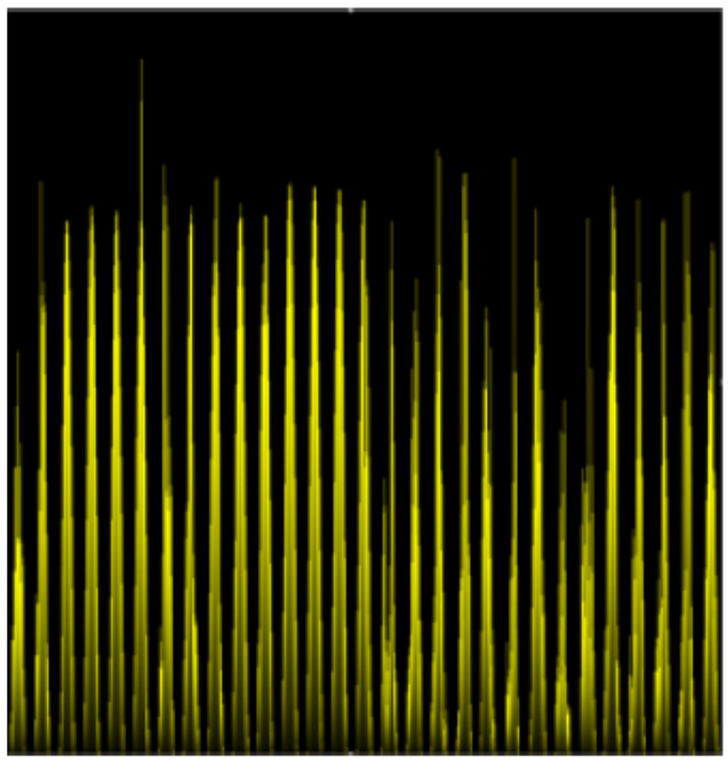
9

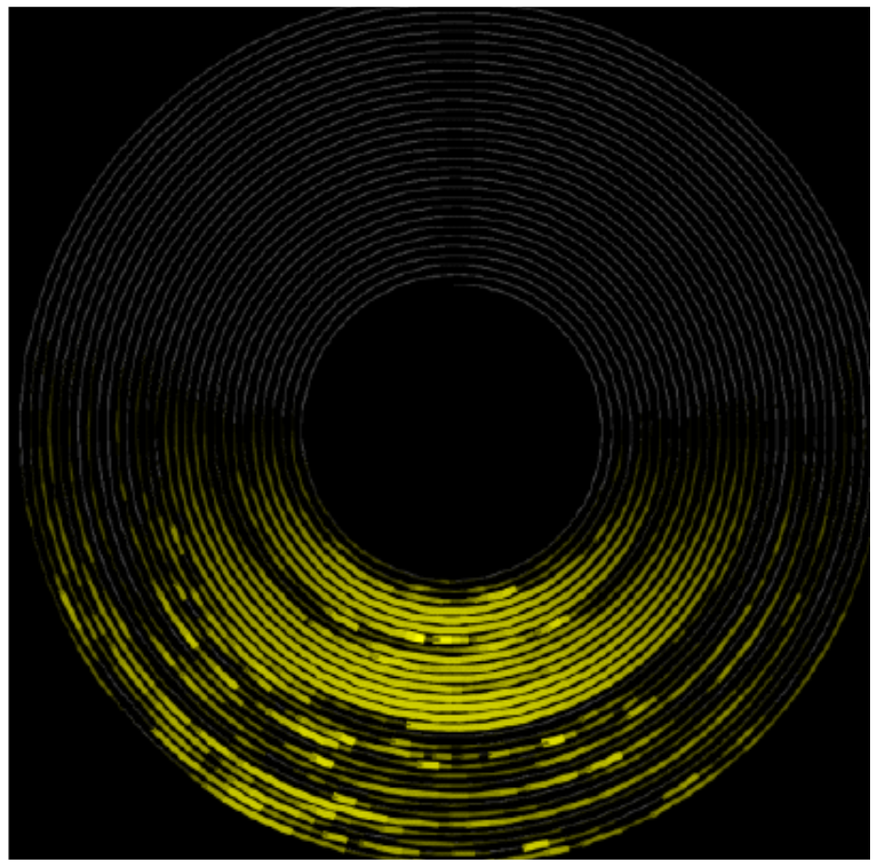
SÉRIES TEMPORAIS – PARTE III

Profa. Raquel C. de Melo Minardi

VISUALIZING TIME-SERIES ON SPIRALS

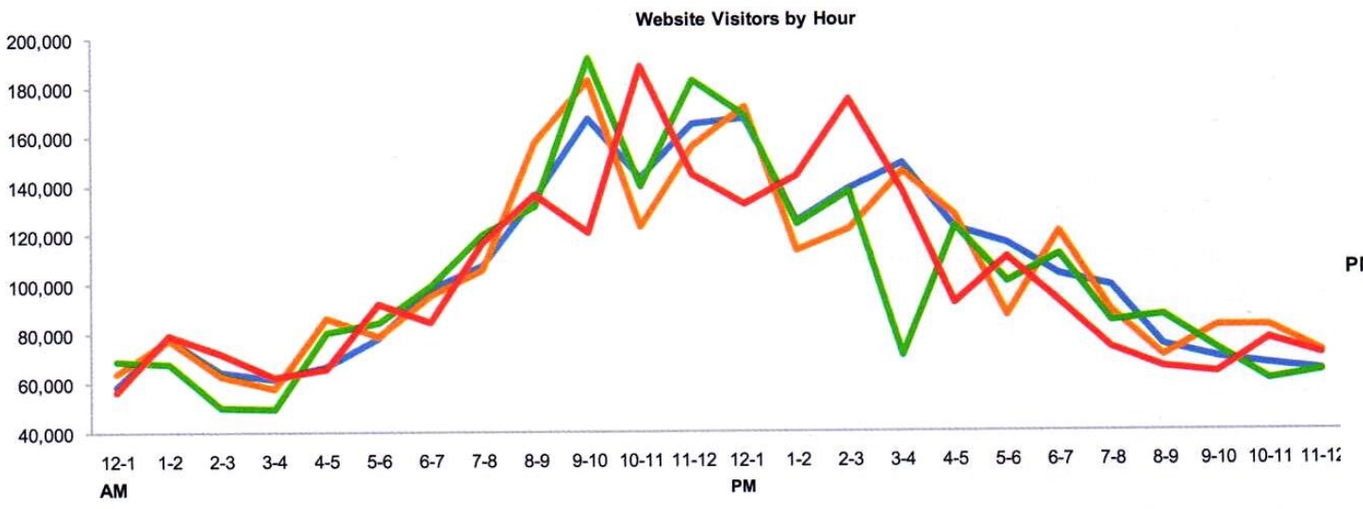
M. Weber, M. Alexa e W. Müller
IEEE Symposium on Information Visualization
2001

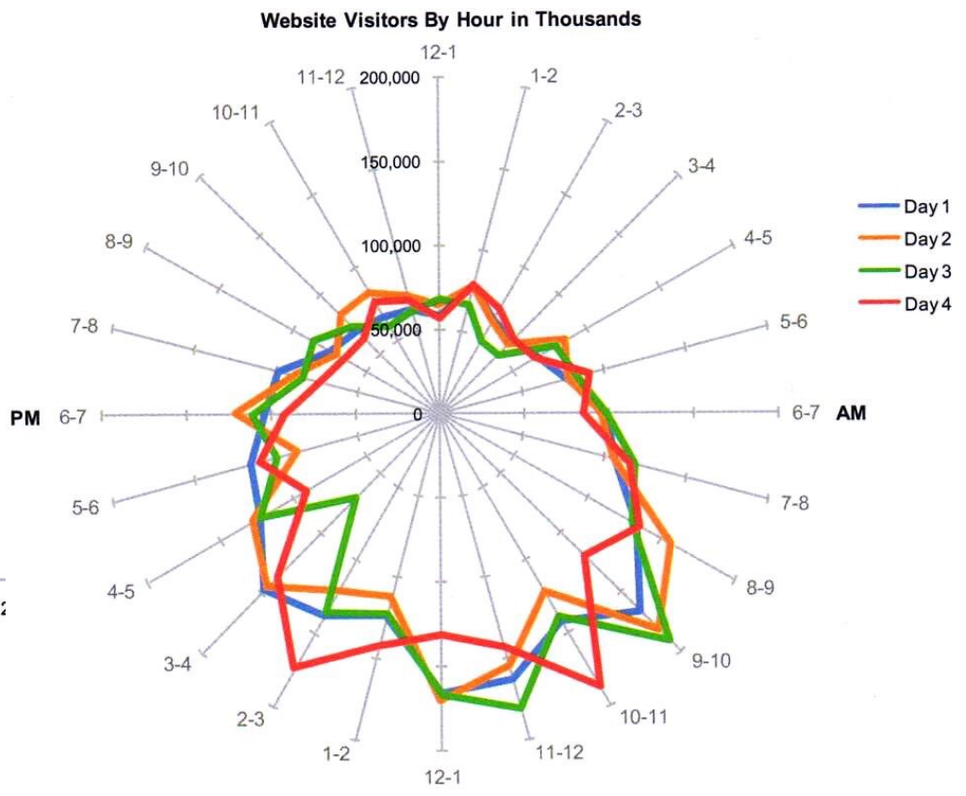




Gráficos de linhas revelam tendências em relacionamentos temporais

Gráficos em ciclo mapeiam linhas em círculos com comportamento periódico com duração conhecida

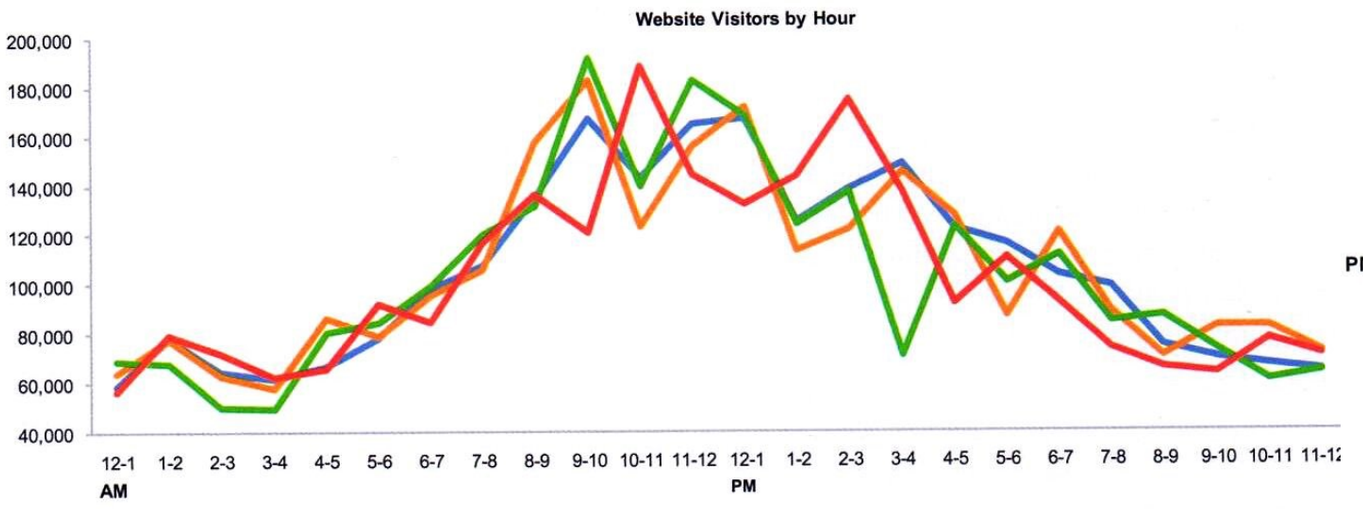


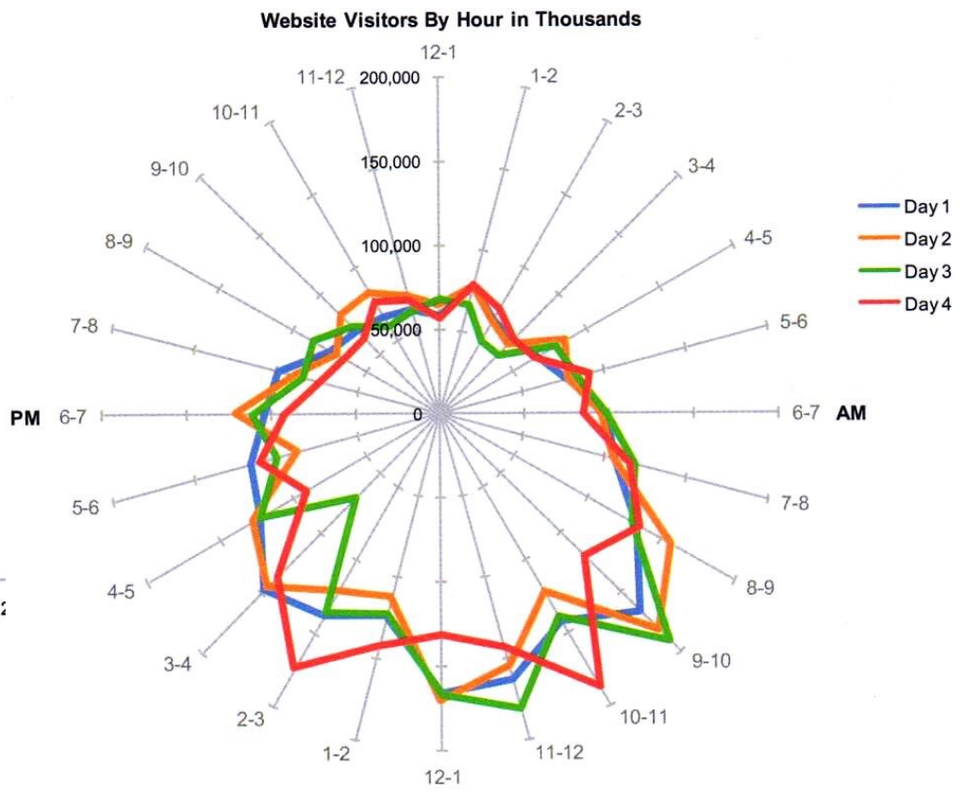


Limitações

Podem mostrar algo entre 2 e 8 diferentes séries

É preciso conhecer a duração do ciclo





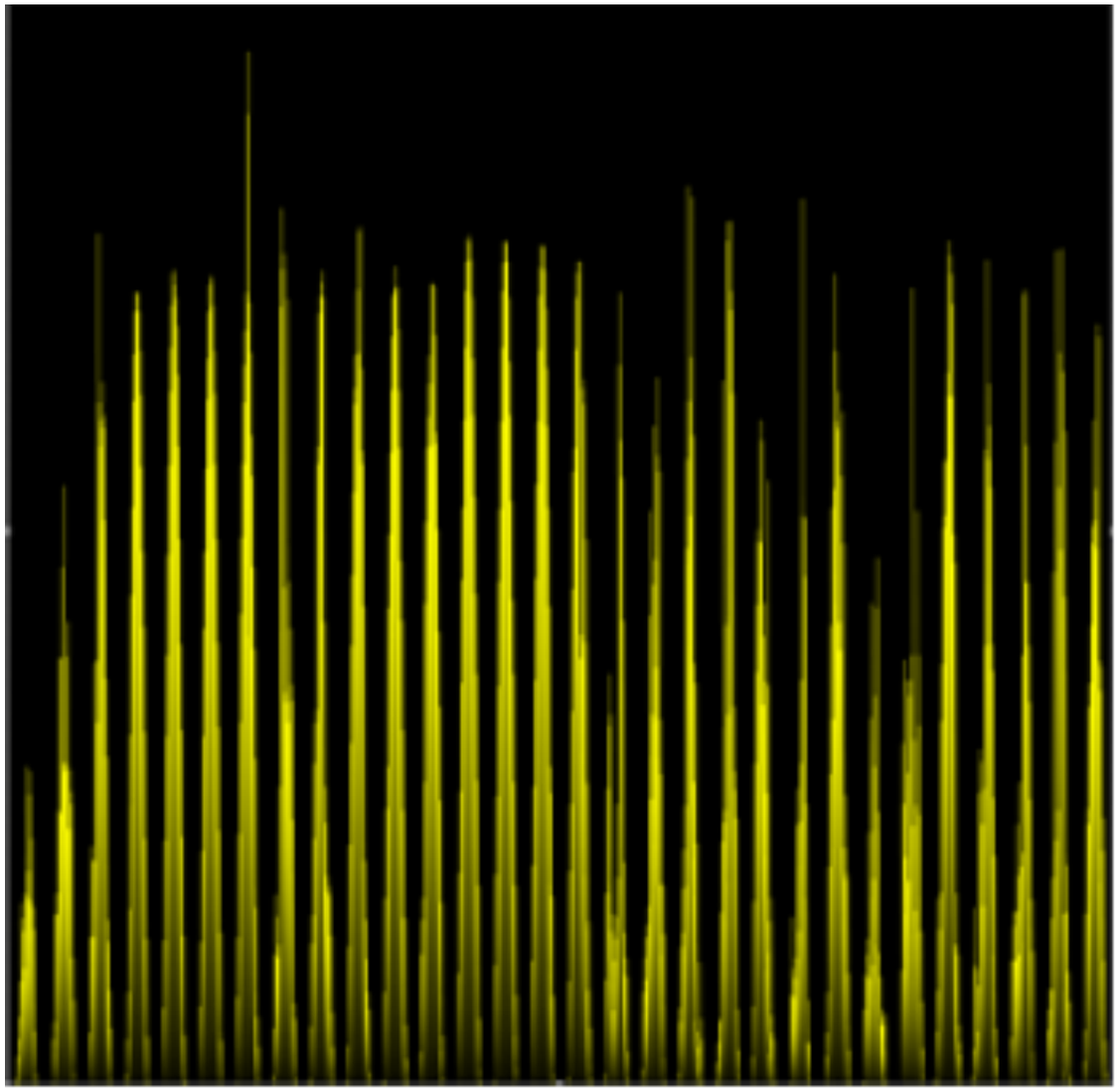
Segundo Weber et al., uma visualização de séries temporais deve:

Possibilitar a **comparação** de dados nominais e quantitativos

Suportar a visualização de **conjuntos de dados volumosos**

Suportar a **comparação de ciclos** em um conjunto de dados

Auxiliar na **identificação de padrões** de intensidades e das durações dos ciclos



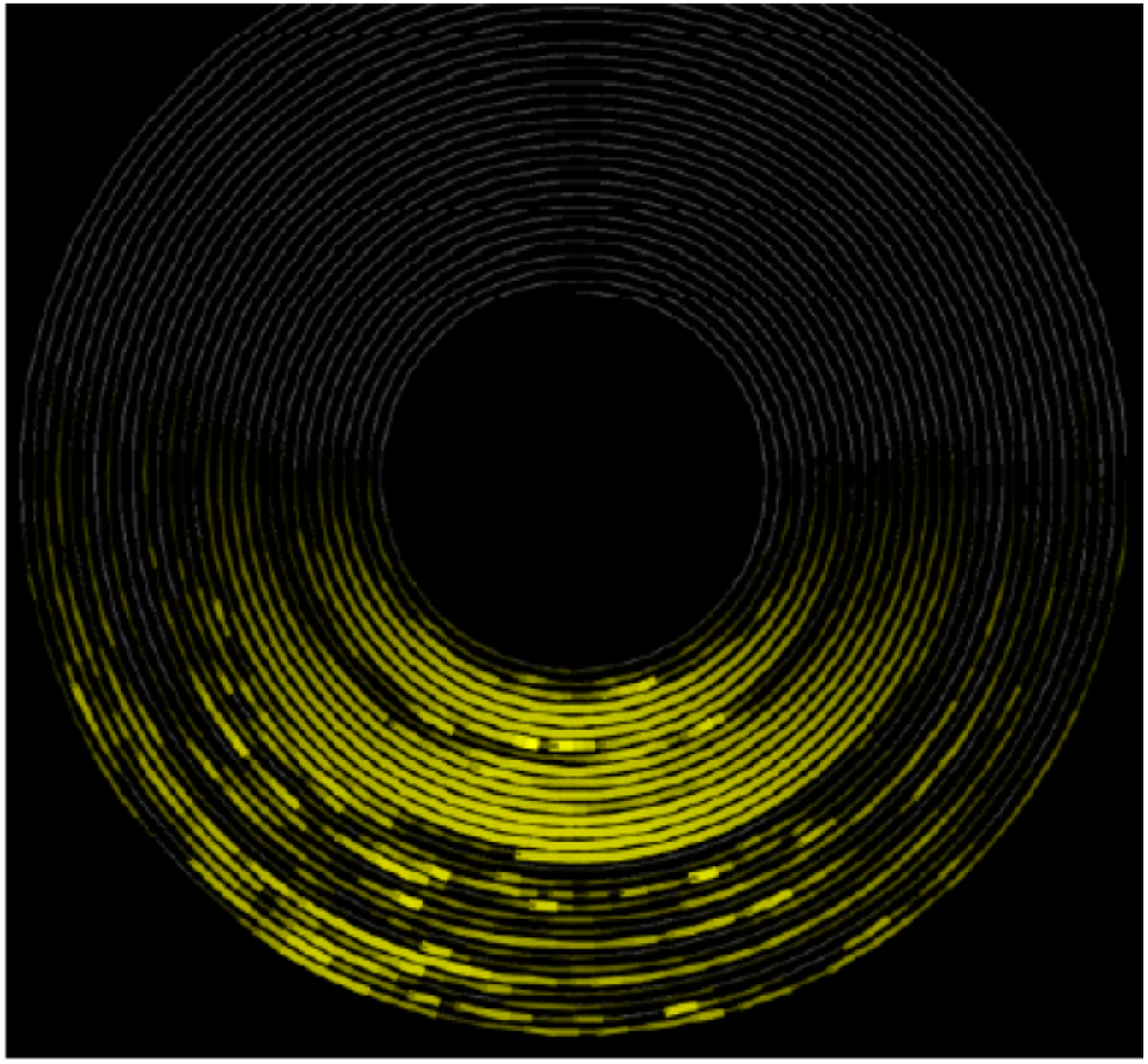
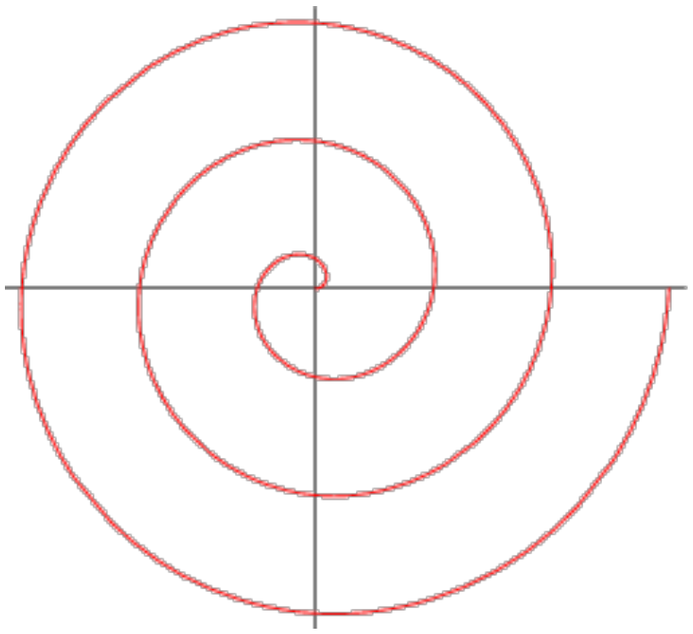


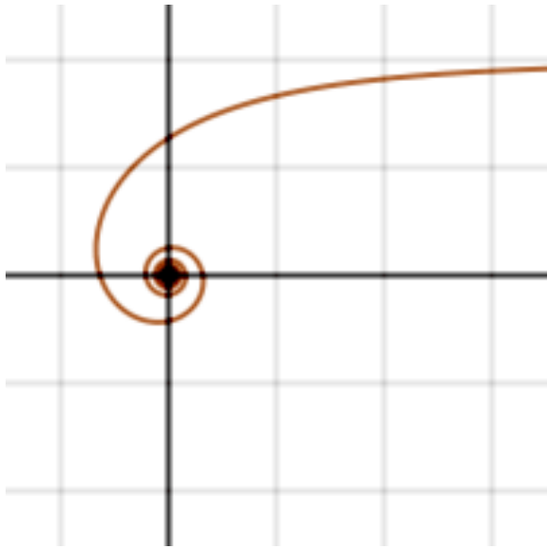
Figure 1:Two visualizations of sunshine intensity using about the same screen real estate and the same color coding scheme. In the spiral visualization it is much easier to compare days, to spot cloudy time periods, or to see events like sunrise and sunset.

