

Next Generation Time-based Visualization Tools for HIV Epidemiology

Yingyan Hua

Advisor: Dr. Nadir Weibel
University of California, San Diego
Computer Science and Engineering
y4hua@eng.ucsd.edu

ABSTRACT

The study of the HIV virus often involves complex clinical, social, and behavioral interactions among infected, uninfected, and at-risk people. Investigating causes and consequences of acquiring HIV is based on the analysis of a combination of data spanning phylogenetic, socio-demographic and geographic information. Better understanding relationships across these data may help to uncover hidden social, sexual, geographic and virologic patterns that could lead to novel insights in terms of intervention or prevention. However, it is hard to have a unified view on these heterogeneous data and the lack of a clear way to visualize these data makes their analysis harder. In this project we investigate these challenges and propose a novel web-based interactive visualization approach to augment molecular epidemiology study based on clusters of HIV infected patients, with a time component. Our approach enables explorations of how HIV clusters change over time by integrating interactive visualizations and dynamic timelines that can be operated to show data at specific times and in specific geographical context. We believe that our tools enable researchers and clinicians to gain novel insights and understanding on the local impact of HIV incidence.

Author Keywords

Avant Garde; HIV; Time-based Visualization;
Geographical Visualization

INTRODUCTION

The human immunodeficiency virus (HIV) is a retrovirus that causes the acquired immunodeficiency syndrome (AIDS) [1, 2] a condition in humans in which progressive failure of the immune system allows life-threatening opportunistic infections and cancers to thrive. According to UNAIDS 2014 global report, around 35 million people were living with HIV and 19 million of those did not know their HIV-positive status [3].

Although these numbers are high, by identifying affected or at-risk populations, it is possible to concentrate on going treatment and prevention more accurately [4]. Any improvement on identifying trends in the epidemiology of HIV could help preventing the transmission of virus and its potential risks. However, the complex socio-demographic and phylogenetic factors that influence HIV transmission make it extremely difficult to identify these populations [4].

Supported by a recent NIH NIDA Avant-Garde award [5], UCSD, and in particular Dr. Davey Smith from the Center for AIDS Research (CFAR) began to track and analyze HIV exploiting its transmission networks. The goal is to concentrate treatments/interventions on the individuals where the most gain in reduced HIV transmission would likely be achieved [6].

As part of this quest, the UCSD CFAR and the Antiviral Research Center (AVRC) at UCSD started to collect a range of clinical, demographic, behavioral and life-style data from the HIV-positive population in San Diego and Tijuana (Mexico) [8]. In order to support the mission of the Avant-Garde award, in this project we propose the use of a set of new web-based visualization tools to enable researchers and clinicians to interact with this HIV data and explore hidden factors in order to better understand HIV transmissions.

MOTIVATION AND BACKGROUND

Most of our work stem from the ongoing research of Dr. Mehta at UCSD who works on the Avant-Garde project and conducts experiments on Molecular Epidemiology of HIV at the San Diego/Tijuana border. According to his report on HIV epidemiology in Mexico, in 1980s, most HIV cases were identified in individuals who previously lived in the US. By 1991, only 44.3% of cases were in previous US residents, while in 2000 the rate dropped to 12% [13]. Figure 1 shows a map highlighting the migration of HIV from Mexico to the US. As expected, the states with high rates of migration like Zacatecas/Michoacan were also the

states in which higher numbers of HIV+ individuals had previously lived in the US [13].

This data serves to motivate our work, specifically researchers such as Dr. Mehta studying these datasets, are looking to answers questions like:

- Can we demonstrate that San Diego and Tijuana HIV epidemics are connected?
- Is there directionality to viral migration across the border?
- Because of HIV Diversity and High fidelity, what's the relationship among them?
- How does an international border affect the spread of infectious disease?

As evident from the data presented above, an internal relationship among different factors behind the HIV migration must exist. Moreover, researchers might assume that this relationship is somehow related to the evolution and changes at the San Diego/Tijuana border area [20]. However it is not easy to explore those relationships and researchers need a better way to look at the data taking into account the perspective of both time and locality.

No existing solution allows researchers to flexibly look at these data, and visualization, as well as interactive tools to explore links across the data have not been widely investigated yet.

We report here on our approach that is the result of an iterative design and development exercise to create tools to visualize HIV data. Our investigations lead to the creation of an interface heavily based on a time component that facilitate exploratory data analysis [21]. In the remainder of this paper we focus on

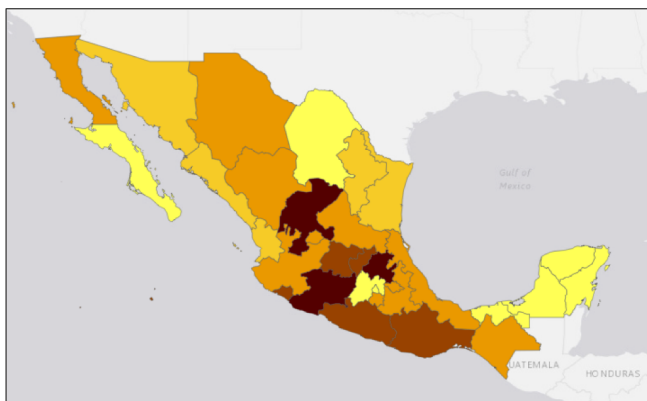


Figure 1. Migration to US: darker colors represent states with higher numbers of people that previously lived in the US.

this purpose and we describe the resulting visualization facilities we built. The use of our tools shows how it is possible to investigate different dimensions of the available data, and how these interact with each other, specifically outlining how geographical distribution of groups changes over time.

DATA COLLECTION

A number of medical sites in the San Diego/Tijuana region are conducting HIV studies that aim at better understanding the local impact of the HIV incidence, as well as a variety of biological, social, and clinical aspects of the epidemics.

