Data Visualisation
Lecture 2 – Perception and Categorisation

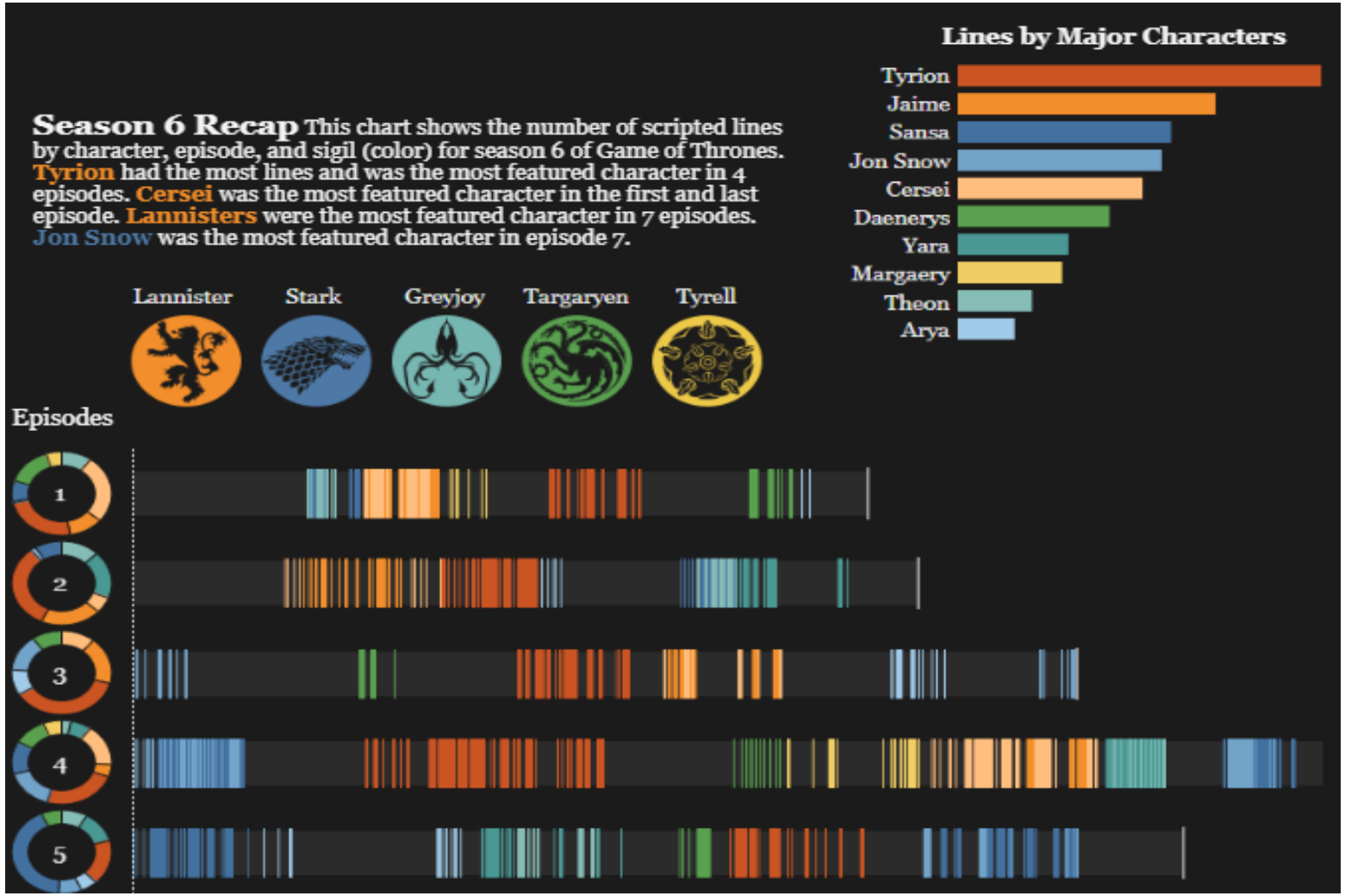
Dr. Cathy Ennis

cathy.ennis@tudublin.ie

Learning Outcomes Lecture 2

- Demonstrate understanding of how humans perceive the world around them on a general level and absorb complex data/information on a specific level.
- Analyse and evaluate how metaphors are used to convey unfamiliar information.
- Analyse and evaluate how mental models aid in the interpretation of complex visual displays.
- Select, formulate and integrate metaphors to suit data-driven tasks
- Design effective visualisations based on principles from perceptual psychology, cognitive science, graphic design and visual art.
- Demonstrate an understanding how visualisation is used in date journalism to communicate complex ideas and stories
- Demonstrate understanding how visualisation is used in story telling

Visualisation of the Week



https://public.tableau.com/en-us/s/gallery/game-thrones-season-6-review?gallery=featured

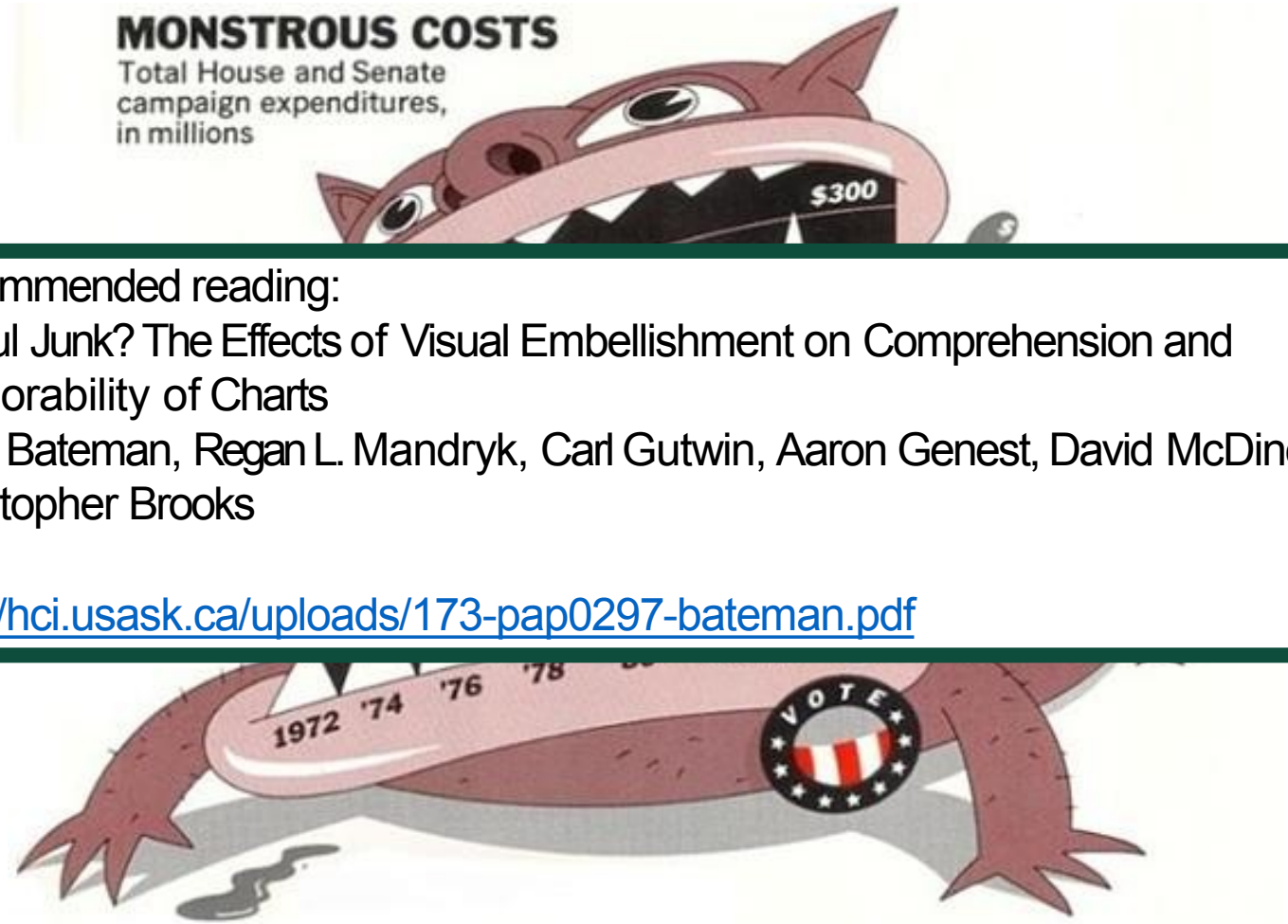
UnVisualisation of the Week

Tufte advocates a minimalist perspective

High data to ink ratio.

https://thedatalass.com/2017/10/17/data-visualization-design-part-3-chart-junk-in-data-science/

UnVisualisation of the Week



Recommended reading:
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

http://hci.usask.ca/uploads/173-pap0297-bateman.pdf

https://thedatalass.com/2017/10/17/data-visualization-design-part-3-chart-junk-in-data-science/

Overview

- Human Perception and memory
- How to categorise visualisations
- Class Exercises

Memory and Human Perception

Remember!

Well-designed visual representations can replace cognitive calculations with **simple perceptual inferences** and **improve comprehension**, **memory**, and **decision making**

By making data more **accessible** and **appealing**, visual representations may also help engage more diverse audiences in exploration and analysis

The challenge is to create **effective and engaging visualisations** that are **appropriate** to the data

Human Memory and Visualisation

- The power of visualisation is the ability to get across lot more complex information than what our memory can generally hold

- There are 3 types of memories that process information in our brain-

1. Iconic memory or Sensory memory
2. Working memory
3. Long term memory

Human Memory and Visualisation

Type of Memory	Iconic Memory	Working Memory	Long Term Memory
Duration	1 second approx.	1 minute approx.	1 second – lifetime
Processing	Pre-attentive processing, before we pay attention	Hold and process about 5-9 chunks of information (Miller's Law)	Information stored through rehersal or repeated application

- Long term memory is where things we memorize or remember are stored
- Iconic and working memories are the ones that interact with visualisations, so let's look at them in depth

Iconic Memory

- When we see a visual, the information remains in the iconic memory for less than a second. We process and store information automatically in this time
- This process is called **pre-attentive processing** and it happens automatically, before we pay attention to the information
- The pre-attentive process detects several visual attributes

Understanding how to make a particular attribute stand out can help us create visuals that emphasize the more important information

Working Memory

- The memory in use when working with a visual
  - Sensory information of interest to us is processed in the working memory
  - Information stays here for about a minute
  - Capacity of our working memory is between 5 to 9 similar items (Miller's Law)

- The capacity of our working memory can be increased by a process called **chunking**
  - Grouping similar items together

Working Memory

- Data visualisations take advantage of chunking
- When information is displayed in visuals that show meaningful patterns, more information can be chunked together
  - When we look at a visual, we can process much more information than when looking at the data in a table
- For a visualisation to be effective, we must not provide more data than what our brains can process
  - It is also important to display the visual on a single location, so we can see it without having to scroll or bounce back and forth between multiple locations

Human Perception for Visualisation?

- In order for our visualisations to be useful, it is important that they are easily interpreted by a viewer
- Bad data visualisations can be difficult to read and understand
- To visualize data effectively, it is useful to follow design principles
  - Derived from an understanding of human perception
- In order to create effective visualisations we must understand perception

What is Human Visual Perception?

