Considerations for Visualizing Comparison

Michael Gleicher

Department of Computer Sciences

University of Wisconsin - Madison

What is this paper?

Considerations for Visualizing Comparison

What is this paper?

Considerations for Visualizing Comparison

4 questions to ask

What is this paper?

Considerations for Visualizing Comparison

4 questions to ask

when designing a visualization or tool

What is this paper?

Considerations for Visualizing Comparison

4 questions to ask

when designing a visualization or tool

?

Dictionary.com definitions compare

compare

[kuh m-**pair**]

Spell Syllables

CITE

Examples Word Origin

See more synonyms on Thesaurus.com

verb (used with object), **compared, comparing.**

1. to examine (two or more objects, ideas, people, etc.) in order to note similarities and differences:
   to compare two pieces of cloth; to compare the governments of two nations.
2. to consider or describe as similar; liken: *"Shall I compare thee to a summer's day?"*.
3. *Grammar.* to form or display the degrees of comparison of (an adjective or adverb).

To examine (two or more objects, ideas, people, etc.) in order to note similarities and differences

OED Oxford English Dictionary
The definitive record of the English language

Quick search:

« Previous | Back to Results | Next » Help on Dictionary Entry | Print | Save | Email | Cite

compare, *v.*[1]

Text size: A A

View as: Outline | Full entry Quotations: Show all | Hide all Keywords: On | Off

Pronunciation: Brit. /kəmˈpɛː/, U.S. /kəmˈpɛ(ə)r/
Forms: Also ME *Sc.* **comper.**
Frequency (in current use): •••••••
Etymology: < Old French *comperer* (from 14th cent. *comparer*) = Provençal *comparar*, Spanish ... (Show More)

1.

a. *trans.* To speak of or represent as similar; to liken. Const. *to*. (With negative, in such phrases as *not to be compared to*, usually implying great inferiority in some respect.) Thesaurus »

1447 O. Bokenham *Lyvys Seyntys* (1835) 9 Seynt Margrete On to that gemme [may] weel comparyd be.
1489 (▸ a1380) J. Barbour *Bruce* (Adv.) I. 403 Off manheid and mekill mycht Till Ector dar I nane comper.
a1538 T. Starkey *Dial. Pole & Lupset* (1989) 31 The one may..be comparyd to the body & the other to the soule.
1611 *Bible* (King James) Prov. iii. 15 All the things thou canst desire, are not to be compared vnto her.
1699 W. Dampier *Voy. & Descr.* I. vii. 125 He compares it to a Sloe, in shape and taste.
1855 W. H. Prescott *Hist. Reign Philip II of Spain* I. I. iv. 113 He greatly offended the Flemings by comparing their ships to muscle-shells.
(Hide quotations)

†**b.** ***to compare***: (a thing) for one to compare, (a thing) to be compared, comparable (*to*, *with*). Thesaurus »

1484 Caxton tr. G. de la Tour-Landry *Bk. Knight of Tower* (1971) lv. 80 Suche men or wymmen be to compare to the wyf of Lothe.
1711 J. Addison *Spectator* No. 161. ¶9 An Imitation of the best Authors, is not to compare with a good Original.
(Hide quotations)

c. *intr.* To draw a com

1597 Shakespeare *Ri*

2.

a. *trans.* To mark or
or place together (act
Const. *with* (or *to*) ar

1509 A. Barclay *Bra*
?1531 J. Frith *Disput*
a1640 R. Burton *An*
1667 Milton *Paradis*
1710 R. Steele *Tatler*
1850 R. W. Emerson *Mo*
in any other [country].
1860 J. Tyndall *Glaciers of Alps* II. x. 283 To compare the motion of the eastern and western halves of the glacier.
1879 G. C. Harlan *Eyesight* viii. 106 This cramping tendency of town as compared to country.

This entry has not yet been fully updated (first published 1891).

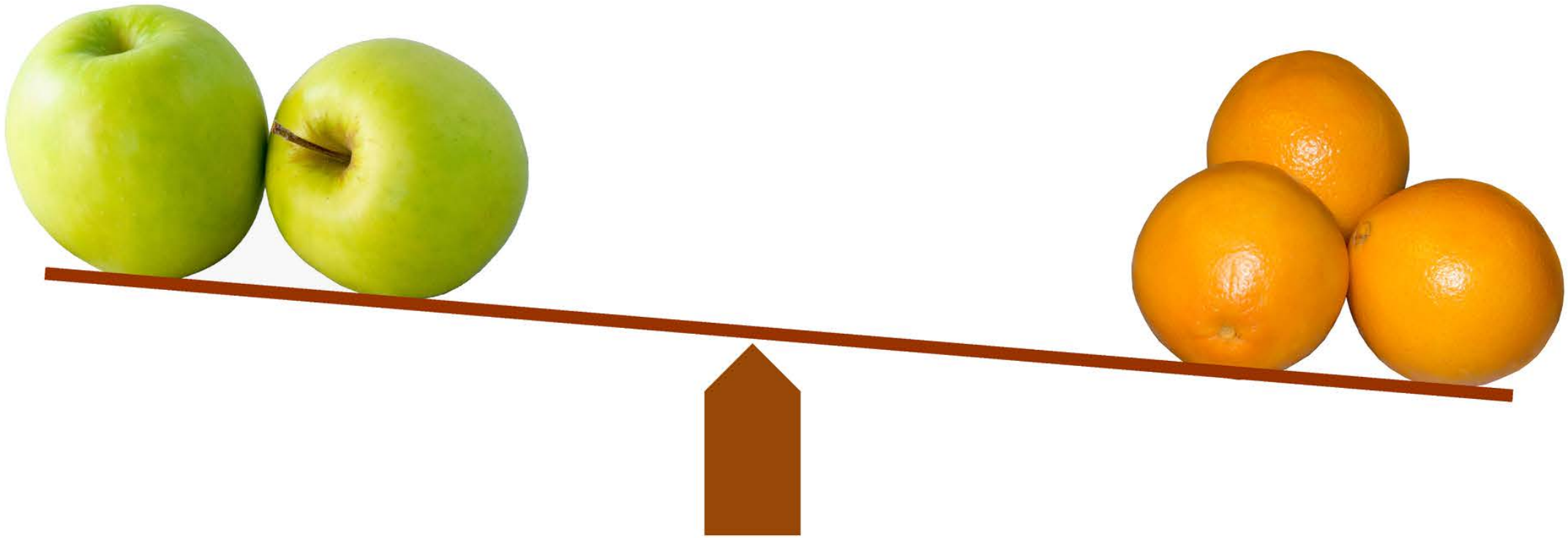
Entry history
Entry profile

Previous version: OED2 (1989)

In this entry:
compare notes, to
compare, to

In other dictionaries:
Oxford Dictionaries
compare: view definition in Oxford Dictionaries
comparen, v. in Middle English Dictionary

To mark or point out the similarities and differences of (two or more things)



How do I think about comparison?

to help me develop tools to help people do it

What is the comparison?

Why is it hard?

How to address the challenges?

Which visual design to use?

What is the comparison?

Comparative Elements
Targets
Actions

Why is it hard?

Comparative Challenges
Number of Targets
Large or Complex Targets
Complex Relationships

How to address the challenges?

Scalability Strategies
Scan Sequentially
Select Subset
Summarize Somehow

Which visual design to use?

Comparative Designs
Juxtapose
Superpose
Explicit Encoding

Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization

Danielle Albers, *Student Member, IEEE*, Colin Dewey, and Michael Gleicher, *Member, IEEE*

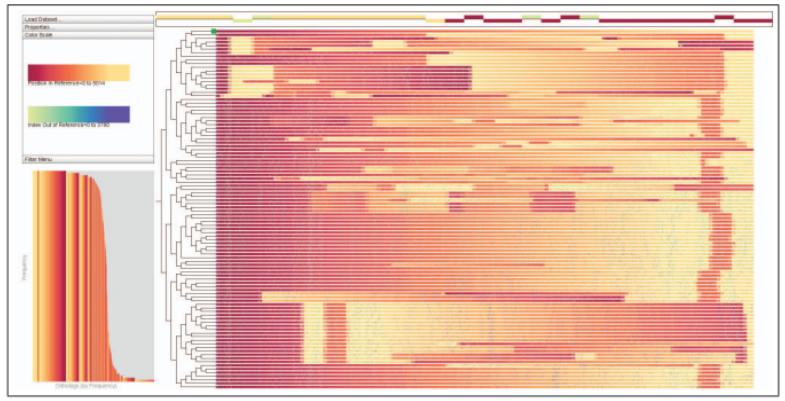


Fig. 1. Sequence Surveyor visualizing 100 synthetic genomes generated by an evolution simulation. Each genome is mapped to a row and genes are ordered by position. Color encodes the position of the gene within the chosen reference sequence (top row, indicated by the green box). Genes are aggregated, with each block's texture reflecting the overall distribution of colors in that block. The dendogram shows the phylogeny of the data set while the histogram shows the frequency distribution of orthology group sizes.

Abstract—In this paper, we introduce overview visualization tools for large-scale multiple genome alignment data. Genome alignment visualization and, more generally, sequence alignment visualization are an important tool for understanding genomic sequence data. As sequencing techniques improve and more data become available, greater demand is being placed on visualization tools to scale to the size of these new datasets. When viewing such large data, we necessarily cannot convey details, rather we specifically design overview tools to help elucidate large-scale patterns. Perceptual science, signal processing theory, and generality provide a framework for the design of such visualizations that can scale well beyond current approaches. We present Sequence Surveyor, a prototype that embodies these ideas for scalable multiple whole-genome alignment overview visualization. Sequence Surveyor visualizes sequences in parallel, displaying data using variable color, position, and aggregation encodings. We demonstrate how perceptual science can inform the design of visualization techniques that remain visually manageable at scale and how signal processing concepts can inform aggregation schemes that highlight global trends, outliers, and overall data distributions as the problem scales. These techniques allow us to visualize alignments with over 100 whole bacterial-sized genomes.

Index Terms—Bioinformatics Visualization, Perception Theory, Scalability Issues, Visual Design.

1 Introduction

Sequence comparison is a fundamental task in the biological sciences. Scientists often need to compare genomic sequences, for example, to understand evolution, to infer common function, or to identify differences. Because sequences are often too long for manual examination, scientists rely on alignment tools that automatically identify matching subsequences. Tools for visualizing these alignments are commonly used when performing sequence comparison. A variety of approaches for displaying and exploring alignments exist, and have been incorporated into a wide variety of tools. Procter et al. [28] presents a recent survey of many popular tools.

The amount of sequence information available is growing rapidly. Scientists are exploring larger numbers of genomes and longer genomes. However, most tools by design focus on providing in-depth exploration of a small set of sequences for predefined tasks. Focusing on low-level details obscures the task of tracing high-level trends in large datasets (cf. Figure 10a). Looking at larger datasets at this fine level of detail is overwhelming, and does not scale to growing datasets.

