

Data Visualization

Dec 2017

Contents

- What is data visualization?
- Why bother?
- Design critiques
- Importance of design
- Visual Variables to Charts to Networks to High Dimensional Data

Visualization

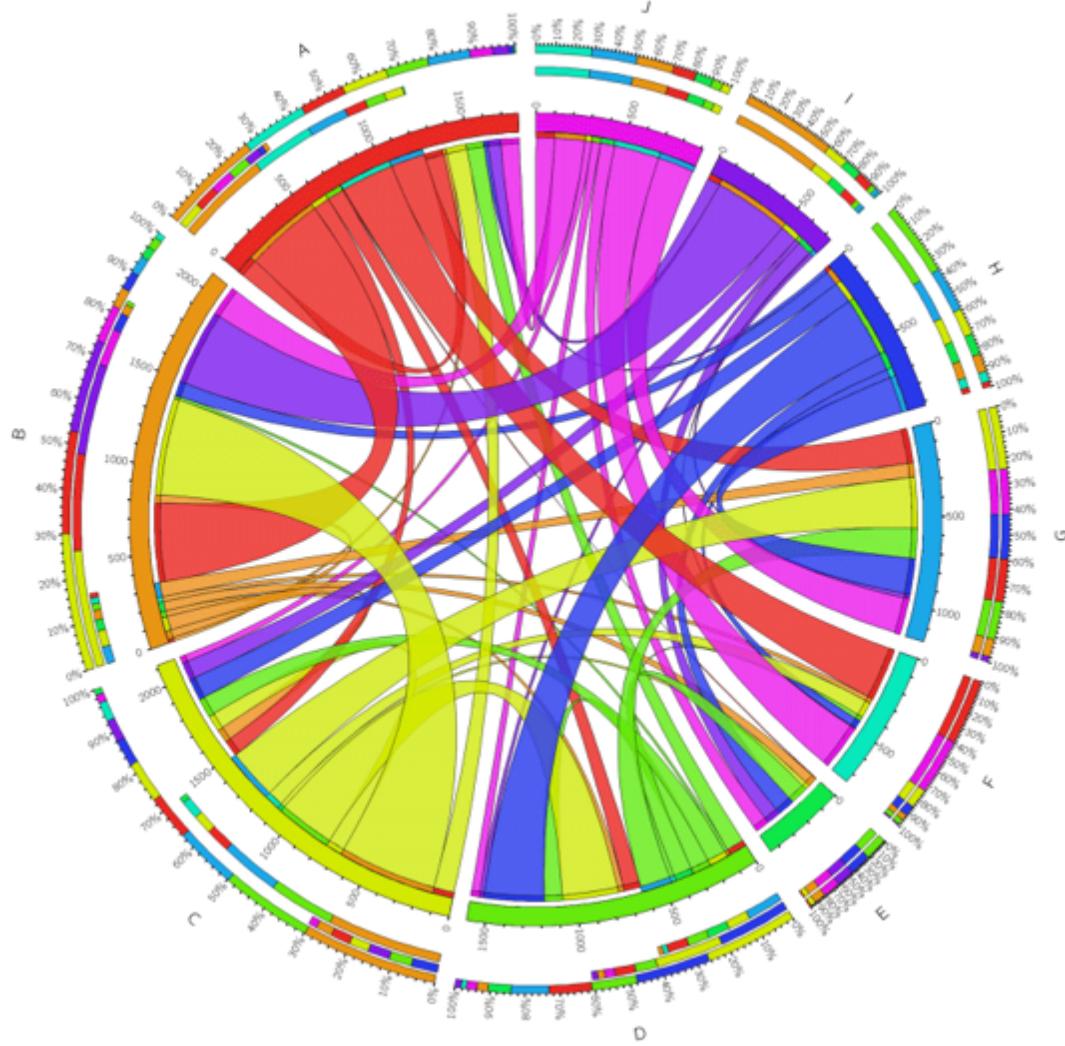
noun vi·su·al·i·za·tion \,vi-zhə-wə-lə-ˈzā-shən, ,vi-zhə-lə-, ,vizh-wə-lə-\

1. formation of mental visual images
2. the act or process of interpreting in visual terms or of putting into visible form
3. the process of making an internal organ or part visible by the introduction (as by swallowing) of a radiopaque substance followed by radiography

Source: Merriam-Webster dictionary

(Data) Visualization

To convey information through graphical representations of data



Goals of visualization

- **Record** information
- **Analyze** data to support decision making
- **Communicate** information to others

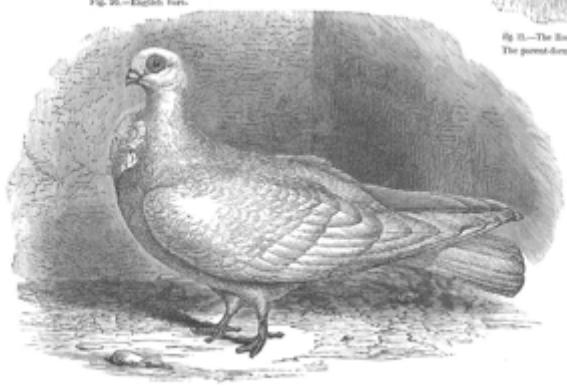
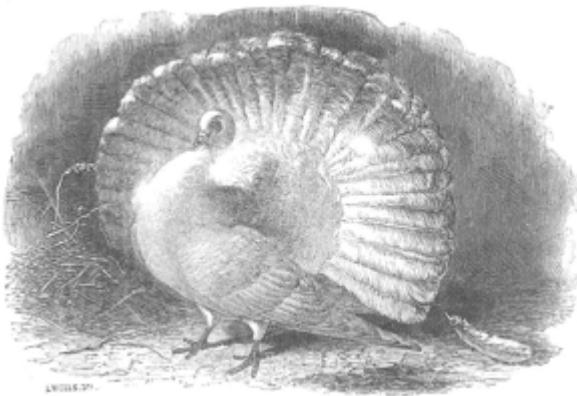
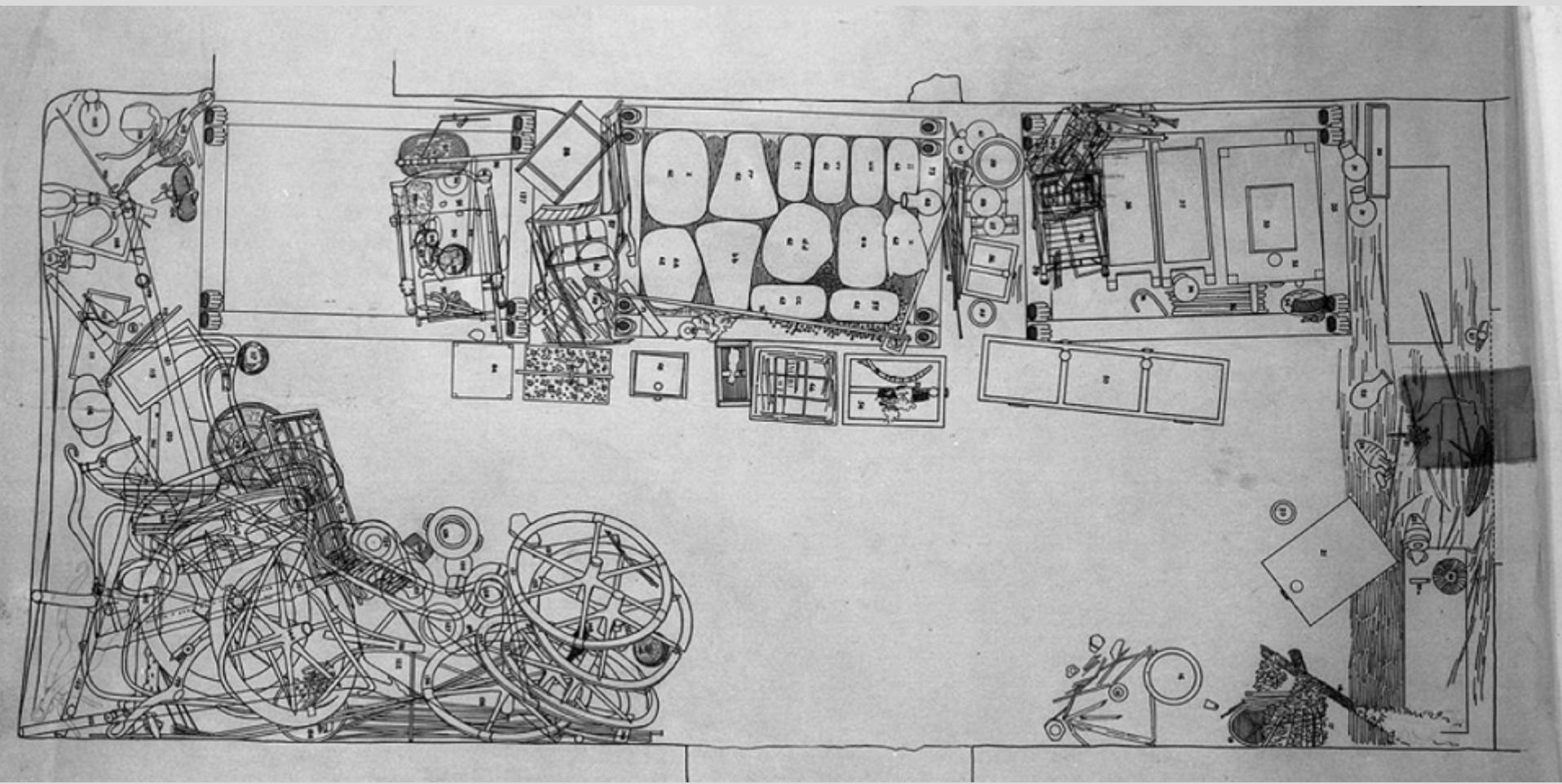


Fig. 22.—African Owl.

Fig. 23.—Blue-faced English Tumbler.



