



# Depth Explorer

A Software Tool for the Analysis of Depth Measures

John Hugg, Eynat Rafalin and Diane Souvaine

Department of Computer Science, Tufts University, Medford MA 02155  
Partially supported by NSF grant CCF-0431027

# Primary Project Goal

To compare, visually and numerically, the performance of different depth measures on real and synthetic data

1. Help determine which depth measure(s) to use
2. Develop new approaches for applying existing depth measures

# Secondary Goals

1. To provide an easy to use, visual tool to introduce new students and researchers to the concept of Data Depth
2. To provide camera-ready data visualizations for researchers in Data Depth and in applied fields

# Motivation

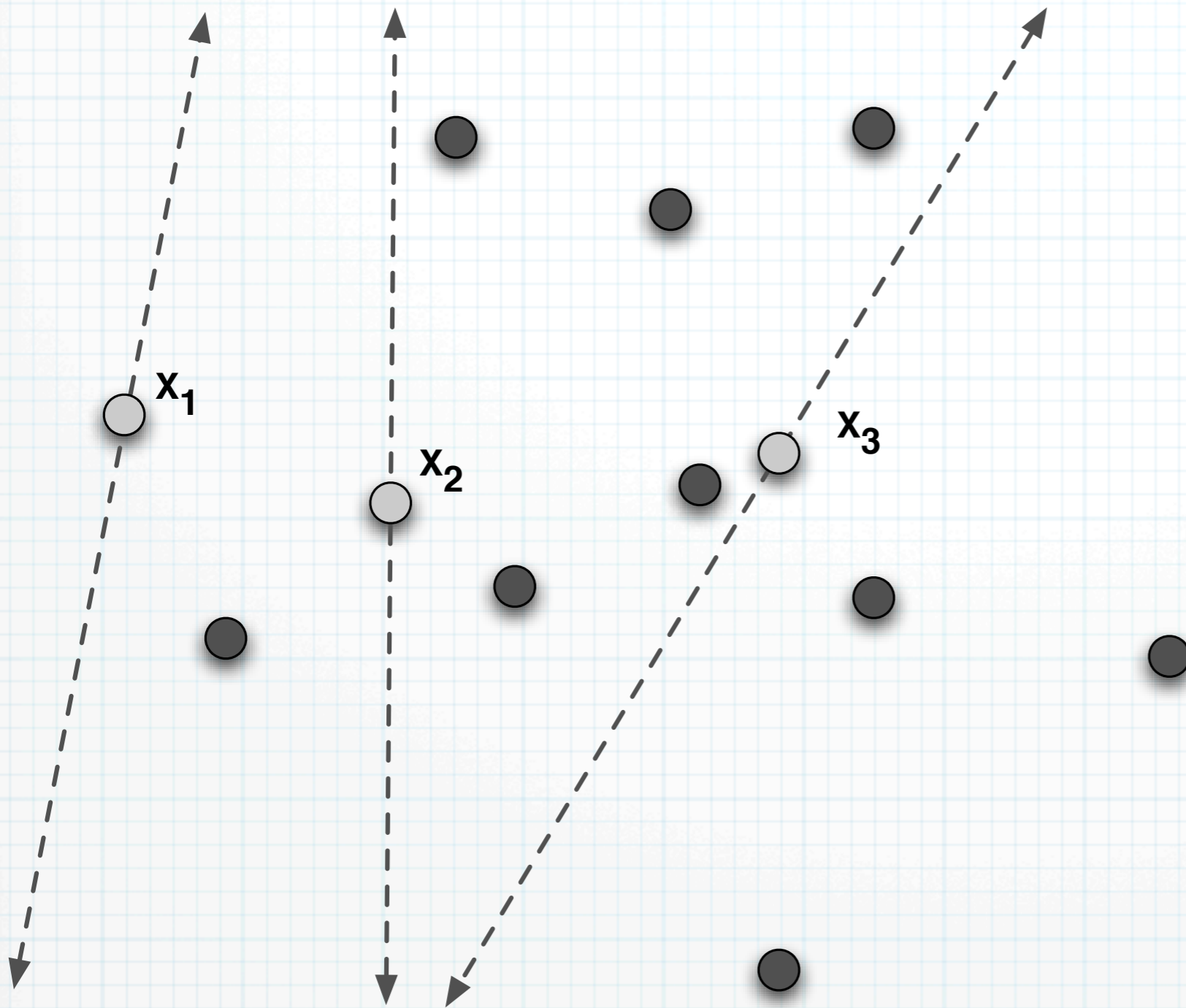
Why Create Depth Explorer?



# The Catalyst

- Request:  
How can Data Depth help analyze a biomedical data set?  
*(5000, 15-dimensional vectors)*
- Unable to use Halfspace depth as complexity is exponential in dimension
- LI depth is one of several depth measures with more efficient performance

