Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

# Parallel Coordinates-REVISITED 

Norm Matloff<br>University of California at Davis (new collaborator: Yingkang Xie)

Bay Area R Users Group
November 12, 2013

## Outline

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Outline

Norm Matloff
University of California at Davis
(new
collaborator: Yingkang Xie)

- What IS parallel coordinates, anyway?


# Outline 

Norm Matloff
University of California at Davis
(new
collaborator:
Yingkang Xie)

- What IS parallel coordinates, anyway?
- SEEMS to be a great tool.


## Outline

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

- What IS parallel coordinates, anyway?
- SEEMS to be a great tool. But has MAJOR problems.


## Outline

Norm Matloff University of California at Davis (new
collaborator: Yingkang Xie)

- What IS parallel coordinates, anyway?
- SEEMS to be a great tool. But has MAJOR problems.
- I will present a novel way to make parallel coordinates usable.


## What IS Parallel Coordinates?

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## What IS Parallel Coordinates?

Norm Matloff University of California at Davis
(new collaborator: Yingkang Xie)

- Attempt to view multidimensional data on 2-dimensiohal screen.


## What IS Parallel Coordinates?

Norm Matloff University of California at Davis
(new collaborator: Yingkang Xie)

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:


## What IS Parallel Coordinates?

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:
- Draw a vertical line for each variable ("parallel coords.").


## What IS Parallel Coordinates?

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:
- Draw a vertical line for each variable ("parallel coords.").
- For each data point, mark a dot on each vertical line, at the value of that variable for that data point.


## What IS Parallel Coordinates?

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:
- Draw a vertical line for each variable ("parallel coords.").
- For each data point, mark a dot on each vertical line, at the value of that variable for that data point.
- For each data point, "connect the dots."


## What IS Parallel Coordinates?

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:
- Draw a vertical line for each variable ("parallel coords.").
- For each data point, mark a dot on each vertical line, at the value of that variable for that data point.
- For each data point, "connect the dots."
- Resulting graph: a jagged line for each of your original data point.


## What IS Parallel Coordinates?

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:
- Draw a vertical line for each variable ("parallel coords.").
- For each data point, mark a dot on each vertical line, at the value of that variable for that data point.
- For each data point, "connect the dots."
- Resulting graph: a jagged line for each of your original data point.
- Can then try to find relations between variables by looking at line patterns.


## What IS Parallel Coordinates?

- Attempt to view multidimensional data on 2-dimensiohal screen.
- Simple idea:
- Draw a vertical line for each variable ("parallel coords.").
- For each data point, mark a dot on each vertical line, at the value of that variable for that data point.
- For each data point, "connect the dots."
- Resulting graph: a jagged line for each of your original data point.
- Can then try to find relations between variables by looking at line patterns.
- The operative word is "try."


## Example: R cars data

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## REVISITED

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

## Example: R cars data



Parallel CoordinatesREVISITED

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

## Example: R cars data



- Each jagged line is one car.


## Norm Matloff

 University of California at Davis (new collaborator: Yingkang Xie)
## Example: R cars data



- Each jagged line is one car.
- Vertical axes are the variables, Cyl, Disp, Hp, etc.


## Example: R cars data



- Each jagged line is one car.
- Vertical axes are the variables, Cyl, Disp, Hp, etc.
- ALREADY hard to interpret!


## Example: R cars data



- Each jagged line is one car.
- Vertical axes are the variables, Cyl, Disp, Hp, etc.
- ALREADY hard to interpret!
- Note: Variables are typically centered and scaled.


## Problems

## Problems

Norm Matloff
University of California at

Davis
(new
collaborator: Yingkang Xie)

## Hard to interpret, except in "small n, small p" data.

## Problems

Norm Matloff
University of California at Davis
(new
collaborator: Yingkang Xie)

Hard to interpret, except in "small n, small p" data. ( $\mathrm{p}=$ number of variables)

## Problems

Norm Matloff University of California at

Davis
(new
collaborator: Yingkang Xie)

Hard to interpret, except in "small n, small p" data.
( $\mathrm{p}=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

## Problems

Norm Matloff University of California at

Davis
(new
collaborator: Yingkang Xie)

Hard to interpret, except in "small n, small p" data.
( $\mathrm{p}=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:

## Problems

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Hard to interpret, except in "small n, small p" data.
( $\mathrm{p}=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:
Allow user to interactively do various permutations of the axes.

## Problems

Hard to interpret, except in "small n, small p" data.
( $p=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:
Allow user to interactively do various permutations of the axes.

Problem 2: Screen clutter!!!!

## Problems

Hard to interpret, except in "small n, small p" data.
( $p=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:
Allow user to interactively do various permutations of the axes.

Problem 2: Screen clutter!!!!
Typical solutions:

## Problems

Hard to interpret, except in "small n, small p" data.
( $p=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:
Allow user to interactively do various permutations of the axes.

