Data Exploration & Visualization

Module 10

Temporal Visualization

DSAA 5024

The Hong Kong University of Science and Technology (Guangzhou)

Data Exploration & Visualization

Module 10: Temporal Visualization

- Time series data
  - Properties, tasks, taxonomy
- Temporal visualization
  - Linear times: Line charts, stacked graph
  - Interval times:
    - Cyclic: calendar view, cyclic layout
    - Linear: Gantt charts, lifelines
- Case study

Time series data

- Definition: Sets of values changing over time.
- Fundamental chronological component to the data set.
  - finance (stock prices, exchange rates)
  - science (temperatures, pollution levels, electric potentials)
  - public policy (crime rates)
- Random sample of 4000 graphics from 15 of world's newspapers and magazines from 1974 - 1980 found that 75% of graphics published were time series

  - Tufte, vol. 1

Data mining

- Data mining domain has techniques for algorithmically examining time series data, looking for patterns, etc.
- Good when objective is known a priori
- But what if not?
  - Which questions should I be asking?
  - Visualization better for that

Time series tasks

- What kinds of questions do people ask about time series data?

- Examples
  - When was something greatest/least?
  - Is there a pattern?
  - Are two series similar?
  - Do any of the series match a pattern?
  - Provide simple, fast access to the series

Time series tasks

- More examples
  - Does a data element exist at a specific time? (Existence of a data element)
  - When does a data element exist on time? Is there any cyclic behavior? (Temporal location)
  - How long is the time span from beginning to end of the data element? (Temporal interval)
  - How fast is a data element changing or how much difference is there from data element to data element over time? (Rate of change)
  - In what order do data elements appear? (Sequence)
  - Do data elements exist together? (Synchronization)

Taxonomy

- Discrete time vs interval time
  - Discrete time points have no duration.
  - Interval time uses an interval scaled time axis like days, months, or years.
    - Data elements d_i are specified by two time points.
- Linear time vs. cyclic time
  - Linear time assumes a starting point. Data elements from past to future.
  - Cyclic time uses a cyclic time axis. Point order of a cyclic time axis is meaningless with respect to a cycle.

Mueller & Schumann, 2003

Taxonomy

- Ordinal time vs. continuous time
  - Ordinal time: time axis is scaled ordinal, ‘before/after’.
  - Continuous time: quantify the time difference between the appearance of two data elements.
- Ordered time vs. branching time vs. time with multiple perspectives
  - Ordered time: events happen on after the other $d_{i-1} \rightarrow \mathrm{d}_i \rightarrow d_{i+1}$
  - Branching time: sequences of actions are foreseen. Multiple alternatives are possible. Typically used in decision making process. $d_i \begin{cases} d_{i+1} \\ d'_{i+1} \\ d''_{i+1} \end{cases}$
  - Time with multiple perspectives: more than one data element d_i for one time step t_i.

Mueller & Schumann, 2003

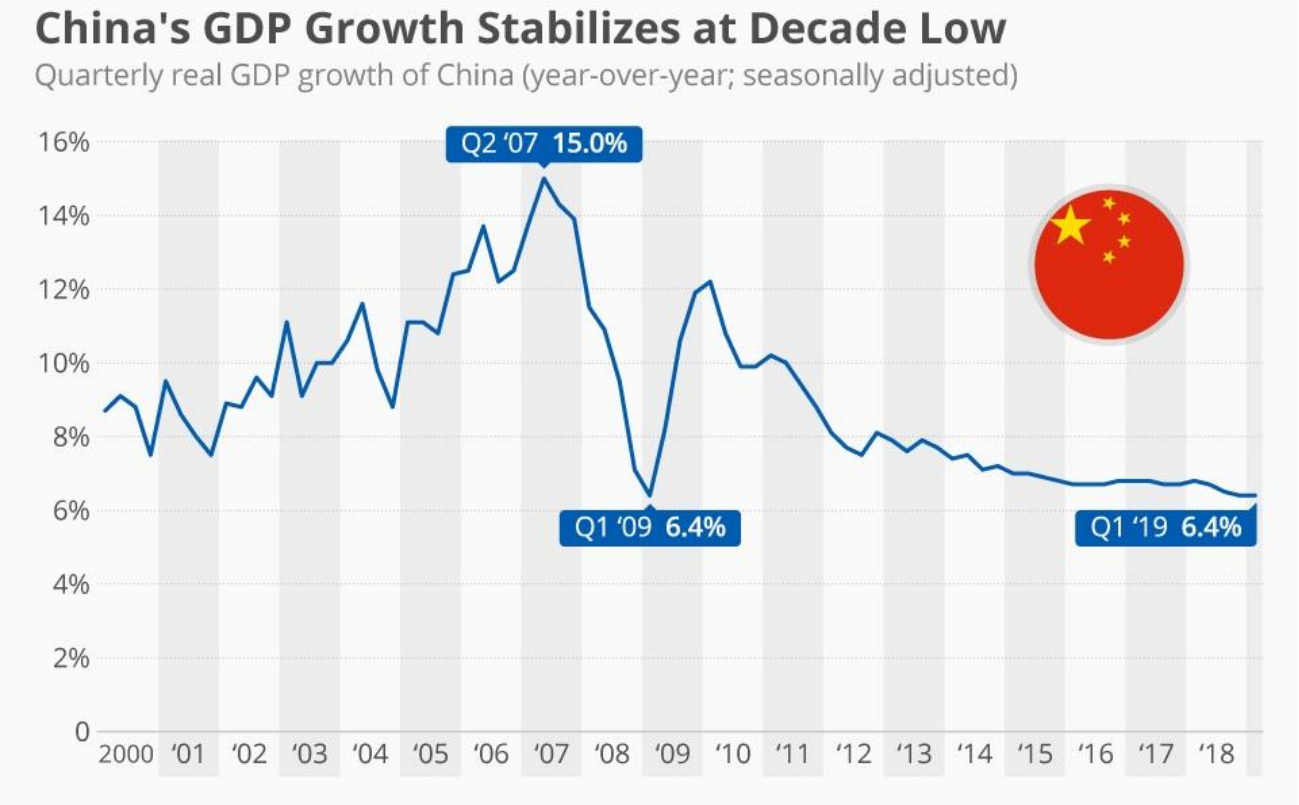
Data Exploration & Visualization

Module 10: Temporal Visualization

- Time series data
  - Properties, tasks, taxonomy
- Temporal visualization
  - Linear times: Line charts, stacked graph
  - Interval times:
    - Cyclic: calendar view, cyclic layout
    - Linear: Gantt charts, lifelines
- Case study

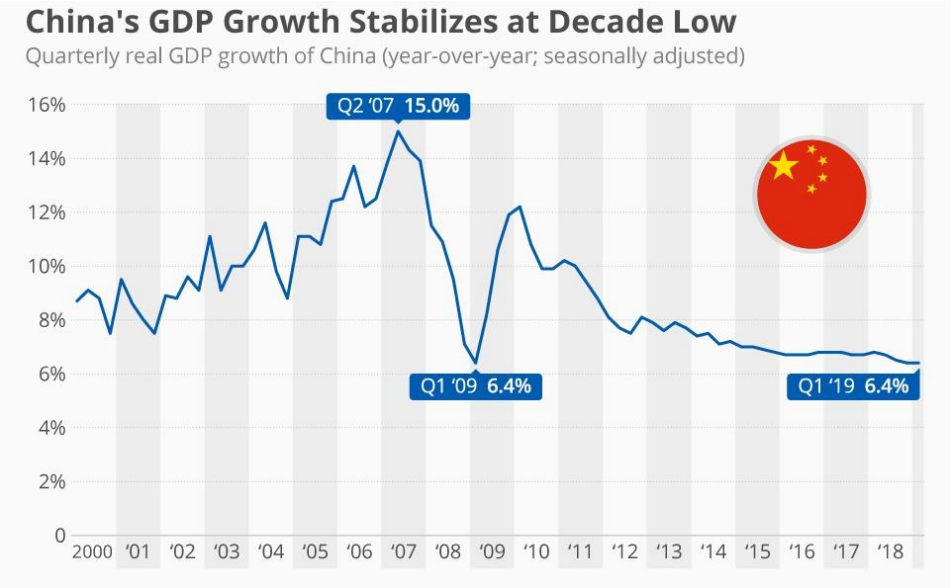
Standard presentation

- Present time data as a 2D line graph with time on x-axis and some other variable on y-axis
  - Focus here is measuring some value over time



Standard presentation

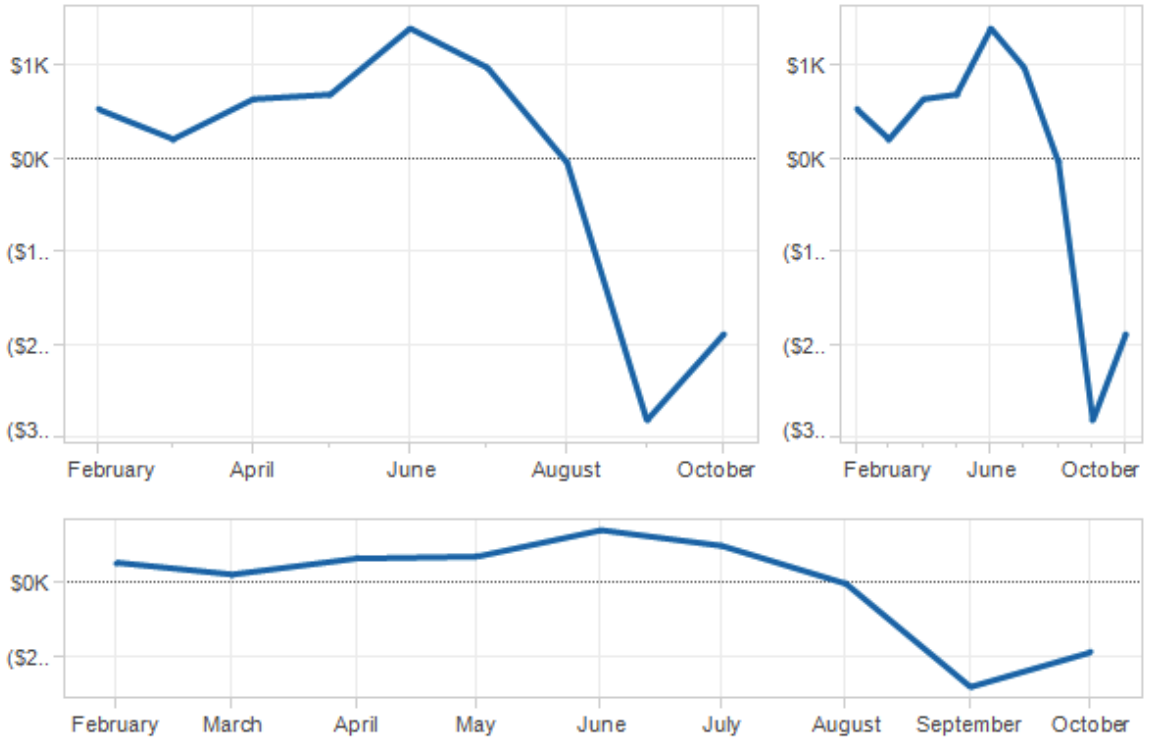
- Data
  - 2 quantitative attributes
  - one key, one value
- Mark
  - Points
  - Line connects marks between them



- Channels
  - Aligned lengths to express quantitative value
  - Separated and ordered by key attribute into horizontal regions
- Tasks
  - Find trend
- Scalability
  - hundreds of key and value levels

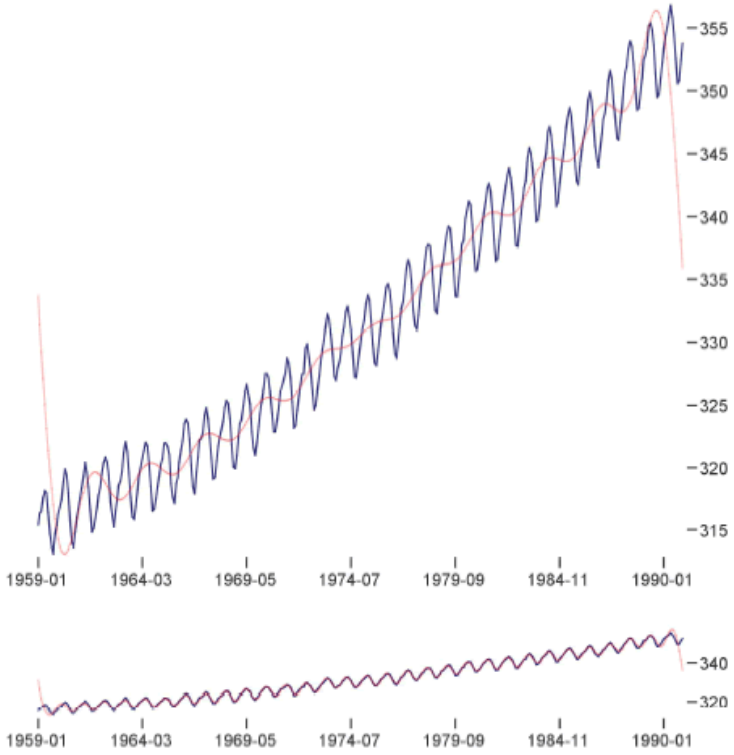
Standard presentation

- The same data can look very different
  - Aspect ratio influences the perception of the data
  - Banking to 45 degree



Standard presentation

- What is the optimal aspect ratio?
  - Banking to 45 degrees [Cleveland, 1988]
  - Multi-scale banking to 45 degrees [Heer & Agrawala, 2006]
  - Local orientation resolution [Wang et al., 2018]



$$\sum_i \frac{|\theta_i(\alpha)|}{n} = 45^\circ$$

where α : aspect ratio of the chart

$\theta_i(\alpha) = tan^{-1}(s_i/\alpha)$

s_i: a line segment

⇩ Take line length into account

$$\frac{\sum_i |\theta_i(\alpha)| l_i(\alpha)}{\sum_i l_i(\alpha)} = 45^\circ$$

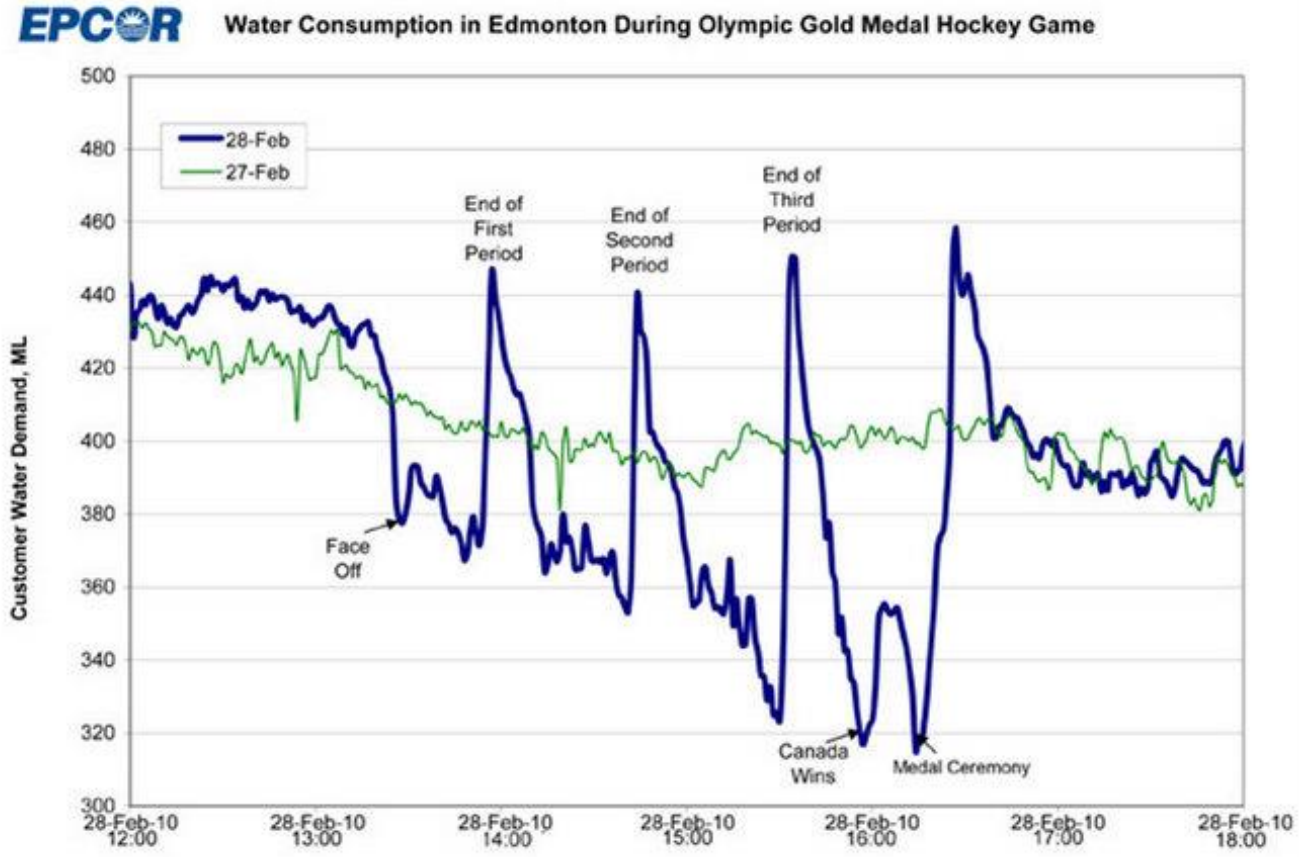
Data for line chart

- What are line charts presenting?
  - 2 quantitative attributes: one key, one value
  - Aligned lengths to express quantitative value
  - Separated and ordered by key attribute into horizontal regions
    - Discrete or interval?
    - Linear or cyclic?
    - Ordinal or quantitative?

