Real-Time Visual Analytics: An Experiential Learning Activity for Undergraduates

Paul Stephens
prs@fsmail.bradley.edu

Jacob Young
jayoung@fsmail.bradley.edu

Bradley University
Peoria IL

Abstract

We describe a "Day of Giving" university fundraising event that can be used to introduce data visualization to undergraduate students. The project involves integrating data sources, creating a Tableau data model, and designing a heat map that can be embedded into a front-end website. Our case provides opportunities to discuss various technological concepts, such as: client/server networks, front end web development, backend database servers, database design, sources of data, data preparation and cleaning, data management, webhooks, real-time data collection, and visual analytics. For the purpose of this paper, we focus on concepts related to sources of data, real-time data collection, visual analytics, and webhooks, as well as security and ethics issues that arise from these activities. Throughout the paper we explain how instructors can choose to implement the activity as a mini-case during a single class session or as a team project over a longer period. Both approaches provide students with experiential learning opportunities in data analytics. First, we outline fundamental concepts for instructors to introduce at the start of the activity. Second, we describe the institution's need for a heat map to display donation amounts by location in real-time. Third, we discuss the tools we used to feed data to the visualization. Fourth, we describe steps for instructors to follow to replicate the project. Lastly, we provide discussion points to explore security and ethics issues related to data analytics.

Keywords: analytics, data visualization, experiential learning, pedagogy

1. INTRODUCTION

A conceptual discussion of how technology works provides the basis for experience. A hands-on experience with technology helps students retain concepts (Christen, 2009; Hsiao, Chen, Lin, Zhuo and Lin, 2019). Such experiences are sometimes difficult to create in a classroom environment due to time constraints (Gillies and Boyle, 2010; Le, Janssen and Wubbels, 2018) and the availability of the appropriate technology isolated from campus networks (Mew, 2016). Entry level, undergraduate courses need to introduce technical concepts to students who often have very limited hands-on experience with technology. And, these students need positive encounters that enhance their technological self-efficacy (MacLeod, Yang, Zhu and Shi, 2018).

A key concept to address when introducing business analytics is the use of data visualization to aid decision making (Adkins, 2016). It is also important for such visualizations to be provided in real-time (Toasa, Maximiano, Reis and Guevara, 2018). While providing students with data visualization experience is generally straightforward, creating hands-on experiences that simulate the collection of real-time data is more challenging. A common example of real-time data collection that will be familiar to students is a dynamic Internet website.

Most universities collect real-time data that they want to share with internal and external stakeholders. An example would be when the Advancement office at the university wants to show the result of a fundraising campaign. Fundraising

events and their results are often shared online publicly. In this paper, we demonstrate how an institutional need can be leveraged into an experiential learning activity, an in-class exercise, or a team project. In our exercise, students create a Tableau heat map for a "Day of Giving" which is updated in real-time as donations flow to the website. The links between the Tableau heat map and the website are created via webhooks using free, open source applications. The event can be introduced as a mini-case and the activity can be replicated as a simulation where the instructor walks the class through the process. The concepts, case, and activity could be integrated into an undergraduate introductory business analytics course, a visual analytics course, or even an ecommerce course.

Prior to facilitating the class activity, instructors should present students with conceptual information in support of the mini-case and in-class project. This can be done in a variety of ways, but we suggest two possibilities: (1) as part of a class lecture or (2) as a flipped learning activity outside of class. Our case provides opportunities to discuss various technological concepts, such as: client/server networks, front end web development, backend database servers, database design, sources of data, data preparation and cleaning, data management, webhooks, real-time data collection, and visual analytics. For the purpose of this paper, we focus on concepts related to sources of data, real-time data collection, visual analytics, and webhooks, as well as security and ethical issues that arise from these activities.

2. THE CONCEPTUAL INTRODUCTION

In the conceptual introduction, instructors should discuss data sources, data preparation, visual analytics, real-time data collection, and webhooks. If presenting the material in class via lecture, we recommend that response system software be used to engage the students in discussion and feedback. If a flipped learning activity is required outside of class time, then we recommend that a graded quiz, implemented through the university learning management system, be developed to ensure that each student covers the material prior to the in-class activity. In this section, we provide a brief synopsis of the conceptual material that should be covered before beginning the in-class project.

