In this issue:

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

Editors

Scott Hunsinger
Senior Editor
Appalachian State University

Thomas Janicki
Publisher
University of North Carolina Wilmington

2020 JISAR Editorial Board

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)
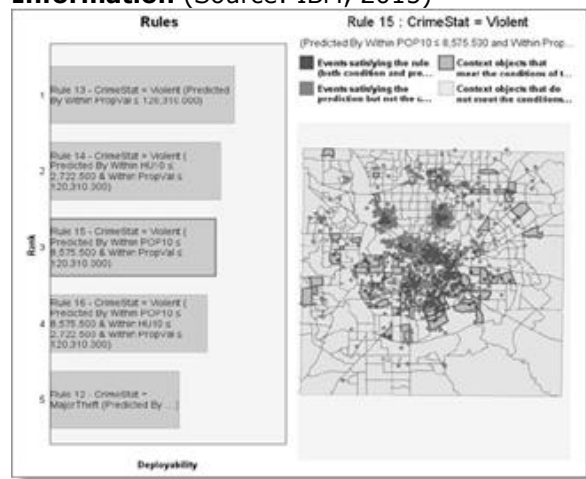
Prescriptive analytics, as well as predictive analytics, focuses on the future and more

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)
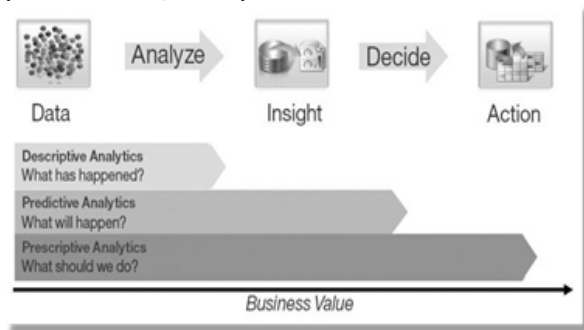
Geospatial Analytics

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)
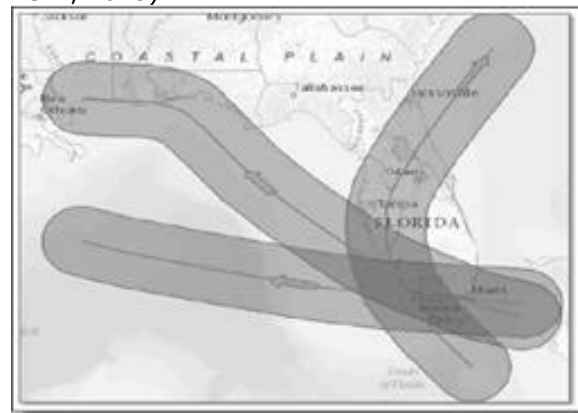
Figure 10. Damage from Flooding Overlay (Source: ESRI, 2013)
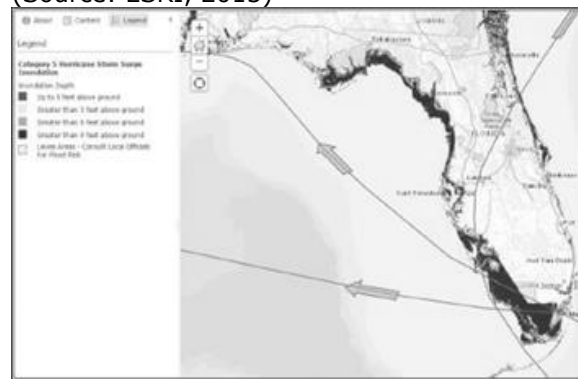
Figure 11. Policy Holders in the Highest Risk Zone (Source: ESRI, 2013)

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	Rank	Neighborhood Name	Life Expectancy
1	Mount Lookout - Columbia Tusc	87.8	27	Fairview - Clifton Heights	74.1
2	North Avondale - Paddock Hills	87.1	28	University Heights	74.0
3	Mount Adams	86.4	29	Evanston - East Walnut Hills	73.5
4	Mount Lookout	85.9	30	Bond Hill	73.0
5	Hyde Park	83.2	31	Carthage	72.9
6	Madisonville	83.1	32	Linwood	72.7
7	Mount Washington	82.9	33	East End	72.6
8	Mount Airy	82.3	34	Corryville	72.6
9	California	81.2	35	Northside	72.6
10	College Hill	81.1	36	North Fairmount-English Woods	72.1
11	West Price Hill	80.4	37	Evanston	71.8
12	Clifton	79.3	38	South Cumminsville - Millvale	71.2
13	Oakley	79.2	39	East Price Hill	70.8
14	Pleasant Ridge	79.1	40	Over the Rhine	70.6
15	Villages at Roll Hill	77.3	41	West End	69.8
16	East Walnut Hills	77.2	42	Walnut Hills	69.6
17	Central Business District	76.9	43	Avondale	68.2
18	Kennedy Heights	76.7	44	Camp Washington	67.8
19	Riverside - Sayler Park	76.5	45	Sedamsville - Riverside	67.0
20	Westwood	76.1	46	Lower Price Hill	66.8
21	Spring Grove Village	76.1	47	South Fairmount	66.4
22	Roselawn	75.1			
23	Hartwell	74.7			
24	Mount Auburn	74.5			
25	Winton Hills	74.5			
26	Sayler Park	74.4			

Figure 6. Chart Depicting the Life Expectancy in Cincinnati (Source: CAGIS, 2015)
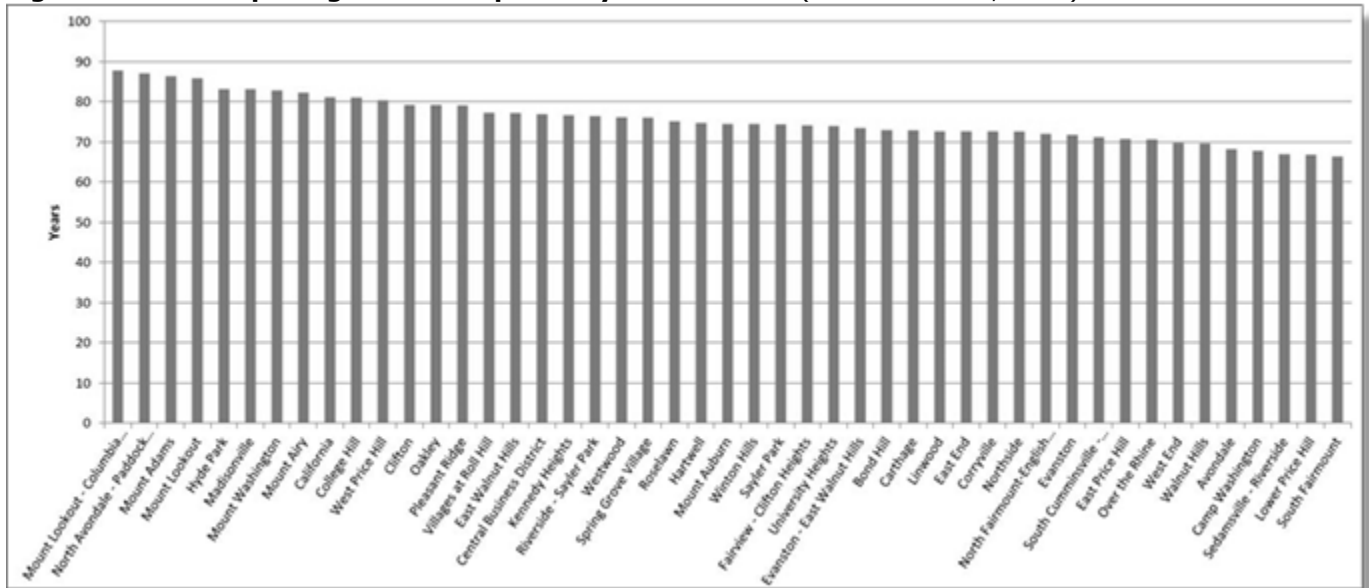
Figure 7. GIS Map of Life Expectancy in Cincinnati (Source: CAGIS, 2015)

Figure 8. Visualization of Total Insured Value (Source: ESRI, 2013)
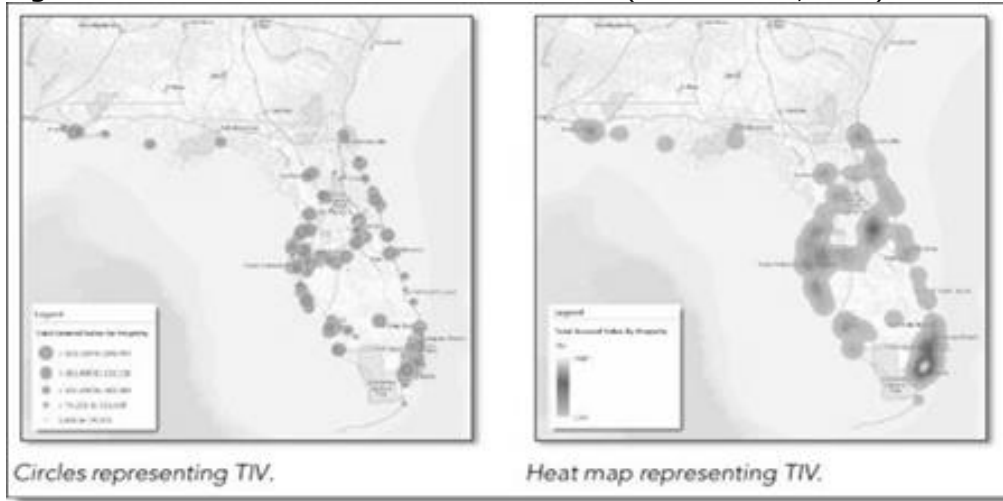
Figure 12. Bicycle Commuters to Work by State (Source: ESRI, 2013)

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

Examining Organizational Security Governance (OSG) Objectives: How strategic planning for Security is undertaken at ABC Corporation?

Sushma Mishra
mishra@rmu.edu
Computer Information Systems Department
Robert Morris University
Moon Township, PA 15108, USA

Abstract

Organizational security governance (OSG) objectives provide a strategic direction to the organization's comprehensive security planning. It is essential that security planning has overarching objectives such that the proposed controls and measures are contextually grounded. Mishra (2015) proposed six fundamental and seventeen means objectives, theoretically and empirically developed, to provide a strategic basis of an organization's security program. The purpose of this study is to examine the proposed six fundamental OSG objectives in an organizational context and assess how it guides the security efforts. A case study was performed to understand the use of the six fundamental objectives in the overall security program. The results suggest that the OSG objectives indeed provide a strategic basis for developing security measures of the organization. Implications are drawn and contributions listed.

1. INTRODUCTION

In today's environment, cyber-attacks are on the rise, and security investment has grown manifolds for strengthening preventive measures. However, organizations are struggling for strategies to address such attacks (Ahmad et al., 2014). The biggest problem with such reactive security planning is that organizational security governance (OSG) objectives are missing and there is a lack of potential implementation strategies for such objectives. It is essential to understand how OSG influences strategic decision-making in information security and ensures that investments in security are not wasted (Tan et al., 2017). There is little guidance, in the research literature, on how organizations develop OSG and implement them.

Mishra (2015) proposed six fundamental and seventeen OSG objectives, grounded in theory and data, which could provide comprehensive security strategic planning and measures to ensure more secure organizations. Fundamental objectives are the core of the OSG program, whereas means objectives support the fundamental objectives. For a more detailed understanding of these objectives, refer to Mishra (2015). The fundamental objectives are essential to achieve string OSG. All means objectives help in achieving one or more fundamental objective for better OSG practices. This study focusses just on the fundamental objectives as these form the core of the OSG program. The six proposed fundamental objectives are F1-Ensure Corporate Controls Strategy, F2-Encourage a Controls- Conscious Culture, F3-Establish Clarity in Policies and Procedures, F4-Maximize Regulatory Compliance, F5-Ensure Continuous Improvements in controls and F6-Enable Responsibility and Accountability in Roles (table

1). These objectives, however, have not been examined in an organizational setting. In this research, a case study is performed to examine and understand the effectiveness of six fundamental objectives proposed by the Mishra (2015). Insert table 1 here

The guiding research questions are:

RQ1: How do the six fundamental OSG objectives guide strategic planning at ABC Corporation?

RQ2: What are the measures used by ABC to enhance its strategic planning using these objectives?

The rest of the paper is organized as follows. The next section presents the organizational context and discusses each fundamental OSG objective in the context of ABC Corporation. Following this section, a discussion is presented about the objectives and its implications for OSG literature. Contributions are listed, and the future direction of research suggested.

2. CASE STUDY

ABC Corporation is the IT department, of a state agency, responsible for all information technology needs of a large city in the southeast USA. The mission statement of the organization explicitly highlight the guiding values of the company, and one of the core missions is to provide security to customer data at all times. In the process or organizing its IT functionality, the organization has identified "managing organizational security governance" as its strategic area of improvement. The management, led by the CIO, believes that strategic planning for security controls based on OSG objectives would improve the efficiency and effectiveness of their core business. ABC Corporation is the key to all information technology changes in the city it serves. As IT evolves, the management aspires to provide more ways of better service delivery and create operational efficiency along the way. The CIO believes that this organization would play an essential part in the process of transformation of ways in which the state agency conducts business. The strategic role of this organization, for the entire city, relies heavily on management's investment in OSG initiatives. The CIO of the organization has initiated several task forces to implement changes to manage the IT architecture. The IT infrastructure is based on the city's business needs and not on the latest technology trends.

The organizational structure includes the CIO as the head of the agency. Five managers directly report to the CIO. There is an application division that controls all in house development. The database systems and the enterprise system are under infrastructure services that are managed by a manager. A training division provides administrative support functions. Project management division looks at the feasibility of all proposed projects and makes decisions about the ones that will get funded. There are well above 100 employees, multiple consultants, and many open positions for hire. The organization targets improvements based on the specific needs of different agencies. These improvements are based on joint maps created with the IT organization and the agency. All the policies, procedures, and controls are transparent as the state agency is under the purview of public scrutiny and audit and compliance requirements on an ongoing basis.

For this study, ten interviews were conducted at ABC Corporation. Participants included top management, middle management, support staff, auditors, and developers. There were secondary sources of data examined, such as security policies, procedures, project reposts, training documents, and audit manuals.

3. RESULTS & DISCUSSION

F1-Regulatory compliance at ABC
Regulatory compliance ensures that the organization meets all the legal and mandatory requirements about security and internal controls. This objective entails formalizing the process of compliance in the organization and promotes the development of controls in accordance with legislation. The objective entails following the regulations in their entirety and using the legislation as a catalyst for the improvement of security governance.

ABC, as a state agency, has HIPPA and e-discovery as its principal regulations to comply with. The agency has a compliance audit, both from internal as well as external auditors. The culture in the organization is such that transparency about processes and availability of information are considered of paramount importance. The CIO is aware of people's right to ask for different types of information about the agency and the use of taxpayer's dollars in operations. By state law, most of the information about the agencies' current and future plans are accessible through the website. The common perception, of employees, of the regulations and the compliance efforts in the organization is that

of a "necessary evil." The middle-level managers and the line staff consider compliance as the "right thing" to do but not necessarily helpful. The sentiment of frustration is understandable, given the mammoth preparations required for compliance. Compliance with laws such as SOX is costly (Bennett & Cancilla, 2005). It needs managerial as well as technical support to create an infrastructure in organizations to meet the demands of this law. Some of the specialized areas that need particular concern for compliance purposes are: data management issues (Volonino et. al., 2004; Farris, 2004, Yugay and Klimchenko, 2004), security of data and system, choice of software development methodologies that could incorporate the compliance issues in its lifecycle, robust documentation for external auditing purposes, versioning and auditibility of electronic record needs and file systems in use (Peterson and Burns, 2005).

The internal IT audit director for the agency considers the regulations as helpful in providing momentum to security and internal controls operations in the organization. Myler and Broadbent (2006) argue that evaluation of compliance with the policies and procedures in an organization and regular follow up of the recommendations are essential. The evaluation process helps in estimating the effectiveness and possible shortcomings of the controls process. Delineating audit controls and tools to determine areas for improvement (Myler and Broadbent, 2006) is what the IT audit director for the City believes in.

The chief agency head did not have a favorable opinion about the regulations, though. As the applications development manager commented:

> "In my personal opinion, compliance is reactive, not proactive. You look at SOX. Enron collapsed, and so many people were ruined or hurt, and then SOX came. So compliance is a vehicle, the way compliance operates today, I don't think that an organization should say ok…we rely on compliance as a mechanism for developing our internal controls. If you do that you are going to be in bad shape".

The general perception of the management about compliance is that it drives the security governance efforts backward. The regulations enforce activities that should already have been a part of the governance program in the first place. This perception is consistent with the majority of research in this area. One of the

most significant managerial issues that regulations imply is for IT governance purposes (Fox, 2004). Effective IT governance would require the management to plan for preparedness for quarterly reporting, security policies, cost management for compliance, and preparation for external audit. These measures need planning and effective internal control assessment (Chin and Mishra, 2013). The administration believes that the preparedness should be there, to begin with, and not inserted as an afterthought while preparing to be compliant.

Compliance acts as a driver in getting all the resources that are required for the agency. The security officers shared how in the name of compliance, they order software, get management's attention, and other required resources. The responsibility of the regulatory compliance efforts for the city does not lie with ABC but the larger agency. The lack of authority for regulatory compliance explains a lot of discontentment with the use of regulations in the organization. Officials at ABC comply with the requests of auditors and supply all the paperwork required. The organization plays a passive role in the City's compliance plan.

Overall, it did appear that regulatory compliance is vital for the agency. Since the prime responsibility of being compliant did not lie with this agency; managers in the organization were candid about it. ABC used compliance as a means to get things from the City, which they would never get otherwise. Also, the organization is in the process of developing new policies and controls. It remains to be seen how these new controls are implemented and assessed. To sum up, compliance is important to ABC but not in the right spirit of the legislation. A summary of regulatory compliance at ABC is presented in the table below (table 2).

F2-Ensuring continuous improvements in controls at ABC
Instituting continuous improvements in controls implementation process has been identified as a fundamental objective for maximizing security governance. The control implementation process should be iterative, continuous, and adaptive. Effective implementation of controls calls for putting the right controls in the right place at the right time, and this can only be achieved through flexible implementation practices.

At ABC, the management identifies the need for a constant reevaluation of controls under

changing business conditions. Regular revalidation of the controls is important, as the Chief Security Officer shared:

> "You have to keep up with it…It's not what you are getting over with…you have to constantly keep up with it…we do have some machines and software that are from over ten years old…but you have to keep up with it….what else can we do..we need constant reevaluation as controls implementation is an evolving process".

This Chief Information Officer at the firm has a similar vision of regularly testing and updating the control structure. The organization provides training and education to the security staff about the changing needs of controls and policies. The security officers are encouraged to attend conferences and seminars in the relevant area to keep abreast with the upcoming trends and technologies in the security area. As one of the security officer said:

> "I am a firm believer that you can put whatever you want in place, but if the end user doesn't own it up, it is not going to work. I have been in seminars where I dealt with fortune 500 companies, people who are making billions of dollars a year as revenue, and they still have the same problem. You know those guys have everything, they have done everything, but it [control implementation] needs to be a constantly evolving process. They have to learn and then reeducate because things change."

It was apparent from the observations at ABC that management understands the importance of the controls implementation process. There were frequent meetings and seminars about security controls and discussions on how controls should be used to overcome common security breaches. However, a sense of disconnect in the attitude of the managers and the operational people, about continuous changes in controls, was felt. The knowledge about the benefits from revalidation of controls is concentrated more on the management side than on the operational side of the organization. The business folks considered control implementation as a technical requirement for the organization and distanced themselves from it. The perception in the non-security staff, working in development or other IT related areas, is that control implementation is primarily the work of security staff. The majority of operational people believe that the security staff

should be responsible for the success or failure of the controls.

The enabling value of security controls has to be clearly articulated. Benefits of security governance should be linked with business objectives so that the stakeholders see the positive impact of security on attaining profits, productivity, and growth. Security governance can help in avoiding negligence and enhance strategic business goals hence acting as a motivator for top management (Wright, 2001). It is crucial to ensure that security controls and security management practices of the organization are regularly reviewed. Such reviews could lead to finding misconfigurations in the systems and identify areas where security protection is such that a single failure could result in prolonged exposures (Wilson, 2007).

F3-Responsibility and accountability structures at ABC

Responsibility and accountability structures ensure that roles are defined in a way that appropriate responsibilities are shared, and stakeholders are held accountable for their actions. The objective prescribes that job descriptions should be not changed abruptly, clear organizational responsibility for compliance should be defined; individuals should be made responsible for appropriate accesses, and transparency about the accountability should be encouraged.

The management at ABC completely identifies with the criticality of having clear responsibility and accountability structure for information systems security governance. The Chief Information Officer said:

> "If you are talking about the outcome of the controls, then to me, its management. The idea of having a documented hierarchy, especially around data, is a must. If you think about it; we publish corporate organizational charts all the time. We should have a controls organizational chart which says, okay, if you are at this level, this is what you get [controls]."

