

Multivariate Data, Visual Encodings, & Basic Charts

CS4460 - Introduction to Information Visualization

Spring, 2019

Alex Endert



Alex Endert

endert@gatech.edu

Today, we will cover

- Data types and characteristics
 - Items, Attributes, Types
- Basic techniques to visualize types
 - Univariate, Bivariate, Multivariate
- Basic Visual Charts
 - Tables
 - Visual Primitives, Encodings

Basic Data Types

- Nominal (categorical)

Equal or not equal to other values

Example: gender

- Ordinal

Obeys < relation, ordered set

Example: freshman, sophomore, junior, senior

- 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 data values

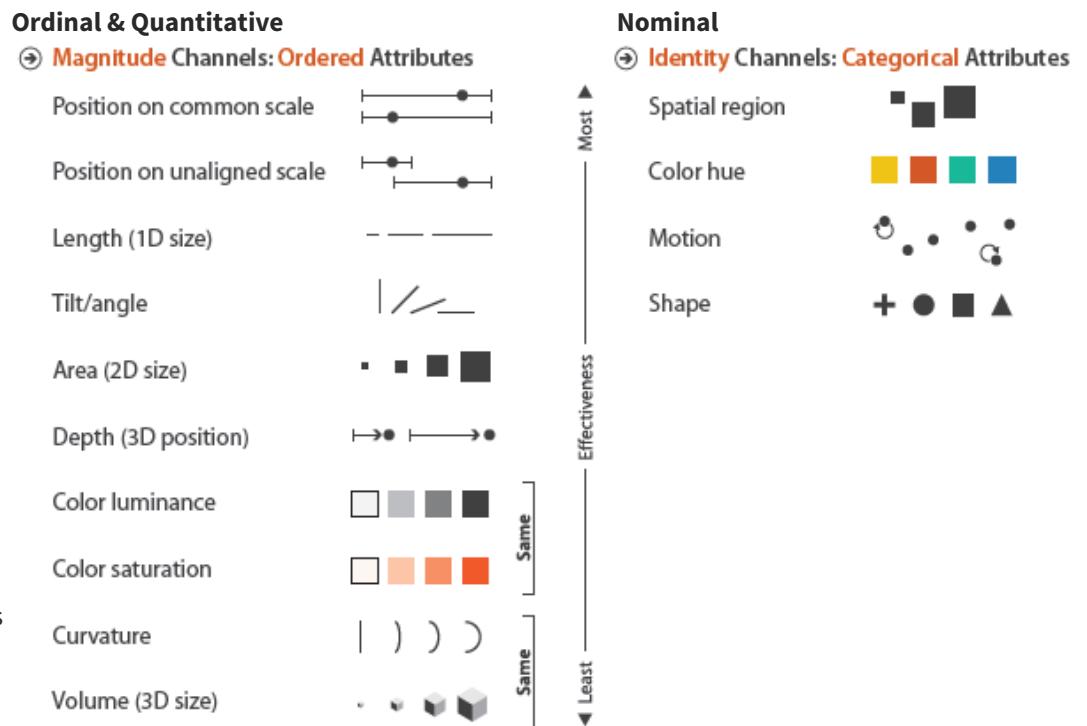
- Size, Shape, Color (HSV), Orientation, Texture, Border, Thickness

- Information Visualizations are made up of **Data Marks placed inside of visual metaphors, and mapped to data values**

Data Type Implies Mark Type

Fig 5.1, Visualization Analysis and Design, Munzer, 2014

Not an exhaustive list



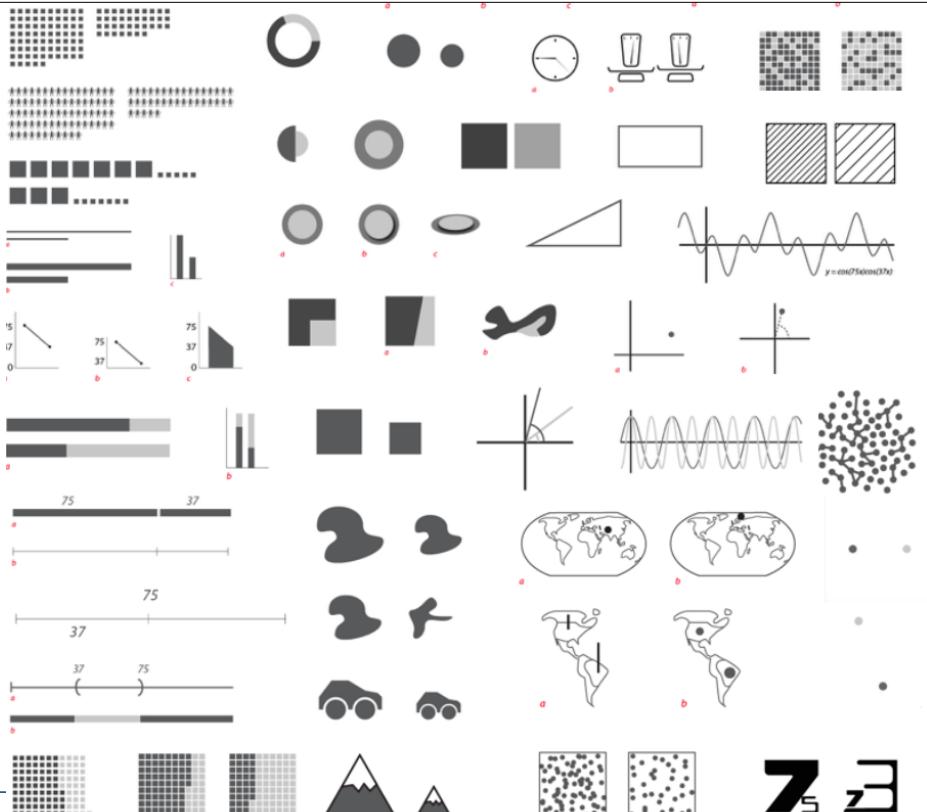
a closer look

④ **Magnitude Channels: Ordered Attributes**

Georgia Tech

Position on common scale	Horizontal line with two dots
Position on unaligned scale	Horizontal line with two dots and a break
Length (1D size)	Horizontal dashed line
Tilt/angle	Vertical line with two slanted lines
Area (2D size)	Four squares of increasing size
Depth (3D position)	Two horizontal lines with dots at ends
Color luminance	Four squares of increasing grayscale value
Color saturation	Four squares of increasing color intensity

More Data Marks



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

Georgia Tech

Basics of Visual Representation

- binding data values to visual glyphs
- emphasize salient features of data for perceptual system
- augments cognitive processes of people to reason about data

a simple example; TableLens

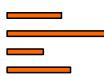
- Spreadsheet is certainly one hypervariate data presentation
- Idea: Make the text more visual and symbolic
- Just leverage basic bar chart idea

Rao & Card
CHI '94

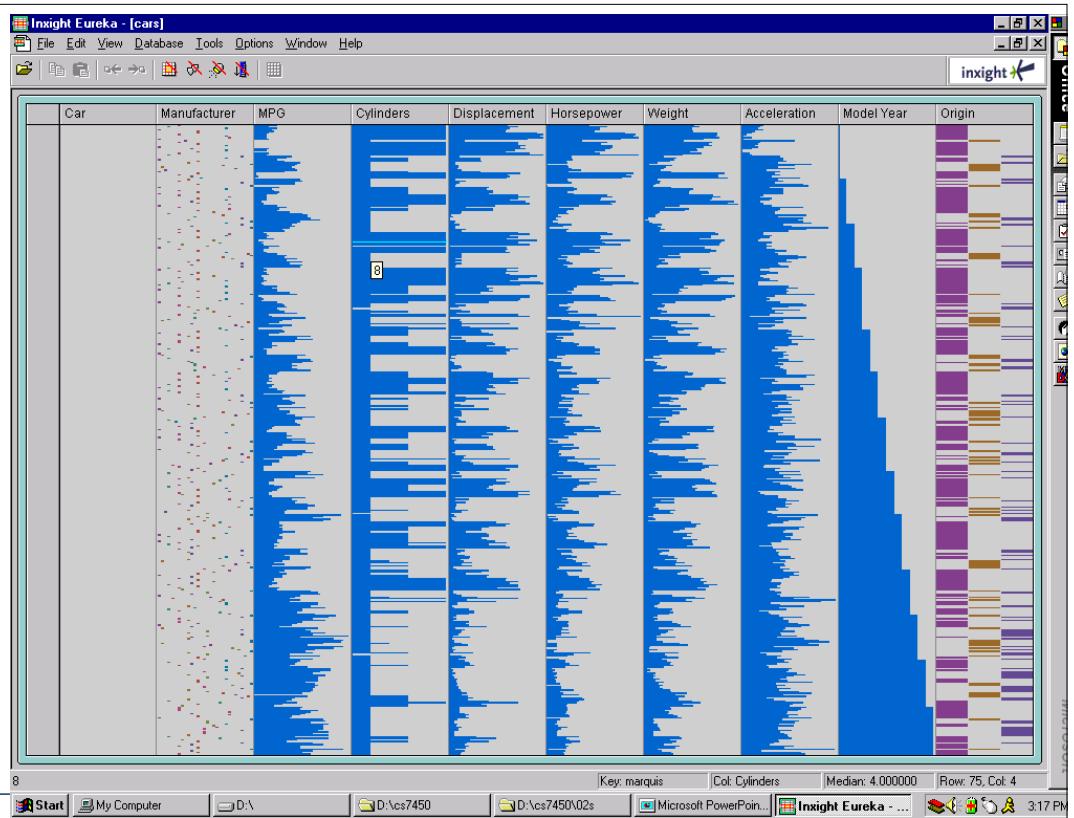
Visual Mapping

	A	B	C	D	E	F
1	Sales rep	Quota	Variance to quota	% of quota	Forecast	Actual bookings
2	Albright, Gary	200,000	-16,062	92	205,000	183,938
3	Brown, Sheryll	150,000	84,983	157	260,000	234,983
4	Cartwright, Bonnie	100,000	-56,125	44	50,000	43,875
5	Caruthers, Michael	300,000	-25,125	92	324,000	274,875
6	Garibaldi, John	250,000	143,774	158	410,000	393,774
7	Girard, Jean	75,000	-48,117	36	50,000	26,883
8	Jones, Suzanne	140,000	-5,204	96	149,000	134,796
9	Larson, Terri	350,000	238,388	168	600,000	588,388
10	LeShan, George	200,000	-75,126	62	132,000	124,874
11	Levenson, Bernard	175,000	-9,267	95	193,000	165,733
12	Mulligan, Robert	225,000	34,383	115	275,000	259,383
13	Tetracelli, Sheila	50,000	-1,263	97	50,000	48,737
14	Wotisek, Gillian	190,000	-3,648	98	210,000	186,352
15						

