Visualization. Basics

Pere-Pau Vázquez
Dept. Computer Science – UPC

Outline

- Tasks and Data types
- The value of visualization
- The visualization mantra
- Takeaways
- Designing visualization systems

Outline

- **Tasks and Data types**
- The value of visualization
- The visualization mantra
- Takeaways
- Designing visualization systems

Tasks and Data Types

- Solving the visualization problem primarily requires answering these questions:
  - What is visualized?
  - Why is it visualized?
  - How is it visualized?

  - The answers to the first two questions determine the answer to the third question

Tasks and Data Types

- What is visualized? $\rightarrow$ Specification of the data
  - The data that has been measured, observed, or computed in the data domain, i.e., the dependent variables
    - includes not only simple data values, but also objects and their relationships, where objects and relations may have arbitrary data attributes associated with them

Tasks and Data Types

- Data Types
  - 1-dimensional
  - 2-dimensional
  - 3-dimensional
  - Temporal
  - Multi-dimensional
  - Tree
  - Network

Tasks and Data Types

- 1-dimensional data:
  - Textual documents, program source code, alphabetical lists of names, floating point values…
    - Organized in a sequential manner!
  - Each item in the collection is a line of text containing a string of characters
  - **Interface design issues:** what fonts, color, size to use and what overview, scrolling, or selection methods can be used
  - **User problems:** find the number of items, see items having certain attributes, or see an item with all its attributes

Tasks and Data Types

- 2-dimensional data:
  - Planar or map data include geographic maps, floorplans, or newspaper layouts, 2D vectors…
  - Each item in the collection covers some part of the total area and may be rectangular or not
  - **User problems:** find adjacent items, containment of one item by another, paths between items
  - **Basic tasks:** counting, filtering, and details-on-demand

Tasks and Data Types

- 3-dimensional data:
  - Real-world objects such as molecules, the human body, and buildings
    - Items with volume and some potentially complex relationship with other items
  - **Users' tasks:** adjacency plus above/below and inside/outside relationships, as well as the basic tasks. Understanding position and orientation when viewing the objects, plus the serious problems of occlusion.
  - Proposed techniques: overviews, landmarks, perspective, stereo display, transparency, and color coding

Tasks and Data Types

- Temporal data:
  - Medical records, project management, historical presentations
  - **Features:** Items have a start and finish time, items may overlap
  - **Tasks:** finding all elements before, after, or during some period or moment, plus basic tasks.

Tasks and Data Types

- Multi-dimensional data:
  - Relational and statistical databases can be manipulated as multi-dimensional data
    - Items with n attributes become points in a n-dimensional space
  - The interface representation can be 2-dimensional scattergrams
    - With additional dimension controlled by a slider or buttons for attribute values when the cardinality is small (e.g. < 10)
    - Disorientation and occlusions appear as main problems
  - **Tasks:** finding patterns, clusters, correlations among pairs of variables, gaps, and outliers
  - Parallel coordinates makes some tasks easier
    - But takes practice for users to comprehend

Tasks and Data Types

- Tree:
  - Hierarchies or tree structures are collections of items with each item having a link to one parent item (except the root).
  - Items and the links between parent and child can have multiple attributes.
  - **Tasks:** Basic tasks (applied to items and links), and tasks related to structural properties, e.g. how many levels in the tree? or how many children does an item have?
  - **Interface representations:** Outline style of indented labels, a node and link diagram, or a treemap (with child items are rectangles nested inside parent rectangles)

Tasks and Data Types

- Network:
  - Collection of items linked to an arbitrary number of other items
  - **Tasks:** Basic tasks applied to items and links, and more specific to networks, e.g. shortest or least costly paths connecting two items or traversing the entire network
  - **Interface representations:** include a node and link diagram, square matrix of the items with the value of a link attribute in the row and column representing a link

Tasks and Data Types

- Data Types
  - *Nominal* (unordered set of names)
    - Examples: Car manufacturers and countries
    - Only test for inequality are possible
  - *Ordinal* (ordered set of non-measureable data)
    - Examples: days of the week and rating scales
    - Tests for equality and "direction (<,>,=, ≠)

Tasks and Data Types

- Data Types
  - *Quantitative* (measured or simulated data)
    - Examples: physical measurements of height, weight, length
    - Full set of arithmetic operations possible
- Data Types. Transformations:
  - Nominal to ordinal data by introducing order, e.g., alphabetical
  - Quantitative to ordinal data by binning (*classification*)

Tasks and Data Types

- Why is it visualized? → Specification of the task
  - What is the necessity?
  - What tasks the user seeks to accomplish with the help of the visualization?

Tasks and Data Types

- Why is it visualized?
  - Three types of abstract tasks
    - explorative analysis
    - confirmative analysis
    - presentation of analysis results

Tasks and Data Types

- *Explorative analysis:* Undirected search. No a priori hypotheses about the data are given.
  - The goal is to get insight into the data, to begin extracting relevant information, and to come up with hypotheses
- *Confirmative analysis:* Visualization is used to prove or disprove hypotheses
  - May come from data exploration or from models associated with the data
- *Presentation:* Step to communicate and disseminate analysis results

Tasks and Data Types

- Visualization tasks (e.g. from Shneiderman, 1996)
  - Overview
  - Zoom
  - Filter
  - Details-on-demand
  - Relate
  - History
  - Extract

Tasks and Data Types

- Overview:
  - Gain an overview of the entire collection.
  - **Strategies:** zoomed out views of each data type to see the entire collection plus an adjoining detail view

Tasks and Data Types

- Zoom:
  - Zoom in on items of interest
  - Users typically have an interest in some portion of a collection, and they need tools to enable them to control the zoom focus and the zoom factor
  - Smooth zooming helps users preserve their sense of position and context
  - Zooming could be on one dimension at a time by moving the zoom bar controls or by adjusting the size of the field-of -view box

Tasks and Data Types

- Filter:
  - Filter out uninteresting items. Dynamic queries applied to the items in the collection
    - One of the key ideas in information visualization
  - By allowing users to control the contents of the display, users can quickly focus on their interests by eliminating unwanted items
  - Sliders, buttons, or other control widgets coupled to rapid display update (less than 100 milliseconds) is the goal
    - Even when there are tens of thousands of displayed items

Tasks and Data Types

- Details-on-demand:
  - Select an item or group and get details when needed
  - Once a collection has been trimmed to a few dozen items it should be easy to browse the details about the group or individual items
  - The usual approach is to simply click on an item to get a pop-up window with values of each of the attributes

Tasks and Data Types

- Relate:
  - View relationships among items

Tasks and Data Types

- History:
  - Keep a history of actions to support undo, replay, and progressive refinement
  - It is rare that a single user action produces the desired outcome
  - Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important

Tasks and Data Types

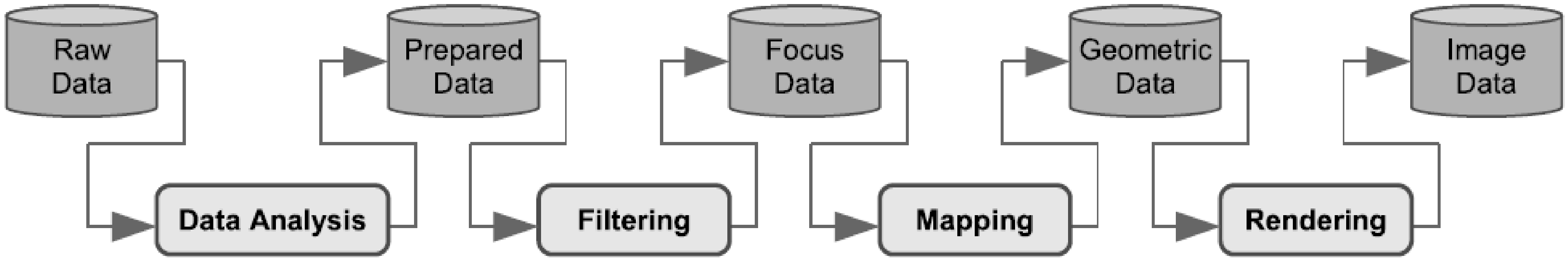
- Extract:
  - Allow extraction of sub-collections and of the query parameters
  - Once users have obtained the item or set of items they desire, it would be useful to be able to extract that set and save it to a file in a format that would facilitate other uses such as sending by email, printing, graphing, or insertion into a statistical or presentation package

Tasks and Data Types

- How is it visualized? $\longrightarrow$ The visualization pipeline
  - Raw data need to be transformed into image data in a data-dependent and task-specific manner
  - Map data to geometry and corresponding visual attributes like color, position, size, or shape, also called **visual variables**

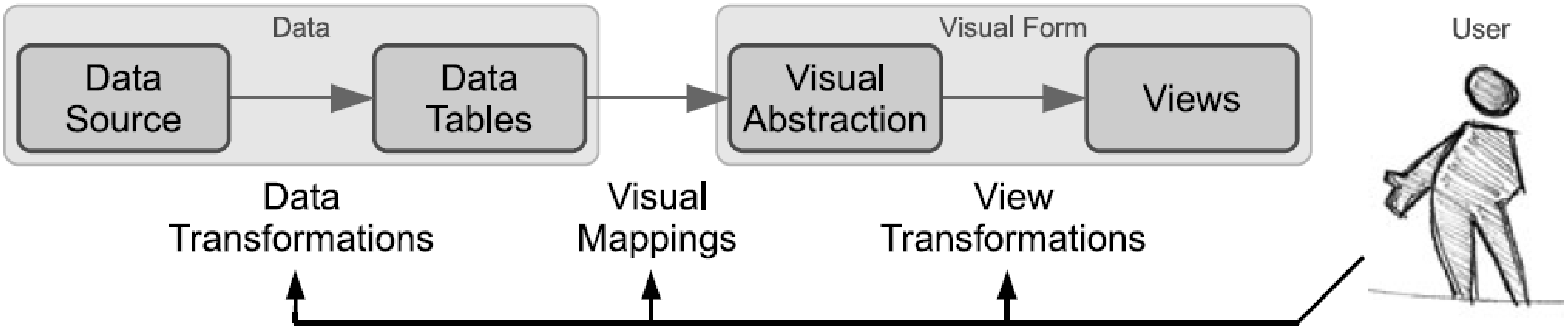
Tasks and Data Types

- The visualization pipeline (e.g. by dos Santos and Brodlie, 2004)



Tasks and Data Types

- The visualization pipeline is not static (from Aigner et al., 2011)



Tasks and Data Types

- Three major criteria have to be satisfied:
  - Expressiveness
  - Effectiveness
  - Appropriateness

Tasks and Data Types

- *Expressiveness:* Show exactly **the information in the data**
  - Nothing more and nothing less
- *Effectiveness:* Take into account the **cognitive capabilities** of the human visual system, and
  - the task, application background, and other context-related information…
- *Appropriateness:* Cost-value ratio that assesses the benefit of the visualization process with respect to achieving the task
  - Mainly **time** (computation) and **space** (screen-space) efficiency

Outline

- *Tasks and Data types*
- **The value of visualization**
- The visualization mantra
- Takeaways
- Designing visualization systems

The value of Visualization

- Is it valuable at all?
  - We can actually **measure** whether a visualization is worth it
  - Satsko’s formula would be:

$$\text{Value} = T + I + E + C$$

The value of Visualization

Value = **T** + I + E + C

T: Time

- Ability to minimize the total **time** needed to answer a wide variety of questions about the data
  - Without formal queries, interaction really helps

The value of Visualization

- Minimizing time, e.g. for “low level” tasks:
  - Retrieve value
  - Filter
  - Compute derived value
  - Find extremum
  - Sort
  - Determine range
  - Characterize distribution
  - Find anomalies
  - Cluster
  - Correlate

The value of Visualization

Value = T + **I** + E + C

I: Insights

- Ability to spur and discover **insights** or insightful questions about the data
  - Would be really difficult with only the data

The value of Visualization

- Getting insights
  - Insight: An individual observation about the data by the participant, a unit of discovery
    - Knowledge-building and model-confirmation
  - Some common features:
    - Complex
    - Deep
    - Qualitative
    - Relevant
    - Specially interesting if it is unexpected

The value of Visualization

Value = T + I + **E** + C

E: Essence

- Ability to convey an overall **essence** or take-away sense of the data
  - The big picture: Whole is greater than the sum of the parts

The value of Visualization

$$\text{Value} = T + I + E + C$$

C: Confidence

- Ability to generate **confidence** and trust about the data, its domain and context
  - Beneficial data analysis process side effects

The value of Visualization

- Visualization can be worth it if…
  - It saves us **time**
  - It gives us new **insights on the data**
  - It is able to communicate the **essence** of the data
  - It helps us to generate **confidence** on the data

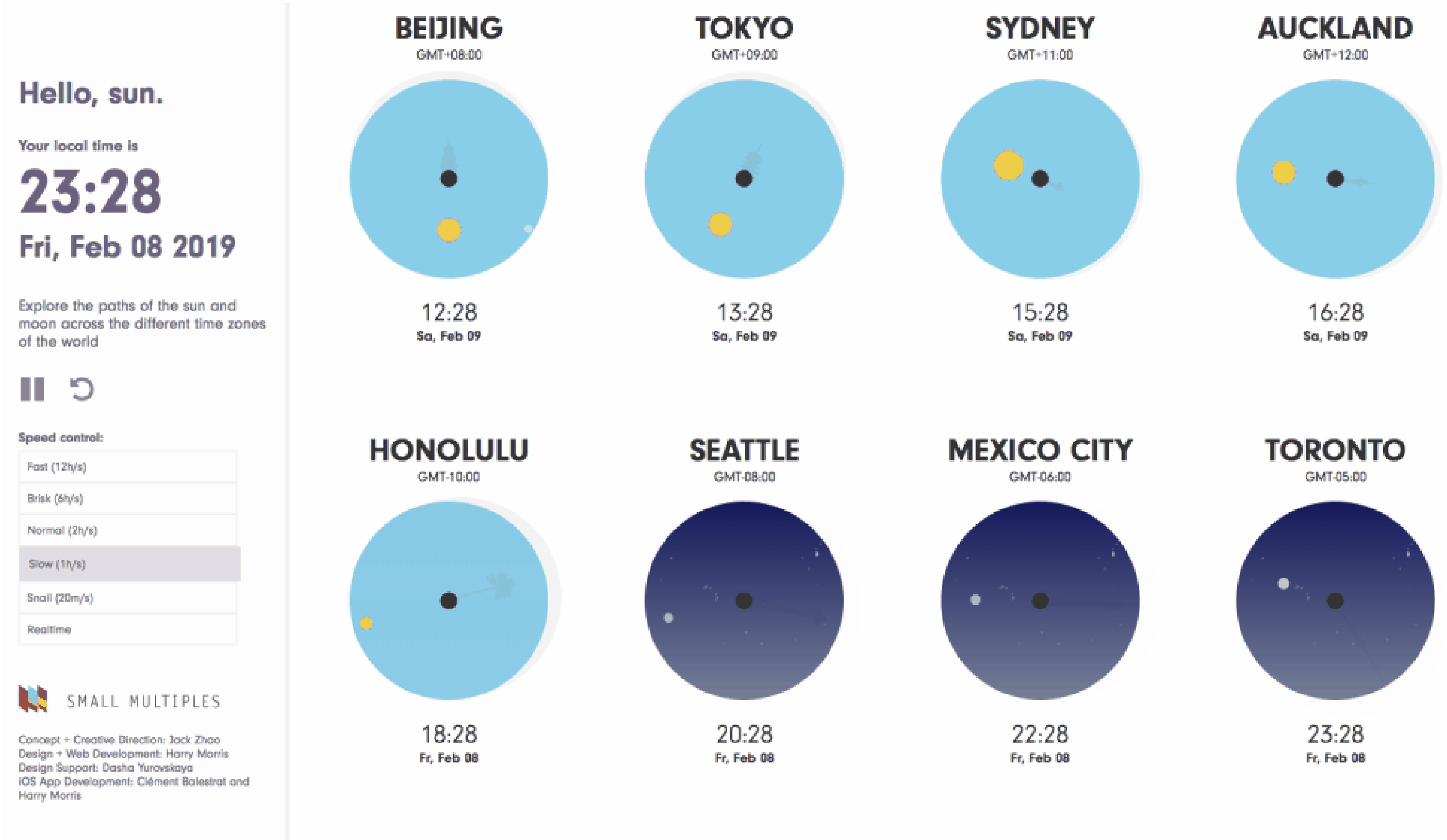
… and the cost of creating and using the visualization is lower than achieving those by other means…

