INTRODUCTION P5 & DATA ANALYSIS CHALLENGE

PETRA ISENBERG

INFOVIS

DATA ANALYSIS

Challenge

BIBLIOMETRICS

Study of measuring and analysing science, technology and innovation

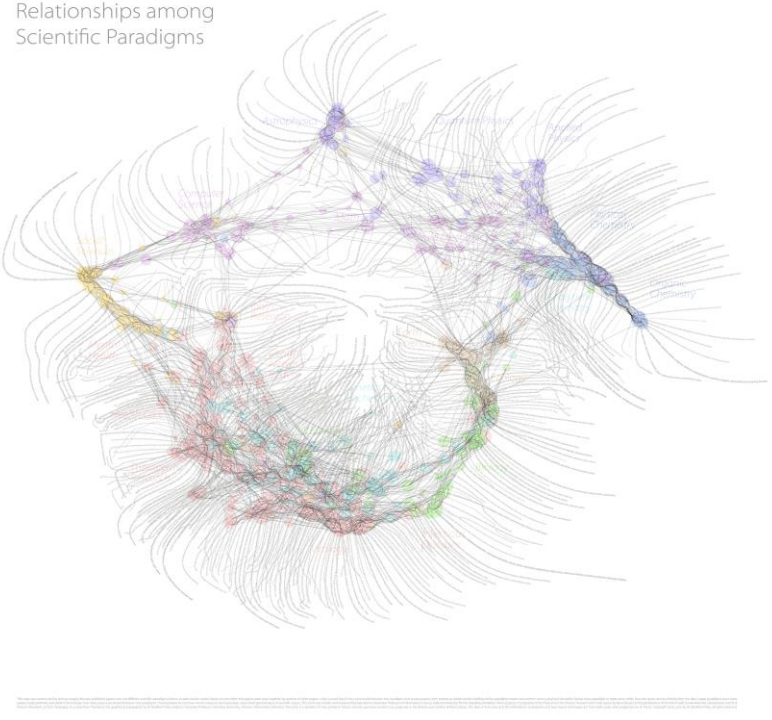
BIBLIOMETRICS

the application of mathematical and statistical methods to books and other media of communication (Pritchard, 1969)

Scientometrics: the science of measuring and analyzing science

WHY?

to understand science



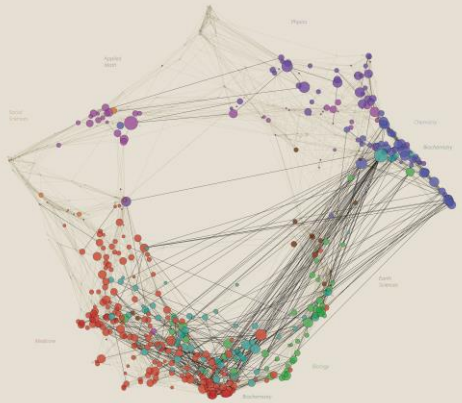
http://wbpaley.com/brad/mapOfScience/

The Scientific Paradigms that Support Patent Generation

This network diagram directs attention to the areas of science that support patents. Each circular node represents scientific publications on a topic (more accurately called a paradigm). There are 775 nodes in all, distributed in a ring-like structure; the largest colored ones are the ones that represent scientific paradigms that generate many patents. General groupings of paradigms include those that concern physics (at approximately 1:00), Chemistry (2:00), nodes that form the jutting peninsula of Organic Chemistry (at 3:00), clusters of nodes abutting with Earth Sciences and Biology (4:00), an interesting span of teal Biochemistry (from 4:30 to 6:00), linking Chemistry to Medicine (from 6:00 to 8:00), the Social Sciences (9:00) and Applied Math (including Computer Science, at 10:00).

The size of each colored node represents the number of patents that build from that area of science. Patents emerge from these general areas of science. There is a concentration of network of research in the computer science (around 12:00), a larger band of science in physics in chemistry (between 12:30 and 3:00), and a very large band of science in the medical sciences that yield patents. There are relatively fewer patents emerging from the social sciences (9:00), astrophysics (12:00), and the earth sciences (5:30).

The edges (lines between nodes) represent papers that explore multiple scientific paradigms; the black edges are the ones where patents are created. Chemistry and medicine seem to have many edges between them. This represents the large number of pharmaceutical patents that must do interdisciplinary work to bridge these disparate scientific paradigms. There are a particularly large number of links between areas of chemistry dealing with analytical methods (at 2:30) and the treatment of infectious diseases (9:00). There are also many patents that build from disparate paradigms within medicine (between 6:00 and 8:00).



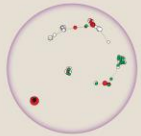
Drilling Down for Additional Insights

Paradigm 365

A paradigm in computer science has been pulled out of the diagram. Additional edges, showing relationships between this paradigm and other areas of science, illustrate that this paradigm links to the social sciences, chemistry, biology and medical science.

Research Communities within Paradigm 365

There are 85 unique research communities within this paradigm that are further grouped into four research areas. The size of each circle represents the amount of research in the area. The color represents performance: white for average, red for high-vitality science, green for patent-producing communities, black for communities that both get patents and practice high-vitality science.



Author Communities within Paradigm 365

The size of each circle represents the amount of research from each the 152 author clusters in this paradigm. The color represents performance, as mentioned at left. Author communities are much more dispersed (only 29% of the authors could be clustered). Authors that are members of clusters are much more likely to be associated with high-performing science.

Themes within Paradigm 365

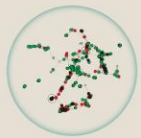
Another way to gain insights into this paradigm is to cluster the keywords in the 3150 scientific publications. The 127 keyword clusters shown below are more interconnected than the other two diagrams. The largest and most interconnected theme (the large black circle near the middle) deals with the use of network diagrams to understand the development of the World Wide Web.

Paradigm 735

One of the largest and most important paradigms in Chemistry has been pulled out of the diagram in order to show its impact in medicine. There are 4982 publications in this paradigm. The strongest edges connecting it and other paradigms are shown to illustrate the wide variety of scientific disciplines with which it shows publications.

Research Communities within Paradigm 735

There are 410 unique research communities within paradigm 735. The relationships between these clusters suggest a relatively complex but highly structured set of research topics. The dispersion of patent-supporting science (green and black) suggest areas to focus on and areas where the science is not as likely to generate patents.



