

Class 2: More Processing, plus Design for Data

FA12/New School
zannah marsh
zmarsh10@gmail.com

Plan for tonight...

- Look at homework!
- More with Processing
- Assignment schedule
- Break
- Lecture on Data Design

Programming!



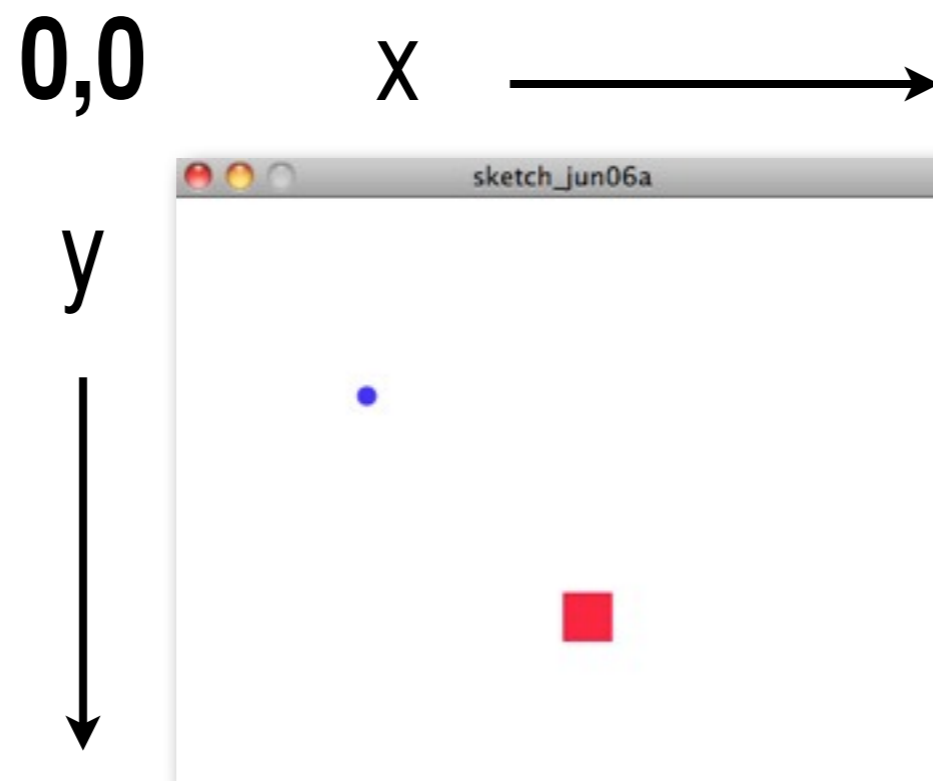


When programming, you need to:

- Write effective, logical instructions
- Understand the **context and tools** that you and the computer share (languages and libraries)
- Understand which tools and contexts you need to **create or explain** to the computer
- Understand the **syntax** the computer needs you to use

Processing Basics

- Visual output— program files are called **sketches**, program “**draws**” to the screen
- All measurements in pixels
- Built in **loop** for animation (change from frame to frame)
- Screen space defined by **coordinate system** (x, y, z)



text in...

...visuals out

```
sketch_jun06a | Processing 1.2.1
sketch_jun06a §
//set the size of your sketch (w,h in pixels)
size(400,300);
//change the color for the background
//0 to 255, 0 is back, 255 is white
background(255);
//turn off the outline around things
noStroke();
//smooth out my shapes
smooth();

//fill red-ish
fill(255,0,50);

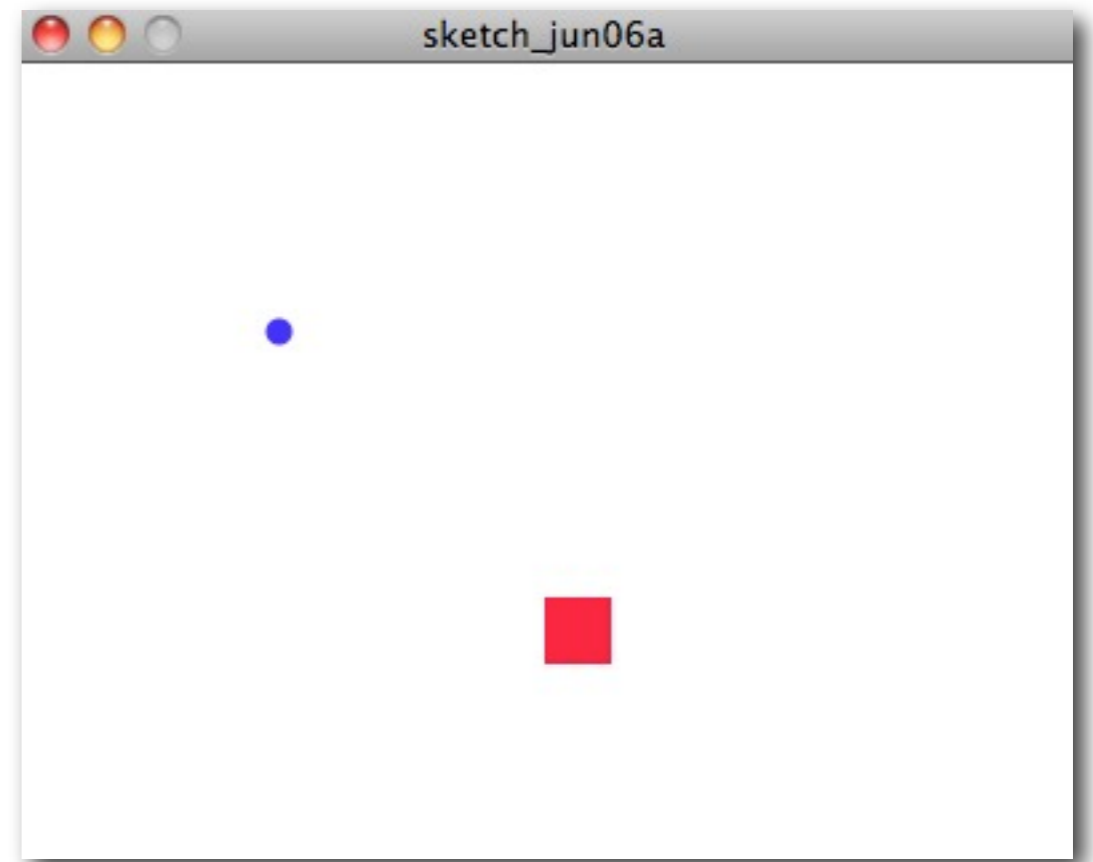
//draw a rectangle at x=200, y=200, with width of 25
//and height of 25
rect(200,200,25, 25);

//fill blue-ish
fill(50,0,250);

//draw an ellipse at x=100, y=200, with a
//width and height of 10
ellipse(100,100,10,10);
```

X →

y ↓



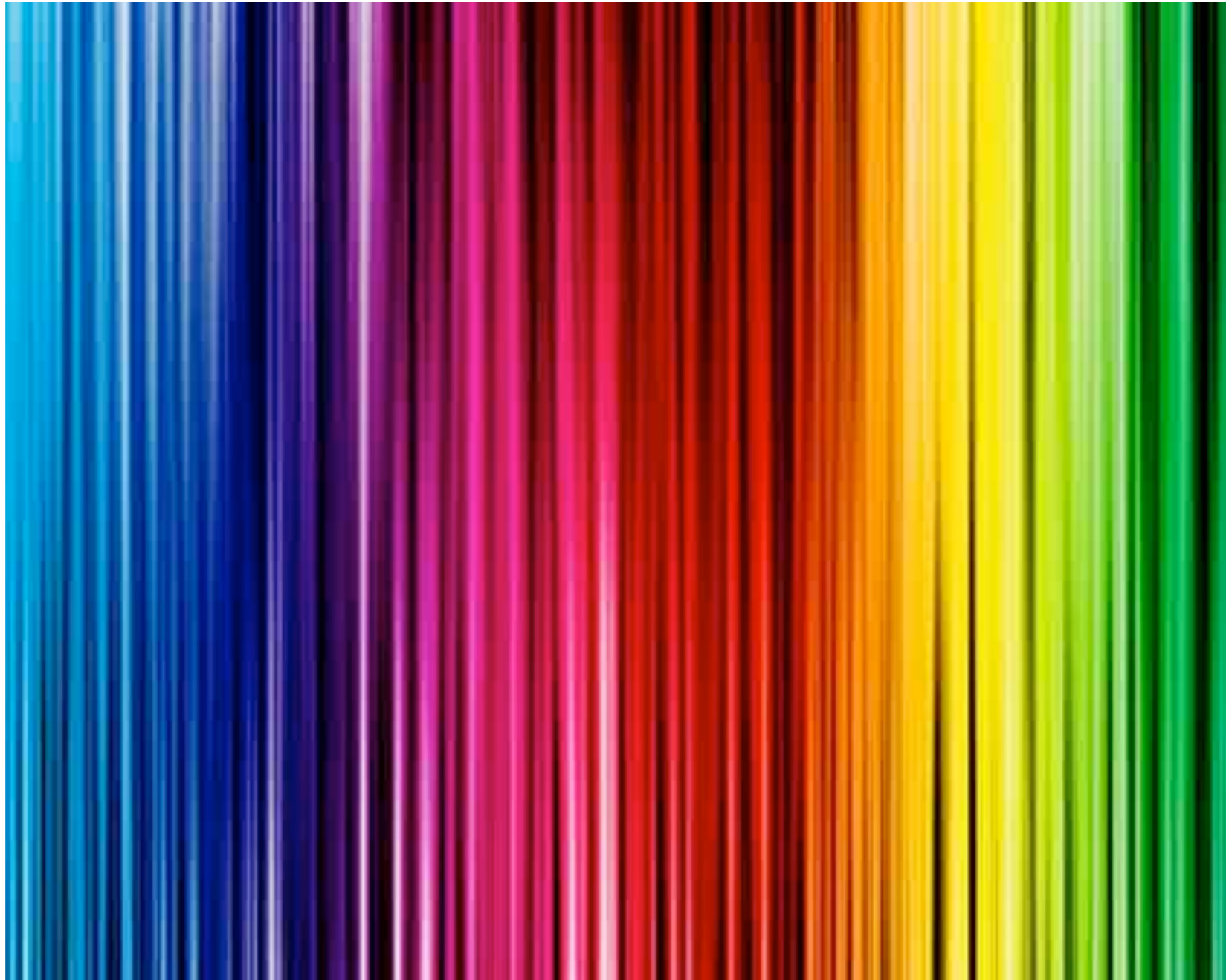
Functions!

- Out-of-the-box reusable chunks of code! Shared tools or methods that you and the computer have at your fingertips
- You write functions **ONCE**, but can use or **call** them multiple times— very handy
- You can give functions extra information to make them more flexible and useful

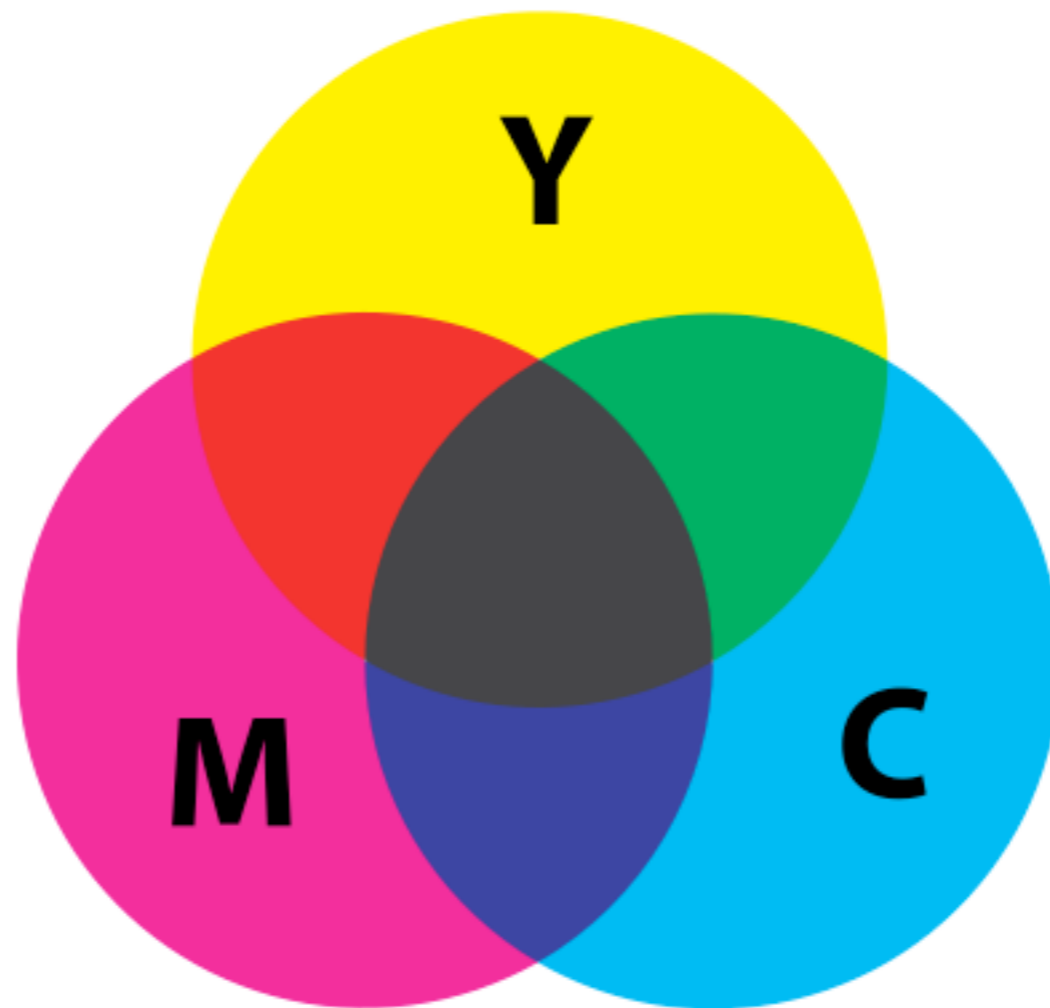
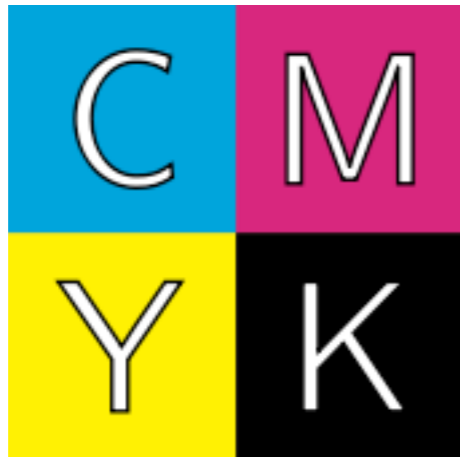
First (Built-in) Functions In Processing

- size(), background()
- rect(), ellipse(), line(), triangle(),
- fill(), stroke(), noStroke(), strokeWeight()
- Remember, functions take **arguments...**
 - integers from 0 to 255 for RGB or grayscale colors
 - integers for x and y positions onscreen, width, height, etc

Color!

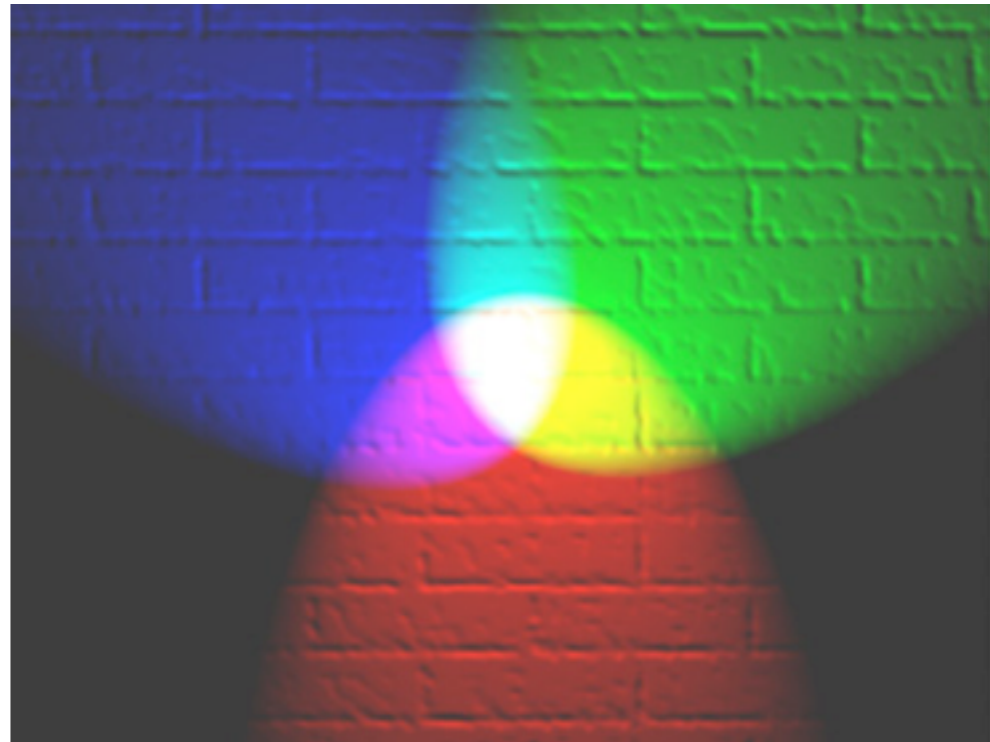
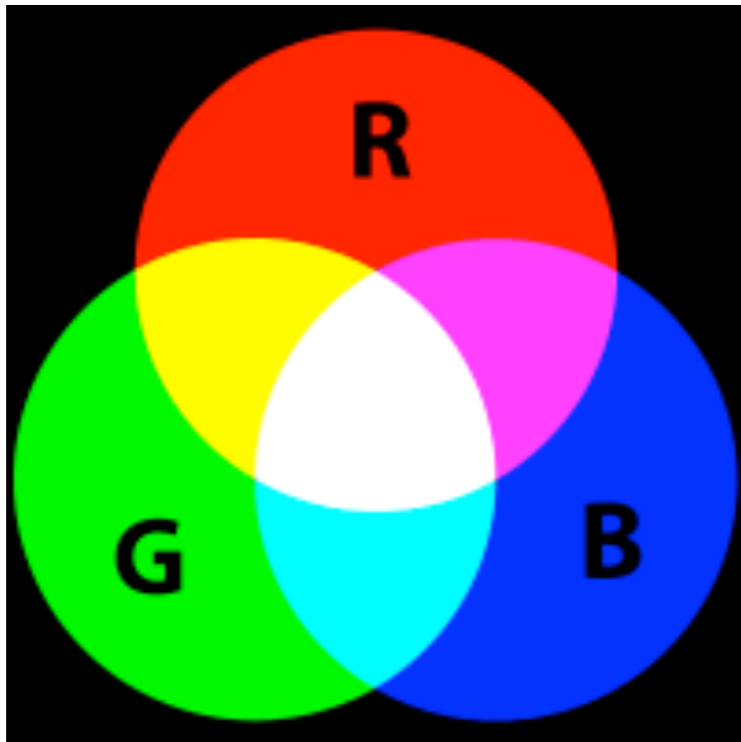


CMYK: *subtractive color*



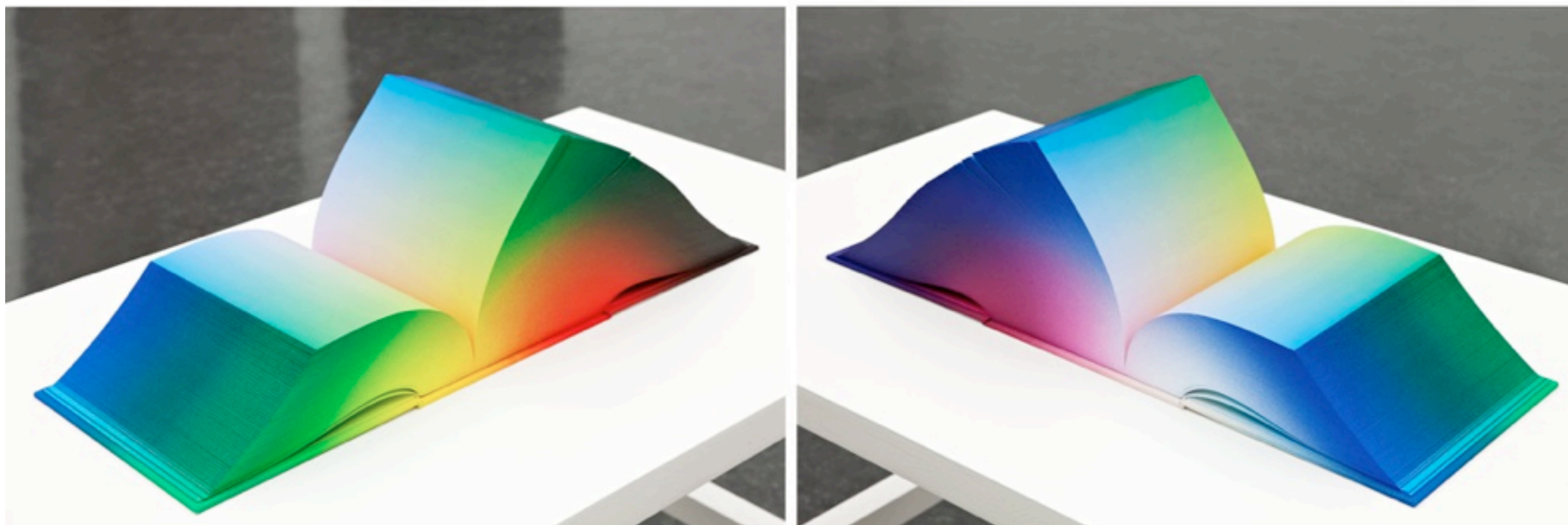
colors (ink) applied to white surface (paper) reduce the light that would normally be emitted by the surface

