Case Study

125 years of words in Scientific American

A visualization by **Moritz Stefaner**

- German data visualization specialist,
  - Participated in important projects
  - Winner of several contests

- Website: https://truth-and-beauty.net/
- Podcast: https://datastori.es/ Data stories with Enrico Bertini
- Interesting, current project: WHO Data Design Language

Step 1: The project - vision and goals

- Dig deep in Scientific American Archive from 1845 to 2020 to show how language evolved through these years.

  5107 editions (1845 to 1993 predigital, scanned and OCR)

  199,694 pages

  110,292,327 words

- Goal:
  - Commemorate the event
  - Reflect on the evolution of language
  - Entertain

Step 1: The project – Audience and medium

- Audience:
  - Scientific American fans
  - Historians
    - but also
  - Plain readers
- Medium:
  - Paper (special article)
  - Digital (interactive, simpler visualization)

Step 2: First ideas

- Is the dataset worth the visualization?

- Possible questions:
  - Trends in issues / through years

  - Unit of analysis
    - Sentence lengths
    - Terms (two or three words)
    - Interpunctuation styles (exclamations, question marks)
    - Persons and places

Step 2: Exploration of data: limitations

- OCR errors, impossible to work with pairs or three words
  - No terms
  - No places
  - No persons

- Issue level full of small word recognition errors

- Decision: *Yearly level, Individual words*

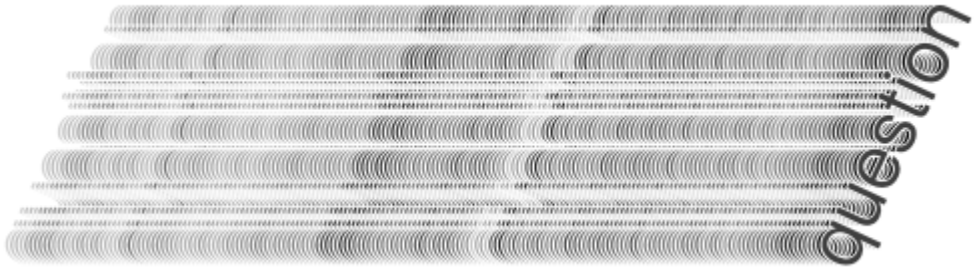
Step 2: Exploration insights

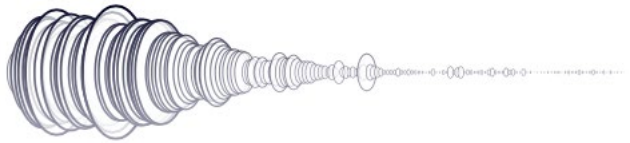
- What or *How*?
  - Few repetitive words (4000), long tail
  - Importance of *frequency*

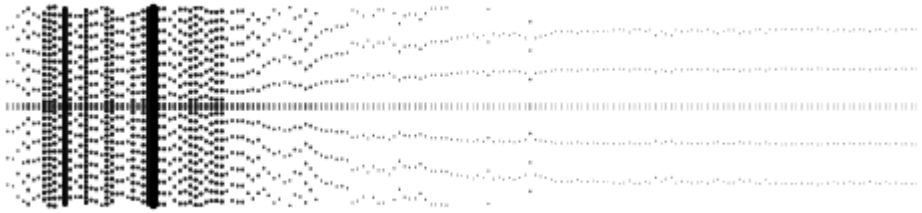
- Absolute or *Relative*?
  - The amount of words has changed a lot over the years