# Halfspace Depth

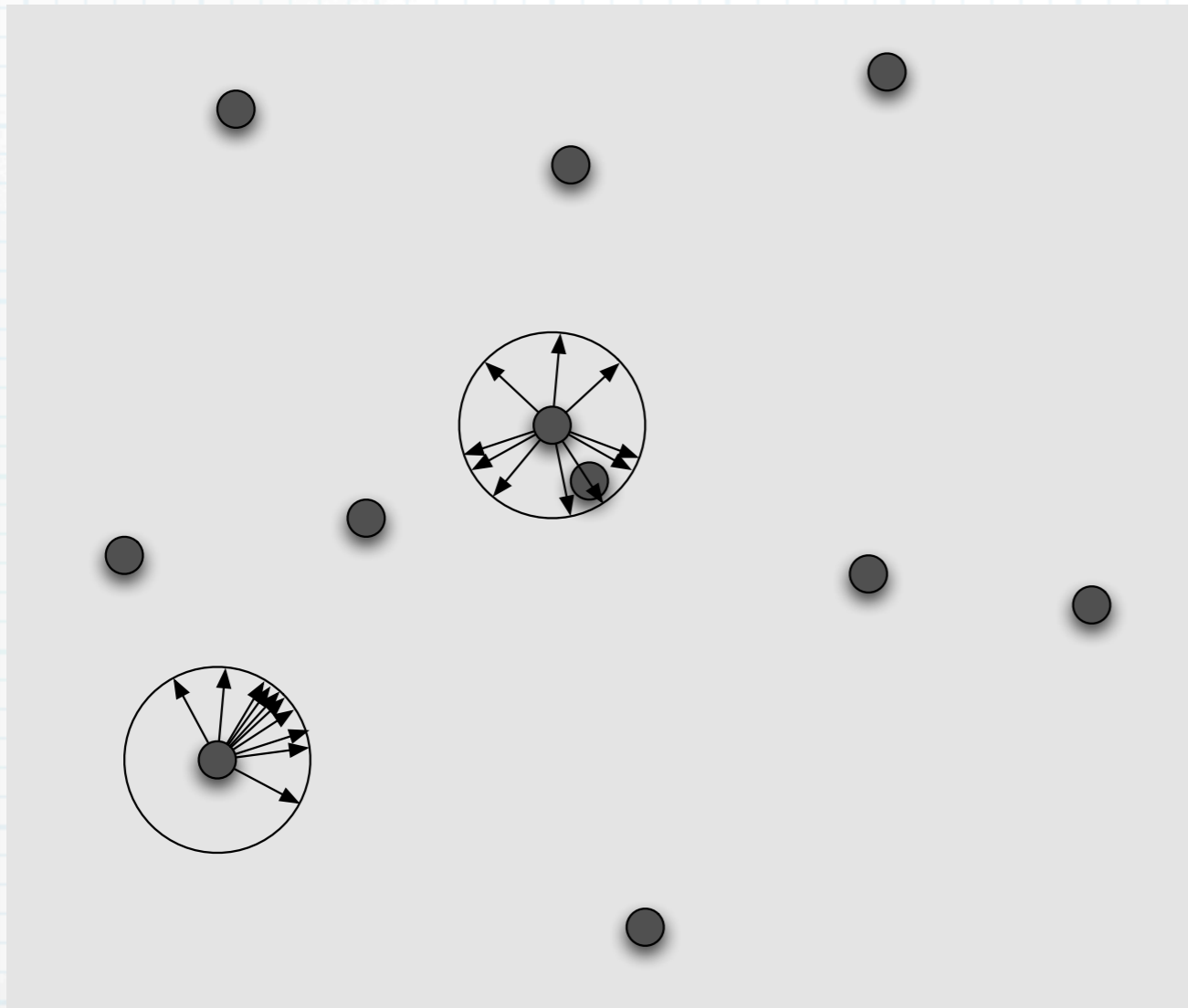


The Half-space depth of a point  $x$  with respect to a data set  $S$  is:

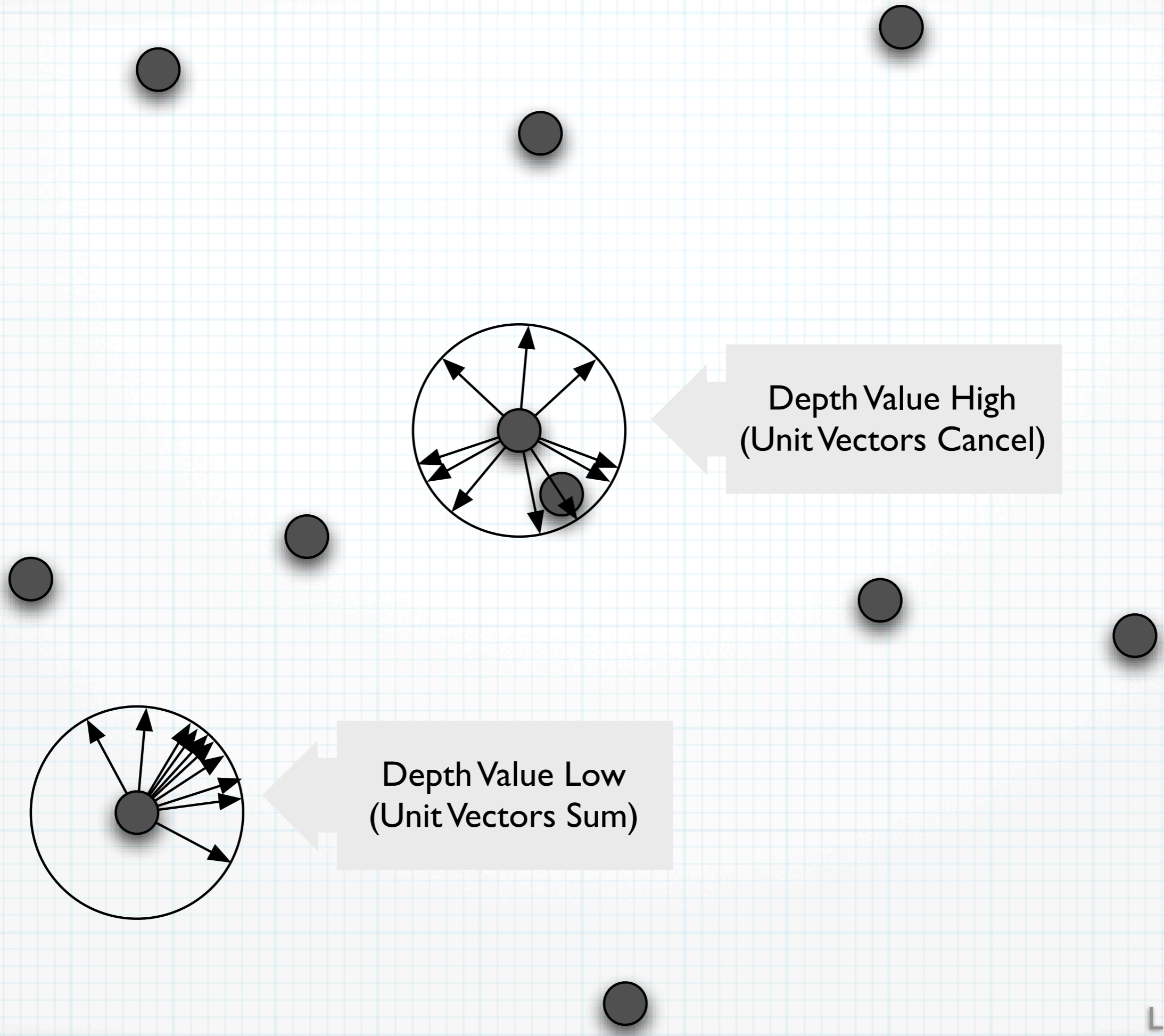
The minimum number of points of  $S$  lying in any closed half-space determined by a line through  $x$

# LI Depth

- The LI depth of a point  $x$  with respect to a data set  $S = \{X_1, \dots, X_n\}$  in  $\mathbb{R}_d$  is one minus the average of the unit vectors from  $x$  to all observations in  $S$



- Complexity is linear in dimension
- Depth is zero *only* at infinity.  
For many other depth measures, depth is uniformly zero at the convex hull.



Depth Value High  
(Unit Vectors Cancel)

Depth Value Low  
(Unit Vectors Sum)



# LI Results on High-D Data

- From an initial K-Means clustering:
  - Set A: Points identified as misclassified points according to LI Depth
  - Set B: Points identified as misclassified after normalizing all 15 dimensions and applying LI Depth
- The two sets of reclassified points were almost completely disjoint: *Why?*

# LI Results on High-D Data

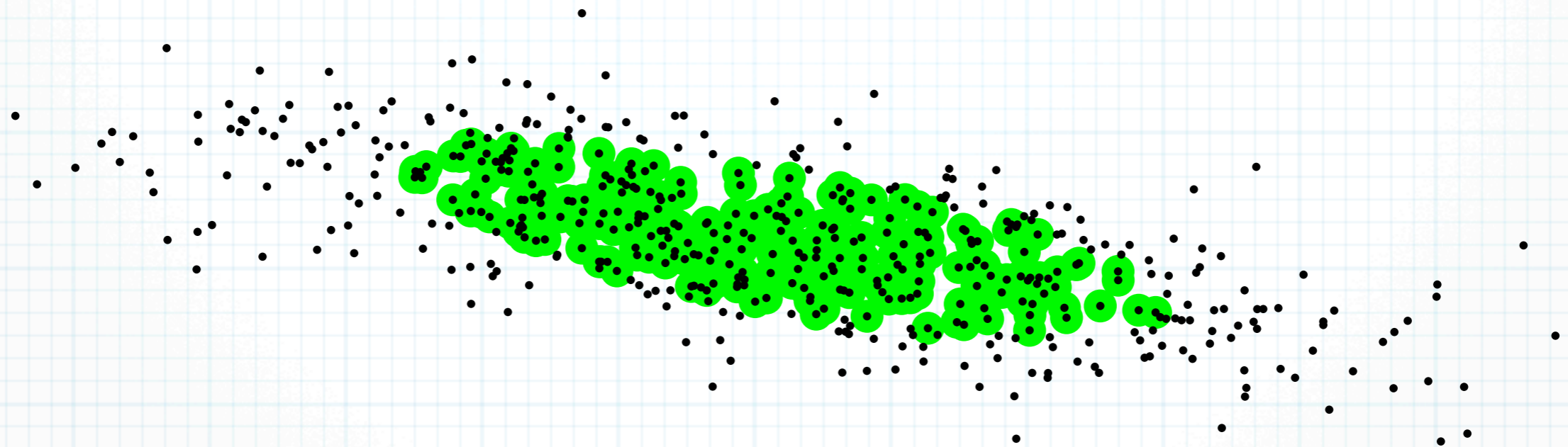
Different ways of using depth give different results

- Are any of the methods appropriate for this data set?
- How do we evaluate the appropriateness of a method?

