Visualization Types II

Text Visualizations

April 9, 2013 – Michael Porath (@poezn)

Assignment 4

Feedback

Narrative

Oxford English Dictionary

nar·ra·tive *noun* \ˈner-ə-tiv, ˈna-rə-\

an account of a
series of events, facts, etc.,
given in order and
with the establishing of connections
between them.

Framework

Guide. Highlight. Interpret

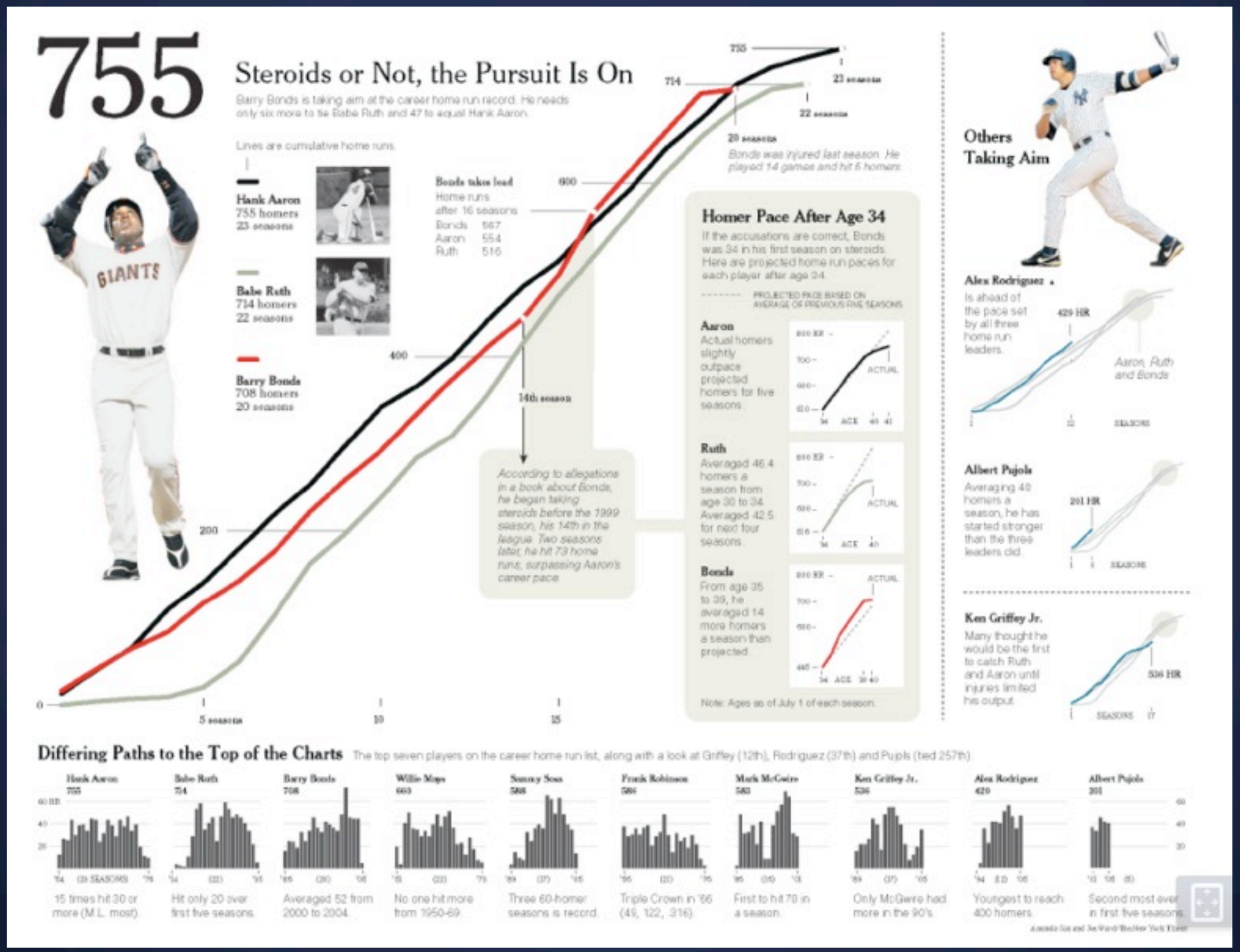
Visual Structure support the story

Messaging tell the story

Interactivity engage the story

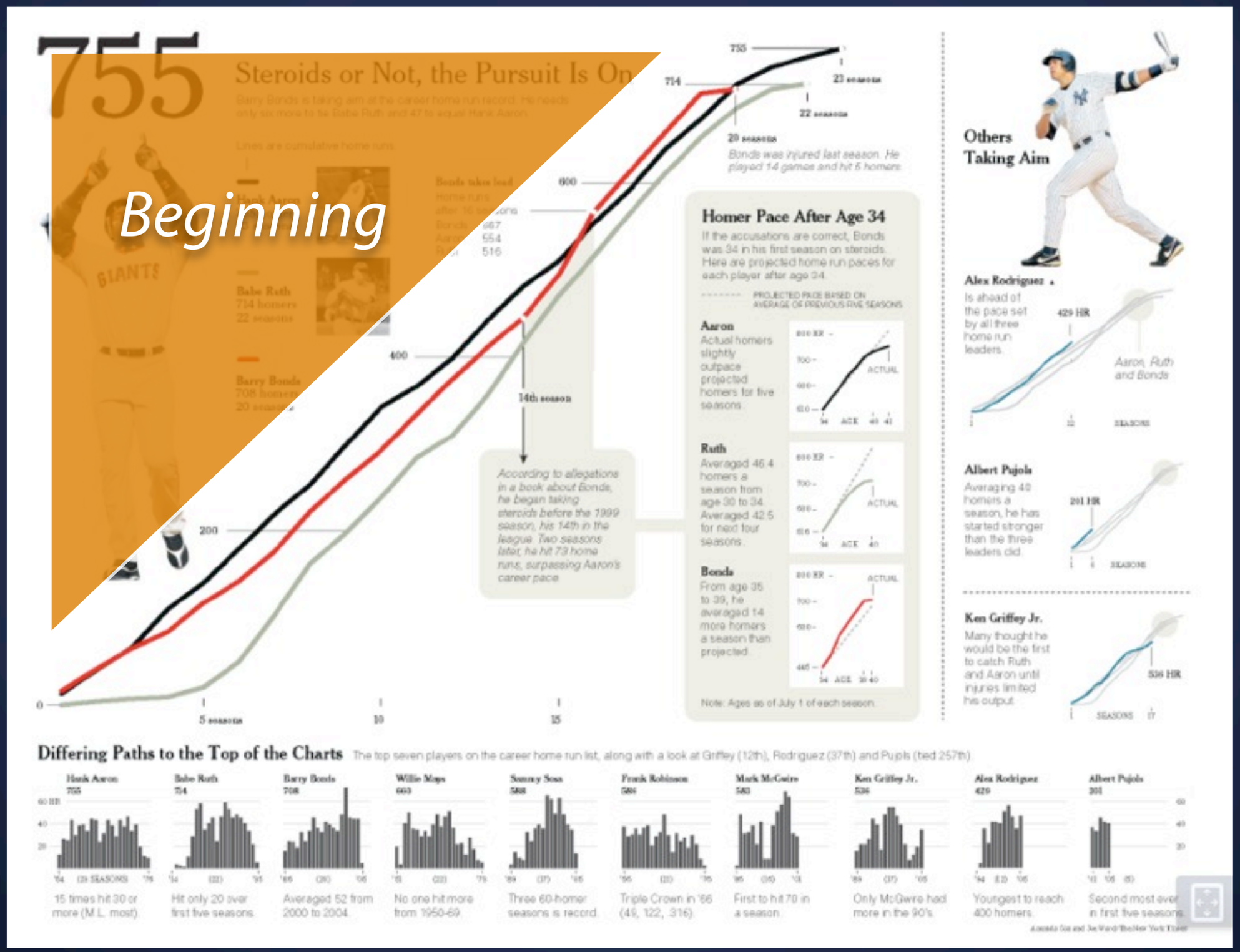
Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



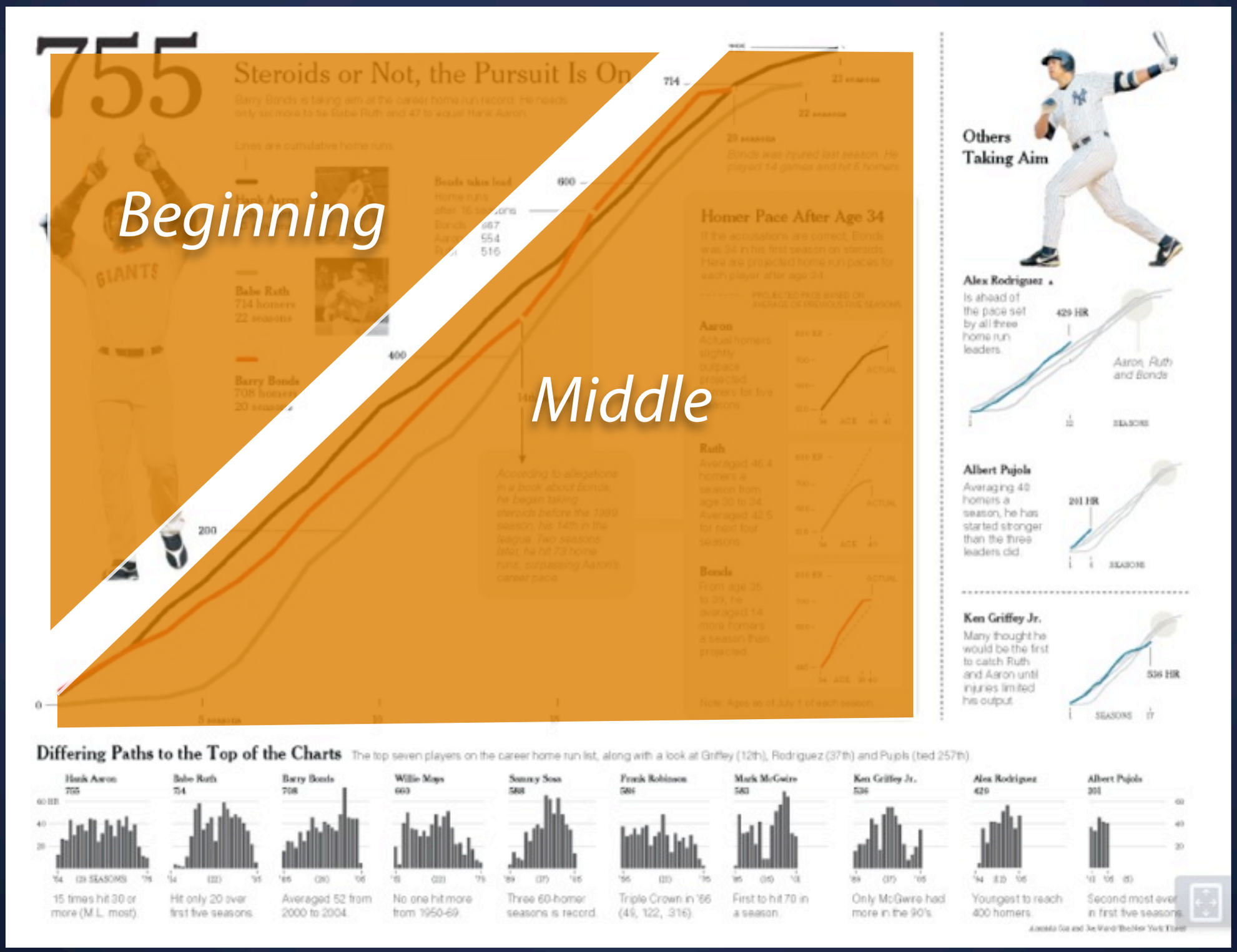
Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