Espiral de Arquimedes



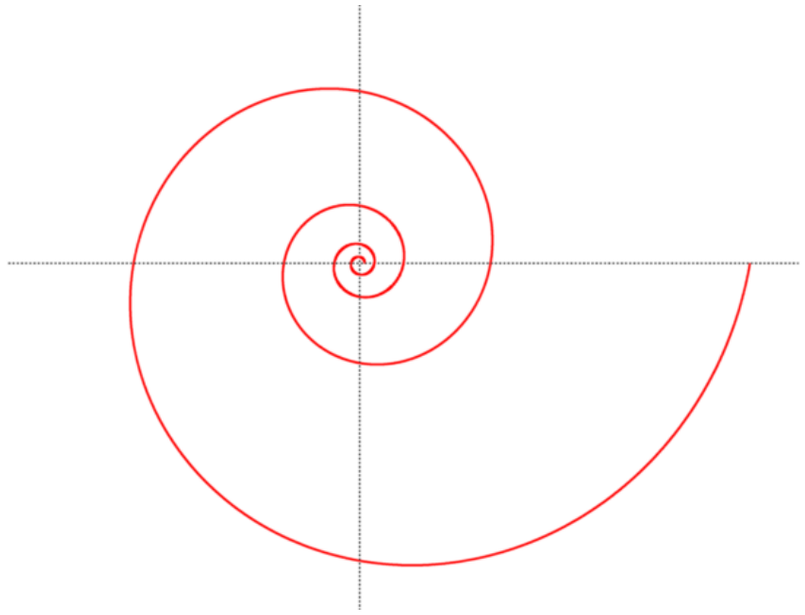
$r = a\,\theta$

Espiral de Hiperbólica



$r = a / \theta$

Espiral Logarítmica

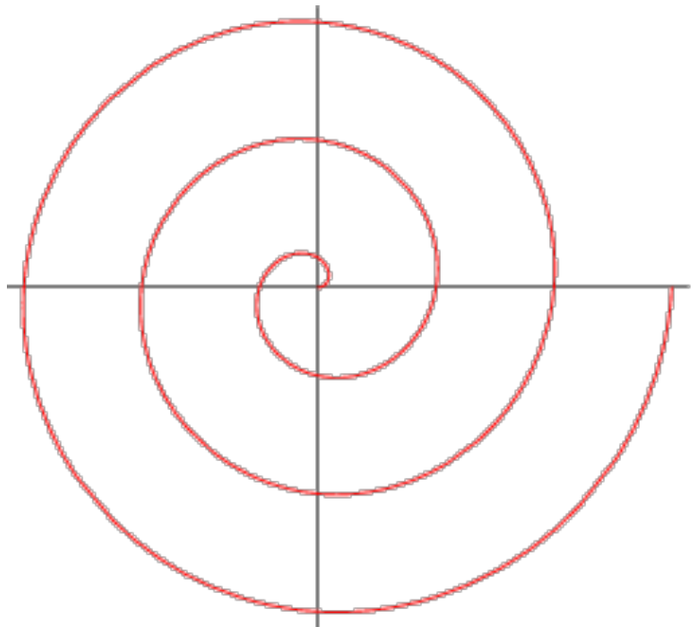


$r = a\,e^{k\,\theta}$

$x = r \cos \theta$ $\qquad$ $r = \sqrt{x^2 + y^2}$

$y = r \operatorname{sen} \theta$ $\qquad$ $\Theta = \arctan(y/x)$

Espiral de Arquimedes



$r = a\,\theta$

Dados de diferentes períodos normalmente têm a mesma importância, logo a espiral de Arquimedes é a mais apropriada

REPRESENTAÇÃO VISUAL

- Cores das linhas
- Estilo das linhas
- Espessura das linhas
- Uso de ícones

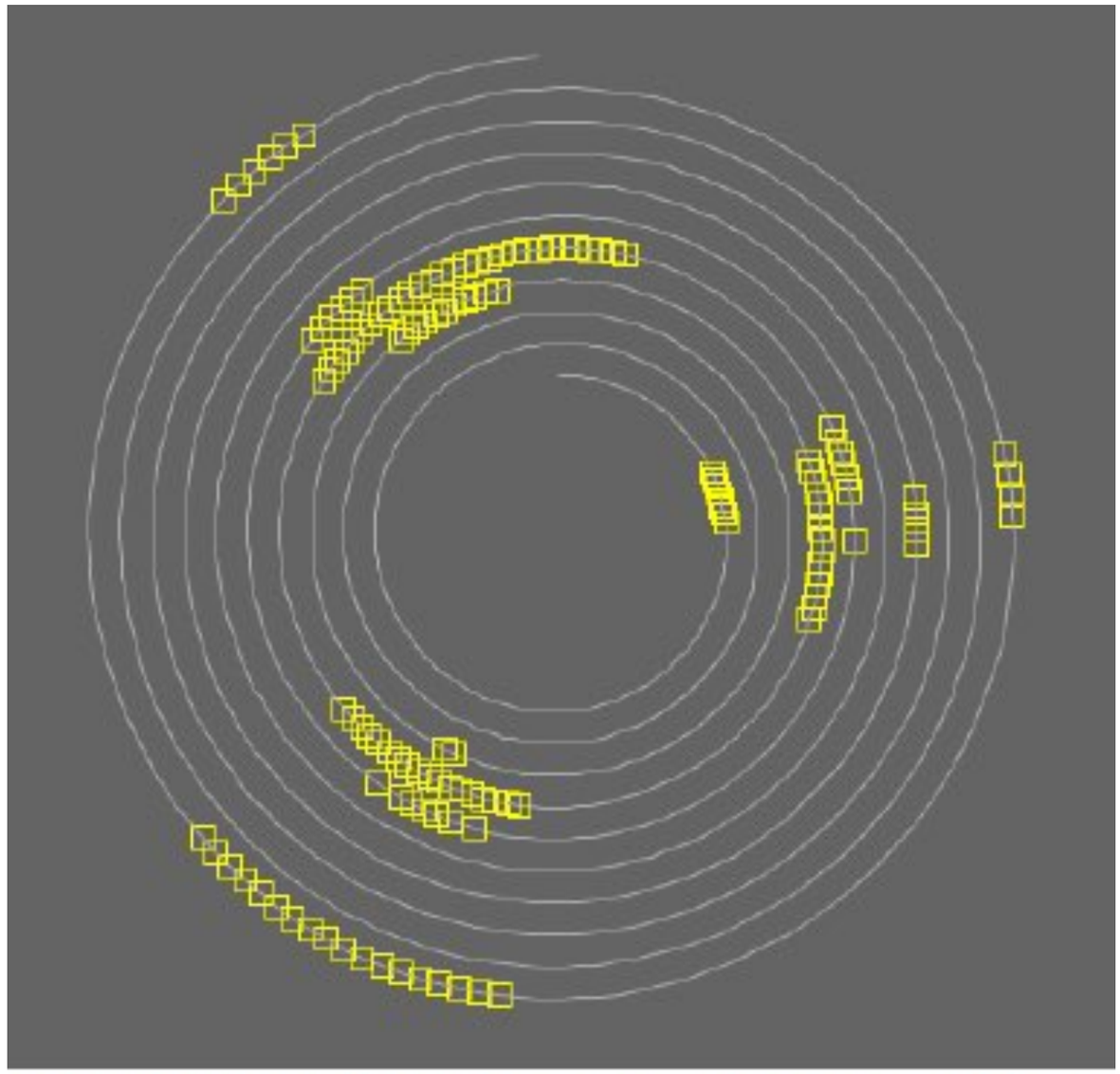
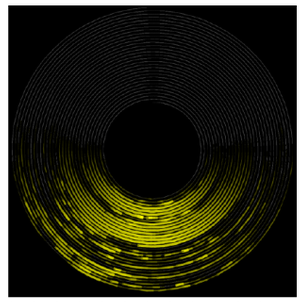


Figure 2:Nominal time-series data presented on a Spiral. The periodic behaviour of the underlying process is revealed.

DETECÇÃO DE CICLOS



- Em muitos casos, conhecemos os ciclos a serem analisados e a visualização tem grande potencial de evidenciar o comportamento periódicos dos dados
- Quando estes ciclos não são conhecidos, os períodos devem ser inferidos dos próprios dados
- Pode-se usar animações para detectar o comprimento do ciclo

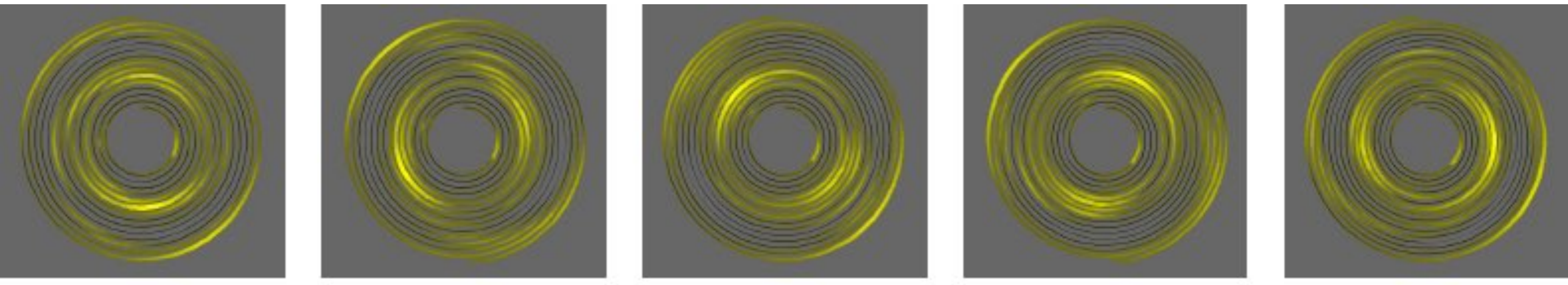


Figure 3:Visualizations of the same data with continously changing cycle length. The period in the data can be found visually, i.e. the visual system is used to detect periodic patterns in the data exploiting the spatial layout on the spiral.

MULTI-ESPIRAL

- Uso de diferentes cores para comparação entre séries
- Possível comparar até 8 séries

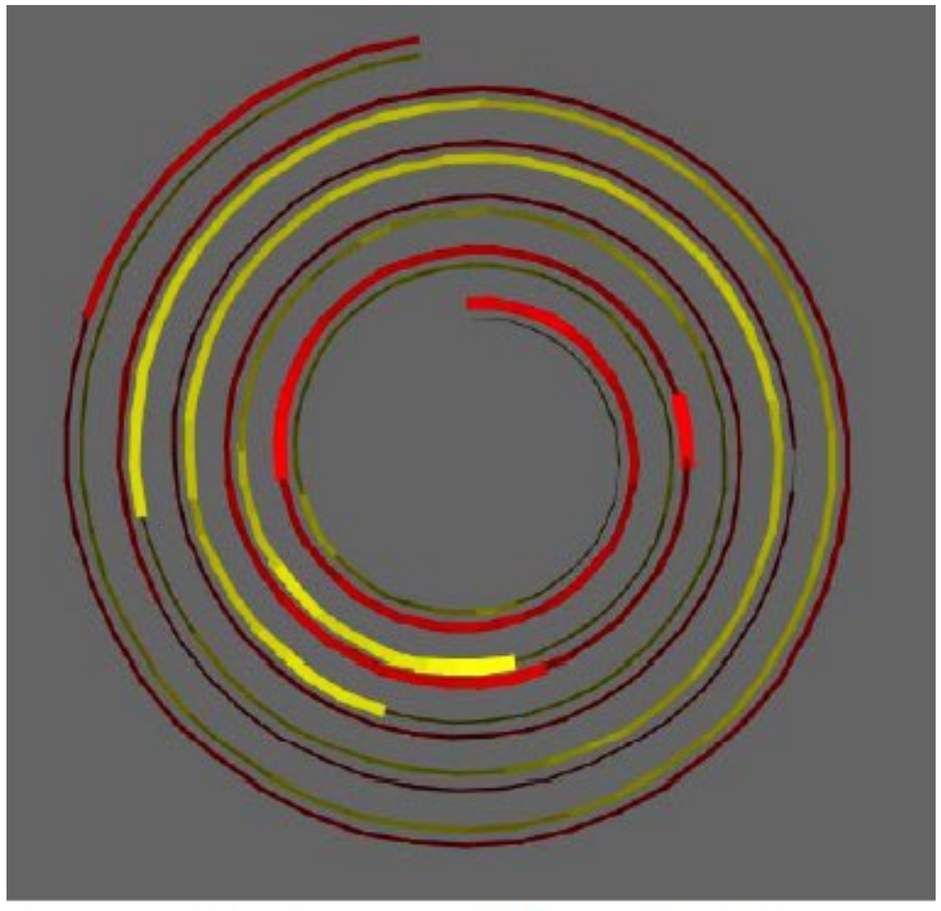


Figure 4:Stock prices of Microsoft (yellow) and Sun Microsystems (red) in five years on parallel spirals.

ESCALAS

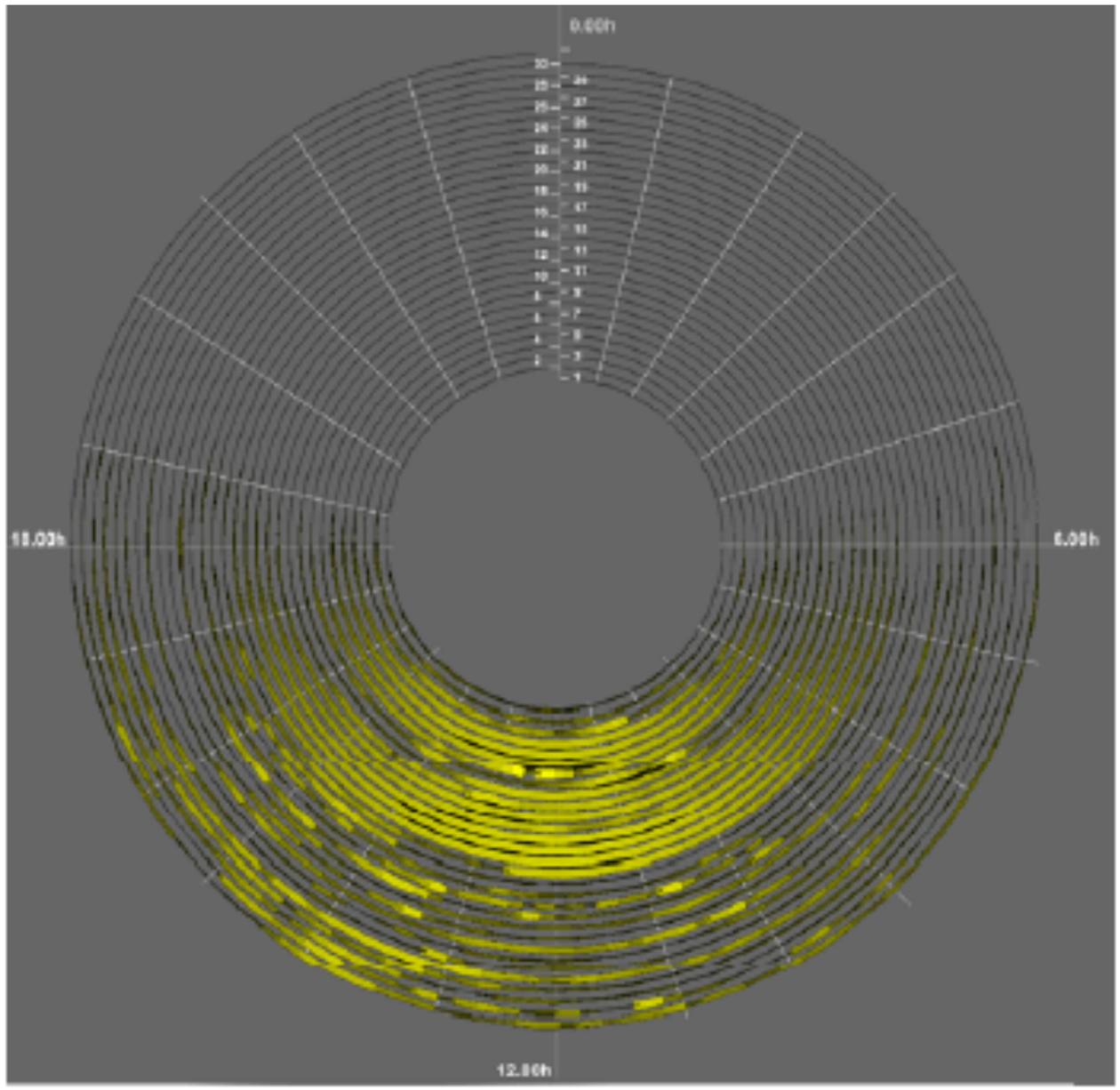
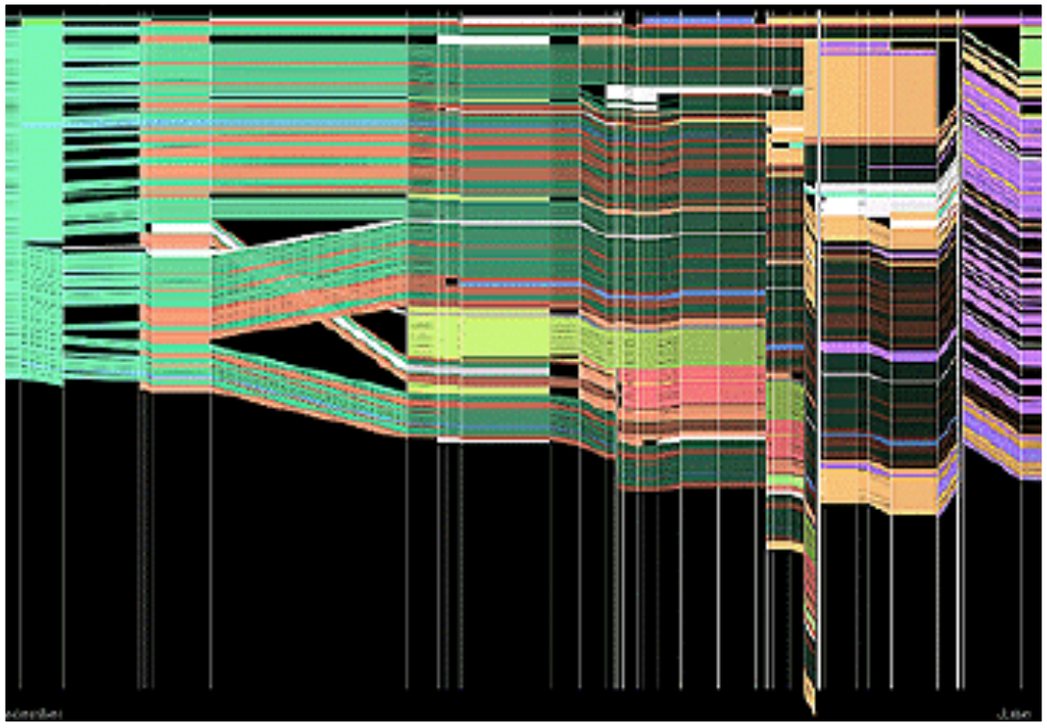


Figure 5:A possible way to add informative scales to the parametric dimensions of a spiral.

STUDYING COOPERATION AND CONFLICT BETWEEN AUTHORS WITH HISTORY FLOW VISUALIZATIONS



F.B. Viégas, M. Wattenberg e K. Dave

Conference on Human Factors in Computer Systems
2004

WIKIPEDIA

English
The Free Encyclopedia
3 600 000+ articles

日本語
フリー百科事典
741 000+ 記事

Deutsch
Die freie Enzyklopädie
1 209 000+ Artikel

Español
La enciclopedia libre
745 000+ artículos

Français
L'encyclopédie libre
1 085 000+ articles

Русский
Свободная энциклопедия
695 000+ статей

Italiano
L'enciclopedia libera
787 000+ voci

Português
A enciclopédia livre
679 000+ artigos

Polski
Wolna encyklopedia
791 000+ haseł

Nederlands
De vrije encyclopedie
679 000+ artikelen

As pessoas estão acostumadas a participar de comunidades online nas quais pessoas com perspectivas conflitantes se encontram e discutem

Na Wikipedia, ao contrário de enciclopédias tradicionais, todo o conteúdo pode ser editado por qualquer pessoa a qualquer momento

Alguns acreditam que esta estratégia é vulnerável a erros, desconhecimento e malícia não sendo uma ferramenta de referência confiável

Como este sistema aberto e vulnerável pode funcionar?

Table 1: 2005 Statistics (SAMPLE05)

Revision Type	Number	Mean time	Median time
All content	901,242	19.1 days	1113 minutes
Mass delete (MD)	4,848	7.0 days	2.9 minutes
MD obscene	105	0.13 days	2 minutes

A Wikipedia fornece os dados de todas as edições que foram realizadas nos artigos

O *History Flow* é uma visualização que mostra os relacionamentos entre as diversas versões de um mesmo artigo

O algoritmo de casamento de seções do texto é

O. Heckel. *A technique for isolating differences between files*. Communications of the ACM, 1978.

A análise exploratória desta visualização revela padrões complexos de cooperação e conflito

REPRESENTAÇÃO VISUAL

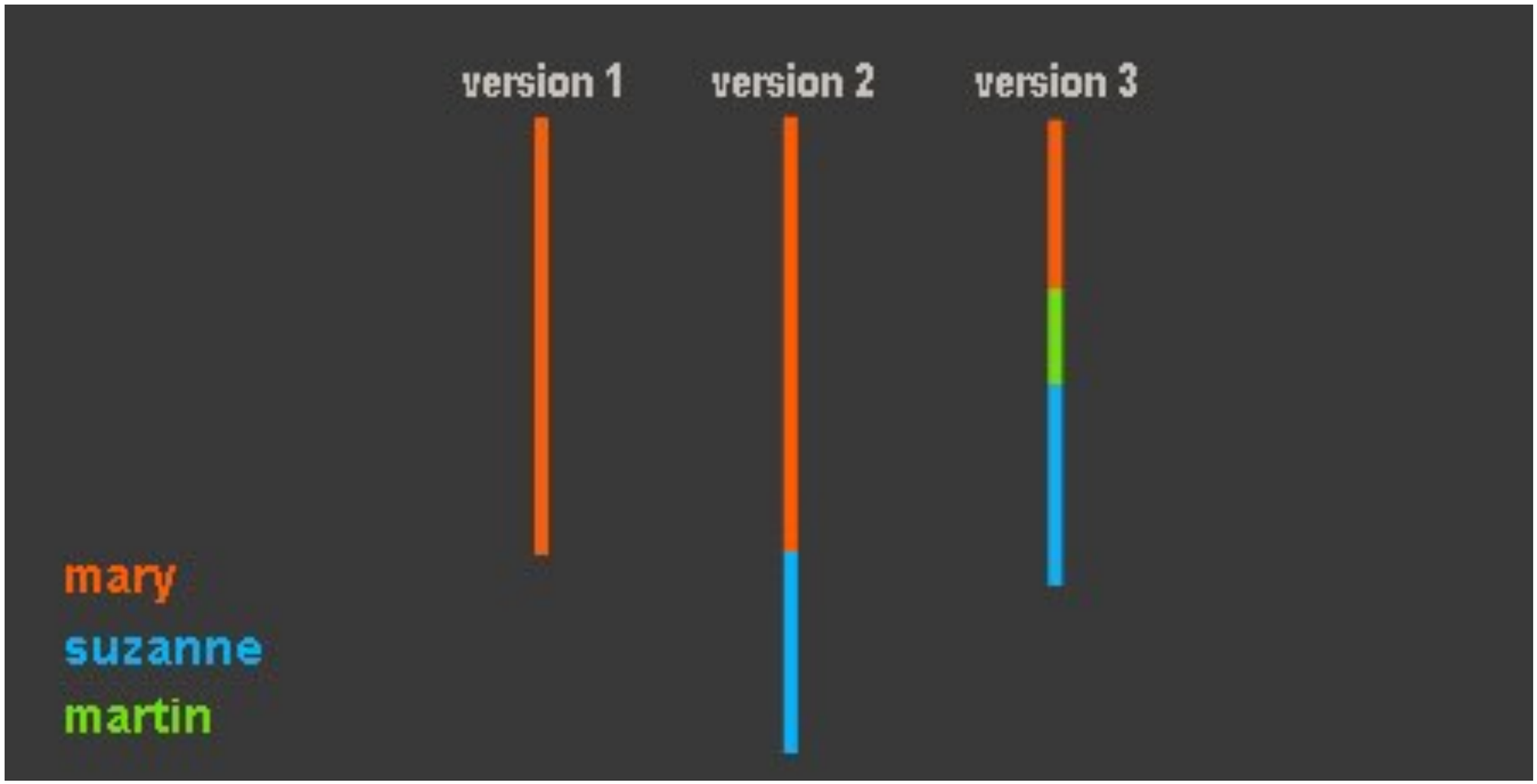
- **Cada versão do documento** é representada como uma **linha** de revisão com comprimento proporcional ao tamanho do texto

REPRESENTAÇÃO VISUAL

- Os **autores** são representados por **cores diferentes**
- **Seções das linhas** de revisão são coloridas de acordo com a cor representativa do **autor**