# Demo I

Early Experiments

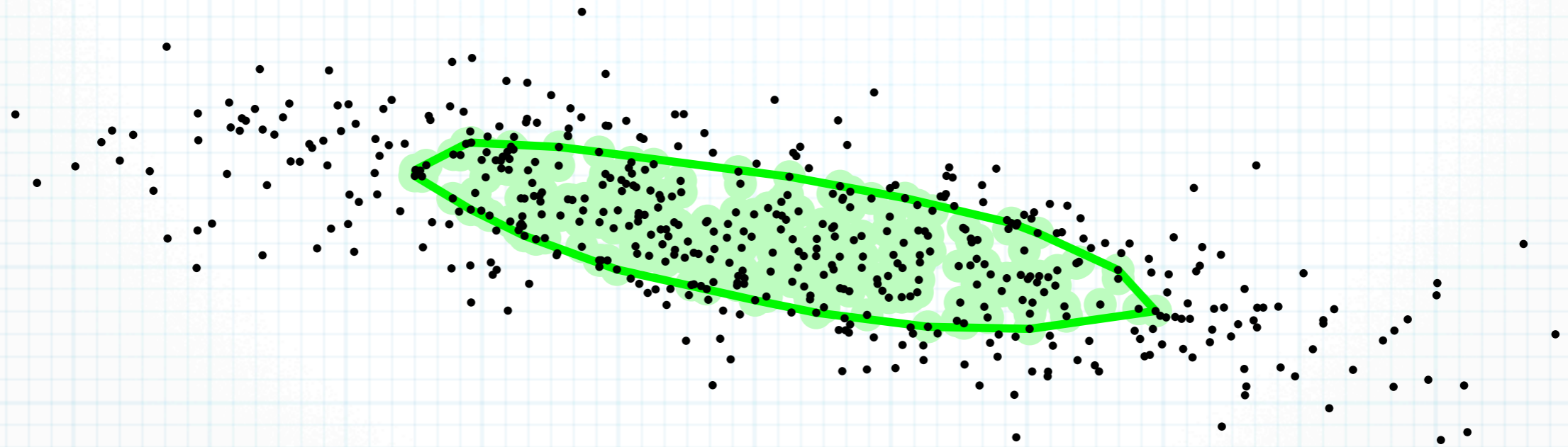
# Depth Contours



- The region enclosed by the contour of depth  $t$  is the set of points such that  $D(x) \geq t$
- For well behaved depth function the contours can be approximated using the convex hull of the point of depth  $t$  [*Liu 2003*]

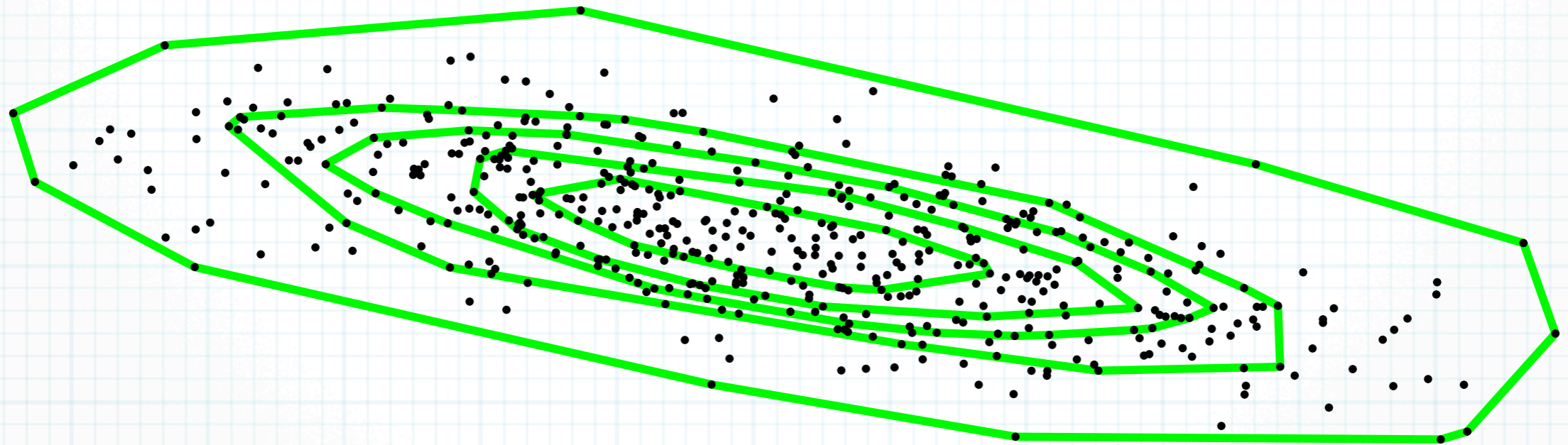


# Depth Contours



- The region enclosed by the contour of depth  $t$  is the set of points such that  $D(x) \geq t$
- For well behaved depth function the contours can be approximated using the convex hull of the point of depth  $t$  [*Liu 2003*]

# Depth Contours

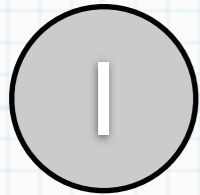


- The region enclosed by the contour of depth  $t$  is the set of points such that  $D(x) \geq t$
- For well behaved depth function the contours can be approximated using the convex hull of the point of depth  $t$  [Liu 2003]

