Multivariate Data – More Overview

CS 4460 - Information Visualization

Jim Foley

Last Revision – August 2016

Some Key Concepts – Quick Review

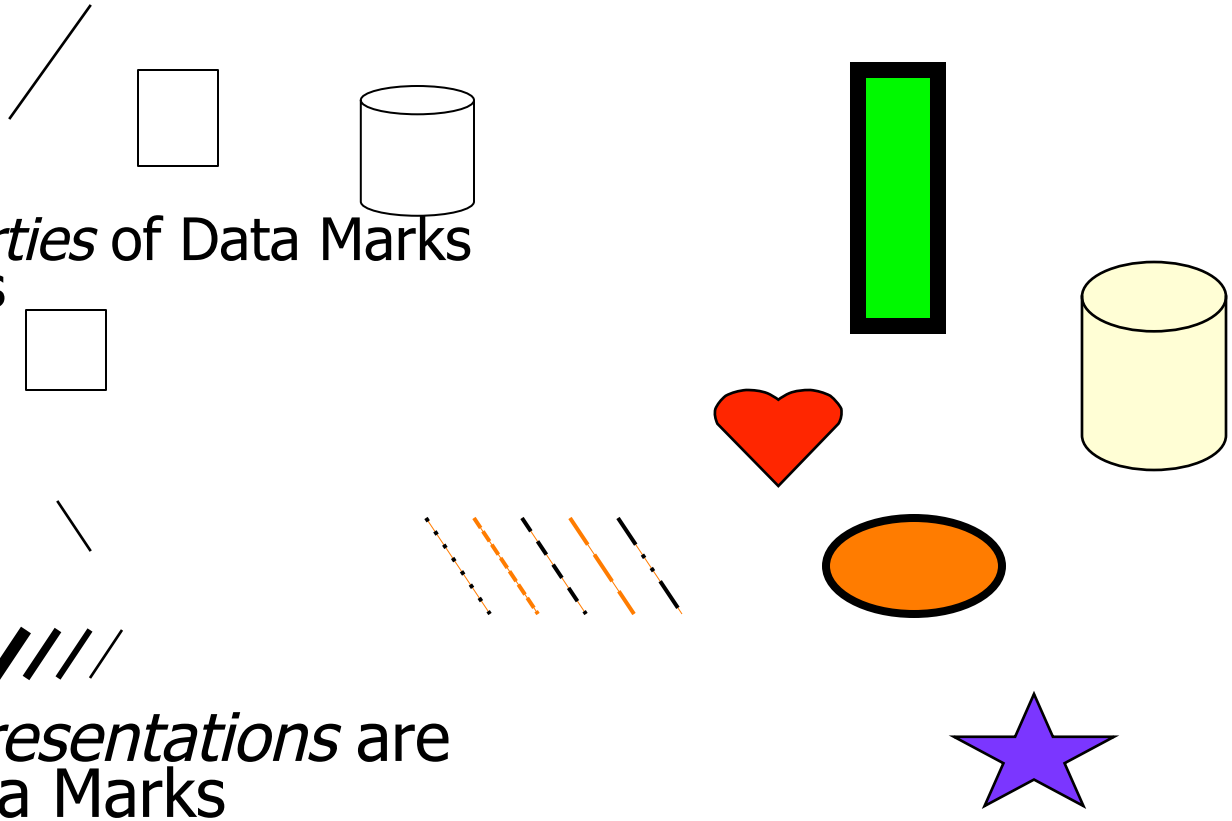
- Data Types
- Data Marks

Basic Data Types

- N-Nominal (categorical)
  - Equal or not equal to other values
  - Example: gender
- O-Ordinal
  - Obeys < relation, ordered set
  - Example: freshman, sophomore, junior, senior
- Q-Quantitative
  - Can do math, equal intervals
  - Examples: distance, weight, temperature, population count, your age

Data Marks

- *Data Marks* are visual primitives in 2D or 3D space
  - Points
  - Lines
  - Areas
  - Volumes
- *Graphical Properties* of Data Marks encode variables
  - Size
  - Shape
  - Color (HSV)
  - Orientation
  - Texture
  - Border
  - Thickness
- *Information Presentations* are built up of Data Marks



J. Foley

Data Type Implies Mark Type

Data Type: Ordinal & Quantitative

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Most ▲ Effectiveness ▼ Least

Same (Color luminance, Color saturation)

Same (Curvature, Volume (3D size))

Data Type: Nominal

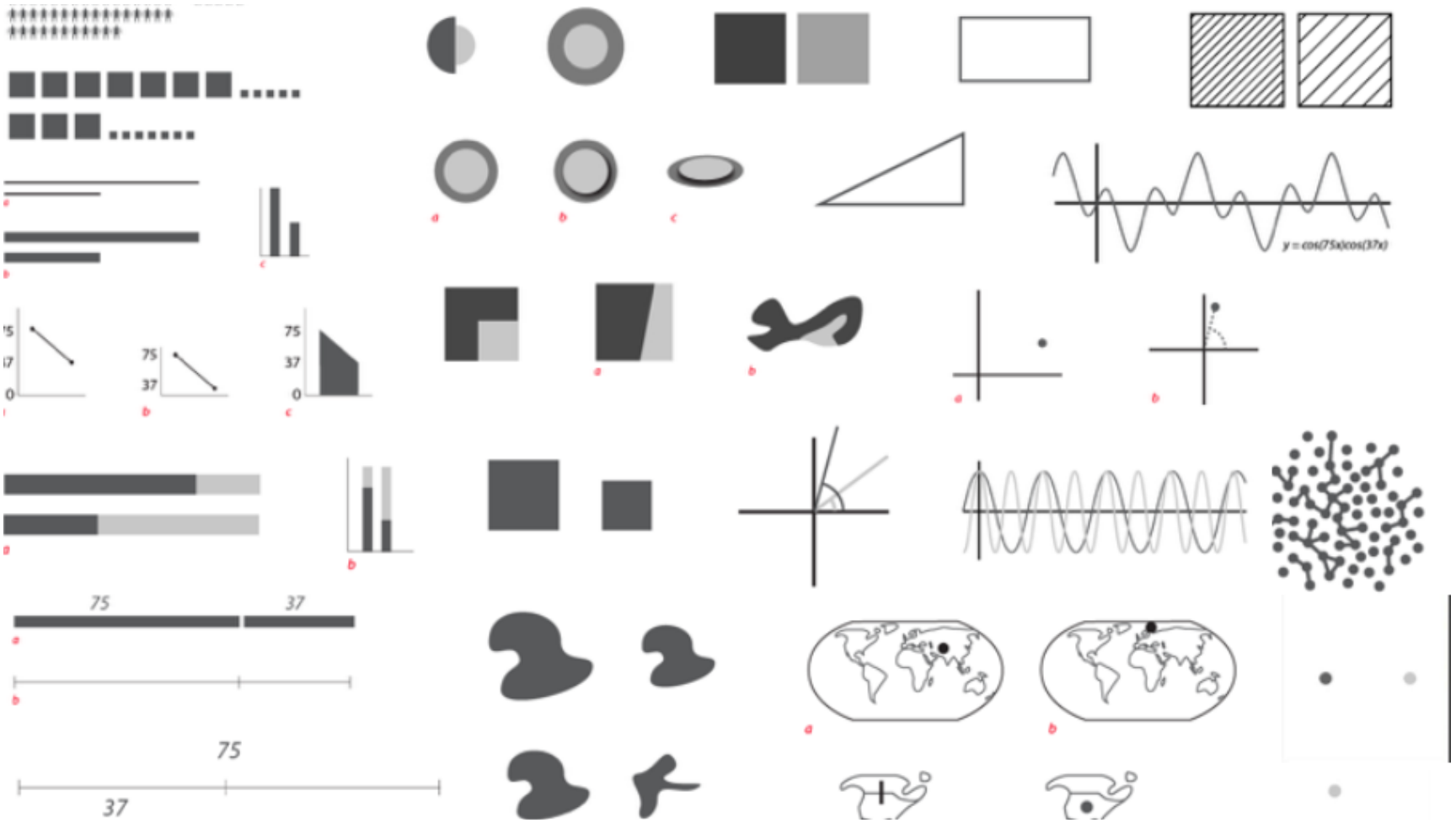
Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

Not an exhaustive list

Fig 5.1, *Visualization Analysis and Design*, Munzer, 2014

More Data Marks



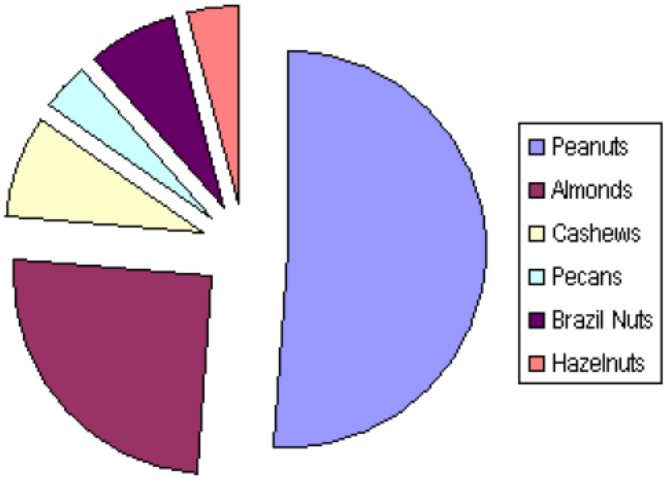
From http://blog.visual.ly/45-ways-to-communicate-two-quantities/

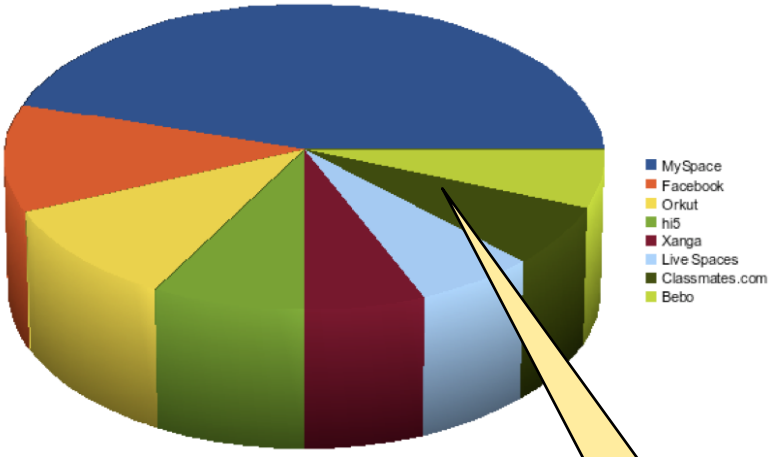
How Many Ways to Visualize Multivariate Data?