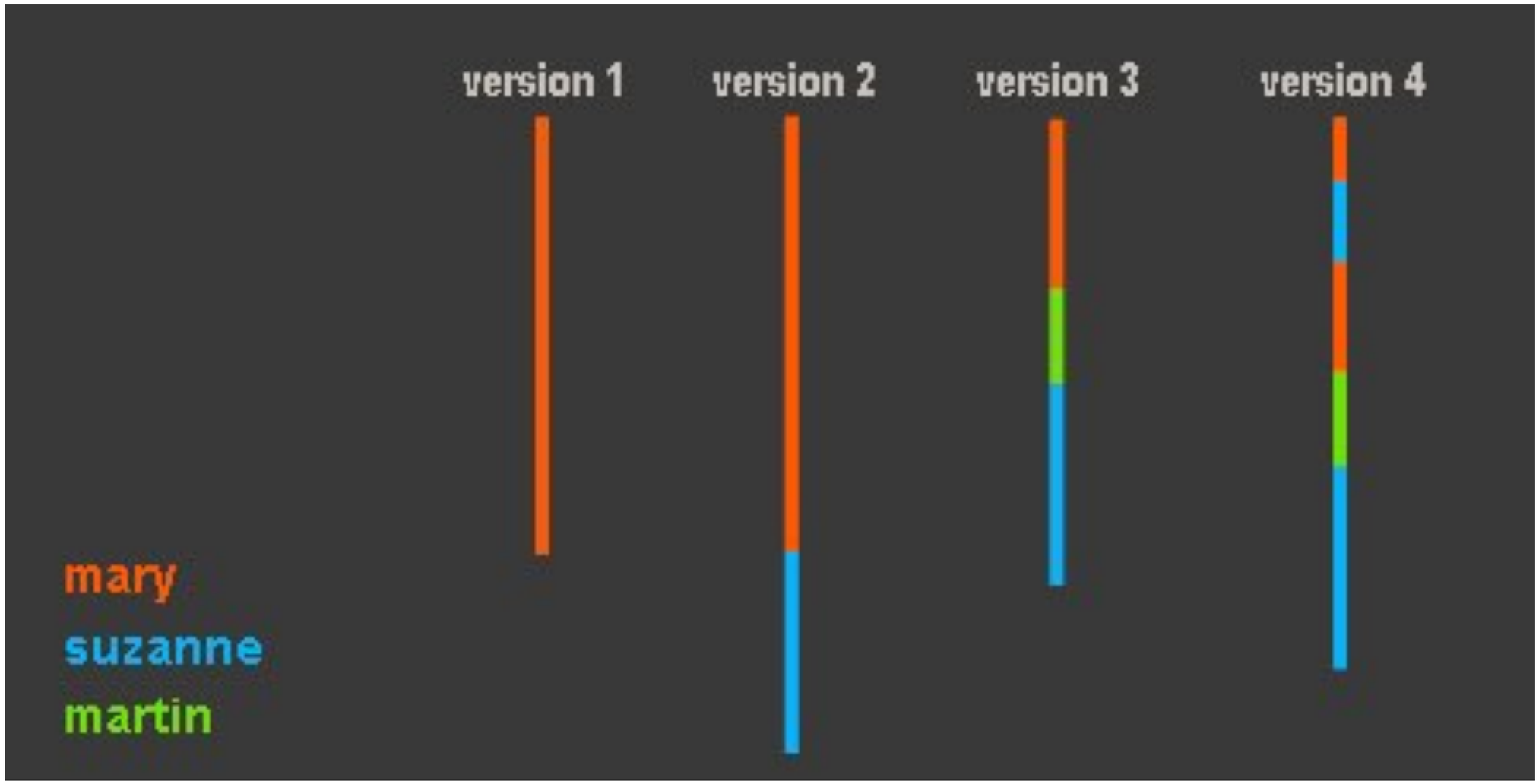
REPRESENTAÇÃO VISUAL

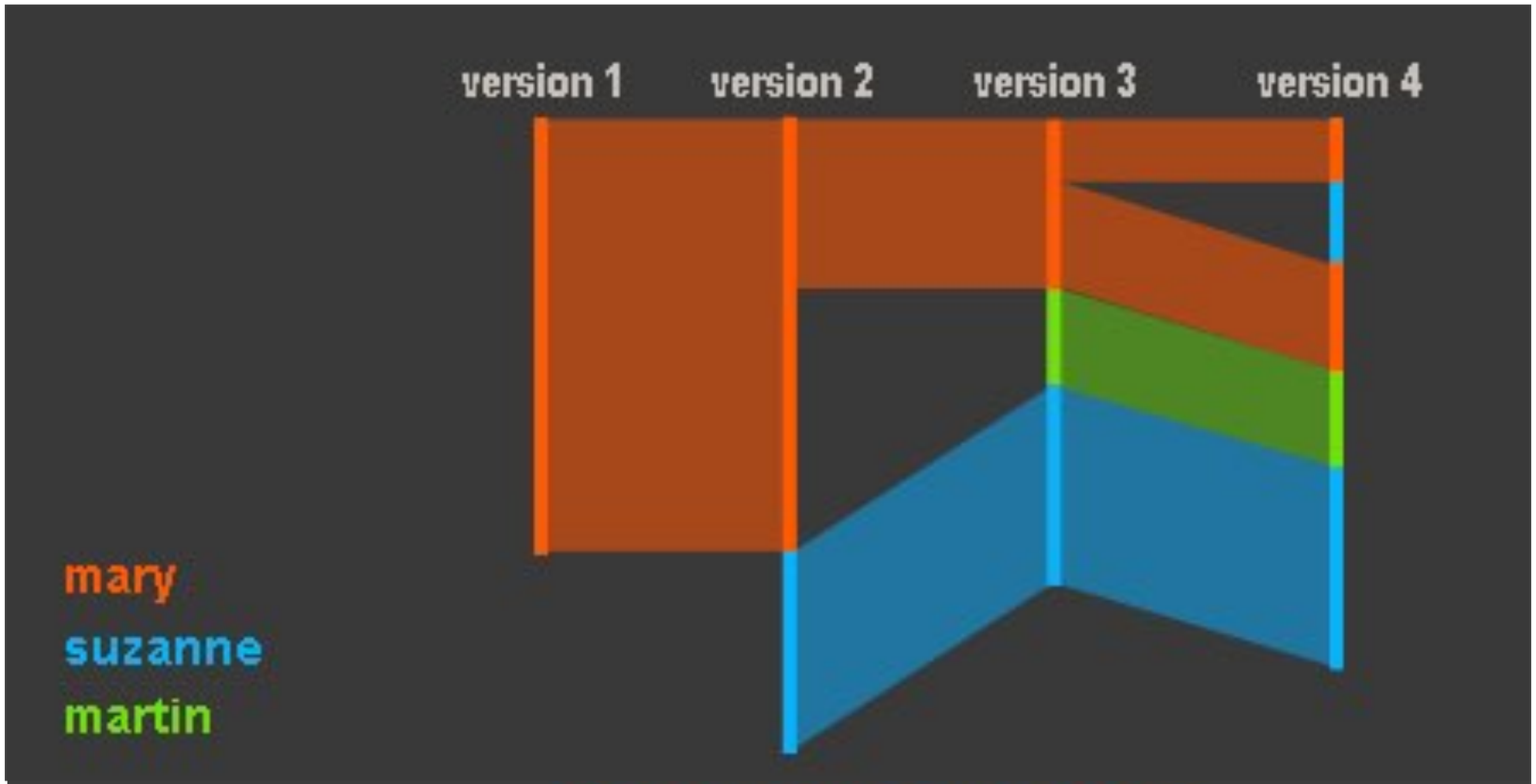
- Os **autores** são representados por **cores diferentes**
- **Seções das linhas** de revisão são coloridas de acordo com a cor representativa do **autor**

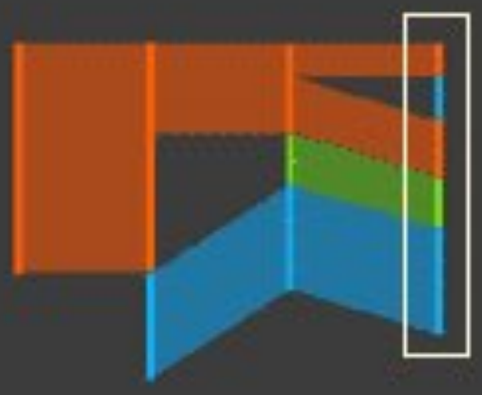


REPRESENTAÇÃO VISUAL

- Os **autores** são representados por **cores diferentes**
- **Seções das linhas** de revisão são coloridas de acordo com a cor representativa do **autor**

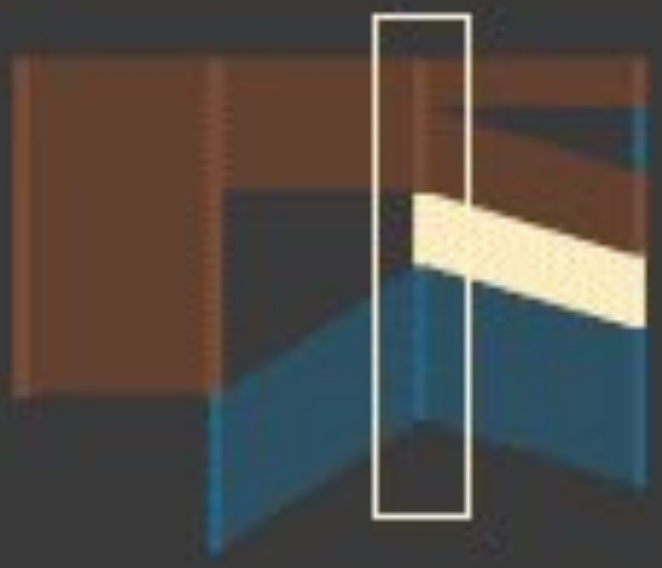






This is a visualization method for seeing the evolution of a document over time. Currently it is meant as a tool for exploratory data analysis in the WikiProject; we ourselves are the target audience. However it would be interesting to develop it further. In particular, it seems possible that it would be useful for looking at the evolution of other documents.

Example: the evolution of the page on "Abortion" on the wikipedia through several dozen version. (this is real data) Time goes left-to-right; document position is on the y-axis; each "streak" is a piece of text that remains the same from version to version.



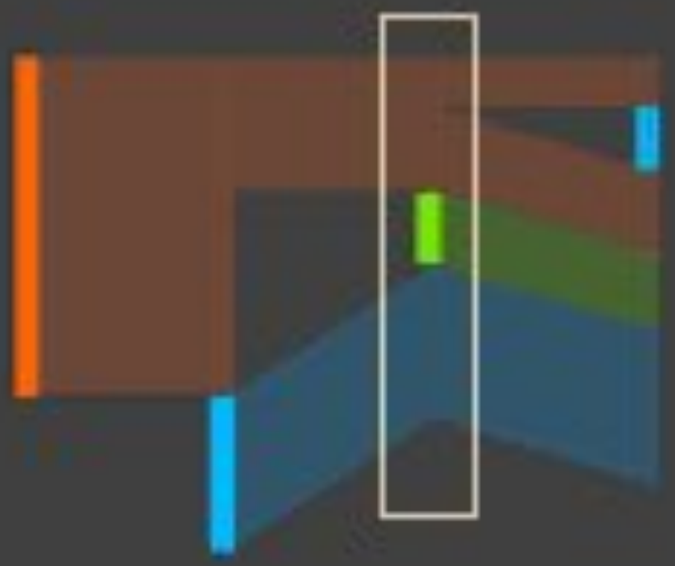
This is a visualization method for seeing the evolution of a document over time. Currently it is meant as a tool for exploratory data analysis in the WikiProject; we ourselves are the target

In particular, it seems possible that it would be useful for looking at the evolution of other documents.

Example: the evolution of the page on "Abortion" on the Wikipedia through several dozen version. (this is real data!) Time goes left-to-right, document position is on the y-axis; each "thread" is a piece of text that remains the same from version to version

Visão por autor

destaca a contribuição por autor ressaltando a persistência de suas contribuições no tempo

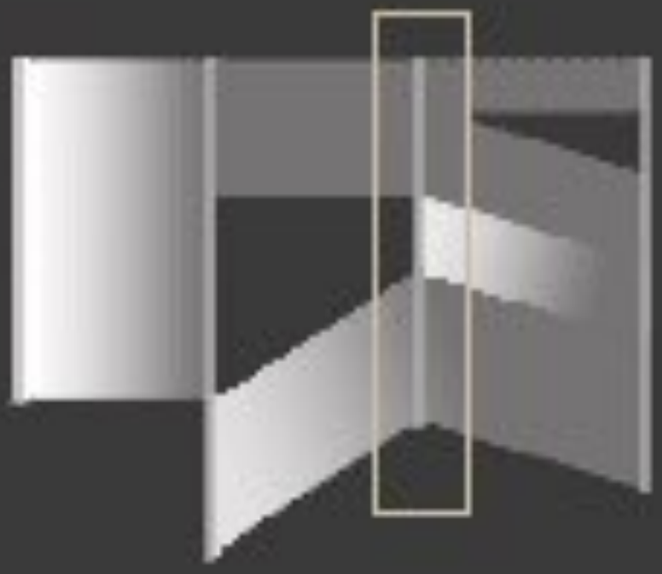


This is a visualization method for seeing the evolution of a document over time. Currently it is meant as a tool for exploratory data analysis in the WikiProject; we ourselves are the target. In particular, it seems possible that it would be useful for looking at the evolution of other documents.

Example: the evolution of the page on "Abortion" on the wikipedia through several dozen version. (this is real data!) Time goes left-to-right, document position is on the y axis; each "streak" is a piece of text that remains the same from version to version.

Visão por contribuições recentes

ressalta o novo conteúdo em cada nova versão dos artigos. Permite identificar os trechos do código mais ativamente editados



This is a visualization method for seeing the evolution of a document over time. Currently it is meant as a tool for exploratory data analysis in the WikiProject; we ourselves are the target In particular, it seems possible that it would be useful for looking at the evolution of other documents.

Example: the evolution of the page on "Abortion" on the wikipedia through several dozen version. (this is real data) Time goes left-to-right; document position is on the y-axis; each "streak" is a piece of text that remains the same from version to version.

Visão por idade

Destaca a persistência das contribuições.

Branco = nova contribuição

Cinza = contribuição antiga

Há a versão igualmente espaçada e outra espaçada de acordo com a data de edição

Interessante pois enfatiza as versões mais estáveis

Revela o ritmo das colaborações

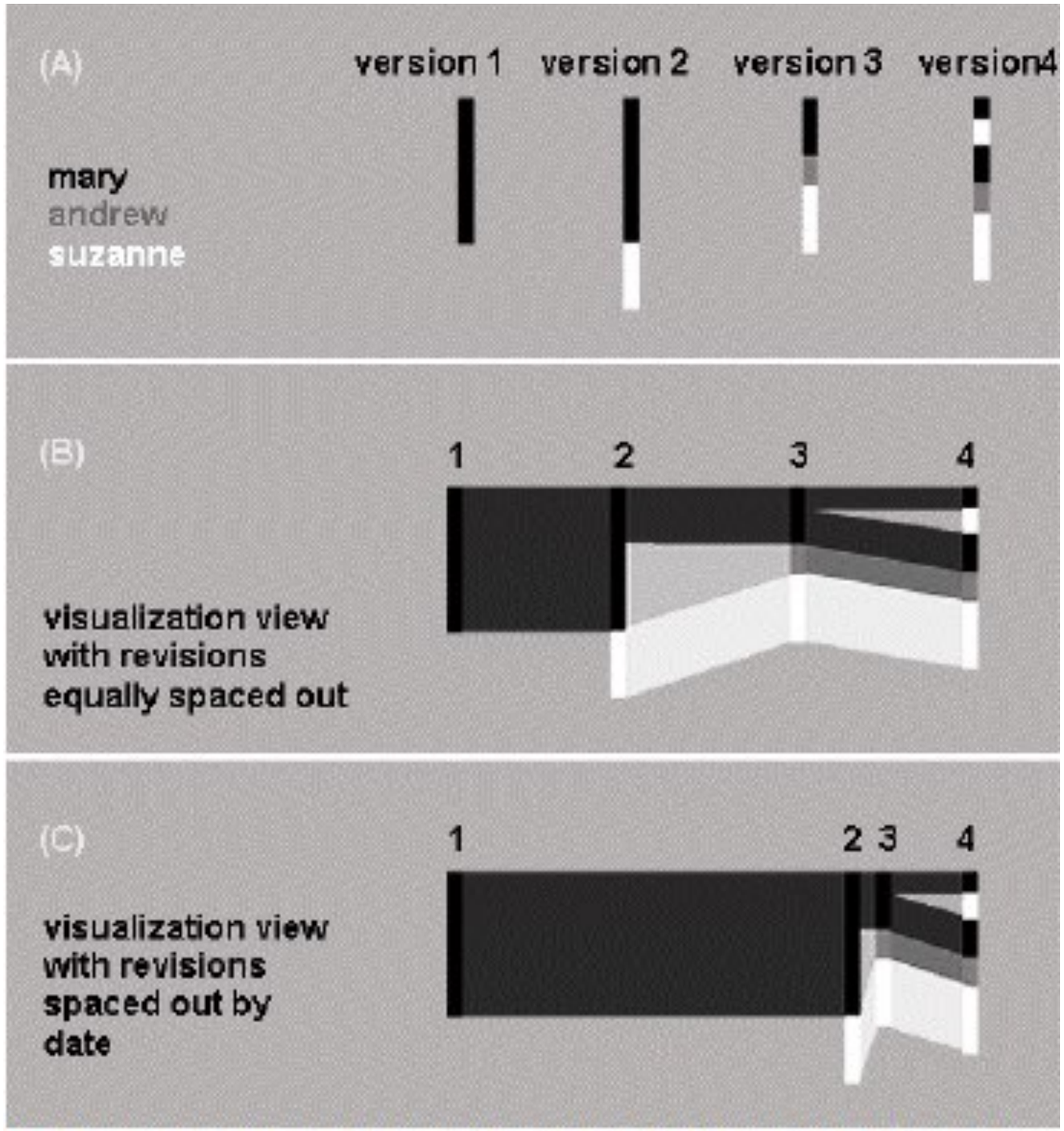
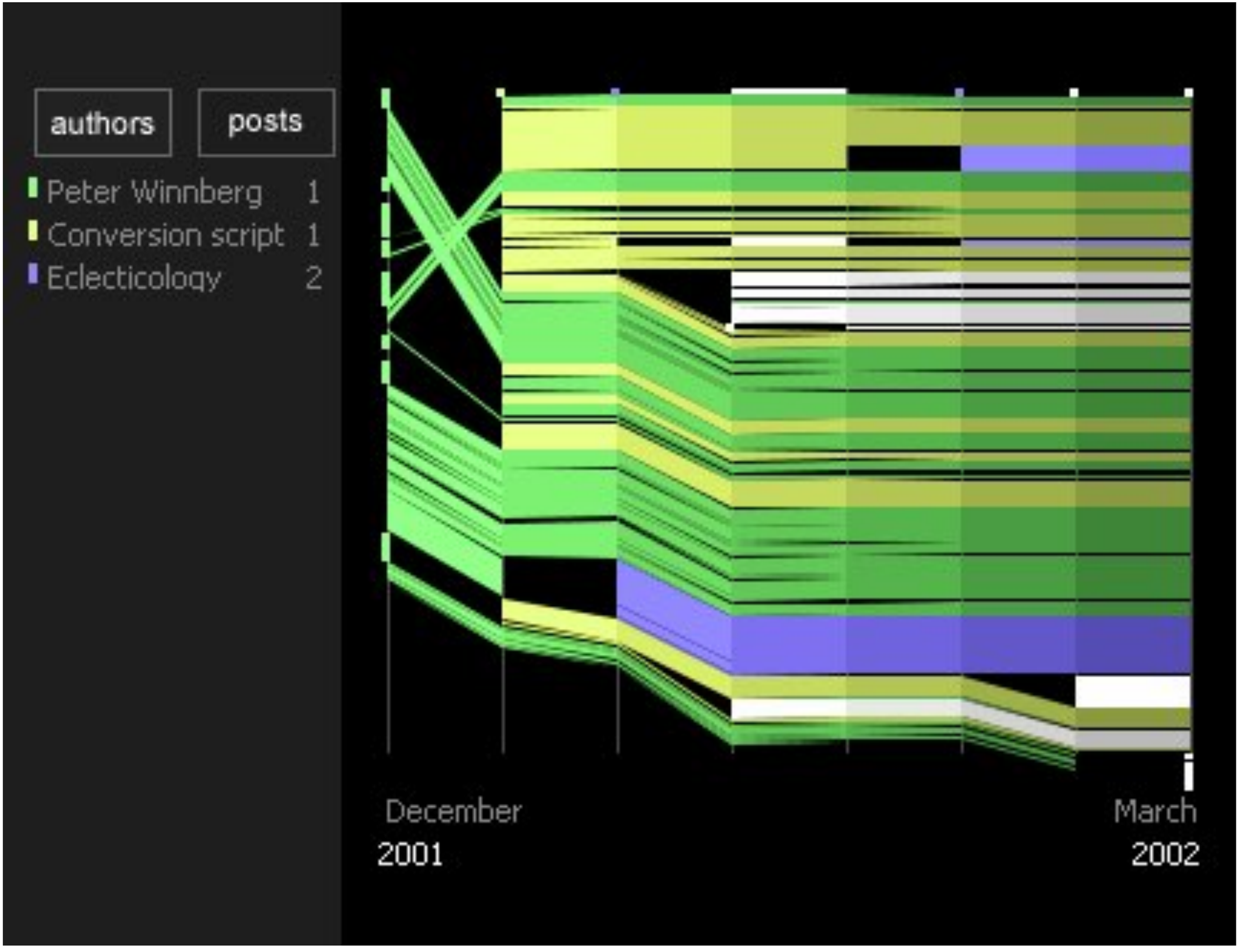
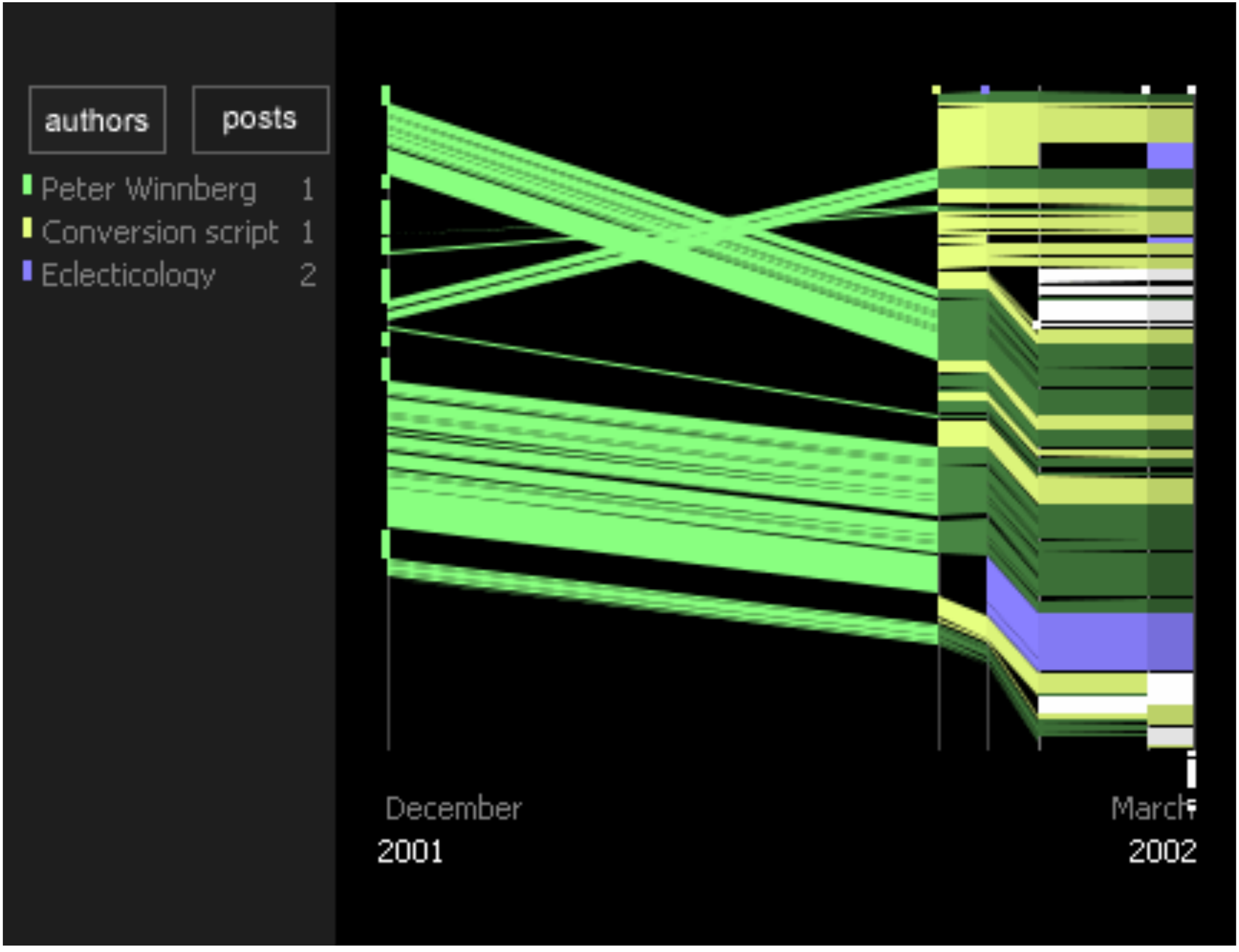
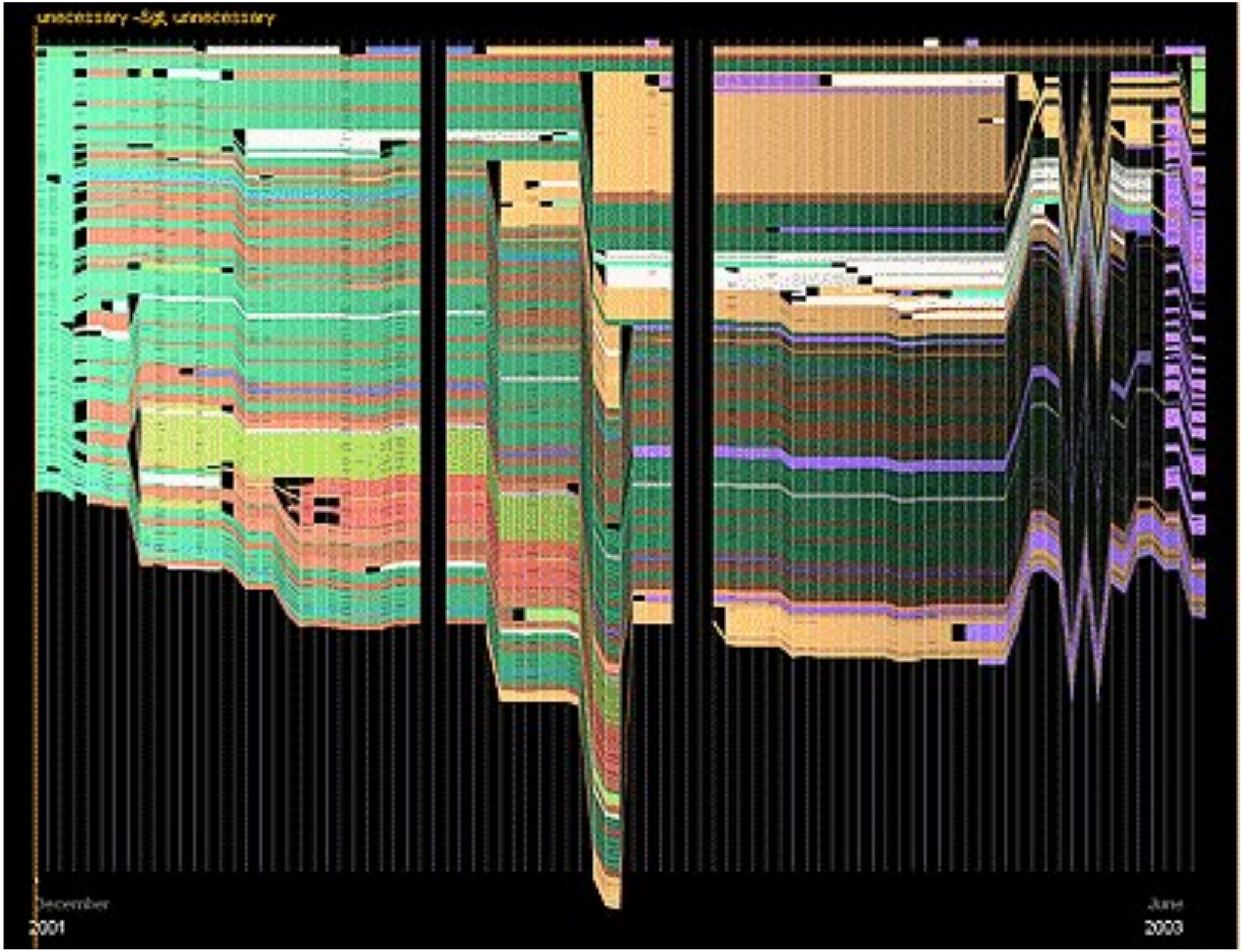