VENUS

MERCUR

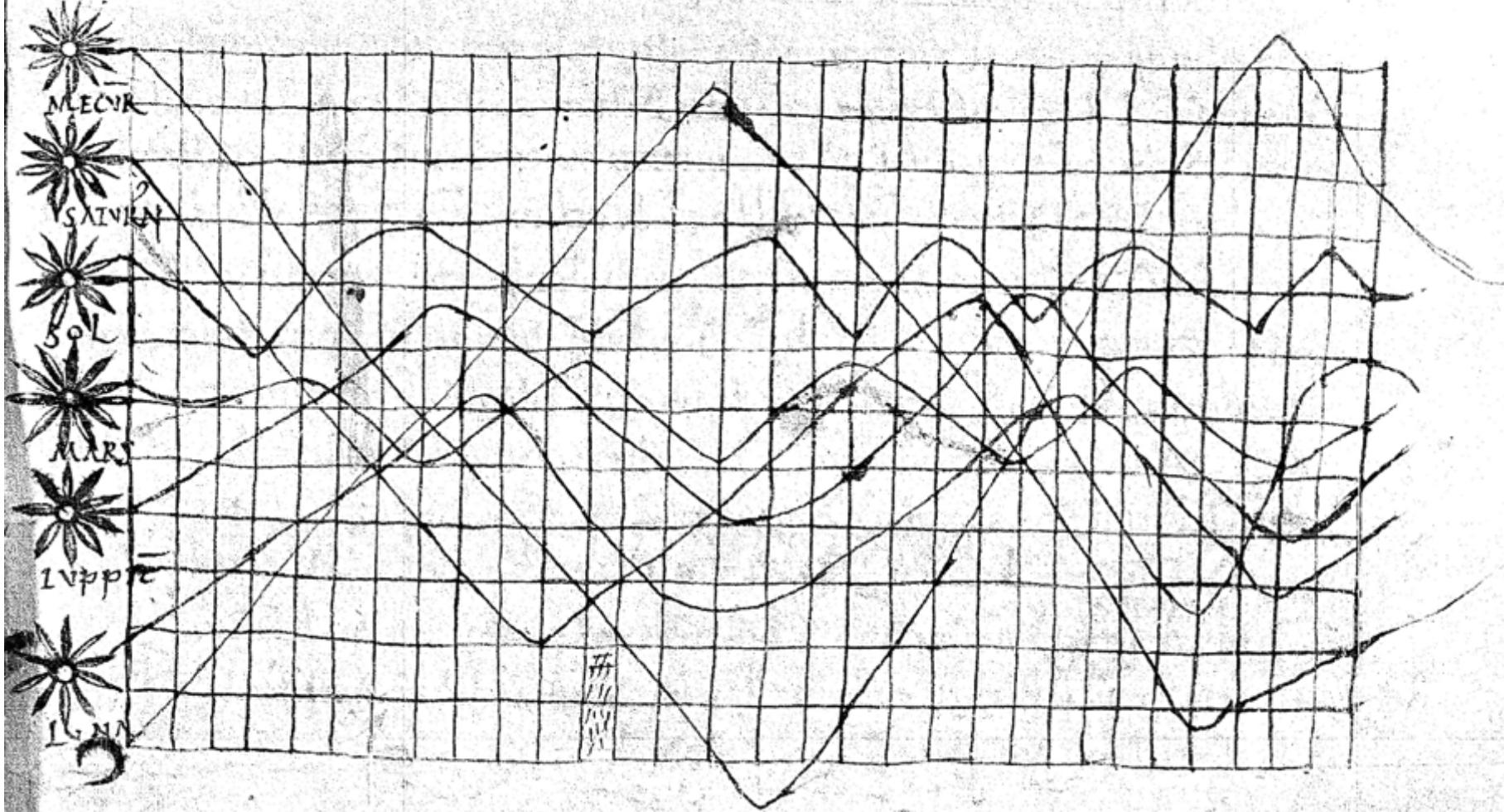
SATURN

SOL

MARS

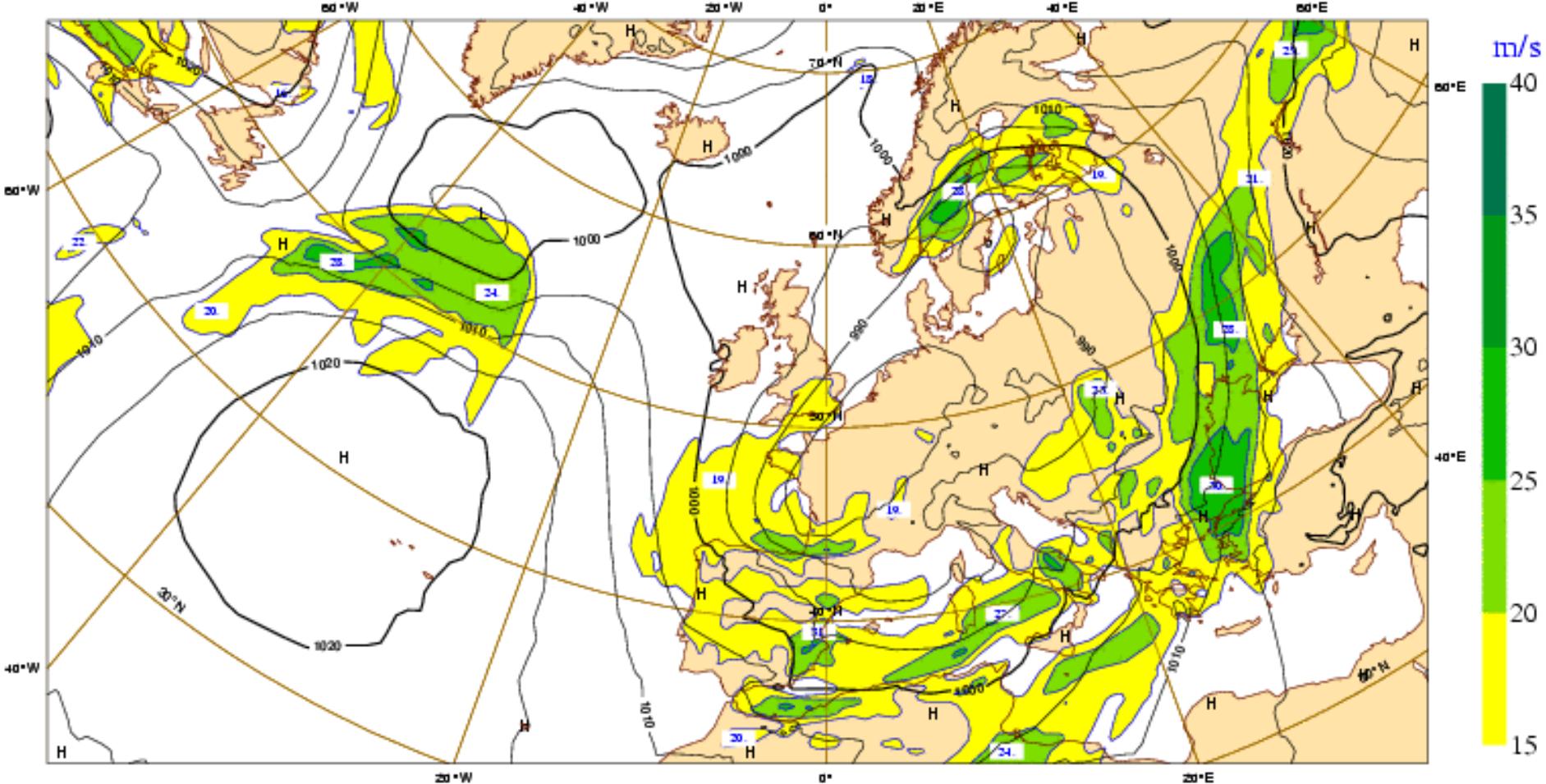
JUPITER

LENA





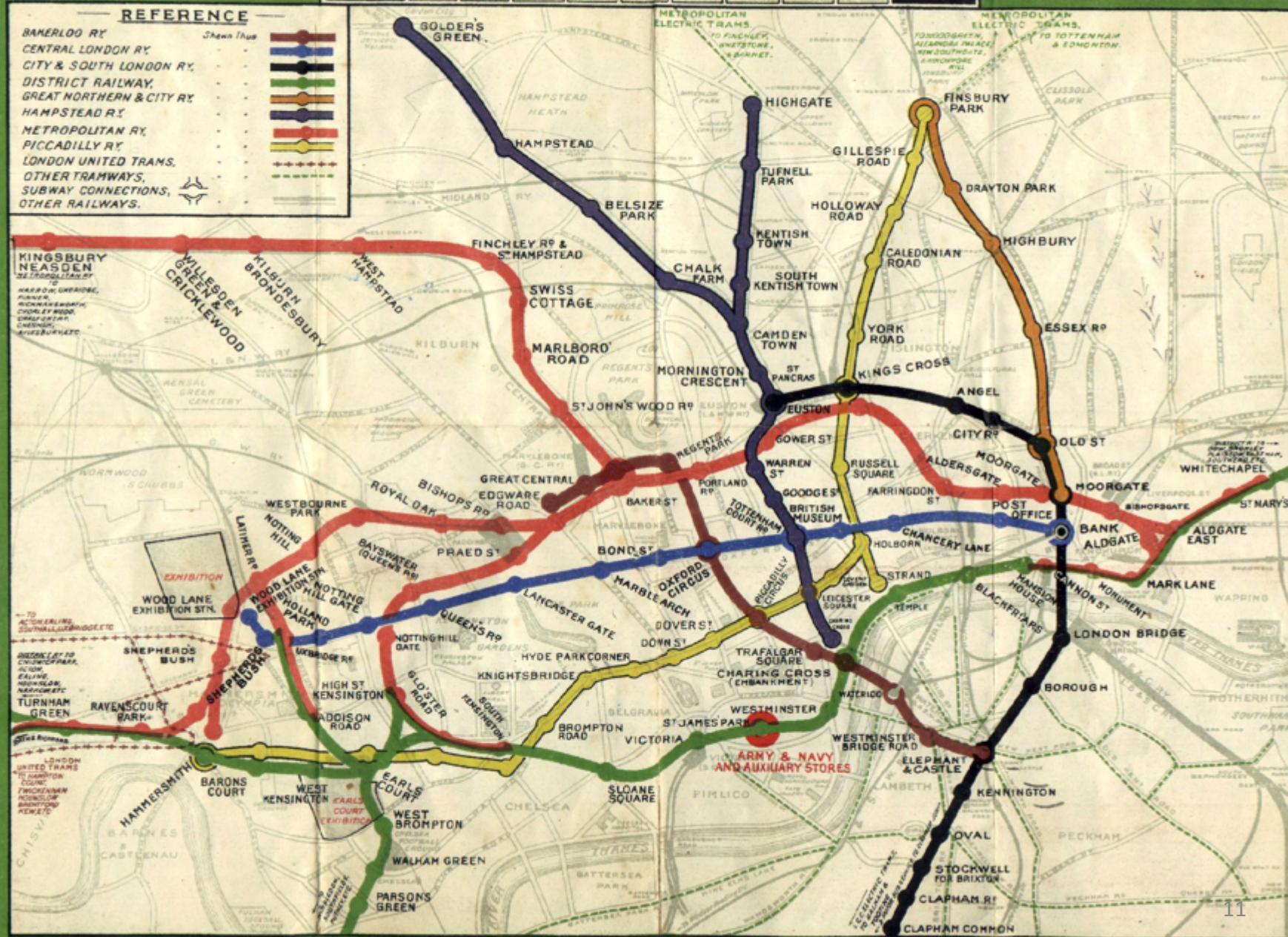
Wednesday 10 November 2010 00UTC ©ECMWF Analysis t+000 VT: Wednesday 10 November 2010 00UTC
Surface: Mean sea level pressure / 850-hPa wind speed



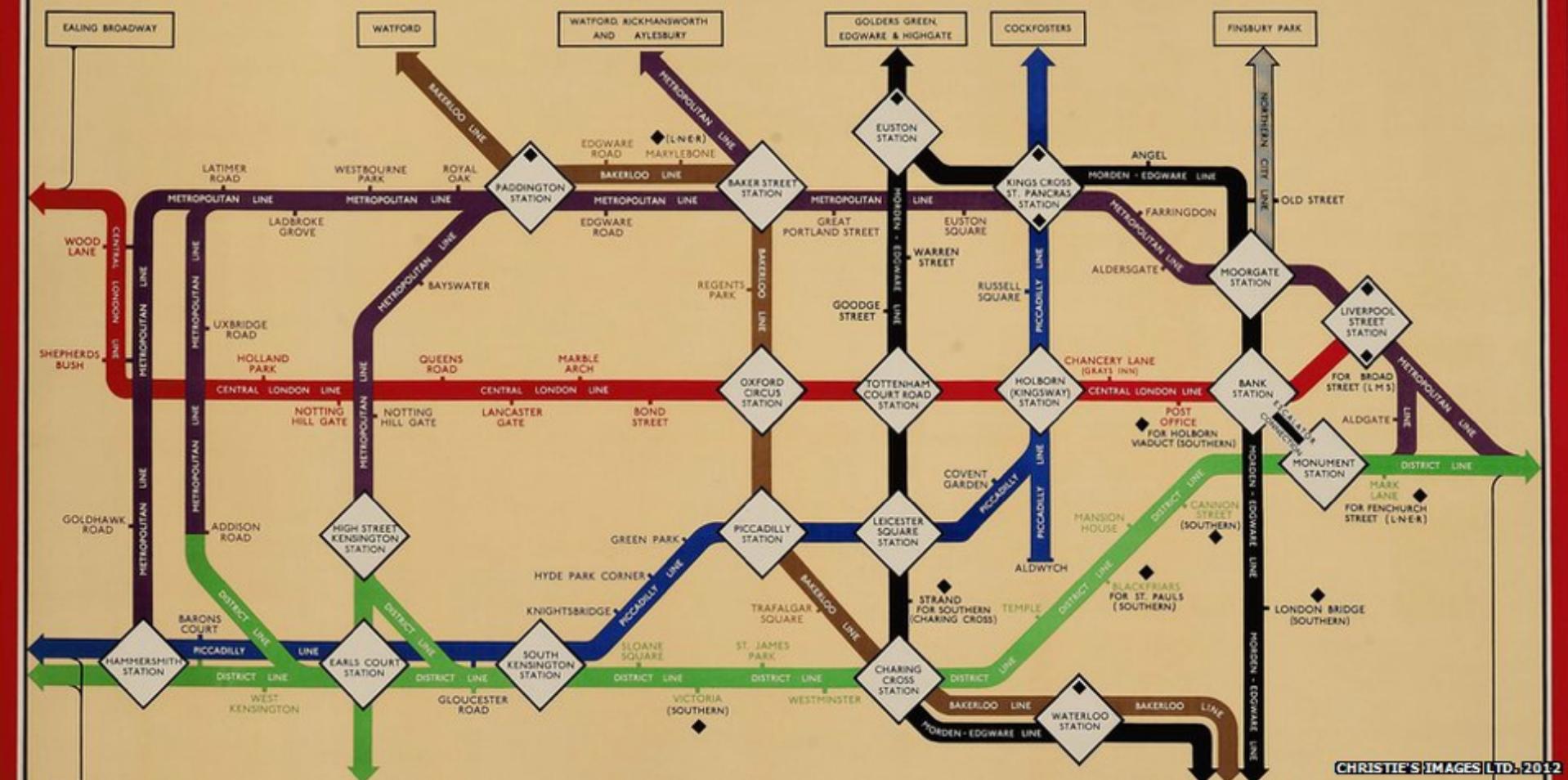
LONDON

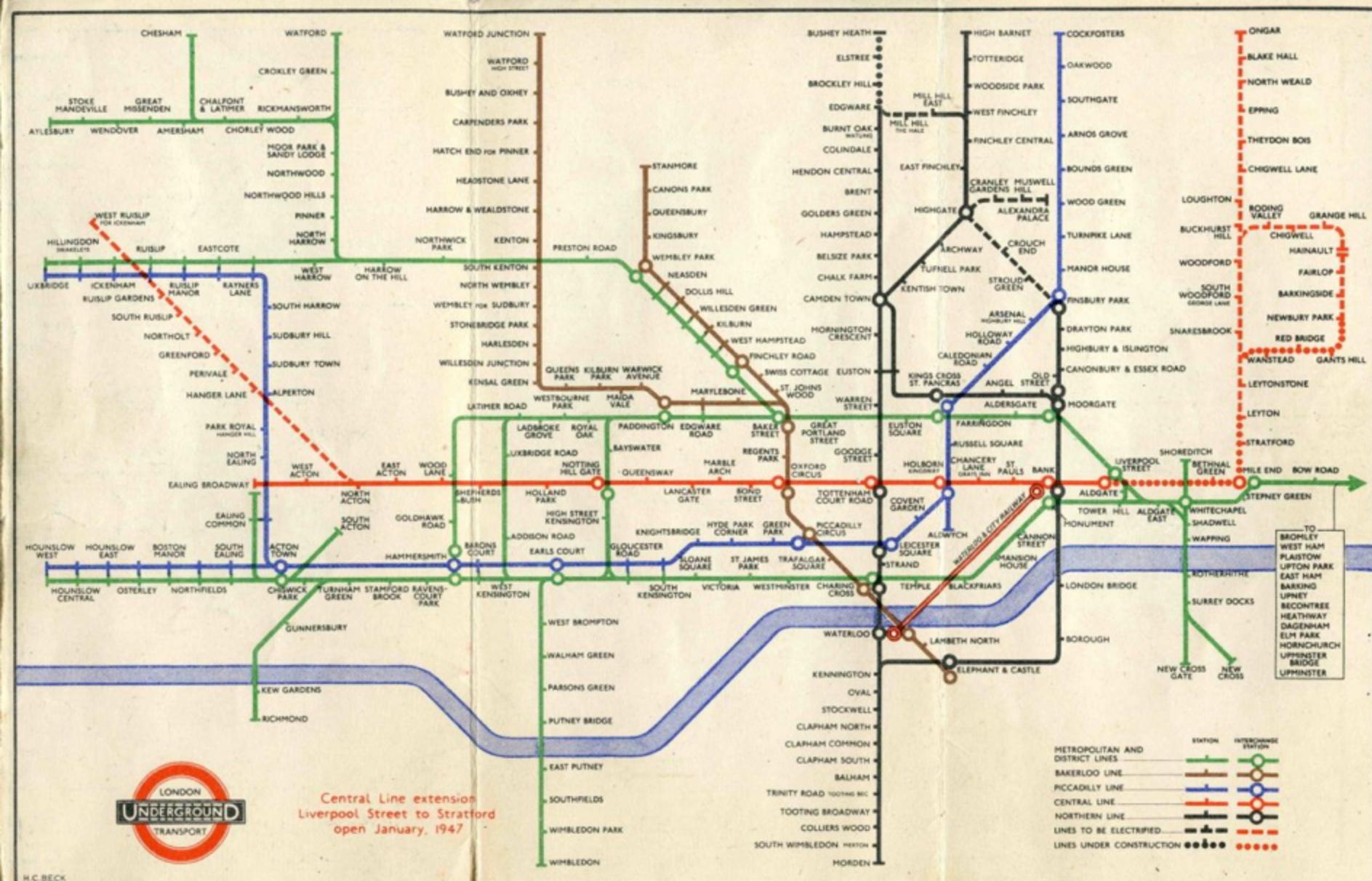
UNDERGROUND

RAILWAYS



CENTRAL AREA MAP





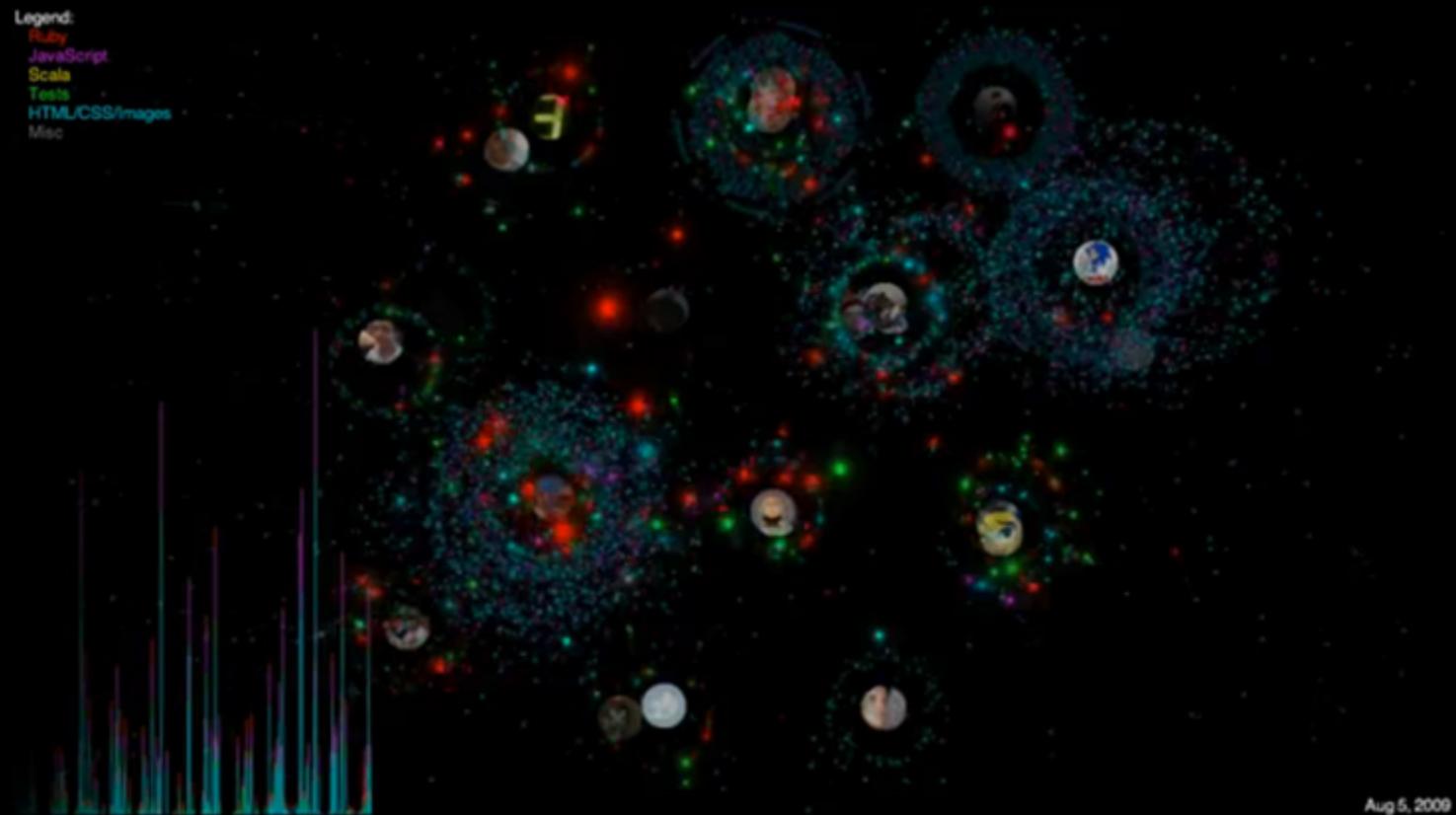


The Best Stats You Have Ever Seen



<https://www.youtube.com/watch?v=usdJgEwMinM>

Storytelling

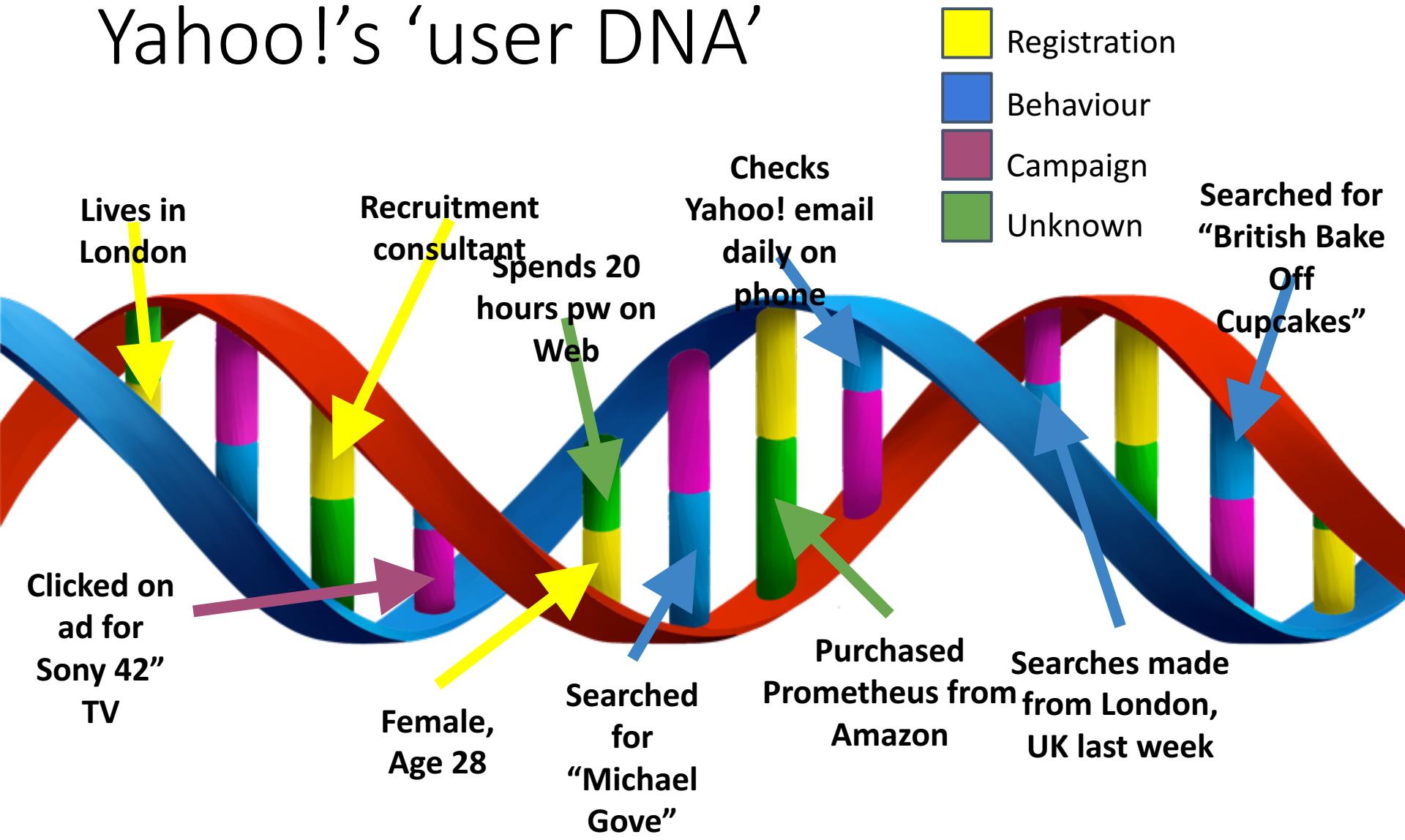


<https://vimeo.com/9225227>

Why data visualization?

