# Improving Big Data Visual Analytics with Interactive Virtual Reality

Andrew Moran, Vijay Gadepally, Matthew Hubbell, Jeremy Kepner

**IEEE High Performance Extreme Computing Conference** 

September 15-17, 2015



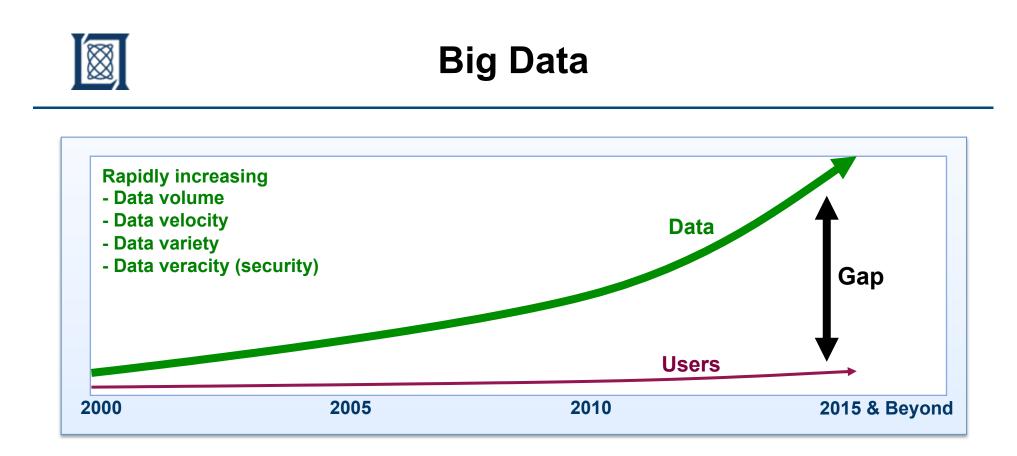
This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract #FA8721-05-C-0002. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Disclaimer: Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government



# Outline

- Introduction
  - Big Data
  - Visualization
  - Virtual Reality
  - Approach
    - Data Extraction
    - Game Configuration
    - Utilized Technologies
  - Results
    - Virtual Environment
    - Analytical Tasks
    - Challenges
  - Summary, Future Work and Next Steps

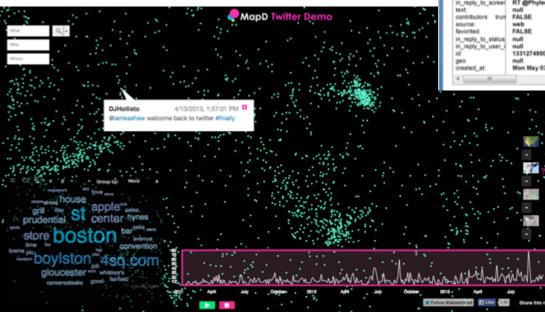


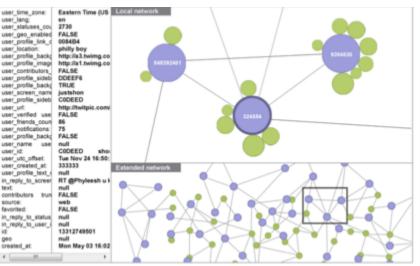
- Unstructured and overly complex
- Difficult to detect patterns
- Challenges in understanding/interaction



## Visualization

- Display Analytics
- Enable Exploration
- Draw Conclusions Faster







## Visualization

Eastern Time (US Local net **Display Analytics** 2730 ٠ user statuses col FALSE iser geo enable profile link 0084844 philly boy ser location **Enable Exploration** ttp://a3.tv ٠ profile imag http://a1.twimg.co FALSE ser contributors DOEEF6 er profile sidel **Draw Conclusions Faster** \_profile\_backg TRUE • user screen nam justshor user\_profile\_sideb CODEED http://twit ser\_uf: user verified use FALSE user friends coun user\_notifications: 75 user profile backs FALSE user name use null CODEED user\_id: user utc offset Tue Nov 24 16:50: 333333 user\_created\_at: user\_profile\_text\_ null in\_reply\_to\_screer RT @Phyleesh text null antributors FALSE source: web FALSE favorited in reply to status null null in\_reply\_to\_user\_i 13312749501 null Mon May 03 16:02 080 reated\_at 4/13/2013, 1:57:01 PM DJHollisto ne back to twitter #final **Two-Dimensional** • **Limited Perception** • Context •



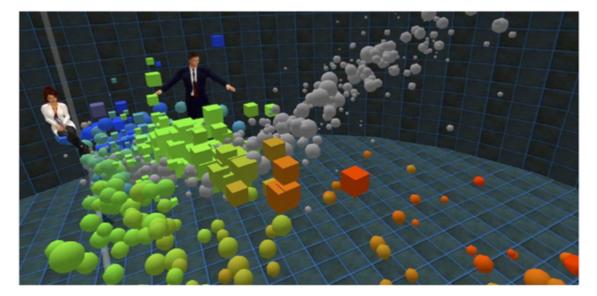
## **Virtual Reality**

- Sustainable Interaction
- Natural Movement
- Less Restriction





- First Person
   Perspectives
- Multi-Dimensional Data
- Immersive Gameplay
- Additional Context





# Outline

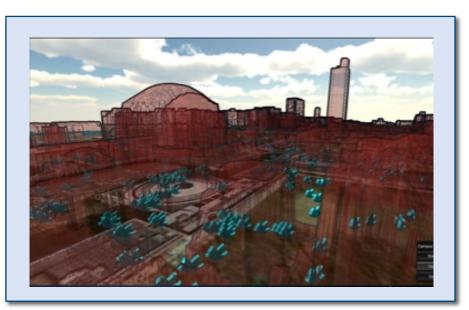
- Introduction
  - Big Data
  - Visualization
  - Virtual Reality



- Approach
  - Data Extraction
  - Game Configuration
  - Utilized Technologies
- Results
  - Virtual Environment
  - Analytical Tasks
  - Challenges
- Summary, Future Work and Next Steps