Step 3: Charts and encodings, first sketches

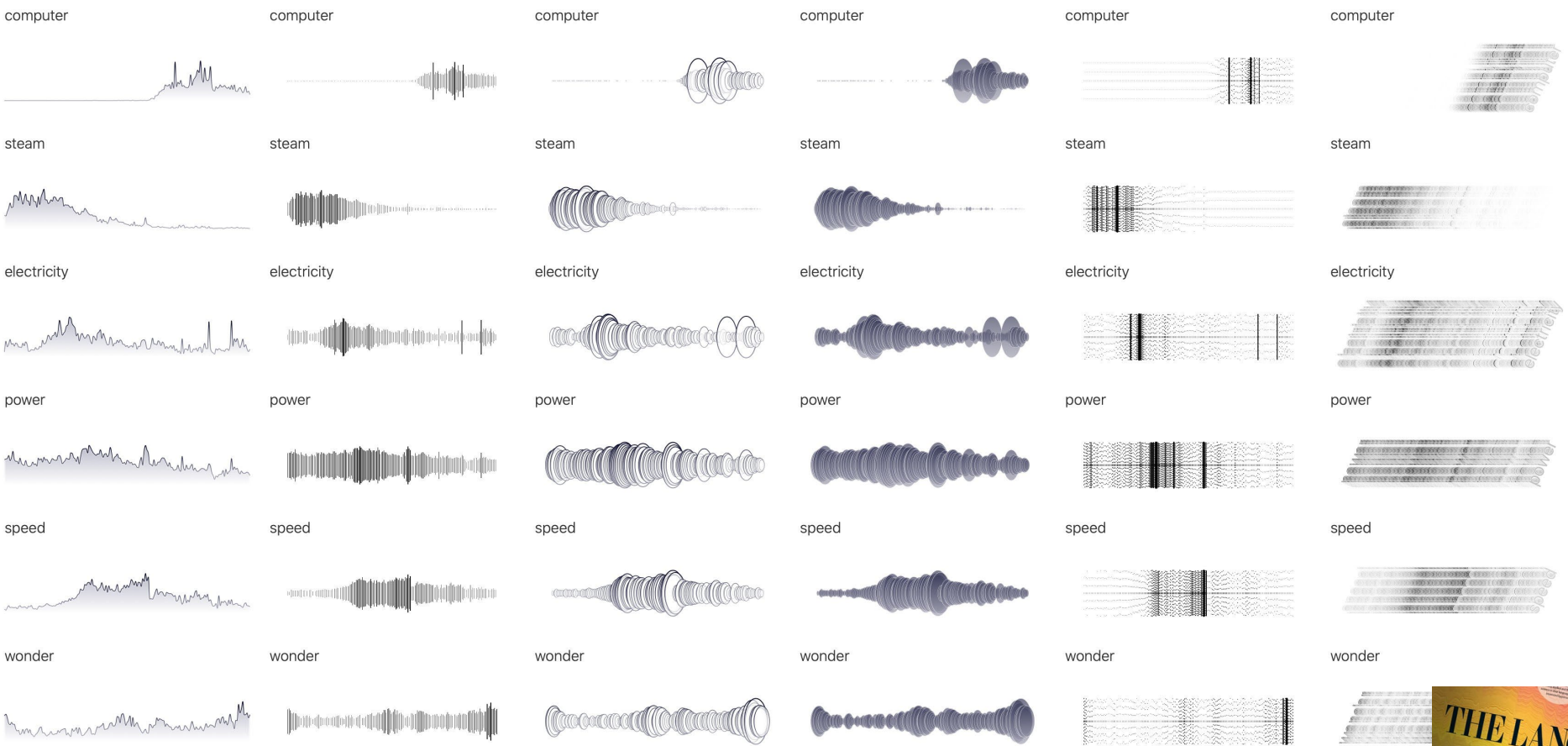
- Trials: Individual words evolution







Step 3: Charts and encodings, more sketches



Step 3: Choosing a chart

- Word clouds
- *Stack area graphs* – For a global view
- *Line charts* – For the evolution of each word
- Animations
- Spatial maps
- Semantic spaces