In this paper, we introduce a different type of tool for exploring large multiple genome alignment datasets: overview visualization. Sequence Surveyor, our prototype system shown in Figure 1, provides flexible views of large datasets. It allows scientists to examine patterns and trends in multiple genome alignment datasets of unprecedented scale, such as a set of 100 bacteria genomes (Figures 11 and 13). Such

- *Danielle Albers is with University of Wisconsin - Madison, E-mail: dalbers@cs.wisc.edu.*
- *Colin Dewey is with University of Wisconsin - Madison, E-mail: cdewey@biostat.wisc.edu.*
- *Michael Gleicher is with University of Wisconsin - Madison, E-mail: gleicher@cs.wisc.edu.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

an example to start

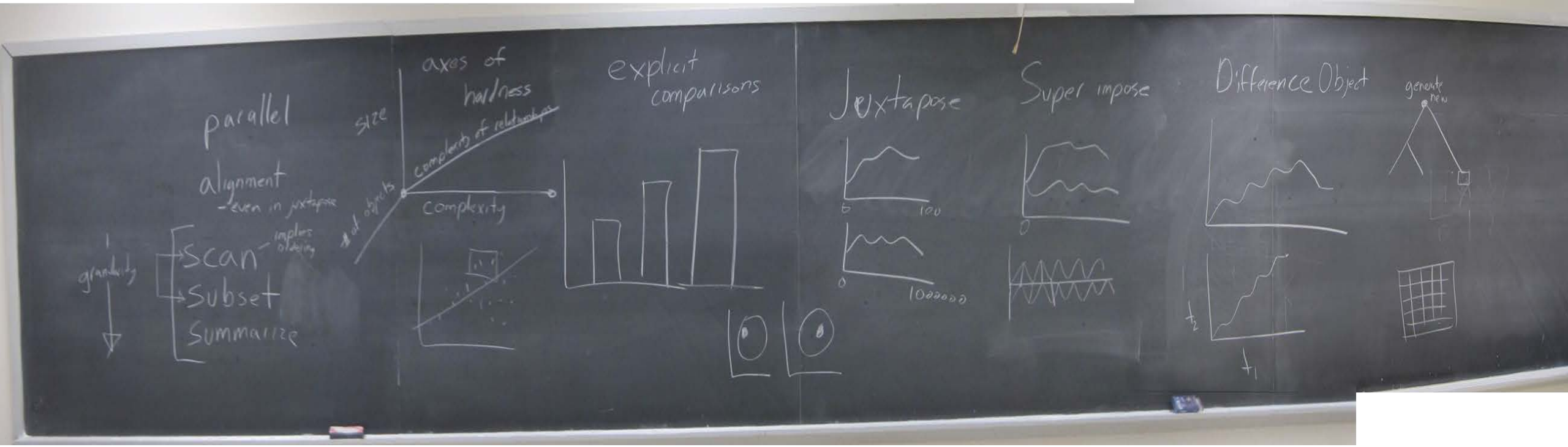
Sequence Surveyor

InfoVis 2011

Albers, D., Dewey, C., & **Gleicher**, M. (2011). Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2392–2401.

History

Your example is from 2011?



My CS838 (Data Visualization) class
March 11, 2010

What is the comparison?

Comparative Elements

Targets

Actions

Added 2015, inspired by Munzner

Why is it hard?

Comparative Challenges

Number of Targets

Large or Complex Targets

Complex Relationships

Reduced to 3 categories after 2008

How to address the challenges?

Scalability Strategies

Scan Sequentially

Select Subset

Summarize Somehow

Jason named this one in 2010

Which visual design to use?

Comparative Designs

Juxtapose

Superpose

Explicit Encoding

Danielle named this one in 2011

Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization

Danielle Albers, *Student Member, IEEE*, Colin Dewey, and Michael Gleicher, *Member, IEEE*

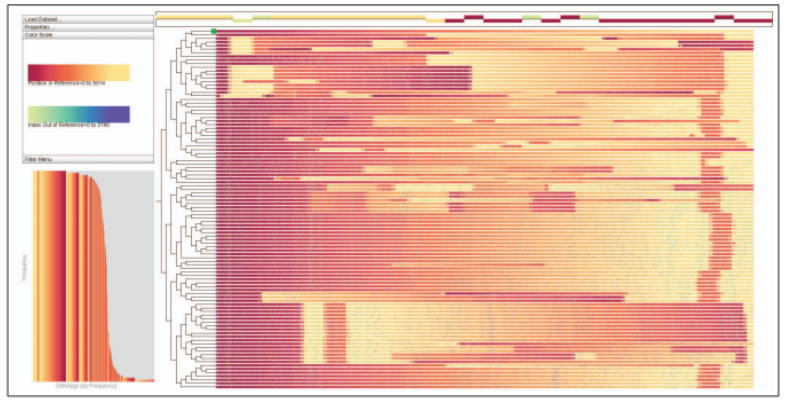


Fig. 1. Sequence Surveyor visualizing 100 synthetic genomes generated by an evolution simulation. Each genome is mapped to a row and genes are ordered by position. Color encodes the position of the gene within the chosen reference sequence (top row, indicated by the green box). Genes are aggregated, with each block's texture reflecting the overall distribution of colors in that block. The dendogram shows the phylogeny of the data set while the histogram shows the frequency distribution of orthology group sizes.

Abstract—In this paper, we introduce overview visualization tools for large-scale multiple genome alignment data. Genome alignment visualization and, more generally, sequence alignment visualization are an important tool for understanding genomic sequence data. As sequencing techniques improve and more data become available, greater demand is being placed on visualization tools to scale to the size of these new datasets. When viewing such large data, we necessarily cannot convey details, rather we specifically design overview tools to help elucidate large-scale patterns. Perceptual science, signal processing theory, and generality provide a framework for the design of such visualizations that can scale well beyond current approaches. We present Sequence Surveyor, a prototype that embodies these ideas for scalable multiple whole-genome alignment overview visualization. Sequence Surveyor visualizes sequences in parallel, displaying data using variable color, position, and aggregation encodings. We demonstrate how perceptual science can inform the design of visualization techniques that remain visually manageable at scale and how signal processing concepts can inform aggregation schemes that highlight global trends, outliers, and overall data distributions as the problem scales. These techniques allow us to visualize alignments with over 100 whole bacterial-sized genomes.

Index Terms—Bioinformatics Visualization, Perception Theory, Scalability Issues, Visual Design.

1 Introduction

Sequence comparison is a fundamental task in the biological sciences. Scientists often need to compare genomic sequences, for example, to understand evolution, to infer common function, or to identify differences. Because sequences are often too long for manual examination, scientists rely on alignment tools that automatically identify matching subsequences. Tools for visualizing these alignments are commonly used when performing sequence comparison. A variety of approaches for displaying and exploring alignments exist, and have been incorporated into a wide variety of tools. Procter et al. [28] presents a recent survey of many popular tools.

The amount of sequence information available is growing rapidly. Scientists are exploring larger numbers of genomes and longer genomes. However, most tools by design focus on providing in-depth exploration of a small set of sequences for predefined tasks. Focusing on low-level details obscures the task of tracing high-level trends in large datasets (cf. Figure 10a). Looking at larger datasets at this fine level of detail is overwhelming, and does not scale to growing datasets.

In this paper, we introduce a different type of tool for exploring large multiple genome alignment datasets: overview visualization. Sequence Surveyor, our prototype system shown in Figure 1, provides flexible views of large datasets. It allows scientists to examine patterns and trends in multiple genome alignment datasets of unprecedented scale, such as a set of 100 bacteria genomes (Figures 11 and 13). Such

- *Danielle Albers is with University of Wisconsin - Madison, E-mail: dalbers@cs.wisc.edu.*
- *Colin Dewey is with University of Wisconsin - Madison, E-mail: cdewey@biostat.wisc.edu.*
- *Michael Gleicher is with University of Wisconsin - Madison, E-mail: gleicher@cs.wisc.edu.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

an example to start

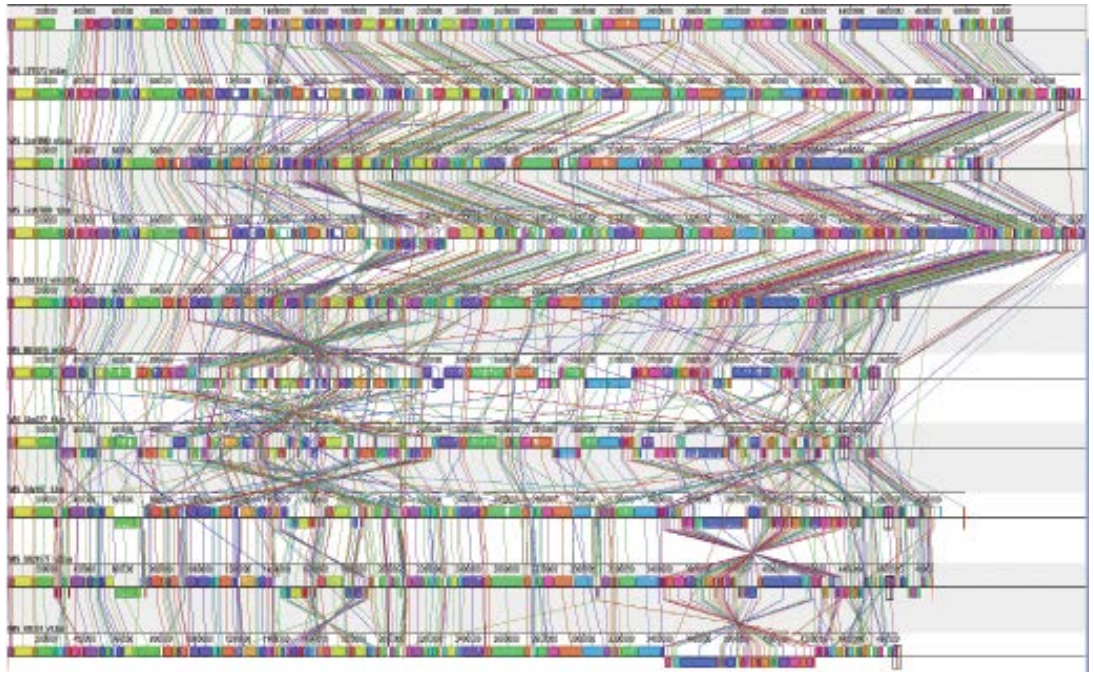
Sequence Surveyor

InfoVis 2011

Albers, D., Dewey, C., & **Gleicher**, M. (2011). Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2392–2401.

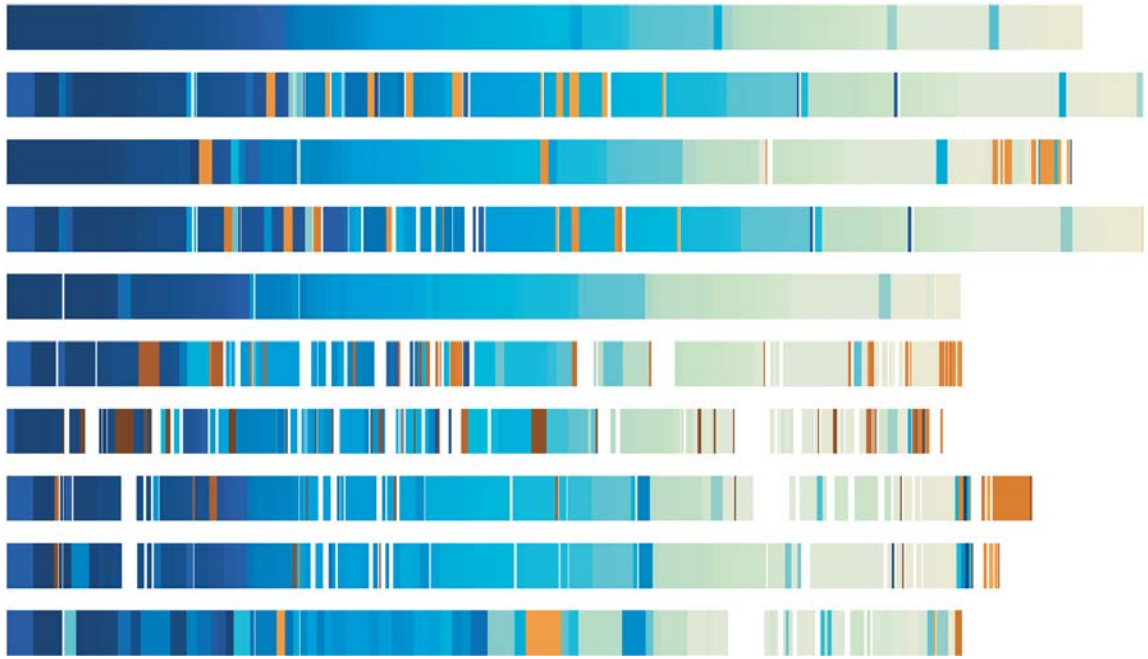
Sequence Surveyor: Scalable Genomic Sequence Comparison

Prior Art: Mauve



Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–403.

Sequence Surveyor



Albers, D., Dewey, C., & **Gleicher**, M. (2011). Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2392–2401.

What is the comparison?

Comparative Elements
Targets
Actions

Why is it hard?

Comparative Challenges
Number of Targets
Large or Complex Targets
Complex Relationships

How to address the challenges?

Scalability Strategies
Scan Sequentially
Select Subset
Summarize Somehow

Which visual design to use?

Comparative Designs
Juxtapose
Superpose
Explicit Encoding

What is the comparison?

Comparative Elements
Targets
Actions

Why is it hard?

Comparative Challenges
Number of Targets
Large or Complex Targets
Complex Relationships

How to address the challenges?

