CS 519: Scientific Visualization

Time Series

Eric Shaffer

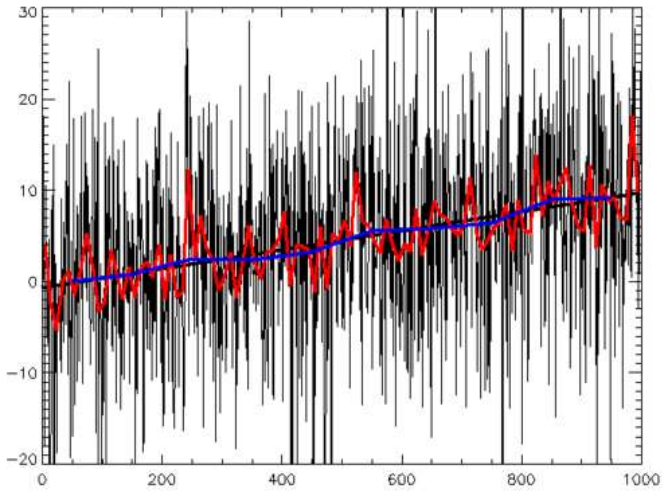
Some slides adapted from work by
Professor Tamara Munzner
University of British Columbia

Time Series

A **time series** is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.

-- **Wikipedia**

$$D = \{(t_1, d_1), \ldots, (t_n, d_n)\}$$



Purpose of Time Series Visualizations

- Some common goals of time series visualizations
  - Detect properties of an unknown function
    - e.g. is there periodicity?
    - e.g. Is the function smooth?
    - e.g. Is there high variance?
  - Determine when was something greatest/least?
  - Are two series similar?
  - Verify that a statistical analysis seems to be correct.
- For which of these is visualization not really appropriate?
- Can you think any other tasks people use time series for?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

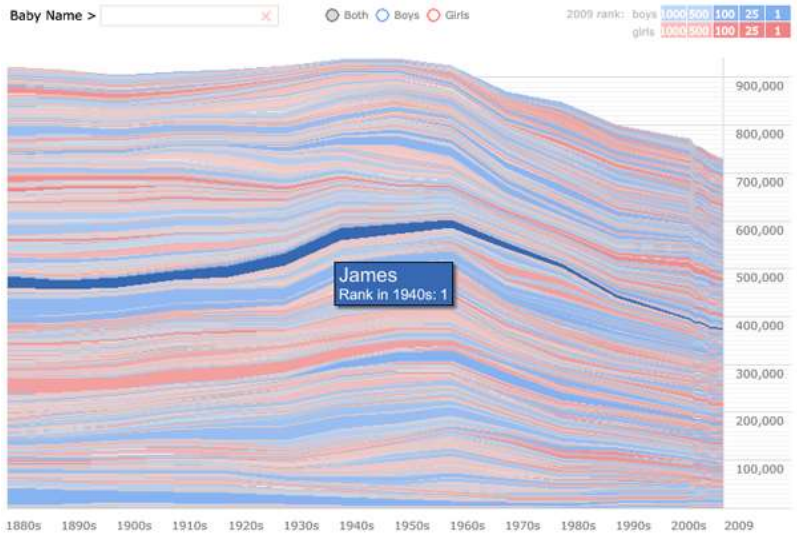
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- Professor Tamara Munzner

Example: ThemeRiver

ThemeRiver: Visualizing Thematic Changes in Large Document Collections,
Susan Havre, Elizabeth Hetzler, Paul Whitney, Lucy Nowell
2002

- A version of a Stacked Graph
- Theme River display multiple time series, each shown as a "stream"
  - Typically, width of a stream the value of the stream at that time.
  - The overall width of the river is the sum of the stream values.
- Color coding allows additional information to be presented

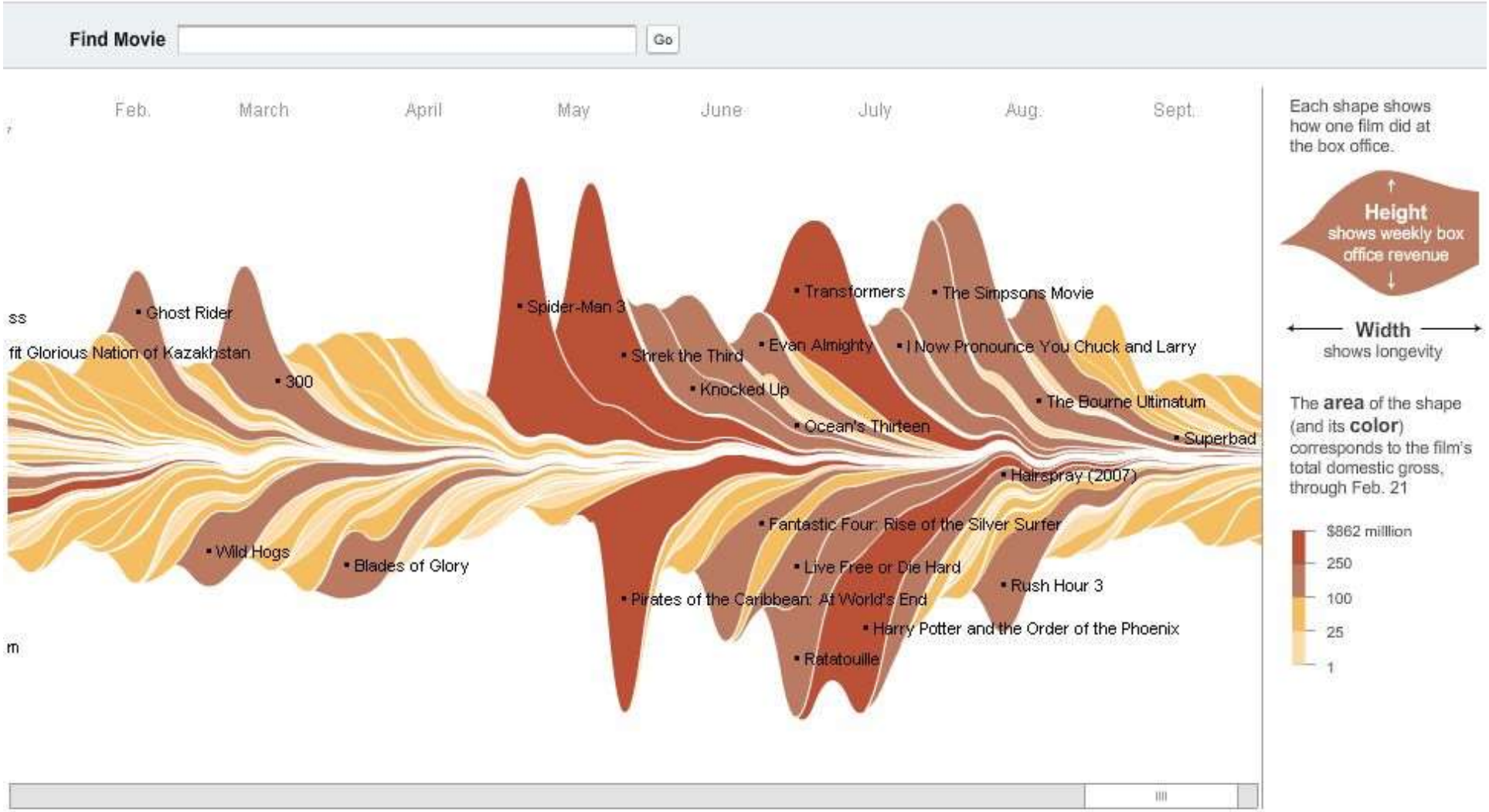


February 23, 2008

SIGN IN TO E-MAIL OR SAVE THIS | FEEDBACK

The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.



