

DimStiller: Workflows for Dimensional Analysis and Reduction



Stephen Ingram, Tamara Munzner



Veronika Irvine, Melanie Tory



Steven Bergner, Torsten Möller

Overview

- Dimensionality Reduction
- Users
- Related Work
- Guidance
- DimStiller

Dimension(ality) Reduction

Dimension Reduction

Filter

Synthesize

Cull

Collect



dimensions are
uninteresting
(weight of spoon)



dimensions are
redundant
(caffeine + jitteriness)



complex combinations
of input dimensions
(nuttness, fruitiness)

http://www.luxuo.com/wp-content/uploads/2009/03/coffee_tongue.jpg

http://commons.wikimedia.org/wiki/File:A_small_cup_of_coffee.JPG

<http://api.ning.com/files/Gbc35LODOU25XeTeFJTf-J4a9yCLyeENOYyDV-WhmYU7-BtAkehxTvfzMzK8RoLf3v0JhACwOAMIYFCbNykIkYB-naNWGTW5/toomuchcoffee.gif>

Dimension Reduction

Filter

Synthesize

Cull



dimensions are
uninteresting
(weight of spoon)

Collect



dimensions are
redundant
(caffeine + jitteriness)



complex combinations
of input dimensions
(nuttness, fruitiness)

http://www.luxuo.com/wp-content/uploads/2009/03/coffee_tongue.jpg

http://commons.wikimedia.org/wiki/File:A_small_cup_of_coffee.JPG

<http://api.ning.com/files/Gbc35LODOU25XeTeFJTf-J4a9yCLyeENOYyDV-WhmYU7-BtAkehxTvfzMzK8RoLf3v0JhACwOAMIYFCbNykIkYB-naNWGTW5/toomuchcoffee.gif>

Dimension Reduction

Filter

Synthesize

Cull

Collect



dimensions are uninteresting (weight of spoon)



dimensions are redundant (caffeine + jitteriness)



complex combinations of input dimensions (nuttness, fruitiness)

http://www.luxuo.com/wp-content/uploads/2009/03/coffee_tongue.jpg

http://commons.wikimedia.org/wiki/File:A_small_cup_of_coffee.JPG

<http://api.ning.com/files/Gbc35LODOU25XeTeFJTf-J4a9yCLyeENOYyDV-WhmYU7-BtAkehxTvfzMzK8RoLf3v0JhACwOAMIYFCbNykIkYB-naNWGTW5/toomuchcoffee.gif>

Dimension Reduction

Filter

Synthesize

Cull

Collect



dimensions are
uninteresting
(weight of spoon)



dimensions are
redundant
(caffeine + jitteriness)



complex combinations
of input dimensions
(nuttness, fruitiness)

http://www.luxuo.com/wp-content/uploads/2009/03/coffee_tongue.jpg

http://commons.wikimedia.org/wiki/File:A_small_cup_of_coffee.JPG

<http://api.ning.com/files/Gbc35LODOU25XeTeFJTf-J4a9yCLyeENOYyDV-WhmYU7-BtAkehxTvfzMzK8RoLf3v0JhACwOAMIYFCbNykIkYB-naNWGTW5/toomuchcoffee.gif>

Dimension Reduction

Filter

Synthesize

Cull



dimensions are
uninteresting
(weight of spoon)

Collect



dimensions are
redundant
(caffeine + jitteriness)



complex combinations
of input dimensions
(nuttness, fruitiness)

http://www.luxuo.com/wp-content/uploads/2009/03/coffee_tongue.jpg

http://commons.wikimedia.org/wiki/File:A_small_cup_of_coffee.JPG

<http://api.ning.com/files/Gbc35LODOU25XeTeFJTf-J4a9yCLyeENOYyDV-WhmYU7-BtAkehxTvfzMzK8RoLf3v0JhACwOAMIYFCbNykIkYB-naNWGTW5/toomuchcoffee.gif>

Synthetic DR Example



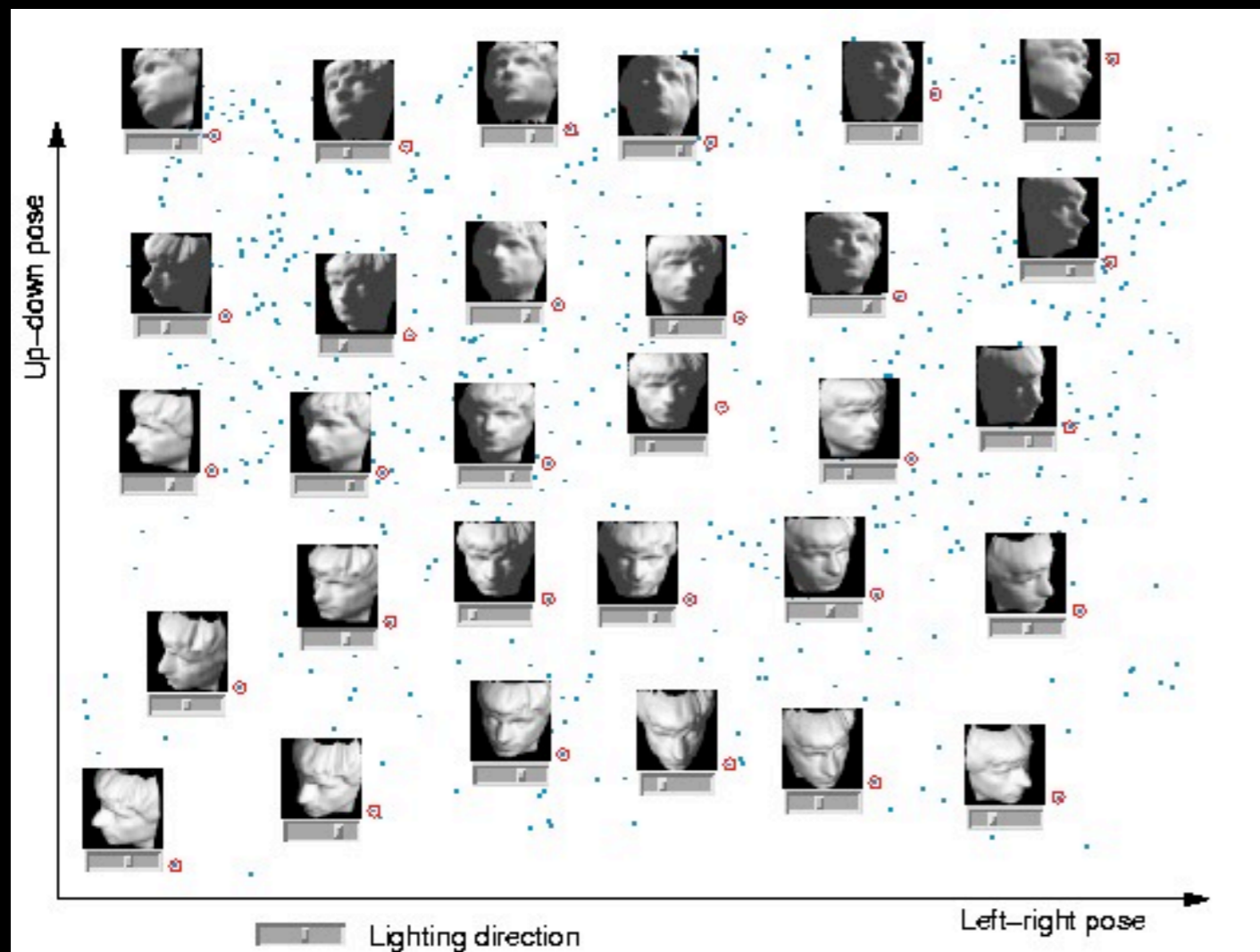
Face Image Dataset:

700 Faces

$35 \times 35 = 1225$ Dimensions

700 x 1225 Dataset

Synthetic DR Example



New Dataset:

700 Faces

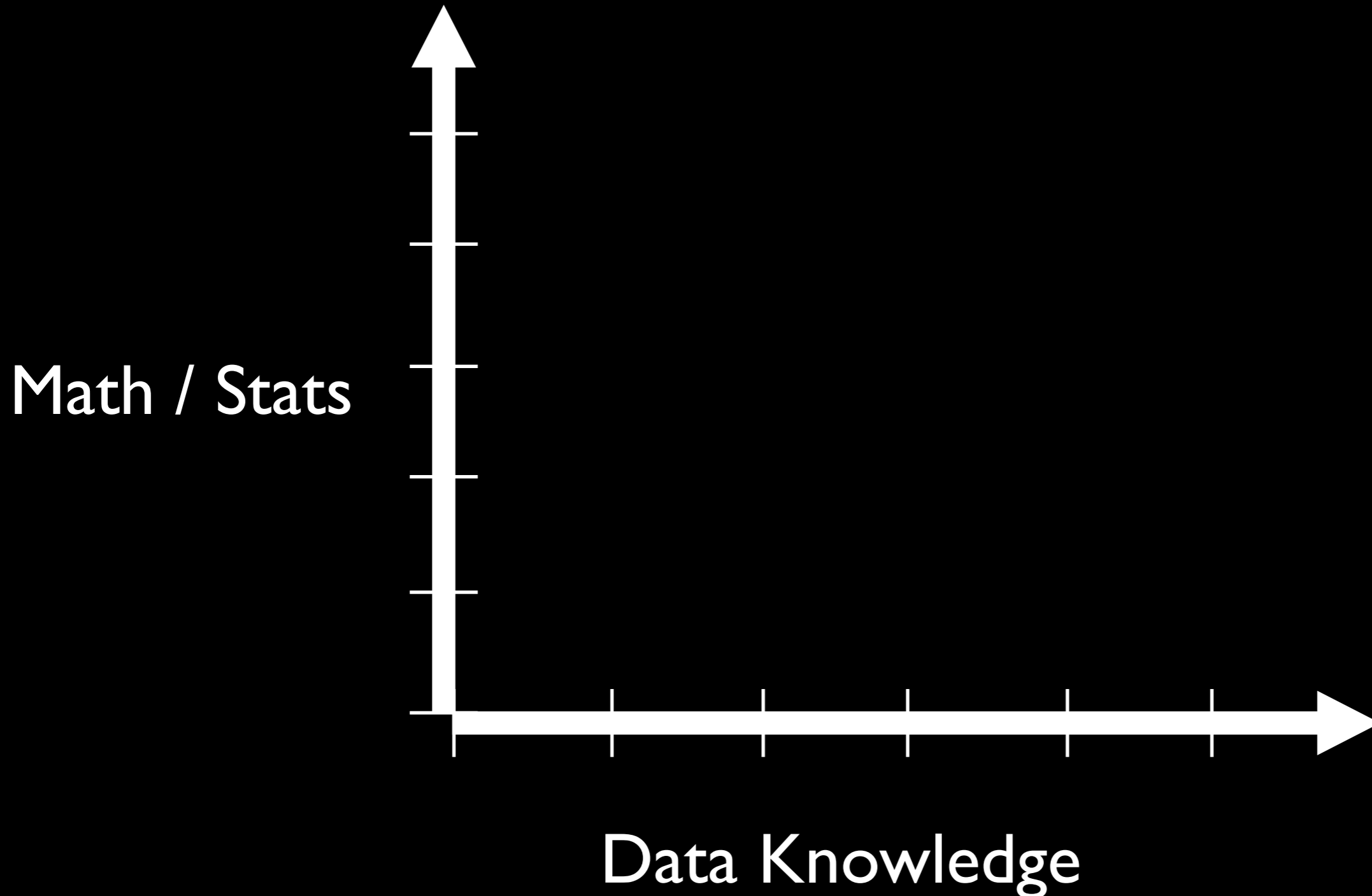
2 Dimensions

700 x 2 Dataset

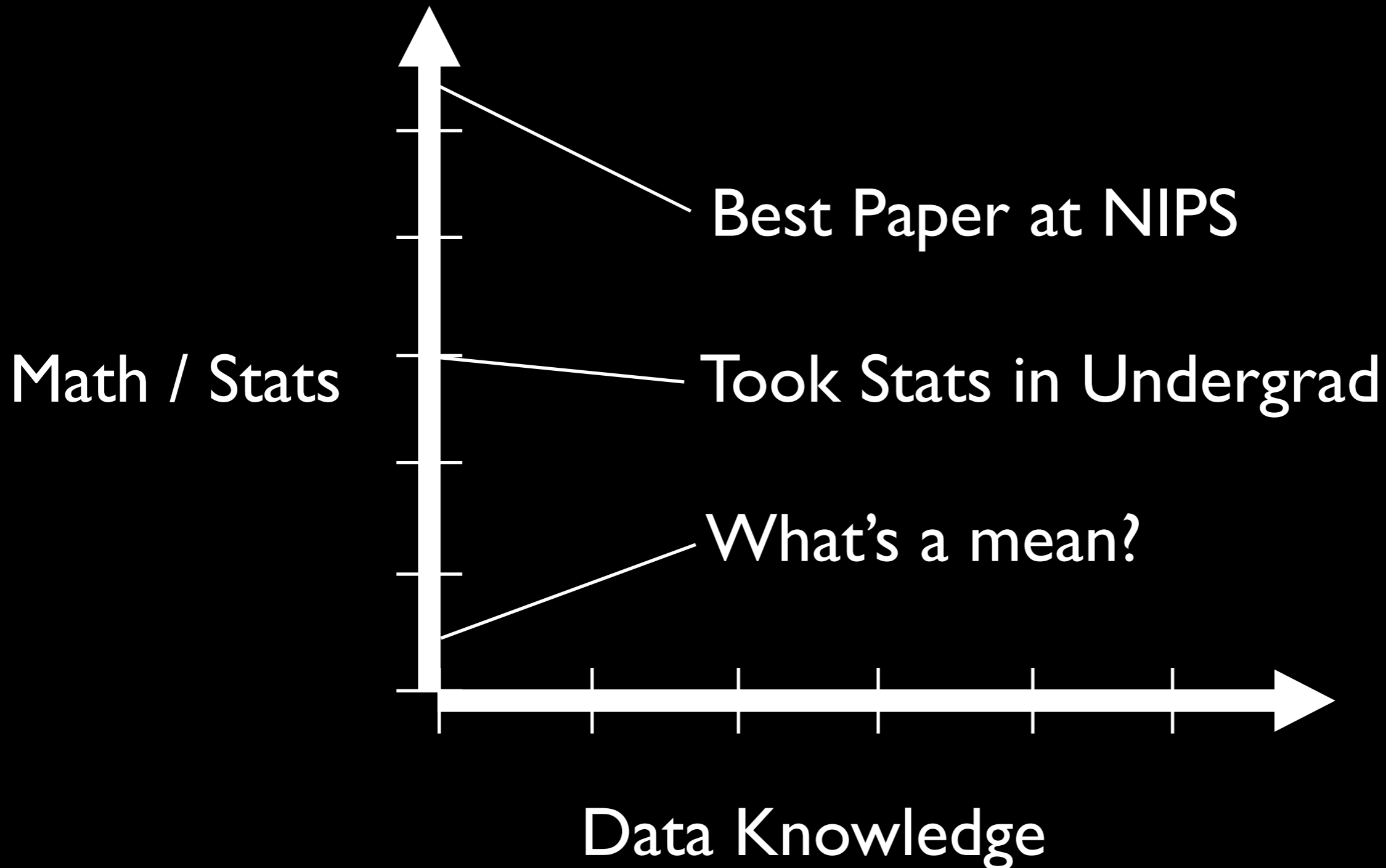
<http://isomap.stanford.edu/web3.jpg>

USERS

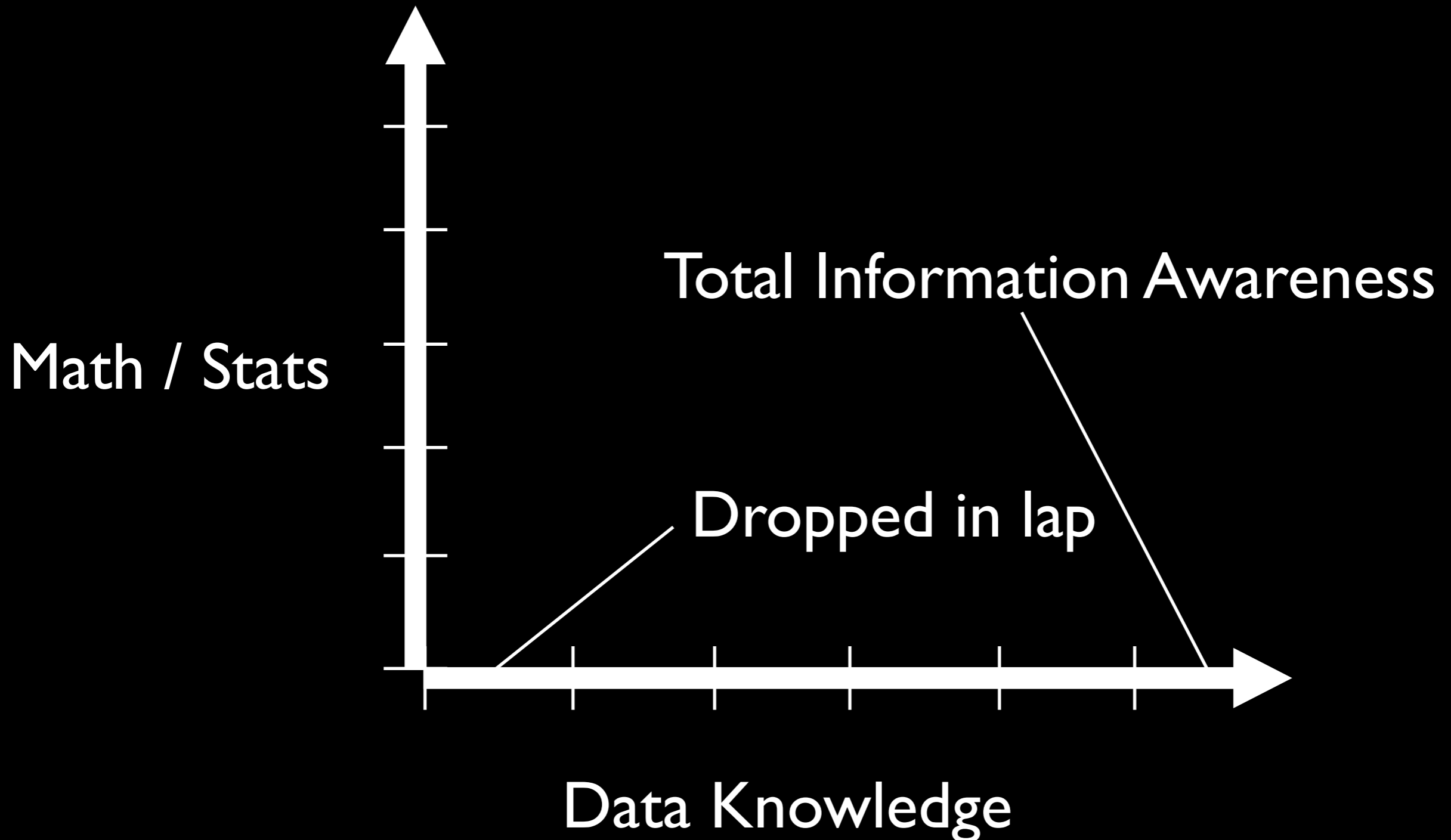
Visual High Dimensional Analysis (VHDA) User Map



