

# Venice Big Data

## Visualizing the Continuum of Data in Venice

An Interactive Qualifying Project

Submitted to faculty of

***Worcester Polytechnic Institute***

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

***Submitted By:***

Joseph Bremner

Austen Couvertier

Frank Egan

***Sponsoring Agencies:***

Reset Venezia

Venice Project Center

***Submitted To:***

Fabio Carrera

N. Aaron Deskins



Date:

<https://sites.Google.com/site/ve16bigdata/>

[ve16-bigd@WPI.EDU](mailto:ve16-bigd@WPI.EDU)

## Authorship

The authors of this report are Joseph Bremner, Austen Couvertier, and Frank Egan.

The specified WPI students were responsible for working on equal parts of this report.

## Acknowledgements

Before expanding upon the specifics of our process, we would like to extend our gratitude to those who have provided us with invaluable aid in designing and completing our project. First and foremost, we would like to thank our ID 2050 instructor, Prof. Melissa Butler, for her extensive reviews on our proposal and for pushing us to design our project with an open mind. We would additionally like to thank Laura Robinson, our project librarian, for her guidance in using our resources efficiently and effectively. We also must thank our sponsor Emanuele Dal Carlo from Reset Venezia, for his interest in ensuring that our efforts are well met and for his crucial advice related to Venetian culture and its relation to technology. We are in debt to several employees of the Venice Project Center, namely Tomaso Minelli, Andrea Toffanello, Sarah Puccio and Piero Toffolo, for being continuously on call when we were in need of help, and for providing technical and logistical aid to us that we surely would have faltered without. We also must express our utmost gratitude to our advisor, Prof. Aaron Deskins, for his enthusiasm and positivity as well as his unwavering advice on polishing our report. Lastly, we need to thank our advisor Prof. Fabio Carrera for being a wealth of inspiration and ideas behind the Venice Project Center, and for having the best interest of the city of Venice at heart. Without the help of the figures mentioned, we certainly would not have been able to attain the level of success that we were able to in completing our project.

## Abstract

The Venice Project Center (VPC) Big Data team's mission was to allow locals, tourist and researchers more interactivity with large datasets related to Venice. Deliverables included: an improved real-time Venice City Dashboard, a new website that presents visualizations of long-term trends, and two new big data visualizations utilizing data from the VPC and Venice Open Data Project (VOD) site. Each deliverable served to allow the potential users greater access to complex data sets and ultimately more knowledge about the city they inhabit.

## Executive Summary

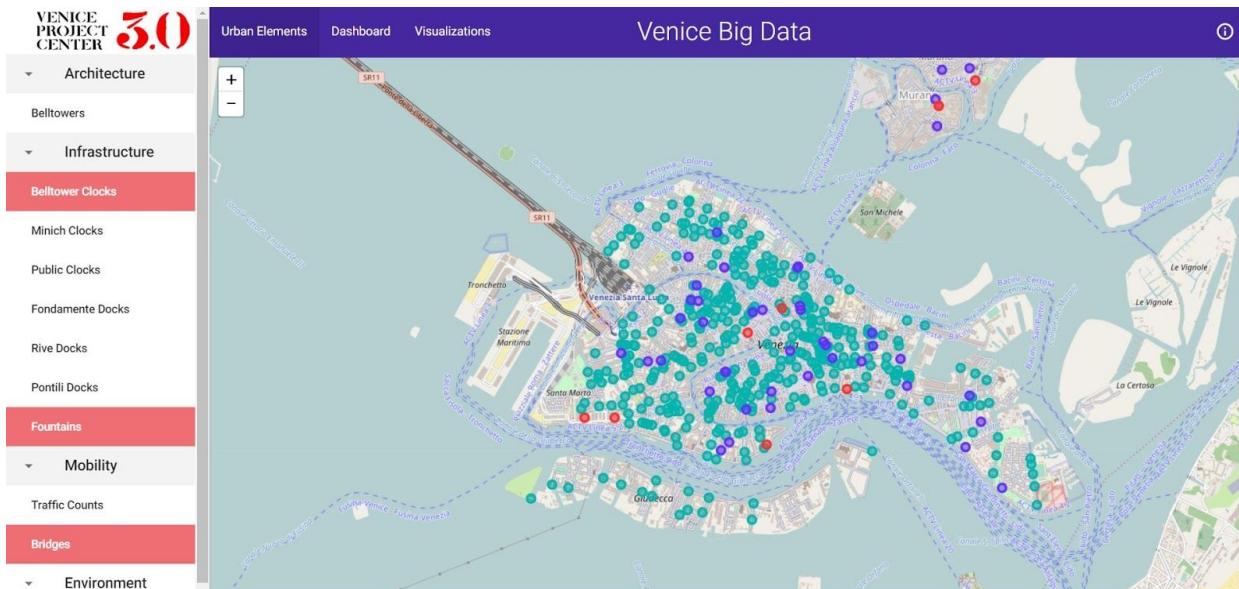
The mission of our project, with the help of the Venice Project Center (VPC) and Reset Venezia, was to give more power to the locals with regards to deciding the future of Venice. Since the introduction of technology and the beginning of data collection, knowledge can be gained by understanding data that has been collected for years. For Venice, these repositories of data knowledge are primarily the Venice Open Data Project site (VOD) and the VPC itself. The VOD currently has 142 datasets across 6 administrations of Venice. Also the VPC has 8 topics for visualizations on VPC.org, the Venice City (CK) Knowledge Dashboard with 14 widgets, and Venipedia with articles on at least 500 different topics. However some of this information: the VPC's visualizations, and the VOD's many data sets, are not readily understandable due to inaccessibility and the large quantity, and varying collection rate and types of data (known as the 3Vs). Thus, the goal of this project is to make data collected and stored by the VPC and VOD more available and usable for the Venetian people.

The goal of this project was completed in stages by following three main objectives:

1. Creating a new data visualization website for the VPC focused solely on Big Data.
2. Creating a new dashboard that has reimaged versions of old widgets as well as new social media widget.
3. Creating two new visualizations, one related to Venetian Budget Spending and the other related to Tourist Arrival Data.

The first two objectives revolve around making the VPC's data more available with the website and dashboard both being connected to VPC.org. Although each object was worked on in parallel, they all were coordinated through the first objective.

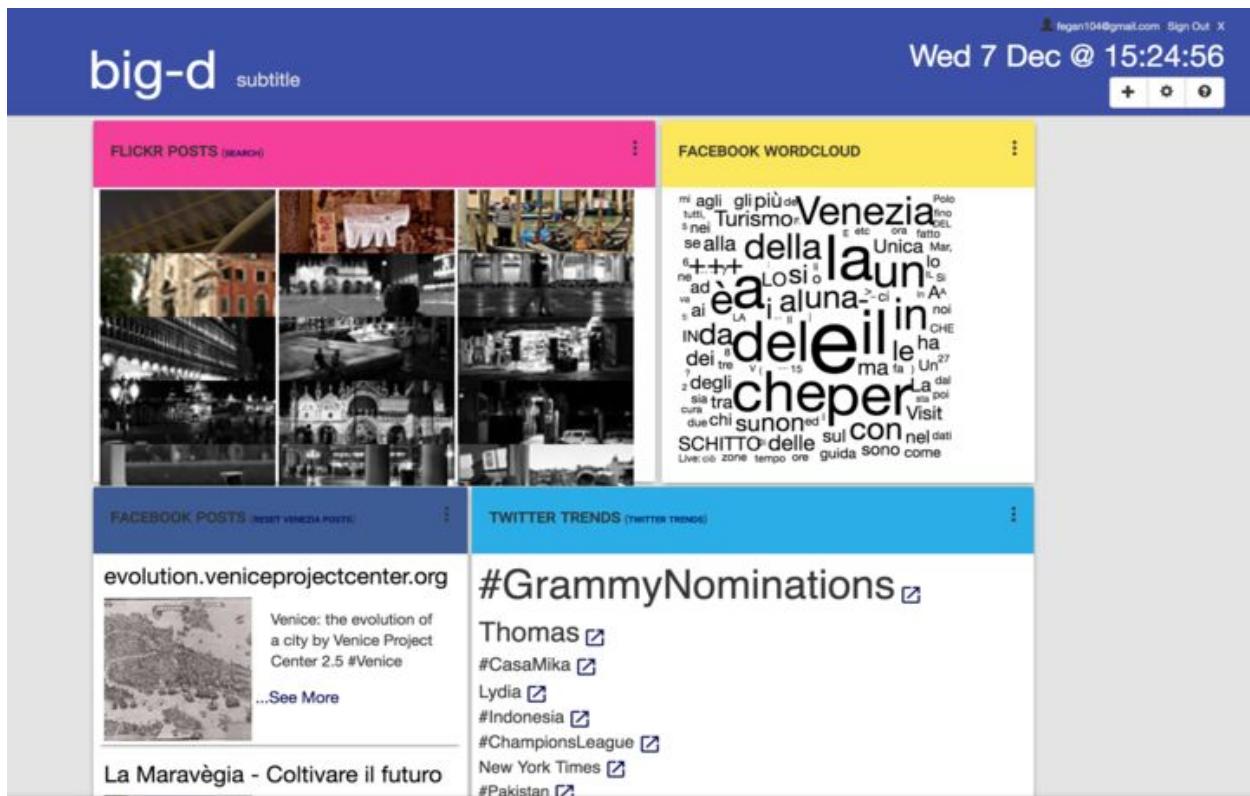
The main features of the visualization website is a toggle section that allows the user to choose between “Urban Elements”, “Dashboard”, and “Visualizations”. The “Urban Elements” section is a way to for the user to see the VPC’s big data related to various structures in Venice such as Bell towers and Bridges. Ultimately, we organized these urban elements by three main sections. After users have decided which type of information they want to see, the menu expands to show what data is available relative to that section and when selected it will be displayed as a layer on a map. This allows for some of the slowly updated Big Data of the VPC to be accessed by the public.



A screenshot of the Urban Elements section of the website.

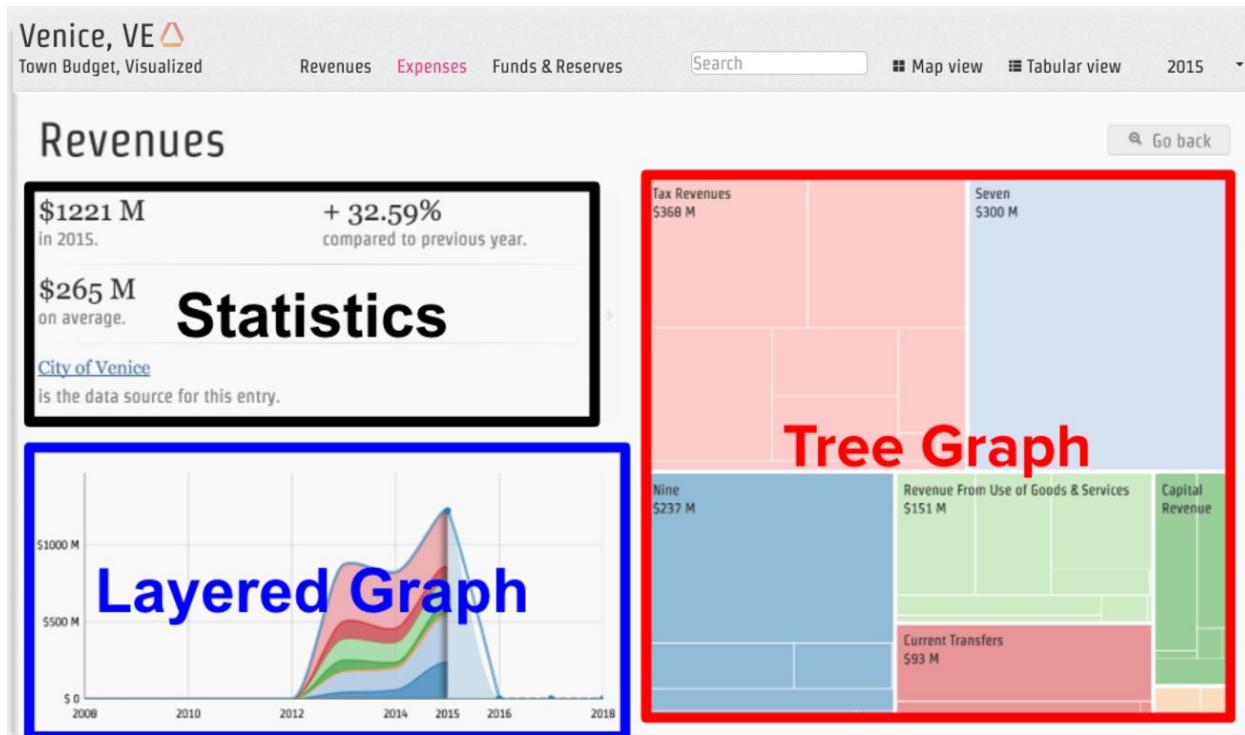
The second toggle section from the main website is the dashboard, which leads to our second objective. Before our work, the VPC hosted a dashboard that had widgets related to various tourist and local data sets such as “Hotel Availability” and “Tide Levels”. However, each widget was created by a different previous team that led to several methods related to the data storage and collection for these widgets. Also, the old dashboard had no widgets related to social media. Through our second object we created four widgets related to social media: Facebook, Twitter, Flickr and the News. The Facebook widget filters posts from various “opinion makers” of Venice such as our

sponsor Reset Venezia. The Twitter and Flickr widgets work very similarly because they both filter the content of their respective website by geolocation of the post. This allows the widgets to display trending topics and pictures related only to tweets and posts written in Venice. Lastly the News widget focused on collecting newspaper articles from the day and producing word clouds depicting repeating words. All four of these widgets, utilize different areas of social media to present all topics relevant to Venice on one platform, making it easier for locals to stay informed.



The updated social dashboard widgets

Lastly, the third section of the website, “Visualizations”, refers to the two visualizations we created ourselves from VOD and VPC data. These visualizations came from Venetian Spending Report (2013-2016) data and Daily Tourist Arrival data collected by the VOD and VPC respectively. The VOD’s budget data culminated in a visualization that contained an interactive tree graph, layered bar graph, statistic section, and tax analysis, which are shown below.



This above image showcases the three main sections of our first visualization. These are the tree graph (red), statistics (black), and the layered line graph (blue). Through each section the user can understand the large data set that is the Venetian Spending report in great detail. Using the Expense section, this visual shows how the tree graph is used to display each section of the budget. Although each section of the visual: "Revenues", "Expense", and "Funds" have different titles, the principle is the same. The title for each Expense section is highlighted in green.

The tree graphic is used to showcase the various sections of the government spending. Both are color coordinated to emphasize the different sections, and the tree graph relates the size of the square to the amount of money for that section, which can be seen in the figure above.

The second visualization, related to tourist arrivals, came from a data set collected by the VPC's own Venice CK Dashboard. Before delving into the functionality, it must be noted that this purely a proof of concept because the logged data needs to be confirmed for accuracy.

The dataset collected by the Venice CK Dashboard's Tourist Arrival widget included real time numbers or extrapolations related to tourist arrival methods such as by plane, boat or bus. From this we correlated this data best to several area line graphs as depicted in the figure below.



The above image showcases how the data from the widget was organized by transportation type. Each mode of transportation received it's own graph and color.

As can be seen the graphs are color coordinated and each section of travel gets it's own graph. The y-axis represents the total number of tourists and the x-axis refers to the date at which the data was collected.

Through all these objectives our project was a great start to satisfying the needs of data availability in Venice. Through our website, the user can access the large sums of real-time and visualization data made available by the VPC as well as access new datasets not visualized by both the VPC and VOD. This website is also easy to access through [Veniceprojectcenter.org](http://Veniceprojectcenter.org), and makes all the data readily available. The new widgets serve to add more dimensions to types of data available to the public because there has been little work with social media in previous years. Also, our new visualizations serve two purposes: to showcase untapped data sources of the VOD as well as the VPC's real-time widget logs and exemplify how new forms of data visualization can make gathering information from data easier. We believe this will inspire future teams to work with untapped data sources of both the VOD and VPC as well work with new visualizations methods to continue to make data more accessible to the people of Venice.

# Table of Contents

<b>Authorship</b>	<b>2</b>
<b>Acknowledgements</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>Executive Summary</b>	<b>5</b>
<b>Table of Contents</b>	<b>10</b>
<b>List of Figures</b>	<b>13</b>
<b>List of Tables</b>	<b>15</b>
<b>1.0 Introduction</b>	<b>16</b>
<b>2.0 Background</b>	<b>19</b>
2.1 Defining Big Data and its Relation to Data Visualization	19
2.1.1 Defining the Three “V”s of Big Data	19
2.1.2 Visualizing Big Data	21
2.1.3 Benefits of Visualizations	22
2.2 Past Efforts in Bringing Big Data to Venice	31
2.2.1 The Venice City Knowledge Dashboard	31
2.2.2 Meeting the Big Data Needs of Researchers	32
2.3 Issues with Accessing Venice’s Current Big Data	34
2.3.1 Shortcomings of the Current Dashboard	34
2.3.2 Shortcomings of Long-Term Data Mediums in Venice	36
2.4 Best Practices in Visualizing Big Data in Other Cities	38
2.4.1 Visualizations that Benefit Venetians	39
2.4.2 Visualizations that Benefit Researchers	45
<b>3.0 Methodology</b>	<b>48</b>
3.1 Creation of an Improved Real-Time Venice City Knowledge Dashboard	50
3.1.1 Designing and Creating our New Dashboard	50
3.1.2 Recycling Elements of the Existing Venice CK Dashboard	52
3.1.3 Adding Social Media Functionality to the New Dashboard	53
3.2 Creating a New Website to Display Long-Term Trends in Venice	55
3.2.1 Creating A List of Use Cases	55
3.2.2 Determining Data Sources to be Visualized	56
3.2.3 Creating a Wireframe of the New Website	57
	10

3.2.4 Creating a Platform Containing All Visualizations	58
3.3 Visualizations: New & Remastered	60
3.3.1 Choosing Data Sources: New & Old	60
3.3.2 Pairing Data to Visualization Method	61
3.3.3 Uploading Data to Visuals and the Internet	63
<b>4.0 Results &amp; Analysis</b>	<b>66</b>
4.1 Venice City Knowledge “Dashboard 3.0”	66
4.1.1 General Design	66
4.1.2 Twitter	66
4.1.3 Flickr	67
4.1.4 Facebook	68
4.1.5 Word Cloud	69
4.2 Venice Big Data website	70
4.2.1. General Design	70
4.2.2 “Visualizations” section	71
4.2.3 “Urban Elements” section	76
4.3 New Visualizations	79
4.3.1 General Design - Venetian Spending Report	80
4.3.2 User Interactivity	80
4.3.3 General Design - Daily Tourist Data	87
4.3.4 Data Organization & Interactivity	88
4.3.5 Aiding Other Project Groups with Visualizing Data	89
<b>5.0 Recommendations &amp; Conclusions</b>	<b>92</b>
5.1 Venice’s Big Data Visualizations	92
5.1.1 Visualizing VOD Data	92
5.1.2 Visualizing Stored Widget Data	93
5.2 Future Improvements to the Venice Big Data Website	94
5.2.1 Improvements to the Urban Elements Section of the Website	94
5.2.2 Other Potential Improvements to the Website	95
5.3 Future Venice Widgets	95
5.3.1 Social Media & Sentiment Analysis	95
5.4 Project Outcomes	96
<b>6.0 Bibliography</b>	<b>99</b>
<b>Appendices</b>	<b>102</b>
Appendix A: Tools Used for Development	102
Appendix A.1 SublimeText Editor	102
Appendix A.2 GitHub	102



# List of Figures

<a href="#">Figure 1: Airbnb Widget</a>	22
<a href="#">Figure 2: NCAA Data Visualization</a>	24
<a href="#">Figure 3: Women's Clothes Size Visualization</a>	26
<a href="#">Figure 4: Consumption of UK Domestic Energy Visualization</a>	28
<a href="#">Figure 5: US Health Map Visualization</a>	30
<a href="#">Figure 6: Venice City Knowledge Dashboard</a>	32
<a href="#">Figure 7: Venice Open Data Spreadsheet</a>	33
<a href="#">Figure 8: Venetian Census Data Visualization</a>	34
<a href="#">Figure 9: Hotels Widget</a>	35
<a href="#">Figure 10: Flagstaff Pedestals Visualization</a>	36
<a href="#">Figure 11: Path to the Hotels Visualization</a>	37
<a href="#">Figure 12: Material Sliders</a>	40
<a href="#">Figure 13: Use of Color</a>	41
<a href="#">Figure 14: Use of Motion</a>	42
<a href="#">Figure 15: Twitter Trends</a>	43
<a href="#">Figure 16: London City Dashboard - Twitter Widget</a>	44
<a href="#">Figure 17: CityBeat Dashboard</a>	45
<a href="#">Figure 18: Russian Government Spending Visualization</a>	46
<a href="#">Figure 19: Dropdown Menu</a>	47
<a href="#">Figure 20: Wireframe of New Dashboard</a>	52
<a href="#">Figure 21: Wireframe of New Website</a>	57
<a href="#">Figure 22: The Data Tab of the VPC Homepage</a>	60
<a href="#">Figure 23: Tourist Arrival Widget</a>	61
<a href="#">Figure 24: Treemap of Venetian Spending Report</a>	62
<a href="#">Figure 25: Tourist Arrival Visualization</a>	63
<a href="#">Figure 26: Raw Venetian Expense Data</a>	64
<a href="#">Figure 27: Raw CSV Expense Data</a>	65
<a href="#">Figure 28: Twitter Widget</a>	67
<a href="#">Figure 29: Flickr Widget</a>	68

<a href="#">Figure 30: Facebook Widget</a>	69
<a href="#">Figure 31: Word Cloud Widget</a>	70
<a href="#">Figure 32: The Navigation Bar</a>	71
<a href="#">Figure 33: Visualization of the Day</a>	72
<a href="#">Figure 34: The Visualization Sidebar</a>	73
<a href="#">Figure 35: Selected Visualization</a>	74
<a href="#">Figure 36: Additional Information</a>	75
<a href="#">Figure 37: Urban Elements</a>	76
<a href="#">Figure 38: Modal Dialog</a>	77
<a href="#">Figure 39: Overlain Data</a>	78
<a href="#">Figure 40: Multiple Overlain Datasets</a>	79
<a href="#">Figure 41: Venetian Spending Report</a>	80
<a href="#">Figure 42: Expenses Section of Budget Visualization</a>	81
<a href="#">Figure 43: Expanded Treemap</a>	82
<a href="#">Figure 44: Budget Sparklines</a>	83
<a href="#">Figure 45: Budget Layered Line Graph</a>	84
<a href="#">Figure 46: Budget Tooltips</a>	86
<a href="#">Figure 47: Tourist Arrival Data</a>	87
<a href="#">Figure 48: Tourist Data Time Slider</a>	88
<a href="#">Figure 49: Housing Team Airbnb Visualization</a>	90
<a href="#">Figure 50: Made Team Business Visualization</a>	91

## List of Tables

<a href="#"><u>Table 1: Use Cases for New Dashboard</u></a>	51
<a href="#"><u>Table 2: Use Cases for the Website</u></a>	56

## 1.0 Introduction

---

Everyday data is stored about our internet browsing, use of public transport, shopping habits, and much more (McAfee & Brynjolfsson, 2012), the field of big data analytics, or more simply the endeavor to understand the various implications of very large data sets (Srinivasa, 2014), aids many different entities at both microscopic and macroscopic levels. Organizations ranging from multinational businesses to small towns are utilize big data in order to improve decision-making, as it allows them to apply “data and analysis” to decisions rather than “intuition and experience” (Lohr, 2012). The 21st century city is an example of a macro-level entity that is capable of benefitting from the use of big data. Similarly to the term “big data”, “smart city” is a popular term that has come to fruition in recent years (Al Nuaimi, Al Neyadi, Mohamed, & Al-Jaroodi, 2015). The notion of a “smart city” is based on the idea that “connecting up, integrating and analyzing the information produced by [technology in cities] ...provides a more cohesive and smart understanding of the city that enhances efficiency and sustainability, and provides rich seams of data that can [be] used to better depict, model and predict urban processes and simulate the likely outcomes of future urban development” (Kitchin, 2014). For a city to become a “smart city” it must undergo a multi-step process (Batty, 2013): the collection of related city data leads to the analysis of the data, which in turn leads to the applications of said analyses in order to solve problems. However, upon its collection, data exists in its most raw form; and in the case of big data, that form may consist of thousands to millions of individual numbers and pieces of information. One way to easily understand large quantities of data is to visualize it in ways humans can derive meaning because humans are inherently visual beings (Tamara van Gog, Fred Paas, Nadine Marcus, Paul Ayres, & John Sweller, 2009).

Venice is just one example of a city that could benefit from the display of its data. For over twenty-five years, the Venice Project Center (VPC) has been collecting data of

all sorts about Venice, collecting information spanning from Venetian art to the city infrastructure (Venice Project Center, 2013). The Venice Project Center is an organization that describes its mission as, “Leaving Venice better than we found it.” The organization was founded in 1988 by Fabio Carrera and has been running Interactive Qualifying Project for WPI students ever since. In 2016, lots of data lies stagnant within the VPC’s database. For example real time information about hotel bookings throughout the city is being stored through the Hotel’s widget but older data isn’t used after it is stored. The VPC currently maintains a general website, a Wiki and a real-time city dashboard (Kirby & Carrera, 2013), all of which display information the VPC-collected data of some sort. Additionally, further measures to increase the availability of data were taken in 2013, culminating in the Venice Open Data website (Ding, Gualdarrama, Horton, & Perrone, 2014), which brought over 100 Venice-related raw data sets to the public. However, not all of the data that has been collected and exists in the VPC’s database has been made ready for display. Essentially, the city government has taken steps towards making Venice into a “smart city”, however, this is an ongoing process, as much of the data that has been collected has simply not yet been presented in ways that people can make meaning out of.

Big data that exists outside of the VPC’s database can, and should, be displayed as well. Currently, the aforementioned dashboard, which was created in 2013 to display real-time data (Kirby & Carrera, 2013), contains widgets, small applications displaying specific information, which highlight various Venice-related information, such as air traffic and daily tourist statistics. The concept of “sentiment analysis” refers to the gauging of how people feel about a topic, often analyzed by “measuring the sentiments embedded in social media posts” (Petulla, 2013), such as the analysis of hashtagged trends on Twitter. The increasing relevance of sentiment analysis is stated by researchers to be resultant of the growth of different forms of social media, such as “reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks” (Liu, 2012). As a result, many city dashboards now display widgets pertaining to sentiment analysis such as London, whose dashboard contains, among others, a simple widget

displaying the top trending hashtags in tweets sent from London (UCL). However, Venice's dashboard fails to display any sentiment-related data at all in its current state.