# The Tool

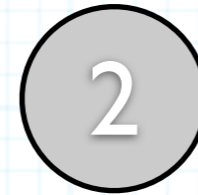
How Depth Explorer Works

# Highest Level Overview

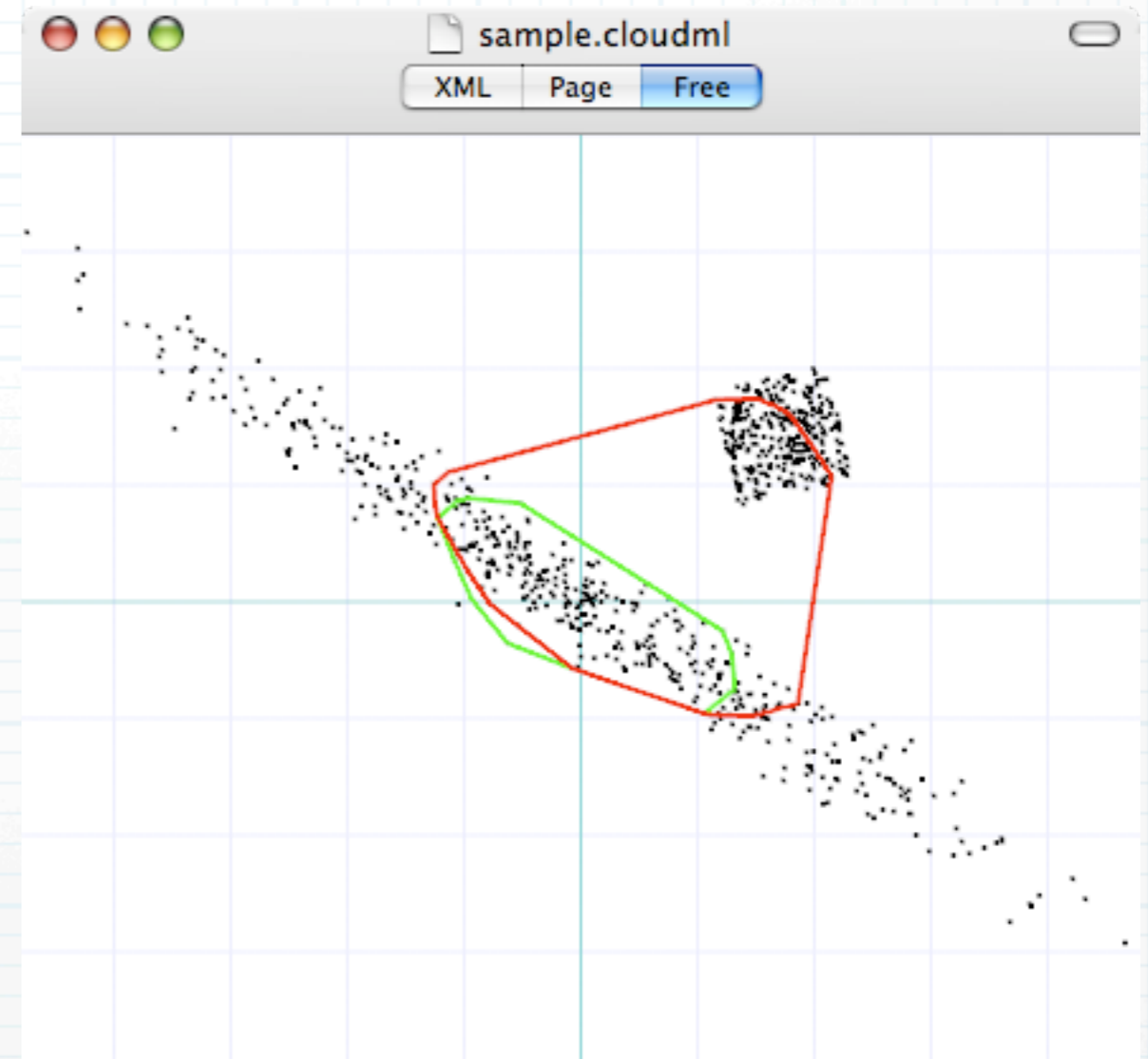


Describe Data & Visualizations in XML

```
<canvas width='6' height='6' margins='.1' minrange='6'>
  <pcabag color='1,0,0' size='.6'>
    <transform angle='13'>
      <transform xtrans='2' ytrans='1'>
        <cloud type='uniform' points='300' />
      </transform>
    <bag type='l1' color='0,1,0' size='.5'>
      <transform xscale='.2' yscale='2' angle='45'>
        <cloud type='normal' points='500' />
      </transform>
    </bag>
  </transform>
</pcabag>
</canvas>
```