Sources: Baseline StudioSystems; Box Office Mojo

Mathew Bloch, Lee Byron, Shan Carter and Amanda Cox

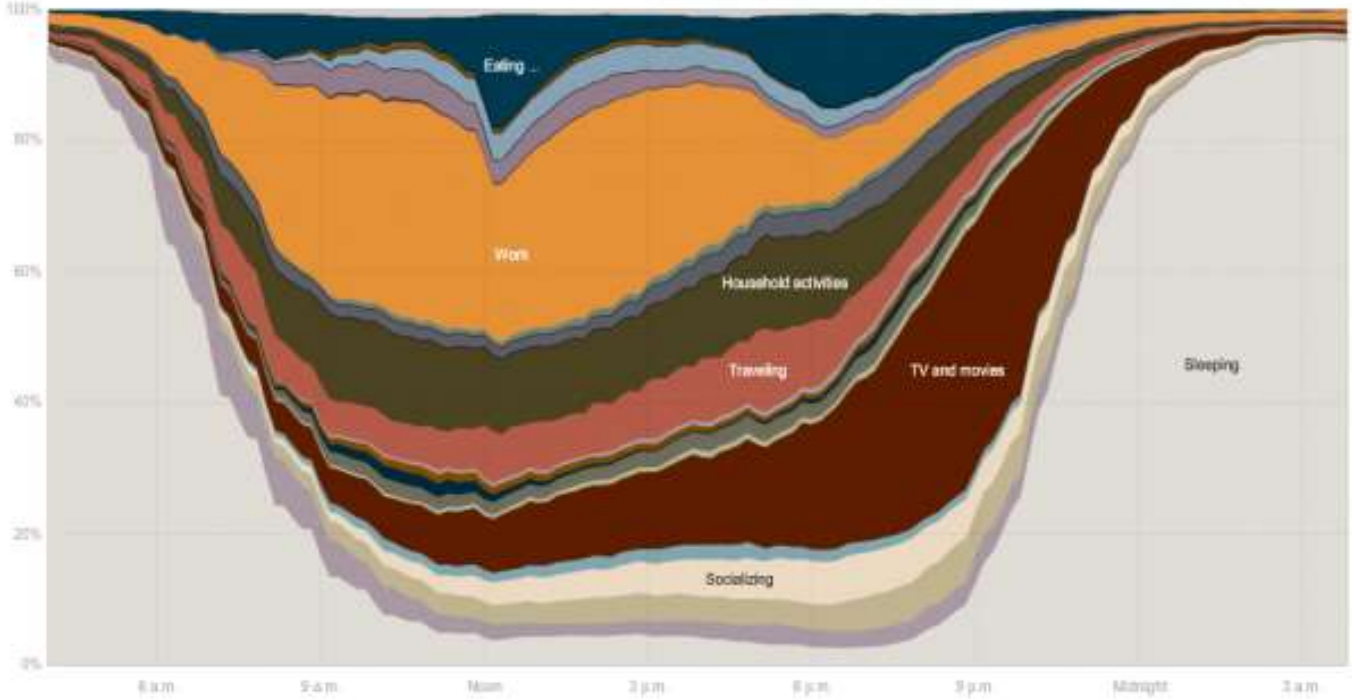
How Different Groups Spend Their Day

The American Time Use Survey asks thousands of American residents to recall every minute of a day. Here is how people over age 15 spent their time in 2008. Related article

Everyone

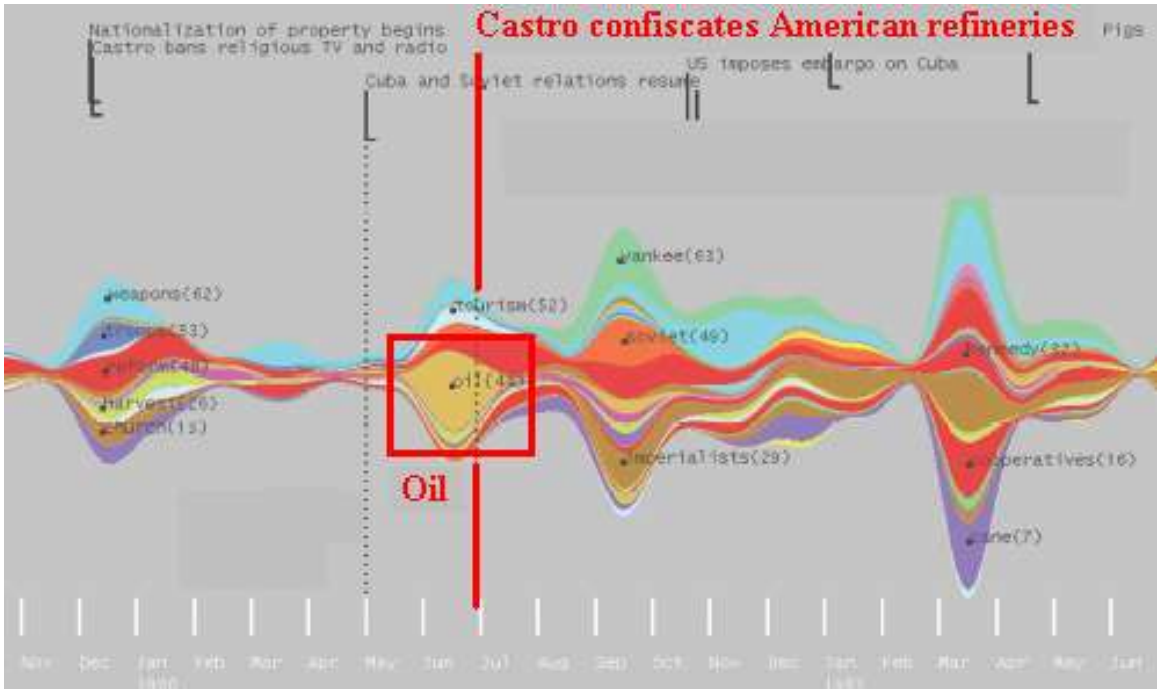
Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-54	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 55+	Advanced	Two+ children



Summary of Visual Components

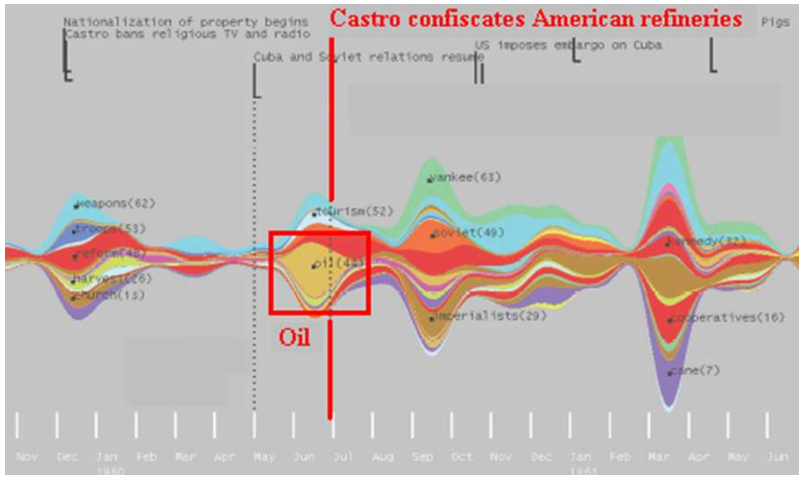
- Stream width
- River width
- Color mapping
- Time line
- Labels

