



WPI

Organizing and Visualizing 30 Years of Data Collection at the Venice Project Center

Griffin Bishop
Drew Ciccarelli
Kristen Dettloff
Christian Elzey
Matthew Schueler
Myles Spencer
Colin Willoughby
Jessie Ying

Fabio Carrera, Advisor
Peter Hansen, Advisor

December 15, 2017

ve17-know@wpi.edu ve17-open@wpi.edu
<https://sites.google.com/site/veb17knopen/>

An Interactive Qualifying Project Report submitted to the faculty of
Worcester Polytechnic Institute in partial fulfillment of the requirements of the
Degree of Bachelor of Science

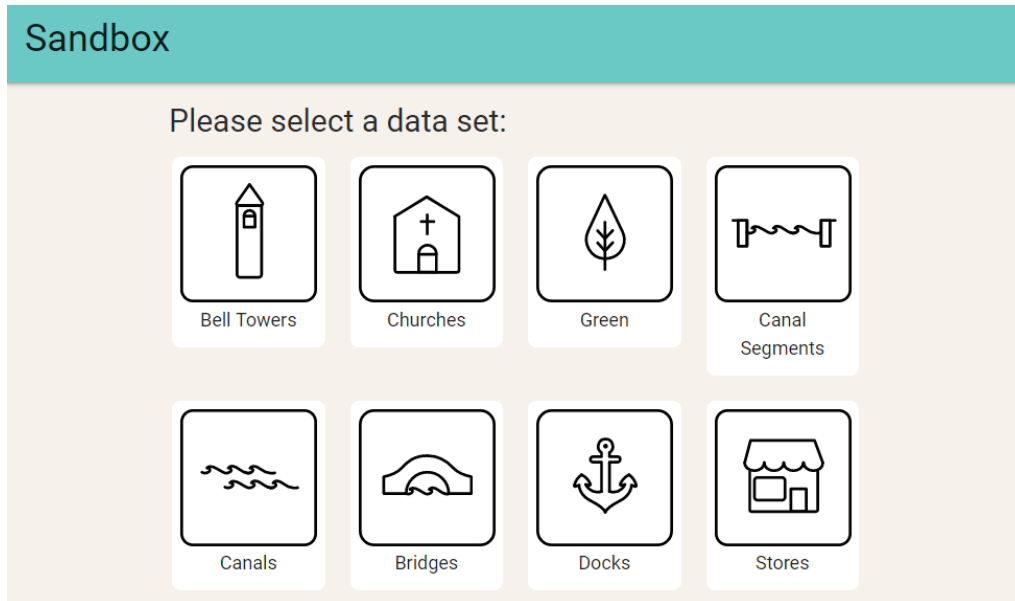
Abstract

Over the past thirty years, the Venice Project Center has collected data on numerous topics relating to the infrastructure, history and culture of the city. Our team's mission was to organize and disseminate this research in a way that is informative and easy to reuse. The culmination of our work resulted in a database of organized and standardized data; a Sandbox Application for creating visualizations of the data; infographics to be included in reference booklets; and a companion website that hosts our interactive visuals. These deliverables provide greater access to the VPC data for anyone who is interested in learning about historic Venice and helps to facilitate future projects to benefit Venice and its citizens.

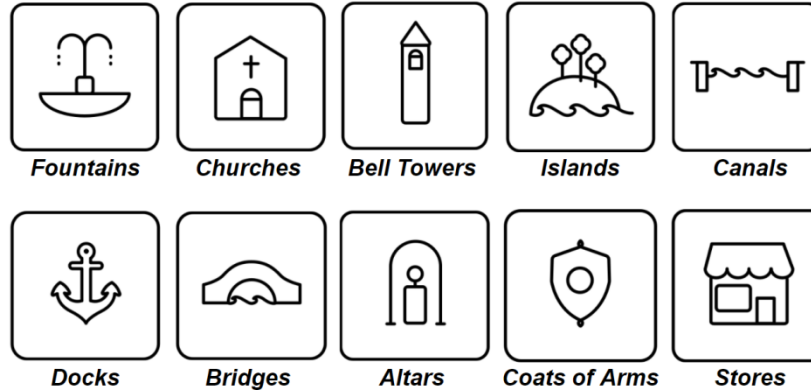
Executive Summary

The Venice Project Center (VPC) has hosted more than more than 200 projects completed by over 800 Worcester Polytechnic Institute (WPI) students since its founding in 1988. Almost every project completed at the VPC has focused on collecting data on the city of Venice, and nearly 400,000 individual data records about items in Venice are now stored in the VPC's data repository. The data and work of the VPC has been recognized by many prestigious publications such as *National Geographic* and *Wired*.

Properly utilizing this data is just as important as collecting it. Many projects have worked to distribute this data so it is readily available for users, and a number of online interactive tools have been created to increase its usability. These tools were created individually for a single purpose and are specific to a single data set. Our team's goal was to make an generic, extensible tool, which we named the Sandbox Application (sandbox.veniceprojectcenter.org), that can create a variety of visualizations for any of the VPC's data sets.

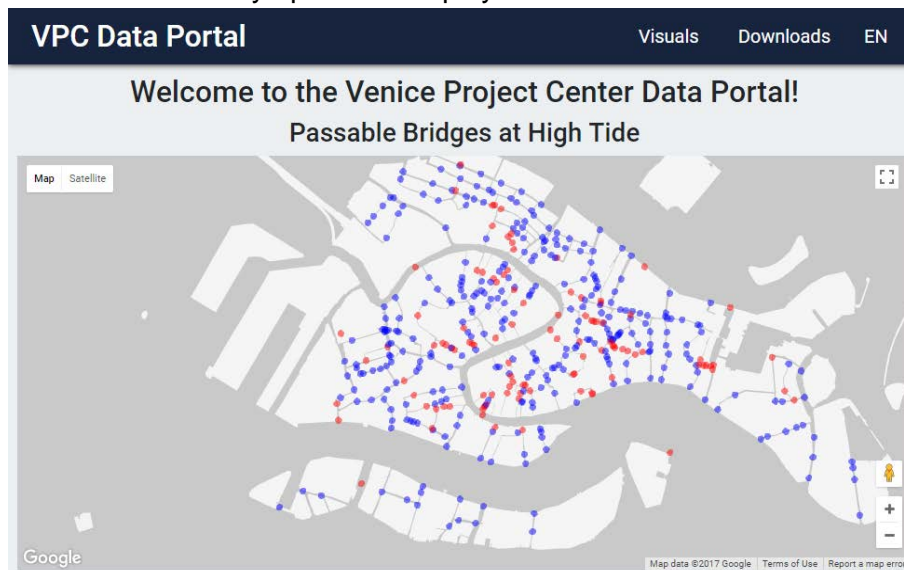


For its 30th Anniversary in 2018 the VPC plans to release a series of 30 booklets, *Conoscere Venezia*, or *Knowing Venice*. Each booklet will focus on one of 30 topics relating to Venice, 10 of which are shown below.

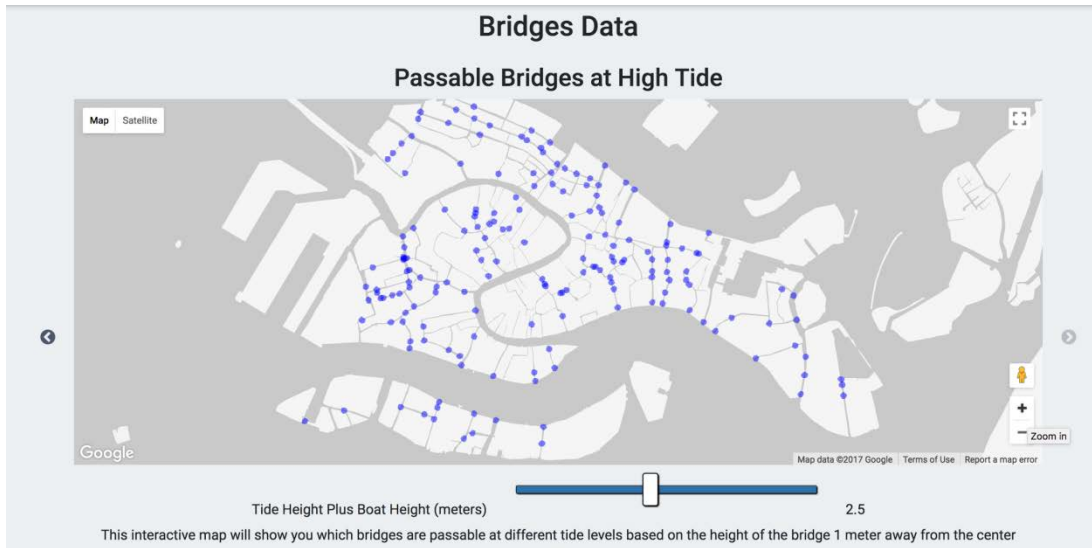


Creating a booklet series of such magnitude will require accurate and complete data, as well as high impact infographics. Our team ensured the completeness of the necessary data and assisted in the creation of static infographics, including map-based data visualizations to be used in the publications.

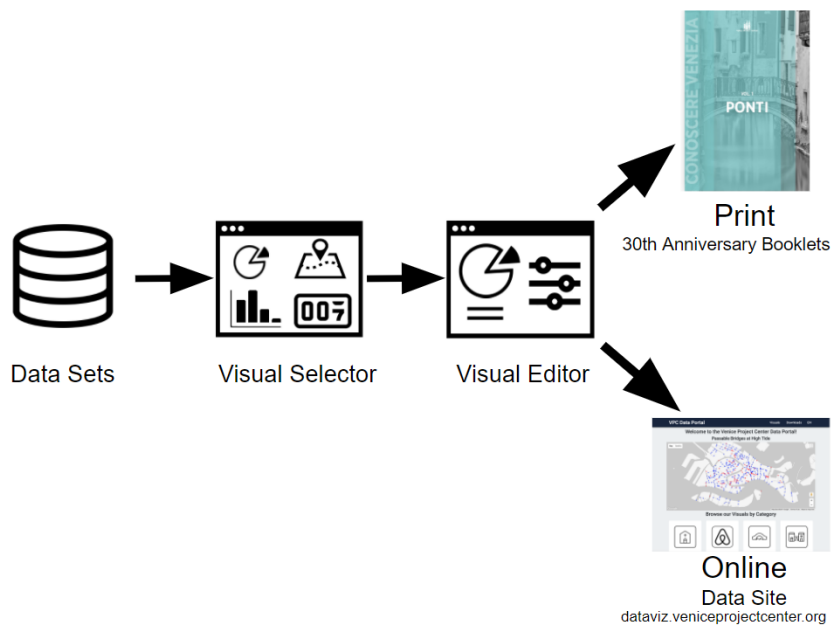
The goal of our project was to organize and disseminate the research produced at the VPC since 1988 in a way that is informative and easy to reuse, by making the data and visualizations accessible to the public. To accomplish our goal, we created both the Sandbox Application, a web-based tool for producing visualizations, and the VPC Data Portal site (shown below) to display finalized interactive visualizations. The two are linked to each other, which allows completed visuals made in the Sandbox Application to be directly published to the data site. The information used by all of our visualizations is pulled live from the VPC's data repository, so it will automatically update to display the latest data.



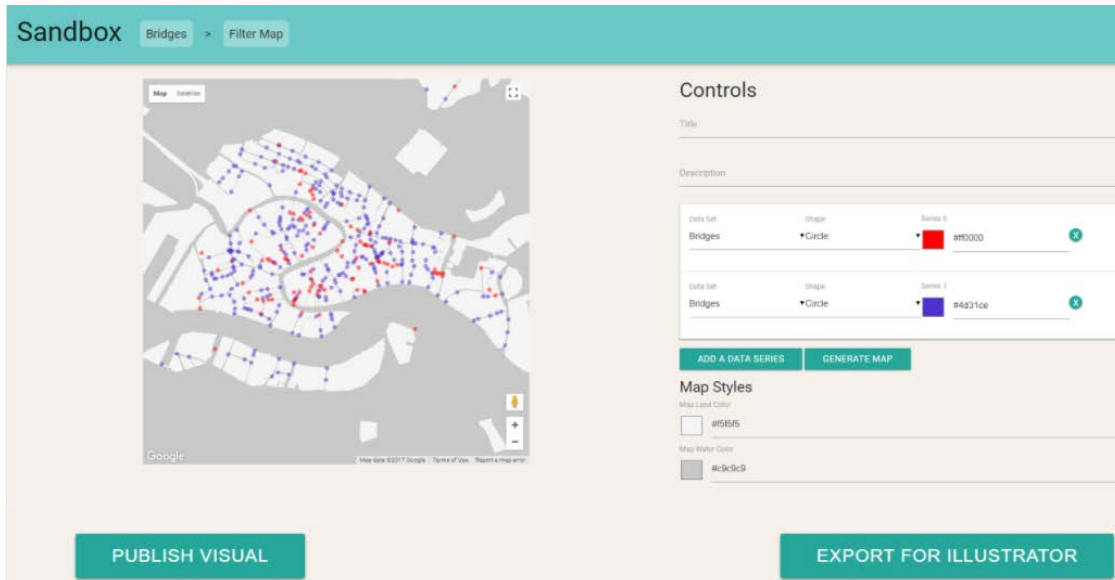
An important use of the Sandbox Application is improving public safety within Venice. For example, a past team collected various height measurements of bridges. This data is especially useful for ambulance drivers who need to know which bridges they can pass under when the water levels of canals rise. Displaying the data as an interactive and customizable visual, like the one we created below, can help the ambulance driver map out the route to take during high tides and could potentially save a person's life. The visual can be adjusted using the slider to display bridges passable by boats of any height at any tide level.



The process for creating a visual like the one pictured above involves the following steps:



The first step in creating a graphic is to select the data set and visualization type that you would like to use. The visual below displays data points on a map that meet the criteria defined by the user. The next step is to customize the graphic using the interface located to the right.



From here there are two options: the visualization can be exported for further editing in a graphic design program or published to our data site. If this graphic is going to be included in one of the booklets, it is necessary to export it and touch it up, resulting in an image like the one shown below. This static visual shows passable bridge as indigo circles and non passable bridges as magenta circles.



Mapa dei ponti, evoluzione storica

Itas di beaurundi quo magnihit et elestem etur simetur si aceaquatur mo mo ipsum si comriat iandesti doliensis dolo im corem unto itilesum ide nes de vellito. Um explicatit ex esto que et officumqui nis in vendis sante vellupt astati ore-pel magnis si conestrum ipsam fugit qui aut omnimpo repedit atibus ipacite vellere imclata cux, nonaou amendi barem

Itas di beaurundi quo magnihit et elestem etur simetur si aceaquatur mo mo ipsum si comriat iandesti doliensis dolo im corem unto itilesum ide nes de vellito. Um explicatit ex esto que et officumqui nis in vendis sante vellupt astati ore-pel magnis si conestrum ipsam fugit qui aut omnimpo repedit ad quatinv entur, sed ma num estate

Ponti / 83



Ponti / 84

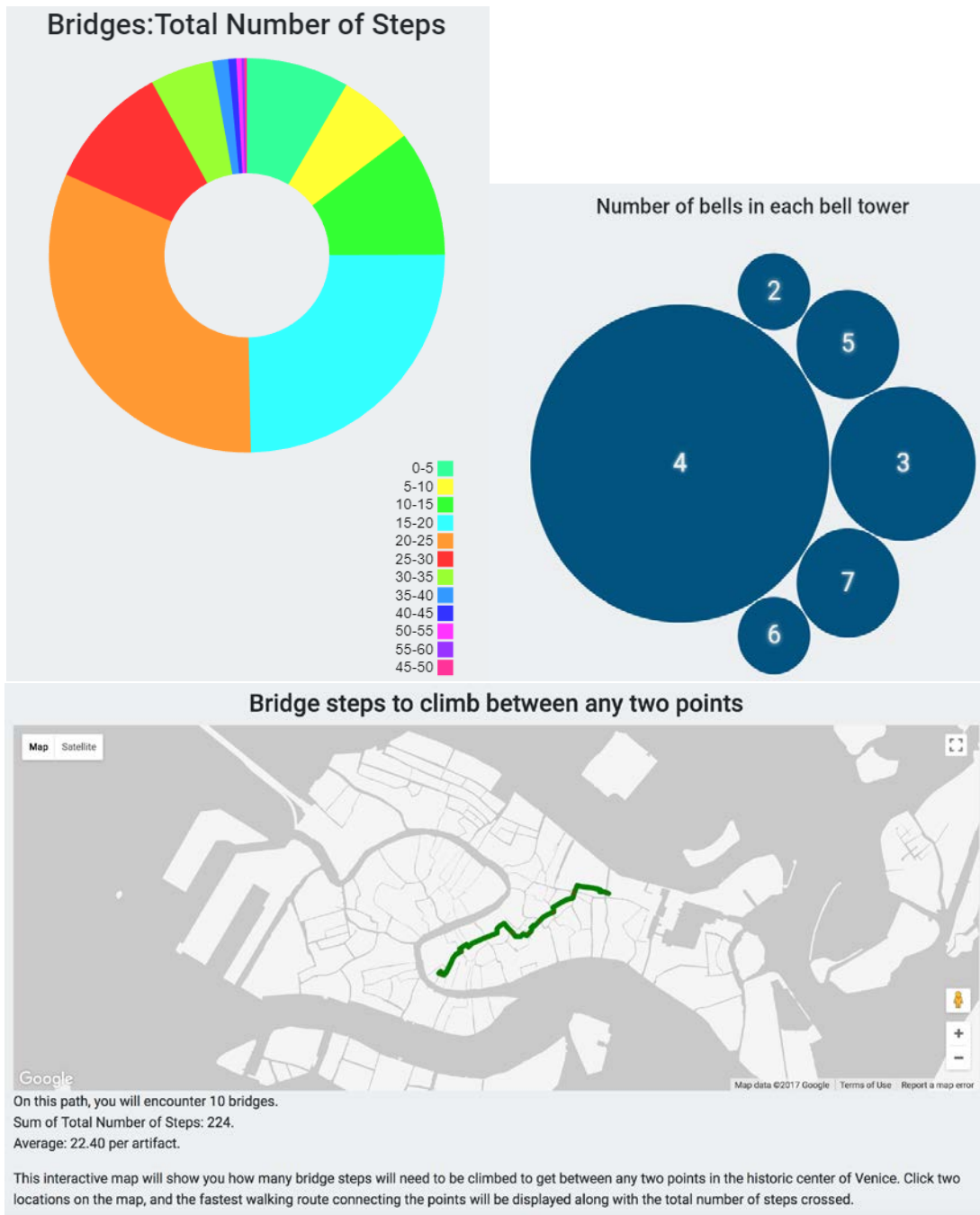


Interactive visuals, on the other hand, will be published to the VPC Data Portal website, which is accessible at dataviz.veniceprojectcenter.org. The interactive visual on passable bridges can be found here along with other visuals we have created. Anyone can embed these visuals in their own website by using the copy code button found at the bottom of each visual

page, as shown in the screenshot below. By allowing embedding, we created an easy way for anyone to disseminate the VPC's data via the visualizations we published.



We used this method to create many infographics for the booklet series and companion website. Below are two static visualizations that could be included in the booklets, followed by a dynamic visualization included in the VPC Data Portal website. The first visual shows bridges grouped by the number of steps. This shows most bridges have between 15-25 total steps. The second graph shows bell towers sorted by how many bells they each contain. Viewers can see that all towers have at least two bells and the most common number of bells per tower is four. The third image is an interactive map that allows users to click on two locations in Venice and then produces the shortest path to get from one point to the other. The visual also details the number of bridges crossed as well as the total number of steps climbed.



In addition, our team developed a methodology for evaluating which data sets contain missing or inaccurate data. We used our tools to identify missing data and show where the object is located in Venice. These look through a data set and identify fields that are missing or contain suspicious values, such as ones that fall outside of the expected value.

Due to time constraints, we could not practically verify all of the data sets through field visits, so we focused on gathering the missing data for bridges, which is the topic of the first booklet. We practiced our methodology in two ways. First, if the bridge was missing geographic coordinates, we looked at an older copy of the data set to find them. Second, we went out to

verify and collect any suspicious or missing data, particularly on the number of steps of various bridges.



The tools we created allow the VPC's data to be published in a way that is more understandable to the general public. The wealth of knowledge the VPC has collected can benefit the city through the creation of comprehensive visuals. Our team recommends using the Sandbox Application to create visualizations for the *Conoscere Venezia* booklets and future projects completed at the VPC. The ability to create visualizations using any data set, embed visuals in other websites, and sync to the database in real-time are key features that make this tool unique.

Table of Contents

Abstract	ii
Executive Summary	iii
List of Figures	xiii
List of Tables	xiv
Authorship and Acknowledgments	xv
1. Introduction	1
2. Background	3
2.1 Impact of the Venice Project Center	3
2.1.1 History and Accomplishments of the VPC	3
2.1.2 Uses of the Data	4
2.1.3 Anniversary Projects	5
2.2 The VPC's Online Resources	6
2.2.1 VPC 2.5	7
2.2.2 Cartography	7
2.2.3 Venipedia	8
2.2.4 Firebase and City Knowledge Console	9
2.3 Volume of VPC Data and Infrastructure	10
3. Project Objectives and Scope	12
4. Validating and Integrating VPC data for publication	13
4.1 Assessing the completeness and correctness of existing data	13
4.2 Collecting missing data	15
4.3 Adding data to Firebase	15
4.4 Results	15
5. Developing a tool to facilitate the production of interactive visuals	16
5.1 Determining visuals that our tool should produce	16
5.2 Building the Sandbox Application	16
5.3 Results	17
6. Using the tool to create specific visualizations for print and online publications	18
6.1 Case study: Creating a visualization from start to finish	19
6.2 Example Visualizations made using the Sandbox Application	22

7. Making the VPC data accessible to the public via an online platform	26
7.1 Overview of the VPC Data Portal	26
7.2 Examples of Interactive Graphics Hosted on Our Site	27
7.3 Data Portal Downloads Page	29
7.4 Data Portal Embedding	29
8. Conclusion and Recommendations	31
8.1 The Future of the Data Site	32
8.2 Recommendations for Additional Visualization Tools	32
8.3 Recommendations for Creating Graphics with the Sandbox Application	32
8.3.1 Creating Infographics for Future Booklet Publications	32
8.3.2 Creating Infographics for VPC Websites	33
8.3.3 Creating Infographics for Future IQPs	33
8.4 Recommendations on Data Integration	33
8.4.1 Creation of a Diagnostic Tool	33
9. Bibliography	34
Appendix A: Data Source Tracking Sheet	36
Appendix B: The Open Data Movement	37
Benefits of Open Data	37
Barriers to Open Data	37
Arup: Shaping a Better World with Open Data	38
Appendix C: Popular Science Publications	39
Appendix D: Data Visualization	40
Importance of Data Visualization and Design	40
Types of Static Infographics	40
Chartjunk: Striking a balance	42
Choosing the Right Visualization	43
Appendix E: CSV Data Set Code	45
Appendix F: CSV Data Set Statistics	47
Appendix G: Types of Infographics we Implemented	48
Appendix H: Sandbox Application Feature Guide	49
Bar Chart	51
Donut Chart	52
Bubble Chart	53

Filter Map	54
Category Map	56
Route Map	57
Scaled Up Number	59
Choropleth Map	60
Bubble Map Chart	61
Donut Chart Map	62
Data View tool	63
Publishing visualizations to the Data Site	64
Exporting as SVG	64
Publish Visual	65
Visual Removal	65

List of Figures

- [Figure 1](#): *Knowing Venice* booklets
- [Figure 2](#): VPC projects completed by year, since 1988
- [Figure 3](#): Data Topics
- [Figure 4](#): The first VPC website homepage, created in 1998
- [Figure 5](#): A screenshot of the current VPC 2.5 Website.
- [Figure 6a](#): Venipedia article created automatically from a template
- [Figure 6b](#): The data set entry corresponding to the entry in Figure 6a.
- [Figure 7](#): Data collected about Venetian stores at the VPC
- [Figure 8](#): Bridge book cover and example page
- [Figure 9](#): Data organization sheet
- [Figure 10](#): Breakdown of average null and empty fields to data fields
- [Figure 11](#): Field Visit to the Ponte dei Meloni
- [Figure 12](#): Process for creating our visualizations
- [Figure 13](#): A visual we created include in a page from the prototype booklet on bridges
- [Figure 14](#): Selecting a data set in the Sandbox Application
- [Figure 15](#): Selecting a visual to use with the bridges data set
- [Figure 16](#): The Sandbox Application filter map configuration screen
- [Figure 17](#): The final bridge height interactive visual.
- [Figure 18](#): Number of bells in each bell tower
- [Figure 19](#): Distribution of Venice population from 2011 census
- [Figure 20](#): Airbnb listings in Venice by occupancy (color) and price (size)
- [Figure 21](#): Distribution of souvenir stores in Venice
- [Figure 22](#): Public art in Venice most in need of restoration
- [Figure 23](#): The VPC Data Portal website main page (dataviz.veniceprojectcenter.org)
- [Figure 24](#): The carousel shown after selecting the bridges data set
- [Figure 25](#): “Passable Bridges at High Tide” visual with a tide plus boat height of 2.25 meters
- [Figure 26](#): “Passable Bridges at High Tide” visual with a tide plus boat height of 2.75 meters
- [Figure 27](#): VPC Data Portal Downloads Page
- [Figure 28](#): Copy embedded code example
- [Figure 29](#): An example embedded visual on a test website
- [Figure 30](#): Process for creating our visualizations
- [Figure 31](#): Data Tracking Sheet
- [Figure 32](#): Publications the VPC has been featured in
- [Figure 33](#): Example of a time-series chart
- [Figure 34](#): Relational graph
- [Figure 35](#): Example of chartjunk
- [Figure 36](#): Another example of chartjunk
- [Figure 37](#): Election results for 1984 European Parliament election
- [Figure 38](#): Scatter plot matrix of mileage, weight, and price of different car models
- [Figure 39](#): Sandbox Application display window
- [Figure 40](#): Visual Selection

- [Figure 41](#): Bar Chart
- [Figure 42](#): Donut Chart
- [Figure 43](#): Bubble Chart
- [Figure 44](#): Example Filter Map
- [Figure 45](#): Example Filter Map with multiple data sets
- [Figure 46](#): Category Map
- [Figure 47](#): Route Map
- [Figure 48](#): Scaled Up Number
- [Figure 49](#): Choropleth Map
- [Figure 50](#): Bubble Chart Map
- [Figure 51](#): Donut Chart Map
- [Figure 52](#): Data View Tool
- [Figure 53](#): Export options in the Sandbox Application
- [Figure 54](#): A Donut Chart SVG being edited in Adobe Illustrator
- [Figure 55](#): Deleting a published visual

List of Tables

[Table 1](#) The Metadata about the Venetian CSV data sets that was generated by the code in Appendix E

Authorship and Acknowledgments

The authors of this proposal are Griffin Bishop, Drew Ciccarelli, Kristen Dettloff and Christian Elzey from the Know Team, and Matthew Schueler, Myles Spencer, Colin Willoughby, and Jessie Ying from the Open Team. Each student specified above contributed equally to this document.

