Improving Big Data Visual Analytics with Interactive Virtual Reality

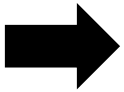
Andrew Moran, Vijay Gadepally, Matthew Hubbell, Jeremy Kepner

IEEE High Performance Extreme Computing Conference

September 15-17, 2015

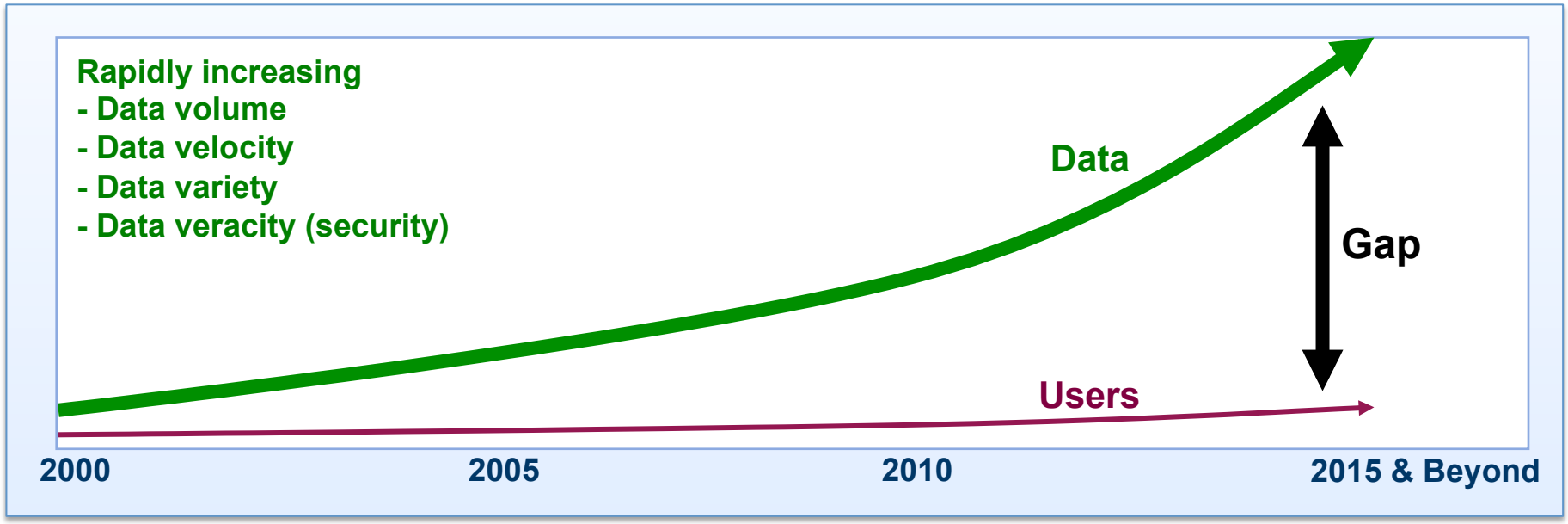
This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract #FA8721-05-C-0002. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

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

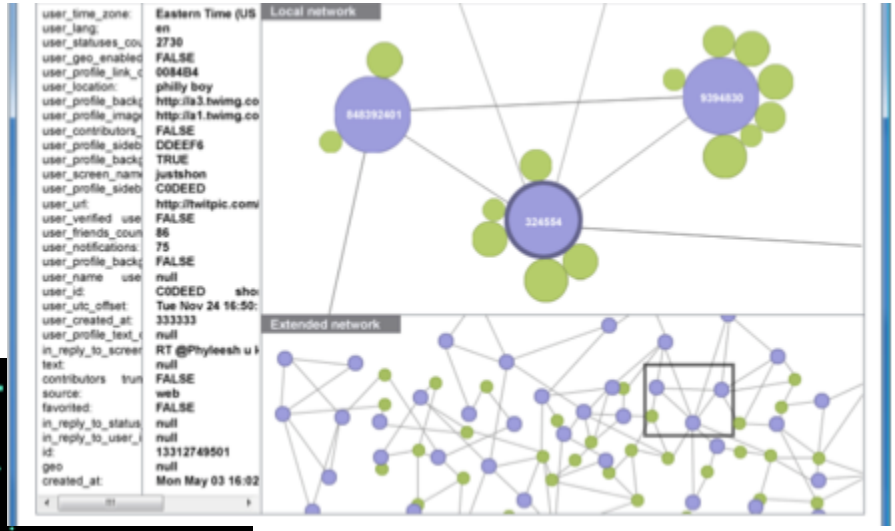
Big Data

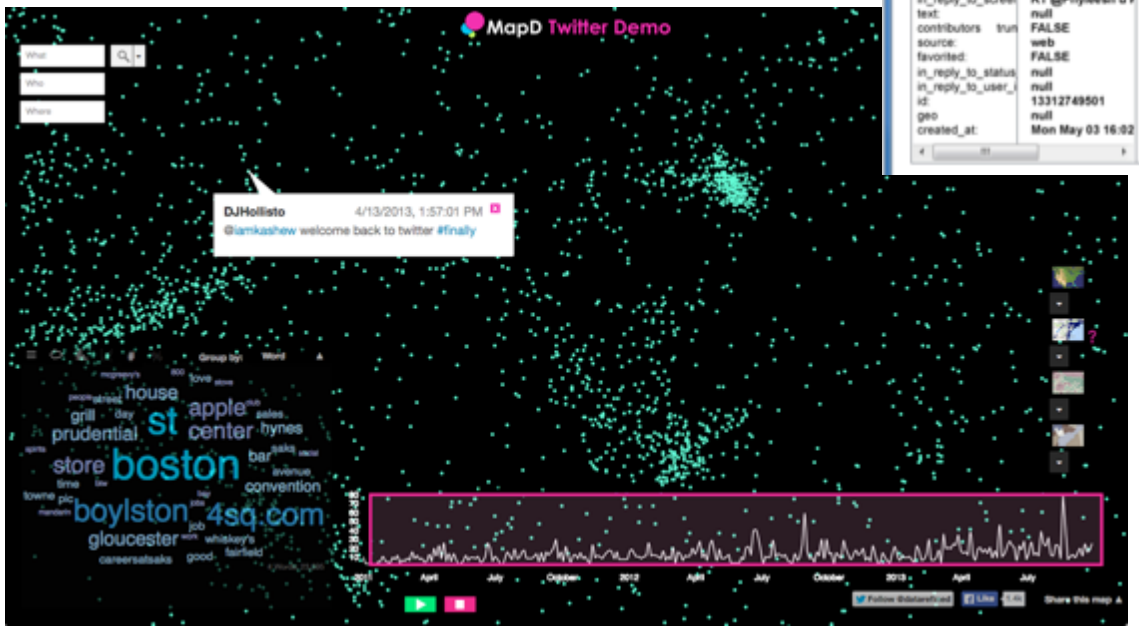


- Unstructured and overly complex
- Difficult to detect patterns
- Challenges in understanding/interaction

Visualization

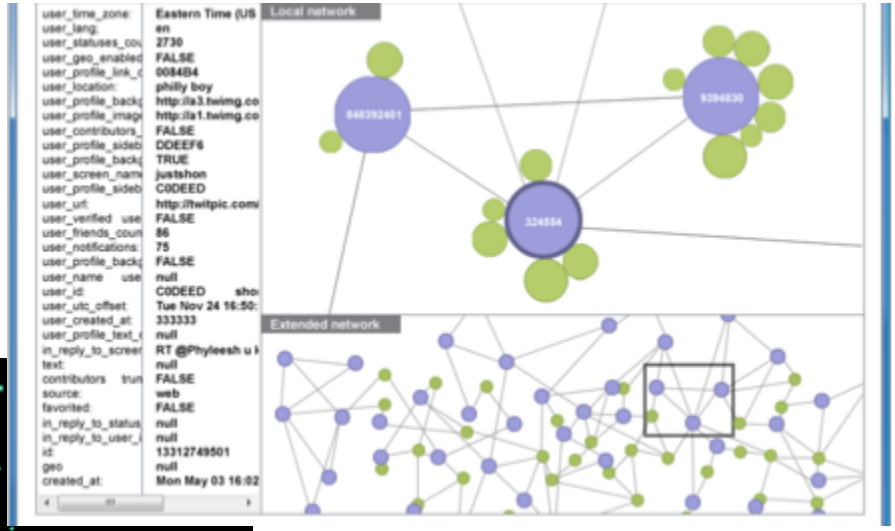
- Display Analytics
- Enable Exploration
- Draw Conclusions Faster

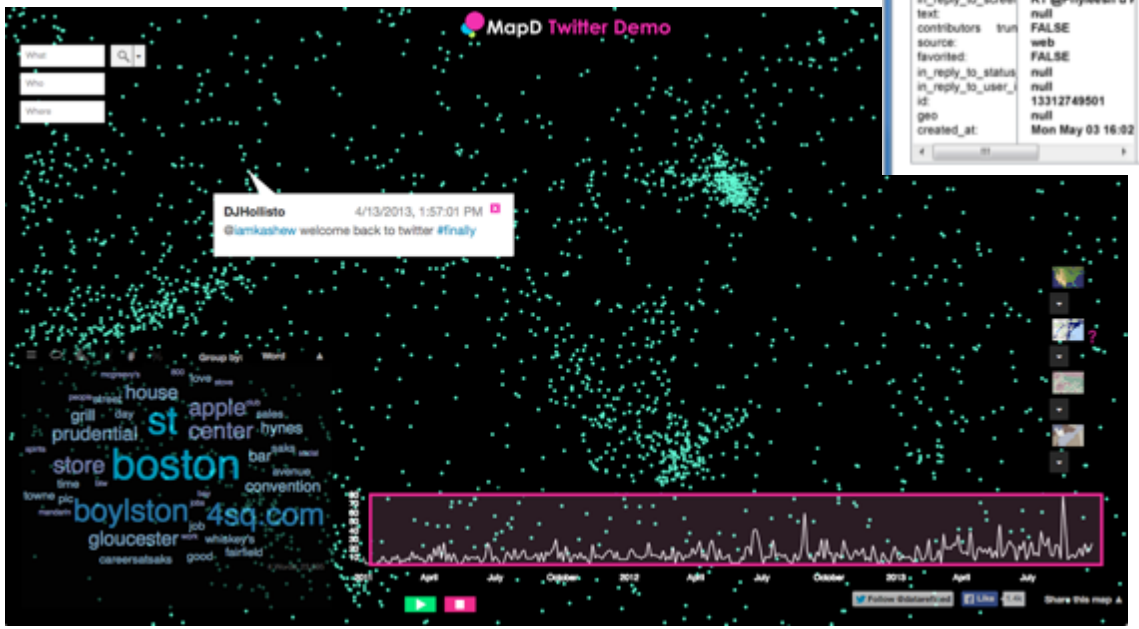




Visualization

- Display Analytics
- Enable Exploration
- Draw Conclusions Faster



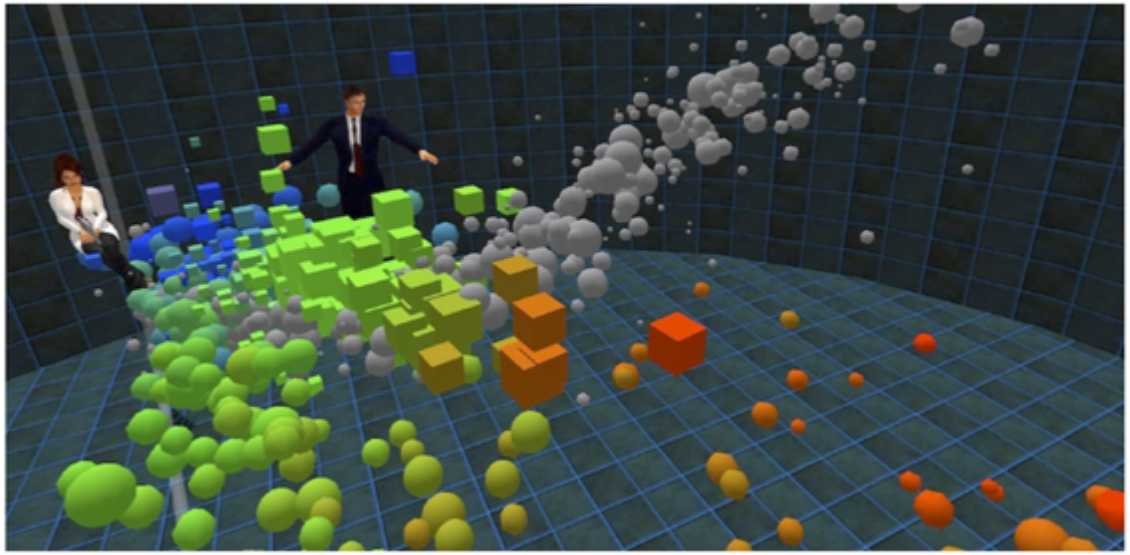


- Two-Dimensional
- Limited Perception
- Context

Virtual Reality

- Sustainable Interaction
- Natural Movement
- Less Restriction

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

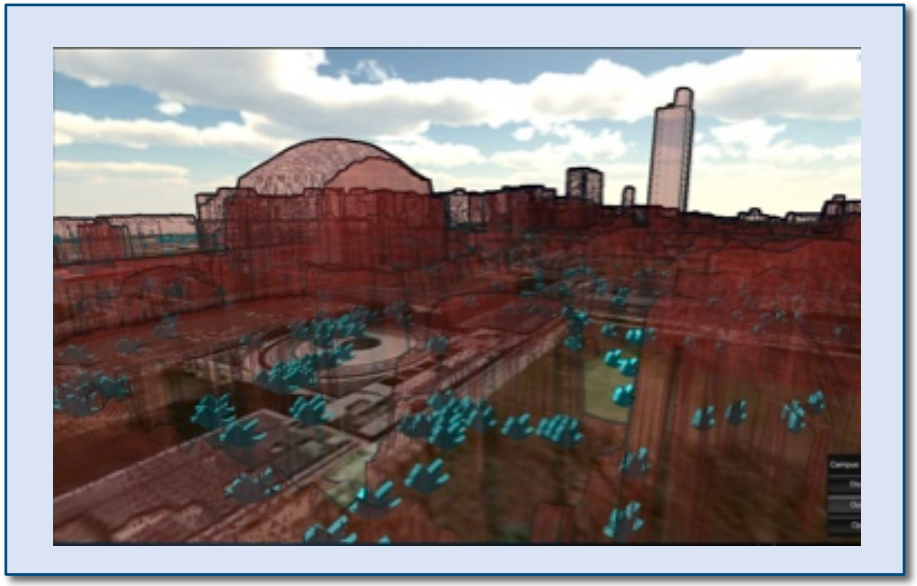


Outline

- 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

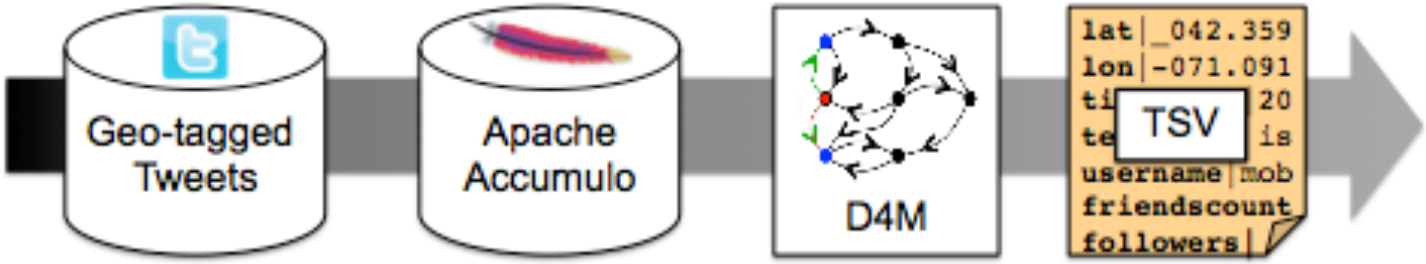
Approach

- 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



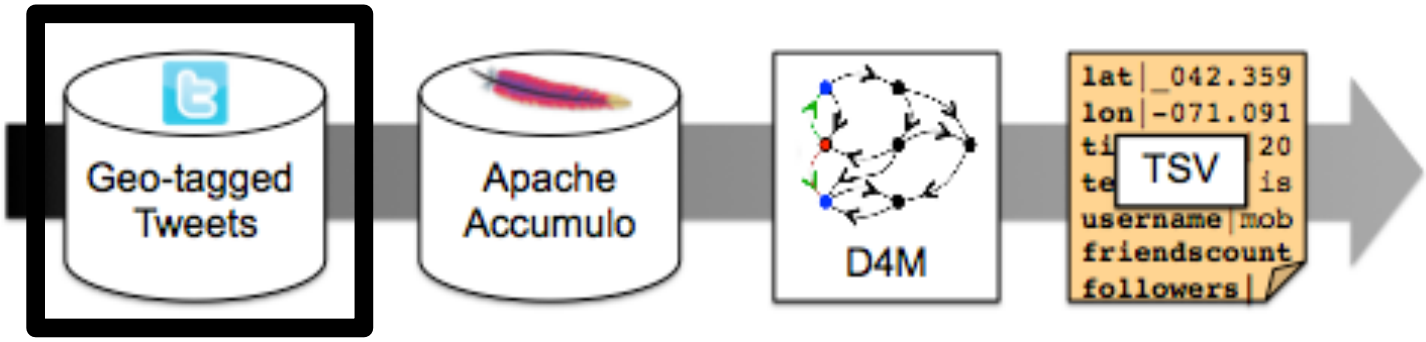
Data Extraction, Tweets

Process Pipeline for Twitter Data



Data Extraction, Tweets

Process Pipeline for Twitter Data

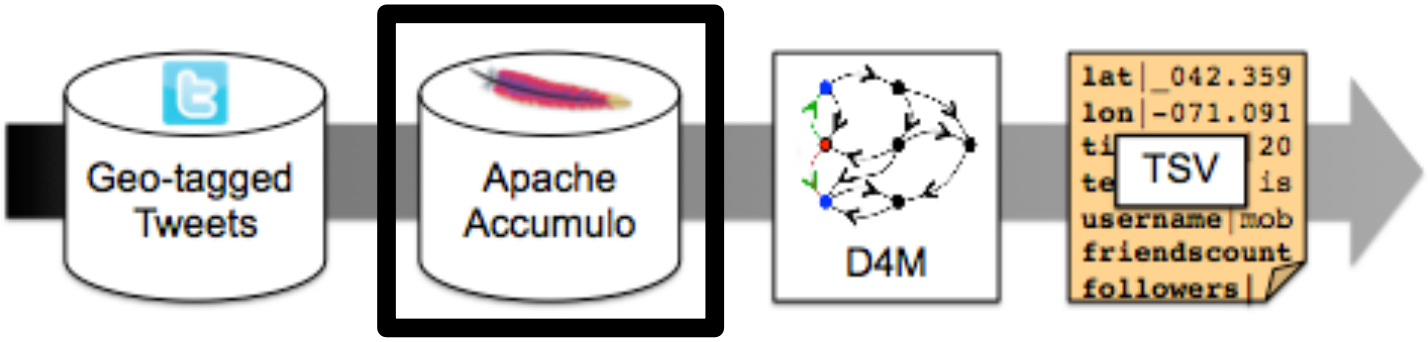


Twitter Decahose

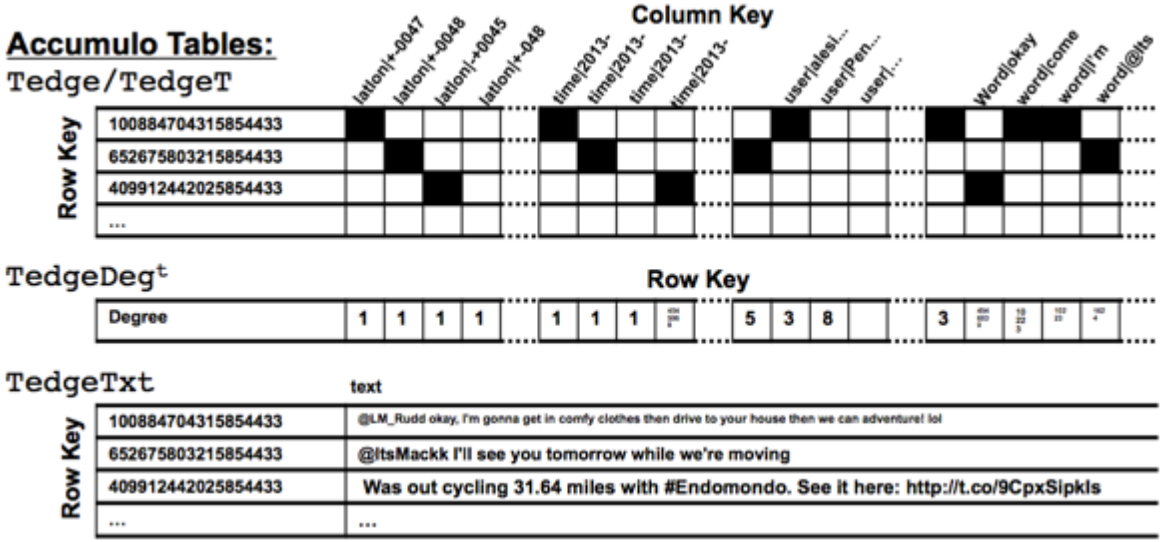
TweetID	Time	Lat	Long	User	Source	Text
334458513407488001	2013-05-15 00:01:38	41.50442327	-71.3080676	alesiatasso	Twitter for iPhone	@LM_Rudd okay, I'm...
334458512308576256	2013-05-15 00:01:37	42.56289604	-84.82935883	skally_pal	Twitter for iPhone	@ItsMackk I'll see you…
334458513755602944	2013-05-15 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 …
…	…	…		…		…