- Limited only by our imagination and creativity
- Here are some of the more common

- Following examples generally do not include geo-coded or time-coded data – more on that later

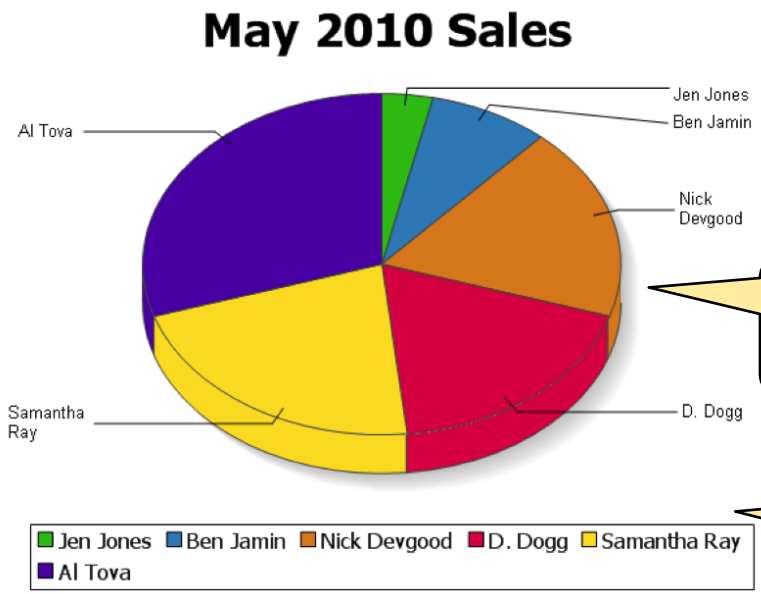
Pie Charts: Usually Bivariate (N=2)





What does 3D add?

What COULD 3D add?



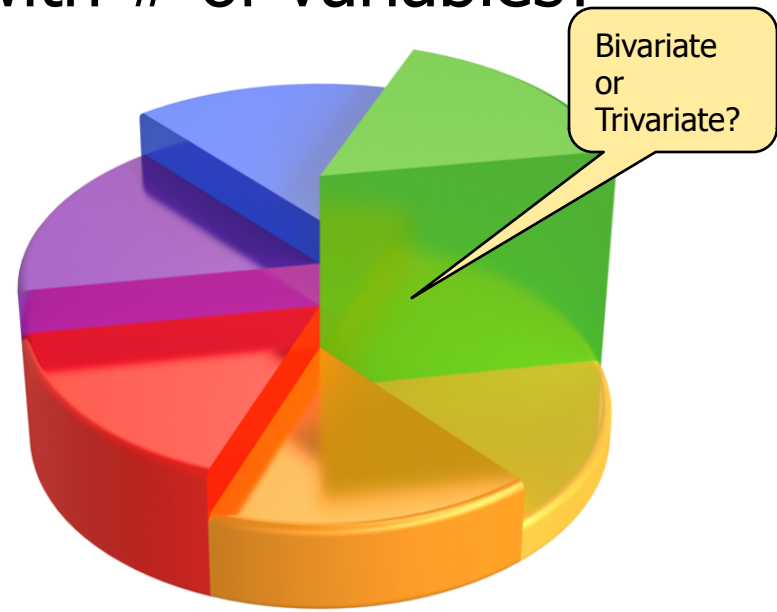
Double key – labels on slices + legend

Legend in same order as slices

How well do pie charts scale with cases?

Pie Charts

- What data types are most commonly depicted with pie charts?
  - Identification of each slice – what data type?
  - Size of each slice – what data type?
- How well do pie charts scale with # of variables?
  - Angle
  - Color (H, S, V)
  - Height
  - Texture
- Would pie chart with 4 variables be useful?



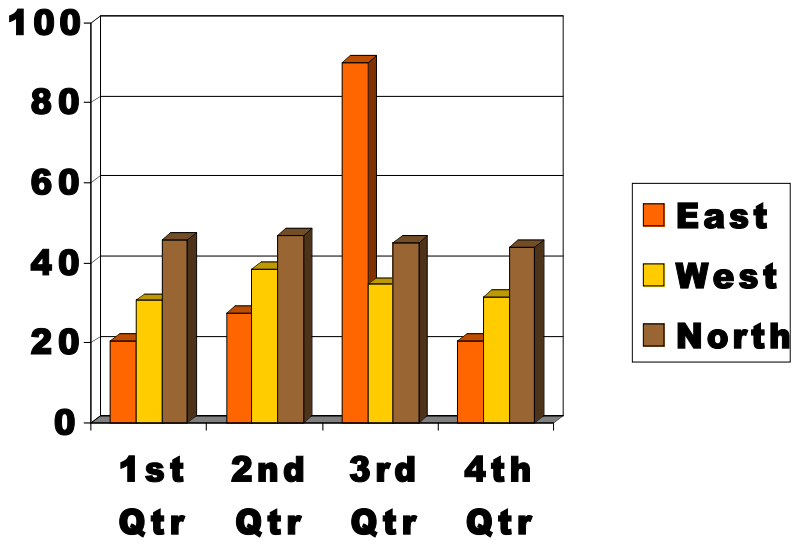
Hypervariate Data N > 3

- Number of well-known visualization techniques exist for data sets of 1-3 dimensions
  - line graphs, bar graphs, scatter plots OK
  - We see a 3-D world (4-D with time)
- What about data sets with more than 3 variables?
  - Often the interesting, challenging ones
  - Could use *additional* data mark properties to encode *additional* data variables.

Bar Chart

Show the relationships between variables' values in a data table

How many variables in the multivariate table?



What are their types?

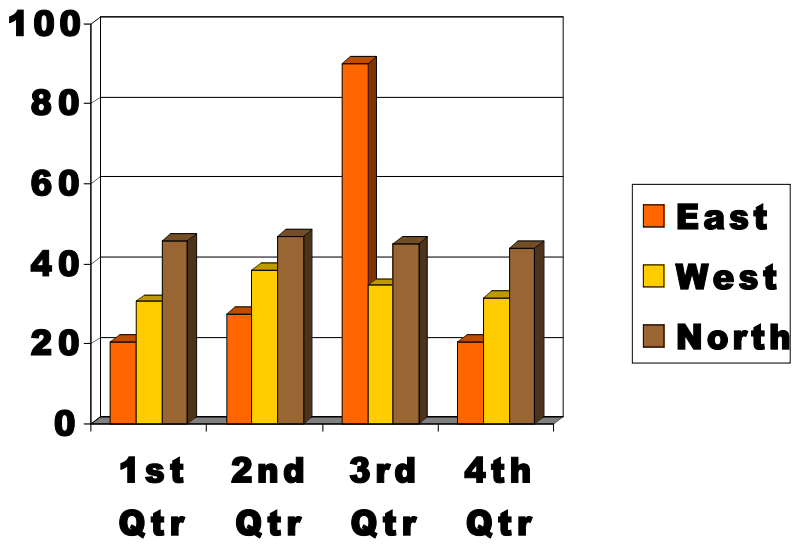
How many cases?

Bar Chart

Show the relationships between variables' values in a data table

How many variables in the multivariate table?

Region
Quarter
Sales



What are their types?

Region – Nominal
Sales – Ratio
Quarter – Ordinal

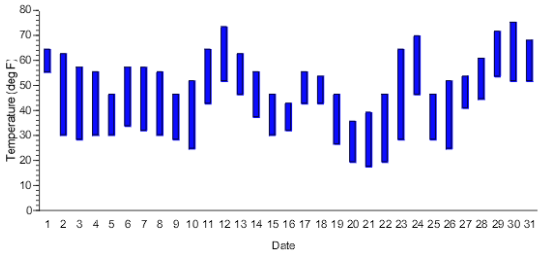
How many cases?

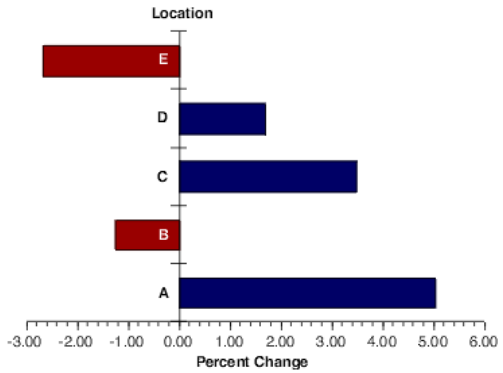
12

The Data Table

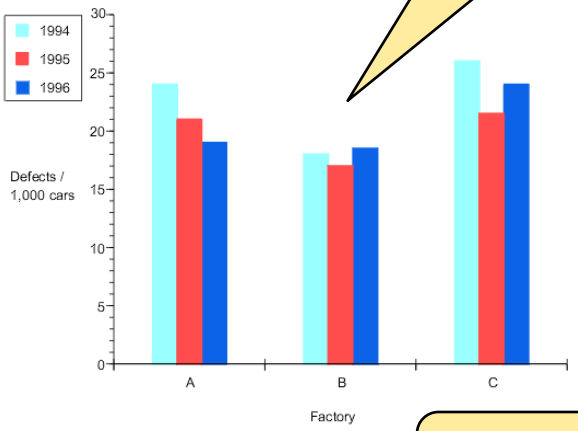
Region	Quarter	Sales
East	1	20
East	2	28
East	3	88
East	4	20
West	1	30
West	2	38
West	3	36
West	4	31
North	1	45
North	2	46
North	3	44
North	4	43