RGB: *additive color*

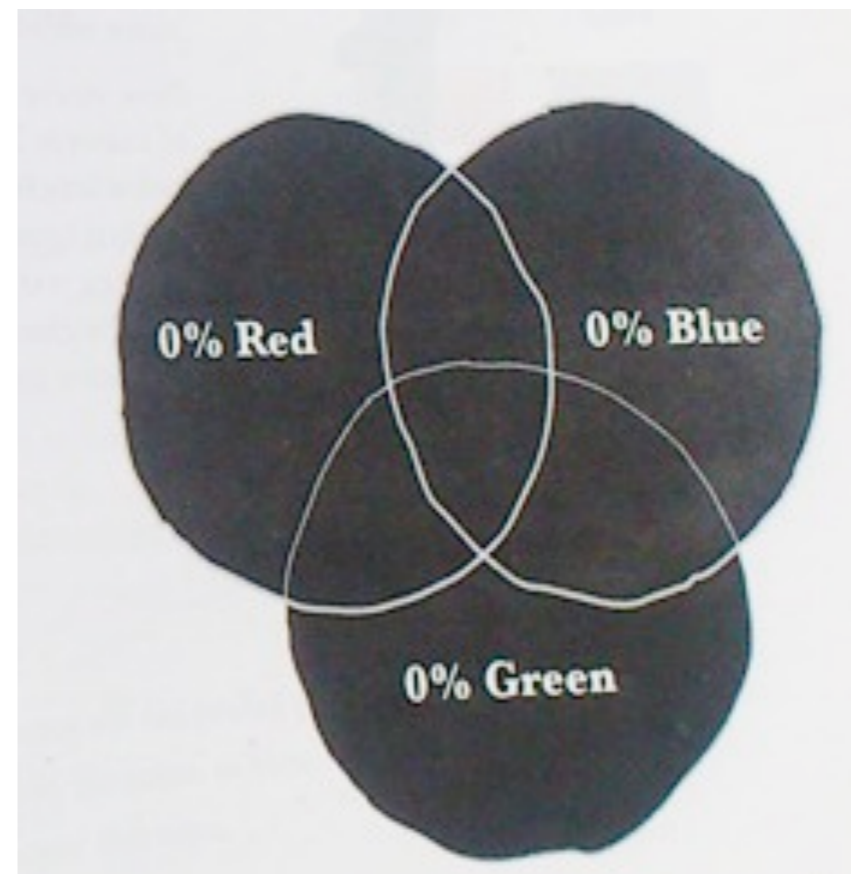
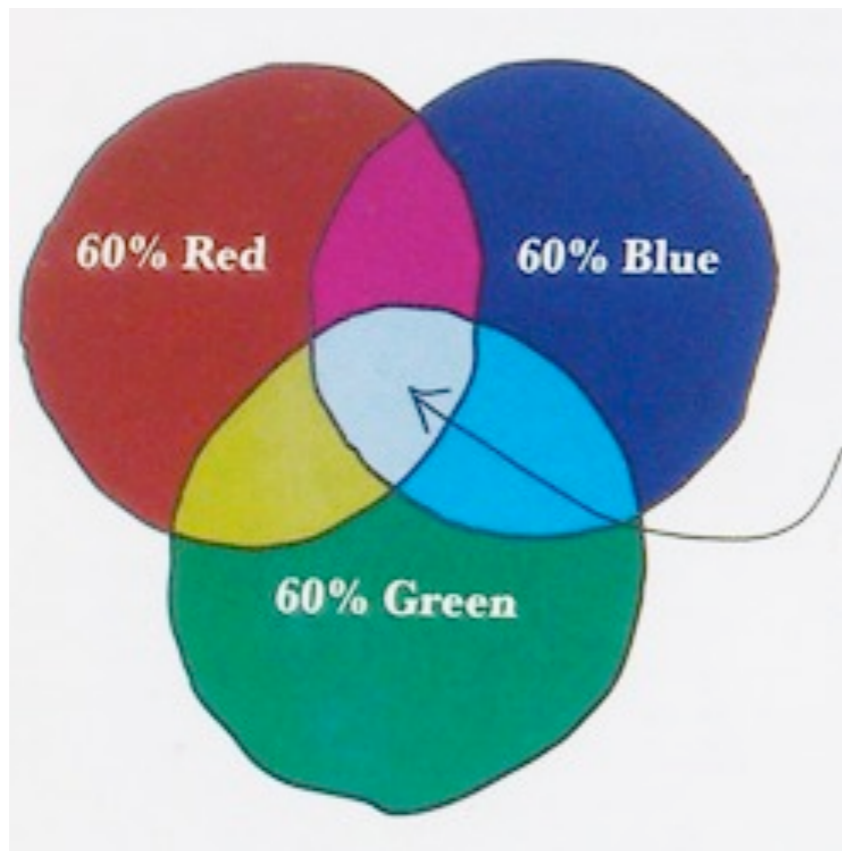
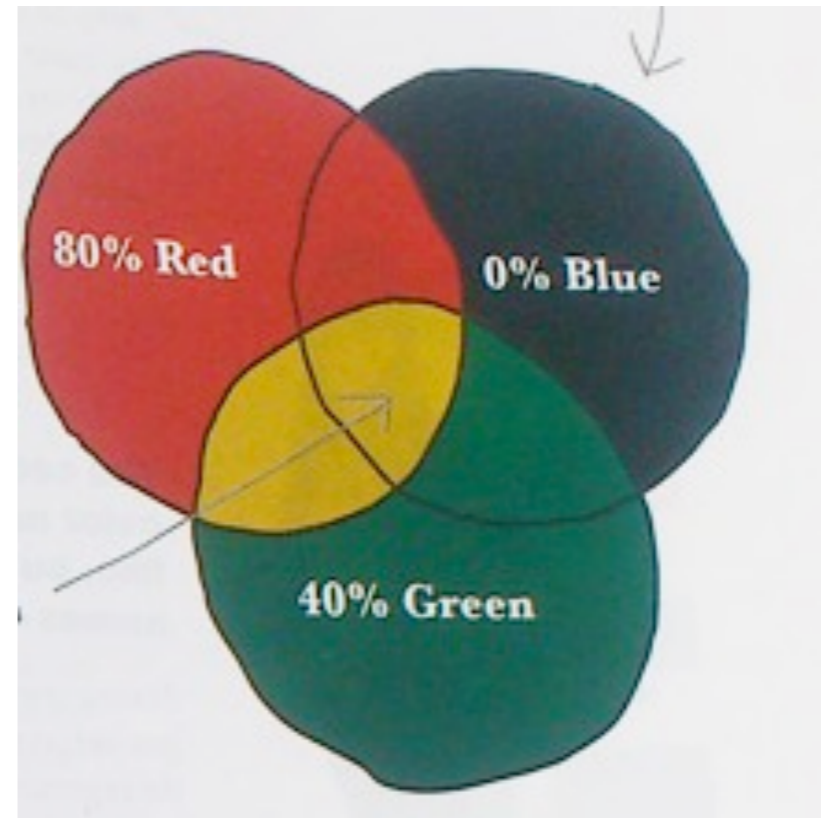
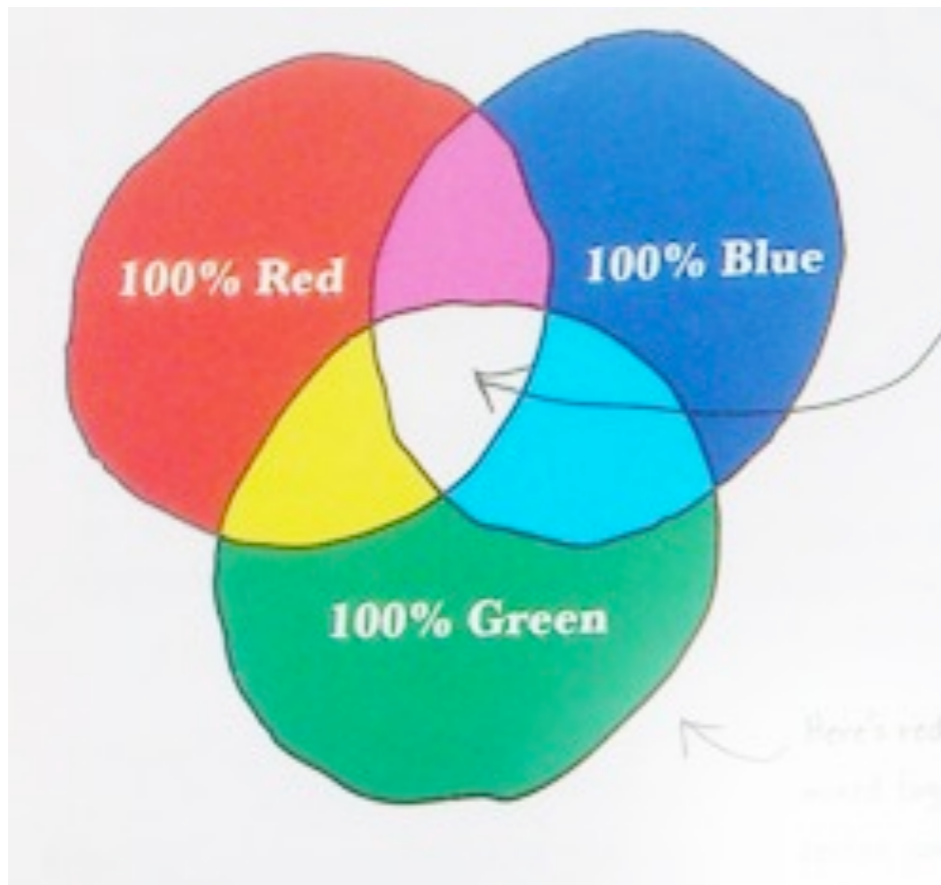


red, green, and blue light mixed (added) together to make a color: 100% of each makes white, 0% of each makes black.

Values for R, G, and B are often each specified as a number between 0 and 255.



RGB colorspace as a book (Tanya Auerbach)



Using void setup() and void draw()

- Both are functions, neither take parameters
- Most processing sketches use them
- setup() does preliminary stuff once
- draw() loops over and over again until you close the sketch, allows for animation
- void means neither of the functions will “return” (give back) anything to you

Using a variable

- A variable is a **place to store some useful data** (a number, a sequence or string of letters that make up a word, or a object (more on these later))
- Variables have a **datatype**, so the computer knows what it can and can't do with the value (int, string, float, etc)
- Variables point to spaces in memory where data is save; values held there can be **changed**
- In Processing, variables must be **declared** and **initialized** before you use them!

Declaring and Initializing a Variable

in two separate steps:

int ypos; ypos = 10;

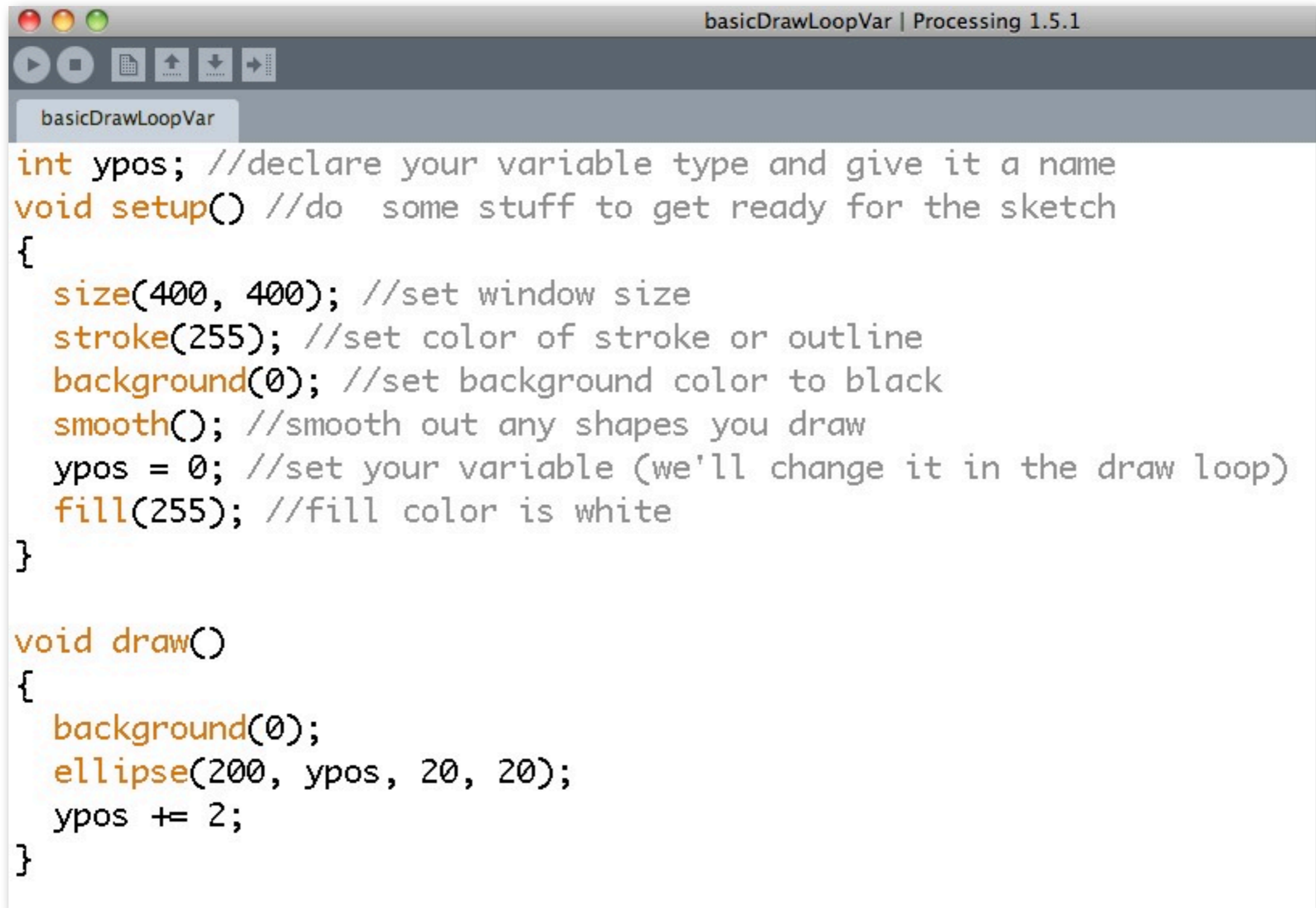
↑ ↑ ↑

data type variable name set it equal to a value
(you make up)

or in one step:

int ypos = 10;

Let's look at some code....

A screenshot of a Processing IDE window. The title bar reads "basicDrawLoopVar | Processing 1.5.1". The window contains a code editor with the following code:

```
int ypos; //declare your variable type and give it a name
void setup() //do some stuff to get ready for the sketch
{
  size(400, 400); //set window size
  stroke(255); //set color of stroke or outline
  background(0); //set background color to black
  smooth(); //smooth out any shapes you draw
  ypos = 0; //set your variable (we'll change it in the draw loop)
  fill(255); //fill color is white
}

void draw()
{
  background(0);
  ellipse(200, ypos, 20, 20);
  ypos += 2;
}
```

Adding a conditional (if statement)

Think in **pseudocode**: make up a human sentence about what you want the computer to do....



- If there is jam in the jar, open the jar
- If there is jam in the jar, open the jar ELSE put down the jar
- If the y-position of the ellipse is off the screen...

If statement

```
if ( some condition)
{
    do this...
}
```

Add text to the screen!

- Use the **PFont** object to store your font info
- Declare it like a regular variable: **PFont** font;
- Create your font using Tools > create font
- Initialize it using the loadFont() function, which takes the font file name as an argument
- Call the function textFont() with your new font variable: textFont(font); this will ready the font for use onscreen

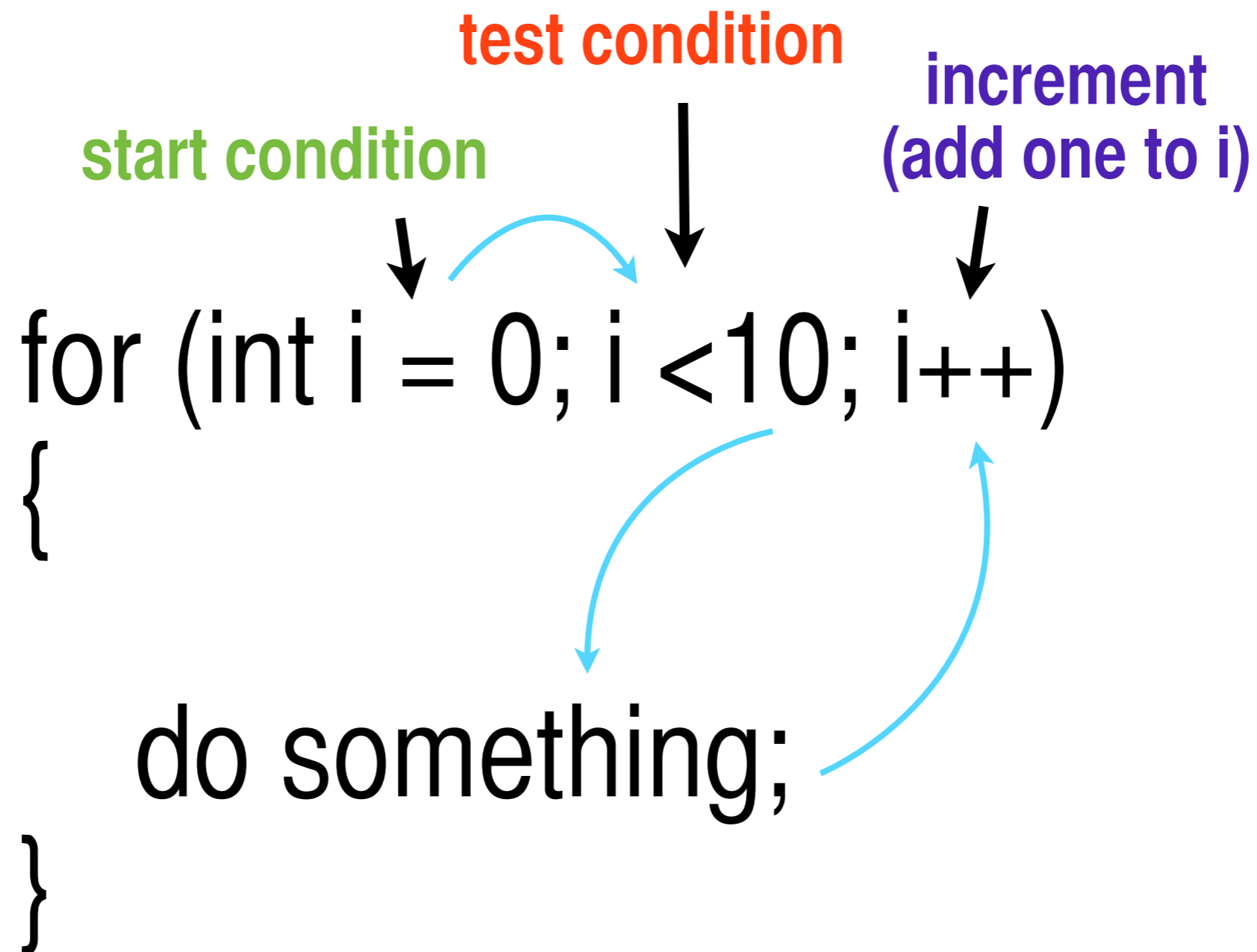
Thinking about loops...

- why are they useful?
- how can we harness them?
- what are their pitfalls?



The mighty For loop!

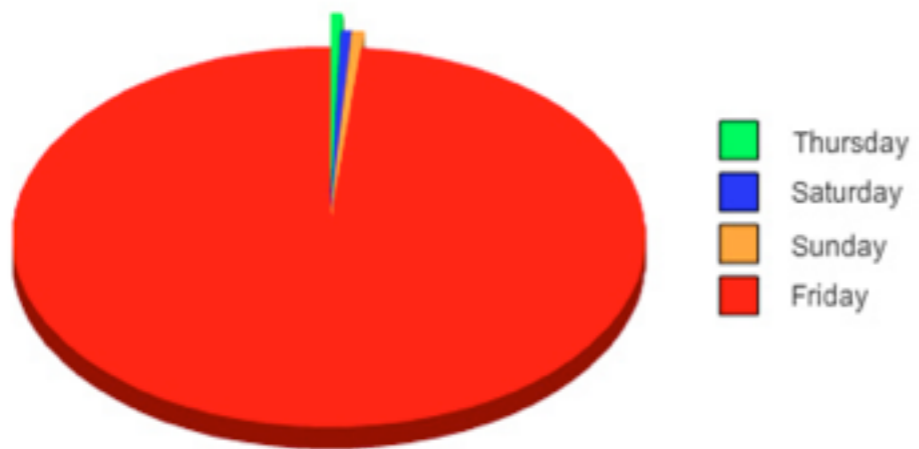
Use a for loop to do something a certain number of times.
Here, i is a variable that keeps track of your loop, like a counter...



