Object Oriented Data Analysis

J. S. Marron

Dept. of Statistics and Operations Research,

University of North Carolina

January 16, 2014

An Aside on Current "Fashion"

Big Data

- Isn't It Just Statistics?

An Aside on Current "Fashion"

Big Data

- Isn't It Just Statistics?
- Yes, But We Need to Remind Folks
- Maybe Bigger Challenge:

Complex Data

Object Oriented Data Analysis

What is the "atom" of a statistical analysis?

- 1st Course: Numbers
- Multivariate Analysis Course : Vectors
- Functional Data Analysis: Curves
- More generally: Data Objects

Object Oriented Data Analysis

Original Thought:

OODA = Mathematical Framework

(containing wide variety
of interesting cases)

Object Oriented Data Analysis

Original Thought:

OODA = Mathematical Framework

Current View:

OODA = Focal Point

Object Oriented Data Analysis

Original Thought:

OODA = Mathematical Framework

Current View:

OODA = Focal Point

{For discussions (interdisciplinary) about tackling serious analyses}

Object Oriented Data Analysis

Original Thought:

OODA = Mathematical Framework

Current View:

OODA = Focal Point

What should be the Data Objects?

Functional Data Analysis

Curves as Data Objects

Important Duality:

Curve Space ←→ Point Cloud Space

Illustrate with Travis Gaydos Graphics

- 2 dim'al curves (easy to visualize)

Functional Data Analysis

UNC, Stat & OR

Curves as Data Objects

Important Duality Concept:

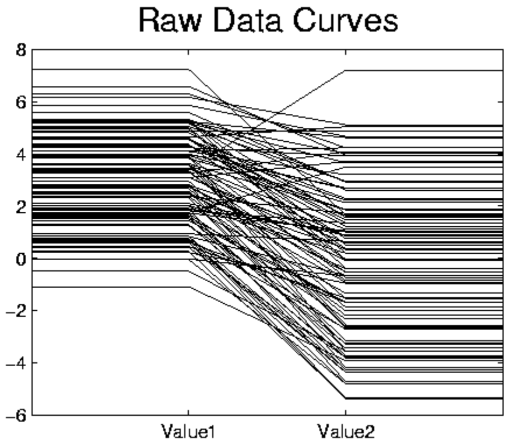
Curve Space ←→ Point Cloud Space

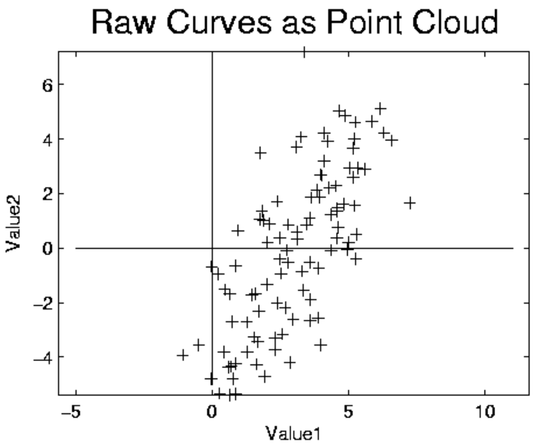
(= Object Space) (= Feature Space)

Illustrate with Travis Gaydos Graphics

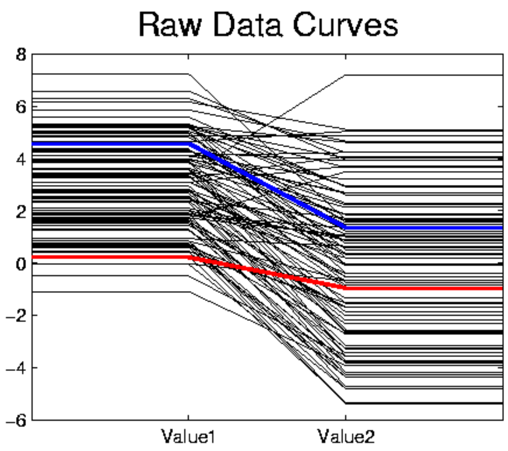
- 2 dim'al curves (easy to visualize)