Gallery of Bar Charts

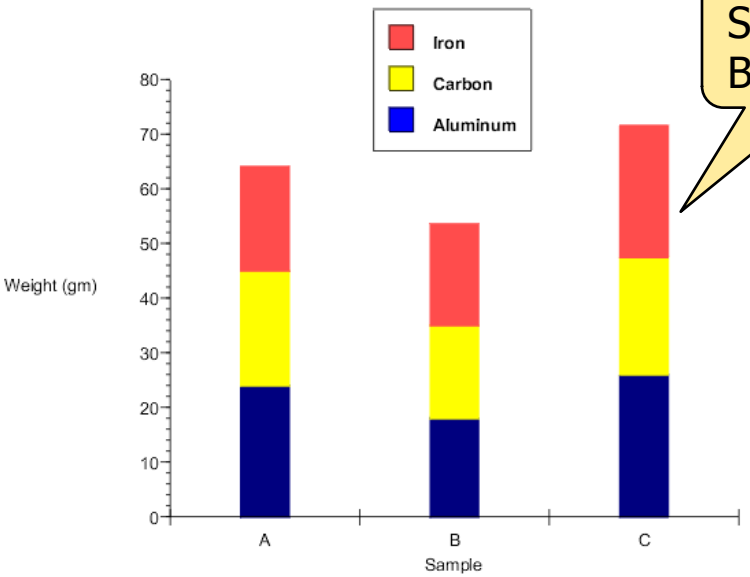




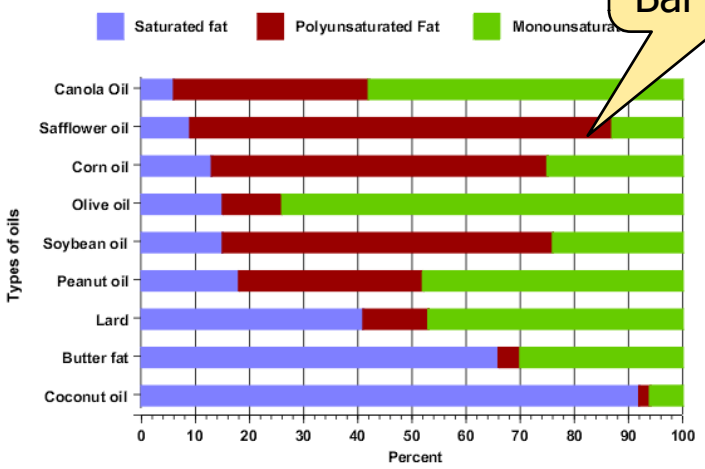
Clustered Bar Chart



Stacked Bar Chart

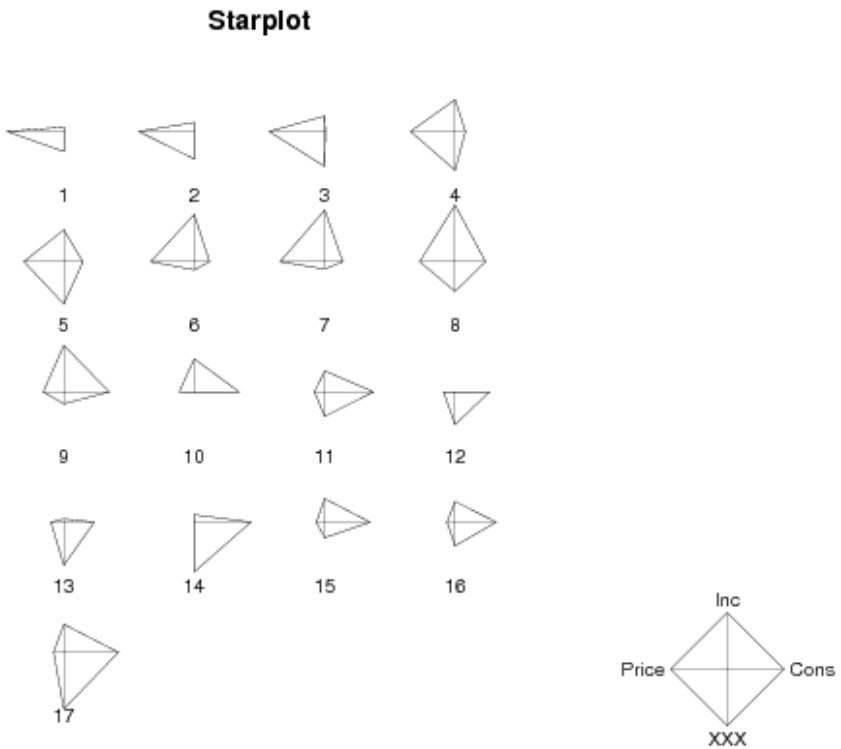


Stacked Bar Chart

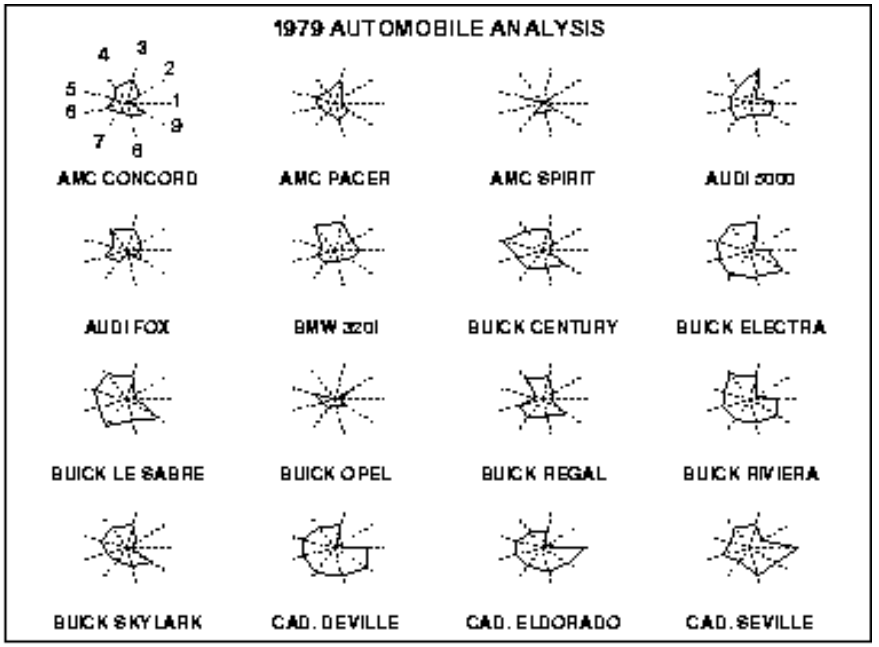


How many variables and cases?

Small Multiples: Star Plot N>3



N = 4 (5 if include case index/number); created at http://www.wessa.net/rwasp_starplot.wasp

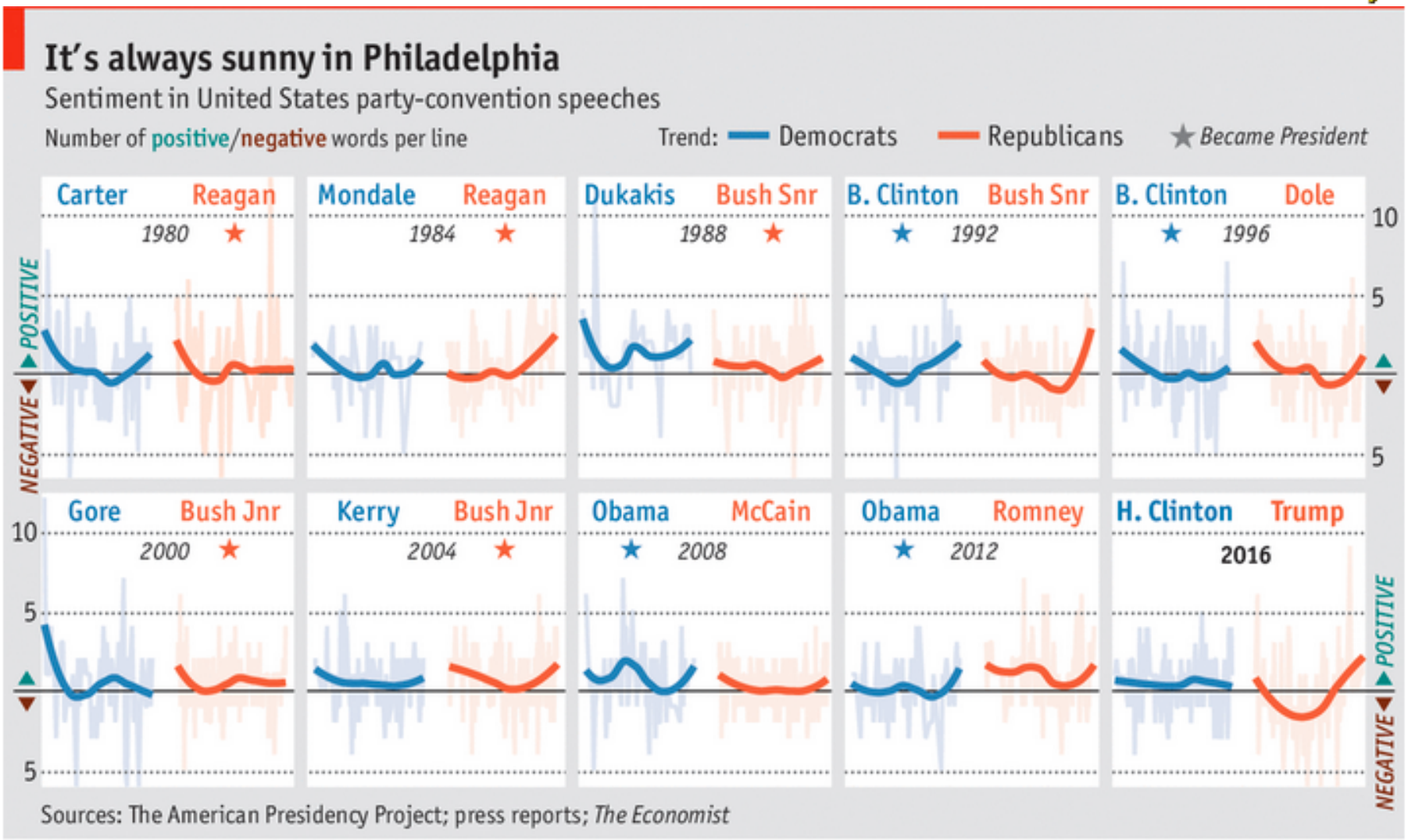


N = 10; Car type + 9 data items

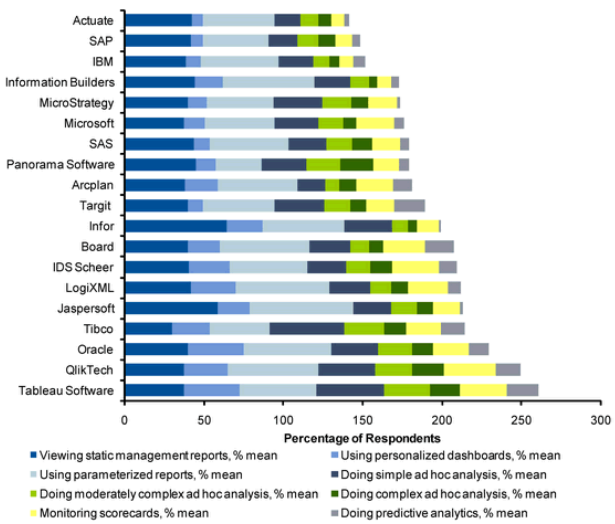
How well scale with
- # cases?
- # variables?