- What if there are multiple values to track?

Multiple lines

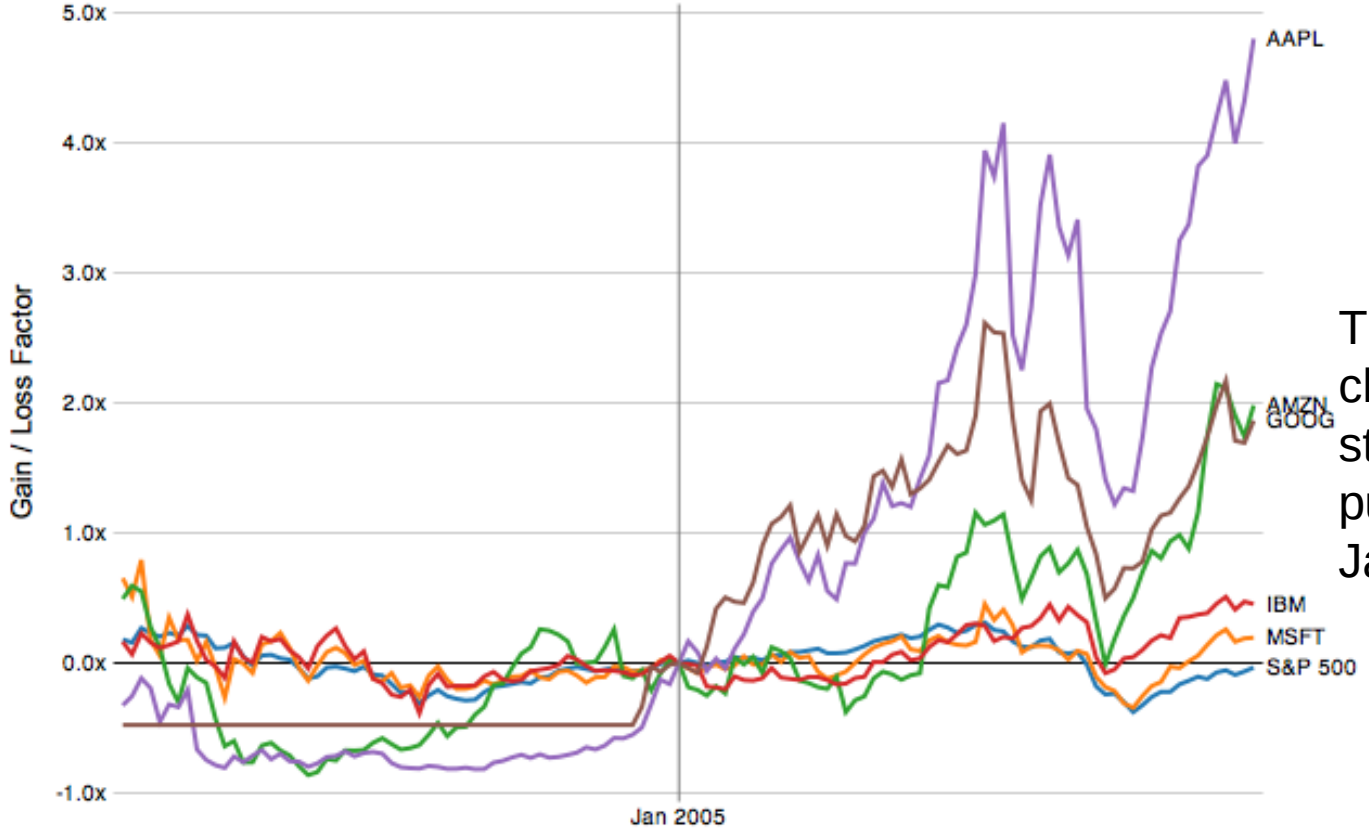
- What If Everybody in Canada Flushed At Once?



https://www.patspicks.com/blog/item/what_if_everybody_flushed_at_once_Edmonton_water_gold_medal_hockey_game/

Multiple lines

- In some cases, raw values are less important than relative changes.

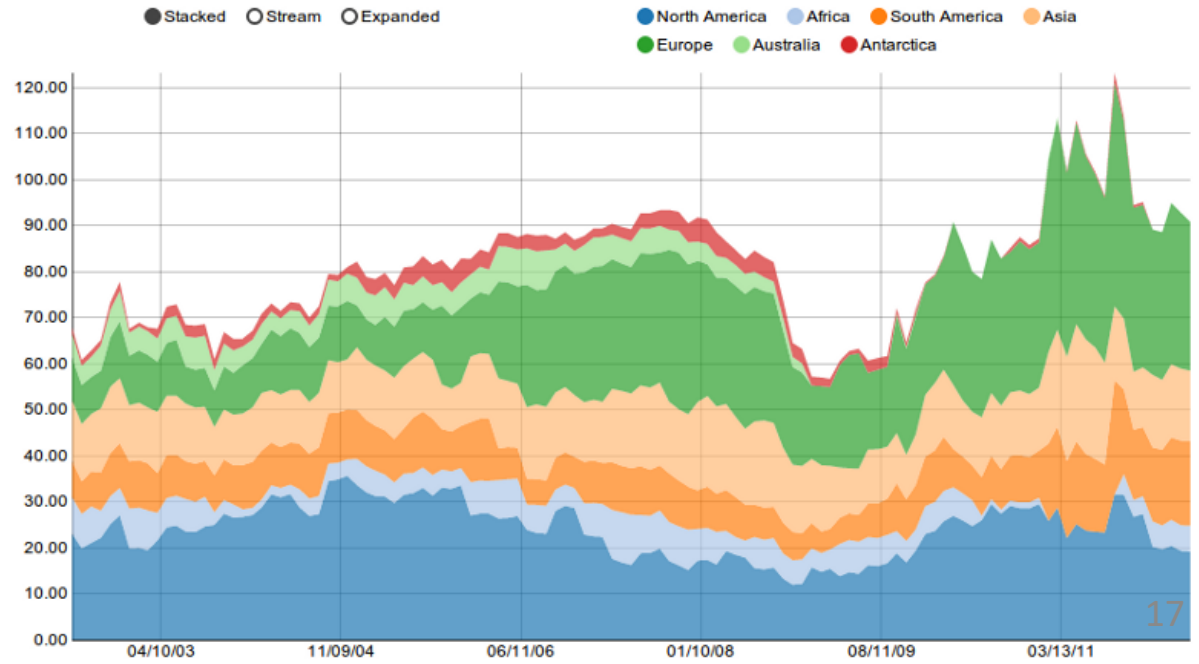


The percentage change of selected stock prices if purchased in January 2005.

Heer et al., 2010

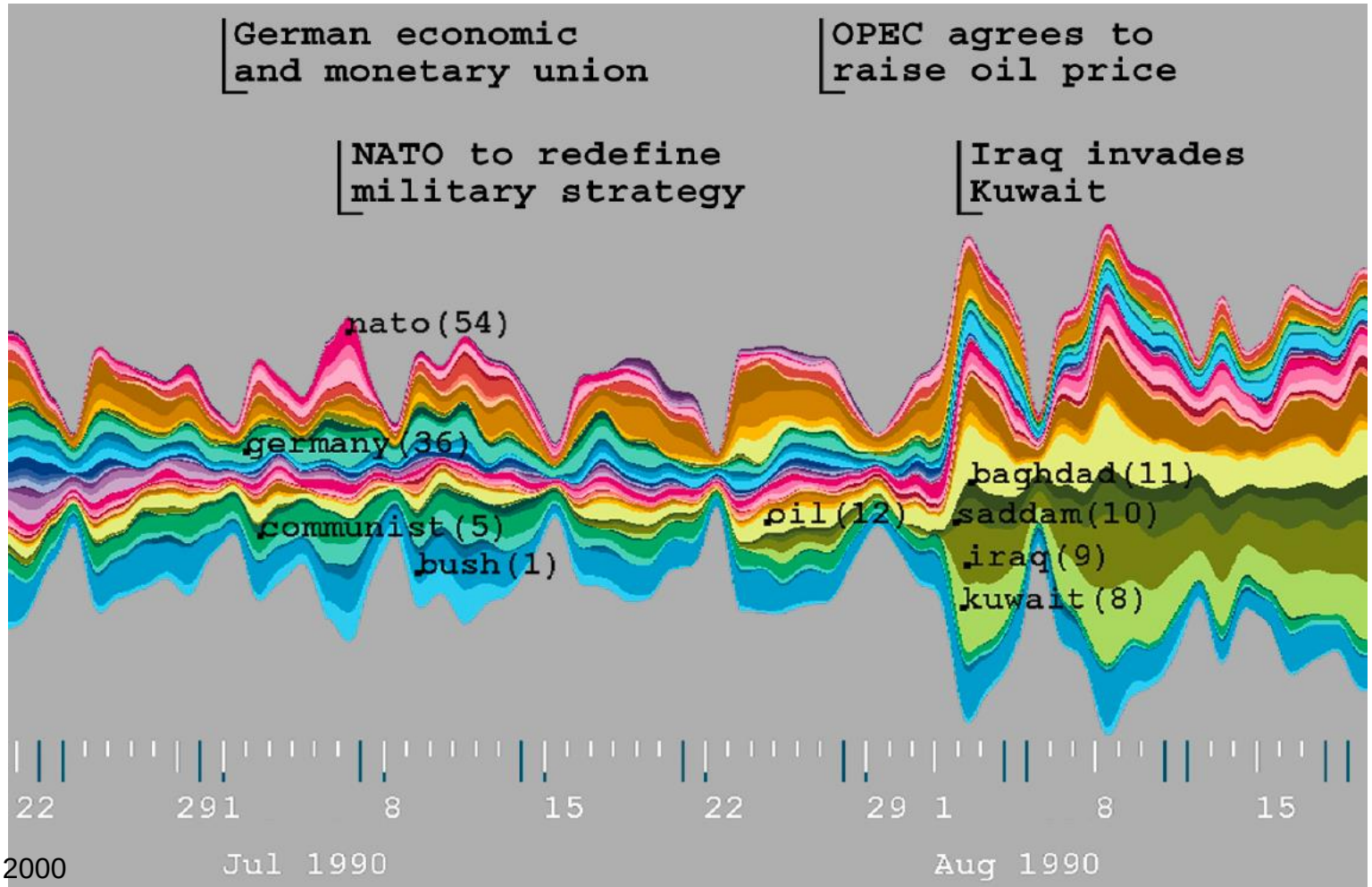
Stacked graph

- Data
  - 1 categorical key attribute
  - 1 ordered key attribute (usually time)
  - 1 quantitative value attribute
- Emphasizing horizontal continuity and part-whole relationship



Streamgraph

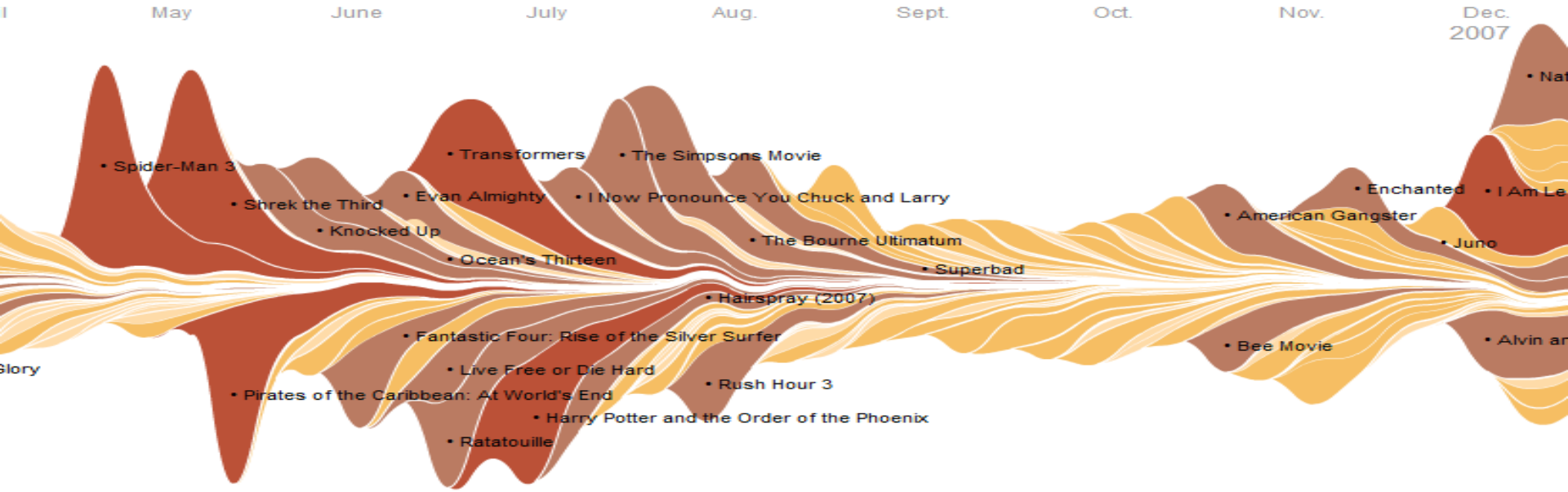
- ThemeRiver



Havre et al., 2000

Streamgraph

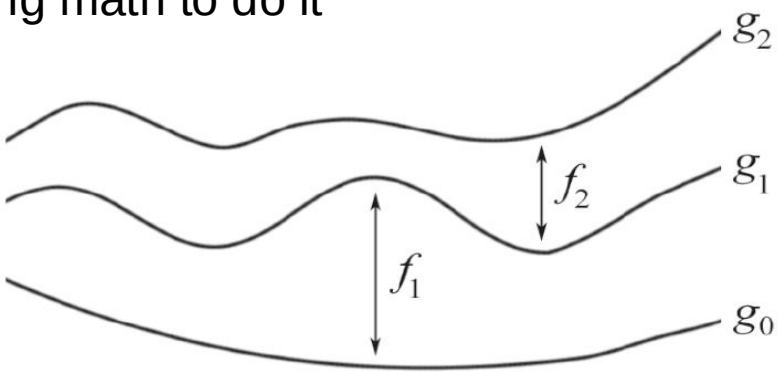
- Generalized stacked graph
- Scalability
  - Hundreds of time keys
  - Dozens to hundreds of categories