As part of this project we take into account data collected from studies on both sides of the USA/Mexico border, in particular the Tijuana "Persons who inject drugs" (PWID) study, the "Female Sex Workers" (FSW) study, as well as data from clients and truck drivers. Data from San Diego is collected through the "Primary Infection Research Consortium" (PIRC), the PWID study, as well as the STAHR and STAHR II studies and the South Bay Primary Infection Surveillance study [13].

The merged data set we are considering in this project consists of data from 1107 HIV-positive patients. However, since the data is stored separately for each study, in order to get access to these data to perform analysis across the sites, the data first needs to be integrated. Previous work in collaboration between the CSE department and AVRC enabled this integration [22]. Moreover, other work in progress [23], enable this data to be served as part of any web application.

Although we are analyzing data across various demographic, clinical, behavioral and social dimensions, including enrollment date, sexual orientation, drug use, etc (for a total of 70 attributes) we will focus most of our analysis around the concept of a "cluster" of HIV sequences. Each HIV-infected individual is identified by a unique sequence of his/her viral population, but those sequences can be grouped together in a cluster depending on their genetic distance.

Any of the available dimensions can be used to investigate the collected data, however for the aim of our project, together with our medical collaborators at the AVRC we decided to focus on the following data: PID (patient ID), SequenceID (sequence ID), SDTJ_015(a particular clustering of the data), PrimaryStudy (medical site name), EnrollmentDate (date that this patient enrolled in medical site), NumPartners (number of partners), Age, Sex, Race, Ethnicity, ZIP code of their residence, drug use, needle sharing, and sexual behavior. We included the other data for completeness, but we rarely interacted with them.

WORKFLOW

Our system is based on a back-end component including a database and an application server that acts as a data pre-processing server, and a front-end component that allows researchers to query the data in the back-end using a variety of time-based visualizations, such as a Motion Chart, and a Geo Chart. Figure 2 shows main components that are related to our system.

Back End Components

Database

Our main data source consists of a database that has been previously integrated from data retrieved and harmonized from the different sites [22]. Through the central database it is possible to view the data online and export it more easily. Figure 3 shows web interface of the database.

Data Pre-processing Server

A data pre-processing server powered by Ruby on Rails framework, sitting on top of the database, allows developer to filter and format the data according to the spec and serves it to front-end web app which

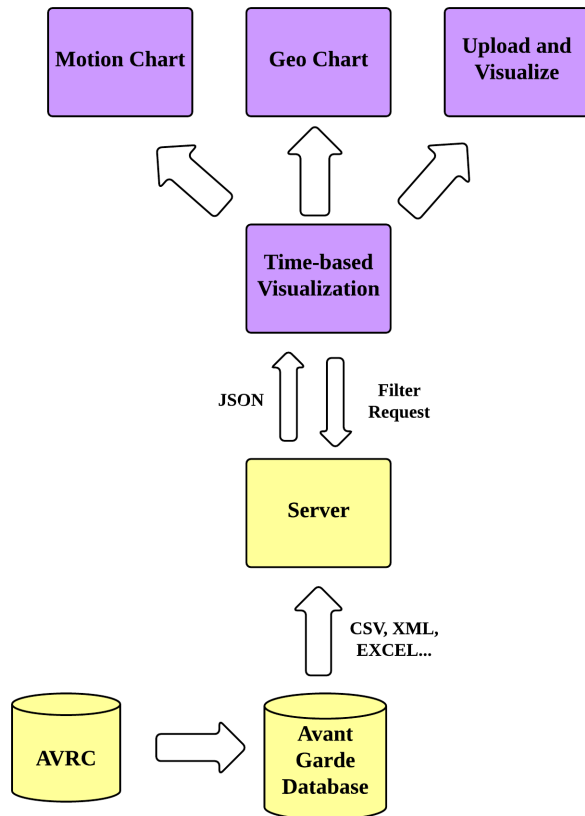


Figure 2. System Architecture. In yellow the back-end components, in purple the front-end.

can further be filtered and queried on the client side. This pre-processing service is built flexible enough that it can not only process the data from the existing databases, but it can also allow data uploading in excel or csv format, as well as taking third-party data sources in JSON and XML format. An example of a JSON data structure served by the pre-processing server is shown below.

```
1 [
2   {
3     "PID" : "50100119",
4     "SequenceID" : "SDHC|AEH020|050100119|1
5       0-25-1996",
6     "PrimaryStudyID" : "05-01-0011-9",
7     "PrimaryStudy" : "AEH020",
8     "EnrollmentDate" : "10/25/96",
9     "Age" : "16",
10    "Sex" : "Female",
11    "Race" : "White",
12    "Ethnicity" : "Unk",
13    "MaritalStatus" : "",
14    "SexualOrient" : "Het",
15    "NumPartners" : "1",
16    "DateCurr" : "10/25/96",
17    "ZIP" : "",
18    "State" : "",
19    "Country" : "US",
20    "City" : "",
21    "SDTJ_015" : "135",
22    "Latitude" : "",
23    "Longitude" : ""
24  },
25  ...
26  },
27  ...
28 ]
```

Front End Components

The app's front end is the most important part of the project, it needs to successfully retrieve the data and visualize them in the expected fashion plus support rich client-side interactivity to allow users to examine the data from different perspective. In particular, our project focuses on allowing the users to explore how HIV clusters change over time by leveraging and integrating interactive visualizations and dynamic time-lines. Through our front end app, the users can pick the data that falls in certain category at specific times and in specific geographical context. As we can see in Figure 4, there are two key modules in our front end app: (1) a motion chart module and (2) a geo chart module. In addition to that, our front end app also allows users to upload and visualize the local data files.