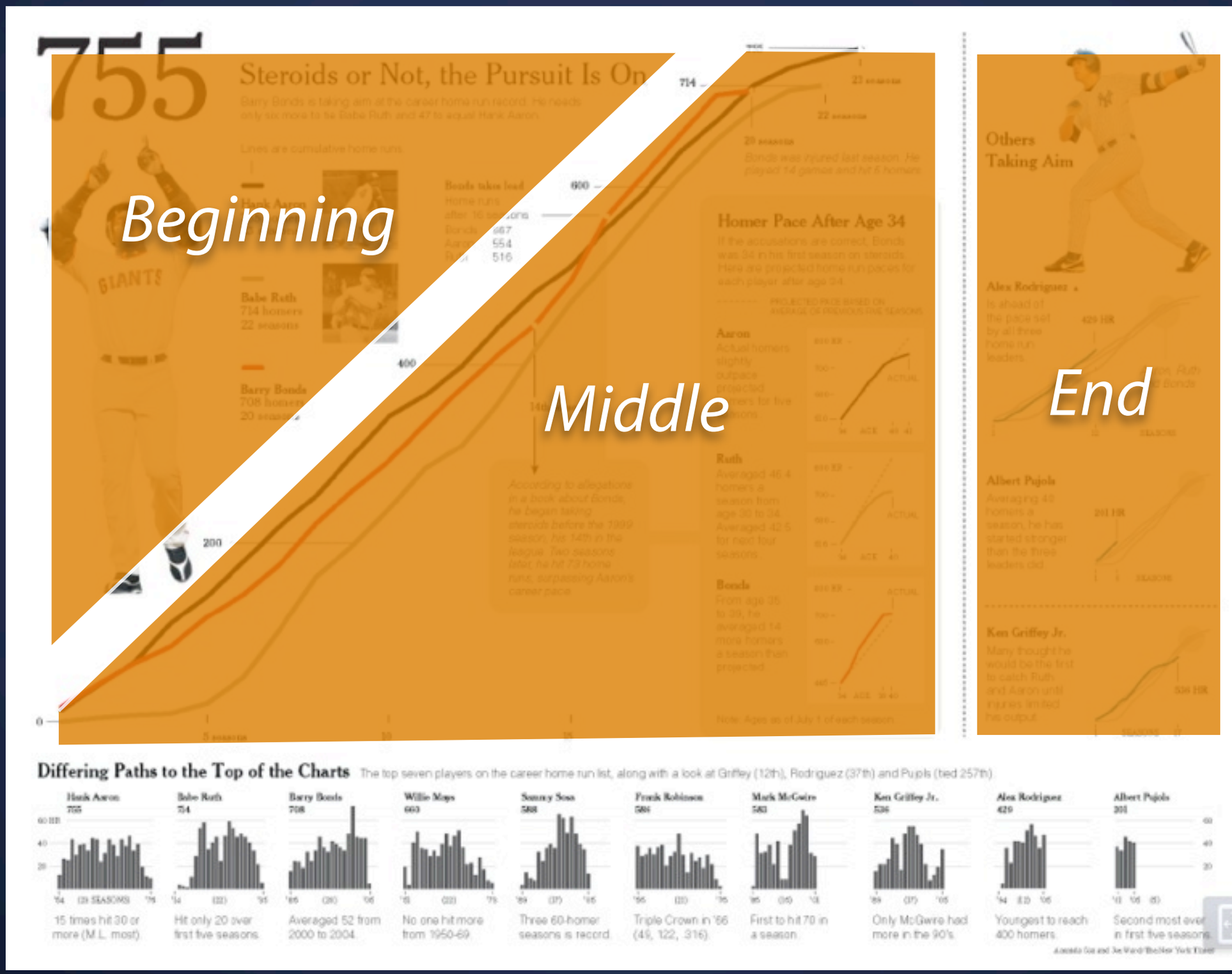
Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



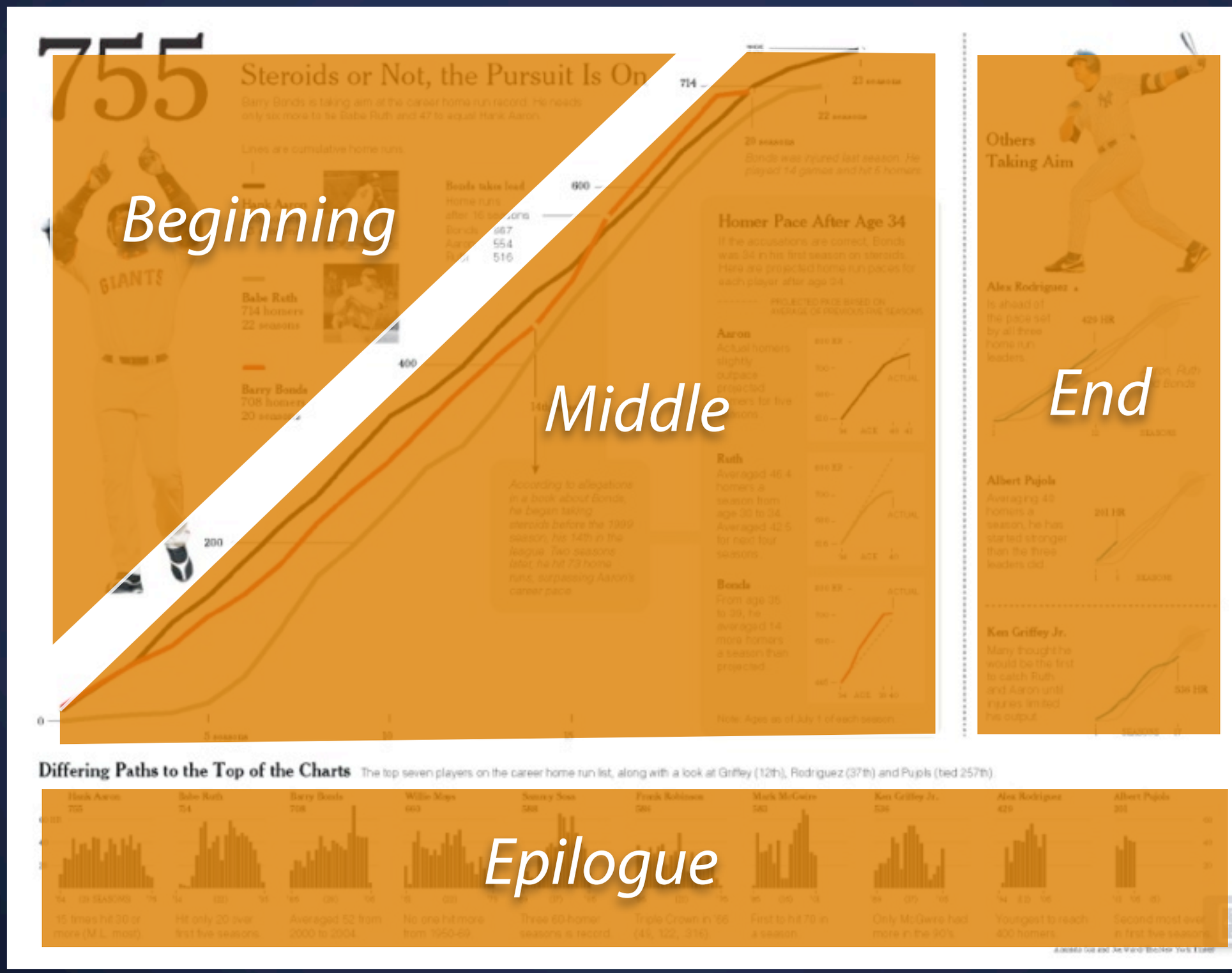
Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



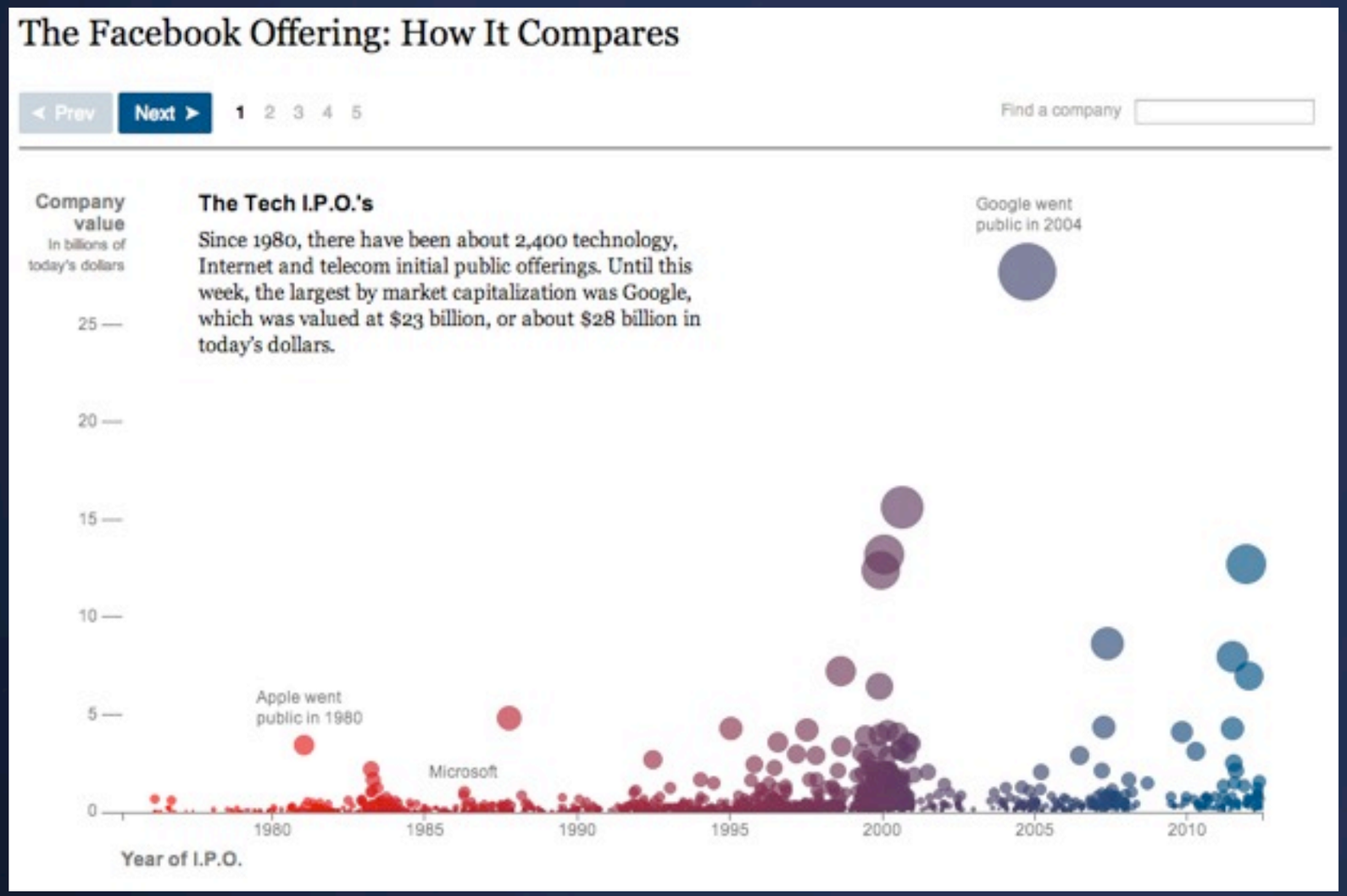
Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



