

WebPlotViz: Browser Visualization of High Dimensional Streaming Data with HTML5

**STREAM2016 Workshop Washington
DC March 23 2016**

Supun Kamburugamuve, Pulasthi Wickramasinghe,
Saliya Ekanayake, Chathuri Wimalasena and Geoffrey Fox

Indiana University



INDIANA UNIVERSITY BLOOMINGTON

SCHOOL OF INFORMATICS AND COMPUTING

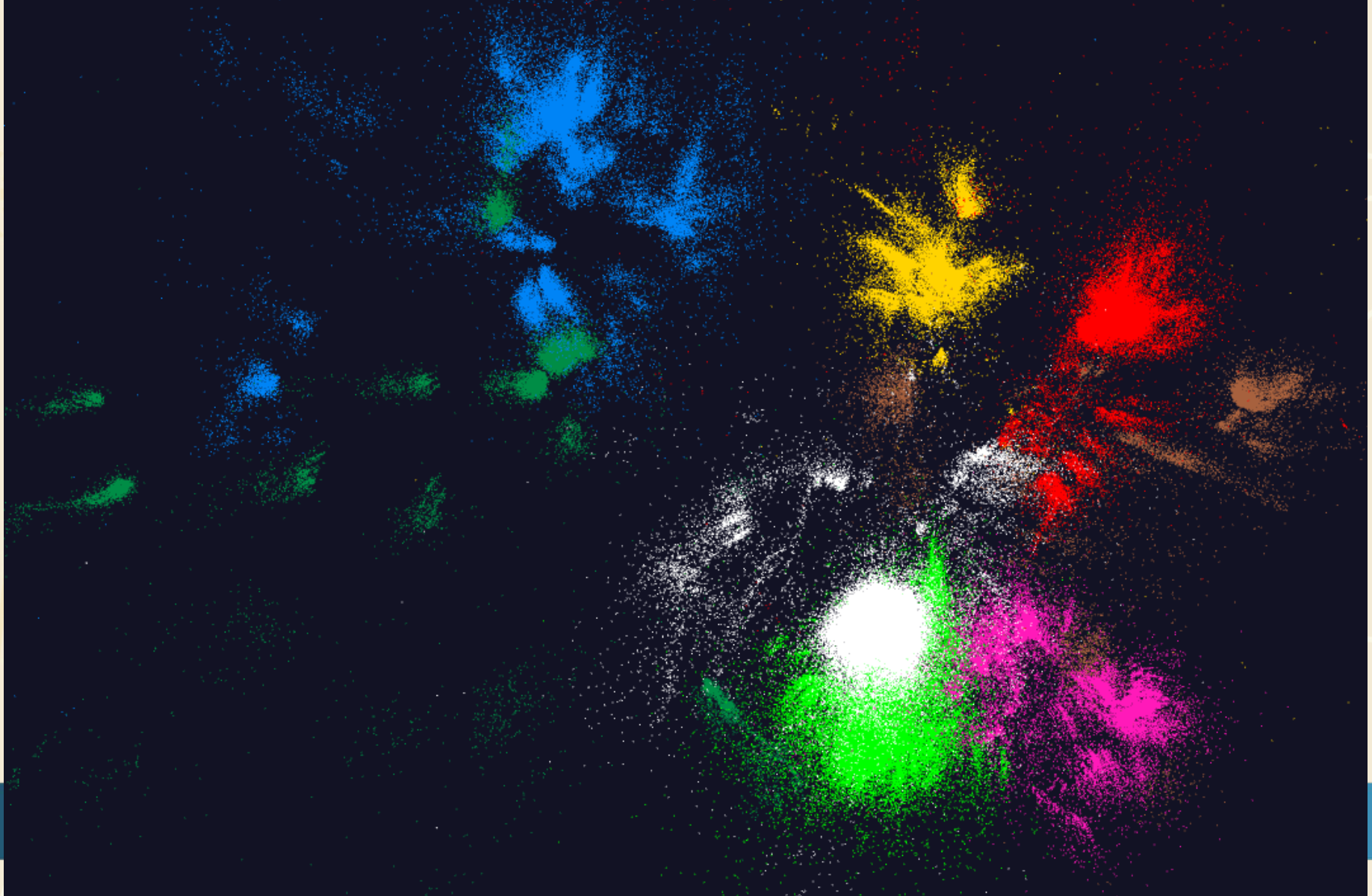
WebPlotViz Basics

- Many data analytics problems can be formulated as study of **points** that are often in some abstract non-Euclidean space (bags of genes, documents ..) that typically have pairwise distances defined but sometimes not scalar products.
- Helpful to visualize set of points to understand better structure
- **Principal Component Analysis** (linear mapping) and **Multidimensional Scaling MDS** (nonlinear and applicable to non-Euclidean spaces) are methods to map abstract spaces to three dimensions for visualization
