DonVis: An Interactive Tool for Donation Information Visualization

Arthur Sun
School of Electrical and Computer Engineering
University of British Columbia
2329 West Mall
Vancouver, B.C. Canada V6T 1Z4
Email: sunb@ece.ubc.ca

Tian Huaying
School of Electrical and Computer Engineering
University of British Columbia
2329 West Mall
Vancouver, B.C. Canada V6T 1Z4
Email: tianhuayingapply@126.com

Abstract—Public donations have already been widely accepted as a way to help those people or community who are in urgent need of help considering their current financial situation and great publicity has been spread about the importance of public donations. Nowadays, we witness an astonishing amount of public donations which has been noted by various non-governmental organization(NGO). However, as for specific donation like classroom donation, which aims at donating to those mostly needed schools in highly-poverty areas in every country. It has not been paid enough emphasis. We introduce DonVis, which is an interactive tool for donation visualization given a specific dataset for input, to visualize the distribution and situation of donations across multiple states and aware people of location where donation is mostly needed.

Keywords: Data Visualization, DonVis, Donation Analysis, Web Donation Visualization

I. INTRODUCTION

Nowadays, there is a growing phenomenon in countries that governments are realizing that helping people who are in urgent need of help is becoming more and more important in building a harmonious society. Moreover, Non-Governmental Organization so called NGO is playing an increasing role in helping people get organized action with government to initiate resume operation towards those people in need. Donation is often regarded as an effective way to ease the living pressure of the needful people. People tend to donate money towards the group which are similar with them because people are group-aware.[1] However, there exists certain circumstances that people in a group or individual who are willing to make a donation but lack a certain channel or target to donate to. They have strong willingness to help others but are at a loss. In this way, it is clear that donation recommendation should play an important role in helping public and people to get a better understanding about where donation is most needed and help is most appreciated.[2] The idea for donation information visualization for us was based on an accidental browse on YouTube about an American doing charity work for Project Hope in China but found out that some poorest areas in China don't have governmental support for funding and there is a huge gap to realize even the subtlest change for them. The American wanted to align with his other friends to donate money to the places where were mostly needed to but lacked

specific information source for where to donate to. We try to find an open dataset from China of donation status for Hope Project but failed and accidentally found the public dataset for US on donors choose. We hope our project can be a stepping stone for people to realize that there are still many places in the world where donation is urgently needed but due to certain conditions, we don't know and visualization can help us find the weak spot. With more open datasets for donation, we can demonstrate more people the most needed place for donation. Suppose there is a very clear picture that donation in L.A for grade three school books is much less sufficient than those in N.Y.U and you are counting on reliable information source to give you recommendation about your donation, will you donate to N.Y.U still rather than L.A? People would change behavior based on the actual results, and donate to a different state where donation is most needed. We want to clarify the goal for the visualization: show recommendation to people about where donation is mostly needed when they have no exact tendency to donate to and use what we learn in class to visualize the dataset, help the public understand the data better, and then give appropriate suggestions on public donation. However, you pointed out that it is hard to believe that anybody would change behavior based on seeing these results, and donate to a different state. This may make some sense under certain circumstances, i.e. it is the school where I attended before, I was born in this state or some other reasons, but according to our investigation in this field (after reading papers about what contributes to public donation behavior), appropriate guidance can make a great difference in public donation. What' s more, a great proportion of people still have strong wills to donate to the places which are in desperate needs of public help. Personal emotion is much less important than we thought, we always donate to where donation is desperately needed. Under such circumstances, we think it is of vital significance to visualize the data.