The CIO believes in the concept of having a "controls chart," which is similar to the organization chart. The controls chart clearly defines the responsibilities of the members regarding security controls. The controls chart is like adding control responsibilities to the organization chart. It helps in documenting the requirements for a role in owning up to the responsibility of controls. As we go up in the

controls chart, roles become more crucial for security governance; the individual higher up should have more controls and accountability associated with their work. Research in security governance suggests that increased awareness and individual accountability can significantly affect how security practices are implemented in an organization (Mellor and Noyes, 2006).

People higher up in hierarchy have greater accessibility to sensitive data and have a higher probability of creating vulnerability in the system. Mellor and Noyes (2006) found that adding personal accountability into roles helps the cause of security governance. The concept of controls chart is not implemented yet at ABC but would be helpful for security governance purposes. As explained by the CIO, it is crucial to understand what is it that we want to protect from a management point of view. If there is clarity in responsibilities and roles, better controls can be associated with the position and the individuals. For example, if the human resource people have a high level of access to crucial personal identifiable data of personnel in the organization, there should be stringent controls for people in this department. As suggested by the CIO, such managers should be audited for their access pattern every quarter to ensure that the managers are doing what they are supposed to do, and security is not being compromised. Given the nature of the sensitive information that human resources people have access to, it makes sense to have better protection and accountability for such people. Research literature suggests that top management should be proactive about responsibility assignment to roles. Myler and Broadbent (2006) argue that corporate boards that undertake the challenge of plugging IT oversights show that they understand the scope of their corporate accountability and responsibility, and are proactive in their leadership duties. If organizations do not ensure that all employees understand their information security roles and responsibilities, it may become challenging to protect the confidentiality, integrity, and availability of information assets (NIST Special Publication 800-16, 1998).

ABC has access to crucial data about the taxpayers in the City. The department has access to DMV data, readings for gas, water and electricity consumption, property details, and tax details about the residents. One of the duties of the department is to ensure that the meter reading for the household utilities is performed correctly as and when required. This operation,

if not completed successfully, could present a severe threat to the integrity of the data recorded. As mentioned by the end user services manager:

> "I think the accountability piece is required. How do they control, say even a meter reading application? How do we ensure that every meter gets read every morning? You have meters that haven't been read, and there has been no consumption on that meter for over a year, and the service is still on, then there is a problem. So put controls and make someone accountable, that's how you guarantee that every meter is being read and the consumption of gas and water is recorded."

Reading a utility meter requires that there is appropriate segregation of duties defined in the organization else the security of the data could be compromised. It is essential to separate developers who make the application from people who read the meters and record the consumption by providing logical access to the groups. Else, it is possible for the developer to change the readings, through the application, for themselves or friends or whoever they deem appropriate.

At ABC, management is concerned about assigning appropriate responsibilities and accountability to users of the systems (see table 4). But it seemed that there is a lack of clarity about roles and responsibilities on many fronts. For example, when discussing the regulatory compliance issues in the organization, there seems to be confusion about who in the agency was responsible for the meeting compliance deadlines. People at ABC meet auditors' request for submitting the required documents.

F4-Corporate control strategy at ABC
This objective suggests establishing a corporate risks management plan and developing controls guidelines using consensus. Controls should be viewed as a cost of doing business. Security controls should be a non-negotiable budget item, and adequate planning for the governance initiatives should be ensured.

The management at ABC believes that for long term strategic planning in the organization, it is essential to have a control strategy. A clear vision about the security governance and each department's own controls plan along with an enterprise level risk assessment plan would go a long way. An information security risk assessment is the staged process by which an

organization's information assets are valued. Here, the vulnerabilities and threats are identified so that they then guide the implementation and monitoring of control strategies and measures (Whitman and Mattord, 2005).

At ABC, there is a lack of agreement between stakeholders on what should be put and how should the controls be deployed and monitored. This disagreement is a direct result of a fundamental lack of planning and understanding about what are the assets and what is that needs to be controlled. A controls strategy can provide a broad vision for the organization in this regard. As shared by the security manager:

> "People should try to at first establish and see what the controls are. That's reflected in your requirements to some degree. People need to know what they want to control. You have to know what you want to control, and the problem is that you don't know what you want to control."

The necessary process of controls development approach needs long term planning and undying commitment on the part of the management. The CIO believes that a strategy about controls needs to be established such that all pieces of governance program come together.

Observations at ABC suggest that a "bottom-up" approach of developing security initiatives would not work in this organization. The operational level management does not have a holistic picture of the role of controls in achieving overall security. The strategic inputs about security governance should flow from the top management to the entire organization. The lack of a control strategy would cause the controls to be laid out without risks analysis and policies which could be expensive and detrimental. With a top-down approach to management, a more appropriate strategy in the shape of long-term policies, efficient procedures, and technical safeguards could be developed (May 2005).

An interesting observation was noted that there is inadequate planning about protecting the human assets in case of an emergency such as fire or flood. Without a sound strategy, efforts will be wasted. Therefore, a structured methodology for developing a strategy will increase the likelihood of success of corporate initiatives (Mishra and Dhillon, 2006). It is a serious issue: what is the strategy about protecting employees and equipment in case of

emergency? The management at ABC seems distressed about the fact that the City does not consider this issue important enough to discuss at high-level meetings. The state of affairs at ABC does substantiate our call for a controls strategy which could plan about things such as this at the corporate level.

There is a growing awareness of the need for such a strategy (Shedden et al., 2006). Information security should be integrated into an organization's overall management plan (Lane, 1985). Firms have to integrate IT strategies with organizational strategies to attain business objectives (Lainhart IV, 2001). In the case of ABC, the management could have an oversight committee that sets an appropriate strategy for IT governance endeavors (Myler and Broadbent, 2006), especially about security events. A summary of the control strategy initiatives at ABC is provided in table 4 below.

F5-Control conscious culture at ABC
A control culture ensures an environment where individuals 'watch out' for each other. This fundamental objective emphasizes the importance of a control culture that creates and sustains connections among various security efforts such as policies, processes, and norms. A "prevention mentality" promoted by the control culture of the organization, helps in minimizing the friction among groups over security issues. It is essential to establish standard codes of conduct for the employees in carrying out their security responsibilities.

The CIO of the organization believes in establishing a culture that needs to consider all the information that ABC has and protects it as something personal for the employees. The CIO explained:

> "I think you need to have a clear core value; a clear company recognized or accepted perspective, the role of having those controls. For example, in my mind, I think you should treat everything, every data you handle like it's your information. Would you leave your wallet out in the middle of the street, on the bench when you go to get a coffee? What type of care would you take if it's yours? That is the kind of care you need to take."

Management espousing similar values as it claims should ultimately lead to the *shared tacit assumptions* of employees becoming aligned with these *espoused values* of the organization, thus progressing towards information security

obedient culture (von Solms, 2006). The management realizes that it is a long and tedious process before a control culture is established. As the chief security officer enunciated:

> "Establishing the concept [the importance of controls] takes much time and commitment, to do that you want to bring that culture and it takes time, and it is just a matter of time and that it will come after you do it for long."

The management feels that establishing a control culture would help the policies and procedures in being appropriately followed, and the management become more involved in the security governance process. The tacit knowledge of information security practices and procedures and the resulting behavior guide the day-to-day activities of the employees in the organization. As a consequence, information security practices and procedures should become part of the corporate culture of an organization (Thomson and von Solms, 2005). Observations at various meetings and informal conversations with the employees suggest that the organization had a control culture where people treat the information as they would treat their property. Maybe it is the beginning of the long and tedious process of establishing a controls consciousness of this nature because the leadership at the organization did seem determined to drive the organization towards control culture. The controls culture is crucial for security governance as it can act as a robust, underlying set of forces that establishes individual and group behavior within an organization. Ideally, corporate culture should incorporate information security controls into the daily routines and implicit practice of employees (Thomson and von Solms, 2006). If the beliefs and attitudes are addressed by the management, it leads to changed actions and behaviors of the employees and synchronizes it with the overall corporate security culture in the organization (Thomson and von Solms, 2008). A summary is presented in table 6.

F6-Clarity in policies and controls at ABC

Policies should be fair, visible, and easily accessible to all in the organization. The clarity in policies communicates management commitment to security governance.

Policies and procedures are organizational laws that determine acceptable and unacceptable conduct within the context of corporate culture (Whitman, 2003). It is a means to communicate management's commitment to security governance efforts (Myler and Broadbent, 2006). ABC emphasizes establishing clarity of policies and controls. The conventional norm is to explain the policies and procedures frequently so that that it makes an impression on the user and stays with them eventually. Usually, the most common reason why employees make mistakes about controls in the organization is the lack of understanding as to what needs to be done. Research suggests that good policies can protect vulnerabilities (Lapke and Dhillon, 2008). Better policies lead to deterrence as policies give the employees responsibility and accountability in the job (Maynard and Ruighaver, 2007). The security team feels that people never come up and ask about policies or controls unless they are in trouble. The management at ABC explains the purpose and scope of the controls proactively before the employees get into trouble.

Research literature in security policy domain argues for revisiting the policies periodically. For instance, it is becoming a huge problem to prevent employees from wasting their time on browsing the Internet during office hours. Policies about personal use of computers during office hours need to be clearly defined. Limited Internet use or unlicensed software usage should be discouraged (Schauer, 2001). Maynard and Ruighaver (2007) maintain than beside the iterative nature, security policies need quality verification periodically. This assessment needs to be carefully managed to ensure a balanced approach and make sure that stakeholders have adequate skills and training to assess quality. The management also believes that policies should be developed as a continuous process so that changing business needs are reflected.

The taxpayers should be able to access the security policies to have confidence in the city's security measures about protecting their data. Also, the current policies have not been made easily accessible to the employees as well. The lack of clarity creates a potential rift in the minds of people about the policies. As the security staff officer explained:

> "We had regulation and policies established but did people know that? Make all the required things accessible to people. Our policies are so hard to find on our website that I don't know how anyone can ever read them. This is serious".

The management is developing a new set of security policies and procedures. It is planned that the security policies would be made accessible to all the citizens at the web site. A central repository of security policy and control resources would be created on the Intranet which would be available to all Agencies Citywide. The management has planned extensive educational sessions to establish the clarity of new policies. It remains to be seen in the future though that how well these measures play out in creating effective security governance. A summary of how clarity of policies and procedures is being accomplished at ABC is presented in table 7 below.

The case study data suggests that all six fundamental OSG objectives are meaningful to ABC Corporation and are guiding the security program of the organization in many ways. The objectives thus examined are relevant to the organizational context and provide a strategic basis for planning current and future security initiatives. Both research questions are adequately addressed in this section.

4. CONCLUSIONS

This study uniquely contributes to research in information security governance domain. The empirical validation of OSG objectives proposed in the literature provides a meaningful way of developing comprehensive security governance program in organizations. It contributes to practitioners in organizations in terms of delivering overarching OSG objectives for strategic planning of security to prevent breaches. It also provides prescriptive measures that could be used to strengthen security initiatives.

5. REFERENCES

Ahmad, A., Maynard, S., and Park, S. "Information security strategies: towards an organizational multi-strategy perspective," Journal of Intelligent Manufacturing), 2012/07/22 2012b, pp 1-14.

Bennet, V. and Cancilla, B. (2005). IT responses to Sarbanes-Oxley. IBM, Retrieved on 09/30/05,http://www-128.ibm.com/developerworks/rational/library/sep05/cancilla-bennet/index.html

Chin, A. and Mishra, S. (2013). Assessing the Impact of Governmental Regulations on Organizational Competitiveness: An Analysis Using Neo Institutional Theory, *Issues in Information Systems*, Volume 14, Issue 1, pp.286-299, 2013

Farris, G. (2004). Mitigating the Ongoing Sarbanes-Oxley Compliance Process with Technical Enforcement of IT Controls. DM Direct Newsletter, DMReview.com, Retrieved on 07/17/06 http://www.dmreview.com/article_sub.cfm?articleId=1014858

Fox, C. (2004). Sarbanes-Oxley- Considerations for a Framework for IT Financial Reporting Controls. Information Systems Control Journal, 1

Lapke, M and Dhillon, Gaurpreet, "Power Relationships in Information Systems Security Policy Formulation and Implementation" (2008). *ECIS 2008 Proceedings*. 119. https://aisel.aisnet.org/ecis2008/119

Lane, V.P. Security of Computer Based Information Systems Macmillan, London, 1985.

Lainhart IV, J. "An IT assurance framework for the future " Ohio CPA Journal (60:1) 2001, pp 19- 23.

Mellor, M., and Noyes, D. "Awareness and accountability in information security training " 6th Annual Security conference The Information Institute, USA Las Vegas, 2007.

Mishra, S. (2015) "Organizational objectives for information security governance: a value focused assessment", Information & Computer Security, Vol. 23 Issue: 2, pp.122-144, https://doi.org/10.1108/ICS-02-2014-0016

Mishra, S. and Dhillion G (2006), "Information Systems Security Governance Research: A Behavioral Perspective", 9th Annual NYS Cyber Security Conference and Annual Symposium on Information Assurance, June 14-15 Albany, NY

Myler, E. and Broadbent, G. "ISO 17799 : Standard for security " The Information Management Journal, November/December 2006, pp 43-52.

Peterson, Z. and Burns, R. (2005). Ext3cow: A Time-Shifting File System for Regulatory Compliance. ACM Transactions on Storage, 1(2), 2005, 190-212

Schauer, P. "Common sense security," Ohio CPA Journal (60:1), January 2001, pp 12-16.

Shedden, P., Ruighaver, T., and Ahmed, A. "Risk management standards: The perception of ease of use," 5th Annual Security Conference, The Information Institute, USA, Las Vegas, 2006.

Tan, T; Maynard, S; Ahmad, A; and Ruighaver, T, "Information Security Governance: A Case Study of the Strategic Context of Information Security" (2017). PACIS 2017 Proceedings. 43. http://aisel.aisnet.org/pacis2017/43

Thomson, K., and Von Solms, R. "Information security obedience: a definition," Computers & Security. (24:69-75) 2005

Volonino, L., Kermis, G., Gessner, G. (2004). Sarbanes-Oxley links IT to Corporate Compliance. Proceedings of the Tenth Americas Conference on Information Systems, New York, 2004

von Solms, B.V. (2006), "Information security-the fourth wave", Computers & Security, Vol. 25 No. 3, pp. 165-168

Whitman, M. "Enemy at the Gate: Threats to Information Security," Communications of the ACM (46:8) 2003, pp 91-95.

Wilson, P. "Governance and security: side by side," Computer Fraud & Security, 2007

Wright, M.A. "Keeping top management focused on security" Computer Fraud & Security (5:1) 2001, pp 12-14.

Yugay, I. and Klimchenko, V. (2004). SOX Mandate Focus on Data Quality and Integration. DM Review Magazine, Dmreview.com, Retrieved on 09/30/05

Appendices and Annexures

	Objective	Key Lessons
F1	Ensure Corporate Controls Strategy	Control strategy aligns the security governance and business objectives Antecedent to complete security and process integrity Provides the departments with control plans
F2	Encourage a Controls-Conscious Culture	Risk consciousness in employees creates a "prevention mentality." Helps in minimizing intergroup rivalry over security governance initiatives Creates an environment where individuals "watch out" for each other
F3	Establish Clarity in Policies and Procedures	Ensure the proper use of the applications and technological solutions instituted Make policies easily accessible Reflect control requirements in the systems Develop visibility of fair policies
F4	Maximize Regulatory Compliance	Meet legal, regulatory and contractual obligations Use compliance as a driver to develop security governance initiatives
F5	Ensure Continuous Improvements in controls	Continuous and iterative control assessment improves the controls environment Understand the organizational context of particular controls Change in roles should be reflected in subsequent controls
F6	Enable Responsibility and Accountability in Roles	Provide clarity in roles and ownership of decisions Promote transparency in roles and avoid sudden changes in responsibility structures

Table 1: Fundamental Objectives for Organizational Security Governance (Mishra 2015)

Objective Name	Evidence from ABC	Measures at ABC
F1-Ensure Regulatory Compliance	"SOX and HIPPA and other kinds of things are to help protect those data but these are guidelines, and they really don't mean anything by themselves because they don't come down and tell you specifically what you are supposed to do."	♦ Compliant with several legislations for the state as well as federal ♦ Internal audit department guides through the process ♦ Develop controls proactively that easily meet compliance requirements

Table 2: Regulatory compliance at ABC

Objective Name	Evidence from ABC	Measures at ABC
F2-Ensure continuous improvements in controls	"We need constant reevaluation as controls implementation is always an evolving process." "I have been in seminars where I dealt with fortune 500 companies, people who are making billions of dollars a year as revenue and they still have the same problem. You know those guys have everything, they have done everything, but it [control implementation] needs to be a constantly evolving process. They have to learn and then reeducate because things change."	♦ Constant reevaluation is done ♦ Considered an iterative process ♦ Attend seminars and conferences and learn about implementation practices form others ♦ Involve people across discipline and other agencies under the city, to help in implementation

Table 3: Continuous improvement in controls at ABC

Objective Name	Evidence from ABC	Measures at ABC
F3-Establish Responsibility and Accountability Structures	"The idea of having a documented hierarchy, especially around data is a must. If you think about it; we publish corporate organizational charts all the time." "So I think accountability piece is required. How do they control, say even a meter reading application? How do we ensure that every meter gets read every morning? You have meters that haven't been read, and there has been no consumption on that meter for over a year, and the service is still on then there is a problem. So put controls and make someone accountable, that's how you guarantee that every meter is being read and the consumption of gas and water is recorded."	♦ clear segregation of roles ♦ developing a controls chart with clear control responsibility and accountability ♦ encourages ownership of information

Table 4: Responsibility and accountability in structures at ABC

Objective Name	Evidence from ABC	Measures at ABC
F4-Ensure Corporate Control Strategy	"People should try to at first establish and see what the controls are. That's some degree reflected in your requirements. People need to know what they want to control. You have to know what you want to control, and the problem is that you don't know what you want to control." "You need to plan ahead and have a strategy about controls and its success. You need to figure out how am I going to be proactive rather than letting a reactive compliance approach drive my internal controls that we use."	♦ Provide more resources ♦ Enhance trust ♦ Proactive controls approach versus reactive approach ♦ Corporate level planning for security governance in advance

Table 5 Controls strategy at ABC

Objective Name	Evidence from ABC	Measures at ABC
F5-Establish Control Conscious Culture	"I think you need to have a clear core value; a clear company recognized or accepted perspective, the role of having those controls. For example, in mind, I think you should treat everything, every data you handle like it's your information. Would you leave your wallet out in the middle of the street, on the bench when you go to get a coffee? What type of care would you take if it's yours? That is the kind of care you need to take" "we cannot have controls everywhere but should have control in the places where we can get the most benefit for the organization."	♦ An environment where individuals watch out for each other ♦ Treat customers' information as if it is your information

Table 6: Control conscious culture at ABC

Objective Name	Evidence from ABC	Measures at ABC
F6-Maximize Clarity in Policies and Controls	"Make the policy and procedures clear and accessible. [Establish] Clarity in policies and controls, transparency in procedures, and gradually standardize the process; everyone knows what it could mean. What you [employee] can do to help & protect yourself without making those costly mistakes, make those very clear and understandable because if people don't understand them and they are not clear, people can't follow them and they make excuses".	♦ Explain the policies repeatedly ♦ Make the policies accessible easily ♦ The continuous iterative process of development ♦ Constant explanation of the benefits

Table 7: Clarity in policies and controls at ABC

An Exploratory Study of the Perceptions of Library Faculty and Patrons on Library Resources

Kyle Herman
kherman@georgiasouthern.edu
Georgia Southern University Libraries
Georgia Southern University
Statesboro, GA USA

Hayden Wimmer
hwimmer@georgiasouthern.edu*
Department of Information Technology
Georgia Southern Univeristy
Statesboro, GA USA

Carl Rebman Jr.
rebman@sandiego.edu
Department of Information Systems
University of San Diego
San Diego, CA USA

Abstract

Traditionally, a library has been defined by four aspects: its collection of books, the building that houses them, the librarians who are experts in retrieving this stored and cataloged information, and the patrons who are the end users of library services. Moving to the future, information systems will play a larger role within the library as physical collections will be circumvented by digital ones. With the growth of digital content, circulation of print material is in decline, and gate counters are showing a trend of decreased traffic. To address these issues academic and public libraries are trying to reinvent themselves both in physical and digital offerings. For example, libraries have added computer labs and teaching spaces, and they are expanding into mobile development, cloud sourcing, semantic library web, and online education. Unfortunately, these changes do not guarantee patrons will use library services. This study seeks to investigate the differences in librarian and patron perspectives on the physical and digital resources housed in the library's information systems. 2x2 factorial design was used on six constructs: system quality, information quality, context quality, user satisfaction, perceived benefit, and intent to use. The results showed that both librarians and patrons felt the utilization of the library building was the best overall indicator of the quality of a library. In addition, both groups felt closing a library would have a detrimental impact on the community. One interesting result that was the patrons indicated they viewed the storing of books as the most important service while the librarians favored buying access to online resources and technology. This suggests public perception of libraries is remaining static and based on an outdated view that libraries are only warehouses of physical books. Library Information Systems will play an ever increasing role in the future of libraries and the communities that they support, but only if they properly marketed.