This project will provide and organize visual representations of big data related to the city of Venice. These visual representations can be anything from tree graphs, geographical maps, or word clouds and the data can span from search histories to recorded boat routes. Once visualized, the data visualizations provided will be delivered through two online websites, aimed at separate audiences. The first medium will be an updated version of the current Venice City Dashboard, which will be aimed at tourists and locals, and will provide real-time data graphics about the city's current conditions, incorporating new elements intended to display trends in Venetian sentiments through the use of data from social media. The second medium will be a website, available to the public but aimed towards researchers, *data.Veniceprojectcenter.org*, which will display old and new visualizations of long-term data related to Venice. We will contribute to the Venice Project Center's ongoing collection of visualizations of long-term data in Venice by creating new interactive graphics of data that has not been visualized yet, while also providing tools for researchers to easily compare data sets in the pursuit of second-order knowledge. The primary sources of the data we will be working with are the VPC's database, as well as major social media outlets, such as Twitter, Instagram, Flickr, and Facebook. Our efforts will allow Venetian locals, tourists and researchers to make sense of otherwise unreachable data so that they can make more informed decisions pertaining to Venice.

## 2.0 Background

In our background, we will begin by describing big data and the benefits of its visualization. We will then briefly address the current problems regarding big data understanding and shift to Venice's efforts to make data available to the public. After this, we will address current issues that plague Venice's means of providing access to its big data, followed by practices that exemplify better ways of providing access to the data in question. Finally, we will address how the current dashboard and VPC/Venipedia can be improved, which will inspire our Methodology section.

### 2.1 Defining Big Data and its Relation to Data Visualization

#### 2.1.1 Defining the Three “V”s of Big Data

Big data has three primary characteristics known as the “Three V’s” (Mauro, Greco, & Grimaldi, 2016), which are listed below and discussed in the following section:

- High **volume** of the data: the sheer amount of data in question,
- **Velocity** of the data: the speed (fast or slow) at which data is moved around, and
- High **variety** of the data: different and inconsistent forms of data.

The study's definition of big data states that these three defining attributes “require specific Technology and Analytical Methods for its transformation into Value” (Mauro et al., 2016). In other words, big data is characterized by attributes that make it difficult to control and view, as each “V” in its definition requires extensive analytical methods to derive any meaning. It is important to understand that according to this definition, big data is not necessarily required to adhere to all three V’s in order to be considered big data; rather the three V’s serve as guidelines with which one may identify data as big data, as it comes in so many different forms.

The first characteristic, volume, refers specifically to the size of the data source. In order to be classified as big data, the volume of the dataset must be on the scale of a terabyte ( $10^{12}$  bytes) of data (Mauro, Greco, & Grimaldi, 2016). An example of big data that is defined by volume is the Veniceprojectcenter.org (VPC). The VPC website is updated to display new information as student based project have been completed. After 28 years of continuous research, VPC.org and the VPC itself now have several terabytes worth of information (Ding et al., 2014). The sheer size of this data set inhibits its comprehension because most analytic methods cannot handle terabytes of data, as a large amount of the data collected isn't fully visualized.

Another factor that must be addressed is the velocity of the data. The velocity of big data corresponds to the specific rate at which these large data sets are being formed or collected. The two main rates that will be relevant to our project are real-time and long-term. Real-time data is data that is stored and processed immediately after creation (Mauro, Greco, & Grimaldi, 2016), whereas long-term data is collected over a longer period of time (monthly, annually, etc.). With respect to Venice, real-time data is exemplified by the current Venice City Knowledge (CK) Dashboard, which uses data sets related to Venice that are updated in real time. These data sets are updated immediately through various outside sources. Specifically, one data set is updated through four different hotel websites as room availability is posted (Kirby & Carrera, 2013). The second measure of big data in terms of its velocity, long-term data, is exemplified by Venice's Venipedia, which is updated annually by WPI students to display the data they collect during their IQPs (Venice project center.2013; Gopalkrishnan, Steier, Lewis, & Guszcza, Aug 12, 2012). Both real-time and long-term are velocities that present issues for understanding big data because the framework of the analysis must account for the rate of new data.

Lastly, the variety of big data must be acknowledged in order to understand its complexity. The variety of big data refers to how many different forms of data are contained in a data set, or related data sets. For the purposes of our project, the variety of the data sets we will be working with will include pictorial, numerical and textual data.

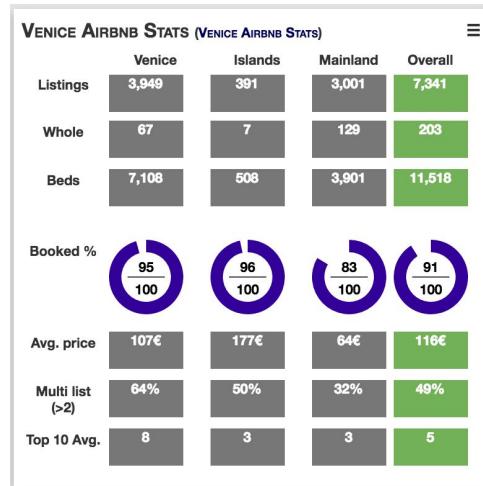
Common examples of variety in big data include social networks, such as Twitter, Facebook and Instagram. To exemplify this, social media includes pictures and various text posts, while also incorporating numerical data through geotags to describe location. These can all be made specific to Venice by the addition of a filter, such as *#Veniceitaly*. These different varieties of data require added work to organize, thus making them more difficult to analyze. To address the aforementioned issues with processing big data, the data must be manipulated so different audiences may benefit from it.

### 2.1.2 Visualizing Big Data

The primary way to gain value from big data is through analytics (Katal, Wazid, & Goudar, 2013). Analytics is a process of comprehending data through methods of analysis. By analyzing big data, entities can infer information that can help them to make informed decisions (K. Michael & K. W. Miller, 2013). However, the three V's of big data serve to increase the difficulty associated with inferring information. To reduce the complexities associated with big data, visualizations can be used to depict the data through graphics, pictures or animations. These mediums can be analyzed in order to extract value from the data.

To accurately analyze big data visualizations, the visualizations must display the information in relevant ways. According to Edward Tufte, an oft-cited pioneer in the field of data visualization, qualities of good data graphics include: “avoiding distorting what the data has to say”, “present[ing] many numbers in a small space”, “be[ing] closely [related] to verbal description of data”, and “serv[ing] a reasonably clear purpose” (Tufte, 1983). Once big data visualizations have met these criteria, value can be derived. In relation to Venice, they have big data visualizations in the form of widgets on the Venice CK Dashboard. A widget is a display or interface that allows users to interact or view data. One specific widget on the CK Dashboard is the “Venice Airbnb Stats” widget, which allow users to see information about all Airbnb apartments in Venice (Kirby &

Carrera, 2013)). Although this widget is not interactive, it does address several of the qualities of a good visualization that are seen in the figure below.



**Figure 1:** Depicted above is the Venice Dashboard's Airbnb widget that displays statistics regarding house listings in Venice. This showcases the readability of the widgets, as all information is immediately available.

First, it displays the numbers related to Airbnbs in a small space. Also, by just showing the numbers, it doesn't distort the original data in any way. Lastly, the data displayed relates entirely to Airbnbs in Venice.

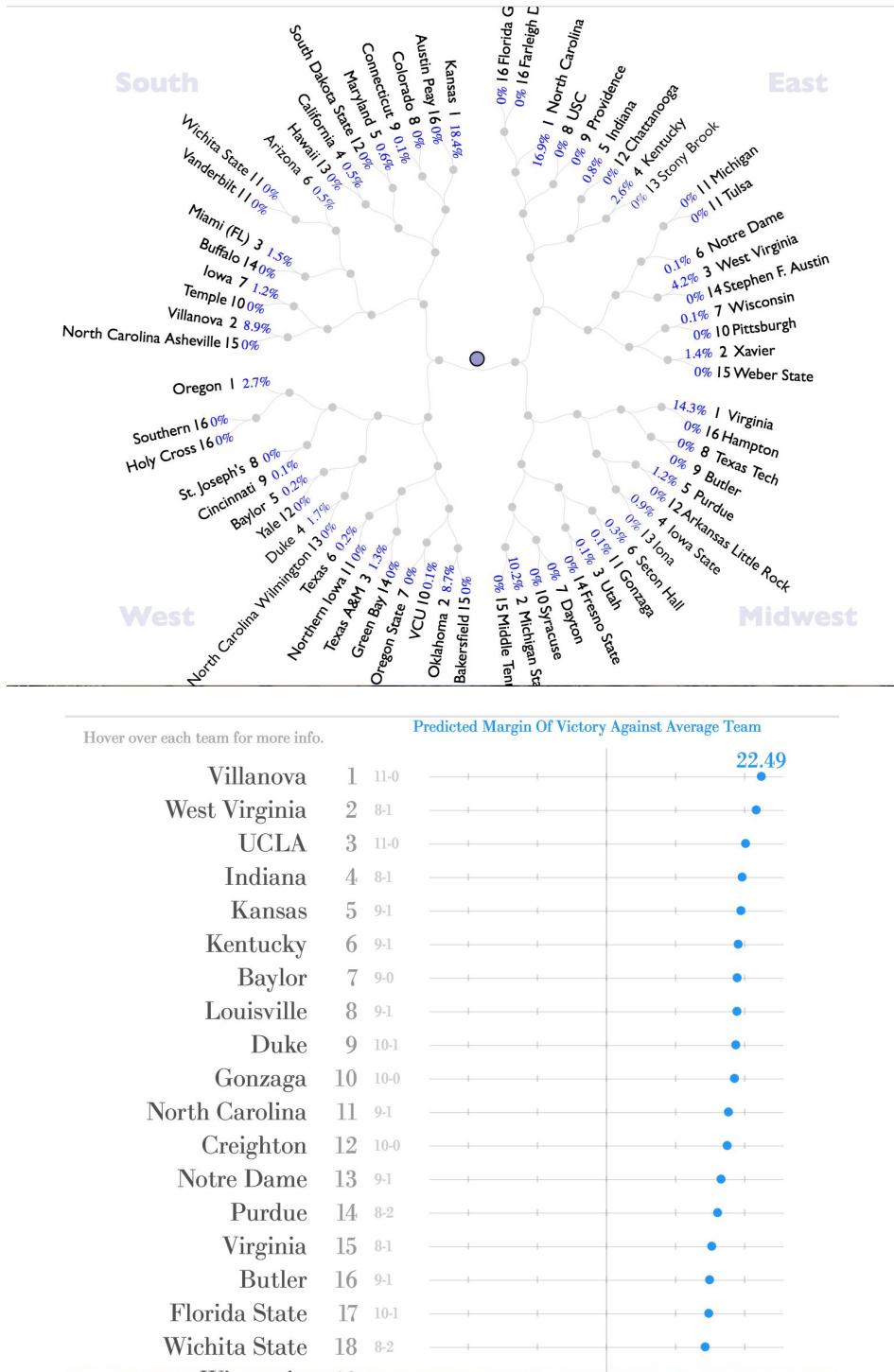
Although the aforementioned example describes just one widget, there exist other widgets on the CK Dashboard, as well as other city dashboards, that display similar data in different ways. With these dashboards in place, citizens and researchers have been able to derive value from data once impossible to analyze. However, since big data visualization and analytics are relatively new, there are still many improvements that can be made (Lohr, 2012). In this regard, big data analytics is a resource that is not being fully utilized. This results in a problem on the global scale because if ordinary people, as well as researchers, lack good visualizations of the data they need to make informed decisions, they will be unable to properly understand a large amount of available information.

### 2.1.3 Benefits of Visualizations

As stated earlier, good visualizations are used to gain value from complex sets of big data. This value manifests through two main benefits: better decision-making and

improved services (Brian Gentile, 2014). Although big data is relatively new, preliminary steps toward visualizing and gaining benefits from big data have been achieved.

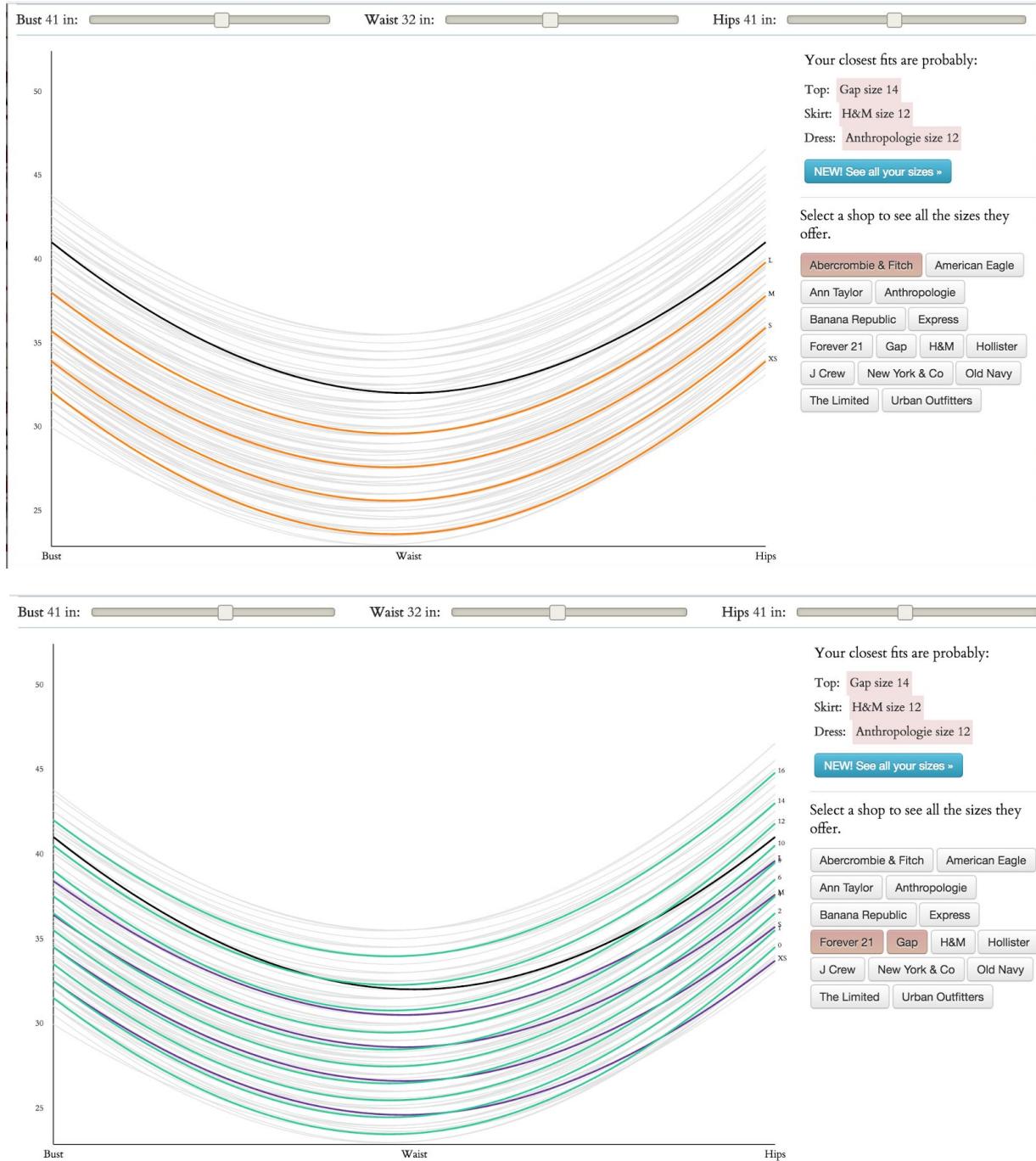
One visualization that exemplifies applications in better decision-making is the NCAA Power Ranking Visualization. In the United States, “March Madness” takes center stage as all Division 1 NCAA basketball teams compete. Attached to this “madness” is a tradition of betting on which teams will win, normally combined with the incentive of a monetary wager (March madness betting to total \$9.2 billion this year.2016). With this added pressure around picking the right teams, big data and its visualization can be used to help the decision process. As shown in Figure 2 below, a visualization was created using large sums of data related to each basketball team in the NCAA to create an interactive graph that displays win percentages at every stage of March Madness (Ed Feng, ).



**Figure 2:** In the above figure, the chances of winning for NCAA basketball teams have been shown in various ways. With the upper image, the ability to click on a specific section of the bracket (championship) and see the percent chance a team has of getting that far in the tournament. Lastly, the bottom image depicts how any NCAA team will do against an “average team” where the points they’ll win or lose by are represented by a number line.

Along with this helpful visualization, it compares every team against a fictitious “average team” which creates a profile that contains the average statistics of every NCAA team. These two visualization optimize the user’s ability to select winning teams because Figure 2 (top) depicts a team’s odds of winning, which can be used to decide how many games a team will likely win. While Figure 2 (bottom) depicts how well a team will do against a statistically average team, which can be used to predict unexpected wins from average ranking teams as some high win percentage teams have only won against worse teams. Together, the user can view the chances of a team winning, which are shown in the statistical visualizations.

Another relatable innovation regarding big data is the inconvenient system of women’s clothes sizing. This system major flaw is the fact that a woman’s pant, shirt or underwear size can vary from company to company, such as Forever21 to GAP (Dockterman, 2015). This inspired the “What Size am I?” visualization pictures in Figure 3 below.

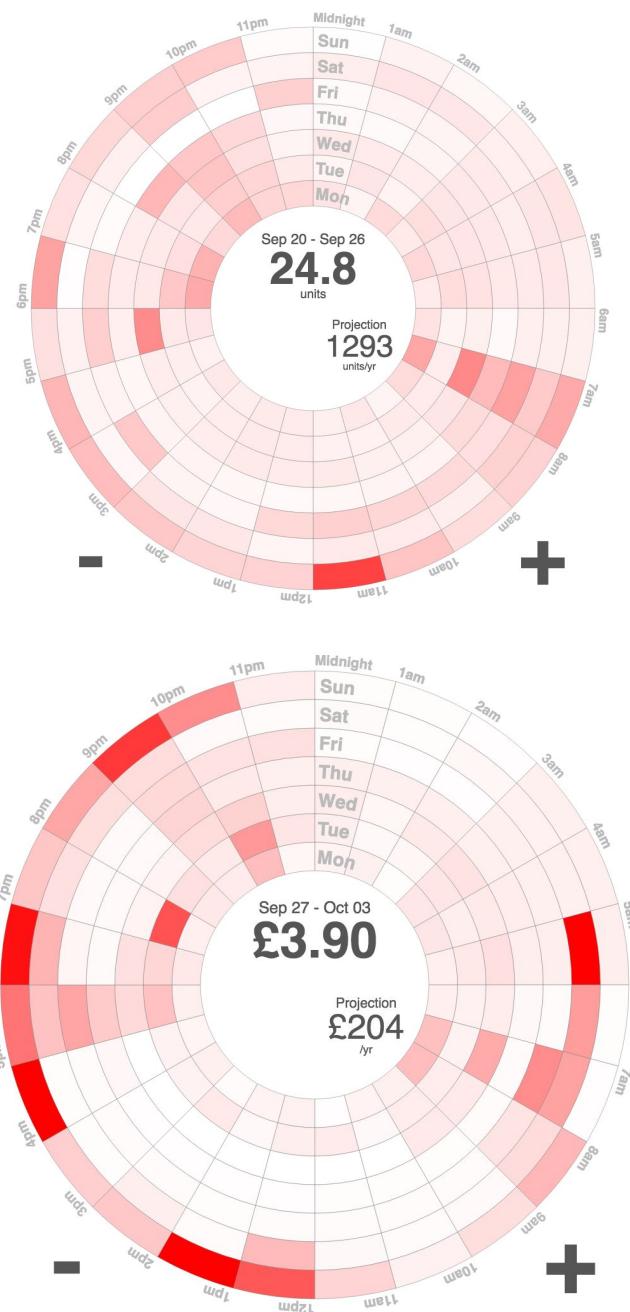


**Figure 3:** This figure depicts the “What Size am I?” visualization. The top image depicts a women with dimensions 41in, 31in, 41in (bust, waist, hips) and it shown by the dark curve. Also selected are the various sizes associated with Abercrombie and Fitch (highlighted orange). Lastly, the bottom image shows the difference in sizes between the GAP and Forever 21 (highlighted dark blue and light green respectively)

By compiling the records of 21 different retail stores, the user can interact with the visualization to highlight the corresponding sizes and recommended stores to shop at

(Powell-Smith, 2012). This visualization also allows the user to toggle between UK and US stores and two measurement systems to ensure accurate results and quality service. Ultimately this allows for an average shopper to take uncertainty out of clothes shopping and make more informed decisions. When combined with the NCAA visualization, the two show unrelated fields both allowing the audience to make more effective and informed decisions by analyzing data through the ease of visualization.

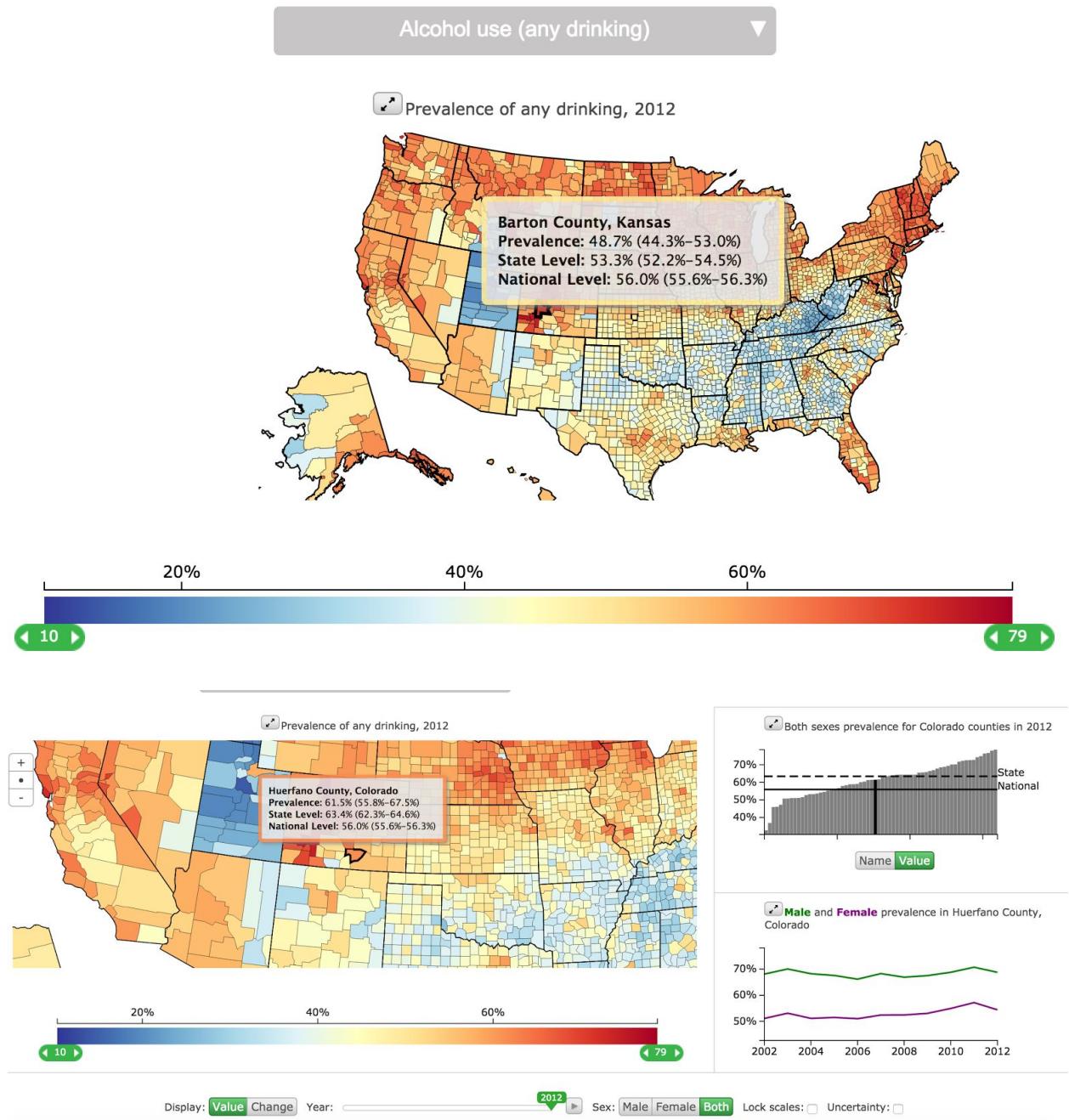
Also, through the use of big data, services offered by companies can be improved. The Domestic Energy Consumption Visualization best shows the main example of this that is displayed in Figure 4.



**Figure 4:** This figure depicts the consumption of domestic energy in the UK the form of this circle. It is divided to represent different hour sections along the edge of the circle and weekdays are represented inside of the circle. Using the plus and minus buttons you can change the week displayed, where the top image is 9.20-9.26 while the bottom image is 9.27-10.3. The visualization uses color gradients, where the darker red a segment the higher the energy consumption at that day and time. Also, the user can toggle between CO2 emission, power usage and monetary usage.

In the above figure, data collected on domestic energy consumption in the UK by “Current Cost CC128” is displayed in a way that allows the user to toggle between

electricity usage, monetary cost, and CO<sub>2</sub> emission (Cook, 2012). Once selected, the circle is broken up by weekday and hour for a specified week that can be changed. Also, the color gradient is used to displays the relative energy consumption of that time period. This visualization allows the user to see the exact economic and environmental impact their usage has on the world. With this information made easy to understand, it could encourage the usage of an alternate energy source because the cost of electricity consumption can be compared from fossil fuel usage to environmentally friendly alternatives. Also, energy companies can be encouraged to redirect energy consumption to match the times of excessive usage (Al Nuaimi et al., 2015). Another example that showcases the ability of big data visualization to improve services is the “US Health Map” (See Figure 5 below).



**Figure 5:** The above figure shows a heat map of the USA as it relates to alcohol where the less affected areas are blue and get redder as it increases. On the top, it is highlighted that the user can have accuracy up to the county level. The lower image showcases a graphic that shows how each county compares to state and national average.

