A City and Its Patents - Factors to Become an "Innovation Hub"

Nathan Caplan, Tingyu Chang, Rohun Iyer, Sarah Sachs

April 30, 2019

Abstract

In recent years there has been great interest around "innovation hubs" and how a nexus of technological development is formed. This capstone project will investigate the factors that facilitate innovation growth within a city. To understand how this process develops, we will analyze more than thirty-five years of patent data in the United States from 1976-2014. Our regression analysis will explore three dimensions that influence the growth of innovation: regulation, socioeconomic diversity, and the spatial geography of the city.

1. Introduction

The first Patent Act of the U.S. Congress in 1790 was titled "An Act to promote the progress of useful Arts."[1] Ever since, the patent has been a key representation of innovation and progress in the United States. Keeping that in mind, this project looks into the ecosystem of innovation in the United States from 1976-2014 using the registry of patents as its foundational aspect. Our goal is to develop an innovation score and understand the urban structure that facilitates the growth of an innovation hub.

There is 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 a technology hub such as Silicon Valley. [2]

Patent analysis is an appropriate approach for investigating the features that advance innovation in a city. Using patents information is a well-established strategy for understanding the development of technology and spread of information as it relates to economic growth. As explained by O hUallachain, an Urban Planning professor, used there has been previous use of patent data to explore topics such as inventive activity, the scale of manufacturing in cities and what factors drive technological advances in a region.[2][3][4][5]

For our project, we collect, engineer, and compare two sets of features: to measure innovation itself and to explain its growth in a city. For the former, we use the thorough patent data from Patentsview.org containing detailed information on every patent assigned to a United States based organization from 1976-2014.[6] Patents and their associated citations have shown to indicate the level of a firm's innovation capacity.[7] By aggregating this data by city, we are able to scale this model to evaluate a city's innovation capacity.[8]

Our second set of features contains socioeconomic, spatial and regulatory data from the past three decades that help explain the foundation upon which a city's innovative culture is built. Previous work on measuring city growth and economic output uses Census data as independent variables. Bettencourt et al. investigated how changes in population and income would support city growth in terms of innovation.[9] The paper also specifically points to patents as an innovation measurement suggesting these features are imperative to our study. Jonathan Quigley touched upon these and other features when exploring the relationship between urban diversity and economic growth. He added that features, such as population density, number of commuters, and economic diversity are involved in or indicate economic and innovative growth in cities.[10]

Urbanist Richard Florida continued to support these claims, saying that cities that 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, the arts -- and other post-higher-education occupations, is a driving force of economic development.[11]

Work from Breandan O' Huallachain and Timothy Leslie explored the spatial relationship of knowledge spillovers on the distribution of "innovation hubs" throughout the United States. Using regression analysis, his work explores patent growth rates and how it spatially converges and exhibits a spillover effect.[2] Our approach will utilize similar geographical influence such as Moran's I and LISA cluster maps to identify regional groupings of "innovation hubs." However, we will go a step forward to identify the network of subject material that has grown in different hubs and how it relates to overall growth. Additionally, we will consider the fractal dimension of land use as it relates to the city structure and how it contributes to innovation.

In 1997, the National Research Council has mentioned R&D investment as one indicator in measuring input of innovation.[12] The Information Technology and Innovation Foundation also mentioned “The Innovation Success Triangle” in one of its reports and one leg of the triangle is strong innovation policy system. It elaborates that such a system includes investment in innovation infrastructures, funding to technology and industry research and active policies on taxation to spur innovation.[13] Therefore, supportive regulations by government play an incentive row for cities to thrive in business and technology development and stimulate innovations to happen within the area. Different types of regulations include federal programs for R&D funding, tax benefits, government subsidies and etc.Therefore, our approach will take regulations in multiple aspects into account in our model to figure out at what extent of time and how they have driven what kind of cities into innovation hubs.

Over the next few pages, we detail the data, methods and limitations in creating our model. First, an in-depth analysis of the data provided on patents and the key features we have extracted and engineered to create a measure for innovation. Next, we expand on the features we have gathered that we believe will have an outstanding effect on innovative abilities of a city and previous literature that supports our hypotheses.

2. Problem Statement

Using the wealth of patent, demographic, geographic and business data available to us, our logistic regression models will identify the key features that determine a future of economic and innovative success of metropolitan statistical areas within the United States. Developing a model to determine and predict future inflection points will give an insight into the future innovative landscape for cities and investors. We will create an interactive application that will allow users to see more in depth visual analysis of the innovative landscape over time. The dimensions of innovation we will explore include the city's dynamic regulatory environment, socioeconomic structure and spatial density.