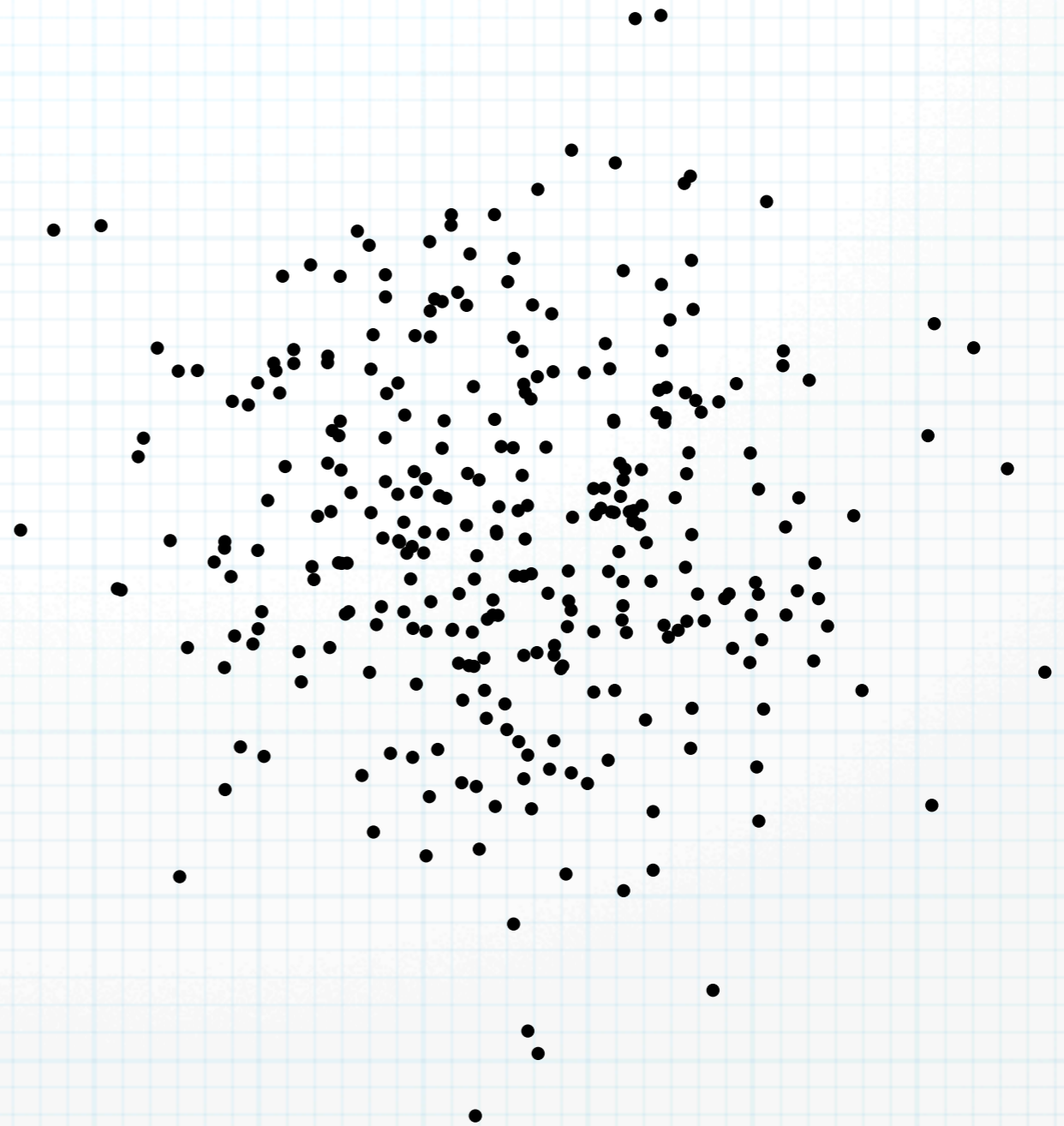
View, Save or Print the Results





# Example

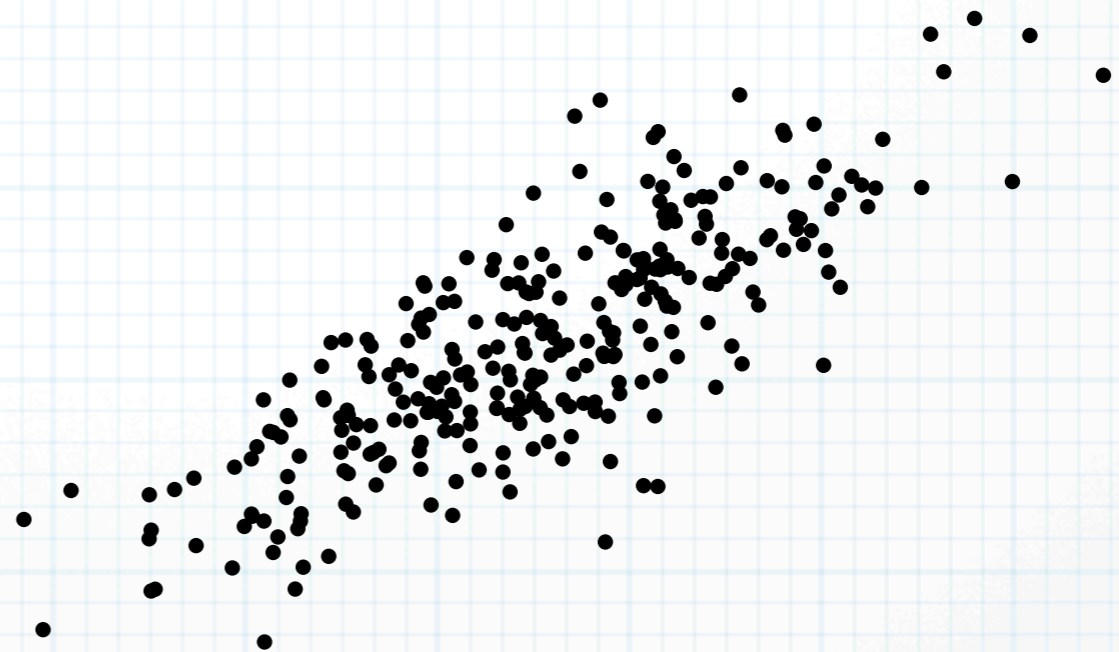
Cloud:  
300 Points,  
Normally  
Distributed



