Next Generation Time-based Visualization Tools for HIV Epidemiology

Yingyan Hua
Advised by: Nadir Weibel

Outline

1. Background and Goal
2. Design, Features and Use Cases
3. Evaluation
4. Conclusion and Future Work

Outline

1. **Background and Goal**
1) Background
2) Challenges
3) Solutions

2. Design, Features and Use Cases

3. Evaluation

4. Conclusion and Future Work

Background: Avant-Garde HIV

1. **HIV (Human Immunodeficiency Virus)**
   1) Around 35 million people are living with HIV
   2) Complex factors(clinical & social) influencing transmission

2. **“Avant-Garde” Award - HIV research**
   1) Concentrate on HIV treatment and prevention
   2) Track and analyze HIV transmission networks

3. **AVRC (Antiviral Research Center) - Data Collection**
   1) UCSD AVRC and medical sites in San Diego (US) and Tijuana (Mexico) area are conducting studies collecting HIV patient data(clinical, demographic and lifestyle)

HIV in San Diego - Tijuana Border

states with high rates of migration were also the states in which higher number of HIV+ individuals had previously lived in the US.

- 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.

- In 2000 the rate dropped to 12%.

Challenges

1. There can be some internal relationships between different factors behind the HIV migration, but how can we find them?

2. No existing solution allows researchers to explore and analyse the data through Molecular Epidemiology* methods.

* Molecular Epidemiology: Using the relationships between viral sequence data sampled from infected individuals, researchers will be able to infer and characterize the spread of HIV.

Our Solution: Time-based Visualization

A Time-based HIV visualization tool that can help answer:

1) How different dimensions of the available data interact with each other
2) How geographical distribution of patients clusters changes over time

Outline

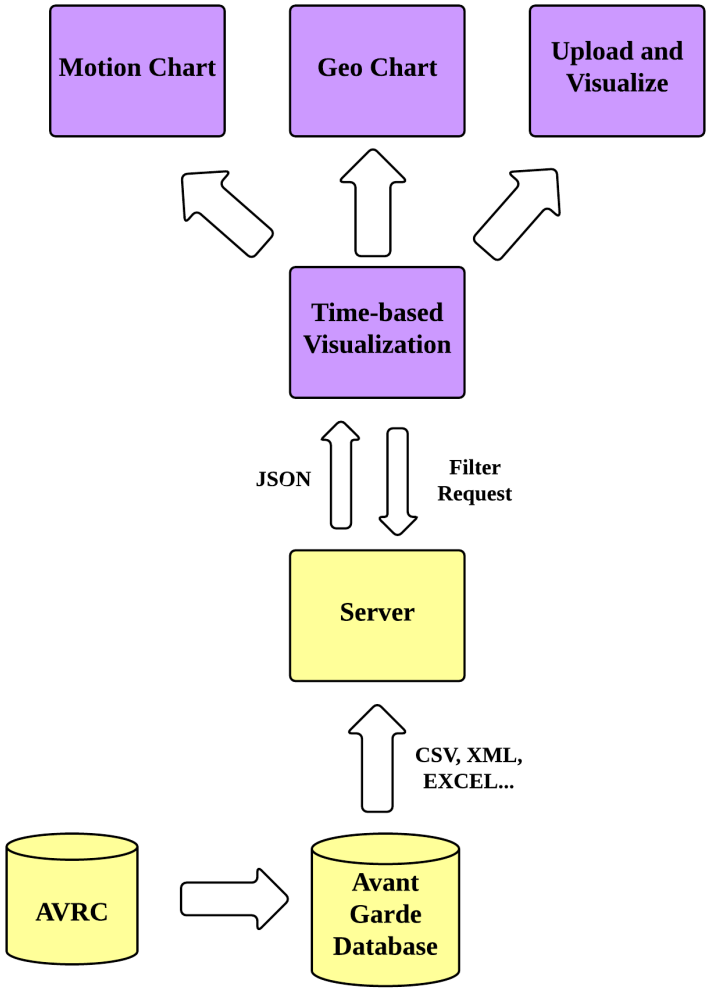
1. Background and Goal

2. **Design, Features and Use Cases**
1) System Architecture
2) Modules
3) Functionalities
4) Use Cases