3. Data and Methods

3.1 How to Measure Innovation?

Since 1976, publicly available patent data has acted as a resource for tracking and analyzing the cutting edge, the products of the future. Using the API provided by PatentsView, our project will use available features from patent applications to create a measure for innovation.[6] Using this measure we will classify cities as innovation hubs. But innovation does not advance equally; different decades experience different innovative advances. To address this, we are accounting for trending technologies and patent properties using patent classifications, types, organizations along with other engineered features to create a LASSO regression to identify key variables that lead to increased overall patent production and classify innovation hubs.

The PatentsView data gives a precise description of every successful patent application. This description includes geographic features such as assignee and inventor locations, patent classification features pulled from World Intellectual Property Organization (WIPO), assignee features such as organization and organization type, and even the full patent abstracts.[14]

Accounting for economist Sadao Nagoaka's work on patents and their role in identifying innovation, features such as patent citation rate, assignee type, and patent type were identified as relevant to determining centers of innovation.[2][15]

The fields help paint a picture of each city for each year through an innovative lens. Each city contains fields on the number of assignee patents, inventor patents, the ratio of inventor to assignee patents, citations for assignee and inventor patents, the ratio of citations to patents, proportion of patents being classified into the 8 high-level WIPO classifications, the type of assignee organization (including US/Foreign corporations, governments and individuals), and the patent type (utility, reissue, design, defensive).

A major benefit of working with patent data and innovation (as is loosely defined by the public), is public perception. In our exploratory analysis, there were various cities we identified that had precipitous drops in success or astronomical rises to become the center of 21st Century progress. Using these cities, we compared these changes with their patent and patent citation outputs to support any fledgling hypotheses on patents as indicators of innovation. Figures 1 and 2 show the varying fortunes of Pittsburgh and San Jose, two urban areas that have seen opposing fortunes in the last 40 years.

There are clear indicators to measure innovative performance. Using features such as normalized citation rates, assigned/invented patent proportions and patent classification distributions over a logistic regression will give us an indicator of the innovative split. Furthermore, Figure 3 shows that there appears to be a slight dispersion of assigned patents and more clustering among the inventor patents, indicating a changing innovative environment.

3.2 Determinants of Innovation:

Using previous studies, we have identified three main determinants of an innovative ecosystem: regulatory, socioeconomic, and spatial. Supportive regulations by government play an incentive role for cities to thrive in business and technology development and stimulate innovations to happen within the area. Different types of regulations include federal programs for R&D funding, tax benefits, and government subsidies. We believe that various aspects of regulatory support are influential features in helping a city become an innovation hub. Literature also suggests that city growth can be measured by changes in demographic and economics factors. We believe that innovative growth may follow this pattern as well.

Additionally, when an innovation hub starts to develop, there is a spatial component that can shape the ecosystem. We plan to look into the urban density of the cities, and classify the fractal patterns that are distinguishable as a factor for innovation. Aspects such as spatial agglomeration, spillover - both micro and macro, accessibility and collaboration will be analyzed with regard to physical space and time.

Hypothesis

We anticipate that cities implemented with supportive regulations are more likely to have significant growth in innovative development. Also, we expect to see changes in demographic diversity, job diversity, increased wages, and increase in population density as factors towards innovation growth, where patents are the innovation. There is a spatial factor that is a key component of a city's likelihood to become an innovation hub. A city’s progress with innovation is influenced by the geospatial structure that exists, such that a direct relationship is present.

3.2a The Regulatory Environment

Economist Knut Blind’s work focused on determining regulations’ influence on innovation. Using regression analysis on different combination of six regulations, the result showed how social, economic and institutional regulations have different impacts on innovation in short-run and long-run.[16] Similarly but more narrowed down into city level, we plan to take multiple public policies and regulations into account when building our model to understand their extent of influence in innovation. Currently, we are looking into federal awards -- money that the federal government has promised to pay to companies, organizations, government entities or individuals by contracts, grants, loans or direct payments. Further, we will look at two smaller scopes: the Empowerment Zone and the Small Business Innovation Research (SBIR) program. The empowerment zone is a tax incentive and public funding program initiated by the US Department of Housing and Urban Development since 1994 that is intended to create job opportunities for economically distressed areas. And the SBIR program is a federal funding program that enables small businesses to get financial awards from federal agencies' R&D budgets which has helped thousands of small businesses with over $100 million every year since 1982.[17]

3.2b Socioeconomic Factors

The United States Census Bureau is the umbrella source of our demographic, employment, and geospatial statistics. Among the thousands of features collected by the Bureau, we have identified many that may indicate economic, multicultural, and innovative growth within a city. Some of these features include population demographics, number of employers and employees by job type, population density, and income. We hope to find relationships between how these features may change over time how innovation growth or decline may result. For the purpose of this research, geographies selected include place, census tract, and county.