Using big data collected by the Institute of Health Metrics and Evaluations, this visualization uses an interactive heat map to display the various unhealthy habits across the United States (Institute of Health Metrics & Evaluations, 2015). It allows the user to

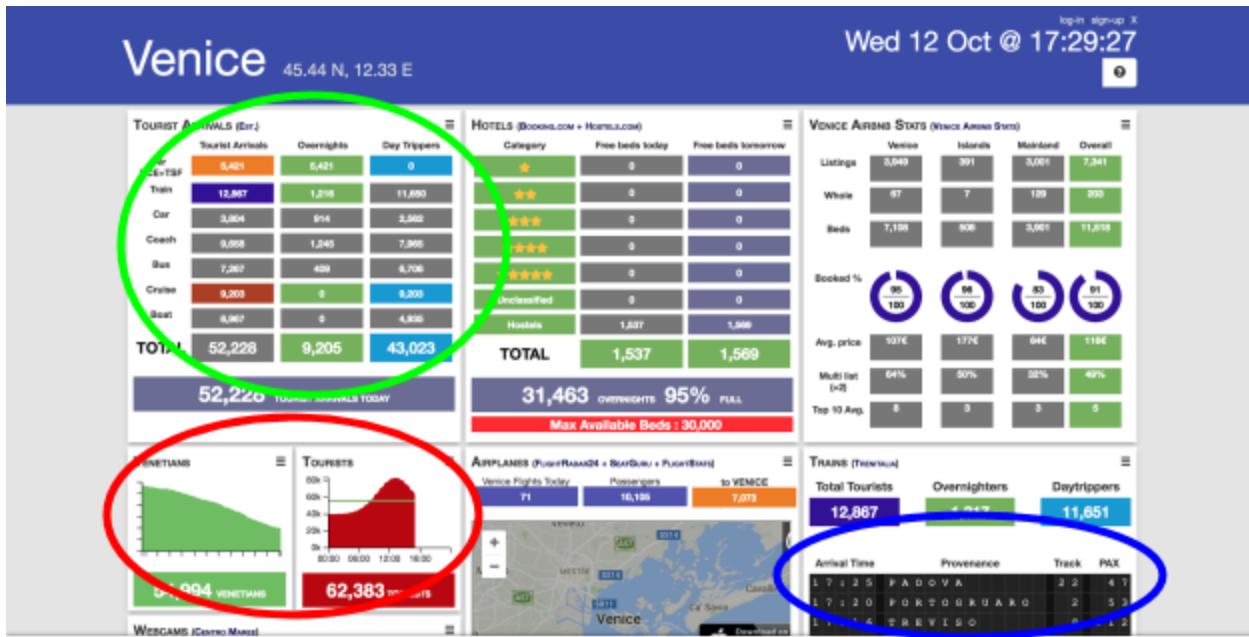
sort between counties, genders, trends, and specific years. This interactivity allows for understanding that could lead to improved and specialized preventative services based on a county's personal trends because the trends are easily recognized before issues can occur (Murdoch & Detsky, 2013). Both this example and the other examples showcase how big data can be visualized in ways that help people, whether they be in energy efficiency and alternatives or preventative and specialized medical services. These avenues are all accessible due, in part, by the ability of the average person to easily understand complex data sets.

## 2.2 Past Efforts in Bringing Big Data to Venice

The city of Venice and the Venice Project Center (VPC) have made several efforts to make Big Data available to its citizens. In this section, we outline some of these past efforts and how they have benefited the citizens, tourists, and researchers of Venice.

### 2.2.1 The Venice City Knowledge Dashboard

To bring the benefits of big data, such as better decision-making or comprehension of information (like Venetian budgets) to the citizens and tourists of Venice, the VPC celebrated its 25th anniversary with the creation of the Venice City Knowledge Dashboard in 2013. Currently, the city of Venice maintains this live dashboard on CityKnowledge.net with real-time widgets and infographics related to many topics, ranging from hotel bookings per day to tide fluctuations to live webcams in various locations throughout the city (Kirby & Carrera, 2013). A snapshot of the dashboard can be seen in Figure 6.



**Figure 6:** Depicted above is the Venice City Knowledge Dashboard that displays big data in graphical (circled red), textual (circled blue), and numerical (circled green) forms

### 2.2.2 Meeting the Big Data Needs of Researchers

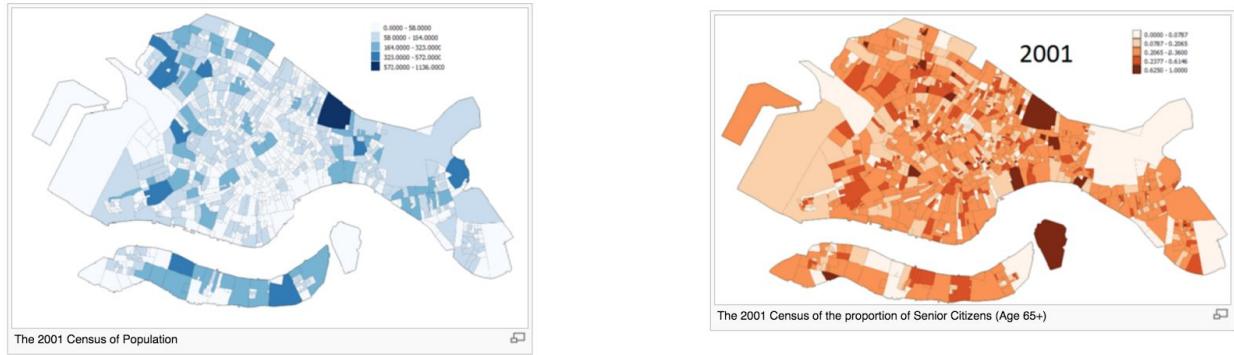
As a way to bring the benefits of big data to researchers, the city of Venice released the Venice Open Data website in 2014. This website provides downloadable resources such as government spending reports and census data (Ding, Gualdarrama, Horton, & Perrone, 2014). These resources are provided in the form of raw data in spreadsheets, which is displayed below in Figure 7.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	DATAESTRAZIONE	DESCRMACROCOMMERCIO	DESCRCOMMERCIO	NUMEROPOSIZIONE	DATAPOSIZIONE	CODVIA	DESCRIVIA	NUMCIV	BARRATO	INDIRIZZOCOMPLETO	DURATA	DATITECNICI	TIPO	DES	TIPO	NUM_STELLE	INSEGNA	NUM_CAMERE_TOT		
2	2016-07-01 C	num 1763	piano 1	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;	F;AFFITTACAMERE;1;CA' SAN MARC							
3	2016-07-01 C	num 1763	piano 2	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;	F;AFFITTACAMERE;1;CA' SAN MARC							
4	2016-07-01 C	num 395	quartiere Di	zona CENTRO STORICO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;A;ALBERGO;3;ALBERGO OLIMPIA;35;58;N.	Camere: 14 singola * 1 - 2 dop							
5	2016-07-01 C	num 38	zona TERRAFERMA;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;A;ALBERGO;4;SMARTHOTEL HOLIDAY;46;90;N.	Camere: 8 singola * 1 - 32 doppia * 2 - 6 tripl								
6	2016-07-01 C	num 4502;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;PERBELINI GRAZIANO;2;N.	Camere: 2 doppia * 2									
7	2016-07-01 C	num 64	quartiere LIDO MALAMOCCO ALBERONI;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;3;alloggi andrea donatella;6;10;N.	Camere: 2 sing								
8	2016-07-01 C	num 398	piano TERR.	quartiere DORSODURO SANTA CROCE SAN POLO GIUDECCA;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;1;ALLOGGI OLIM								
9	2016-07-01 C	num 6068	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;1;CASA CRISTIANA;4;9;N.	Cam								
10	2016-07-01 C	num 684	quartiere DORSODURO SANTA CROCE SAN POLO GIUDECCA;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;3;ALLOGGI GARLAUTO MARIA;2									
11	2016-07-01 C	num 2344	piano 1	quartiere DORSODURO SANTA CROCE SAN POLO GIUDECCA;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;1;ALLOGGI LOCA								
12	2016-07-01 C	num 3547	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;P;CASA PER FERIE;RESIDENZA UNIVERSITARIA AB									
13	2016-07-01 C	num 6068	zona CENTRO STORICO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;A;ALBERGO;3;SANTA MARINA;20;33;N.	Camere: 7 singola * 1 - 13 doppia * 2								
14	2016-07-01 C	num 940;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;CA' DI ROSE;2;3;N.	Camere: 1 singola * 1 - 1 Suite * 4									
15	2016-07-01 C	num 5700;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;CATTAPAN PAULINO;3;10;N.	Camere: 1 doppia * 2 - 2 quadrupla * 4									
16	2016-07-01 C	NUM 5	INT. 10 P.I. QUARTIERE MESTRE CENTRO;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;santhi alloggi;;									
17	2016-07-01 C	num 3539	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;P;CASA PER FERIE;;CASA CARDINAL PIAZZA;24;48;									
18	2016-07-01 C	num 4410	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;;HOUYAK RACHID;2;7;7;N.									
19	2016-07-01 C	num 112	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;ZANIO CARLA;3;6;N.	Camere: 3 doppia * 2								
20	2016-07-01 C	num 2221	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;1;CA' TINTORETTO;4;7;N.									
21	2016-07-01 C	num 4603	piano 1	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;2;CA' D'ORO;3;7								
22	2016-07-01 C	num 244	quartiere F / zona TERRAFERMA;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;A;ALBERGO;3;TITAN INN;23;47;N.	Camere: 22 doppia * 2 - 1 tripla * 3								
23	2016-07-01 C	num 29	quartiere BURANO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;1;LOCANDA CIPRIANI;6;9;N.	Camere: 3 singola * 1 - 3 doppia * 2								
24	2016-07-01 C	num 3489	QUARTIERE CENTRO STORICO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;A;ALBERGO;4;14;39;N.	Camere: 3 singola * 1 - 4 doppia * 2 - 7 Suite * 4								
25	2016-07-01 C	num 4114;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;STEFANI ANDREA;3;7;N.	Camere: 2 doppia * 2 - 1 tripla * 3									
26	2016-07-01 C	num 3362;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;BONARDO PAOLA;1;2;N.	Camere: 1 doppia * 2									
27	2016-07-01 C	num 3056;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;BUSOLIN SAMUELE;2;4;N.	Camere: 2 doppia * 2									
28	2016-07-01 C	num 6114;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;PASTRONG ARKADIU;SZ;3;5;N.	Camere: 1 singola * 1 - 2 doppia * 2									
29	2016-07-01 C	num 2/J;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;BUSETTO PAOLO;1;2;N.	Camere: 1 doppia * 2									
30	2016-07-01 C	num 240	quartiere F / zona TERRAFERMA;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;R;RESIDENCE;3;TITIAN INN;15;31;N.	Camere: 14 appartamento a 2 letti * 2 - 1								
31	2016-07-01 C	num 1861;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;R;RESIDENCE;3;TITIAN INN;15;31;N.	Camere: 14 appartamento a 2 letti * 2 - 1									
32	2016-07-01 C	num 689;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;FRONTERRE' GIUSEPPE;2;4;N.	Camere: 2 doppia * 2									
33	2016-07-01 C	num 3385/I	quartiere SAN MARCO-SANT'ELENA-CASTELLO-CANNAREGIO;ANNUALE;Tipologia Struttura Ricettiva.	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;B;BED AND BREAKFAST;SURATI MARCO;3;6;N.									
34	2016-07-01 C	num 883;ANNUALIF;Tinfinia Struttura Bifamiliare	Num.	Stelle.	Insegna.	Num.	Camere.	Capacità ricettiva.	Descriz.	Camere;F;AFFITTACAMERE;2;DOMUS I LINGUERRI;4;7;N.	Camere: 1 singola * 1 - 3 donna * 2									
2016-07-01_CM_OpenData_Attrivita																				

**Figure 7:** Above is a sample spreadsheet taken from the VOD site. It shows the complexity of the data through the numerous headings, inconsistent formatting and large amount of data to be observed.

As can be seen, there are several reasons why this raw data cannot be easily understood. Initially, all the information is cluttered and unorganized. Also, regarding volume, some spreadsheets are several thousand rows long and have up to 20 columns when formatted properly. Lastly, there are simple issues like the language barrier associated with headings and further still the jargon used as heading titles. For these reasons the data presents many challenges for anyone to immediately understand.

One resource that the Venice Project Center has contributed to big data in Venice is an online encyclopedia, *Venipedia*. Venipedia provides a wealth of information on thousands of subjects related to Venice, such as data related to each of the bell towers, and histories of each fountain in the city (Worcester Polytechnic Institute, 2016). The Venice Project Center also offers several static graphic visualizations of the data they have collected, such as population trends of census reports in the 20th century (Venipedia, Census Trends., 2016). The visualization shown in Figure 8 utilizes a visualization technique called “geographic color mapping.” This technique is good for scalar values, such as population, because the graphic best shows the context that data describes. In this case, the map lays out the areas that population data represents.



**Figure 8:** On the left is 2001 census data by population where the blue gradient describes the number of people, increasing as the gradient darkens. On right is the 2001 census data by percentage of senior citizens where the red gradient represents a range of percentages and increases as the gradient darkens.

## 2.3 Issues with Accessing Venice's Current Big Data

Despite Venice's efforts to display big data through the CK Dashboard, the VOD project, the VPC website ([Veniceprojectcenter.org](http://Veniceprojectcenter.org)), and Venipedia, each effort has its challenges. In this section, we will analyze the issues associated with the VOD, VPC website, Venipedia and dashboard.

### 2.3.1 Shortcomings of the Current Dashboard

In addressing the lack of information made available to the inhabitants of Venice, the VPC created the current Venice CK Dashboard. Although it still gathers and filters information well, it has accumulated several technical issues over time, and lacks certain relevant information. In this section, we discuss the specific technical issues of the dashboard, as well as its current exclusion of social media.

Throughout our experiences working with the dashboard, we have determined four main technical issues: outdated widgets, general aesthetic and organization of data, lack of interactivity, and an inefficient mobile outlet. With regards to outdated widgets, several widgets that currently exist on the website present visualizations that fail to display data due to a miscommunication error between the widget and stored dataset (Kirby & Carrera, 2013).



**Figure 9:** Above is the Venice CK Dashboard Hotels Widget. Currently it displays no data for the star level of the hotels but correctly depicts total overnights.

This prevents users from viewing the data relevant to them, as the data isn't visible. Further, six of the fifteen major widgets have little to no visualizations, and instead just show numbers determined from various sources (Kirby & Carrera, 2013). Additionally, the layout of the data visualizations has very little consistency in terms of its design, resulting in a generally confusing assortment of information. The current dashboard lacks interactivity as well, as only two of the fifteen main widgets displayed on the current dashboard allow mouse interaction to manipulate the data (Kirby & Carrera, 2013). This prevents the user from tailoring the presented data to their own needs. This, along with the outdated widgets and general organization of the dashboard, impede the average user from gaining as much real knowledge from the dashboard as they could. Lastly, mobile functionality is not supported with the current

dashboard, as no efforts have been made to provide a mobile browser-friendly version that accounts for the small size of a phone screen. This prevents the user from access the real-time data when it is needed away from a desktop.

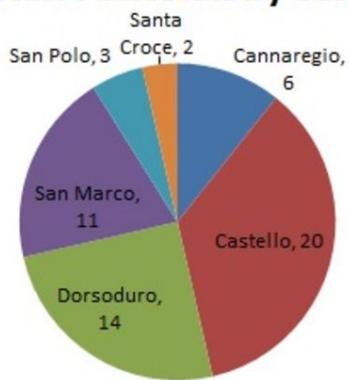
In its current form, the Venice CK Dashboard displays one sole widget related to social media: a widget designed to display Instagram posts geolocated to Venice, which currently does not work. Aside from this, there are no visuals that represent data collected from social media. While opinions - which are often broadcasted through social media - lack objectivity by nature, this does not detract from their relevance to decision-making on both small and large scales, as the choices people and organizations make are “largely based on how others see and evaluate the world” (Liu, 2012). This emphasizes the importance of including subjective data obtained from social media, such as Instagram, Twitter, Facebook and Flickr, on a real-time dashboard among quantitatively measured data.

### 2.3.2 Shortcomings of Long-Term Data Mediums in Venice

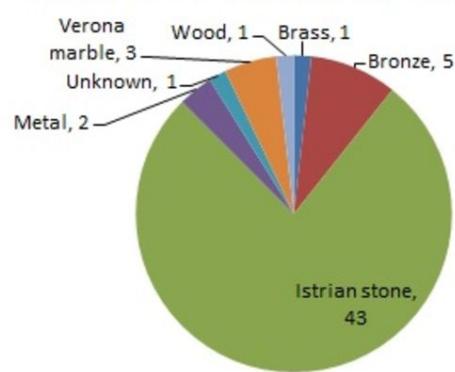
The city of Venice has access to many great long-term data resources that are provided through VOD and Venipedia. However we have identified several areas that could be improved upon, which will be discussed in this section.

One long-term data resource that could be improved upon is the Venice Open Data (VOD) Project. From its databases currently available on the VOD, only one has been properly visualized, and this occurred on the VPC website. This visual is the Flagstaff data collected and displayed in Figure 10 below (Ding, Gualdarrama, Horton, & Perrone, 2014).

**Flagstaff Pedestals by Sestiere**



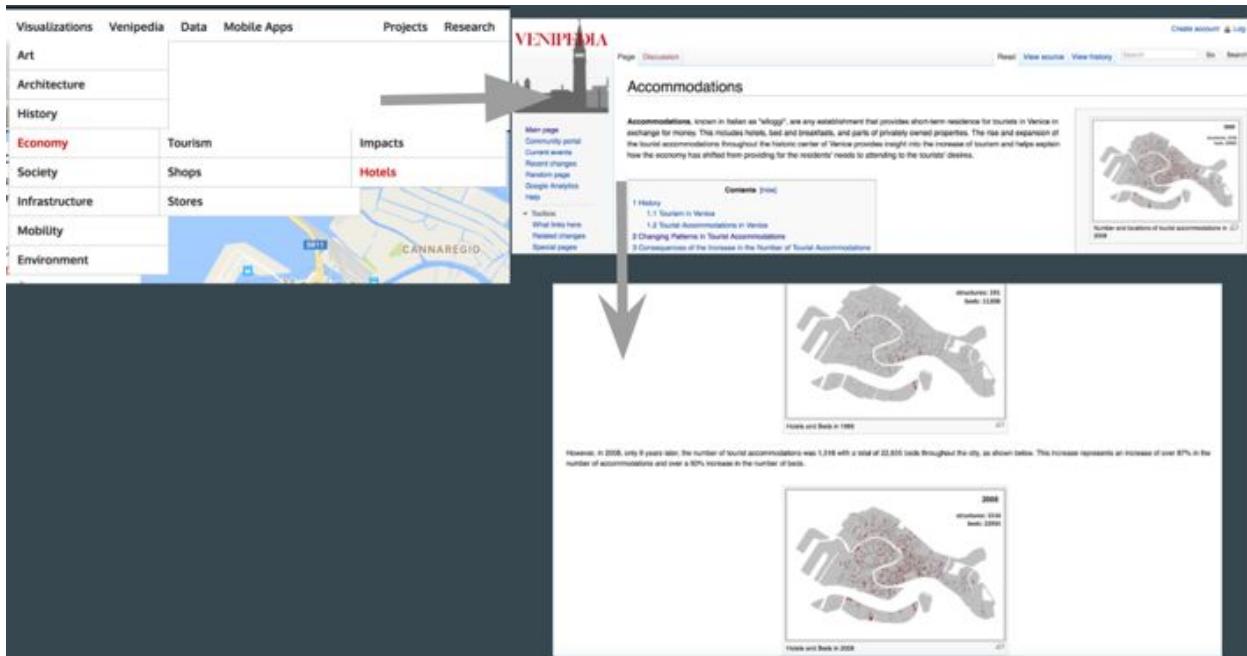
**Flagstaff Pedestals by Material**



**Figure 10:** Depicted above is the flagstaff data collected by a previous IQP Team and uploaded to the VOD. This is a simple pie chart that sorts that number of Flagstaff by Sestiere (left) and Material (right). It was created with data collected to showcase how the VOD and VPC's datasets can be combined.

While the VOD project accomplished exactly what it intended on accomplishing - the public display of raw data - this method of big data display could certainly be improved upon, as the initiative of the VOD Project was to make government data available to the people (Ding et al., 2014). However, in its current form it is difficult to derive meaning as it consists of 138 spreadsheet based datasets. Further visualization of the data would be useful.

Two other sources of long-term data implemented by entities in Venice are Venipedia, an online encyclopedia detailing different aspects of Venice, and the Venice Project Center website, which discusses students' projects over the years. These sites have accumulated two main shortcomings over the years. These shortcomings are the lack of interactive visualizations and the lack of organization of the data visualizations throughout the sites.



**Figure 11:** In the above figure, the path needed to reach the Hotels visualization is shown. The user starts at the VPC homepage and goes through three menus. After which the user is directed to a Venipedia page where the visual is further down on the page. The visuals display hotels and beds in Venice displayed by red dots.

For example, clicking on “Hotels” under the Visualizations tab on the VPC website leads to a section of a Venipedia article which features 2 static colored maps that display the numbers of hotels and beds in Venice in 1999 and in 2008. Although this shows the change in hotels between the two years, this could also be displayed more interactively with a visualization that updates the dots that represent hotels as you change a time domain. This example also speaks to the second issue, because this ineffective visual is only accessible through several subdirectories. These shortcomings can lead to confusion and issues in reaching the sources of data that Venice and the VPC want to make available to the public.

## 2.4 Best Practices in Visualizing Big Data in Other Cities

In this section, we discuss uses of big data in other cities. We have discussed several issues with Venice’s means of displaying its data to everyday Venetians and

researchers alike. Herein, we will move to discuss ideal methods of data visualization that combat the specified issues with the display of Venice's data, such as broken widgets or raw nonvisualized data. By identifying methods of solving problems that are similar to those that we have outlined in Section 2.3, we will ideally be able to create a framework that could be make better visualizations for the big data in Venice.

#### 2.4.1 Visualizations that Benefit Venetians

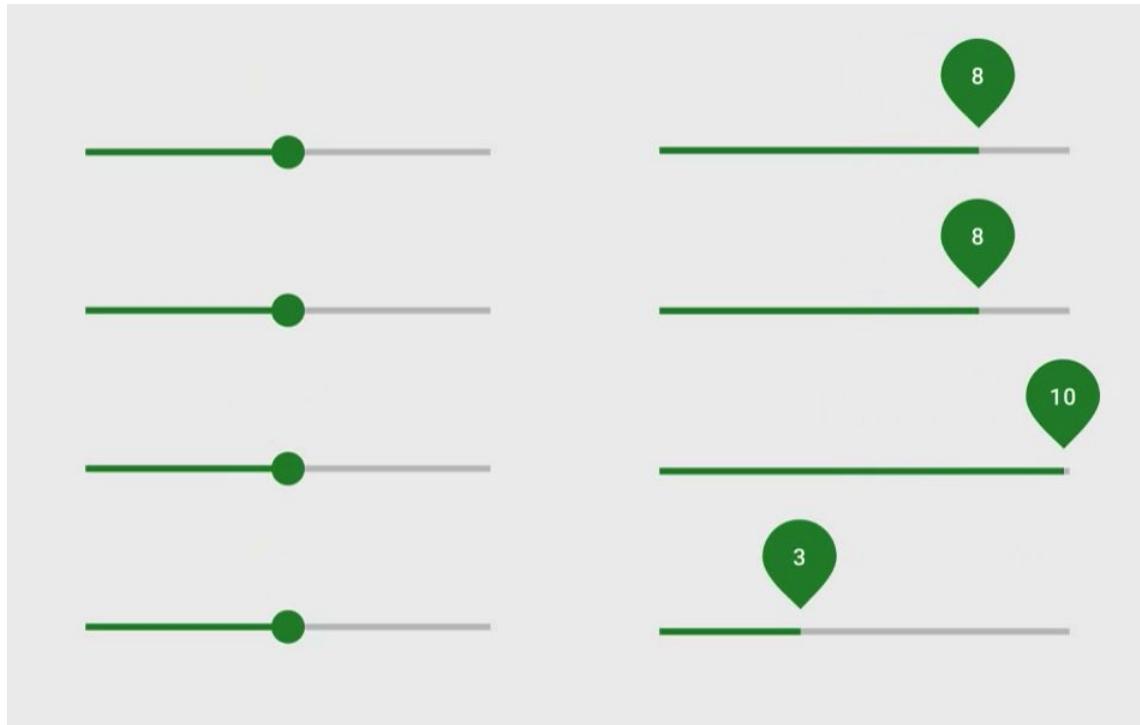
Data visualizations designed for locals and tourists are focused on summarizing and displaying important real-time events and news that are relevant to their city. This data can be reliably obtained from social media, as it provides a constant, nearly instantaneous supply of information. In this section, we discuss some techniques for displaying this data as well as mining it for information and context.

Dashboard data visualizations are set to update in real-time, the sources of the data being fed to the visualizations of the data may change at any point, potentially breaking the source of the data graphics. Essentially, dashboards and other means of real-time data visualization must be maintained in order to account for potential changes in the sources of the data.

The visual appearance of a website in question is also important to the user's perception and trust of the website. Empirical analyses of studies have shown that if websites are easy to use, fast and are visually appealing, users are more likely to trust their validity (Flavián, Guinalíu, & Gurrea, 2006). To address the inconsistent design quality across the web Google has published a set of guidelines known as "Material Design" a specification for how to design effective interfaces.

