

# PROGRAMMING INTERACTIVE VISUALIZATIONS

WEEK 1: PERCEPTION AND PRINCIPLES,  
VIS ON THE WEB

# WHAT WE ARE DOING IN THIS COURSE

Visualization Overview: why, what and how

Color

Data Types:

- Hierarchical
- Statistical
- Geospatial
- Categorical
- Graph
- Text
- High-Dimensional

Interaction design principles

Multiple coordinated views

Our web stack (python, HTML, CSS, SVG, JavaScript, D3)

Using D3 to change a page and create visualizations interactively

Maps in D3 and adopting vis examples

Interaction between Visualizations

Python back-ends

# WHAT WE DID LAST TIME

Preliminaries for the course and introductions

## Visualization Overview

- Finishing week 1
- Perception
- Data vis on the web

## Lab on HTML and SVG

# WHAT WE ARE DOING TODAY

Where we left off

- Good practice in basic visualization

Color

Hierarchical data

Lab: intro to D3

# ANNOUNCEMENTS

Project specification will be posted as soon as possible. Probably tonight.

Preview is that it will be groups of 4-5 people working with data of your choice to create an interactive visualization

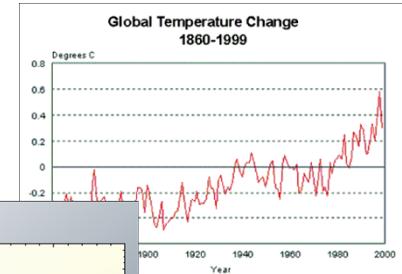
We need to start forming groups right away

**Due this weekend: post to the forums about project interests and data you have available**

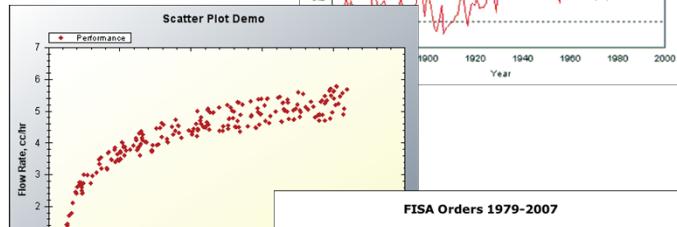
You will then have a week to actually form groups and then I will finalize them and make sure everyone has a group.

# BASIC GRAPH TYPES

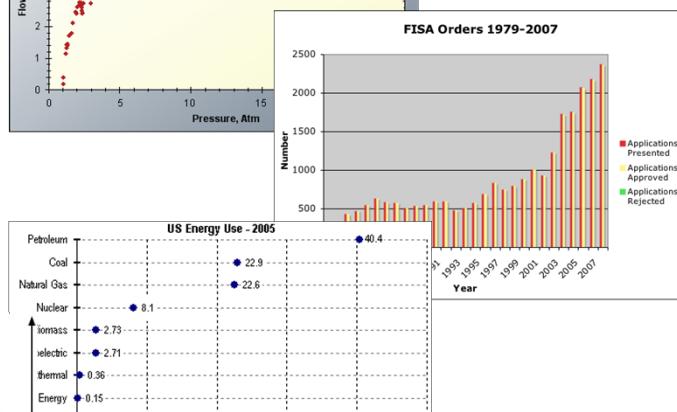
Line



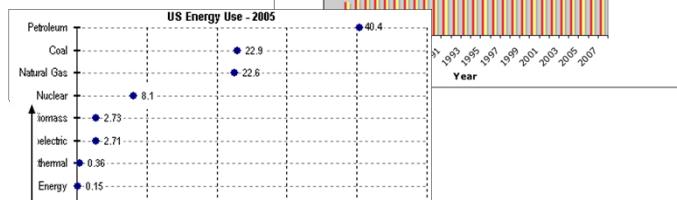
Scatter Plot



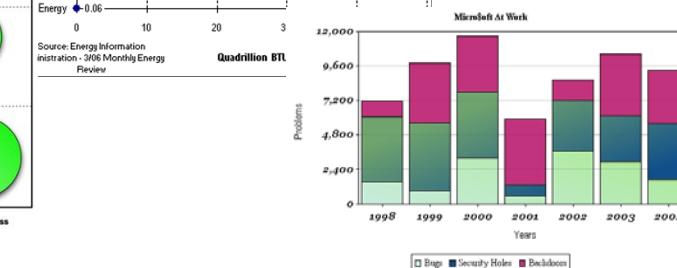
Column (bar)



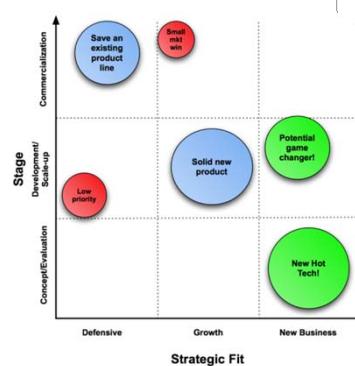
Dot-Chart (like horizontal bar)



Stacked Bar



Area



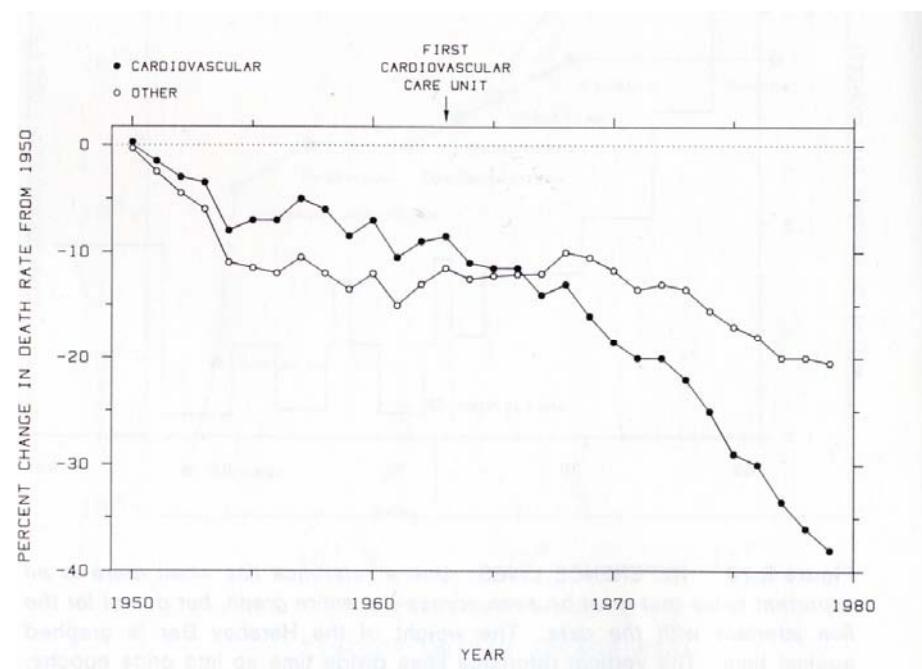
Bubble

# LINE GRAPHS

Points are connected

- Connections have meaning

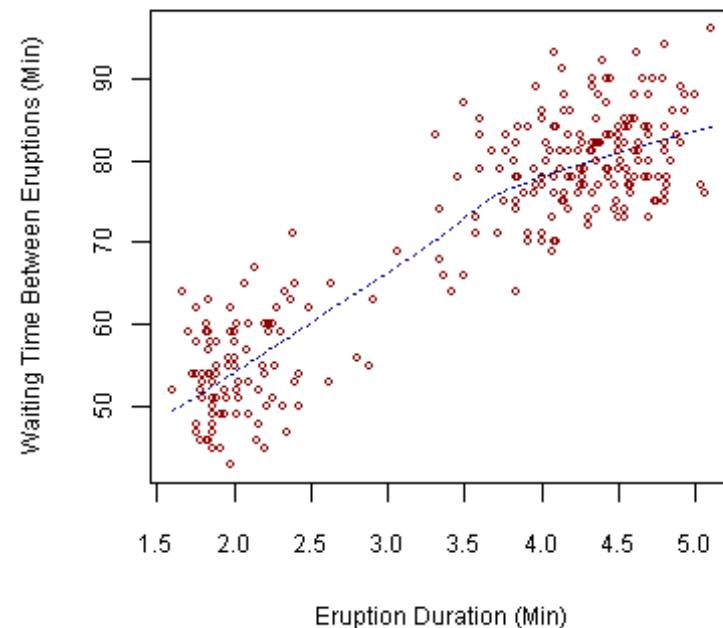
Continuous change and rate of change



# SCATTER PLOTS

- Data are not a series
  - May be more than one y-value for each x-value
- Stochastic data!

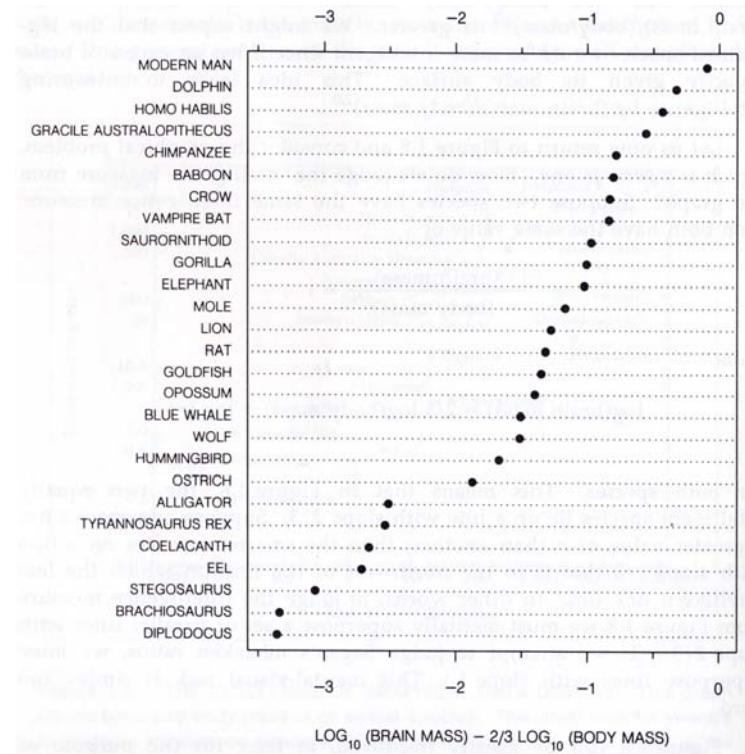
**Old Faithful Eruptions**



# ORDERING IS OFTEN CRITICAL!

For categorical axes, never blindly use the default order

- The order should follow some aspect of the data
- Unless alphabetical order is important for reading the graph!



# EXERCISE

- How many variables can you fit into a line graph?
- What about a bar chart?
- Compare to scatter plot?
- For 2-dimensional data, are these two visualizations the same?

## EXERCISE

- Scatter / Bar / Line have different uses
- Scatter can have multiple Y values at the same X
- Line implies some connection
- Bar is (generally) the best option for categorical data in X

# BAR VS LINE

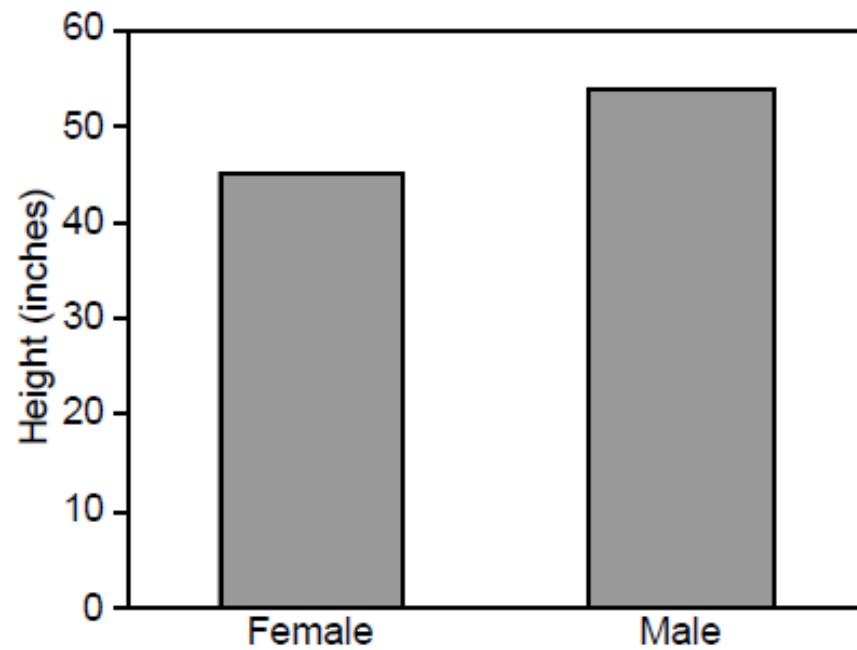


Image courtesy of Barbara Tversky

# BAR VS LINE

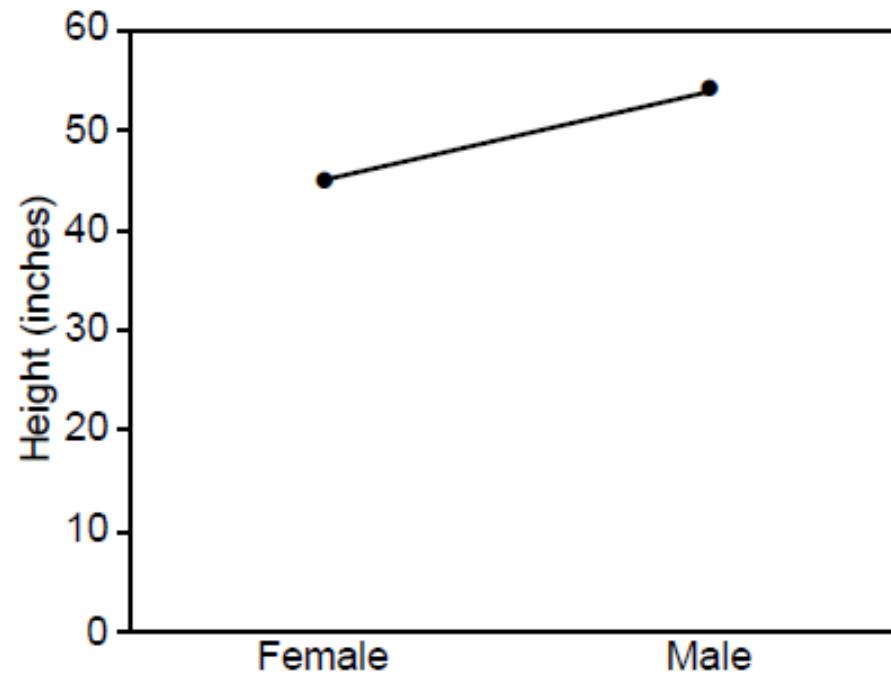


Image courtesy of Barbara Tversky

# PRINCIPLES FOR GOOD VISUALIZATIONS

Principles for Graphical Integrity, from Tufte (abbreviated list)

Representation of numbers as physically measured on the graphic (amount of ink) should be directly proportional to the actual numerical quantity.

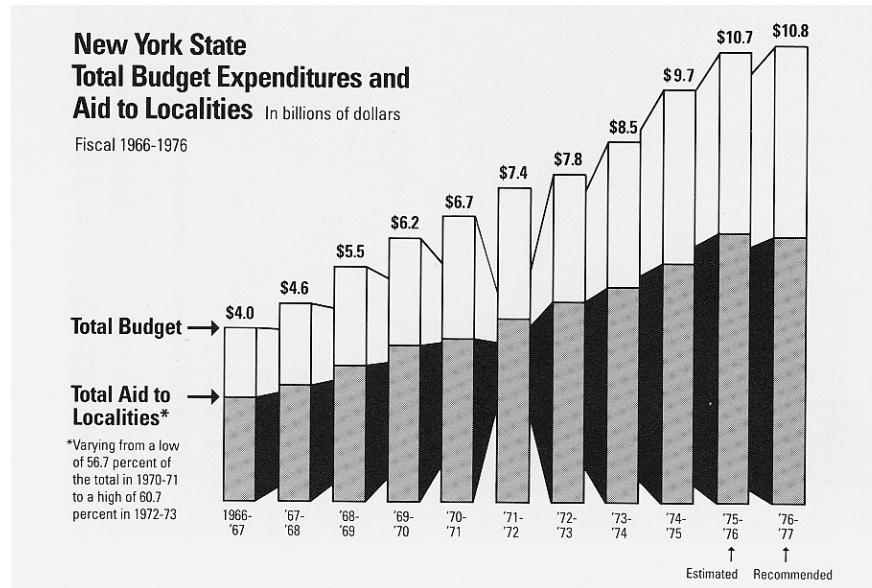
Clear, detailed, and thorough labelling should be used to defeat distortion and ambiguity. Include explanations and labels of important events.

For time series about money, always normalize for inflation

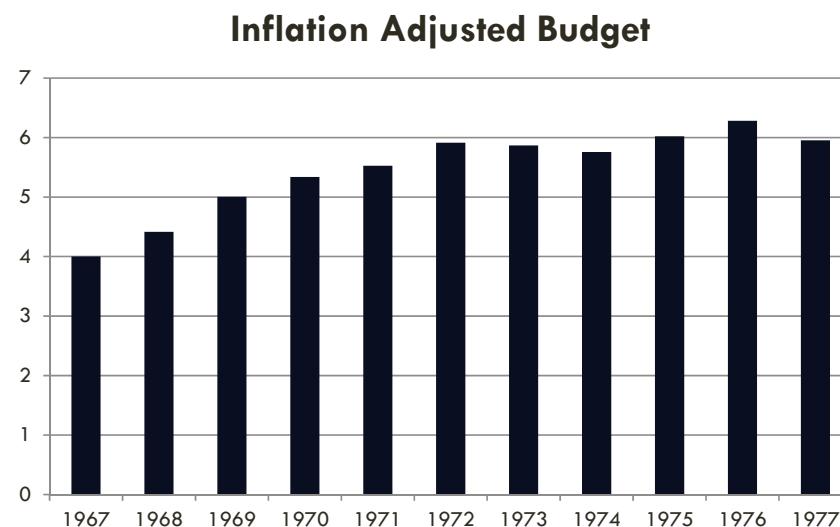
Do not show data out of context

Also, will add – for data where population is a factor, normalize with ‘per capita’

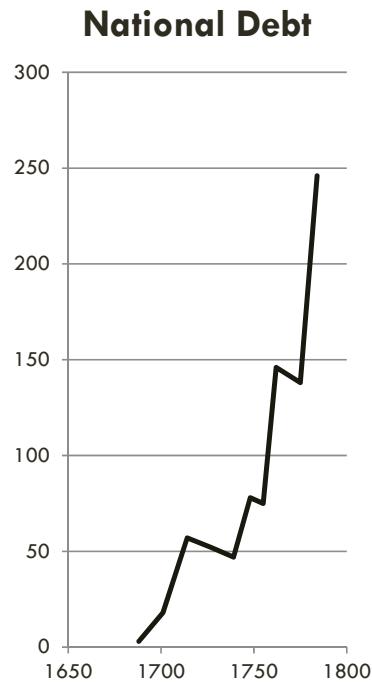
# CONSIDER THIS EXAMPLE



# REMOVE JUNK AND ADJUST FOR INFLATION



# ASPECT RATIO AND EXAGGERATION

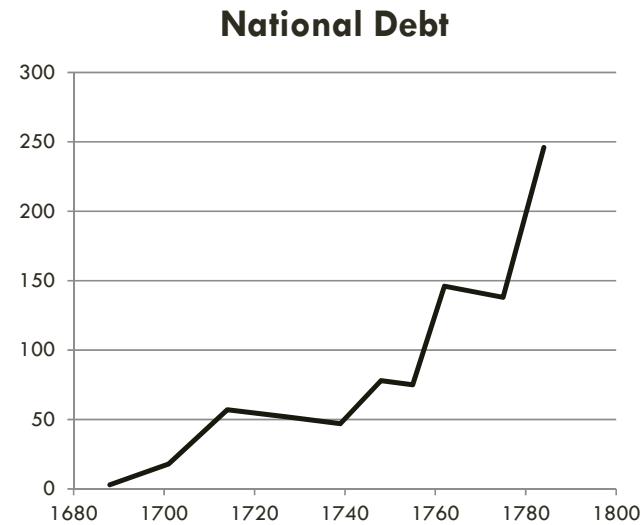


Remember: the aspect ratio of the graph is under our control!