Using the Census and American Community Survey (ACS) APIs, we extracted features for 2015 and 2010 at the census tract levels for the entire US and aggregated these findings to specific cities using the open source OSMNX city shape-files available, even down to the townships (such as Armonk, NY, headquarters of IBM) level. Demographic and household data can be collected decennially going back to 1970 from the IPUMS National Historic GIS , however we hope to find more in-depth data similar to that of the census and ACS.[18]

Bettencourt et al. touch on the necessary steps to measure the new growth outputted from cities. Now that a majority of the world's population lives in cities, there are now multiple ways to measure growth than just simply population. Although assumptions have been made that cities are the center of innovation, until his work there had yet to be a quantified relationship. He finds that population growth and income are a key factors in innovation growth, which they list as patent output.[9]

Jonathan Quigley discusses the relationship between urban diversity and economic growth. He states that increased density and commuters equate to city sprawl, where density indicates more people living in a confined space and commuters indicates people are living further away to work in the city.[10] Florida and Gates discuss the benefits to innovation with the rise in demographic diversity, economic diversity, and of the 'creative class'.[11]

3.2c Spatial Factors

Research from Economist Xavier Tinguely supports the use of spatial determinants for exploring the development of innovation and level of competitiveness in a city. Where clusters of spatial agglomeration exist as a driving factor for further innovation.[19] Furthermore, the work of Gerald Carlino and Robert Hunt, from the Federal Reserve Bank of Philadelphia, support the spillover effect of hubs that are near each other. [20]

Using this, we will first identify the spatial agglomeration of the cities. How many cities identified are within a to-be-determined radius to each other. Then using land use and land cover data we will use the proportion of spatial area that is devoted to each component, allowing us to understand the fractal patterns of density and likelihood for stakeholders from a multidisciplinary background to come together to innovate. We will use spatial autocorrelation and regression analyses to identify if a spillover effect is present through the geographical distribution and growth of invention. Using patent data as our proxy to measure innovation, we will explore the spatial autocorrelation that occurs. With fractal geometry we will determine the relationship between innovation and the physical structure of the city, supported by the work of Batty and Longley, who are both urban planners and Professors of Geographic Information Science.[21]

3.3 Data

For federal awards, data are available from 2001 to 2018 and each award has 260 columns of information including award amount, basic information of recipient including city, funding agency. We will break down the data into different types of organization and fields by recipients and use amount of funding as well as number of awards for each city in our model to identify the importance of federal funding in spurring a city's innovativeness.

For the empowerment zone, data are available at city level (binary outcome for each city) and we will include that in our model to see if this factor aids cities turning into innovation hubs throughout these 20 years.

For the SBIR program, data are available from 1983 to 2019 and contain company information including city, award amount, funding agency and topic/field for each awarded project. We will use average amount of funding received and number of businesses awarded for each city in our model to measure if this program has encouraged innovation within cities.

Demographic and socio economic features at the census tract, city, or county, depending on the geographies are available from US Census Bureau APIs. These features, which include population, racial and ethnic demographics, native vs. foreign born populations, median income by gender, and occupation data, will be used to calculate two composites as defined by Urbanist Richard Florida. Of the features mentioned, minus the latter two, they will be used to calculate a diversity index where we normalize the demographic features to the total population. We will do the same for the income and occupational data to create an economic diversity index normalized to total jobs and median income.[11]

The U.S. Conterminous Wall-to-Wall Anthropogenic Land Use Trends will provide us with a 60-m raster dataset that covers the conterminous United States throughout five time periods: 1974, 1982, 1992, 2002, and 2012. Main characteristics will provide the proportion of land use such as residential, commercial/industrial, transportation and recreation areas.[22] We will use this to identify how the land within the city has changed throughout time with relation to patent subject and development.