 - Both run well in parallel and give great results
- In past used custom client visualization but recently switch to commodity HTML5 web viewer **WebPlotViz**



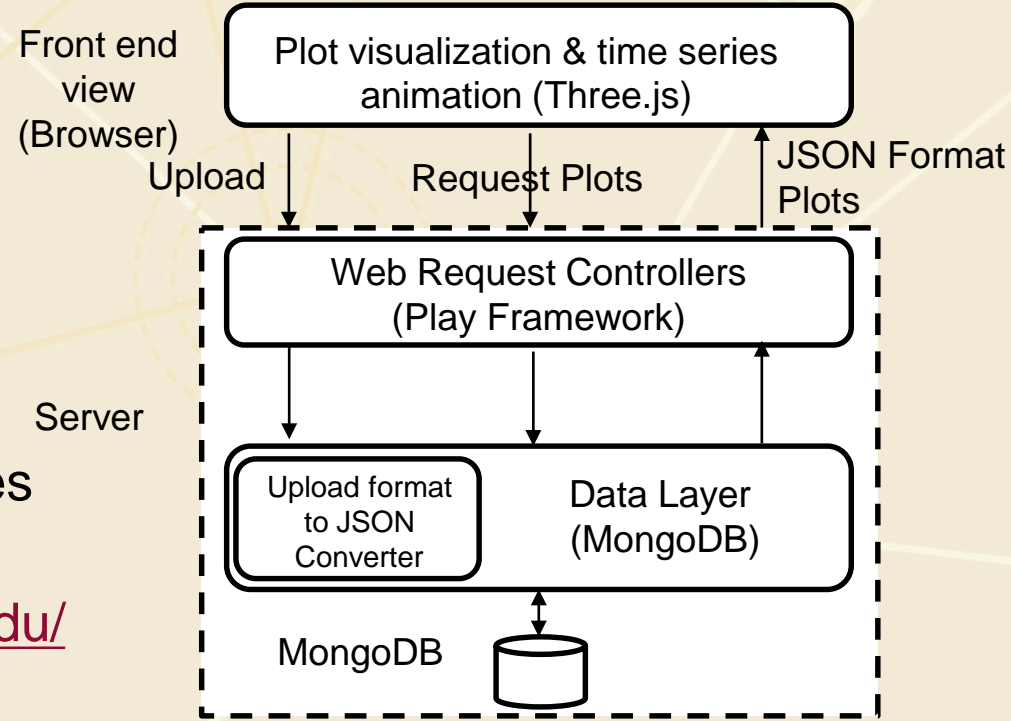
Basic WebPlotViz non Streaming example

- 446K gene sequences mapped to 3D



WebPlotViz Basics II

- Supports visualization of **3D point sets** (typically derived by mapping from abstract spaces) for **streaming and non-streaming case**
 - Simple data management layer
 - 3D web visualizer with various capabilities such as defining color schemes, point sizes, glyphs, labels
- Core Technologies
 - MongoDB management
 - Play Server side framework
 - Three.js
 - WebGL
 - JSON data objects
 - Bootstrap Javascript web pages
- Open Source
<http://spidal-gw.dsc.soic.indiana.edu/>
- ~10,000 lines of extra code