3. Evaluation

4. Conclusion and Future Work

System Architecture



- AVRC:
HIV patient data in distributed medical sites

- Avant-Garde Database:
harmonize and integrate data from medical sites

- Server:
filter and format the data from database

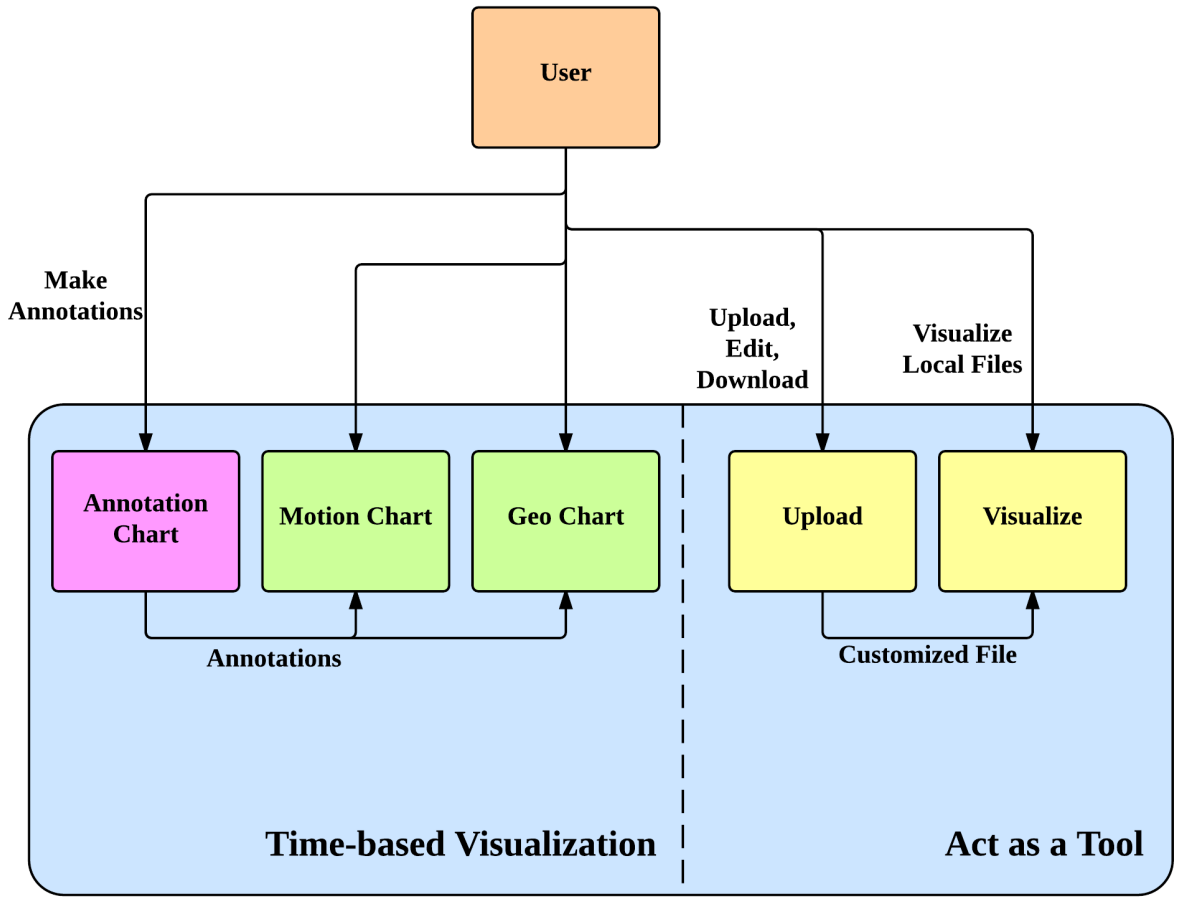
- Front-end:
time-based visualization

Data Structure

Here is a simple example of the data format.

```
[
  {
    "cluster_id":"123",
    "date":19960902,
    "city":"",
    "numPatient":2,
    "numMale":2,
    "numFemale":0,
    "patients":[
      "05-01-0004-8",
      "05-01-0004-8"
    ],
    "totalAge":62,
    "totalPartners":2,
    "derivative":0,
    "lat":"32.742892",
    "lon":"-117.12774",
    "zip":"92104"
  },
...
]
```

Front End Modules



- Motion chart:
How different dimensions of data interact with each other over time.

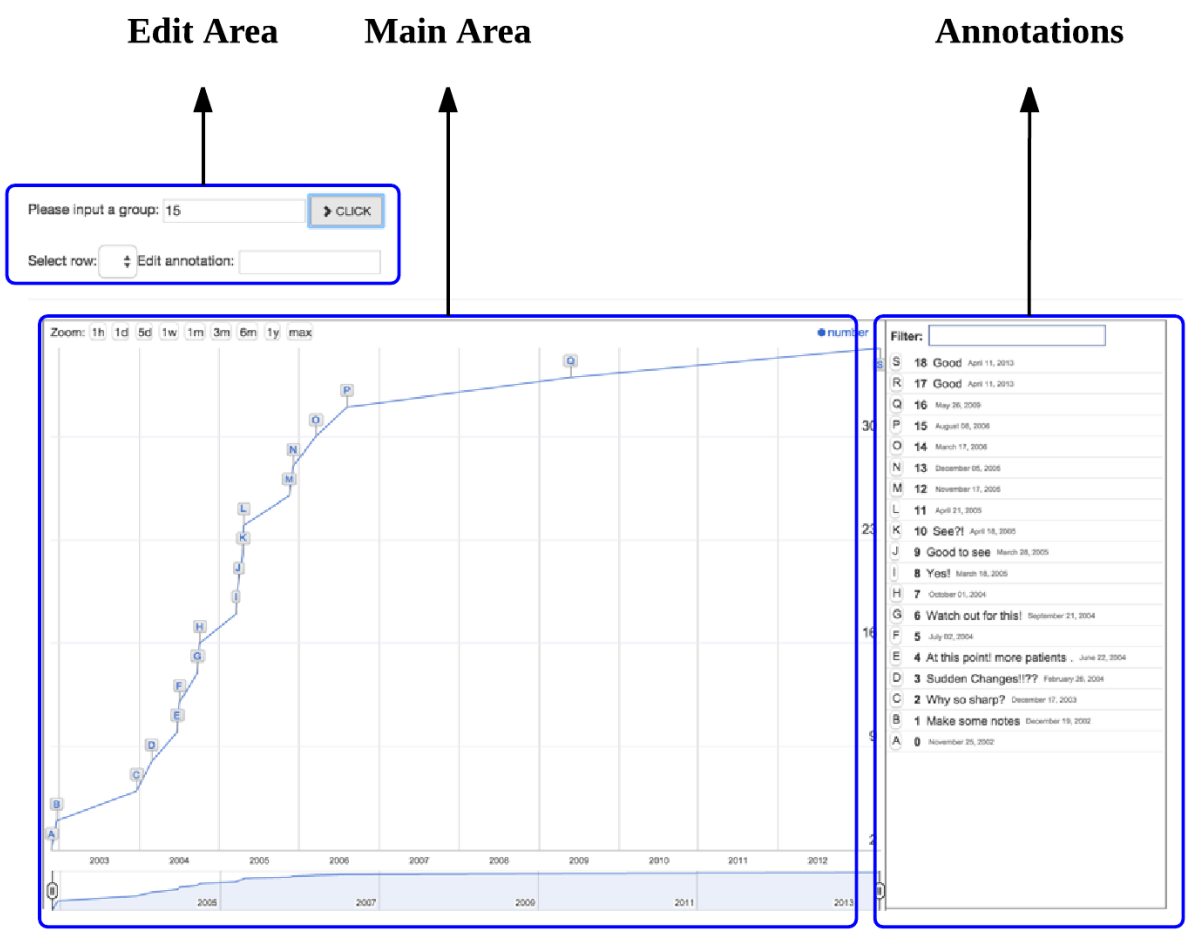
- Geo chart:
How geo distribution of clusters change over time.