- Vision is the first component of the human sensorial system and it is acquired fastest
- Our vision system has parallelism capacity
  - while it focuses on one point, the visual system can reach a wide surrounding area
- The human sensorial system is stimulated by a continuous flow of events
  - Resulting in a neural excitement called **sensation**
  - This process is extremely **fast and efficient**
- This continuous flow of sensations generates what it is known as perception
- Perceptions are different according to the characteristics of the stimulus
  - They are **interpreted**, allowing the brain to extract knowledge
  - The act of thinking, processing and making comparisons is known as **cognition** and is **much slower and less efficient**

Human Visual Perception and Data Visualisation

- Data visualisation is an external aid to support our working memory
- Visual perception is selective
  - If we tune our awareness to everything, we will be overwhelmed. So we selectively pay attention to things that catch our attention
- Our eyes are drawn to familiar patterns
  - We see what we expect to see. So visualisation must take into account what people know and expect
- Our working memory is very limited
  - We can hold a very limited amount of information in our memory when looking at a visual
- We want more visual perception and less cognition

Gestalt Theory

“It is not possible to have knowledge of the whole through the parts, but of the parts through the whole; the whole is more than the sum of its parts”

- Through the perception of the totality, the brain can realize, decode and assimilate an image or a concept

- It proposes a number of laws that which explain the structural and functional principles of the perceptive field:

Some Laws of Gestalt Theory

- Proximity: Objects that are close together are perceived as a group
- Similarity: Objects that share similar attributes (e.g., colour or shape) are perceived as a group
- Enclosure: Objects that appear to have a boundary around them (e.g., formed by a line or area of common colour) are perceived as a group
- Closure: Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such
- Continuity: Objects that are aligned together or appear to be a continuation of one another are perceived as a group
- Connection: Objects that are connected (e.g., by a line) are perceived as a group

Categorisation of Visualisations

Short Glossary of Terms

- **Chart**: Something that shows qualitative information (e.g., flow charts)
- **Data Dimension**: One single channel of data (e.g. a stock graph may comprise dates and prices - each is a unique dimension)
- **Designer**: The creator of a visualisation

"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

Short Glossary of Terms

- **Encoding**: The visual property applied to a dimension of data that encodes the information into a visual medium for decoding by the reader

- **Graph**: Something that shows quantitative information (e.g., pie graphs and bar graphs)

- **Infographic**: visualisations that are manually generated around specific data, tend to be data-shallow, and are often aesthetically rich

"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

Short Glossary of Terms

- **Reader**: The consumer of a visualisation, often someone other than the designer. The reader has information needs that are meant to be satisfied by the visualisation

- **Visual Property**: A characteristic that you can see (e.g. colour, size, location, thickness, and line weight)

- **Variability of a data dimension**: The values that are present or allowed in a single data dimension (e.g. integers vary discretely; position can vary continuously; categories are finite (and discrete, though may be hierarchical); numbers are infinite)

"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

Remember!

Well-designed visual representations can replace cognitive calculations with **simple perceptual inferences** and **improve comprehension**, **memory**, and **decision making**

By making data more **accessible** and **appealing**, visual representations may also help engage more diverse audiences in exploration and analysis

The challenge is to create **effective and engaging visualisations** that are **appropriate** to the data

Classifications of visualisations

1. Complexity
2. Infographics vs Data visualisation
3. Exploration vs Explanation
4. Informative vs Persuasive vs Visual Art

"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

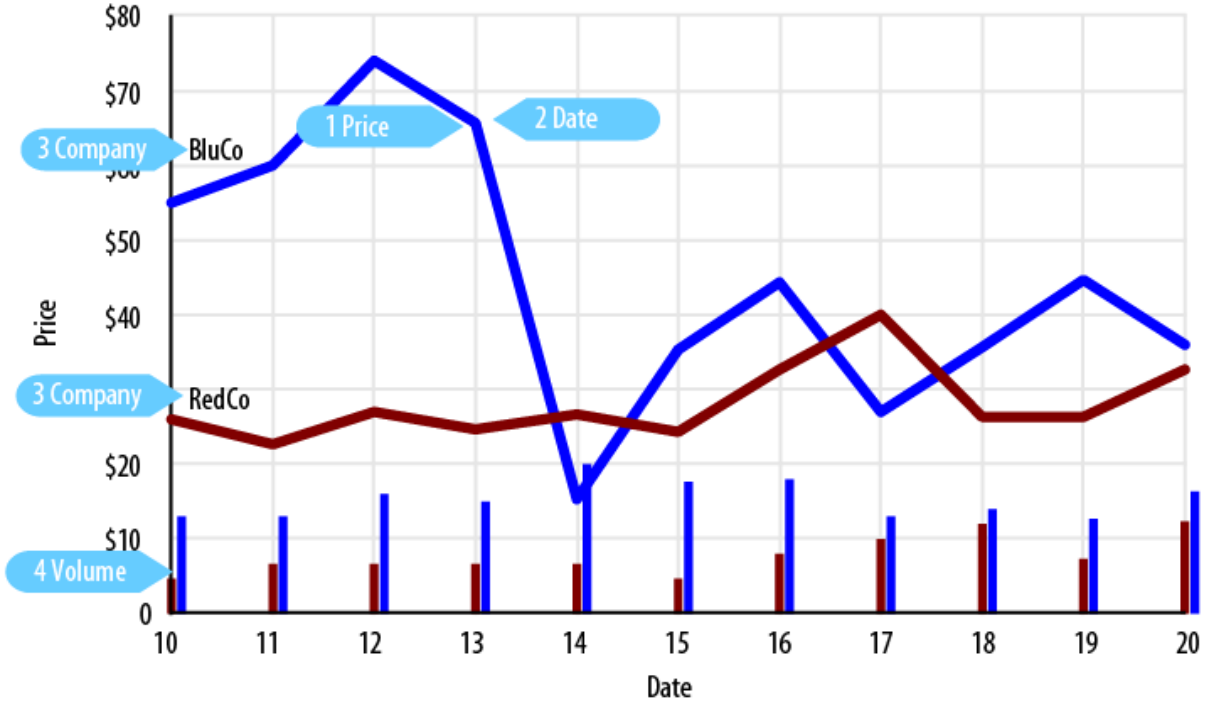
Classifications of visualisations

1. Complexity
2. Infographics vs Data visualisation
3. Exploration vs Explanation
4. Informative vs Persuasive vs Visual Art

“Designing Data visualisations”, N. Iliinsky & J. Steele, O'Reilly Media, 2011

Complexity

- Classify by counting how many different *data dimensions* are represented
- The number of discrete types of information that are visually encoded in a diagram

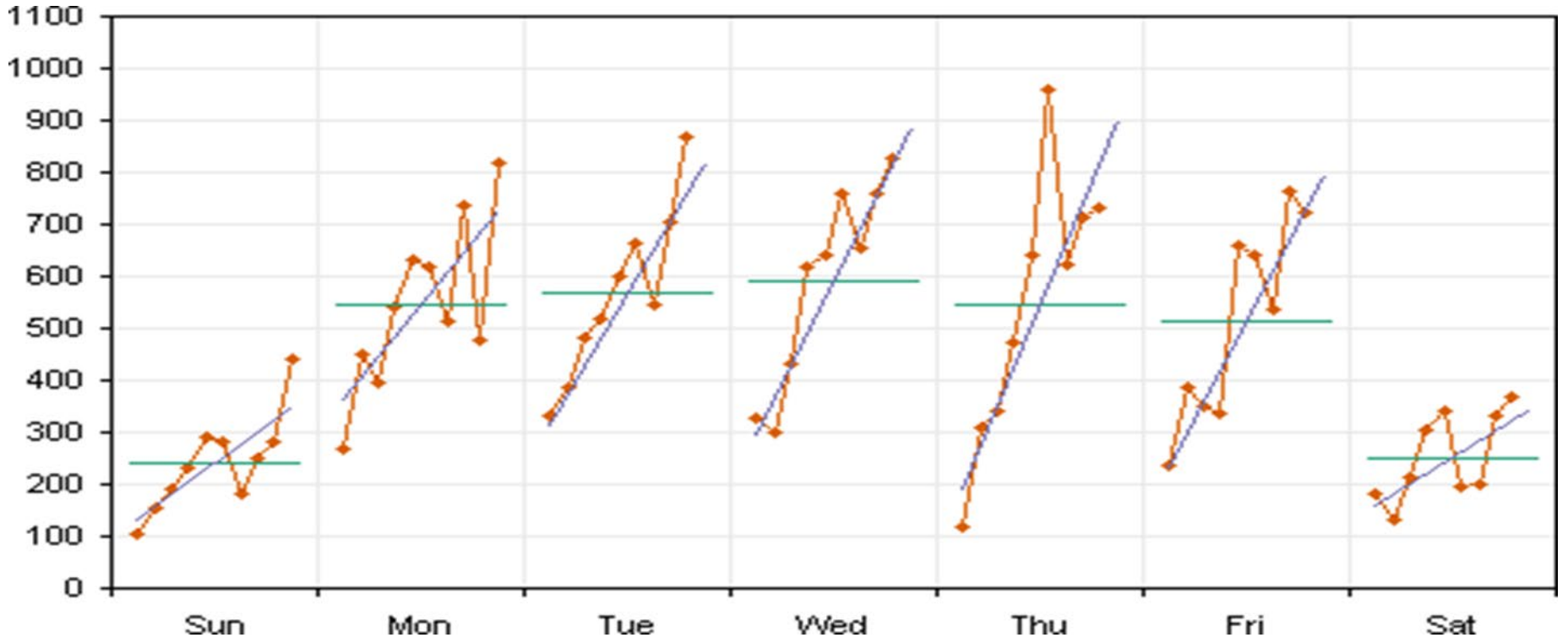


Complexity

- For example, a simple line graph may show the price of a financial instrument on different days: 2 data dimensions

Yahoo Finance: http://finance.yahoo.com/q/bc?s=%5EFTSE

Complexity



Cycle Plot – Days of the Week mapped over 10 weeks

Complexity

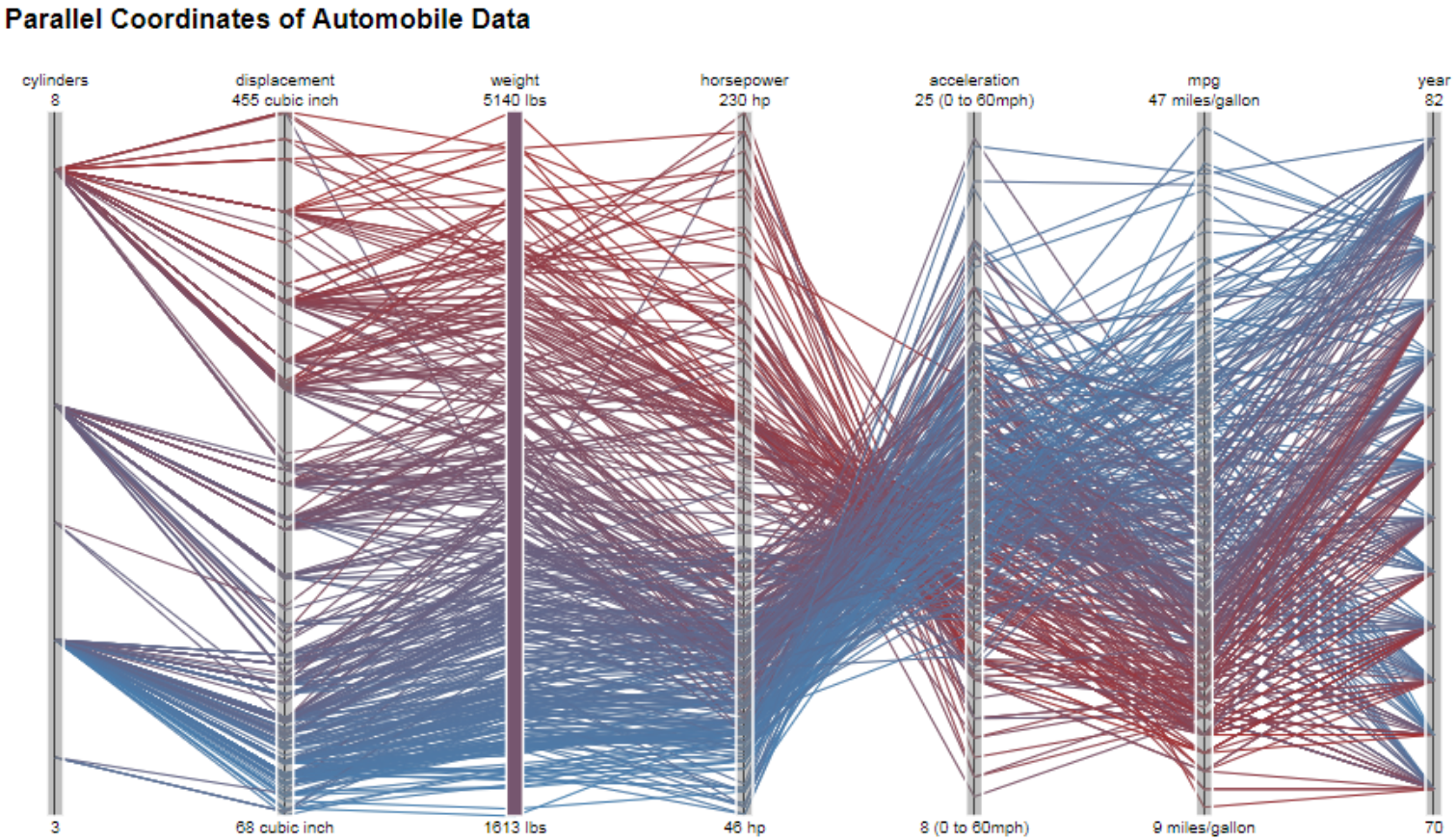
- As visualisations become more complex:
  - They are more challenging to design well
  - Can be more difficult to learn from