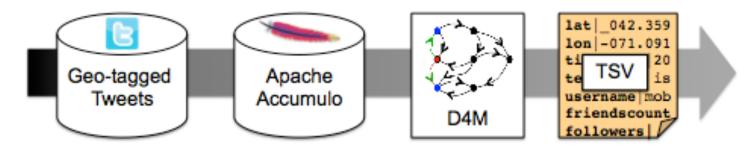


- Bridge the gap between data volume and user interaction through immersive technologies
- Enhance data visualization with additional context
- Appeal to the growth of 3D game development and virtual reality
- Scenario: Visualize Twitter at MIT
  - Why Twitter?
    - Large Data Set
    - Popularity
  - Why MIT?
    - Familiarity
    - Geographical Basis



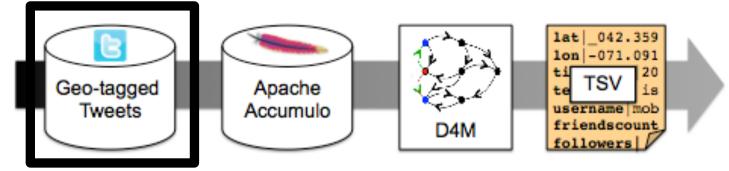


#### **Process Pipeline for Twitter Data**





#### **Process Pipeline for Twitter Data**

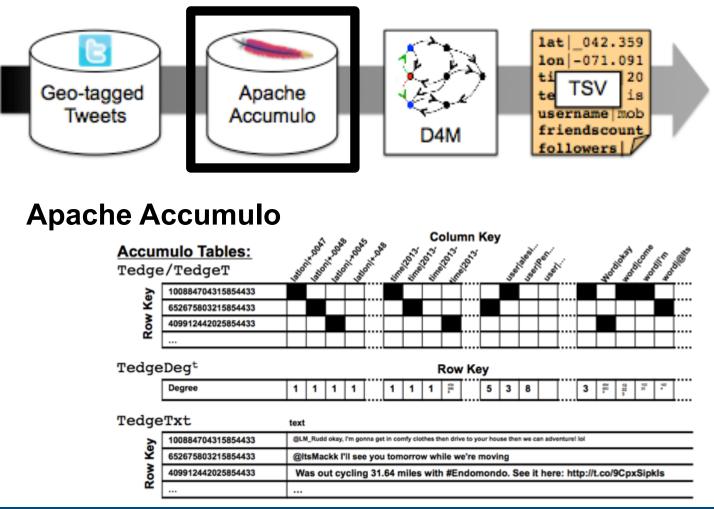


#### **Twitter Decahose**

TweetID	Time	Lat	Long	User	Source	Text
334458513407488001	2013-05-15 00:01:38	41.50442327	-71.3080676	alesiatasso	Twitter for iPhone	@LM_Rudd okay, I'm
334458512308576256	2013-05-15 00:01:37	42.56289604	-84.82935883	skally_pal	Twitter for iPhone	@ltsMackk I'll see you
334458513755602944	2013-05-15 00:01:38	41.39560229	-81.73950863	AmbahDee	Twitter for iPhone	@NataliaProkop come
334458515798245377	2013-05-15 00:01:38	37.19058402	-93.31629432	RockSteady	Twitter for iPhone	@ktp35 I didn't think to
334458515383005184	2013-05-15 00:01:38	35.6337457	139.6607298	karate_h	Twitter for Android	"""@IngatanSekolah:
334458520244219904	2013-05-15 00:01:39	42.2041867	-87.8126571	yobebeau	Endomondo	Was out cycling 31.64

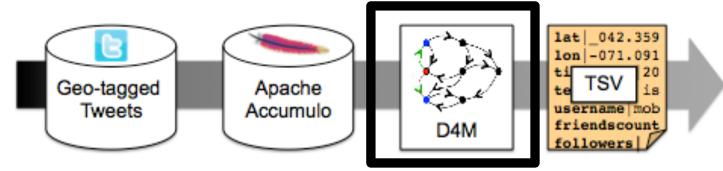


#### **Process Pipeline for Twitter Data**





#### **Process Pipeline for Twitter Data**



#### D4M (Dynamic Distributed Dimensional Data Model)

## Restricted within MIT

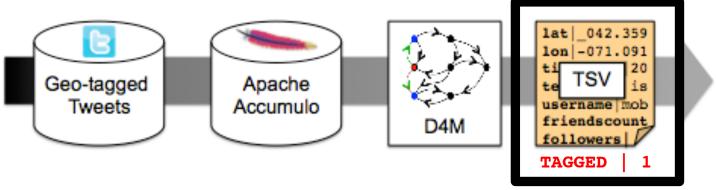
```
Amit = Tedge(:, ['lon|,-71.090,:,-71.099,'] ...
['lat|,42.350,:,42.357,']);
```

#### Contains the word "danger"

```
Atagged = Amit(:, StartsWith('word|danger,');
```



#### **Process Pipeline for Twitter Data**



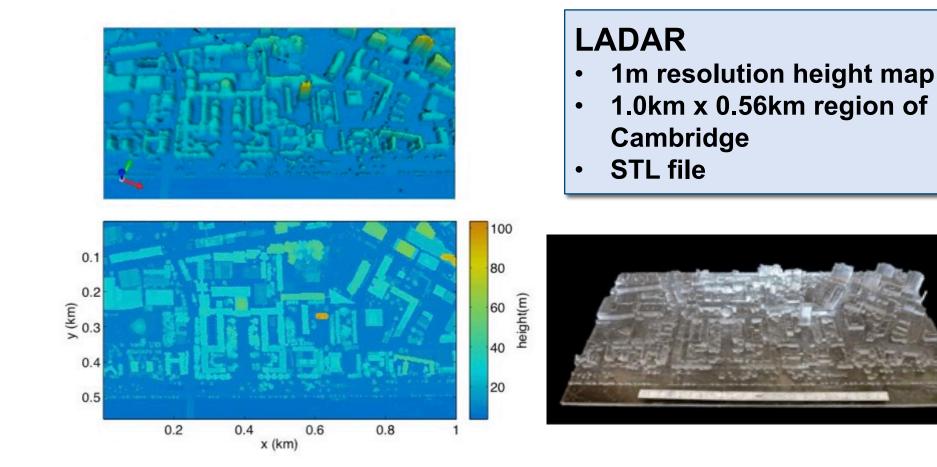
D4M (Dynamic Distributed Dimensional Data Model)

