JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Volume 13, Issue. 2
July 2020
ISSN: 1946-1836

In this issue:

Information Systems & Computing Academic Professionals

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer reviewed academic journal published by ISCAP, Information Systems and Computing Academic Professionals. Publishing frequency is three issues a year. The first date of publication was December 1, 2008.

JISAR is published online (http://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (http://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is about 40%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of EDSIG who perform the editorial and review processes for JISAR.

Journal of Information Systems Applied Research

Digital Transformation and Geospatial Analytics

Ileana Abot
aboti@xavier.edu

Thilini Ariyachandra
ariyachandrat@xavier.edu

Business Analytics and Information Systems Department
Xavier University
Cincinnati, Oh 45207, USA

Abstract

Digital transformation is no longer a mirage that evades organizations. Many companies have touted to have achieved transformation driven by the digital disruptions stemming from big data, advanced analytics and the internet of things. Companies are now inundated with data and different forms of analytics that they should implement to maximize business value. While many see geospatial analytics as the representation of data in a map, it actually involves a deeper integration of location data with businesses strategies and processes. Geospatial analytics has become a key element in gaining quick benefits from the implementation of organizational analytical infrastructure. These analytical products are often the most visible visuals within organizations that provide evidence of the value of analytics to organizational success and transformation. Case examples described in this study presents evidence for the adoption of geospatial analytics in any business area or industry to help with the analytical transformation of an organization.

Keywords: Geospatial analytics, Digital transformation, data visualization, Predictive analytics, Prescriptive analytics, Organizational success

1. INTRODUCTION

On average, more than 2.5 quintillion bytes of data are generated daily (Data Republic 2019). The exponential growth we have seen in data in the past few decades have impacted every industry. From early adoptors such as those in the banking industry, to utilities, fashion design as well as healthcare, industries have begun to adopt analytics as a means to survive. Every organization today feels the pressure to collect, store and analyze the data for use in strategic and everyday decision making. The value of data in staying ahead of the competition and transforming an organization has become well recognized.

Almost 80 percent of an organization's data is based on location or geography (Fernandes, 2006). Therefore, geospatial analytics has become crucial to any organization to gain the competitive advantage. Yet, many organizations are not leveraging this location based data to make their decisions. According to Hempfield (2014), one of the primary reasons is that few BI platforms offer advanced geospatial functionality. In addition, when it comes to data analytics, most of the literature focuses on generated outputs such as graphs, charts, and diagrams reflecting trends or patterns but often disconnected from their geographic component. Geospatial analytics helps organizations to model, analyze, and interpret business location data to make more informed decisions. Eventually, businesses may need to answer some of the following questions: which would be the best site location for a new distribution center or manufacturing facility, how many customers can I reach within a certain distance, or what is the risk of interrupting the business due to a national disaster?

Geographic Information Systems (GIS) allows businesses to answer all these questions and, furthermore, enables them to predict future scenarios. Organizations are beginning to realize the benefits of better incorporating location data into existing analytics. The industry growth statistics further confirms this as the geographic information systems market is expected to grow from $38.65 billion in 2017 to $174.65 billion by 2027 (Lamb 2018). This paper will review the different types of data analytics and identify/explore the reasons why businesses should focus on the analysis of location data (geospatial analytics). An in depth review of business case scenarios will help to emphasize the need for GIS and geospatial analytics to gain business value. Finally, a final section summarizing the findings and future challenges will be presented.

Different Types of Analytics

Within the business intelligence framework, analytics is recognized as adding visible value to a business: "unless an organization takes action on analytics, such as using it as part of a business process, the true value of the technology will not be realized" (Russom, et al., 2015). Others suggest "analytics" as the latest evolution in the maturing and growth in decision support and now business intelligence since 2010 (Watson, 2012).