II. PREVIOUS WORK

As for now, there are a growing trend of rapid developments in US that an increasing number of startups and companies are focusing on providing data strategy and recommendation to Non-Governmental Organization because of the emerging

trend of big data analytics, which enables organization to have a better understanding over where the help is mostly needed. Websites such as Guide Star[3], which specializes in providing enterprise solution for company and has the world's largest source of information on NGO donation. Charity Navigator[4] is a global platform which provides online donation foundation for various kinds of donation projects which are currently happening all around the world. American Institute of Philanthropy[5] is the largest donation decision recommendation website in US which provides accurate watchdog information to help donors make informed decision. It provides charity rankings for the best NGO organization and top-rated charities community which helps people to know where they can make a donation to the nearest NGO organization to help those in need. However, it is obvious that above web organizations only make recommendation for the public to get awareness of certain NGOs and where certain NGOs requires donation. Althoff, Tim and Leskovec, Jure [6] investiaged online community donation project like donorschoose to study an online crowdfunding platform which enables people to get an detailed over view of how people direct their donations. They explore various factors impacting donor retention to identify different groups of donors and quantify their propensity to return for subsequent donations and return to a conclusion that donors are more likely to return if they had a positive interaction with the receiver of the donation. It is quite obvious that with the development of web technologies, donation choices for people are becoming more and more diverse, especially with website like donnorschoose, which makes crowfunding possible without the traditional help of the NGO. For certain donation specific visualization application, Tavi Nathanson, Ephrat Bitton and Ken Goldberg [7] proposed a domain specific dashboard recommendation system that recommends non-profit organizations specific donation to users in the form of a portfolio of donation amounts, which is derived from a certain database for data integration and clustering. Recommendations are made using Eigentaste 2.0 constant-time collaborative filtering algorithm in combination with a weighted portfolio of recommendations, which is fast in data processing but lacks specific visualization interface to show an overview donation status from NGOs.

White, Howard and McGillivary, Mark[8] proposed Suits' index and McGillivray's adjusted performance index to help the public evaluating donor allocative performance with respect to the average income of recipients. This proves to be an effective way to compare the performance of different donation objects, but it is far complicated for ordinary people to use and get a clear picture in a short time. Tableau Corporation [9]proposed an enterprise visualization solution for NGOs to show them show fund raising trend distribution based on Google Map and relevant each state's funding states. It is a big project with dataset covering all major NGOs around the world and pinpoint every location that requires funding based on user's filter and search. Murray[10] mentioned in his book about interactive visualization technology called D3.js, which is a popular data-driven visualization open-source framework

based on javascript, it achieves rapid rendering speed with support for most of the existing web framework, which is a huge step towards web visualization technology.

III. DonVis

We combine the latest web technologies D3.js with open dataset from donnorschoose regarding the distribution of classroom donation situation in US to provide the public with web-based donation recommendation with interactive method just by clicking the mouse with several times. DonVis is characterized as simplicity with complexity, which uses interactive method between the user and the back-end database to achieve real-time data refreshing and loading with hard-to-notice delay. DonVis also provides a new insight into specific donation target: classroom donation, which is used to be government-sponsor funding project but due to certain factors, abundant funding is not available and requires the public to give immediate help to those children. Recommendation is also an advantage in DonVis because it is quite clear that people are more acceptable to illustrative pictures than descriptive words and DonVis combines multiple dimensions of data encoding into concrete illustrative pictures to offer people with better idea of the overall picture of current donation situation. Donation recommendation should be important for the public because we started the idea for donation information visualization based on an accidental browse on YouTube about an American doing charity work for Project Hope in China but found out that some poorest areas in China don't have governmental support for funding and there is a huge gap to realize even the subtlest change for them. The American wanted to align with his other friends to donate money to the places where were mostly needed to but lacked specific information source for where to donate to. We try to find an open dataset from China of donation status for Hope Project but failed and accidentally found the public dataset for US on donors choose. We hope our project can be a stepping stone for people to realize that there are still many places in the world where donation is urgently needed but due to certain conditions, we don't know and visualization can help us find the weak spot. With more open datasets for donation, we can demonstrate more people the most needed place for donation.

A. Scenario

You are an warm-hearted donator who are now eager to donate your money towards a state, which you have not decided yet and wanted to find out where money is most needed among all places.

So at top level, you choose the state you want to see in the first place such as California(CA) as filter to filter the detailed data for CA. On top immediately shows you the total amount of donation number plus donated money altogether to help you get the first glance about the overall number.

