

Welcome to Transforming Data!

FALL 2012/New School
zannah marsh
zmarsh10@gmail.com

What we'll do this semester...

- Learn to identify, “read”, and critically evaluate data visualizations and infographics
- Try making our own data visualizations and infographics, both as analog and digital representations
- Learn some **basic** programming techniques and concepts in order to make digital, interactive data displays

Data

- what is it?
- where does it come from?
- how do we use it?

2.3

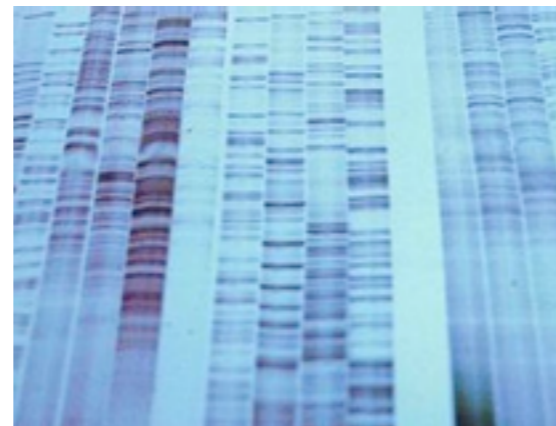
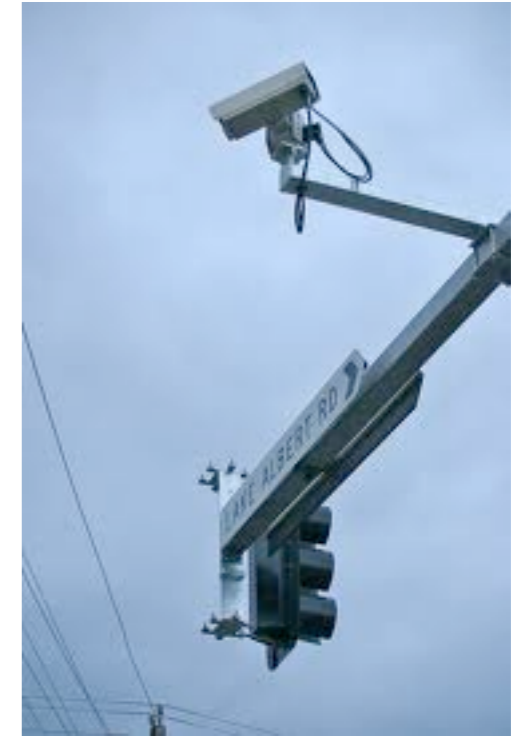
**“...a mark or trace that represents
a portion of the real world...”**

-sharon daniel, the database: an aesthetics of dignity

2.3

Src	Eqid	Version	Datetime	Lat	Lon	Magnitude	Depth	NST	Region
uw	06281051	0	"Monday, June 28, 2010 10:51:48 UTC"	46.8308	-119.6931	2.5	2.10	36	"Washington"
ci	10727357	2	"Monday, June 28, 2010 10:42:52 UTC"	32.8272	-116.2242	2.0	10.80	74	"Southern California"
us	2010yba2	6	"Monday, June 28, 2010 10:40:20 UTC"	33.4046	142.4072	4.6	35.00	24	"off the east coast of"
ci	10727349	2	"Monday, June 28, 2010 10:25:38 UTC"	32.5868	-115.7342	2.3	7.00	38	"Baja California, Mexico"
ak	10057383	1	"Monday, June 28, 2010 10:17:04 UTC"	61.6263	-147.9067	2.0	24.50	21	"Southern Alaska"
ci	10727341	2	"Monday, June 28, 2010 10:06:27 UTC"	32.2223	-115.4355	2.6	12.40	20	"Baja California, Mexico"
nc	71416520	1	"Monday, June 28, 2010 09:37:15 UTC"	38.8065	-122.8113	1.7	2.40	32	"Northern California"
us	2010ybay	6	"Monday, June 28, 2010 09:32:20 UTC"	-23.0119	-66.2953	4.6	209.60	61	"Jujuy, Argentina"
ci	10727333	2	"Monday, June 28, 2010 09:30:46 UTC"	32.7062	-115.9413	1.6	2.30	13	"Southern California"
us	2010ybax	6	"Monday, June 28, 2010 09:28:40 UTC"	-16.1279	-176.2618	4.7	355.00	17	"Fiji region"
ci	10727317	2	"Monday, June 28, 2010 09:13:59 UTC"	32.7293	-115.9703	1.6	5.00	24	"Southern California"
ci	10727309	2	"Monday, June 28, 2010 09:04:15 UTC"	32.6610	-115.7558	1.4	4.40	14	"Southern California"
ci	10727301	2	"Monday, June 28, 2010 08:48:19 UTC"	32.7183	-115.9967	2.3	3.10	73	"Southern California"
ci	10727285	2	"Monday, June 28, 2010 08:38:17 UTC"	32.6838	-115.9237	1.9	3.20	45	"Southern California"
nn	00313087	1	"Monday, June 28, 2010 08:14:13 UTC"	38.1800	-118.2180	1.8	0.00	33	"Nevada"
ci	10727269	2	"Monday, June 28, 2010 08:09:32 UTC"	32.1898	-115.3162	2.4	12.10	17	"Baja California, Mexico"
ak	10057380	1	"Monday, June 28, 2010 07:35:14 UTC"	59.8500	-150.6964	2.6	25.00	22	"Kenai Peninsula, Alaska"
ci	10727261	2	"Monday, June 28, 2010 07:34:50 UTC"	34.2508	-117.5123	1.5	13.60	49	"Southern California"
ci	10727253	2	"Monday, June 28, 2010 07:13:54 UTC"	32.6500	-115.8835	1.5	6.90	19	"Southern California"
us	2010ybas	6	"Monday, June 28, 2010 07:04:47 UTC"	-23.4401	179.5444	5.0	525.10	59	"south of the Fiji Islands"
nc	71416470	0	"Monday, June 28, 2010 06:47:36 UTC"	41.2387	-121.9353	1.6	5.00	8	"Northern California"
nm	hnw0628a	B	"Monday, June 28, 2010 06:36:35 UTC"	35.2218	-91.8580	3.2	0.10	21	"Arkansas"
nc	71416465	0	"Monday, June 28, 2010 06:34:01 UTC"	36.5105	-121.0948	1.2	5.60	12	"Central California"
ci	10727237	2	"Monday, June 28, 2010 06:33:39 UTC"	32.9187	-116.2892	1.2	6.90	36	"Southern California"
us	2010ybam	7	"Monday, June 28, 2010 06:27:31 UTC"	-13.4831	166.9852	5.1	35.00	32	"Vanuatu"
ci	10727229	2	"Monday, June 28, 2010 06:26:07 UTC"	32.6658	-115.9263	1.3	8.40	24	"Southern California"
ci	10727221	2	"Monday, June 28, 2010 06:18:03 UTC"	32.7420	-115.9822	1.6	1.90	29	"Southern California"
ci	10727197	2	"Monday, June 28, 2010 05:17:30 UTC"	33.3920	-116.4110	1.2	12.70	48	"Southern California"
ci	10727157	2	"Monday, June 28, 2010 04:54:26 UTC"	32.6760	-115.8592	1.3	5.40	24	"Southern California"
ci	10727149	2	"Monday, June 28, 2010 04:52:57 UTC"	32.7107	-115.9443	1.5	4.20	24	"Southern California"
ci	10727133	2	"Monday, June 28, 2010 04:46:44 UTC"	32.7083	-115.9335	1.6	9.40	15	"Southern California"
nc	71416410	0	"Monday, June 28, 2010 04:40:45 UTC"	38.7878	-122.7285	1.1	2.40	9	"Northern California"
hv	60156541	1	"Monday, June 28, 2010 04:28:43 UTC"	19.2133	-156.0687	2.9	43.60	31	"Hawaii region, Hawaii"
ci	10727125	2	"Monday, June 28, 2010 04:06:15 UTC"	32.3480	-116.5843	2.0	42.40	14	"Baja California, Mexico"
ak	10057363	1	"Monday, June 28, 2010 04:05:27 UTC"	59.5407	-152.8517	2.5	98.00	34	"Southern Alaska"
ci	10727109	2	"Monday, June 28, 2010 04:02:02 UTC"	32.6818	-115.9665	1.5	5.60	17	"Southern California"
ak	10057362	1	"Monday, June 28, 2010 04:01:28 UTC"	63.5298	-151.0304	1.3	0.10	8	"Central Alaska"
ci	10727101	2	"Monday, June 28, 2010 03:44:01 UTC"	32.6940	-115.9387	1.9	5.70	42	"Southern California"
us	2010ybaj	7	"Monday, June 28, 2010 03:43:39 UTC"	-36.6168	-73.0013	4.7	35.00	48	"offshore Bio-Bio, Chile"
ci	10727093	2	"Monday, June 28, 2010 03:39:32 UTC"	32.7000	-115.9528	3.1	2.20	82	"Southern California"
ak	10057360	1	"Monday, June 28, 2010 03:19:03 UTC"	60.8122	-146.8105	1.6	17.60	9	"Southern Alaska"
ak	10057357	1	"Monday, June 28, 2010 03:15:38 UTC"	61.5813	-149.6276	1.8	118.90	9	"Southern Alaska"
ci	10727069	2	"Monday, June 28, 2010 03:13:37 UTC"	33.0675	-115.9982	1.5	3.20	38	"Southern California"

Where does data come from?



How much data?

Data inflation		
Unit	Size	What it means
Bit (b)	1 or 0	Short for “binary digit”, after the binary code (1 or 0) computers use to store and process data
Byte (B)	8 bits	Enough information to create an English letter or number in computer code. It is the basic unit of computing
Kilobyte (KB)	1,000, or 2^{10} , bytes	From “thousand” in Greek. One page of typed text is 2KB
Megabyte (MB)	1,000KB; 2^{20} bytes	From “large” in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4MB
Gigabyte (GB)	1,000MB; 2^{30} bytes	From “giant” in Greek. A two-hour film can be compressed into 1-2GB
Terabyte (TB)	1,000GB; 2^{40} bytes	From “monster” in Greek. All the catalogued books in America’s Library of Congress total 15TB
Petabyte (PB)	1,000TB; 2^{50} bytes	All letters delivered by America’s postal service this year will amount to around 5PB. Google processes around 1PB every hour
Exabyte (EB)	1,000PB; 2^{60} bytes	Equivalent to 10 billion copies of <i>The Economist</i>
Zettabyte (ZB)	1,000EB; 2^{70} bytes	The total amount of information in existence this year is forecast to be around 1.2ZB
Yottabyte (YB)	1,000ZB; 2^{80} bytes	Currently too big to imagine

The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures. Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.

Source: *The Economist*

UCSD researchers report that in 2008 the average American household was exposed to 3.6 zettabytes of information (or 34 gigabytes per person per day).

- data must retain a **connection** to its source to be meaningful
- data is only useful if you have **more than one** value
- data must be **interpreted** to be meaningful

- the act of measuring something **fixes its state**
- data's true origins are often **invisible**
- data has **authority**

Visualization

- what is it?
- why create one?
- how can we make good ones?
- why now, more than ever?

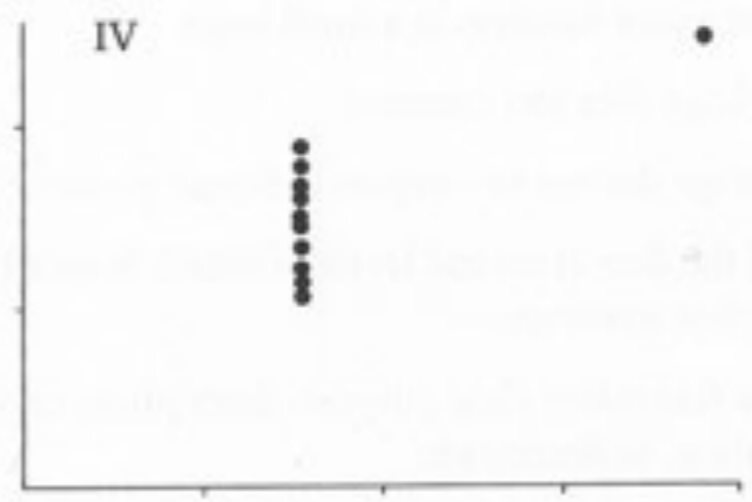
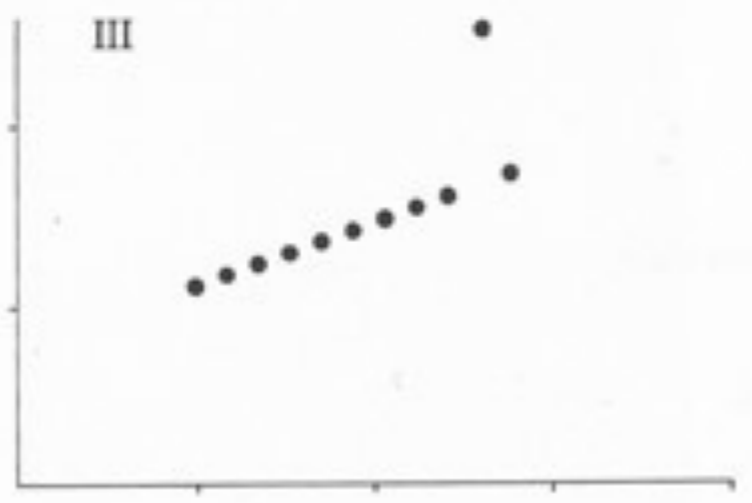
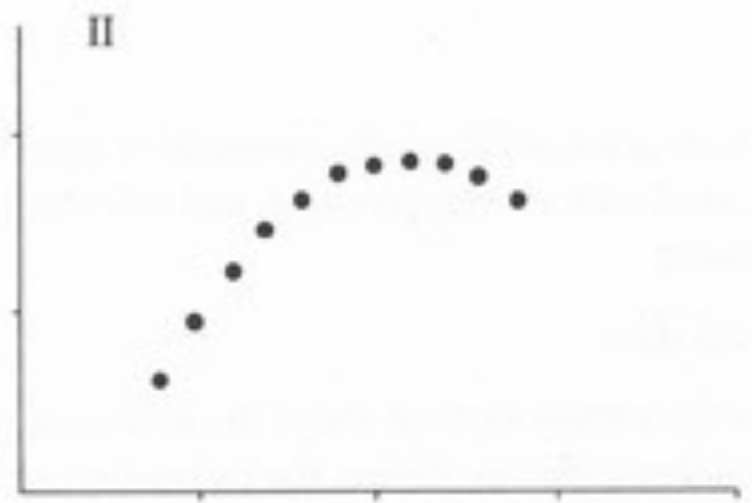
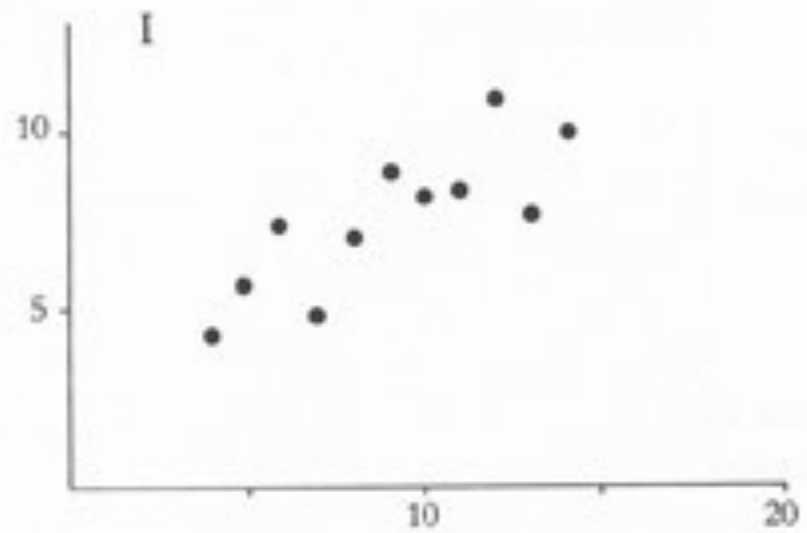
Definitions of Visualization