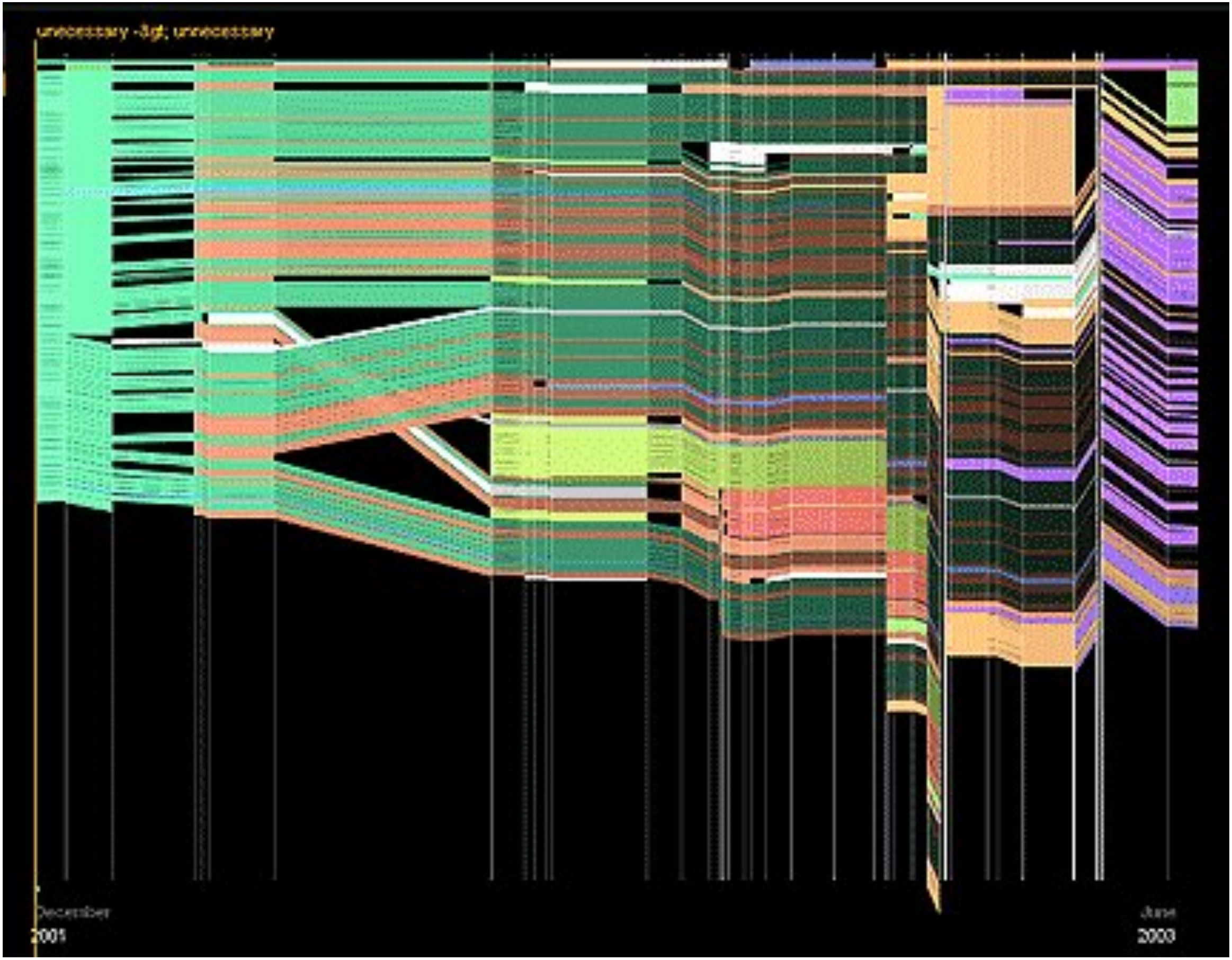
Fig 2: explanation of history flow's visualization mechanism


authors
posts
Peter Winnberg 1
Conversion script 1
Eclecticology 2
December
2001
March
2002


authors
posts
Peter Winnberg 1
Conversion script 1
Eclecticology 2
December
2001
March
2002



Páginas que tratam de temas **controversos** muitas vezes sofrem **vandalismo**, o que fica evidente com as faixas em preto mostrando deleções


unnecessary -3g#; unnecessary
December
2001
June
2003

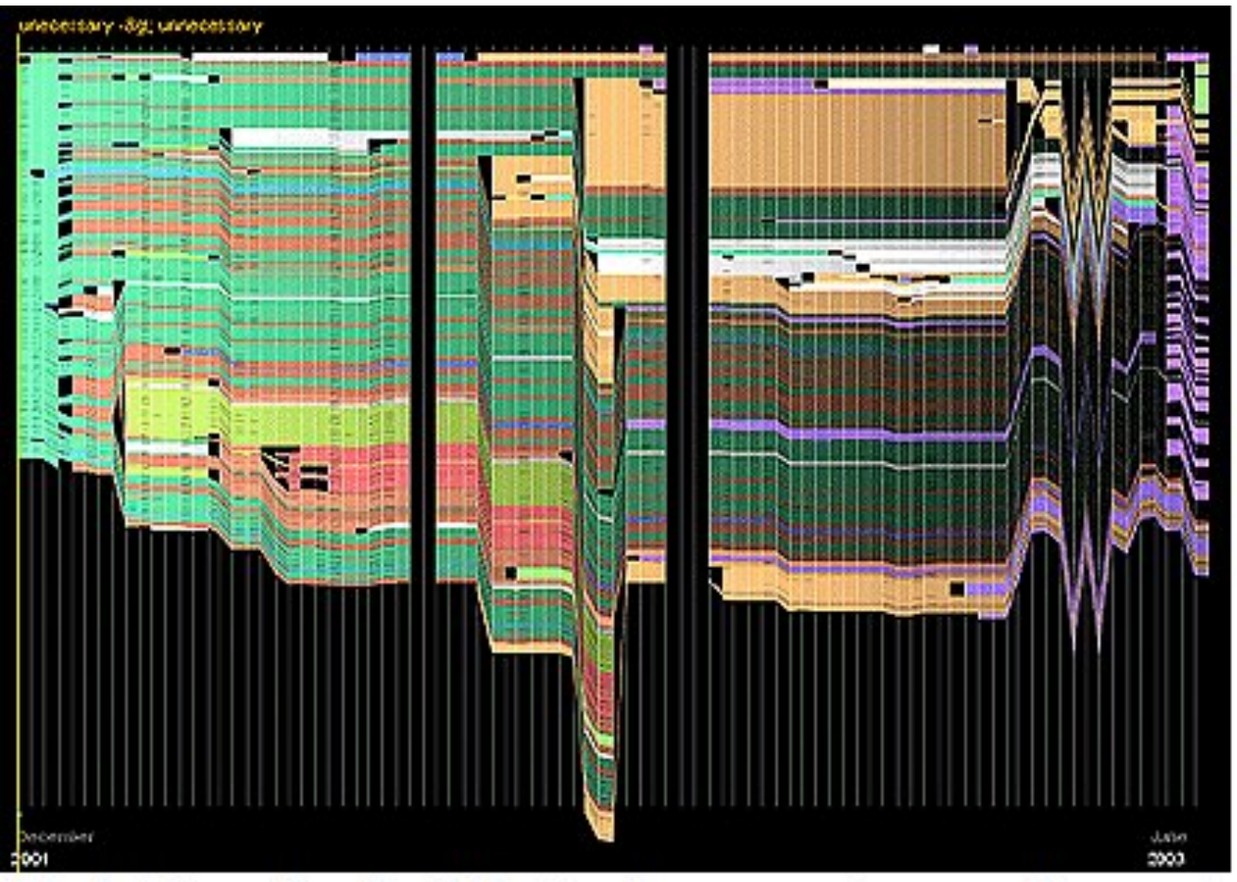


Fig 4: history flow for "Abortion" page, versions equally spaced.

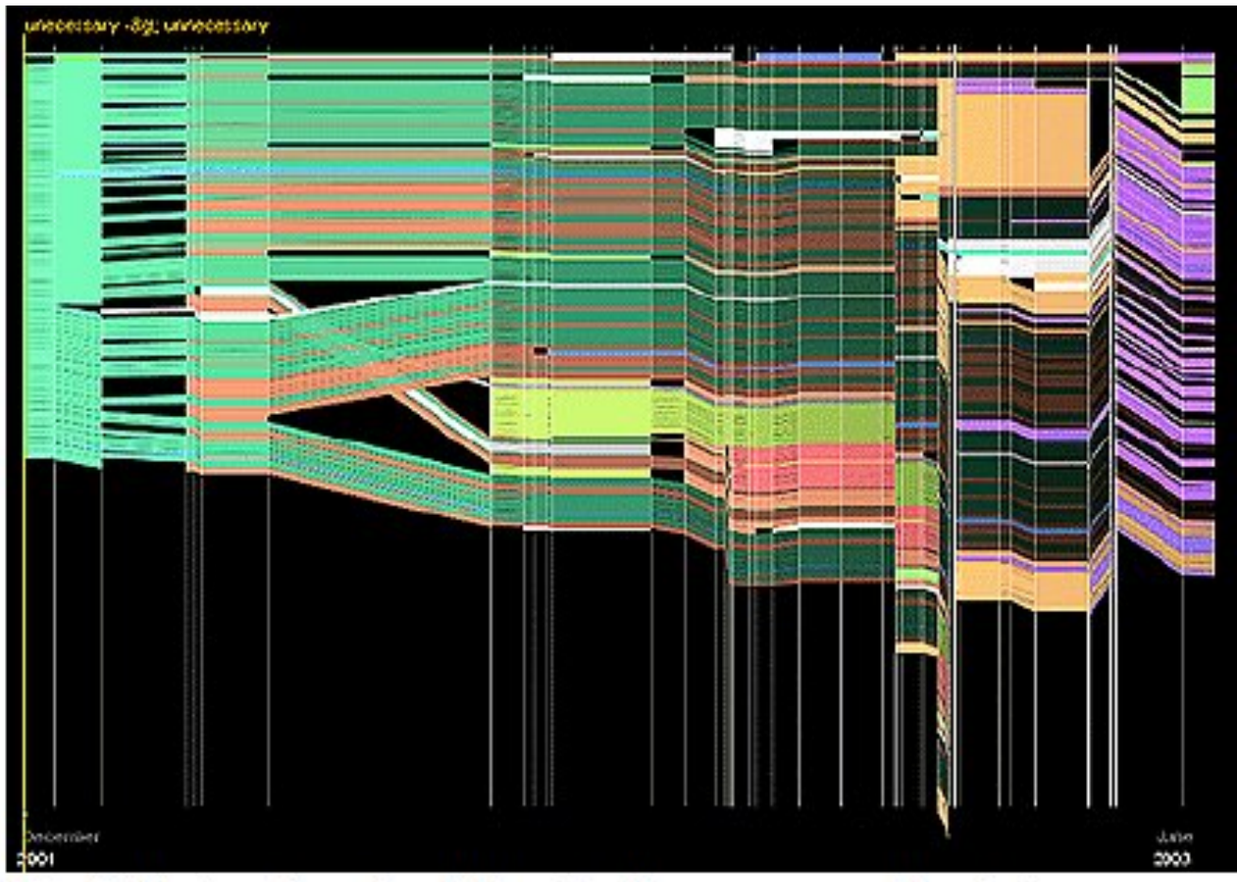
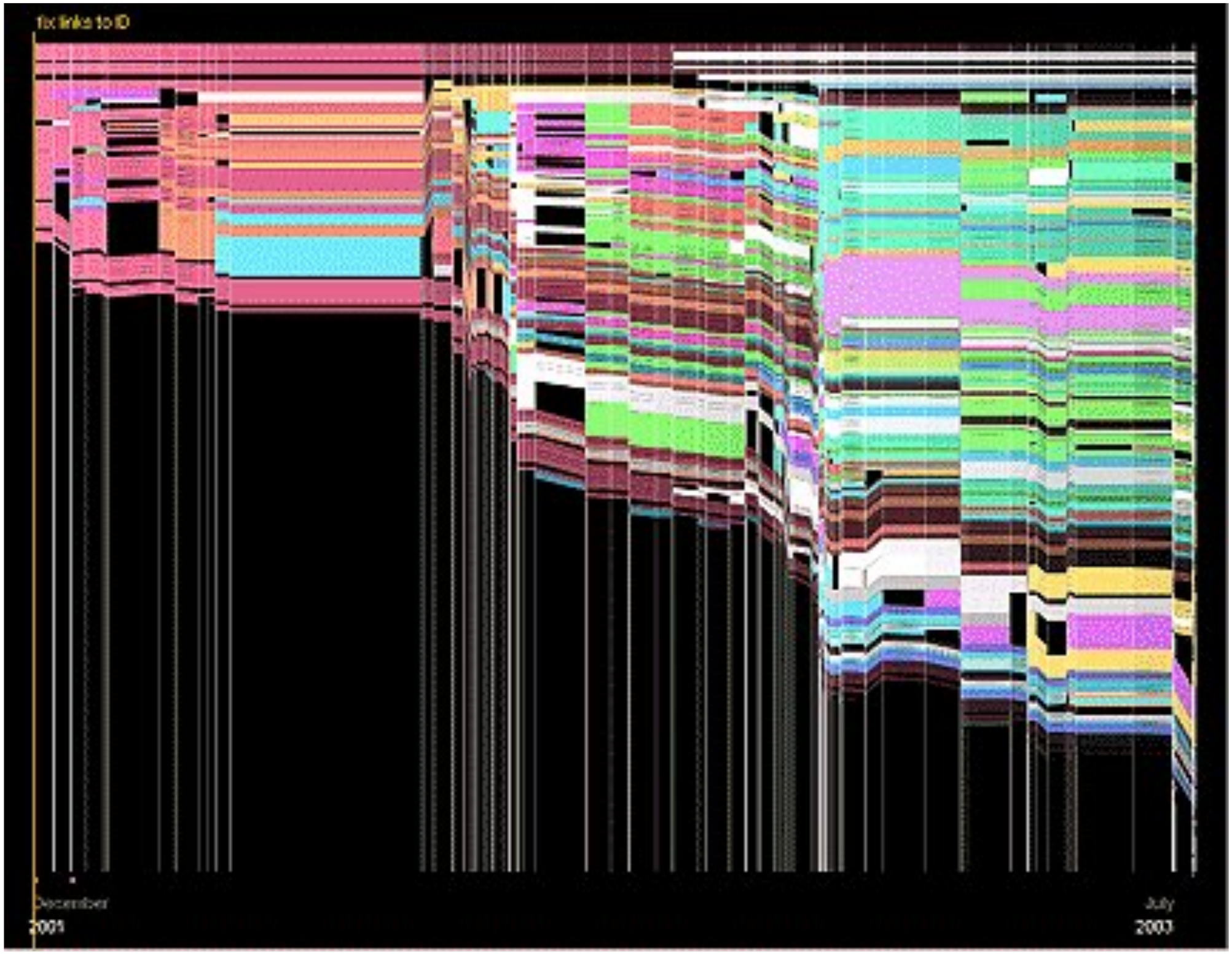
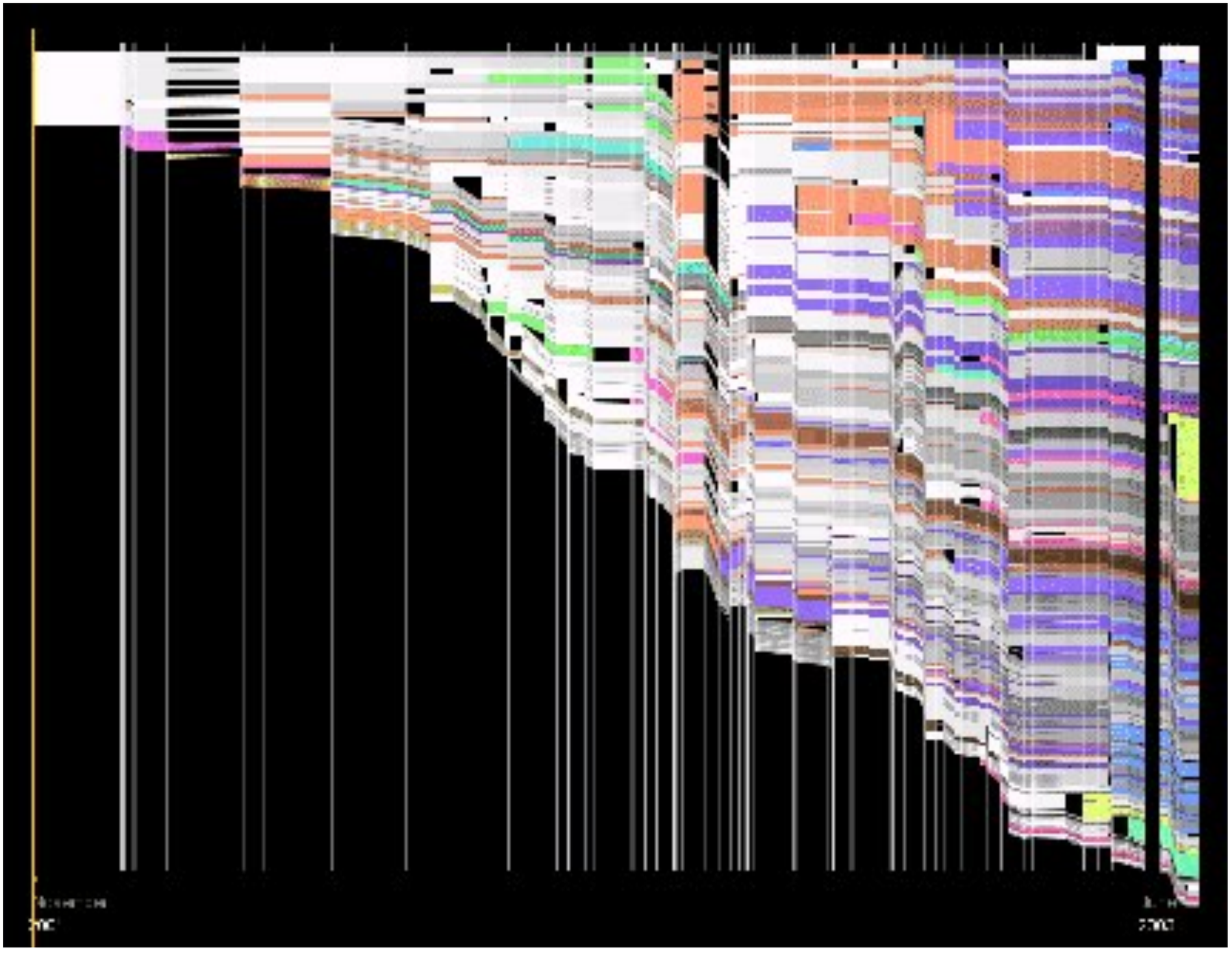


Fig 5: history flow for "Abortion" page, spaced by date

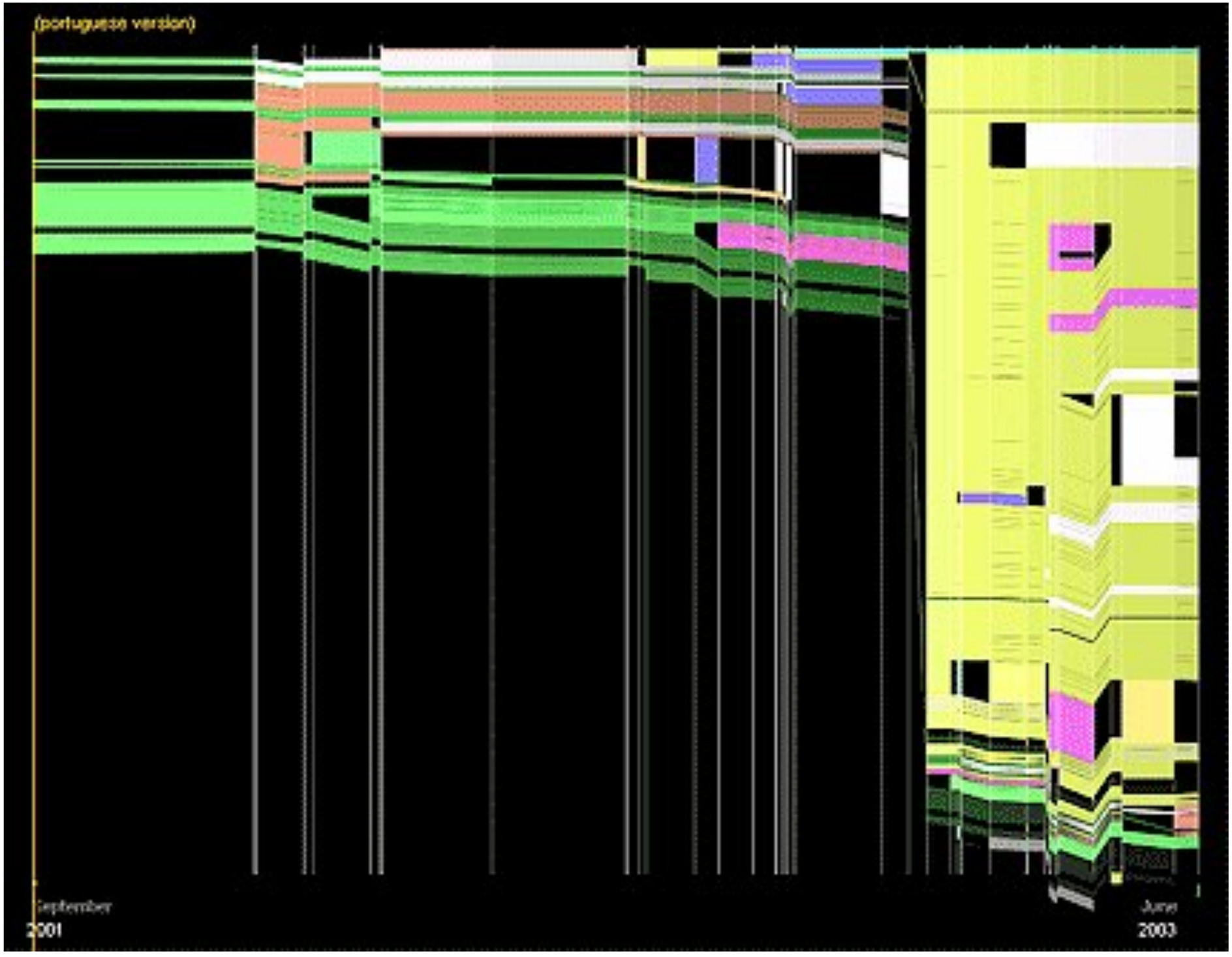


No history flow, pode-se ver o histórico de contribuições por autor e juntamente pode-se ver os autores anônimos branco