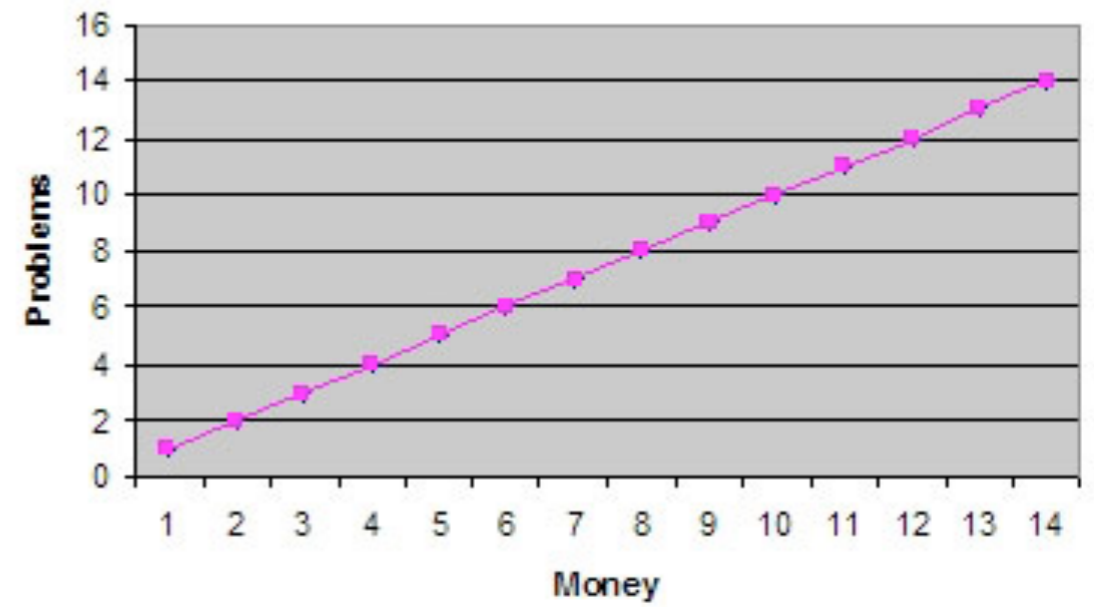
Data vis/infographic
presentation schedule...

Lecture: Design for Data

We So Excited

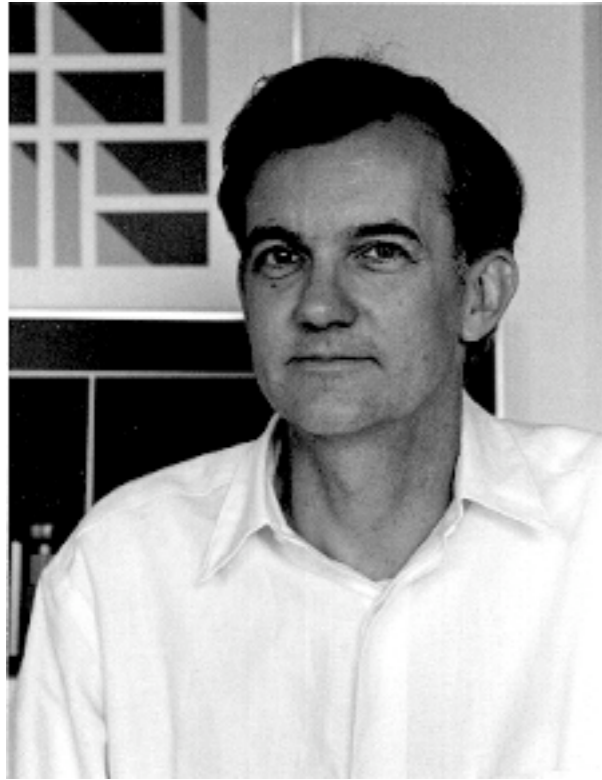


Relationship Between Money and Problems



Effective Visualizations

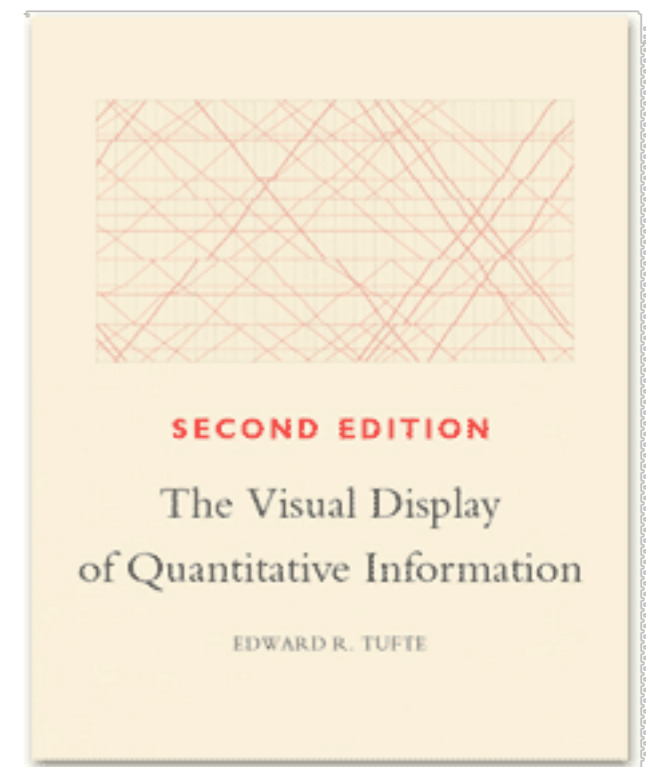
- Enable a specific task
- Show something previously unseen (and hopefully true!)
- Off-load cognitive work
- Leverage strengths of human visual system



Edward Tufte

American statistician and professor emeritus of political science, statistics, and computer science at Yale University.

“...[give] to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.”



Event! ET Modern

507 W. 20th St. NY NY

EDWARD TUFTE

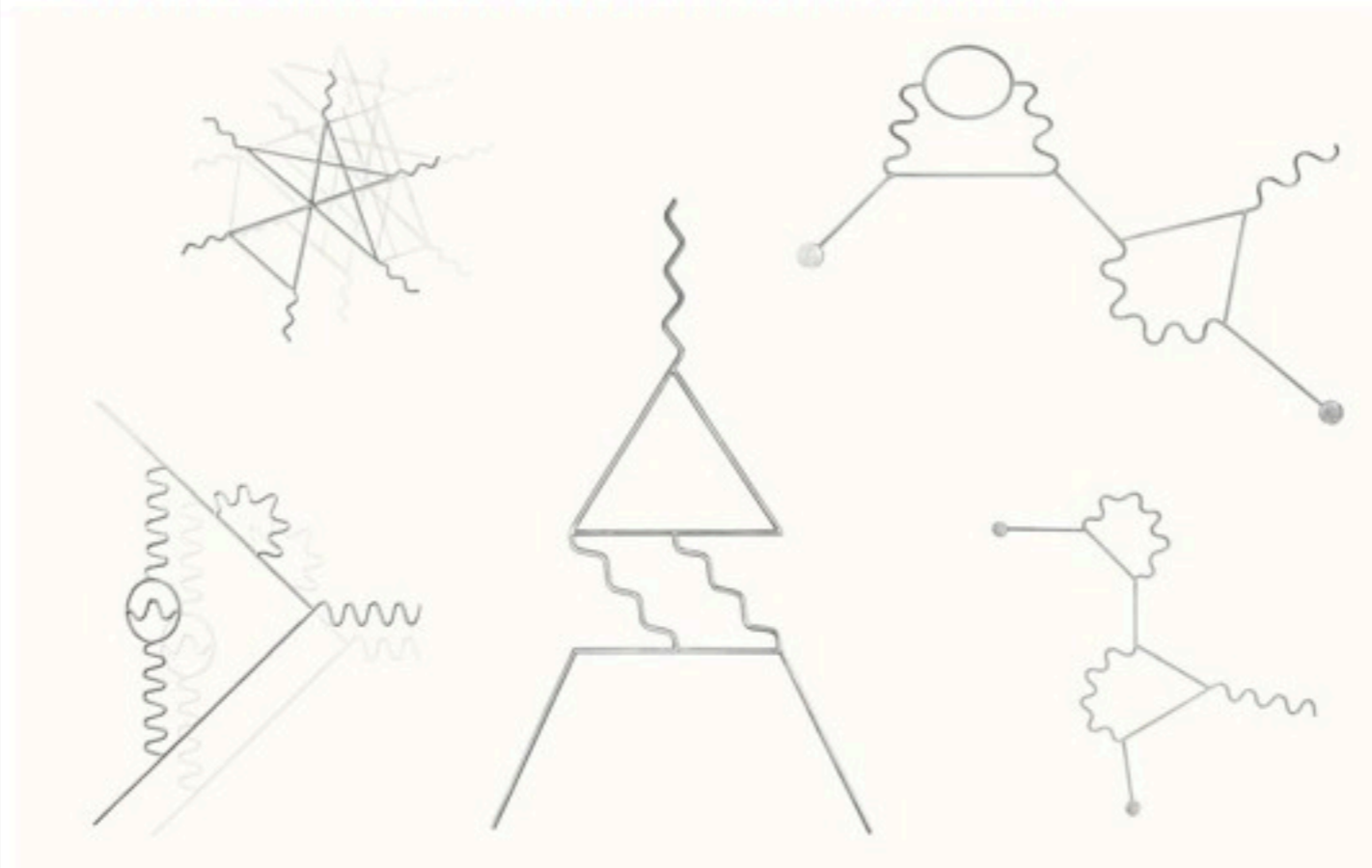
ALL POSSIBLE PHOTONS

The Conceptual and Cognitive Art of Feynman Diagrams

Opening September 15, Saturday 1-7pm

FREE EVENT

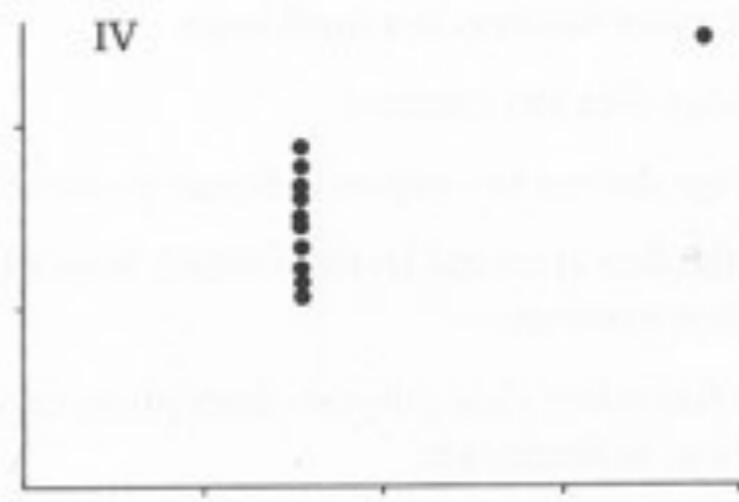
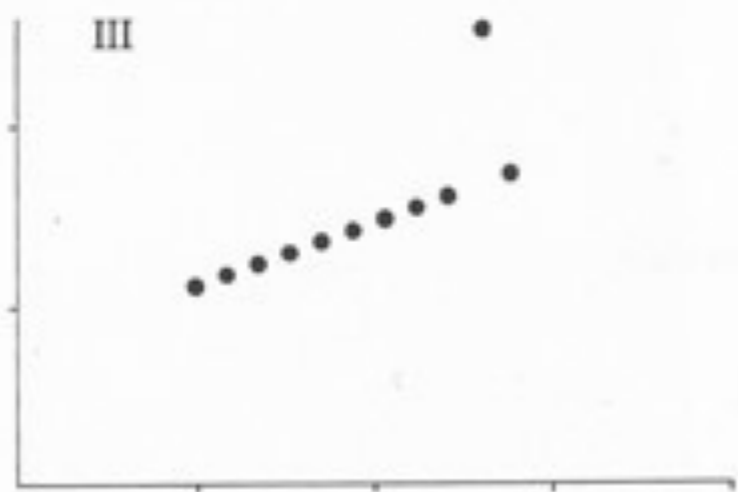
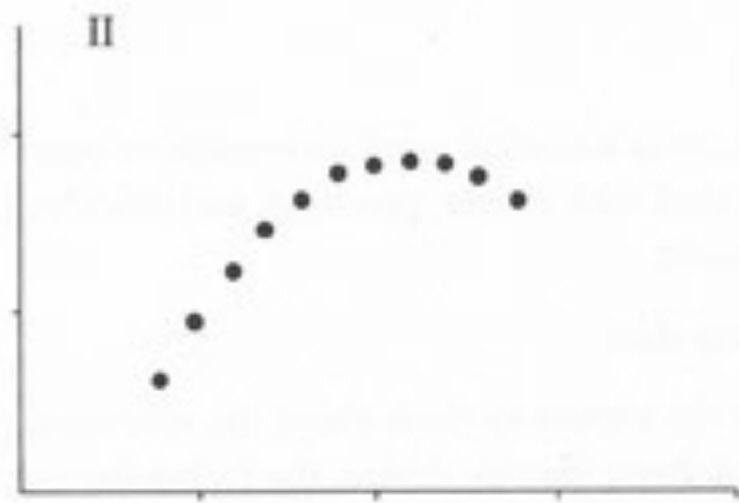
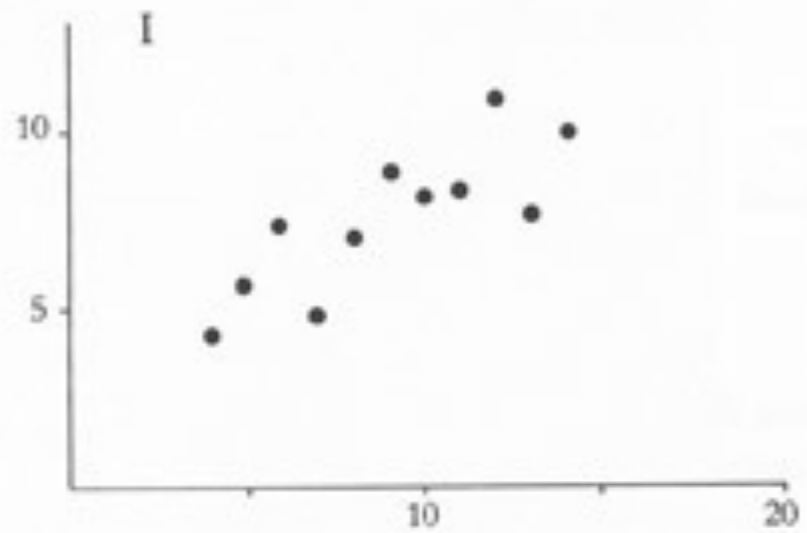
Edward Tufte artist tours at 1.15pm at 3.15pm and at 5.15pm



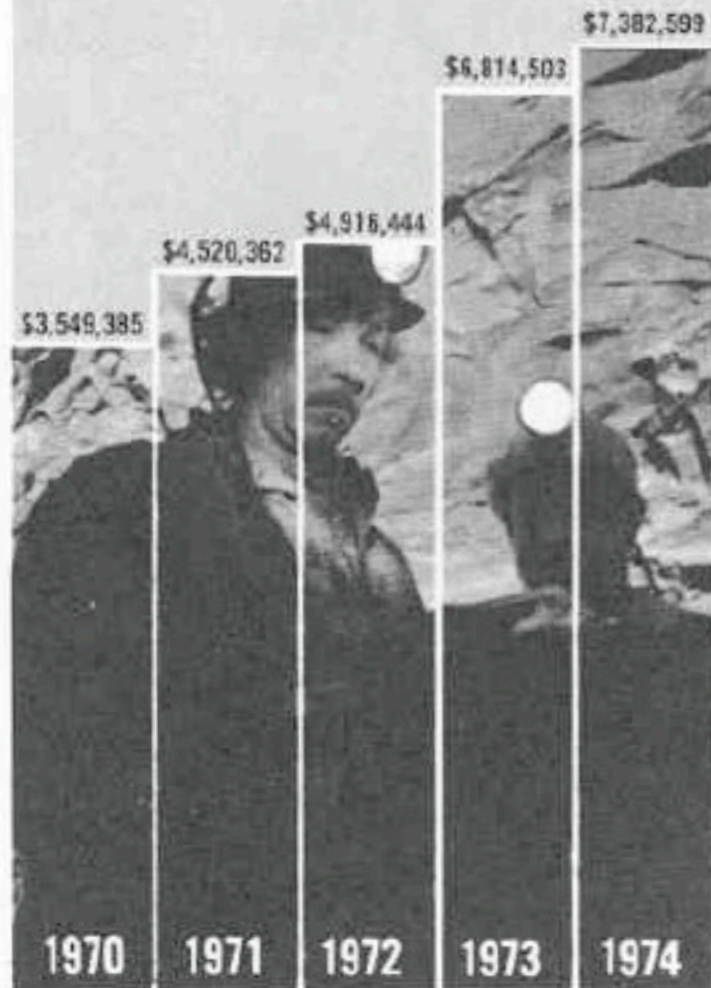
Looking at visualizations...

- what's the question being answered? what's the task this should help you complete?
- what's your immediate impression of the data? what do you *know* at a glance?
- who made this? what is the source of the data?

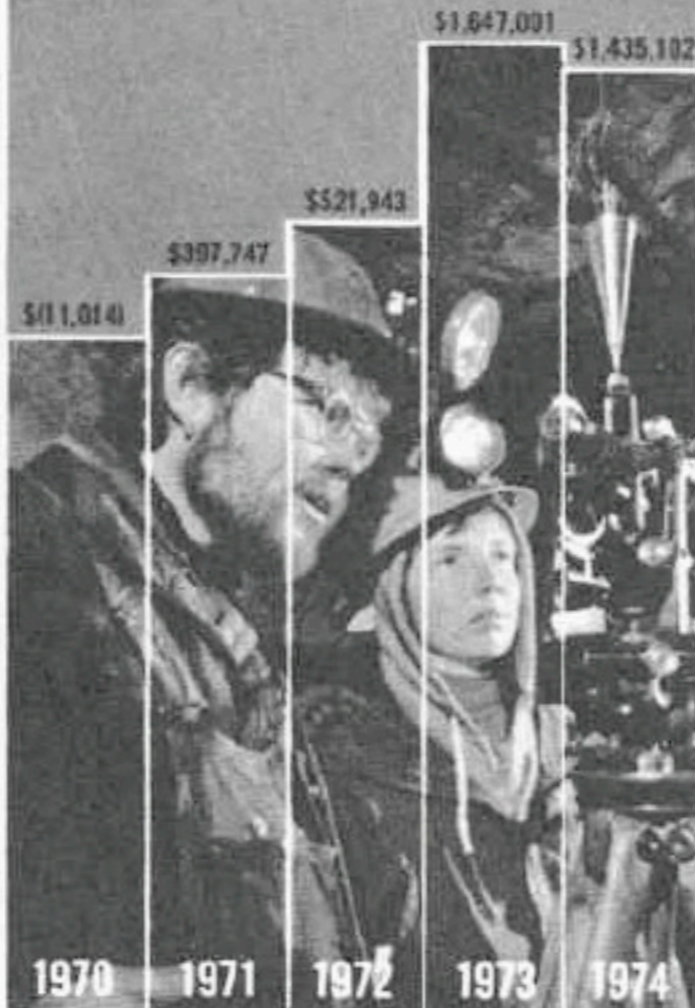
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



