Parallel
Coordinates—
REVISITED

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

Parallel Coordinates—REVISITED

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

Bay Area R Users Group
November 12, 2013

Parallel
Coordinates—
REVISITED

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

Outline

Parallel
Coordinates—
REVISITED

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

Outline

- What IS parallel coordinates, anyway?

Parallel
Coordinates—
REVISITED

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

Outline

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

Parallel
Coordinates—
REVISITED

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

Outline

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

Parallel
Coordinates—
REVISITED

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

Outline

- 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.

Parallel
Coordinates—
REVISITED

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

What IS Parallel Coordinates?

Parallel
Coordinates—
REVISITED

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

What IS Parallel Coordinates?

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

Parallel
Coordinates—
REVISITED

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

What IS Parallel Coordinates?

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

- Simple idea:

Parallel
Coordinates—
REVISITED

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

What IS Parallel Coordinates?

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

- Simple idea:

    - Draw a vertical line for each variable ("parallel coords.").

Parallel
Coordinates—
REVISITED

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

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.

Parallel
Coordinates—
REVISITED

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

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."

Parallel
Coordinates—
REVISITED

Norm Matloff
University of
California at
Davis
(new
collaborator:
Yingkang Xie)
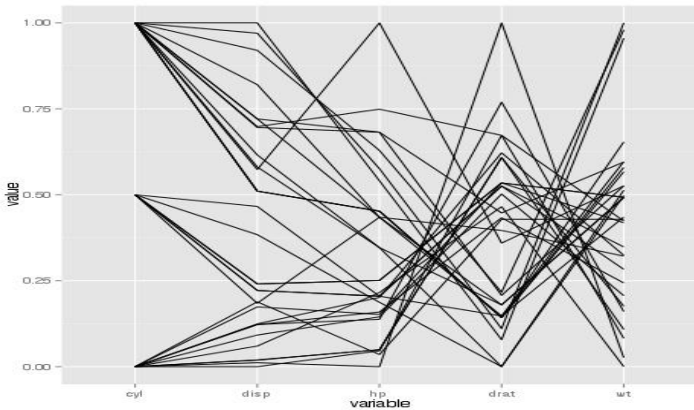
What IS Parallel Coordinates?

- 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.

Parallel
Coordinates—
REVISITED

Norm Matloff
University of
California at
Davis
(new
collaborator:
Yingkang Xie)
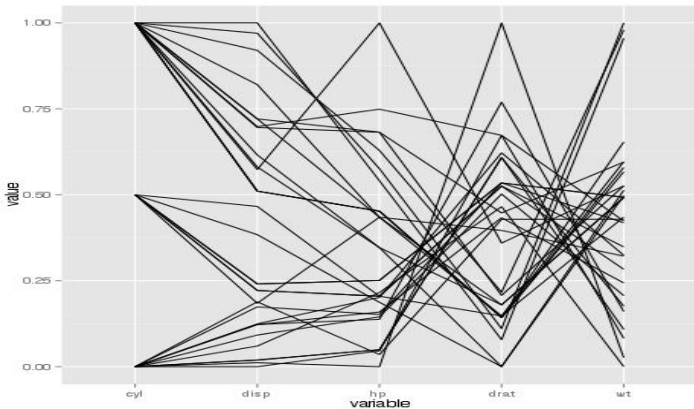
What IS Parallel Coordinates?

- 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.

Parallel
Coordinates—
REVISITED

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

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."