Change quantitative
values to bars

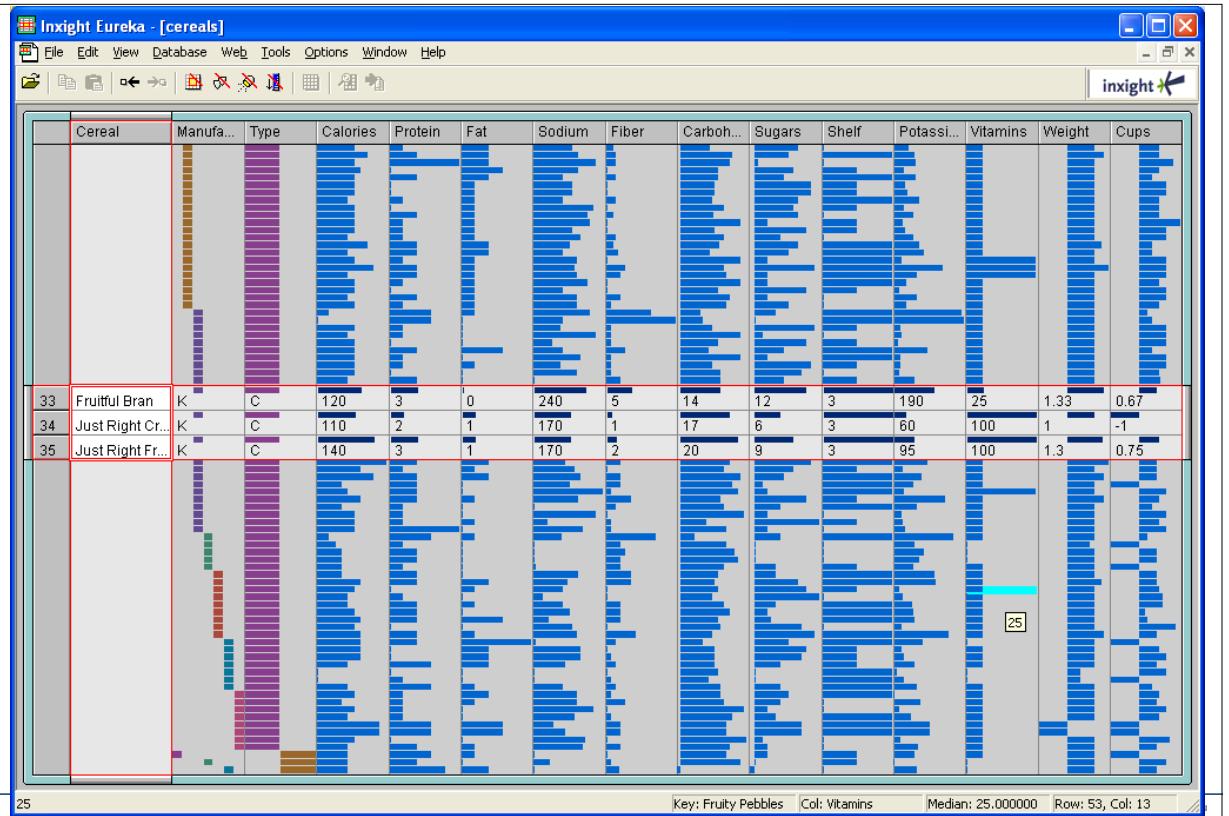


Instantiation



Details

Focus on item(s) while showing the context



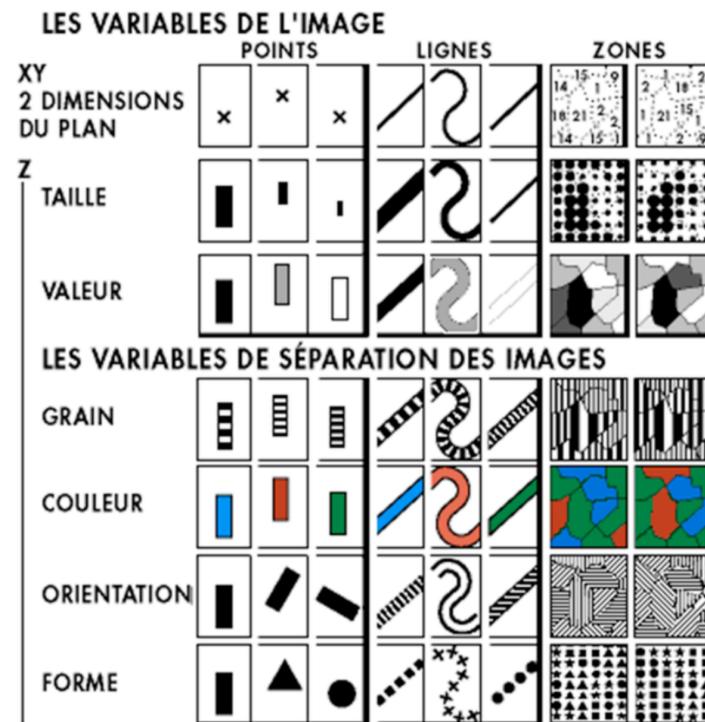
Tricky Part

	A	B	C	D	E	F	G	H	I
1	Cereal	Manufactur	Type	Calories	Protein	Fat	Sodium	Fiber	Carbo
2	Frosted Mini-Wheats	K	C	100	3	0	0	0	3
3	Raisin Squares	K	C	90	2	0	0	0	2
4	Shredded Wheat	N	C	80	2	0	0	0	3
5	Shredded Wheat 'nBran	N	C	90	3	0	0	0	4
6	Shredded Wheat spoon	N	C	90	3	0	0	0	3
7	Puffed Rice	Q	C	50	1	0	0	0	0
8	Puffed Wheat	Q	C	50	2	0	0	0	1
9	Maypo	A	H	100	4	1	0	0	0
10	Quaker Oatmeal	Q	H	100	5	2	0	0	2.7
11	Strawberry Fruit Wheats	N	C	90	2	0	15	3	
12	100% Natural Bran	Q	C	120	3	5	15	2	
13	Golden Crisp	P	C	100	2	0	45	0	
14	Smacks	K	C	110	2	1	70	1	
15	Great Grains Pecan	P	C	120	3	3	75	3	
16	Cream of Wheat (Quick)	N	H	100	3	0	80	1	
17	Com Pops	K	C	110	1	0	90	1	
18	Muesli Raisins, Dates, & R	R	C	150	4	3	95	3	
19	Apple Jacks	K	C	110	2	0	125	1	

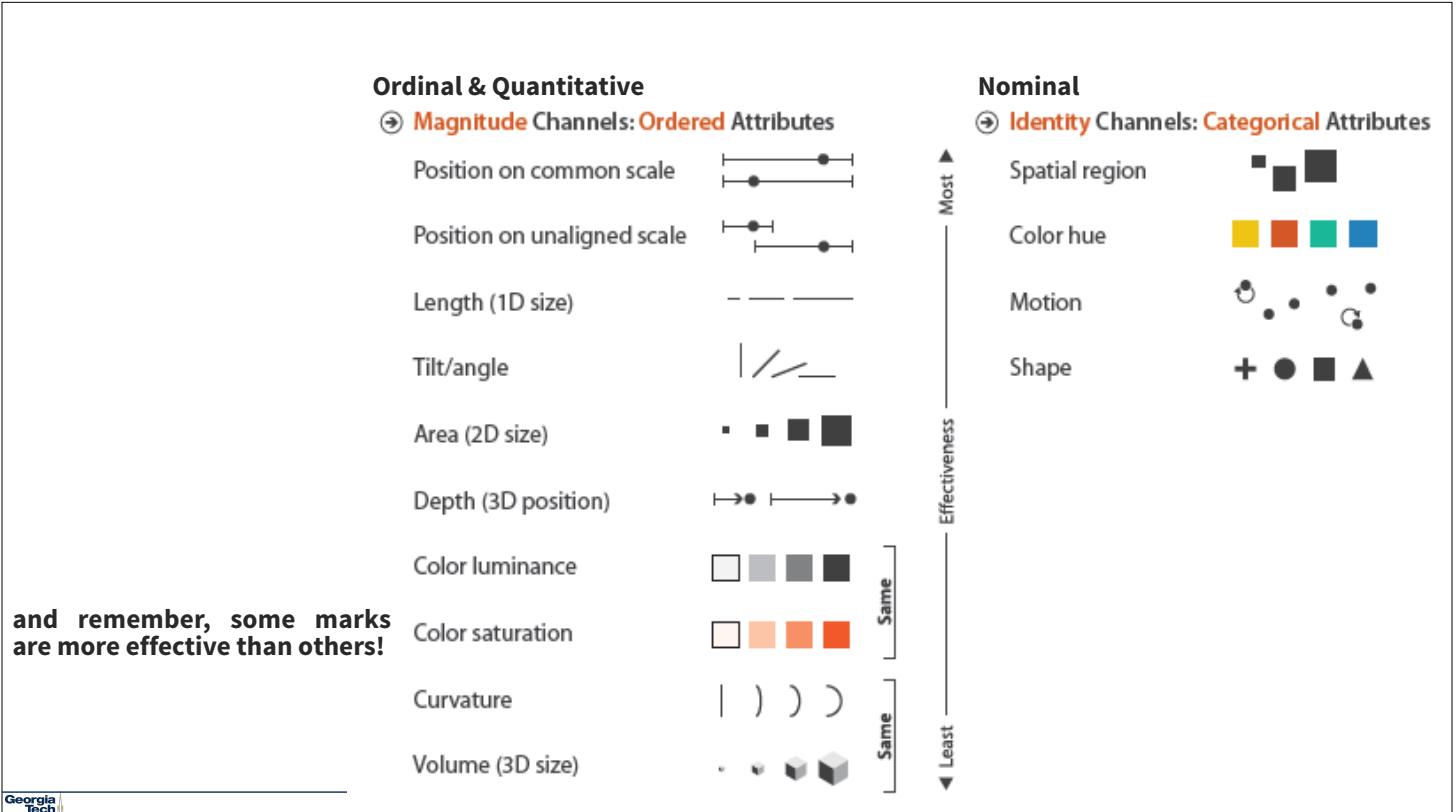
What do you do for nominal data?