You may be wondering what time is the peak time for donation and what time is the lowest time for donation. Then you get a quick glance at the multiple linked view to get a quick search and find out that peak donation for CA is from 2005 and then

drastically drop from 2006. You may want to check a fixed time interval so you use your mouse to set the starting time and end time so you see the total amount of donation as well as detailed relevant information

You place the pointer over a specific resource to check the amount of money donated such as books, the system then automatically change all the funding status, donation grade level as well as poverty level and state comparison accordingly by the time you choose. Then you can get a full understanding about the donation state year by year in detail for your own reference and help you to make recommendation for donation: what state, which grade as well as which resources to donate given the historical trend.

IV. GENERAL INTRODUCTION

A. Dataset

The dataset is an open dataset called classroom dataset from crowdfunding website Donorschoose with downloadable format in CSV where the table is a two-dimension 104,8576 record table which 43 columns of information made up of 500MB in size. Since we want to visualize state-specific donation by grade, year, resource type, we defined the key words we are going to do abstraction which are derived from each column of data from the dataset, we need to convert a large table into multiple graphs. For the load speed for visualization in webpage, we now use 10,000 lines of data instead of choosing them all.

The way we choose the 10,000 lines of data is by randomly choose from the pool but keep every state with 16,667 lines of data respectively for fair comparison.

1) School State: California(CA), Illinois(IL), North Carolina(NC), South Carolina(SC), New York(NY), Texas(TX)
2) Resource Type: Book, Other, Supplies, Technology, Organized Trips
3) Poverty Level: Low, Moderate, High, Highest
4) Year: 2003,2004,2005,2006,2007,2008
5) Funding Status: Lack, Plentiful
6) Grade Level: Grade 3-5, 6-8, 9-12, PreK
7) Donation Amount: Counted in US Dollars

B. DonVis Outline

1) What: Data: Table(database) with 7 attributes for 10,000 lines; Using SQL Query to return table subset
2) What: Derived: Quantitative attributes for every original line: query relevance for the 7 original attributes plus 1 overall relevance attributes: Belonging Group
3) Why: Tasks: Characterize state-specific distribution of donation within different attributes: year, resource type, poverty level, funding status. Finding donation differences within different state group, find correlation between attributes within the same group or different group. Compare different attributes in the same group.
4) How: Encode: multiple linked list, multiple linked bar chart, multiple linked pie chart, horizontal spatial position for donation distribution by year, vertical spatial

position for ordered multiple attributes. categorical and ordered luminance
5) Manipulate: Navigante with pan/scroll/click
6) How: Facet: Multi-correlated form with linked highlighting and shared attribute aggregation. overview-detail with selection in overview then populate detail view.
7) How: Reduce: Filtering, dynamic aggregation

V. IMPLEMENTATION

A. Design

1) *Analyze:* We want to use existing dataset of classroom donation projects provided by Donorschoose to make people understand and discover the fact that although the notion of donation is widely-spread and supported by public, yet most people don't acquire certain knowledge about the fact that some area processes high volume of donation while other area processes relatively low volume of donation and requires people's attention. Based on the goal of the previous step, we are planning to discover and derive the most needed states for public donation and what are most needed for donation: books, classroom or others given by dataset we choose and show a trend of donation given by year and by state.

We aim to provide users with explore for the easiest way to control the interactive visualization to give users the access to choose what information they want to see and give them a bird's overview with the overall information. So we are going to use dynamic changing table to allow user to choose which state and what time duration they want to choose to show the data they are interested in. The visualization will details searching for states and years with regard to their location, often beginning from an overview of everything.

2) *Targets:* Targets: For high-level target we want to achieve is to show a trend, which is a high-level characterization of a pattern in the data and we want to show the trend of donation in current six states to give people recommendation when they want to donate.

For low-level targets down to an attribute, we want to classify the distribution of each individual key words we extracted from dataset: resource type, poverty level, funding status, etc which is correlated and have dependences with previous chosen state and year, while poverty level and funding status have mutual correlation with each other concluded form the dataset.