- Big data
- Limits to cognition/understanding
- Visual queries are easier to understand (even with summarized statistics, abstract visual cues are fast)

Yahoo!'s 'user DNA'





Registration



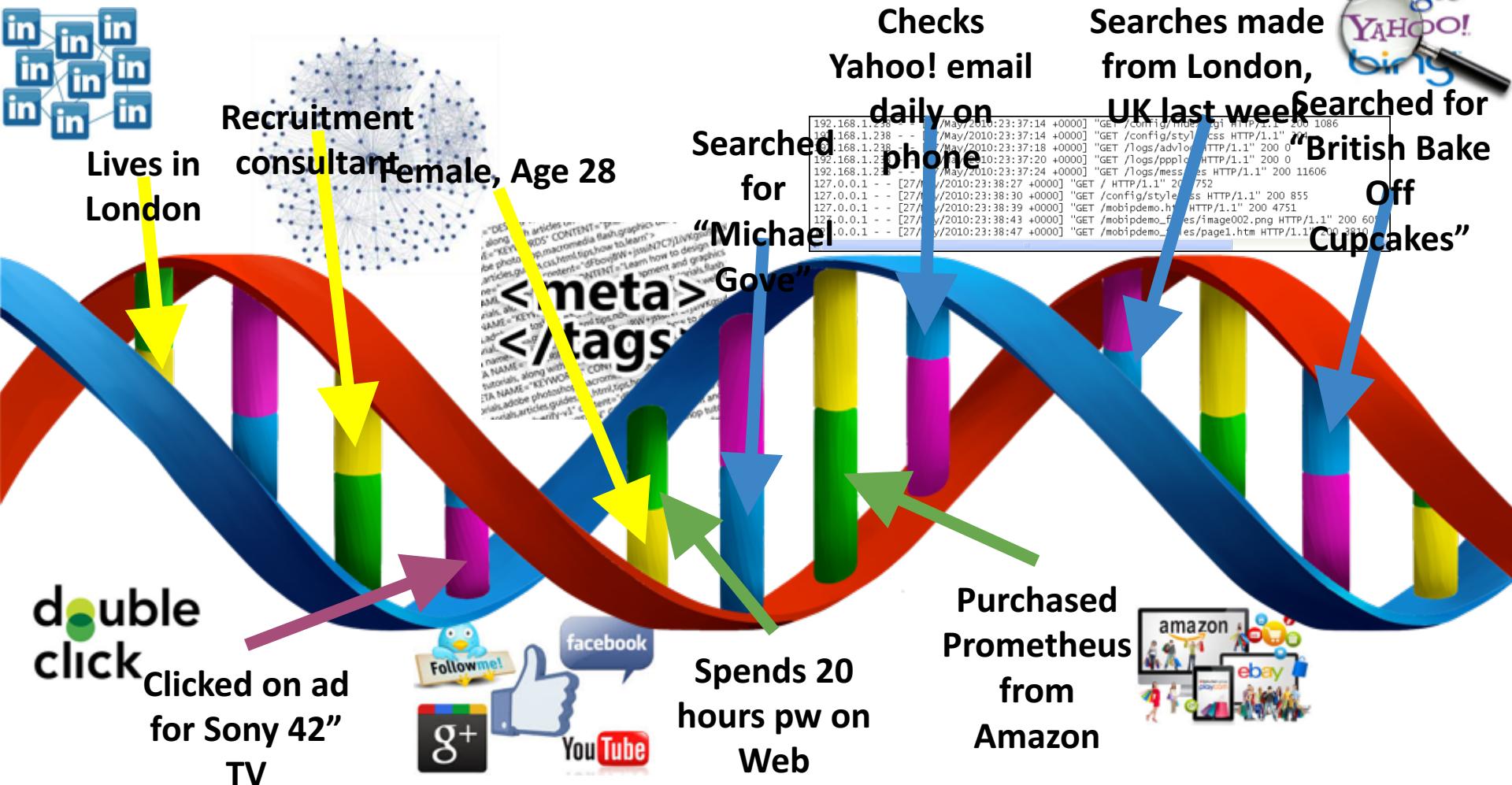
Behaviour



Campaign



Unknown

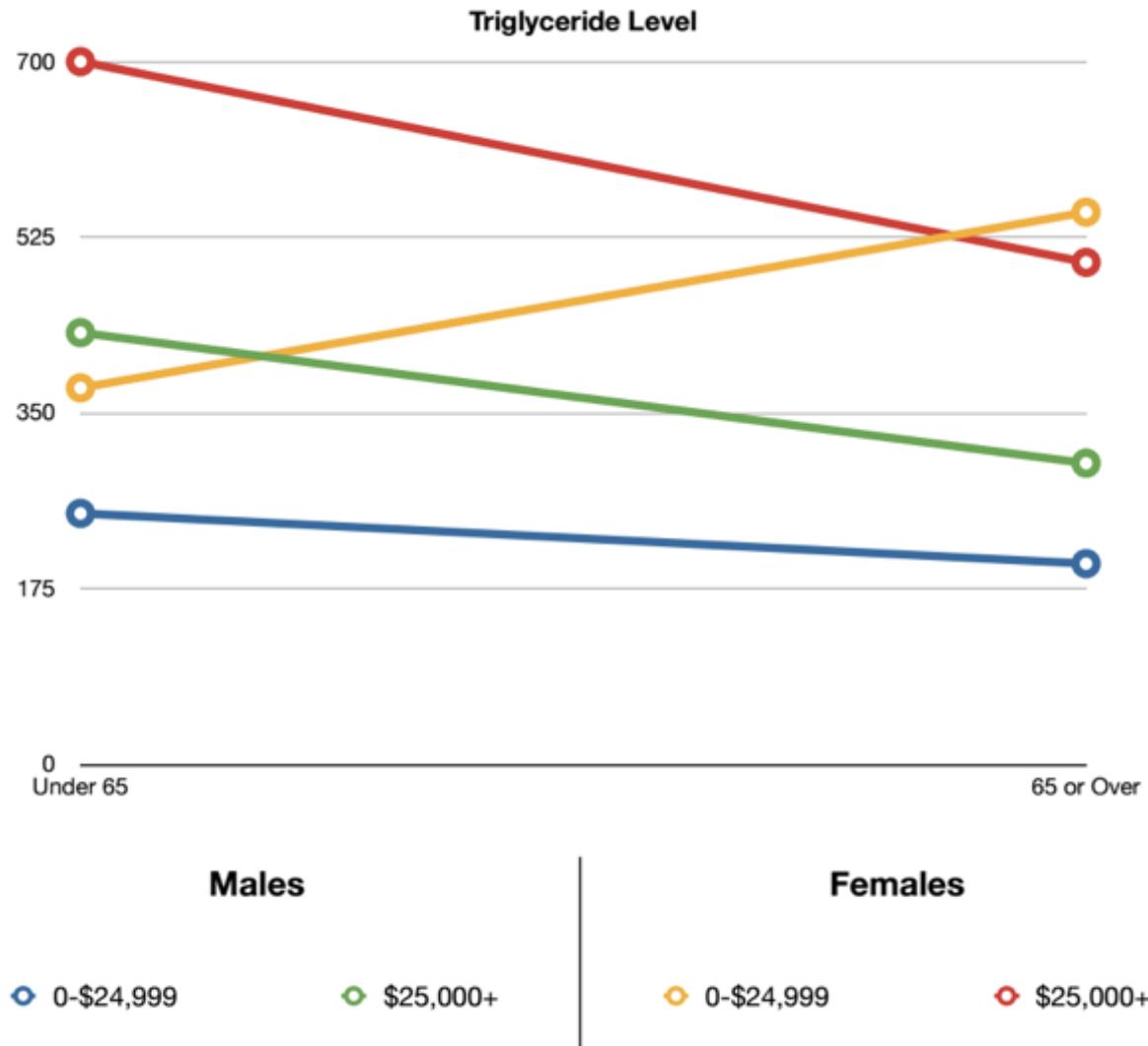


Limits to cognition

- Which gender or income level group shows different effects of age on cholesterol levels?

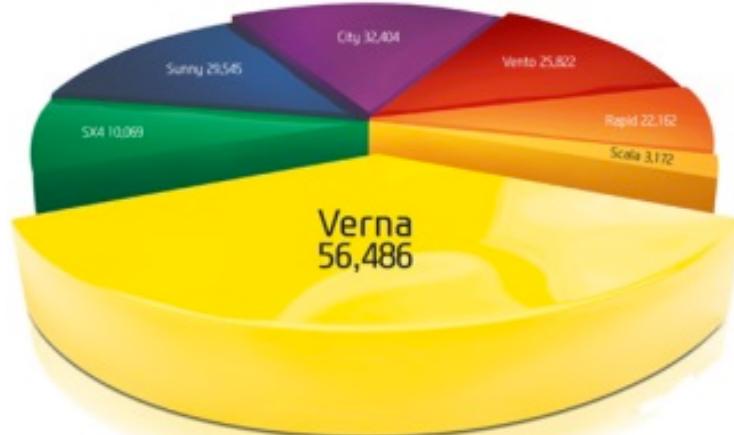
| Income Group | Males | | Females | |
|--------------|----------|------------|----------|------------|
| | Under 65 | 65 or Over | Under 65 | 65 or Over |
| 0-\$24,999 | 250 | 200 | 375 | 550 |
| \$25,000+ | 430 | 300 | 700 | 500 |

Limits to cognition



Critique

Critique



JAN-DEC 2012
Cumulative Sales Figures
Source: SIAM Data

The new FLUIDIC
VERNA
It sets you thinking

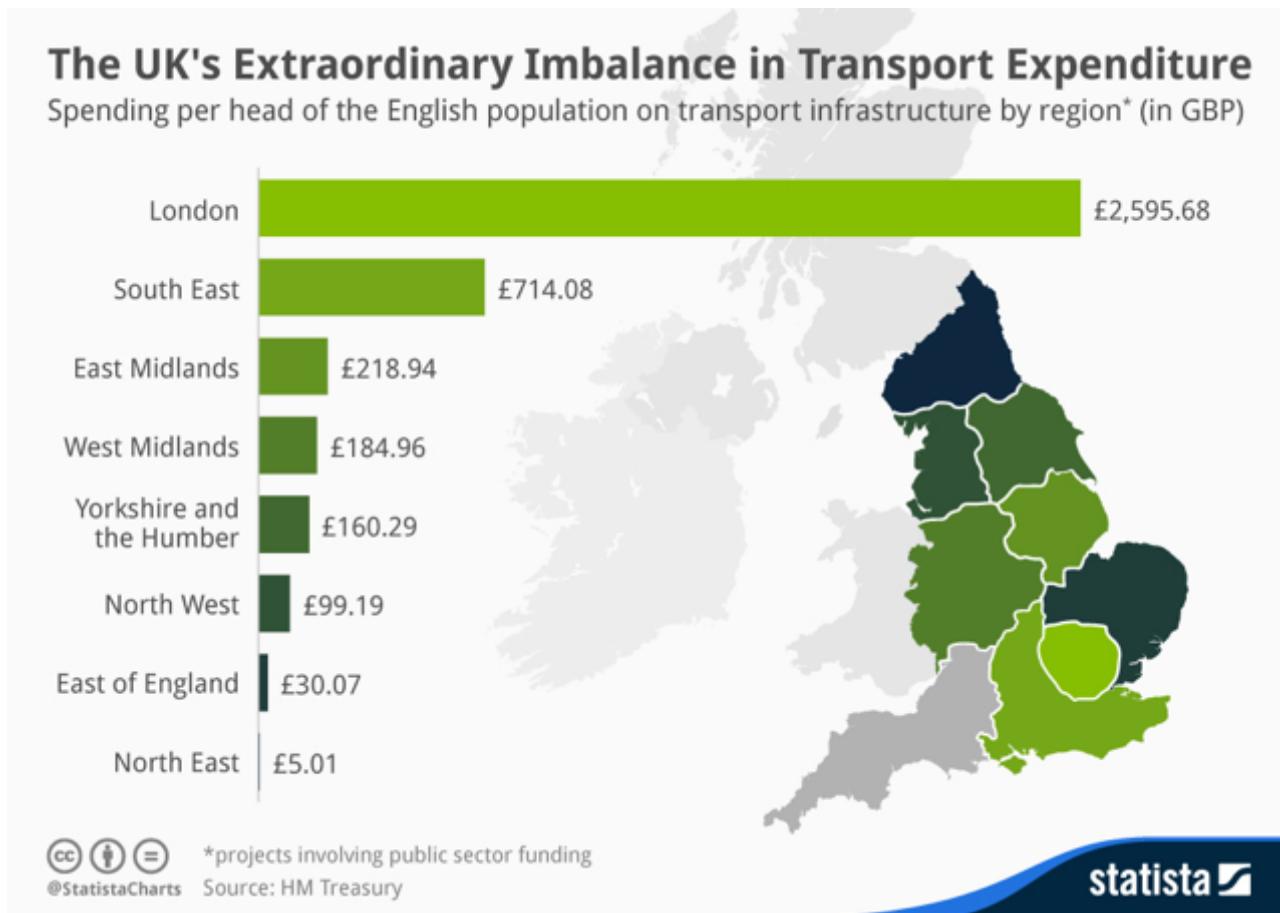
The chartbuster rules.

The Verna stays the undisputed No. 1 in its segment.

With its inimitable styling and superior design, the Verna has emerged as the largest selling car in its segment by a large margin. And it's not just the car that's made us the leader; it's also discerning people like you. No wonder the competition's been left behind. Far far behind.