The value of Visualization

- Examples. **Saving time** (e.g. best/worst performing shares)

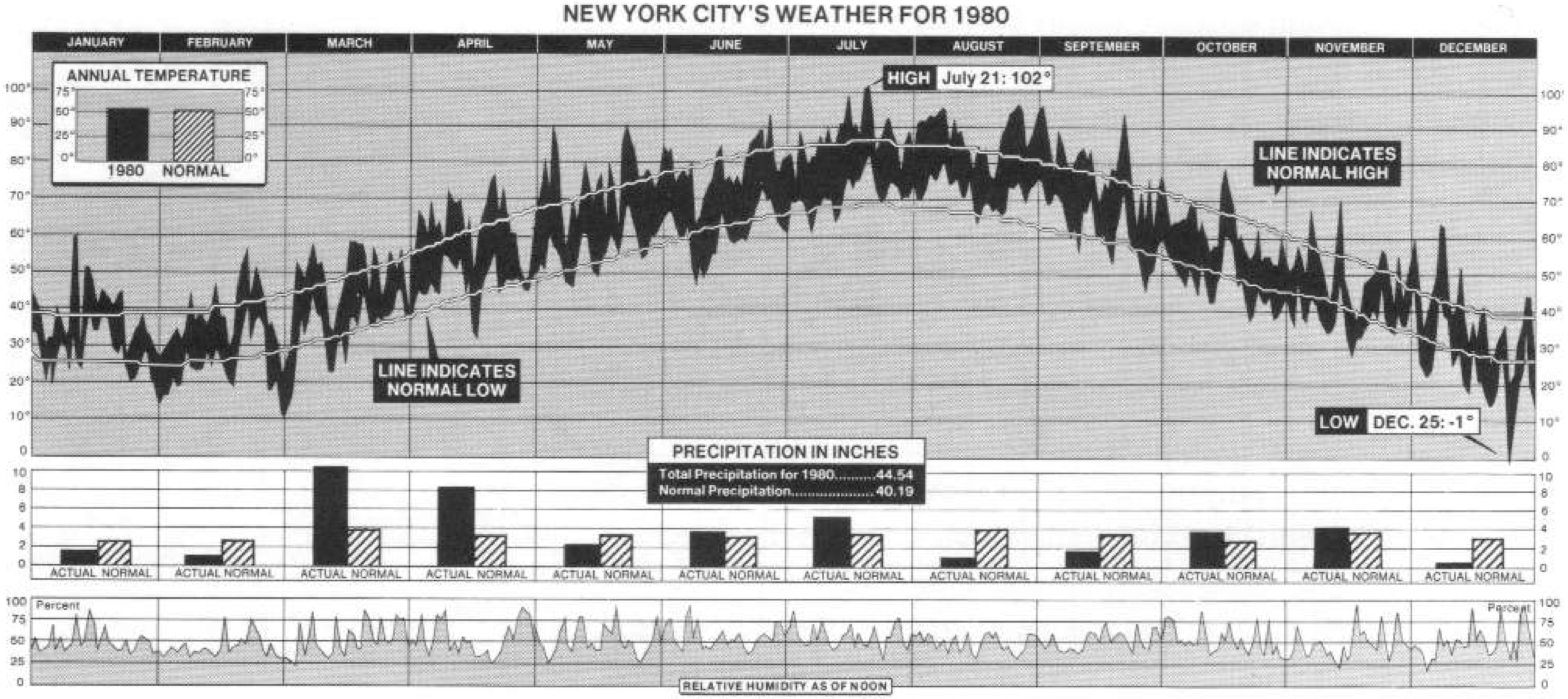
The value of Visualization

- Examples. **Getting insights** on the paths of sun and moon



The value of Visualization

- Examples. The **essence** of the data



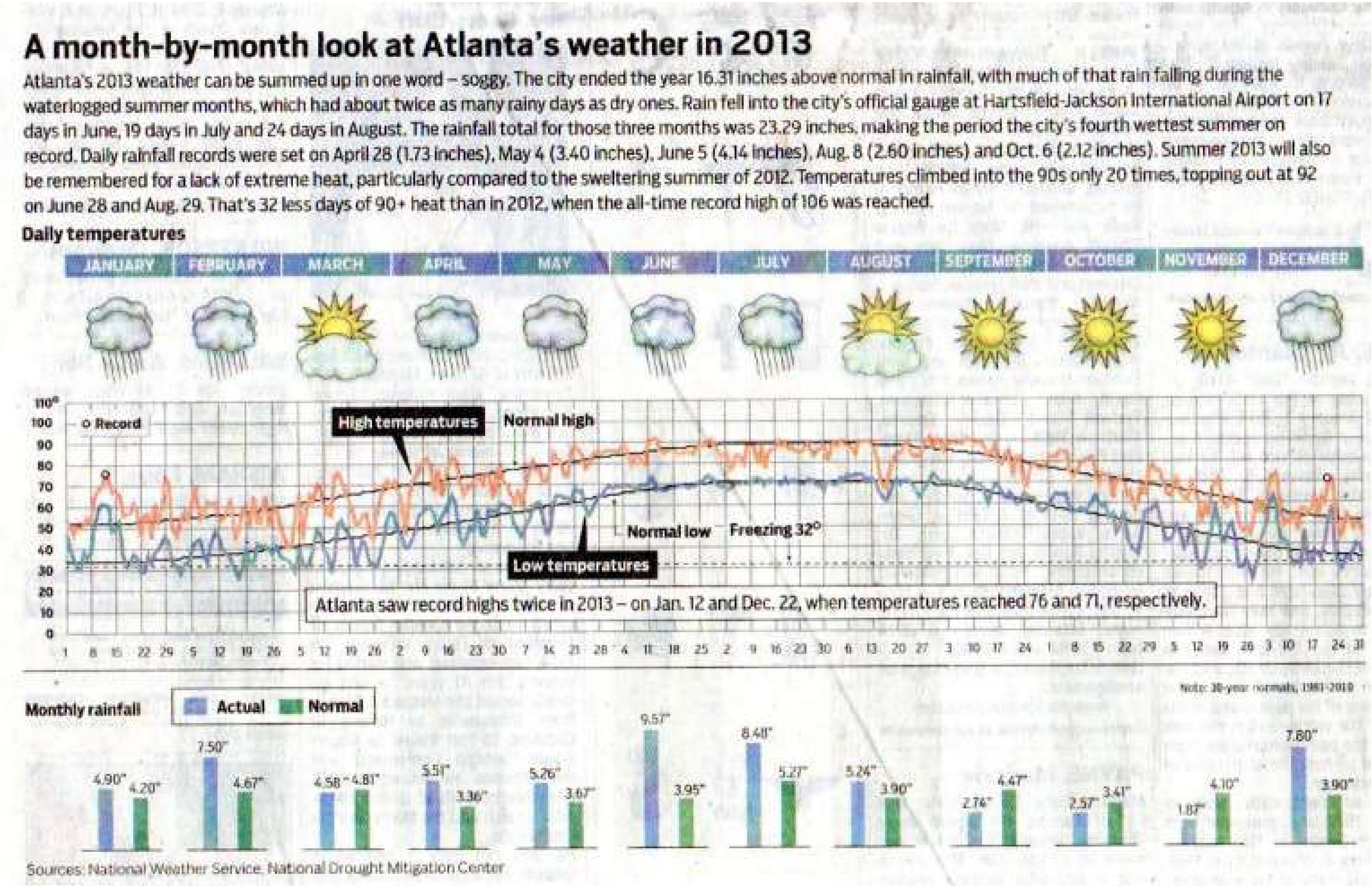
From E. Tufte, The Visual Display of Quantitative Information

New York Times, January 11, 1981, p. 32.

The value of Visualization

- From E. Tufte, The Visual Display of Quantitative Information. Data values:
  - 365 High, Low, Avg high, & avg low temp values for each day
  - 365 Precipitation & Humidity for each day
  - 12 Precipitation & Avg precipitation for each month
  - 1 Precipitation for the year & Avg precipitation per year
  - 1 Highest & Lowest temp (& day) for the year
  - 1 Avg daily temp for the year
  - 1 Avg daily temp per year

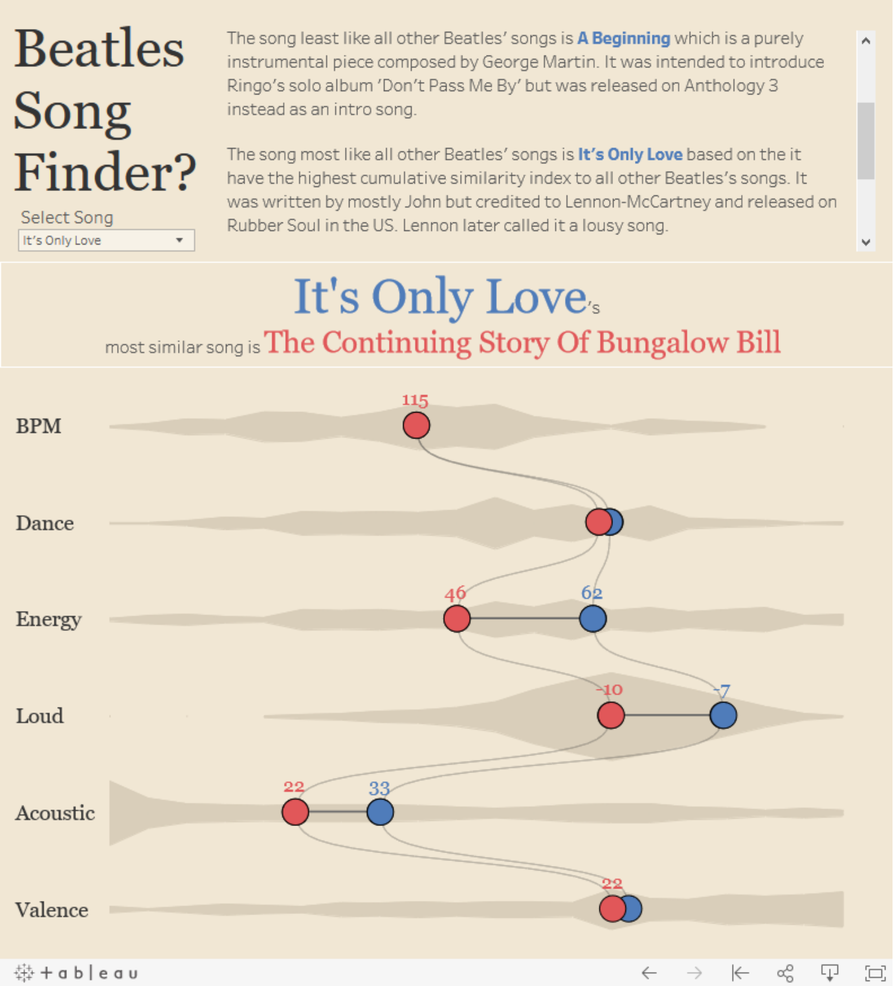
The value of Visualization



The value of Visualization

- Examples. Providing **confidence** on the data (e.g. values)

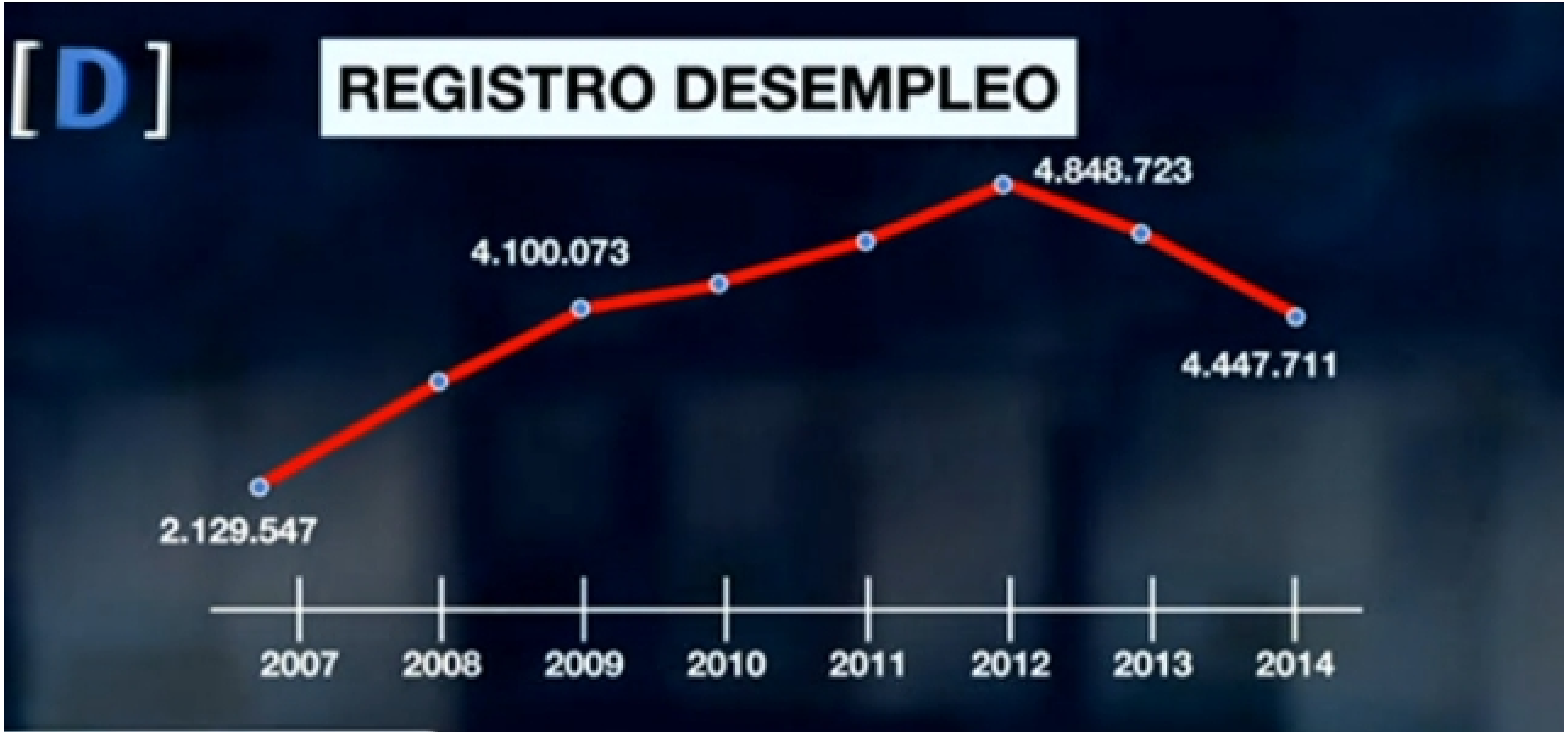
Finding most similar Beatles' songs pairs using the Euclidean distance between the song with 6 attributes with data from Spotify



From http://duelingdata.blogspot.com/

The value of Visualization

- Examples. Lack of **confidence**



The value of Visualization

- Examples. Lack of **confidence**

Outline

- *The value of visualization*
- *Tasks and Data types*
- **The visualization mantra**
- Takeaways
- Designing visualization systems

The Visualization Mantra

- Schneiderman's Mantra
  - overview first, zoom and filter, then details-on-demand
  - overview first, zoom and filter, then details-on-demand
  - overview first, zoom and filter, then details-on-demand
  - overview first, zoom and filter, then details-on-demand
  - overview first, zoom and filter, then details-on-demand
  - overview first, zoom and filter, then details-on-demand
  - overview first, zoom and filter, then details-on-demand

The Visualization Mantra

- Schneiderman's Mantra

1. Overview first
2. Zoom and filter
3. Then Details on Demand

The Visualization Mantra

- Schneiderman’s Mantra

1. Overview first: Provide **big picture** of the data, no details
2. Zoom and filter
3. Then Details on Demand

The Visualization Mantra

- Schneiderman’s Mantra

1. *Overview first: Provide big picture of the data, no details*
2. Zoom and filter: Focus on a **particular area** of the data
3. Then Details on Demand

The Visualization Mantra

- Schneiderman's Mantra

1. *Overview first: Provide big picture of the data, no details*
2. *Zoom and filter: Focus on a particular area of the data*
3. Then Details on Demand: Only when requested, **details of single data items**

Outline

- *The value of visualization*
- *Tasks and Data types*
- *The visualization mantra*
- **Takeaways**
- Designing visualization systems

Takeaways

- “Visualization is a method of computing. It transforms the symbolic into the geometric, **enabling researchers to observe their simulations and computations.** Visualization offers a method for **seeing the unseen.** It enriches the process of scientific discovery and fosters profound and unexpected insights. In many fields it is already revolutionizing the way scientists do science.”