Data Sources, Preparation, and Cleansing
A discussion on data sources should include where data can be found, various data formats, types of data, data messiness, and the need to clean and prepare data prior to analysis. Issues like the size of the data source, the accuracy of the data, data security, and data redundancy should be discussed. Instructors should also introduce relational databases and referential integrity concepts. Something simple like a Microsoft Excel workbook can be used to illustrate all the potential problems that lead to data difficulties and the need to clean and prepare data. An overview of a tool like Tableau Prep could also be used to discuss data cleansing and preparation.

Visual Analytics
Next, instructors should provide background on the purpose and function of visual analytics. Data visualization helps people more easily understand information contained in data as opposed to viewing raw data or a mathematical algorithm. Visualization also helps people grasp difficult concepts and identify patterns in data. A variety of resources are available to illustrate the power of information visualization (e.g., Ware, 2012). The forecasting tool built into Microsoft Excel could be used to illustrate these points.

Leveraging Real-Time Data Collection
Instructors should then explain what "real-time" means in terms of computer systems. A point to emphasize is that a system processing event generates a signal that indicates an event has occurred. That signal can act as a trigger to access and gather data associated with the event. To extend the conversation, the importance of real-time data analytics to decision making can be illustrated in examples like stock trading, hospital patient care, inventory management, or fraud detection. The most common type of real-time data visualization is a dynamic Internet website, such as database driven websites. A simple example that students can understand is a baseball website, where the data being displayed changes with every pitch during a game. Students can also relate to social media examples. Illustrating how businesses scrape twitter data in real time to perform analysis on trending hashtags in order to refine their marketing campaigns is a relevant business example that students understand.

Extract-Transform-Load (ETL) Webhooks
Students also need to understand that a webhook is a method used to deliver data from one application to another in real-time. The system events that serve as the triggering events for the webhook are described while introducing the importance of real-time data analytics. Webhooks

look for a specific event (e.g., when a customer completes a purchase on an ecommerce site). A couple of examples can be mentioned, such as MailChimp, which uses a webhook to signup visitors of a website to a newsletter, or PayPal, which uses a webhook to tell QuickBooks that a client has paid their invoice.

3. THE MINI-CASE AND IN-CLASS ACTIVITY

The idea for an in-class data visualization activity emerged from an undergraduate independent study where one of our students completed a similar project based on an actual need at the university. Internal civic engagement projects not only aid both the university and the students but can often be turned into in-class activities that provide experiential learning opportunities.

The mini-case that we developed centers around a university advancement event called a "Day of Giving". Most universities are non-profit organizations that rely on donations to help fund the operations of the academy. A "Day of Giving" is one tactic institutions use to garner donations. The idea is to set aside a special day to celebrate the university and to recognize all the people who care for and support the institution. Stakeholders are then targeted for donations on that day. A running total of donations is captured and presented to the public through a website.

We discuss how alumni and other friends of the university are located throughout the United States. The university advancement team recognizes that a geographic heat map of donor generosity, presented in real-time, would be an interesting addition to their "Day of Giving" website. The purpose of this visual would be two-fold. First, it represents the geographic breadth of giving, and second, it creates a sense of visual competitiveness. You can imagine a potential donor viewing the map and thinking "my state, my city, is giving the most. We are winning." However, the advancement team is not sure how to bring the idea to fruition, so they use their internal contacts to ask for help. A management information systems instructor offers to have undergraduate students create the heat map at no additional cost to the university.

4. TECHNICAL ASPECTS OF THE PROJECT

The independent study project consisted of four parts using four different tools. First, the university was contracted with MobileCause, a fundraising service that provided a front-end website that acted as the "day of giving" interface

to capture donations and location data. To maintain consistency, we emulated the MobileCause website used by the university for their actual day of giving.