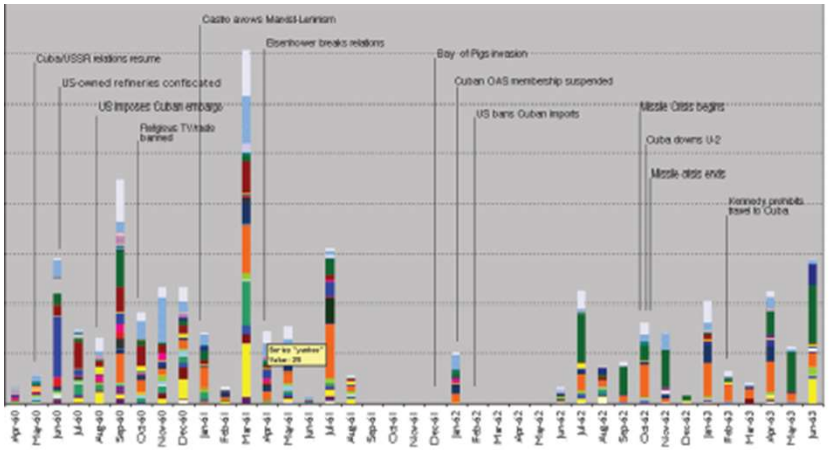


Fidel Castro's speeches 1960-1961

Some Specific Capabilities

- Compare streams quickly
  - Are they correlated?
- Show global and local evolution over time
- Users preferred continuous curves to histograms



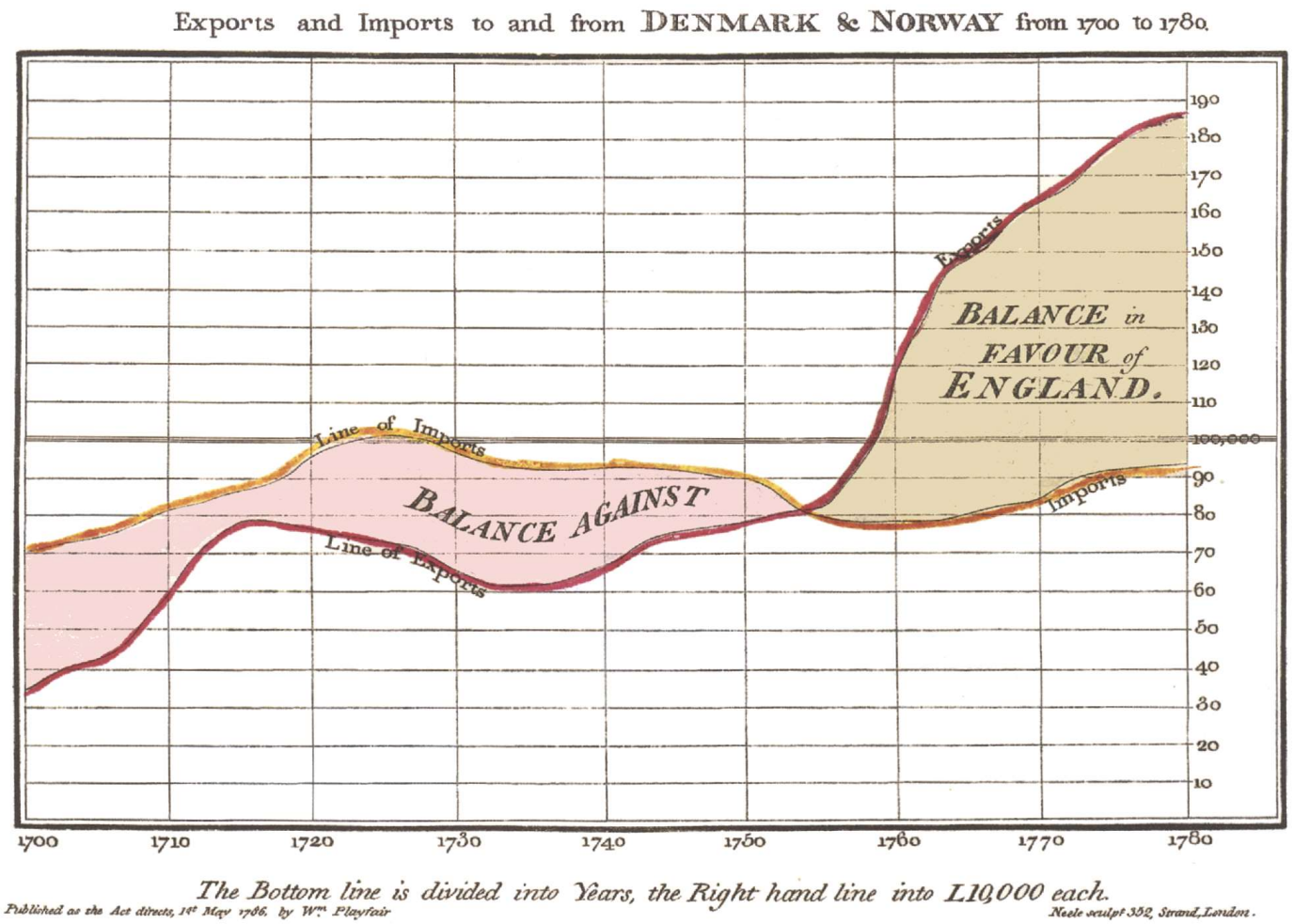


User Evaluations

From the original paper:

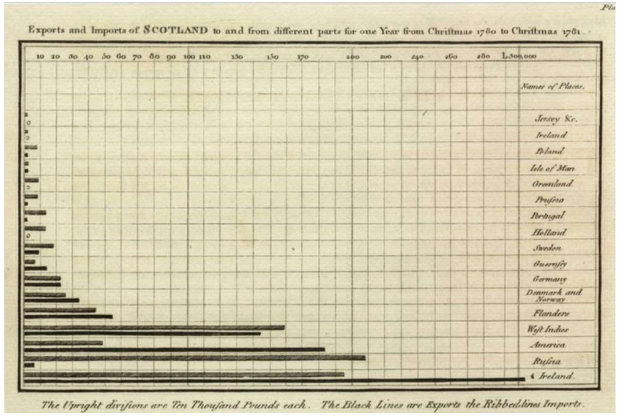
- Strong points:
  - Intuitive
  - Fast visual appreciation of change
  - Fast visual comparison of streams
- Criticisms
  - No ability to reorder streams
  - Colors sometimes chosen in way that streams hard to differentiate
  - Interpolation between data points could be misleading

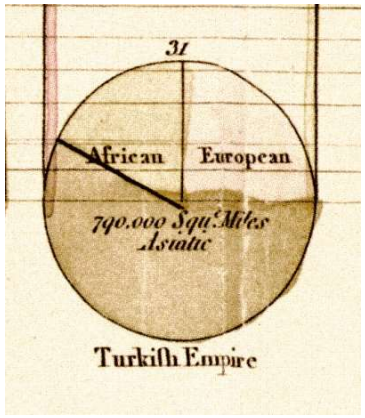
William Playfair's Time Series Visualization (1786)



William Playfair
(22 September 1759 – 11 February 1823)