Data Extraction, Tweets

Process Pipeline for Twitter Data

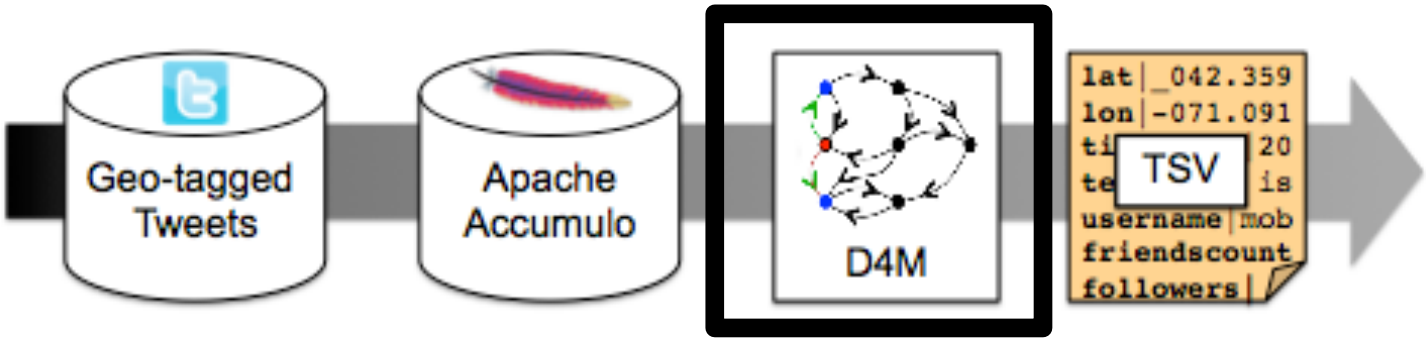


Apache Accumulo



Data Extraction, Tweets

Process Pipeline for Twitter Data



D4M (Dynamic Distributed Dimensional Data Model)

- **Restricted within MIT**

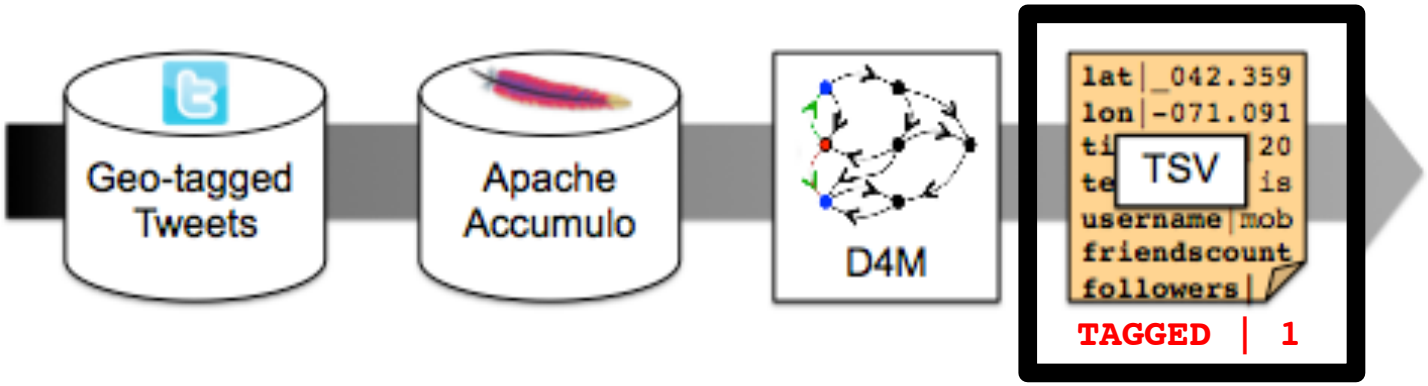
```
Amit = Tedge(:, ['lon|,-71.090,:,-71.099,'] …
                ['lat|,42.350,:,42.357,']);
```

- **Contains the word "danger"**

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

Data Extraction, Tweets

Process Pipeline for Twitter Data



D4M (Dynamic Distributed Dimensional Data Model)

↓

TSV

- **Restricted within MIT**

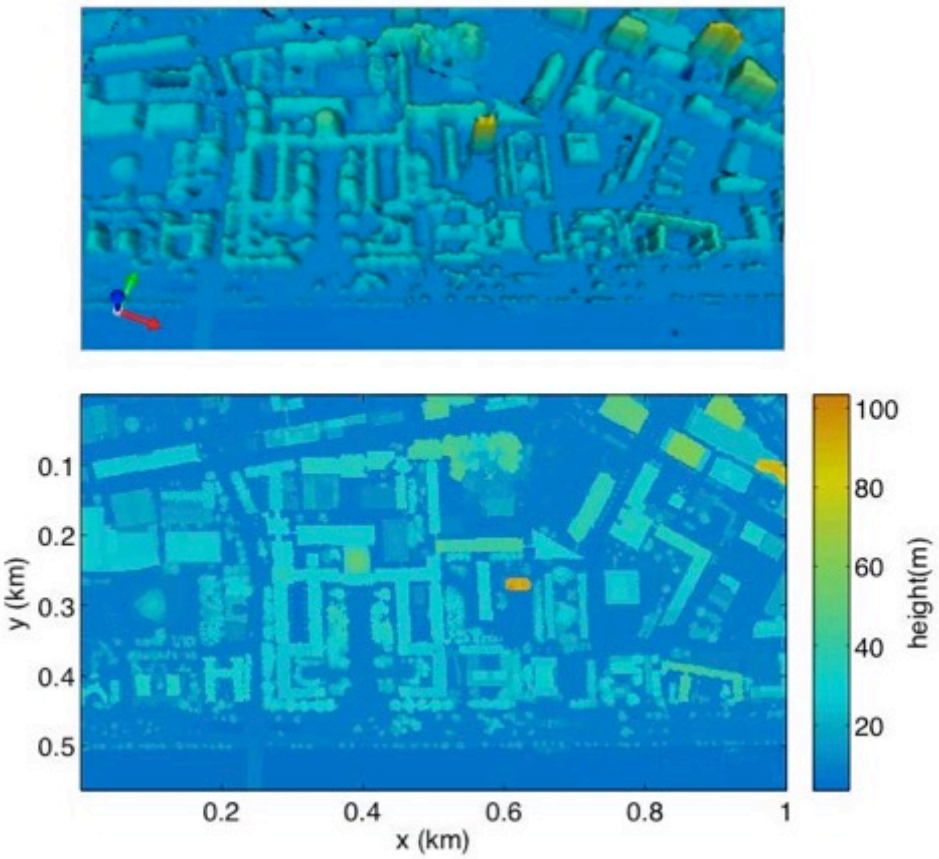
```
Amit = Tedge(:, ['lon|,-71.090,:,-71.099,'] …
                ['lat|,42.350,:,42.357,']);
```

- **Contains the word "danger"**

TAGGED | 1

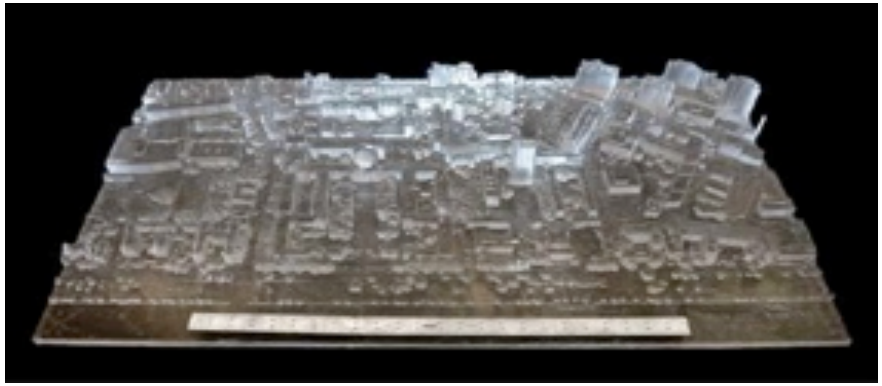
```
Atagged = Amit(:, StartsWith('word|danger,');
```

Data Extraction, MIT

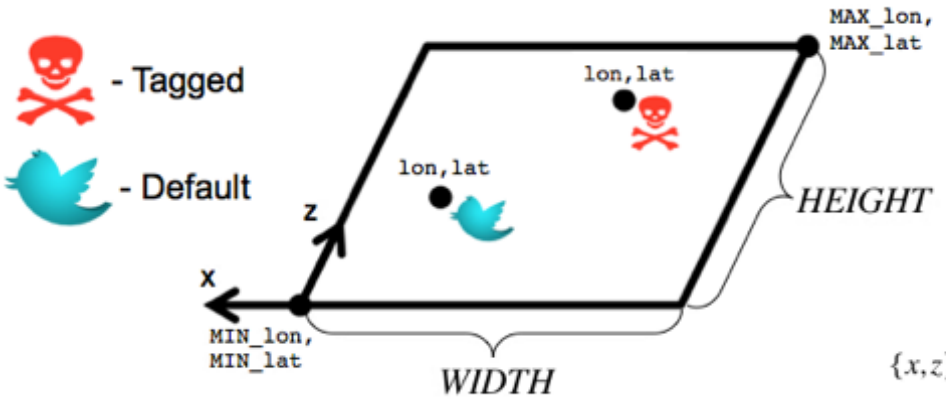


LADAR

- 1m resolution height map
- 1.0km x 0.56km region of Cambridge
- STL file



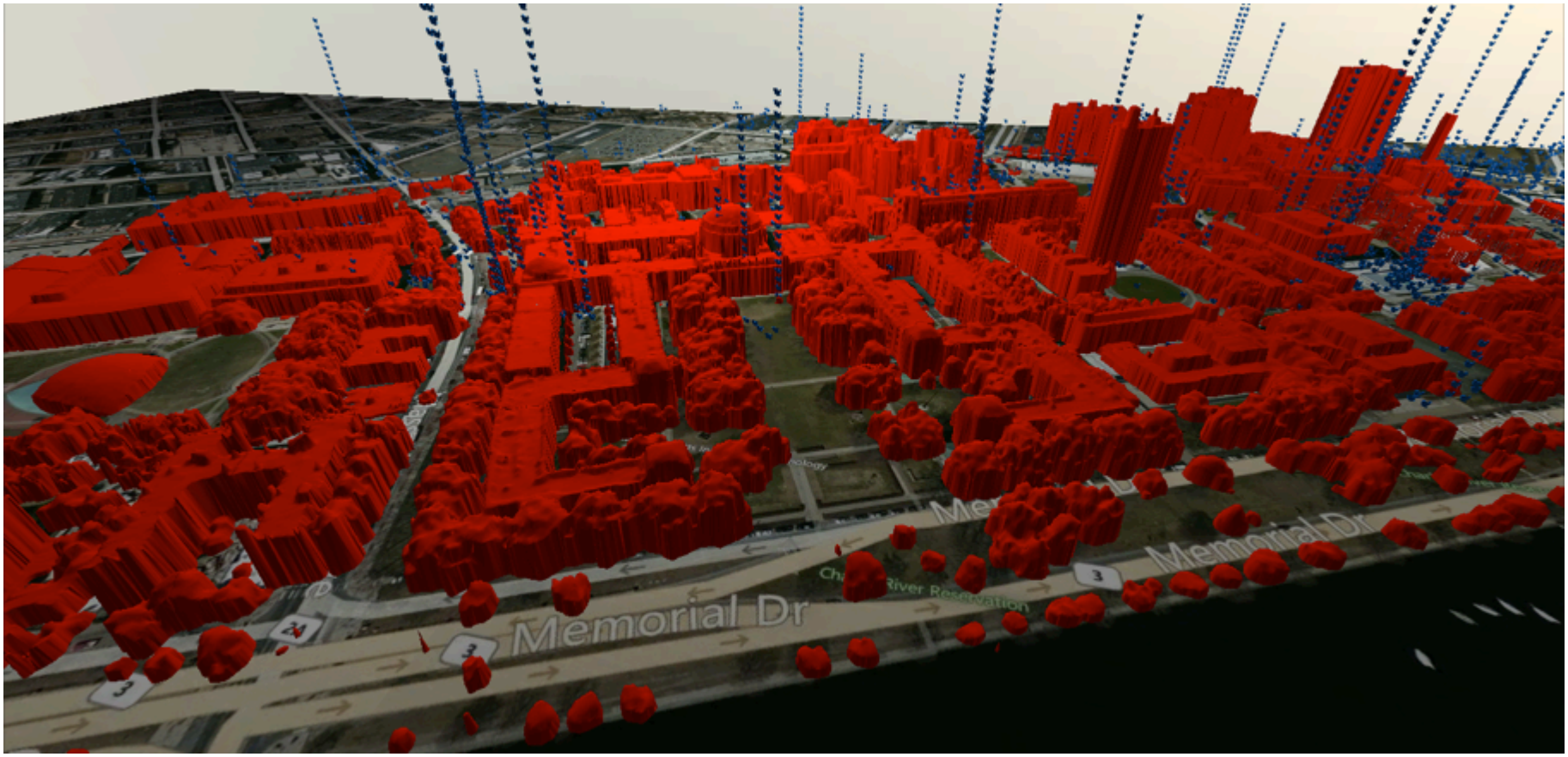
Game Configuration



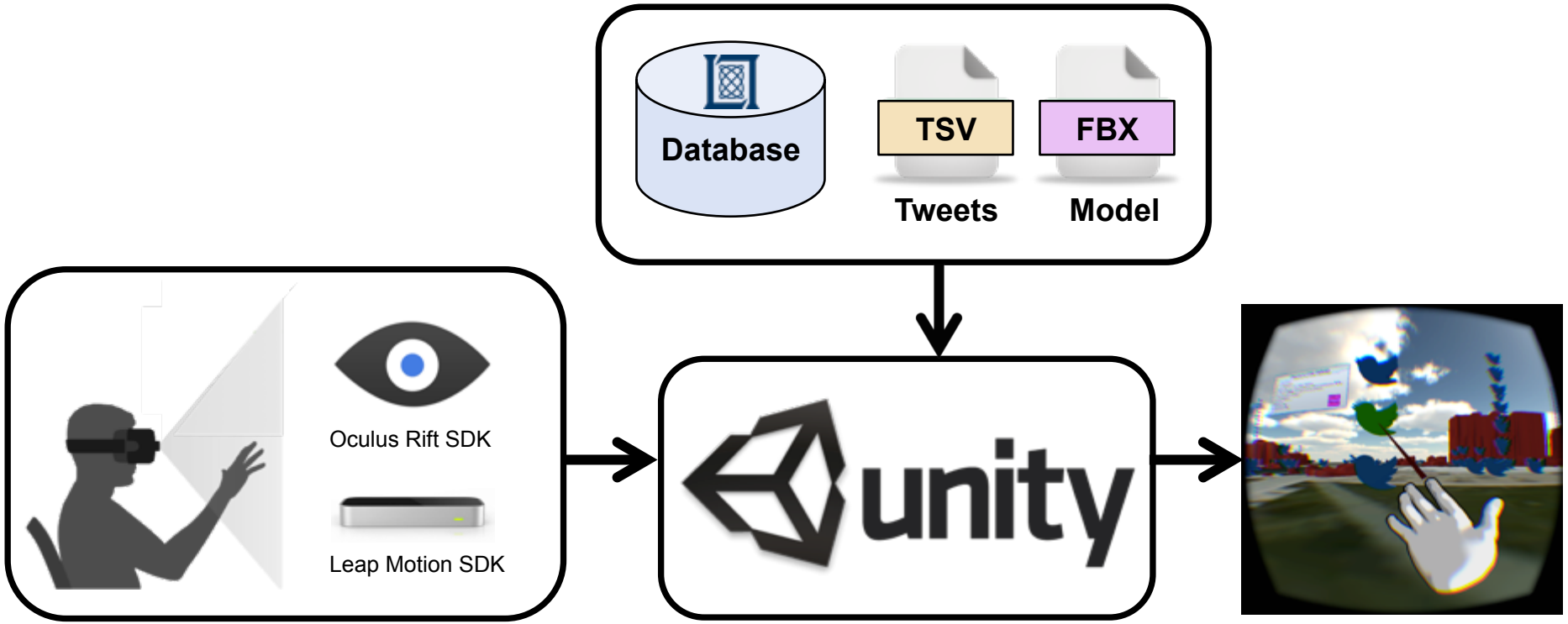
$$\{x,z\} = \frac{MAX\_\{lon,lat\} - \{lon,lat\}}{MAX\_\{lon,lat\} - MIN\_\{lon,lat\}} \times \{WIDTH, HEIGHT\}$$