- Customized data upload & visualize

Part0. Annotation

- Annotations for motion chart and geo chart

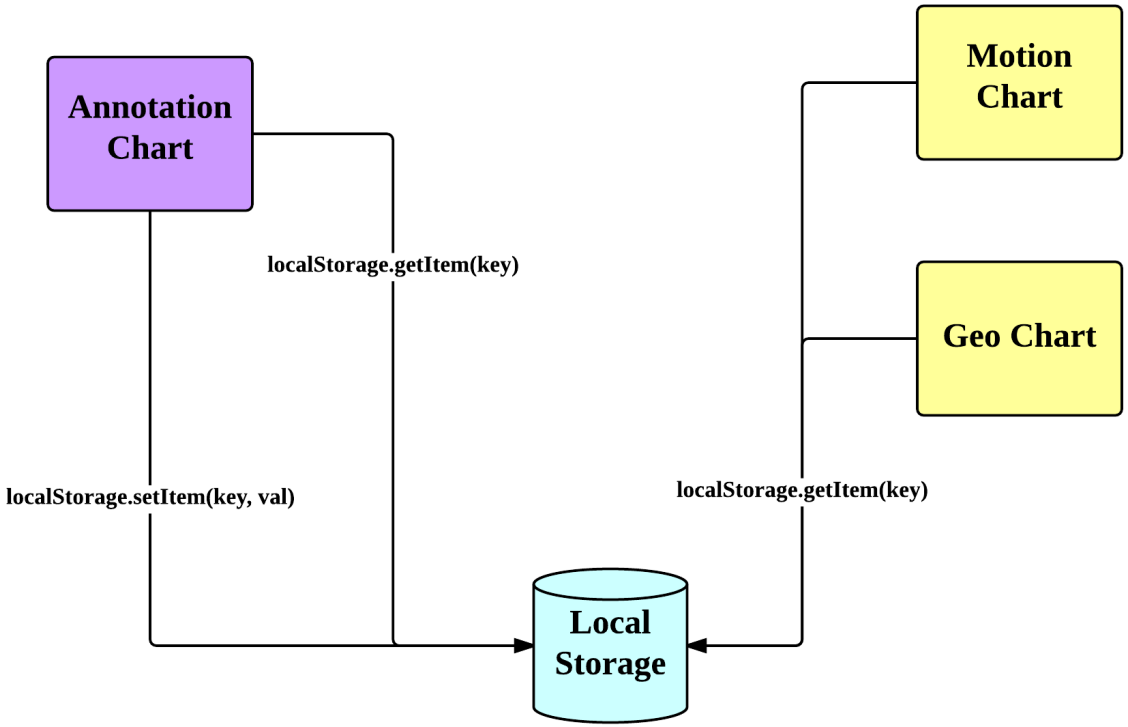
Annotation Chart Layout



- Show line for each cluster
- View/insert/edit annotations

<u>Problem: how to store them?</u>

Solution: Local Storage



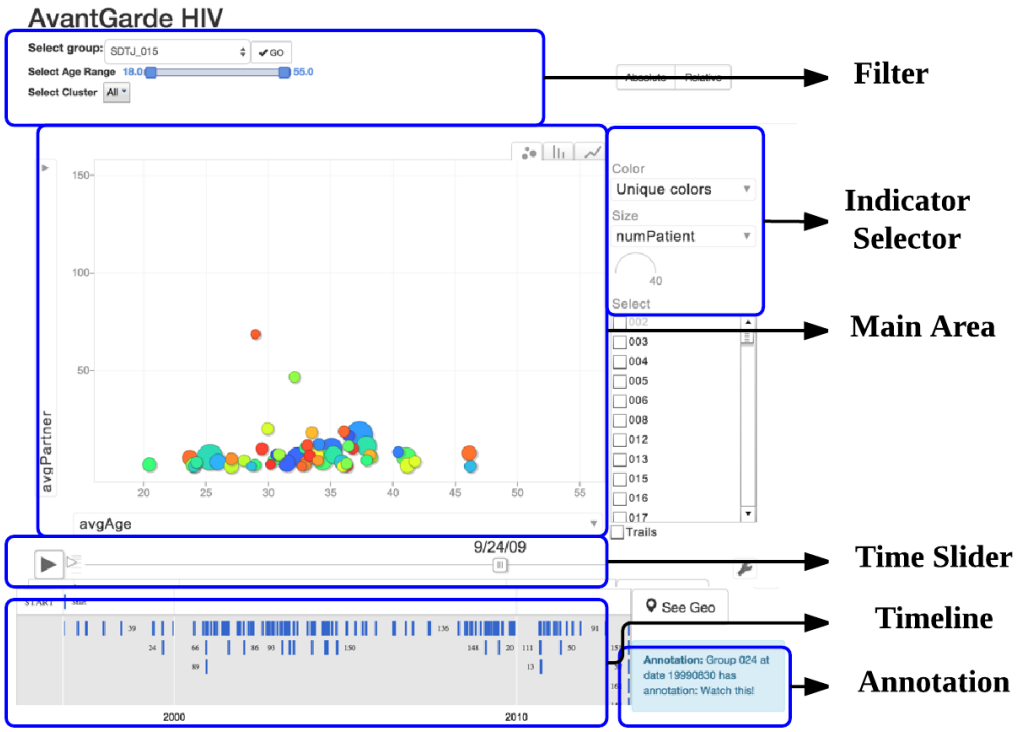
- Large storage space
- On the client side
- Persists beyond a page refresh
- Is not transmitted to the server

