# *Information Visualization*

## Multi-Dimensional Data Visualization

Nadia BOUKHELIFA
nadia.boukhelifa@telecom-paristech.fr


# *after today you will…*

- have learned techniques for MD data visualization
- have learned relationships between MD visualization techniques
- have understood the problems of MD data visualization

# *WHAT IS MULTIDIMENSIONAL DATA?*

# *Data sets*

- Data comes in many different forms
- Typically, not in the way you want it

- **Heterogeneous** data often seen as multiple dimensions of elements extracted by *different* patterns or needs.

# *A data set!*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 164 | dodge-st.-reg | 16.0 | 8. | 400.0 | 170.0 | 4668. | 11.5 | 75. | 1. |
| 165 | fiat-124-spor | 15.0 | 8. | 350.0 | 145.0 | 4440. | 14.0 | 75. | 1. |
| 166 | fiat-124-tc | 16.0 | 8. | 318.0 | 150.0 | 4498. | 14.5 | 75. | 1. |
| 167 | fiat-124b | 14.0 | 8. | 351.0 | 148.0 | 4657. | 13.5 | 75. | 1. |
| 168 | fiat-128 | 17.0 | 6. | 231.0 | 110.0 | 3907. | 21.0 | 75. | 1. |
| 169 | fiat-128 | 16.0 | 6. | 250.0 | 105.0 | 3897. | 18.5 | 75. | 1. |
| 170 | fiat-131 | 15.0 | 6. | 258.0 | 110.0 | 3730. | 19.0 | 75. | 1. |
| 171 | fiat-strada-cu | 18.0 | 6. | 225.0 | 95.00 | 3785. | 19.0 | 75. | 1. |
| 172 | fiat-x1.9 | 21.0 | 6. | 231.0 | 110.0 | 3039. | 15.0 | 75. | 1. |
| 173 | ford-country | 20.0 | 8. | 262.0 | 110.0 | 3221. | 13.5 | 75. | 1. |
| 174 | ford-country | 13.0 | 8. | 302.0 | 129.0 | 3169. | 12.0 | 75. | 1. |
| 175 | ford-country | 29.0 | 4. | 97.00 | 75.00 | 2171. | 16.0 | 75. | 3. |
| 176 | ford-escort-2 | 23.0 | 4. | 140.0 | 83.00 | 2639. | 17.0 | 75. | 1. |
| 177 | ford-escort-4 | 20.0 | 6. | 232.0 | 100.0 | 2914. | 16.0 | 75. | 1. |
| 178 | ford-f108 | 23.0 | 4. | 140.0 | 78.00 | 2592. | 18.5 | 75. | 1. |
| 179 | ford-f250 | 24.0 | 4. | 134.0 | 96.00 | 2702. | 13.5 | 75. | 3. |
| 180 | ford-fairmon | 25.0 | 4. | 90.00 | 71.00 | 2223. | 16.5 | 75. | 2. |
| 181 | ford-fairmon | 24.0 | 4. | 119.0 | 97.00 | 2545. | 17.0 | 75. | 3. |
| 182 | ford-fairmon | 18.0 | 6. | 171.0 | 97.00 | 2984. | 14.5 | 75. | 1. |

# *Schema*

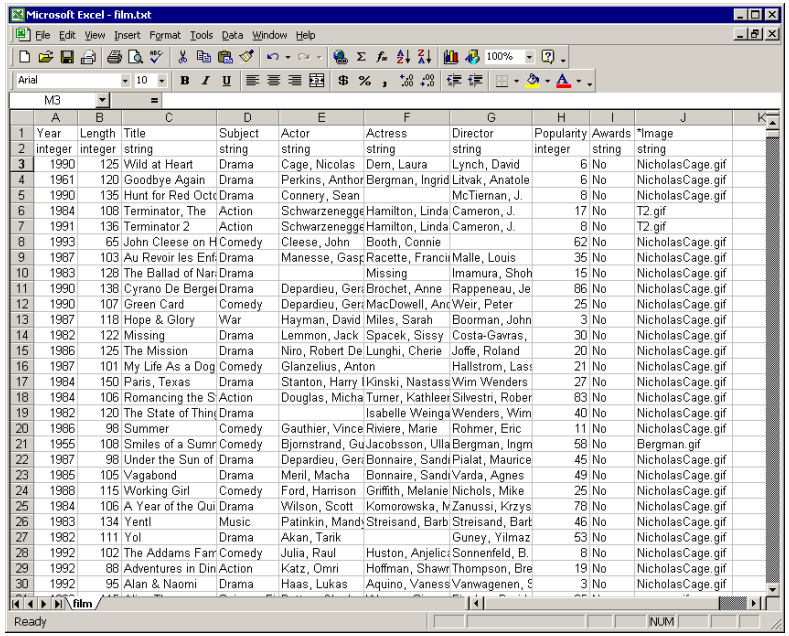
- Cars
  - brand
  - model
  - year
  - cost
  - size
  - weights
  - miles per gallon



# *Data Tables*

- Often, we take raw data and transform it into a form that is more workable

- Main idea:
  - Individual items are called cases (rows)
  - Cases have **variables** (columns)



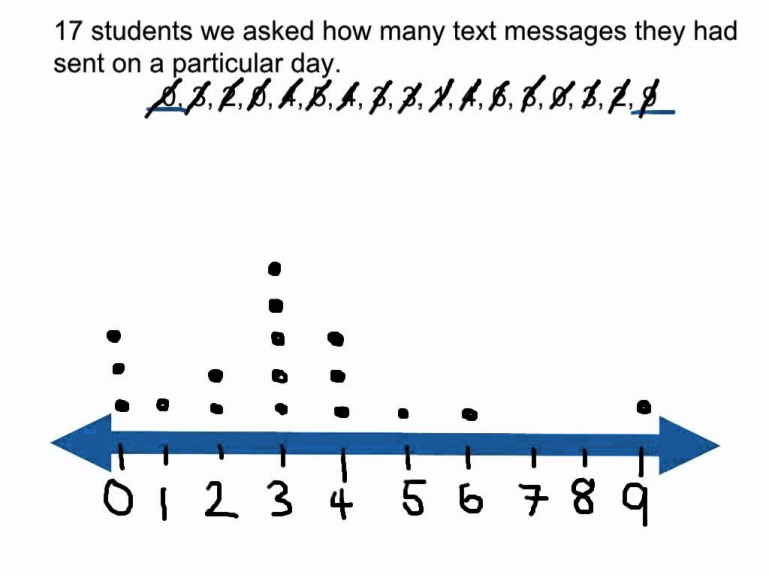

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Year | Length | Title | Subject | Actor | Actress | Director | Popularity | Awards | *Image |
| 2 | integer | integer | string | string | string | string | string | integer | string | string |
| 3 | 1990 | 125 | Wild at Heart | Drama | Cage, Nicolas | Dern, Laura | Lynch, David | 6 | No | NicholasCage.gif |
| 4 | 1961 | 120 | Goodbye Again | Drama | Perkins, Anthor | Bergman, Ingrid | Litvak, Anatole | 6 | No | NicholasCage.gif |
| 5 | 1990 | 135 | Hunt for Red Octo | Drama | Connery, Sean | | McTiernan, J. | 8 | No | NicholasCage.gif |
| 6 | 1984 | 108 | Terminator, The | Action | Schwarzenegge | Hamilton, Linda | Cameron, J. | 17 | No | T2.gif |
| 7 | 1991 | 136 | Terminator 2 | Action | Schwarzenegge | Hamilton, Linda | Cameron, J. | 8 | No | T2.gif |
| 8 | 1993 | 65 | John Cleese on H | Comedy | Cleese, John | Booth, Connie | | 62 | No | NicholasCage.gif |
| 9 | 1987 | 103 | Au Revoir les Enf | Drama | Manesse, Gasp | Racette, Francii | Malle, Louis | 35 | No | NicholasCage.gif |
| 10 | 1983 | 128 | The Ballad of Nar | Drama | | Missing | Imamura, Shoh | 15 | No | NicholasCage.gif |
| 11 | 1990 | 138 | Cyrano De Berge | Drama | Depardieu, Ger | Brochet, Anne | Rappeneau, Je | 86 | No | NicholasCage.gif |
| 12 | 1990 | 107 | Green Card | Comedy | Depardieu, Ger | MacDowell, And | Weir, Peter | 25 | No | NicholasCage.gif |
| 13 | 1987 | 118 | Hope & Glory | War | Hayman, David | Miles, Sarah | Boorman, John | 3 | No | NicholasCage.gif |
| 14 | 1982 | 122 | Missing | Drama | Lemmon, Jack | Spacek, Sissy | Costa-Gavras, | 30 | No | NicholasCage.gif |
| 15 | 1986 | 125 | The Mission | Drama | Niro, Robert De | Lunghi, Cherie | Joffe, Roland | 20 | No | NicholasCage.gif |
| 16 | 1987 | 101 | My Life As a Dog | Comedy | Glanzelius, Anton | | Hallstrom, Las | 21 | No | NicholasCage.gif |
| 17 | 1984 | 150 | Paris, Texas | Drama | Stanton, Harry I | Kinski, Nastass | Wim Wenders | 27 | No | NicholasCage.gif |
| 18 | 1984 | 106 | Romancing the S | Action | Douglas, Micha | Turner, Kathleer | Silvestri, Rober | 83 | No | NicholasCage.gif |
| 19 | 1982 | 120 | The State of Thin | Drama | | Isabelle Weinga | Wenders, Wim | 40 | No | NicholasCage.gif |
| 20 | 1986 | 98 | Summer | Comedy | Gauthier, Vince | Riviere, Marie | Rohmer, Eric | 11 | No | NicholasCage.gif |
| 21 | 1955 | 108 | Smiles of a Sumr | Comedy | Bjornstrand, Gu | Jacobsson, Ulla | Bergman, Ingm | 58 | No | Bergman.gif |
| 22 | 1987 | 98 | Under the Sun of | Drama | Depardieu, Ger | Bonnaire, Sandi | Pialat, Maurice | 45 | No | NicholasCage.gif |
| 23 | 1985 | 105 | Vagabond | Drama | Meril, Macha | Bonnaire, Sandi | Varda, Agnes | 49 | No | NicholasCage.gif |
| 24 | 1988 | 115 | Working Girl | Comedy | Ford, Harrison | Griffith, Melanie | Nichols, Mike | 25 | No | NicholasCage.gif |
| 25 | 1984 | 106 | A Year of the Qui | Drama | Wilson, Scott | Komorowska, M | Zanussi, Krzys | 78 | No | NicholasCage.gif |
| 26 | 1983 | 134 | Yentl | Music | Patinkin, Mand | Streisand, Barb | Streisand, Barb | 46 | No | NicholasCage.gif |
| 27 | 1982 | 111 | Yol | Drama | Akan, Tarik | | Guney, Yilmaz | 53 | No | NicholasCage.gif |
| 28 | 1992 | 102 | The Addams Fam | Comedy | Julia, Raul | Huston, Anjelic | Sonnenfeld, B. | 8 | No | NicholasCage.gif |
| 29 | 1992 | 88 | Adventures in Din | Action | Katz, Omri | Hoffman, Shawr | Thompson, Bre | 19 | No | NicholasCage.gif |
| 30 | 1992 | 95 | Alan & Naomi | Drama | Haas, Lukas | Aquino, Vaness | Vanwagenen, S | 3 | No | NicholasCage.gif |

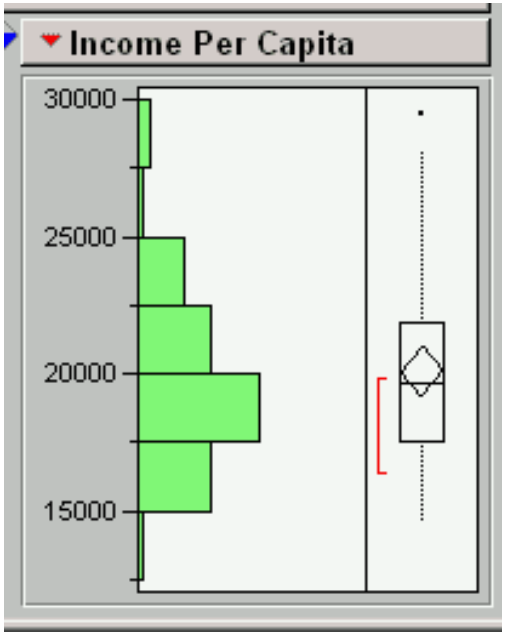
# *Multi-Dimensional data*

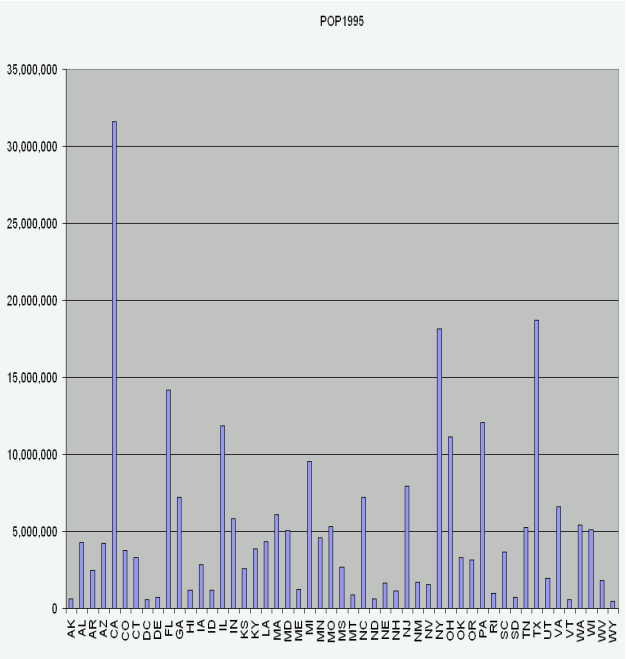
- Each variable defines a data dimension
- Small # of dimensions easy
  - Data mapping, Cleveland's rules
- Easy stuff
  - Univariate:  1 variable
  - Bivariate : 2 variables
  - Trivariate: 3 variables

# *Univariate*

- Dot plot
- Bar chart
- Histogram



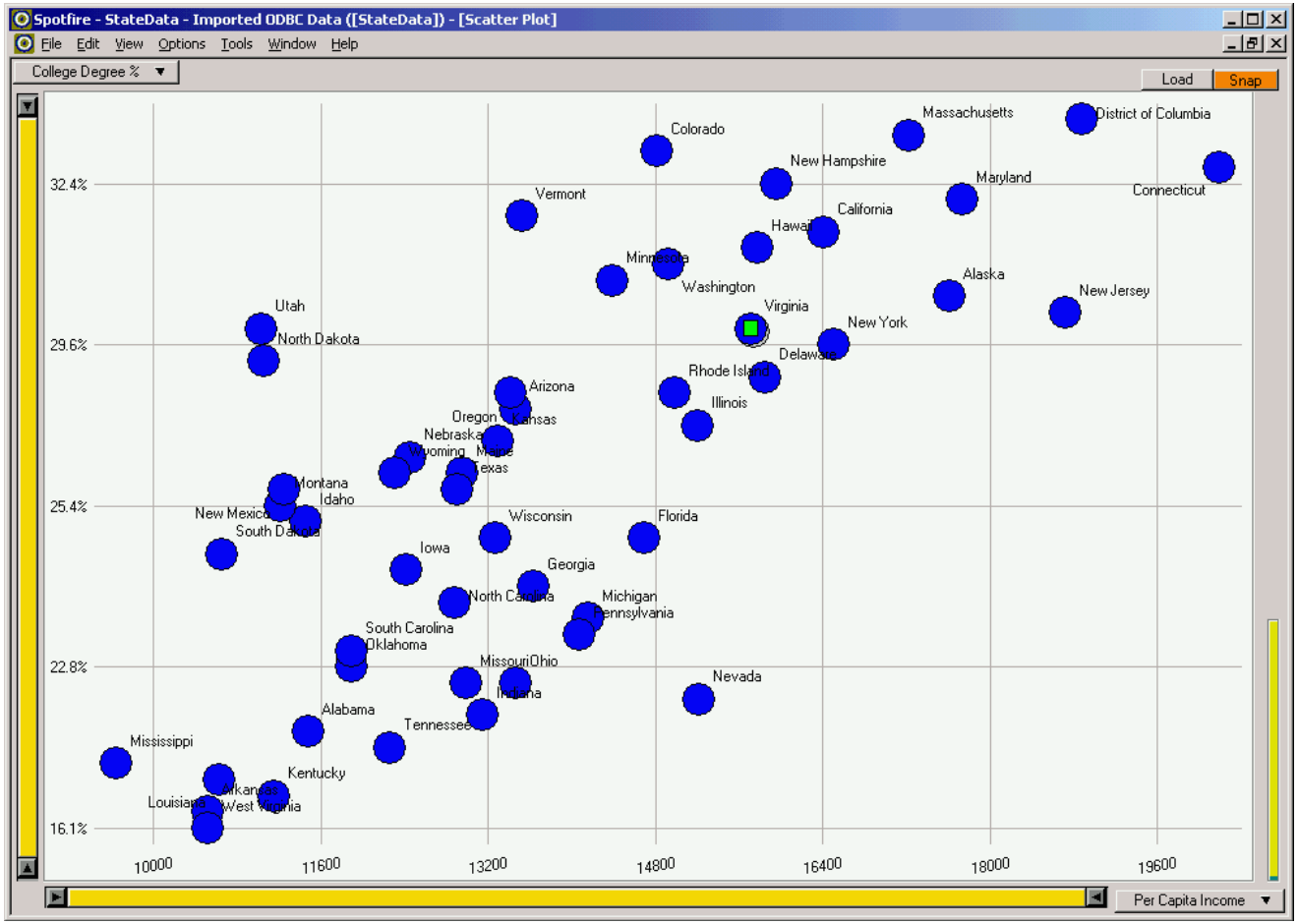




http://i.ytimg.com/vi/_zurDAF1Fw4/maxresdefault.jpg

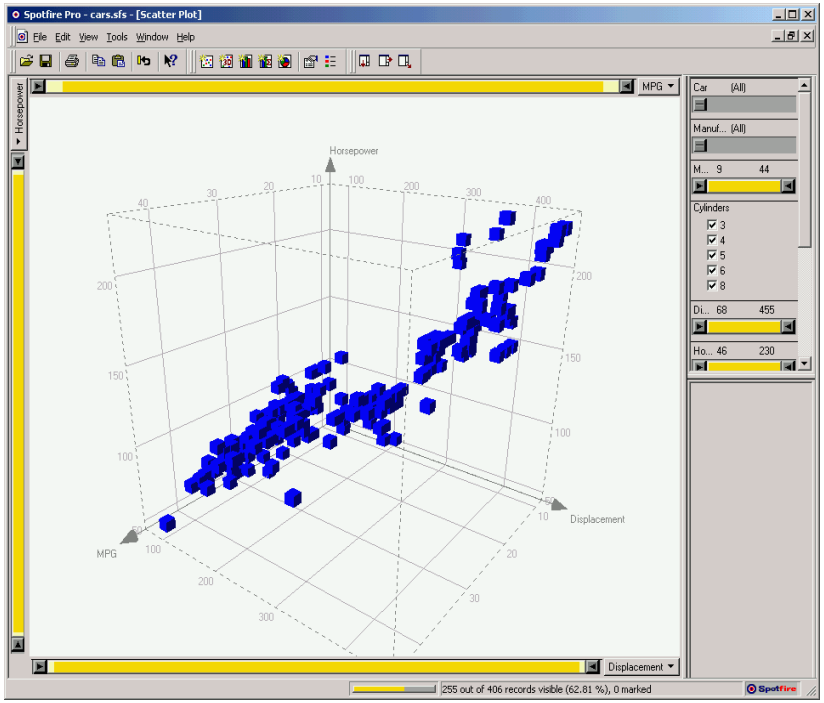
# *Bivariate*

- Scatterplot



# *Trivariate*

- 3D scatterplot, 3x barchart,
  2D plot+color, 2D plot+size



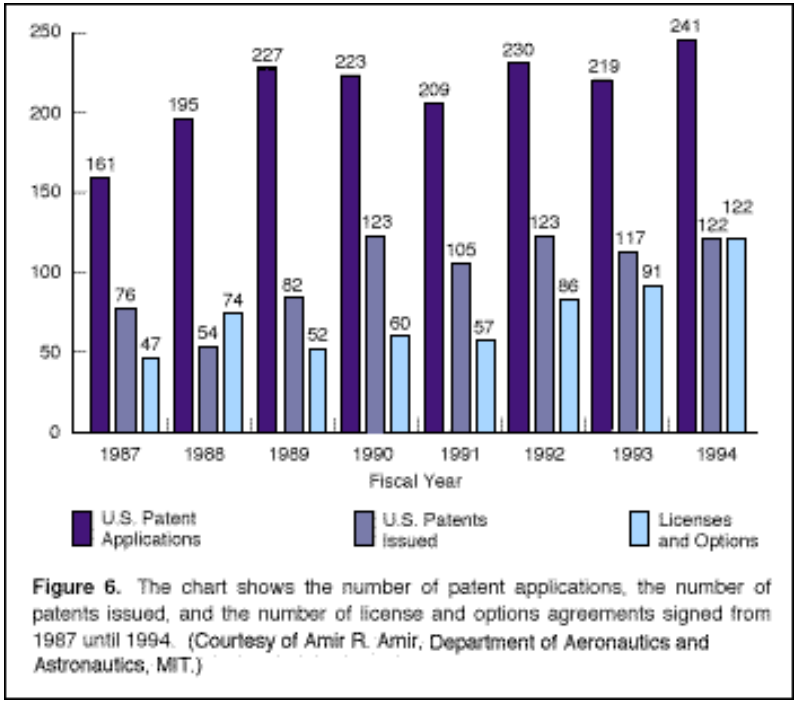
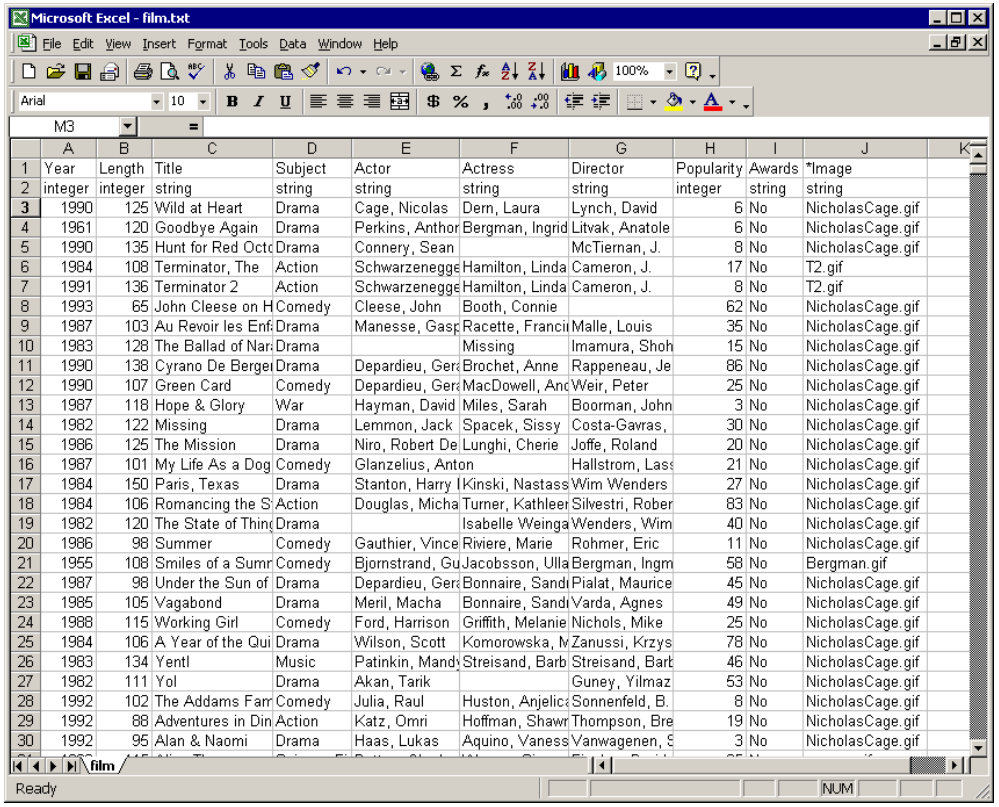




**Figure 6.** The chart shows the number of patent applications, the number of patents issued, and the number of license and options agreements signed from 1987 until 1994. (Courtesy of Amir R. Amir, Department of Aeronautics and Astronautics, MIT.)

# *Multi-Dimensional data*

- What about many dimensional data? n-D

What does 10-D space look like?