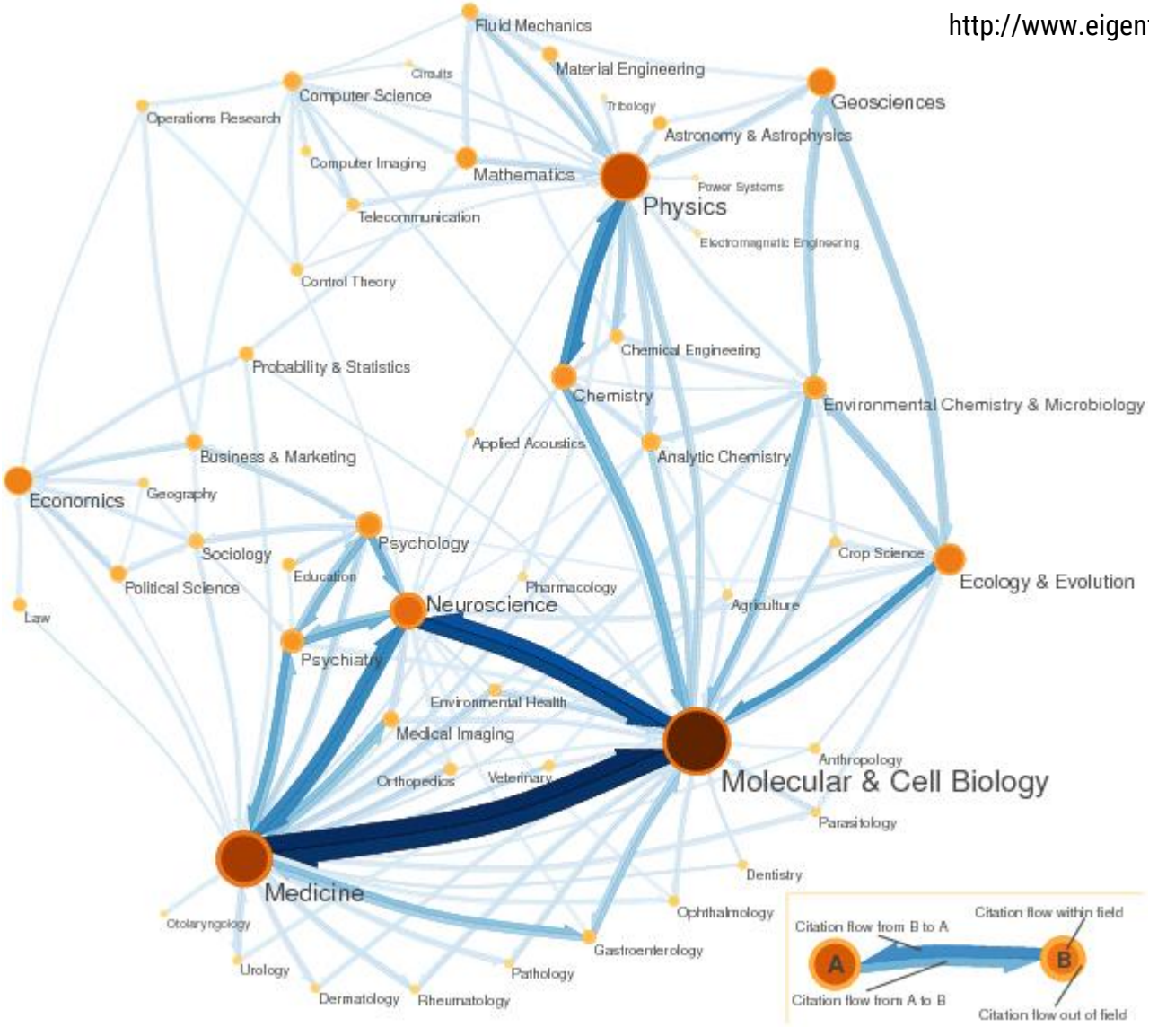
Author Communities within Paradigm 735

The size of each circle represents the amount of research from the 762 author clusters in this paradigm. Author communities are more dispersed that the paradigm shown above (only 28% of the authors are members of an author community). However, authors that are members of these clusters are much more likely to be associated with high-performing science.

Themes within Paradigm 735

The 890 clusters of keywords shown below form a very tight star pattern. This suggests a central theme to this research area. The largest and most interconnected theme in the paradigm deals with pancreatitis. Overall, the paradigm deals with many of the new work in chemistry that is being applied to drug development.

http://www.eigenfactor.org/map/maps.php



WHY?

- to understand science
- to manage science / research
  - ranking of scholarly output of researchers / institutions
  - identifying the centers of excellence

WHY IMPORTANT?

- Globalization of research
- Availability of large databases
- Increased research output → need for awareness
- Quickly evolving research fields

HOW WILL WE ANALYZE SCIENCE?

- through the study of scientific publications
- in the domains of Visual Analytics and Visualization
- by building our own tools

SCIENTIFIC PUBLICATIONS

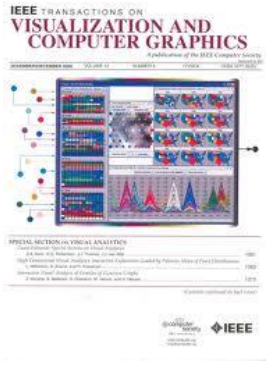
Why are they there?

1. Sharing scientific results/methods/processes
2. To show research performance
3. To allow validation of findings
4. To gain prestige and recognition

PUBLICATION VENUES

Conferences vs. Journals

- journals typical publication venues in most sciences
- in computer science (some) conference publications are highly regarded (with acceptance rates <25%)



RESEARCH QUESTIONS

- Simple & boring
  - Numbers of papers at IEEE VIS 2015
- Boring
  - Numbers of papers by P. Isenberg in 2015
- Interesting (unfortunately not simple)
  - In the domain of visual analytics growing or shrinking?
  - Are visual analytics and visualization the same community?
  - Are research interests of specific researchers changing?
  - What are new research trends in visual analytics?
  - To which university should I go to do a PhD in visual analytics?
  - Who are good reviewers for a certain topic?
  - Who should be in the program committee of VAST / VIS 2020?
  - How does a change in affiliation impact a researcher's interests?
  - I there a relation between affiliation and citations?

Exploring the Placement and Design of Word-Scale Visualizations

Pascal Goffin, Wesley Willett, Jean-Daniel Fekete *Senior Member, IEEE* and Petra Isenberg

Abstract—We present an exploration and a design space that characterize the usage and placement of word-scale visualizations within text documents. Word-scale visualizations are a more general version of sparklines—small, word-sized data graphics that allow meta-information to be visually presented in-line with document text. In accordance with Edward Tufte's definition, sparklines are traditionally placed directly before or after words in the text. We describe alternative placements that permit a wider range of word-scale graphics and more flexible integration with text layouts. These alternative placements include positioning visualizations between lines, within additional vertical and horizontal space in the document, and as interactive overlays on top of the text. Each strategy changes the dimensions of the space available to display the visualizations, as well as the degree to which the text must be adjusted or reflowed to accommodate them. We provide an illustrated design space of placement options for word-scale visualizations and identify six important variables that control the placement of the graphics and the level of disruption of the source text. We also contribute a quantitative analysis that highlights the effect of different placements on readability and text disruption. Finally, we use this analysis to propose guidelines to support the design and placement of word-scale visualizations.

Index Terms—Information visualization, text visualization, sparklines, glyphs, design space, word-scale visualizations

1 Introduction

Small high-resolution data graphics, included alongside words or word sequences in text documents, can often communicate information that could not be succinctly conveyed by the text itself. Examples include small stock charts embedded next to the name of a company, game statistics next to the name of a soccer team, or weather trends next to the name of a city. Traditionally, most of these *"word-scale visualizations"* have consisted of small line charts and bar charts and been placed in-line with text. Edward Tufte terms these word-scale visualizations "sparklines" [30], and provides some guidelines for their visual design. However, Tufte provides little guidance for *placing* word-scale visualizations with respect to text, suggesting only that they be placed in a "relevant context"—usually just after the word that they complement. However, the space of design and placement options for word-scale visualizations is actually quite large, and the consequences of placement decisions, in particular, are not well-understood.

In this paper, we provide design considerations for placing word-scale visualizations associated with words or word sequences (what we refer to as "entities") in a document. Our work is motivated by a close collaboration on digital note-taking with historians in the digital humanities. When visiting an archive, the historians we work with regularly take detailed notes on their findings. In these notes, they specifically tag entities such as the people, locations, or dates that occur in their document sources. The goal of tagging these entities is to help historians build an understanding of how entities relate to one another, where else the same entities appear in their notes, and what kinds of metadata are associated with them. Embedding this information using word-scale visualizations is a promising approach, because these small visualizations can add additional information in-context without distracting attention from the primary reading task.

In prior work, sparklines have typically been placed before or after the word they are related to. However, this is often not possible for the kinds of notes taken by our historians—e. g. when adding information to scanned documents and other immutable texts. Placing word-scale visualizations in-line with text may also be undesirable in other situations, as it requires reflowing the text and restricts the visualization's maximum height to that of the font—making visualizations hard to read when small font sizes were chosen. In-line visualizations can also disrupt sentences, making the text more difficult to read.

To better understand the options available for integrating word-scale visualizations in text documents, we outline a design space of possible placements relative to the text. In doing so, we relax some aspects of Tufte's original sparkline definition, imposing less restrictive size requirements and allowing the small visualizations to extend beyond strictly "word sized." Also, while Tufte did not restrict sparklines to specific visual encodings, the term "sparkline" does inherently suggest a "line-based" data encoding such as a line chart. In contrast, we specifically allow a variety of encodings, including geographical maps, heat maps, pie charts, and more complex visualizations and, thus, chose the term *word-scale visualizations*. We also formalize the notion of an *entity*—a concrete piece of text with associated metadata that can be encoded in a word-scale visualization. This explicit connection between an entity and a word-scale visualization directly affects the options for placing the visualization, and allows us to formally characterize the spatial relationship between text and graphic.