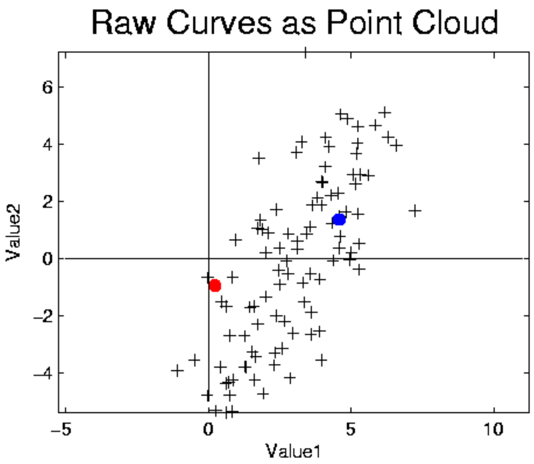
Functional Data Analysis, Toy EG I



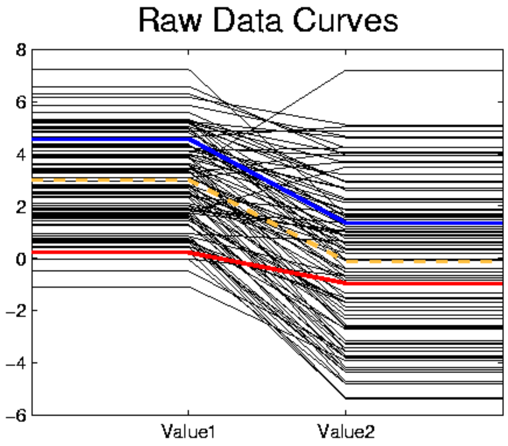


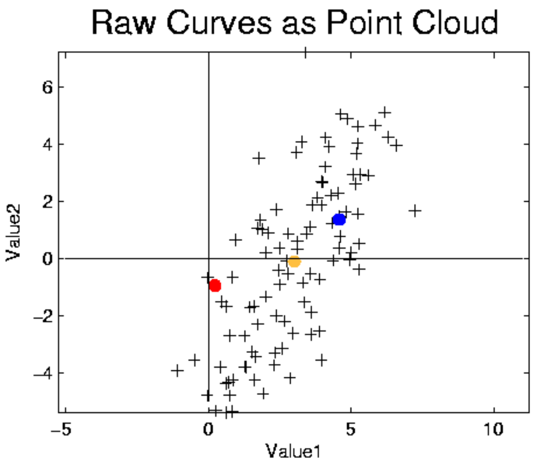
Functional Data Analysis, Toy EG II



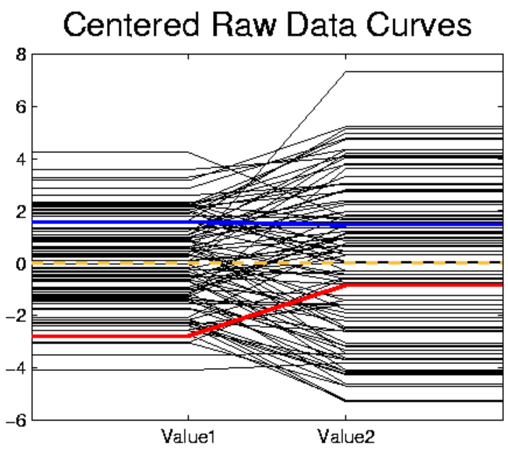


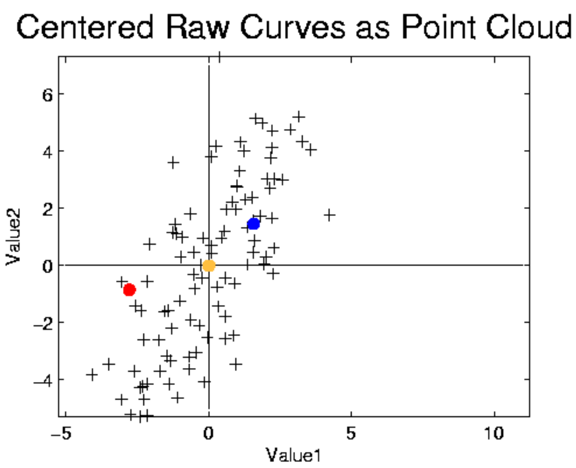
Functional Data Analysis, Toy EG III



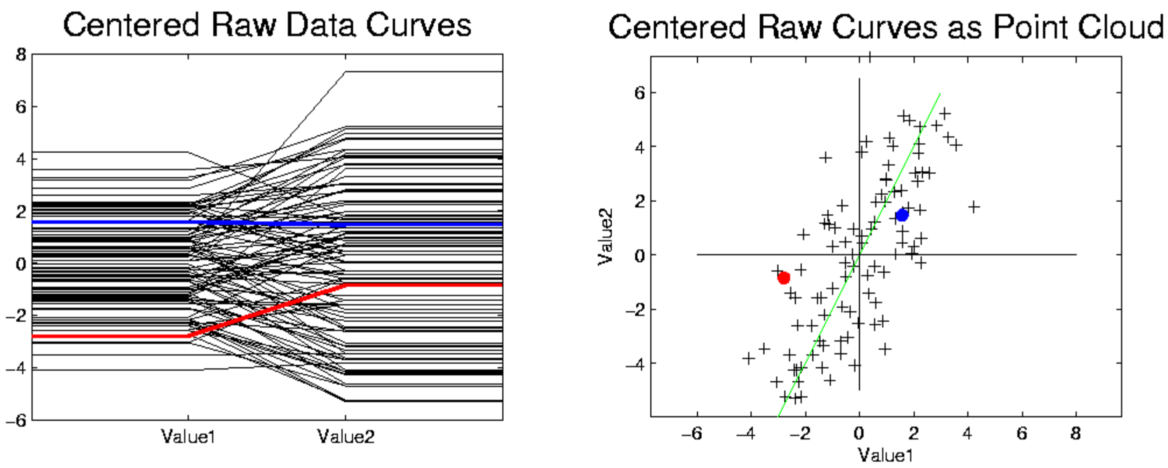


Functional Data Analysis, Toy EG IV

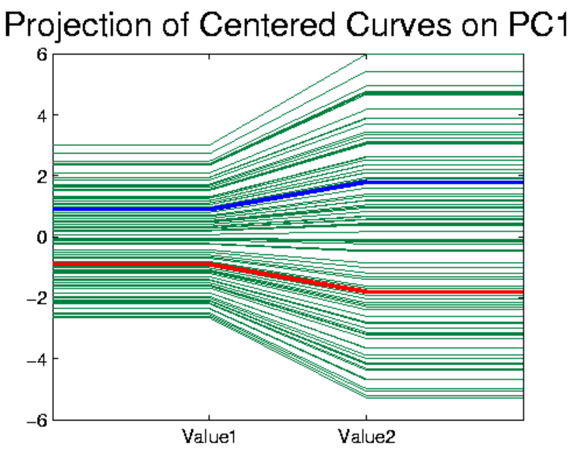


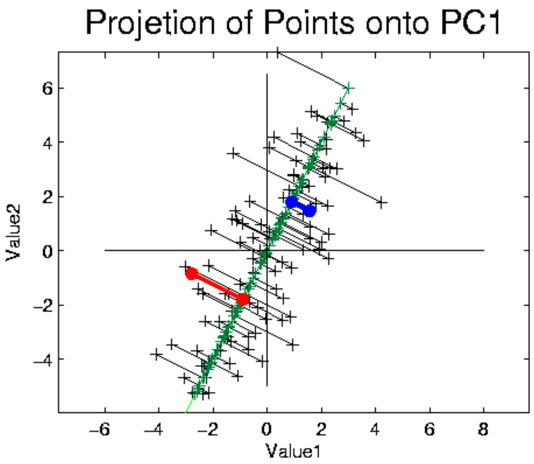


Functional Data Analysis, Toy EG V

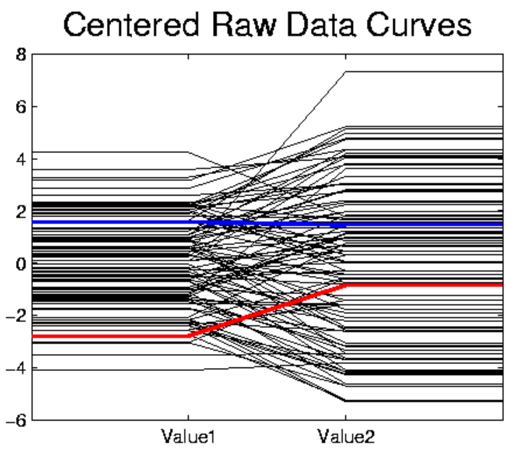


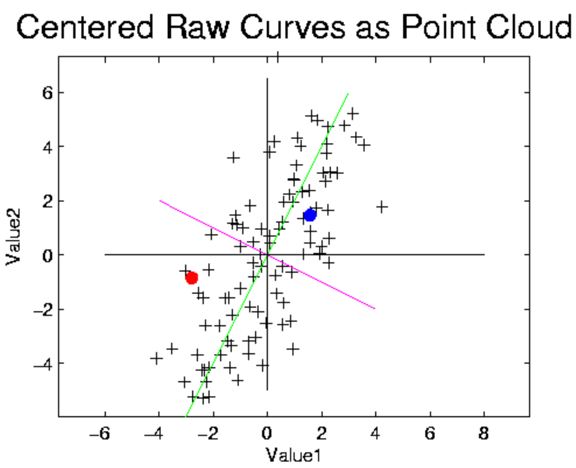
Functional Data Analysis, Toy EG VI





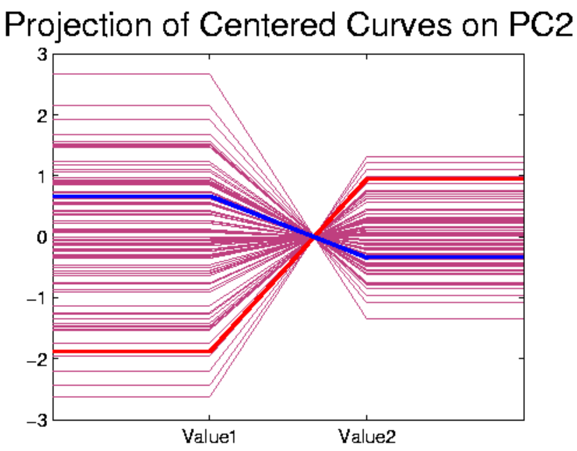
Functional Data Analysis, Toy EG VII

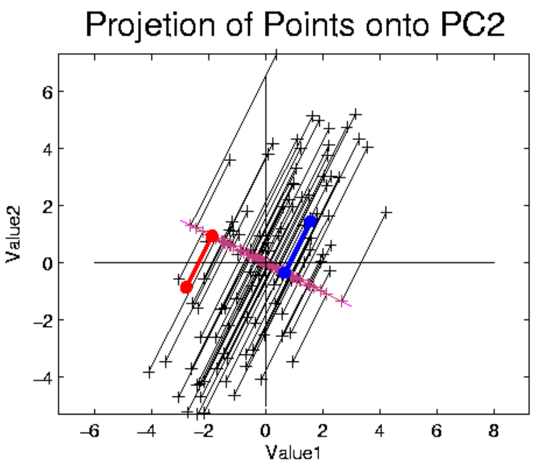




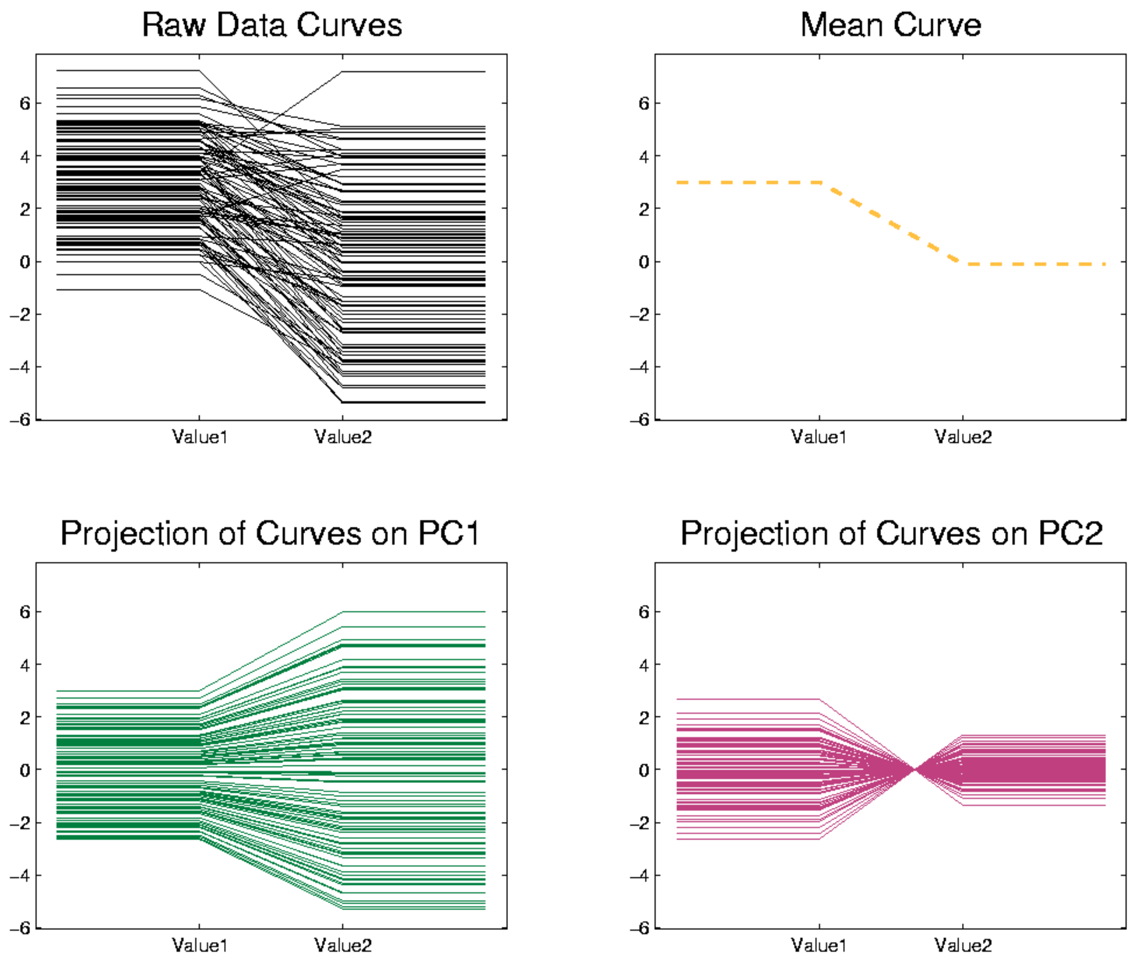
Functional Data Analysis, Toy EG VIII

UNC, Stat & OR

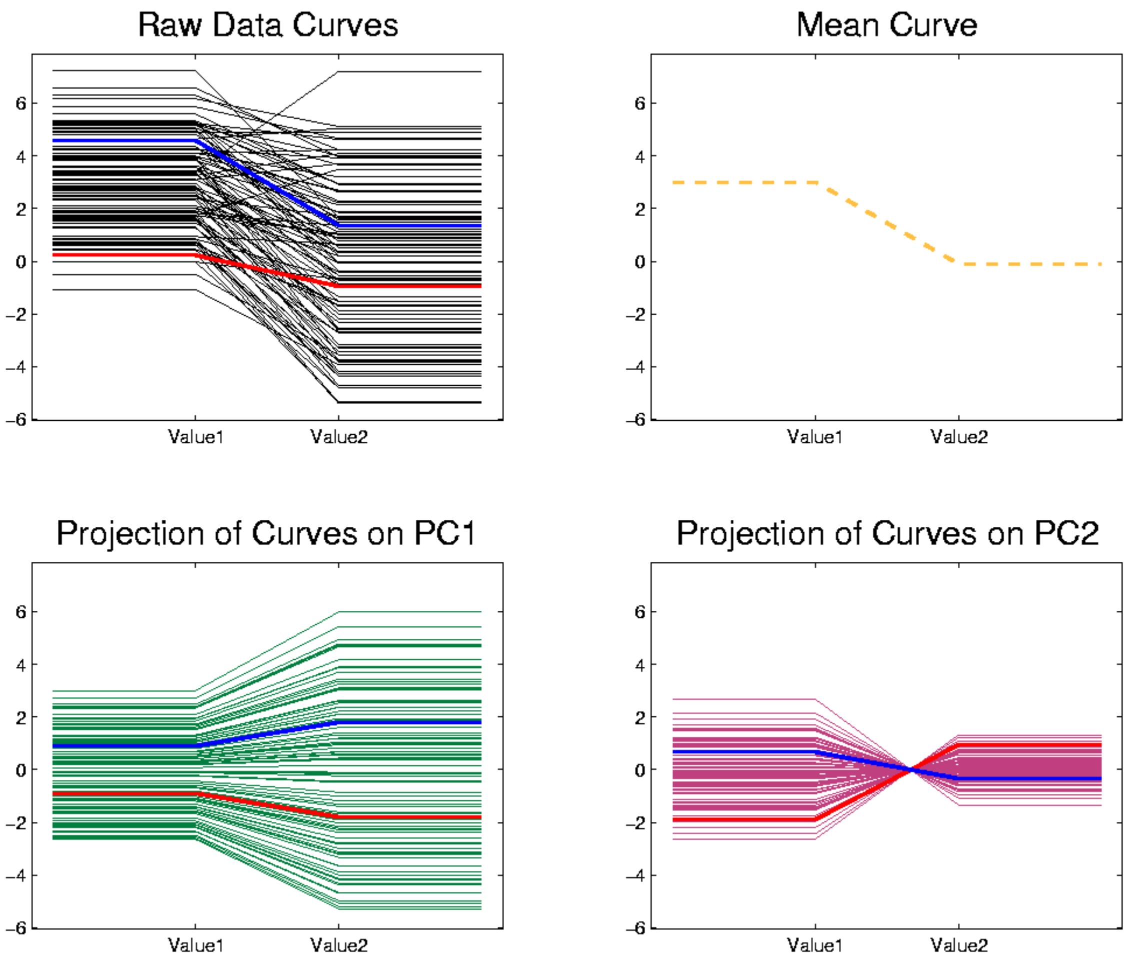




Functional Data Analysis, Toy EG IX

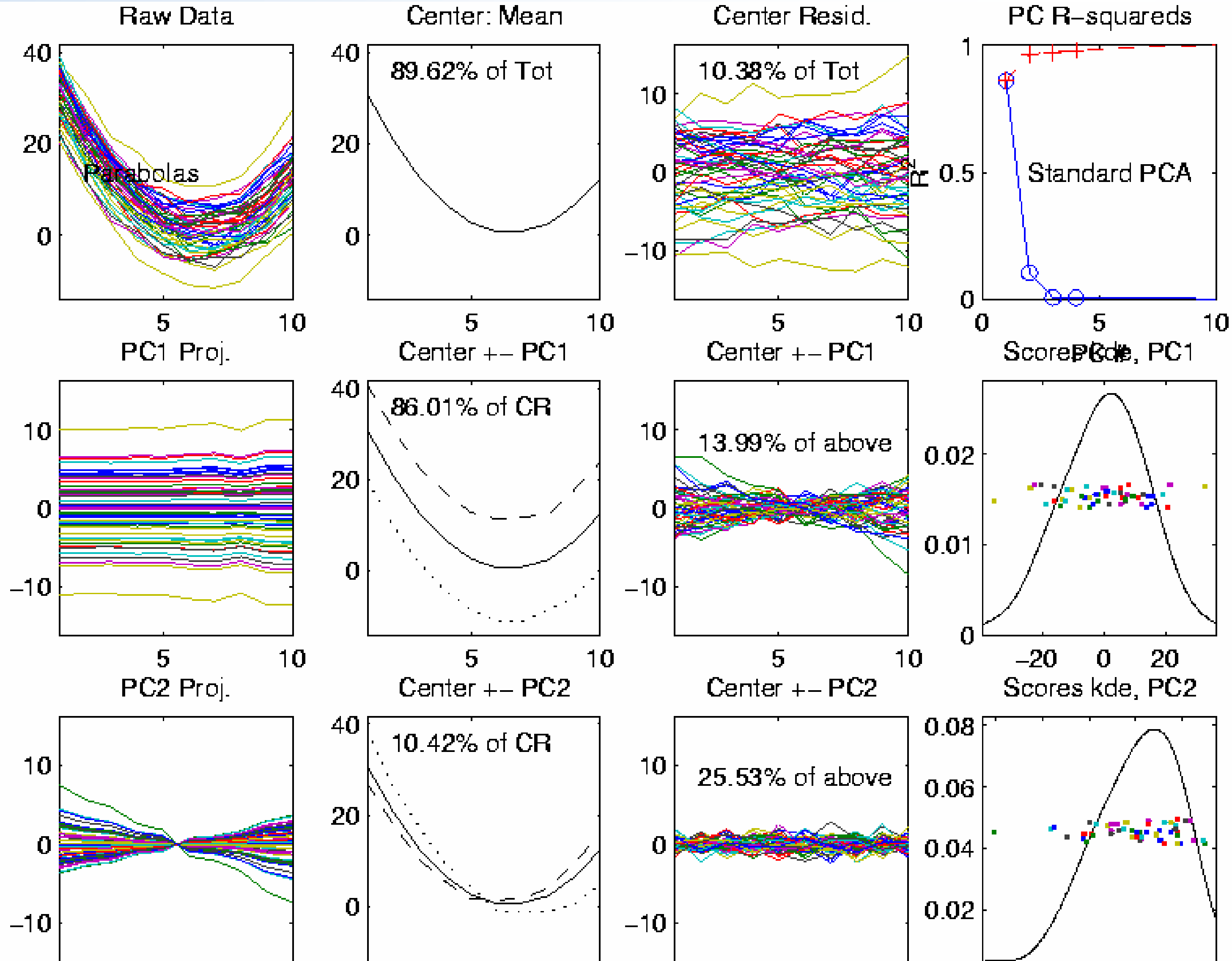


Functional Data Analysis, Toy EG X



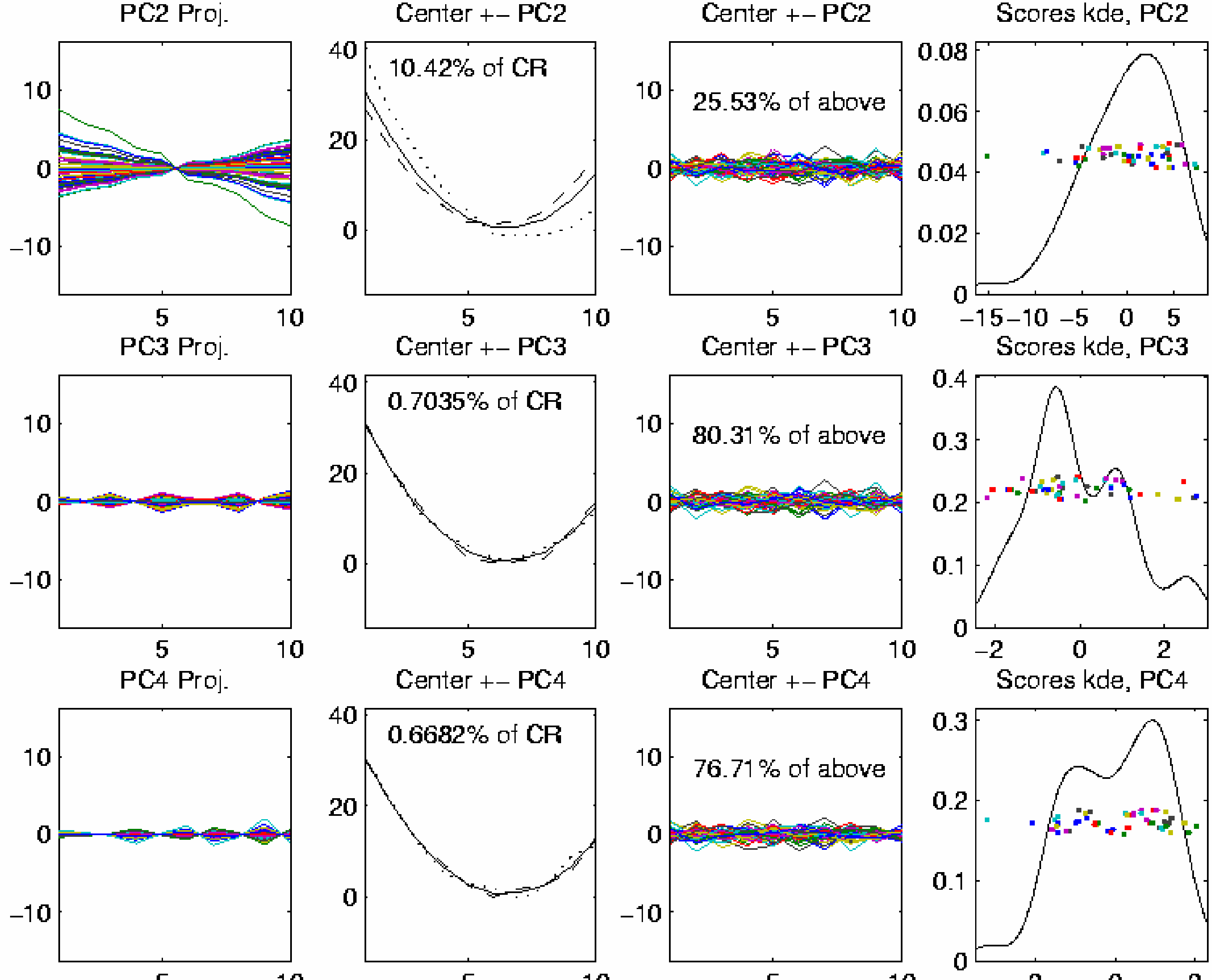
Functional Data Analysis, 10-d Toy EG 1

UNC, Stat & OR



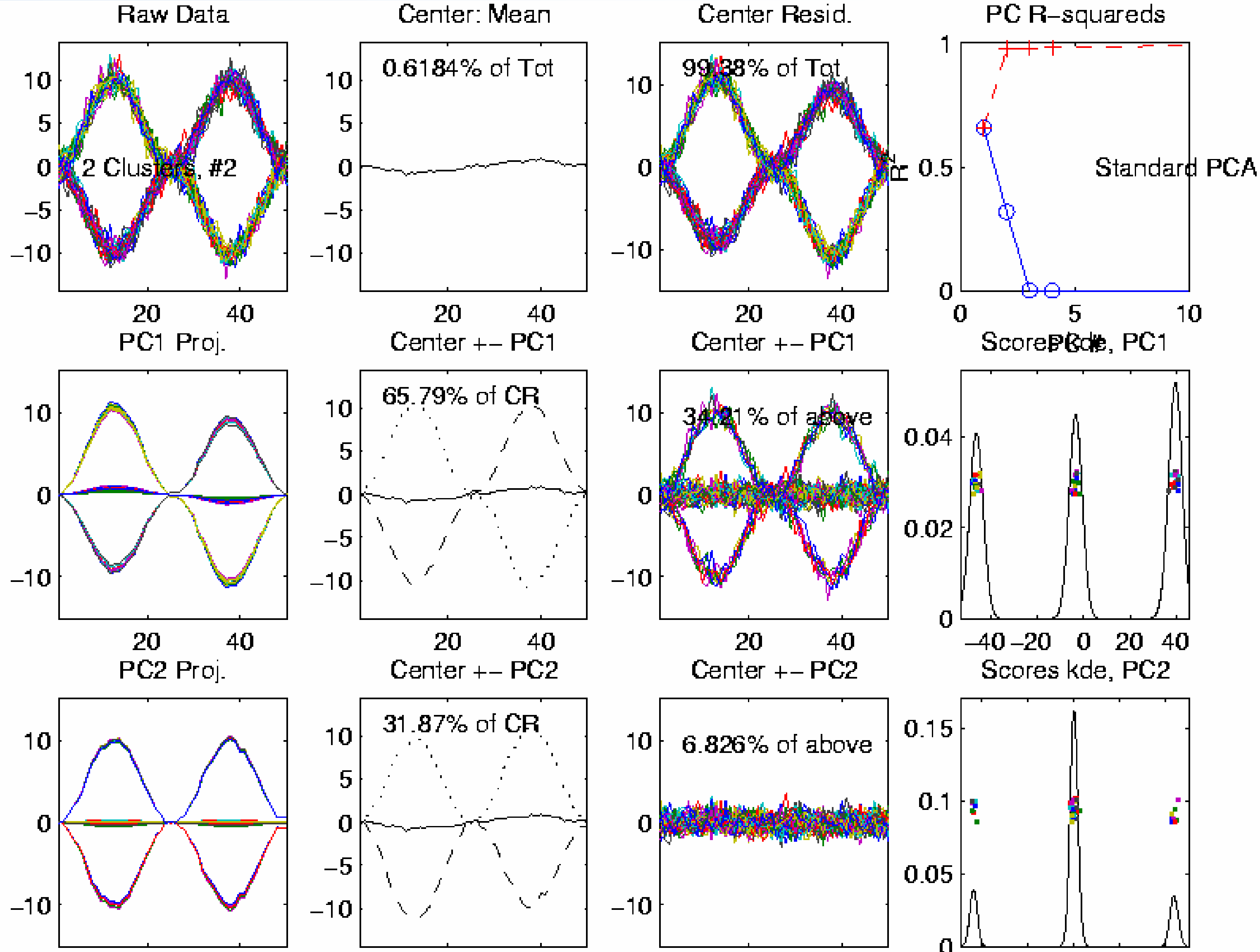
Functional Data Analysis, 10-d Toy EG 1

UNC, Stat & OR



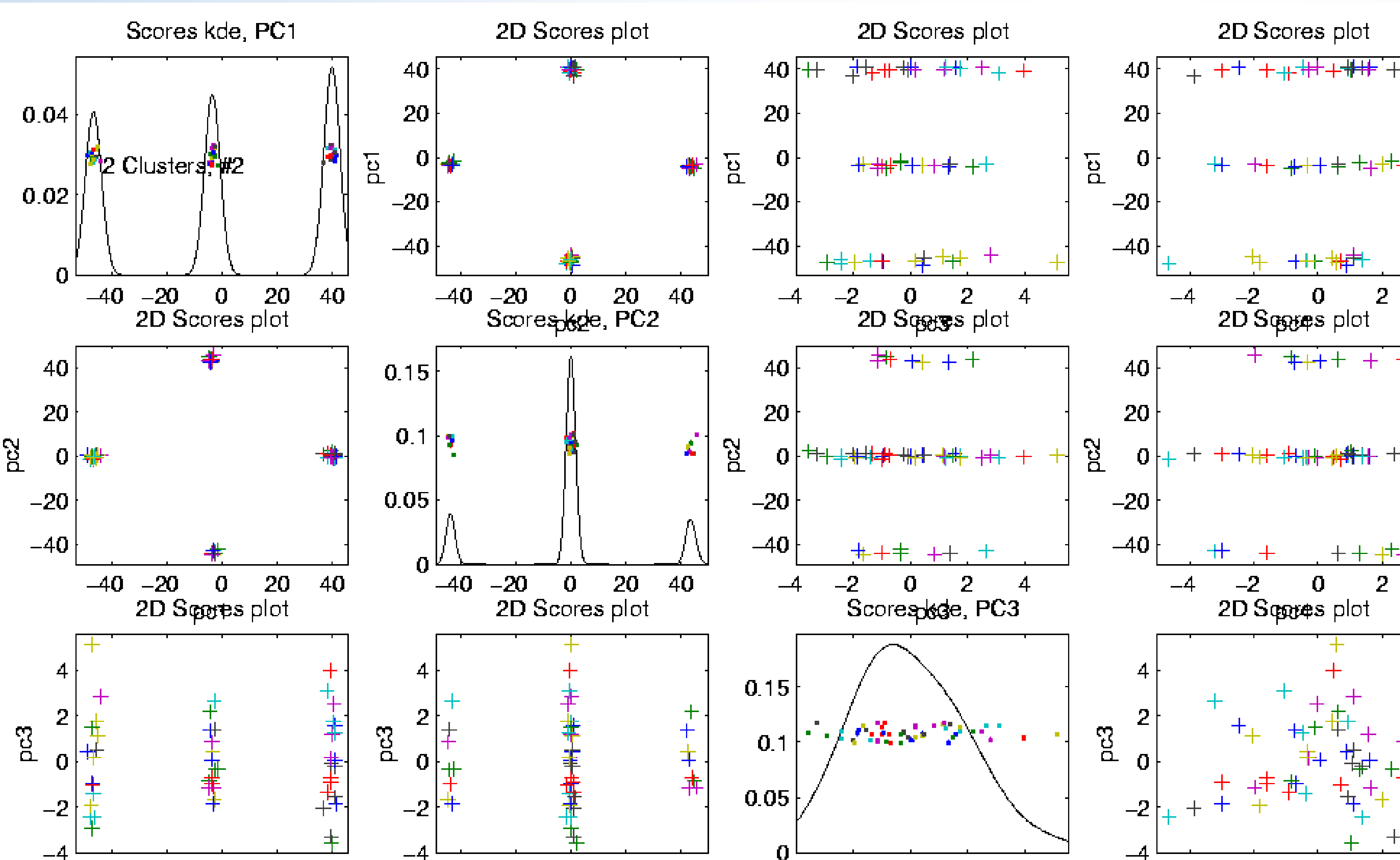
Functional Data Analysis, 50-d Toy EG 2

UNC, Stat & OR



Functional Data Analysis, 50-d Toy EG 2

UNC, Stat & OR



Principal Component Analysis

More Than *Dimensionality Reduction*

UNC, Stat & OR

Principal Component Analysis

More Than *Dimensionality Reduction*:

- Visualization
  - Relationships Between Objects (Scores)
  - Drivers of Relationships (Loadings)

Principal Component Analysis

More Than *Dimensionality Reduction*:

- <u>Visualization</u>
  - Relationships Between Objects (Scores)
  - Drivers of Relationships (Loadings)
- <u>Summarization</u>
  - Lower-d Representation
  - E.g. $n << d$

Principal Component Analysis

UNC, Stat & OR

More Than *Dimensionality Reduction*:

- Visualization
  - Relationships Between Objects (Scores)
  - Drivers of Relationships (Loadings)
- Summarization
  - Lower-d Representation
  - E.g. $n << d$
- Careful about *Information Loss*

Functional Data Analysis

Interesting Data Set:

- Mortality Data
- For Spanish Males (thus can relate to history)
- Each curve is a single year
- x coordinate is age
- Mortality = # died / total # (for each age)
- Study on log scale

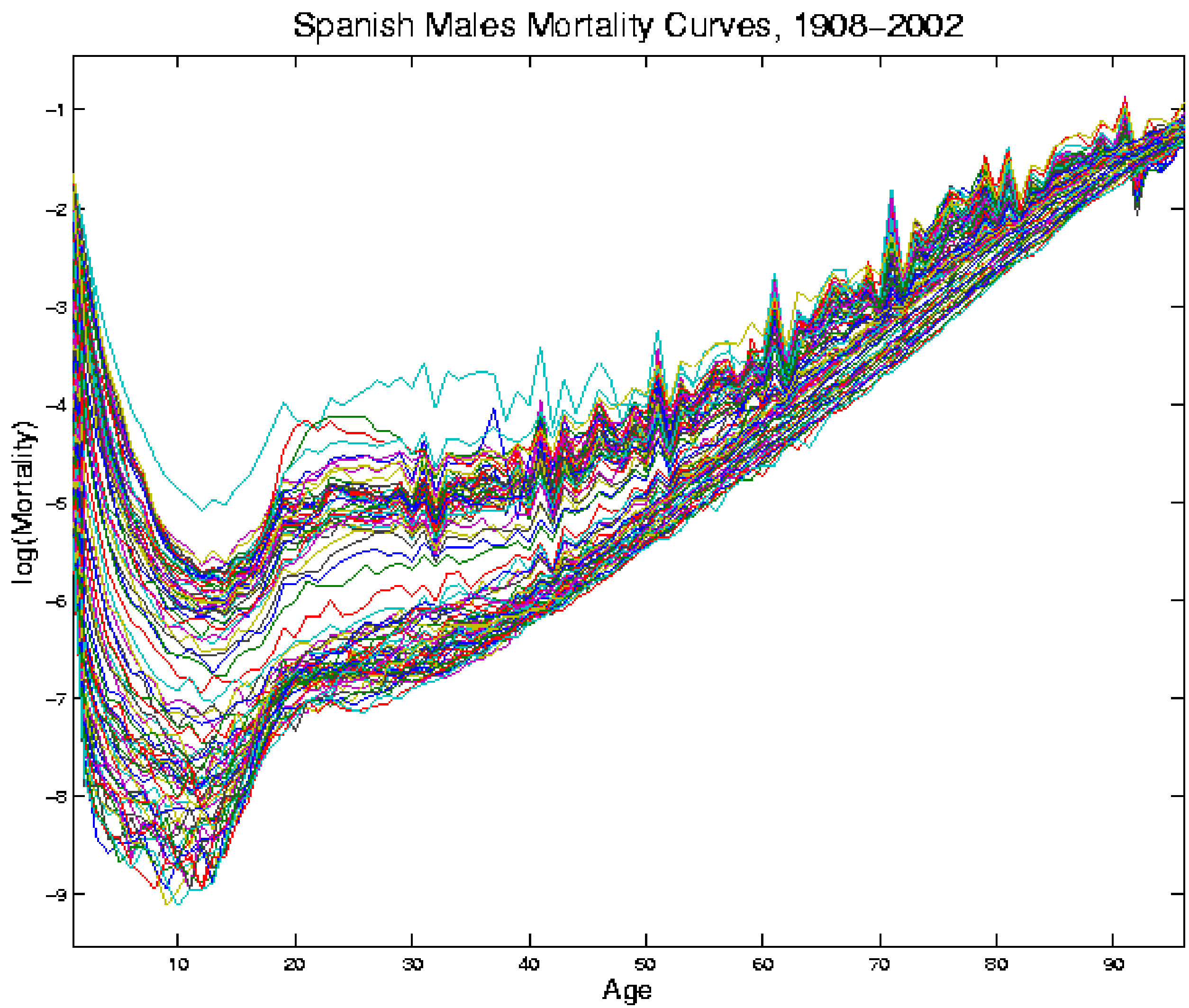
Another *Data Object* Choice

Mortality Time Series

Conventional Coloring:

Rotate Through (7) Colors

Hard to See Time Structure



Mortality Time Series

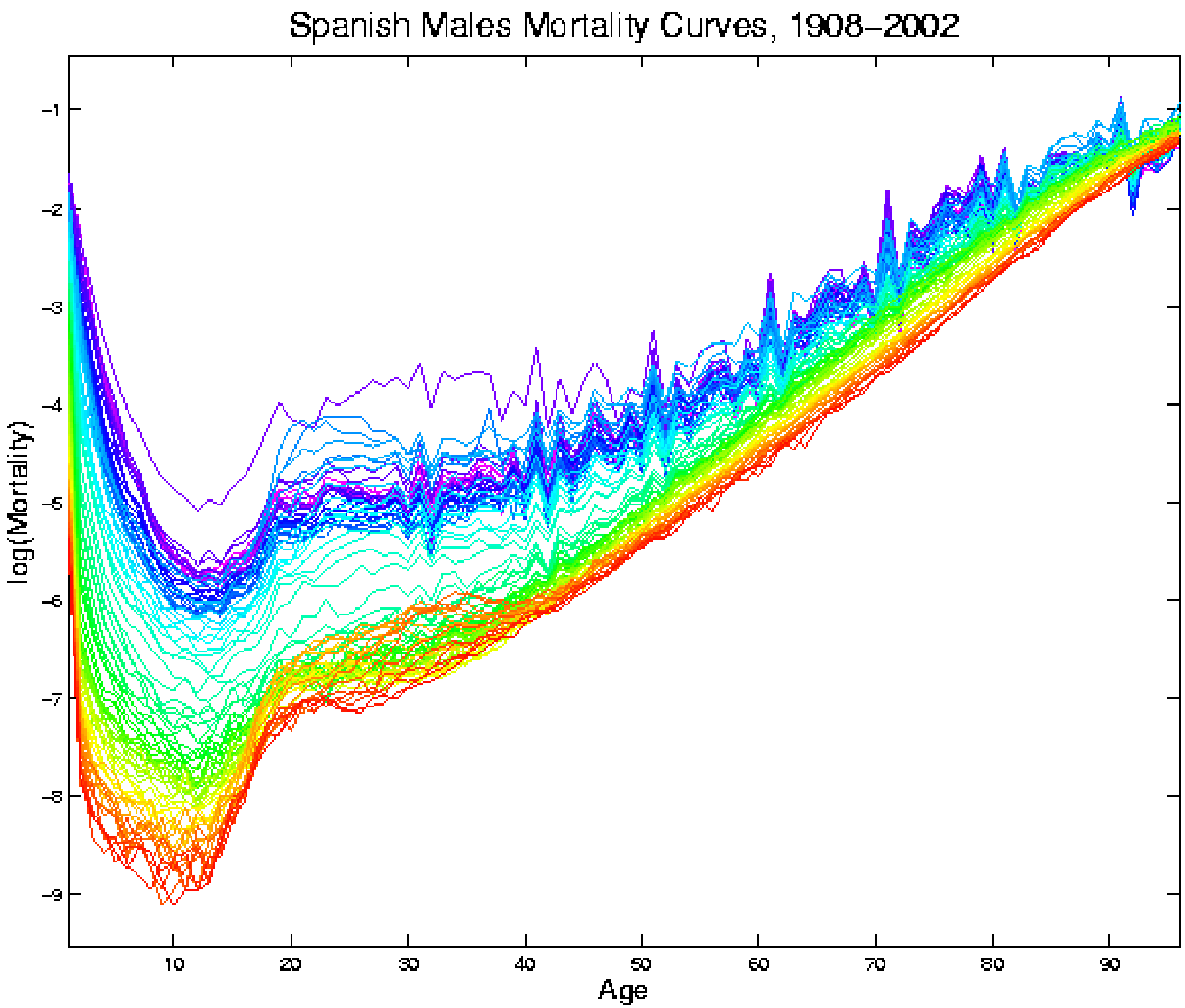
UNC, Stat & OR

Improved Coloring:

Rainbow Representing Year:

Magenta = 1908

Red = 2002



Mortality Time Series

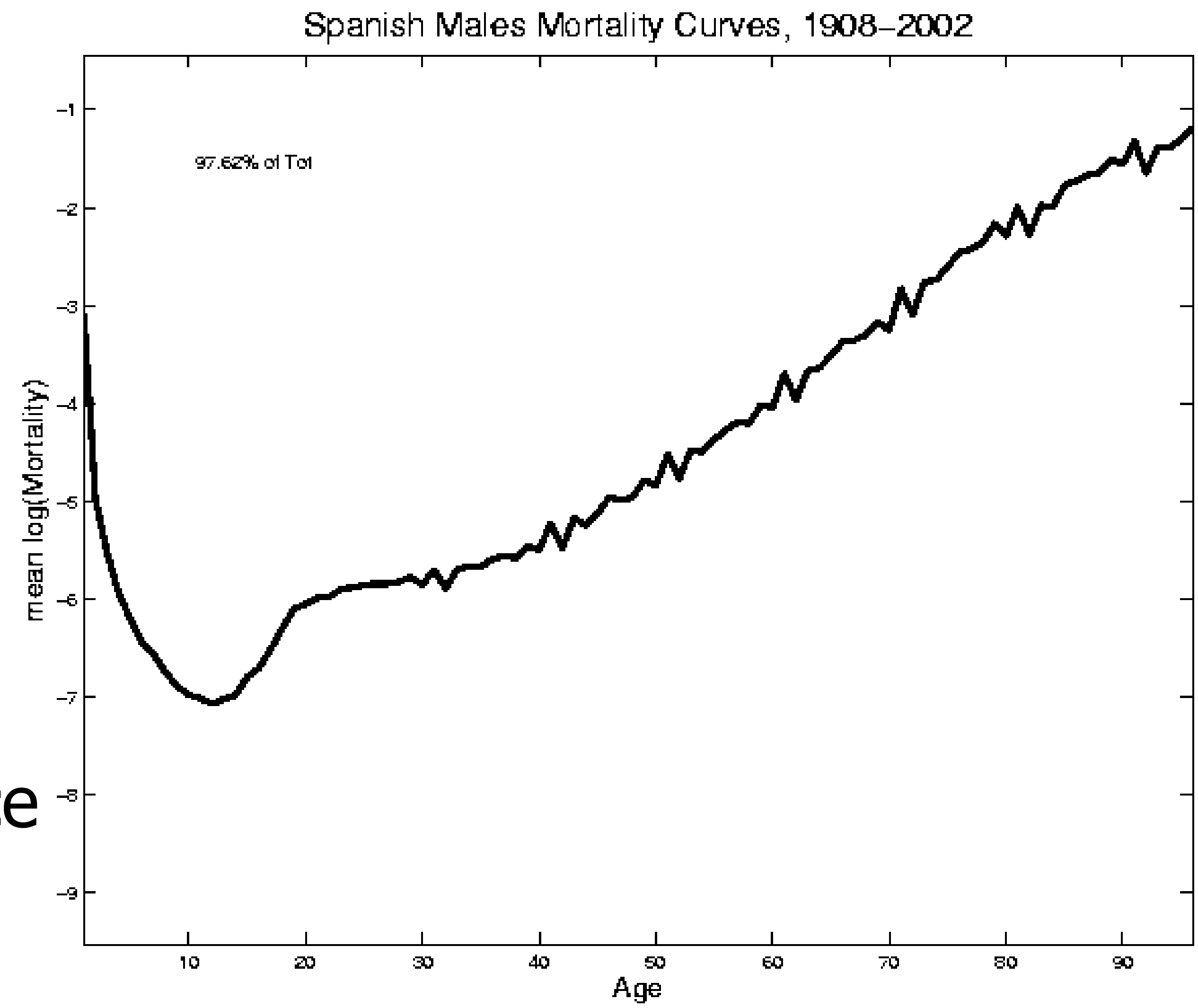
UNC, Stat & OR

Find Population Center

(Mean Vector)

Compute in Feature Space

Show in Object Space



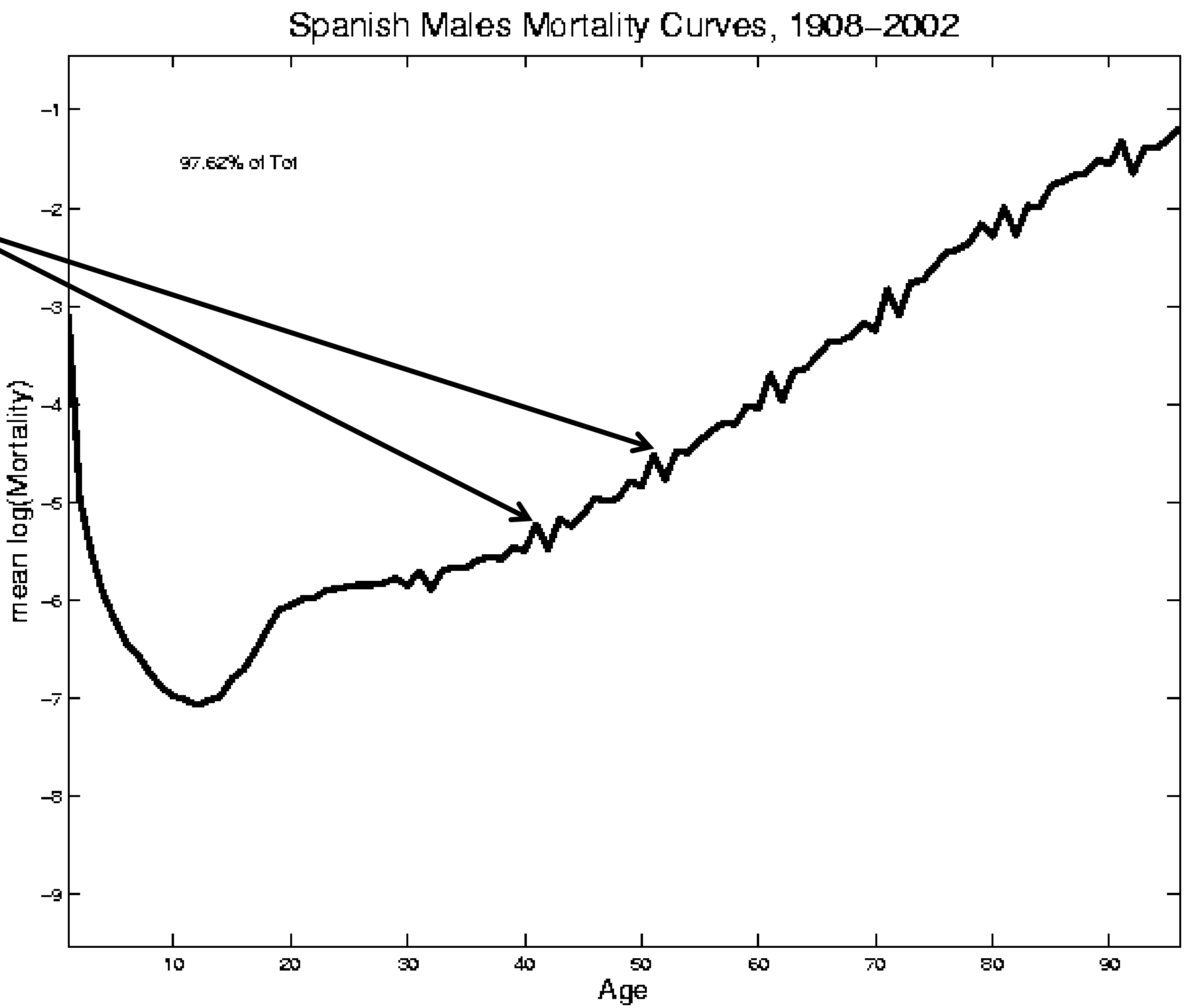
Mortality Time Series

UNC, Stat & OR

Blips Appear
At Decades

Since Ages
Not Precise
(in Spain)

Reported as
"about 50",
Etc.