Game Configuration

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



Utilized Technologies



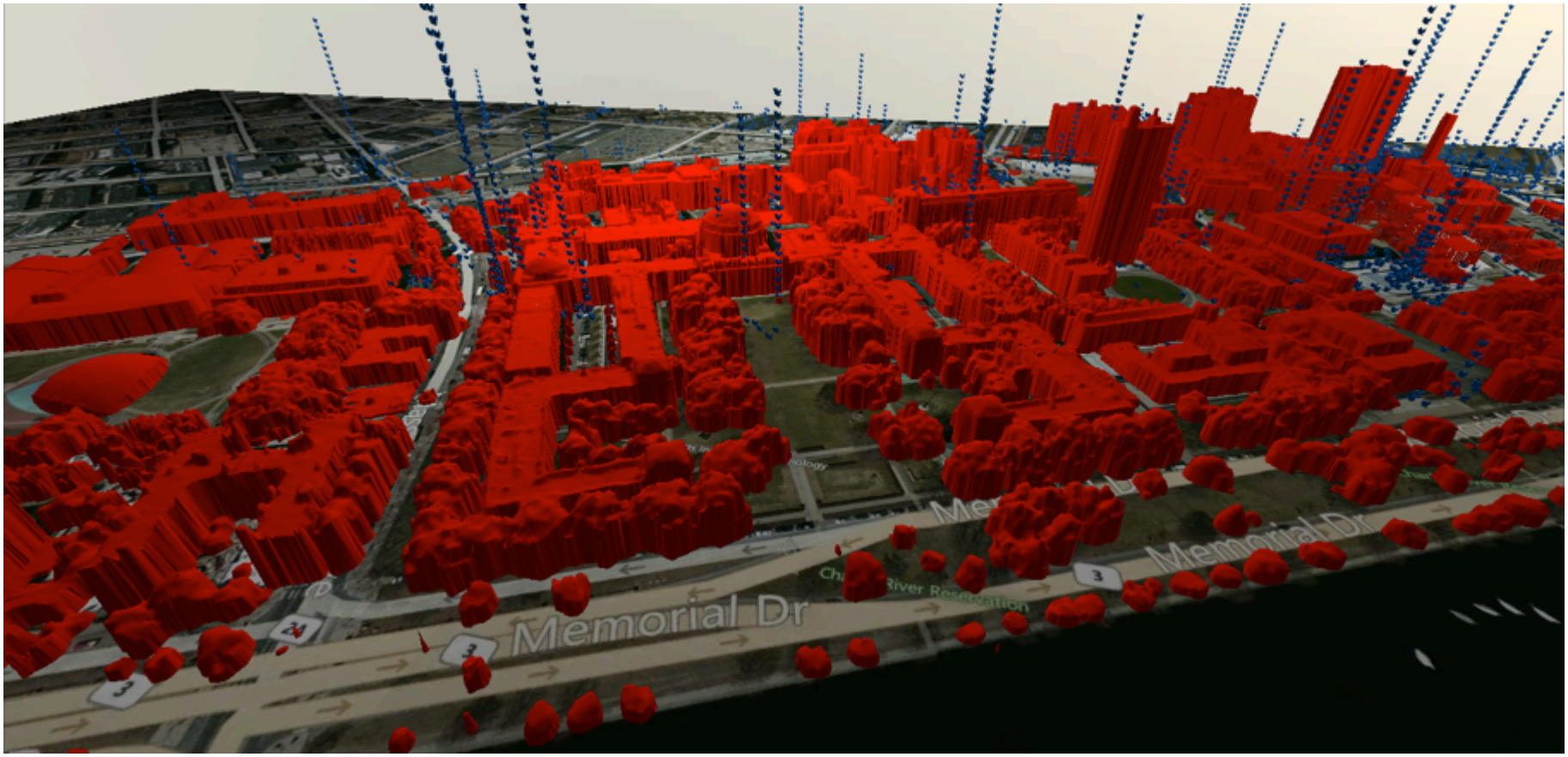
"3D Input Meets 3D Output" **– Why choose Unity3D?**

- **Game Engine**
- **Physics Engine**
- **Extensive Integration**
- **Flexibility**

Outline

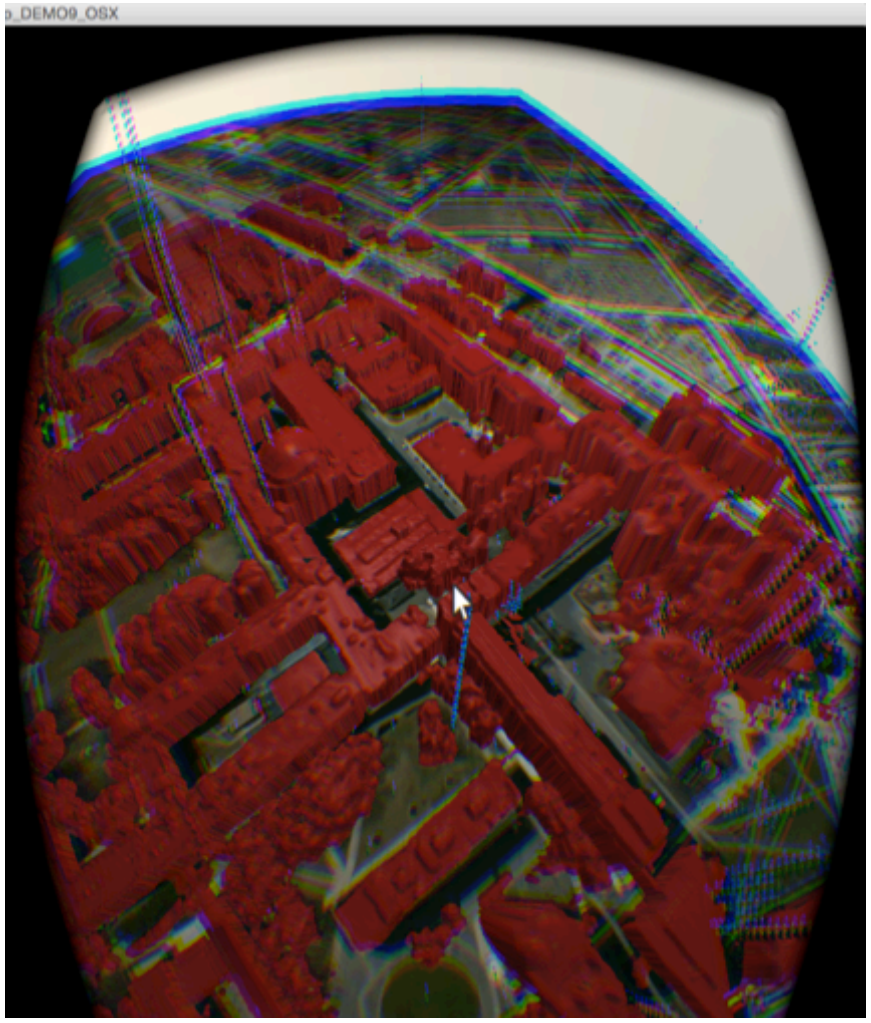
- Introduction
  - Big Data
  - Visualization
  - Virtual Reality
- Approach
  - Data Extraction
  - Game Configuration
  - Utilized Technologies
- **Results**
  - **Virtual Environment**
  - **Analytical Tasks**
  - **Challenges**
- Summary, Future Work and Next Steps