Parallel
Coordinates—
REVISITED

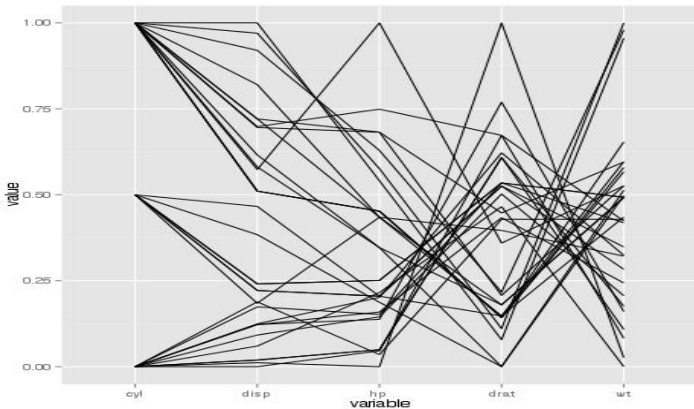Norm Matloff
University of
California at
Davis
(new
collaborator:
Yingkang Xie)

Example: R cars data

Parallel
Coordinates—
REVISITED

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

Example: R cars data

Parallel
Coordinates—
REVISITED

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

Example: R cars data

- Each jagged line is one car.

Parallel
Coordinates—
REVISITED
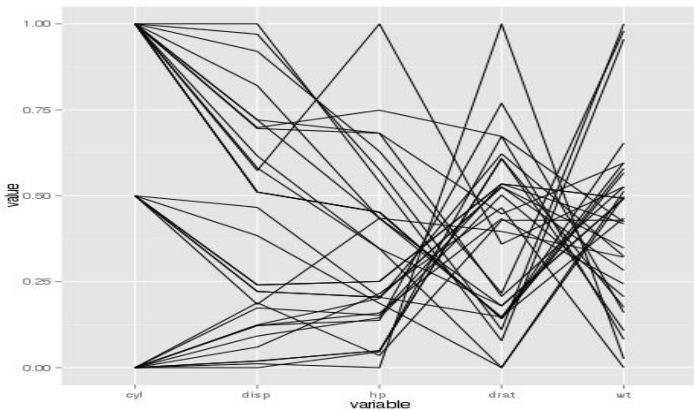
Norm Matloff
University of
California at
Davis
(new
collaborator:
Yingkang Xie)

Example: R cars data

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

Parallel
Coordinates—
REVISITED

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.
- ALREADY hard to interpret!

Parallel
Coordinates—
REVISITED

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

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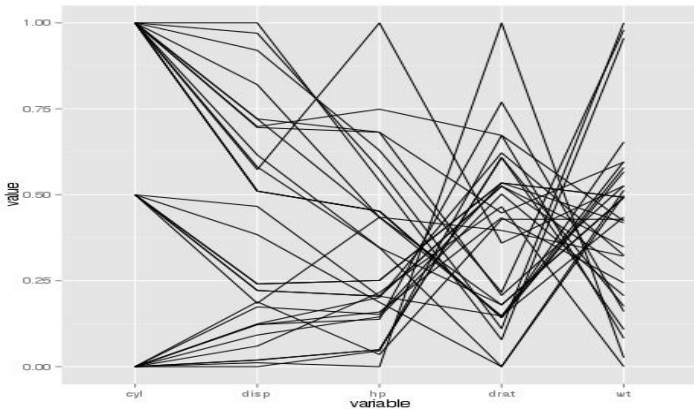
- Each jagged line is one car.
- Vertical axes are the variables, Cyl, Disp, Hp, etc.
- ALREADY hard to interpret!
- Note: Variables are typically centered and scaled.

Parallel
Coordinates—
REVISITED

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

Problems

Parallel
Coordinates—
REVISITED

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

Problems

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

Parallel
Coordinates—
REVISITED

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

Problems

Hard to interpret, except in "small n, small p" data.
(p = number of variables)

Parallel
Coordinates—
REVISITED

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

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**

Parallel
Coordinates—
REVISITED

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

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:

Parallel
Coordinates—
REVISITED

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

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.

Parallel
Coordinates—
REVISITED

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

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!!!!

Parallel
Coordinates—
REVISITED

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

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:

Parallel
Coordinates—
REVISITED

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

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. α blending (making pixels less dark).

Parallel
Coordinates—
REVISITED

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

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. α blending (making pixels less dark).
2. Plotting line density instead of lines (equiv. to α).