This team was originally two separate teams, previously referred to as 'Knowing Venice' and 'Venice 3.Open', but we combined to form a single team called 'Knowing Venice'. The initial project proposals and preparations were done separately by each team, but upon our arrival in Venice, we merged the two reports because of how closely related they were. The two teams were working together closely, essentially completing the same project but with different deliverables (print booklets and online platform), so we incorporated both deliverables into the objectives for our combined project.

Our team would like to thank Professors Carrera and Hansen for advising our project and providing invaluable feedback on how to improve it every step of the way. Professor Carrera shared his vast knowledge on the city of Venice and history of the VPC, allowing us to greatly improve the quality of our background section and project as a whole. Professor Hansen read through and edited our paper and presentations, giving us timely feedback to ensure they were in proper form.

We would also like to thank the staff at the Venice Project Center for assisting us this term: Nicola Musolino for providing technical support when we launched our web applications on the Venice Project Center website, Andrea Toffanello for gathering and creating many of the icons and images related to the data topics and booklets, Piero Toffolo for helping us gather missing geographic data, and Sarah Puccio for translating our slides into Italian. Without them, we would not have been able to successfully complete our project.

1. Introduction

The Venice Project Center (VPC), founded in 1988 by Fabio Carrera, has hosted more than 200 projects over the past three decades. Almost every project completed at the VPC has focused on collecting, organizing, and publishing data, on topics ranging from bell locations and audio, to information on tourist traffic on specific streets. The project center has been featured in prestigious publications such as *Wired* and *National Geographic* for its data-driven approach to solving problems.

In celebration of its upcoming 30th Anniversary, the VPC will be publishing a 30 booklet series, *Knowing Venice*, or *Conoscere Venezia*, to showcase its work. Sample booklet covers can be seen in Figure 1 below. Each of the booklets will highlight a specific topic of Venice, such as bridges, canals, stores and demographics (a complete list can be found in Appendix A).



Figure 1 *Knowing Venice* booklets

The VPC has over 10,000 individual records of data about items in Venice, which has allowed them to address many issues facing Venice as a cultural heritage site. Having such a large collection of data also presents various challenges, such as potentially incomplete or outdated data sets. As a result, a structure has to be put in place so that information can be located and efficiently used.

While past projects have started addressing the VPC's data organization issues, the data sets still need to be standardized and validated before being made accessible publicly. A thorough analysis must then be done to determine which data sets need to be filled in and integrated into modern storage formats.

The goal of our project was to organize and disseminate the research produced at the VPC since 1988 in a way that is informative and easy to reuse, by making the data and

visualizations accessible to the public. This involves the contribution of static visualizations to the 30th Anniversary booklet series as well as the creation of its companion website with interactive visualizations.

2. Background

Over the past 30 years, over 800 students from Worcester Polytechnic Institute (WPI) have completed more than 200 projects at the Venice Project Center. These projects have collected a large volume of data consistent with the mission of the Venice Project Center. The mission of the VPC is to preserve and enhance the rich cultural heritage of Venice while enabling it to adapt to the 21st century. The mission follows from the circumstances surrounding the project center's founding, and has evolved through three decades of projects that have kept up with the forefront of technology. Throughout the years, WPI student projects have not only collected data, but also distributed it using the best technology that was available at the time. Similarly, our project aims to preserve and distribute this data through modern data visualization techniques.

2.1 Impact of the Venice Project Center

Venice is at risk of losing its cultural heritage. A massive increase in tourism alongside a massive decrease in the native Venetian population over the past 60 years has lead some to say that the city may be progressing towards a “social, anthropological, and historical disaster.” (Giuffrida, 2017) To prevent this, measures needed to be taken to help preserve Venice's unique culture and history.

2.1.1 History and Accomplishments of the VPC

As a native Venetian, WPI professor Fabio Carrera is a personal stakeholder in the current and future development of Venice. During his time at WPI as an undergraduate, Carrera participated in the Global Projects Program and saw what a positive impact this program can have on a community. He founded the Venice Project Center in 1988 to address the impending problem of cultural loss. Since then, the project center has focused on preserving the city's heritage through data collection and dissemination. Figure 2 shows the distribution of over 200 projects completed in Venice over these years. Data collected by these projects was on a range of topics, some of which are shown in Figure 3 below.

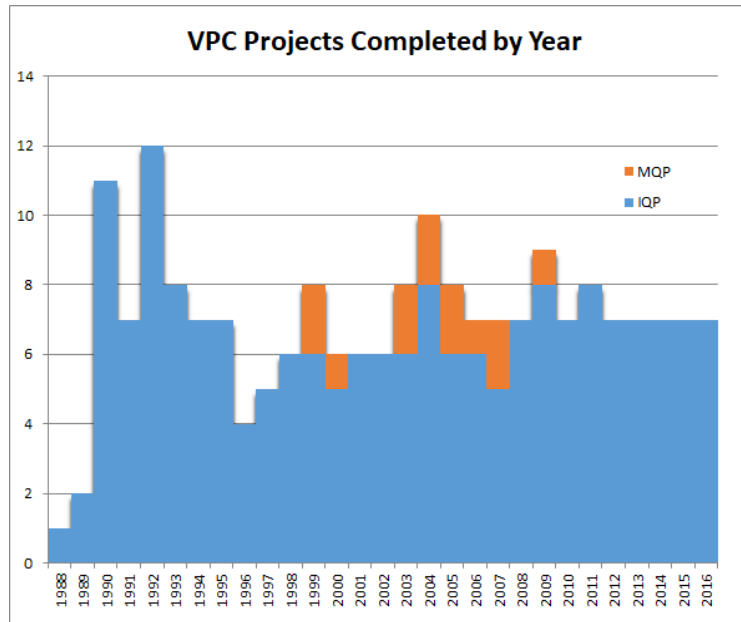


Figure 2 VPC projects completed by year, since 1988

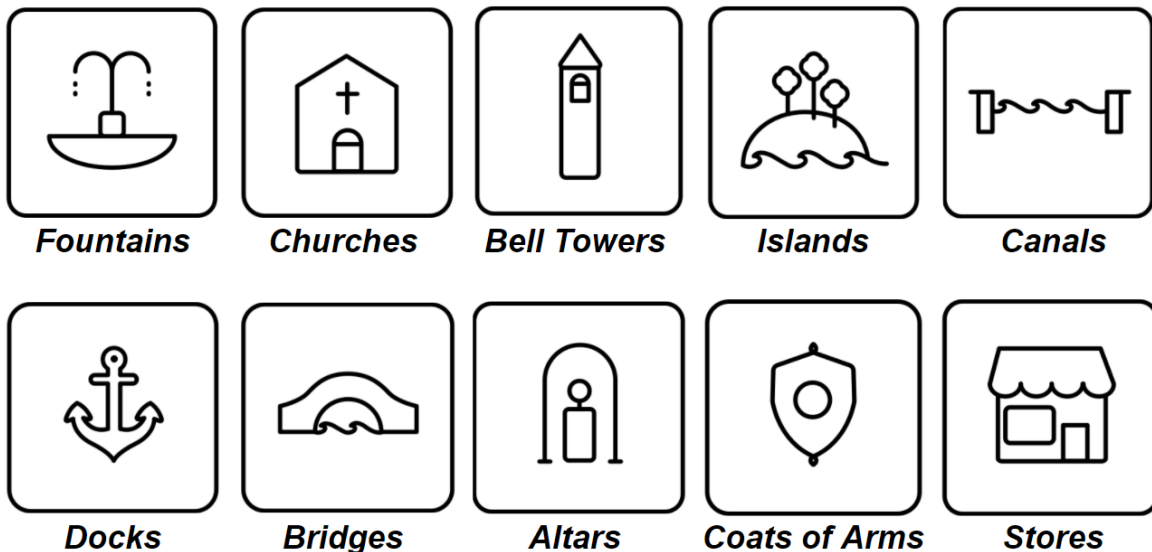


Figure 3 Data Topics

Three decades of project work has established the Venice Project Center as a reliable, encyclopedic source of information about Venice. These efforts have not gone unnoticed; numerous prestigious publications, including National Geographic, Wired, and Smithsonian, have recognized the impacts of the Venice Project Center on the preservation of the city's culture and heritage (see Appendix C).

2.1.2 Uses of the Data

The data collected at the VPC on each of the topics about Venice is useful for a number of real world applications. In particular, it can improve the public safety of Venetians. During

high tides, the water levels of canals rise, causing some of the bridges to become impassable by boats. This can be a huge problem for ambulances, which need to know the shortest path to reach someone in need. The *Canals, Bridges and Urban Maintenance* team collected data on the heights of bridges, which would be helpful to determine the clearance height of bridges at different water levels (Bossalini et al., 2013). Having a way to display this data can help ambulance drivers map out their route during high tides and could potentially save a person's life.

Another use of the VPC data is determining how handicap accessible the islands in Venice are. This can be accomplished by looking at the data collected on boat stops and the bridges connecting the islands. The islands with boat stops as well as islands connected to an accessible island by a bridge with zero steps are easily accessible to handicapped individuals. These example applications of the VPC's data show its potential to improve Venice.

2.1.3 Anniversary Projects

There have been three occasions for VPC anniversary projects, each collecting data and making it available to the public in the forms available at the time. These projects were essential to keeping all VPC databases up to date and publicly available.

The Venice Project Center's Tenth Anniversary documented the early history of the VPC, the problems faced by Venetians, and what WPI students had done to address these problems (Behmke et al., 1998). It sorted the completed projects based on the problems they addressed into one of five categories: Infrastructure Studies, Heritage Preservation, Lagoon Environment Studies, Technology and Sustainable Development, and Strategic Planning. In the end, the tenth anniversary team produced a report, a website containing some of the project information, and a CD of the information and website (Behmke et al., 1998). Figure 4 shows the homepage of their site. While the tenth anniversary project organized the Center's first ten years of work, their deliverables did not include any data analysis or visualizations and did not effectively make VPC data available to the public.



Figure 4 The first VPC website homepage, created in 1998

The project *Showcasing Twenty Years of Venice Project Center Results Using Interactive Online Infographics* increased the usability of VPC data by creating some interactive infographics of the data that anyone could understand. The greatest contribution made by this project group was the Venice 2.0 site, which combined many of the VPC’s digital resources into one location (Sargent et al., 2011). The principle feature of this site was the large interactive graphic, which combined historical maps and VPC data to display the different aspects of Venice at different time periods.

25 Years of Venice Knowledge Online worked to create a more organized online database and mobile application. Their group added the Venice Dashboard, created a VPC specific app to improve data accessibility, and organized and consolidated data using the City Knowledge Console, which is described in Section 2.2.4. Finally, the team created a plan for a crowdfund campaign to raise money for an effort to digitize old Venetian manuscripts and records, preserving their contents (Brann et al., 2014).

2.2 The VPC’s Online Resources

The VPC faces a challenge common to many businesses and institutions: the distribution of data. The creation of websites and other online tools enables the VPC to combat this challenge and share its information with the world. One major advantage of the publications of the VPC are that they are all in English which allows for the data to be reached by a greater audience.

2.2.1 VPC 2.5

The Venice Project Center has had a web presence since 1998, when its first website was created by WPI students as a part of the tenth anniversary project. Over the past 19 years, the VPC has continued to further its online presence by launching online tools and websites such as updated versions of the VPC website, geographical information system (GIS) exploration tools, and Venipedia, each of which will be described in greater detail below. The current website, VPC 2.5, is shown in Figure 5. It was created for the 25th anniversary of the Center's founding. It features an interactive map and a dashboard with widgets including live radar of airplanes, tourist data and complaints. The dashboard provided a clearer understanding of the data, making use of it in a more productive way. Although most of the information available directly on the site is up to date, there are many broken links within the site. For example, users are currently unable to download the data sets as advertised within the website. The current VPC 2.5 site also hosts and links to the various web tools and resources the Center has developed in recent years.

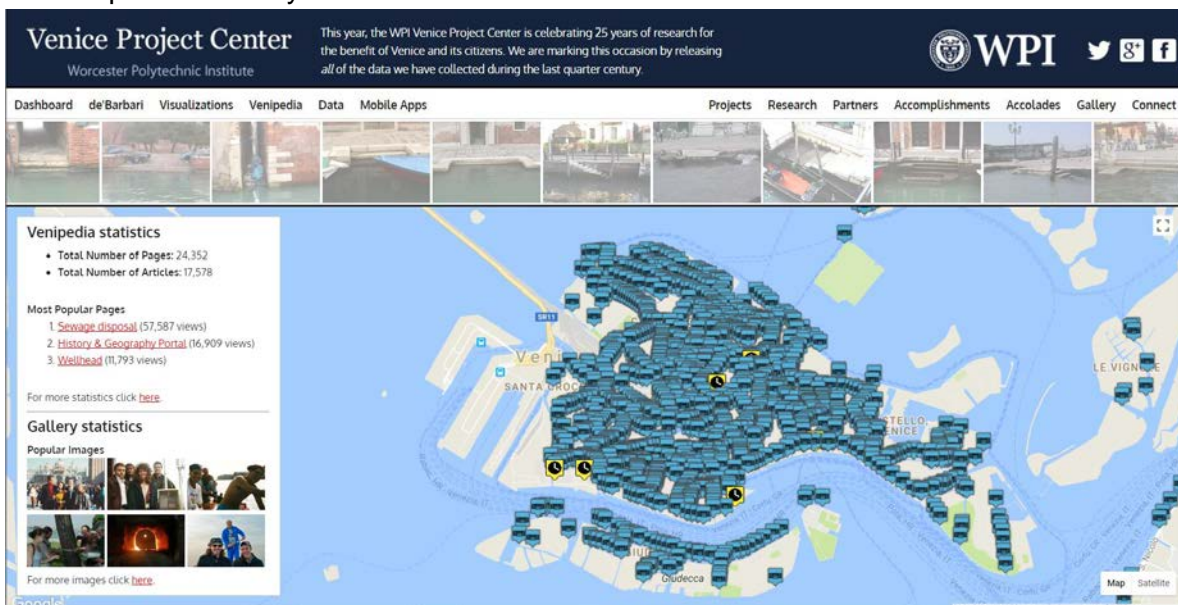


Figure 5 A screenshot of the current VPC 2.5 Website.

2.2.2 Cartography

A large amount of the data created by the Venice Project Center is geographic in nature, involving the locations of specific objects and landmarks in the city. A number of past projects have done work on cataloging and mapping city features, including former landmarks that no longer exist, such as demolished churches. This valuable geographical data paints an interesting picture of Venice's rich history, and the VPC has taken steps to make this information accessible online. The Cartography section of the Venice Project Center website shows the location of all of the bell towers, wellheads, islands, demolished churches, canals and other city features the VPC has collected information on, superimposed on detailed modern and historic maps of Venice. Layers depicting the locations of various city features were created using pre-existing data sets. This online map viewer was designed to be very user-friendly, allowing for easy searching and browsing of VPC geographical information. In 2014, a follow-up

project added additional maps to the online viewer, allowing users to view the locations of city landmarks on five historic maps from different time periods as well as a modern map (Sadowski et al., 2014). This team recommended further integrating the Cartography online viewer with other online tools created by the Venice Project Center, such as Venipedia, but this has not yet been implemented.

2.2.3 Venipedia

Venipedia is a Wiki-based online encyclopedia dedicated to the city of Venice. Professor Carrera created Venipedia in 2007 to fill a knowledge gap that existed between the English and Italian language. Its entries on the city of Venice present information gathered by the Venice Project Center online in the familiar Wiki format (“Help:About”). Venipedia’s “About” page explains, “The information that Venipedia provides is intended to aid academics, scientists, scholars, city planners, and other professionals who may have a need for primary data on Venice” (“Help:About”).

Venipedia is unique in that it hosts large amounts of specialized, quantitative data on physical aspects of the city of Venice, much of which cannot be easily found anywhere else online (“Help:About”). The majority of data presented on Venipedia was created by Venice Project Center sponsored projects, but as with most Wiki-style websites, user contributions are allowed and encouraged. Unlike Wikipedia, which typically requires that contributors attribute all information to published sources, Venipedia allows for original research to be submitted, in order to further build a repository of useful data on the city (Scannel et al., 2011).

Since its introduction, Venipedia has grown substantially in size and in scope, and now contains over 500 pages on a large variety of topics relating to the city of Venice (“Help:About”). To ensure the site is as useful a resource as possible, two past projects have focused on improving the user-friendliness and quality of content of Venipedia. In 2010, a team of students focused on cleaning up the design of the site, fixing technical issues causing a number of broken links, and overall increasing the accessibility of Venipedia for those unfamiliar with the Venice Project Center.

Two years later, another project focused on addressing the challenges typical of sites hosting user-generated content that Venipedia was facing. Low-quality articles, spam, and a lack of consistency in the data reduced the effectiveness and reliability of Venipedia (Tsiros et al., 2013). This team completely rebuilt Venipedia from the ground up, selectively transferring only the highest quality articles to the new site. Anti-spam systems were put into place, and detailed guidelines and a help section for contributors were added to improve the quality and consistency of new articles (Tsiros et al., 2013). Overall, this project better positioned Venipedia for future sustainability.

Most of the pages on Venipedia are automatically generated from existing data sets using templates. As a result, many pages contain poor formatting and awkward wording, with some containing blank spaces that appear when gaps exist in the source data sets. The data used to generate these pages is stored inside of the VPC’s Firebase. Figure 6a is an example of a Venipedia page generated from the contents of the data set entry shown in Figure 6b.



Figure 6a. Venipedia article created automatically from a template

	A	B	C	D	E	F	G	H	I	J	
1	Artifact Code	Page Title	Year Assessed	Shape	Length	Width	Type of Artifact	Materials	Condition Score	Condition Category	Of
631	GIGL_B7	Church Floor Artifact - GIGL_B7	2004	Square	89cm	89cm	Tomb	Red Marble and White Marble	0.2554	Intermediate Damage	the floor be found Maria Zo Square pi 89cm (ler Red Mart marking i

Figure 6b. The data set entry corresponding to the entry in Figure 6a.

2.2.4 Firebase and City Knowledge Console

The Venice Project Center has started to move all of its raw data sets into a single database called Firebase. Firebase is a database created by Google to store large amount of data online. While most of the VPC's data has been transferred into this system, many fields remain empty. One can easily find tables where columns contain mostly null or obviously inaccurate values. The City Knowledge Console (CKC) is a way to access and edit the information stored in the database. It is not easily accessible to the public, however, making it not intuitive to use.

2.3 Volume of VPC Data and Infrastructure

The VPC has accumulated a massive amount of data from over 200 projects completed before 2016. Its collection includes information on a plethora of topics, including bridges, churches, and stores. The bridges data set contains information on 465 bridges in Venice, including data on various height measurements, canals crossed, and the number of steps. We can use this data to find that the total number of steps of all the bridges is 8635, which is equivalent to climbing the San Marco bell tower 50 times.

In addition, the Venipedia site currently contains 27,120 total pages of information on dozens of topics. Each of these topics can have thousands of pages. For example, the stores category consists of 4980 individual pages, one for each store in Venice. As seen in Figure 7, these numbers have only grown over the years. In addition, the site contains 708 uploaded files, which include various images, graphs, and maps (Venipedia).

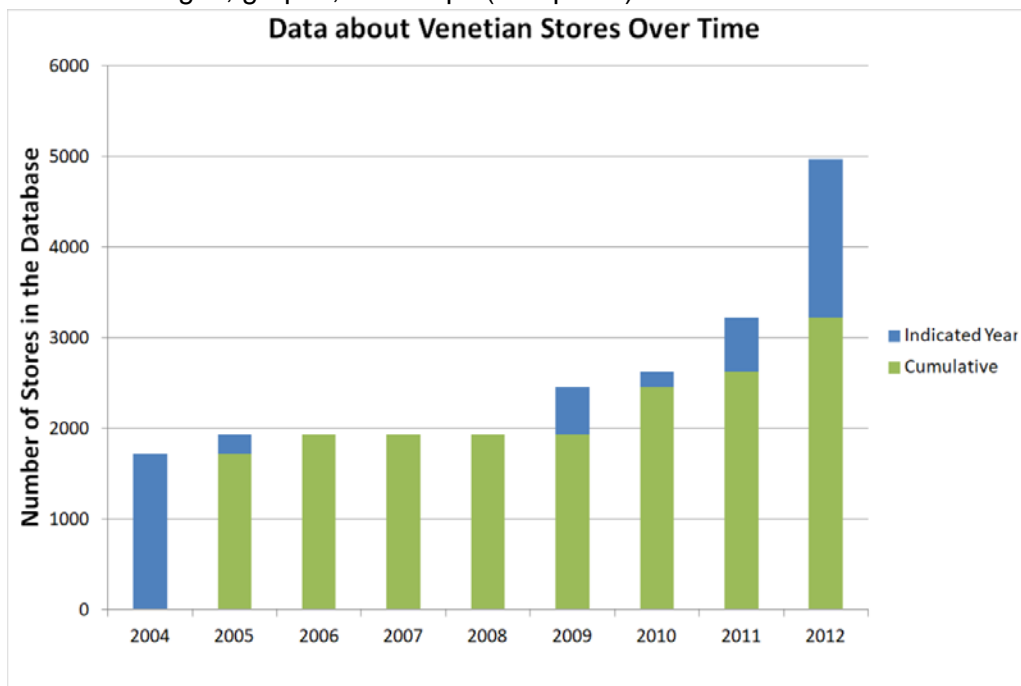


Figure 7 Data Collected about Venetian Stores at the VPC

The data stored in the database is split between 19 tables, each corresponding to a variety of topics. Across all of the tables, there are nearly 10000 rows of data, and most of the tables have 30 or more columns. An average table consists of about 49000 fields of data, and is nearly half a megabyte in size. The largest table has nearly 5000 lines, over 100000 fields, and is over 2 megabytes large.

Unfortunately, a sizable chunk of each of these tables has not been filled in with data. Approximately 26% of the total fields are either empty or filled with null data. One of the tables is missing nearly 75% of its data. For the script used to generate this metadata, see Appendix E. For the full results of this code, see Appendix F.

Although the data collected by students at the VPC is very thorough, there are still improvements that can be made to the online deliverables. We will work to address the issues of data availability, structure, and validity with Venipedia, the VPC databases, and the VPC 2.5 site by improving the format, organization, and content of these services. Because people often

understand data better when it is presented visually, the new content will include easy-to-understand charts and graphs that will allow for easy use of the VPC data.

3. Project Objectives and Scope

The purpose of our project was to organize and disseminate the research produced at the Venice Project Center since its founding in a way that is informative and easy to reuse. The objectives of this project were to:

- Validate and integrate VPC data for publication in print and online
- Develop a tool to facilitate the production of interactive visuals
- Use the tool to create specific visualizations for print and online publications
- Make the VPC data visualizations accessible to the public via an online platform

Our project took place during the seven weeks we were in Venice, from October 23rd to December 16th, and was limited to looking at the VPC data, specifically within the time frame of 1988 to 2017. To complete this project, we examined online sources including the data from the CKC (City Knowledge Console), the VPC website, Venipedia pages, and past IQP projects. We were able to make a number of deliverables, all described in latter sections, which include sheet containing the locations of data, a Sandbox Application site to produce visuals, and a data site to house visuals we have published.

With the 30th Anniversary coming up next year, the VPC will be publishing a series of 30 booklets. Our project prepared the VPC for this feat by making a way to showcase all of the work that has been completed there. Our deliverables made it possible to display the data that has been collected at the VPC, through the production of the data visualizations. Below in Figure 8 is the cover of the bridge book.



Figure 8 Bridge Book Cover and Example Page

4. Validating and Integrating VPC data for publication

The VPC has accumulated a large amount of data over the past thirty years. Over time, it has stored data in many different platforms and formats, often migrating its data to make use of the most modern technology available. The VPC now uses Firebase to store its data on the urban features of Venice.

We classified the current status of the VPC data stored in Firebase using the spreadsheet pictured in Figure 9 below, and began by validating, collecting and integrating existing VPC data sets that were used for the online interactive visualizations.