Visual Variables

Position
Size
Value
Color
Texture
Orientation
Shape



Bertin, Jacques. "Semiology of graphics: diagrams, networks, maps." (1983).



How many visual encodings to use in 1 vis?

- Can only show limited amount of data attributes
 - map each feature to a visual variable
- That means for lots of complex “big” data, we need methods to aggregate or combine data attributes
 - even then, we may still need more help
 - e.g., dimension reduction - will cover this later

A note about terminology

- We have a problem in vis, we can't make up our mind about what to call things
- **Data items**
 - also commonly called data **cases**
- **Data attributes**
 - also often referred to as data **features**, data **dimensions**, data **variables**
- **Data marks**
 - also often called **visual encodings**, visual **marks**, **graphical encodings**
- We will use these terms interchangeably throughout this course

a note about color

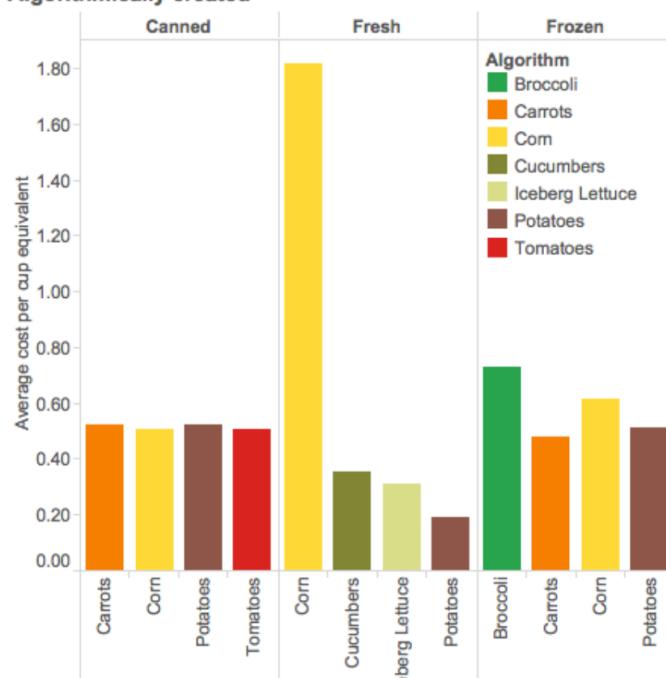
- color is tricky
- lots of research on color ramps, color theory, etc.
- I won't go into details, but will say this:
 - chose wisely
 - ordering of color should be used when data has order
 - separation in color ramp should indicate separation in data
 - there are limits to the number of items/colors people can distinguish between (~7-9)
 - less is more!

picking a color scheme

The screenshot shows the Color Brewer 2.0 interface. At the top, it says "Number of data classes: 3" and "Nature of your data: sequential". Below that, there are two sections: "Multi-hue:" and "Single hue:", each showing a grid of color swatches. Under "Multi-hue:", there are four rows of six swatches each, with colors ranging from light blue to dark purple. Under "Single hue:", there are three rows of five swatches each, with colors ranging from dark blue to black. To the right of these is a map of the United States where counties are colored according to a sequential color scheme. A legend on the left indicates a "3-class BuGn" scheme with three categories: light blue (#e5f5f9), medium blue (#99d8c9), and dark blue (#2ca25f). There are also checkboxes for "colorblind safe", "print friendly", and "photocopy safe". On the far left, there's a Georgia Tech logo.

another interesting way to pick a color scheme

Algorithmically created



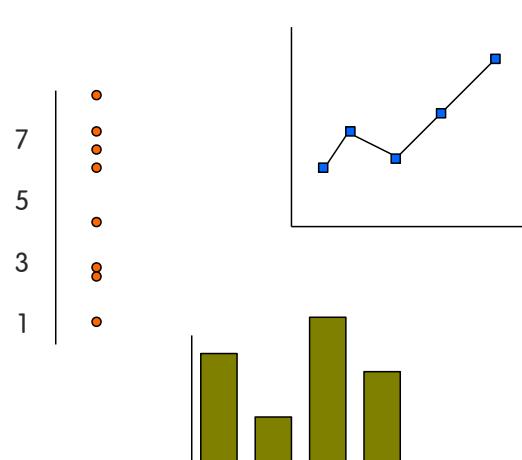
Setlur, Vidya, and Maureen C. Stone. "A Linguistic Approach to Categorical Color Assignment for Data Visualization." *Visualization and Computer Graphics, IEEE Transactions on* 22.1 (2016): 698-707.

Back to Data

- Number of variables per class
 - 1 - **Univariate** data
 - 2 - **Bivariate** data
 - 3 - **Trivariate** data
 - >3 - **Hypervariate** data

Univariate Data

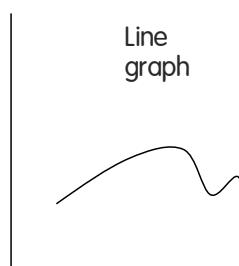
Representations



Label	Variable1
A	4
B	8
C	3
D	1

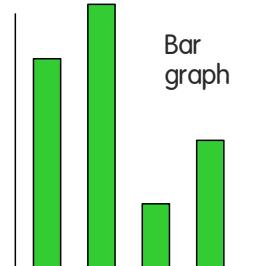
What Goes Where?

- In **univariate** representations, we often think of the data items as being shown along one dimension, and the value in another (the only attribute)



Y-axis is quantitative variable

See changes over consecutive values



Y-axis is quantitative variable

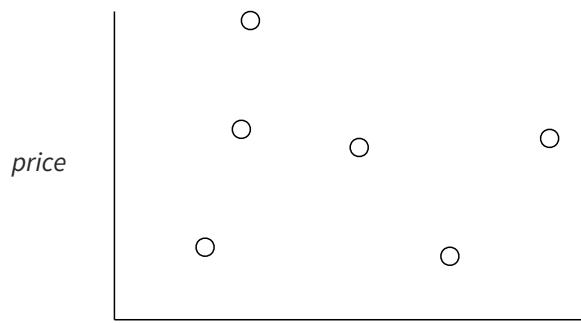
Compare relative point values

Bivariate Data

Scatterplot is common

Two variables, want to see **relationship**

Is there a linear, curved or random pattern?



Each mark is now a data item

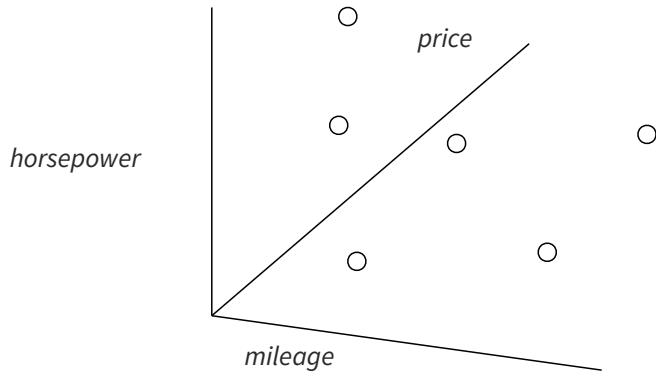
Label	price	mpg
Car A	\$25000	35
Car B	\$80000	16
Car C	\$17500	33
Car D	\$22500	38

*not exactly mapped to vis shown on left

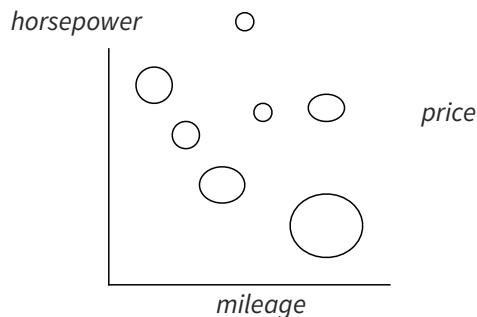
Trivariate Data

3D scatter plot is possible

but remember, 3D often isn't great

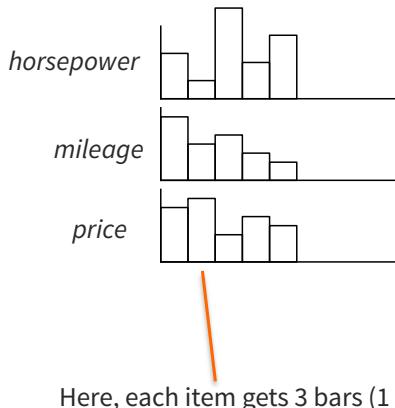


Alternative Representation



Still use 2D but have visual mark property represent third variable (size)

Alternative Representation



Represent each variable
in its own explicit way

Here, each item gets 3 bars (1 for each attribute)

Hypervariate Data $N > 3$

- Recall: Number of well-known visualization techniques exist for data sets of 1-3 attributes
 - line graphs, bar graphs, scatter plots OK
 - we just went through a few
- What about data sets with more than 3 attributes?
 - Often the interesting, challenging ones
 - Could use additional data mark properties to encode additional data variables

Let's try it

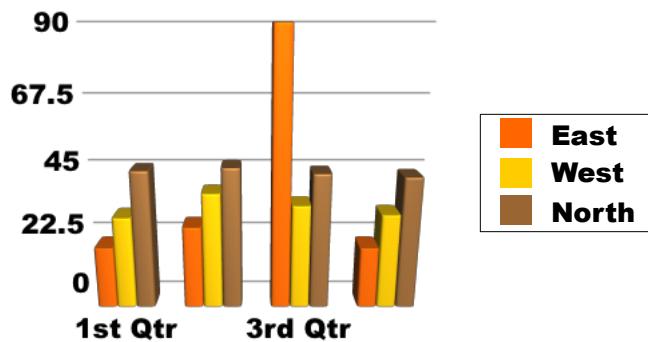
Partner up and answer these questions

How many variables in the multivariate chart?

What are their types?

How many cases, items?

What's done poorly in this vis?