Messaging

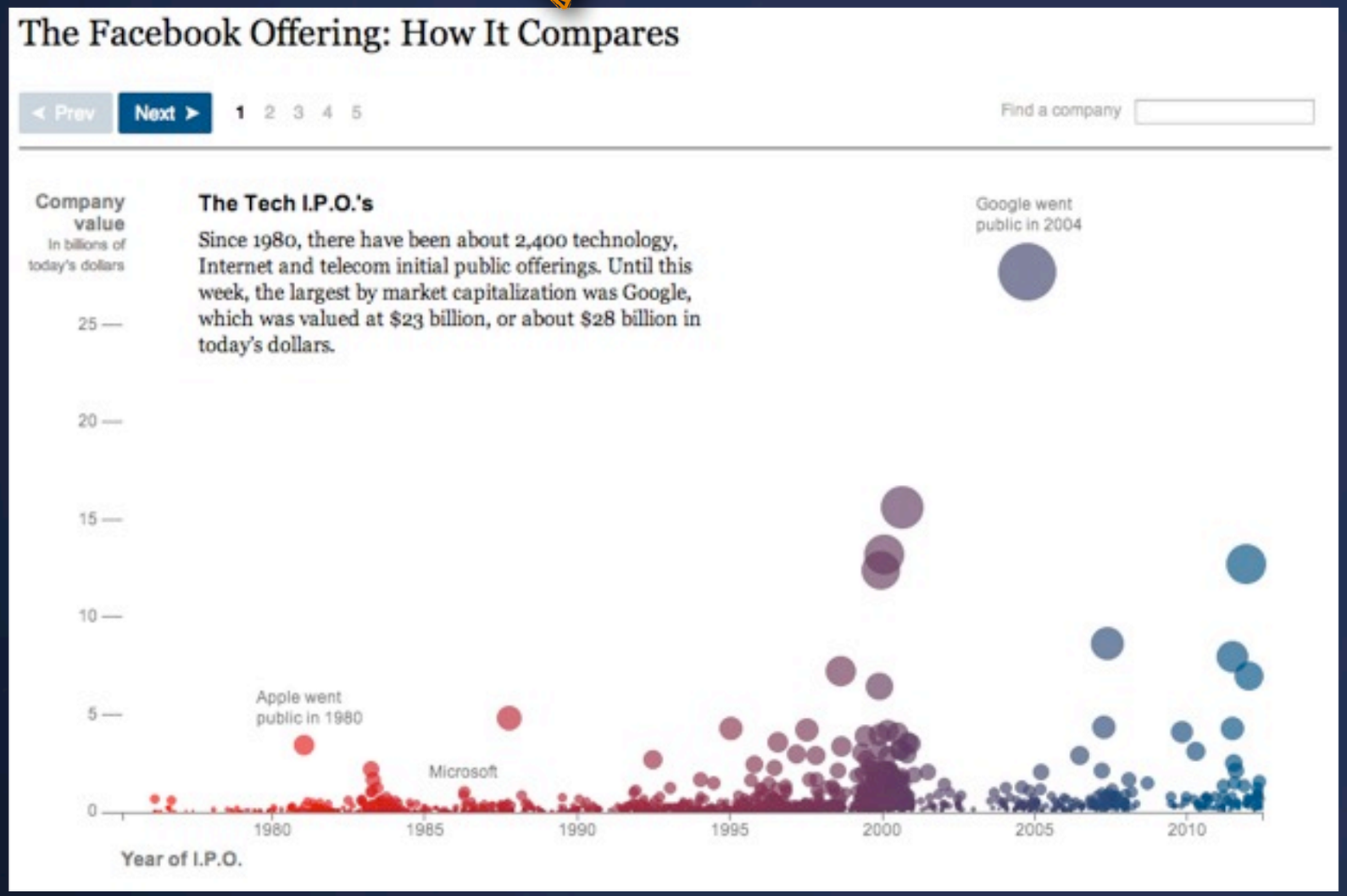
Tell the story; not JUST with visuals



Messaging

Tell the story; not JUST with visuals

Headline

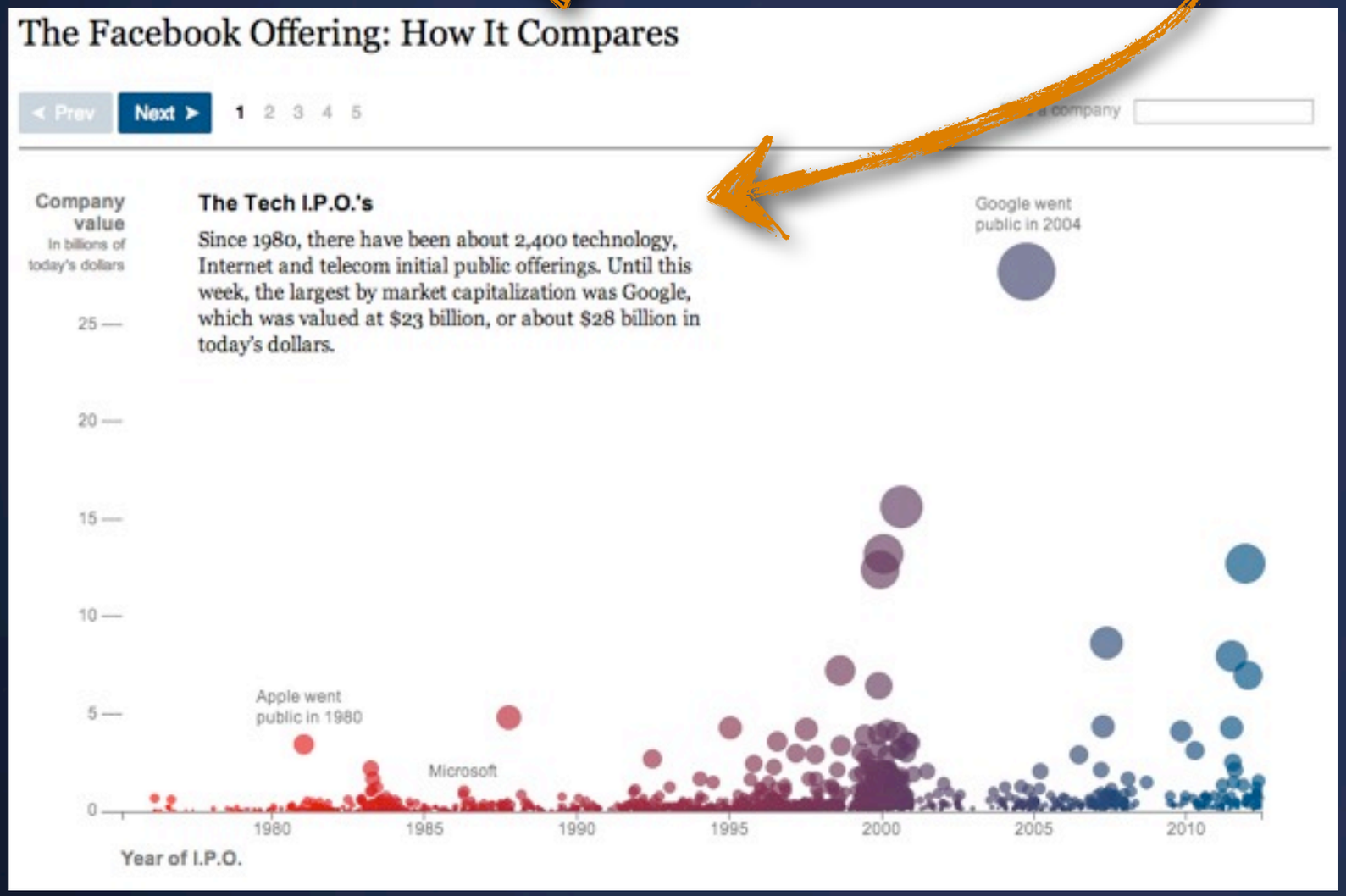


Messaging

Tell the story; not JUST with visuals

Headline

Caption



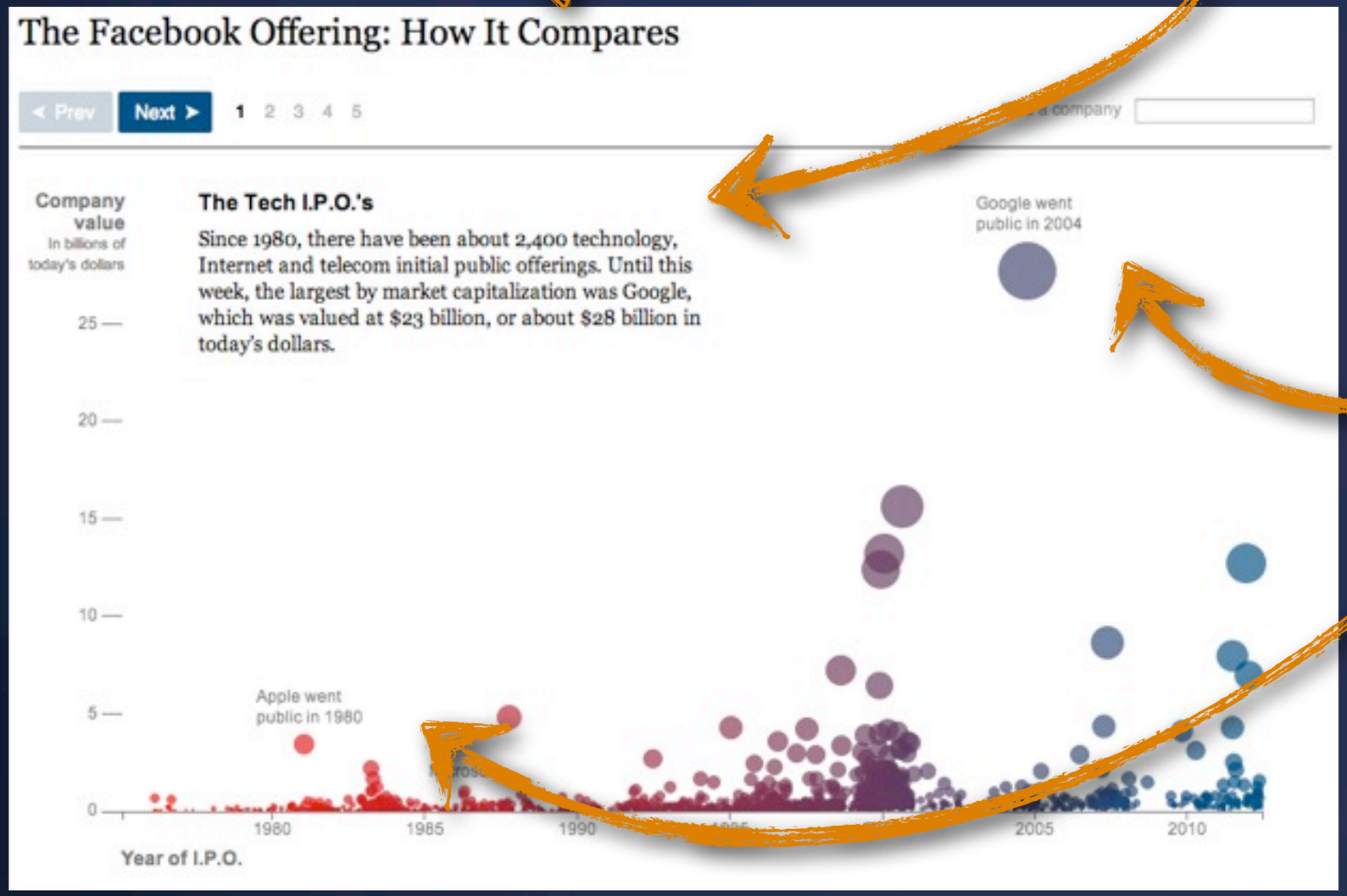
Messaging

Tell the story; not JUST with visuals

Headline

Caption

Annotation

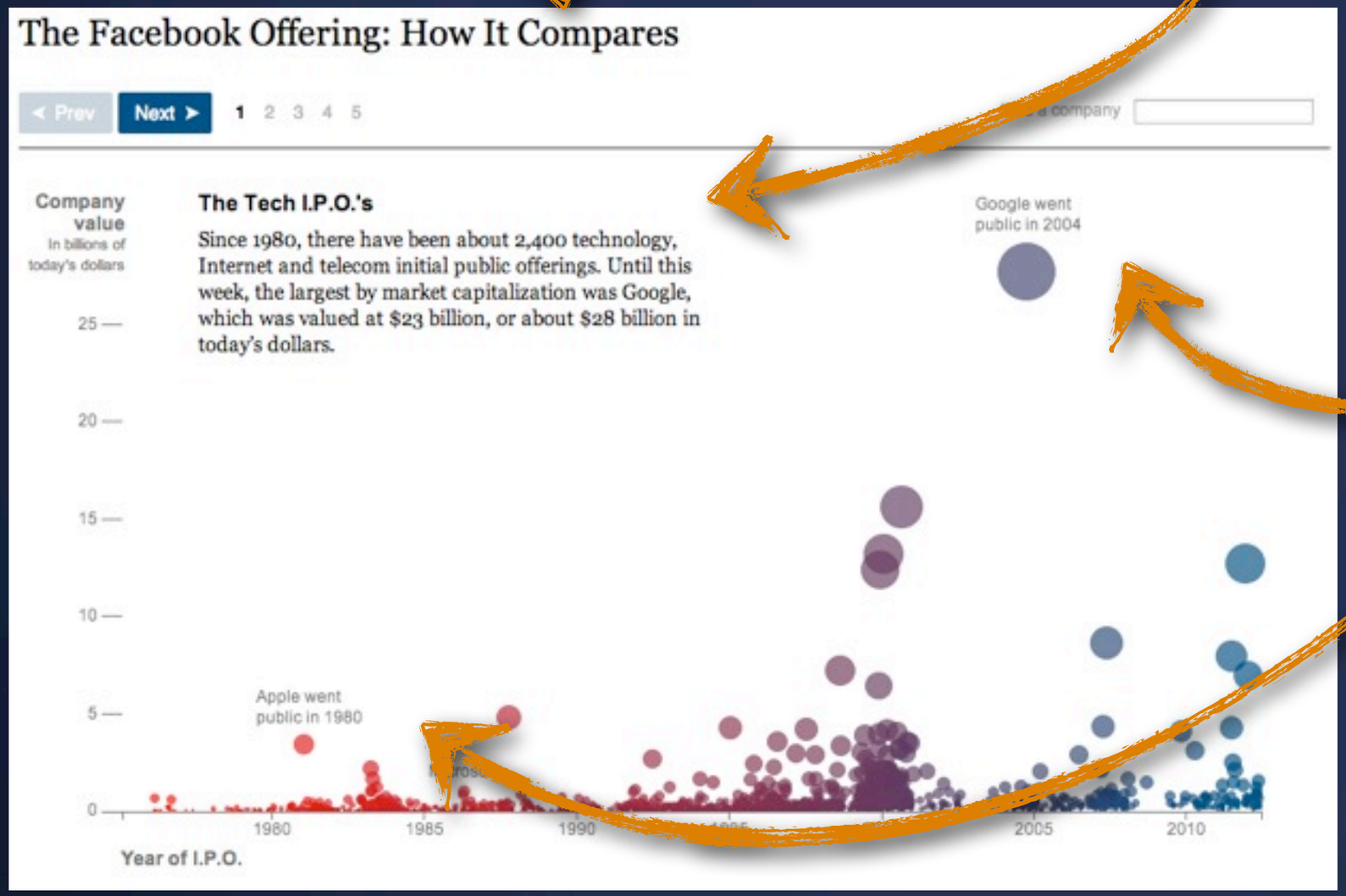


Messaging

Tell the story; not JUST with visuals

Headline

Caption