Second, we used Stitch, an open source extract-transform-load (ETL) application that allows for webhook integration. Stitch collects the data "ping" from the front-end website and delivers the data to our data warehouse. Stitch is a good choice for this project because the service is free for the first five million rows of data. Given the limited scope of our project and in-class activity, we were not concerned with exceeding the five million row limit.

Third, we used data.world as our cloud data warehouse. We chose data.world to host the data because it was free to use and integrate with Stitch. The free individual account allows for 3 projects/databases with a 100MB limit per project/database.

Lastly, we selected Tableau as our data visualization tool. We connected Tableau to our data.world database and built the data model using the geographic data. Once everything was fully configured, the heat map was embedded into the front-end website (see Figure 1).
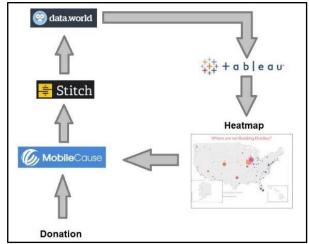

Figure 1: Project Process Map

Various aspects of the project could be built either ahead of time or as part of the in-class project depending on the needs of the course. The front end (donation website) and back end (Tableau data model and heat map) can be prepared for the students or the students could build them from scratch or finish either end as needed. For our in-class activity, we created the front-end website ahead of time, identified the additional

data sources for the Tableau model, then cleaned and prepped those sources. The students can then create the data integration, build the Tableau data model, and design the heat map.

5. IN-CLASS PROJECT PROCESS

To leverage the project into an in-class activity, we begin by previewing a donation website that consists of two simple pages and emulates an online donation website. The homepage provides an explanation of the imaginary day of giving and is where the heat map will be displayed. The second page contains a donation form.

Depending on how hands-on you want the project to be, you can use a single website or replicate the website for each student in the class. An online web hosting service provided us with a free developer account so each student in the class had their own development website.

Stitch Integration
The first step to get Stitch functional is to connect it to the donation website by adding a new integration. Stitch will integrate with a variety of sources, such as BigCommerce or Shopify. For example, with Shopify, you would create a unique integration name and provide the name of the Shopify shop you are using. The integration process between the two sources is automated. Alternatively, if you have your own web server and website, Stitch can generate a webhook URL that can be used via an HTTP POST in a PHP script. Webhooks can be generated through the integration feature on the Stitch website (see Figure 2). Either method creates an active link for real-time data collection.

The webhook integration requires a name and primary key. The primary key field should be left blank. By leaving the primary key field blank, Stitch generates a unique primary key for each data record scraped from the donation website. Once saved, Stitch generates a webhook URL that links to the front-end donation website. Data collected from the front-end website will be delivered to the webhook URL. Stitch then feeds the scraped data to a data warehouse.

Data.World Data Warehouse
After creating a Data.World account, students can create a new dataset and give it a name. Each student should then integrate their Stitch account through both Stitch and Data.World. Using the destination link in Stitch, a form will pop up and the student will enter their Data.World account ID. The data is automatically stored in JavaScript Object Notation (JSON) format. Stitch will test the connection; if the connection is successful, then the process is complete.

Finally, permissions must be granted from data.world to allow Stitch access to the data. First, go to "Your Integrations" from the data.world site. Next, edit the integration and check if it is enabled. If it is not enabled, the permissions may need to be granted. If needed, both Stitch and data.world have extremely helpful documentation that can walk you through each step of the process.

After the connection is enabled in both Stitch and data.world, the data can flow in real-time from the donation website, through Stitch, and into the data.world database. We have found that the process from donation to data.world takes roughly 15 minutes. You can have the students enter dummy data through the donation website or you can create a script and trigger it to feed some preset data into the system. In Data.World, the data can be viewed in the user account accessed via the Internet. In addition, the data can be queried, if so desired. For example, you could ask students to write a query to count the number of donations over $100. We didn't do this for the project because the main purpose of our data.world connection was its ability to integrate easily with Tableau.

Integration with Tableau
Just like with Stitch, to integrate Tableau with data.world, go to "Your Integrations" and connect to Tableau. As needed, check and enable permissions. Using the web data connector in Tableau, a connection is made to the data.world URL.

