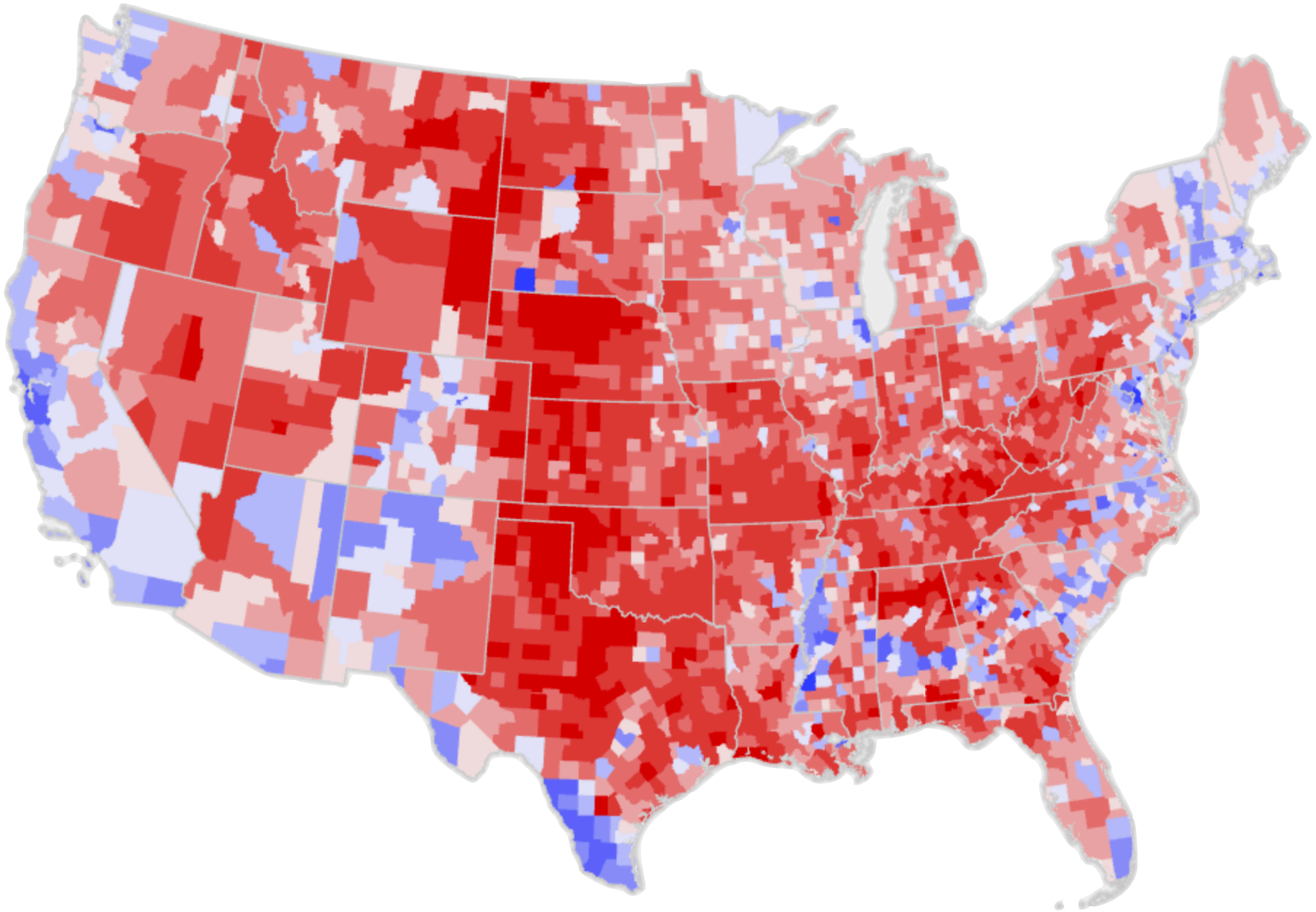
Information Visualization

Introduction

Dr. David Koop

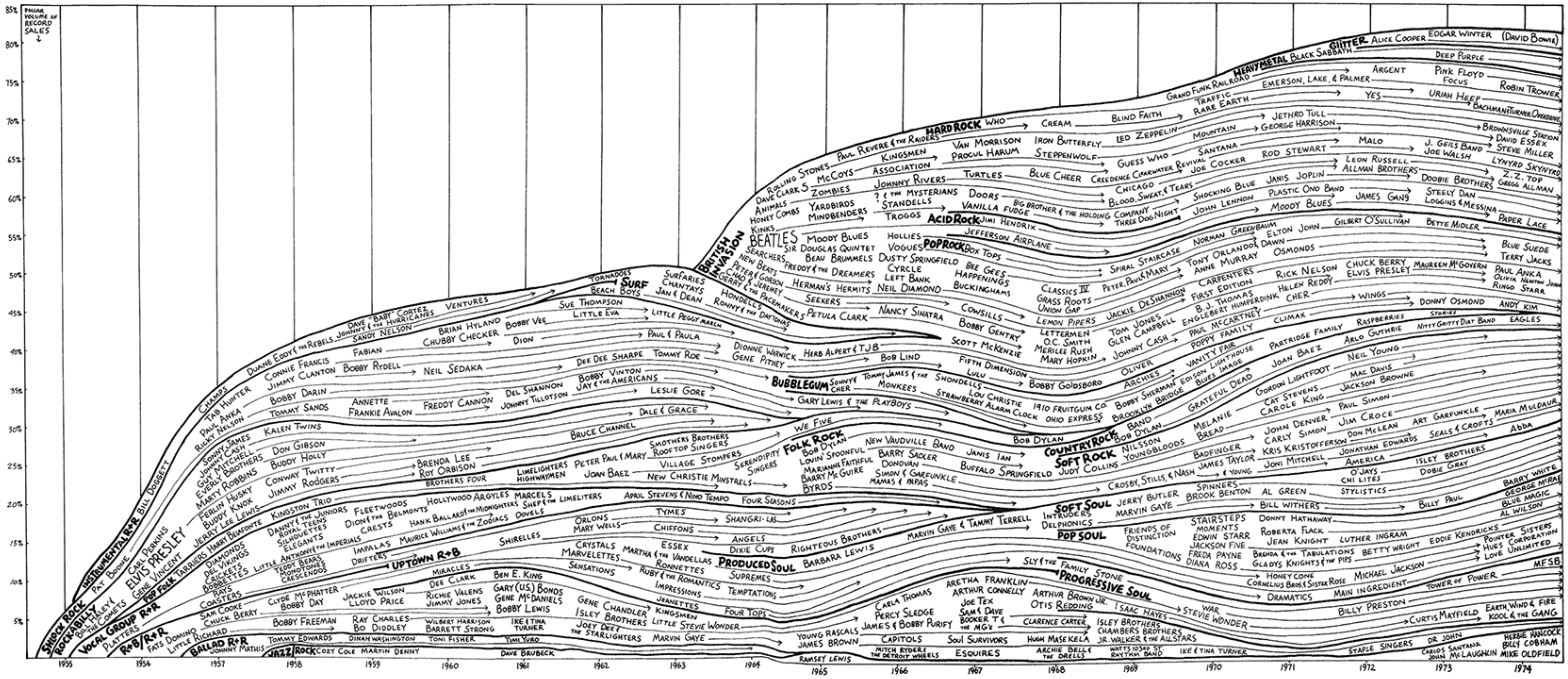
What is Data Visualization?

Data Visualization



[K. Field]

Data Visualization



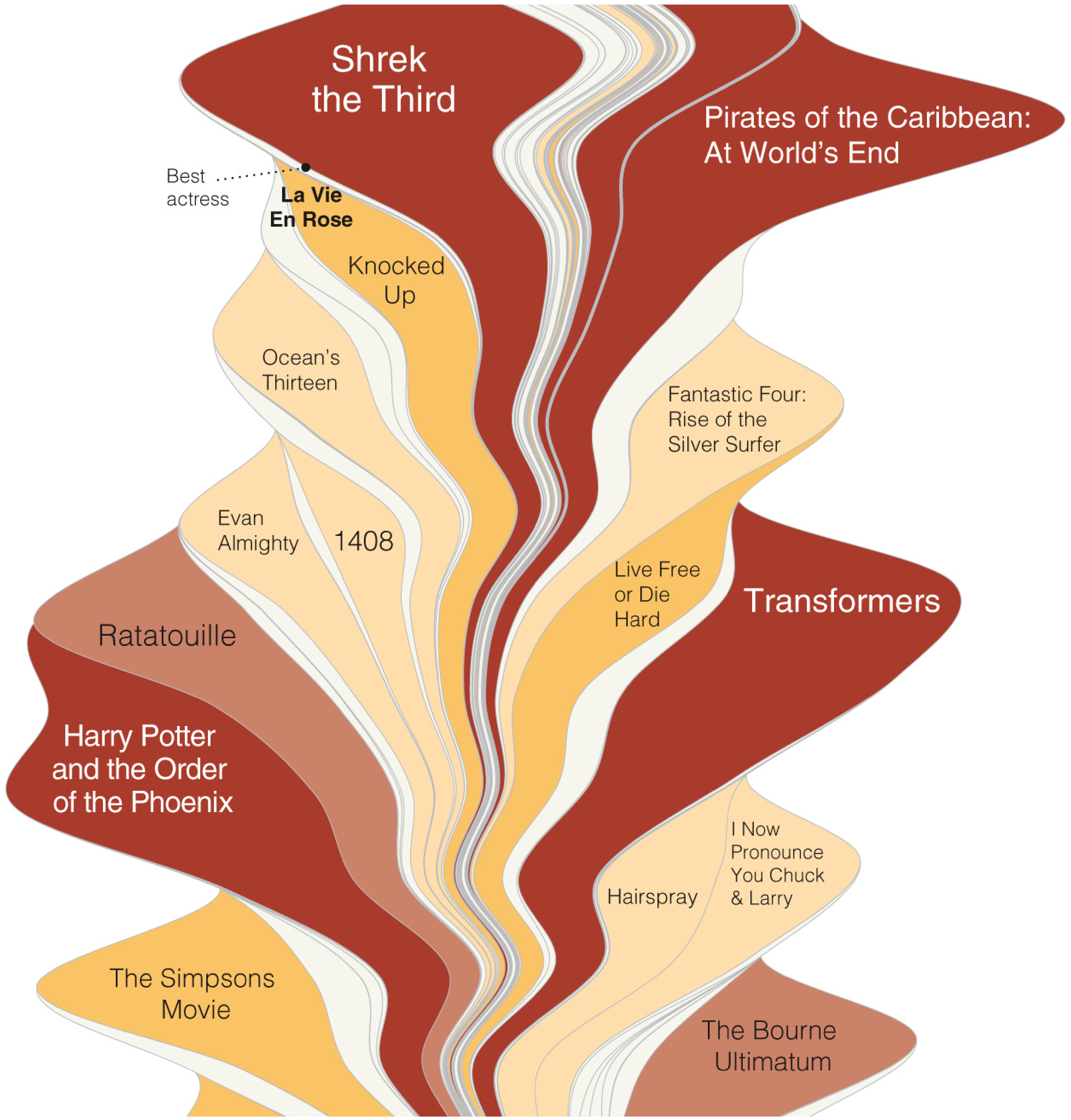
[Rock 'N' Roll is Here to Pay, R. Garofalo, 1977 (via Tufte)]

What is **Information** Visualization (InfoVis)?