Priority	Topics	Team	Dataset	GIS Layer	Venipedia Template	IQPs
1	Bridges	VE17-KNOW	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		www.venipedia.org/wiki/index.php?title=Bridges	Canals, Bridges and Urban Maintenance
2	Venice Islands	VE17-OPEN	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Island	The Rebuilding of the Minor Islands
4	Fountains	VE17-SAVE	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Fountain	Public Art Preservation in Venice: No
5	Canals	VE17-KNOW	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Canal	Canals, Bridges and Urban Maintenance
6	Docks	VE17-OPEN	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Dock	Monitoring and Analysis of the Cargo
7	Boat Traffic	VE17-MOVE			http://www.venipedia.org/wiki/index.php?title=Canal	The Moto Ondoso Index: Assessing H
8	Moto Ondoso	VE17-MOVE			http://www.venipedia.org/wiki/index.php?title=Moto_Ondoso	The Moto Ondoso Index: Assessing H
9	Demographics	VE17-STAY				Vacation Rentals and Residential Hou
10	Housing	VE17-STAY				Safe and Sustainable Tourism: Manag
11	Tourism	VE17-STAY			http://www.venipedia.org/wiki/index.php?title=Demographics	
12	Coats of Arms	VE17-SAVE	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Coats_of_Arms	
13	Patere and Formelle	VE17-SAVE	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Patere	
14	Street Altars	VE17-SAVE	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Street_altars	
15	Lagoon Islands	VE17-GREEN	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Lagoon_Islands	Place Knowledge/Organization Data on
16	Bell Towers	VE17-KNOW	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Bell_towers	Preserving Venetian Bell Towers Thro
17	Churches	VE17-OPEN	https://drive.google.com/file/d/0B45SQVkvb38CMV9uaGFYQIM52g/view?usp=sharing		http://www.venipedia.org/wiki/index.php?title=Church	Ecclesiastical Architecture of Venice:
18	Convents	VE17-OPEN	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Convent	Ecclesiastical Architecture of Venice:
19	Wells	VE17-SAVE	https://drive.google.com/drive/folders/0B45SQVkvb38CMV9uaGFYQIM52g		http://www.venipedia.org/wiki/index.php?title=Wells	Preserving Venetian Wellheads (2000
20	Stamps	VE17-OPEN				

Figure 9 Data Organization Sheet

4.1 Assessing the completeness and correctness of existing data

To validate the existing VPC data, we first checked it for completeness and correctness. We defined completeness as a measure of how much data is present in each data set, and correctness as how much of the data is valid (contains a value that makes sense for that type of data). This means that for each piece of data we needed to have a standard way of determining if an entered value was within a reasonable expected range of values, and also a way to determine when data was missing. For example, having a bridge height entered as 1,000 meters in the database does not make any logical sense in the context of Venice where the average bridge height is about 2.5 meters. It is also not possible to have a negative number of objects.

There were three different values that could indicate nonexistent data: empty, null, and zero. A number of the cells were simply empty and had nothing to indicate value in them. Others contained the value "NULL," which is normally used to represent something actually having no value. A few other cells had a numerical zero in them. The problem with this system was that

using empty values and zero values is ambiguous. An empty cell could mean that the data exists and has simply not been filled in at the current point in time. The number zero could be the actual value of the field, for instance an elevation or count of objects. Figure 10 shows the percentage of NULL and empty fields to data fields for the CKC data sets at the start of our project.

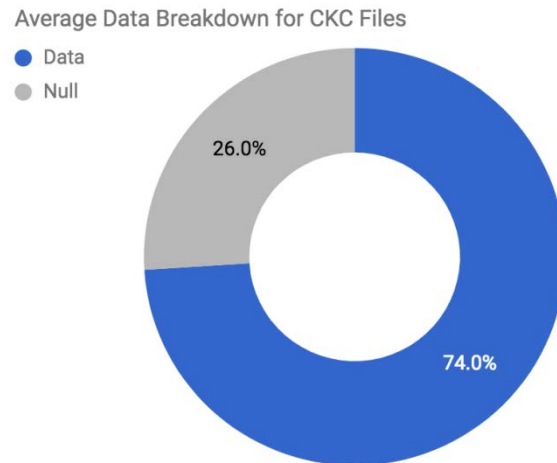


Figure 10 Breakdown of Average Null and Empty Fields to Data Fields

For this reason, we used “NULL” to define the proper lack of data for a field. To make this change, we searched through the data sets to find rows with a zero, empty, or “NULL” value and replaced them with “NULL.” We also had to make judgement calls in some cases where the value 0 may have been correct for that particular field, so we didn’t incorrectly change the value to “NULL.” For example, in the Bridges data set, a number of entries had a value of zero listed for the total number of steps. Initially believing that this indicated a missing value, we were surprised to find during a field visit that some bridges, such as the Ponte dei Meloni pictured in Figure 11 below, actually have no steps.



Figure 11 Field Visit to the Ponte dei Meloni

4.2 Collecting missing data

While validating the existing data, we came across pieces of data that we had to fill in ourselves, either because the data was not present in the database or the current value fell outside the acceptable bounds that we defined. For each field that needed to be filled in, there were two possible cases. The first possibility was that the data had already been collected by someone working for the VPC in the past. In this case, we searched through old IQP projects, offline documents stored physically at the VPC, and old versions of data sets stored either online or offline. We decided to first check these sources, since it took less time to search through text than to go out looking for data ourselves. One example of this was the locations of the bridges, which was an incomplete data set missing 66 latitude longitude pairs. To complete these, we had to go through the QGIS layer containing the bridge locations and extract them from the attribute table. The second possibility was that the data had not already been collected. In these cases, we went out into the city to collect the data ourselves. One example of this was going out to collect step data for the bridges that were missing this information.

4.3 Adding data to Firebase

Once we collected the missing data, we entered it into the VPC's Firebase database. This entailed updating the data sets present in the database using the commands defined by Google in the documentation. In order to help facilitate the analysis and visualization of future data, we also created a new entry for each data set that defines categories for it. These categories will allow future projects to more easily sort the data sets and determine which types of charts and graphs will work for each data set. Different data sets required different visuals, so sorting and categorizing the data by topic made this process easier. In the case where data sets may be related in some way (such as with canals and bridges), we made sure to include in the data a two-way link between the related items. For example, a bridge would have an entry about which canal it expands over, while a canal would have a list of bridges that cross it.

4.4 Results

In order to complete all of our goals for the project in time, we chose to focus only on completing the bridges data set, as it was to be used for the first booklet in the Knowing Venice series. We established the guideline of using "null" to represent missing data. By using the methodology above, we were successfully able to validate and complete the data set, and integrated that data into firebase so it could be used by both our visualization tools and future VPC projects.

5. Developing a tool to facilitate the production of interactive visuals

To make the VPC data more understandable, we produced visualizations of the data sets. Before we created these visuals, we analyzed the data sets to determine the most effective ways of presenting them, and created the Sandbox Application to streamline the visualization process.

5.1 Determining visuals that our tool should produce

The Data Visualization Project (datavizproject.com) is a website produced by the infographic company Ferdio that contains a list of over 150 different types of visuals, with descriptions and examples, that can be filtered by input data types. We used the information on the Data Visualization Project site to create a list of possible visualizations that we could create. Next we determined what visuals best represent the VPC's data sets, by considering feasibility, how many uses each visualization would have, and how interesting they would be. We then compiled a list of all the visualizations that could be potentially used for the data sets (see Appendix G).

5.2 Building the Sandbox Application

In order to automate the production of interactive and static visuals, we created a web-based app that provides controls to customize and export selected charts, graphs, and maps for publication. The creation of this web app, titled the Sandbox Application, involved the use of Firebase, D3.js, Google Maps Application Programming Interface (API), and Materialize CSS. The development of all the components of the Sandbox Application was the most time consuming component of our project.

D3.js is a Javascript visualization library that will allow us to easily create interactive visuals while writing minimal code. The real power behind D3.js is its data management and standardization capabilities which allow for immense flexibility when creating the visual. Because of this flexibility, it is very easy to integrate other Javascript libraries such as the Google Maps API.

The Google Maps API allows for the creation of custom Google Maps embed in web pages. Customizations include the look and feel of the map, color, heat map overlays, and markers with expandable info box which can be added to the map to create a truly unique Google Maps experience. By utilizing Google Maps API's custom look and data overlays as well as D3.js's data management, we will be able to produce map visuals of the geographic information contained in the data sets.

Materialize is a styling framework that aids with consistent site-wide design for forms and elements, such as buttons, input fields, sliders, and fonts.

5.3 Results

We were successful in creating the Sandbox Application for automated visual production. The published site can be accessed at sandbox.veniceprojectcenter.org. We were able to create automation for 11 visuals: Bar chart, donut chart, bubble chart, scaled up number, data view, route map, bubble chart map, category map, donut chart map, choropleth map, and filter map. Descriptions of these can be found in Appendix H. All of the visuals have some interactive component to them when viewed in a browser. A full list of visuals that can be created using the Sandbox is as follows:

There are two options for a finished visualization; it can be exported as an SVG, or published to the data site. Exporting a visualization as an SVG allows for further refinement in Adobe Illustrator or other editing software for inclusion in printed publications. Publishing it to the data site provides a simple way for an infographic to reach the public. You can see Chapter 7 for more information on the data site.

6. Using the tool to create specific visualizations for print and online publications

For its 30th anniversary, the Venice Project Center will be publishing a series of 30 booklets, titled the *Knowing Venice* or *Conoscere Venezia* series, with each volume focusing on a single topic pertaining to the city of Venice. These booklets will serve as authoritative resources on a variety of topics on which the VPC has collected data.

This series, as part of the popular science genre, is targeted to a broad audience that includes scholars, researchers, government officials, and members of the general public, such as Venetian locals. These booklets will be written by Fabio Carrera, designed by Andrea Toffanello, and published by Giovanni Di Stefano. Initially, the booklets will be published in Italian, followed by versions translated into English.

Our team contributed to this publication effort by building the Sandbox Application, which is used to create informative illustrations to be included in the booklets. Our process for creating visualizations using our tool is shown below in Figure 12. Our map visualizations are especially useful for these publications since a good deal of the information in the book is geographical-based.

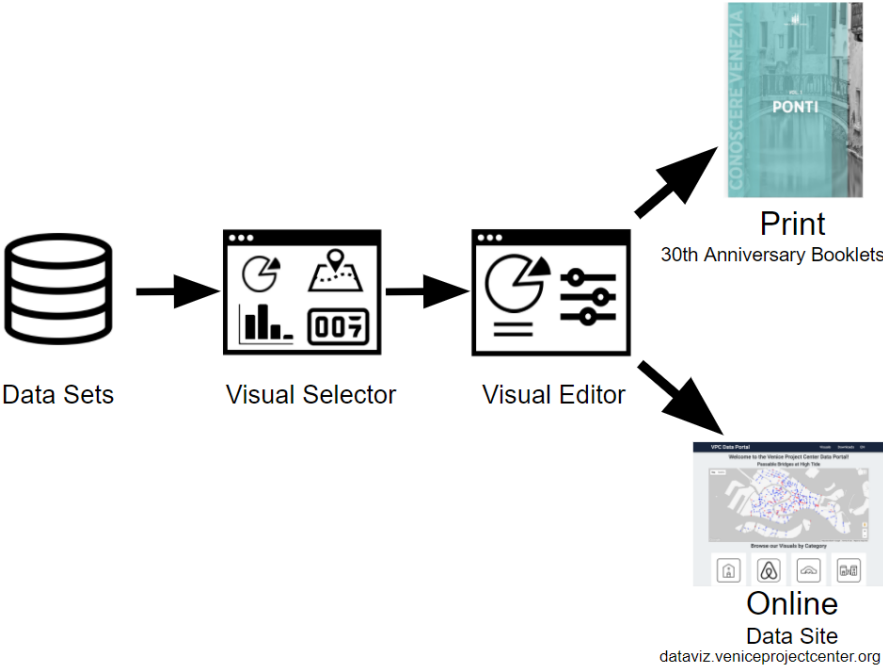


Figure 12 Process for creating our visualizations

The static infographics to be included in the booklets generally represent data that stays consistent over time, while some data sets, such as stores, often change and are best kept online where the infographics can be created in real time as the data is updated. Below, in Figure 13, is a mock static infographic for a booklet is shown.

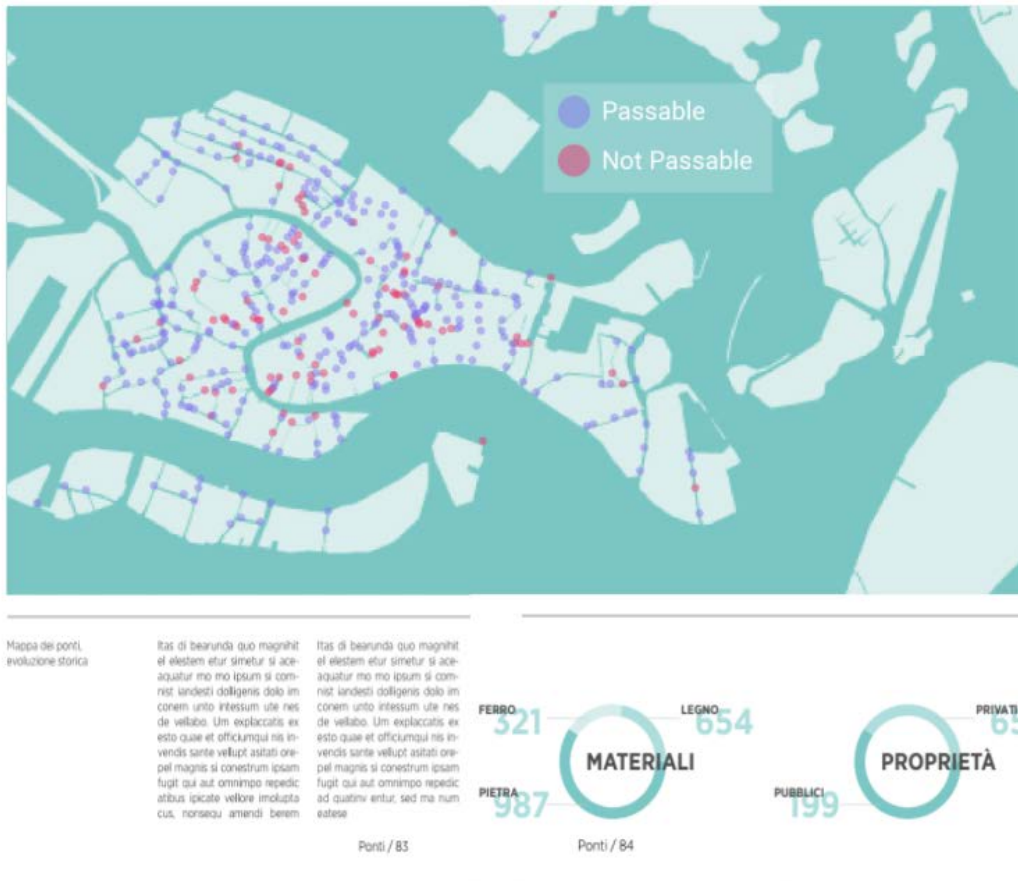


Figure 13 A visual we created include in a page from the prototype booklet on bridges

6.1 Case study: Creating a visualization from start to finish

As discussed in section 2.1.2, knowing bridge heights around the city can be essential information for an ambulance driver when determining which bridges they are able to pass under to reach someone in need. Trying to view this data in a spreadsheet manually would be an time-consuming process, and figuring out which bridges on your route are blocked due to current tide levels would take a long time to figure out. Using the Sandbox Application, you are able to generate a visual with an interactive slider that is able to let an ambulance boat driver set the tide+boat height to a specific level on the fly, and view all bridges they are able to pass under on a map. To begin, visit sandbox.veniceprojectcenter.org, and you will be presented with a list of datasets to create visuals for (seen in Figure 14). Select the data set “Bridges”.

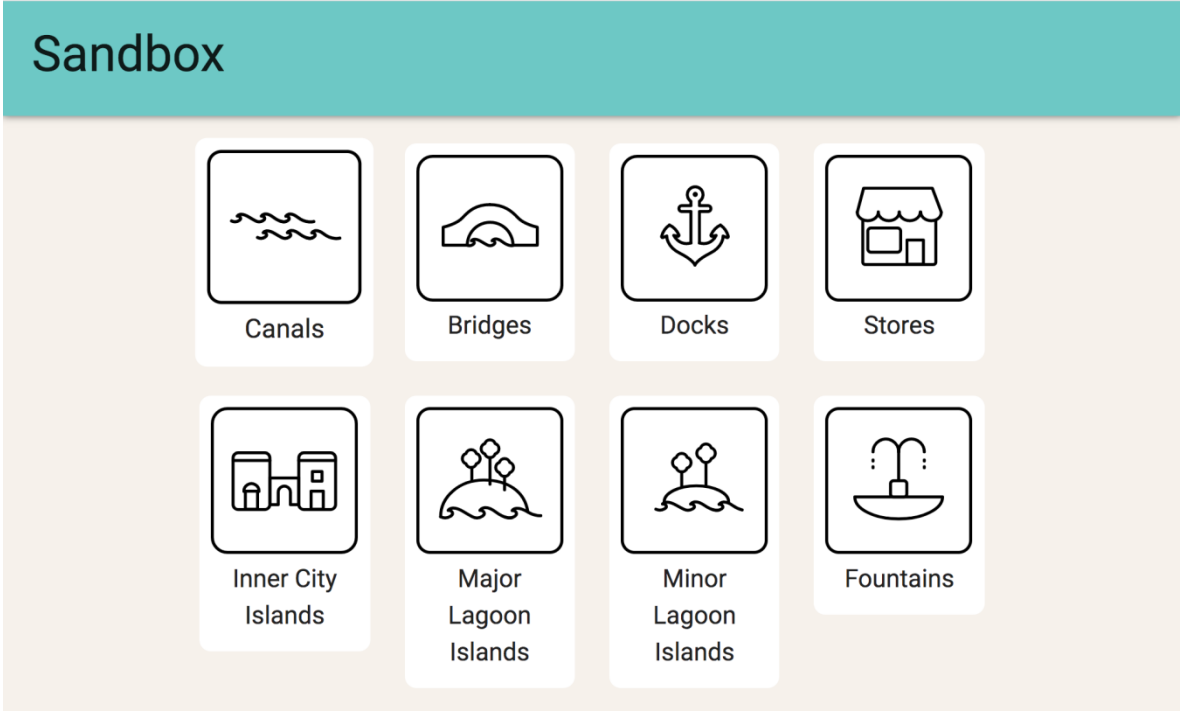


Figure 14 Selecting a data set in the Sandbox Application

Next, you are presented with the different types of visuals you can produce with this data set (seen in Figure 15). Select “Filter Map”.

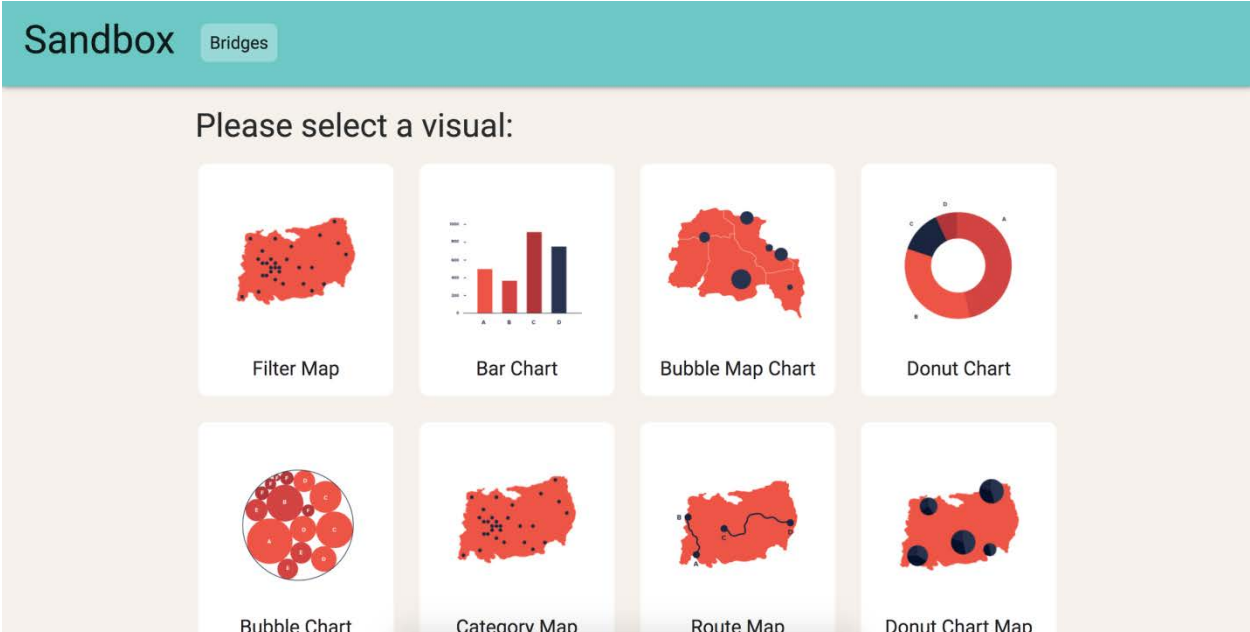


Figure 15 Selecting a visual to use with the Bridges data set

Finally, you are presented with the Filter Map controls on the right, as well as a space for the preview of your visual on the left. Make sure the dropdown on the right is using the Bridges

data set, then under the “Select a property” dropdown, click “Circles”. Using the color picker, you can select the color you want the bridge circles to be displayed as, in this case blue.

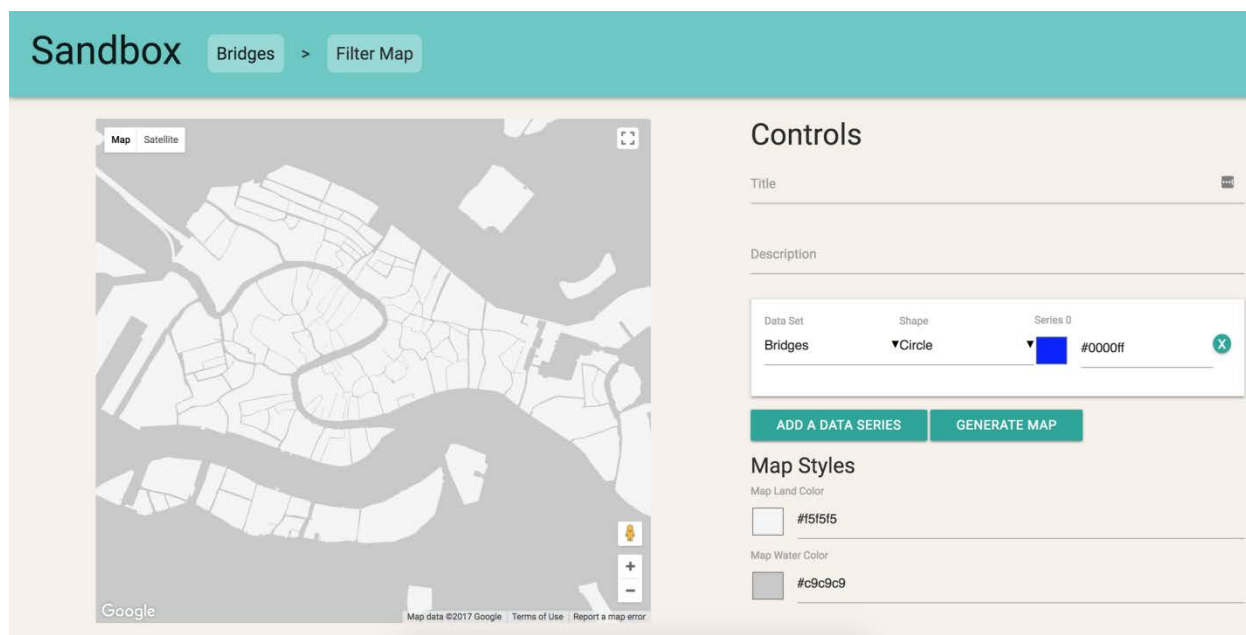


Figure 16 The Sandbox Application Filter Map configuration screen

Next, click any of the whitespace on the white card, and you will be presented with the available filters. Scrolling to “Numerical Filters” use the “Select numeric property” and select “Height 1m North”, then in the “Operation” dropdown click “>”. Instead of defining a static value to filter by, click the checkbox below the filter to enable a slider option, and enter the values 0.5, 0.25, and 5 (lower bound, slider step size, upper bound). Be sure to give the slider a label as well in the text box below. In this case the label should be “Tide Height + Boat Height (m)”. Finally, scroll down and click the “Generate Map” button, and you can see the resulting visual displayed on the left side of the screen, as displayed in Figure 16. The visual is now ready to be downloaded as an SVG or published to the data portal.

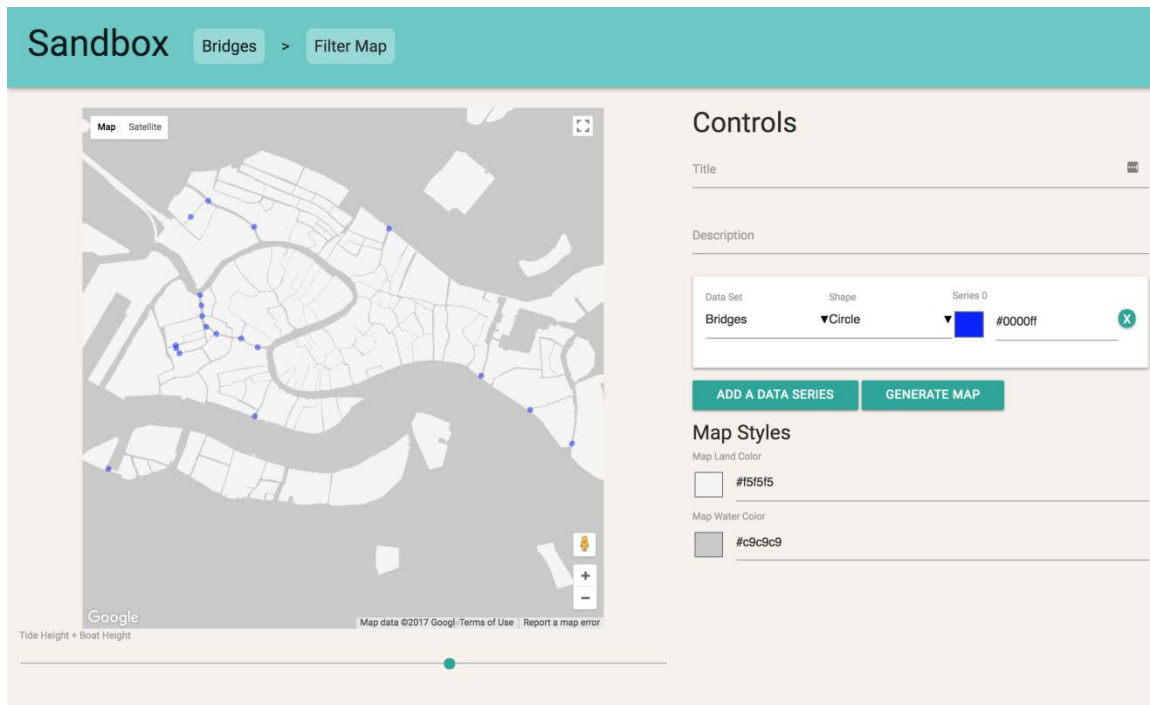


Figure 17 The final bridge height interactive visual

As previously mentioned, graphics like this map have important real world applications. The bridge height visual specifically can be useful for ambulance boats during high tides as the rising water levels can cause some of the bridges to become impassable by boats. Having a way to display this data can help ambulance drivers map out their route during high tides and could potentially save a person's life. For example, in Figure 17 above, the slider has been set to 3.5 m, which would correspond with the record high canal level of 194 cm plus the ambulance boat height of 1.5 m. As you can see, there are only a couple bridges that would be passable at this tide level, so this information could be life-saving if it prevents an ambulance boat driver from taking a blocked path.