Let's try it

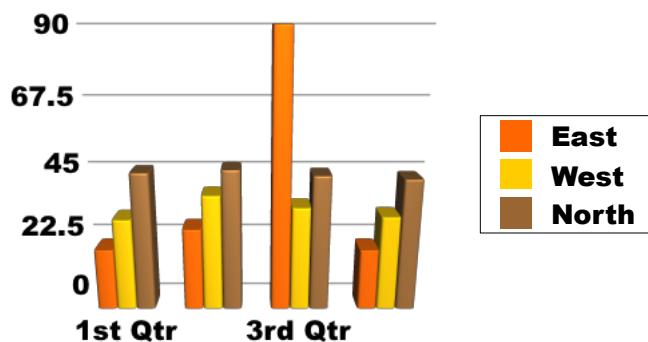
Partner up and answer these questions

How many variables in the multivariate chart? 3 (trivariate)

What are their types?

How many cases, items?

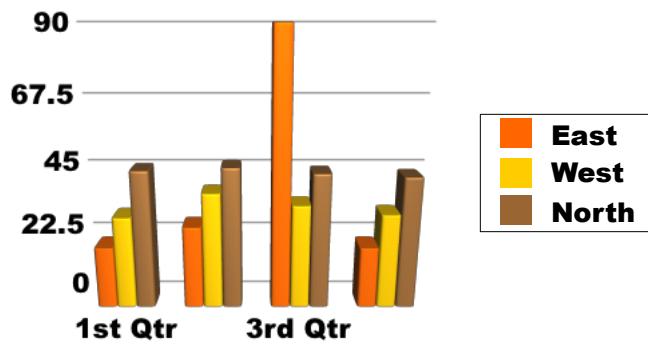
What's done poorly in this vis?



Let's try it

Partner up and answer these questions

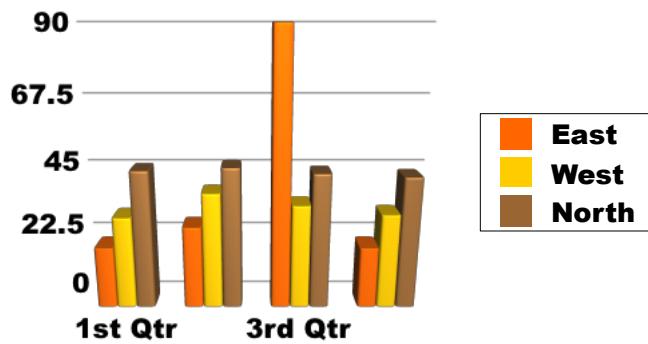
- How many variables in the multivariate chart? 3 (trivariate)
Region – Nominal
Sales – Ratio
Quarter – Ordinal
- What are their types?
- How many cases, items?
- What's done poorly in this vis?



Let's try it

Partner up and answer these questions

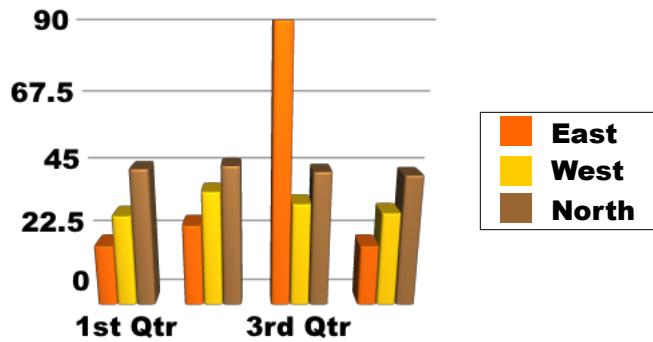
- How many variables in the multivariate chart? 3 (trivariate)
Region – Nominal
Sales – Ratio
Quarter – Ordinal
- What are their types?
- How many cases, items? 12
- What's done poorly in this vis?



Let's try it

Partner up and answer these questions

- How many variables in the multivariate chart? 3 (trivariate)
Region – Nominal
Sales – Ratio
Quarter – Ordinal
- What are their types?
- How many cases, items? 12
- What's done poorly in this vis? 3d? label axes?

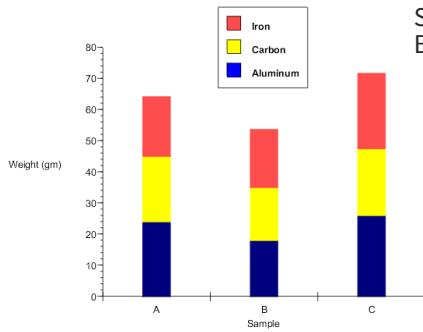
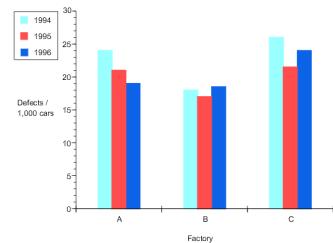
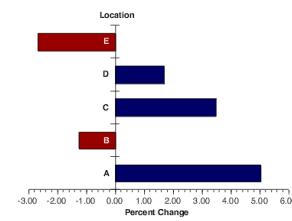
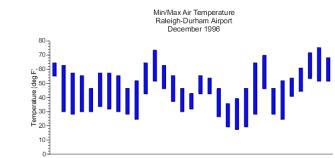


The Data Table

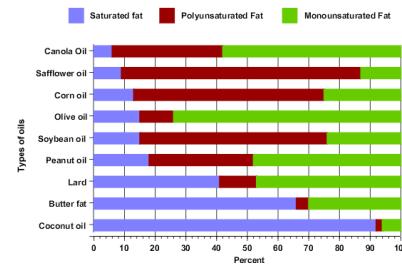
*	Region	Quarter	Sales
	East	1	20
	East	2	28
	East	3	88
	East	4	20
	West	1	30
	West	2	38
	West	3	36
	West	4	31
	North	1	45
	North	2	46
	North	3	44
	North	4	43

* and you would typically have an ID for each item in a separate column, which would not count as one of the attributes

Gallery of Bar Charts



Stacked Bar Chart



Stacked Bar Chart

How many variables and cases? You should be able to answer this. Try it at home.

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

It's always sunny in Philadelphia

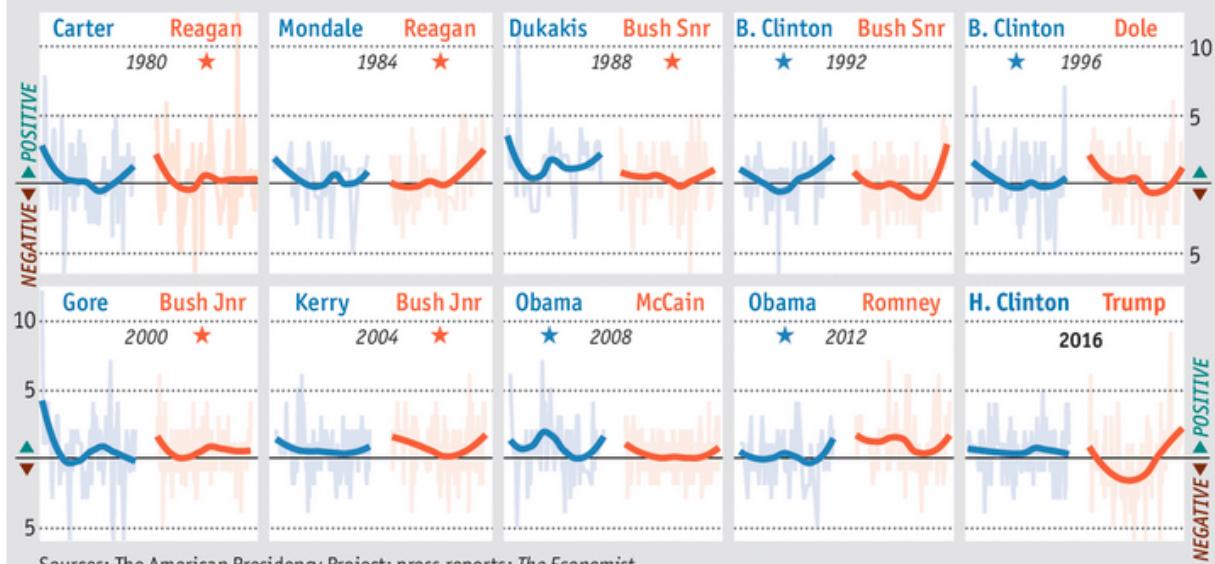
Sentiment in United States party-convention speeches