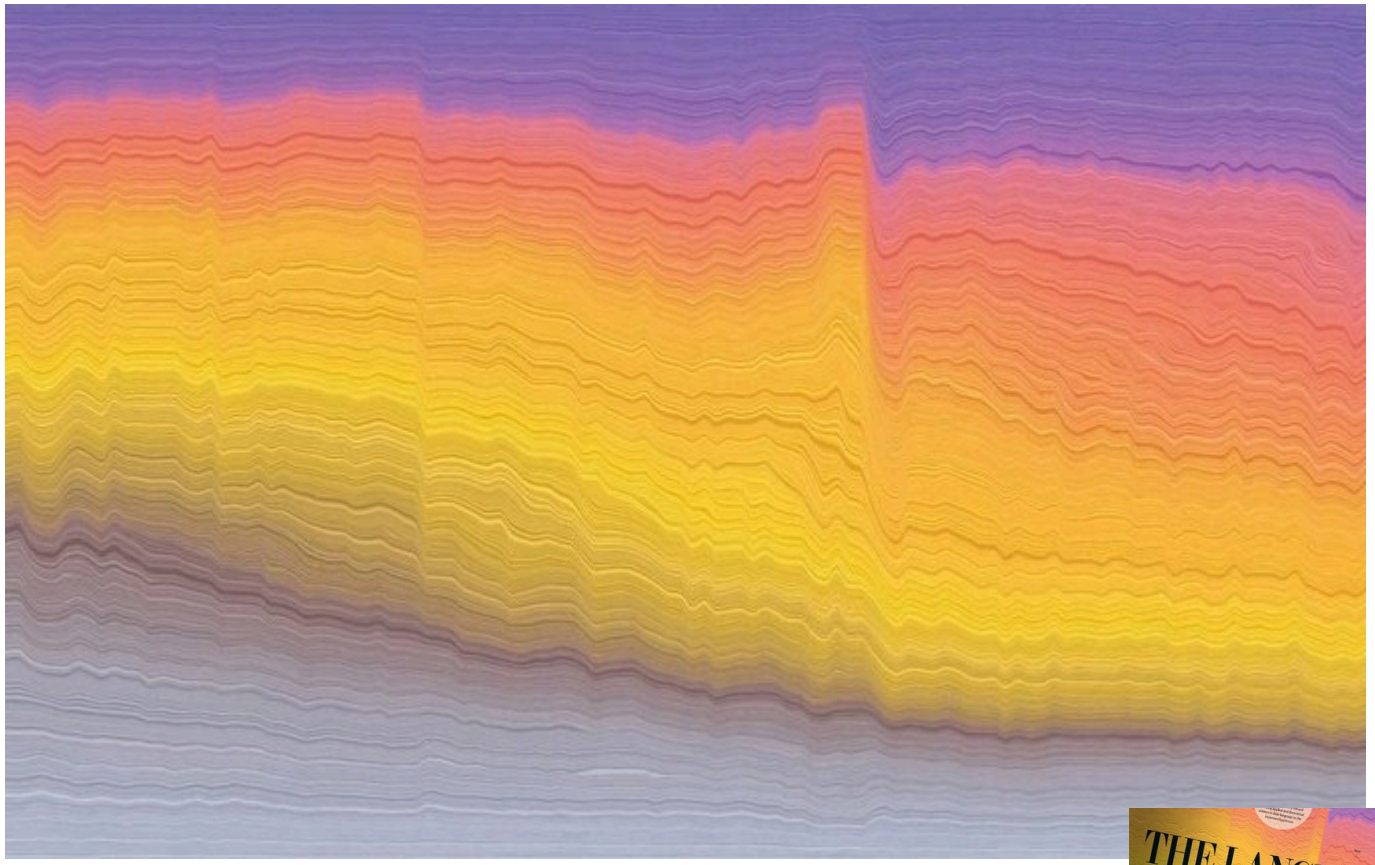
- Small multiples

Step 4: Implementation

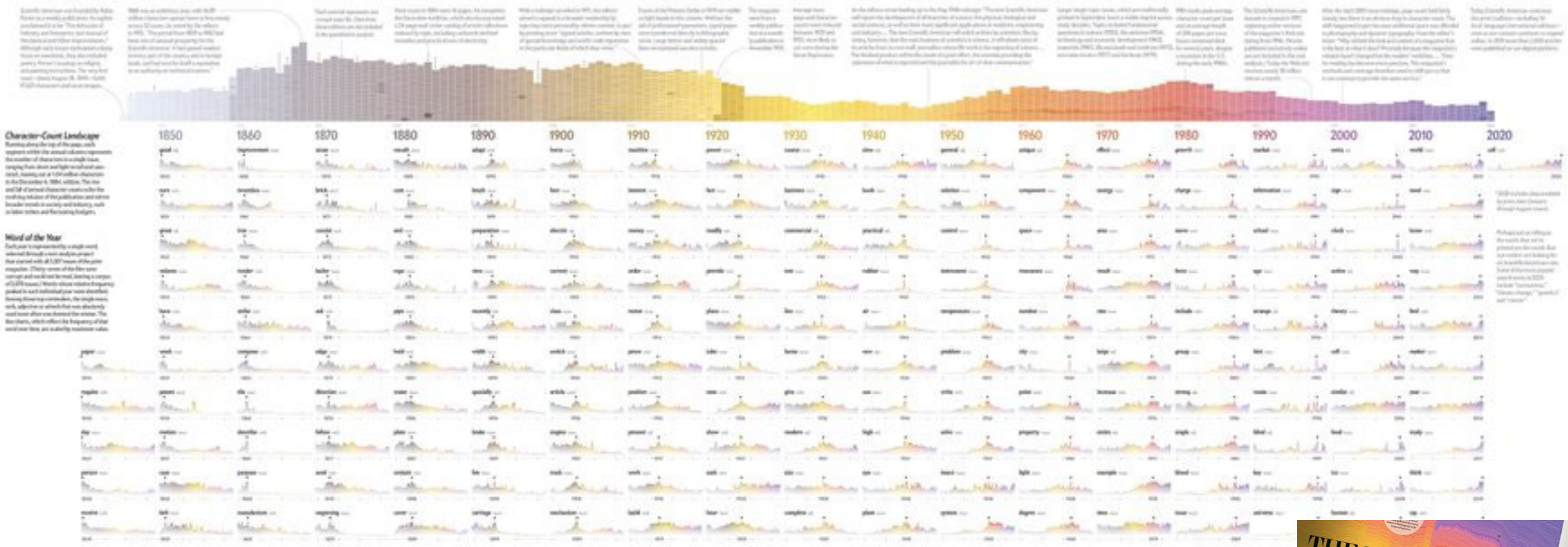
- Taking into account medium restrictions and possibilities
- Clarity, design, final touches
- Colour encoding: for aesthetical and meaning reasons, from vintage brown trough yellow, orange to modern purple

4: Final cover view, on paper: Stacked area chart

Stack area chart where newer words area stacked on top of the older ones – showing the major breaks and shifts in the vocabulary akin to soil layers

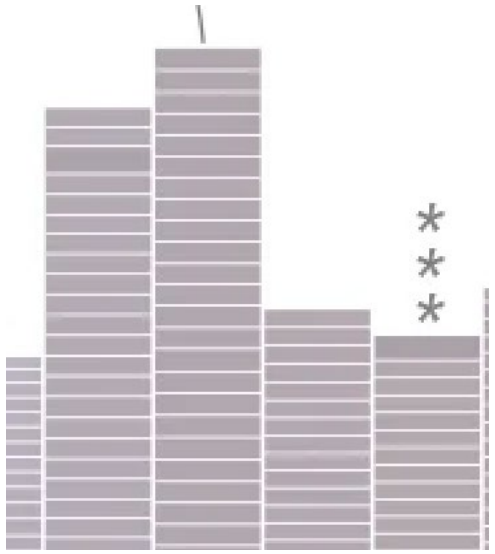


4: Final inside view, on paper



4: Paper, character-count Landscape

- Each segment within the annual column represents the number of characters in a single issue, ranging form short and light to tall and saturated, maxing out at 1.04 million characters in the December 6, 1884, edition.