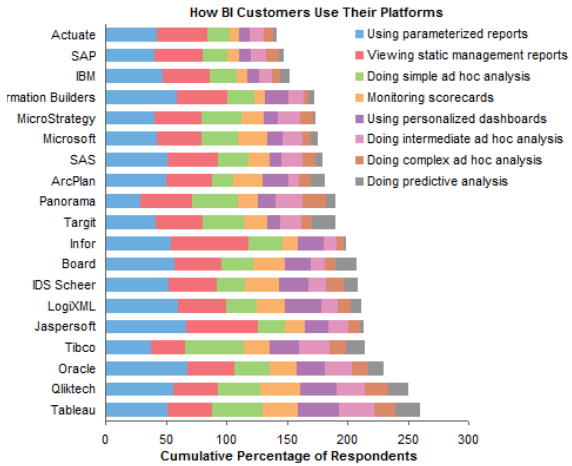
Small Multiples

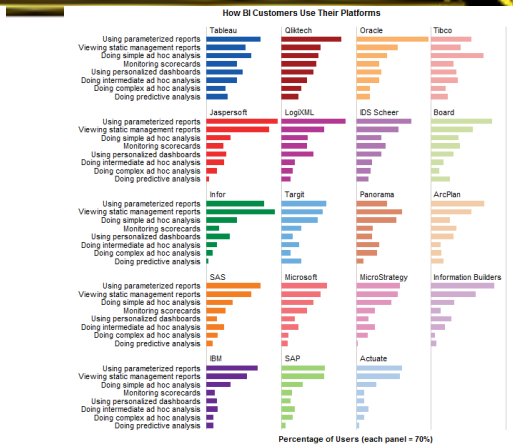
http://www.economist.com/blogs/graphicdetail/2016/07/daily-chart-19

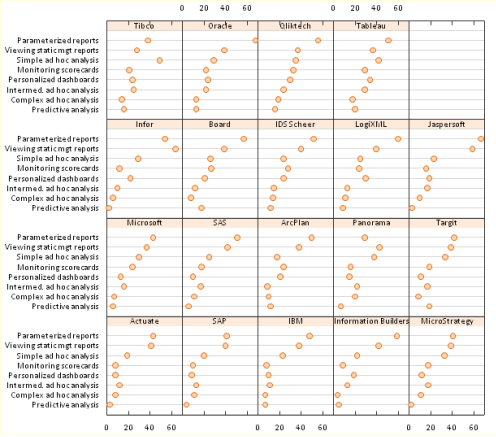


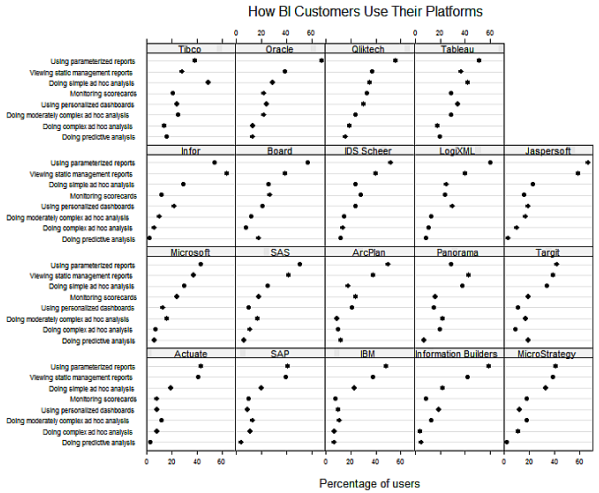
Many Ways to Present Same Data

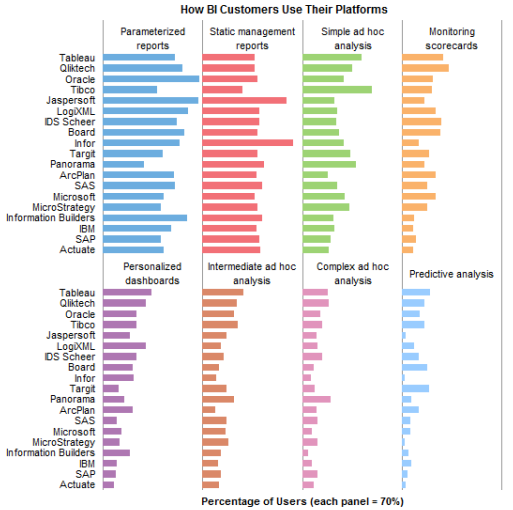






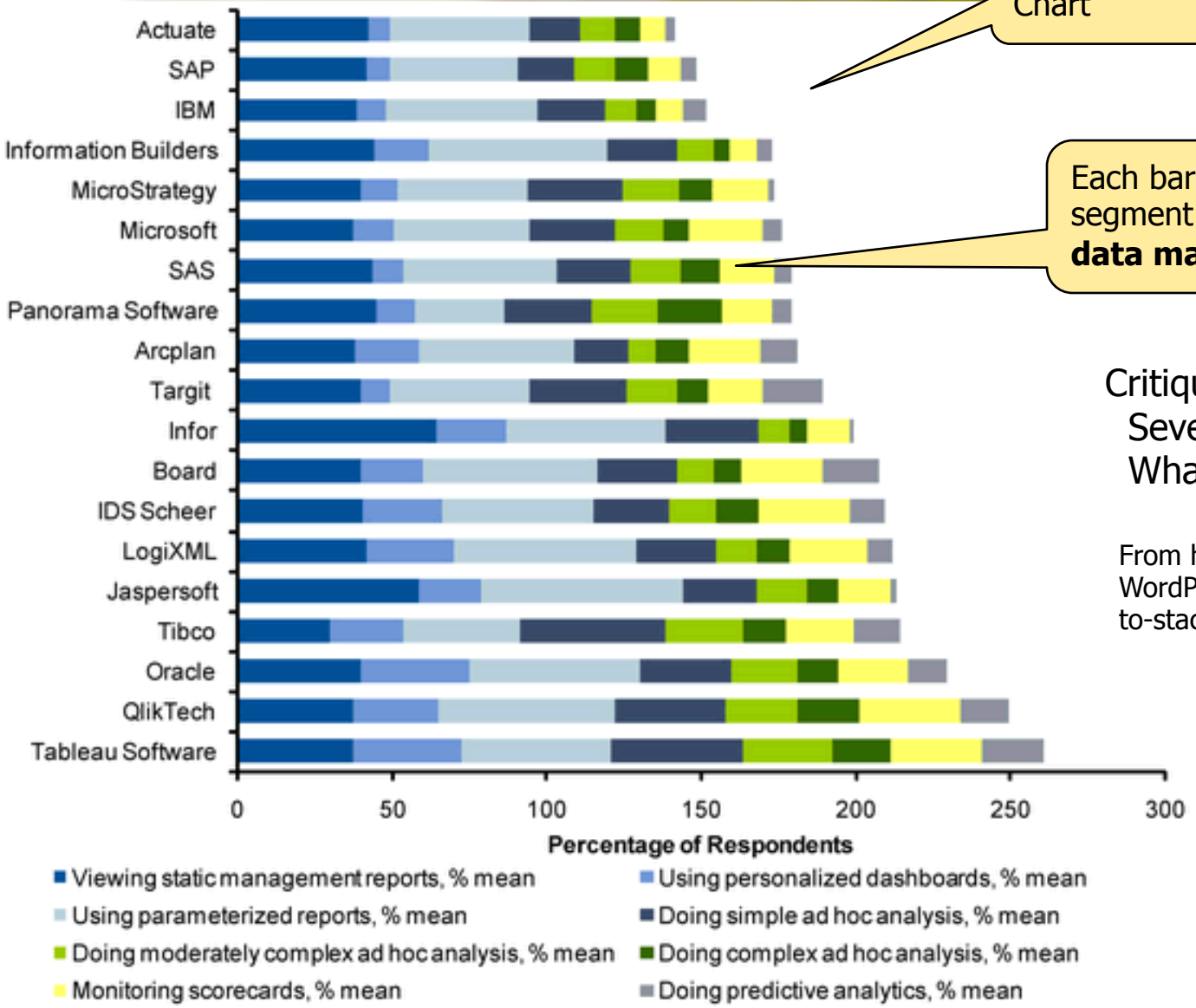






See in detail on next PPts

How Many Variables?