No history flow, pode-se ver o histórico de contribuições por autor e juntamente pode-se ver os autores anônimos branco

Artigo sobre a Microsoft



No history flow, pode-se ver o histórico de contribuições por autor e juntamente pode-se ver os autores anônimos branco

Artigo sobre o Brasil

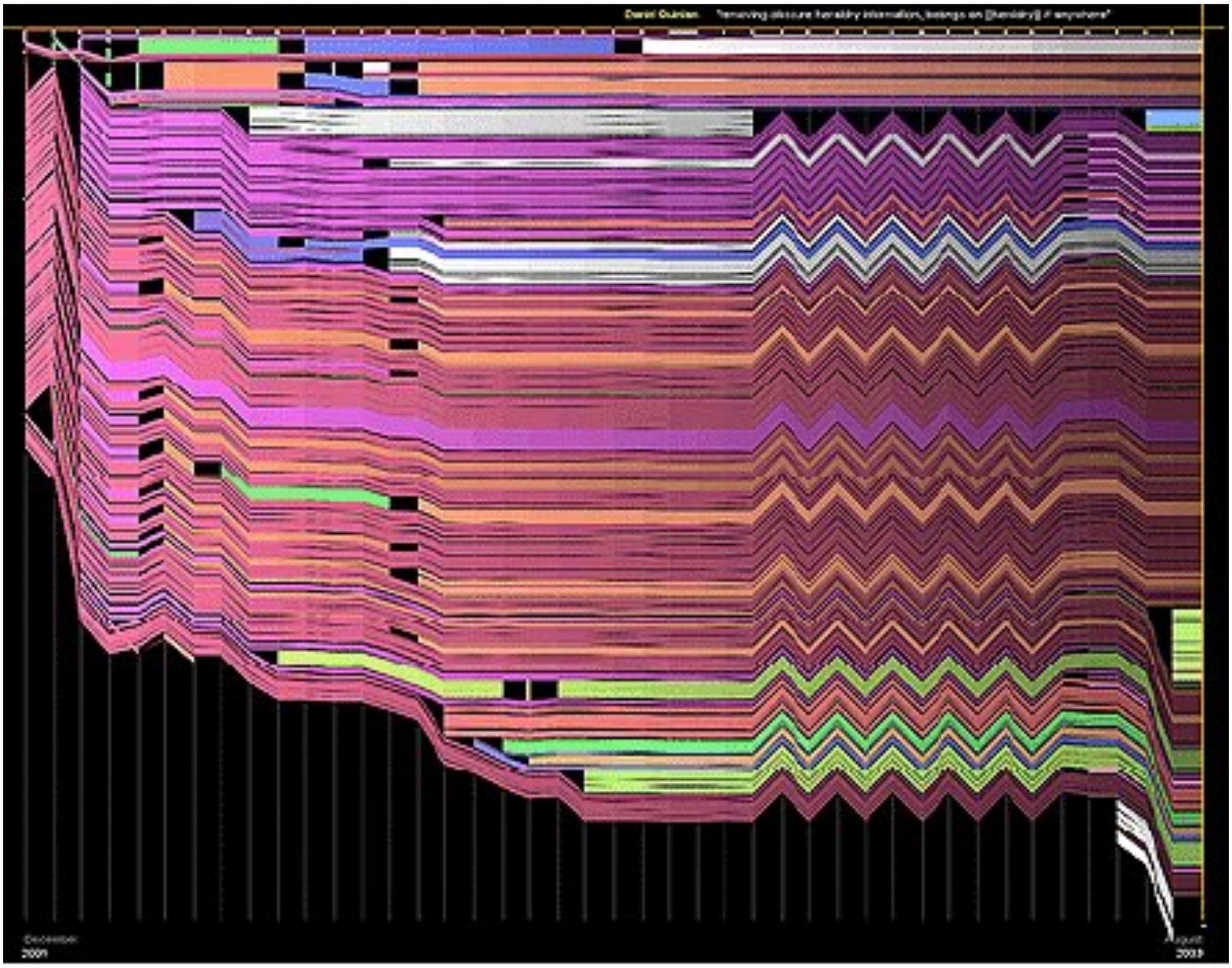


Fig 6: "Chocolate" page spaced out by number of versions; we can see the zigzag pattern of an edit war.

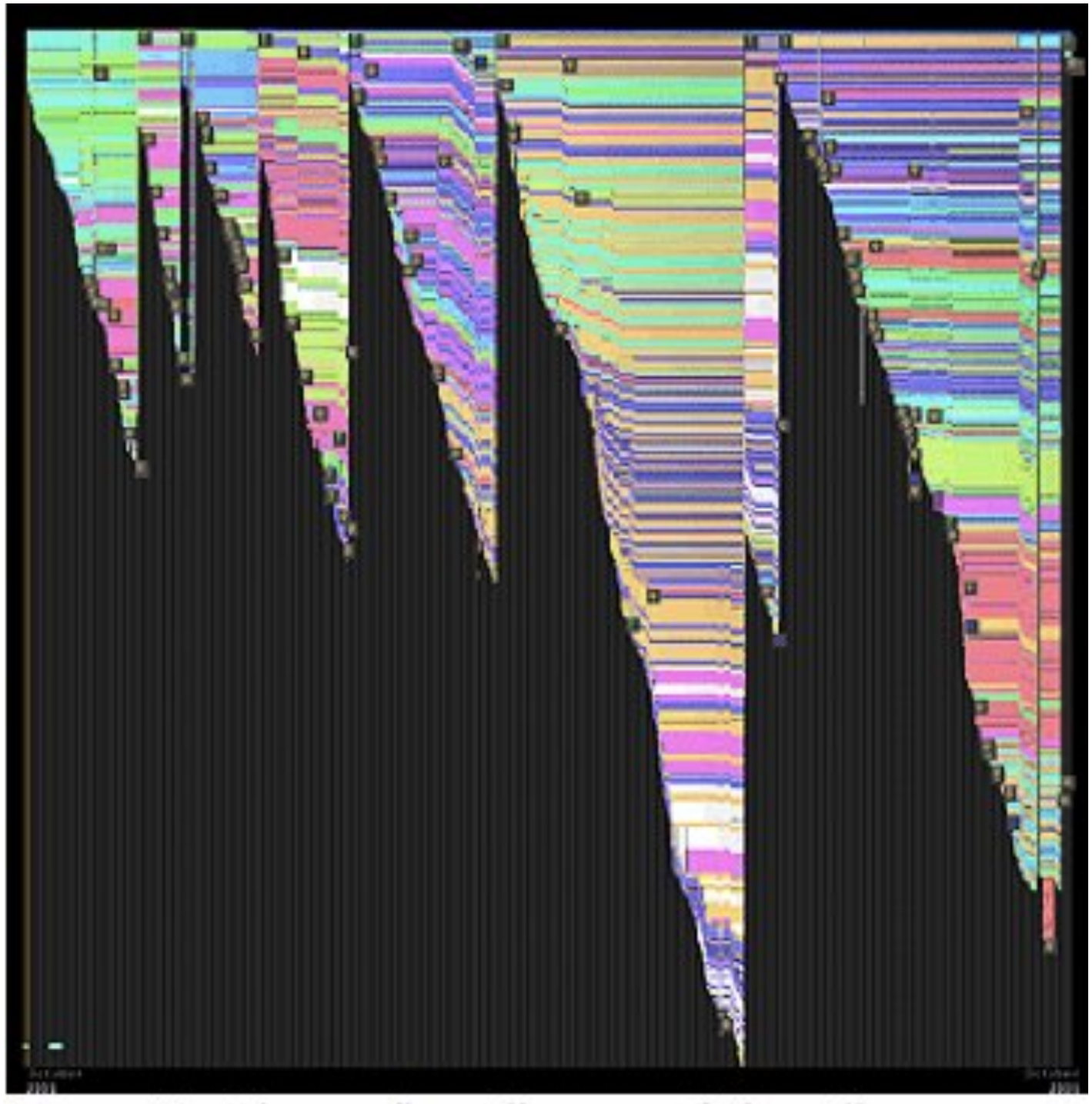
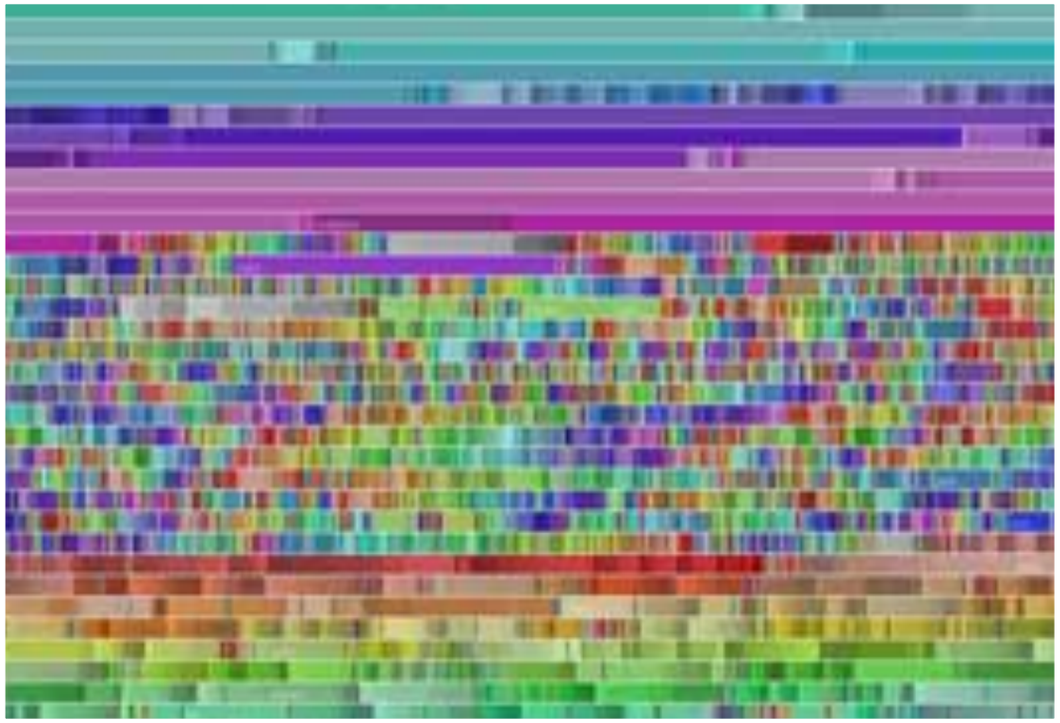


Figure 7: History flow diagram of the talk page on "God," showing the cyclical nature of archiving operations. As the increasing length of the page reaches unwieldy proportions, the discussion is archived and the live talk page returns to a more manageable size.

VISUALIZING ACTIVITY ON WIKIPEDIA WITH CHROMOGRAMS



M. Wattenberg, F. Viégas e K. Hollenbach

Interact
2007

O objetivo da visualização é analisar os hábitos de edição de **usuários administradores da Wikipedia**

Administradores são usuários com permissões especiais para deleção e proteção de páginas

Normalmente são usuários ativos, confiáveis e influentes

DADOS

- Históricos das edições de 509 administradores

- Os históricos consistem em uma sequência de eventos com horário de ocorrência

- Os eventos são os títulos dos artigos e comentários opcionais

- Comentários curtos e seguindo alguma padronização

DESAFIOS

- Volume: 100.000 eventos

- Diversos: 80.000 tokens distintos

- Estrutura irregular

REPRESENTAÇÃO VISUAL

- A primeira letra indica a cor
- A segunda letra indica a saturação
- A terceira letra indica o brilho

a

Time	Page	Comment
May 21, 9:32 am	Sphere	Add cite
May 22, 10:56 am	Sphere	New Intro
May 22, 1:23 pm	sphere.png	Copyright
May 22, 2:54 pm	Helix	Spell check
May 22, 3:00 pm	Mathematics	Revert
May 24, 11:21 am	Fields Medal	2006 data
May 24, 11:25 am	Talk:Fractal	List of proposed changes
May 24, 11:27 am	Talk:Fractal	List of proposed changes
May 25, 10:13 am	Sphere	Fix tpyo
May 25, 10:23 am	mercator.png	Copyright

b **Chromogram of comments: Timeline View**

c **Block View**

d **Block View (compressed, no labels)**

Fig. 2. Creating a Chromogram

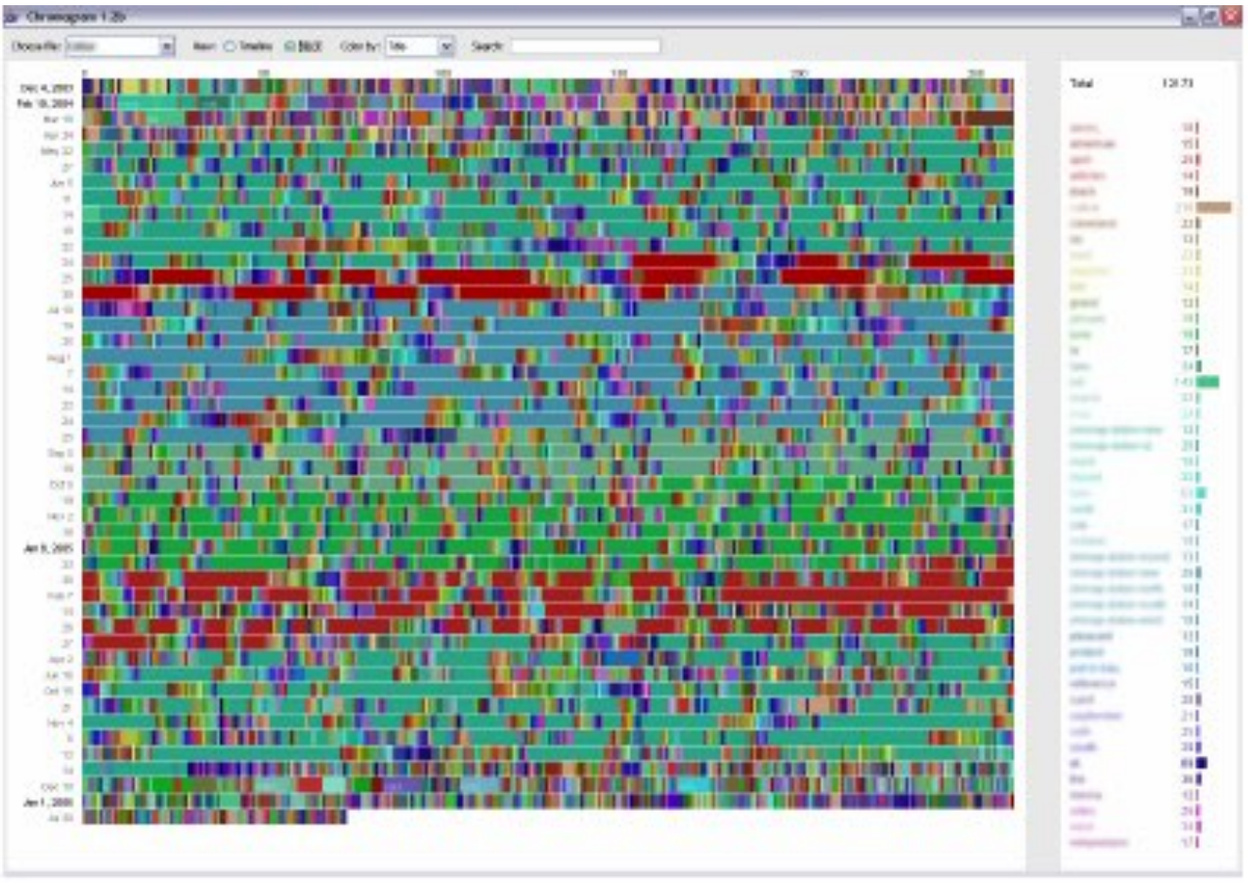


Fig. 3. The Chromogram Application: Block View

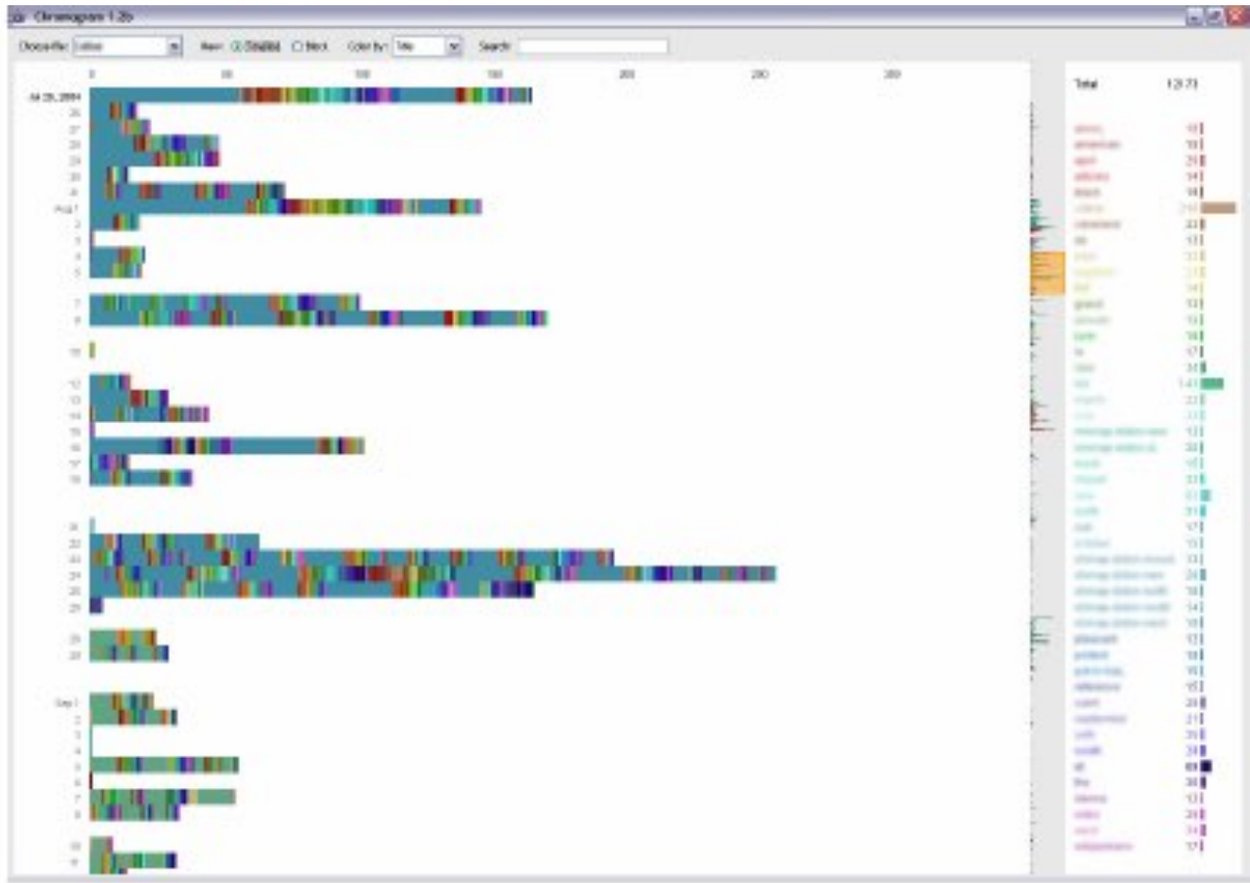


Fig. 4. The Chromogram Application: Timeline View, same data as Fig. 3.

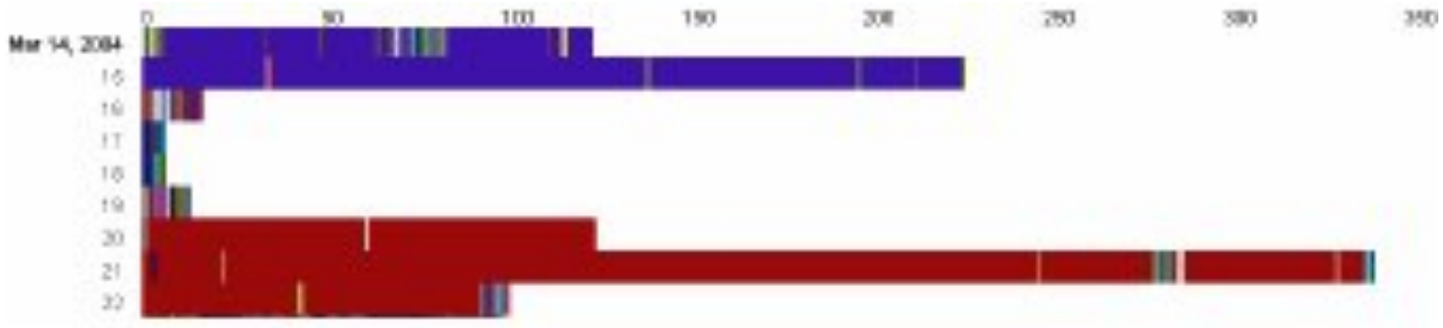


Fig. 5. Timeline comment chromogram with activity bursts. Blue: typo removal. Red: adding disambiguation messages.

As atividades acontecem quase sempre em **blocos**

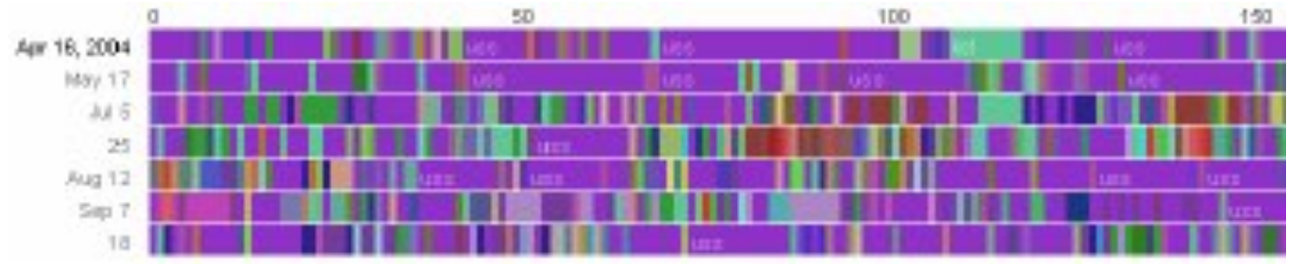


Fig. 6. Title chromogram: purple edits relate to U.S. ships, green to lists of ships.

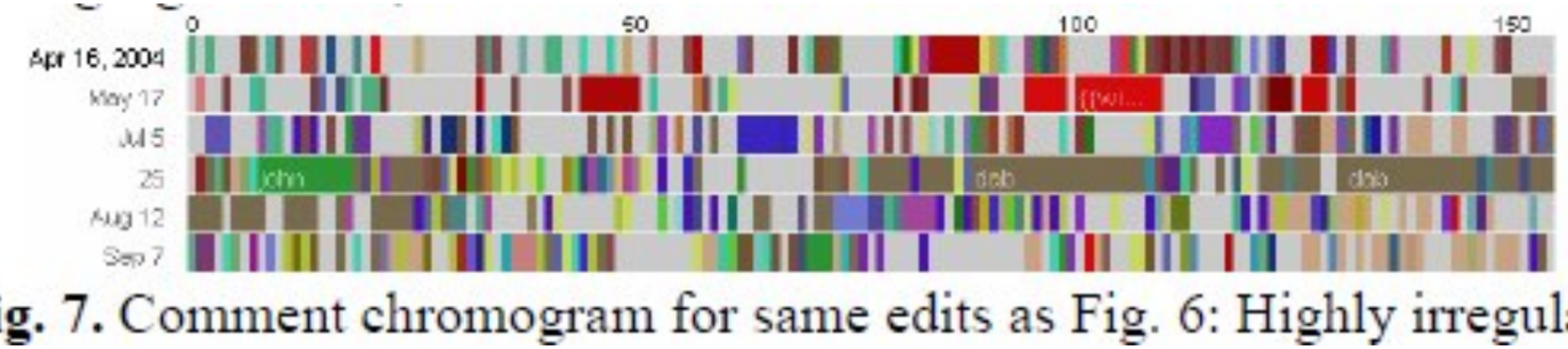


Fig. 7. Comment chromogram for same edits as Fig. 6: Highly irregular.