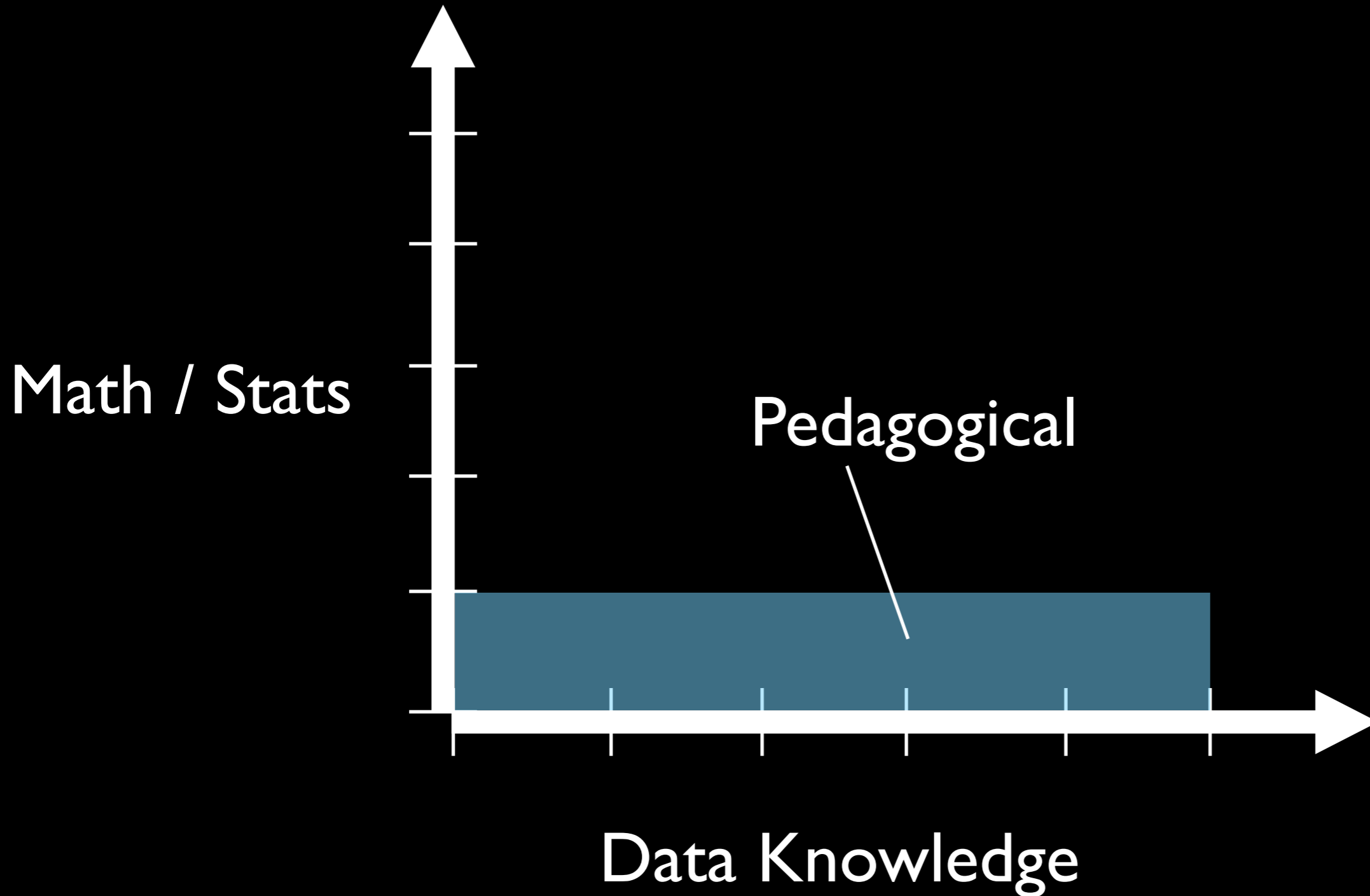
VHDA User Map



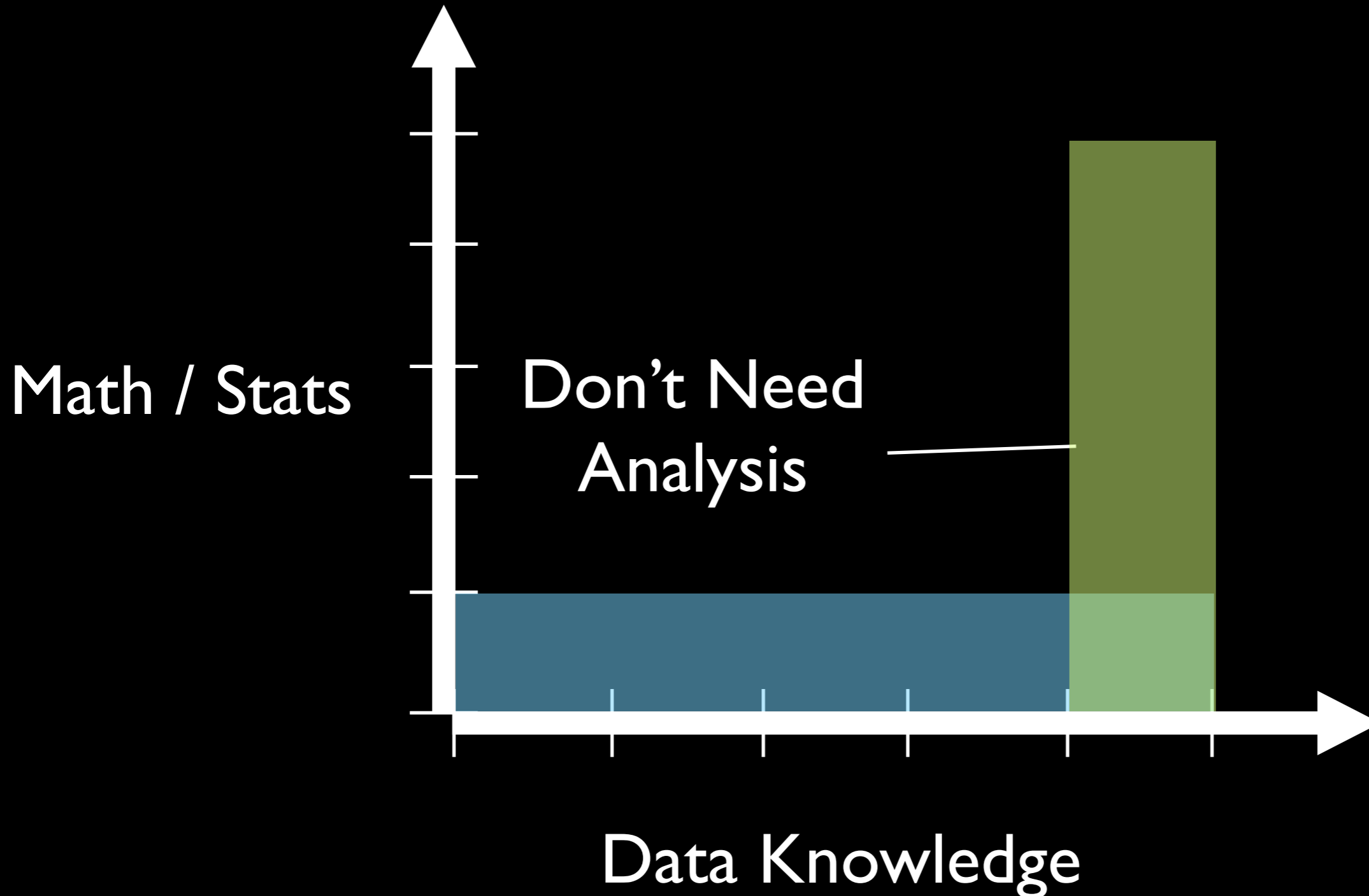
VHDA User Map



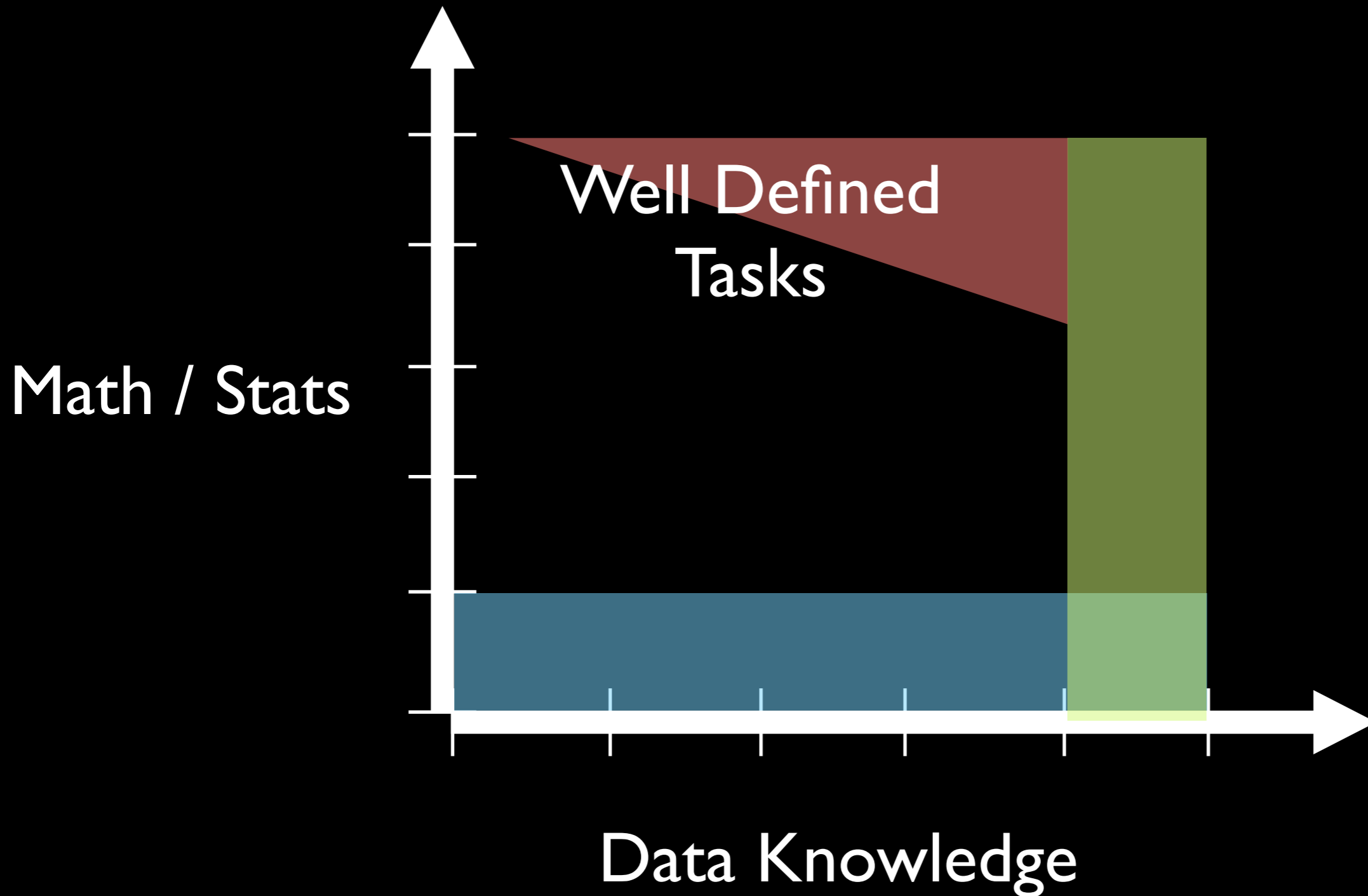
VHDA User Map



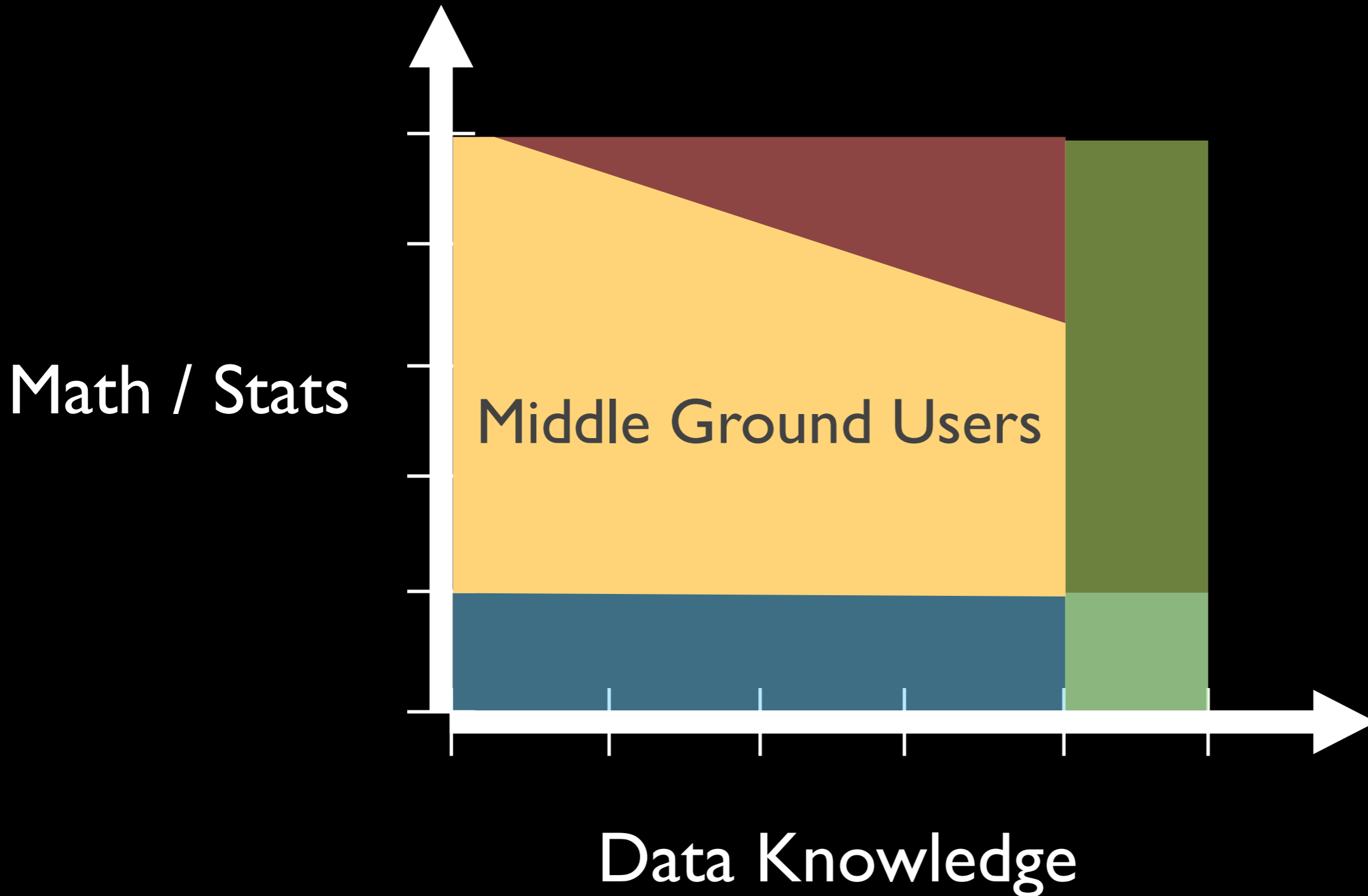
VHDA User Map



VHDA User Map

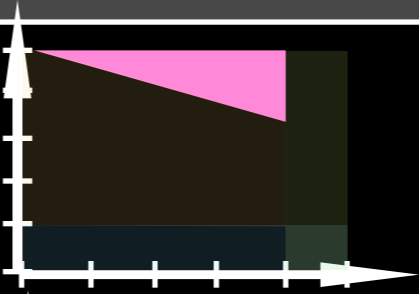
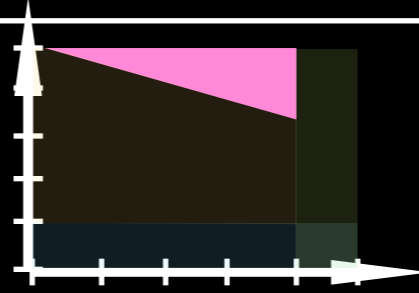
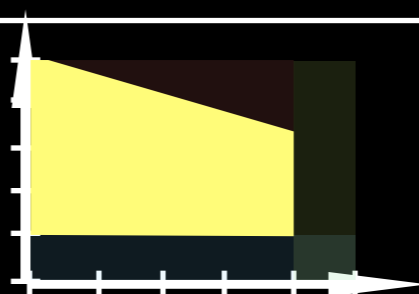
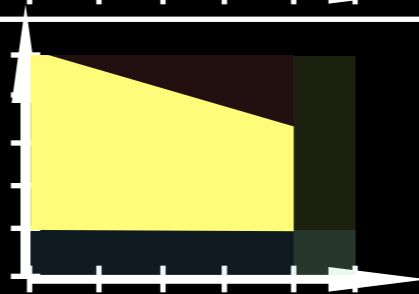