We begin our discussion by reviewing related work on small-scale and text visualizations. Then, in Section 3 we introduce the design space, its focus, and dimensions. Section 4 details several placement options and discusses trade-offs between word-scale visualization placement options. In Section 5 we discuss three examples that demonstrate the importance of the association between word-scale visualization and entity for the purpose of layout and interaction. Finally, in Section 6 we provide an in-depth analysis that examines how various placement options affect word-scale visualization placement in real documents. Based on this analysis, we provide recommendations that can help designers choose the right word-scale visualization given their own constraints.

2 Related Work

Our work relates closely to four research areas: (a) the use of sparklines and the design of word-scale visualizations (b) the integration of meta-data within text documents, (c) research on labeling in visualization, and (d) the readability of texts and visualizations.

2.1 Sparklines and Small-Scale Visualizations

According to Tufte [30] sparklines are "small, intense, simple, word-sized graphics with typographic resolution" that can be included anywhere a word or number can be—e. g. in a sentence, table, headline, map, spreadsheet or graphic. Tufte presents several examples of these embeddings. One example shows sparklines embedded in-line with text in order to provide metadata for a single word, for example glucose measurements next to the word glucose. In another, sparklines

- *Pascal Goffin is with Inria. E-mail: pascal.goffin@inria.fr.*
- *Wesley Willett is with Inria. E-mail: wesley.willett@inria.fr.*
- *Jean-Daniel Fekete is with Inria. E-mail: jean-daniel.fekete@inria.fr.*
- *Petra Isenberg is with Inria. E-mail: petra.isenberg@inria.fr.*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

References

[1] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. Rule-based visual mappings–with a case study on poetry visualization. *Computer Graphics Forum*, 32(3):381–390, 2013.

[2] E. Bertini, M. Rigamonti, and D. Lalanne. Extended excentric labeling. *Computer Graphics Forum*, 28(3):927–934, 2009.

[3] S. Bird, E. Klein, and E. Loper. *Natural Language Processing with Python*. O'Reilly Media Inc., 2009.

[4] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics 2013–State of the Art Reports*, pages 39–63. The Eurographics Association, 2012.

[5] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[6] U. Brandes, B. Nick, B. Rockstroh, and A. Steffen. Gestaltlines. *Computer Graphics Forum*, 32(3):171–180, 2013.

[7] B.-W. Chang, J. D. Mackinlay, P. T. Zellweger, and T. Igarashi. A negotiation architecture for fluid documents. In *Proceedings of the Conference on User Interface Software and Technology (UIST)*, pages 123–132. ACM, 1998.

[8] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988.

[9] M. C. Dyson. How physical text layout affects reading from screen. *Behaviour & Information Technology*, 23(6):377–393, 2004.

[10] J.-D. Fekete and C. Plaisant. Excentric labeling: Dynamic neighborhood labeling for data visualization. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 512–519. ACM, 1999.

[11] S. Few. Time on the horizon, Last read: March 2014. `http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf`.

[12] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 3237–3246. ACM, 2013.

[13] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Sparklificator, Last read: June 2014. `http://inria.github.io/sparklificator/`.

[14] B. Greenhill, M. Ward, and A. Sacks. The separation plot: A new visual method for evaluating the fit of binary models. *American Journal of Political Science*, 55(4):991–1002, 2011.

[15] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):701–708, 2006.

[16] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 1303–1312. ACM, 2009.

[17] M. R. Jakobsen and K. Hornbæk. Transient visualizations. In *Proceedings of the Conference on Computer-Human Interaction (OzCHI)*, pages 69–76. ACM, 2007.

[18] B. Lee, N. H. Riche, A. K. Karlson, and S. Carpendale. SparkClouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1182–1189, 2010.

[19] J. Pearson, G. Buchanan, and H. Thimbleby. Improving annotations in digital documents. In *Research and Advanced Technology for Digital Libraries*, pages 429–432. Springer, 2009.

[20] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2506–2515, 2013.

[21] P. Pirolli and S. Card. Information foraging. *Psychological Review*, 106(4):643–675, 1999.

[22] H. Reijner. The development of the horizon graph. In *Proceeding of Workshop From Theory to Practice: Design, Vision and Visualization Extended Abstracts of IEEE VisWeek*. Citeseer, 2008.

[23] T. Ropinski, S. Oeltze, and B. Preim. Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers & Graphics*, 35(2):392–401, 2011.

[24] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda. Two-tone pseudo coloring: Compact visualization for one-dimensional data. In *Proceedings of the Conference on Information Visualization (InfoVis)*, pages 173–180. IEEE, 2005.

[25] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[26] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2249–2258, 2011.

[27] M. Stone. In color perception, size matters. *IEEE Computer Graphics and Applications*, 32(2):8–13, 2012.

[28] J. Talbot, J. Gerth, and P. Hanrahan. An empirical model of slope ratio comparisons. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2613–2620, 2012.

[29] E. R. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.

[30] E. R. Tufte. *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.

[31] M. O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4):194–210, 2002.

[32] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.

[33] D. Yoon, N. Chen, and F. Guimbretière. TextTearing: Opening white space for digital ink annotation. In *Proceedings of the Conference on User Interface Software and Technology (UIST)*, pages 107–112. ACM, 2013.

[34] P. T. Zellweger, S. H. Regli, J. D. Mackinlay, and B.-W. Chang. The impact of fluid documents on reading and browsing: An observational study. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 249–256. ACM, 2000.