Parallel
Coordinates—
REVISITED

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

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. α blending (making pixels less dark).
2. Plotting line density instead of lines (equiv. to α).
3. Look at random subset of the data.

Parallel
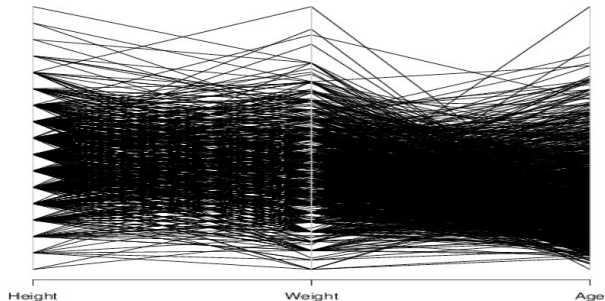Coordinates—
REVISITED

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

Example of Clutter

Parallel
Coordinates—
REVISITED

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.)

Parallel
Coordinates—
REVISITED

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.)

Parallel
Coordinates—
REVISITED

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

Another Example of Clutter

Parallel
Coordinates—
REVISITED

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

Another Example of Clutter

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

Parallel
Coordinates—
REVISITED

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

Another Example of Clutter

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

Parallel
Coordinates—
REVISITED

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

Alpha Blending May Not Help

Parallel
Coordinates—
REVISITED

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

Alpha Blending May Not Help

α blending may not help:

Parallel
Coordinates—
REVISITED

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

Alpha Blending May Not Help

α blending may not help:

Parallel
Coordinates—
REVISITED

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

Yikes!

Parallel
Coordinates—
REVISITED

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

Yikes!

Comments:

- Yikes!

Parallel
Coordinates—
REVISITED

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

Yikes!

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

Parallel
Coordinates—
REVISITED

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

Yikes!

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!

Parallel
Coordinates—
REVISITED

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

Yikes!

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:

Parallel
Coordinates—
REVISITED

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

Yikes!

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 ρ

Parallel
Coordinates—
REVISITED

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

Yikes!

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 ρ
  - $<$ shape $\Rightarrow$ positive ρ

Parallel
Coordinates—
REVISITED

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

Yikes!

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 ρ
    - $<$ shape $\Rightarrow$ positive ρ
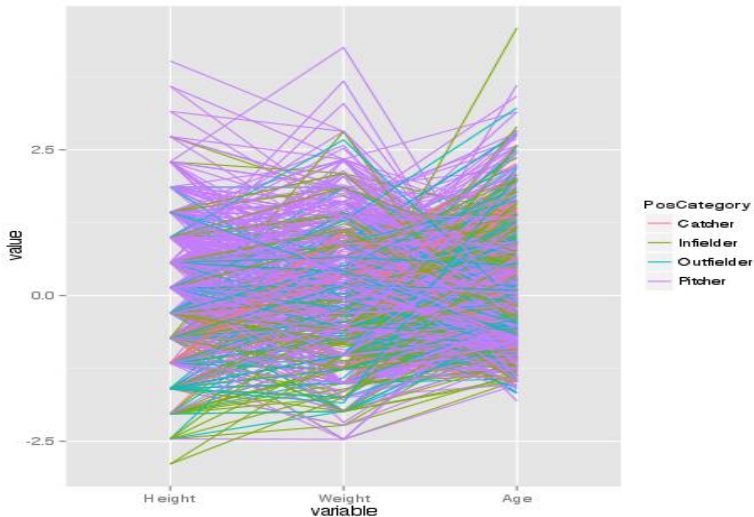   - Nice theory, form projective geometry, etc.

Parallel
Coordinates—
REVISITED

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

Example of Clutter, cont'd.

Parallel
Coordinates—
REVISITED

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

Example of Clutter, cont'd.

Grouping by player position doesn't help much:

Parallel
Coordinates—
REVISITED

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

Example of Clutter, cont'd.