These visuals are also completely generalizable, so this slider application could be used with any data set and any numeric column to show interactively filtered data points on a map.

6.2 Example Visualizations made using the Sandbox Application

You can utilize our tool to create a variety of visualizations on any of the myriad of topics the VPC has collected data on. The flexibility of our Sandbox Application allows for a better understanding of all this data, which is as simple as choosing which data set you want to visualize on its first screen. We were able to create a number of visuals to represent the VPCs data. After looking at the data collected on the bell towers, we determined an informative visual could be created on the number of bells in each tower. From the visual we created, shown in Figure 18, the conclusion that a majority of the bell towers have 4 bells, while none have a single bell can be drawn.

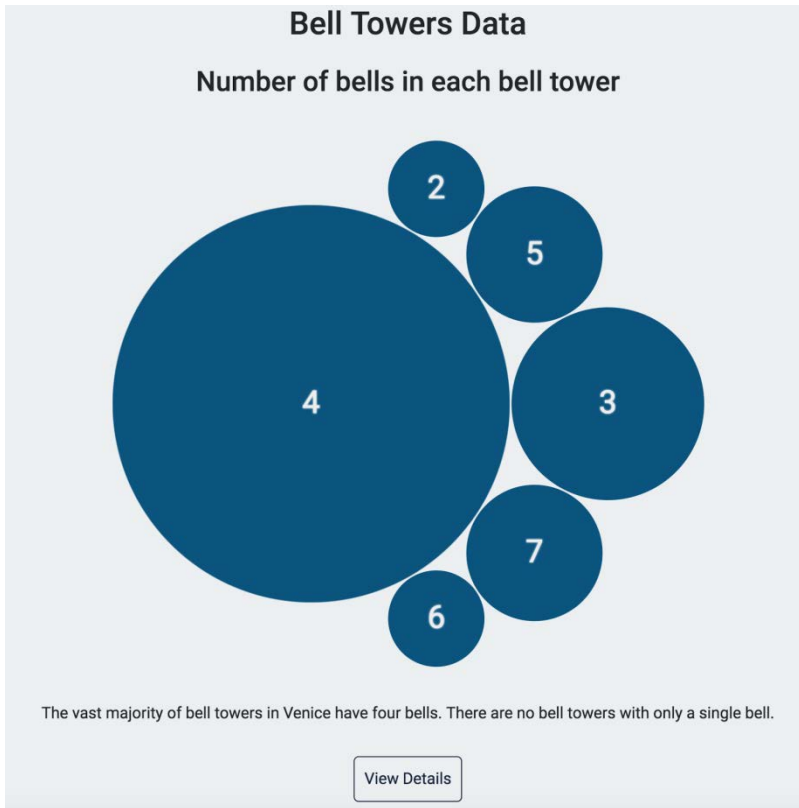


Figure 18 Number of bells in each bell tower

After looking through the data on the islands of Venice, we determined that a visualization representing the data from the most recent census would result in a graph of the distribution of the Venetian population, seen in Figure 19. This can be repeated with the data from successive censuses, which would allow for the analysis of the population over time.



Figure 19 Distribution of Venice population from 2011 census

The Stay team was able to collect data on the Airbnbs that exist throughout Venice. We were able to use our Sandbox Application to create a visual that uses this data and plots the data based off of two attributes, cost and annual occupancy rate, shown in Figure 20. The occupancy rate was represented by color, whereas the cost was represented by the size of the circles on the map. The final visualization can be used to find out where tourists are staying and where they are spending their money.

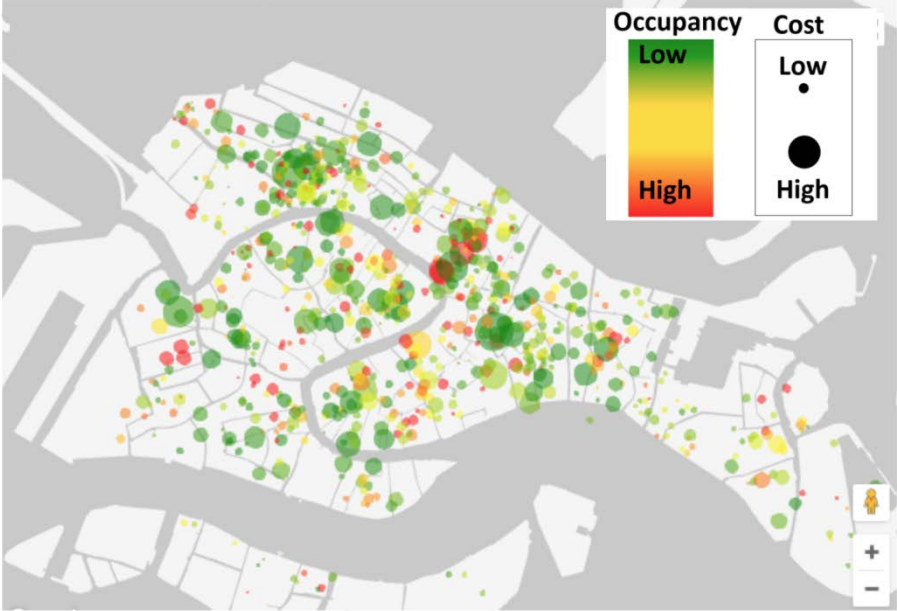


Figure 20 Airbnb listings in Venice by occupancy (color) and price (size)

Looking at the data the VPC has collected on the stores within Venice, the categorization of these by type of store would make an interesting and useful visual, especially looking at the souvenir stores. Figure 21 shows the distribution of these, which shows a concentration along areas with heavy tourist traffic, such as near the Rialto and in San Marco.



Figure 21 Distribution of souvenir stores in Venice

The data collected on the public art throughout Venice includes their condition. We were able to take this data and create a map of which pieces of public art are most in need of restoration, seen in Figure 22. This is useful, as it shows which areas are in need of resources to maintain and repair the art.

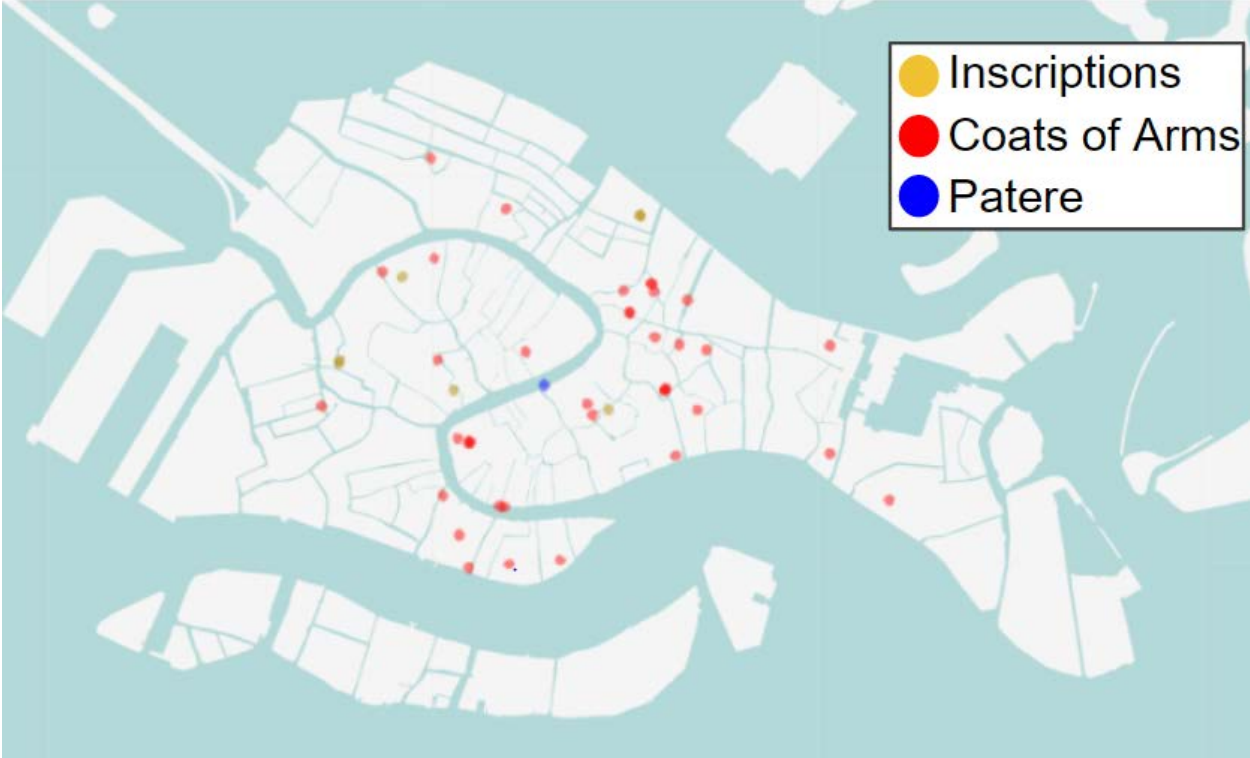


Figure 22 Public art in Venice most in need of restoration

7. Making the VPC data accessible to the public via an online platform

One of our main objectives in this project was to increase the accessibility and ease of use of the vast quantities of data the Venice Project Center has collected. The Sandbox application, as described above, provides a robust way to create visuals using the full extent of the VPC data. This application will be used to create graphics for each of the 30 *Knowing Venice* booklets that will be produced for the VPC's 30th anniversary. Once these data visualizations are created, they must be accessible to the public in a way that is engaging and easy to use. To accomplish this, we created the 'VPC Data Portal' website, with a section for each of one of the topic of the booklets. This website will function as a companion to the booklets, showing off the interactive versions of the graphics included.

7.1 Overview of the VPC Data Portal

This data portal website (dataviz.veniceprojectcenter.org) exists as a showcase for interactive visualizations for various data sets. In order to create these visualizations, we used our Sandbox Application. The website features much of the same code as the Sandbox Application site but with all the configuration tools removed from the user interface leaving only the visual. The site starts by reading a Firebase stored configuration file, which holds information telling which data set, visual, and customizations to use. It uses these parameters, rather than user entered ones, to render the visual in real time on the site. The benefits of rendering the visual real-time rather than inserting a static image is that if the data in Firebase ever changes, the visuals will all immediately and automatically reflect the changes. The slashpage of the data portal can be seen in Figure 23 below.

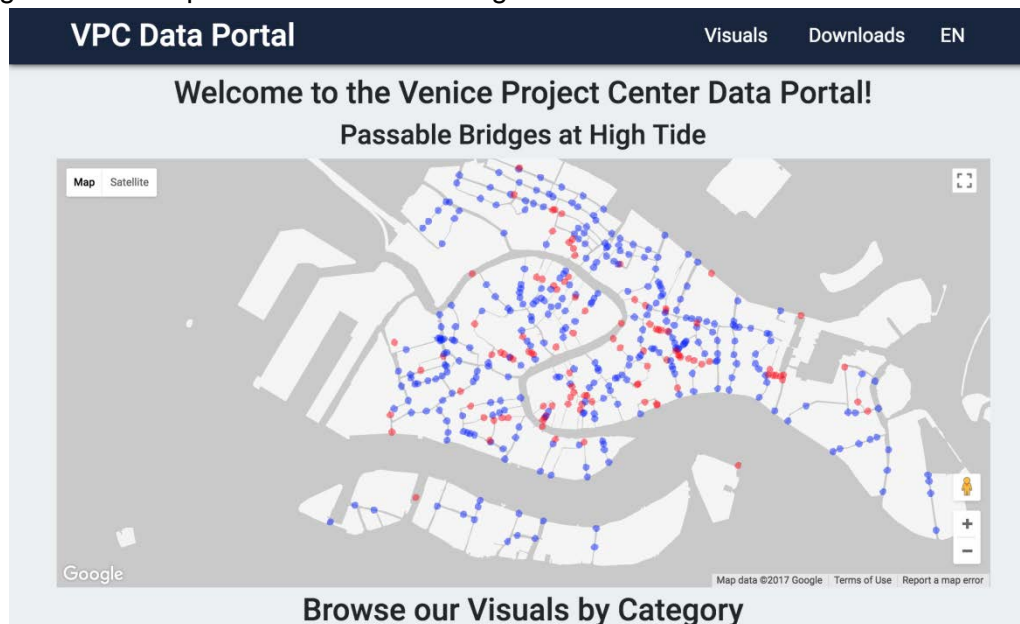


Figure 23 The VPC Data Portal website main page (dataviz.veniceprojectcenter.org)

We used the Firebase deployment environment to host our website. First, we set up a basic project in Firebase using the web console. Then, after developing the website locally, we added the Firebase configuration files to the project using the Firebase command-line utility. We then deployed our code to the Firebase server and domain that was connected to the project we had created before. At this point, we were able to simply deploy our code from the command line and within minutes the hosted site would reflect the changes we had made, which made the development process very easy.

7.2 Examples of Interactive Graphics Hosted on Our Site

After clicking any of the categories of data visualizations on the main page, the user is presented with a carousel of graphics that showcase the full extent of the selected data set. Shown below is the carousel shown after clicking the bridges data set.

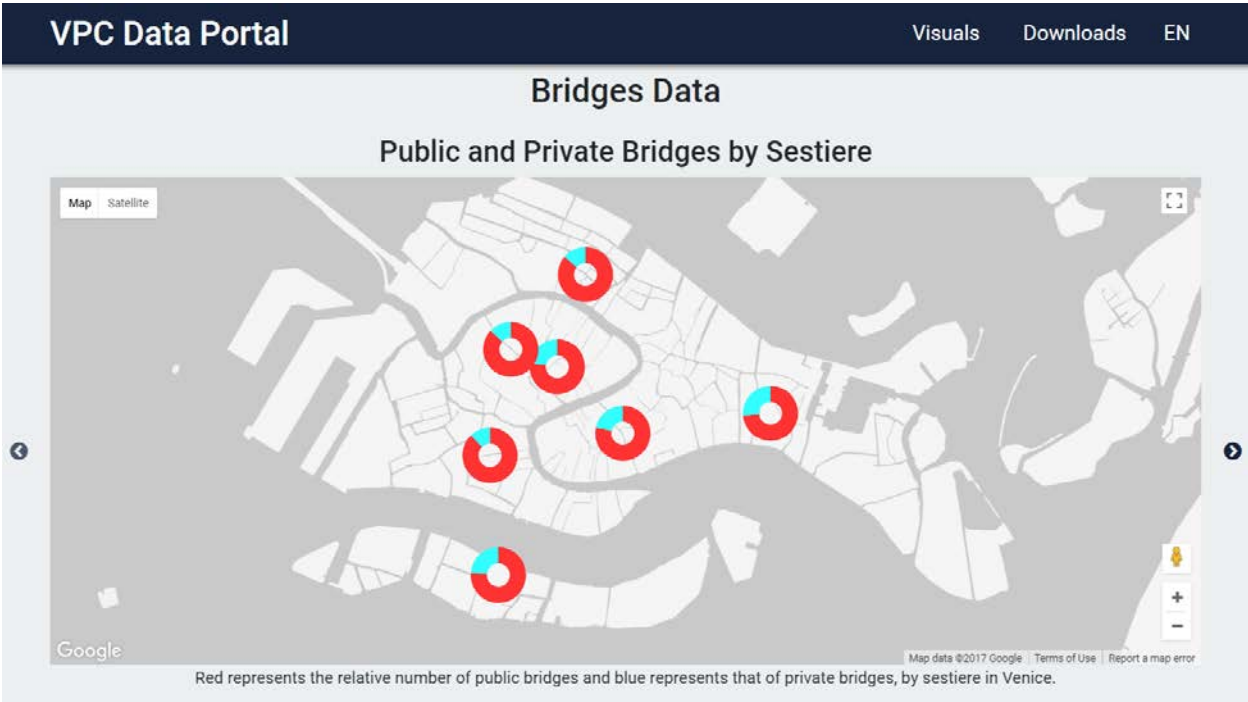


Figure 24 The carousel shown after selecting the bridges data set

As one can see, the arrows to the left and the right of the screen may be clicked to move through the visualizations available for the data set selected. In Figure 24 shown above, the carousel is on a visualization entitled, 'Public and Private Bridges by Sestiere'. This was created using the Donut Chart Map visual in the Sandbox. It shows the relative proportion of public bridges (red) to private bridges (blue) in each of the sestieres.

Another example of an interactive graphic from the bridges data set is shown below. It was created using the Filter Map visual in our sandbox site, as shown in the section 6, and intuitively displays which bridges are passable at a given tide level. The slider allows one to adjust the filter for bridge height: by adjusting the slider to the sum of your boat height and the tide level, you can see which bridges your boat can fit under.

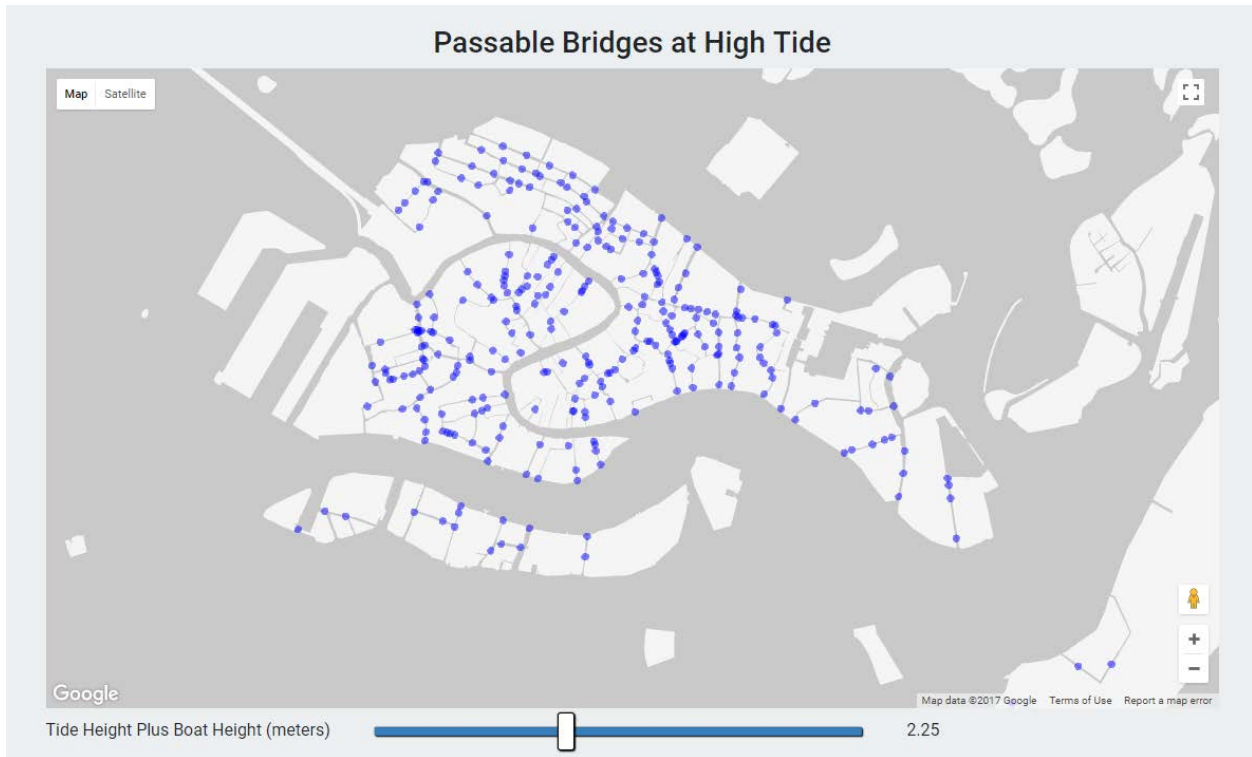


Figure 25 Passable Bridges at High Tide visual with a tide plus boat height of 2.25 meters

Figure 25 shows which bridges are passably by a 1.5 meter boat, such as an ambulance boat as mentioned earlier in the report, at a decently high tide of 75cm. As one can see, a large number of the bridges are still passable. This changes as tide is increased.

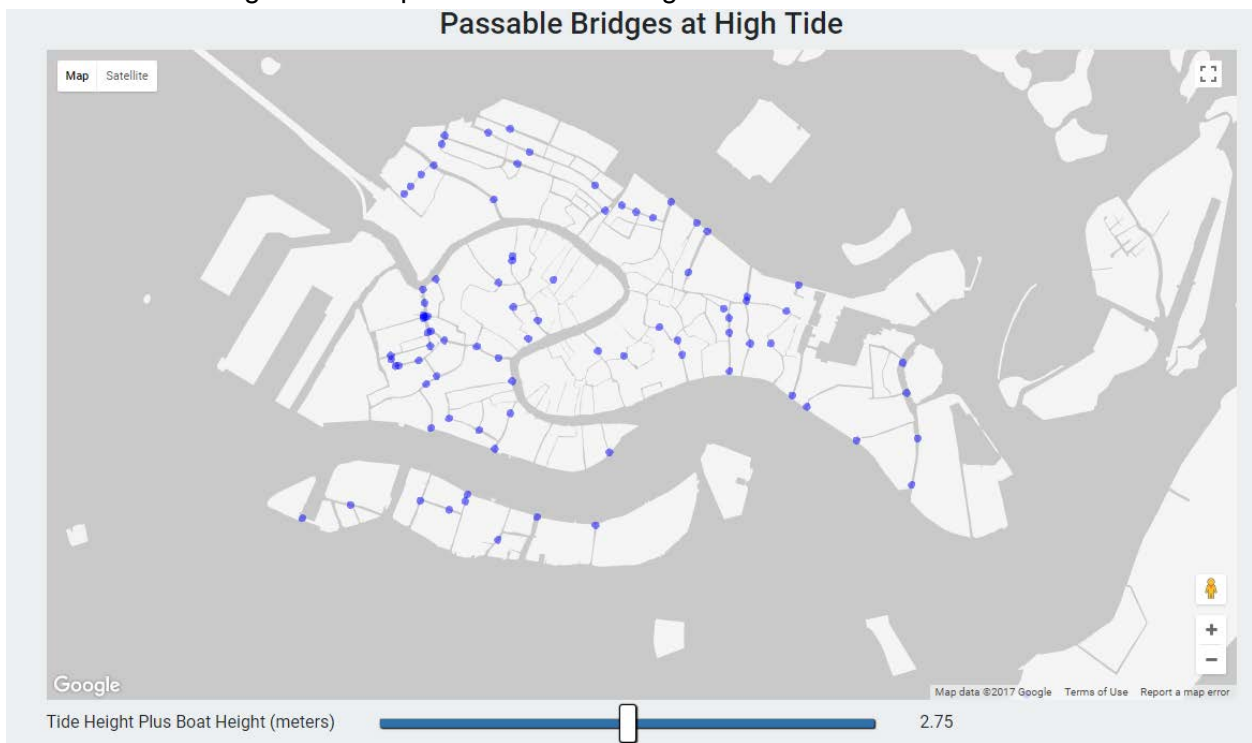


Figure 26 Passable Bridges at High Tide visual with a tide plus boat height of 2.75 meters

Figure 26 shows the bridges passable for the ambulance at a very high tide of 125cm. This tide is a good size *Acqua Alta*, but still likely to occur every few years. As you can see, there are significantly fewer bridges that the ambulance could go under to reach its destination. By using our tool, lives can be saved as ambulances are assured to take the fastest possible route to their destination.

7.3 Data Portal Downloads Page

In order to accomplish our goal of making the VPC data available to anyone who may wish to use it, we created a downloads page that allows anyone to receive the current copy of any VPC data set. The page looks similar to our Sandbox data set selector, except when a user selects a category, the database is downloaded in their browser in JSON format. At the time of writing, the data sets are licensed under the Creative Commons Attribution-ShareAlike 4.0 International License, meaning other organizations can use the data as long as they credit the VPC. The downloads page can be seen in Figure 27.