- For that reason, visualisations with no more than 3 or 4 dimensions of data are the most common
  - Visualisations with six, seven, or more dimensions can be found but may be more difficult to read!

Complexity

- 2 main challenges to designing more complex visualisations:
  - The more dimensions you need to encode visually, the more individual visual properties you need to use
  - There are relatively few well-known conventions, metaphors, defaults, and best practices to rely on
  - There is more of a burden on the designer to make good choices that can be easily understood by the reader
  - However, there are smart ways to think about and represent the data

Complexity



https://homes.cs.washington.edu/~jheer/files/zoo/ex/stats/parallel.html

Classifications of visualisations

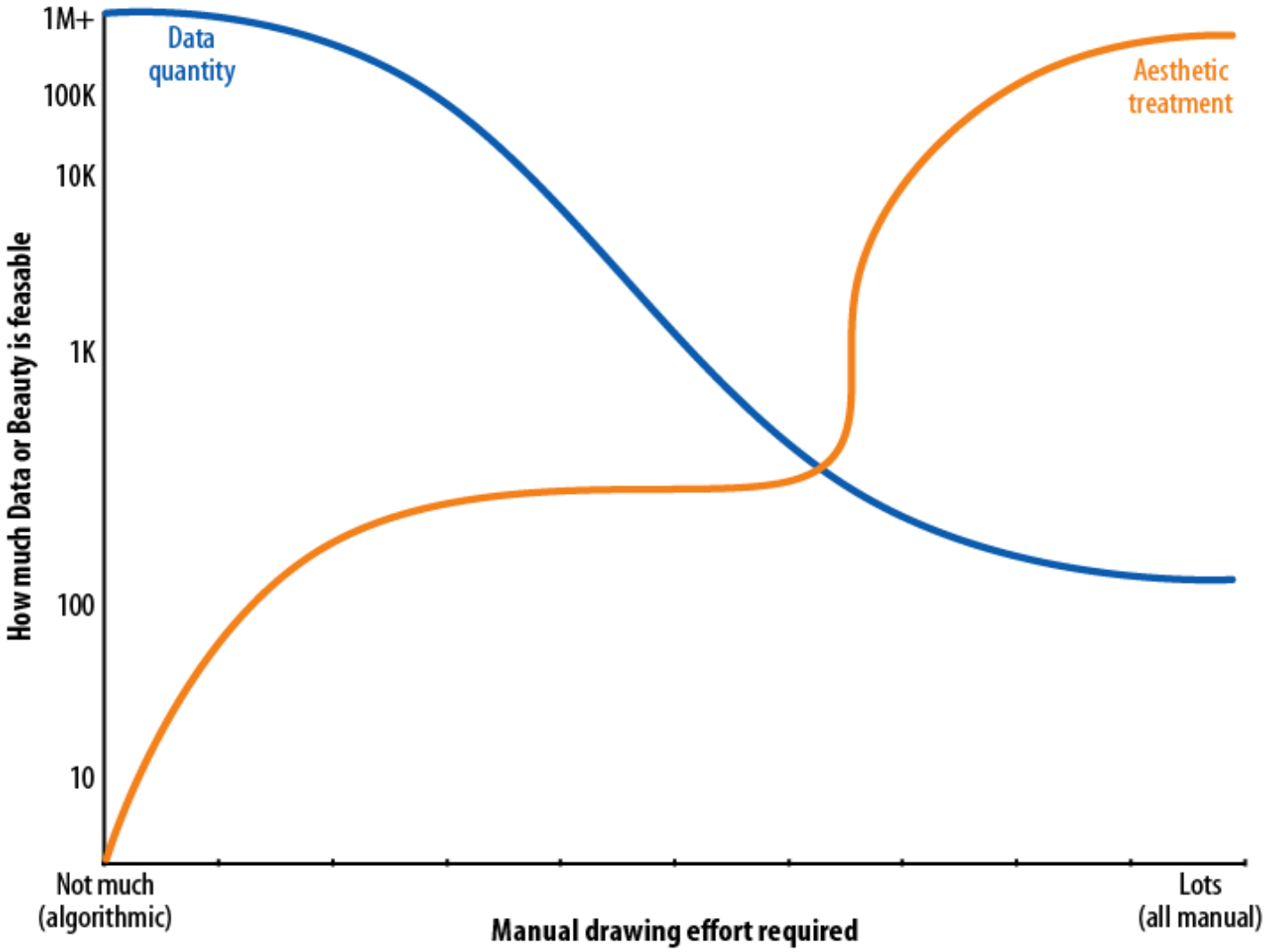
1. Complexity
2. Infographics vs Data visualisation
3. Exploration vs Explanation
4. Informative vs Persuasive vs Visual Art

“Designing Data visualisations”, N. Iliinsky & J. Steele, O'Reilly Media, 2011

Infographics vs Data visualisation

- **Infographics** and **data visualisation** are slightly nebulous terms often used interchangeably, but a distinction is worthwhile
  - **Infographic** used to refer to representations of information perceived as casual, funny, or frivolous
  - **Data visualisation** used to refer to designs perceived to be more serious, rigorous, academic or scientific!