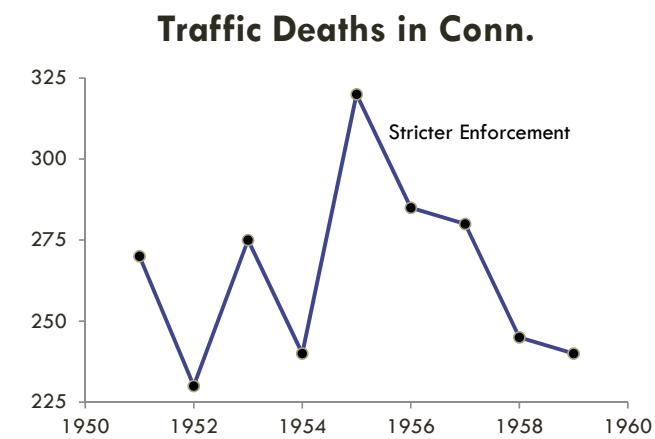
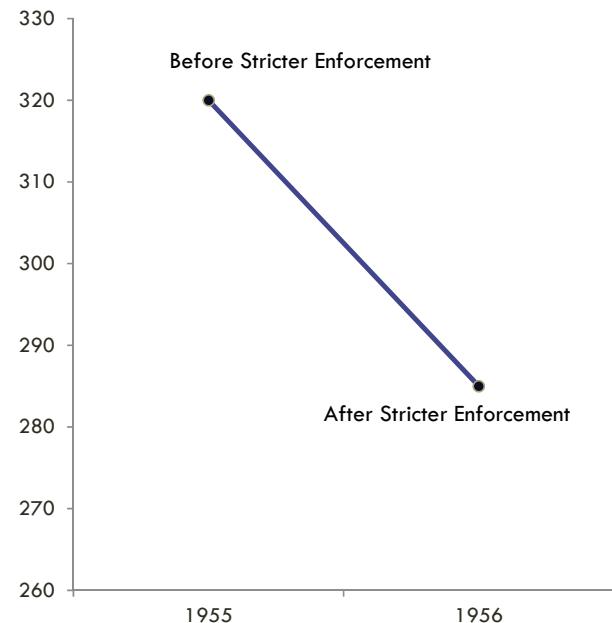
# THE BANKING TO 45° “RULE”

Aspect ratio set for median 45° slope

Often yields better slope comparisons

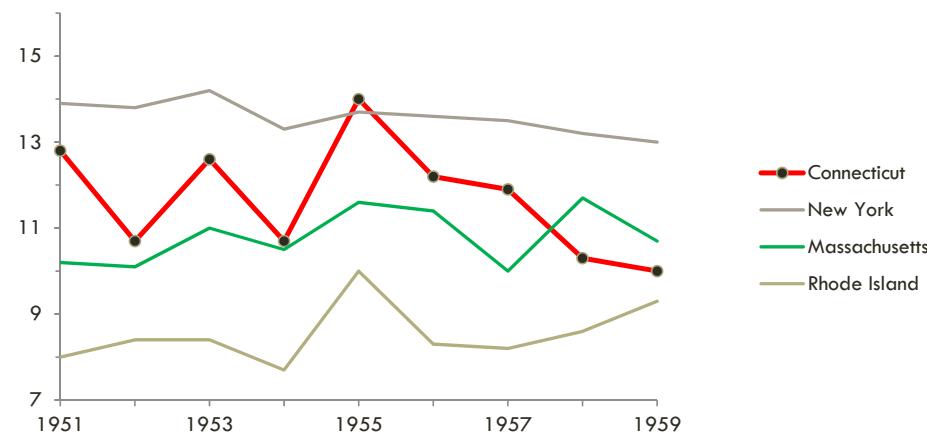


# DATA DISTORTION

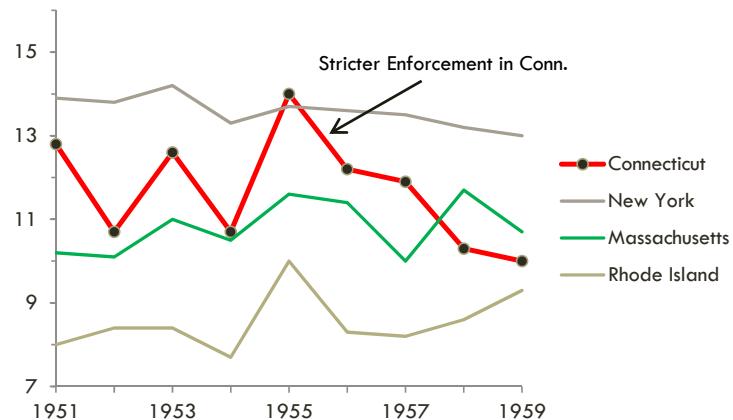
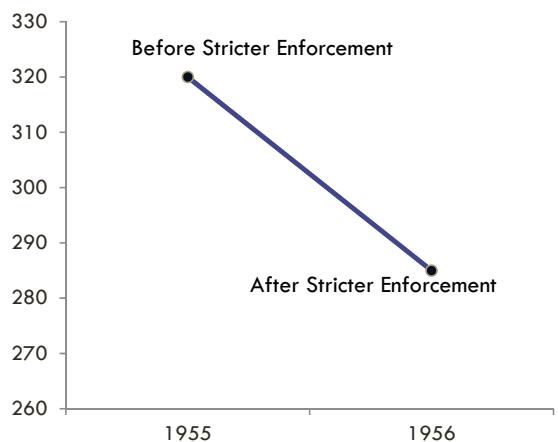


# INCLUDE MORE DATA

What about neighboring states?

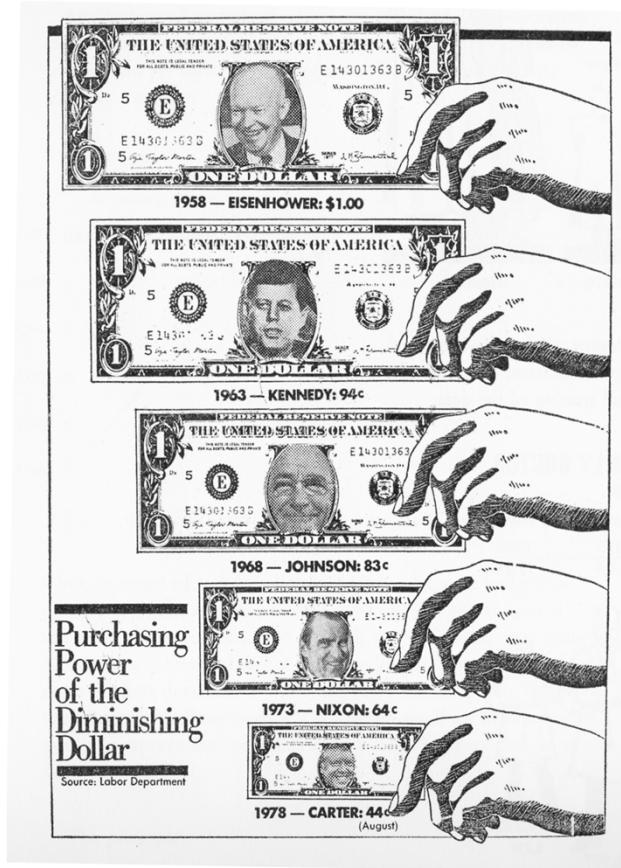


# COMPARE



# EXAMPLE

Purchasing power of the dollar displayed by size



Example from Edward Tufte

# PERCEPTION?

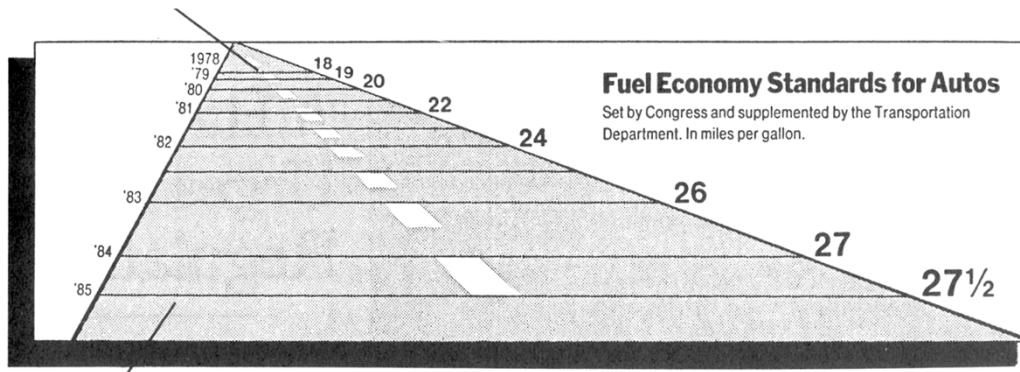
Human perception muddies the water here a bit

Two variables (width, height) represent the same thing

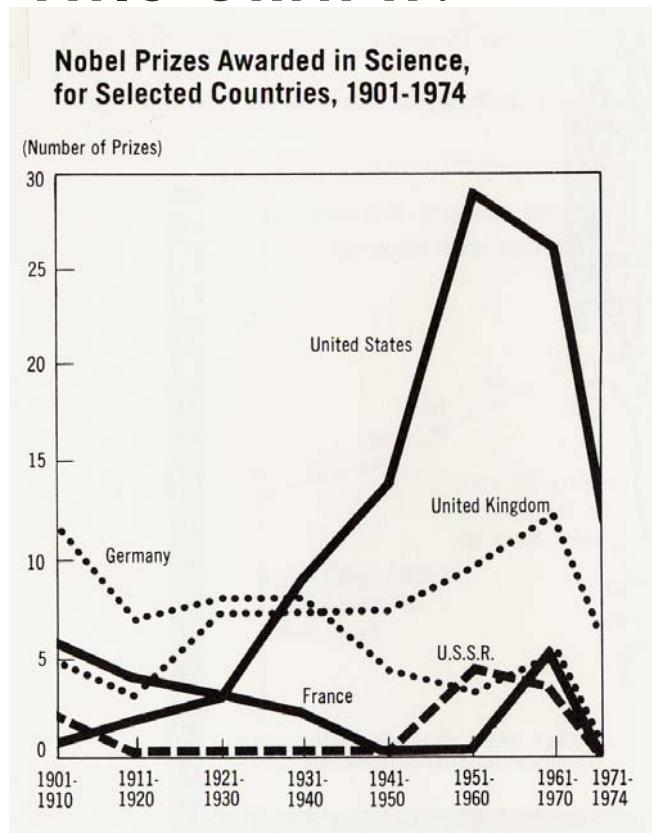
Brain sees area, but the area differences are way off (2x for 1978)



# ANALYZE THIS GRAPH

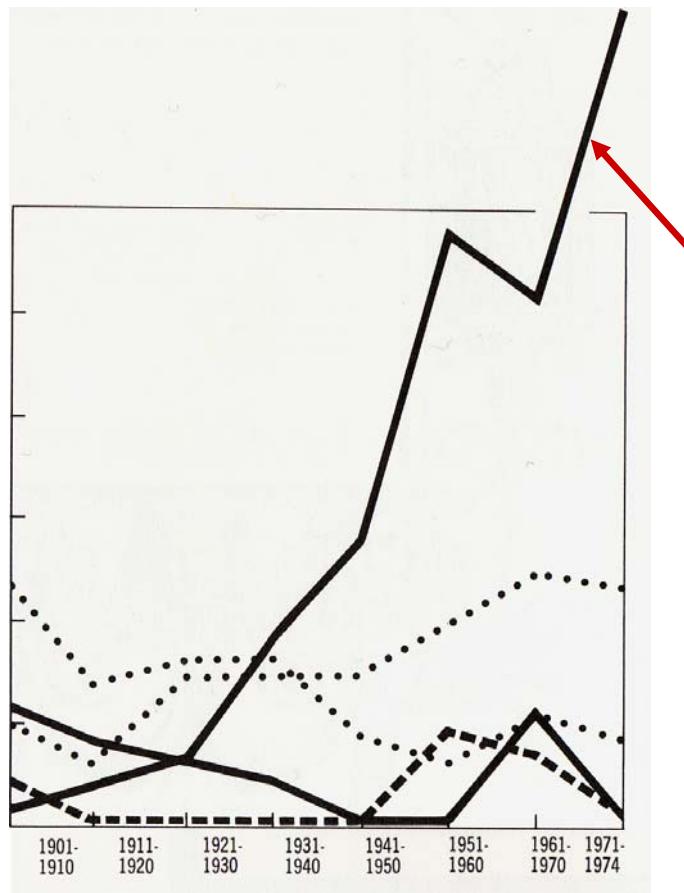


# PROBLEM WITH THIS GRAPH?



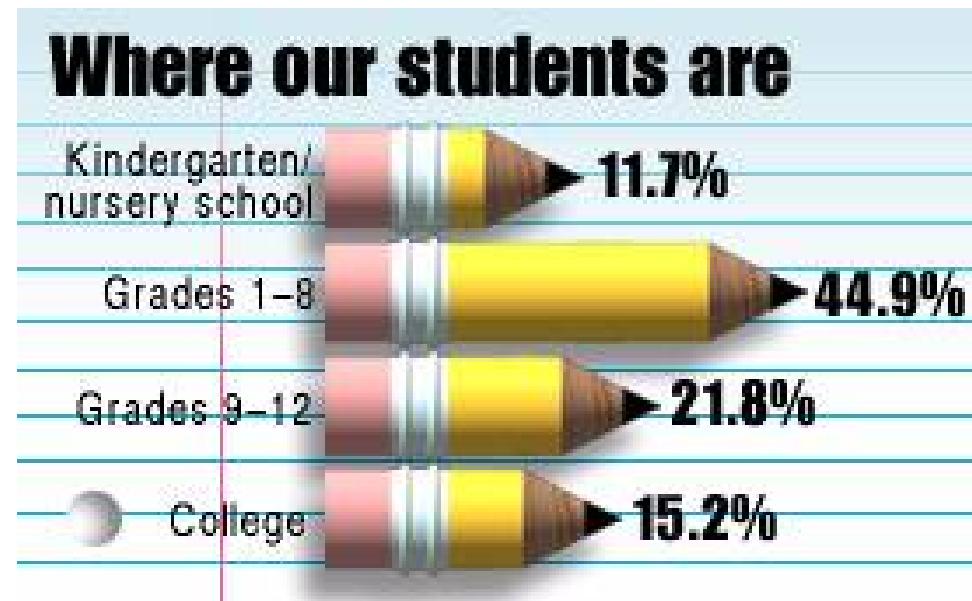
COPYRIGHT © 2014 BY DR. JOHN McDONALD

# DATA DISTORTION!



COPYRIGHT © 2014 BY DR. JOHN McDONALD

# ANALYZE THIS GRAPH



# CHART JUNK VS NON-DATA

Tufte: "A large share of ink on a graphic should present data-information, **the ink changing as the data change**. Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented."

Tufte: "The **interior decoration of graphics** generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies — to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, it is all non-data-ink or redundant data-ink, and it is often **chartjunk**."

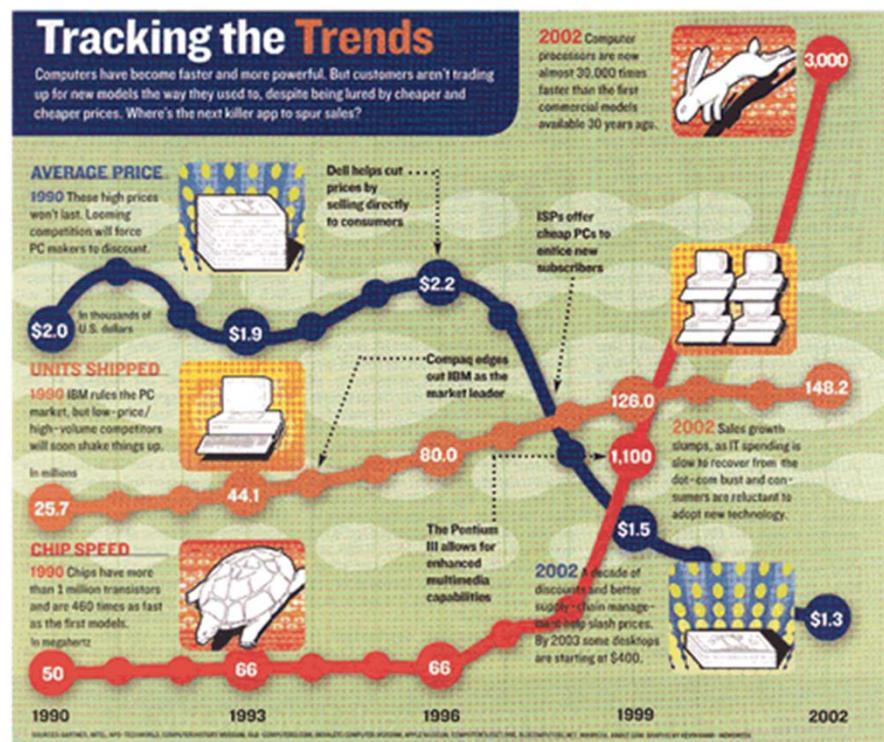
# DATA VS. NON-DATA INK

**Data-Ink/Graphics** = Ink that is used to display information about the data

**Non-Data Ink** = Ink that is used for other purposes

- Major source of clutter
- Draws attention away from data
- Uses space that can be dedicated to more information about the data

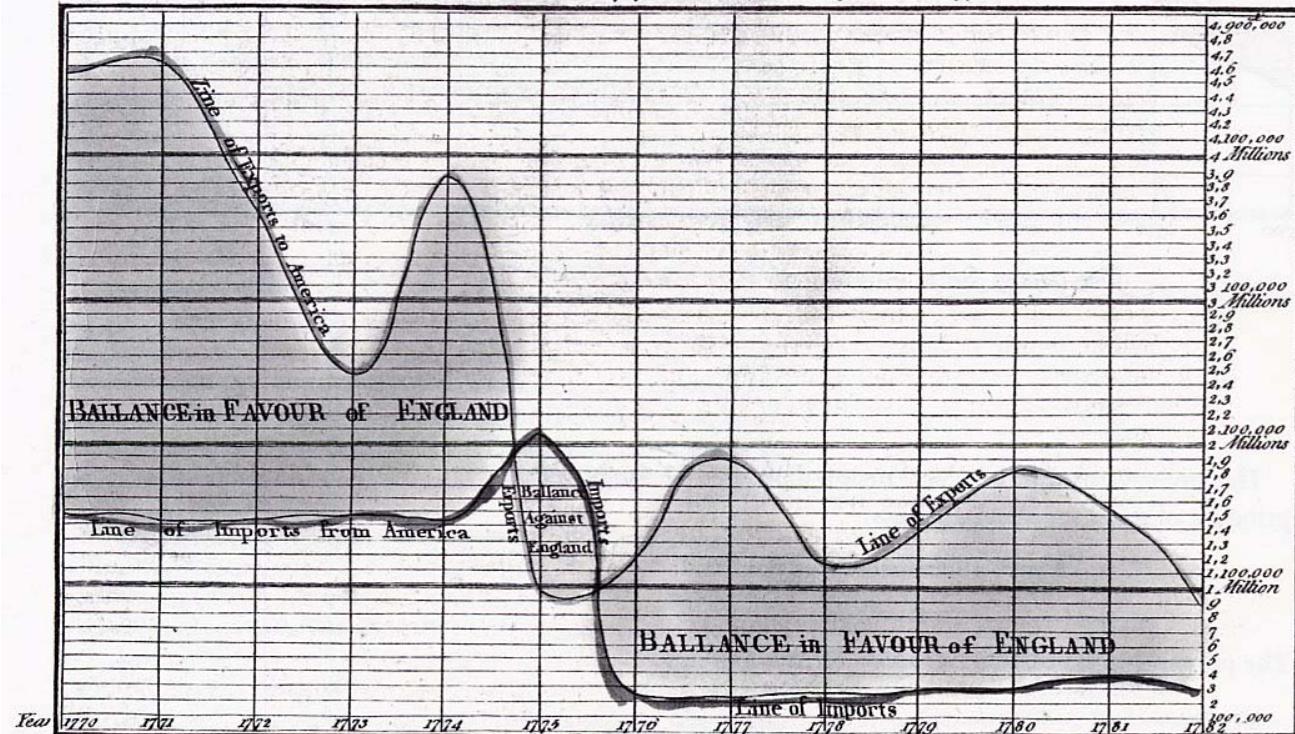
# CHART JUNK



Stephen Few, Tracking the Trends, Common Mistakes in Data Presentation, *intelligent enterprise*, August 7, 2004

# EXAMPLE

*CHART of IMPORTS and EXPORTS of ENGLAND to and from all NORTHAMERICA  
From the Year 1770 to 1782 by W. Playfair*



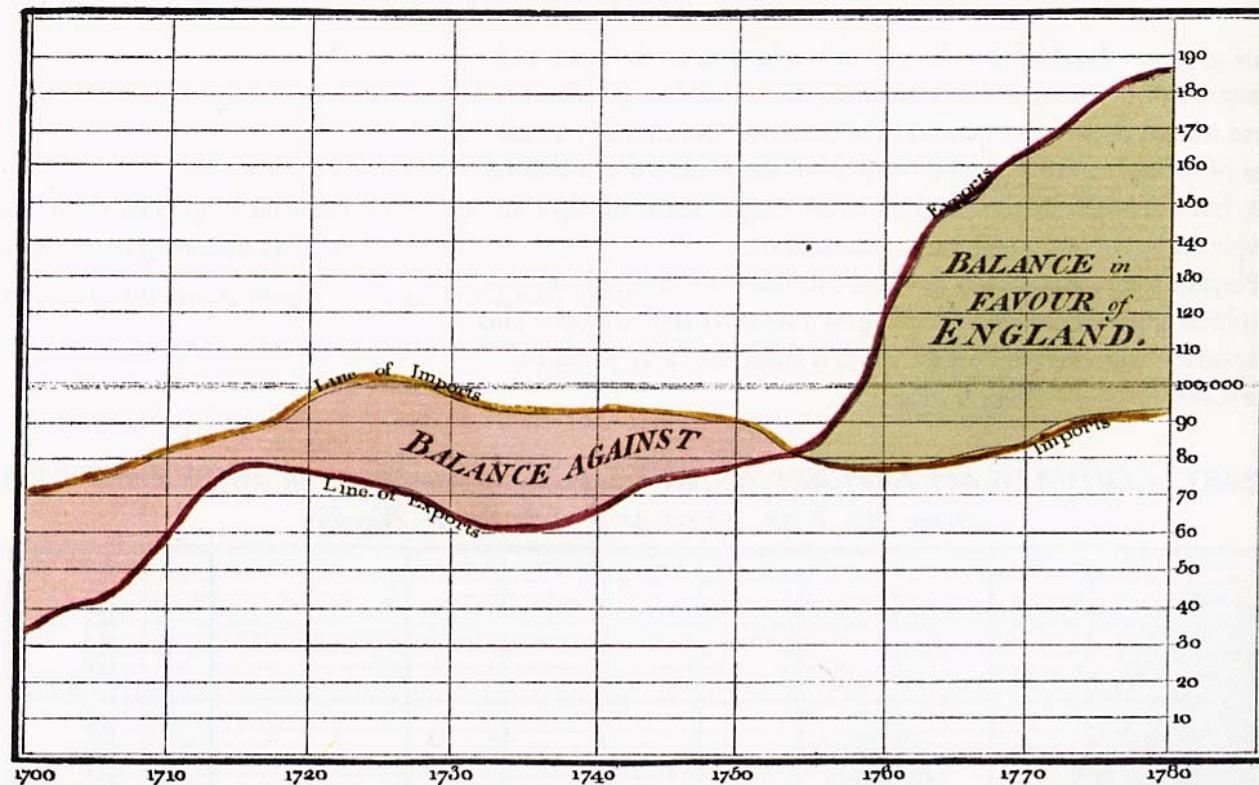
The Bottom Line is divided into Years the right-hand Line into HUNDRED THOUSAND POUNDS

J. Finnie Sculp.

Published as the Act directs 20<sup>th</sup> Aug<sup>r</sup>. 1785.