Em vários casos, os administradores editam diferentes artigos alternadamente (roxo e verde). Comumente os artigos relacionam-se ao mesmo tema

O tipo de atividade é altamente variável

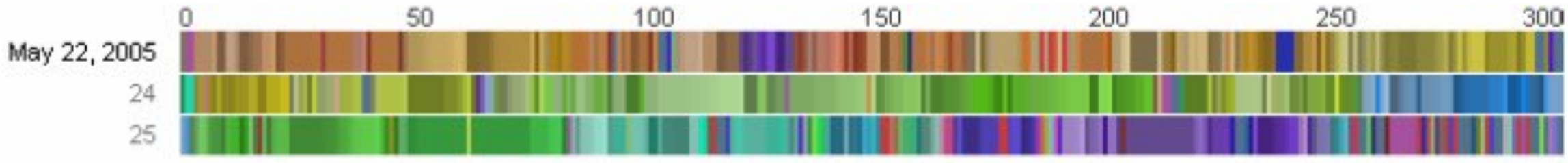


Fig. 8. Rainbows: Alphabetical order effects in a title chromogram for 900 edits

900 edições ao longo de 4 dias?

Alguma ideia do tipo de tarefa realizada?

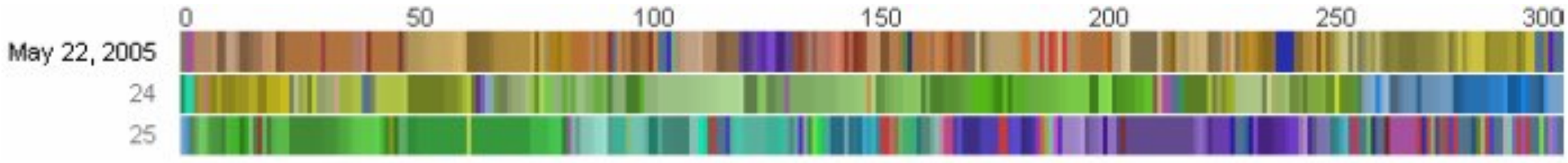


Fig. 8. Rainbows: Alphabetical order effects in a title chromogram for 900 edits

Edição de temas em ordem alfabética

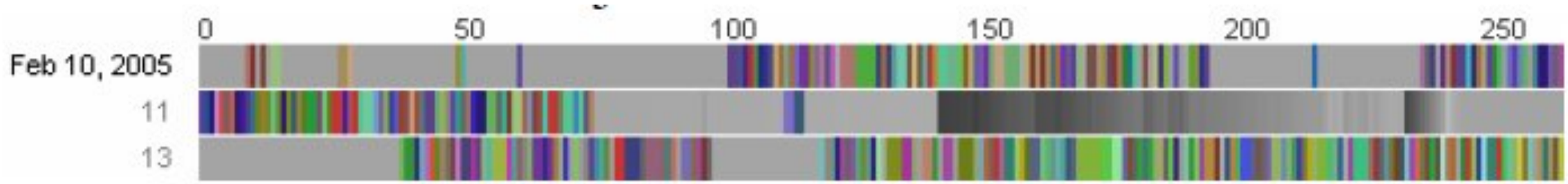


Fig. 9. Title chromogram: edits on pages with numerical titles (in gray)

Edição de temas cujos títulos se iniciam com números como por exemplo “1922 in Literature”

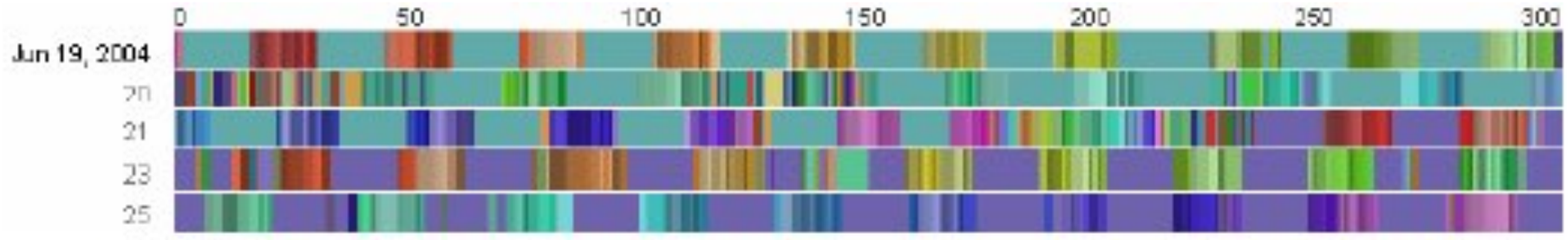


Fig. 11. Title chromogram: Switching between images (purple and light blue) and articles (rainbows)

Administradores sistematicamente fazem o upload de várias imagens e depois as inserem no texto

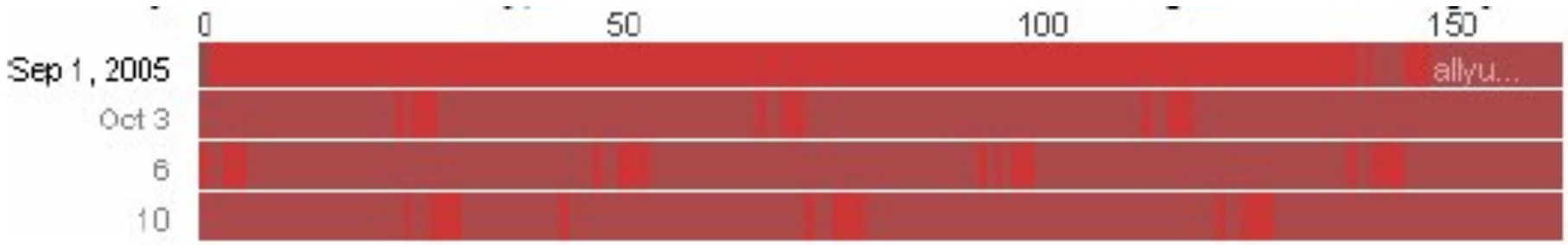


Fig. 12. Title chromogram for "AFD bot". Note the small range of colors and regular rhythm of edits.

Administradores comumente escrevem programas para realização de tarefas repetitivas

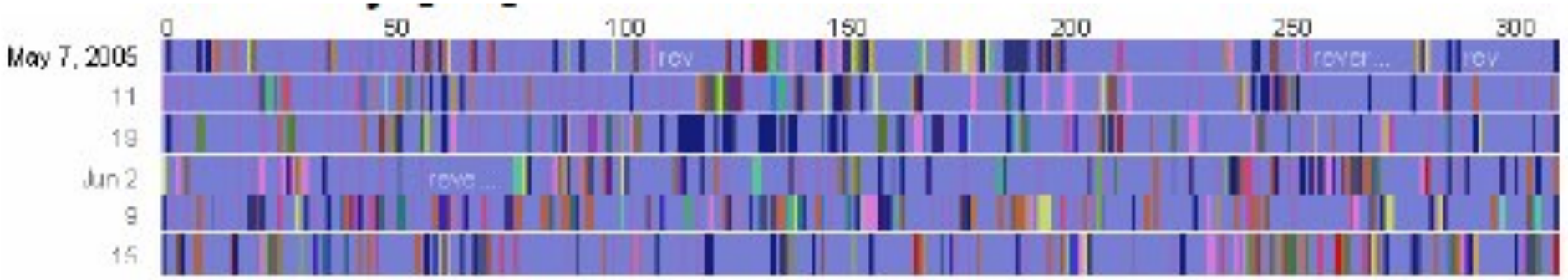


Fig. 13. Comment chromogram: Reverting vandalism and other bad edits

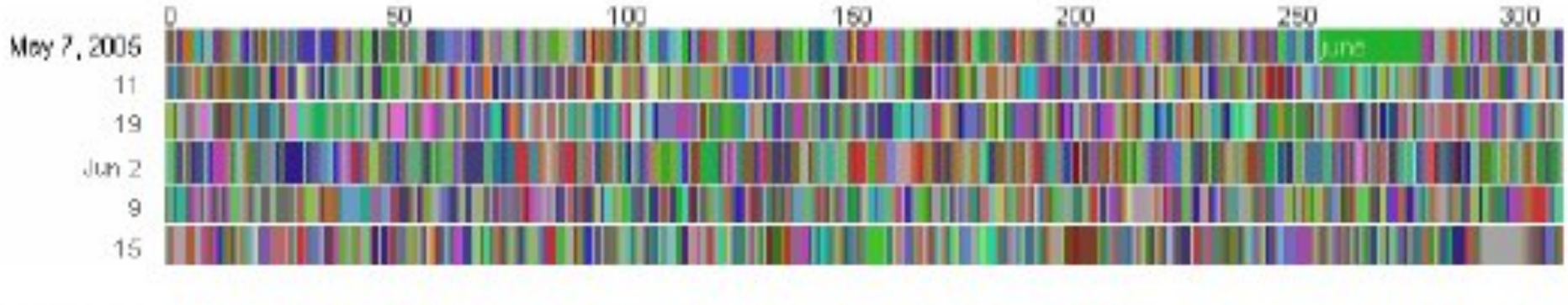
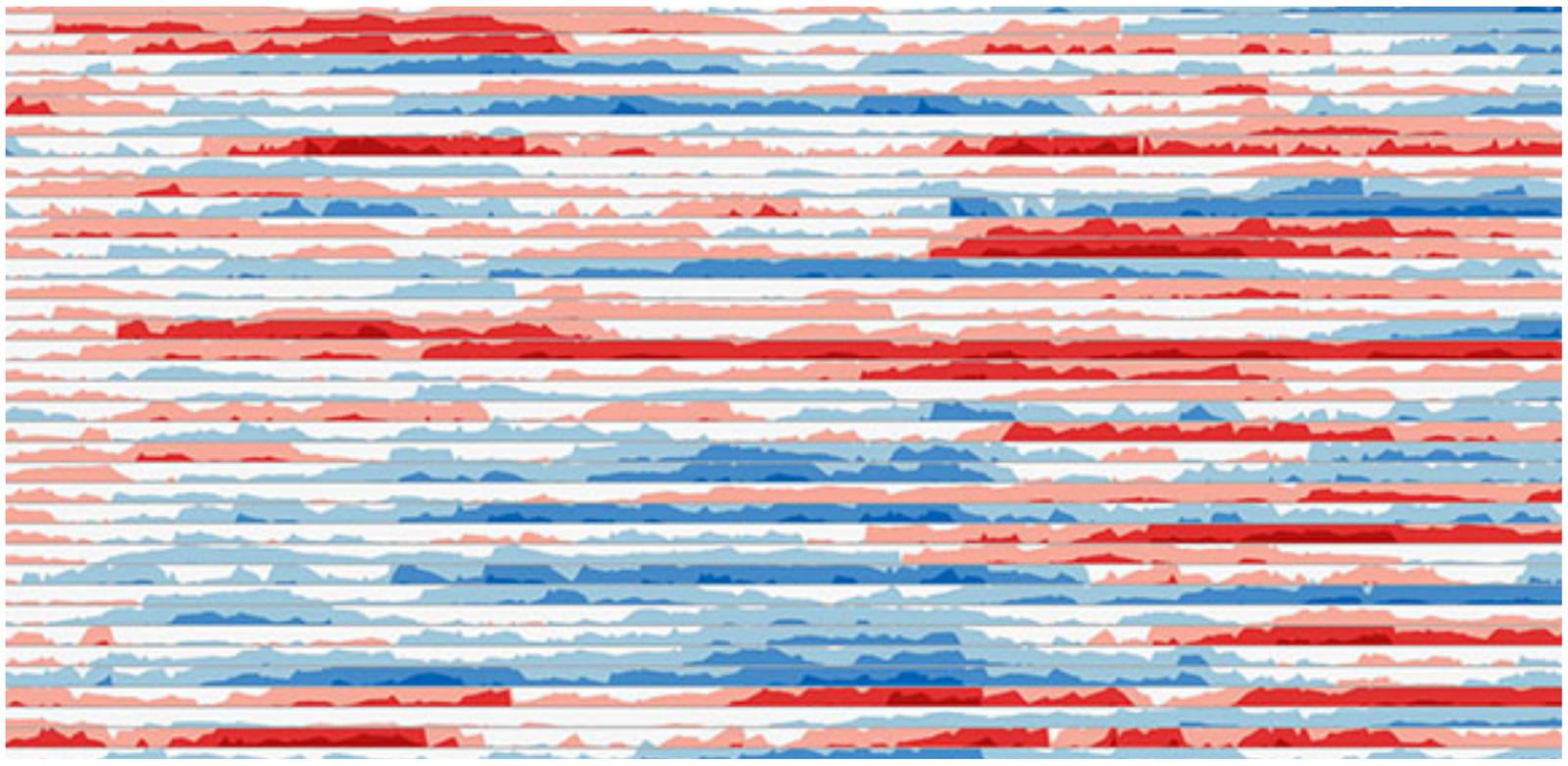


Fig. 14. Title chromogram for same edits in Fig. 13.

Reversão de vandalismo em artigos variados

SIZING THE HORIZON: THE EFFECTS OF CHART SIZE AND LAYERING ON THE GRAPHICAL PERCEPTION OF TIME SERIES VISUALIZATIONS

J. Heer and N. Kong e M. Agrawala
Conference on Human Factors in Computer Systems
2009



O objetivo é a visualização e comparação de múltiplas séries temporais

De acordo com o princípio de Tufte sobre maximizar a densidade de dados

Figure 1. (a) Filled line chart. Area between data values on line and zero is filled in. **(b) "Mirrored" chart.** Negative values are flipped and colored red, cutting the chart height by half. **(c) 2-band horizon graph.** The chart is divided into bands and overlaid, again halving the height.

Como o espelhamento e o uso de camadas para representação de séries temporais afeta a habilidade de análise?

Como os valores negativos espelhados são interpretados?

O desempilhamento mental das camadas do gráfico interferem com a estimativa dos valores?

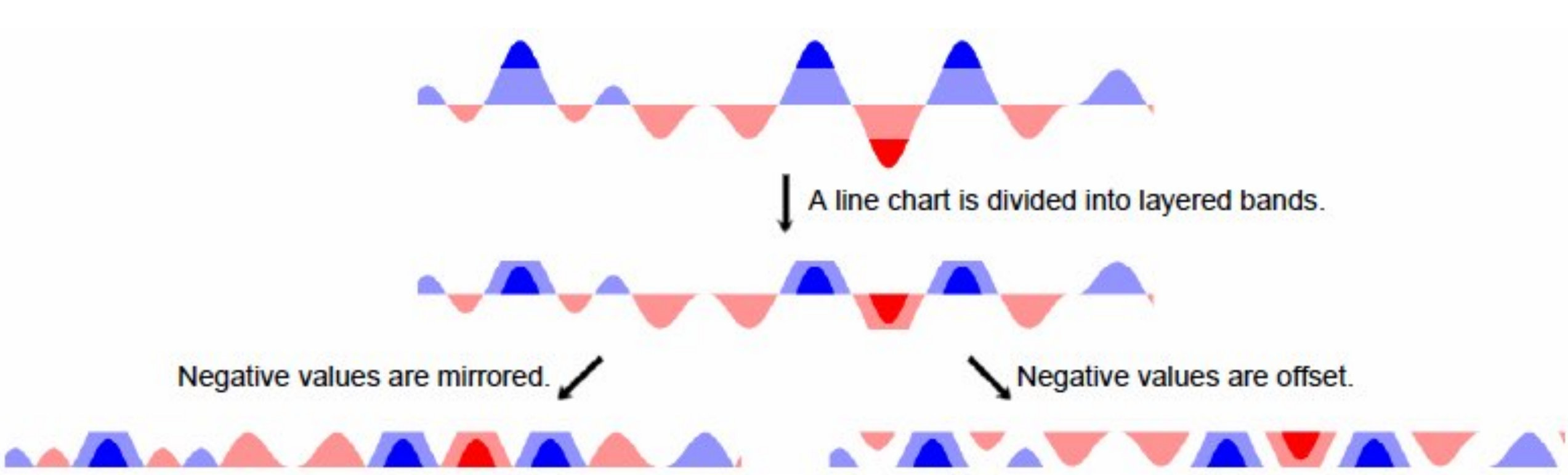


Figure 2. Horizon graph construction. A normal line chart is divided into bands defined by uniform value ranges. The bands are then layered to reduce the chart height. Negative values can be mirrored or offset into the same space as positive values.

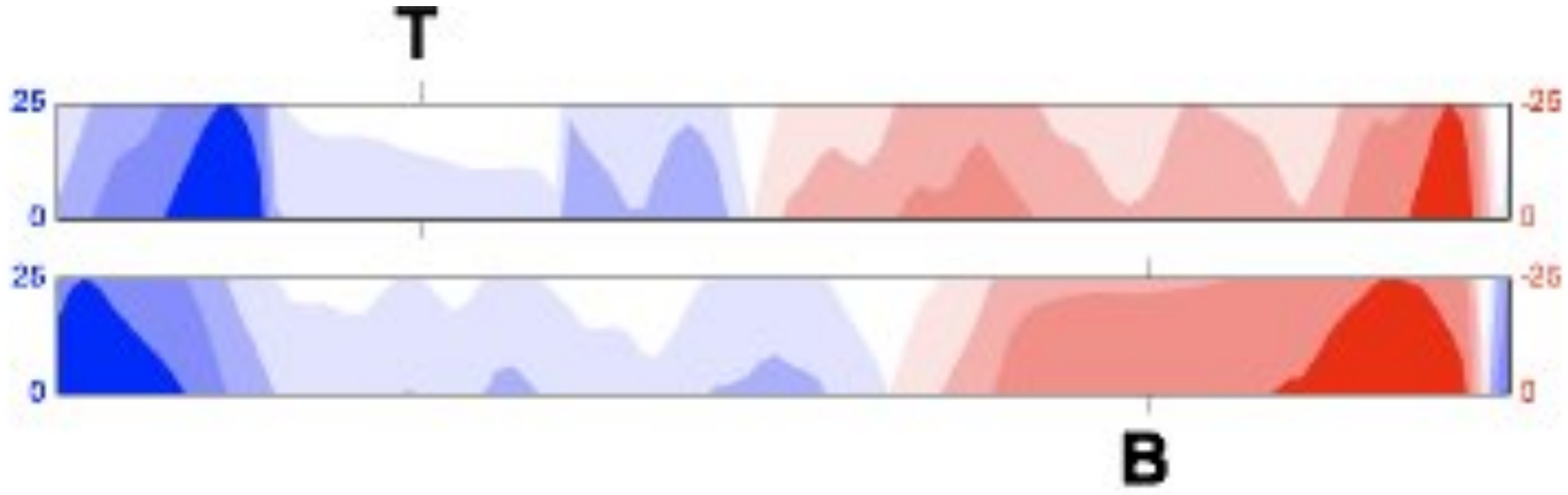


Figure 3. Example trial with a 4-band mirrored graph.
Each band covers 25 values; the total range is [-100, 100].
Subjects reported if T or B was larger, and by how much.

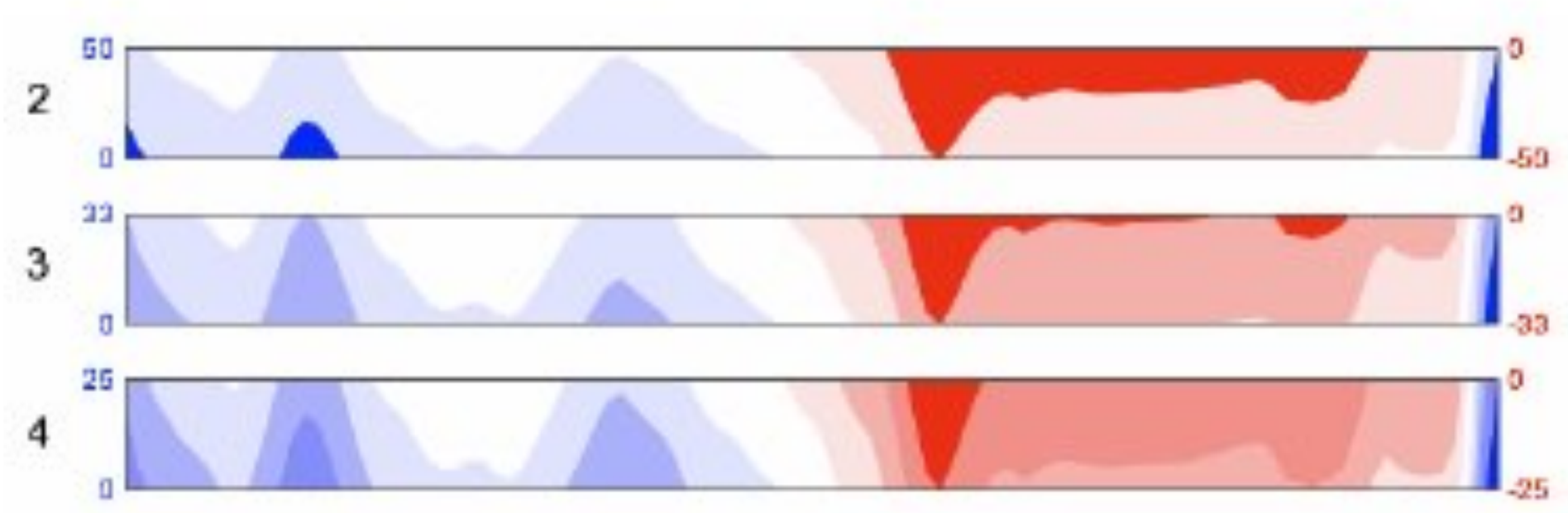


Figure 4. Offset horizon graphs with 2, 3, and 4 bands.

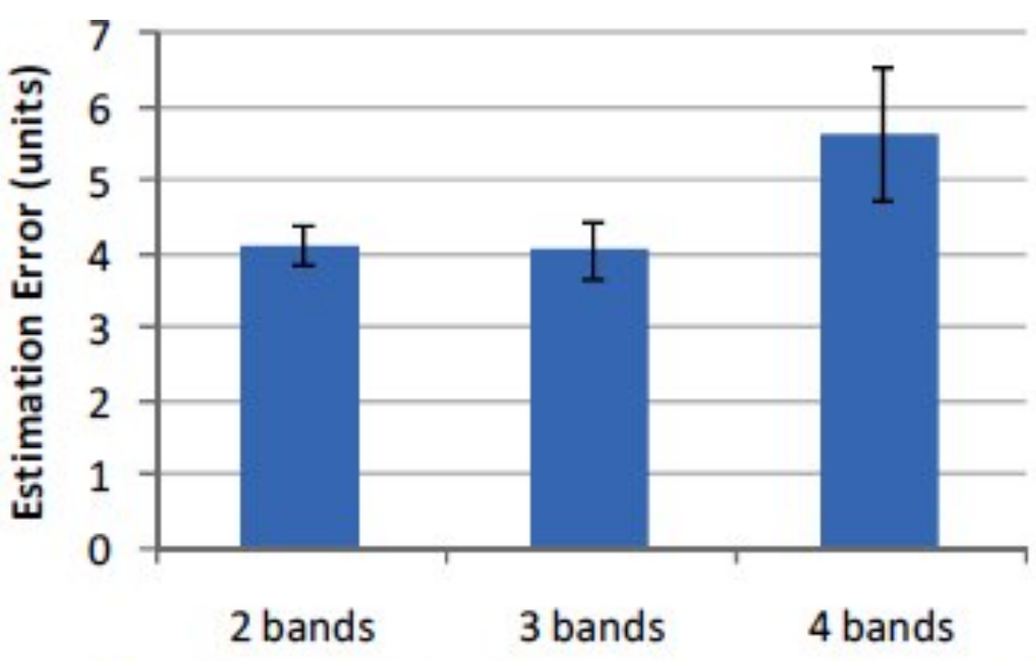


Figure 5. Estimation Error by Band Count. 4-band charts have significantly higher error than 2- or 3-band charts.

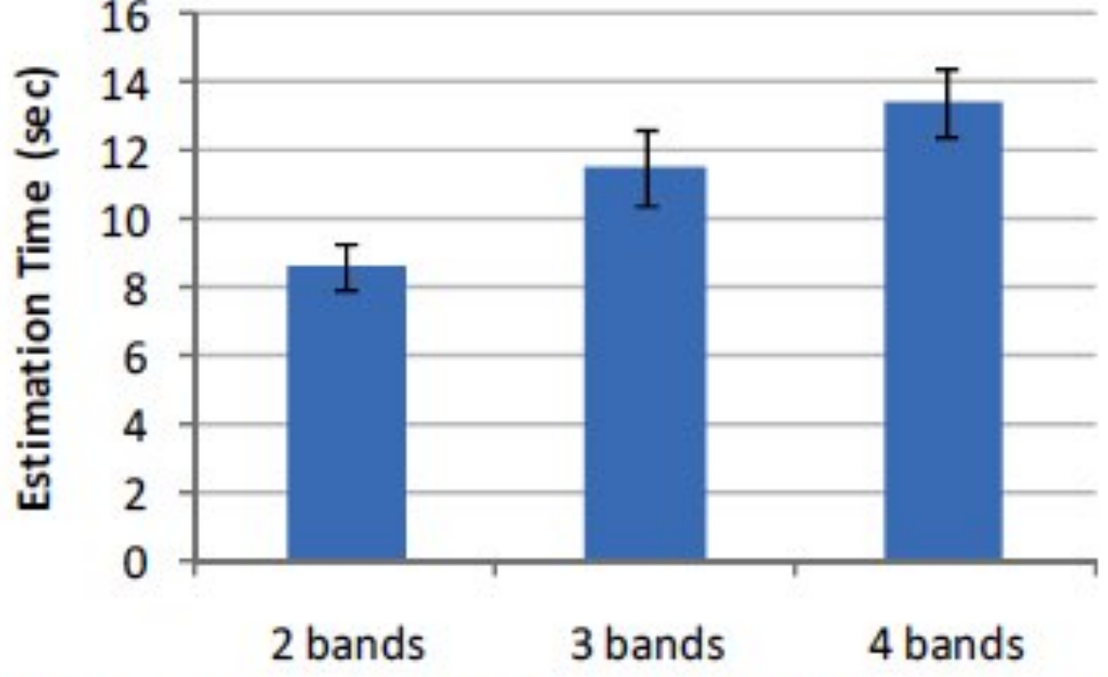


Figure 6. Estimation Time by Band Count. Estimation time increases significantly with each additional band.