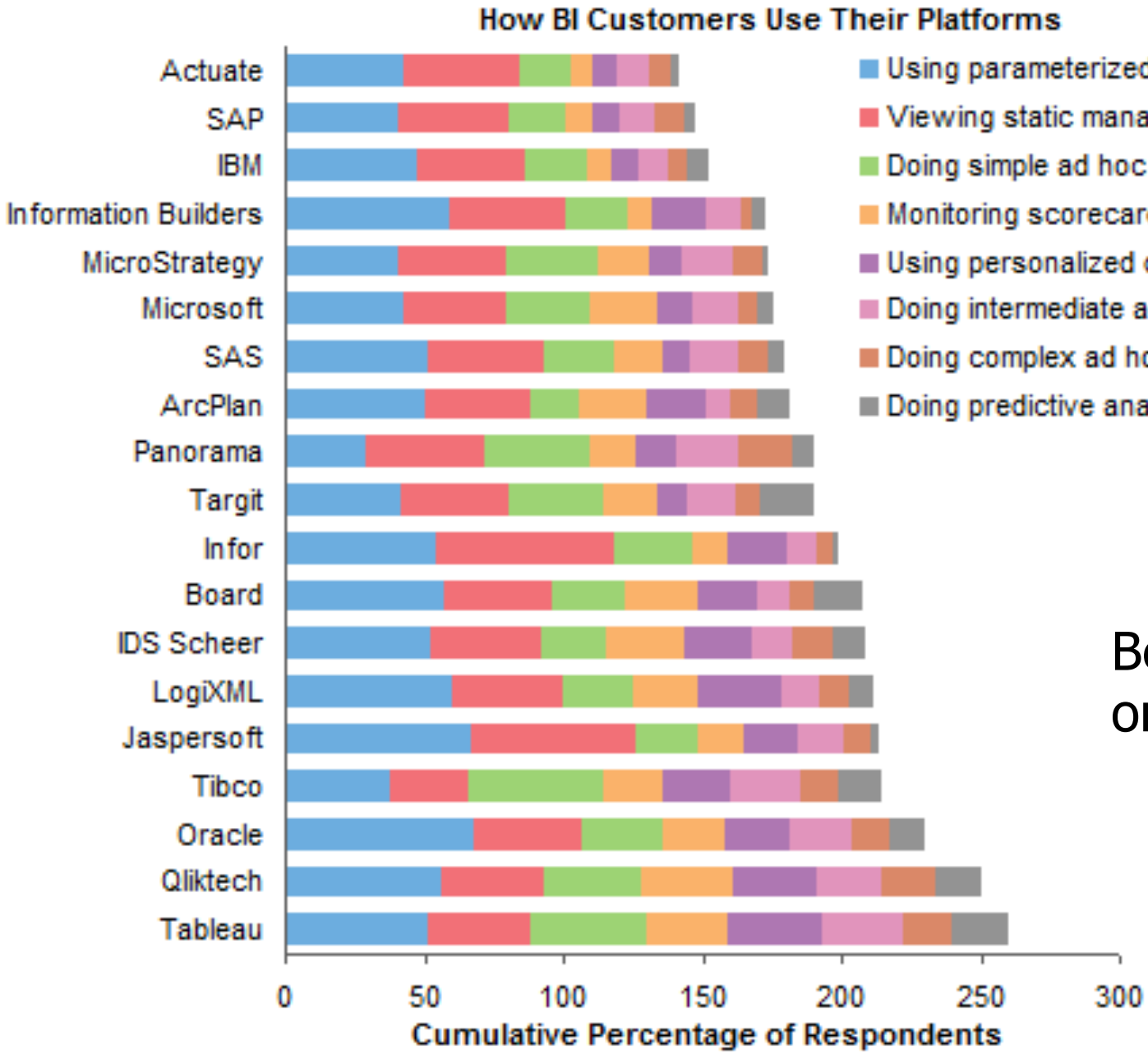


Stacked Bar Chart

Each bar segment is a **data mark**

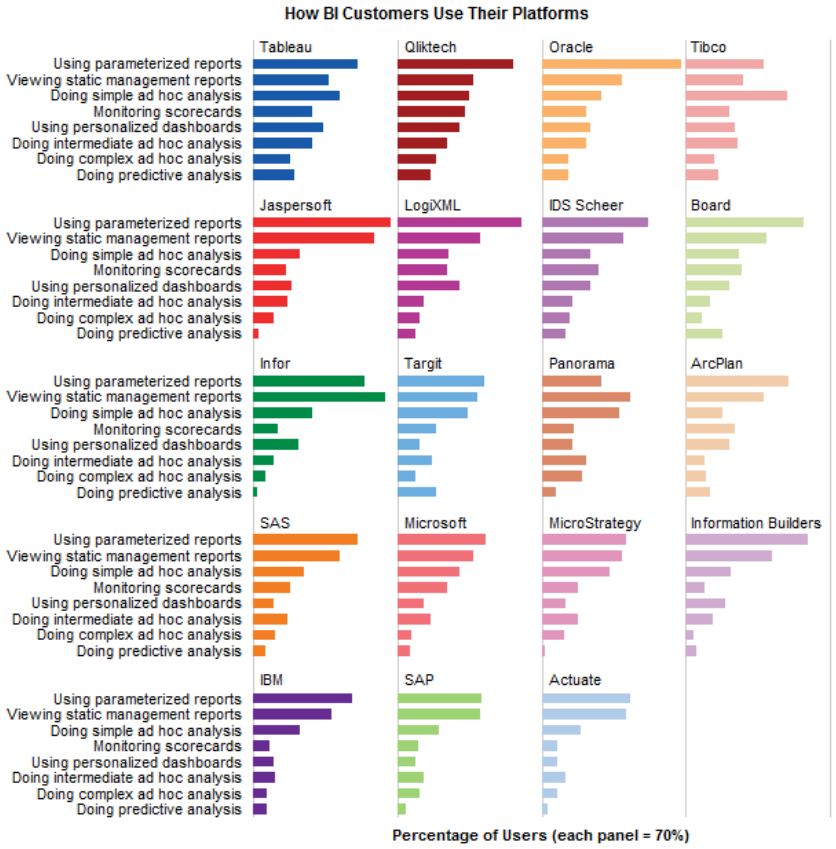
Critique Time -
Several issues!!
What are they?

From http://peltiertech.com/WordPress/trellis-plot-alternative-to-stacked-bar-chart/

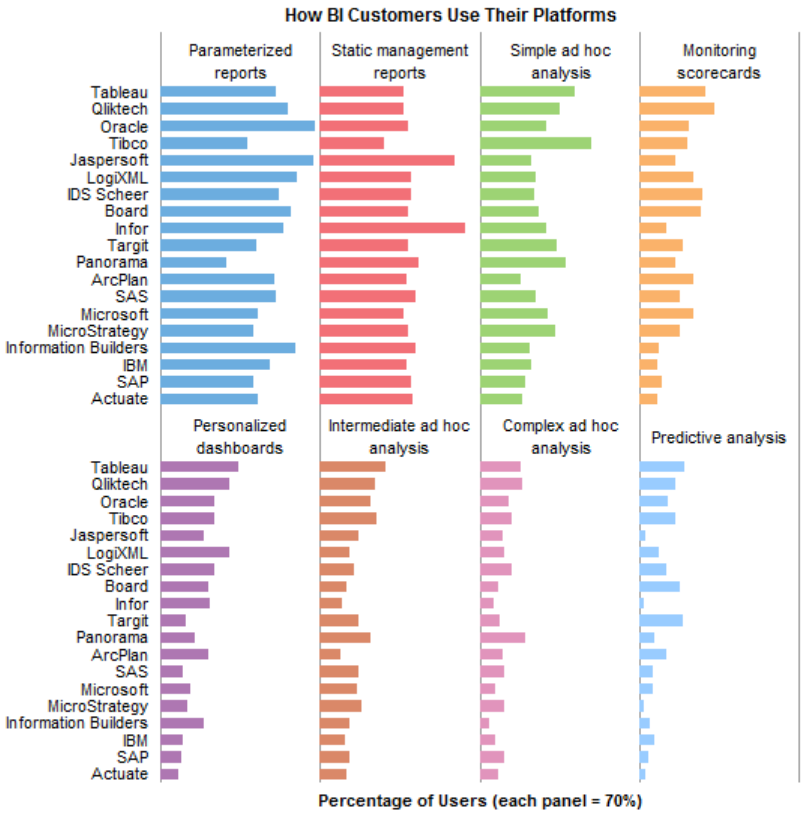


Better? Why or why not?