Compared to Statistical Graphics?

Streamgraphs

- Visualize movie ticket sales by time
- Stack films that are in theaters on top of each other
- Area = the total sales
- "You can see Oscar contenders attract a smaller audience than the holiday and summer blockbusters and kind of slowly build an audience." — N. Yau, FlowingData

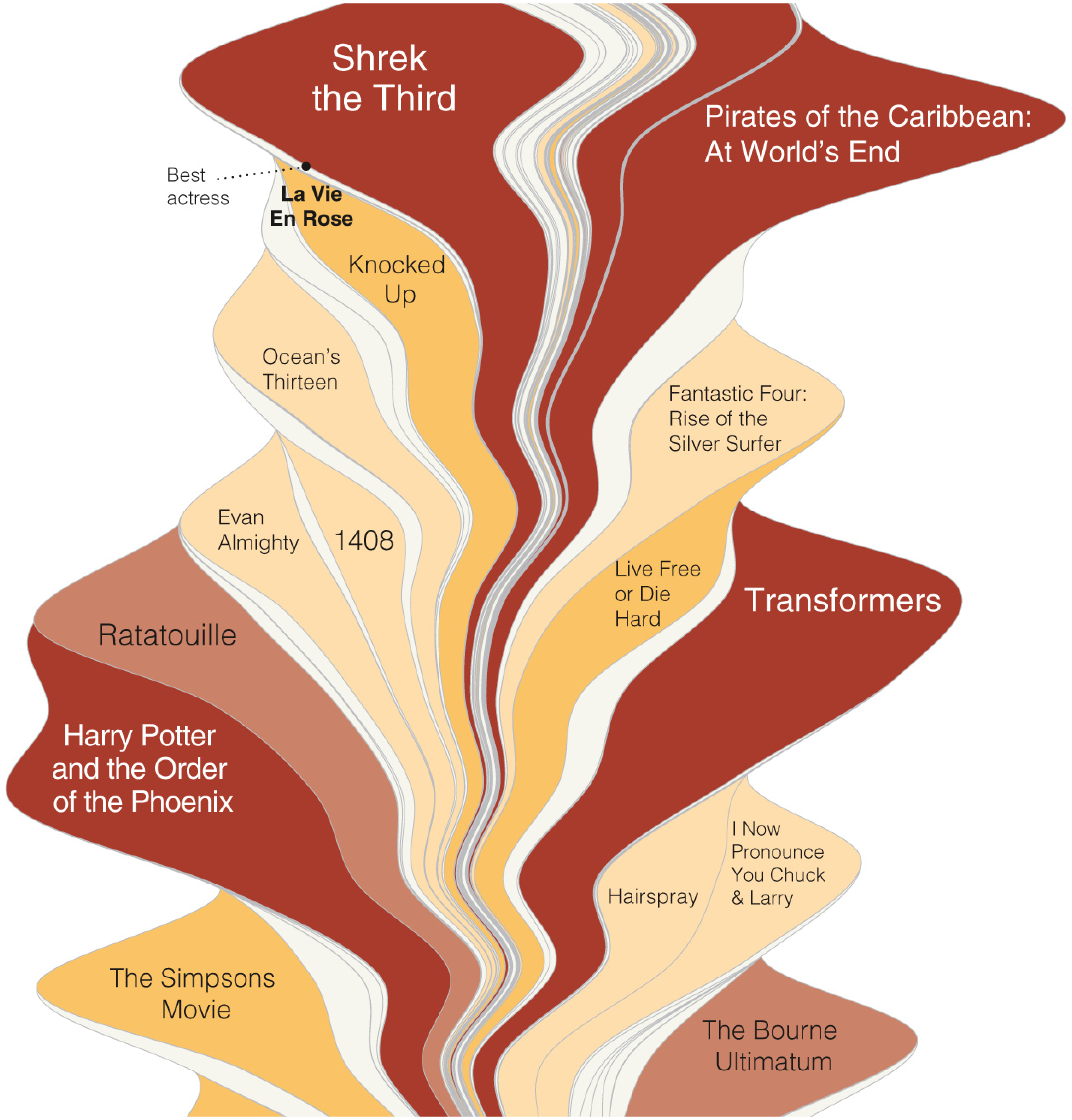


[Byron and Wattenberg, 2012]

Streamgraphs

- [Gelman & Unwin] Instead use two plots:
  1. Total sales over time
  2. Trajectories for individual movies

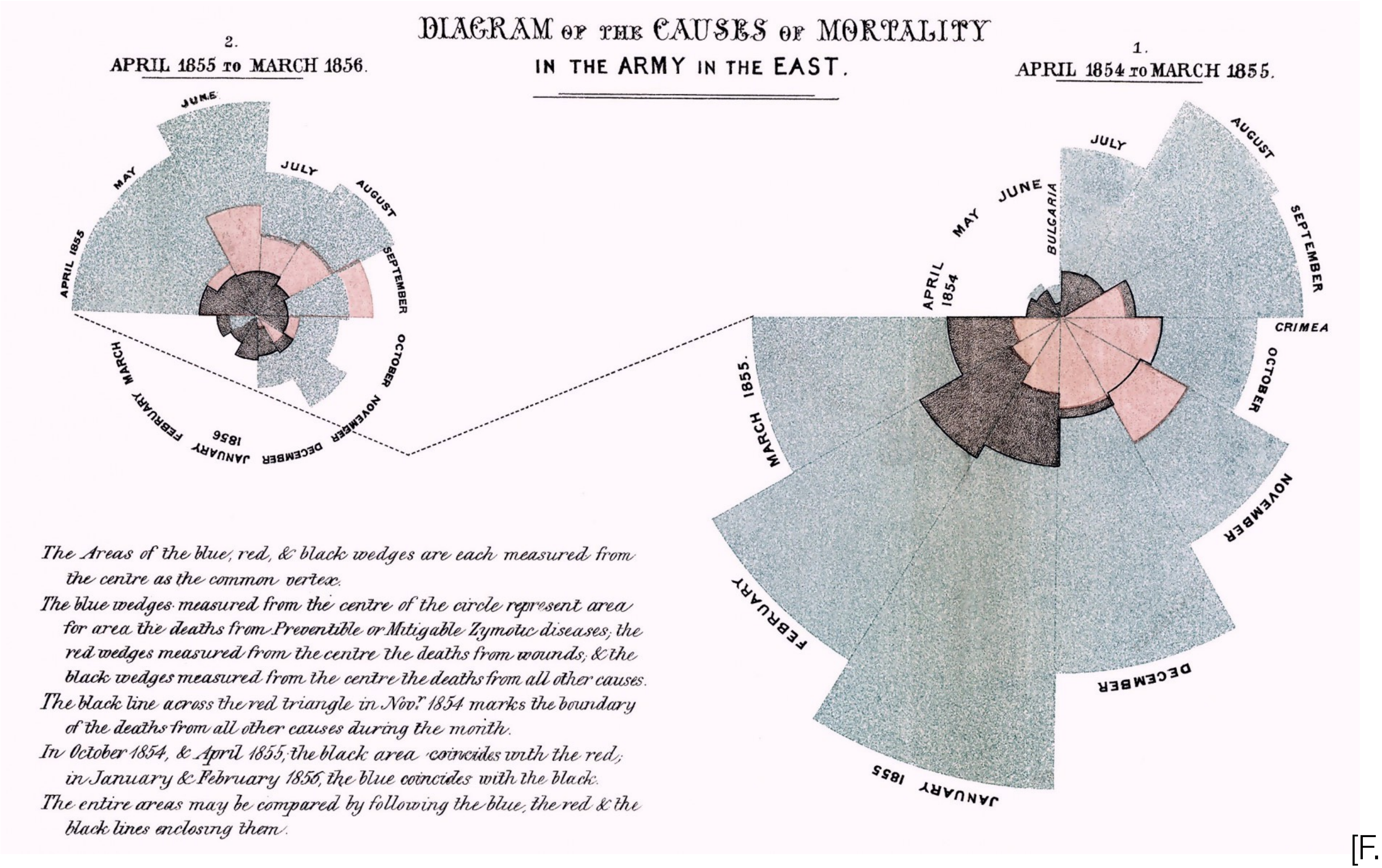
- "Discussion burst out across the Web . . . that I am convinced would not have come about if instead of a Streamgraph, they used say,a stacked bar chart." — N. Yau



[Byron and Wattenberg, 2012]