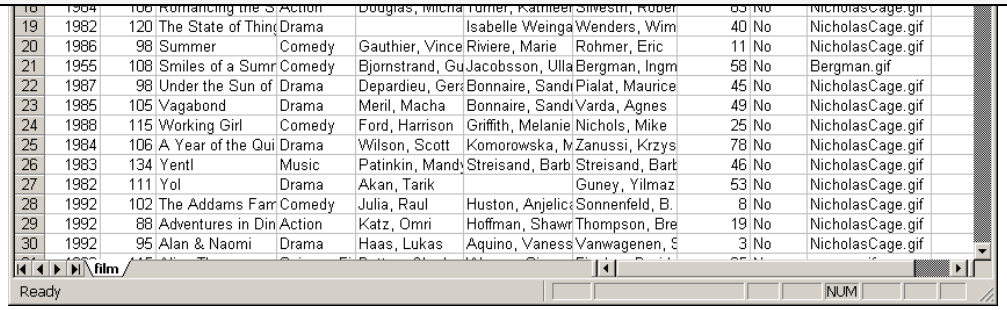



| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Year | Length | Title | Subject | Actor | Actress | Director | Popularity | Awards | *Image |
| 2 | integer | integer | string | string | string | string | string | integer | string | string |
| 3 | 1990 | 125 | Wild at Heart | Drama | Cage, Nicolas | Dern, Laura | Lynch, David | 6 | No | NicholasCage.gif |
| 4 | 1961 | 120 | Goodbye Again | Drama | Perkins, Anthor | Bergman, Ingrid | Litvak, Anatole | 6 | No | NicholasCage.gif |
| 5 | 1990 | 135 | Hunt for Red Octo | Drama | Connery, Sean | | McTiernan, J. | 8 | No | NicholasCage.gif |
| 6 | 1984 | 108 | Terminator, The | Action | Schwarzenegge | Hamilton, Linda | Cameron, J. | 17 | No | T2.gif |
| 7 | 1991 | 136 | Terminator 2 | Action | Schwarzenegge | Hamilton, Linda | Cameron, J. | 8 | No | T2.gif |
| 8 | 1993 | 65 | John Cleese on H | Comedy | Cleese, John | Booth, Connie | | 62 | No | NicholasCage.gif |
| 9 | 1987 | 103 | Au Revoir les Enf | Drama | Manesse, Gasp | Racette, Francii | Malle, Louis | 35 | No | NicholasCage.gif |
| 10 | 1983 | 128 | The Ballad of Nar | Drama | | Missing | Imamura, Shoh | 15 | No | NicholasCage.gif |
| 11 | 1990 | 138 | Cyrano De Berger | Drama | Depardieu, Ger | Brochet, Anne | Rappeneau, Je | 86 | No | NicholasCage.gif |
| 12 | 1990 | 107 | Green Card | Comedy | Depardieu, Ger | MacDowell, And | Weir, Peter | 25 | No | NicholasCage.gif |
| 13 | 1987 | 118 | Hope & Glory | War | Hayman, David | Miles, Sarah | Boorman, John | 3 | No | NicholasCage.gif |
| 14 | 1982 | 122 | Missing | Drama | Lemmon, Jack | Spacek, Sissy | Costa-Gavras, | 30 | No | NicholasCage.gif |
| 15 | 1986 | 125 | The Mission | Drama | Niro, Robert De | Lunghi, Cherie | Joffe, Roland | 20 | No | NicholasCage.gif |
| 16 | 1987 | 101 | My Life As a Dog | Comedy | Glanzelius, Anton | | Hallstrom, Lass | 21 | No | NicholasCage.gif |
| 17 | 1984 | 150 | Paris, Texas | Drama | Stanton, Harry I | Kinski, Nastass | Wim Wenders | 27 | No | NicholasCage.gif |
| 18 | 1984 | 106 | Romancing the S | Action | Douglas, Micha | Turner, Kathleer | Silvestri, Rober | 83 | No | NicholasCage.gif |
| 19 | 1982 | 120 | The State of Thin | Drama | | Isabelle Weinga | Wenders, Wim | 40 | No | NicholasCage.gif |
| 20 | 1986 | 98 | Summer | Comedy | Gauthier, Vince | Riviere, Marie | Rohmer, Eric | 11 | No | NicholasCage.gif |
| 21 | 1955 | 108 | Smiles of a Sumr | Comedy | Bjornstrand, Gu | Jacobsson, Ulla | Bergman, Ingm | 58 | No | Bergman.gif |
| 22 | 1987 | 98 | Under the Sun of | Drama | Depardieu, Ger | Bonnaire, Sandi | Pialat, Maurice | 45 | No | NicholasCage.gif |
| 23 | 1985 | 105 | Vagabond | Drama | Meril, Macha | Bonnaire, Sandi | Varda, Agnes | 49 | No | NicholasCage.gif |
| 24 | 1988 | 115 | Working Girl | Comedy | Ford, Harrison | Griffith, Melanie | Nichols, Mike | 25 | No | NicholasCage.gif |
| 25 | 1984 | 106 | A Year of the Qui | Drama | Wilson, Scott | Komorowska, M | Zanussi, Krzys | 78 | No | NicholasCage.gif |
| 26 | 1983 | 134 | Yentl | Music | Patinkin, Mandy | Streisand, Barb | Streisand, Barb | 46 | No | NicholasCage.gif |
| 27 | 1982 | 111 | Yol | Drama | Akan, Tarik | | Guney, Yilmaz | 53 | No | NicholasCage.gif |
| 28 | 1992 | 102 | The Addams Fam | Comedy | Julia, Raul | Huston, Anjelic | Sonnenfeld, B. | 8 | No | NicholasCage.gif |
| 29 | 1992 | 88 | Adventures in Din | Action | Katz, Omri | Hoffman, Shawr | Thompson, Bre | 19 | No | NicholasCage.gif |
| 30 | 1992 | 95 | Alan & Naomi | Drama | Haas, Lukas | Aquino, Vaness | Vanwagenen, S | 3 | No | NicholasCage.gif |

# *Multi-Dimensional data*

- What about many dimensional data?  n-D

*Note the difference between spatial dimensions (1,2,3D) and data dimensions (n-D)*



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 19 | 1982 | 120 | The State of Thin | Drama | | Isabelle Weinga | Wenders, Wim | 40 | No | NicholasCage.gif |
| 20 | 1986 | 98 | Summer | Comedy | Gauthier, Vince | Riviere, Marie | Rohmer, Eric | 11 | No | NicholasCage.gif |
| 21 | 1955 | 108 | Smiles of a Sumr | Comedy | Bjornstrand, Gu | Jacobsson, Ulla | Bergman, Ingm | 58 | No | Bergman.gif |
| 22 | 1987 | 98 | Under the Sun of | Drama | Depardieu, Ger | Bonnaire, Sandi | Pialat, Maurice | 45 | No | NicholasCage.gif |
| 23 | 1985 | 105 | Vagabond | Drama | Meril, Macha | Bonnaire, Sandi | Varda, Agnes | 49 | No | NicholasCage.gif |
| 24 | 1988 | 115 | Working Girl | Comedy | Ford, Harrison | Griffith, Melanie | Nichols, Mike | 25 | No | NicholasCage.gif |
| 25 | 1984 | 106 | A Year of the Qui | Drama | Wilson, Scott | Komorowska, M | Zanussi, Krzys | 78 | No | NicholasCage.gif |
| 26 | 1983 | 134 | Yentl | Music | Patinkin, Mandy | Streisand, Barb | Streisand, Barb | 46 | No | NicholasCage.gif |
| 27 | 1982 | 111 | Yol | Drama | Akan, Tarik | | Guney, Yilmaz | 53 | No | NicholasCage.gif |
| 28 | 1992 | 102 | The Addams Fam | Comedy | Julia, Raul | Huston, Anjelic | Sonnenfeld, B. | 8 | No | NicholasCage.gif |
| 29 | 1992 | 88 | Adventures in Din | Action | Katz, Omri | Hoffman, Shawr | Thompson, Bre | 19 | No | NicholasCage.gif |
| 30 | 1992 | 95 | Alan & Naomi | Drama | Haas, Lukas | Aquino, Vaness | Vanwagenen, S | 3 | No | NicholasCage.gif |

film

Ready

NUM

# *Terminology …*

- Studied long before computer science, by statisticians and psychologists.

- *The term «dimensionality»* is overloaded: multidimensional, multivariate, hyperdimensional, hypervariate …

# *Terminology …*

- Variables can be dependent or independent:

  $y = f(x)$

  *y:* ***dependent*** *variable*

  *x:* ***independent*** *variable*

# *Terminology …*

- Variables can be dependent or independent:

  $y = f(x)$

  *y:* ***dependent*** *variable*

  *x:* ***independent*** *variable*

- **Multidimensionality** refers to the dimensionality of the **independent** variables

- **Multivariate** refers to the dimensionality of the **dependent** variables

P. C. Wong and R. D. Bergeron, 1997. 30 Years of Multidimensional Multivariate Visualization

# *Multivariate Multi-Dimensional Example*

- Three dimensional cube showing temperature and pressure at various locations is **3d2v**
  - **3 independent variables:** XYZ
  - **2 dependent variables:** temperature and pressure

# *Warning!*

- We do not always know the dependent / independent variables

- For the sake of simplicity, we use the term "multi-dimensional" to refer to dimensionality of both dependent and independent variables

multivariate = multi-dimensional

**3d2v ➔ 5D**

# *Genome Databases*


http://www.extremetech.com/extreme/151133-the-quest-for-the-1000-genome/3

# *Customer Transactions Data*


| Product ID | # of instances of cart add | # of unique instances of cart add | cart add Total Rs. Value | # of instances of cart removal | # of Unique instances of cart removal | cart remove Total value in Rs. | product views | unique product views | Value of products view in Rs. | Item revenue in Rs. | cart add % to product view | cart remove % to cart add |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 134454 | 455 | 381 | 461745 | 27 | 25 | 28773 | 14167 | 11540 | 14339133 | 52985.43 | 6% | 33 |
| 134455 | 833 | 622 | 425667 | 8 | 5 | 3992 | 11171 | 8718 | 5714329 | 23819.47 | 1% | 38 |
| 134456 | 435 | 340 | 570065 | 9 | 6 | 11691 | 12198 | 10215 | 16032202 | 80380.28 | 2% | 47 |
| 134457 | 122 | 101 | 121878 | 2 | 2 | 1998 | 6926 | 6017 | 6919074 | 4415.45 | 2% | 18 |
| 134458 | 239 | 196 | 609761 | 2 | 2 | 4998 | 11250 | 9733 | 28343750 | 57435.56 | 1% | 54 |
| 134459 | 293 | 235 | 206307 | 8 | 7 | 5592 | 7171 | 6140 | 5044929 | 10508.89 | 3% | 29 |
| 134460 | 314 | 249 | 288586 | 24 | 20 | 21576 | 6806 | 5275 | 6247294 | 32582.69 | 8% | 42 |
| 134461 | 88 | 78 | 43912 | 0 | 0 | 0 | 4912 | 4324 | 2451088 | 882.2 | 0% | 9 |
| 134462 | 256 | 210 | 342504 | 3 | 1 | 3987 | 3972 | 2901 | 5336358 | 27026.61 | 1% | 86 |
| 134463 | 45 | 37 | 40455 | 0 | 0 | 0 | 5712 | 4424 | 5214288 | 10331.1 | 0% | 7 |
| 134464 | 196 | 154 | 159604 | 36 | 29 | 29364 | 3638 | 2830 | 2948162 | 10594.49 | 18% | 44 |
| 134465 | 90 | 67 | 227410 | 7 | 1 | 17493 | 5723 | 5005 | 14644277 | 11045.3 | 8% | 40 |
| 134466 | 180 | 141 | 169470 | 6 | 4 | 5454 | 4939 | 4068 | 4645161 | 15272.2 | 3% | 34 |
| 134467 | 26 | 19 | 10894 | 0 | 0 | 0 | 3134 | 2845 | 1321426 | 370.65 | 0% | 3 |
| 134468 | 173 | 128 | 279737 | 4 | 3 | 6436 | 3023 | 2142 | 4888847 | 19916.26 | 2% | 93 |
| 134469 | 166 | 136 | 267784 | 4 | 2 | 6436 | 2706 | 1937 | 4386384 | 12803.31 | 2% | 99 |
| 134470 | 196 | 149 | 264474 | 3 | 2 | 3987 | 2840 | 2141 | 3810840 | 16450.98 | 2% | 93 |
| 134471 | 36 | 27 | 26364 | 4 | 2 | 2796 | 3200 | 2753 | 2402490 | 5747 | 11% | 8 |
| 134472 | 44 | 36 | 21956 | 0 | 0 | 0 | 3391 | 3032 | 1692109 | 882.2 | 0% | 6 |
| 134473 | 116 | 87 | 159500 | 3 | 3 | 4125 | 1821 | 1457 | 2503123 | 0 | 3% | 88 |
| 134474 | 35 | 25 | 17465 | 0 | 0 | 0 | 3413 | 2552 | 1703087 | 2205.51 | 0% | 5 |

http://www.hongkiat.com/blog/secure-online-transactions-consumer-guide/

http://quirksblog.com/blog/2014/09/26/buyers-or-sellers-who-is-driving-customer-to-customer-web-transactions/

http://www.tatvic.com/blog/cart-analysis/

# *Personal Data – Fitness Bands*


| Date | Activity | Group | Start | End | Duration | Distance | Steps | Calories |
|---|---|---|---|---|---|---|---|---|
| 5/10/14 | walking | walking | 2014-05-10T09:34:04-04:00 | 2014-05-10T09:39:40-04:00 | 336 | 0.257 | 552 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T10:00:00-04:00 | 2014-05-10T10:00:41-04:00 | 41 | 0.02 | 63 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T12:33:26-04:00 | 2014-05-10T12:35:59-04:00 | 153 | 0.111 | 238 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T12:46:41-04:00 | 2014-05-10T12:50:19-04:00 | 218 | 0.148 | 317 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T13:00:47-04:00 | 2014-05-10T13:10:03-04:00 | 556 | 0.371 | 796 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T14:04:51-04:00 | 2014-05-10T14:07:22-04:00 | 151 | 0.071 | 152 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T15:09:11-04:00 | 2014-05-10T15:15:57-04:00 | 406 | 0.279 | 599 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T16:03:50-04:00 | 2014-05-10T16:11:30-04:00 | 460 | 0.319 | 683 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T17:09:15-04:00 | 2014-05-10T17:10:46-04:00 | 91 | 0.034 | 109 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T18:05:09-04:00 | 2014-05-10T18:09:44-04:00 | 275 | 0.204 | 437 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T19:04:54-04:00 | 2014-05-10T19:40:34-04:00 | 2140 | 1.373 | 2945 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T20:00:00-04:00 | 2014-05-10T20:20:13-04:00 | 1213 | 0.808 | 1734 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T21:24:02-04:00 | 2014-05-10T21:25:49-04:00 | 107 | 0.037 | 118 | 0 |
| 5/10/14 | walking | walking | 2014-05-10T23:18:10-04:00 | 2014-05-10T23:18:36-04:00 | 26 | 0.014 | 43 | 0 |
| | | | | | 6173 | 4.046 | 8786 | 0 |

# World DataBank

## World Development Indicators

TABLE | CHART | MAP | DOWNLOAD

After selecting variables, please click one of the above options

### DIMENSION FILTERS

**Hierarchy**

Collapse all | Expand all | Unselect all

- Income
- Lending
- Region

*Aggregates*

### COUNTRY (Available:214| Selected:1)

### SERIES (Available:1300| Selected:0)

### TIME (Available:| Selected:0)

### YOUR CURRENT SELECTION

**DATABASE**

World Development Indicators

Change database

**COUNTRY (1)**

Remove all | Sort

*Drag to rearrange the order*

France

**SERIES (0)**

**TIME (0)**

# World DataBank

## World Development Indicators

TABLE | CHART | MAP | DOWNLOAD

After selecting variables, please click one of the above options

### DIMENSION FILTERS

**Hierarchy**

Collapse all | Expand all

- Topic
  - Education
  - Environment
  - Economic Policy & Debt
  - Financial Sector
  - Health
  - Infrastructure
  - Labor & Social Protection
  - Poverty
  - Private Sector & Trade
  - Public Sector

### COUNTRY

(Available:214| Selected:1)

### SERIES

(Available:1300| Selected:0)

Type keywords to filter

( | A | B | C | D | E | F | G | H | I | L | M | N | O | P | Q | R | S | T | U | V | W |

Select all | Unselect all | Invert selection

(

Top

- (%) Benefits held by 1st 20% population - All Social Insurance
- (%) Benefits held by 1st 20% population - All Social Protection
- (%) Benefits held by 1st 20% population - All Social Safety Nets
- (%) Benefits held by 1st 20% population - Unemp benefits and ALMP
- (%) Generosity of All Social Insurance
- (%) Generosity of All Social Protection
- (%) Generosity of All Social Safety Nets
- (%) Generosity of Unemp benefits and ALMP
- (%) Program participation - All Social Insurance
- (%) Program participation - All Social Protection
- (%) Program participation - All Social Safety Nets
- (%) Program participation - Unemp benefits and ALMP

A

Top

- ARI treatment (% of children under 5 taken to a health provider)
- Access to electricity (% of population)
- Adjusted net enrollment rate, primary (% of primary school age children)
- Adjusted net enrollment rate, primary, female (% of primary school age children)
- Adjusted net enrollment rate, primary, male (% of primary school age children)
- Adjusted net national income (annual % growth)
- Adjusted net national income (constant 2005 US$)
- Adjusted net national income (current US$)
- Adjusted net savings, excluding particulate emission damage (% of GNI)
- Adjusted net savings, excluding particulate emission damage (current US$)
- Adjusted net savings, including particulate emission damage (% of GNI)

### TIME

(Available:| Selected:0)

### YOUR CURRENT SELECTION

**DATABASE**

World Development Indicators

Change database

**COUNTRY (1)**

**SERIES (0)**

Remove all | Sort

*Drag to rearrange the order*

**TIME (0)**

1300 indicators

214 countries

| | Benefits | Generosity | Participation | Savings | Interest | Bird species | Birth Rate | C02 Emissions | … |
|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | | | | | | | | | |
| Albania | | | | | | | | | |
| Algeria | | | | | | | | | |
| Andorra | | | | | | | | | |
| Angola | | | | | | | | | |
| Argentina | | | | | | | | | |
| Armenia | | | | | | | | | |
| Aruba | | | | | | | | | |
| Australia | | | | | | | | | |
| … | | | | | | | | | |

# World DataBank

World Development Indicators

TABLE | CHART | MAP | DOWNLOAD

After selecting variables, please click one of the above options

## DIMENSION FILTERS

### Hierarchy

Collapse all | Expand all

- Year

## COUNTRY (Available:214| Selected:1)

## SERIES (Available:1300| Selected:0)

## TIME (Available:54| Selected:0)

SELECT YEARS | FUNCTIONS

Availability Range: **Year 1960 - 2013**

Please drag the slider handle below to select the range

SEARCH

Select all | Unselect all | Invert selection

- 2013
- 2012
- 2011
- 2010
- 2009
- 2008
- 2007
- 2006
- 2005
- 2004
- 2003
- 2002
- 2001
- 2000
- 1999
- 1998
- 1997
- 1996
- 1995
- 1994
- 1993
- 1992
- 1991

## YOUR CURRENT SELECTION

**DATABASE**

World Development Indicators

Change database

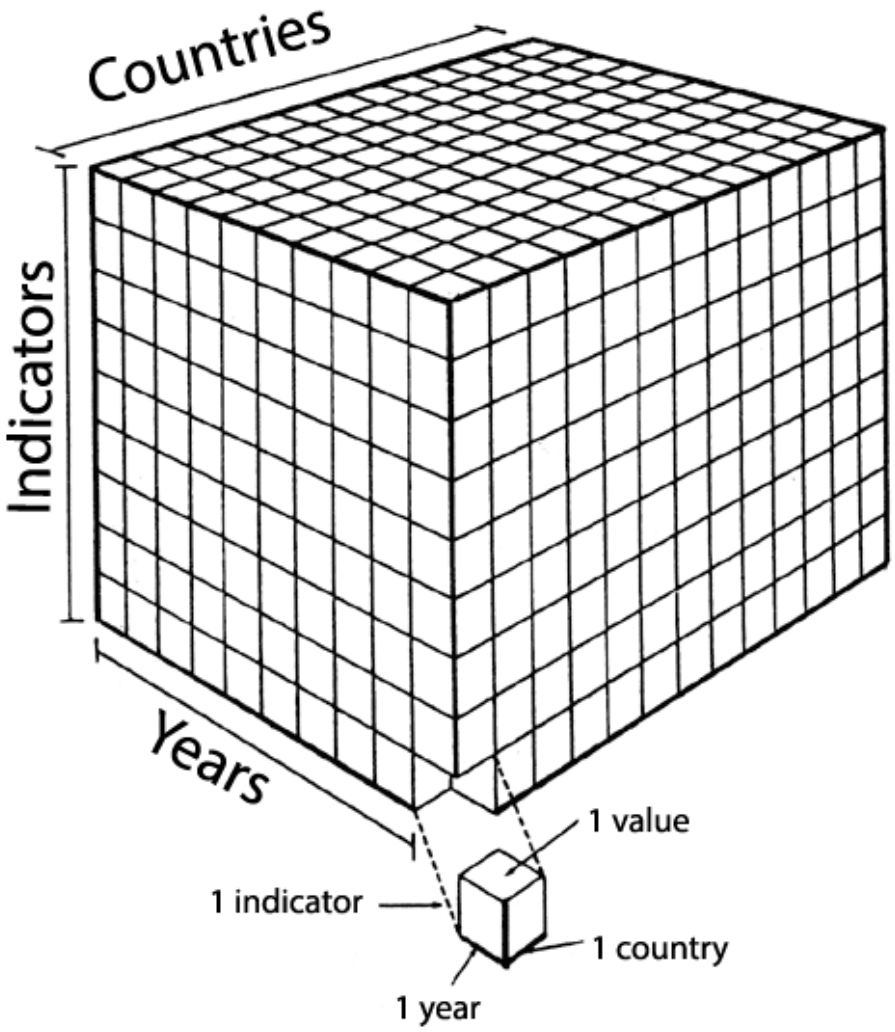
**COUNTRY (1)**

**SERIES (0)**

**TIME (0)**

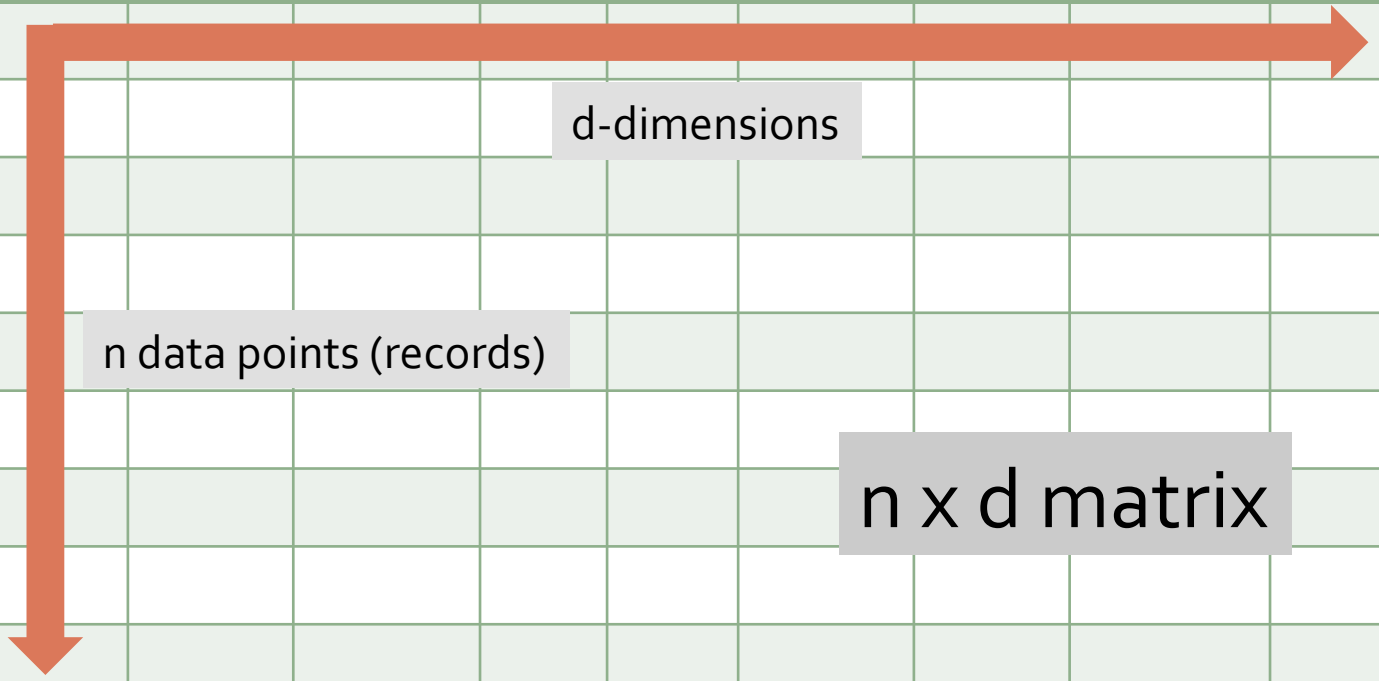
Remove all | Sort | Ascending

# *the data cube*



# *multi-dimensional data – our view today*

| | Benefits | Generosity | Participation | Savings | Interest | Bird species | Birth Rate | C02 Emissions | … |
|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | | | | | | | | | |
| Albania | | | | | | | | | |
| Algeria | | | | | | | | | |
| Andorra | | | | | | | | | |
| Angola | | | | | | | | | |
| Argentina | | | | | | | | | |
| Armenia | | | | | | | | | |
| Aruba | | | | | | | | | |
| Australia | | | | | | | | | |
| … | | | | | | | | | |



# *VISUALIZING MULTI-DIMENSIONAL DATA*

# *Data-Type Taxonomy*

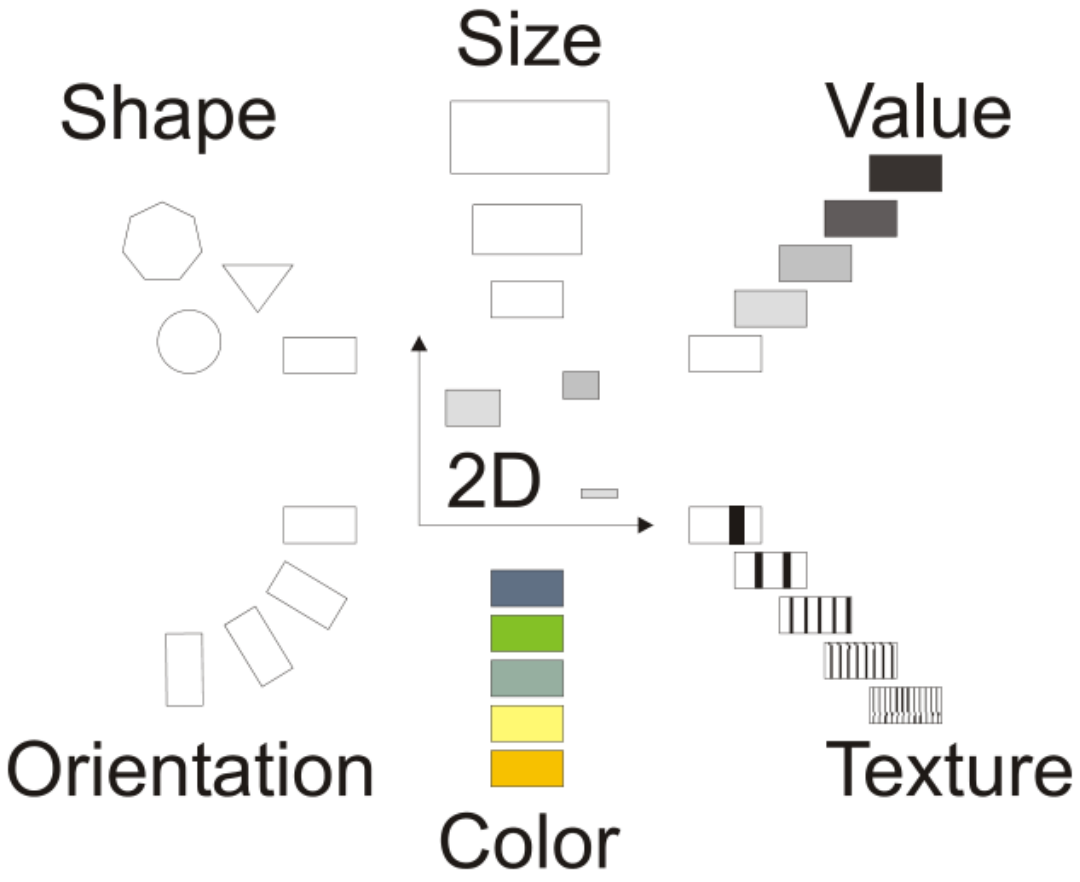
- 1D: list, …
- 2D: maps, …
- 3D: volumes, …
- Temporal: timelines , …
- Hierarchies/trees: directories, …
- Network/Graphs: web, …
- **Multi-dimensional: databases, …**

[Shneiderman 1996]