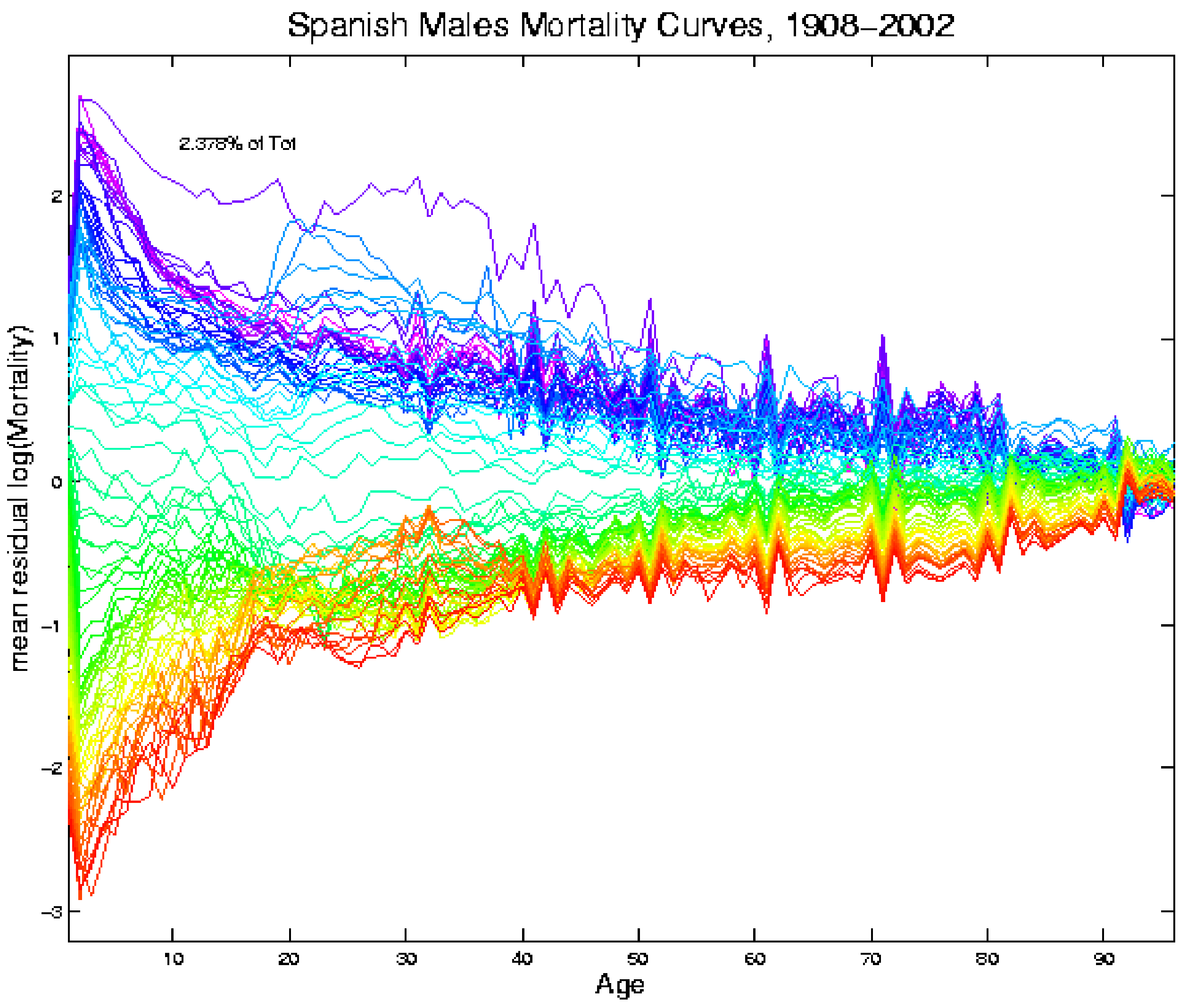


Mortality Time Series

UNC, Stat & OR

Mean Residual

Object Space View of Shifting Data To Origin In Feature Space

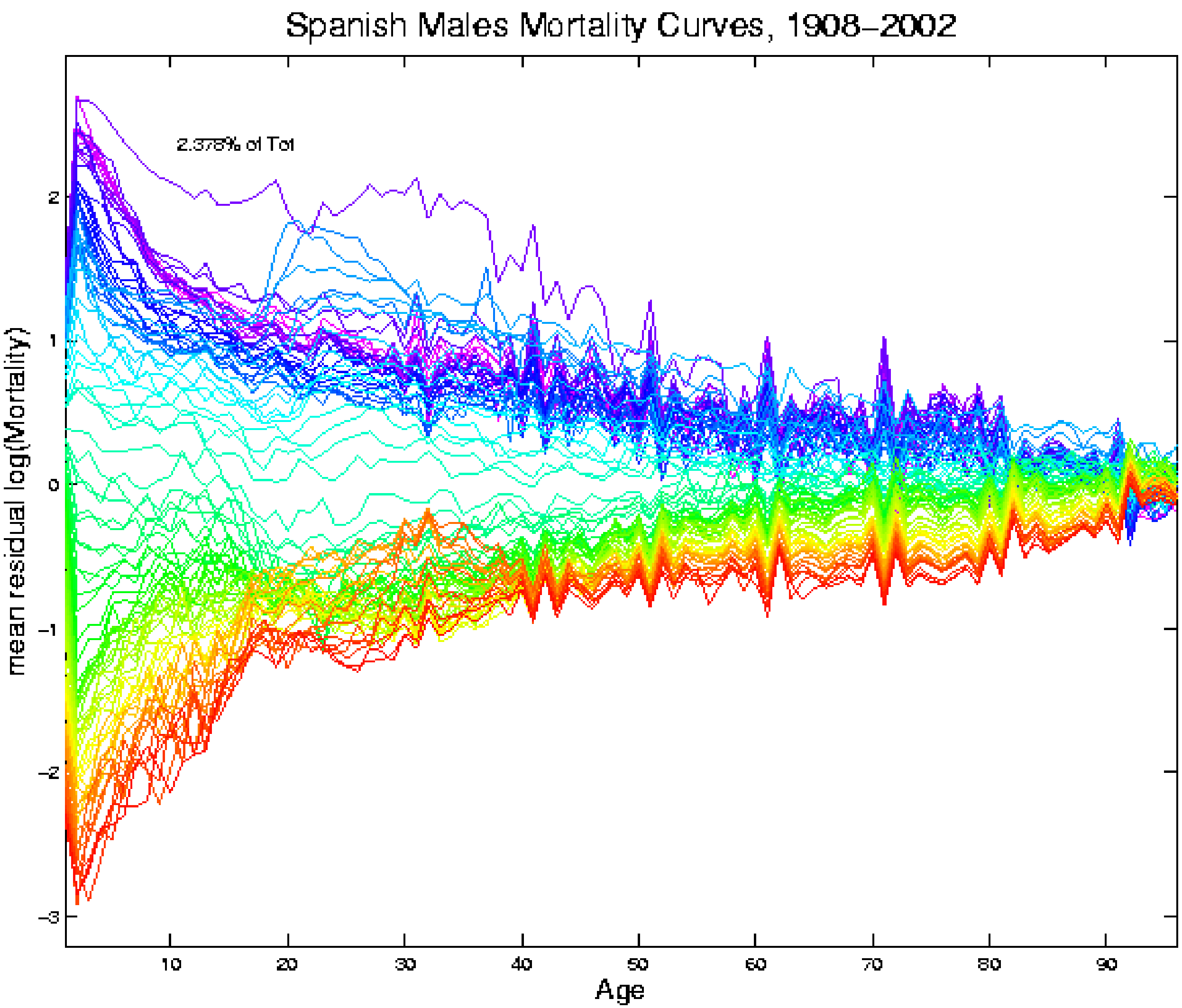


Mortality Time Series

UNC, Stat & OR

Shows:

Main Age Effects in Mean, Not Variation About Mean

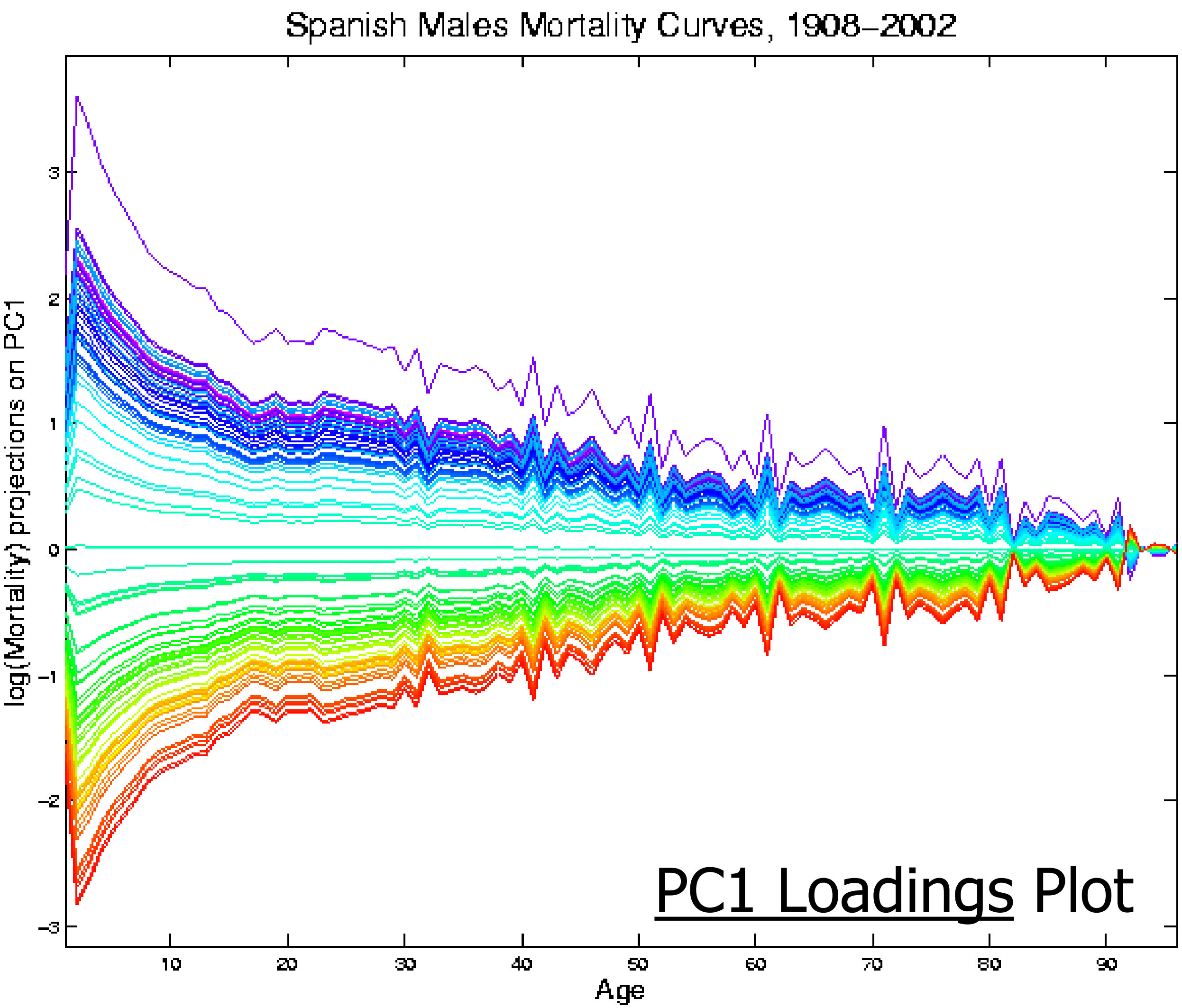


Mortality Time Series

UNC, Stat & OR

Object Space View of Projections Onto PC1 Direction

Main Mode Of Variation: Constant Across Ages



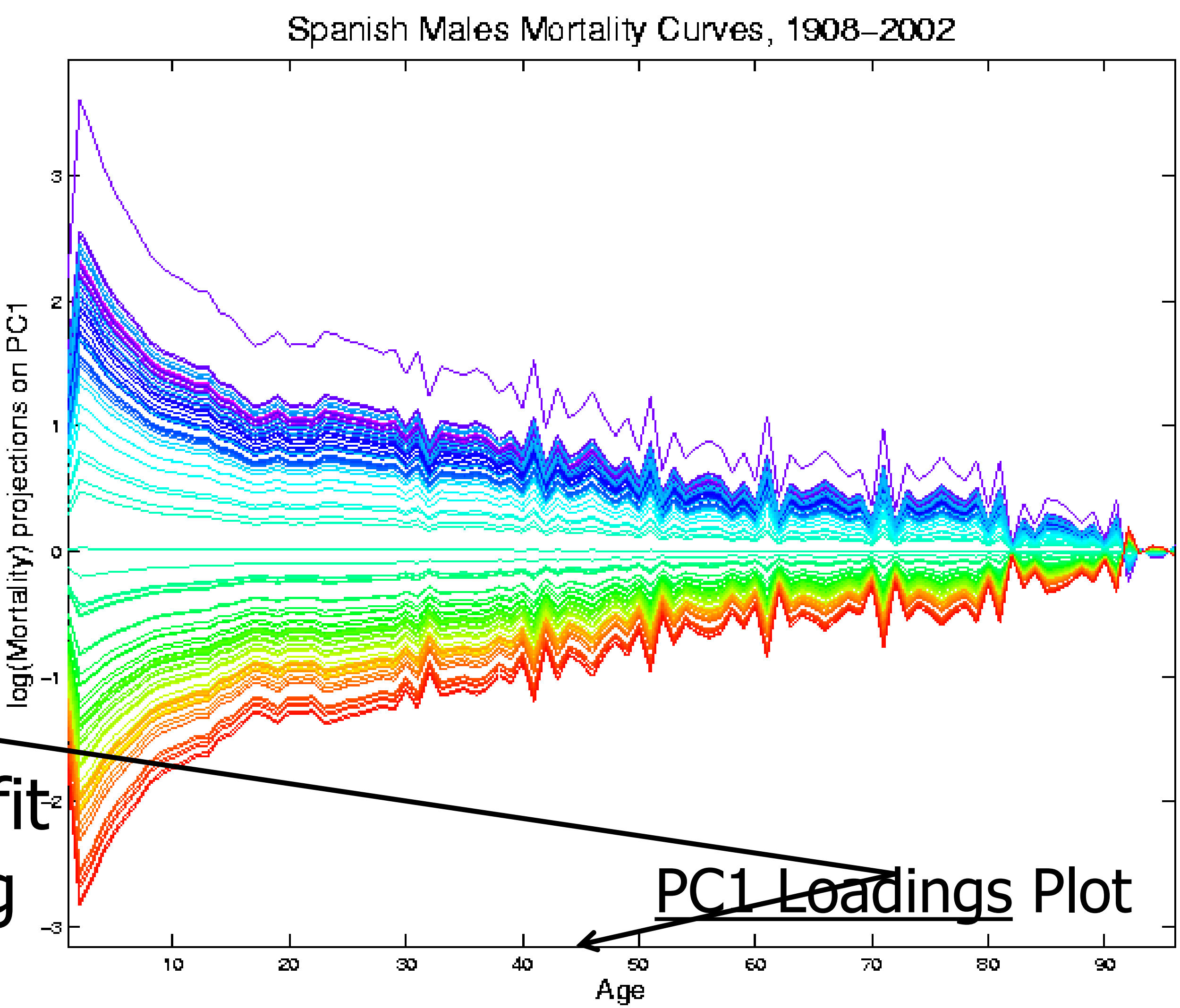
Mortality Time Series

UNC, Stat & OR

Shows *Major* Improvement Over Time

(Pub. Health, Medicine,…)

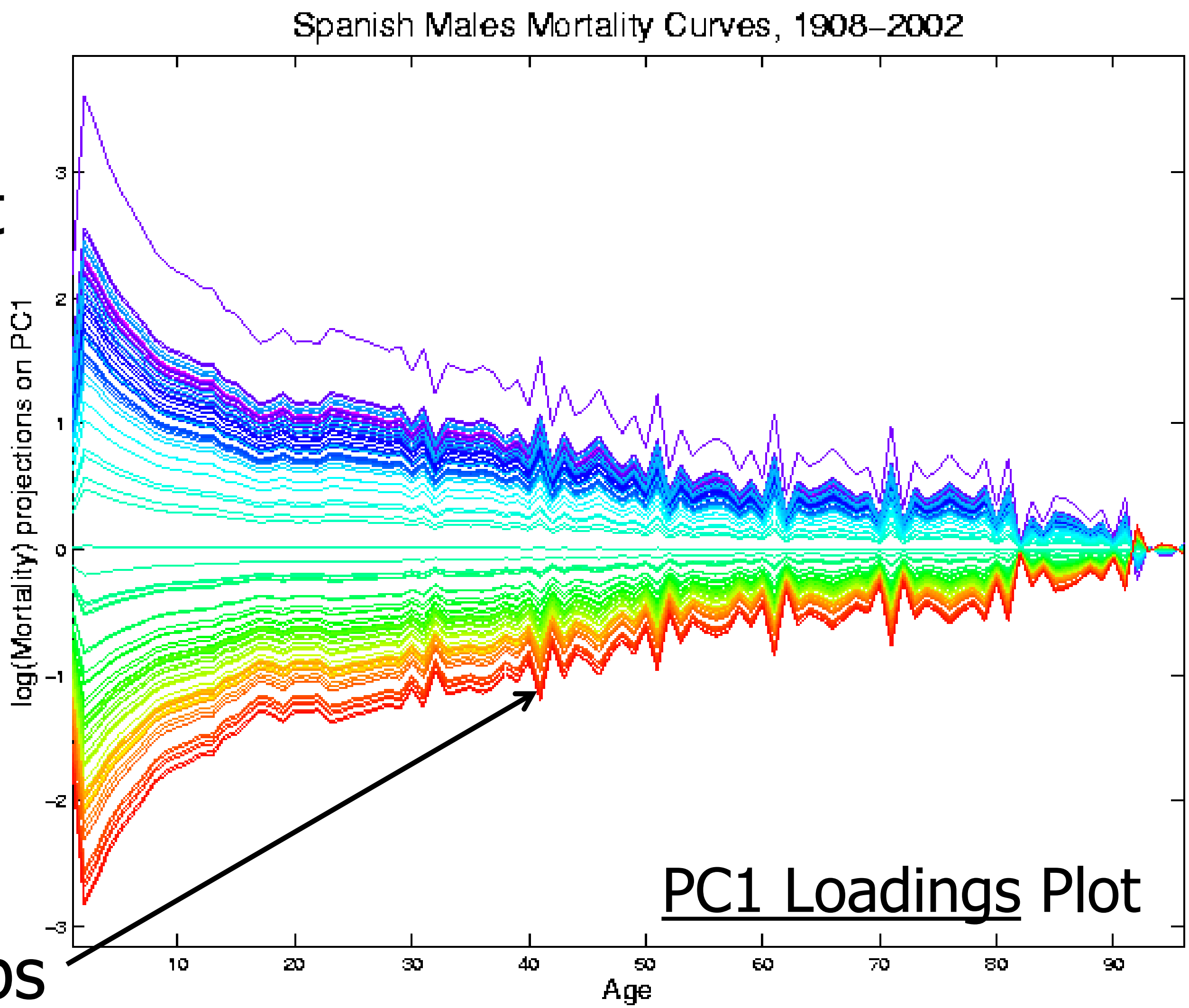
<u>Loadings:</u> Biggest Benefit For the Young