Critique



Critique



Critique

Click a Category

[Clear](#)

- Automobile
- Clothing and Personal
- Discretionary
- Electric
- Entertainment
- Fees
- Groceries
- Home
- Insurance
- Misc
- Phone
- Restaurant
- Water
- Wireless

Month Range



My Alerts

Today

- [My Fund Fact Sheet 9/17](#)
- [Stock screening criteria has identified IBM. Get a quote.](#)
- [Investment & Spend Summary](#)
- [Home Equity Line of Credit New Rate - 6% for 30k \(3 hrs ago\)](#)
- [My Credit Card Statement](#)

My Merchants



My Spend By Payment Method



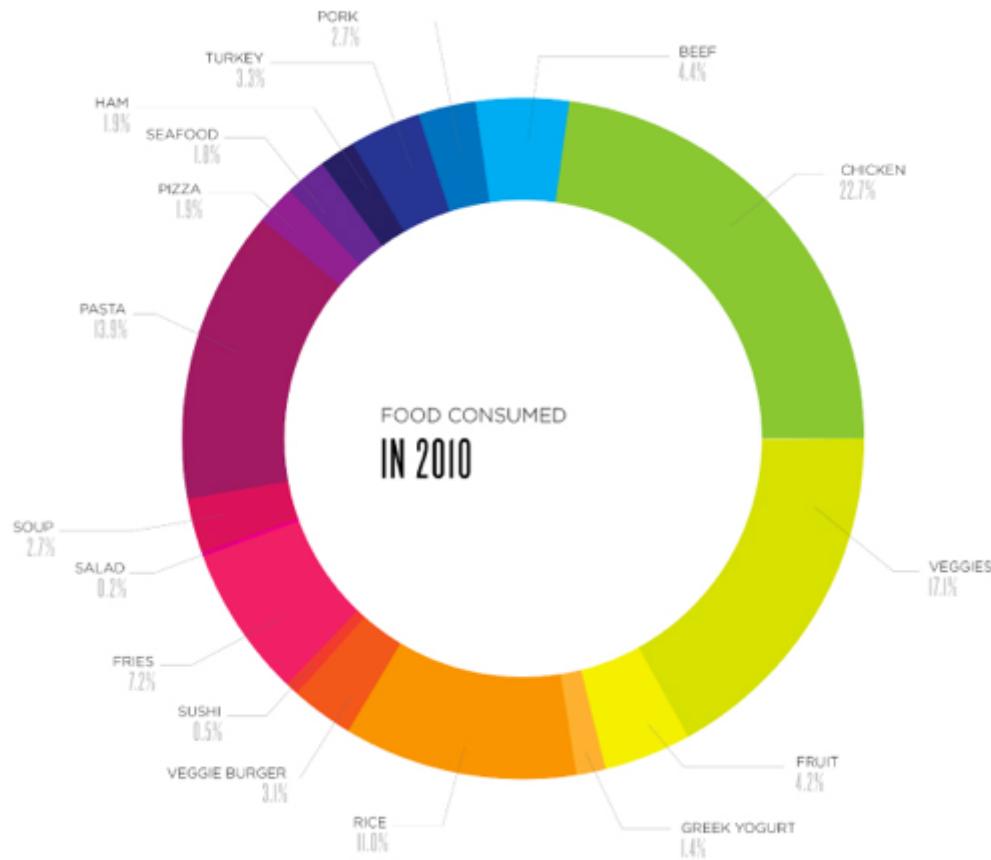
My Spend This Period



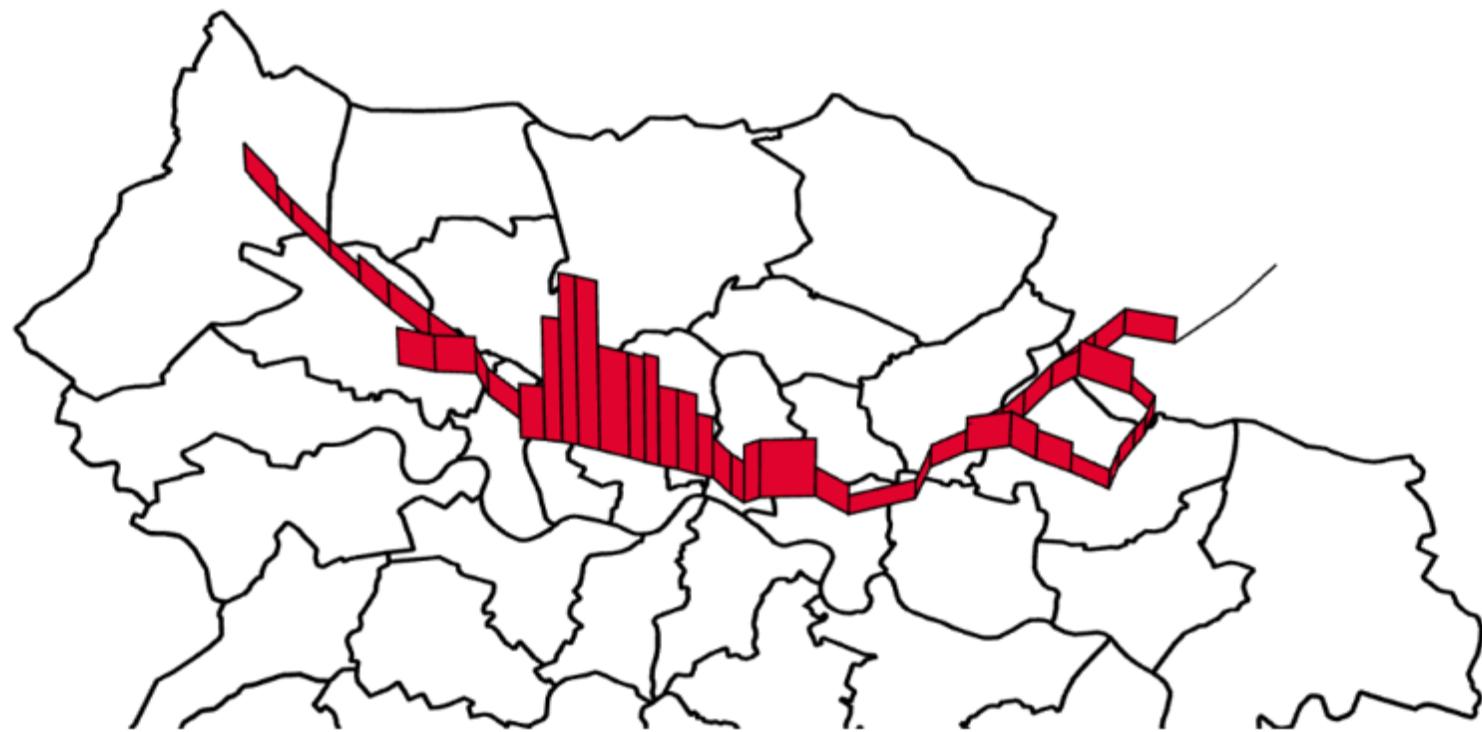
My Spend By Account



Critique



Critique



Importance of Design

LINE

A line is a mark between two points. There are various types of lines, from straight to squiggly to curved and more. Lines can be used for a wide range of purposes: stressing a word or phrase, connecting content to one another, creating patterns and much more.

COLOR

Color is used to generate emotions, define importance, create visual interest and more. CMYK (cyan/magenta/yellow/black) is subtractive; RGB (red/green/blue) is additive.

Some colors are warm and active (orange, red); some are cool and passive (blue, purple).

There are various color types (primary to analogous) and relationships (monochromatic to triad) worth learning more about as well.

TEXTURE

Texture relates to the surface of an object; the look or feel of it. Concrete has a rough texture; drywall has a smooth and subtle texture. Using texture in design is a great way to add depth and visual interest. Printed material has actual, textile texture while screen material has implied texture.

SHAPE

Height + width = shape. We all learned basic shapes in grade school - triangles, squares, circles and rectangles. Odd or lesser seen shapes can be used to attract attention.

There are three basic types of shape: geometric (triangles, squares, circles etc), natural (flowers, animals, trees, people), and abstracted (icons, stylizations, graphic representations etc).

VALUE

Value is how light or how dark an area looks. A gradient, shown above, is a great way to visualize value - everything from dark to white, all the shades in-between, has a value. Use value to create depth and light; to create a pattern; to lead the eye; or to emphasize.

ELEMENTS OF DESIGN
quick reference sheet

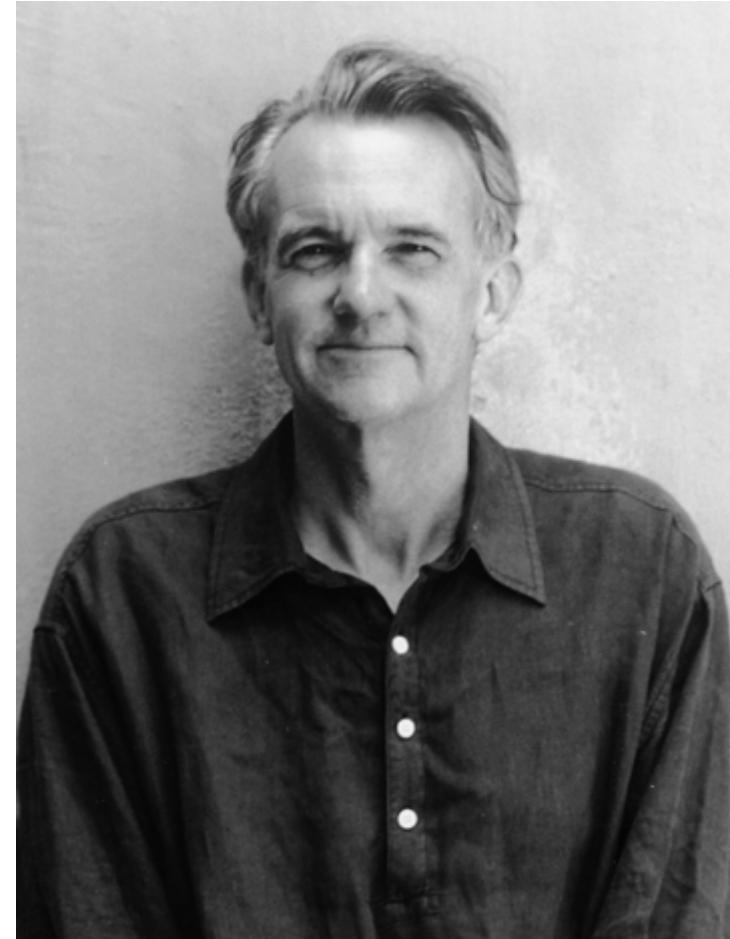
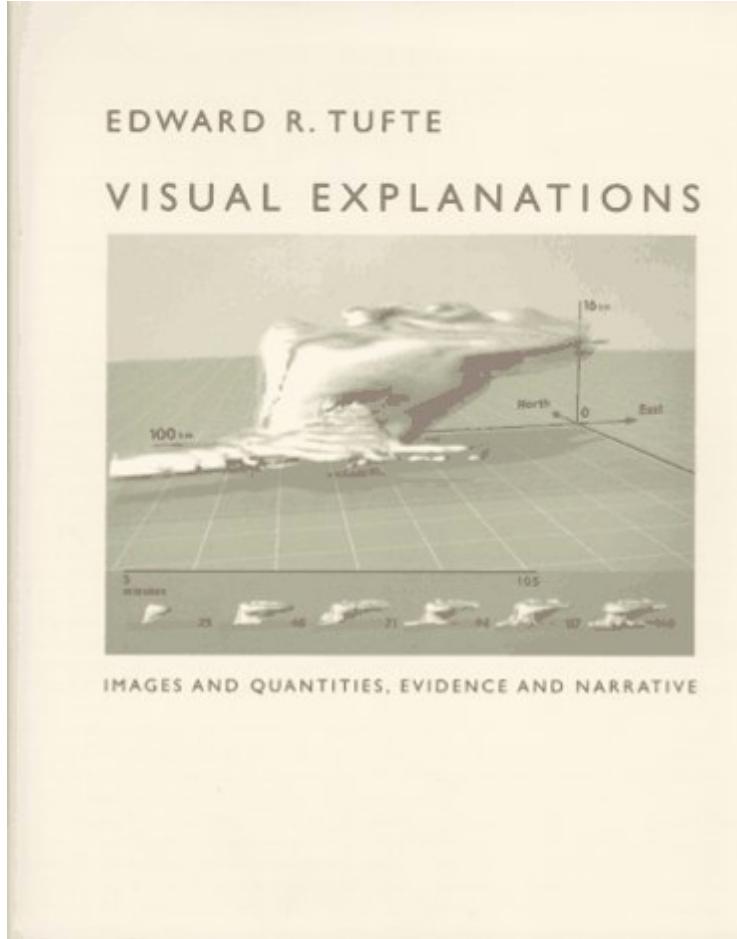
SIZE

Size is how small or large something is: a small shirt vs. an extra large shirt, for example. Use size to define importance, create visual interest in a design (via contrasting sizes), attract attention and more.

SPACE

Space is the area around or between elements in a design. It can be used to separate or group information. Use it effectively to give the eye a rest; define importance; lead the eye through a design and more.

Edward Tufte

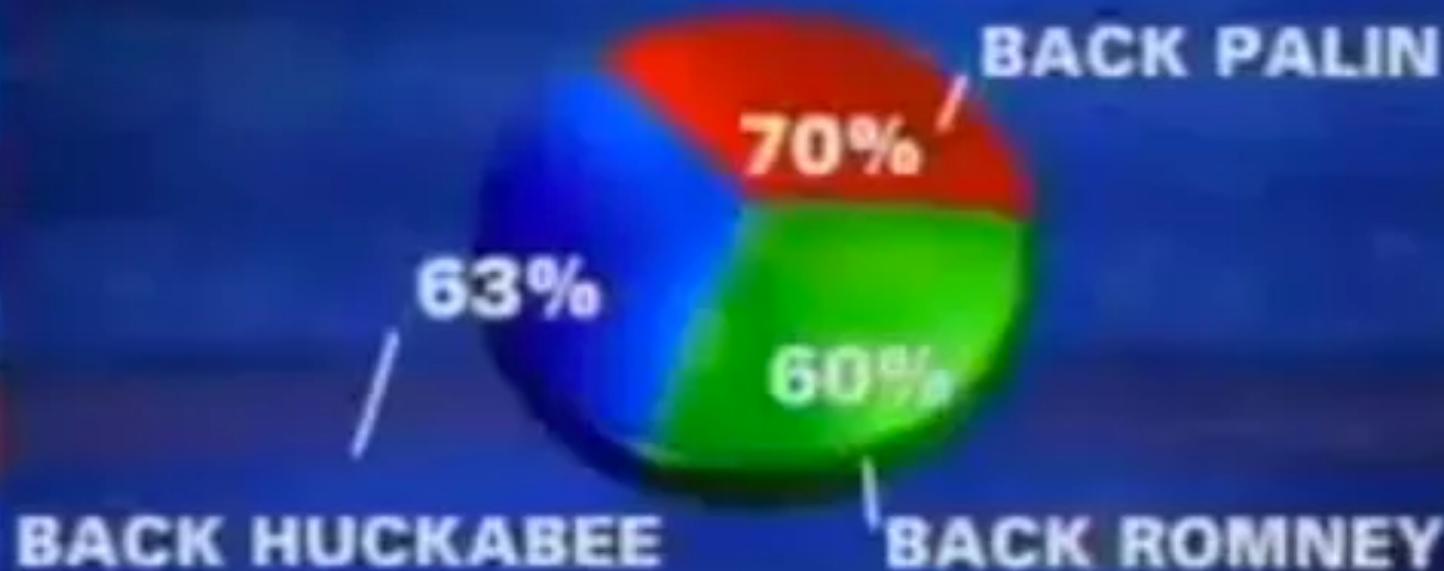


Design Principles

- Graphical integrity
 - Always tell the truth about the data
 - Use clear, detailed labels to defeat distortion and ambiguity
 - Show data variation, not design variation
- Design aesthetics
 - Be clear and precise
 - Maximize data-ink ratio (i.e. reduce redundant display of information)
 - Avoid chartjunk
 - Maximize data density

2012 PRESIDENTIAL RUN

GOP CANDIDATES

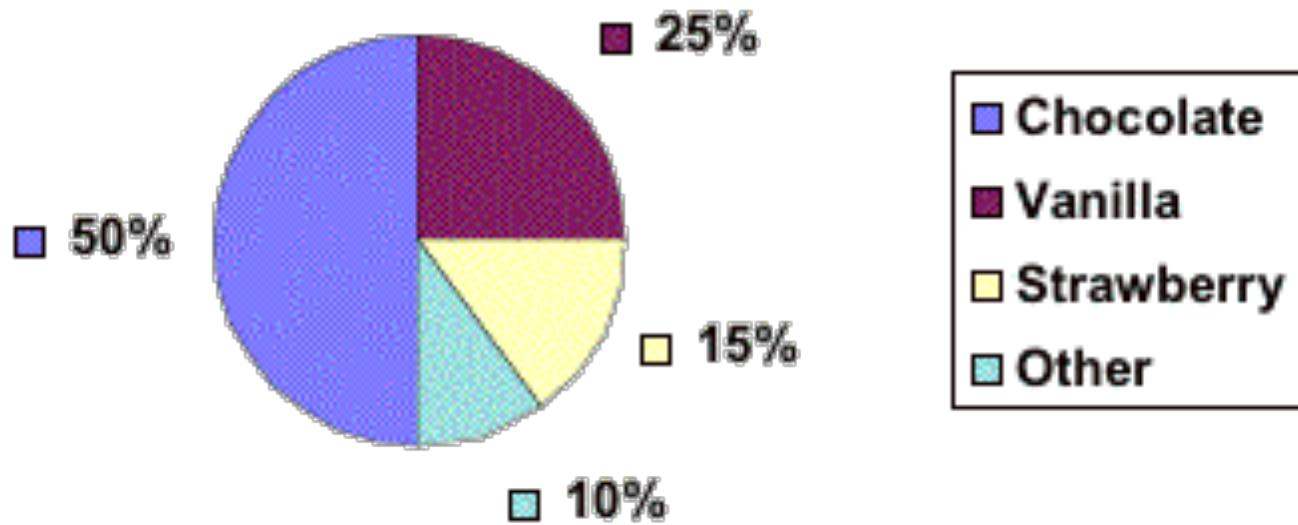


SOURCE: OPINIONS

DYNAMIC

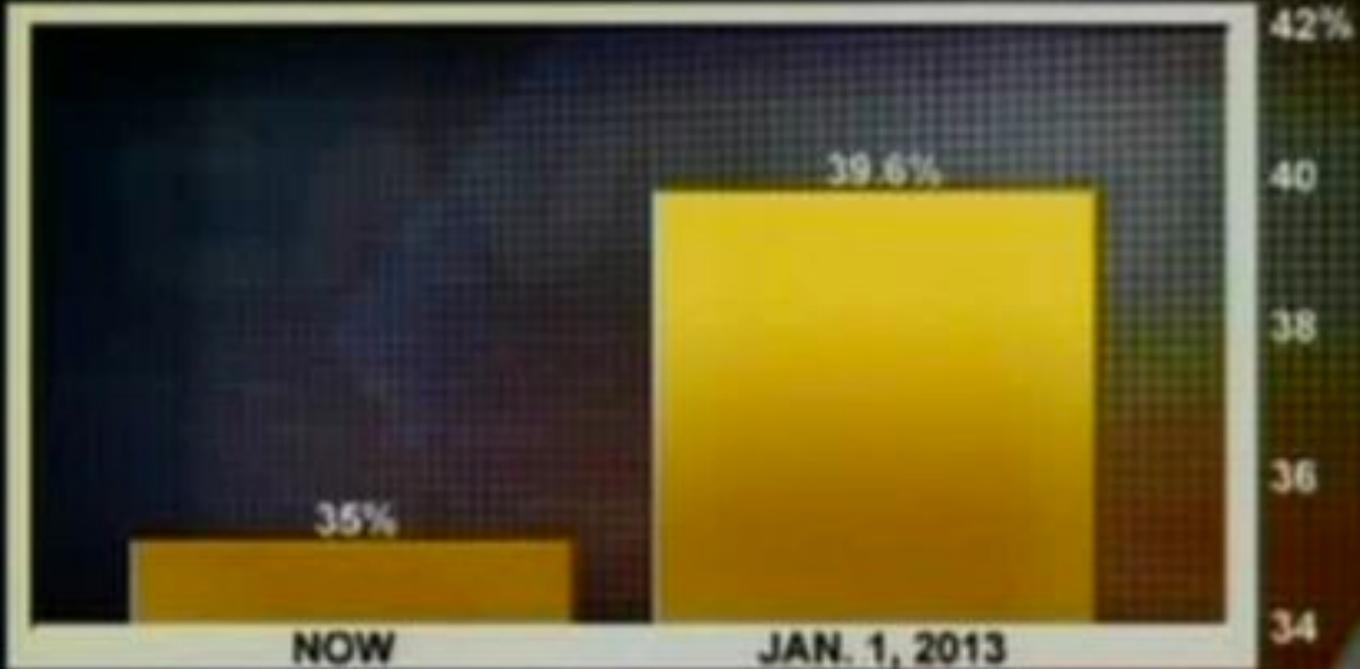
Graphical integrity

Favorite Ice Cream



IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01 p ET

FOX
BUSINESS

TOP STORIES TECHNOLOGY CONSUMER
WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