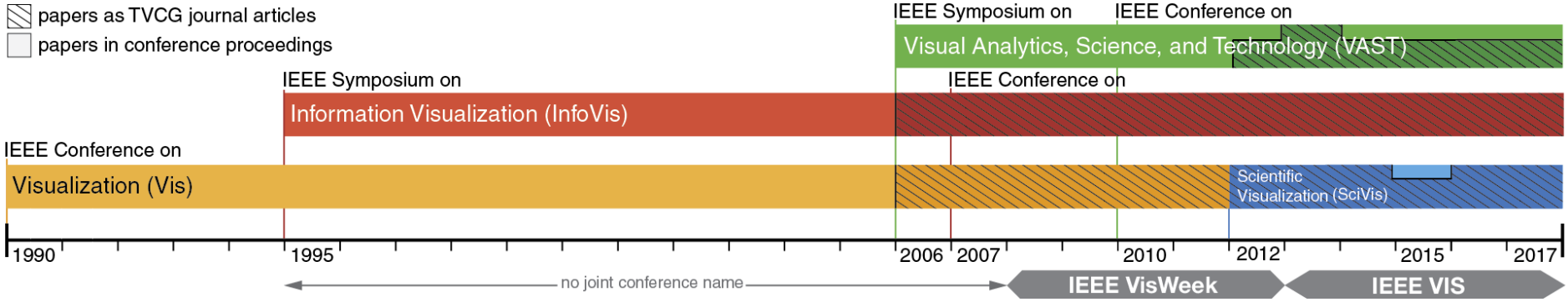
DATA SET

13 columns, >2800 rows

Conference	Year	Paper Title	Paper DOI	Link	First page	Last page	Paper type: C=conference paper, J = journal paper, M=miscellaneous (capstone, keynote, VAST challenge, panel, poster, ...)	Abstract	Author Names	Author Affiliation	References	Author Keywords
Vis	2000	Topology preservin	10.1109/VISUAL.2000.885703	http://dx.doi.org/10.	259	266	C	Multiresolution meth	Thomas Gerstner;Renato Pajarola	Dept. of Appl. Math., Bonn Univ.,	10.1109/VISUAL.1996.568127;10.1	tetrahedral grid ref
Vis	2000	Isosurfacing in high	10.1109/VISUAL.2000.885704	http://dx.doi.org/10.	267	273	C	Visualization algorith	Praveen Bhaniramka;Rephael Wel	Silicon Graphics Comput. Syst., I	10.1109/VISUAL.1992.235222;10.1	
Vis	2000	Semi-regular mesh	10.1109/VISUAL.2000.885705	http://dx.doi.org/10.	275	282	C	We present a novel r	Zoë J. Wood;Peter Schröder;Davic	California Inst. of Technol., Pasad	10.1109/VISUAL.2000.885703	Semi-regular meshes,
Vis	2000	Scanline surfacing:	10.1109/VISUAL.2000.885706	http://dx.doi.org/10.	283	289	C	A standard way to se	David M. Weinstein	Sch. of Comput., Utah Univ, Salt	10.1109/VISUAL.1997.663887;10.1	separating surfaces,
Vis	2000	Navigating high-dim	10.1109/VISUAL.2000.885707	http://dx.doi.org/10.	291	296	C	Throughout the desig	Helen Wright;Ken Brodlie;Tim Dav	Dept. of Comput. Sci., Hull Univ,	10.1109/VISUAL.1999.809921;10.1	Computational steeri
Vis	2000	Visualization of mu	10.1109/VISUAL.2000.885708	http://dx.doi.org/10.	297	302	C	Multi-dimensional en	R. R. Johnson	Dept. of Comput. Sci., Utah Univ, Salt Lake City, UT, USA|c|		Multidimensional Vis
Vis	2000	Real-world relativity	10.1109/VISUAL.2000.885709	http://dx.doi.org/10.	303	310	C	This paper describes	Daniel Weiskopf;Daniel Kobras;Ha	Inst. of Astron. & Astrophys., Tubi	10.1109/VISUAL.1990.146368	image-based renderir
Vis	2000	Visualizing geodes	10.1109/VISUAL.2000.885710	http://dx.doi.org/10.	311	318	C	One of the main rese	Ingrid Hotz;Hans Hagen	Dept. of Comput. Sci., Kaiserslaut	10.1109/VISUAL.1992.235196	geodesics, visualiza
Vis	2000	Geometric compre	10.1109/VISUAL.2000.885711	http://dx.doi.org/10.	319	326	C	The compression of	Olivier Devillers;Pierre-Marie Gan	Inst. Nat. de Recherche en Inf. et	10.1109/VISUAL.1997.663902;10.1	geometry, compressio
Vis	2000	Toward a compellin	10.1109/VISUAL.2000.885712	http://dx.doi.org/10.	327	333	C	In 1998 we introduce	Wei-Chao Chen;Herman Towles;Li	Dept. of Comput. Sci., North Caro	10.1109/VISUAL.1996.568125;10.1	telepresence, tele-i

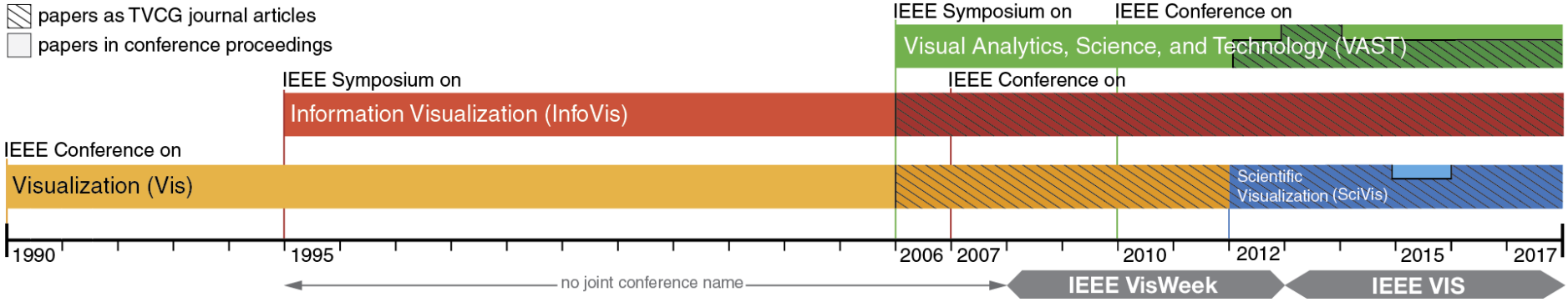
http://www.vispubdata.org/

CONFERENCE



{InfoVis, Vis, SciVis, VAST}

YEAR



{1990 - 2015}

TITLE

Exploring the Placement and Design of Word-Scale Visualizations

Pascal Goffin, Wesley Willett, Jean-Daniel Fekete *Senior Member, IEEE* and Petra Isenberg

Abstract—We present an exploration and a design space that characterize the usage and placement of word-scale visualizations within text documents. Word-scale visualizations are a more general version of sparklines—small, word-sized data graphics that allow meta-information to be visually presented in-line with document text. In accordance with Edward Tufte's definition, sparklines are traditionally placed directly before or after words in the text. We describe alternative placements that permit a wider range of word-scale graphics and more flexible integration with text layouts. These alternative placements include positioning visualizations between lines, within additional vertical and horizontal space in the document, and as interactive overlays on top of the text. Each strategy changes the dimensions of the space available to display the visualizations, as well as the degree to which the text must be adjusted or reflowed to accommodate them. We provide an illustrated design space of placement options for word-scale visualizations and identify six important variables that control the placement of the graphics and the level of disruption of the source text. We also contribute a quantitative analysis that highlights the effect of different placements on readability and text disruption. Finally, we use this analysis to propose guidelines to support the design and placement of word-scale visualizations.

Index Terms—Information visualization, text visualization, sparklines, glyphs, design space, word-scale visualizations

1 Introduction

Small high-resolution data graphics, included alongside words or word sequences in text documents, can often communicate information that could not be succinctly conveyed by the text itself. Examples include small stock charts embedded next to the name of a company, game statistics next to the name of a soccer team, or weather trends next to

alization's maximum height to that of the font—making visualizations hard to read when small font sizes were chosen. In-line visualizations can also disrupt sentences, making the text more difficult to read.

To better understand the options available for integrating word-scale visualizations in text documents, we outline a design space of possible placements relative to the text. In doing so, we relax some aspects

PAPER DOI

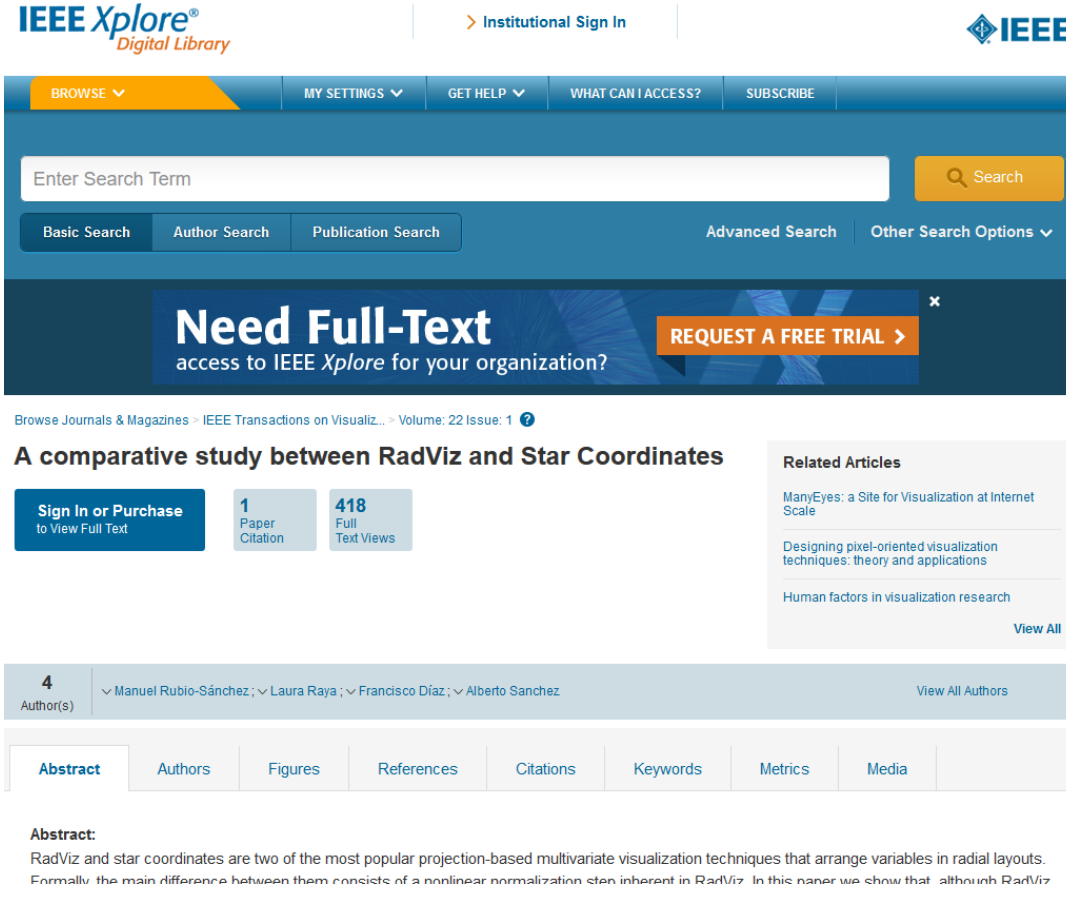
- A persistent identifier used to uniquely identify objects.
- Particularly used for electronic documents such as journal articles.