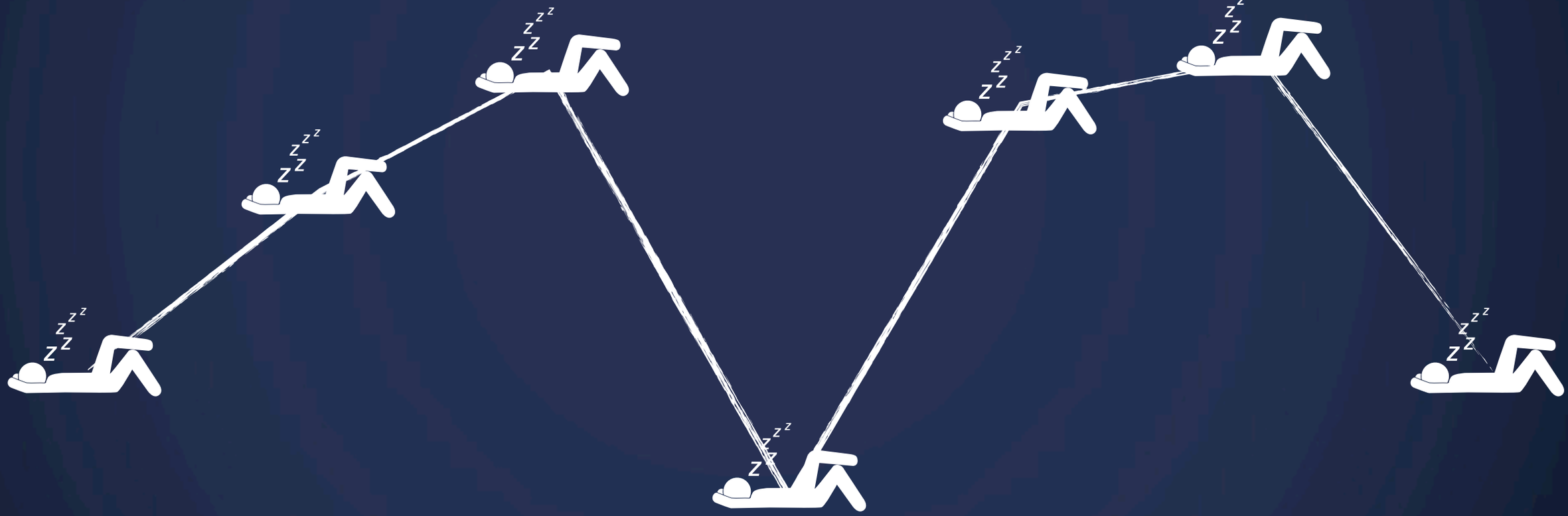


Annotation

weave text *into* the graphic!

Icons

Be **very** careful!



Sleep designed by DonBLC 123 from The Noun Project

Icons

Be **very** careful!

Sleep

hours per night

Midterm due
the next day

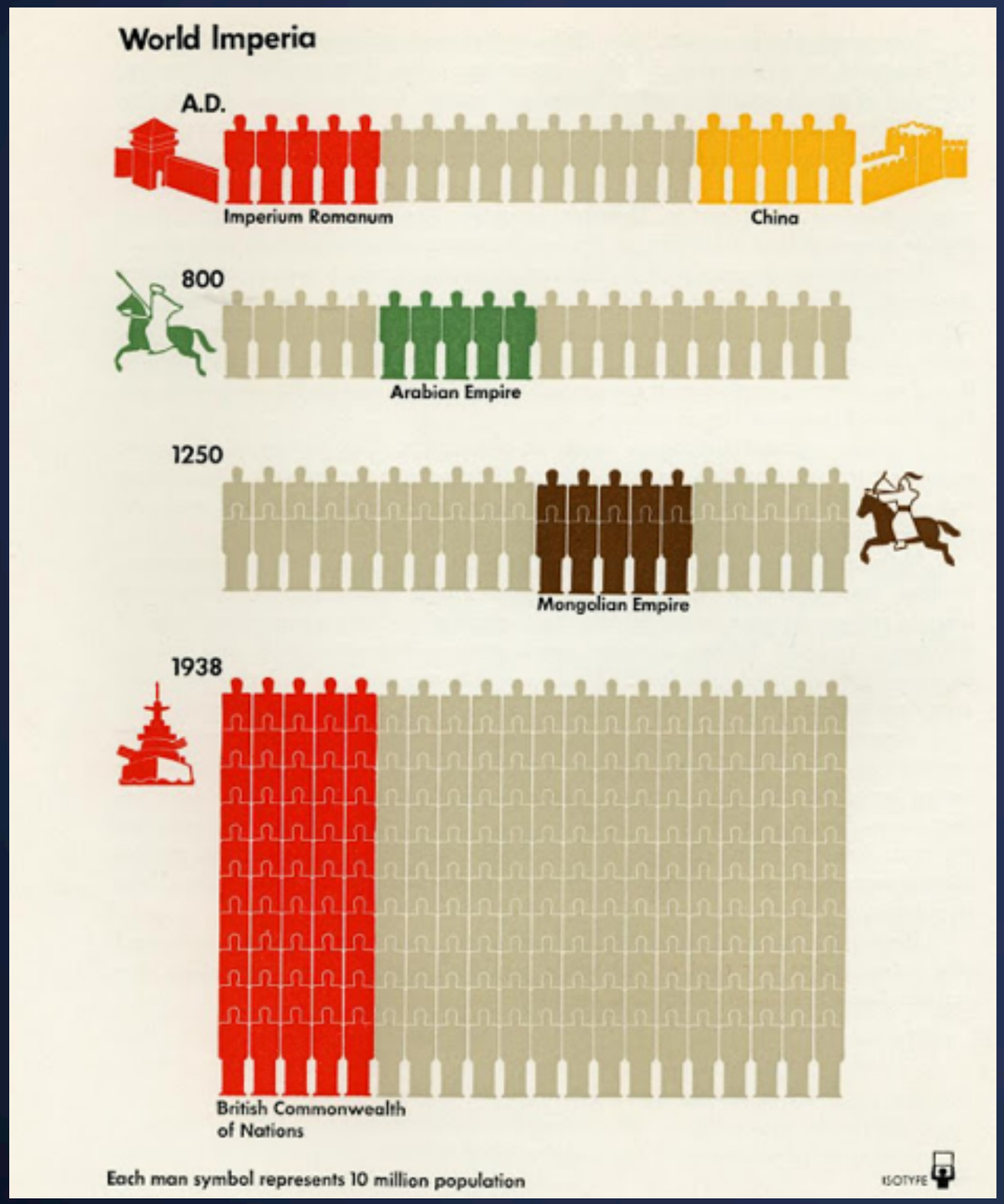
8

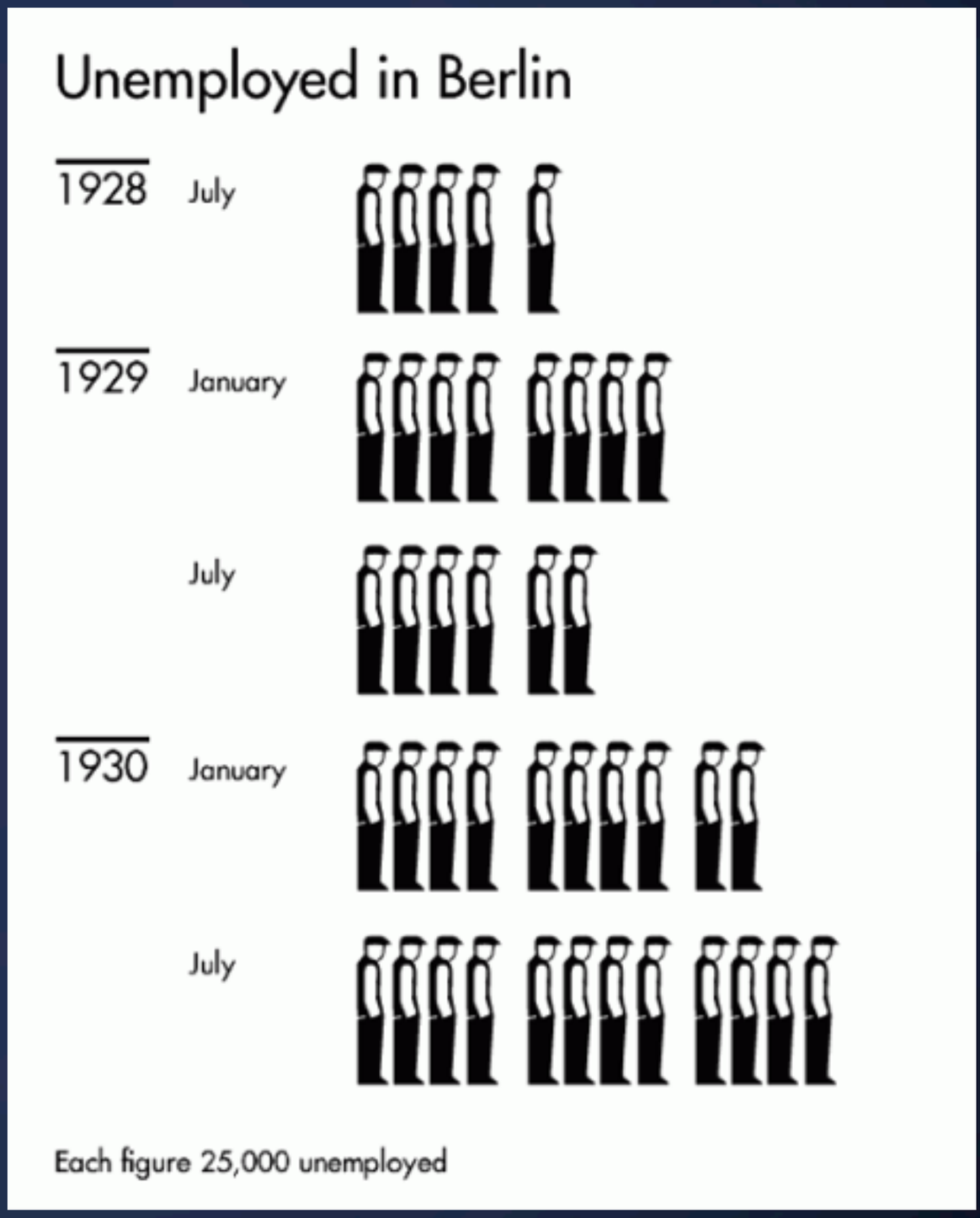
6

Sleep designed by DonBLC 123 from The Noun Project

Icons

Pictograms / Unit Chart: Otto Neurath / Gerd Arntz

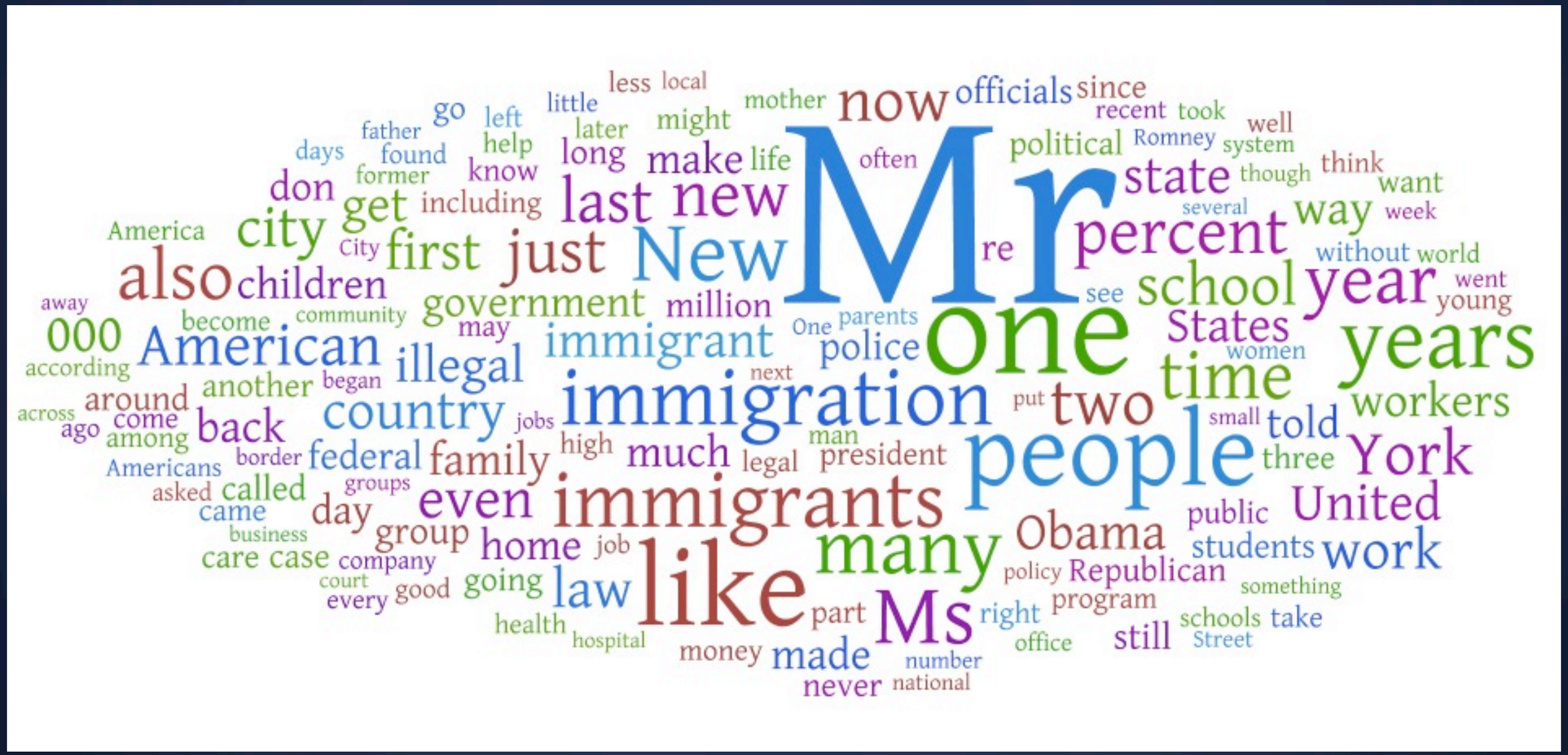




When I say “Text Visualization” you say... ?