Restricted within MIT
 Amit = Tedge(:, ['lon|,-71.090,:,-71.099,'] ...
 ['lat|,42.350,:,42.357,']);

 Contains the word "danger"
 TAGGED | 1
 Atagged = Amit(:, StartsWith('word|danger,');



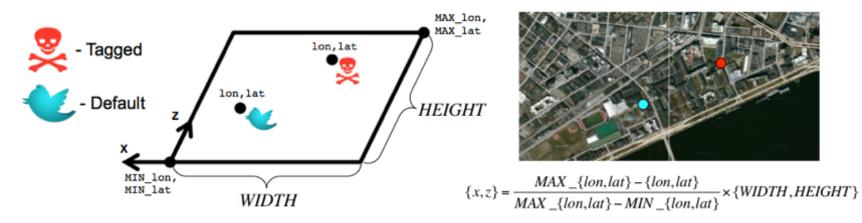
## **Data Extraction, MIT**





## **Game Configuration**

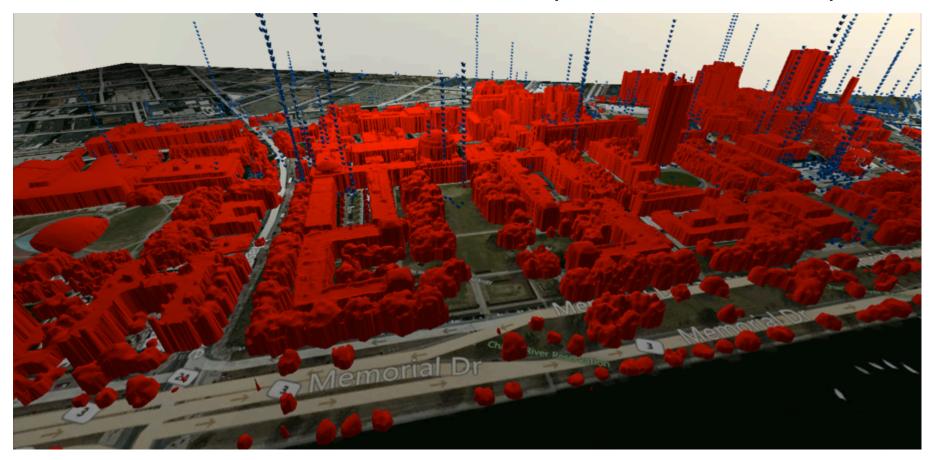






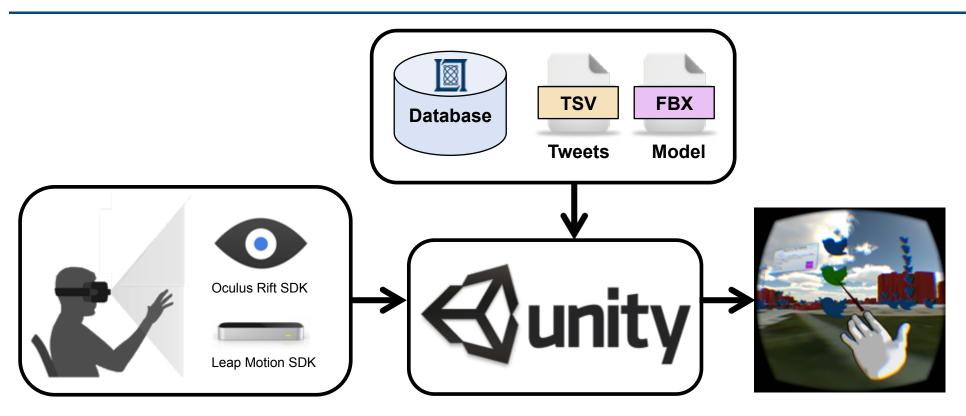
## **Game Configuration**

• 3D Rendition of MIT with ~6000 tweets (Oct. 2013 – Feb. 2014)





## **Utilized Technologies**



#### "3D Input Meets 3D Output" – Why choose Unity3D?

Game Engine 
 Physics Engine 
 Extensive Integration 
 Flexibility



# Outline

- Introduction
  - Big Data
  - Visualization
  - Virtual Reality
- Approach
  - Data Extraction
  - Game Configuration
  - Utilized Technologies

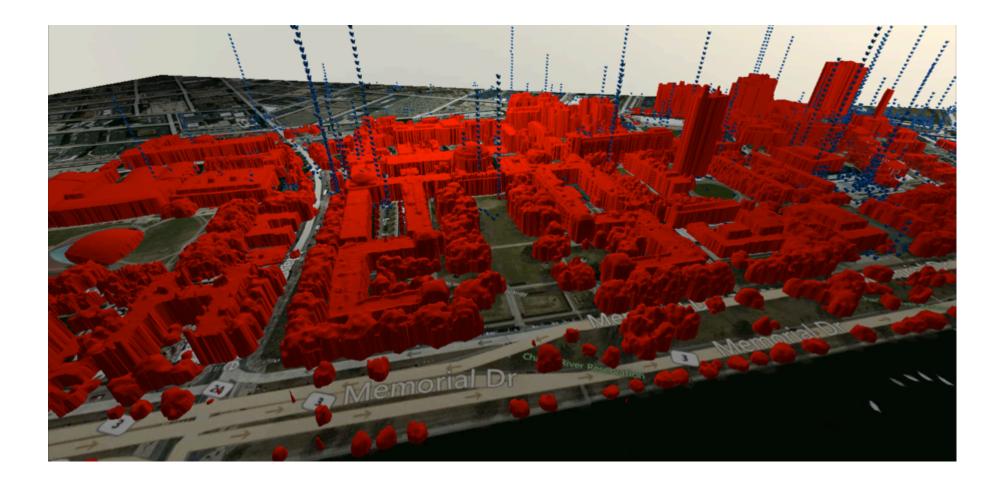


### Results

- Virtual Environment
- Analytical Tasks
- Challenges
- Summary, Future Work and Next Steps