- the process of constructing a visual image in the mind
- **a graphical representation of data or concept**

Uses of Visualizations

- **record**- store information
- **analyze**- detect patterns
- **confirm**- verify hypotheses
- **communicate**- convey info, convince
- *****enable a task*****

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



Visualization as Cognitive Tool

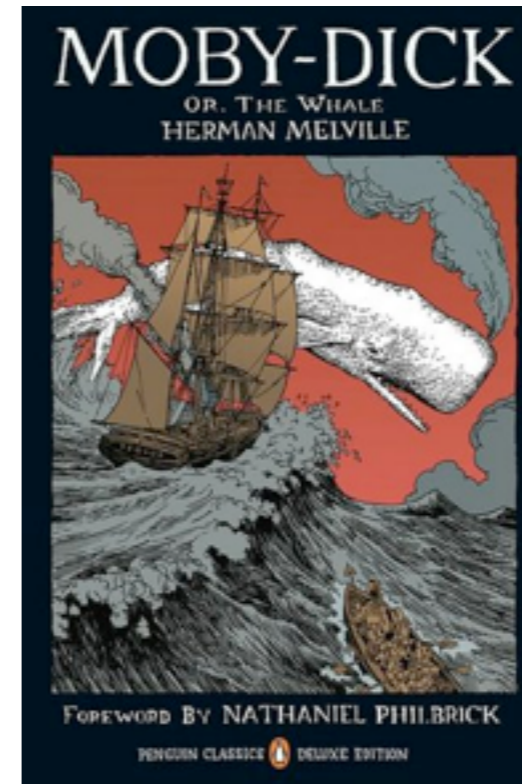
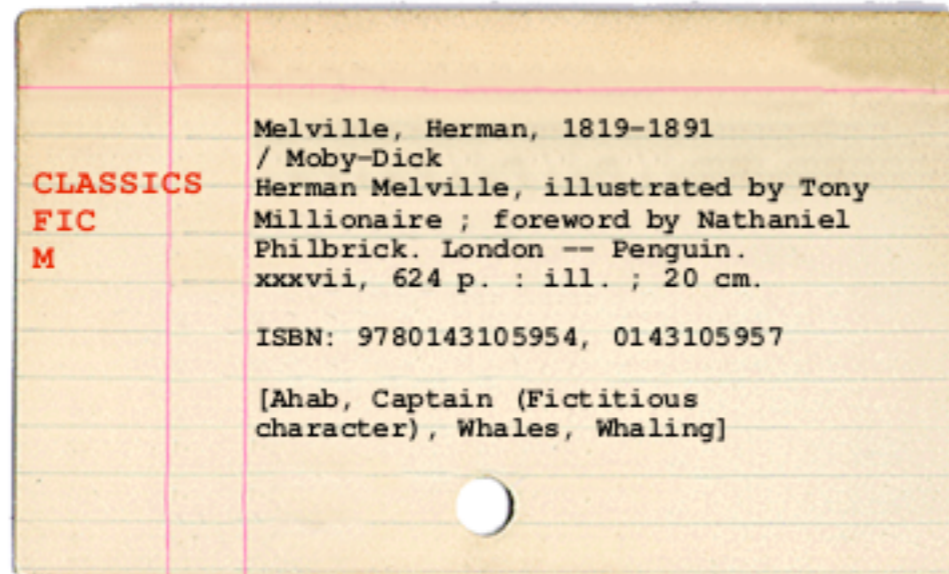
user's visual system
& decision-making
mechanism

information resources,
processing & computing
power

```
graph TD; A[user's visual system & decision-making mechanism] --> C[information visualization]; B[information resources, processing & computing power] --> C;
```

**information
visualization**

Data vs. Metadata



<http://www.blyberg.net/card-generator/>

DATA-VIZ flow chart

the WORLD is your data-source...

collect via observation or technology

methods?
processes?
omissions?
errors?
ownership?
agenda?

"RAW" data

data entry - errors? omissions?

organized DATABASE
DATA SET
search/retrieve

structure?
relationships?
hierarchy?
maintenance?
delay?

process (FRY)

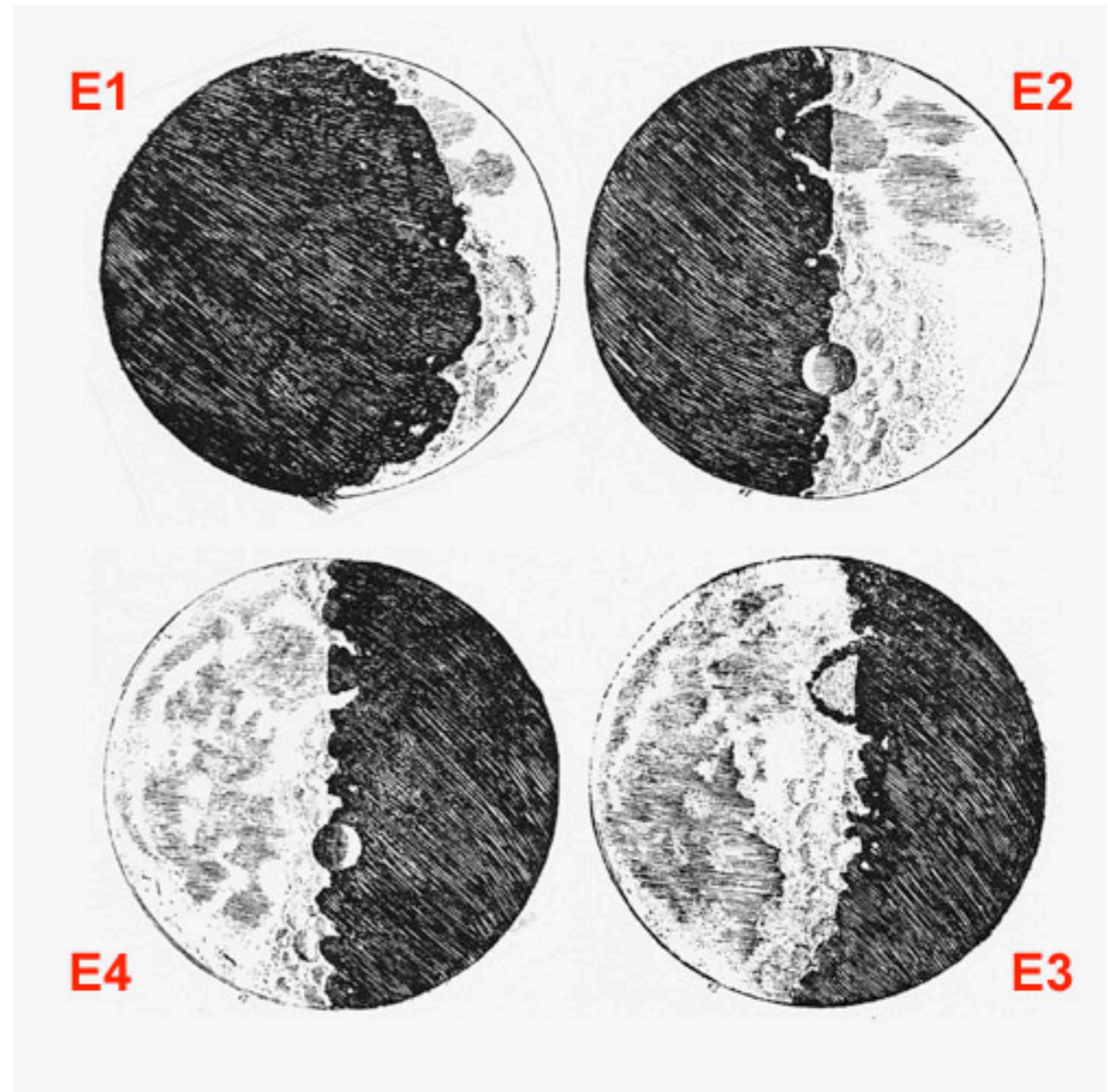
- parse / "clean"
- mine
- represent / TRANSLATE
- refine / DESIGN
- interact...

form vs. content?

yay!
visual system
interact
USER
semantic/cultural context

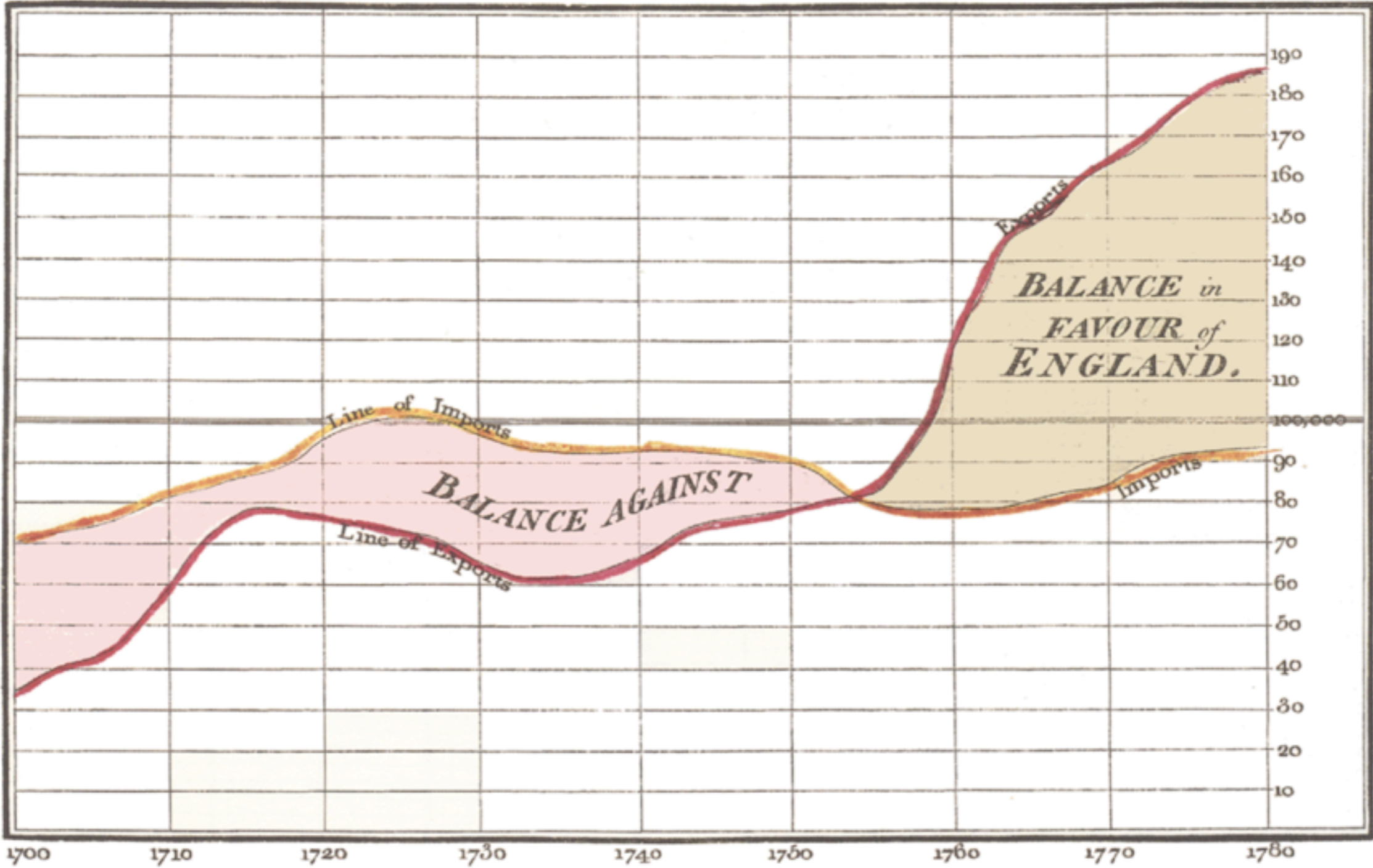
VISUALIZATION
graphical display of info
many formats
many platforms

Early Visualizations



left, cave painting at Lascaux
right, Galileo's moon phase drawings

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



The Bottom line is divided into Years, the Right hand line into L10,000 each.

Published as the Act directs, 1st May 1786. by W^m Playfair

Neale sculp^t 352, Strand, London.

William Playfair, Exports and Imports Chart, 1786



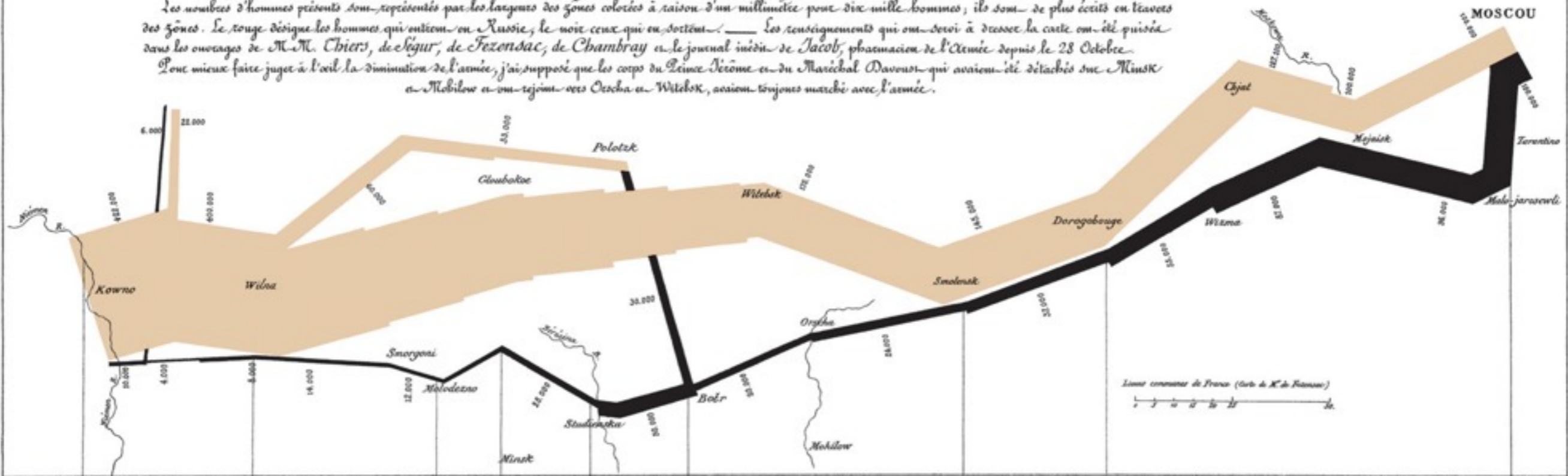
John Snow, Cholera Map (1854)

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.

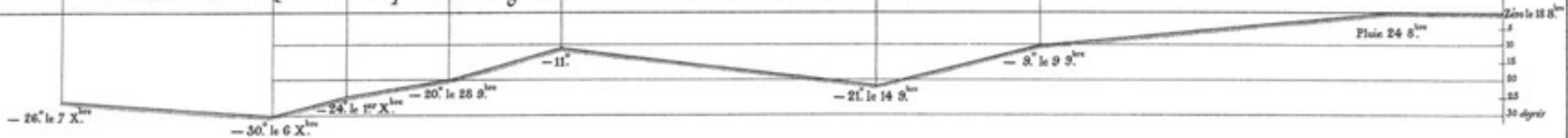
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont péri en Russie, le noir ceux qui en sont restés. — Les renseignements qui ont servi à dresser 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 Davout qui avaient été détachés sur Minsk et Mohilew et qui avaient rejoint vers Orescha 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.