# *Challenges*

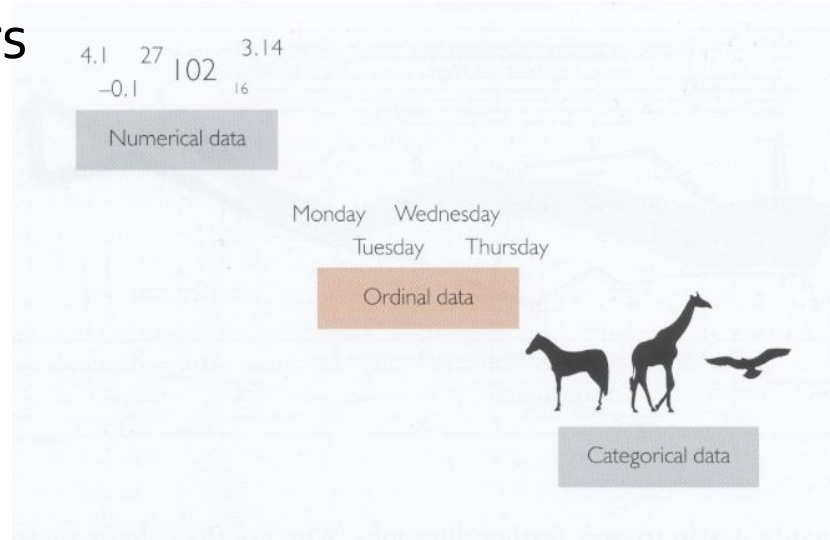
- Mapping from ND to 2D
- Clutter on screen
- Difficult user navigation in data space

# *recap – last lecture*

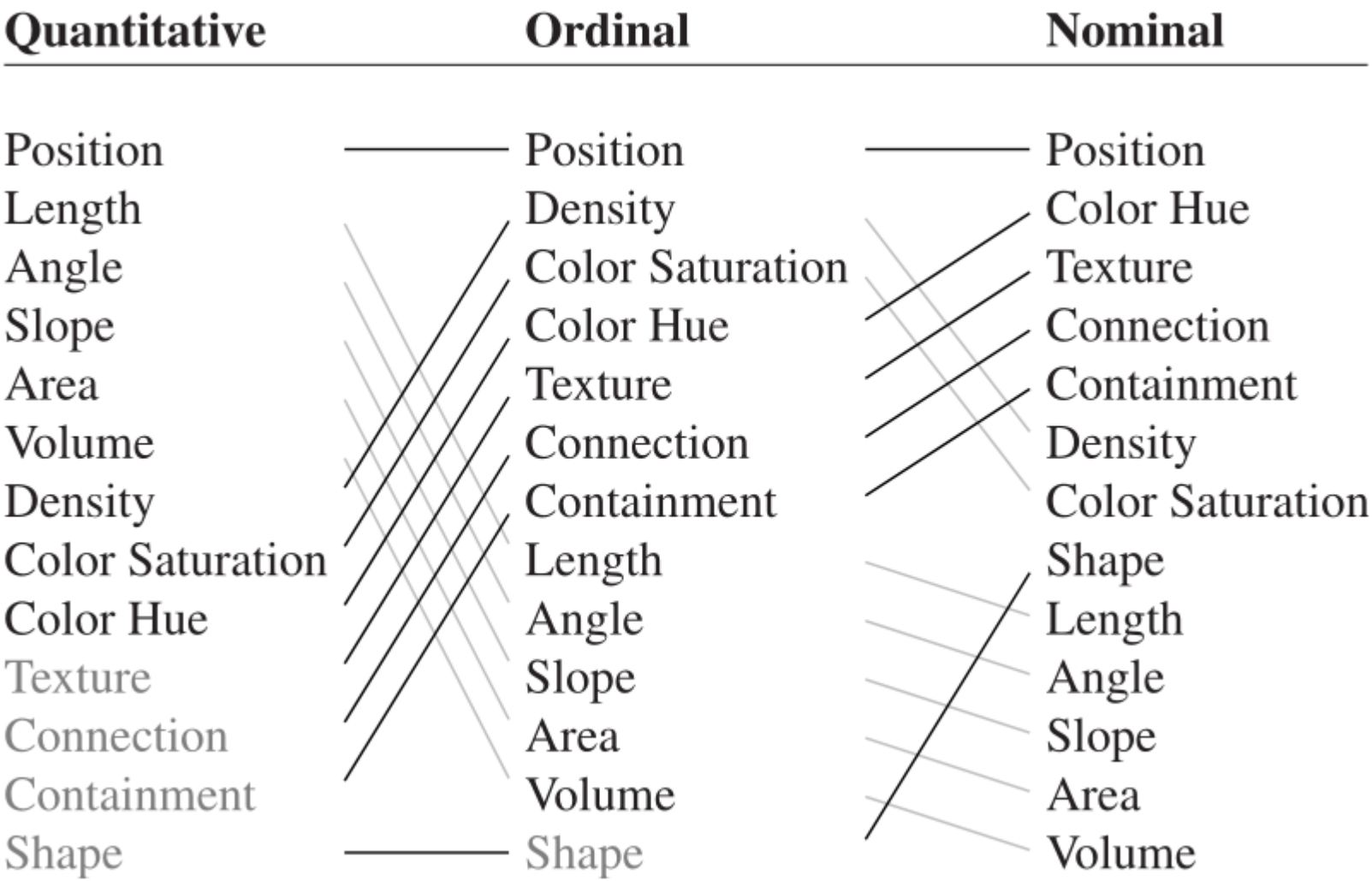


# *recap: Variable Types*

- **N-Nominal** (equal or not equal to other values)
  - Example: gender, hair color (blond, brown, black, red)
- **O-Ordinal** (obeys < relation, ordered set)
  - Example: soccer leagues, rainbow colors
- **Q-Quantitative** (can do math on them)
  - Example: age, Photoshop colors



# *recap - visual encodings*




Mackinlay 1986

# *the problem*

1300 indicators →

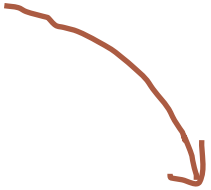
214 countries ↓

| | Benefits | Generosity | Participation | Savings | Interest | Bird species | Birth Rate | C02 Emissions | … |
|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | | | | | | | | | |
| Albania | | | | | | | | | |
| Algeria | | | | | | | | | |
| Andorra | | | | | | | | | |
| Angola | | | | | | | | | |
| Argentina | | | | | | | | | |
| Armenia | | | | | | | | | |
| Aruba | | | | | | | | | |
| Australia | | | | | | | | | |
| … | | | | | | | | | |

Hundreds or thousands of dimensions, only limited number of visual variables!

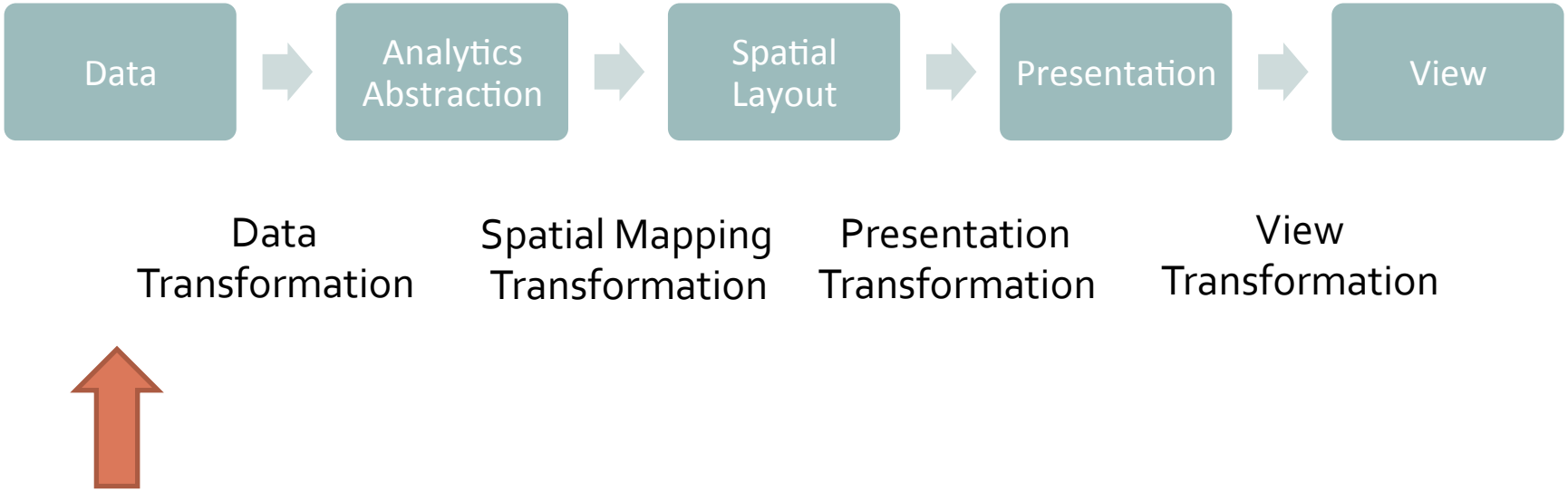
# *two main approaches*

1. **Attempt to visualize all dimensions**
2. Visualize only parts of the data
   → add interaction



*Interaction Lecture*

# *how do we visualize data points?*




1. Find a Dataset

# *how do we visualize data points?*


Data Transformation

Spatial Mapping Transformation

Presentation Transformation

View Transformation

2. Transform the data
to arrive at data you want to visualize

# *transforming MD data*

- Filter the data
  - decide to show LESS of the data
  - in practice you almost always do this
- Use a dimensionality reduction technique
  - E.g. principal component analysis, multi-dimensional scaling
  - often used for exploring (dis)similarities in data

# *how do we visualize data points?*


Data Transformation

Spatial Mapping Transformation

Presentation Transformation

View Transformation

3. Your remaining possibly transformed dimensions

# *how do we visualize data points?*


Data Transformation

Spatial Mapping Transformation

Presentation Transformation

View Transformation

4. Choose how to put the data points on the drawing plane

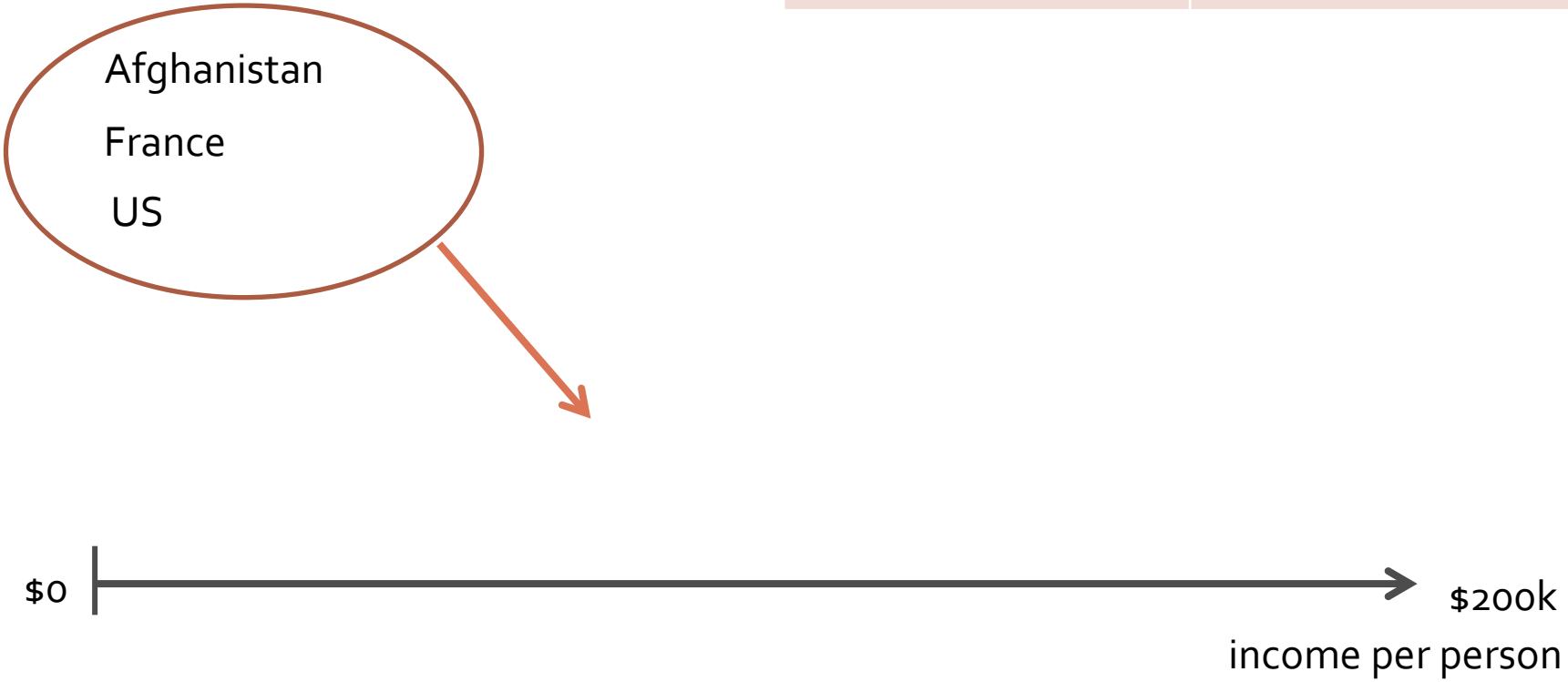
# *1. find a layout*

- Let's start with one dimension: income per person
- 1-D Data points: Afghanistan, France, US

| | Income per person |
|---|---|
| Afghanistan | 850 |
| France | 29500 |
| US | 41000 |

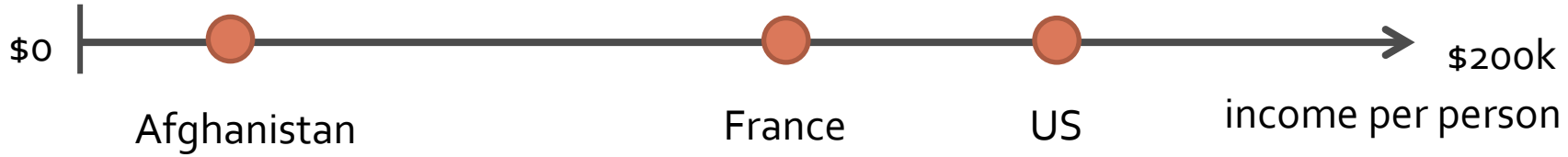
# *1. find a layout*

| | Income per person |
|---|---|
| Afghanistan | 850 |
| France | 29500 |
| US | 41000 |



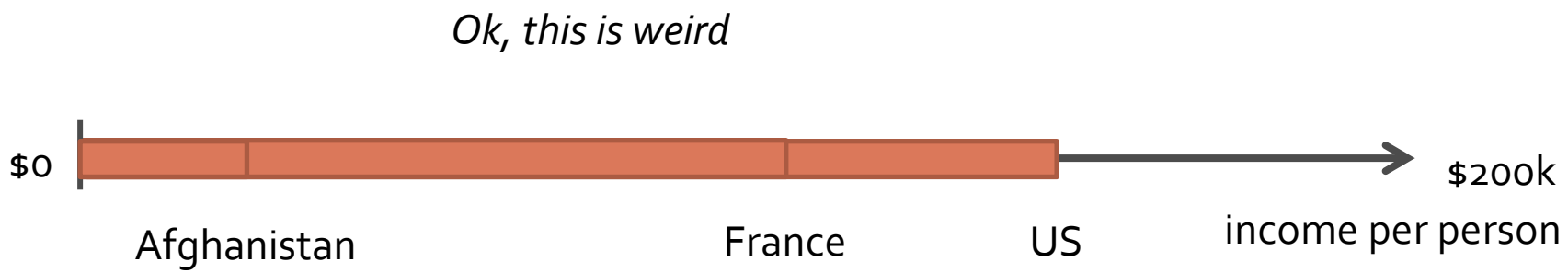
# *2. choose a visual encoding & mark*

- E.g. position + circle



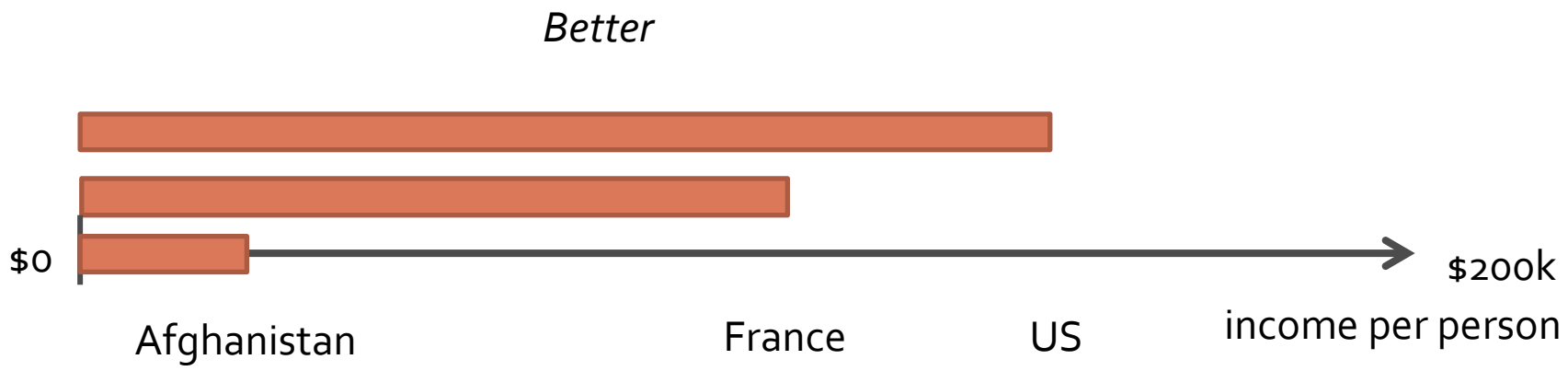
# *2. choose a visual encoding & mark*

- E.g. length + rectangle



# *2. choose a visual encoding & mark*

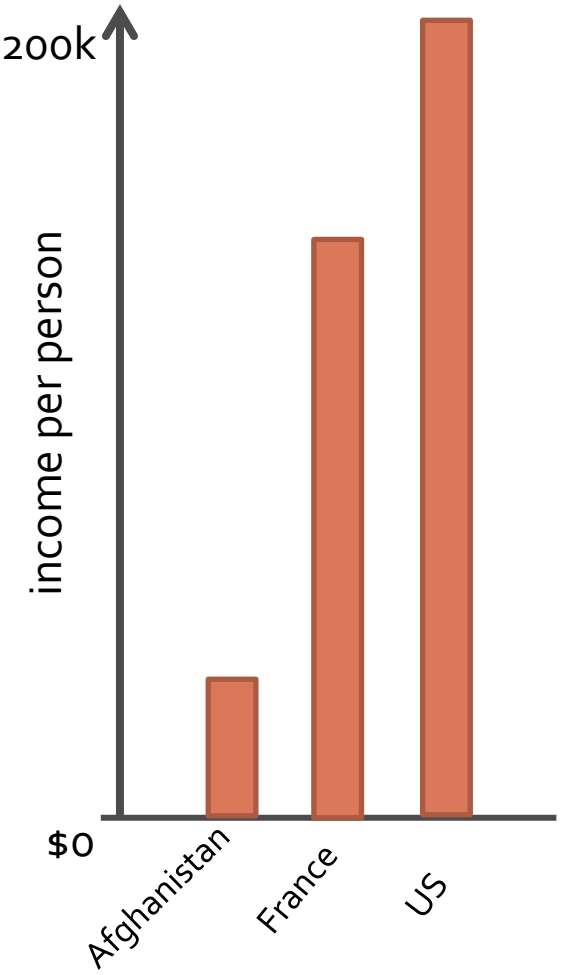
- E.g. length + rectangle



# *2. choose a visual encoding & mark*

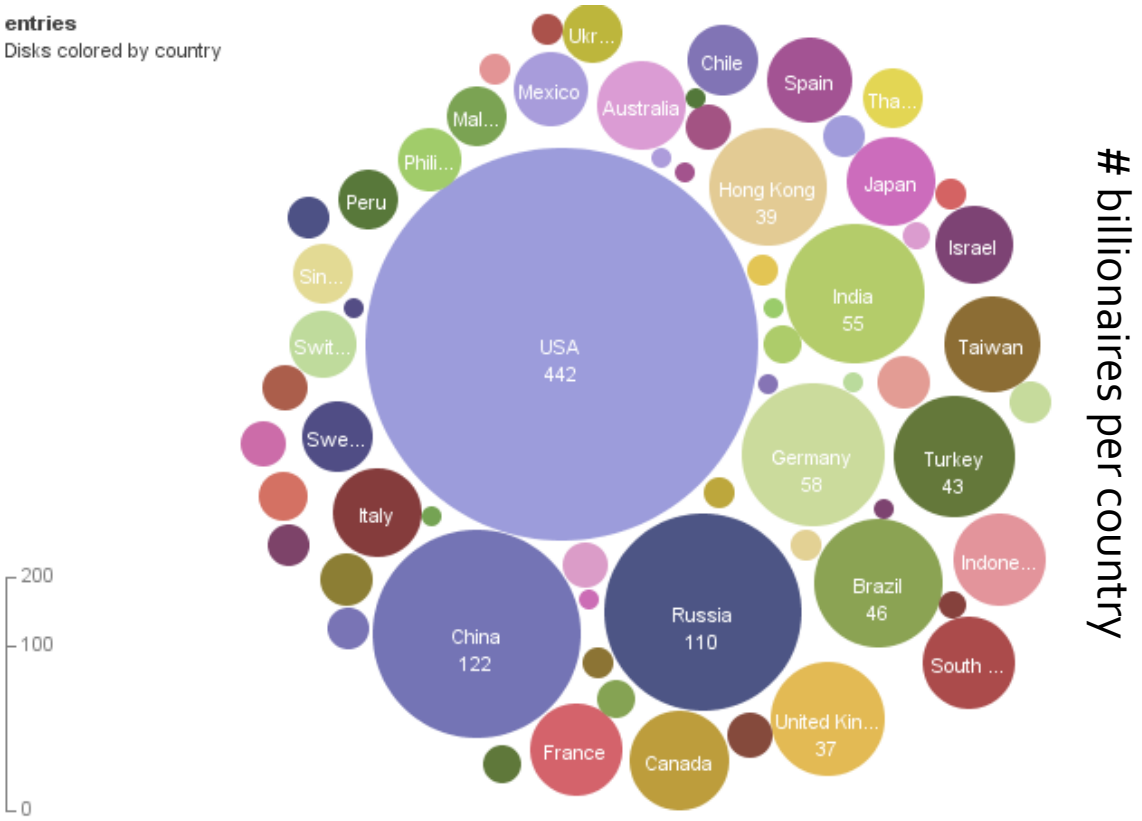
- E.g. length + rectangle

*Even better*

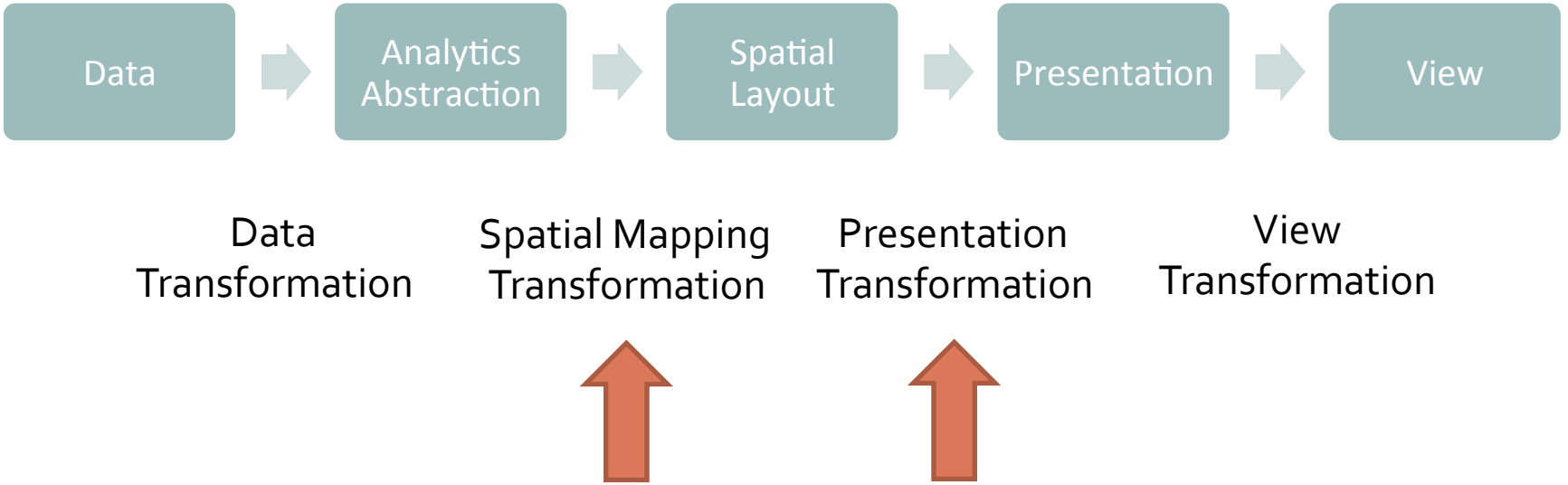


# *bubble chart*

- Spatial layout does not have to be ordered and can be computed by a variety of algorithms
  (what's most effective is another question)



# *Choosing spatial layouts, marks, and visual encodings often goes hand-in-hand*



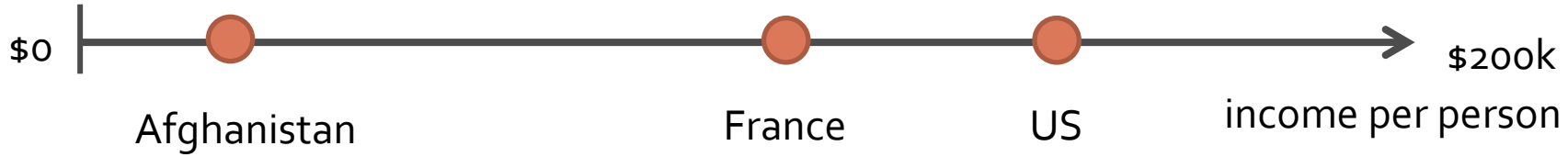

4. Choose how to put the data points on the drawing plane

5. Choose additional visual encodings

# *1. find a layout*

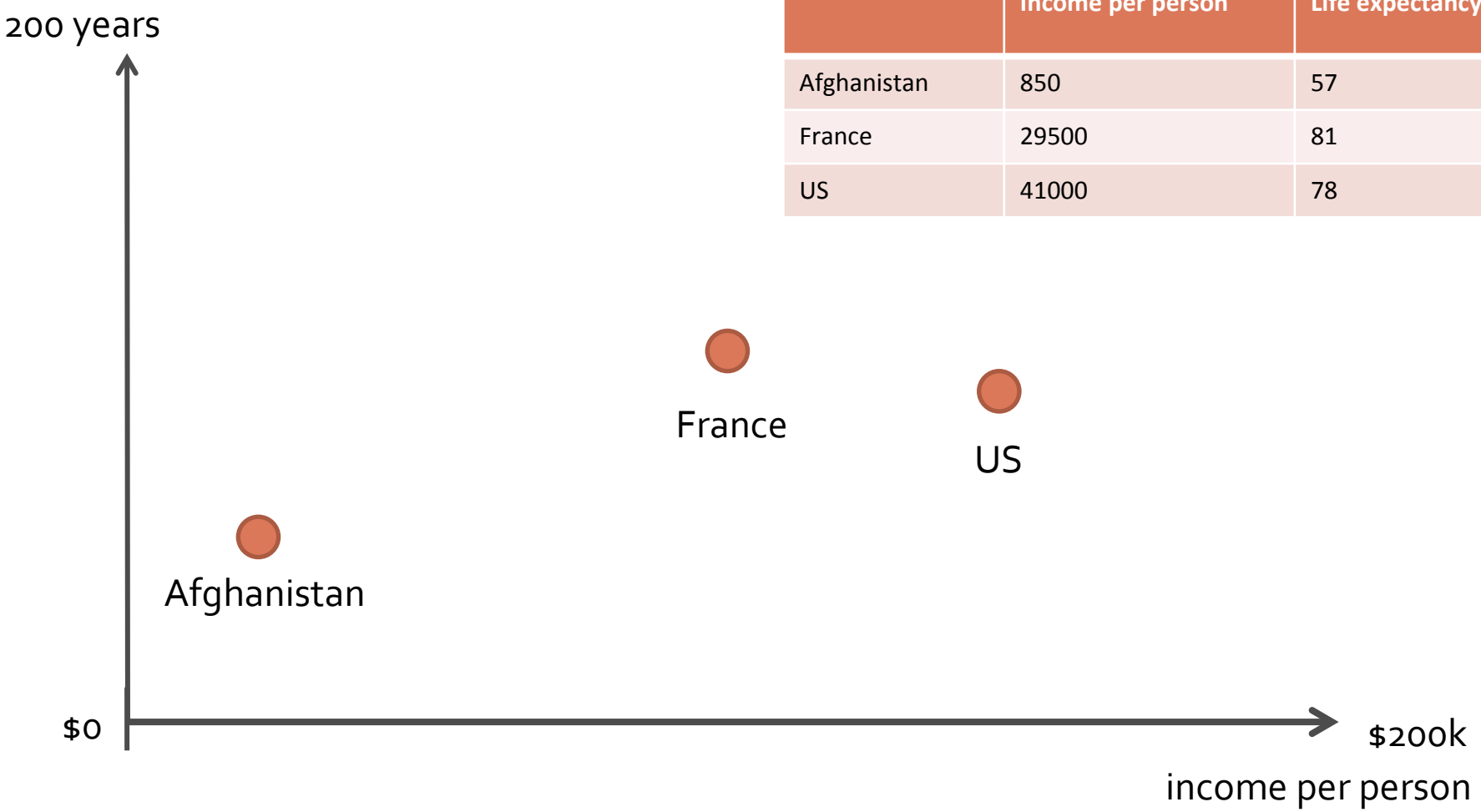
| | Income per person | Life expectancy |
|---|---|---|
| Afghanistan | 850 | 57 |
| France | 29500 | 81 |
| US | 41000 | 78 |

How do we extend this to 2 data dimensions?