Keywords: academic library, higher education, semantic library web, library space, user groups

1. INTRODUCTION

Prior research on the topic of the future of academic libraries only looked to fellow librarians for input. Up until the end of the 20th century, libraries told patrons what they had in their collection, and if the library did not have it or could not acquire it through an inter-library loan service, then it did not exist from the patrons' perspective.

With the proliferation of digital resources and broadband internet access, patrons have become more knowledgeable and if the library does not have what they want, then they can and often do go elsewhere. Thus, academic libraries are facing an ongoing change in user needs and expectations of information services. In addition, the perception of libraries and their staff have changed from being viewed as gatekeepers of information to just another option to access it.

When considering how library services and resources are perceived the question of how a library should measure success in a digital world is raised. In addition to counting books checked out, number of reference transactions, and gate counters, libraries are now measuring the count of digital articles/journals accessed or downloaded, the number of library website visitors, and the number of online librarian-patron interactions.

To stay relevant, libraries must continue to evolve to meet their patrons' needs and continue to offer the resources they find valuable. Often there is a dichotomous difference between what the patrons and the library faculty and staff perceive as an important resource.

This paper reports on the resources patrons find important and what the library group deems significant. This is key as financial resources are limited, choices need to be made, and future directions planned. As public libraries are facing dwindling and scare resources, academic libraries are responsible to set the pace and future trends. Information systems are increasingly becoming a critical and cornerstone of an academic library. The future of the library is not in stacks of books but in reimagining library spaces and improving on digital content via the library's information systems.

To guide our work, we ask two overarching research questions:

- If a patron believes that libraries have value and contribute to their community, is that a good indicator that they will use their library and its information system to locate resources?
- Is physically visiting a library a good indicator as to whether a patron will use a library's online resources?

The rest of the paper is as follows. First is review of relevant literature followed by an explanation of research methodology. Next is a reporting of survey results. The article concludes with a discussion of limitations and future research.

2. LITERATURE REVIEW

Libraries are in a constant struggle between increasing costs, old preconceptions, and perceived value by the patrons. The increase of the digital age also calls into question the quality of library services. The following is a brief literature of past library definitions, perceived patron value, and quality.

Defining a library

Traditionally, a library's identity has been tied to the kind of collection that it holds, and intuitively, people see libraries as a collection of books. Many patrons view the building as "the library" and not the resources and information that is held there. As collections increasingly go digital, the bond between users and the physical library will diminish (Sennyey, 2009). However, the environment that houses the collection is just as important as the information itself. The same is true for a digital environment where a well-designed webpage instills confidence in the information or services that it provides

Challenges facing the Library

In 2004, Brewer, Hook, Welburn-Simmons, and Williams believed that 1) higher education institutions would face a significant long-term budget decrease, 2) scholarly journals would continue to increase in subscription costs, and 3) there will be a need to raise salaries to hire qualified librarians.

In 2004, historical data showed that libraries have faced increasing costs of scholarly information with annual inflation rates of 6 – 12%, and shortages of trained librarians has led to increased costs of recruitment and retention (Brewer, 2004). These challenges hold true today where academic libraries continue to face

stagnant budgets but increased expenditures on subscription costs and salaries.

Creating large stocks of physical books does not create value externally where it once did in the 20th century. With circulation of print materials in decline, libraries are moving physical books to long-term storage in high-density locations. This availability of open space has allowed many academic libraries to house other departments such as coffee shops, writing centers, and student support centers. Gayton (2008) warned that if the activities of non-library organizations are not closely coordinated with those of the library, then the library could be at risk of becoming another office or classroom building. Furthermore, if the public views libraries as synonymous with physical books then moving all the books to high-density storage facilities off site could be detrimental to the library's identity.

The transition of information from print to electronic is being followed by the transition from purchased to open access content. Over time, this means that the library's collection of purchased materials, in both print and electronic formats, could be less important.

Patron perception

User's perceptions of library resources are influenced by where they use a resource and how they find it. (Sennyey, 2009). Kiran wrote that traditionally, the quality of an academic library has been described in terms of its collection and measured by the size of its holding and the counts of its use, i.e. number of visitors and number of items checked out. However, researchers in library usage and service are now looking at information needs, users' wants and perceptions of library services to define the quality of a library.

A 2005 OCLC environmental study found that users perceive libraries as a source of physical books but turn elsewhere when they desire digital information. The study found that Google has become the default digital library for most users (Perceptions of libraries and information resources: a report to the OCLC membership, 2005). The underlying reason for this is people navigate to and discover things on the internet using Google or another search engine. A paradox for libraries is that as they increase their digital collections, these resources will be discoverable by search engines and the library's overall importance will decrease from the patron's perspective (Sennyey, 2009).

The plight of academic libraries is rooted in the idea that to the public they are little more than storage facilities for print material. However, Gayton (2008) argued that patrons come to the library not just for the intellectual resources but also for space in which they engage those resources. Gayton (2008) believed they come for and value the "communal" experience of seeing and being seen by others engaged in the same serious studious activity.

Measuring quality

Sapp and Gilmour (2002) wrote that the measures of quality of research libraries should transcend the traditional counting of volumes and instead look at their ability to match user needs with relevant information. Until recently, the status of a research library was defined by the extent of its collection. Now with the flood of information available online and the value of the physical book diminishing, it is necessary to ask what the users' needs are and how are they changing. This then can provide context for reconsidering library services and collections (Gilmour, 2003). Historically, a library, public and/or academic, has been described in terms of its collection and measured by the size of its holdings. Today, a library's value is determined by how successfully patrons have access to needed information, regardless of format and location of the resource. Connecting patrons to the information they need is the fundamental objective of libraries (Nitecki, 1996).

3. RESEARCH METHODOLGY

This study sought to add to previous studies regarding patrons, and library faculty and staff perspectives on the physical and digital resources of the library. The method of investigation was an online survey in which participants were asked to answer questions on a 5 –point Likert scale or in rank order.

Survey questions were developed based on information discovered from the literature review, and also adapted from Wu & Wang 2006 (Wang, 2006). The survey measured six constructs: 1) system quality, 2) information quality, 3) context quality, 4) user satisfaction, 5) perceived benefit, and 6) intention to use.

Two groups of respondents were surveyed, library employees and non-library employees (patrons) and were related to the constructs of two library styles, physical and digital. In order to analyze the data, we create a 2x2 factorial design (Delaney & Maxwell, 2004) and conducted a multi-variate analysis of variance

(Hair, 2010). It was administered online using Qualtrics, and the estimated survey duration was 10 - 15 minutes.

One of the main intentions of the study was to measure the differences in perspectives between the two groups. This study utilized six different constructs, which are listed below and are as follows:

1. **SYSTEM QUALITY** - measured if the physical library was easy to user, user friendly, and had attractive features.
2. **INFORMATION QUALITY -** was the information available at a suitable time, was the information meaningful, and did the library make it easy to create documents
3. **CONTEXT QUALITY –** Did the physical library provide a complete knowledge portal with link to information sources for more detail
4. **USER SATISFACTION –** asked if the respondent was satisfied with knowledge and information needs, efficiency, effectiveness, and overall satisfaction.
5. **PERCEIVED BENEFIT –** would or does the library help in acquiring new knowledge or innovative ideas, accomplish tasks more efficiently, enhance quality of work, or job performance.
6. **INTENTION TO USE –** would the respondent use the library to make decisions, record knowledge, communicate knowledge

4. RESULTS AND DISCUSSION

An email invitation was sent out to approximately 55 library faculty and staff members, and Amazon Mechanical Turk was utilized for the non-library patron group. Of the 55 faculty invitations, 41 usable responses were collected for a response rate of 74.5%. The study received 32 participants to represent the non-library patron group.

The demographic questions from the survey showed that the library group is on average older than the non-library group. The library group had ten responses for the 45 - 54 age range, and seventeen out of the thirty-one respondents described themselves as 45 years old or older. The non-library group had eleven responses for the 25 - 34 age range, and twenty-seven responses out of thirty for 44 and younger.

The differences between genders were minimal. The library group responded as having eighteen females and thirteen males, while the non-library responded as having fourteen females and sixteen males. Twenty-two individuals in the library group and twenty-seven in the non-library reported "white" as their ethnicity. The non-library employee (patron) group had six individuals with an associate degree; thirteen with a bachelor's degree, and the remainder had some college or responded that they had a high school education.

For questions Lib6 - Lib13 listed in Appendix A, an Independent Samples T-Test was used. This test compares the means of the two independent groups in order to determine whether there is statistical evidence that the association population means are significantly different.

This answers the question as to whether the difference between the means is statistically significant or whether the difference is due to sampling error. Two difference variances are obtained: Equal variances assumed, and Equal variances not assumed. Levene's Test for Homogeneity of Variances was then used to determine which variance to use. If the significance is < .05, then Equal variances not assumed was used. If the significance is > .05, then Equal variances assumed was used.

The P-value listed in the "Sig (2-tailed)" column was then inspected. If the P-value was > .05 then the results fail to reject the null hypothesis and the difference is due to chance. If the P-value was < .05 then the results allow you to reject the null hypothesis ("SPSS Tutorials: Independent Samples t Test," 2018).

The results for question Lib6 are displayed in graph form in Figure 1 and are statistically significant. See Figure 2 for the T-test results for question Lib6.

This shows the data on how often survey takers physically visit a library with the non-library group indicating that they predominately never or almost never physically visit a library. The results for the library group are as expected in that they very often physically visit a library.
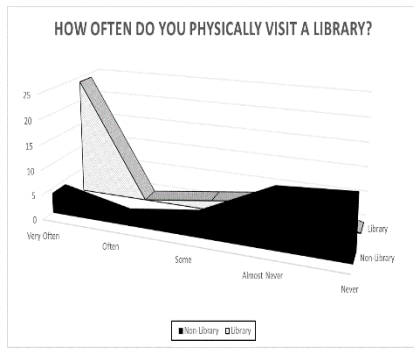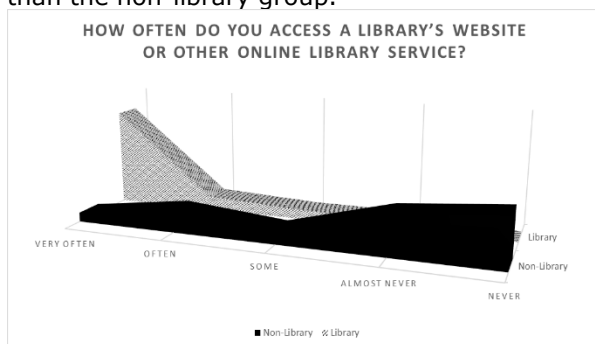
Figure 1: How often do you visit a library?

Independent Samples Test									
		Equality of Variances		t-test for Equality of Means					
						Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the
		F	Sig.	t	df				Lower	Upper
Lib6	Equal variances not assumed			-6.949	53.735	0.000	-2.260	0.325	-2.912	-1.608

Figure 2: T-Test results for "How often you visit a library?"

The results for question Lib7 are displayed as a graph in Figure 3 and the T-test results are shown in Figure 4. The library group visits a library website on a much more regular basis than the non-library group.


Figure 3: How often do you access a library's website?

Independent Samples Test									
		Equality of Variances		t-test for Equality of Means					
						Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the
		F	Sig.	t	df				Lower	Upper
Lib7	Equal variances not assumed			-7.662	52.151	0.000	-2.269	0.296	-2.863	-1.675

Figure 4: T-Test results for How often do you access a library's website?

The results for question Lib8 are shown as a graph in Figure 5 and the T-test results are shown in Figure 6. Figure 5 supports the data shown in Figure 3 that the non-library group never or almost never visits a library's website to initially search for digital information. Search engine usage is high for the non-library

employee (patron) group and for the library employee group.
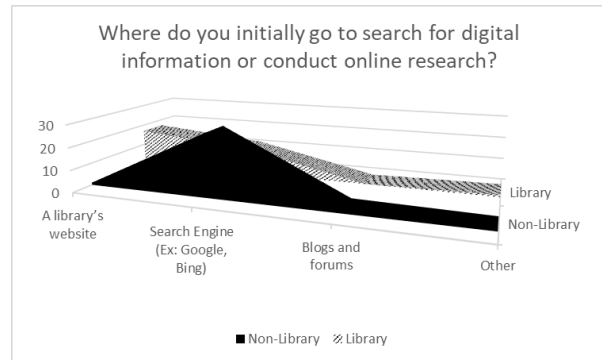
Figure 5: Initial search for digital information

Independent Samples Test									
		Equality of Variances		t-test for Equality of Means					
						Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the
		F	Sig.	t	df				Lower	Upper
Lib8	Equal variances not assumed			3.950	39.082	0.000	0.516	0.131	0.252	0.780

Figure 6: T-test results for Initial search for digital information

The results from questions Lib9, Lib10, Lib11, and Lib12 were not statistically significant and are not shown. The results from question Lib13 were significant and are shown as a graph in Figure 7 and the T-test results are shown in Figure 8.

Both groups agreed that shutting down a library would have a high impact. This indicates that although the non-library employee (patron) group may not visit a library physically or digitally as indicated in Figure 1 and 3 respectively, they still believe that the library provides value to the community or organization.
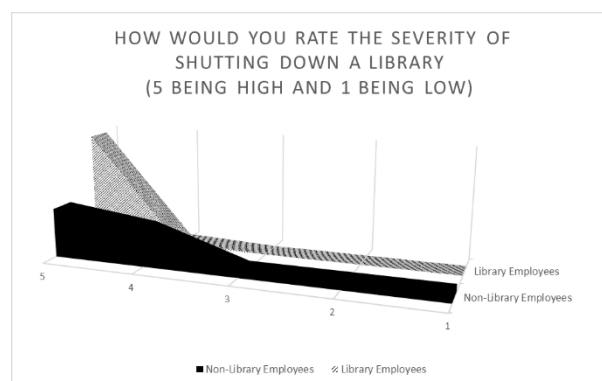
Figure 7: Severity of shutting down a library.

Independent Samples Test									
		Equality of Variances		t-test for Equality of Means					
						Sig. (2-tailed)	Mean Difference	Std. Error Difference	Interval of the
		F	Sig.	t	df				Lower	Upper
Lib13	Equal variances not assumed			-4.105	29.233	0.000	-0.932	0.227	-1.396	-0.468

Figure 8: T-test results for Severity of shutting down a library

Results of our analysis showed that there are significant differences primarily between the library employee group and non-library employee (patron) group with little significant differences between perceptions of physical and digital libraries.

A summary can be viewed in Table 1. Following recommendations by Hair (2010), we report on the Wilks Lambda. For the Library and Non-library groups, System Quality was significant with $F = 5.344$, $p<0.001$, Wilk's $\lambda = 0.873$. Information quality was significant with $F = 3.182$, $p<0.05$, Wilk's $\lambda = 0.895$.

Context Quality and User satisfaction revealed no significant difference. Perceived benefit was significant with $F = 5.056$, $p<0.001$, Wilk's $\lambda = 0.843$.

Finally, intent to use was significant with $F = 2.843$, $p<0.05$, Wilk's $\lambda = 0.884$. The differences between perceptions of the physical and digital library were not as strong with weak support only for information quality with $F = 2.194$, $p<0.1$, Wilk's $\lambda = 0.925$.

	Library vs Non-Library	Physical vs Digital Lib
System Quality	Significant (p< 0.01)	Not Significant
Information Quality	Significant (p< 0.05)	Significant (p< 0.
Context Quality	Not Significant	Not Significant
User Satisfaction	Not Significant	Not Significant
Perceived Benefit	Significant (p<0.01)	Not Significant
Intent to Use	Significant (p< 0.05)	Not Significant

Table 1 – Summary of Statistical Analysis of 2X2 Factorial Design

5. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This work examined perceptions of physical and digital libraries from the perspectives of library employee and non-library employees (patrons). It was found that both groups believe "the" library adds value to their community regardless of whether they visit it themselves. Based on the responses, the researchers hypothesize that if a person doesn't physically visit a library then it's likely that they won't use the library's online resources as well. This hypothesis encompasses

how they search for information on the internet and whether they use a library's online catalog.

From the statistical analysis of the data, a 2x2 factorial design was used on six constructs: system quality, information quality, context quality, user satisfaction, perceived benefit, and intent to use. We found significant differences in the library employee and non-library employee (patrons) groups on system quality, information quality, perceived benefit, and intention to use. Additionally, a significant difference was found among physical and digital libraries only on information quality. The library group responses indicate that they value the resources and services that a library offers more strongly than the non-library employee (patron) group.

The findings of this study are interesting though they are based on a small sample size. In addition, the library group's response may not be representative of all universities. Further research should be conducted to see if there are differences due to size of the university and Carnegie Mellon ranking.

6. REFERENCES

Delaney, S., & Maxwell, H. (2004). *Designing Experiments and Analyzing Data*. Mahwah, NJ: Lawrence Erdlam Associates Inc.

Gayton, J. T. (2008). Academic Libraries: "Social" or "Communal?" The Nature and Future of Academic Libraries. *The Journal of Academic Librarianship, 34*(1), 60-66.

Gilmour, G. S. a. R. (2002). A Brief History of the Future of Academic Libraries: Predictions and Speculations from the Literature of the Profession, 1975 to 2000--part one, 1975 to 1989. *Libraries and the Academy, 2*(4), 553-576. doi: https://doi.org/10.1353/pla.2002.0086

Gilmour, G. S. a. R. (2003). A Brief History of the Future of Academic Libraries: Predictions and Speculations from the Literature of the Profession, 1975 to 2000--part two, 1990 to 2000. *Libraries and the Academy, 3*(1), 13-34. doi: https://doi.org/10.1353/pla.2003.0008

Hair, J. F. (2010). *Multivariate data analysis*: Pearson College Division.

Hansoon, J. (2015). Documentality and legitimacy in future libraries - an analytical

framework for initiated speculation. *New Library World, 116*(1/2), 4-14. doi: https://doi.org/10.1108/NLW-05-2014-0046

Joseph M. Brewer, S. J. H., Janice Welburn-Simmons, and Karen Williams. (2004). Libraries Dealing with the Future Now. *Library Faculty Research and Publications*(234), 1-9.

Lewis, D. W. (2007). A Strategy for Academic Libraries in the First Quarter of the 21st Century. *College & Research Libraries, 68*(5), 418-434. doi: http://dx.doi.org/10.5860/crl.68.5.418

Marcum, M. A. J. a. J. W. (2010). Sustainability Challenge for Academic Libraries: Planning for the Future. *College & Research Libraries, 71*(2), 160-259. doi: https://doi.org/10.5860/0710160

Nitecki, D. A. (1996). Changing the Concept and Measure of Service Quality in Academic Libraries. *The Journal of Academic Librarianship, 22*(3), 181-190. doi: https://doi.org/10.1016/S0099-1333(96)90056-7

Perceptions of libraries and information resources: a report to the OCLC membership. Dublin, Ohio: OCLC Online Computer Library Center, 2005. Web.. https://lccn.loc.gov/2006367596.

Pongracz Sennyey, L. R., and Caroline Mills. (2009). Exploring the future of academic libraries A definitional approach. *The Journal of Academic Librarianship, 35*(3), 252-259.

Rosa, C. 2005. College students' perceptions of libraries and information resources : a report to the OCLC membership. Retrieved from https://www.oclc.org/content/dam/oclc/reports/pdfs/studentperceptions.pdf

SPSS Tutorials: Independent Samples t Test. (2018).

Wang, J.-H. W. a. Y.-M. (2006). Measuring KMS success: A respecification of the DeLone and McLean's model. *Information & Management, 43*(6), 728-739. doi: https://doi.org/10.1016/j.im.2006.05.002

Wu, J.-H., & Wang, Y.-M. (2006). Measuring KMS success: A respecification of the DeLone and McLean's model. *Information & management, 43*(6), 728-739.