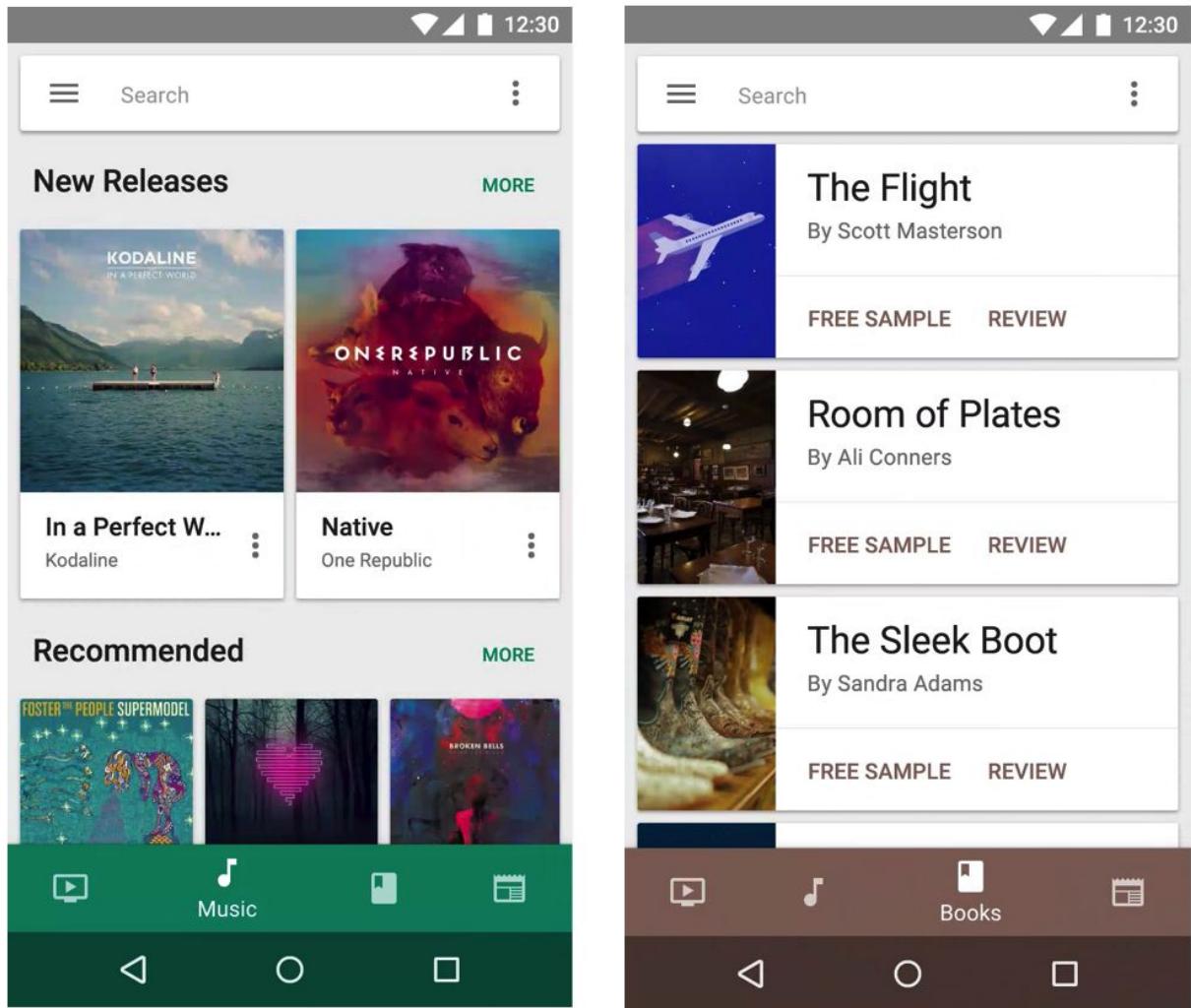
Material Design aims to optimize visual appearance as well as user accessibility (Google, ), 2014). These specifications are manifest in certain frameworks, such as Angular.js and Materialize CSS, that allow programmers to create websites that apply techniques to enhance user experiences on a variety of levels, such as the use of grids, imagery, and motion, which are "not only visually pleasing, but also create a sense of hierarchy, meaning, and focus" (Chang, Louie, Mark, & Wang, ) 2014). The three

principles of Material Design are: “material as the metaphor”, “bold, graphic, intentional”, and “motion provides meaning”. The first rule, “material is the metaphor”, states that any graphic object must obey the same consistent rules throughout your application. For example, buttons and images should not just pop in and out of existence. They should naturally move into place like things in the real world.



**Figure 12:** An example of material sliders. This showcases the first principle as the sliders to the left directly correspond to the bars on the right. On top of being interactive, this exemplifies the fact that these bar graphs (pictured right) do not appear from nowhere and are instead related to the sliders.

The second principle, “bold, graphic, intentional”, states that the choice of color or typeface should be used to build a sense of hierarchy. For example if we used a shade of green throughout our financial data visualizations using the same shade of green in our population or tidal visualizations would be confusing.



**Figure 13:** An example of how color provides context. In this case the green color (left) is associated with music while the brown color (right) is associated with books. Although other sites may color code differently, the colors allow the user to associate sections with colors for ease of use and understanding.

Finally the third principle, “motion provides meaning”, gives the user a sense of control and familiarity. The whole application should be responsive to user taps and provide feedback to the user such as ripple effect on user clicks.



**Figure 14:** An example of how buttons react with ripples. This shows how “motion provides meaning”. The “ripples” around the word “Focus” on the second button are part of an animation that activates when the button is clicked on.

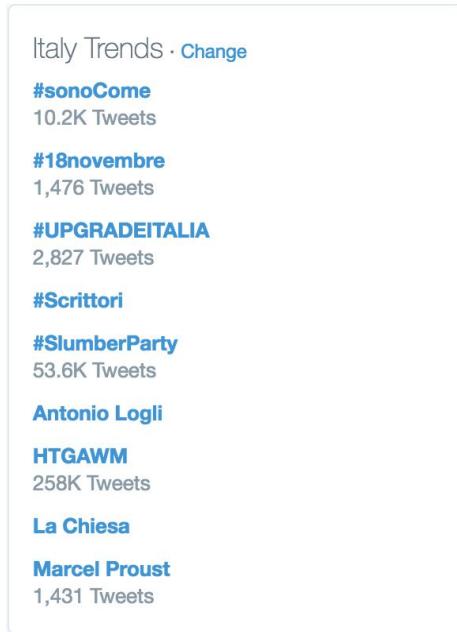
By using these frameworks, programmers can deliver more interactive and visually pleasing technologies, which can in turn increase the user’s trust in the product.

Social media is an ubiquitous source of real-time information, and thus can be displayed on real-time dashboards in order to deliver a more comprehensive user experience. However, because there is a large amount of social media output at any given time, it is necessary to explore ways to filter what data from social media is most relevant to the city in question. From this filter, the information should be displayed to locals and tourists that will interact with the dashboard.

This notion of inferring valuable information from the subjective content written in social media posts is generalized in the field of *sentiment analysis*, or opinion mining (Liu, 2012). To provide a more in-depth view of sentiment analysis, the term is defined as “the field of study that analyzes people’s opinions, sentiments, evaluations, attitudes, and emotions from written language” (Liu, 2012). The growing importance of sentiment analysis is parallel to the growth of social media, as “for the first time in human history, we now have a huge volume of opinionated data recorded in digital form” (Liu, 2012), which can in turn be analyzed.

One method to filter social media data is by displaying trending hashtags. On social networks such as Twitter, users can insert a *hashtag* in their posts, defined by a pound symbol (#) that is directly in front of a word or phrase in a tweet, turning the word

or phrase into a link that can be clicked on, leading to many other tweets that inserted the same hashtag. The simplest example of this sort of implementation lies on Twitter's website itself, which filters the hashtags by number of mentions in the user's area, and displays the links on the user's homepage on a menu entitled "Trends" (depicted in Figure 15 below).



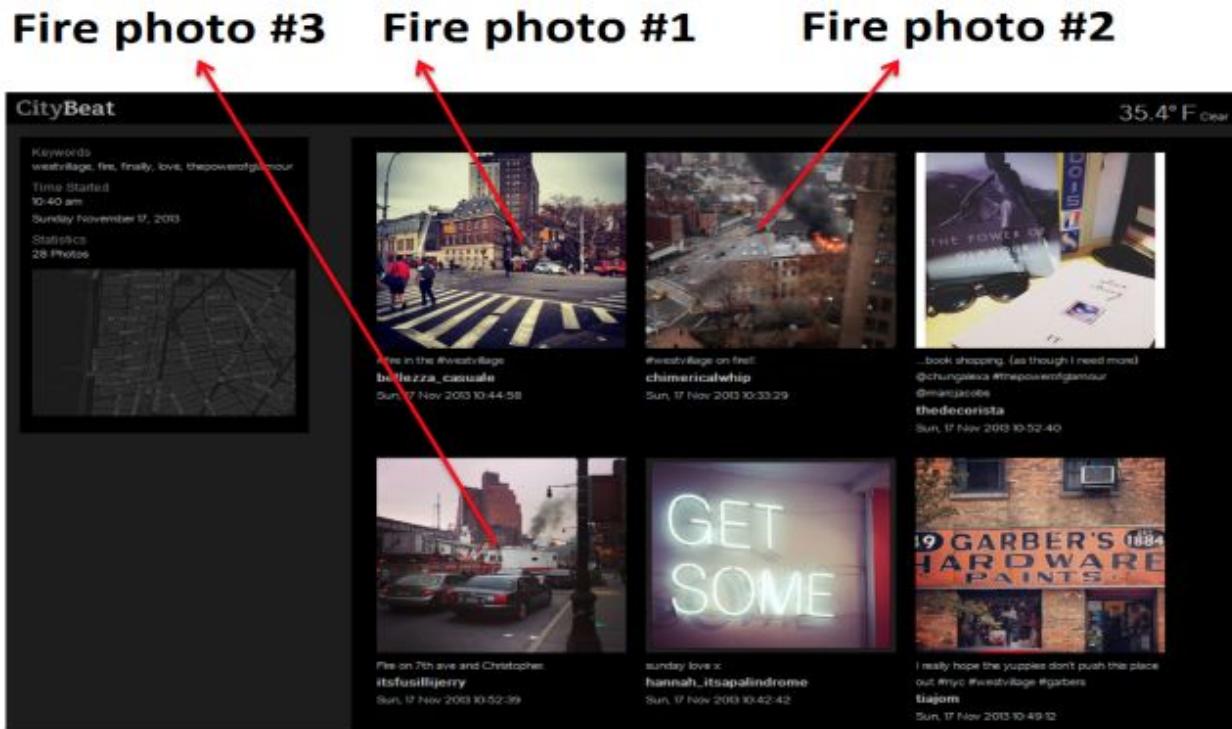
**Figure 15:** Showcased above is Twitter's own trend tracker. This particular tracker is already set to Italy and allows the user to see trends in order of number of tweets. The exact number is immediately below the trends and the trends are linked to a page with description.

This type of implementation is an example of sentiment analysis because it allows the user to not only see what topics people are thinking about, but also read their opinions on said topics. The London City Dashboard (seen in Figure 16) is an example of a city dashboard that implements a trending feature as well, in the form of a widget (UCL). The widget displays trending hashtags that are sorted by their number of mentions, with the most trending hashtags being displayed in larger font sizes. The hashtags are geolocated to the city of London using the Twitter application-processing interface (API), ensuring that all the tweets in question are sent from London. This allows the user to view what other London locals and tourists are discussing the most on Twitter.



**Figure 16:** London City Dashboard Twitter trends for London. As can be seen, the trends are sorted by size with the large ones being more relevant in tweets. Also, the trends aren't limited to just hashtags as some keywords show such as "England"

The CityBeat dashboard is a more in-depth example of social media visualization in the form of sentiment analysis, which is aimed at locals and tourists, in this case with respect to New York City (Xia et al., 2014). The purpose of this dashboard is to analyze trending topics in New York City-geolocated posts on social networks such as Twitter and Instagram, in order to detect ongoing activities in the city. Essentially, the dashboard is able to analyze sentiments in great magnitude in order to understand and detail concrete events. This is shown in Figure 17 as the CityBeat dashboard recognizes the three pictures of fire, from social media, are at the same area and concludes that the building in the photos is on fire. This illuminates the power of sentiment analysis using social media, as methods can be applied to social media in order to turn subjective information into objective data.



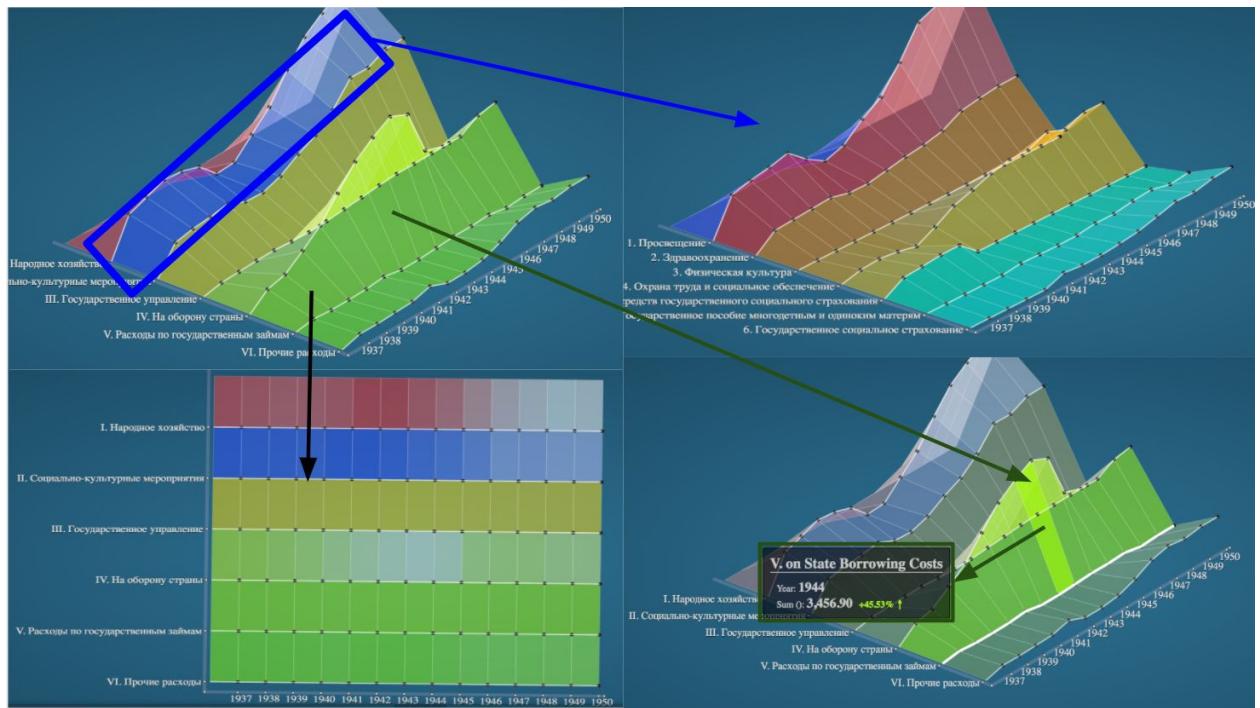
**Figure 17:** A figure showing the CityBeat Dashboard with three separate images of the same fire all tweeted about and collected on the dashboard. This lead to the sentiment analysis that deduced that a fire was occurring in real time.

#### 2.4.2 Visualizations that Benefit Researchers

In order for researchers to make meaning of long-term data and subsequently apply it to their research, it must first be presented in a way that allows researchers to understand it. A key point from Edward Tufte, an esteemed pioneer in the field of data visualization, is that “graphics reveal data” (Tufte, 1983). While it is beneficial to provide public access to data in its raw form, so that researchers can at least view the data, it is also necessary to create abstractions of the data, so that everyone researchers and everyday viewers alike may better understand the data.

While creating visualizations is a crucial start in making meaning out of data, applying interactivity to these visualizations is a method of applying even greater meaning to the data (Telea & Safari Books Online, 2015). One example of an interactive big data visualization is the history of the Russian government’s spending, shown below in Figure 18. In this interactive visualization, the data is displayed in three dimensions: the year, the specified department of the government, and the amount spent. In order to

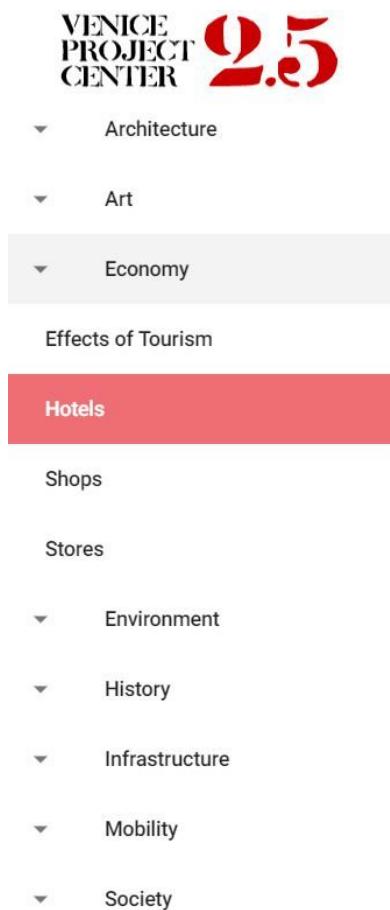
maximize user accessibility to the data, the data is displayed as a three-dimensional model that the user can adjust and view from different angles. The third dimension, the amount of money spent, is designated both by height and by lightness of color as shown, allowing the user to make sense of the three dimensions even without looking at the three-dimensional portrayal of the data. These sorts of graphics can be created through the use of JavaScript libraries such as D3.js, which essentially provides easy ways for programmers to create complex and interactive data visualizations (*Introduction to D3.*, 2015).



**Figure 18:** 3D and top-down views of the same interactive visualization of Russian government budgeting over a 14-year frame of time. Depicted in top left is the initial visualization, where each section of spending is outlined in a different color and shown from 1937-1950. Highlighted in blue is the functionality that allows the user click on a section and then the graphic changes to show the subsections of that section (top right). Highlighted with dark green is the user ability to hover over a specific section and then see the exact name of the expense and the change from previous years (bottom right). Lastly, highlighted with the black arrow is the ability of the user to rotate the visualization from a 3D view to a two dimensional view where the colors now display a gradient to showcase the change in spending to those sections.

Another important aspect of data visualizations is organization. Organizing the visualizations in an intuitive way ensures that researchers are able to find all relevant graphics about a given subject. This ensures a consistency through all of the VPC's

data. This also has the advantage of making sure relevant visualizations aren't overlooked and forgotten. One example of proper data visualizations structuring can be seen below, in Figure 19, where visualizations are organized by broad topics at the top level and a pull-down menu opens for each expanding to show all visualizations of that topic. This showcases the fact that the organization allows the user to first think of the general type of data they want to see and, at a click, only display those visuals.



**Figure 19:** Our organization of visuals matches the VPC's layout as a dropdown menu is highlighted. The user is able to click the general section they want more information (Economy in dark gray). From this point, all visualizations related to Economy are shown allowing easy access to the exact information the user was looking for

## 3.0 Methodology

---

Our mission was to aid Venetian locals, tourists, and researchers by creating, manipulating and organizing interactive, easy-to-understand visualizations of Venice-related data. These various graphics were presented through two primary forms. One medium was a new website that functioned as an updated version of the current Venice City Dashboard, which was primarily catered to locals and tourists, and displays real-time information about the city. The other medium was a website populated with interactive visualizations of long-term data, which is available to the public, but is focused towards researchers, intended to provide meaningful and reliable sources of long-term information about Venice. To accomplish our mission, we identified three main objectives:

1. To produce an updated version of the current Venice city dashboard, which displays real-time graphics about the city
2. To create a new website devoted to the presentation of interactive infographics displaying long-term trends about Venice
3. To allow the user to gain access to more information by producing a new visualization based on budget spending and redesigning an existing visualization.

In order to complete our first objective of implementing a new and improved dashboard, we utilized the functionality of the current City Knowledge dashboard framework, which allows authorized users to create new dashboards and populate them with widgets. After we created the new dashboard, we filled it in with a combination of existing widgets, some of which we reworked, and new widgets that we created. We will expand on our methodology for our first objective in Section 3.1.

As opposed to those on the improved dashboard, the visualizations displayed on the new Venice Big Data site are diverse and do not all follow the same format like widgets do. Thus, we created an entirely new website to display these visualizations, rather than adhering to an existing schema. The website was then organized by areas of interest, and populated with visualizations of Venice-related data. Our methodology for this second objective will be discussed further in Section 3.2.

Our third objective focuses on the fact that big data visualizations have improved greatly since the founding of the VPC (Telea & Safari Books Online, 2015) and the constant need for new information to be visualized. To satisfy both of these areas we will implement a visualization related to new Venetian budget data and reuse data from an existing visualization or widget. The former will showcase the power of new visualizations on new information while also keeping the information relevant to the people of Venice because understanding where tax dollars is relevant to any taxpayer. The latter showcases the power of new visualizations to draw forth more information from data already visualized. Also, the tourist data being visualized is always relevant to Venice because the constant rise in tourist numbers creates issues in all aspects of local life. (Mack, 2012)

In order to complete our project, we used resources that are freely available on the web, stored in databases owned by the VPC, and published by the Venetian government. Our project resources range from data collected hundreds of years ago, such as Venetian census data, to rapidly paced real time data, such as tweets about the city. Lastly, we must effectively use data visualization as it relates to big data. Our project specifically focused on visualizing two types of data: real time data from the city such as Facebook shares or Instagram posts, and larger static datasets, such as Venetian spending reports. The following sections describe the methods we carried out to achieve each objective listed above.

### 3.1 Creation of an Improved Real-Time Venice City Knowledge Dashboard

Our first step in improving the dashboard was selecting a robust and effective platform for creating our website. We researched several open source web dashboards but due to the fact that we were inheriting such a large project it made sense to keep with the same code style and platform that existed on the previous dashboard. To proceed further, we divided our methodology with respect to the new dashboard into three primary categories: designing and creating our new dashboard, making improvements to existing features on the dashboard, and implementing new features and infographics to improve the robustness and information output of the new dashboard.

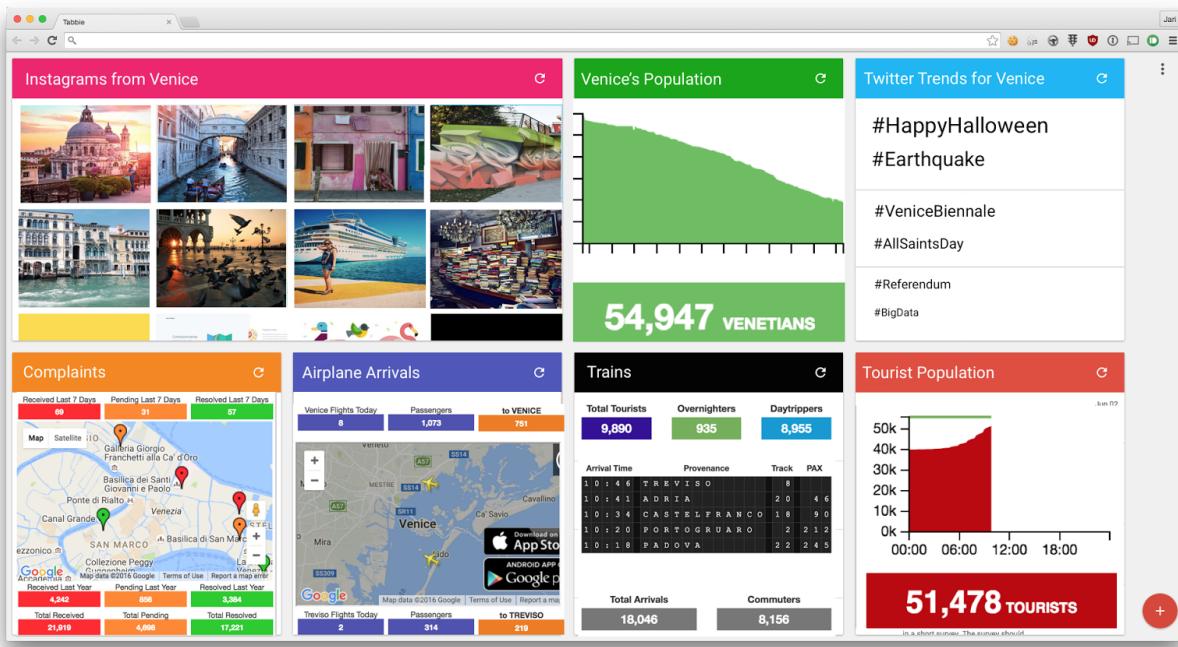
#### 3.1.1 Designing and Creating our New Dashboard

In order to create the new dashboard, we began by brainstorming methods of functionality and design. In software engineering, one common method of breaking down a larger goal into more specific aspects of user functionality is the creation of a table of use cases. Each use case is a specific intended way through which the user can interact with the end product. The “Entry Criteria” field refers to the criteria that must be in place in order for the specific functionality to take place, whereas “Exit Criteria” denotes the required state for the action to be considered finished. Below, we detail our table of use cases for the new dashboard.

Title	Steps	Entry Criteria	Exit Criteria
Open Widget Information	1. User selects “About” section on widget 2. Box appears, containing information about widget’s history and functionality	Widget exists and has information that is not currently displayed	New information has been displayed to the user
Drag Widget	1. User drags widget to a different part of the screen 2. Widgets are reorganized based on the user’s movement	There is an available place for the widget to be moved to	Widgets are reorganized
Toggle Between Numeric & Graphical Visualization	1. User requests a toggle from graphic to numerical or vice versa 2. Widget updates visualization to new form	Widget has both numerical and graphical visualizations	Widget has been updated with selected visualization type
Reorganization Widgets	1. User requests subdomain of widgets 2. Only widgets of selected subdomain appear	Widgets organized by topic. Directory.	Widgets of subdomain have been displayed
Link to Source	1. User requests source for widget 2. User is redirected to website used for data scraping	Widgets. Link to website used for data scraping	User is redirected to appropriate webpage
Zoom Into Visualization	1. User requests to zoom into graphic 2. Graphic updates with minimized axes bounds	There is data in the graphic and the graphic is not zoomed in all the way.	The visualization has been zoomed into.
Zoom Out of Visualization	1. User requests to zoom out of graphic 2. Visualization updates with enlarged bounds for axes	There is data on the visualization and it is not zoomed out all the way.	The visualization has been zoomed out of.
Social Sharing (Tweet Button)	User requests to display information on social media User is redirected to social media platform with link to widget/infographic/etc.	Widget/Infographic. Button that redirects user	The user is redirected to requested social media platform with link

**Table 1:** This figure showcases how our team organized our thoughts before coding any of our widgets. As can be seen we gave the functionality a name (Title), shortly described how it should work (Steps), and then decided what the widget would need to work and what it would produce (Entry & Exit Criteria). This is used in software engineering frequently and allowed us to ensure our widgets contained all the needed functions.

With regards to design, we used Balsamiq, a graphic design tool used to easily wireframe mockups of websites, in order to visualize what our dashboard could look like. When designing our mock-ups we consulted Google's Material Design Guidelines in regard to color choices, structuring of non-homogenous data, and actions available across widgets. For the color choices we wanted to keep each widget toolbar color bold and consistent with whatever source it came from such as Twitter Blue for the Twitter widget. For the layout of our widgets we decided to keep widgets to different sizes that fit the data best. For example the Instagram widget has the plenty of space so you can easily see the images being posted however for the airplanes widget a single square was enough.



**Figure 20:** Our wireframe of the new dashboard, based on Material Design standards. Using these standards we were able to envision how the widgets would line up and be displayed. This allowed us to see exactly what we wanted to code for before we started. This ensures less errors, for example using the wireframe it depicts relative sizing of widget and how many we should make in one line on the screen.

### 3.1.2 Recycling Elements of the Existing Venice CK Dashboard

The previous Venice CK dashboard contained many widgets that could be useful for our new website. We recycled widgets displaying regularly updated data about

Venice from the previous dashboard - such as those displaying the date, local time, incoming and outgoing flights at Venice airports, and the weather forecast. In order to accomplish this, we were provided by the Venice Project Center with the code for the existing widgets, which we were able to transfer to the new dashboard.