# REDESIGN

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.

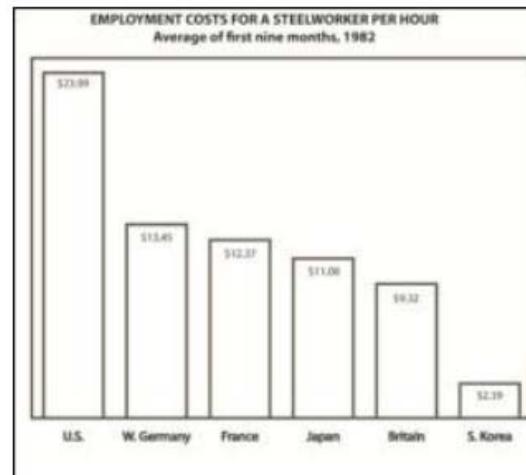
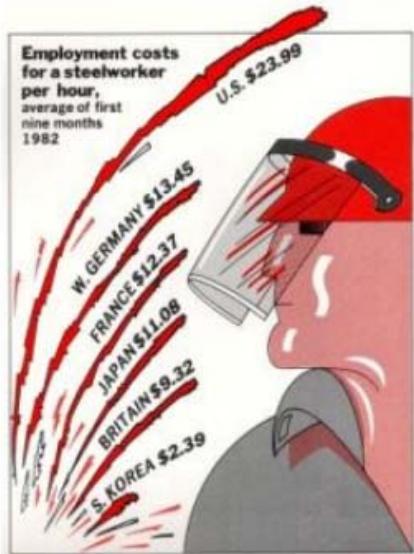


The Bottom line is divided into Years, the Right hand line into £10,000 each.

Published at the Act directed, 10<sup>th</sup> May 1786, by W<sup>m</sup> Playfair

No. 66, Sculpt. 352, Strand, London.

# CHART JUNK



# CHART JUNK

Is chart-junk always a bad thing?

*Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts*  
by Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, Christopher Brooks

*Benefitting InfoVis with Visual Difficulties* by Hullman, Adar and Shah.

# COLOR

# COLOR IS NOT GOOD AT

Fine distinctions

Exact measurements

Proportional differences

Yet, color is one of the most powerful visualization tools

# ROLES OF COLOR

Label

To draw a contrast

Dichotomy

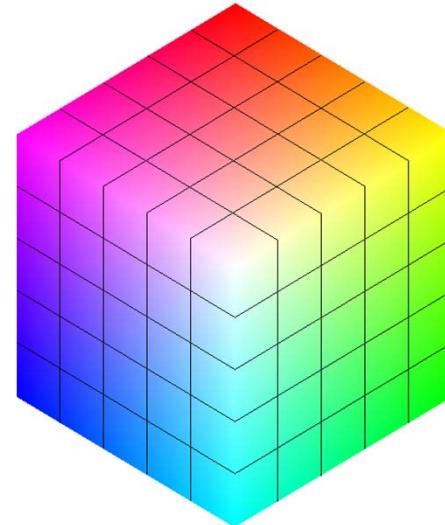
Stand-outs

Group values

# WHICH COLORS DO I USE?

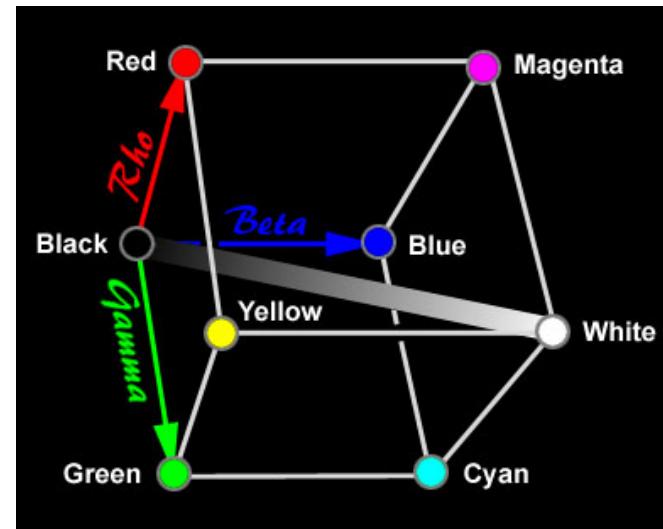
Computers use RGB color

- Almost useless for Visualization



# WHICH COLORS DO I USE?

How do you go smoothly from red to green to blue??



# COLOR TERMINOLOGIES

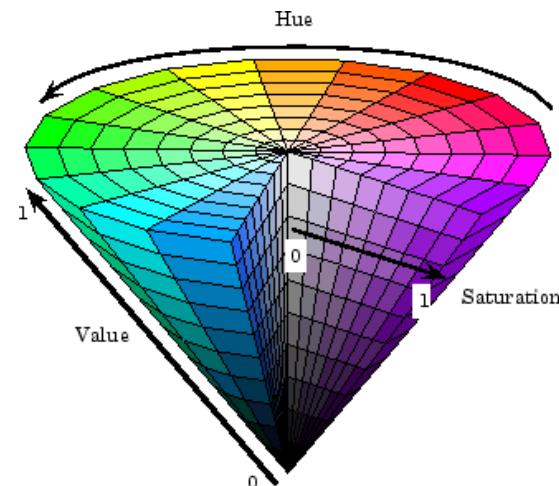
- **Hue:** distinguishes among colors such as red, green, purple, and yellow
- **Saturation:** refers to how pure the color is, how much white/gray is mixed with it
  - red saturated; pink unsaturated  
  - royal blue saturated; sky blue unsaturated  
  - Our eyes can distinguish about 128 fully saturated hues
- **Lightness:** perceived achromatic intensity of reflecting object
- **Brightness:** perceived intensity of a self-luminous object, such as a light bulb, the sun, or a CRT
- Our eyes can distinguish approximately 7 million colors when the samples are placed side-by-side

# WHICH COLORS DO I USE?

HSV is more natural

- Hue, Saturation, Value

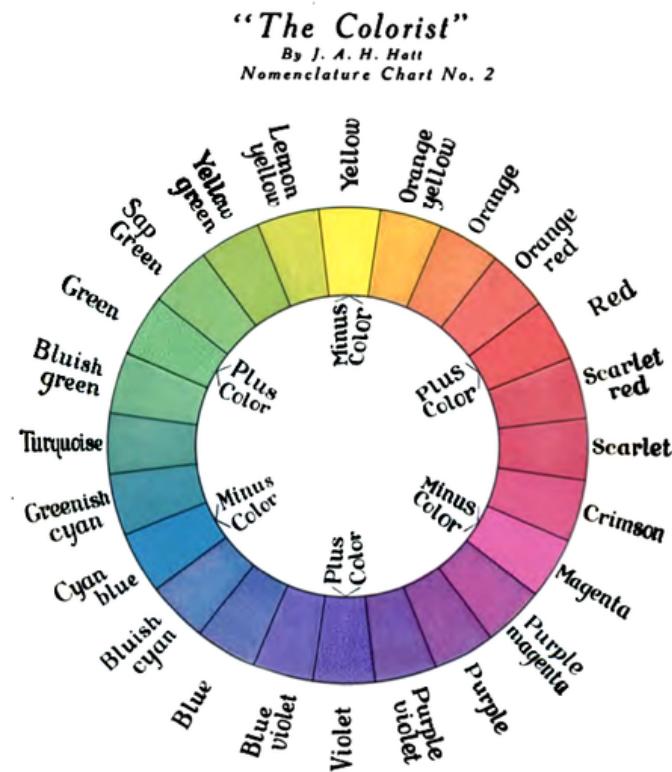
Can easily blend along all three axes!



# BUILDING COLOR SCHEMES

The color wheel = “hue”

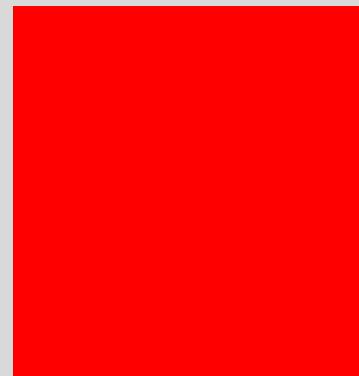
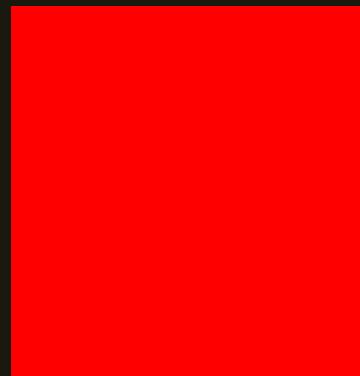
- Analogous colors are adjacent
- Contrastive colors are opposite



Showing proposed names for hues 15° apart.  
Colors opposite each other are complimentary.

D. VAN NOSTRAND CO.  
Publishers, N. Y.

# SUBTLETIES IN COLOR PERCEPTION



# SUBTLETIES IN COLOR MAPPING

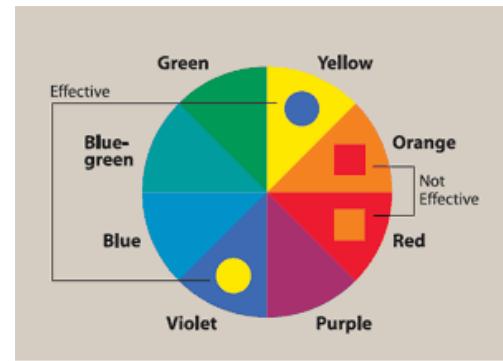
## Contrast is Key

- Want to make data stand out

Can you read me?

# CREATING COLOR CONTRAST

Use colors from opposite sides of the HSV cone



©2005 Lighthouse International. All rights reserved.

# QUESTION

Why don't we judge **precise** value well from color?

# WHERE IS THE MIDDLE VALUE?



# WHERE IS THE MIDDLE VALUE?

A    B    C    D    E    F    G    H    I    J

# WHERE IS THE MIDDLE VALUE?



WHERE DOES EACH FADE TO BLACK



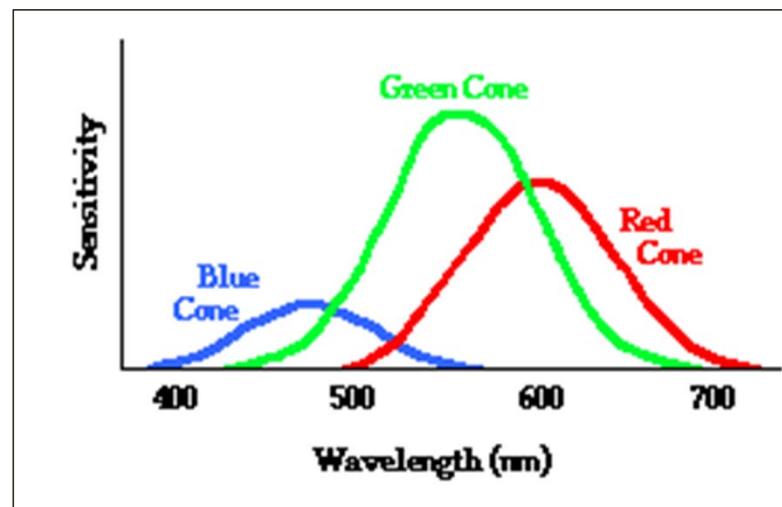
# ARE ALL COLOR CHANGES EQUAL?

Which “band” of color looks the widest?



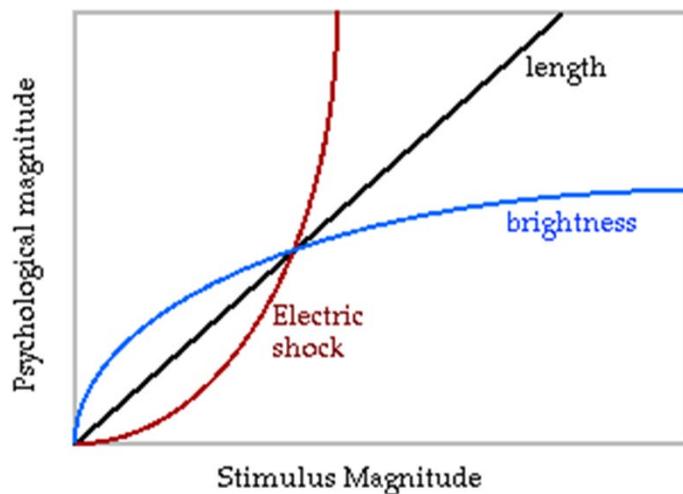
# COLOR PERCEPTION

Cones are most sensitive to green



# NON-UNIFORM PERCEPTION OF INTENSITY

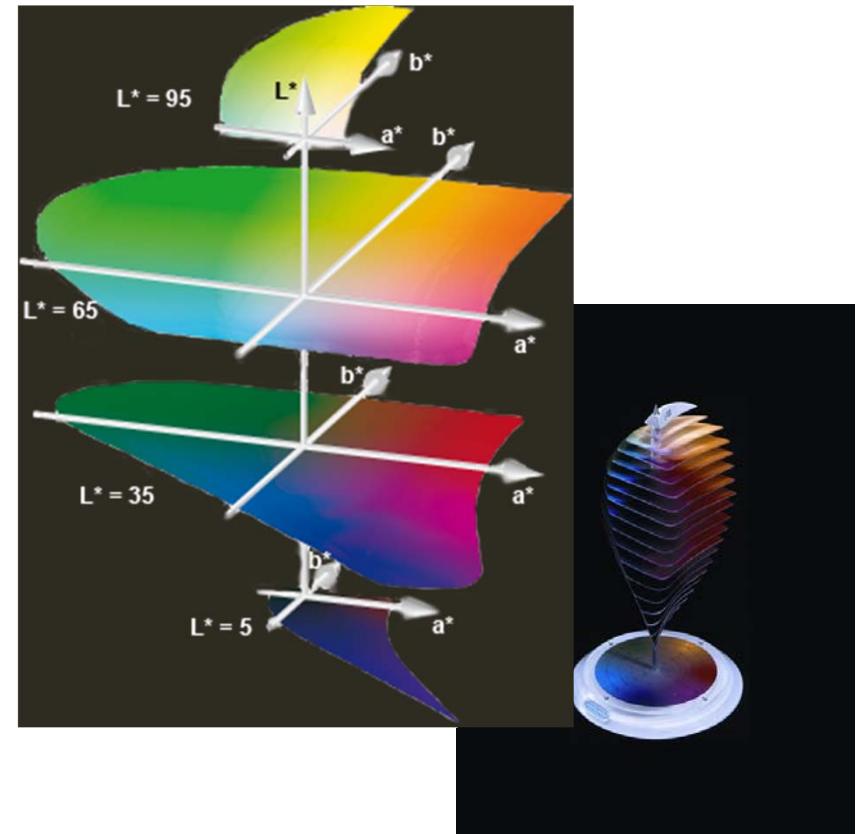
Steven's Power Law – the relationship between the magnitude of the stimulus and the perceived intensity follows the general form of ( $I$  = magnitude,  $k$  = constant,  $\psi$  = psychophysical function)



$$\psi(I) = kI^a,$$

# CIE LAB – A PERCEPTION BASED COLOR SPACE

- CIE Lab was introduced in 1976
  - popular for use in measuring reflective and transmissive objects
  - Based on the three color receptors of the human eye(red, green and blue)
- Three components:
  - $L^*$  is luminosity
  - $a^*$  is red/green axis
  - $b^*$  is yellow/blue axis
- Mathematically described space and a perceptually uniform color space



# CHOOSING A COLOR SCHEME

Use the power of suggestion (e.g. food data: bananas can be in yellow)

Be careful of cultural differences

Be aware of color blindness

Don't use too many colors!

Use the right type of mapping:

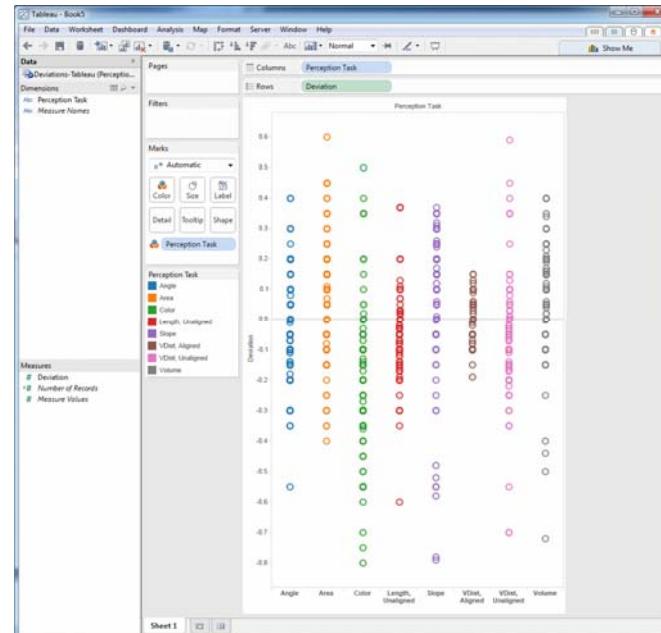
- Nominal/categorical/discrete – one color per discrete thing, use to draw contrast
- Sequential – values ranging from low to high map to a color range; low hue shift
- Divergent – values range from one extreme to neutral to another extreme

# DISCRETE COLOR SCHEMES

## Qualitative Color Palettes

### Goals

- Contrast without clutter
- Colors can't be close!
- Many are unmanageable



# HOW MANY COLORS?

Tufte (a bit optimistic ☺)

“For encoding abstract information, however, more than 20 or 30 colors frequently produce not diminishing but **negative** returns.”