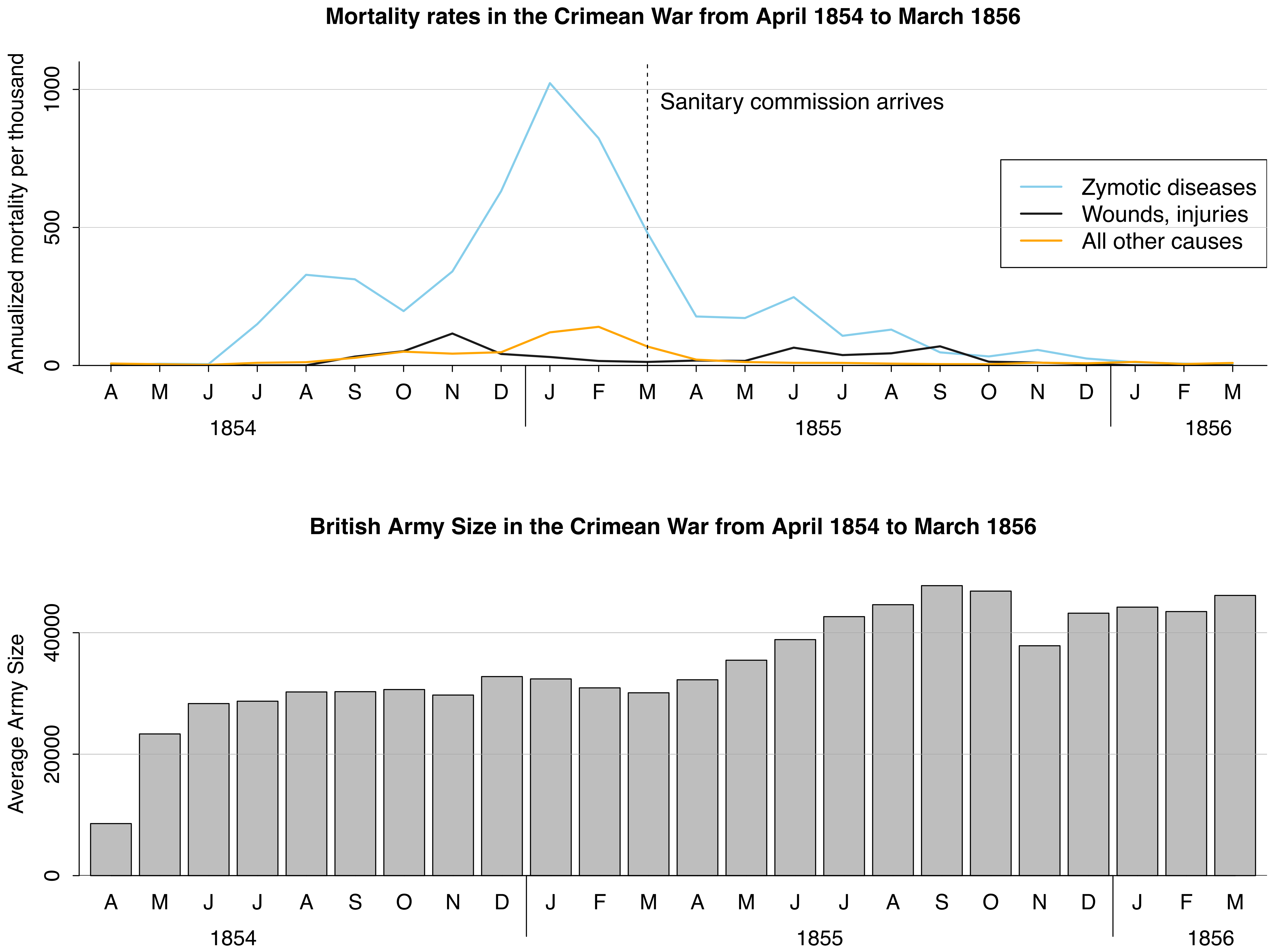
"That Puzzle-Solving Feeling" — [Gelman & Unwin]

Nightingale's Coxcomb Diagram



[F. Nightingale, 1858]

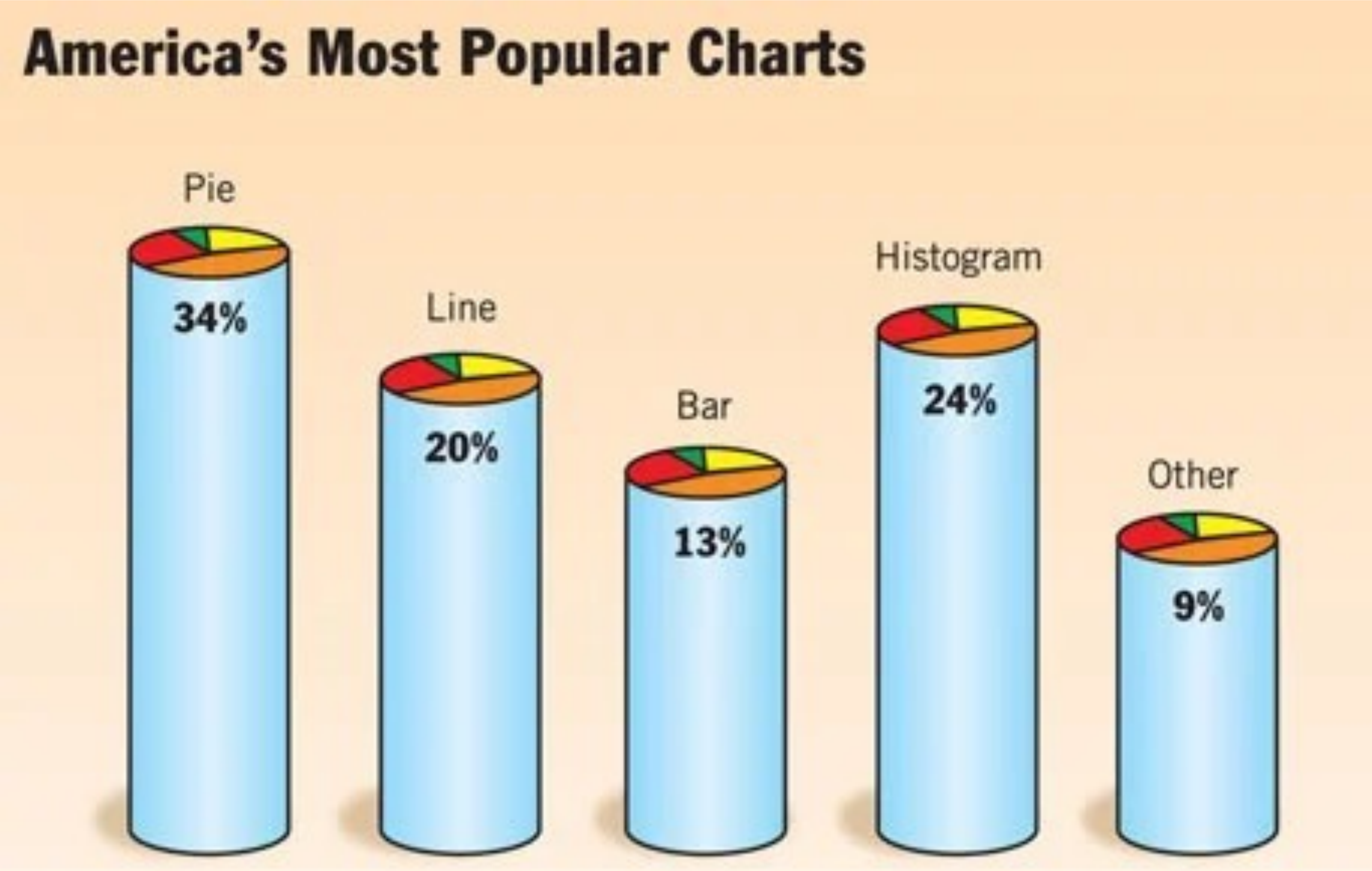
Gelman and Unwin's Remake



[Gelman and Unwin, 2014]

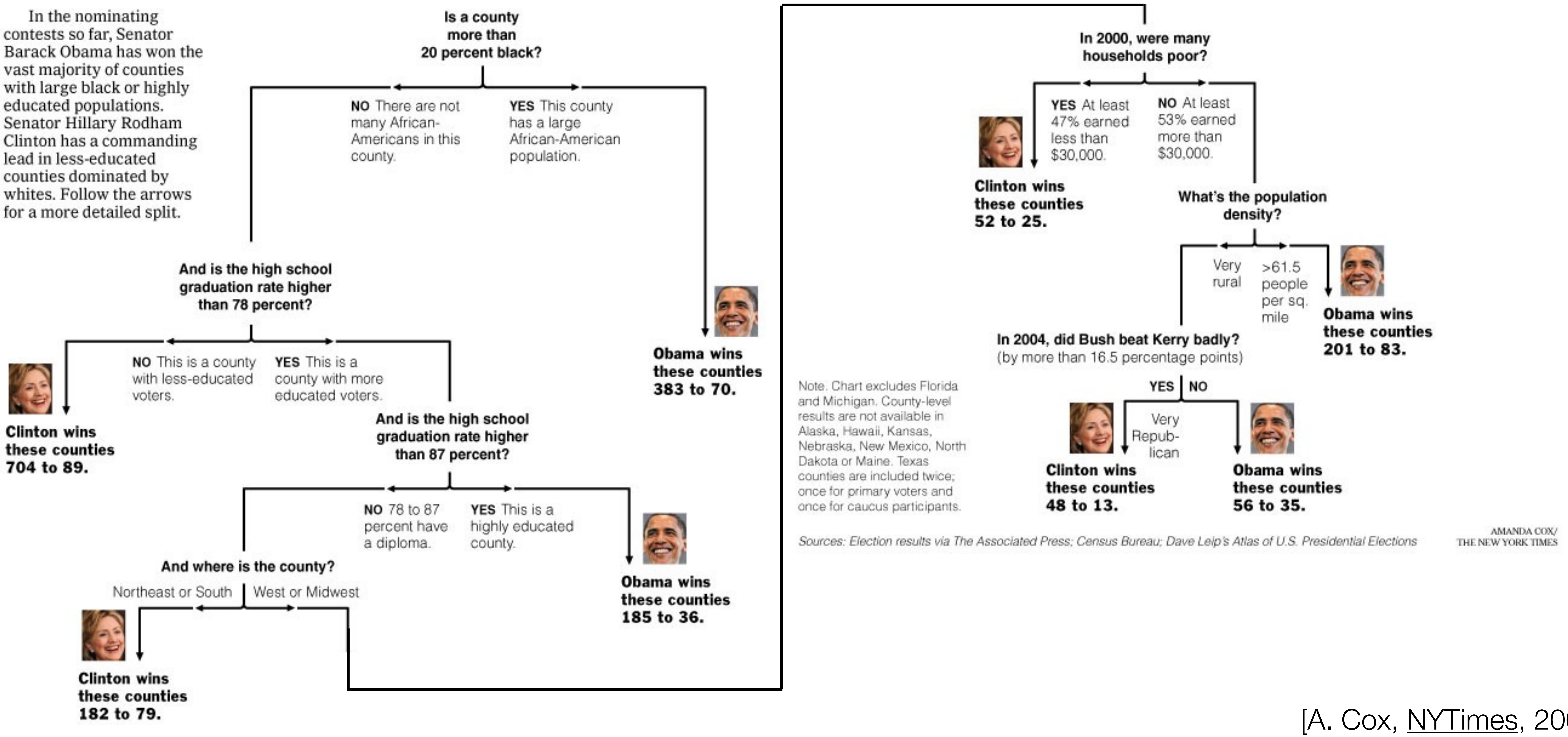
Compared to Infographics?