Part1. Motion chart

- How different indicators interact with each other over time

- Sudden changes in timeline

Motion Chart Layout

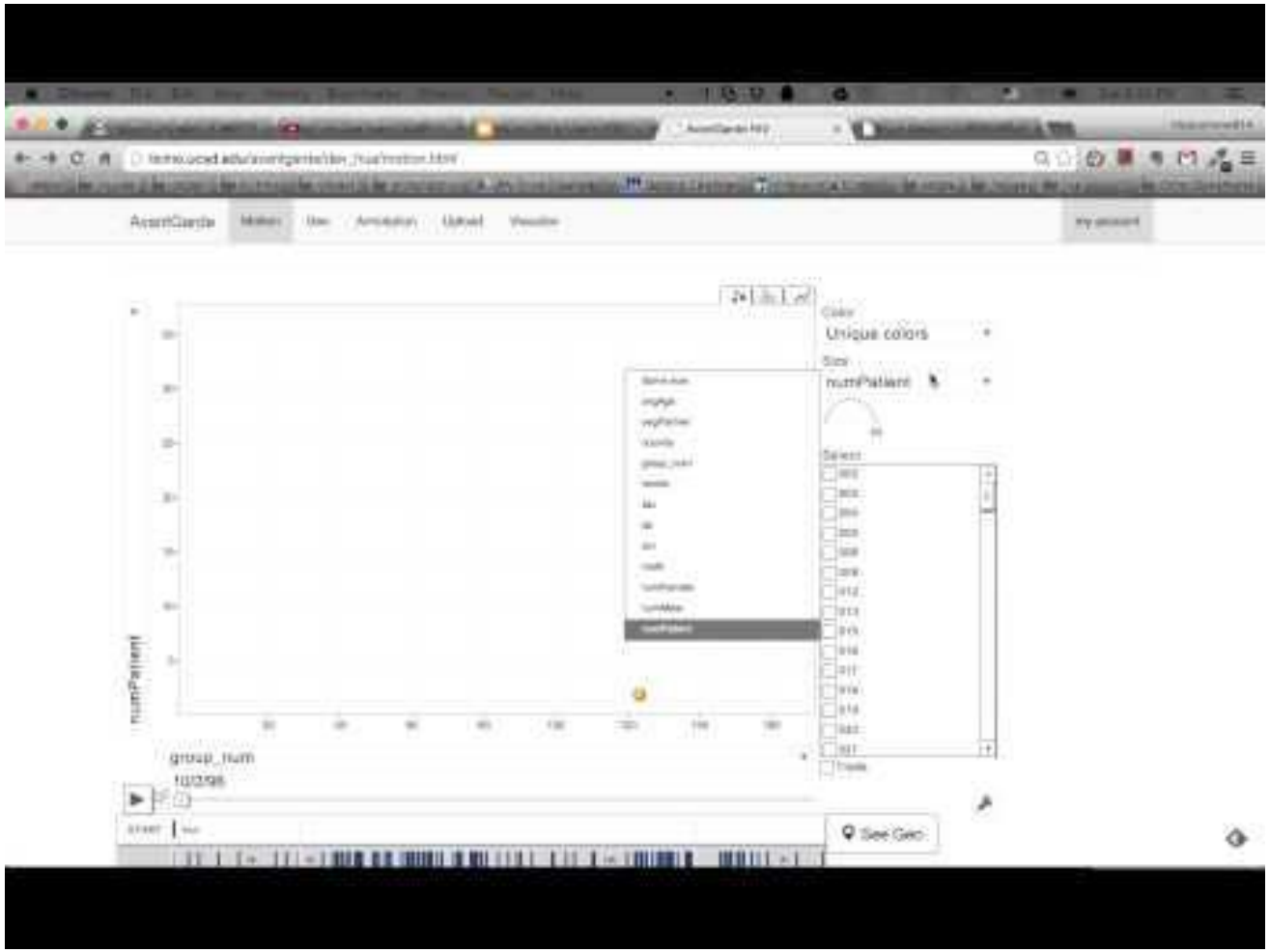


group	enrollDate	zipcode	lat	lon	avgAge	numPatient	numMale	numFemale	avgPartner	group_num
123	Oct 2, 1996	92104	32.742892	-117.12774	31	2	2	0	1	123
123	Apr 11, 2013	92116	32.76235	-117.12246	29.5	4	4	0	NaN	123
139	Mar 5, 1997	92123	32.803799	-117.13595	29	2	2	0	6	139
151	Apr 11, 2013	92103	32.746018	-117.16715	26	2	2	0	1	151
077	Dec 3, 1997	92114	32.706954	-117.0542	26	2	2	0	NaN	77
077	Apr 11, 2013	92101	32.719601	-117.16246	32.5	4	4	0	NaN	77
039	Jun 8, 1998	92127	33.022384	-117.10694	39	2	0	2	1	39
039	Apr 11, 2013	92105	32.741256	-117.0951	48.5	4	2	2	1	39
057	Apr 11, 2013	92139	32.677286	-117.05241	26	2	2	0	15	57
161	Apr 11, 2013	92103	32.746018	-117.16715	33	2	2	0	10	161
153	May 20, 1999	92110	32.766482	-117.20077	31	2	2	0	NaN	153
002	May 27, 1998	92102	32.714892	-117.12537	46	2	2	0	NaN	2
141	Apr 11, 2013	92103	32.746018	-117.16715	38	2	2	0	NaN	141
143	Aug 26, 1999	92103	32.746018	-117.16715	45	2	2	0	NaN	143
024	Aug 30, 1999	92115	32.7603	-117.07031	45	2	2	0	1	24
024	Nov 9, 1999	92104	32.742892	-117.12774	38.5	4	4	0	1.5	24
036	Jan 27, 2000	92103	32.746018	-117.16715	28.5	4	4	0	NaN	36
002	Apr 11, 2013	92102	32.714992	-117.12537	37	4	4	0	NaN	2
112	Apr 11, 2013	92264	33.798669	-116.51887	44	2	2	0	90	112
024	May 17, 2000	92083	33.179587	-117.24461	40.33	6	6	0	2.67	24
024	Jul 12, 2000	92104	32.742892	-117.12774	35.75	8	8	0	32	24
024	Jul 26, 2000	92110	32.766482	-117.20077	35.8	10	10	0	26	24
026	Jul 26, 2000	92104	32.742892	-117.12774	32	2	2	0	3	26
024	Aug 1, 2000	92103	32.746018	-117.16715	37.5	12	12	0	22.67	24
072	Oct 31, 2000	92168	33.016928	-116.846046	25	2	2	0	5	72
072	Nov 13, 2000	92111	32.804972	-117.16871	29	4	4	0	4	72

Data Table

- **Filter data**
- **Select indicator**
- **Play**
- **Show sudden changes**
- **Show annotations**

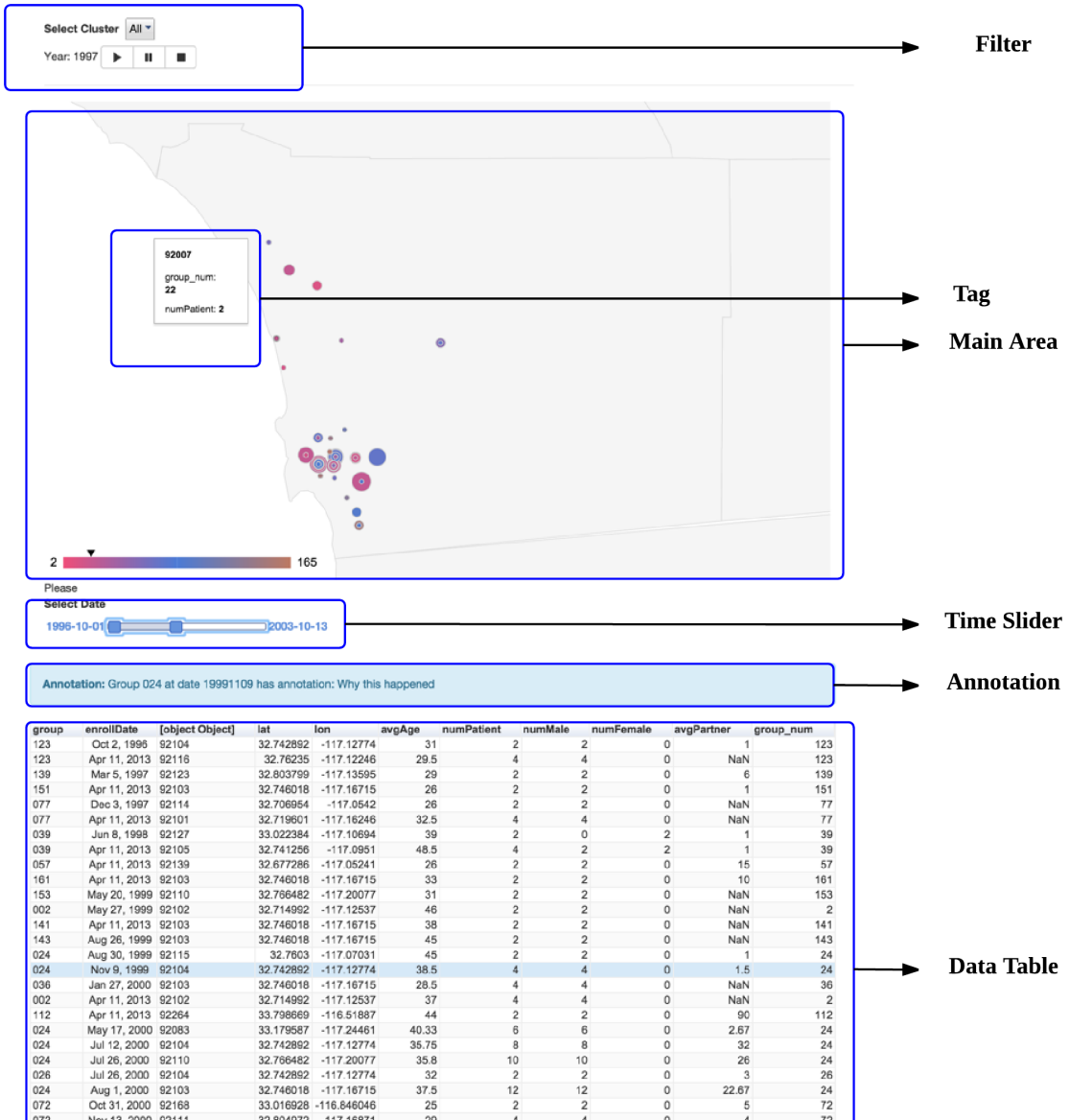
Use Case - How heroin usage affects HIV growth