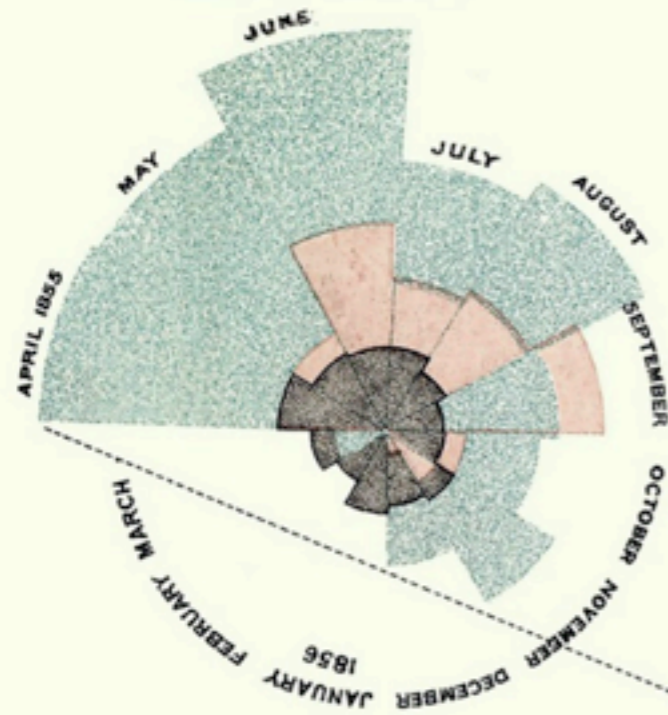


Les Cosaques passent au galop le Niemen gelé.

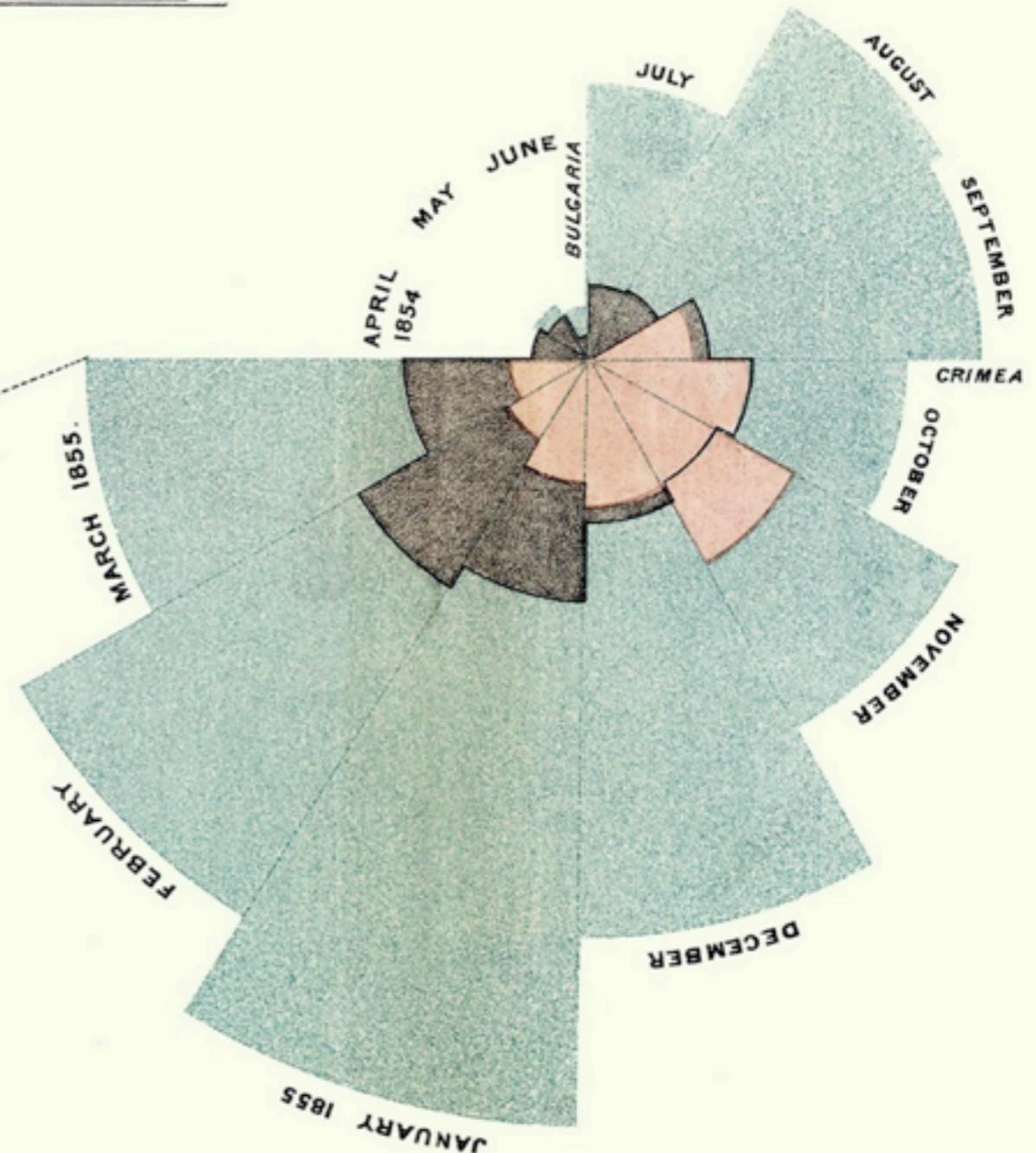
Joseph Minard
 Map of Napoleon's Russian 1812 Campaign (1869)

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

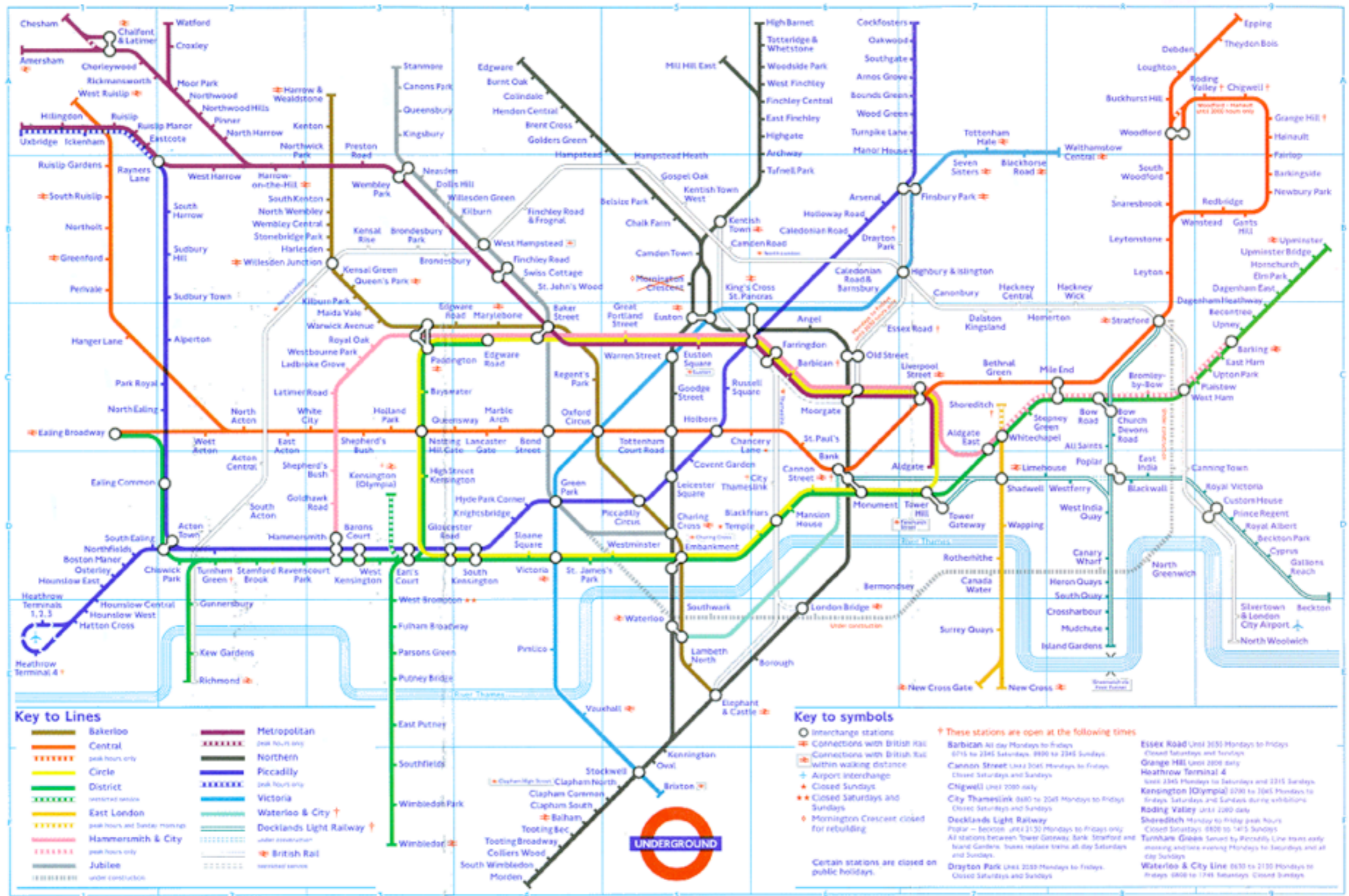
The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black.

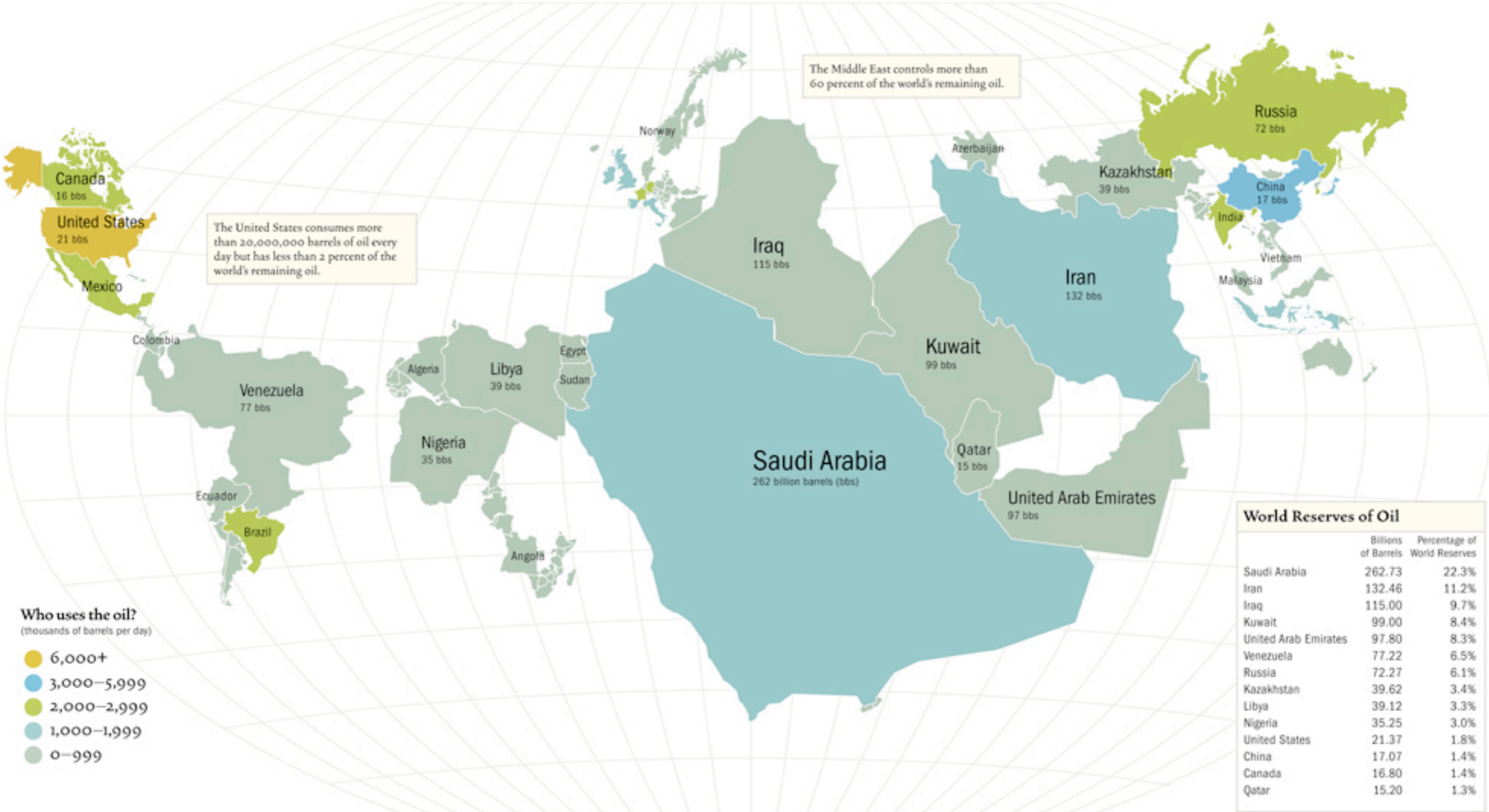
The entire areas may be compared by following the blue, the red & the black lines enclosing them.



Tube Map, 1927



Who has the oil?



Each country's size is proportional to the amount of oil it contains (oil reserves); Source: BP Statistical Review Year-End 2004 & Energy Information Administration