- Born in Scotland during the era known as the Enlightement
- Founder (?) of graphical statistics
  - Invented the bar chart and pie chart

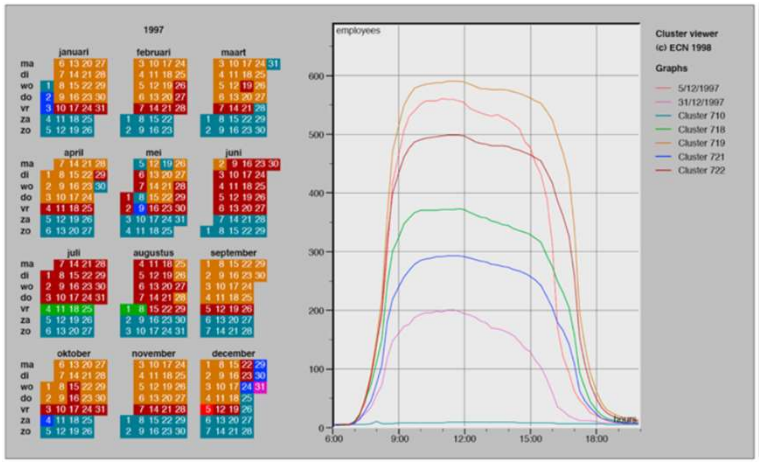




Playfair had a variety of careers. He was in turn a millwright, engineer, draftsman, accountant, inventor, silversmith, merchant, investment broker, economist, statistician, pamphleteer, translator, publicist, land speculator, convict, banker, ardent royalist, editor, blackmailer and journalist. On leaving Watt's company in 1782, he set up a silversmithing business and shop in London, which failed. In 1787 he moved to Paris, taking part in the storming of the Bastille two years later. He returned to London in 1793, where he opened a "security bank", which also failed. From 1775 he worked as a writer and pamphleteer and did some engineering work. --**Wikipedia**

Cluster-Calendar

- Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.
- Focus on univariate time series data
- Challenges:
  - Data can be large
    - Can't be displayed in single view
  - Patterns occur multiple time scales
    - Days, weeks, months….
  - Vis is used for hypothesis formation
    - Interaction required



Cluster Calendar: Using More Dimensions

- We wish to detect *day patterns* (i.e. when do similar days occur)
- Consider making function 2D $f(day, hour) = d_i$
- Display shown here is problematic… we'll discuss later

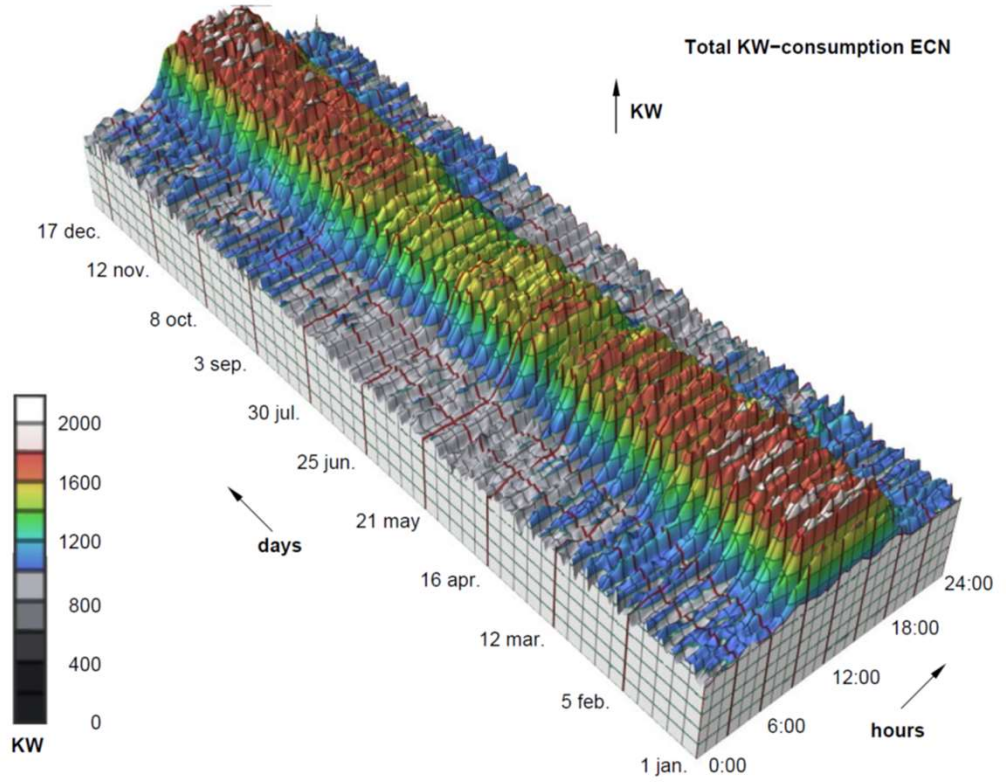
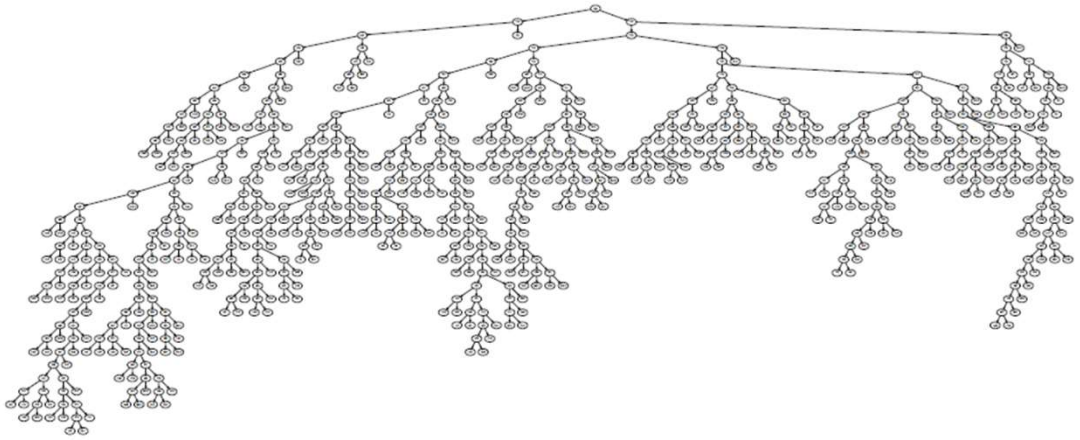


Figure 1. Power demand by ECN, displayed as a function of hours and days

Better Approach: Clustering Day Patterns

- Use bottom-up clustering to cluster day patterns
- Start with M days…each in a single cluster
  - Compute a measure of distance between each cluster
  - Merge the least different pair of clusters
  - Iterate….
- Produces a binary tree of 2M-1 clusters



Measuring Distance

- Suppose we have 2 day patterns $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$
- Can measure distance as average geometric distance

$$d_{rms} = \sqrt{\sum (y_i - z_i)^2 / N}$$

- Can employ a normalized version

$$d_{nm} = \sqrt{\sum (y_i / y_{\max} - z_i / z_{\max})^2 / N}$$

  - How is it normalized?
  - What sort of patterns matches will it pick out?