McCormick, B.H., T.A. DeFanti, M.D. Brown, ***Visualization in Scientific Computing,*** Computer Graphics 21(6), November 1987

Takeaways

- "… exploring data and information …"
- "… to gain understanding and insight into the data."
- "… relying on the humans' powerful ability to visualize."
- "… in such a way as to allow the user to extract significant features and results quickly and easily."

K.W. Brodlie, L.A. Carpenter, R.A. Earnshaw, J.R. Gallop, R.J. Hubbard, A.M. Mumford, C.D. Osland, P. Quarendon, ***Scientific Visualization, Techniques and Applications***, *Springer-Verlag, 1992*

Takeaways

- Provide a frame of reference
- Visualization allows people to offload tasks that are cognitive demanding to the perceptual system, using carefully designed images as a form of *external memory*
  - We excel at perception
- It also provides means to use our pattern matching capabilities
  - We also excel at pattern matching
- Working area (portion of screen, sheet of paper with partial results…) is **augmenting memory**

Takeaways

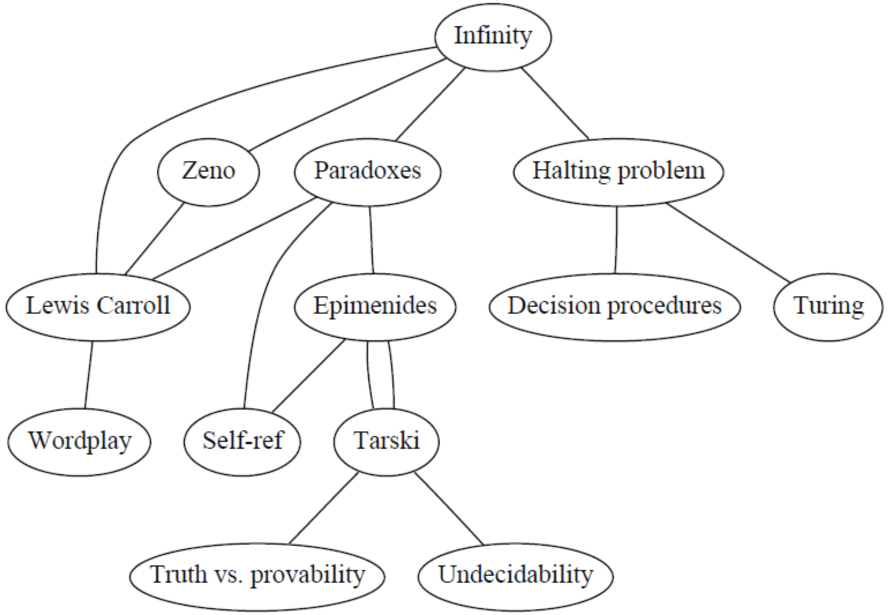
- Simple example. Understanding relationships between topics.
  - Given this list:

Infinity - Lewis Carroll	Epimenides - Self-ref
Infinity - Zeno	Epimenides - Tarski
Infinity - Paradoxes	Tarski - Epimenides
Infinity - Halting problem	Halting problem - Decision procedures
Zeno - Lewis Carroll	Halting problem - Turing
Paradoxes - Lewis Carroll	Lewis Carroll - Wordplay
Paradoxes - Epimenides	Tarski - Truth vs. provability
Paradoxes - Self-ref	Tarski - Undecidability

  - Find all the topics two hops away from the target topic *Paradoxes*

Takeaways

- Simple example. Understanding relationships between topics.
  - What if we show the information in a different way?



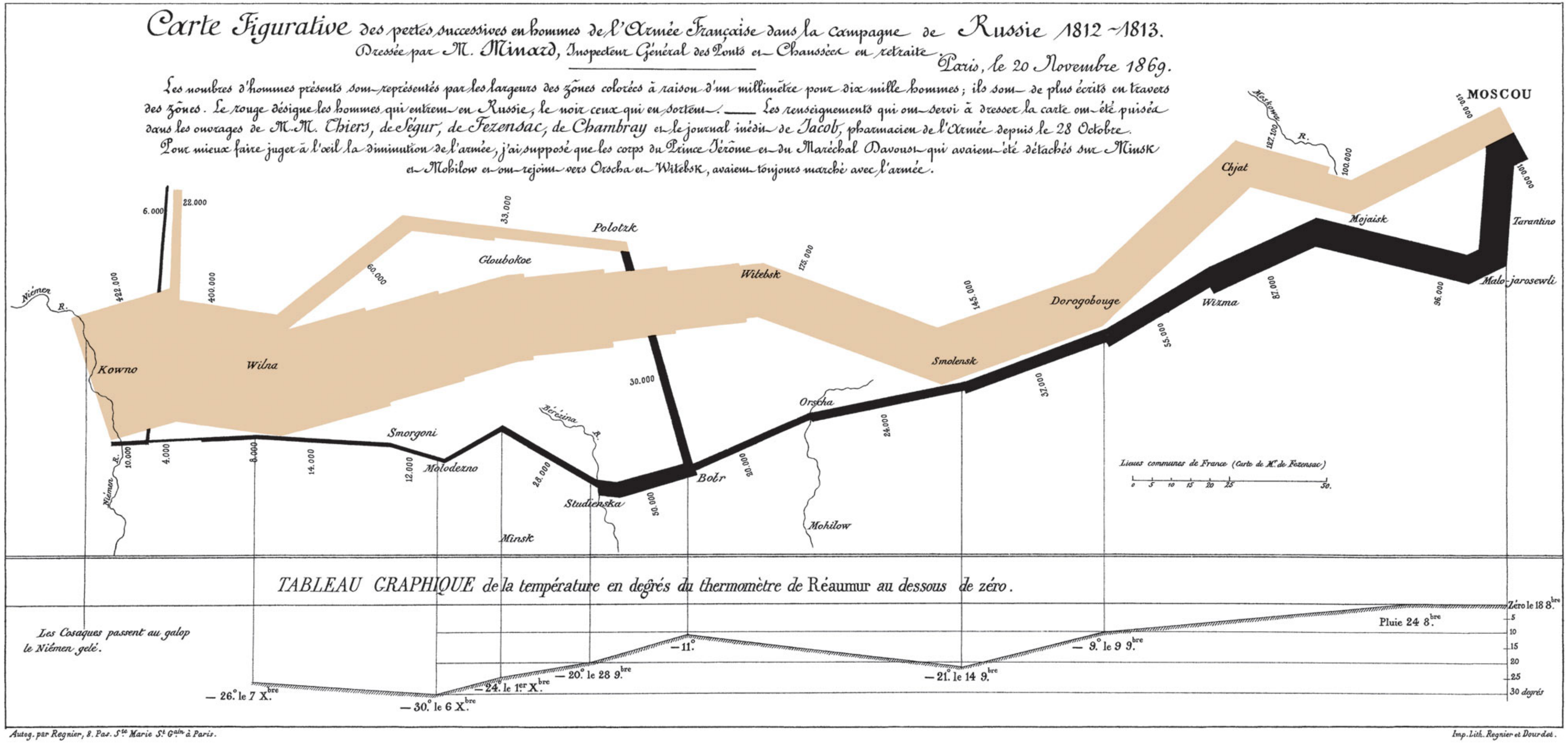
  - Find all the topics two hops away from the target topic *Paradoxes*

Takeaways

- Main applications of visualization
  - Presentation: Explanatory
  - Hypothesis testing: Confirmatory
  - Analysis: Exploratory