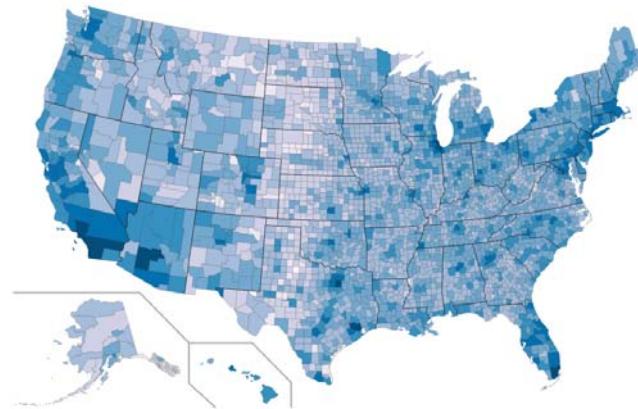
- Contrasting colors ... not continuous shading

# SHADES OF DATA

## Continuous/Sequential Color Palettes

- Show a single value
- Highs and lows stand out

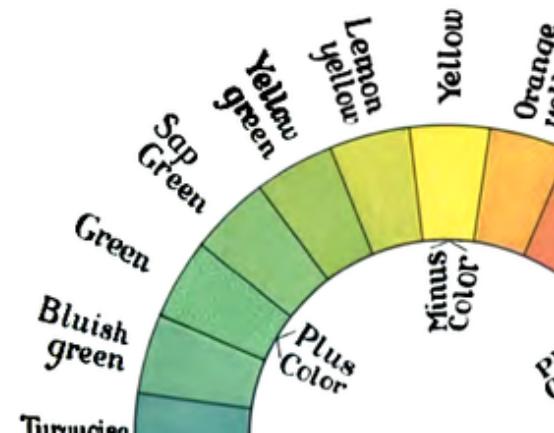
Shades of a single color



# SHADES OF DATA

Blend between adjacent colors  
on the wheel

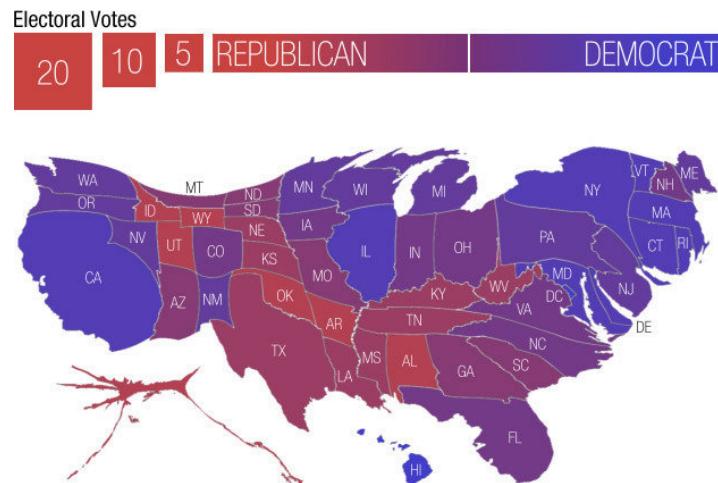
- A bit more contrast



# EMPHASIZING BOTH EXTREMES

Divergent Color Palettes

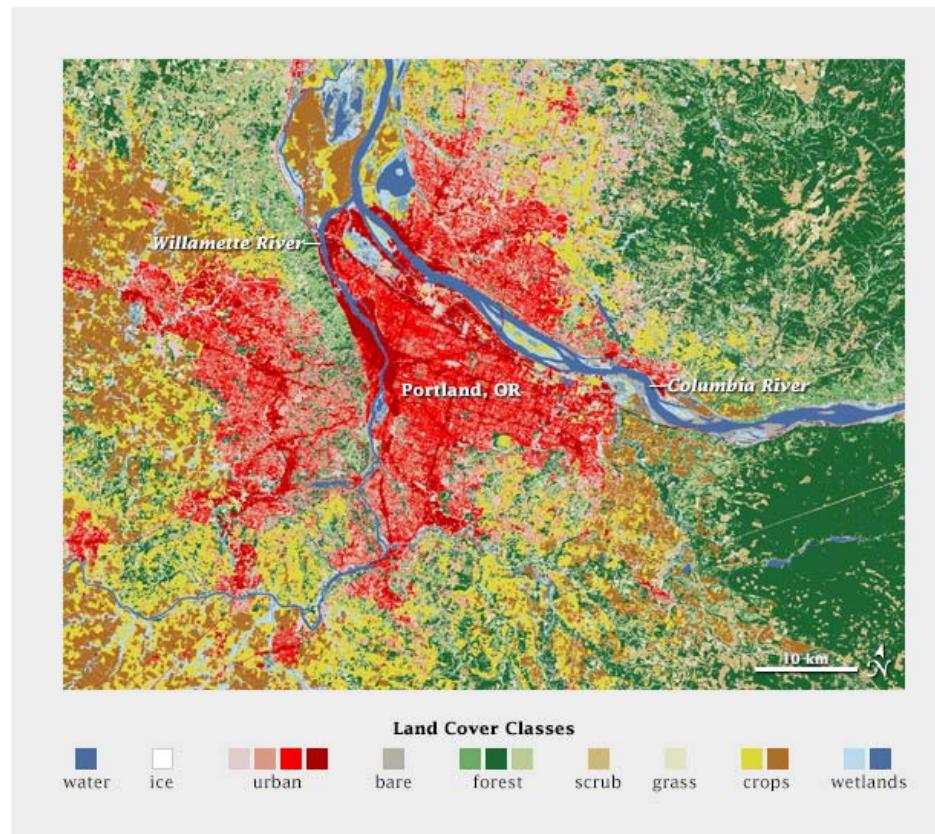
Blend between contrasting colors



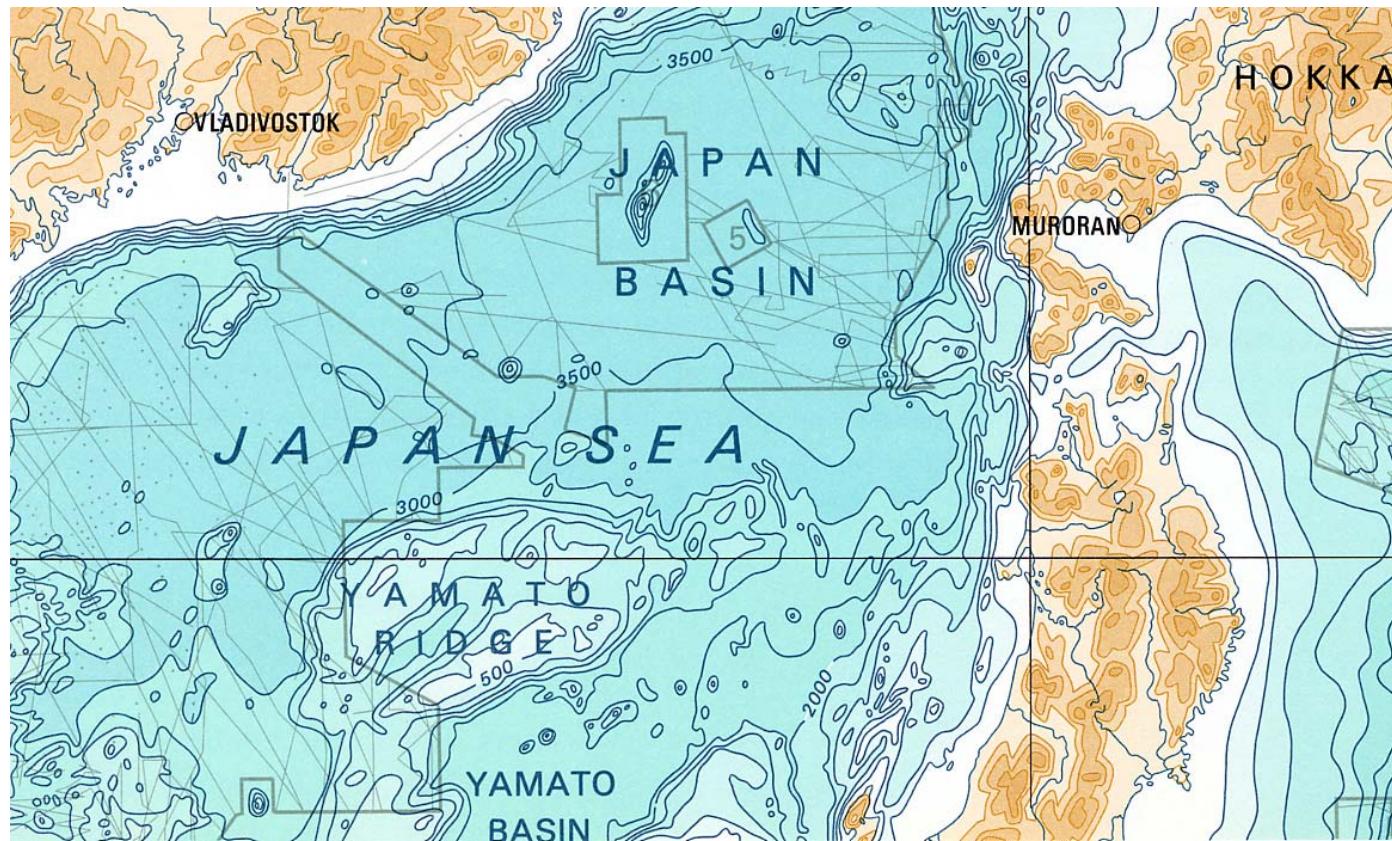
# CHOOSING A COLOR SCHEME



# CHOOSING A COLOR SCHEME



# HOW IS COLOR BEING USED?



# DESIGN PRINCIPLES

Use adjacent colors in the cone to suggest subtle differences

- Gradually change color saturation or hue

Pastels can be effective to show continuity

- They blend well with little visible edge

# CHOOSING A COLOR SCHEME

Use complimentary\* colors to draw contrasts

- Sharply different colors draw out different extremes in value

Use when both extremes are equally important in the data

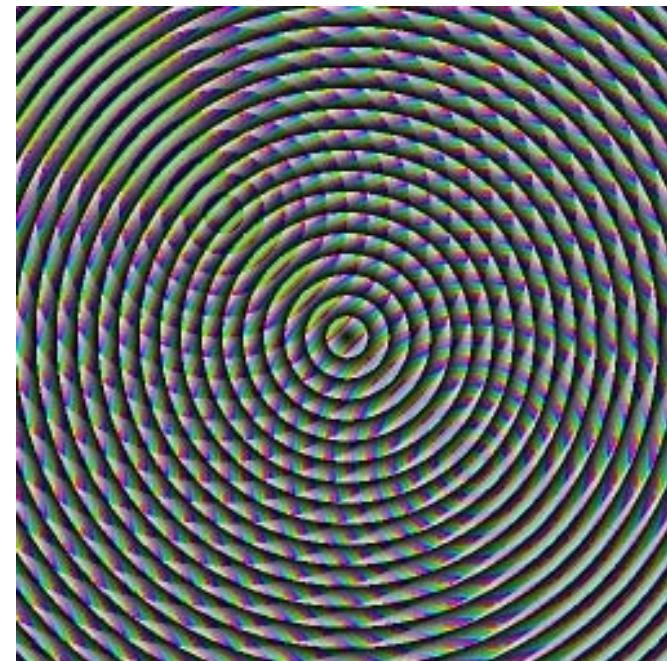
\* opposite side of color wheel

# CHOOSING A COLOR SCHEME

Beware of misleading color illusions!

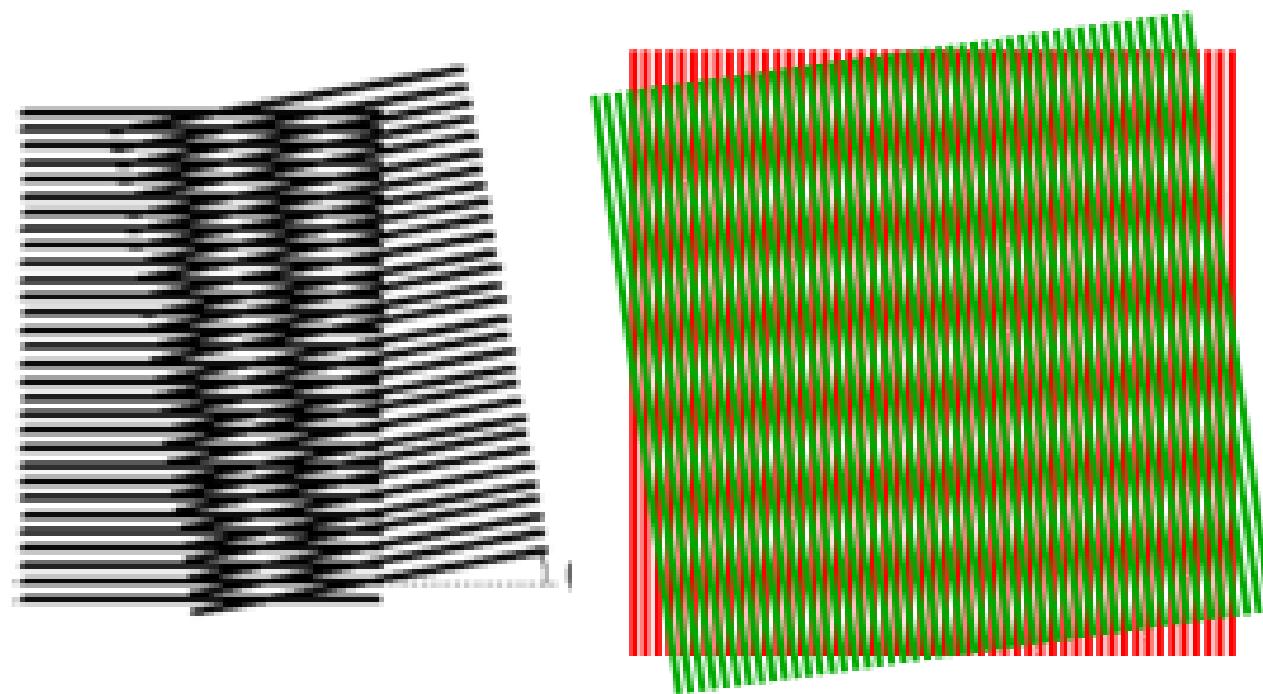
- **Moire Patterns**

Test at different sizes

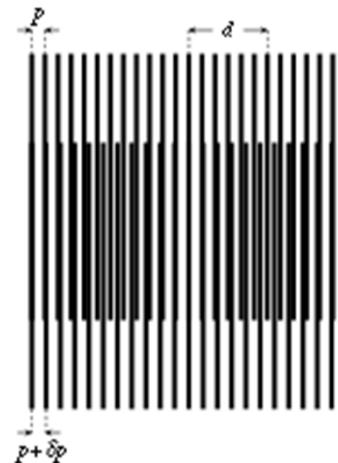
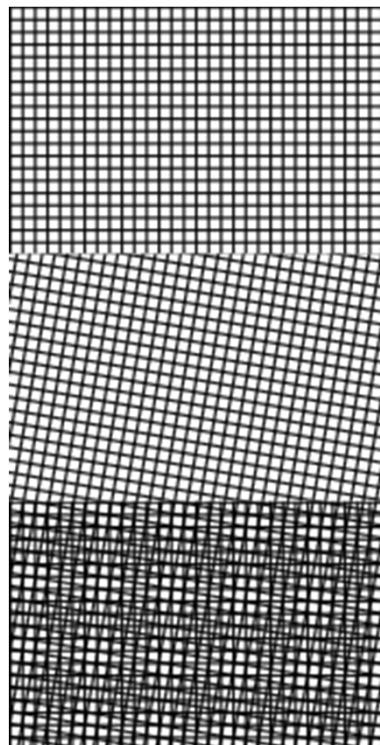


<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/moirepatterns/>

# MOIRE



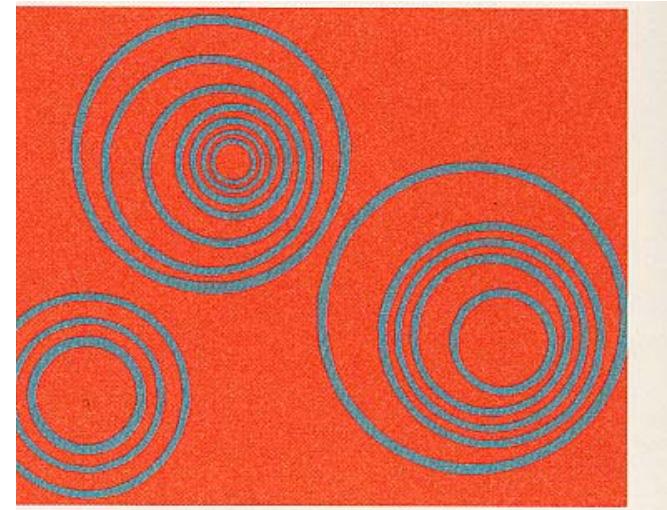
# MOIRE



# GUIDELINE

Avoid bright primary colors in large areas

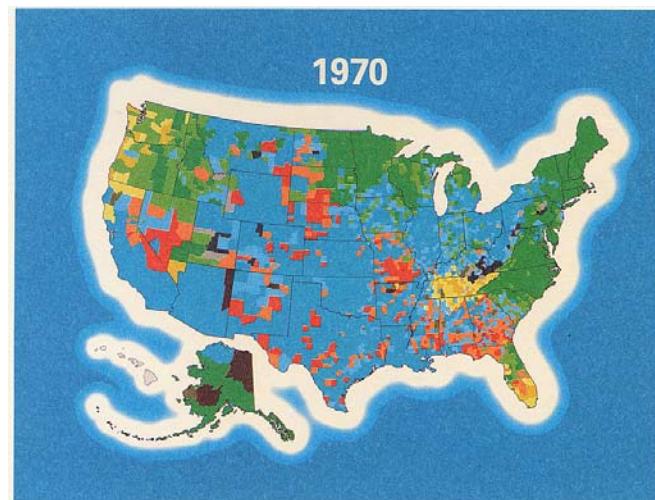
- Uncomfortable “ringing” effects



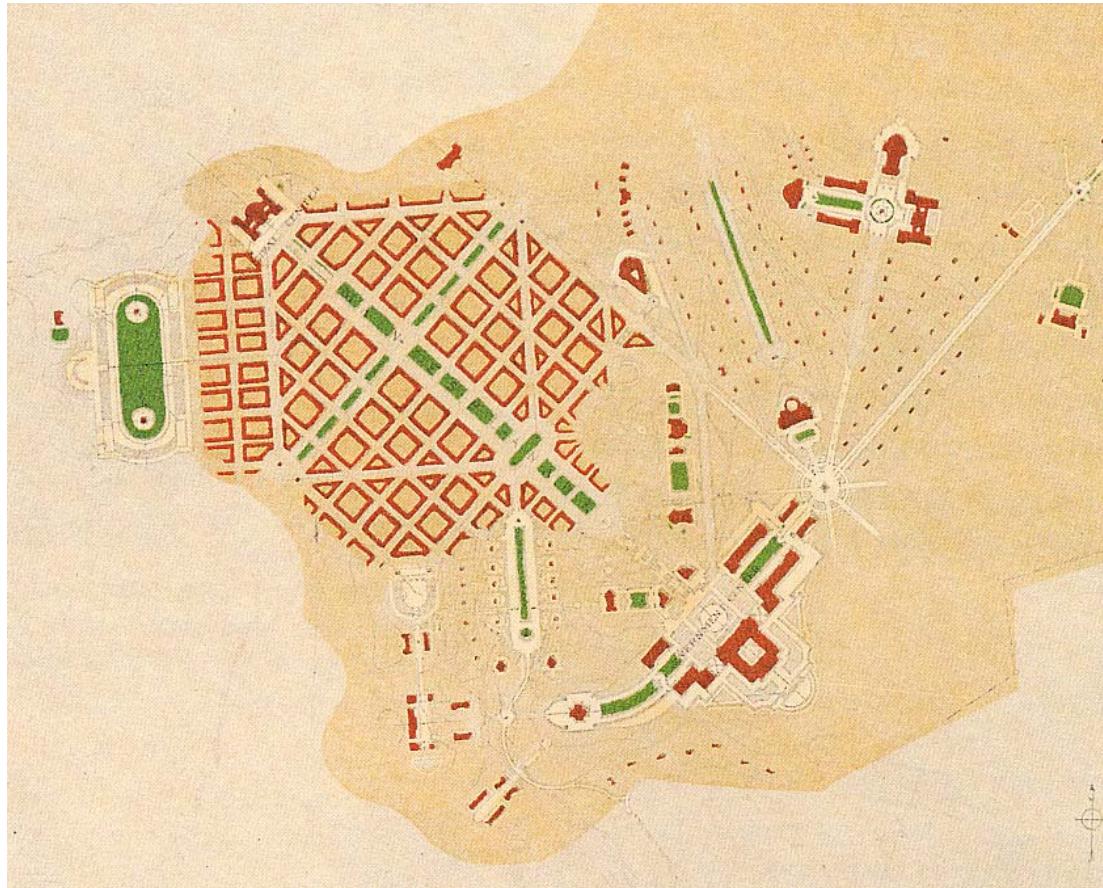
# GUIDELINE

Avoid light/bright colors mixed with white

- Unpleasant
- Especially if used for large areas



# GOOD EXAMPLE

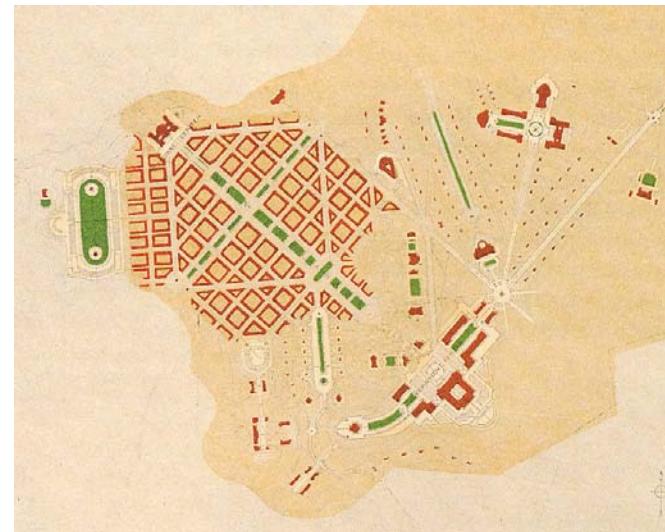


# GUIDELINE

Use bright/primary colors for small important features

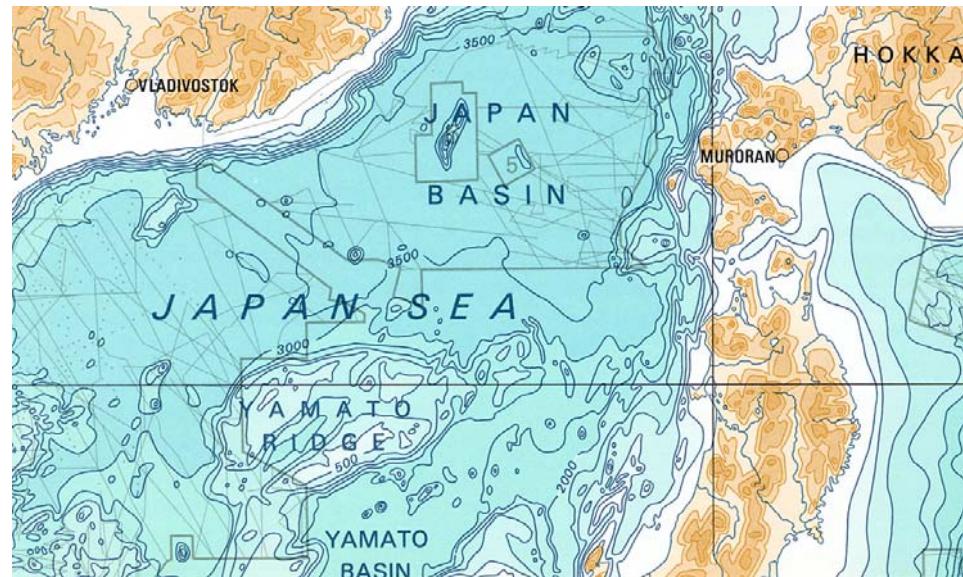
Large background

- Softer
- Less saturated
- “do their work most quietly!”