Part2. Geo Chart

- How geographical distribution of clusters changes over time

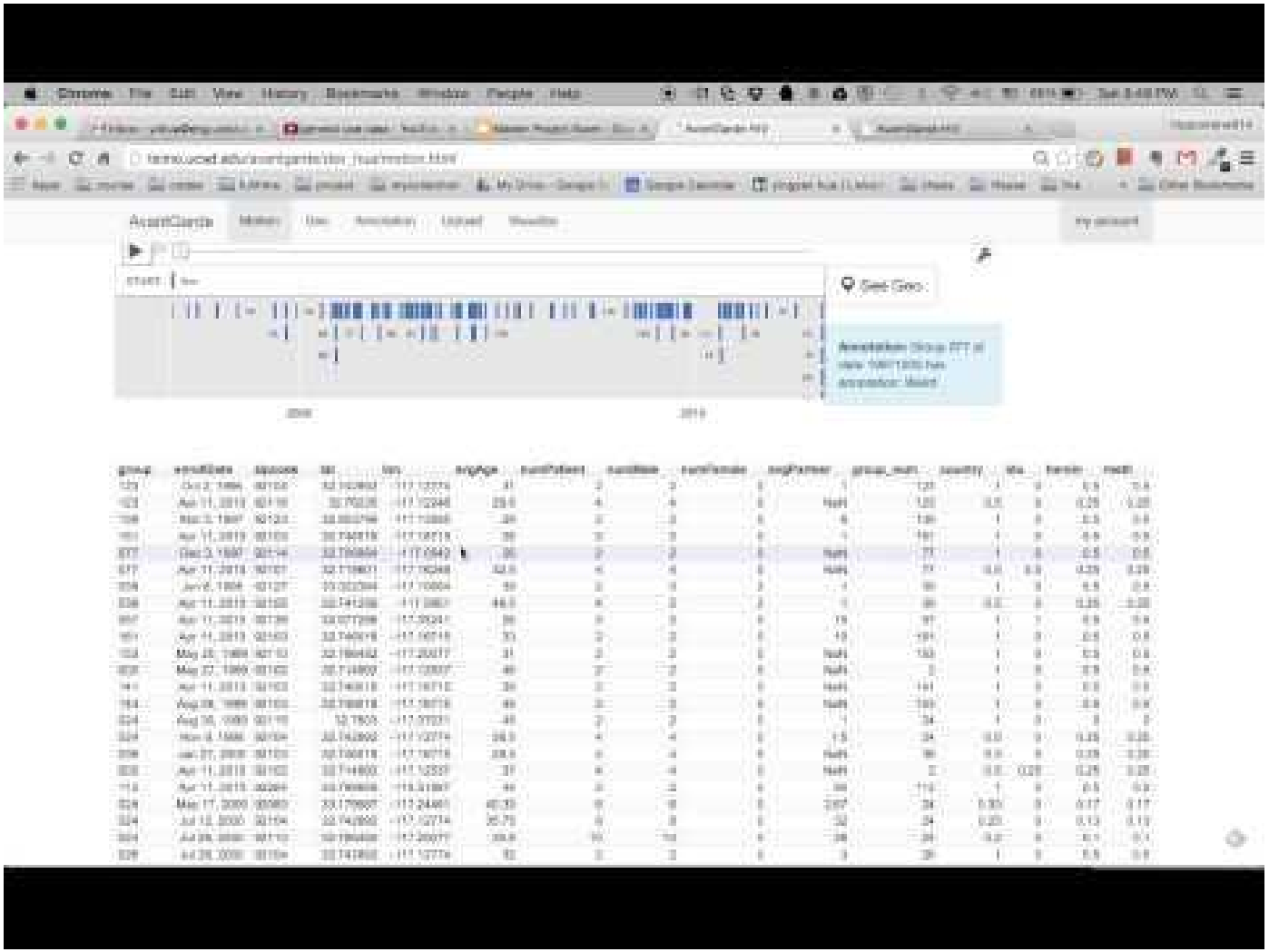
Geo Chart Layout



group	enrollDate	[object Object]	lat	lon	avgAge	numPatient	numMale	numFemale	avgPartner	group_num
123	Oct 2, 1995	92104	32.742892	-117.12774	31	2	2	0	1	123
123	Apr 11, 2013	92116	32.76235	-117.12246	29.5	4	4	0	NaN	123
139	Mar 5, 1997	92123	32.803799	-117.13595	29	2	2	0	6	139
151	Apr 11, 2013	92103	32.746018	-117.16715	26	2	2	0	1	151
077	Dec 3, 1997	92114	32.706954	-117.0542	26	2	2	0	NaN	77
077	Apr 11, 2013	92101	32.719601	-117.16246	32.5	4	4	0	NaN	77
039	Jun 8, 1998	92127	33.022384	-117.10694	39	2	0	2	1	39
039	Apr 11, 2013	92105	32.741256	-117.0951	48.5	4	2	2	1	39
057	Apr 11, 2013	92139	32.677286	-117.05241	26	2	2	0	15	57
161	Apr 11, 2013	92103	32.746018	-117.16715	33	2	2	0	10	161
153	May 20, 1999	92110	32.766482	-117.20077	31	2	2	0	NaN	153
002	May 27, 1999	92102	32.714992	-117.12537	46	2	2	0	NaN	2
141	Apr 11, 2013	92103	32.746018	-117.16715	38	2	2	0	NaN	141
143	Aug 26, 1999	92103	32.746018	-117.16715	45	2	2	0	NaN	143
024	Aug 30, 1999	92115	32.7603	-117.07031	45	2	2	0	1	24
024	Nov 9, 1999	92104	32.742892	-117.12774	38.5	4	4	0	1.5	24
036	Jan 27, 2000	92103	32.746018	-117.16715	28.5	4	4	0	NaN	36
002	Apr 11, 2013	92102	32.714992	-117.12537	37	4	4	0	NaN	2
112	Apr 11, 2013	92264	33.798669	-116.51887	44	2	2	0	90	112
024	May 17, 2000	92083	33.179587	-117.24461	40.33	6	6	0	2.67	24
024	Jul 12, 2000	92104	32.742892	-117.12774	35.75	8	8	0	32	24
024	Jul 26, 2000	92110	32.766482	-117.20077	35.8	10	10	0	26	24
026	Jul 26, 2000	92104	32.742892	-117.12774	32	2	2	0	3	26
024	Aug 1, 2000	92103	32.746018	-117.16715	37.5	12	12	0	22.67	24
072	Oct 31, 2000	92168	33.016928	-116.846046	25	2	2	0	5	72