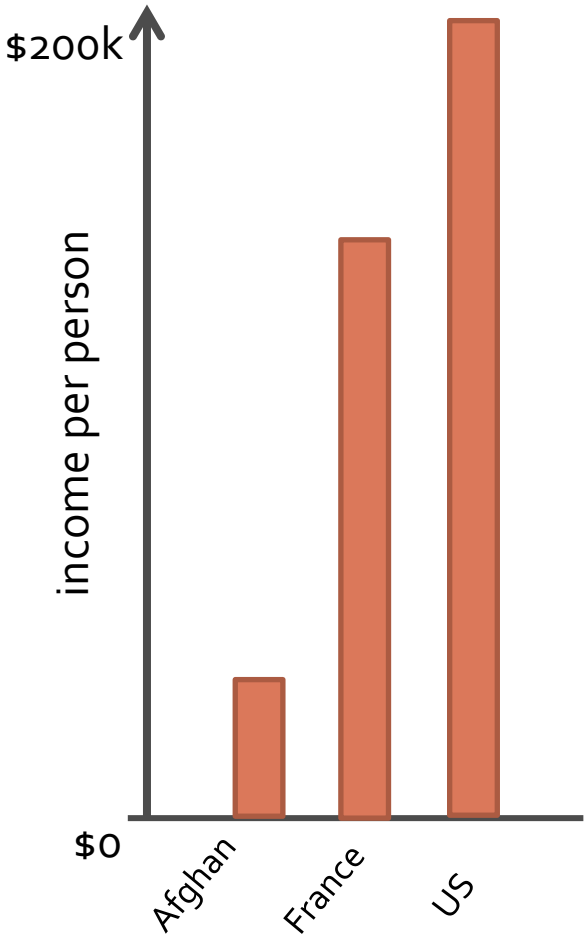


# *1. find a layout*

| | Income per person | Life expectancy |
|---|---|---|
| Afghanistan | 850 | 57 |
| France | 29500 | 81 |
| US | 41000 | 78 |



# *how do you extend this to 2D?*




| | Income per person | Life expectancy |
|---|---|---|
| Afghanistan | 850 | 57 |
| France | 29500 | 81 |
| US | 41000 | 78 |

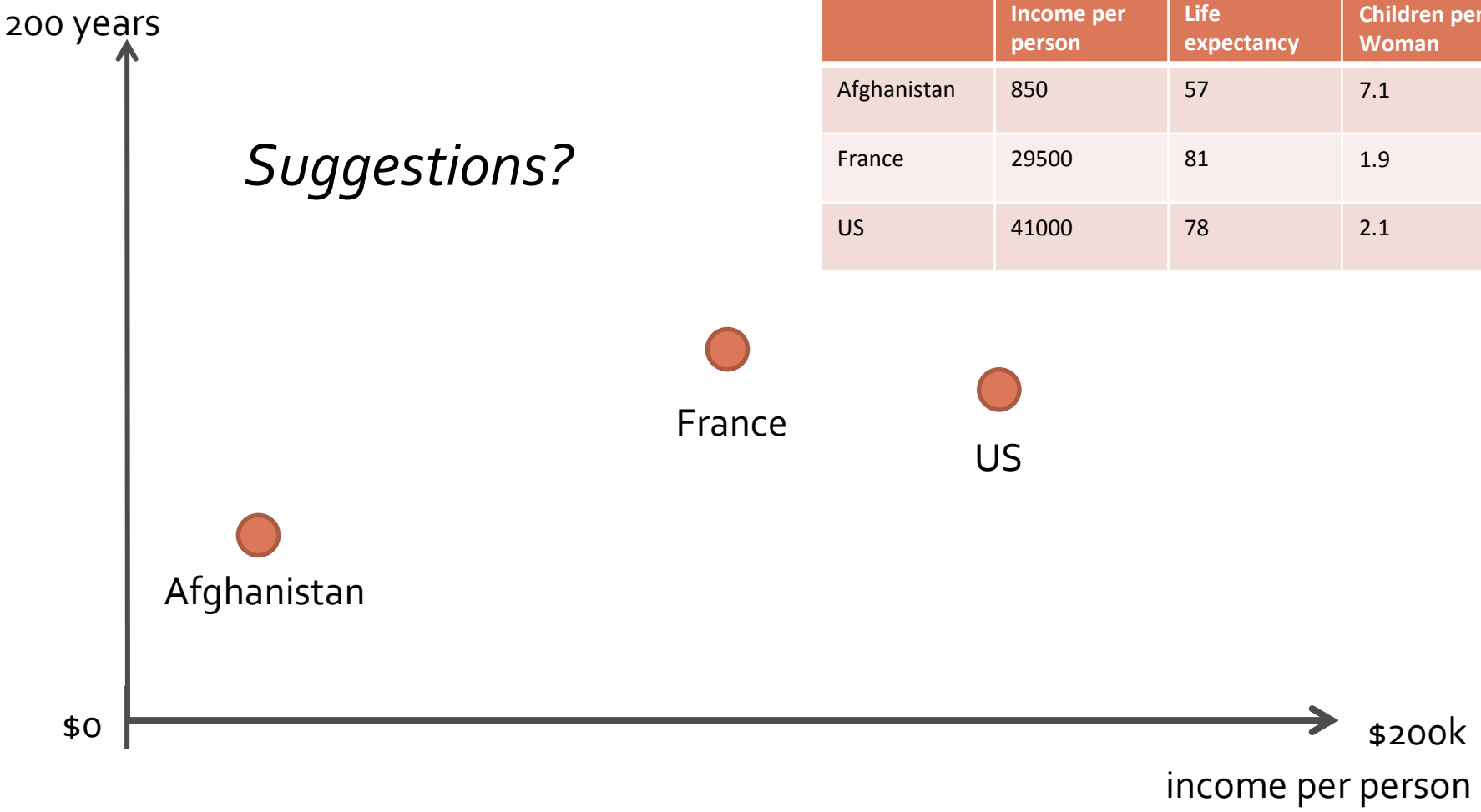
*Think about it!*

# *adding a third dimension*

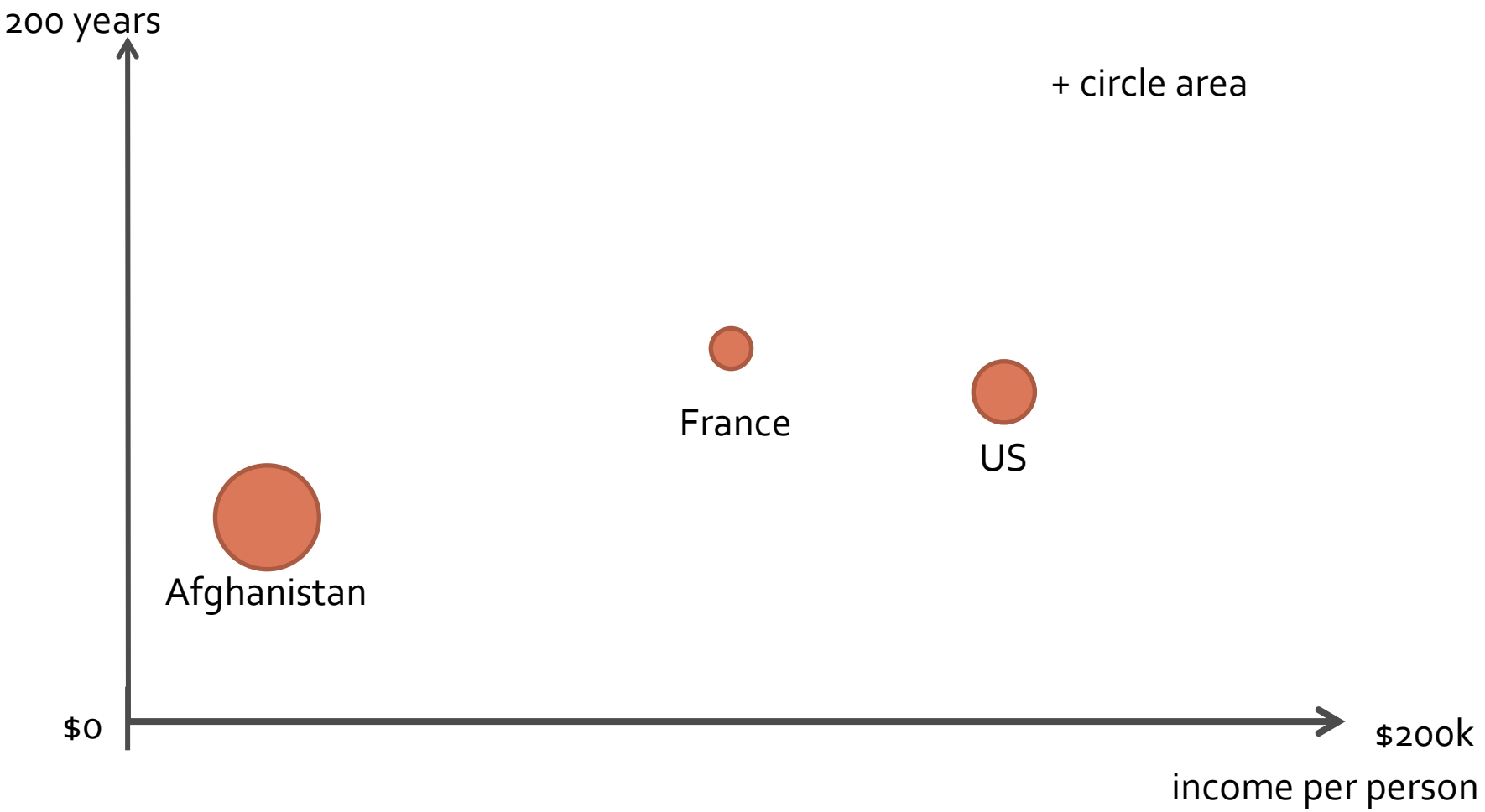
| | Income per person | Life expectancy | Children per Woman |
|---|---|---|---|
| Afghanistan | 850 | 57 | 7.1 |
| France | 29500 | 81 | 1.9 |
| US | 41000 | 78 | 2.1 |

# *adding a third dimension*

| | Income per person | Life expectancy | Children per Woman |
|---|---|---|---|
| Afghanistan | 850 | 57 | 7.1 |
| France | 29500 | 81 | 1.9 |
| US | 41000 | 78 | 2.1 |

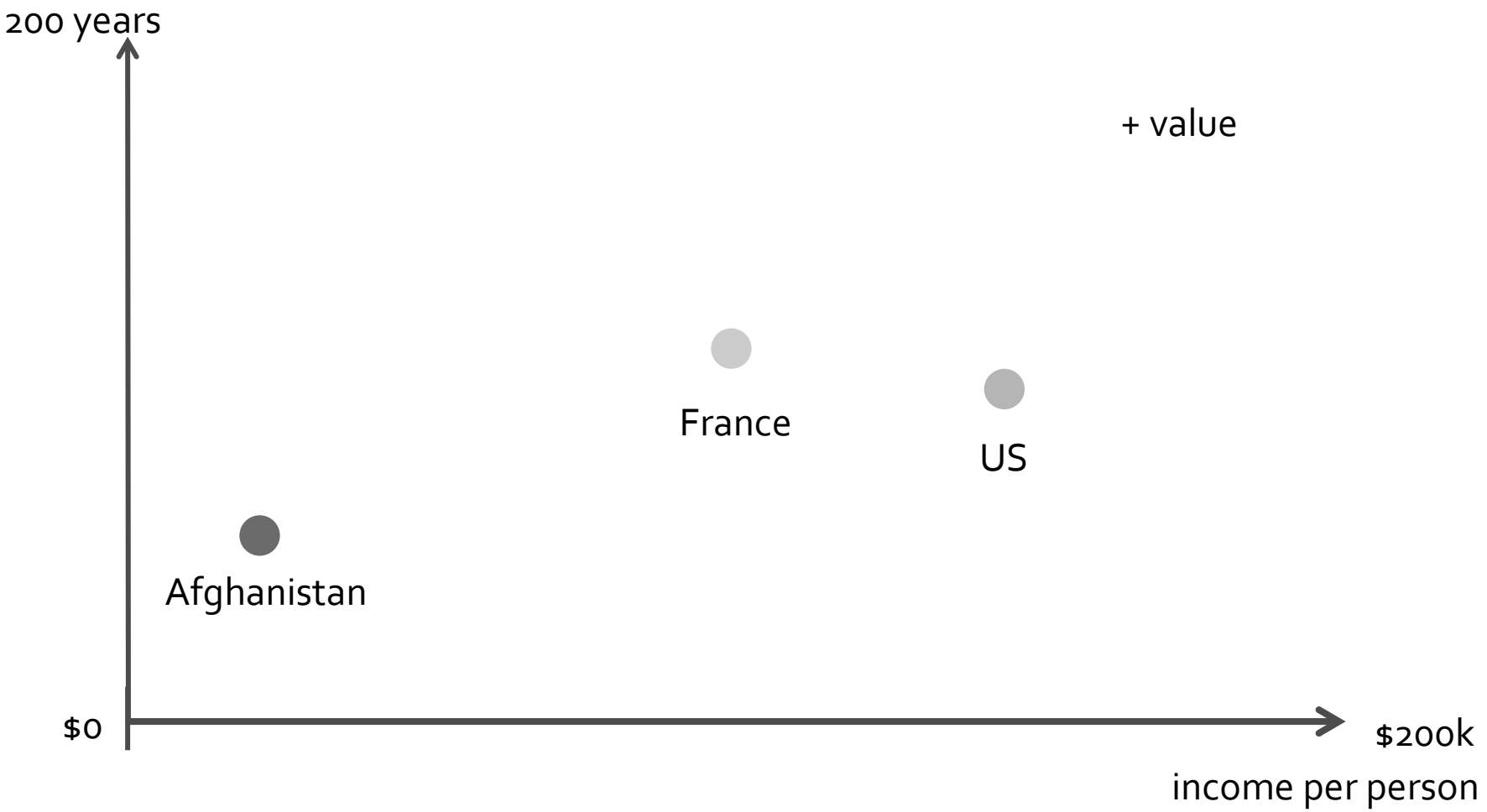


# *add another visual encoding*



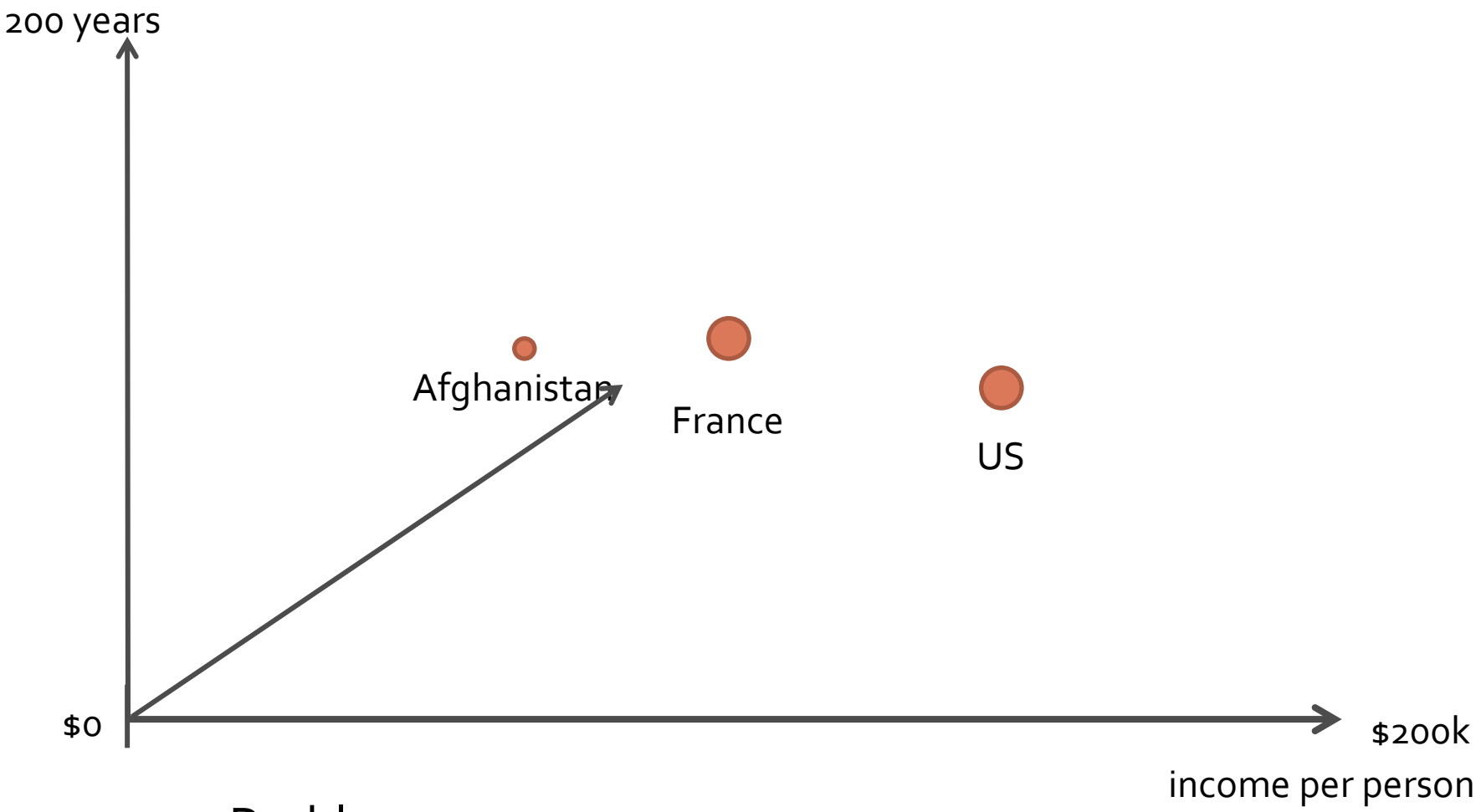

Problem:
Does not scale well to more dimensions

# *add another visual encoding*




Problem:
Does not scale well to more dimensions

# *add an axis*

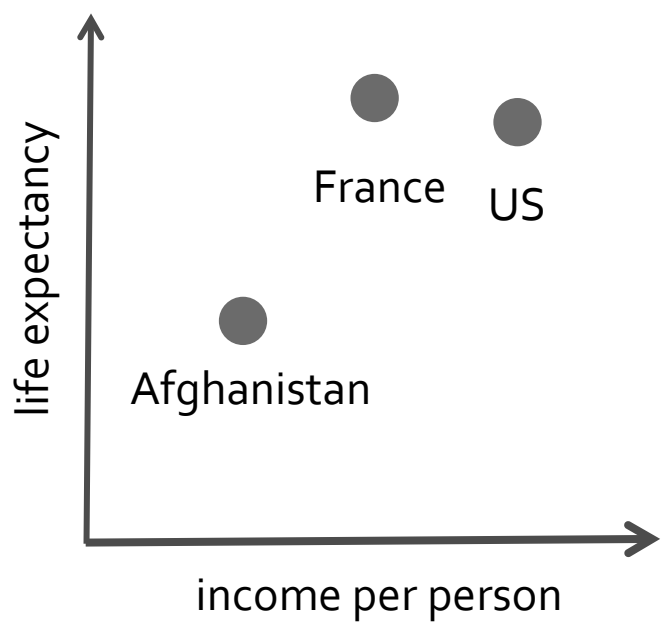


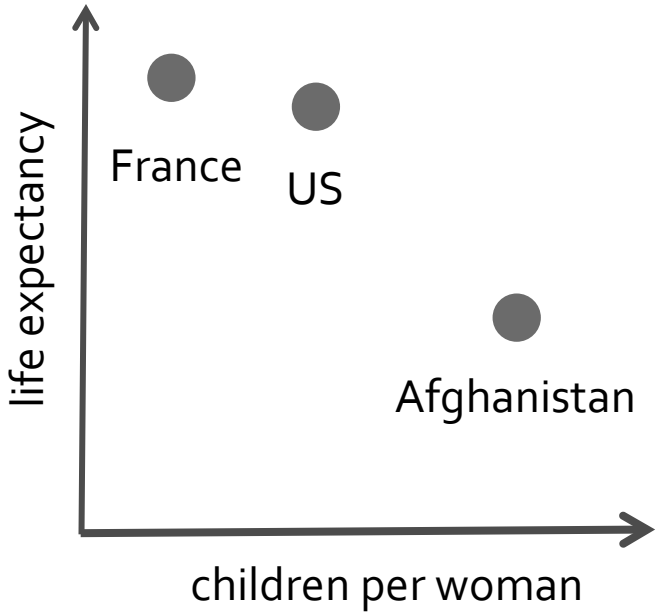

Problems:

Occlusion, perspective distortion, does not scale

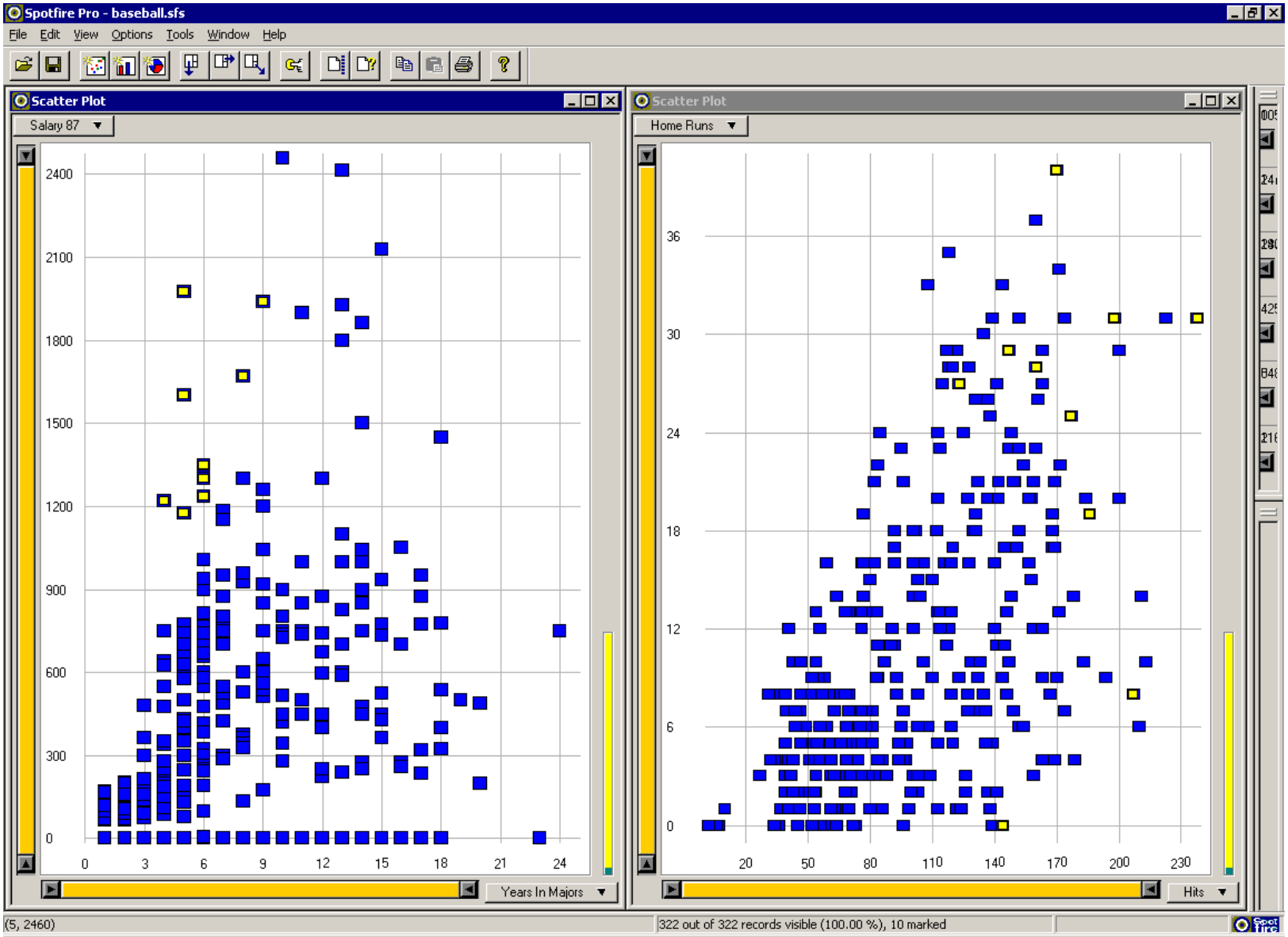
→ Not usually recommended

# *add an axis*


life expectancy
France
US
Afghanistan
income per person


life expectancy
France
US
Afghanistan
children per woman

# *Linked Views with Brushing and Linking*



# *back to our original problem: 1. find a layout*

- We can also decide to display the whole data table and augment it

# *displaying the data table itself*

- Heatmap on individual cells
- (Excel does this)