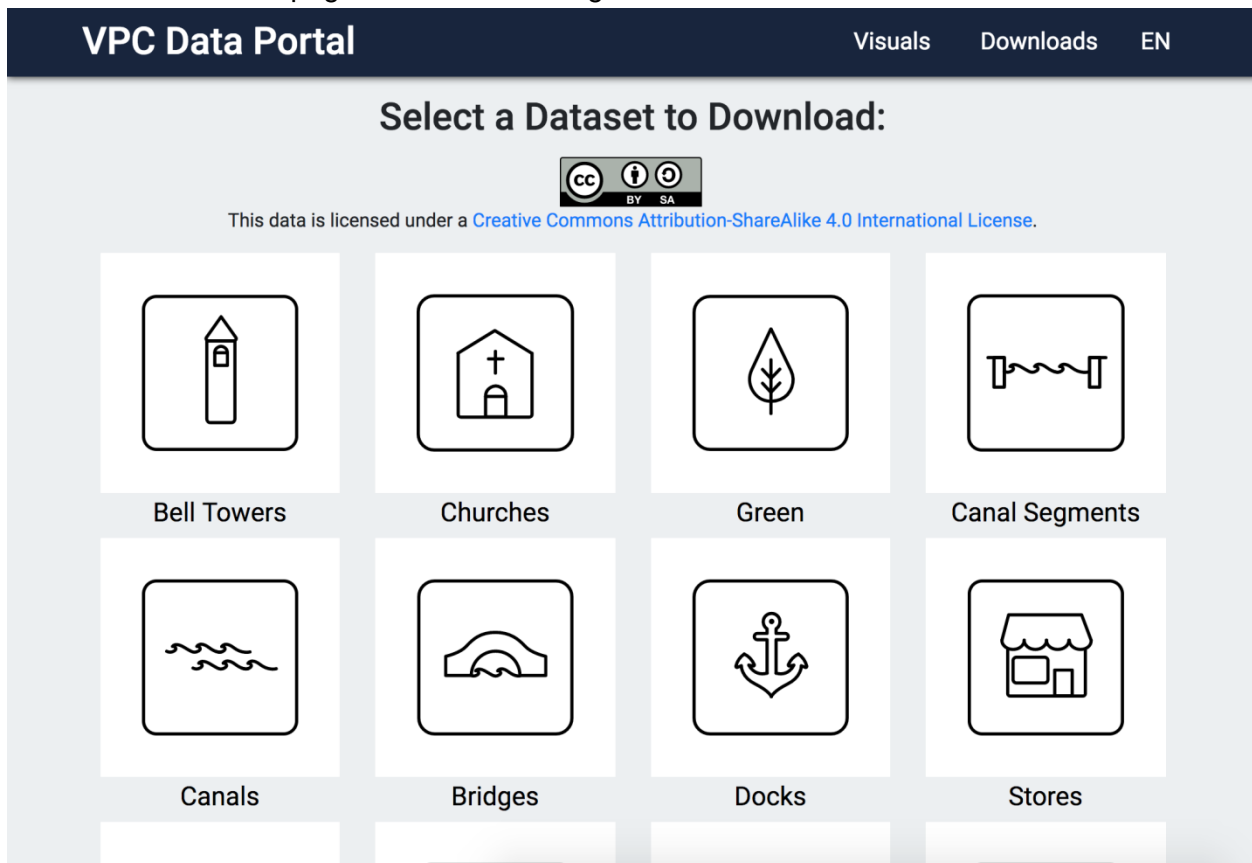


Figure 27 VPC Data Portal Downloads Page

7.4 Data Portal Embedding

The final feature of the data site is the ability for anyone to take the embedded code of any visualization and embed it in their own website. Each visual has a copy code button on its

page as shown below in Figure 28. This greatly helps disseminate the information collected by Venice Project Center, as anyone who thinks the data could be useful for visitors of their site can place the visual on their site, and the data backing the visual will stay up to date with the Venice Project Center databases. An example of an embedded visual can be seen in Figure 29.



Figure 28 Copy embedded code example

[Home](#) [Articles](#) [About](#) [Contact](#)

Welcome to Knowing Venice Team's example website!

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam efficitur felis a massa suscipit, at fermentum tellus cursus. Morbi pretium elit non mi pharetra, vel convallis arcu euismod. Nulla consequat blandit varius. Suspendisse tempus leo sed fermentum ultrices.

Passable Bridges at High Tide

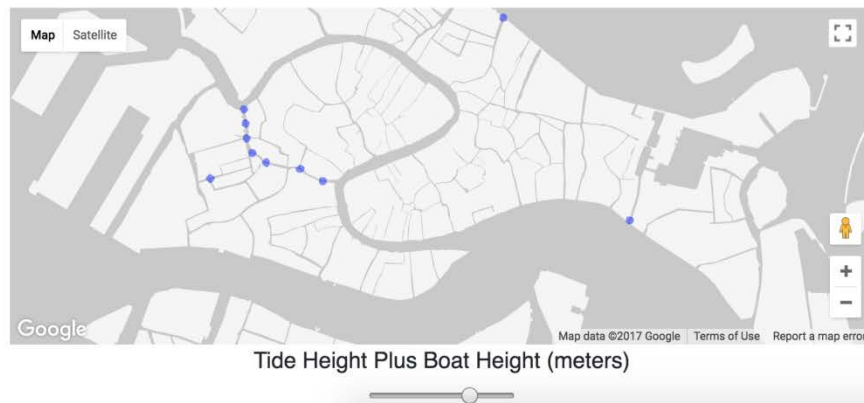


Figure 29 An example embedded visual on a test website

8. Conclusion and Recommendations

The goal of this project was to make the data collected over nearly 30 years by students at the Venice Project Center more accessible online and in print. We were able to make tools to aid in the production of visuals of this data that can be used both in the 30th Anniversary booklet series as well as online, in an interactive form. We created a Sandbox Application site in which we were able to create visualization as well as our data site to which these are published.

We were also able to create an easy portal for downloading up-to-date copies of the VPC data repositories. There are many social, economic, and technical benefits that result from opening data for public use. Some of these benefits entail empowering startups, aiding government policy, and creating transparency for the public. By creating an easy portal for downloading the latest copy of our data, we've created opportunities for creating more jobs, and allow other developers to create applications for Venice's benefit.

We have concluded that our Sandbox Application is a superior way to create visualizations of VPC data as it updates automatically, uses a consistent style, and produces high quality infographics. Below Figure 30 summarizes the processes we followed to create both static and interactive graphics with the Sandbox Application. The most important of these features is the automatic updates. A common problem across many VPC tools is that they use outdated data or broken aspects. By using our tool, the use of current data is assured and only one tool would need to be fixed.

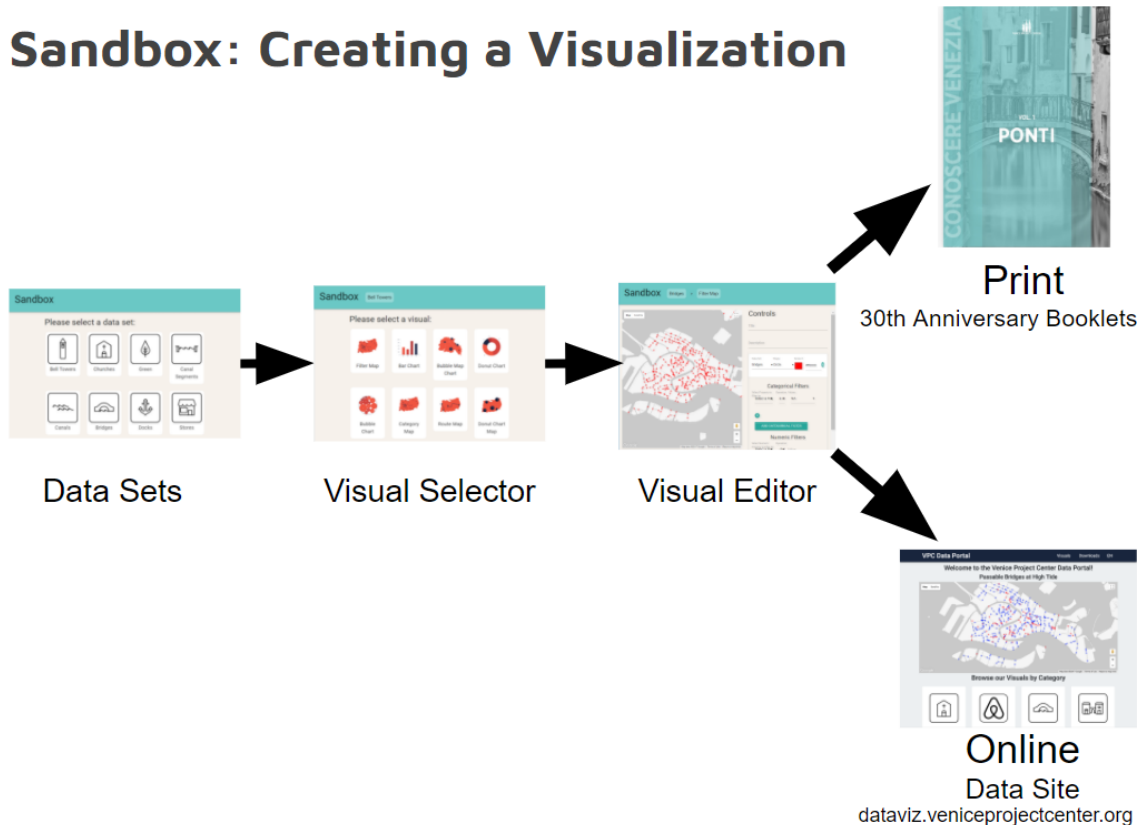


Figure 30 Process for creating our visualizations

8.1 The Future of the Data Site

Our website, dataviz.veniceprojectcenter.org, includes the visualizations published from the Sandbox Application as a companion to the booklets. Although there is the option to have the site in Italian or English, all of the visuals are published with the exact text entered by a visualization creator. One suggestion for a completely translated site would require the addition of the translation in the Sandbox Application as well. On the data site, within the articles tab, there is a sample article on bridges. The Venice Project Center should continue to upload interesting articles in both Italian and English to the site. The site could also be expanded to include easy access to all of the functional interactive web tools created by past projects, such as the DeBarbieri map viewer and the Bells site.

8.2 Recommendations for Additional Visualization Tools

With only seven weeks in Venice, there was a limited amount of visuals we were able to create. There are a few other visuals that we were not able to complete that could be created in the future. Similar to the path finding visual we created, a future team may be able to map out all of the possible boat routes and do a path finding visual for boats. This would be complex, as there are many aspects you have to account for when creating this, such as the direction of traffic in the canals, the width of the canals when other boats are docked on the edge of the canal, and the height of the bridges that the boat would have to pass under. Another recommendation for a future teams contribution would be the ability to further combine data sets. Our filter map has the ability to integrate more than one data set for the visualization, but it would be beneficial to add this feature to all of the visuals.

8.3 Recommendations for Creating Graphics with the Sandbox Application

Our Sandbox Application can create visualizations for nearly any VPC endeavour. With our selection of any Firebase Data Sets, many Visualization types, and embedding/exporting capabilities, any visualization can be made and added anywhere. Some uses for these graphics include future booklet publications, the VPC website, and future IQP projects.

8.3.1 Creating Infographics for Future Booklet Publications

At the time of our project, only the bridges booklet was in the works, which is why much of our focus was on the bridges data. The creation of our Sandbox Application, however, allows for the creation of visuals for any data set. For each topic of the booklets, it is possible to customize a visualizations to the topic of the booklet and then export them for a graphic design program, such as Adobe Illustrator.

We suggest that the filter map be utilized for most data sets. It provides the most flexibility for selecting specific data sets, and can provide useful data for any data set that has latitude or longitude. The scaled up number is a good visual for any data set. It allows a user to create a large number with an optional label of the count, sum, or average of any data set. This

visual is usually only appropriate for exporting as an SVG, as it creates a single labeled large number.

The stacked bar chart is helpful when grouping the data by certain criteria makes sense. For example, one can make each bar represent a sestiere and each portion of a bar represents the number of bridges of a certain condition in each sestiere. The donut chart is useful for showing percentage data, such a bridge material, church usage, etc. The chart should use either the legend or always display labels as the graphics will be static.

8.3.2 Creating Infographics for VPC Websites

The VPC websites will use primarily interactive graphics. Designers should focus on our map graphics such as the filter map, route map, and bubble chart map. All of these have interactive features making them preferable for websites. When using the donut chart, the on hover display labels is a great option as it add a dynamic aspect to an otherwise static visualization.

8.3.3 Creating Infographics for Future IQPs

Future IQP groups can utilize our tool to easily create visualizations for their projects. They can create static infographics for papers and presentations, such as filter maps, donut charts and bar charts. They can also use our dynamic visualizations for any websites or web tools they create. Our filter map, route map, and bubble chart map are especially useful for interactive purposes through an online platform. Using our tool is preferable to other methods as the visualization will update automatically as the VPC's database is changed

8.4 Recommendations on Data Integration

We recommend that users follow our format of using NULL to represent missing data. By establishing this as a standardized process, all of the data sets will be easier to iterate through. For now, the Data View and Filter Map tools allow users to identify missing data in a certain column and where the object is located.

8.4.1 Creation of a Diagnostic Tool

To help with identifying missing or incomplete data, we recommend creating a diagnostic tool (similar to the filter map and data view tools) to look through the data sets. This would look through a data set and identify any fields that are missing values (NULL or empty) or contain values that are suspicious, such as values that fall far outside of the range of all other values in the column. This would allow for a more automated way of validating data sets, instead of manually going through each data set to check for completeness like our team had originally done.

9. Bibliography

- Arup. Retrieved from <https://www.arup.com>
- Benefits of open data. Retrieved from <https://www.europeandataportal.eu/en/using-data/benefits-of-open-data>
- Behmke, James M.; Corrado, Tanya M.; Flores, Gabriel M.; Looft, Fred J. (1998). *Venice Project Center's Tenth Anniversary: The Documentation, Interpretation, and Presentation of its History*. Retrieved from https://www.google.com/url?q=https://drive.google.com/file/d/0B0ELG5jkw_b4UDdVWHhpZkFScUE/edit?usp%3Dsharing&sa=D&ust=1512915974866000&usg=AFQjCNECfDnr6vwcMvWLDPH76rDIFADExQ
- Bobell, Patrick Anthony; McCarthy, Lauren Barbara; Poganski, David Michael; Tsiros, Alyssa B. (2012). *Venipedia: An English Language Data-Driven Wiki Dedicated to Venice*. Retrieved from <https://web.wpi.edu/Pubs/E-project/Available/E-project-121412-163337/>
- Bossalini, Dominick Christopher; Gardner, Brian William; Mathews, Tyler Douglas; Wills, Mitchell Edwin. (2013). *Canals, Bridges and Urban Maintenance*. Retrieved from <https://web.wpi.edu/Pubs/E-project/Available/E-project-122013-134409/>
- Brann, Kristen Lee; Giola, Gregory Thomas; Kirby, Dylan Joseph; Richtmyer, William Hames. (2013). *25 Years of Venice Knowledge Online*. Retrieved from <https://web.wpi.edu/Pubs/E-project/Available/E-project-122113-100840/>
- Broks, Peter. (2005). *Understanding Popular Science*. McGraw-Hill.
- Carrera, Fabio. (2014). *Venice project center 25th anniversary*. Retrieved from <https://www.slideshare.net/carreraf/venice-project-center-25th-anniversary>
- Chhim, Sopheaktra; Hctor, Thomas Aniello; Pryor, John William Rogers; Sadowski, Bryan Joseph. (2014). *The Evolution of Venice Through Mapping Technology: A Venice Project Center Collaboration with the Atlante della Laguna*. Retrieved from <https://web.wpi.edu/Pubs/E-project/Available/E-project-121914-112605/>
- data.gov. *U.S. Open Data Action Plan*. (2014). United States White House. Data Visualization Project. Retrieved from <http://datavizproject.com/>
- Dowse, R., & Ehlers, M. (2005). *Medicine labels incorporating pictograms: Do they influence understanding and adherence?* *Patient Education and Counseling*, 58(1), 63-70. doi:S0738-3991(04)00225-3 [pii]
- European parliament election, 1984 (Italy). (2017). Retrieved from [https://en.wikipedia.org/wiki/European_Parliament_election,_1984_\(Italy\)](https://en.wikipedia.org/wiki/European_Parliament_election,_1984_(Italy))
- Finelli, Thomas; O'Brien, Catherine; Scannell, Kevin. (2010). *Venipedia: A Modern Knowledge and Data Wiki Dedicated to the City of Venice*. Retrieved from <https://web.wpi.edu/Pubs/E-project/Available/E-project-121710-165119/>
- Fry, B. (2004). *Computational Information Design* (PhD thesis). Massachusetts Institute of Technology. Cambridge, MA.

Gibson, Daniel Gilbert; Brunelli, John Paul; Orsi, Edward Paul; Sargent, Amanda Rose. (2009). *Showcasing Twenty Years of Venice Project Center Results Using Interactive Online Infographics*. Retrieved from <https://web.wpi.edu/Pubs/E-project/Available/E-project-121710-153620/>

Giuffrida, A. (2017, July 22). *'Imagine Living with this Crap': Tempers in Venice Boil Over in Tourist High Season*.

Gurstein, M. B. (2011). *Open Data: Empowering the Empowered or Effective Data Use for Everyone?* *First Monday*, 16(2)

Help:About. (n.d.). Retrieved September 23, 2017, from <http://www.venipedia.org/wiki/index.php?title=Help%3AAbout>

Introduction to graphs in stata | stata learning modules. Retrieved from <https://stats.idre.ucla.edu/stata/modules/graph8/intro/introduction-to-graphs-in-stata/>

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). *Benefits, adoption barriers and myths of open data and open government*. *Information Systems Management*, 29(4), 258-268.

Kirk, A., & Safari Books Online. (2016). *Data visualization: Representing information on modern web : Unleash the power of data by creating interactive, engaging, and compelling visualizations for the web : A course in three modules*. Birmingham, England: Packt Publishing. Retrieved from <https://proquest.safaribooksonline.com.ezproxy.wpi.edu/9781787129764>

Lima, R. C., De Castro Andrade, R., Monat, A. S., & Spinillo, C. G. (2014). *The relation between online and print information graphics for newspapers*. Paper presented at the , 8519(3) 184-194. doi:10.1007/978-3-319-07635-5_19

Rohrer, M. W. (2000). *Seeing is believing: The importance of visualization in manufacturing simulation*. *Simulation Conference, 2000. Proceedings*. Winter, 2 1211-1216.

Statistics. Retrieved from <http://www.venipedia.org/wiki/index.php?title=Special:Statistics>

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, Conn: Graphics Press.

Venice Project Center. Retrieved from <http://veniceprojectcenter.org>

Venipedia - An English Language Data-Driven Wiki Dedicated to Venice. Retrieved from <http://www.venipedia.org>

Walker, R. (2015). *The Importance of Data Visualization and Automation in the Realm of Big Data*. ACI Information Group. Retrieved from <http://wpi.aci.info.ezproxy.wpi.edu/view/14c57908e4a000e0009/14f61633d090001193f>

Appendix A: Data Source Tracking Sheet

For each of the thirty topics, it is necessary to organize the data so it can be easily retrieved. The information contained within the spreadsheet will include any data that has been previously collected by the VPC and past IQPs, GIS layers if they exist, the City Knowledge data set, if there is a Venipedia template for that topic, as well as any previous IQP reports that contain relevant information. The organization of this data will be completed as a combined effort with the other teams. This spreadsheet is displayed in Figure 31.

<https://docs.google.com/spreadsheets/d/1t-0bPOhL-8u1b51Pdo45jh630iHkVDm7lgKAV4Wkwll/edit#gid=697225169>

	A	B	C	D	E	F	G
1	Priority	Topics	Team	Dataset	GIS Layer	Name on CK console	Venipedia Template
2	1	Bridges	VE17-KNOW	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			http://www.venipedia.org/wiki/index.php?title=Bridge http://www.venipedia.org/wiki/index.php?title=Bridges
3	1	Venice Islands	VE17-OPEN	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			http://www.venipedia.org/wiki/index.php?title=Island http://www.venipedia.org/wiki/index.php?title=Islands
4	2	Lagoon Islands	VE17-GREEN	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			
5	2	Bell Towers	VE17-KNOW	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			http://www.venipedia.org/wiki/index.php?title=Bell_tower http://www.venipedia.org/wiki/index.php?title=Bell_towers
6	1	Fountains	VE17-SAVE	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			
7	1	Canals	VE17-KNOW	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			http://www.venipedia.org/wiki/index.php?title=Canal http://www.venipedia.org/wiki/index.php?title=Canals
8	1	Docks	VE17-OPEN	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			http://www.venipedia.org/wiki/index.php?title=Dock http://www.venipedia.org/wiki/index.php?title=Docks
9	3	Archeological sites	VE17-KNOW				
10	2	Churches	VE17-OPEN	https://drive.google.com/file/d/0B4z5QVXwb8CMV9uGFjY0IM5z/view?usp=sharing			http://www.venipedia.org/wiki/index.php?title=Church http://www.venipedia.org/wiki/index.php?title=Churches
11	2	Convents	VE17-OPEN	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			http://www.venipedia.org/wiki/index.php?title=Convent http://www.venipedia.org/wiki/index.php?title=Convents
12	2	Wells	VE17-SAVE				
13	3	Streets	VE17-KNOW				http://www.venipedia.org/wiki/index.php?title=Street http://www.venipedia.org/wiki/index.php?title=Streets
14	2	Stores	VE17-OPEN	https://drive.google.com/file/d/0B4z5QVXwb8CMV9uGFjY0IM5z/view?usp=sharing			http://www.venipedia.org/wiki/index.php?title=Stores
15	1	Boat Traffic	VE17-MOVE				
16	1	Moto Ondoso	VE17-MOVE				
17	1	Demographics	VE17-STAY				
18	1	Housing	VE17-STAY				
19	1	Tourism	VE17-STAY				
20	1	Coats of Arms	VE17-SAVE	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			
21	1	Patere and Formelle	VE17-SAVE	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			
22	1	Street Altars	VE17-SAVE	https://drive.google.com/drive/folders/0B4z5QVXwb8CMV9uGFjY0IM5z			
23	2	Statues and Monuments	VE17-SAVE				

Figure 31 Data Tracking Sheet

Appendix B: The Open Data Movement

The Open Data Movement is a movement that aims to publish local, regional, and national data, with the hopes that individuals and organizations will use that data for their benefit (Gurstein, Michael B., 2011). The types of data being published can be anything that might be useful to society, such as weather, traffic, tourist, business, political, and geographical data. Many of the projects created at the VPC have dealt with data collection and organization, and the goals of these projects align closely with those of the Open Data Movement. In this section, we address the benefits and barriers of open data, as well as a case study of a company that uses open data to provide sustainable solutions for city architecture, exemplifying the benefits that open data can achieve.

Benefits of Open Data

The benefits of open data can be broken down into three sub-categories: social, economic, and technical. The social benefits include transparency of government operations, accountability, insights into the public sector, and equal access of data for everyone regardless of social status. The United States is one of the leaders in this area, publishing data in categories including but not limited to climate, energy, finance, and manufacturing on its website data.gov. One example of the use of this data is an app that allows people to view the health inspection scores for local restaurants, hospitals, hotels, convenience stores and supermarkets (data.gov). This benefits society by motivating businesses to improve their health conditions, which will create better quality products and services for all. The economic benefits include availability of information to investors and stimulation of competitiveness. If an investor is able to see the need for a certain product or service to exist, having data available to back their observations can help them make an informed decision for their company. As a hypothetical example, if an investor saw that there was high demand for water taxis by tourists, but also noticed a rise in the sales of personal boats, the entrepreneur might think to develop an Uber-like solution for water taxis that allows anyone with a boat to give people rides around the city easily using a mobile app. The addition of open data also stimulates competitiveness because it creates an environment where startups are able to use existing data infrastructure to create their product, rather than needing to spend money doing their own collection and research. The information availability means these startups can compete with bigger existing companies, as opposed to an environment where the amount of private data and intellectual property a large company has can make it prohibitively difficult to compete as a new player in the market. One technical benefit is that optimizations can be made to administrative and research processes because open data can be reused, resulting in time and money saved. For example, the EU estimates they can save billions every year through open data related optimizations (Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk, 2012; "Benefits of Open Data").

Barriers to Open Data

With these benefits also come some barriers that hinder the open data movement's success. The presentation and organization of the data can undermine its use, because

although it is easy to publish large amounts of data, drawing conclusions and making connections can be difficult for the average person due to the data's complexity. Information quality is also critical to ensure proper extrapolation and to maintain trust by the parties using the data. Values can be incorrect or missing altogether, and certain checks for validation might be necessary before the data can be used effectively.

The technical aspects of how data is stored can also present some barriers, especially when the size of the collection is large. Some of the technical barriers typically encountered are lack of a good supporting infrastructure, fragmentation of data, and legacy data formats. An organization's data can become fragment with the introduction of newer database technologies because they may not have the resources to go through their existing data sets and update them to the newest standard, but they start to use the new standard for any newly data added nonetheless. Over time, this will leave a database in a fragmented state, which will cause the data to be difficult to access, and can create bugs in websites that provide access to it (Janssen et al., 2012).

Arup: Shaping a Better World with Open Data

Among the numerous companies working to advance the Open Data Movement's objectives, a good example is a UK-based firm called Arup. Arup is an independent company consisting of designers, planners, engineers, and technical experts, with the mission of using open data to help solve the world's problems. One of the problems they solve are mobility issues in cities: "With the continually increasing number of people living in cities, city authorities are facing significant challenges with regards to ensuring future levels of mobility... cities can use [the] data to transform city operations and urban mobility to enhance economic, social and environmental well being" ("Arup"). In Sydney, for example, Arup used Geographic Information Systems (GIS) in combination with over 100,000 data points of underground utilities in order to map a collision-free path for the new Light Rail transportation system. By making use of public utility data, they were able to automate an otherwise time consuming process of efficient and collision-free path mapping. This same concept can be applied to the VPC data, which will hopefully assist independent organizations and the government in addressing infrastructural issues around Venice.

Appendix C: Popular Science Publications

The Venice Project Center seeks to make its research available to the broadest possible audience by creating booklets in the genre of popular science. Popular science publications are designed to make scientific information and research accessible to a general audience (Broks, 2005). Well known popular science magazines include *American Scientist*, *Discover*, *Science Illustrated*, and *Popular Science* (seen in Figure 32). Proper scientific journals, books and other publications appeal to a very specific audience often composed of researchers, scientists and academia. This genre can promote a better understanding of scientific topics among the general public, so it is a good choice for the VPC's goals (Broks, 2005).



Figure 32. Publications the VPC has been featured in

The Venice Project has collected and made available extensive information on Venice and the surrounding area. Our project continued to help visualize and disseminate information in new forms. We continued the trend of previous projects to organize VPC data from Venipedia, the 2.5 Website, project reports, and firebases. We then utilized this data to assist in creating high-impact infographics for booklet publications and their companion website.

Appendix D: Data Visualization

Large data sets are often difficult to understand and therefore useless without some form of processing. One of the easiest and most effective ways of making sense of a data set is through visualizations (Walker, 2015). In this section, we outline the benefits associated with creating visuals for data, as well as some best practices in visualisation techniques.

Importance of Data Visualization and Design

Visualizations allow for a better understanding of large amounts of data while ensuring it can be easily and quickly understood. The human brain is wired to quickly understand visual stimuli, with far less cognitive effort required than abstract data. People are able to almost immediately identify and process visual patterns allowing easy processing and understanding of data (Fry 33). A study done on the comprehension of medicine labels showed that 25% more people understood labels with both text and visuals than labels with text alone (Dowse, R., & Ehlers, M. 2005).