VHDA User Map



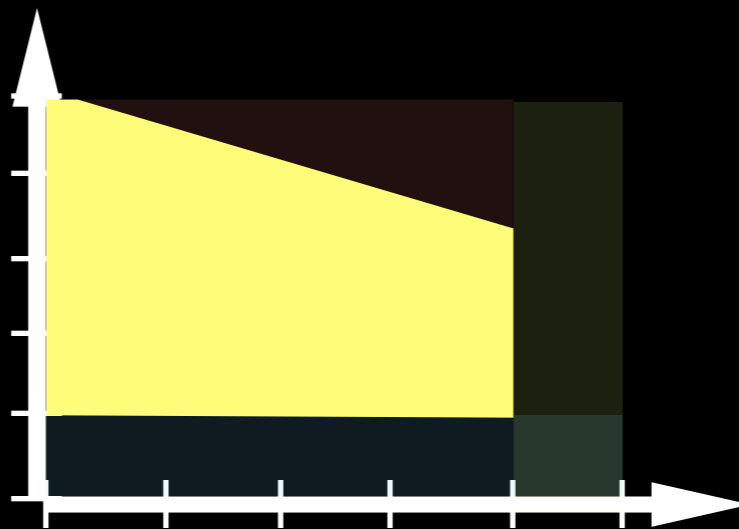
RELATED WORK

Other Systems

Tool	Target Users	Limitations
Matlab, R, etc.	 <p>A 2D plot with a vertical y-axis and a horizontal x-axis. A pink shaded area is bounded by the y-axis, a horizontal line at the top, and a diagonal line sloping downwards from left to right. A dark blue shaded area is a horizontal bar at the bottom. A dark green shaded area is a vertical bar on the right side.</p>	Needs Power Users
DR Toolkits	 <p>A 2D plot with a vertical y-axis and a horizontal x-axis. A pink shaded area is bounded by the y-axis, a horizontal line at the top, and a diagonal line sloping downwards from left to right. A dark blue shaded area is a horizontal bar at the bottom. A dark green shaded area is a vertical bar on the right side.</p>	Only Less Programming
XMDVTool, GGobi	 <p>A 2D plot with a vertical y-axis and a horizontal x-axis. A yellow shaded area is bounded by the y-axis, a horizontal line at the top, and a diagonal line sloping downwards from left to right. A dark blue shaded area is a horizontal bar at the bottom. A dark green shaded area is a vertical bar on the right side.</p>	No Guidance Beyond Vis
Johansson & Johansson 2009	 <p>A 2D plot with a vertical y-axis and a horizontal x-axis. A yellow shaded area is bounded by the y-axis, a horizontal line at the top, and a diagonal line sloping downwards from left to right. A dark blue shaded area is a horizontal bar at the bottom. A dark green shaded area is a vertical bar on the right side.</p>	No Synthetic DR

Hole In Prev Work

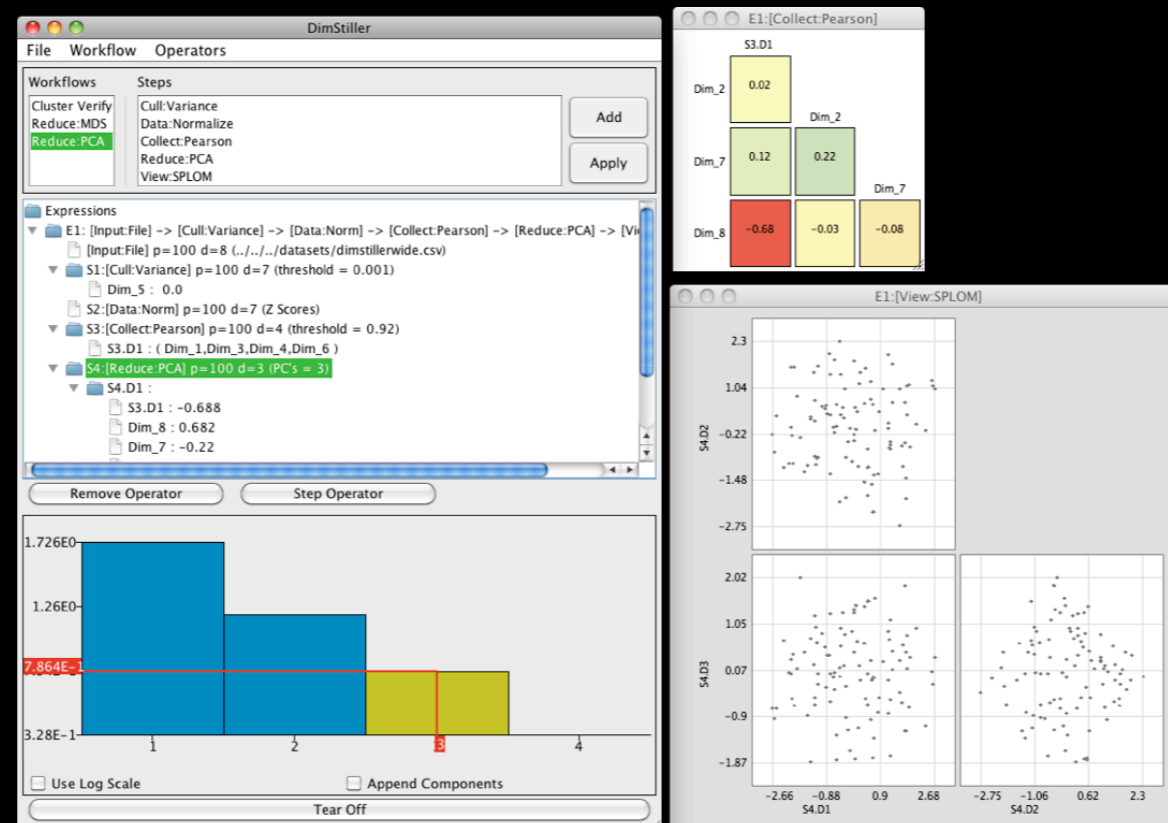
- Access To Range Of DR Algos



- Guidance For Middle Ground Users

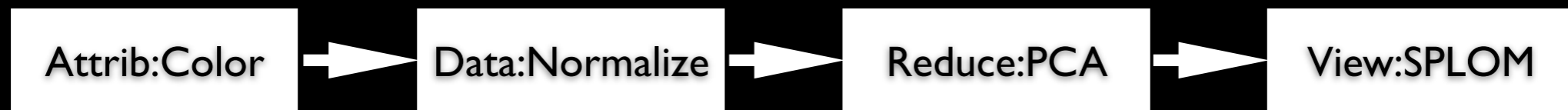
Contributions

Design and Implementation of DimStiller

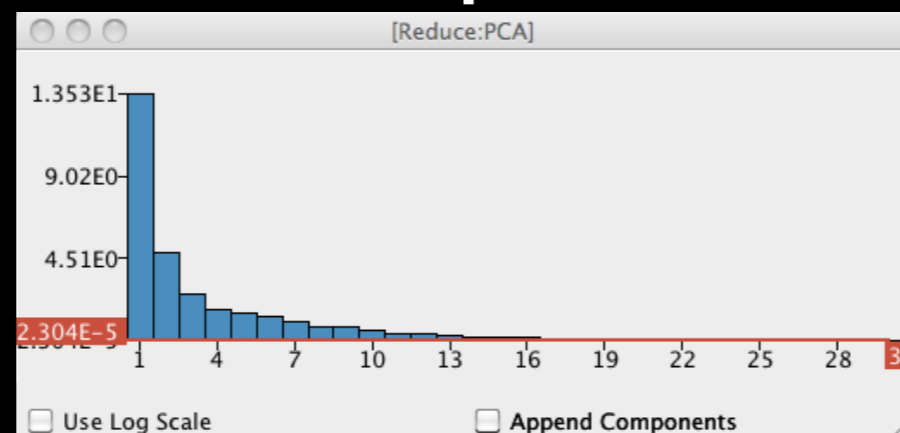


Global and Local Guidance

Global : Workflows



Local: Operators



GUIDANCE

Sloppy,
Misunderstood

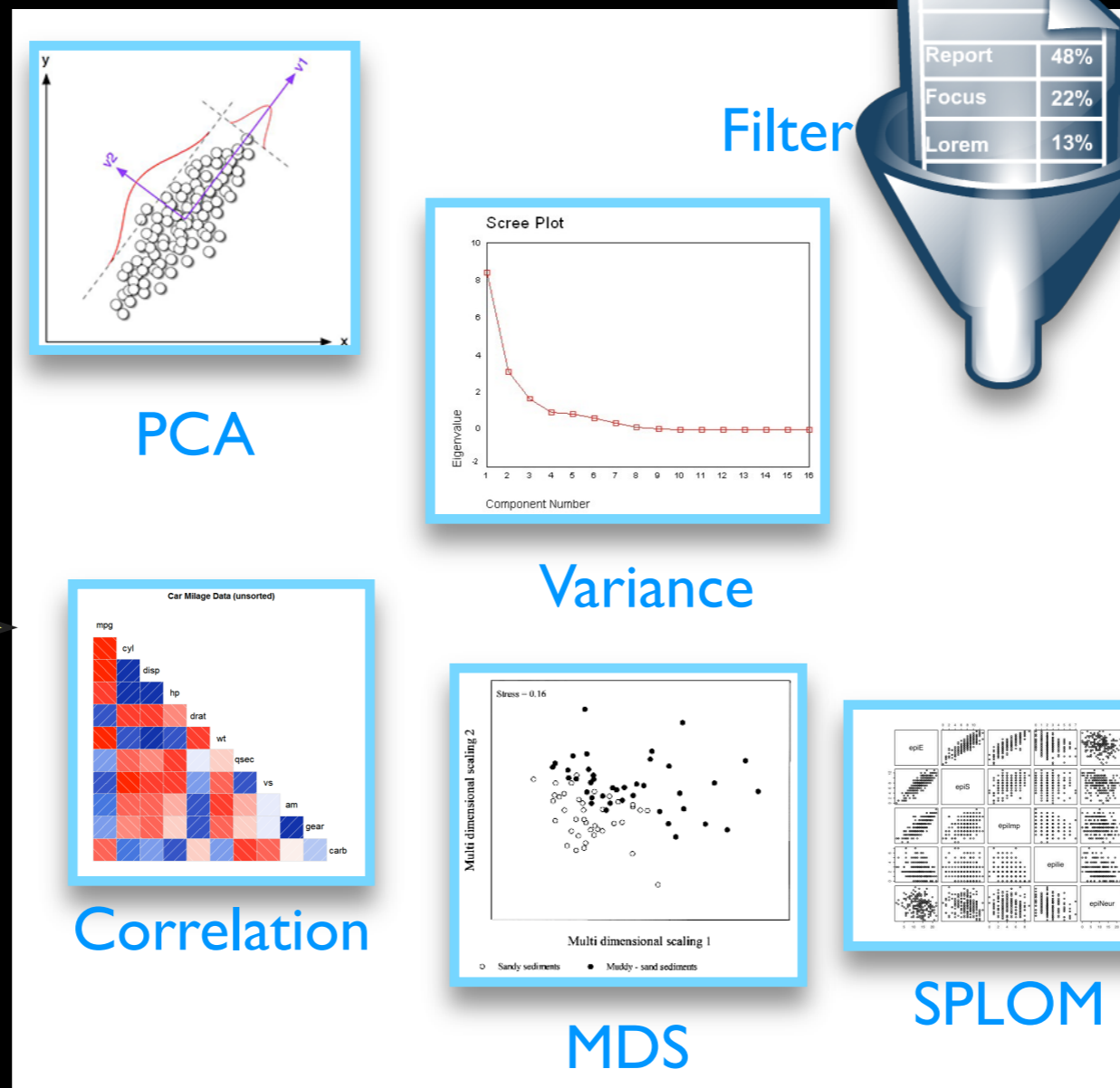


Compact,
Evocative

Operator Space

Which Operations and What Order?

Sloppy,
Misunderstood



Operator Space

<http://www.cs.cornell.edu/courses/cs322/2008sp/schedule.html>

<http://www.statmethods.net/advgraphs/images/corrgram3.png>

http://en.wikibooks.org/wiki/File:Scree_plot_for_the_initial_dataset_Figure_36.jpg

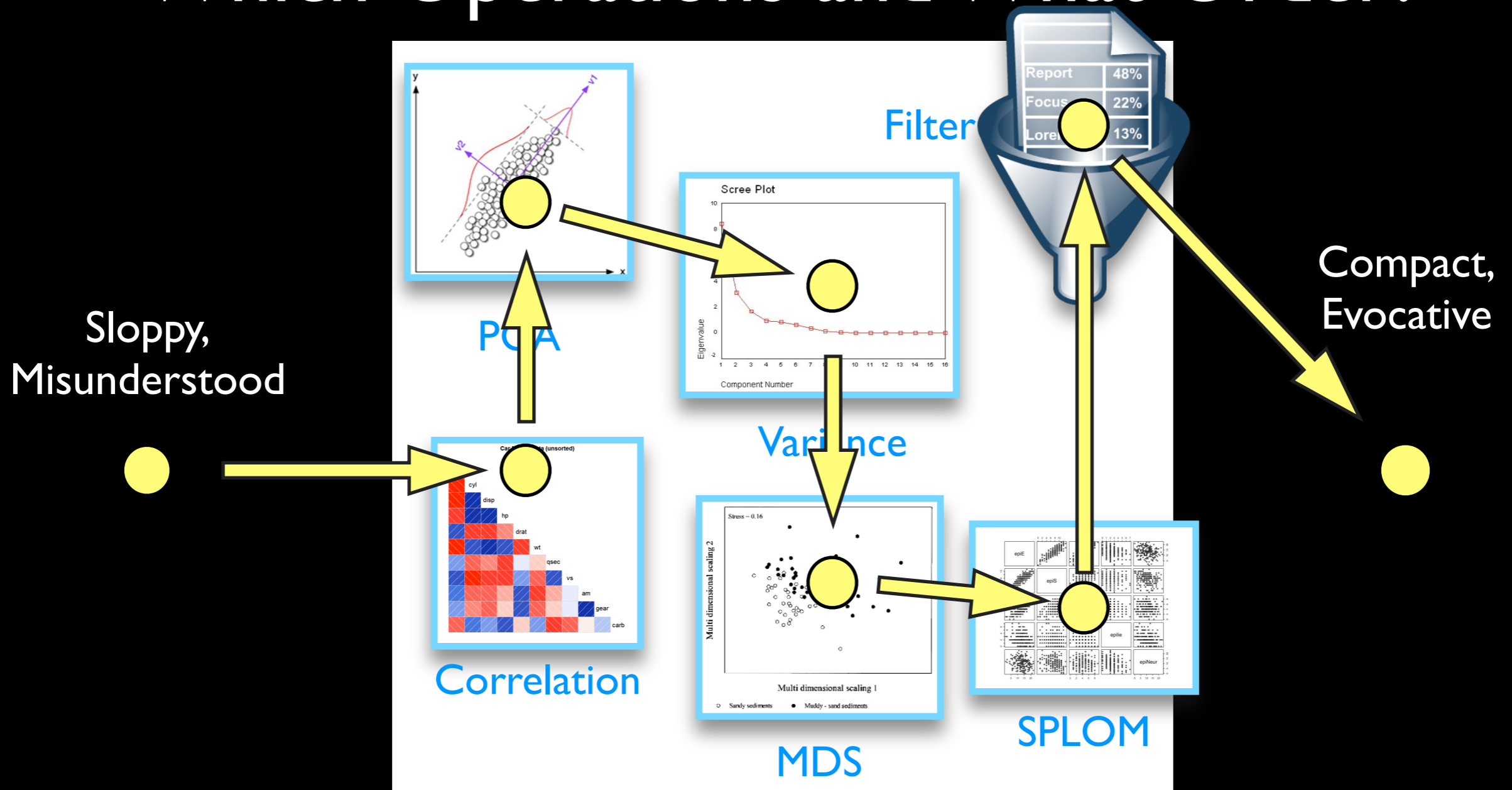
http://www.scielo.cl/scielo.php?pid=S0716-078X2001000200019&script=sci_arttext

http://www.iconfinder.com/icondetails/44818/400/data_filter_icon?r=1

<http://www.personality-project.org/R/>

Global Guidance

Which Operations and What Order?