APPENDIX A

Lib6. How often do you physically visit a library?
Scale (1-5, with 1 being very often and 5 being never)

Lib7. How often do you access a library's website or other online library services?
Scale (1-5, with 1 being very often and 5 being never)

Lib8. Where do you initially go to search for digital information or conduct online research?
A library's website
- Search Engine (Ex: Google, Bing)
- Blogs and forums
- Other (write in response)

Lib9. How strongly do you feel that print material should be digitized and retired?
    Scale 1-5 (1 strong feeling and 5 no feeling)

Lib10. How strongly do you feel that libraries should focus their attention on providing information not easily accessed via the web?
    Scale 1-5 (1 being high focus and 5 being no focus)

Lib11. How strongly do you feel that libraries should have quiet space/floors?
    Scale 1-5 (1 strong feeling and 5 no feeling)

Lib12. How do you prefer to read?
- Physical/hard copy
- Digital (backlight)
- Digital (non-backlight)
- Audiobooks
- I don't have a preference
- I don't read on a regular basis

Lib13. How would you rate the severity of shutting down a library (academic or community)?
    Scale 1-5 (1 being high, 5 being low)

Developing a GIS Dashboard Tool
to Inform Non-Profit Hospitals of Community
Health Needs at the Neighborhood Level

Evan Copello*
n00692543@unf.edu

Jason Smith*
smalls257@outlook.com

Anna Mease˄
Mease.Anna@mayo.edu

Karthikeyan Umapathy*(C)
k.umapathy@unf.edu

Dan Richard*
drichard@unf.edu

Ann-Marie Knight˄
Knight.AnnMarie@mayo.edu

Monica Albertie˄
Albertie.Monica@mayo.edu

Hana Hamadi*
n00968253@unf.edu

Aaron Spaulding˄
Spaulding.Aaron@mayo.edu

Emma Apatu⌘
apatue@mcmaster.ca

*University of North Florida, Jacksonville, FL – 32224, USA
˄Mayo Clinic, Jacksonville, FL – 32224, USA
⌘McMaster University, Hamilton, ON – L8S4L8, Canada
(C) – Corresponding Author

Abstract

The objective of this paper is to describe the methods used to develop geographic information systems (GIS) dashboard tool and explain how it can assist nonprofit hospitals to identify priority neighborhoods. Multiple data sources from the 500 Cities Project databases were analyzed, and two online dashboards were created. The first dashboard is a hospital-specific composite dashboard, and the second is a comparison dashboard of health outcomes identified by both the hospital and the county's community health needs assessment focused on neighborhood-level disparities. Hospital-specific health outcomes were Stroke, Diabetes, and Coronary Heart Disease. County-specific health outcomes were Obesity, Dental, and Mental Health. All of the six health outcomes were standardized, rescaled, and weighted within the final composite score. Tableau was used for developing the dashboards and geographically mapping the analyzed data. The maps were developed specifically for a large hospital in Florida; however, this methodology can be utilized by other hospitals across the US. City-specific data is essential to ensure the accuracy of community health needs. The development of an interactive, comprehensive map using Tableau is a useful tool for visualizing target neighborhoods for community health outreach. The integration of community needs assessment findings into the development of composite scores allows hospitals in the US to use this tool to inform community health outreach strategy adequately.

Keywords: GIS, Dashboards, Data Science, Decision-Making, Health Outcomes, Community Health.

1. INTRODUCTION

In 2011, nonprofit hospitals saved approximately $24.6 billion due to tax exemptions (Rosenbaum, Kindig, Bao, Byrnes, & Colin O'Laughlin, 2015). To regulate the use of this money and to justify the continued usage of tax-exempt status, the 2010 Patient Protection and Affordable Care Act (ACA) implemented stricter criteria for nonprofit hospitals to demonstrate community benefits. Nonprofit hospitals must not only conduct a Community Health Needs Assessment (CHNA) every three years but also provide evidence that they are addressing these needs within the target community (IRS, 2013). As a result, hospitals must gather data and develop strategies that can address community health outcomes (IRS, 2013). As such, this study describes the process by which a hospital can articulate and depict neighborhood-level health needs to allow hospitals to meet the needs of the communities they serve.

Studies that report on nonprofit hospitals progress in meeting federal CHNA requirements are substantially lacking in the literature. However, one recent national cross-sectional study attempted to address this gap by examining the progress of tax-exempt hospitals across the US in meeting four CHNA implementation activities: "strategy formulation to address identified needs, participation in the development of community-wide plans, planning for the provision of community benefits, and budget development to address identified needs." The authors found that only a minority, n= 574 (36%) of hospitals met all four actions (Cramer, Singh, Flaherty, & Young, 2017). This staggering finding indicates that there are many barriers associated with CHNA activities.

ACA legislation requires that hospitals collaborate with local health professionals and organizations (Pennel, Burdine, Prochaska, & McLeroy, 2017). While the collaborative activities in and of themselves may not pose a barrier, the areas in which hospital administrators should pursue and spend limited hospital resources are not always clear. Conducting a needs assessment, and defining population needs requires both quantitative and qualitative data at various population levels. However, while state, city, and county data sources are often available, neighborhood-level data are often lacking or unavailable altogether. While there is currently one data-driven tool available to hospitals to use (Community Commons) (Chow, Jaffee, & Snowden, 2003), the tool doesn't allow hospitals a macro -view of local health outcomes based on CHNA and hospital identified health needs. The usage of localized data can guide the CHNA and ultimately inform hospital-sponsored outreach in a more targeted way that better meets community needs and is a better utilization of hospital resources.

In this paper, we describe two geographic information systems (GIS) dashboards developed to assist a large hospital in the State of Florida in viewing health outcomes of different neighborhoods in Duval County. The nonprofit

hospital intends to use the GIS dashboards to select communities to implement its WellnessRx program created in response to CHNA needs.

2. DESIGN SCIENCE RESEARCH METHODOLOGY

Design science is a problem-solving paradigm involving the creation, analysis, and evaluation of artifacts designed as a solution to real-world problems. Design science research (DSR) methodology can be effective when creating an innovative information-based artifact strategically and holistically. The study reported in this paper uses design science research methodology to govern the design, development, and evaluation of the GIS dashboard.

We describe our adoption of the design science research methodology following the framework described in Peffers et al. (2007).

Problem Identification
A fundamental guideline for practicing design science research is the creation of an artifact that addresses an important and unresolved real-world problem (Hevner, March, Park, & Ram, 2004). In this paper, we address the problem of selecting a neighborhood to implement public health initiatives using a data-driven approach as opposed to gut-driven decisions. Section 3 discusses the problem identification step of the DSR methodology.

Define the Objectives of a Solution
The objective of this study is to develop a GIS Dashboard that addresses the identified wicked problem. The goals of the artifact are discussed in section 3. An important aspect of defining the objectives of a solution is identifying key ingredients that are needed to design the artifact. In regards to our study, those key ingredients are the relevant data sources. We worked along with the stakeholders to identify data sources pertinent to develop the GIS dashboard, as discussed in section 4.

Design & Development
Creation of an innovative artifact is central to the DSR outcome. Hevner et al. (2004) have described several forms a DSR artifact can take. In this study, we developed composite scoring methods to assign community health indicators to neighborhoods to assist with the decision-making process. We implemented the composite scoring methods along with other relevant information in a GIS dashboard tool. The dashboard tool was developed in five iterations

with each iteration being a two weeks sprint. Section 5 discusses the composite scoring methods, and section 6 provides an overview of the GIS dashboard. GIS dashboard can be accessed from the URL: http://bit.ly/17fldssgmayotableau

Demonstration
Demonstrating how to the use of the artifact to solve the identified problem is an essential step of the DSR process. Demonstration steps help in ensuring the stakeholders have required resources to solve the problem using the designed artifacts. In regards to this study, we demonstrated the artifacts to the stakeholders during the three-month development period. We also created a YouTube video to show how to use the GIS dashboard to select a neighborhood based on community health indicators. The demonstration video can be accessed from https://www.youtube.com/watch?v=clIBj9jCQ6 w.

Evaluation
Assessing the utility of the artifact in regards to solving the problem is a required activity of DSR methodology. Hevner et al. (2004) have described multiple forms of evaluating DSR artifacts. In this paper, we performed a summative and interpretive form of evaluation. We used software walkthrough technique to illustrate the utility of the GIS dashboard to the stakeholders as well as identify anomalies and defects with the dashboard product. Results of the walkthrough are discussed in section 6.

Communication
Disseminating the importance of the problem and innovativeness of the solution to an appropriate audience is the final set of actions of a design science research. We have presented the problem and GIS dashboard in the open to the public forum to the stakeholder and community at large in the northeast Florida region. Publication of this paper is a valid form of diffusing the resulting knowledge gained from this study.

3. DECISION-MAKING SUPPORT NEED

Mayo Clinic in Florida is a non-profit hospital that served over 1.3 million people. In response to their 2016 CHNA, Mayo Clinic Florida launched Wellness Rx in collaboration with New Town Success Zone (NTSZ), a grassroots community neighborhood initiative focused on connecting families with available community resources within the New Town community in Jacksonville, FL. Wellness Rx is a community-led wellness

program designed to empower and educate New Town residents with information to improve their overall health. Through civic participation, the NTSZ has created a model of engagement that encourages self-responsibility, accountability, and community-driven action around health and wellness priorities.

Mayo Clinic wanted to expand the Wellness Rx program to other community neighborhoods. However, the identification of the target area posed limitations because there was no health outcomes data available that could help identify a community based upon neighborhood geography. Thus, Mayo Clinic did not have any reliable means to make decisions on which neighborhoods to implement Wellness Rx programs. Therefore, Mayo Clinic formed a partnership with The Florida Data Science for Social Good (FL-DSSG) program to develop a data tool to address this need.

The Florida DSSG (https://dssg.unf.edu/) program was founded in 2017 and was created to provide students at the University of North Florida with hands-on opportunities to develop data-driven solutions for local public sector organizations, including nonprofits and government agencies. The FL-DSSG offers a summer internship that recruits students from diverse disciplines to work alongside community partners on social good projects that blend data science and technology design.

The overall goal of the Wellness Rx and DSSG project was to create a Geographic Information Systems (GIS) based dashboard tool to visually represent health outcomes data at a neighborhood level which could then be used to identify and evaluate other neighborhoods for future interventions. Given that few repositories provide hospital administrators ability to develop data maps of community health needs, this paper focuses on describing the resultant GIS dashboard utilized at Mayo Clinic Jacksonville to help identify community needs at the neighborhood level.

4. DASHBOARD DATA SOURCES

Data for this study was obtained from multiple sources. First, we used the 500 Cities Project databases developed by the Centers for Disease Control and Prevention (CDC) and the Robert Wood Johnson Foundation (CDC-500Cities, 2016). The 500 Cities Project data recorded health outcomes in 500 cities around the United States at the Census Tract level, for the year 2014. The City of Jacksonville was one of the 500 cities that participated in the project.

However, only 164 out of the total 173 census tracts for Duval County were available. The absent census tracts include tracts 139.02, 139.05, 140.01, 140.02, 141.01, 141.02, 142.02, 142.03, and 142.04. The data is open-source, but was prepared and provided to the research team by the Florida Department of Health in Duval County. The 500 Cities project provided 28 variables within three broad categories (see Appendix A), each consisting of multiple variables (k), namely: Health Outcomes (k = 13), Prevention Measures (k = 10), and Unhealthy Behaviors (k = 5).

Second, to build the maps, we used data from:
- The University of Florida GeoPlan Center supplied Brownfield maps, as well as community service centers.
- The Florida Department of Environmental Protection: yielded additional Brownfield maps.
- City of Jacksonville Property Appraiser: added average housing values by census tract.
- Zillow: provided neighborhood-name correlates for census tracts.
- American Community Survey (ACS): census demographics (education level, household income, and ethnicity)
- Behavioral Risk Factor Surveillance System (BRFSS): health indicators (binge-drinking, asthma, and smoking rates).

The resultant maps were built using geographical data layered into Tableau. Google Code™ Application Programming Interfaces (APIs) provided the base map of the city as well as an interactive Google Street View™ that allows the user to pick a census track and virtually travel around the area.

5. COMMUNITY HEALTH INDICATOR COMPOSITE SCORES

In this study, we developed two community health indicators using composite scoring mechanism. We used a total of 6 health variables to create two composite scores. First, the Wellness Rx program identified three variables of interest, Stroke, Diabetes, and Coronary Heart Disease. These three primary community health variables served as the foundation for a "Mayo Composite" to identify the health of the community. Second, the Community Health Needs Assessment (CHNA) for Duval County identified three other variables as "Targeted Negative Health Outcomes": Obesity, Poor Mental Health, and No Dental Visit in the Past Year. These variables were added to

the Mayo Composite measures to create an overall "CHNA Composite."

To create composite scores, standardization, rescaling, and weighting was necessary. Our first step to compare data was to standardize each outcome. Standardization was done to allow for comparison between the different health outcomes. This process allowed us to combine different outcomes into a single index or composite score. We converted each of the health variables into standardized values (z-scores). We then rescaled the values on a scale of 0-10 to allow easy comparison across variables and neighborhoods using the formula listed in Equation 1.

Rescaled Value =(X1-MIN([a,b]))/(MAX([a,b])-MIN([a,b]))*(11-1)

Equation 1. Rescaling Formula for All Variables

"X1" represents the value to be rescaled in the original metric and [a,b] represents the total range of the data whereby $X1 \mid a \leq x \leq b$. The "(11-1)" portion of the equation set the largest scale value as 10. The automatic low-end of the rescale will always be "0" with this formula. After the variables were standardized and rescaled, they were combined to create the Mayo composite and the CHNA composite. These composite scores were used to gain an overview view perspective of a group of relevant health disparities, presented by Census Tract.

Initially, there were five separate composites, but after running a Pearson Correlation (r) analysis in SPSS 24, we opted to keep only two of them, as they all were highly correlated to each other (r > .800, p < .001). The decision was made to keep composites that would be useful for Mayo Clinic and their partnering organizations.

The last step was to weight the composites. Combining the outcomes would create equally weighted composites. For the Mayo Composite, the three health outcomes would have 33.3% weight in the composite, making them equal. To highlight the community's needs, we wanted to present the three other variables within the CHNA Composite properly. To do this, we ran a Pearson Correlation between Stroke, Coronary Heart Disease, Diabetes, Obesity, Dental, and Poor Mental Health, with the Mayo Composite. Table 1 shows the weightages used for the CHNA Composite.

These weights are then inserted into the composite formulas to reflect the data. Equation 2 shows the formula used for calculating the weighted composite scores. These were the final composites that were used to display community health indicators for the selected neighborhood in the Tableau dashboards.

Variables	Correlation	Equation	Weight
Diabetes	.970	.970/5.145	18.85%
Stroke	.948	.948/5.145	18.43%
Heart Disease	.929	.929/5.145	18.06%
Obesity	.867	.867/5.145	16.85%
Dental	.748	.748/5.145	14.54%
Mental Health	.683	.683/5.145	13.28%
Total	5.145		100%

Table 1. CHNA Weighting Methodology

Weighted Mayo Composite: Heart Disease(.3333) + Stroke(.3333) + Diabetes(.3333)

Weighted CHNA Composite: Diabetes(.1885) + Stroke(.1843) + Heart Disease(.1806) + Obesity(.1685) + Dental(.1454) + Poor Mental Health(.1328)

Equation 2. Weighted Composite Scores

6. GIS DASHBOARD

Several programs were used for data analyses to build the actual repository of information contained in the dashboard. All of the data was stored and maintained in Microsoft Excel. For statistical data analyses, we used the Statistical Package for Social Sciences, version 24 (IBM, 2016), and to display the results in series of maps geographically, we used Tableau 10 (Tableau, 2018).

The completed GIS tool has two dashboards (Appendix B), the first a composite and the second a comparison. To distinguish between health outcome severity, shading is used in both dashboards with darker shading representing higher composite scores or worse health outcomes. Each dashboard contains different additional features. In the composite dashboard, health outcomes can be viewed individually or combined. The dashboards were successful at providing visual data regarding the specific health outcomes that were most prevalent in each census tract. There is also a street view feature providing a virtual walk-through of a census tract. Census demographics (education level, household income, and ethnicity) and health indicators (binge-drinking, asthma,

smoking rates, etc.) are listed to the right of the composite map. The GIS tool aims to help nonprofit hospitals identify information regarding a particular health indicator at the census tract level.

GIS dashboards offer an interactive feature whereby, if the mouse is placed over an area, census tract number, neighborhood name and numeric representation of composite score are displayed. Lastly, community resources available within each health tract are provided. The comparison dashboard allows you to select multiple census tracts at once and compare them by shaded color, composite, and outcome scores. Along with the selected comparison, a summary table will be produced, providing a comparison of the highest negative health outcomes. Screenshots of the final tableau dashboard tool can be found in Appendix B.

7. WALKTHROUGH OF THE GIS DASHBOARD

A walkthrough is an informal software inspection method used for evaluating as well as educating a software product to an audience. Software inspection methods, including walkthrough, are described in the IEEE 1028-2008 - Standard for Software Reviews and Audits (IEEE-1028, 2008). Following the walkthrough purposes described in the IEEE 1028-2008 standard, we used walkthrough technique to evaluate conformance to stakeholder expectations and evaluate the usability of the dashboard.

Walkthrough technique adopts a step-by-step narrative use of the artifact (Light, Burgess, & Duguay, 2018). Walkthroughs reveal the user engagement process with the artifact to the intended goals. User walkthroughs can help designers identify users' departure from intended procedures or engagement process (Light et al., 2018). Tweaking the user interface and potentially functional design in response to identified user engagement departures can help in improving the user experience of the artifact.
To facilitate the walkthrough procedures, we demonstrated both composite and comparison dashboards to key stakeholders from Mayo Clinic.
The step-by-step illustrated use of the dashboards is shown in Appendix B. In this section; we discuss the results of the walkthrough.

After the walkthrough demonstrations, stakeholders stated that both dashboards conform with their need of viewing health outcomes for neighborhoods and selecting neighborhoods to implement Wellness Rx program. During the demonstration, stakeholders used the dashboard for a variety of scenarios, and outcome & composite index options. Stakeholders identified that data for two neighborhoods did not fit with their knowledge of the community. When we investigated the dataset, discovered an issue with the merging process. We rectified the error, and rerun the merging process, and subsequently, the dashboard data source was updated. Some stakeholders stated the dashboard did not work in the Internet Explorer and Edge browsers. Since the dashboard utilizes Google Maps and other related tools; we recommended stakeholders to use Chrome and Firefox browsers.

8. DISCUSSION

Many tax-exempt hospitals are faced with the perennial issue of determining how to develop a CHNA implementation strategy that aligns with priority community health needs. The complexity of deciding where to begin, what stakeholders to engage, and what data to use can stymie the process. However, localized data represented visually via a GIS dashboard tool can maximize nonprofit hospitals investment in community benefit through a targeted approach (Newbold, 2018). In this paper, we described how Mayo Clinic Jacksonville, partnered with FL-DSSG supported by the University of North Florida, developed an online data dashboard that aids hospital administers' decision-making in the CHNA pre-planning process. Hospitals around the nation can use a similar tool as a means to start conversations with their CHNA implementation teams and to analyze health disparities at the neighborhood level.

There are some limitations to the GIS dashboard. For instance, the 500 cities data is not available each year; therefore, the dashboard only serves as a snapshot of a particular point in time. To assess the success of any interventions, it is suggested that there be a pre and post-test given to program participants in each census tract. The data cannot be relied on to show change over time. The dashboard primarily relies on 500 cities data from 2014. It can be argued that trends have already changed prior to interventions. Despite these challenges, we were able to develop a GIS dashboard that provides neighborhood-level community health data that was otherwise not readily.

Practical Implications

The GIS dashboard described in this paper provides a more localized understanding of Jacksonville neighborhoods and details current health outcomes and resources. Knowing which resources are available within a census tract helps health workers know what support they may have within an area and what they need to focus on developing. When identifying where to take health programs, this data tool provides a faster approach to identifying target communities and is based solely on health data. Basing decisions on health data alone removes a lot of the subjectivity humans bring to decision-making. This can enable nonprofit hospitals to coordinate efforts to improve population health. This dashboard can be updated with new census data each year, providing consistency in available health data and the opportunity to track progress within an area. It allows hospitals to evaluate the impact of their community outreach efforts.

The tool proved to be useful for Mayo Clinic Wellness Rx program, and overall fills a void for available neighborhood-level data in the Jacksonville area. This tool is publicly available and may be used by other non-profit hospitals and health organizations in the Jacksonville area. Additionally, the methodology is provided so that other health organizations and hospitals across the U.S. can replicate this dashboard for their city.

In conclusion, this paper addresses one of the CHNA needs of hospitals in the U.S. by developing a GIS dashboard containing community health outcomes and capability to filter neighborhoods based on composite scores of the health variables. We developed the dashboard following the design science research methodology. In the paper, we describe the composite scoring method included within the dashboard. We use walkthrough technique to confirm that the dashboard meets the expectations of the stakeholders. Results of the walkthrough reveal some issues which were addressed to satisfactorily meets all of the stated needs of the stakeholder.