Visualizations also allow for the compression of the data and therefore the communication of multiple dimensions at once (Walker, 2015). Humans naturally process visuals better than they do numbers alone because it happens in the preconscious, similar to breathing. This frees up the conscious for the critical thinking and analysis needed to draw meaningful conclusions, increasing understanding and promoting deeper learning (Lima et al., 2; Rohrer, 2000).

Types of Static Infographics

Edward R. Tufte's book, *The Visual Display of Quantitative Information* is often considered the most authoritative text on the subject of infographics and data visualization. This book describes in detail the most common and important types of infographics, and how they use data to effectively tell a story.

The most frequently used form of graphic design is a time-series plot, in which data is plotted against a horizontal time scale (Tufte 28). A time-series chart is used to display complex statistical data in which the time dimension on the chart provides relevant context to increase the understanding of the data presented (Tufte 30). Tufte uses the image below in Figure 33 as an example of a successful use of a time-series chart that tells a story by organizing a large number of data into a single image.

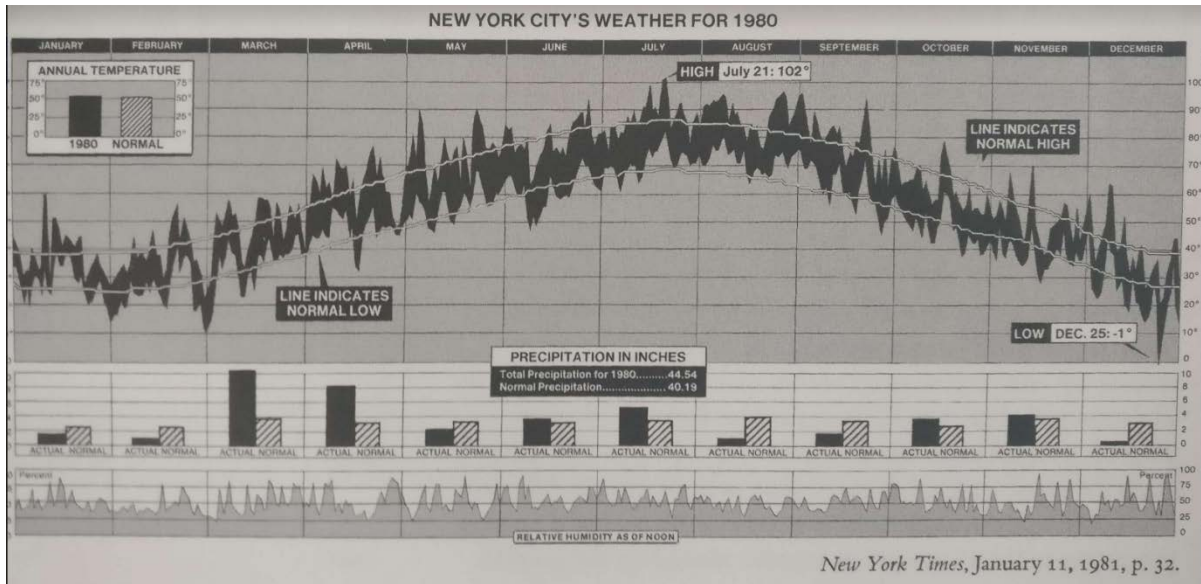


Figure 33 Example of a time-series chart

Tufte emphasizes how the simple passage of time alone is not a good explanatory variable for the data being presented, and that instead, the time axis should drive the data by providing context and meaning for any noticeable patterns in the data (Tufte 37).

The second most common infographic is a relational graph, making up about 40% of published graphics in scientific literature (Tufte 47). These charts link two or more variables together, encouraging the viewer to think about the causal link between the plotted variables (Tufte 47). These charts are often constructed similarly to time-series charts, except that one dependent variable X can be compared to any independent Y variable, not just time. These charts are often very simple and to the point, and provide a way for an author to clearly display a link between X and Y. An example of an effective relational graph from a 2001 Venice Project Center-sponsored IQP is shown in Figure 34.

Non-Deviant Factories: Temp vs kWh

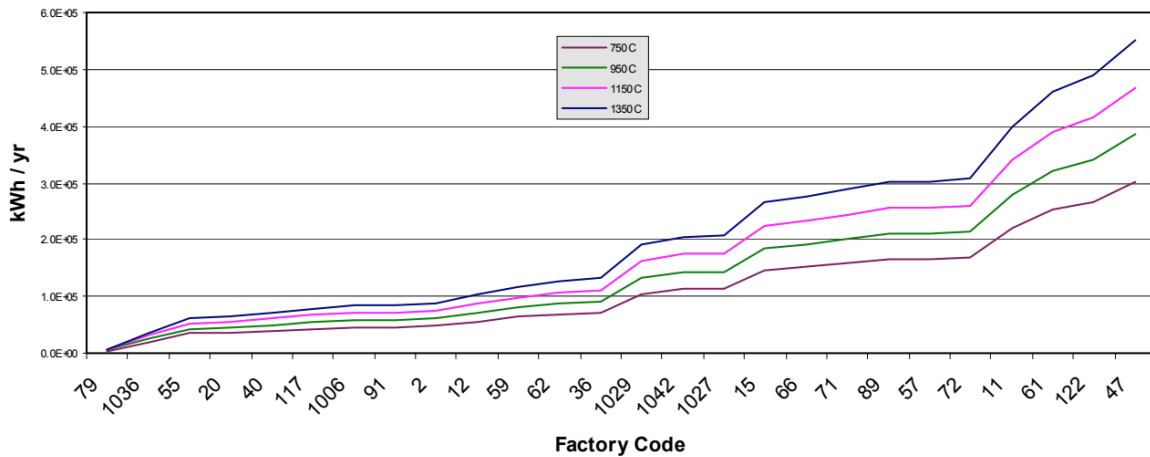


Figure 34 Relational graph

Chartjunk: Striking a balance

Chartjunk is a phrase coined by Edward Tufte to describe visual components of infographics that do not enhance the understanding of the information presented, but rather create an extreme distraction for the viewer (Tufte 107). Chartjunk is created by the excessive and unnecessary use of artistic style in a graphic that often adds visual hindering of the viewer's comprehension of the image. Two examples of real published chartjunk cited by Tufte are displayed in the figures below.

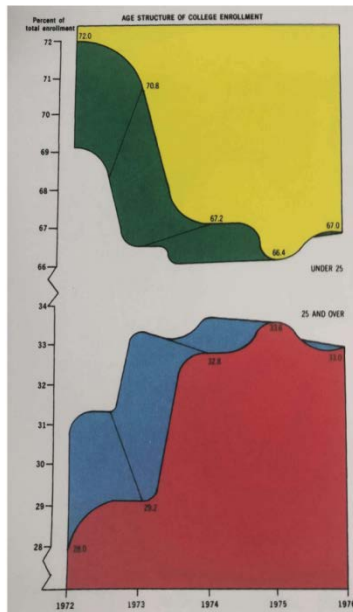


Figure 35 Example of chartjunk

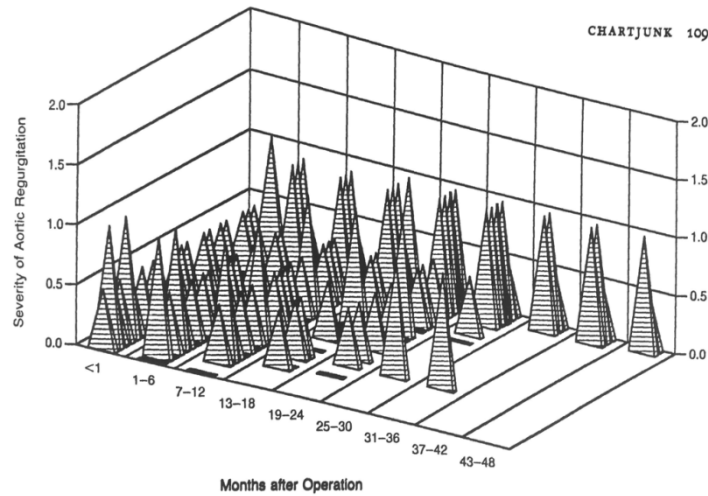


Figure 2. Serial Echocardiographic Assessments of the Severity of Regurgitation in the Pulmonary Autograft in 31 Patients. The numerical grades were assigned according to the severity of regurgitation, as follows: 0, none; 0.5, trivial; 1.0 to 1.5, mild; 2.0, moderate; and 3.0, severe.

Figure 36 Another example of chartjunk

Figure 35 uses four colors and a confusing, unnecessary 3D perspective to present only ten total data points, and Figure 36 uses cross hatching and a 3D perspective. It is clear that the information contained in these charts could be easily understood by the viewer if it were

presented clearly in basic time-series and relational graphs, respectively. The use of effective infographics will be able to display the data that would otherwise be too difficult to explain well and comprehensively.

Choosing the Right Visualization

In order to determine the best way to visualize data, one must consider four factors: the reason, the audience, the data, and the look. The reason for the visualization is important to consider because in order to create a visual that effectively communicates substantial information about the data, thought must be put into the purpose of the visualization (Kirk & Safari Books Online, 2016).

If the goal is to explain the data, then the audience is very important to consider. What is understandable to PhD students may not make sense to the general public. For election results, a map such as Figure 37 (European parliament election) is great for the general public, but for an analyst there isn't enough detail. A bar chart of electoral votes by state and party could be a more useful representation. Another thing to consider is the specific message one wants to convey. A clear message is one of the defining traits that makes a visual explanatory.

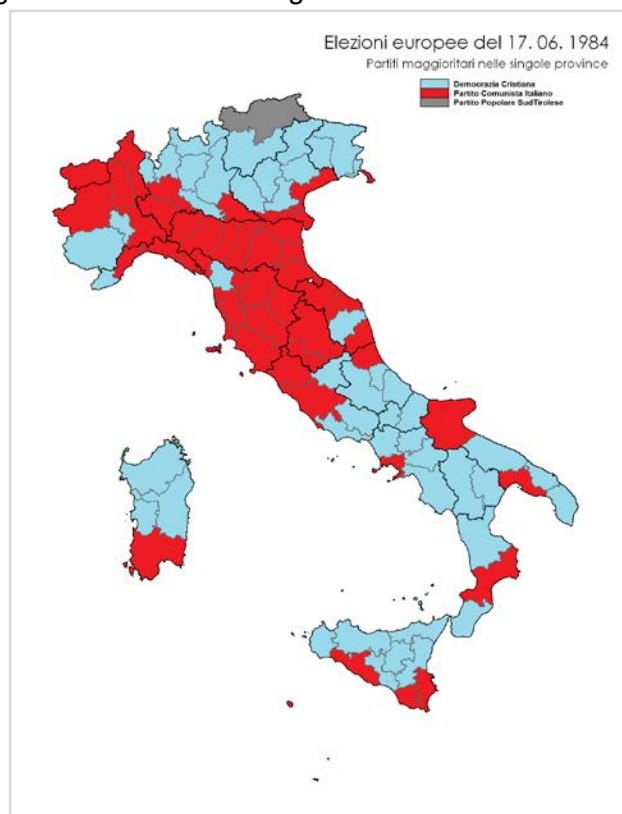


Figure 37 Election results for 1984 European Parliament election

Exploration of data is an iterative process and has different requirements from explanation. While the audience must be taken into consideration for exploration, it is common that those exploring data will be experts in the field. With exploration, the person visualizing the data may not know exactly what they are looking for or trying to convey. Because of this, visuals for exploration must show the data in multiple different configurations. Interactive visuals really

lend themselves to data exploration due to the fact that they can be dynamically configured and display the data in a multitude of different ways (Kirk & Safari Books Online, 2016). While interactive visuals are ideal, exploration through static visuals is also possible. One example is a scatter plot matrix such as Figure 38 (“Introduction to graphs in stata | stata learning modules”). Graph matrixes allow for multiple configurations of data to be displayed and compared at once.

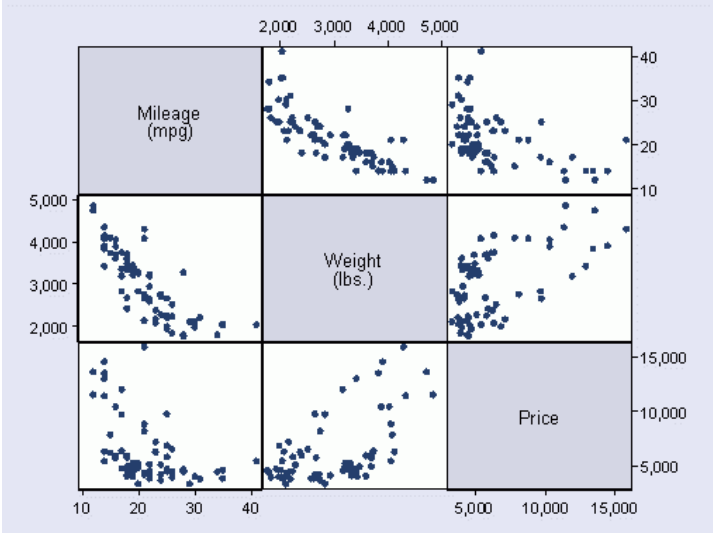


Figure 38 Scatter plot matrix of mileage, weight, and price of different car models

The last thing to consider when creating a visualization is the look and feel of the visualization. The main parts that make up this aspect are color, font, spacing and imagery, which will vary based on audience and the reason for the visualization. For a preschooler visuals should be fun and colorful, but for a corporate office they should look professional and focus on the information being presented. A bar graph about water pollution will send a much stronger message if it is made using piles of dead fish than if it is made using plain blue bars. The visual look draws people in and makes them want to consume the data.

Appendix E: CSV Data Set Code

```
import csv
import os
import numpy as np
from tabulate import tabulate

output = []
for file in os.listdir():
    if file.split('.')[1] == 'csv':
        with open(file, encoding="utf8") as csvfile:
            reader = csv.reader(csvfile)
            count = 0
            items = 0
            linelen = []
            filllen = []
            maxline = 0
            try:
                for line in reader:
                    count += 1
                    fields = 0
                    tfields = 0
                    for field in line:
                        tfields += 1
                        if field != '' and field != 'Null':
                            fields += 1
                            items += 1
                    linelen.append(fields)
                    filllen.append(100*fields/tfields)
                    maxline = max(maxline, tfields)
            except:
                print('ERROR\n\n')
            finally:
                output.append([file, count, maxline, np.mean(linelen),
                    np.mean(filllen), items, os.path.getsize(file)/1000])

sumline = ['Sum',sum([x[1] for x in output]),'---','---','---',sum([x[5]
for x in output]),'{0:.3f}'.format(sum([x[6] for x in output]))]
avgline = ['Avg','{0:.3f}'.format(np.mean([x[1] for x in
output])), '{0:.3f}'.format(np.mean([x[2] for x in
output])), '{0:.3f}'.format(np.mean([x[3] for x in
output])), '{0:.3f}'.format(np.mean([x[4] for x in
output])), '{0:.3f}'.format(np.mean([x[5] for x in
output])), '{0:.3f}'.format(np.mean([x[6] for x in output]))]

for line in output:
    line[3] = '{0:.3f}'.format(line[3])
```

```
line[4] = '{0:.3f}'.format(line[4])

output.append(['\n', '\n', '\n', '\n', '\n', '\n'])
output.append(sumline)
output.append(avgline)
print(tabulate(output, headers=["Filename", "Lines", "Max Line Size", "Avg
Items/Line", "Avg Line Fill", "Items", "Filesize (kB)"],
floatfmt=".2f"), '\n')
```

Appendix F: CSV Data Set Statistics

Data Set Name	Lines	Max Line Size	Avg Items/Line	Avg Line Fill %	Items	Filesize (kB)
Belltowers	100	250	65.450	26.180	6545	200.90
Canals	231	21	16.550	78.808	3823	267.95
Churches	145	26	24.821	95.464	3599	162.73
City Islands	128	36	27.227	75.629	3485	57.83
Coats Of Arms	1099	74	59.103	79.869	64954	924.58
Convents	60	48	30.167	62.847	1810	294.13
Fountains	182	55	49.154	89.371	8946	496.38
Major Lagoon Islands	36	36	26.722	74.228	962	17.30
Minor Lagoon Islands	37	33	28.838	87.387	1067	22.35
Patere	495	74	60.495	81.750	29945	466.12
Ponti	473	126	108.397	86.030	51272	642.03
River	1628	73	59.659	81.725	97125	763.24
Stores	4968	61	21.942	35.971	109008	2027.06
Street Altars	110	74	59.818	80.835	6580	102.77
Sum	9692	---	---	---	389121	6445.36
Avg	692.28 6	70.500	45.596	74.007	27794. 357	460.38

Table 1: The Metadata about the Venetian CSV data sets that was generated by the code in Appendix E

Appendix G: Types of Infographics we Implemented

1. Donut Chart (percent of a count)
2. Filter Map
3. Scaled Up Number
4. Data View
5. Stacked Bar Chart
6. Choropleth Map
7. Bubble Chart Mpa
8. Category Map
9. Bubble Chart
10. Donut Chart Map
11. Route Map

Appendix H: Sandbox Application Feature Guide

The Sandbox Application used to create static and interactive visualizations can be accessed by visiting sandbox.veniceprojectcenter.org. Upon visiting the site, the user will be prompted to select a data set from the available data sets stored in Firebase. An image of the home screen is shown in Figure 39 below.

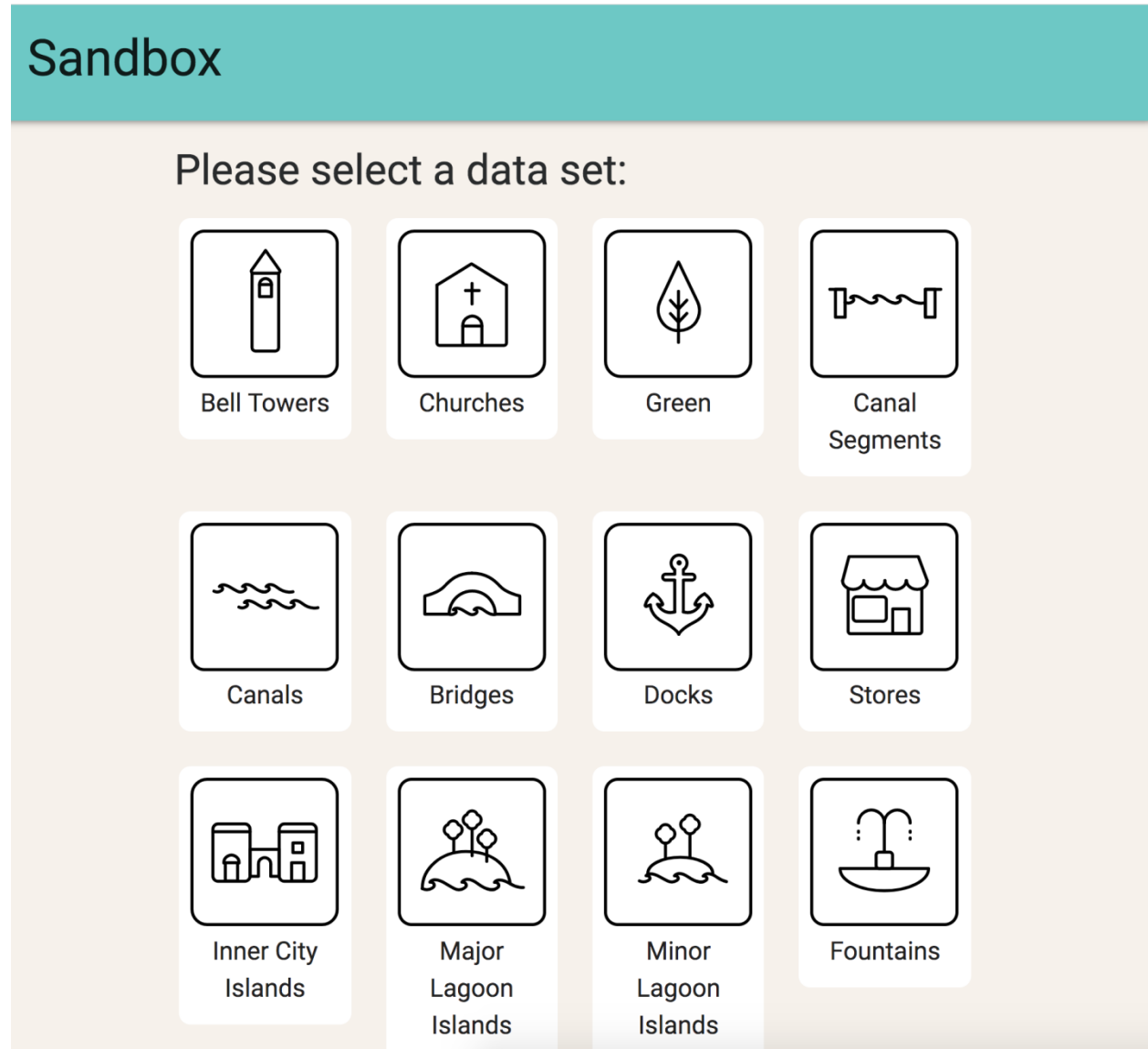


Figure 39 Sandbox Application display window

After selecting a data set to base a visualization on, the list of available visualizations is displayed. The user is prompted to select a visual type to create, as in Figure 40.

Please select a visual:

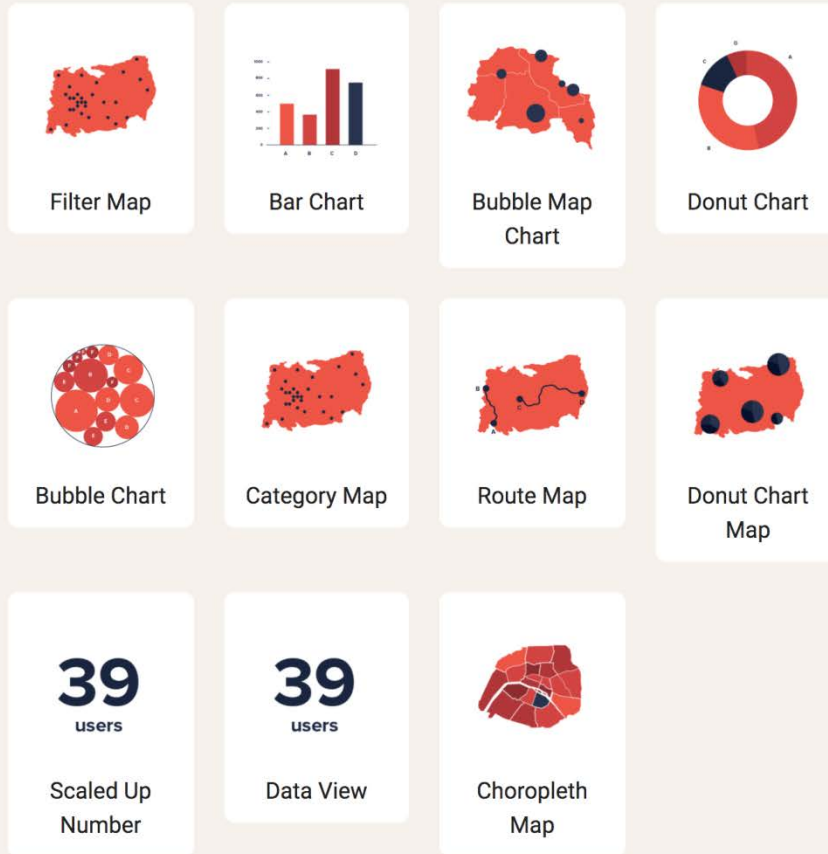


Figure 40 Visual selection

The different visualization options available in the Sandbox Application are described below.

Bar Chart

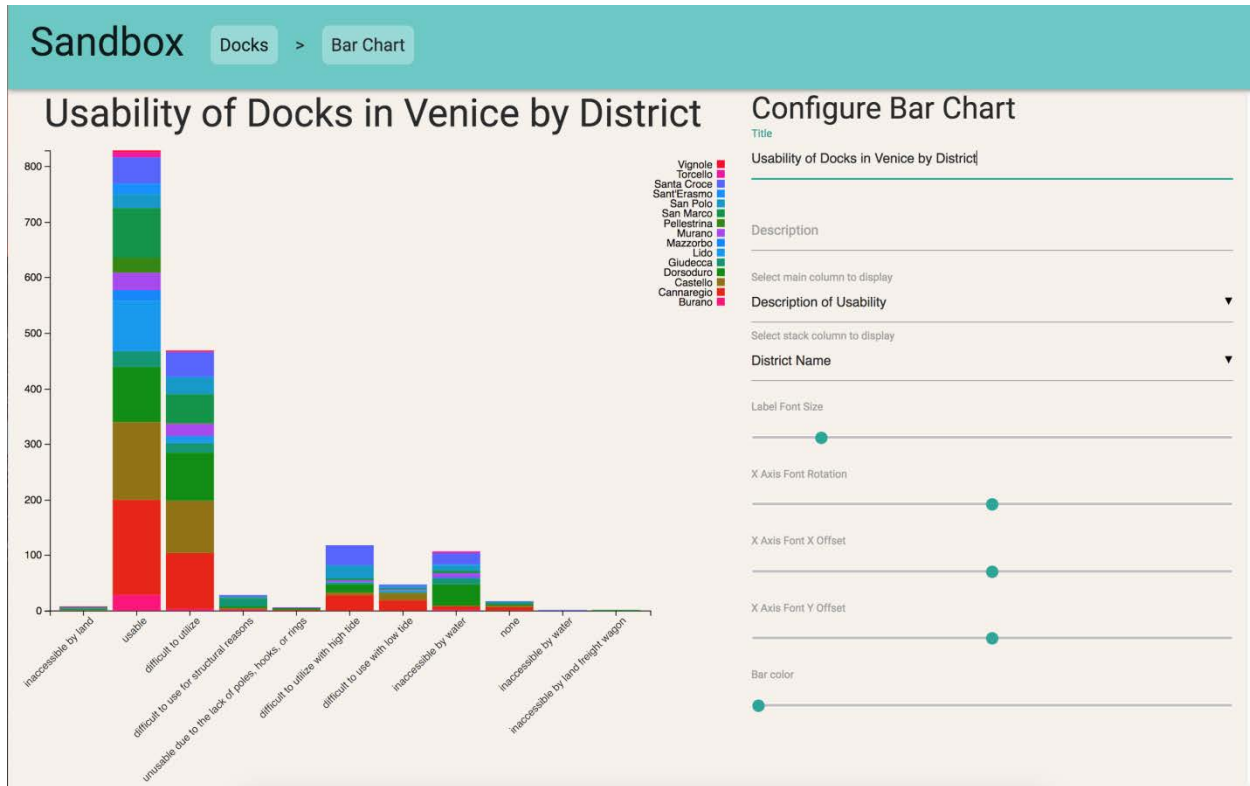


Figure 41 Bar chart

Selecting Bar Chart from the visual screen brings up the bar chart creation tool. An example of the bar chart creation tool is shown in Figure 41 above, with the Docks data set selected.