3) *Reduce:* As for the original dataset,there are more than 1048576 records,but we only need 10000 attributes in our project. As a result,we need to reduce. Reduction is one of five major strategies for managing complexity in visualizations.Typically, static data reduction idioms only reduce what is shown, as the name suggests. However, in the dynamic case, the outcome of changing a parameter or a choice may be an increase in the number of visible elements.We will use attribute filtering and item filtering to reduce attributes and items. The design choice of filtering

is a straightforward way to reduce the number of elements shown: some elements are simply eliminated. Then we will use summarize as the idiom for reduce stage because after a user select specific state and time during, requests will be sent to back-end database for sum() and select() calculation for summation of donation, the different categories of most needed goods, etc, which will give a comprehensive view of what the viewer has chosen.

4) Aggregation: Data aggregation is conducted in the back-end database with SQL command line for key words classification

1) Categorize every column of data by state and do SUM() for total donation for the state and using a number to visualize called Total Donation
2) Categorize previous state-specific data into chronological order using GROUP BY() and enable data-duration selection by real-time GROUP BY() query into database
3) Classify resource type by SELECT() given previous step to show detailed resource donation status by year during and state
4) Classify funding status by ORDER BY() into completed or expired given previous step
5) Classify donation of different grades, poverty levels and state in the same way mentioned above in the sketch below

B. Visual Encoding

1) Number of Donations: number of donations, multiple linked view is used to show the overall donation by a specific time duration, which can give clear information to the users. The target users can control the time frame,choose what time range they want to see corresponding tendency in several aspects. The central benefit of the multiple linked view is in seeing how a region that is contiguous in one view is distributed within another. The rationale behind multiform encoding across views is that a single monolithic view has strong limits on the number of attributes that can be shown simultaneously without introducing too much visual clutter. Although simple abstract tasks can often be fully supported by a single view of a specific dataset, more complex ones often cannot. With multiple views as opposed to a single view, each view does not have to show all of the attributes. They can each show only a subset of the attributes, avoiding the visual clutter of trying to superimpose too many attributes in a single view.

2) Grades, Poverty Level, Resources: For composition of resources and donation by grades, class, poverty levels, bar chart can best show the percentage of composition by different categories with clear view of respective number regards to it. The reason that we don't use line chart is because it is largely due to the fact that line charts also use connection marks to emphasize the ordering of the items along the key axis by explicitly showing the relationship between one item and the next. Thus, they have a stronger implication of trend relationships, making them more suitable for the abstract task of spotting trends. As is mentioned in this course, bar charts are the best-known example for an idiom which shows one key and one value attribute, and are convenient to lookup and compare values.The key attribute like poverty levels and class, separate the marks along the vertical spatial axis. The value attribute, like amount of donation, express the value with aligned horizontal spatial position and line marks. Bar charts encourage discrete comparisons, which is exactly what we want,while line graphs encourage trend assessments.

3) Funding: Pie chart for funding status can best illustrate the proportion of the completed and expired since there are just two statuses and using other visualization method will not be so efficient and clear to see. Pie charts encode a single attribute with area marks and the angle channel. The most useful property of pie charts is that they show the relative contribution of parts to a whole,which is exactly what we want.In figure(4),we do not care about the exact number of completed funding and expired funding,we just want to have a whole overview of funding status,so pie chart is the most suitable choice. As a result,we choose bar charts to show the funding status.

4) Multiple View: We use small multiple views partitioned on condition and Juxtapose.With multiple views as opposed to a single view,each view does not have to show all of the attributes; they can each show only a subset of the attributes, avoiding the visual clutter of trying to superimpose too many attributes in a single view. Even if two views show exactly the same set of attributes, the visual channels used to encode can differ.

C. Detailed System Implementation
VI. DONVIS CURRENT CHALLENGES
A. Encoding Challenges

For InfoVis technique, we are still trying to get to understand how to best display various data aggregation result in group with the best match of visualization technique to best show the result for viewer as well as the best display in the webpage to help viewer better understand the overall picture because the previous proposal we proposed exists certain faults which we haven't considered.