GEOGRAPHIC VIEW BY RANKING

1896 1900 1904 1908 1912 1920 1924 1928 1932 1936 1956 1960 1964 1968 1972 1976

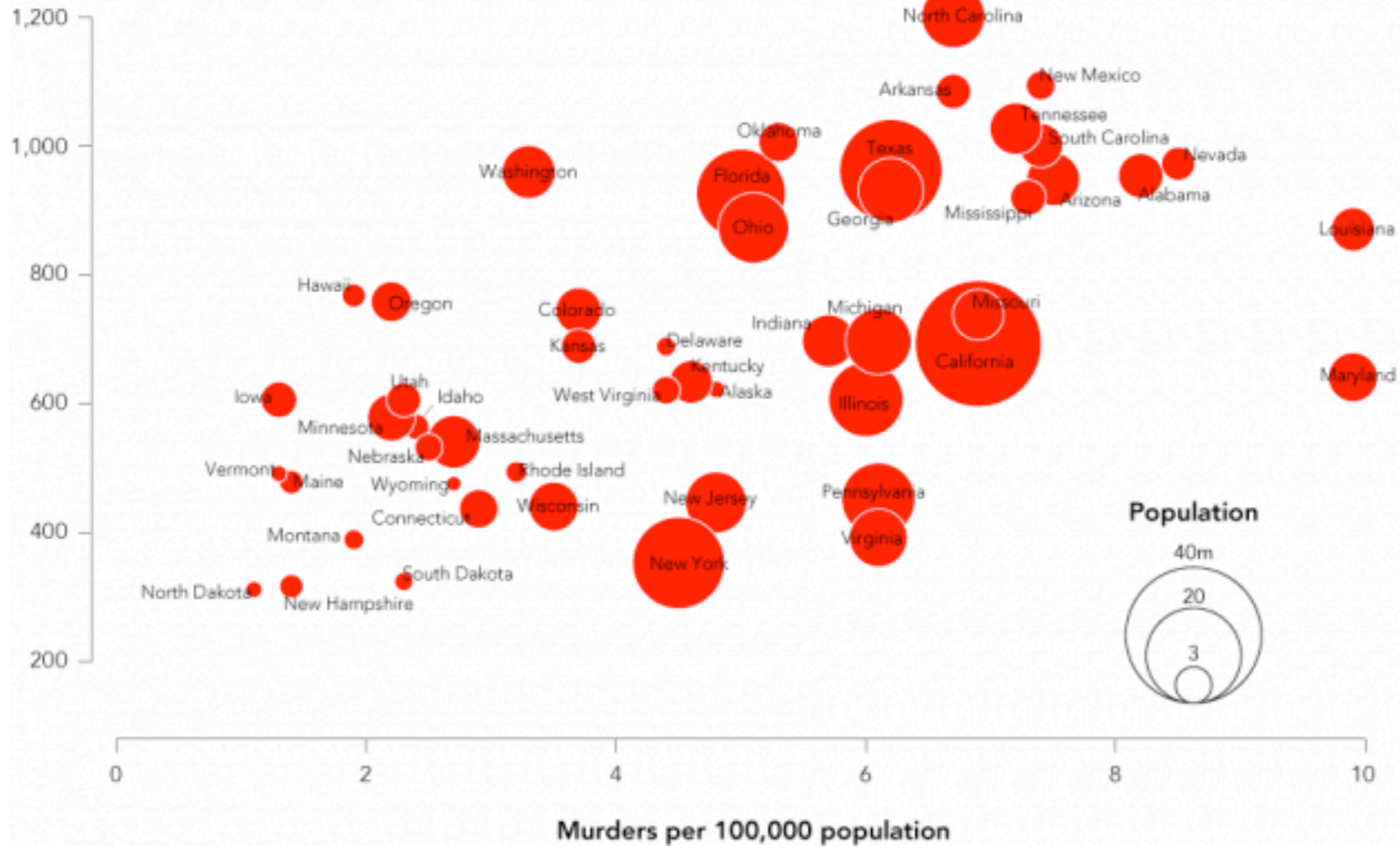


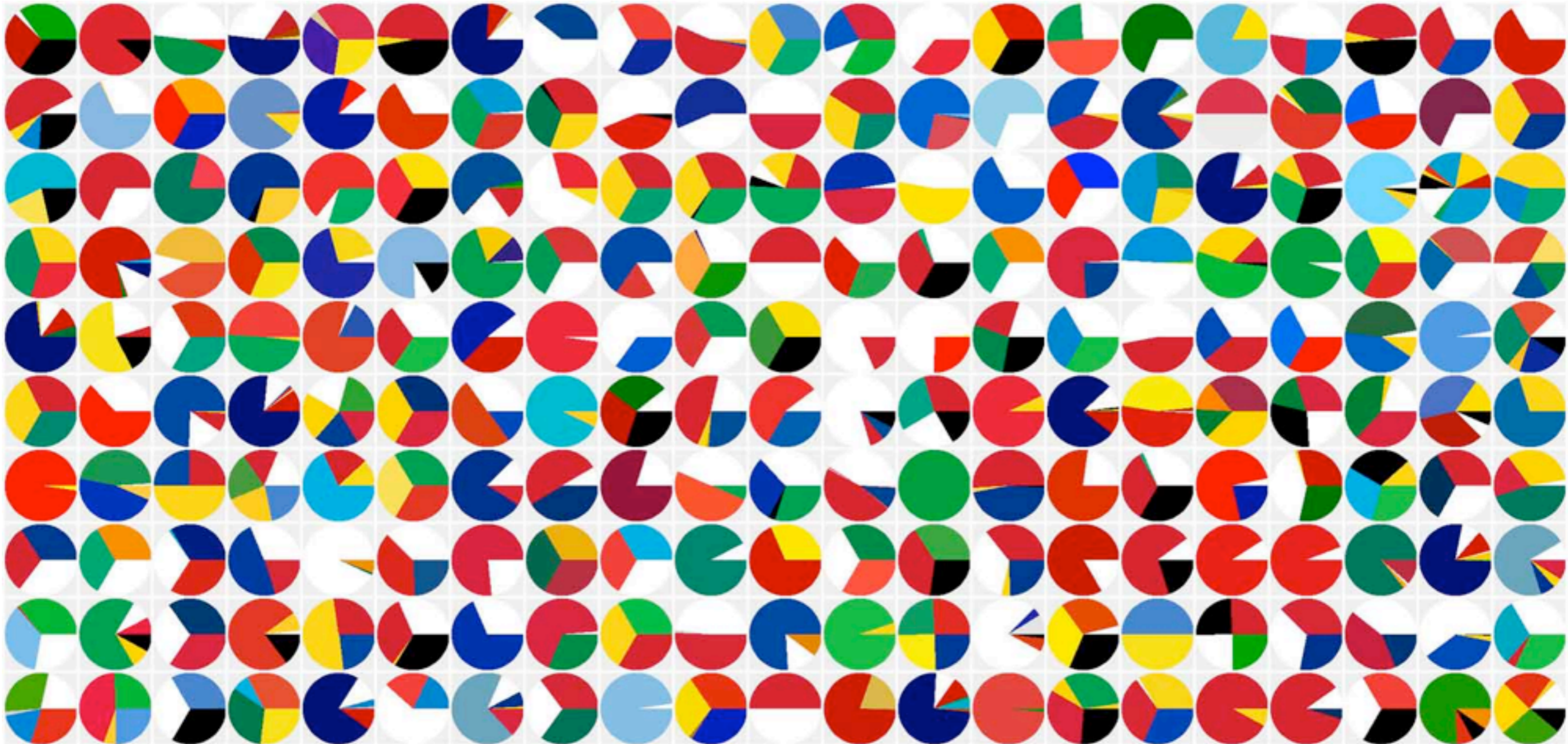
Source: International Olympic Committee

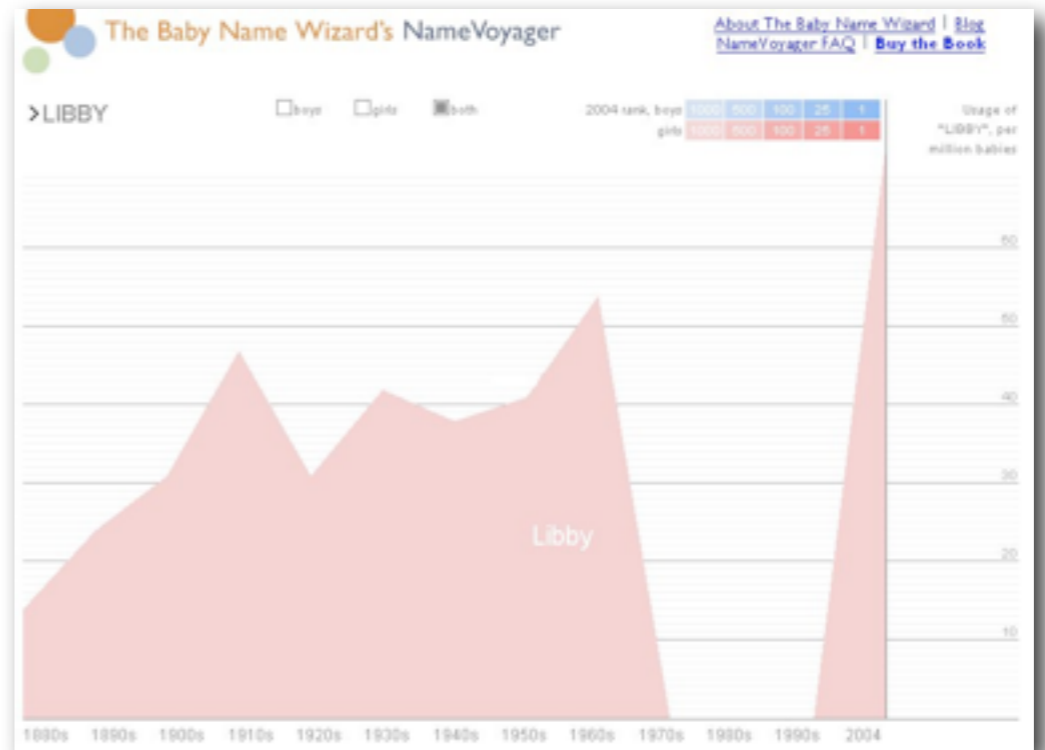
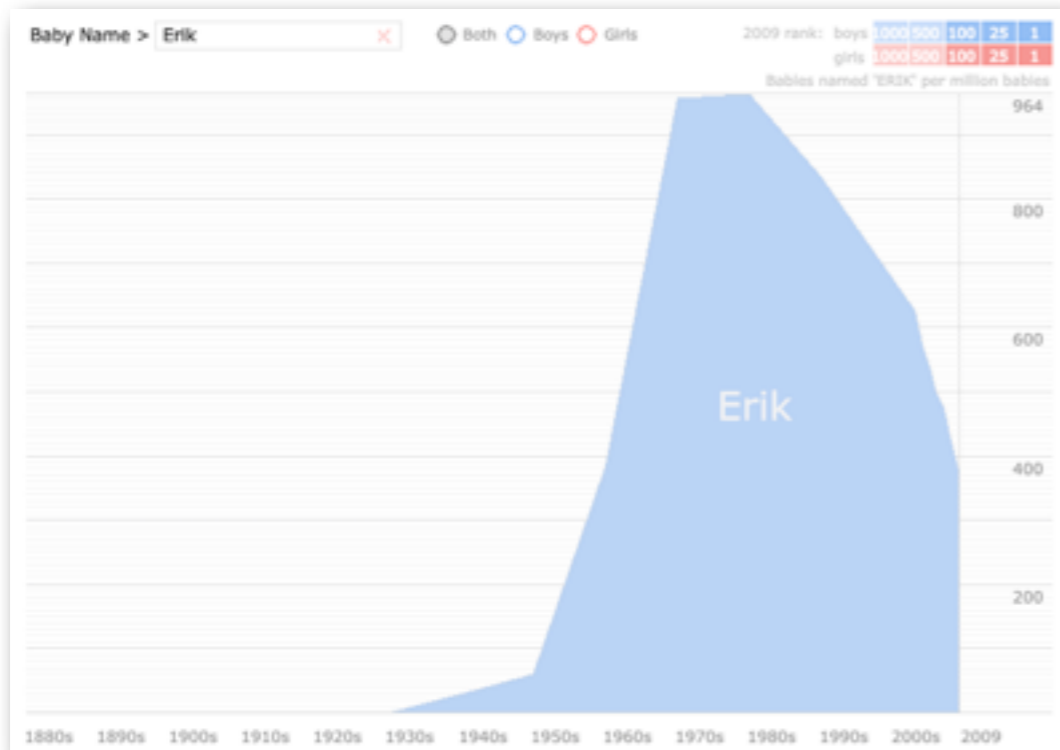
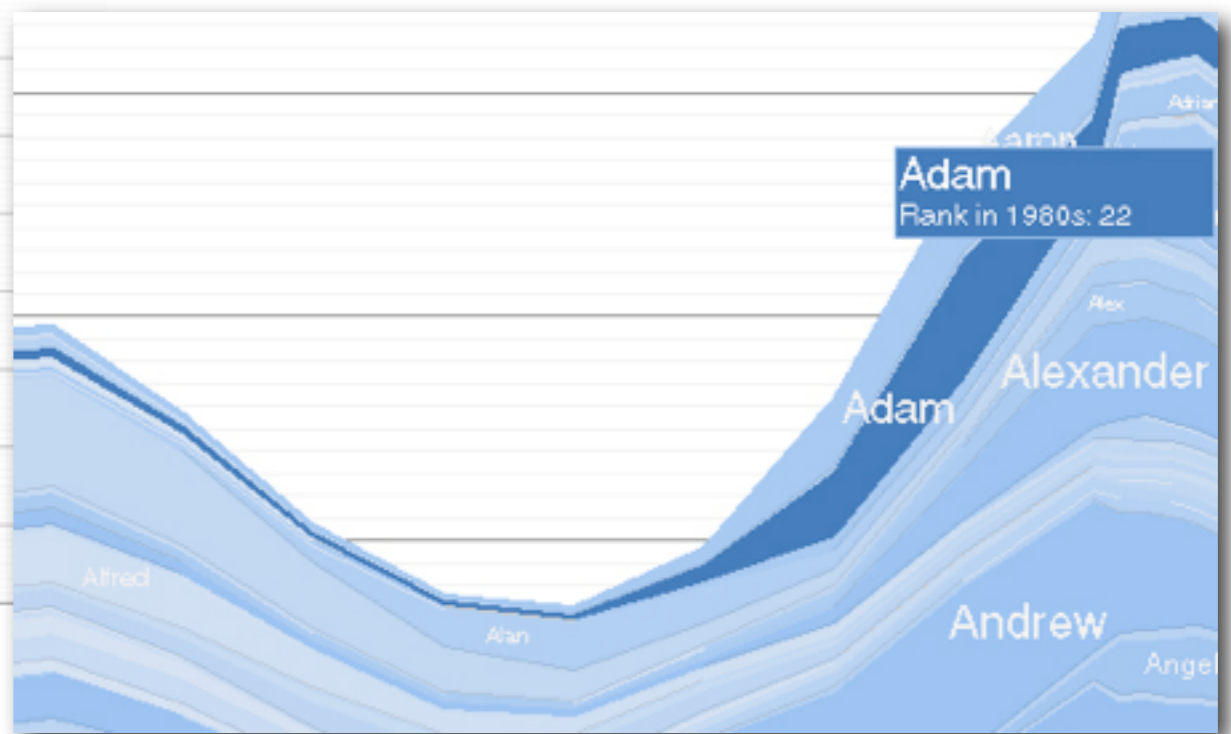
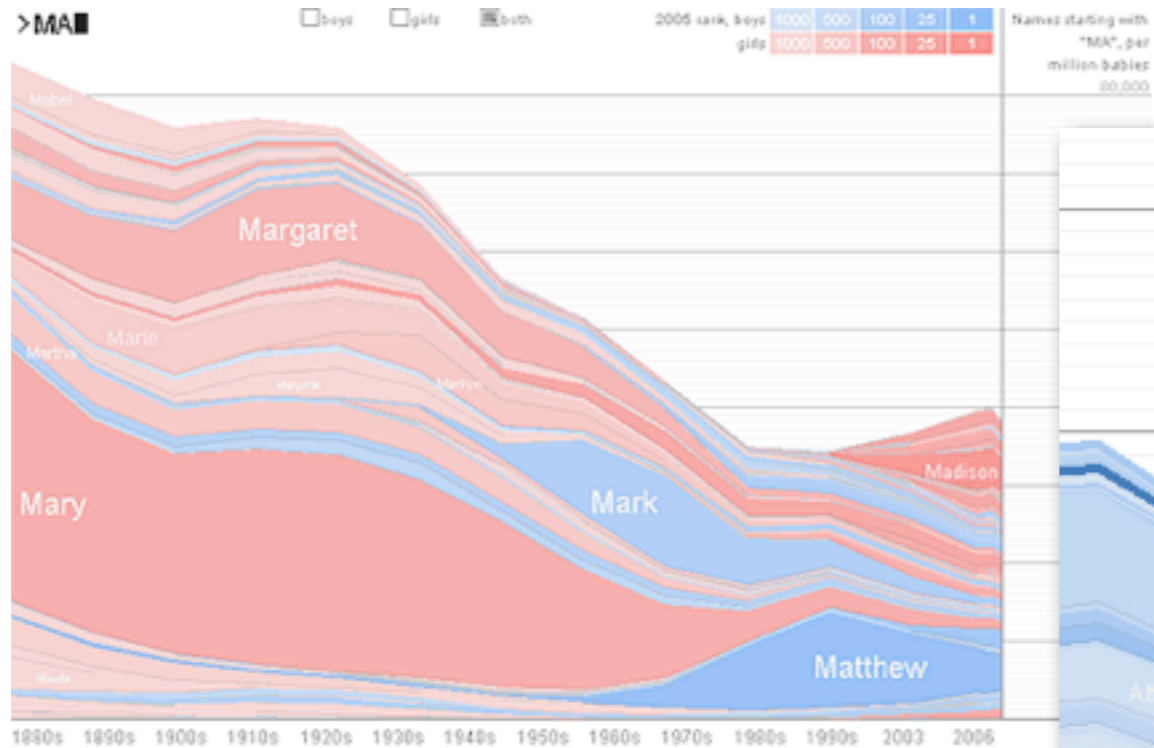
Lee Byron, Amanda Cox and Matthew Ericson/The New York Times