To create a bar chart:

1. From the home screen, select the desired data set and then select “Bar Chart” from the visual selection screen.
2. Enter a title and description for the chart, if desired.
3. Select what data attribute will be used to separate the data into columns. In the example above, “Description of ” was selected.
4. If creating a stacked bar chart, select the data attribute to break the columns into smaller sections. In the example above, “Step pavement material” was selected.
5. Use the sliders to adjust the label font size, label offset, and label rotation so that the labels are easily readable.
6. Drag the Bar Color slider to change the color of the bars in the chart. Note that this can only be done to regular bar charts, not stacked bar charts.
7. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Donut Chart

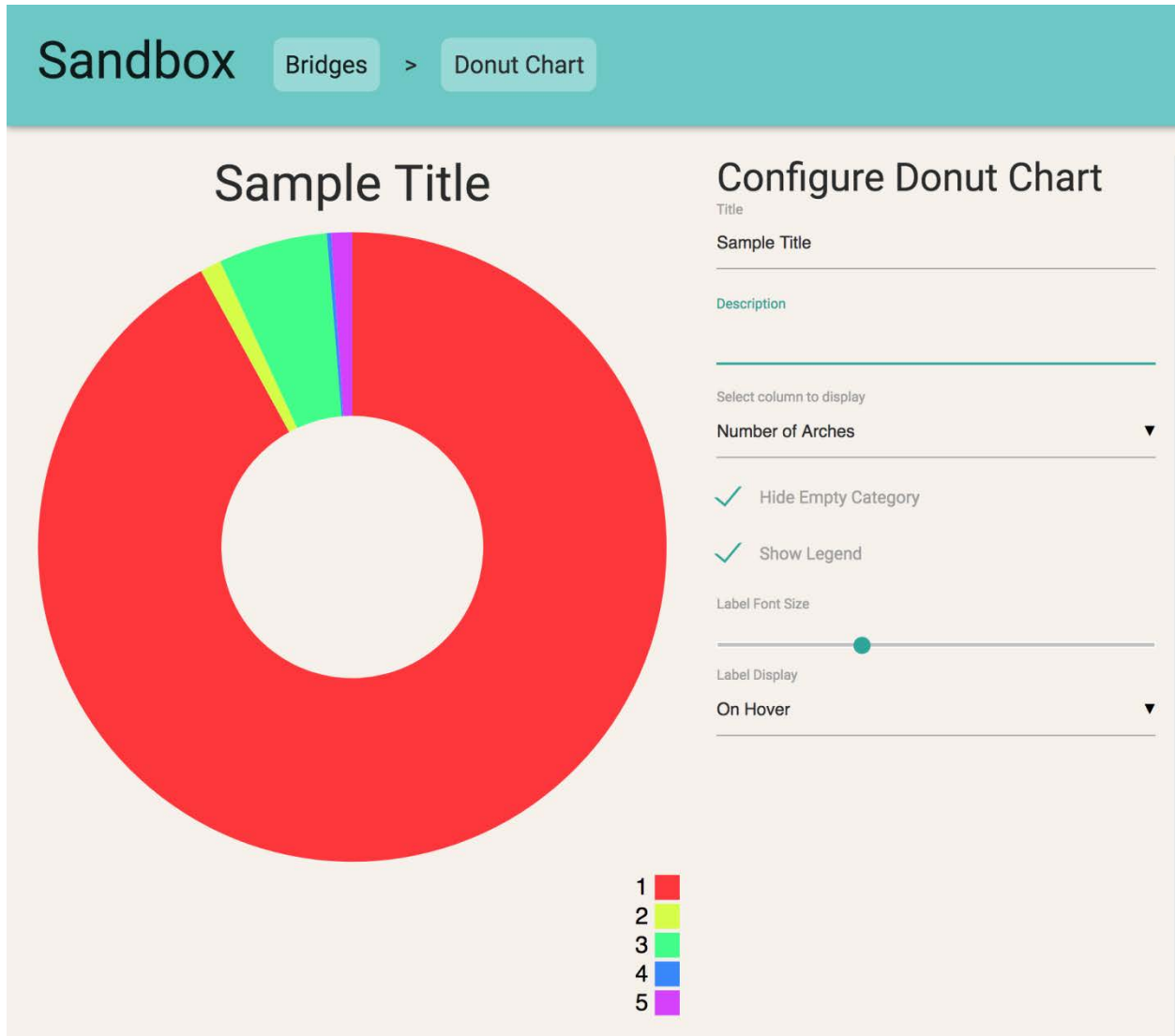


Figure 42 Donut Chart

Selecting Donut Chart from the visual screen brings up the bar chart creation tool pictured in Figure 42 above.

To create a donut chart:

1. From the home screen, select the desired data set and then select “Donut Chart” from the visual selection screen.
2. Enter a title and description for the chart, if desired.
3. Select what attribute of the current data set to display on the donut chart. In the above example, “Height 1m North (m)” was selected.
4. If the data displayed is categorical in nature, the elements in the data set will be grouped into bins of a specified range, as shown in the example above. Enter a start value for the first group, in this example “0.0”, as well as the desired size of the categorical group, “.5”.

5. Check the “show legend” box if you would like the legend to be displayed.
6. Check the “hide empty category” box to prevent zero and null values from being displayed.
7. Select a Label Display option from the drop down menu. With the selected settings shown above, the labels will appear only when a chart segment is hovered over.
8. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Bubble Chart

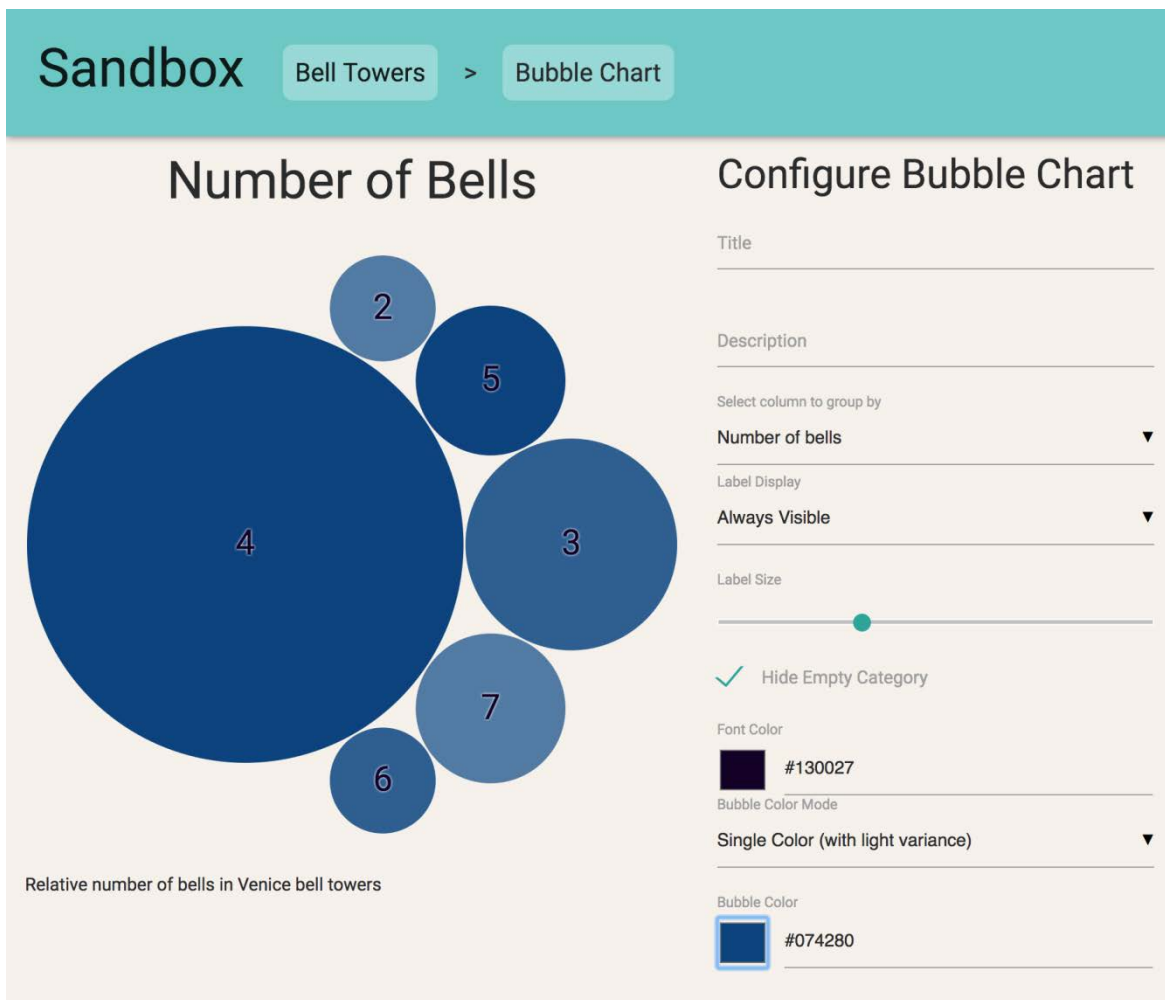


Figure 43 Bubble Chart

Selecting Bubble Chart from the visual screen brings up the bar chart creation tool pictured in Figure 43 above.

To create a donut chart:

1. From the home screen, select the desired data set and then select “Bubble Chart” from the visual selection screen.
2. Enter a title and description for the chart, if desired.

3. Select which data attribute should be used to create the chart. In the above example, “District Name” was selected.
4. Select a Label Display option. “Always Visible” is selected in the example shown above.
5. Use the slider to adjust the font size of the labels, and select a font color for the label
6. Select a color mode for the bubbles in the chart. Available options are Single Color, Single Color with Light Variance, and Manual Color.
7. Click the color picker labelled “bubble color” to choose a color for the chart. In the Manual Color mode, each bubble is manually colored by clicking on a bubble, then utilizing the color picker to fill the selected bubble.
8. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Filter Map

The Filter Map application is the one of the most flexible visualization tools in the Sandbox Application. It allows for any of the VPC’s data set entries containing geographic information to be selectively superimposed on a map of Venice. A example filter map, showing the location of souvenir stores in Venice, is displayed in Figure 44 below.

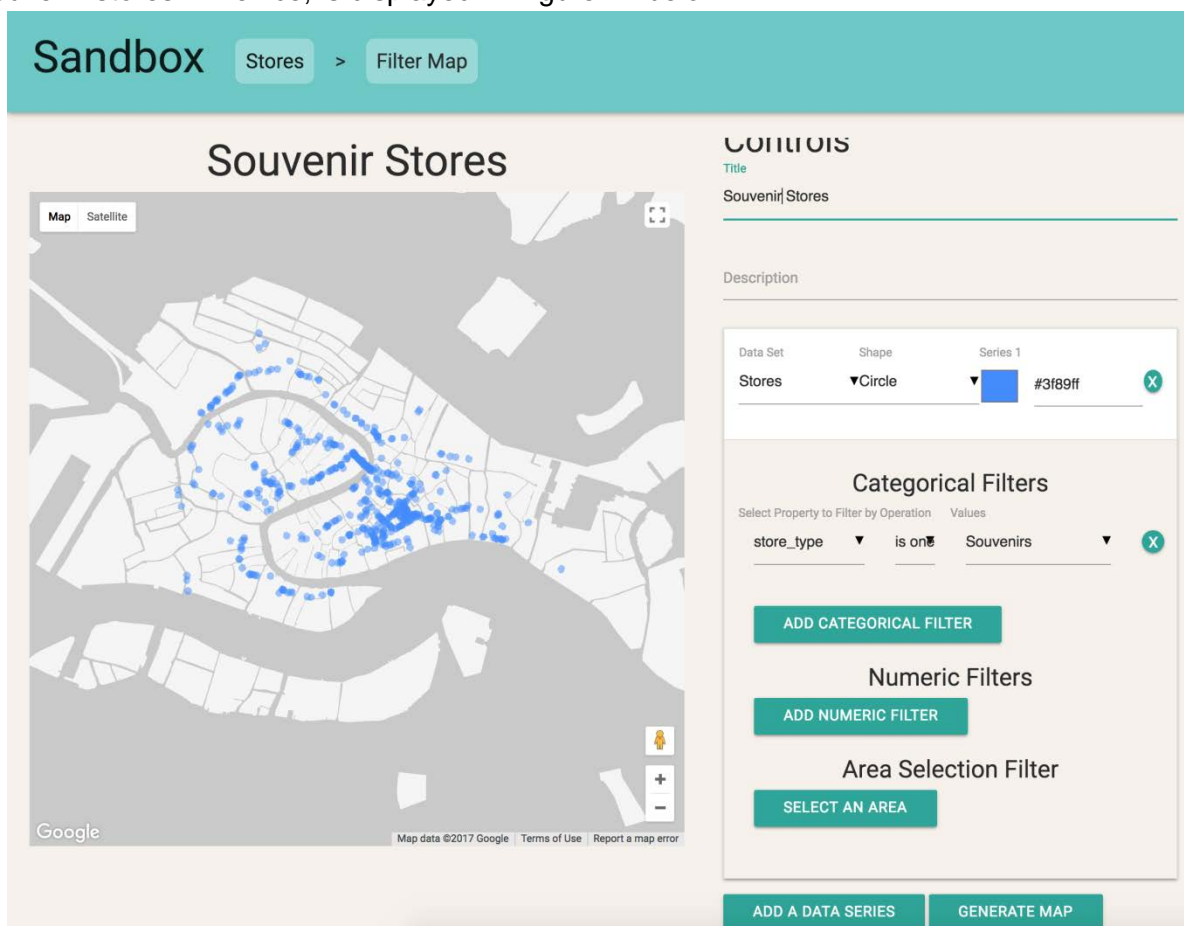


Figure 44 Example Filter Map

The filter map application also allows for multiple data columns and data sets to be superimposed on the same map. An example of a filter map, visualizing data from the Churches data set as well as the Bell Towers data set, is displayed in Figure 45.

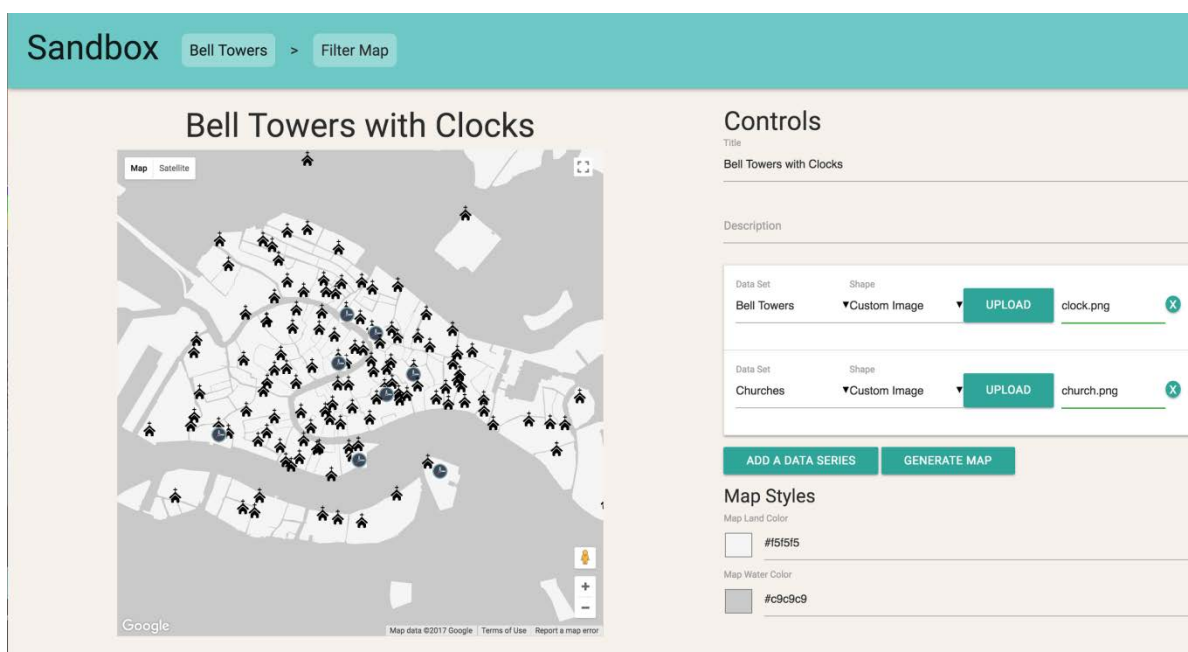


Figure 45 Example Filter Map with multiple data sets

To create a filter map:

1. From the home screen, select the desired data set and then select “Filter Map” from the visual selection screen.
2. Enter a title and description for the chart, if desired.
3. Select a data set for the filter to pull data from. For the first set of filters, the data set selected on the Home screen is the default selection here. When additional filter sets are desired, such as when multiple sets of data are to be displayed on the same map, the proper data set will need to be chosen.
4. Choose a shape and color to display the data points for the selected data, or select custom image to prompt the upload of any image.
5. Click within the white box of the data set to open the options for filters.
6. For a categorical filter, select a property to filter by, operation, and value by which to filter any of the categorical properties.
7. For a numeric filter, select a property to filter by, operation, and type in a value to by which filter any of the numerical properties.
8. To add an additional filter, click the button to add either another categorical or numeric filter.
9. To add an area selection filter, click select an area and continue to click the outline of the desired area on the map. Make sure to make the final click at the same point of the first, which may require zooming into the map on the first point. The map will only show data points residing inside this polygon.

10. To display an additional category of data from the map, click “Add a data series” and repeat steps 2 through 9.
11. To style the map, you can change the colors of the land and the water on the map by selecting the colors under “Map Styles” which will prompt a window to choose the desired color.
12. To view the final map, click “Generate Map”
13. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Category Map

The Category Map tool provides a quick way to visualize all the entries in a data set, colored by category. The Filter Map should be used where more control and customization is required, or where more than one data set needs to be visualized on the same map. The example below shows the location of every bridge, with each color representing a different number of arches. An example category map showing number of arches is shown in Figure 46.

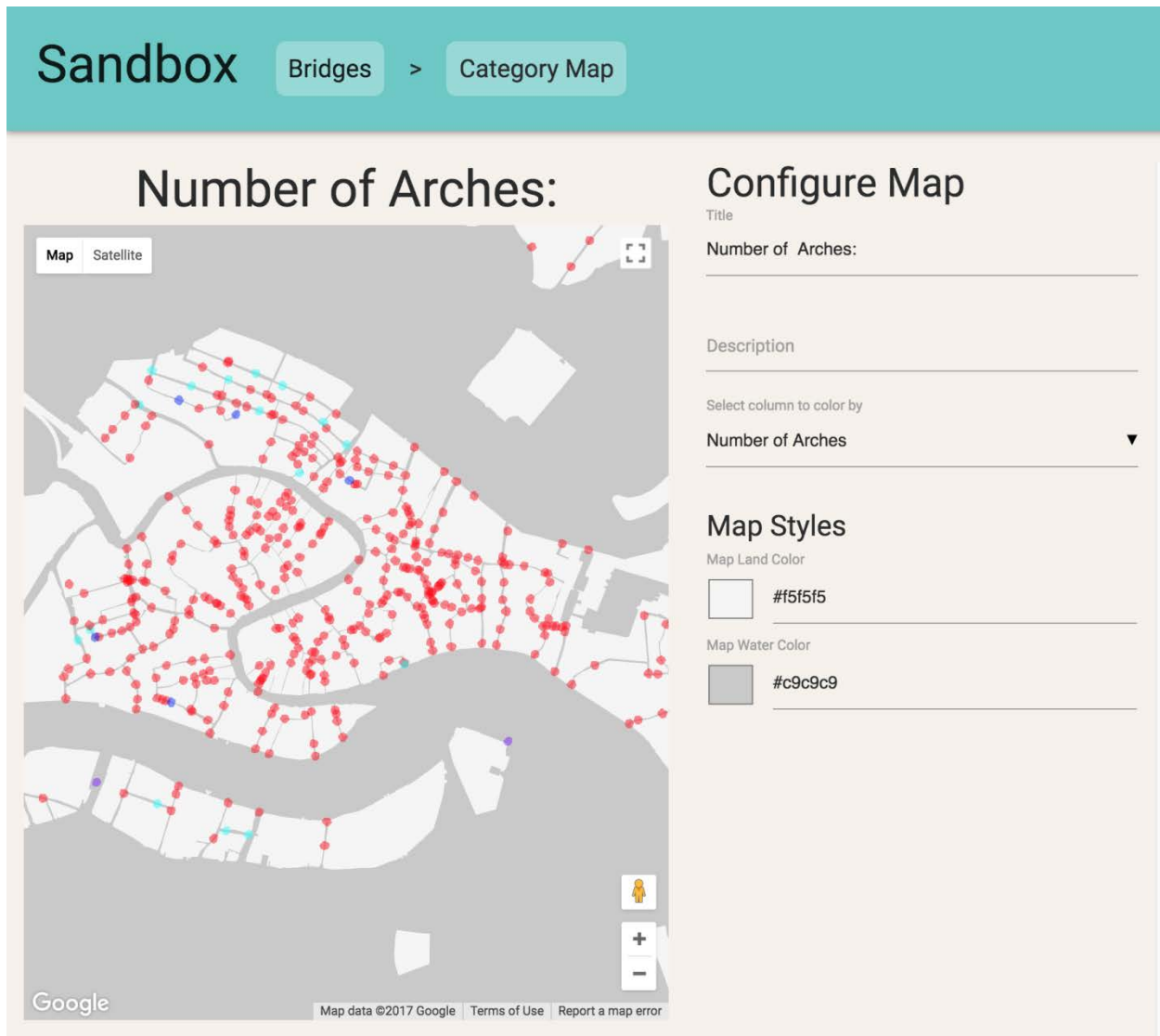


Figure 46: Category Map tool

To create a category map:

1. From the home screen, select the desired data set and then select “Category Map” from the visual selection screen.
2. Enter a title and description for the chart, if desired.
3. Select the column by which to group the data sets into categories
4. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Route Map

The Route Map tool is an interactive application that allows for the user to select two points on a map, corresponding to the origin and destination of a walking trip within the historic center of Venice. At present, this map only works with the Bridges data set. After two points are clicked

on the map, a route will appear, which matches the fastest route that the Google Maps smartphone application would suggest. Superimposed on the route will be the locations of all the bridges that would be crossed. The number of encountered bridges is displayed. The Route Map application also sums up all entries in a selected column of the Bridges data set. This feature is most useful for summing up the Total Number of Steps column entries for the bridges encountered on the route, so a user can know how many stair steps they will climb during that journey. Selecting Route Map from the visual screen brings up the route map creation tool pictured in Figure 47 below. Route maps are currently only available when the Bridges data set is selected.

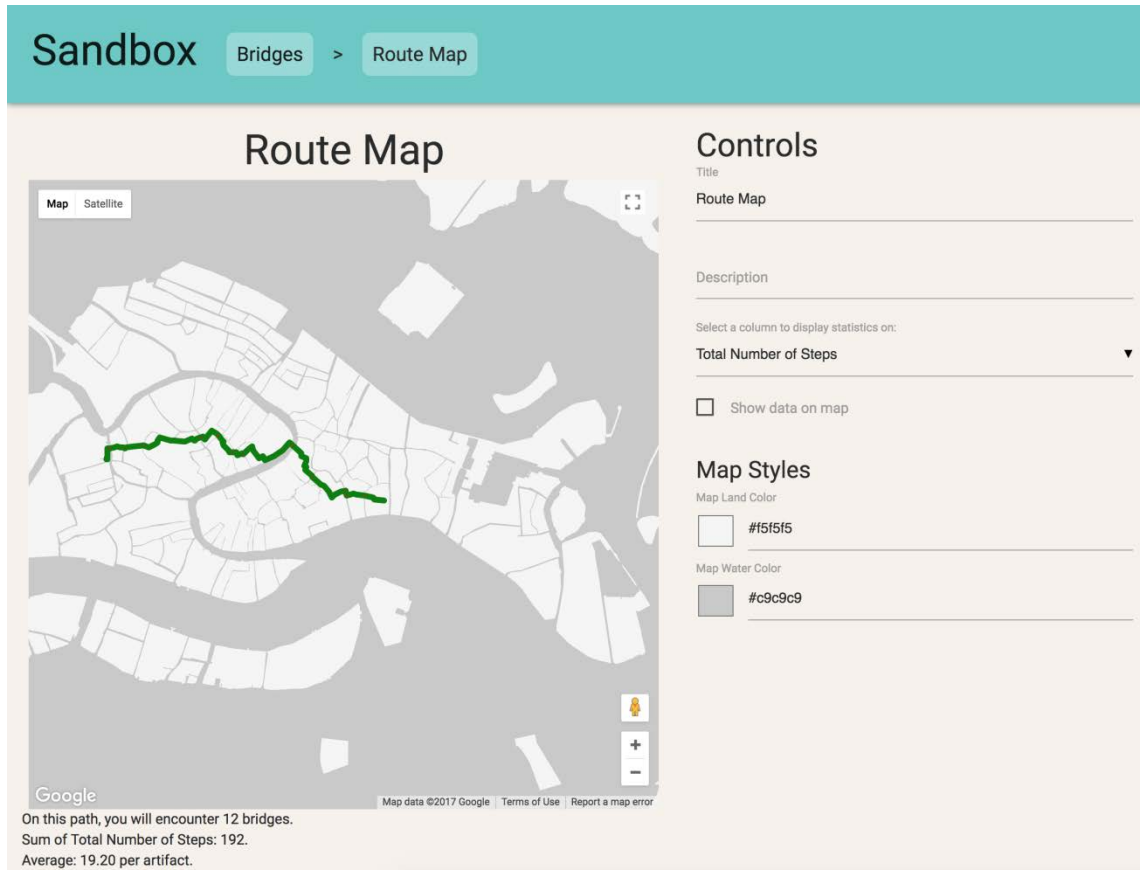


Figure 47 Route Map

To create a route map:

1. Select what attribute of the current data set to display under the route map. In the example above, "Total Number of Steps" was selected.
2. Select a start and finish point, from which a path will be created highlighting each of the objects within the data set that fall near the path. The number of objects will be displayed underneath the map as well as the sum and average of the chosen attribute. This attribute can be changed by the drop down in the controls.
3. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Scaled Up Number

The Scaled Up Number tool simply generates a large number displaying the sum, average, or count of entries in a data set column. This allows for simple calculations on numerical information stored in data sets to be performed. A large number visual is generated, displaying this information in a very clear form. This large number can also be exported as an SVG or embedded in a site, if desired. The output of the Scaled Up Number tool is shown below.