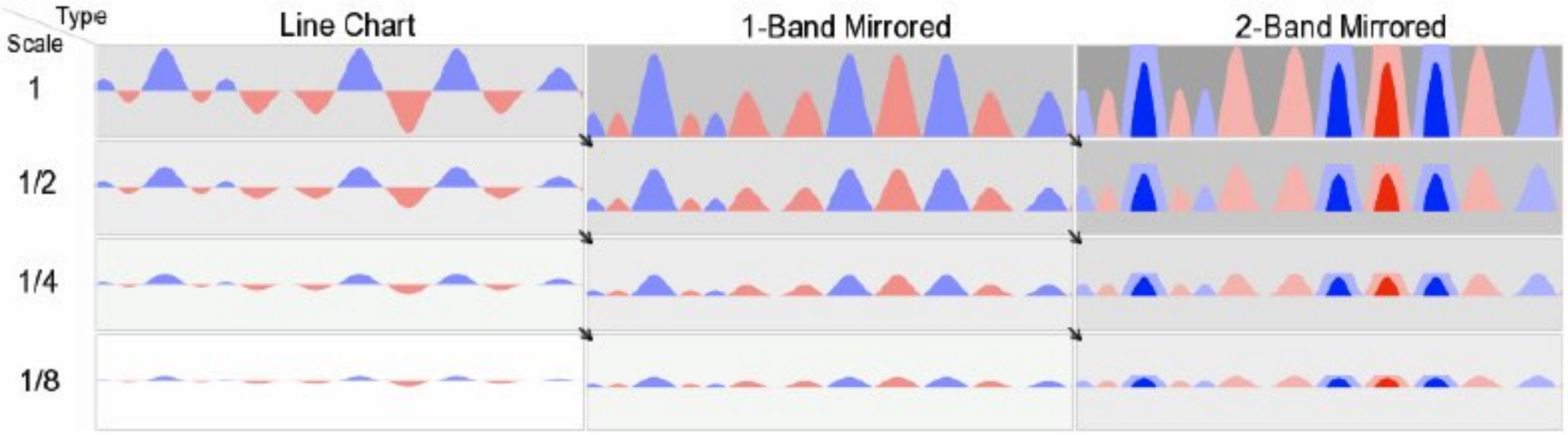


Figure 7. Chart Type and Scale Conditions in Experiment 2. We crossed 3 chart types and 4 chart heights. The diagonally adjacent cells indicated by arrows and shading have the same *virtual resolution*: the un-mirrored, un-layered size of the chart.

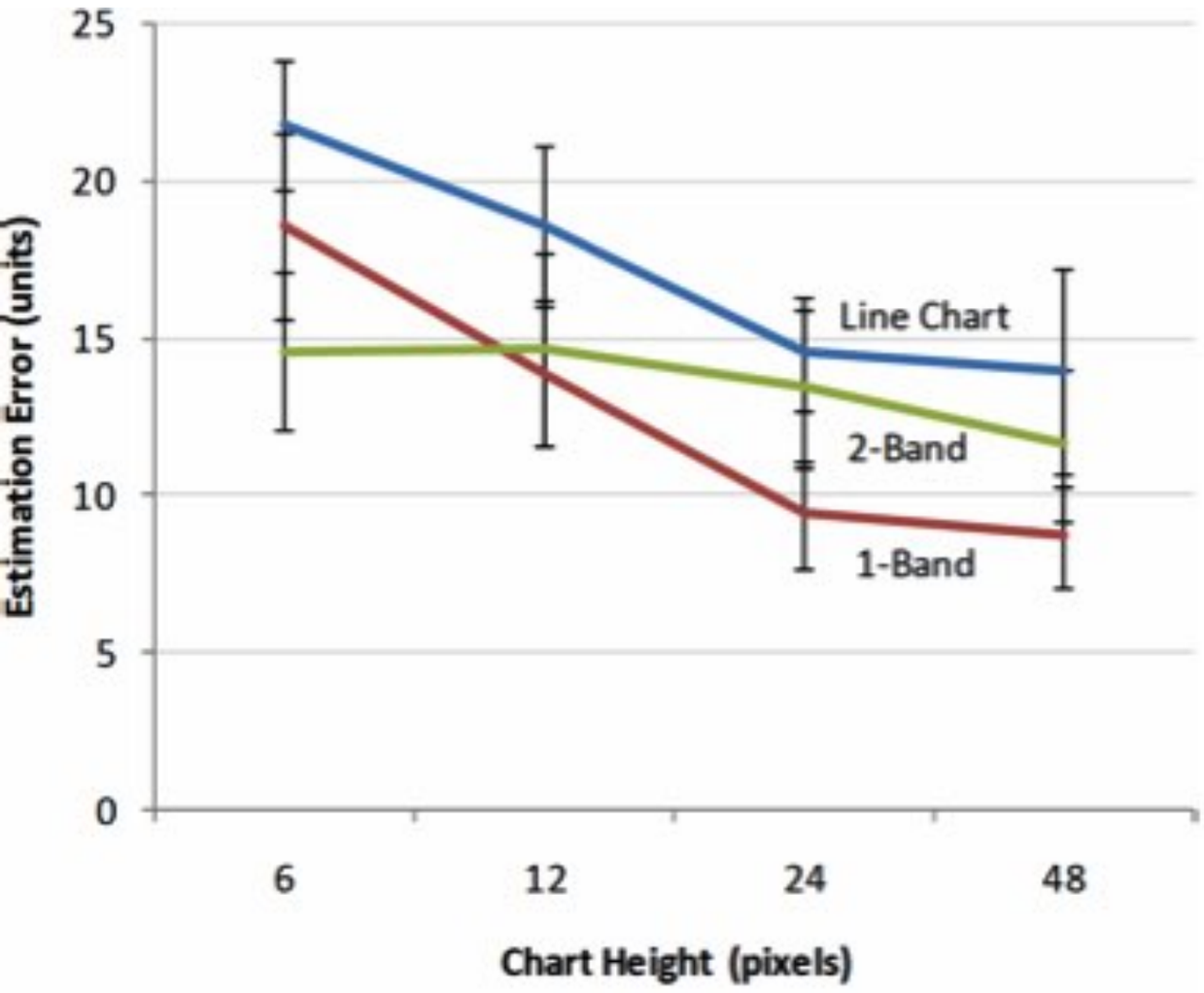


Figure 8. Estimation Error by Chart Type and Height. The 2-band mirror chart crosses the 1-band case at a chart height of 12 pixels (scale factor 1/4).

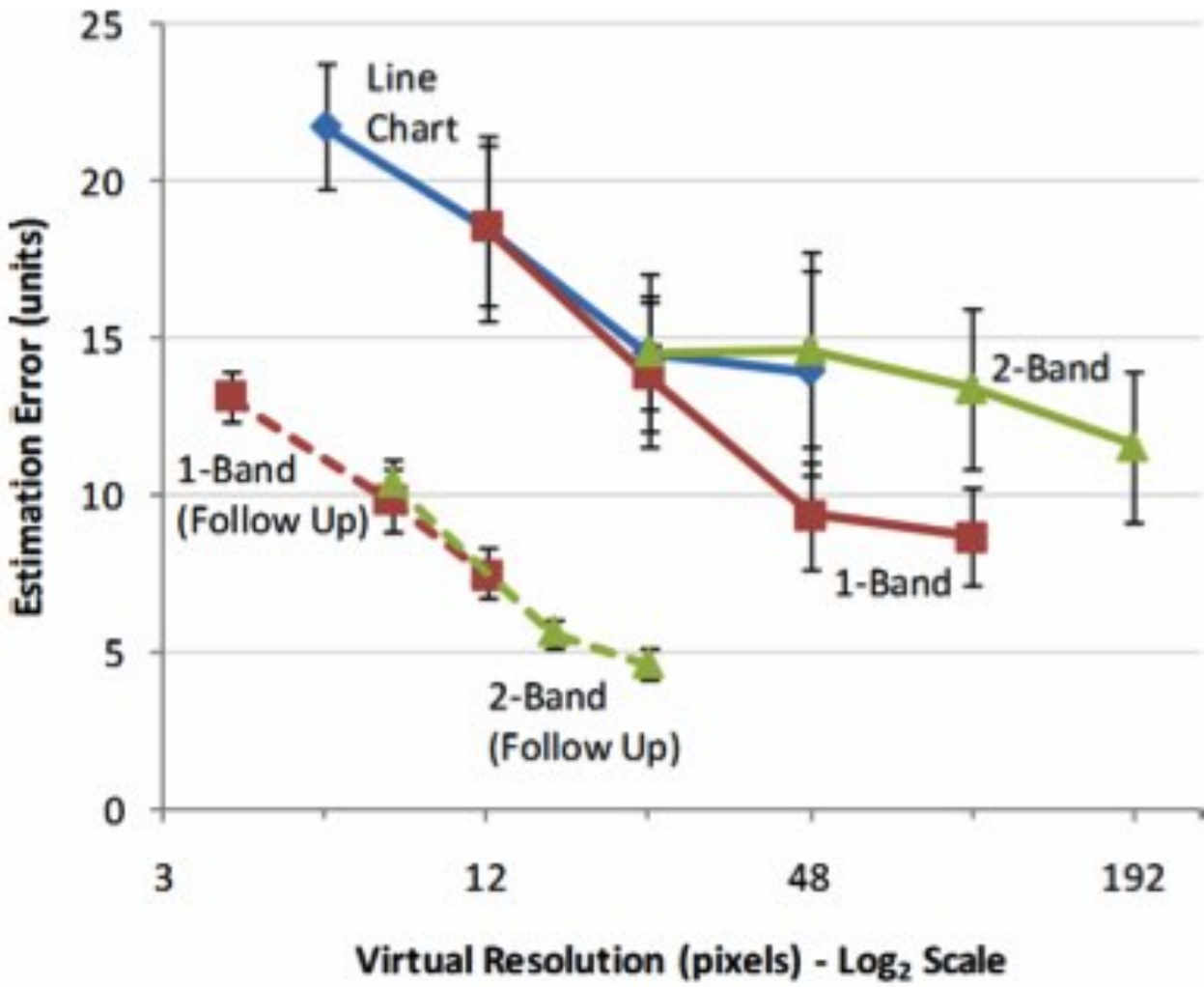


Figure 9. Estimation Error by Chart Type and Virtual Resolution. Error levels hold relatively stable at high virtual resolutions, but increase linearly at smaller resolutions.

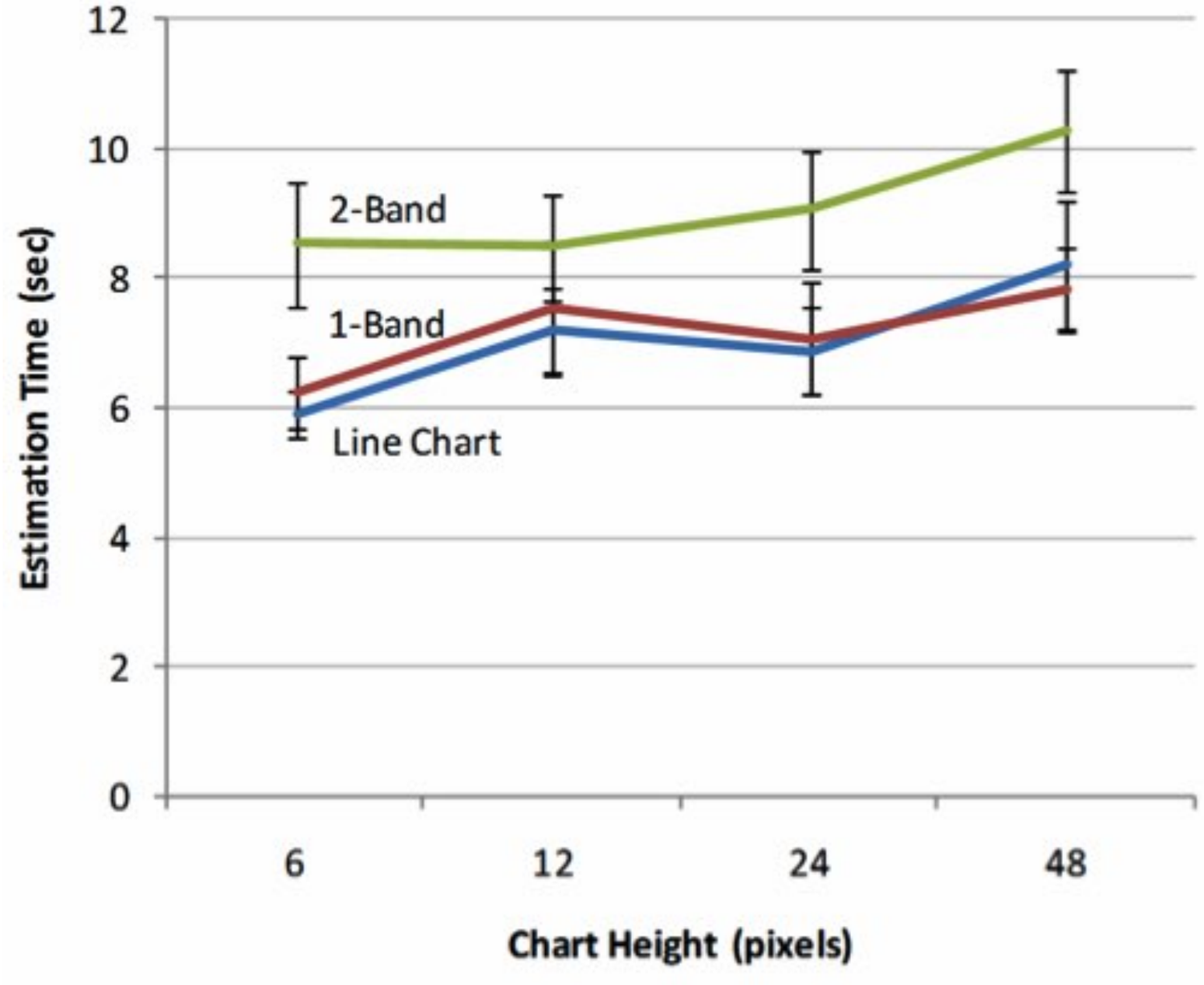


Figure 10. Estimation Time by Chart Type and Height.
Line and 1-band mirror charts result in similar estimation times. Both are significantly faster than 2-band charts.

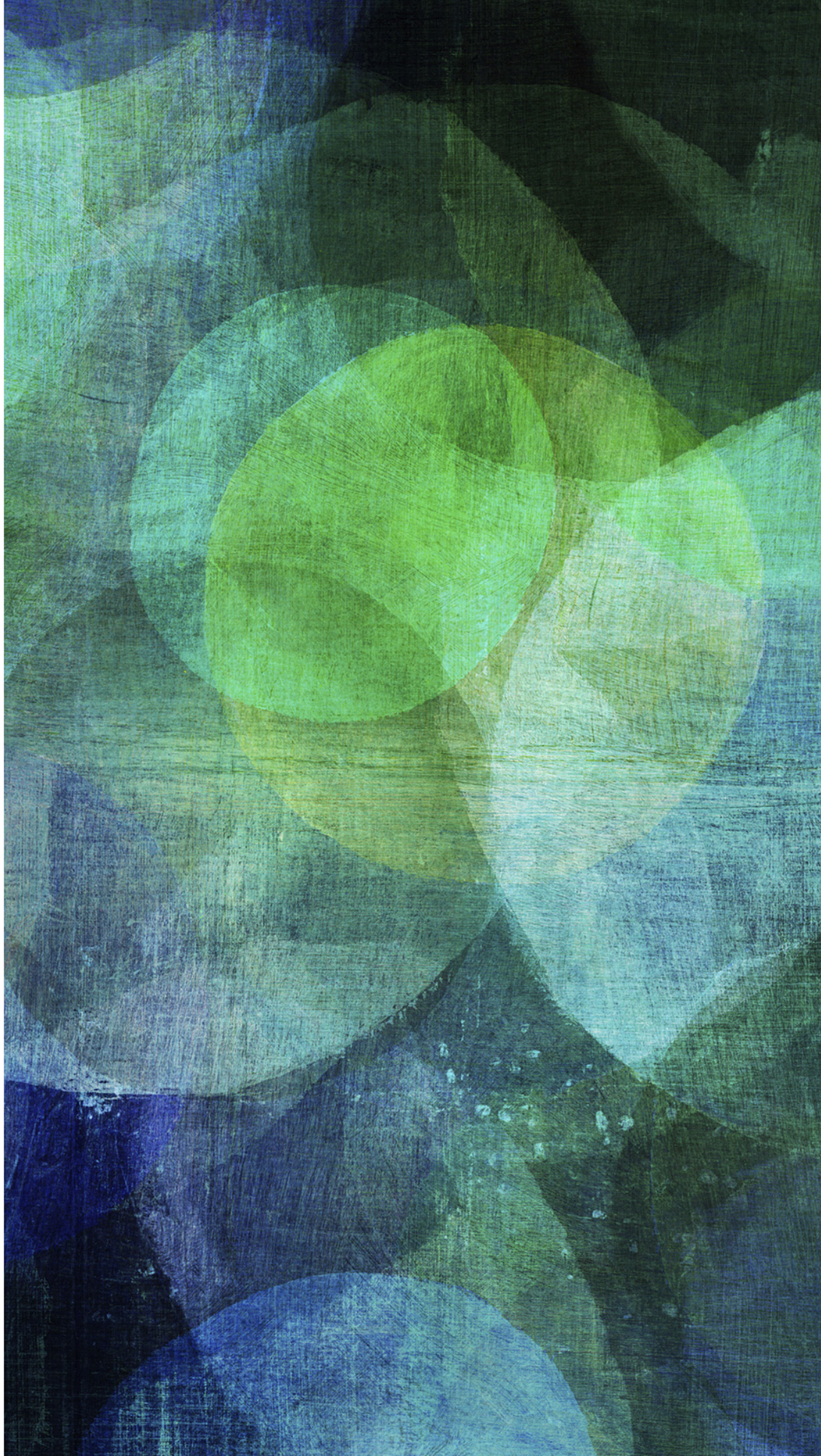
AVALIAÇÃO

- O espelhamento não tem efeito sobre o tempo e a acurácia

- As camadas são benéficas quando o tamanho do gráfico diminui (gráficos com altura menor que 6.8mm tiveram melhores resultados com 2 faixas)

- Não é recomendável usar 4 faixas ou mais: os usuários relataram cansaço para estimar os valores

VISUALIZAÇÃO DE SÉRIES TEMPORAIS E MINERAÇÃO DE DADOS



Principais tarefas de mineração de dados aplicadas a séries temporais

- **Agrupamento:** encontrar grupos de séries temporais similares
- **Classificação:** assinalar uma ou mais classes a uma nova série temporal
- **Detecção de anomalias**
- **Indexação:** dada uma série temporal e uma métrica de similaridade, encontrar a série temporal mais próxima

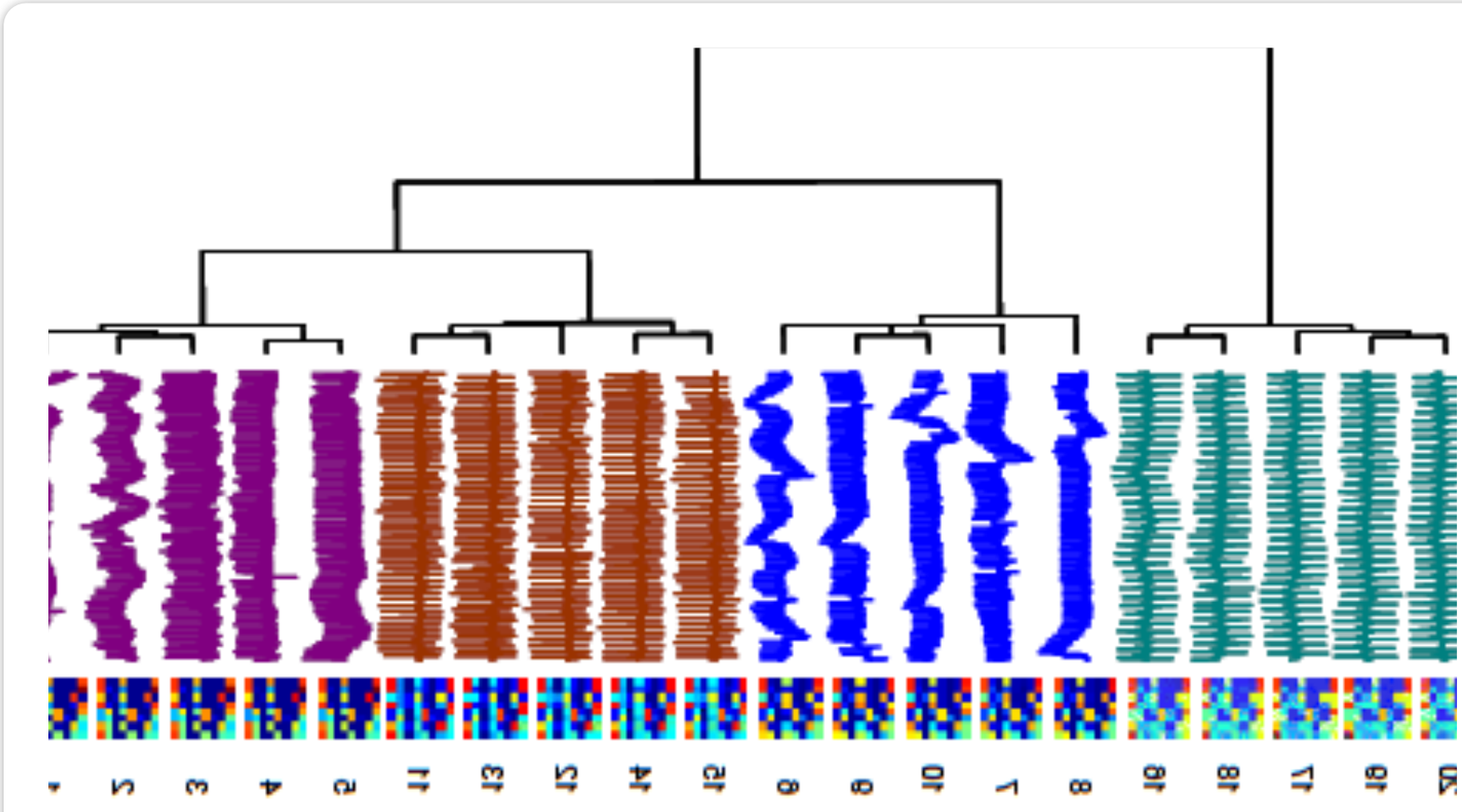
PRINCIPAIS TAREFAS

- **Agrupamento:** encontrar grupos de séries temporais similares
- **Classificação:** assinalar uma ou mais classes a uma nova série temporal
- **Detecção de anomalias**
- **Indexação:** dada uma série temporal e uma métrica de similaridade, encontrar a série temporal mais próxima

TIME-SERIES BITMAPS: A PRACTICAL VISUALIZATION TOOL FOR WORKING WITH LARGE TIME SERIES DATABASES

N. Kuman, N. Lolla, E. Keogh, S. Lonardi e C. Hatanamahatana

SIAM
2005



OBJETIVOS

- Mineração de dados de séries temporais facilitando a navegação em grandes volumes de séries temporais
- Extração de características de séries temporais e representação visual em bitmaps
- Agrupamento e detecção de regularidades e / ou anomalias em várias séries temporais

COMO COMPARAR SÉRIES TEMPORAIS

- Distância euclidiana no espaço n-dimensional

- Dynamic time warping (DTW)