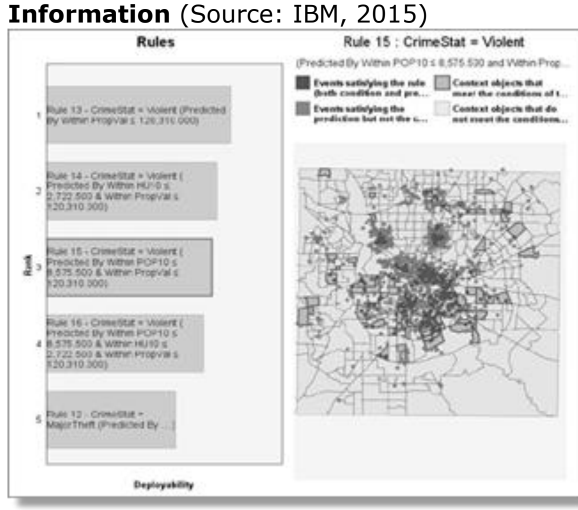
Figure 1. Dashboards (Source: Rainer, 2013)

There are three types of analytics: descriptive, predictive, and prescriptive. Descriptive analytics, the most traditional form of analytics, focuses on "what has occurred" and is backward looking. Reporting, dashboards, and data visualization are the typical applications used. Dashboards facilitate access to timely information and are very user friendly (See Figure 1). Data visualization technologies refer to the utilization of visual formats for the presentation of alternatives or results (See Figure 2).

Figure 2. Data Visualization Technologies (Source: ESRI, 2013)

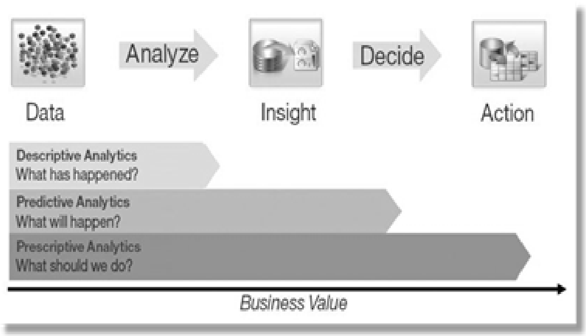
Predictive analytics forecasts "what will occur in the future". The most common methods applied are regression analysis and machine learning (Watson, 2012). This type of analytics is widely applied in marketing when trying to understand customers' needs and preferences. Discovery analytics also falls under this classification and it refers to the relationship findings in vast data sources such as big data. Golden path analysis is a fairly new technique used to identify patterns based on behavioral data. If companies can anticipate a behavior, they could potentially intercede and modify the predicted behavior See Figure 3).

Figure 3. Example of Predicted Crime Hotspots Based on Demographic Information (Source: IBM, 2015)



Prescriptive analytics, as well as predictive analytics, focuses on the future and more specifically on the actions to be taken given a certain scenario; it identifies the optimal solution. According to Kart, in 2015, "predictive and prescriptive analytics will be incorporated into less than 25% of business analytics project, but will deliver at least 50% of the business value". "Prescriptive analytics delivers decisions, not just predictions" (Kart, 2015). It helps answering questions such as which product should I offer to each customer? How do I price my services? Should I acquire a company or enter into a new market? These are all questions related with strategic, tactical, and operational decision-making. Figure 4 summarizes the level of business value added when using each type of analytics.

Figure 4, Analytics and Business Value (Source: IBM, 2015)



Geospatial Analytics

Geospatial analytics is the practice of incorporating data with spatial characteristics in an organization's analytics phase; it adds the dimensions of time and space. Geospatial analytics is helpful in many business areas and can be applied to any industry. Although businesses may understand the value of geospatial analysis, they may not have access to the right tools to maximize its benefits. Furthermore, although almost 80 percent of an organization's data is based on location or geography (Fernandes, 2006) it has been underutilized in most business intelligence implementations. A research study concluded that spreadsheets were used 49% of the time to analyze location data; Geographic Information Systems (GIS) were used only 23% of the time (IBM, 2015).