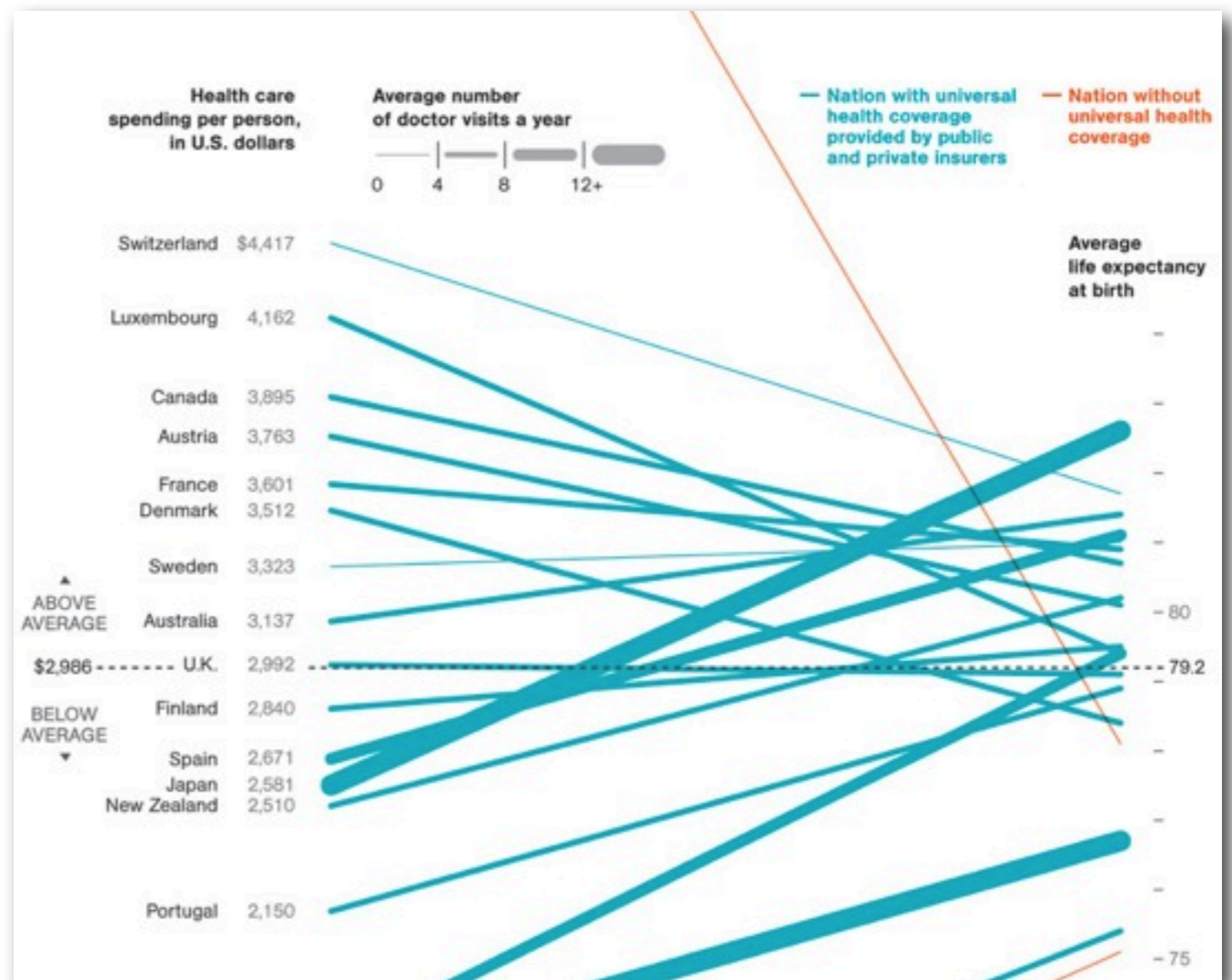
http://www.nytimes.com/interactive/2008/08/04/sports/olympics/20080804_MEDALCOUNT_MAP.html

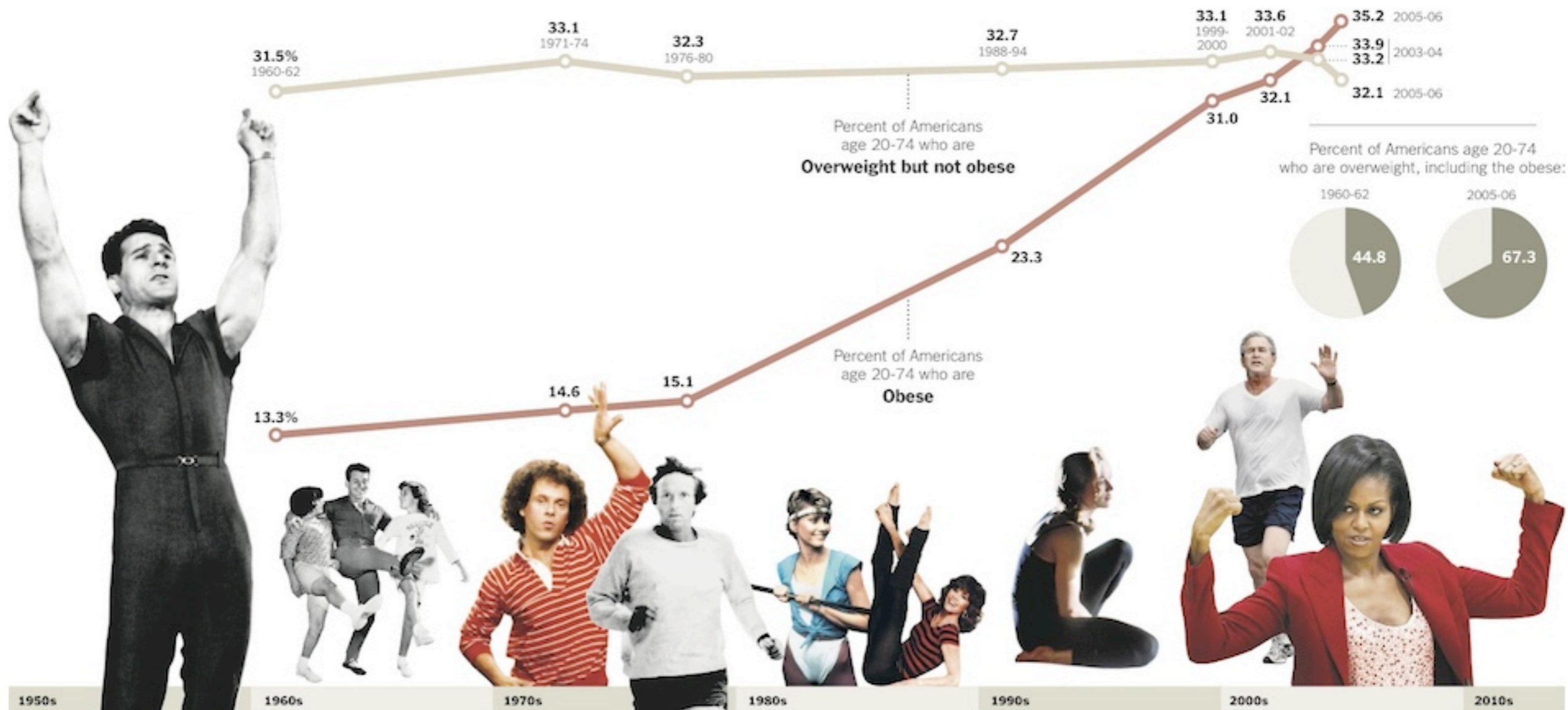
Burglaries per 100,000 population











1950s 1960s 1970s 1980s 1990s 2000s 2010s

1951: "The Jack LaLanne Show" begins. ▶

1961: The Kennedy administration organizes a conference on physical fitness and revamps the President's Council on Youth Fitness. (Above: LaLanne with students, early 1960s.)

1968: Kenneth H. Cooper publishes "Aerobics," which promotes running, swimming and other such exercises to improve the heart and lungs.

1974: Richard Simmons opens the Slimmons exercise studio.

1981: Olivia Newton-John releases the sexually suggestive exercise-themed song "Physical."

1977: James F. Fixx publishes "The Complete Book of Running," which helps popularize competitive running.

1982: Jane Fonda releases her first exercise video, "Jane Fonda's Workout."

1996: Madonna takes up yoga to get back in shape following the birth of her daughter.

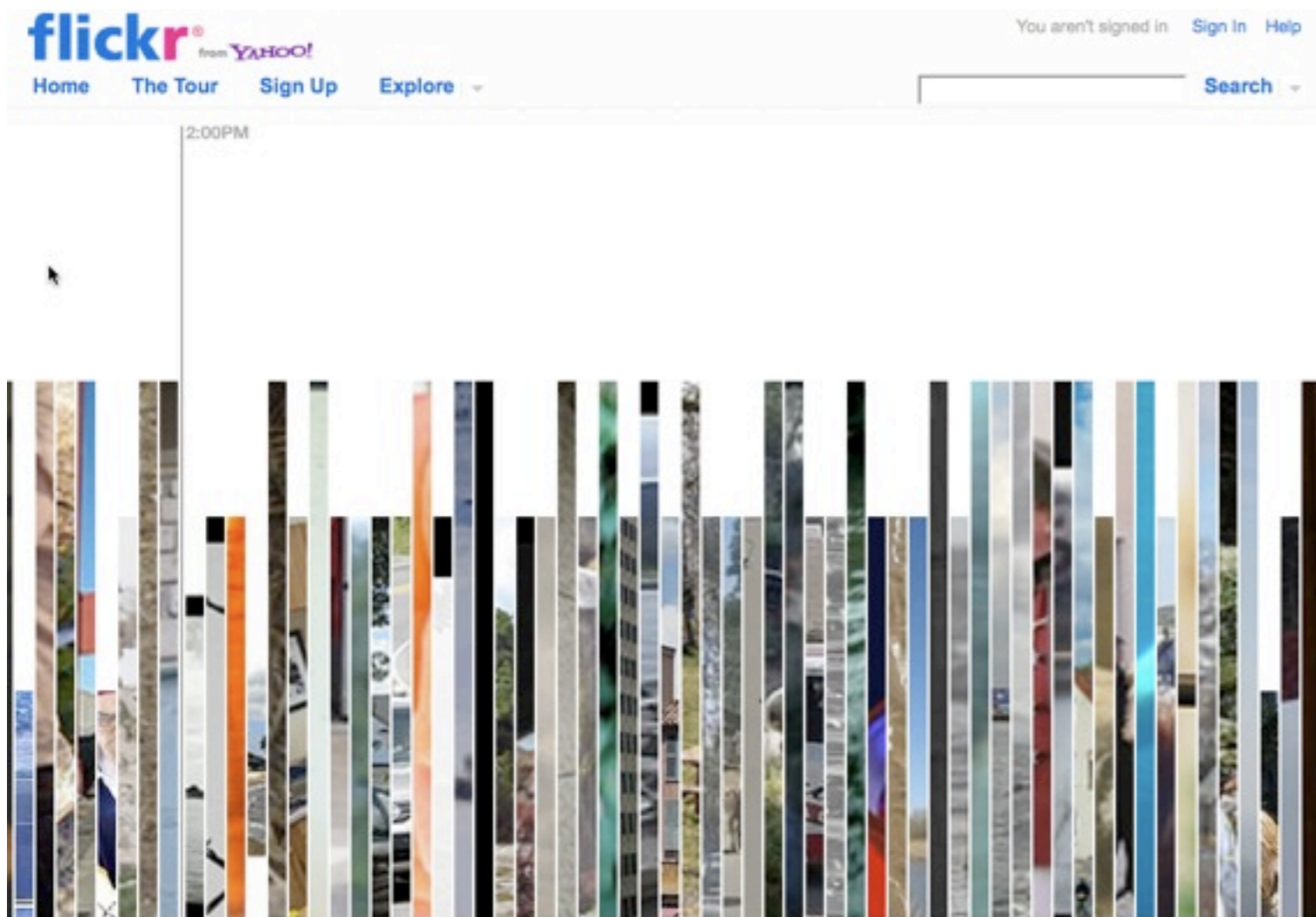
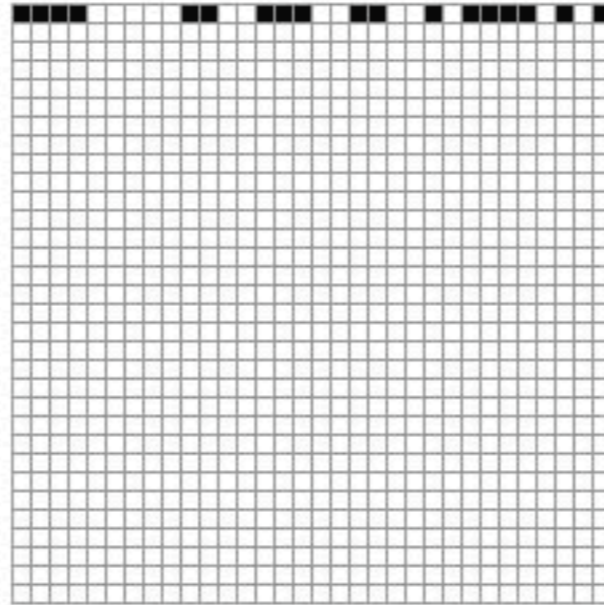
2002: George W. Bush tells Runner's World magazine that his running times became "faster right after the war began."

2006: Nintendo introduces the Wii, ushering in "lounge fitness." ▶

2010: Michelle Obama introduces Let's Move, a campaign to fight childhood obesity.