America's Most Popular Charts



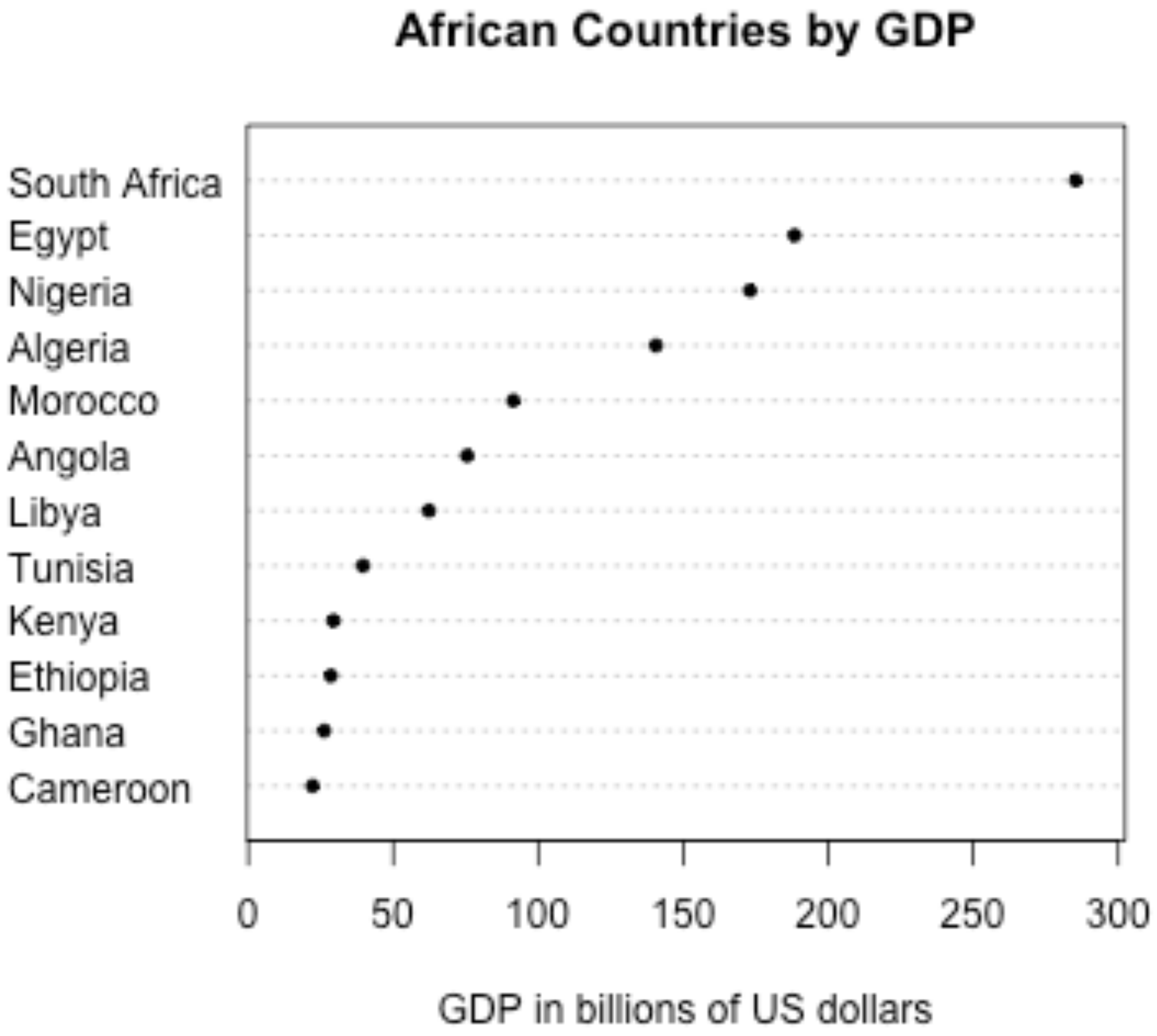
[The Onion, 2007]

Decision Tree: The Obama-Clinton Divide



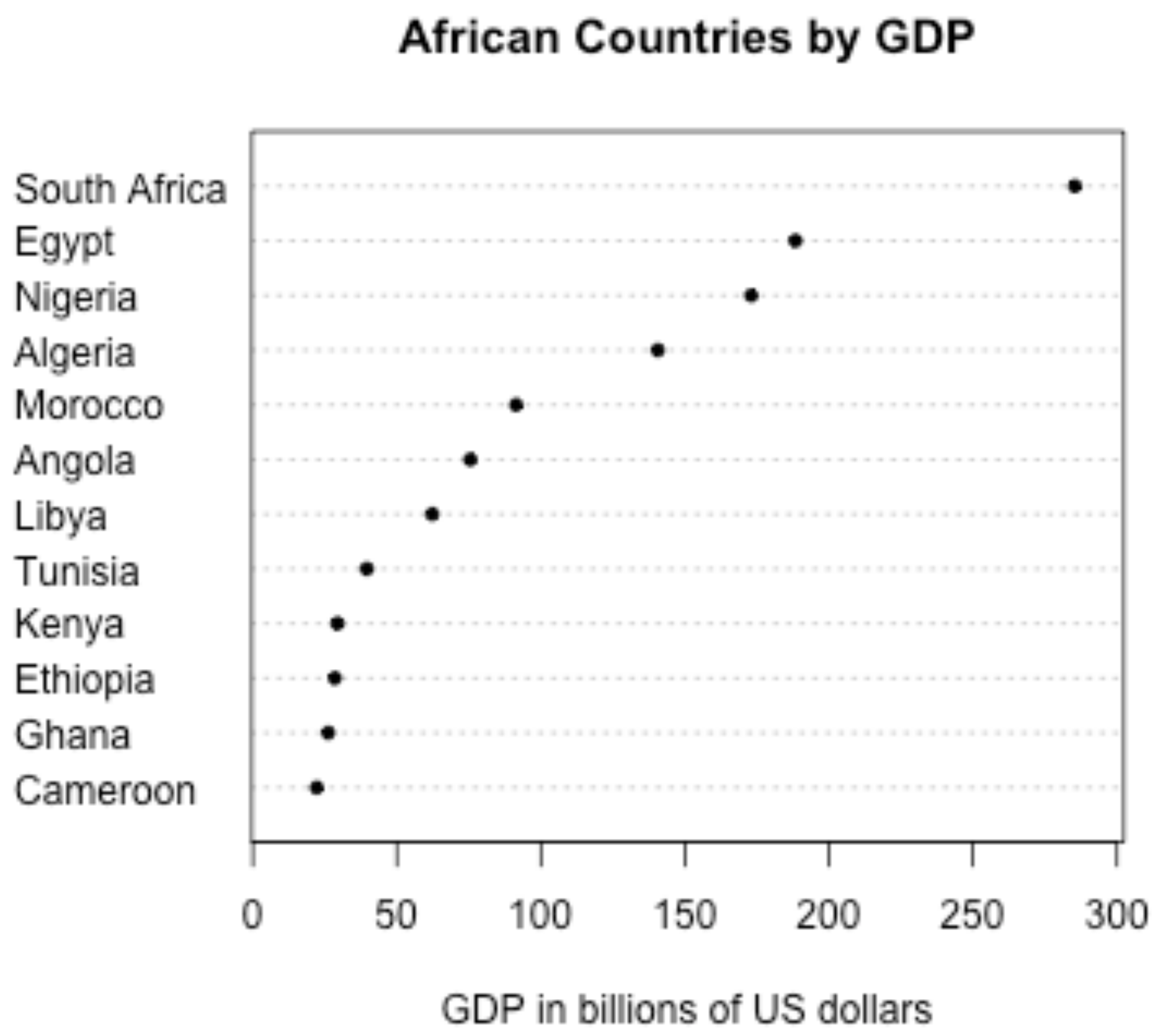
[A. Cox, NYTimes, 2008]