Scalability Strategies
Scan Sequentially
Select Subset
Summarize Somehow

Which visual design to use?

Comparative Designs
Juxtapose
Superpose
Explicit Encoding

Question 1:
What is the comparison?

Question 1:

~~What is the comparison?~~

What are the elements of the comparison?

The Elements of a Comparison . . .

To examine
(two or more objects, ideas, etc.)
in order to note
similarities and differences

To mark or point out the
similarities and differences of
(two or more things)

The Elements of a Comparison . . .

To examine
(**two or more objects, ideas, etc.**)
in order to note
similarities and differences

To mark or point out the
similarities and differences of
(**two or more things**)

Targets — Set of things being compared

The Elements of a Comparison . . .

To **examine**
(**two or more objects, ideas, etc**.)
in order to note
similarities and differences

To **mark** or **point** out the
similarities and differences of
(**two or more things**)

Targets — Set of things being compared

Action — What to do

The Elements of a Comparison . . .

To **examine**
(**two or more objects, ideas, etc.**)
in order to note
similarities and differences

To **mark** or **point** out the
similarities and differences of
(**two or more things**)

Targets — Set of things being compared

Action — What to do with the **relationship** among them

Question 1A:
The Elements: Targets

Do you know what you are comparing?

Explicit Comparisons – the system has the set of targets

Implicit Comparisons – the system may not know all the targets

compare against an implicit baseline

compare against the user's knowledge

Questions of naming

Question 1A:
The Elements: Targets

Do you know what you are comparing?

Explicit Comparisons – the system has the set of targets

Implicit Comparisons – the system may not know all the targets

compare against an implicit baseline

compare against the user's knowledge

Questions of **naming**

Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization

Alignment Visualization

Danielle Albers, *Student Member, IEEE*, Colin Dewey, and Michael Gleicher, *Member, IEEE*

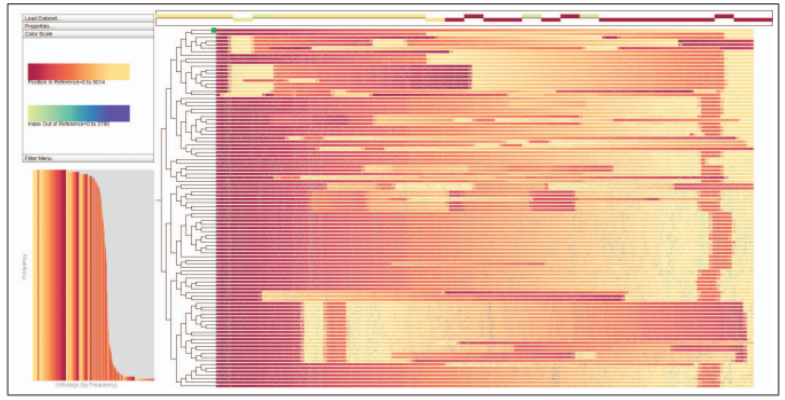


Fig. 1. Sequence Surveyor visualizing 100 synthetic genomes generated by an evolution simulation. Each genome is mapped to a row and genes are ordered by position. Color encodes the position of the gene within the chosen reference sequence (top row, indicated by the green box). Genes are aggregated, with each block's texture reflecting the overall distribution of colors in that block. The dendogram shows the phylogeny of the data set while the histogram shows the frequency distribution of orthology group sizes.

Abstract—In this paper, we introduce overview visualization tools for large-scale multiple genome alignment data. Genome alignment visualization and, more generally, sequence alignment visualization are an important tool for understanding genomic sequence data. As sequencing techniques improve and more data become available, greater demand is being placed on visualization tools to scale to the size of these new datasets. When viewing such large data, we necessarily cannot convey details, rather we specifically design overview tools to help elucidate large-scale patterns. Perceptual science, signal processing theory, and generality provide a framework for the design of such visualizations that can scale well beyond current approaches. We present Sequence Surveyor, a prototype that embodies these ideas for scalable multiple whole-genome alignment overview visualization. Sequence Surveyor visualizes sequences in parallel, displaying data using variable color, position, and aggregation encodings. We demonstrate how perceptual science can inform the design of visualization techniques that remain visually manageable at scale and how signal processing concepts can inform aggregation schemes that highlight global trends, outliers, and overall data distributions as the problem scales. These techniques allow us to visualize alignments with over 100 whole bacterial-sized genomes.

Index Terms—Bioinformatics Visualization, Perception Theory, Scalability Issues, Visual Design.

1 Introduction

Sequence comparison is a fundamental task in the biological sciences. Scientists often need to compare genomic sequences, for example, to understand evolution, to infer common function, or to identify differences. Because sequences are often too long for manual examination, scientists rely on alignment tools that automatically identify matching subsequences. Tools for visualizing these alignments are commo
for displaying and exploring alignments exist, and have been incor
rated into a wide variety of tools. Procter et al. [28] presents a rec
survey of many popular tools.

The amount of sequence information available is growing rapidly. Scientists are exploring larger numbers of genomes and longer genomes. However, most tools by design focus on providing in-depth exploration of a small set of sequences for predefined tasks. Focusing on low-level details obscures the task of tracing high-level trends in large datasets (cf. Figure 10a). Looking at larger datasets at this fine level of detail is overwhelming, and does not scale to growing datasets.

In this paper, we introduce a different type of tool for exploring large multiple genome alignment datasets: overview visualization. Sequence Surveyor, our prototype system shown in Figure 1, provides flexible views of large datasets. It allows scientists to examine patterns and trends in multiple genome alignment datasets of unprecedented scale, such as a set of 100 bacteria genomes (Figures 11 and 13). Such

- *Danielle Albers is with University of Wisconsin - Madison, E-mail: dalbers@cs.wisc.edu.*
- *Colin Dewey is with University of Wisconsin - Madison, E-mail: cdewey@biostat.wisc.edu.*
- *Michael Gleicher is with University of Wisconsin - Madison, E-mail: gleicher@cs.wisc.edu.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

1 Introduction

Sequence comparison is a fundamental task in the biological sciences. Scientists often need to compare genomic sequences, for example, to understand evolution, to infer common function, or to identify differences. Because sequences are often too long for manual examination, scientists rely on alignment tools that automatically identify matching

Question 1B:
The Elements: Actions

Verbs on relationships

Try to be more specific than “examine” or “compare”

Truth in Advertising: I didn’t have this worked out in 2010

Question 2:
Why is this comparison hard?

If it isn't hard, you probably don't need to think about it (much)

Abstractly

Too many targets to compare

Large or Complex Targets

Complex Relationships

Challenges of Scale!

Sequence Surveyor

Question 2:
Why is this comparison hard?

If it isn't hard, you probably don't need to think about it (much)

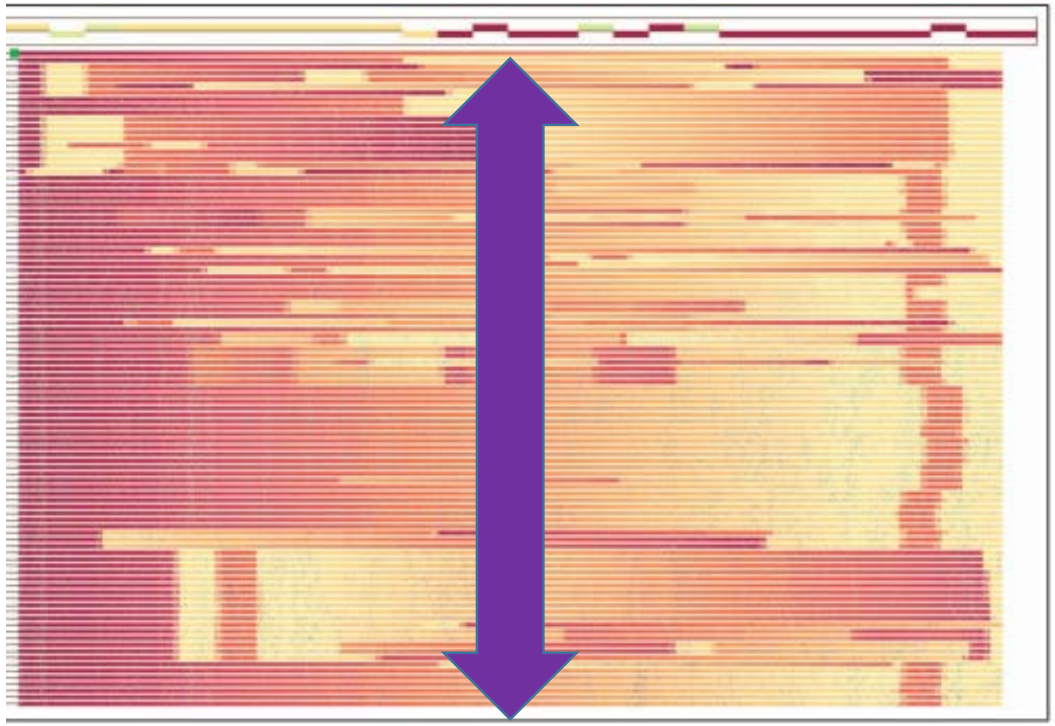
Abstractly

Too many targets to compare

Large or Complex Targets

Complex Relationships

Sequence Surveyor



Question 2:
Why is this comparison hard?

If it isn't hard, you probably don't need to think about it (much)

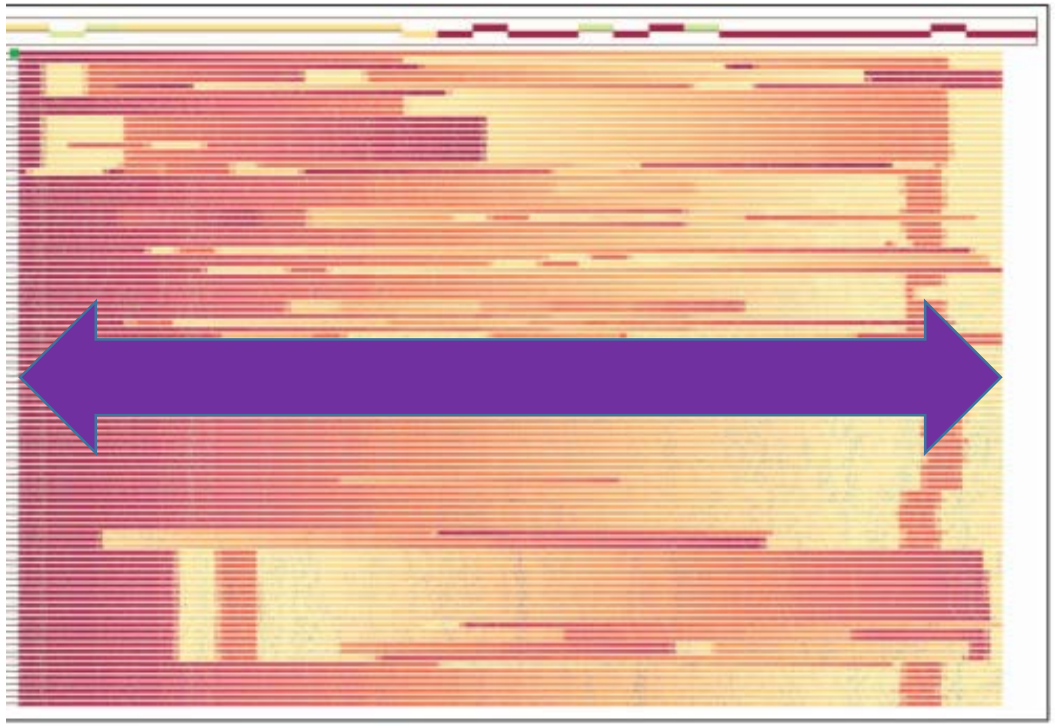
Abstractly

Too many targets to compare

Large or Complex Targets

Complex Relationships

Sequence Surveyor



Question 2:
Why is this comparison hard?