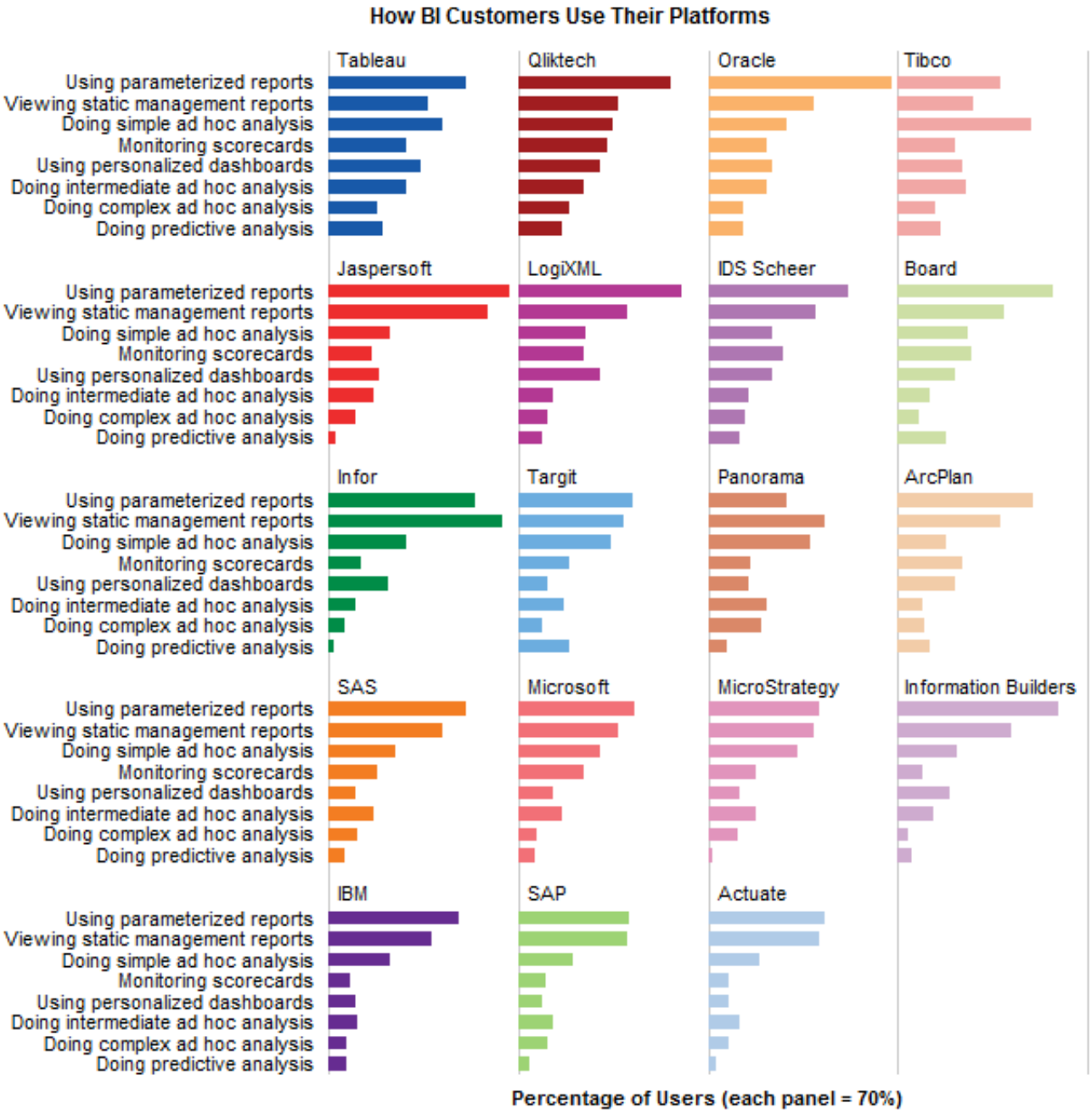
Small Multiples – two Variations



Small multiple for each of 4 SW platforms

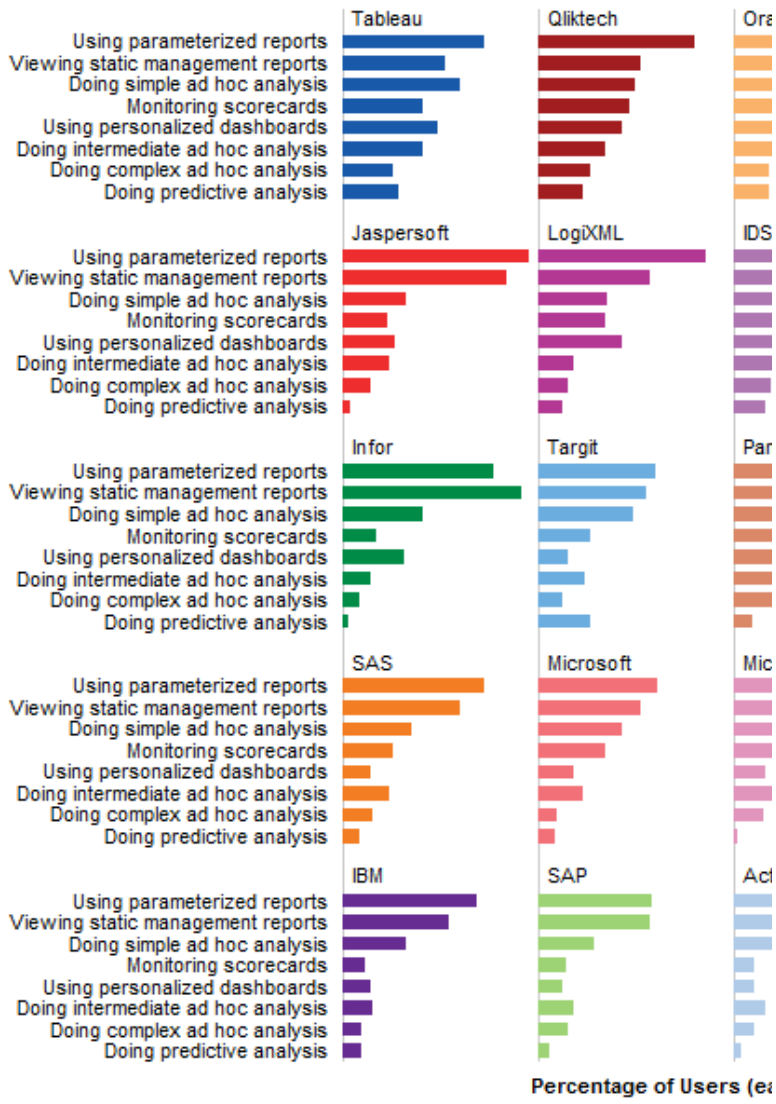


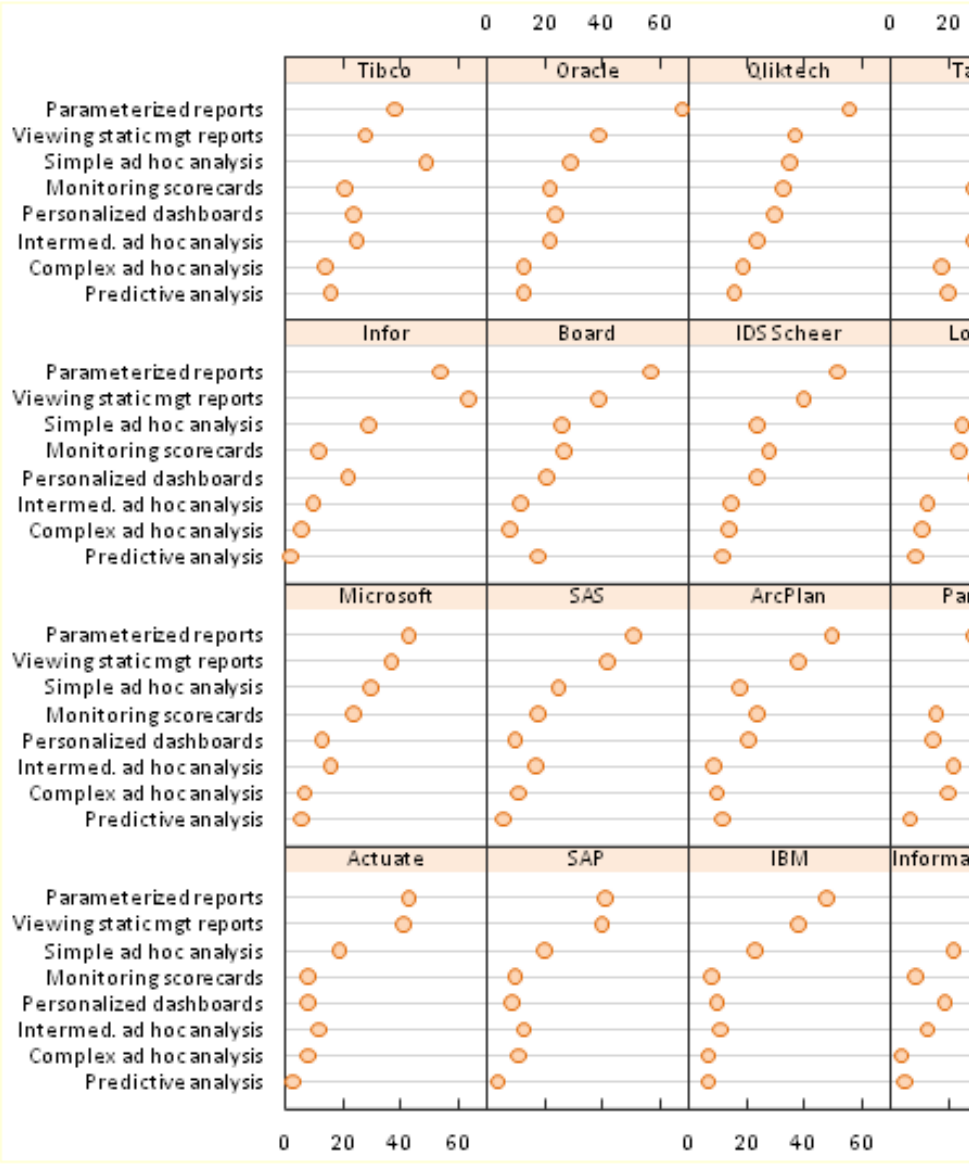
Small multiple for each of 8 uses



Enlarged: Small multiple for each of 4 SW platforms

Marks Instead of Bars





Sparklines

	A	B	C	D	E	F	G	H
1	Salesperson	May	June	July	Aug.	Sept.	Oct.	
2	Albertson, Kathy	$3,947.00	$557.00	$3,863.00	$1,117.00	$8,237.00	$8,690.00	
3	Allenson, Carol	$4,411.00	$1,042.00	$9,355.00	$1,100.00	$10,185.00	$18,749.00	
4	Altman, Zoey	$2,521.00	$3,072.00	$6,702.00	$2,116.00	$13,452.00	$8,046.00	
5	Bittiman, William	$4,752.00	$3,755.00	$4,415.00	$1,089.00	$4,404.00	$20,114.00	
6	Brennan, Michael	$4,964.00	$3,152.00	$11,601.00	$1,122.00	$3,170.00	$10,733.00	
7	Carlson, David	$2,327.00	$4,056.00	$3,726.00	$1,135.00	$8,817.00	$18,524.00	
8	Collman, Harry	$3,967.00	$4,906.00	$9,007.00	$2,113.00	$13,090.00	$13,953.00	
9	Counts, Elizabeth	$4,670.00	$521.00	$4,505.00	$1,024.00	$3,528.00	$15,275.00	
10	David, Chloe	$3,379.00	$3,428.00	$3,973.00	$1,716.00	$4,839.00	$13,085.00	
11	Davis, William	$5,363.00	$1,562.00	$2,945.00	$1,176.00	$9,642.00	$13,714.00	
12	Dumlao, Richard	$3,275.00	$2,779.00	$7,549.00	$1,101.00	$5,850.00	$15,065.00	

"Magic Quadrant"

How many variables?

⇒How many columns in table?