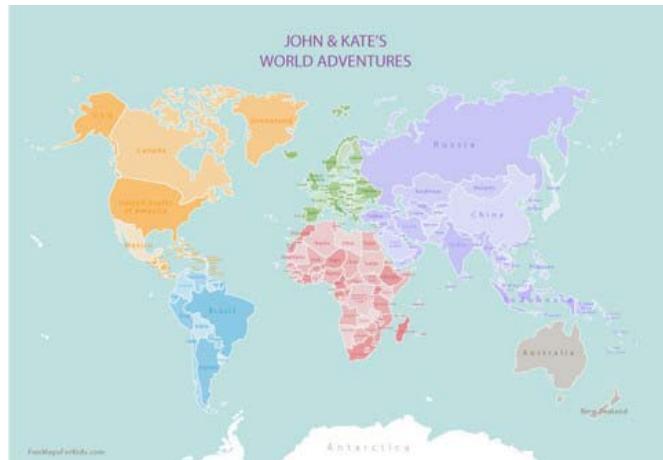


# GUIDELINE

Curves can often be used to augment subtle color changes!



# PRIMARIES VS. PASTELS

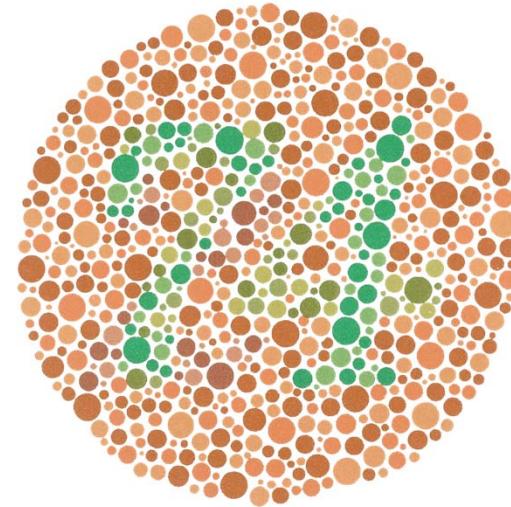


# ANOTHER CONCERN

## Color blindness

- Inability to distinguish between certain primary colors
- Red-green is most common

Can adversely affect visualization decoding

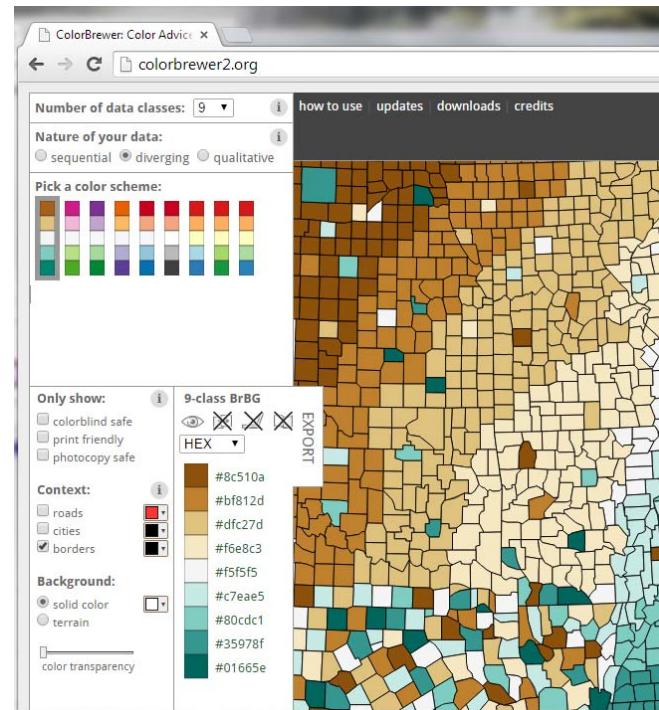


<http://ishiharatest.blogspot.com/2011/03/ishihara-color-blindness-test.html>

# THE COLOR BREWER

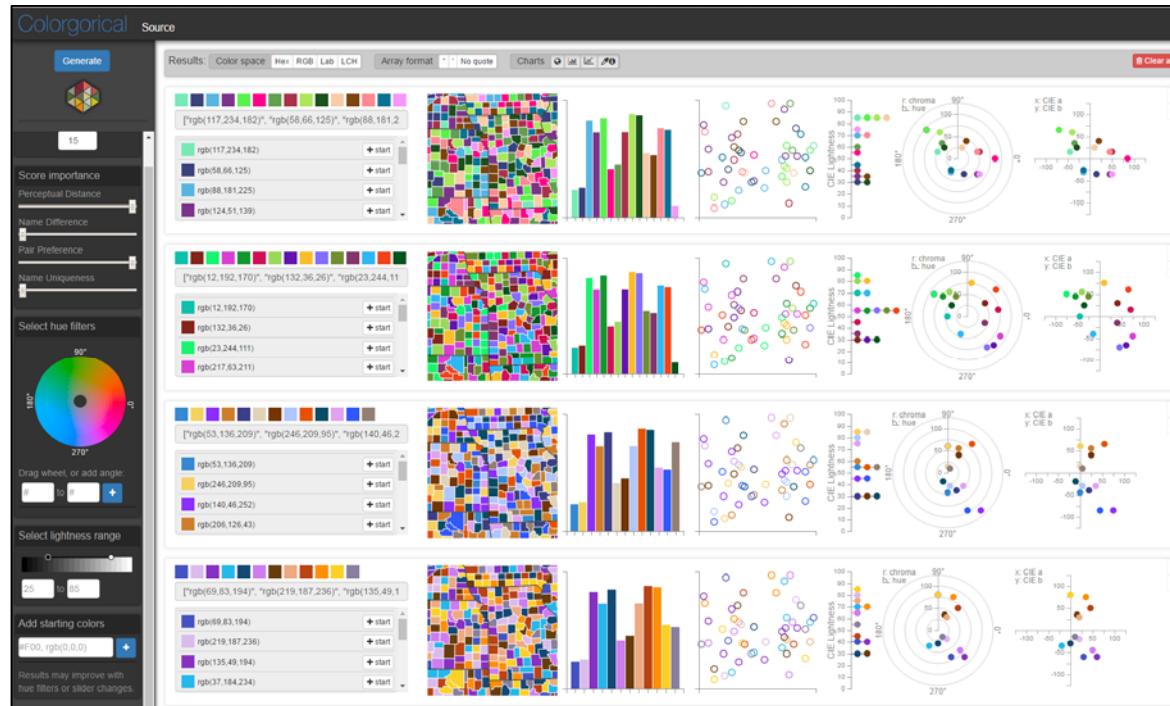
Allows you to explore color schemes

- Also built-in to D3



Colorbrewer.org

# COLORGORICAL



<http://vrl.cs.brown.edu/color>

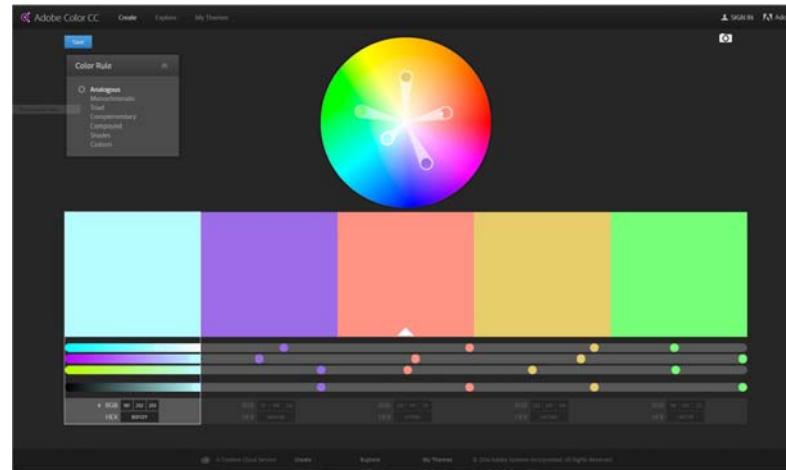
**Quite new – presented a few months ago!**

# ADOBETOOLS

Tools in their powerful professional design software

Free online tool (right)

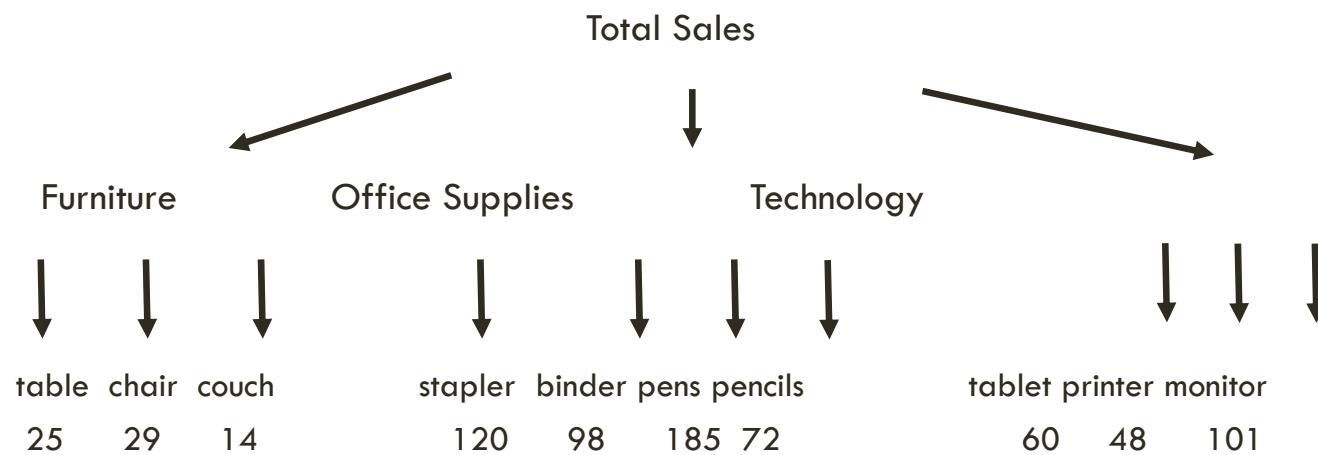
Built for designers



<https://color.adobe.com/create/color-wheel/>

# HIERARCHY

# HIERARCHICAL DATA



<http://www.theinformationlab.co.uk/2014/12/16/treemap-vs-bar-chart-end-treemap/>

# HIERARCHICAL DATA REPRESENTATION

Can be in specially formatted file

For us, usually a table:

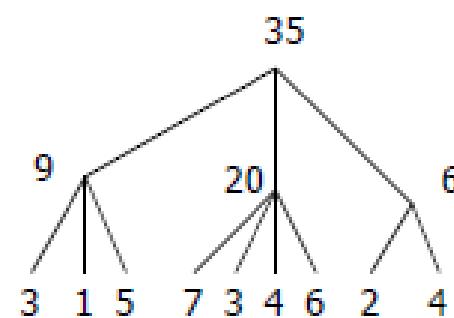
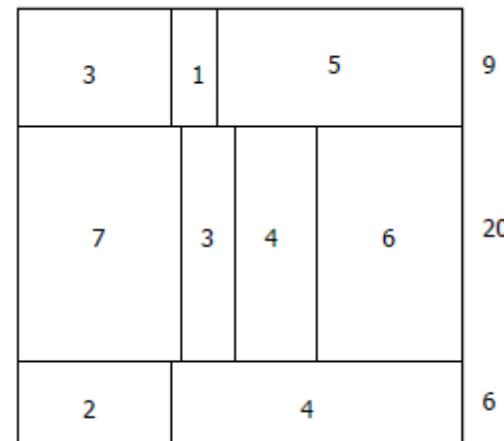
| Category | Sub-Cat. | Sub-Sub-Cat. | Value |
|----------|----------|--------------|-------|
|          |          |              |       |

e.g.

| Make   | Form  | Model   | Price |
|--------|-------|---------|-------|
| Honda  | SUV   | CR-V    | 25000 |
| Honda  | Sedan | Accord  | 28000 |
| Honda  | Sedan | Civic   | 20500 |
| Toyota | SUV   | RAV-4   | 24500 |
| Toyota | Sedan | Corolla | 20000 |
| Toyota | Sedan | Camry   | 28500 |

# TYPES OF HIERARCHY VISUALIZATIONS

- Space Filling
- Node-Link

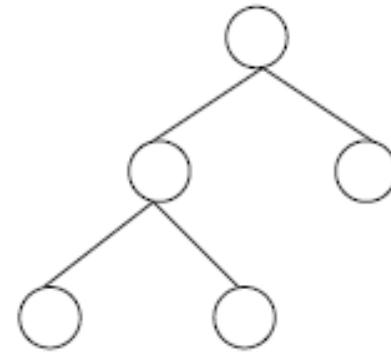
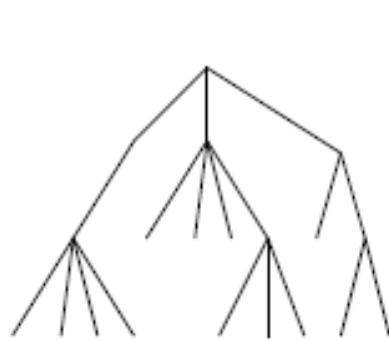


# NODE-LINK STRUCTURES

- Usually represented as a graph (or a tree)
  - Directed
  - Acyclic
- Ordering (top-bottom / left-right / inside-out) represent ancestral relationships

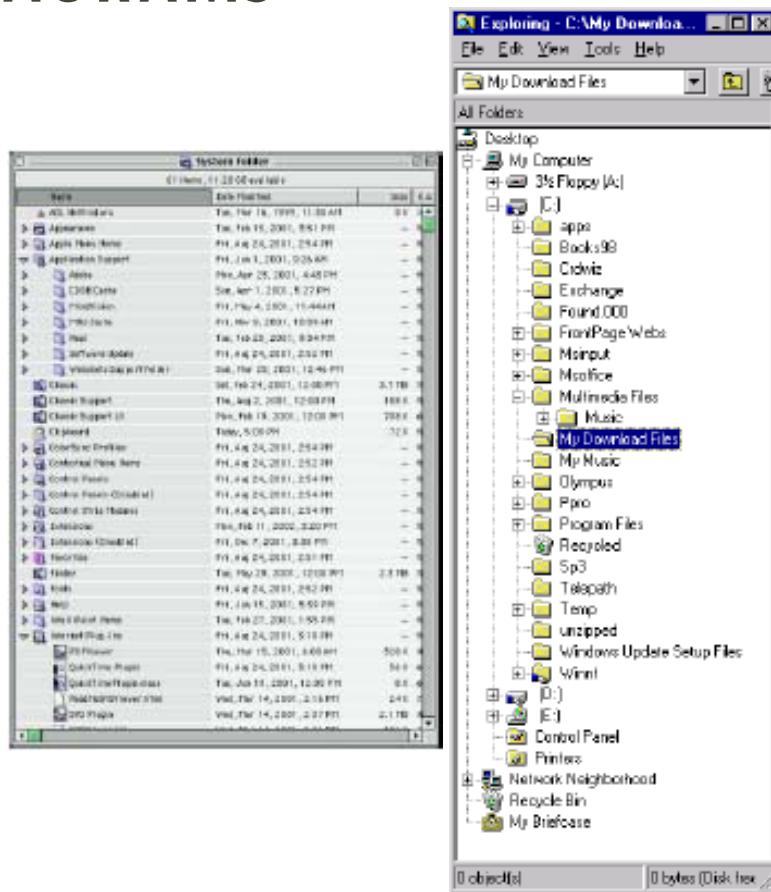
# COMMON NODE-LINK DIAGRAMS

- Top-Down ordering, root node on top, leaf nodes at the bottom.



# COMMON NODE-LINK DIAGRAMS

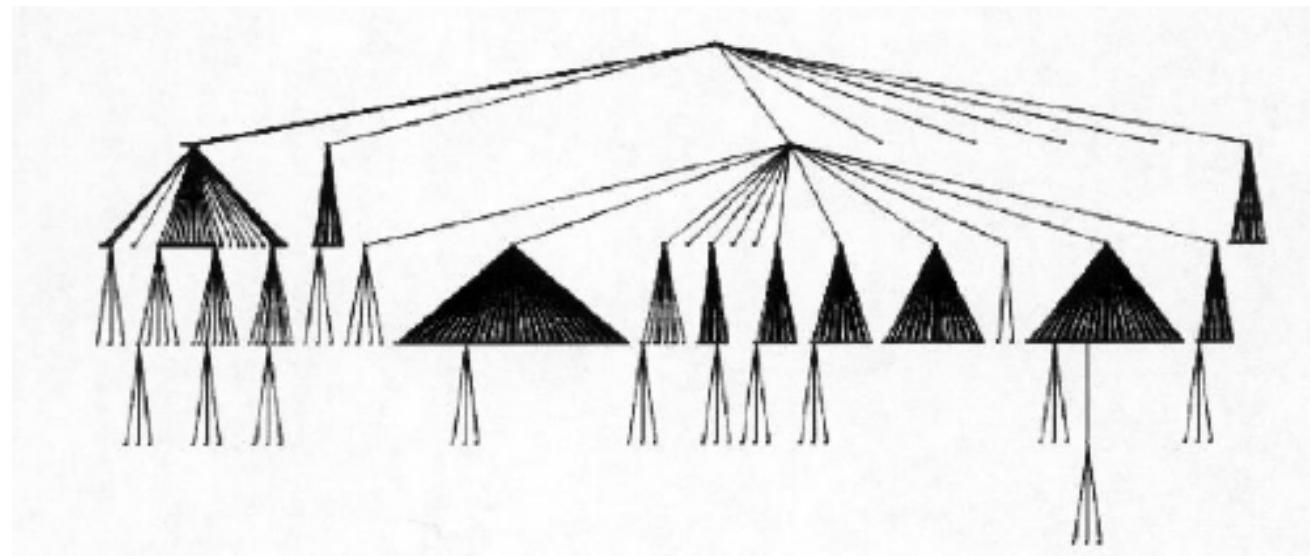
- Left-Right ordering
  - Indented – root node to the left, leaf nodes to the right.