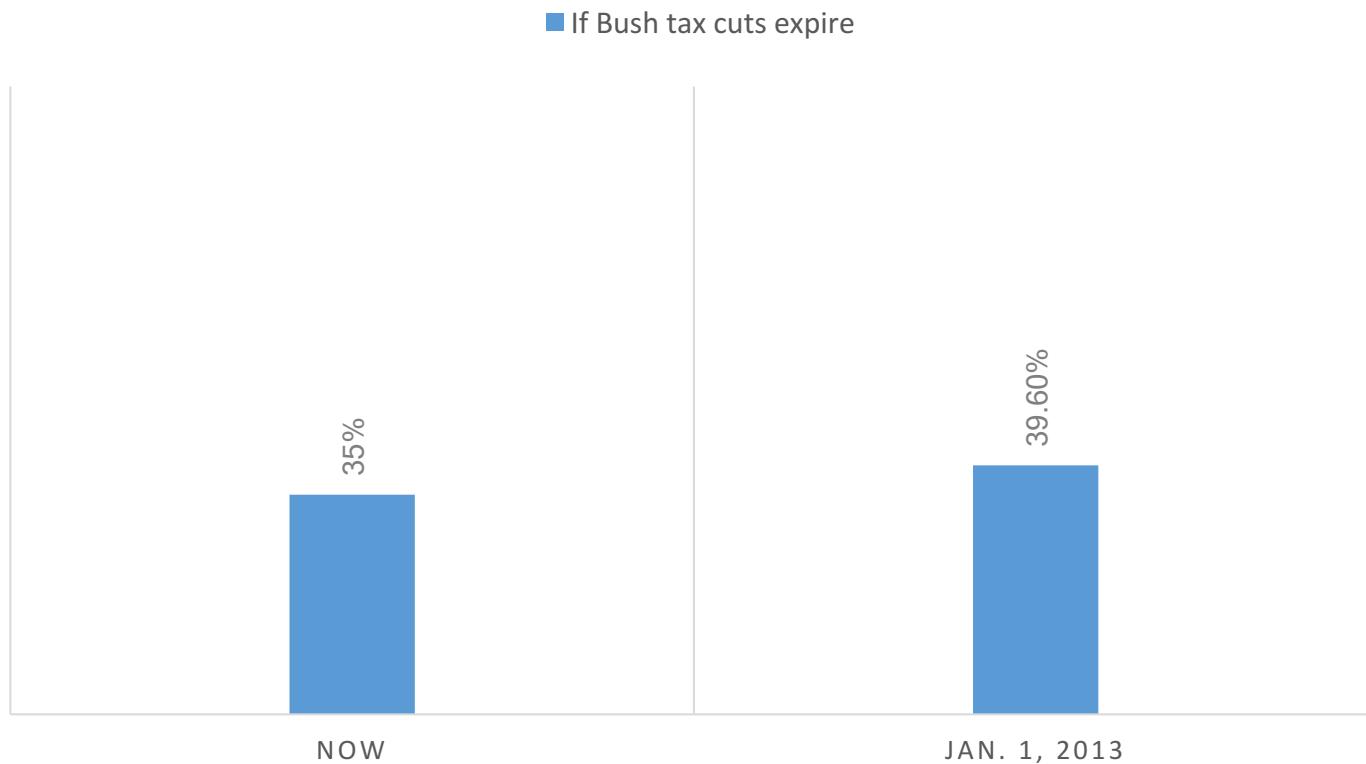
DOW 13008.68 ▲ 64.33

S&P 1379.32 ▲ 5.98

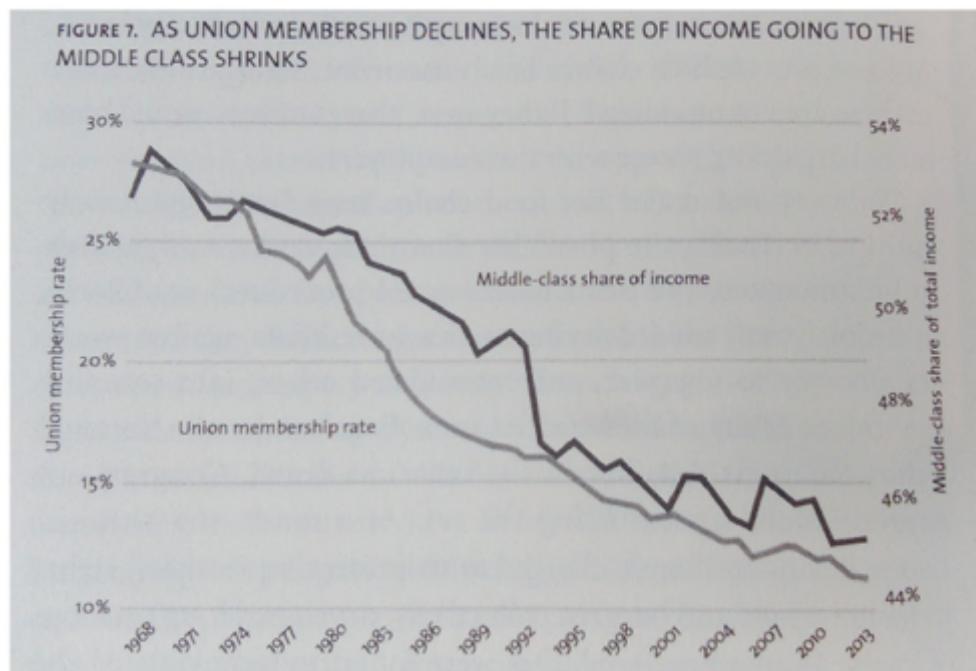
NASDAQ 2939.52 ▲ 6.32

Graphical integrity

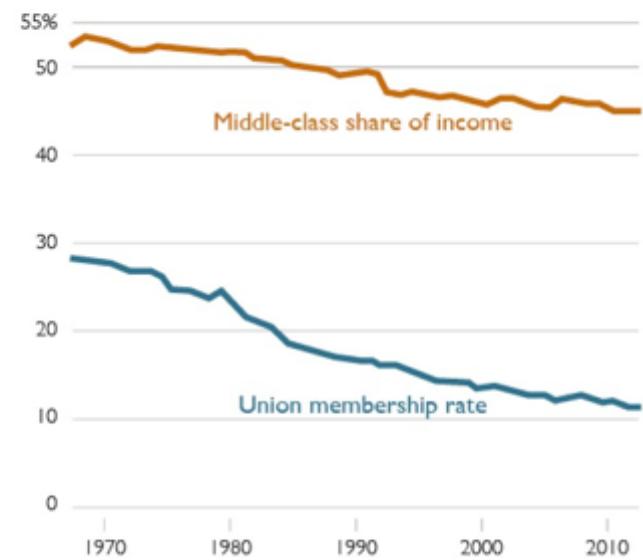
IF BUSH TAX CUTS EXPIRE



Scale distortions

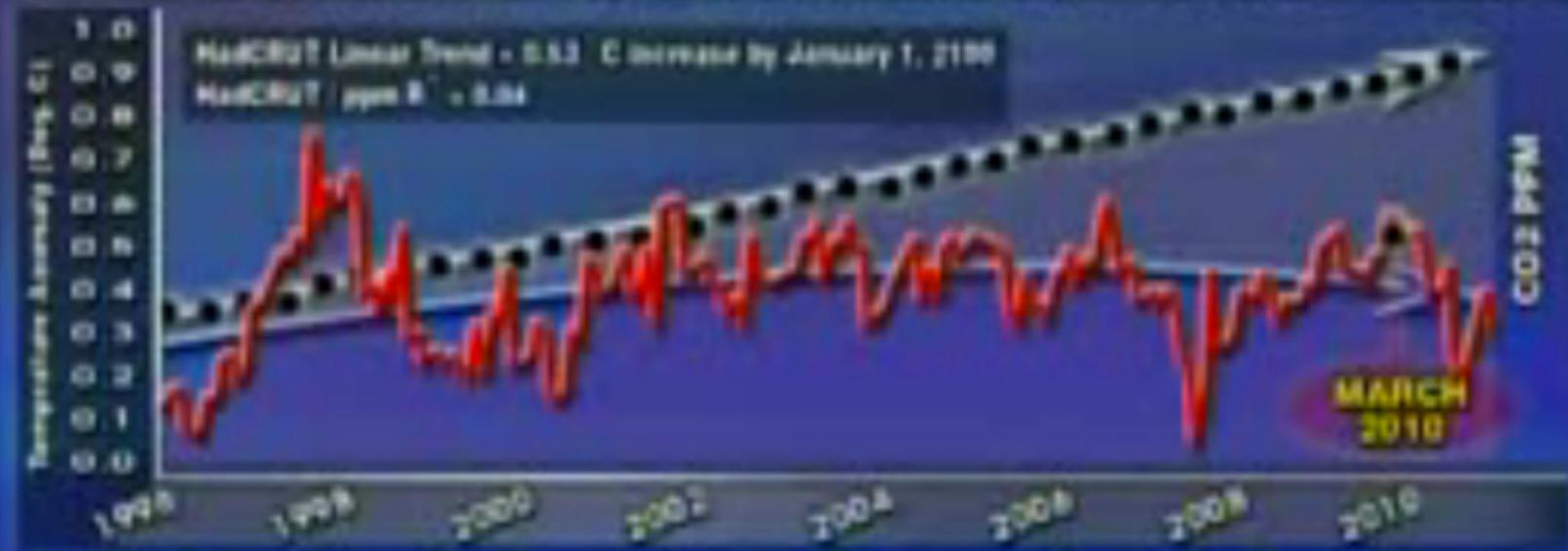


NEW VERSION



HADCRUT GLOBAL WARMING INSIGNIFICANT

15 YEARS ENDING MAY 2011 - MONTHLY ANOMALIES AND CO2 LEVELS

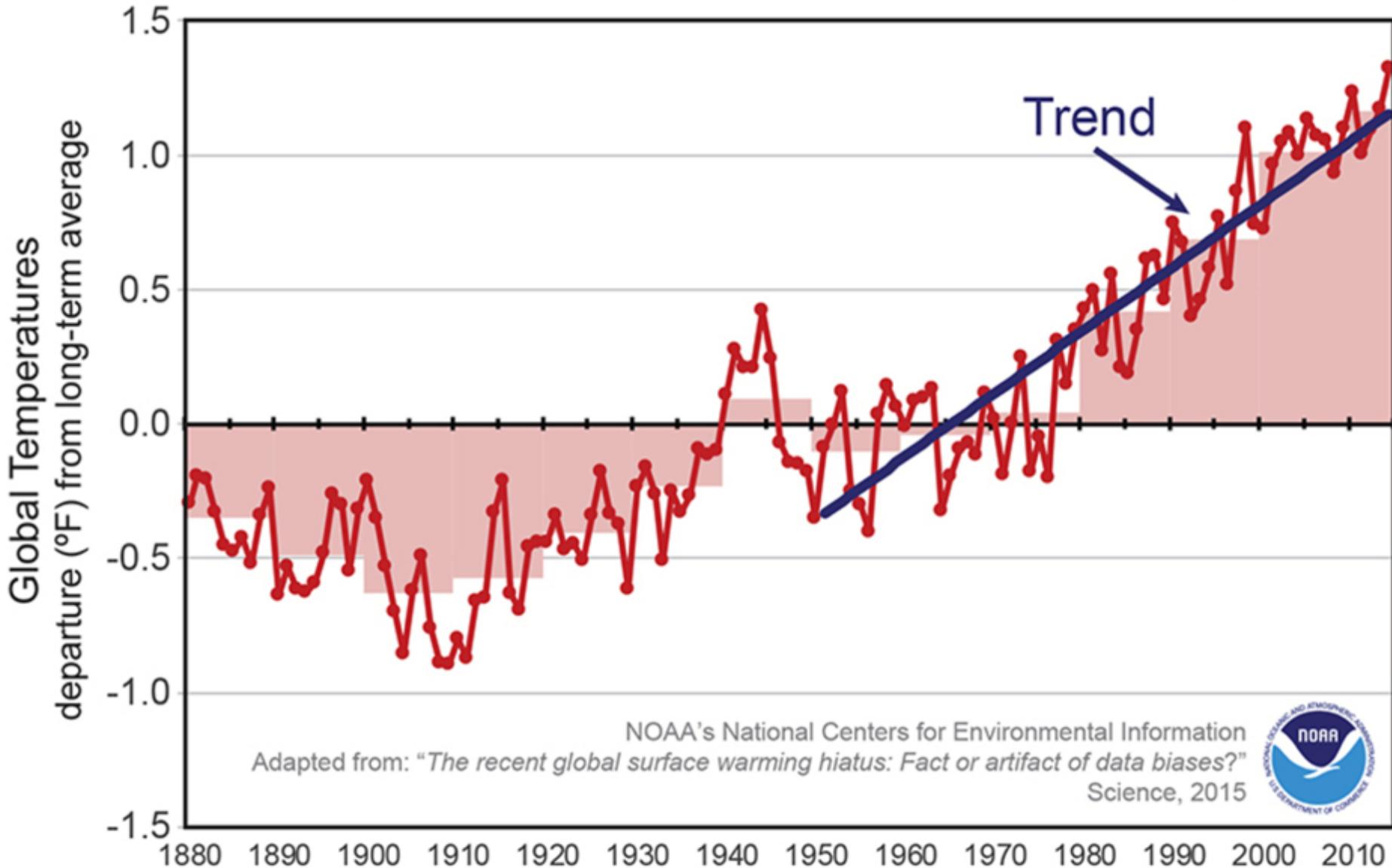


SOURCE: NOAA

FOX
NEWS
3:54 PT

D IN RECENT YEARS BY COLOMBIA'S U.S.-BACKED NYC 88.65

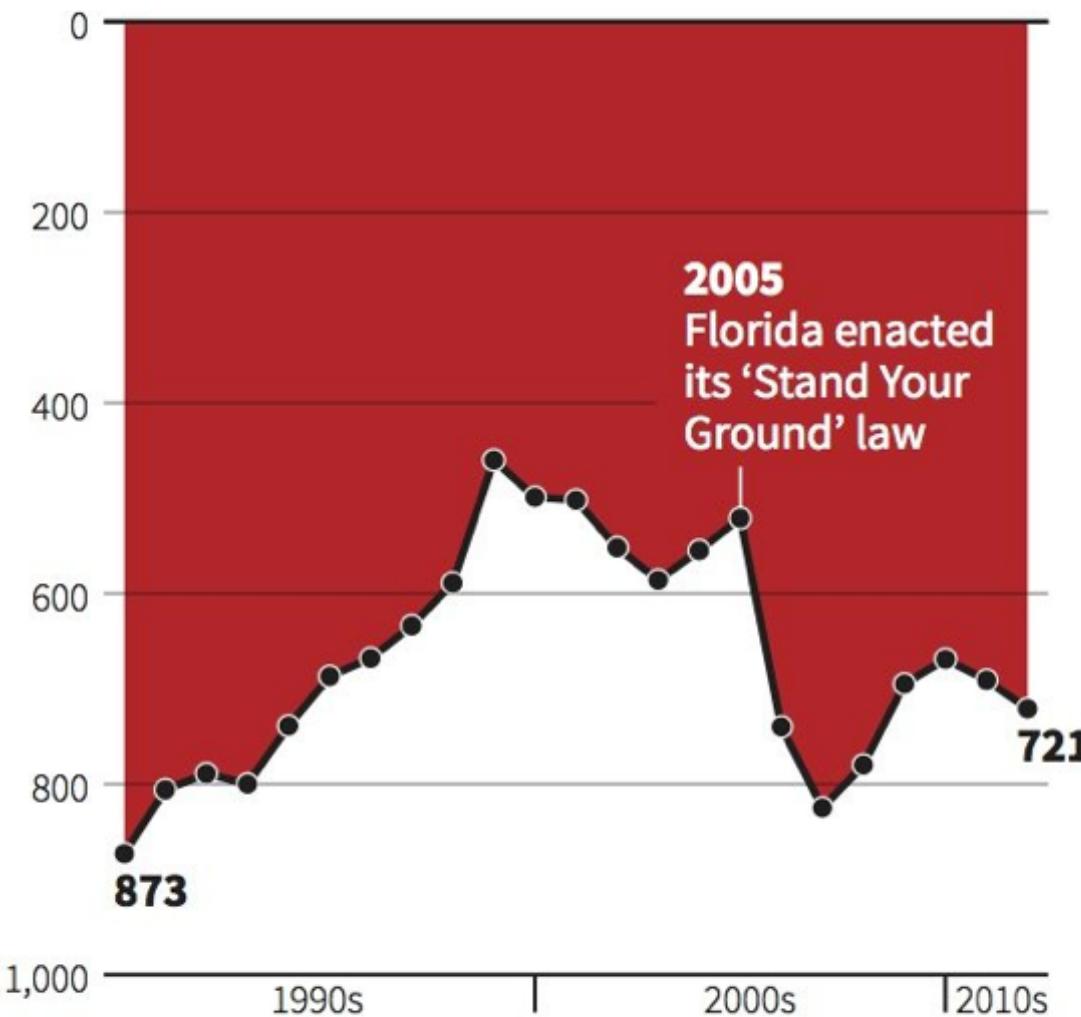
No Slow Down in Global Warming



Contrary to much recent discussion, the latest corrected analysis shows that the rate of global warming has continued, and there has been no slow down.