Export CSV		Visualize All							
Survey ID	Site Subject ID	Site Name	Date of Interview	Baseline Age (in years)	Question Duration (in months)	Gender	Marital Status	Sexual Orientation	
1	A025	Amigo	2011-07-11 00:00:00	60	4	Male	Common Law	Bisexual	
2	A053	Amigo	10/21/2011	36	4	Male	Married	Bisexual	
3	A188	Amigo	08/02/2012	27	4	Male	Never Married	Bisexual	
4	EC029	EICuete	04/06/2011	34	6	Female	Single	Heterosexual	
5	EC043	EICuete	04/08/2011	35	6	Male	Married	Heterosexual	
6	EC060	EICuete	04/11/2011	33	6	Male	Married	Heterosexual	
7	EC092	EICuete	04/13/2011	30	6	Male	Single	Heterosexual	
8	EC135	EICuete	04/19/2011	39	6	Male	Married	Heterosexual	
9	EC190	EICuete	04/26/2011	35	6	Male	Single	Heterosexual	
10	EC231	EICuete	05/26/2011	30	6	Male	Married	Heterosexual	
11	EC245	EICuete	06/01/2011	33	6	Male	Married	Heterosexual	
12	EC263	EICuete	06/07/2011	38	6	Male	Married	Heterosexual	
13	EC267	EICuete	06/28/2011	34	6	Female	Married	Heterosexual	
14	EC269	EICuete	07/04/2011	35	6	Female	Separated	Heterosexual	
15	EC344	EICuete	08/19/2011	45	6	Male	Single	Heterosexual	
16	EC345	EICuete	08/18/2011	29	6	Male	Single	Bisexual	
17	EC347	EICuete	08/24/2011	35	6	Male	Married	Heterosexual	
18	EC348	EICuete	08/24/2011	29	6	Female	Married	Heterosexual	
19	EC408	EICuete	09/20/2011	44	6	Male	Married	Heterosexual	
20	EC413	EICuete	09/21/2011	31	6	Male	Single	Heterosexual	

Figure 3. Database

Motion Chart

We use the motion chart and an associate timeline component to allow users to see how different indicators interact with each other over time. Users are able to mix-and-match data categories and create a variety of motion graphics within a single chart [9]. Users can also see sudden changes in the data (e.g. The first enrolled patient in any of the clusters) as time points visualized on the timeline. By clicking on one time point, users will see the corresponding data in the motion chart at that particular time. Users may also filter the data for example by groups and average age; our front end implementation will automatically render the corresponding data in the motion chart as

well as in the underlying data table. Figure 5 shows the layout of our motion chart visualization, including filters menu and the resulting data table.

Geo Chart

We also introduced a geo chart to help researchers to figure out how different groups distribute geographically at certain time and how their distribution changes over time.

A geo chart is rendered within the browser using SVG or VML [10]. The geographic location of different clusters are plotted using the latitude and longitude of the point at the center of the ZIP code. The size of circles in the chart denote the number of patients of that cluster in that area and different colors denote different clusters. With time changing, the distribution of clusters changes accordingly.

Users can click on the “play”, “pause” and “reset” button to start, pause and reset the time-based evolution of the geo chart. Additionally, with the time changing, the underlying rows in the data table will change accordingly. By dragging the time slider at certain time point, users will see the geographical distribution of the data at that time. When users hover over the cluttered marker, more information is shown. Finally, if users click on the marker visualized on the map, the corresponding row will be highlighted in the data table. Our geo chart also enables users to see annotations and filter data. Figure 6 shows a specific layout of the geo chart.