10.1109/TVCG.2015.2467471

= your unique key to each paper in the database

LINK

- A link to the digital library of the publisher of the paper
- The paper can be read/bought here



FIRST PAGE – LAST PAGE

- can be used to deduce page count
- likely not clean data

PAPER TYPE

- J = Journal
  - the most prestigious type
  - a full scientific paper (8-10 pages usually)
- C = Conference
  - a full scientific paper (8-10 pages usually)
- M = Miscellaneous
  - a poster (2 pages)
  - a talk abstract (1-2 pages)
  - NOT a full paper

ABSTRACT

a short summary of the paper content

Exploring the Placement and Design of Word-Scale Visualizations

Pascal Goffin, Wesley Willett, Jean-Daniel Fekete *Senior Member, IEEE* and Petra Isenberg

Abstract—We present an exploration and a design space that characterize the usage and placement of word-scale visualizations within text documents. Word-scale visualizations are a more general version of sparklines—small, word-sized data graphics that allow meta-information to be visually presented in-line with document text. In accordance with Edward Tufte's definition, sparklines are traditionally placed directly before or after words in the text. We describe alternative placements that permit a wider range of word-scale graphics and more flexible integration with text layouts. These alternative placements include positioning visualizations between lines, within additional vertical and horizontal space in the document, and as interactive overlays on top of the text. Each strategy changes the dimensions of the space available to display the visualizations, as well as the degree to which the text must be adjusted or reflowed to accommodate them. We provide an illustrated design space of placement options for word-scale visualizations and identify six important variables that control the placement of the graphics and the level of disruption of the source text. We also contribute a quantitative analysis that highlights the effect of different placements on readability and text disruption. Finally, we use this analysis to propose guidelines to support the design and placement of word-scale visualizations.

Index Terms—Information visualization, text visualization, sparklines, glyphs, design space, word-scale visualizations

1 INTRODUCTION

AUTHORS

- Firstname Lastname
- Separated by ;
- First author often the project lead
- Last author often the advisor

Exploring the Placement and Design of Word-Scale Visualizations

Pascal Goffin, Wesley Willett, Jean-Daniel Fekete *Senior Member, IEEE* and Petra Isenberg

Abstract—We present an exploration and a design space that characterize the usage and placement of word-scale visualizations within text documents. Word-scale visualizations are a more general version of sparklines—small, word-sized data graphics that allow meta-information to be visually presented in-line with document text. In accordance with Edward Tufte's definition, sparklines are traditionally placed directly before or after words in the text. We describe alternative placements that permit a wider range of

AUTHOR KEYWORDS

- added by the authors to a paper
- think of as tags describing the content

Exploring the Placement and Design of Word-Scale Visualizations

Pascal Goffin, Wesley Willett, Jean-Daniel Fekete *Senior Member, IEEE* and Petra Isenberg

Abstract—We present an exploration and a design space that characterize the usage and placement of word-scale visualizations within text documents. Word-scale visualizations are a more general version of sparklines—small, word-sized data graphics that allow meta-information to be visually presented in-line with document text. In accordance with Edward Tufte's definition, sparklines are traditionally placed directly before or after words in the text. We describe alternative placements that permit a wider range of word-scale graphics and more flexible integration with text layouts. These alternative placements include positioning visualizations between lines, within additional vertical and horizontal space in the document, and as interactive overlays on top of the text. Each strategy changes the dimensions of the space available to display the visualizations, as well as the degree to which the text must be adjusted or reflowed to accommodate them. We provide an illustrated design space of placement options for word-scale visualizations and identify six important variables that control the placement of the graphics and the level of disruption of the source text. We also contribute a quantitative analysis that highlights the effect of different placements on readability and text disruption. Finally, we use this analysis to propose guidelines to support the design and placement of word-scale visualizations.

Index Terms—Information visualization, text visualization, sparklines, glyphs, design space, word-scale visualizations

REFERENCE

- which other VIS paper is cited from this particular paper
- based on DOI and separated by ;

10.1109/VAST.2010.5652433;10.1109/INFVIS.1998.729559;10.1109/VISUAL.1997.663916;10.1109/TVCG.2013.182;10.1109/TVCG.2014.2346258;10.1109/TVCG.2008.173

REFERENCES

[1] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. Rule-based visual mappings–with a case study on poetry visualization. *Computer Graphics Forum*, 32(3):381–390, 2013.

[2] E. Bertini, M. Rigamonti, and D. Lalanne. Extended excentric labeling. *Computer Graphics Forum*, 28(3):927–934, 2009.

[3] S. Bird, E. Klein, and E. Loper. *Natural Language Processing with Python*. O'Reilly Media Inc., 2009.

[4] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics 2013–State of the Art Reports*, pages 39–63. The Eurographics Association, 2012.

[5] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[6] U. Brandes, B. Nick, B. Rockstroh, and A. Steffen. Gestaltlines. *Computer Graphics Forum*, 32(3):171–180, 2013.

[7] B.-W. Chang, J. D. Mackinlay, P. T. Zellweger, and T. Igarashi. A negotiation architecture for fluid documents. In *Proceedings of the Conference on User Interface Software and Technology (UIST)*, pages 123–132. ACM, 1998.

[8] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988.

[9] M. C. Dyson. How physical text layout affects reading from screen. *Behaviour & Information Technology*, 23(6):377–393, 2004.

[10] J.-D. Fekete and C. Plaisant. Excentric labeling: Dynamic neighborhood labeling for data visualization. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 512–519. ACM, 1999.

[11] S. Few. Time on the horizon, Last read: March 2014. http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf.

[12] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 3237–3246. ACM, 2013.

[13] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Sparklificator, Last read: June 2014. http://inria.github.io/sparklificator/.

[14] B. Greenhill, M. Ward, and A. Sacks. The separation plot: A new visual method for evaluating the fit of binary models. *American Journal of Political Science*, 55(4):991–1002, 2011.

[15] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):701–708, 2006.

[16] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 1303–1312. ACM, 2009.

[17] M. R. Jakobsen and K. Hornbæk. Transient visualizations. In *Proceedings of the Conference on Computer-Human Interaction (OzCHI)*, pages 69–76. ACM, 2007.

[18] B. Lee, N. H. Riche, A. K. Karlson, and S. Carpendale. SparkClouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1182–1189, 2010.

[19] J. Pearson, G. Buchanan, and H. Thimbleby. Improving annotations in digital documents. In *Research and Advanced Technology for Digital Libraries*, pages 429–432. Springer, 2009.

[20] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2506–2515, 2013.

[21] P. Pirolli and S. Card. Information foraging. *Psychological Review*, 106(4):643–675, 1999.

[22] H. Reijner. The development of the horizon graph. In *Proceeding of Workshop From Theory to Practice: Design, Vision and Visualization Extended Abstracts of IEEE VisWeek*. Citeseer, 2008.

[23] T. Ropinski, S. Oeltze, and B. Preim. Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers & Graphics*, 35(2):392–401, 2011.

[24] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda. Two-tone pseudo coloring: Compact visualization for one-dimensional data. In *Proceedings of the Conference on Information Visualization (InfoVis)*, pages 173–180. IEEE, 2005.

[25] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[26] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2249–2258, 2011.

[27] M. Stone. In color perception, size matters. *IEEE Computer Graphics and Applications*, 32(2):8–13, 2012.

[28] J. Talbot, J. Gerth, and P. Hanrahan. An empirical model of slope ratio comparisons. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2613–2620, 2012.

[29] E. R. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.

[30] E. R. Tufte. *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.

[31] M. O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4):194–210, 2002.

[32] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.

[33] D. Yoon, N. Chen, and F. Guimbretière. TextTearing: Opening white space for digital ink annotation. In *Proceedings of the Conference on User Interface Software and Technology (UIST)*, pages 107–112. ACM, 2013.

[34] P. T. Zellweger, S. H. Regli, J. D. Mackinlay, and B.-W. Chang. The impact of fluid documents on reading and browsing: An observational study. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 249–256. ACM, 2000.