# PROBLEMS WITH TREES...

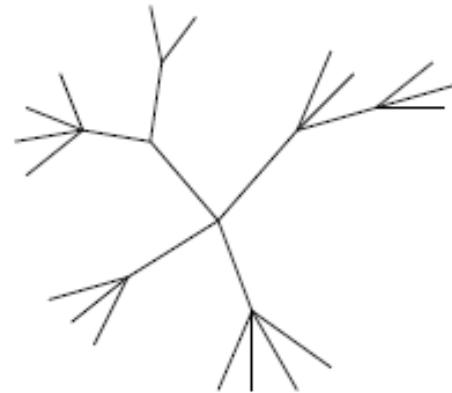
- In top-down designs, the horizontal screen real estate becomes scarce quickly.
  - Assume a balanced binary tree, at level  $n$ , there are  $2^n$  nodes.
- Trees are not guaranteed to be balanced. One branch can be much deeper than others, causing vertical real estate usages to become disproportional.

# HOW TO DRAW THIS TREE?

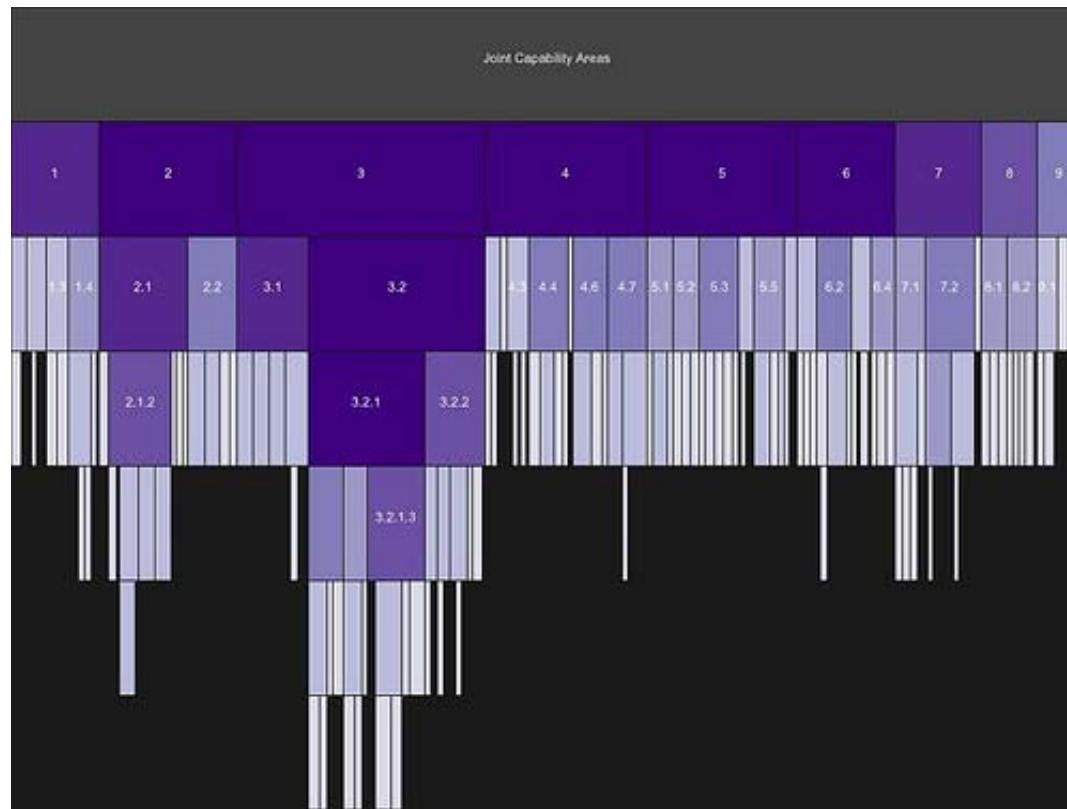


# COMMON NODE-LINK DIAGRAMS

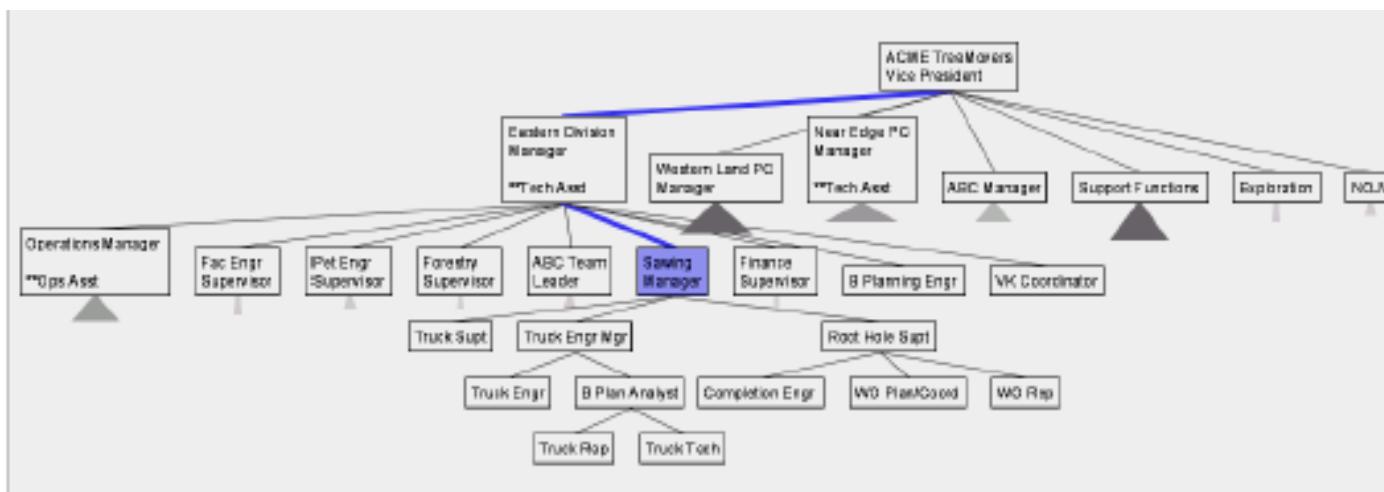
- Inside-Out Ordering.
- Root node in the center, leaf nodes to the outside.



# ICICLE PLOT



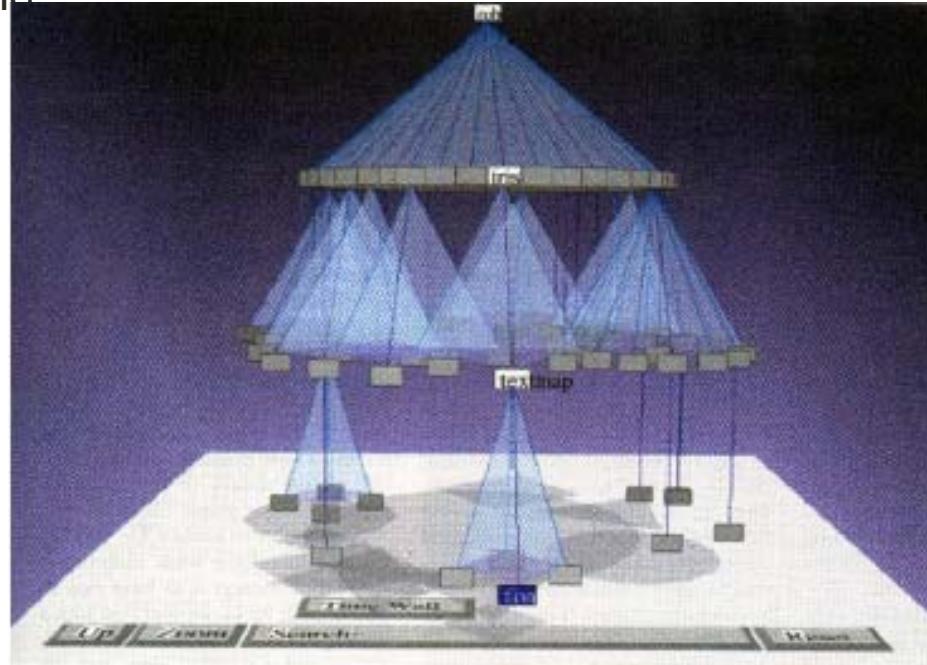
# SPACETREE



Grosjean, Plaisant, BedersonInfoVis „02

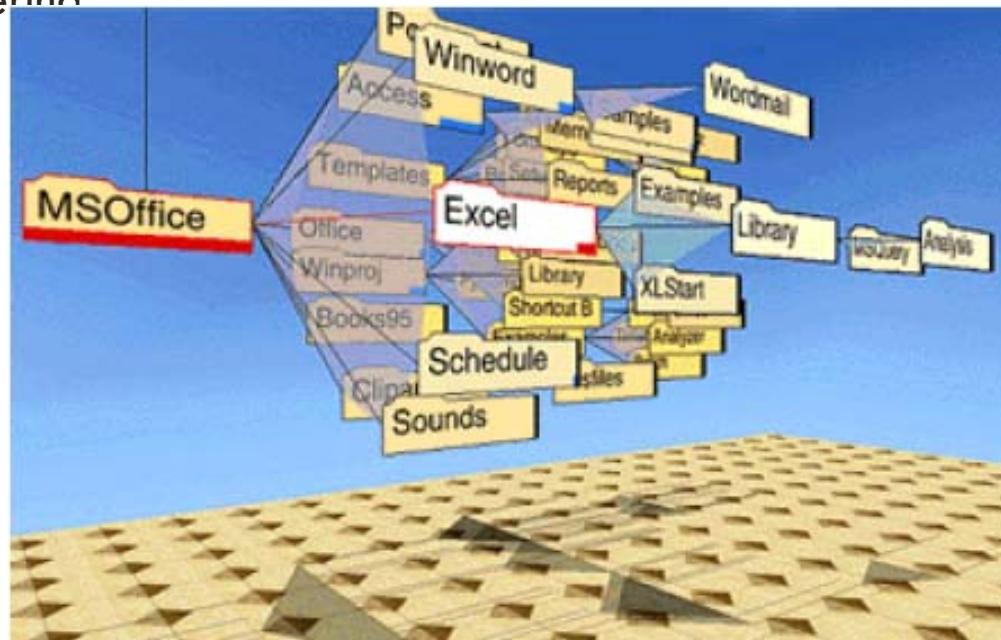
# CONE TREES

- Top-Down Ordering



# CONE TREES

- Left-Right Ordering



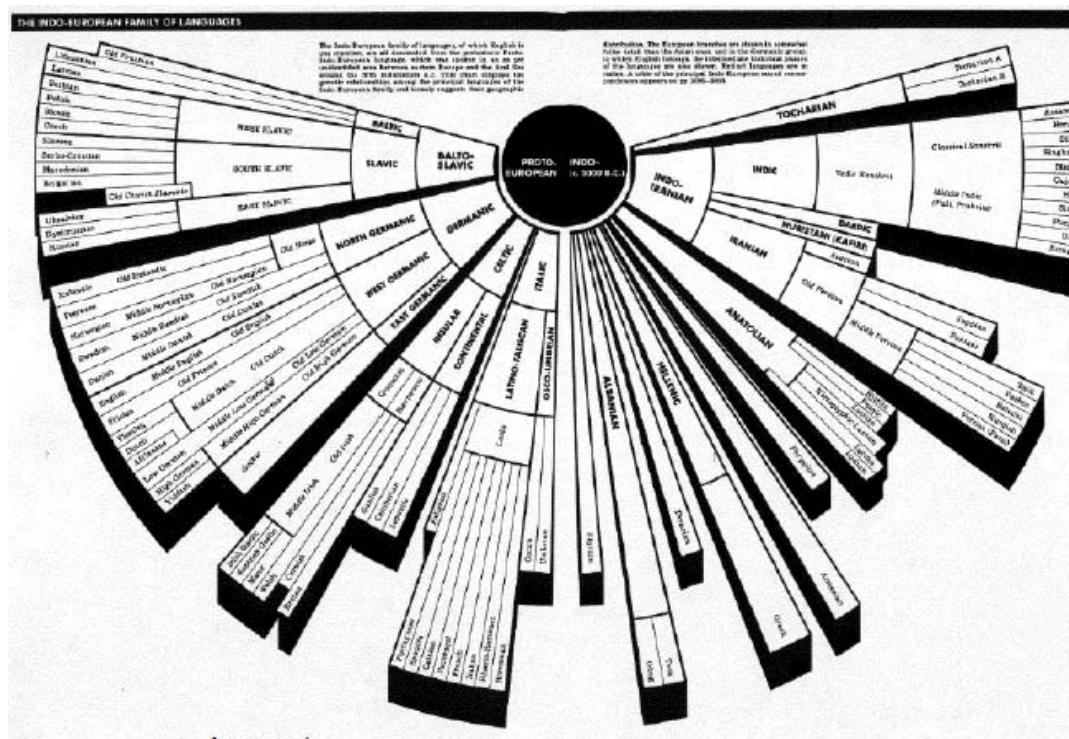
# NON-SQUARE LAYOUT?

That was basically a circular view of a tree

How about circular space filling

- Pie chart fears
- Angle to size mapping is not what we're trying to communicate

# RADIAL VIEW

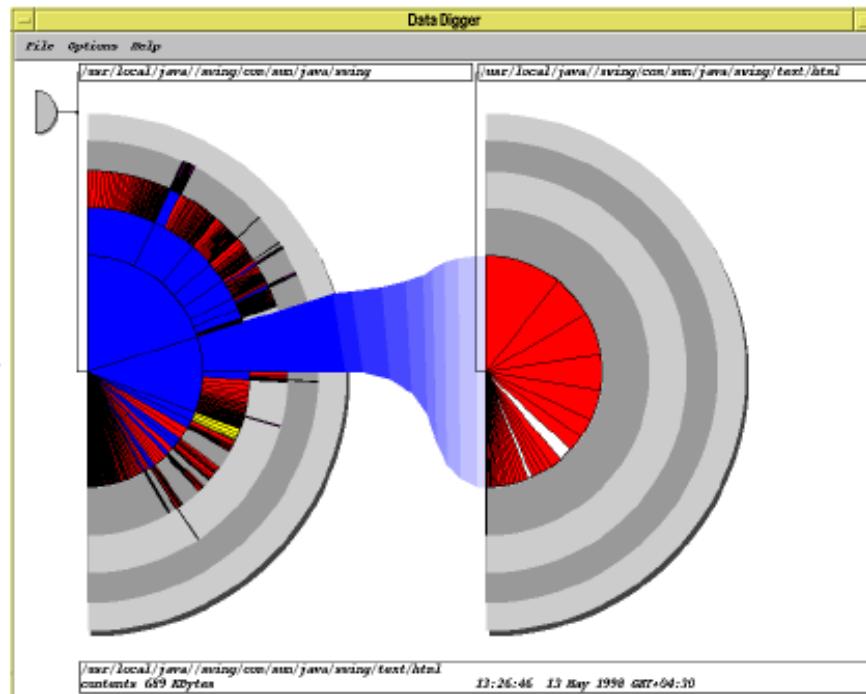


Appears in:  
*American Heritage Dictionary*, 3rd Ed. Houghton Mifflin, 1992

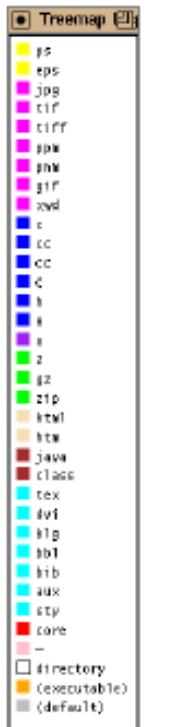
# RADIAL SPACE-FILLING

Chuah  
InfoVis '98

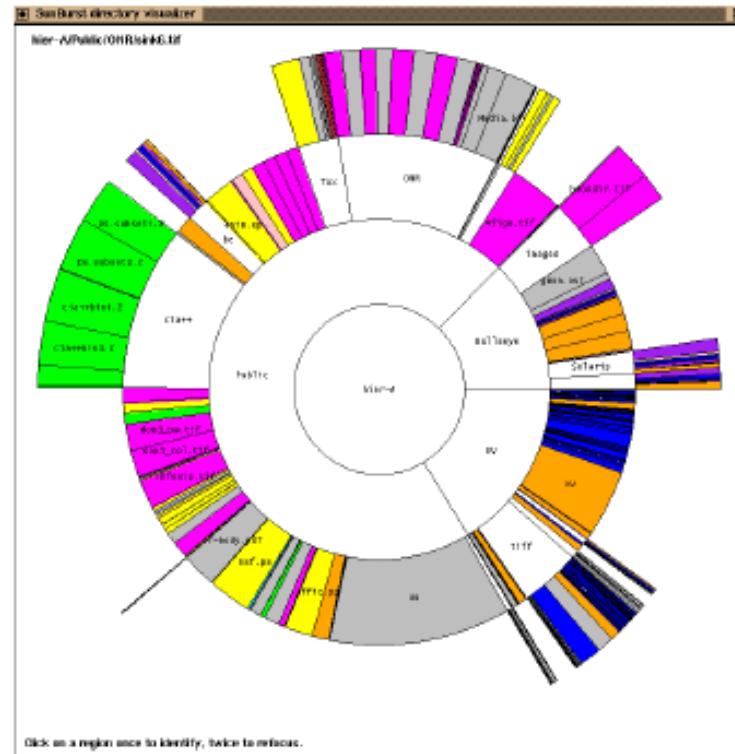
Andrews &  
Heidegger  
InfoVis '98



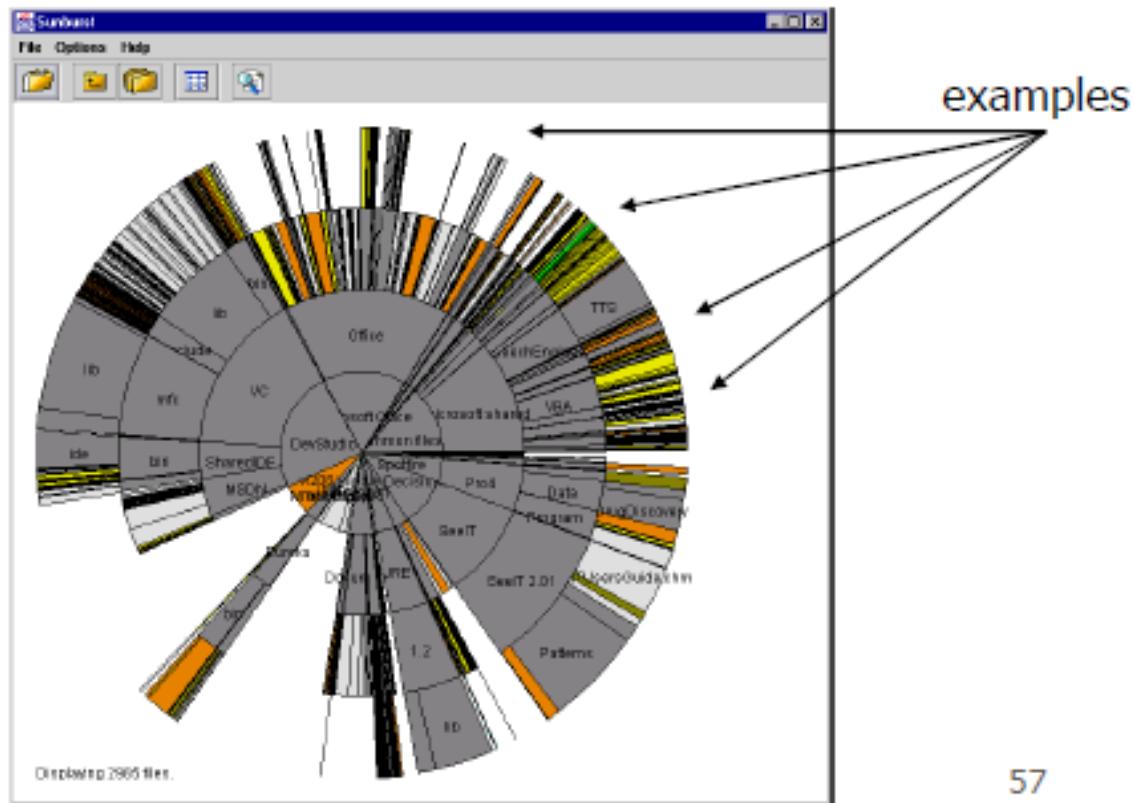
# SUNBURST



Demo



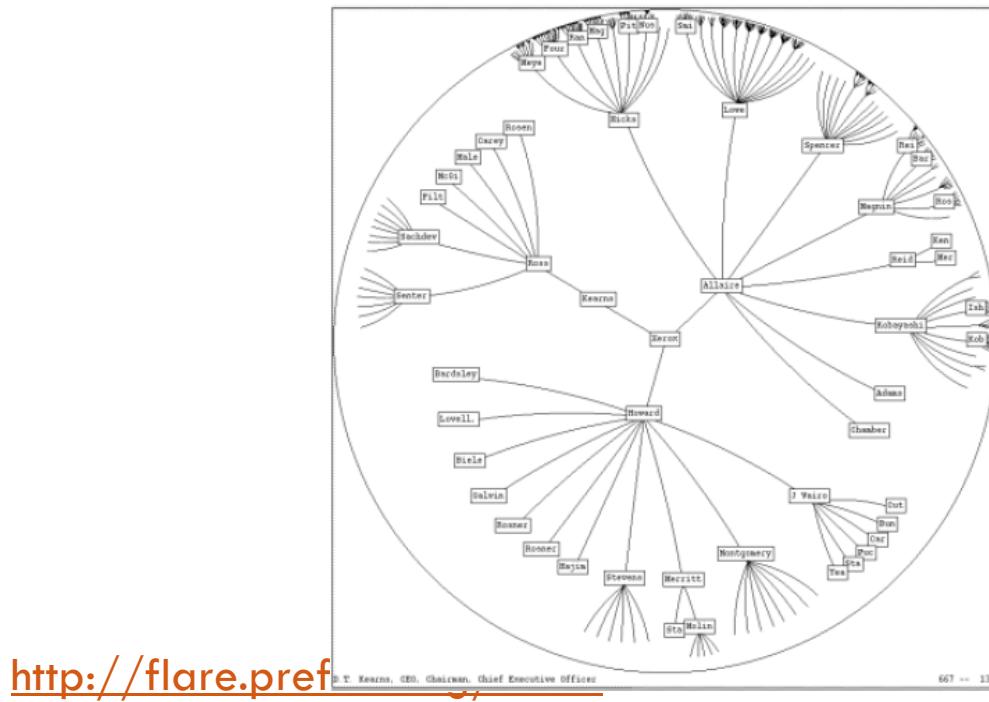
# PROBLEMS WITH RADIAL SPACE-FILLING



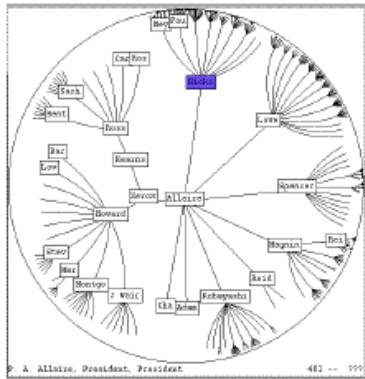
# HYPERBOLIC TREES

- Focus + Context Technique
  - Detailed view blended with global view
- Lays out the hierarchy on a hyperbolic plane, and then mapped to a disk
- Inside-outside ordering