9. ACKNOWLEDGMENTS

10. REFERENCES

CDC-500Cities. (2016). 500 cities project: Local data for better health | home page | CDC. Retrieved from https://www.cdc.gov/500cities/index.htm

Chow, J. C., Jaffee, K., & Snowden, L. (2003). Racial/ethnic disparities in the use of mental health services in poverty areas. American Journal of Public Health, 93(5), 792-797. 10.2105/AJPH.93.5.792

Cramer, G. R., Singh, S. R., Flaherty, S., & Young, G. J. (2017). The progress of US hospitals in addressing community health needs. American Journal of Public Health, 107(2), 255-261. 10.2105/AJPH.2016.303570

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105. 10.2307/25148625

IBM. (2016). SPSS software. Armonk, NY: IBM.IEEE-1028. (2008). IEEE 1028-2008 - IEEE standard for software reviews and audits. Retrieved from https://standards.ieee.org/standard/1028-2008.html

IRS. (2013). Community health needs assessments for charitable hospital. Retrieved from https://www.govinfo.gov/content/pkg/FR-2013-04-05/pdf/2013-07959.pdf

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. New Media & Society, 20(3), 881-900. 10.1177/1461444816675438

Newbold, P. A. (2018). Developing data to support effective coordination of nonprofit hospital community benefit investments. Journal of Healthcare Management, 63(4), 281–282. 10.1097/JHM-D-18-00108

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77. 10.2753/MIS0742-1222240302

Pennel, C. L., Burdine, J. N., Prochaska, J. D., & McLeroy, K. R. (2017). Common and critical components among community health assessment and community health improvement planning models. Journal of Public Health Management and Practice, 23, S21. 10.1097/PHH.0000000000000588

Rosenbaum, S., Kindig, D. A., Bao, J., Byrnes, M. K., & Colin O'Laughlin. (2015). The value of the nonprofit hospital tax exemption was $24.6 Billion in 2011. Health Affairs, 34(7), 1225-1233. 10.1377/hlthaff.2014.1424

Tableau. (2018). Tableau desktop. Seattle, WA, USA: Tableau. Retrieved from https://www.tableau.com/products/desktop

Appendix A

Below table provides a listing of the 500 Cities project measures obtained from the Behavioral Risk Factor Surveillance System (BRFSS).

| Category	Measures
Unhealthy Behaviors	Binge drinking among adults aged ≥18 years
	Current smoking among adults aged ≥18 years
	No leisure-time physical activity among adults aged ≥18 years
	Obesity among adults aged ≥18 years
	Sleeping less than 7 hours among adults aged ≥18 years
Health Outcomes	Arthritis among adults aged ≥18 years
	Current asthma among adults aged ≥18 years
	High blood pressure among adults aged ≥18 years
	Cancer (excluding skin cancer) among adults aged ≥18 years
	High cholesterol among adults aged ≥18 years who have been screened in the past 5 years
	Chronic kidney disease among adults aged ≥18 years
	Chronic obstructive pulmonary disease among adults aged ≥18 years
	Coronary heart disease among adults aged ≥18 years
	Diagnosed diabetes among adults aged ≥18 years
	Mental health not good for ≥14 days among adults aged ≥18 years
	Physical health not good for ≥14 days among adults aged ≥18 years
	All teeth lost among adults aged ≥65 years
	Stroke among adults aged ≥18 years
Prevention	Current lack of health insurance among adults aged 18-64 years
	Visits to doctor for a routine checkup within the past year among adults aged ≥18 years
	Visits to dentist or dental clinic among adults aged ≥18 years
	Taking medicine for high blood pressure control among adults aged ≥18 years with high blood pressure
	Cholesterol screening among adults aged ≥18 years
	Mammography use among women aged 50-74 years
	Papanicolaou smear use among adult women aged 21-65 years
	Fecal occult blood test, sigmoidoscopy, or colonoscopy among adults aged 50-75 years
	Older adults aged ≥65 years who are up to date on a core set of clinical preventive services by age and sex

Table 2. Variables from the 500 Cities Project

Appendix B

WALKTHROUGH OF COMPOSITE DASHBOARD

User task: Browse and select a neighborhood based on composite and health indicator interests.

Step 1: Open the Composite Dashboard in the Browser
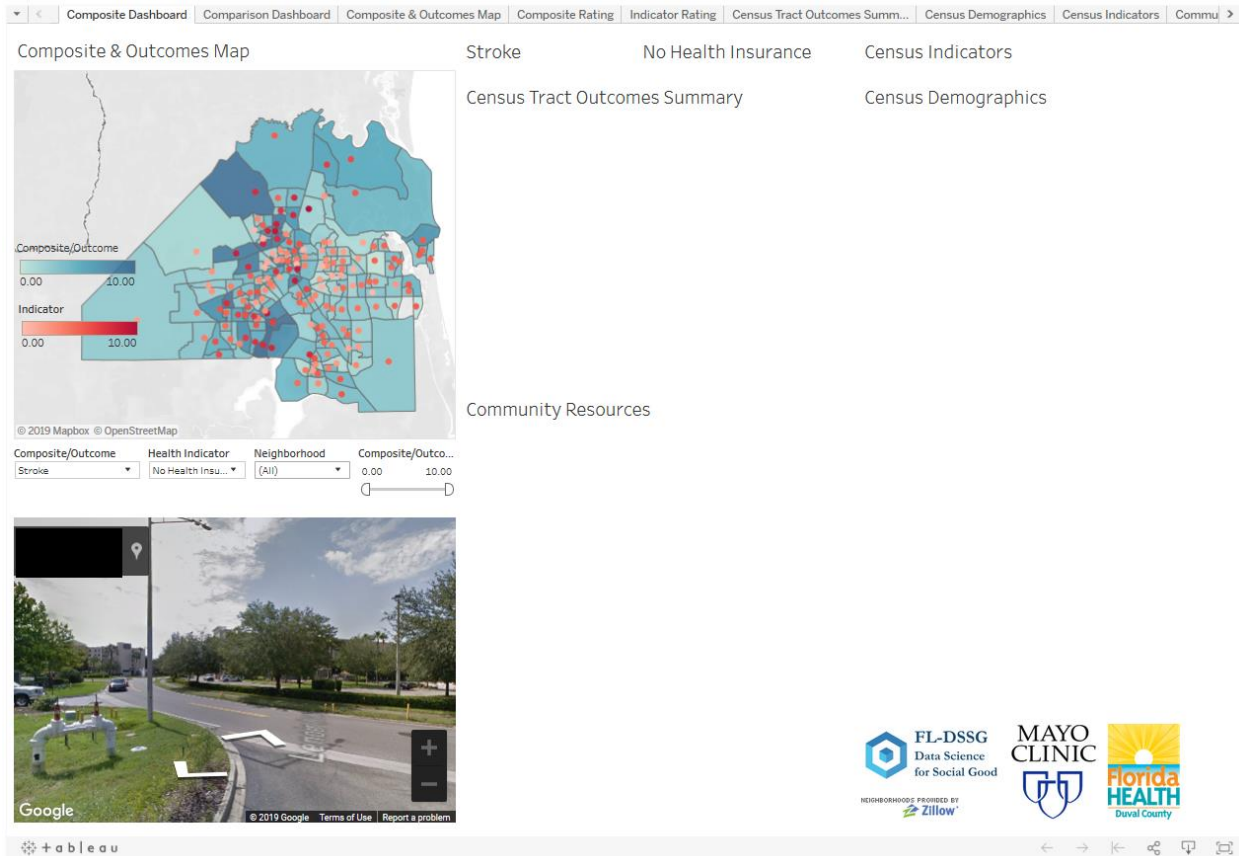

Figure 1. Composite Dashboard Landing Screen

Users can access the composite dashboard by visiting the URL, http://bit.ly/17fldssgmayotableau in the browser. We recommend users to use Firefox version 60 or above or Chrome version 60 or above browsers. Upon visiting the URL, the composite dashboard will be loaded. The left top portion of the composite dashboard shows a map of neighborhoods in the Duval county with options to select composite indexes, health indicators, neighborhood, and composite score ranges. Bottom left portion displays the Google street view of a selected neighborhood. Right top portion displays community health data available for the selected neighborhood, while bottom right portion displays community resources available for the selected neighborhood. Figure 1 shows the landing screen for the composite dashboard.

Step 2: Adjusting Composite and Outcome Options

Below the Composite & Outcomes Map in the top left section of the dashboard, users have options to select a composite index, a health indicator, a neighborhood, or a composite/outcome filter range of interest. Suggested sequence of actions is selecting a composite index, then a health indicator, and then the range for the composite filter. Figure 2 shows composite & outcomes map when Mayo Composite index is selected, figure 3 shows the map when families with children below the poverty

level are selected as the health indicator, and figure 4 shows the map when the composite score range is filtered between 7 and 10. Figure 5 shows the composite dashboard for the above-specified selections. Figure 6 shows the composite dashboard for CHNA index as the composite index, no health insurance as a health indicator, and 7 and 10 as the composite filter values. After selecting a neighborhood, users can browse and view census indicators, health summary data for the census tract selected neighborhood resides, census demographics, and listing of community resources. Thus, the composite dashboard can be used for selecting and viewing data available for a neighborhood.
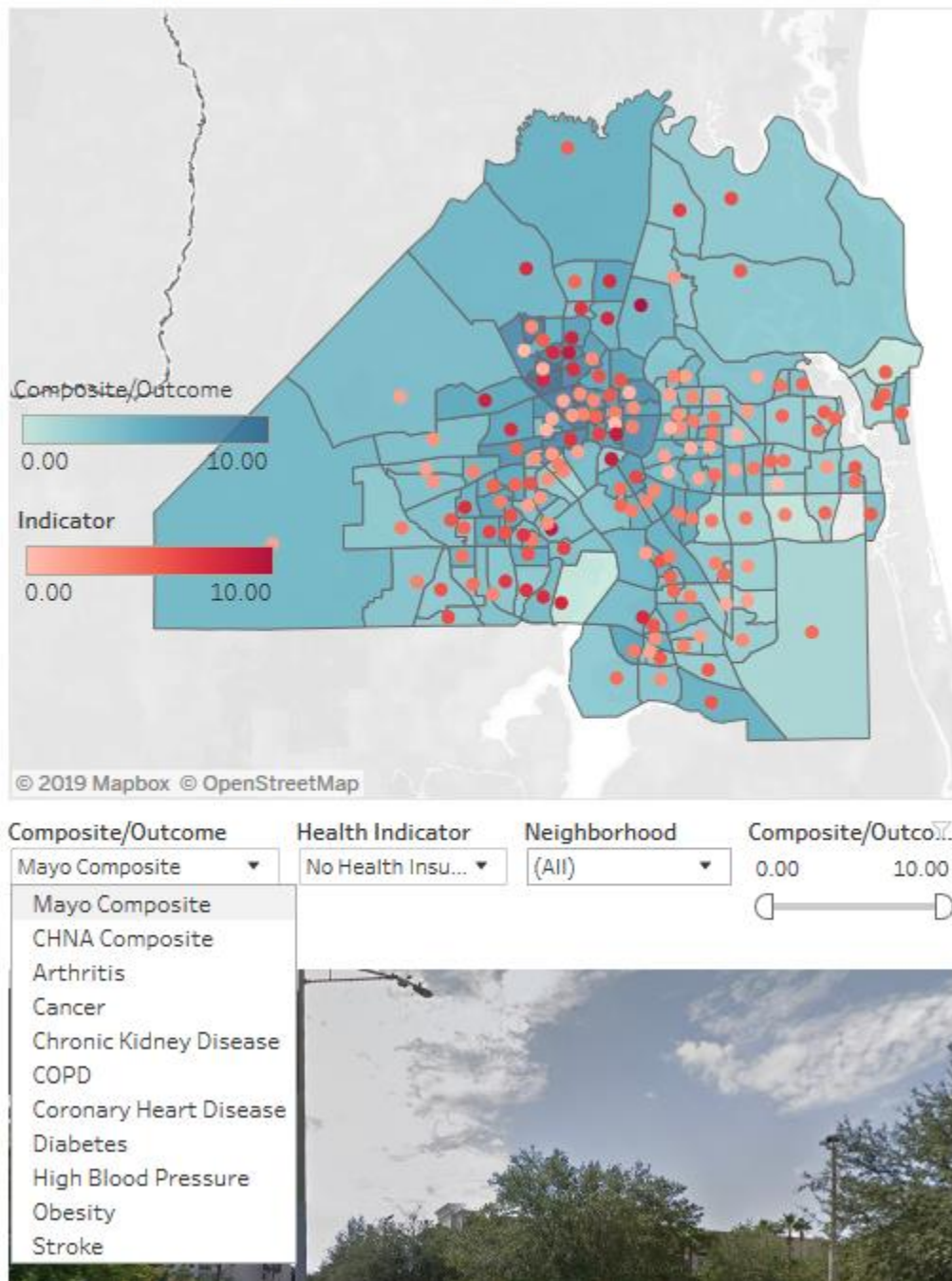
Figure 2: Composite & Outcome Index Options
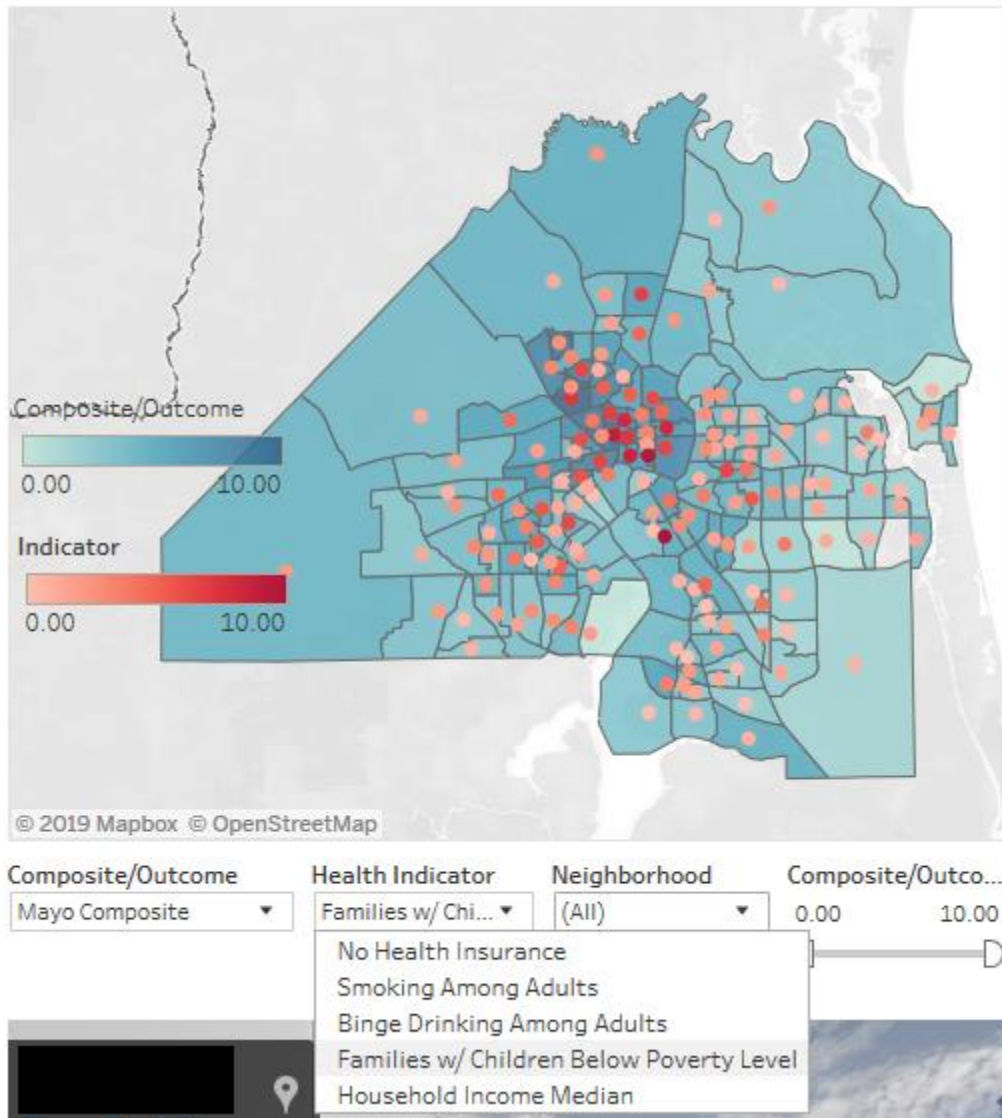
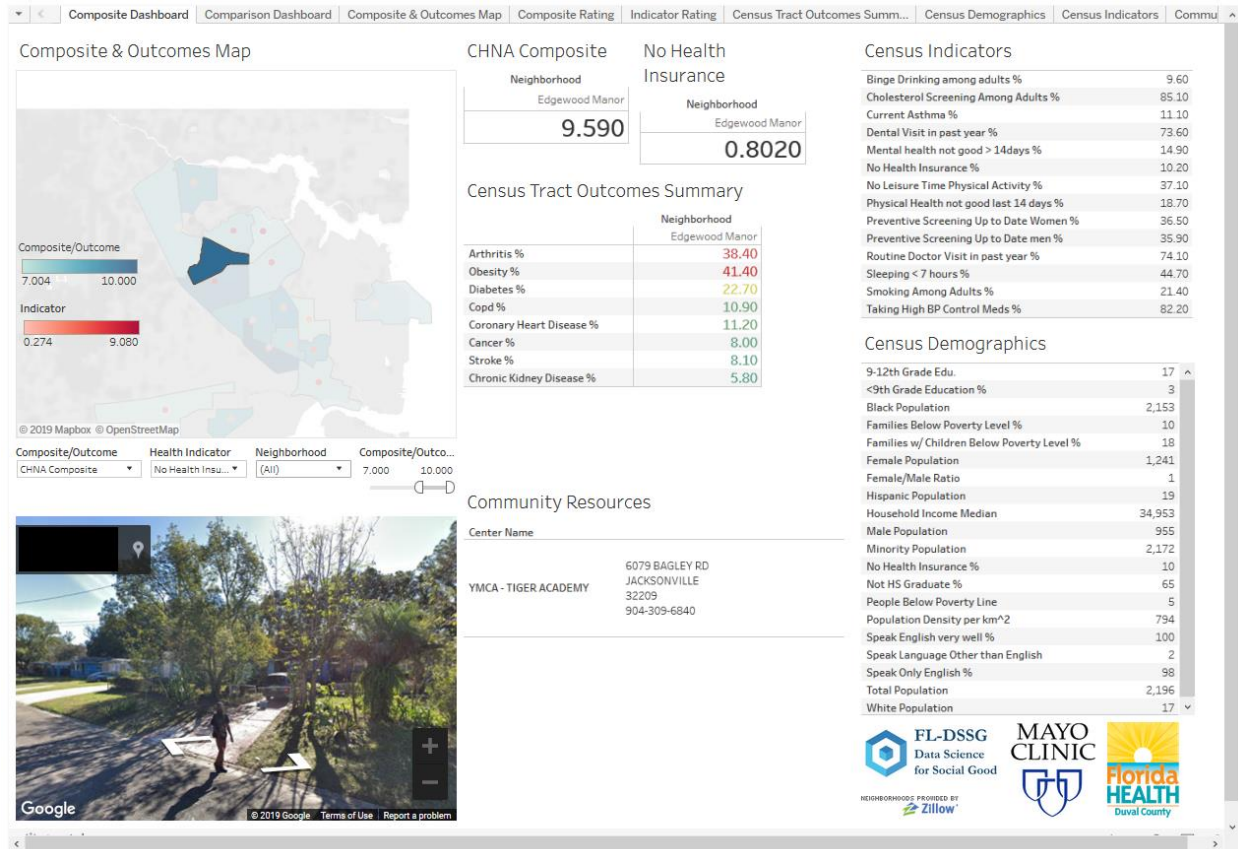**Figure 3: Health Indicator Options**
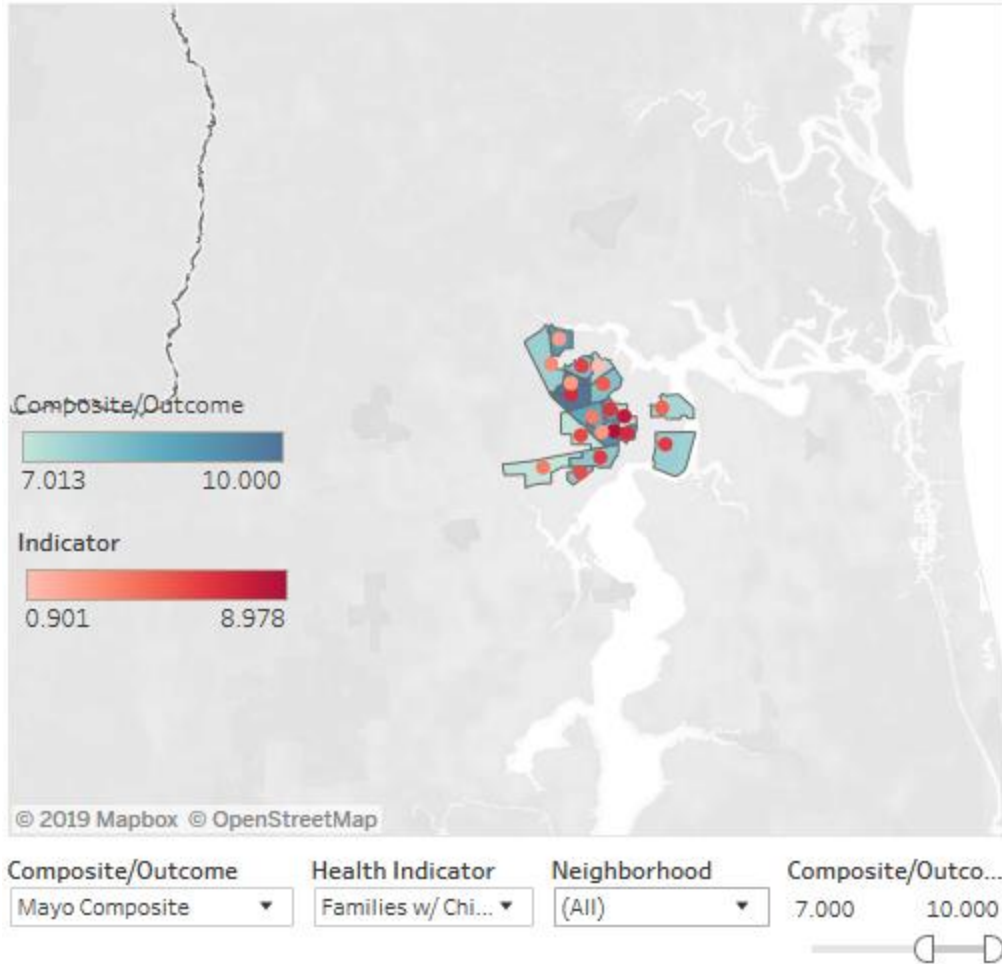
Figure 4: Composite & Outcome Filter
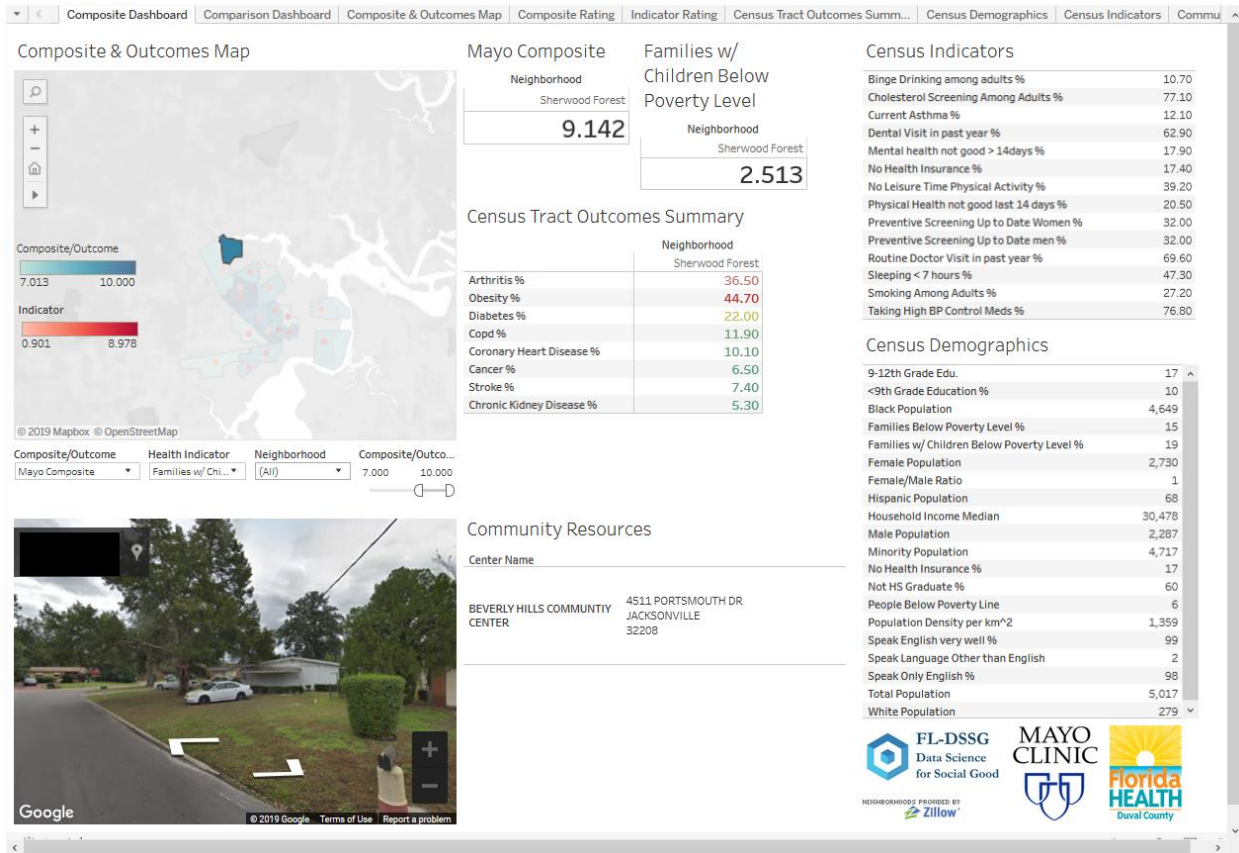
Figure 5. Composite Dashboard with Community Health Data and Resources available for a Selected Neighborhood based on Mayo Composite Index
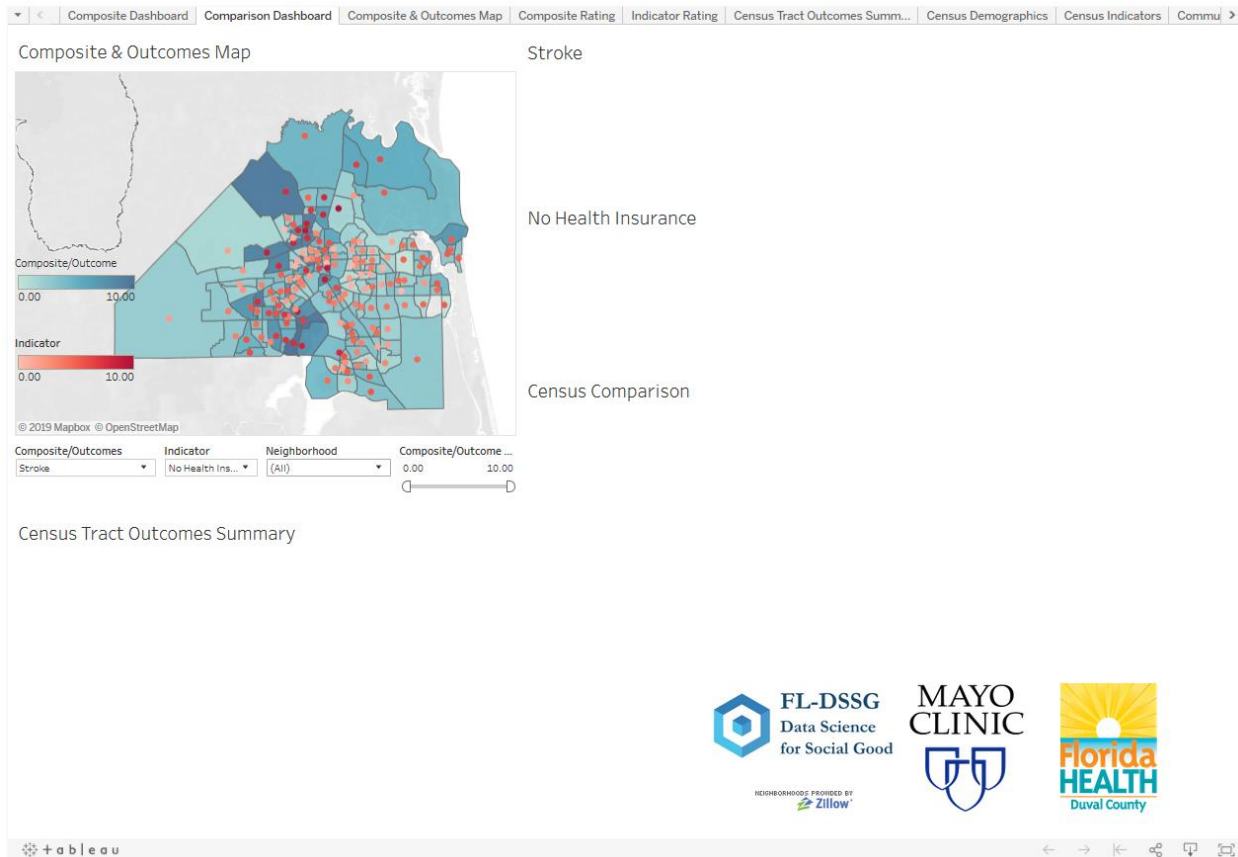
Figure 6. Composite Dashboard with Community Health Data and Resources available for a Selected Neighborhood based on CHNA Composite Index