DATASET 2

- Paper DOI
- A gender per author (male, female, unisex, unknown)
- Position in author list per paper
- Region in author list (F=first, M=middle, L=last author)
- Certainty of gender identification

Paper.DOI	Author.Names	gender	certainty	Author.Position.Per	PositionRegion
10.0000/00000001	Marc Abrams	male	0,8	2	L
10.0000/00000001	Randy L. Ribler	male	0,8	1	F
10.0000/00000002	Donna J. Cox	female	1	1	F
10.1109/INFVIS....	Hua Su	unisex	1	4	L
10.1109/INFVIS....	Huw Dawkes	male	0,8	3	M
10.1109/INFVIS....	Lisa Tweedie	female	0,875	2	M
10.1109/INFVIS....	Robert Spence	male	0,857142857142...	1	F
10.1109/INFVIS....	Gene Golovchins...	male	0,875	3	L
10.1109/INFVIS....	Klaus Reichenbe...	male	0,8	1	F
10.1109/INFVIS....	Thomas Kamps	male	0,833333333333...	2	M
10.1109/INFVIS....	W. Wright	male	1	1	F
10.1109/INFVIS....	Peter Lüders	male	1	1	F
10.1109/INFVIS....	Rolf Ernst	male	0,833333333333...	2	L
10.1109/INFVIS....	Joe Mattis	male	0,857142857142...	3	M
10.1109/INFVIS....	John Kolojejchick	male	0,857142857142...	4	L
10.1109/INFVIS....	Mei C. Chuah	female	0,8	1	F
10.1109/INFVIS....	Steven F. Roth	male	1	2	M
10.1109/INFVIS....	Dean F. Jerding	male	1	1	F
10.1109/INFVIS....	John T. Stasko	male	0,857142857142...	2	L
10.1109/INFVIS....	Anne Schur	female	1	6	M
10.1109/INFVIS....	D. Lantrip	male	1	4	M

RESEARCH QUESTIONS

What can we do with this data?

GENDER DIVERSITY IN VIS

The influence of gender on visualization research is unknown

For the first time we will make available to you a dataset of researchers and their gender

GENDER DIVERSITY IN VIS

Build visualizations that give an insight on the differences of the genders and the influence on the research being conducted

GENDER DIVERSITY IN VIS

As this question is completely unknown so far we want to share what you come up with with the community

→ We will build web-based visualizations

CODING ENVIRONMENT

p5.js

Download * Start * Reference * Libraries * Learn * Community

Hello! p5.js is a JavaScript library that starts with the original goal of Processing, to make coding accessible for artists, designers, educators, and beginners, and reinterprets this for today's web.

Using the original metaphor of a software sketchbook, p5.js has a full set of drawing functionality. However, you're not limited to your drawing canvas, you can think of your whole browser page as your sketch! For this, p5.js has addon libraries that make it easy to interact with other HTML5 objects, including text, input, video, webcam, and sound.

p5.js is a new interpretation, not an emulation or port, and it is in active development. An official editing environment is coming soon, as well as many more features!

---

p5.js was created by Lauren McCarthy and is developed by a community of collaborators, with support from the Processing Foundation and NYU ITP. © Info.

Processing

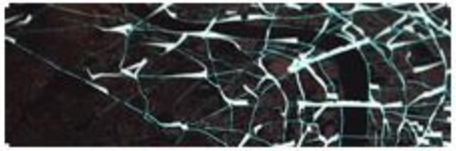
Welcome to Processing 3! Dan explains the new features and changes; the links Dan mentions are on the Vimeo page.

» Download Processing
» Browse Tutorials
» Visit the Reference

Processing is a flexible software sketchbook and a language for learning how to code within the context of the visual arts. Since 2001, Processing has promoted software literacy within the visual arts and