Number of positive/negative words per line

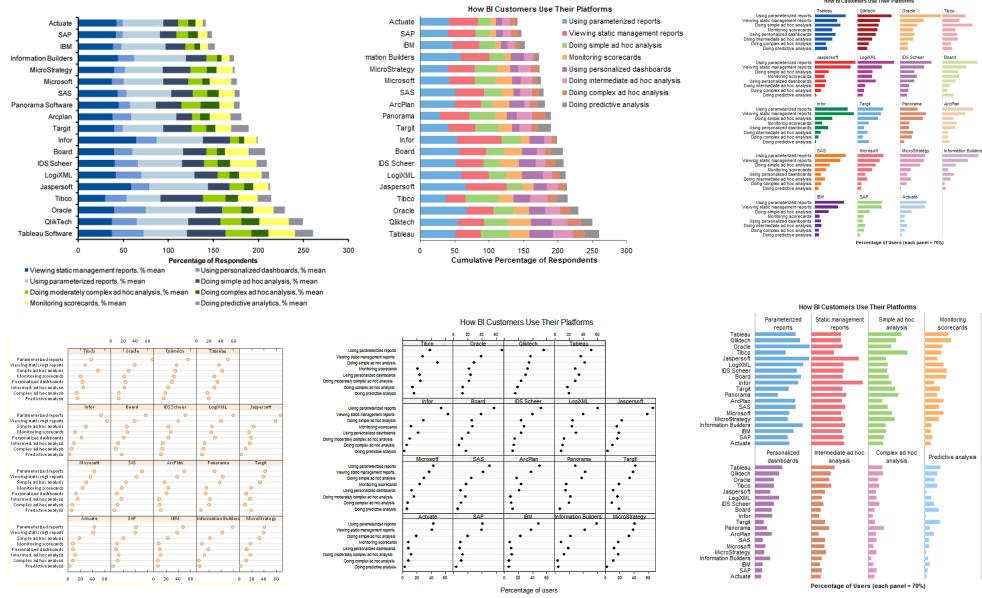
Trend: — Democrats

— Republicans

★ Became President



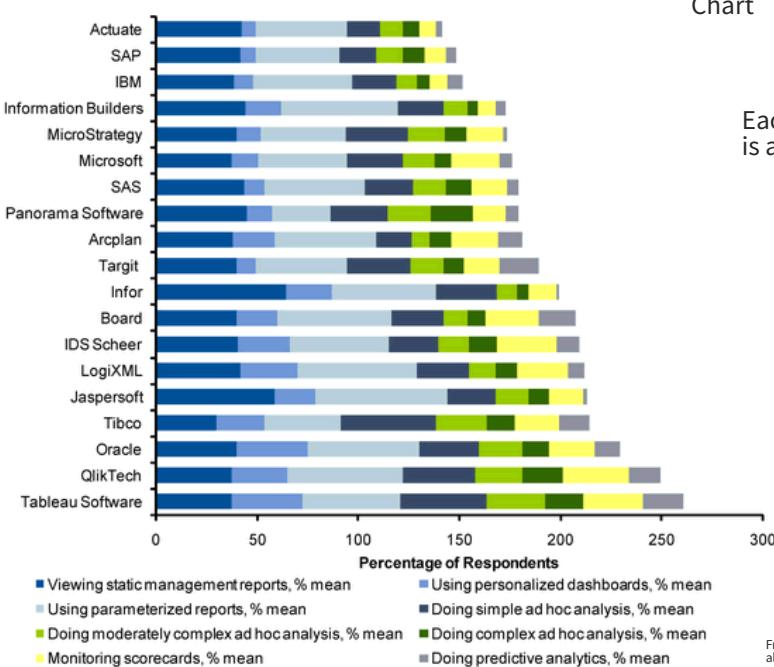
Many Ways to Present Same Data



See in detail on next slides

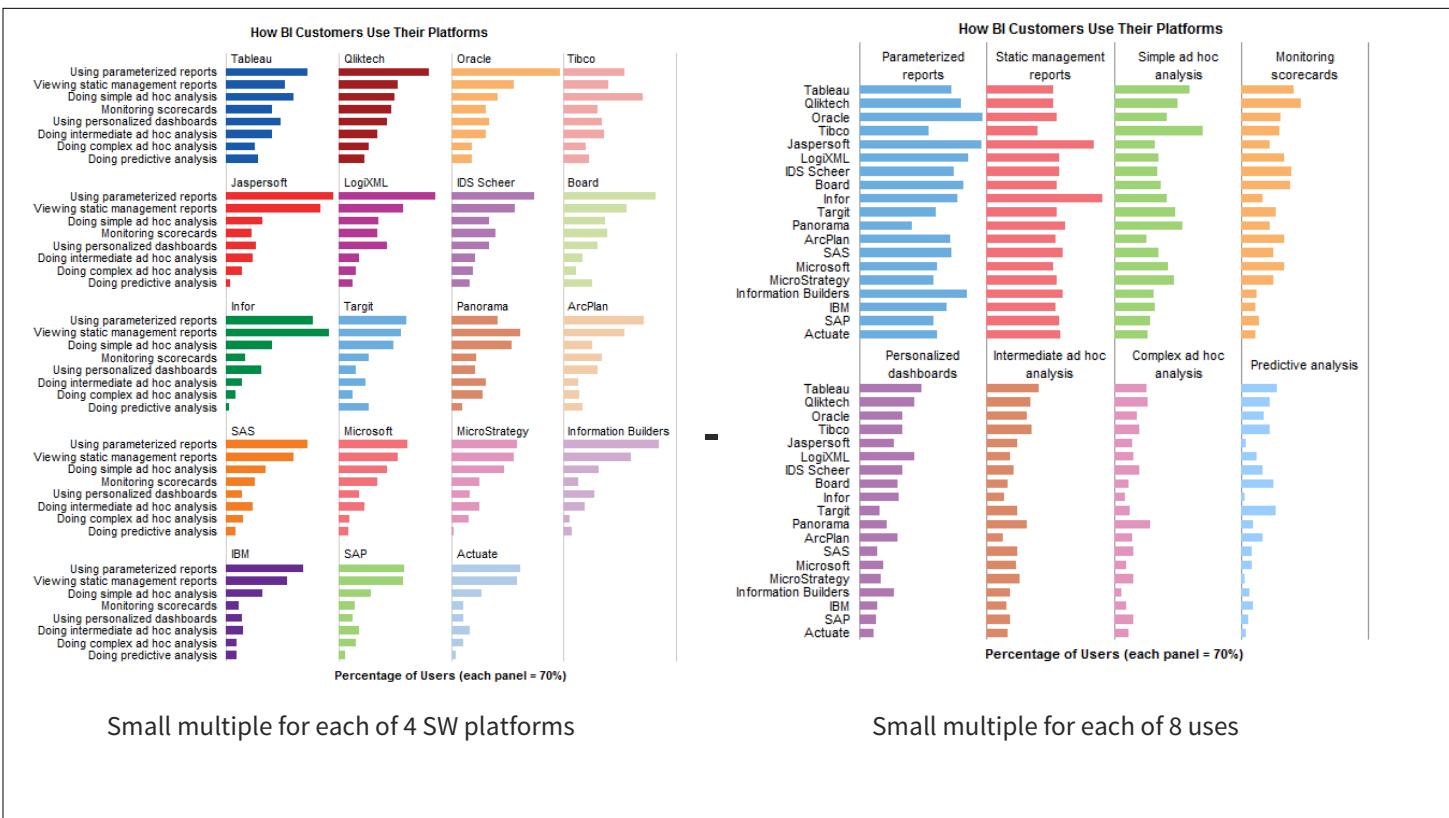
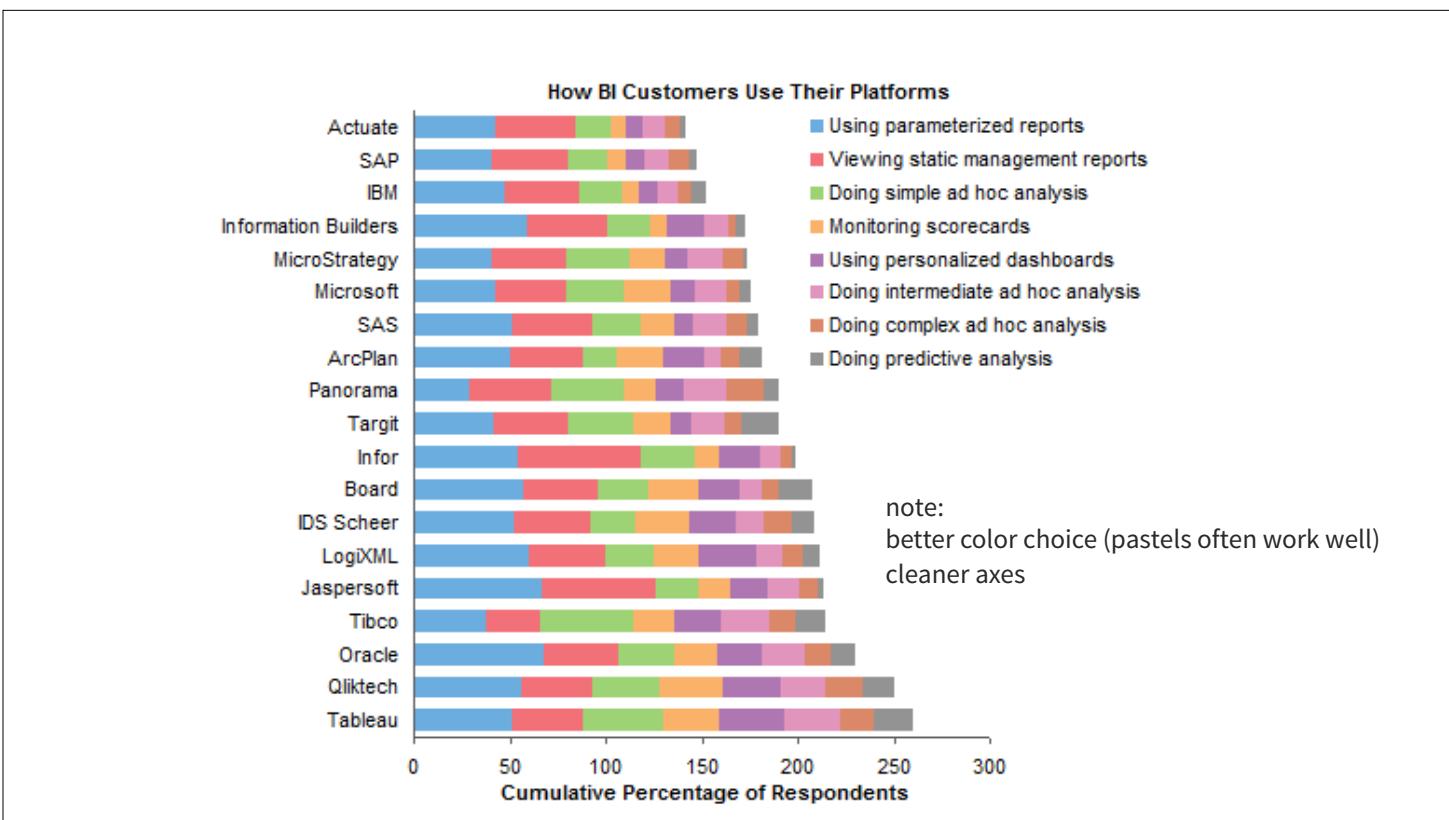
How BI Customers Use Their Platforms