In addition to the widgets themselves, we decided to keep all other features from the previous dashboard. For example, each widget contained an “About” pop-up, detailing information about the creators of the widget, as well as its sources of data and how it functions. Of course, it is important that users are aware of what they are looking at, so we opted to keep this feature. However, because several widgets’ About sections were missing critical information, we looked at each section individually and made changes to ensure consistency between the About sections and to guarantee that users have enough information to know where the data comes from and how it is manipulated to form the widget. We added an  icon to each widget toolbar so that the user can easily access the About section. The main other piece of functionality that we kept from the previous dashboard is the ability for a user to drag widgets around the screen. This way, users can customize the layout of the dashboard to their tastes, simply another layer of interactivity with the dashboard.

### 3.1.3 Adding Social Media Functionality to the New Dashboard

In addition to incorporating old widgets on our dashboard, we included visualizations of new data that had not yet been displayed in order to create a more comprehensive dashboard. One source of real-time information that is accessible via an API and is included on many modern dashboards is social media. As such, we gathered Venice-related posts from different social media sources and displayed them on the dashboard. We identified Twitter, Flickr and Facebook as social networks that we would explore ways to create new widgets with. Throughout the course of our project, we created widgets based on each of these social networks, which we discuss further below.

To get the data for each of our widgets we created a Go script to run on Heroku that is a platform that hosts and runs projects for free. We chose to write our script in Go because it is the most performant language available on Heroku. This would allow us to parse lots of data very quickly and perform the needed analysis. Once the Facebook, Twitter, and Flickr data has been collected on Heroku it is sent to the dashboard where to populate each widget (See Appendices). The first widget we made displays in real time the Twitter trends for the city of Venice. This widget was selected because it used building blocks of all other widgets HTML, JavaScript, and CSS in a simple manner where we only needed to make simple request for trends in our area and then displaying that data in a list. Twitter widget displays it. Users can click on each trending topic and view tweets related to that topic.

The next social media widget that we planned on creating, “Instagrams”, would display Instagram posts related to Venice that are organized by several filters. Through working with the Instagram API, we discovered that in order to get public photos posted in Venice we would need to have each individual user login to Instagram on the widget. This would take away from the “at-a-glance” nature of the dashboard so we deemed the Instagram widget unfeasible. From there we moved to collecting Flickr photos to display on the dashboard. We then added functionality so that the user is able to click on a photo, which enlarges the photo and displays the caption.

The last social media widget we made was the Facebook Feed widget. This widget takes Facebook posts from popular Venetian Facebook pages and displays them to the user. This widget gives a feel for what topics Venetians are currently discussing.

Finally, we created a widget that allows the user to view a simple word cloud, giving the user a general sense of which Venice-related words are being posted about among various platforms. In order to do this, we wrote a script that takes Facebook posts that are being displayed on the dashboard at any given time, compiles them into and feeds the text into a word cloud generator provided by D3.js, which displays them in a word cloud where the size of each word is proportional to its frequency.

## 3.2 Creating a New Website to Display Long-Term Trends in Venice

We decided on *Venice Big Data* as the name for the website on which we displayed our visualizations of long-term trends in Venice. We identified five main tasks that reached this objective:

1. Creating a list of use cases.
2. Determining data sources to be visualized.
3. Pairing new data sources with visual representations.
4. Creating a wireframe of the website.
5. Creating a platform that contains all visualizations.

### 3.2.1 Creating A List of Use Cases

First, we decided how the new website would work. To do this, we created a list of use cases. A use case is a function that users can interact with on an application, such as clicking on a button to update or refresh the data in a graph. As such, creating a table of intended use cases for an application aids developers because it allows them to organize all desired functionality into one table. The list includes five columns that are labeled *Title*, *Steps*, *Entry Criteria*, and *Exit Criteria*. *Title* refers to the name of the function. *Steps* are normally one to two short bullets that describe how the function is used. *Entry* and *Exit Criteria* describe the structures needed to execute the function and what should be left after the function, respectively. This method of detailing use cases was chosen because understanding the functionality we want early aided us in limiting the vast amount of data that can be visualized. Additionally, knowing our desired functionality helped us to sculpt our wireframe, and ultimately, the new website. This method is common in software development, and transitioned well for our website building purposes.

Our initial table of use cases for the website is as follows:

Title	Steps	Entry Criteria	Exit Criteria
Pan Visualization	<ol style="list-style-type: none"> <li>1. User requests a valid panning movement of the visualization</li> <li>2. Visualization slides from one place to another</li> </ol>	Visualization exists and has data off-screen that can be panned to.	Visualization has been panned.
Zoom Into Visualization	<ol style="list-style-type: none"> <li>1. User requests to zoom into graphic</li> <li>2. Graphic updates with minimized axes bounds</li> </ol>	There is data in the graphic and the graphic is not zoomed in all the way.	The visualization has been zoomed into.
Zoom Out of Visualization	<ol style="list-style-type: none"> <li>1. User requests to zoom out of graphic</li> <li>2. Visualization updates with enlarged bounds for axes</li> </ol>	There is data on the visualization and it is not zoomed out all the way.	The visualization has been zoomed out of.
Load New Visualization	<ol style="list-style-type: none"> <li>1. User selects a new category of visualization</li> <li>2. WebApp reloads with new visualization</li> </ol>	There is more than 1 visualization category available for display.	The current visualization has been changed to the selected visualization.
Download Data From Visualization	<ol style="list-style-type: none"> <li>1. User requests to download data source used for visualization.</li> <li>2. WebApp provides a download link for the selected data source.</li> </ol>	The data source is available for download and the data is legally able to be shared.	The user has been provided a download link.

**Table 2:** A table of use cases for all dashboard functionality. Similar to our table in 3.1, this includes a name for the functionality (Title), a two-step layout of how it should work (Steps), and what the widgets needs to work and what it will produce (Entry & Exit Criteria). This served as a way to constantly check that our desired functionalities are implemented as correctly and fully as possible.

### 3.2.2 Determining Data Sources to be Visualized

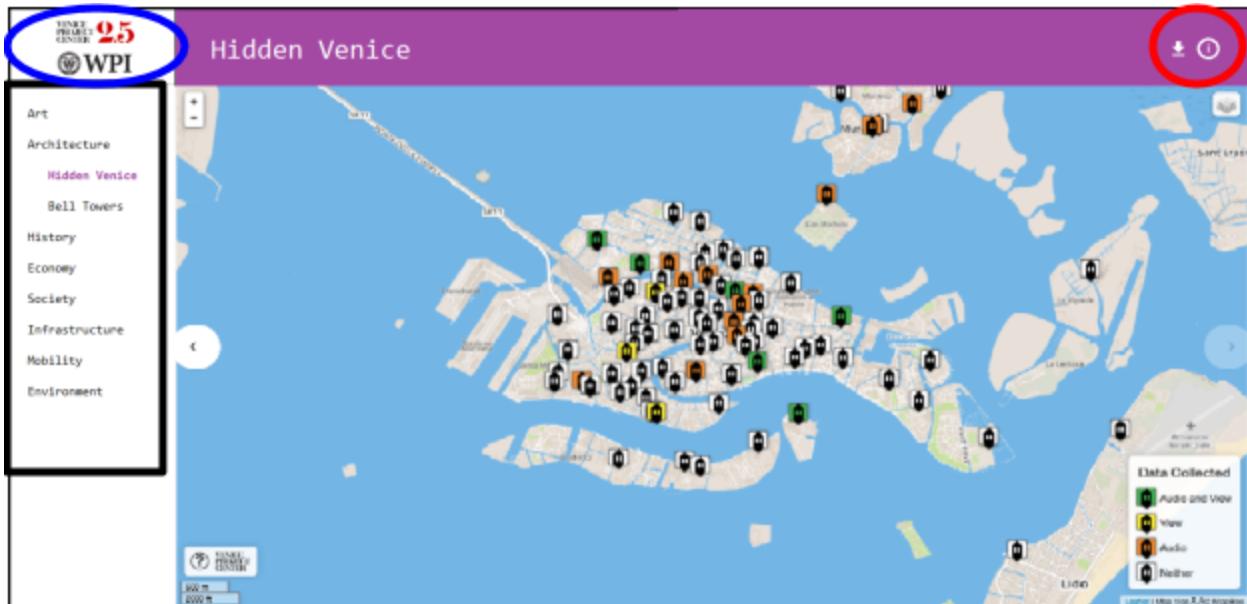
Our next step in the implementation of the new website was to decide which data visualizations to include. As it stood before and still stands, the VPC website contains a section entitled “Visualizations”, which contains 24 digital data visualizations created by Venice Project Center teams in past years.

Initially, we created a table highlighting each existing data visualization on Veniceprojectcenter.org, listing their apparent pros and cons, in an attempt to narrow down the existing visualizations to ones we deemed worthy of being transferred over to the new website. However, we ultimately decided to move each visualization on the website over to the new site. While the visualizations do vary somewhat in quality and

level of user interactivity, all of them provide useful information, and updating each and every one of them to more advanced versions proved to be beyond the scope of our project. Before considering new visualizations to create, we agreed to start by moving all of the existing visualizations on Veniceprojectcenter.org over to the new site.

### 3.2.3 Creating a Wireframe of the New Website

With both use cases and all visualizations finalized, we created a wireframe for our website. A wireframe is a drawing or primitive visualization of the website the developer plans on creating. It allows the developer to have a concrete image or roadmap of what code needs to be written. Our previous steps further this because we created several wireframes that display aspects of the functionality from the use cases on visualizations that already exist. One such wireframe is pictured below:



**Figure 21:** The first mockup of our big data visualization website. Again, this follows the structure of the wireframe in 3.1. The image serves to have a hard picture of how the website should look which helps guide the coding. For example, the image shows the menu of visualizations as a collapsible (black), which is very specific to code for. Some other key features are the download and info button (red), and the various logos (blue).

The above image represents a rough draft of our website that includes: a download link and about page in the upper right corner (circled red), a drop-down menu bar on the left (circled black), and a logo that links to the VPC website in the upper left (circled blue)

hand corner. The drop-down menu bar allows users to select whichever visualization they wish to view. Lastly, the “more info” button allows the user to get a sense of how the data was collected and displayed before they decided to use the download button to work with the raw data themselves. While the final website ended up looking reasonably different from the wireframe, this mockup guided us largely in our design. We will discuss elements of our mockup that remained, as well as ones that were changed, in the final website in the next section.

### 3.2.4 Creating a Platform Containing All Visualizations

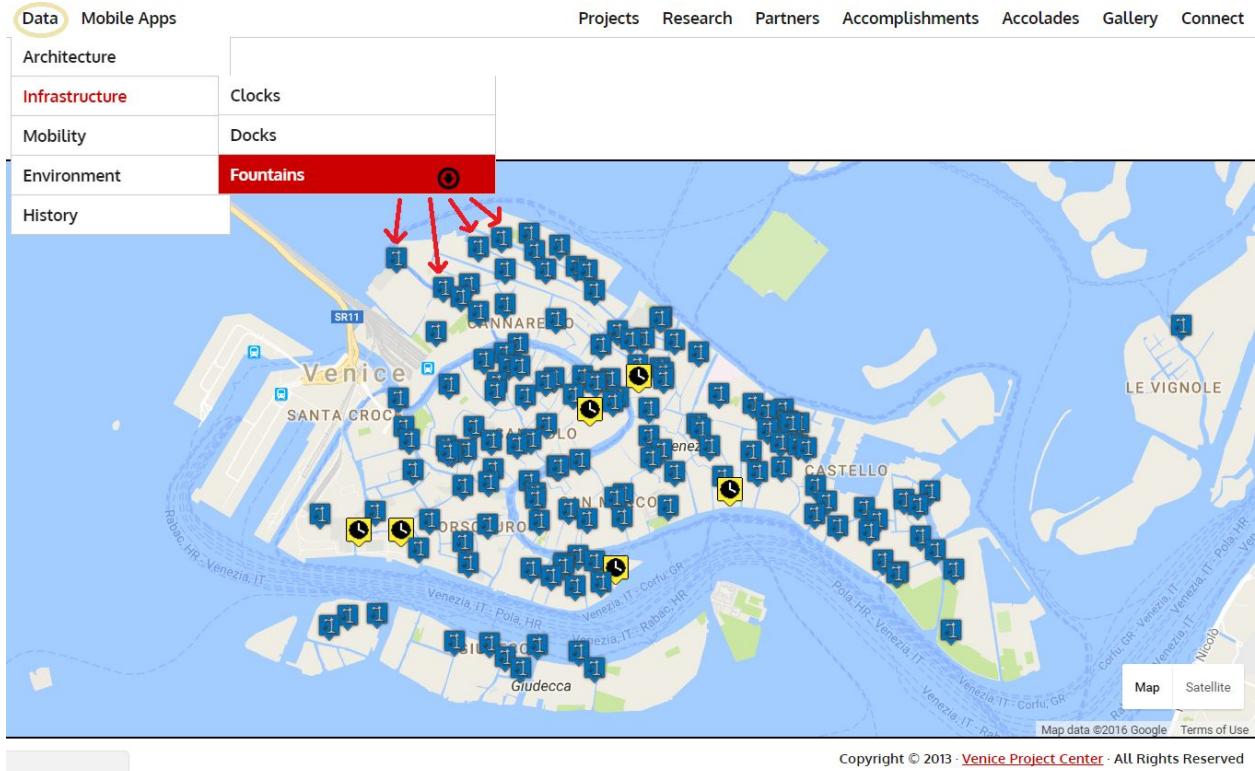
After going over our mockup with our advisors, we began writing the code for the final website. To create our website, we used the Materialize CSS framework (previously mentioned in Section 2.4.1), on a recommendation from Tomaso Minelli. Materialize is a library of HTML, CSS and JavaScript functions, which allows programmers to more easily program more modern-looking websites, with an emphasis on simplicity and user interactivity.

While refining the layout of our website, we ultimately decided to split the site into two primary sections: “Visualizations”, which provides some context as well as links to visualizations created by the VPC and Venice IQP teams in the past, and “Data”, which displays a digital map, upon which users can toggle layers of different points of interest in Venice, such as bridges, canals, bell towers and the like. These sections are modeled after their respective sections on the current Venice Project Center website, and will be expanded upon more in the Results section, specifically Section 4.2. To write our code, we used the Sublime Text editor, which is extremely useful for programming as it natively supports many common programming languages, and provides many features that work to neaten the appearance of code and to ease otherwise tedious aspects of coding (See Appendices).

While coding our website, we had several meetings with our collaborator Tomaso Minelli, who is responsible for much of the data logging and website upkeep of the VPC. Through these meetings, we were able to get a better sense of direction in terms of how

to write our code, particularly for the “Data” section. More specifically, while the map layer functionality does exist on the Venice Project Center website (as shown below), Mr. Minelli and the VPC envisioned the scaffolding for this mapping tool being reworked for the new website in a way that would make it easier for future groups to add layers to the website.

In order to help the VPC accomplish this, we utilized the Leaflet.js API. Leaflet is a JavaScript library for interactive maps; while it did exist when the map functionality was introduced on the VPC website, it has increased vastly in robustness and ubiquity since then. The map functionality on the VPC website uses the Google Maps API, which is more limited than Leaflet, providing justification for the switch to the new API. Additionally, we re-worked the back end so that it takes in its map data (geoJSON, JavaScript objects which specify map coordinates as some of their attributes) directly from Firebase, Google’s database system for developers, which is used by the VPC. This way, people adding to the website in the future will easily be able to add layers to the map by adding the geoJSON to Firebase, and making slight modifications to the code.



**Figure 22:** The “Data” tab, shown in the top left corner, leads the user to a list of drop-down menus containing layers, such as “Fountains” in this image, that can be toggled on and off. This functionality was transferred over to the new Venice Big Data website. However, we rewrote the back-end code so that future groups will have an easier time adding new layers to the map.

### 3.3 Visualizations: New & Remastered

Since big data has recently been proven to be a great resource, visualizations have become increasingly better to tap into big data’s stored knowledge. (Telea & Safari Books Online, 2015) To allow the Venetian public to have a more thorough access to information we have decided to create two new visualization. Both will be displayed on our new website, one of which working with completely new data and the other remastering old data from the VPC.

#### 3.3.1 Choosing Data Sources: New & Old

When choosing our data sources we focused on two main issues of Venice: the economy and tourism (Ross, 2015). After searching through the Venice Open Data (VOD) website, we were able to download budget spending data from the years 2013 to 2016. This was also a completely new data set that the VPC has yet to visualize. Also,

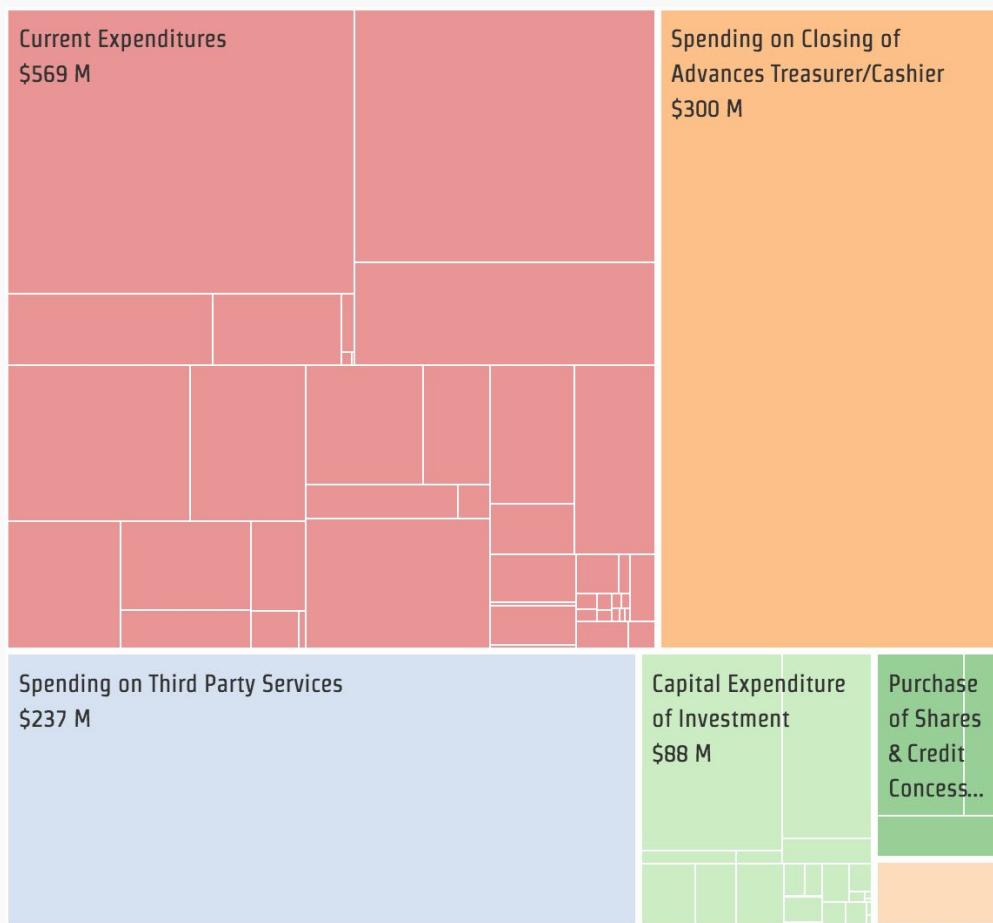
this detailed dataset of spending, in Figure 23 below, was chosen due to its several terabytes of data and large temporal range. Both of these qualities related back to big data and provided a great deal of information.



**Figure 23:** Above is the Tourist Arrivals Widget that inspired our dataset from the VPC. All sections that are colored (such as Air & Cruise) are collected in real-time from other widgets while the gray ones rely on approximations based on the other real-time numbers. However, this is just the front end, as the data is updated in the display, the older data is stored and logged acting as an untapped source for visualizations.

### 3.3.2 Pairing Data to Visualization Method

To create our visualization, we used a combination of D3.js, Google Sheets, and Firebase. One style of visualization we decided on was the tree graph, which relies on data being organized by several levels to show how the data points are related. This is best shown in Figure 24 below.

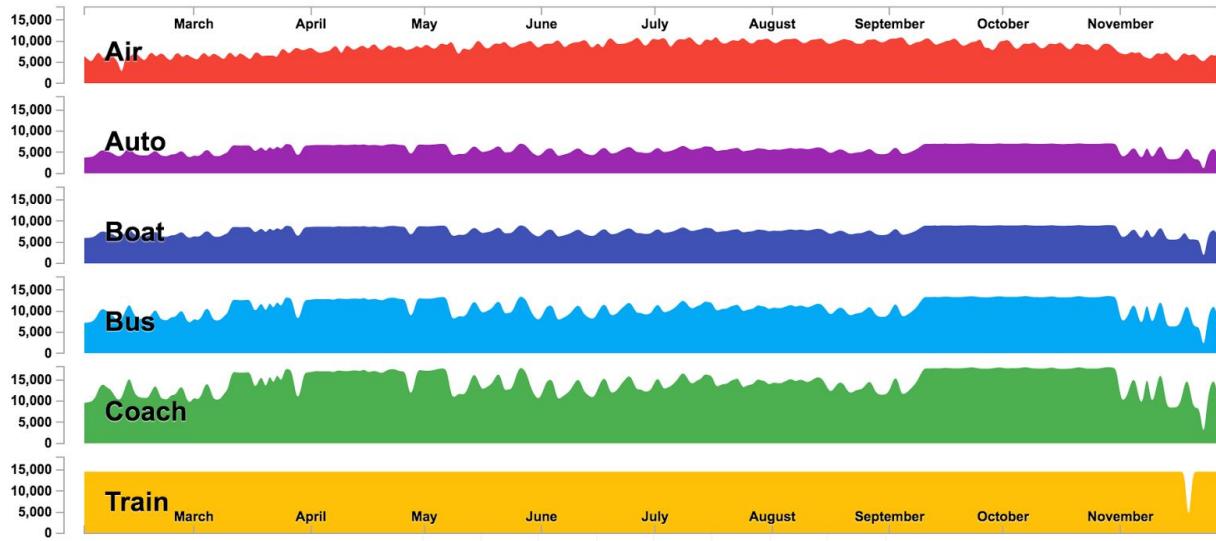


**Figure 24:** Above is a tree graph taken from the finished Venetian Spending Reports visual. As can be seen each major square represents a section of the budget and the size corresponds directly to the amount of money in that sector. Also, each square (where applicable) is made up several tinier square that represent the subsections of that major section. This is best seen in the “Current Expenditures” (Red) section where those square are also sized based on amount.

Ultimately, we related this sort of visualization to spending data. After looking through the several spreadsheets made available, we determined different Titles, Missions, and Statements sort the data. To create the tree graph, we relied on GitHub with contained much of the needed code (See Appendices). The other style of visualization we chose was an area line graph that is pictured in Figure 25 below.

Next, in order to find a data set related to tourism, we focused on existing data collected by the VPC. Throughout the VPCs collection of data related to tourism, we decided on data collected by the Daily Tourist Arrivals widget. As shown in Figure 25 below, this data is already collected from several sources and displayed as a widget.

However, this widget only displays recent data while collecting and storing data since its creation. Thus, transitioning this data to a new visualization will showcase how old data can be visualized while giving the user more information, as the data will be displayed across time.



**Figure 25:** This figure showcases the design of our area line graphs. As you can see each section of arrivals would get it's own section and color. From this graph the user can see trends in individual modes of arrival as well as compare different sections due to this organization. Ideally, the user will be able to also select a section of time to be examined as well. For example, the user would highlight a specific week and each graph would change to display just that data.

This format worked best as the graphs being directly above and below each other allows for direct comparisons. Also, the area graphic allows the user to clearly see the trends across any time frame as the graph increases and decreases.

### 3.3.3 Uploading Data to Visuals and the Internet

After we had decided on the visualizations to be used for each data set, we had to format the data such that the data would be displayed. In the case of the budget spending, the code acquired from GitHub included a python script that converts CSV files to JSON. After this conversion, the JSON objects are used to fill out the various sections of the graph. In order to utilize this script, we used Google sheet functions and add-ons to combine the separate sheets into a properly formatted file type. The organization of the process is described in Figure 26 below with pictures of the sheets

as they were worked on. First the data was parsed through for unique Title/Mission/Program combinations. The those combinations were used to sum all the expenses from those specific sections. Lastly the excel sheet was formatted to match the order required by the visualization, as shown in Figure 26.

Year	Title	Mission	Program	Chapter Description	Confirmed/Committed to Competence
2013	1	1		1 ACQUISTO BIGLIETTI E TESSERE DI CIRCOLAZIONE	0.00
2013	1	1		1 ACQUISTO BIGLIETTI E TESSERE DI CIRCOLAZIONE	0.00
2013	1	1		1 ACQUISTO BIGLIETTI E TESSERE DI CIRCOLAZIONE	0.00
2013	1	1		1 ACQUISTO BIGLIETTI E TESSERE DI CIRCOLAZIONE	0.00
2013	1	1		1 ACQUISTO BIGLIETTI E TESSERE DI CIRCOLAZIONE	0.00
2013	1	1		1 ACQUISTO BIGLIETTI E TESSERE DI CIRCOLAZIONE	0.00
2013	1	1		1 ACQUISTO DI SERVIZI PER FORMAZIONE E ADDESTRAMENTO	0.00
2013	1	1		1 ACQUISTO MATERIALE INFORMATIVO	0.00
2013	1	1		1 ACQUISTO MATERIALE VARIO	0.00
2013	1	1		1 ACQUISTO MATERIALE VARIO	0.00
2013	1	1		1 ACQUISTO TESSERE E BIGLIETTI DI CIRCOLAZIONE	0.00
2013	1	1		1 ALTRE SPESE PER SERVIZI NON SANITARI	214,713.01
2013	1	1		1 ALTRE SPESE PER SERVIZI NON SANITARI	0.00
2013	1	1		1 ALTRI BENI DI CONSUMO	30,951.56
2013	1	1		1 ALTRI BENI DI CONSUMO	28,730.25
2013	1	1		1 ALTRI BENI DI CONSUMO	7,003.52
2013	1	1		1 ALTRI BENI DI CONSUMO	6,616.46
2013	1	1		1 ALTRI BENI DI CONSUMO	4,917.00
2013	1	1		1 ALTRI BENI DI CONSUMO	1,990.29
2013	1	1		1 ALTRI BENI DI CONSUMO	186.66
2013	1	1		1 ALTRI BENI DI CONSUMO	6,715.23

+ ⌂   
 2013 ⏺   
 2014 ⏺   
 2015 ⏺   
 Unique Combos   
 Addition Function   
 Final Formatting   
 Functions ⏺   
  

**Figure 26:** Pictured in the excel sheet is the 2013 Expenses Data, where most of the work is done in several of the tabs highlighted in the bottom left. In the black box is the set of raw data (2013-2015). The red box contains the “Unique Combos” section which takes the various “Title”/“Mission”/“Program” combinations and filters keeps only the first occurrence. After that, the blue section is the “Addition Function” which uses the unique combinations to sum all the values that are connected to those combinations. Lastly, the green box is “Final Formatting” makes the final sheet format match the format needed for the visualization. The yellow box highlights “Functions”, which provides the excel functions utilized and how they work.