One of the issues we encountered is queries created in Tableau Desktop do not establish a live connection to the data.world data. In Tableau Server or Tableau Online, extracts can be refreshed every 15 minutes to match the feed through the pipeline. At first, we considered this a problem, but we then realized it was an opportunity to discuss the differences among Tableau versions, such as how they work in organizations that use Tableau Server versus the desktop version.

This step also provides the opportunity to discuss Tableau Bridge, a resource that can maintain connections to data sources that Tableau Online cannot reach directly.

Tableau Data Model
The data collected in the database includes the zip code for each person submitting a donation. Our design is based on showing donations by the 210 recognized media markets in the United States. Therefore, we need a data source that creates a relationship between the donations and the media markets. A free source of all U.S. zip codes is easy to find online. However, finding a free source that links each media market to each zip code currently doesn't exist. A listing of media markets is available, but a considerable amount of data preparation is necessary to create the one-to-many relationships needed for the data model. We don't take students through this process for the in-class project, but we do explain it. Depending upon your needs, it would be possible to include this step, but you would have to expand the project across multiple class periods. Given that we are introducing much of this material conceptually in a lower level undergraduate class, we don't feel it is necessary. Although, a brief explanation might be useful for those interested in replicating the process as a class exercise.

The zip code database (see Figure 3) lists every zip code in the United States and its designated market area (DMA). These markets are geographical areas where all residents receive the same broadcast offerings, including television stations and radio stations. The bulk of the Zip code data is available for free from ZipCodes.org (https://www.unitedstateszipcodes.org/zip-code-database/) and downloadable as an Excel workbook. But, the source doesn't come with the DMA data. We added that data ourselves. For the purposes of the exercise, the two Excel workbooks (zip codes and media gps) are cleansed, prepped and ready to go. At this point in the exercise, instructors can briefly revisit the concept of cleaning and prepping data sources.

As for the designated market areas workbook, that information is also downloadable from a variety of web sources. But, the GPS coordinates are not included. Instructors can review the data prepping process by demonstrating a free online geocoding tool, such as mapdevelopers.com used to find the latitude and longitude of the most prominent city in each DMA. This city served as the central point where the heatmap would place the mark to signify a donation in that DMA (see Figure 4). Both workbooks are available to interested parties upon request.

If instructors provide the students with the workbooks, then a data model consisting of two one-to-many relationships can be constructed in Tableau (see Figure 5). The transaction report is the link to the data.world database. The basic entity relationship diagram is shown in Figure 6.

Tableau Heat Map
Creating the simple heat map in Tableau is straightforward. Using the default circle as the mark, the sum of the collected donation amounts is used for the size, and the DMA is used for the color. The latitude and longitude from the Media Market spreadsheet are placed in the columns and rows to generate the map. We restrict it to the lower 48 states in the U.S. (see Figure 7). At this point, the instructor can feed additional data through the pipeline via the donation website and students can watch the map update in real-time.

Adding the Heat Map to the Website
The final step involves adding the heat map to the donation website. Using Tableau Public or Server, you generate some code to embed into the website. If you are using an online developer platform, you can simply edit the page and paste the code into the appropriate space.

Given the time restraints of a single class, we recommend having the instructor demonstrate this process. Therefore, the Tableau Public or Server account, setup, and design should be in place prior to class. During the demonstration, the instructor can explain the purpose of Tableau Public and/or Server, and how they are used.

Final Product
At this point, for our exercise, the final product is just a dynamic heat map embedded on a single web page without any other information. The dummy "donation" website is a work in process and will develop into something more meaningful as we continuously improve the exercise.

This exercise evolved from an independent study where one of our students worked on an actual day of giving project for our university. If you want an idea of what the finished project might look like as we continue to refine our website for the exercise, this URL will take you to that project https://www.bradley.edu/sites/building-bradley/

6. SECURITY AND DATA ETHICS

Due to the rise of big data and the development of increasingly powerful analytical tools, we believe that these issues must be integrated throughout the entire information systems curriculum as much as possible. Therefore, once the data visualization aspects of the in-class

activity have been completed, we also encourage instructors to discuss issues related to security and data ethics.