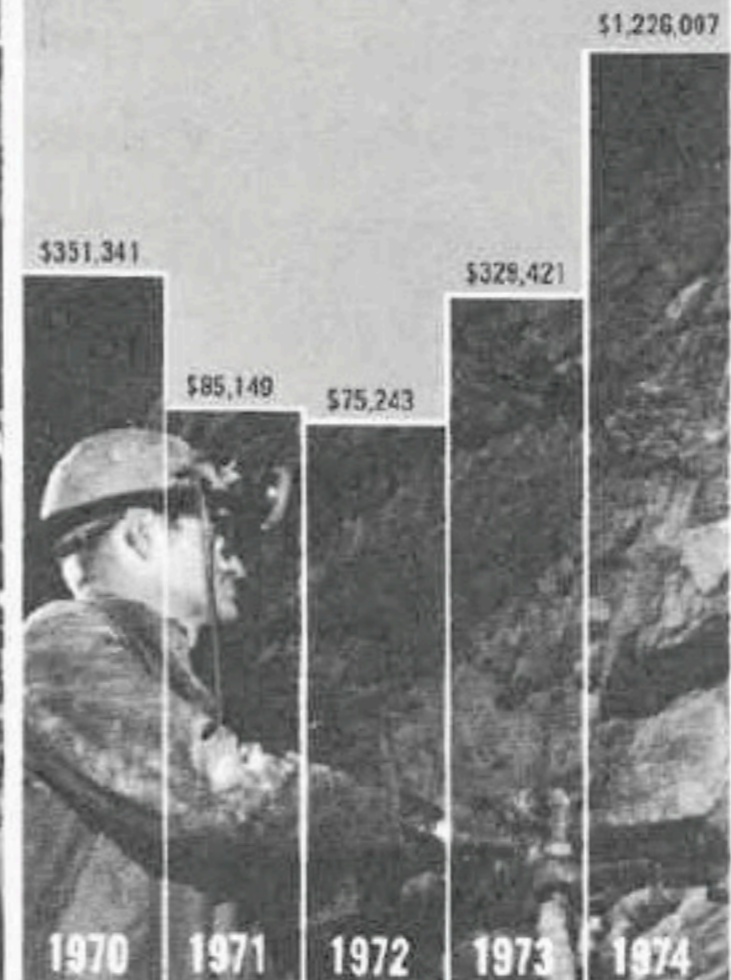
OPERATING REVENUES



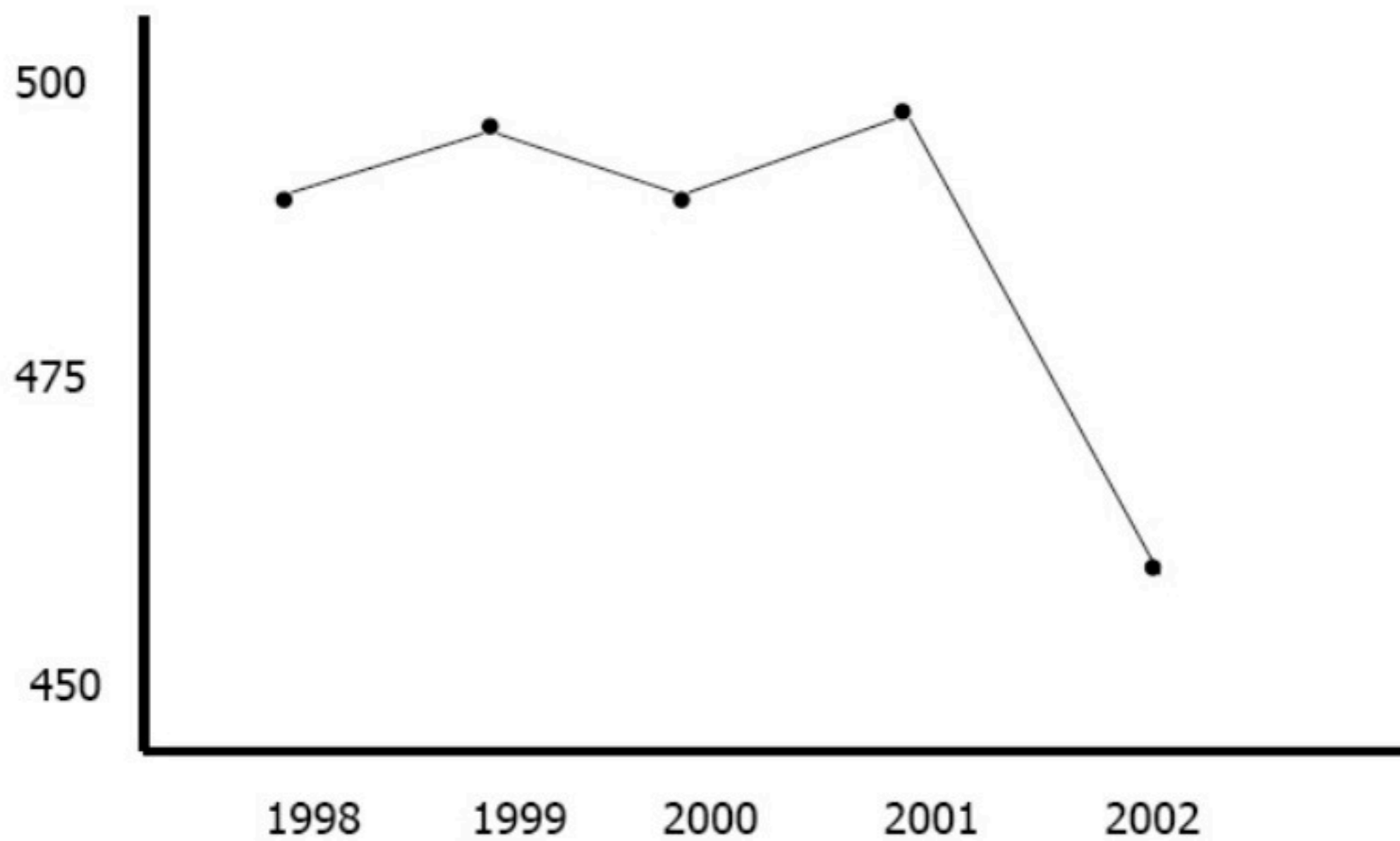
NET INCOME (LOSS)



EXPLORATION & DEVELOPMENT EXPENDITURES

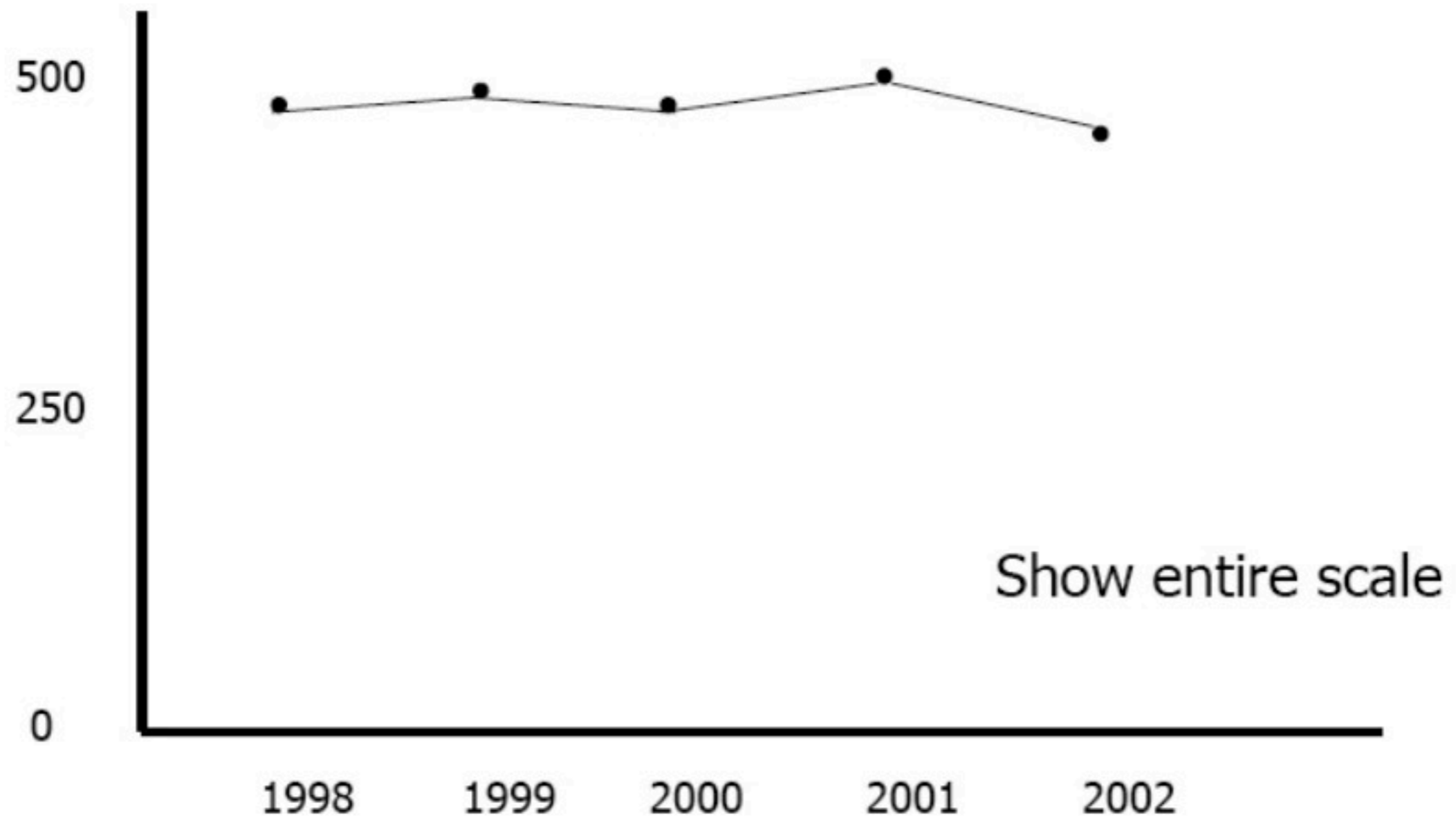


First Impressions?

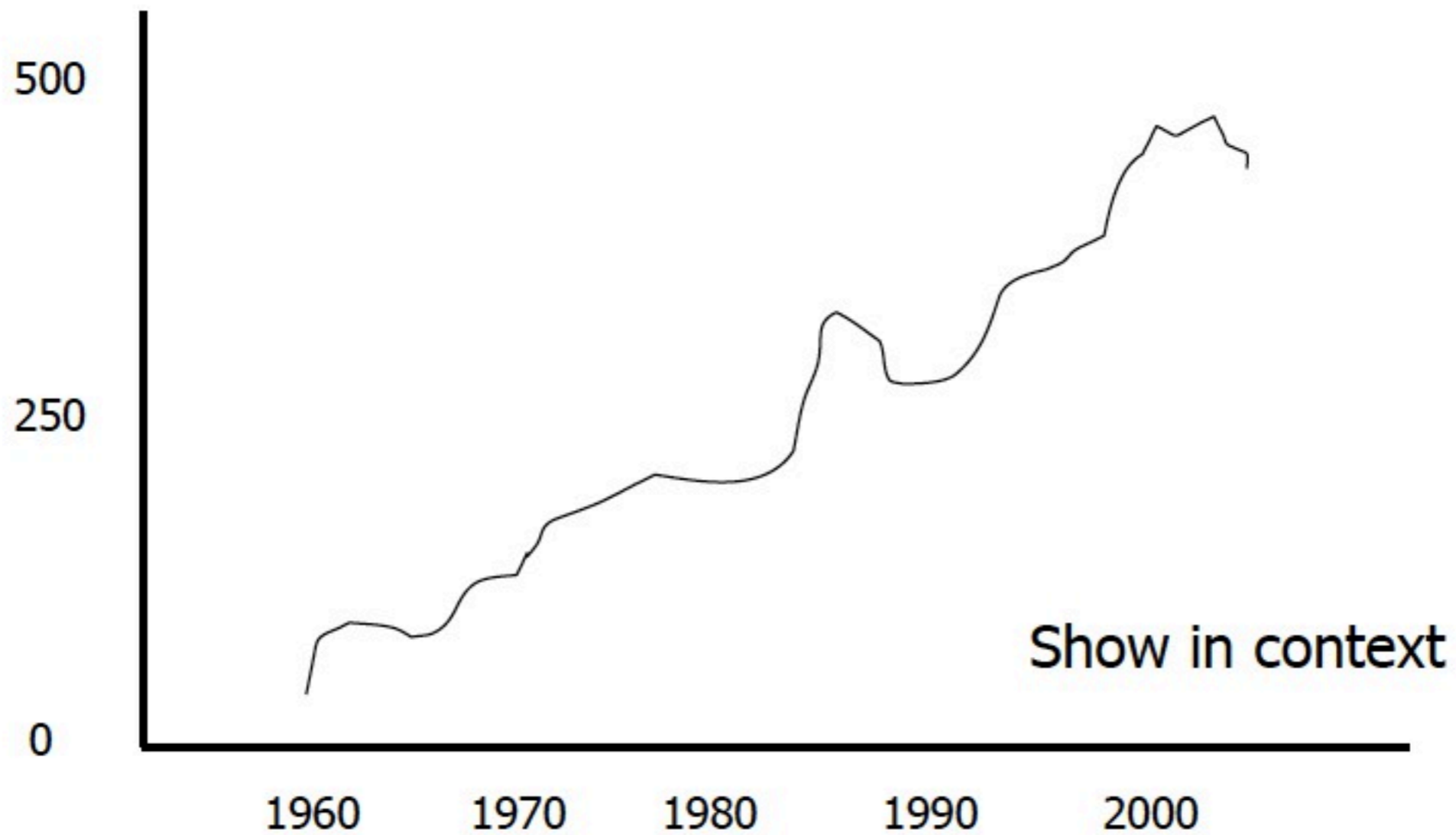


graphic from J Stasko, GA Tech

Start at zero



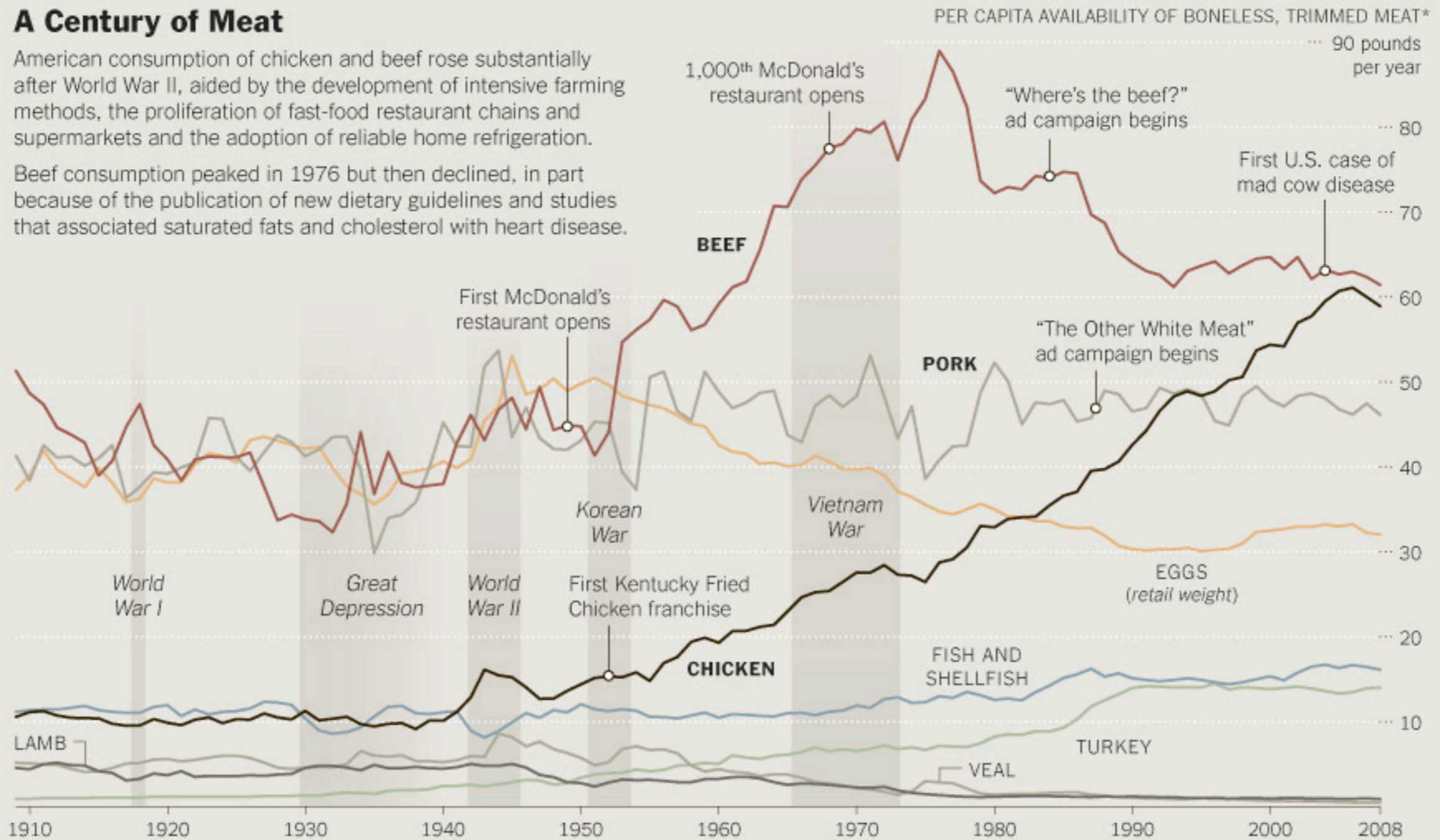
Tell the whole story



A Century of Meat

American consumption of chicken and beef rose substantially after World War II, aided by the development of intensive farming methods, the proliferation of fast-food restaurant chains and supermarkets and the adoption of reliable home refrigeration.

Beef consumption peaked in 1976 but then declined, in part because of the publication of new dietary guidelines and studies that associated saturated fats and cholesterol with heart disease.

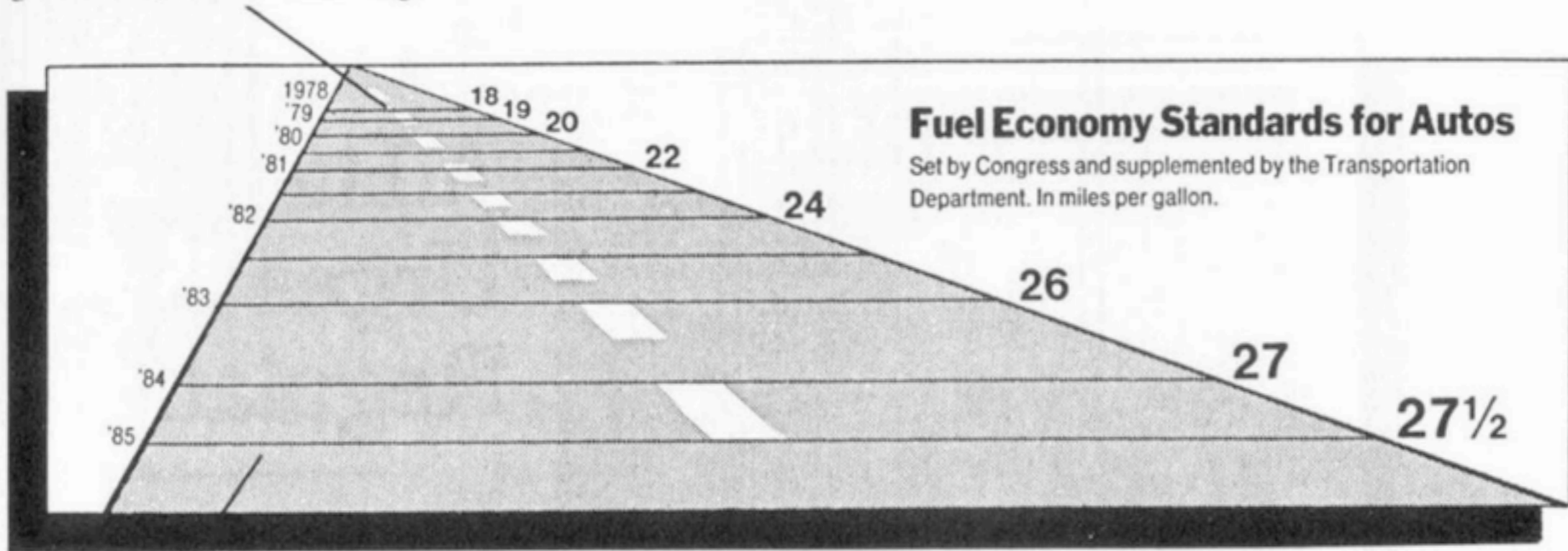


*Note: per capita availability of boneless meat is a proxy for human consumption, and is lower than retail weight or carcass weight. Bones, offal and game are excluded.