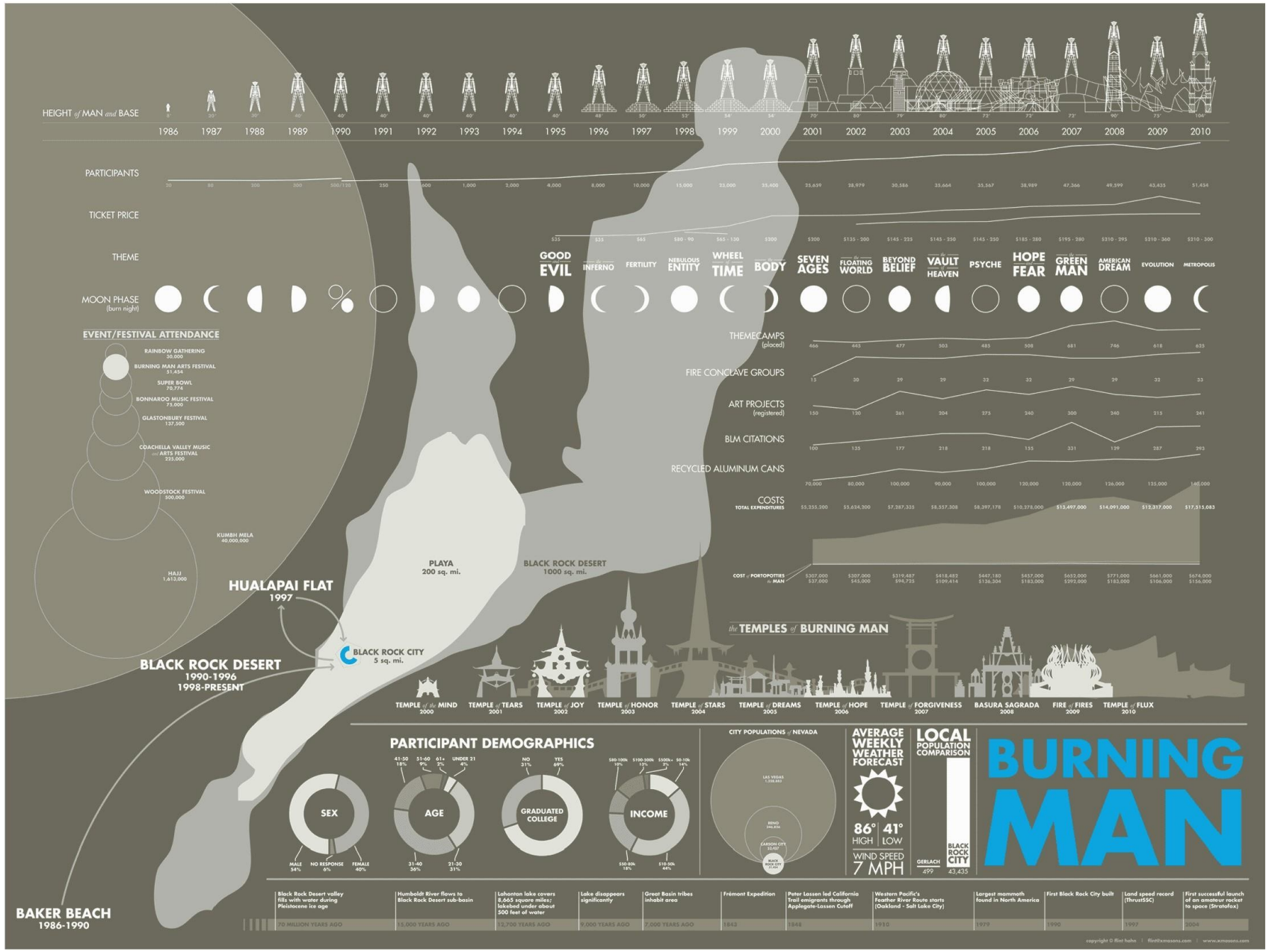
Infographics vs Data visualisation



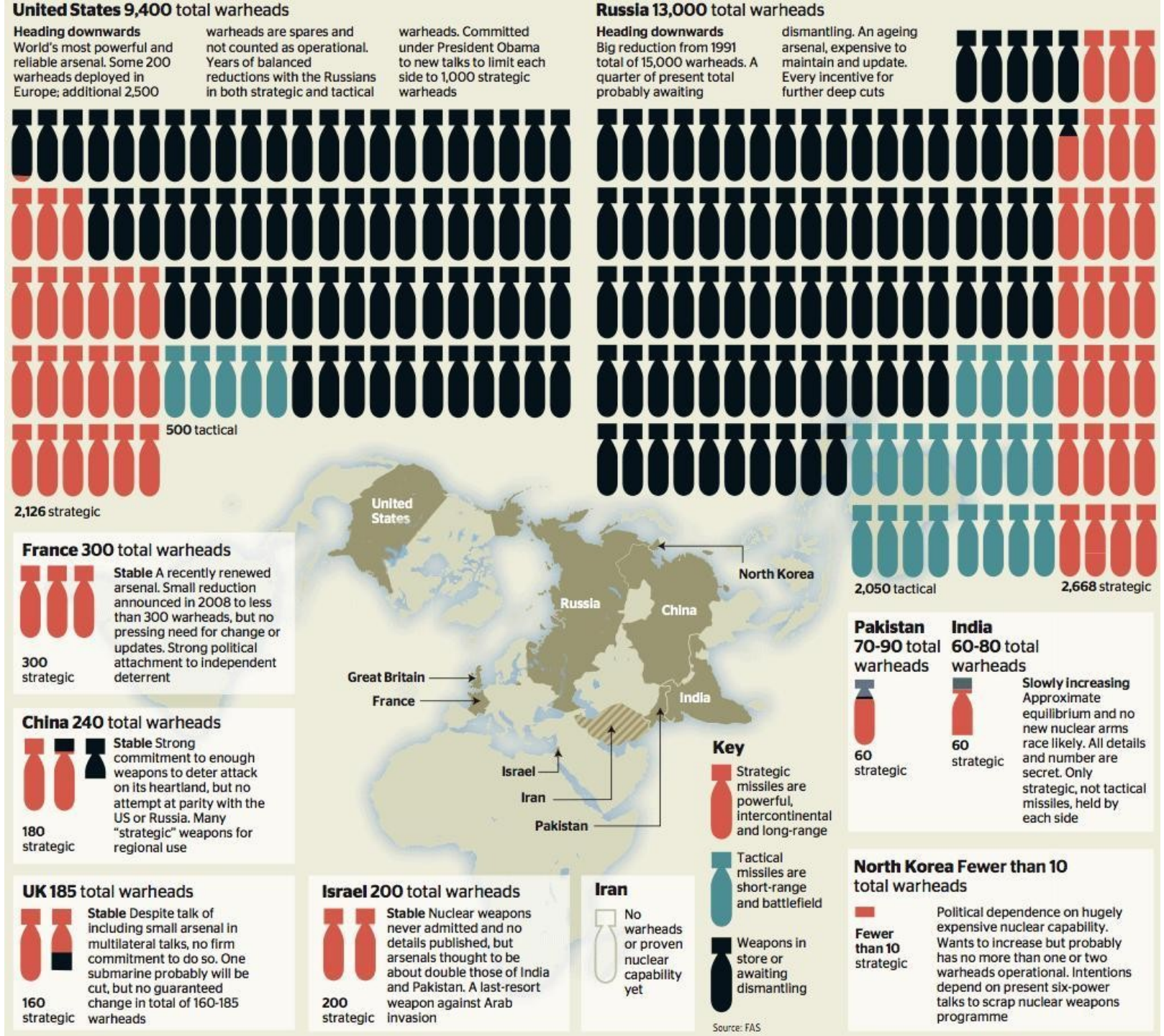
Infographics

- Iliinsky & Steele suggest that the term infographics is useful for referring to visual representation of data that is:
  - Manually drawn (and therefore a custom treatment of the information)
  - Specific to the data at hand (and therefore non- trivial to recreate with different data)
  - Aesthetically rich (strong visual content meant to draw the eye and hold interest)
  - Relatively data—poor (because each piece of information must be manually encoded)

"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011



Burning Man, Flint Hahn (2010)

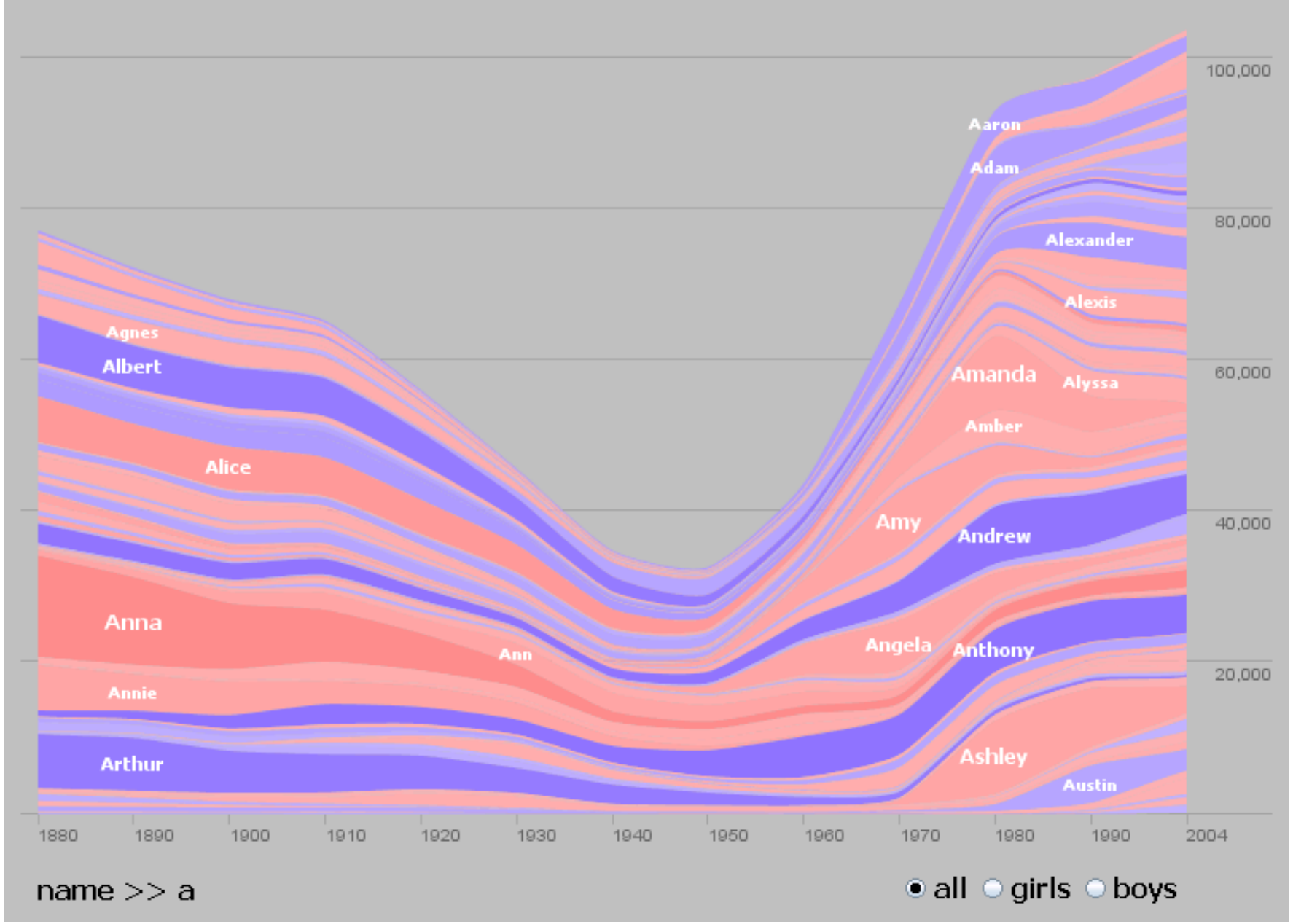

United States 9,400 total warheads
Heading downwards
World's most powerful and reliable arsenal. Some 200 warheads deployed in Europe; additional 2,500 warheads are spares and not counted as operational. Years of balanced reductions with the Russians in both strategic and tactical warheads. Committed under President Obama to new talks to limit each side to 1,000 strategic warheads
500 tactical
2,126 strategic
Russia 13,000 total warheads
Heading downwards
Big reduction from 1991 total of 15,000 warheads. A quarter of present total probably awaiting dismantling. An ageing arsenal, expensive to maintain and update. Every incentive for further deep cuts
2,050 tactical
2,668 strategic
United States
Russia
China
North Korea
Great Britain
France
India
Israel
Iran
Pakistan
France 300 total warheads
300 strategic
Stable A recently renewed arsenal. Small reduction announced in 2008 to less than 300 warheads, but no pressing need for change or updates. Strong political attachment to independent deterrent
China 240 total warheads
180 strategic
Stable Strong commitment to enough weapons to deter attack on its heartland, but no attempt at parity with the US or Russia. Many "strategic" weapons for regional use
UK 185 total warheads
160 strategic
Stable Despite talk of including small arsenal in multilateral talks, no firm commitment to do so. One submarine probably will be cut, but no guaranteed change in total of 160-185 warheads
Israel 200 total warheads
200 strategic
Stable Nuclear weapons never admitted and no details published, but arsenals thought to be about double those of India and Pakistan. A last-resort weapon against Arab invasion
Iran
No warheads or proven nuclear capability yet
Key
Strategic missiles are powerful, intercontinental and long-range
Tactical missiles are short-range and battlefield
Weapons in store or awaiting dismantling
Source: FAS
Pakistan 70-90 total warheads
60 strategic
India 60-80 total warheads
60 strategic
Slowly increasing
Approximate equilibrium and no new nuclear arms race likely. All details and number are secret. Only strategic, not tactical missiles, held by each side
North Korea Fewer than 10 total warheads
Fewer than 10 strategic
Political dependence on hugely expensive nuclear capability. Wants to increase but probably has no more than one or two warheads operational. Intentions depend on present six-power talks to scrap nuclear weapons programme

GetBulb


47%
47%
51%
66%
68%
6%
Kids 2-11
Teens 12-17
Adults 18-34
Adults 35-54
Adults 55+
Increase from 2012
Ammonium
Beryllium
Calcium
Propargite
Sodium
2012
2011
2010
women
men
high street
59%
28%
shopping centres
42%
18%
advertising
51%
49%
social media
59%
1%

Data visualisation

- By contrast, the term data visualisation is useful for referring to any visual representation of data that is:
  - Algorithmically drawn: may have custom touches but is largely rendered with the help of computerized methods
  - Easy to regenerate with different data: the same form may be repurposed to represent different datasets with similar dimensions or characteristics
  - Maybe aesthetically barren: data is not decorated
  - Relatively data-rich: large volumes of data are welcome and viable, in contrast to infographics