## **Virtual Environment**

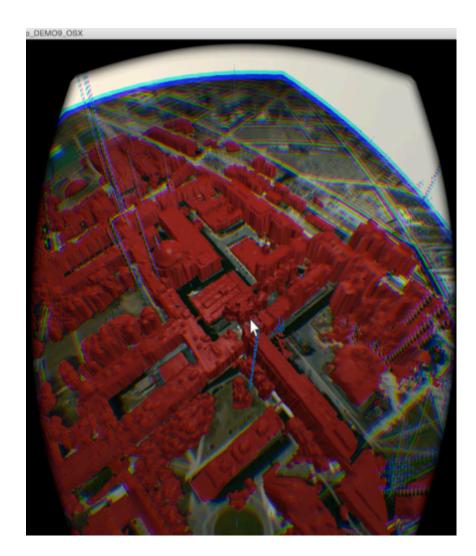


LINCOLN LABORATORY MASSACHUSETTS INSTITUTE OF TECHNOLOGY



## Navigation/Exploration

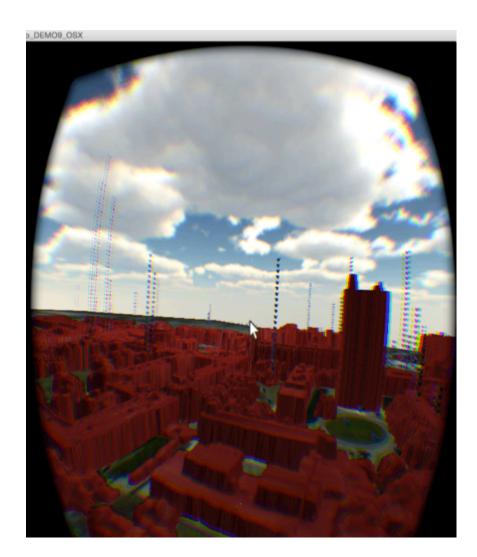
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





## Navigation/Exploration

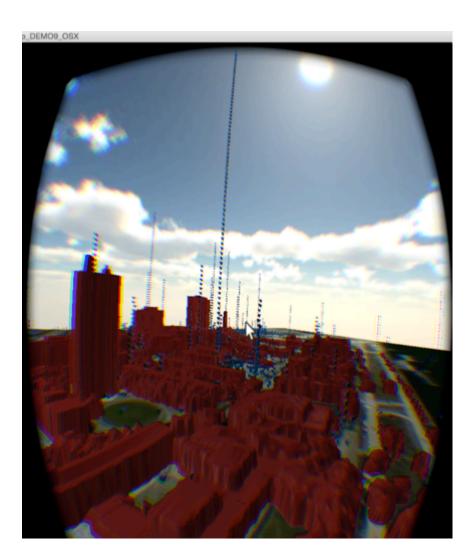
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