Sources: U.S. Department of Agriculture (data); news and company reports; "Putting Meat on the American Table," by Roger Horowitz

JONATHAN CORUM/THE NEW YORK TIMES

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



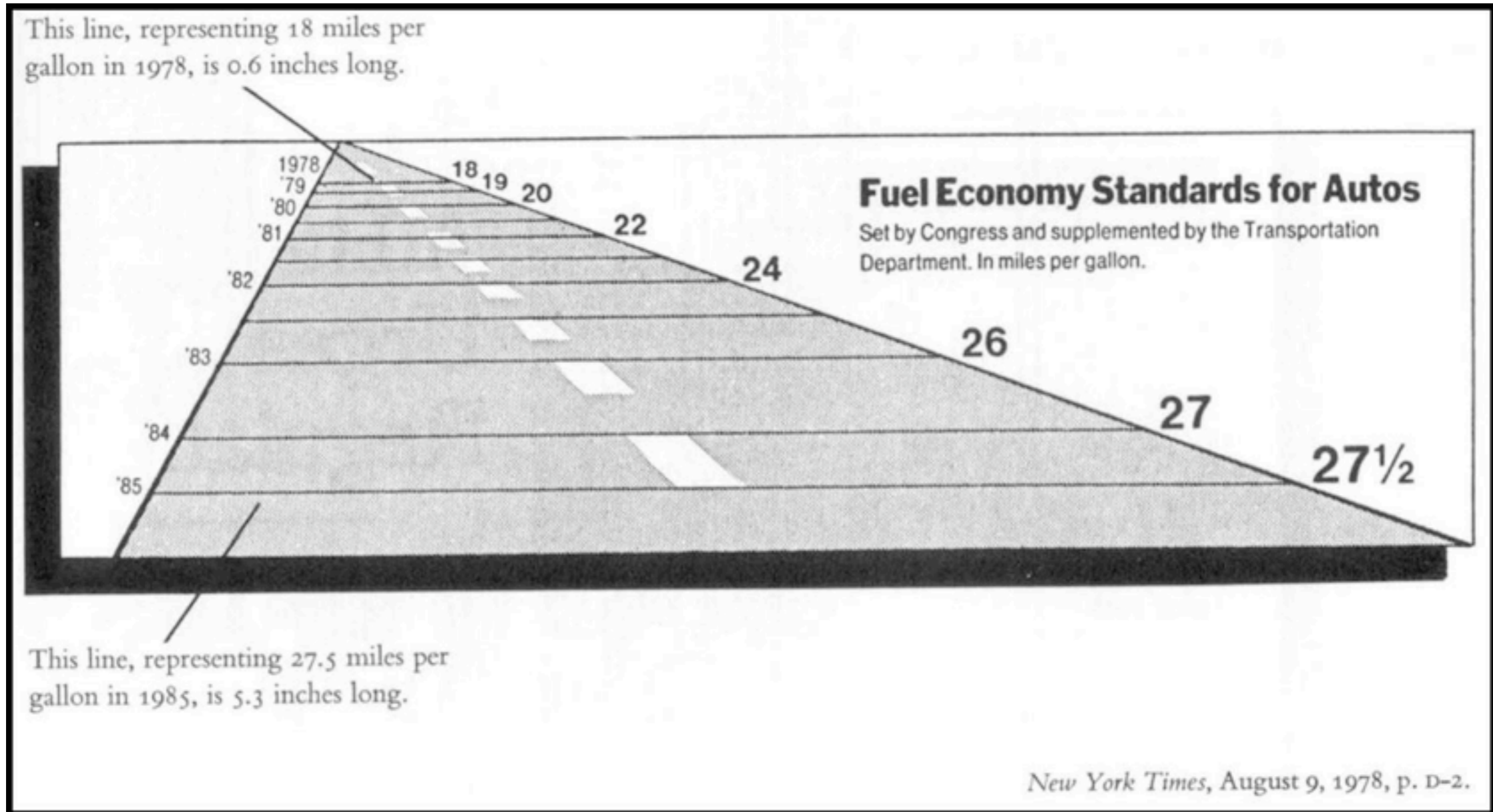
Fuel Economy Standards for Autos

Set by Congress and supplemented by the Transportation Department. In miles per gallon.

This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Data Distortion

$$\text{Tufte's "Lie Factor"} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$



in the data: $27.5 - 18 / 18.0 * 100 = 53\%$ change
 in the graphic: $5.3 - 0.6 / 0.6 * 100 = 783\%$ change

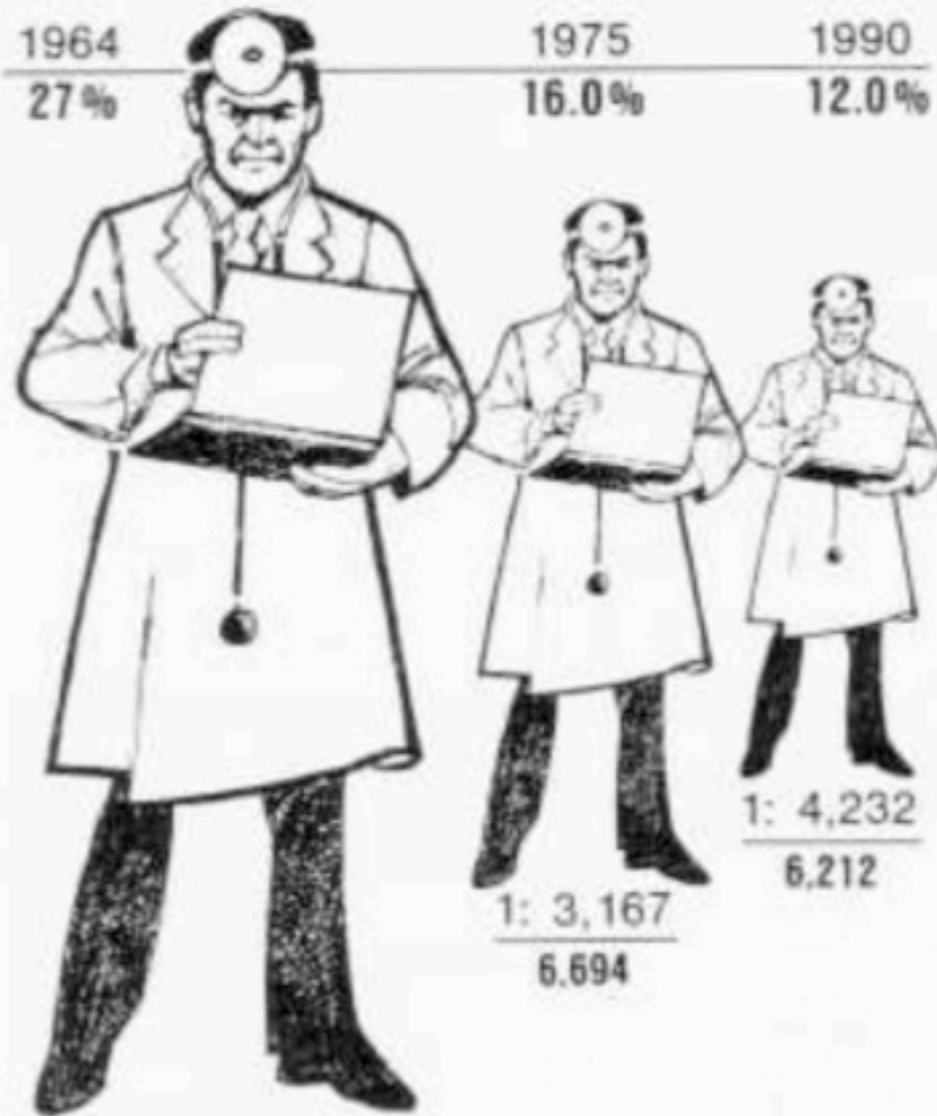
783/53 = Lie Factor of 14.8

THE SHRINKING FAMILY DOCTOR

In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



1: 2,247 RATIO TO POPULATION
8.023 Doctors

1: 3,167
6,694

1: 4,232
6,212

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

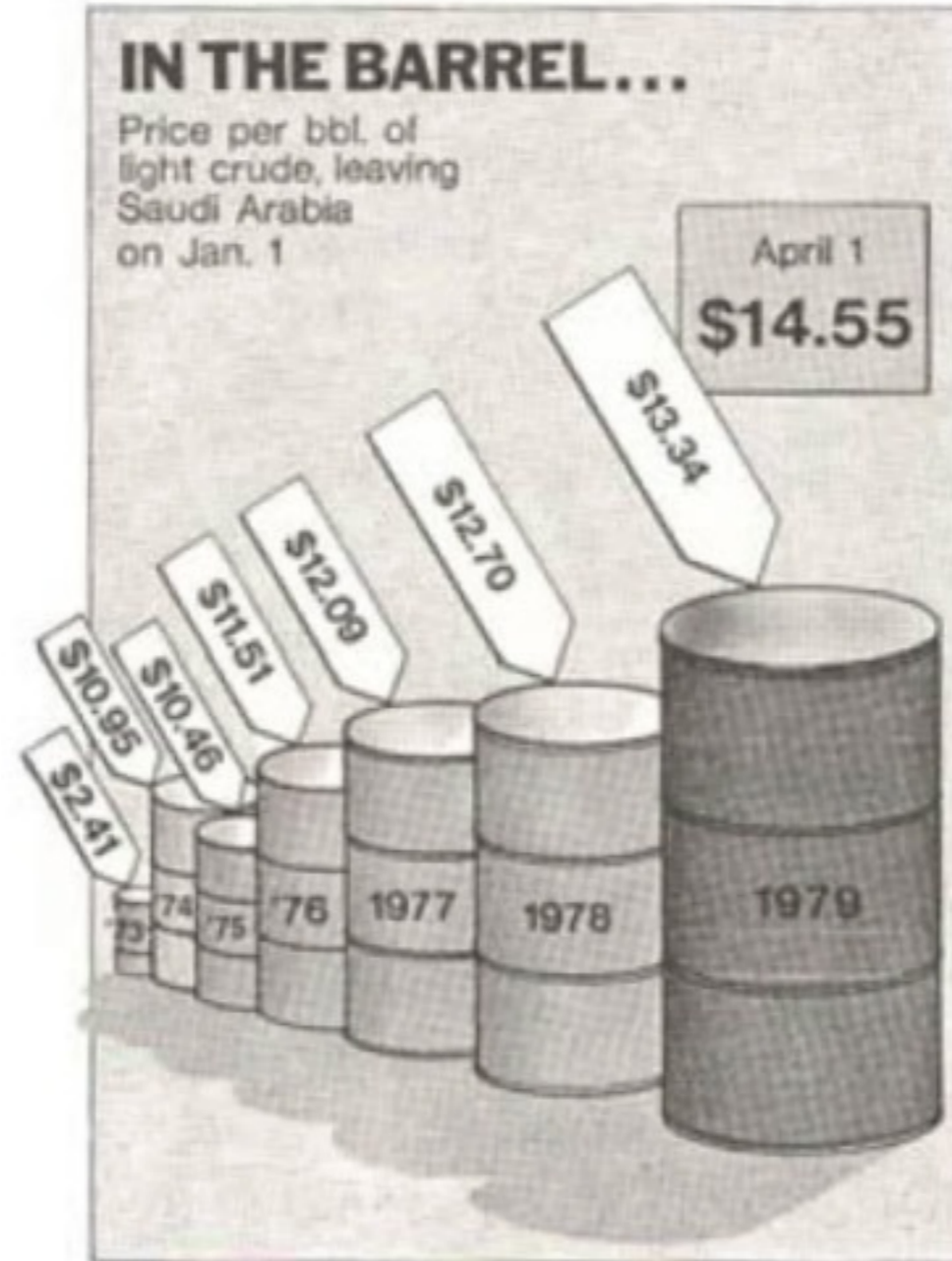
1964	1975	1990
27%	16.0%	12.0%



- distorted size coding— area
- large area for single point of data
- figures in diminishing perspective
- skewed horizontal scale

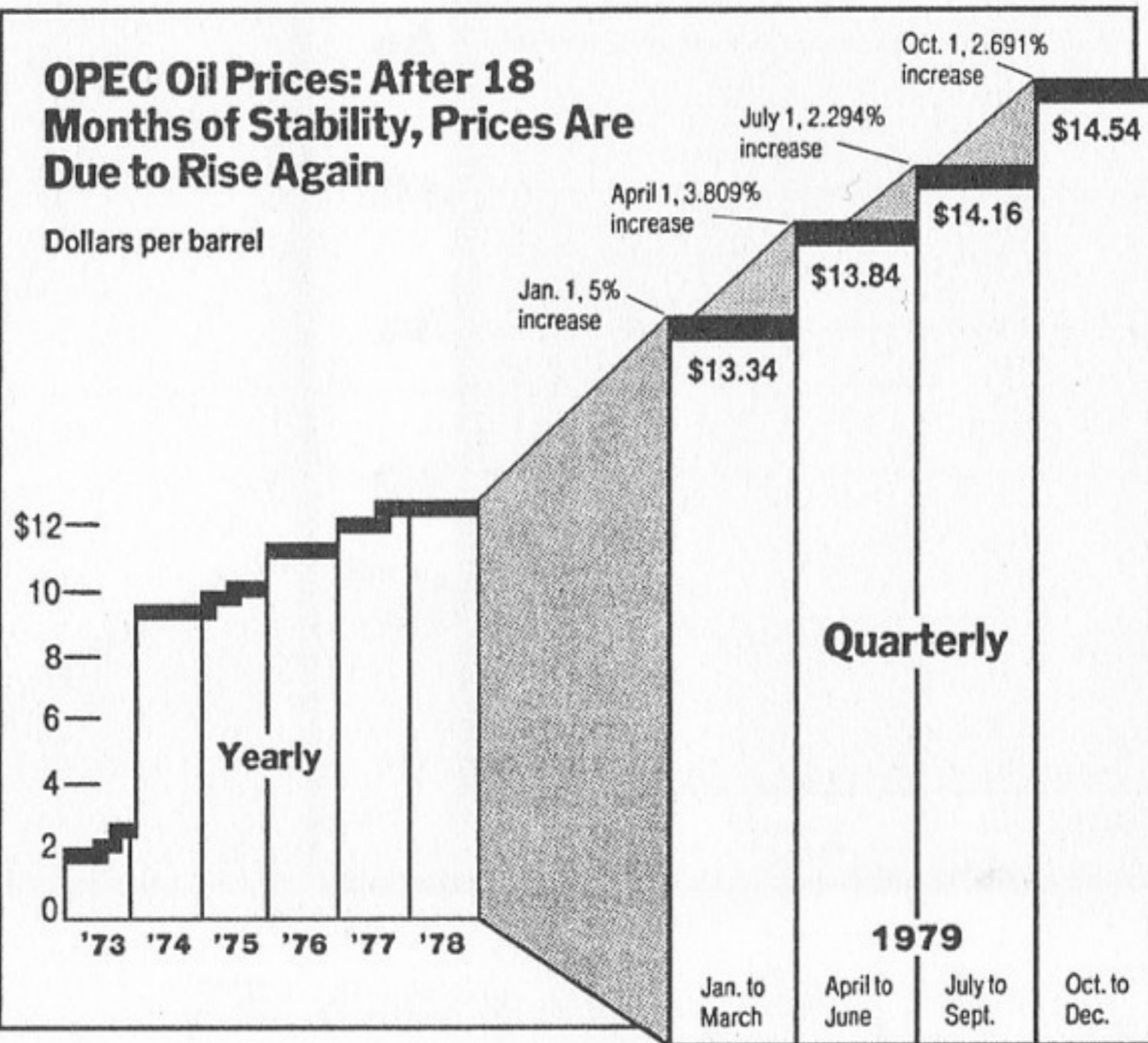
“Lie Factor” = 2.8

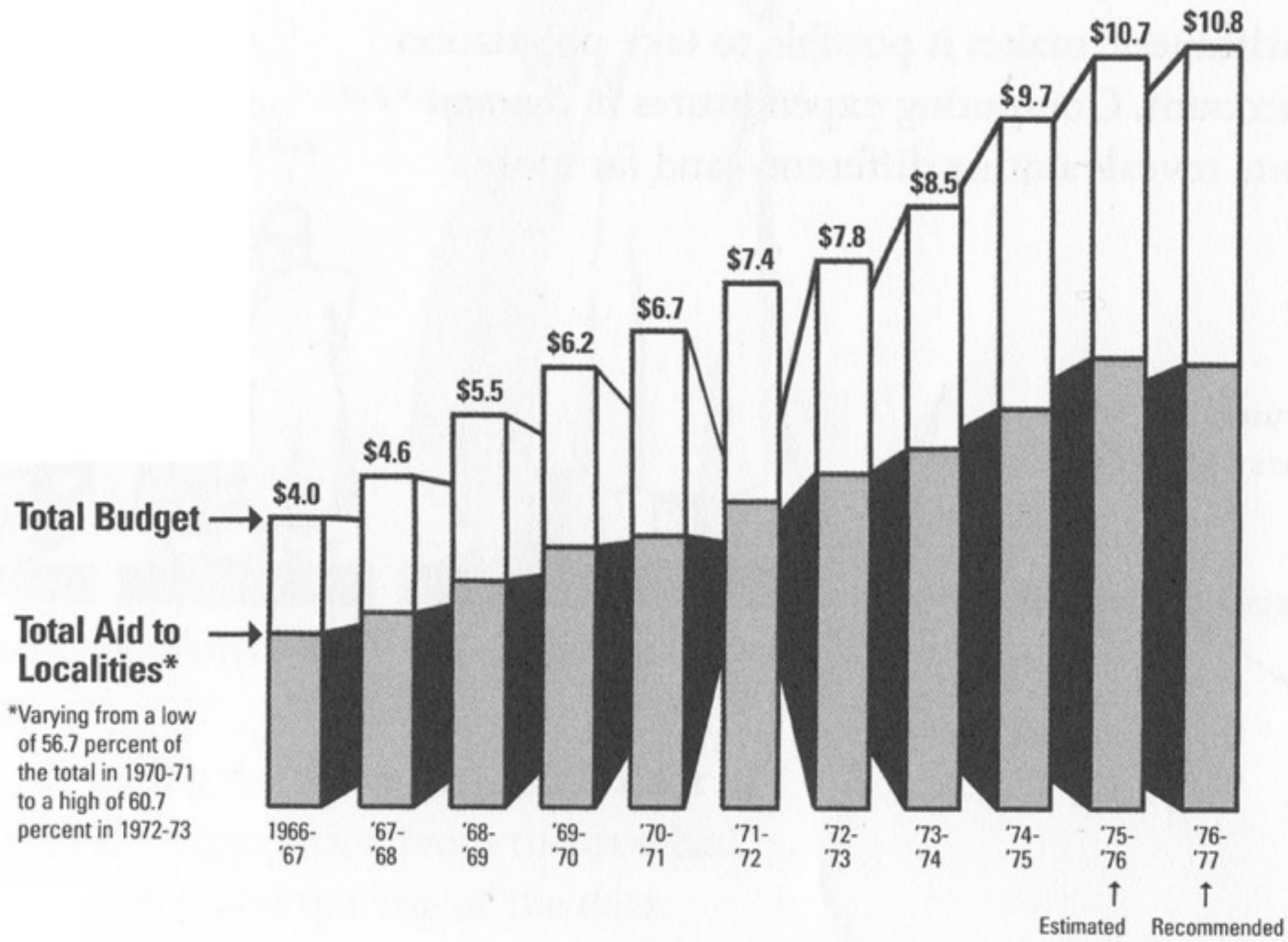
More Distortions...



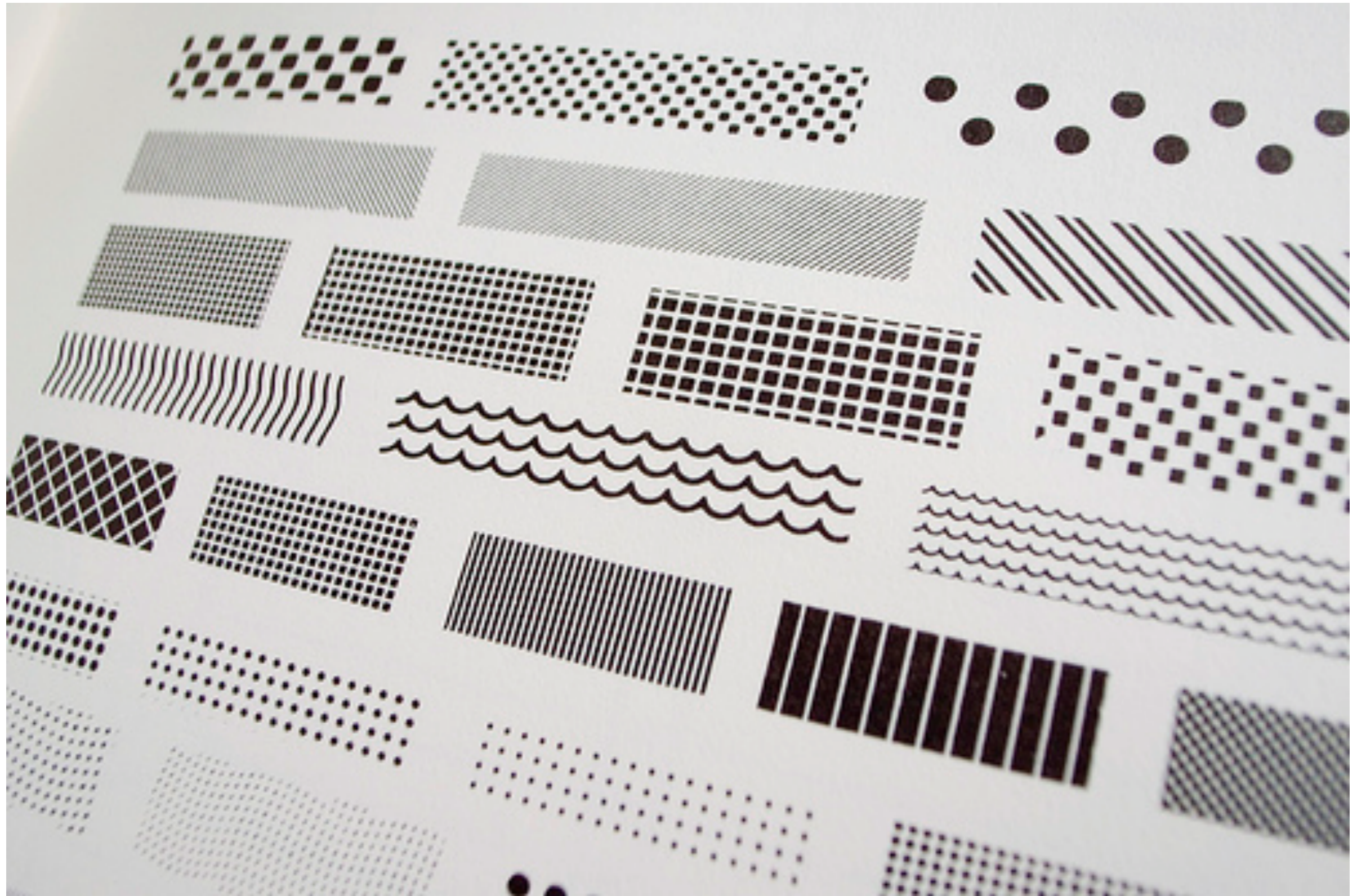
OPEC Oil Prices: After 18 Months of Stability, Prices Are Due to Rise Again