LEVEL1	LEVEL2	LEVEL3	TOOLTIP	SOURCE	SOURCE URL	2013	2014	2015	LEVEL
Current Expenditures	One	One				19499390.68	15555582.5	13177133.63	3
Current Expenditures		Two				3497875.56	3504427.12	3262182.55	3
Current Expenditures		Three				21403493.28	22560642.68	21870773.51	3
Current Expenditures		Four				88177517.02	84521484.23	84276577.38	3
Current Expenditures		Five				4752059.64	4955169.41	6065305.66	3
Current Expenditures		Six				9437132.41	9852881.62	8626670.24	3
Current Expenditures		Seven				6725899.29	6101839.52	6883396.47	3
Current Expenditures		Eight				10410006.65	10570245.89	12169016.77	3
Current Expenditures		Ten				5825350.09	5984539.18	5210120.27	3
Current Expenditures		Eleven				10126434.98	10202092.75	16588120.3	3
Current Expenditures		One Total				179855159.6	173808904.9	178129296.8	2
Current Expenditures	Two	One				5891642	4935538.96	3383860.91	3
Current Expenditures		Two Total				5891642	4935538.96	3383860.91	2
Current Expenditures	Three	One				26698815.97	24758889.66	22878894.01	3
Current Expenditures		Three Total				26698815.97	24758889.66	22878894.01	2
Current Expenditures	Four	One				13674584.8	13660725.17	13062753.56	3
Current Expenditures		Two				4457856.15	4527698.48	4797385.79	3
Current Expenditures		Six				8210817.69	8152641.73	7443272.67	3
Current Expenditures		Seven				1110377.02	1136086.85	1030904.55	3

**Figure 27:** This figure depicts the final formatting of our data before it is exported as a CSV (Comma Separated Values) file. LEVEL 1,2,3 correspond to Title, Mission, and Program respectively and the numbers from the Addition Function are listed under the 2013, 2014 and 2015. In the middle are Tooltip, Source and Source URL. Tooltip refers to any information that would be displayed when the box is hovered over and the Source and Source URL are for crediting the data source (which will be done at a different section of the visual). Lastly, there is a LEVEL section which determines how embedded in the tree the data should be. 3 refers to a sub-subsection, 2 refers to a subsection and 1 refers to the main section that is displayed at the start of the graphic.

With regards to the tourist data, we worked closely with Tomaso Minelli to translate the stored log data from JSON to CSV. The data was initially stored on the VPC's Firebase and was immediately ready as a JSON file. With the JSON file we used a python script to parse through the year's worth of data for total tourist arrivals. The python script, which in our case, is a code that accepts a JSON file and converts it to CSV quickly created a correctly formatted file for use. After the file was created we attached it to the visualization. Unfortunately, since the framework of the VPC has been reworked since Tomaso has arrived, the data collected is not entirely accurate so this visualization serves as a proof of concept. Specifically, it shows that the data collected by widgets can be used for visualizations.

## 4.0 Results & Analysis

---

Within this section, we will discuss the different tools we built throughout the course of our project. Ultimately, we were able to produce a website dedicated to the display of Venice Project Center data visualizations, an updated Venice City Knowledge dashboard displaying real-time data about the city, and visualizations of Venice-related data from the both the VPC's and VOD's datasets. Accordingly, we will divide this section into three subsections, in a similar manner as our methodology.

### 4.1 Venice City Knowledge “Dashboard 3.0”

One major deliverable for our project is the Venice Dashboard 3.0. The Venice City Dashboard is easily the VPC's most visible projects with newspaper stories being written about it almost daily. In this section we discuss our contributions.

#### 4.1.1 General Design

We kept the overall feel of each widget very similar, but provided a more readable font and bold colors for widget toolbars. Additionally all the new widgets resize to match the bounds of the widgets instead of being fixed sizes.

#### 4.1.2 Twitter

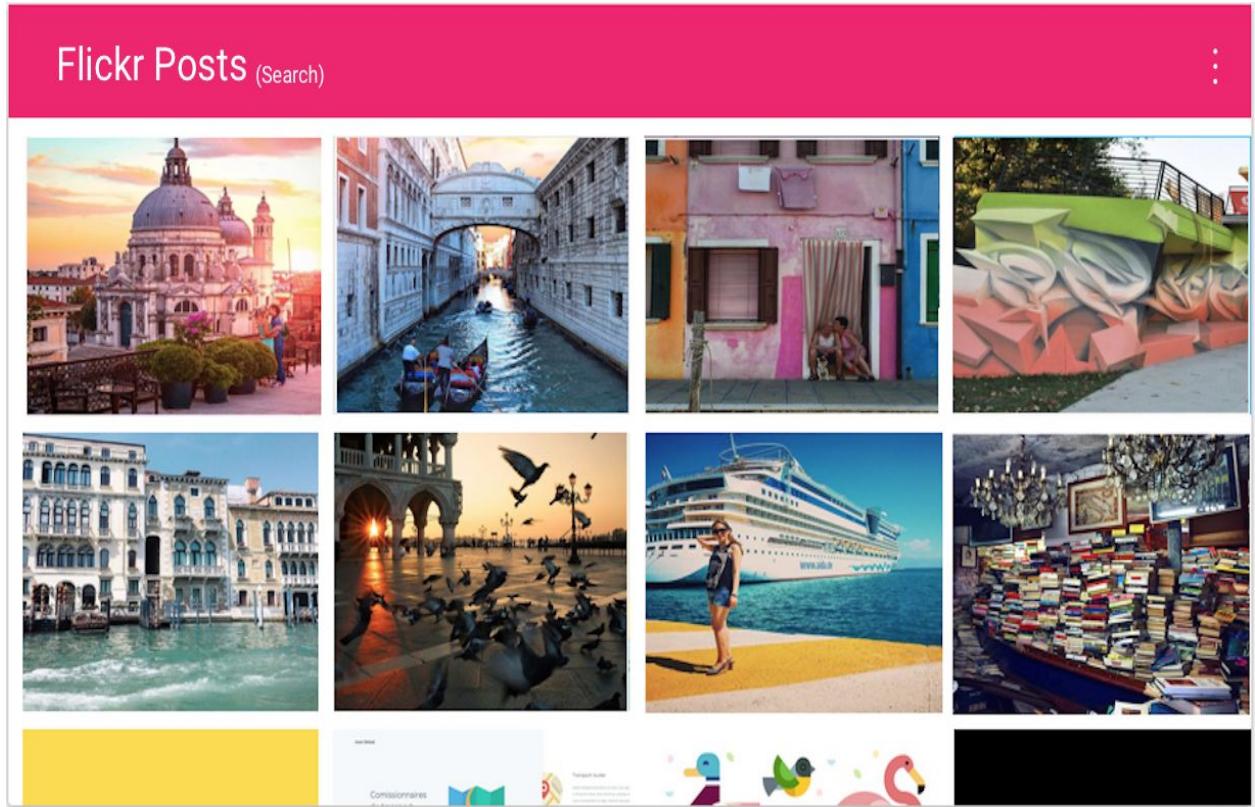
The Twitter widget allows users to scroll through Twitter trends for Venice. Each trend is scaled proportionally to how popular the trend is in Venice. Additionally, each trend can be opened in Twitter to see what Venetians are saying out it.

The image shows a mobile application interface. On the left, there is a 'TWITTER TRENDS' section with a blue header. Below it, a list of trending topics is displayed, each with a small icon. The topics include 'Ronaldo', 'Diego Costa', '#Gentiloni', '#TorinoJuve', '#ITALYISPROUDOFLouis', 'Gonzalo', 'Bonucci', '#domenicalive', '#inmezzora', '#PalermoChievo', '#arenagiletta', and '#bimbumbam'. A green arrow points from the '#Gentiloni' entry in this list to the right side of the screen. On the right, there is a detailed view of a tweet from 'Paolo Gentiloni' (@PaoloGentiloni). The tweet features a photo of him and several other people wearing blue hats. Below the photo, his stats are listed: 5,556 tweets, 548 following, and 105K followers. The tweet itself reads: 'Ministro degli Esteri e della cooperazione internazionale Italian Foreign Affairs Minister'. Below this, three other tweets are shown from users 'mattia feltri', 'Michele Dell'Orco', and 'AlessandroDiBattista', each with their own profile picture, name, and a brief message.

**Figure 28:** The Twitter is showcased to the left. As can be seen, the titles of the trends are sorted by size where the larger size corresponds to more mentions. From this widget the user can click on a link that will bring the user to the page associate with the topic.

#### 4.1.3 Flickr

The Flickr widget displays the most recent pictures people have posted from Venice. Users can scroll through and see what spots are popular for tourists to photograph while visiting Venice.



**Figure 29:** The Flickr widget displaying recent photos.

#### 4.1.4 Facebook

The Facebook widget pulls a summary of posts from the pages of many respected Venetian opinion makers and displays them for users to scroll through. Each post also provides a link to read the full post if the user wishes. This widget provides an easy way for users to see what is important in Venice right now without having to find Facebook pages themselves.



**Figure 30:** The Facebook widget collects posts from opinion makers in Venice such as Reset Venezia, the VPC and Lo Shitto. These post are then displayed with a picture, title and quick description. After the description the user can be prompted to "...See More" which will send them to the actual post.

#### 4.1.5 Word Cloud

The Word Cloud Widget could be thought of as a more condensed version of the Facebook widget. The word cloud widget takes data from all the Facebook posts mentioned in the Facebook Widget and counts each unique word and displays each word scaled proportionally to its frequency. Users can see important trends emerge from these opinion makers as fast as they can post to Facebook.



**Figure 31:** This is a word cloud generated from Facebook posts. As can be seen, it also takes from local opinion makers from the Facebook widget.

## 4.2 Venice Big Data website

Our second primary deliverable was the Venice Big Data website, which can be reached at <http://bigdata.Veniceprojectcenter.org>. Here, we will discuss each aspect of functionality that we included on the site, and go into detail about how these functional elements will benefit users.

### 4.2.1. General Design

The website is split into two main sections: the “Visualizations” section and the “Urban Elements” section. These sections are parallel to their respective sections on the Veniceprojectcenter.org website. Intuitively, “Visualizations” refers to the data visualizations being shown, which are presented through a variety of mediums. The front page of this section of the website is the section that the user is brought to first. This section of the website is dedicated to the display of data visualizations created by past, present and future Venice Project Center teams, and we will expand on specific elements of this section’s functionality in Section 4.2.2.

“Urban Elements” in this instance refers to map data, specifically data collected by the VPC that highlights the locations of different aspects of Venice. As stated in our methodology, we strove to transfer over the map functionality from the “Data” section of

Veniceprojectcenter.org, which allows users to toggle different layers on top of a digital map of the island of Venice. Ultimately, we were able to do so, and we will provide a more in-depth analysis of this section in Section 4.2.3.

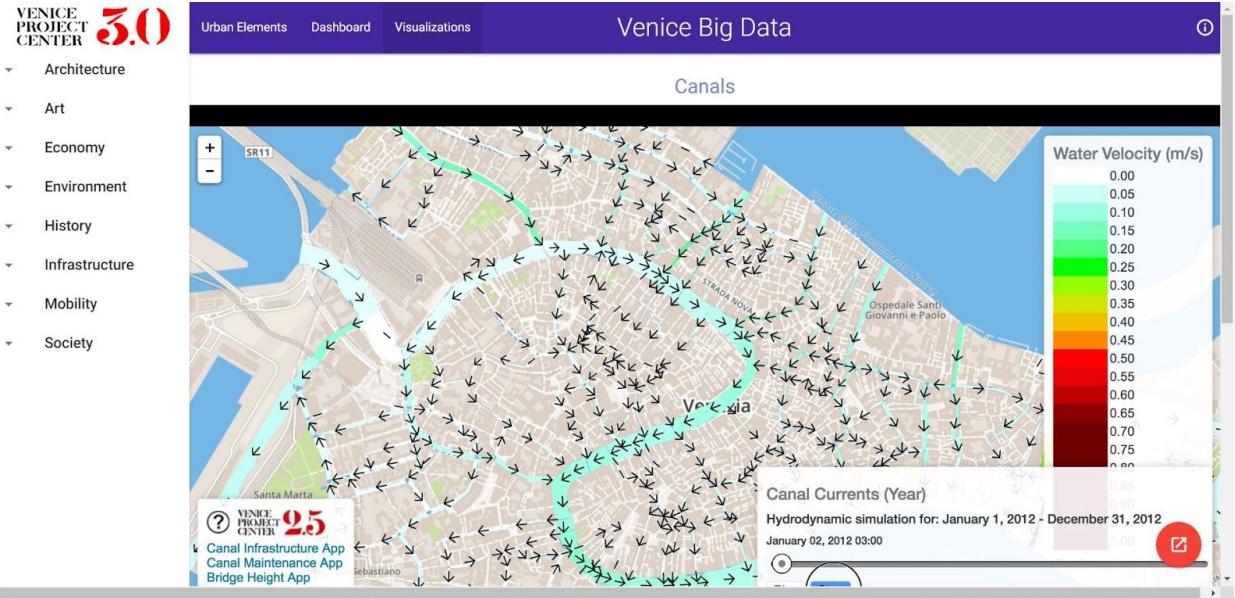
Additionally, the website does include a section entitled “Dashboard”, next to “Urban Elements” and “Visualizations” on the top bar. Because the real-time dashboard, which was focused on in Section 4.1, is a separate entity from the Venice Big Data website, this link simply opens the dashboard in a separate tab. However, because the updated dashboard was a primary deliverable of our project, as well as a reliable source of Venice-related big data, we deemed it important enough to be linked to from our website.



**Figure 32:** The navigation bar at the top of each page on the website. The two main sections of the website, “Visualizations” and “Urban Elements”, as well as a link to the updated City Knowledge dashboard, can be seen. On the right, an information button can be seen, which causes an information modal to pop up on-screen.

#### 4.2.2 “Visualizations” section

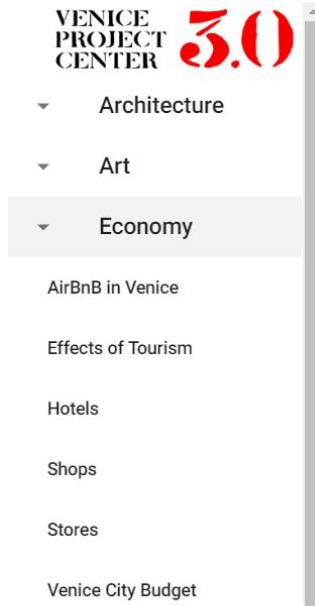
One primary section of the Venice Big Data website is the “Visualizations” section. When users select this section, they are brought to a front page, as shown below:



**Figure 33:** An example “visualization of the day”. As of currently, the website selects a random visualization and displays it on the front page each time it is loaded. Users can see a screen capture displaying how to interact with the visualization, and the floating action button (FAB) in the bottom right corner, when clicked, will open the visualization in a new tab.

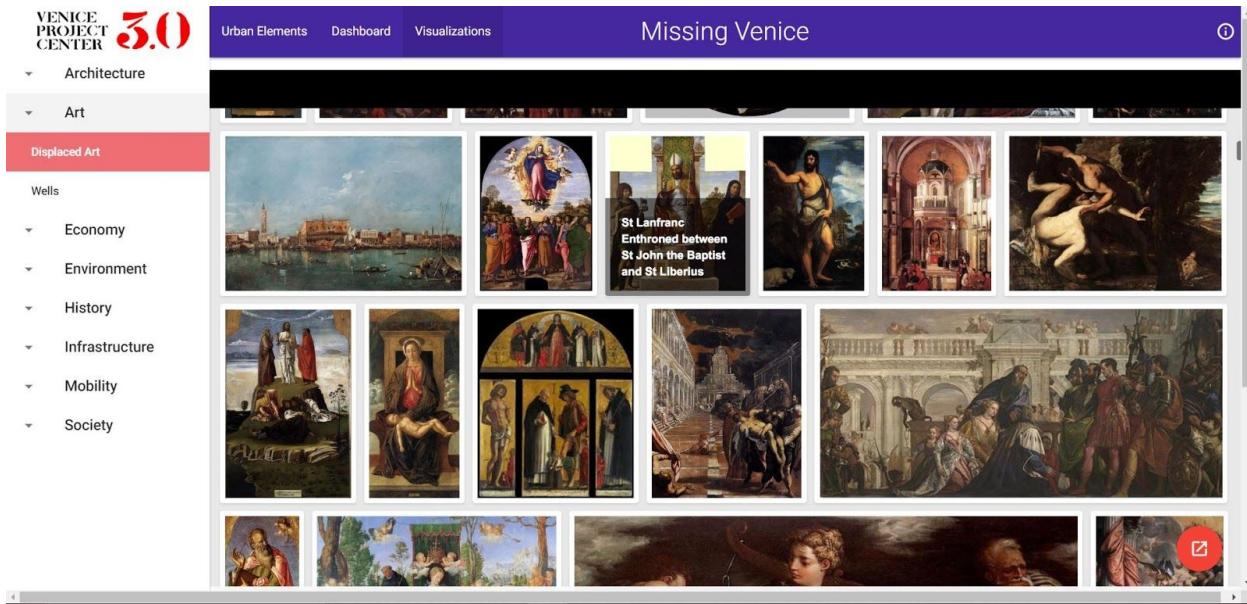
As such, the front page of the “Visualizations” section contains the title of the section, and points the user to the left of the page. Additionally, there is a section on the front page entitled “Visualization of the Day”, which we will address in depth in Section 4.2.3.

To the aforementioned left, the sidebar lists each data visualization that was previously linked to on [Veniceprojectcenter.org](http://Veniceprojectcenter.org), as well as the data visualizations created by IQP teams in 2016. The visualizations are organized according to their respective general topics, as they were previously organized on [Veniceprojectcenter.org](http://Veniceprojectcenter.org). The general topics are architecture, art, economy, environment, history, infrastructure, mobility and society. On the new website, however, the general topics, as well as the titles of the visualizations contained within them, are alphabetized in order to allow the user to more easily find specific visualizations. This simple aspect of our design should prove to be especially helpful in the future, when more visualizations are inevitably added to the website. This sidebar is shown below:



**Figure 34:** The sidebar for the new visualization website. This serves to showcase how the visualization can be organized to increase accessibility.

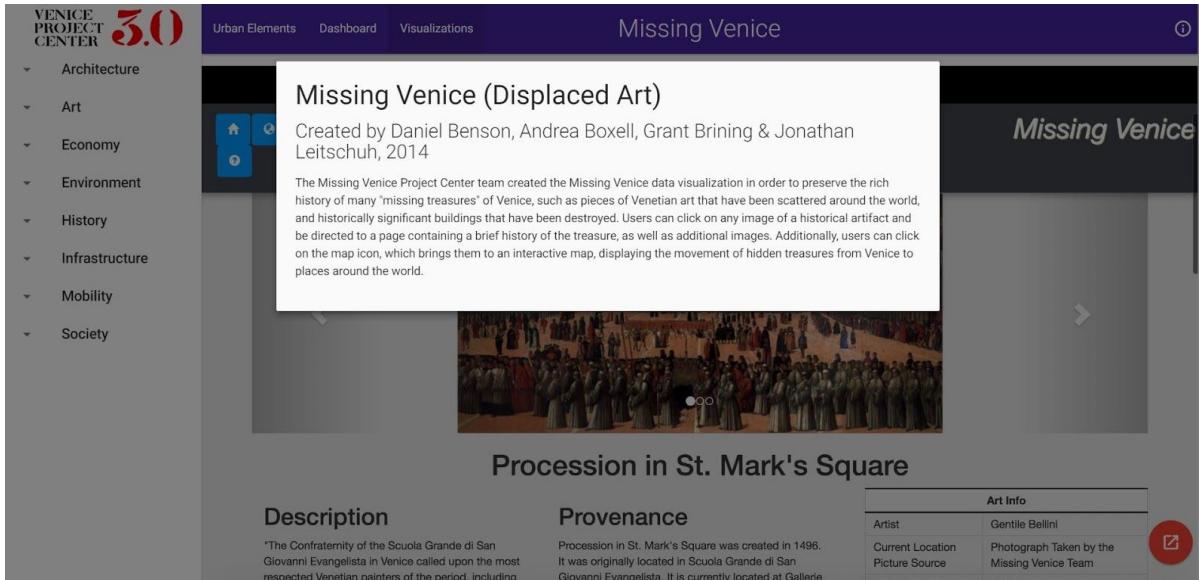
When users select the general topic they are interested in, a dropdown menu appears, listing all relevant visualizations in that particular domain. Users are then free to click on any visualization that they are interested in. Upon doing so, they are brought to a page on the website dedicated to that specific visualization, as shown:



**Figure 35:** Selecting a visualization from the menu opens a page displaying a brief tutorial video displaying how to interact with the visualization.

As seen, the page is titled identically to the visualization it represents (in this instance, bell towers). The primary focus of this page is a short screen capture, which is intended to give the user a visual understanding of the visualization's functionality. This screen capture is programmed to play as soon as the page opens, but it can be paused, stopped and replayed at the user's will. A floating action button is present in the corner of the screen on each visualization page, which allows users to open the visualization in question in a new tab when they are ready to view it.

Upon clicking the info button, users are shown a brief textual description of the purpose and functionality of the data visualization, as well as the authors of said visualization, as shown below:



**Figure 36:** When the information button (the circled “i”) in the top right-hand corner of the screen is clicked, a modal pops up on the screen as such, displaying textual information about the data visualization in question. A “modal” is the term for this type of pop-up box in web design.

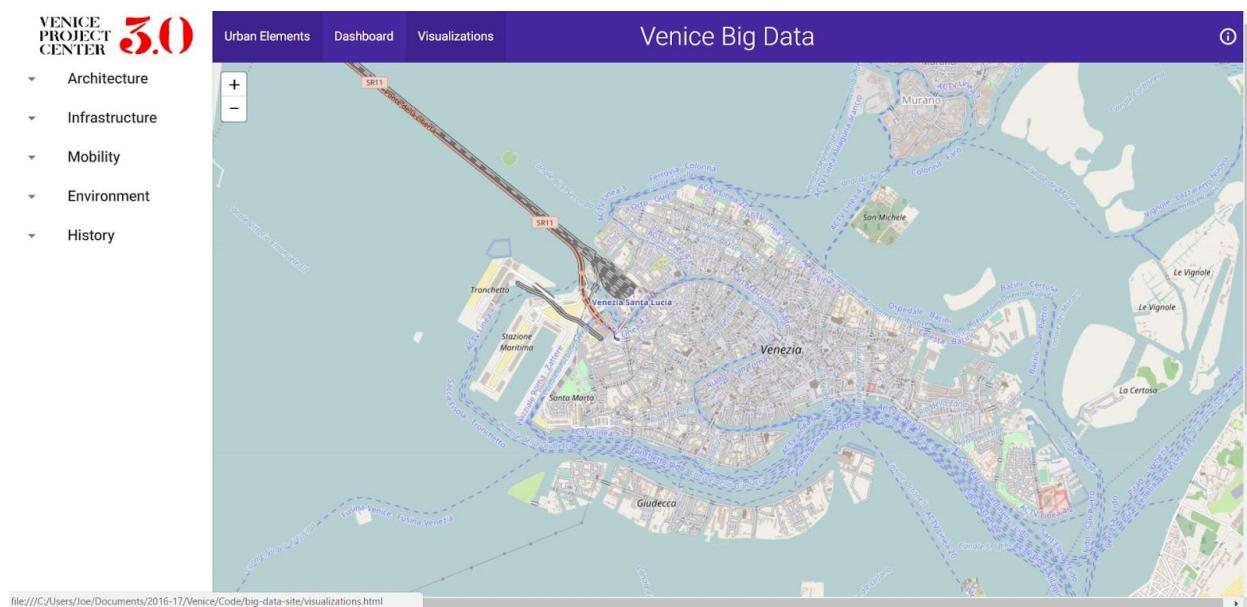
It is necessary to justify the purpose of these “buffer” pages, which exist between the user’s selection of the visualization on the website and the visualization itself, because a logical argument against this design would be that users could simply select the visualization, which would simply direct them to the visualization itself. A notable issue with this approach involves the long loading times characteristic of many visualizations. As many of the existing visualizations display large quantities of data in complex and abstract formats, some of them take a long time to load, in some cases, over two minutes. By including this buffer page, the user is very quickly able to see what the visualization does, and can reliably gain some general information about the visualization, before deciding to go ahead and interact with it.

Additionally, this design approach allows the visualizations to be rooted in some context. Previously, on the [Veniceprojectcenter.org](http://Veniceprojectcenter.org) website, users selecting the “Hidden Venice” visualization for the first time would be unaware what they would be looking at, knowing little information besides the title of the visualization. Past project groups spent large amounts of time working on the data visualizations in question, and as such, they

deserve to be rooted in some context, so that users can be aware of what they will be looking at, and how to interact with it.

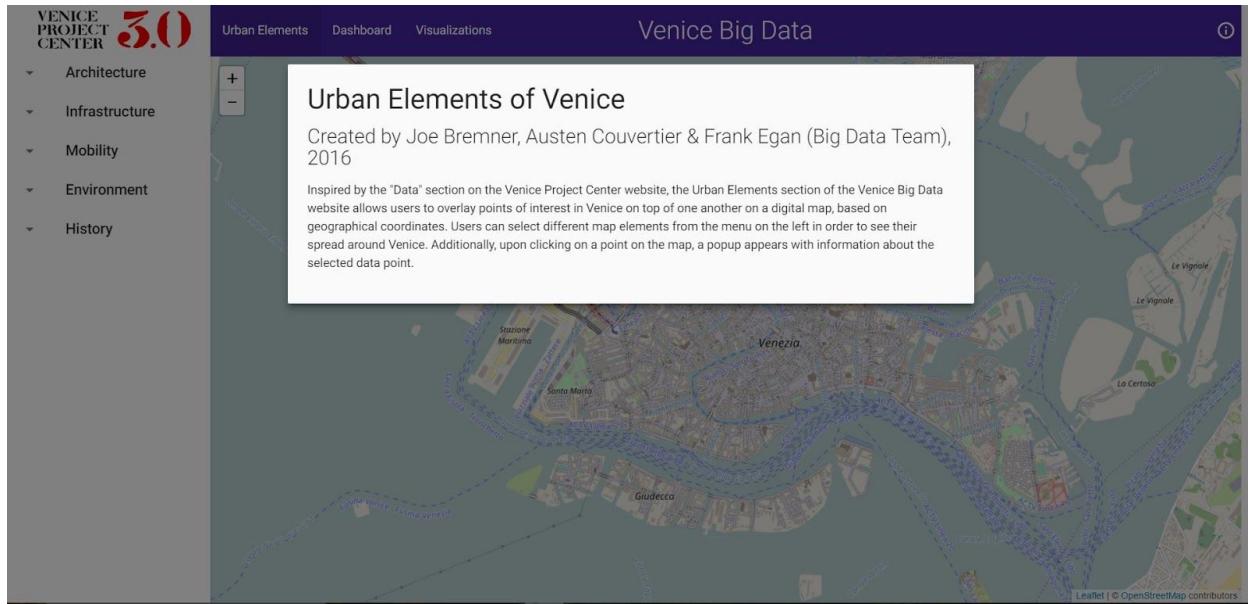
#### 4.2.3 “Urban Elements” section

The second and last main section of our website, entitled “Urban Elements”, is an ongoing repurposing of the “Data” section of the main VPC website. When the “Urban Elements” button is clicked on, the user is brought to a page that displays a visual map, which is automatically set to display the island of the historical center of Venice, but can be zoomed out and moved around to show other areas as well.