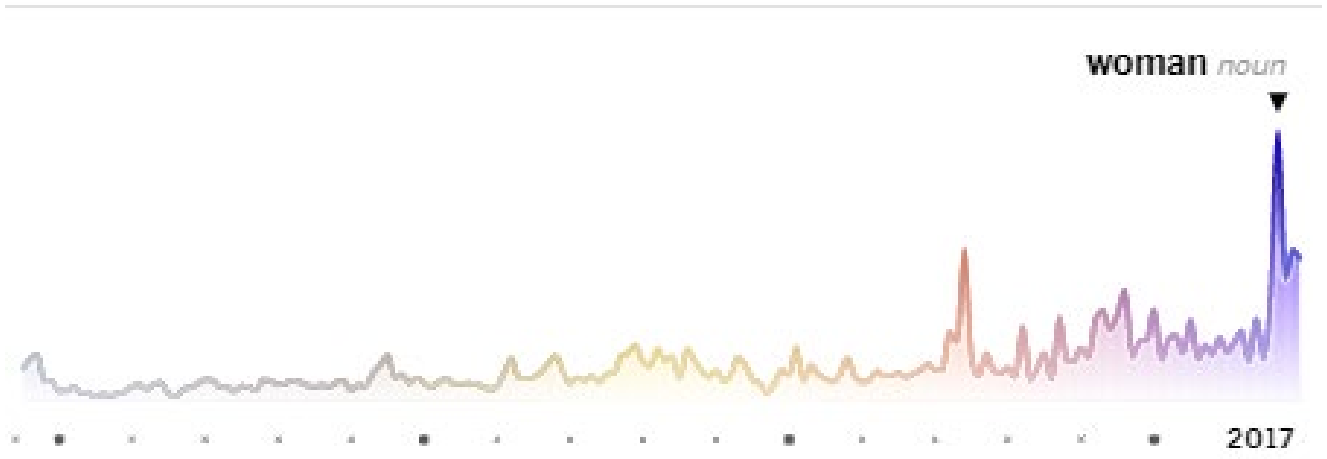


4: Paper, Word of the year

Each year is represented by a single word,

Selected through a text-analysis project of the print magazine. Words whose relative frequency peaked in each individual year were identified. The single noun, verb, adjective or adverb that was absolutely used most often was deemed the winner.

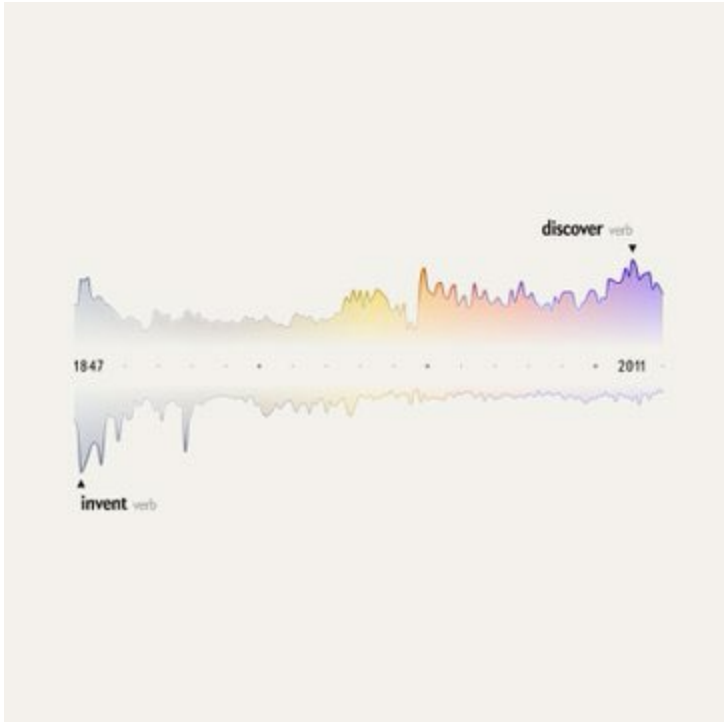
The line charts, which reflect the frequency of that word over time, are scaled by maximum value.