Byron and Wattenberg, 2008

Streamgraph

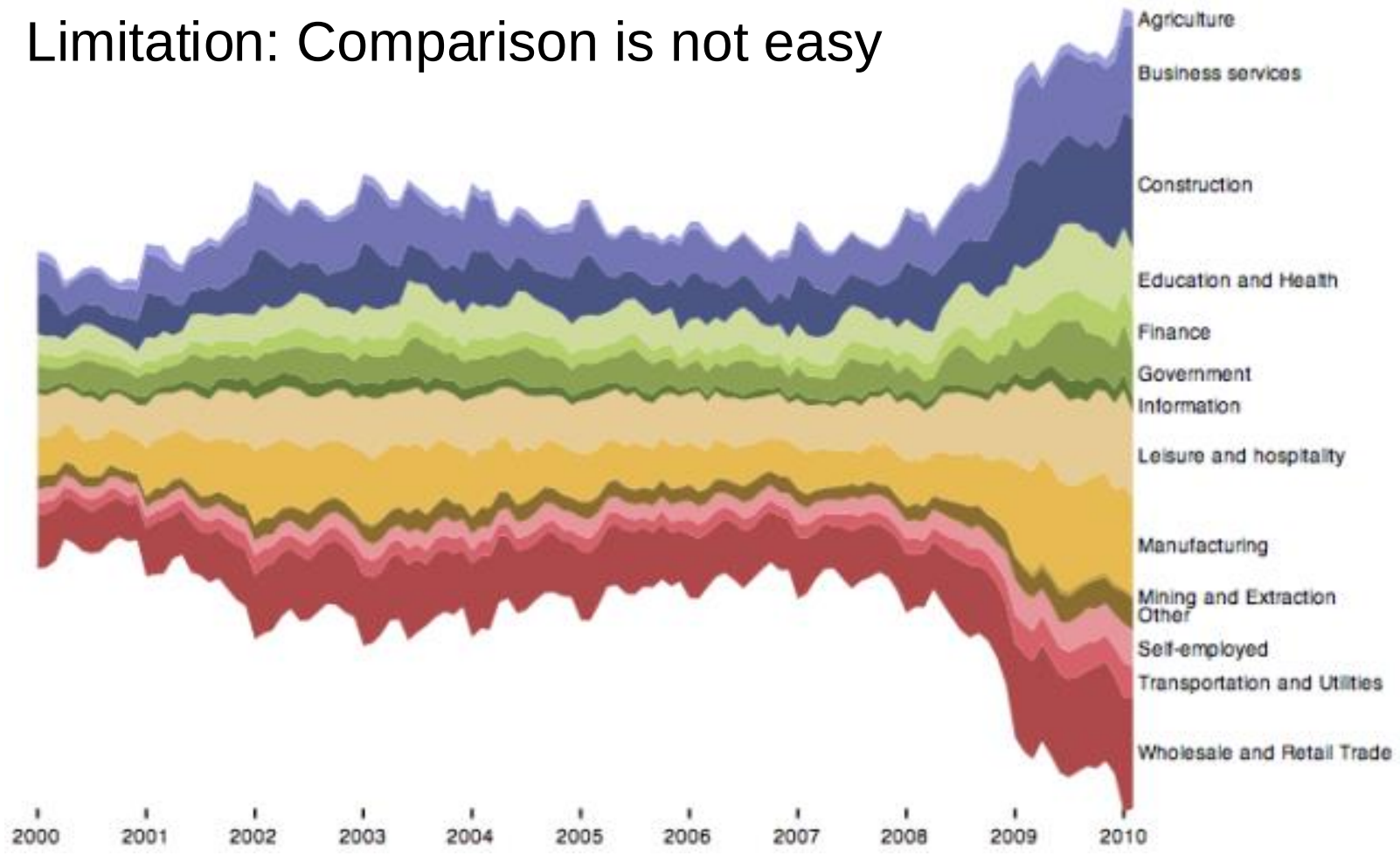
- Design issues
  - Curve shape
    - Wiggle, symmetry, balance
    - Some interesting math to do it
  - Color choice
  - Labeling
  - Layer ordering



Byron and Wattenberg, 2008

Streamgraph

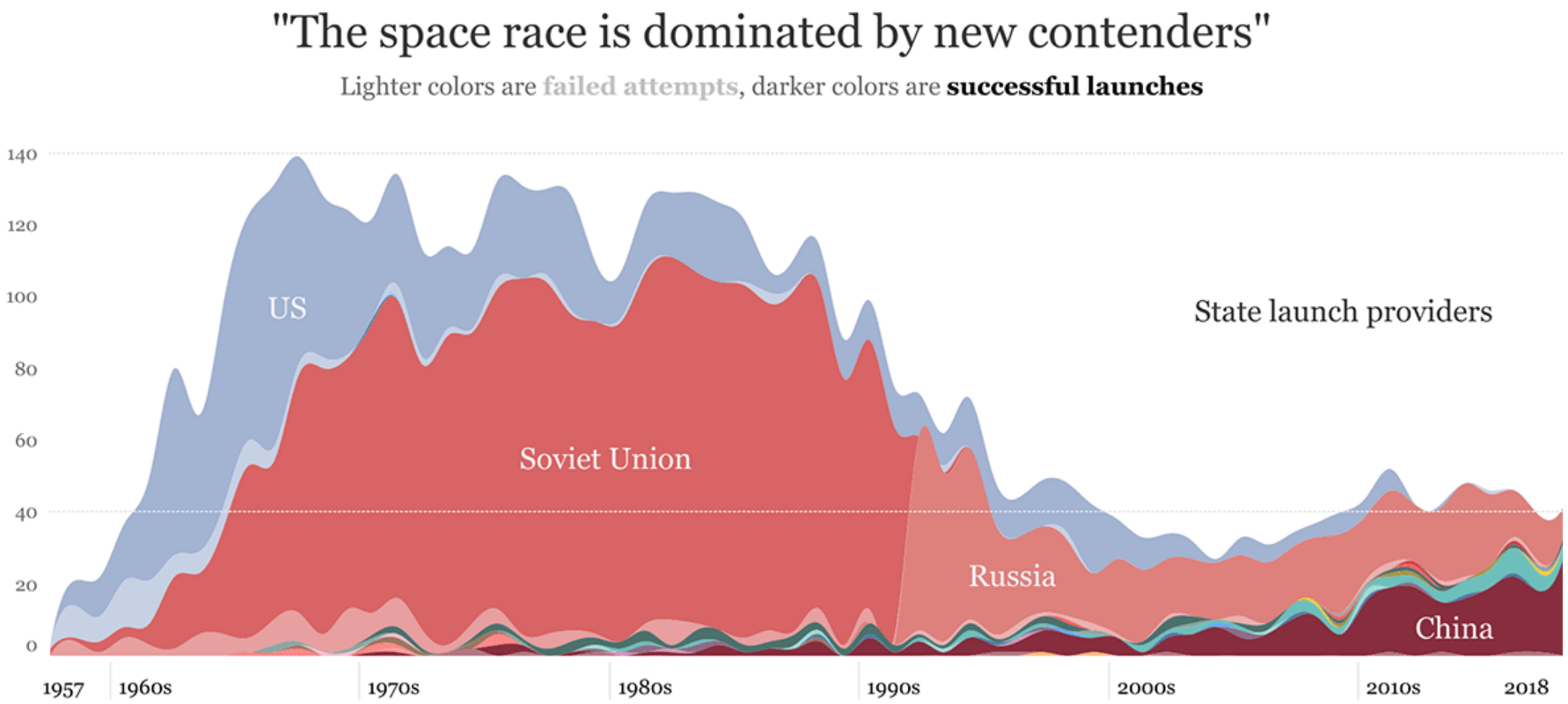
- Limitation: Comparison is not easy



Heer et al., 2010

Streamgraph

- Limitation: Comparison is not easy
  - Align the baseline
  - Sort the order



Different data

- What if you want to show two continuous variables over time?
  - And not just use two lines

Gapminder

- Bubble chart + animation
  - Strengths?
  - Weaknesses?

Alternative design

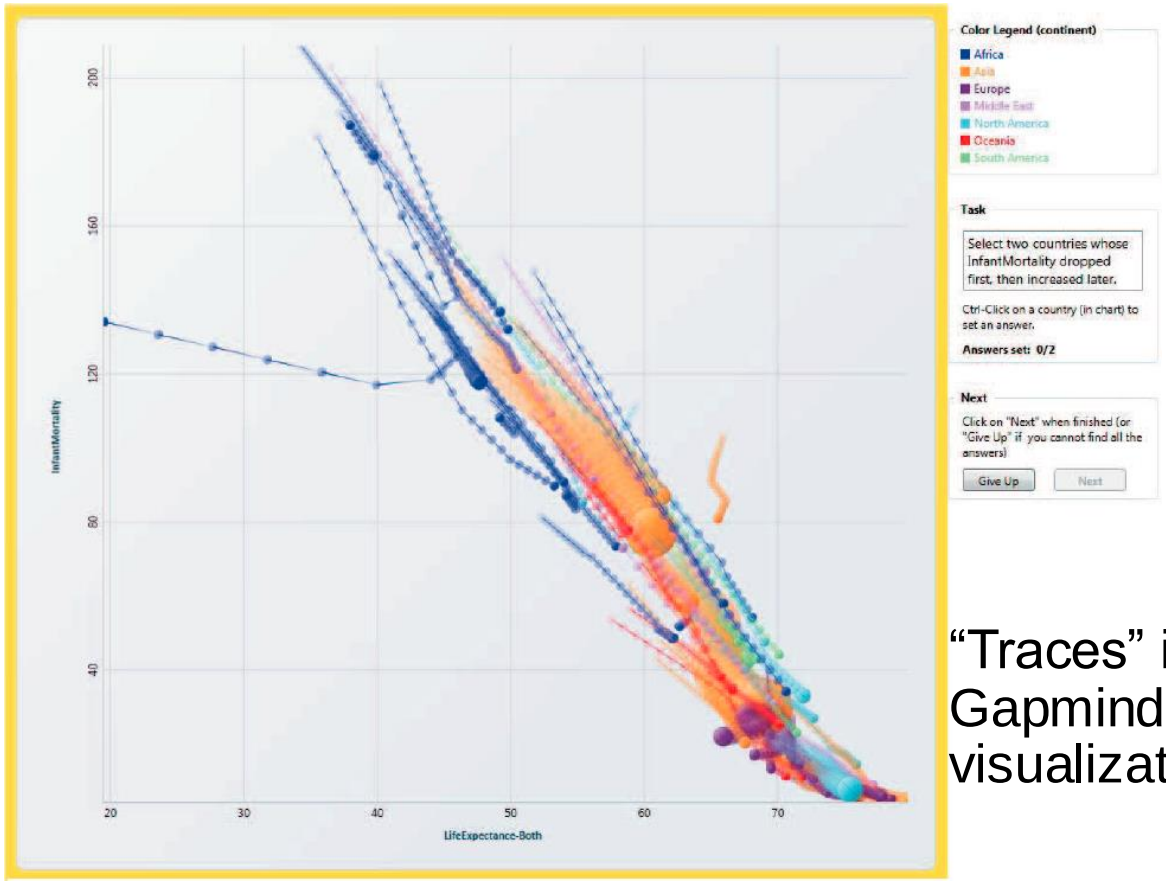
- How do we address weakness?
  - **Presentation**: while animation it is the fastest technique for presentation and participants find it enjoyable and exciting, it does lead to many participant errors.
  - **Analysis**: both static depictions of trends are significantly faster than animation, and the small multiples display is more accurate.

- How to get ride of time slider?

Robertson et al., 2008

Connected scatterplots

- Visualize two related time series in a scatterplot and connects the points with a line in temporal sequence.



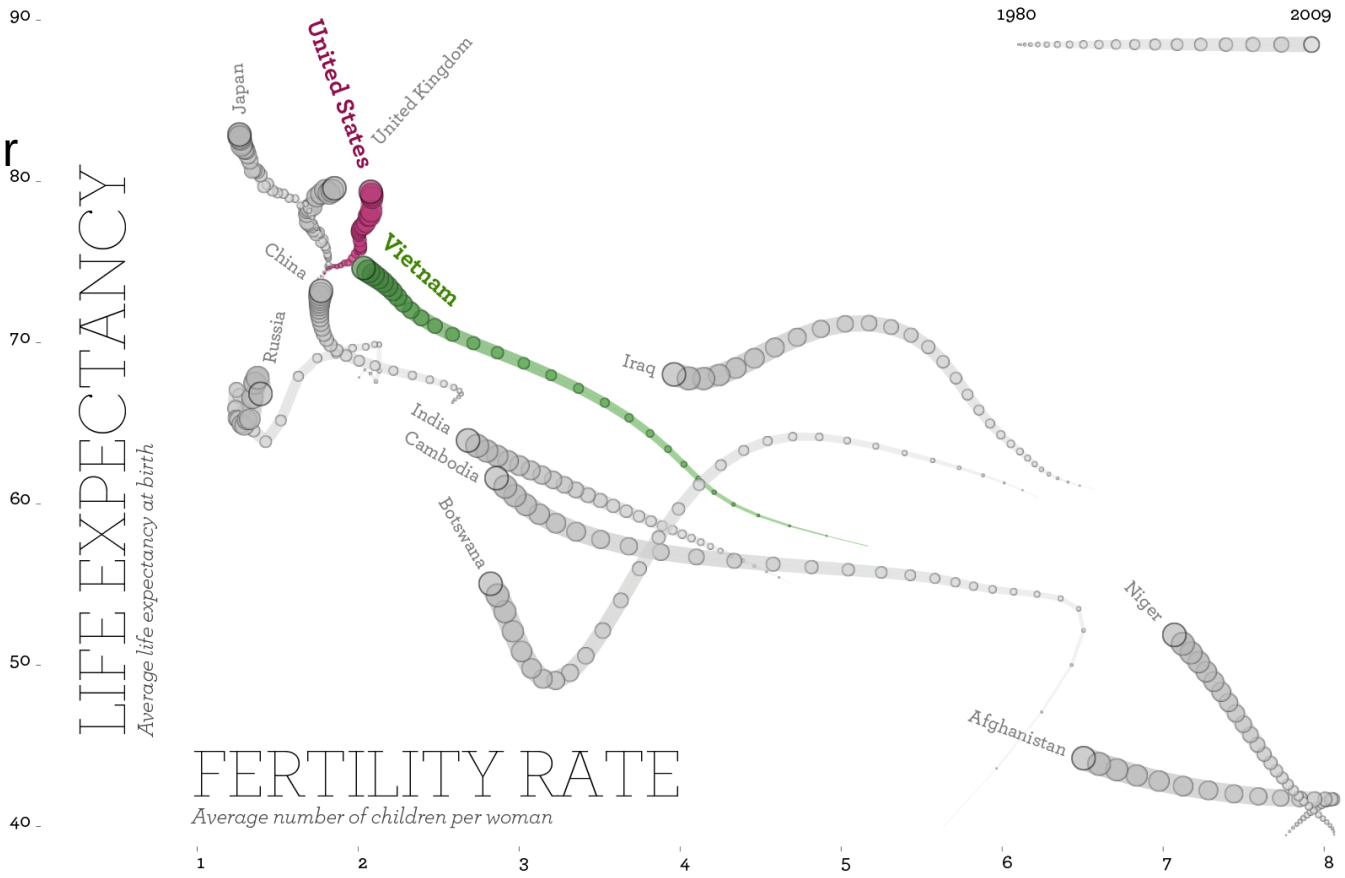
"Traces" in Gapminder-style visualization