# 2D HYPERBOLIC BROWSER



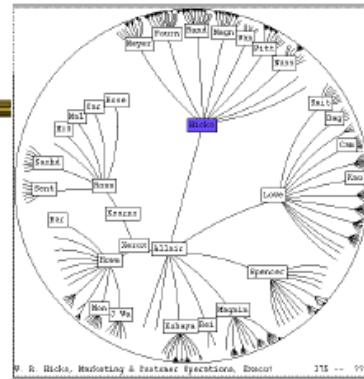
# ANIMATION IN HYPERBOLIC TREE



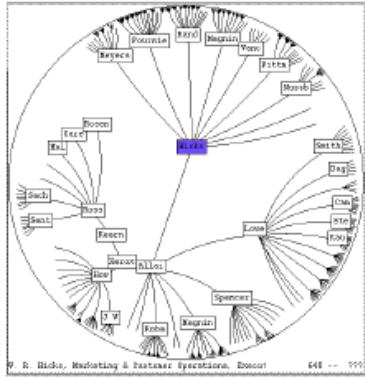
1



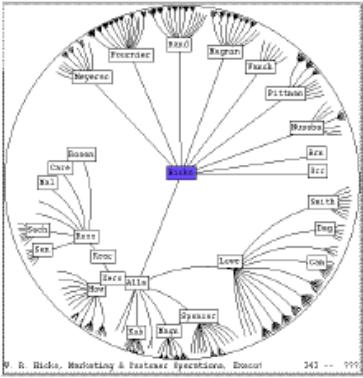
2



2



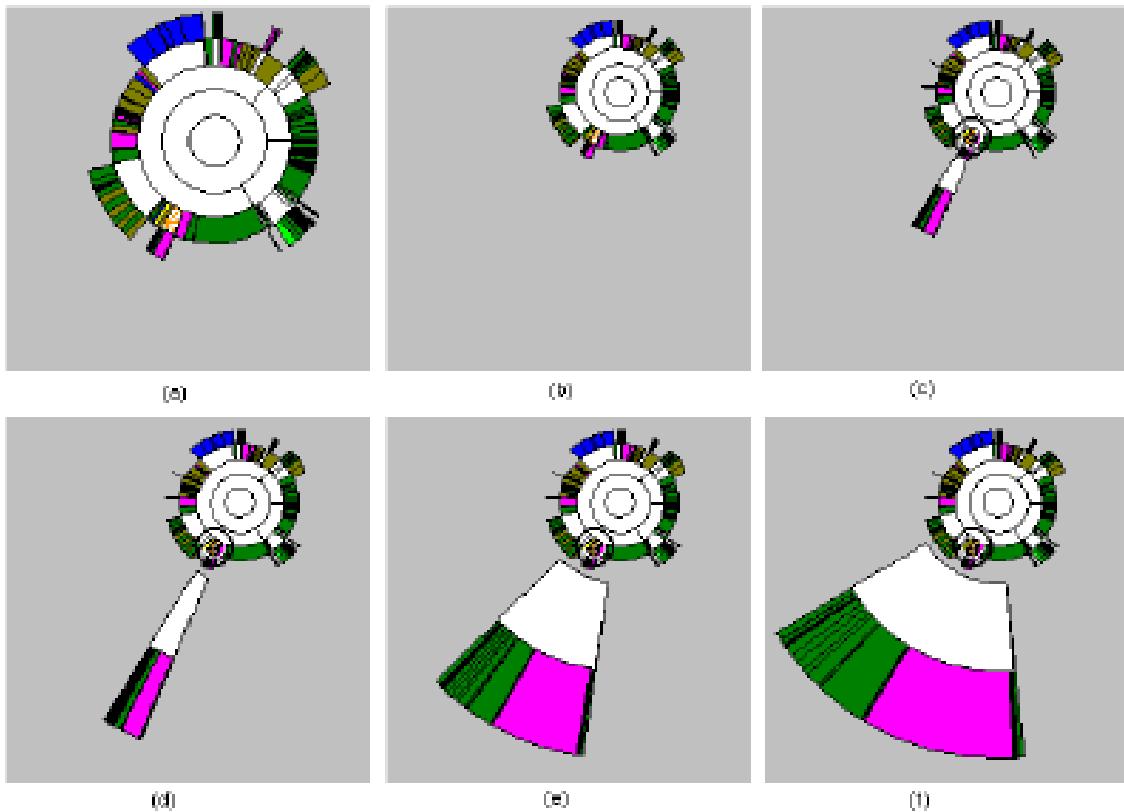
4



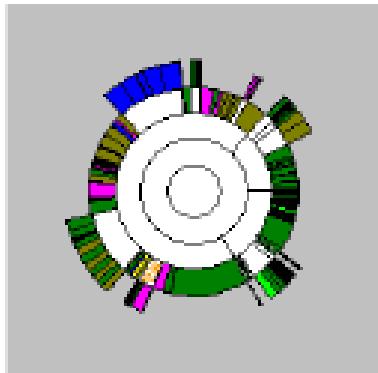
5

Clicking on the blue node brings it into focus at the center

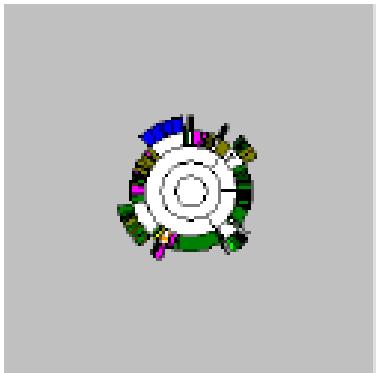
# ANGULAR DETAIL



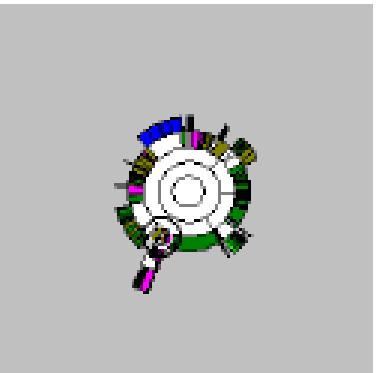
# DETAIL OUTSIDE



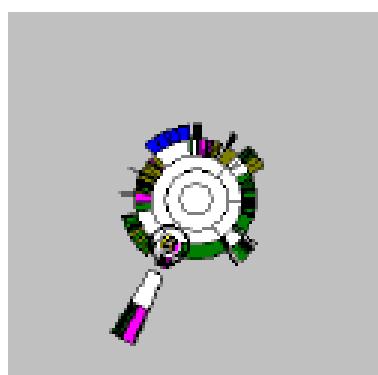
(a)



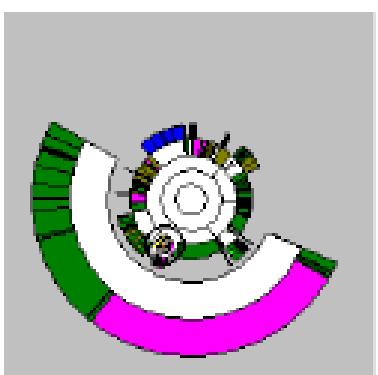
(b)



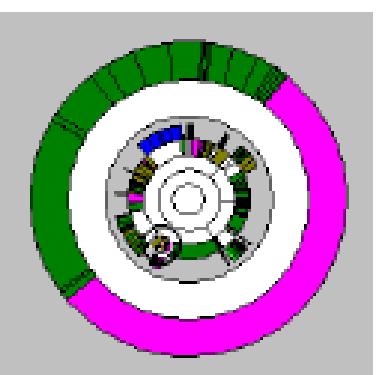
(c)



(d)

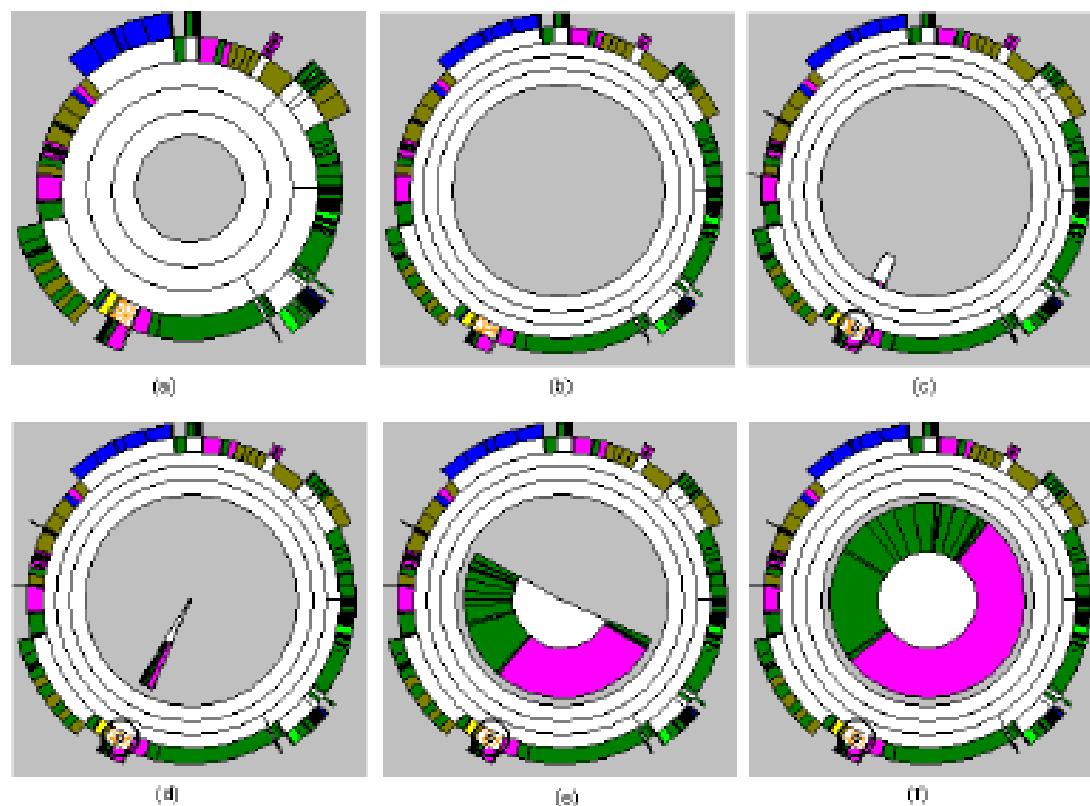


(e)



(f)

# DETAIL INSIDE



# CIRCLE PACKING

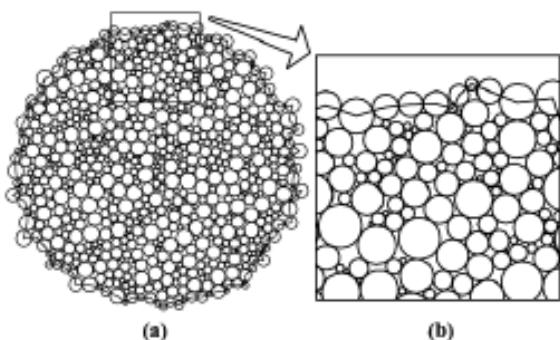


Figure 4. Packing 1000 circles with random radii

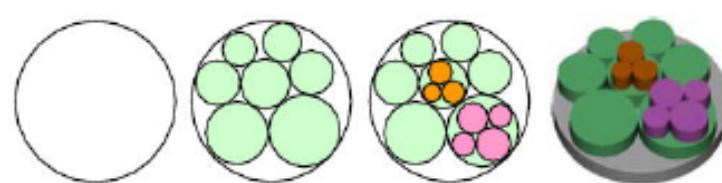
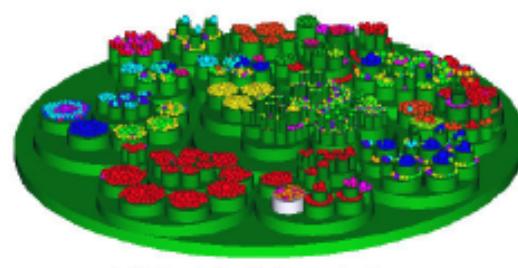
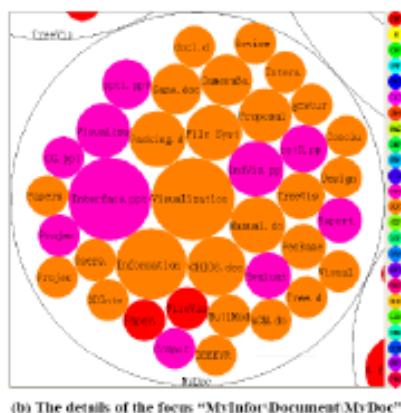
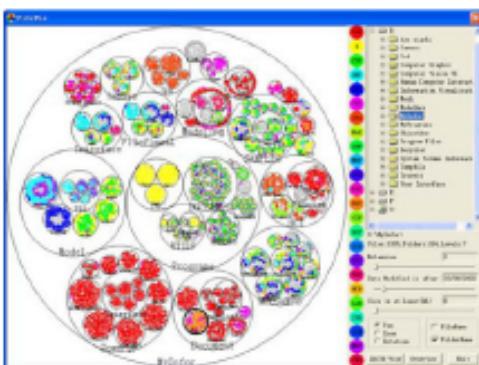
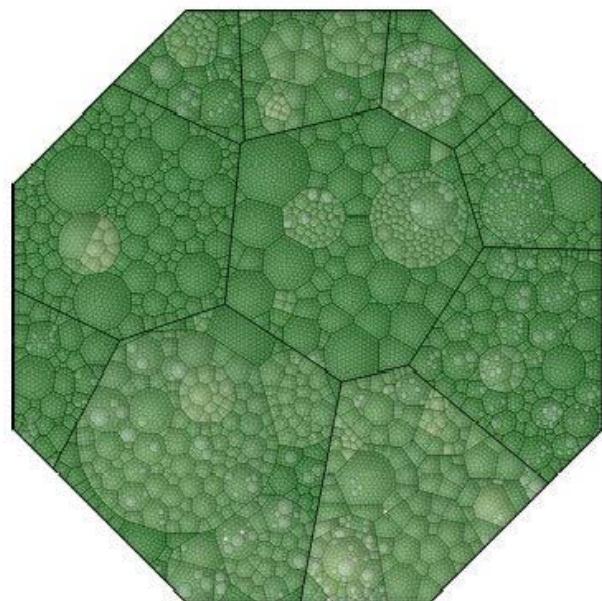
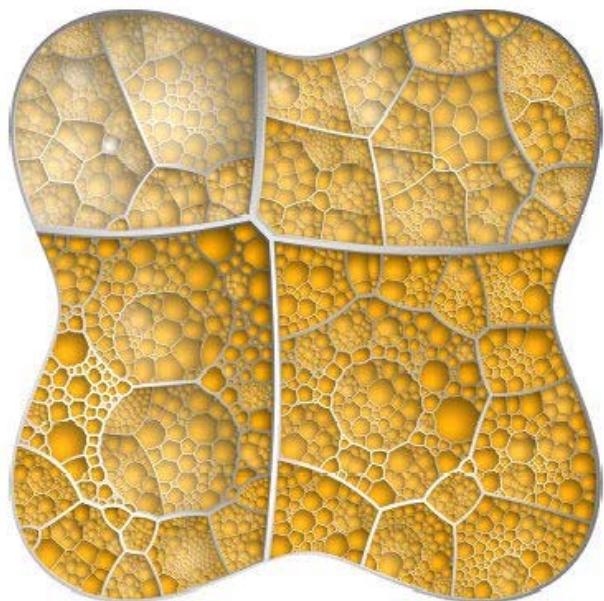


Figure 5. Pack circles into a circle recursively



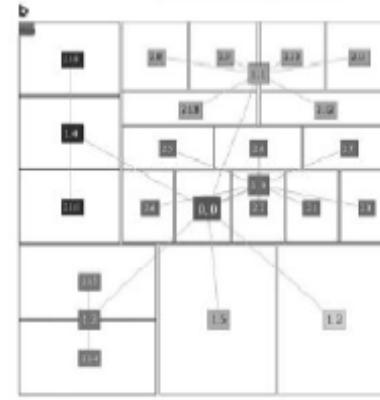
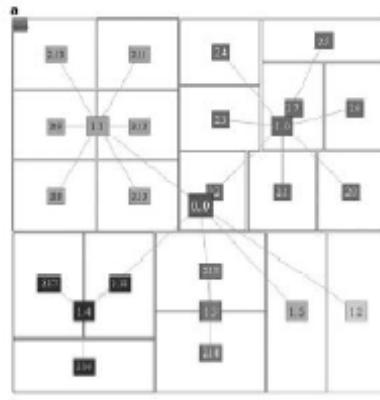
# VORONOI TREEMAPS



# HYBRID (NODE-LINK + SPACE-FILLING)

EnCon

- Explicit combination of node-link and treemap-like techniques
- Partition space into hierarchical regions, then draw node link into that



# OVERVIEW

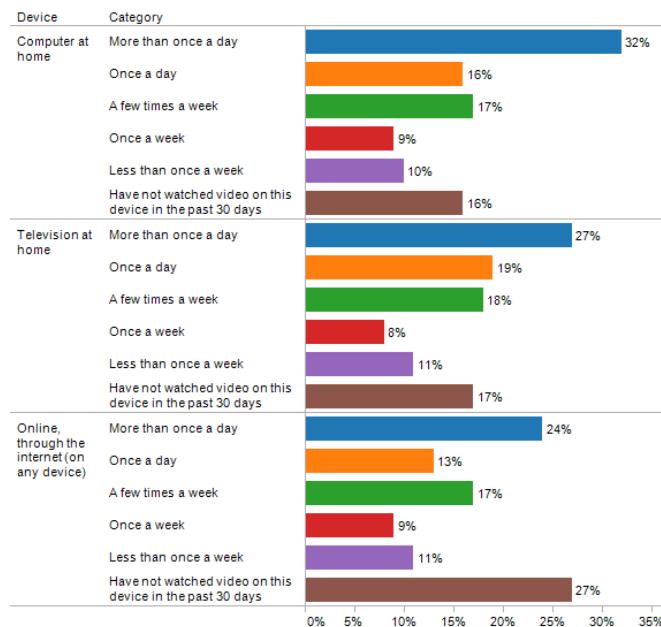
- Turns out that there are TONS more!
- IEEE InfoVis 2010, Best Poster Award: “**A Visual Survey of Tree Visualization**”
- <http://vcg.informatik.uni-rostock.de/~hs162/treeposter/poster.html>