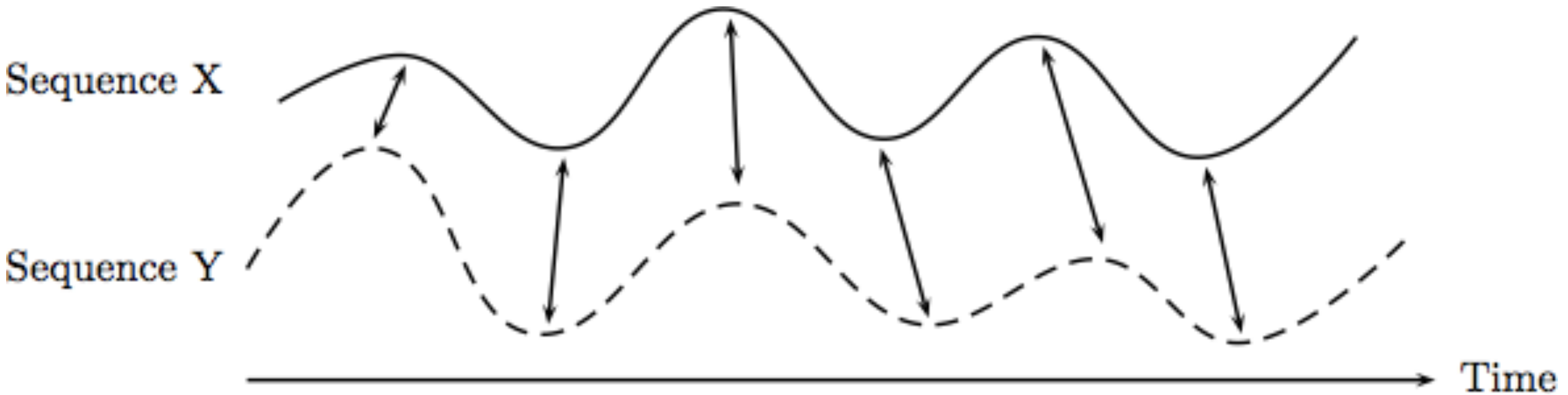
- Symbolic Aggregate ApproXimation (SAX)

DYNAMIC TIME WARPING (DTW)

Dynamic time warping (DTW) é uma técnica comumente usada para alinhamento entre séries temporais

Um par de sequências é distorcido de forma não linear para ser casado

DTW foi usado inicialmente para comparar padrões de fala em sistemas de reconhecimento automático de fala



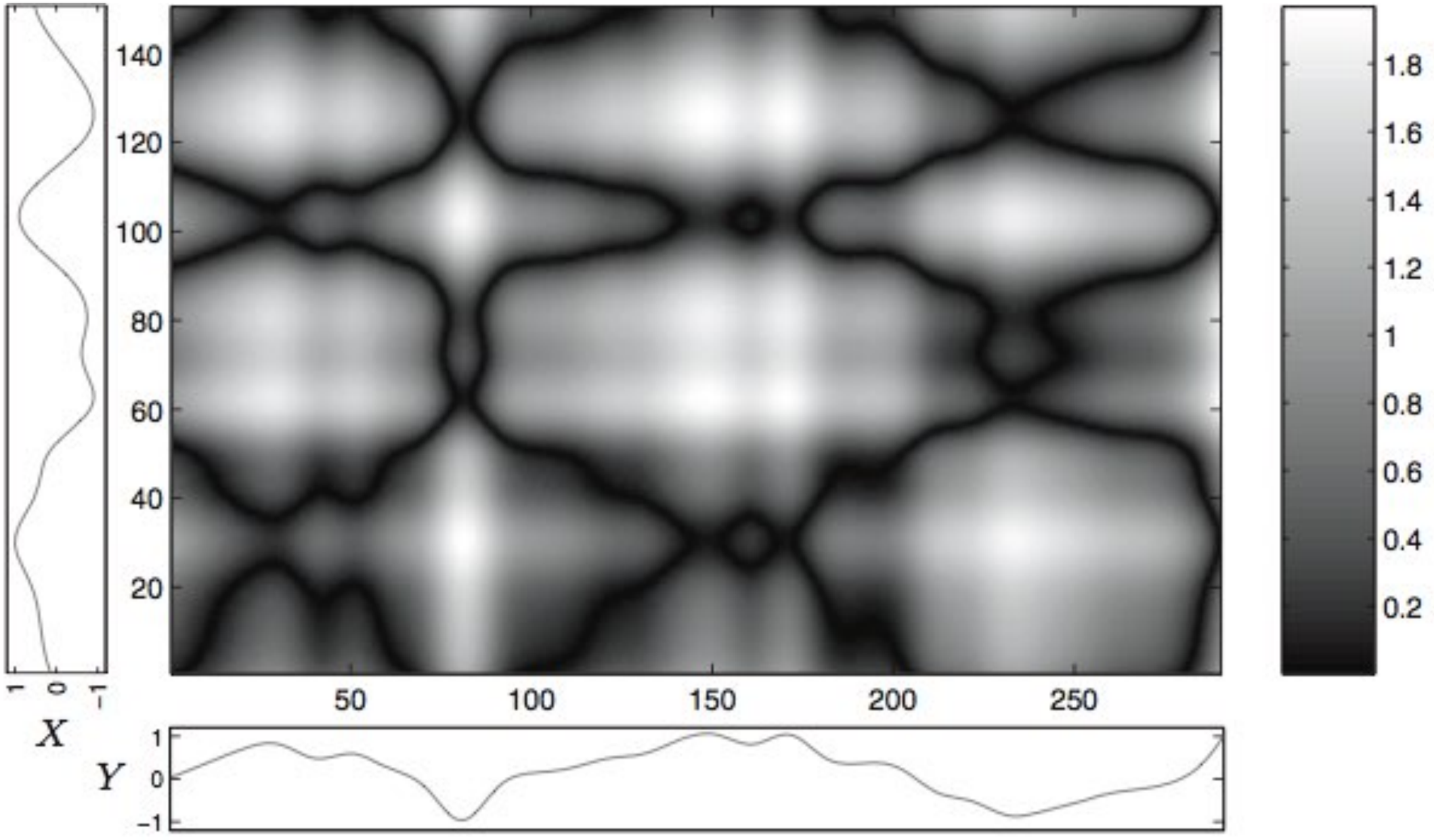


Fig. 4.2. Cost matrix of the two real-valued sequences X (*vertical axis*) and Y (*horizontal axis*) using the Manhattan distance (absolute value of the difference) as local cost measure c. Regions of low cost are indicated by *dark colors* and regions of high cost are indicated by *light colors*

Algorithm: OPTIMALWARPINGPATH

Input: Accumulated cost matrix D.
Output: Optimal warping path p^*.

Procedure: The optimal path $p^* = (p_1, \ldots, p_L)$ is computed in reverse order of the indices starting with $p_L = (N, M)$. Suppose $p_\ell = (n, m)$ has been computed. In case $(n, m) = (1, 1)$, one must have $\ell = 1$ and we are finished. Otherwise,

$$p_{\ell-1} := \begin{cases} (1, m-1), & \text{if } n = 1 \\ (n-1, 1), & \text{if } m = 1 \\ \text{argmin}\{D(n-1, m-1), & \\ \qquad D(n-1, m), D(n, m-1)\}, & \text{otherwise,} \end{cases} \tag{4.6}$$

where we take the lexicographically smallest pair in case "argmin" is not unique.

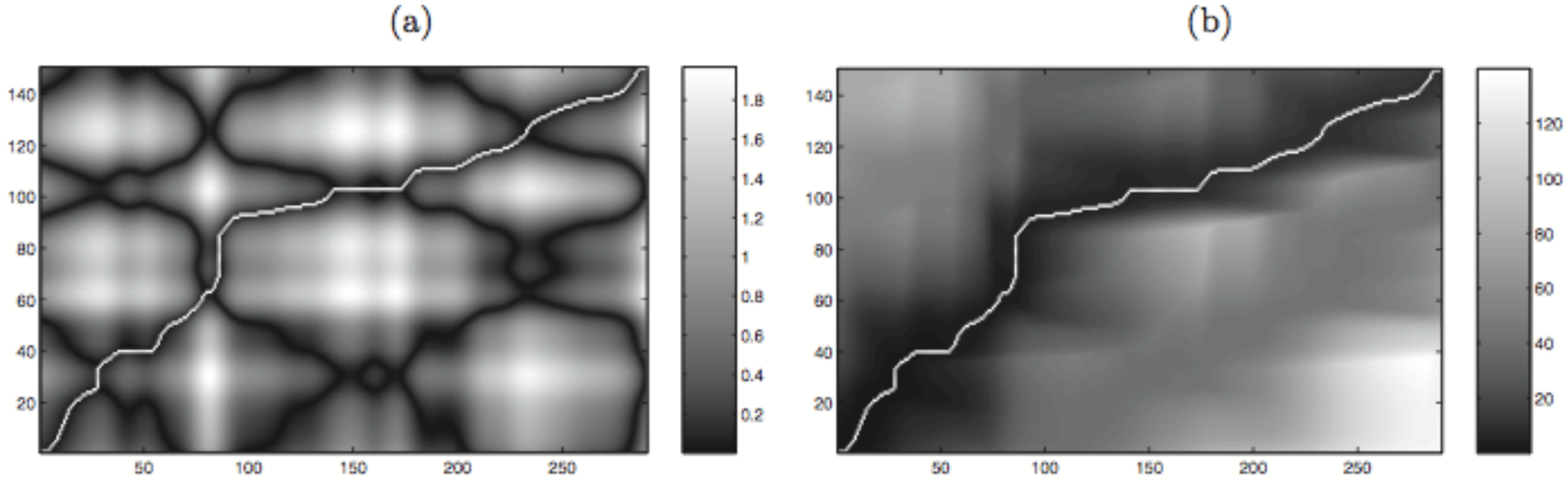


Fig. 4.4. **(a)** Cost matrix C as in Fig. 4.2 and **(b)** accumulated cost matrix D with optimal warping path p^* (*white line*)

SYMBOLIC AGGREGATE APROXIMATION (SAX)

Consiste em agregar diversos símbolos em uma matriz $2^q \times 2^q$, na qual cada posição apresenta uma correspondência *1:1* para a frequência de uma subsequência de tamanho *q*

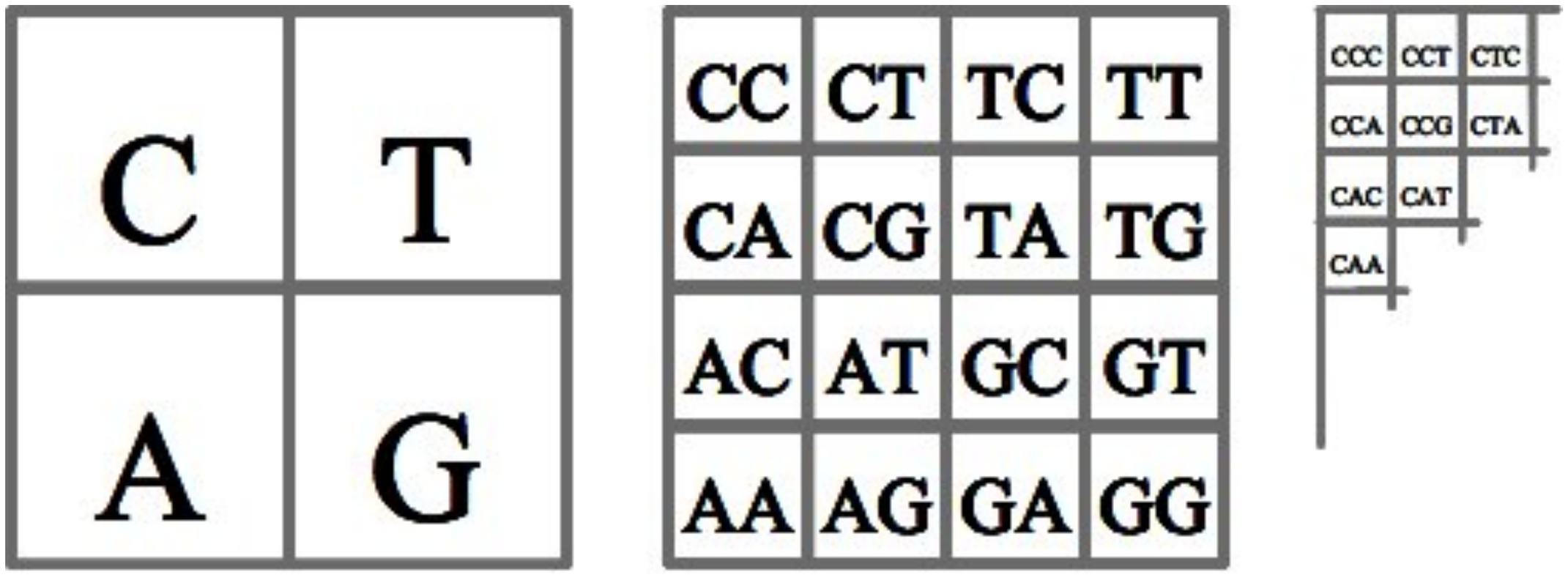


Figure 5: The quad-tree representation of a sequence over the alphabet {A,C,G,T} at different levels of resolution

Pan troglodytes

Elephas maximus

Homo sapiens

Loxodonta africana

The gene sequences of mitochondrial DNA of four animals, used to create their own icons using a chaos game representation. Note that *Pan troglodytes* is the familiar Chimpanzee, and *Loxodonta africana* and *Elephas maximus* are the African and Indian Elephants respectively. The file icons show that humans and chimpanzees have similar genomes, as do the African and Indian elephants.

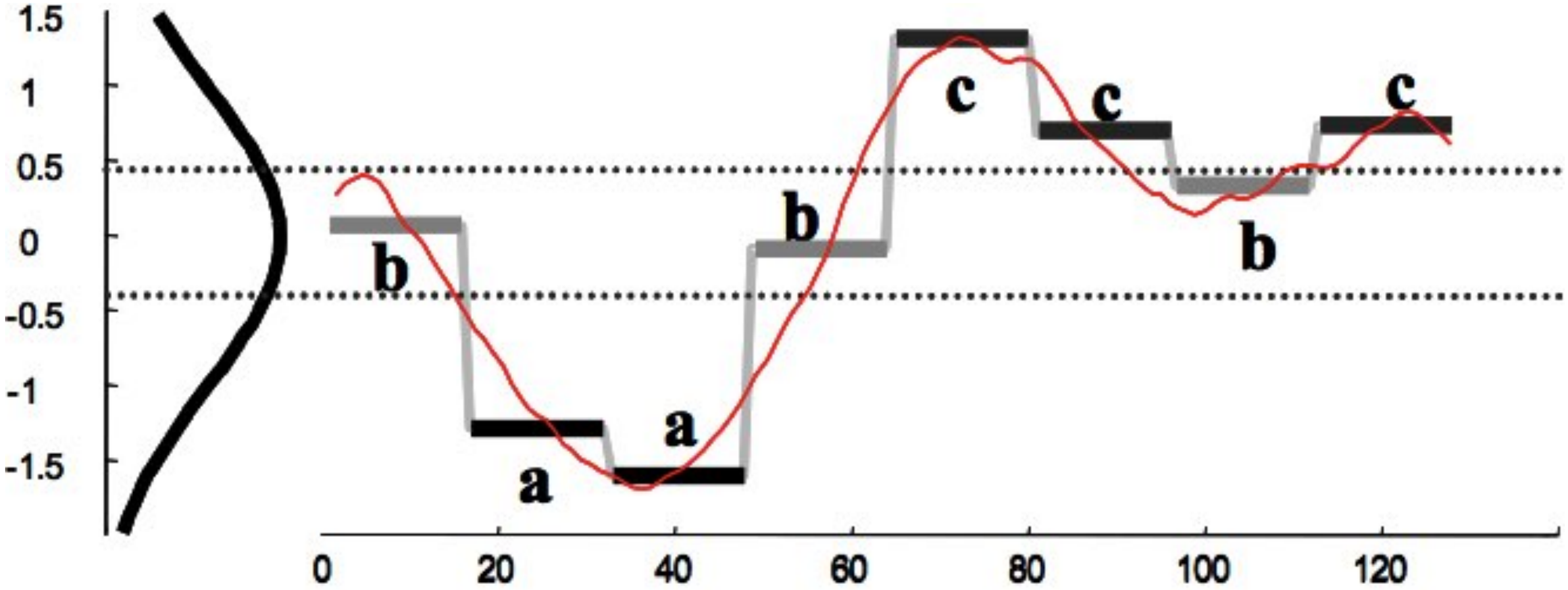


Figure 7: A real valued time series can be converted to the SAX word **baabccbc**. Note that all three possible symbols are approximately equally frequent.

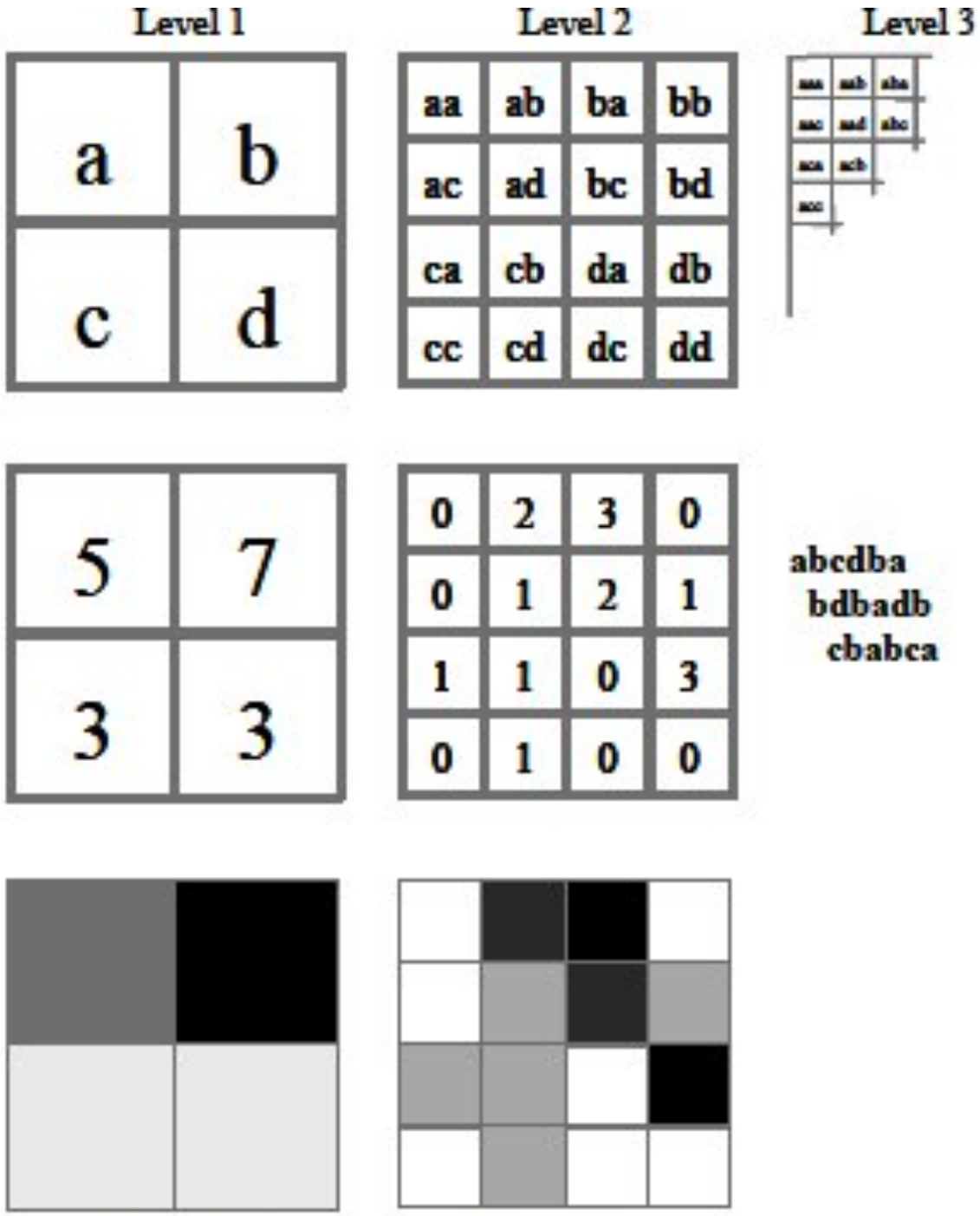
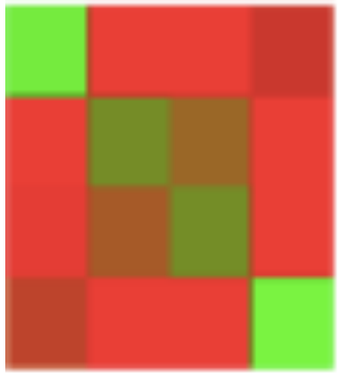
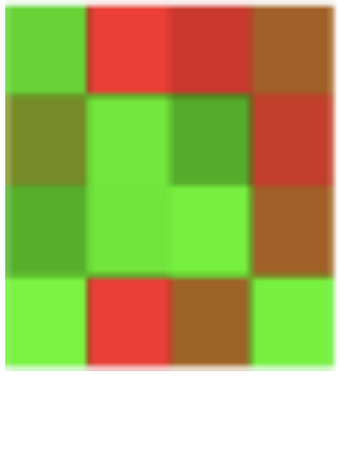


Figure 8: *Top*) The four possible SAX symbols are mapped to four quadrants of a square, and pairs, triplets etc are recursively mapped to finer grids. *Middle*) We can extract counts of symbols from a SAX representation and record them in the grids. *Bottom*) The recorded values can be linearly mapped to colors, thus creating a square bitmap.



example_a.dat

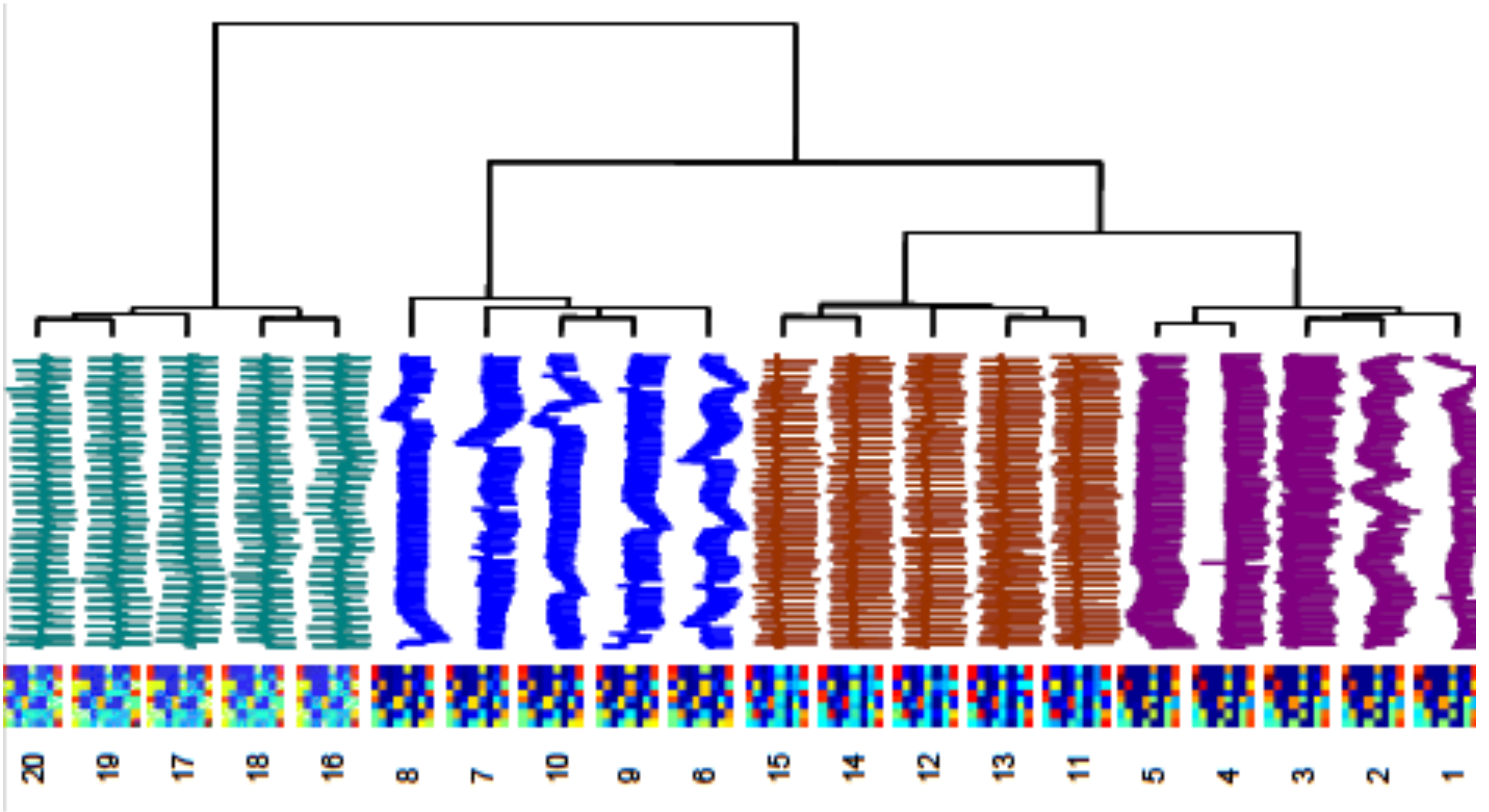
example_b.dat



example_c.dat

example_d.dat

Four time series files represented as time series bitmaps. While they are all examples of EEGs, example_a.dat is from a normal trace, whereas the others contain examples of spike-wave discharges. The fact that there is some difference between one dataset and all the rest is immediately apparent from a casual inspection of the bitmap representation.


20
19
17
18
16
8
7
10
9
6
15
14
12
13
11
5
4
3
2
1