If it isn't hard, you probably don't need to think about it (much)

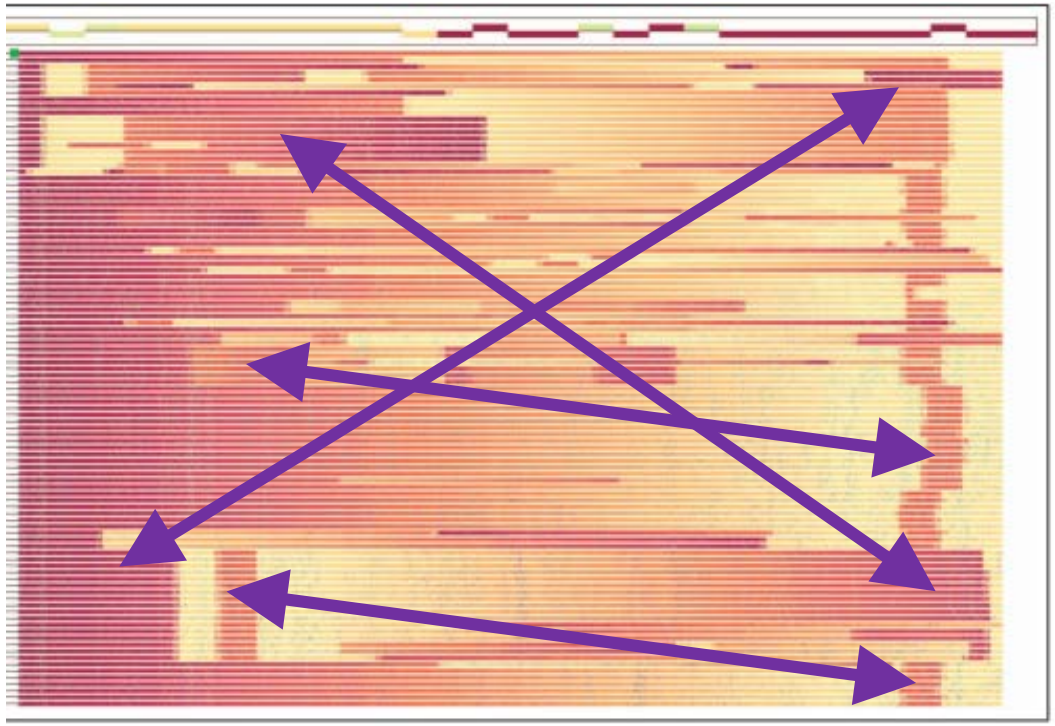
Abstractly

Too many targets to compare

Large or Complex Targets

Complex Relationships

Sequence Surveyor



Question 2:
Why is this comparison hard?

If it isn't hard, you probably don't need to think about it (much)

Abstractly

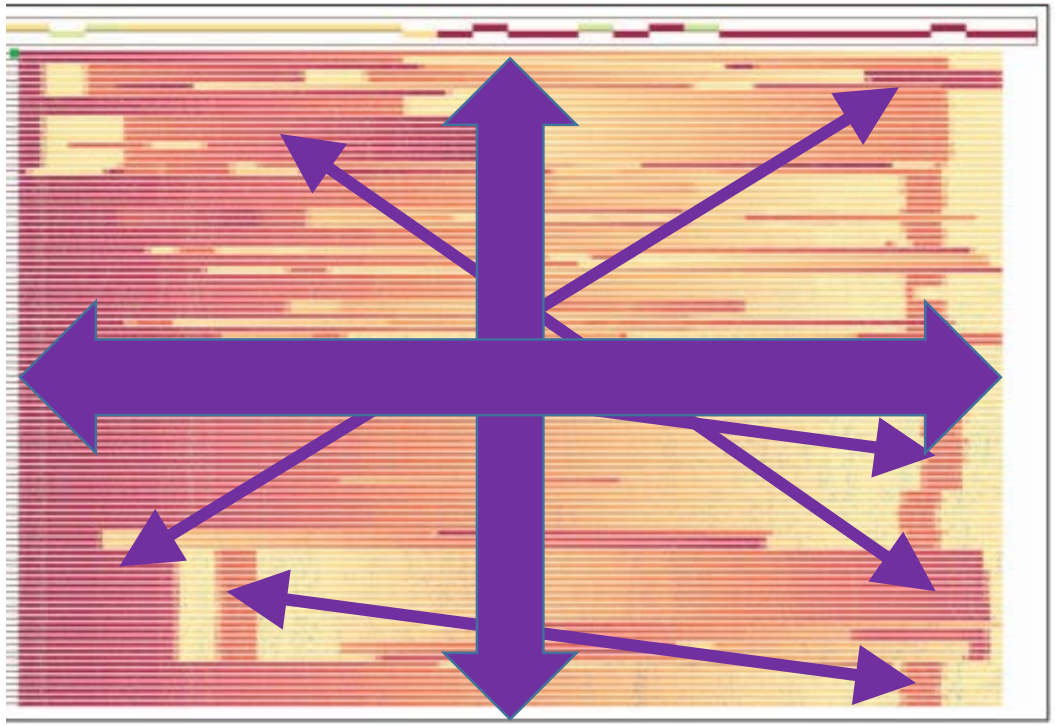
Too many targets to compare

Large or Complex Targets

Complex Relationships

Sequence Surveyor has all three!

Sequence Surveyor



Question 3:
What is your strategy for those challenges?

Abstractly

Scan Sequentially

Select Subset

Summarize Somehow

Sequence Surveyor

Scalability Strategies!

Question 3:
What is your strategy for those challenges?

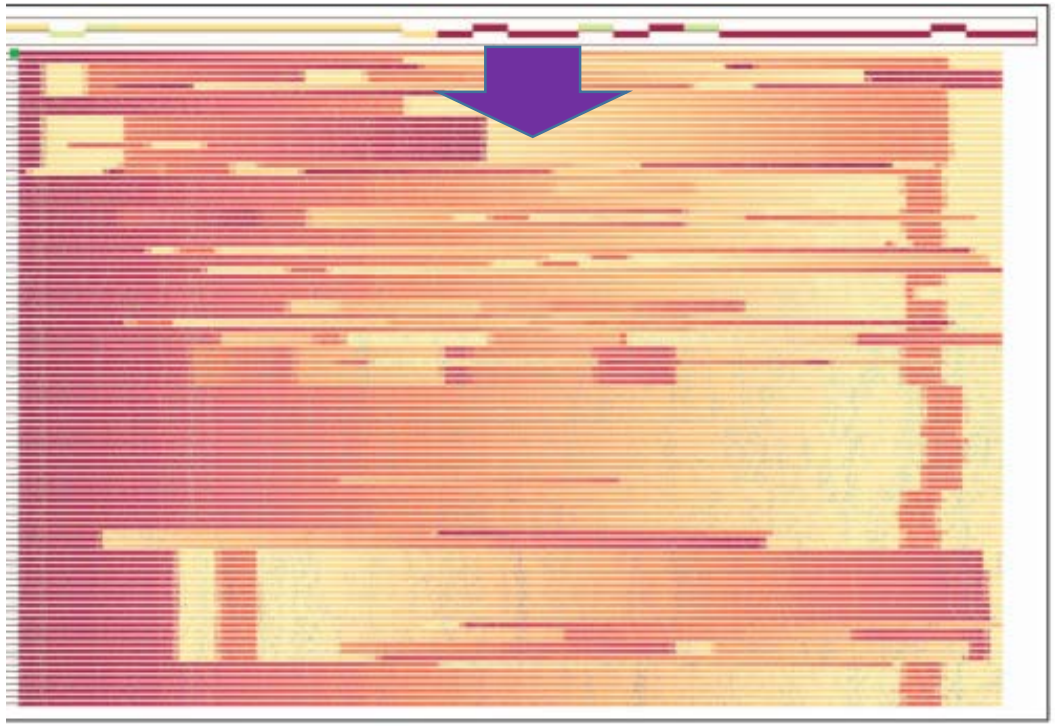
Abstractly

Scan Sequentially

Select Subset

Summarize Somehow

Sequence Surveyor



Question 3:
What is your strategy for those challenges?

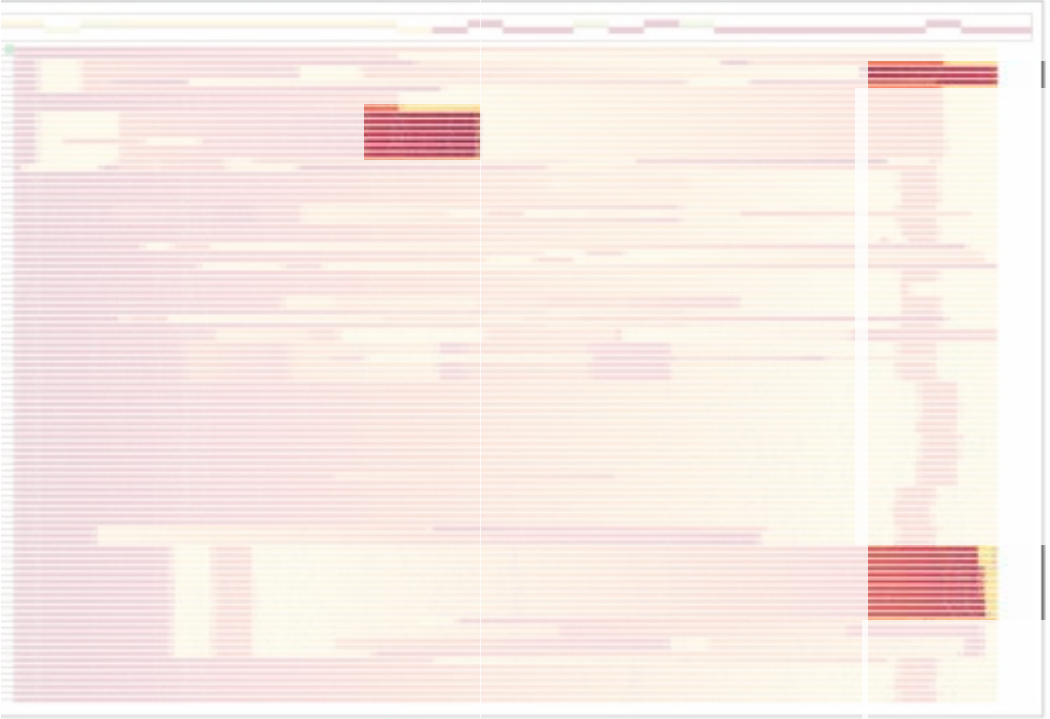
Abstractly

Scan Sequentially

Select Subset

Summarize Somehow

Sequence Surveyor



Question 3:
What is your strategy for those challenges?

Abstractly

Scan Sequentially

Select Subset

Summarize Somehow

Sequence Surveyor

Question 3:
What is your strategy for those challenges?

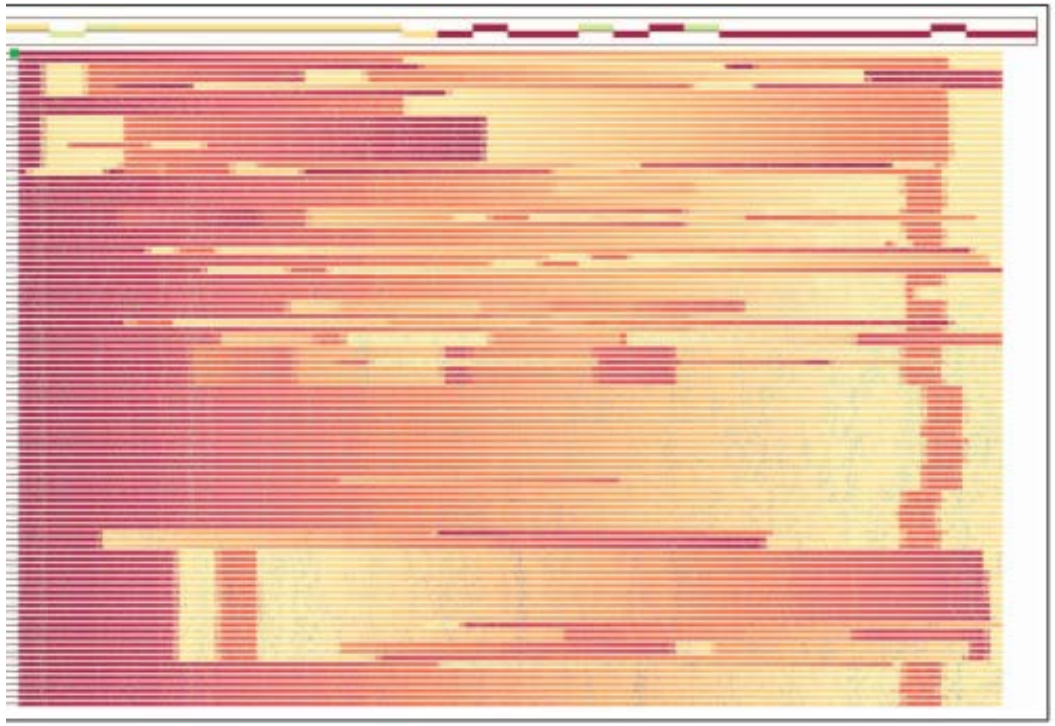
Abstractly

Scan Sequentially

Select Subset

Summarize Somehow

Sequence Surveyor

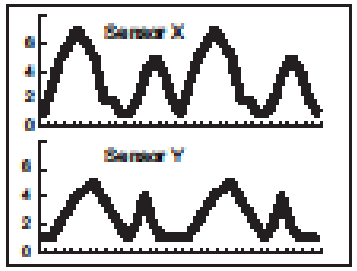


Sequence Surveyor used all 3 strategies

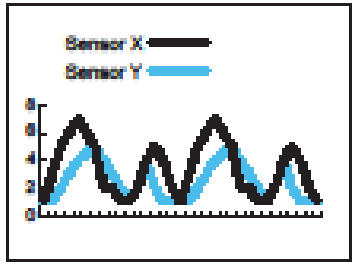
Question 4:
What Visual Design for Comparison?