Measuring Distance

- Suppose we have 2 day patterns $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$
- If we are only interested in peak values

$$d_{ma} = | \, y_{\max} - z_{\max} \, |$$

- If we want to determine if two patterns match except for an offset

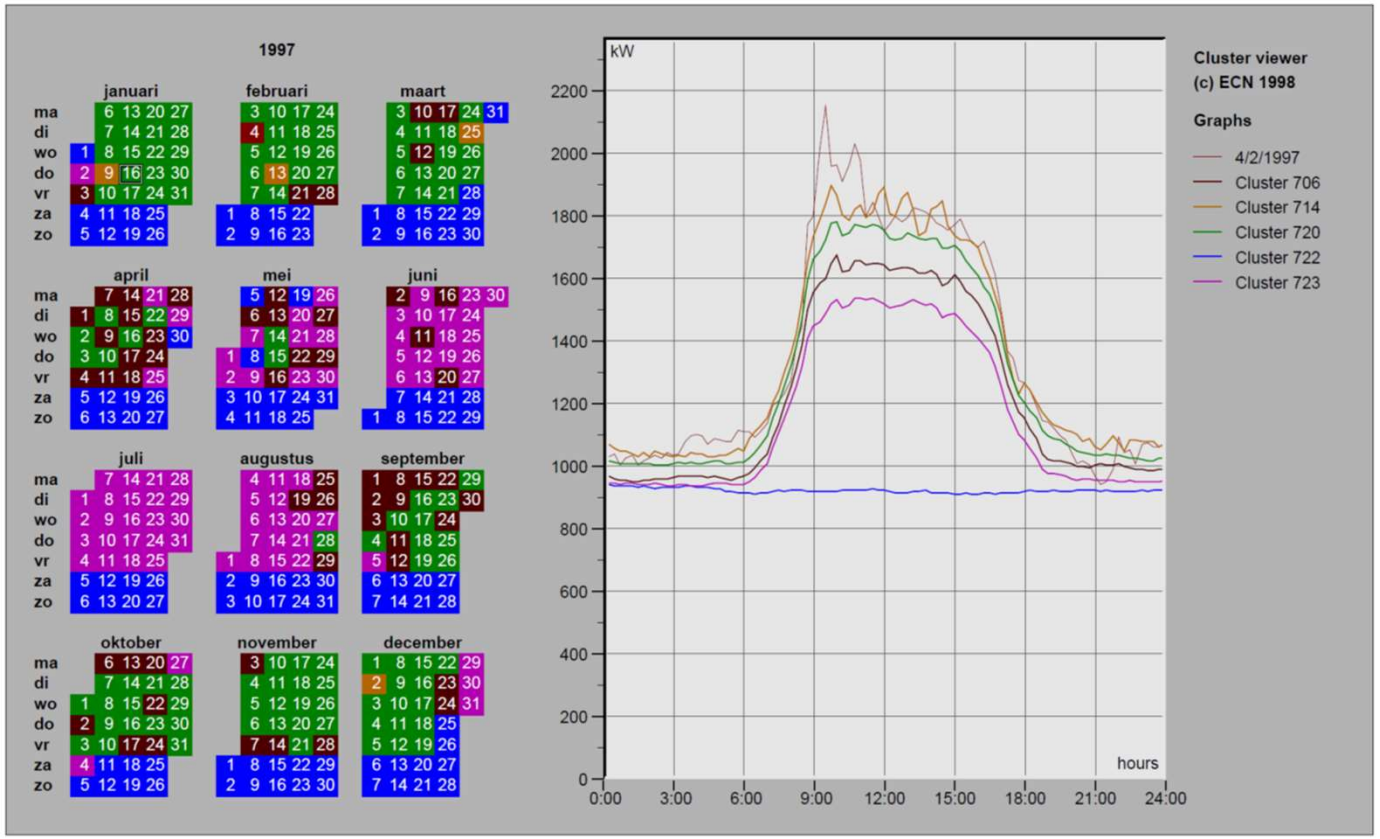
$$d_{sh} = \sqrt{\sum (y_i - z_i - \Delta)^2 / N},$$

with

$$\Delta = \sum (y_i - z_i) / N.$$

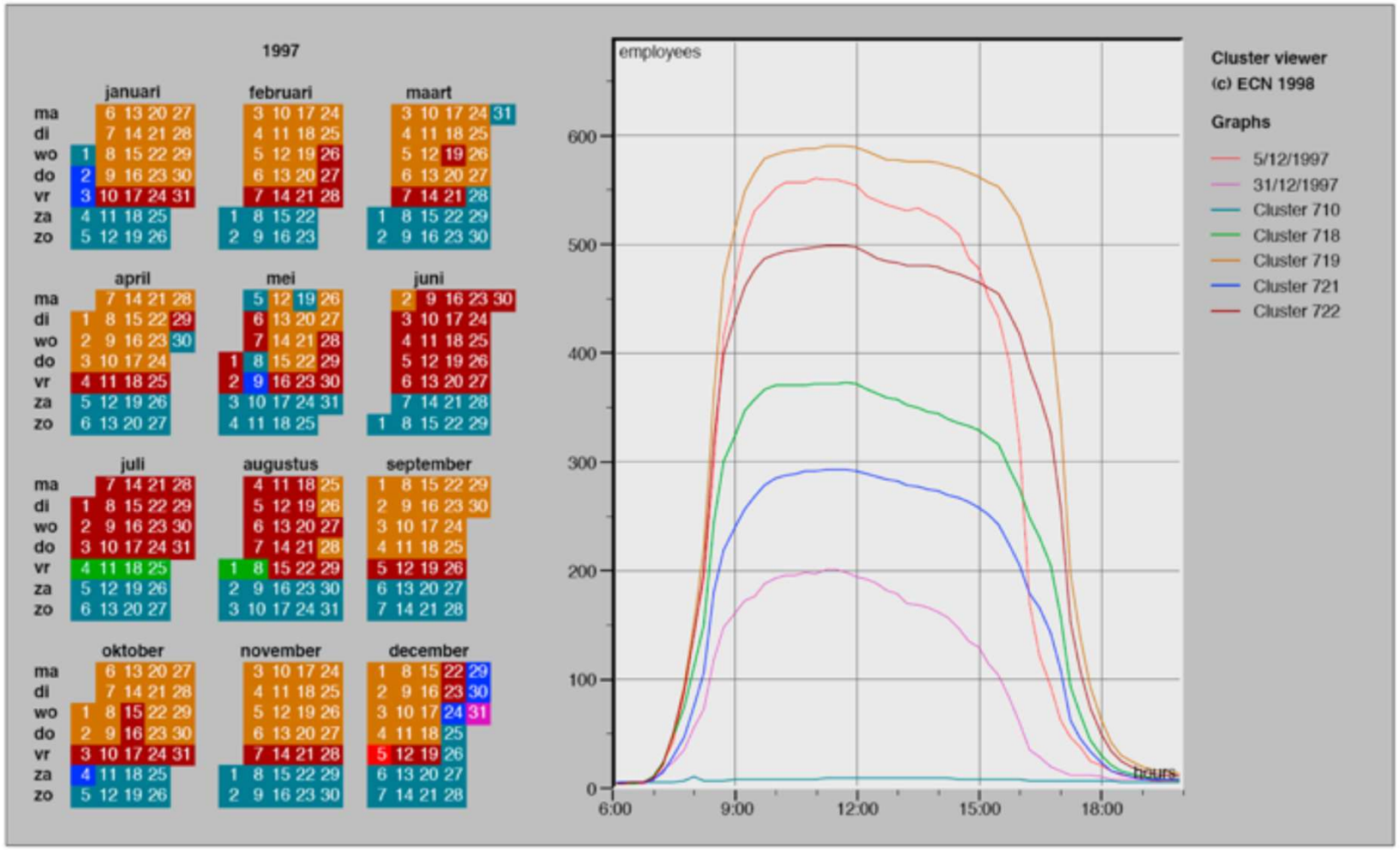
Visualization

- Clusters shown via color code on calendar
- Avg hourly pattern of a cluster is shown as a graph on the right