Grouping by player position doesn't help much:

Parallel
Coordinates—
REVISITED

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

My Way

Parallel
Coordinates—
REVISITED

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

My Way

My approach:

Parallel
Coordinates—
REVISITED

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

My Way

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

Parallel
Coordinates—
REVISITED

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

My Way

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

- "Typical" means highest estimated multivariate density.

Parallel
Coordinates—
REVISITED

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

My Way

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

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

Parallel
Coordinates—
REVISITED

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

My Way

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

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

Parallel
Coordinates—
REVISITED

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

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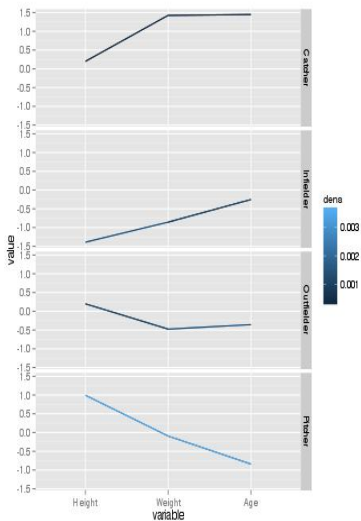
My approach: **Plot only a few "typical" lines.**

- "Typical" means highest estimated multivariate density.

- No clutter.

- Far-apart variables problem ameliorated.

- (Not related to *parallel coordinate density plots*.)

Parallel
Coordinates—
REVISITED

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

My Way

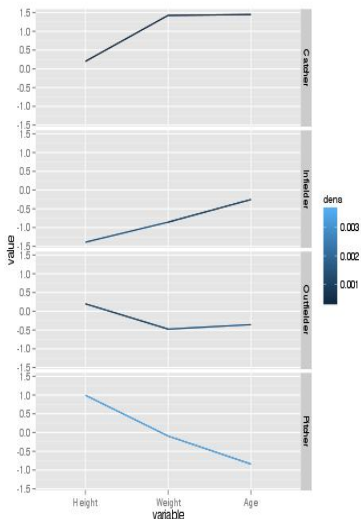My approach: **Plot only a few "typical" lines.**

- "Typical" means highest estimated multivariate density.

- No clutter.

- Far-apart variables problem ameliorated.

- (Not related to *parallel coordinate density plots.*)

Parallel
Coordinates—
REVISITED

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

Baseball Data, My Way

Parallel
Coordinates—
REVISITED

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

Baseball Data, My Way

Parallel
Coordinates—
REVISITED

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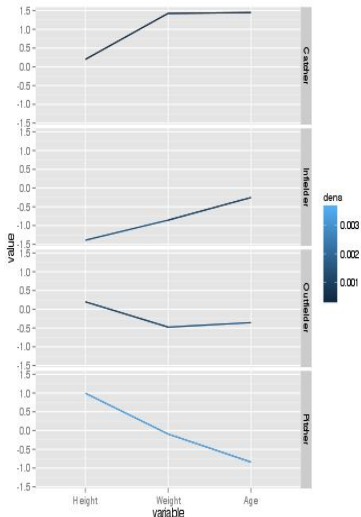Davis
(new
collaborator:
Yingkang Xie)

Baseball Data, My Way

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

Parallel
Coordinates—
REVISITED

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

Baseball Data, My Way

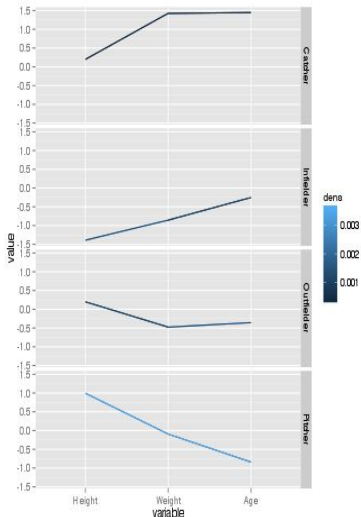- "The monkeys stand for honesty, giraffes are insincere, elephants are kindly but they're dumb"—Simon & Garfunkel song

- Pitchers are typically tall, thin, young.