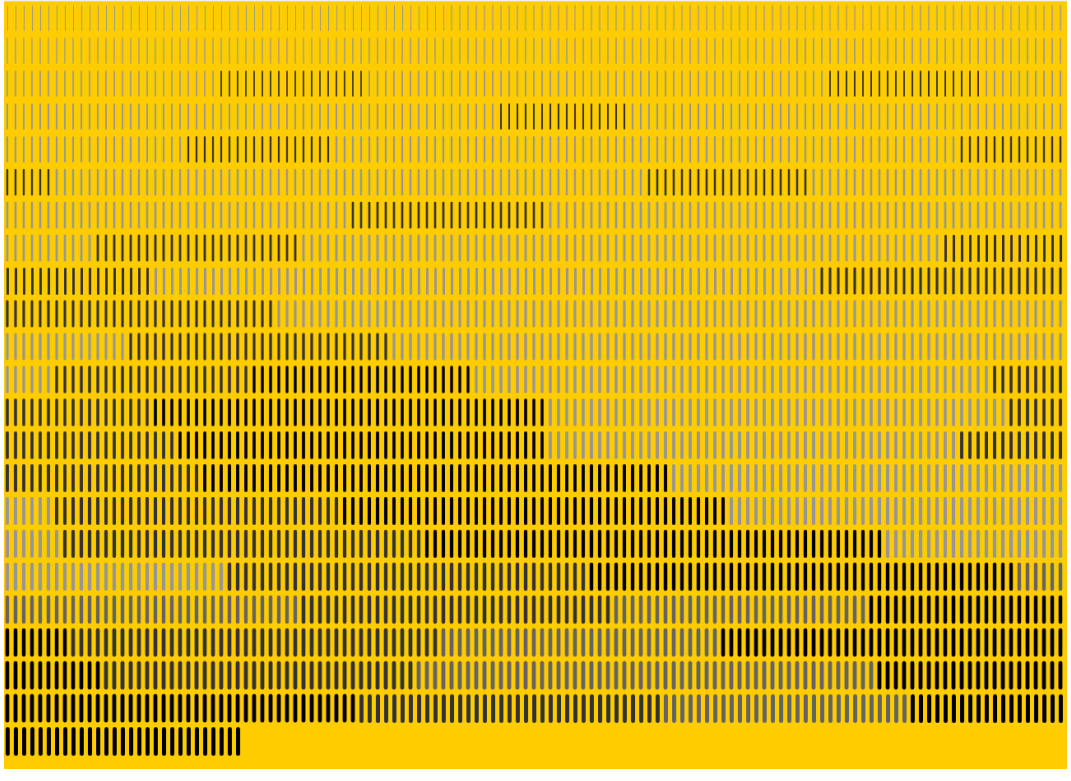
» Exhibition

Fluid Leaves
by Reinoud van Laar



cf.city flows
by Till Nagel and Christopher Pietsch

WHAT WE WILL BE BUILDING TODAY



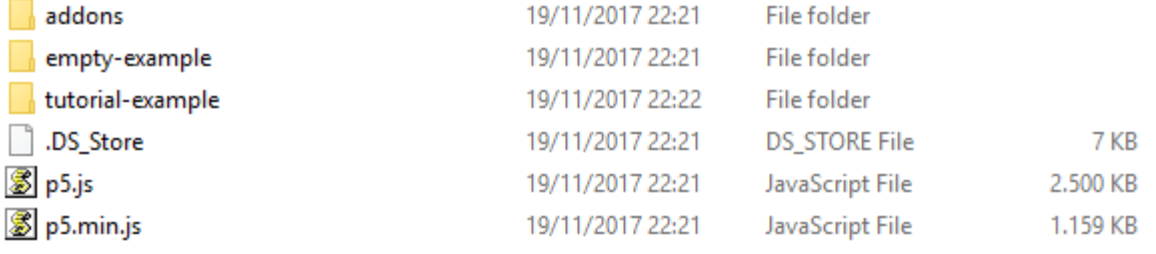
DOWNLOAD

Get your favorite text editor

On windows, e.g. Notepad++

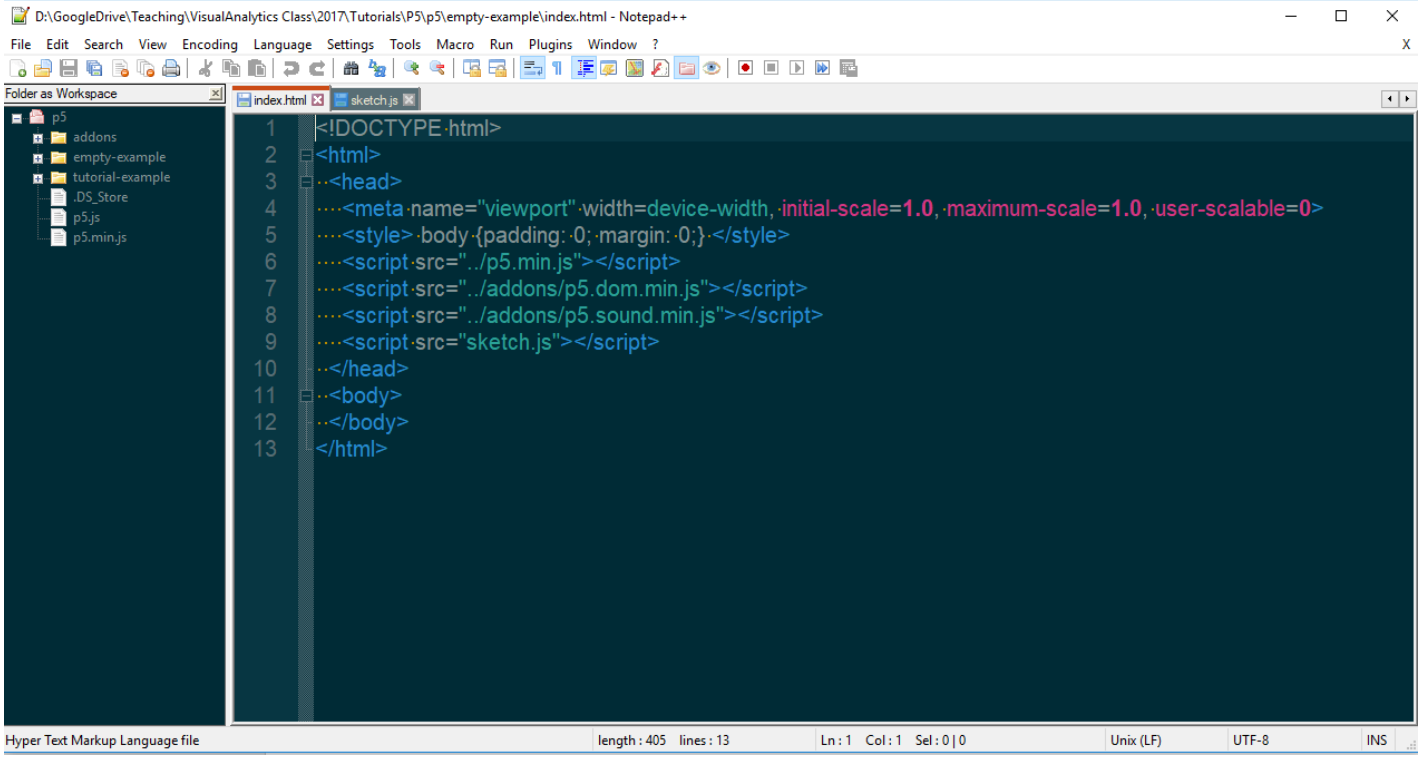
P5 COMPLETE

- Extract into a folder
- Copy the empty example
- Rename the empty example to something useful, e.g. “tutorial-example”

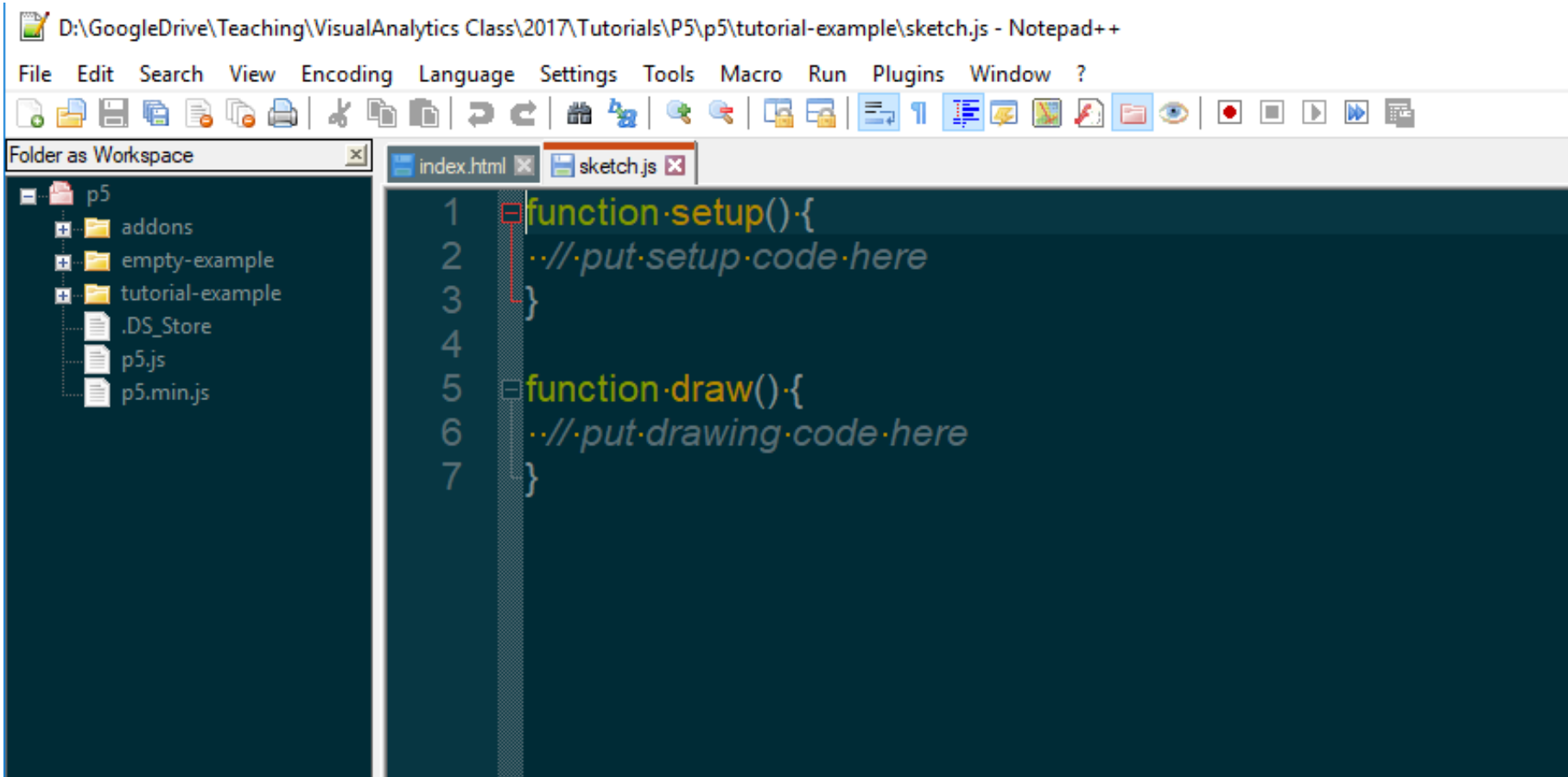


OPTIONAL - NOTEPAD++

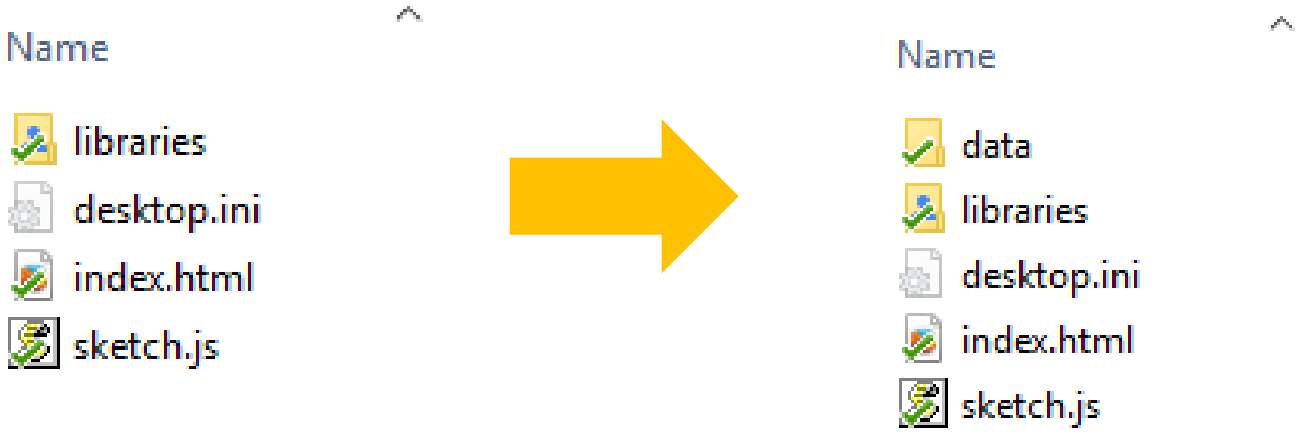
- File -> Open folder as workspace



START


D:\GoogleDrive\Teaching\VisualAnalytics Class\2017\Tutorials\P5\p5\tutorial-example\sketch.js - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
Folder as Workspace
p5
addons
empty-example
tutorial-example
.DS_Store
p5.js
p5.min.js
index.html
sketch.js
1 function setup() {
2 // put setup code here
3 }
4
5 function draw() {
6 // put drawing code here
7 }