Gun deaths in Florida

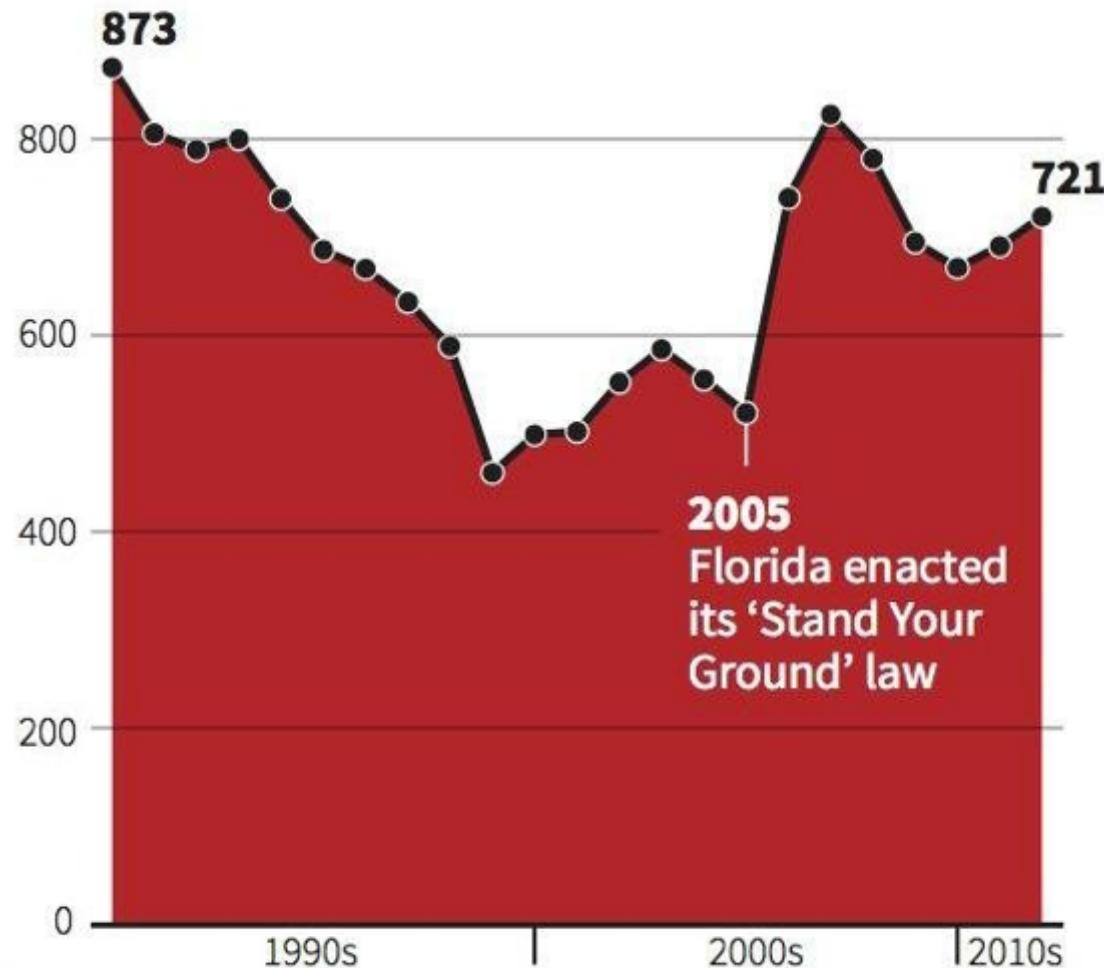
Number of murders committed using firearms