Parallel
Coordinates—
REVISITED

Norm Matloff
University of
California at
Davis
(new
collaborator:
Yingkang Xie)
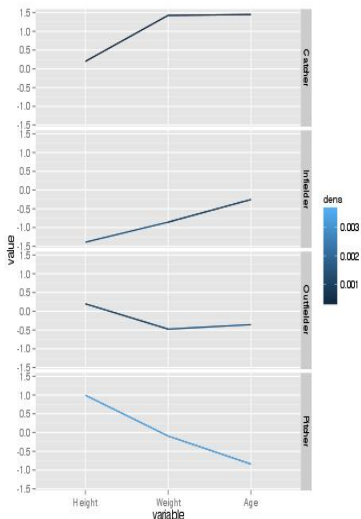
Baseball Data, My Way

- "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.

Parallel
Coordinates—
REVISITED

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

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.

Parallel
Coordinates—
REVISITED

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

Example: UCI Wine Quality Data

Parallel
Coordinates—
REVISITED

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

Example: UCI Wine Quality Data

Variables are chemical measures, plus a quality variable.

Parallel
Coordinates—
REVISITED

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

Example: UCI Wine Quality Data

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

Parallel
Coordinates—
REVISITED

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

Example: UCI Wine Quality Data

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

Parallel
Coordinates—
REVISITED

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

Example: UCI Wine Quality Data

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

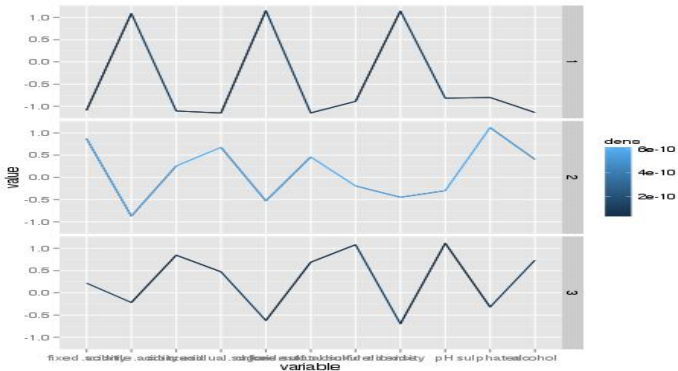Need domain expert to intrepret, but distinction between
quality groups is clear.

Parallel
Coordinates—
REVISITED

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

How Many to Plot?

Parallel
Coordinates—
REVISITED

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

How Many to Plot?

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

Parallel
Coordinates—
REVISITED

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

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.

Parallel
Coordinates—
REVISITED

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

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 m = 1.

Parallel
Coordinates—
REVISITED

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

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 m $= 1$.
- m smaller $\Rightarrow$ higher "typicalness"

Parallel
Coordinates—
REVISITED

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

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 m = 1.
- m smaller ⇒ higher "typicalness"
- m larger ⇒ better chance to see ρ's among the most typical data points

Parallel
Coordinates—
REVISITED

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

More Features

Parallel
Coordinates—
REVISITED

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

More Features

More features of the software (currently experimental):

Parallel
Coordinates—
REVISITED

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

More Features

More features of the software (currently experimental):

- Cluster hunting.

Parallel
Coordinates—
REVISITED

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

More Features

More features of the software (currently experimental):

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

Parallel
Coordinates—
REVISITED

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

More Features

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.

Parallel
Coordinates—
REVISITED

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

More Features

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 <u>least</u> typical lines.

Parallel
Coordinates—
REVISITED

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

Computation

Parallel
Coordinates—
REVISITED

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

Computation

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

Parallel
Coordinates—
REVISITED

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.

Parallel
Coordinates—
REVISITED

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.

- Use R's FNN ("fast nearest neighbor") library for some
  speed.

Parallel
Coordinates—
REVISITED

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.
- Use R's FNN ("fast nearest neighbor") library for some
  speed.
- Use parallel computing for a lot more speed.