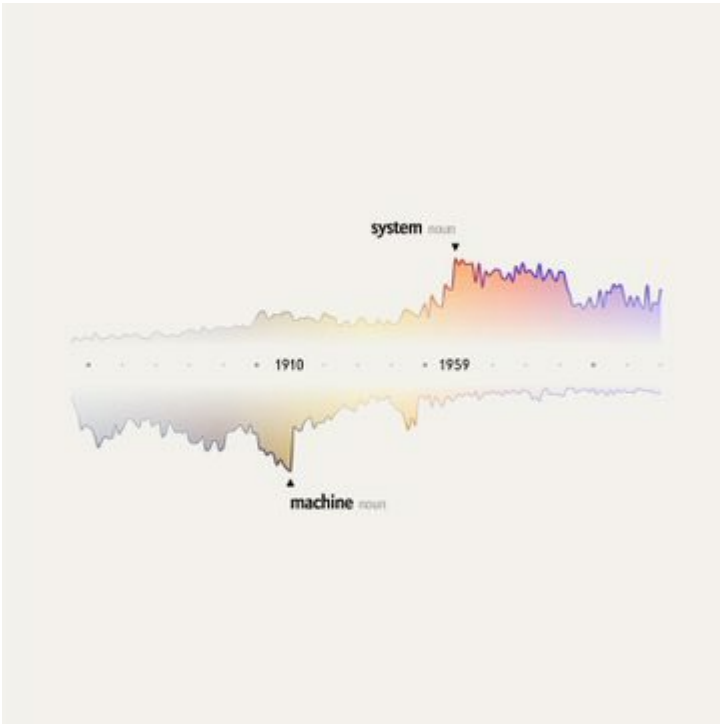


4: Paper, Comparison of words

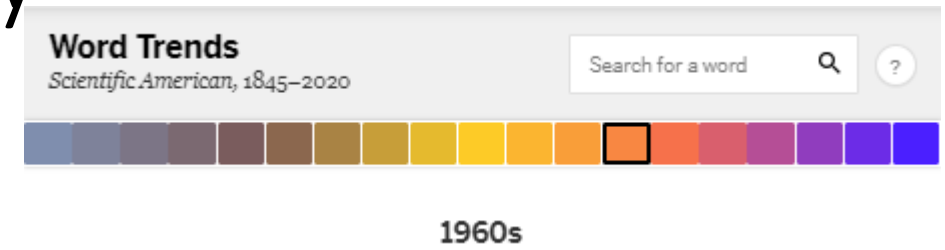
It also features a couple of hand-picked juxtapositions, to show how the magazine's perspective has changed — from "machine" to "system" or "invent" to "discover".