Robertson et al., 2008

Connected scatterplots

Moritz Stefaner

Remixing Rosling

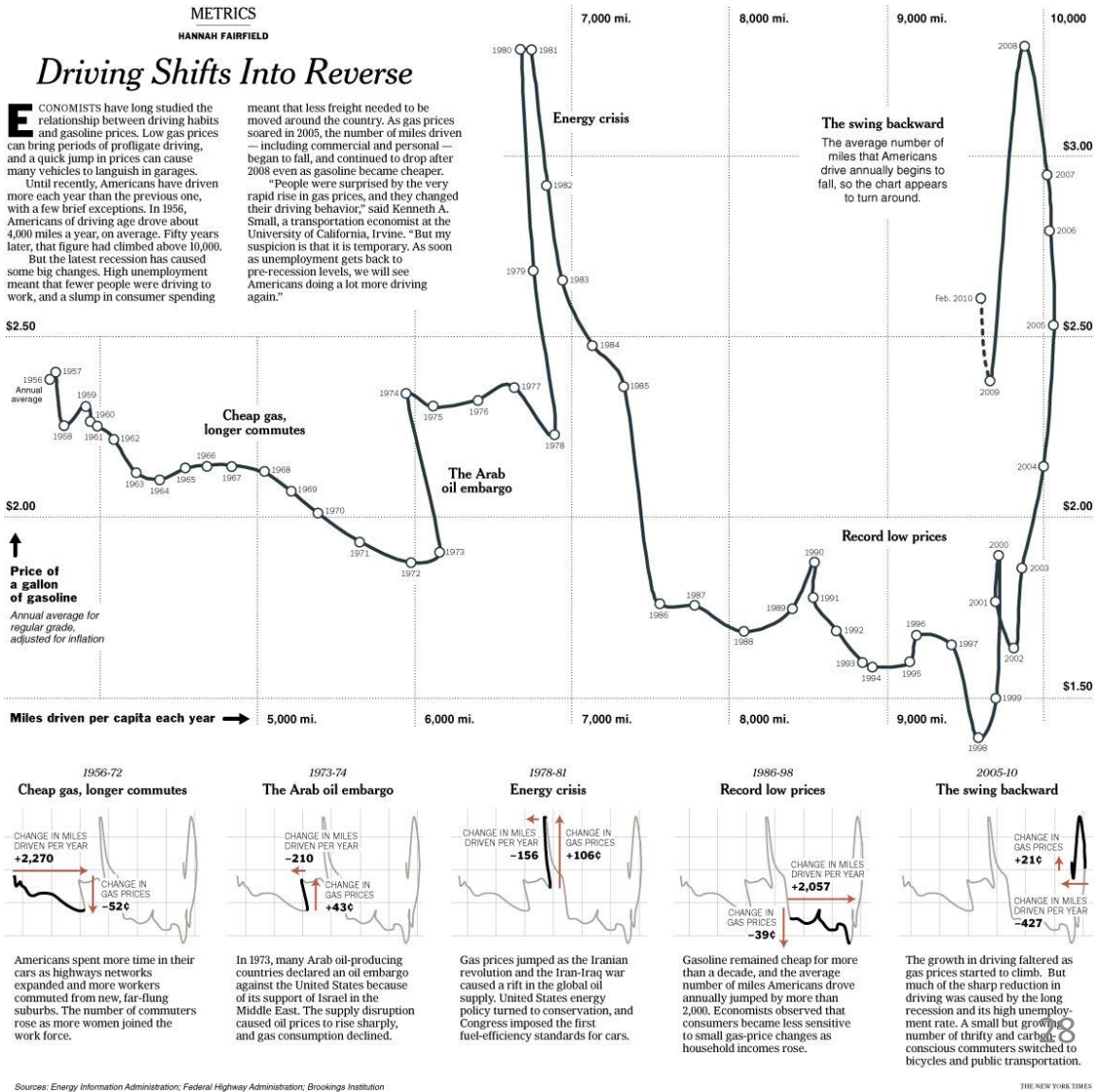


http://truth-and-beauty.net/projects/remixing-rosling/

Connected scatterplots

Hannah Fairfield

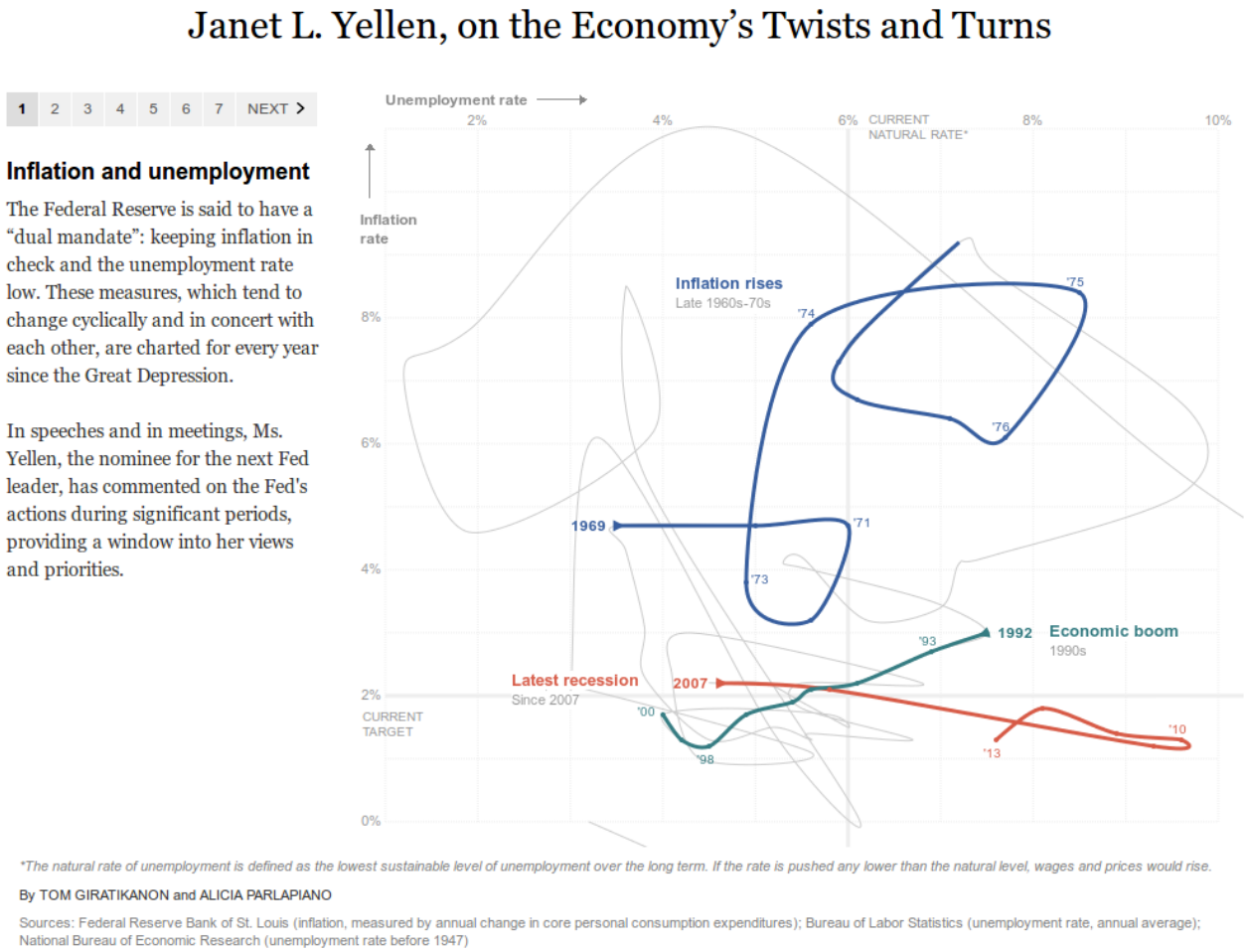
Relationship between driving habits and gasoline prices



Connected scatterplots

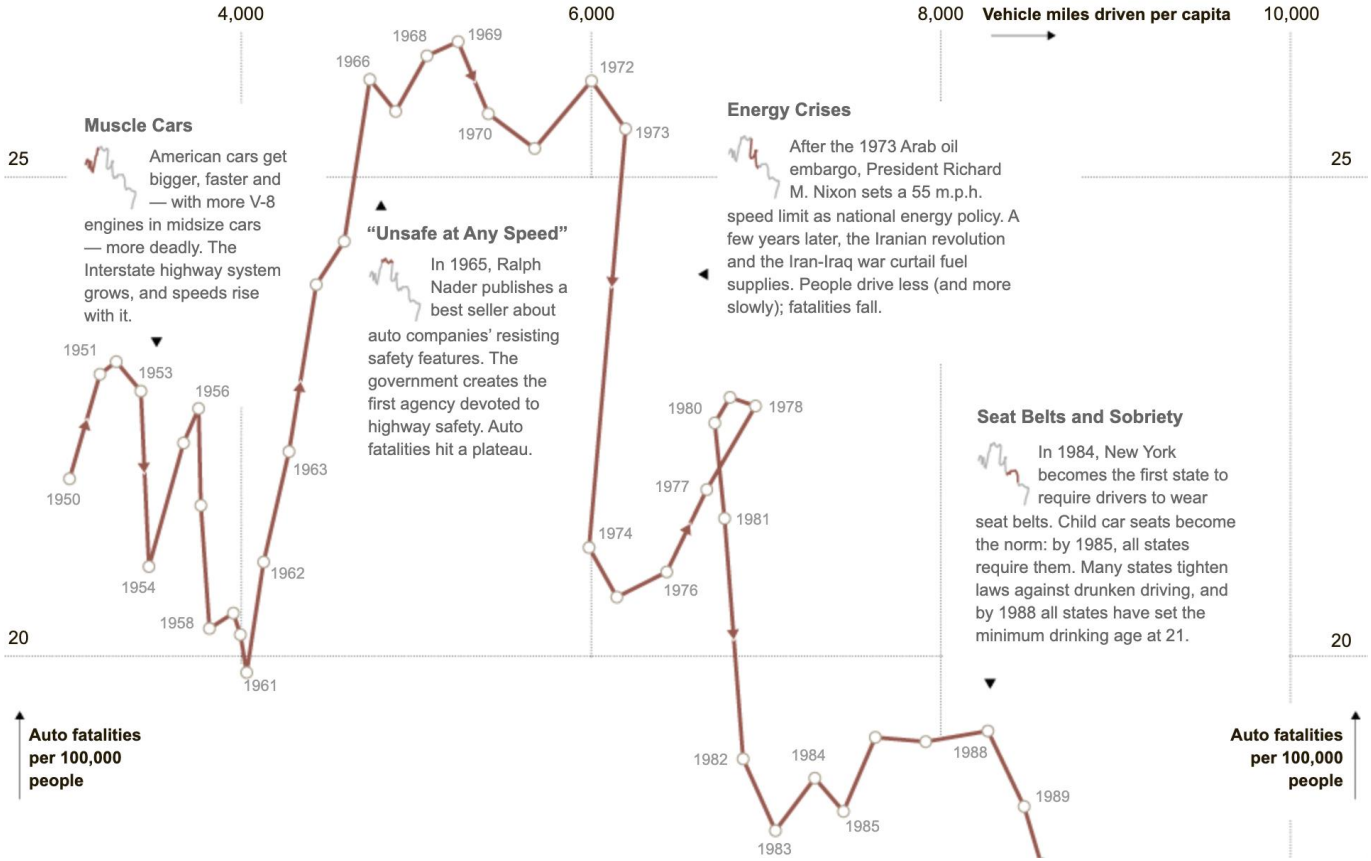
Janet L. Yellen

The measures to keep inflation in check and the unemployment rate low since 1941.