Mortality Time Series

UNC, Stat & OR

Shows *Major* Improvement Over Time



<u>PC1 Loadings</u> Plot

And Change In Age Rounding Blips

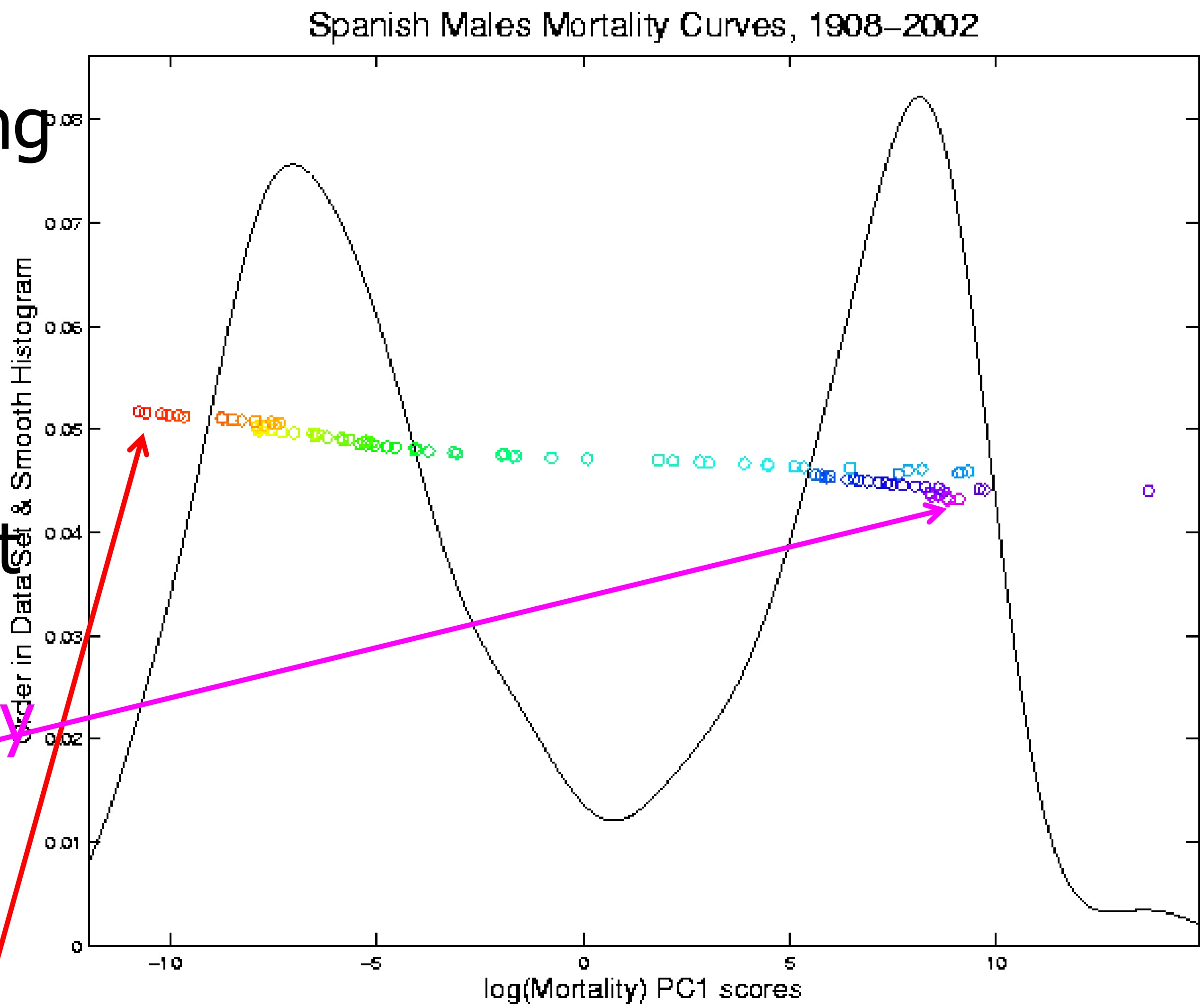
Mortality Time Series

Corresponding
PC 1 Scores

Again Shows
Overall
Improvement

High Mortality
Early

Lower Later



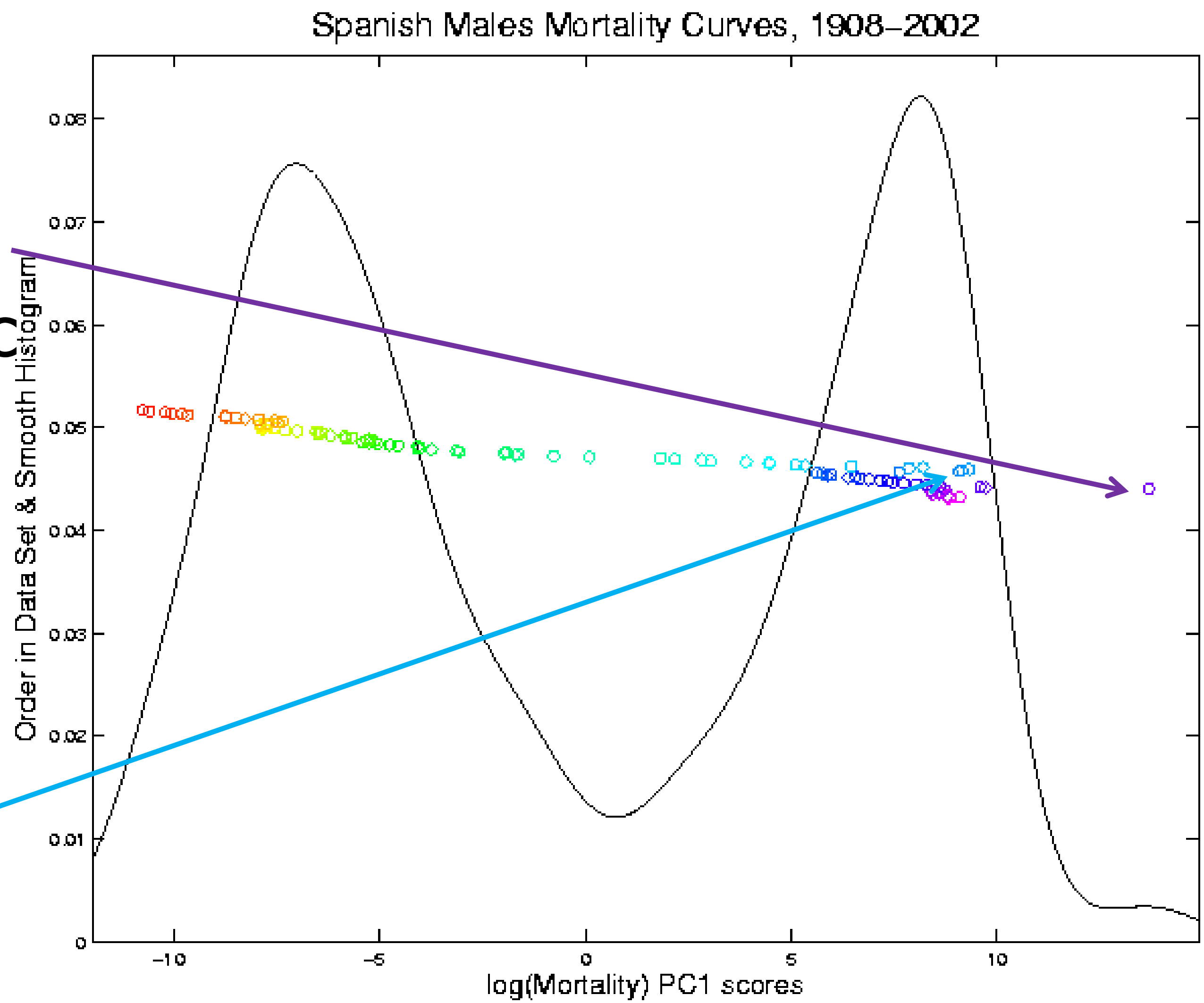
Mortality Time Series

UNC, Stat & OR

Outliers

1918 Global Flu Pandemic

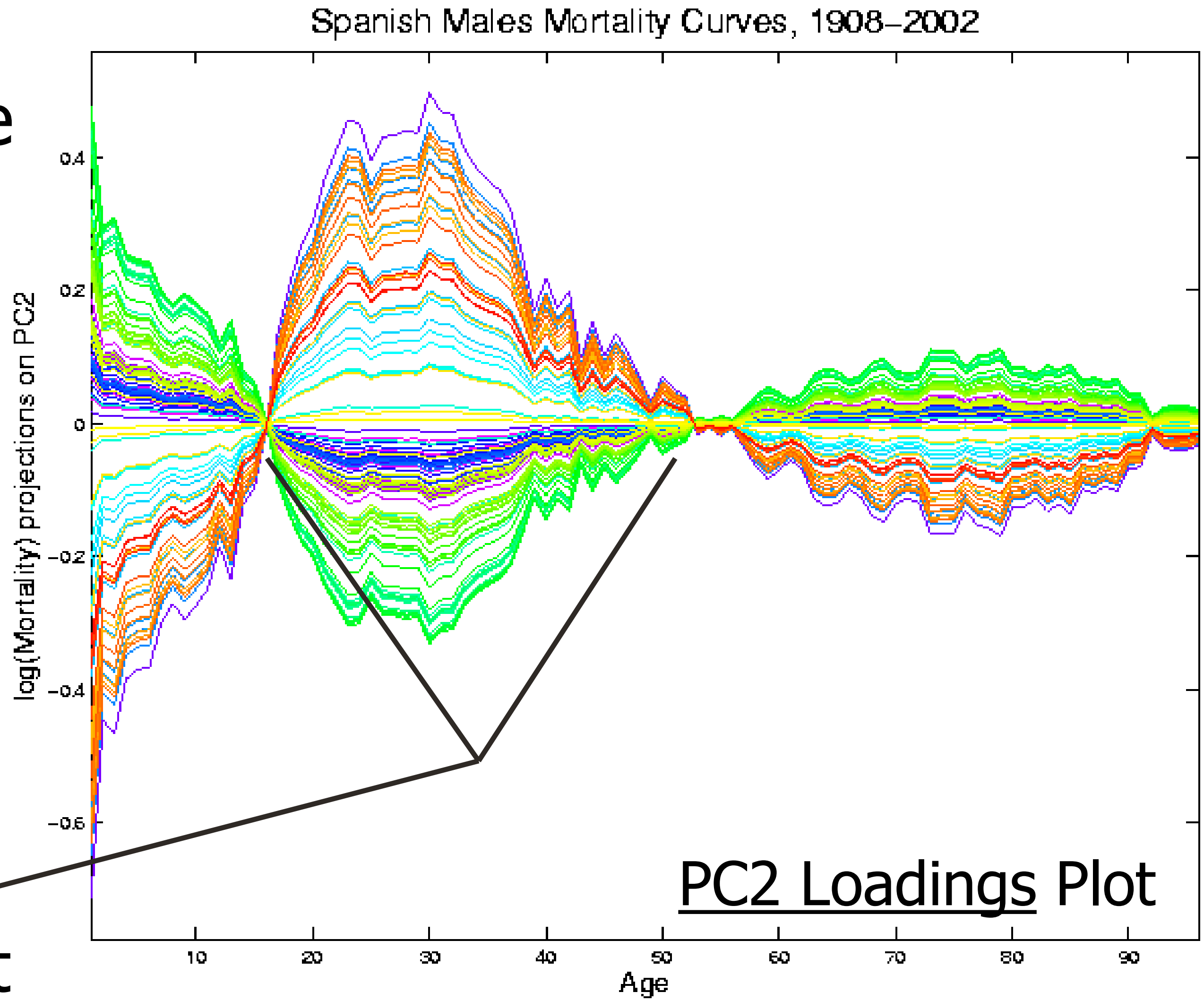
1936-1939 Spanish Civil War



Mortality Time Series

Object Space View of Projections Onto PC2 Direction

2nd Mode Of Variation: Difference Between 20-45 & Rest



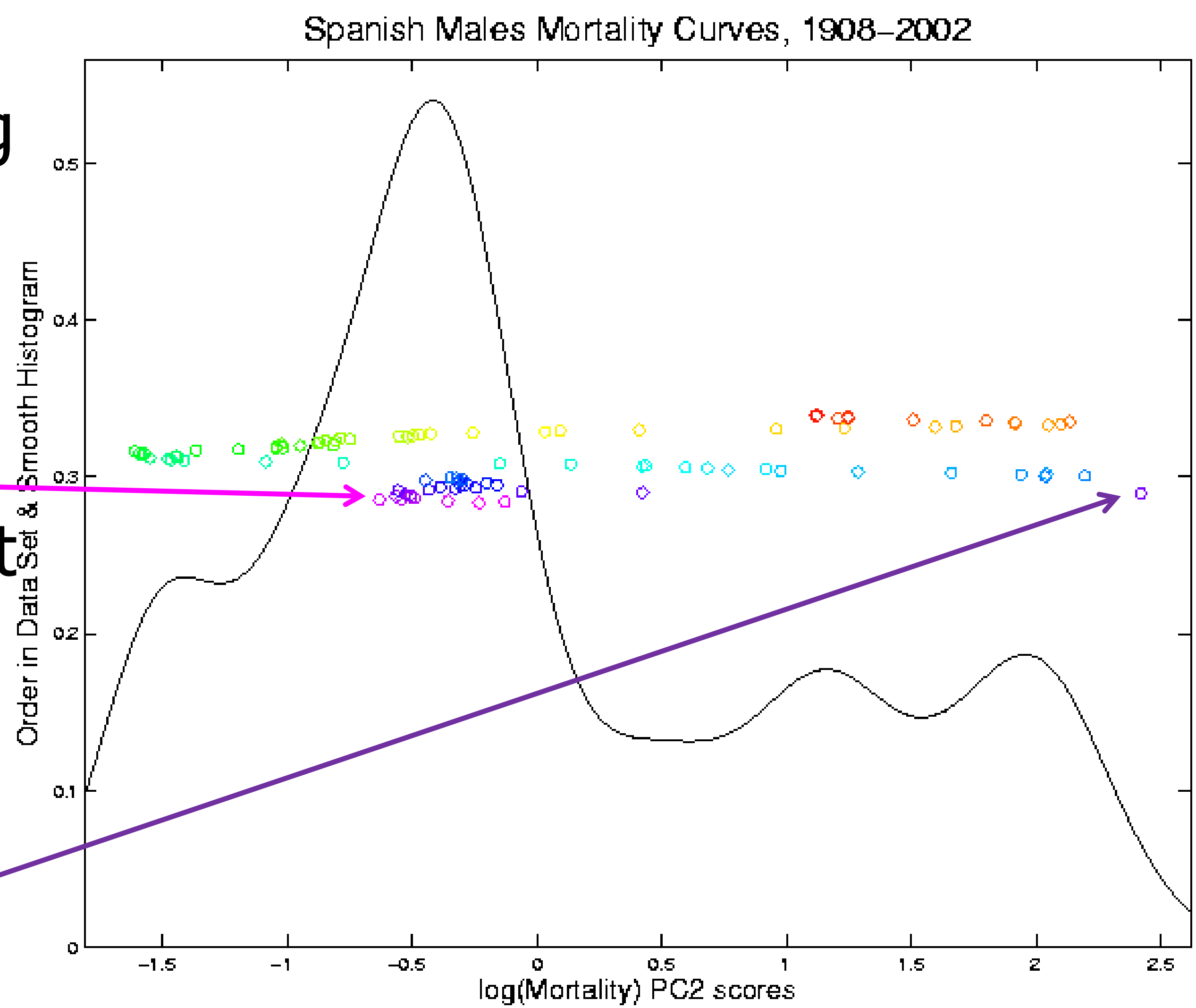
Mortality Time Series

UNC, Stat & OR

Explain Using
PC 2 Scores

Early
Improvement

Pandemic
Hit Hardest



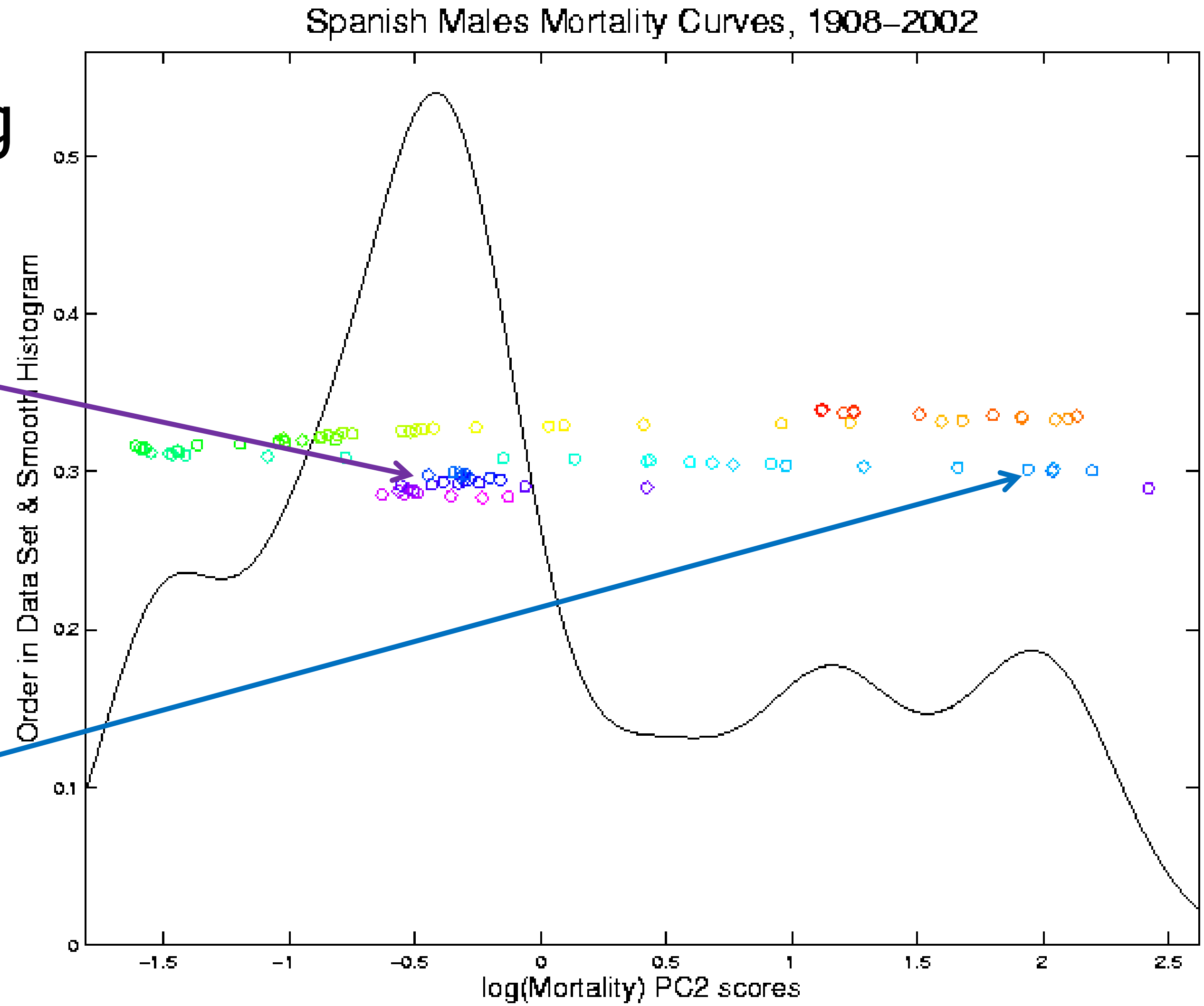
Mortality Time Series

UNC, Stat & OR

Explain Using
<u>PC 2 Scores</u>

Then better

Spanish Civil War Hit Hardest

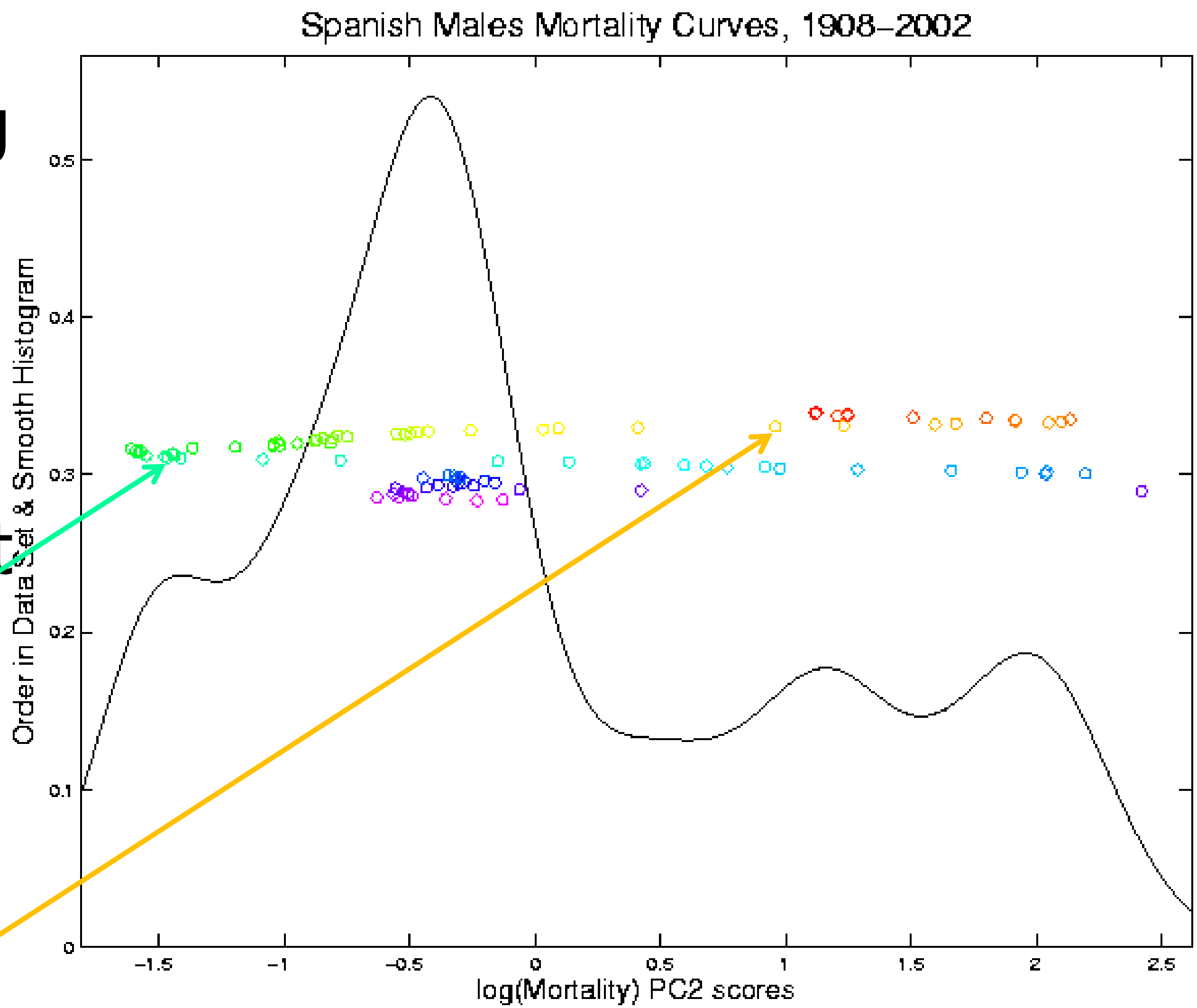


Mortality Time Series

Explain Using
PC 2 Scores

Steady
Improvement
To mid-50s

Increasing
Automotive
Death Rate



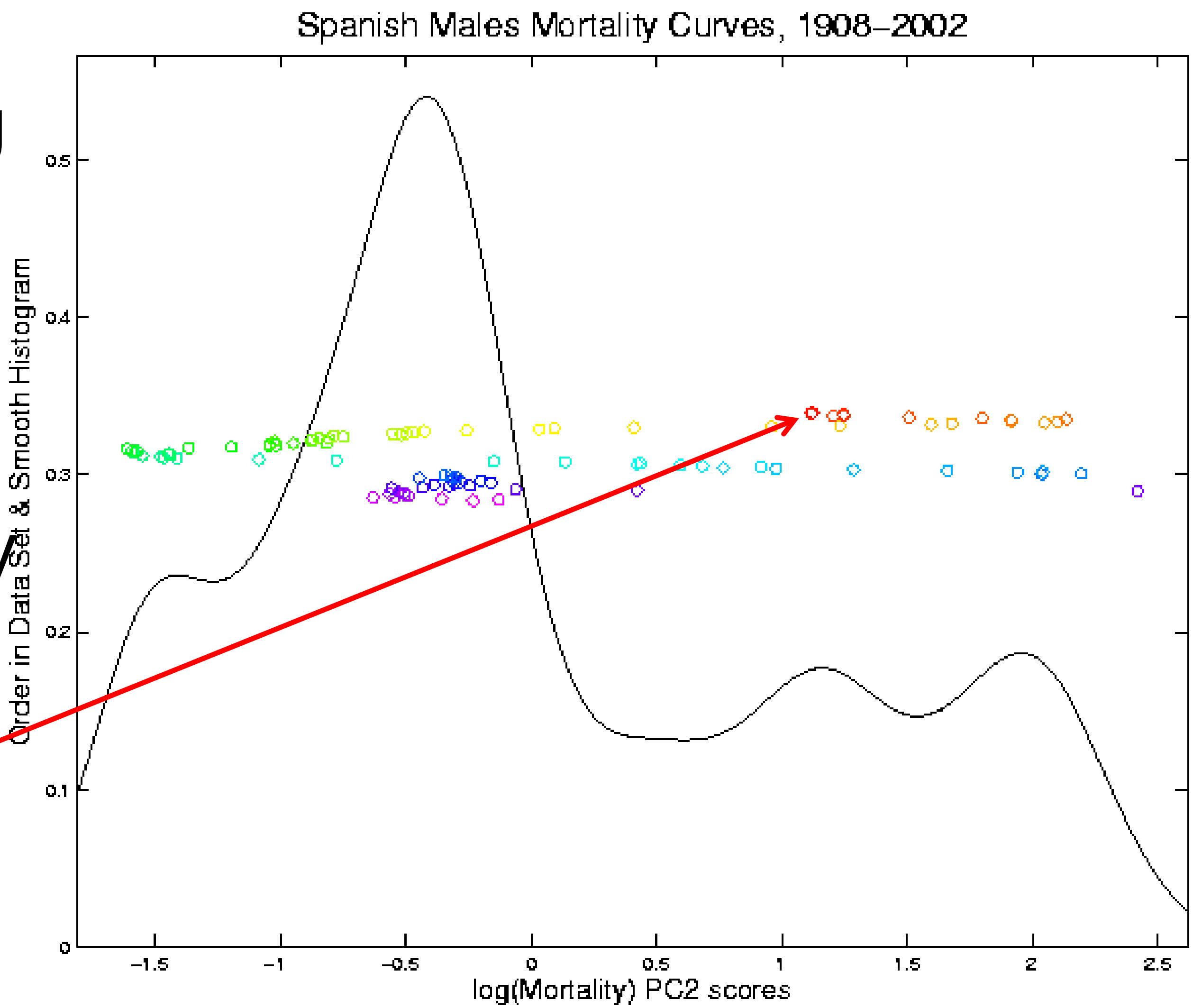
Mortality Time Series

UNC, Stat & OR

Explain Using

PC 2 Scores

Corner Finally Turned by Safer Cars & Roads



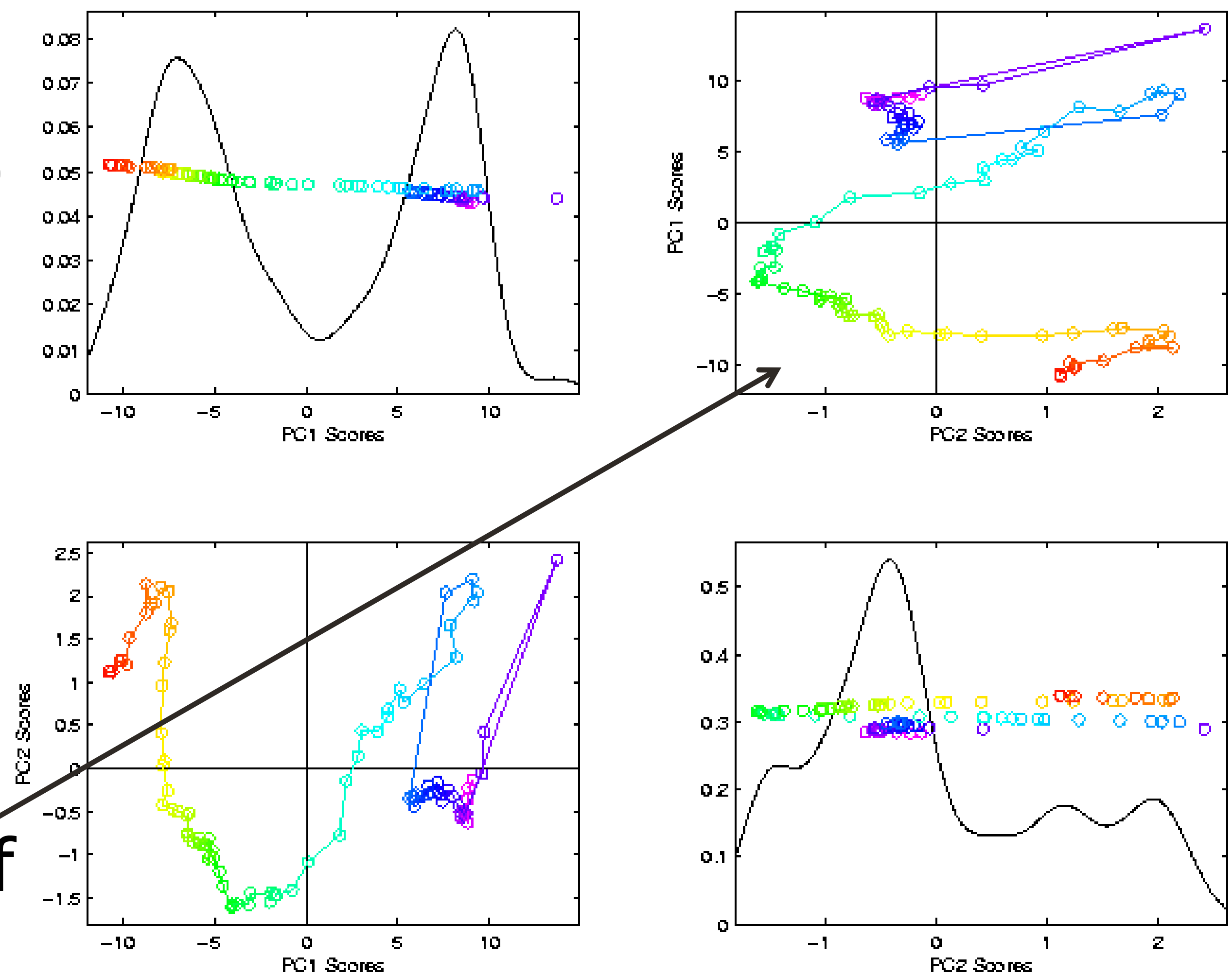
Mortality Time Series

UNC, Stat & OR

Feature (Point Cloud) Space View

Connecting Lines Highlight Time Order

Good View of Historical Effects



Functional Data Analysis

Interesting Real Data Example

- Genetics (Cancer Research)
- RNAseq (Next Gener'n Sequen'g)
- Deep look at "gene components"

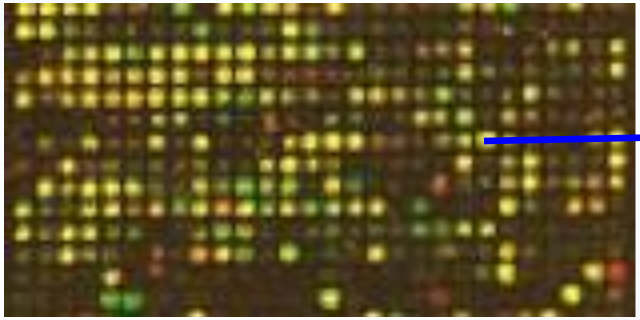
Microarrays: Single number (per gene)

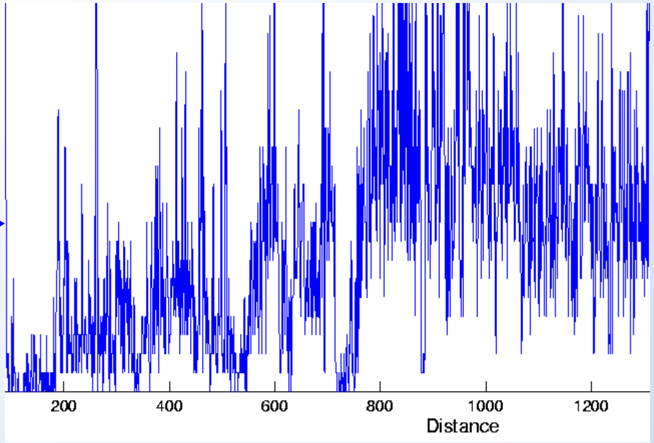
RNAseq: Thousands of measurements

Functional Data Analysis

Interesting Real Data Example

- Genetics (Cancer Research)
- RNAseq (Next Gener'n Sequen'g)
- Deep look at "gene components"





Functional Data Analysis

Interesting Real Data Example

- Genetics (Cancer Research)
- RNAseq (Next Gener’n Sequen’g)
- Deep look at “gene components”

- Gene studied here: CDKN2A
- Goal: *Study Alternate Splicing*
- Sample Size, n = 180
- Dimension, d = ~1700

Functional Data Analysis

UNC, Stat & OR

Simple

1^{st}

View:

Curve

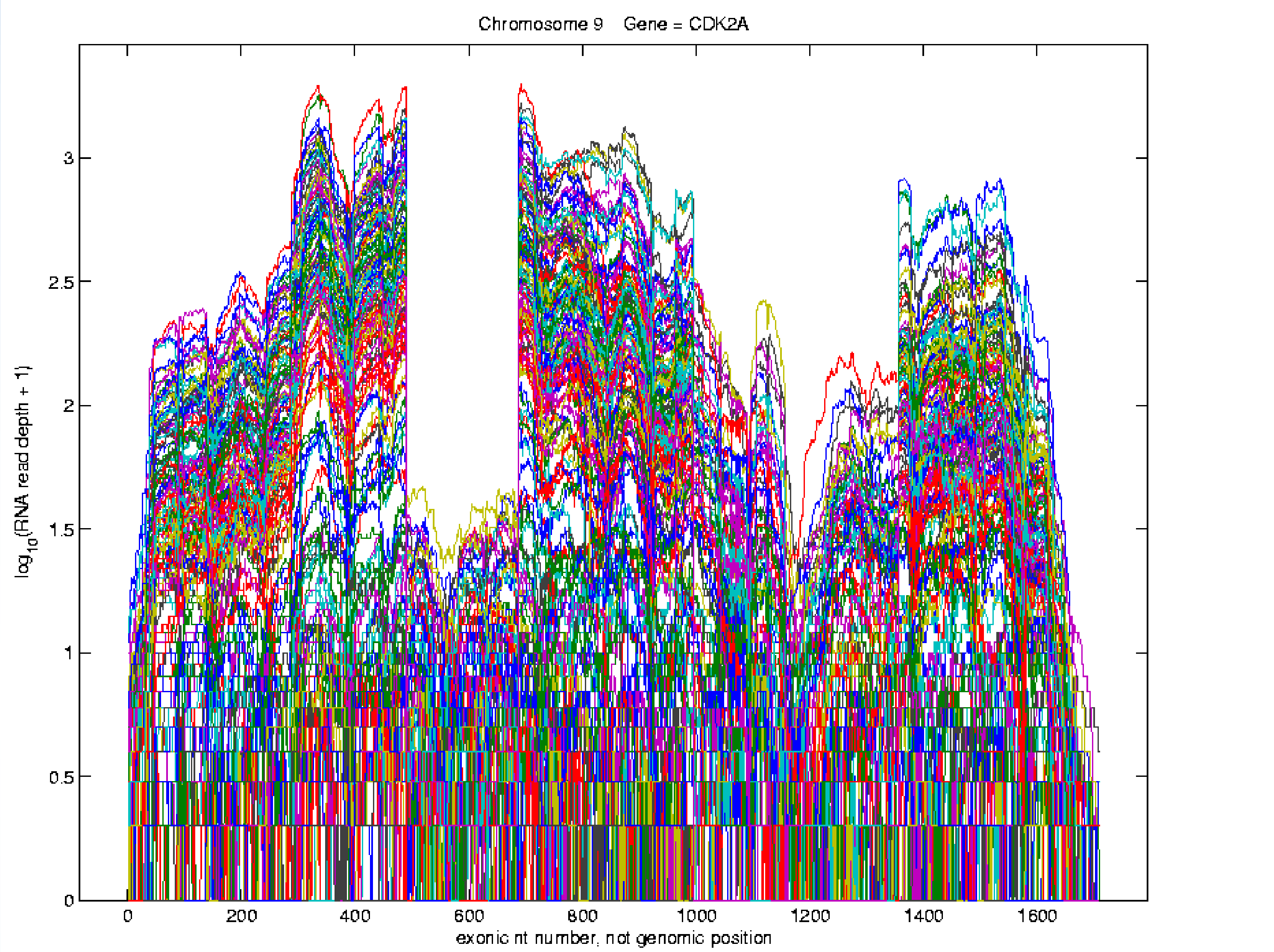
Overlay

(log

scale)

Thanks to

Matt Wilkerson



Functional Data Analysis

UNC, Stat & OR

Often

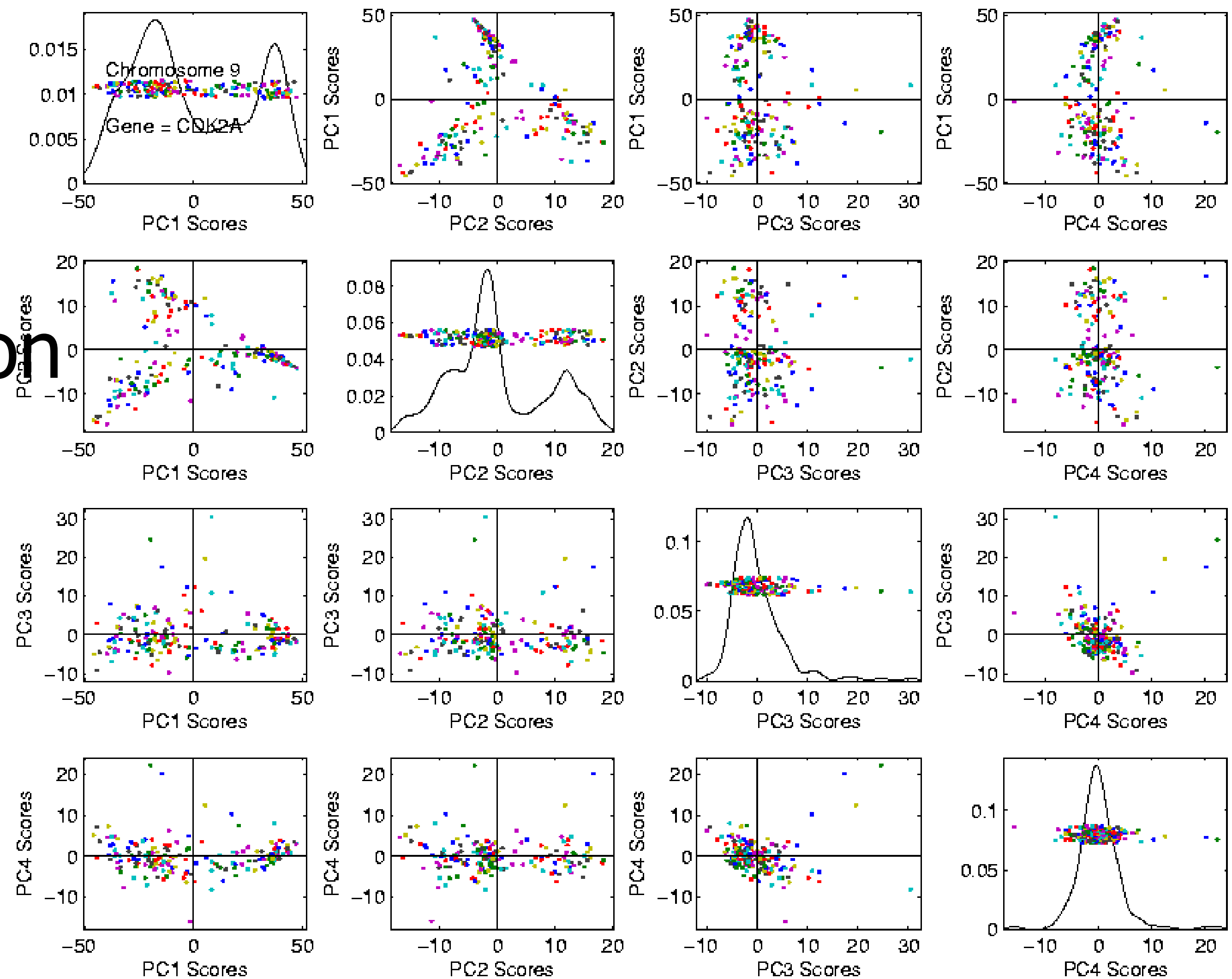
Useful

Population

View:

PCA

Scores



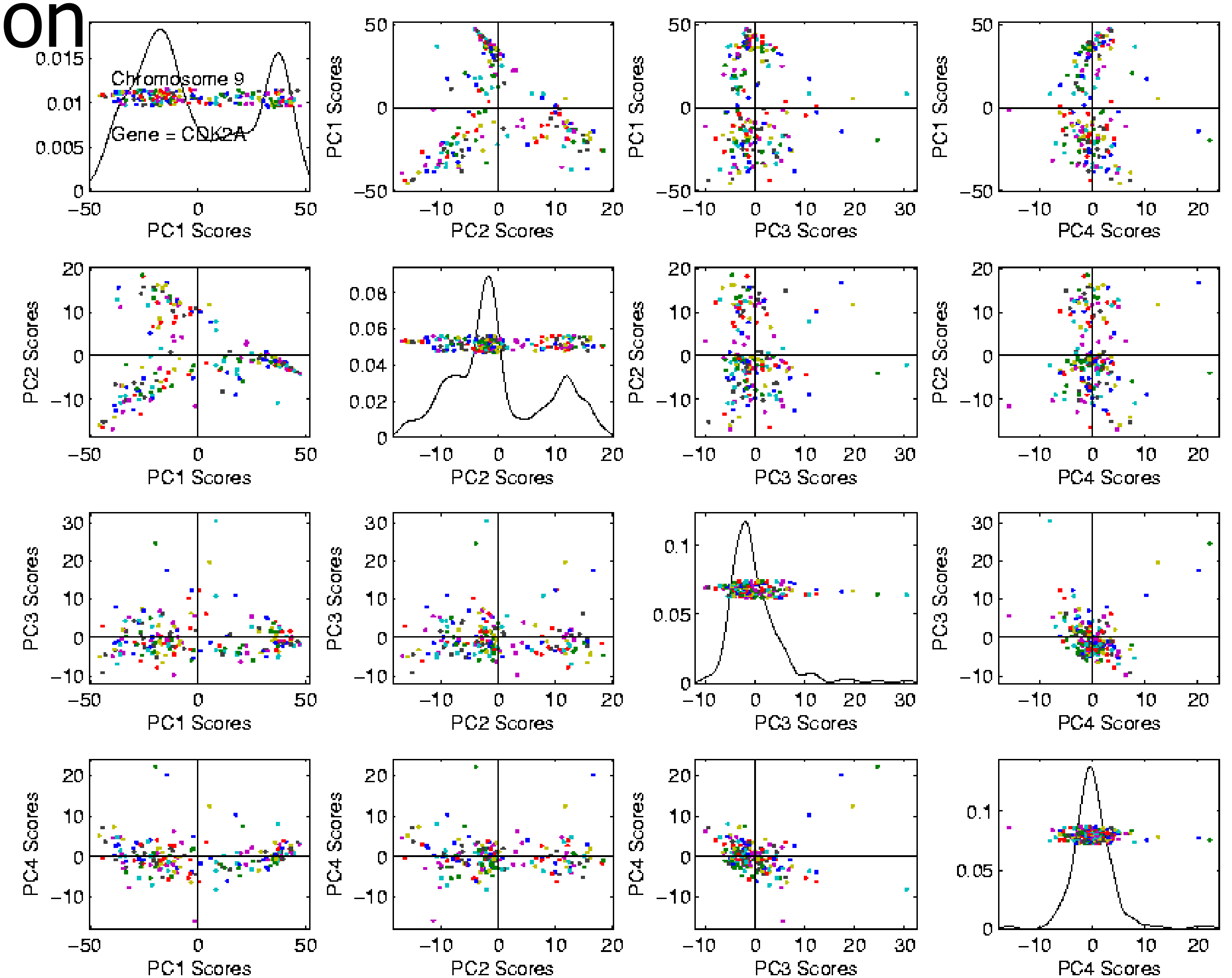
Functional Data Analysis

Suggestion

Of

Clusters

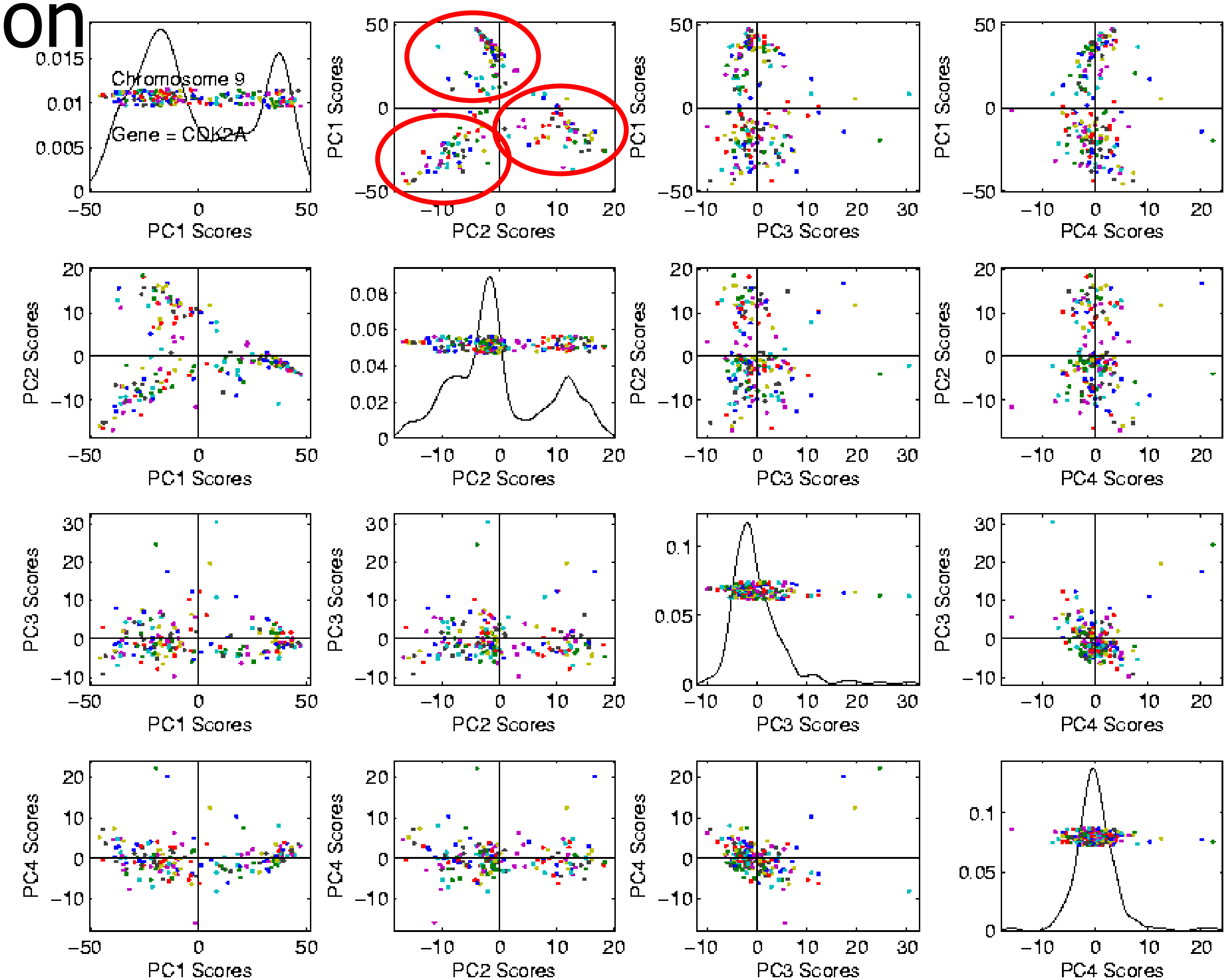
???



Functional Data Analysis

Suggestion Of Clusters

Which Are These?



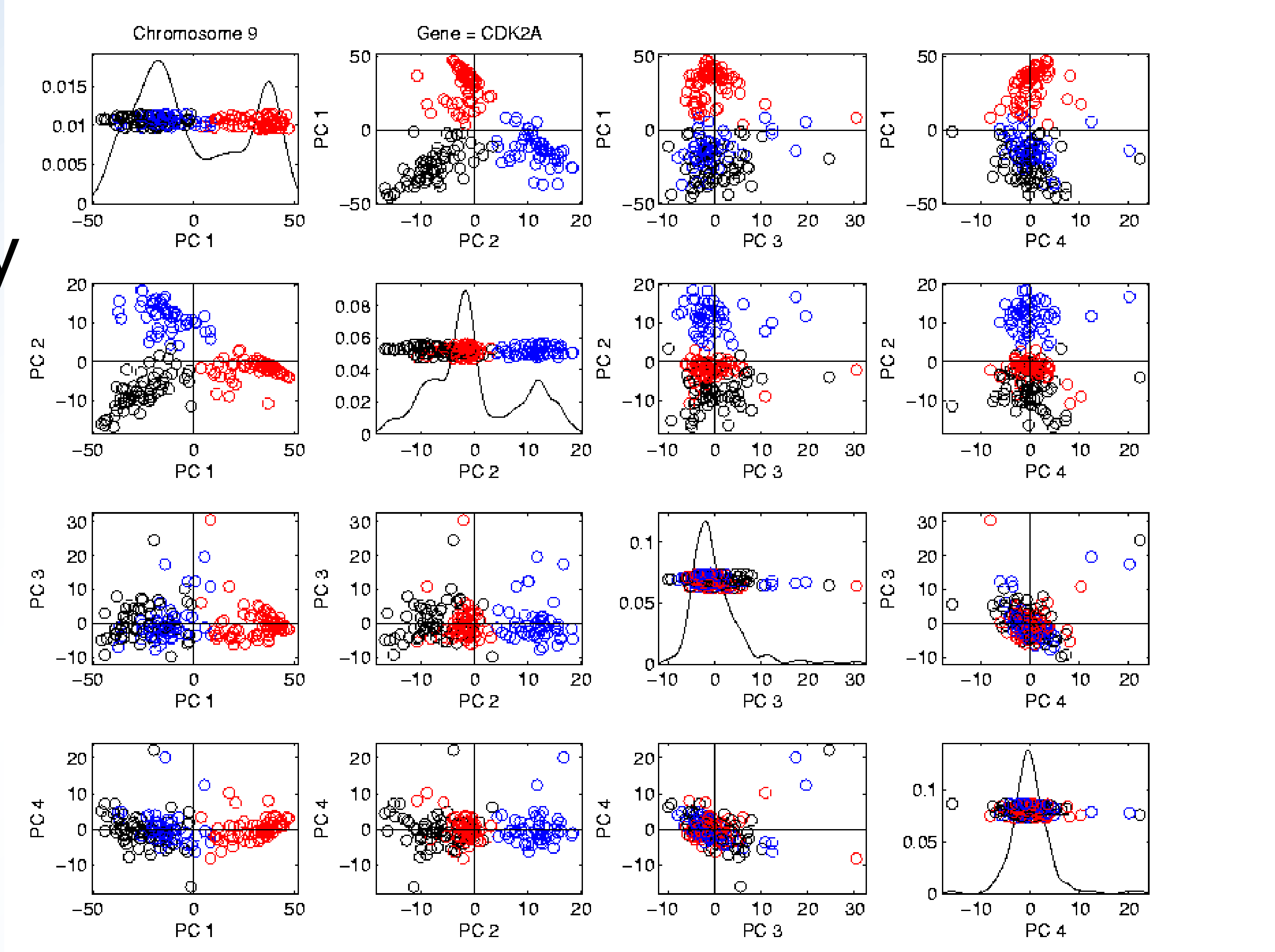
Functional Data Analysis

UNC, Stat & OR

Manually

Brush

Clusters



Functional Data Analysis

UNC, Stat & OR

Manually

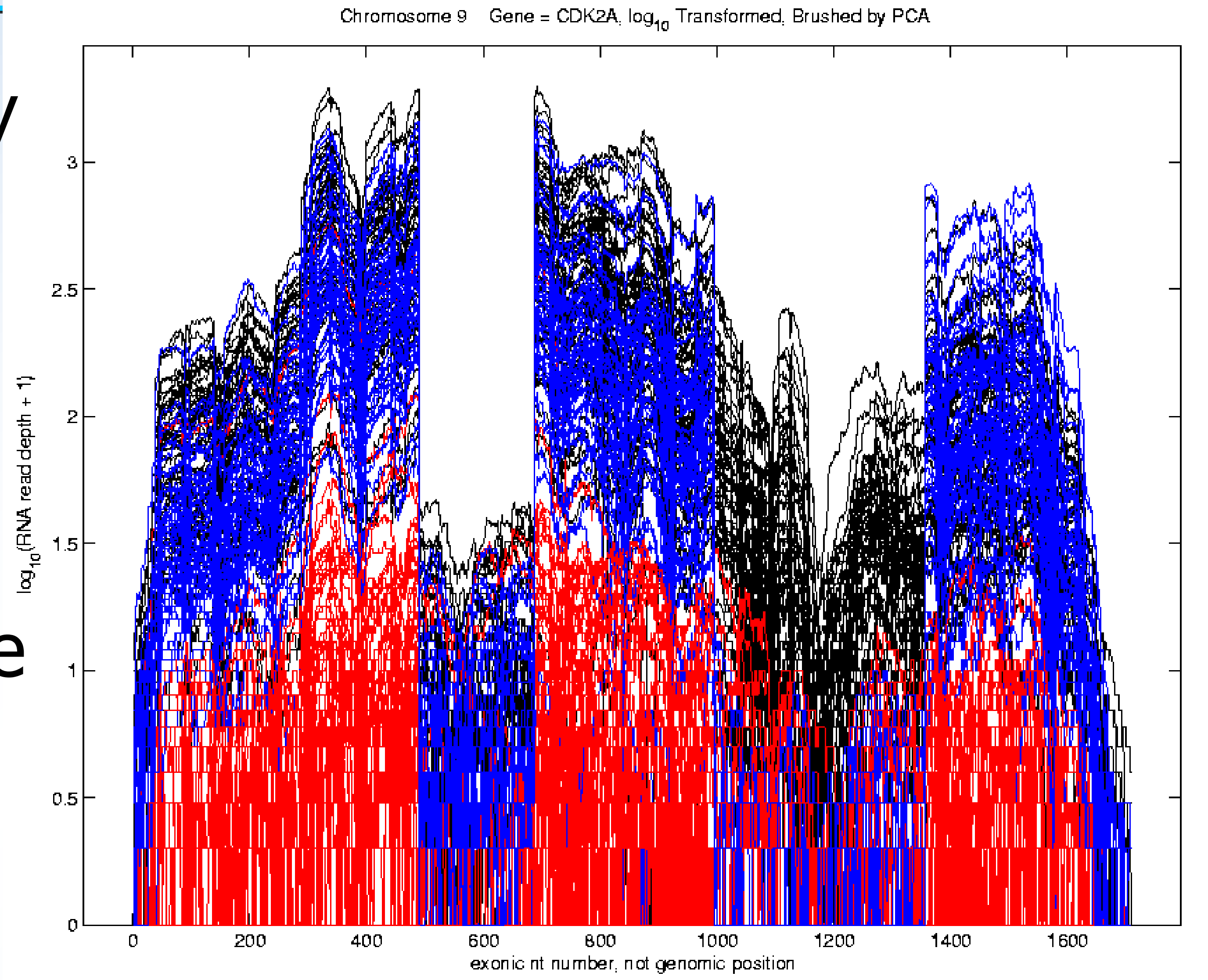
Brush

Clusters

Clear

Alternate

Splicing



Functional Data Analysis

Important Points

✓ PCA found *Important Structure*

✓ In High Dimensional Data Analysis

Object Oriented Data Analysis

What is the "atom" of a statistical analysis?

- 1st Course: Numbers
- Multivariate Analysis Course : Vectors
- Functional Data Analysis: Curves
- More generally: Data Objects

Object Oriented Data Analysis

Examples:

- Medical Image Analysis
  - Images as Data Objects?
  - Shape Representations as Objects
- Gene Expression (Microarrays – RNAseq)
  - Just multivariate analysis?

Object Oriented Data Analysis

Typical Goals:

- Understanding population variation
  - Visualization
  - Principal Component Analysis +
- Discrimination (a.k.a. Classification)
- "Vertical Integration" of Data Types

Major Statistical Challenge, I:

High Dimension Low Sample Size (HDLSS)

- Dimension d >> sample size n
- "Multivariate Analysis" nearly useless
  - Can't "normalize the data"
- Land of Opportunity for Statisticians
  - Need for "creative statisticians"

Aside on Terminology

Personal suggestion:

High Dimension Low Sample Size (HDLSS)

- Dimension: d
- Sample size n

Versus: "Small n, large p"

- Why p? (parameters??? predictors???)
- Only because of statistical tradition…

Major Statistical Challenge, II:

- Data may live in *non-Euclidean space*
  - Lie Group / Symmet'c Spaces (manifold data)
  - Trees/Graphs as data objects
- Interesting Issues:
  - What is "the mean" (pop'n center)?
  - How do we quantify "pop'n variation"?

Statistics in Image Analysis, I

First Generation Problems:

- Denoising
- Segmentation
- Registration

(all about single images)

Statistics in Image Analysis, II

Second Generation Problems:

- Populations of Images
  - Understanding Population Variation
  - Discrimination (a.k.a. Classification)
- Complex Data Structures (& Spaces)
- HDLSS Statistics

HDLSS Statistics in Imaging

Why HDLSS (High Dim, Low Sample Size)?

- Complex 3-d Objects Hard to Represent
  - Often need d = 100's of parameters
- Complex 3-d Objects Costly to Segment
  - Often have n = 10's cases

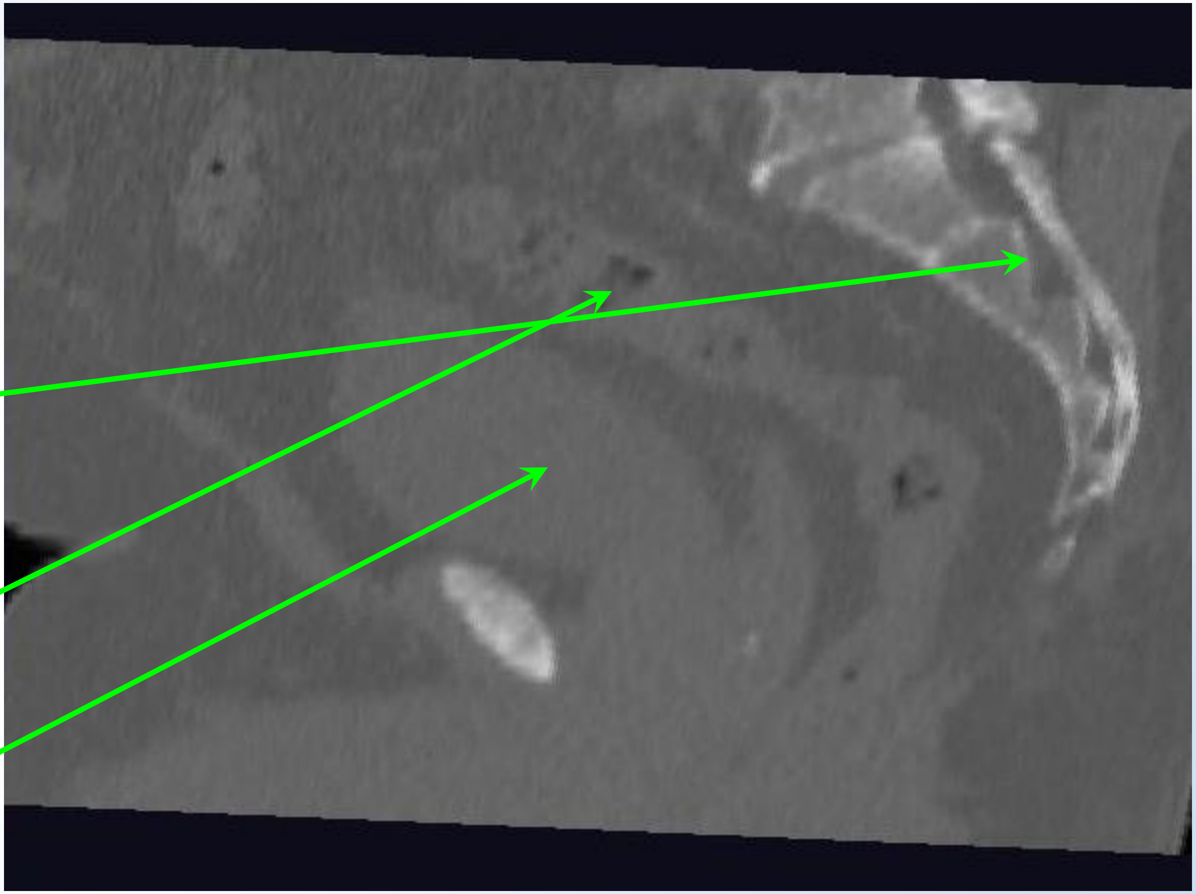
Medical Imaging – A Challenging Example

- Male Pelvis
  - Bladder – Prostate – Rectum
  - How do they move over time (days)?
  - Critical to Radiation Treatment (cancer)
- Work with 3-d CT
- *Very* Challenging to Segment
  - Find boundary of each object?
  - Represent each Object?

Male Pelvis – Raw Data

One CT Slice

(in 3d image)



Coccyx

(Tail Bone)

Rectum

Bladder

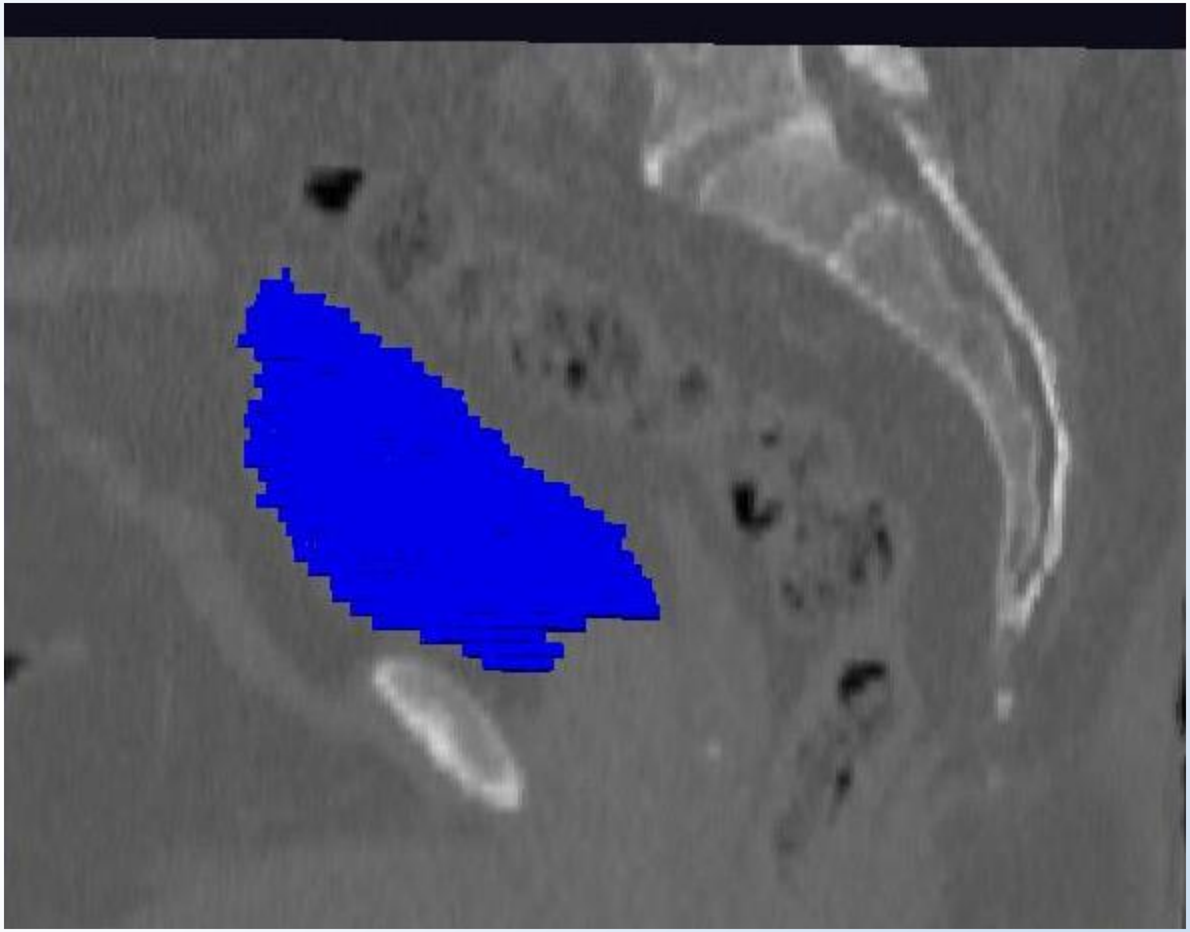
Male Pelvis – Raw Data

Bladder:

manual segmentation

Slice by slice

Reassembled



Male Pelvis – Raw Data

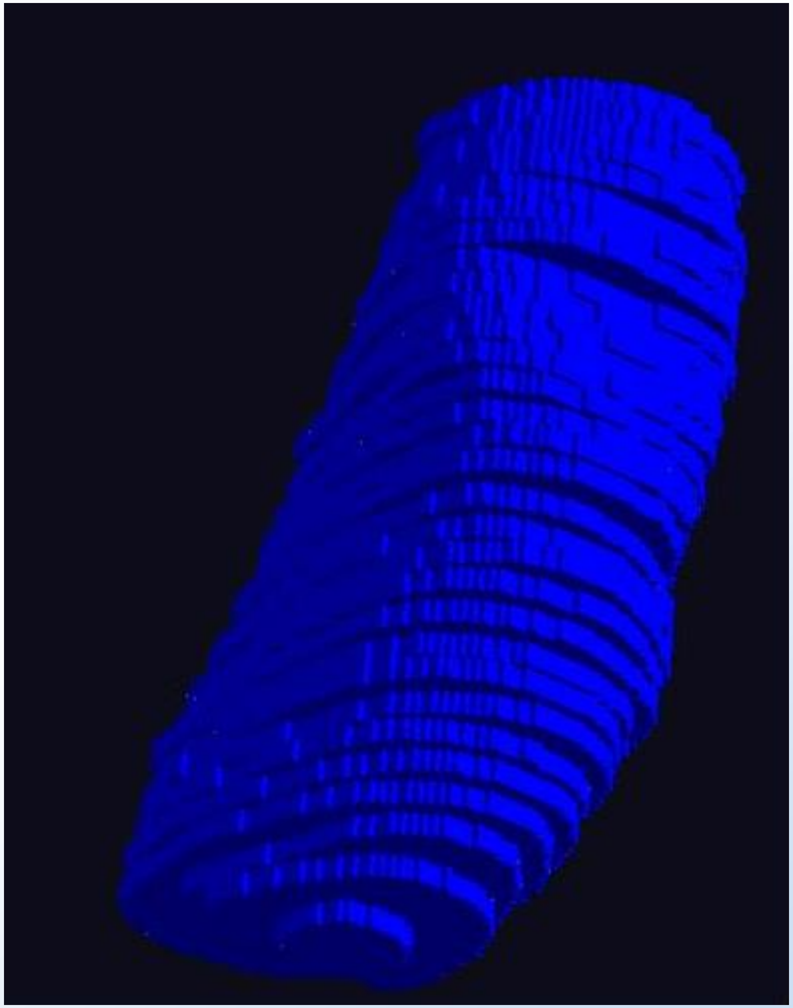
Bladder:

Slices:

Reassembled in 3d

<u>How to represent?</u>

Thanks: Ja-Yeon Jeong



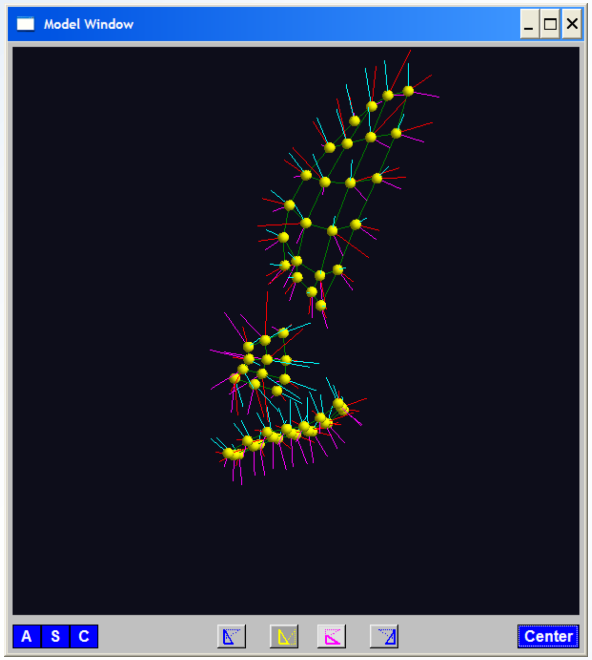
Object Representation

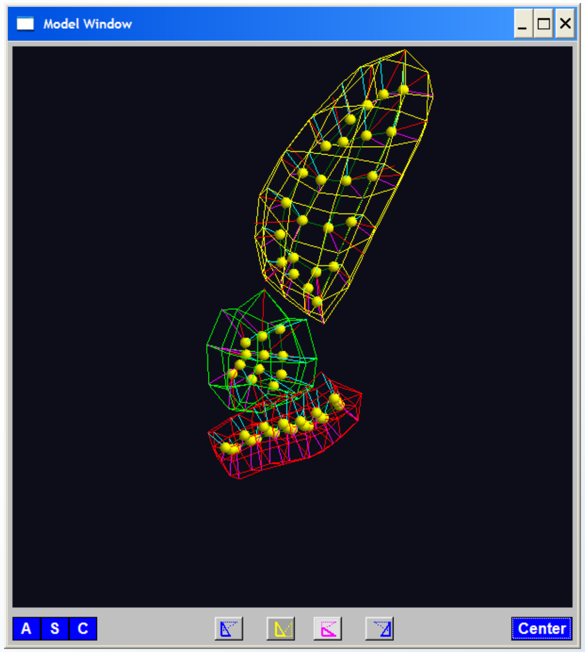
- Landmarks (hard to find)
- Boundary Rep'ns (no correspondence)
- Medial representations
  - Find "skeleton"
  - Discretize as "atoms" called M-reps