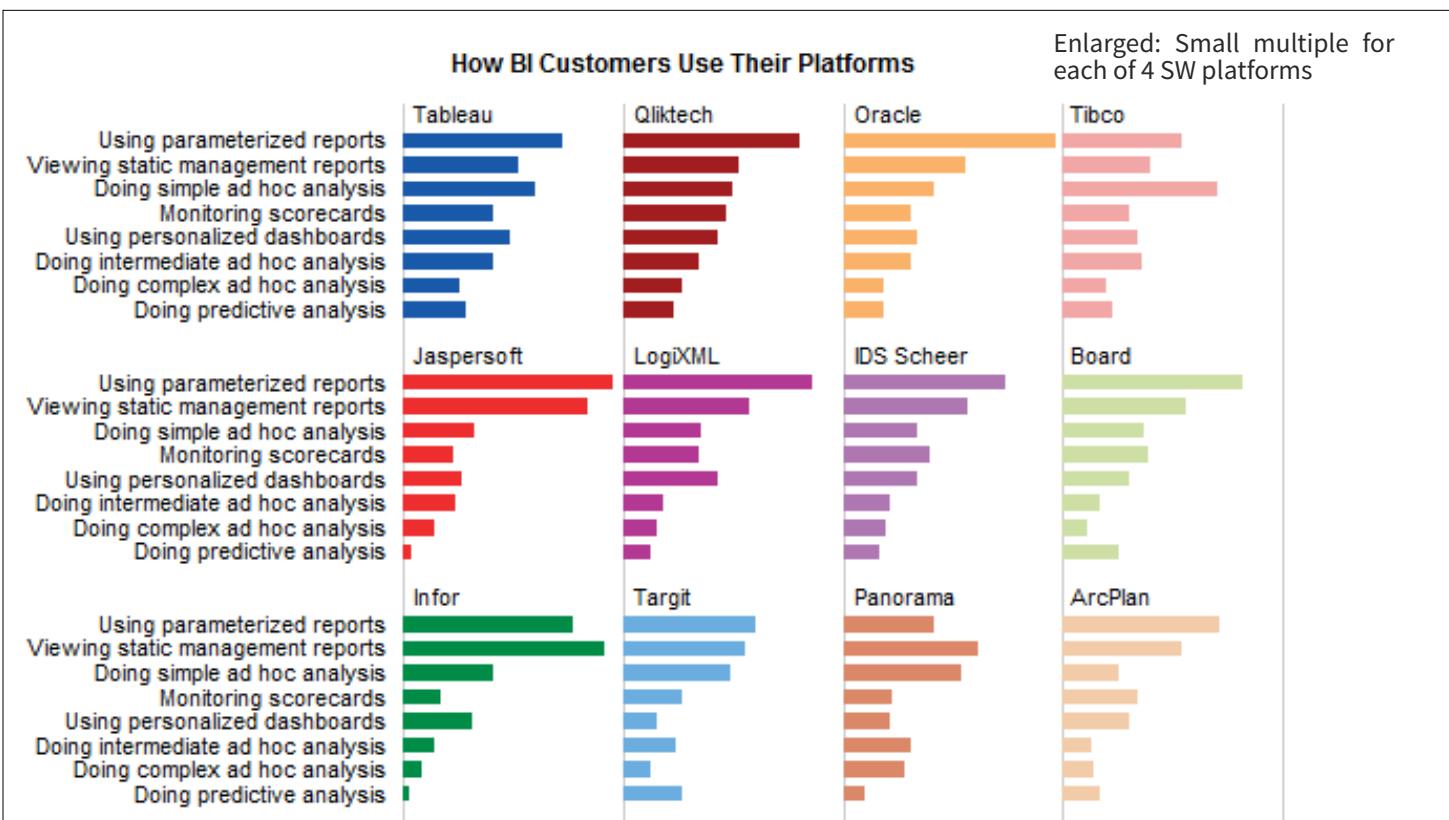
Stacked Bar Chart



Each bar segment
is a data mark

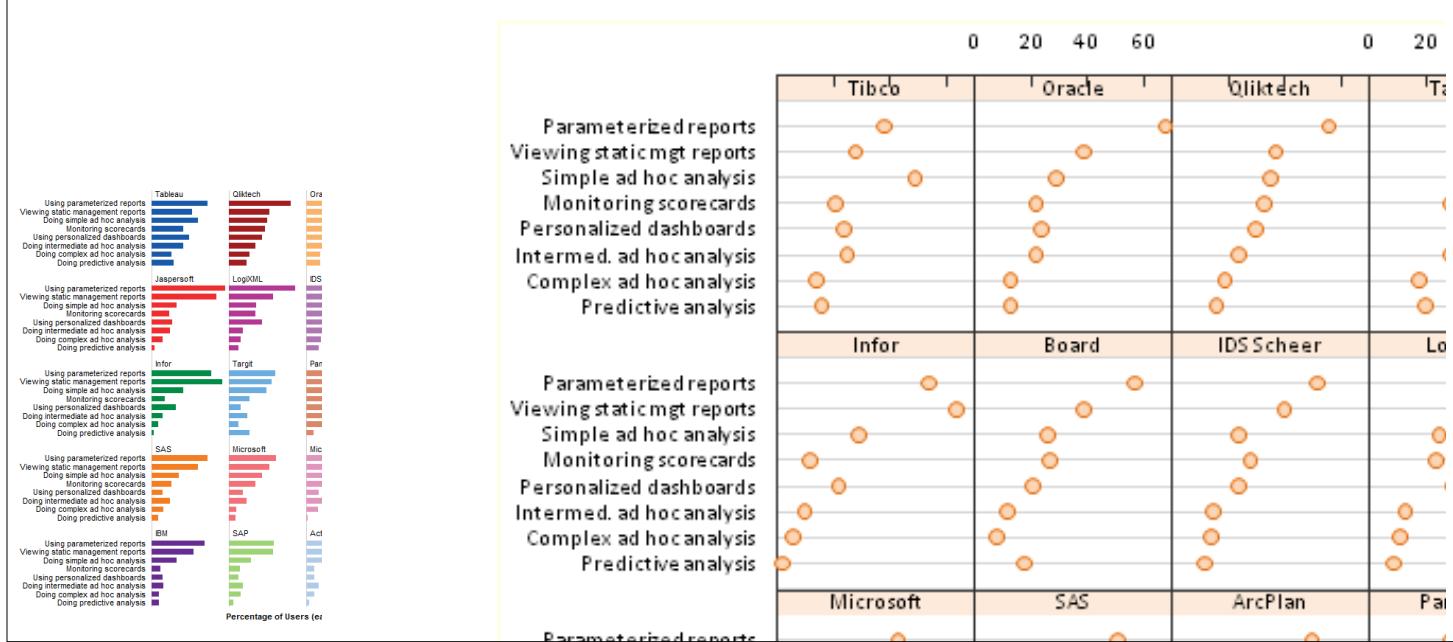
From <http://peltiertech.com/WordPress/trellis-plot-alternative-to-stacked-bar-chart/>





Marks Instead of Bars

primary encoding is now position along a common axis



let's try it

- Work with someone near you
- Get out a piece of paper
- Visualize these two numbers:
 - 11
 - 27
- Take 5 minutes, be prepared to come up and show alternatives
- Be prepared to talk about the **visual marks/encodings** you used

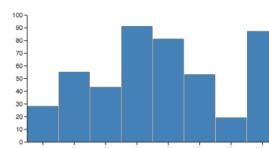
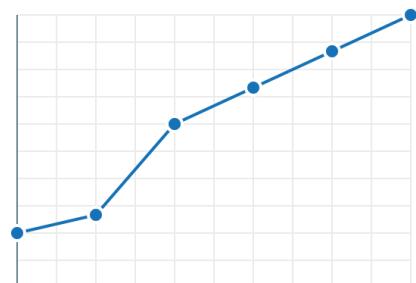
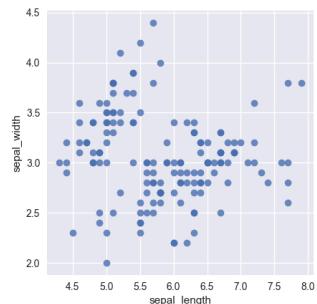
Basic Charts

Few's Selection & Design Process

- Determine your message and identify your data
- Determine if a table, or graph, or both is needed to communicate your message
- Determine the best means to encode the values
- Determine where to display each variable
- Determine the best design for the remaining objects
 - Determine the range of the quantitative scale
 - If a legend is required, determine where to place it
 - Determine the best location for the quantitative scale
 - Determine if grid lines are required
 - Determine what descriptive text is needed
- Determine if particular data should be featured and how

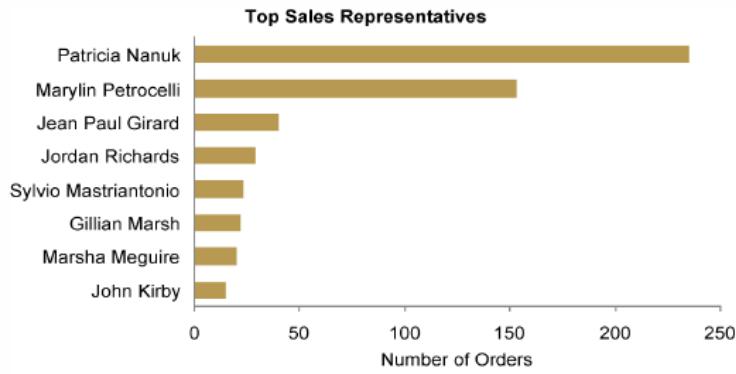
Points, Lines, Bars, Boxes

- **Points**
 - Useful in scatterplots for 2-values
 - Can apply shapes, color for additional variable to encode
- **Lines**
 - Connect values in a series
 - Show changes, trends, patterns
 - Not for a set of nominal or ordinal values
- **Bars**
 - Emphasizes individual values
 - Good for comparing individual values (when along common axis)
- **Boxes**
 - Shows a distribution of values