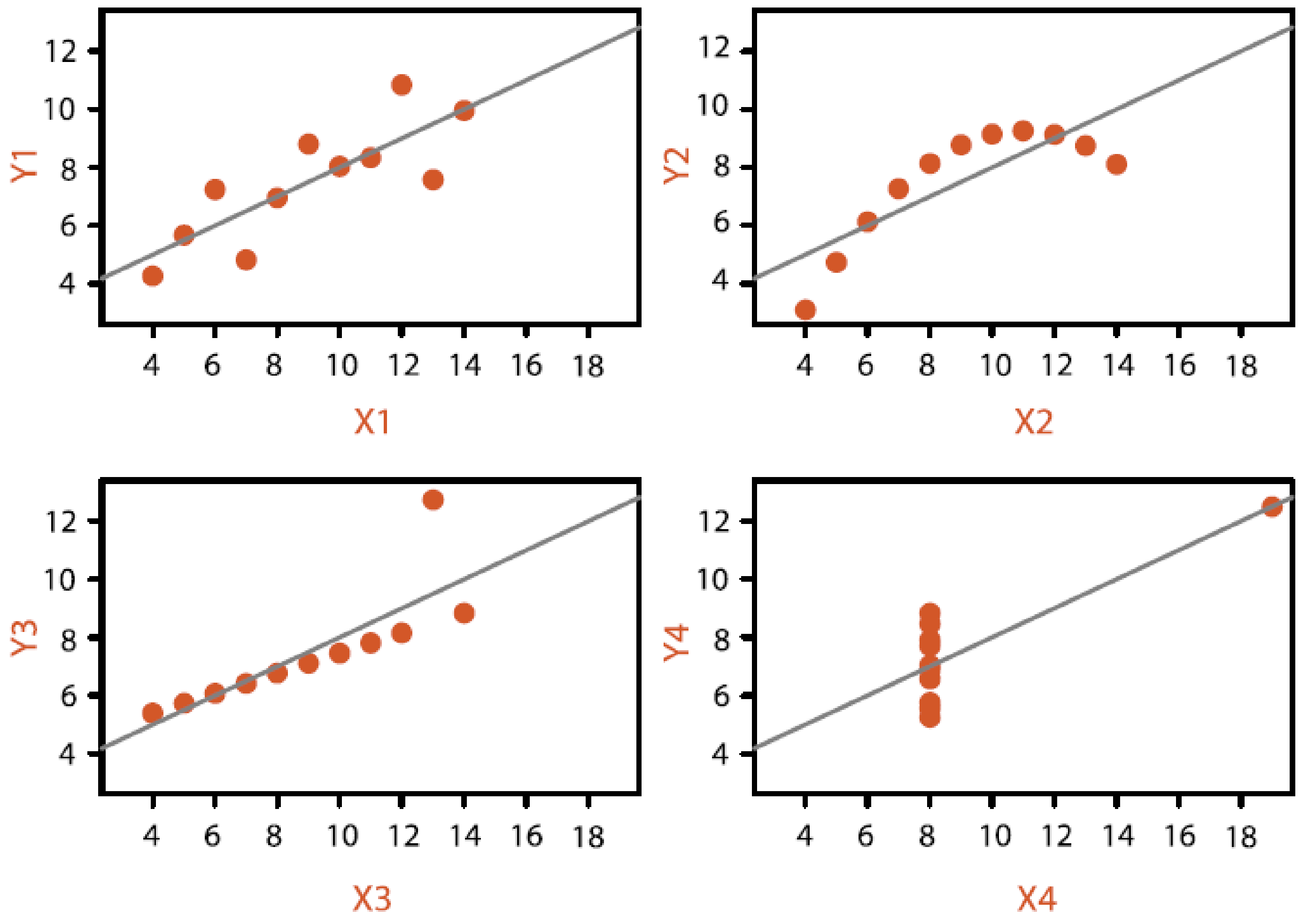
Takeaways

- Explanatory



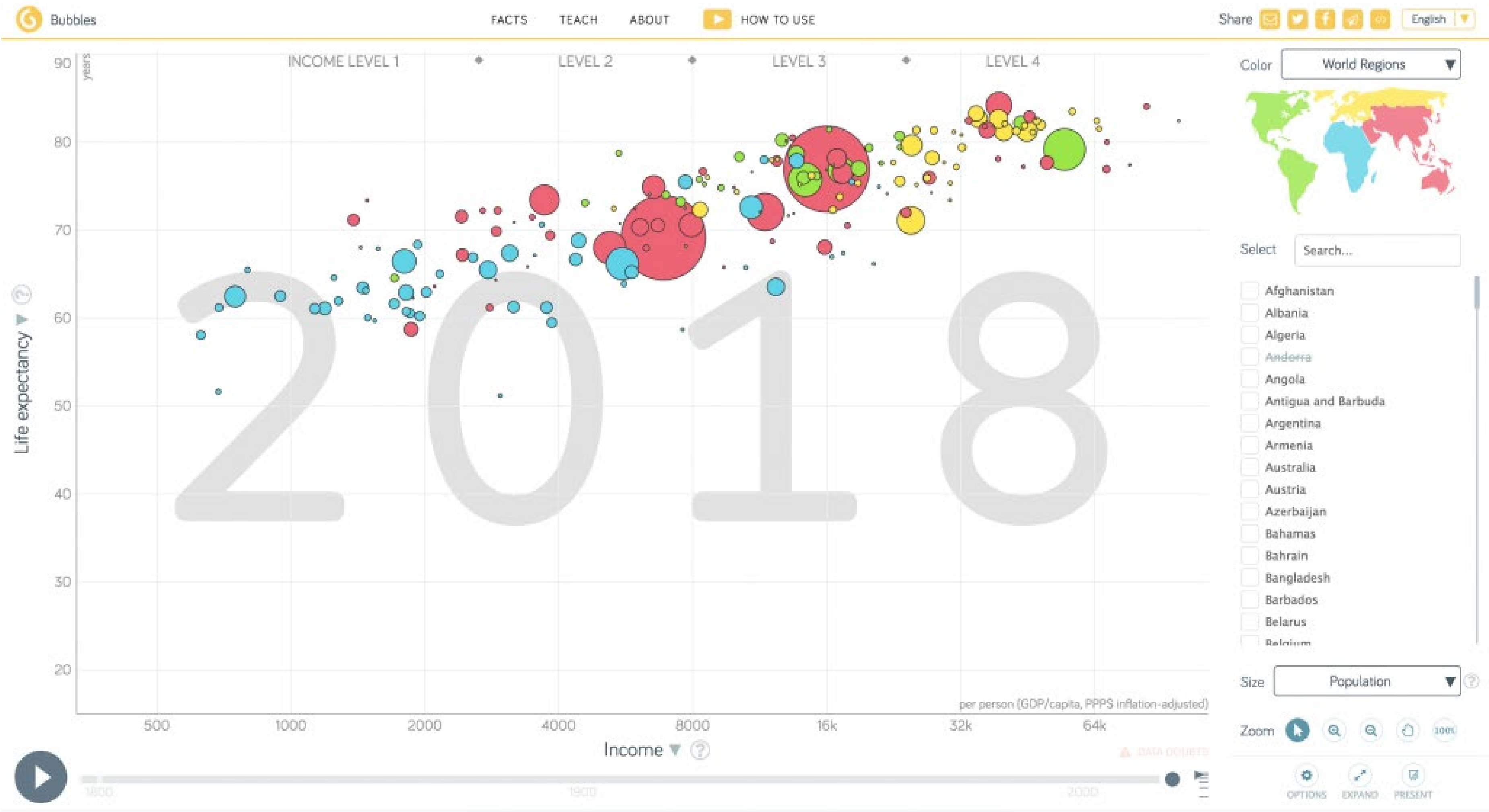
Takeaways

- Analysis
  - Confirmation



Takeaways

- Analysis
  - Exploration



Takeaways. Presentation

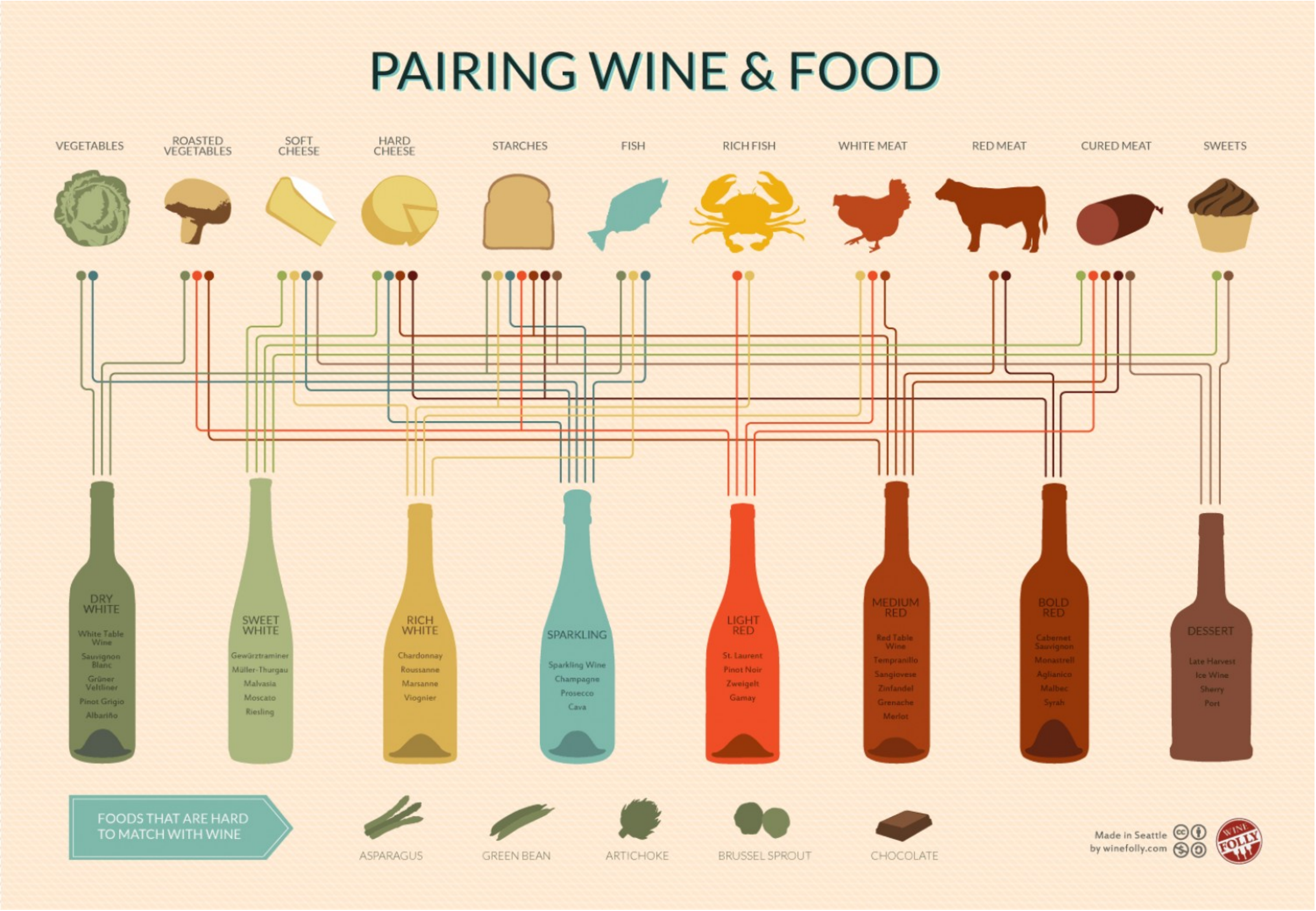
- Presentation
  - Communicate data and ideas
  - Explain and inform
  - Provide evidence and support
  - Influence and persuade
    - See also Rosling's 2009 TED talk "Let my dataset change your mindset" http://www.ted.com/talks/hans_rosling_at_state#t-236936

Takeaways. Presentation

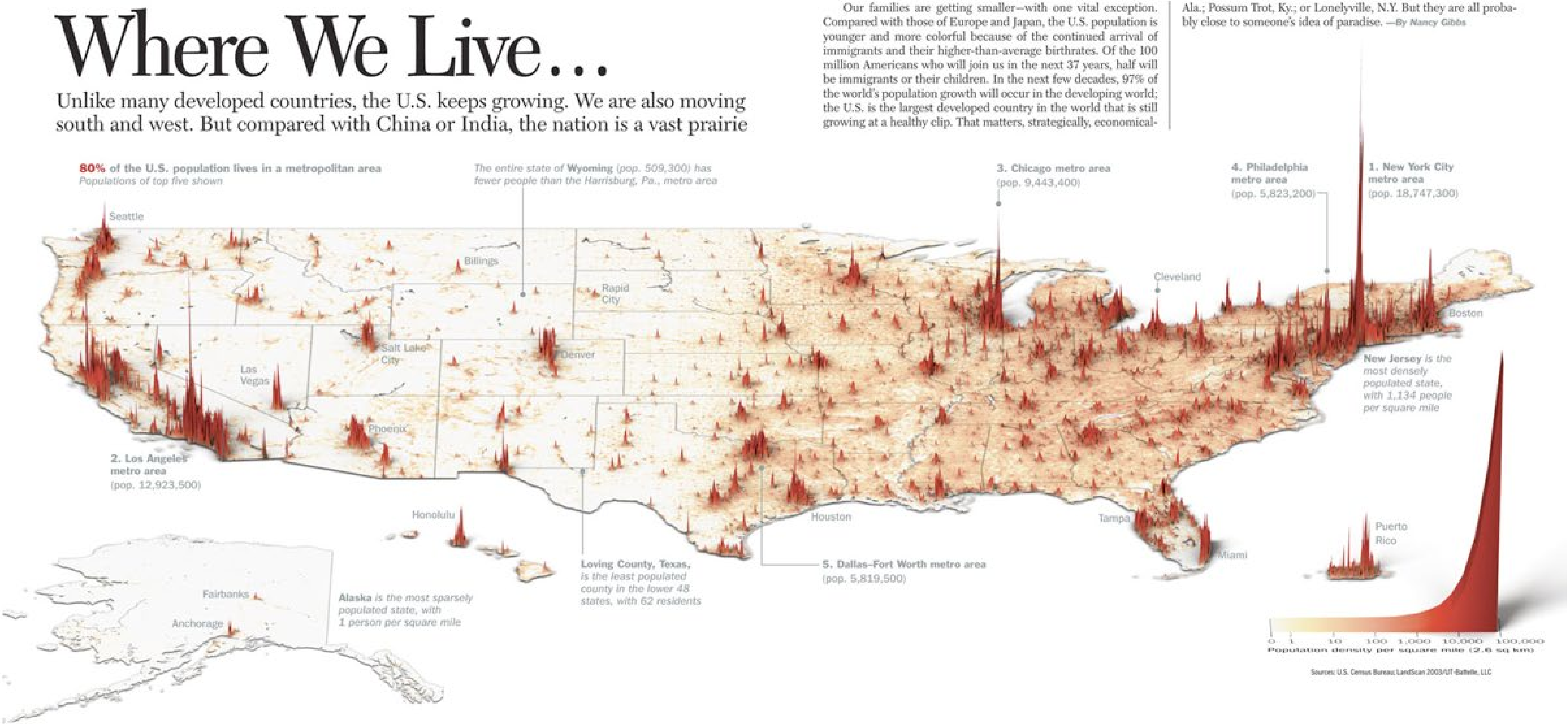
- Presentation. Infographics
  - A type of visualization that has become very popular lately
  - Note that sometimes there is a balance between accuracy and catching the eye of the reader
    - Take a look at the following example…

Takeaways. Presentation

- Presentation. Infographics



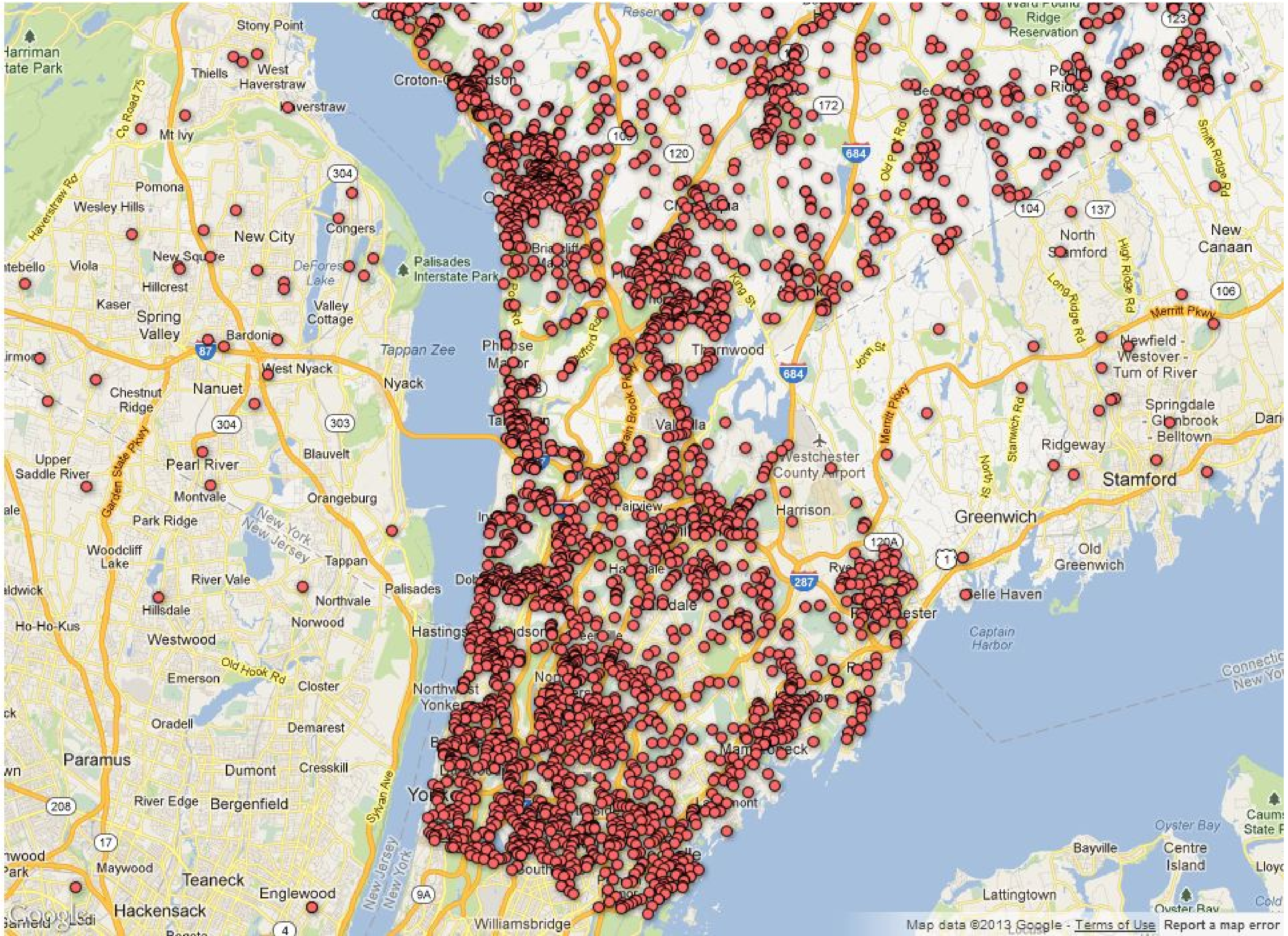
Takeaways. Presentation



Takeaways. Presentation

- Frequent presentation goals
  - Clarify
  - Focus
  - Highlight
  - Simplify
  - Persuade
- Commonly only showing a few variables of the data
  - Yet its impact may be HUGE!!!