A geographic information system is a system that allows users to manage, analyze, and display geographically referenced information. It helps users to answer questions and solve problems, to understand and interpret data revealing patterns and trends: to perform geospatial analytics. So why is it important to use GIS or perform geospatial analytics? The example in Figure 5 shows the charts and diagrams generated after a study performed by the City of Cincinnati Health Department. They have determined the average life expectancy in 47 neighborhoods in Cincinnati. Figure 5 located in the Appendix summarizes the findings: residents of Mount/Lookout – Columbia Tusculum live more than 20 years longer than the residents of South Fairmont. These findings reflect significant health inequities that should be addressed.

How can the department identify the reasons and, more importantly, prepare an action plan to modify this pattern? One of the first steps would be to overlay demographic data (See figure 6 and 7 in the Appendix). The information of age, income level, and household structure could help understand the causes of such life expectancy differences. Crime data would be another useful source for identifying reasons of low life expectancy rates: crime related deaths could be significantly higher in specific neighborhoods comparing to others. Finally, any environmental facts could explain the initial findings: high child mortality due to the lack of access to health care or high percentage of homes affected by lead paint. These type of answers cannot be provided by charts and diagrams, only geospatial analysis can guide the department's search for answers.

The data warehousing institute (TDWI, 2015) best practices report for emerging technologies indicates that, based on 344 respondents, only 36% are using geospatial analysis for analytics and, more importantly, almost as many, 30%, plan to incorporate it into their analytics processes within the next 3 years. When it comes to predictive analytics, 49% are currently using it and 37% plan to within the next 3 years. The numbers for prescriptive analytics are lower: 28% currently in use and 37% in the future. This indicates that geospatial analytics is still an immature emerging technology.

Case Study 1: Risk Assessment

A national insurance company, offering insurance policies for auto, life, fire, and property, wants to better understand the liability risk the company may experience from claims due to damage from wind and flooding from hurricanes in the State of Florida (ESRI, 2015). Florida is known for receiving more hurricane activity than any other state and it has a large amount of insurance claims from properties damages from hurricanes. When hurricanes make landfall, they damage buildings, trees, and anything in their path. They are classified from 1 to 5, with 5 being the one causing the most damage. In coastal areas like the Atlantic Ocean, the water damage from flooding is added to the wind damage due to the storm surge. Some of the questions that the company needs answered are: where are the customers located? How does the company best reach them? What are the risks based on location? Will those risks impact costs? At what level should premiums be set? How can the company maximize profit? How much loss can be expected should a catastrophic event occur?

Initially, the mapping of the Total Insured Value (TIV) by property helps to visualize the grouping of policy holders. Also, setting the symbology to graduated symbols relative to the amounts of TIV is useful to reveal the amounts of property values at risk. Figure 8 in the Appendix reveals two patterns: there are more policy holders in the larger metro areas such as Miami and Tampa, and the TIV seem to be higher in the same areas. The map on the right of Figure 8, called a heat map, indicates clustering: areas with higher TIV values appear yellow and the ones with lower TIV values appear blue.

An interesting further analysis would be to overlay the paths of hurricanes most likely to occur. The National Oceanic and Atmospheric Administration (NOAA) compiles this type of information. By buffering such paths, it is possible to simulate the paths generating the heaviest damage; 90 kilometers is usually the width of maximum potential damage. Figure 9 depicts the zones of potential damage if the previous paths were to occur and, therefore, the insurance company can set the property insurance rates accordingly. Overlaying damage produced by flooding would also be necessary to assess the correct insurance premium amounts. Coastal areas, where most of the population resides, have increased risk for experiencing such damage. Figure 10 shows areas in dark grey where the Inundation depth could be greater than 9 feet. Finally, by combining the information of both types of damages, it is possible to identify the policy holders living in these areas and are at a higher risk. Although premium amounts are affected by many factors, risk from location accounts for the highest percentage. By reviewing the data on these current policy holders, the average monthly premium is $262.09. The company sees this amount as not sufficient to cover the costs and decides to add a correction factor to offset the company's claims risk.