We will use the World Topographic Map from ESRI to capture the city boundaries. It has additional characteristics such as water, physiographic features, parks, landmarks, transportation, and buildings that could be used for further analysis. [23]

4. Risks and Mitigation

Federal awards data and the SBIR data have some repeated parts which could result in covariance between independent variables in the model. We will take this into account when interpreting our model and use Gaussian Process and ANCOVA to control the effect of the possible covariance.

There are some risks in using census data. If we chose to use annual data, data would be more recent and telling among all years. However, annual data is the least reliable due to the amount of extrapolations taken by the bureau. We instead chose to only look at five-year data as a result to compromise the risks associated with using annual data.

Creating boundaries for the cities spatial density might pose a risk. There might be some overlap between areas that cannot be determined by their legal boundaries. Additionally, the cultural and administrative boundaries may differ.

While patents have shown to be decent measures of innovation not all inventive ideas are necessarily patented (software is covered by copyright and not patent).[15] Another risk lies in using citations as a measure of patent value as the average number of citations per patent has increased dramatically with the advent of the searchable databases as seen in Figure 4.[24] To address this, we are using normalized citation rates for individual years, however this may still raise issues on determining the value of patents. Another issue is the incredible rise in the number of patents in the past decade versus the late 1970s and 1980s. This rise could have an effect when determining the value of rapid output of patents in cities across separate sectors.

5. Team Roles

Carrying forward, each team member will be working on feature engineering within their data focus and expanding their literature. After this step, the data should be ready to test and create a robust regression model using the specified features and indices above (further specification in Gantt chart below). The focuses are as follows:

1. Rohun - Patents and measuring innovation
2. Christine - Public policy and regulatory features
3. Nathan - Socioeconomic features
4. Sarah - Geospatial features