Operator Space

<http://www.cs.cornell.edu/courses/cs322/2008sp/schedule.html>

<http://www.statmethods.net/advgraphs/images/corrgram3.png>

http://en.wikibooks.org/wiki/File:Scree_plot_for_the_initial_dataset_Figure_36.jpg

http://www.scielo.cl/scielo.php?pid=S0716-078X2001000200019&script=sci_arttext

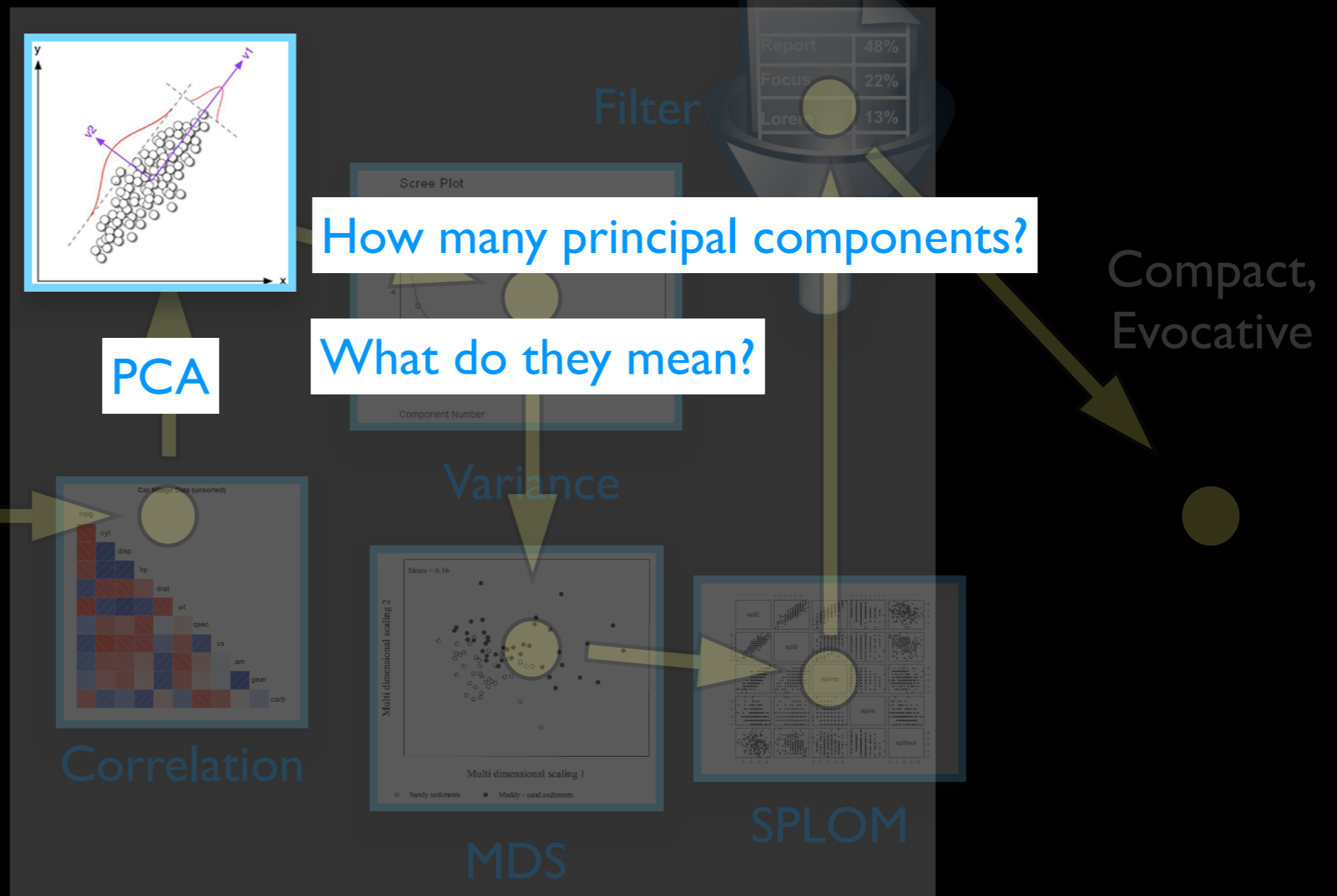
http://www.iconfinder.com/icondetails/44818/400/data_filter_icon?r=1

<http://www.personality-project.org/R/>

Local Guidance

What to do with a given operator?

Sloppy,
Misunderstood



Operator Space

<http://www.cs.cornell.edu/courses/cs322/2008sp/schedule.html>

<http://www.statmethods.net/advgraphs/images/corrgram3.png>

http://en.wikibooks.org/wiki/File:Scree_plot_for_the_initial_dataset_Figure_36.jpg

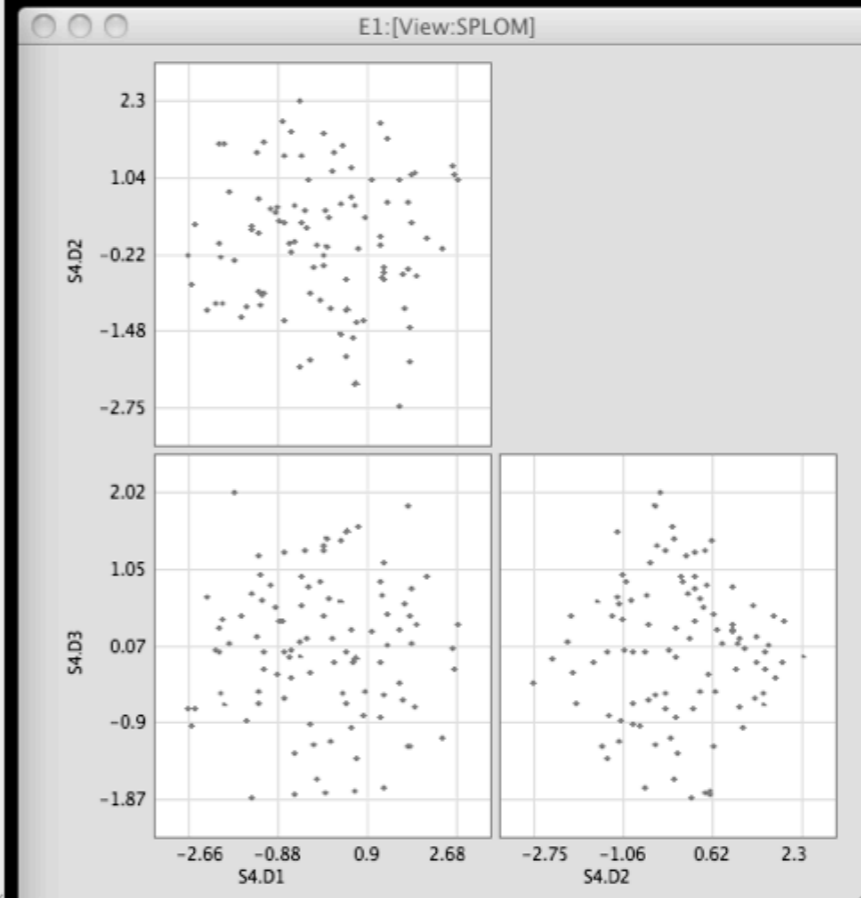
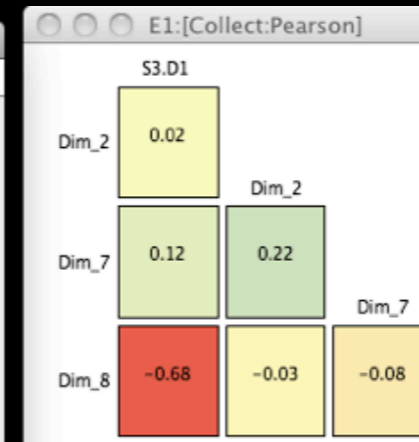
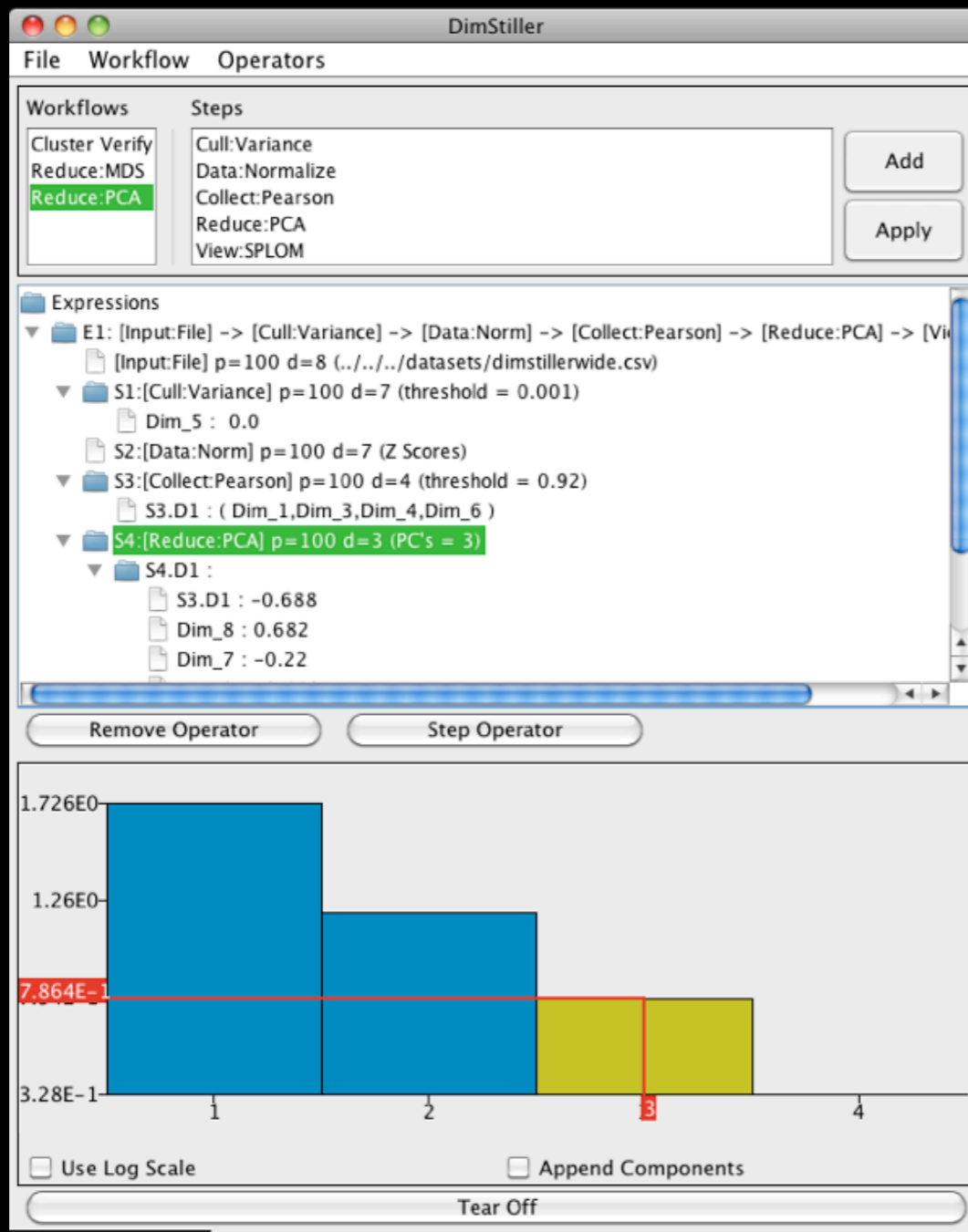
http://www.scielo.cl/scielo.php?pid=S0716-078X2001000200019&script=sci_arttext

http://www.iconfinder.com/icondetails/44818/400/data_filter_icon?r=1

<http://www.personality-project.org/R/>

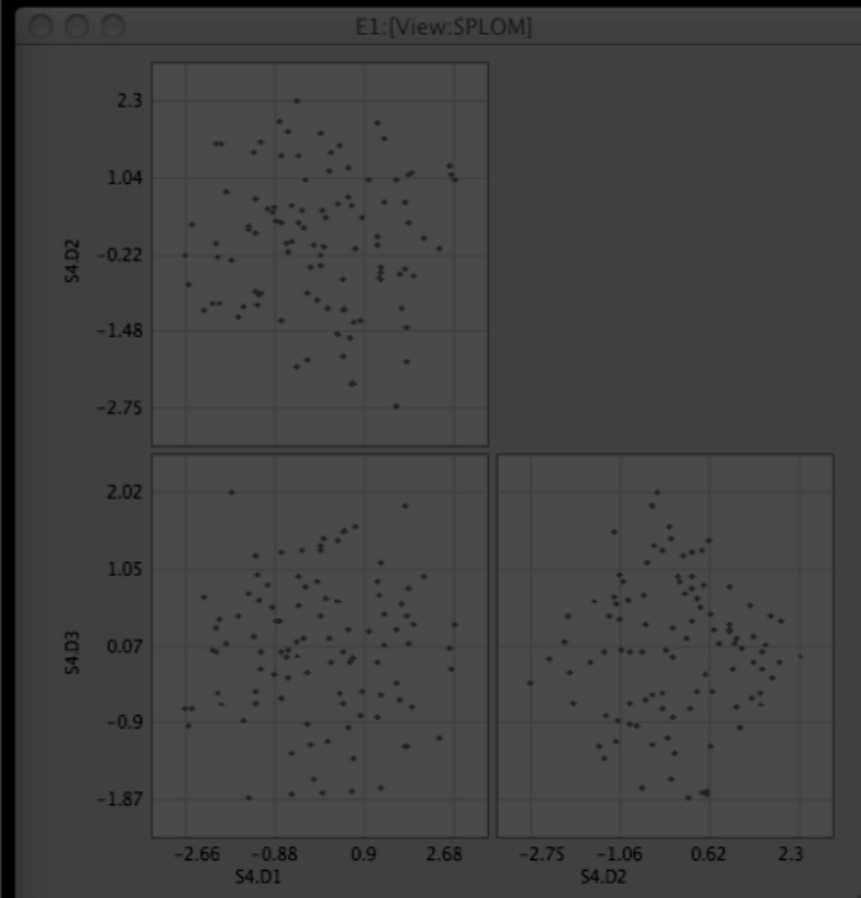
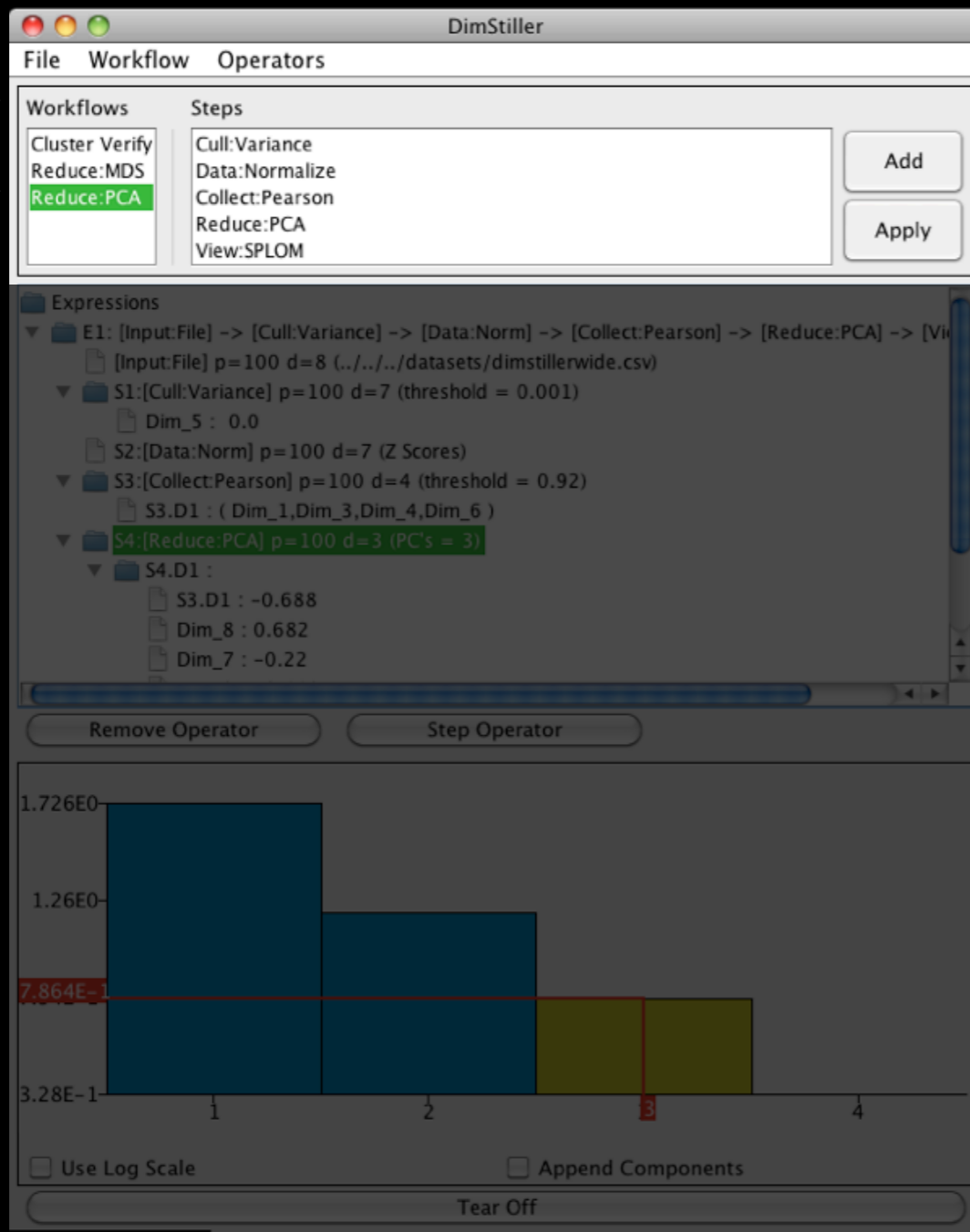
DIMSTILLER