| Rank | Player | Team | Pos | G | AB | R | H | 2B | 3B | HR | RBI | BB | SO | SB | CS | AVG | OBP | SLG | OPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Posey, B | SF | C | 148 | 530 | 78 | 178 | 39 | 1 | 24 | 103 | 69 | 96 | 1 | 1 | 0.336 | 0.408 | 0.549 | 0.957 |
| 2 | Cabrera, M | DET | 3B | 161 | 622 | 109 | 205 | 40 | 0 | 44 | 139 | 66 | 98 | 4 | 1 | 0.33 | 0.393 | 0.606 | 0.999 |
| 3 | McCutchen, A | PIT | CF | 157 | 593 | 107 | 194 | 29 | 6 | 31 | 96 | 70 | 132 | 20 | 12 | 0.327 | 0.4 | 0.553 | 0.953 |
| 4 | Trout, M | LAA | CF | 139 | 559 | 129 | 182 | 27 | 8 | 30 | 83 | 67 | 139 | 49 | 5 | 0.326 | 0.399 | 0.564 | 0.963 |
| 5 | Beltre, A | TEX | 3B | 156 | 604 | 95 | 194 | 33 | 2 | 36 | 102 | 36 | 82 | 1 | 0 | 0.321 | 0.359 | 0.561 | 0.921 |
| 6 | Braun, R | MIL | LF | 154 | 598 | 108 | 191 | 36 | 3 | 41 | 112 | 63 | 128 | 30 | 7 | 0.319 | 0.391 | 0.595 | 0.987 |
| 7 | Mauer, J | MIN | C | 147 | 545 | 81 | 174 | 31 | 4 | 10 | 85 | 90 | 88 | 8 | 4 | 0.319 | 0.416 | 0.446 | 0.861 |
| 8 | Jeter, D | NYY | SS | 159 | 683 | 99 | 216 | 32 | 0 | 15 | 58 | 45 | 90 | 9 | 4 | 0.316 | 0.362 | 0.429 | 0.791 |
| 9 | Molina, Y | STL | C | 138 | 505 | 65 | 159 | 28 | 0 | 22 | 76 | 45 | 55 | 12 | 3 | 0.315 | 0.373 | 0.501 | 0.874 |
| 10 | Fielder, P | DET | 1B | 162 | 581 | 83 | 182 | 33 | 1 | 30 | 108 | 85 | 84 | 1 | 0 | 0.313 | 0.412 | 0.528 | 0.94 |
| 11 | Hunter, T | LAA | CF | 140 | 534 | 81 | 167 | 24 | 1 | 16 | 92 | 38 | 133 | 9 | 1 | 0.313 | 0.365 | 0.451 | 0.817 |
| 12 | Butler, B | KC | 1B | 161 | 614 | 72 | 192 | 32 | 1 | 29 | 107 | 54 | 111 | 2 | 1 | 0.313 | 0.373 | 0.51 | 0.882 |
| 13 | Cano, R | NYY | 2B | 161 | 627 | 105 | 196 | 48 | 1 | 33 | 94 | 61 | 96 | 3 | 2 | 0.313 | 0.379 | 0.55 | 0.929 |
| 14 | Pacheco, J | COL | 3B | 132 | 475 | 51 | 147 | 32 | 3 | 5 | 54 | 22 | 61 | 7 | 2 | 0.309 | 0.341 | 0.421 | 0.762 |
| 15 | Craig, A | STL | RF | 119 | 469 | 76 | 144 | 35 | 0 | 22 | 92 | 37 | 89 | 2 | 1 | 0.307 | 0.354 | 0.522 | 0.876 |
| 16 | Scutaro, M | SF | 3B | 156 | 620 | 87 | 190 | 32 | 4 | 7 | 74 | 40 | 49 | 9 | 4 | 0.306 | 0.348 | 0.405 | 0.753 |
| 17 | Wright, D | NYM | 3B | 156 | 581 | 91 | 178 | 41 | 2 | 21 | 93 | 81 | 112 | 15 | 10 | 0.306 | 0.391 | 0.492 | 0.883 |
| 18 | Jay, J | STL | CF | 117 | 443 | 70 | 135 | 22 | 4 | 4 | 40 | 34 | 71 | 19 | 7 | 0.305 | 0.373 | 0.4 | 0.773 |
| 19 | Murphy, D | TEX | LF | 147 | 457 | 65 | 139 | 29 | 3 | 15 | 61 | 54 | 74 | 10 | 5 | 0.304 | 0.38 | 0.479 | 0.859 |
| 20 | Rios, A | CWS | CF | 157 | 605 | 93 | 184 | 37 | 8 | 25 | 91 | 26 | 92 | 23 | 6 | 0.304 | 0.334 | 0.516 | 0.85 |
| 21 | Gonzalez, C | COL | LF | 135 | 518 | 89 | 157 | 31 | 5 | 22 | 85 | 56 | 115 | 20 | 5 | 0.303 | 0.371 | 0.51 | 0.881 |

Imagesource: dataremixed.com

# *table lens*

Ramana Rao, Stuart K. Card, Peter Pirolli (Xerox Parc) – 1994-96

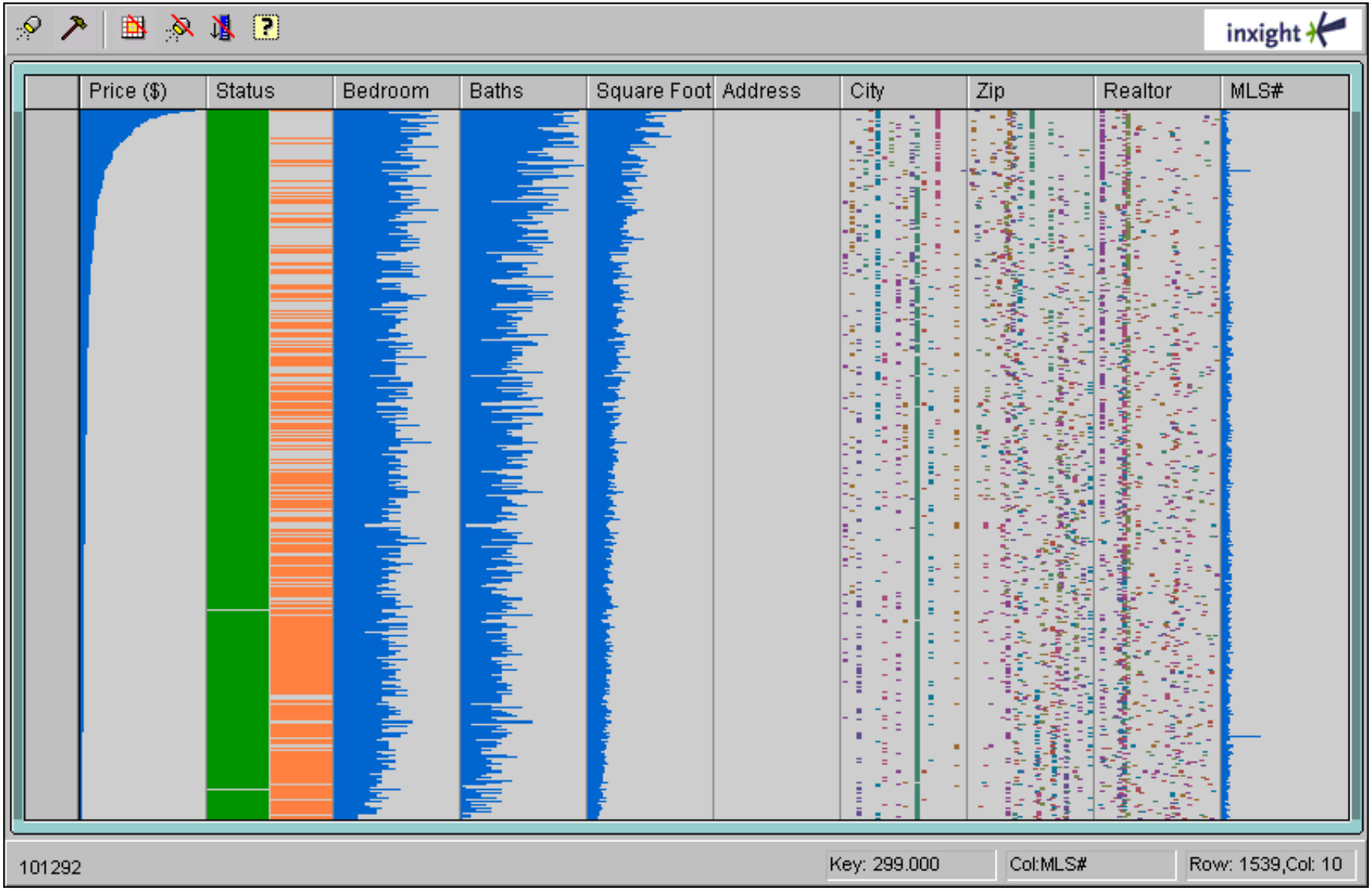



Image source: http://www.sapdesignguild.org/goodies/controls/TableLens.htm

# *Classification of MD Vis Techniques*

- Geometric
- Icon-based
- Pixel-oriented
- Hierarchical
- Graph-based

# *Classification of MD Vis Techniques*

- Geometric
- Icon-based
- Pixel-oriented
- Hierarchical
- Graph-based

# Classification of MD Vis Techniques

## Geometric techniques

# *scatterplot matrix*

- This idea scales relatively well

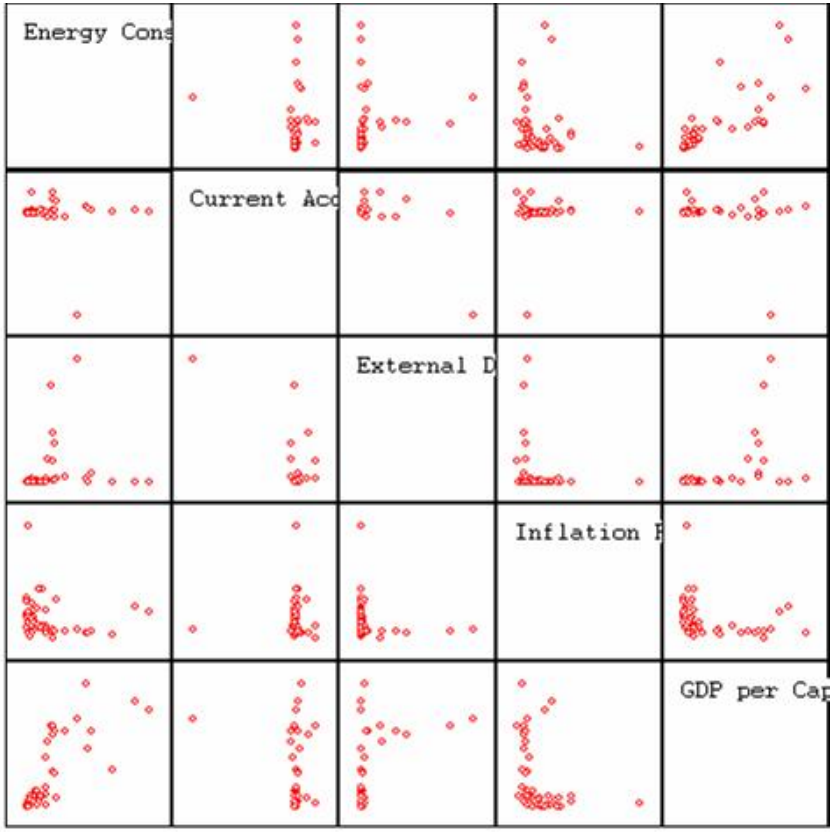
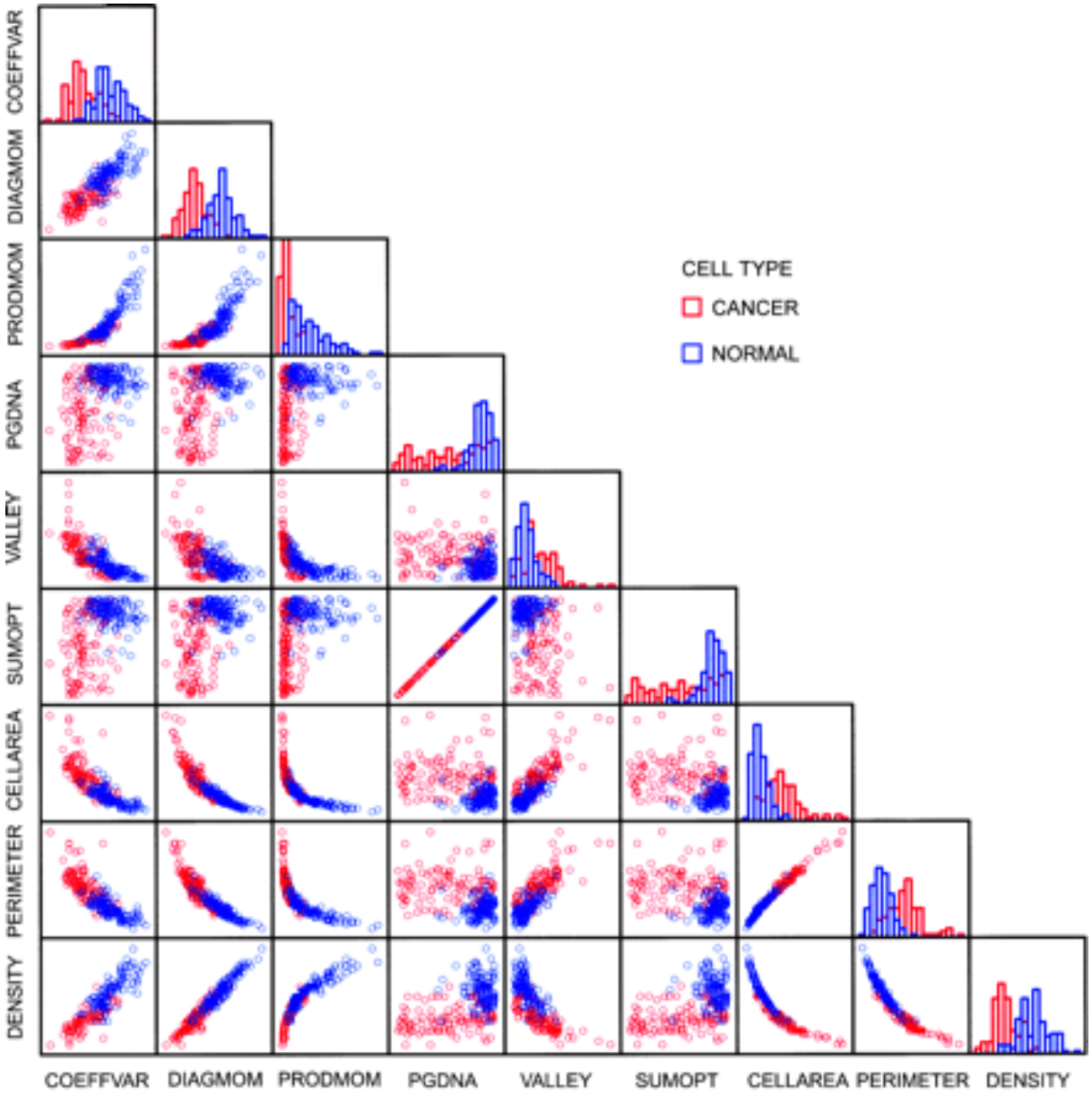
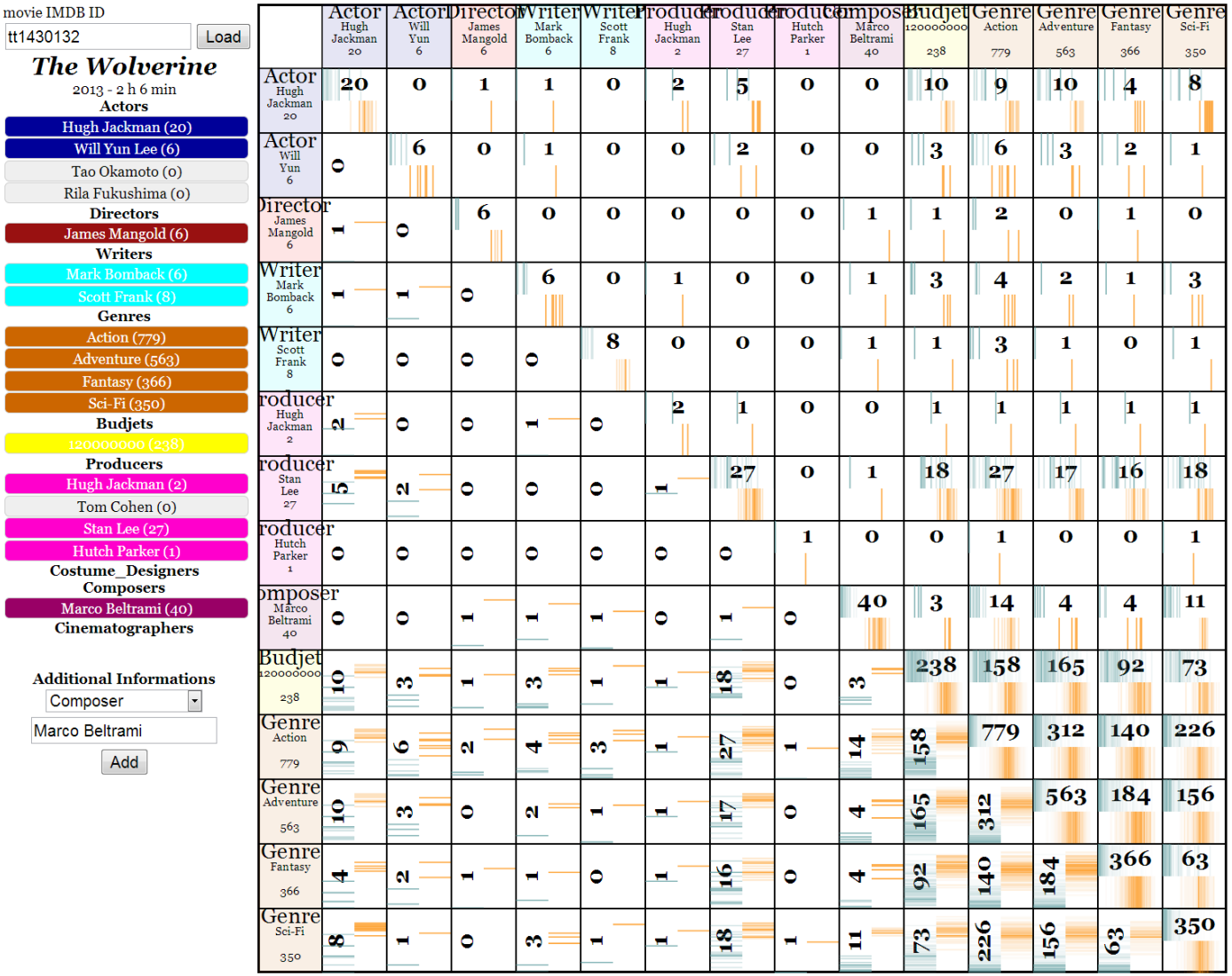



Image Source: Wikipedia


COEFFVAR
DIAGMOM
PRODMOM
PGDNA
VALLEY
SUMOPT
CELLAREA
PERIMETER
DENSITY
CELL TYPE
CANCER
NORMAL

# *scatterplot matrix*




By Charles Perin

# *ScatterDice*

# *small multiple* *(sometimes called trellis chart, lattice chart, grid chart, panel chart)*

- Series of related graphs/charts/plots typically layed out in a grid-structure
- A Scatterplot matrix is a type of Small Multiple – but you can create and use many types of graphs and charts in this manner from your MD datasets

Concept introduced by Edward Tufte

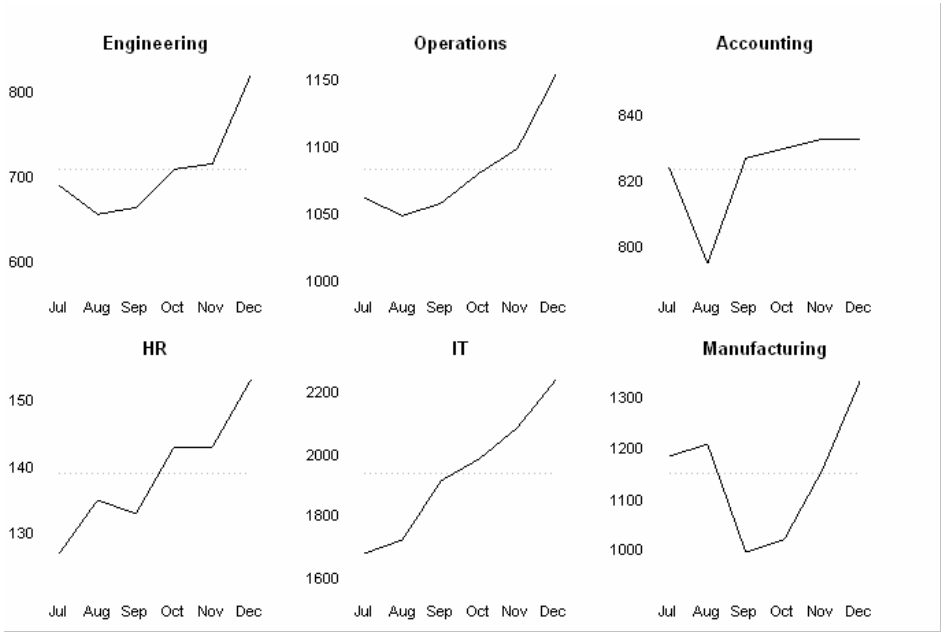
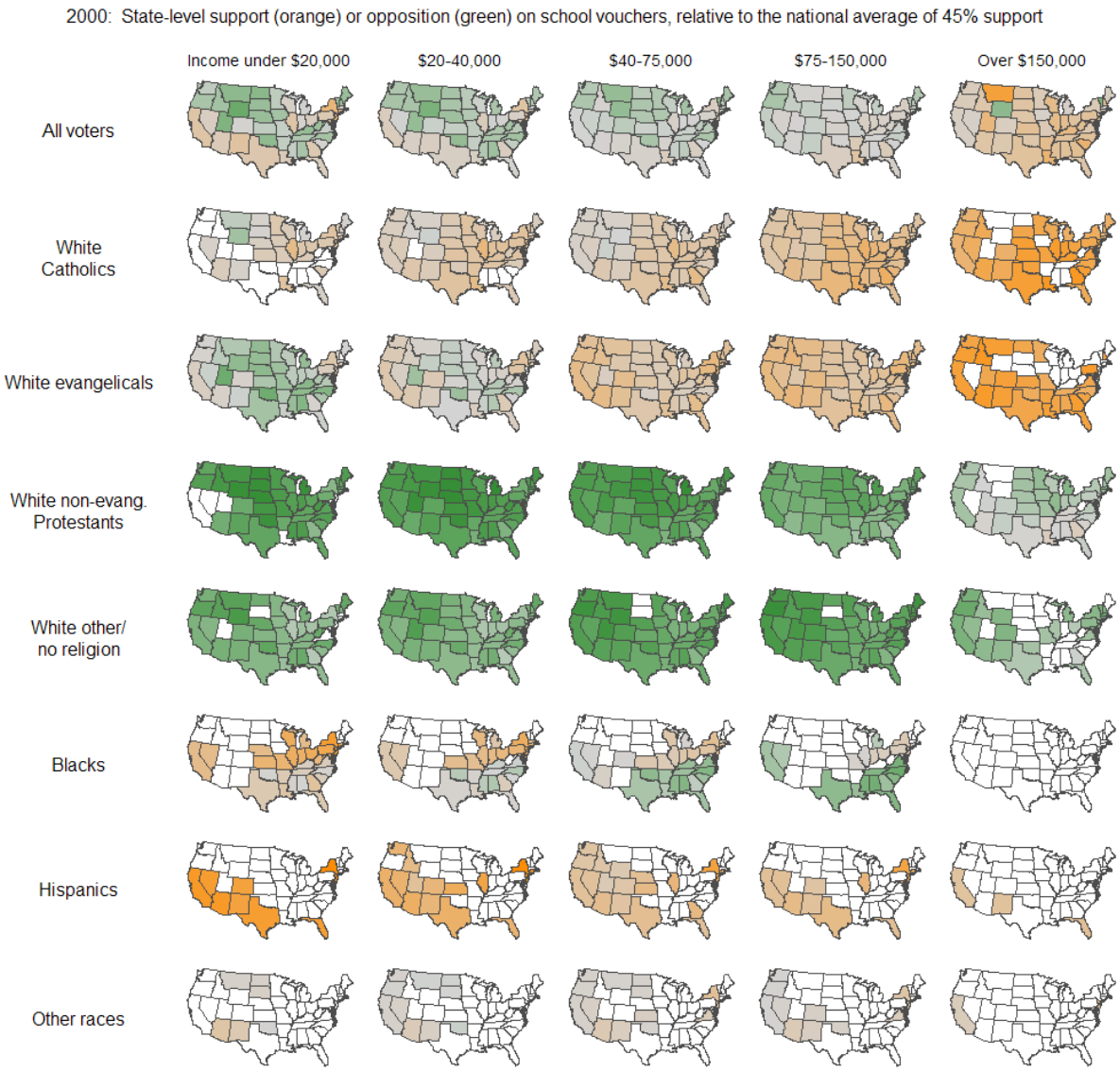