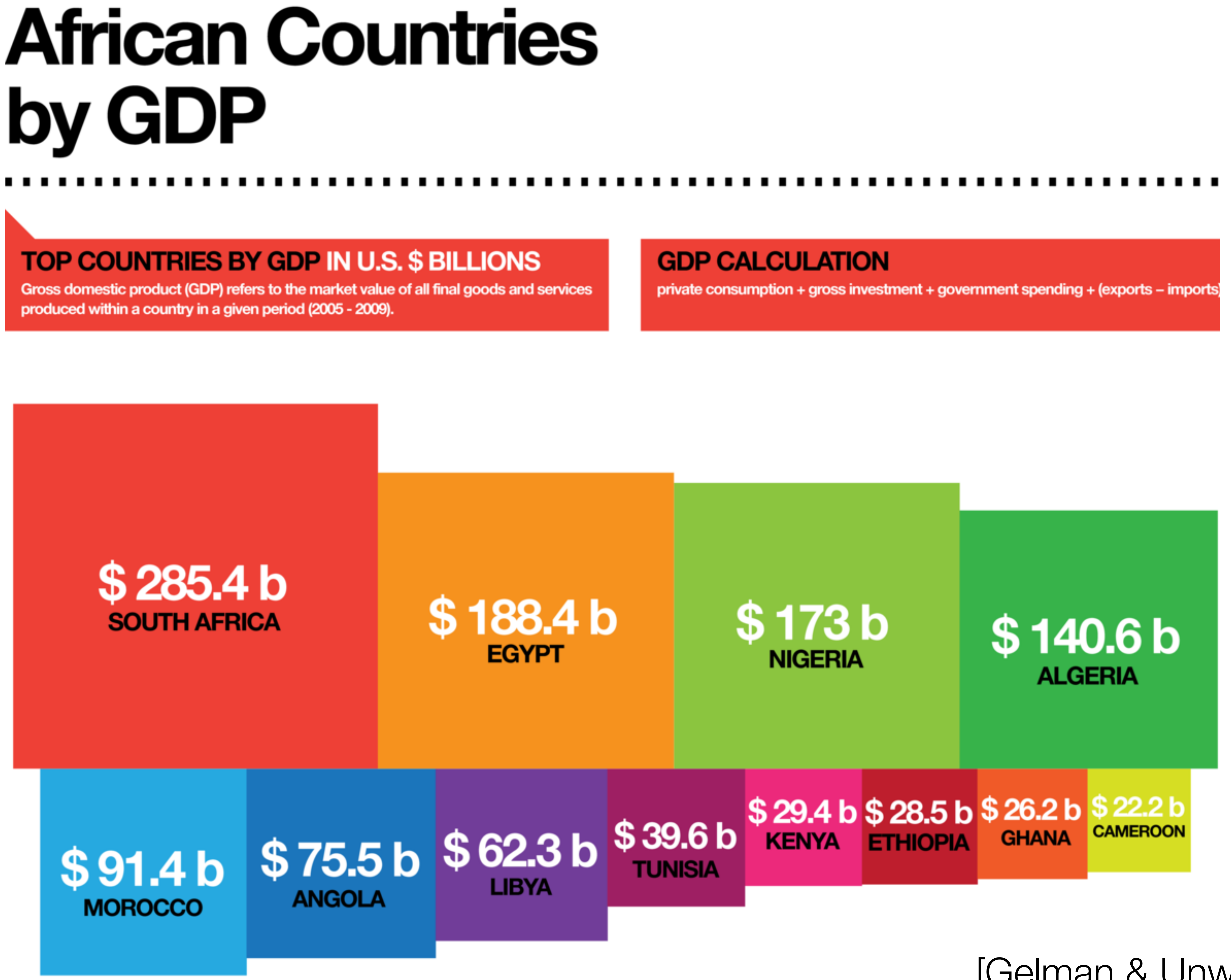
Infographics Embellish Boring Plots?



[Gelman & Unwin]

Infographics Embellish Boring Plots?

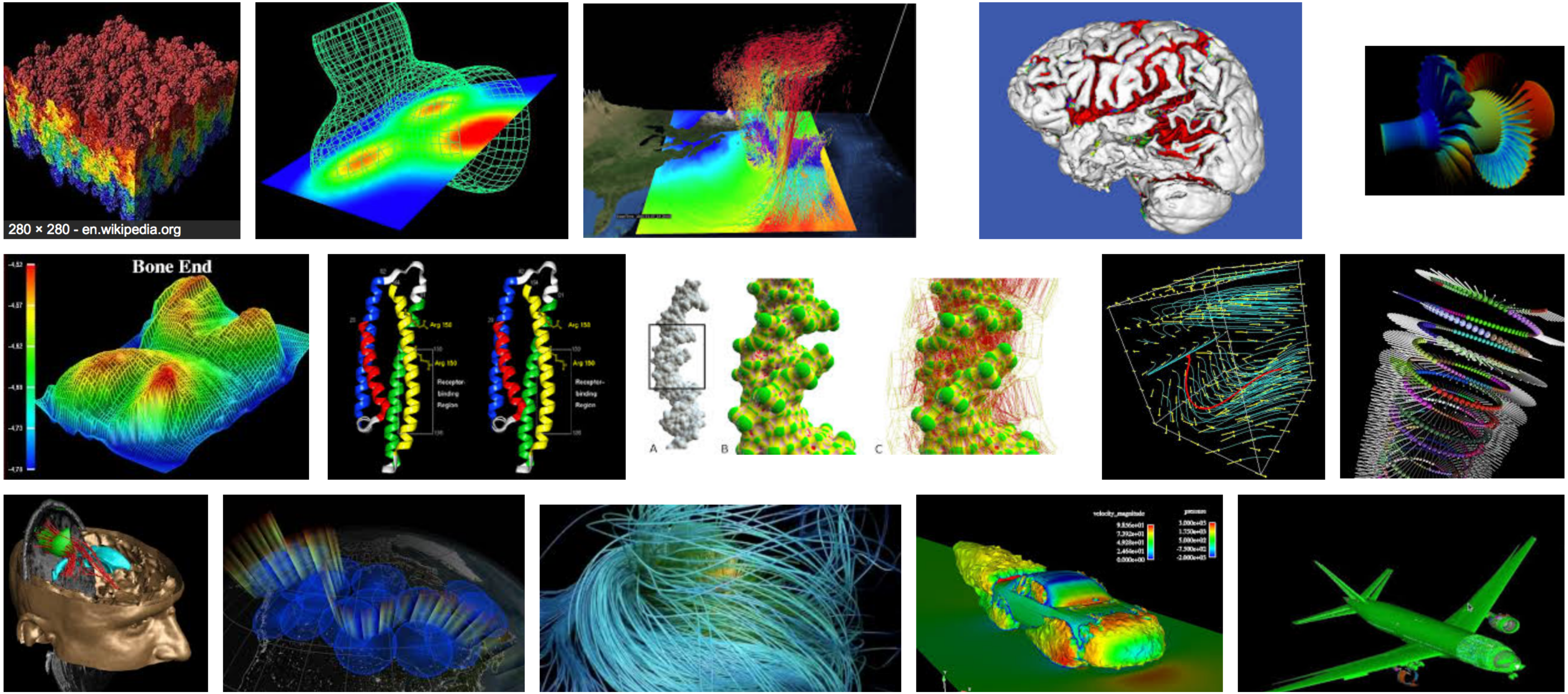




[Gelman & Unwin]

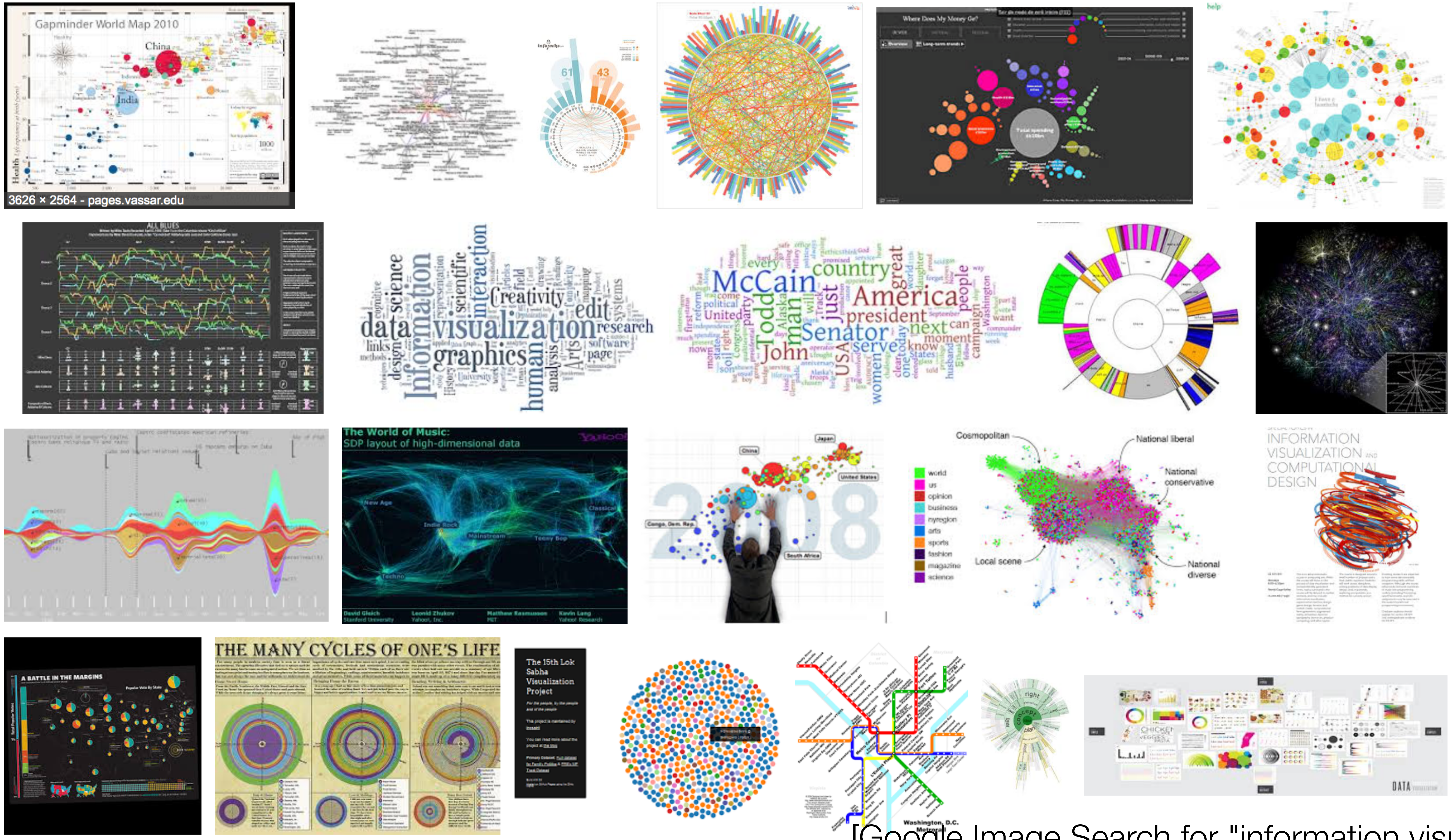
Compared to Scientific Visualization?