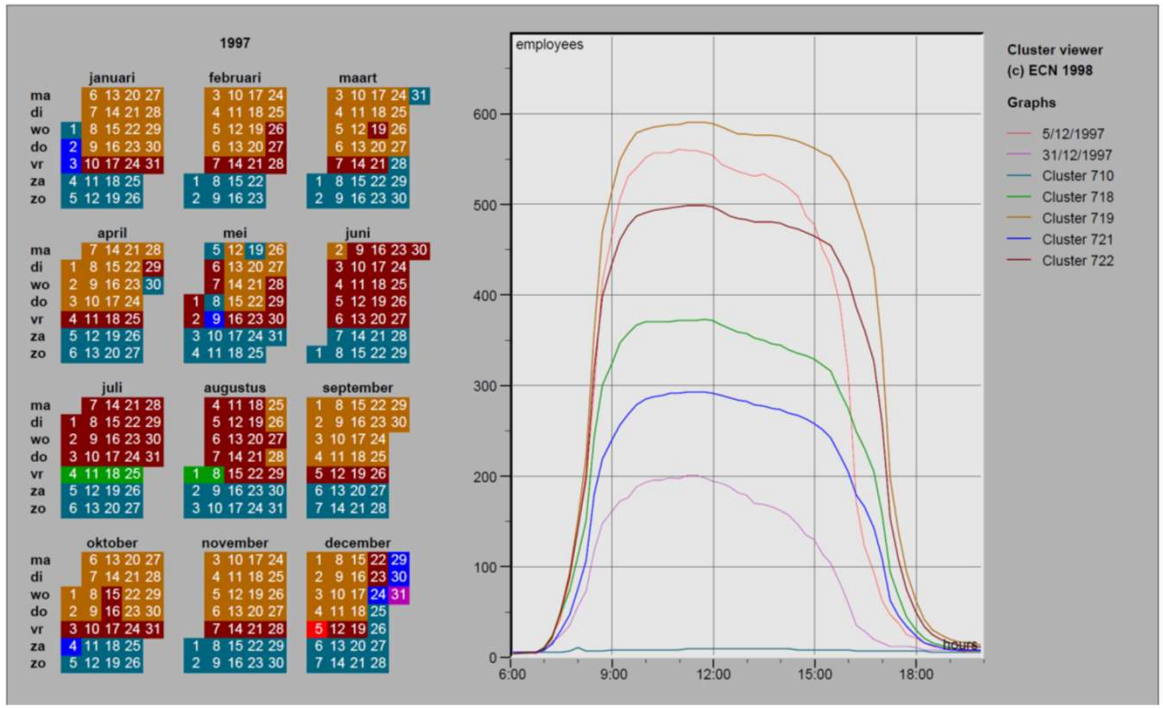
Cluster-Calendar

- What can we discover from looking at the number of employees present each hour of each day at some research institute in the Netherlands?



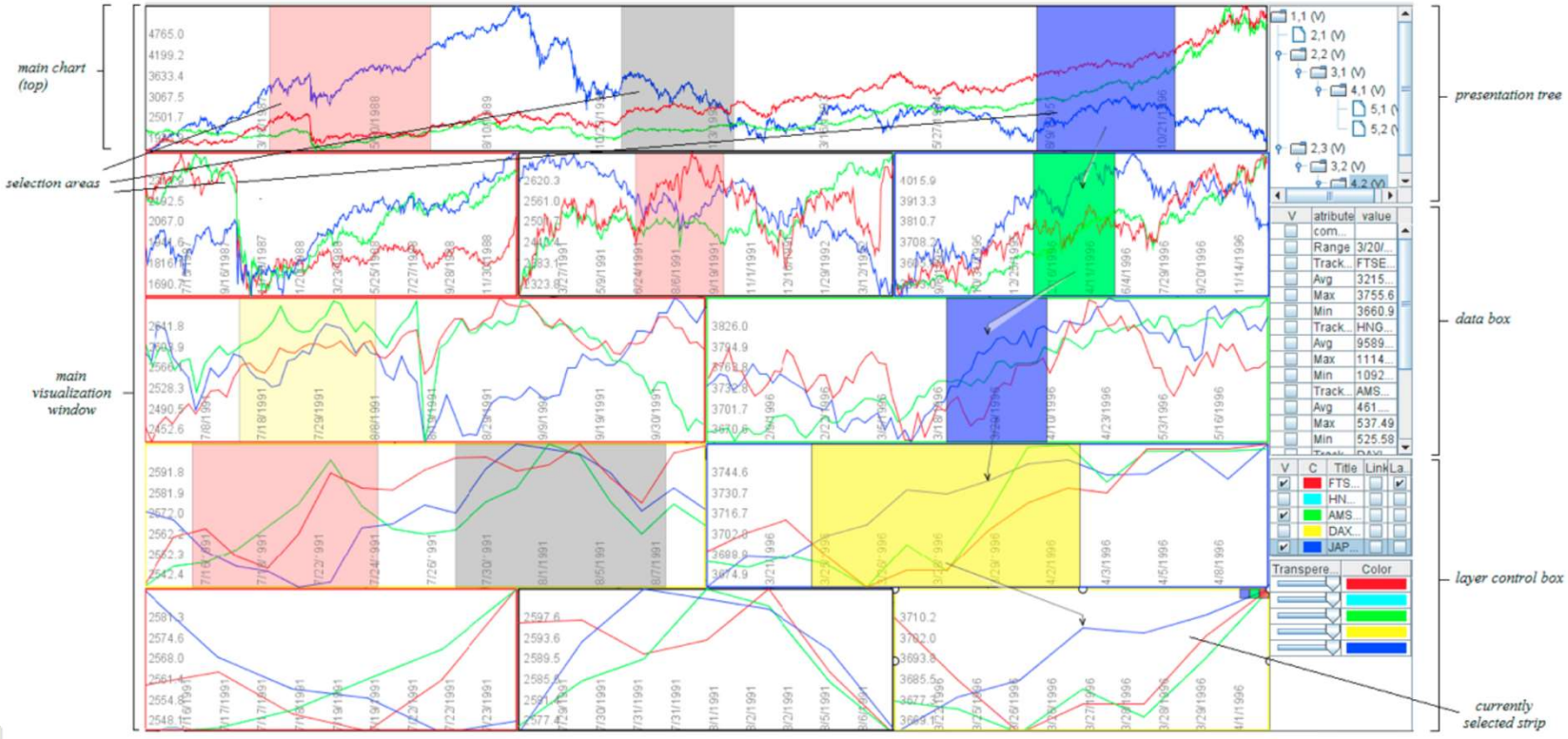
What can we learn?

- Most people work 9 to 5ish
- People take vacation on Fridays and in summer
- Holidays and weekends very visible
- Anything else?