Dollars per barrel

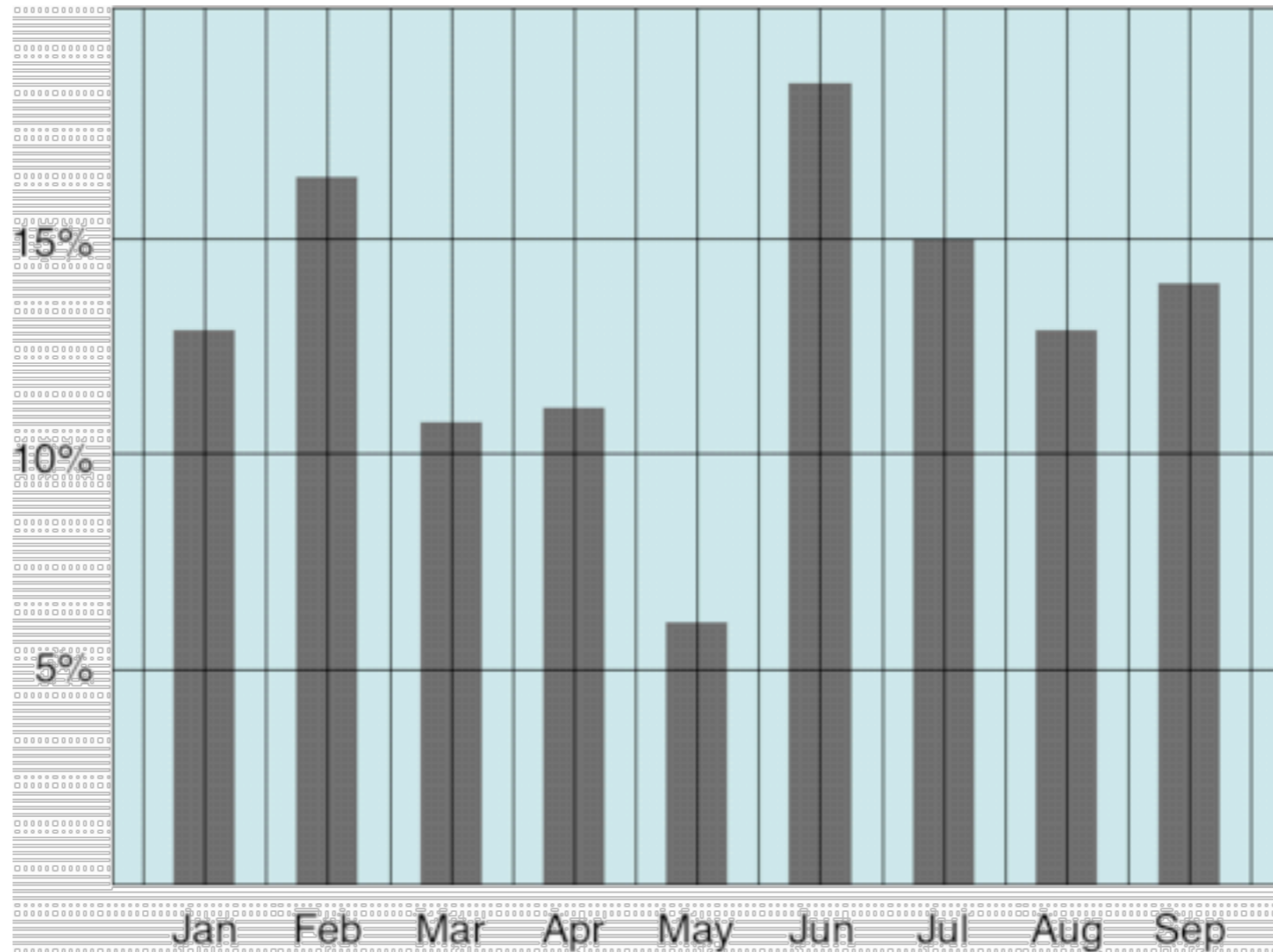




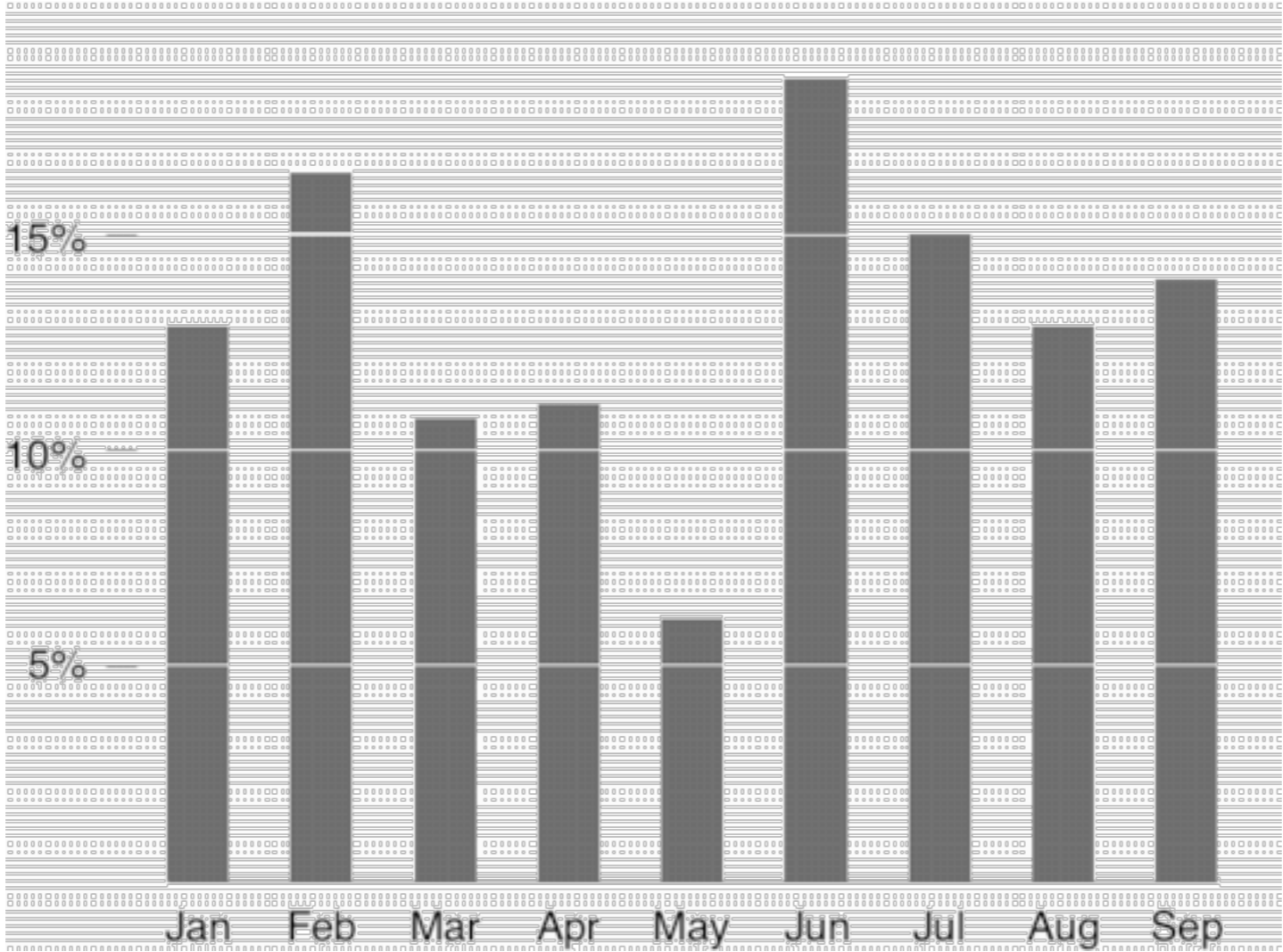
Data-Ink vs. Non-Data Ink

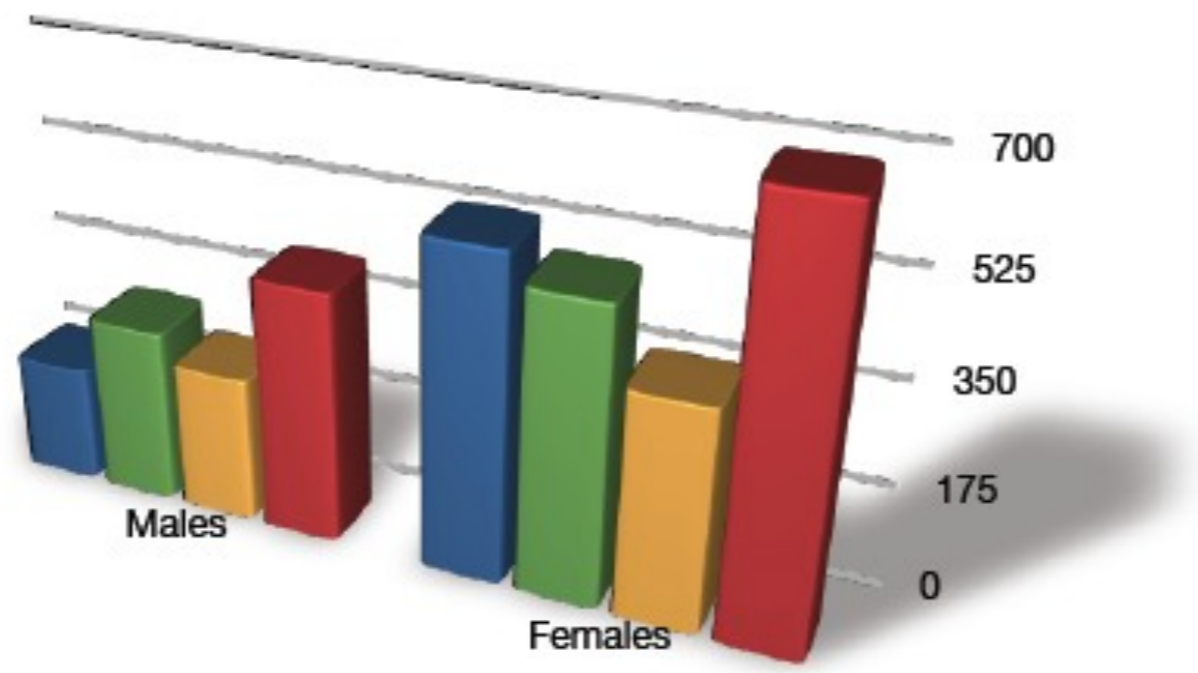


Data-ink ratio = data-ink / total ink used



Less Ink, More Data

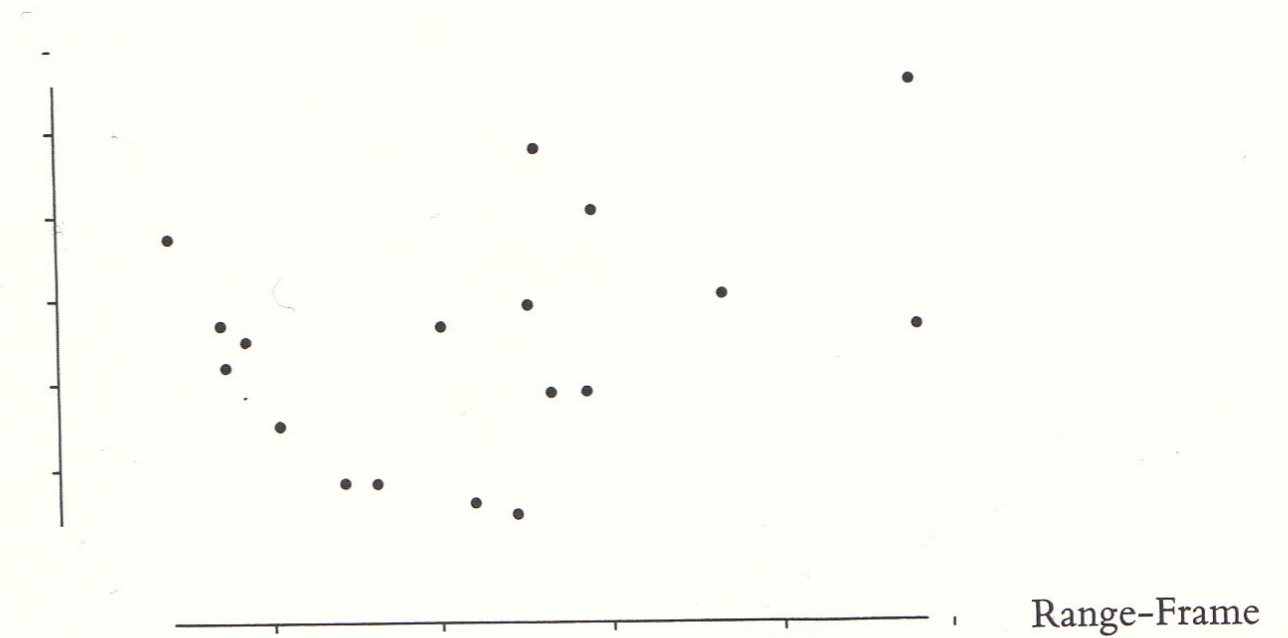
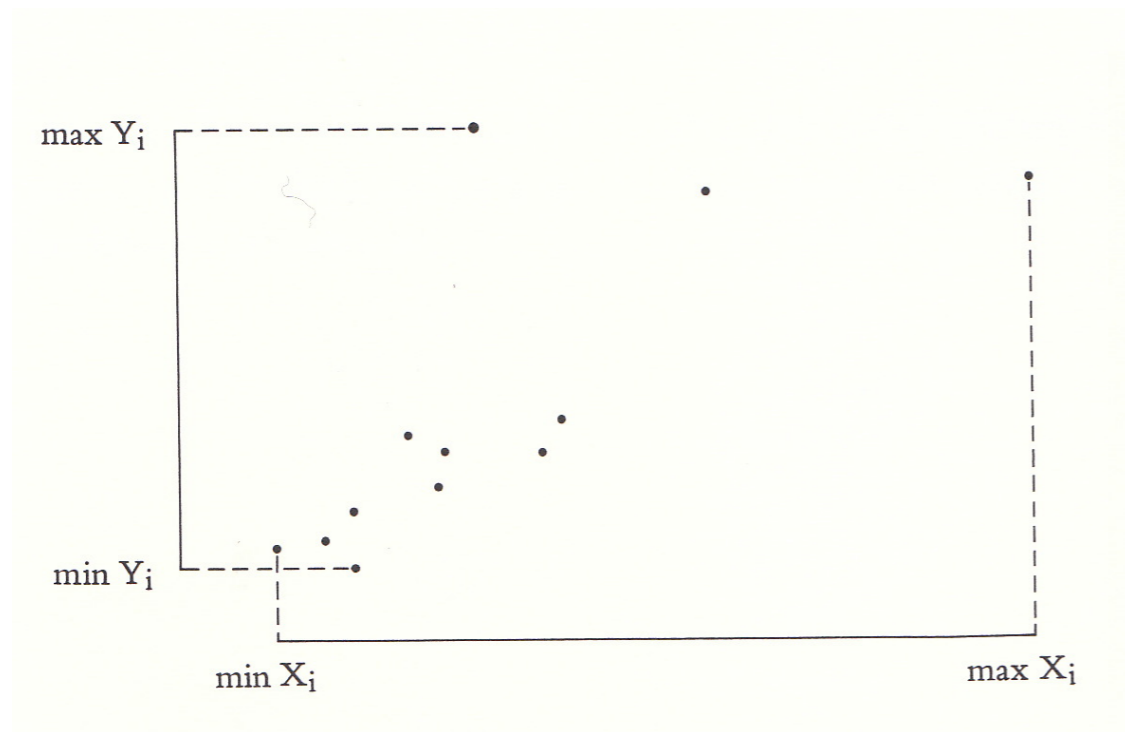
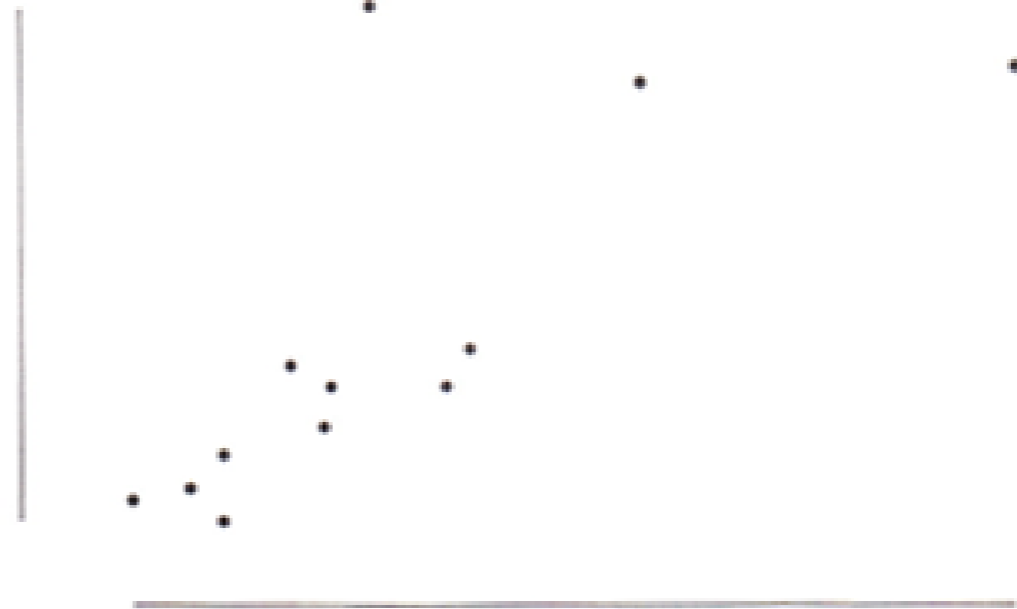
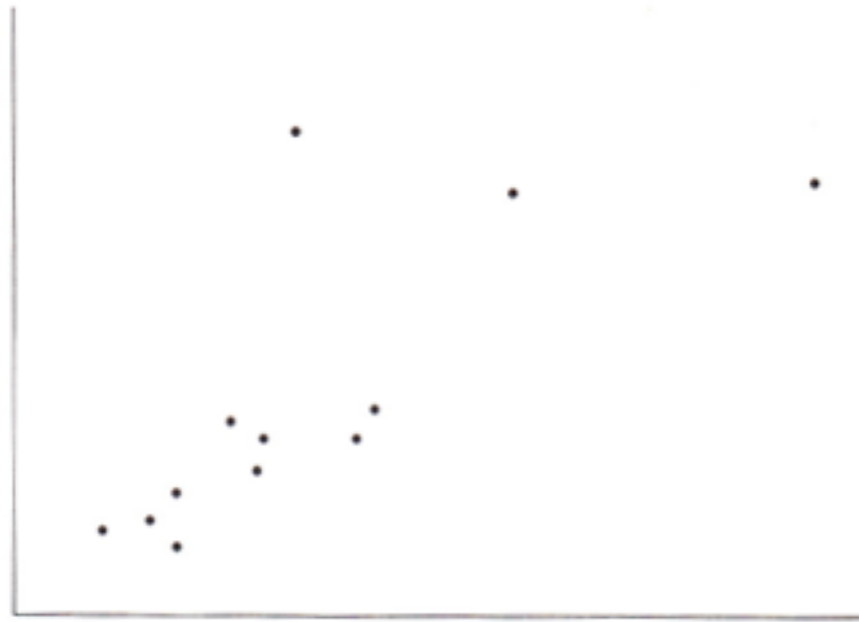