This case study summarizes the application of the three types if analytics reviewed initially. The visualization of the TIV by property falls under the descriptive analytics: where are my customers and what is their total insured value? The process of adding the hurricane paths and storm surge inundation areas can be categorized as predictive analytics: what would be the damage cost in the most likely storm scenario? Finally, the analysis of the current monthly premiums and the estimate of future increases to offset the cost of claims demands action and, therefore, fall under the prescriptive analytics classification.

Figure 9. Buffered Hurricane Paths (Source: ESRI, 2013)

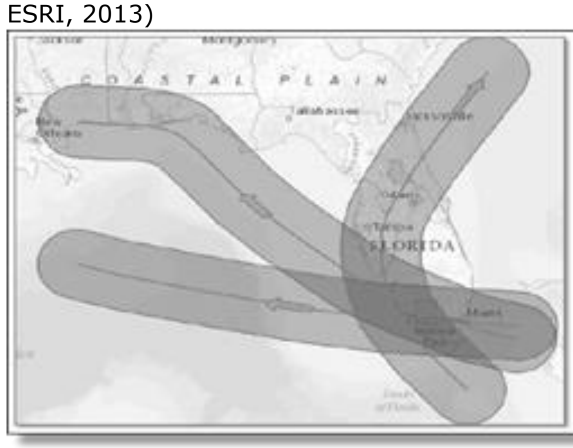


Figure 10. Damage from Flooding Overlay (Source: ESRI, 2013)

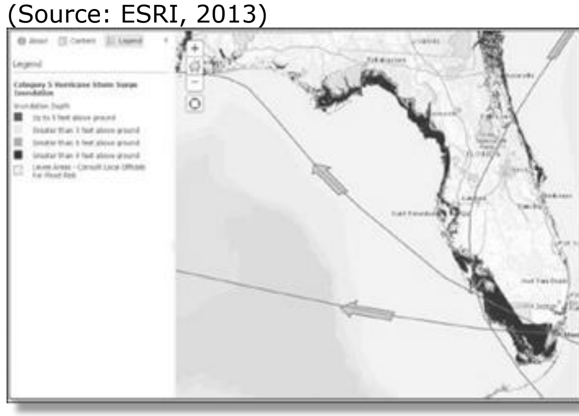


Figure 11. Policy Holders in the Highest Risk Zone (Source: ESRI, 2013)

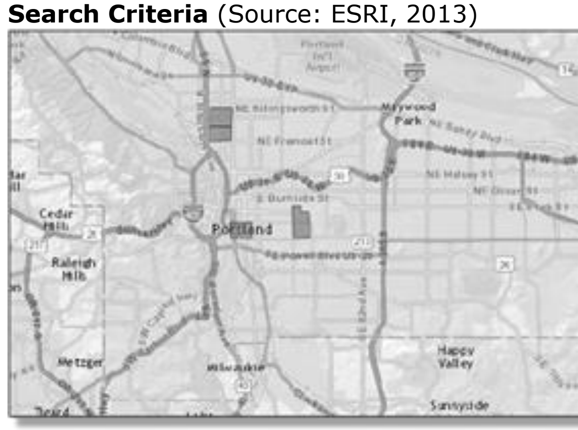
Case Study 2: Finding Potential Markets

A fairly new small business, which promotes commuting to work by bicycle, has created a new product for bicycle commuters: glasses with a display that provides real time information without the need of a smart phone (ESRI, 2015). The device includes GPS, providing commuters with information such as speed, direction, distance, alerts on weather conditions, traffic updates, special offers for coffee, etc. This company would want to market such product in an area with a large number of bike commuters, of ages between 25 and 40, with a moderate disposable income ($25,000 to $35,000). In order to identify the places where large number of potential customers are located, the following data would be needed: demographic data (population, age, and income level), information on cities with high number of bicycle commuters, and area business types. The goal is to find the best neighborhoods to market the product. Figure 12 in the Appendix shows the map of all states ranked by the percentage of people who bicycle to work. The state that has the highest percentage is the District of Columbia, however, the total state's population is small comparing to the rest. The second with a high percentage is Oregon: 2.34%. By zooming in to Oregon, one can analyze the percentages by county (Figure 13 in Appendix). The counties with darkest red color are potential areas. Not only the highest percentages are useful, also the actual total number of commuters are important: Multnomah County has 18,000 bicycle commuters.