Wordle!

Visualizing Word Frequencies



Wordle

Color

Wordle

Context

Wordle

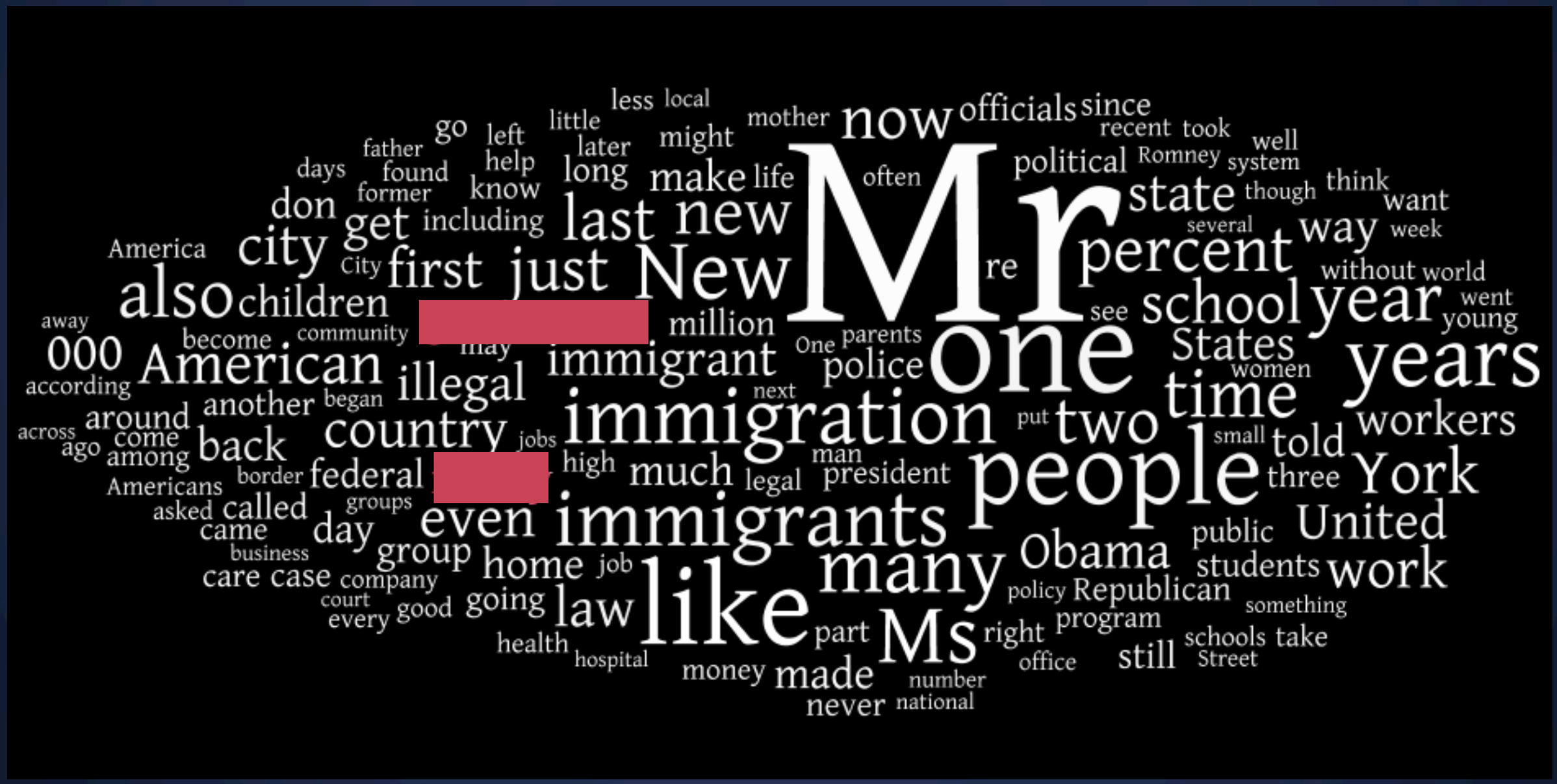
Context

Wordle

Which one has more mentions?

Wordle

Which one has more mentions?



What is text data?

What can we visualize?

What is text data?

What can we visualize?

Documents

Articles, books, novels

Emails, web pages, blogs

Computer programs, logs

What is text data?

What can we visualize?

Documents

Articles, books, novels

Emails, web pages, blogs

Computer programs, logs

Document Collections

Messages (emails, blogs, comments, tweets)

Academic publications

What is text data?

What can we visualize?

Documents

Articles, books, novels

Emails, web pages, blogs

Computer programs, logs

Document Collections

Messages (emails, blogs, comments, tweets)

Academic publications

Meta Data

Tags

Protagonists/ Characters

Text length

language

release date

What tasks? Discuss.

What is the reader's task/objective when looking at a visualization of textual data?

Why visualize text?

Goals, tasks, objectives

Why visualize text?

Goals, tasks, objectives

Understanding get the "gist" of a document

Why visualize text?

Goals, tasks, objectives

Understanding get the "gist" of a document

Grouping cluster for overview or classification

Why visualize text?

Goals, tasks, objectives

Understanding get the "gist" of a document

Grouping cluster for overview or classification

Comparing compare document collections, or inspect evolution of collection over time

Why visualize text?

Goals, tasks, objectives

Understanding get the "gist" of a document

Grouping cluster for overview or classification

Comparing compare document collections, or inspect evolution of collection over time

Correlating compare patterns in text to those in other data

Text as data

Properties of text

Text as data

Properties of text

Text is highly dimensional (each word is a dimension)

Text as data

Properties of text

Text is highly dimensional (each word is a dimension)

Semantics are important:

Text as data

Properties of text

Text is highly dimensional (each word is a dimension)

Semantics are important:

Order April, February, January, March, May

Text as data

Properties of text

Text is highly dimensional (each word is a dimension)

Semantics are important:

Order April, February, January, March, May

Membership Tennis, Running, Swimming, Hiking

Text as data

Properties of text

Text is highly dimensional (each word is a dimension)

Semantics are important:

Order April, February, January, March, May

Membership Tennis, Running, Swimming, Hiking

Hierarchy, Synonyms, Antonyms, etc.

Text Processing

The ‘Data Cleaning’ for text data

Tokenization

Segment text into terms.

- Remove stop words?
  a, an, the, of, to, be
- Numbers and symbols?
  #i247, @berkeleyischool
- Entities?
  San Francisco, O’Connor, U.S.A.

Stemming

Group together different forms of a word

- Porter stemmer?
  *visual*ization(s), *visual*ize(s), *visual*ly
  ➜ ***visual***
- Lemmatization? goes, went, gone
  ➜ ***go***

Document-Level Visualizations

Word Frequency Count

wordcount.org

WordCount™ ©2003 Jonathan Harris | Number27 | Help

Tag Clouds

See beginning of the lecture

Word Frequency Visualizations

Limitations?

Word Frequency Visualizations

Limitations?

Focus on unigrams

Health

Care

⇔ Health Care

Word Frequency Visualizations

Limitations?

Focus on unigrams

Health

Care

⇔ Health Care

Often favors frequent (TF) or rare (IDF) terms

Word Frequency Visualizations

Limitations?

Focus on unigrams

Health

Care

⇔ Health Care

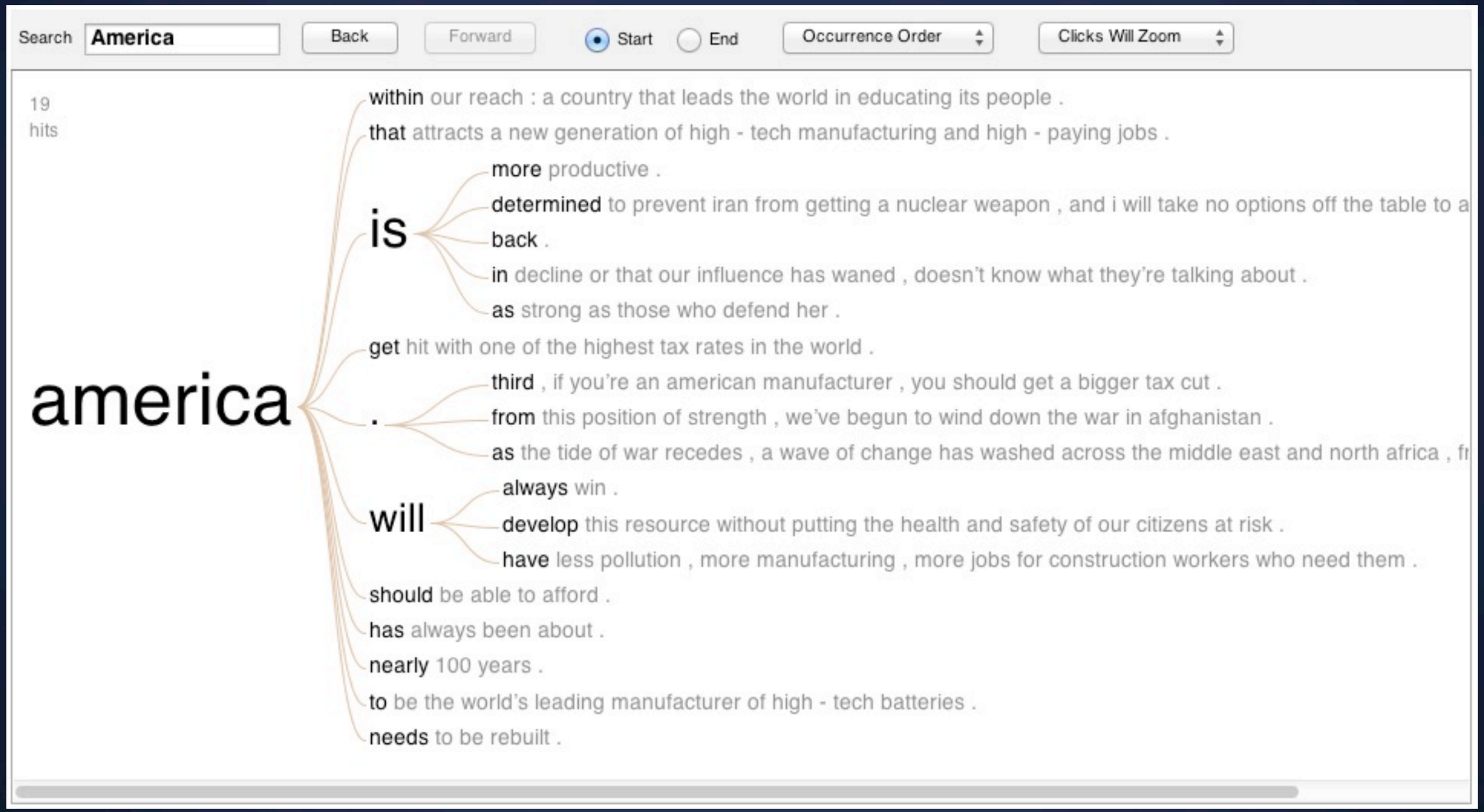
Often favors frequent (TF) or rare (IDF) terms