For example, students will have access to the entire donation database, which includes the donor names and complete addresses. Yet, the only information needed to create the heat map is the donation amount and donor zip code. This provides an opportunity for instructors to demonstrate how providing analysts with full access to sensitive and identifiable data, especially data that is not needed for the task at hand, can increase the risk of a breach. Instructors can then explain the need for access controls, encryption, anonymity, and data minimization.

There are also several questions that can be discussed from a data ethics perspective. For example, many donors might wish to make an anonymous gift. If a donor elects to keep their contribution anonymous, should their donation data be included in the heat map? What if only one alumnus lives in a given zip code? Are there any other data sources that, when combined with the heat map data, might reveal the identity of an anonymous donor? Further, what disclosures should be provided on the donation website?

To reduce the risks associated with big data and data analytics, students must develop the ability to identify ethical issues and potential threats to privacy and security prior to accessing data. Including an ethical component and discussing how technological advancements can sometimes lead to negative outcomes will help students recognize and mitigate similar risks in the future.

7. CONCLUSION

For many years, pedagogical research has recognized the value of learning with technology rather than simply learning about technology (Schuldt, 1991; Podeschi, 2016). Today, researchers recognize hands-on, class projects as active, experiential learning tools for students well suited to a flipped classroom approach (Wu, Manohar and Acharya, 2016). Our data visualization activity presents such an opportunity for instructors. We can see it being used in an introductory business analytics course, a database course, or even as part of a survey course on electronic commerce. It should be noted that our experience with the exercise is nascent. We haven't tried it out in multiple courses but the plan is to do that and collect additional information in the process.

Yet, we would posit that the in-class exercise allows individual instructors to have some flexibility in how they manage their time constraints, how much setup to complete prior to class, and how they want students engaged in the project. For our purposes, we see this project as appropriate for students learning about these concepts and tools for the first time. We want to keep it simple and limit the amount of actual hands-on work. Given the variation in student abilities, their technological self-efficacy, and the possibility of technological glitches, this exercise can be accomplished in a single 75-minute class period.

The purpose of our exercise is to give meaning to the conceptual discussion by letting students perform some of the tasks necessary in making the project work without needing multiple class periods. At the same time, we recognize the opportunity to turn this into a full-scale project that could be required as an assignment outside of class time. We believe such a project would be more appropriate for advanced undergraduate students.

8. REFERENCES

Adkins, J. K. (2016). Progression of a Data Visualization Assignment. *Information Systems Education Journal,* 14(6) pp 20-26.

Christen, A. (2009). Transforming the classroom for collaborative learning in the 21st century. Techniques: Connecting Education and Careers, 84(1), 28-31.

Gillies, R., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. Teaching and Teacher Education, 26, 933–940.

Hsiao, H., Chen, J., Lin, C., Zhuo, P., & Lin, K. (2019). Using 3D printing technology with experiential learning strategies to improve preengineering students' comprehension of abstract scientific concepts and hands-on ability. Journal of Computer Assisted Learning, 35(2), 178–187.

MacLeod, J., Yang, H. H., Zhu, S., & Shi, Y. (2018). Technological Factors and Student-to-Student Connected Classroom Climate in Cloud Classrooms. Journal of Educational Computing Research, 56(6), 826–847.

Mew, L. (2016). Information Systems Education: The Case for the Academic Cloud. *Information Systems Education Journal,* 14(5) pp 71-79.

Podeschi, R. (2016). Building I.S. Professionals through a Real-World Client Project in a Database Application Development Course. *Information Systems Education Journal,* 14(6) pp 34-40.

Schuldt, B.A. (1991). "Real –World" Versus "Simulated" Projects In Database Instruction. Journal Of Education For Business, 67(1).

Toasa R., Maximiano M., Reis C. and Guevara D. (2018). Data visualization techniques for real-time information — A custom and dynamic dashboard for analyzing surveys' results, *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, Caceres, pp. 1-7.