# Example

Transform:  
Scale Y-Values to 30%  
Rotate 30 Degrees

Cloud:  
300 Points  
Gaussian



# Example

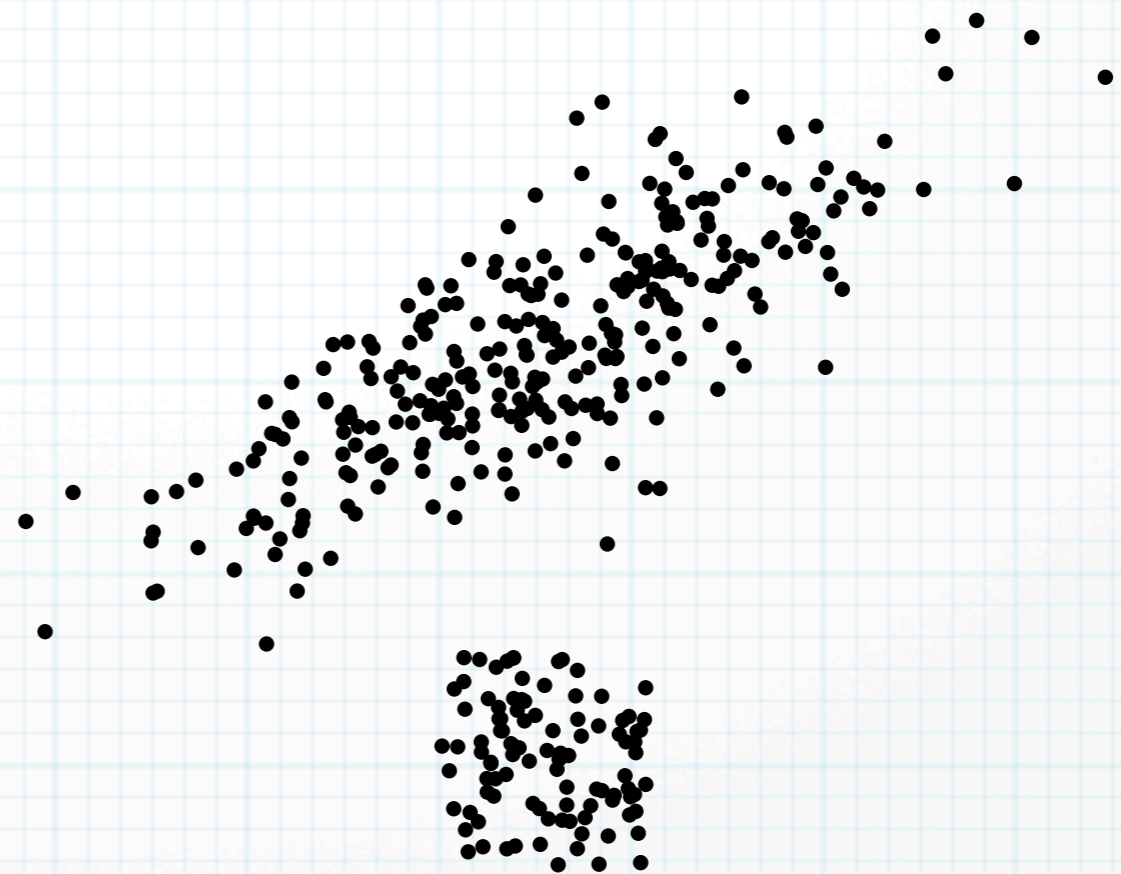
## Compose

Transform:  
Scale Y-Values to 30%  
Rotate 30 Degrees

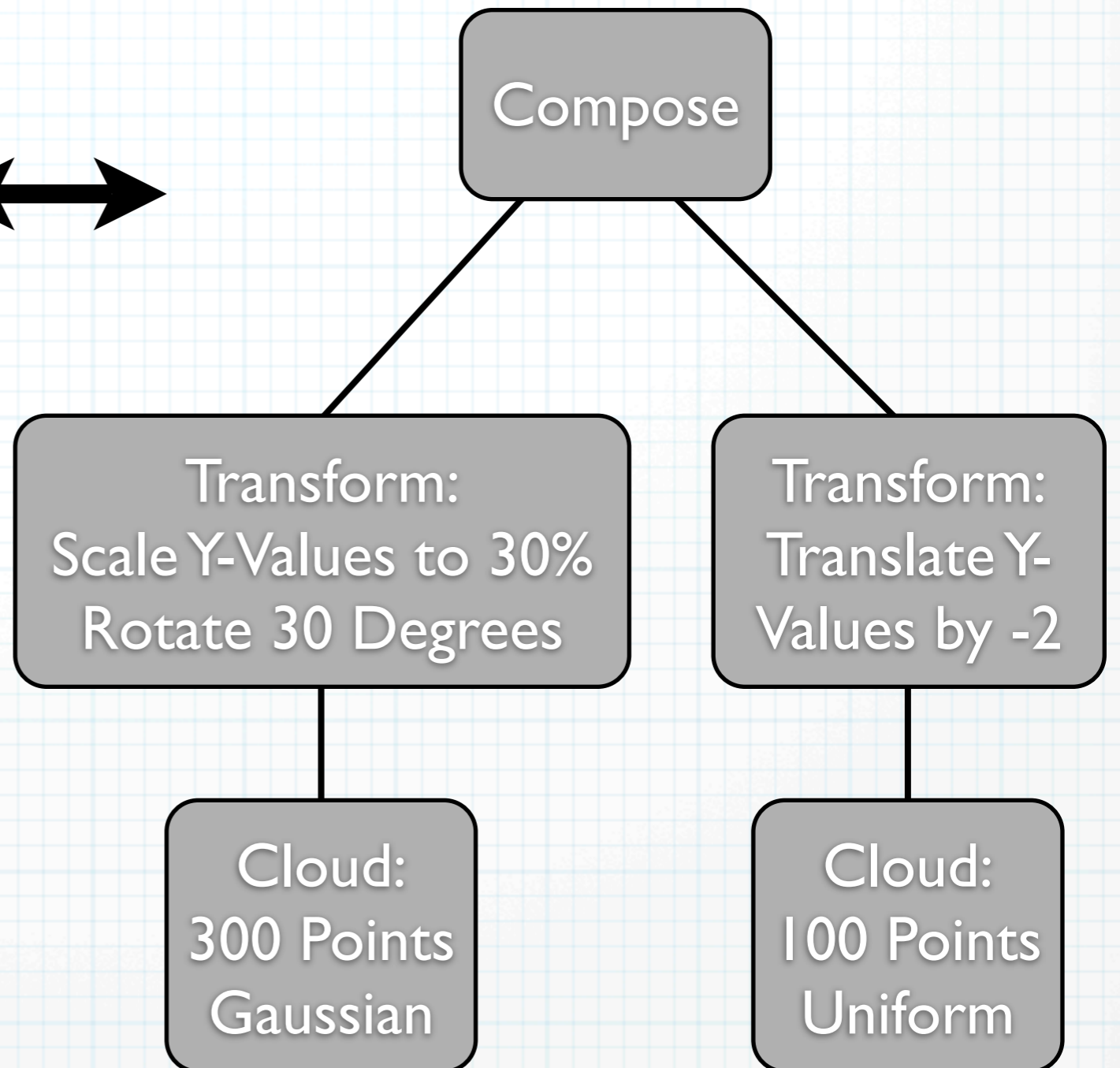
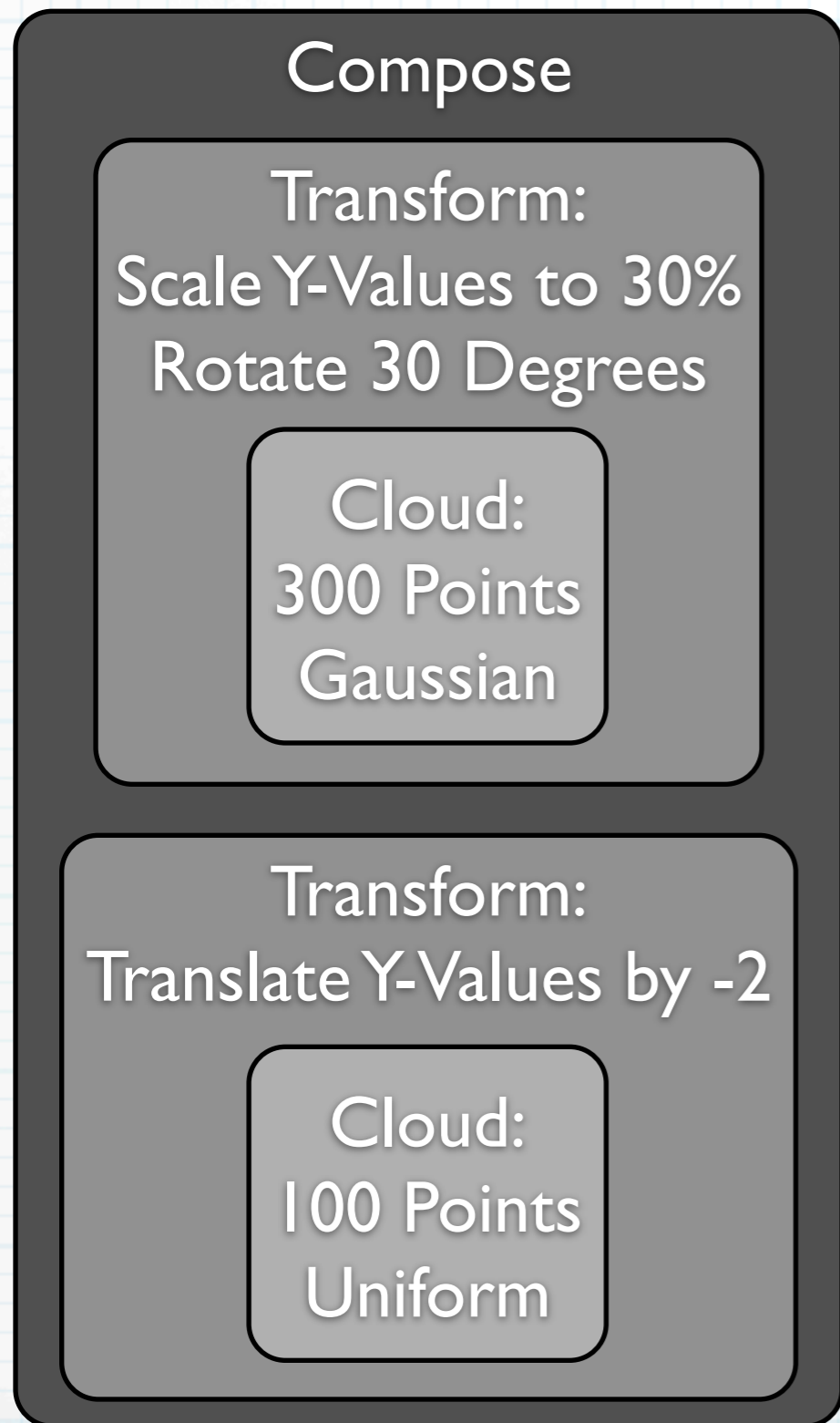
Cloud:  
300 Points  
Gaussian