In a further refined analysis, demographic data is overlaid on the same map. The demographic layer is set based on the three filter ranges: ages between 25 and 40, income between $25,000 and $34,999, and bicycle to work percentage higher than 18%. Figure 14 depicts the results: four census tracts within Portland downtown area (Multnomah County). Finally, a deeper search for population characteristics in these census tracts could refine the product features, price, placement, and promotion. Tapestry segmentation information could guide in answering questions such as how to price the product? Should it target people with high incomes and include more features or make it affordable to a wider group of people with basic features? Based on where they get their news and shop, the company could plan the placement of the product for sale (bike shops, online, retail stores) and promotion strategies.

Since in this case study all questions focus on future events, all steps involved represent predictive and prescriptive analytics. Initially, the company predicted where the potential customers are located, and finally specific courses of action for determining price and promotions derive from a further analysis of tapestry segmentation characteristics.

Figure 14. Census Tracts that Meet the Search Criteria (Source: ESRI, 2013)



2. CONCLUSION

Many authors agree on the advantage of using analytics not only to facilitate the understanding of facts and improve the decision making process, but also to add business value. Depending on the kind of analytics applied, a different degree of value would be added.

Geospatial analytics is not only the representation of data in a map, it involves a deeper integration of location data with businesses strategies and processes. Although a vast amount of literature advocates the use of geospatial analytics, the detailed case studies review presented in this paper intends to go beyond the mere outlining of benefits. Hopefully this paper presents a compelling argument in favor of the implementation of geospatial analysis in any business area or industry to gain business value.

Although 35% of the companies participating in the 2015 TDWI survey currently use geospatial analysis in their day-to-day analytics process, another 30% plans to incorporate it in the upcoming years. This evidences interest in this emerging technology, however, obstacles such as cost and the small availability of options within the existing BI solutions in the market seem to preclude a further expansion.

Currently, with the advent of big data and the internet of things (IoT), additional challenges are presented ahead. These new sources of data; characterized by the large volume, velocity, and variety; pose an additional burden on the current BI systems and data management techniques. Watson states in his tutorial on big data: "a key to deriving value from big data is the use of analytics" (Watson, 2012). Halper views "geospatial and IoT as one of the next evolutions in big data" (Halper, 2015). There is no doubt, there is business value rendered when applying geospatial data either in traditional or big data analytics.

3. REFERENCES

CAGIS. 2015. CAGIS Portal Maps. *CAGIS ArcGIS Online*.

ESRI. 2015. The Location Advantage. *ESRI Online Training.*

ESRI. 2013. Put the Where in your Enterprise: Deliver Better Outcomes with Esri Location Analytics. *ESRI Online.*

Fernandes, Louella. 2006. IT Analysis – Putting Business Intelligence on the Map. *Quocirca Ltd. online*.

Forbes Insights. 2015. The Eureka Moment: Location Intelligence and Competitive Insight. *Forbes Online*.

Halper, Fern. 2015. The Changing Face of Geospatial Analytics. *TDWI Research Online*.

Halper, Fern. 2013. Seven Use Cases for Geospatial Analytics. *TDWI Research Online*.

Halper, Fern. 2013. Three Ways to Use Geospatial Data in Analytics. *TDWI Blog*.

Hempfield, Clarence. 2014. A Marriage of Necessities: Why Integrating Spatial Analysis with Major BI Platforms is Essential. *Business Intelligence Journal*, Vol. 19, No. 4: 28-34.

IBM. 2015. Geospatial Analytics with IBM SPSS Modeler. *IBM Online*.