Source: Florida Department of Law Enforcement

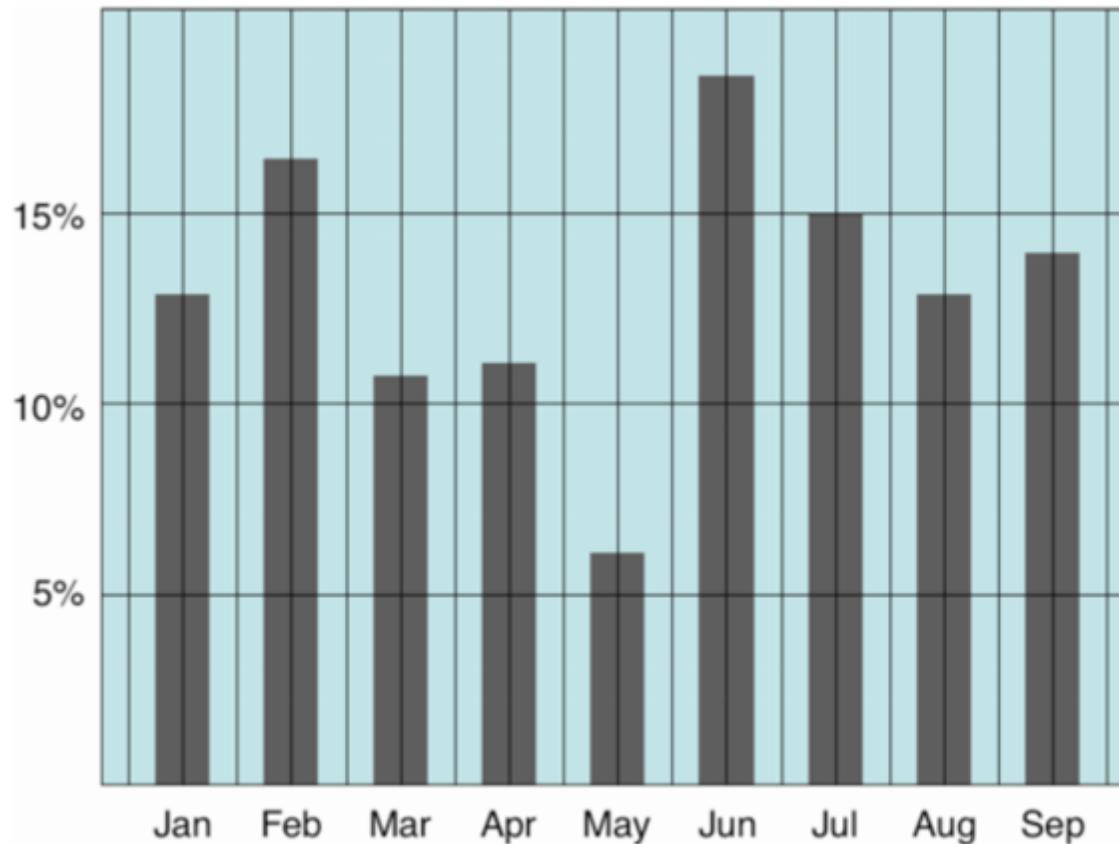
Gun deaths in Florida

Number of murders committed using firearms

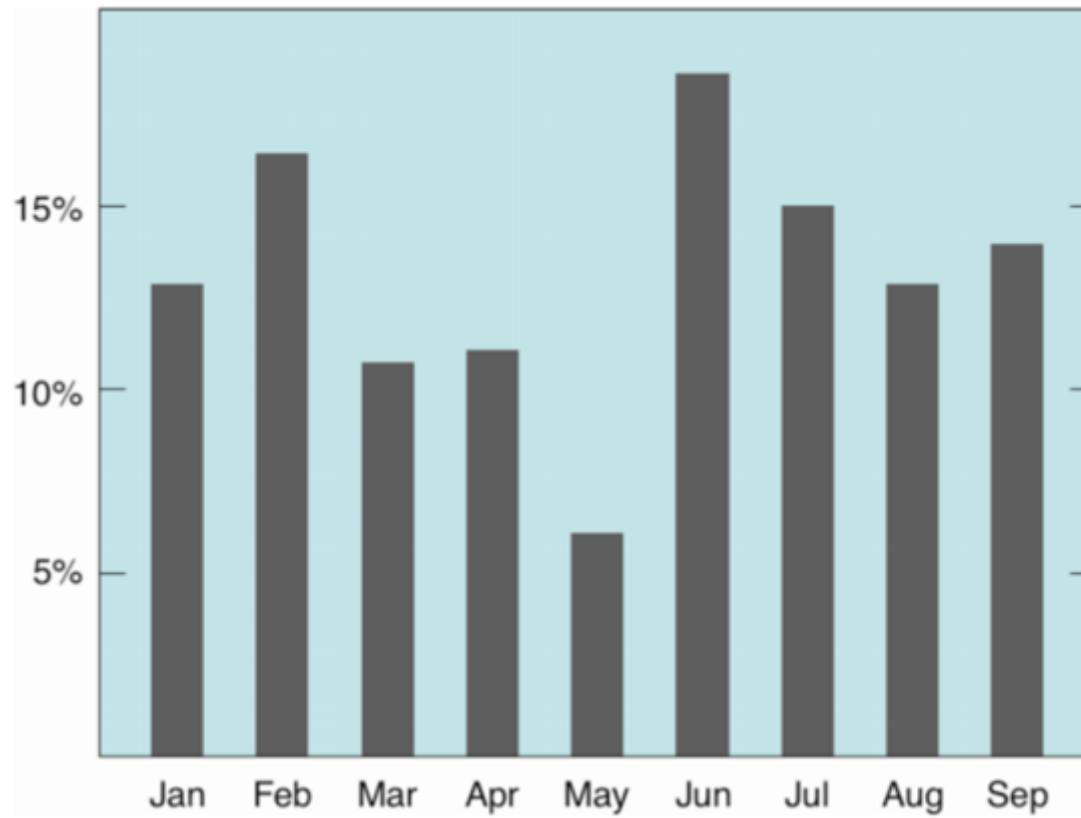


Source: Florida Department of Law Enforcement

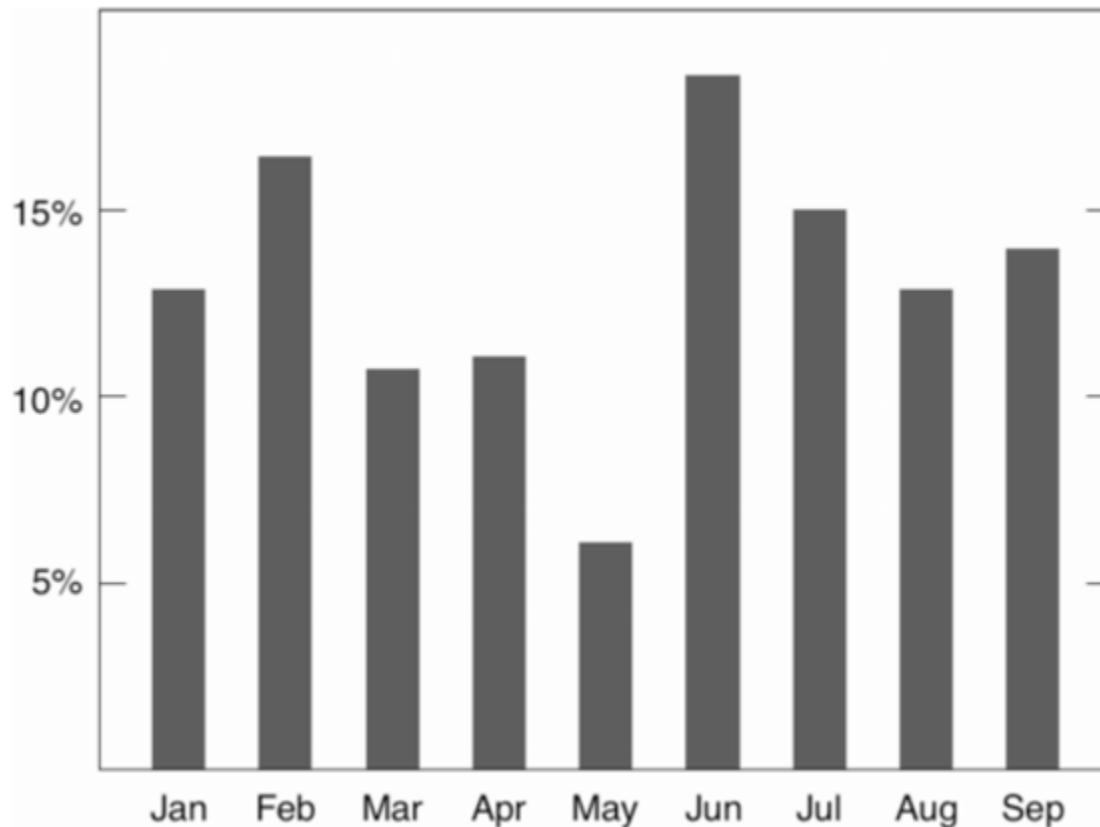
Avoid chartjunk



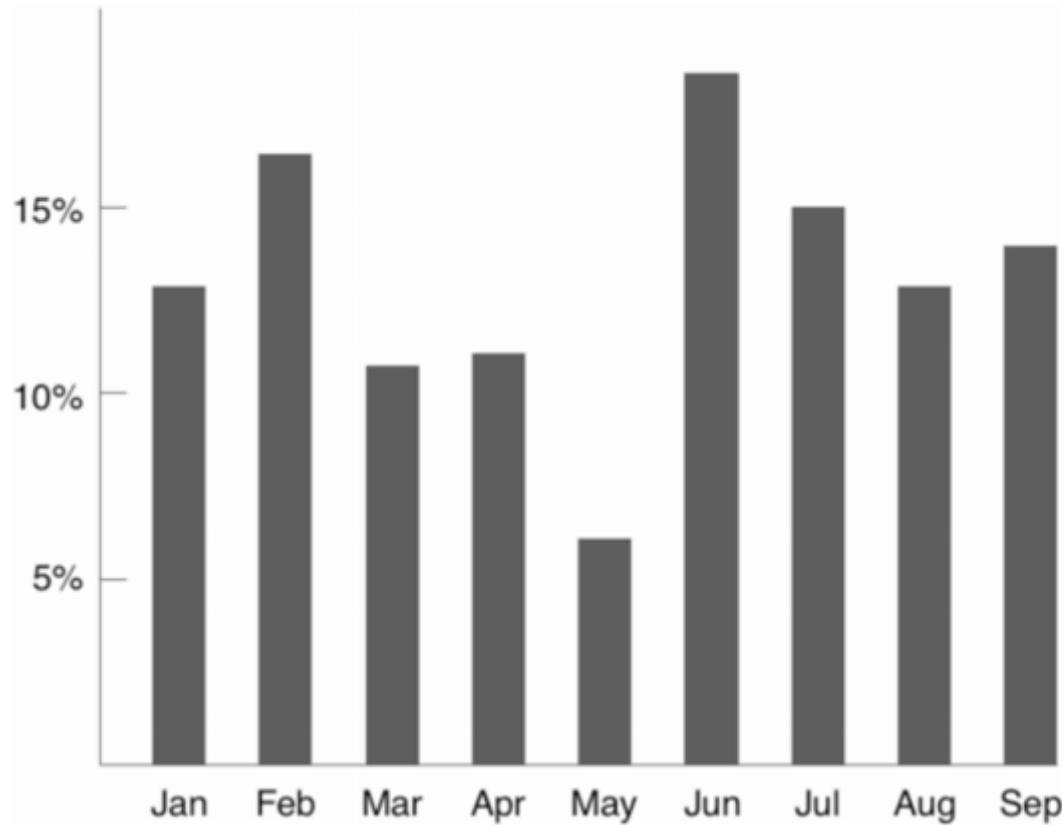
Avoid chartjunk



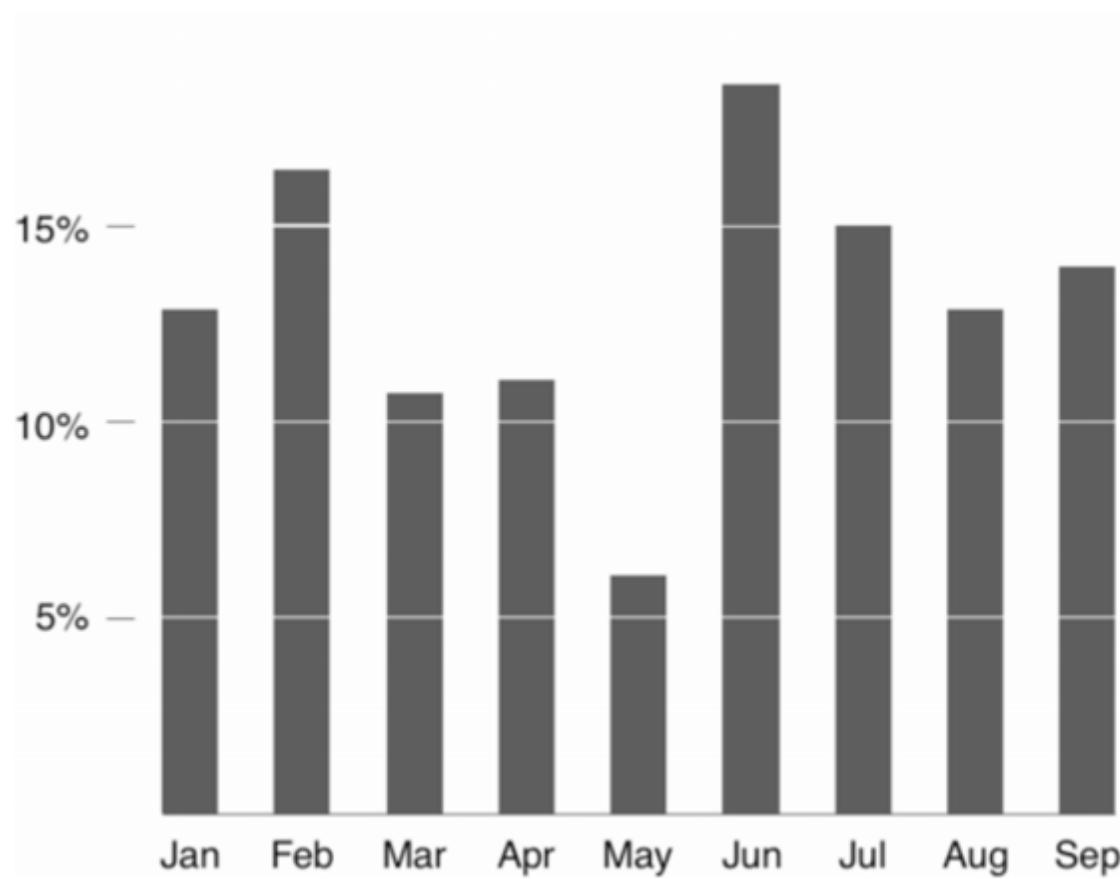
Avoid chartjunk



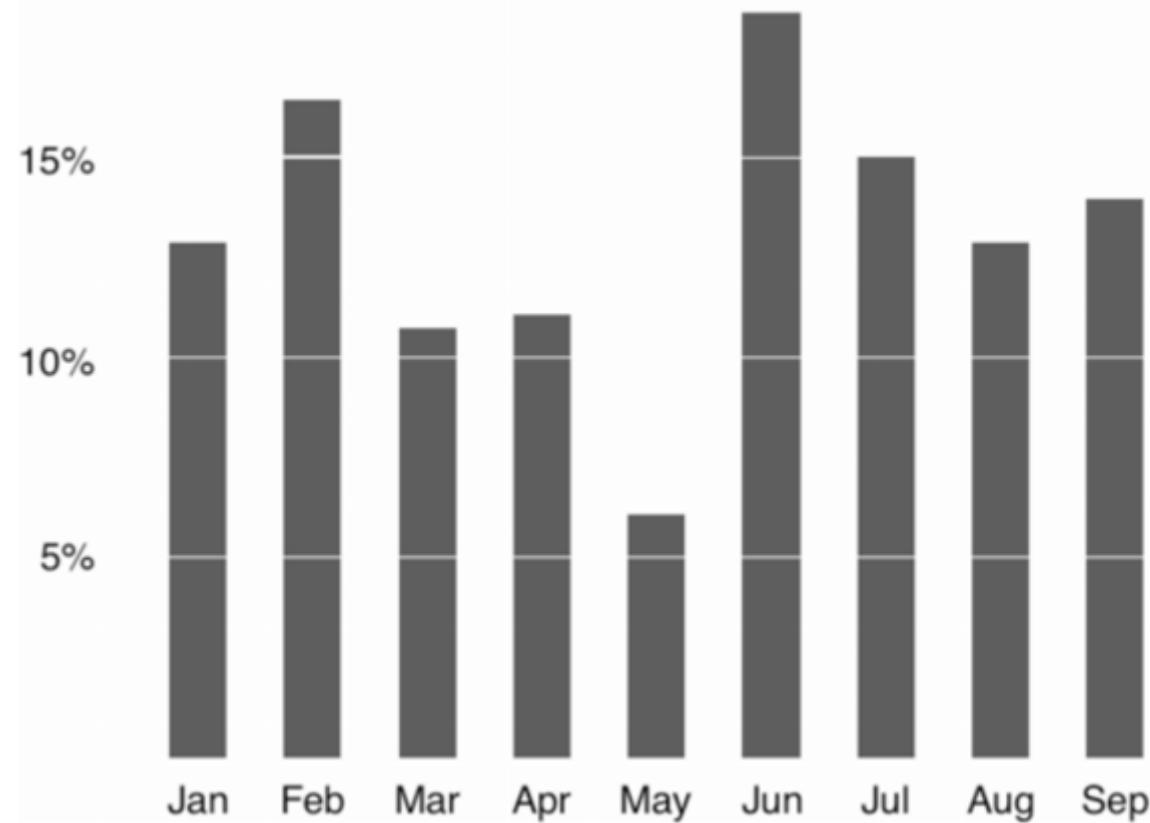
Avoid chartjunk



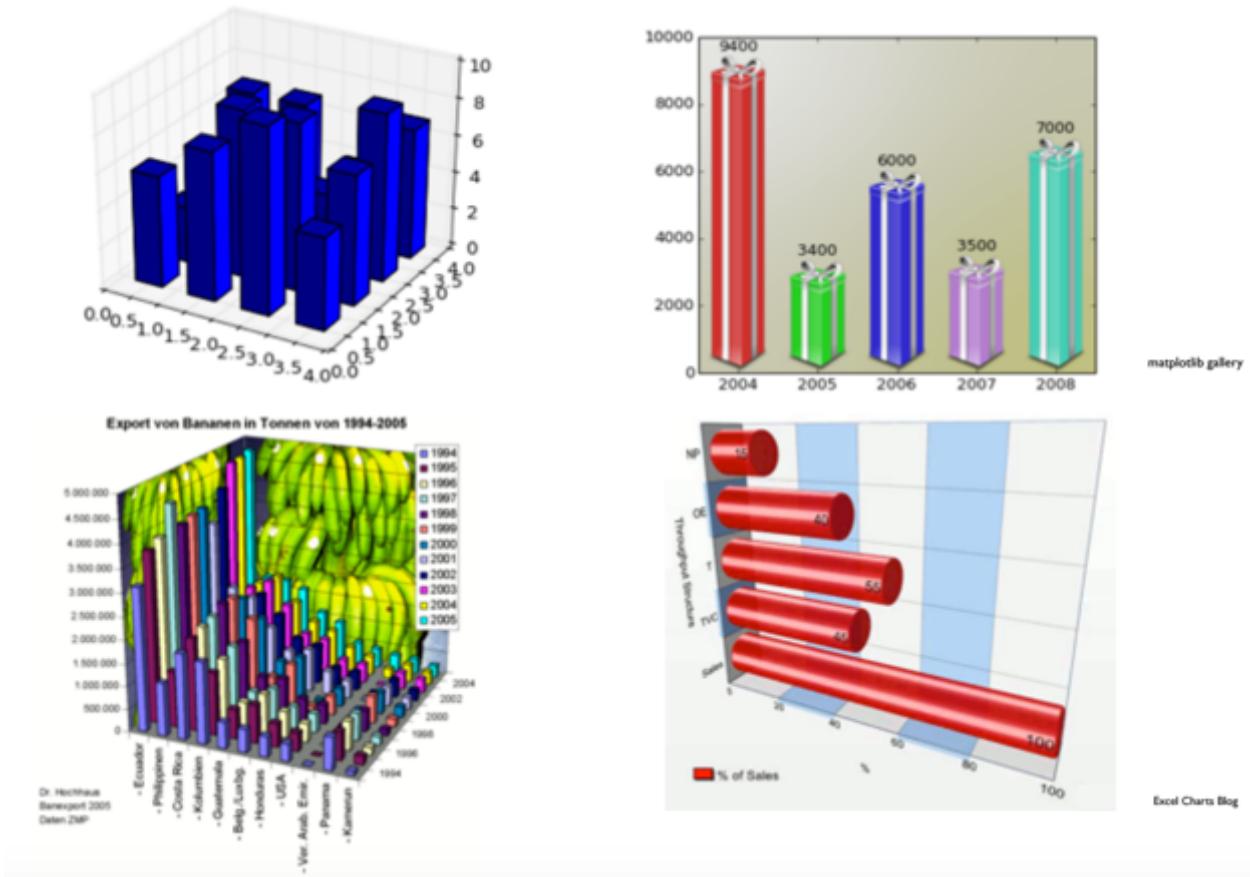
Avoid chartjunk



Avoid chartjunk



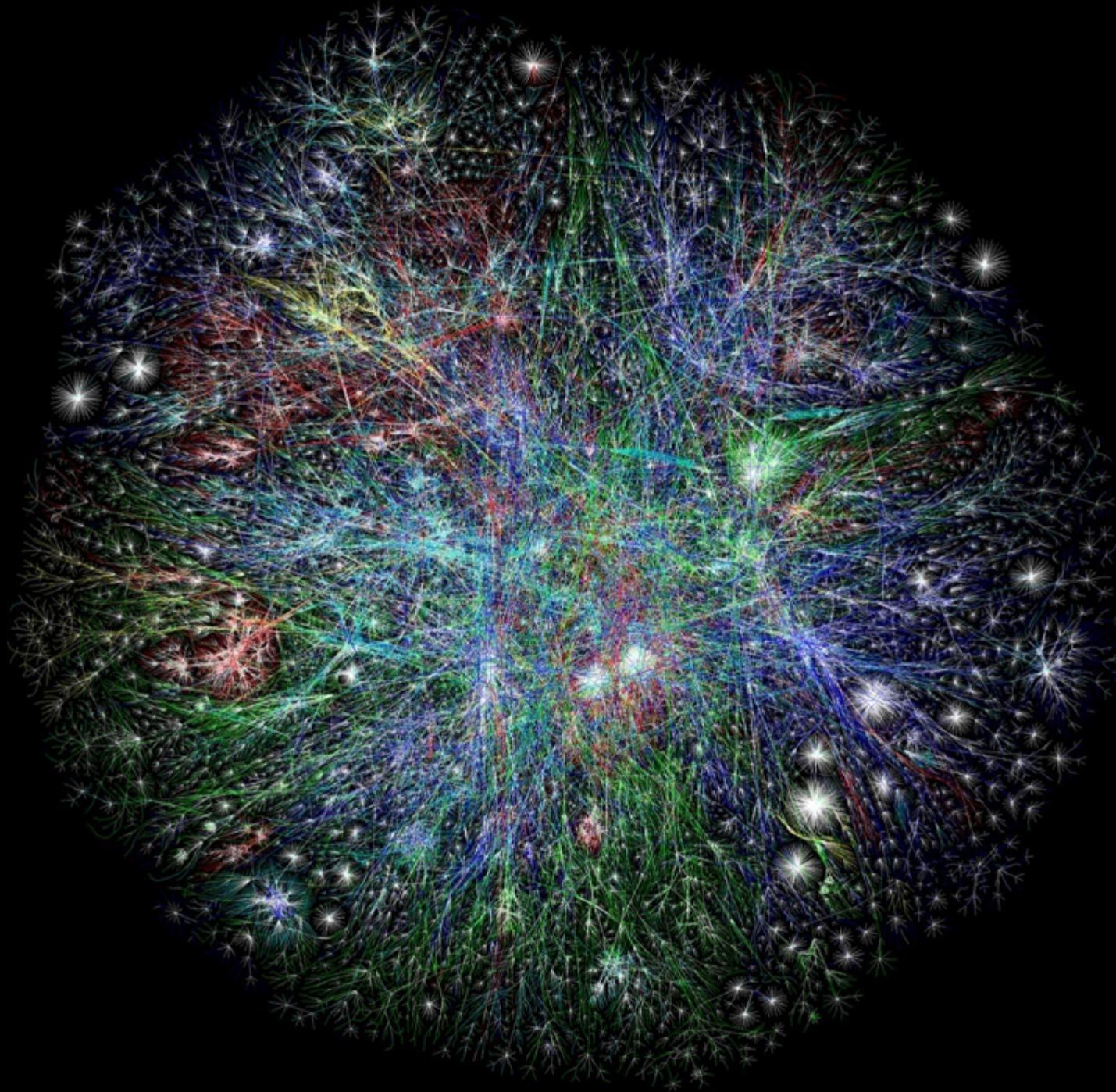
Examples of bad charts

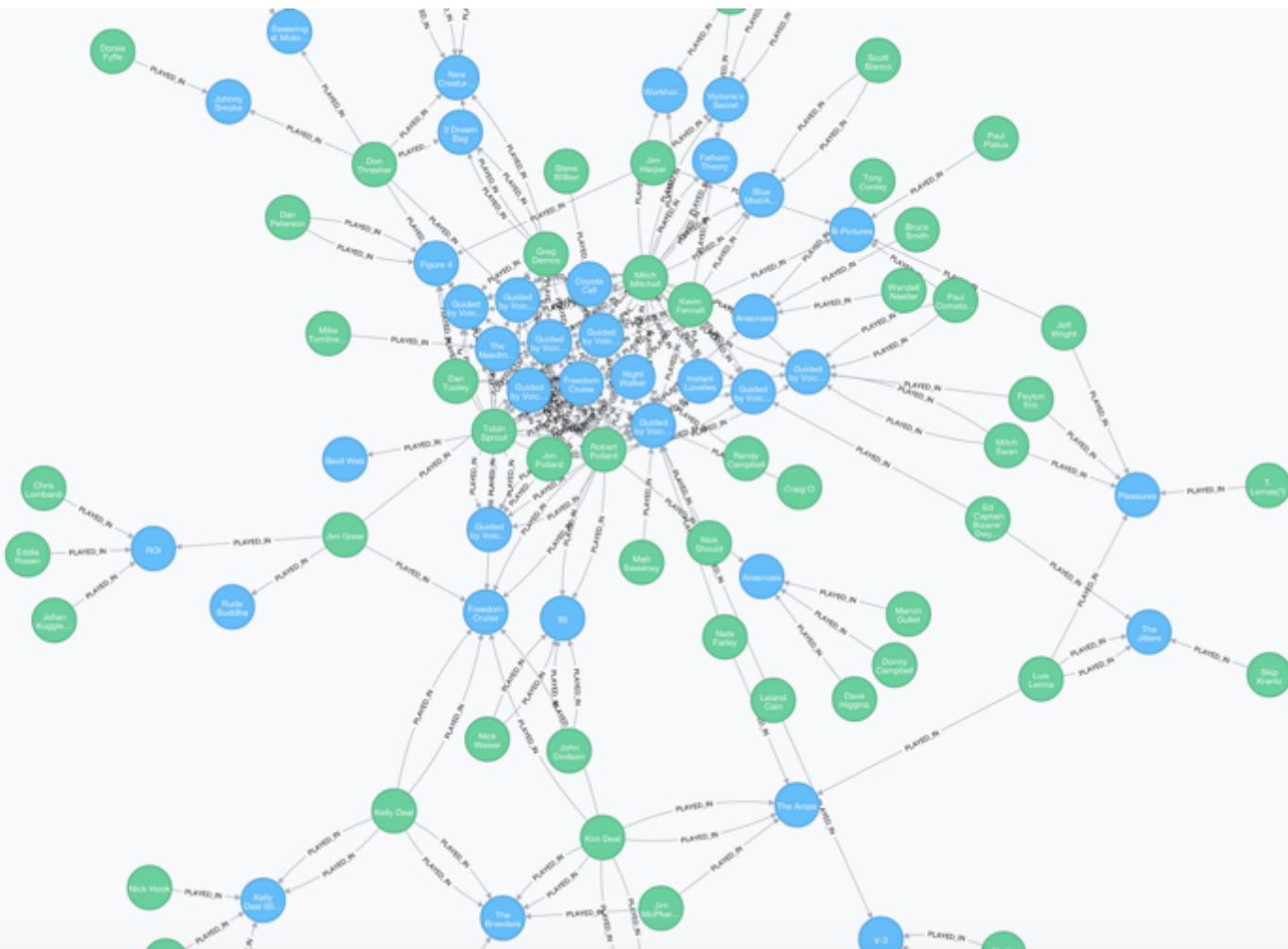


Visual Variables

- Basic model for stuff we measure (data):
 - From Science paper “On The Theory of Scales and Measurements”, basically:
 - Nominal scale – categorization
 - Ordinal scale – continuous, comparable
 - Interval scale – kind of like ordinal but no true zero
 - Ratios – can measure ratios and proportions
 - Also, entities, relationships and attributes
- => Jacques Bertin’s Visual Variables

| | <i>Points</i> | <i>Lines</i> | <i>Areas</i> | <i>Best to show</i> |
|------------------------|---------------|--|------------------|---|
| <i>Shape</i> | | <i>possible, but too weird to show</i> | <i>cartogram</i> | <i>qualitative differences</i> |
| <i>Size</i> | | | <i>cartogram</i> | <i>quantitative differences</i> |
| <i>Color Hue</i> | | | | <i>qualitative differences</i> |
| <i>Color Value</i> | | | | <i>quantitative differences</i> |
| <i>Color Intensity</i> | | | | <i>qualitative differences</i> |
| <i>Texture</i> | | | | <i>qualitative & quantitative differences</i> |





<http://neo4j.com/graphgist/b9833a175976a959d7d6>

Charts

- Many many different kind of charts...