⇒Any ancillary information

Magic Quadrant

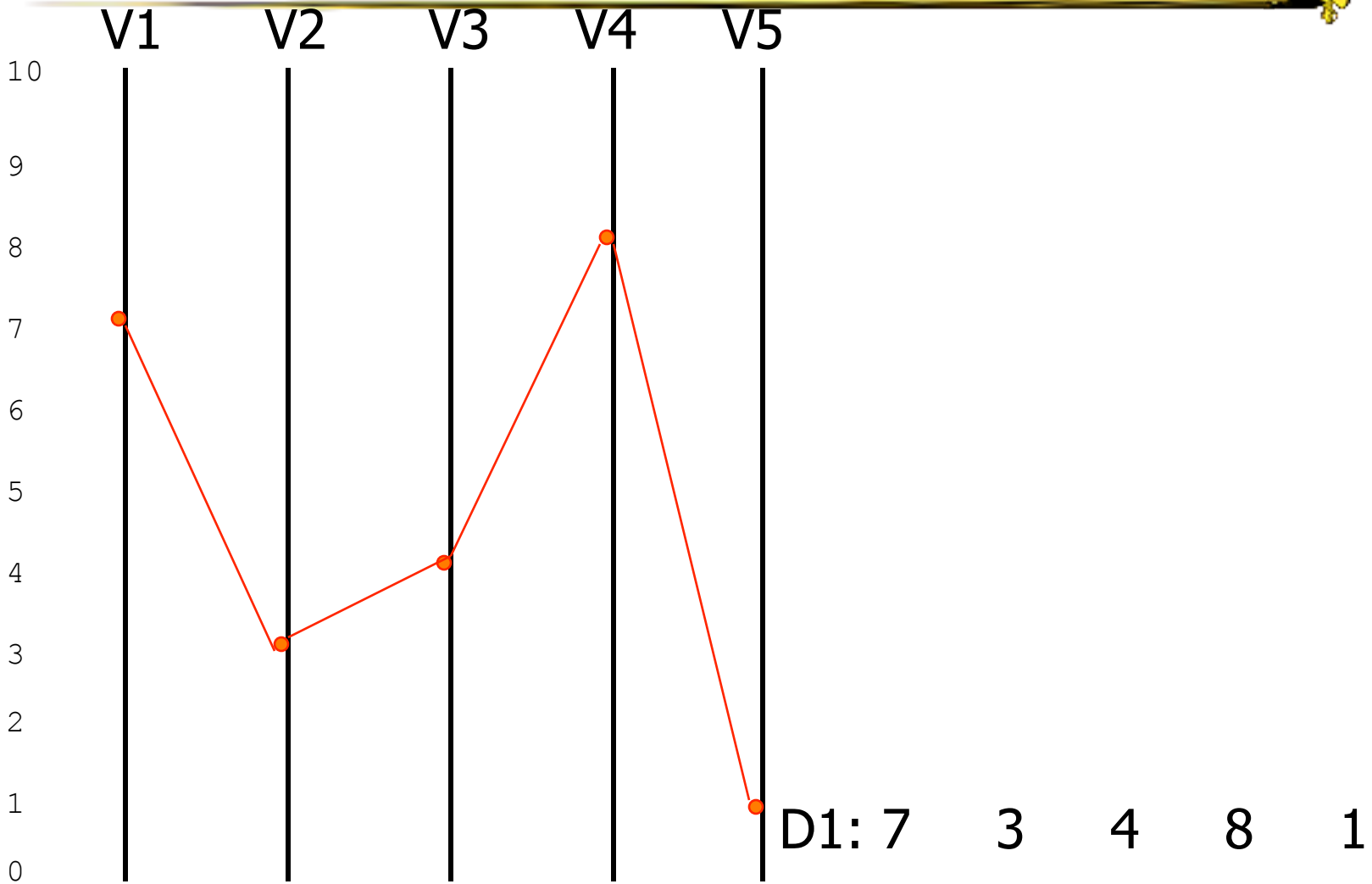
Figure 1. Magic Quadrant for Business Intelligence and Analytics Platforms