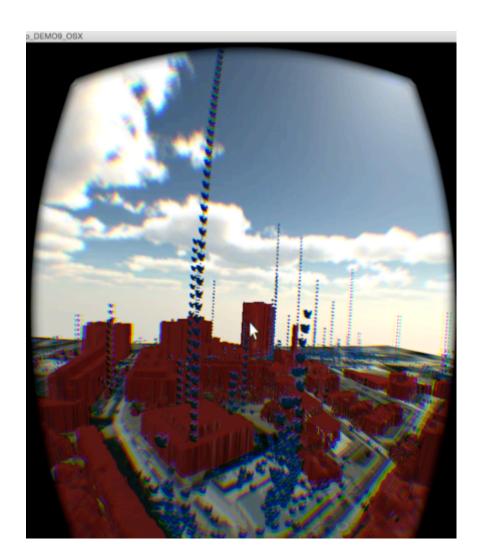
#### **Navigation/Exploration**

- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



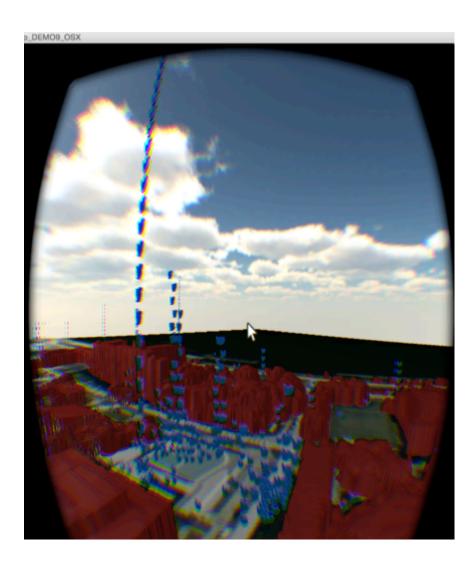


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



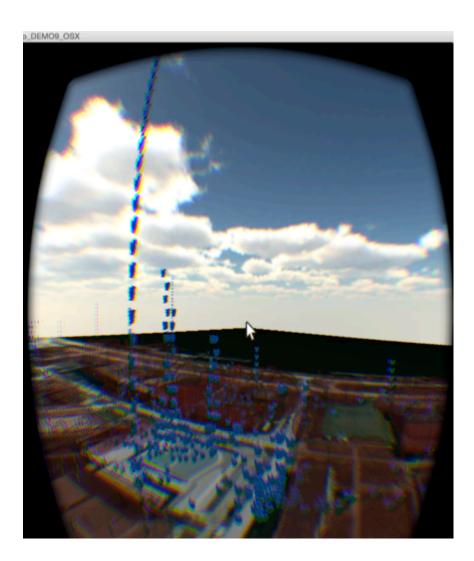


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



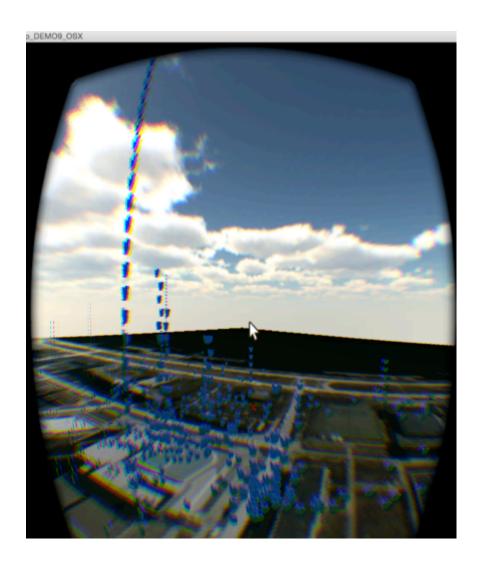


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



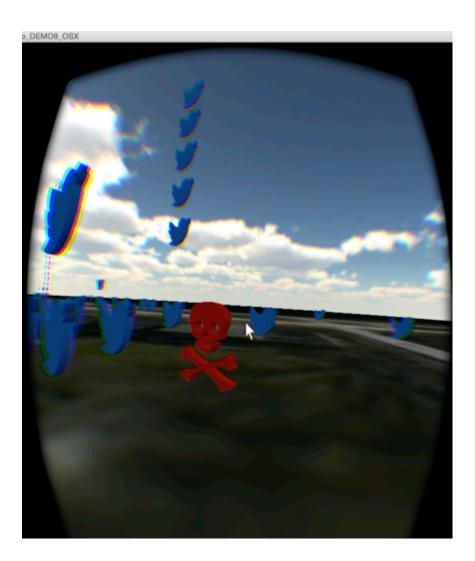


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



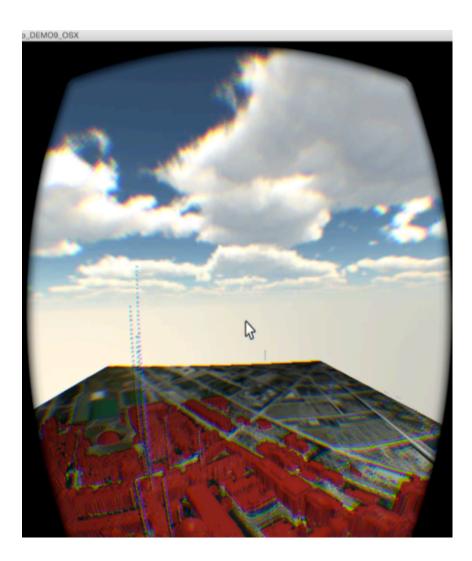


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking

D_DEMO9_OSX	
Record ID# 174	
142.955577274055 1301-771055771442534	
trestern 2010-10-251052235 0007	
tox: 24ttaRctess ummm who are you again? #strangenta	
Manage and	- Haren
Frenchercourt 32	
Nexte	
PERVOUS	