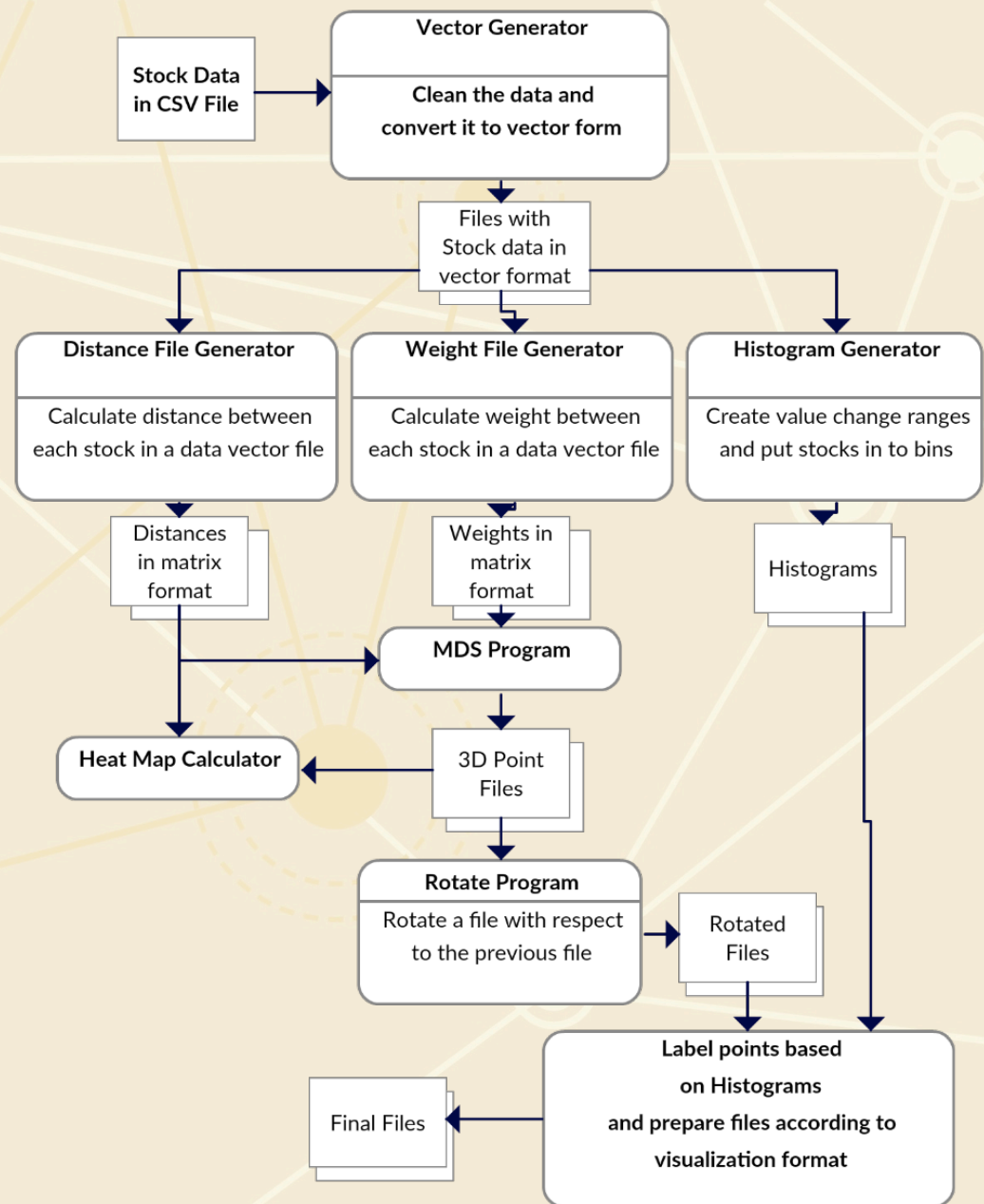
Stock Daily Data Streaming Example

- Typical streaming case considered. Sequence of “collections of abstract points”; cluster, classify etc.; map to 3D; visualize
- Example is collection of around 7000 distinct stocks with daily values available at ~2750 distinct times
 - Clustering as provided by Wall Street – Dow Jones set of 30 stocks, S&P 500, various ETF’s etc.
- The Center for Research in Security Prices (CSRP) database through the Wharton Research Data Services (wrds) web interface
- Available for free to the Indiana University students for research
- 2004 Jan 01 to 2015 Dec 31 have daily Stock prices in the form of a CSV file