Connected scatterplots

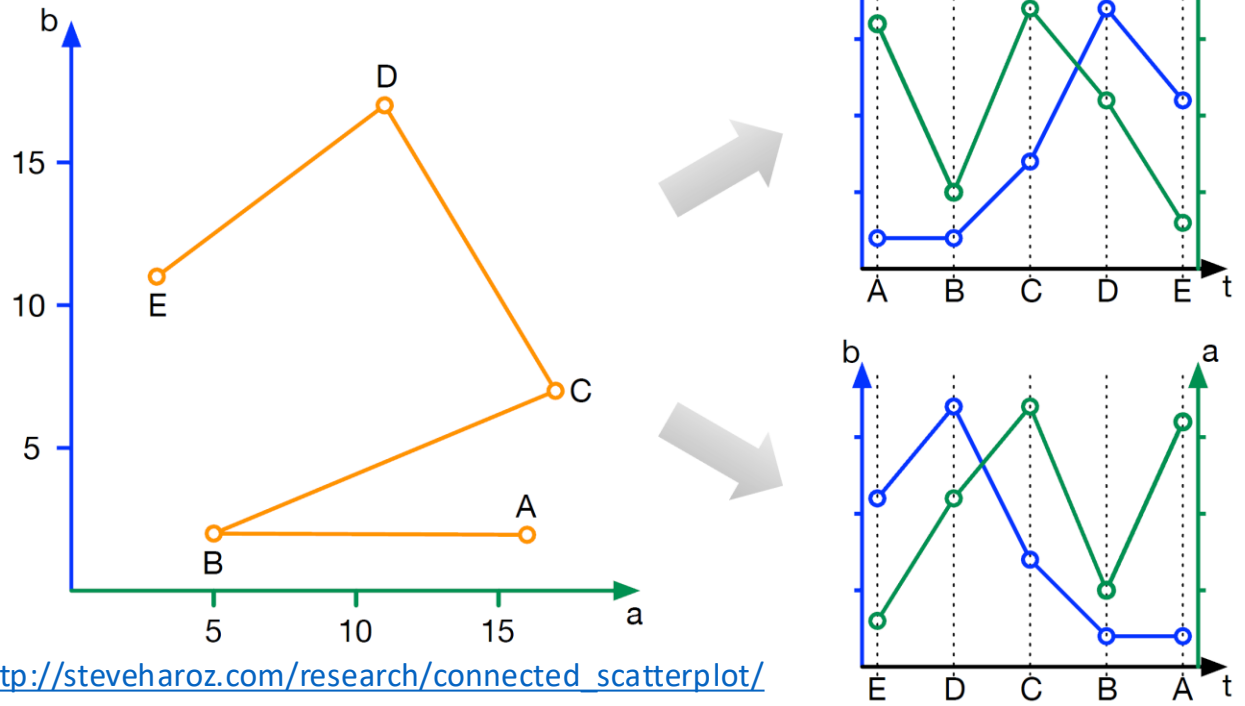
Driving safety, in fits and starts



New York Times

Connected scatterplots

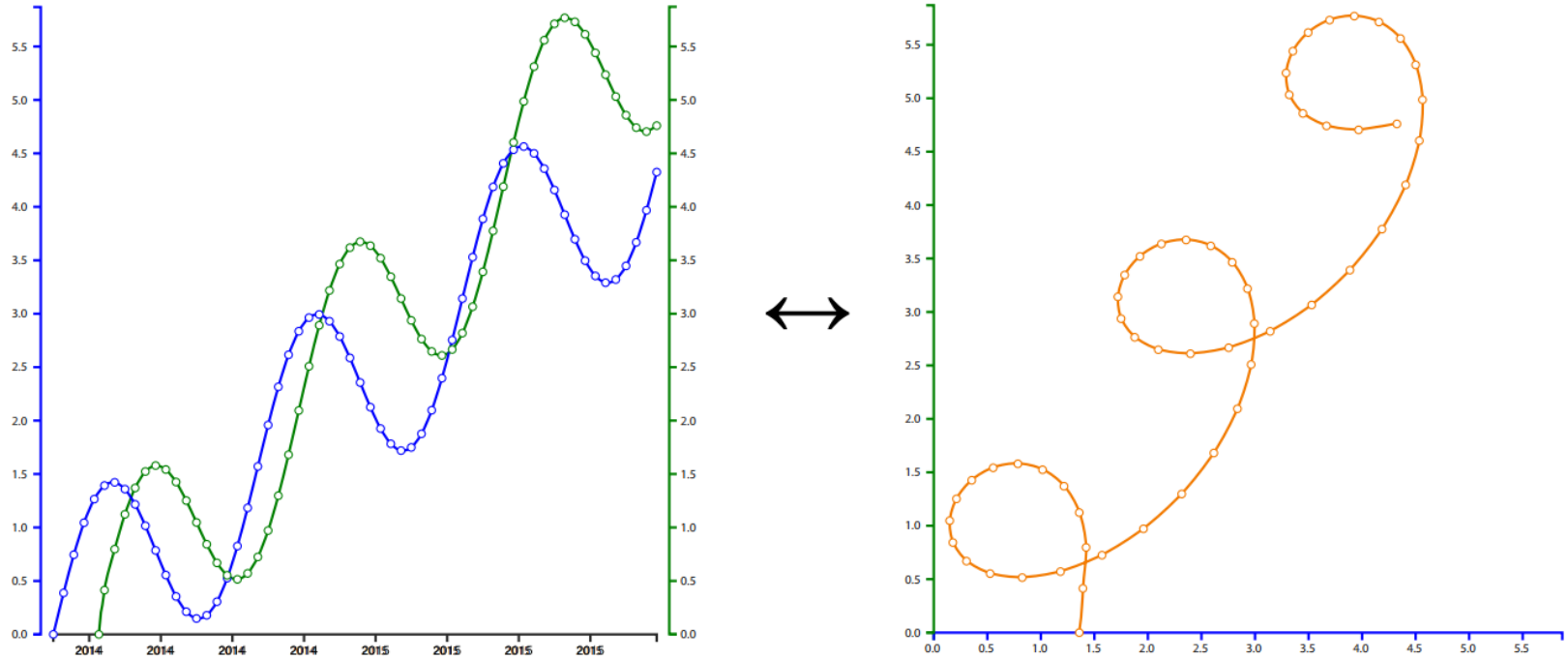
- While connected scatterplots is an effective tool for journalist, it has certain limitations
  1. Viewers can confuse order and direction.
  2. Viewers are less likely to report correlation



Haroz et al., 2016 http://steveharoz.com/research/connected_scatterplot/

Connected scatterplots

- While connected scatterplots is an effective tool for journalist, it has certain limitations
  3. Known concepts have new visual features



Haroz et al., 2016 http://steveharoz.com/research/connected_scatterplot/

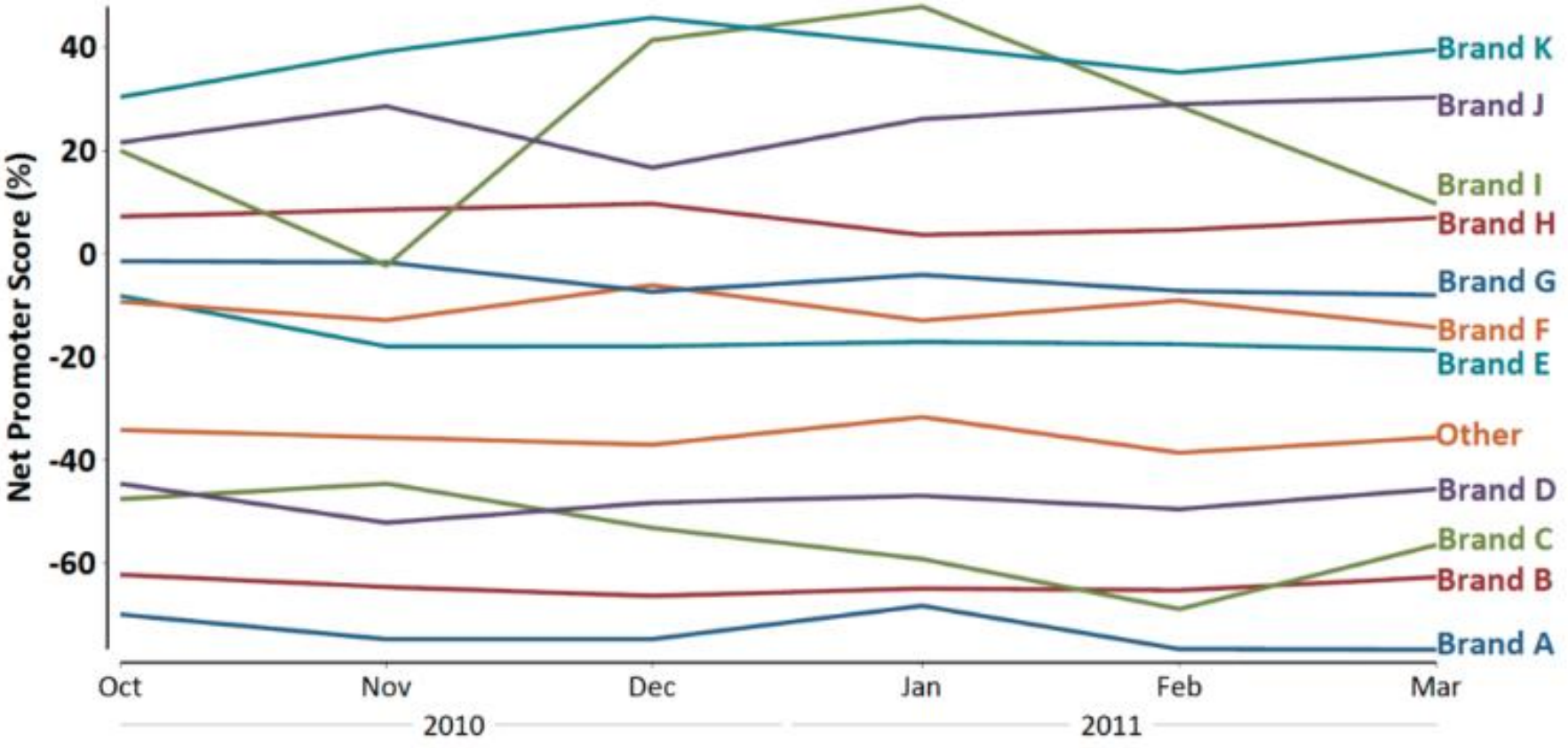
Small multiples

Small multiples visualization shows trace lines for each country separately.

Robertson et al., 2008

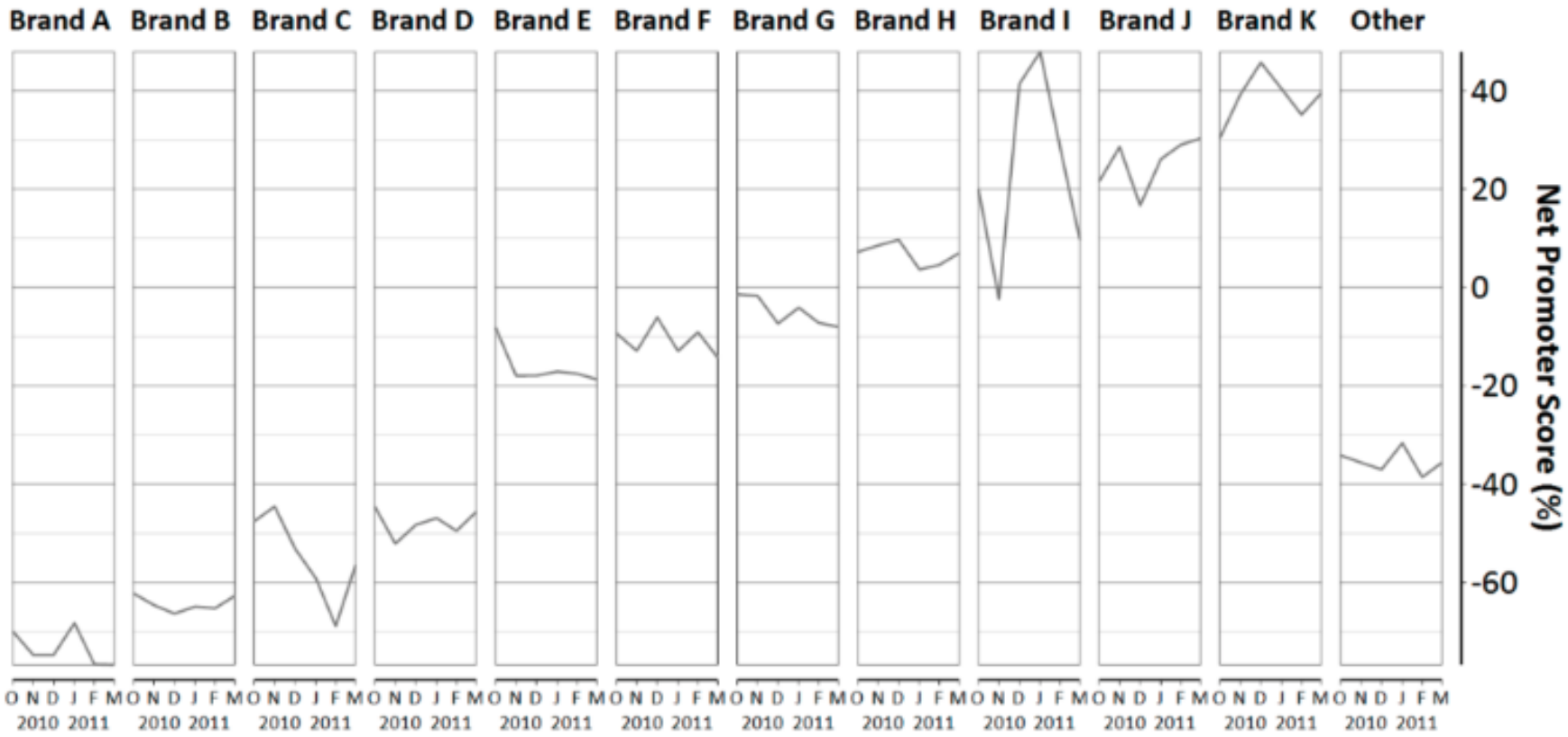
Small multiples

- Showing each series in its own chart.



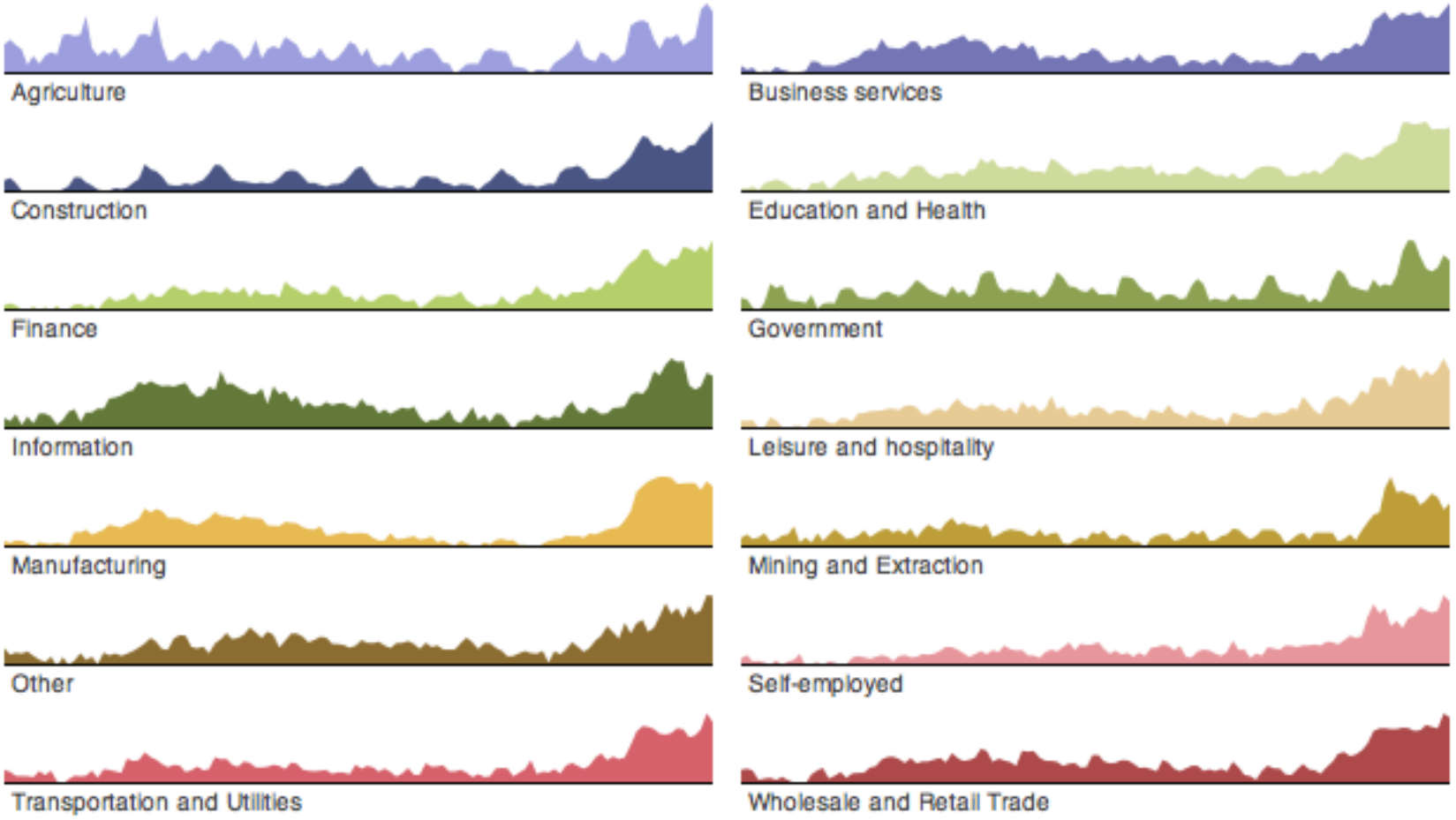
Small multiples

- Showing each series in its own chart.
  - The individual graphics have been ordered to make it easy to see the average difference between the brands.