Abstractly

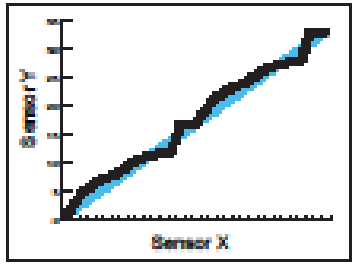
Sequence Surveyor



Juxtaposition



Superposition

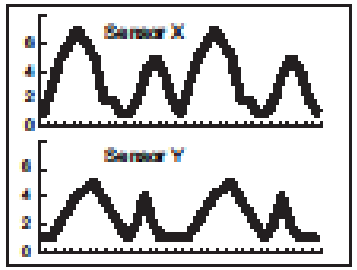


Explicit Encoding

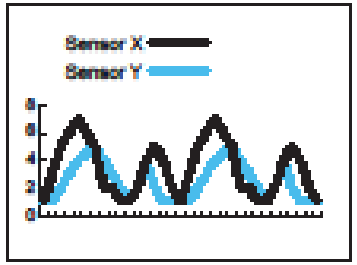
Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, *10*(4), 289–309.

Question 4:
What Visual Design for Comparison?

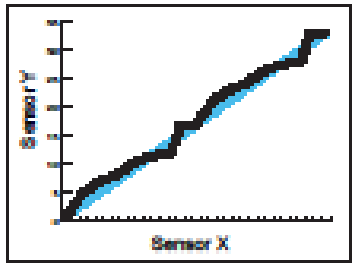
Abstractly



Juxtaposition



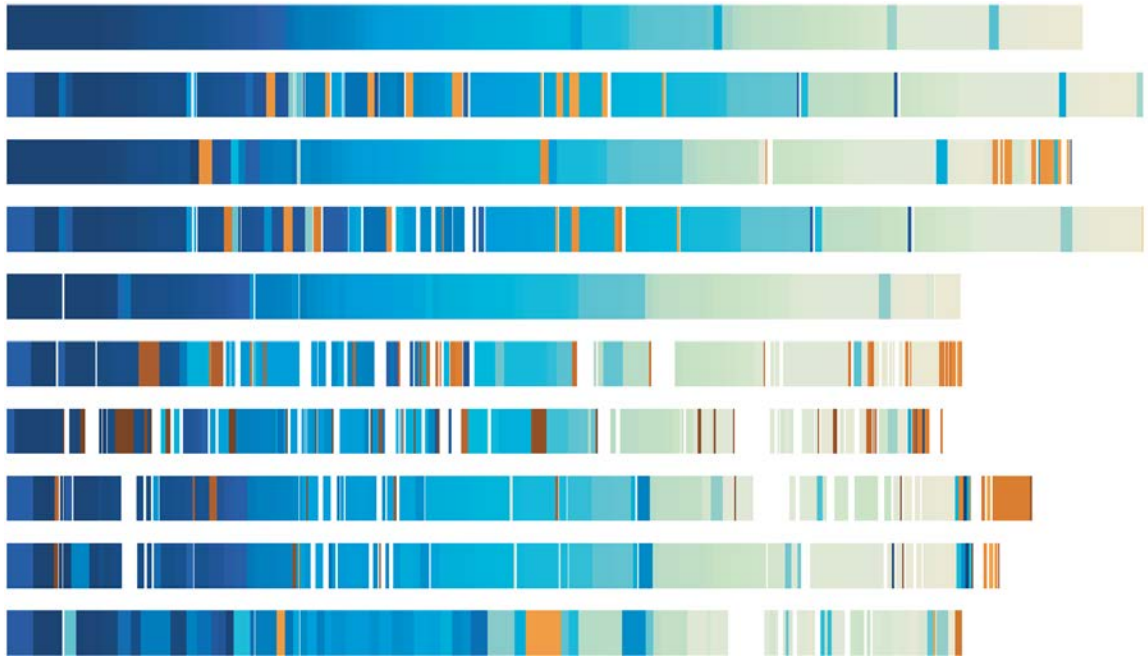
Superposition



Explicit Encoding

Sequence Surveyor

Juxtaposition

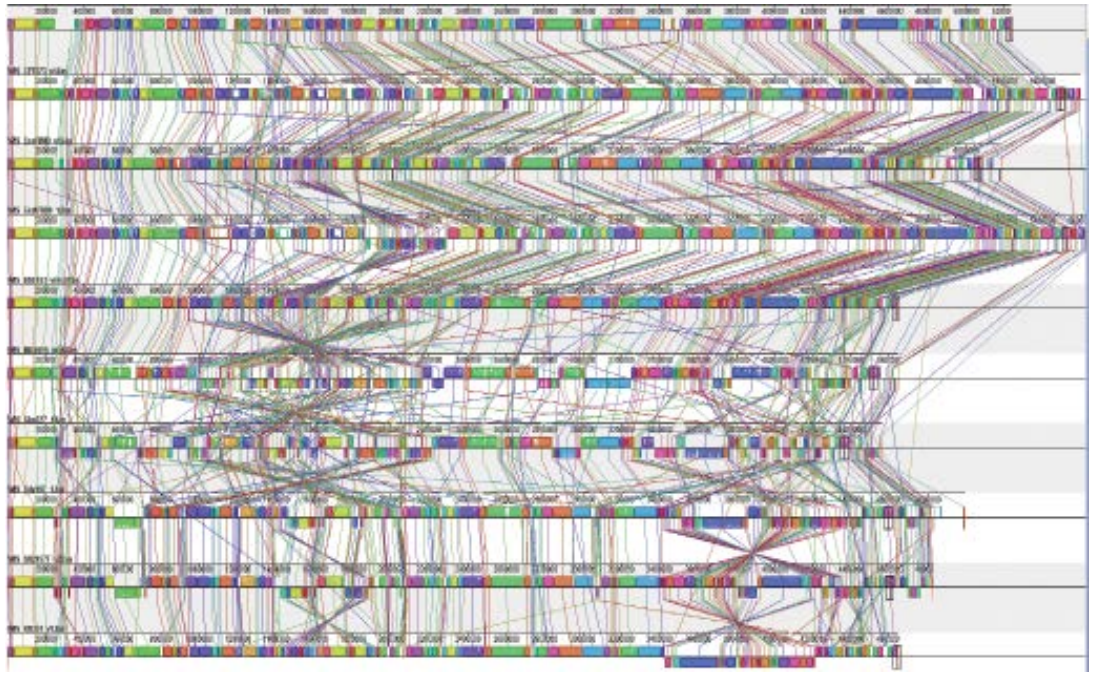


Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, *10*(4), 289–309.

Comparison with the standard

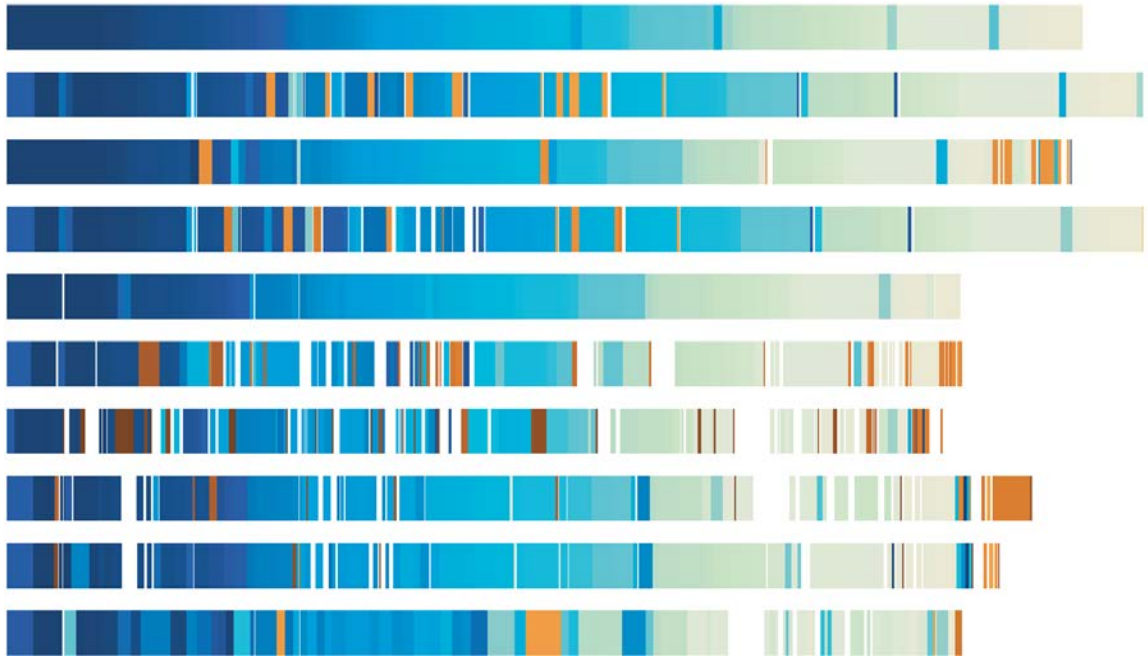
Improve scalability with a different strategy and design

Mauve



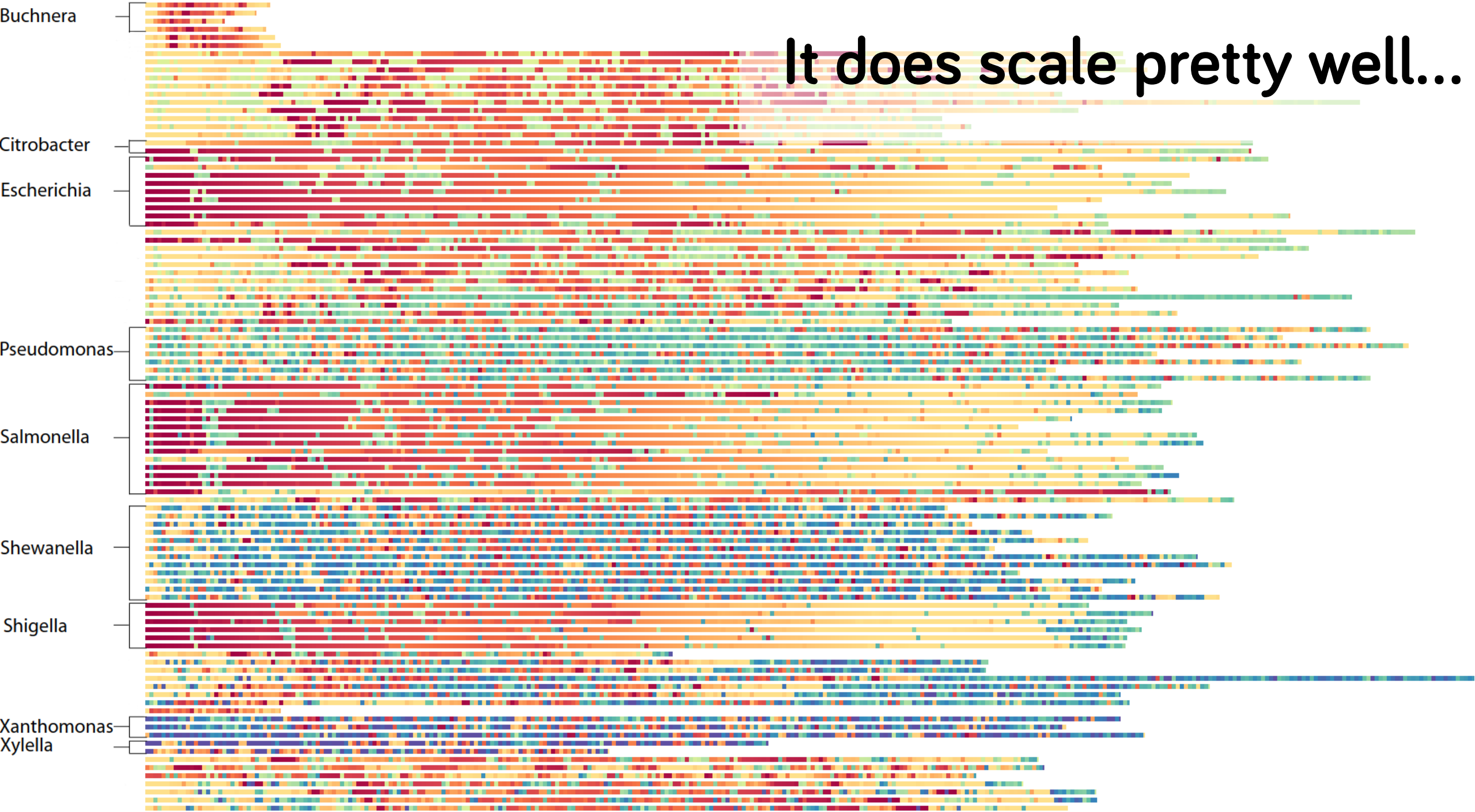
Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–403.