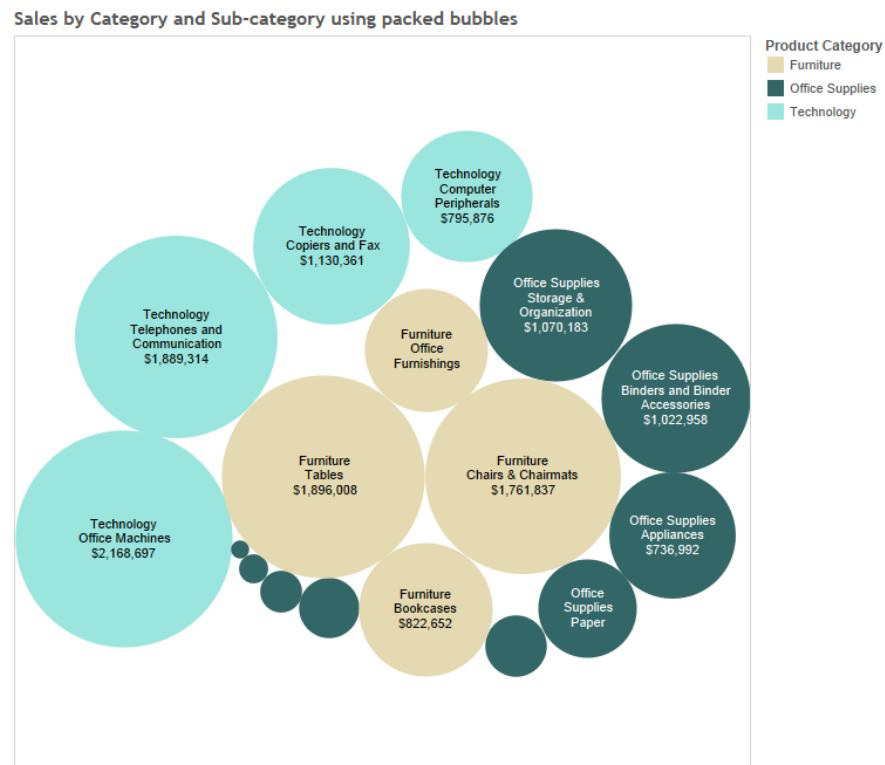
# HEIRARCHICAL BAR CHART

## Data by subcategories

- If scale is %, then proportion is easy to read off
- Distance can make bars harder to compare



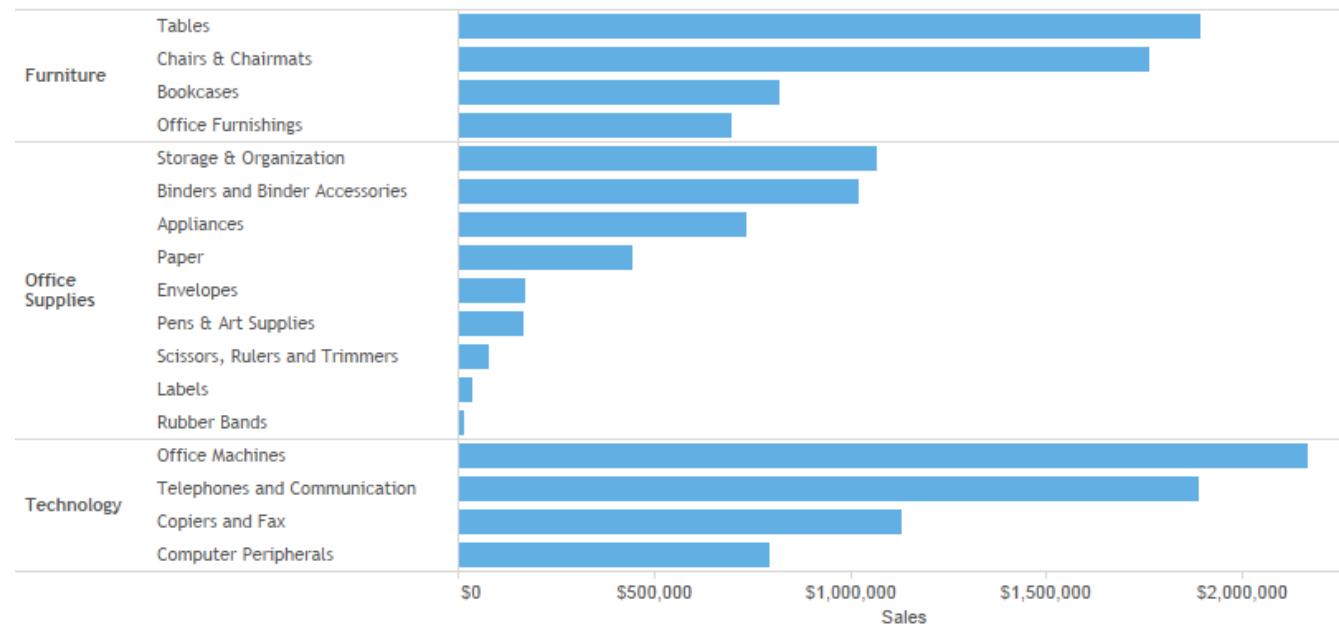
# REVISITING PACKED BUBBLES



<http://www.datarevelations.com/circles-labels-colors-legends-and-sankey-diagrams-ask-these-three-questions.html>

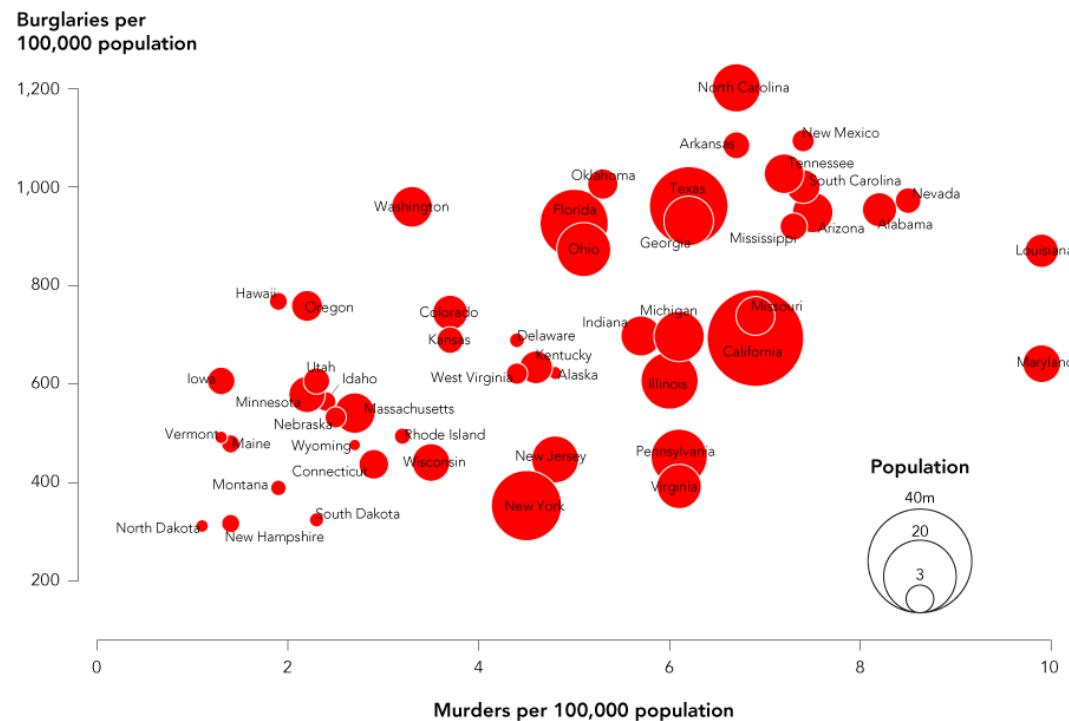
# REVISITING PACKED BUBBLES

Sales by category and sub-category



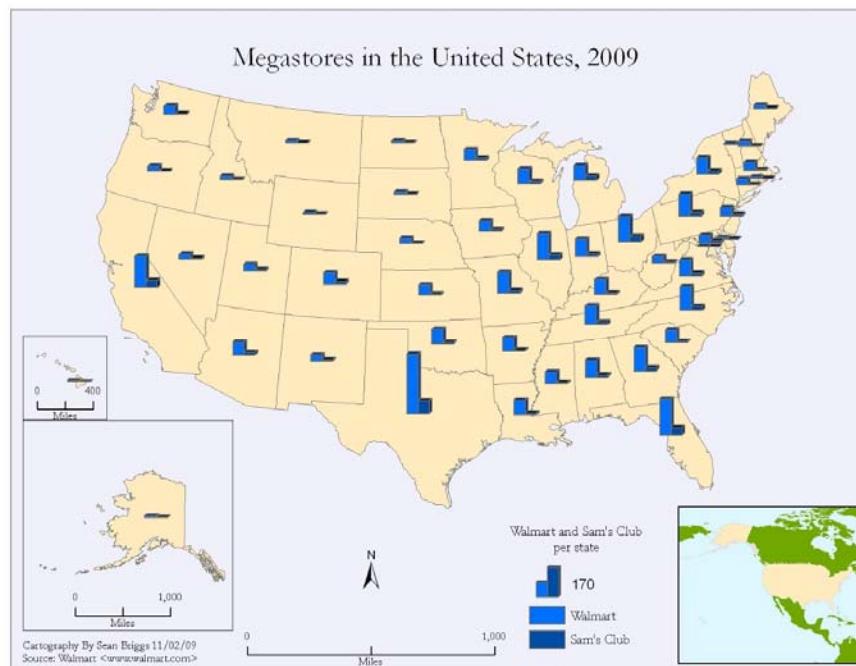
<http://www.datarevelations.com/circles-labels-colors-legends-and-sankey-diagrams-ask-these-three-questions.html>

# ANOTHER USE FOR BUBBLES



<https://flowingdata.com/2010/11/23/how-to-make-bubble-charts/>

# EVER SEEN SOMETHING LIKE THIS?



# COMPARE

African-Americans are more heavily represented in the Southeastern states  
%African-Americans, by state, 2002

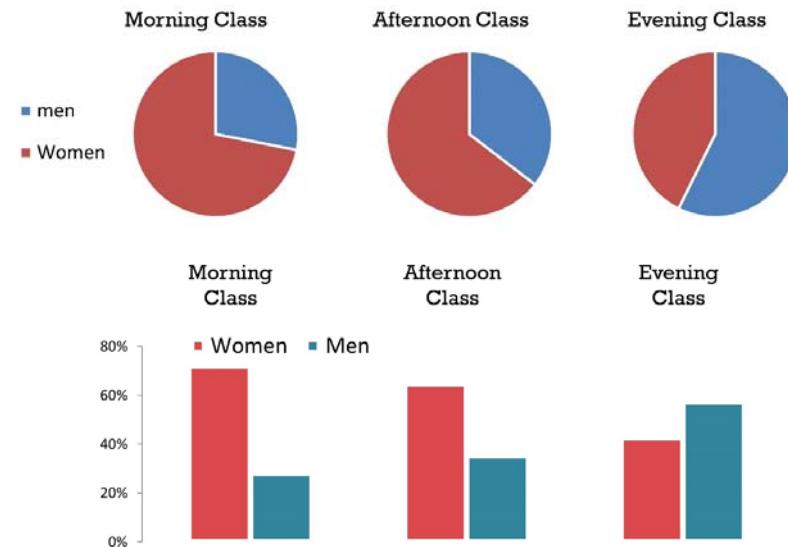


Megastores in the United States, 2009



# DISCUSSION ...

Women are more likely to attend the day classes, while men are more commonly found in the evening class

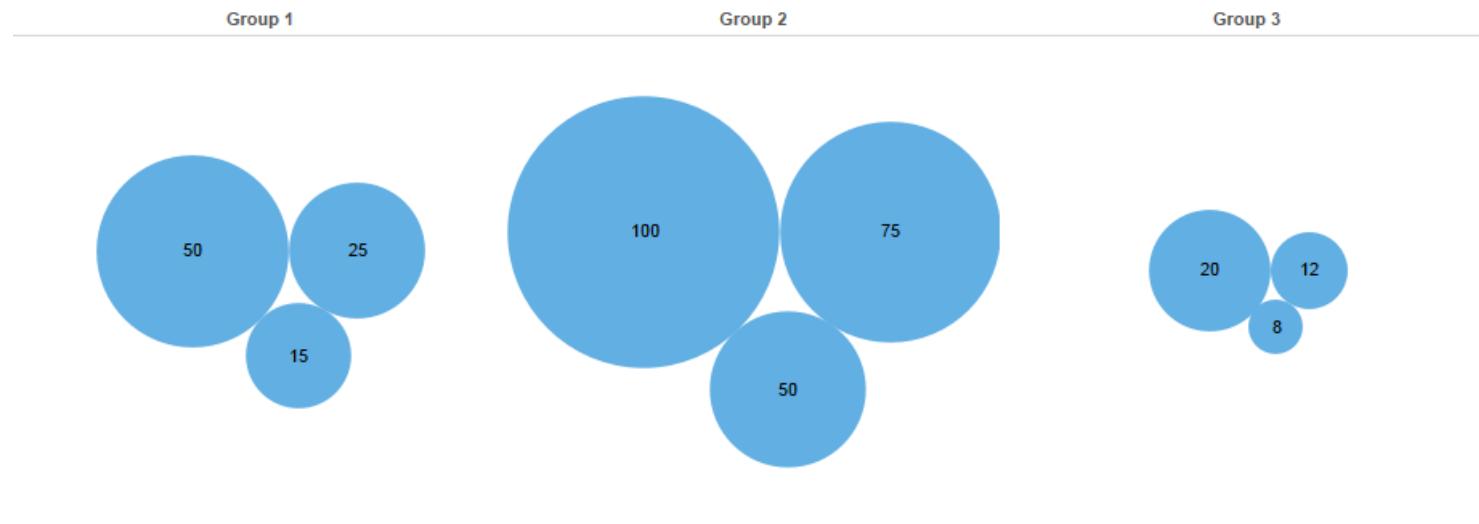


Bruce Gabrielle,

<http://speakingppt.com/2013/03/18/why-tufte-is-flat-out-wrong-about-pie-charts/>

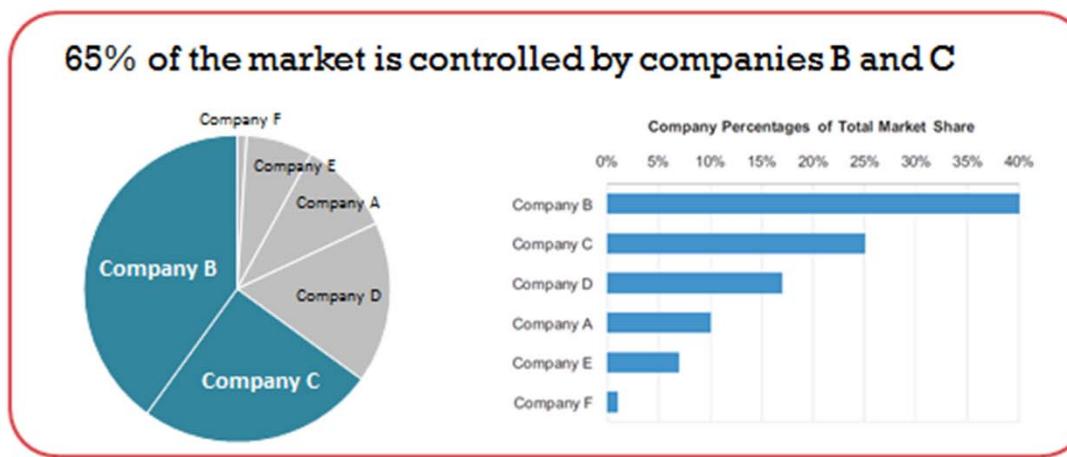
# WHICH DATA ARE GRAPHED CORRECTLY?

Bubbles -- which group is right?



<http://www.datarevelations.com/circles-labels-colors-legends-and-sankey-diagrams-ask-these-three-questions.html>

# EXCEPTIONS



Bruce Gabrielle,

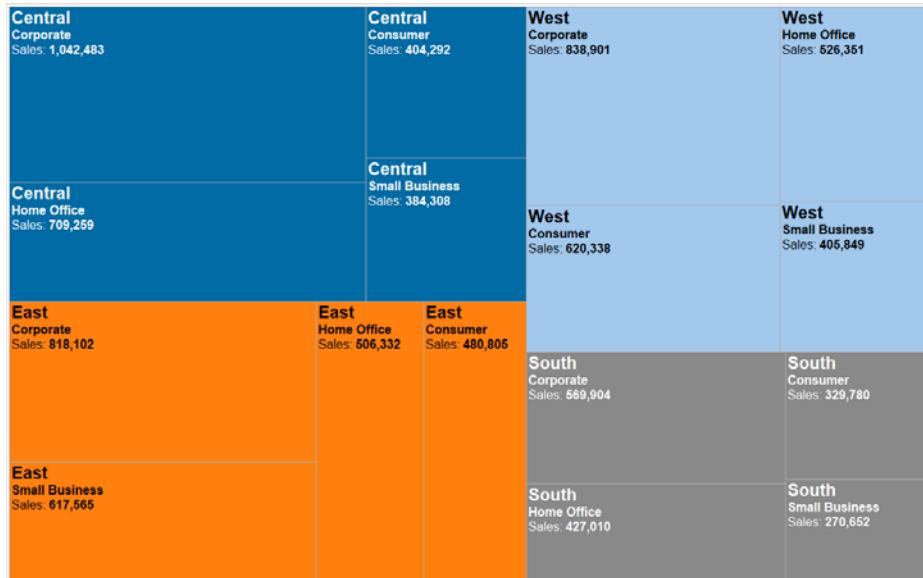
<http://speakingppt.com/2013/03/18/why-tufte-is-flat-out-wrong-about-pie-charts/>

# THE TREE MAP

Shows hierarchy with containment metaphor

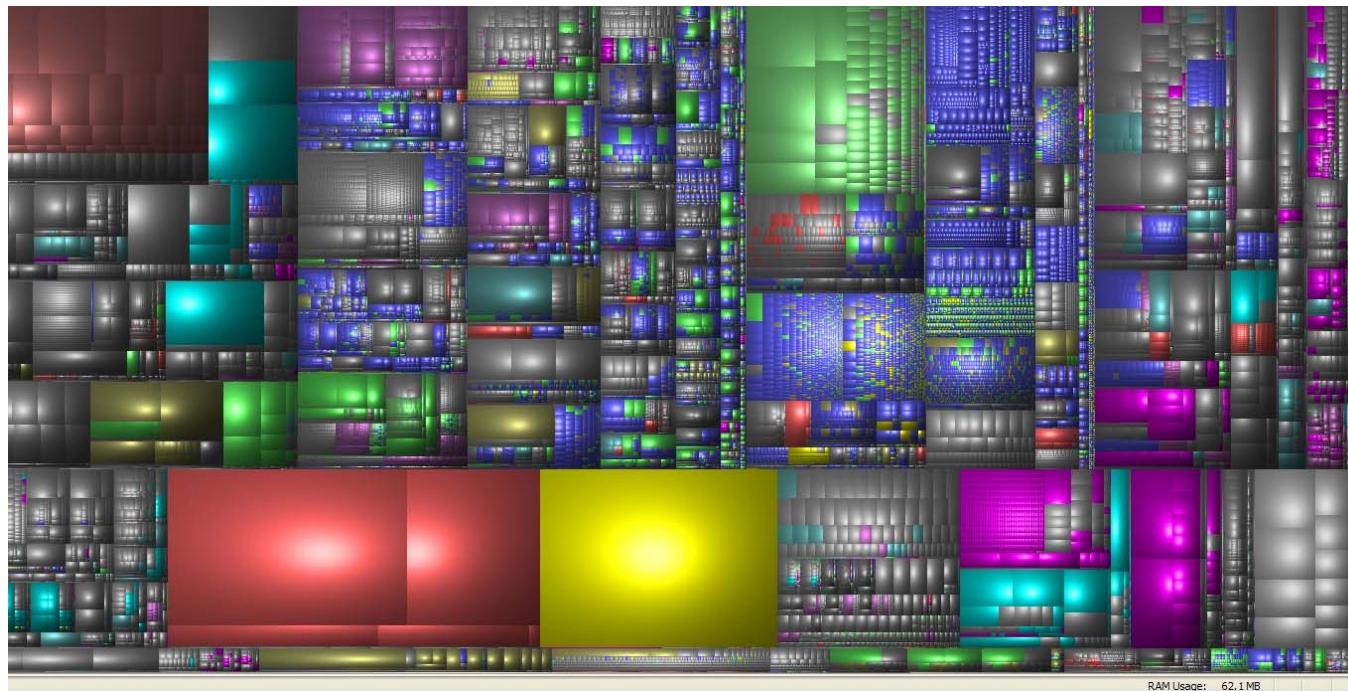
Draws recursively

Core feature in Tableau, R has “treemap” pkg.



[Theinformationlab.com](http://Theinformationlab.com)

# THE TREE MAP



# SPACESNIFFER