Ware, C. (2012). Information Visualization: Perception for Design (3 ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Wu, P. Y., Manohar, P., Acharya, S. (2016). The Design and Evaluation of Class Exercises as Active Learning Tools in Software Verification and Validation . *Information Systems Education Journal,* 14(4) pp 4-12.
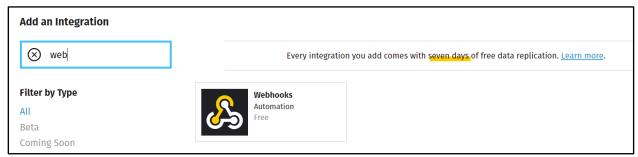
Appendix A: Additional Figures

Add an Integration

⊗ web

Every integration you add comes with seven days of free data replication. Learn more.

Filter by Type

All

Beta

Coming Soon

Webhooks
Automation
Free

Figure 2: Webhook Integration with Stitch

zip	type	primary_city	state	county	timezone	area_code	country	DMA
36066	STANDARD	Prattville	AL	Autauga County	America/Chicago	334	US	698
36068	PO BOX	Prattville	AL	Autauga County	America/Chicago	334	US	698
36003	STANDARD	Autaugaville	AL	Autauga County	America/Chicago	334	US	698
36008	PO BOX	Booth	AL	Autauga County	America/Chicago	334	US	698
36067	STANDARD	Prattville	AL	Autauga County	America/Chicago	334	US	698
36749	STANDARD	Jones	AL	Autauga County	America/Chicago		US	698
36051	STANDARD	Marbury	AL	Autauga County	America/Chicago		US	698
36006	STANDARD	Billingsley	AL	Autauga County	America/Chicago		US	698

Figure 3: Zip Code Spreadsheet with DMA

Rank	Market	State	Latitude	Longitude	DMA
165	Abilene-Sweetwater	Texas	32.45464	-99.7383	662
152	Albany, GA	Georgia	31.5776	-84.1763	525
59	Albany-Schenectady-Troy	New York	42.6664	-73.7987	532
47	Albuquerque-Santa Fe	New Mexico	35.10555	-106.647	790
179	Alexandria, LA	Louisiana	31.29345	-92.4702	644
208	Alpena	Michigan	45.07448	-83.4383	583
131	Amarillo	Texas	35.20038	-101.828	634
147	Anchorage	Alaska	61.14834	-149.195	743
10	Atlanta	Georgia	33.76265	-84.4231	524
105	Augusta-Aiken	Georgia	33.36446	-82.0708	520
40	Austin	Texas	30.30718	-97.756	635
122	Bakersfield	California	35.35015	-119.031	800
26	Baltimore	Maryland	39.30098	-76.6106	512
155	Bangor	Maine	44.83227	-68.7906	537

Figure 4: Media Market Spreadsheet

○ Live   ○ Extract

Connections   Add

Donation Data
Microsoft Excel

media_markets_...al_prepped (3)
Microsoft Excel

transaction_report_201904... ▾ —◑|— zip_code_database —◑|— Media_GPS

Figure 5: Tableau Data Model

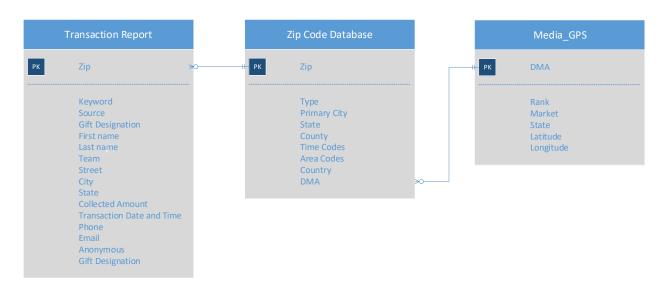| Transaction Report	Zip Code Database	Media_GPS
PK Zip	**PK** Zip	**PK** DMA
Keyword Source Gift Designation First name Last name Team Street City State Collected Amount Transaction Date and Time Phone Email Anonymous Gift Designation	Type Primary City State County Time Codes Area Codes Country DMA	Rank Market State Latitude Longitude

Figure 6: Entity Relationship Diagram

Figure 7: Example Tableau Heat Map Output