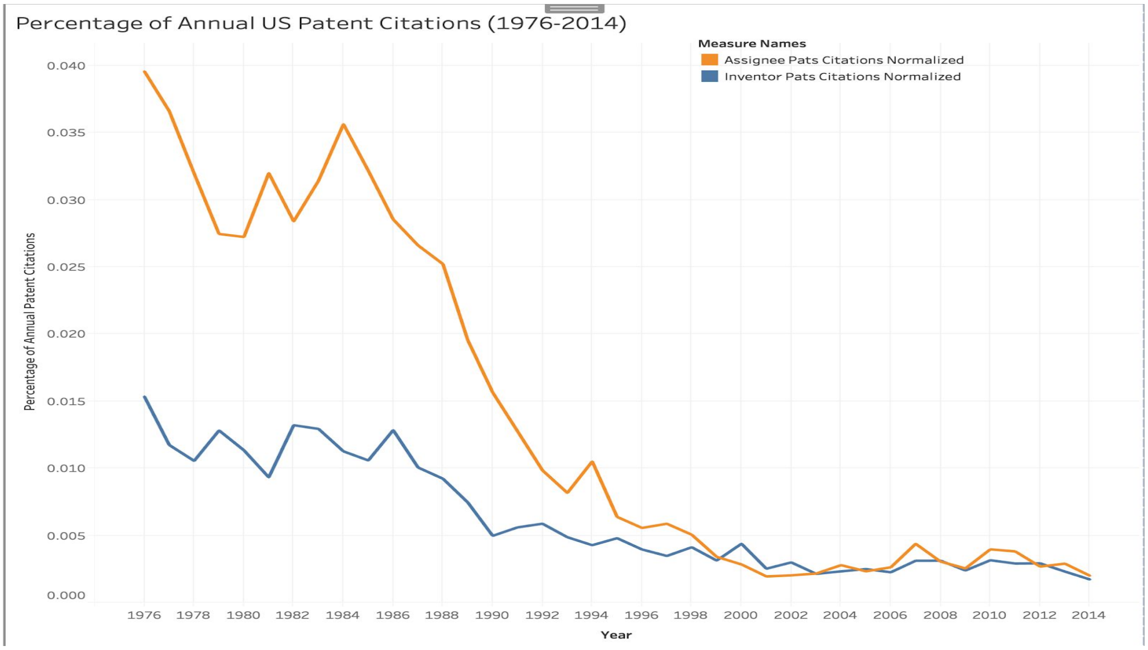
References

1. “U.S. PATENT ACT -- OVERVIEW,” *Legal Information Institute*. [Online]. Available: https://www.law.cornell.edu/patent/patent.overview.html.
2. Breandán 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
3. Sokoloff, K. L. Inventive activity in early industrial America: Evidence from patent records 1790–1846. The Journal of Economic History, 48: 813–50. 1988.
4. Pred, A. R.. The spatial dynamics of U.S. urban-industrial growth, 1800–191, Cambridge, MA: MIT Press.1966.
5. 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.
6. “Why Explore Patent Data?,” *PatentsView*. [Online]. Available: http://www.patentsview.org/api/doc.html.
7. Riitta Katila. Using Patent Data to Measure Innovation Performance. *International Journal of Business Performance Management*. 2000.
8. David Audretsch Zoltan Acs. Patents as a Measure of Innovative Activity. *Kyklos*. 1989.
9. 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.
10. J. M. Quigley, “Urban Diversity and Economic Growth,” *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 127–138, Apr. 1998.
11. 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.
12. "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.
13. Robert D. Atkinson, Understanding the U.S. National Innovation System, The Information Technology and Innovation Foundation. June 2014. [Online]. Available: Link
14. *International Patent Classification (IPC)*. [Online]. Available: https://www.wipo.int/classifications/ipc/en/. [Accessed: 30-Mar-2019].
15. 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.

16. Knut Blind, The influence of regulations on innovation: A quantitative assessment for OECD countries. October 2011.
17. "Dashboard," *Dashboard*. [Online]. Available: https://www.sbir.gov/awards/annual-reports. [Accessed: 30-Apr-2019].
18. "Census Data API:" *api.census.gov*, 28-Nov-2011. [Online]. Available: https://api.census.gov/data.html.
19. Tinguely, X.The Nature of the Innovation Process and the New Geography of Innovation. In *The New Geography of Innovation*. Palgrave Macmillan. 2013.
20. 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
21. M. Batty and P. Longley, *Fractal cities: a geometry of form and function*. London: Academic Press, 1994.
22. 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.
23. "World Topographic Map," *arcgis.com*. [Online]. Available: https://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f. [Accessed: 30-Apr-2019].
24. Mario Pianta Daniele Archibugi. Measuring Technological Through Patents and Innovation Surveys. *Institute for Studies on Scientific Research*. 1996.