DimStiller



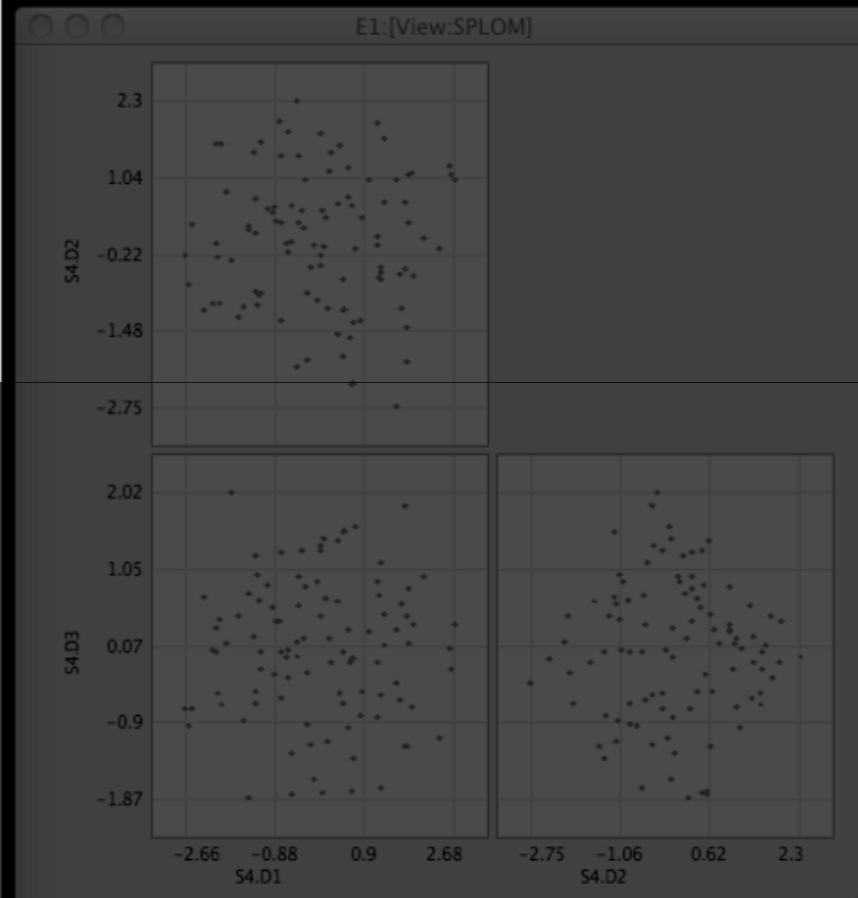
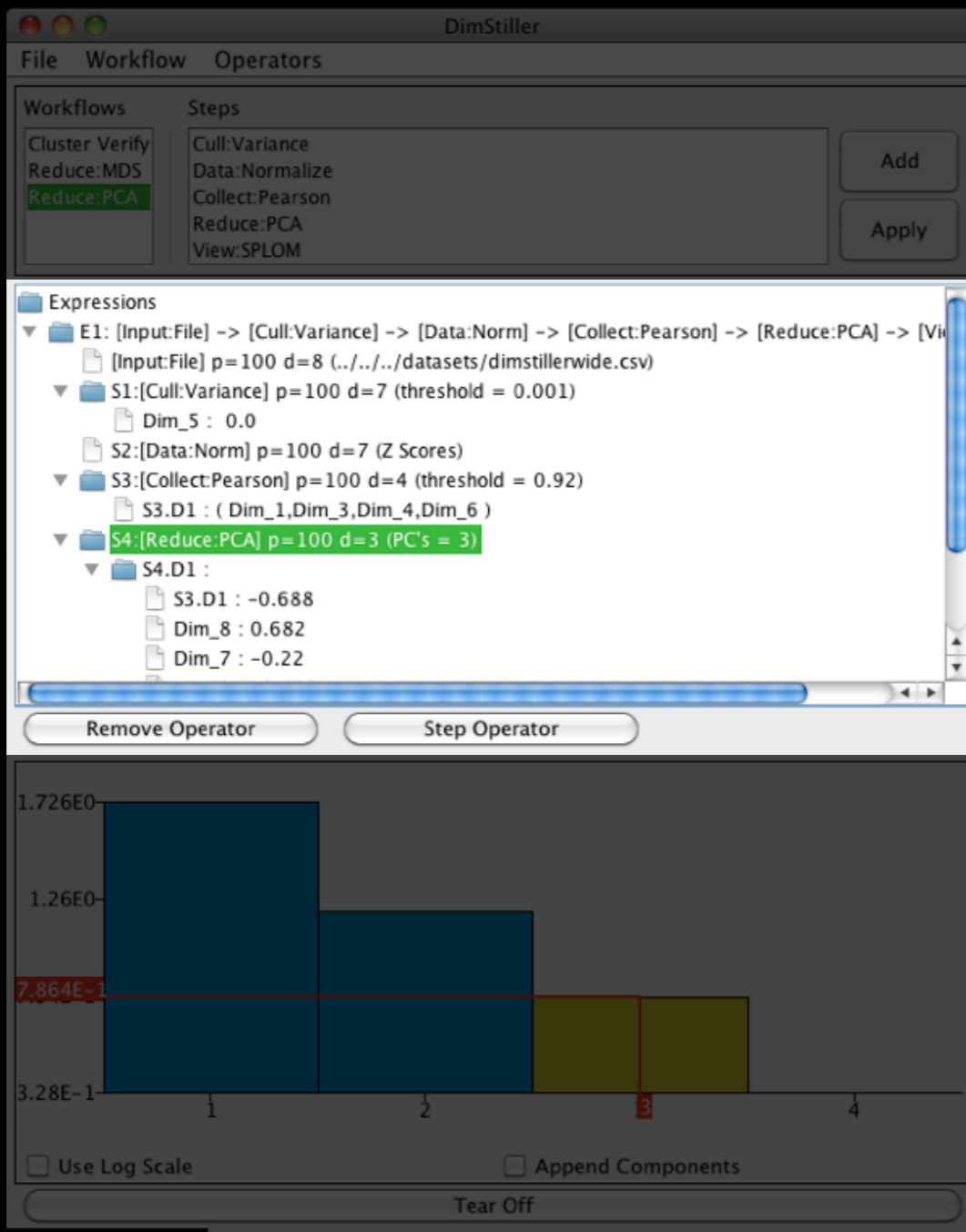
DimStiller

Workflow
Selector



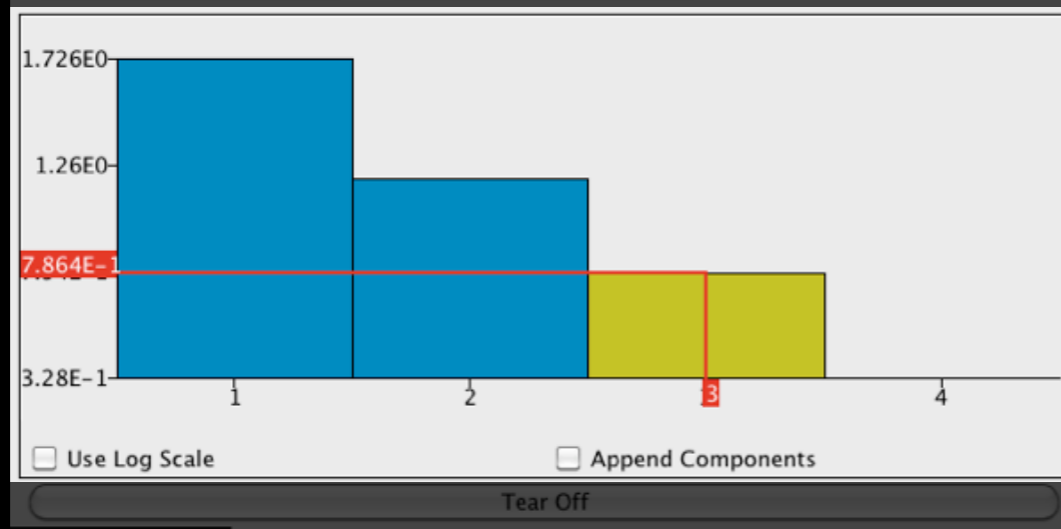
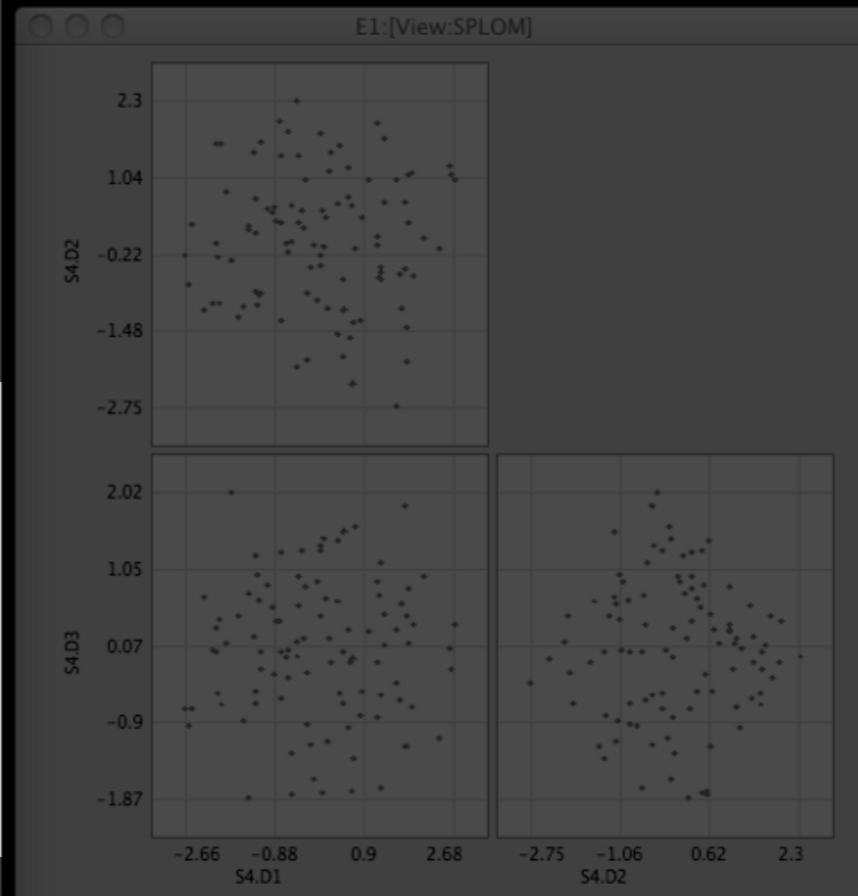
DimStiller

Expression
Tree



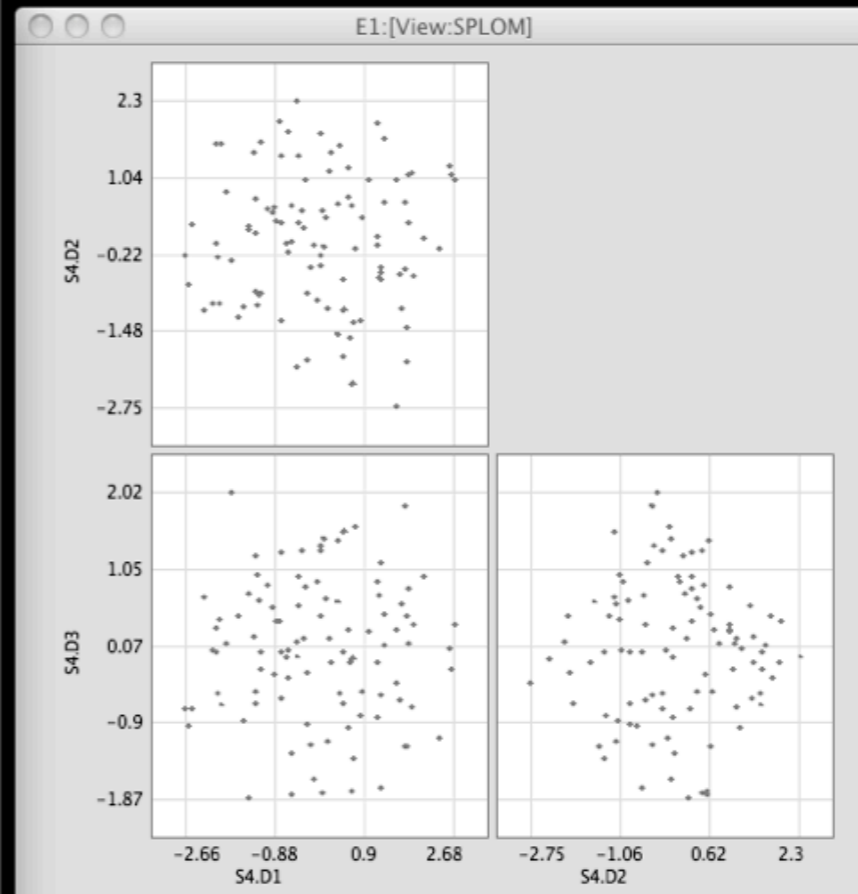
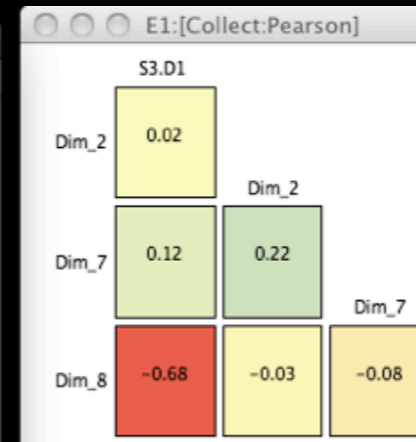
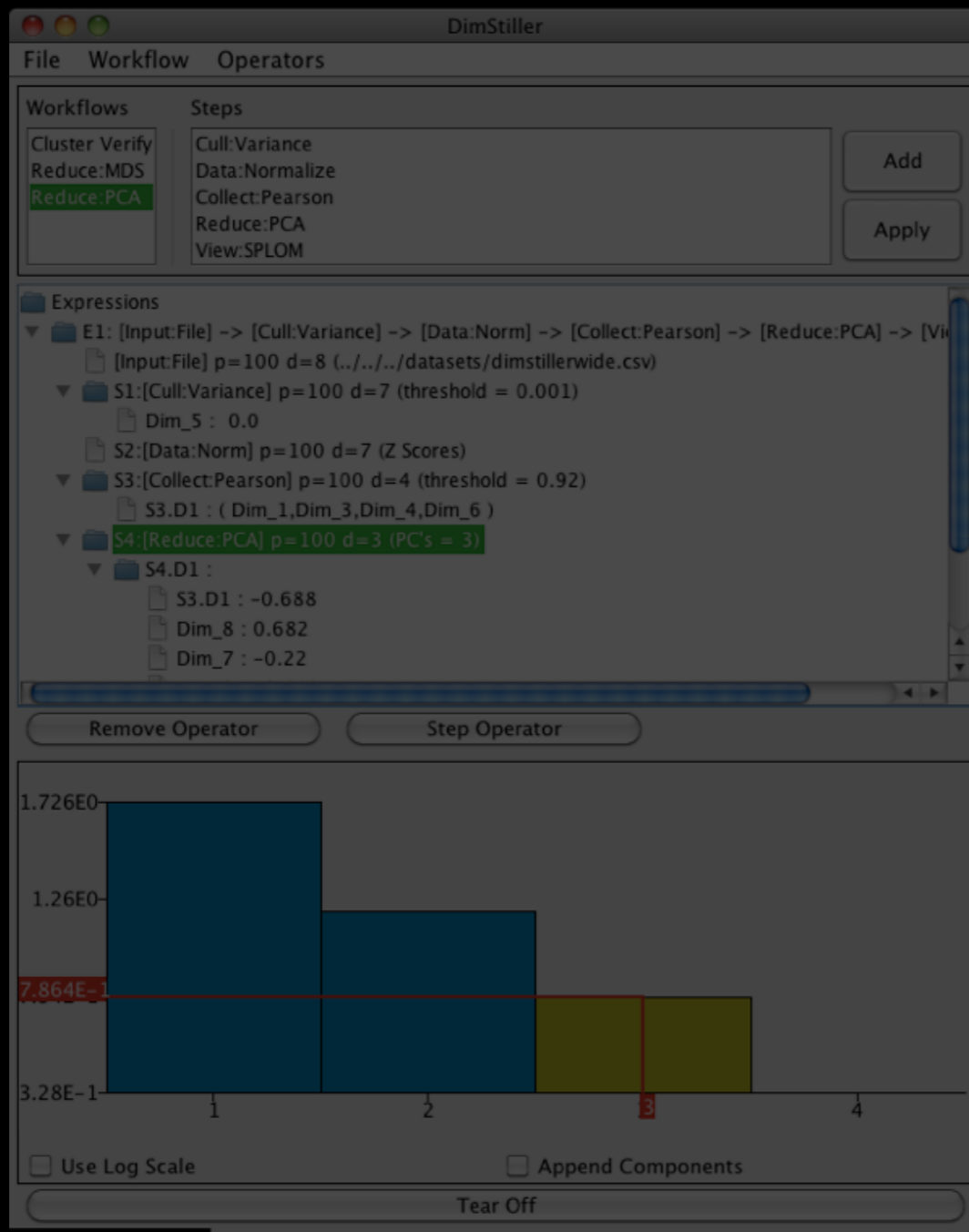
DimStiller

The screenshot shows the DimStiller application window with a menu bar (File, Workflow, Operators) and a main workspace. The workspace is divided into 'Workflows' and 'Steps' sections. The 'Steps' section contains a list of operations: Cull:Variance, Data:Normalize, Collect:Pearson, Reduce:PCA, and View:SPLOM. Below this is an 'Expressions' tree showing a sequence of operations: E1: [Input:File] -> [Cull:Variance] -> [Data:Norm] -> [Collect:Pearson] -> [Reduce:PCA] -> [View:SPLOM]. The 'Reduce:PCA' step is highlighted in green, and its parameters are shown as 'S4:[Reduce:PCA] p=100 d=3 (PC's = 3)'. Below the expressions are buttons for 'Remove Operator' and 'Step Operator'.