- We use the information
 - ID, Date, Symbol, Factor to Adjust Volume, Factor to Adjust Price, Price, Outstanding Stocks



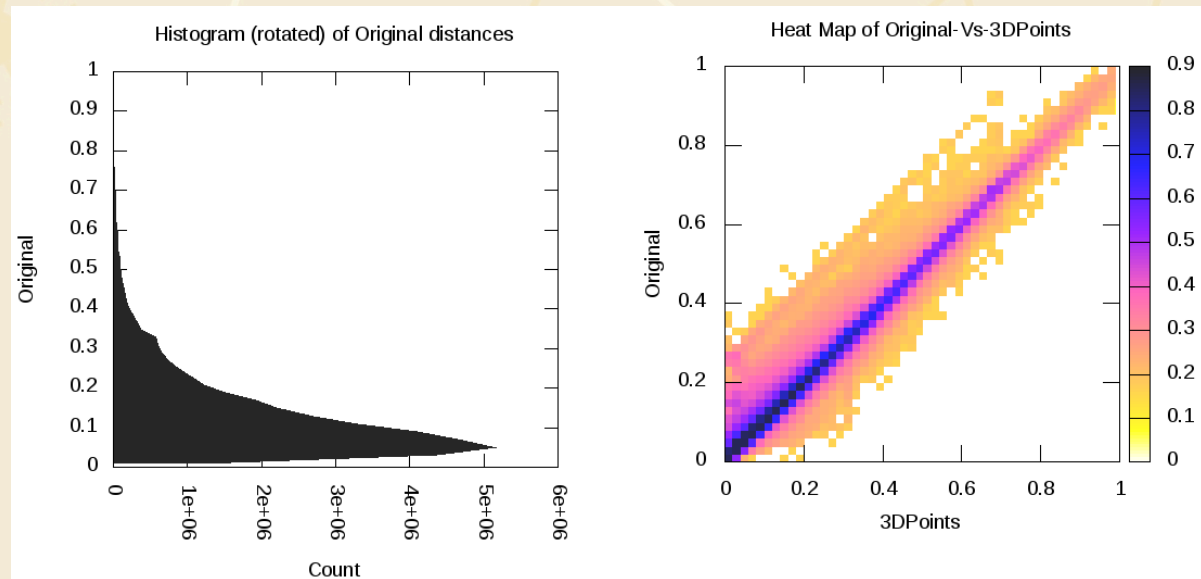
Stock Problem Workflow

- Clean data
- Calculate distance between stocks
- Calculate distance between stocks (Pearson Correlation as missing data)
- Map 250-2800 dimensional stock values to 3D for each time
- Align each time
- Visualize
- Will move to Apache Beam to support custom runs



Few Notes on Mapping to 3D

- MDS performed separately at each day – quality judged by match between abstract space distance and mapped space distance
 - Pretty good agreement as seen in heat map averaged over all stocks and all days