Virtual Environment



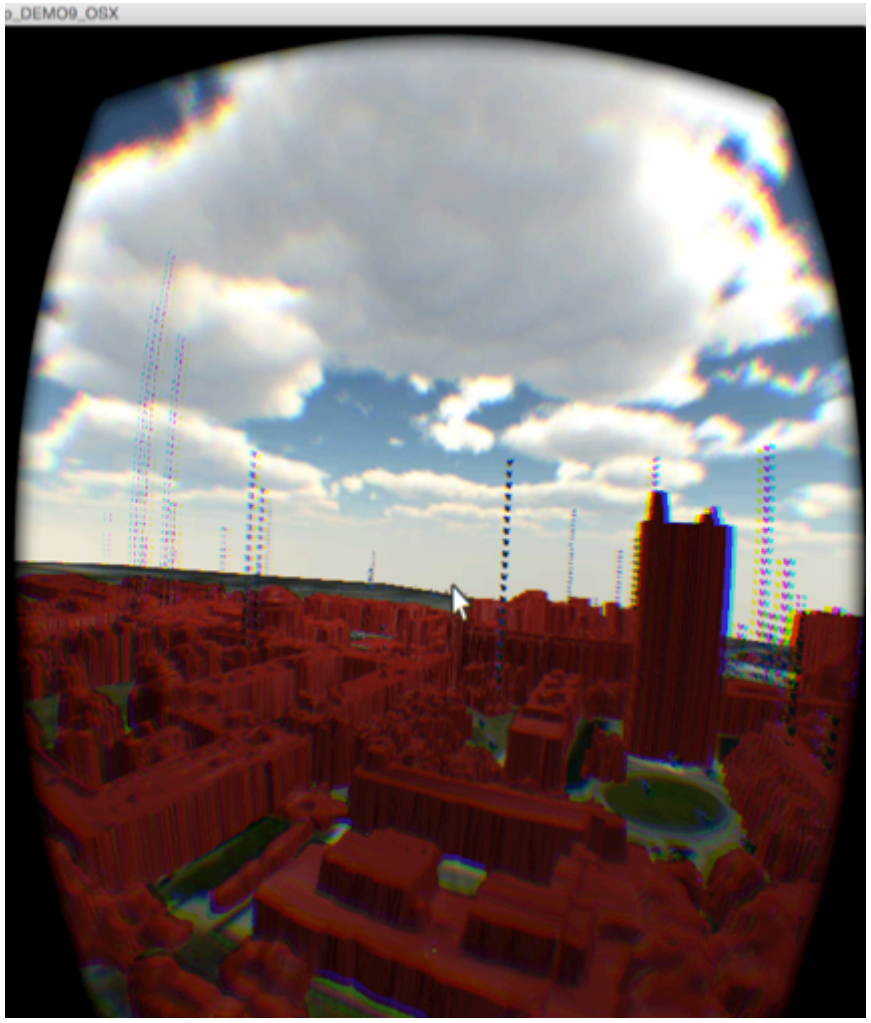
Analytical Tasks

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



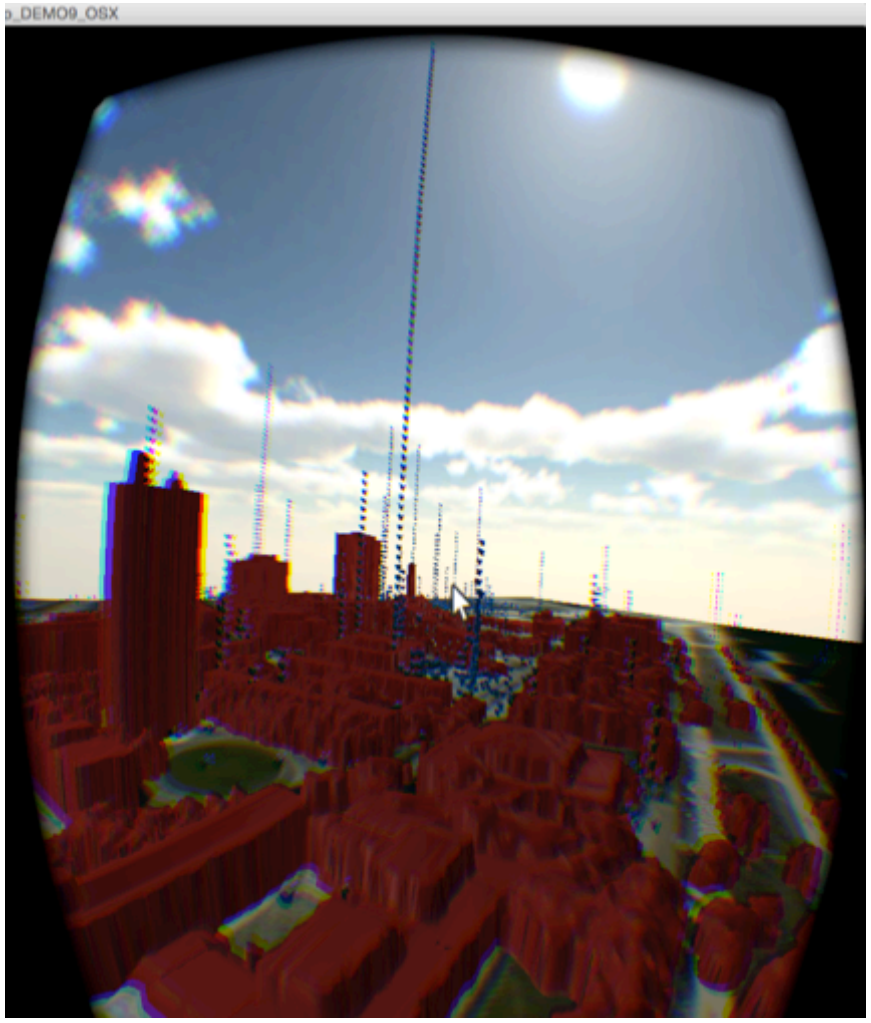
Analytical Tasks

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



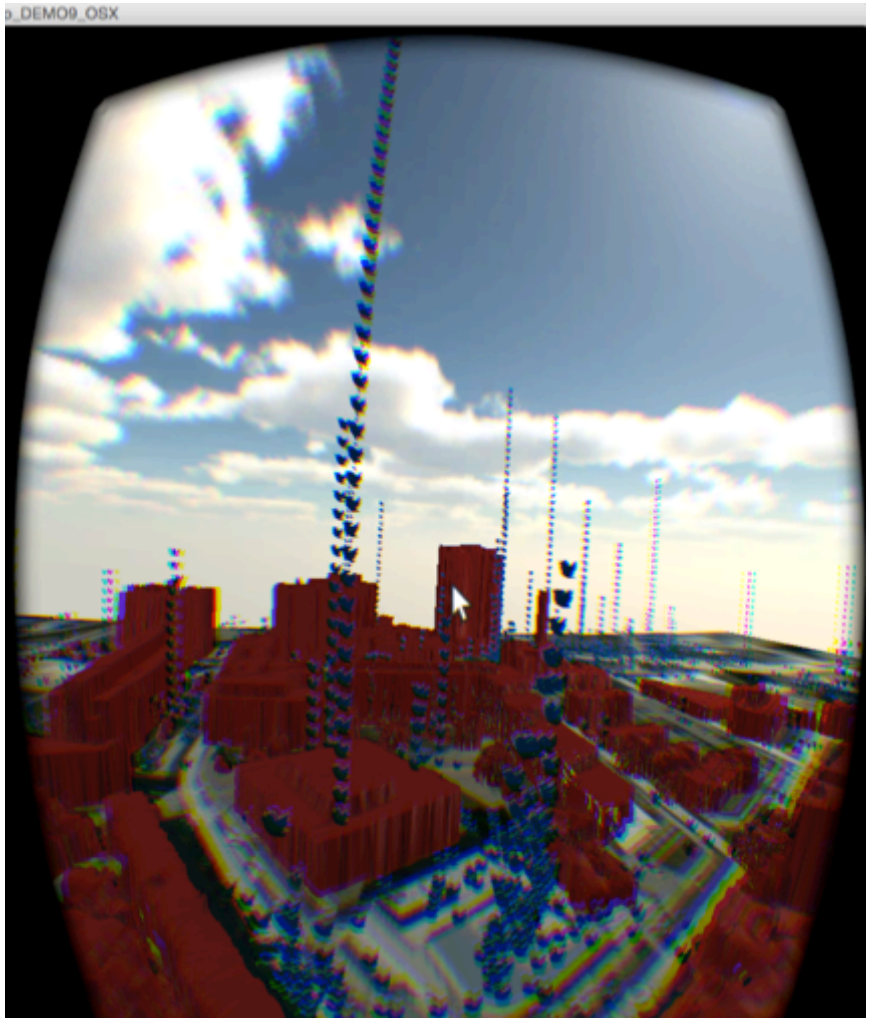
Analytical Tasks

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



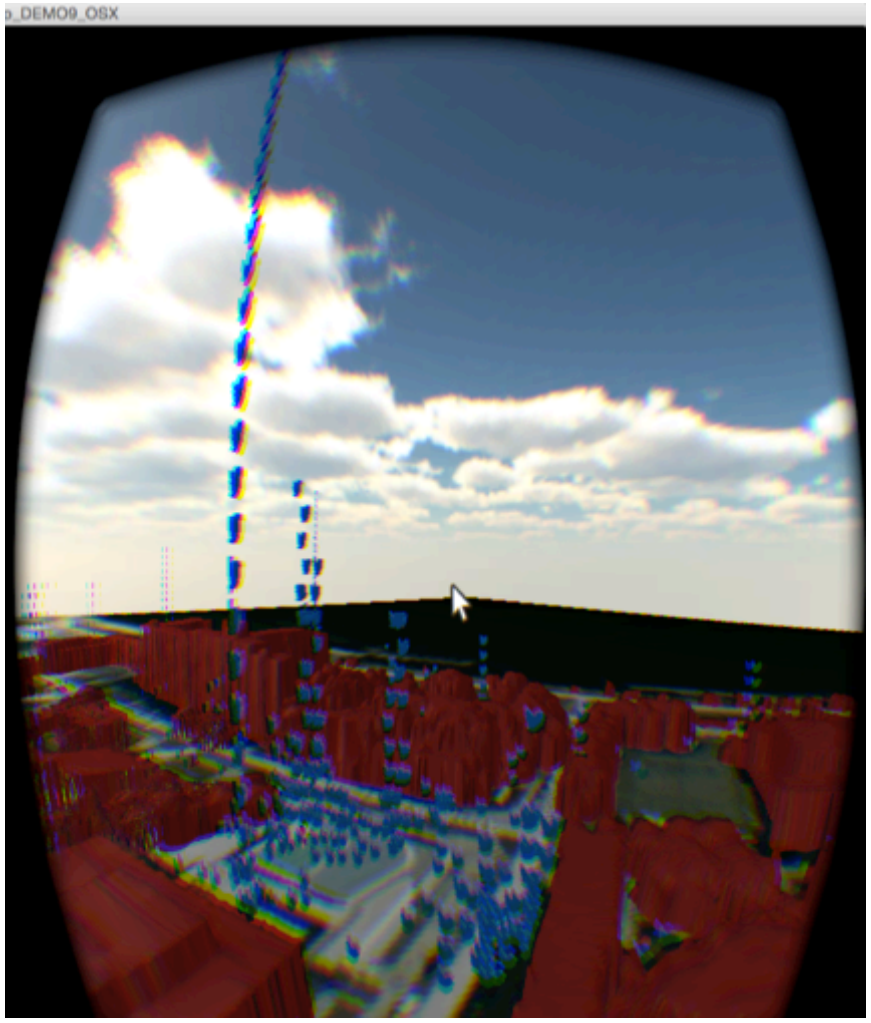
Analytical Tasks

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



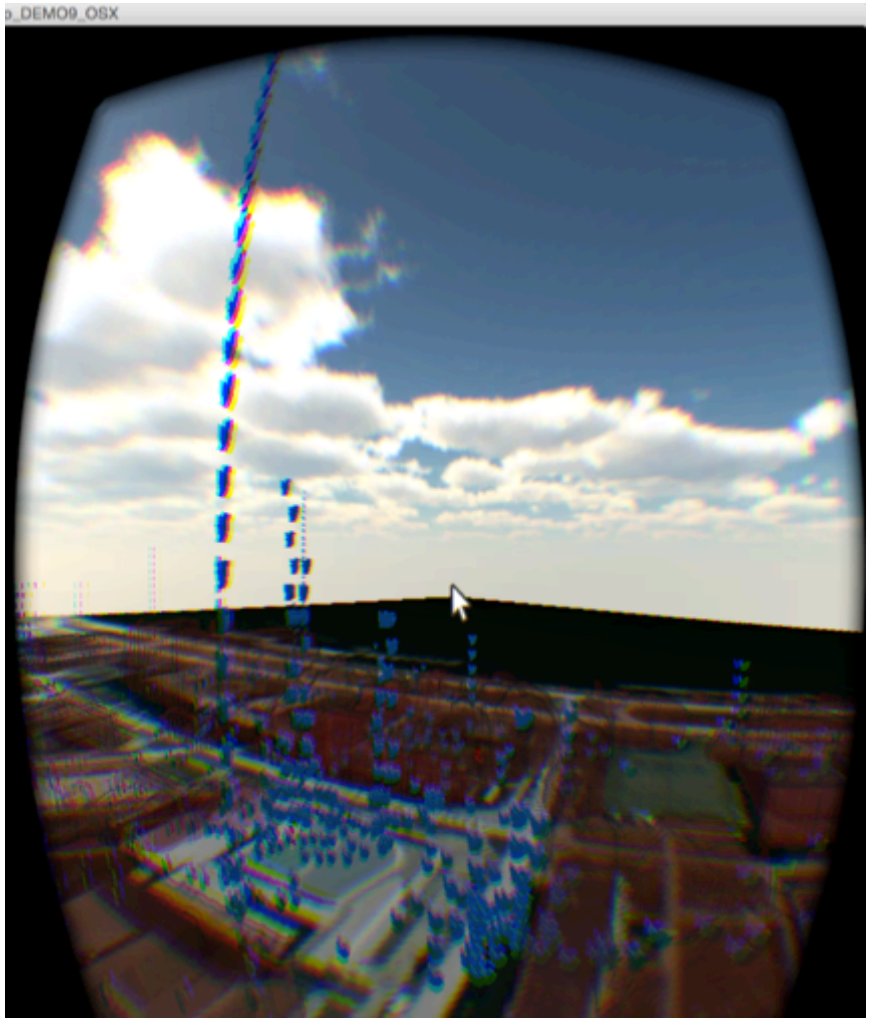
Analytical Tasks

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



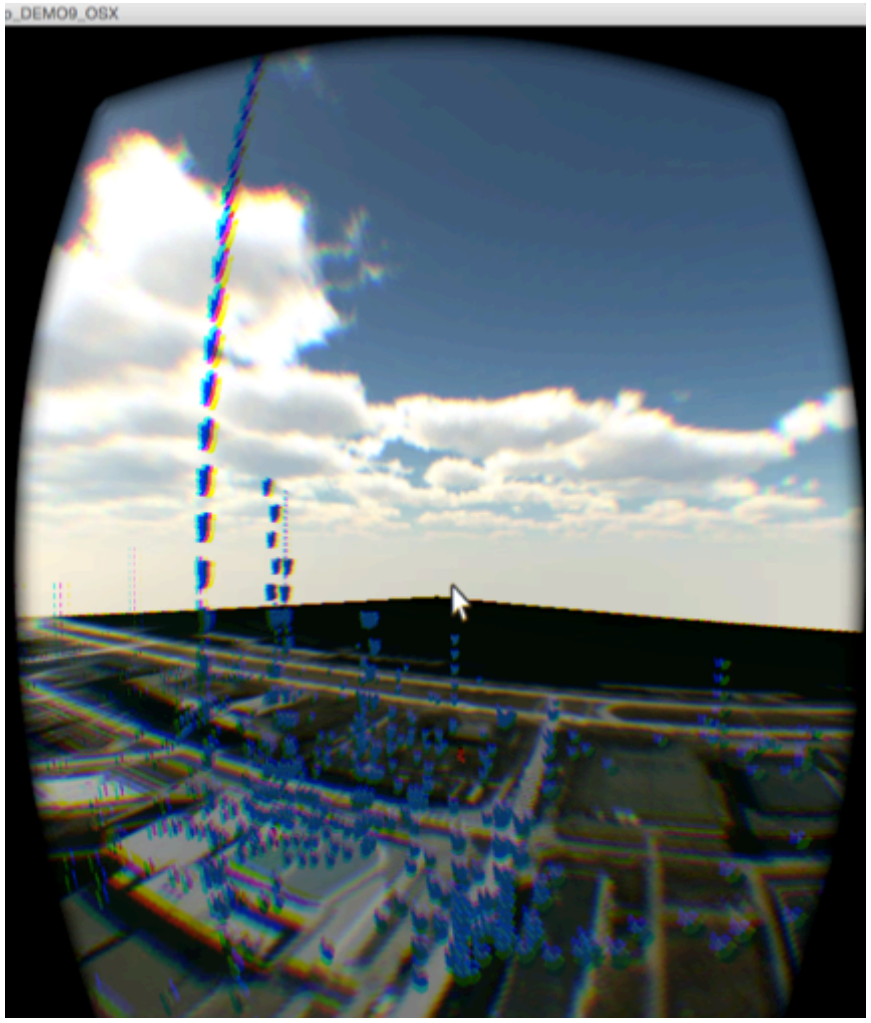
Analytical Tasks

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



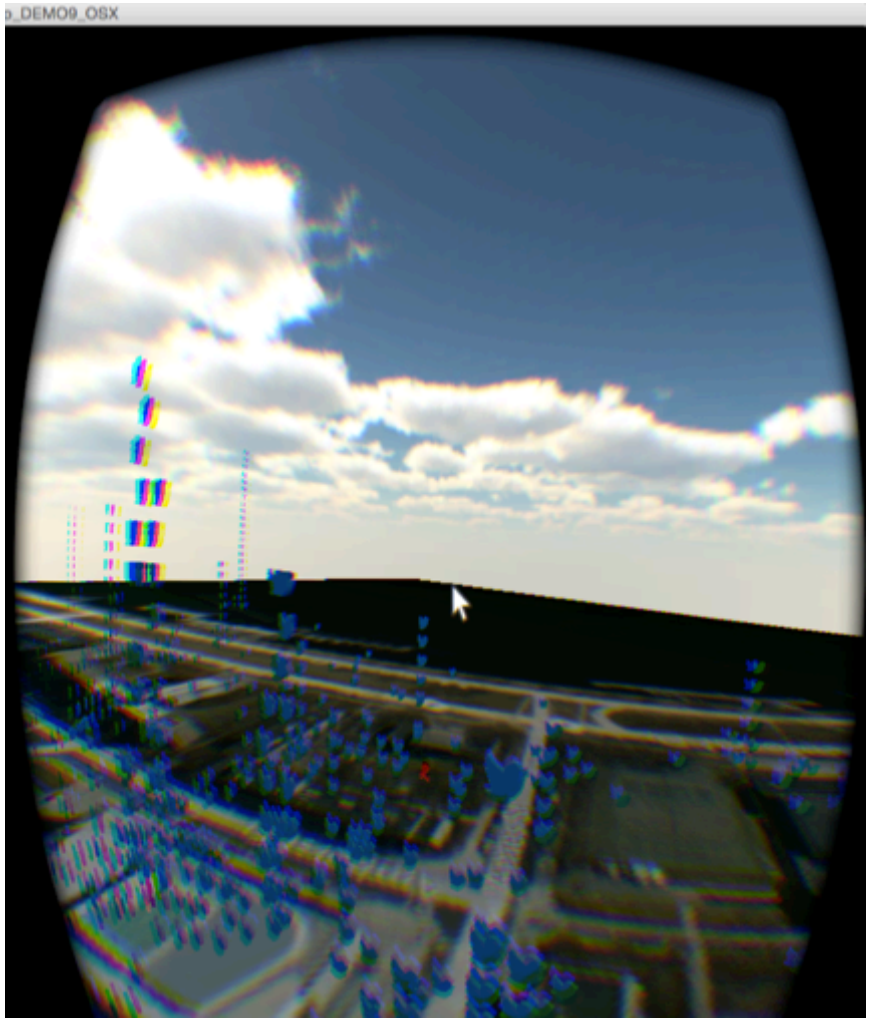
Analytical Tasks

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