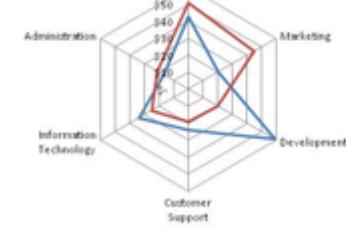
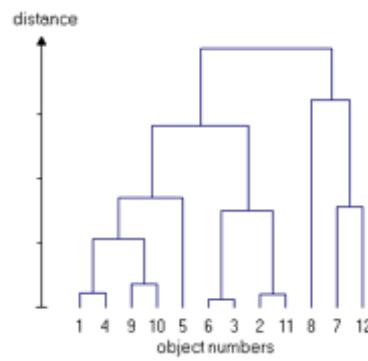
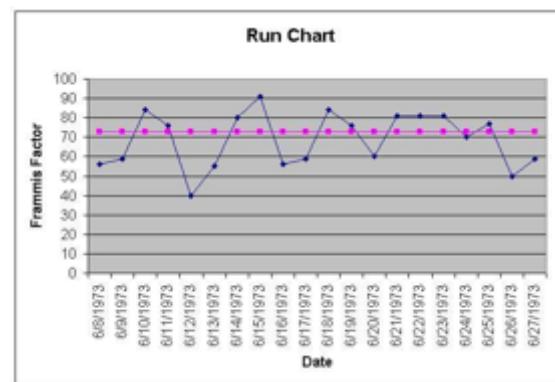
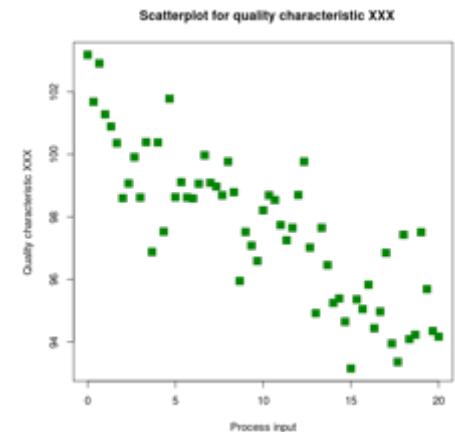
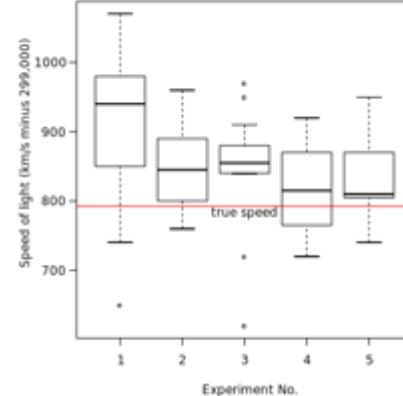
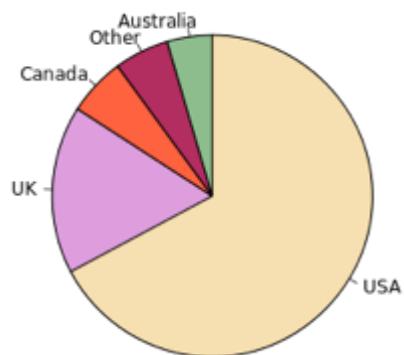
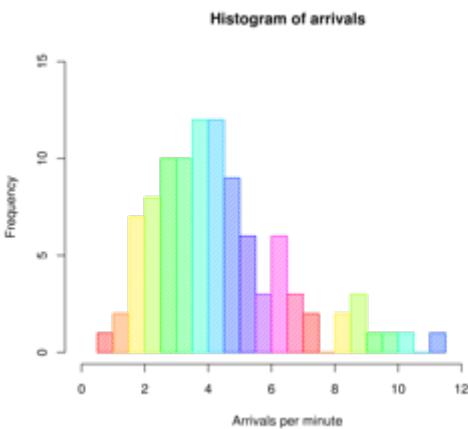
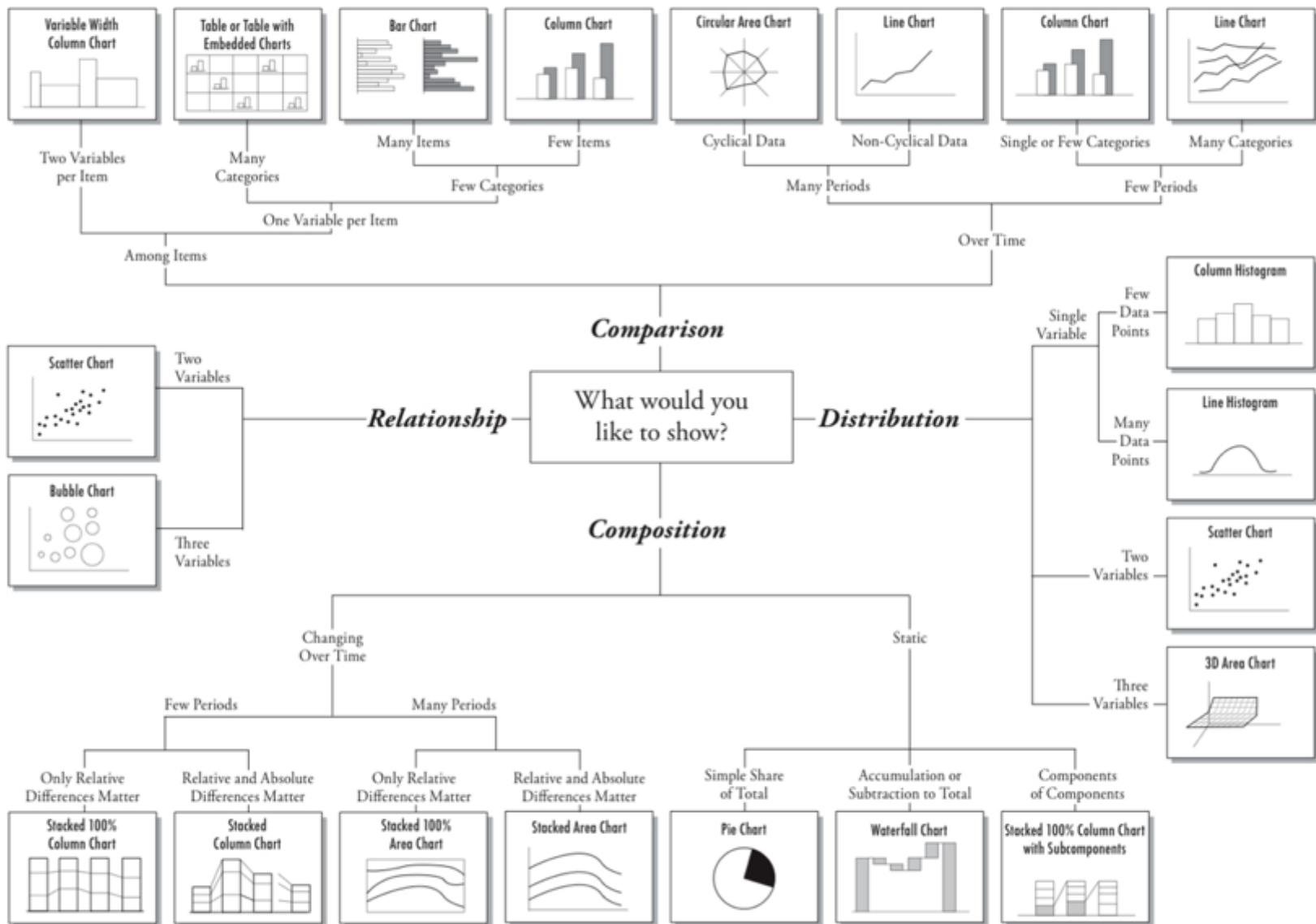
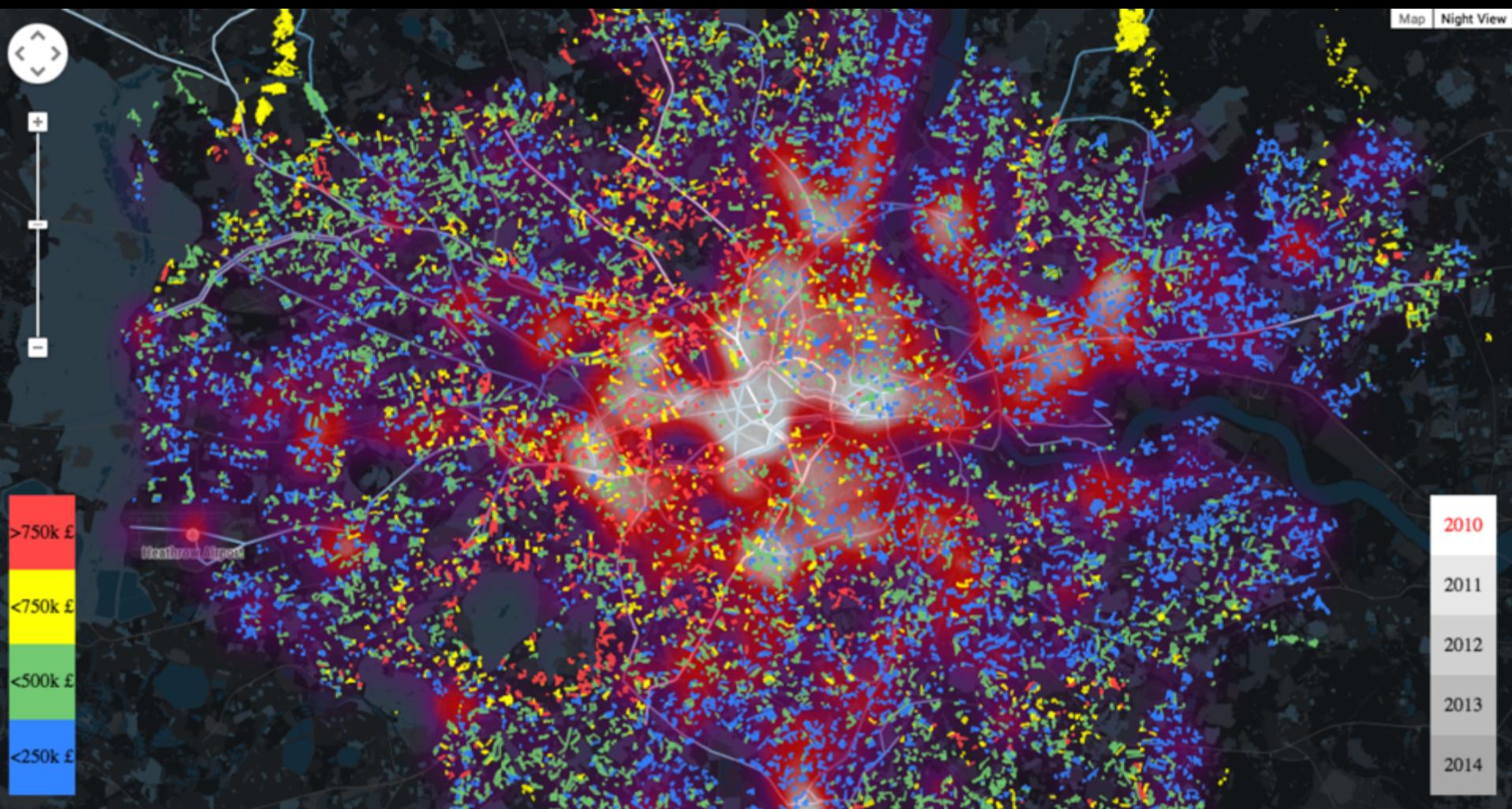


Chart Suggestions—A Thought-Starter



High dimensional data

- Heatmaps
- Connected charts
- Animations

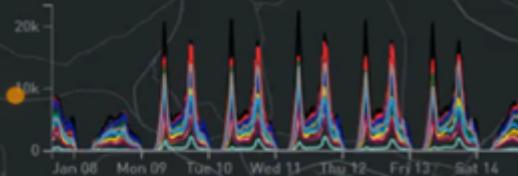
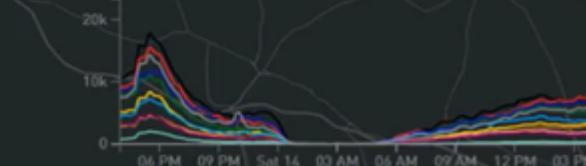


Energy & Environment of Future Cities: Human Mobility



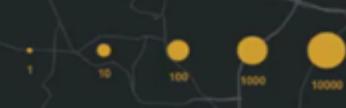
14/01/2012 15:45

Usage of London tube stations
during the Olympic Games 2012



- Hammersmith and City Line
- Circle Line
- Central Line
- Bakerloo Line
- Jubilee Line
- Victoria Line
- Piccadilly Line
- Northern Line

Passengers at station

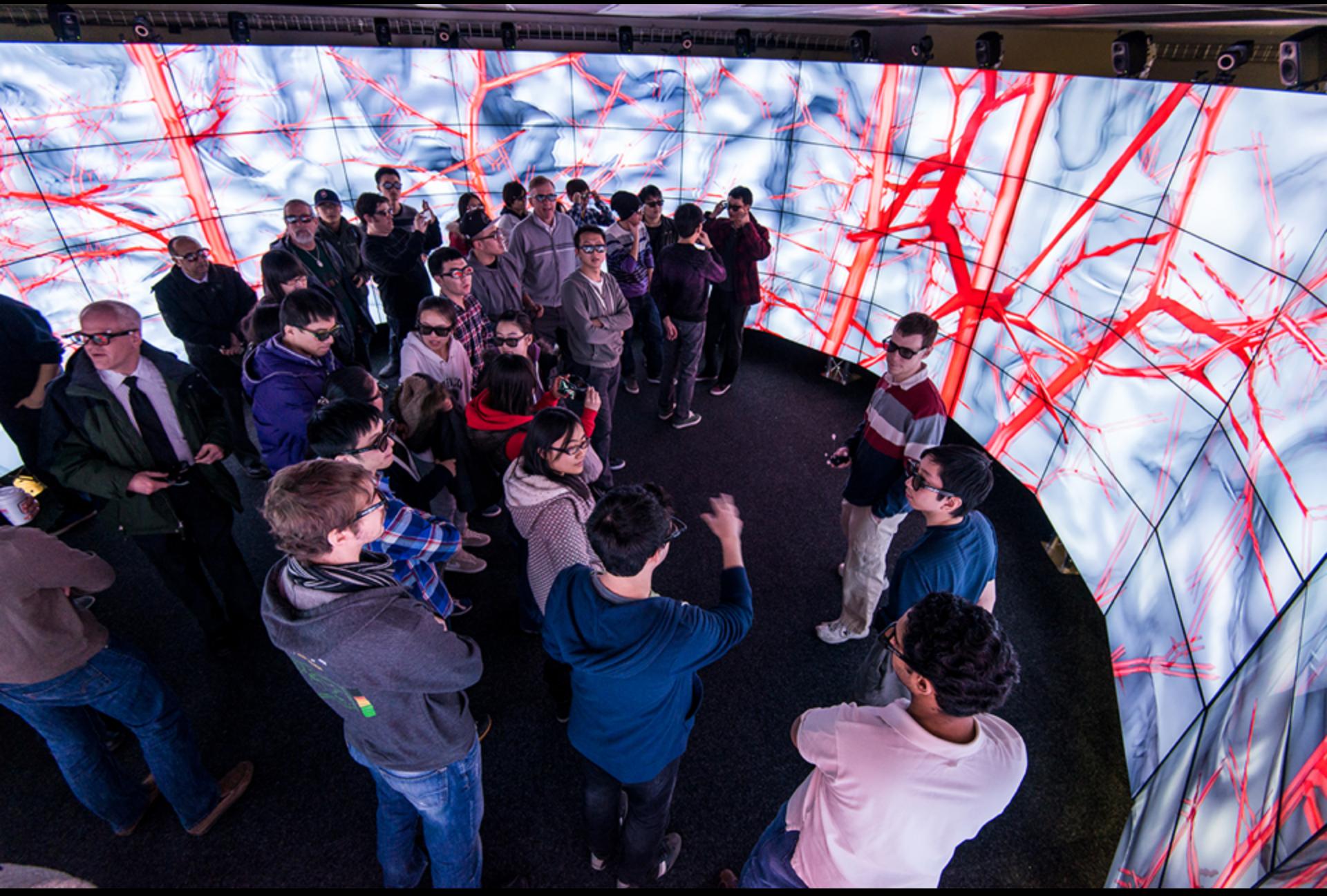


NYC version <https://www.youtube.com/watch?v=g6EaMQDHu7Q>

Visualization Systems

- Much visualization tooling is currently centered around how to present information on standard displays
 - Web
 - Desktop/laptop
 - Tablets/mobile
- What about on big devices/systems?

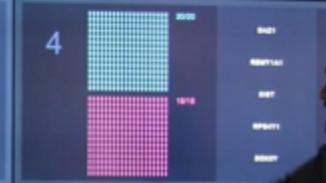




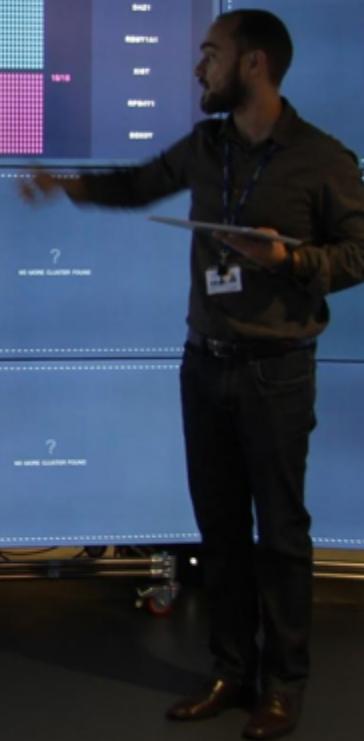




Asthma Precision Medicine using the U-BIOPRED cohort with eTRIKS/tranSMART



Genome Exploration





How does this fit in with the rest of CLO?

- Visualization could plug into different stages of your processing pipeline in data science/experimental exercises.
- Important for how to turn your big data analyses into summarized and presentable artefacts (typically for human consumption)
- Externalizing data representation into visual artefacts to use our visual cognition compliments our mind's cognitive abilities to extract meaning from data.

Visualization tools and frameworks

- Code/scripting frameworks
 - D3 (JS, very very widely used), bokeh, matplotlib (Python), R, and loads loads more....
- Tools
 - Tableau, Circos, MATLAB, Octave and more...

...lab time!