- **Filter data**
- **Play/stop/reset**
- **Show annotations**

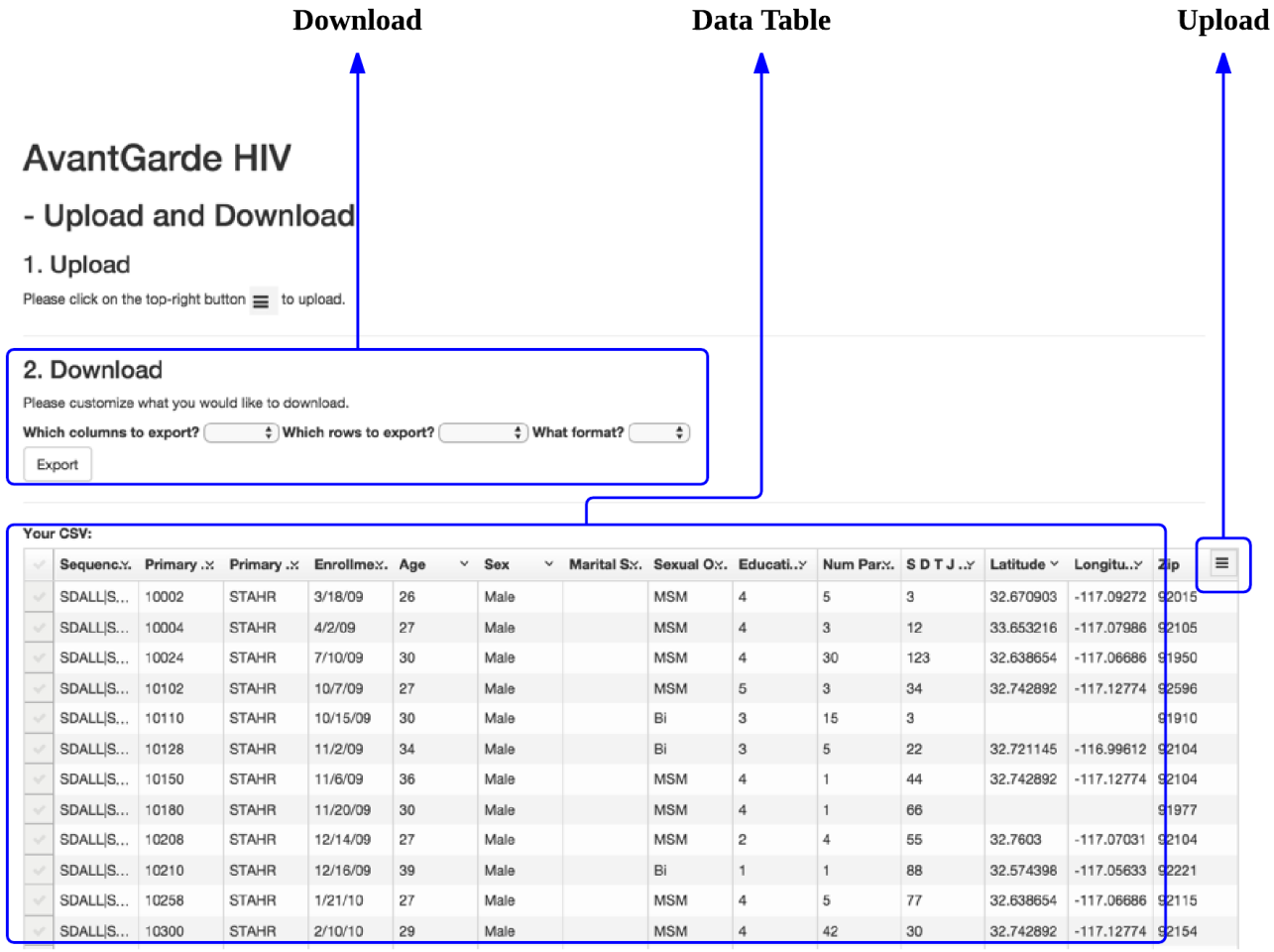
Use Case - General usage scenario



Part3. Upload and Visualize

- Act as a tool for users to upload, edit, download and visualize local files

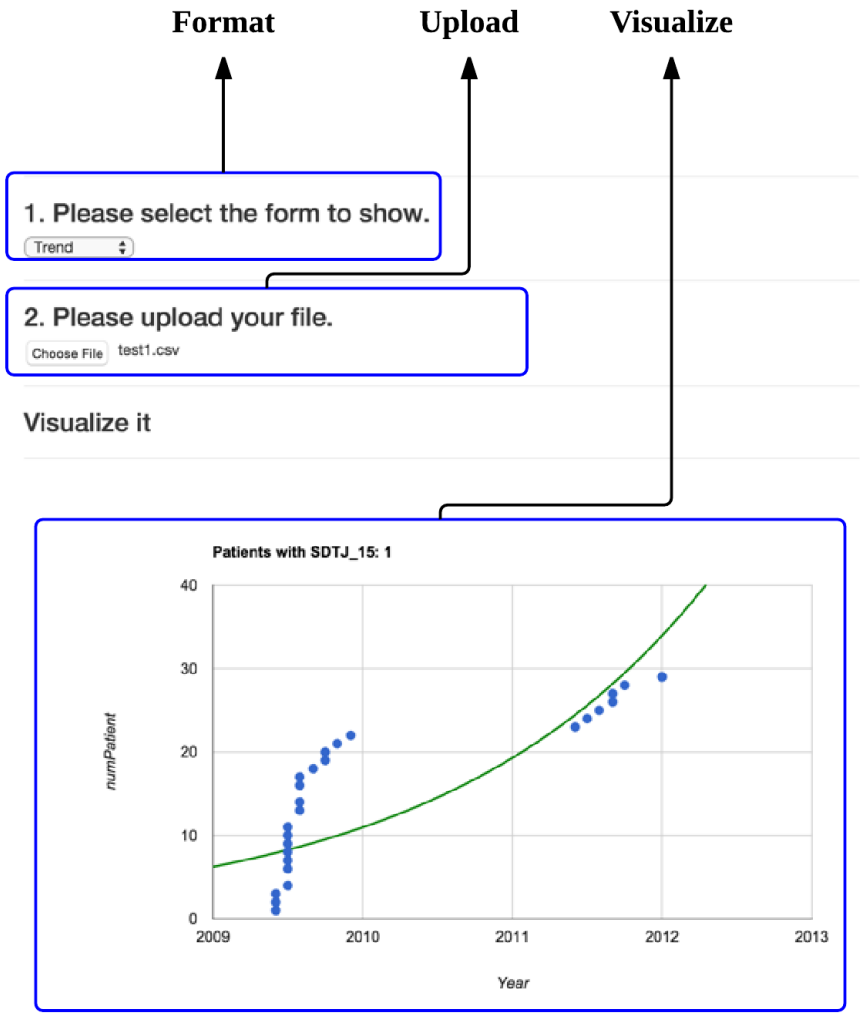
Upload Layout



Sequenc..	Primary ..	Primary ..	Enrollme..	Age	Sex	Marital S..	Sexual O..	Educati..	Num Par..	S D T J ..	Latitude	Longitu..	Zip
SDALL|S...	10002	STAHR	3/18/09	26	Male		MSM	4	5	3	32.670903	-117.09272	92015
SDALL|S...	10004	STAHR	4/2/09	27	Male		MSM	4	3	12	33.653216	-117.07986	92105
SDALL|S...	10024	STAHR	7/10/09	30	Male		MSM	4	30	123	32.638654	-117.06686	91950
SDALL|S...	10102	STAHR	10/7/09	27	Male		MSM	5	3	34	32.742892	-117.12774	92596