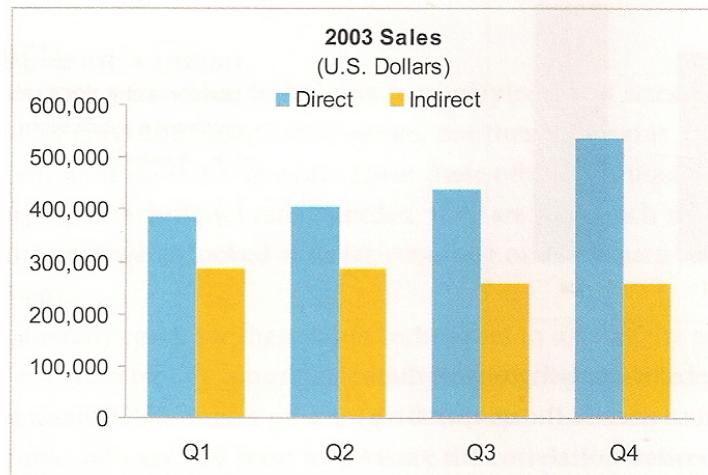
Vertical vs. Horizontal Bars

- Horizontal can be good if long labels or many items



Multiple Grouped Bars

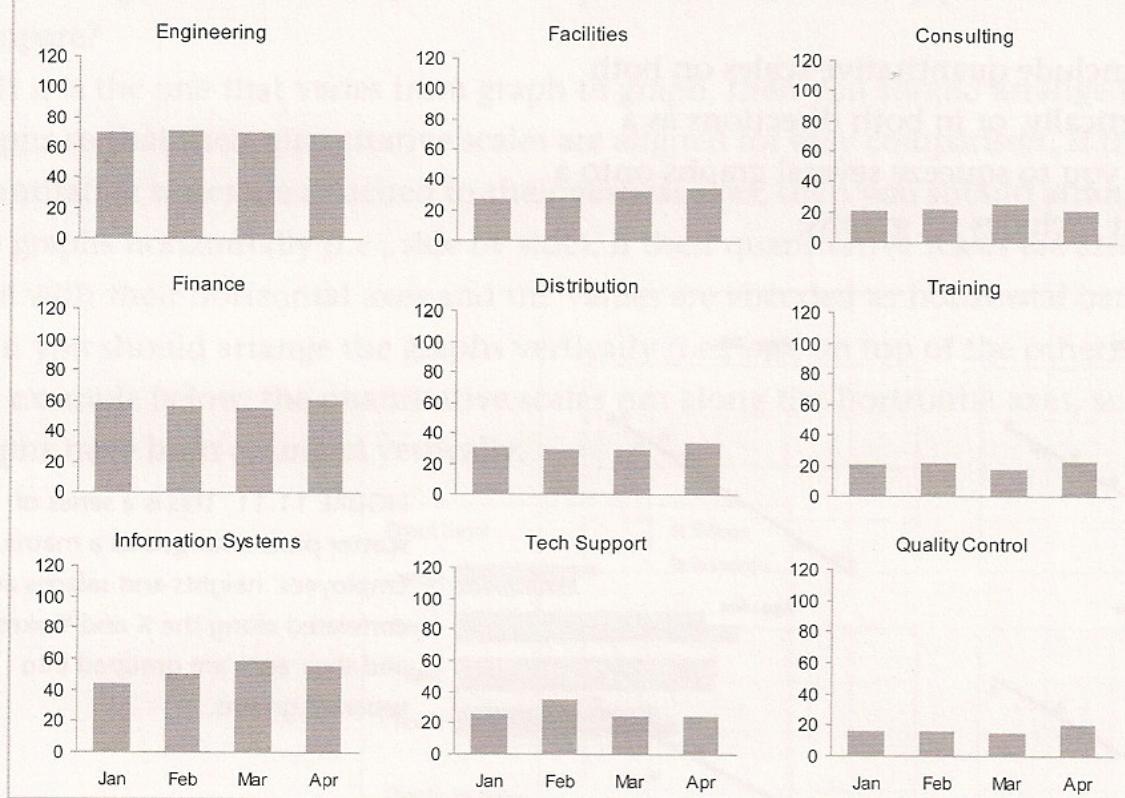
- Can be used to encode another variable



Multiple Graphs

- Can distribute a variable (\$) across graphs too

Sometimes called a trellis display



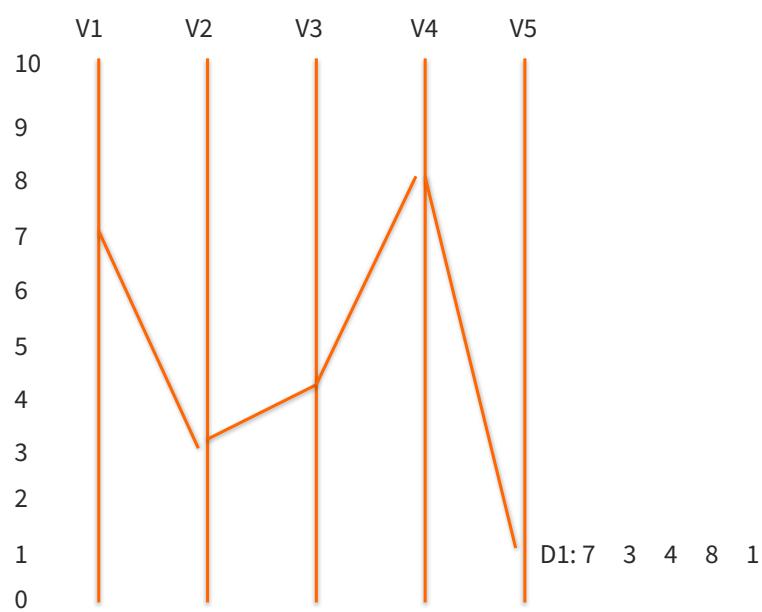
What about Hypervariate data?

Parallel Coordinates

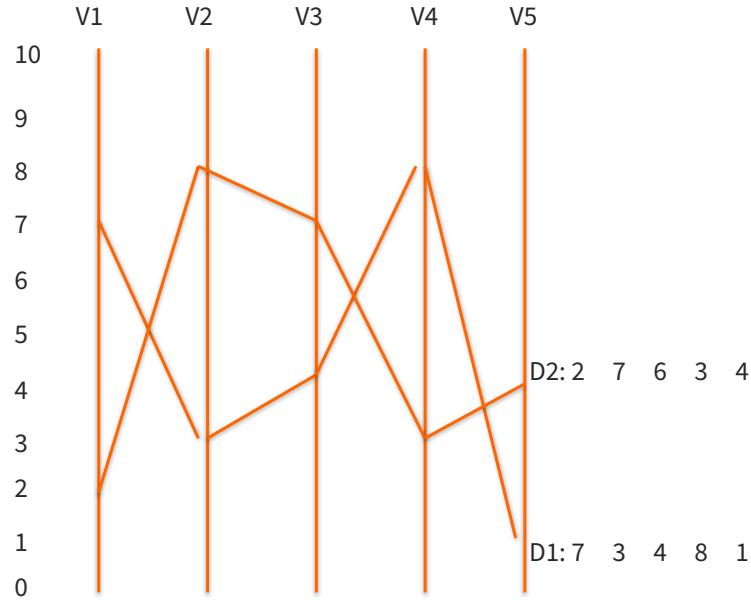
- Given this data table

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

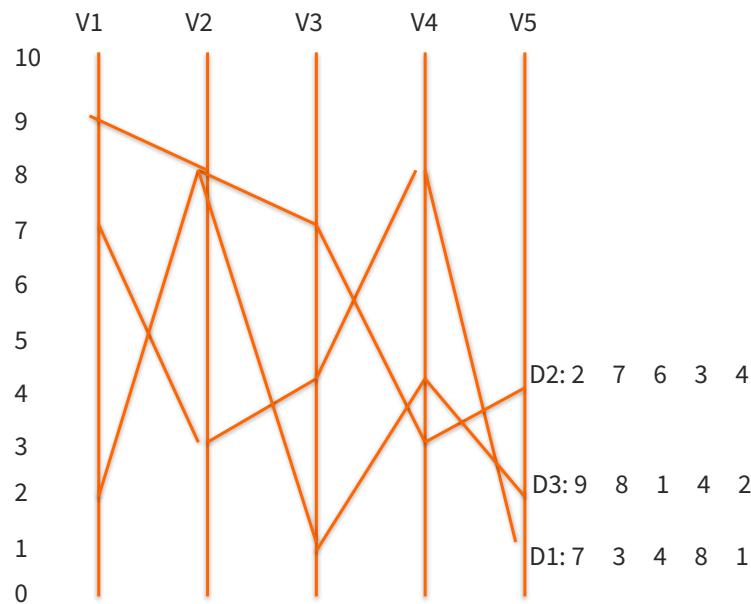
Parallel Coordinates



Parallel Coordinates

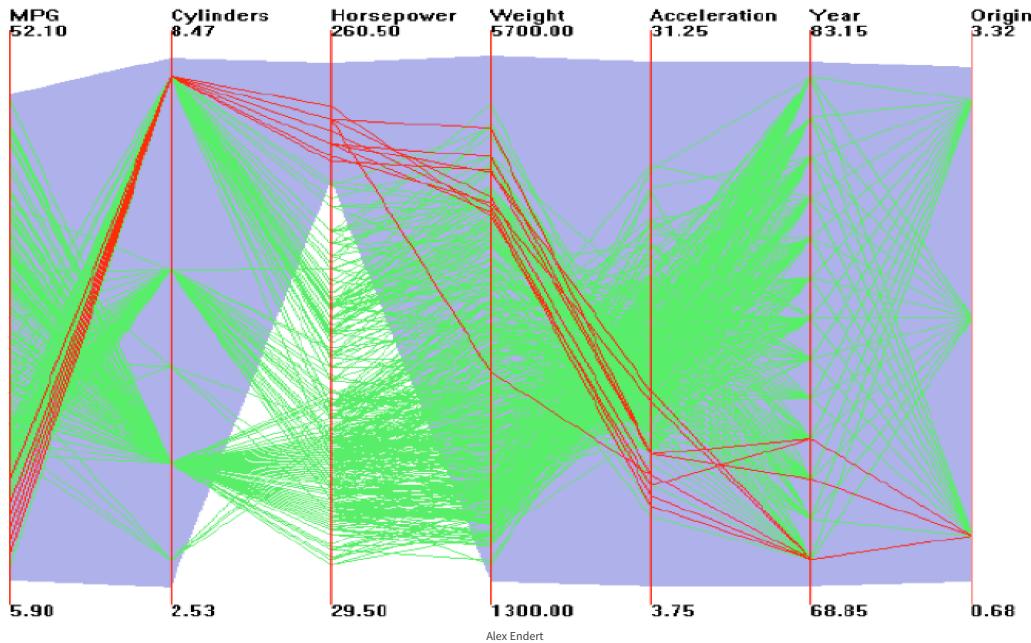


Parallel Coordinates



Automobile Data in Parallel Coords

How well do parallel
coordinates scale with
cases?
variables?