IBM. 2015. Geospatial Analytics: Enrich your Predictions with Geographic and Spatial Insights. *IBM Online*.

Kart, Lisa. 2015. How to Get Started With Prescriptive Analytics. *Gartner Online*.

Kent, Steve. 2010. Integrating Geospatial Analysis. *Business Intelligence Journal*, Vol. 15, No. 3: 14-18.

Lamb, James. 2018. Geospatial Analytics Market Size, Share, Report, Analysis, Trends & Forecast to 2026. Accessed from: http://www.strategymrc.com/report/geospatial-analytics-market.

Loshin, David. 2012. Using Location Information for Geospatial Analytics. *TDWI Research Online*.

Russom, Philip, David Stodder, and Fern Halper. 2015. Best Practices Report: Emerging Technologies for Business Intelligence, Analytics, and Warehousing. *TDWI Research Online*.

Sallam, Rita, Bill Hostmann, Kurt Schlegel, Joao Tapadinhas, Josh Parenteau, and Thomas W.

Oestreich. 2015. Magic Quadrant for Business Intelligence and Analytics Platforms. *Gartner Online*.

SAS Institute Inc. 2014. Data Visualization Techniques: From Basics to Big Data with SAS Visual Analytics. *Datawatch Online*.

Watson, J. Hugh. 2012. Big Data: Concepts, Technologies, and Applications. *Teradata University Network Online*.

Appendix

Figure 5. Life Expectancy Findings Table (Source: CAGIS, 2015)

Rank	Neighborhood Name	Life Expectancy
1	Mount Lookout - Columbia Tusc	87.8
2	North Avondale - Paddock Hills	87.1
3	Mount Adams	86.4
4	Mount Lookout	85.9
5	Hyde Park	83.2
6	Madisonville	83.1
7	Mount Washington	82.9
8	Mount Airy	82.3
9	California	81.2
10	College Hill	81.1
11	West Price Hill	80.4
12	Clifton	79.3
13	Oakley	79.2
14	Pleasant Ridge	79.1
15	Villages at Roll Hill	77.3
16	East Walnut Hills	77.2
17	Central Business District	76.9
18	Kennedy Heights	76.7
19	Riverside - Sayler Park	76.5
20	Westwood	76.1
21	Spring Grove Village	76.1
22	Roselawn	75.1
23	Hartwell	74.7
24	Mount Auburn	74.5
25	Winton Hills	74.5
26	Sayler Park	74.4

Rank	Neighborhood Name	Life Expectancy
27	Fairview - Clifton Heights	74.1
28	University Heights	74.0
29	Evanston - East Walnut Hills	73.5
30	Bond Hill	73.0
31	Carthage	72.9
32	Linwood	72.7
33	East End	72.6
34	Corryville	72.6
35	Northside	72.6
36	North Fairmount-English Woods	72.1
37	Evanston	71.8
38	South Cumminsville - Millvale	71.2
39	East Price Hill	70.8
40	Over the Rhine	70.6
41	West End	69.8
42	Walnut Hills	69.6
43	Avondale	68.2
44	Camp Washington	67.8
45	Sedamsville - Riverside	67.0
46	Lower Price Hill	66.8
47	South Fairmount	66.4

Figure 6. Chart Depicting the Life Expectancy in Cincinnati (Source: CAGIS, 2015)