Takeaways. Presentation



Takeaways. Presentation

- The journal was attacked
  - Names of the children and the schools where they went were published in newspapers…

Takeaways. Analysis

- Explore the data
- Assess a situation
- Determine how to proceed
- Decide what to do

Takeaways. Analysis

- Other methods for data analysis
  - Statistics
  - Database & information retrieval
  - Data mining
  - Machine learning

Takeaways. Analysis

- Typical objectives
  - Show many variables
  - Illustrative overview and detail
  - Facilitate comparison

- Presentation might choose some parts
  - Analysis will focus on **all of them**

Takeaways. Analysis

- Need of exploratory analysis applications:
  - Typically difficult to show information so that it can be interpreted properly
  - Visualization very useful in exploratory data analysis
    - Don’t know what you’re looking for (unknown insights)
    - Don’t have a priori questions (don’t know what the data looks like)
    - Want to know what questions to ask (further analysis)
  - Still requires effective communication

Takeaways

- Other way of presenting it… Visualization serves for …
  - … exploration
    - Nothing is known,
      Vis. used for gathering knowledge on the data
  - … analysis
    - There are hypotheses,
      Vis used verification or falsification
  - … presentation
    - "everything" known about the data,
      Vis. used for communication of results

Often together: Tools for Exploratory Analysis

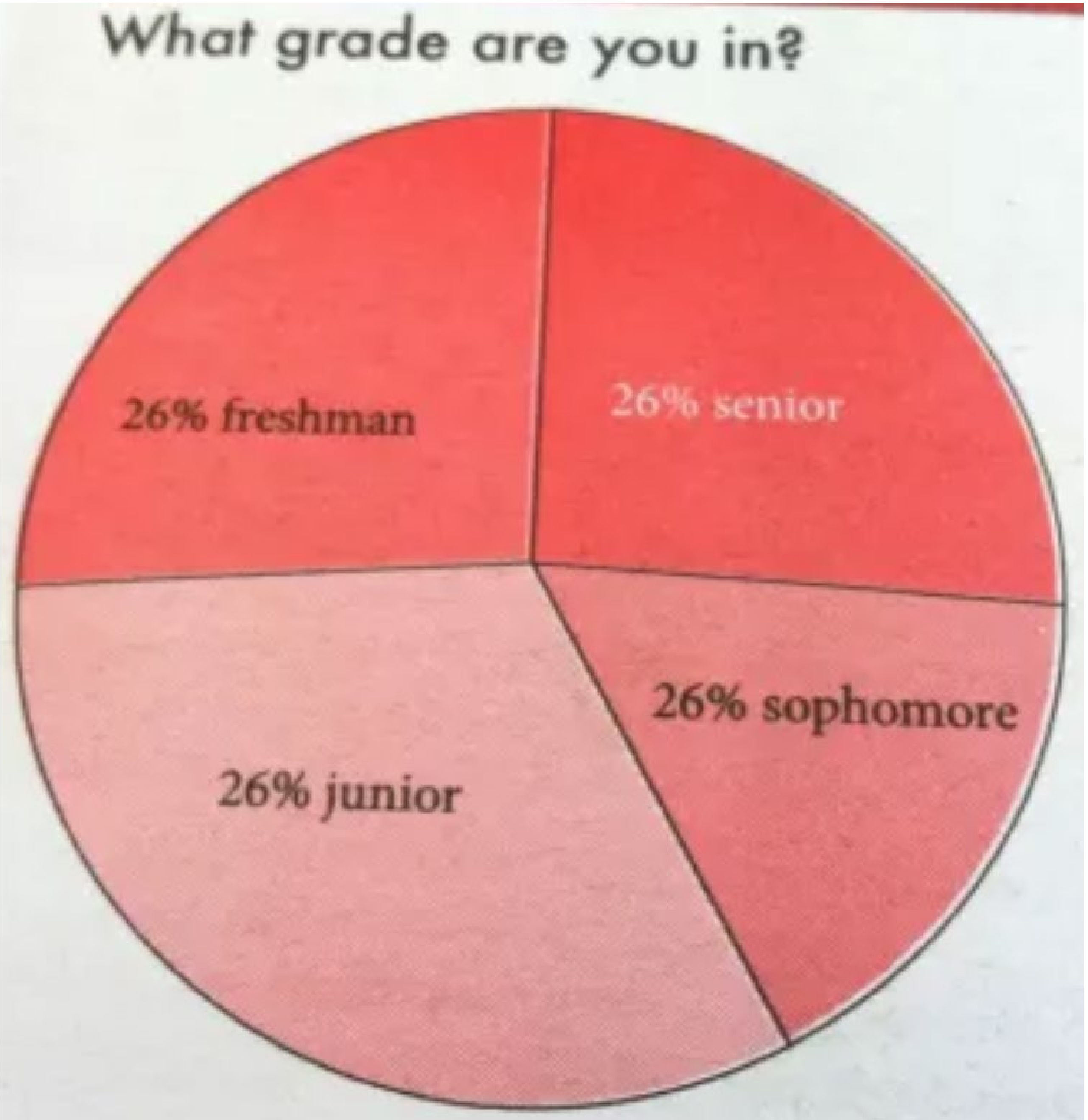
Takeaways

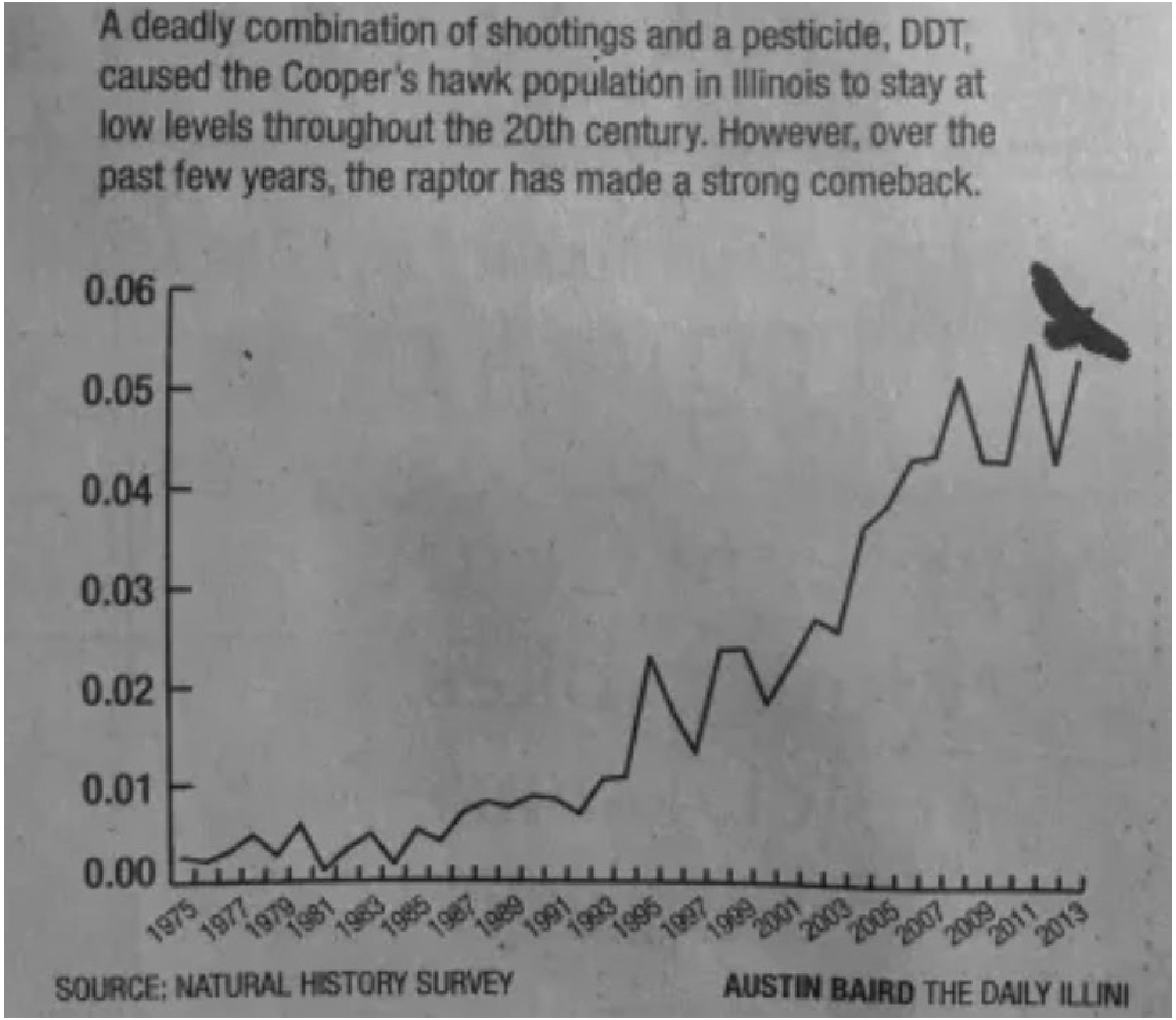
- The value of visualization comes determined by its purpose
  - It actually may be the best way to "…to allow the user to extract significant features and results quickly and easily."

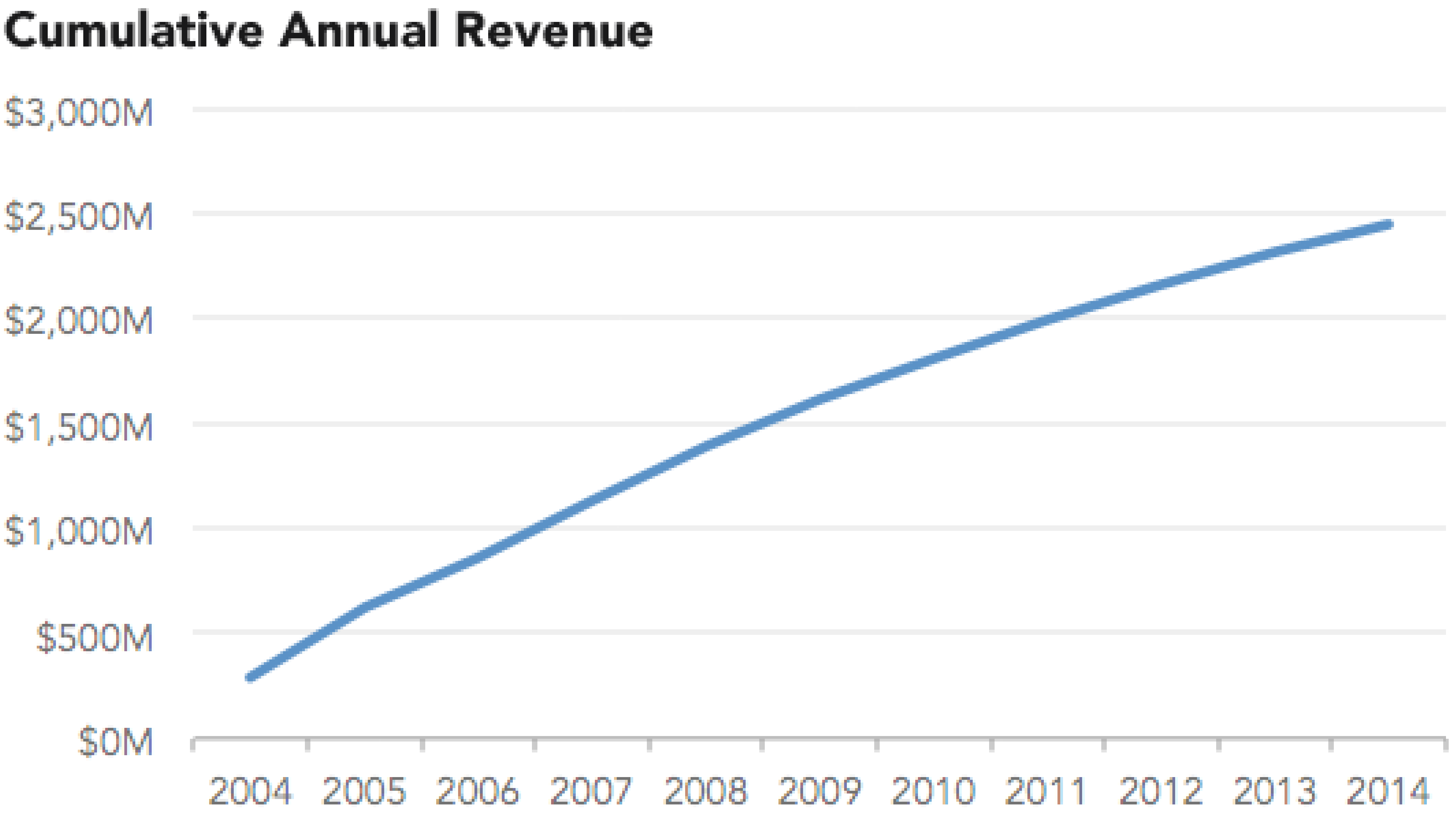
- Let's see more examples…

Takeaways

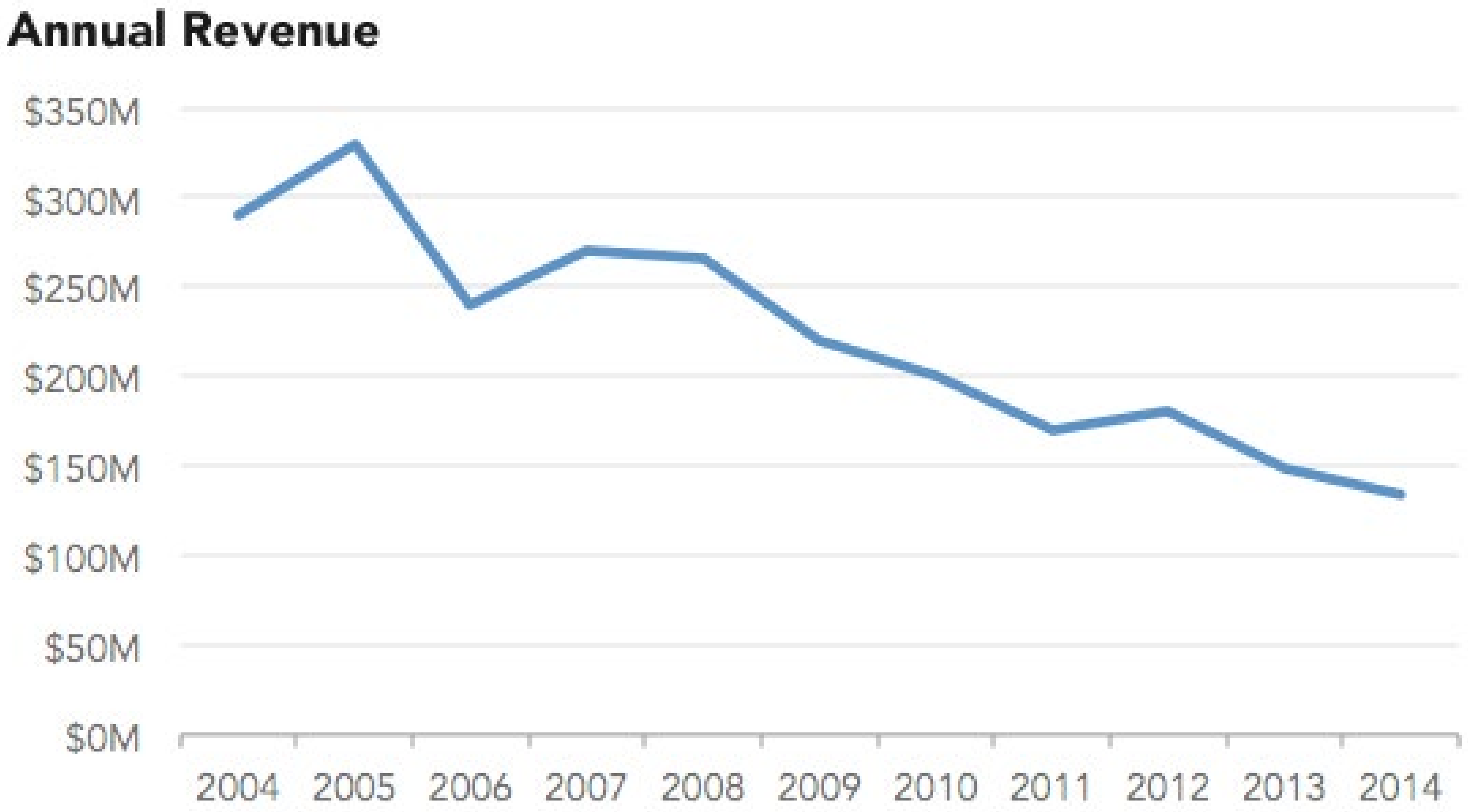

MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions
$300
250
200
150
100
50
1972
'74
'76
'78
'80
'82 est.
VOTE