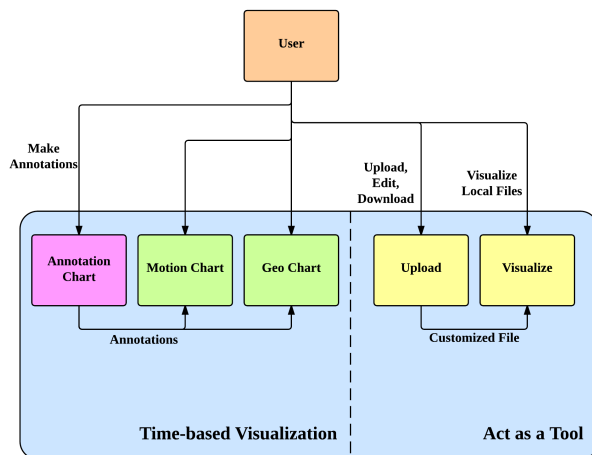


Figure 4. Modules

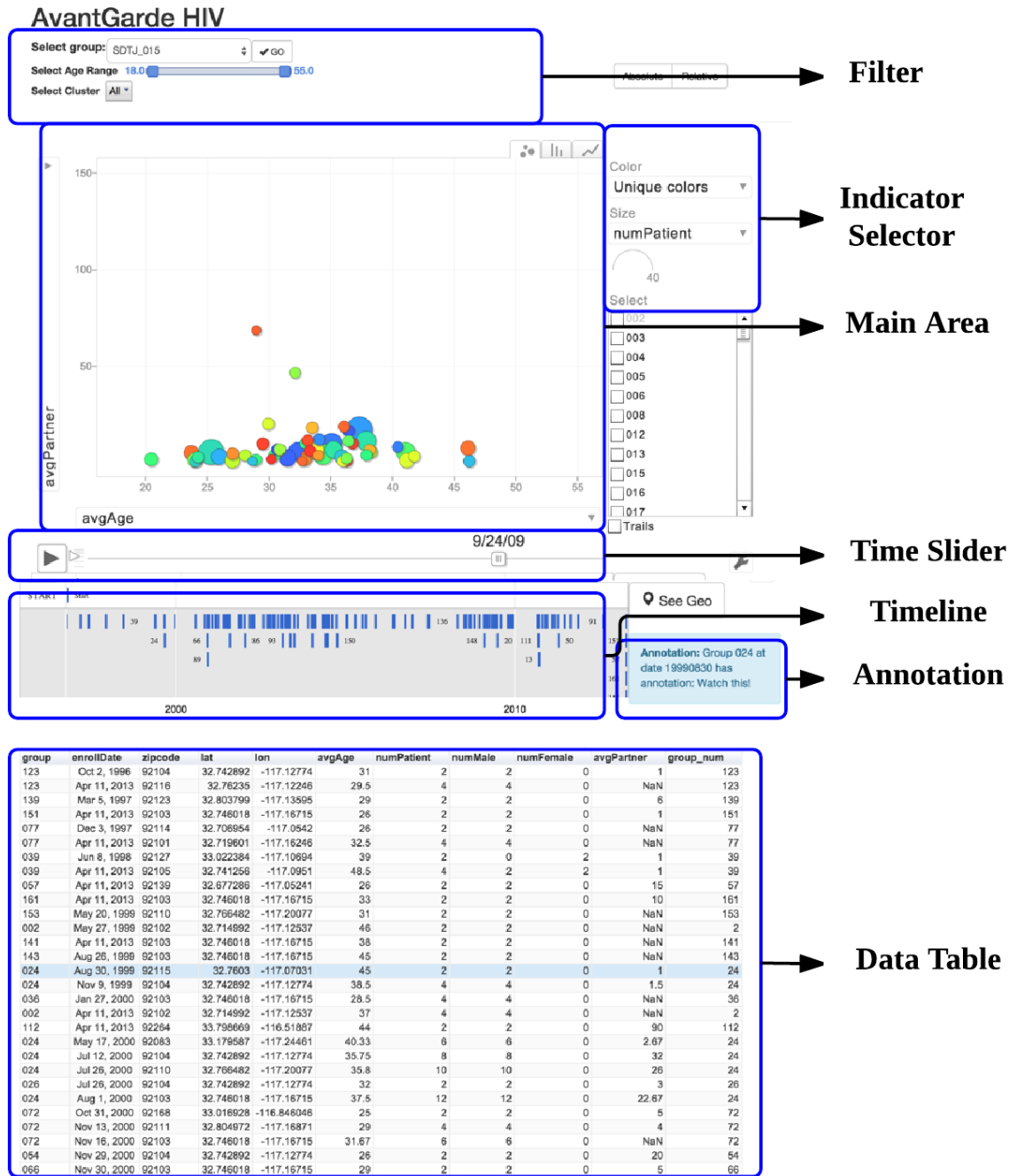


Figure 5. Motion Chart Layout. With the motion chart, users are able to filter data by “Filter”, select size and color indicators by “Indicator Selector”, select time by “Time Slider” and “Timeline” and see corresponding chart and data in “Main Area” and “Data Table”.

Upload and Visualize

In case users have additional data to analyze, our system allows users to upload local files, edit them online and visualize data using our time-based visualizations. The upload and visualize page are shown separately in Figure 7 and Figure 8.

Annotations

When users notice interesting data points in the chart, they might want to make annotations for future use. To accommodate this requirement, we implemented the annotation feature in our front end app and synchronized the annotation data across the motion chart as well as the geo chart. On the