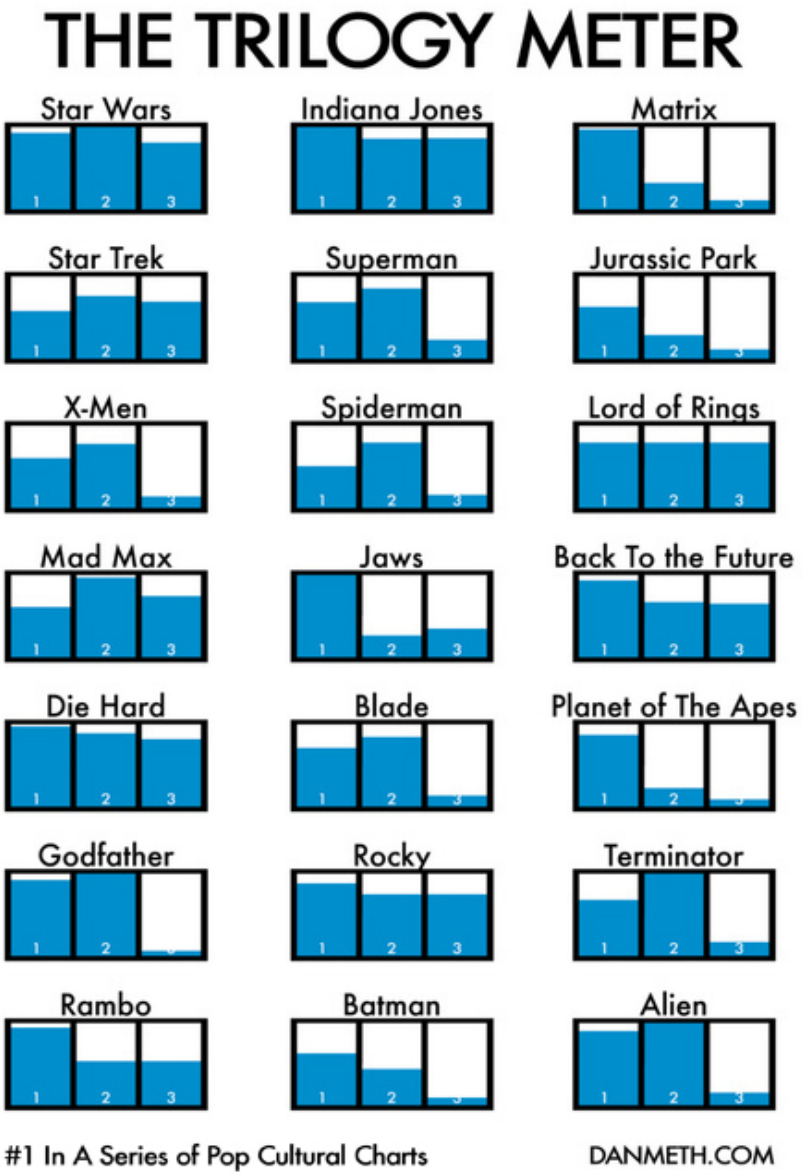

Image Source: Wikipedia

# *other examples: sm + maps*




source: http://andrewgelman.com/2009/07/15/hard_sell_for_b/

# *other examples: sm + bar charts*



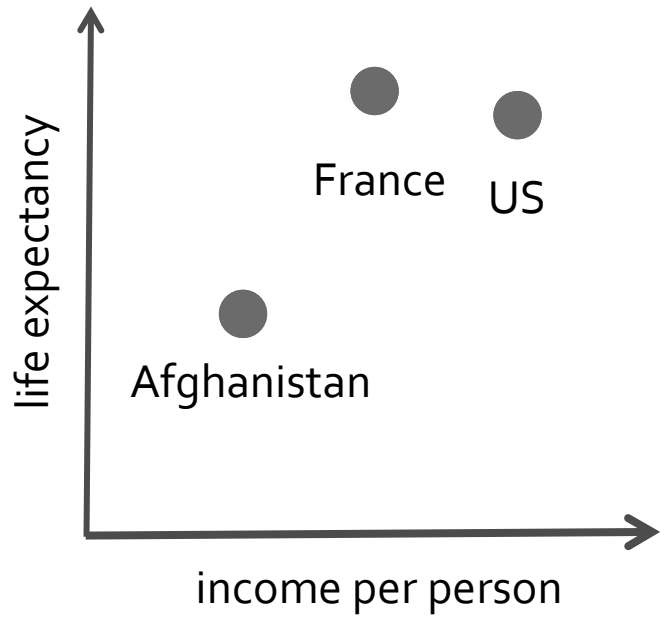

http://danmeth.com/post/7747620/my-trilogy-meter-1-in-a-series-of-pop-cultural

# *design choices to consider for small multiples*

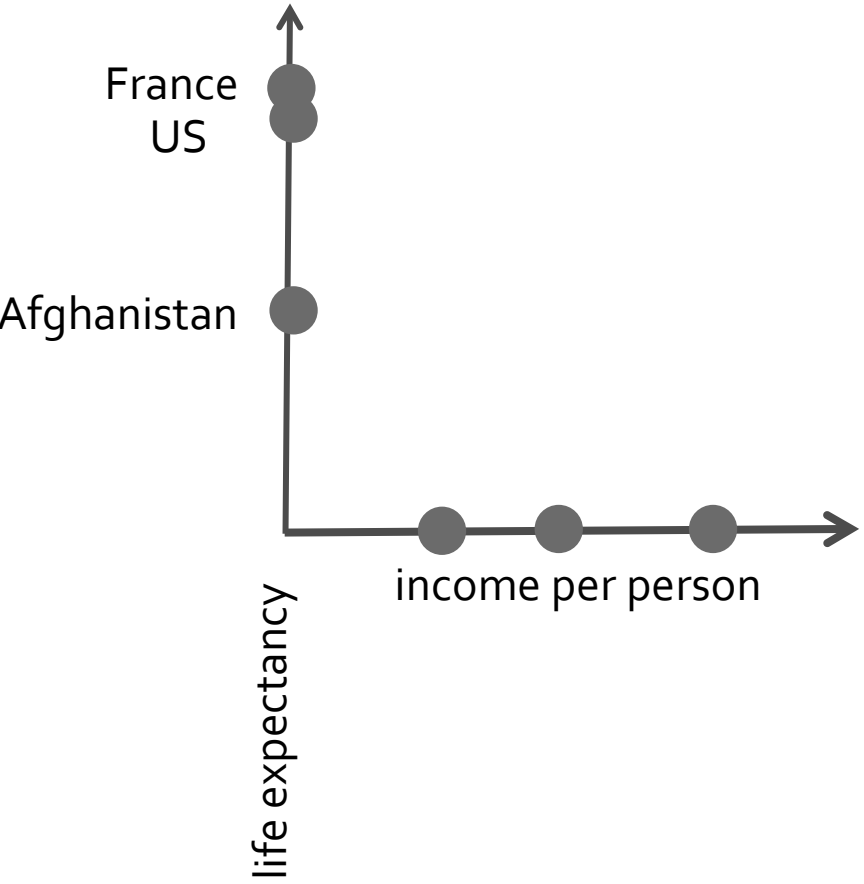
- Order the individual charts logically
- Use appropriate charts for your data
  - use comparable scales
  - use comparable encodings
- Use the same size for all charts & choose an appropriate size

# *Parallel Coordinates*

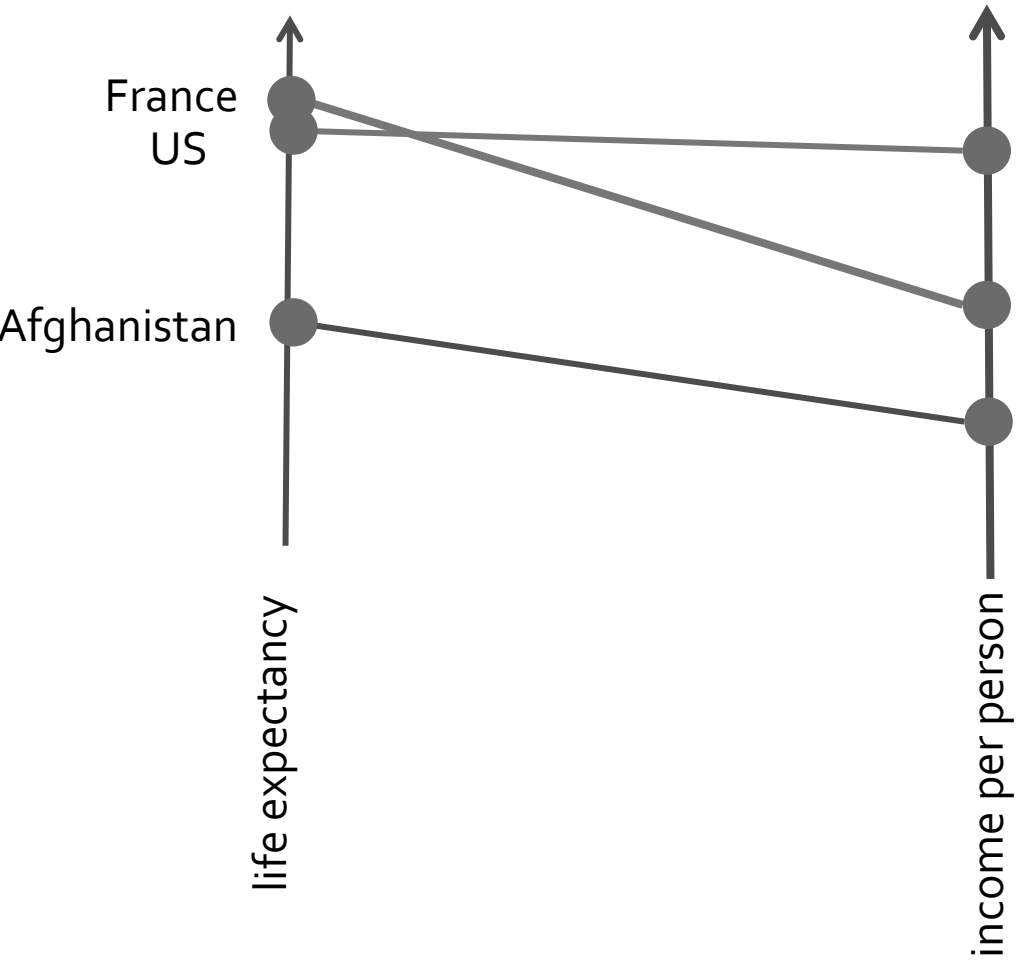
Back to our original example



# *Parallel Coordinates*



# *Parallel Coordinates*



# *Parallel Coordinates*

## MULTIDIMENSIONAL DETECTIVE

Alfred Inselberg*, Multidimensional Graphs Ltd†
&
Computer Science Department
Tel Aviv University, Israel
aiisreal@math.tau.ac.il

### Abstract

The display of multivariate datasets in parallel coordinates, transforms the search for *relations* among the variables into a 2-D pattern recognition problem. This is the basis for the application to *Visual Data Mining*. The Knowledge Discovery process together with some general guidelines are illustrated on a dataset from the production of a VLSI chip. The special strength of parallel coordinates is in modeling **relations**. As an example, a simplified Economic Model is constructed with data from various economic sectors of a real country. The visual model shows the interelationship and dependencies between the sectors, circumstances where there is competition for the same resource, and feasible economic policies. Interactively, the model can be used to do trade-off analyses, discover sensitivities, do approximate optimization, monitor (as in a Process) and Decision Support.

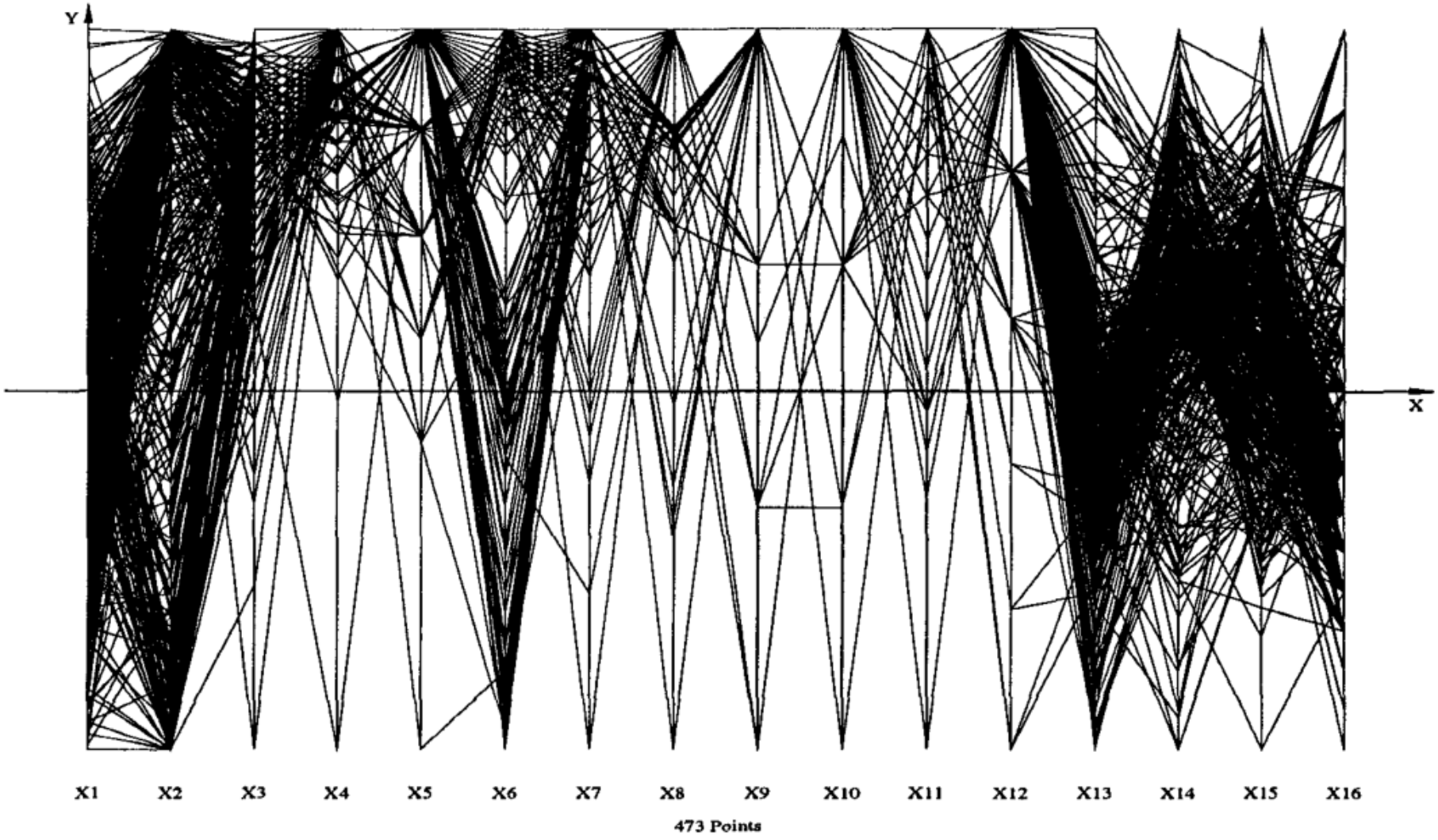
### Introduction

In Geometry parallelism, which does not require a notion of angle, rather than orthogonality is the more fundamental concept. This, together with the fact that orthogonality "uses-up" the plane very fast, was the inspiration in 1959 for "Parallel" Coordinates. The systematic development began in 1977 [4]. The goals of the program were and still are (see [6] and [5] for short reviews) the visualization of multivariate/multidimensional problems without loss of information and having the properties:

1. Low representational complexity. Since the number of axes, N equals the number of dimensions (variables) the complexity is $O(N)$,
2. Works for any $N$,
3. Every variable is treated uniformly (unlike "Chernoff Faces" and various types of "glyphs"),
4. The displayed object can be recognized under projective transformations (i.e. rotation, translation, scaling, perspective),
5. The display easily/intuitively conveys information on the properties of the N-dimensional object it represents,
6. The methodology is based on rigorous mathematical and algorithmic results.

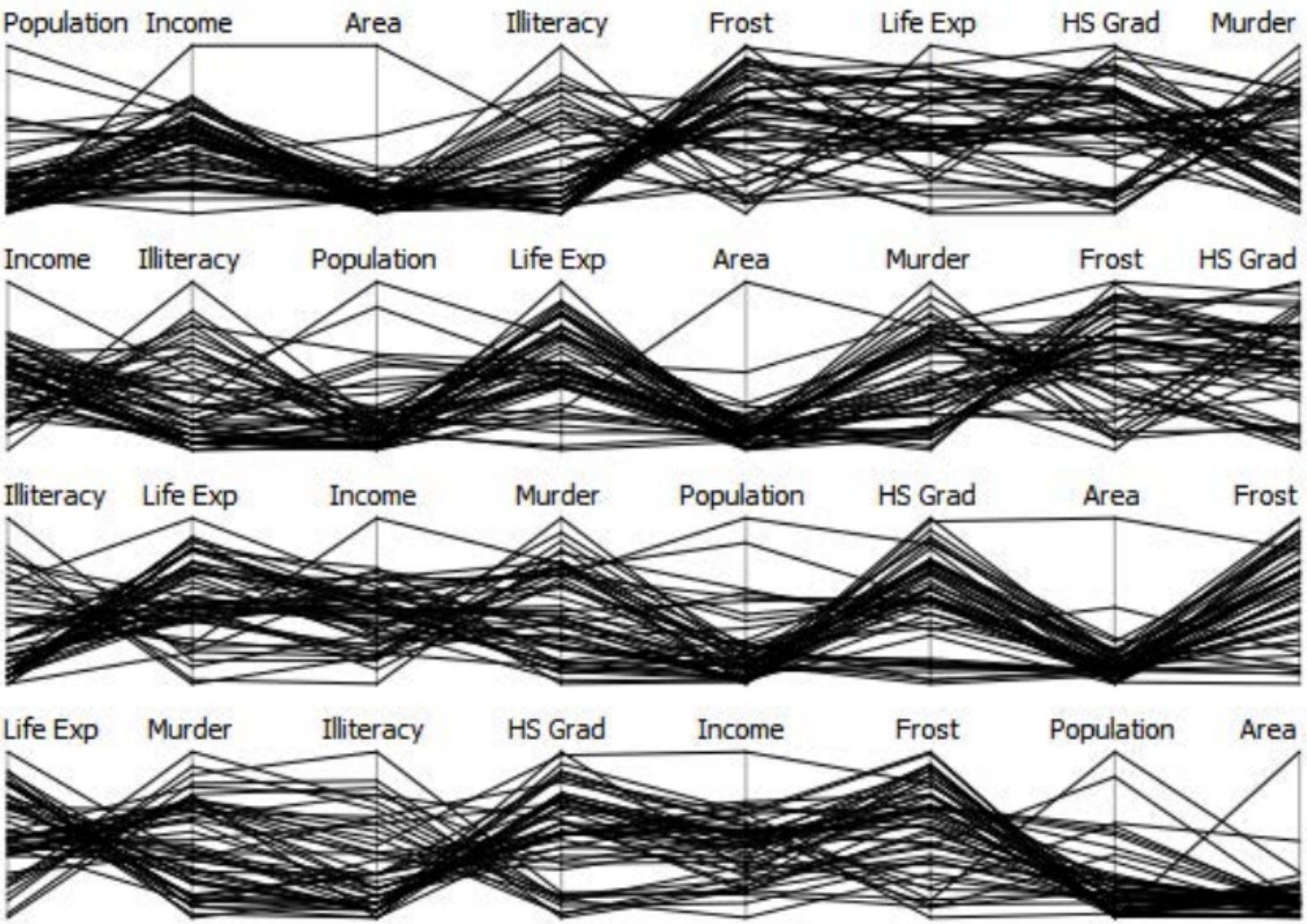
Parallel coordinates (abbr.||-coords) transform multivariate relations into 2-D patterns, a property that is well suited for Visual Data Mining.

*Senior Fellow San Diego SuperComputing Center
†36A Yehuda Halevy Street, Raanana 43556, Israel




Original Example from Inselberg 1997

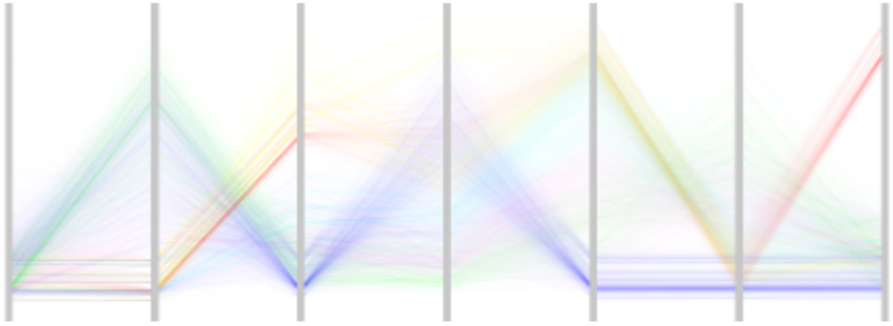
# *the order of axes matters*



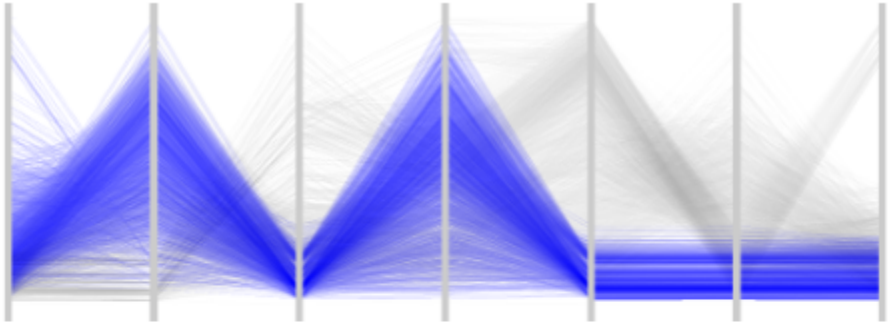

Eurographics 2013, STAR Report
J. Heinrich, D. Weiskopf

# *reduce clutter - highlight clusters*

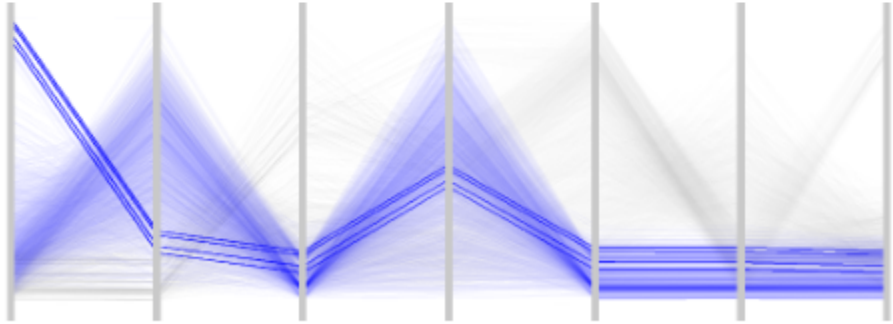
Lots of work on this. For example:



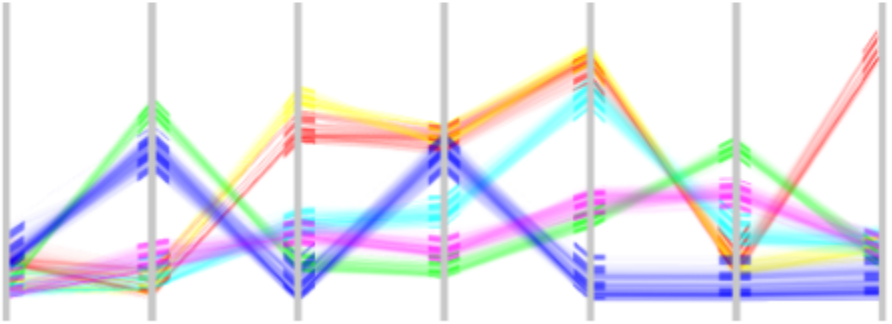
(a) A linear transfer function has been applied to the high-precision texture in order to prevent cluttering and to provide overview of the data.



(b) A logarithmic transfer function is applied to a selected cluster. The structure is preserved and emphasis is put on the low density regions.



(c) Local cluster outliers are enhanced. A square root transfer function is used and the outliers are visible even through high-density regions.

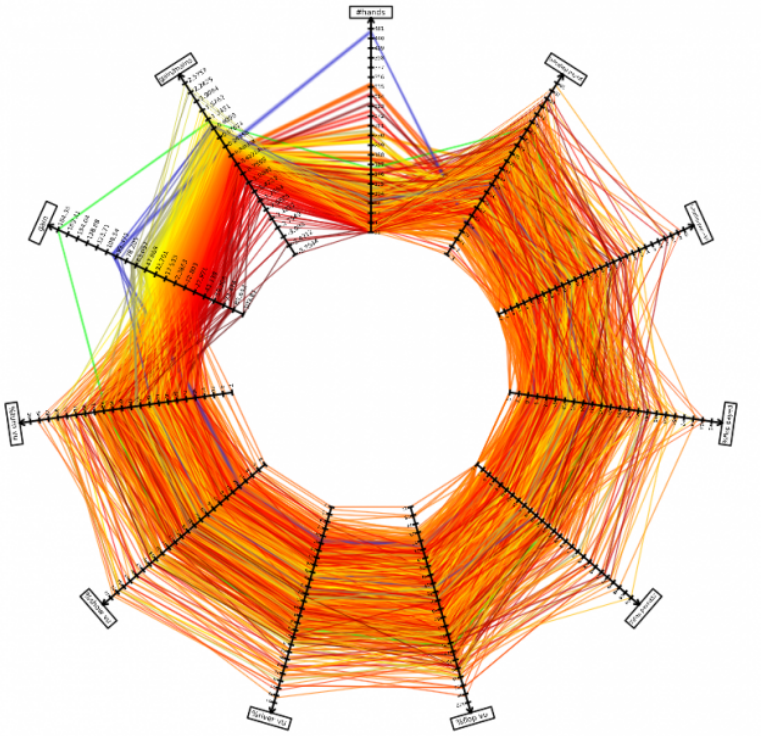


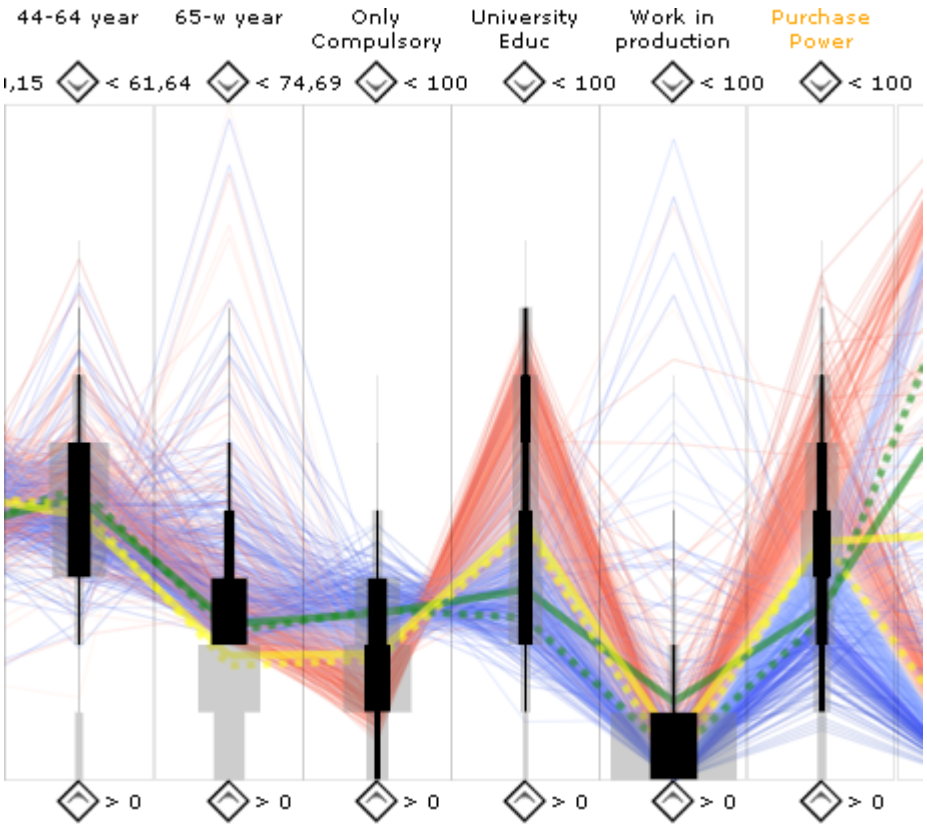
(d) A complementary view of the clusters with uniform bands. 'Feature animation' presents statistics about the clusters and acts as a guidance.