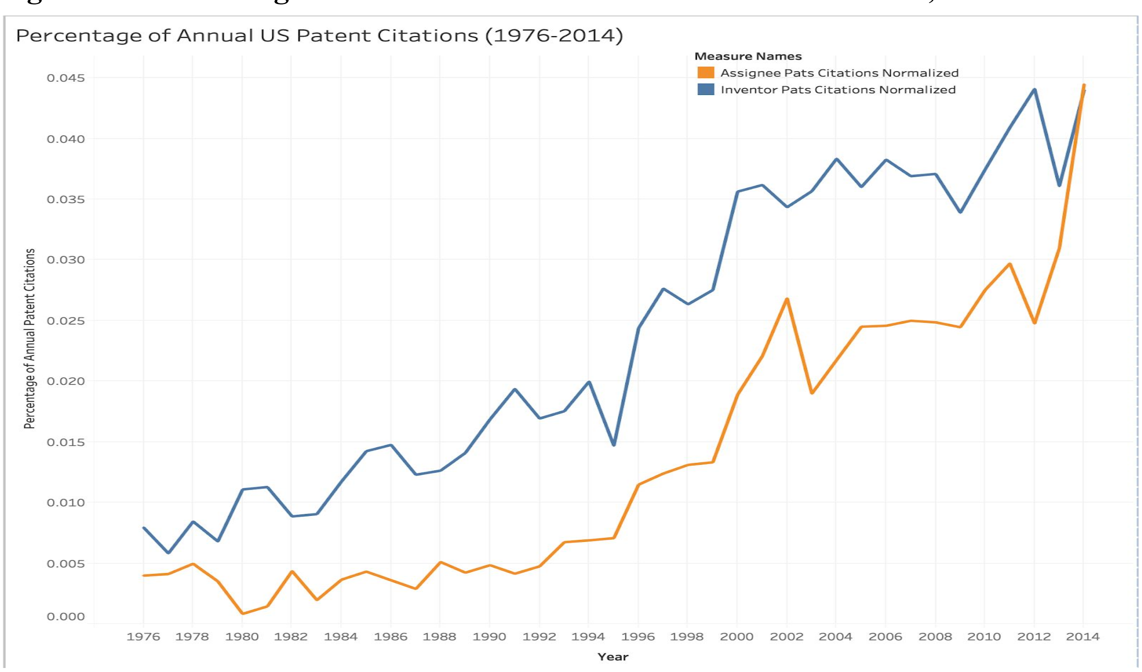
Appendix

Figure 1A: Percentage of Annual US Patent Citations from Pittsburgh, PA



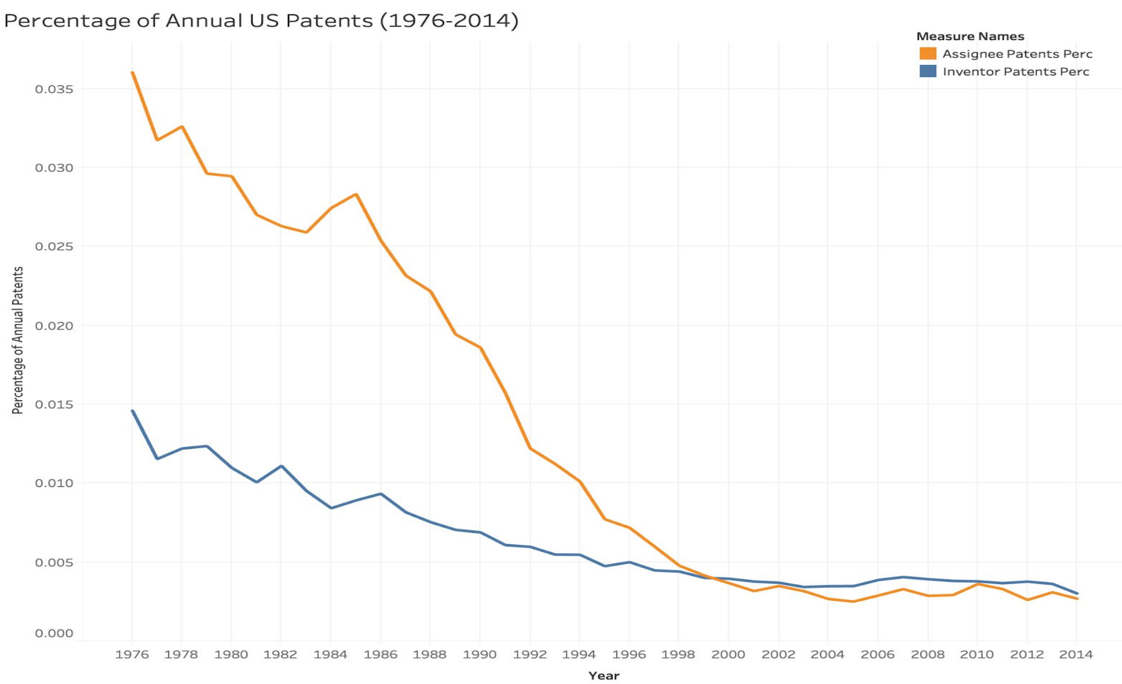
Notes: The plot above shows the share of assigned and invented patent citations coming out of Pittsburgh. In 1976, around 4% of all patent citations in the United States were assigned to organizations in Pittsburgh, whereas in 2014 that number has dropped to 0.2%. Likewise, patent citations with inventors based in Pittsburgh has dropped from about 1.5% to 0.2%.