SciVis



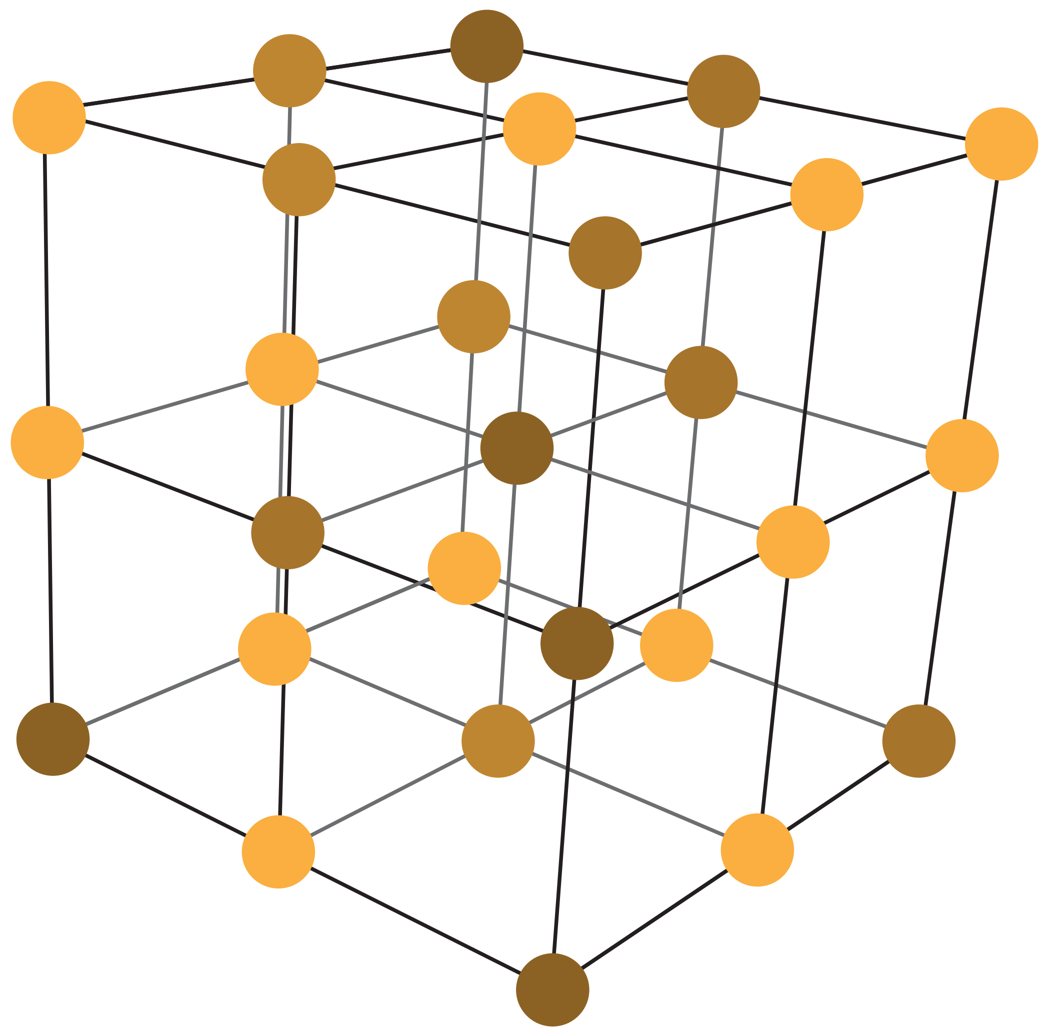
[Google Image Search for "scientific visualization", 2017]

InfoVis



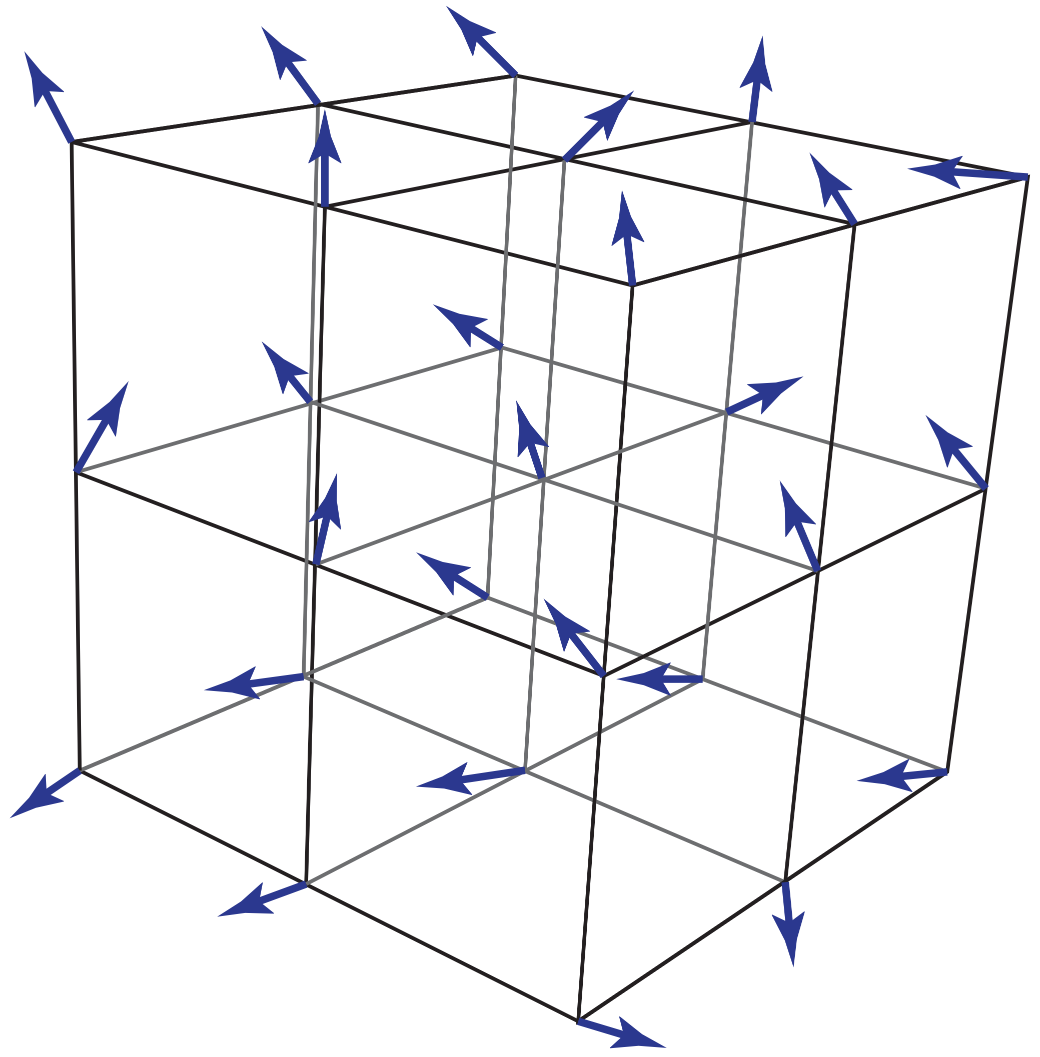
[Google Image Search for "information visualization", 2017]