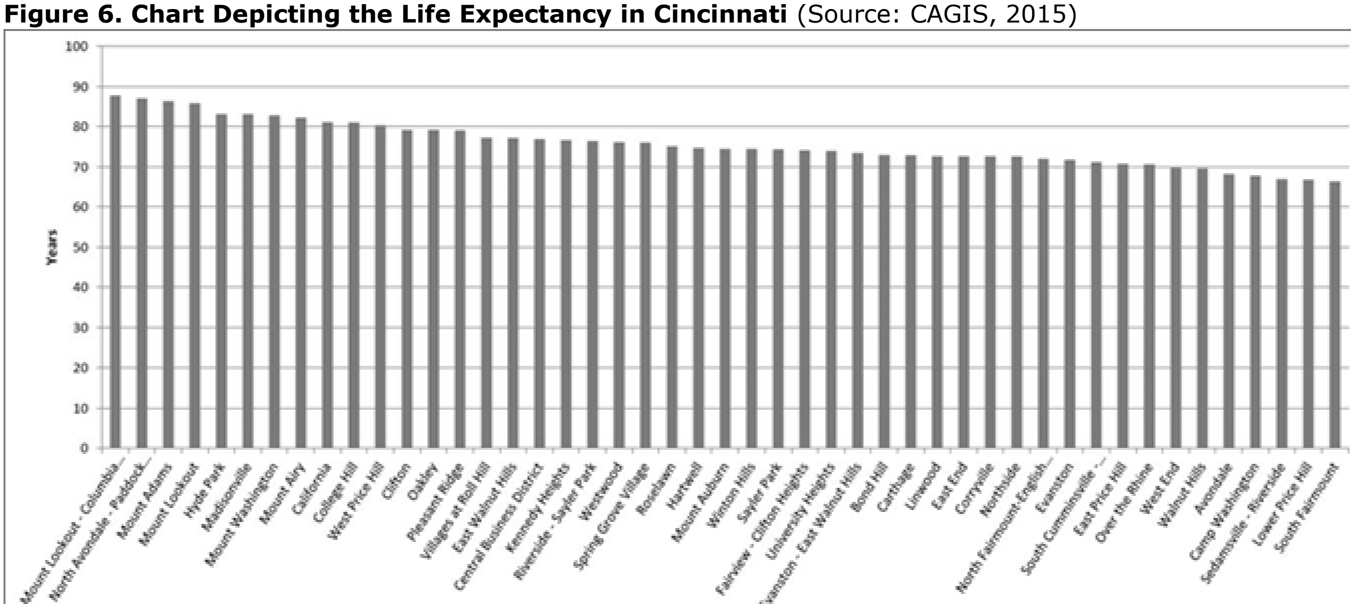
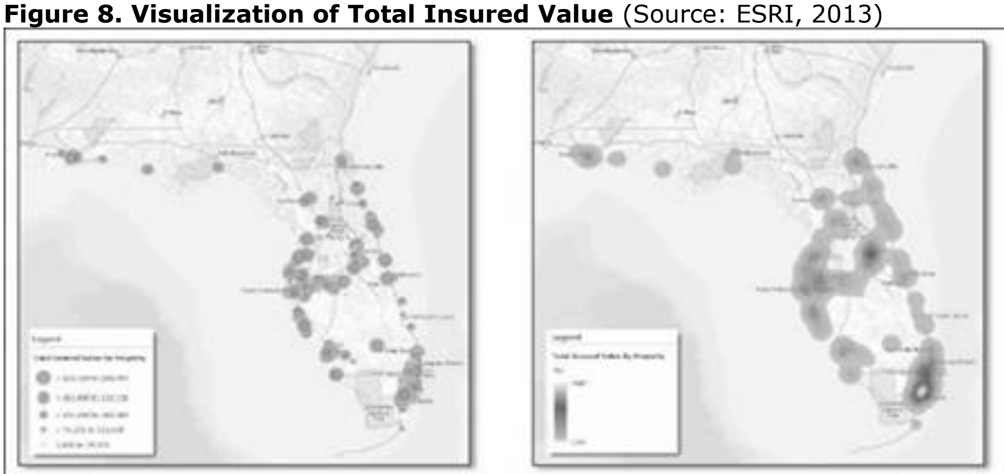


Figure 7. GIS Map of Life Expectancy in Cincinnati (Source: CAGIS, 2015)

Figure 8. Visualization of Total Insured Value (Source: ESRI, 2013)



Circles representing TIV. *Heat map representing TIV.*

Figure 12. Bicycle Commuters to Work by State (Source: ESRI, 2013)

Figure 13. Bicycle Commuters to Work by County (Source: ESRI, 2013)