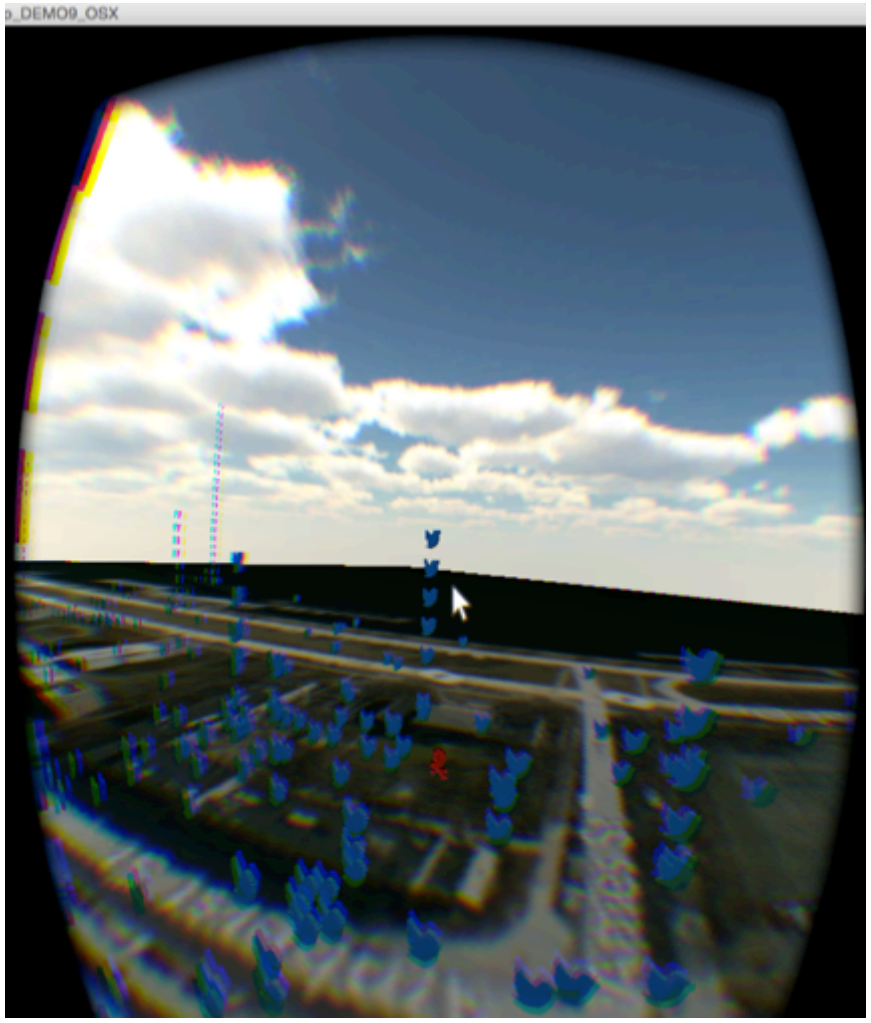
Analytical Tasks

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



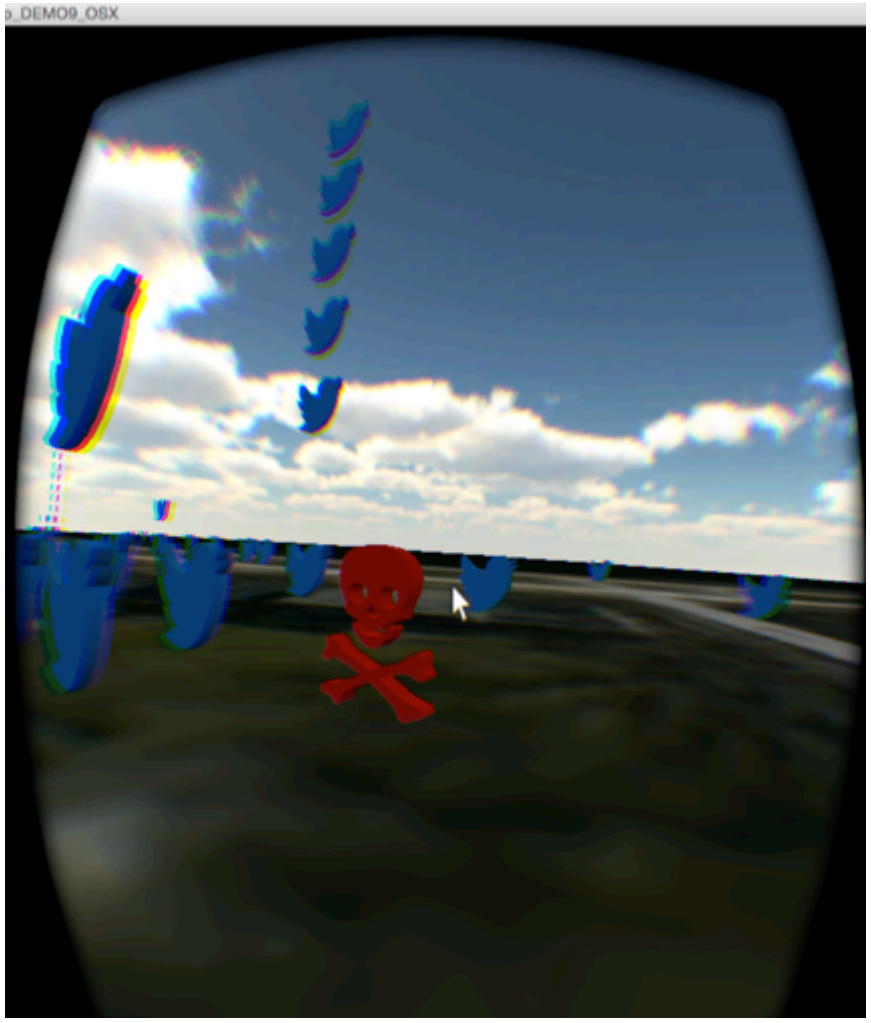
Analytical Tasks

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



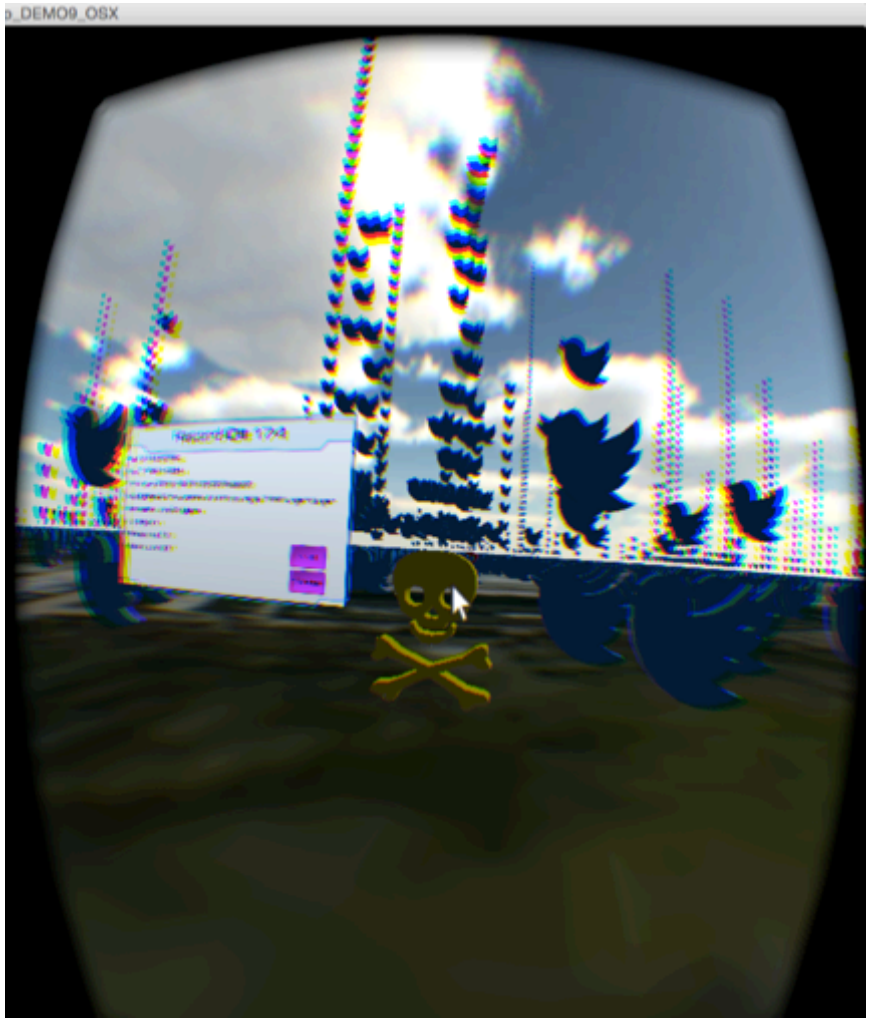
Analytical Tasks

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



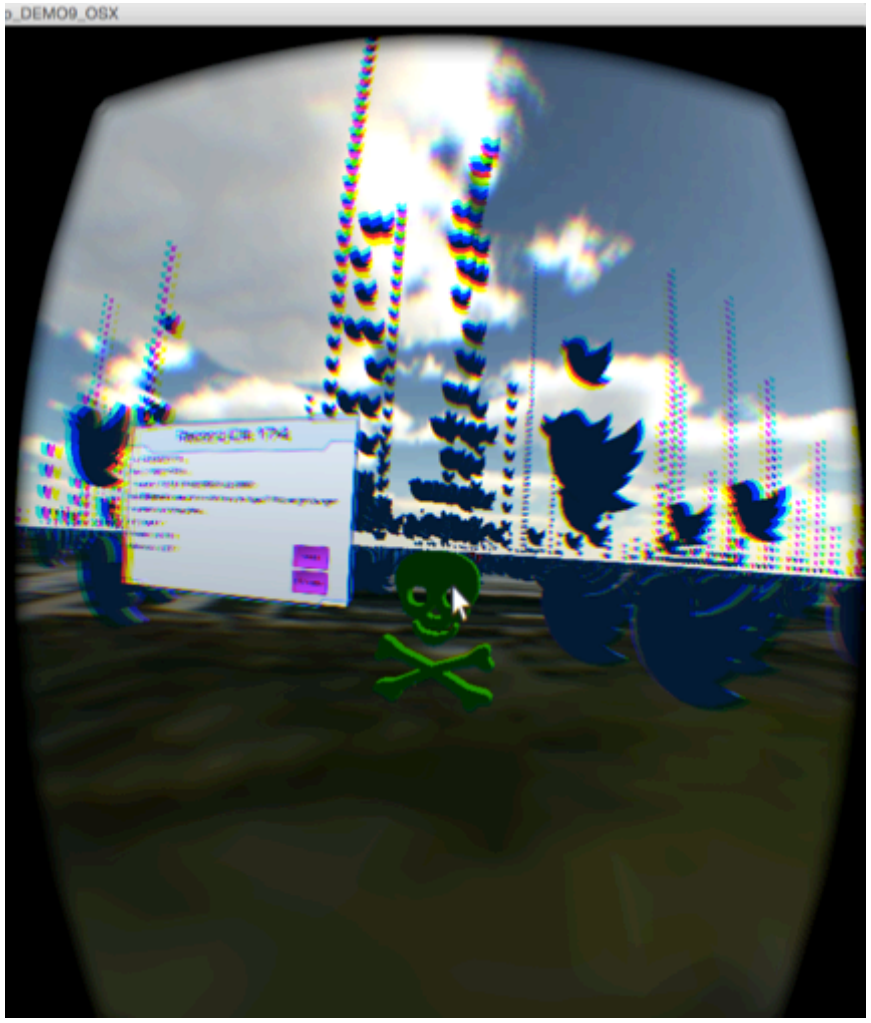
Analytical Tasks

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



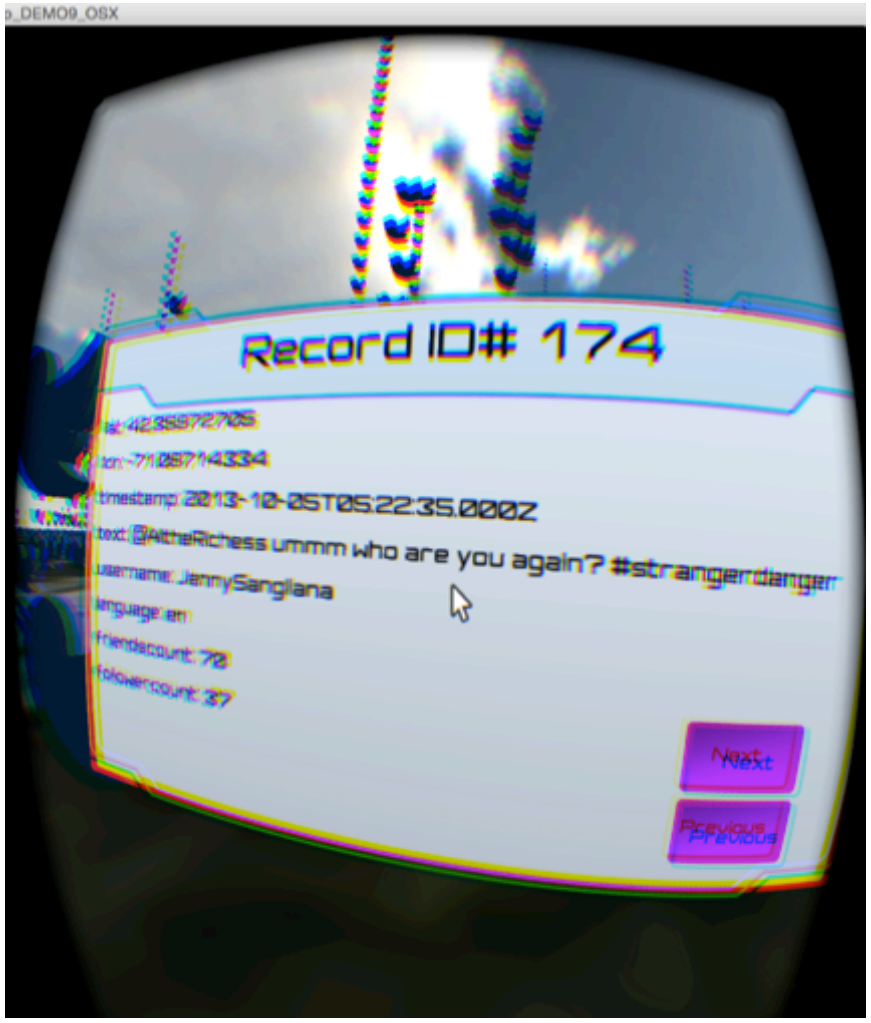
Analytical Tasks

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



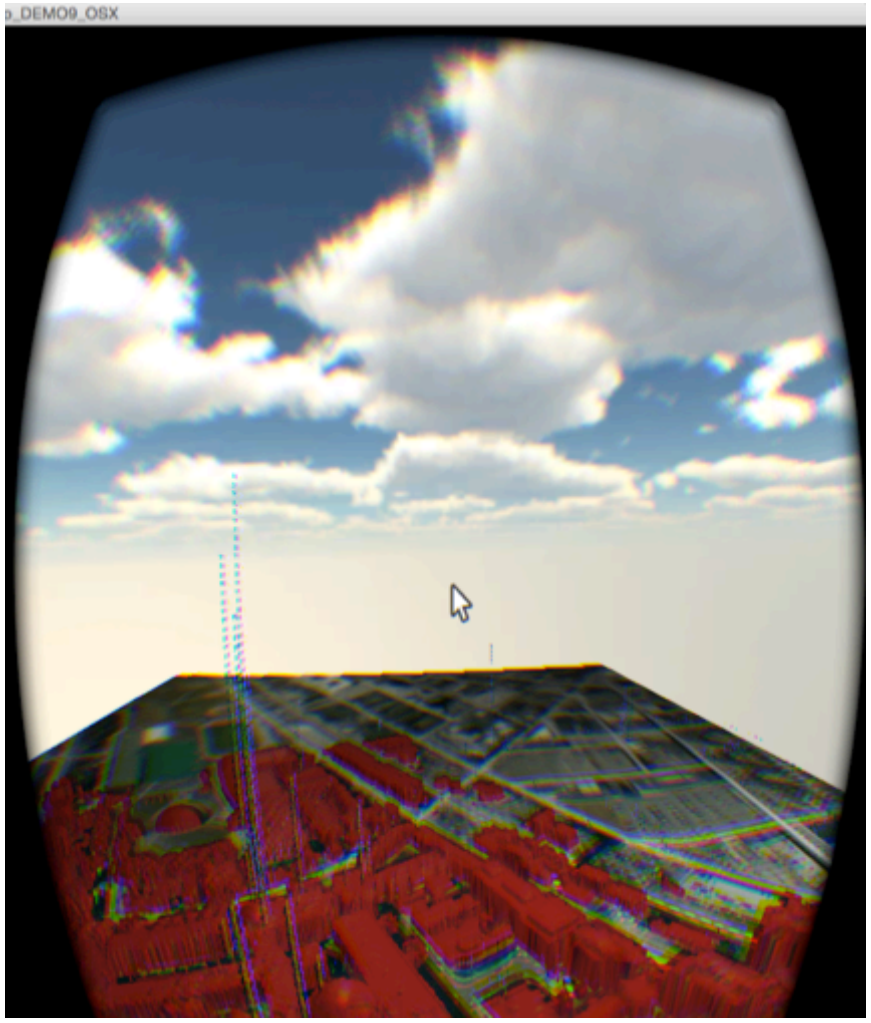
Analytical Tasks

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



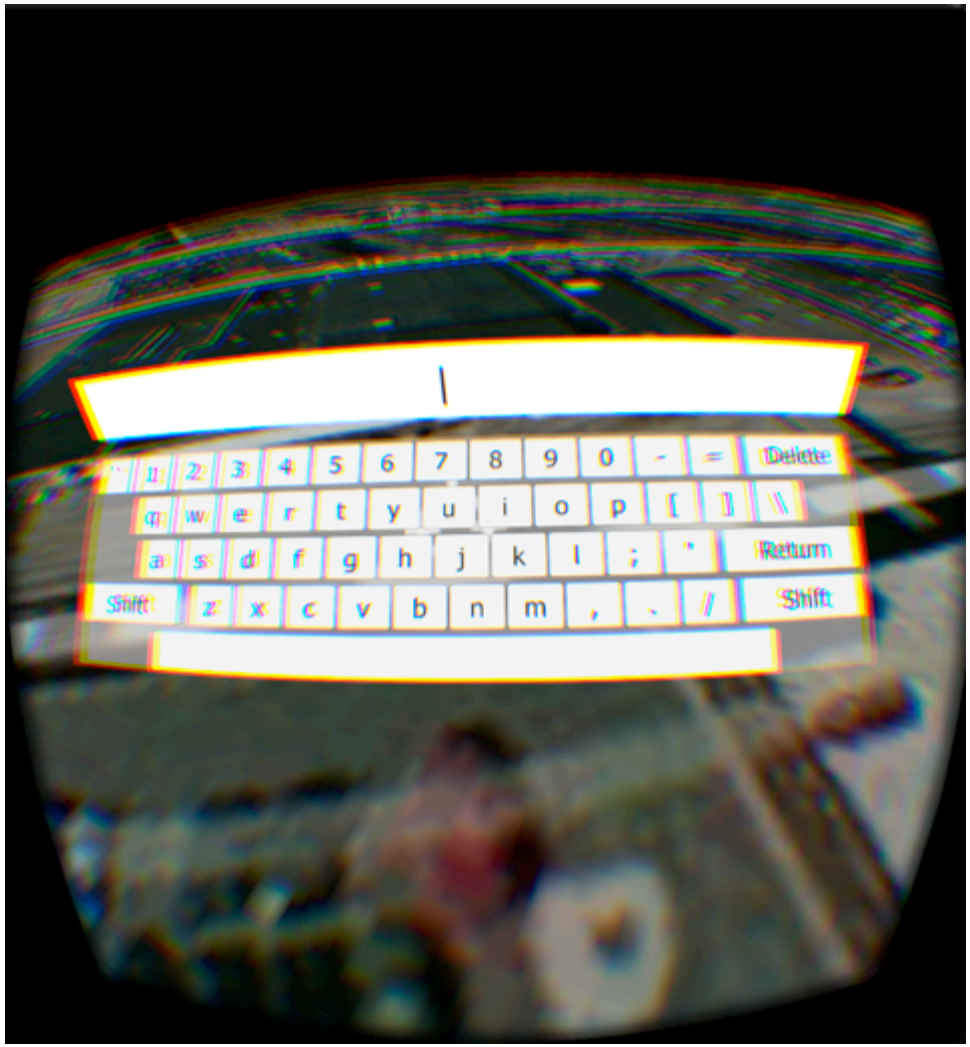
Analytical Tasks

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