What grade are you in?
26% freshman
26% senior
26% sophomore
26% junior

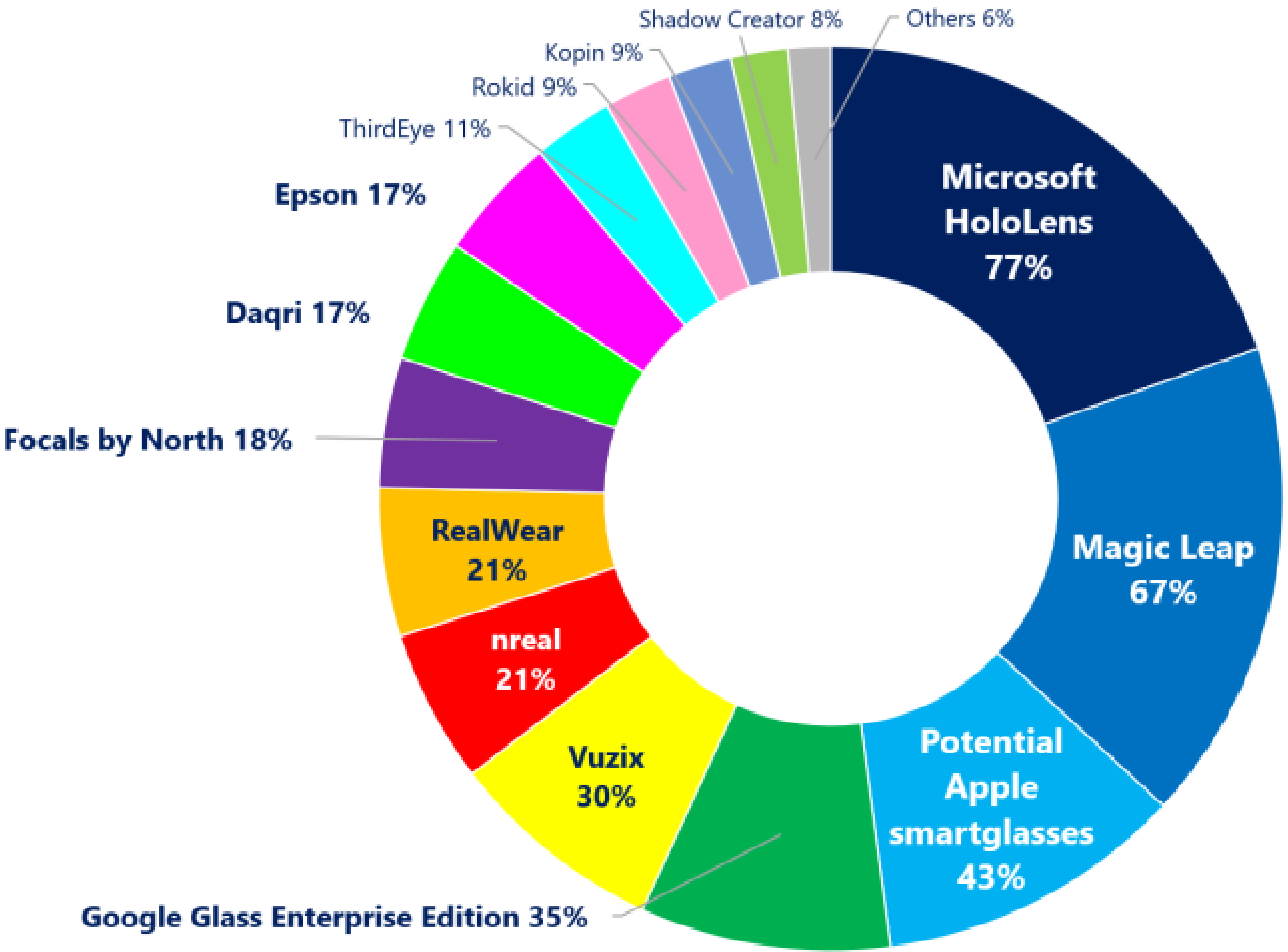

A deadly combination of shootings and a pesticide, DDT, caused the Cooper's hawk population in Illinois to stay at low levels throughout the 20th century. However, over the past few years, the raptor has made a strong comeback.
0.06
0.05
0.04
0.03
0.02
0.01
0.00
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
SOURCE: NATURAL HISTORY SURVEY
AUSTIN BAIRD THE DAILY ILLINI



Cumulative Annual Revenue
$3,000M
$2,500M
$2,000M
$1,500M
$1,000M
$500M
$0M
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014


Annual Revenue
$350M
$300M
$250M
$200M
$150M
$100M
$50M
$0M
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

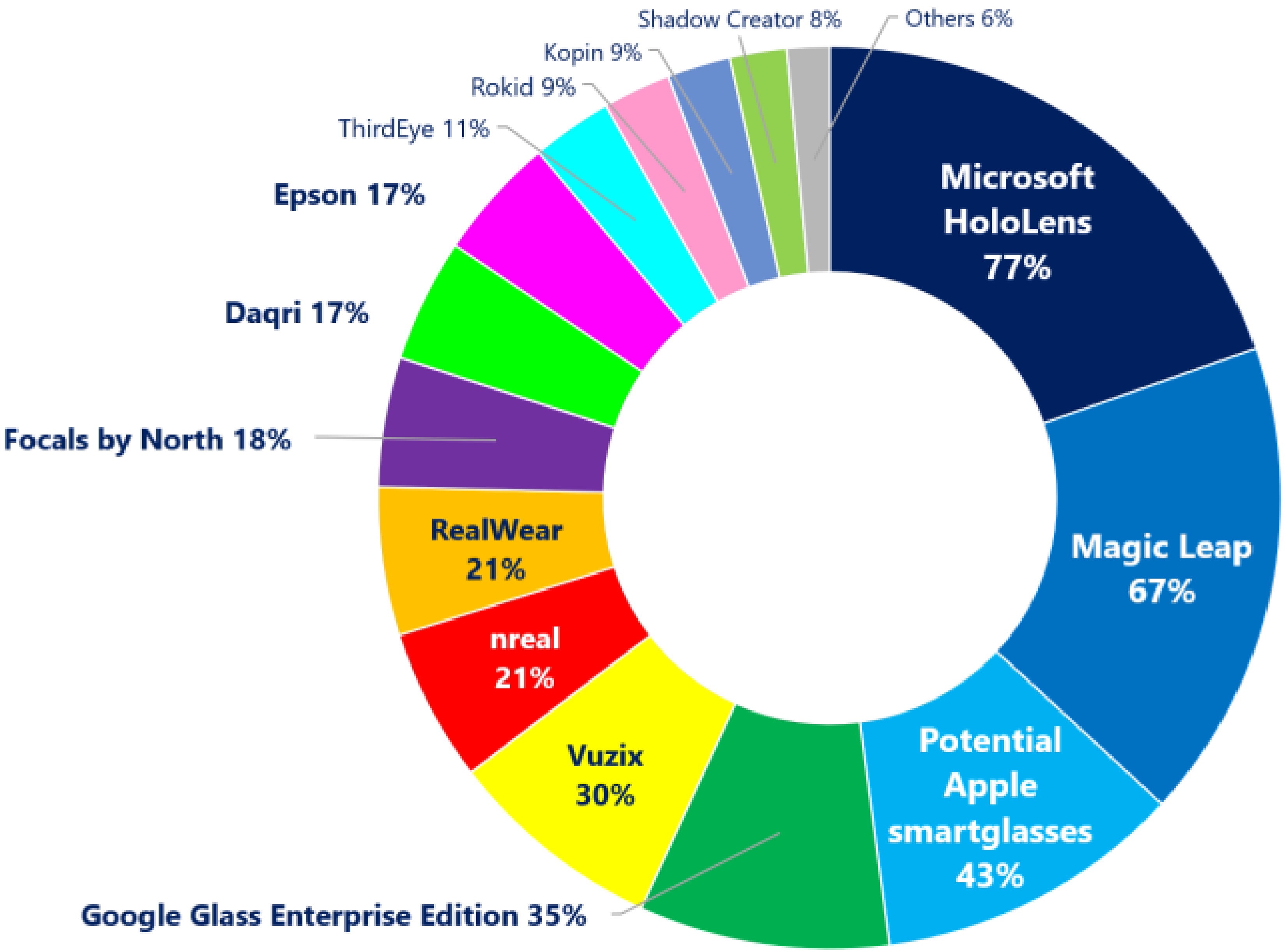
Industry smartglasses platform focus



Digi-Capital

Notes: Percentages not cumulative, "Others" include DreamGlass, EverySight, Glassear, IMMY, Iristick, Lenovo, Letinar, Lumus, Optinvent, Realmax, Snap Spectacles, Tilt Five

Industry smartglasses platform focus

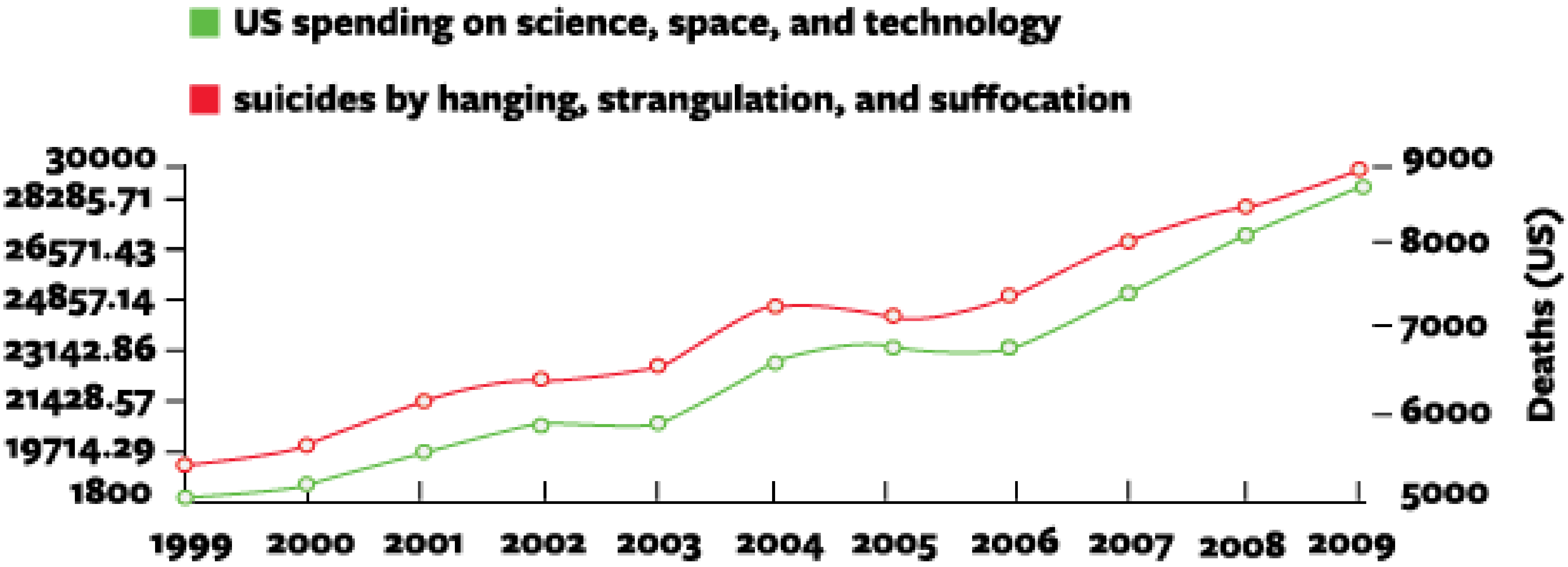


389% total in this pi
chart?!?
Digi-Capital is asking
people to pay $150
for this report!

Digi-Capital

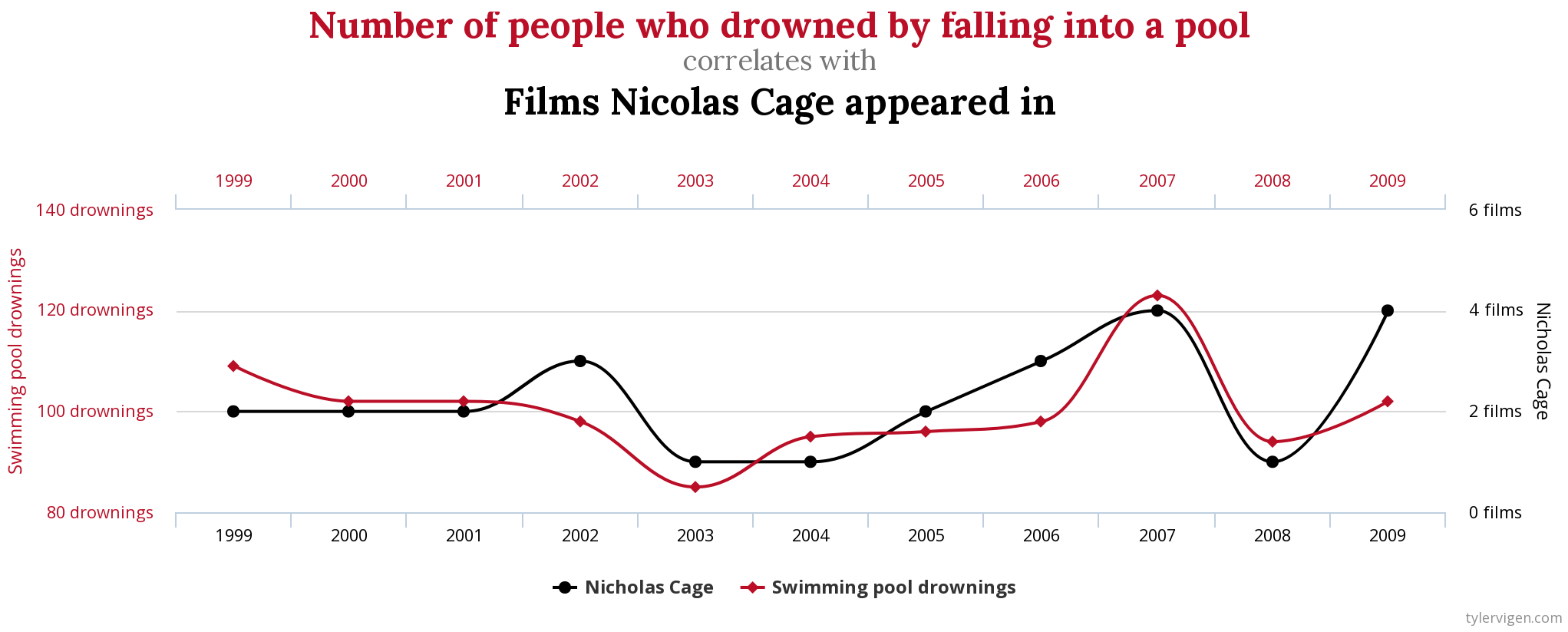
Notes: Percentages not cumulative, “Others” include DreamGlass, EverySight, Glassear, IMMY, Iristick, Lenovo, Letinar, Lumus, Optinvent, Realmax, Snap Spectacles, Tilt Five