Transform:  
Translate Y-Values by -2

Cloud:  
100 Points  
Uniform

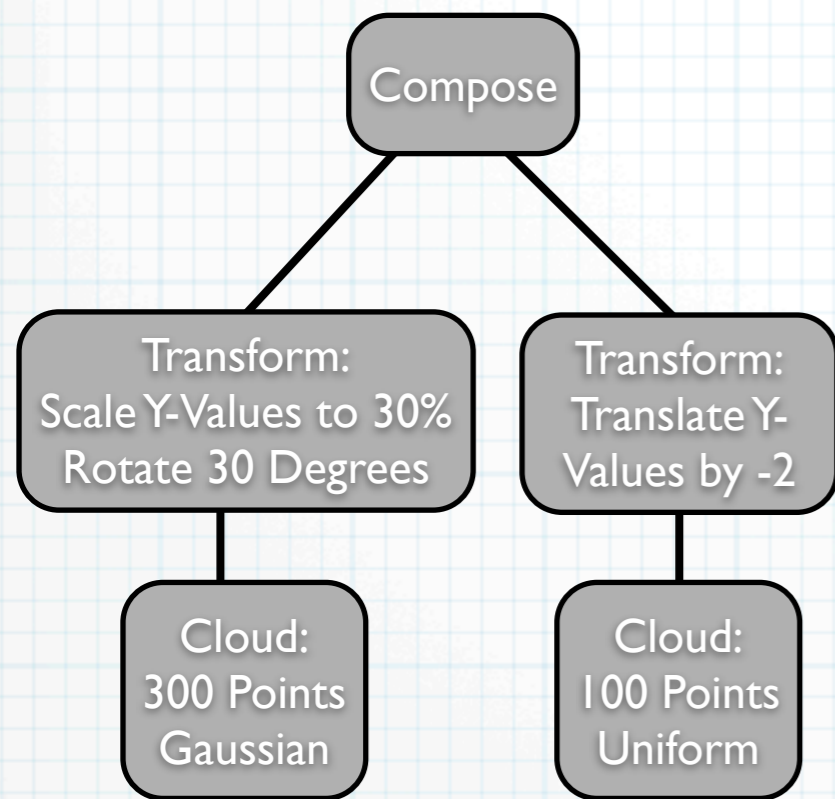


# Example





# Example



```
<compose>
```

```
<transform angle='30' yscale='.3'>
```

```
<cloud type='normal' points='300' />
```

```
</transform>
```

```
<transform ytrans='-2'>
```

```
<cloud type='uniform' points='100' />
```

```
</transform>
```

```
</compose>
```

# Demo 2

Using Depth Explorer

# Version 1.1 Available Now

- Available free on the web for Mac OS X 10.4
- Supports 7 Depth Measures  
Convex Hull Peeling, Halfspace, LI, PCA-LI,  
Proximity (3 Kinds)
- Can Highlight Points or Draw Contours to  
Indicate Depth
- Supports XML Saving, PDF-Export, Online Help  
and many other Desktop Application Features

# Extending Depth Explorer

## You Need:

- A Mac OS X Development System  
*Windows & Linux versions in development*
- The Depth Explorer Source Code Package
- A C/C++ Implementation of your Depth Measure or Visualization  
*Support for additional languages in development*

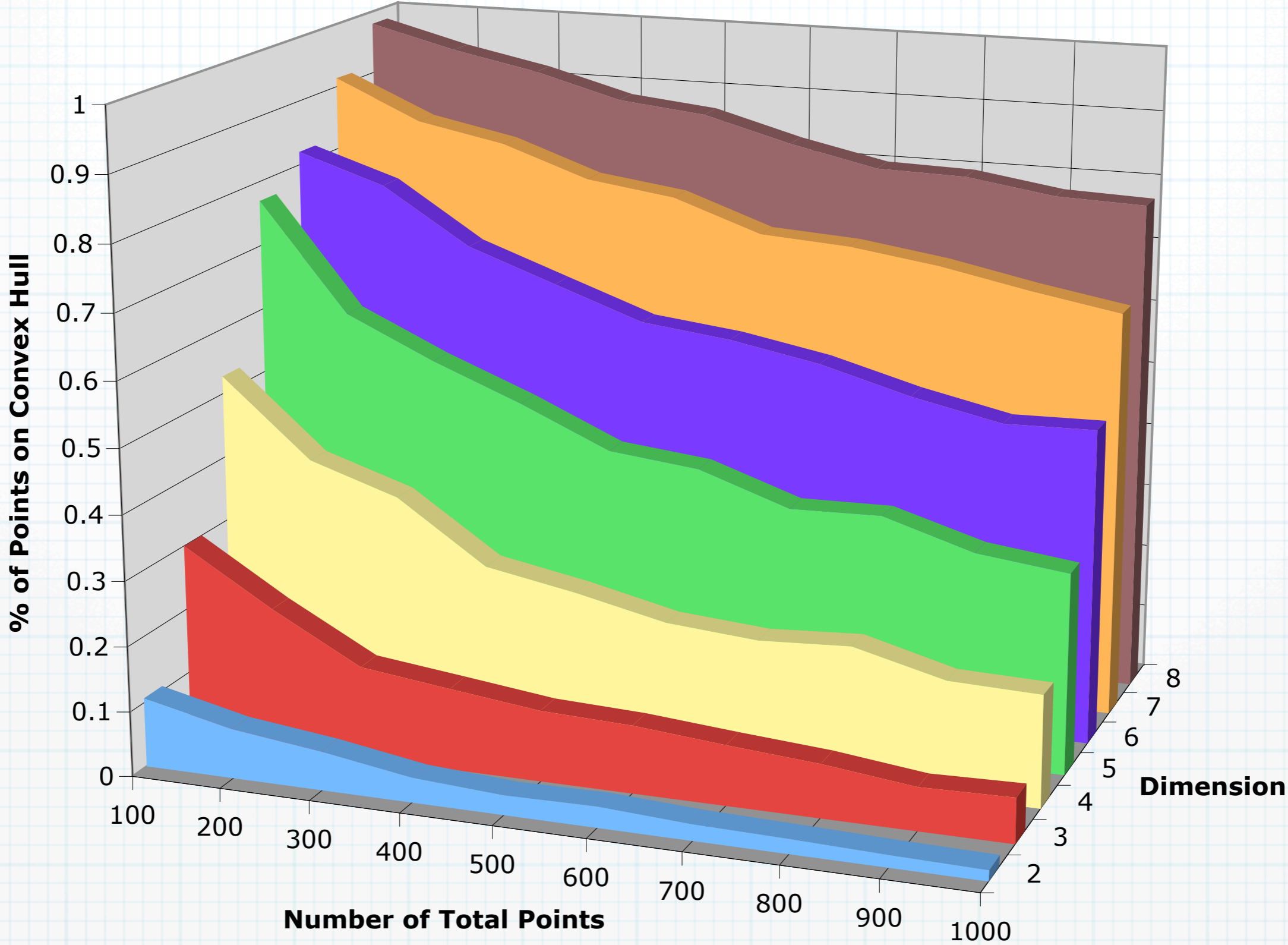
*Follow the included guide to glue it all together!*