Sequence Surveyor



Albers, D., Dewey, C., & **Gleicher**, M. (2011). Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2392–2401.

It does scale pretty well...



Should I think about comparison?

Maybe… If it helps

What is the comparison?

Comparative Elements
Targets
Actions

Why is it hard?

Comparative Challenges
Number of Targets
Large or Complex Targets
Complex Relationships

How to address the challenges?

Scalability Strategies
Scan Sequentially
Select Subset
Summarize Somehow

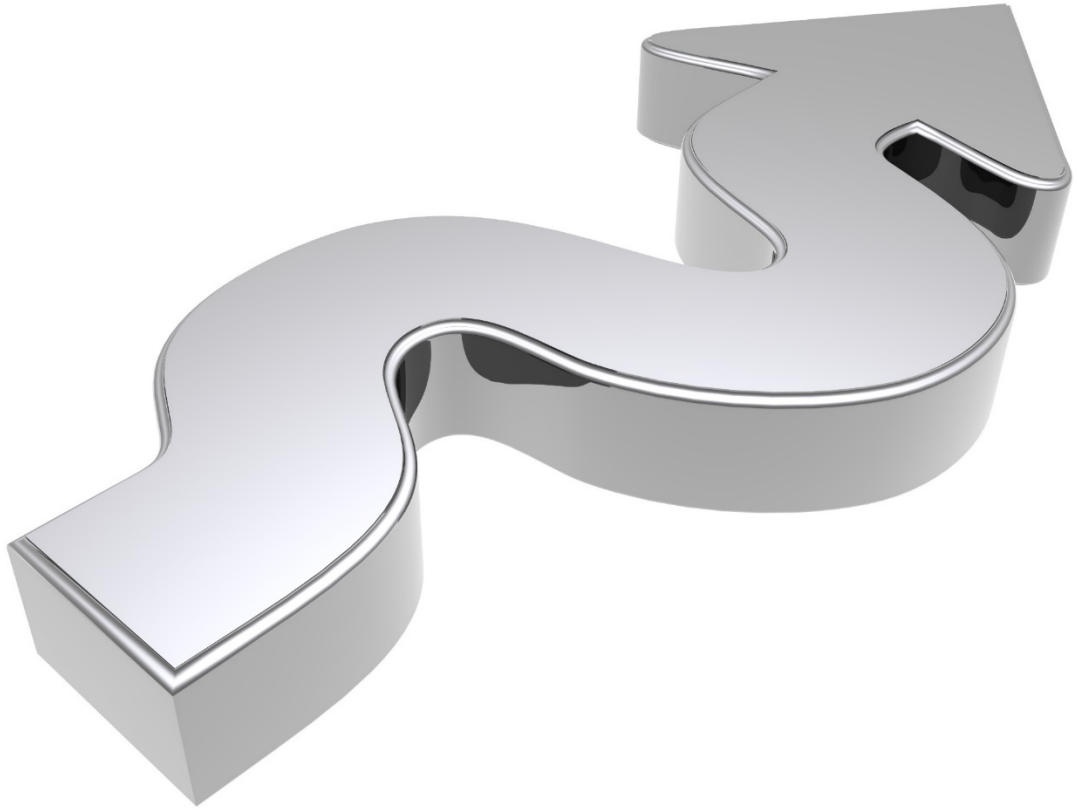
Which visual design to use?

Comparative Designs
Juxtapose
Superpose
Explicit Encoding

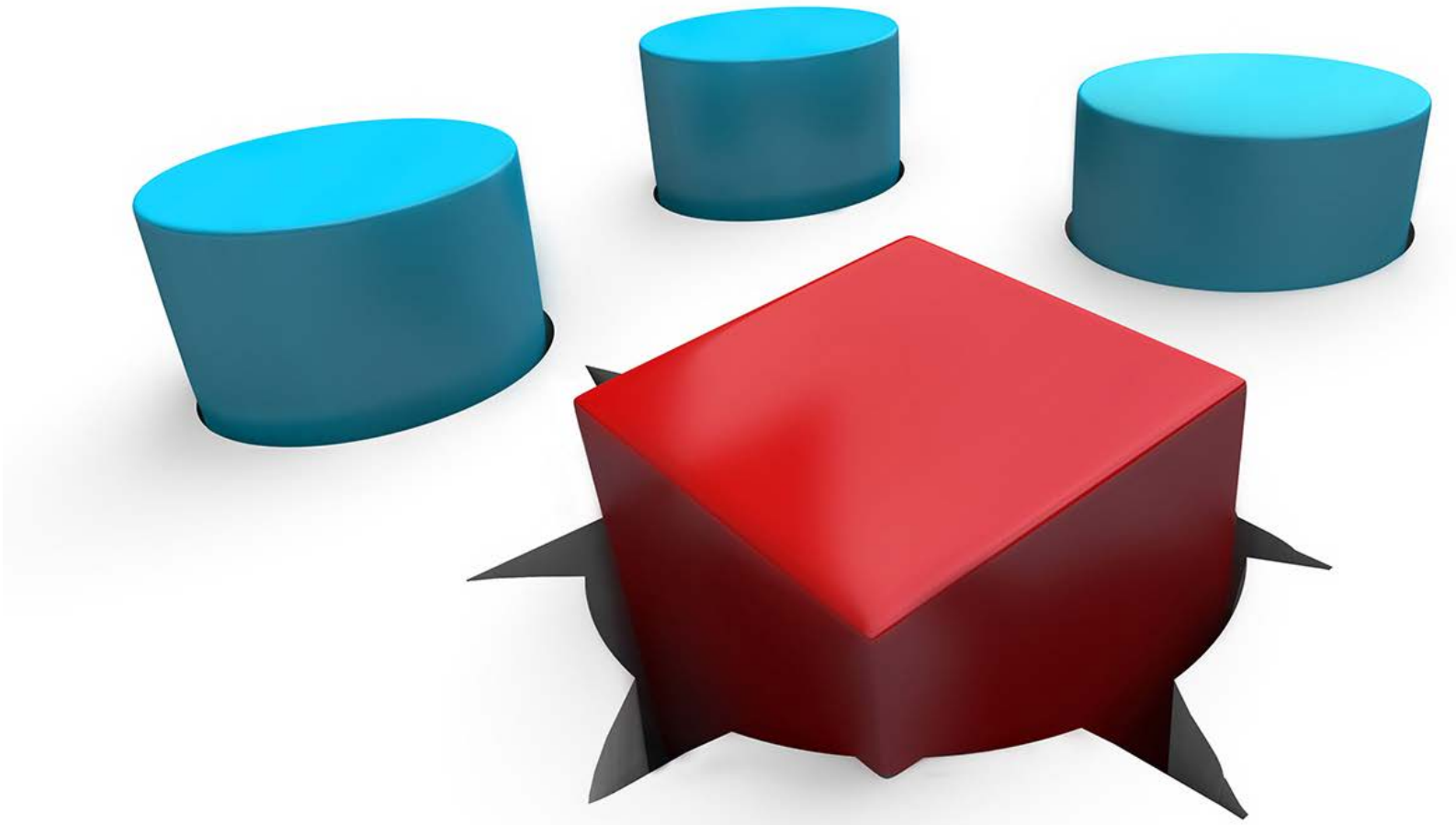
Do I have to follow this order?

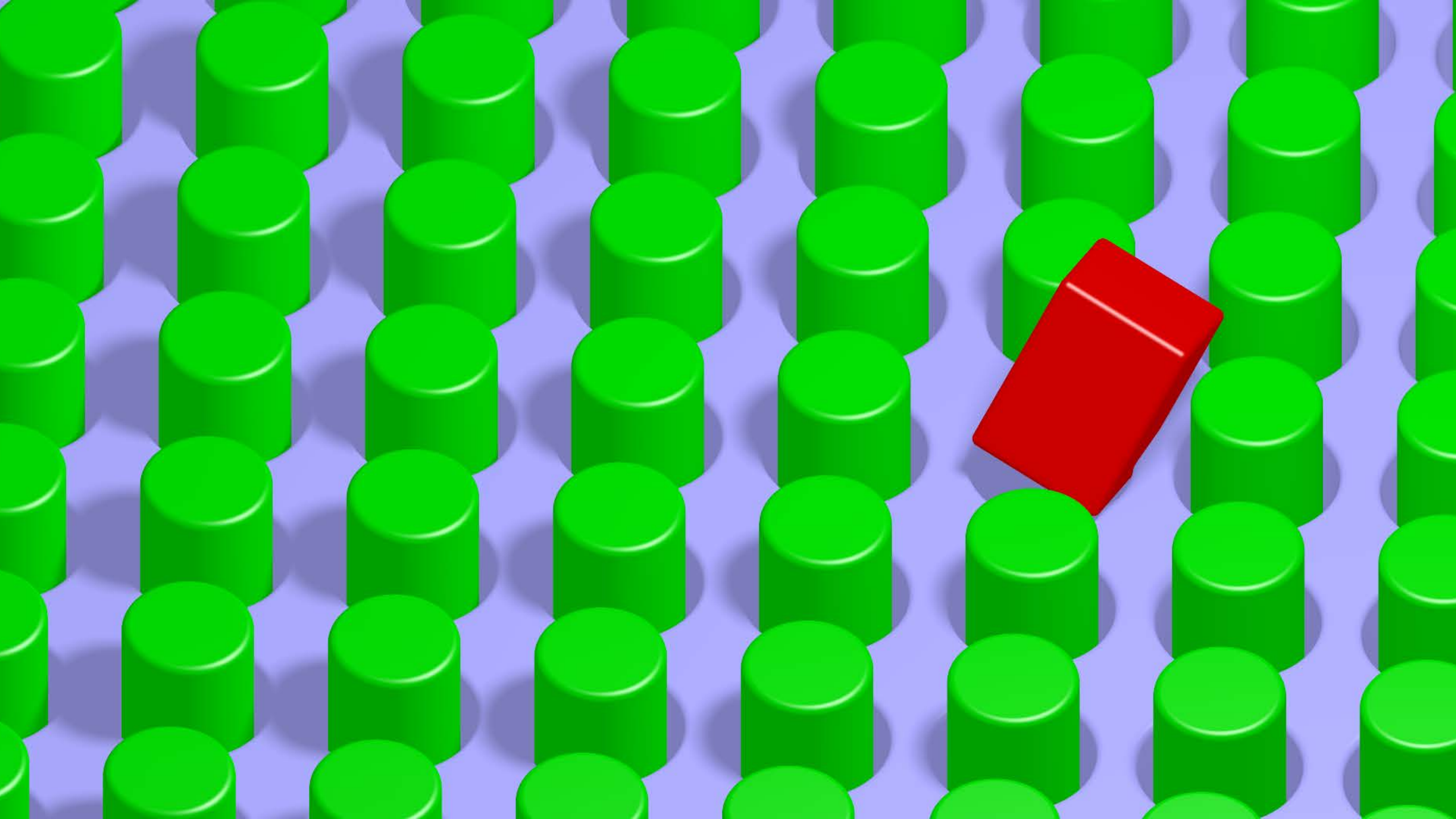
Yes! Follow them in logical order as a process!