Small multiples

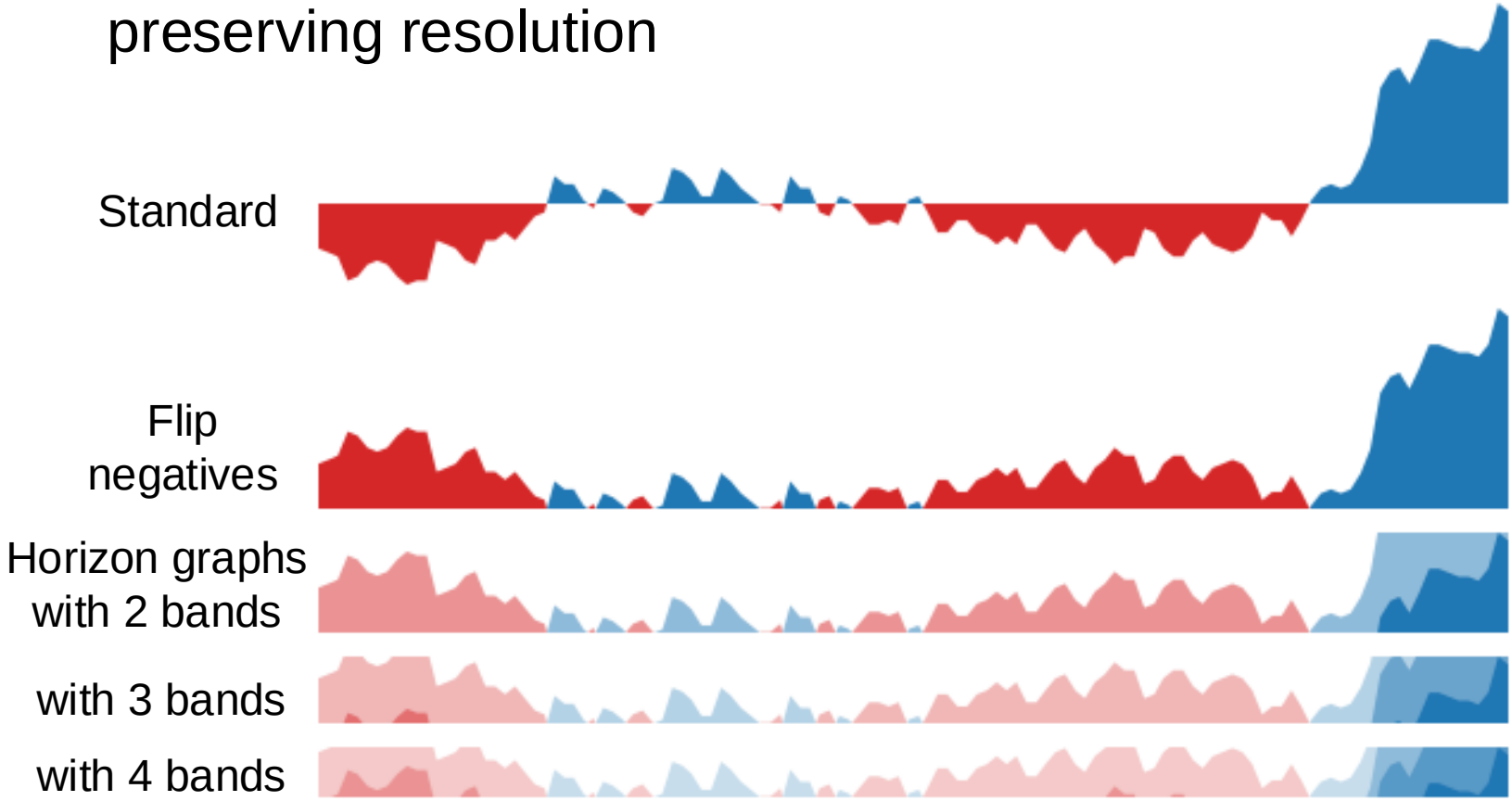
- Showing each series in its own chart.



Heer et al., 2010

Horizon graphs

- Increasing the data density of a time-series view while preserving resolution



Heer et al., 2009

Fundamental tradeoff

- Is the visualization time-dependent, i.e., changing over time (beyond just being interactive)
  - Static
    - Shows history, multiple perspectives
    - Allows comparison
  - Dynamic (animation)
    - Gives feel for process & changes over time, has more space to work with

Data Exploration & Visualization

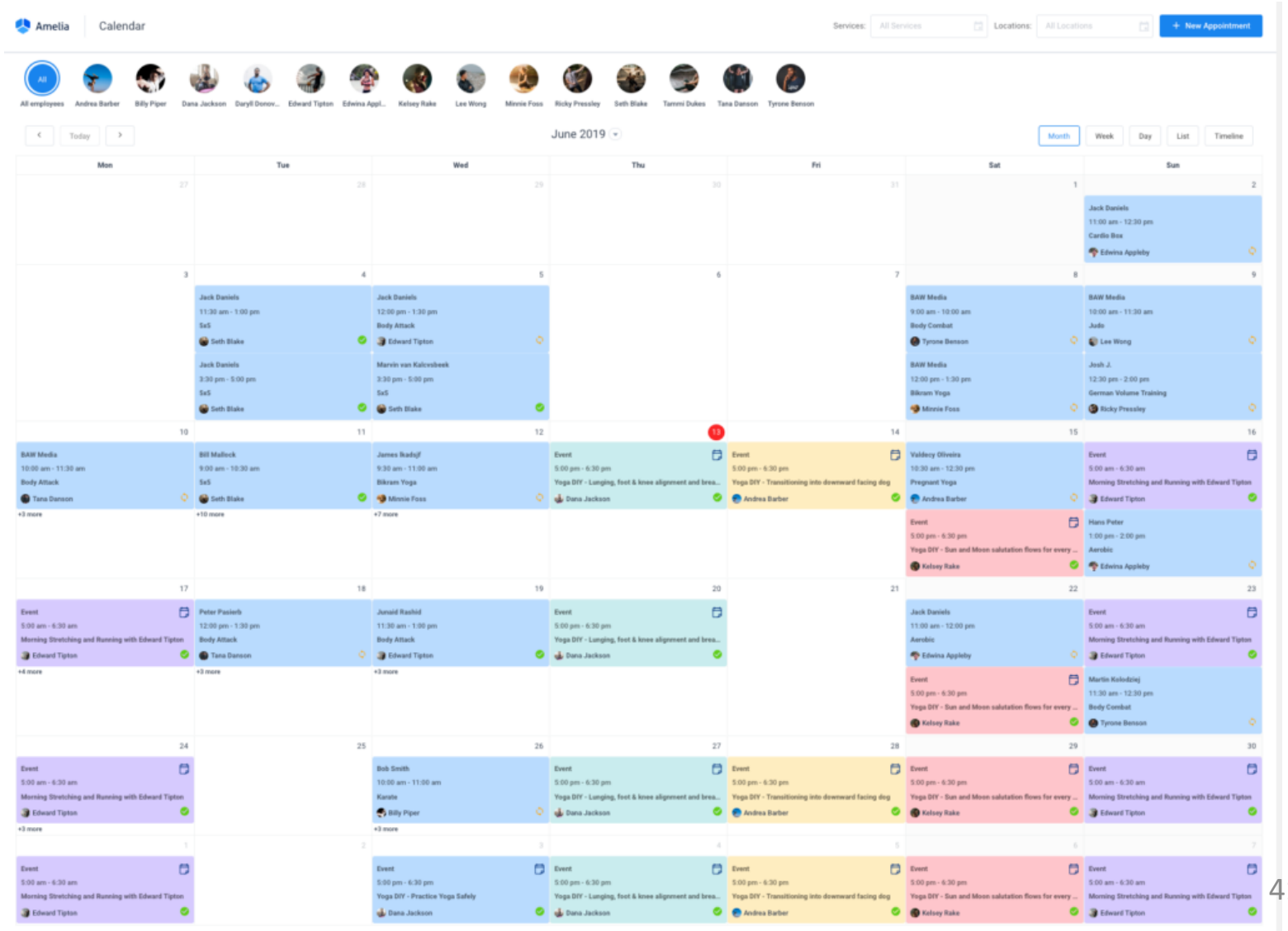
Module 10: Temporal Visualization

- Time series data
  - Properties, tasks, taxonomy
- Temporal visualization
  - Linear times: Line charts, stacked graph
  - Interval times:
    - Cyclic: calendar view, cyclic layout
    - Linear: Gantt charts, lifelines
- Case study

Different data

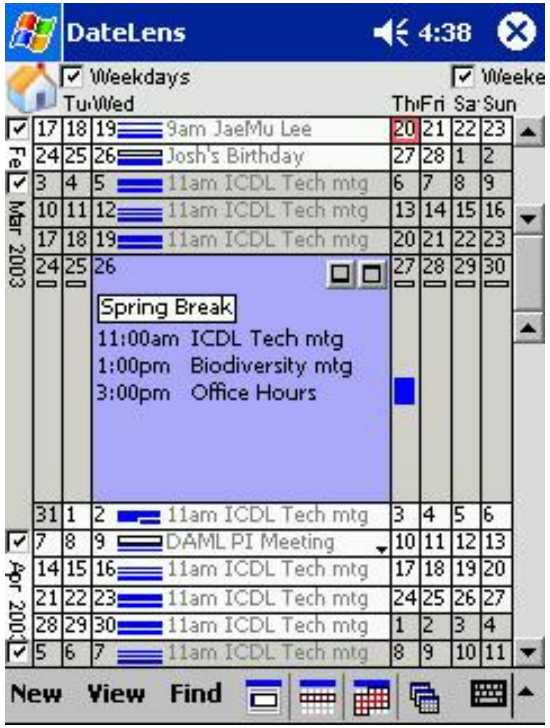
- Nominally-typed events occurring over time with **durations**
  - Interval time
- Do days/weeks/months/years matter?
  - Yes ⇒ cyclic
  - then…