WALKTHROUGH OF COMPARISON DASHBOARD

User task: Browse and select two or more neighborhoods for comparing based on the community health outcomes

Step 1: Open the Comparison Dashboard in the Browser

Figure 7. Comparison Dashboard Landing Screen

Users can access the comparison dashboard by visiting the URL, http://bit.ly/17fldssgmayotableau in the browser. We recommend users to use Firefox version 60 or above or Chrome version 60 or above browsers. Upon visiting the URL, the users will have to click on the comparison dashboard to load the comparison dashboard. Similar to the composite dashboard, left top portion of the comparison dashboard shows a map of neighborhoods in the Duval county with options to select composite indexes, health indicators, neighborhood, and composite score ranges. Bottom left portion displays census tract outcomes summary for the selected neighborhoods. The right portion of the dashboard displays selected composite outcomes, health indicators, and census comparison data for the selected neighborhoods. Figure 7 shows the landing screen for the comparison dashboard.

Step 2: Selecting neighborhoods for comparison

Similar to the composite dashboard, users can filter neighborhoods of interest by adjusting composite indexes, health indicators, and composite outcome options. Users can select a neighborhood by clicking on a shaded neighborhood area and can select multiple neighborhoods by pressing control and clicking on neighborhoods. Figure 8 shows an example of three neighborhoods selected for comparison.
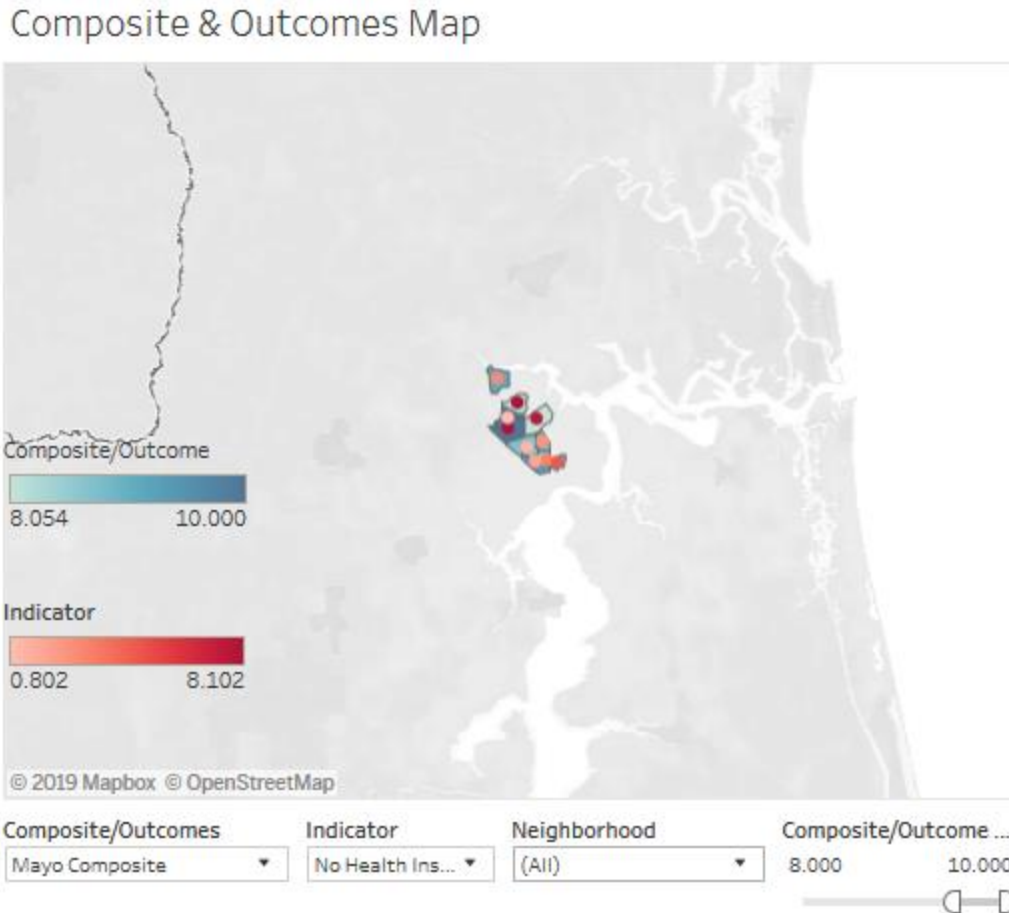
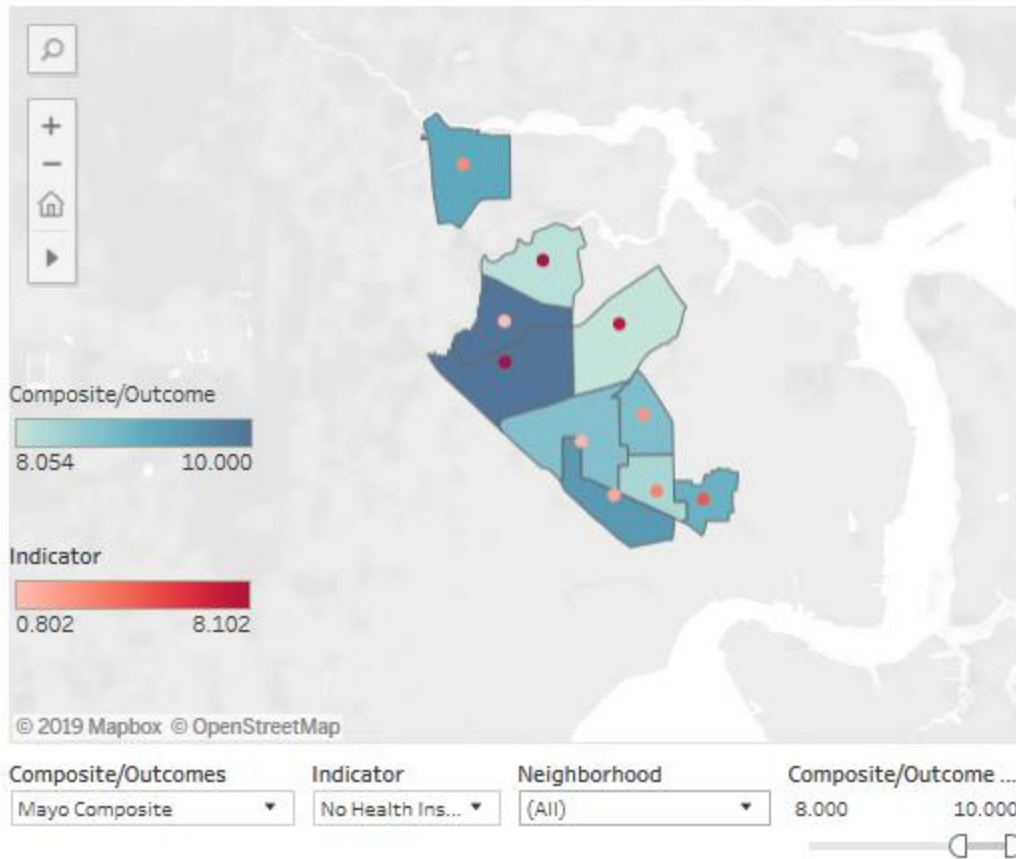**Figure 8. Neighborhoods Selected for Comparison**

Step 3: Determining a neighborhood to implement Wellness Rx program

Given that mayo composite index, no insurance as a health indicator and composite scores in the range of 8 and 10 as the manipulation values, 10 neighborhoods were filtered matching the provided options. Three neighborhoods of interest were selected for comparison. For easier selection of neighborhoods, the map can be zoomed and panned using options available on the map as shown in figure 9. Figure 10 shows a zoomed map of neighborhoods matching the filtering criteria. Users can hover the mouse over to see neighborhood name and composite score value for the selected index, as shown in figure 11. For this illustration, three neighborhoods were selected for comparison, and the resulting dashboard is displayed in figure 12. As the Wellness Rx program focusses on heart diseases, diabetes, and stroke, two of three selected neighborhood seems to be ideal for implementing the program.
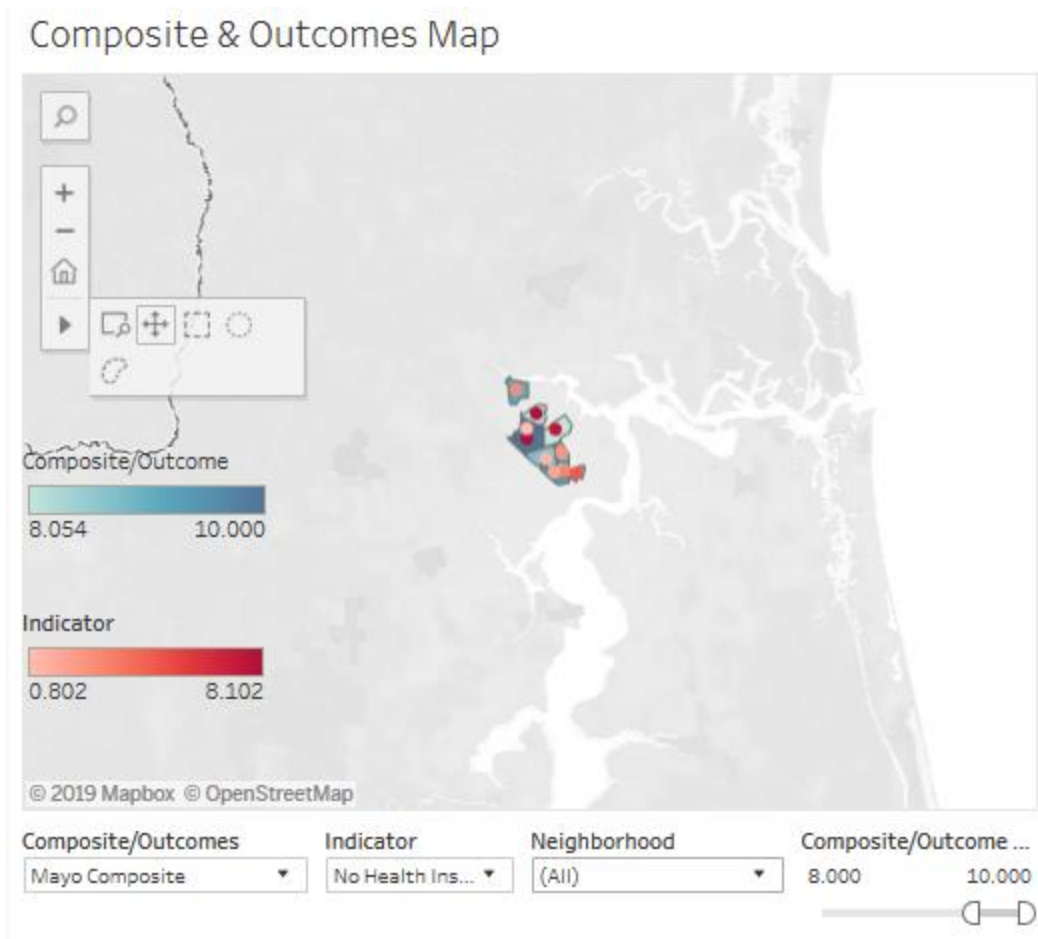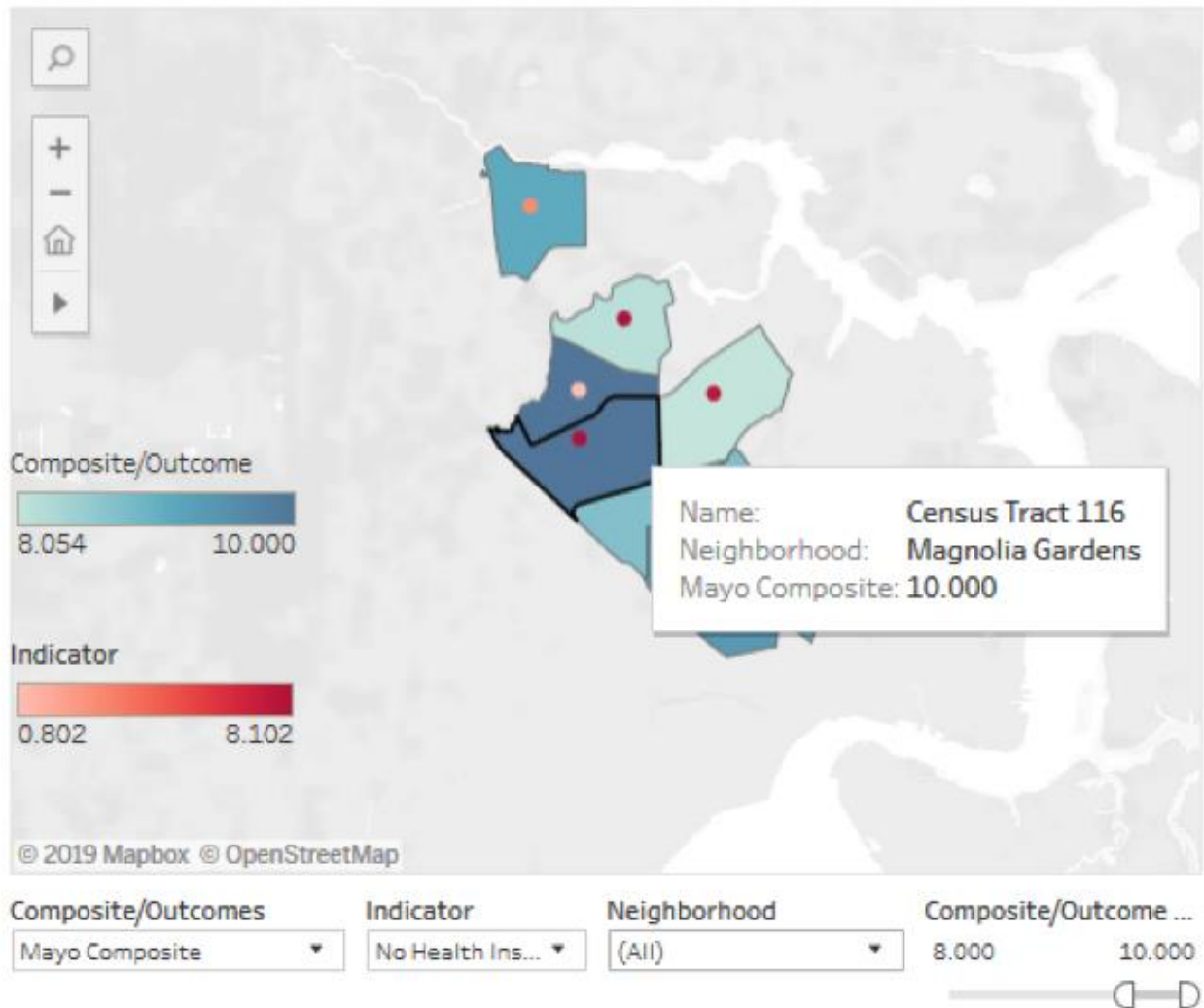
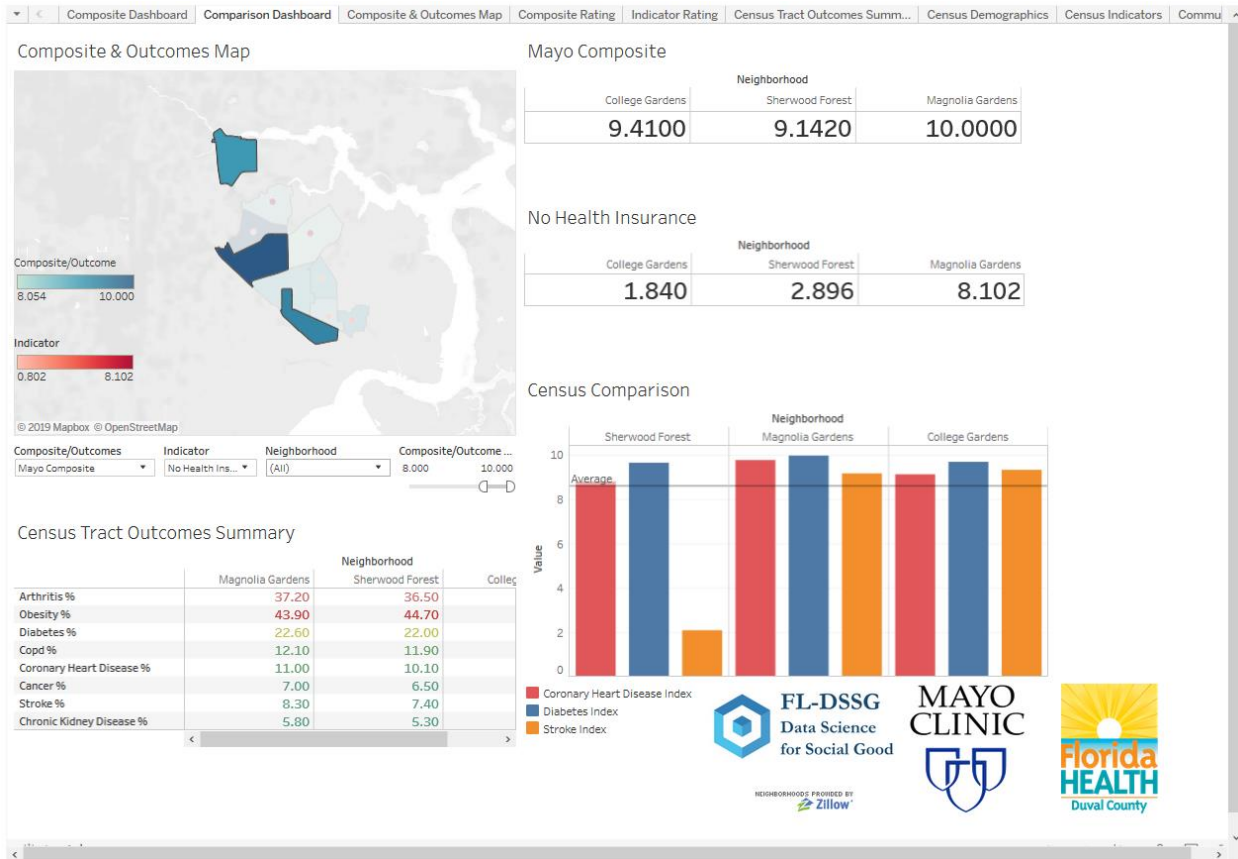**Figure 9. Options to Pan and Zoom the Neighborhood Map**