Stack Zooming

- Stack Zooming for Multi-Focus Interaction in Time-Series Data Visualization. Javed and Elmqvist. Proc PacificVis 2010, p 33-40.
- https://youtu.be/dK0De4XPm5Y



Stack Zooming Example

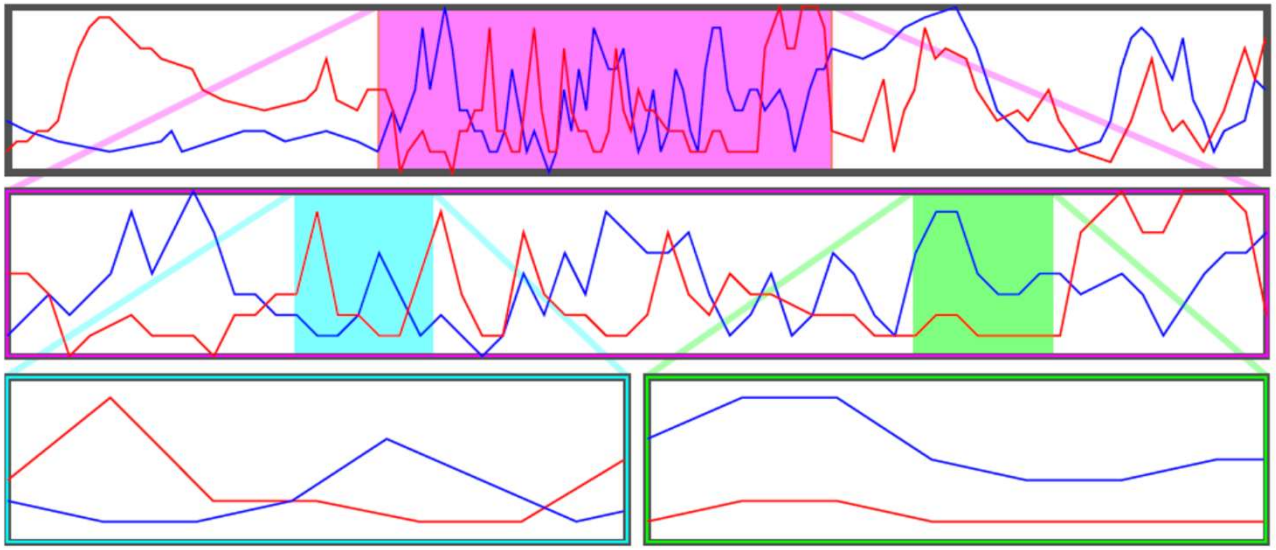
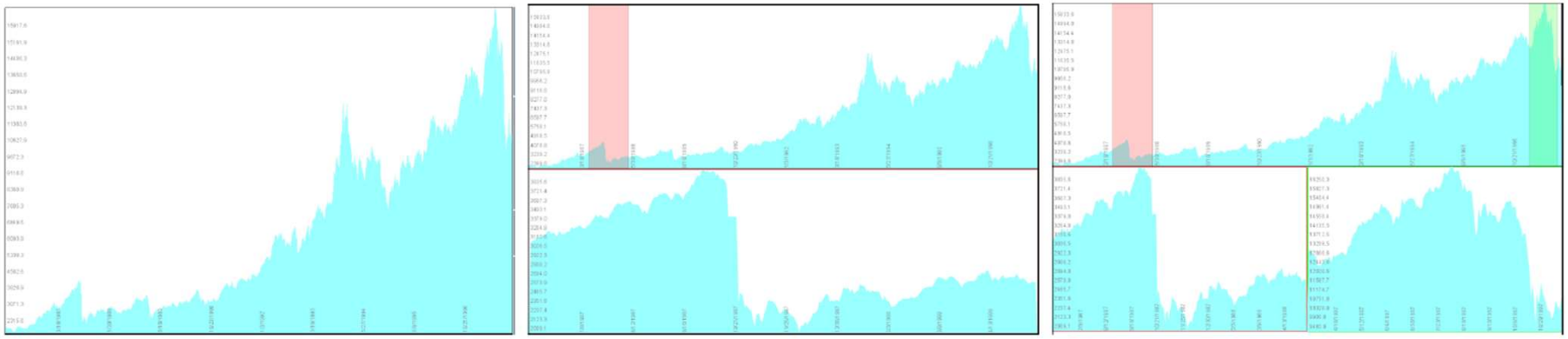


Figure 1: The stack zooming technique for line graphs. The analyst has focused on a period of radical changes in the main timeline (top).

Stack Zooming Example



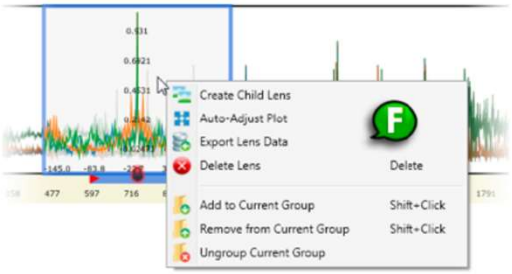
(a) Main strip showing the whole dataset (June 1986 to December 1997).

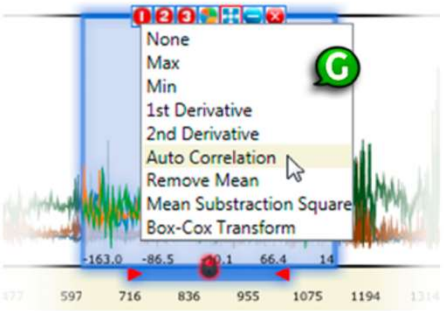
(b) Creating a child strip for Black Monday (October 19, 1987).

(c) Creating a second child strip for the Asian financial crisis (starting July 1997).