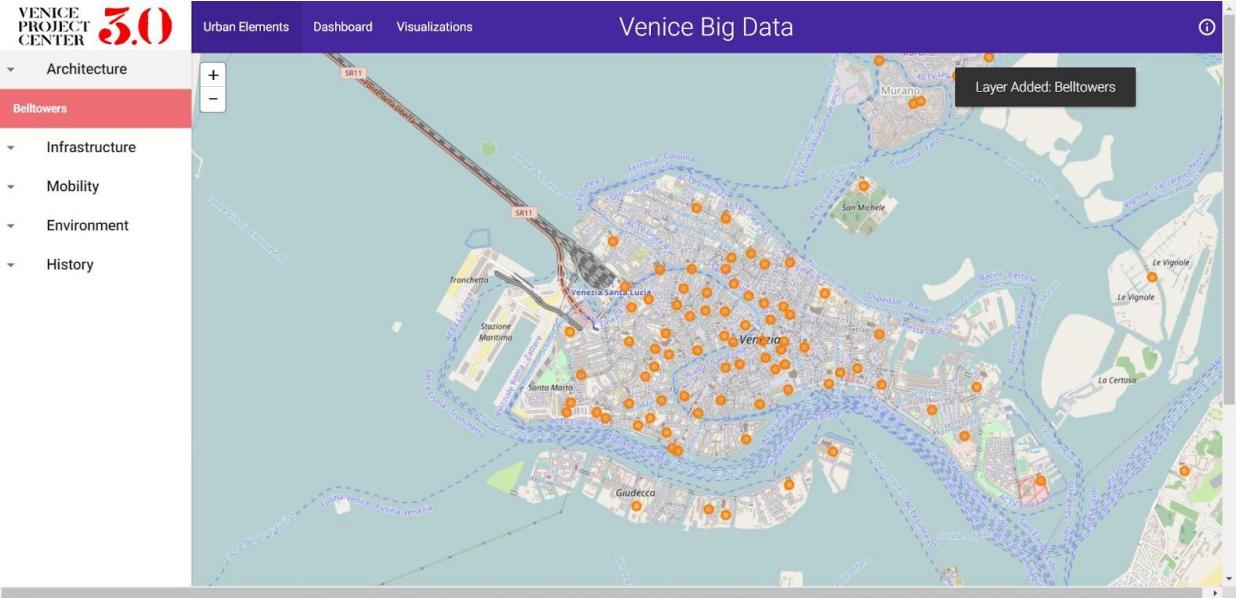
**Figure 37:** The front page of the “Urban Elements” section of the website. A digital map takes up the majority of the screen, with the menu of urban elements on the left. Users can zoom in and zoom out of the map using a mouse, touch pad or the plus and minus buttons in the top left corner, and can click and drag the map around as well. Regardless, the map will always default to this view of Venice.

When users click on the information button in the top right-hand corner of the screen, a modal pops up on the screen, displaying information as such:



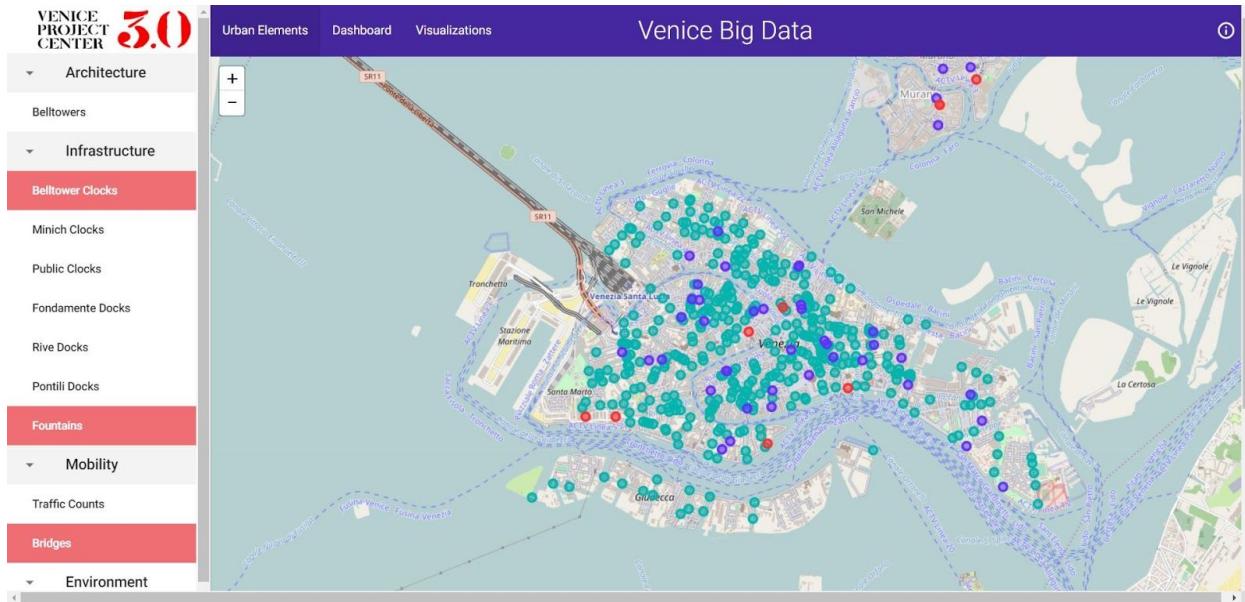
**Figure 38:** The modal dialog for the Urban Elements section of the Venice Big Data website.

Similarly to the Visualizations section of the website, users can click on any of the menu elements at left, causing them to drop down. In the Urban Elements section, however, the menu elements are layers that users can overlay on top of one another on the digital map.



**Figure 39:** By selecting layers from the left-hand side of the screen, the selected layer turns red, the points of interest show up on a map based on their geographical coordinates, and a “toast” pops up in the top right hand corner of the screen for a moment, simply stating that a layer has been added. Likewise, when users click these buttons again, the button turns back from red to white, the layer is removed, and a toast pops up on the screen, this time informing the user that the layer has been removed.

Users can also select multiple map layers from different sections to be shown at one time. This way, users can see how far away from one another certain locations are, and get a visual perspective of where places are in relation to one another.



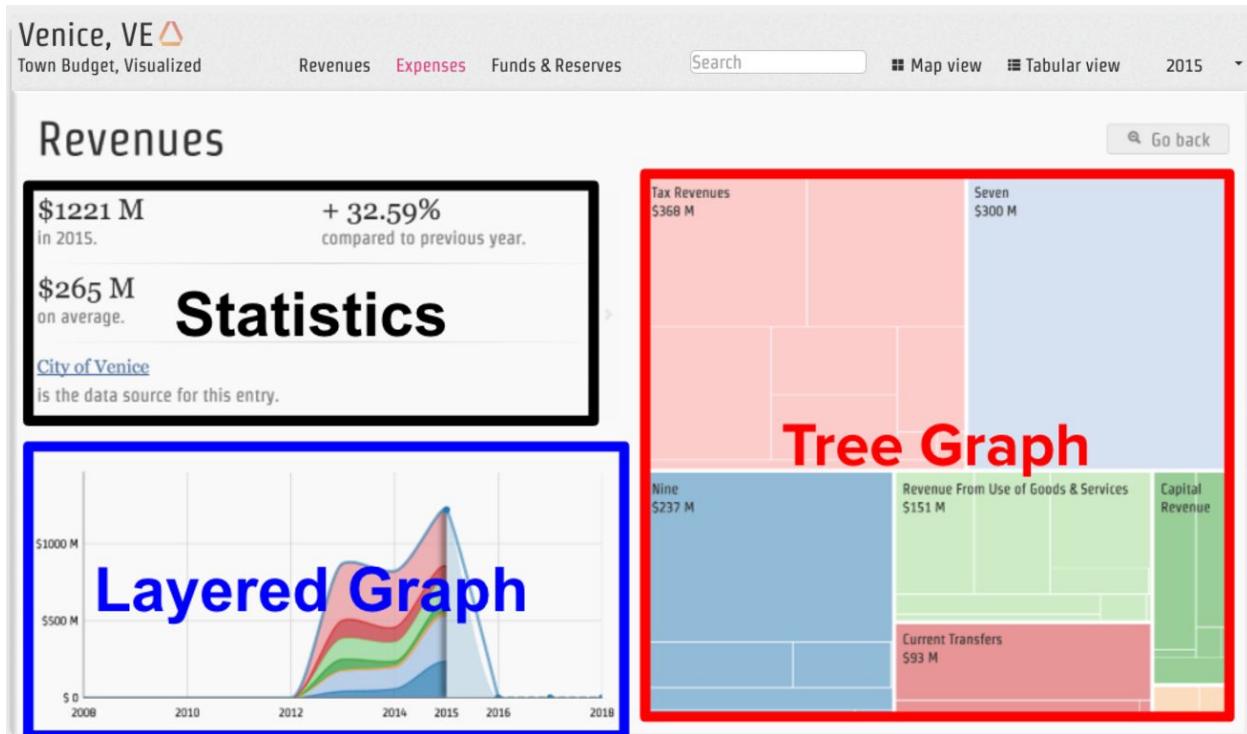
**Figure 40:** Here, the user has selected multiple layers from the menu on the left, which are identified by being highlighted red. Each layer is identified by a different color on the map.

While we were not ultimately able to add any more mapping features beyond the ones discussed due to time constraints, we will be discussing potential improvements that can be made to this feature in the future in our recommendations in Section 5.

### 4.3 New Visualizations

Our third deliverable culminated in the production of visualizations for the Venetian Spending Report (2013-2016) and Daily Tourist Population Data. These two sources were chosen because the new spending data highlights the versatility of visualization techniques with new data sets. While, the existing tourist data showcases how new visualization techniques can be used to derive more knowledge from data previously visualized. This section serves to outline how the final products are designed and describe their various features.

#### 4.3.1 General Design - Venetian Spending Report



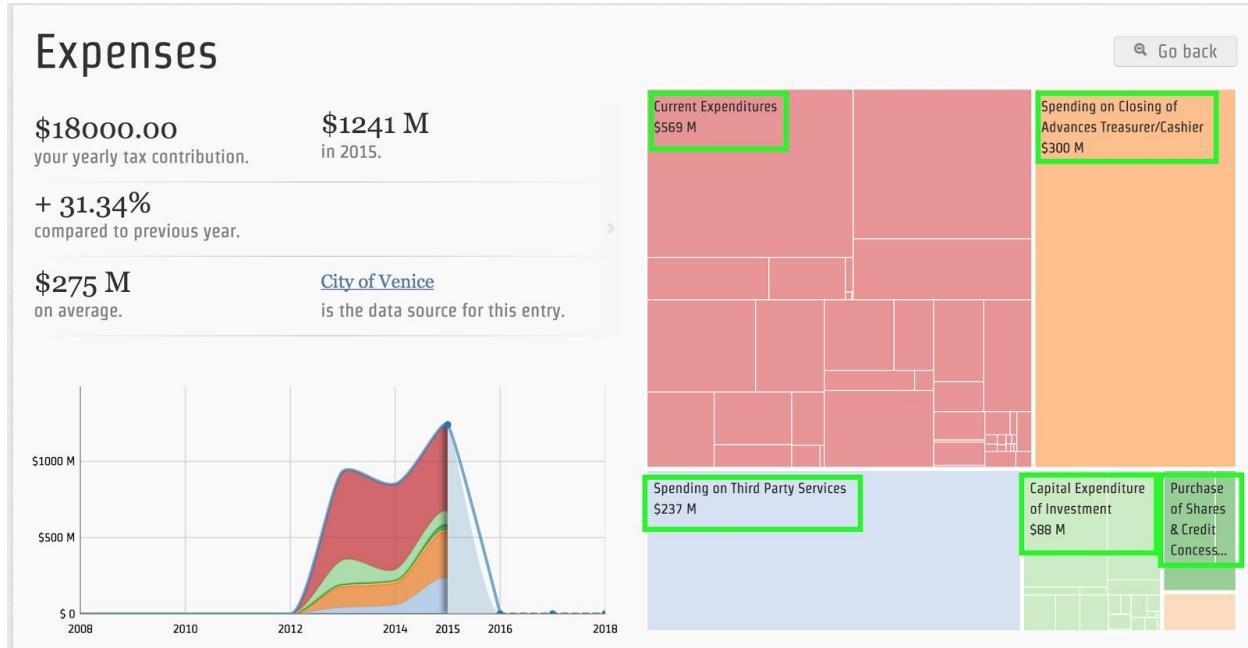
**Figure 41:** This figure depicts the main screen of the Venice Budget Visualization. In the bottom left corner there is a layered line graph that is color coordinated to the different sections of the budget with the statistic section above it. On the right half of the screen, the same data is displayed on the tree graph with section titles and totals displayed.

As can be seen above in Figure 40, the visualization of Venetian spending is a combination of a tree graph and layered line graph. In the upper right hand corner there is a toggle option for “Map View”, “Tabular View” and a year selector. Also, toward the top center of the display there is tab system for “Revenue”, “Expenses”, and “Funds & Reserves”. On the left side of the display, there is a link to our data source ([dati.venezia.it](http://dati.venezia.it)), a link to our website from 4.2 and a changing display of statistics. Lastly, most code related to this display was derived from GitHub.

#### 4.3.2 User Interactivity

On top of the aesthetic design, we included three layers of user interactions to ensure the most information can be extracted from the budget data. These layers include: the branched tree graph, sliding layered line graph, user tax input.

The main area of interactivity comes with the branched tree graph which is initially broken up by the 7 different “Title” sections of the budget data as can be seen in Figure 41.



**Figure 42:** A screenshot of the Expenses section of the budget visualization to highlight the section. Each section, “Revenues”, “Expenses”, and “Funds”, are comprised of 7 different Titles, however the specific names are different for each section.

From each section, the user is able to click on the area and they will be redirected to another tree graph that lays out the “Mission” section that makes up the initial “Title” section. Also the user can hover over any section of the tree graph and it will prompt a quick description of the type of spending associated with that section. This transition from “Title” to “Mission” can be seen in below.

## Funds

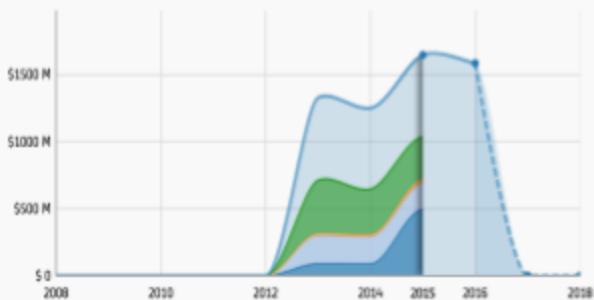
Go back

\$1647 M  
in 2015.

+ 24.05%  
compared to previous year.

\$528 M  
on average.

[City of Venice](#)  
is the data source for this entry.



## Two

Go back

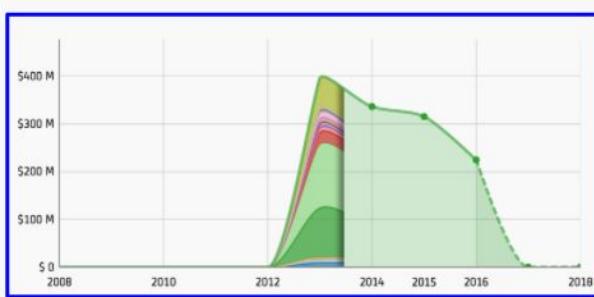
\$397 M  
in 2013.

+ 100%  
compared to previous year.

29.86%  
of total funds.

\$116 M  
on average.

[City of Venice](#)  
is the data source for this entry.



**Figure 43:** As can be seen the top picture showcases the tree graph and the “Two” section (red). Once the user has clicked onto the “Two” section the screen changes to depict the lower image of the visual. The “Two” is now the title and the tree graph is changed to show the subsections of two (red). Also, the layered line graph is updated to show how the subsections of Two have changed over time (blue).

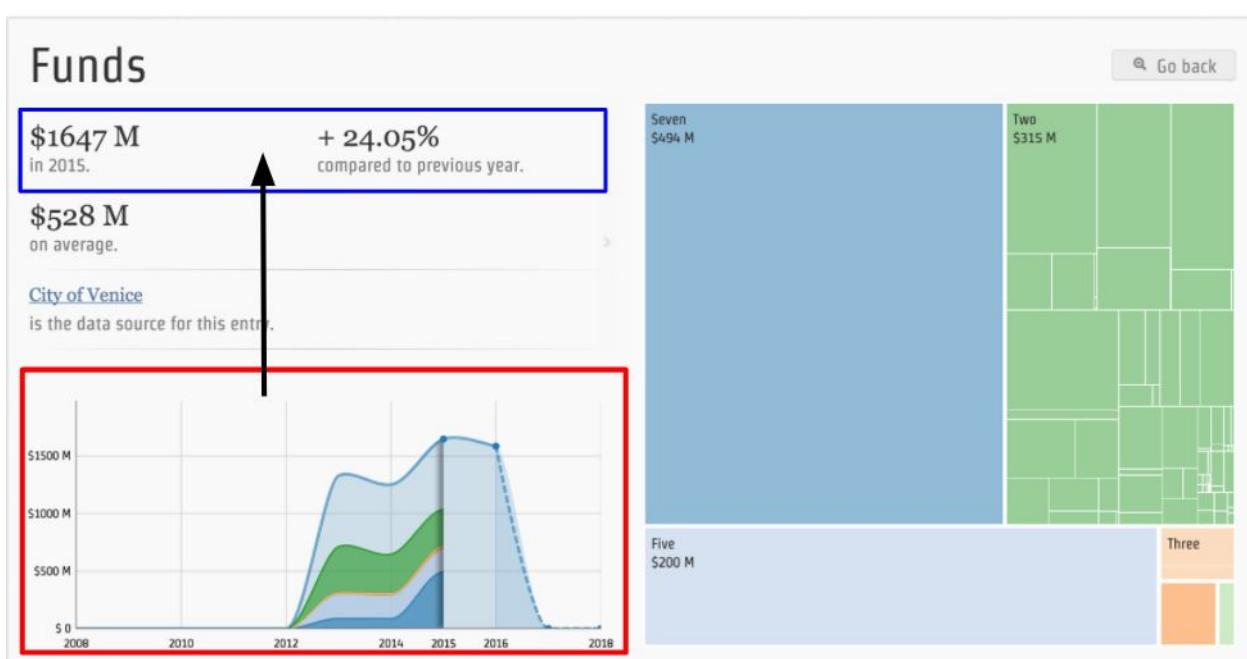
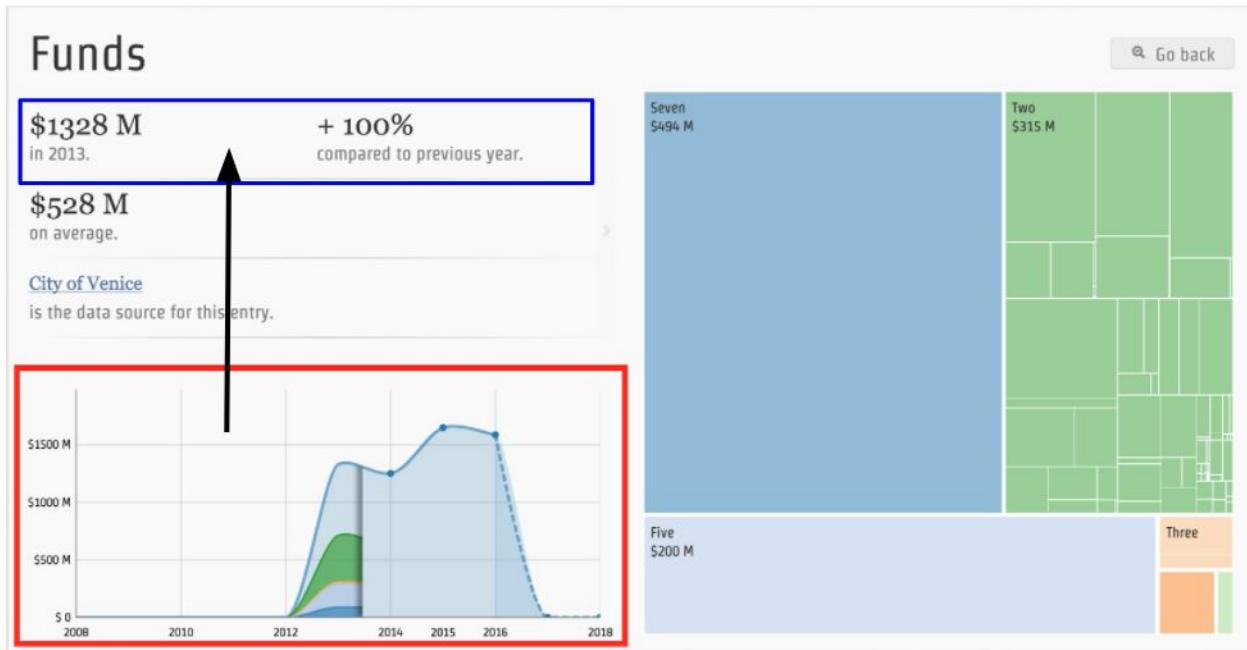
Also, the display can be changed from a treemap to individual lines, known as “Tabular View”. This feature, shown in Figure 42, displays how individual subsections of the budget have change over time. Lastly, the user can also toggle between “expense”, “revenues”, and “funds & reserves” which change the data being presented on the treemap.

Name	Growth	Impact	Amount
▶ Funds	+ 24.05%	100%	\$ 1,647,167,847 ➤
▶ Seven	+ 81.98%	29.99%	\$ 493,950,000 ➤
▶ Five	0%	12.14%	\$ 200,000,000 ➤
▶ Four	- 44.38%	0.74%	\$ 12,265,602 ➤
▶ Three	+ 88.7%	0.83%	\$ 13,669,022 ➤
▶ Two	- 6.62%	19.15%	\$ 315,491,004 ➤
Zero	+ 100%	0.24%	\$ 3,982,373 ➤

**Figure 44:** The above image highlights how the tree graph can be changed to spark lines for each section subsection and subsubsection of the budget. Moving the small black dot, which will change the statistic for “Impact” and “Amount” for the year selected, can interact with these sparklines..

All these varying features working with the tree graph allows the user to quickly views the exact expenses or sources of revenue the government has reported for the past 3 years.

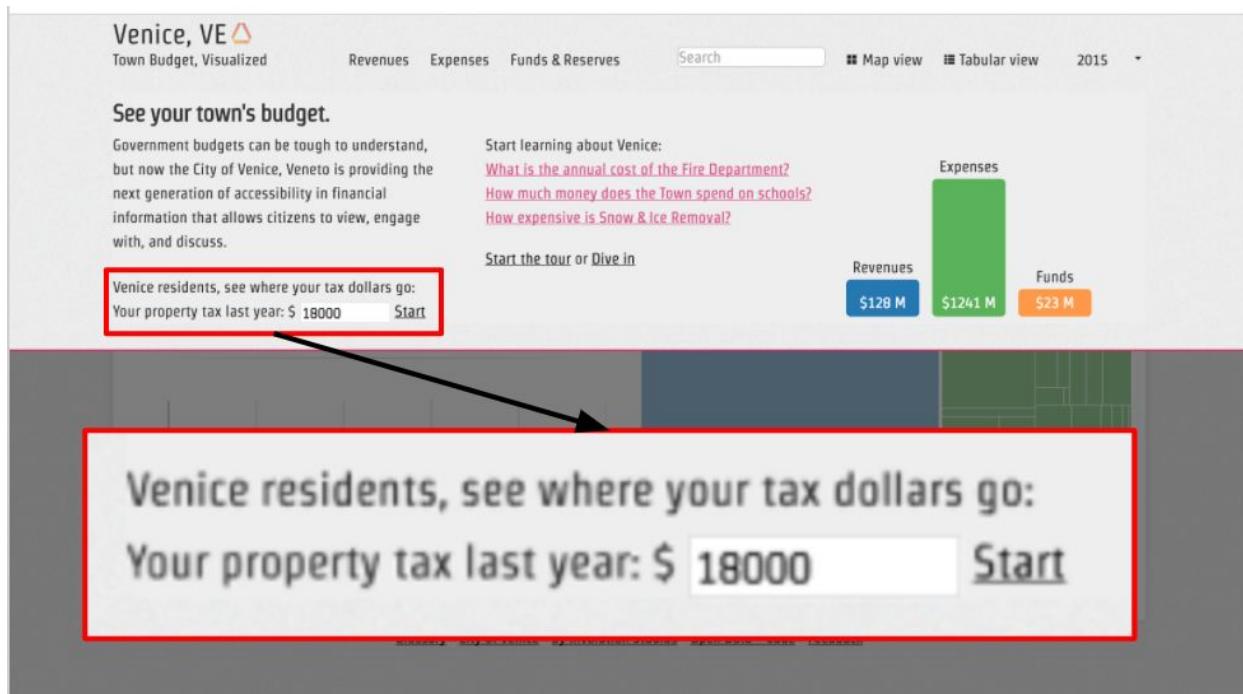
The next area of interactivity is the sliding layered line graph which is color coordinated by the major “Title” sections of the budget data. This allows the user to change the data being analyzed to any year from 2013 to 2016 as seen below.



**Figure 45:** The top screen depicts the layered graph (red) set to 2013 and the corresponding statistic data (blue is highlighted). Below shows how the graphic changes when the user slides the graph to the 2015.

Ultimately this sliding feature changes statistics that appear on the center right side of the display that compare budget data by year. This allows the user to see how the government has specifically changed the size of the budget throughout the years.