http://www.babynamewizard.com/voyager#prefix=&sw=both&exact=false

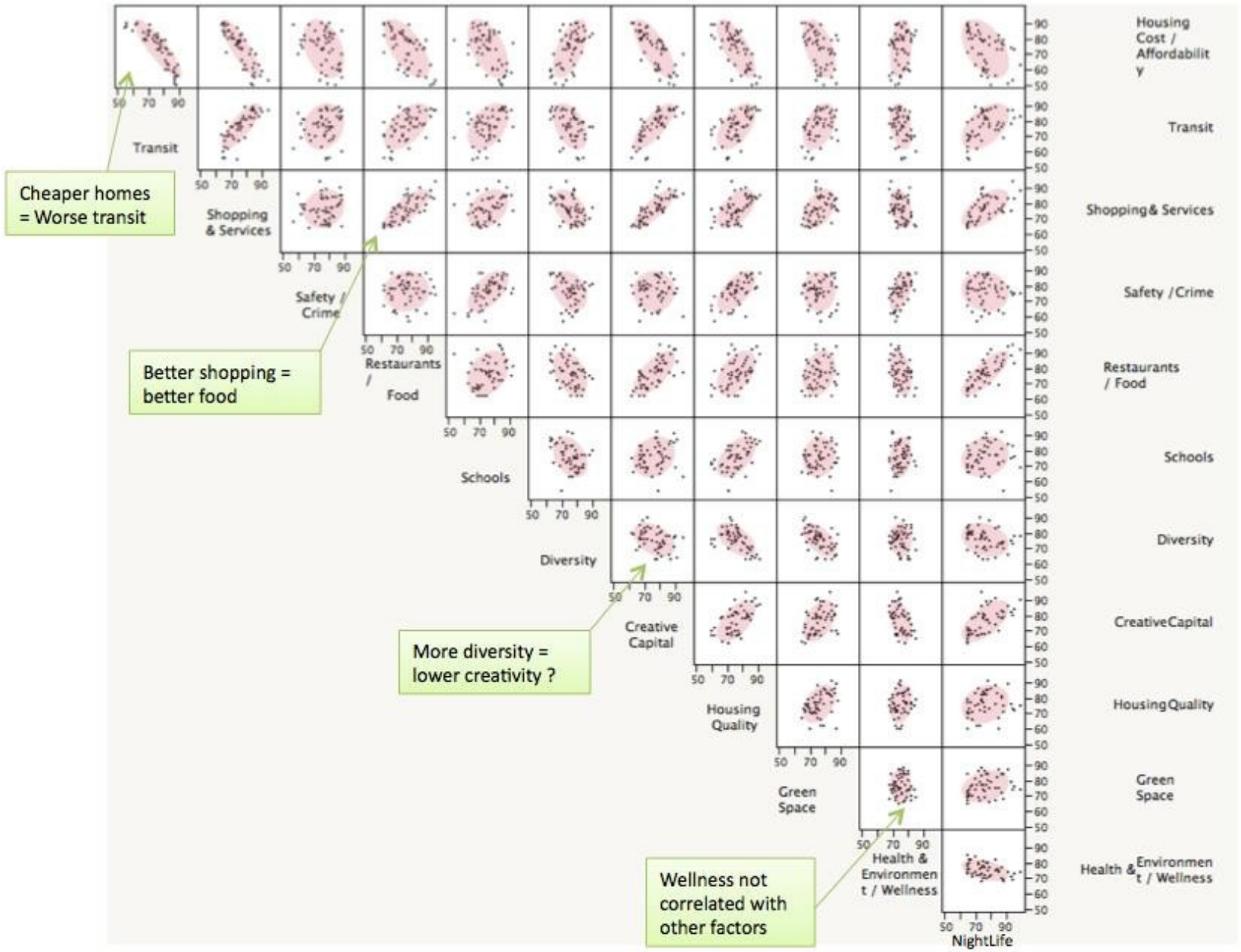
Classifications of visualisations

1. Complexity
2. Infographics vs Data visualisation
3. Exploration vs Explanation
4. Informative vs Persuasive vs Visual Art

“Designing Data visualisations”, N. Iliinsky & J. Steele, O'Reilly Media, 2011

Exploration

- Exploratory data visualisations are appropriate when you have lots of data and you are not sure what's in it
  - When you need to get a sense of what's inside your data set, translating it into a visual medium can help you quickly identify its features, including interesting curves, lines, trends, or anomalous outliers
- This type of visualisation is typically part of the data analysis phase, and is used to find the story the data has to tell you

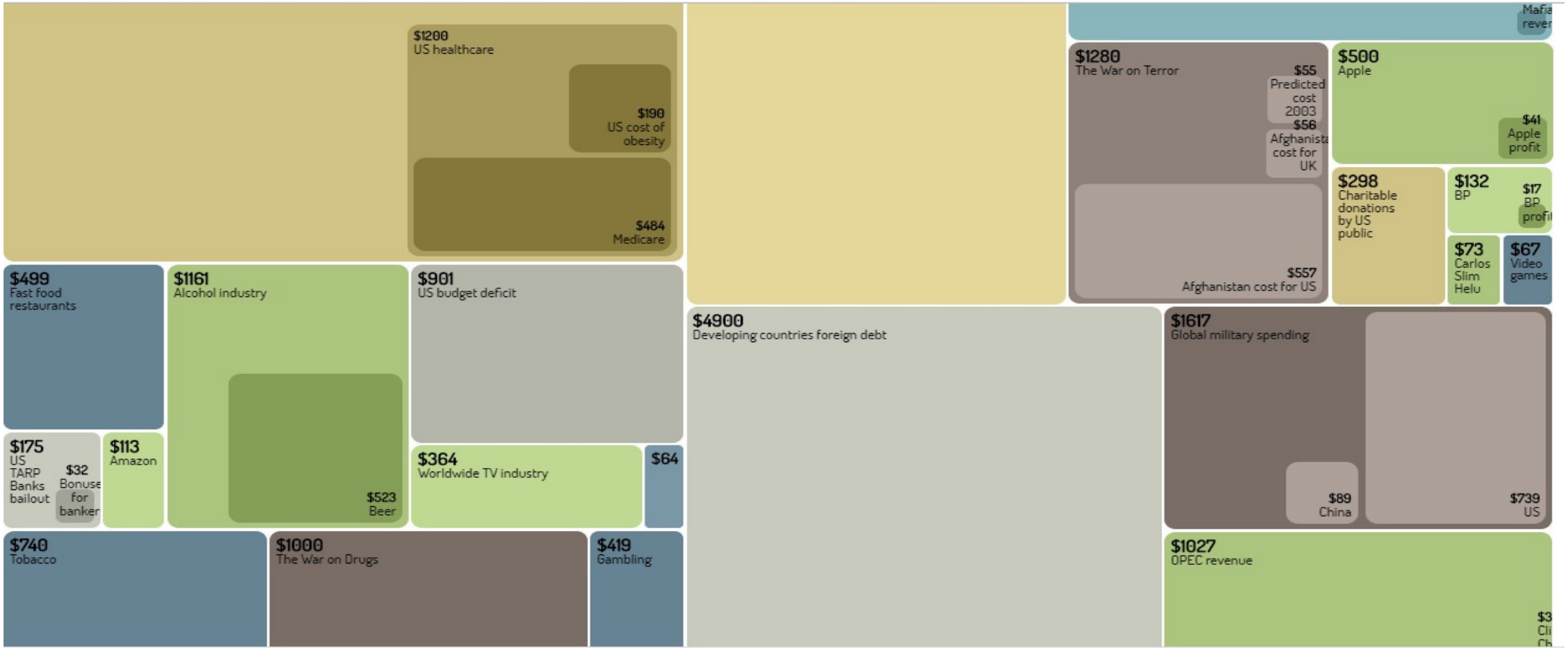


http://junkcharts.typepad.com/junk_charts/2010/06/the-scatterplot-matrix-a-great-tool.html

Explanation

- Explanatory data visualisation is appropriate when you already know what the data has to say, and you are trying to tell that story to somebody else
  - Whoever your audience is, the story you are trying to tell is known to you at the outset, and therefore you can design to specifically accommodate and highlight that story
- Explanatory data visualisation is essentially part of a presentation phase - classic slideware!