Ignores additional information

Grammar / part-of-speech
Position within document
Recognizable entities

Word Tree

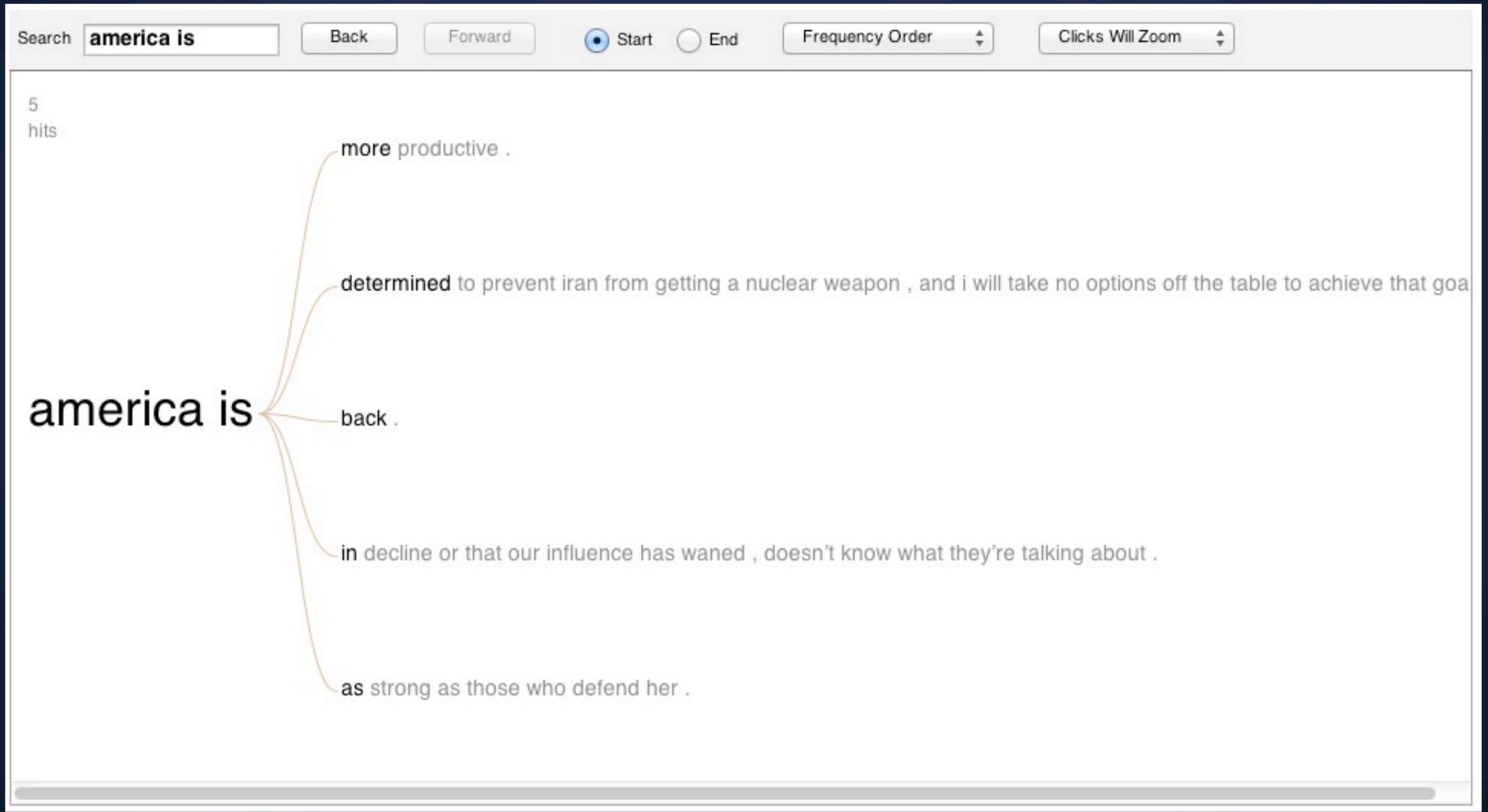
Considering n-grams (SOTU 2012)



http://www-958.ibm.com/software/analytics/labs/manyeyes/#vis=500404

Word Tree

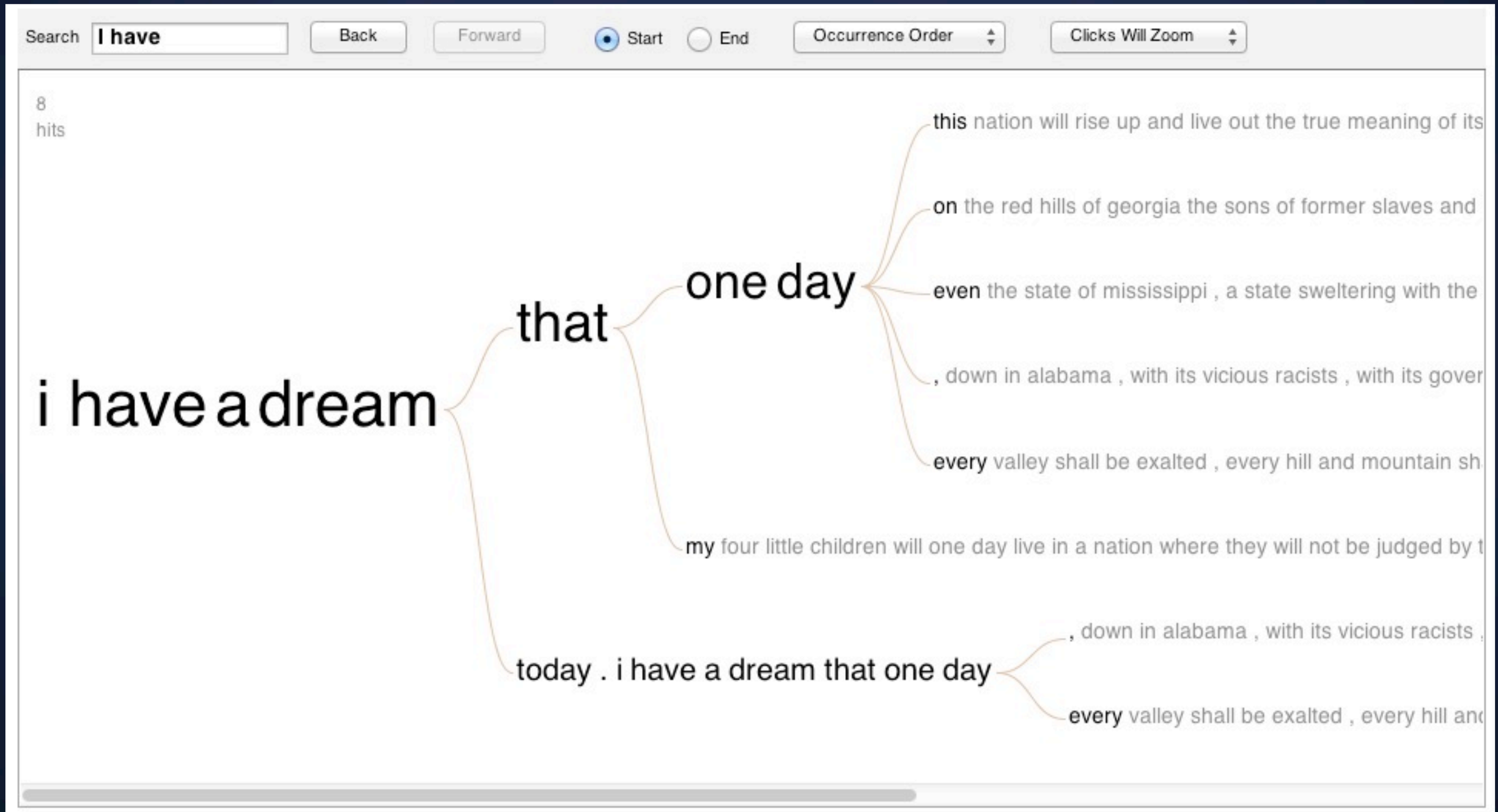
Considering n-grams (SOTU 2012)



http://www-958.ibm.com/software/analytics/labs/manyeyes/#vis=500404

Word Tree

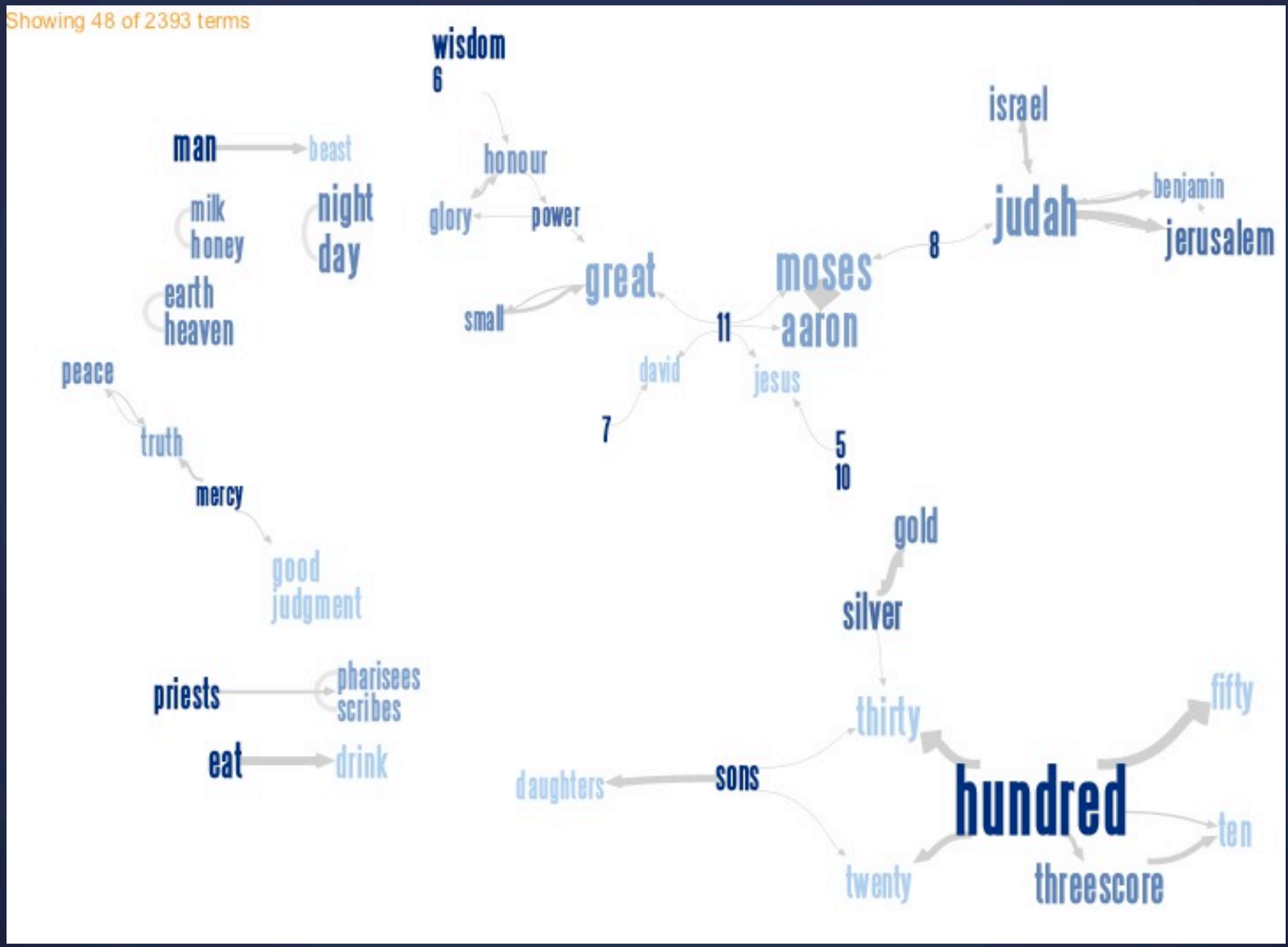
Visualizing rhetorical devices / repetition