Problem 2: Screen clutter!!!!
Typical solutions:

1. $\alpha$ blending (making pixels less dark).

## Problems

Hard to interpret, except in "small n, small p" data.
( $p=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:
Allow user to interactively do various permutations of the axes.

Problem 2: Screen clutter!!!!
Typical solutions:

1. $\alpha$ blending (making pixels less dark).
2. Plotting line density instead of lines (equiv. to tonal, like $\alpha$ ).

## Problems

Hard to interpret, except in "small n, small p" data.
( $p=$ number of variables)
Problem 1: Hard to see relation between "far apart" variables

Typical solution:
Allow user to interactively do various permutations of the axes.

Problem 2: Screen clutter!!!!
Typical solutions:

1. $\alpha$ blending (making pixels less dark).
2. Plotting line density instead of lines (equiv. to tonal, like $\alpha$ ).
3. Look at random subset of the data.

## Example of Clutter

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Example of Clutter

Norm Matloff University of California at

Davis
(new
collaborator: Yingkang Xie)

Example: Baseball Player data-height, weight, age (courtesy of UCLA Stat. Dept.)

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

## Example of Clutter

Example: Baseball Player data-height, weight, age (courtesy of UCLA Stat. Dept.)



## Another Example of Clutter

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Another Example of Clutter

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Example: Wine Quality data-various chemical measures (UCI Repository)

## Another Example of Clutter

Example: Wine Quality data—various chemical measures (UCI Repository)


## Alpha Blending May Not Help <br> Much

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Alpha Blending May Not Help

## Much

$\alpha$ blending may not help much:

## Norm Matloff

University of California at

## Davis

(new
collaborator: Yingkang Xie)

## Alpha Blending May Not Help Much

$\alpha$ blending may not help much:


## Yikes!

## Norm Matloff University of California at <br> Davis <br> (new <br> collaborator: <br> Yingkang Xie)

## Yikes!

## Norm Matloff University of California at <br> Davis <br> (new <br> collaborator: <br> Yingkang Xie)

## Comments:

## Yikes!

Norm Matloff University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Comments:

- Yikes!


## Yikes!

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Comments:

- Yikes!
- "Don't let the picture intimidate you!"-A. Inselberg, one of the pioneers of parallel coordinates, speaking in general of cluttered p.c. plots


## Yikes!

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Comments:

- Yikes!
- "Don't let the picture intimidate you!"-A. Inselberg, one of the pioneers of parallel coordinates, speaking in general of cluttered p.c. plots
- But it IS intimidating!


## Yikes!

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Comments:

- Yikes!
- "Don't let the picture intimidate you!"-A. Inselberg, one of the pioneers of parallel coordinates, speaking in general of cluttered p.c. plots
- But it IS intimidating!
- Can TRY to exploit geometric properties:


## Yikes!

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Comments:

- Yikes!
- "Don't let the picture intimidate you!"-A. Inselberg, one of the pioneers of parallel coordinates, speaking in general of cluttered p.c. plots
- But it IS intimidating!
- Can TRY to exploit geometric properties:
- $X$ shape $\Rightarrow$ negative $\rho$


## Yikes!

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Comments:

- Yikes!
- "Don't let the picture intimidate you!"-A. Inselberg, one of the pioneers of parallel coordinates, speaking in general of cluttered p.c. plots
- But it IS intimidating!
- Can TRY to exploit geometric properties:
- $X$ shape $\Rightarrow$ negative $\rho$
- $<$ shape $\Rightarrow$ positive $\rho$


## Yikes!

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

Comments:

- Yikes!
- "Don't let the picture intimidate you!"-A. Inselberg, one of the pioneers of parallel coordinates, speaking in general of cluttered p.c. plots
- But it IS intimidating!
- Can TRY to exploit geometric properties:
- $X$ shape $\Rightarrow$ negative $\rho$
- $<$ shape $\Rightarrow$ positive $\rho$
- Nice theory, form projective geometry, etc.


## Example of Clutter, cont'd.

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Example of Clutter, cont'd.

Norm Matloff
University of California at Davis
(new
collaborator: Yingkang Xie)

## Grouping by player position doesn't help much:

## Example of Clutter, cont'd.

## Grouping by player position doesn't help much:



Norm Matloff
University of California at

Davis
(new
collaborator: Yingkang Xie)

## My approach:

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

My approach: Plot only a few "typical" lines.

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

My approach: Plot only a few "typical" lines.

- "Typical" means highest estimated multivariate density.


## My Way

Norm Matloff University of California at Davis (new
collaborator: Yingkang Xie)

My approach: Plot only a few "typical" lines.

- "Typical" means highest estimated multivariate density.
- No clutter.


## My Way

Norm Matloff University of California at Davis (new
collaborator: Yingkang Xie)

My approach: Plot only a few "typical" lines.

- "Typical" means highest estimated multivariate density.
- No clutter.
- Far-apart variables problem ameliorated.