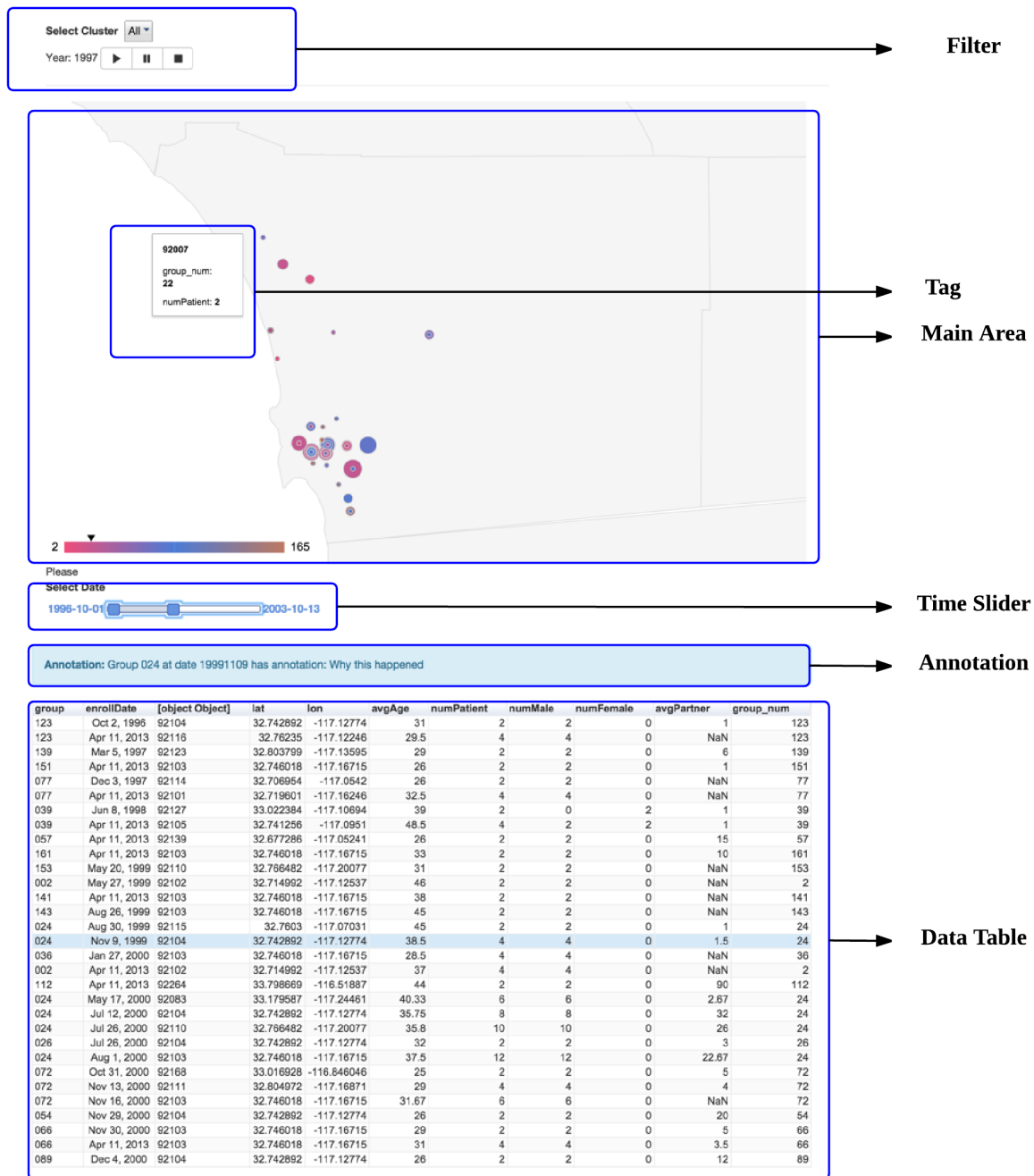


Figure 6. Geo Chart Layout. With geo chart, users are able to filter data by “Filter”, select time by “Time Slider” and see corresponding chart and data in “Main Area” and “Data Table”. A tag will be displayed when users hover over a marker in “Main Area”.

annotation page, by selecting a particular cluster, users can edit the annotations and our front end app will store it in user’s browser. On both motion chart and geo chart page where we have the result data table, when users click on one row in the table, they can see corresponding annotations on the page as well. The Annotation page is illustrated in Figure 9.

DESIGN AND IMPLEMENTATION

Since our goal was to investigate a prototype research tool, we did not receive a comprehensive and specific software spec in terms of what features or functionalities the researchers needed. Our approach to design was therefore based on explorations and prototyping, including conducting experiments against the dataset. The set of functionalities we developed in this project emerged step by step through



Figure 7. Upload Layout

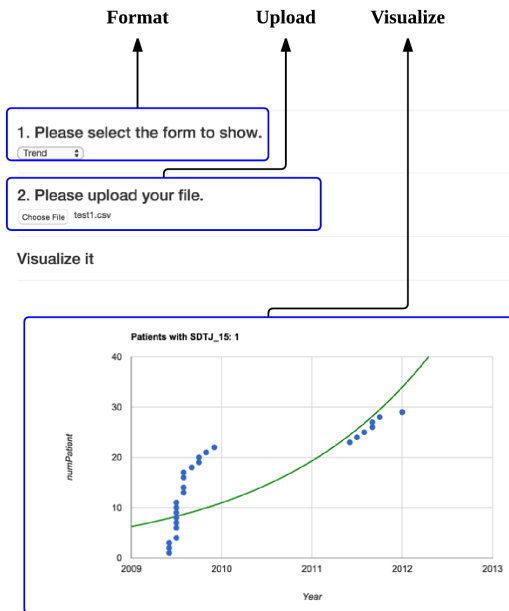


Figure 8. Visualize Layout

continuous interactions with the researchers at the UCSD medical center and Prof. Weibel. We performed weekly iterations on the tool and gradually refined the spec as well as the requirement down the road.

After our initial investigations, it became clear that researchers wanted to have a tool that supports data visualization examples that combined time-based analysis with other visualization implementations. For example, we wanted to develop our tool in such a way to allow the researchers to see how the patients data change in certain geographical locations while the time progress. Most of the existing toolkits or libraries can only visualize the data in one

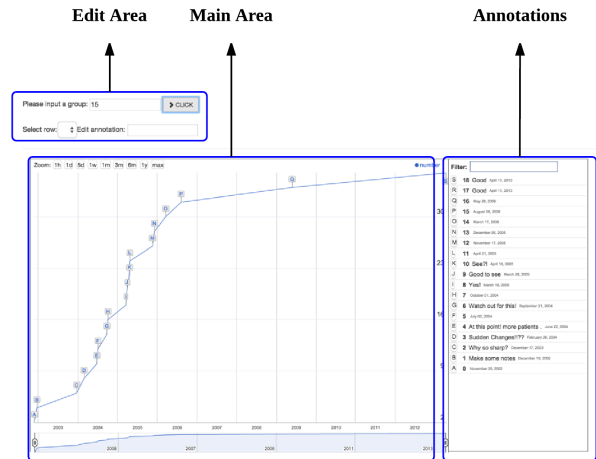


Figure 9. Annotation Chart Layout

single dimension, but we wanted to introduce a more flexible way to look at this data, combining and leveraging the unique power of the single visualizations.

Implementation

Software Development Methods

The nature of this project required an agile-like approach for the development and design process where developers can build certain prototyping functionality first, then immediately get the feedback and iterate based on that. Over the course of 6 months, We held regular weekly meetings with Prof. Weibel and maintained ongoing conversations with the medical center researchers in order to demonstrate and discuss the newly built features and visualizations and retain feedback to further progress towards our final tool. This development workflow helped us avoiding the implementation of unnecessary or erroneous features and ensured we were always in-sync with the researchers.

Data Pre-processing

The original raw datasets we worked with required a lot of pre-processing to be suitable for the development purpose. We implemented our back-end service to support continuous clean up and data formatting based on the rules and the spec from the researchers. After the pre-processing, data is made available by our data service to the front-end application through a RESTful API in a structured JSON format.

We now describe an example use case, for the time visualization. Each JSON object denotes a patient data and among all the data attributes, we want to include the SDTJ_15 column – which stands for a unique HIV transmission cluster id – to enable specific clustering functionality to occur in the front-end application.