ChronoLenses

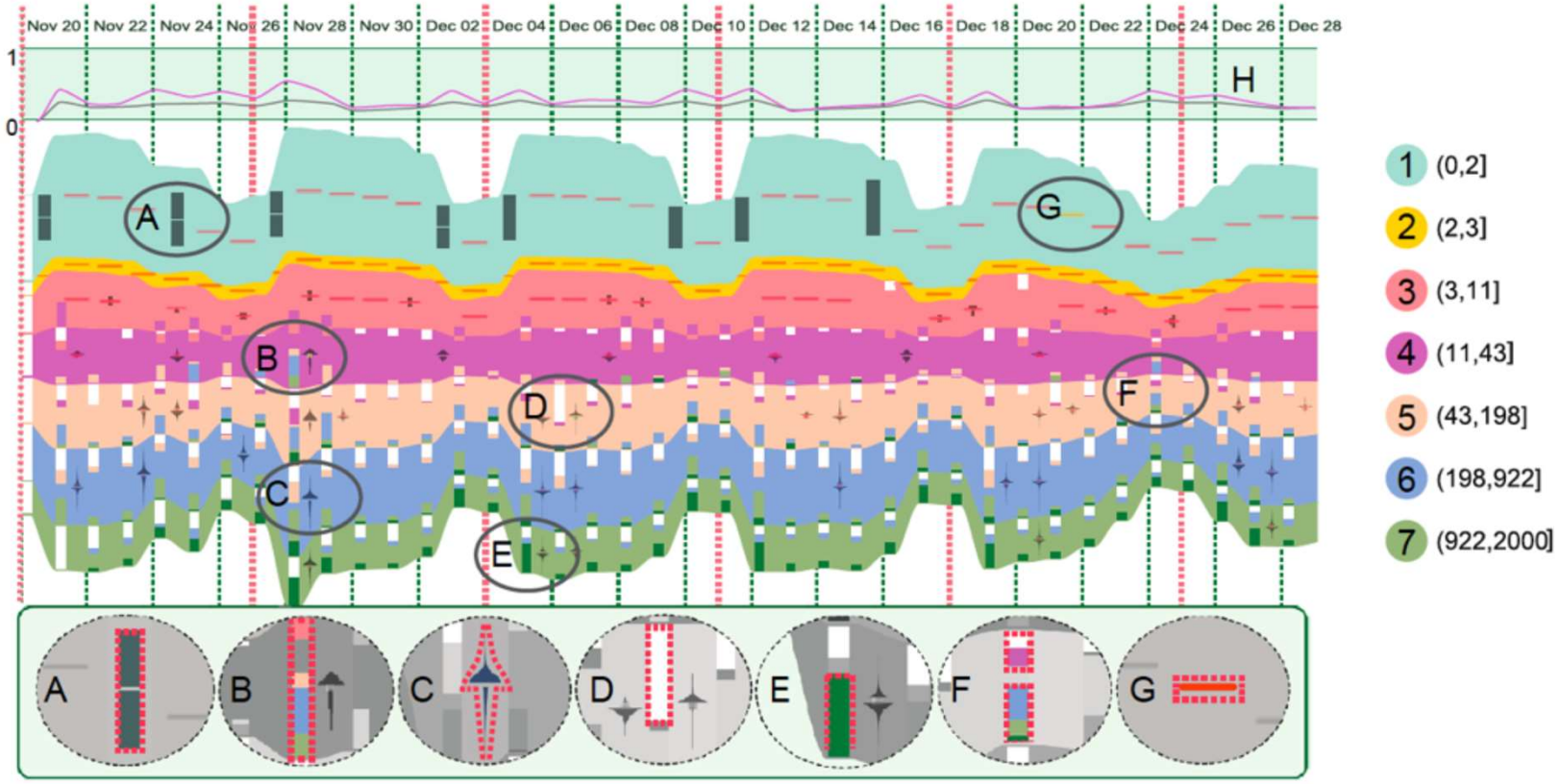
- Exploratory Analysis of Time-Series with ChronoLenses. Zhao, Chevalier, Pietriga, and Balakrishnan. IEEE TVCG 17(12):2422-2431(Proc. InfoVis 2011).
- https://youtu.be/k7pI8ikczqk





RankExplorer

- RankExplorer: Visualization of Ranking Changes in Large Time Series Data. Shi, Cui, Liu, Xu, Chen and Qu. IEEE TVCG 12(18):2669-2678 (Proc. InfoVis 2012)
- https://youtu.be/rdgn1qcZ2A4



RankExplorer Example

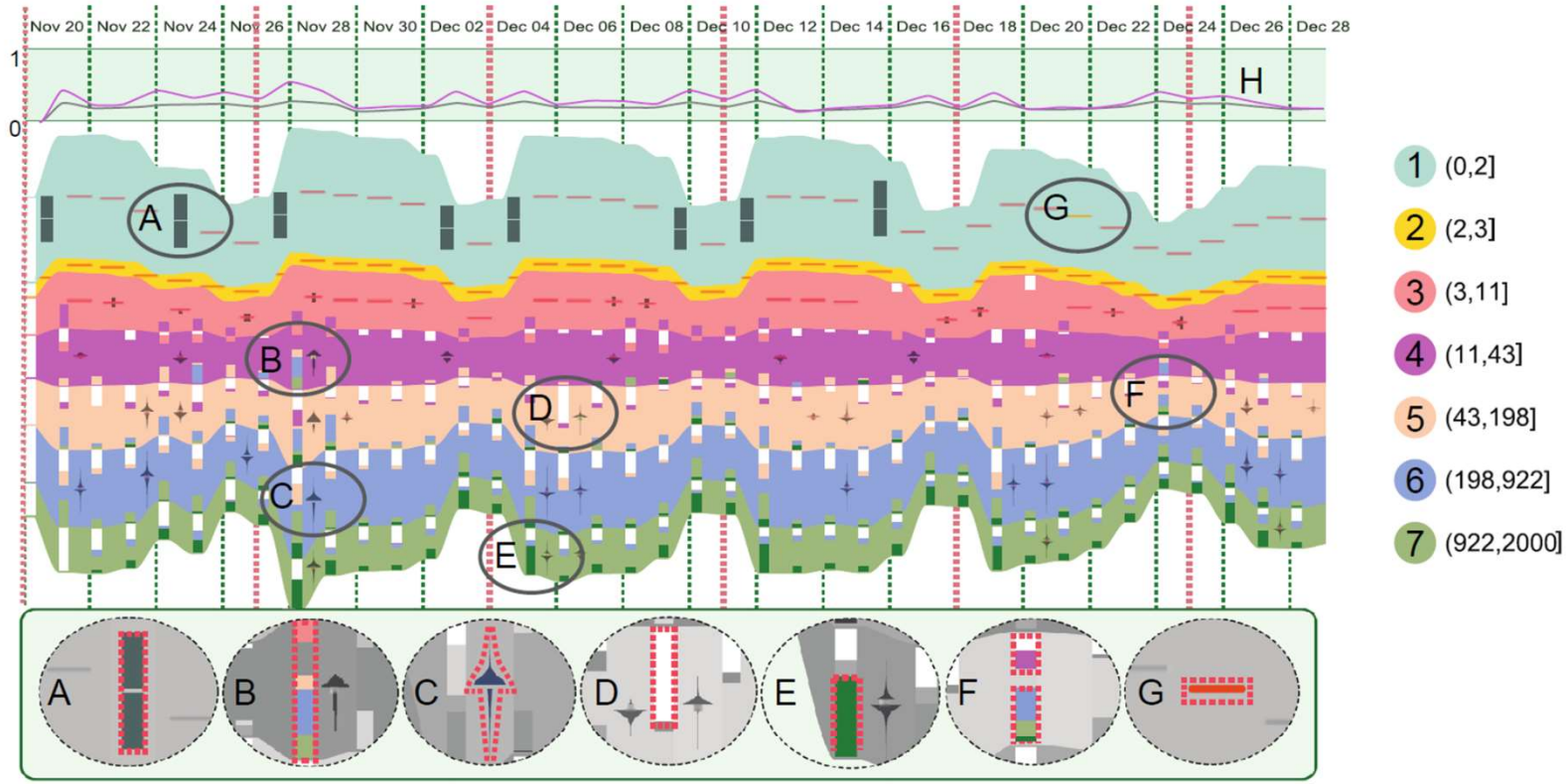


Fig. 1. RankExplorer visualization of the top 2000 Bing search queries from Nov. 20 to Dec. 29 in 2011. All queries are divided into seven categories. The width of each layer at a time point encodes the total query count at that time. The color bar and glyphs encode the content changes in each ranking category. From the color bar, we can observe: 1) the change between layers (the bar segments with the colors of other layers in B and F); 2) new queries coming in (the white segment in D); 3) recurring queries (the dark green segment in E). From the changing glyphs, we can see: 1) a non-change pattern (only red line in G); 2) a swap pattern (the two equal-height segments in A represent that the two queries swap their rankings); 3) a shift pattern (the increasing part is significantly larger than the decreasing part in C). From the trend curve (H), we can see the degree of ranking change over time.

LiveRAC

- LiveRAC - Interactive Visual Exploration of System Management Time-Series Data. McLachlan, Munzner, Koutsofios, and North. Proc. Conf. on Human Factors in Computing Systems (CHI) 2008
- http://youtu.be/ld0c3H0VSkw