Revealing Structure within Clustered Parallel Coordinates Displays, InfoVis 2005

# *Other variants*

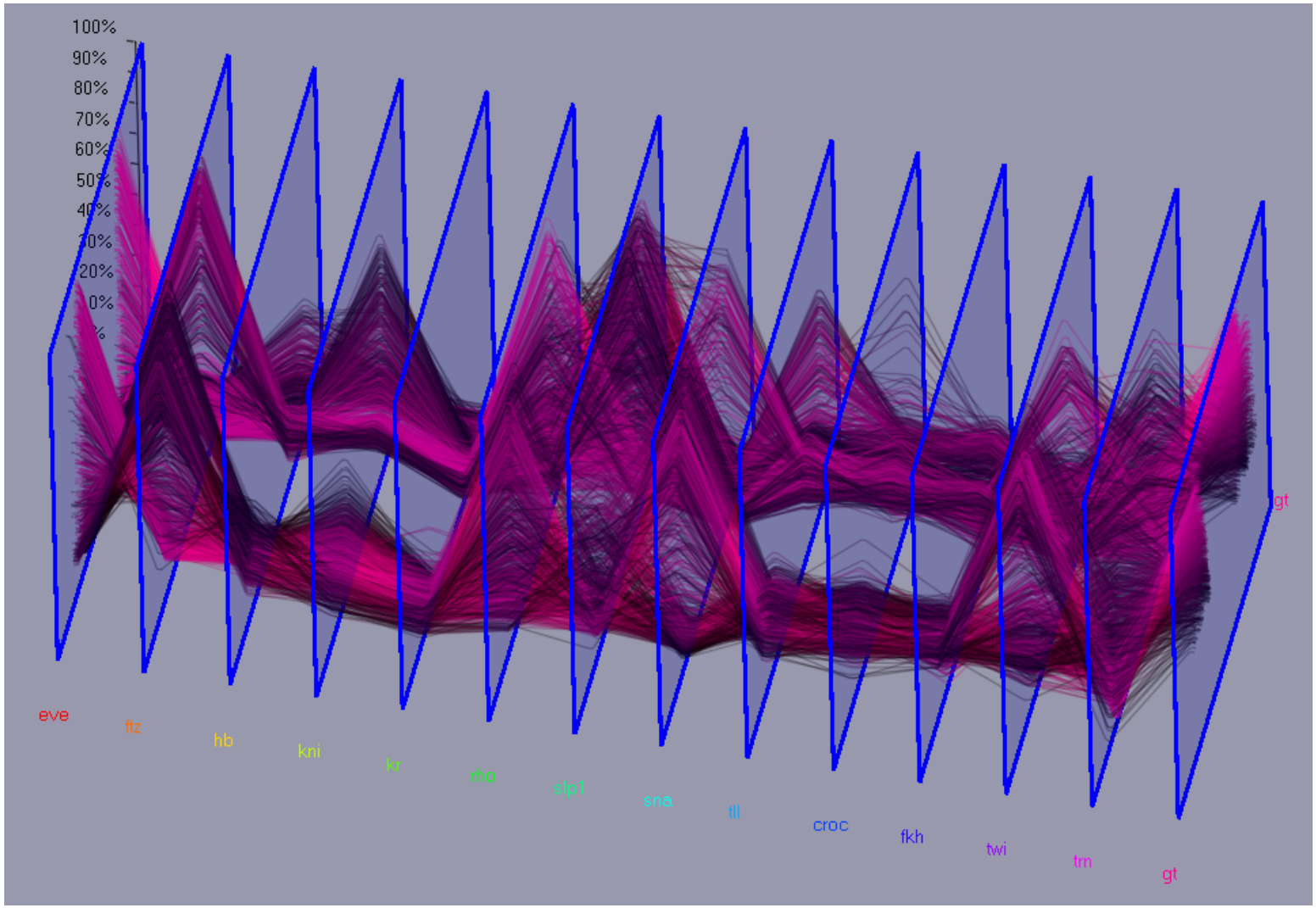
- Other layout, tulip (circular)
- PC with summary stats






http://tulip.labri.fr/TulipDrupal/?q=node/321

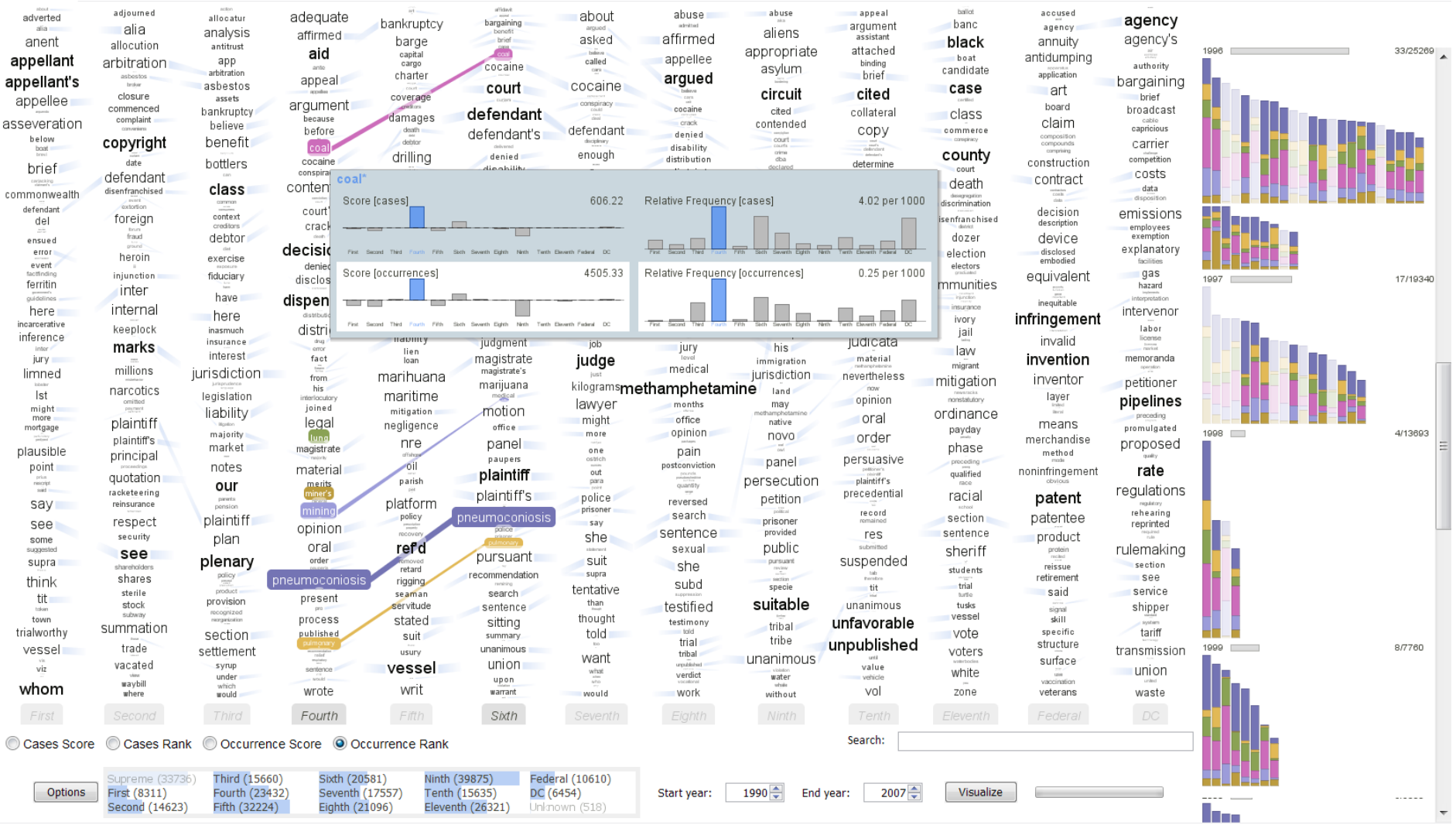
# 3D Parallel Coordinate




http://www-vis.lbl.gov/Events/SC05/Drosophilia/Parallel-Coordinated.png

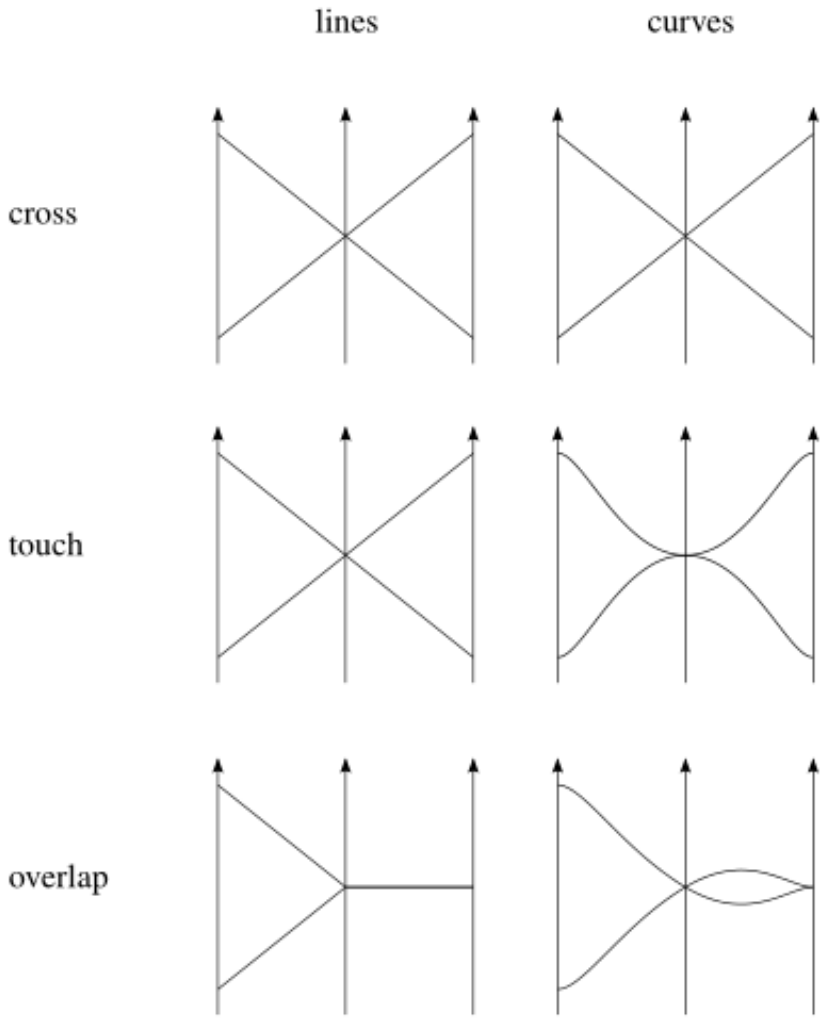
# *combine with other visualization techniques*

Parallel Tag Clouds to Explore Faceted Text Corpora (Collins et al., VAST 2009)



# *how to draw the lines*

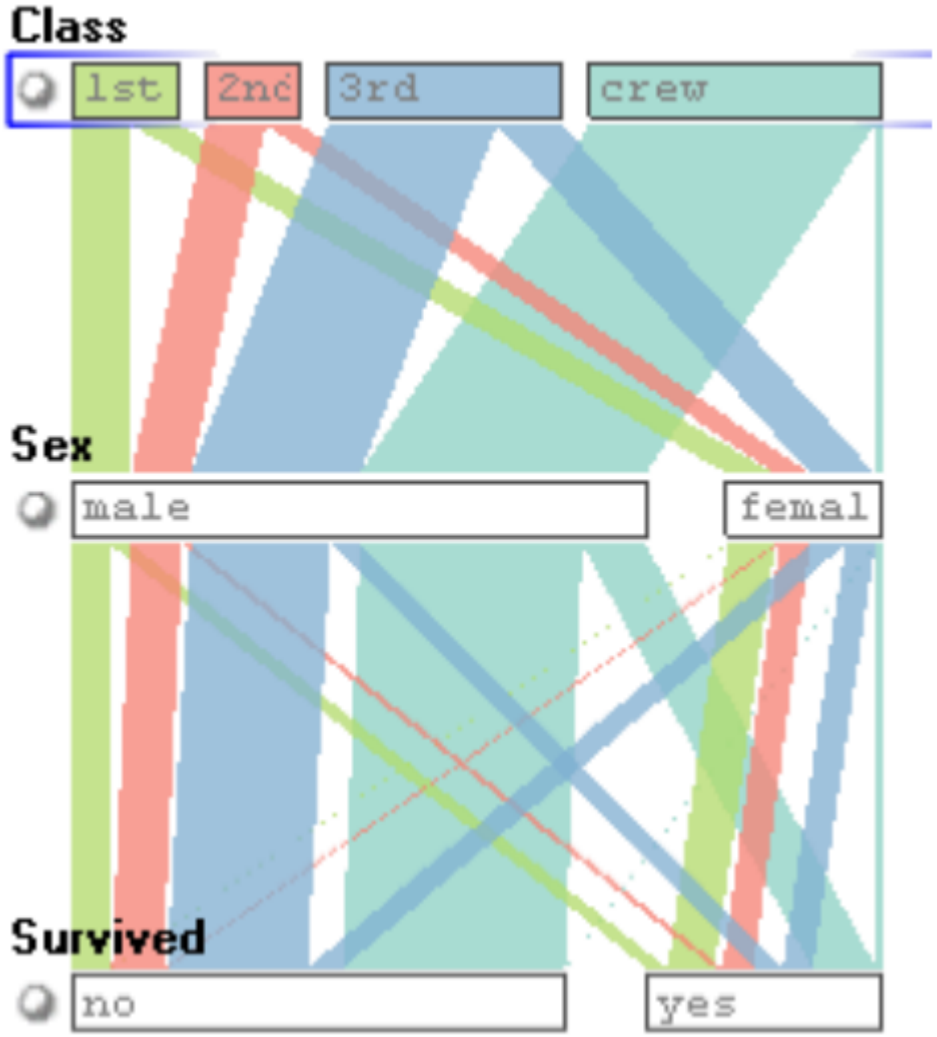
Goal: avoid ambiguity




Eurographics 2013, STAR Report
J. Heinrich, D. Weiskopf

*star glyphs + parallel coordinates*

# *Parallel sets – categorical data*




Fabian Bendix, Robert Kosara, Helwig Hauser, Infovis 05

# *there is much more on this…*

- Start here if you want more information



## State of the Art of Parallel Coordinates

J. Heinrich and D. Weiskopf

Visualization Research Center, University of Stuttgart

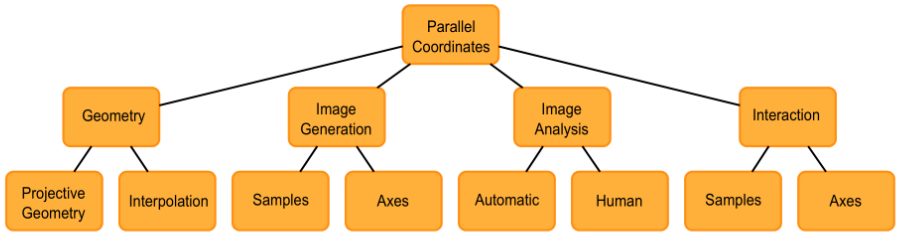



Figure 1: Taxonomy of topics for parallel coordinates in the scientific literature. The first-level nodes each represent a section in this paper, where the scope and definition of each topic will be explained.

**Abstract**

*This work presents a survey of the current state of the art of visualization techniques for parallel coordinates. It covers geometric models for constructing parallel coordinates and reviews methods for creating and understanding visual representations of parallel coordinates. The classification of these methods is based on a taxonomy that was established from the literature and is aimed at guiding researchers to find existing techniques and identifying white spots that require further research. The techniques covered in this survey are further related to an established taxonomy of knowledge-discovery tasks to support users of parallel coordinates in choosing a technique for their problem at hand. Finally, we discuss the challenges in constructing and understanding parallel-coordinates plots and provide some examples from different application domains.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

# *Paralle Coordinates*

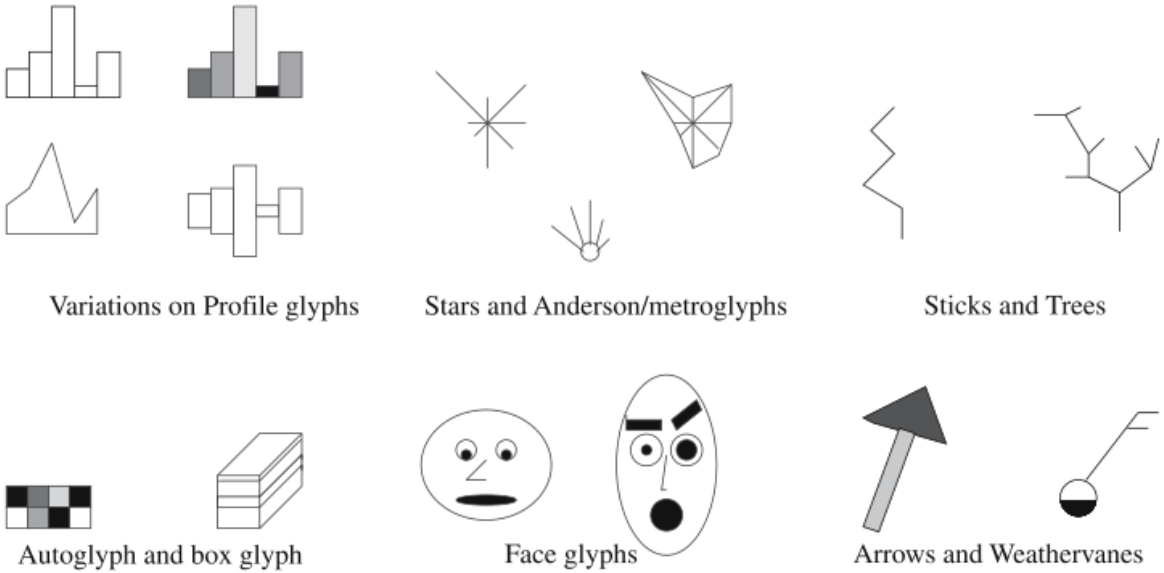
- Try it out
  - XmdvTool  http://davis.wpi.edu/%7Exmdv/index.html
  - Parvis http://www.mediavirus.org/parvis/
  - Macrofocus http://www.macrofocus.com/public/products/infoscope.html

# Classification of MD Vis Techniques

## Icon-based

# *glyphs*

- Small composite visual representations of multi-dimensional data points
- Often used in small multiple setting
- Characterized generally by lack of reference structures (grid lines, axes labels, …)




From Ward, 2002
A taxonomy of glyph placement strategies for multidimensional data visualization

# *mappings from dimension to visual encoding*

- One-to-one:
  - each dimension maps to a distinct and different visual encoding

# *example: chernoff faces*

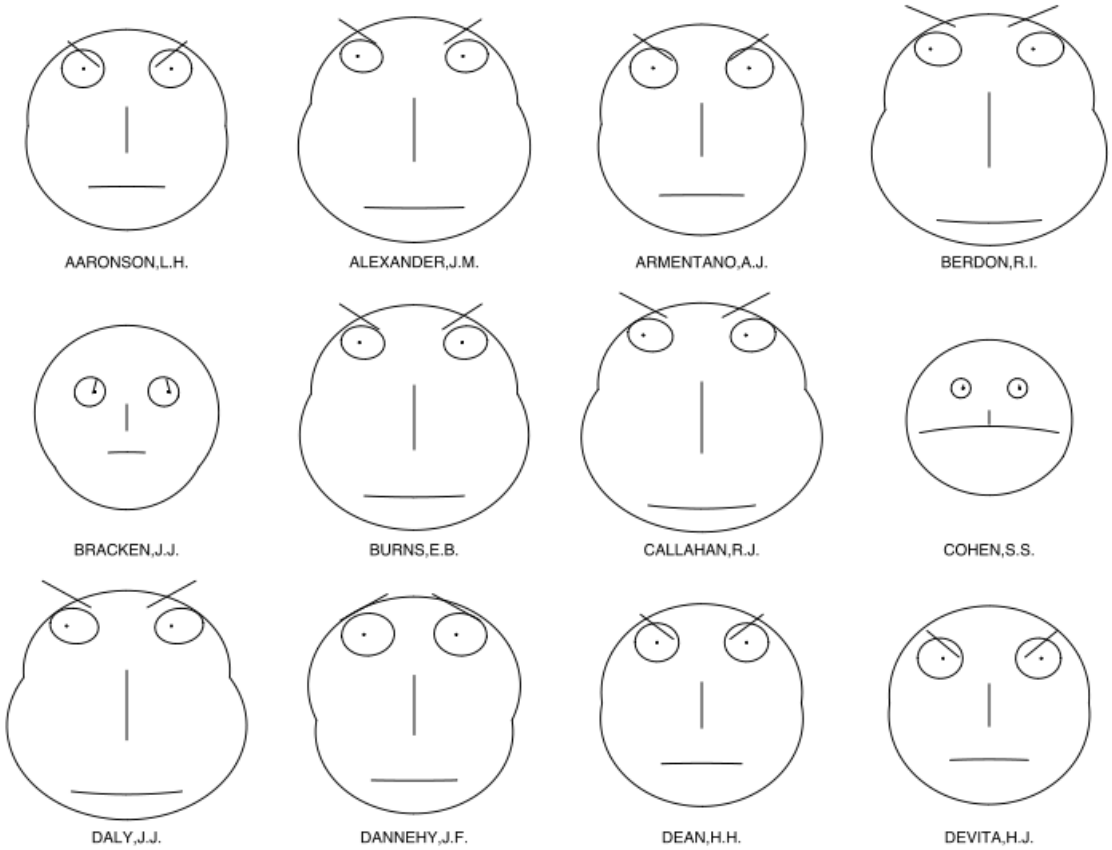


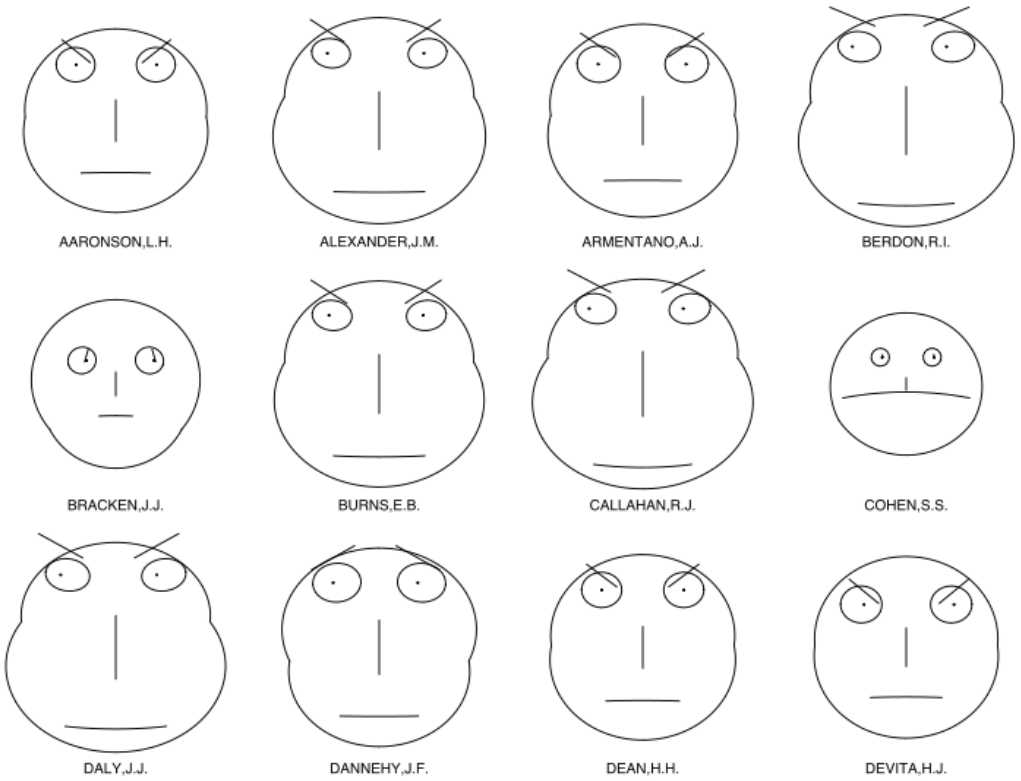

Image source: Wikipedia

Herman Chernoff,
The Use of Faces to Represent Points in K-Dimensional Space Graphically, 1973.