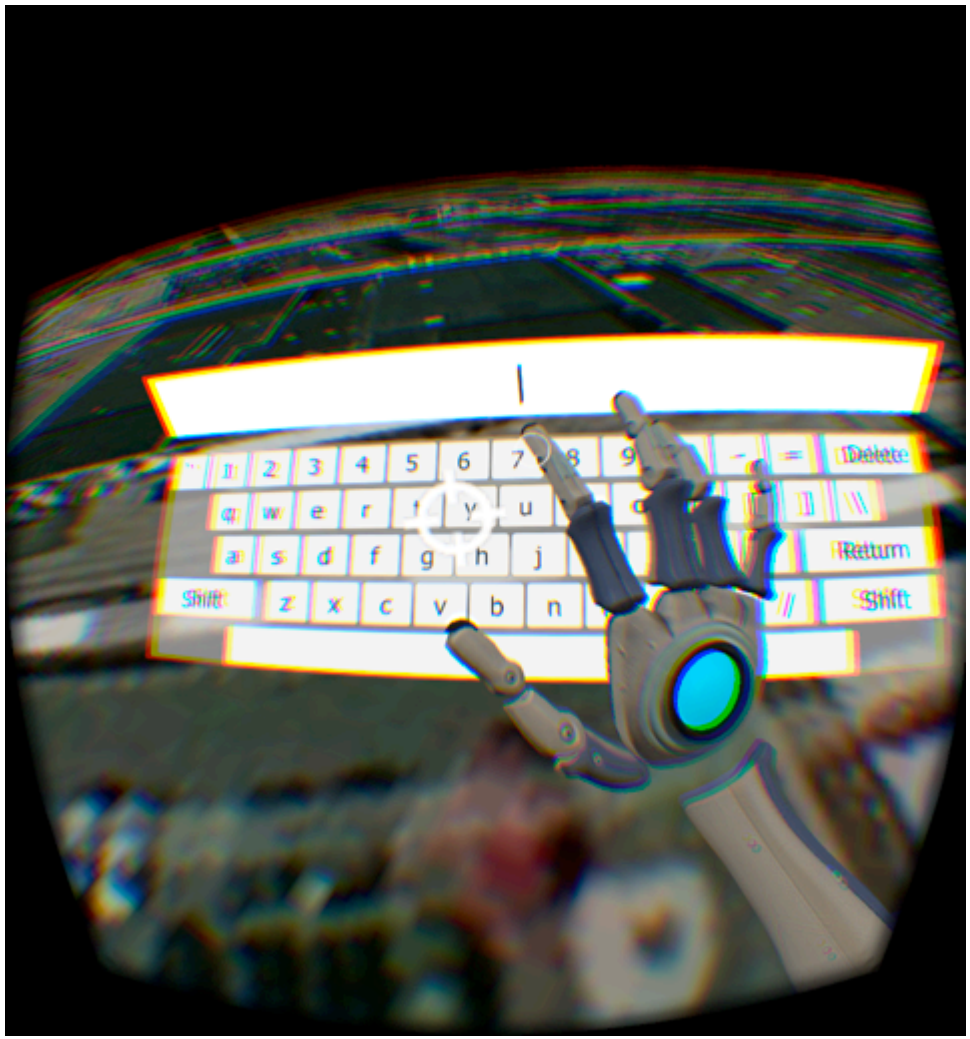
Analytical Tasks

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



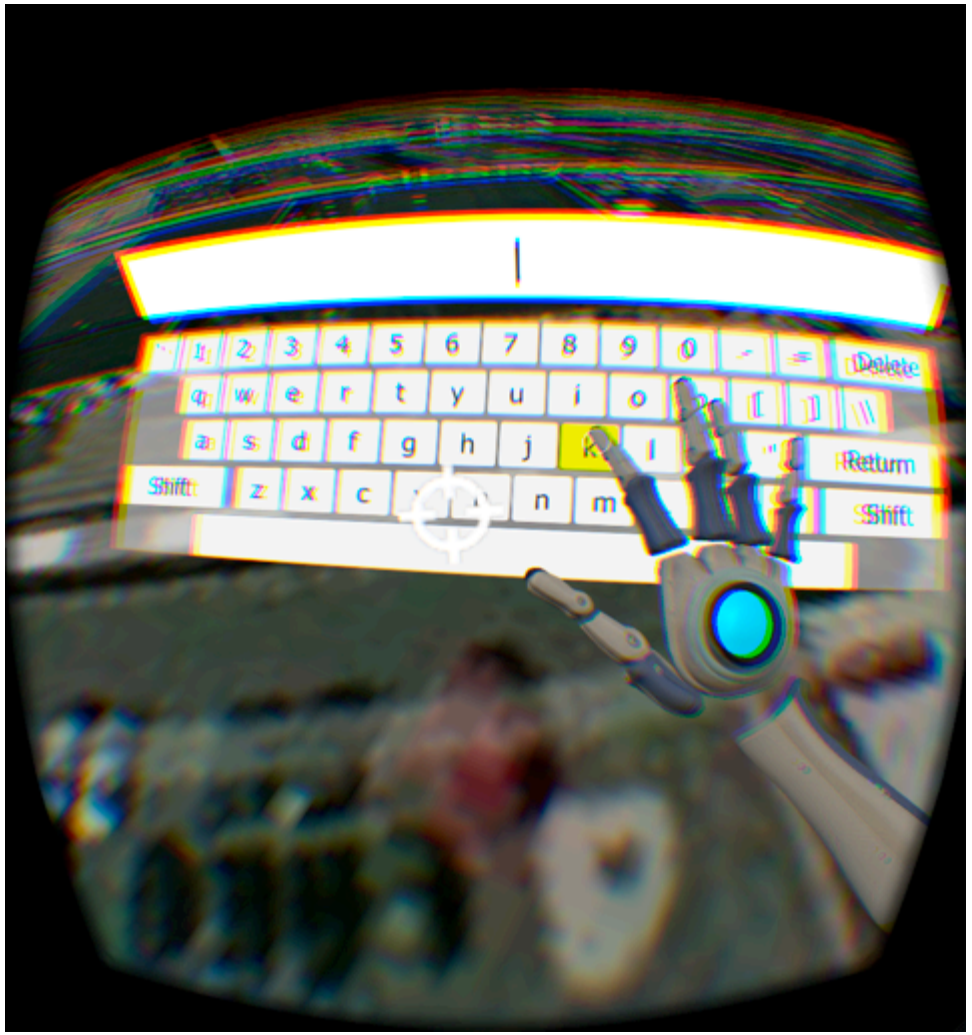
Analytical Tasks

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



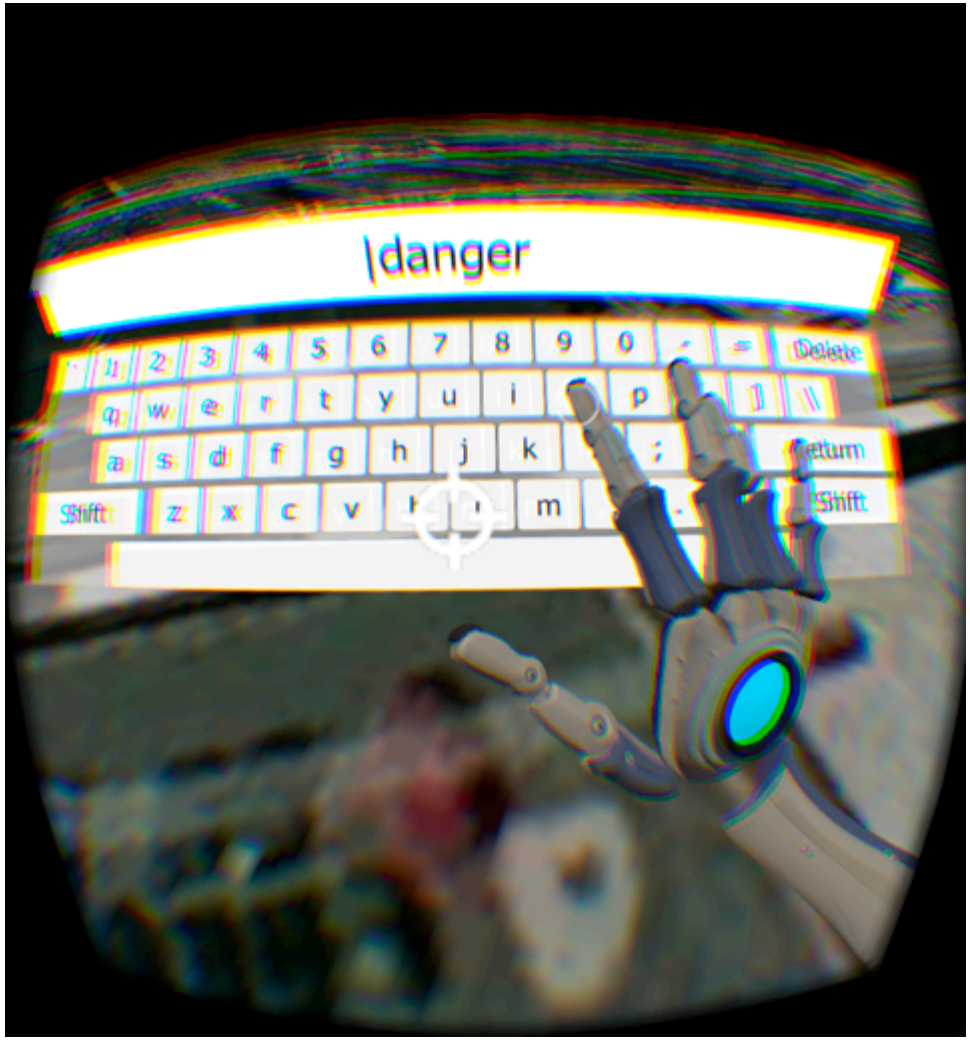
Analytical Tasks

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



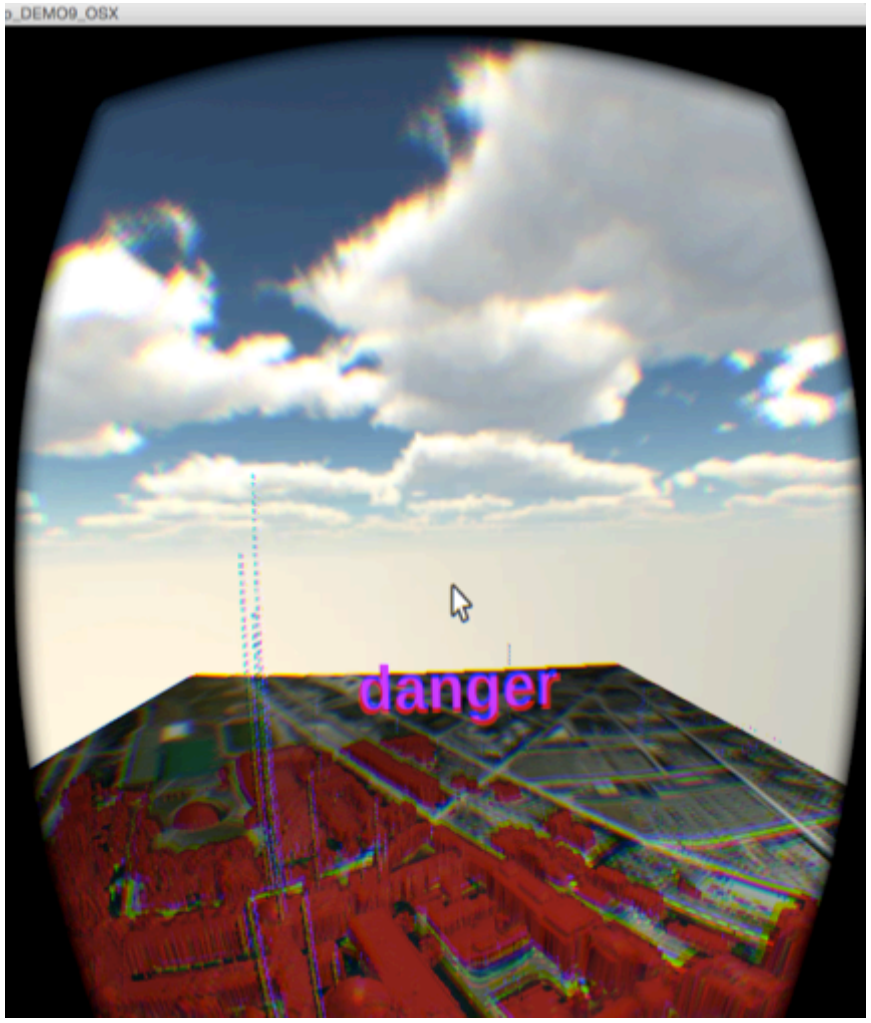
Analytical Tasks

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



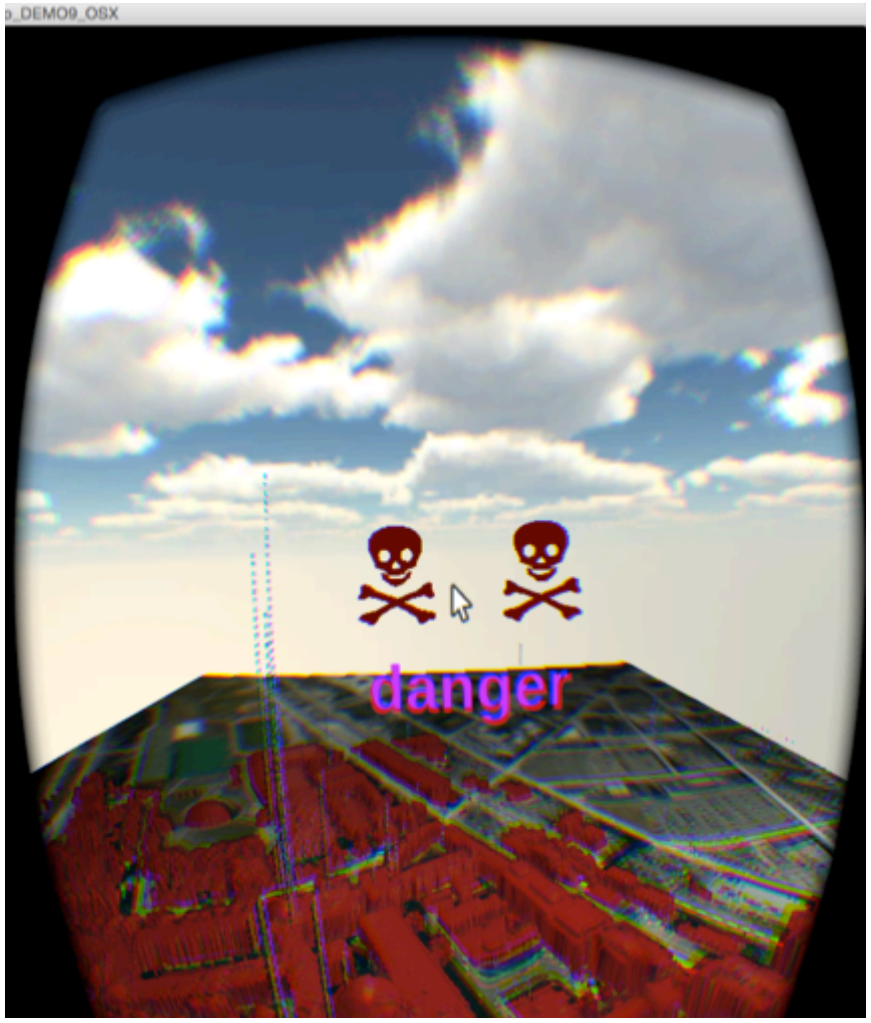
Analytical Tasks

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



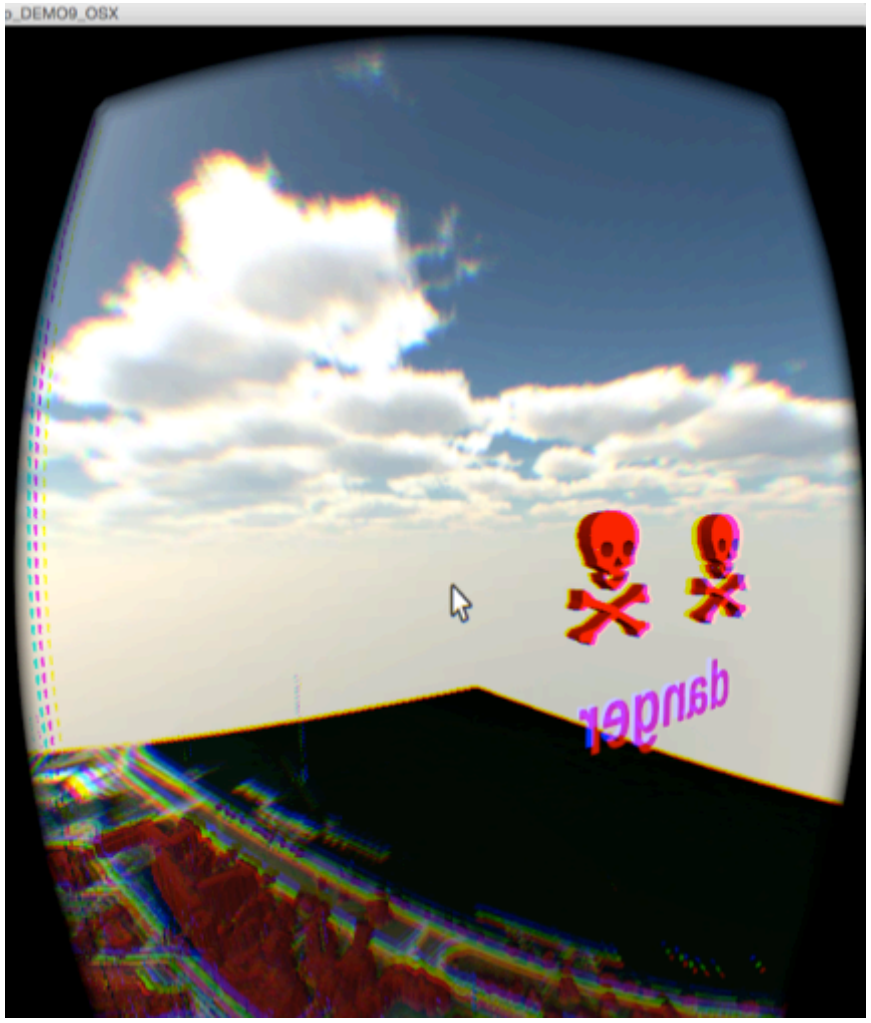
Analytical Tasks

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



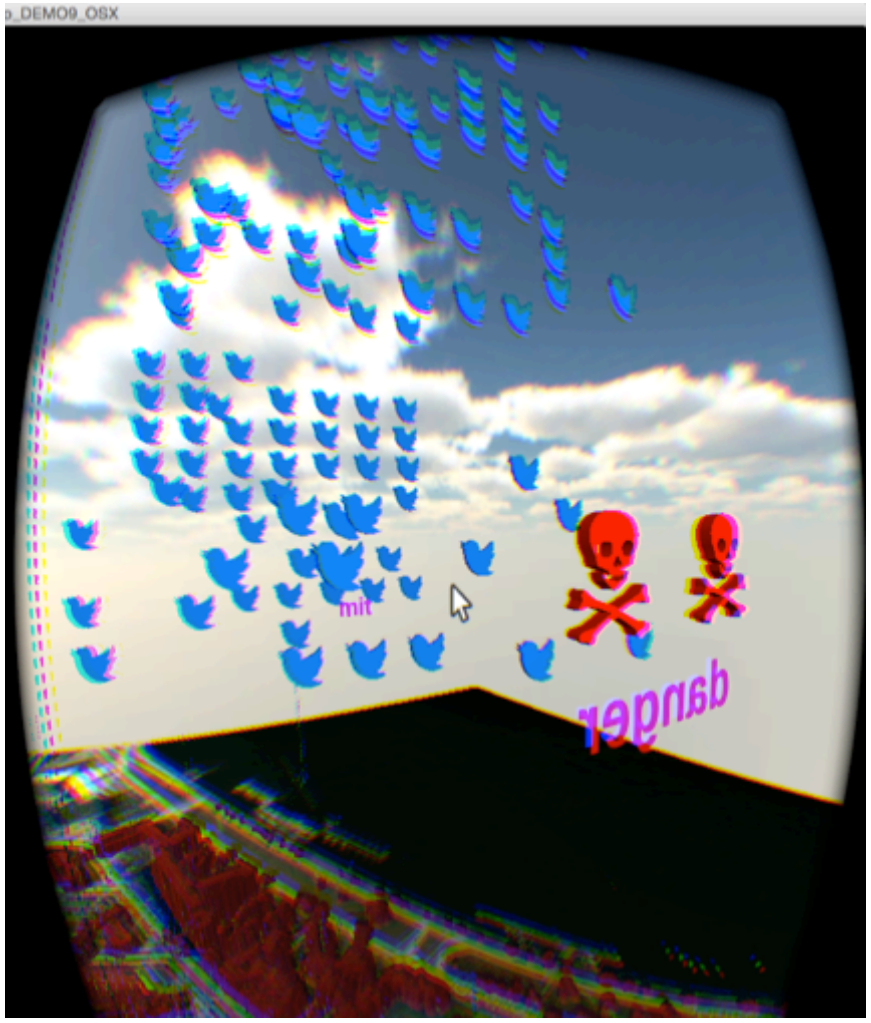
Analytical Tasks

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