Additionally, we also want to include the ZIP code attribute so we can visualize the patients that have the same cluster id but different geographical context. The described example would for example allow the researchers to investigate how the HIV gets transmitted across the San Diego and Tijuana border area. Below we show the resulting JSON structure for one record of this specific data query.

```

1  [
2    {
3      "cluster_id": "123",
4      "date": 19960902,
5      "city": "",
6      "numPatient": 2,
7      "numMale": 2,
8      "numFemale": 0,
9      "patients": [
10       "05-01-0004-8",
11       "05-01-0004-8"
12     ],
13     "totalAge": 62,
14     "totalPartners": 2,
15     "derivative": 0,
16     "lat": "32.742892",
17     "lon": "-117.12774",
18     "zip": "92104"
19   },
20   ...
21 ]

```

Data Visualization

This customizable data format allows us to accommodate all the 3rd party visualization toolkit that we have adopted. For the motion chart functionality, we combined Google motion chart as well as Google timeline chart to implement the rich interactive interface. For geographic visualization, we leveraged Google geochart [10] which allows us to render patient's geographical distribution and demonstrate how this progress over time.

For the data annotation and manual data uploading and processing, we adopted AngularJS [14] which enables us to add a number of rich functionality to our front-end app. Researchers can make changes and see them on the fly as a single page web application. With the help of "Angular UI Grid" [15] and "AngularJS Google Chart Tools directive" [16], we can achieve dynamic data binding and update when the user performs any actions on the client-side.

In addition to the Javascript frameworks mentioned above, we also leveraged the HTML5 local storage to allow users to save their annotations and configurations within their browser so that they could reuse this information in later sessions. Although the local

storage of all annotations as (cluster_id - date, annotation) pairs is effective, compared to the ideal server-side implementation, our browser local storage solution might not be able to save a large quantity of data and is not synchronized if the user switch to another computer or device. Figure 10 shows how the different data visualizations interact with the in-browser local storage.

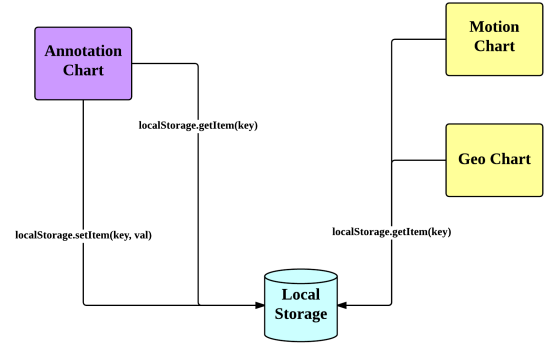


Figure 10. Interactions between visualization component with Local Storage

USE CASES

In the previous sections we illustrated the implementation of the main modules and functionalities of our interactive time-based visualization tool. However, in order to really demonstrate how researchers can apply our application in their real world studies or as a complementary tool for data analysis, we would like to present a major usage scenario and some concrete example use cases here.

Usage Scenario

Our general assumption is that data collected from patients have been time-stamped and all contain information about when the data has been stored in the data repository. Researchers want to see how the data progresses and changes over time from a variety of different perspectives. Our tool facilitates this analysis by enabling interactive time-based visualization of this data.

While investigating the data in terms of cluster groups a researcher notices an outlier cluster (id #15) that looks suspicious to him so he makes annotations on it. He then starts to look at the relationship between patients' average age, number of males individuals, and number of patients in different clusters over time. In the motion chart, he selects avgAge and avgMale as x-axis and y-axis respectively. He selects "unique color" as the option to color the currently visualized clusters and numPatients as the size indicator. After the researcher clicks on the "play" button of the motion chart and sees all the data

changing over time, he notices that cluster #15 has a dramatic increase in November 2004. In order to know more about the patients enrolled in cluster #15 at that time, he clicks on the corresponding time point in the timeline and the motion chart scrolls to that particular point and shows the corresponding information.

At this time, he would like to see how patients of cluster #15 are distributed geographically. He clicks on the “See Geo” button and the application redirects him to the geo chart page that show the cluster distribution at that particular time point. He now clicks on the “play” button on the geo chart page to see how the geographical distribution of clusters changes over time. With the help of our app, he is able to see that most patients in cluster #15 in 2004 were located in the San Diego area and after that time point, patients of cluster #15 started to spread to other places in the US and Mexico. From this observation, he can infer that 2004 might have been the starting time for a new spread of HIV patients from cluster #15 across the San Diego-Tijuana border area. By further investigate the available data he can now try to understand the reasons of this spread.

While the above scenario describes a plausible situation, to reinforce our point, we now introduce two more concrete use cases where medical researchers have applied our approach and tools, or plan to do so in the near future.

How IDU affects the growth of HIV infections

In this use case the researchers are interested in exploring the relationship between IDU (injecting drug users) and HIV patients, and specifically investigate whether IDU usage is a major contributing factor to HIV infection. The researchers start by setting IDU, heroin usage or methamphetamine usage as “color indicator”. This approach allows a researcher to pick

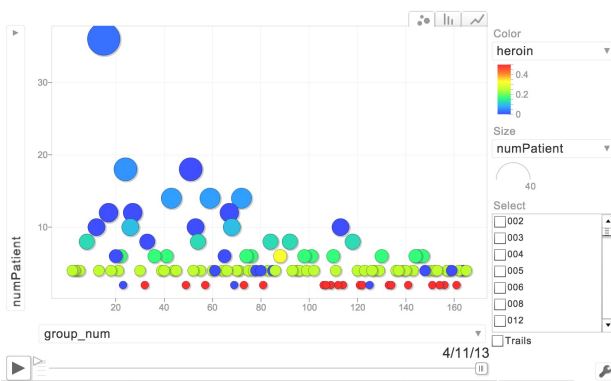


Figure 11. Motion Chart with Heroin Usage as Color Indicator

heroin usage as color indicator to see how the heroin usage affected the overall spread and growth of the HIV patients over time. When looking at the chart, he notices that with the time progressing, some clusters with small number of patients have a higher heroin usage while some large clusters have a lower heroin usage (see Figure 11). This will enable him to further explore the relationship between different patient clusters and the drug usage.