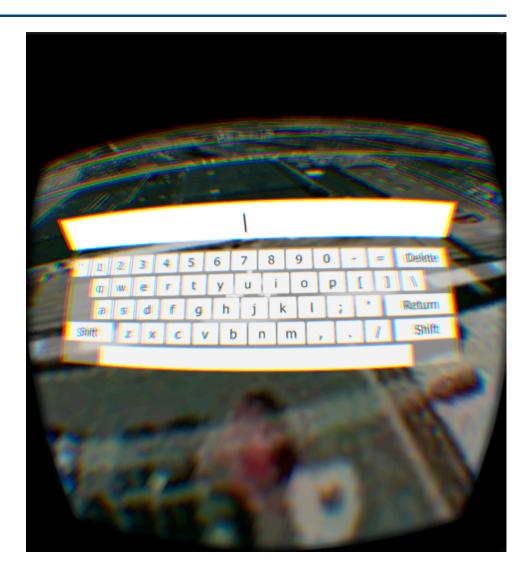


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



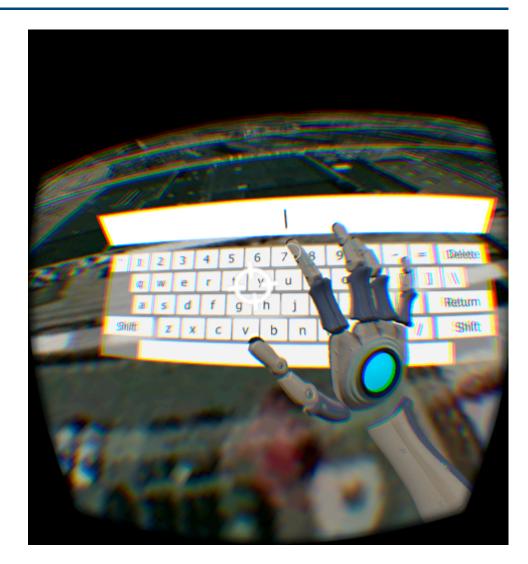


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



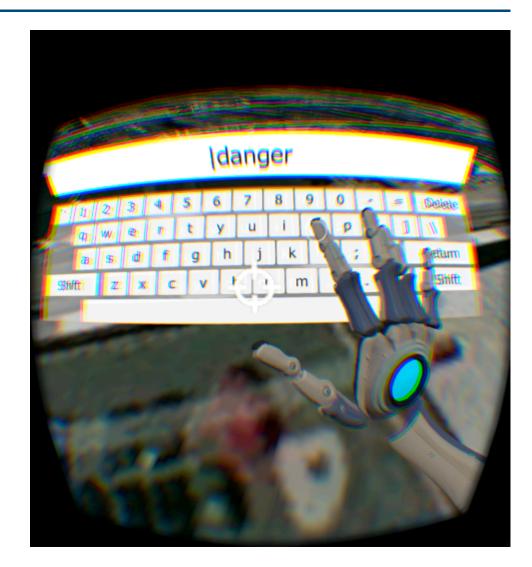


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



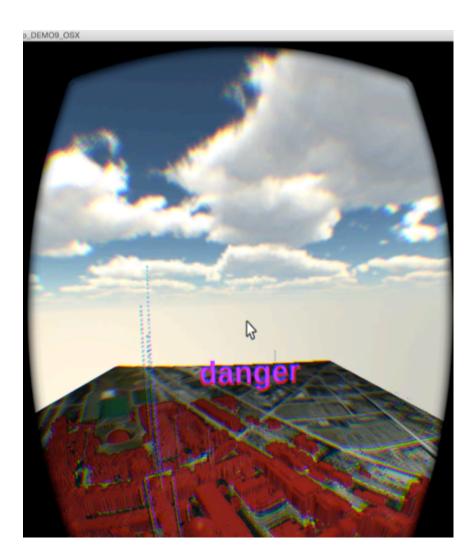


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



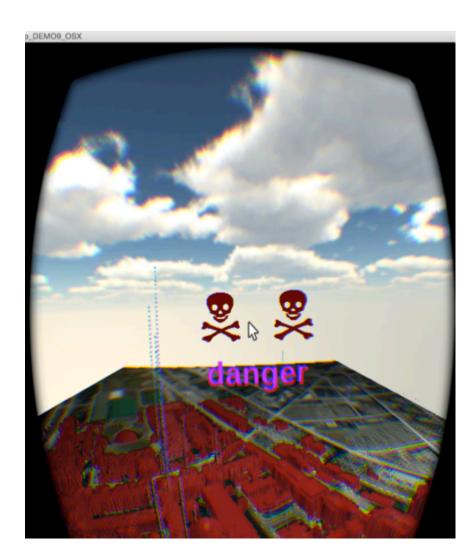


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



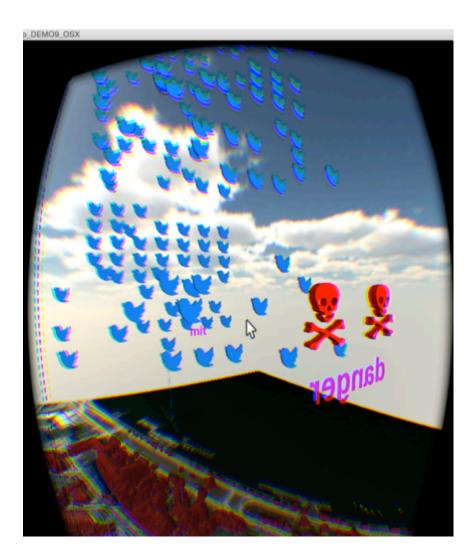


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



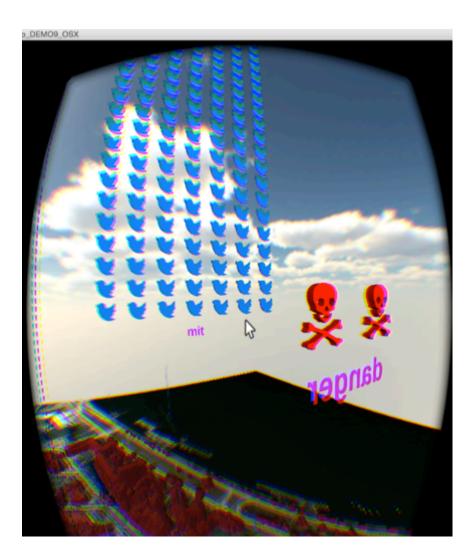