Lastly, when the user first interacts with the visualization, they are given the option to fill out their “yearly property tax”. From this single bit of data, the visualization is able to determine how your tax dollars are spent across all fields. This analysis is not limited to the main “Title” sections, as it will track through to even the subsection of “Mission” and “Program” as depicted below.



**Figure 46:** The above visual shows how the user inputs and then sees their tax contributions. Upon opening the visualization, the user is prompted to enter their tax data last year (which is not saved). After this, the user can either hover over any section of the tree graph to see the exact Euro contribution or use the statistics section to see the percentage they pay to each section

Also, the user can toggle the statistic information, which is derived from the sliding layered line graph to see the exact percentage of their tax dollars being spent in

certain budget areas. All of these extrapolation based on the user's taxes allows the user to determine exactly how their hard earned money is being spent.

Although our project wasn't focused on analysis, this visualization can be used by all of our audiences to do various degrees of meta-analysis. With respect to locals, the most prominent feature is the Individual Contribution feature. This allows the user to understand exactly how the City spends their money and could potentially lead to more informed citizen to government debate over new proposed budgets. Next, tourists can use this as a simple tool to understand how their arrivals impact the economy of Venice. Lastly, research can use the layered line graph and the tree graph to easily see trends in government spending and budgeting. This allows researchers to see trends in large data sets easier and than download the data sets to be worked with further.

#### 4.3.3 General Design - Daily Tourist Data

One other data set that we visualized was the daily tourist arrivals into Venice. Each mode of transportation has it's own area graph which represents the number of arrivals that day

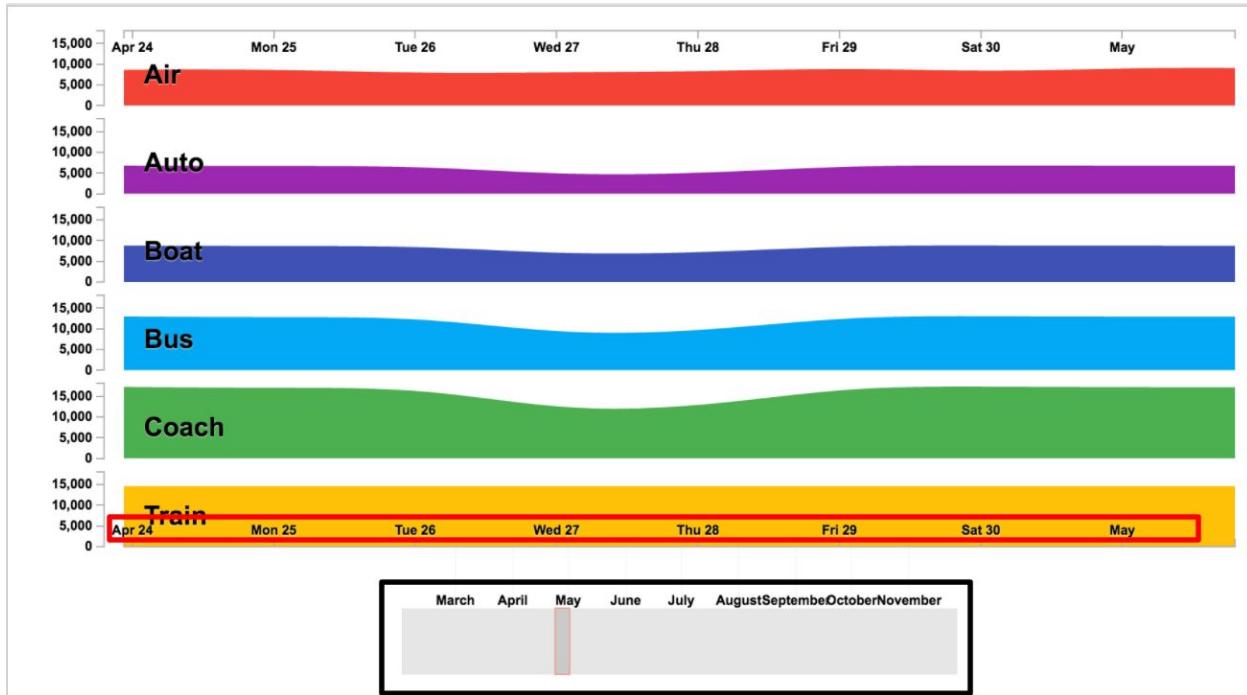


**Figure 47:** Our interactive daily tourist arrival widget. Each mode of transportation is given a separate graph and unique color. The y-axis is the number of tourist while the x-axis is the time domain. For the data we could obtain that domain is February to November 2016.

This is best shown in Figure 46 above, as each graph is titled and color coordinated. This data taken directly from the dashboard widget however it does currently lack functionality to consistently update. Currently the data is available for February to November of 2016.

#### 4.3.4 Data Organization & Interactivity

The data collected by this widget has yet to be represented, so we made the main functionality focus around seeing trends on different time frames. Because the data is organized by storage data and mode of transportation, the user has the ability to change the domain (time segment) of the graphs to see the trends. These trends could be as small as hourly, weekly, monthly or the year. This is best depicted by Figure 47 below which, when compared to Figure 46 above, show the domain shrunk to just a week (highlighted red) and new graphs to represent this.

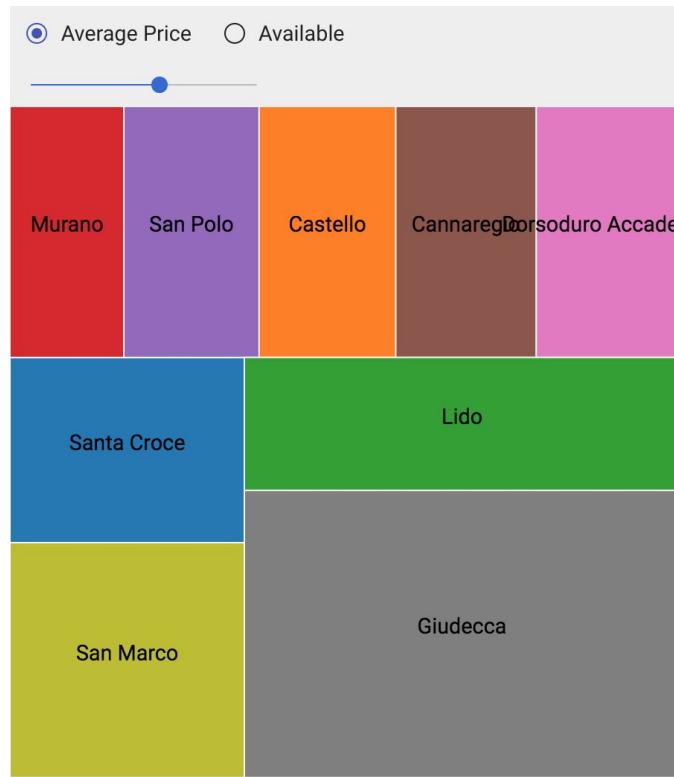


**Figure 48:** Compared to Figure 46, this shows a shortened range due to the time bar (black). As can be seen in red, the data is now filtering for only the last week of April and the first Day of May. Using the time bar and selecting a section, which is darker grey and slightly outlined, accomplished this. This segment can now be moved to show this time frame at any other points in the data.

Also, we added this functionality through a time bar on the bottom (highlighted black). This allows the user to select a time segment and then drag that specific time range to any other time points. One possible avenue of application for this visualization would be to track specific tourist arrival trends. As more data is added to the visualization the user can look through weekly trends across years or even monthly trends. With this information, more accurate times for peak hour of travel can be determined for any time of the year. Ultimately, this would lead to increased revenue as the city could charge more and also improve services when arrivals are the highest.

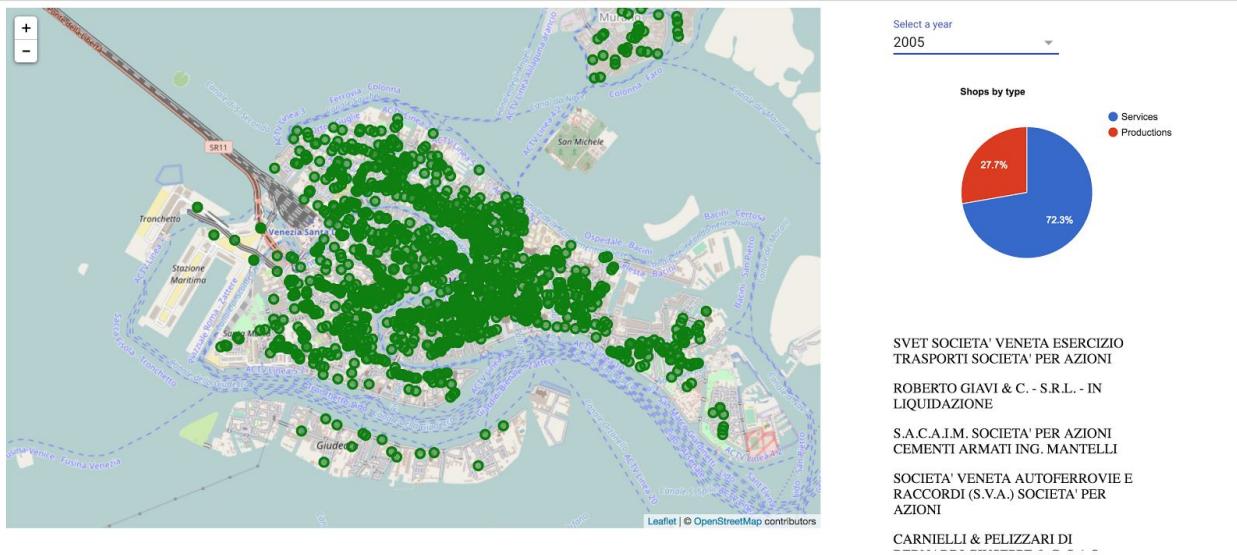
#### 4.3.5 Aiding Other Project Groups with Visualizing Data

Lastly, we aided several other teams by creating data visualizations for their project purposes, namely the Housing team and the Made in Venice team. For the housing team, we created a visualization that allows users to organize the sestieri by either average Airbnb rental price per night or by available units. There is also a date slider that allows users to move through time and see trends since the beginning of data collection.



**Figure 49:** Our Housing team Airbnb data visualization by sestieri.

The visualization that we created for the Made in Venice team was based on data regarding Venetian local businesses. The data sets identified local businesses in Venice by their title, whether they were producers or services, and their geographical coordinates. The first data set was collected in 1990, with new data sets being collected every 5 years afterward (i.e. 1995, 2000, etc.) As a result, the Made team wanted their data visualization to allow the user to toggle the year, which would modify different elements of the visualization. Our final result is shown below:



**Figure 50:** The data visualization we made for the Made in Venice team.

As shown, the user can toggle the year by using the button in the top right-hand corner. Upon doing so, multiple events will occur. First, each local business in Venice corresponding to the year in question will show up on the map based on their geographical coordinates. Additionally, a pie chart, seen on the right of the screen, dynamically adjusts in order to show what percentage of businesses in Venice are services and what percentage are producers. Lastly, as seen in the bottom right-hand corner, the businesses are listed based on the year as well. By interacting with this data visualization, users will be able to obtain a deeper perspective on the growth of local businesses in Venice. Both this visualization and the visualization for the Housing team will be linked to through the Venice Big Data website.

## 5.0 Recommendations & Conclusions

Throughout the course of our project, we have recognized several areas of improvement for future big data projects in Venice. This chapter should serve as a list of recommendations for future big data projects in Venice. Also, we will surmise all that we have accomplished during our time at the VPC.

### 5.1 Venice's Big Data Visualizations

Venice's real-time dashboard has been worked on by many students, with each leaving a different style with its own impact on the project. A strict set of guidelines known as a "Style Guide" would help provide consistent cohesive structure throughout the project. This would make it easier to maintain it would also make it easier for new developers to join the project.

As for the visualizations website, much of the data was manually imported to the Firebase real time database that backs all of the VPC's data. Providing a friendlier interface for entering this data would be of benefit to nearly all teams in the future.

#### 5.1.1 Visualizing VOD Data

Also completed during our project was the visualization depicting Venetian government spending from 2013 to 2016. As mentioned in Methodology, we searched through the 142 datasets across 6 different administrations of Venice. Out of all of these, the spending reports were only 18 datasets in 1 administration of Venice. This leaves 87% of datasets non-visualized, despite being related to the environment, population and culture (Ross, 2015). Although those only make half of the administrations of data available, those are very relevant when it comes to presenting data that is relatable to locals, tourist and researchers. Ultimately, a future team could utilize D3.js in order to produce new visualizations related to government verified data

from the VOD. From there, the visualization can be added to the new Big Data website as well to expand the datasets stored by the VPC.

Another starting point, besides a completely new data set, would be to improve on the work of our budget visualization. By the time future groups come in, new expense, funds, and revenues data will be available for the visualization. This can easily be added and will also allow the visualizations extrapolation function to better predict future budget. Also, future teams could also search for older budget data. Specifically they could start by looking for 2008 to 2012 Budget data that will completely fill the visualization. Lastly, our visualization needs more descriptive names for the “Mission” and “Program” names of the budget data. This may be found by working with the VOD, but we were unable to procure this data.

### 5.1.2 Visualizing Stored Widget Data

Lastly, our second visualization was made to represent data stored by the “Daily Tourist Arrival” widget. However, all widgets currently on the dashboard have had their gathered data stored by VPC since their creations. This provides any future team with a great depth of unused data that is very relevant to the City of Venice. The depth or size of this data has been increasing for every day the widgets are refreshed and the data is immediately relevant because the widgets are designed to appeal to tourists and locals. Now future teams can see our visual and be inspired by other data sets collected through widgets to make visualizations that will display long term trends not shown by the correspond real-time widget. For example, a future team could use “Daily Complaint” data to assess city response time based on season or simply how the density of complaints changes across time.

Also, our own visualization can be improved. By working with Tomaso Minelli, a future team could ensure that the logged data from widgets is accurate and change our visualization for a proof of concept to a final product. Checking the numbers collected by the VPC to those collected by the Venetian Government can start this project.

## 5.2 Future Improvements to the Venice Big Data Website

While we are proud of what we are able to accomplish with the Venice Big Data website, we fully acknowledge that there are many ways that it could be improved. This section details potential improvements we've identified.

### 5.2.1 Improvements to the Urban Elements Section of the Website

As stated in Section 4.2.3, the Urban Elements section of the website works seamlessly, as users are able to overlay different layers of Venetian urban elements on top of one another on the Leaflet map as they please. However, we believe that certain elements could be added to make this section even user-friendlier. As previously stated, the map points corresponding to each layer that appears on the map are distinguished by different colors. One aspect of functionality that we believe could be added to the Urban Elements section is a dynamically changing key or legend that tells the user which layers correspond to which colors. This way, users will always be aware of what exactly they are looking at, especially when they have many layers open on top of one another at the same time.

Additionally, we seek to eventually provide a map functionality that is just as robust as the Data section on the VPC website, after being transferred over to Leaflet. One important feature of the map functionality on the VPC website is that whenever users click on a data point on the map, a pop-up appears displaying a photo of the point of interest, its name and a link to its respective Venipedia article, if applicable. While we would have liked to implement this feature, we were not able to, due to time constraints. In order to do so, future groups would have to manipulate the geoJSON on Firebase to include fields specifying the objects' names, photos and links to Venipedia articles. Once this is accomplished, the Leaflet API allows developers to fairly easily implement pop-ups based on geoJSON attributes. Ultimately, we think that both of these aspects of functionality, if implemented, could be beneficial to users, and also provide inspiration for future elements that could be added to this section of the website.

### 5.2.2 Other Potential Improvements to the Website

In addition to the Urban Elements section, we believe that there are other aspects of the website that could be improved upon in the future. As previously stated, the website was developed using Materialize CSS, which automatically recognizes when users open the website on a mobile device. Thus, our website responds accordingly and resizes its features when opened on a mobile device. However, as the optimization of our website for mobile accessibility was beyond the scope of our project, there are still some issues with the website when being viewed on mobile, such as the overlap of some text in the navigation bar. These kinds of issues will be able to be fixed with relative ease by future teams.

## 5.3 Future Venice Widgets

Although our project consisted of three deliverables related to a website, dashboard & widgets, and new visualizations, we feel that the most projects can stem from our work with the latter two. With respect to the dashboard, there is potential for more widgets related to social media and the area of sentiment analysis. Also, our new visualizations opened two new avenues for the VPC: data that has been collected through widgets and VOD data.

### 5.3.1 Social Media & Sentiment Analysis

After seven-weeks of intensive work, we have created new widgets that allow the VPC to now store and display data related to social media. More specifically, we worked with data related to Facebook, Twitter, Flickr and Venetian News. After collection, we filtered all applicable data to show posts and news related to important issues. To continue with this, a future project can be revolved around analyzing or cross relating this data. Regarding analysis, there is an opening for sentiment analysis, which is when data is used to showcase the feelings of the public. For example, this can be accomplished through a filter system that searches Facebook post for keywords associated with

emotions (i.e. “happy” relates to “good” and “positive”). This can be done for several of the other widgets that display text. Lastly, a future team could focus on correlating the data collected by these widgets. For example, as the Facebook widget current stands it takes in posts related to important groups in Venice such as “Reset Venezia” and “”. With this widget showcases topics brought up by these organization, this could be compared to the relative importance of the topics on other platforms such as Twitter.

## 5.4 Project Outcomes

The mission of our project was to make complex datasets more available to the public through visual and digital means. Specifically, this mission resulted in objectives centered on the creation of a big data website, social media widgets, and big data visualizations. All these objectives were reached at the end the seven-week term by working with the VPC and VOD’s used and unused data sources.

The first objective used to complete this mission was the creation of a big data website. This new website showcases all of the data and visualizations collected by the VPC in one easily accessible section. Also, several key features such as a “Visualization of the Day” and visual demo make the visualizations more accessible than they ever were. Specifically, the “Visualization of the Day” appears on the home screen and displays a random visualization that can be changed upon refreshing the screen. This ensures some of the lesser-known visualization are still showcased and also may teach the user more about Venice. Further, the visual demo is associated with every visualization and showcases the functionality of the visualization before the user interacts with it. This addresses the fact that some visualizations have long loading times, so the demo allows the user to ensure that this visualization is the one they were looking for. These two examples exemplify how we created the website in order to give more information to the people by showcasing all visualizations and data fully. Also, the big data website serves as central hub to new widgets and visualizations.

The second objective, creation of new widgets, also worked toward the completion of our mission. Before our arrival, the VPC’s dashboard had only one widget

related to social media. We ultimately improved on this by creating our four social media widgets because social media is the way that any audience can get their news. The social media that was used to create these widgets was: Facebook, Flickr, Twitter and the News. We started by geo-locking all the data we analyzed, keeping it specific to Venice, because we only wanted news that was relevant to our local audience. Also, through working with the data from these website, the new widgets are able to update in real-time showcasing only the most relevant information. Using various Facebook opinion makers and the News we were able to create two widgets: one that showcases top posts from these groups and the other that parses through their keywords and creates a word cloud. Twitter was used to create a list of trending topics that increase in size as more people mention them. Also, both the Facebook post and Twitter trends are linked to their topics so the user can read more about them. Lastly Flickr was used to create a collection of images taken of Venice to highlight the beauty of Venice and some of the key spots people have been going to. Through these widgets we supplied any dashboard user with up-to-date news, trends, and even visuals of the city they live in. This ultimately allows the user to stay informed with everything that is happening in the City of Venice.

Our final outcome from the project was the creation of two new visualizations related to Tourist Arrival data and Venetian Spending Reports. We chose these data sets because both tourism and the economy are intertwined and always relevant to the city of Venice. The Tourist Arrival data was taken directly from the VPC's Tourist Widget. This showcased the ability to visualize an untapped resource (the widget's logs) as well as keep the information relevant to concerns of Venice. The visualization to best display this data was several area line graphs where each mode of arrival had its own graph. Further, the user could select a certain length of time (several hours, week, 9 days and more) to see more detailed trends and then move this time length to any other points of data. This allowed for data and information that hadn't been seen before to be readily available and interactive to allow the user to gain their own knowledge. Lastly, our Venetian Spending Report data was taken from the VOD's database and displayed

using tree graphs and layered line graphs. These two forms of visualization allow the budget to be easily understood as complex budget subsections are simplified to squares that have size related to the allotted funds. We also added interactivity as the user can click on any section and the tree graph will update to a tree graph that displays the subsections of that section. This allows the user easy access to any part of the budget. Finally, we added a very local specific layer of interactivity regarding “Individual Contribution”. This allows the user to input their tax expenses and then see exactly where their money is going. We felt that would provide more personal information and also give the user a sense of contribution and ownership of the city. Through both these visualizations we made data that hasn’t been worked with before more readily available to the public and utilized simple interactivity to allow for the user to glean even more information.

Together each objective works together to complete our mission and make information more available to the public of Venice as well as tourist and researchers. In doing this, we also started to create a more complete digital image of the city by showcasing key data about social media, government spending, and tourism. We finalized all of this through a website that serves as the starting point for all this new and older data information.

## 6.0 Bibliography

- Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 1.
- Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274-279.
- Brian Gentile. (2014). The top 5 business benefits of using big data. Retrieved from <http://data-informed.com/top-5-business-benefits-using-data-visualization/>
- Chang, A., Louie, K., Mark, A. & Wang, A. Materialize CSS. Retrieved from <http://materializecss.com/>
- Cook, P. (2012). *Visualization of domestic energy consumption*
- Ding, R., Gualdarrama, J. P., Horton, R. S., & Perrone, M. J. (2014). Venetian open data: Providing greater access to the Venice project center's data resources.
- Dockterman, E. (2015). Inside the fight to take back the fitting room. Retrieved from <http://time.com/how-to-fix-vanity-sizing/>
- Ed Feng. NCAA tournament predictions. Retrieved from <http://thepowerrank.com/ncaa-tournament-predictions/>
- Google. *Material design*
- Gopalkrishnan, V., Steier, D., Lewis, H., & Guszcza, J. (Aug 12, 2012). Big data, big business. 7-11. doi:10.1145/2351316.2351318
- Institute of Health Metrics & Evaluations. (2015). US health map&nbsp;
- K. Michael, & K. W. Miller. (2013). Big data: New opportunities and new challenges

[guest editors' introduction]. *Computer*, 46(6), 22-24. doi:10.1109/MC.2013.196

Kirby, D., & Carrera, F. (2013). Venice dashboard. Retrieved from

<http://dashboard.cityknowledge.net/#/Venice>

Lohr, S. (). The age of big data. *New York Times*

Lohr, S. (2012, Aug. 11,). How big data became so big. *The New York Times* Retrieved from

<http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html>

Mack, B. (2012, 11 September). Tourism overwhelms vanishing Venice. *World*,

March madness betting to total \$9.2 billion this year. (2016, Mar 15,). *Targeted News Service (TNS)*

Mauro, A. D., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, 65(3), 122-135. doi:10.1108/LR-06-2015-0061

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Retrieved from <https://hbr.org/2012/10/big-data-the-management-revolution>

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama*, 309(13), 1351-1352. doi:10.1001/jama.2013.393

Powell-Smith, A. (2012). What size am I?

Ross, W. (2015, May 14,). The death of Venice: Corrupt officials, mass tourism and soaring property prices have stifled life in the city. *The Independent* Retrieved from <http://www.independent.co.uk/news/world/europe/the-death-of-Venice-corrupt-officials-mass-tourism-and-soaring-property-prices-have-stifled-life-in-10251434.html>

Srinivasa, S. (2014). Big data analytics.

Tamara van Gog, Fred Paas, Nadine Marcus, Paul Ayres, & John Sweller. (2009). The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review*, 21(1), 21-30.  
doi:10.1007/s10648-008-9094-3

Telea, A., & Safari Books Online. (2015). *Data visualization: Principles and practice* (Second ed.). Boca Raton: CRC Press, Taylor & Francis Group. Retrieved from  
[http://wpi.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtlz0EUrE4wTQRNAFoagFAZMR6ImRsmmlloajZknmKQYppuCtVkiXRjDARrYRYvgiHrHWBaEedeQc1lswAObzEFvQARFqJYIJtk7-\\_t7BzAzMFgaghYAuQaGIkRmlYvAuMDNgBQ26gxt2OBSUDzpjlsLwgE6kmchNkYE0FbU8QYmBKzRNm4EY6TICEQcglsSRRoSyzGLRRErK9UpRBys01xNIDF2hOPHSoJh5YV5hZmFsYizHwJolWuOeVgDfCpUgwKJgbGxubpxoIJaUkAvtAaYZJZsapyebJwAxqkWxgnGYhySCMxSRJBllkUdiMVryhOTDnAWtmnPIWRsB-iKWIFFZTpRm4gO0GU8hIhAwDS0lRaaosOEQAi4alxw](http://wpi.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtlz0EUrE4wTQRNAFoagFAZMR6ImRsmmlloajZknmKQYppuCtVkiXRjDARrYRYvgiHrHWBaEedeQc1lswAObzEFvQARFqJYIJtk7-_t7BzAzMFgaghYAuQaGIkRmlYvAuMDNgBQ26gxt2OBSUDzpjlsLwgE6kmchNkYE0FbU8QYmBKzRNm4EY6TICEQcglsSRRoSyzGLRRErK9UpRBys01xNIDF2hOPHSoJh5YV5hZmFsYizHwJolWuOeVgDfCpUgwKJgbGxubpxoIJaUkAvtAaYZJZsapyebJwAxqkWxgnGYhySCMxSRJBllkUdiMVryhOTDnAWtmnPIWRsB-iKWIFFZTpRm4gO0GU8hIhAwDS0lRaaosOEQAi4alxw)

Tufte, E. R., 1942. (1983). *The visual display of quantitative information*. United States: Venice project center. (2013). Retrieved from [Veniceprojectcenter.org](http://www.veniceprojectcenter.org)

# Appendices

## Appendix A: Tools Used for Development

### Appendix A.1 SublimeText Editor

SublimeText is very lightweight text editor for Windows, Mac, and Linux. We selected this editor for several reasons.

1. It's lightweight nature allows for rapid development to quickly deploy our work
2. It's cross platform nature allows for each team member to work in the same environment.

### Appendix A.2 GitHub

GitHub is a web-based repository service which utilizes the git version control system. Using GitHub allows all of the members of the team to work on the same code at the same time, on different computers. All of the code is backed up in the cloud, so everyone can access it from anywhere. In addition, GitHub tracks all changes made to the code and allows users to revert to previous versions if something is accidentally deleted or commit a change which breaks part of the application. Lastly, GitHub is cross platform, so again, all members of the team can use it regardless of operating system.

### Appendix A.3 Heroku

Heroku is a Backend as a Service (BaaS). This means that it allows you to deploy projects and scripts to their servers that can handle all of our backend needs without having to handle every pitfall associated with hosting your own applications.