Parallel Coordinates N>3

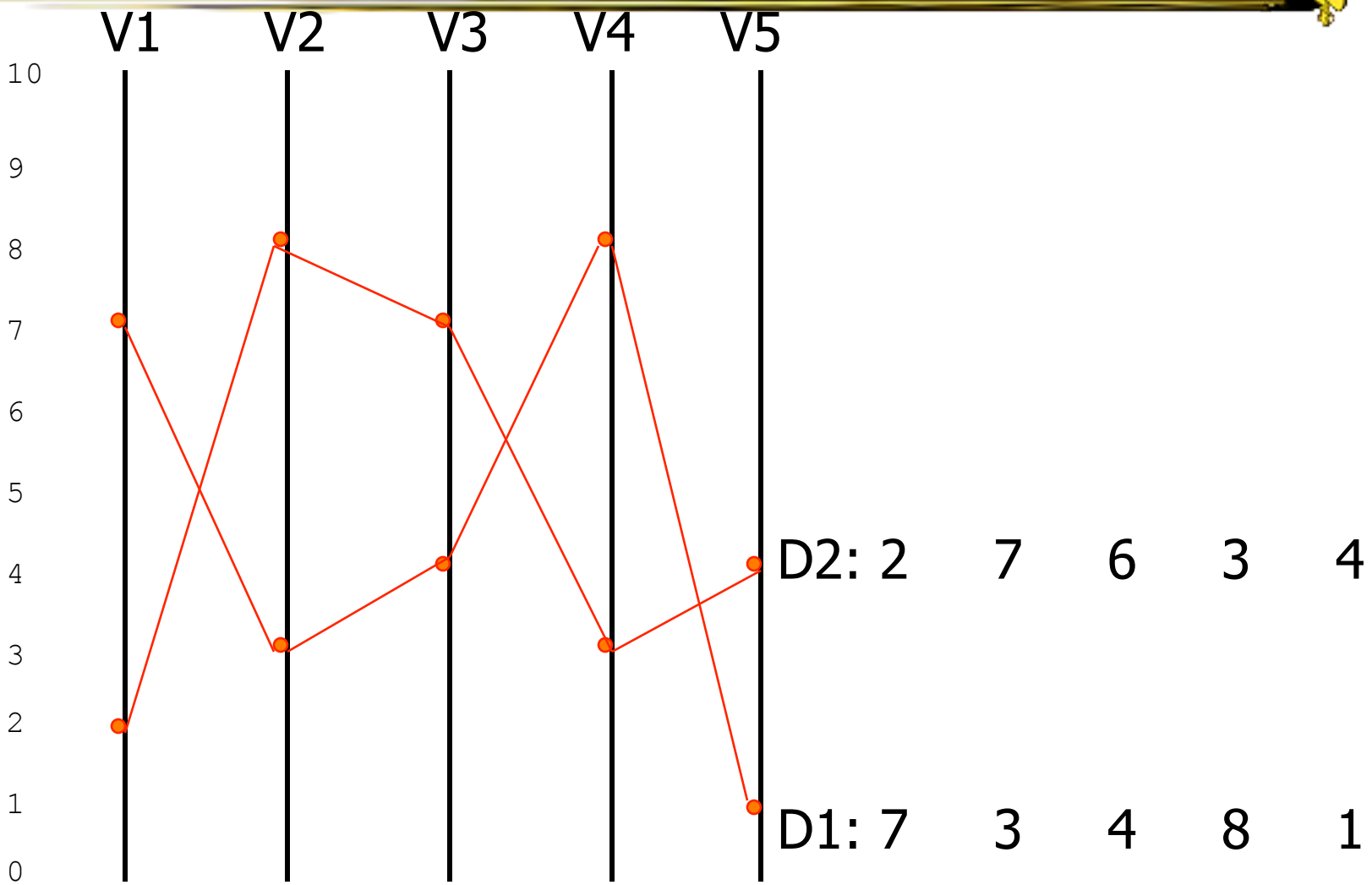
- Given this data table

	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

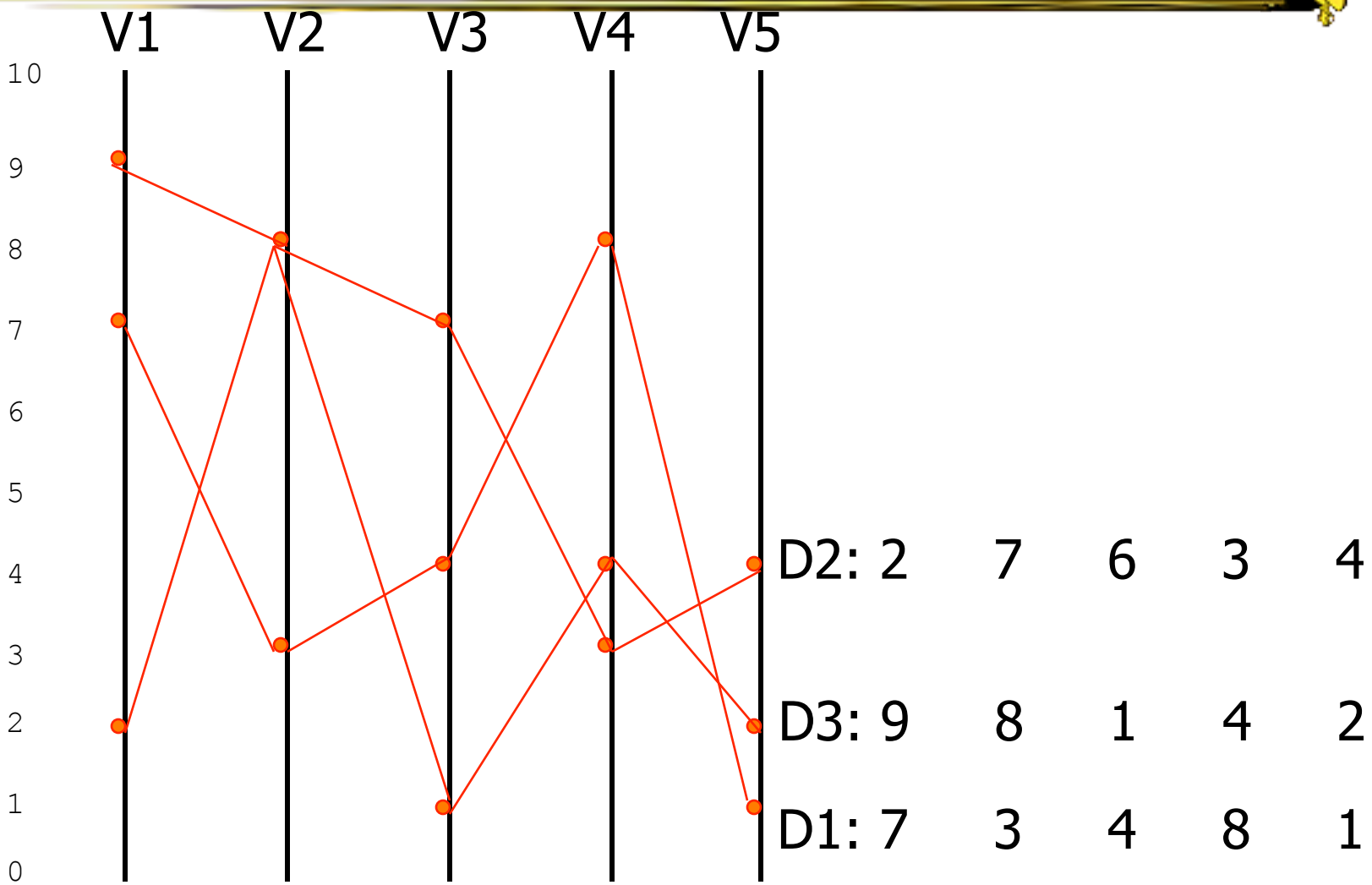
Parallel Coordinates



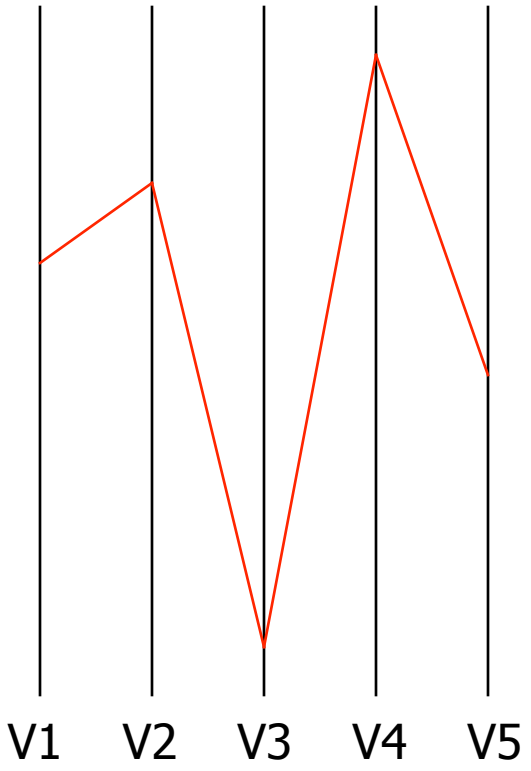
Parallel Coordinates



Parallel Coordinates



Parallel Coordinates



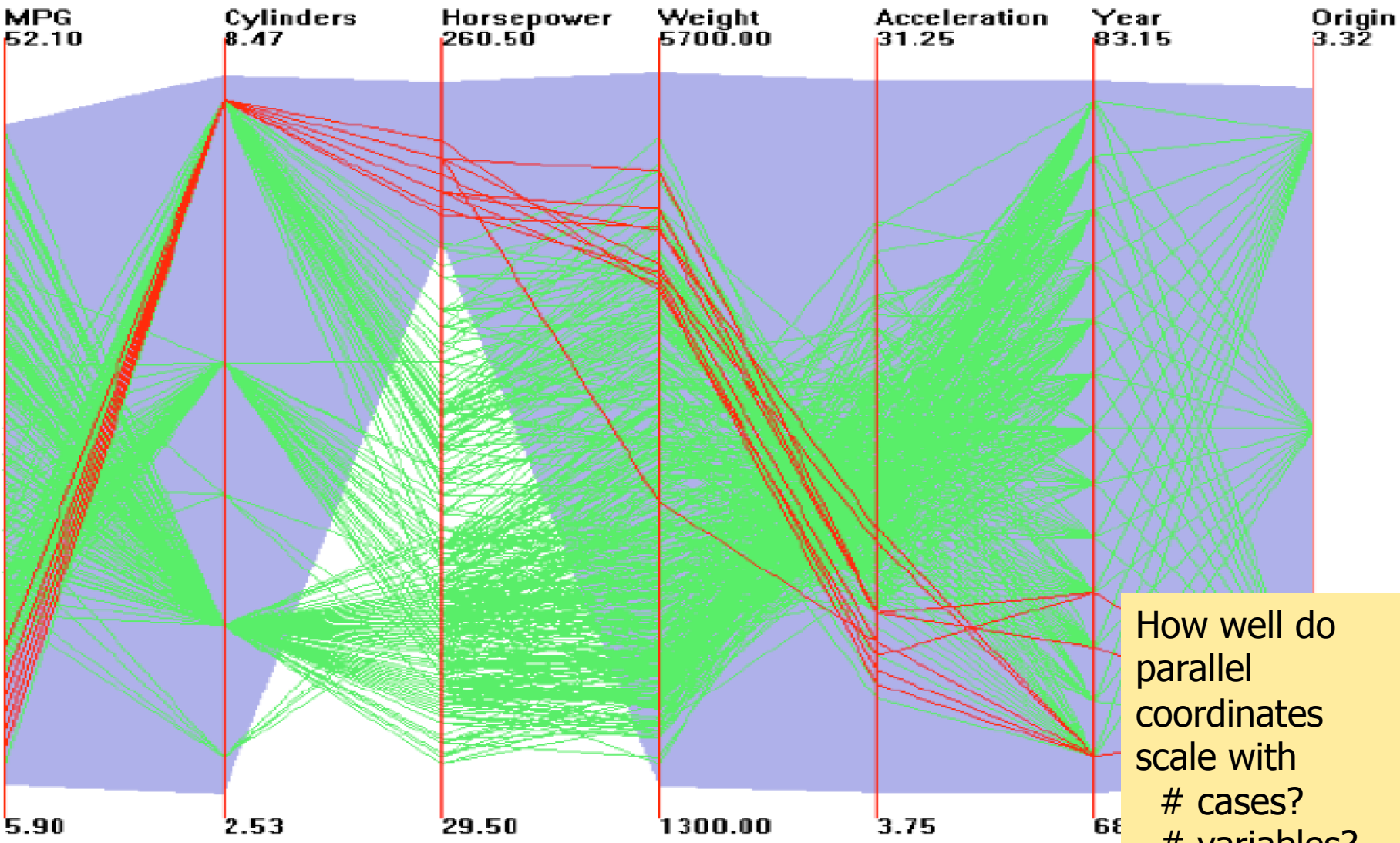
Encode variables V1, V2, etc along horizontal row

Vertical line specifies different values that variable can take

Data points (D1, D2, etc) represented as polyline

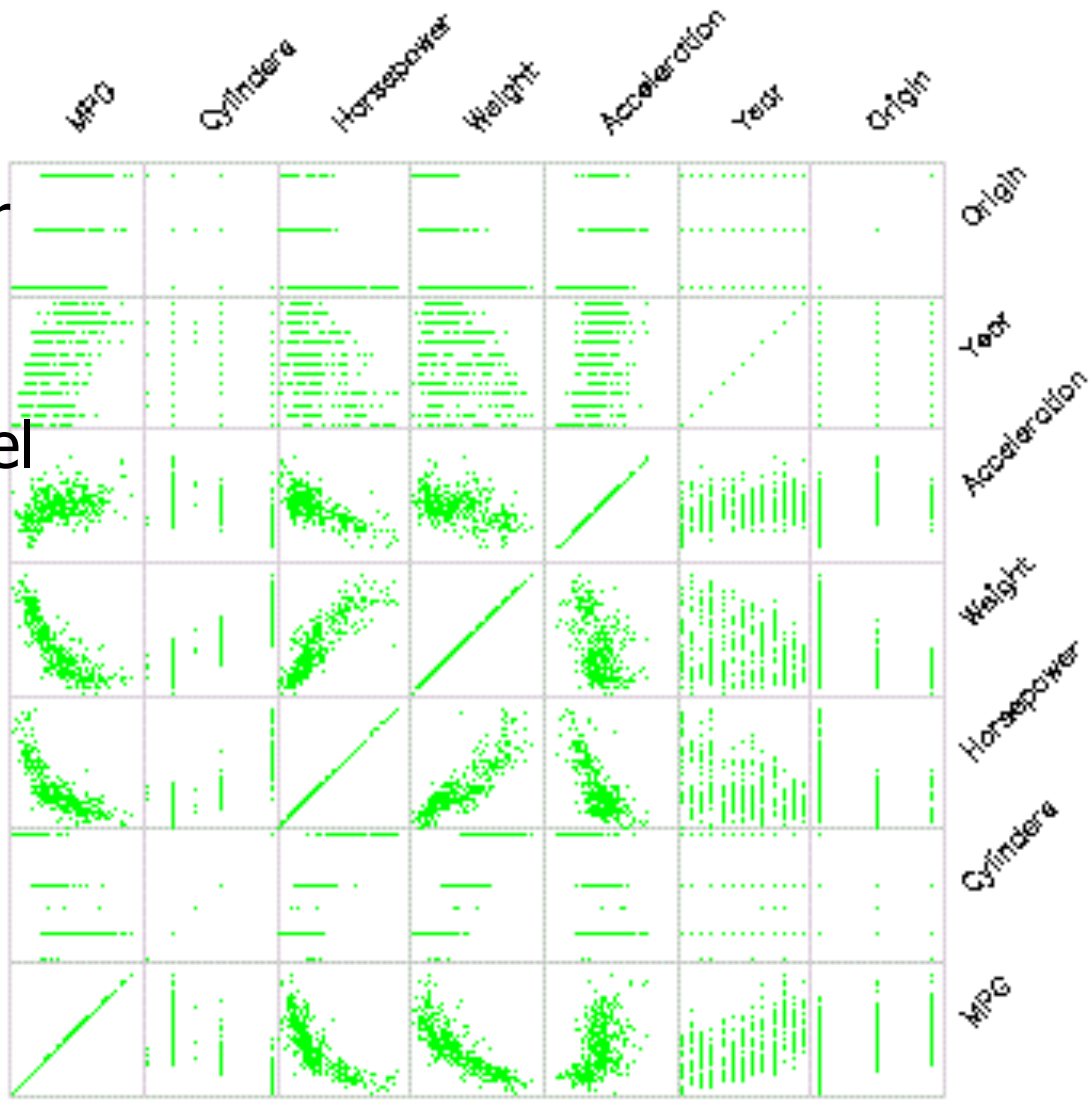
How differ from star plot?

Automobile Data in Parallel Coords



How well do parallel coordinates scale with
- # cases?
- # variables?

Automobile Data in Scatterplot Matrix



Small multiples: each pair of variables in scatterplot

How compare with parallel coordinates

- Seeing trends?
- Scale with # variables?
- Scale with # cases?

Takeaways – what are they?

- Work with a neighbor to write down three key points

- Now share them with other neighbors

Some Key Points

- Data types & marks
- Lots of ways to vis multivariate data
- Questions to ask about any vis
  - How many variables, what data types?
  - How many cases
  - How effective?
    - Absolute terms
    - Relative to alternatives
  - How does it scale up
    - # cases
    - # variables