No! Ask any question at any time!



Does everything fit into a category?





Does it help to think about these categories?

Wrong Question:
Is my problem Comparison?

Just about anything can be viewed as comparison

Not everything benefits from being viewed this way

Serendip: Topic Model-Driven Visual Exploration of Text Corpora

Eric Alexander, Joe Kohlmann, Robin Valenza, Michael Witmore, and Michael Gleicher, *Member, IEEE*

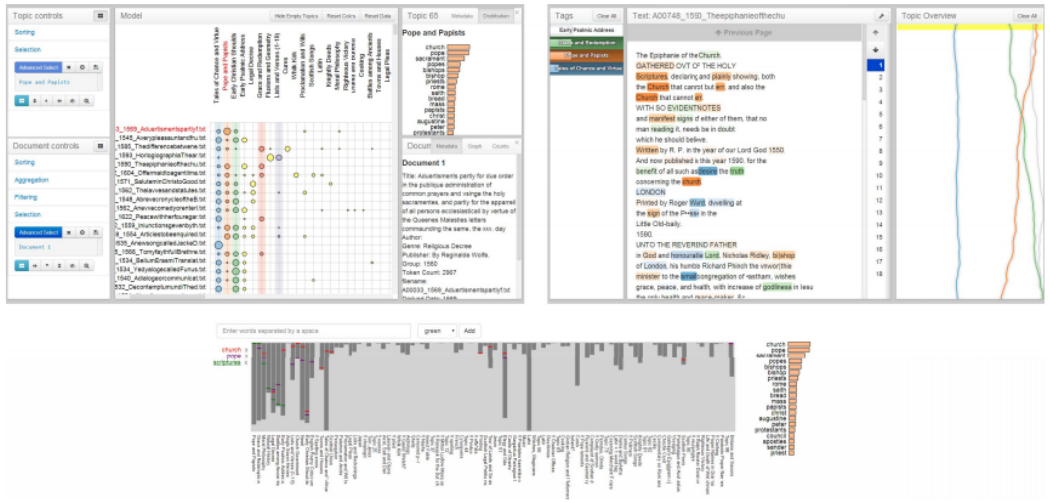


Fig. 1. The three main views of Serendip: CorpusViewer, TextViewer, and RankViewer.

Abstract— Exploration and discovery in a large text corpus requires investigation at multiple levels of abstraction, from a zoomed-out view of the entire corpus down to close-ups of individual passages and words. At each of these levels, there is a wealth of information that can inform inquiry—from statistical models, to metadata, to the researcher's own knowledge and expertise. Joining all this information together can be a challenge, and there are issues of scale to be combatted along the way. In this paper, we describe an approach to text analysis that addresses these challenges of scale and multiple information sources, using probabilistic topic models to structure exploration through multiple levels of inquiry in a way that fosters serendipitous discovery. In implementing this approach into a tool called Serendip, we incorporate topic model data and metadata into a highly reorderable matrix to expose corpus level trends; extend encodings of tagged text to illustrate probabilistic information at a passage level; and introduce a technique for visualizing individual word rankings, along with interaction techniques and new statistical methods to create links between different levels and information types. We describe example uses from both the humanities and visualization research that illustrate the benefits of our approach.

Index Terms—Text visualization, topic modeling.

1 Introduction

Exploration and discovery in large text corpora can be a daunting task. Corpora can easily grow to thousands or more texts, ranging in length from short snippets to long books. The task is further complicated by the range of questions that can be asked of such corpora, broad both in subject (making comparisons across time, genre, author, etc.) and in level of detail (corpus, document, passage, even word). Discoveries must often connect multiple subjects and levels of inquiry. Fortunately, there is considerable information to aid these inquiries. Beyond the texts themselves, there are statistical summaries of content, document metadata, and analysts' explicit and implicit knowledge of the documents and their context. However, mixing these different types of information across scales of inquiry is challenging. The information types, and the existing tools that support their use, generally focus solely on a particular scale.

In this paper, we introduce a topic modeling tool for text exploration that is designed to address the issues of inter-mixing scales of inquiry and information types. Our core idea is that to enable fluent fusion, a system must provide not only a set of views for looking at the data from multiple viewpoints, but also connections between the different types of information allowing a reader to move smoothly across scales, data types, and research questions. To achieve this, we have had to adapt existing views to work with different types of text corpora data, develop new views that address some unmet needs, and introduce statistical methods that help connect between different object types. The resulting system enables users to explore questions about collections

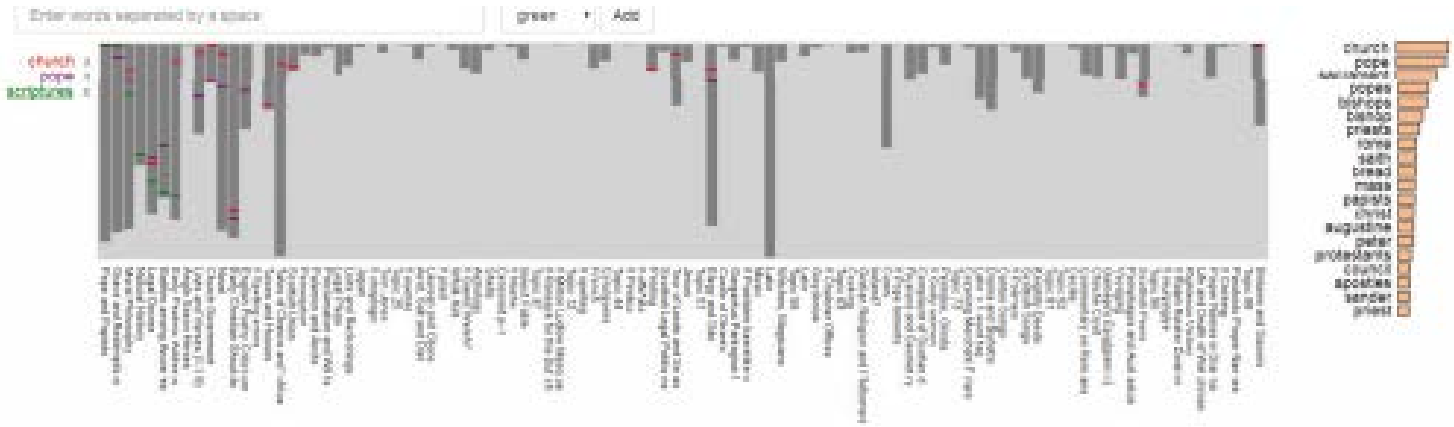
- *E. Alexander, J. Kohlmann, and M. Gleicher are with the Department of Computer Sciences at the University of Wisconsin-Madison. E-mail: ealexand@cs.wisc.edu, kohlmannj@gmail.com, gleicher@cs.wisc.edu.*
- *R. Valenza is with the Department of English at the University of Wisconsin-Madison. Email: valenza@wisc.edu.*
- *M. Witmore is with the Folger Shakespeare Library in Washington, D.C. Email: mwitmore@folger.edu.*

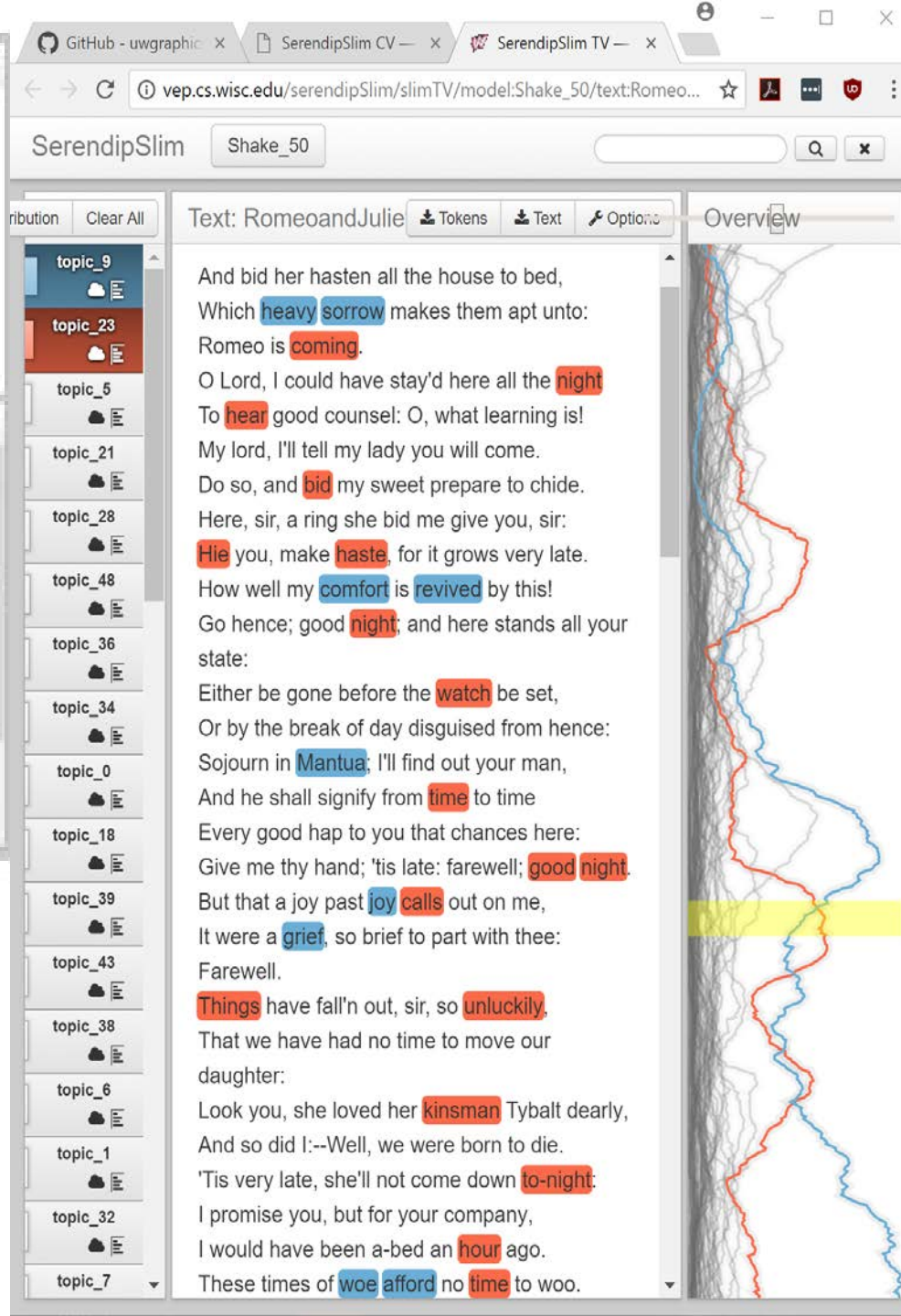
IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

Serendip, VAST '14

Alexander, E., Kohlmann, J., Valenza, R., Witmore, M., & Gleicher, M. (2014). Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*


Topic controls
Model
Topic 65
Pope and Papists
Document controls
Document 1




GitHub - uwgraphic
SerendipSlim CV —
SerendipSlim TV —
vep.cs.wisc.edu/serendipSlim/slimTV/model:Shake_50/text:Romeo...
SerendipSlim
Shake_50
Clear All
Text: RomeoandJulie
Tokens
Text
Overview
topic_9
topic_23
topic_5
topic_21
topic_28
topic_48
topic_36
topic_34
topic_0
topic_18
topic_39
topic_43
topic_38
topic_6
topic_1
topic_32
topic_7
And bid her hasten all the house to bed,
Which heavy sorrow makes them apt unto:
Romeo is coming.
O Lord, I could have stay'd here all the night
To hear good counsel: O, what learning is!
My lord, I'll tell my lady you will come.
Do so, and bid my sweet prepare to chide.
Here, sir, a ring she bid me give you, sir:
Hie you, make haste, for it grows very late.
How well my comfort is revived by this!
Go hence; good night; and here stands all your state:
Either be gone before the watch be set,
Or by the break of day disguised from hence:
Sojourn in Mantua; I'll find out your man,
And he shall signify from time to time
Every good hap to you that chances here:
Give me thy hand; 'tis late: farewell; good night.
But that a joy past joy calls out on me,
It were a grief, so brief to part with thee:
Farewell.
Things have fall'n out, sir, so unluckily,
That we have had no time to move our daughter:
Look you, she loved her kinsman Tybalt dearly,
And so did I:--Well, we were born to die.
'Tis very late, she'll not come down to-night:
I promise you, but for your company,
I would have been a-bed an hour ago.
These times of woe afford no time to woo.