Serendipity + Curation criteria

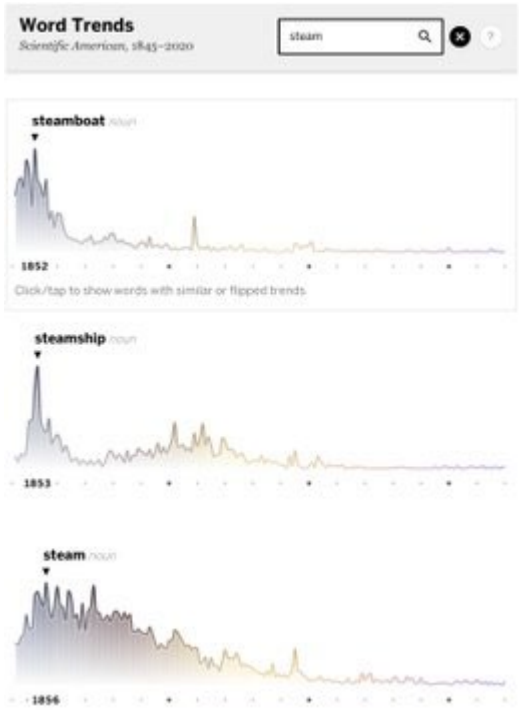




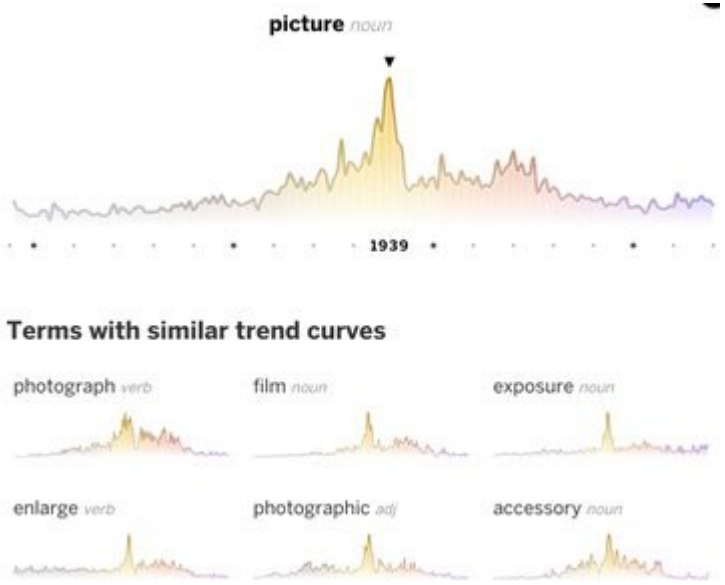
4: Digital, interactivity



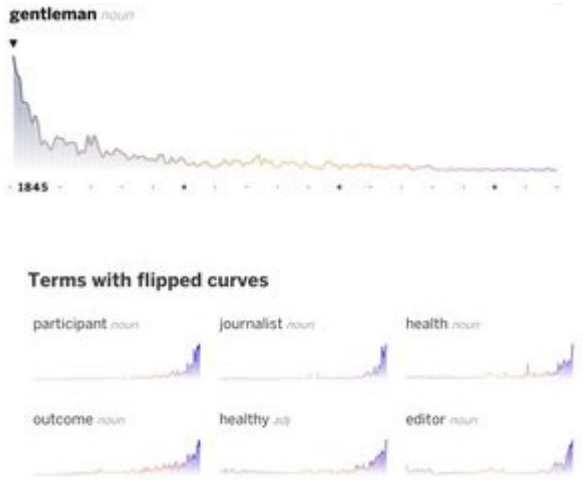
Search function



Similar trend curves



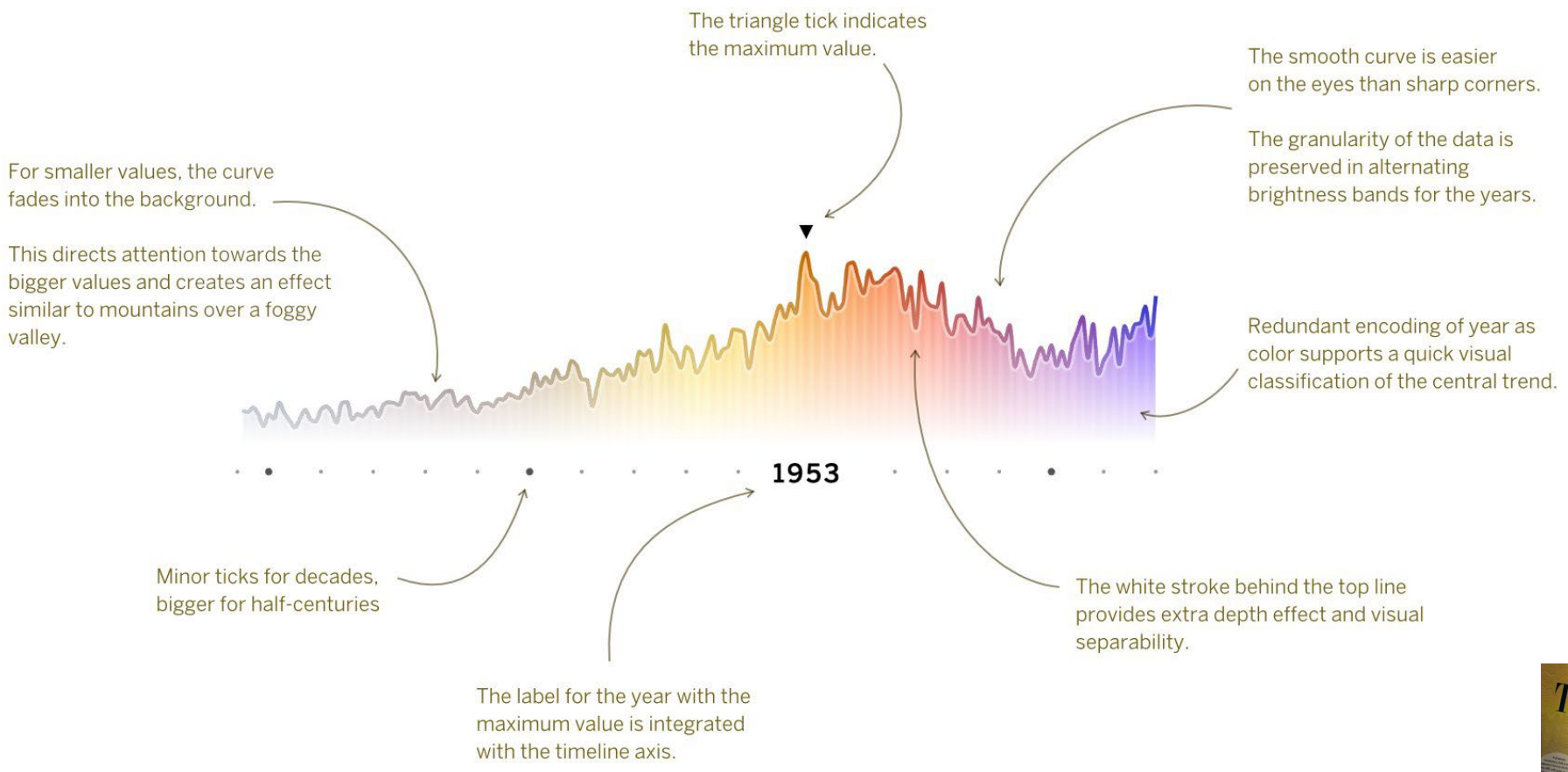
Similar / Flipped trend curves



4: Refining and finishing details

Word Trends
in Scientific American 1845–2020

Design Details



References

- S2E7 - JEN CHRISTIANSEN & MORITZ STEFANER - 175 years of word usage patterns in Scientific American, within Explore Explain, by Andy Kirk https://www.youtube.com/watch?v=EvWEEoh0d8E
- Moritz Stefaner, Lorraine Daston, Jen Christiansen “The Language of Science: How the words we use have evolved over the past 175 years” Scientific American, September 1, 2020 https://www.scientificamerican.com/article/the-language-of-science/
- Moritz Stefaner, How to turn 175 years of Words in Scientific American into an Image: a data designer explains the art and science of analyzing and charting text from 5107 issues of this magazine. Scientific American, September 1, 2020 https://www.scientificamerican.com/article/how-to-turn-175-years-of-words-in-scientific-american-into-an-image/
- Moritz Stefaner, Explore 175 Years of Words in Scientific American. Scientific American, August 18, 2020 https://www.scientificamerican.com/article/explore-175-years-of-words-in-scientific-american/
- Moritz Stefaner, report on Twitter on the making-off https://twitter.com/moritz_stefaner/status/1295733727704875010

- https://mobile.twitter.com/hashtag/SciAm175
- https://twitter.com/hashtag/sciencewords