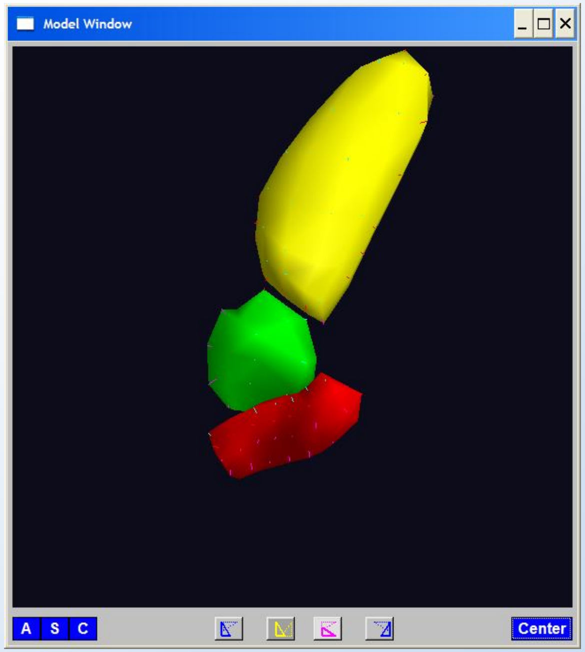
3-d m-reps

Bladder – Prostate – Rectum (multiple objects, J. Y. Jeong)

- Medial Atoms provide “skeleton”
- Implied Boundary from “spokes” → “surface”







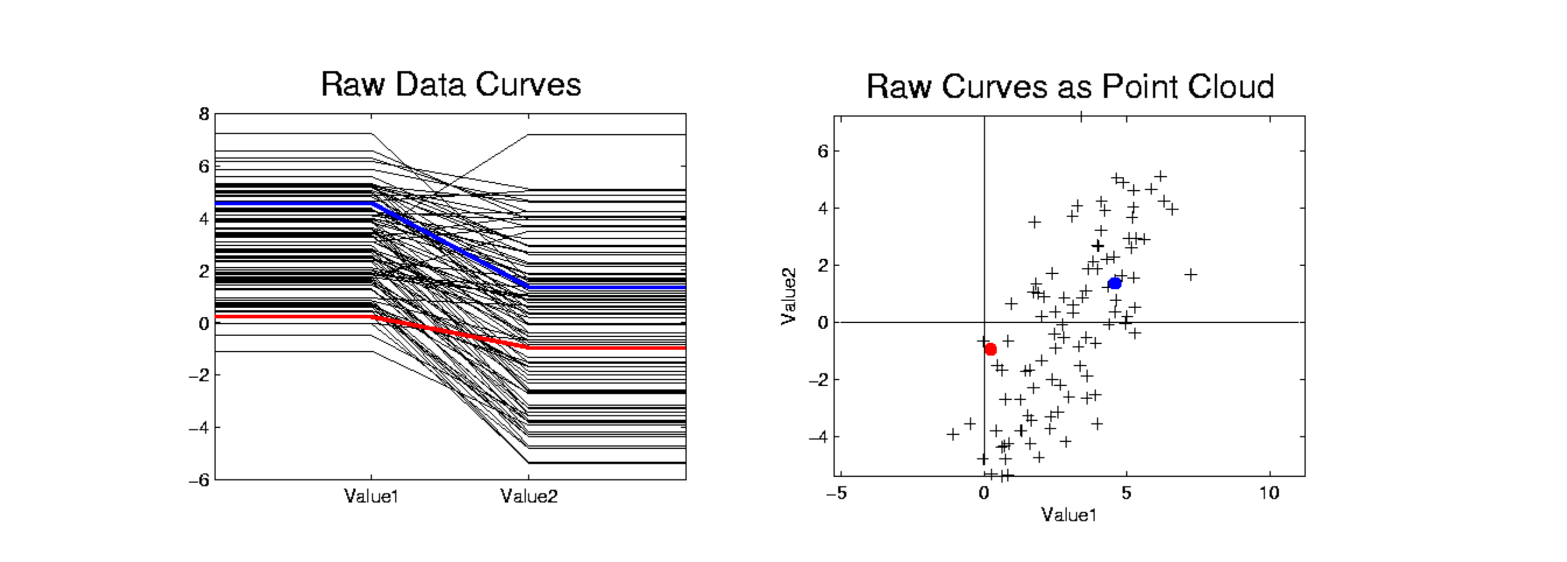
3-d m-reps

M-rep model fitting

- Easy, when starting from *binary* (blue)
- But very expensive (30 – 40 minutes technician's time)
- Want *automatic approach*
- Challenging, because of poor contrast, noise, …
- Need to borrow information across training sample
- Use Bayes approach: prior & likelihood → posterior
- ~Conjugate Gaussians, but there are issues:
  - Major HLDSS challenges
  - Manifold aspect of data

Illuminating Viewpoint

Object Space ←→ Feature Space



Focus here on collection of *data objects*

Here conceptualize *population structure* via "point clouds"

Data Lying On a Manifold

Major issue: m-reps live in $\Re^3 \times \Re^+ \times S^2 \times S^2$

(locations, radius and angles)

E.g. "average" of: $2^\circ, 3^\circ, 358^\circ, 359^\circ =$???

Data Lying On a Manifold

Major issue: m-reps live in $\Re^3 \times \Re^+ \times S^2 \times S^2$

(locations, radius and angles)

E.g. "average" of: $2^\circ, 3^\circ, 358^\circ, 359^\circ =$???

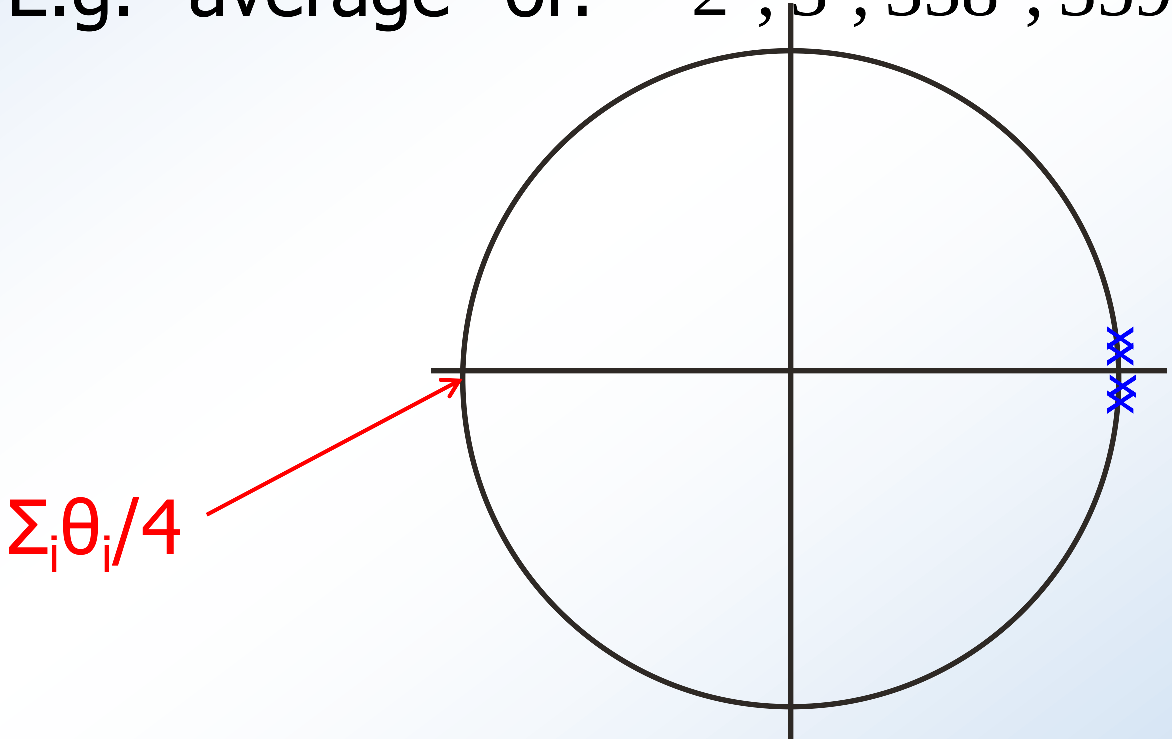
$\Sigma_i \theta_i / 4$?

Data Lying On a Manifold

Major issue: m-reps live in $\Re^3 \times \Re^+ \times S^2 \times S^2$

(locations, radius and angles)

E.g. "average" of: $2^\circ, 3^\circ, 358^\circ, 359^\circ =$???

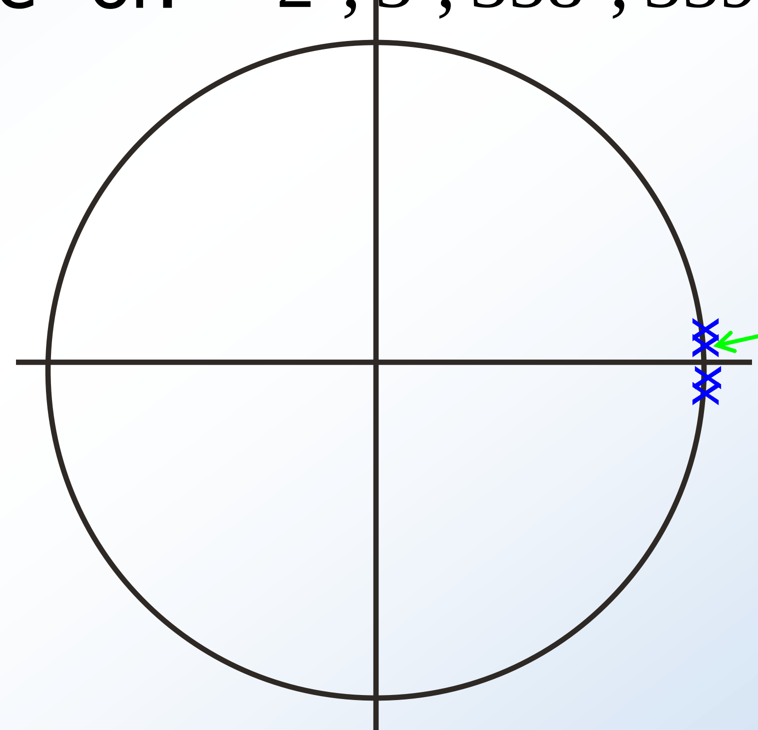


Data Lying On a Manifold

Major issue: m-reps live in $\Re^3 \times \Re^+ \times S^2 \times S^2$

(locations, radius and angles)

E.g. "average" of: $2^\circ, 3^\circ, 358^\circ, 359^\circ =$???



Should Use Unit Circle Structure

Data Lying On a Manifold

Major issue: m-reps live in $\Re^3 \times \Re^+ \times S^2 \times S^2$

(locations, radius and angles)

E.g. "average" of: $2^\circ, 3^\circ, 358^\circ, 359^\circ =$???

Natural Data Structure is:

Lie Groups ~ Symmetric spaces

(smooth, curved manifolds)

Mildly Non-Euclidean Space

UNC, Stat & OR

Useful View of Manifold Data: Tangent Space

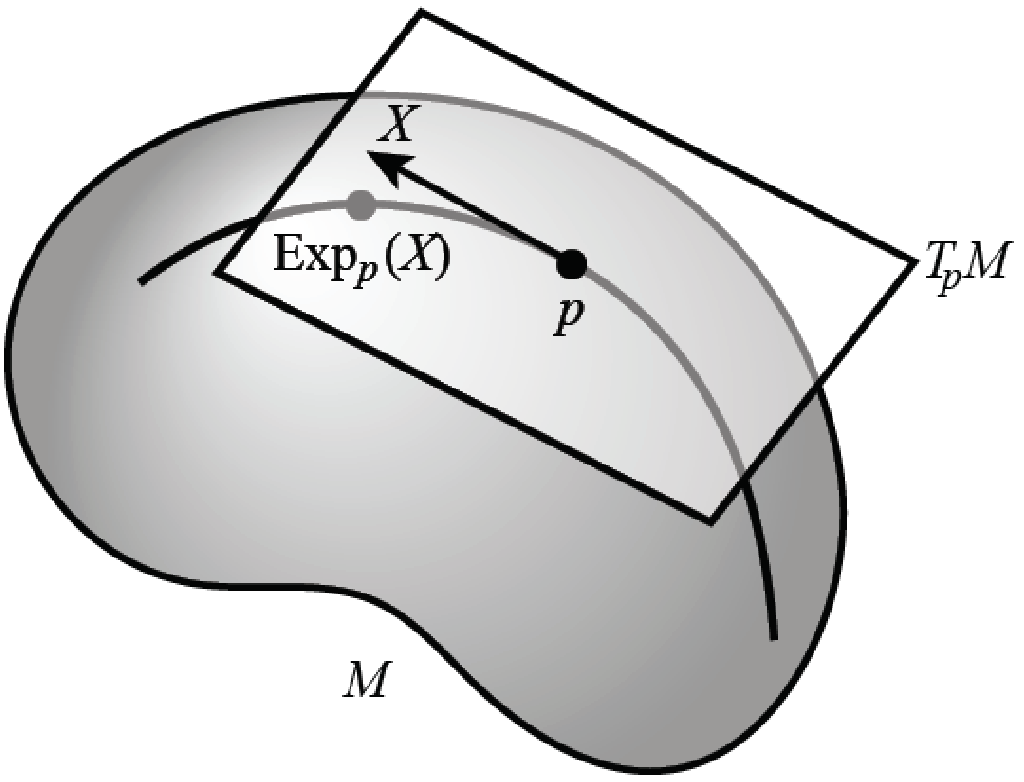


Figure 2.2: The Riemannian exponential map.

Thanks to P. T. Fletcher

Mildly Non-Euclidean Space

Useful View of Manifold Data: Tangent Space

At each point, ∃
Approximating
Tangent Plane

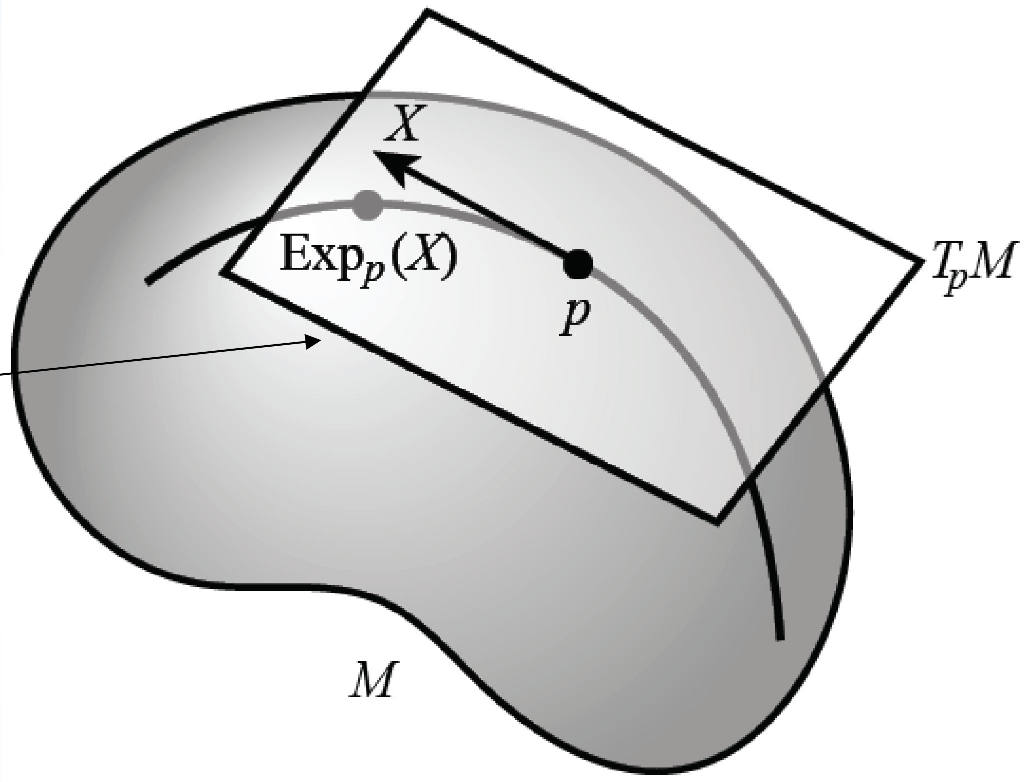


Figure 2.2: The Riemannian exponential map.

Thanks to P. T. Fletcher

Mildly Non-Euclidean Space

Useful View of Manifold Data: Tangent Space

At each point, ∃ Approximating Tangent Plane

Reason for terminology "mildly non Euclidean"

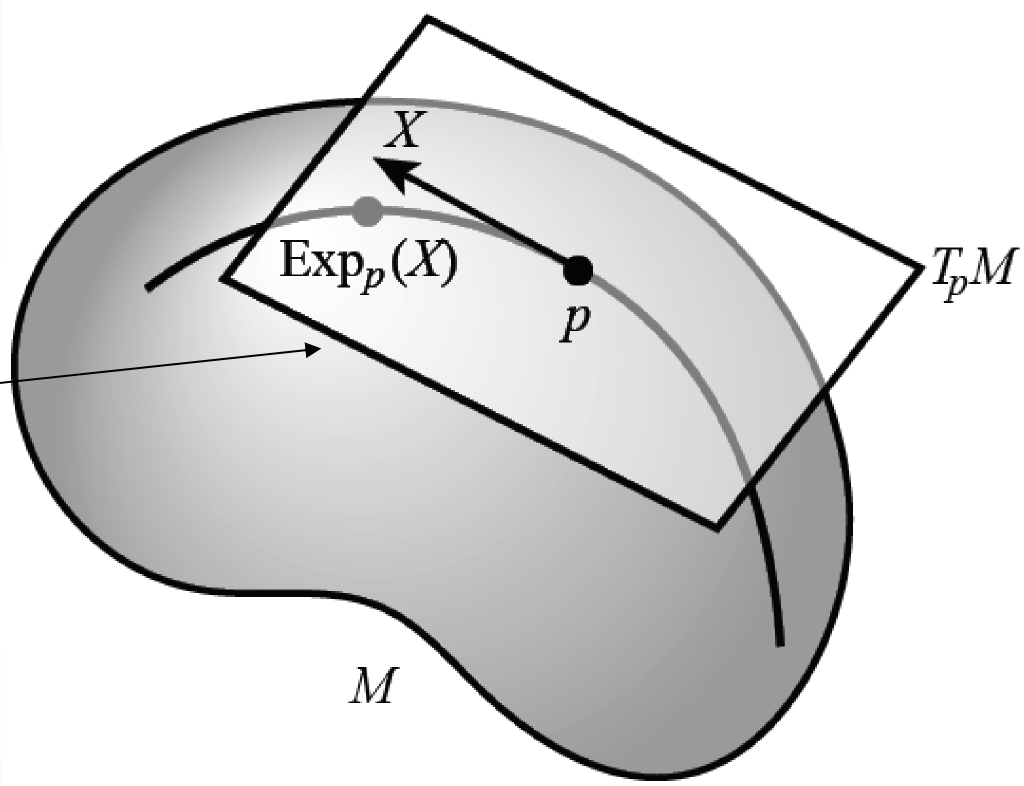


Figure 2.2: The Riemannian exponential map.

Thanks to P. T. Fletcher

Mildly Non-Euclidean Space

UNC, Stat & OR

Useful View of Manifold Data: Tangent Space

Useful Data

Center Point:

Geodesic Mean

(= Frèchet Mean)

(= Barycenter)

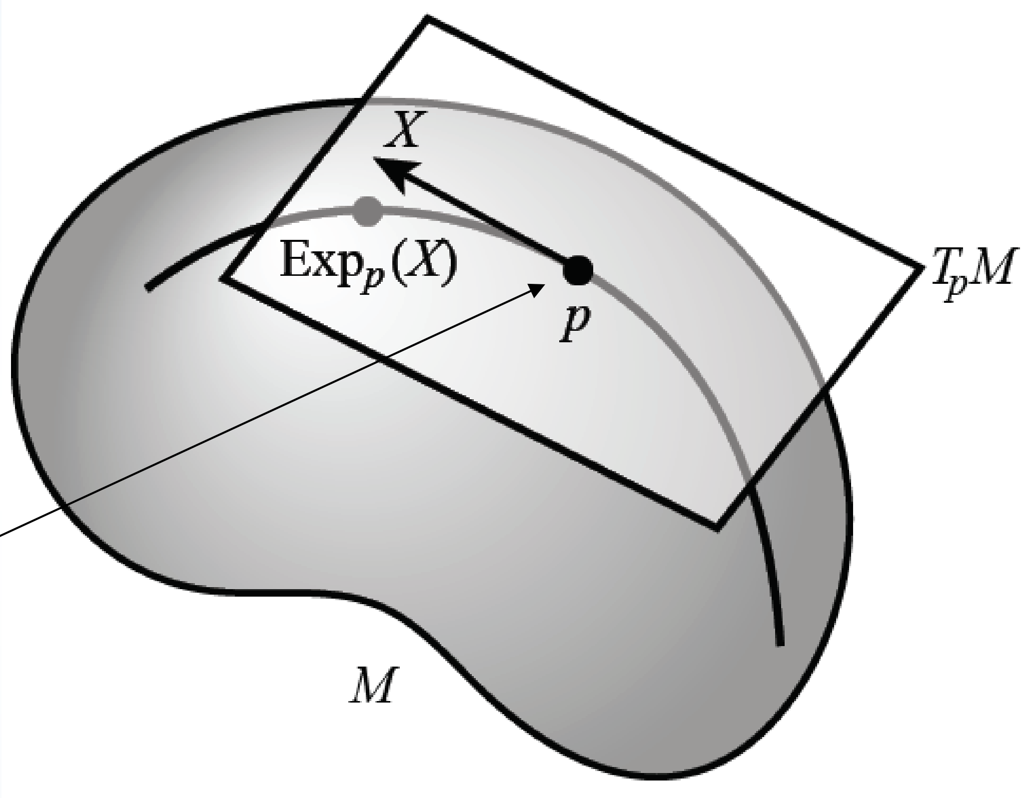


Figure 2.2: The Riemannian exponential map.

Thanks to P. T. Fletcher

Geodesic Mean

For $X_1, \cdots, X_n$ in any metric space:

$$\text{Mean} = \ argmin_x \sum_{i=1}^{n} d(x, X_i)^2$$

($x =$ point with least square distance to data)

Geodesic Mean

For $X_1, \cdots, X_n$ in <u>any metric space</u>:

$$\text{Mean} = \; argmin_x \sum_{i=1}^{n} d(x, X_i)^2$$

($x =$ point with least square distance to data)

Geodesic Mean (on Manifolds):

$d =$ Geodesic Distance

(Along Manifold Surface)

Data Lying On a Manifold

PCA on non-Euclidean spaces?

(i.e. on Lie Groups / Symmetric Spaces)

T. Fletcher: Principal Geodesic Analysis

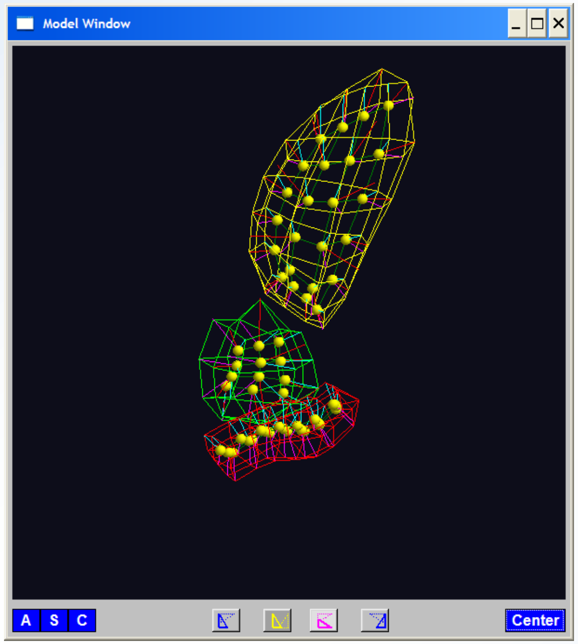
Idea: replace "linear summary of data"
With "geodesic summary of data"…

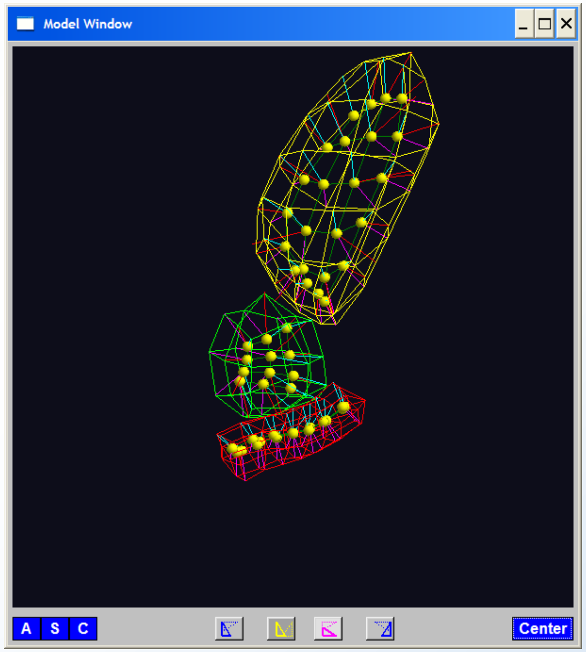
PGA for m-reps, Bladder-Prostate-Rectum

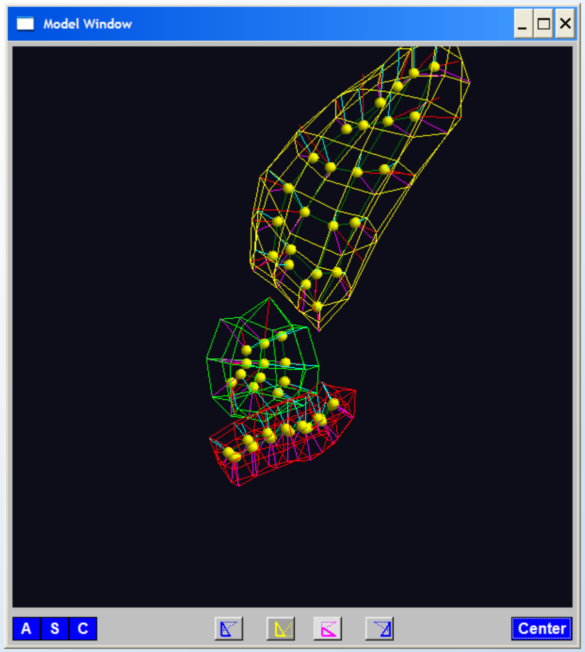
Bladder – Prostate – Rectum, 1 person, 17 days

PG 1 PG 2 PG 3

(analysis by Ja Yeon Jeong)





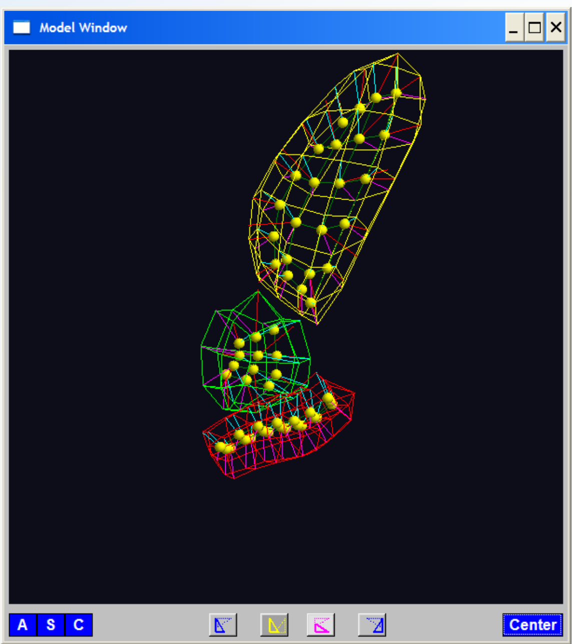


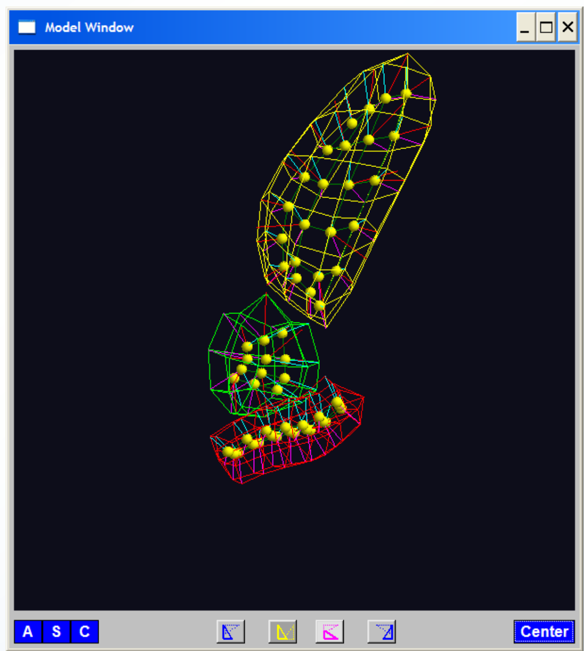
PGA for m-reps, Bladder-Prostate-Rectum

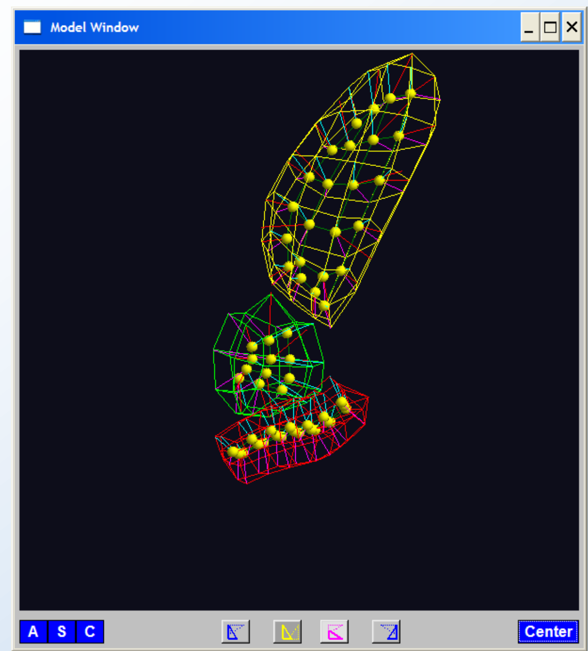
Bladder – Prostate – Rectum, 1 person, 17 days