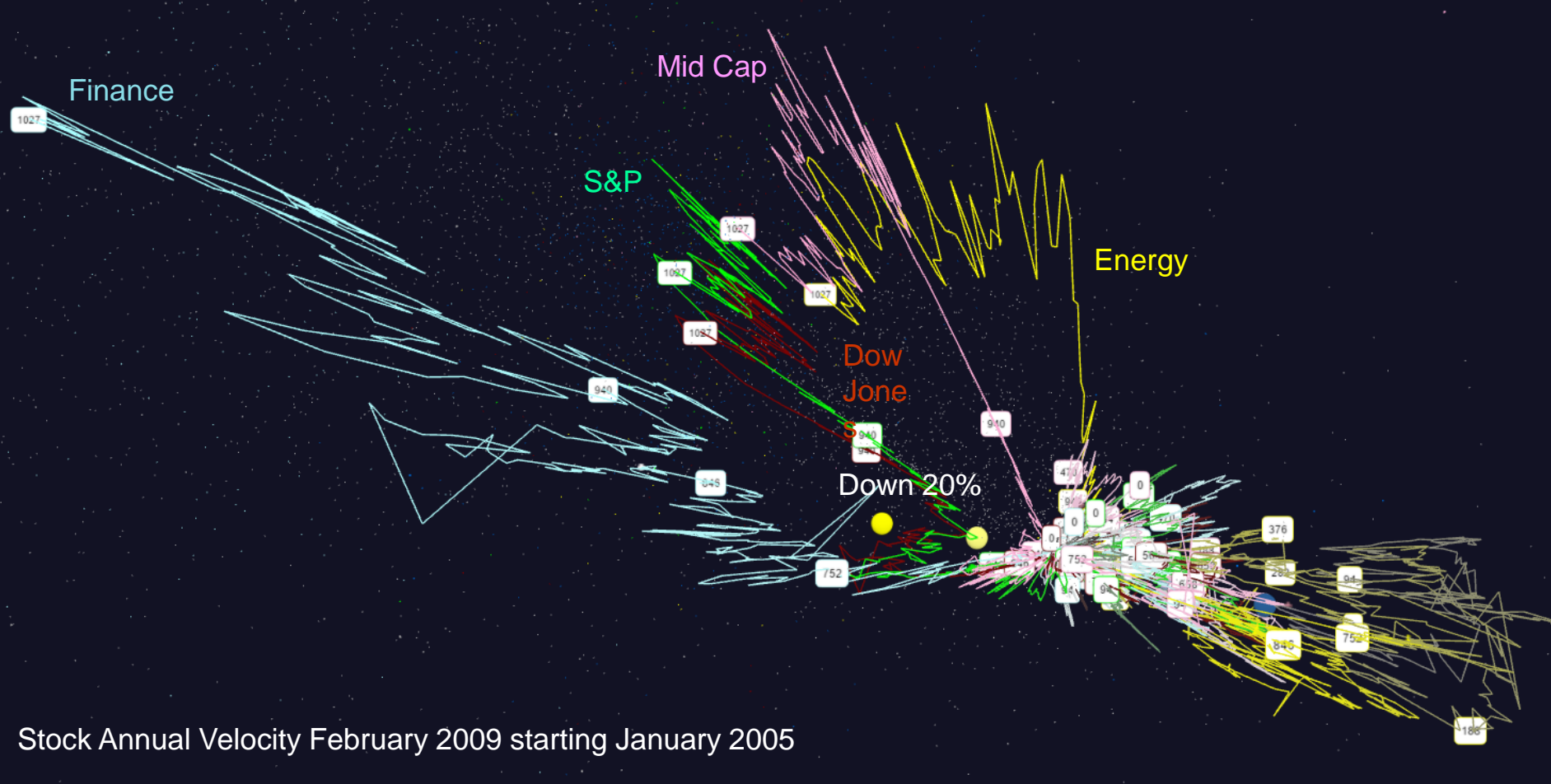


- Each day is mapped independently and is ambiguous up to global rotations and translations
 - Align each day to minimize day to day change averaged over all stocks