## My Way

Norm Matloff University of California at Davis (new

My approach: Plot only a few "typical" lines.

- "Typical" means highest estimated multivariate density.
- No clutter.
- Far-apart variables problem ameliorated.
- (Not related to parallel coordinate density plots.)


## My Way

Norm Matloff University of California at Davis (new

My approach: Plot only a few "typical" lines.

- "Typical" means highest estimated multivariate density.
- No clutter.
- Far-apart variables problem ameliorated.
- (Not related to parallel coordinate density plots.)


## Baseball Data, My Way

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

Parallel CoordinatesREVISITED

Norm Matloff University of California at

## Davis

(new
collaborator: Yingkang Xie)

## Baseball Data, My Way



## Baseball Data, My Way

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

- "The monkeys stand for honesty, giraffes are insincere, elephants are kindly but they're dumb"-Simon \& Garfunkel song


## Baseball Data, My Way

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

- "The monkeys stand for honesty, giraffes are insincere, elephants are kindly but they're dumb"-Simon \& Garfunkel song
- Pitchers are typically tall, thin, young.


## Baseball Data, My Way

- "The monkeys stand for honesty, giraffes are insincere, elephants are kindly but they're dumb"-Simon \& Garfunkel song
- Pitchers are typically tall, thin, young.
- Catchers typically are much heavier, older.


## Baseball Data, My Way



- "The monkeys stand for honesty, giraffes are insincere, elephants are kindly but they're dumb"-Simon \& Garfunkel song
- Pitchers are typically tall, thin, young.
- Catchers typically are much heavier, older.
- Infielders typically shorter, thinner, younger.


## Example: UCI Wine Quality Data

Norm Matloff University of California at

## Davis

(new
collaborator: Yingkang Xie)

## Example: UCI Wine Quality Data

Variables are chemical measures, plus a quality variable.

## Example: UCI Wine Quality Data

Variables are chemical measures, plus a quality variable. I broke quality into 3 broad ranges.

## Example: UCI Wine Quality Data

Variables are chemical measures, plus a quality variable. I broke quality into 3 broad ranges.


## Example: UCI Wine Quality Data

Variables are chemical measures, plus a quality variable. I broke quality into 3 broad ranges.


Need domain expert to intrepret, but distinction between quality groups is clear.

## How Many to Plot?

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

Norm Matloff University of California at Davis
(new
collaborator: Yingkang Xie)

## How Many to Plot?

## We plot "a few of the most typical lines."

## How Many to Plot?

We plot "a few of the most typical lines."

- My software asks the user to define "a few," in an argument m .

Norm Matloff University of California at Davis
(new
collaborator:

## How Many to Plot?

We plot "a few of the most typical lines."

- My software asks the user to define "a few," in an argument m.
- Above plots had $\mathrm{m}=1$.


## How Many to Plot?

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

We plot "a few of the most typical lines."

- My software asks the user to define "a few," in an argument m .
- Above plots had $m=1$.
- m smaller $\Rightarrow$ higher "typicalness"


## How Many to Plot?

Norm Matloff University of California at Davis (new collaborator: Yingkang Xie)

## More Features

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## More Features

Norm Matloff
University of California at Davis (new
collaborator: Yingkang Xie)

## More features of the software (currently experimental):

## More Features

Norm Matloff
University of California at Davis (new
collaborator: Yingkang Xie)

# More features of the software (currently experimental): 

- Cluster hunting.


## More Features

Norm Matloff
University of California at Davis (new

More features of the software (currently experimental):

- Cluster hunting. Instead of finding global max of density, plot jagged line at each local maximum


## More Features

Norm Matloff University of California at Davis (new

More features of the software (currently experimental):

- Cluster hunting. Instead of finding global max of density, plot jagged line at each local maximum
- Outlier hunting.


## More Features

Norm Matloff University of California at Davis (new
collaborator: Yingkang Xie)

More features of the software (currently experimental):

- Cluster hunting. Instead of finding global max of density, plot jagged line at each local maximum
- Outlier hunting. Plot the least typical lines.


## Computation

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Computation

Norm Matloff
University of California at Davis
(new
collaborator:
Yingkang Xie)

- R package available at http://heather.cs.ucdavis.edu/bdgraphs.html

Norm Matloff
University of California at

Davis
(new
collaborator:
Yingkang Xie)

## Computation

- R package available at http://heather.cs.ucdavis.edu/bdgraphs.html
- Use k-NN density estimation.


## Computation

Norm Matloff University of California at Davis (new
collaborator: Yingkang Xie)

- R package available at http://heather.cs.ucdavis.edu/bdgraphs.html
- Use k-NN density estimation.
- Use R's FNN ("fast nearest neighbor") library for some speed.


## Computation

- R package available at http://heather.cs.ucdavis.edu/bdgraphs.html
- Use k-NN density estimation.
- Use R's FNN ("fast nearest neighbor") library for some speed.
- Use parallel computing for a lot more speed.