Dimensional Reordering

Changing the order of dimensions creates new visual patterns

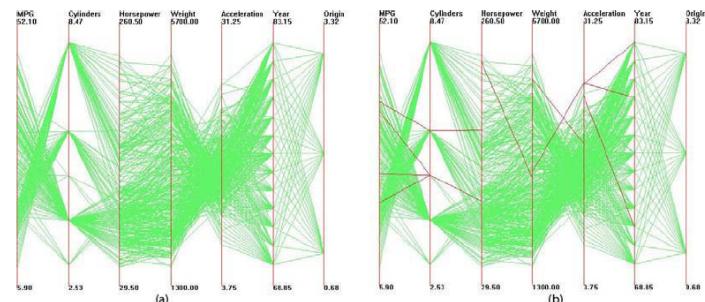
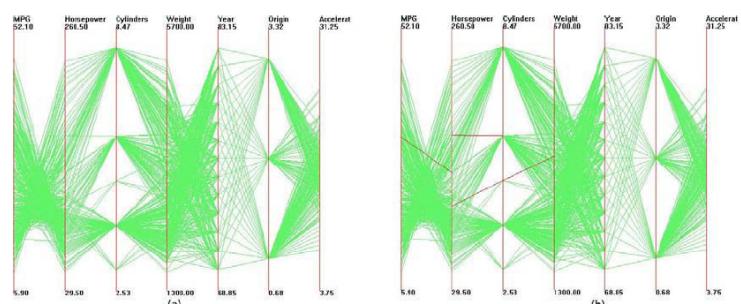
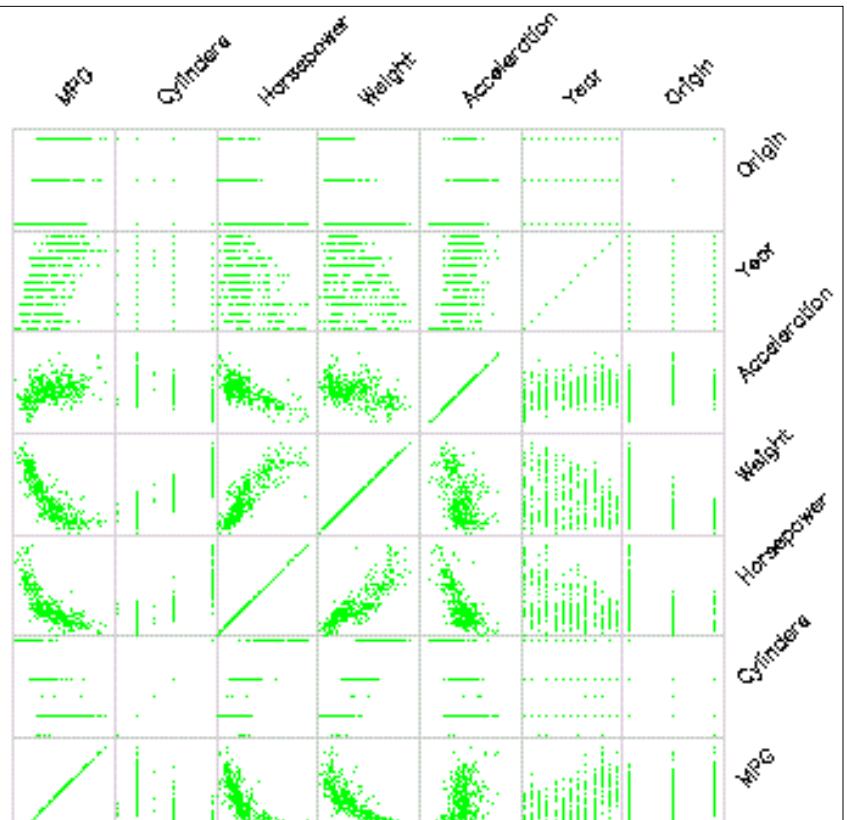


Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).



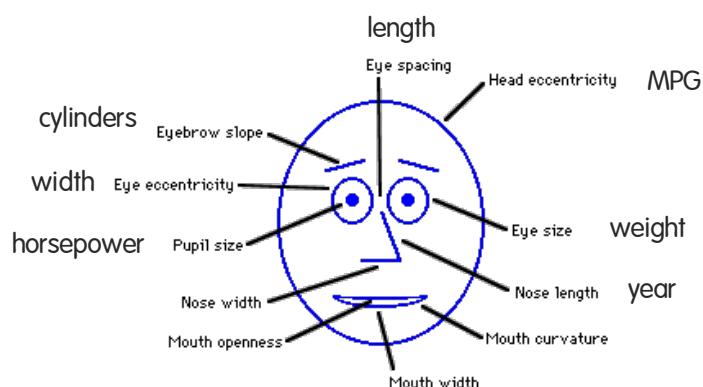
Automobile Data in Scatterplot Matrix

Small multiples: each pair of variables in scatterplot

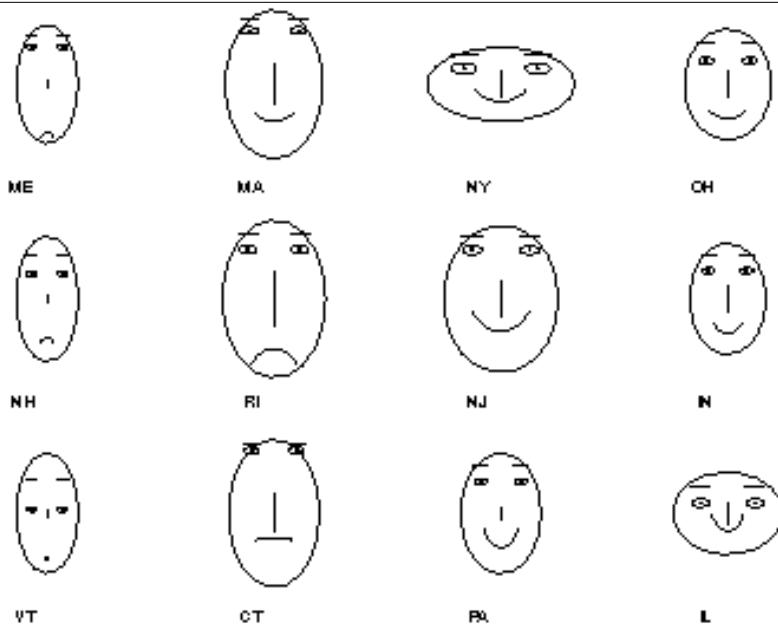


Chernoff Faces

Encode different variables' values in characteristics of human face

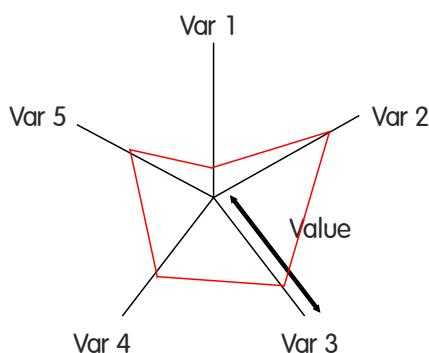


Examples



<http://www.cs.uchicago.edu/~wiseman/chernoff/>
<http://hesketh.com/schampeo/projects/Faces/chernoff.html>

Star Plots



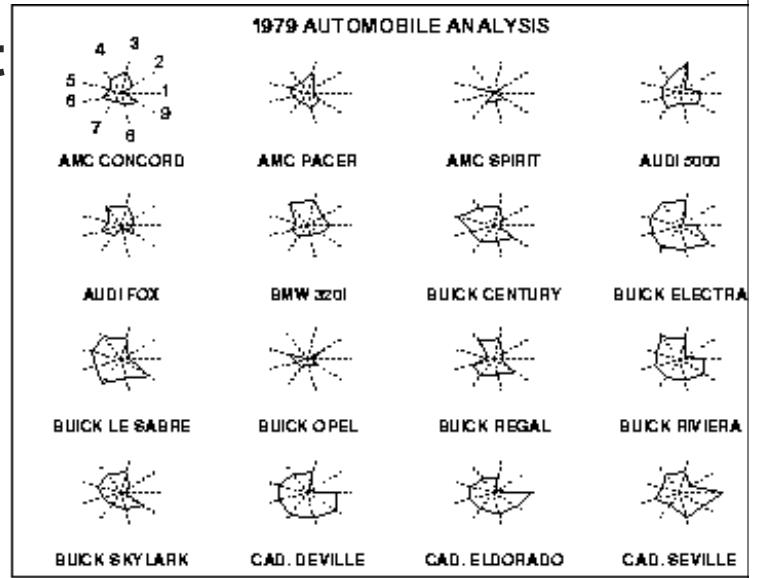
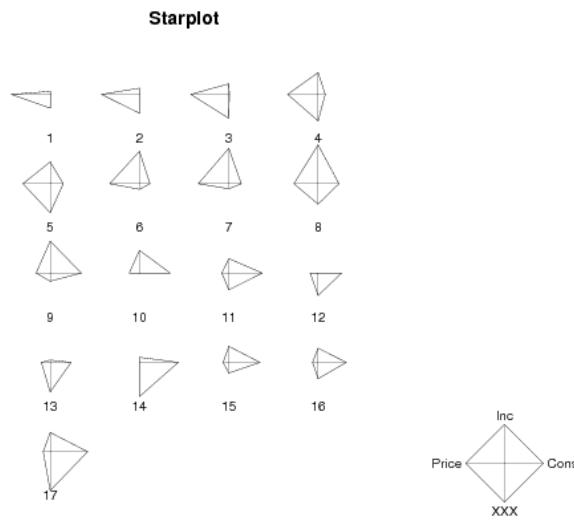
Space out the n
variables at equal
angles around a
circle

Each “spoke” encodes
a variable’s value

Alternative Rep.

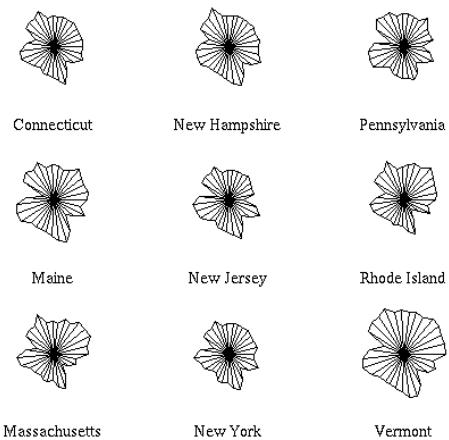
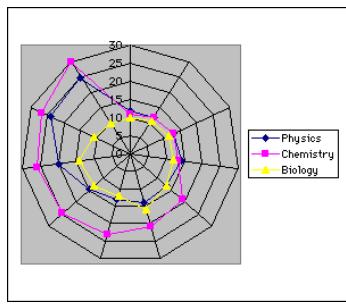
Data point is now a “shape”

Small Multiples: Star Plot



N = 4 (5 if include case index/number); created at

Star Plot examples



<http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/compl.html>

example I just saw the other day

not exactly a star plot, but similar principles



endert@gatech.edu

Selection of Key Points from Today

- Data types & marks
 - ranking of basic visual encodings
- Lots of ways to vis multivariate data
 - should be able to draw these given 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