B. Technical Challenges

The actual dataset we downloaded from Donorschoose is more than 500MB now and contains 104,8576 records, which are still currently increasing and in order to get a comprehensive view for overall status for current donation, we extract 10,000 data records into our MongoDB database so far and the reason we still used database is because before the next step of visualization, we need to do various simple calculation including summation, aggregation , etc. for all the records and there will exist huge delay for display if we do this in the front-end web page. For front-end web page system, we have

set up the Node. JS and finished running simple testing page on it and succeeded in connecting MongoDB with Node.JS to display dataset key-value stores. However, there still exist some problems to be solved: 1.Time delay for webpage to load the whole dataset is quite slow, which may take up to 10s to load the whole page and this requires a solution 2.Finding out a way to solve the complexity of having to do every data aggregation and complex coding for each chart. 3.Finding out a solution for convenient data slicing and dicing in the front-end system to reduce complexity 4.Optimize the layout for the page and arrange it into a acceptable page view

VII. PROJECT TIMELINE

1) 10.10-10.20: Arthur and Huaying: Project Scope Investigation and discussion, domain knowledge discussion and implementation possibility survey
2) 10.20-10.25: Arthur: Task Abstraction and specification. Huaying: domain knowledge deep investigation
3) 10.26-10.30: Arthur: Encoding technique verification and confirmation. Huaying: Design pattern planning and trying
4) 11.1-11.5: Arthur: Dataset configuration in Mango DB for testing environment. Huaying: HTML CSS learning for basic web development
5) 11.6-11.10: Arthur: Dataset data slicing and computation and connect to front-end Node.JS server. Huaying: D3, CSS, HTML learning and plotting the whole canvas
6) 11.10-11.20: Arthur and Huaying: Web page layout development and data loading performance tuning
7) 11.20-11.25: Arthur and Huaying: Finish draft version of the whole visualization web
8) 11.25-11.31: Arthur and Huaying: Improve the visualization web based on actual performance
9) 11.31-End: Arthur and Huaying: Paper writing

VIII. CURRENT PROJECT UPDATE

As for now, we have strictly followed the timeline we previously planned before and now we have finished most part of the project realizaiton including back-end side data filter and front-end side web engine setup and performance tuning. However, for the layout of the page, there still exists some trivial problems and we are currently working on them.

REFERENCES

[1] Karen Page Winterich, Vikas Mittal, and William T Ross. Donation behavior toward in-groups and out-groups: The role of gender and moral identity. *Journal of Consumer Research*, 36(2):199–214, 2009.

[2] Beth Breeze. How donors choose charities: the role of personal taste and experiences in giving decisions. *Voluntary Sector Review*, 4(2):165–183, 2013.

[3] GuideStar. Donation recommendation, 4 2015. URL `http://www.guidestar.org/Home.aspx`.

[4] Charitynavigator. URL `http://www.charitynavigator.org`.

[5] American Institute of Philanthropy. American institute of philanthropy. URL `https://www.charitywatch.org/home`.

[6] Tim Althoff and Jure Leskovec. Donor retention in online crowdfunding communities: A case study of donorschoose. org. *International World Wide Web Conferences Steering Committee*, pages 34–44, 2015.

[7] Tavi Nathanson, Ephrat Bitton, and Ken Goldberg. Donation dashboard: a recommender system for donation portfolios. *Proceedings of the third ACM conference on Recommender systems*, pages 253–256, 2009.

[8] Howard White and Mark McGillivray. How well is aid allocated? descriptive measures of aid allocation: a survey of methodology and results. *Development and Change*, 26(1):163–183, 1995.

[9] Tableau. Tableau non-profit visualization technique, 4 2015. URL `http://www.tableau.com/solutions/non-profit`.

[10] Scott Murray. *Interactive data visualization for the Web*. O'Reilly Media, Inc., 2013.