Figure 10. Zoomed view of the Neighborhood Map

Figure 11. Neighborhood Information Presented when Mouse Hovering

Figure 12. Comparison of Selected Neighborhoods

A Longitudinal Study of Google Privacy Policies

Alan Peslak
arp14@psu.edu
Information Sciences and Technology
Penn State University
Dunmore, PA 18512 USA

Lisa Kovalchick
kovalchick@calu.edu
Mathematics, Computer Science and Information Systems
California University of Pennsylvania
California, PA 15419 USA

Mauri Conforti
maconforti1@geisinger.edu
Geisinger Health System
Scranton, PA 18508 USA

Abstract

Google has come a long way since its founding by Larry Page and Sergey Brin in 1998. From a Stanford University PhD research project, to the world leader in Internet search engines, Google has changed and grown exponentially. "Google it" has become a household phrase. Its core search engine product has approximately 246 million unique US visitors per month. Its market share in the US, in search engines, is estimated at 63%. In addition, it has expanded product offerings to include the ubiquitous YouTube Internet video platform, Google Home smart speakers, Chrome web browser, Android operating system, Pixel phones, and Gmail, among other products. As a result, Google collects a tremendous amount of data from its users. With its growing popularity and the growing user privacy concerns, following the recent data breaches, Google is constantly updating its privacy policy. Our manuscript examines Google privacy policies from 2000 to present day. The policies were collected via the Internet Archive and represent a selected day from each available year from 2000 through 2018. A comprehensive qualitative, linguistics and sentiment analysis on these policies was performed. Our review finds significant similarities and differences in Google privacy policies over the years. Overall, Google privacy policies have become increasingly wordy and legalistic, but also more positive in sentiment and more personal in approach. Though the grade level of the documents has declined slightly they remain at a 12th grade level. Implications and opportunities for further research are presented.

Keywords: Sentiment analysis, Google, Linguistic analysis, LIWC

1. INTRODUCTION

Through the use of the web, people are now able to easily search for any type of information by typing a few keywords into an Internet search engine and the information they are seeking is at their fingertips. Google is the most popular search engine in the US. According to Statista (2019), in the month of December 2018, there were 246 million unique visitors to the search

engine home page. This is startling considering there were 328 million people in the US in December 2018 (Census, 2019).

Google also has an estimated 63% market share in the United States for Internet searches. In the global market, their market share is even higher, estimated at 90%. (Statista, 2019). "Seventy-four percent of U.S. adults currently say they use Google in a typical week" (Jones, 2018).

The original Google product was the online search engine, but its product lines have expanded greatly over the years and they all share their data across platforms. Their products now include YouTube, which has been accessed by 68% of the population. They have developed online advertising, and sponsored searches. (Statista, 2019). As reported by Kelly (2018), Google "is the world's largest digital advertising company". They control the Android operating system, which has an 86% market share, globally, for operating systems. They own and control Google Maps, Gmail, the Chrome web browser and Chrome OS. They have in home devices such as Chromecast, Google Home, and Google Now and Assistant. Recently, they expanded into mobile phones with Nexus and Pixel brands (Statista, 2019).

Kelly (2018) performed a detailed journalistic study of Google and found many interesting and chilling statistics; she reports that "Google collects far more data than Facebook".
Have you ever been sitting next to someone talking about making a purchase, maybe for something like a new pressure washer, then, the next time you take a look at the web, there it is, an advertisement for a pressure washer? As Kelly (2018) reports, this is not coincidence; she quotes a study by Digital Context Next which reports "Google can collect data even if you aren't using your phone. The study says that a dormant Android phone with Chrome running in the background sent location data to Google servers 340 times in one 24-hour period. ….But Google collected two-thirds of its data without any input at all from users in the researcher's experiment."

If you are concerned about your privacy and want to limit the information that Google collects, you may think this would be a simple task. All you need to do is to adjust the privacy settings in the myriad of Google products that you use. But hold on, it is not that simple. "A recent investigation by the Associated Press found that the company continued to record location data, even after a user disables the Location History option. Google said the data is used to improve services but has updated the wording of the setting to make it clear that location information is still collected." (Kelly, 2018).

Kennedy chronicled some incredible statistics in 2008, statistics which have only grown over the past decade. In 2008, Google processed 40,000 search queries each second or 1.2 trillion searches yearly, worldwide. In 2008, Google processed over 20 petabytes of data per day through an average of 100,000 MapReduce jobs spread across its massive computing clusters. A petabyte is a million gigabytes. This is a massive amount of data being collected and stored about its myriad of users.

As a result of this huge data collection, there is increasing concern over what use Google makes of this data. States attorneys general and the federal government have been investigating what and how Google stores and shares its data, and specifically, how Google shares information across all its programs and devices further reducing privacy. The 2018 Google Privacy Policy specifically states "We may combine the information we collect among our services and across your devices for the purposes described above."

With the emergence and rapid increase in Internet usage, along with the recent increase in privacy concerns due to data breaches that exposed the personal information of many Americans, some states have created recommendations with regard to online privacy policies. In 2004, California was the first state in the U.S. to enact online privacy policy legislation, the California Online Privacy Protection Act (CalOPPA), which requires commercial websites and online services to post a privacy policy. Since CalOPPA went into effect, the California Attorney General has set forth several recommendations regarding the construction of such privacy policies. One such recommendation addresses the readability of such policies; it recommends the use of "plain, straightforward language. Avoiding technical or legal jargon" (Harris, 2014). Turow, J., Hennessy, M., & Draper, N. (2018) begin to explore the problems with privacy policies and examine Americans' misunderstanding of the function of privacy policies.

In order to explore the current status of Google privacy, we need to examine the privacy policy that they post and to which they are legally

required to adhere. Our study reviews the evolution of this policy from 2000 to the present day. By examining the policy and its evolution, we hope to gain insight into how the policy has changed and what its implications are for the myriad of Google products users.

2. LITERATURE REVIEW

Over the years, a number of studies have examined the readability, content, and complexity of various website privacy policies. Graber, D'Alessandro, and Johnson-West (2002) examined 80 Internet health websites and found that these privacy policies are "not easily understandable by most individuals in the United States and do not serve to inform users of their rights." Jensen and Potts (2004) analyzed 64 website privacy policies and found that "only 6% of policies are readable by the most vulnerable of the population, and that 13% of policies were only readable by people with a post-graduate education". Proctor, Ali, and Vu (2008) examined the privacy policies of 100 websites and found that "although the readability analysis showed that a person with 13 years of education should be able to comprehend the policies, college students were able to answer correctly only about 50% of questions they were asked about specific policies".

Several studies have examined snapshots of website privacy policies for various reasons. Opsahl (2010) examined snapshots of Facebook's privacy policy over the years to demonstrate what he perceived as the disappearance of user's privacy. Warzel and Ngu (2019) examine snapshots of Google privacy policies over the past 20 years, in order to demonstrate how the Internet has changed and to attempt to understand underlying reasons for major changes to the Google privacy policy over the years.

Sentiment evaluation and linguistic analysis are commonplace techniques of studies in communication analysis. The use of word frequency and word analysis, though not perfect, is well established in the literature as a tool for corpus analysis including Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). The utilization of linguistic analysis and especially the use of a LIWC (Linguistic and Word Count) software program for research functions has been substantial. Back, Kufner, and Egloff (2011), Cordova, Cunningham, Carlson, and Andrkowski (2001), and Robinson, Navea, and Ickes (2013) all used LIWC analysis.

Holtman et al. (2018) used LIWC to study linguistic patterns of narcissism and corelated these with sports, second-person pronouns and swear words. Hawkins et al. (2017) used LIWC to study dream content.

LIWC software (Pennebaker, Booth, Boyd, and Francis, 2015) is described as such: "The way that the **L**inguistic **I**nquiry and **W**ord **C**ount program works is fairly simple. Basically, it reads a given text and counts the percentage of words that reflect different emotions, thinking styles, social concerns, and even parts of speech" (Pennebaker Conglomerates, 2015).

Overall, sentiment analysis has been an increasingly important qualitative analysis tool. As Liu (2012) defines: "Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes,". Sentiment analysis is the review of written or other forms of communication or qualitative data to determine a quantifiable and comparable measure of some form of feeling in the communication or data. Pak and Paroubek (2010) studied Twitter feeds for sentiment analysis. Pang and Lee (2008) analyzed whether textual information had a positive or negative sentiment.

Google has also been a valid and frequent subject of study in the literature. For example, Wu and Brynjolfsson (2015) studied Google Trends to predict changes in housing prices and sales. Many studies have been performed on privacy policies of Internet sites. The authors previously studied the Privacy Policies in several manuscripts (Peslak, 2016, Peslak, 2017, Peslak, 2018). After a comprehensive Google Scholar search, we could find no instances of sentiment analysis or qualitative mining of Google privacy policies in the literature.

3. METHODOLOGY

To study the Google privacy policies over time, it was first necessary to obtain past and current Google privacy policies. The current Google privacy policy was obtained from their website, and past policies were retrieved from the Internet Archive (https://archive.org/web/) also known as the Wayback Machine. As noted, we selected one dated page from each available year from 2000 to 2018 (2001-2003 were unavailable) and retrieved the archived privacy policy from Google from that date. Though there may have been years where the policy may have changed at other times within the year, we believed a once a year selection provided a

reasonable representation of the volatility of Google privacy.

Once we extracted the policies, we next needed to determine how to analyze these policies. We chose three areas of analysis based on past work and other qualitative relevant literature.

The three general areas studied include: Overall Content, Specific Word and Key Word Content, and Sentiment and Linguistic Analysis. To analyze Overall Content, we utilized several tools; Microsoft Word was used to determine reading grade level, complexity, and word count. For sentiment and linguistic analysis, two tools were used. LIWC was used to determine key variables over time including clout, analytic, tone, and authenticity. IBM Watson Sentiment Analysis was used for sentiment (positive/negative) evaluation to determine degree of positive and negative content. For keyword and other specific content, Voyant Tools was utilized, as well as, specific author reviews of each policy. Another tool used in the study was Microsoft Excel for charting and other analyses.

In addition, we imported the privacy policies into LIWC (Linguistic and Word Count). LIWC software produces unique measures for linguistic analyses. For the most part these are expressed by a percentage of total words mapping to the dictionary category of each measure. The exceptions are several relating to word counts, as well as, calculated emotional measures. Appendix 1 includes all the measures, including LIWC variables used. Analytic reflects logical thinking versus narrative, authentic reflects honest versus guarded, and tone reflects upbeat versus sad.

4. RESULTS

The first finding of our study is that the Google privacy policy has become more difficult to read. One measure of complexity is the reading grade level of the policies. As Figure 1 depicts, Google has always had a high reading grade level for its privacy policies. (Note that all raw data in charts is shown in Appendix 1). Although reading grade level decreases slightly over the years, the expectation is still at a 12th grade level or above. According to the Clear Language Group (n.d.) "For the general public, text should be written at the 8th grade level or lower". An example of online documents with much better readability index come from a study by Leroy et al. (2008), who found online consumer health sites have a readability index of 10.5. Google's own guide to "Finding your way around YouTube" instructions

only requires a Flesch-Kincaid reading grade level of 9.3 (Google, 2019). The current Google privacy policy now includes video explanations and additional imagery; thus it seems that Google may be attempting to make the content of its privacy policy more accessible to the general public.
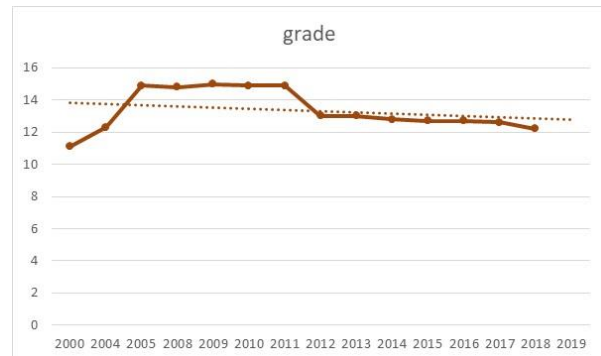

Figure 1. Reading Grade Level

Our readability findings are aligned with previous findings. Protcor et. al. (2008) studied the privacy policies of 100 different websites and found that the policies were at a 13th grade reading level. Jensen and Potts (2004) examined 64 website privacy policies and found that the average reading grade level was 14.15. Graber et. al. (2002) studied the privacy policies of 80 Internet health websites and found that, of the sites that had a privacy policy, the average readability level was that of grade 14.

The increasing complexity of the Google privacy policy over the years is shown by an ever-increasing word count (see Figure 2). In 2000, the word count of the privacy policy was 657, with just over one page of information. In 2009, the word count increased to 2,140, with four pages. Multiple new sections were added since 2000, such as Choices for Personal Information, Information Sharing, Information Security, Data Integrity, and Accessing and Updating Personal Information, and Enforcement. Also, in 2000, sections included What Information Do We Collect and the "Google and Cookies" sections had three paragraphs that expanded to fourteen paragraphs in 2009, under the heading Information We Collect and How We Use It.
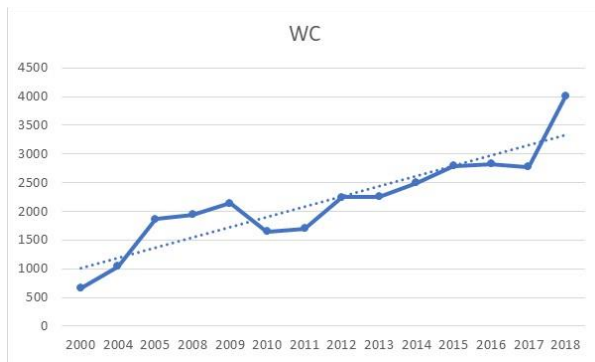
Figure 2. Word Count

In addition, the 2009 policy added uses for personal information including: Customized Content and Advertising, Auditing, Research and Analysis, Ensuring Technical Function, Protecting the Rights or Property of Google or Our Users, and Developing New Services. There is also a notice that Personal Information can be processed in the United States or on servers in other countries.
Choices for Personal Information in 2009, include: Your Consent and Changes to the Privacy Policy.

Google acquired the Internet advertising company DoubleClick in 2008 and shortly afterwards, the Information Sharing section of their privacy policy increased from one paragraph to four paragraphs previously titled, With Whom Does Google Share Information? Additional new information includes:
- Require consent to share personal information
- Personal information can be used by trusted businesses on Google's behalf.
- Legal requests, terms of service enforcement, security reasons.
- Can be transferred if a merger or sold after notice is given.
- Third party sharing will not identify who you are, just your interest.

Several other sections were added in 2009, including the following. The Information Security section describes security measures in place, to restrict unauthorized access to personal data. The Data Integrity section describes how personal information is used according to the privacy policy and reviewed for accuracy; users must update their own information, when needed. The Accessing and Updating Personal Information section describes your access to personal information and the procedures used to correct or delete it. The Enforcement section, previously titled Who Can I Ask if I Have

Additional Questions, informs users how to contact Google with questions or concerns and how they reply.

In 2010, the word count decreased to 1,652. Sections condensed include: Introduction, Choices for Personal Information's title was changed to Choices, and Changes to this Privacy Policy. The omitted items were Gadgets, Links, and Data Integrity. A Unique Application Number section was added. Some external services identify you with this ID, but this number is not linked to your Google personal information.

In 2018, the word count reached 4,009, the peak word count of all years. This increase came as Google re-wrote its privacy policy, in response to the Europe Union's (EU) General Data Protection Regulation (GDPR). "Simply put, the GDPR mandates a baseline set of standards for companies that handle EU citizen's data to better safeguard the processing and movement of citizens' personal data" (De Groot, 2019).