How policy change affects cross-border HIV transmission

In 2001, a new border-control policy regarding the San Diego/Tijuana border came into practice which resulted in a more strict border crossing check and scrutiny. By leveraging our time-based visualization, we can see the change of HIV patients distribution from this policy. We use our motion chart and use “country” as color indicator.

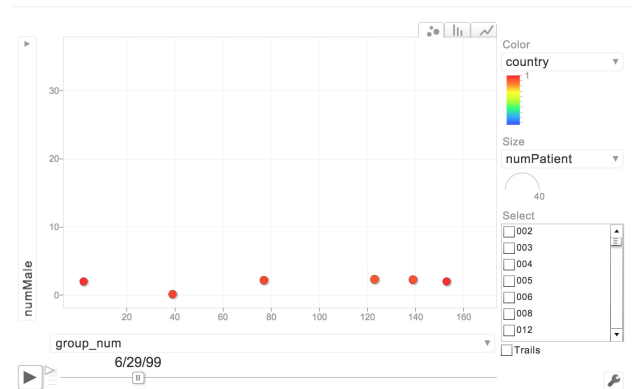


Figure 12. Motion Chart with Country as Color Indicator before 2001(June, 1999)

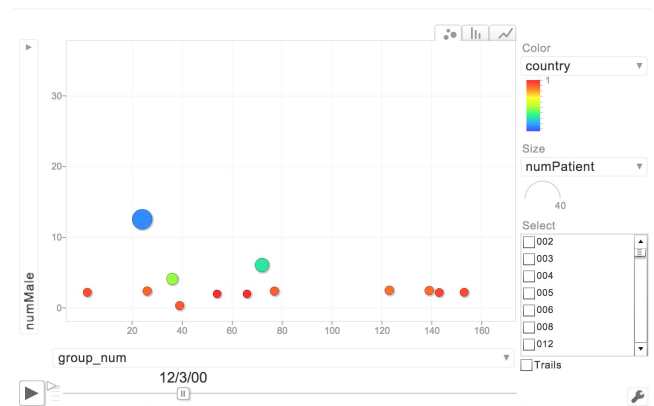


Figure 13. Motion Chart with Country as Color Indicator before 2001(Dec., 2000)

Before 2001 (Figure 12 and 13) the colors changed significantly, which demonstrates how easily patients could travel across the border back and forth. However, when we look at the data after 2001 (Figure 14 and 15) we can see there is a significant

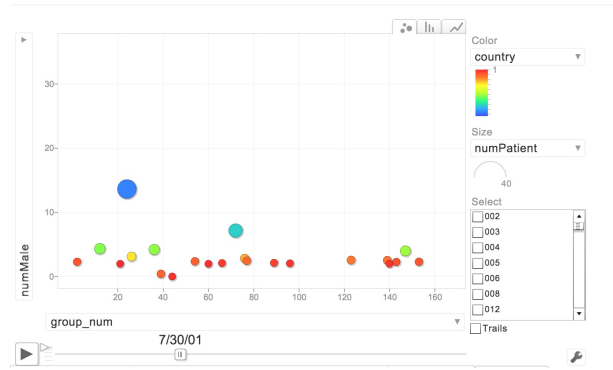


Figure 14. Motion Chart with Country as Color Indicator after 2001(July, 2001)

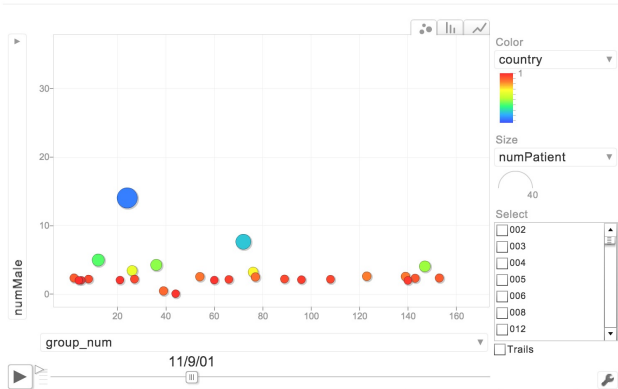


Figure 15. Motion Chart with Country as Color Indicator after 2001(Nov., 2001)

decrease in terms of the variation of patients cluster, which means the border control policy actually helped reducing the spread and the transmission of virus across the border. This is informative for both medical researchers as well as government officials.

EVALUATION

Evaluating the iterative and user-centered design of a system like our interactive time-based visualization tool, is quite different from establishing the success of projects that either have a well-adopted benchmark to run against, or a numeric way to measure and evaluate the performance. Our evaluation was mainly based on continuous deployment of our refined prototype and its use by the medical researchers we collaborated with. Across the six months of iterative development, we engaged in a variety of meetings and discussion interviews where we evaluated our approach and prototype with several HIV researchers and staff responsible for collecting and analyzing data.

To validate our approach we outline below three key events that were part of our evaluation and describe in detail how these helped with the further development of our tool.

1. Initial Hypothesis Evaluation

In the early stage of our project, we met with a group of five HIV researchers and applied the design thinking methodology [24] to brainstorm ideas and features the researchers would like to have in the tool. The meeting focused in particular on evaluating different time-based visualization approaches we were proposing.

During the meeting, we demonstrated our ongoing project and the researchers responded with a number of “I wish” and “I like” kind of feedback. They specifically focused on what data they could provide and how they would like it to be visualized. For instance, they suggested that we should allow the users to group the cluster in different ways to have more flexibility instead of just clustering by virus sequence ID. They also suggested to integrate the annotation feature into the visualization page so that the users can seamlessly see the annotations on both geo and motion chart.

2. Research Presentation

One of our medical researchers collaborators, Dr. Mehta, demonstrated the tool in a research presentation in front of 50 medical researchers at the UCSD Medical Center - Hillcrest. During his presentation, he used our time-based visualization tool to show how the number of patients, and the average age of patients in different clusters progressed over time across the San Diego/Tijuana border.