SciVis → Fields



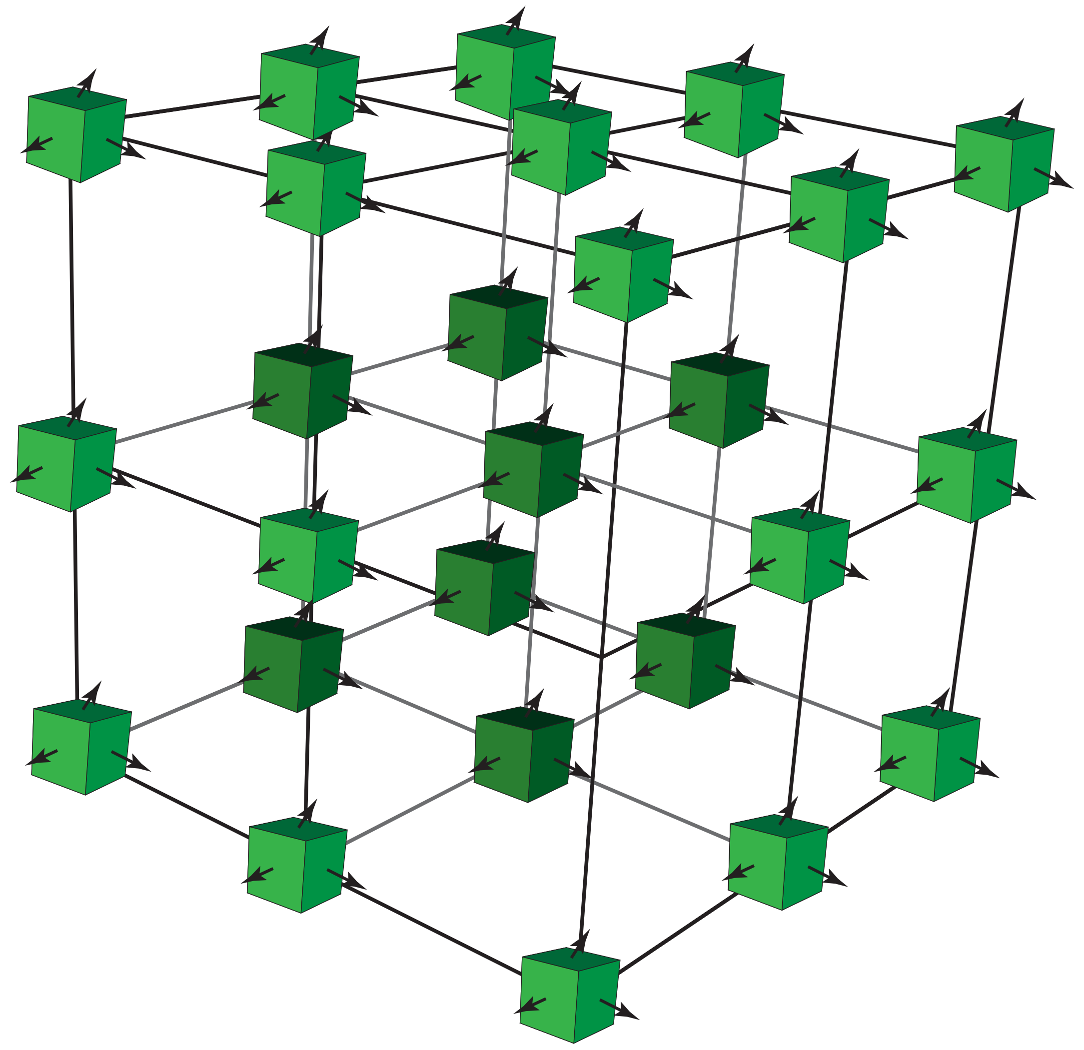
Scalar Fields

(Order-0 Tensor Fields)



Vector Fields

(Order-1 Tensor Fields)



Tensor Fields

(Order-2+)

Each point in space has an associated...

$$s_0$$

Scalar

$$\begin{bmatrix} v_0 \\ v_1 \\ v_2 \end{bmatrix}$$

Vector

$$\begin{bmatrix} \sigma_{00} & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_{11} & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Tensor

Visualization Taxonomy Structure

	Display Attributes: *Given*	*Constrained*	*Chosen*
Continuous	Images (e.g., medical) Fluid / gas flow, pressure distributions Molecular structures (distributions of mass, charge, etc.) Globe – distribution data (e.g., elevation levels)	Distortions of given / continuous ideas (e.g., flattened medical structures, 2D geographic maps, fish-eye lens views) Arrangement of numeric variable values	Continuous (high-dimensional) mathematical functions Continuous time-varying data, when time is mapped to a spatial dimension Regression analyses
Discrete	Classified data / images (e.g., segmented medical images) Air traffic positions Molecular structures (exact positions of components) Globe – discrete entity data (e.g., city locations)	Distortions of given / discrete ideas (e.g., 2D geographic maps, fish-eye lens views) Arrangement of ordinal or numeric variable values	Discrete time-varying data, when time is mapped to a spatial dimension Arbitrary entity-relationship data (e.g., file structures) Arbitrary multi-dimensional data (e.g., employment statistics)

[Tory & Möller]

Visualization Taxonomy Structure

SciVis

	Given	*Constrained*	*Chosen*
		Display Attributes	
Continuous	Images (e.g., medical) Fluid / gas flow, pressure distributions Molecular structures (distributions of mass, charge, etc.) Globe – distribution data (e.g., elevation levels)	Distortions of given / continuous ideas (e.g., flattened medical structures, 2D geographic maps, fish-eye lens views) Arrangement of numeric variable values	Continuous (high-dimensional) mathematical functions Continuous time-varying data, when time is mapped to a spatial dimension Regression analyses
Discrete	Classified data / images (e.g., segmented medical images) Air traffic positions Molecular structures (exact positions of components) Globe – discrete entity data (e.g., city locations)	Distortions of given / discrete ideas (e.g., 2D geographic maps, fish-eye lens views) Arrangement of ordinal or numeric variable values	Discrete time-varying data, when time is mapped to a spatial dimension Arbitrary entity-relationship data (e.g., file structures) Arbitrary multi-dimensional data (e.g., employment statistics)

[Tory & Möller]

Visualization Taxonomy Structure

SciVis

	Display Attributes: *Given*	*Constrained*	*Chosen*
Continuous	Images (e.g., medical) Fluid / gas flow, pressure distributions Molecular structures (distributions of mass, charge, etc.) Globe – distribution data (e.g., elevation levels)	Distortions of given / continuous ideas (e.g., flattened medical structures, 2D geographic maps, fish-eye lens views) Arrangement of numeric variable values	Continuous (high-dimensional) mathematical functions Continuous time-varying data, when time is mapped to a spatial dimension Regression analyses
Discrete	Classified data / images (e.g., segmented medical images) Air traffic positions Molecular structures (exact positions of components) Globe – discrete entity data (e.g., city locations)	Distortions of given / discrete ideas (e.g., 2D geographic maps, fish-eye lens views) Arrangement of ordinal or numeric variable values	Discrete time-varying data, when time is mapped to a spatial dimension Arbitrary entity-relationship data (e.g., file structures) Arbitrary multi-dimensional data (e.g., employment statistics)

InfoVis

[Tory & Möller]

Kosara's Definition of Information Visualization

- **It is based on (non-visual) data**. "The data to be visualized must come from outside the program, and the program must be able (at least in principle) to work on different data sets. Also, visualization is not image processing or photography; if the source data is an image and is used as an image in the result, it is not being visualized."
- **It produces an image**. "Clearly, each visualization has the goal of producing one or more images from the data, and the visual must be the primary means of communicating the data. Other media can be part of a visualization, but the visualization must be able to stand on its own."

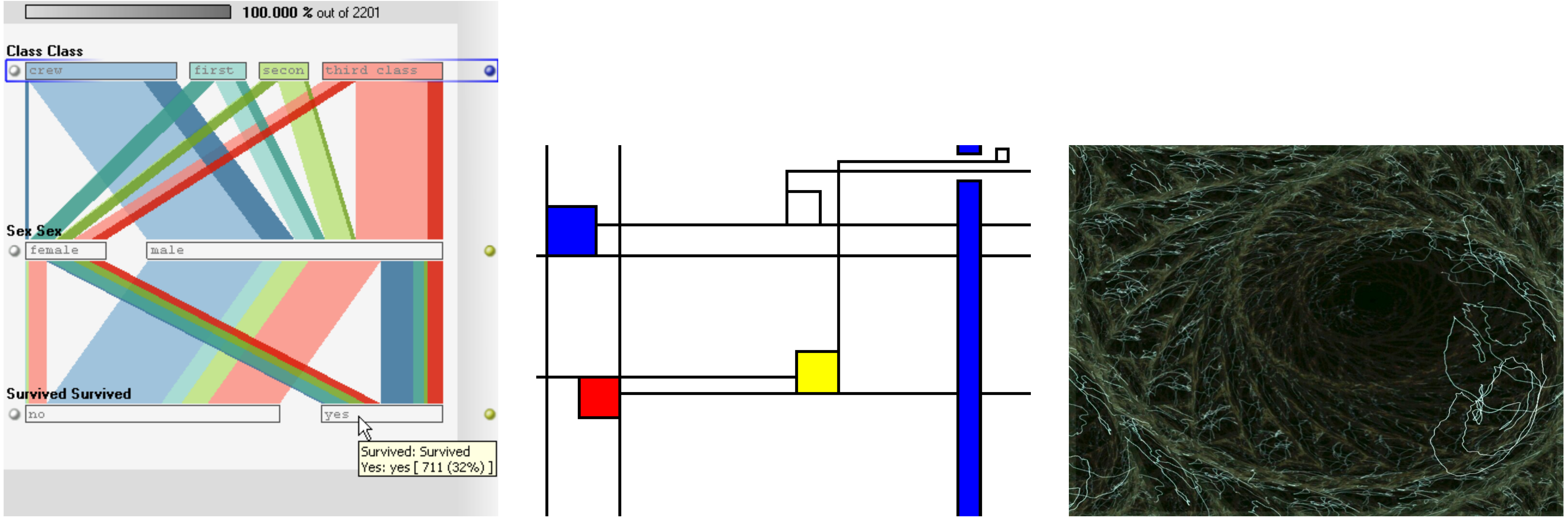
[R. Kosara]

Kosara's Definition of Information Visualization

- **The result is readable and recognizable.** "There are many ways to transform data into images, most of which do not allow the viewer to understand the underlying data. A visualization must produce images that are readable by a viewer, even if that requires training and practice. Visualization images must also be recognizable as such, and not appear to be something else. The use of additional elements (or even “eye candy”) is certainly possible, but must not take precedence over the communication goals of the visualization."