http://www-958.ibm.com/software/analytics/labs/manyeyes/#vis=500405

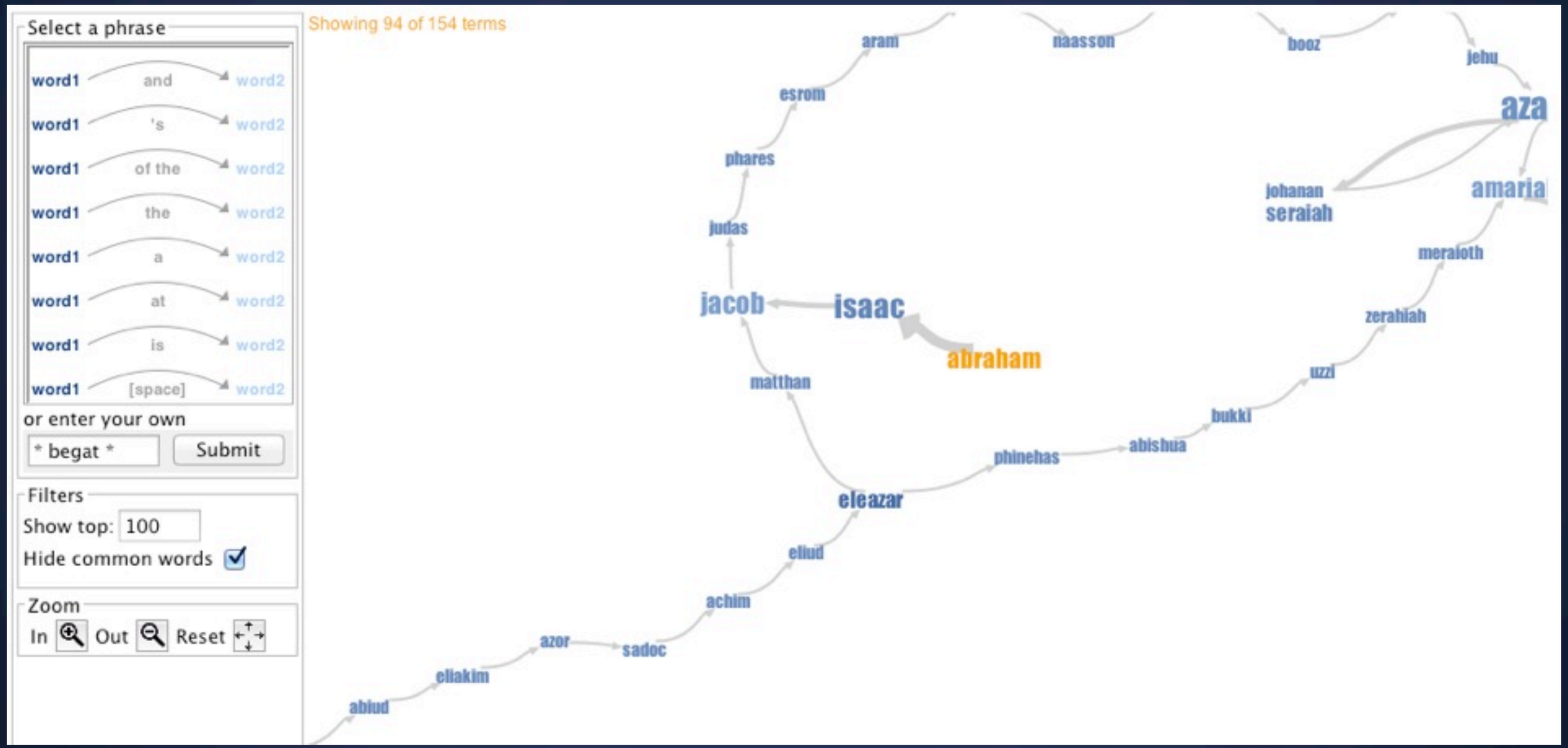
Phrase Nets

Relationship between words (Here: The Bible)



Phrase Nets

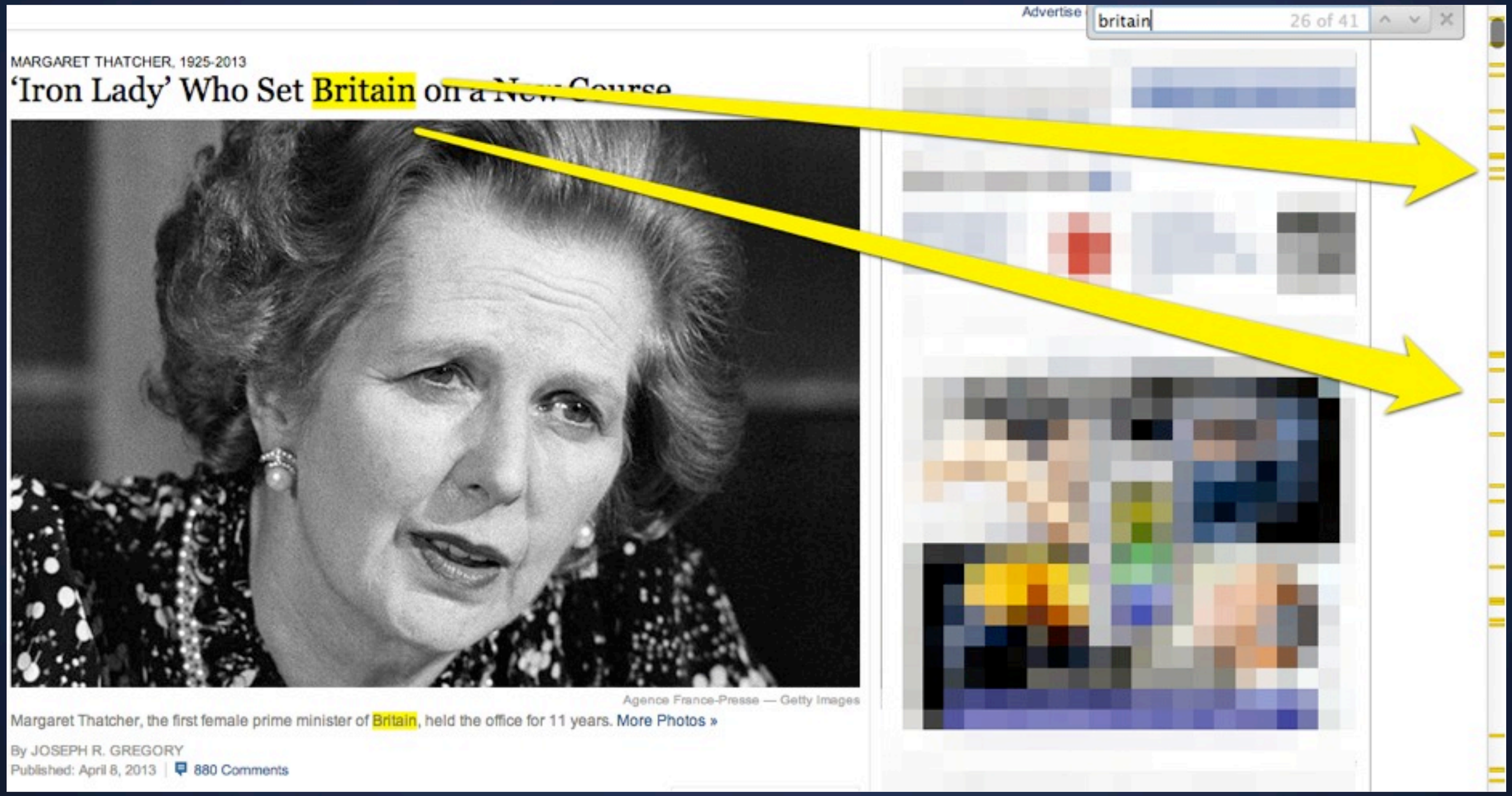
The Bible: X - begat - Y



http://www-958.ibm.com/software/analytics/manyeyes/visualizations/who-begat-whom-in-the-bible