New sections that were added include the following (*in italics*). *Why Google Collects Data*, which states this is done to provide, customize, and deliver better services.

Your Privacy Controls, which defines controls the user can access to manage, review, and adjust privacy settings, such as links to Privacy Check Up and Product Privacy Guide.

Compliance & Cooperation with Regulators, which is reviewed frequently, to ensure compliance. Even though servers may be outside the country with different protection laws, Google provides the same protection, no matter where the server is located.

About This Policy, which explains that all the services offered by Google are covered by the privacy policy, such as YouTube and third-party sites. However, the privacy policy does not apply to the practices of other companies or vendors.

Related Privacy Practices, which added 18 links for additional information about Google privacy notices.

Sharing Your Information, which is on the 2010 policy with less content, than 2018. One of the new sections pertains to Domain Administrators. This policy is different, since it refers to students and employees under an organization that uses Google services, instead of individual users.

Next, we examine the analytic score of the various Google privacy policies. Pennebaker et. al. (2015) state that a high analytic score indicates formal, logical and hierarchical thinking; whereas, a low analytic score indicates more informal and personal thinking. As shown in Figure 3, a major decrease in the analytic score was found from 2011 to 2012, where a more personalized experience emerged. Google began using the words "you", "your", and "our users" to give a more individual-centered experience.
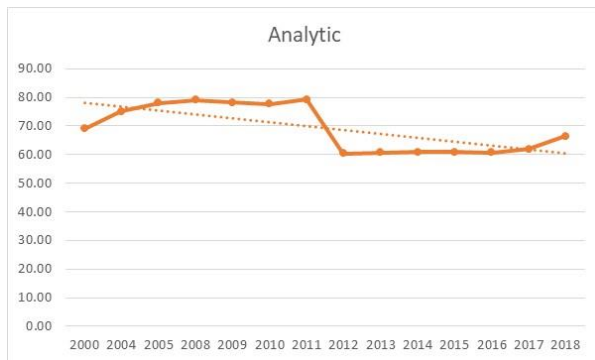
Figure 3. Analytic Scores

A re-worded 2012 introduction personalizes their services, products, and websites.
- "We strive to develop innovative services to better serve our users"
- "We recognize privacy is an important issue, so we design and operate our services with the protection of your privacy in mind"

Several sections were added in the How We Use Information We Collect section giving a more individualized experience.
- "We use the information we collect ……to offer you tailored content – like giving you more relevant search results and ads."
- Information from cookies and technologies like pixel tags is used "to improve your user experience and the overall quality of our services".
- Google combines personal information from its services "to make it easier to share things with people you know".

In 2018, the analytic rating finished slightly below where it began on the chart in 2000. The personalized tone shown in 2012 gradually changed to a more legalistic approach with less personal touch by 2018.

According to Pennebaker et. al. (2015), a high authenticity score indicates a more honest and personal text; whereas, a low authenticity score indicates a more guarded and impersonal text. As shown in Figure 4, authenticity scores decreased significantly over the 18-year span except for one year, with a more guarded, legalistic, less personal approach. In 2000, the score was 48.0, in 2004, it increased slightly to 51.1, in 2018, the most recent privacy policy, the score fell to 19.4. In 2004, they added some verbiage using "our" and "your" in the introduction and the addition of "we design and operate our services with the protection of your privacy in mind." This most likely accounts for the slight increase in the numbers. In 2012, authenticity dropped slightly, with the addition of the following sections: Transparency and Choices, Information Security, and Application of the Privacy Policy.
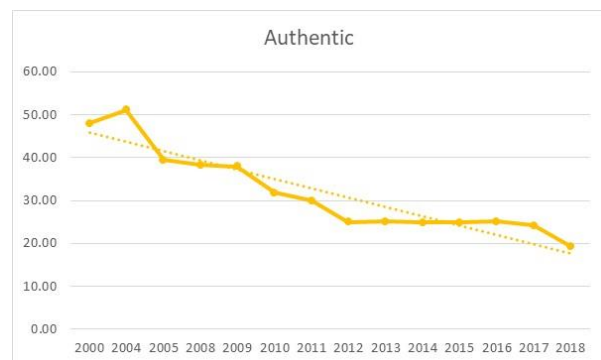

Figure 4. Authentic Scores

In 2018, as a result of the re-write to address the EU's GDPR, several new sections were added addressing security concerns, which most likely affected the authenticity score. Your Privacy Controls describes the controls for privacy management on Google. Links in the introduction are Privacy Checkup, Product Privacy Guide, and Google Account to Review and Update Information. Additional links included are: Activity Controls, Ad Settings, Information You Share, My Activity, Google Dashboard, Your Personal Information About You, Shared Endorsements and Export Your Data. In this section alone, in the content, 2018's policy has 26 links to assist in controlling your privacy compared to 0, in 2000 and 11, in 2010. Exporting & Deleting Your Information describes controls allowing you to export or delete some or all of your data. There are measures in place to protect the information you are deleting from being maliciously deleted. Related Privacy Practices has 18 additional links for access to more specific resources on how Google practices and their privacy policies such as Chrome, Payments, Privacy Checkup,

Google's Safety Center, Technology and Principles, and How Google Uses Data When You Use Our Partners Sites or Apps, among others.

Over the years, the sentiment of Google privacy policies has shown a positive, friendlier, more enjoyable tone. Figure 5 shows the increase in sentiment based on IBM Watson. The exception is the 2018 privacy policy sentiment score, which decreased after 8 years of consistent increases. The highest increases are from 2000 to 2004. In 2004, specific positive passages were added:

- "when we require personally identifying information, we will inform you about the types of information we collect and how we use it."
- "we hope this will help you make an informed decision about sharing your personal information with us."
- "we may share the information submitted … in order to provide you with a seamless experience"
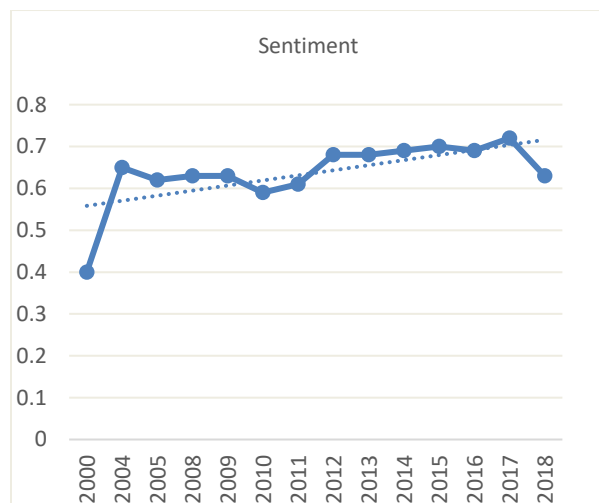

Figure 5. Sentiment Scores

From 2017 to 2018, there is a slight decrease in sentiment taking the sentiment to the side of a more negative experience.

The following 2017 passages that give a positive, friendly tone were removed in the 2018 policy.

- Using Google services with your information "we can make those services even better"
- "we tried to keep it as simple as possible, but if you're not familiar with terms like cookies, IP addresses, pixel tags and browsers, then read about them first. Your privacy matters to Google so whether you are new to Google or a long-time user,

please do take time to know our practices – and if you have any questions, contact us."

A common tool for visualizing large text material is a word cloud. A word cloud shows the relative frequency of words in a document by the size of the word, itself. The word clouds for the collected Google privacy policies of 2000 and 2018 are displayed in Figure 6 and 7, respectively.
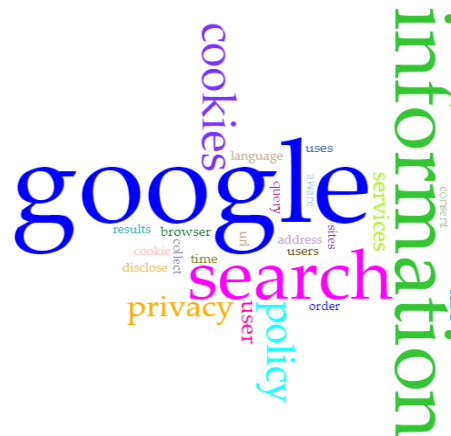

Figure 6. 2000 Google Privacy Word Cloud

Figure 7. 2018 Google Privacy Word Cloud

A review of the two word clouds shows large differences between the original and current privacy policy. Initially, cookies frequently appeared in the policy; however, in 2018, the term cookies, has much less prominence. Information and services dominate in 2018; whereas, in 2000, Google and search were very prominent. Use and account also become much more prominent in 2018, reflecting that detailed description of use of data and the inclusion of an account, in the recent policy.

Another fertile area for data visualization is network graphs that show collocation of words.

These are readily available via Voyant Tools. A "Collocated Graph represents keywords and terms that occur in close proximity, as a force directed network graph." Keywords are shown in blue and collocated words (words in proximity), are shown in orange.

Figures 8 and 9 depict the collocated graphs for the collected Google privacy policies for the years of 2000 and 2018, respectively.
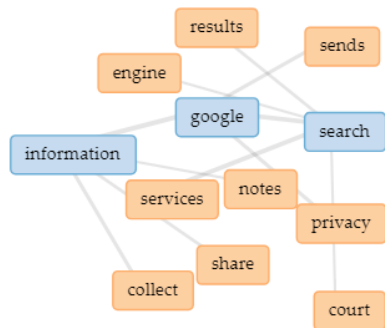

Figure 8. 2000 Collocation Graph
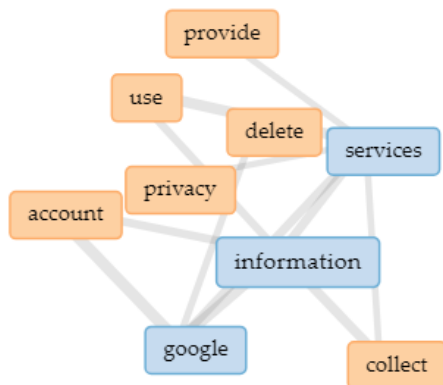
Figure 9. 2018 Collocation Graph

In 2000, the keywords were Google, information, and search. Google Collocated words were Google sends, Google privacy, and Google privacy court. With the keyword Information, collated words were Information collect, Information share, and Information notes. Search engine, Search results, Search privacy, Search privacy court, and Search services are the collocated Search words.

In 2018, we have an entirely different graph. Search is no longer a keyword. This, perhaps, suggests the de-emphasis on just the Google search engine, as their product offerings have expanded. Instead of search, we now have Services as a keyword. We also have dropped Share as a word in the graph; it seems to have been entirely replaced by use. Some new Google collocates are: Google account, Google delete, Services provide, Services collect, Information privacy, Information account, and Services collect, use, delete.

The final chart, displayed in Figure 10, comes from Voyant tools and is a charting of the five fair information practices and their inclusion in the privacy policies over the years. Though terms do not show the entire story, they do provide some measure of how organizations' privacy policies are easily mappable to Federal Trade Commission Fair Information Practices. The Y axis is shown as relative frequency, to adjust for the increasing word count over the years. In 2000, the FTC (2000) published a document entitled "Fair information practices" and suggested a voluntary standard for privacy policies, recommending what they should contain. The 5 areas to be included were: access, security, notice, choice, and enforcement. Over the years, the inclusion of these terms has varied widely for Google privacy policies. Notice, as a specific term was important in early years, but has dwindled significantly over the years. Security was initially not mentioned but has been moderately included all other years. Access was not noted until 2004 but rose strongly and has remained the most frequent term, in most years. Choice was first included in 2010 but has low mentions, in most years. Finally, enforcement, as a term, started in 2005 and was included until 2014, when it was dropped altogether.
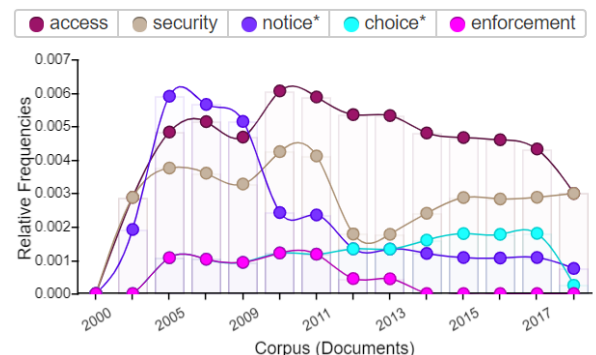
Figure 10. Fair Information Practices

5. SUMMARY AND CONCLUSIONS

In this paper, the authors have examined Google privacy policies from 2000 to 2018. We have performed a detailed review through a variety of

qualitative and data visualization tools. The results can be used by students, faculty, practitioners, and researchers, to understand the evolution of Google privacy over time. The study can also be used as a basis for comparing other privacy policies. The study can serve as a model for comparing any written documents for similarities and differences.

Some of the key findings from this study include:

- The complexity of the Google policy has increased over time, due to a more than sixfold increase in word count, as well as, the continuation of a 12th grade reading level.
- The Google privacy policy has evolved to a more personalized document, suggesting an attempt to make the document appear friendlier.
- The policy sentiment also improved over the years using more positive and enjoyment laden words; however, this is offset by a 50% decrease in authenticity score, which actually depicts a more legalistic and guarded content.
- There were no signficiant differences in tone over the period.
- Google has further obfuscated their privacy policy, by adding links to other pages. For example, the 2018 policy has 26 links to assist in controlling your privacy compared to 0, in 2000, and 11, in 2010.
- The word clouds have similarities between 2000 and 2018, but also have many key differences. Services supplanted search and use became a very frequent used word. Cookies virtually were elimiated as a keyword, by 2018.
- Collocation network graphs show differences in keywords and words in proximity for 2000 and 2018. Some of these differences include importance of services in 2018 and the rise of use, delete, and account in 2018; along with the elimination of share as a significant collocated word from 2000 to 2018.
- Of the five fair information principles, access is the most frequently mentioned, followed by security, choice, and notice. Enforcement has been dropped from the Google privacy policy.

Overall there has been great change in Google privacy policies over the past two decades. The greatest takeaway for the authors is that, even though the policy gives the appearance and words associated with friendlines and positive sentiment, the policy is actually highly legalistic and allows for cross platform sharing and using of data for primarily any purpose. As noted by Kelly, the information we provide to Google allows the organization to track our day from beginning to end.

Although websites such as Google offer their services to users free-of-charge, these services actually do come at a cost to the user. This cost is realized in the collection of user data, which is utilized by advertisers and marketers, and other groups. Google's free services come at a high price for privacy. Additional research is proposed to further explore these polcies and their full impact on consumers, users, and society, at large. The authors plan to further pursue this fertile area with future studies.

6. REFERENCES

Back, M. D., Küfner, A. C., & Egloff, B. (2011). Automatic or the people? Anger on September 11, 2001, and lessons learned for the analysis of large digital data sets. *Psychological Science*, *22*(6), 837-838.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, *28*(2), 15-21.

Clear Language Group (n.d.). Readability. Retrieved from http://www.clearlanguagegroup.com/readability/

Cordova, M. J., Cunningham, L. L., Carlson, C. R., & Andrykowski, M. A. (2001). Social constraints, cognitive processing, and adjustment to breast cancer. *Journal of consulting and clinical psychology*, *69*(4), 706.

De Groot, J. (2019, July 15). What is the General Data Protection Regulation? Understanding & Complying with GDPR Requirements in 2019 [Blog post]. Retrieved from https://digitalguardian.com/blog/what-gdpr-general-data-protection-regulation-understanding-and-complying-gdpr-data-protection

Federal Trade Commission. (2000). *Privacy Online: Fair Information Practices in the Electronic Marketplace, A Report to Congress*. Available at http://www.Federal Trade Commission.gov/reports/privacy2000/privacy2000.pdf

Google (2000 -2018) Google Privacy Policy.

Google (2019) Find your way around YouTube https://support.google.com/youtube/answer/2398242?hl=en&ref_topic=9257409

Graber, M.A., D'Alessandro, D.M., & Johnson-West, J. (2002). Reading level of privacy policies on Internet health Web sites. Journal of Family Practice, 51, 642-645.

Harris, K.D. (2014, May). Making your privacy practices public. Retrieved from https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf

Hawkins, I. I., Raymond, C., & Boyd, R. L. (2017). Such stuff as dreams are made on: Dream language, LIWC norms, and personality correlates. *Dreaming*, *27*(2), 102.

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C., große Deters, F., ... & Mehl, M. R. (2018). Linguistic Markers of Narcissism: An Exploratory LIWC Analysis of 15 Samples. *PsyArXiv. August 8*.

Hutchinson, Baird, and Garg (2016) Federal Trade Commission. (2000). *Privacy Online: Fair Information Practices in the Electronic Marketplace, A Report to Congress*. Available at http://www.Federal Trade Commission.gov/reports/privacy2000/privacy2000.pdf

Jensen, C., & Potts, C. (2004). Privacy policies as decision-making tools: An Evaluation of online privacy notices. CHI 2004, 6, 471-478.

Jones, J. (2018, April 11) Facebook Users' Privacy Concerns Up Since 2011. Retrieved from https://news.gallup.com/poll/232319/facebook-users-privacy-concerns-2011.aspx

IBM (2019) IBM Watson Natural Language Understanding https://www.ibm.com/watson/services/natural-language-understanding/

Internet Archive (2019) https://archive.org/web/

Kelly H. (2018, August 21). Google's data collection is hard to escape, study claims https://money.cnn.com/2018/08/21/technology/google-data-collection/index.html

Kennedy, N. (2008) Google processes over 20 petabytes of data per day https://www.niallkennedy.com/blog/2008/01/google-mapreduce-stats.html

Leroy, G., Helmreich, S., Cowie, J. R., Miller, T., & Zheng, W. (2008). Evaluating online health information: Beyond readability formulas. In *AMIA Annual Symposium Proceedings*(Vol. 2008, p. 394). American Medical Informatics Association.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Microsoft Word Readability Statistics (2019) Microsoft Corp.

Opsahl, K. (2010, April 28). Facebook's Eroding Privacy Policy: A Timeline. *Electronic Frontier Foundation*. Retrieved from https://www.eff.org/deeplinks/2010/04/facebook-timeline

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, *2*(1-2), 1-135.

Pennebaker Conglomerates (2015) *LIWC How It Works*. http://liwc.wpengine.com/how-it-works/

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015*). Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Peslak, A. (2018). Facebook Fanatics: A Linguistic and Sentiment Analysis of the Most "Fanned" Facebook Pages. Journal of Information Systems Applied Research, 11(1) 23-33.

Peslak (2017). Sentiment Analysis and Opinion Mining: Current State of the Art and Review of Google and Yahoo Search Engines' Privacy Policies V10 N3 Pages 38-47 December 2017 Journal of Information Systems Applied Research

Peslak, A. R. (2016). Do Social Networking Sites Privacy Policies Differ? A Linguistic Analysis Of The Ten Most Popular Social Networking Sites *Issues in Information Systems* (17), 7.

Proctor, R.W., Ali, M.A., & Vu, K.L. (2008). Examining usability of web privacy policies. *International Journal of Human-Computer Interaction*, 24(3), 307-328.

Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, 0261927X13476869.

Sinclair, Stéfan and Geoffrey Rockwell, 2016. *Voyant Tools*. Web. http://voyant-tools.org/

Statista (2019) Google - Statistics & Facts https://www.statista.com/topics/1001/google/

Turow, J., Hennessy, M., & Draper, N. (2018). Persistent misperceptions: Americans' misplaced confidence in privacy policies, 2003–2015. *Journal of Broadcasting & Electronic Media*, *62*(3), 461-478.

US Census (2018) National and State Population Estimates https://www.census.gov/newsroom/press-kits/2018/pop-estimates-national-state.html

Voyant Tools (2019). http://www.voyant-tools.org

Warzel, C., & Ngu, A. (2019, July 10). Google's 4,000-word privacy policy is a secret history of the Internet. The *New York Times*. Retrieved from https://www.nytimes.com/interactive/2019/07/10/opinion/google-privacy-policy.html

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In *Economic analysis of the digital economy* (pp. 89-118). University of Chicago Press.

Editor's Note:

This paper was selected for inclusion in the journal as an CONISAR 2019 Distinguished Paper. The acceptance rate is typically 7% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2019.

Appendices
Appendix 1. Word Count, LIWC variables, Grade Level

Year	WC	Analytic	Authentic	Sentiment	Grade Level
2000	657	69.02	48.00	0.4	11.1
2004	1044	75.11	51.12	0.65	12.3
2005	1866	77.98	39.47	0.62	14.9
2008	1947	79.04	38.25	0.63	14.8
2009	2140	78.24	37.93	0.63	15
2010	1652	77.75	31.81	0.59	14.9
2011	1703	79.23	29.97	0.61	14.9
2012	2246	60.35	25.04	0.68	13
2013	2256	60.80	25.11	0.68	13
2014	2499	60.86	24.85	0.69	12.8
2015	2789	60.88	24.83	0.7	12.7
2016	2829	60.75	25.06	0.69	12.7
2017	2779	62.00	24.21	0.72	12.6
2018	4009	66.42	19.38	0.63	12.2