Stock Velocity Bear Market

You can look at many things. We look at values and velocities (value change over window – one year here). Can study over different ranges. 6500 points each display but can use glyphs and trajectories to study particular stocks or collections thereof

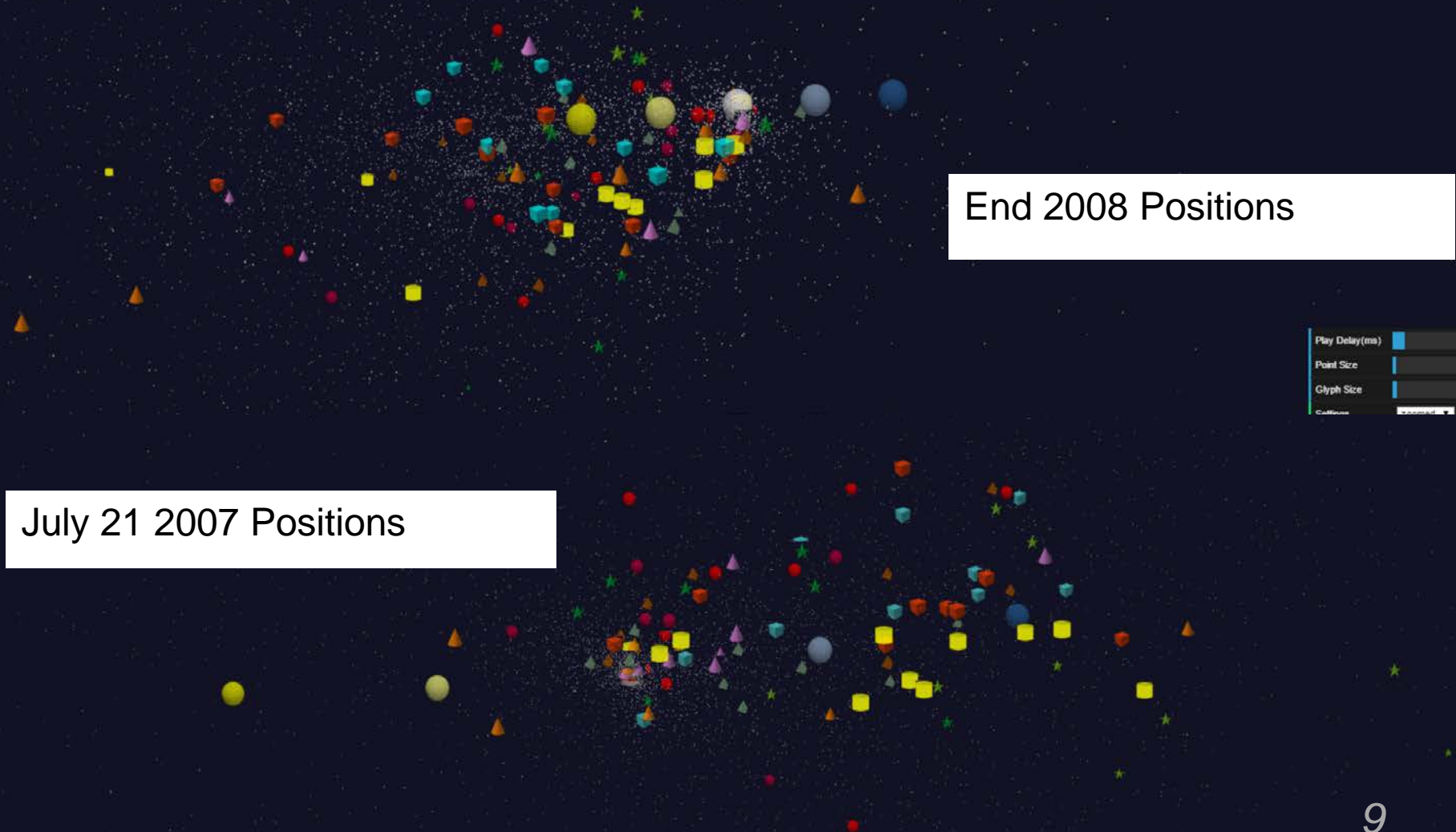


Stock Annual Velocity February 2009 starting January 2005



Top 10 stocks highlighted with glyphs

3



End 2008 Positions

July 21 2007 Positions

Play Delay (ms)

