent Graduate
Feature Importance
0.18
0.16
0.14
0.12
0.10
2002
2004
2006
2008
2010
2012
Year


SBIR Award Mean
Feature Importance
0.25
0.20
0.15
0.10
2002
2004
2006
2008
2010
2012
Year