SDALL|S...	10110	STAHR	10/15/09	30	Male		Bi	3	15	3			91910
SDALL|S...	10128	STAHR	11/2/09	34	Male		Bi	3	5	22	32.721145	-116.99612	92104
SDALL|S...	10150	STAHR	11/6/09	36	Male		MSM	4	1	44	32.742892	-117.12774	92104
SDALL|S...	10180	STAHR	11/20/09	30	Male		MSM	4	1	66			91977
SDALL|S...	10208	STAHR	12/14/09	27	Male		MSM	2	4	55	32.7603	-117.07031	92104
SDALL|S...	10210	STAHR	12/16/09	39	Male		Bi	1	1	88	32.574398	-117.05633	92221
SDALL|S...	10258	STAHR	1/21/10	27	Male		MSM	4	5	77	32.638654	-117.06686	92115
SDALL|S...	10300	STAHR	2/10/10	29	Male		MSM	4	42	30	32.742892	-117.12774	92154

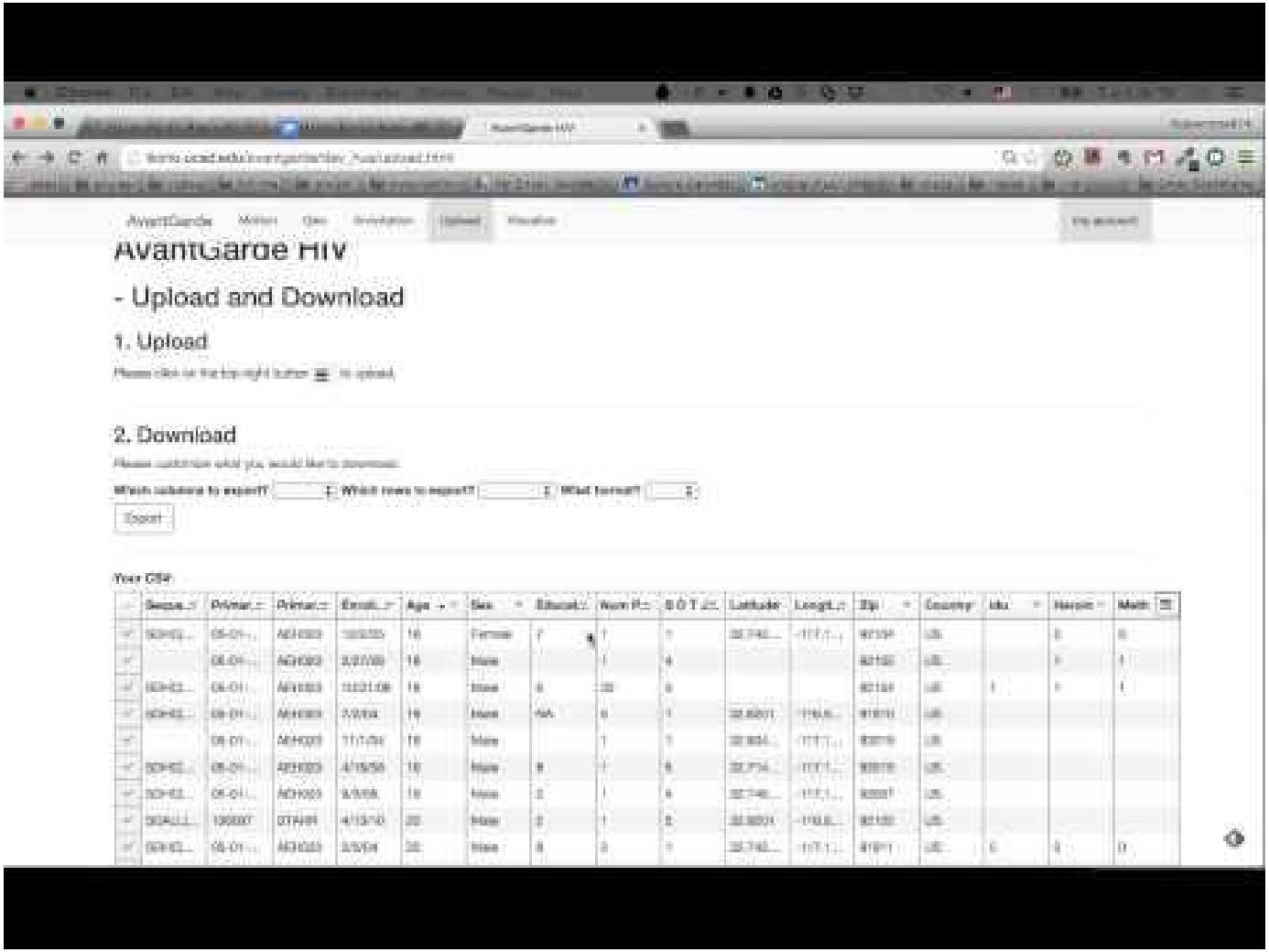
- Upload file
- Sort/edit/select data
- Download