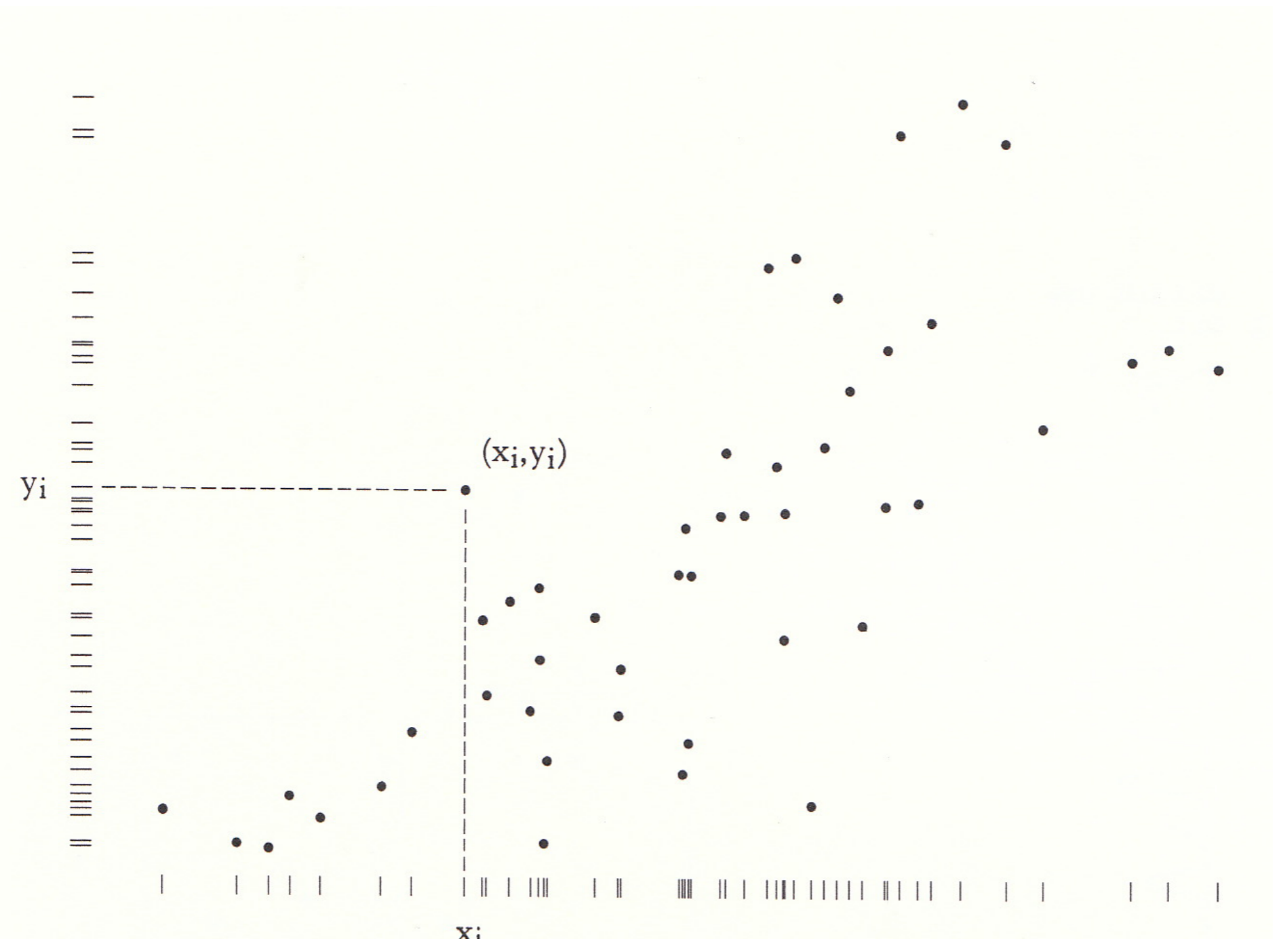


■ 0-\$24,999 ■ \$25,000+ ■ 0-\$24,999 ■ \$25,000+

Redesign of the Scatterplot

Consider the standard bivariate scatterplot:





The fringe of dashes in the dot-dash-plot can connect a series of bivariate scatters in a *rugplot* (since it resembles a set of fringed rugs—and covers the statistical ground):

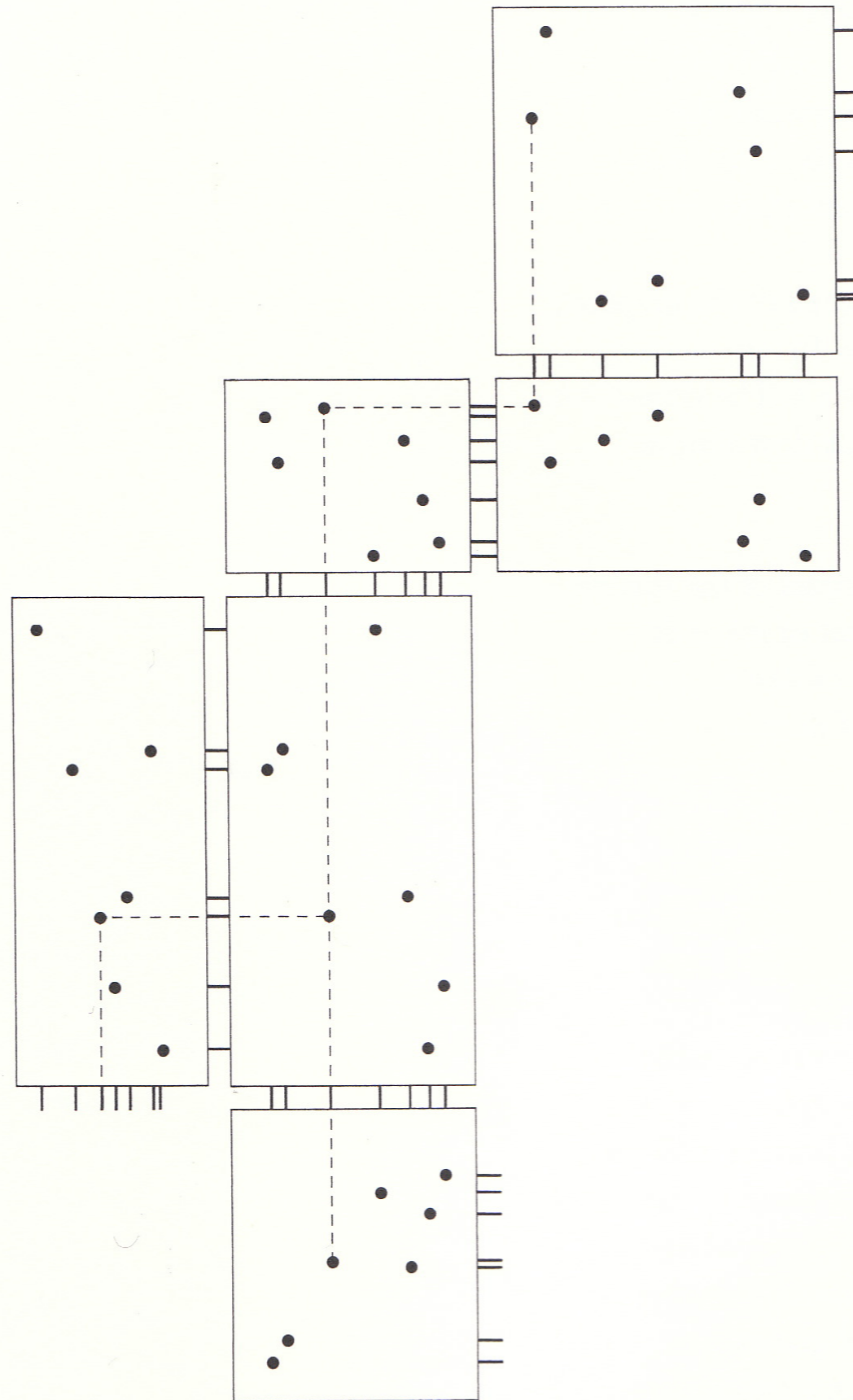
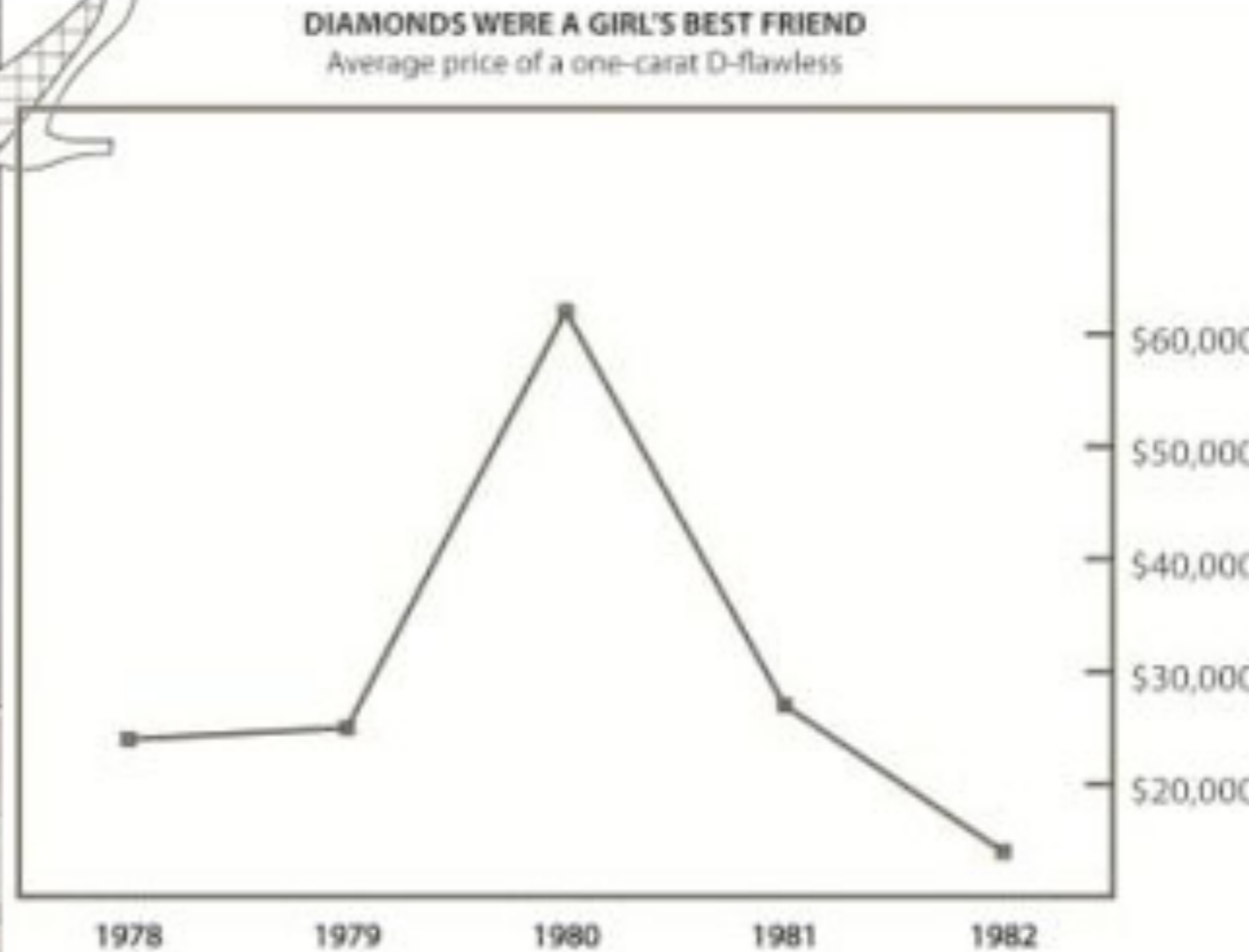
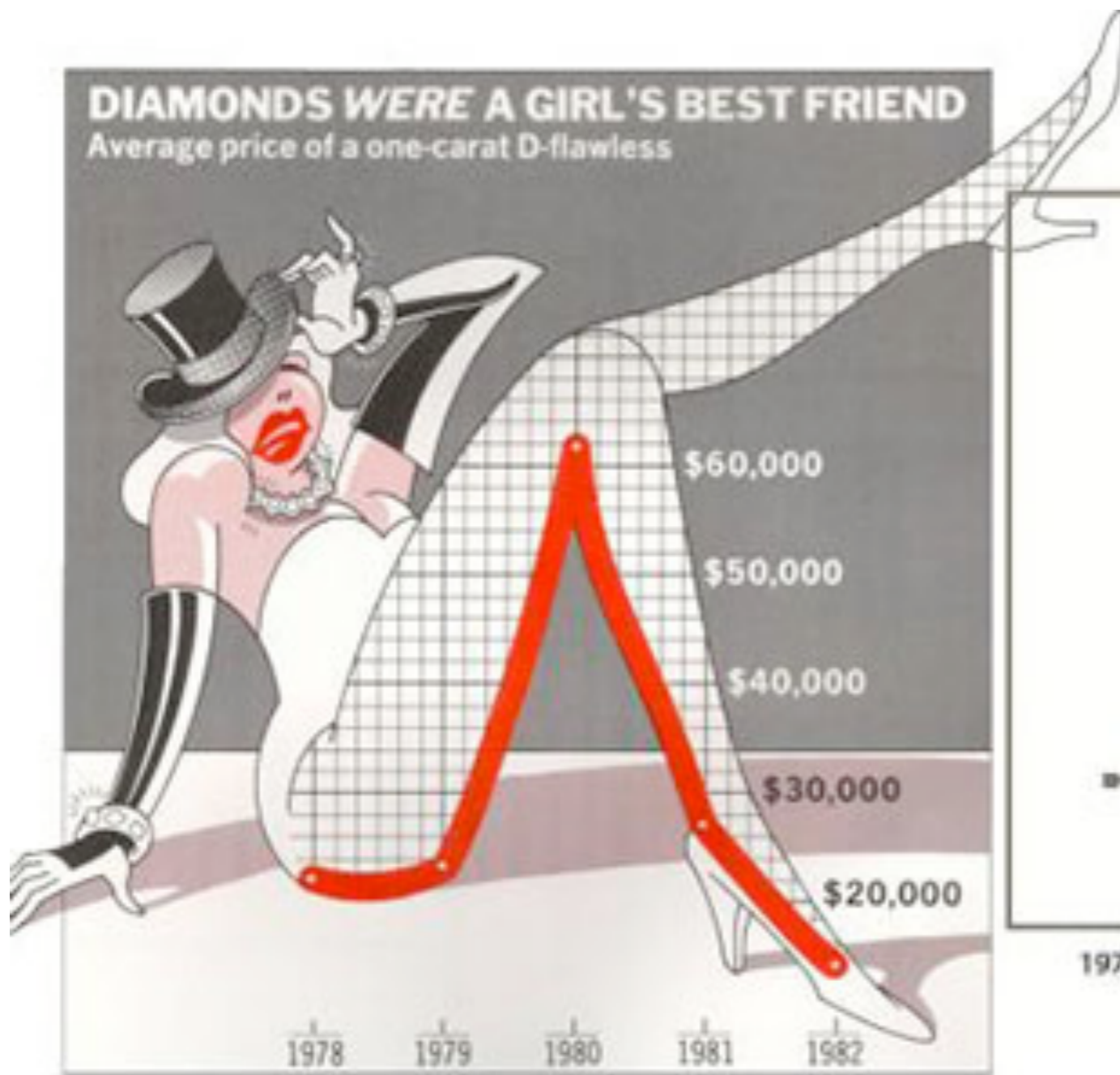


Chart Junk – no data, just frill

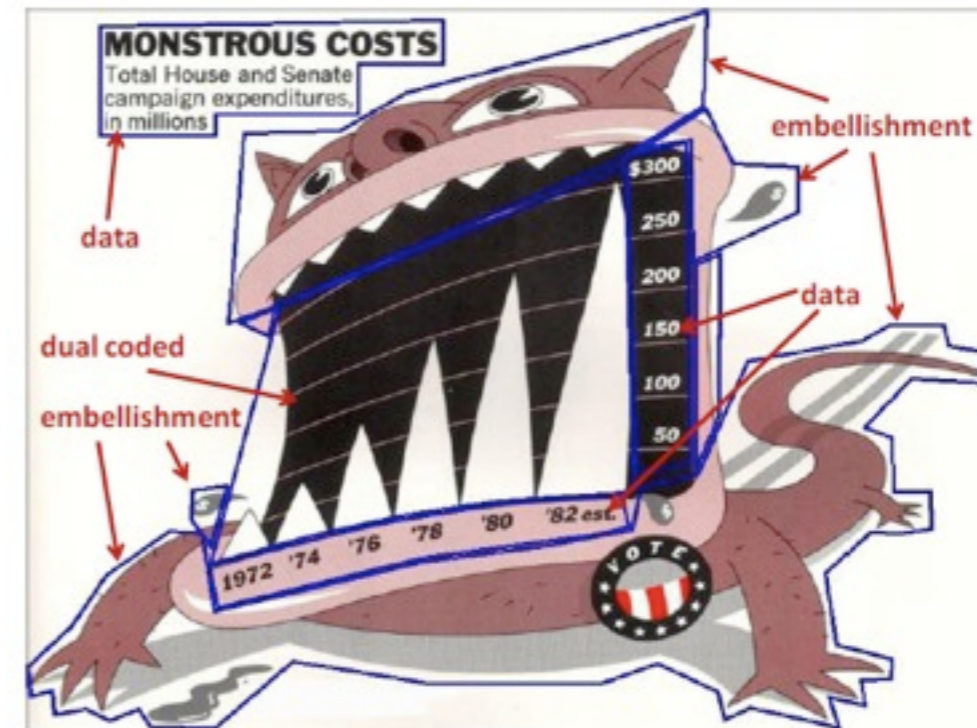


Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

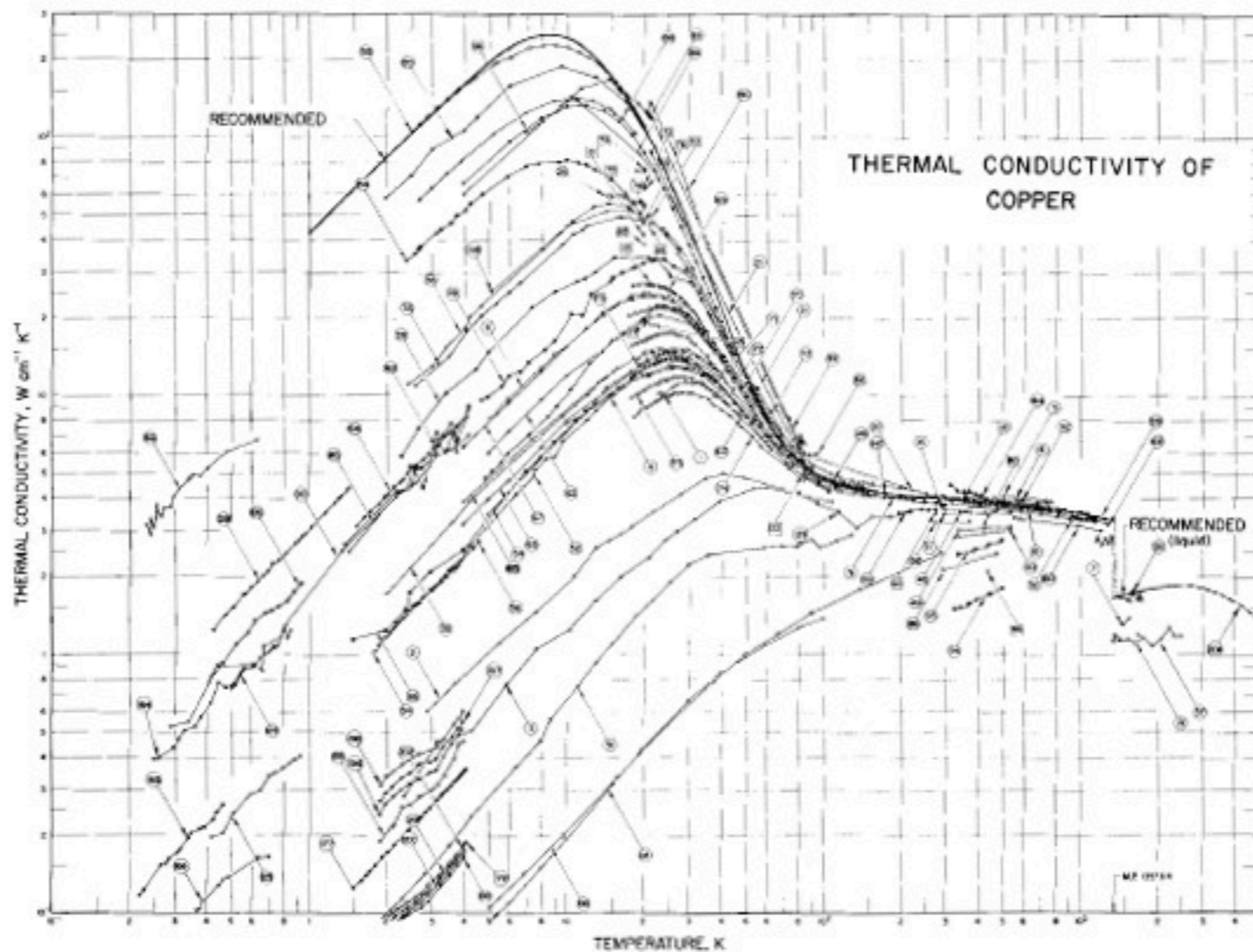
Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

From the abstract:

... We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.



$$\text{Data Density} = \frac{\text{Number of entries in the data array}}{\text{Area of data graphic}}$$

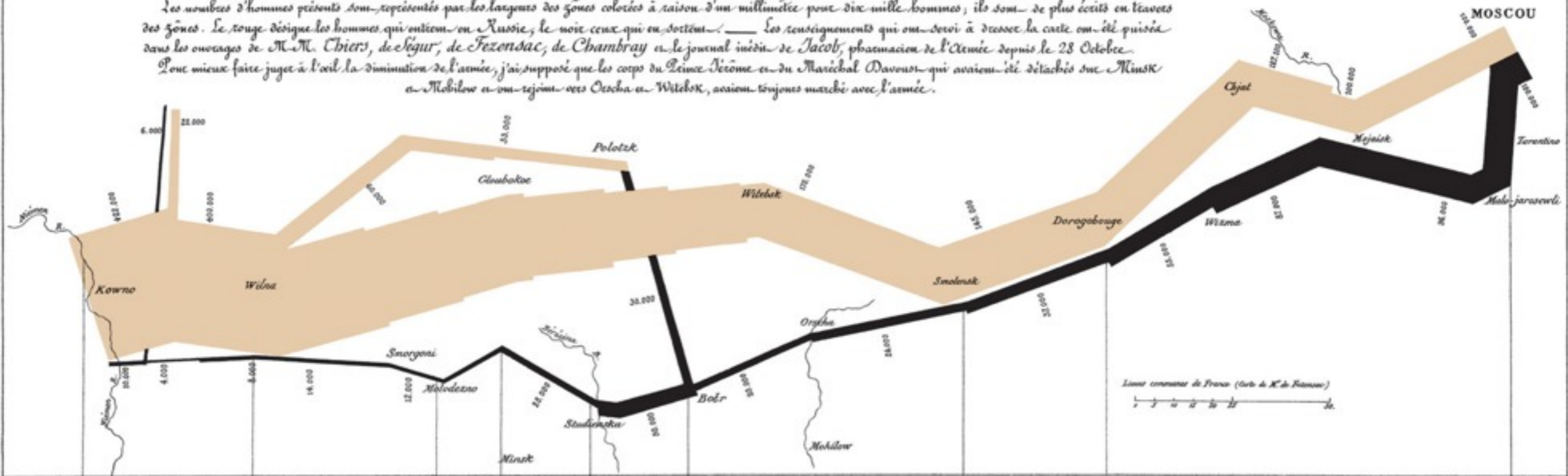


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

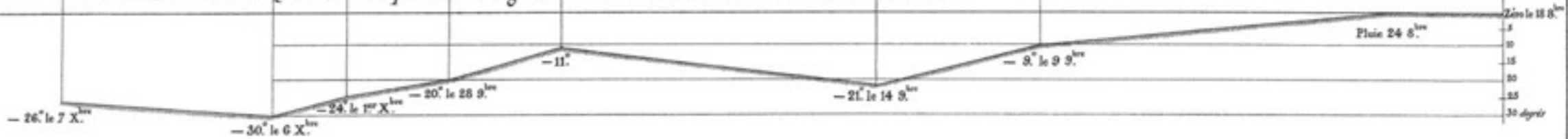
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont péri en Russie, le noir ceux qui en sont restés. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilew et qui avaient rejoint vers Oescha et Witebsk, avaient toujours marché avec l'armée.