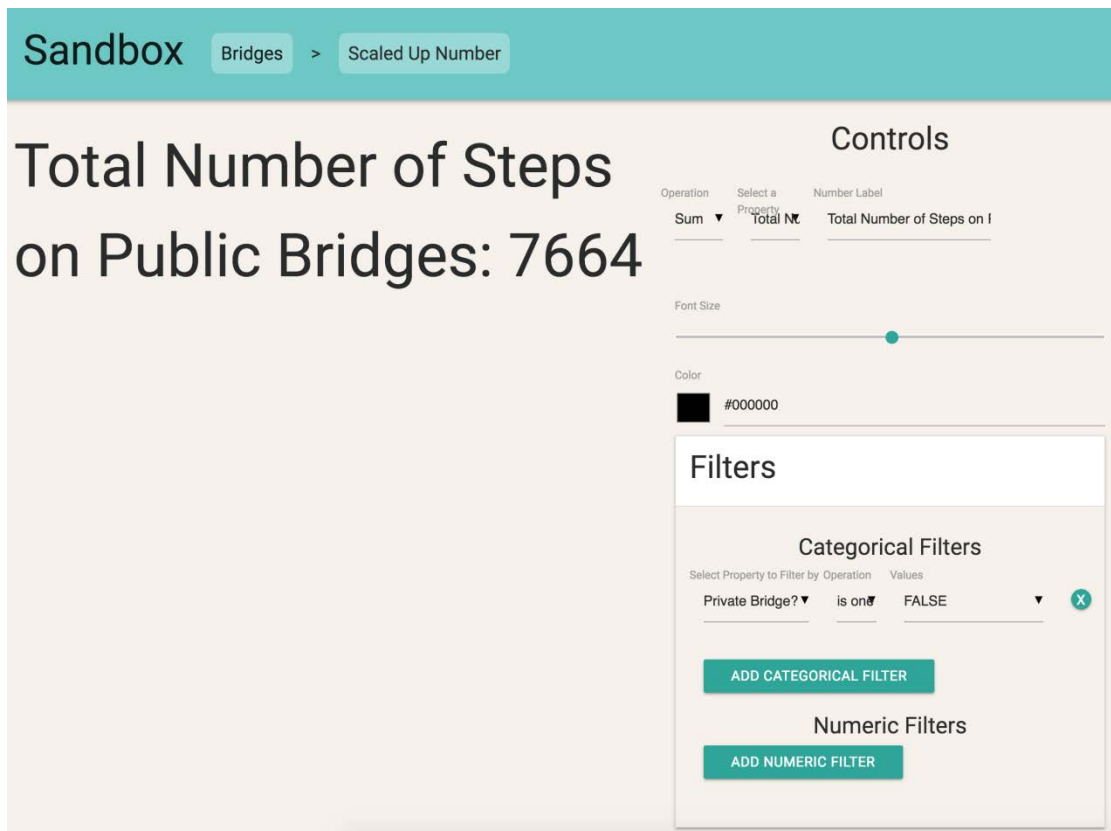


Figure 48 Scaled Up Number tool

To create a scaled up number:

1. From the home screen, select the desired data set and then select “Scaled Up Number” from the visual selection screen.
2. Select the property (data set column) desired.
3. Choose an operation. Sum sums the column selected above. Average returns the mean of the column selected above. Count returns the number of entries in the data set that have data in the selected column.
4. Enter a label for the number, describing what is being displayed.
5. Click the “Filter” button to filter what data entries are factored into the operation selected above. In the example above, a categorical filter was applied to exclude private bridges from the sum.

- Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Choropleth Map

A choropleth map is a great way to visualize how a measurement or statistic varies with geography. A map, in Venice in our case, is divided up into a large number of small geographic cells defined by predefined boundaries, and each cell is colored to reflect some measurement or statistic pertaining to the geographic area inside the boundaries. In Venice, dividing the map up by island was the logical choice for defining the boundaries of the cells. The Choropleth Map tool works well with the data sets related to the islands in the Venice lagoon, such as the Inner City Islands data set. Data from government censuses is particularly well suited for visualization in a choropleth map. An example, showing population density by island, is displayed in Figure 49 below.

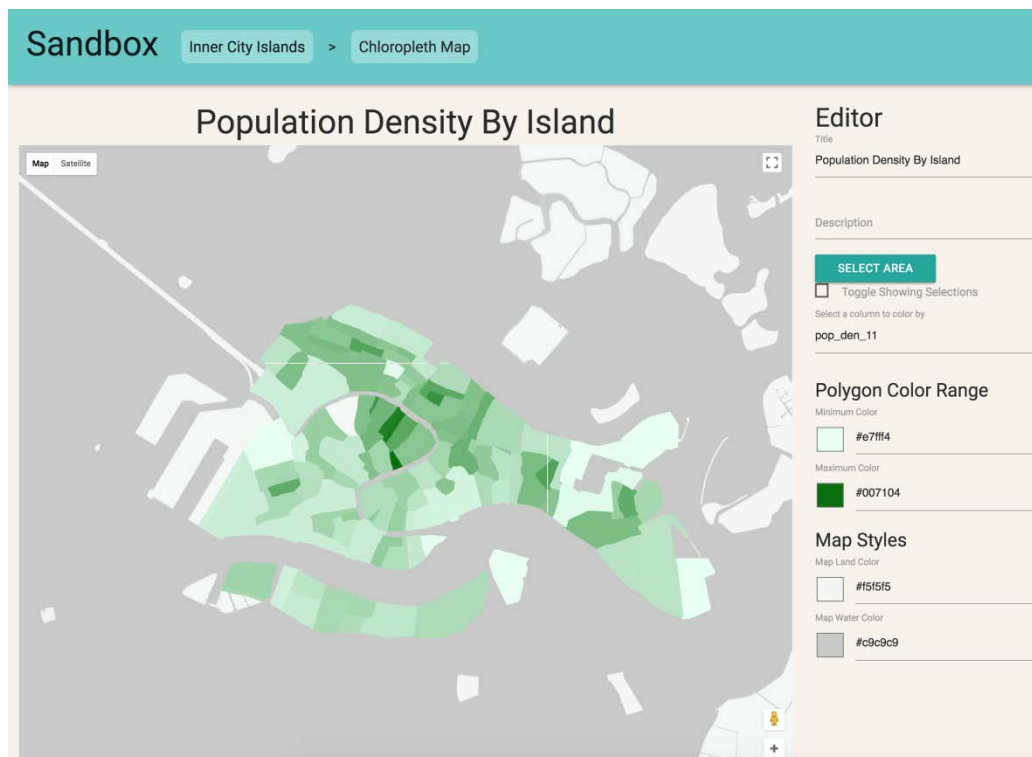


Figure 49 Choropleth map

To create a choropleth map,

- From the home screen, select the desired data set and then select “Bubble Chart” from the visual selection screen.
- Enter a title and description for the chart, if desired.
- Select which data attribute should be used to create the chart. In the above example, “pop_den_11” was selected, which corresponds to the population density of each island as recorded in the 2011 census.

4. Select the Minimum Color and Maximum Color. These colors should be selected so that the distinction between a high value and a low value is clear.
5. Click “generate map.”
6. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Bubble Map Chart

A Bubble Map Chart is capable of displaying multidimensional data on a map. It is similar to a Filter Map in that the location of each entry in a data set is displayed on a map, however, up to two additional data fields can be displayed in addition to the location fields by changing the size and color of each dot on the map. Our tool is capable of doing this automatically for the selected data fields. An example of a bubble map chart, showing the location, cost and occupancy rate of the AirBnB locations in Venice, is shown in Figure 50 below.

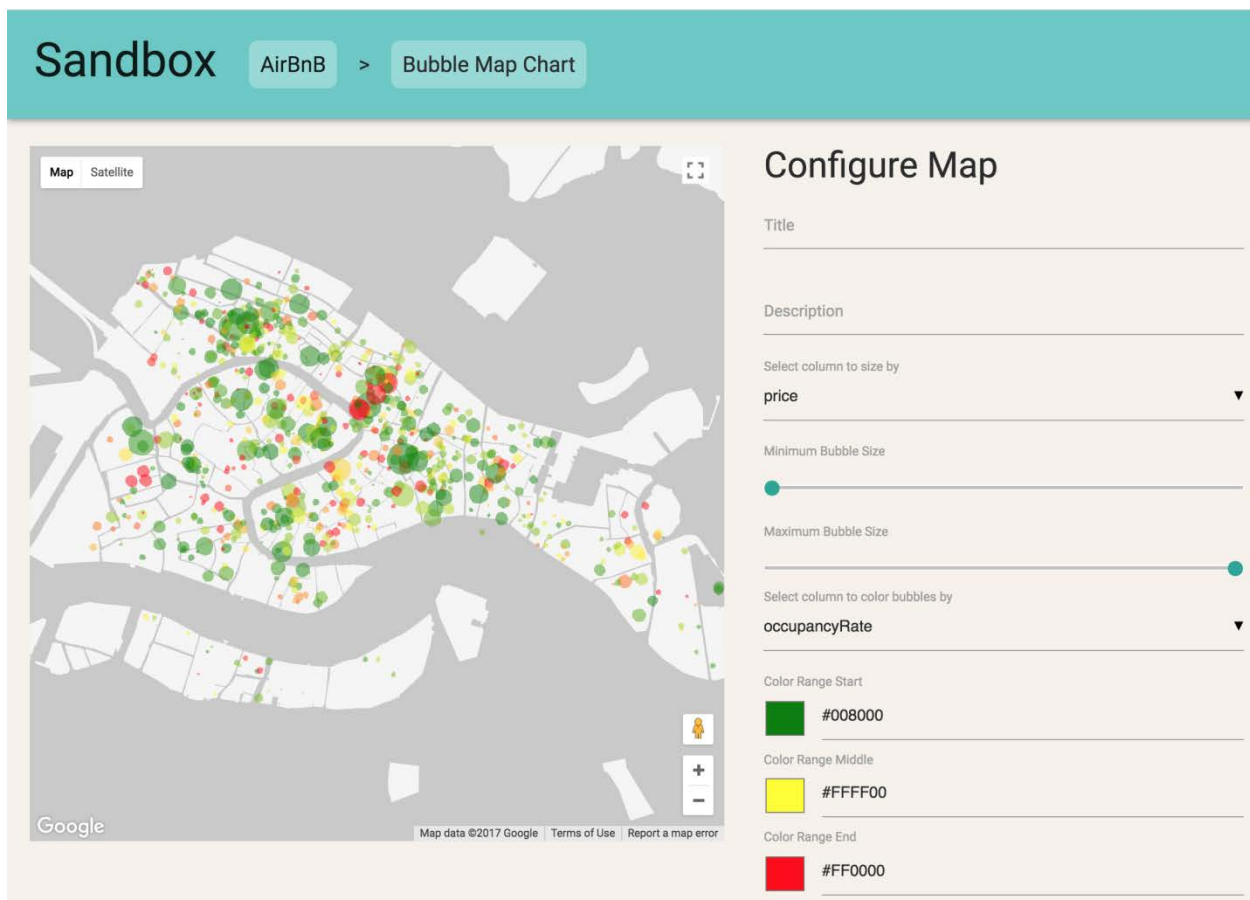


Figure 50 Bubble chart map

To create a bubble map chart:

1. From the home screen, select the desired data set and then select “Donut Chart” from the visual selection screen.
2. Enter a title and description for the chart, if desired.
3. Select a column by which to size the circles. In the example above, this is the price of the AirBnB listing.
4. Select a column by which to color the circles by. In the example above, this is the occupancy rate.
5. Select a range for the color of circles.
6. Once the visualization is displayed as desired, it can be exported as a SVG or published to Firebase. When the visualization is published to Firebase, it automatically appears on the VPC Data Portal website.

Donut Chart Map

A donut chart map allows for entries a data set to be grouped by a category, where data from a selected column is placed in a donut chart, and all of the donut charts are placed on a map at the average lat/long location of the entries in the data set. An example showing the ratio of handicap accessible islands to nonaccessible ones is shown in Figure 51 below.

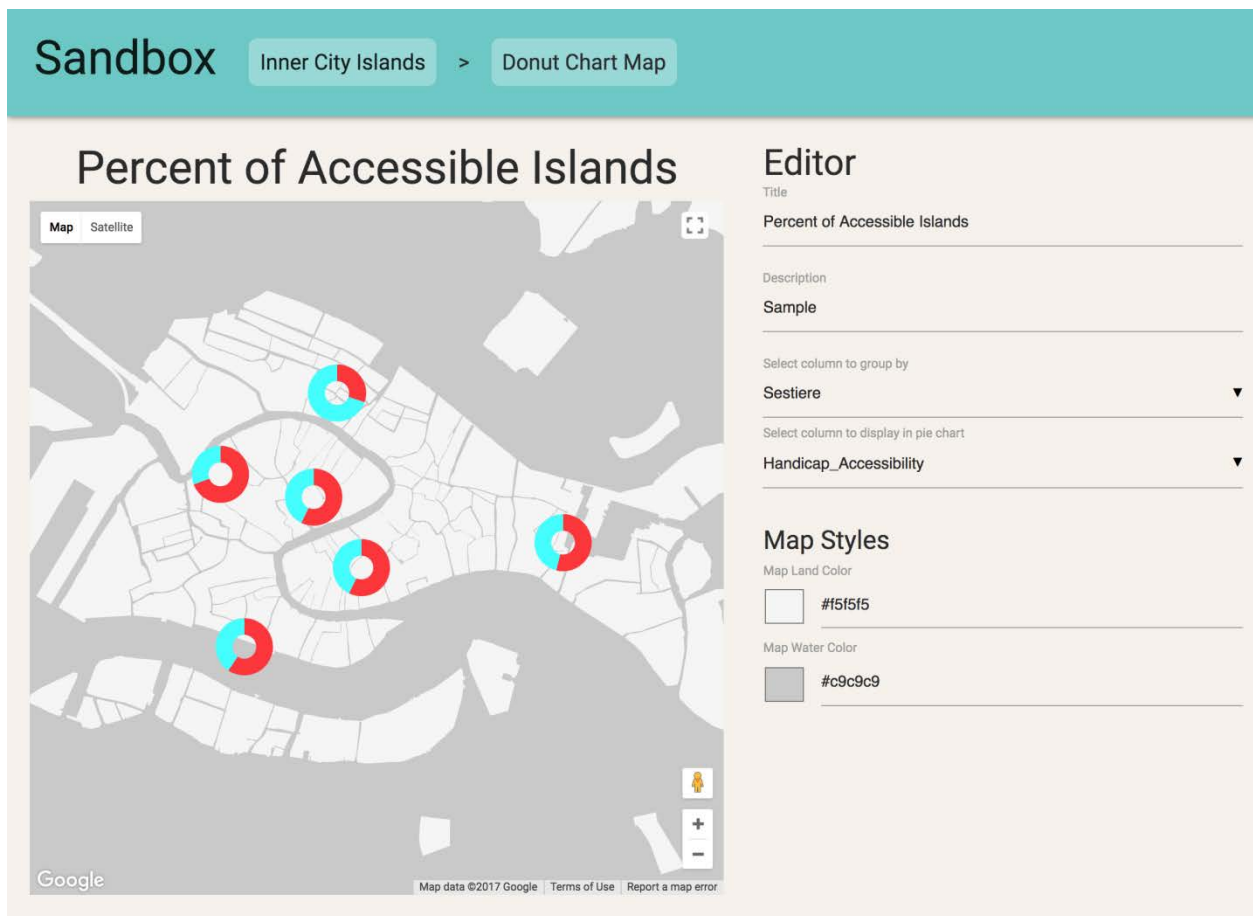


Figure 51 Donut chart map

To create a donut chart map:

1. From the home screen, select the desired data set and then select “Donut Chart” from the visual selection screen.
2. Enter a title and description for the chart, if desired.
3. Select the column by which to group the data set entries into donut charts. This should be geographic information, such as island name or sestiere.
4. Select the data set column to display in the pie charts.
5. Once the map is displayed as desired, it can be exported as a SVG, published to Firebase, or exported as a config file by using one of the buttons on the bottom of the page.

Data View tool

The Data View tool does not create a visual. Instead, it can be used to inspect the contents of a data set so the user can get an idea of what the actual entries in the data set look like. Another useful application for this tool is finding all entries in a data set that match certain criteria, defined using the filters. In the example below, the Data View tool is used to print the names of all bridges with zero steps. This tool can also be used to troubleshoot when a visual is not appearing as expected for a certain data set. Figure 52 shows how the Data View tool can be used to find any bridges with no steps.

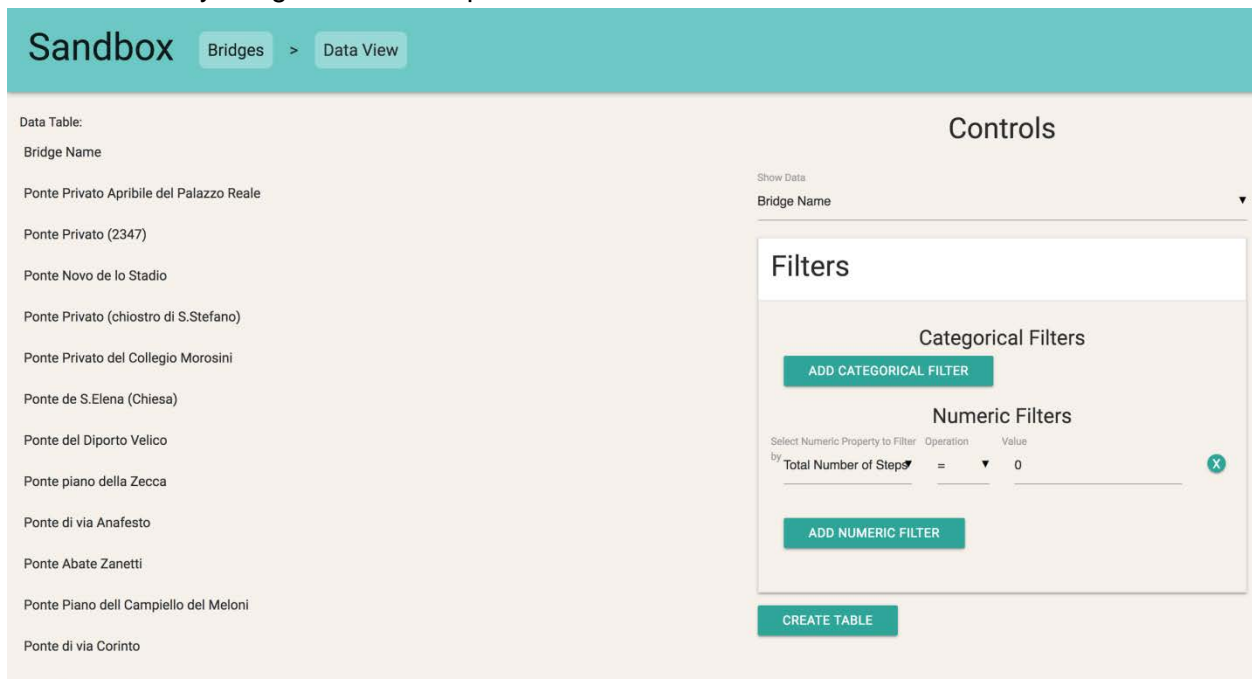


Figure 52: Data View tool

To use this tool:

1. From the home screen, select the desired data set and then select “Data View” from the visual selection screen.
2. Check the checkboxes under the “Show Data” drop down menu corresponding to the columns that are to be displayed in the data table.

3. Click the Filters button and apply filters, if desired.
4. Click the Create Table button and the desired data will be displayed.

Publishing visualizations to the Data Site

Once a static or interactive visualization has been created using the Sandbox Application, it is ready to be published. It can be exported in two different ways. It can be exported as a download of an editable SVG file, or it can be published directly to the VPC Data Portal site accessible at dataviz.veniceprojectcenter.org. Both of these export options is accessible by buttons located at the bottom of each Sandbox Application visual creation tool pages, as shown below in Figure 53.

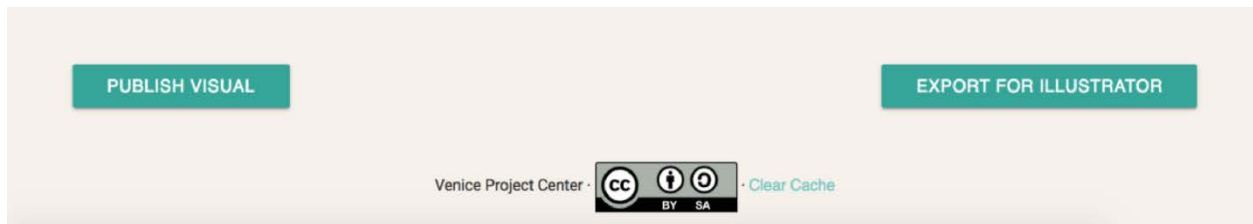


Figure 53: Export options in the Sandbox Application

Exporting as SVG

All visualizations produced using the Sandbox Application can be exported to an editable SVG file for easy importing into graphics design software such as Adobe Illustrator. This is the best option for those wishing to include a static visualization in a printed publication. The visualization can be edited to match the color scheme or design theme of the larger printed publication, and easily included. An SVG of a donut chart created using the Sandbox Application is shown below.

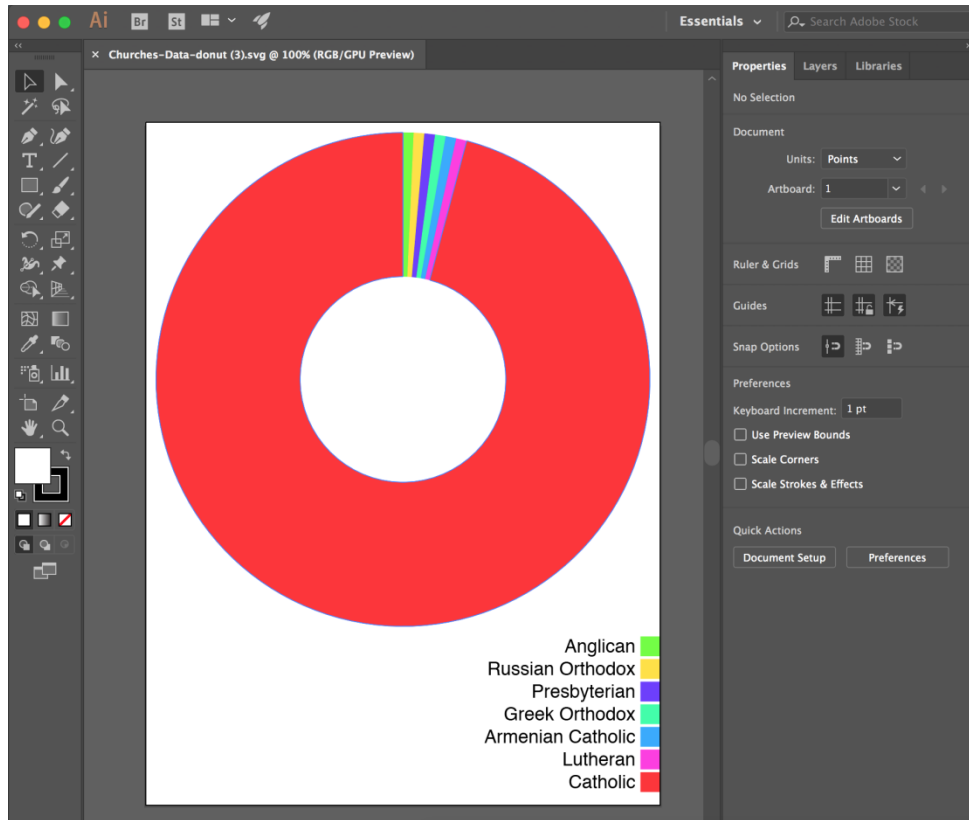


Figure 54: A Donut Chart SVG being edited in Adobe Illustrator

Publish Visual

Interactive visualizations can be exported from the Sandbox Application by publishing them to the VPC Data Portal website. To do this, click the “Publish Visual” button. This will bring up the Login screen shown below. This feature is reserved for authorized users as appointed by the Venice Project Center only. It is necessary to log in with valid credentials. This is to prevent visitors to the Sandbox Application from creating spam on the VPC Data Portal site. When an authorized user enters their username and password and clicks the “Login and Publish” button, the data needed to create that visualization is automatically pushed into Firebase, and the visual automatically appears on the dataviz.veniceprojectcenter.org site. Once the visualization appears on the Data Portal site, it is available to the general public, and can be embedded in any webpage using the HTML Code Snippet feature shown below. More information about the Data Portal site is available in Chapter 7.

Visual Removal

Published visuals can be removed from the Data Portal on the manage configs page (linked in the footer of the Sandbox Application). From there a list of all the visual configs is displayed (see Figure 55) and clicking on one will prompt for authentication and then delete the visual config. Deleting a visual config will immediately remove it from the Data Portal.

Sandbox

Click on a config to delete it:

Public Art in Greatest Need of Restoration (c. 2011)

Erratic Sculpture Coats of Arms

Filter Map

Passable Bridges at High Tide

Bridges

Filter Map

Map of usable water fountains

Fountains

Filter Map

Figure 55 Deleting a published visual