Visualize Layout



- **Select format**
- **Upload file**
- **Visualize it**

Use Case - Customize and visualize local data



Outline

1. Background and Goal

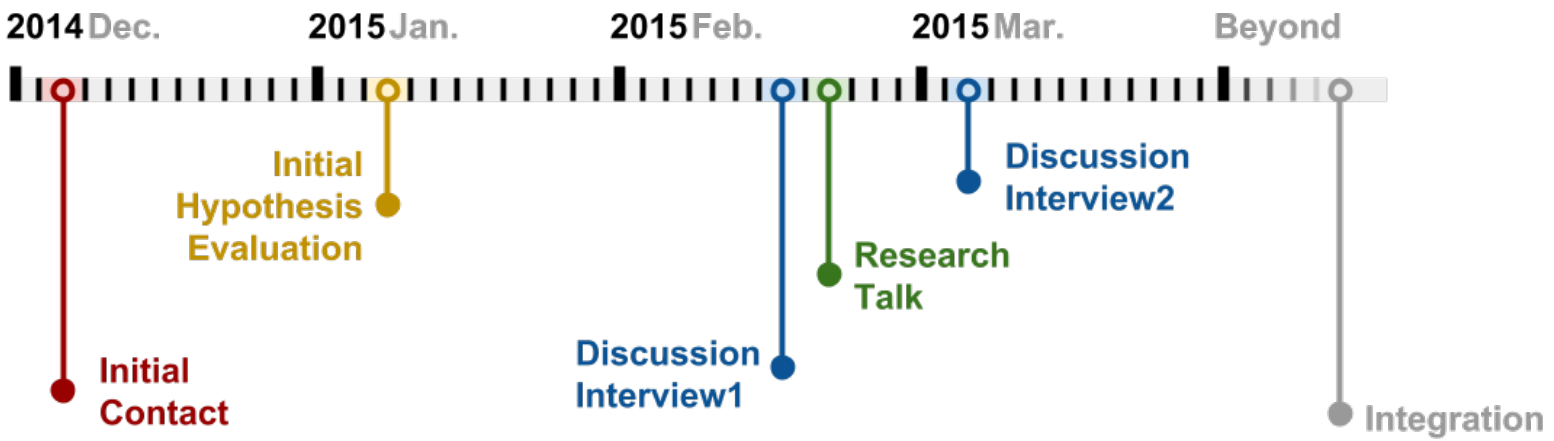
2. Design, Features and Use Cases

3. **Evaluation**
1) Prototype - design thinking
2) Researchers talk
3) Discussion interview

4. Conclusion and Future Work

Evaluation - Timeline

Initial Hypothesis Evaluation

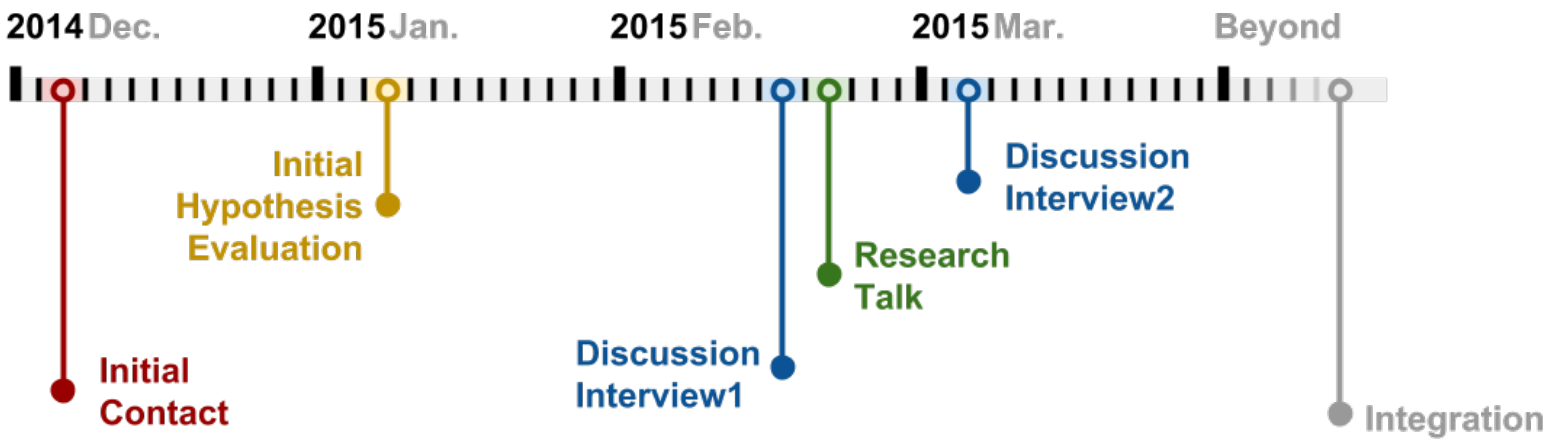


The researchers' feedback confirmed our hypothesis that our initial implementation of the motion chart is an effective way to capture and analyse the data change over time.

Discussion Interview 1

We demonstrated the geo chart to Dr Mehta and he thought by using geo chart they will able to capture and analyse the regional patient data change. And we started to take use cases into account.

Research Talk



During the meeting, other researchers expressed their interest in this tool and thought this could help with their research if it is more flexible with data.

Discussion Interview 2

Dr. Mehta agreed that the new feature "relative growth" could illustrate data more comprehensively and we evaluated different approaches to further combine our tool with other visualization approaches.

Outline

1. Background and Goal

2. Design, Features and Use Cases

3. Evaluation

4. **Conclusion and Future Work**

1) Conclusion
2) Future Work

Conclusion

- We built a novel **web-based interactive visualization tool with time component**.

- Allow users to **interact with data** much more easily and makes it possible for them to **visualize their own data**.

- The researchers and clinicians are now able to **explore novel insights and understanding** on the local impact of HIV incidence with this tool.

Future Work

- Explore further how we can test and evaluate our visualization tool

- Investigate how to automatically leverage new data context related to a new set of data.

- Integrate other data visualization approaches (Rich Interactive Data Visualization Tool for HIV)

Acknowledgements

- Prof. Nadir Weibel
- Dr. Sanjay Mehta
- Yu Xia & Sandy Law

For continuous aid,feedback and suggestions during the development process.

Reference

[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/index.php?option=com_content&view=article&id=392:dr-davey-smith-receives-a-2012-nida-avant-garde-award-for-hivaids-research&catid=81:current-news&Itemid=95
[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.

Q & A

Thanks!

Feel free to contact me: y4hua@eng.ucsd.edu