PG 1 PG 2 PG 3

(analysis by Ja Yeon Jeong)





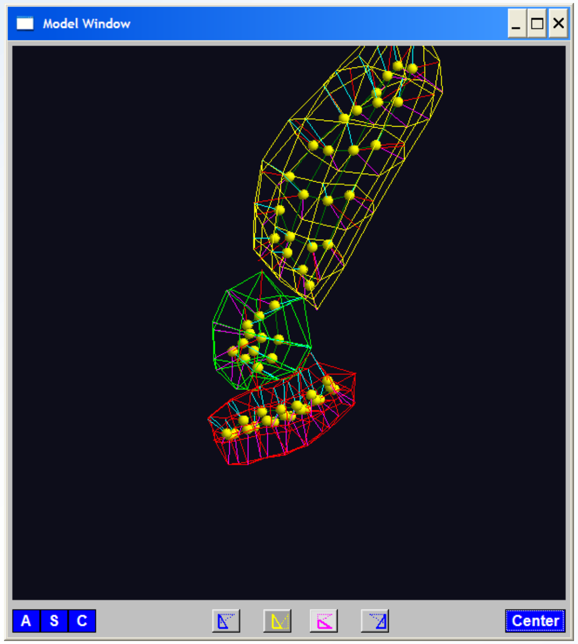


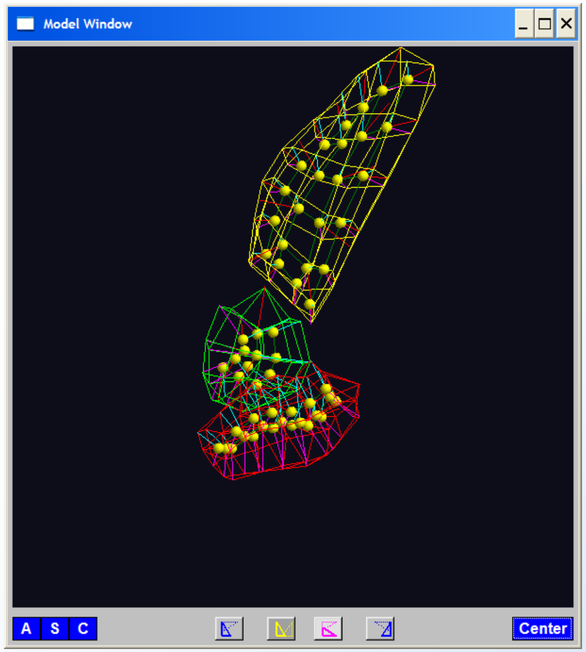
PGA for m-reps, Bladder-Prostate-Rectum

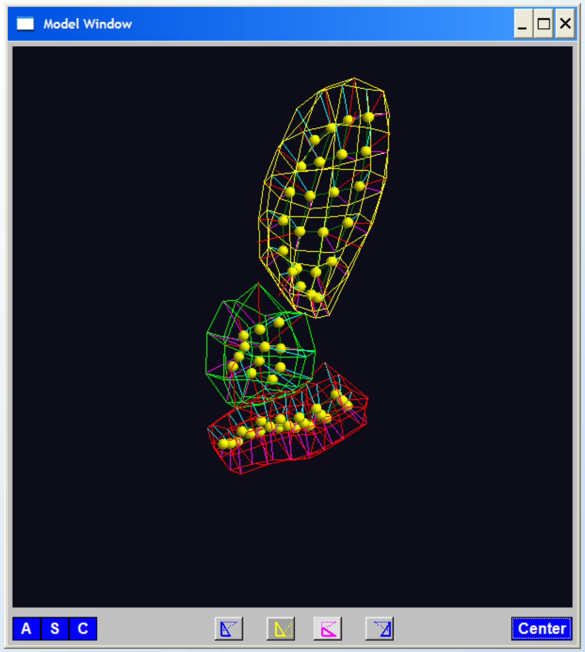
Bladder – Prostate – Rectum, 1 person, 17 days

PG 1 PG 2 PG 3

(analysis by Ja Yeon Jeong)







PCA Extensions for Data on Manifolds

- Fletcher (Principal Geodesic Anal.)
  - Best fit of geodesic to data
  - Constrained to go through geodesic mean

PCA Extensions for Data on Manifolds

- Fletcher (Principal Geodesic Anal.)
  - Best fit of geodesic to data
  - Constrained to go through geodesic mean

Counterexample:

Data on sphere, along equator

PCA Extensions for Data on Manifolds

- Fletcher (Principal Geodesic Anal.)
  - Best fit of geodesic to data
  - Constrained to go through geodesic mean
- Huckemann, Hotz & Munk (Geod. PCA)
  - Best fit of <u>any</u> geodesic to data

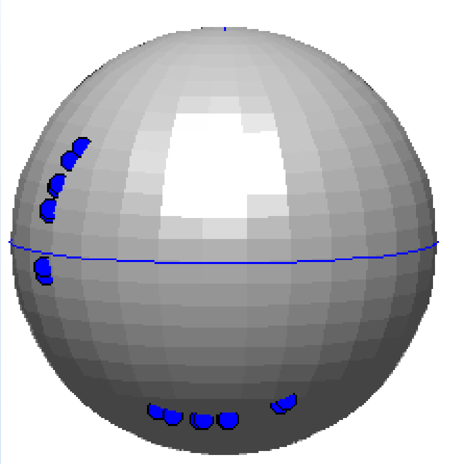


PCA Extensions for Data on Manifolds

- Fletcher (Principal Geodesic Anal.)
  - Best fit of geodesic to data
  - Constrained to go through geodesic mean
- Huckemann, Hotz & Munk (Geod. PCA)
  - Best fit of <u>any</u> geodesic to data

Counterexample:

Data follows Tropic of Capricorn



(thanks to Ja-Yeon Jeong)

PCA Extensions for Data on Manifolds

- Fletcher (Principal Geodesic Anal.)
  - Best fit of geodesic to data
  - Constrained to go through geodesic mean
- Huckemann, Hotz & Munk (Geod. PCA)
  - Best fit of any geodesic to data
- Jung, Foskey & Marron (Princ. Arc Anal.)
  - Best fit of any circle to data

    (motivated by conformal maps)

PCA Extensions for Data on Manifolds

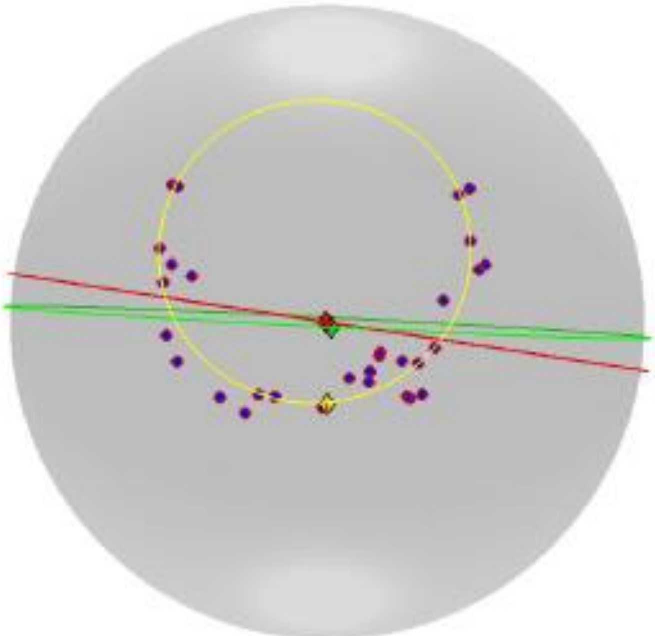
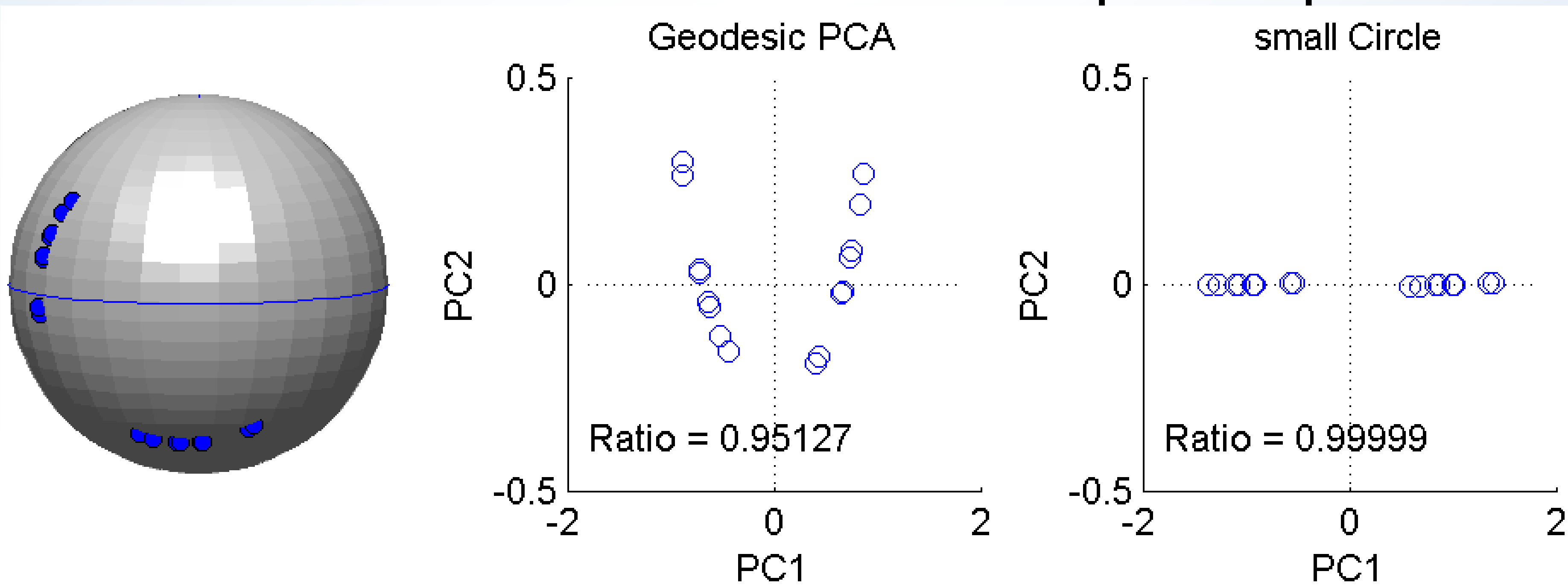


Figure: Generalization of PCA on S^2. Yellow: fitted small circle, Green: great circle found by Geodesic PCA (Huckemann), Red: great circle found by PGA (Fletcher). μ (PC mean, or geodesic mean) is depicted as yellow (green, or red, respectively) diamond.)

Principal Arc Analysis

Jung, Foskey & Marron

- Best fit of <u>any circle</u> to data
- Can give better fit than geodesics
- Observed for simulated m-rep example



Landmark Based Shape Analysis

Currently popular approaches to PCA on S^k:

- Early: PCA on projections
- Fletcher: Geodesics through mean
- Huckemann, et al: Any Geodesic

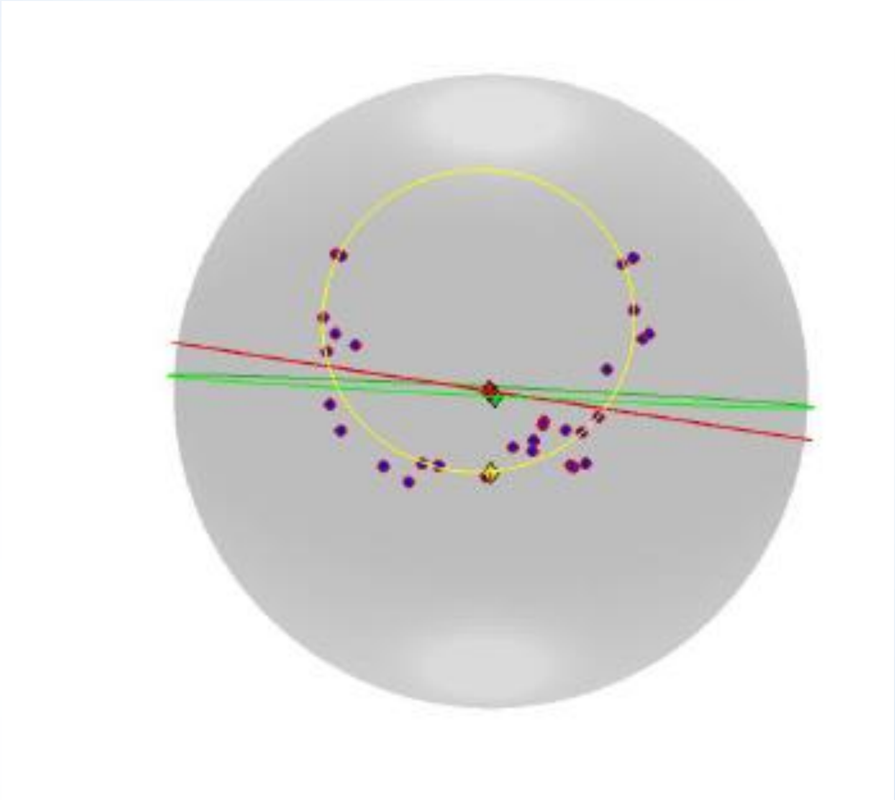
New Approach (Jung, Dryden, Marron):

Principal Nested Sphere Analysis

Principal Nested Spheres Analysis

Main Goal:

Extend Principal Arc Analysis (S^2 to S^k)



Principal Nested Spheres Analysis

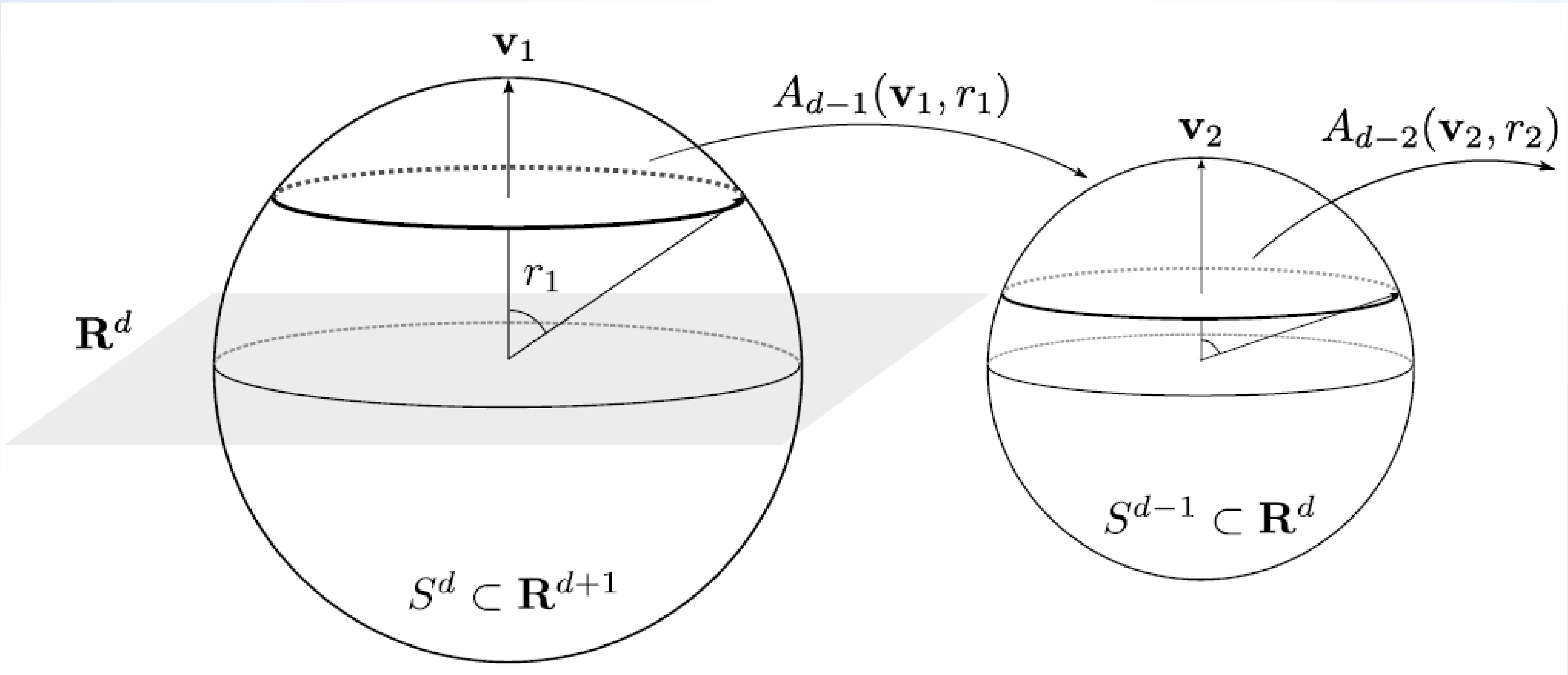
Main Goal:

Extend Principal Arc Analysis (S^2 to S^k)

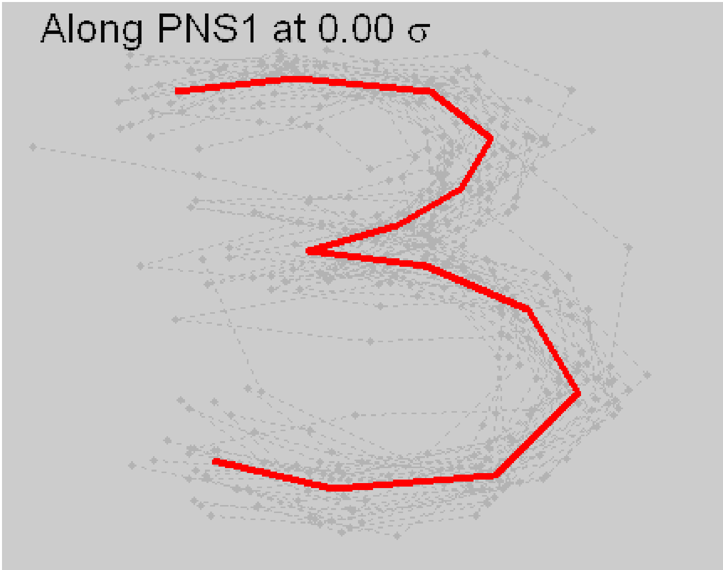
Jung, Dryden & Marron (2012)

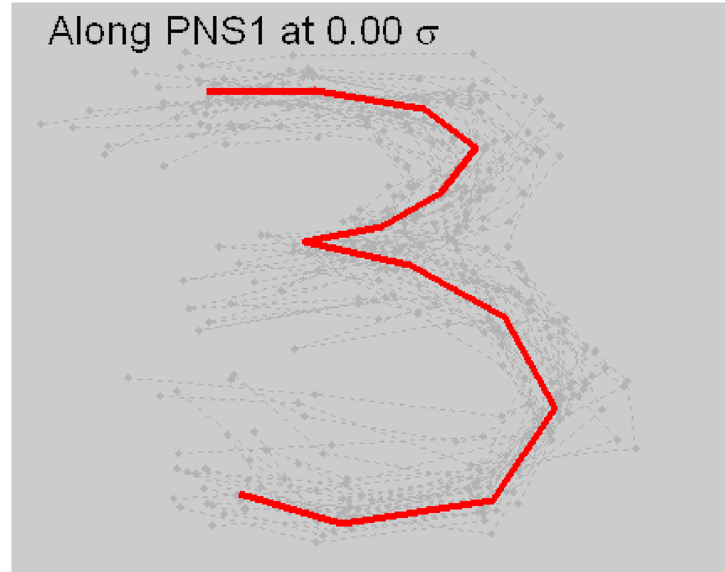
Principal Nested Spheres Analysis

Top Down Nested (small) spheres



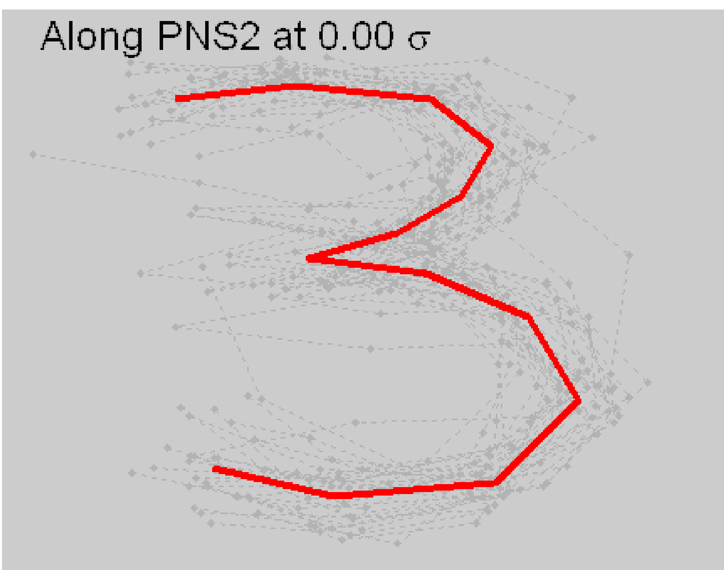
Digit 3 data: Principal variations of the shape

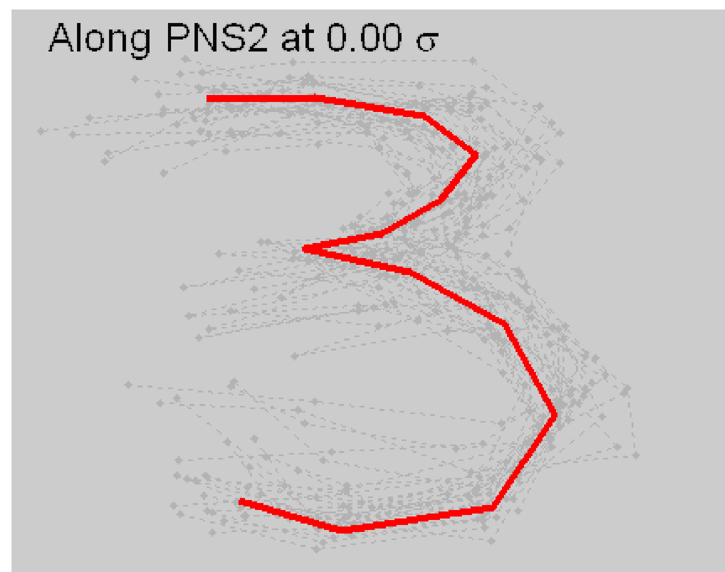




Princ. geodesics by PNS

Principal arcs by PNS





Principal Nested Spheres Analysis

Main Goal:

Extend Principal Arc Analysis (S^2 to S^k)

Jung, Dryden & Marron (2012)

Impact on Segmentation:

- PGA Segmentation: used ~20 comp's
- PNS Segmentation: only need ~13
- Resulted in visually better fits to data

Principal Nested Spheres Analysis

Main Goal:

Extend Principal Arc Analysis (S^2 to S^k)

Jung, Dryden & Marron (2012)

Important Landmark: This Motivated

Backwards PCA

Principal Nested Spheres Analysis

Key Idea:

Replace usual *forwards* view of PCA

With a *backwards* approach to PCA

Terminology

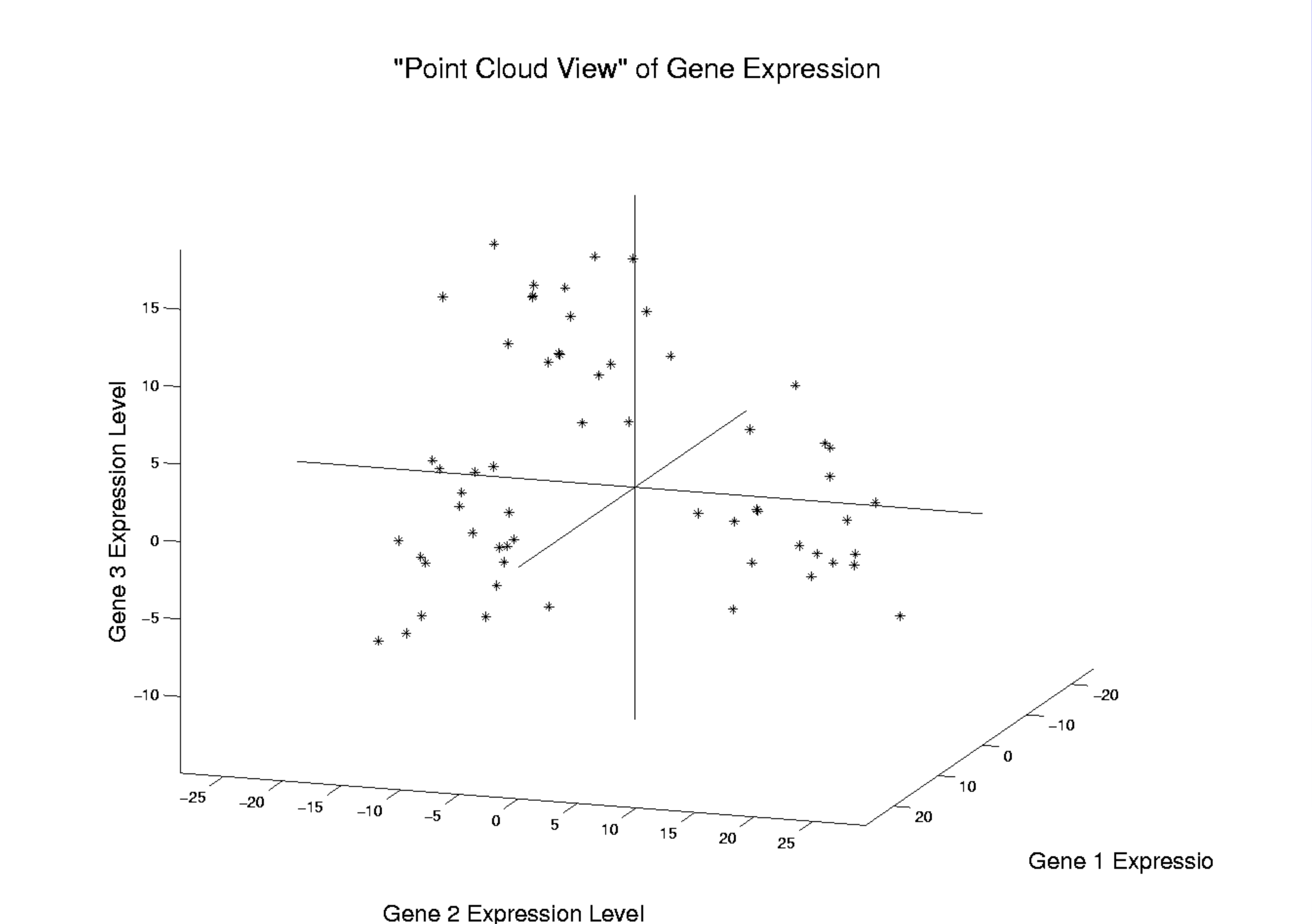
Multiple linear regression:

$$Y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_k x_{ik}$$

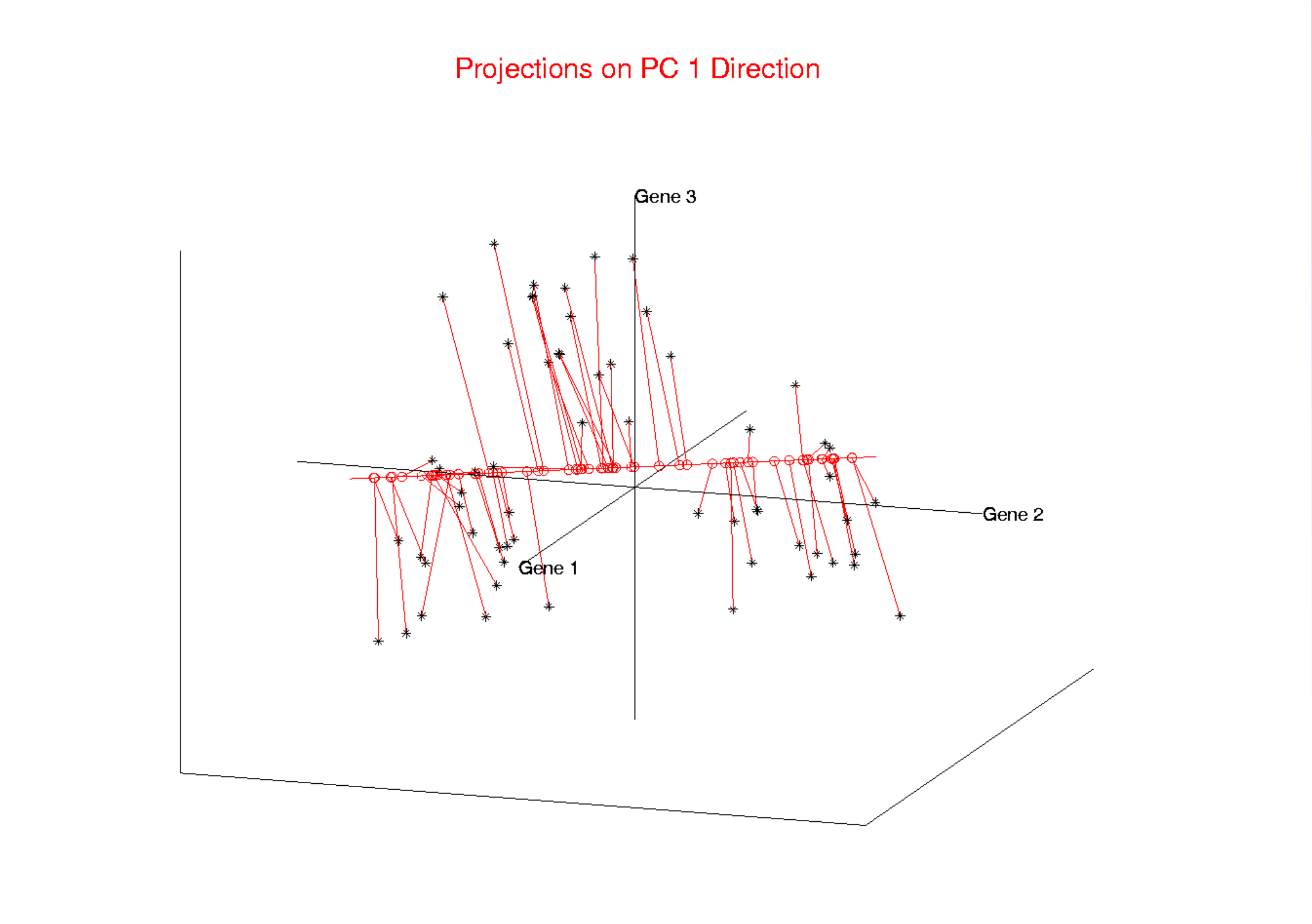
Stepwise approaches:

- *Forwards:* Start small, iteratively <u>add</u> variables to model
- *Backwards:* Start with all, iteratively <u>remove</u> variables from model

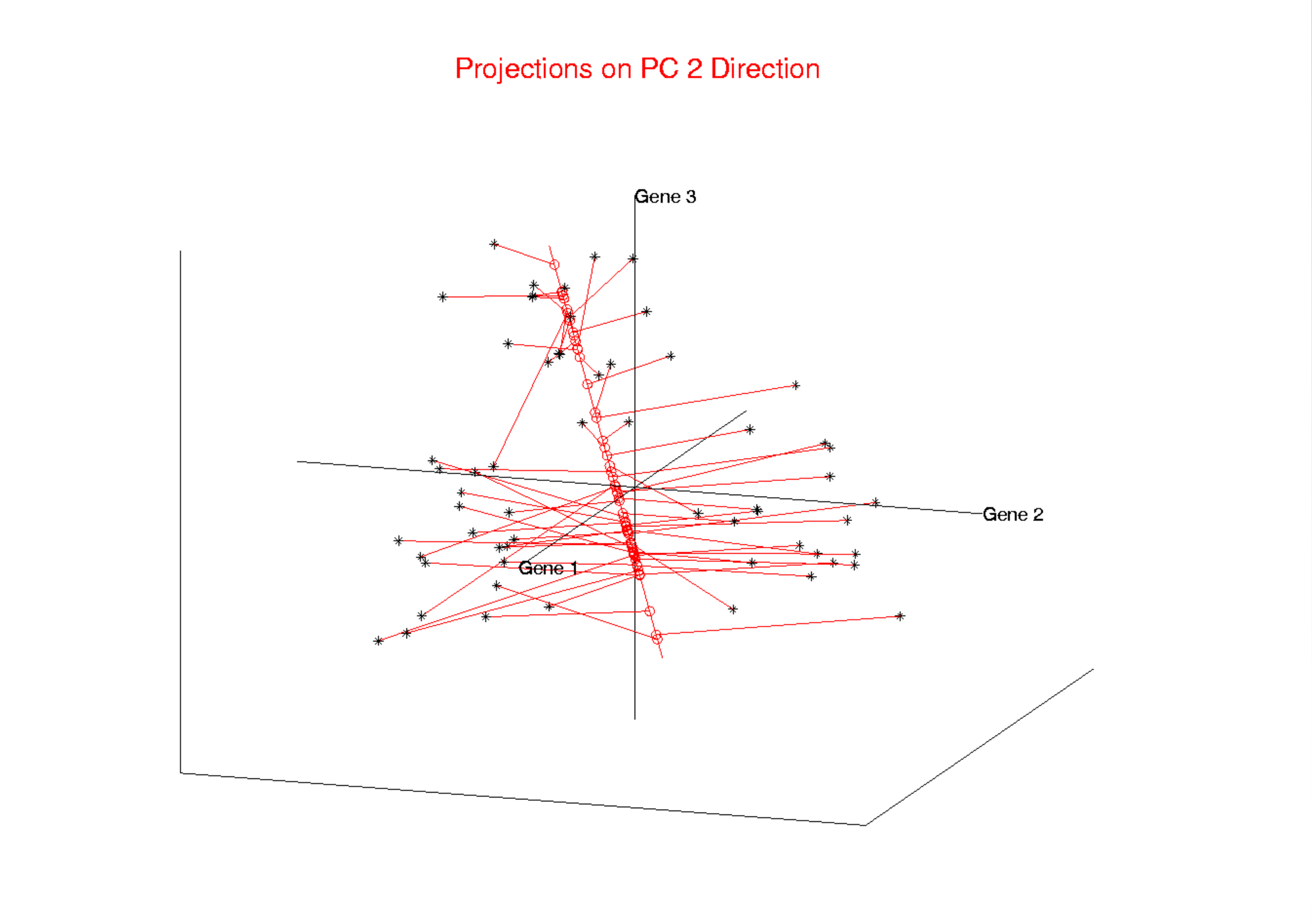
Illust'n of PCA View: Recall Raw Data



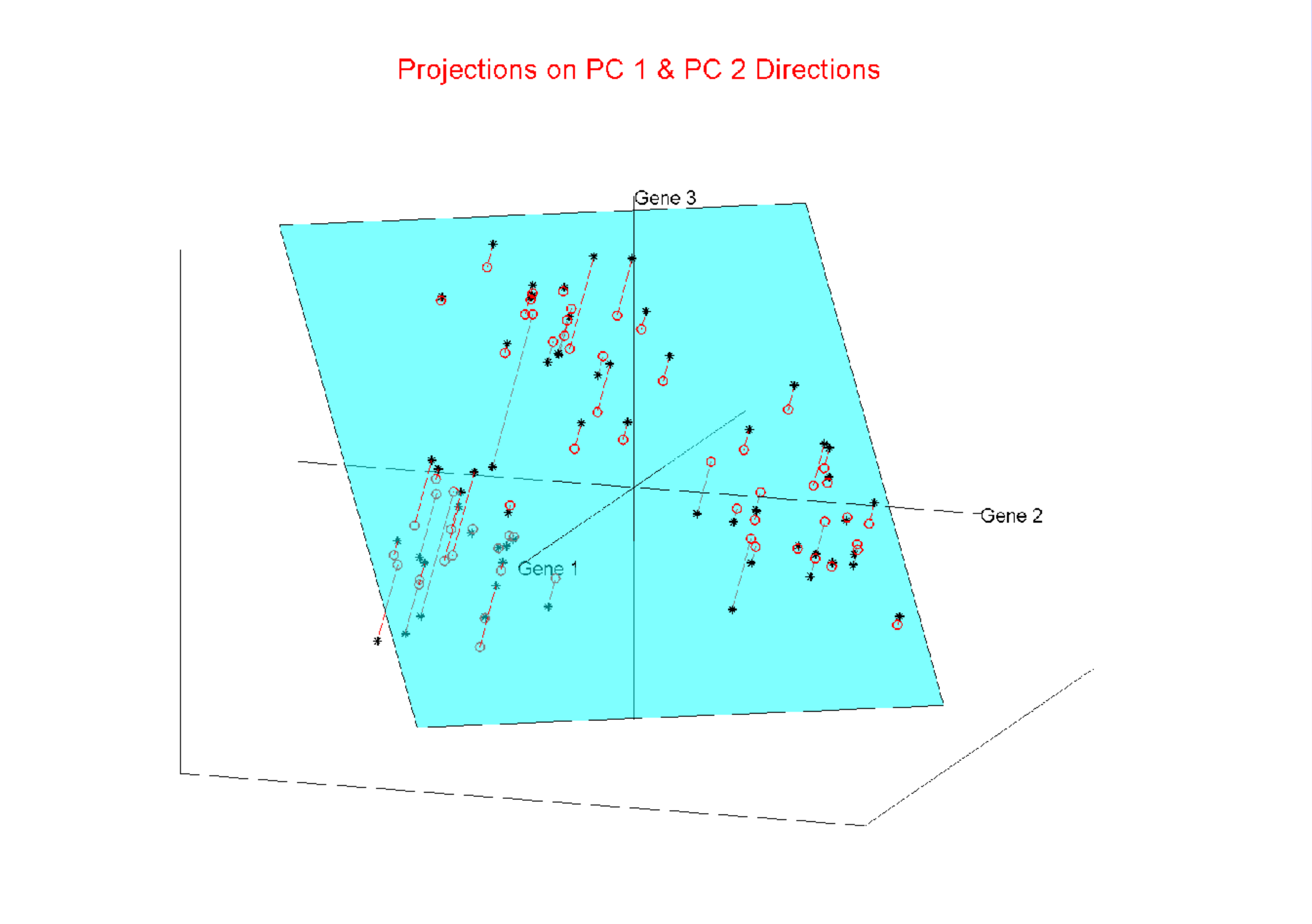
Illust'n of PCA View: PC1 Projections



Illust'n of PCA View: PC2 Projections



Illust'n of PCA View: Projections on PC1,2 plane



Principal Nested Spheres Analysis

Replace usual *forwards* view of PCA

Data → PC1 (1-d approx)

→ PC2 (1-d approx of Data-PC1)

→ PC1 U PC2 (2-d approx)

⋱

→ PC1 U … U PCr

(r-d approx)

With a *backwards* approach to PCA

Data → PC1 U … U PCr (r-d approx)

→ PC1 U … U PC(r-1)

. . .

→ PC1 U PC2 (2-d approx)

→ PC1 (1-d approx)

An Interesting Question

How generally applicable is

Backwards approach to PCA?

An Interesting Question

How generally applicable is

Backwards approach to PCA?

Where is this already being done???

An Interesting Question

How generally applicable is

Backwards approach to PCA?

Potential Application: **Principal Curves**

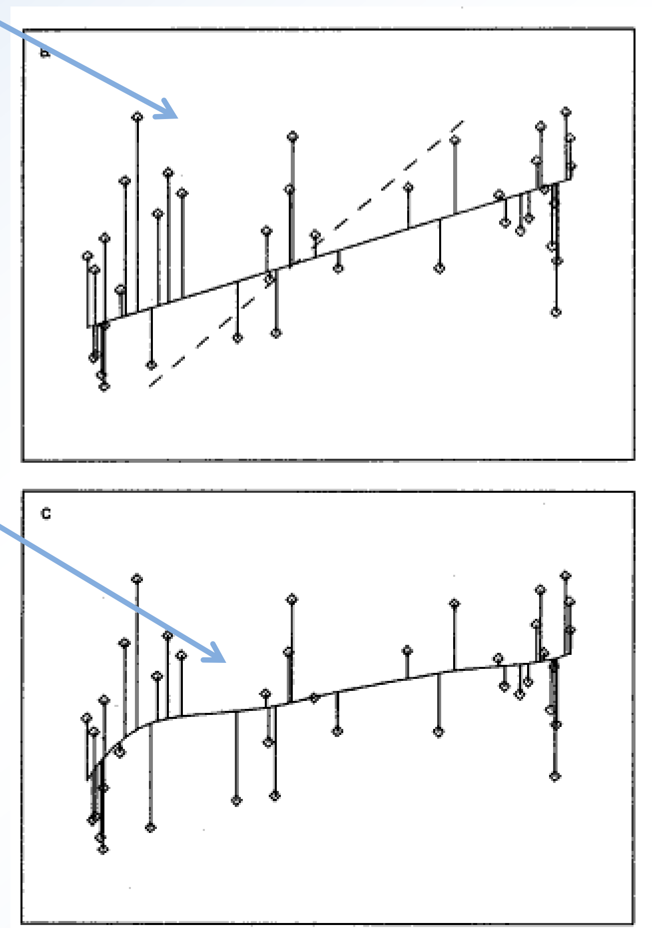
Hastie & Stuetzle, (1989) *JASA*

(Foundation of Manifold Learning)

1st Principal Curve

Linear Reg'n

Usual Smooth

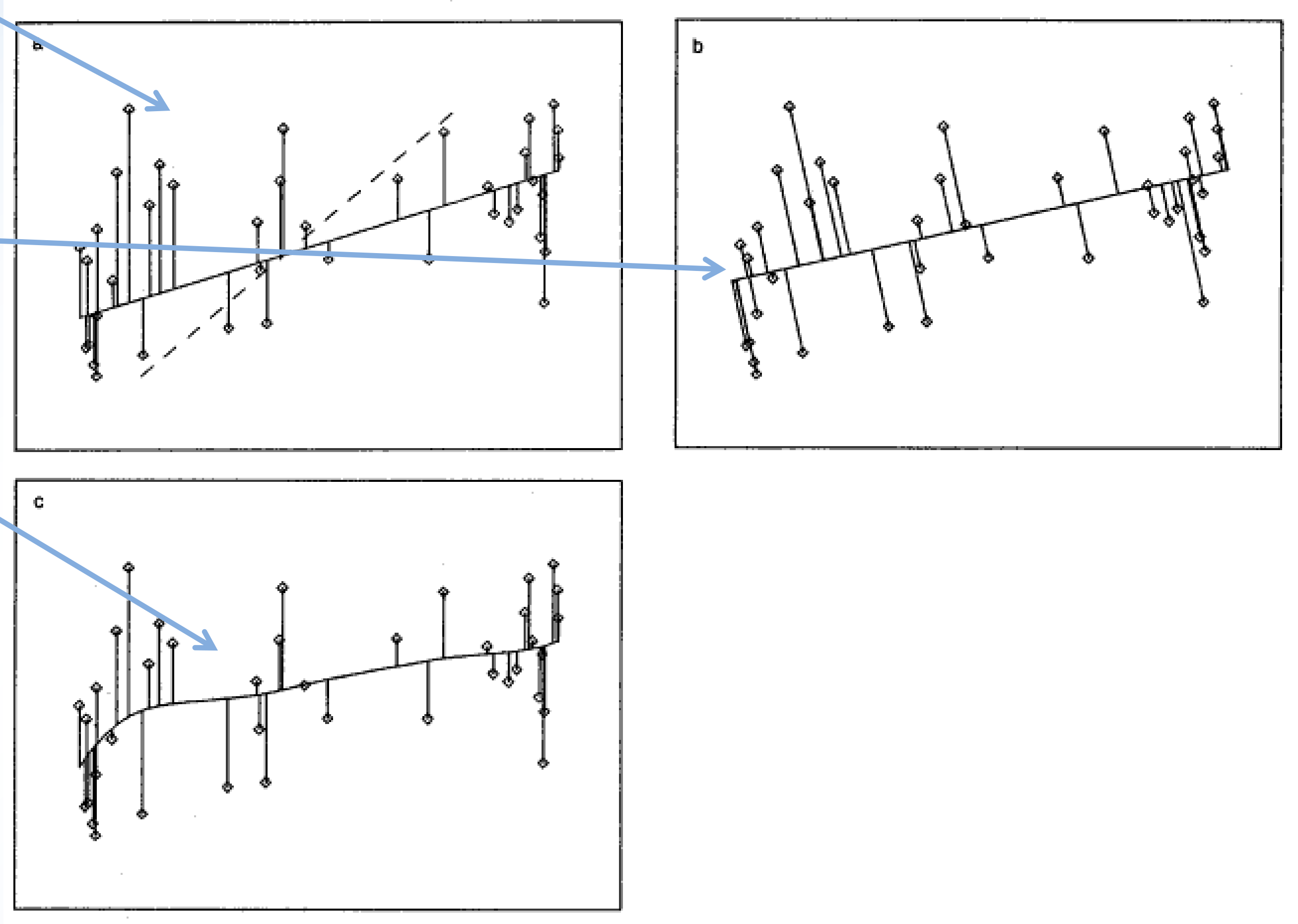


1st Principal Curve

Linear Reg'n

Proj's Reg'n

Usual Smooth



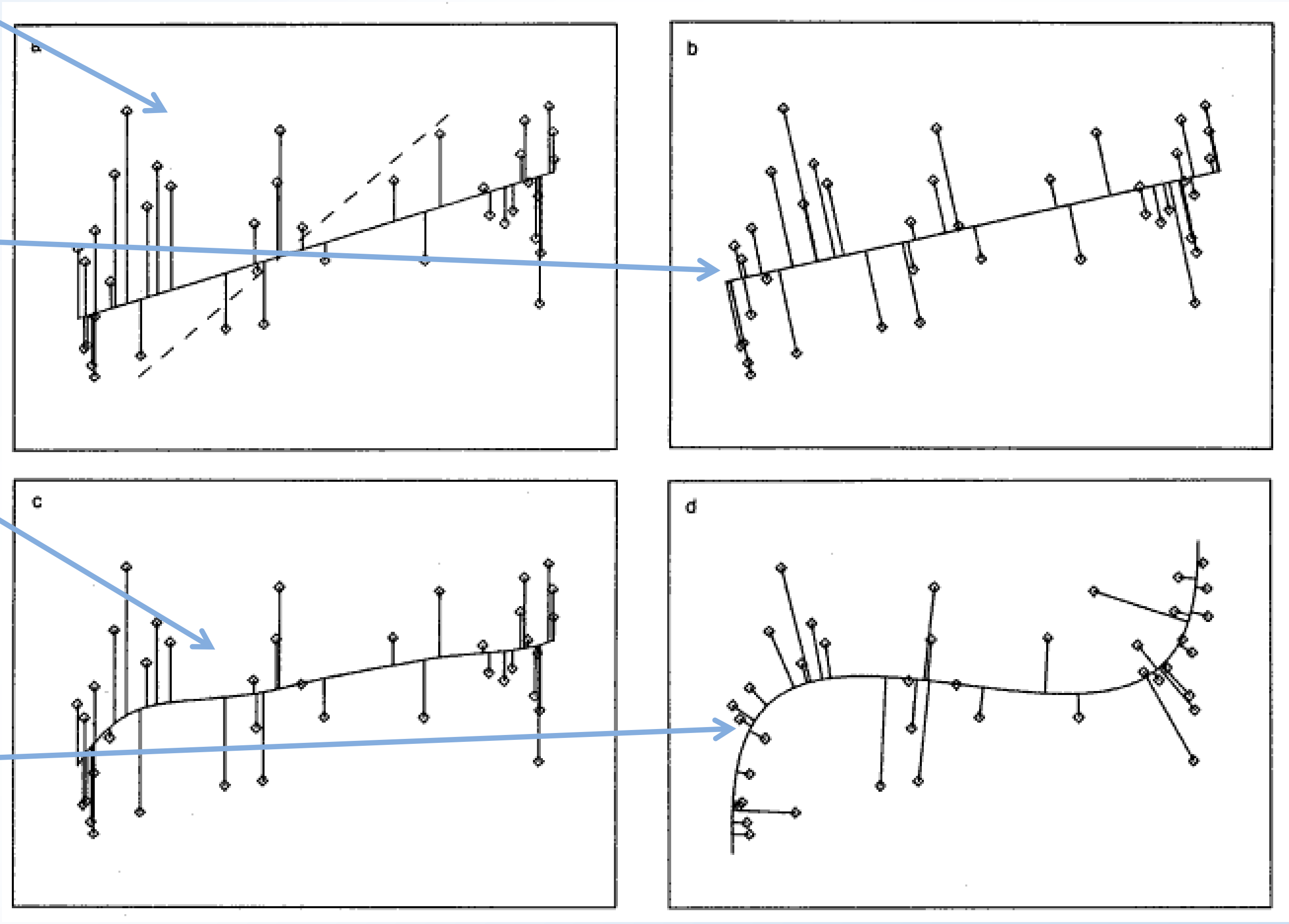
1st Principal Curve

Linear Reg'n

Proj's Reg'n

Usual Smooth

Princ'l Curve



An Interesting Question

How generally applicable is

Backwards approach to PCA?

Potential Application: **Principal Curves**

Perceived Major Challenge:

How to find <u>2^{nd} Principal Curve</u>?

Backwards approach???

An Interesting Question

How generally applicable is

Backwards approach to PCA?

Another Potential Application:

Nonnegative Matrix Factorization

= PCA in Positive Orthant

Current Approach, Lee et al (1999):

<u>Not Nested</u>, $(k = 3 \quad \not\approx \quad k = 4)$

An Interesting Question

How generally applicable is

Backwards approach to PCA?

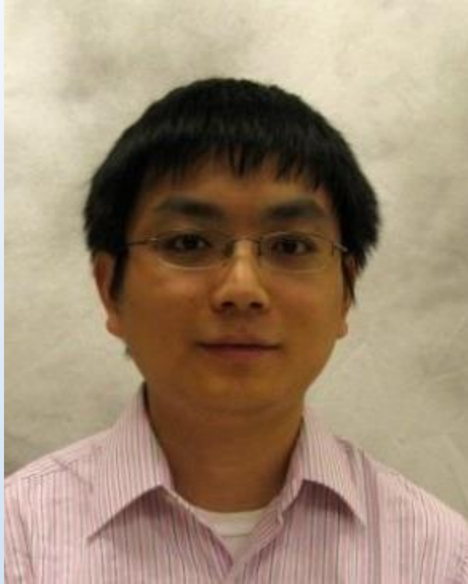
Another Potential Application:

Nonnegative Matrix Factorization

= PCA in Positive Orthant

(Backwards Nested Approach:

Lingsong Zhang)



An Interesting Question

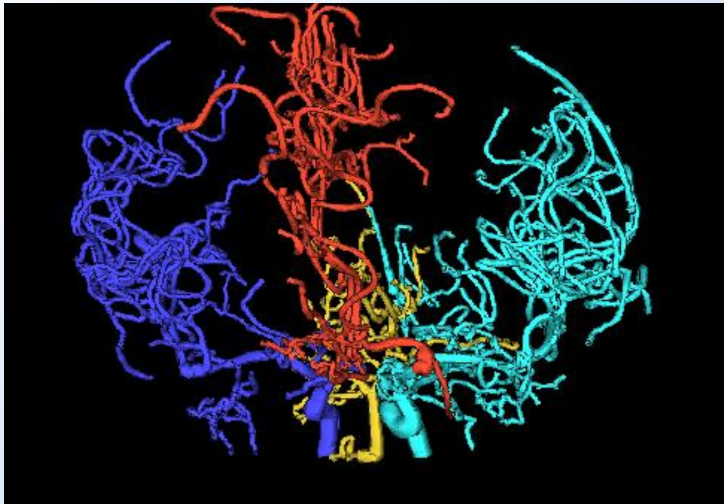
How generally applicable is

Backwards approach to PCA?

Another Potential Application:

Trees as Data

(early days)



An Interesting Question

How generally applicable is

Backwards approach to PCA?

An Attractive Answer

An Interesting Question

How generally applicable is

Backwards approach to PCA?

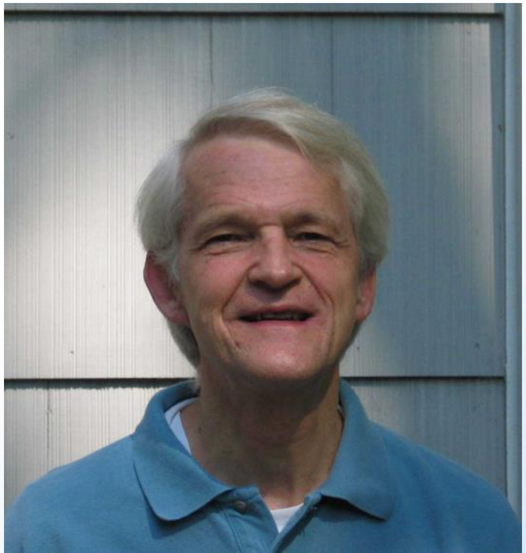
An Attractive Answer:

James Damon, UNC Mathematics

Geometry

Singularity

Theory



An Interesting Question

How generally applicable is

Backwards approach to PCA?

An Attractive Answer:

James Damon, UNC Mathematics

Key Idea: Express Backwards PCA as

<u>Nested Series of Constraints</u>

General View of Backwards PCA

Define *Nested Spaces* via Constraints

Satisfying More Constraints $\Rightarrow$

$\Rightarrow$ Smaller Subspaces

General View of Backwards PCA

Define *Nested Spaces* via Constraints

E.g. SVD

(Singular Value Decomposition =

= Not Mean Centered PCA)

(notationally very clean)

General View of Backwards PCA

Define *Nested Spaces* via <u>Constraints</u>

E.g. SVD

Have k Nested Subspaces:

$$S_1 \subseteq S_2 \subseteq \cdots \subseteq S_d$$

General View of Backwards PCA

Define *Nested Spaces* via <u>Constraints</u>

E.g. SVD $S_k = \left\{x : x = \sum_{j=1}^{k} c_j \overrightarrow{u_j}\right\}$

k-th SVD Subspace

Scores

Loading Vectors

General View of Backwards PCA

Define *Nested Spaces* via <u>Constraints</u>

E.g. SVD $S_k = \left\{x : x = \sum_{j=1}^{k} c_j \overrightarrow{u_j}\right\}$

Now Define:

$$S_{k-1} = \{x \in S_k : \langle x, \overrightarrow{u_k} \rangle = 0\}$$

General View of Backwards PCA

Define *Nested Spaces* via Constraints

E.g. SVD $S_k = \left\{ x : x = \sum_{j=1}^{k} c_j \overrightarrow{u_j} \right\}$

Now Define:

$$S_{k-1} = \left\{ x \in S_k : \langle x, \overrightarrow{u_k} \rangle = 0 \right\}$$

Constraint Gives *Nested* Reduction of Dim'n

General View of Backwards PCA

Define *Nested Spaces* via Constraints

- Backwards PCA

  Reduce Using Affine Constraints

General View of Backwards PCA

Define *Nested Spaces* via Constraints

- Backwards PCA
- Principal Nested Spheres

Use Affine Constraints (Planar Slices)

In Ambient Space

General View of Backwards PCA

Define *Nested Spaces* via Constraints

- Backwards PCA
- Principal Nested Spheres
- Principal Surfaces

    Spline Constraint Within Previous?

General View of Backwards PCA

Define *Nested Spaces* via Constraints

- Backwards PCA
- Principal Nested Spheres
- Principal Surfaces

Spline Constraint Within Previous?

{Been Done Already???}

General View of Backwards PCA

Define *Nested Spaces* via <u>Constraints</u>

- Backwards PCA
- Principal Nested Spheres
- Principal Surfaces
- Other Manifold Data Spaces

Sub-Manifold Constraints??

(Algebraic Geometry)

General View of Backwards PCA

Define *Nested Spaces* via Constraints

- Backwards PCA
- Principal Nested Spheres
- Principal Surfaces
- Other Manifold Data Spaces
- Tree Spaces

Suitable Constraints???

General View of Backwards PCA

Why does Backwards Work Better?

General View of Backwards PCA

Why does <u>Backwards</u> Work Better?

- Natural to *Sequentially Add Constraints*

(I.e. Add Constraints,

Using Information in Data)

General View of Backwards PCA

Why does Backwards Work Better?

- Natural to *Sequentially Add Constraints*
- Hard to *Start With Complete Set,*
  And Sequentially Remove

Carry Away Concept

OODA is more than a "framework"

It Provides a Focal Point

Highlights Pivotal Choices:

What should be the Data Objects?

How should they be Represented?