DATA & LIBRARIES FOLDERS



COPY DATA FILE

- Into data folder
- Copy p5-min.js into libraries folder

- If you want to use chrome, start a webserver
- E.g. python -m http.server
  (python 3)

CHANGE HTML FILE

```
<!DOCTYPE html>
<html>
  <head>
    <meta name="viewport" width=device-width, initial-scale=1.0, maximum-scale=1.0, user-scalable
    <style> body {padding: 0; margin: 0;} </style>
    <script src="libraries/p5.min.js"></script>
    <script src="sketch.js"></script>
  </head>
  <body>
  </body>
</html>
```

```
var w = 1300;
var h = 1000;

function setup() {

  createCanvas(w,h);
  noLoop();
  background(255,204,0);
}

function draw() {

}
```

Ctrl+Shift+R for reloading a refreshed js

```
function preload()
{
   table = loadTable("data/IEEE VIS papers 1990-2016 - Main dataset.csv", "csv", "header");
}

function setup() {

  createCanvas(w,h);
  noLoop();
  background(255,204,0);

  console.log(table.getRowCount() + " total rows in table.");
  console.log(table.getColumnCount() + " total columns in table.");
```

```
function draw() {

  var spacing = 10;
  var x = 0;
  var y = 5;
  var length = 10;
  var lineheight = 30;

  for (var i = 0; i < table.getRowCount(); i++)
   {

     x = x + spacing;

     if (x > w - spacing) {
        x = x % w + spacing;
        y = y + lineheight + 10;
     }

     line(x , y, x, y + lineheight)
   }

}
```

```
4  var table;
5
6  var yearCol;
7  var conferenceCol;
8  var minYear;
9  var maxYear;
10
11 var minWidth = 1;
12 var maxWidth = 5;
13
14 var fills = [0,50,100,150,200];
15 var conferences = ["VAST","InfoVis","SciVis","Vis"];
```

```
function setup() {

  createCanvas(w,h);
  noLoop();
  background(255,204,0);

  console.log(table.getRowCount() + " total rows in table.");
  console.log(table.getColumnCount() + " total columns in table.");

  yearCol = table.getColumn("Year");
  minYear = min(yearCol);
  maxYear = max(yearCol);

  minWidth = 1;
  maxWidth = 5;

}
```

```
  for (var i = 0; i < table.getRowCount(); i++)
  {
    x = x + spacing;

    if(x > w - spacing)
    {
      x = x % w + spacing;
      y = y + lineheight + 10;
    }

    currentYear = yearCol[i]
    currentWidth = map(currentYear,minYear,maxYear,minWidth,maxWidth);

    strokeWeight(currentWidth);

    line ( x , y, x, y + lineheight);
  }
}
```

```
function setup() {

  createCanvas(w,h);
  noLoop();
  background(255,204,0);

  console.log(table.getRowCount() + " total rows in table.");
  console.log(table.getColumnCount() + " total columns in table.");

  yearCol = table.getColumn("Year");
  minYear = min(yearCol);
  maxYear = max(yearCol);

  minWidth = 1;
  maxWidth = 5;

  conferenceCol = table.getColumn("Conference");

}
```

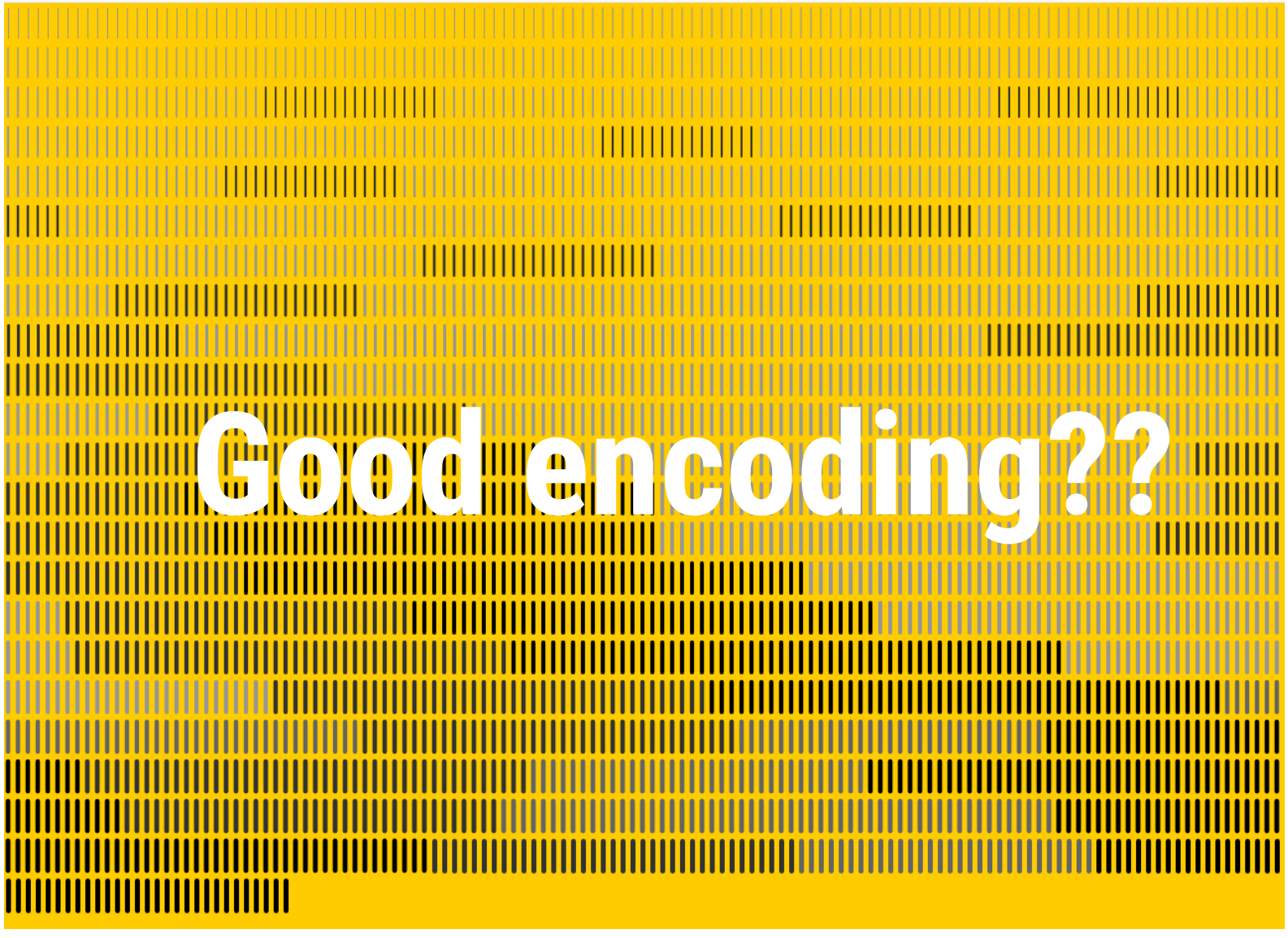
```
    currentYear = yearCol[i]
    currentWidth = map(currentYear,minYear,maxYear,minWidth,maxWidth);

    strokeWeight(currentWidth);

    var conf = conferenceCol[i];
    var index = conferences.indexOf(conf);
    var strokeColor = fills[index];

    stroke(strokeColor);

    line ( x , y, x, y + lineheight);
}
```


Good encoding??