Takeaways

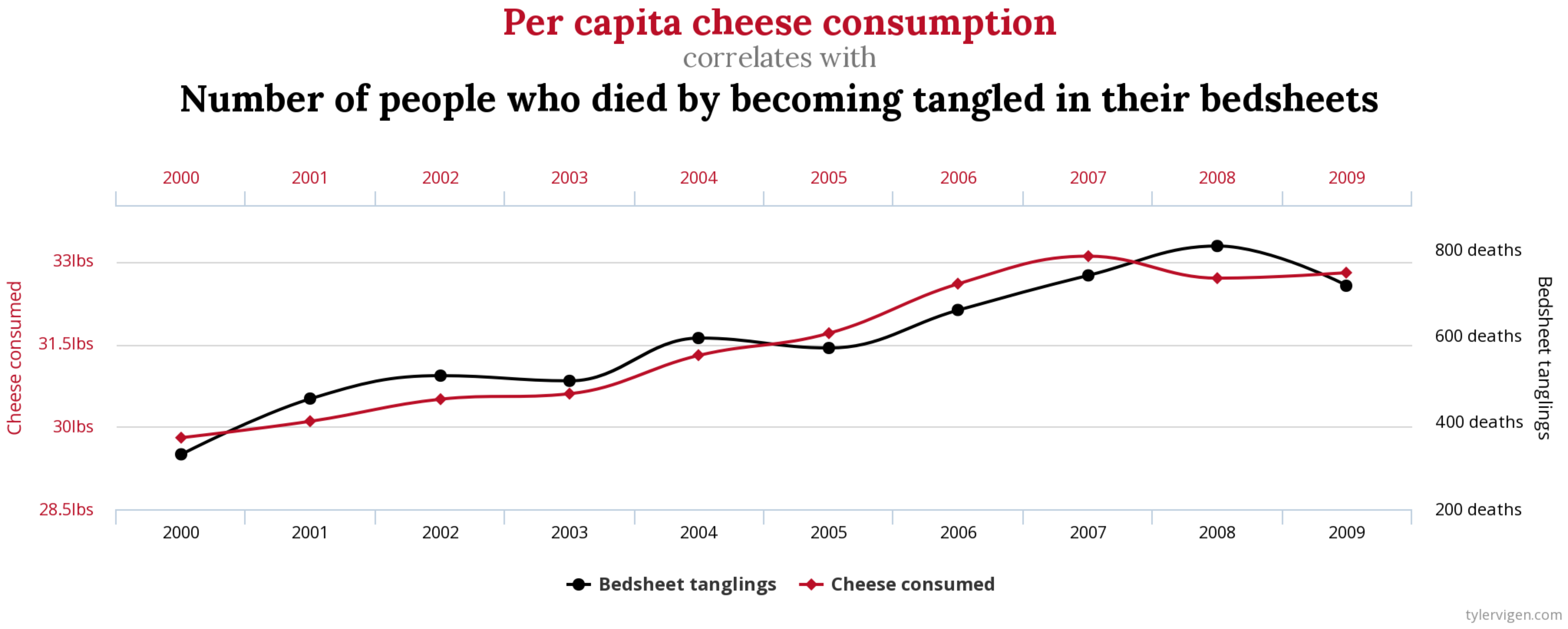


US spending on science, space, and technology
suicides by hanging, strangulation, and suffocation
30000
28285.71
26571.43
24857.14
23142.86
21428.57
19714.29
1800
9000
8000
7000
6000
5000
Deaths (US)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009

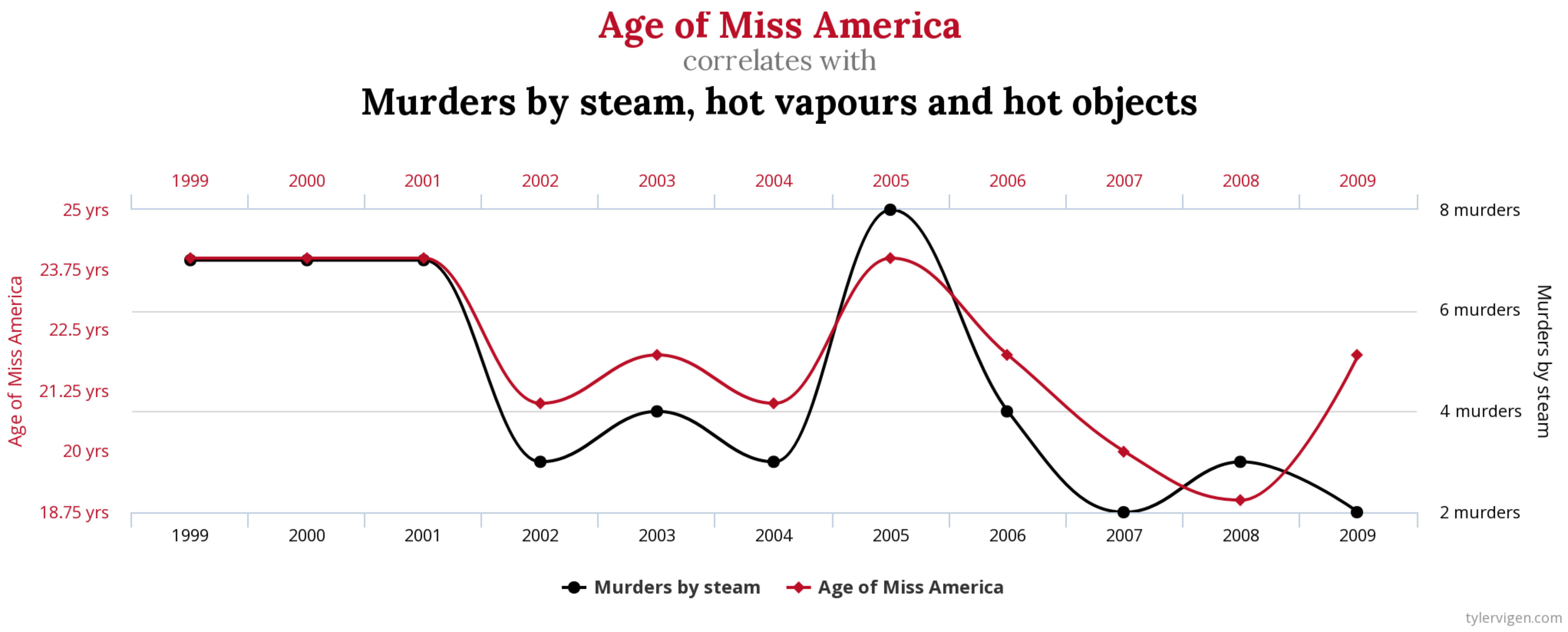
Takeaways


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drownings
120 drownings
100 drownings
80 drownings
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
Nicholas Cage
Swimming pool drownings
tylervigen.com

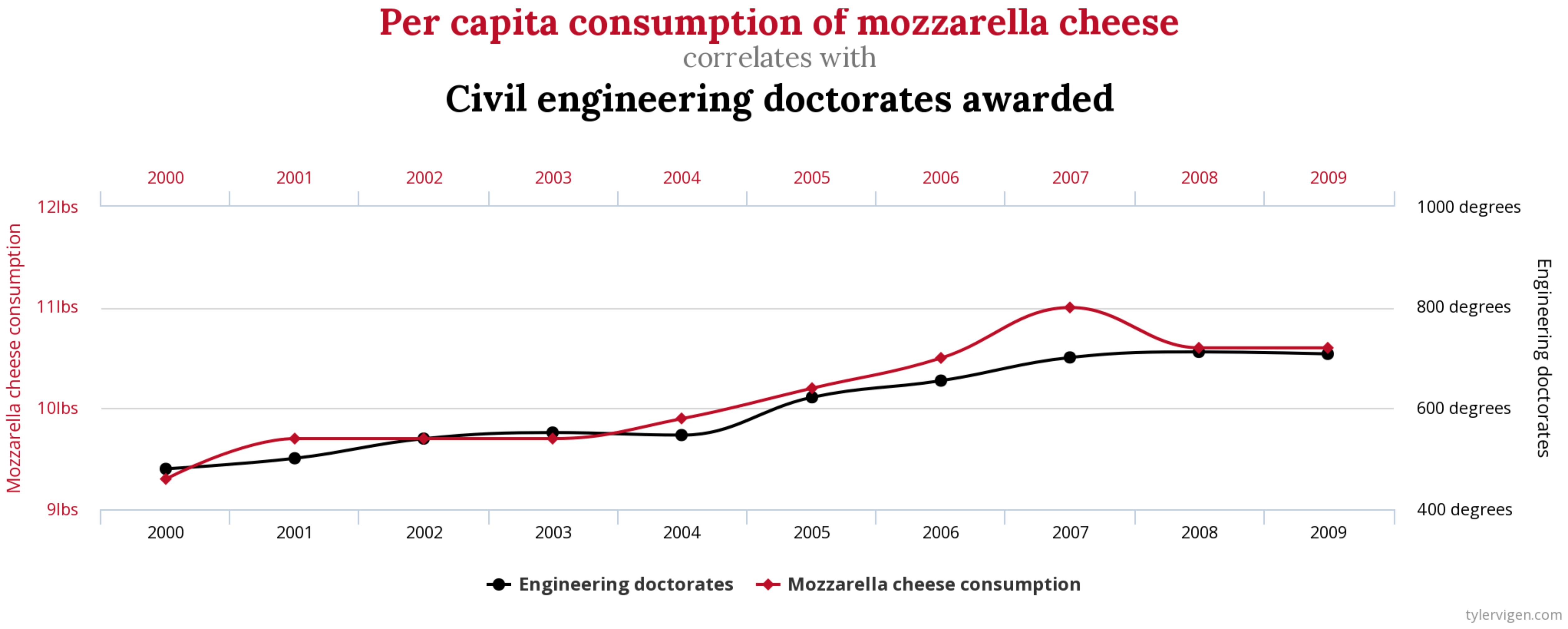
Takeaways


Per capita cheese consumption
correlates with
Number of people who died by becoming tangled in their bedsheets
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
Cheese consumed
33lbs
31.5lbs
30lbs
28.5lbs
800 deaths
600 deaths
400 deaths
200 deaths
Bedsheet tanglings
Bedsheet tanglings
Cheese consumed
tylervigen.com

Takeaways

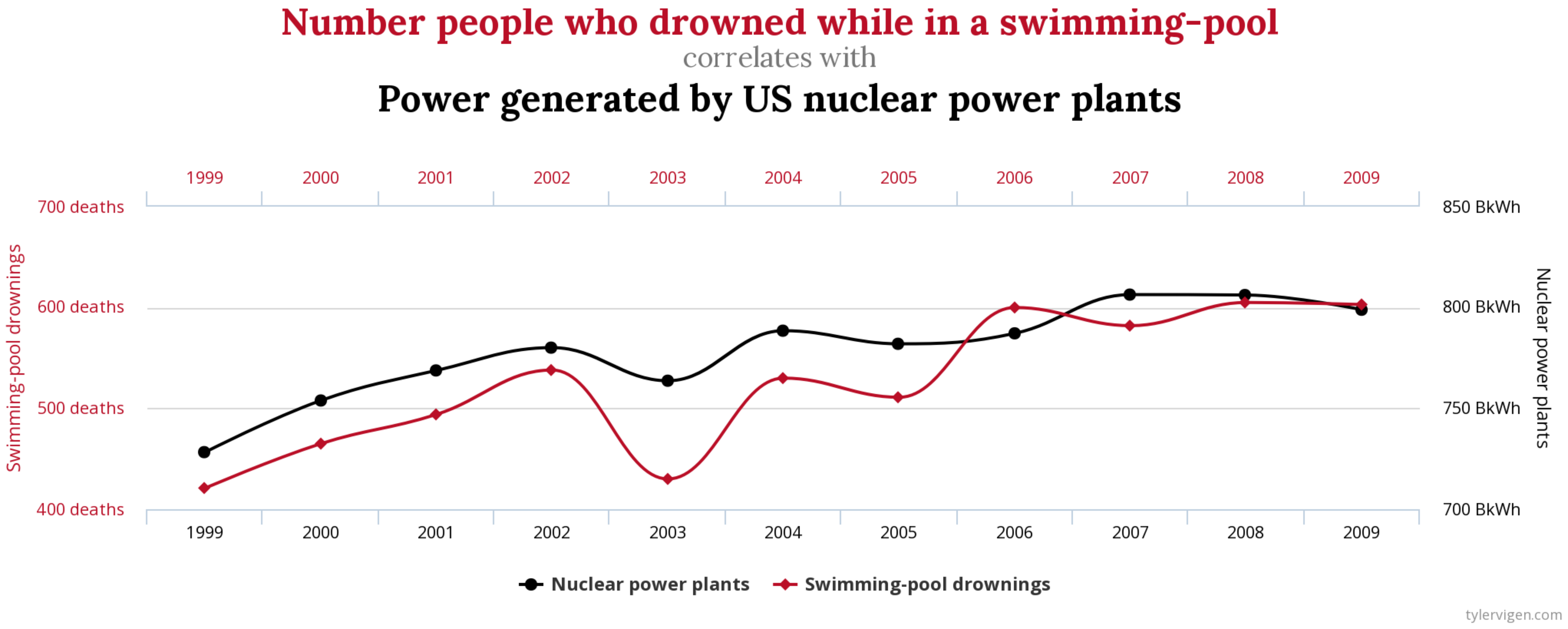

Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
25 yrs
23.75 yrs
22.5 yrs
21.25 yrs
20 yrs
18.75 yrs
Age of Miss America
8 murders
6 murders
4 murders
2 murders
Murders by steam
Murders by steam
Age of Miss America
tylervigen.com

Takeaways



Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
12lbs
11lbs
10lbs
9lbs
Mozzarella cheese consumption
1000 degrees
800 degrees
600 degrees
400 degrees
Engineering doctorates
Engineering doctorates
Mozzarella cheese consumption
tylervigen.com

Takeaways



Number people who drowned while in a swimming-pool
correlates with
Power generated by US nuclear power plants
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
700 deaths
600 deaths
500 deaths
400 deaths
Swimming-pool drownings
850 BkWh
800 BkWh
750 BkWh
700 BkWh
Nuclear power plants
Nuclear power plants
Swimming-pool drownings
tylervigen.com

Outline

- *Tasks and Data types*
- *The value of visualization*
- *The visualization mantra*
- Takeaways
- **Designing visualization systems**

Designing visualization systems

- Limitations:
  - Computational capacity
  - Human perceptual and cognitive capacity
  - Display capacity

Designing visualization systems

- Computational capacity:
  - Computer time and memory are limited resources
  - If the visualization system needs to deliver interactive response, algorithms must be efficient enough
    - Better a fraction of a second than minutes or hours

Designing visualization systems

- Human perceptual and cognitive capacity:
  - Memory and attention are finite resources

Designing visualization systems

- Memory notoriously limited
  - Both for long-term recall and for shorter-term working memory
  - Surprisingly little information is internally stored in visual working memory
    - Vulnerable to *change blindness*
      - Large changes not noticed if attending to something else
- Vigilance is a highly limited resource
  - Ability to perform visual search tasks degrades quickly
    - Far worse results after several hours than in the first few minutes

Designing visualization systems

- Display capacity
  - Often not enough space for the desired result
    - Screen resolution is not large enough to show all desired information simultaneously.
  - The *information density* of a is a measure of the amount of information encoded versus the amount of unused space
    - Must carefully choose appropriate amount of information
    - Also referred to (at least partially) as Ink-ratio (see any of Tufte's books, e.g. "Envisioning information")

Designing visualization systems

- Display capacity. Tradeoffs information density:
  - Much information
    - Reduces the need for navigation
    - Reduces the need for exploration
    - User may be overwhelmed by visual clutter
  - White space
    - Tradeoff between clutter and wasting space, find sweet spot between dense and sparse