Billion-dollar-o-gram

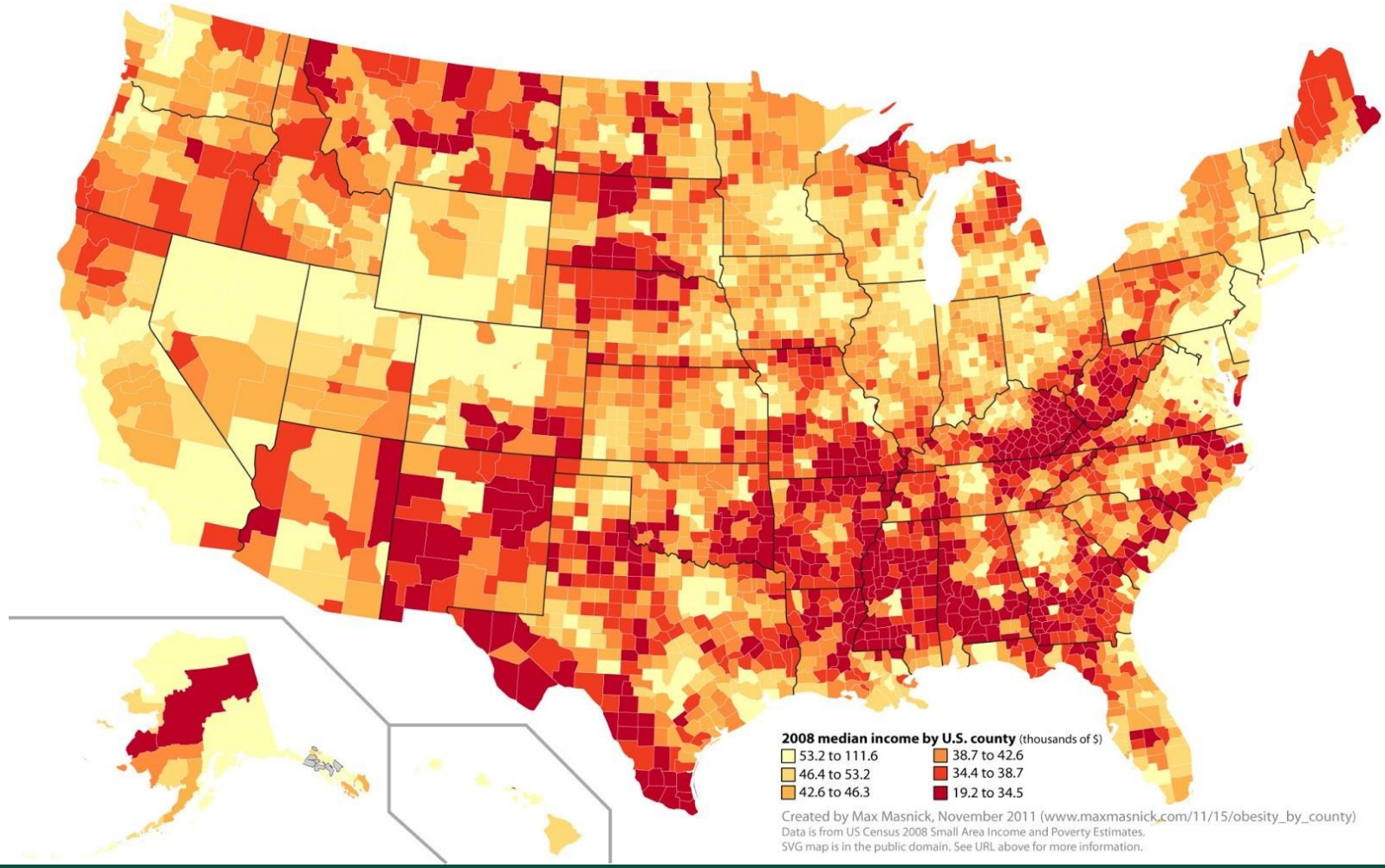


http://www.informationisbeautiful.net/visualisations/billion-dollar-o-gram-2013/

Hybrids: Exploratory & Explanatory

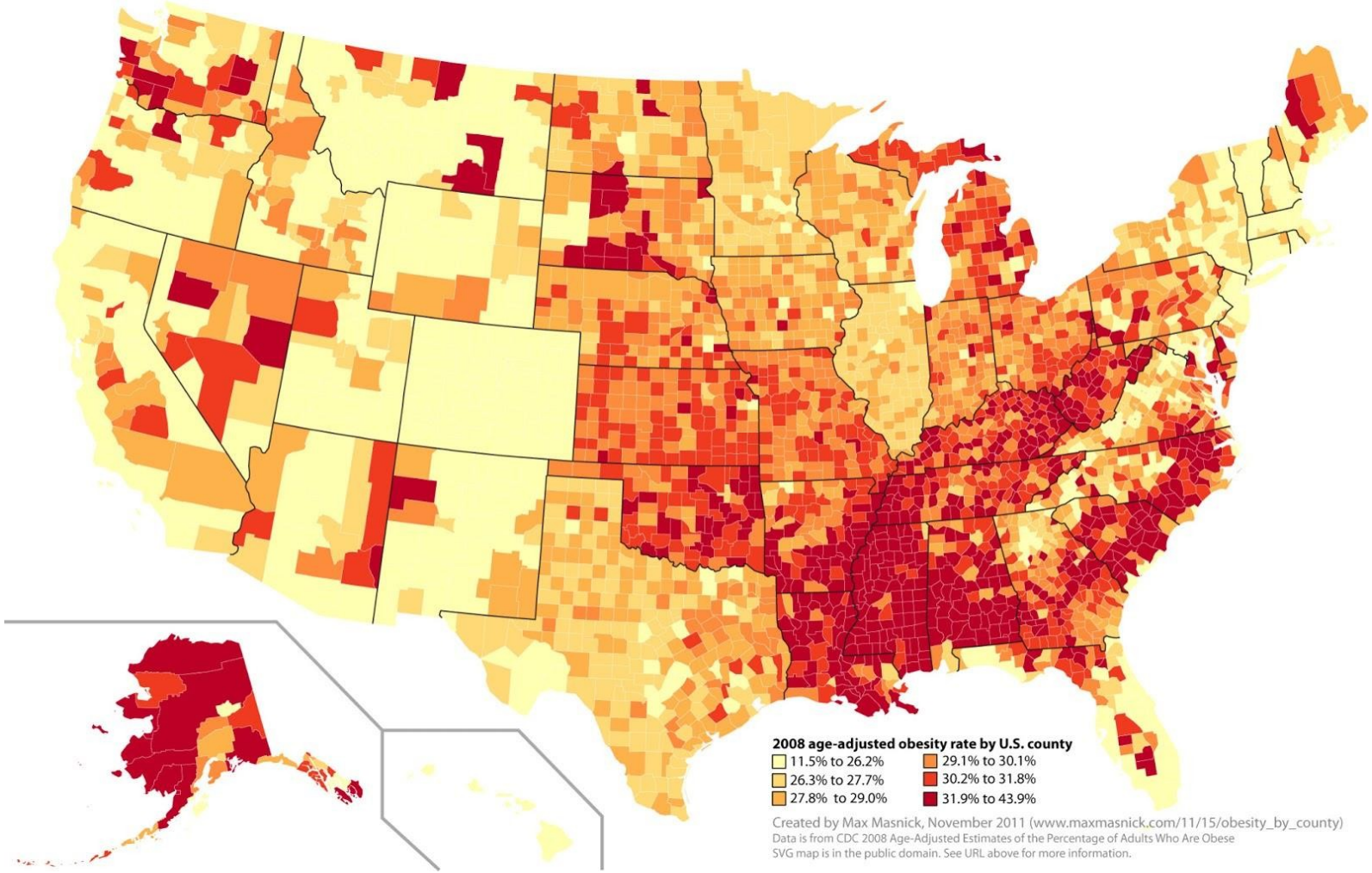
- Hybrid category: a curated dataset that is nonetheless presented with the intention to allow some exploration on the reader's part
  - These visualisations are usually interactive via some kind of graphical interface that lets the reader choose and constrain certain parameters
  - Sometimes allows the reader discover insights the designer of the visualisation hasn't come across yet!
- In these hybrid designs there is a certain freedom- of-discovery aspect to the information presented, but it is usually not totally raw; it has been distilled and facilitated in some way

Income Map USA



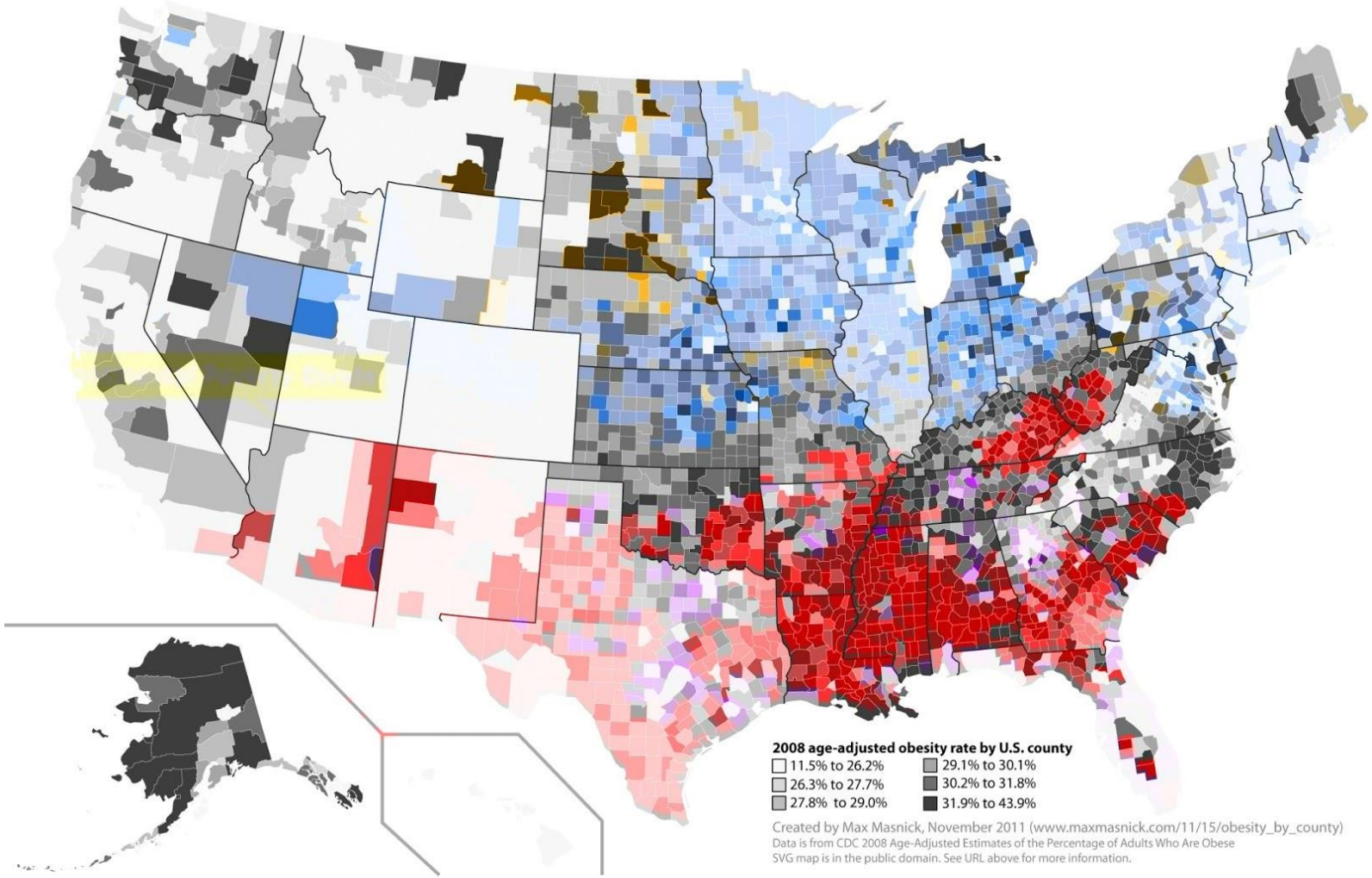
http://asserttrue.blogspot.ie/2013/02/poverty-and-obesity-in-america-how-they.html

Obesity Map USA



http://asserttrue.blogspot.ie/2013/02/poverty-and-obesity-in-america-how-they.html

Correlated Map



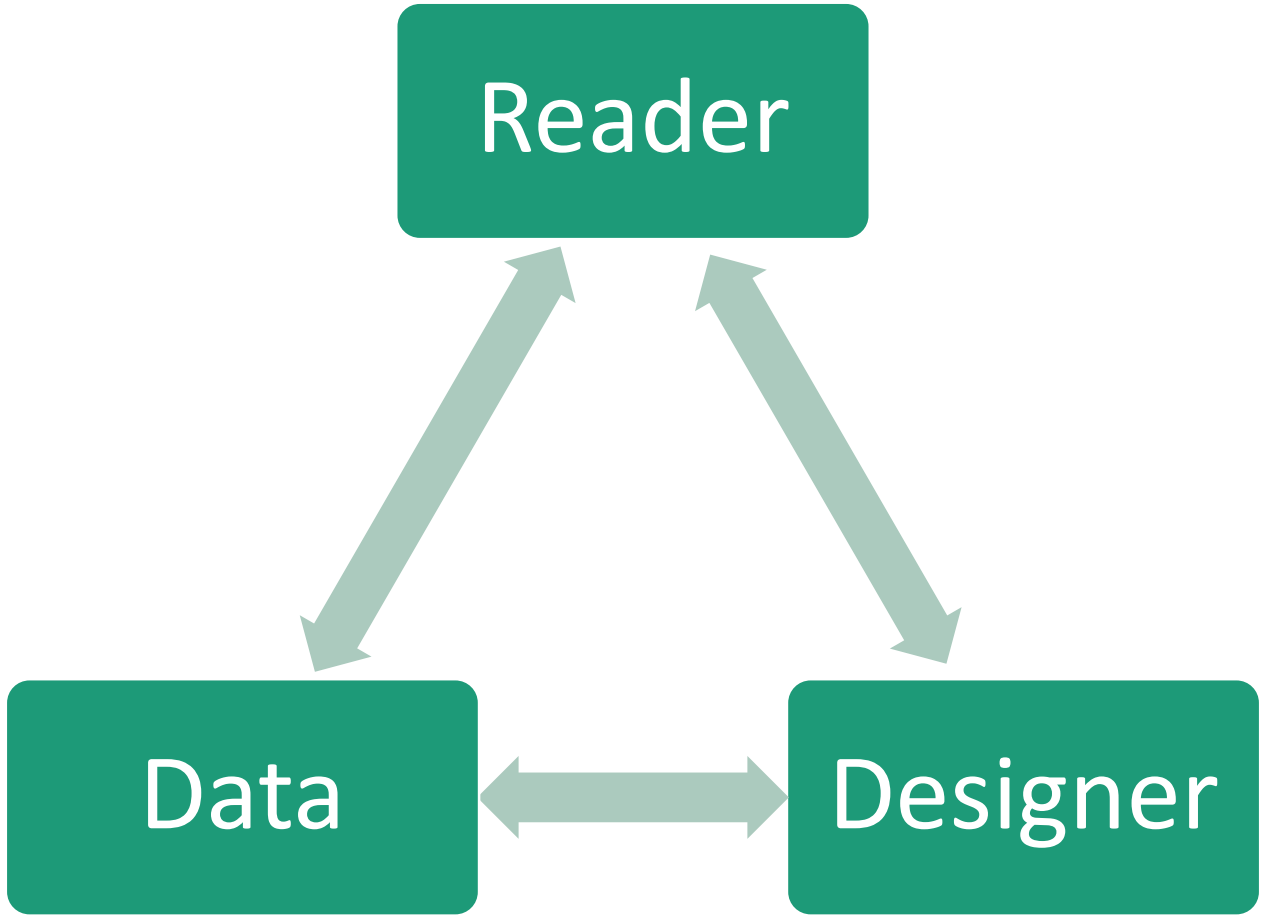
http://asserttrue.blogspot.ie/2013/02/poverty-and-obesity-in-america-how-they.html

Classifications of visualisations

1. Complexity
2. Infographics vs Data visualisation
3. Exploration vs Explanation
4. Informative vs Persuasive vs Visual Art