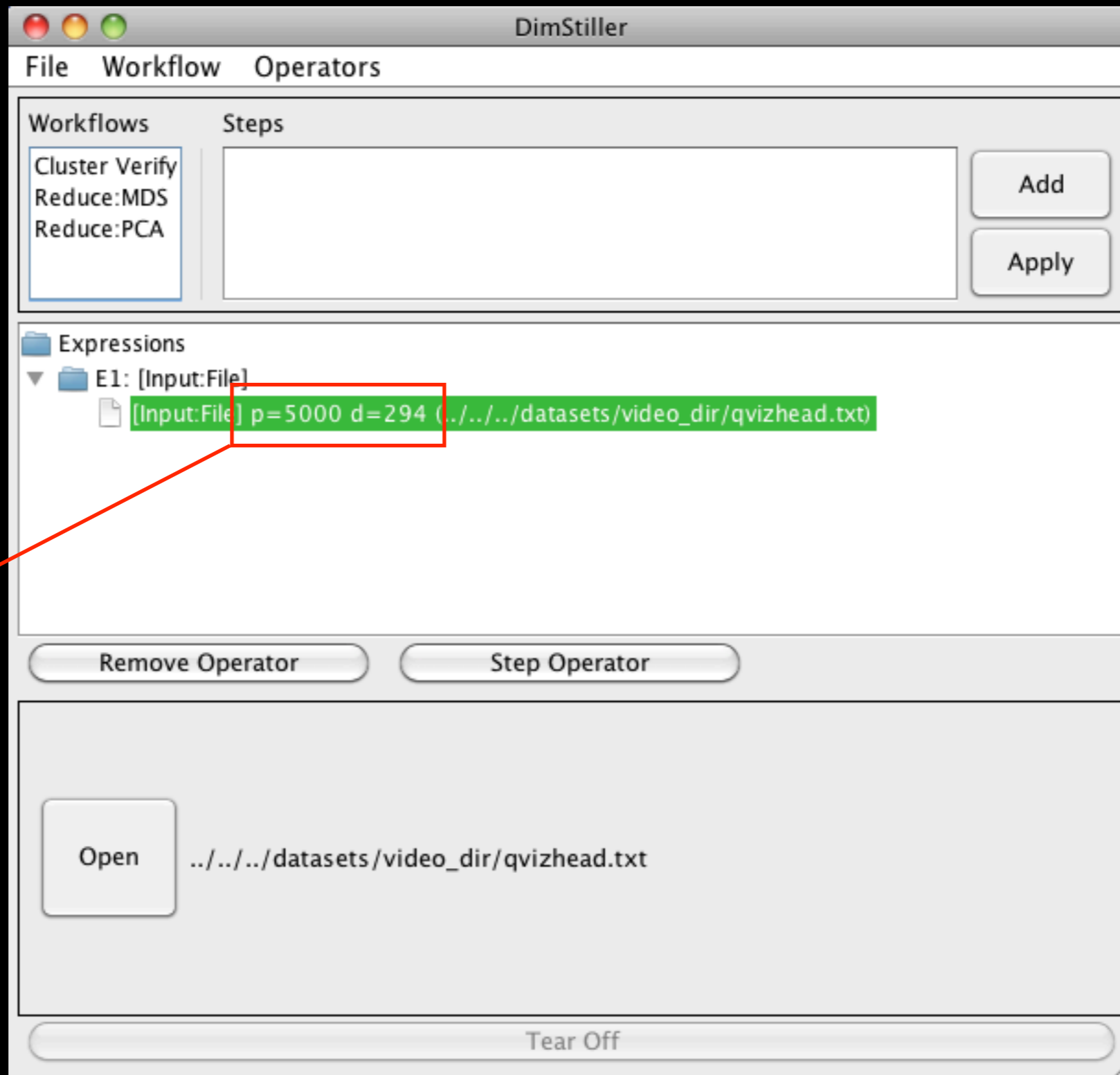
Operator Control

DimStiller



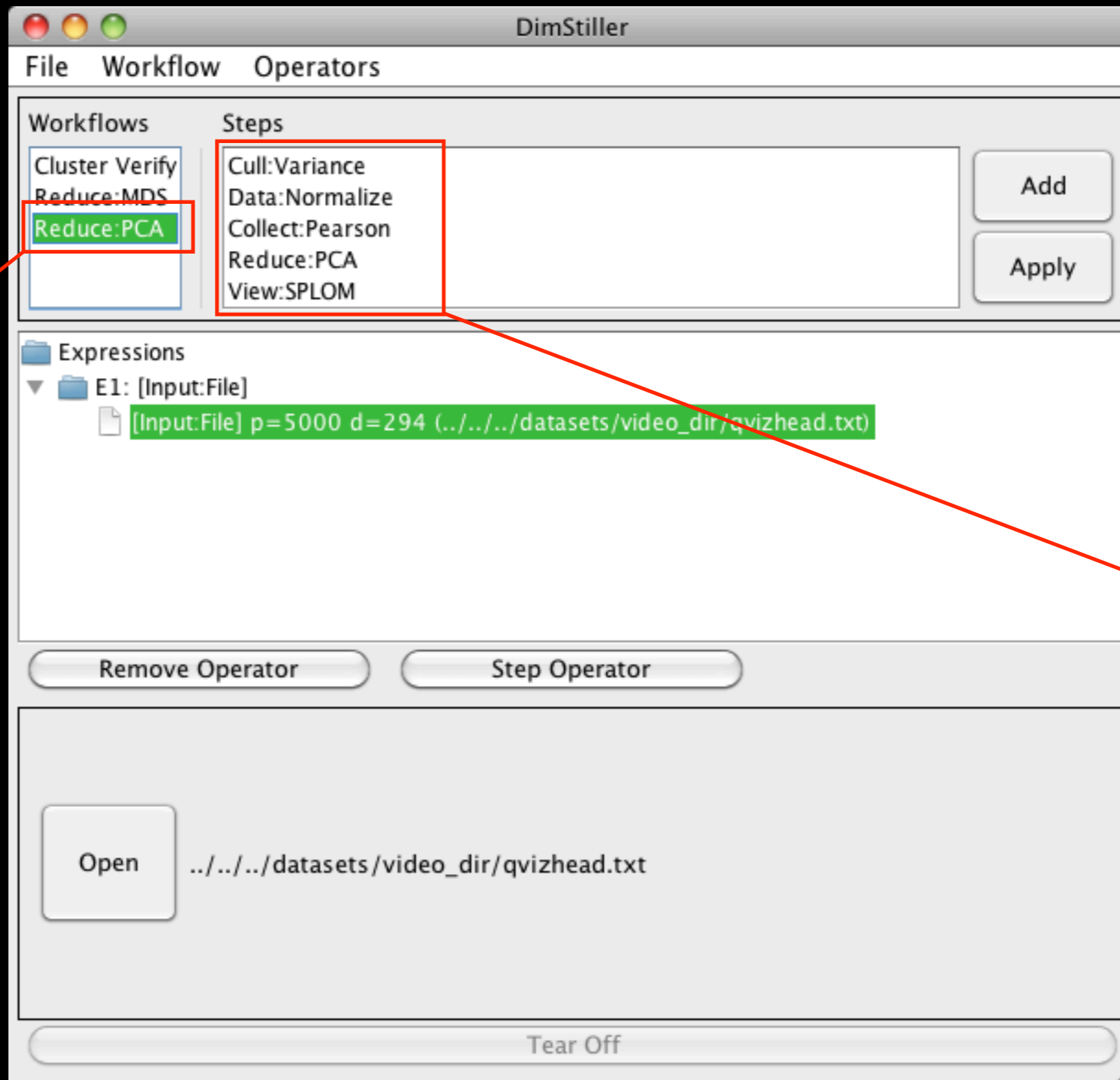
Operator Views

EXAMPLE



5000 pts
294 dim

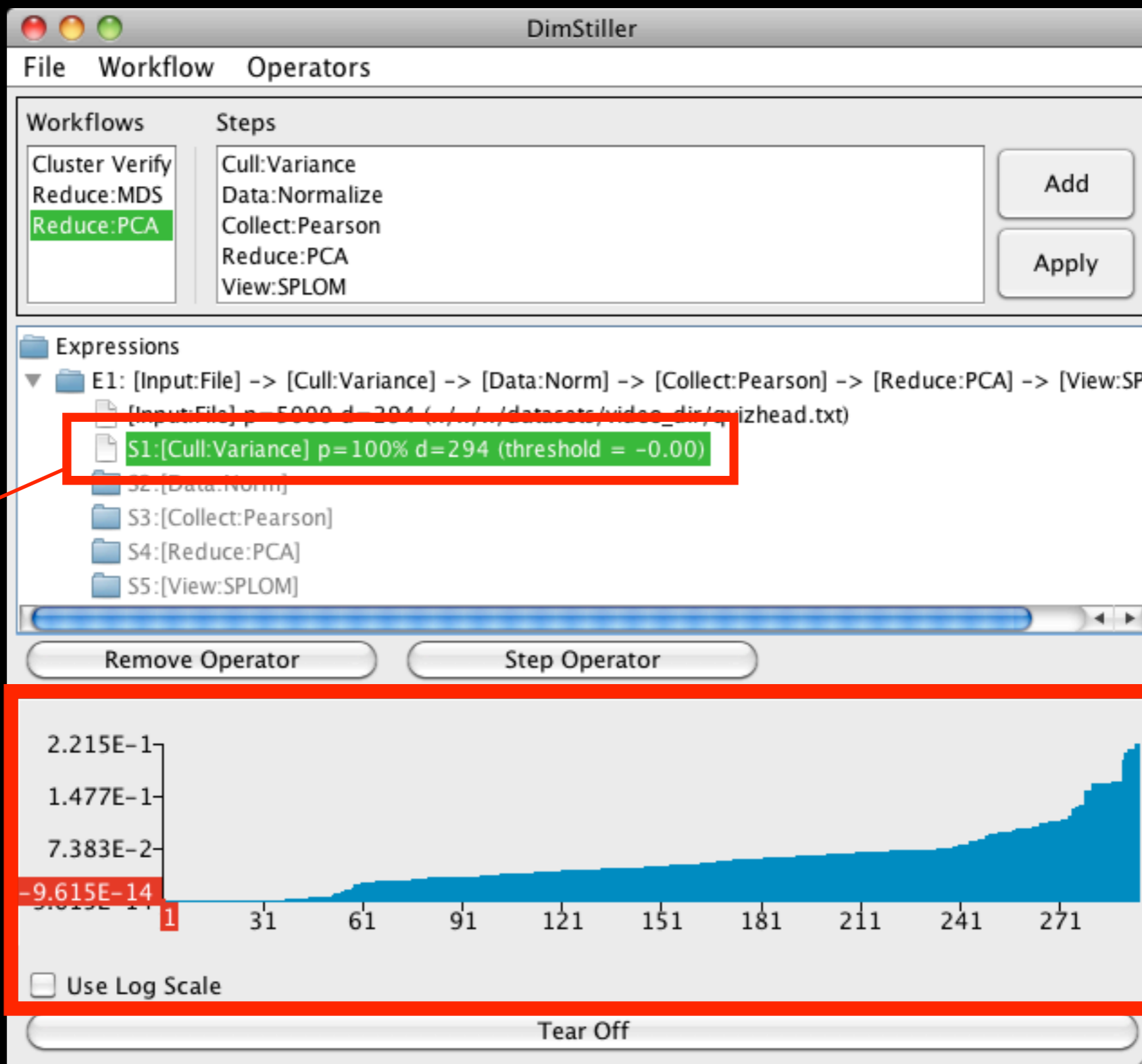
294 DIMS



Select
Reduce:PCA
Workflow

View
Operator
List Here

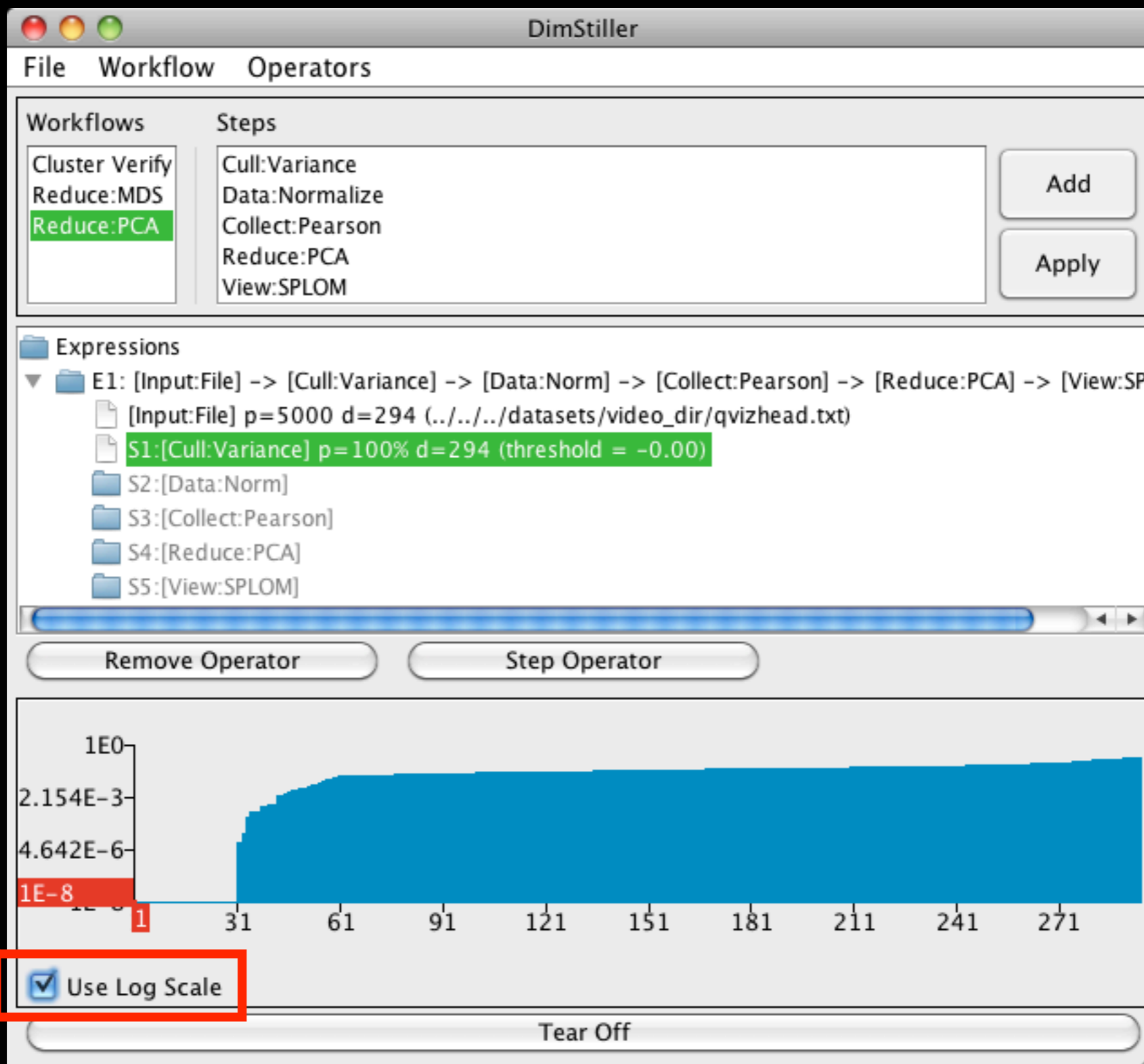
294 DIMS



Cull:Variance
Operator

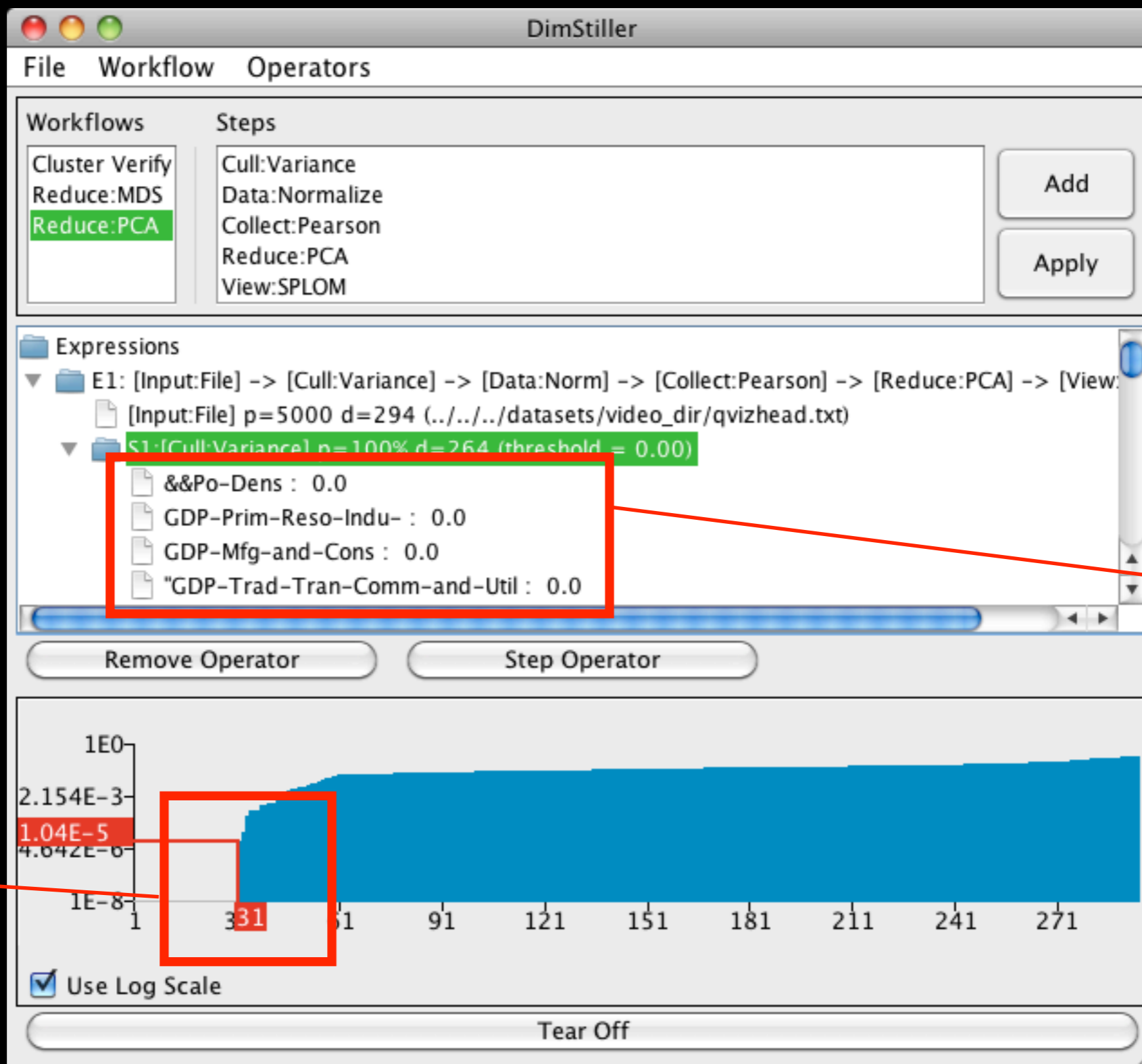
Scree Plot
of Variances

294 DIMS



Log-scale
for better
Visibility

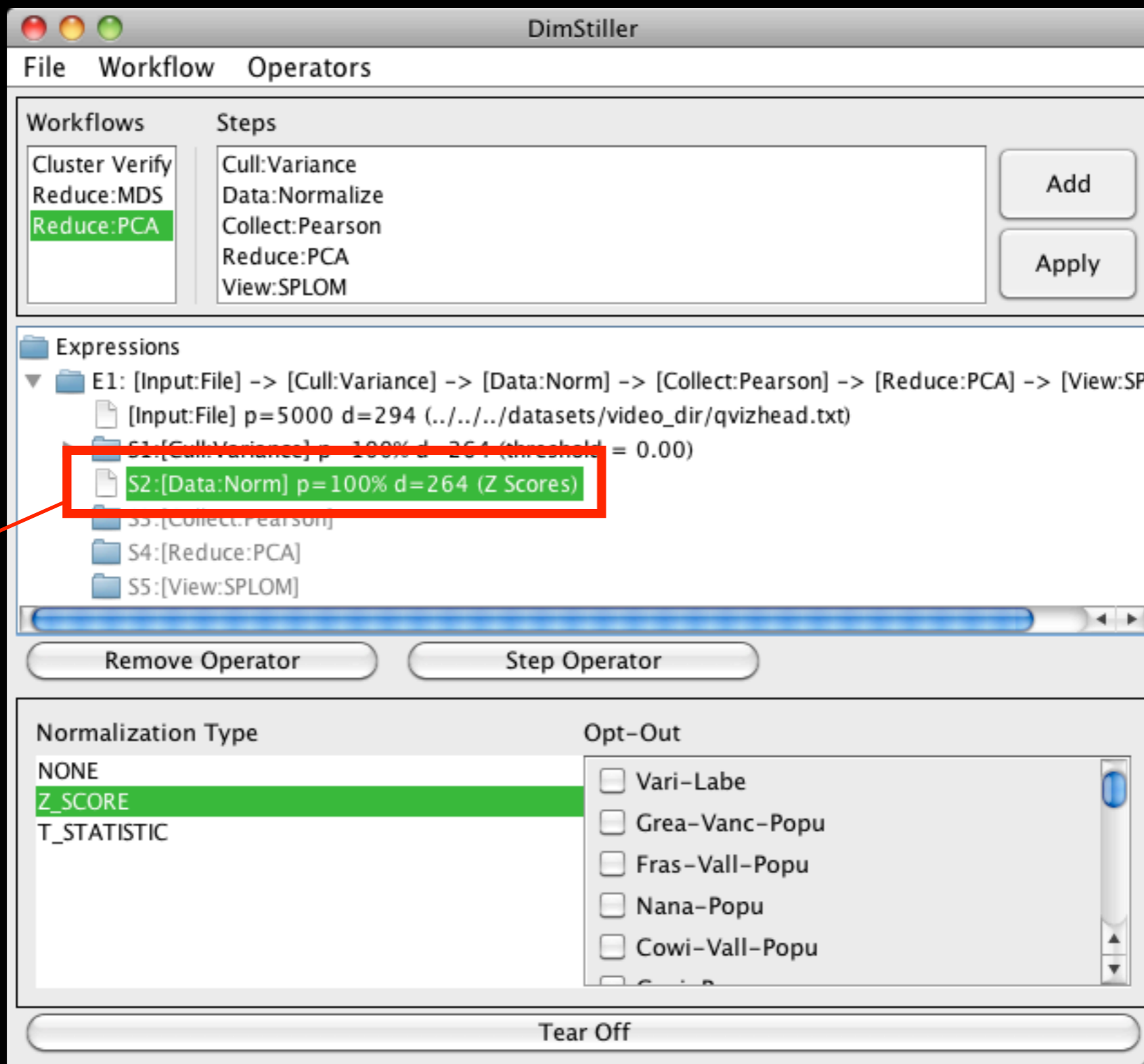
294 DIMS



List of Culled Dims

Choose first nonzero dimension (31)

264 DIMS



Data:Norm
Operator

264 DIMS

DimStiller

File Workflow Operators