Designing visualization systems

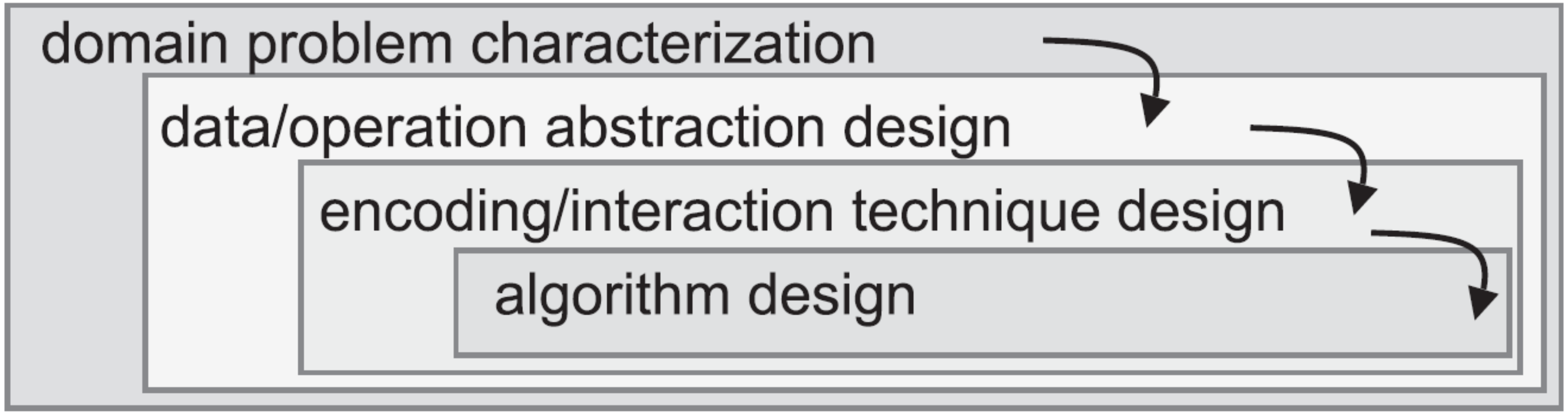
- Visualization design is a Human-Centered task
  - “Overview first, zoom and filter, details on demand”

Shneiderman, 1996

- Overviews help the user notice regions where further investigation might be productive
  - Here comes the hugely important problem of interaction

Designing visualization systems

- Visualization design can be split into a cascading set of layers



Designing visualization systems

- Problem characterization:
  - A given dataset will surely have many possible visual encodings
  - Choosing the right visual encoding can be guided by the specific needs of some intended user
  - Different questions, or *tasks*, require very different visual encodings
    - Recall the Instagram visualization example before
      - While it allows to understand what happens with cat pictures, it does not give you clues to increase the number of "likes" to your own pictures

Designing visualization systems

- Problem characterization:
  - Understanding the requirements of some target audience is a tricky problem
    - One commonly works with a group of target users over time
    - Users know they need to somehow view their data
      - They cannot directly articulate their needs as clear-cut tasks
        - » In terms of operations on data types
    - Iterative design including gathering information from the target is a must

Designing visualization systems

- Data abstraction
  - Once the domain problem has been identified, it must be abstracted into a more generic representation
  - Problems of different domains may map to the same visualization abstraction
  - Operations can be performed for the abstraction
    - Generic operations: Sorting, filtering, characterizing trends and distributions, finding anomalies, correlation, outliers…
    - Specific operations: Following a path in a graph…
  - Abstraction involves transformations from the raw data to the derived dimensions
    - Often different type from the original data
      - Tabular data to a graph, graph to a tree…

Designing visualization systems

- Technique and algorithm design:
  - Select appropriate visual encoding
  - Determine (& implement) adequate interaction techniques

Designing visualization systems

- Validation
  - Each layer can be validated
    - Each of them have different validation requirements
  - **Problem domain:** Is the target audience performing tasks that would benefit from the visualization tool?
    - Different measures could be applied, but some of them (e. g. adoption rate) can only be evaluated once the tool has been developed
  - **Abstraction:** Are the developed data types usable to solve the problem.
    - A field study might assess whether the audience is using the tool and how.
    - Images might be analysed quantitatively and qualitatively.

Designing visualization systems

- Validation
  - **Visual encoding:** Do the visual encoding communicate effectively the abstraction?
    - Ensure no perceptual and cognitive principles are violated.
    - Formal studies can also be carried out, take measures
  - Algorithm/technique design: Verify the designed algorithm to visualize and render is faster, takes less memory
    - Could analyse & compare computational complexity
    - Time in performance can be measured

Designing visualization systems

- Interaction
  - Problem:
    - You have a lot of data (& attributes) to understand
  - Do you?
    - Pack all the data into one complex representation
    - Spread the data into multiple coordinated views
    - ***Use interaction to reveal different subsets of the data***

Designing visualization systems

- Interaction
  - "The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means."

S. Few, *Now you see it*

- Two key aspects of data visualization
  - Representation
  - Interaction
    - Interaction is Vital
      - Engage in a dialog with your data

Designing visualization systems

- Why interact?
  1. Select
  2. Explore
  3. Reconfigure
  4. Encode
  5. Abstract/Elaborate
  6. Filter
  7. Connect

Designing visualization systems

- How we do interact?
  - Traditional – Desktop: keyboard, mouse
  - New – Tablet: fingers and multitouch

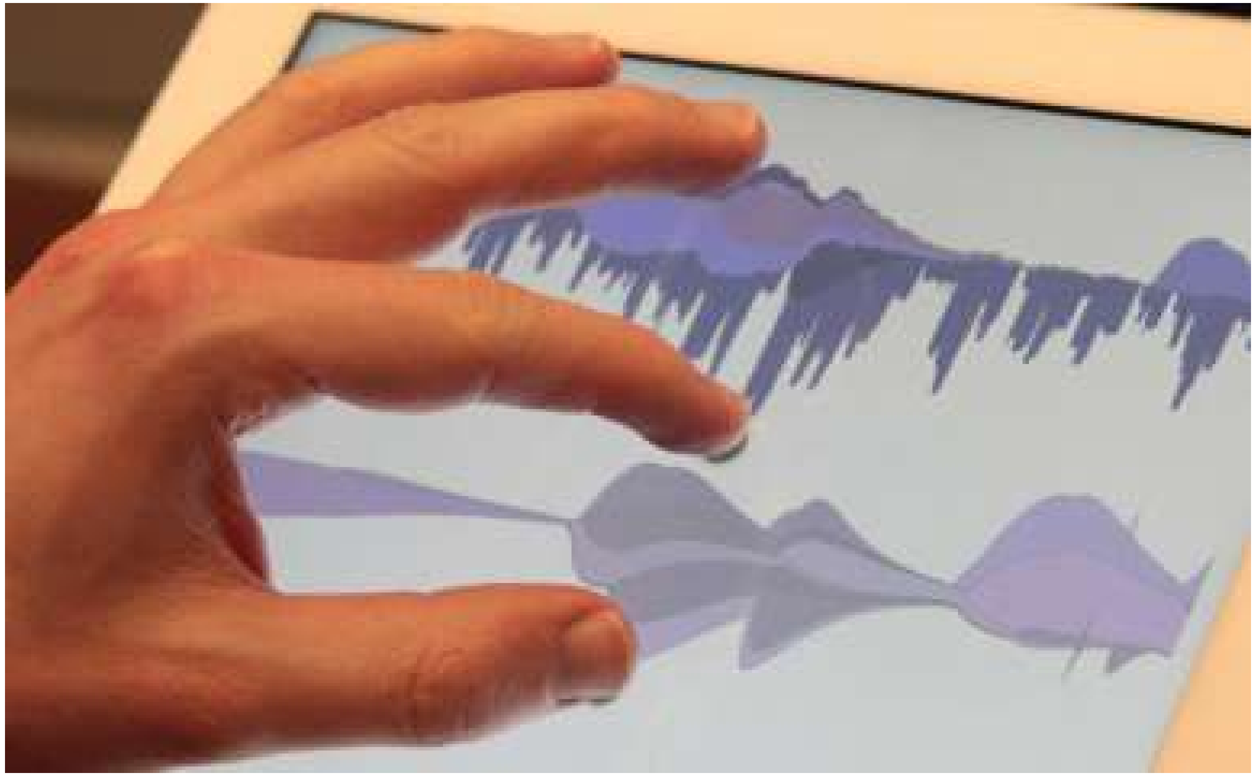
Designing visualization systems

- Interacting in a mobile
  - Volume Rendering in Android
  - Transfer function definition with touch
  - Samsung Galaxy S (4.3”)

Designing visualization systems

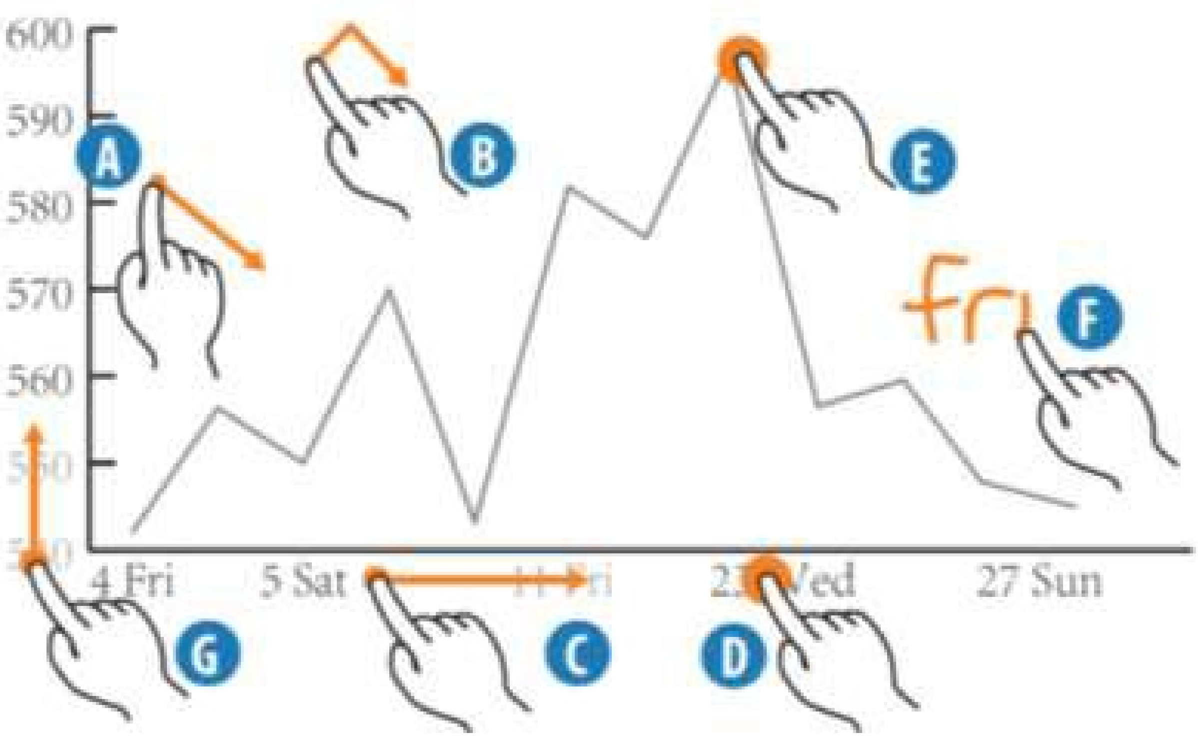
Designing visualization systems

- Interacting in a tablet
  - TouchWave (Baur, Lee & Carpendale, 2012)

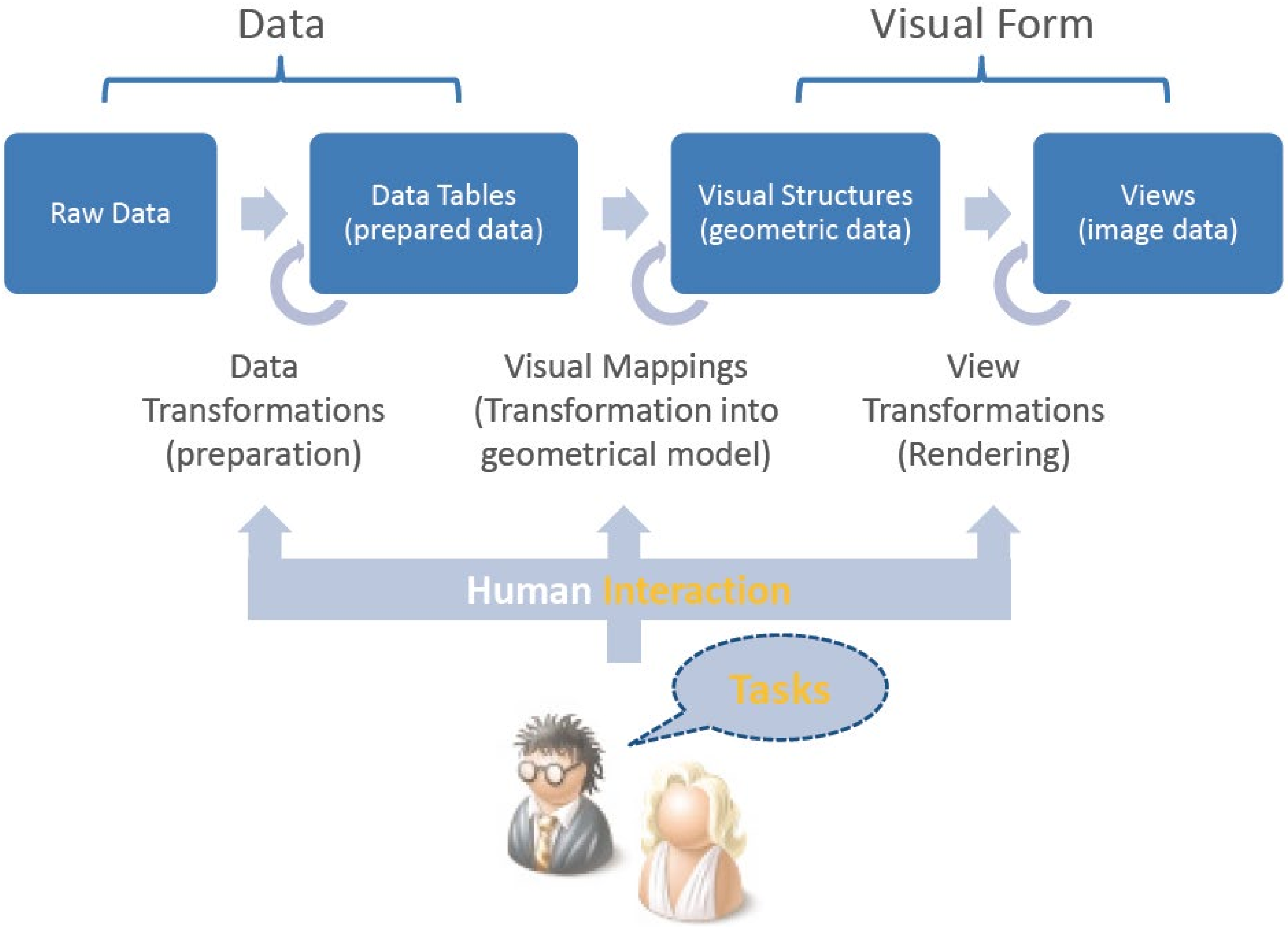


Designing visualization systems

- Interacting in a tablet
  - TouchWave (Baur, Lee & Carpendale, 2012)



Designing visualization systems



Designing visualization systems

- Raw Data: idiosyncratic formats
- Data Transformations: Mapping raw data into an organization appropriate for visualization
- Data Tables: relations (cases by variables) + metadata
- Visual Mappings: Encoding abstract data into a visual representation
- Visual Structures: spatial substrates + marks + graphical properties
- View Transformations: Changing the view or perspective onto the visual presentation
- Views: Graphical parameters (position, scaling, clipping, …)
- Human Interaction: User influence at any level

Thanks

- The information and some of the images come from different presentations/papers/course notes/books from:
  - A. Vilanova, T. Munzer, M. E. Groeller, M. Hadwiger, D. Weiskopf, T. Möller, R. Peikert, P. Muigg, J. Stasko, S. Oeltze-Jafra, J. Bertin, E. Tufte… and/or other papers
  - Thanks to all of them!

Visualization. Basics

Pere-Pau Vázquez
Dept. Computer Science – UPC