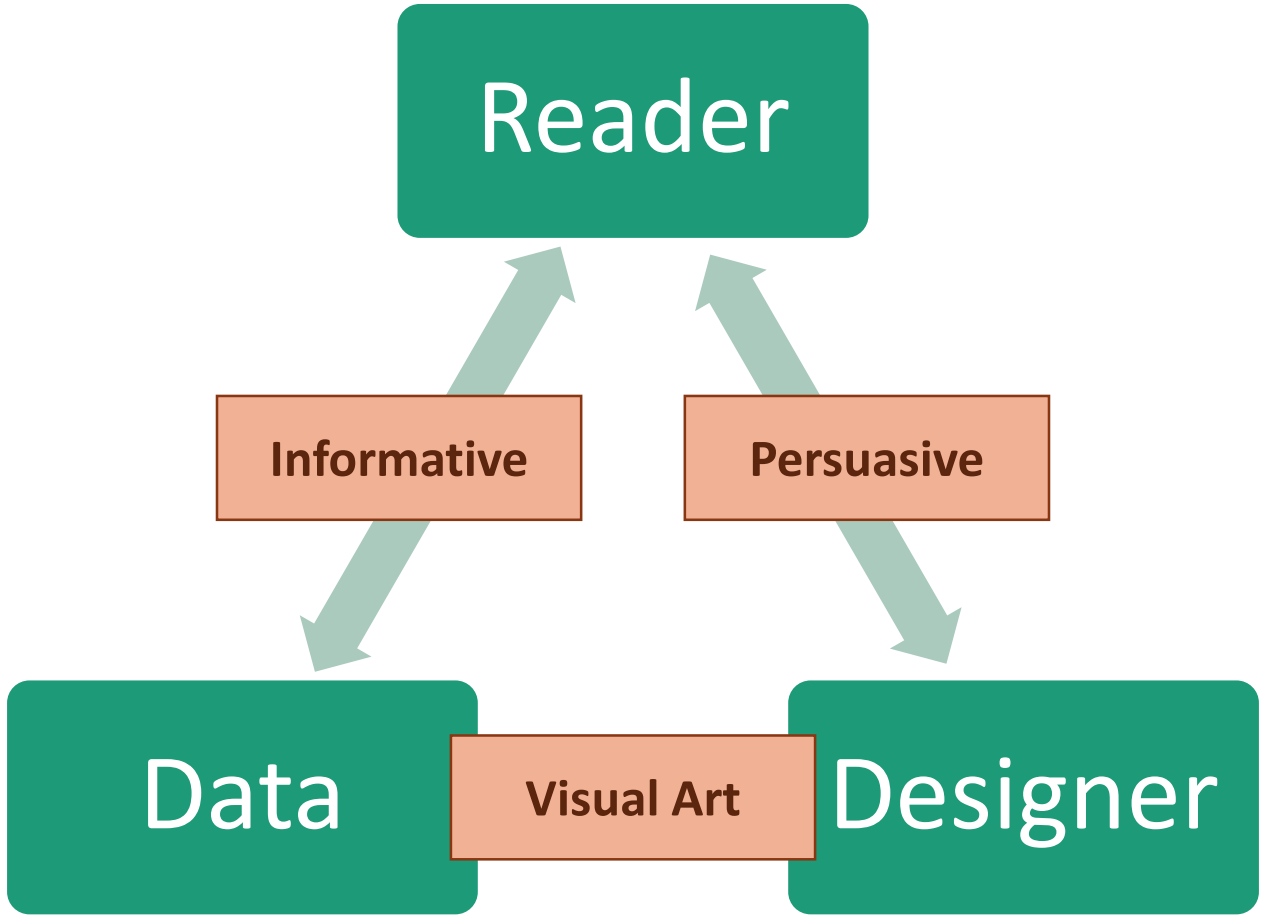
"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

The Designer-Reader-Data Trinity



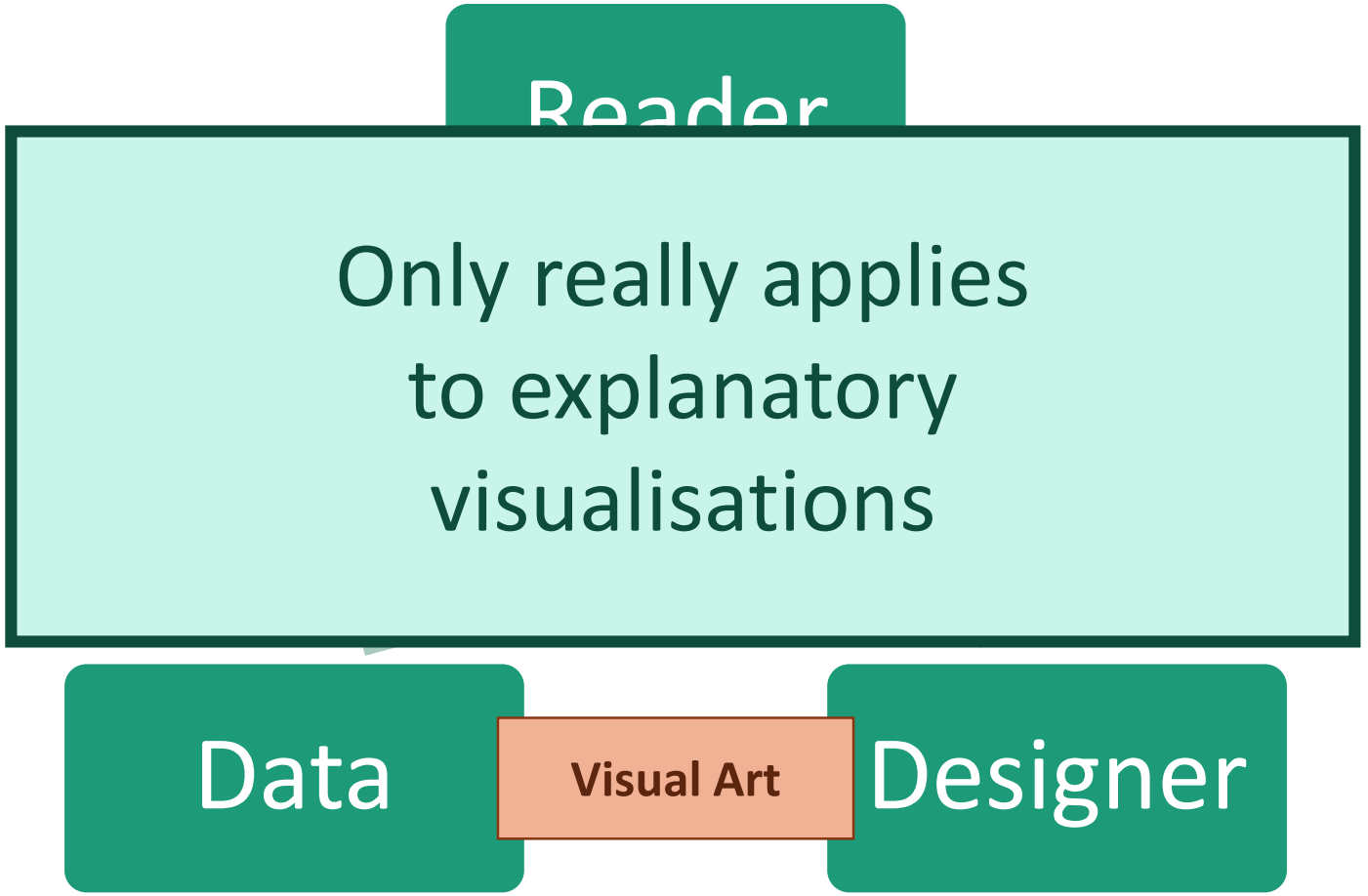
"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

The Designer-Reader-Data Trinity



"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

The Designer-Reader-Data Trinity



"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

The Designer-Reader-Data Trinity

- We can think of explanatory data visualisation as being supported by a three-legged stool consisting of the designer, the reader, and the data
- Each of these "legs" exerts a force, or contributes a separate perspective, that must be taken into consideration for a visualisation to be stable and successful
- Each of the three legs of the stool has a unique relationship to the other two
- While it is necessary to account for the needs and perspective of all three in each visualisation project, the dominant relationship will ultimately determine which category of visualisation is needed

Informative

- **An informative visualisation primarily serves the relationship between the reader and the data.**

- It aims for a **neutral presentation** of the facts in such a way that will educate the reader (though not necessarily persuade him).

- Informative visualisations are often associated with broad data sets, and seek to distil the content into a manageably consumable form.