Workflows

- Cluster Verify
- Reduce:MDS
- Reduce:PCA

Steps

- Cull:Variance
- Data:Normalize
- Collect:Pearson
- Reduce:PCA
- View:SPLOM

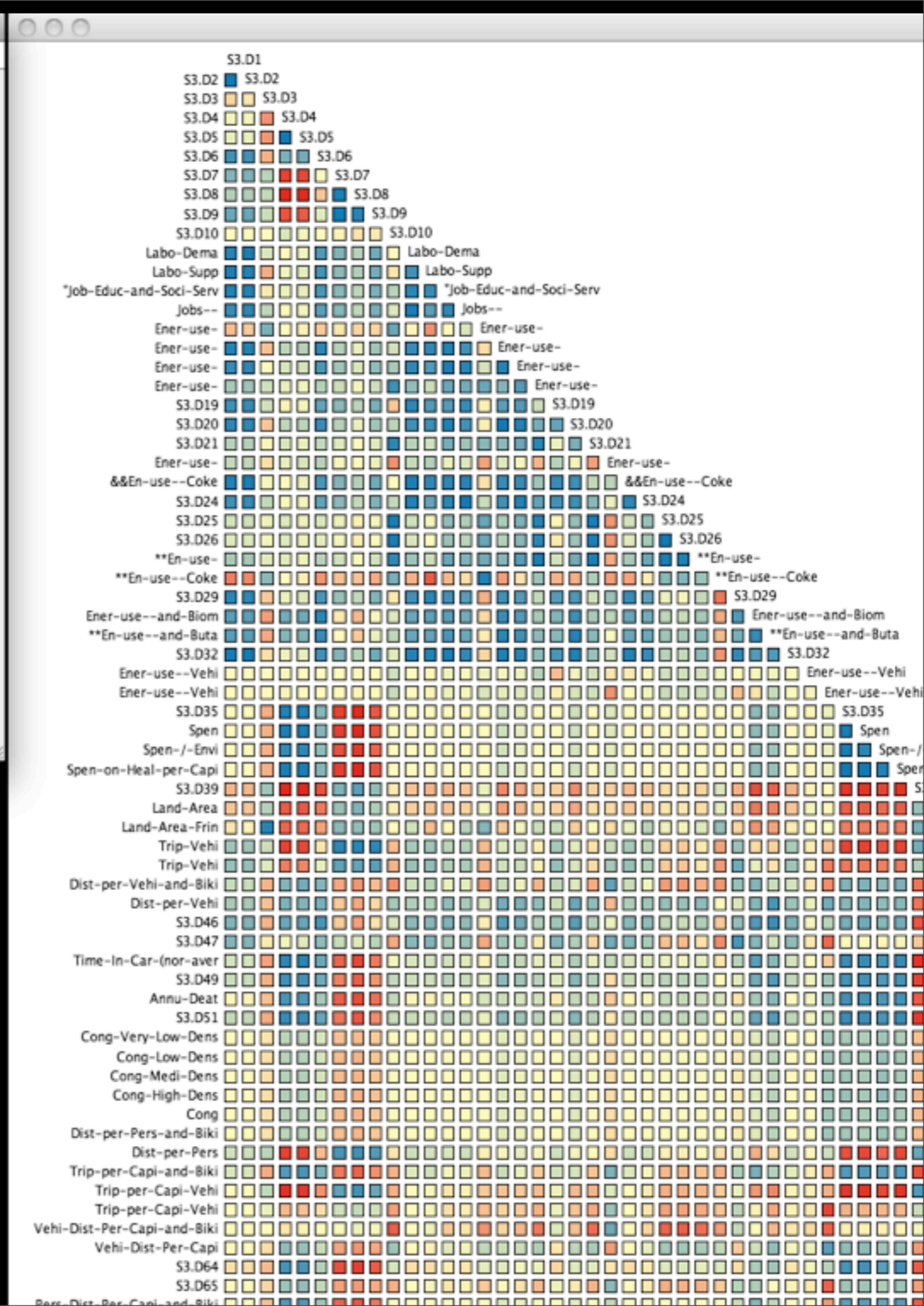
Expressions

- E1: [Input:File] -> [Cull:Variance] -> [Data:Norm] -> [Collect:Pearson] -> [Reduce:PCA] -> [View:SP]
- S1:[Cull:Variance] p=100% d=264 (threshold = 0.00)
- S2:[Data:Norm] p=100% d=264 (Z Scores)
- S3:[Collect:Pearson] p=100% d=146 (threshold = 1.00)**
- S4:[Reduce:PCA]
- S5:[View:SPLOM]

Pearson's Coefficient Cutoff

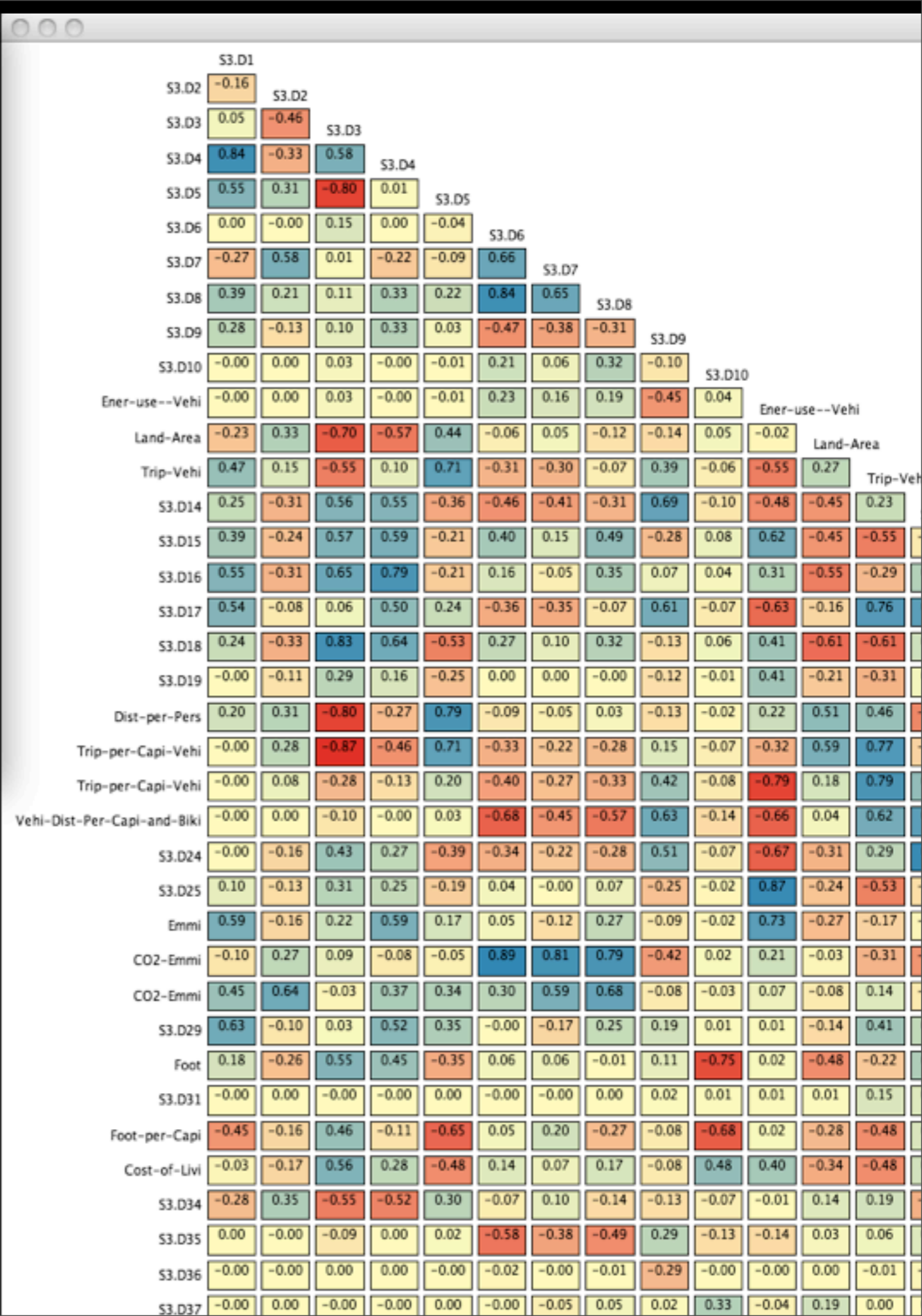
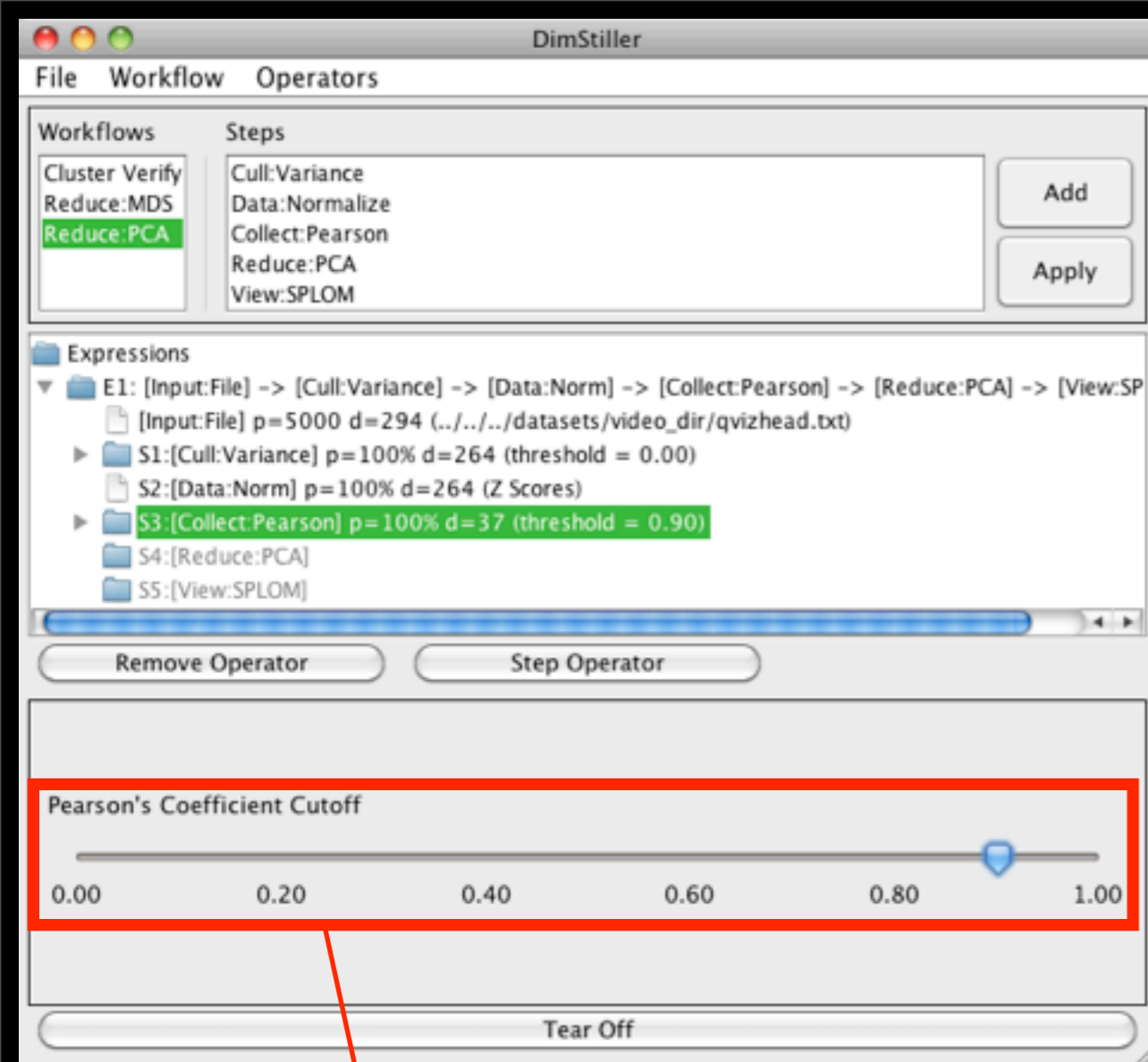
0.00 0.20 0.40 0.60 0.80 1.00