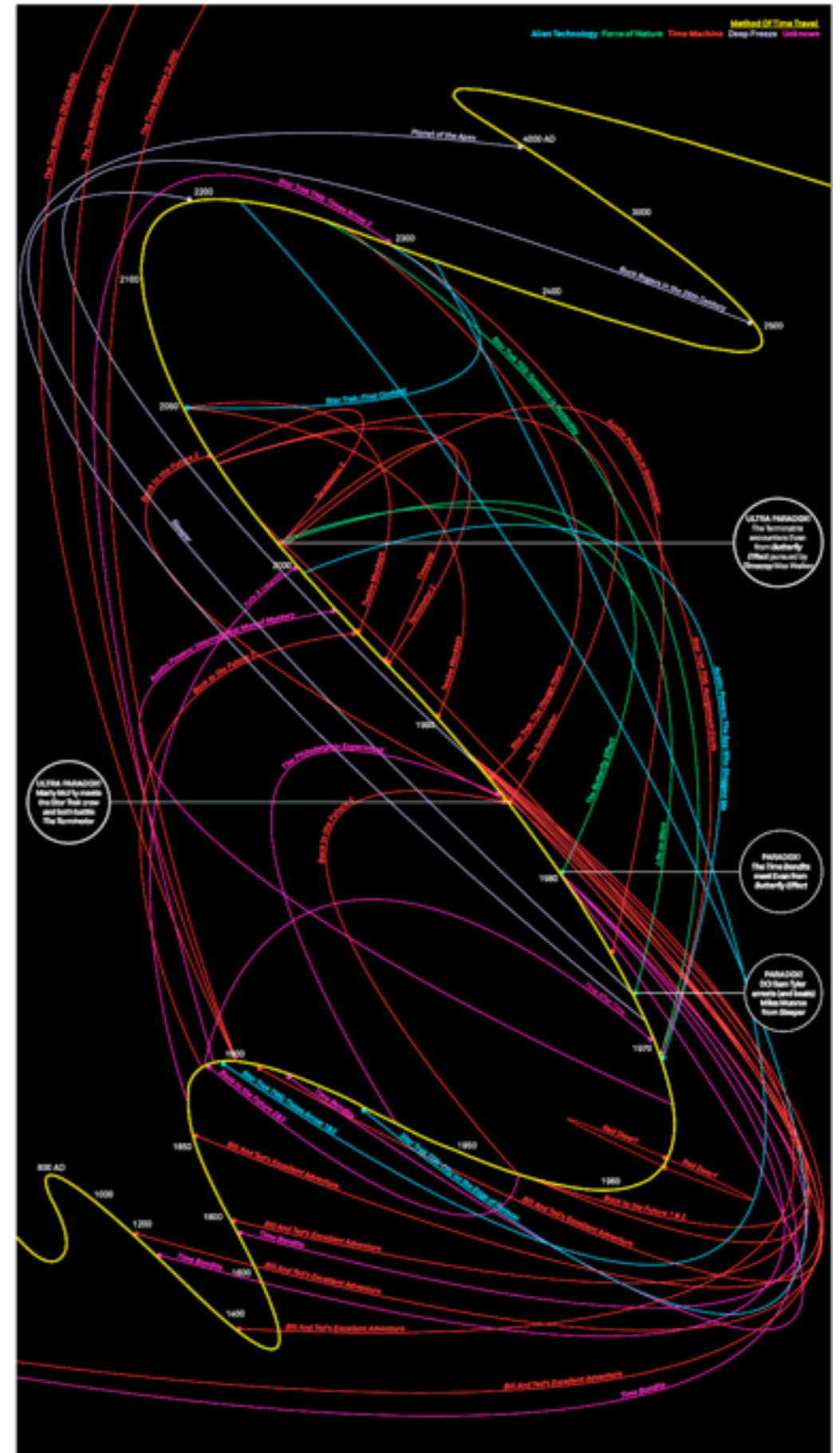


Time

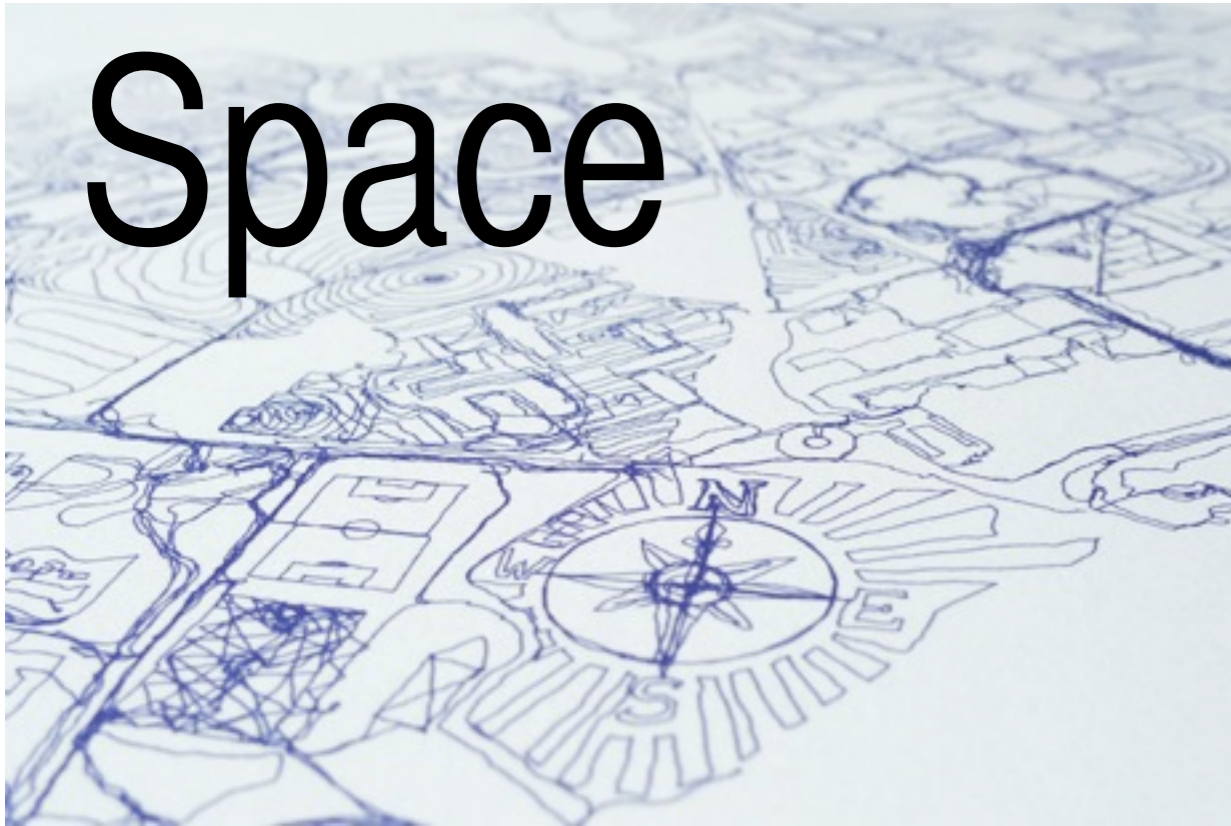


Timelines

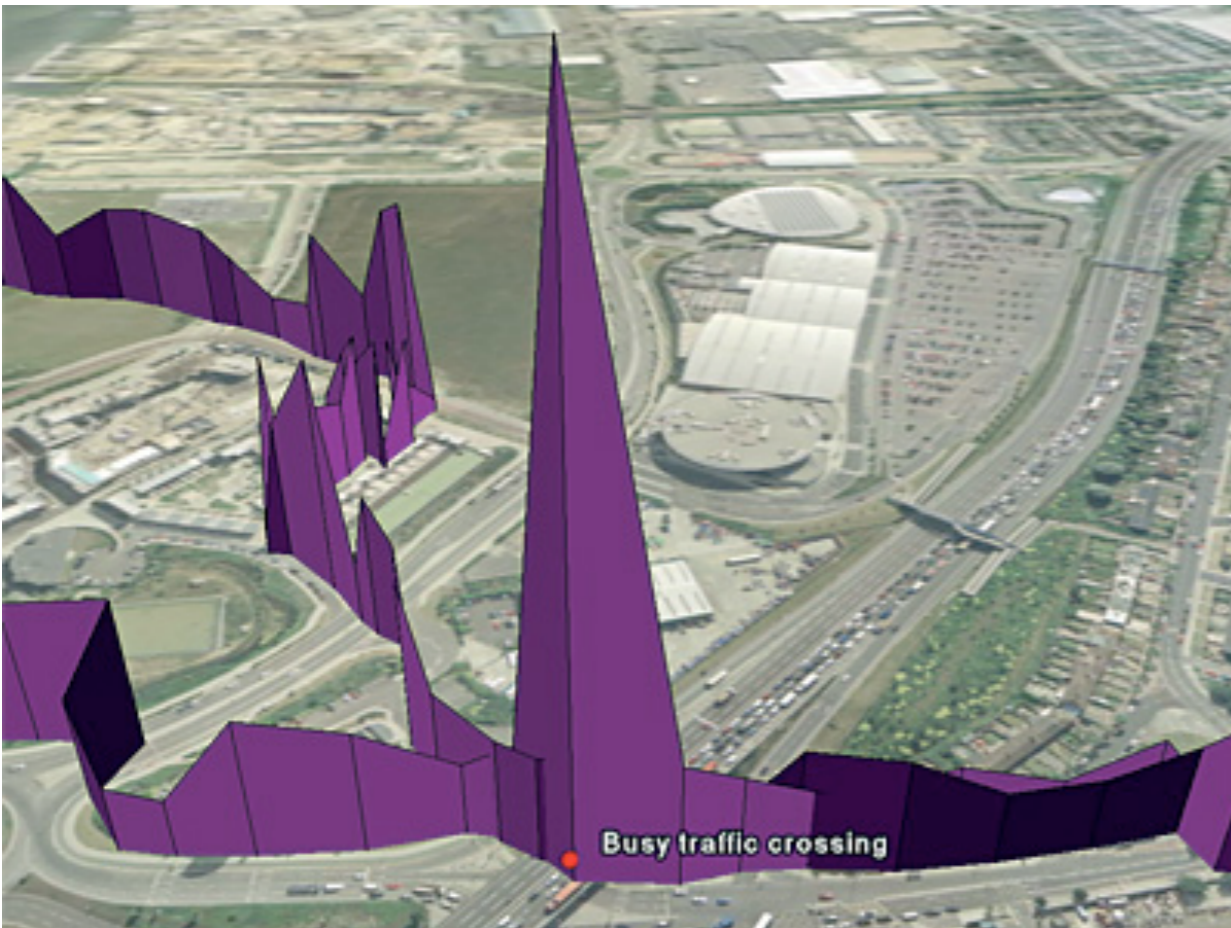
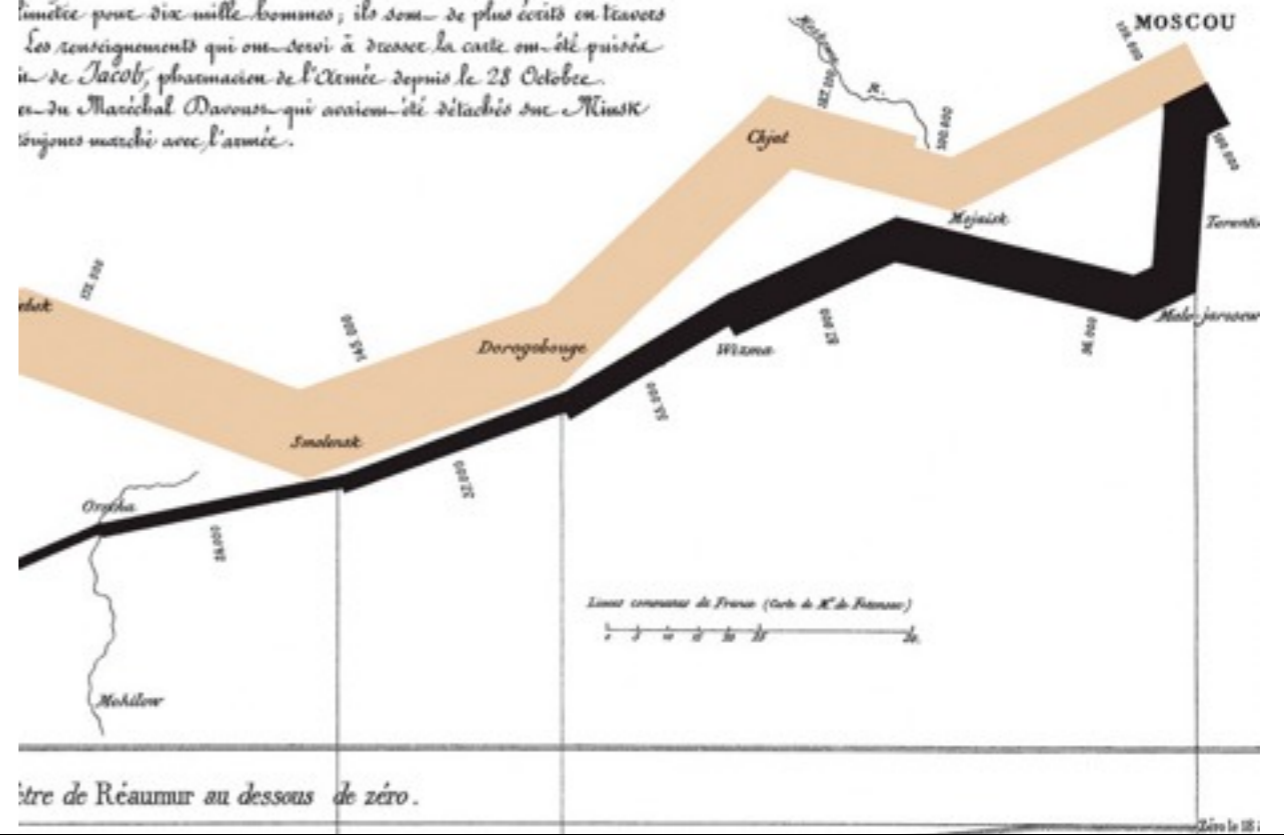
Time travel in popular film and tv



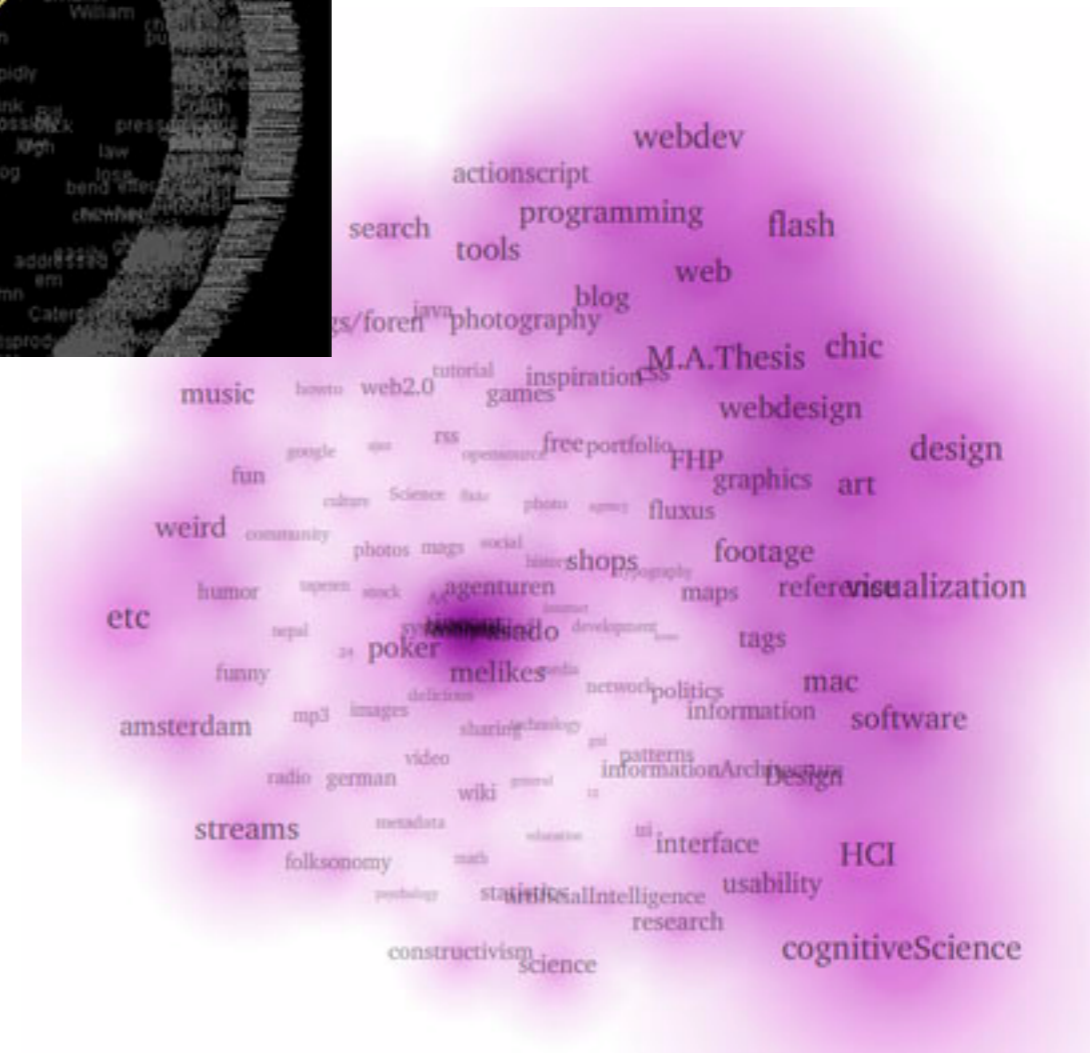
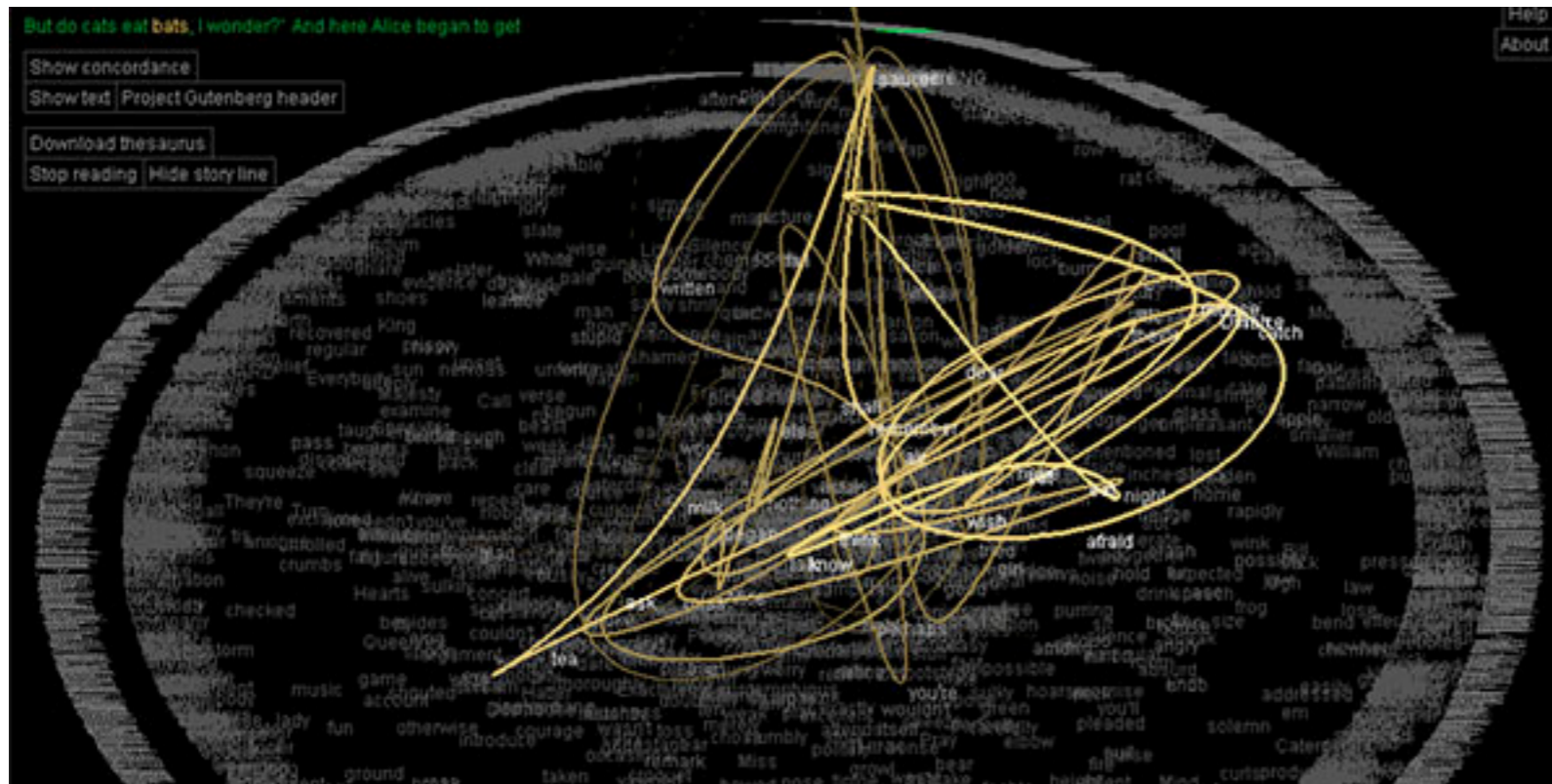
Space



— dans la campagne de Russie 1812-1813.
— Chausseur en retraite
Paris, le 20 Novembre 1869.
fonctio pour dix mille hommes; ils sont de plus écrits en lettres
Les renseignements qui ont servi à dresser la carte ont été puisés
in de Jacob, pharmacien de l'armée depuis le 28 Octobre.
et du Maréchal Davoust qui avaient été détachés sur Minsk
toujours marchés avec l'armée.