Lignes convenues de France (Carte de M. de France)
 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

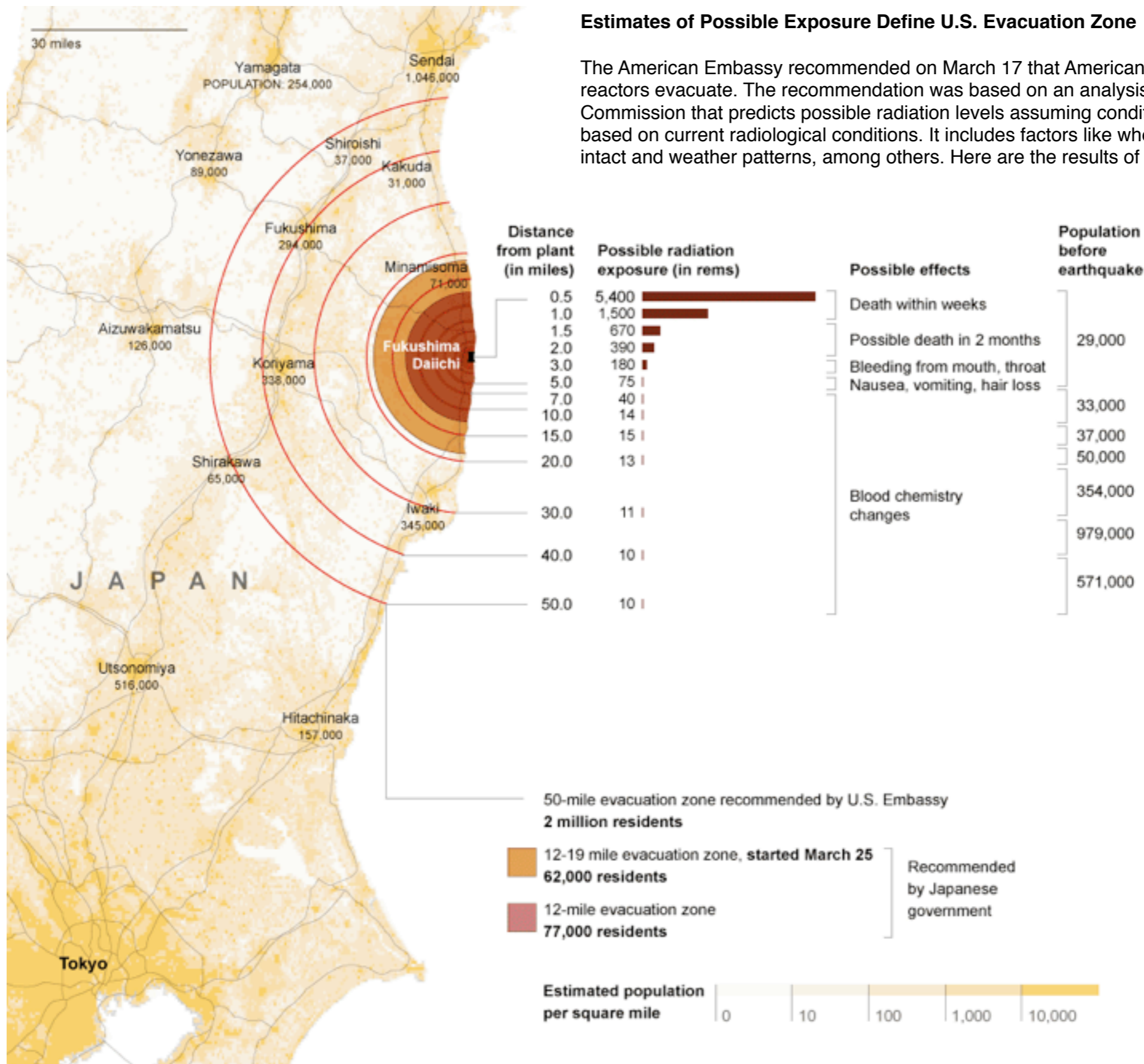


Les Cosaques passent au galop le Niémen gelé.

Joseph Minard
 Map of Napoleon's Russian 1812 Campaign (1869)

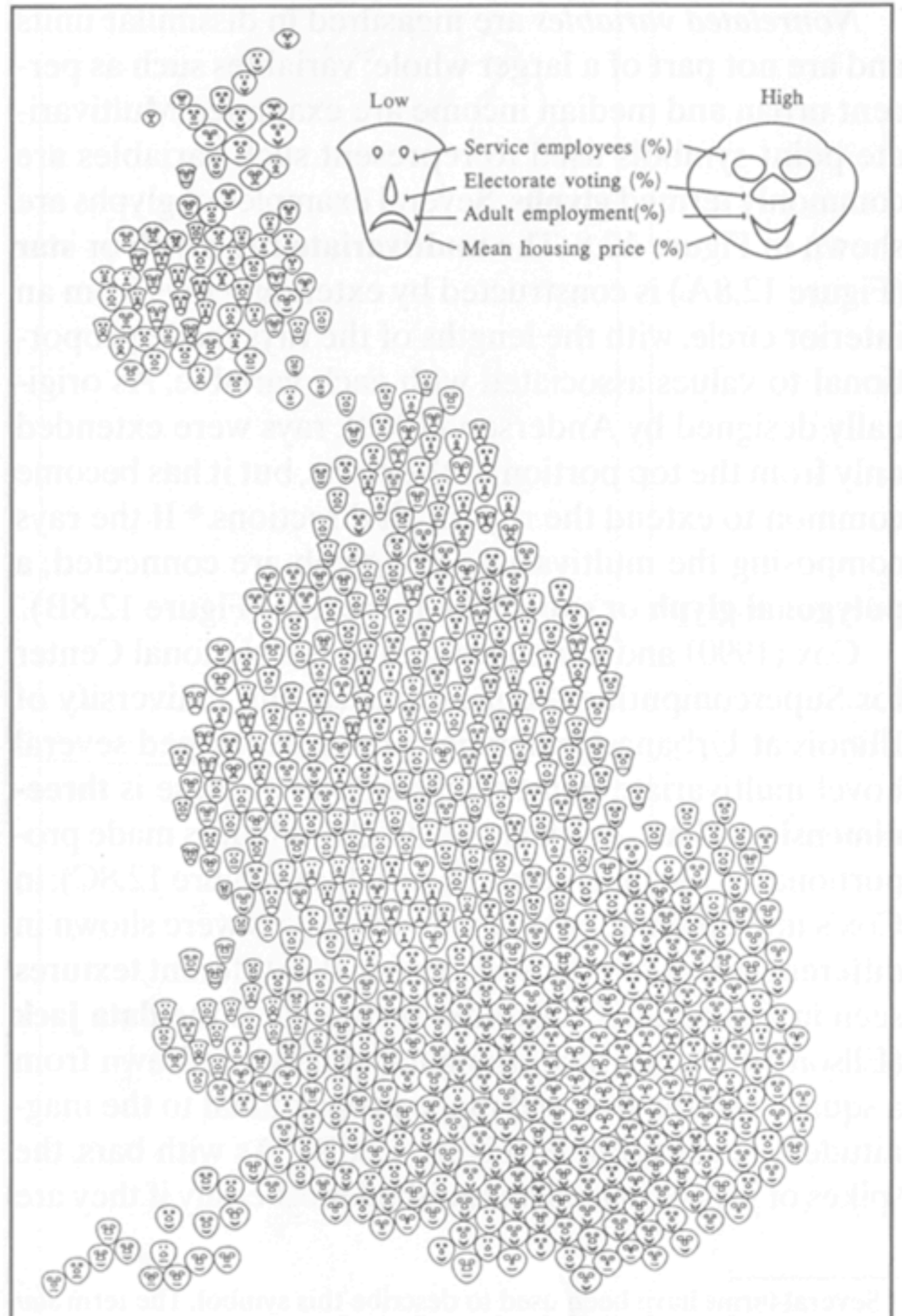
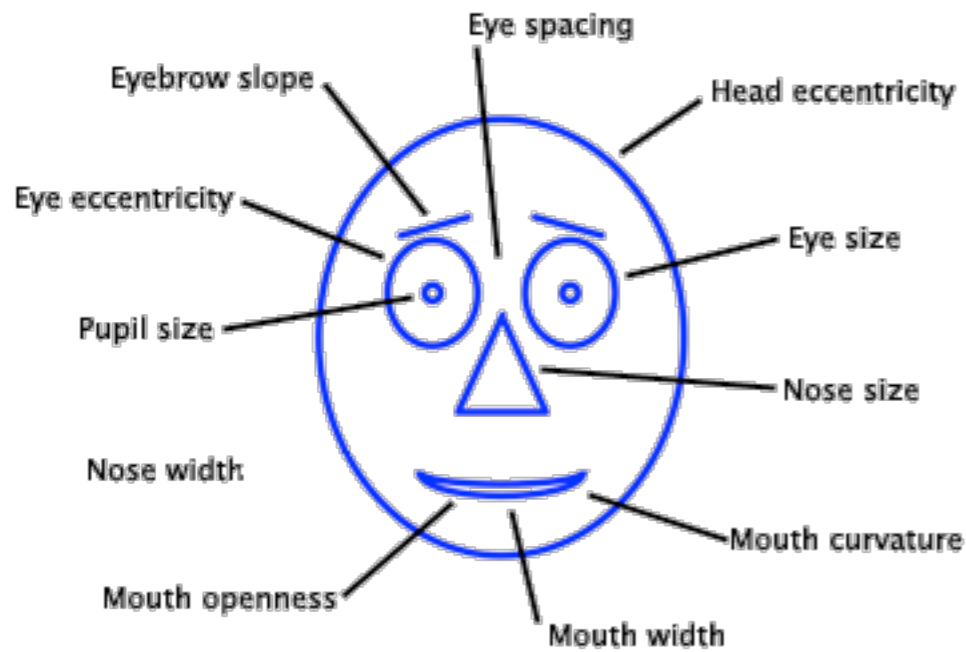
Estimates of Possible Exposure Define U.S. Evacuation Zone

The American Embassy recommended on March 17 that Americans within 50 miles of the Fukushima reactors evacuate. The recommendation was based on an analysis by the Nuclear Regulatory Commission that predicts possible radiation levels assuming conditions at the plant degrade. It is not based on current radiological conditions. It includes factors like whether containment vessels remain intact and weather patterns, among others. Here are the results of the analysis on March 16.



data density

(chernoff faces)





Manny Acta
Washington Nationals
.451 season winning percentage



Buddy Bell
Kansas City Royals
.426



Bud Black
San Diego Padres
.546



Bruce Bochy
San Francisco Giants
.438



Bobby Cox
Atlanta Braves
.519



Terry Francona
Boston Red Sox
.593



Ron Gardenhire
Minnesota Twins
.488



Phil Garner
Houston Astros
.443



Bob Geren
Oakland Athletics
.469



John Gibbons
Toronto Blue Jays
.512



Fredi González
Florida Marlins
.438



Ozzie Guillen
Chicago White Sox
.444



Clint Hurdle
Colorado Rockies
.552



Tony La Russa
St. Louis Cardinals
.481



Jim Leyland
Detroit Tigers
.543



Grady Little
Los Angeles Dodgers
.506



Pete Mackanin
Cincinnati Reds
.513



Joe Maddon
Tampa Bay Devil Rays
.407



Charlie Manuel
Philadelphia Phillies
.549



John McLaren
Seattle Mariners
.512



Bob Melvin
Ariz. Diamondbacks
.556



Lou Piniella
Chicago Cubs
.525



Willie Randolph
New York Mets
.543



Mike Scioscia
L.A. Angels of Anaheim
.580



Joe Torre
New York Yankees
.580



Jim Tracy
Pittsburgh Pirates
.420



Dave Trembley
Baltimore Orioles
.430



Ron Washington
Texas Rangers
.463



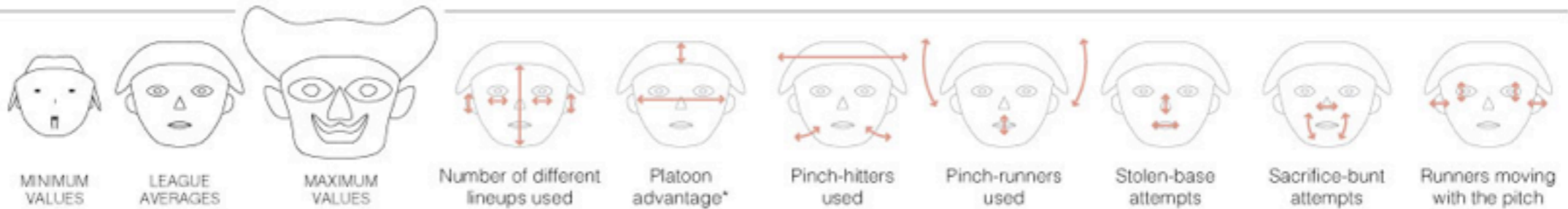
Eric Wedge
Cleveland Indians
.593



Ned Yost
Milwaukee Brewers
.512

SMILE IF YOU BUNT

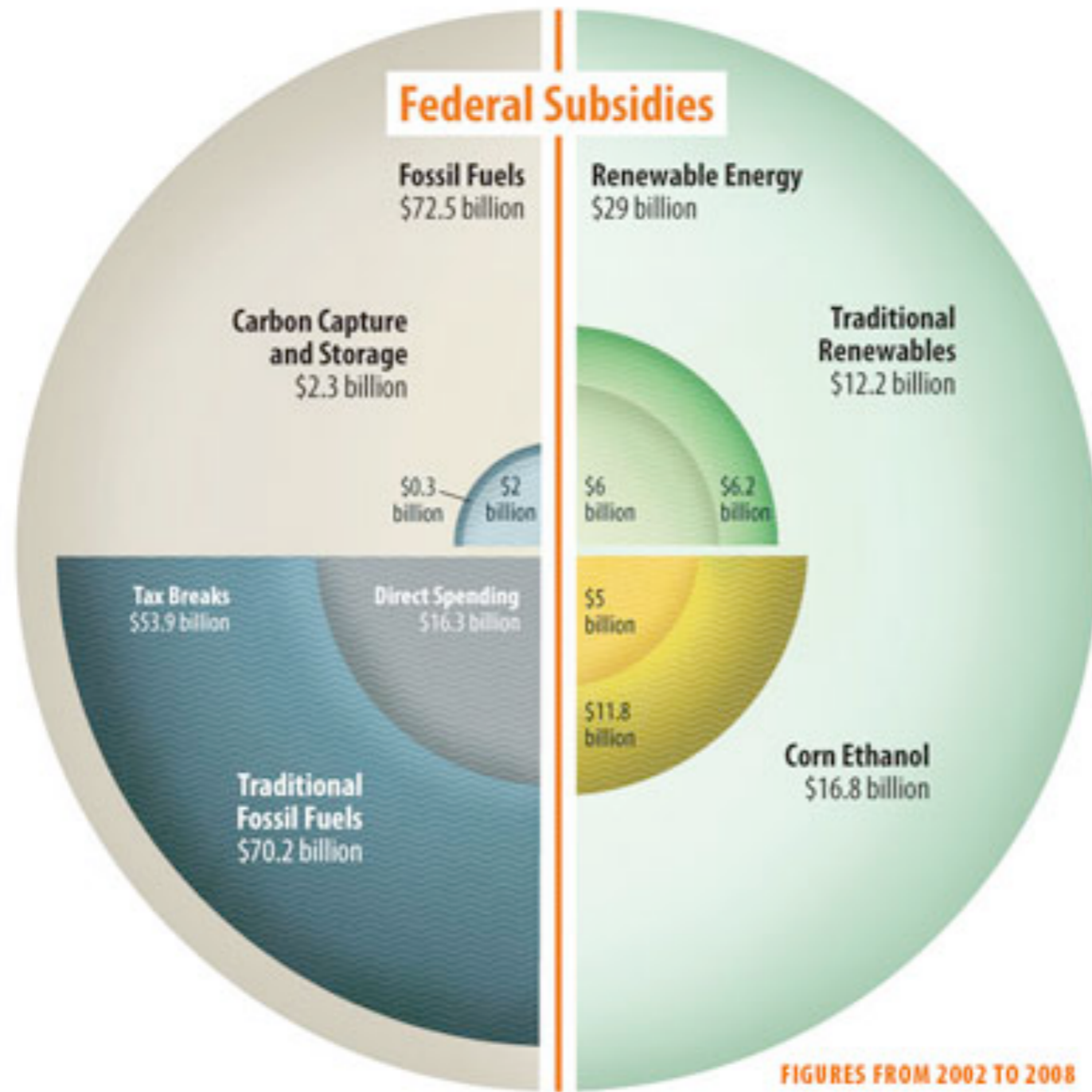
Steve C. Wang, an associate professor of statistics at Swarthmore College, charted baseball managers from the 2007 season as Chernoff faces, a method of using the heights, widths and angles of facial features to represent different sets of numbers.



*Percentage of players who had the advantage of batting against an opposite-handed pitcher at the start of the game.

Note: Because different rules cause National League managers to use more pinch-hitters, for example, each manager's rates are compared with his league's average.

JONATHAN CORUM/
THE NEW YORK TIMES



Peter and Maria Hoey (Source: Tommy McCall/Environmental Law Institute) via Phil

THERE'S NO PLACE LIKE HOME:

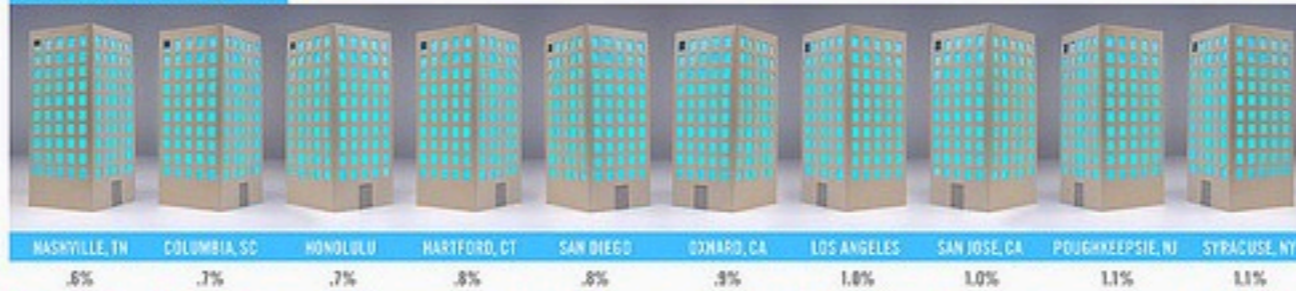
As the number of foreclosures continue to mount, vast swaths of American cities are becoming deserted. In some places, more than a quarter of available rentals sit empty, while other cities have largely resisted the trend. This is a look at the 10 most vacant and most occupied metropolitan areas for both houses and rental apartments. It looks like you could get a great deal in Memphis.

% OF HOMES VACANT



NATIONAL AVERAGE: 2.5%

10 CITIES WITH MOST OCCUPIED HOMES



10 CITIES WITH MOST VACANT HOMES



10 CITIES WITH MOST OCCUPIED RENTAL PROPERTIES



% RENTAL HOMES VACANT



NATIONAL AVERAGE: 10.6%

10 CITIES WITH MOST VACANT RENTAL PROPERTIES



Where good data vis goes wrong

- Scale
- Distortion
- Context
- Labeling
- Design variation vs. data variation
- Chart junk