Tear Off



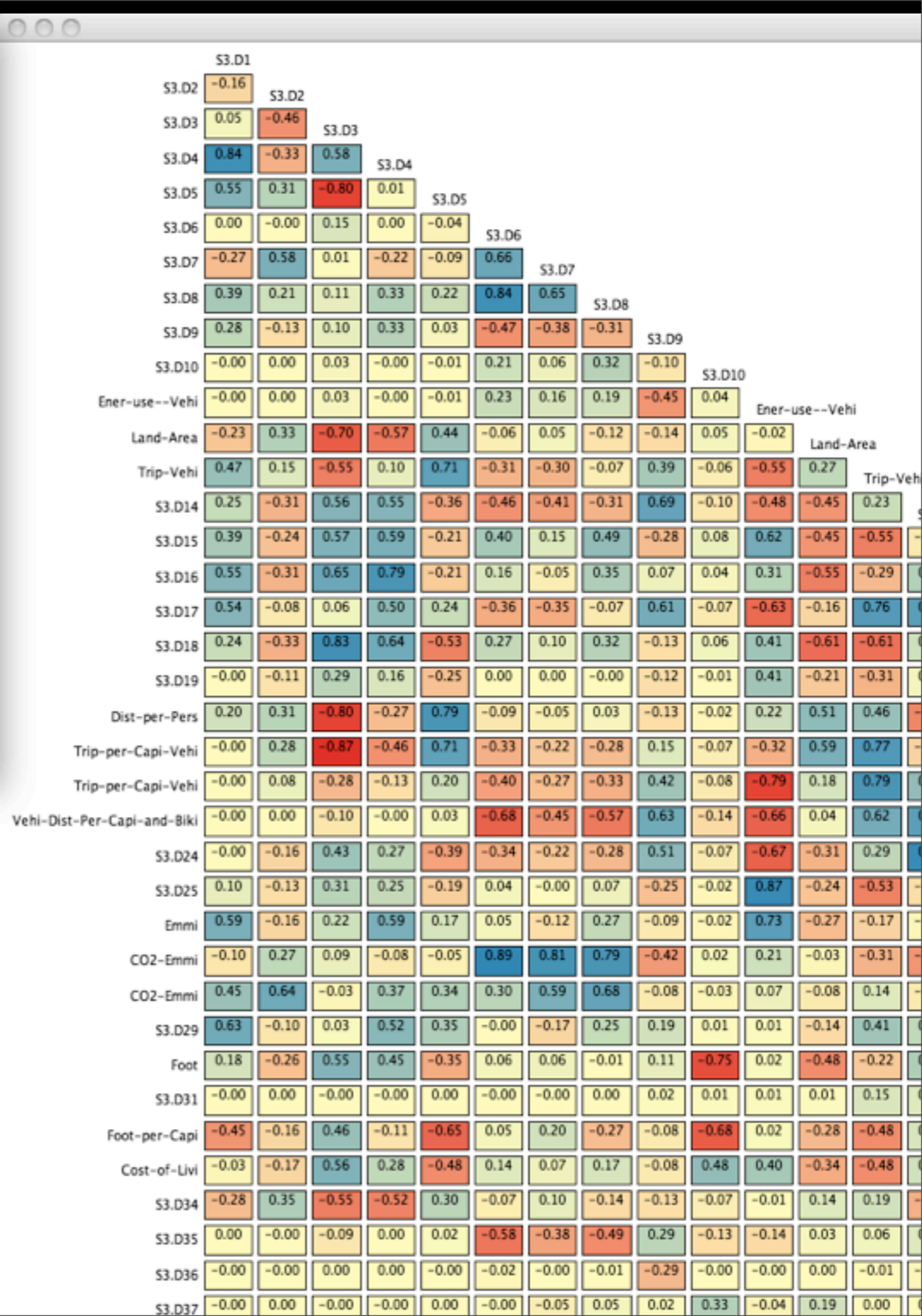
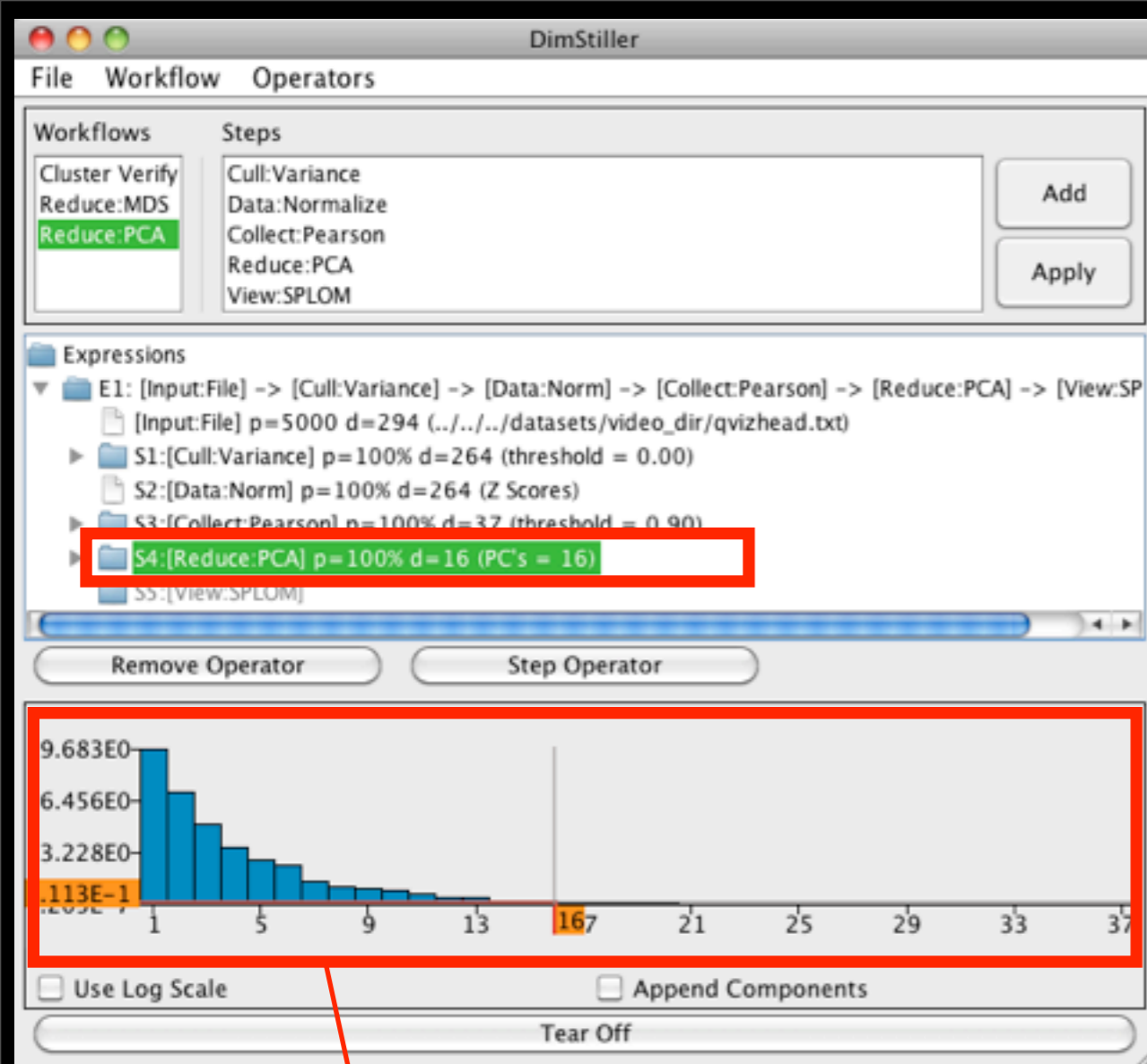
Correlation Slider
set to 1.0

146 DIMS



Correlation Slider
set to 0.9

37 DIMS



DimStiller

File Workflow Operators

Workflows

Cluster Verify
Reduce:MDS
Reduce:PCA

Steps

Cull:Variance
Data:Normalize
Collect:Pearson
Reduce:PCA
View:SPLOM

Add
Apply

Expressions

E1: [Input:File] -> [Cull:Variance] -> [Data:Norm] -> [Collect:Pearson] -> [Reduce:PCA] -> [View:SP

[Input:File] p=5000 d=294 (..././././datasets/video_dir/qvizhead.txt)

S1:[Cull:Variance] p=100% d=264 (threshold = 0.00)

S2:[Data:Norm] p=100% d=264 (Z Scores)

S3:[Collect:Pearson] p=100% d=37 (threshold = 0.90)

S4:[Reduce:PCA] p=100% d=16 (PC's = 16)

S5:[View:SPLOM] p=100% d=16

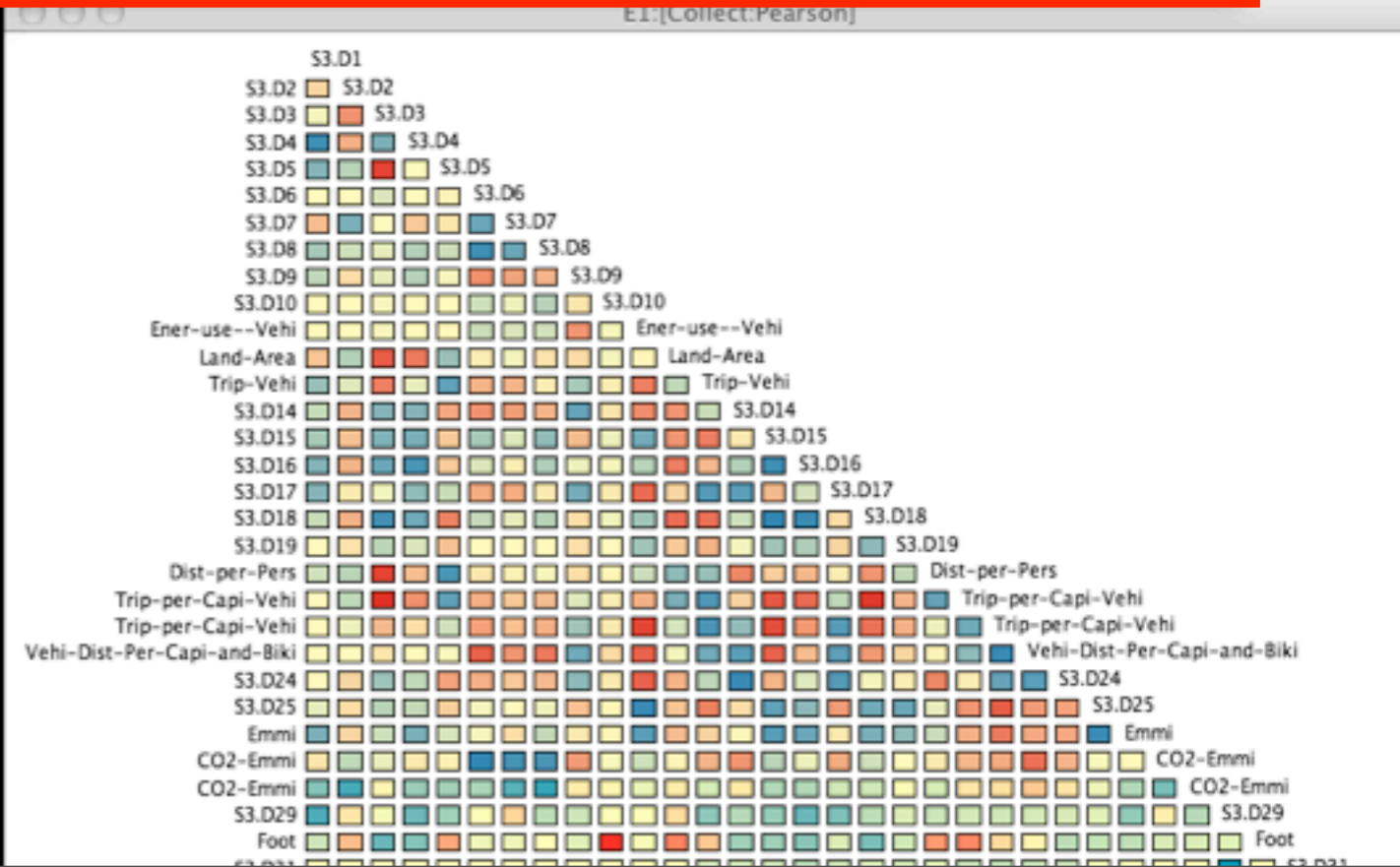
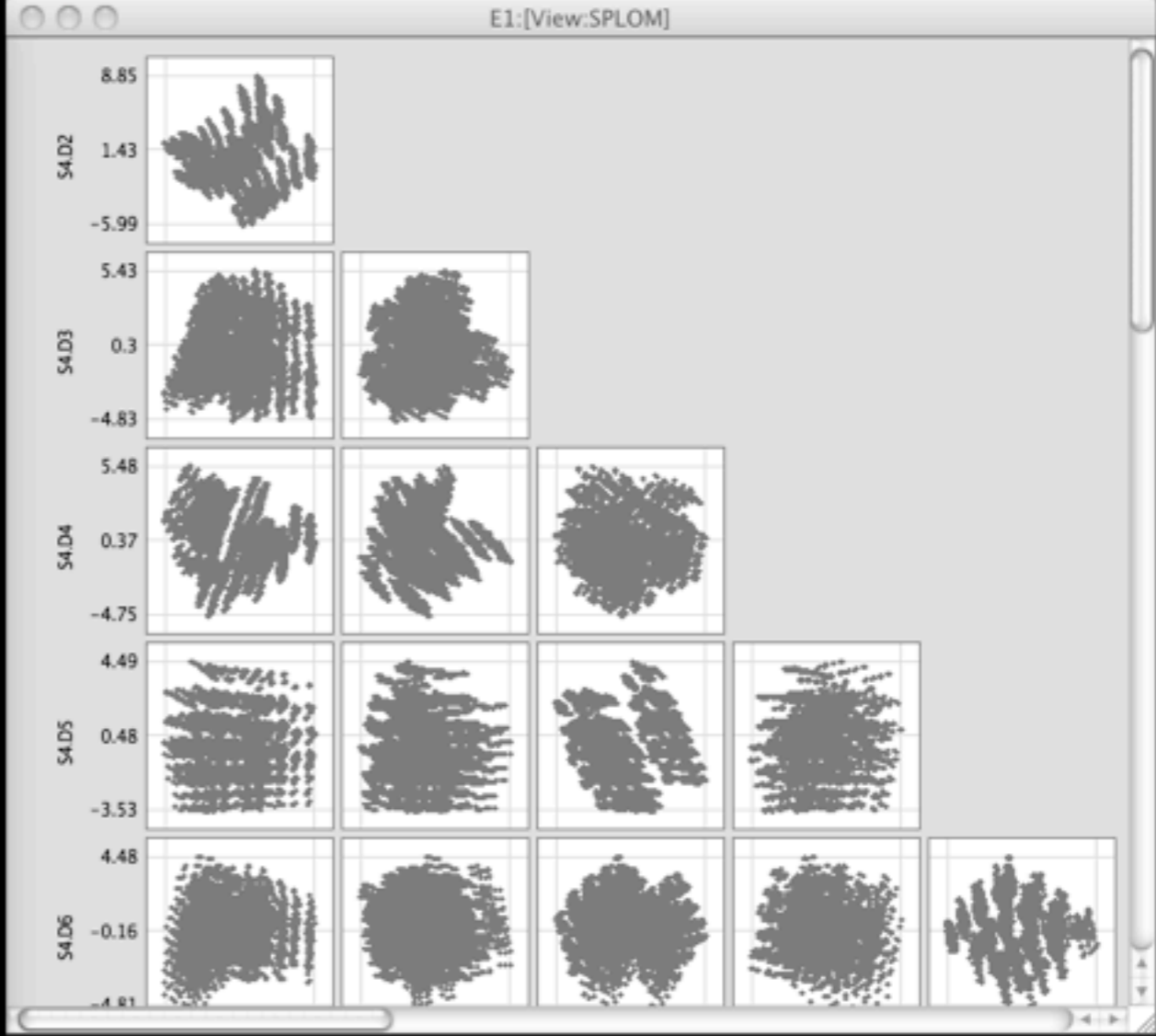
Remove Operator Step Operator

Cull Highlighted Cull NonHighlighted

Point Size 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Bubble Plot Freeze Axes

Manual Plot Size 50 100 150 200 250 300 350 400 450 500

Tear Off



Manageable SPLOM

16 DIMS

Operators & Workflows

Operator Families

Family Name	Operators
Cull	Variance, Name
Collect	Pearson's
Reduce	PCA, MDS
View	SPLOM, Histo
Attrib	Color, Cluster
Filter	Value

Custom Workflows

- Three Workflows Given
- Freeform Experimenting With Operators
- Custom Workflows after Success

Conclusions

- Presented the design and implementation of the DimStiller software
- Provided Global and Local guidance to open up dimensionality reduction for middle ground users
 - beyond experts in math AND data

Thanks!

- Download DimStiller at ...
<http://www.cs.ubc.ca/~sfingram/dimstiller>
- Doing Dim Reduction? Let me know!
sfingram@cs.ubc.ca
- Funded By NSERC