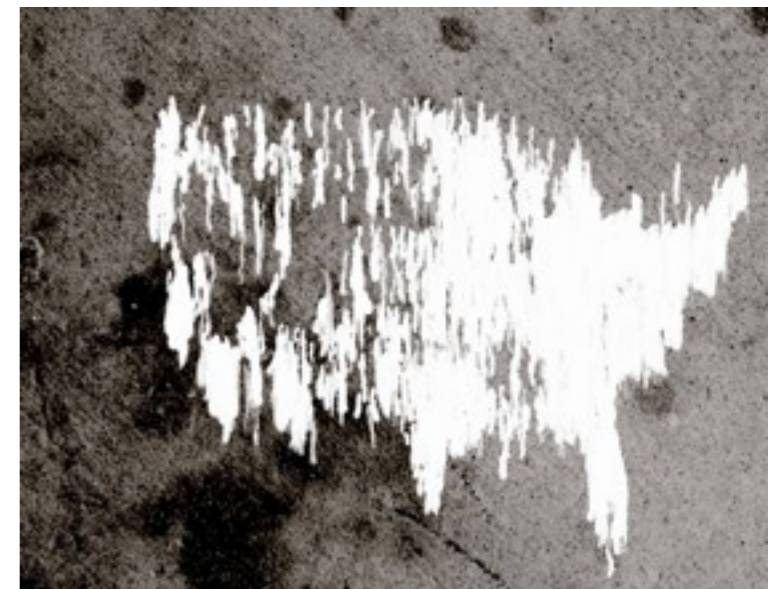
Text



Video!

Hans Rosling and Gap Minder

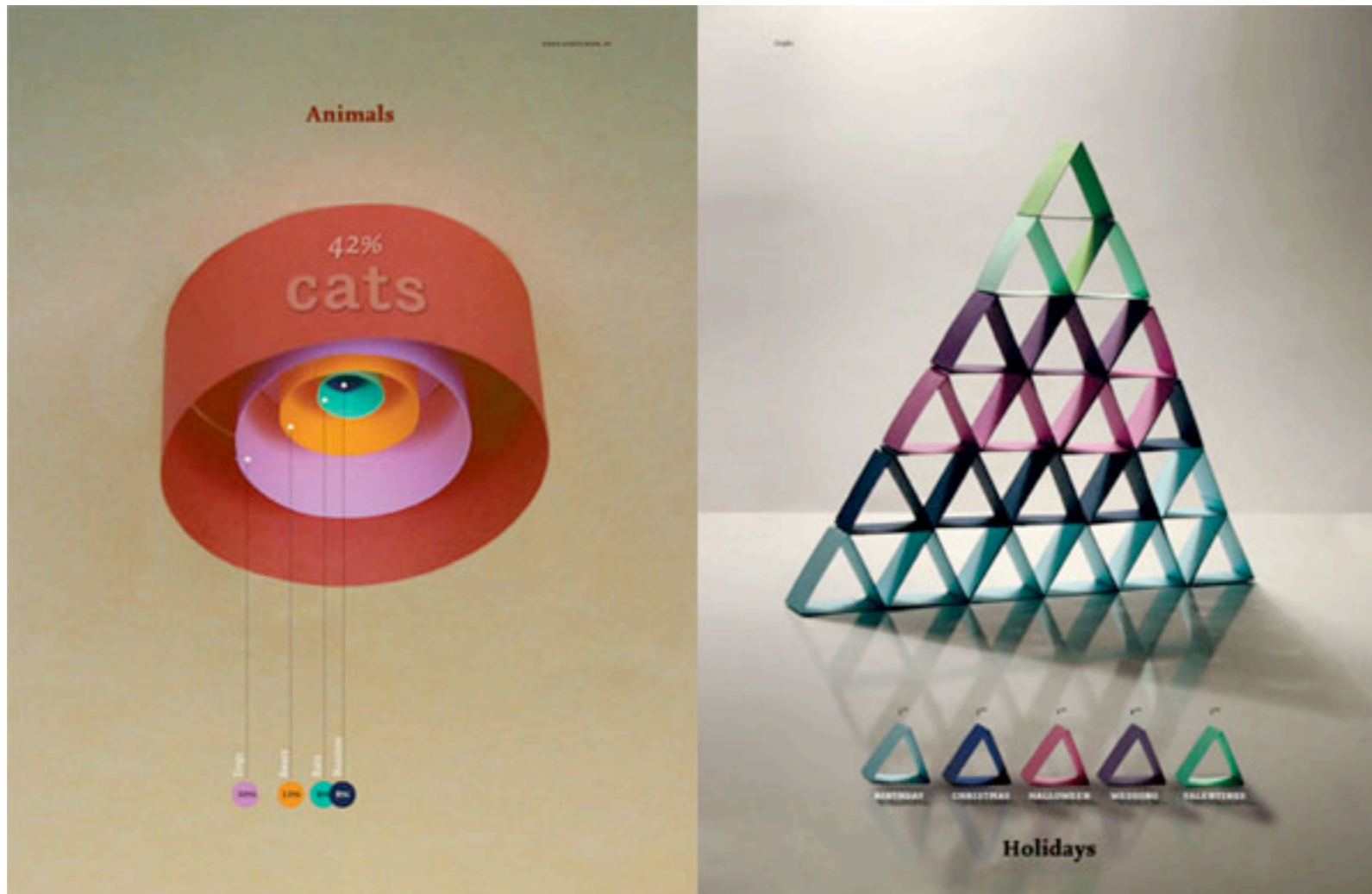
Introductions...



First Assignment

- Sketch two or three “analog” representations of aspects of our class data. Ask yourself a question about this data, and see if you can answer it.
- Make two “screen drawings” in Processing, using shape functions and colors.
- Animate your screen drawing in a loop, using a variable.

analog examples



gretchen nash <http://www.gretchenetc.com/deargretchen.html>



sarah illenberger

<http://www.sarahillenberger.com/stories.php?c=neon&n=2>

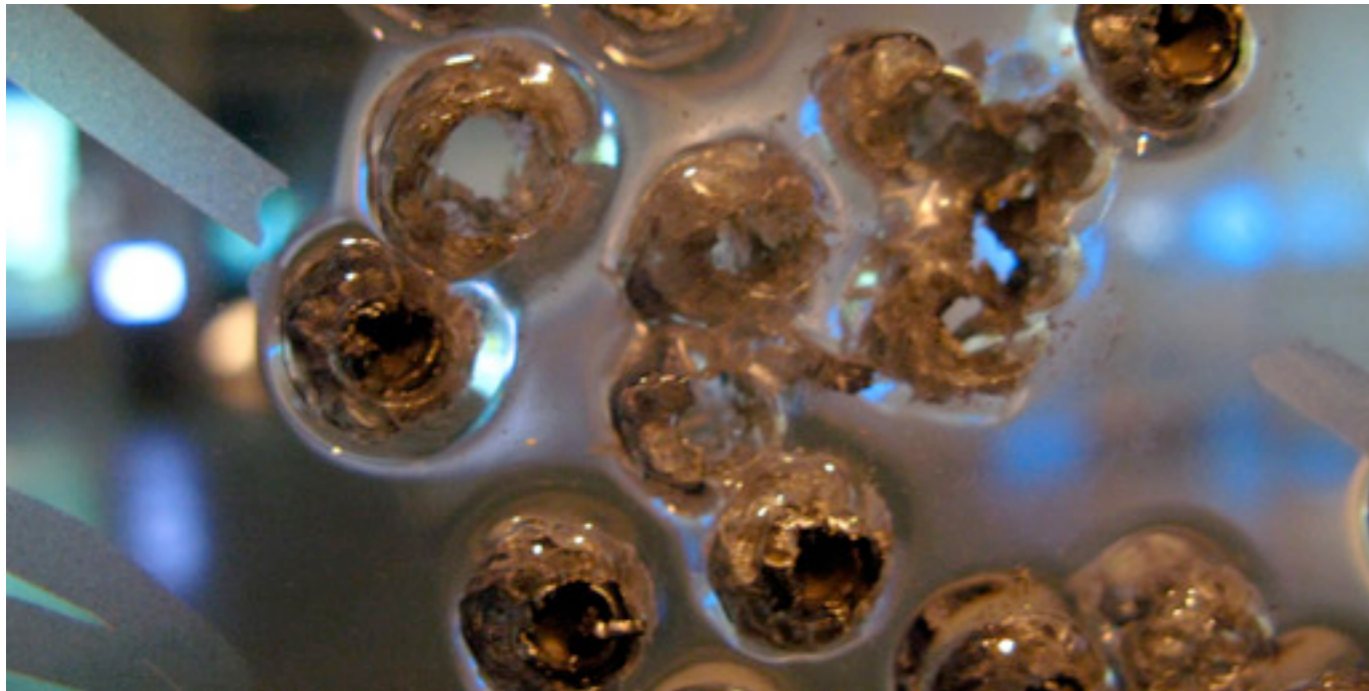




valentina de filippo <http://www.vale-n-tina.com/#138408/Relationship-matters>



corriette schoenaerts

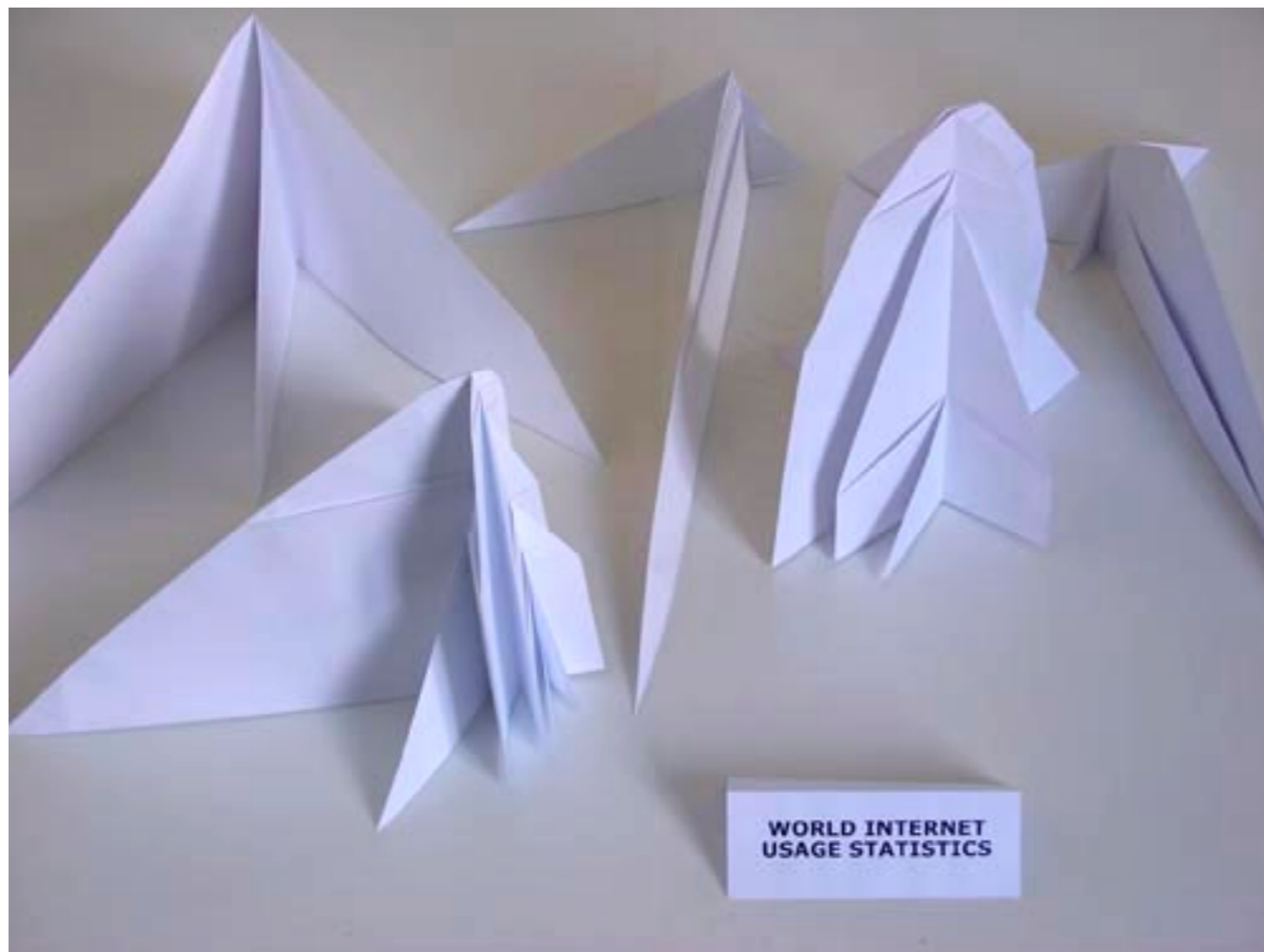


Instances of Use of United States Armed Forces Abroad,
1798–2006
.22 bullets shot into polycarbonate
157 inches x 78 inches x .5 inch
4 meters x 2 meters x 1 centimeter