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



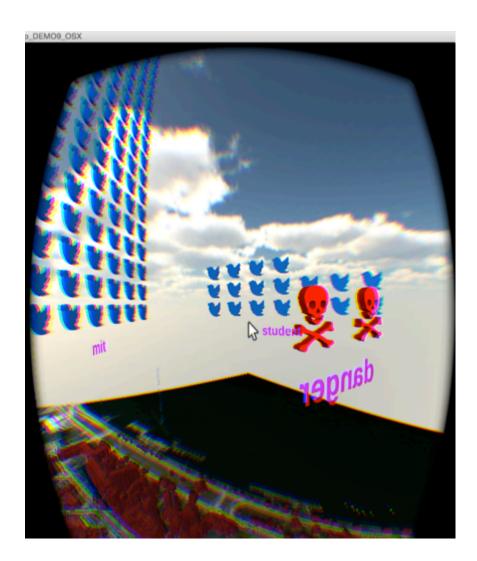


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



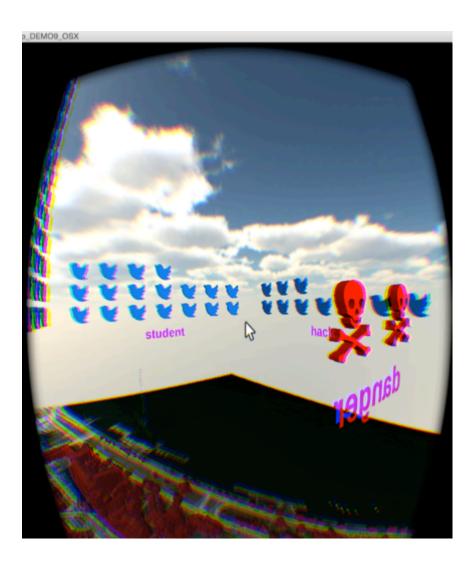


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



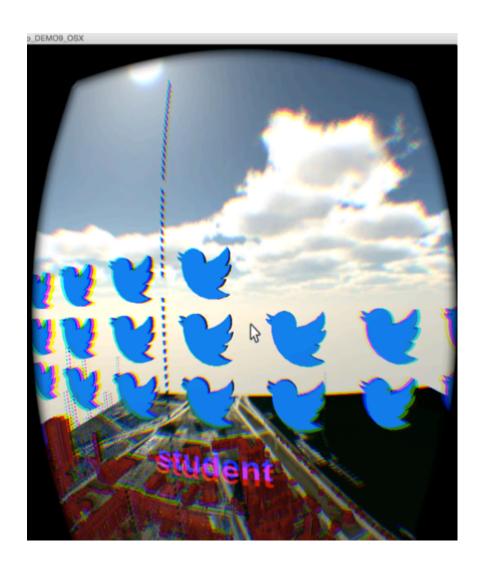


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



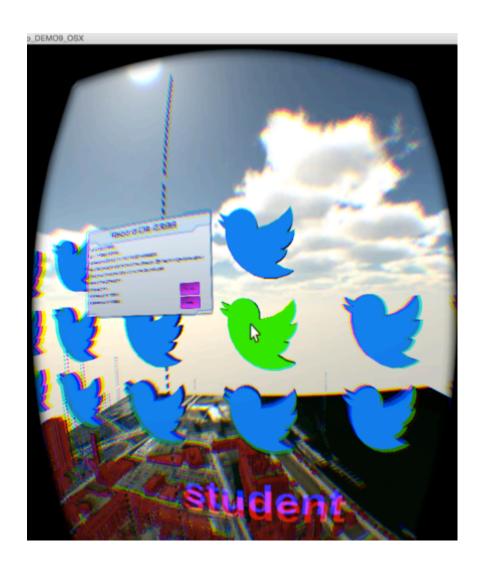


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
  - Tracking



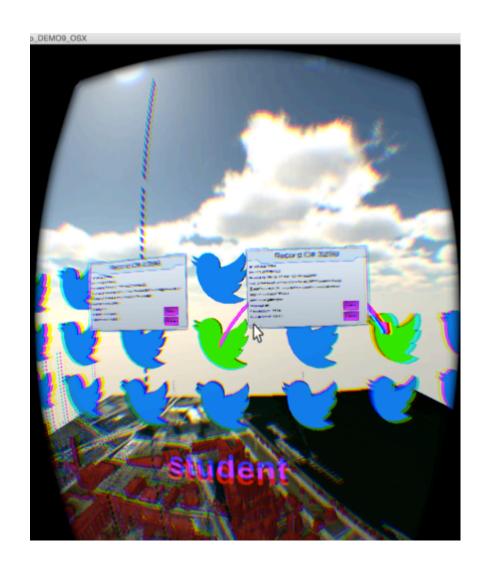


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
  - Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
  - Tracking