# *example: chernoff faces*

- features of a human face encode data values

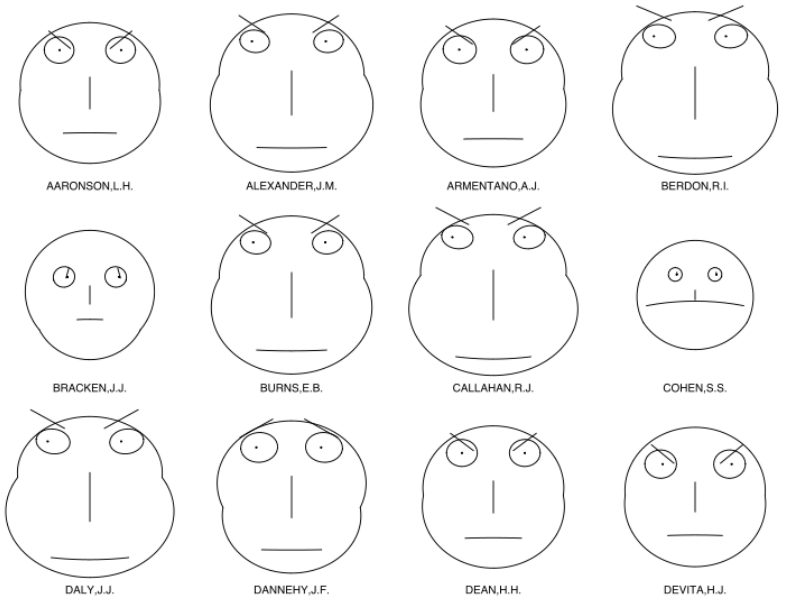
- 10 Parameters:
  - Head Eccentricity
  - Eye Eccentricity
  - Pupil Size
  - Eyebrow Slope
  - Nose Size
  - Mouth Vertical Offset
  - Eye Spacing
  - Eye Size
  - Mouth Width
  - Mouth Openness



# *chernoff faces*

- reasoning: humans are good at differentiating faces and reading face features

- problem: chernoff faces have generally found not to be very effective

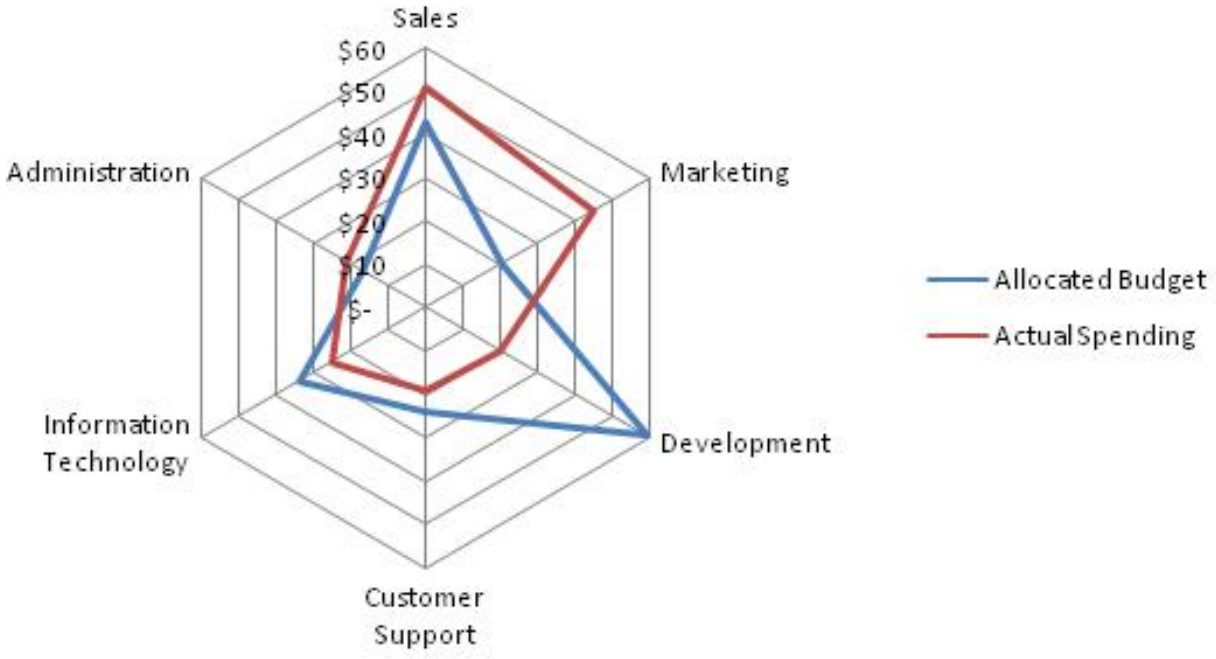


# *mappings from dimension to visual encoding*

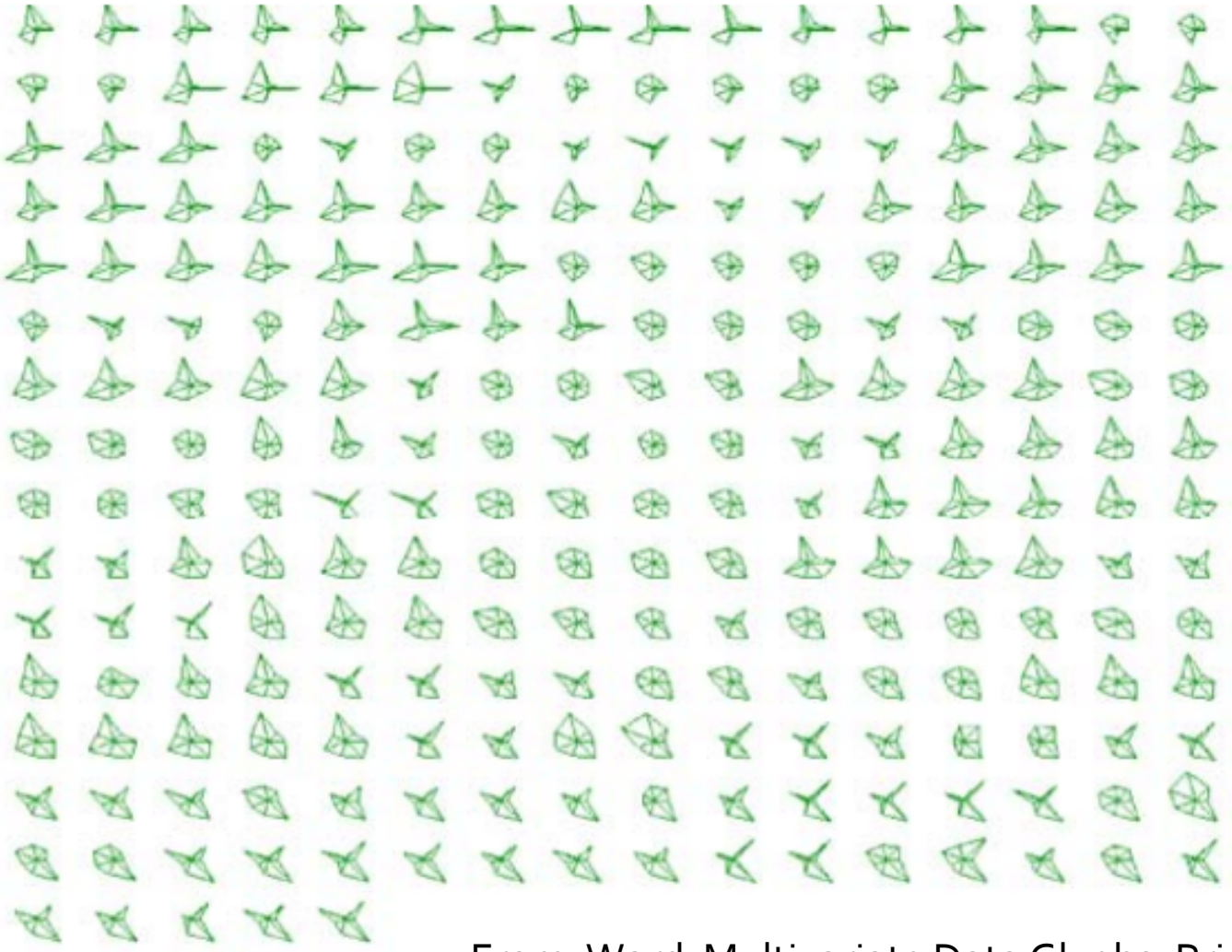
- One-to-one:
  - each dimension maps to a distinct and different visual encoding
- One-to-many
  - one dimension encoded in more than one way
  - used to improve accuracy & ease of interpreting the data (e.g. for color blind viewers)
- Many-to-one
  - several or all dimensions use the same visual encoding, separated in space, orientation, or otherwise

# *star glyphs*

- Lay out dimension in radial fashion
- Draw each point as a closed polyline



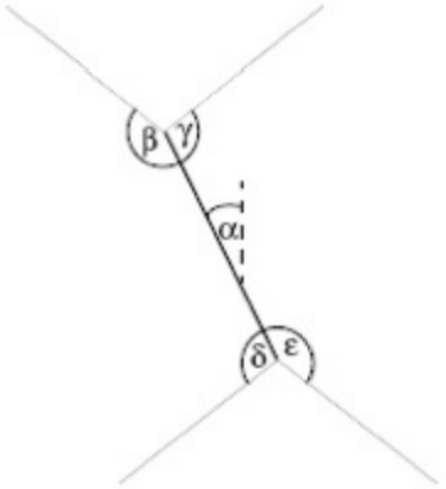
# *star glyphs*

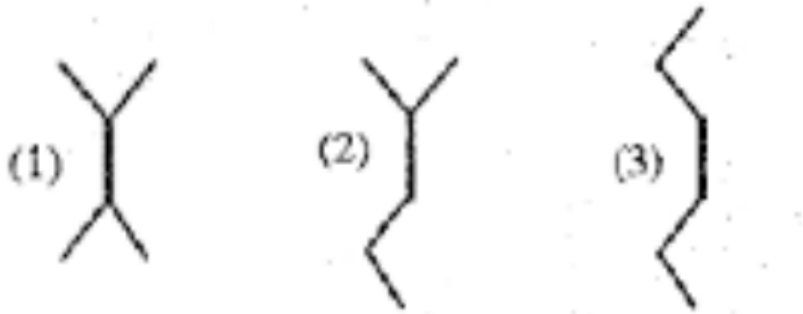


From: Ward  Multivariate Data Glyphs: Principles and Practice. Handbook of Data Visualization (2008)

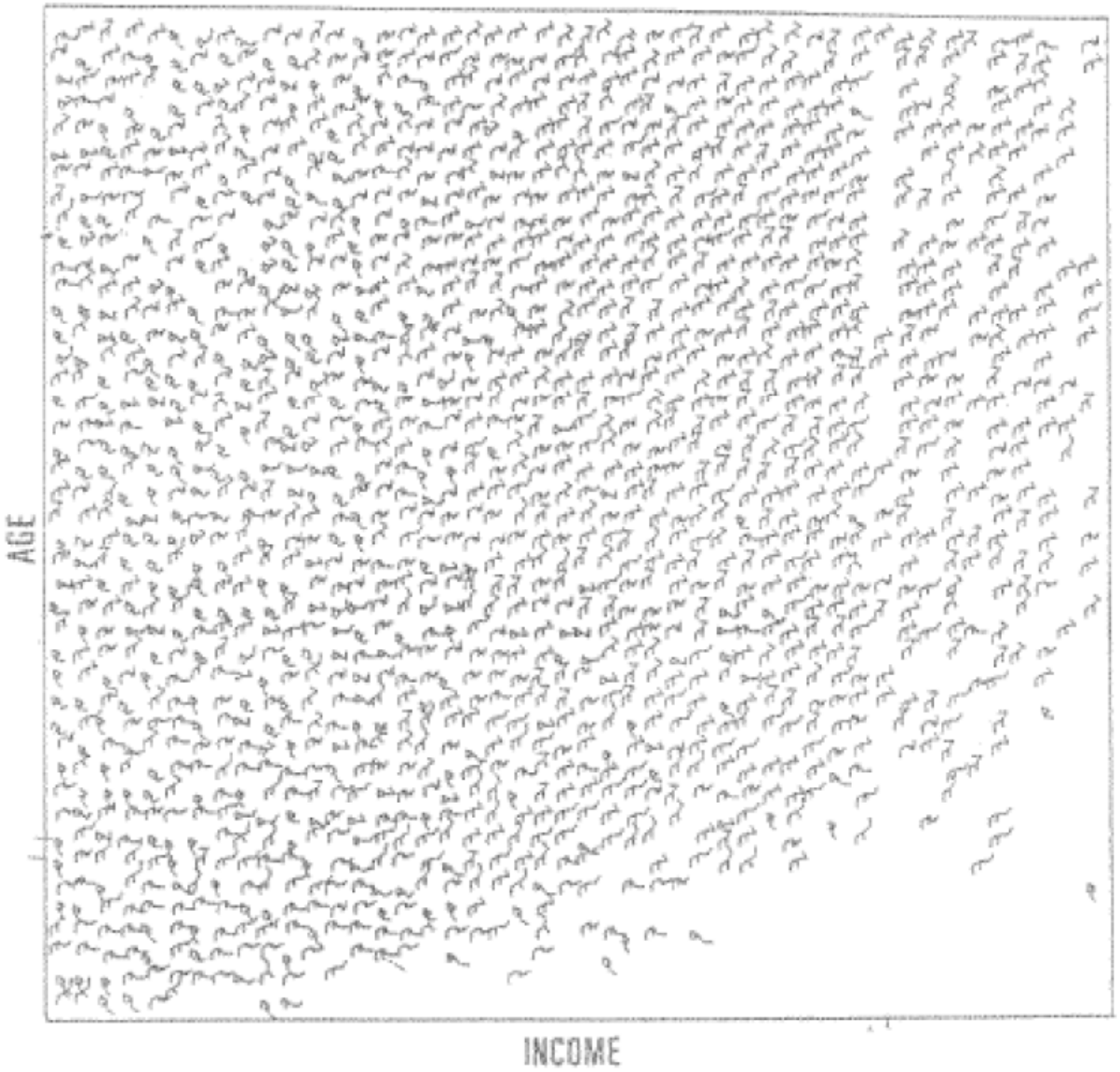
# *Stick figure Icons (Pickett & Grinstein)*

- two attributes are mapped to the display axes
- remaining attributes are mapped to lengths of limbs or angles between them
- Idea: Texture pattern in visualization shows certain characteristics.



# *Stick figure Icons (Pickett & Grinstein)*




Census data showing age (y), income (x), education, salary, language, marital status etc

# *Glyph-Based MD Visualization*

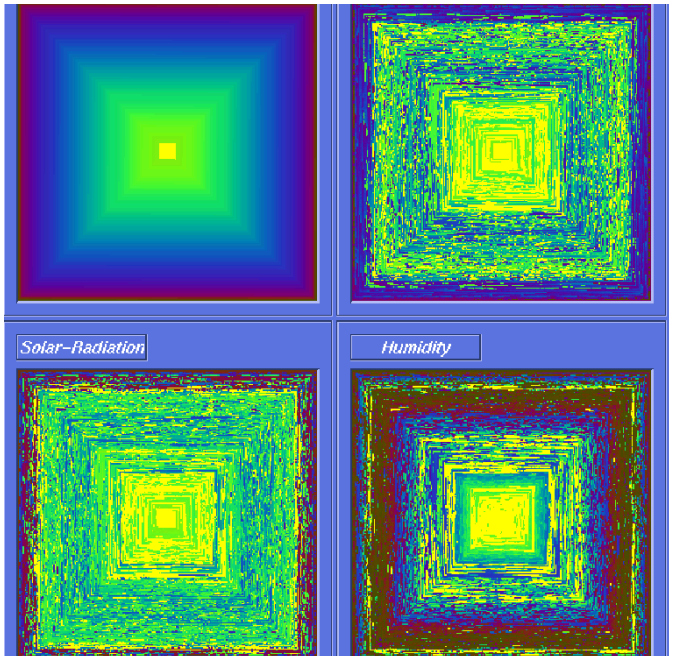
- Pros
  - Provide holistic overview of the information space
  - Exploit the human powerful ability of perceiving (texture)
  - patterns and human face characteristics (Chernoff)
  - Direct metaphor of Chernoff-face-like icons (e.g. houses)
  - may prove to be intuitive for novice users
- Cons
  - Glyphs must be learned
  - Only suitable for small to medium data sets
  - stick figures give a rather broad overview and may be difficult to interpret
  - Mappings may introduce biases in interpretation (e.g. the head shape of a Chernoff-face may be easier to perceive and compare than length of nose)

# Classification of MD Vis Techniques

## Pixel-oriented Techniques

# *VisDB*

- Database of data items, each of n dimensions
- Issue a query that specifies a target value of the dimensions
- Often get back no exact matches
- Want to find near matches




- Relevance factor
- metadata

Taken from:
D. Keim, H-P Kriegel, "VisDB Database Exploration Using Multid Vis", IEEE CG&A, 1994.

# *Technique*

- Calculate relevance of all data points
- Sort items based on relevance

- Use spiral technique to order the values
- Color items based on relevance

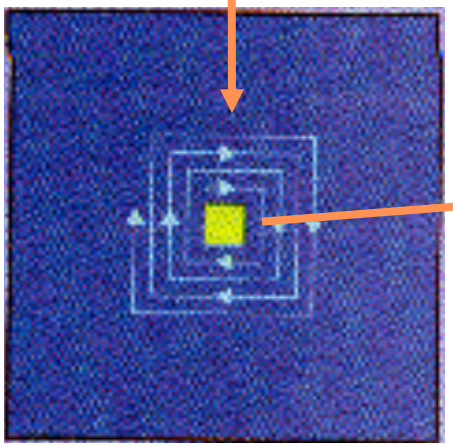
High Empirically established Low

# *Display Methodology*

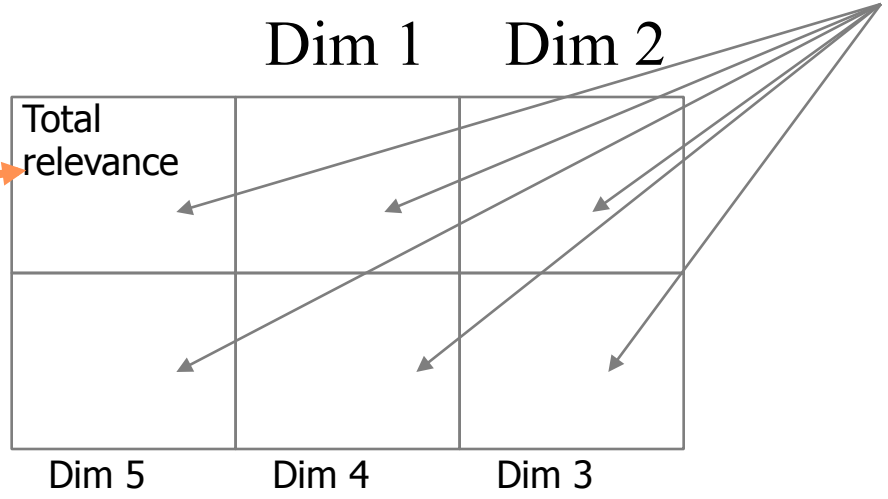
Items ordered by total relevance

Highest relevance value in center, decreasing values grow outward

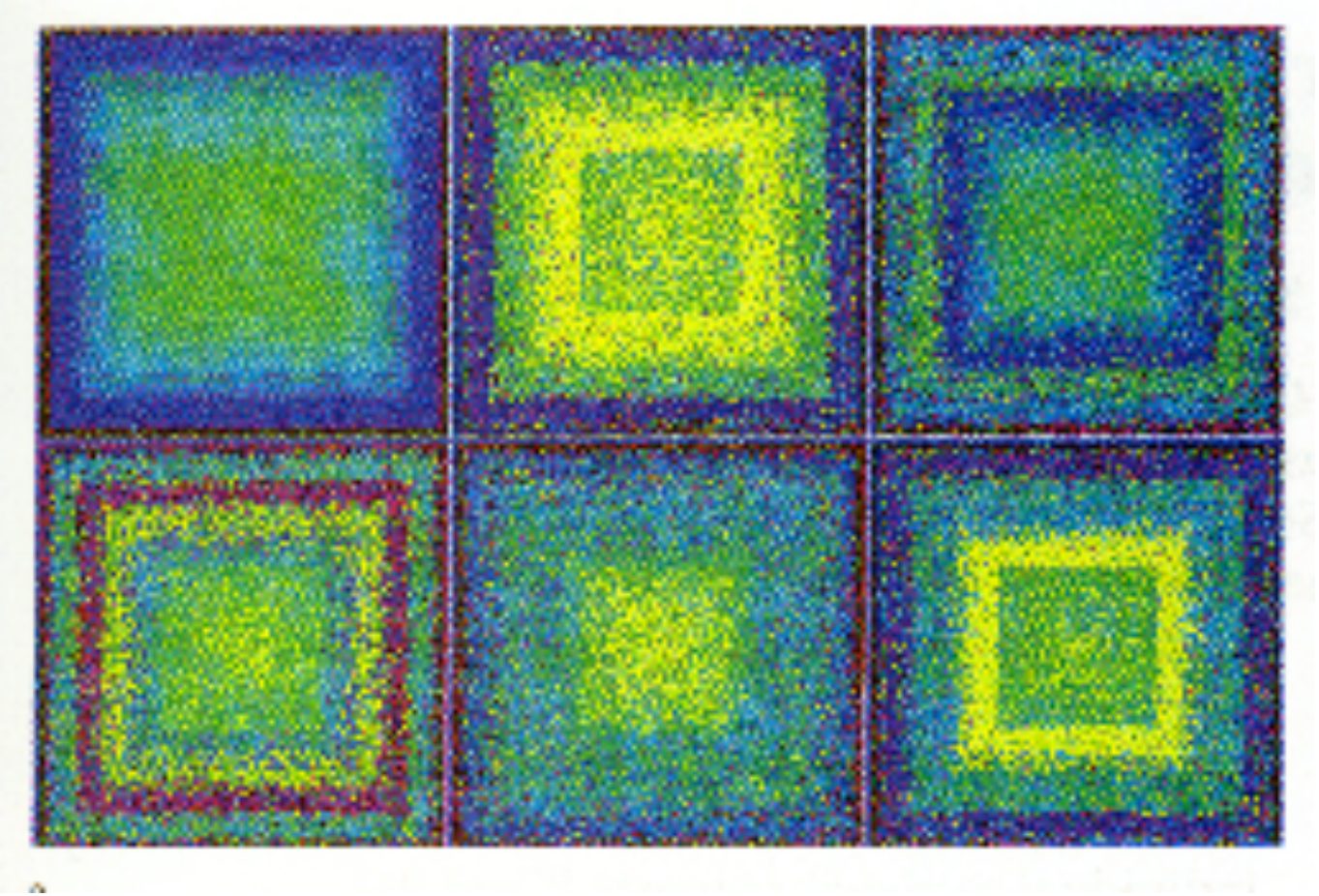
Same item appears in same place in each window



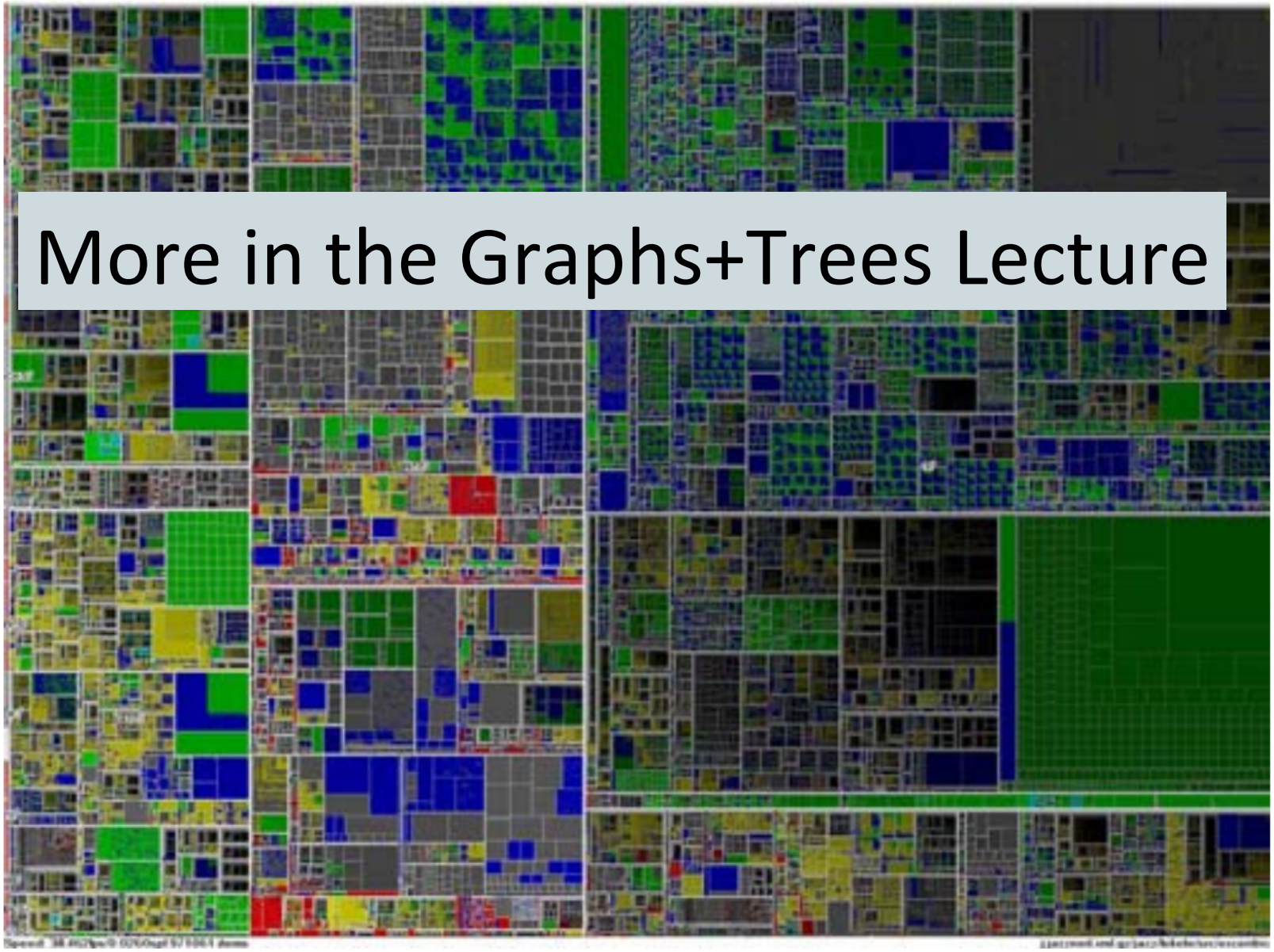
Spiral in each window



# *VisDB Example*



# *Directory Structure using treemap …*



# ***Back to**: two main approaches*

1. Attempt to visualize all dimensions
2. **Visualize only parts of the data**
   → interaction

# ***Back to**: two main approaches*

1. Attempt to visualize all dimensions
2. **Visualize only parts of the data**
   → interaction
   → Dimension Reduction

# ***Back to**: two main approaches*

1. Attempt to visualize all dimensions
2. **Visualize only parts of the data**
   → interaction
   → Dimension Reduction

which dimensions to choose?

# *Dimension Reduction - Approaches*

- Variable / feature selection
- Combined dimensions: PCA, MDS
- Dimension clustering

# *Feature selection*

- the process of selecting a subset of relevant features
- redundant features are those which provide no more information than the currently selected features
  - simple to perform
  - intuitive to understand
  - continuous process

# *PCA*

- *Statistical procedure that uses orthogonal transformations*

- *It 'discovers' new variables, called 'Principle Components' PCs which account for the majority of the **variability** in the data.*

- Enables description of information with considerably fewer variables than original data set.

http://www.dsea.unipi.it/Members/balestrinow/CP/file/TD_16_6_davies_pca.pdf

# *MDS*

- *Multidimensional scaling (MDS) maps the distances between items in a high dimensional space into a lower (e.g., two) dimensional space.*

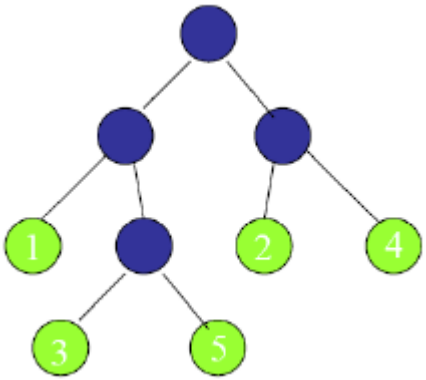
- Relocate *n items in a 2D dimensional space as points* so that the differences between pairs of points in this reduced space match the ordered differences between the subjects in original dimension

- Matching distances: minimize the difference of distances in original and projection data space

http://www.cs.kent.edu/~zhao/infovis14/lectures/MultiVis2.pdf

# *Hierarchical Dimension Reduction*

- 5-D dataset
- Similar dimensions form clusters
- Clusters are grouped into larger clusters in a dimension hierarchy




http://www.cs.kent.edu/~zhao/infovis14/lectures/MultiVis2.pdf

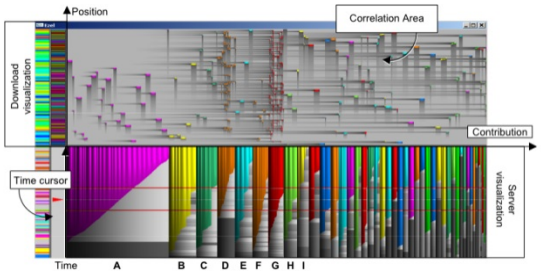
# *summary*

- there are many different kinds of md data
  - today we covered mostly general md-quantitative data
  - MANY more techniques exist that also mix types of data
  - many techniques are designed to solve a domain-specific problem or to help domain experts
  - almost all remaining lectures will also deal with md data

some md examples from lecture 1


graphs

time




text

# *general md-data vis strategies*

- Avoid over-encoding, reduce the problem space
  - ask yourself or your users what needs to be seen
  - ask what we can learn from the data you have, focus on this aspect
  - ask yourself what questions you want your viewers to answer
- Use space and small multiples intelligently
- User interaction to generate relevant views
- Invest time in **interaction** design (see lecture)

Content of this slide adapted from Jeffrey Heer

# *today you…*

- learned about techniques for MD data visualization
- learned relationships between MD visualization techniques
- understood the problems of MD data visualization

# *Acknowledgements*

Slides adapted or inspired from slides authored by:

- Petra Isenberg
- Fred Vernier
- Chris North
- Ye Zhao
- Penny Rheingans
- Remco Cheng