<http://elahi.sjsu.edu/>

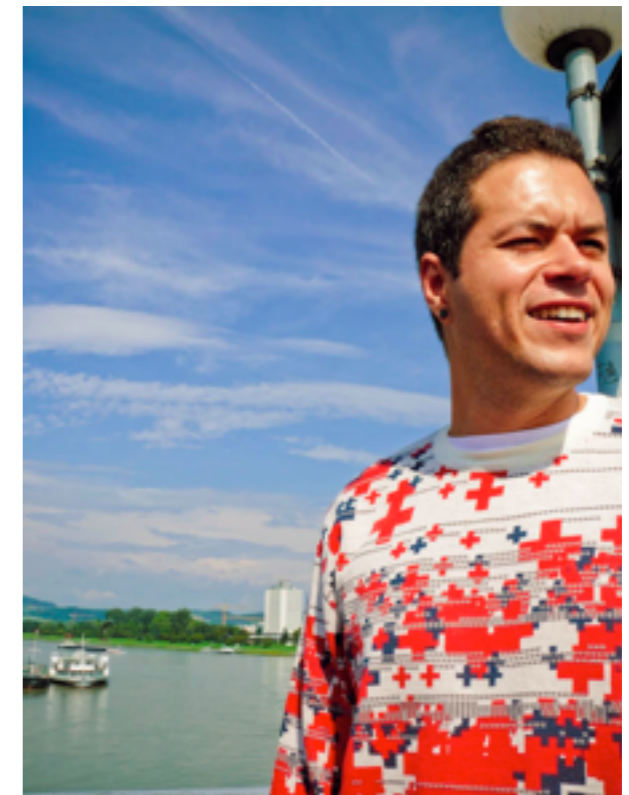
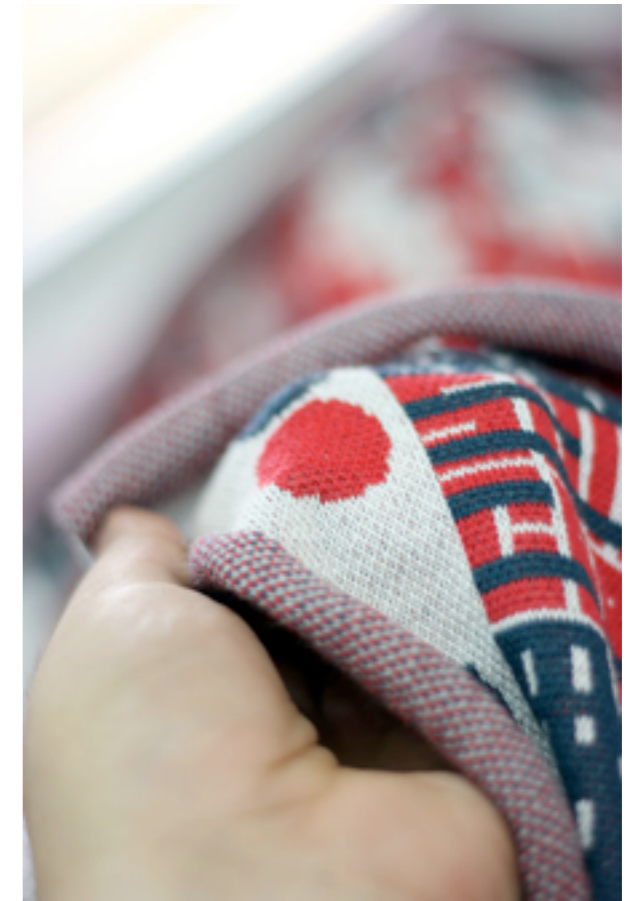
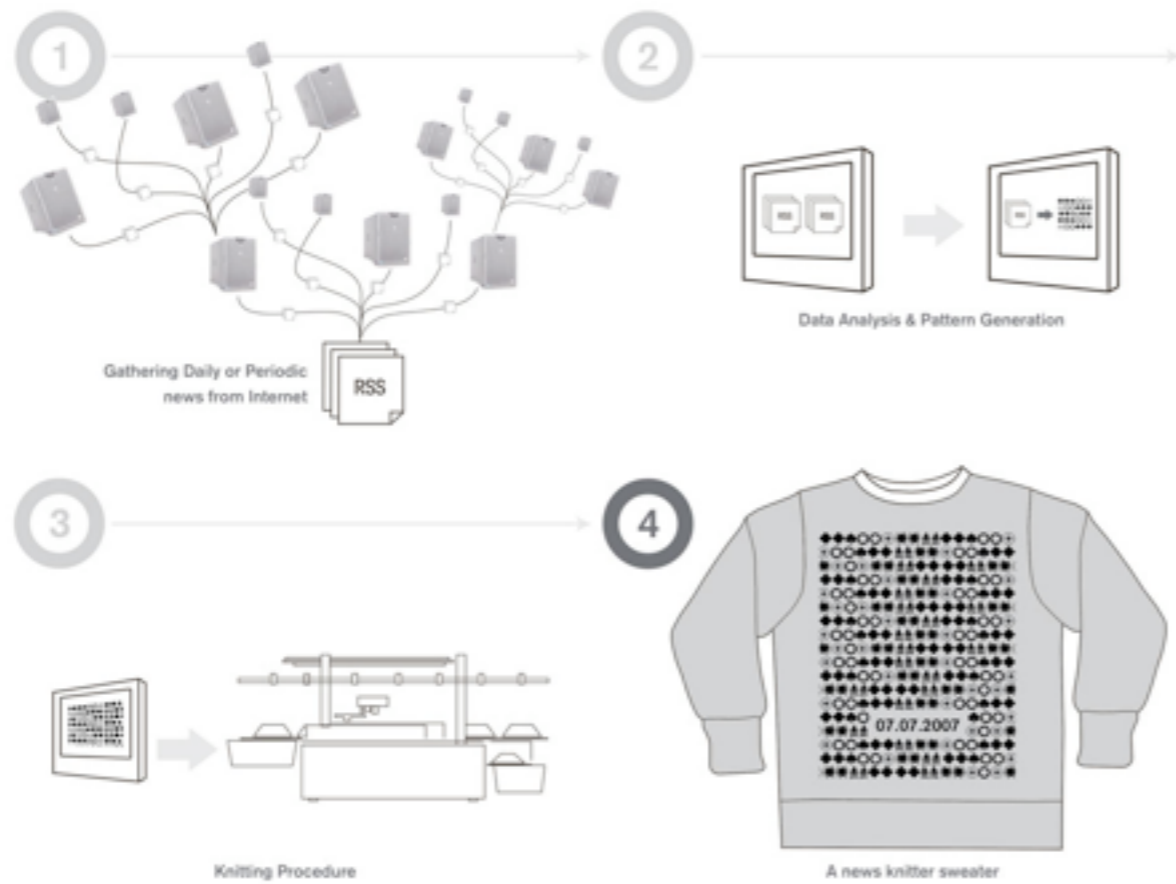


johann volkmer <http://www.faltjahr2010.de/loop.html>



Etienne Cliquet
World Internet Usage Statistics



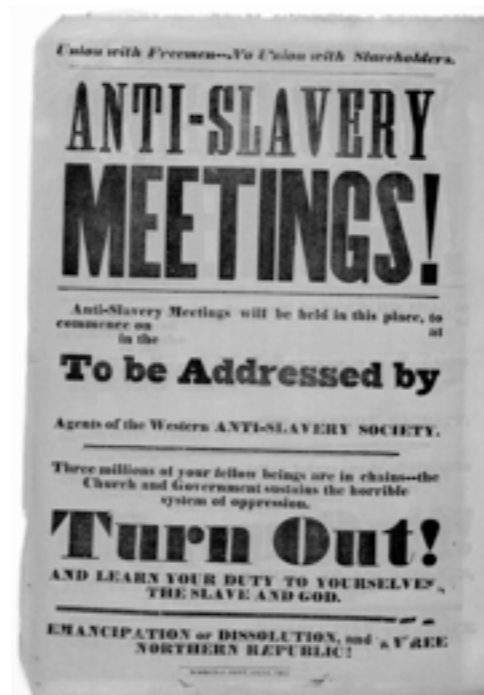


casual data *news knitter* <http://www.casualdata.com/newsknitter/>

Collaboration with Brooklyn Historical Society!



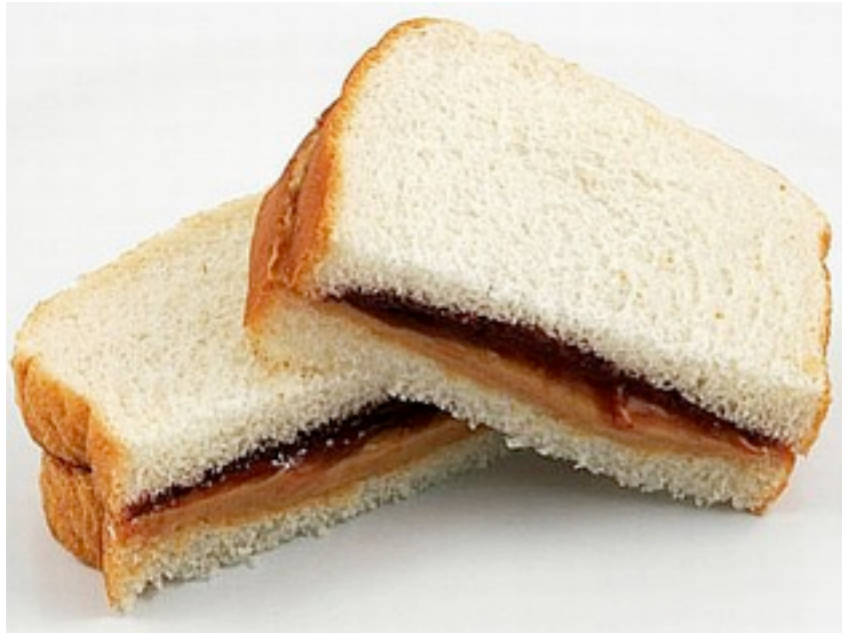
BROOKLYN
HISTORICAL
SOCIETY



Programming!



- When we program, we're **writing** instruction sets for the computer
- The computer will do **EXACTLY** as you say
- Computers are stupid: they will not guess or infer what you mean



YOU need to:

- 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

Functions!

- Out-of-the-box reusable chunks of code! Shared tools or methods that you and the computer have at your fingertips
- You can give functions extra information to make them more flexible and useful
- Let's imagine a function called "openjar" for our sandwich making task, and write instructions for it

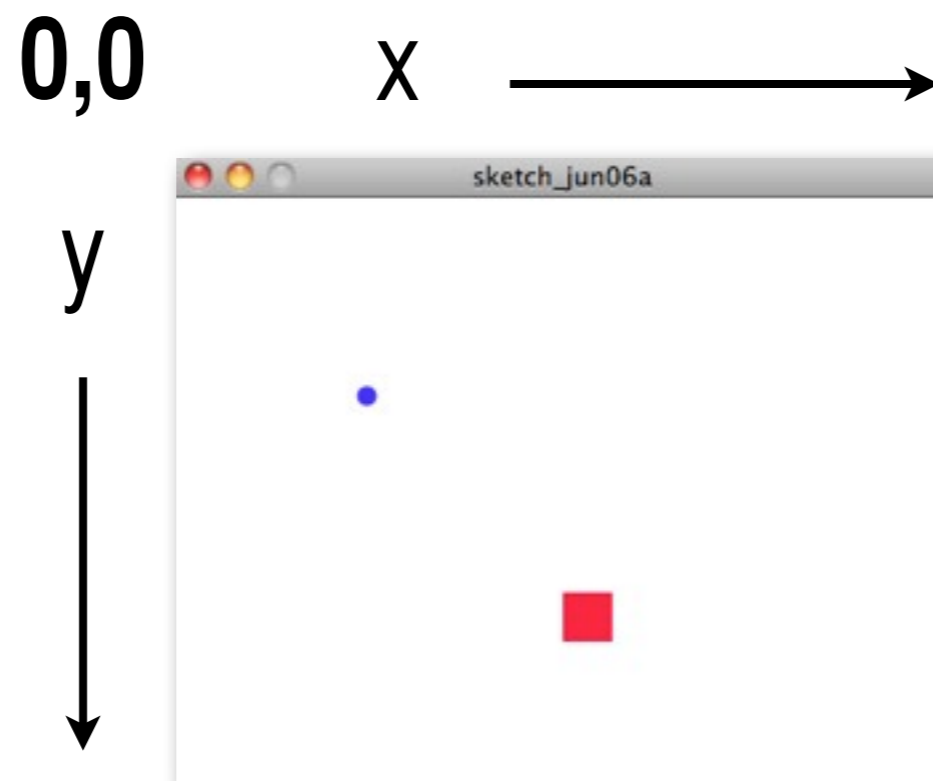
Now, on to Processing!



- Open source programming language for making visuals (text in → visuals out)
- Enables interaction
- Based on Java
- “High Level” Programming Language
- Cross-platform, free!
- download and reference at processing.org, comes with IDE

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)



Processing Syntax

- **Again, computers are dumb**
- Punctuation, spelling and capitalization matters!
- end your lines with “.”
- functions take arguments inside () like this:
openJar(jamJar)

First Functions In Processing

- size(), background()
- rect(), ellipse(), line(), triangle(),
- fill(), stroke()
- colors in B&W or RGB: values from 0 to 255
- alpha values for opacity
- Remember, functions take **arguments!**

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);

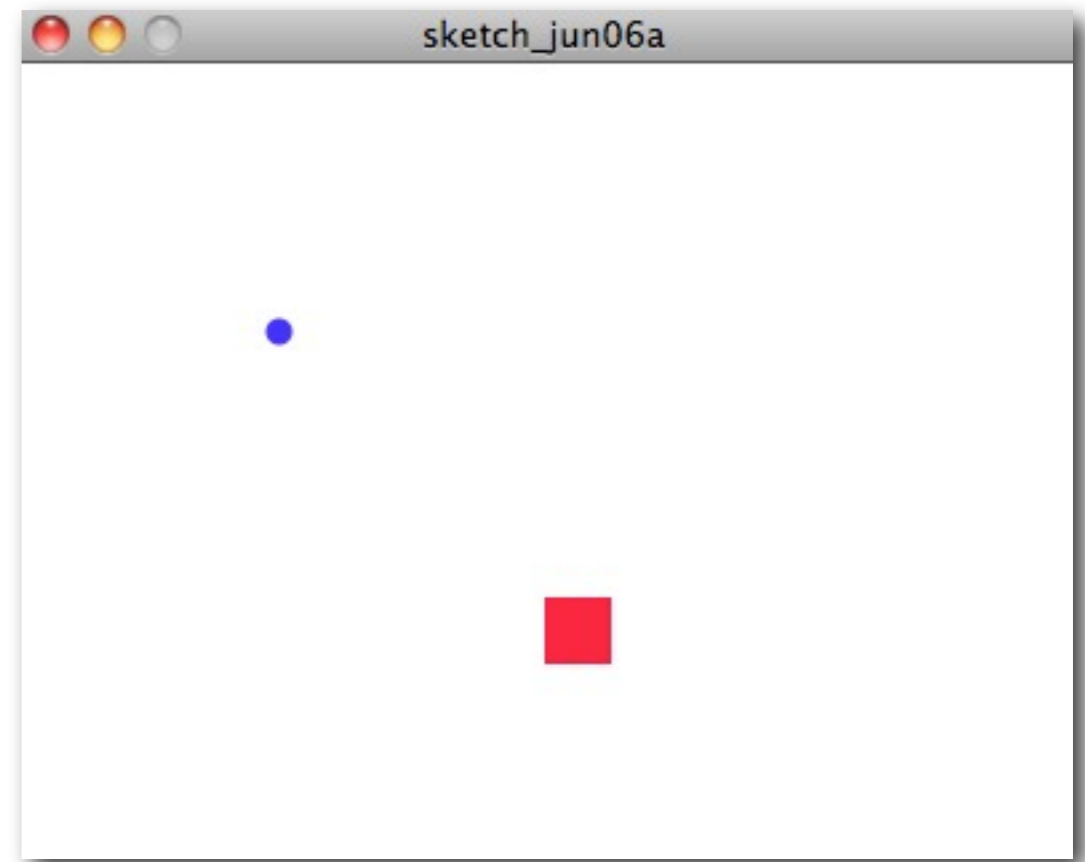
//draw a rectangle at x=200, y=200, with width of 25
//and height of 25
rect(200,200,25, 25);

//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 ↓



Let's Try It!

Using void setup() and void draw()

- Both are functions, neither take parameters
- setup() does preliminary stuff once
- draw() loops over and over again until you close the sketch
- 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** (like a number, a sequence or string of letters that make up a word, or an 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 saved; values held there can be **changed**
- In Processing, variables must be **declared** and **initialized** before you use them!

```
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()
{
  ellipse(200, ypos, 20, 20); //use the variable to draw the ellipse
  ypos += 2; //update the variable
}
```

what will this sketch do?