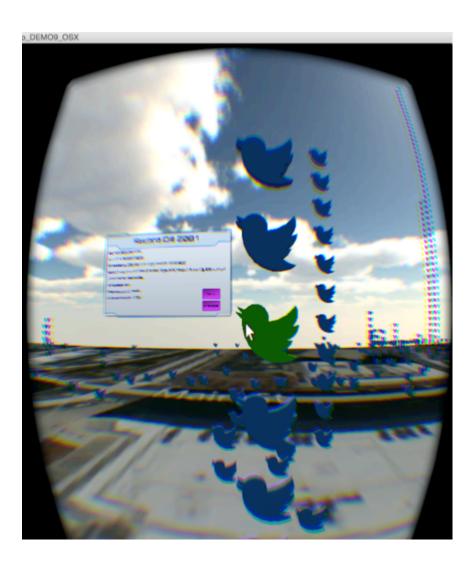


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
  - Tracking



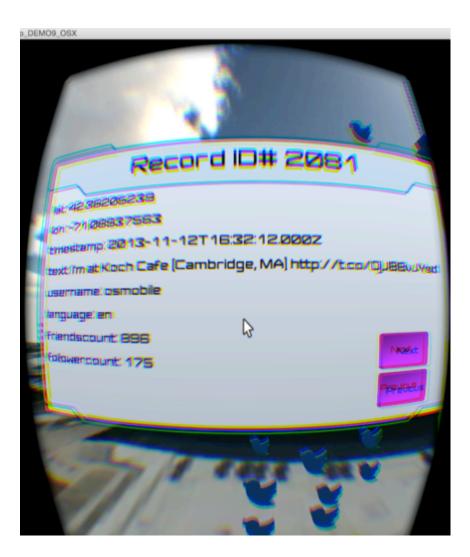


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



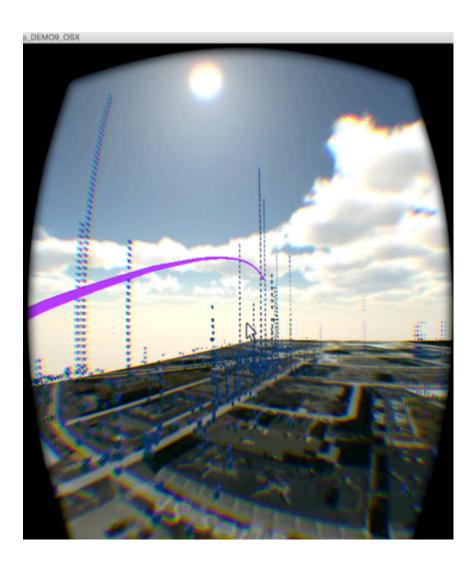


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



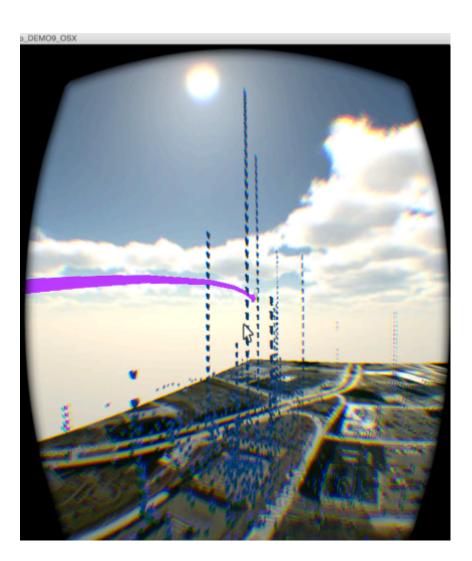


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



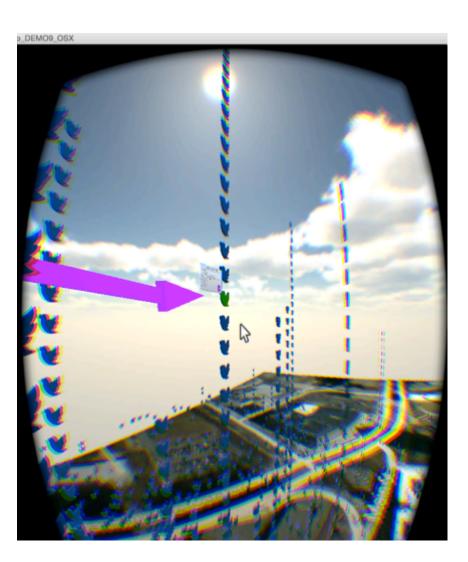


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



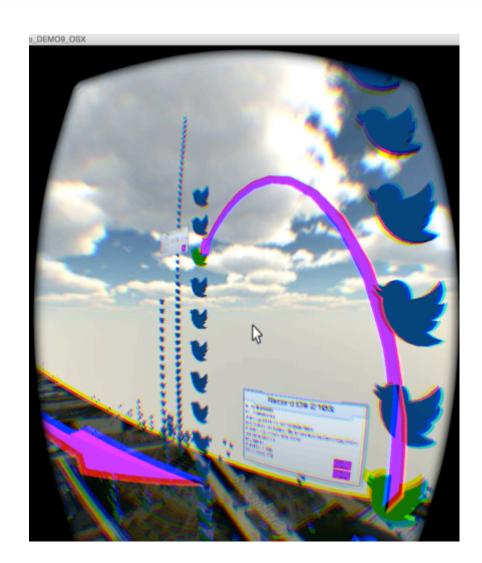


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking

Record ID# 2103	
NC 412 1922020596 201: 7/4 2000005702 201: 7/4 200005702 201: 7/4 2010 2010 7: 12 142 302 005 00002	
text://mili/kandall/Square - Qchampionskiendal (Cambridge:/wkjtw/ 2010/esi/100/s1E02LEXSC	
usermenti comobile languagettern Friendiscount: 19595	
Ronowerscolung: 11755	k



- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



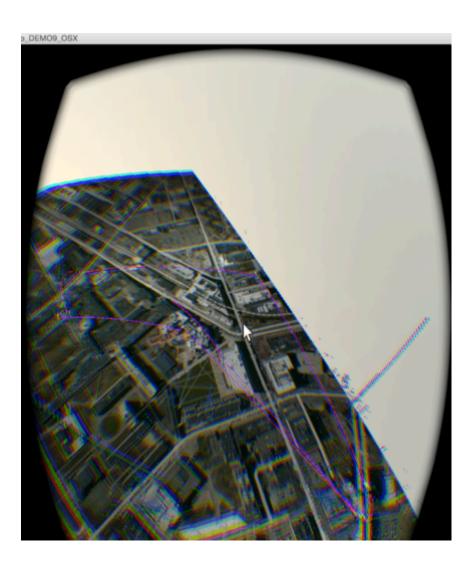


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



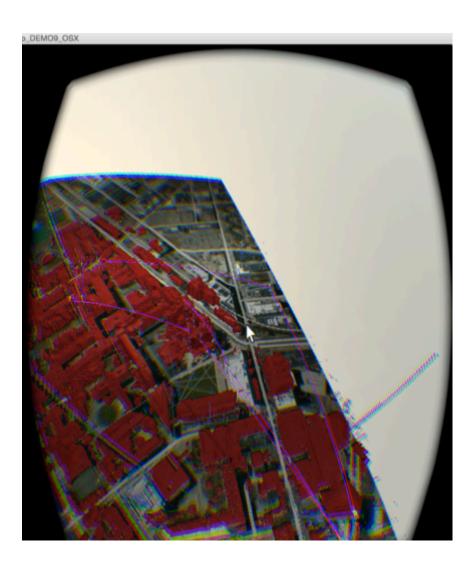


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking



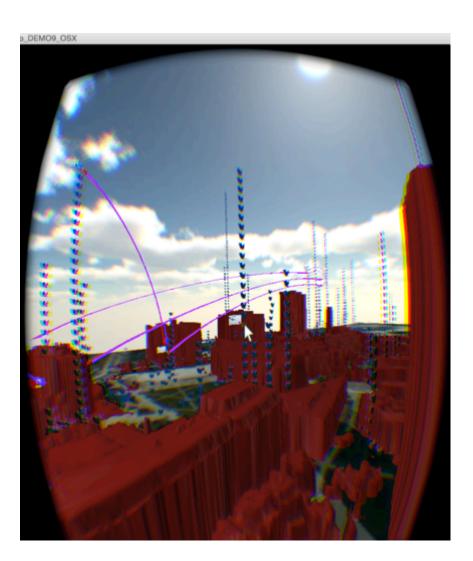


- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





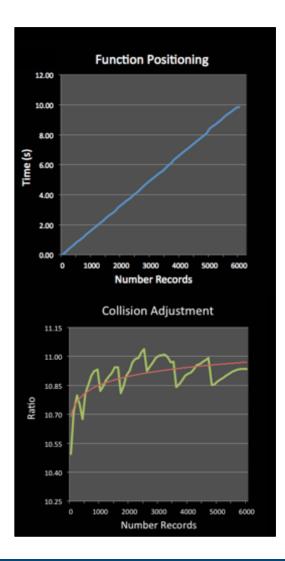
- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- Relationships
- Tracking





## Challenges

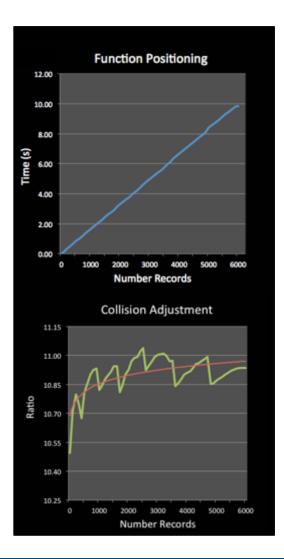
- Rendering Models
- Camera LOD
- Occlusion Layers
- Processing Frame Rate
- Collider Activation





## Challenges

- Rendering Models
- Camera LOD
- Occlusion Layers
- Processing Frame Rate
- Collider Activation
  - <u>Problem</u>: Further instantiation leads to more collisions
  - <u>Solution</u>: Convert to using geohashed locations





# Outline

- Introduction
  - Big Data
  - Visualization
  - Virtual Reality
  - Overview of Approach
- Approach
  - Data Extraction
  - Game Configuration
  - Utilized Technologies
- Results
  - Virtual Environment
  - Analytical Tasks
  - Challenges



Summary, Future Work and Next Steps



#### Summary

#### Goal

- Promote improved visualization and interaction with Big Data
- Achieve better situational awareness

#### Approach

- Collect and Visualize Twitter Data at MIT
- Utilize Unity3D Game Engine, showcase emerging technologies

#### Results

- Immersive virtual environment
- Data overlaid on geospatial domain
- Dynamically perform analytical tasks
- Ongoing/Future Work
  - Performance enhancements/optimizations
  - Usability & GUI Improvement
  - Additional Analytics
  - Live Data



#### Backup