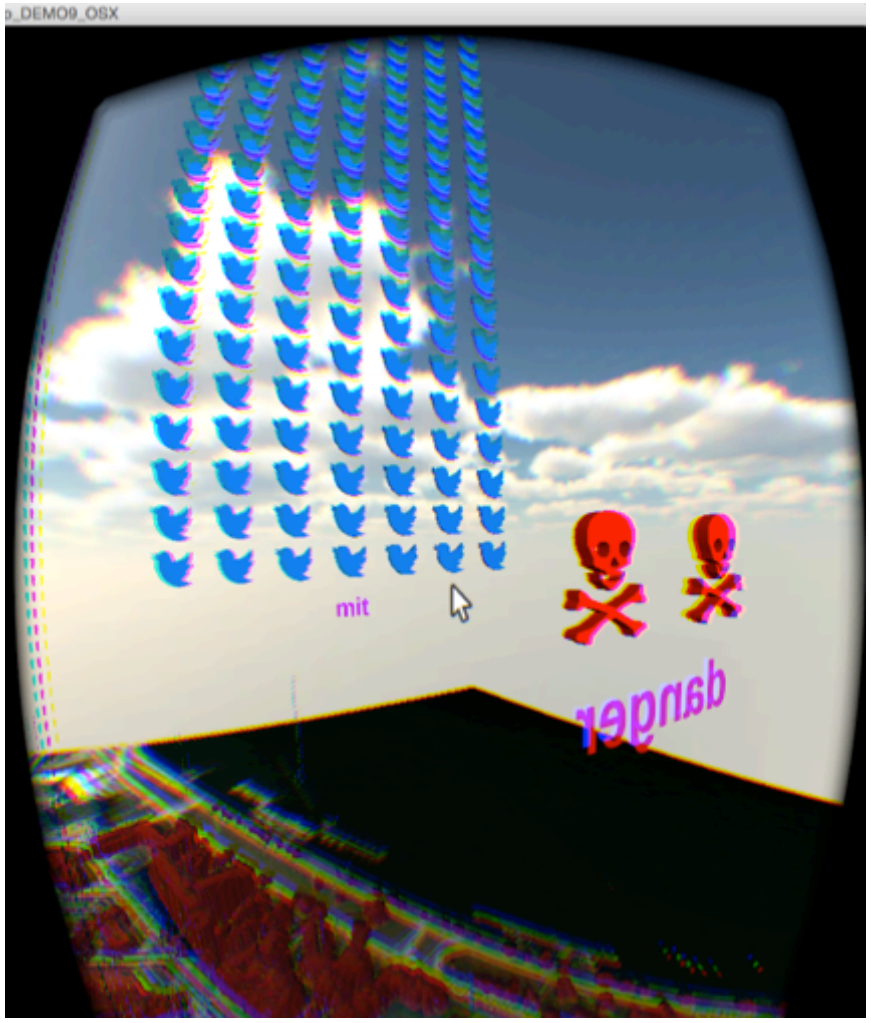
Analytical Tasks

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



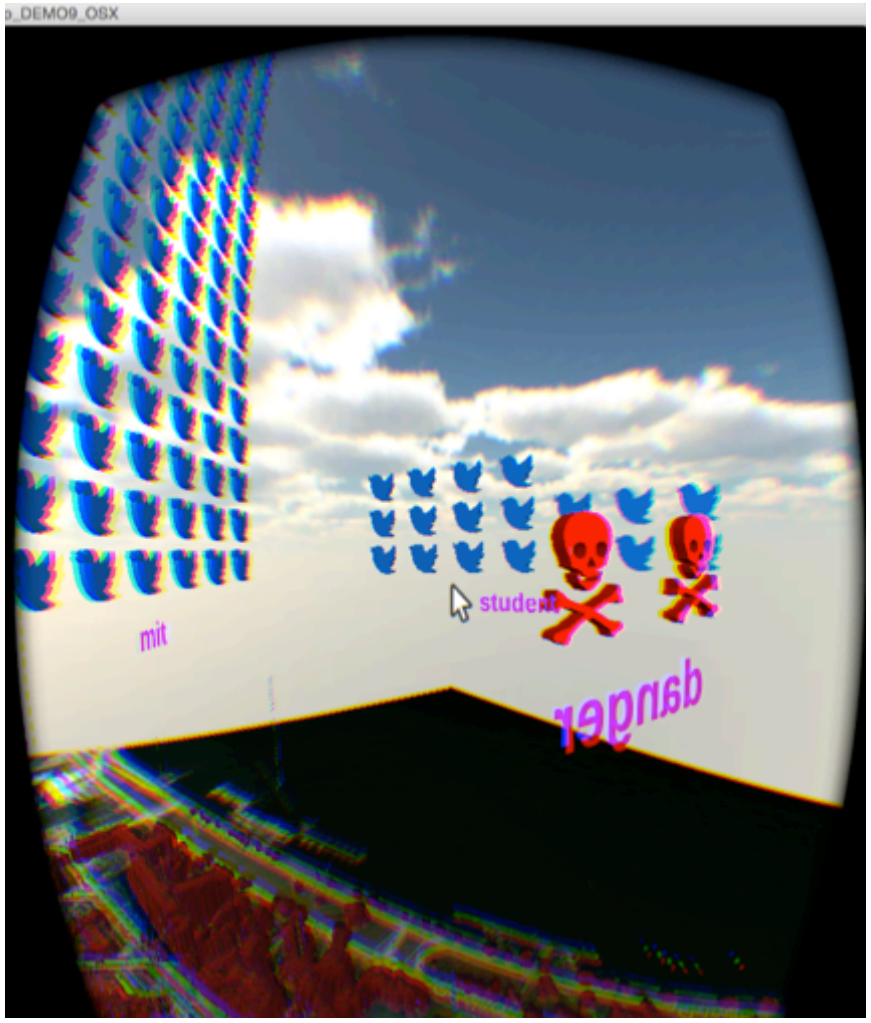
Analytical Tasks

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



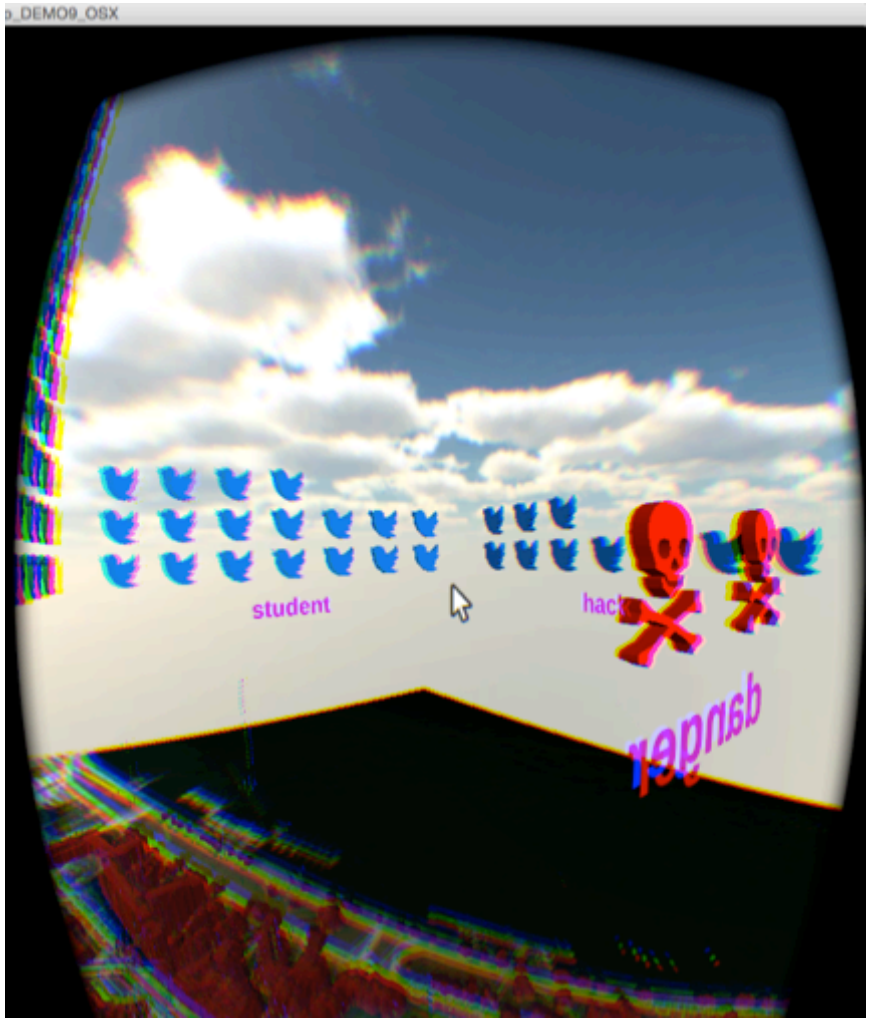
Analytical Tasks

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



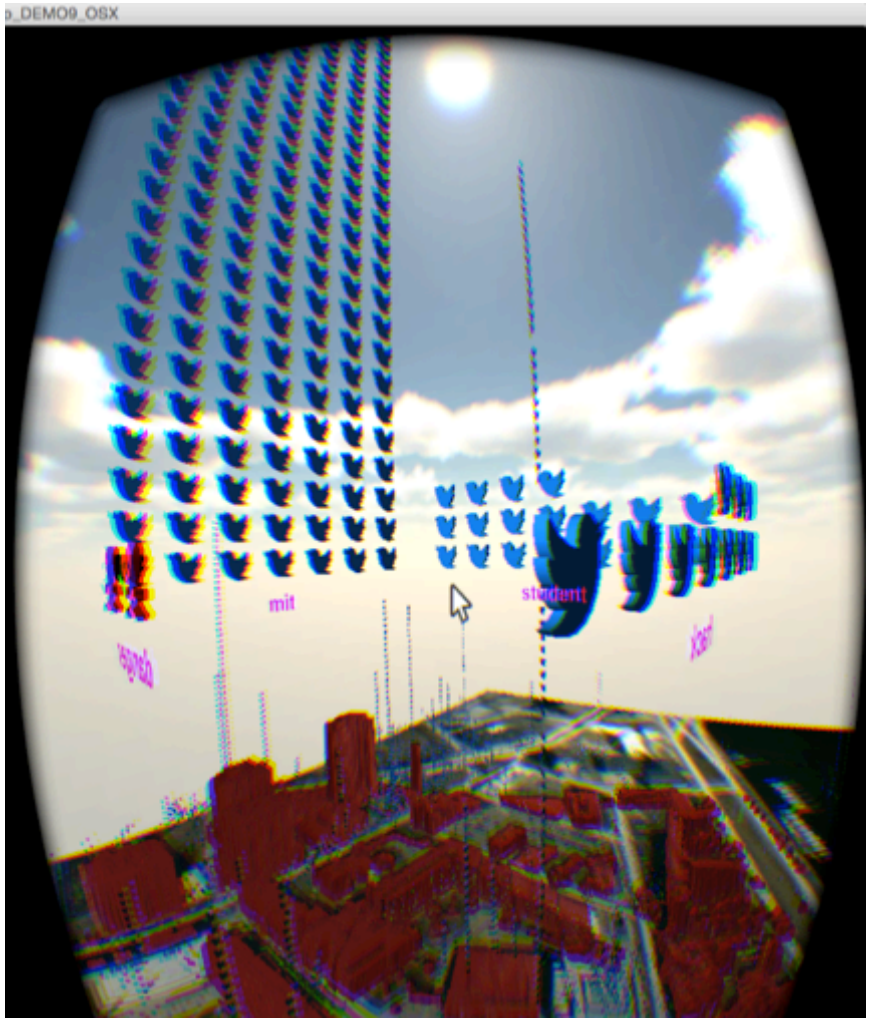
Analytical Tasks

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



Analytical Tasks

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



Analytical Tasks

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

Analytical Tasks

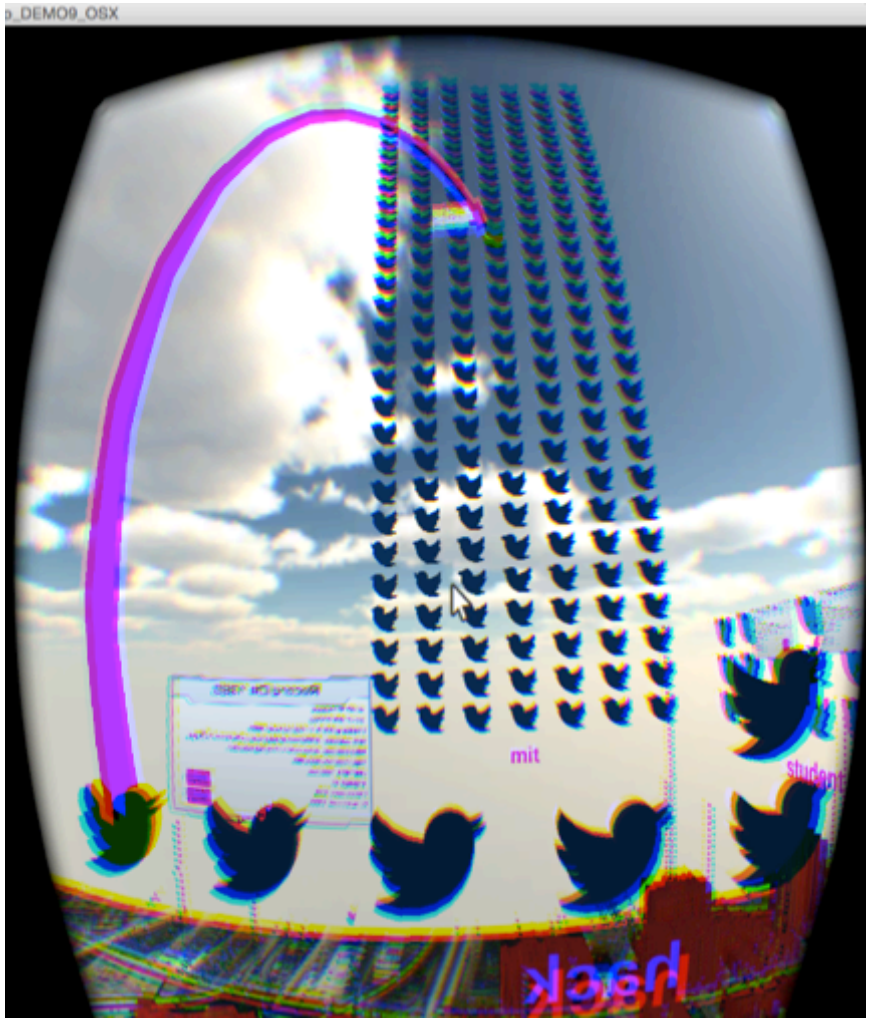
- Navigation/Exploration
- Cluster Recognition
- Filtering
- Identification
- Selection
- Querying
- **Relationships**
- Tracking