Is this comparison?

No!

A tool for exploring a topic model!

We didn't describe it as comparison

Tool for looking at one topic model

Unclear how users think about it

Yes!

Comparison thinking really helped!

We did think about comparison

Tool for using topic models

Our users had comparison tasks

~~Is this comparison?~~ I don't care!

No!

A tool for exploring a topic model!

A survey of comparison would have missed this.

parison

model

Unclear how users think about it

Yes!

Comparison thinking really helped!

We did

Tool for

It's a great example of comparison ideas

Our users had comparison tasks

Question 1A:
The Elements: Targets

Do you know what you are comparing?

Explicit Comparisons – the system has the set of targets

Implicit Comparisons – the system may not know all the targets

compare against an implicit baseline

compare against the user's knowledge

Questions of **naming**

Question 1A:
The Elements: Targets

Start with user tasks / questions

Does the model match my expectations?

What documents are similar?

How do groups of documents differ?

What words indicate these differences?

How are words used differently?

Where in texts are these differences?

Do the patterns match other things I know?

Question 1A:
The Elements: Targets

What is being compared? – Comparison Targets

Does the **model** match my expectations?

What **documents** are similar?

How do **groups of documents** differ?

What **words** indicate these differences?

How are **words** used differently?

Where in **texts** are these differences?

Do the **patterns** match other things I know?

Question 1B:
The Elements: Actions

What to do with the relationship? Comparison Actions (Verbs)

Does the **model** match my expectations?	**Measure/Quantify** relationship
What **documents** are similar?	**Identify** similar things
How do **groups of documents** differ?	**Measure/Quantify** relationship
What **words** indicate these differences?	**Dissect** a difference
How are **words** used differently?	**Identify** meaningful differences
Where in **texts** are these differences?	**Contextualize** the relationships
Do the **patterns** match other things I know?	**Identify** similar things

Actions:
Are these the right “names”?

The important things:

- More specific than “compare”
- Actionable verbs on relationships

I find this list useful to start with

Disclaimer:

It’s not as evolved as a “real” task taxonomy

Identify

Measure/Quantify/Summarize

Dissect

Connect

Contextualize

Communicate/Illuminate

Question 2:
Why is this comparison hard?

If it isn't hard, you probably don't need to think about it (much)

Abstractly

Too many targets to compare

Large or Complex Targets

Complex Relationships

Serendip

Challenges of Scale!

Question 3:
What is your strategy for those challenges?

Abstractly

Scan Sequentially

Select Subset

Summarize Somehow

Solutions for Challenges of Scale!

Only Scalability Challenges?

Many different comparisons

Challenges from the kind (not scale)

- Hard target types (implicit)
- Hard action types (dissection)
- Hard combinations (dissect implicit)

Tasks influence scalability challenges

Solutions must respond to both!

Serendip Comparison Example: Compare Groups

Task Challenge:

Implicit targets – what groups?

Strategy: make explicit

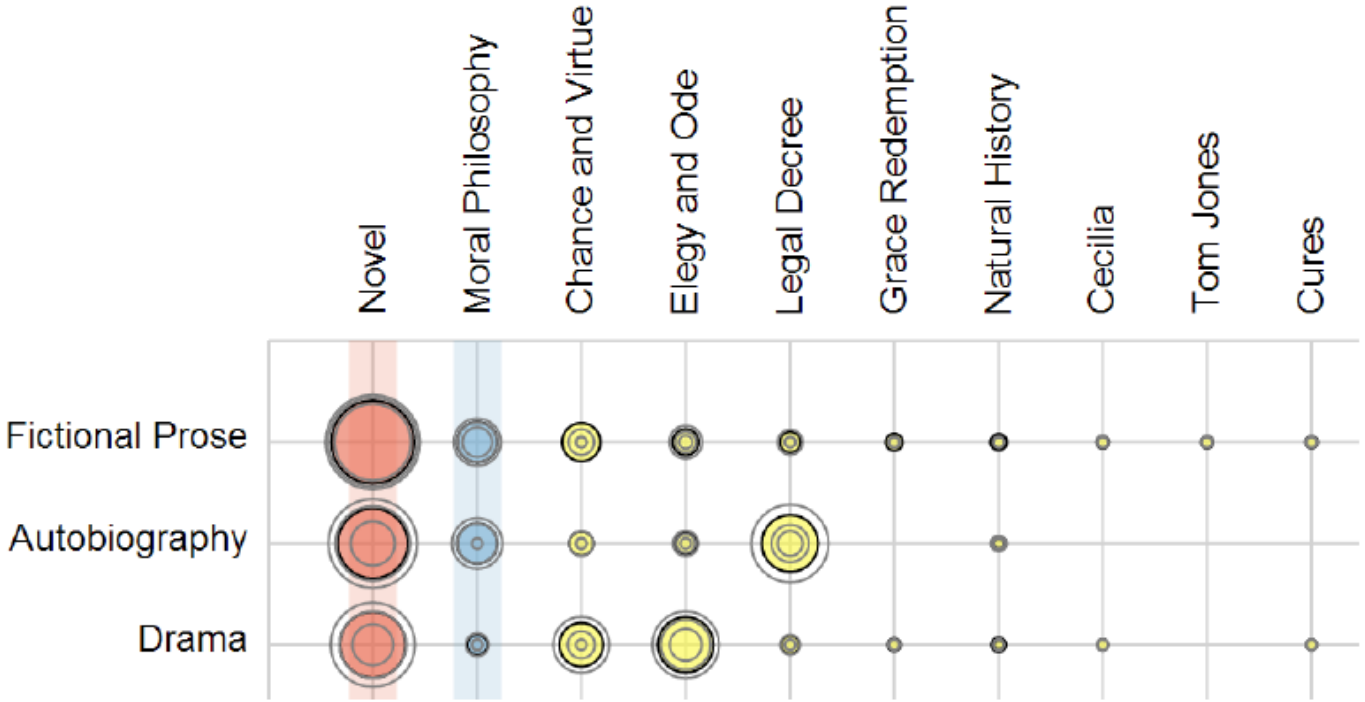
Design: user specifies groups

Scalability Challenge:

Lots of documents

Strategy: summarize

Design: how to present statistics



Serendip Comparison Example: Where does this happen? (contextualize)

Task Challenge:

Contextualize – fit user knowledge

Strategy: show in context

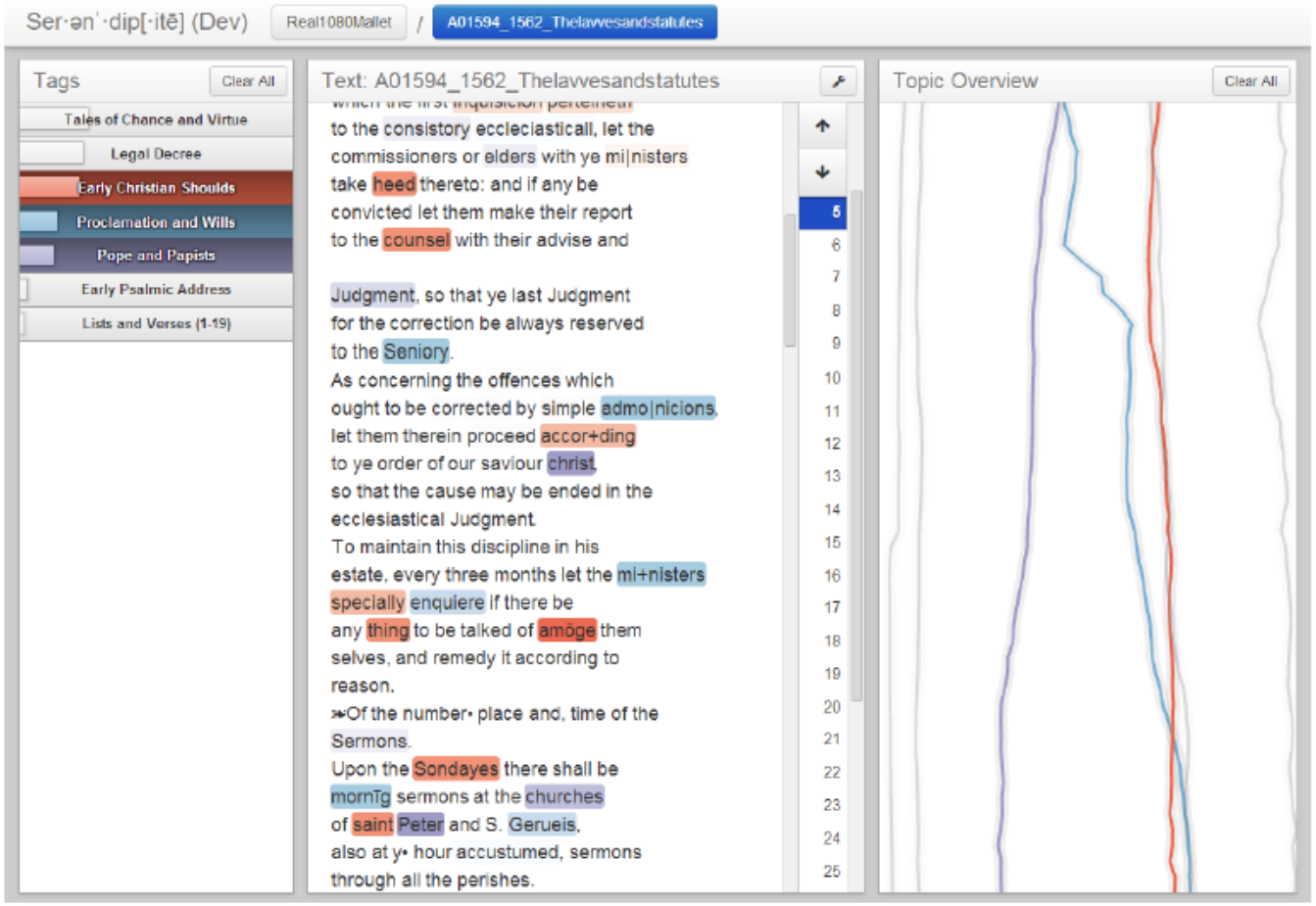
Design: use text as scaffold

Scalability Challenge:

Long Documents

Strategy: summarize

Design: overview + detail



Some Things I cannot give you (yet)

What are the strategies for implicit comparisons?

This is how to think about the problems

How do you choose the solutions?

What are the connections cognitive issues and decision making?

Does it apply to scalability issues more broadly?

Sarikaya, Alper. *Exploring Visual Summaries*.
Ph.D. Thesis. University of Wisconsin – Madison.
August 2017

What is the comparison?

Comparative Elements
Targets
Actions

Why is it hard?

Comparative Challenges
Number of Targets
Large or Complex Targets
Complex Relationships

How to address the challenges?

Scalability Strategies
Scan Sequentially
Select Subset
Summarize Somehow

Which visual design to use?

Comparative Designs
Juxtapose
Superpose
Explicit Encoding

Summary

4 questions to think about comparison

1. What comparative elements?
2. What scalability challenges?
3. What scalability strategy?
4. What visual design?

I find this a helpful process in design

Thank you!

To you for listening

To my students who helped with the ideas and built the systems

To Chuck and Steve, co-PIs on the project

To everyone who listened and gave feedback

To the photographers of the CC0 licensed art

This work was supported in part by NSF Award 1162037. Domain applications were supported by NIH and the Andrew Mellon Foundation

Considerations for Visualizing Comparison

Michael Gleicher gleicher@cs.wisc.edu

Department of Computer Sciences

University of Wisconsin - Madison