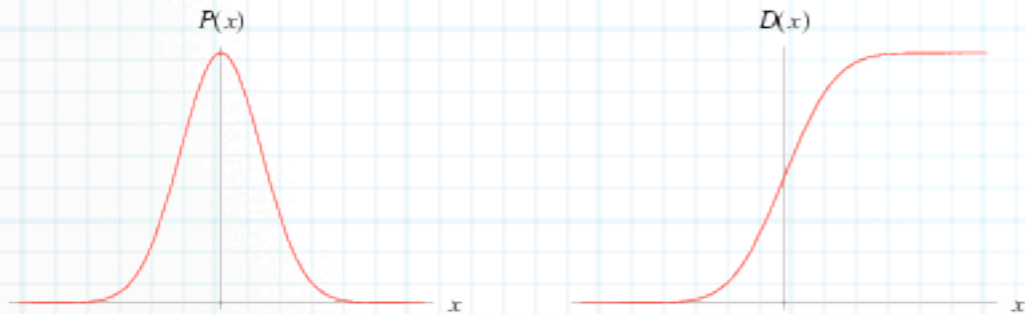
# High-Dimensional Note

Important note for working with depth in High-D

# Extreme Point Fraction vs. Dimension for Uniformly Random Points in a Hypercube



# Where this Applies



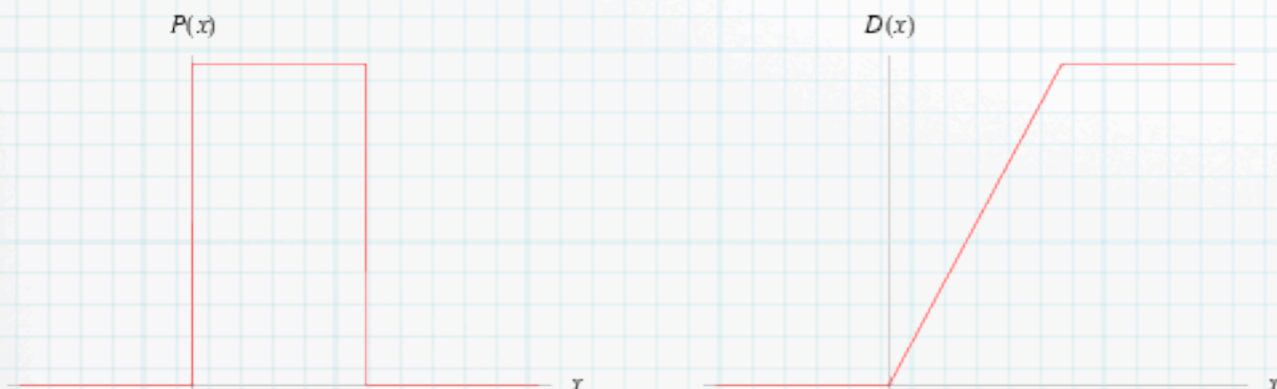
*Gaussian*

Seems Not

?

*Somewhere  
In-Between*

Unclear



*Uniform*

Yes



# Consequently...

- Many depth measures assign a depth of zero to points on the convex hull.
- These depth measures may be less effective on certain high-dimensional distributions.
- LI Depth has non-zero and non-uniform value at convex hull vertices, making it applicable in higher dimensions for more data-sets.
- Our refinement, PCA-Scaling LI Depth, also shares this helpful property.



# Ongoing Work

# 2006

- Windows & Linux clients
- 3D Visualizations and data sets
- Support additional depth measures

# 2007

and beyond

- Quantitative as well as visual results
- Expand domain into clustering and classification
- Allow use of DE functionality from other programming languages and tools
- Develop DE user and developer communities

# If You're Interested in DE...

Email Me: [jhugg@cs.tufts.edu](mailto:jhugg@cs.tufts.edu)

- What prevents you from using DE?
- What features would you find most valuable?
- Would you like to stay in contact?

# For More Info

<http://www.cs.tufts.edu/r/geometry/>

[jhugg@cs.tufts.edu](mailto:jhugg@cs.tufts.edu)


Computational Geometry at Tufts

06/02/2006 01:20 PM

Depth Explorer

06/02/2006 01:21 PM

## Computational Geometry at Tufts



**Research**

- Computational Statistics & Data depth
  - [Data Depth \(general\)](#)
  - Proximity Depth
  - [The Depth Explorer Software Tool](#)
  - [Half-Space Depth Contours](#)
  - [Half-Space Depth query using  \$O\(\log n\)\$  point location](#)
  - [Simplicial Depth](#)
- Topological Sweep
  - [Topological Sweep in Degenerated Cases](#)
  - [Topologically Sweeping the complete graph](#)
  - [LMS Regression using Guided Topological Sweep](#)

[Crossing number for points and segments](#)  
Dynamic Computation of The Ham-Sandwich Cut  
[Finding Knots and Links in Vector Fields](#)  
[Sphere Fitting in High Dimensions](#)

[Past Research](#)  
[Project Proposals](#) (Tufts students and faculty only)

**Related sites**  
[Artificial Intelligence at Tufts](#)

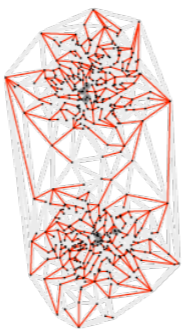
**People**

**Current**

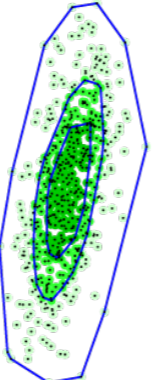
- [Prof. Diane Souvaine](#)
- [Eynat Rafalin](#)
- John Hugg
- [Alexandra Lauric](#)
- [Kathryn Seyboth](#)

**Former**

- Victoria Brumberg
- Michael Burr
- Ryan Coleman
- Elena Jakubiak
- Alok Lal
- Marcia Lazo
- Jeff Lindy
- Janet Luan
- Kim Miller
- Tim Mitchell
- Nikolai Shvertner
- Ori Taka



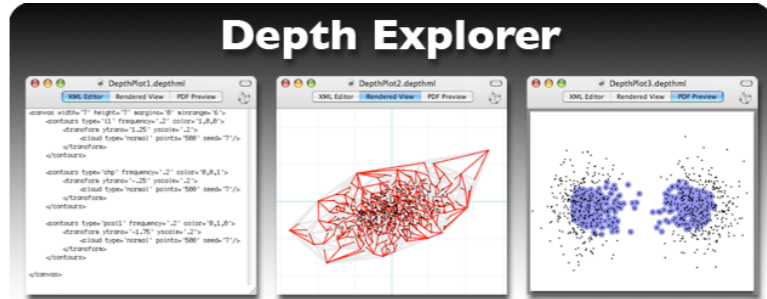
Computation of Proximity Depth using the Delaunay Triangulation.



Data points assigned to one third quantiles according to Halfspace Depth with overlaid contours.

## Tufts University Computational Geometry Research Group

### Depth Explorer

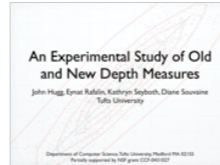


**What is Depth Explorer?**

Depth Explorer is a visual interactive tool for learning about Data Depth and evaluating depth measures.

Depth Explorer can generate diverse data sets or load existing ones. Through multiple visualizations, the performance of different depth measures can be compared visually, quickly and effectively.

**Depth Explorer at ALENEX 06**



Depth Explorer 1.0 made its debut at [ALENEX 06](#).

The [slides](#) (pdf) of the presentation have been made available here.

The full [text](#) (pdf) of the paper from the proceedings is also available.

**Depth Explorer Documentation**

Please note: This documentation is a work in progress. Expanded documentation will be a primary focus of incremental point releases.

[Data Depth Background](#)      [Example Documents](#)  
[Getting Started](#)              [XML Specification](#)  
[How Depth Explorer Works](#)      [Source Documentation](#)

**Download Depth Explorer**

[1.0 Application Disk Image](#)

**To Install:**

When the disk image mounts, copy the Depth Explorer application to your Applications folder.

**System Requirements:**

- Macintosh Laptop or Desktop
- G4 or G5 processor
- 256 MB memory
- OS X 10.4

**Download Source Code**

[1.0 Source Disk Image](#)

Depth Explorer is released under the terms of the [GNU Public License](#).

*Those wishing to explore or modify the Depth Explorer source may wish to wait until the next version (1.1). This revision, planned for February, will primarily bring improvements to the source and source documentation.*

**Contact the Author**