Calendar view

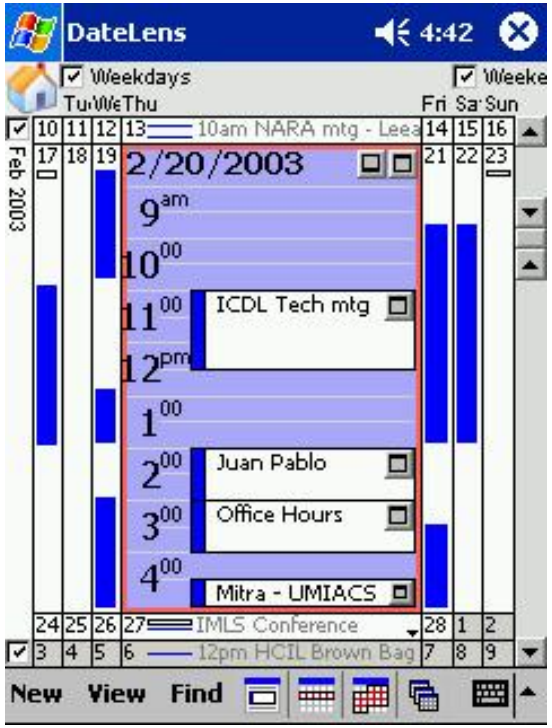


More context

- How do we see more context/overview?
  - Focus view



One day in three month



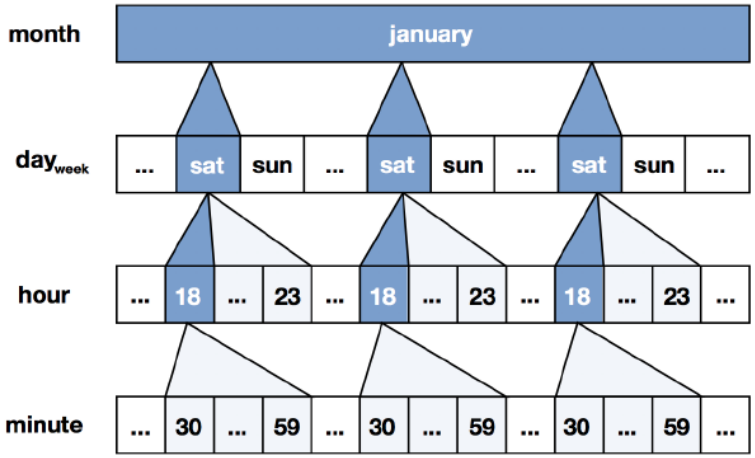
One day in a month

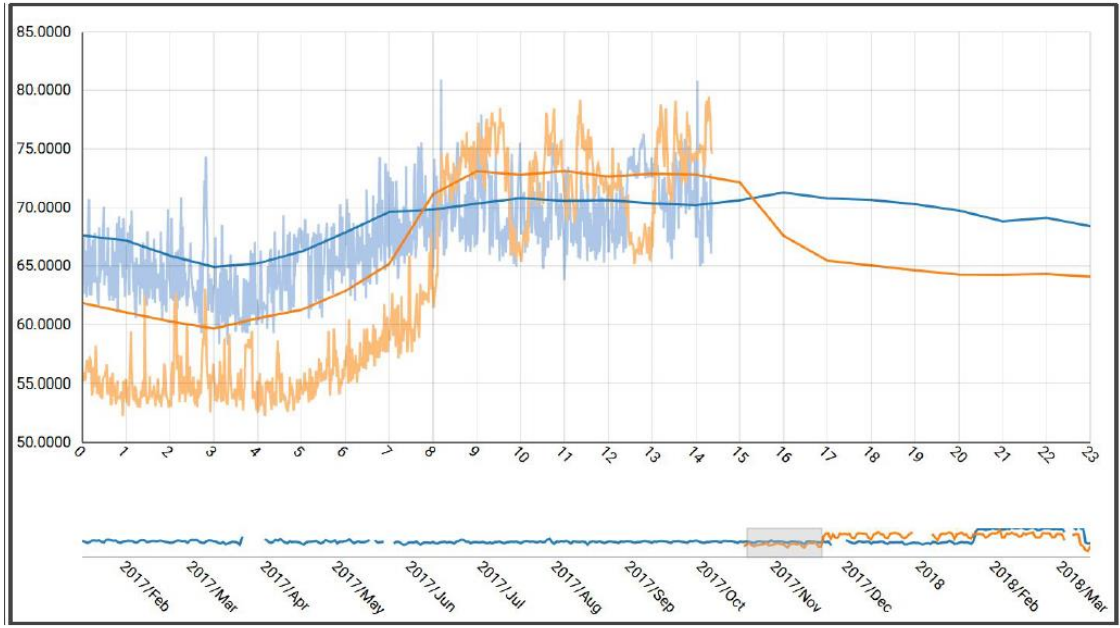
Single day

Benjamin et al., 2004

More context

- How do we see more context/overview?
  - Time lattice
  - Drill down ↔ roll up



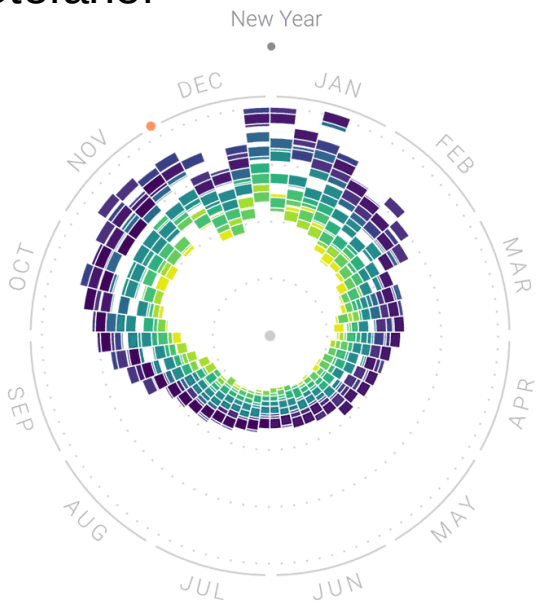


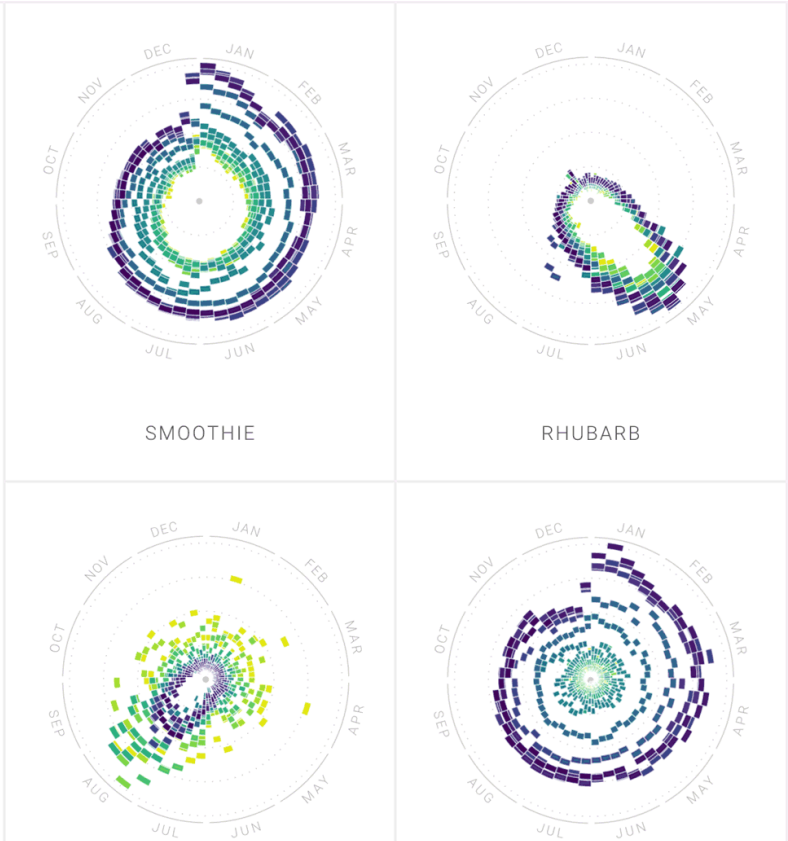
Miranda et al., 2018

Alternative design

- Cyclic layout for cyclic time series

Moritz Stefaner





The Rhythm of Food

http://truth-and-beauty.net/projects/the-rhythm-of-food

Different data

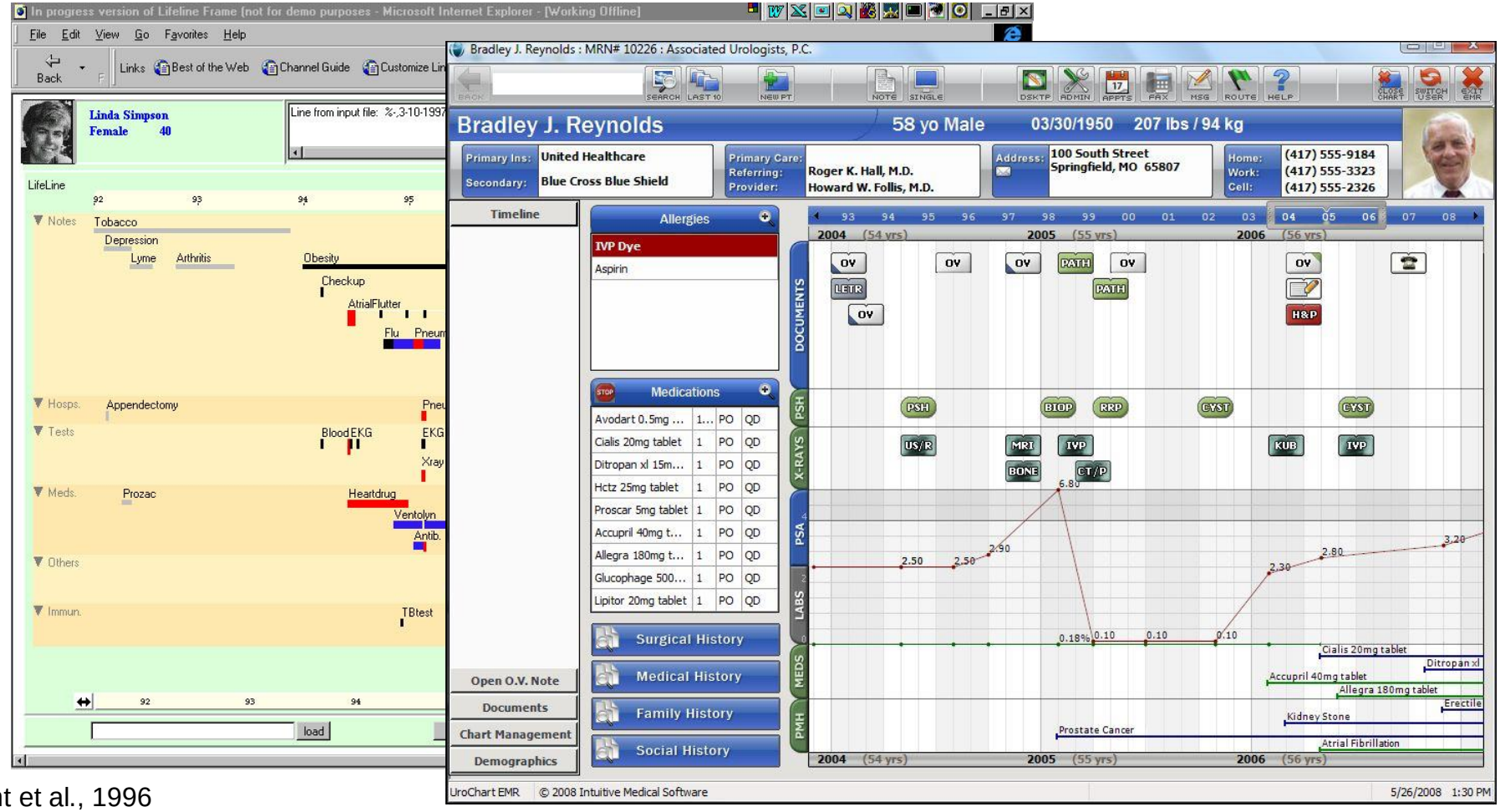
- Nominally-typed events occurring over time with **durations**
  - Interval time
- Do days/weeks/months/years matter?
  - No ⇒ linear
  - then…

Gantt charts

- Potential tasks:
  - Put together complete story
  - Garner information for decision-making
  - Notice trends
  - Gain an overview of the events

Lifelines

- Visualize personal history in some domain



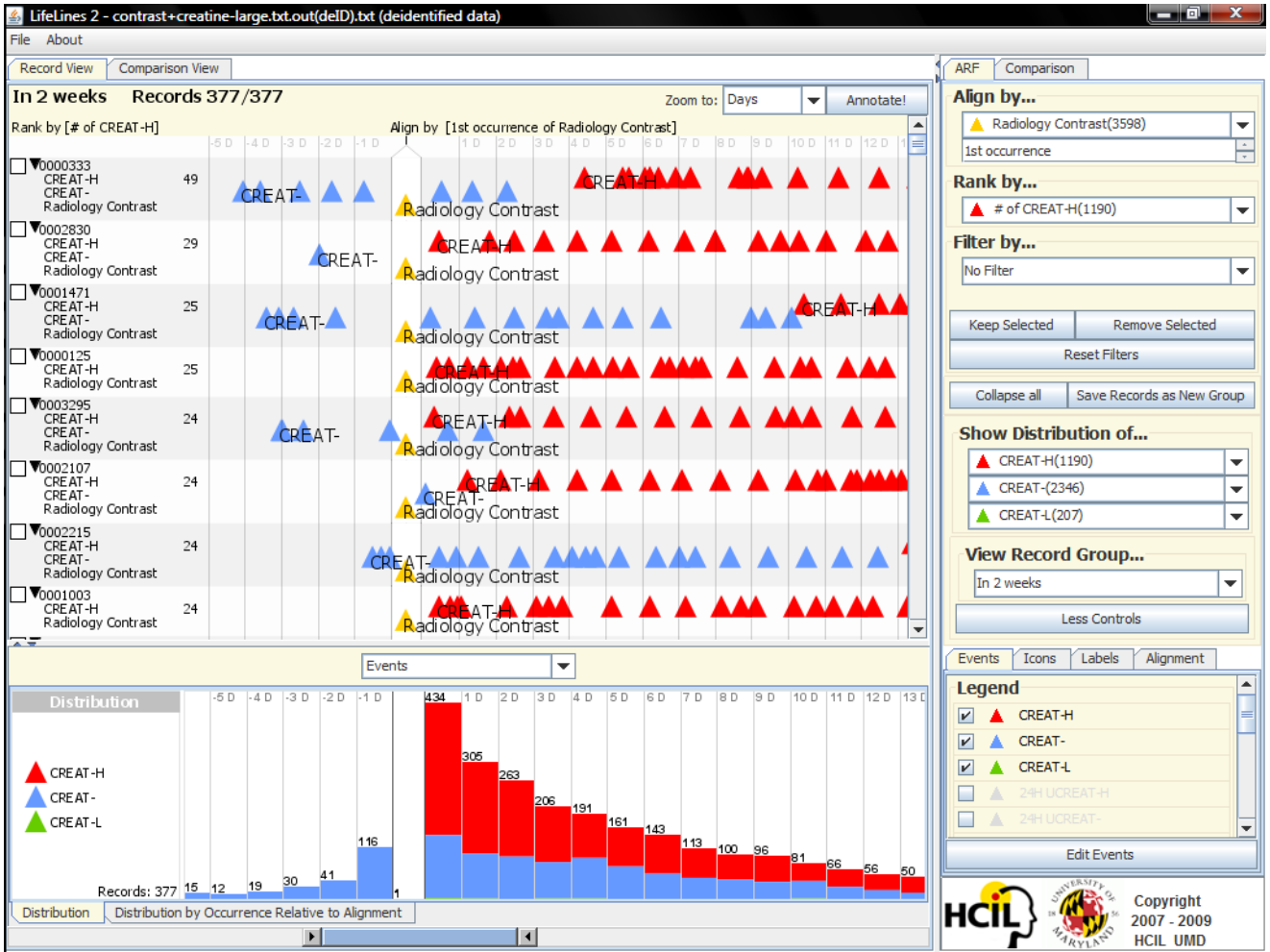
Plaisant et al., 1996

Challenges

- Scalability (could be thousands of tests)
- Can multiple records be visualized in parallel (well)?
  - Comparisons
    - What trends do you see in the last 8 EKGs?
    - Compare the 8 people who all seem to have the same problem
- Support alignment, rank, and filter
- Medical application:
  - Look for temporal coincidence of two events
    - First pneuomonia and asthma attack
  - Medical professionals don't want to fool with zooming and panning

Lifelines2

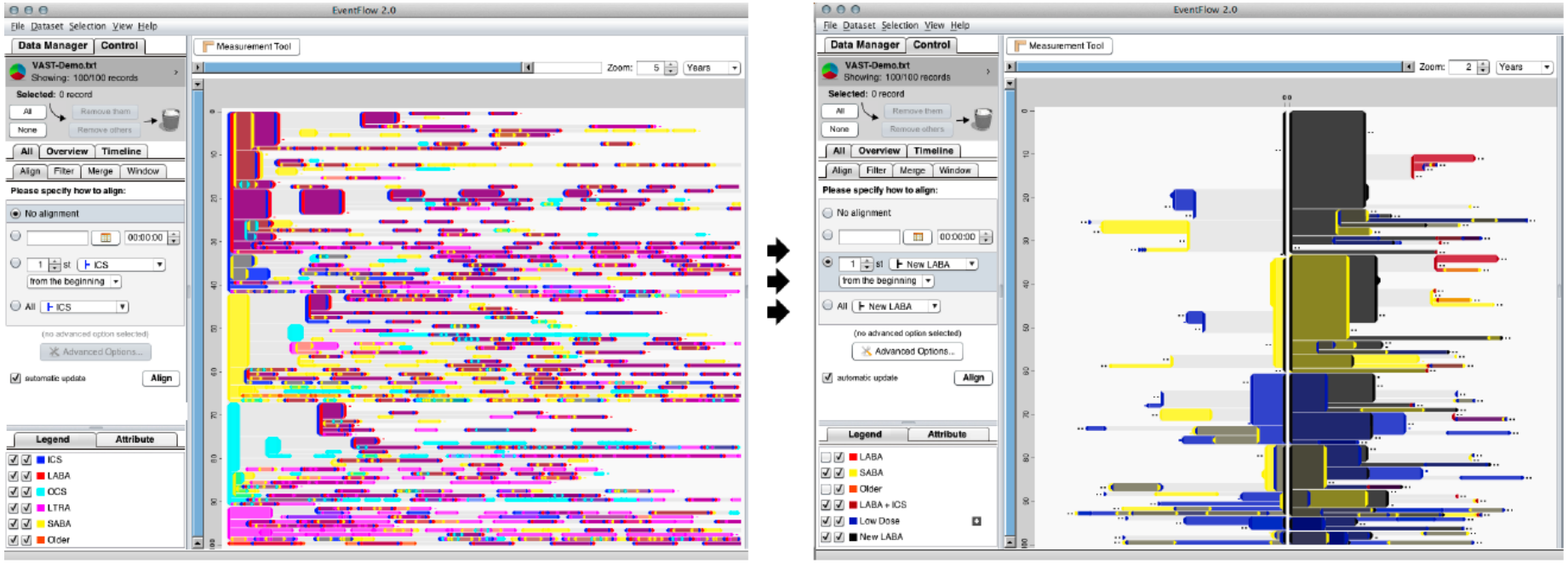
- Focus on alignment along events



Wang et al., 2008

EventFlow

- Smart aggregations to show overviews of large collections of events



Transform

Monroe et al., 2013

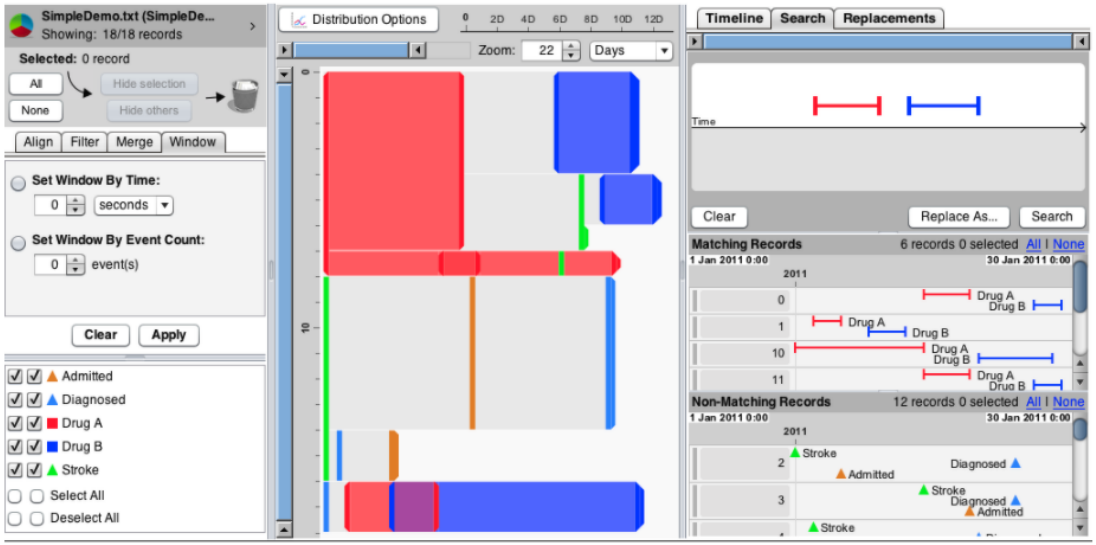
HCIL projects

Summary of HCIL Projects in **Temporal Visualizations**

Quick Links: LifeLines2, LifeLines, Similan, LifeFlow, EventFlow, PatternFinder, PatternFinder in Amalga, TimeSearcher 1-3, Learning Historian, LifeLines (original)

EVENTFLOW : EXPLORING POINT AND INTERVAL EVENT PATTERNS

The HCIL's ongoing work with temporal event records has produced powerful tools for analyzing and exploring patterns of point-based events (Lifelines2, LifeFlow). However, users found that point-based events limited their capacity to solve problems that had inherently interval attributes, for example, the 3-month interval during which patients took a medication. To address this issue, EventFlow extends its predecessors to support both point-based and interval-based events. Interval-based events represent a fundamental increase in complexity at every level of the application, from the input and data structure to the eventual questions that a user might ask of the data. Our goal was to accomplish this integration in a way that appeared to users as a simple and intuitive extension of the original LifeFlow tool. With EventFlow, we present novel solutions for displaying interval events, simplifying their visual impact, and incorporating them into meaningful queries.