Analytical Tasks

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

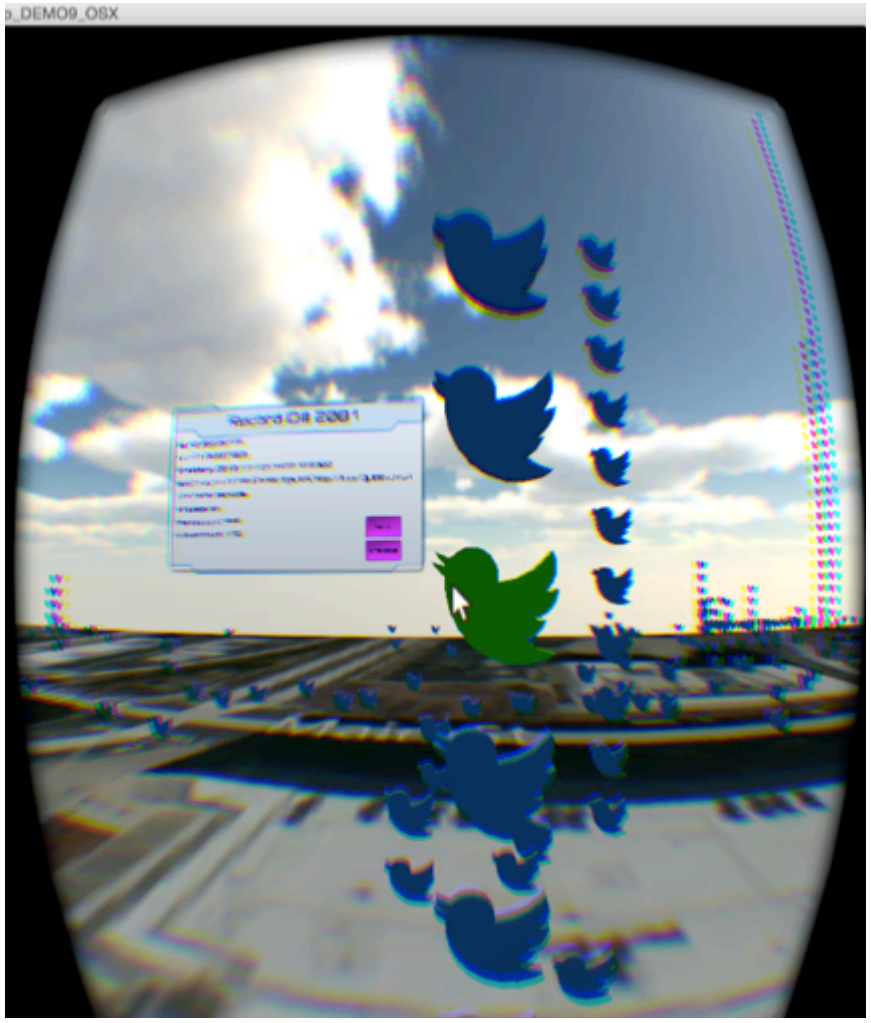
Analytical Tasks

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



Analytical Tasks

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

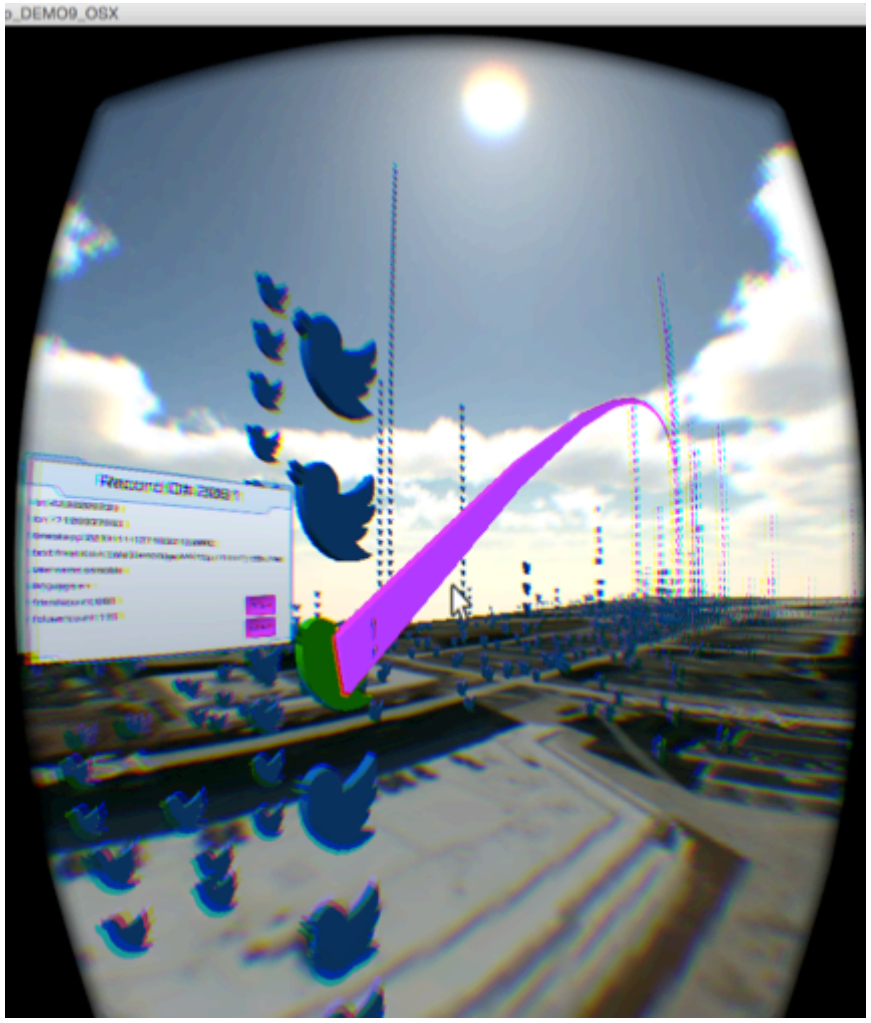


Analytical Tasks

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

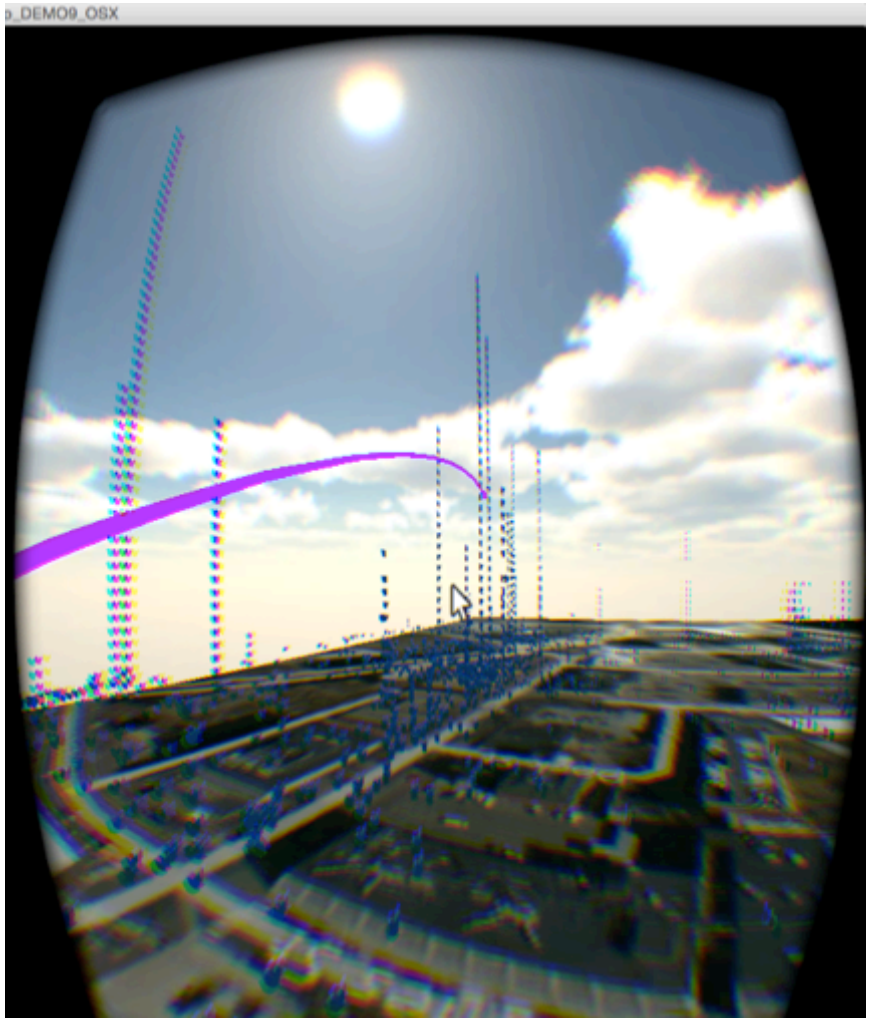
Analytical Tasks

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



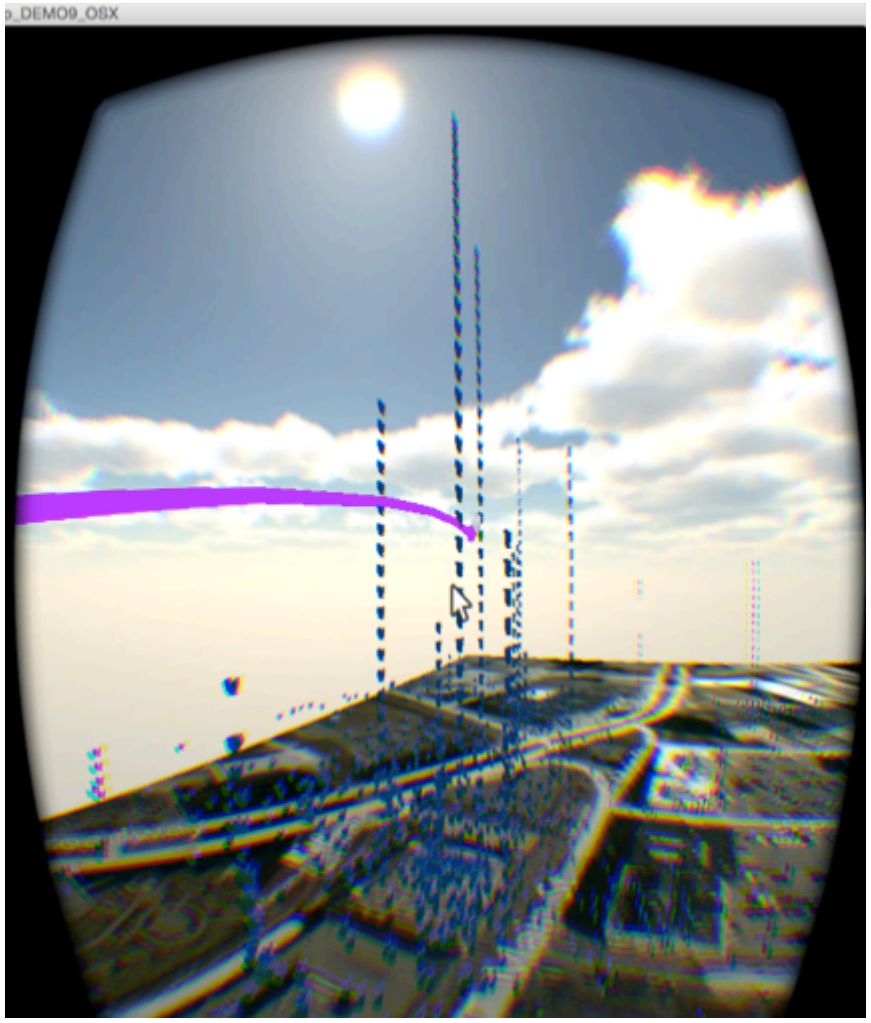
Analytical Tasks

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



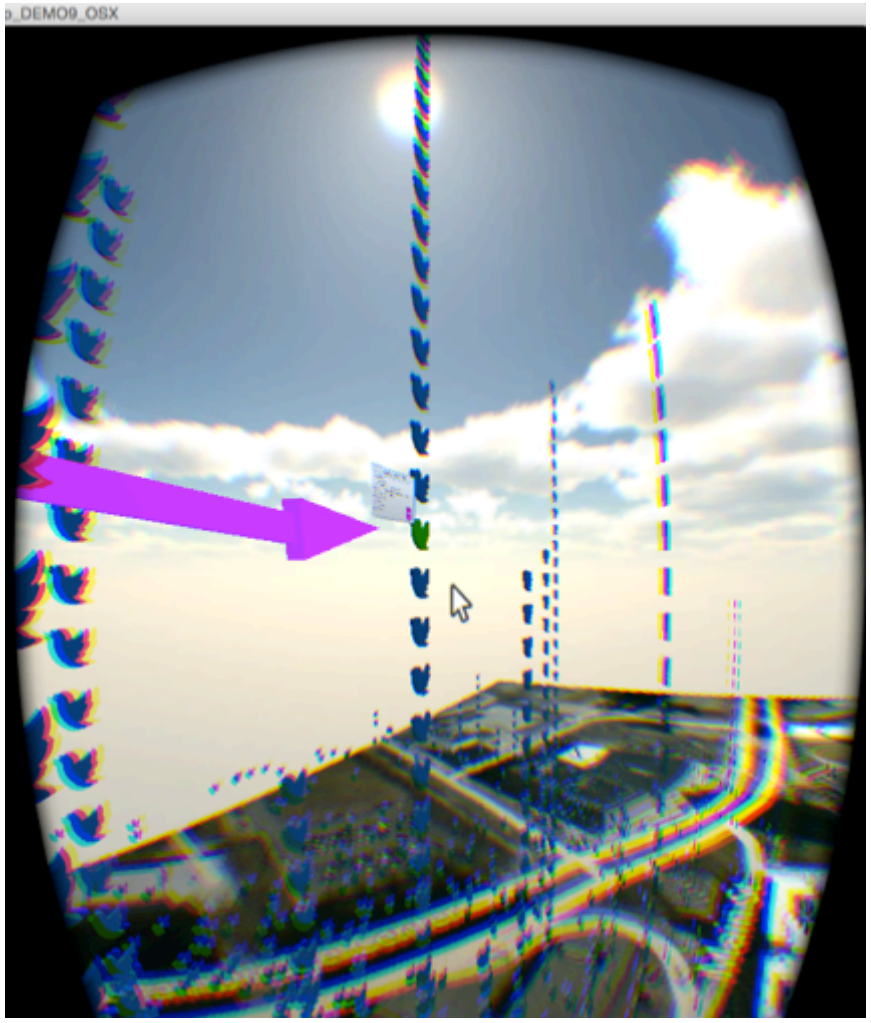
Analytical Tasks

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



Analytical Tasks

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

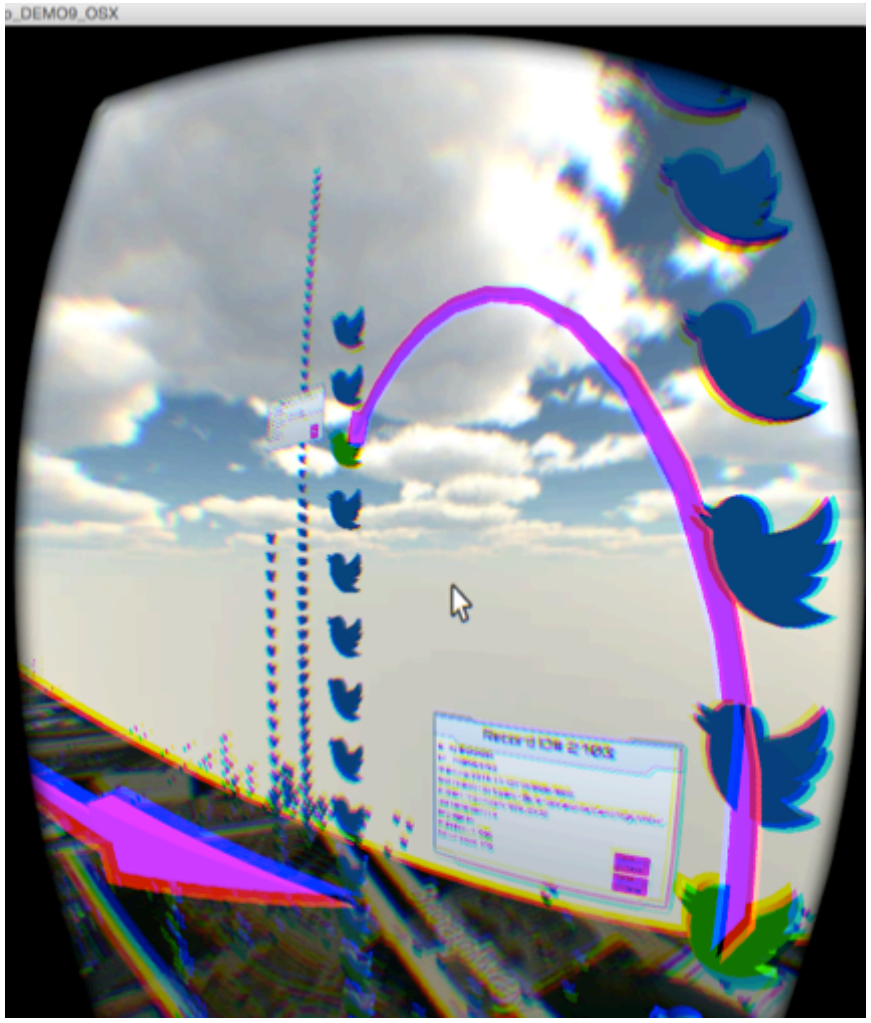


Analytical Tasks

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

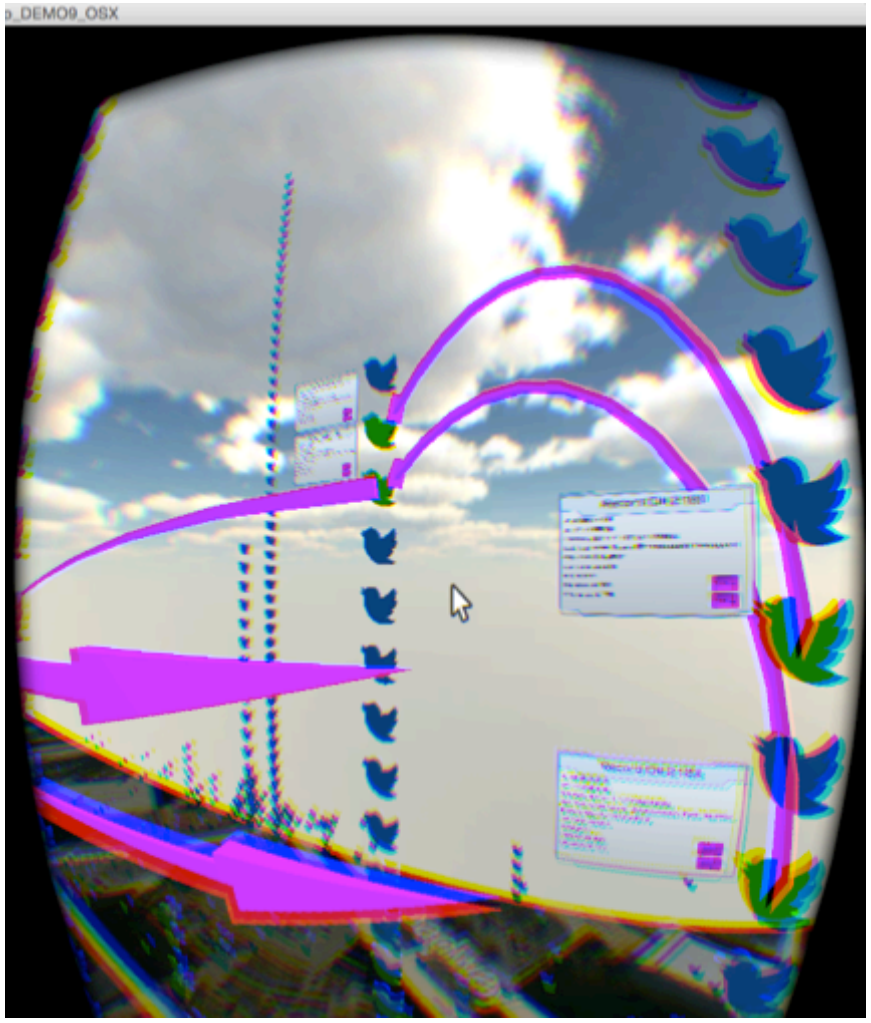
Analytical Tasks

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



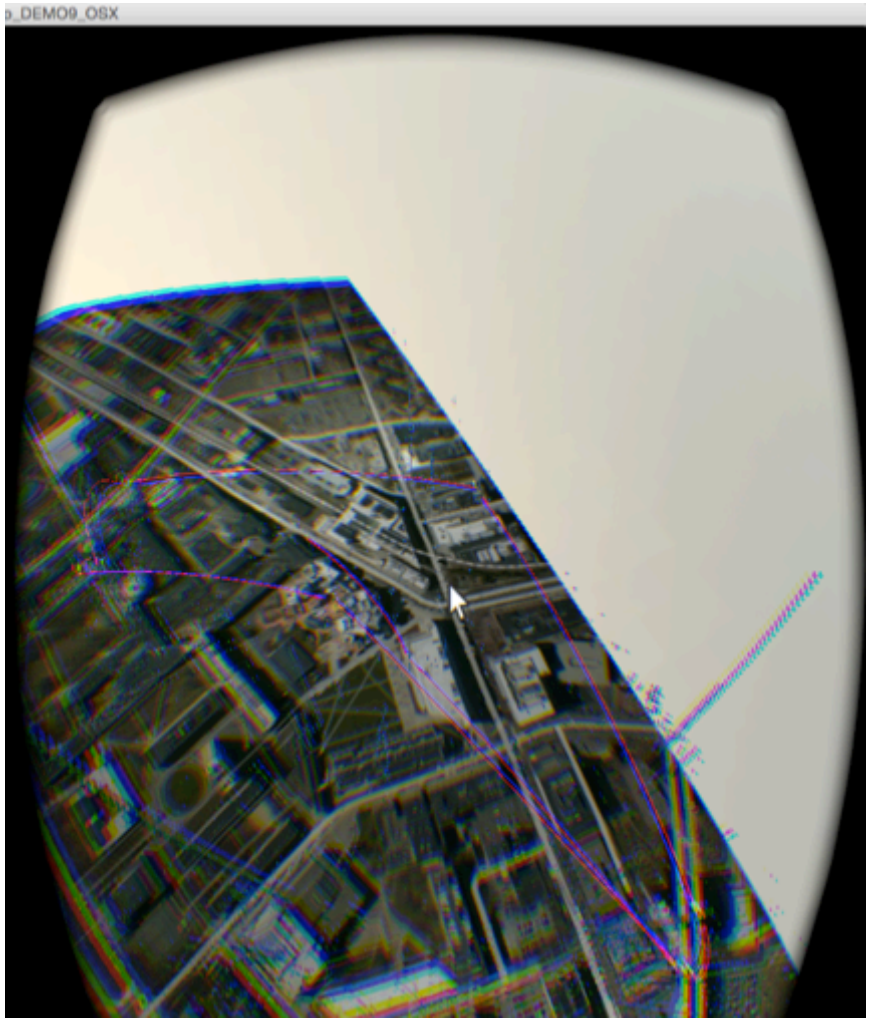
Analytical Tasks

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



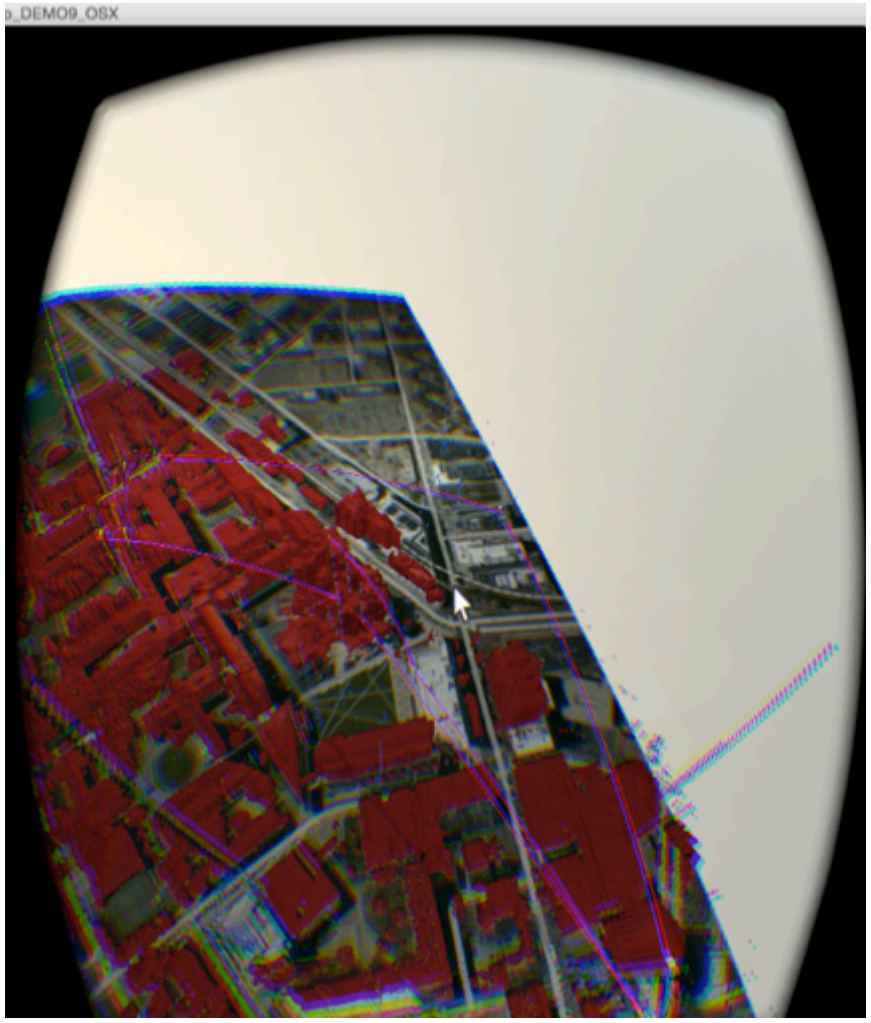
Analytical Tasks

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



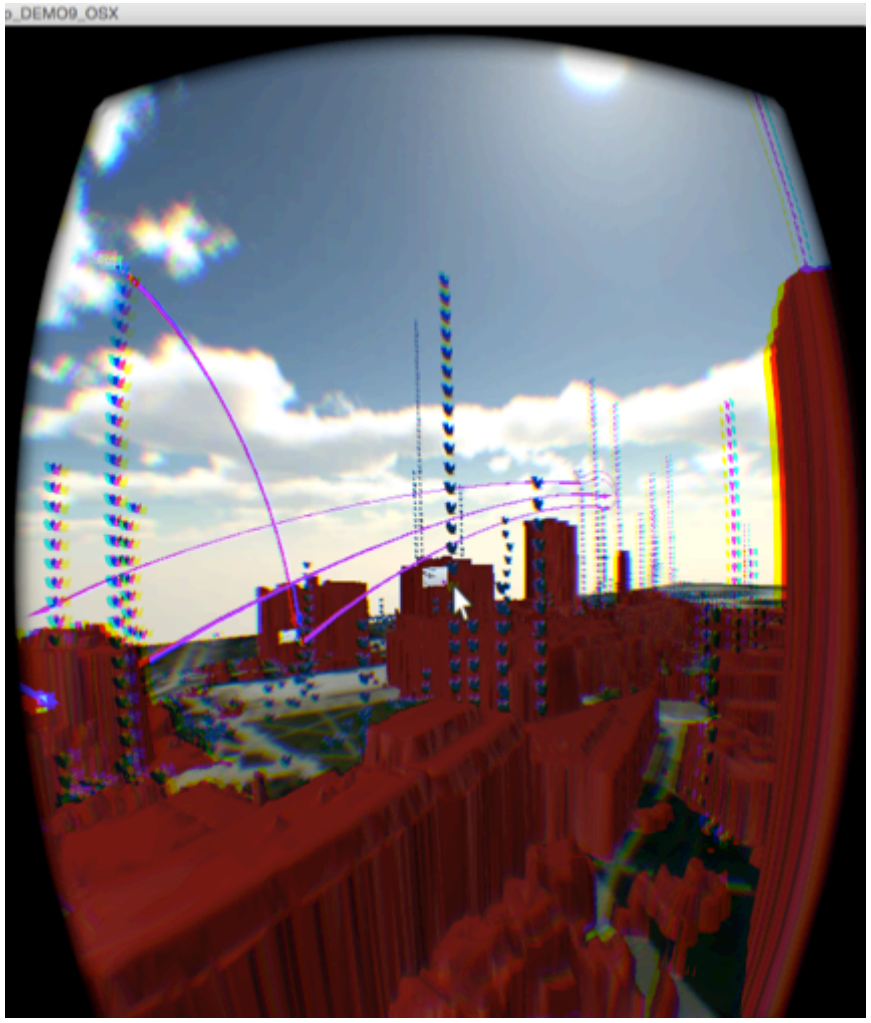
Analytical Tasks

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



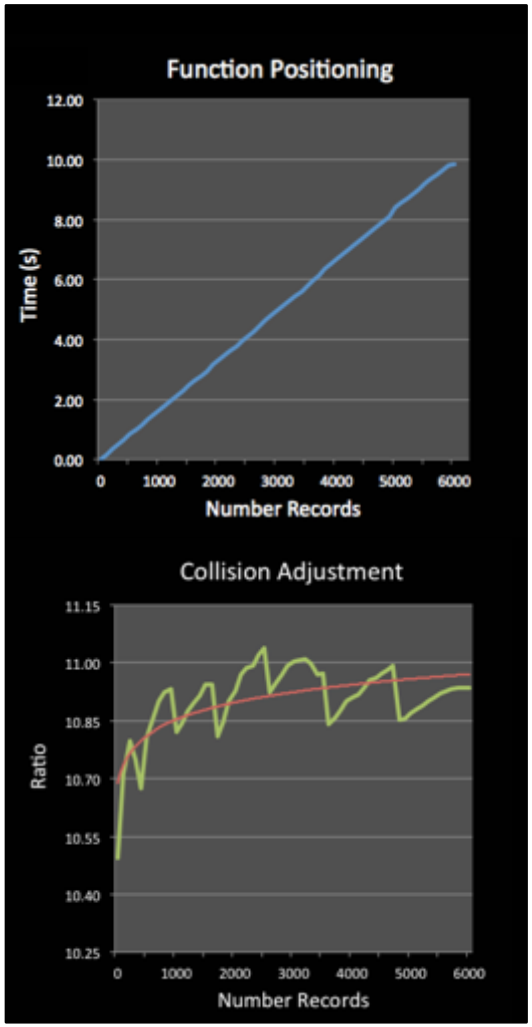
Analytical Tasks

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



Challenges

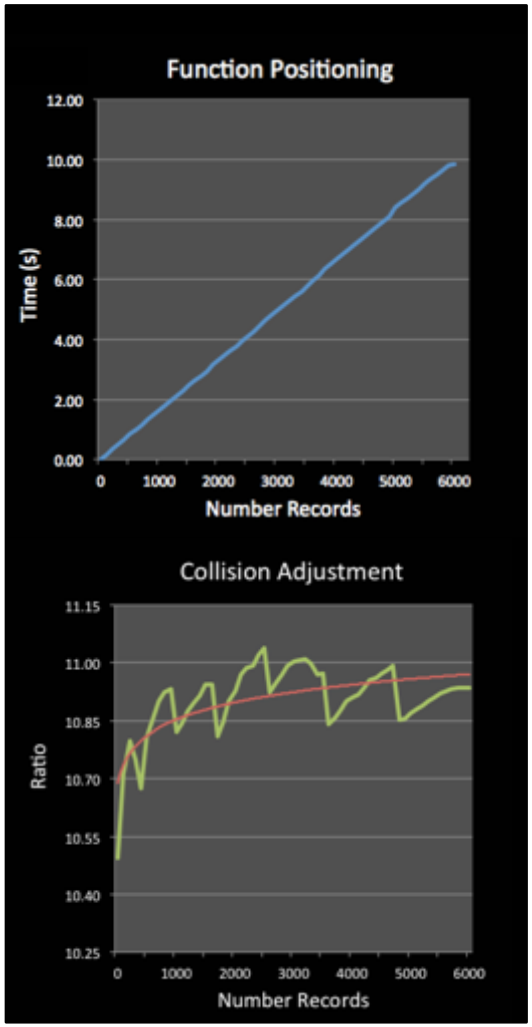
- Rendering Models
- Camera LOD
- Occlusion Layers
- Processing – Frame Rate
- Collider Activation



Challenges

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

- Problem: Further instantiation leads to more collisions
- Solution: 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