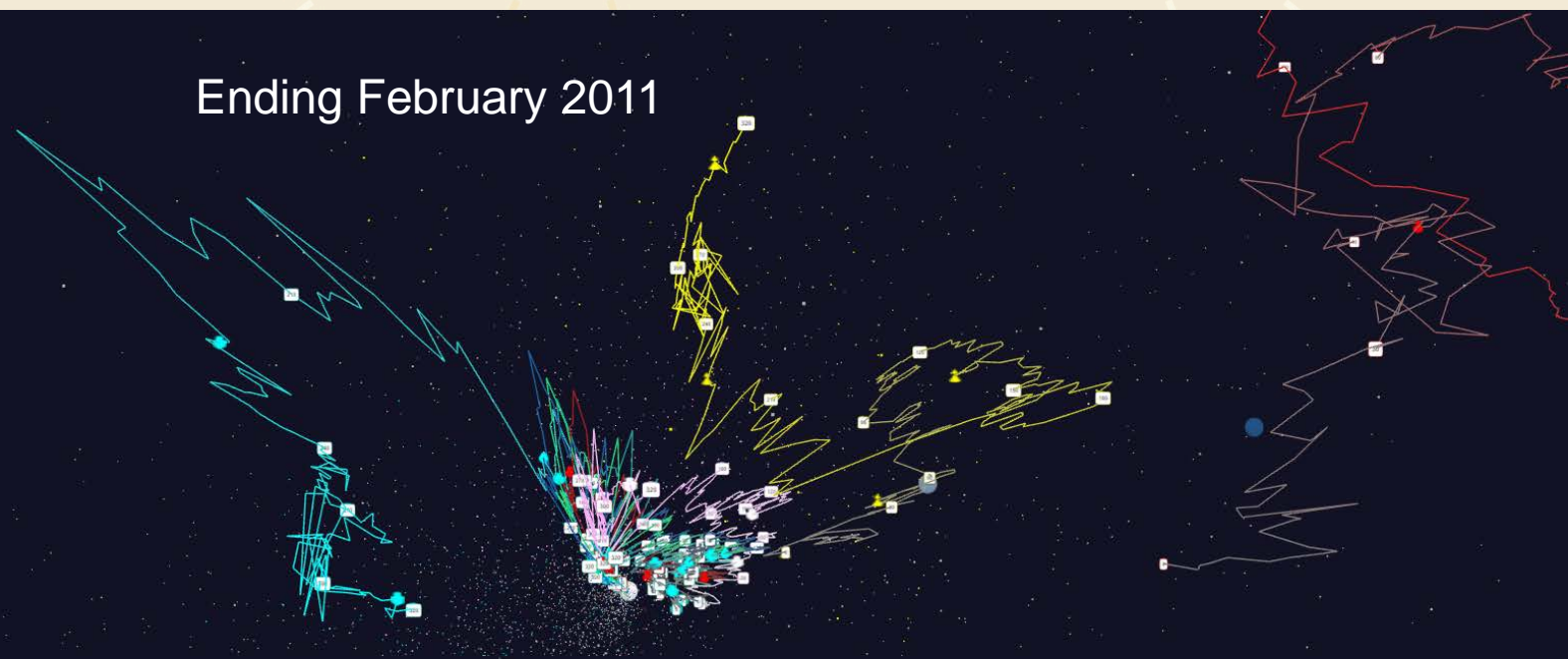
Point Size

Glyph Size

Call Home

Relative Changes in Stock Values starting January 2004

Ending February 2011



Ending December 2015