The feedback about his demonstration was quite positive and a number of researchers at the meeting expressed their interest in using the visualization tool in their own research and agreed that this approach would help them to understand their data better.

At the same time they also raised up a number of questions related to the prototype and provided feedback for improvement. For instance, in terms of the motion chart, they mentioned that the indicators of color in the chart were not legible and could be better illustrated.

3. Discussions Interview

In order to get additional feedback for our prototype we held several rounds of iterations and discussion interviews with the medical researchers, in particular Dr. Mehta.

Dr. Mehta conveyed the idea that our tool could satisfy most of his current needs but it would be helpful to combine this tool into other toolkit that he’s been using in research. He also gave several suggestions such as enriching the data displayed in the information window on the geo chart.

Towards the end of the project we held another discussion interview with Dr. Mehta to evaluate our work including pros and cons as well as discuss future work. We laid out specific use cases for which our tool has been used or could be used. Additionally we evaluated different approaches to further combine our tool with other visualization approaches.

After each round of evaluation of researchers against our tool, we gradually clarified our implementation goals and the spec. The feedback for our prototype tool was in general positive and many additional improvements were suggested and discussed.

CONCLUSION

In this project we proposed a novel web-based interactive visualization tool to augment the HIV molecular epidemiology study based on clusters of HIV infected patients, with a time component. Specifically, our approach enables explorations of how HIV clusters change over time by integrating interactive visualizations and dynamic timelines that can be operated to show data at specific times and in specific geographical context so that researchers can explore in real-time many facets of the local transmission network [19].

According to the evaluation, the researchers and clinicians are now able to explore novel insights and understanding on the local impact of HIV incidence with this tool.

Future Works

One of our major future works is to explore how we can test and evaluate our visualization tool in an objective manner. We plan to present the prototype to more potential users namely researchers in a variety of medical sites, working with them closely to gain more insights into how to make our system even better.

Additionally, we would also like to investigate how to automatically leverage new data context related to a new set of data. For instance, in the future data will cover also treatment related information showing patients' medications and therapy. We envision researchers to utilize our tool to observe the change of growth in clusters with different treatments so as to figure out how effective the treatments are. We would also want to build a more flexible and robust data pre-processing system to be able to interface with other data sources with more complicated data format.

Finally, we would like to explore other potential ways to visualize the data in a time-based fashion and to see how we can integrate other data visualizations approaches effectively. Our group is already working on a Rich Interactive Data Visualization Tool for HIV

Epidemiology [18, 23] and we plan to integrate this new tool and the presented project into the Avant-Garde platform of the AVRC.

ACKNOWLEDGEMENTS

I would like to especially thank Prof. Nadir Weibel for the continuous help, instructions and feedback. I would like to thank also Dr. Sanjay Mehta for helping us choose relevant data fields, explaining workflow process for HIV researchers and providing feedback. At the same time, I would like to thank Sandy Law for the flexible access to the Avant-Garde database and Yu Xia for his effort on the back-end as well as for providing reliable server support.

REFERENCES

1. Weiss, Robin A. "How does HIV cause AIDS?." *Science* 260.5112 (1993): 1273-1279.
2. Douek, Daniel C., Mario Roederer, and Richard A. Koup. "Emerging concepts in the immunopathogenesis of AIDS." *Annual review of medicine* 60 (2009): 471.
3. UNAIDS report. <http://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2014/july/20140716prgapreport>.
4. Sandy Law. Avant-Garde HIV Research: Harmonizing and Visualizing Patient Data.
5. Avant-garde award. <http://www.drugabuse.gov/about-nida/organization/offices/office-nida-director-od/aids-research-program-arp/avant-garde-award-hivaids-research>.
6. Dr. Davey Smith Receives a 2012 NIDA Avant-Garde Award. <https://www.charterresource.ucsd.edu>.
7. Avant-Garde Program for HIV/AIDS Research. <http://cfar.ucsd.edu/cfar-grants/extramural-grants/avant-garde-program-for-hiv-aids-research>.
8. The UC San Diego AntiViral Research Center (AVRC). <http://avrc.ucsd.edu/>.
9. Motion chart. <https://developers.google.com/chart/interactive/docs/gallery/motionchart>.
10. Geochart. <https://developers.google.com/chart/interactive/docs/gallery/geochart>.
11. How the Drought is Shrinking California's Reservoirs [Visualization]. <http://blogs.kqed.org/lowdown/2014/03/18/into-the-drought-californias-shrinking-reservoirs/>.
12. Design thinking. <https://chat2lrn.wordpress.com/2013/09/17/design-thinking-what-does-it-mean-to-ld/>.
13. Sanjay Mehta. Molecular Epidemiology of HIV in San Diego and Tijuana. <http://cfar.ucsd.edu/links/downloads/disparities-in-hiv-care/MehtaCFARDisparities2.pdf>.

14. AngularJS. <https://angularjs.org/>.
15. Angular UI Grid. <http://ui-grid.info/>.
16. AngularJs Google Chart Tools directive..
<http://bouil.github.io/angular-google-chart/#/fat>.
17. The past, present & future of local storage for web applications.
<http://diveintohtml5.info/storage.html>.
18. Data-Driven Documents. <http://d3js.org>.
19. Nadir Weibel. Dev Grant Research Plan..
20. Little, Susan J., et al. "Using HIV networks to inform real time prevention interventions." PloS one 9.6 (2014): e98443.
21. Tukey, John W. "Exploratory data analysis." Reading, Ma 231 (1977): 32.
22. Sandy Law. "Avant-Garde HIV Research: Harmonizing and Visualizing Patient Data.".
23. Yu Xia. "Rich Interactive Data Visualization Tool for HIV Epidemiology". work-in-progress.
24. Design Thinking.
http://en.wikipedia.org/wiki/Design_thinking.