Figure 1B: Percentage of Annual US Patent Citations from San Jose, CA



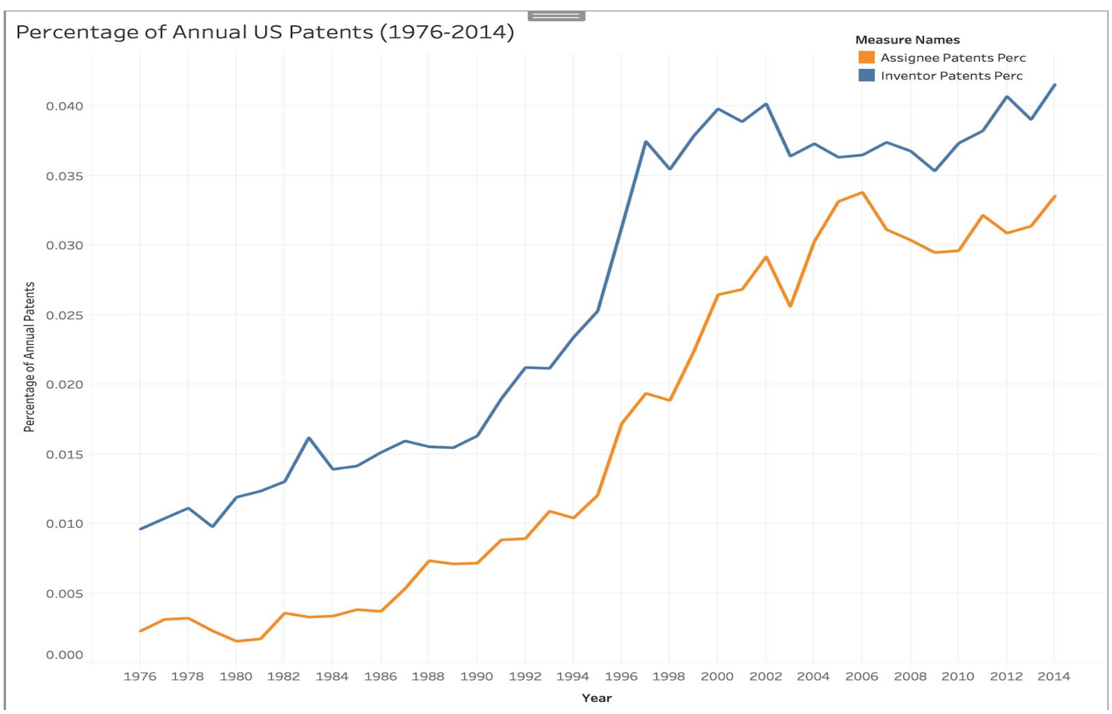
Notes: The plot above shows the share of assigned and invented patent citations coming out of San Jose. It shows the opposite trend displayed Pittsburgh's plot. In 1976, San Jose had around 0.4% and 0.7% of assigned and invented patents, respectively. These numbers both climb to 4.4% in 2014.

Figure 2A: Percentage of Annual US Patent from Pittsburgh, PA



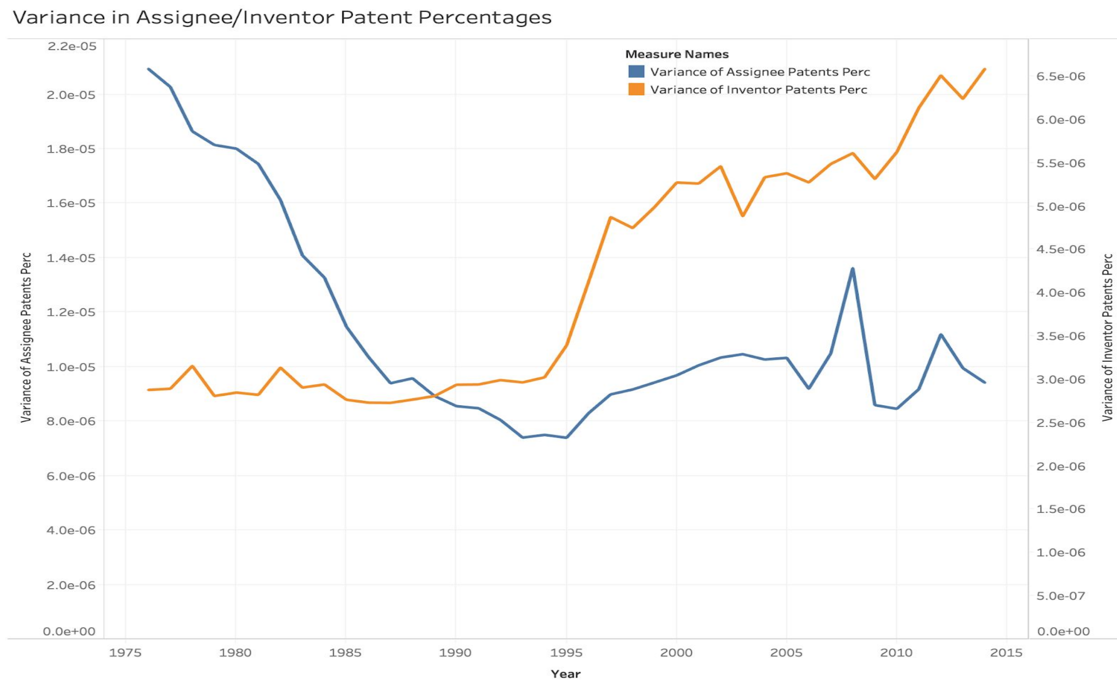
Notes: The plot above shows the share of assigned and invented patents coming out of Pittsburgh. In 1976, around 3.5% of all patents in the United States were assigned to organizations in Pittsburgh, whereas in 2014 that number has dropped to 0.3%. Likewise, patents with inventors based in Pittsburgh has dropped from about 1.5% to 0.3%.