Google’s Share Price

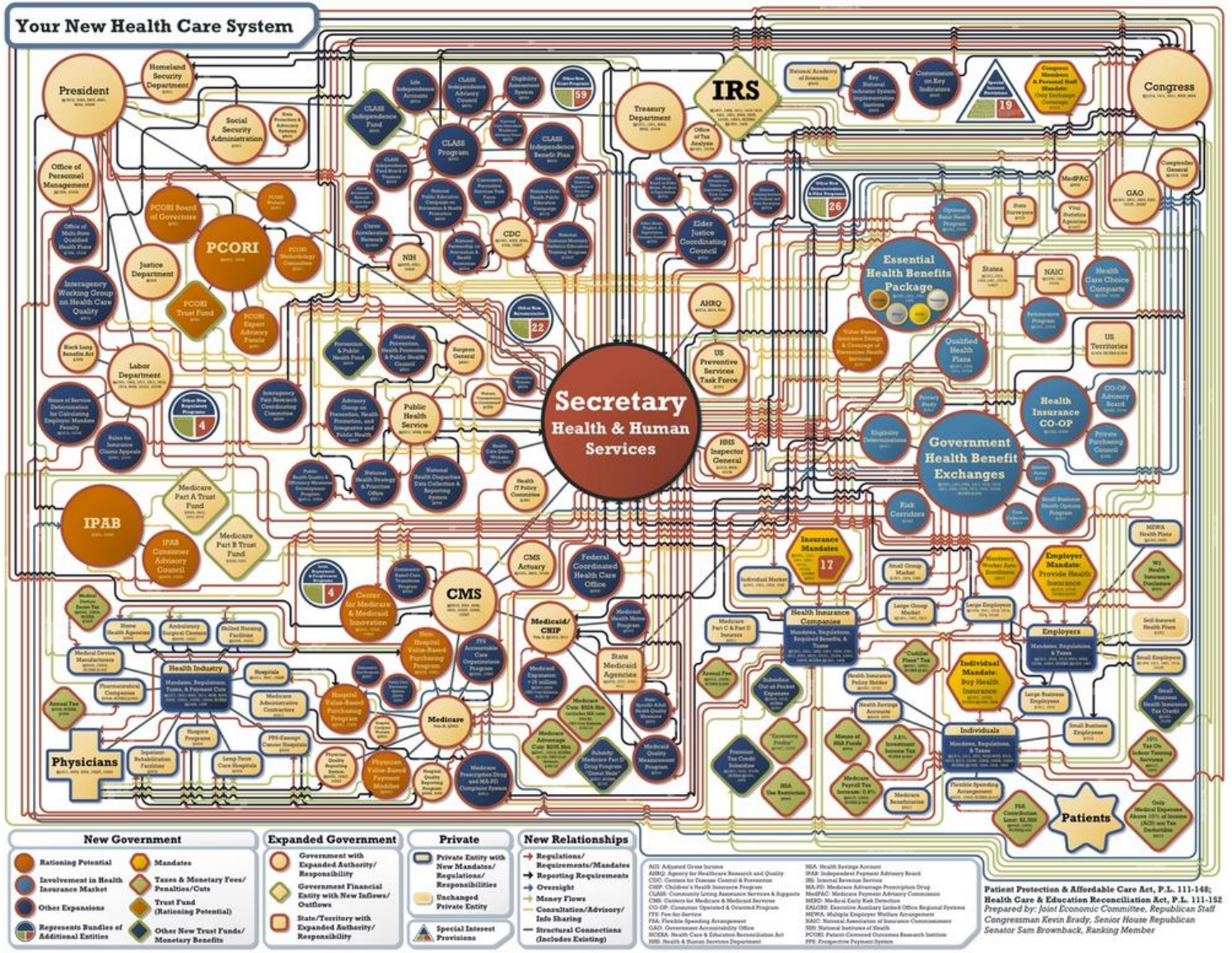
Persuasive

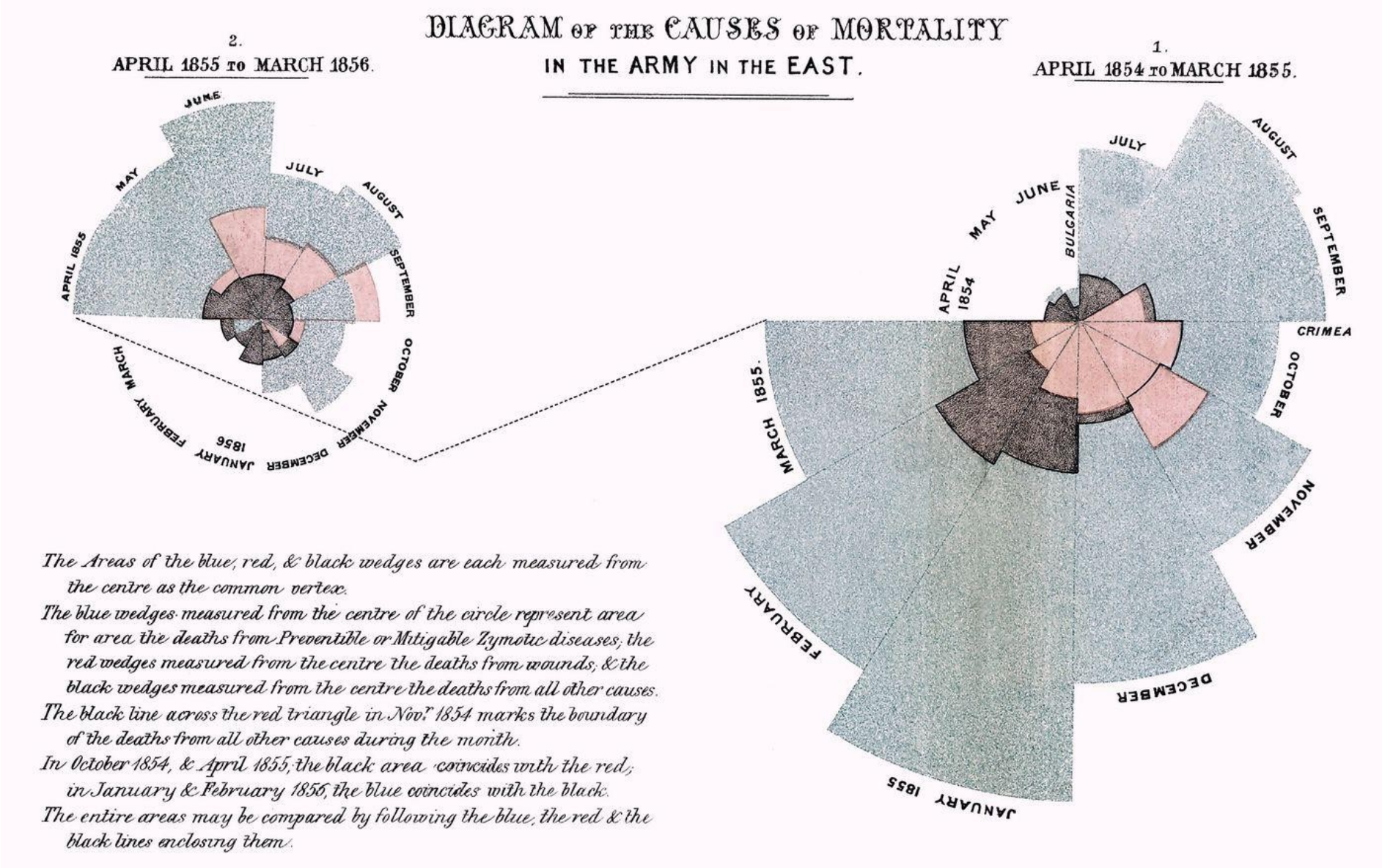
- A persuasive visualisation primarily serves the relationship between the designer and the reader.

- **It is useful when the designer wishes to change the reader's mind about something.**

- In this category of visualisation, the data represented is specifically chosen for the purpose of supporting the designer's point of view, and is presented carefully so as to convince the reader of same.

Persuasive


Your New Health Care System
President
Homeland Security Department
Social Security Administration
Office of Personnel Management
PCORI
PCORI Board of Governors
PCORI Trust Fund
PCORI Expert Advisory Panels
Justice Department
Interagency Working Group on Health Care Quality
Labor Department
Secretary
Health & Human Services
IRS
Treasury Department
Congress
CLASS Program
CLASS Independence Benefit Plan
CLASS Independence Fund
CDC
NIH
Elder Justice Coordinating Council
AHRQ
US Preventive Services Task Force
Public Health Service
HHS Inspector General
Essential Health Benefits Package
States
NAIC
Health Care Choice Compacts
US Territories
Qualified Health Plans
Government Health Benefit Exchanges
Health Insurance CO-OP
Risk Corridors
IPAB
IPAB Consumer Advisory Council
Medicare Part A Trust Fund
Medicare Part B Trust Fund
CMS
CMS Actuary
Center for Medicare & Medicaid Innovation
Federal Coordinated Health Care Office
Medicaid/ CHIP
State Medicaid Agencies
Medicare
Physicians
Health Industry
Insurance Companies
Insurance Mandates
Employer Mandate: Provide Health Insurance
Employers
Individual Mandate: Buy Health Insurance
Individuals
Patients
MedPAC
GAO
New Government
Rationing Potential
Involvement in Health Insurance Market
Other Expansions
Represents Bundles of Additional Entities
Mandates
Taxes & Monetary Fees/ Penalties/Cuts
Trust Fund (Rationing Potential)
Other New Trust Funds/ Monetary Benefits
Expanded Government
Government with Expanded Authority/ Responsibility
Government Financial Entity with New Inflows/ Outflows
State/Territory with Expanded Authority/ Responsibility
Private
Private Entity with New Mandates/ Regulations/ Responsibilities
Unchanged Private Entity
Special Interest Provisions
New Relationships
Regulations/ Requirements/Mandates
Reporting Requirements
Oversight
Money Flows
Consultation/Advisory/ Info Sharing
Structural Connections (Includes Existing)
Patient Protection & Affordable Care Act, P.L. 111-148;
Health Care & Education Reconciliation Act, P.L. 111-152
Prepared by: Joint Economic Committee, Republican Staff
Congressman Kevin Brady, Senior House Republican
Senator Sam Brownback, Ranking Member



http://thisisstatistics.org/florence-nightingale-the-lady-with-the-data/

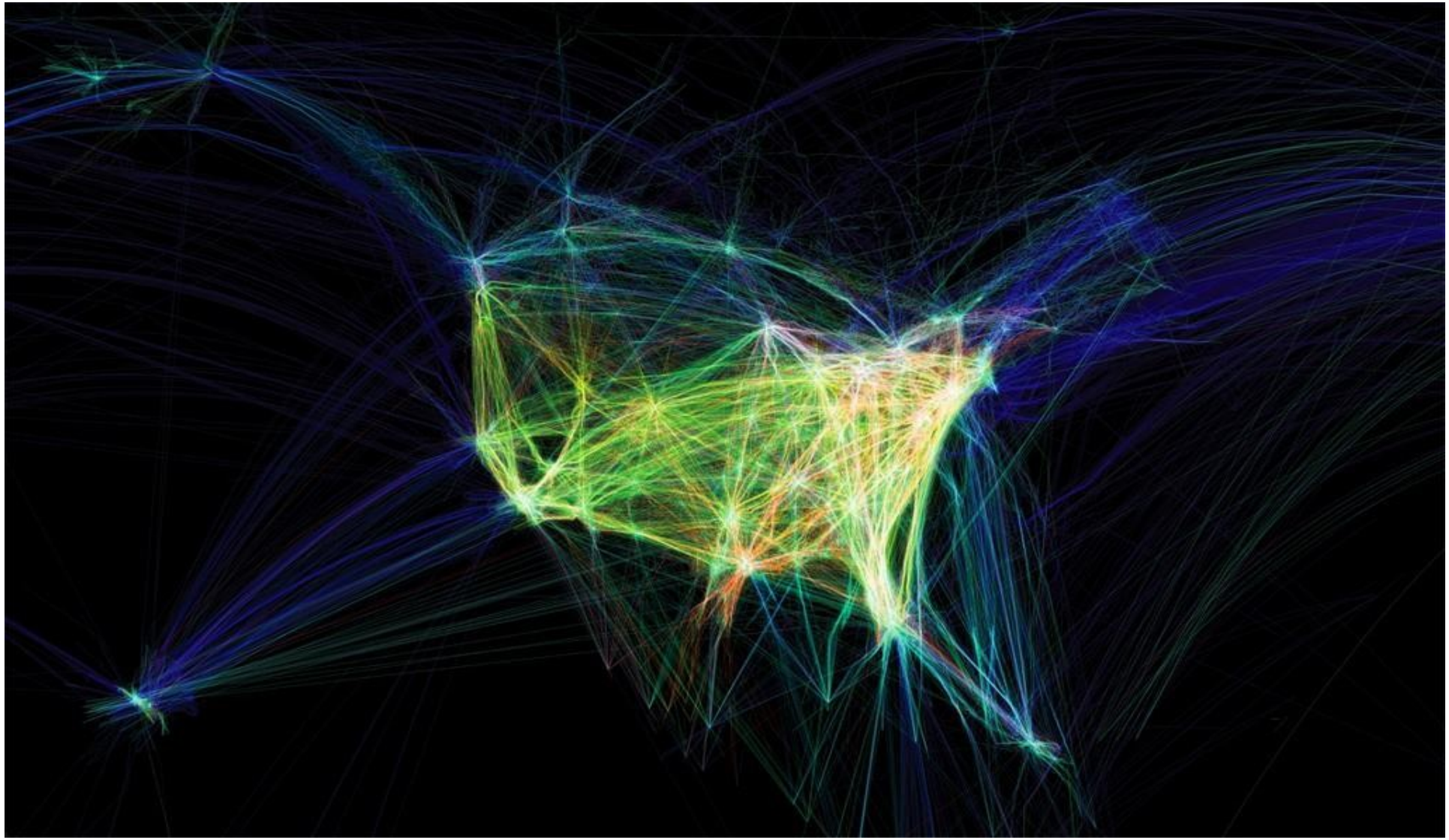
Visual Art

- The third category, visual art, primarily serves the relationship between the designer and the data.

- Visual art is unlike the previous two categories in that it often entails **unidirectional encoding of information, meaning that the reader may not be able to decode the visual presentation to understand the underlying information.**

- The designer may intend only to condense it, translate it into a new medium, or make it beautiful; she may not intend for the reader to be able to extract anything from it other than enjoyment.

Fight Patterns, Aaron Koblin



Facebook Visualised

Classifications of visualisations

1. Complexity
2. Infographics vs Data visualisation
3. Exploration vs Explanation
4. Informative vs Persuasive vs Visual Art

"Designing Data visualisations", N. Iliinsky & J. Steele, O'Reilly Media, 2011

Classifications of visualisations

- There are many different ways to classify visualisations
- Classifications can be useful when discussing/planning visualisation projects
- Classifications are also useful when determining the appropriate visualisation techniques to deploy

Interesting Reading

- Open Mind Blog.
  - USA Temperature: can I sucker you?
  - Link: https://tamino.wordpress.com/2018/08/08/usa-temperature-can-i-sucker-you/

Video

- Weather visualisation
  - https://youtu.be/q01vSb_B1o0?t=46

Exercise 1

Discuss in groups of 3 your good and bad visualisation examples.

Write a list of good features and bad features identified in the examples.

Upload them to Brightspace group discussion.

Exercise 2

Suggest a visualisation of Dublin's traffic congestion consider 3 variables:

- Time
- Volume
- Geography

Instructions:

- groups of 3
- Take 30 mins
- Sketch on board when ready

Thanks To

- Marisa Llorens-Salvador, John McAuley, Colman McMahon and Brian Mac Namee for an earlier version of these lecture notes