http://www.cs.umd.edu/hcil/temporalviz/

Data Exploration & Visualization

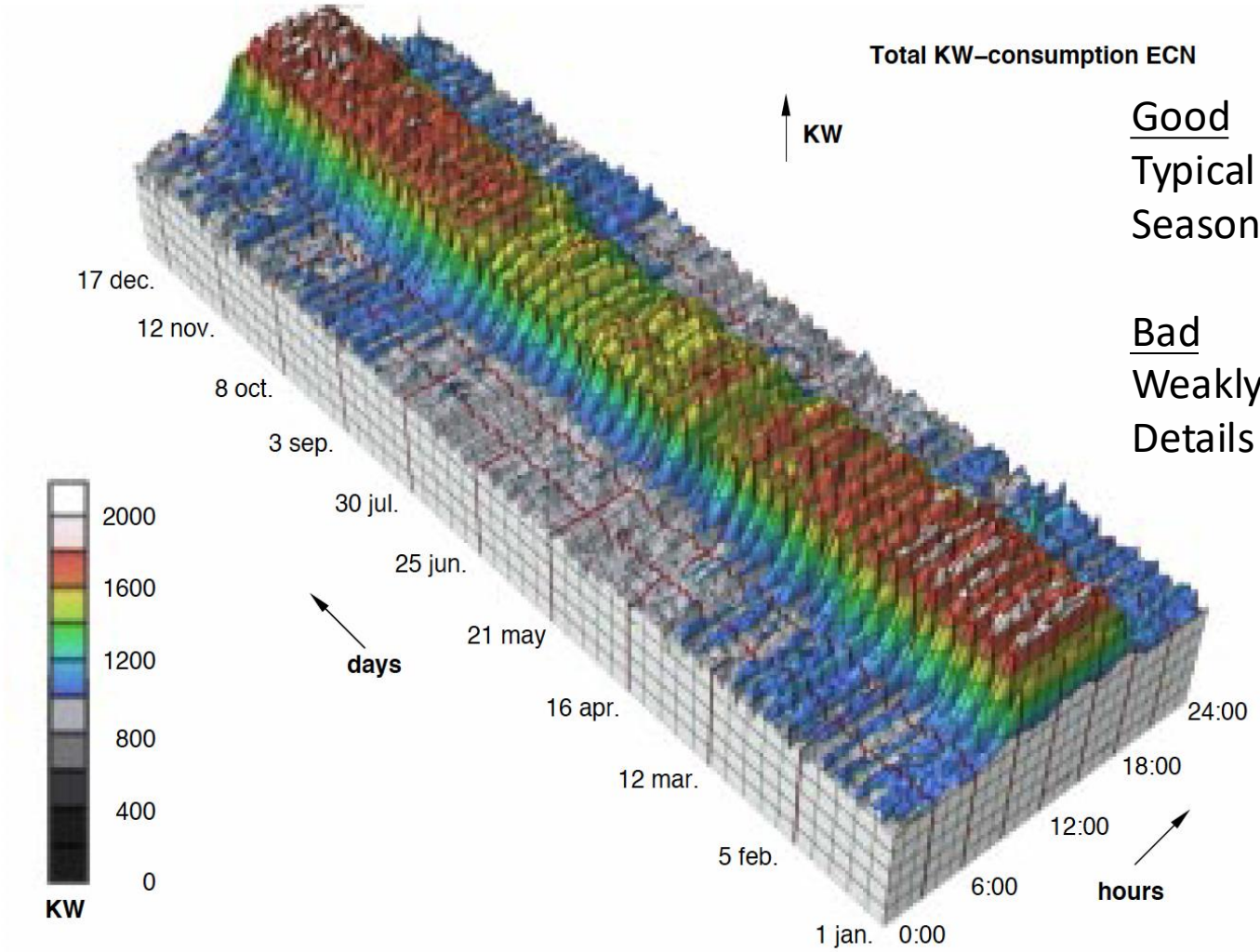
Module 10: Temporal Visualization

- Time series data
  - Properties, tasks, taxonomy
- Temporal visualization
  - Linear times: Line charts, stacked graph
  - Interval times:
    - Cyclic: calendar view, cyclic layout
    - Linear: Gantt charts, lifelines
- Case study

Case study 1

- Tasks:
  - Understand patterns of presence/resource usage/events over time
  - Show this large amount of data in an easily understandable and query-able manner
- Scenarios:
  - Workers punch in and punch out of a factory
    - Want to understand the presence patterns over a calendar year
  - Power plant electricity usage over a year

Alternative design



Good
Typical daily pattern
Seasonal trend

Bad
Weakly pattern
Details

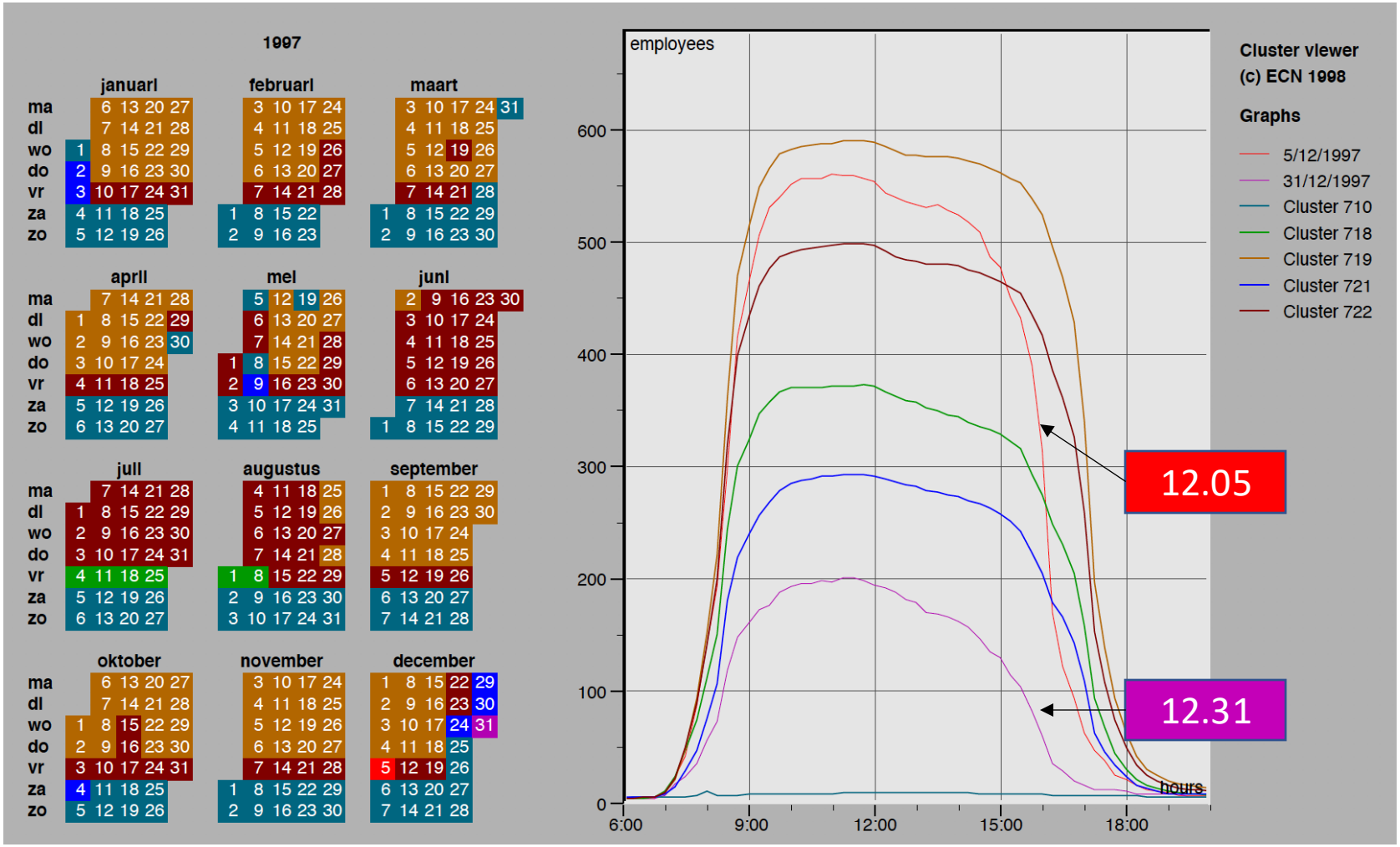
van Wijk and van Selow, 1999

Approach taken

- Before visualization: Cluster analysis
  - Find two most similar days, make into one new composite
  - Keep repeating until some preset number left or some condition met

- Visualization?

Visualization

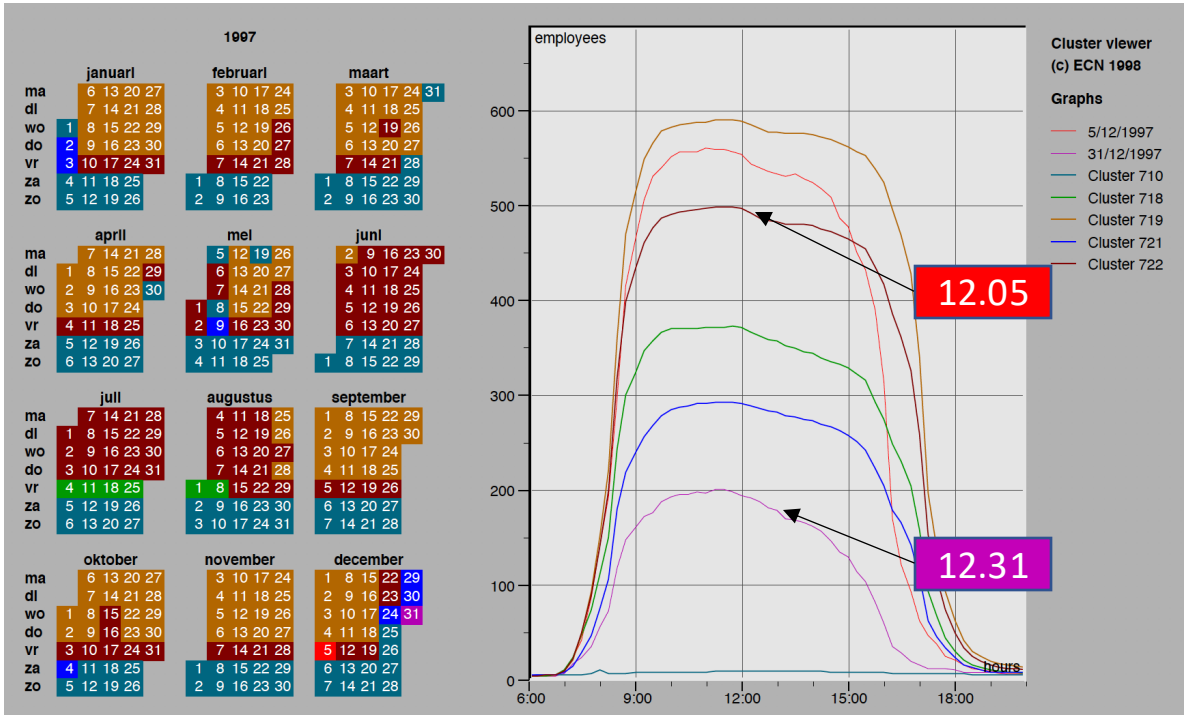


Characteristics

- Unique types of days (individual or cluster) get their own color
- Contextually placed in calendar and line graph for it is shown
- Stop clustering when a threshold met or at a predetermined number of clusters
- Interactions
  - Click on day, see its graph
  - Select a day, see similar ones
  - Add/remove clusters

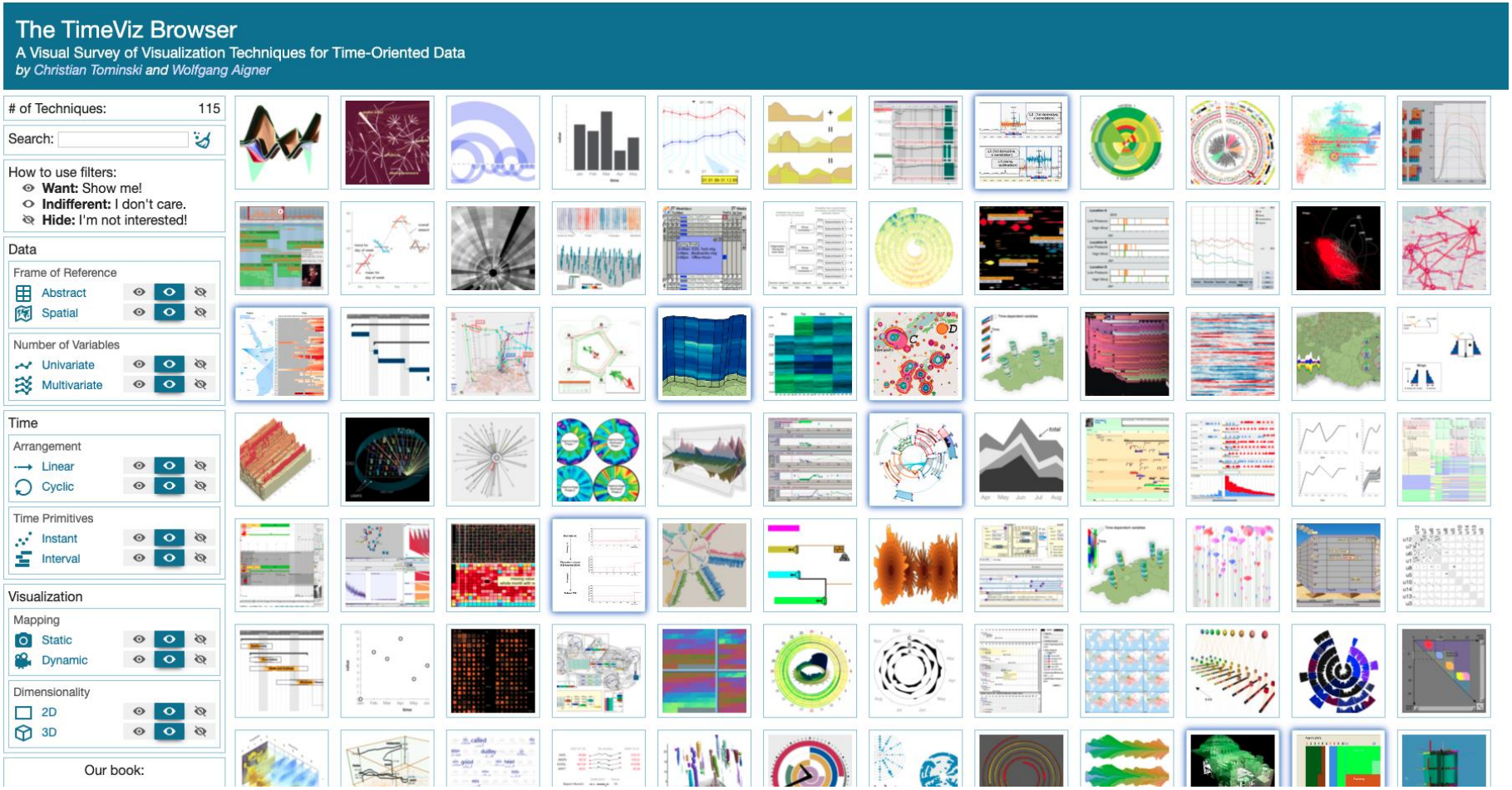
Insights

- Traditional office hours followed
- Most employees present in late morning
- Just a few people work holidays when the holidays occurred



- Many people take off day after holiday
- Many people leave at 4pm on December 5

Summary



https://vcg.informatik.uni-rostock.de/~ct/timeviz/timeviz.html