Figure 2B: Percentage of Annual US Patent from San Jose, CA



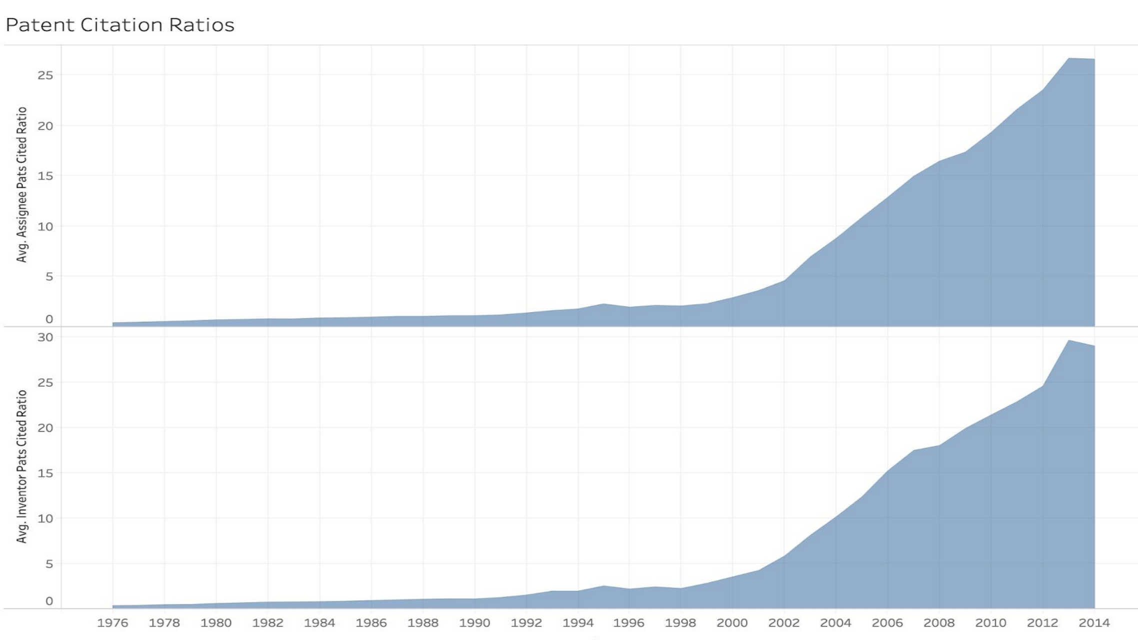
Notes: The plot above shows the share of assigned and invented patent citations coming out of San Jose. It shows the opposite trend displayed Pittsburgh's plot. In 1976, San Jose had around 0.2% and 1% of assigned and invented patents, respectively. These numbers climb to 3.4% and 4.2% of assigned and invented patents in 2014, respectively.

Figure 3: Variance in Assignee and Inventor Patents per City from 1976-2014



Notes: The variance in city patent percentages have shown opposing trends over the last 40 years.

Figure 4: Citation to Patent Ratio from 1976-2014



Notes: As seen here, the average citation to patent ratio has drastically increased. This is attributed to the advent of searchable databases making patent searches easier and more efficient.

Gantt Chart

Person	Task	May				June				July			
		1	2	3	4	1	2	3	4	1	2	3	4
CATEGORY		**Goals**											
Sarah	Join spatial data with patent data	-											
	Aggregate land use data		-	-									
	Calculate proportion of land use per city				-	-							
	Analyze spatial autocorrelation						-						
Rohun	Create regression function to classify innovation	-	-	-	-								
	Use other features to detect future innovation			-	-	-	-						
	Create interactive visual						-	-	-	-			
Nathan	Data wrangling for socioeconomic data	-	-										
	Development of diversity indeces		-	-									
	Apply indeces to predictive innovative city models				-	-	-	-					
Christine	Data wrangling for regulatory data	-	-	-	-								
	Join regulatory data with patent data			-	-	-	-						
	Work on covariation for multiple data						-	-	-				
All - Begin Final Report								-	-	-	-	x	