[R. Kosara]

Pragmatic <-> Artistic Visualization



a) b) c)

Parallel Sets, etc.	Informative Art	Music Visualization, etc.
Recognizable Readable	Not Recognizable Readable	Not Recognizable Not Readable

Pragmatic Visualization ⟷ Artistic Visualization

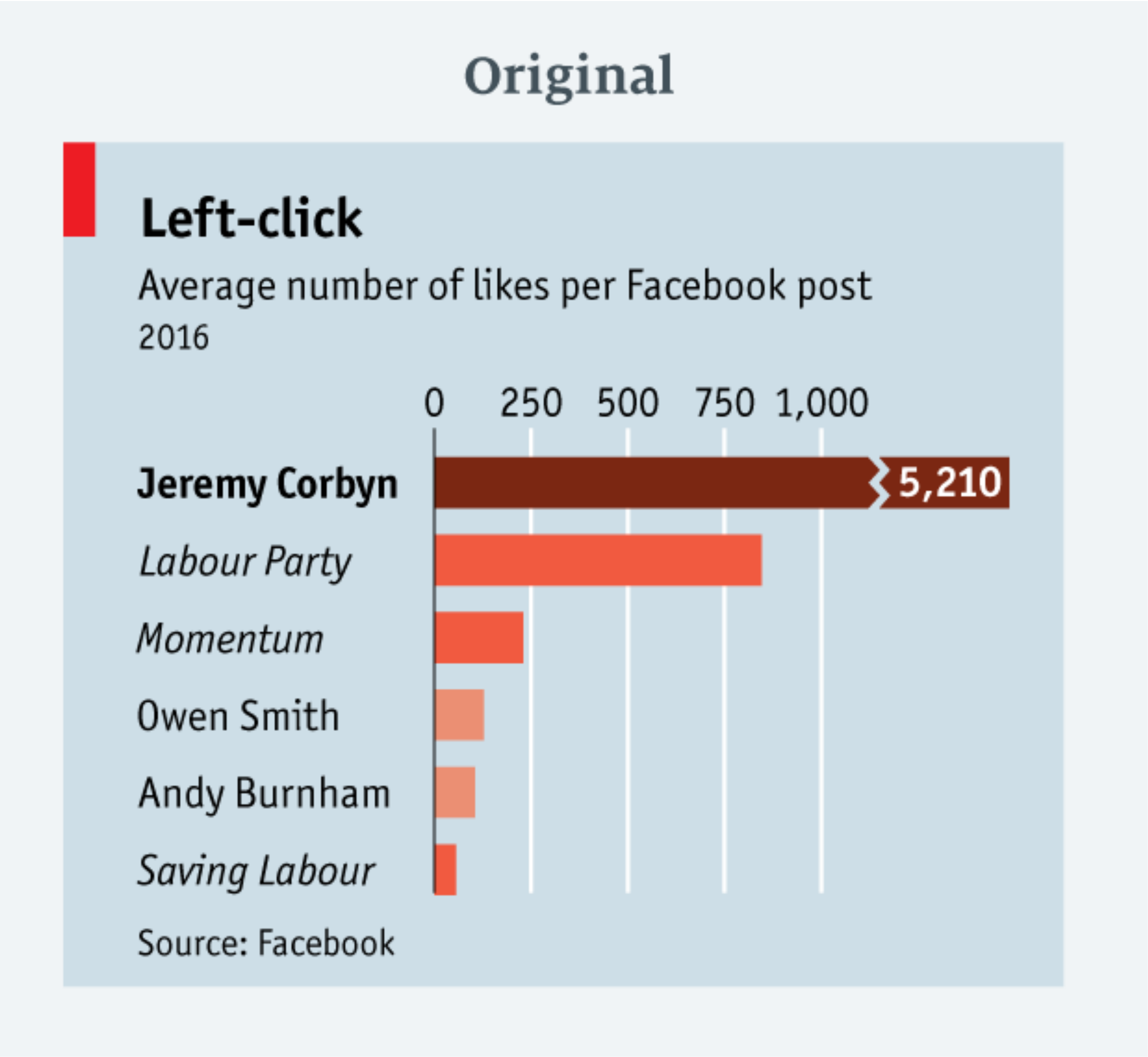
Utilitarian ⟷ Sublime

[R. Kosara]

Rules

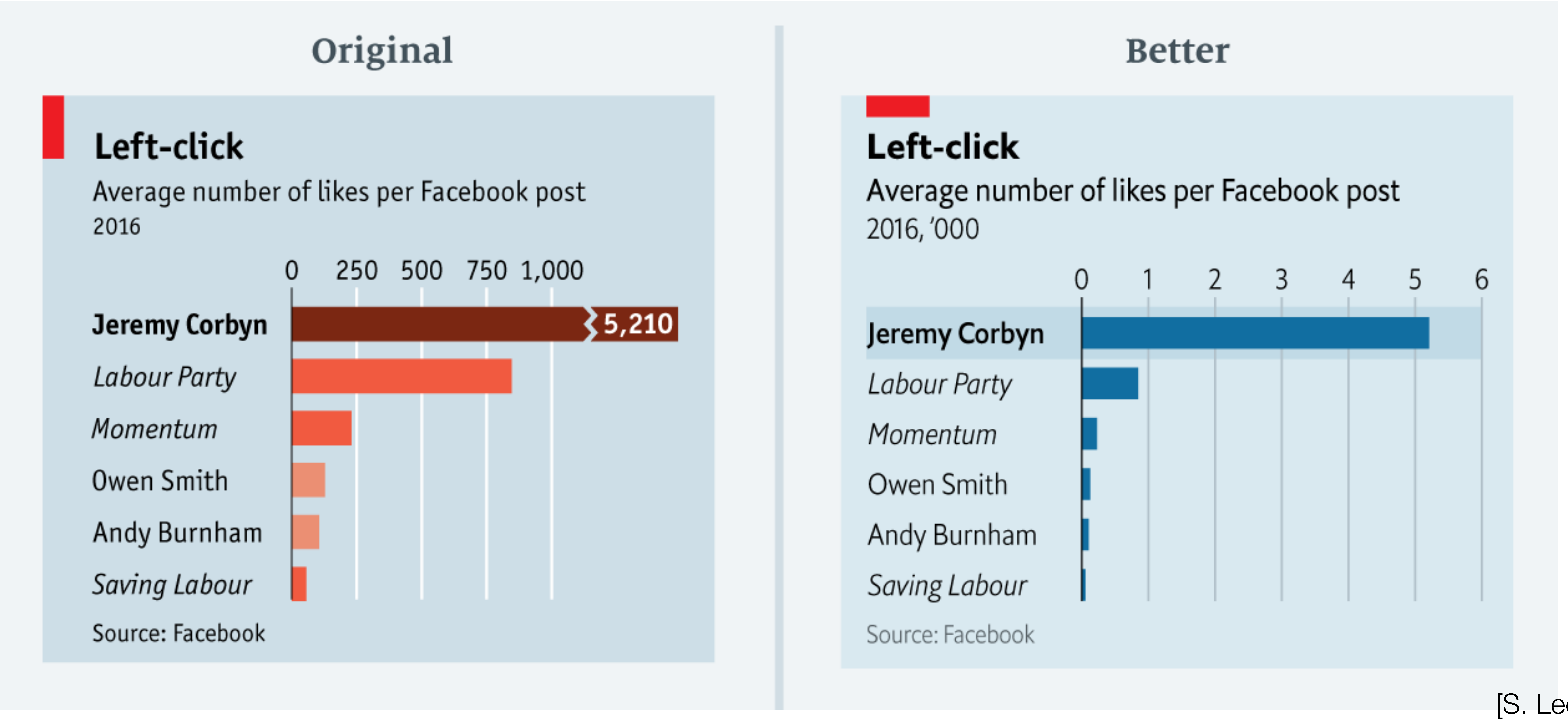
- We saw many rules in CSCI 627 (Data Visualization)
- How do we use those to think about visualization?

Visualization Mistakes



[S. Leo]

Visualization Mistakes



[S. Leo]

Can we break the rules?

Benefitting InfoVis with Visual Difficulties

	Cognitive efficiency	Visual difficulties
Cognitive operations	Minimize the cognitive steps required to process visualization	Induce constructive, self-directed cognitive activity on the part of the user
Data-ink ratio	Maximize the ratio of data to ink	Design representations that are most likely to engage a user to actively process the information

[J. Hullman et al.]

Benefitting InfoVis with Visual Difficulties

Organization	Choose the format which makes important information most visually salient	Choose the format that best stimulates deep cognitive reflection on the important data
Animation	Use animation to quickly and intuitively visualize important processes	Use static representations to induce interval visualization processing around causal mechanisms; consider animation in cases where mental animation lies beyond users' capacities
Labeling	Use labels rather than legends to optimize immediate clarity	Use legends to stimulate deeper reflection on data

[J. Hullman et al.]

Administrivia

About Me

- Research Interests
  - Visualization
  - Computational Provenance
  - Geospatial Analysis
- Research Projects
  - Dataflow Notebooks
  - Geospatial Trajectory Data
  - Provenance for Web Applications
- See my web page for more information
  - http://faculty.cs.niu.edu/~dakoop/

About You

- Research Papers?
- Data Visualization
- Tools? JavaScript, D3, Tableau, Others?
- Research Experience?
- What topics do you want to see covered?

About this course

- Course web page is authoritative:
  - http://faculty.cs.niu.edu/~dakoop/cs628-2021fa/
  - Schedule, Readings, Assignments will be posted online
  - Check the web site before emailing me
- Lectures: TuTh 9:30-10:45am in PM 252
- This is an Advanced (Tier 2) Graduate Course
  - Present and discuss cutting-edge topics
  - Work on research problems
- Requires **participation**: readings and discussions

About this Course

- "Focus on advanced theory and methods for manipulating and visualizing the data of non-physical systems… Emphasis on the advanced study of the latest information visualization techniques developed by the research community. A computer programming background is required. Extensive laboratory work."

About this course

- Course Registration:
  - Make sure you have registered for the course
  - Email me if you are not registered but are interested in taking the course
- Review of course policies:
  - Plagiarism and academic honesty
  - If you have any concerns or questions, please email me as soon as possible
- If you are not sure if this course is a good fit, please email me or talk to me

Office Hours & Email

- Office hours will be held in person
  - Tu: 1:45-3pm, Th: 10:45am-12pm, or by appointment
- Please adhere to university regulations (Protecting the Pack)
- You do not need an appointment to stop by during scheduled office hours
- If you wish to meet virtually, please schedule an appointment
- If you need an appointment, please email me with **details** about what you wish to discuss and times that would work for you
- Many questions can be answered via email. **Please consider writing an email before scheduling a meeting.**

Expectations

- Be engaged:
  - Active participation
  - Constructive participation
- Work independently: self-directed and sustained
- Work collaboratively: learn from each other
- Put effort into the course:
  - Must put significant work in **each** week
  - Do not try to do everything before a deadline
  - Best effort (success or failure)

Next Class

- Review of Core Visualization Topics
- Topics Survey
- Bring Your Ideas