Google Chrome

In-page search: Britain



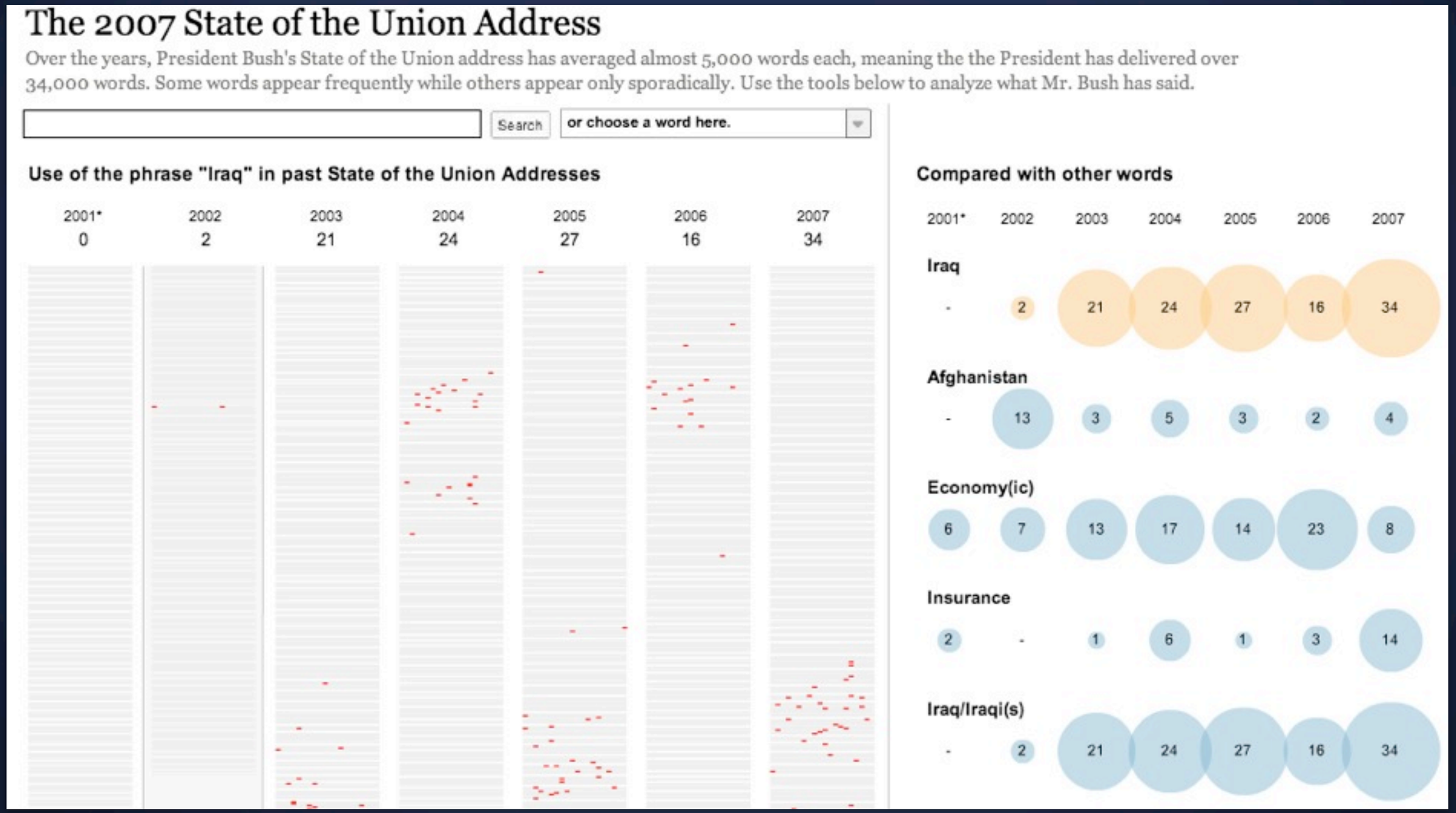
Google Chrome

In-page search: Britain

Collection-Level Visualizations

George W. Bush's speeches

State of the Union / NY Times



http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html

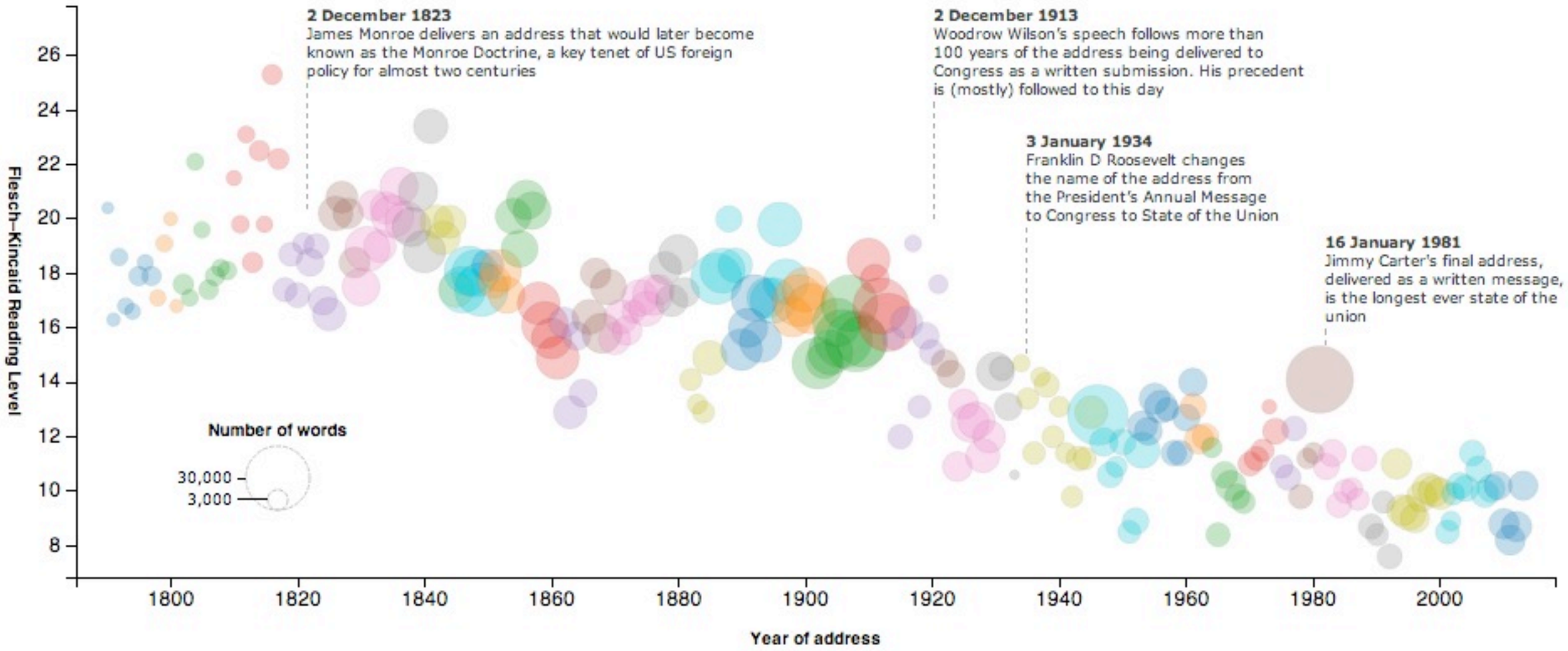
Reading Level

State of the union speeches over time

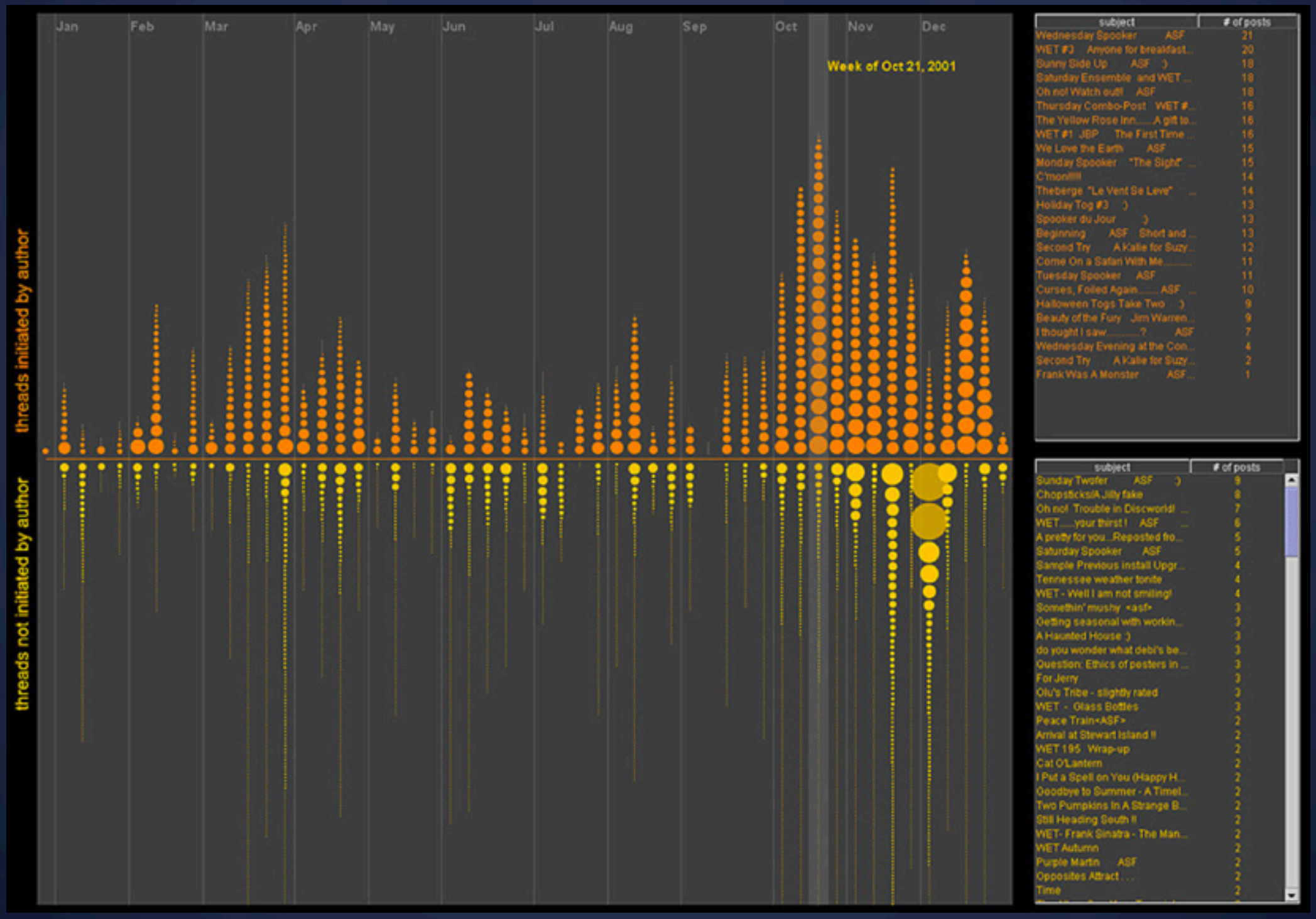
The state of our union is ... dumber:

How the linguistic standard of the presidential address has declined

Using the Flesch-Kincaid readability test the Guardian has tracked the reading level of every state of the union



Newsgroup Usage

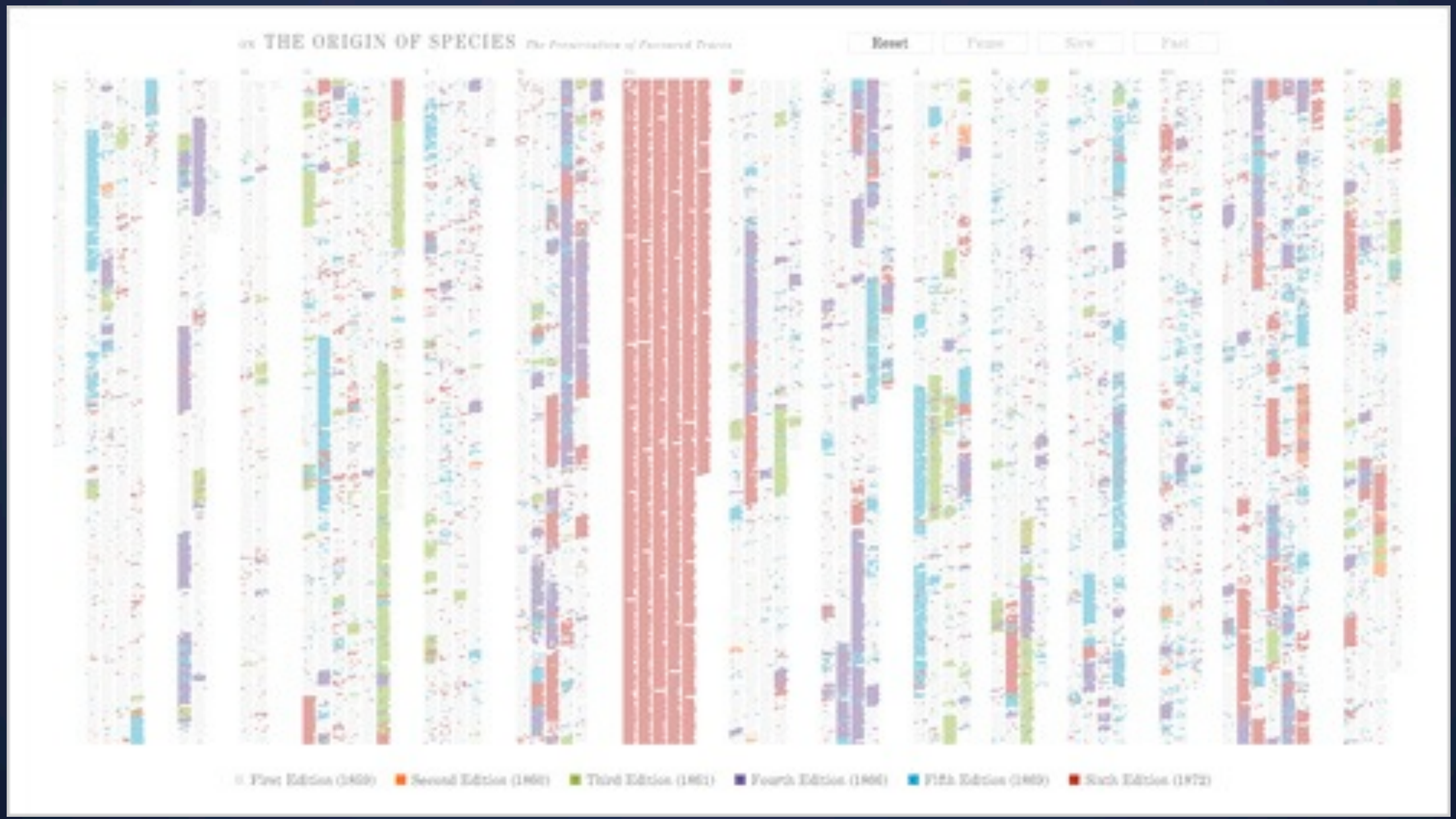


Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces (Viegas)

Evolving Documents

Document-Level Visualizations

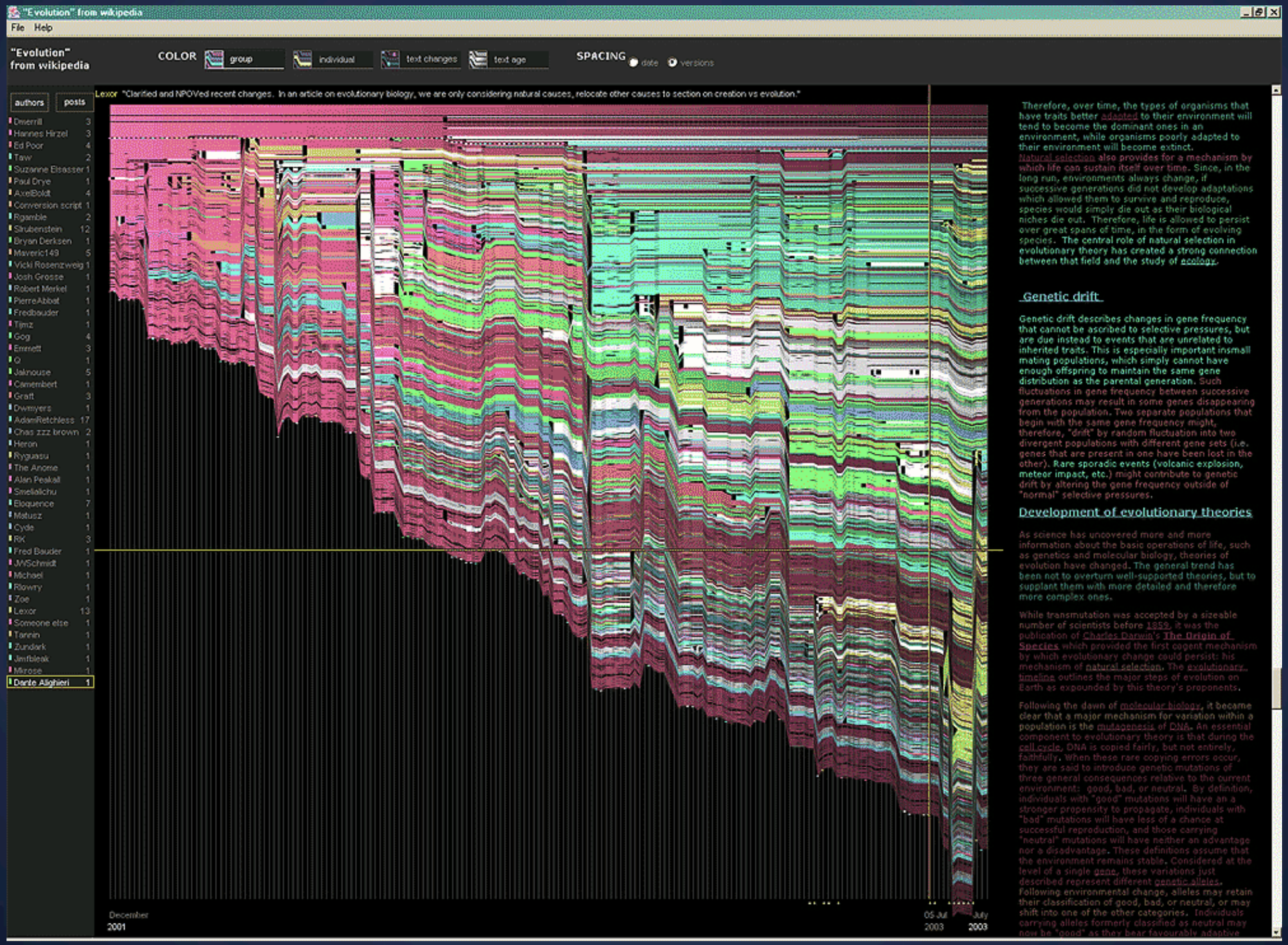
On the Origin of Species: The Preservation of Favoured Traces



http://benfry.com/traces/

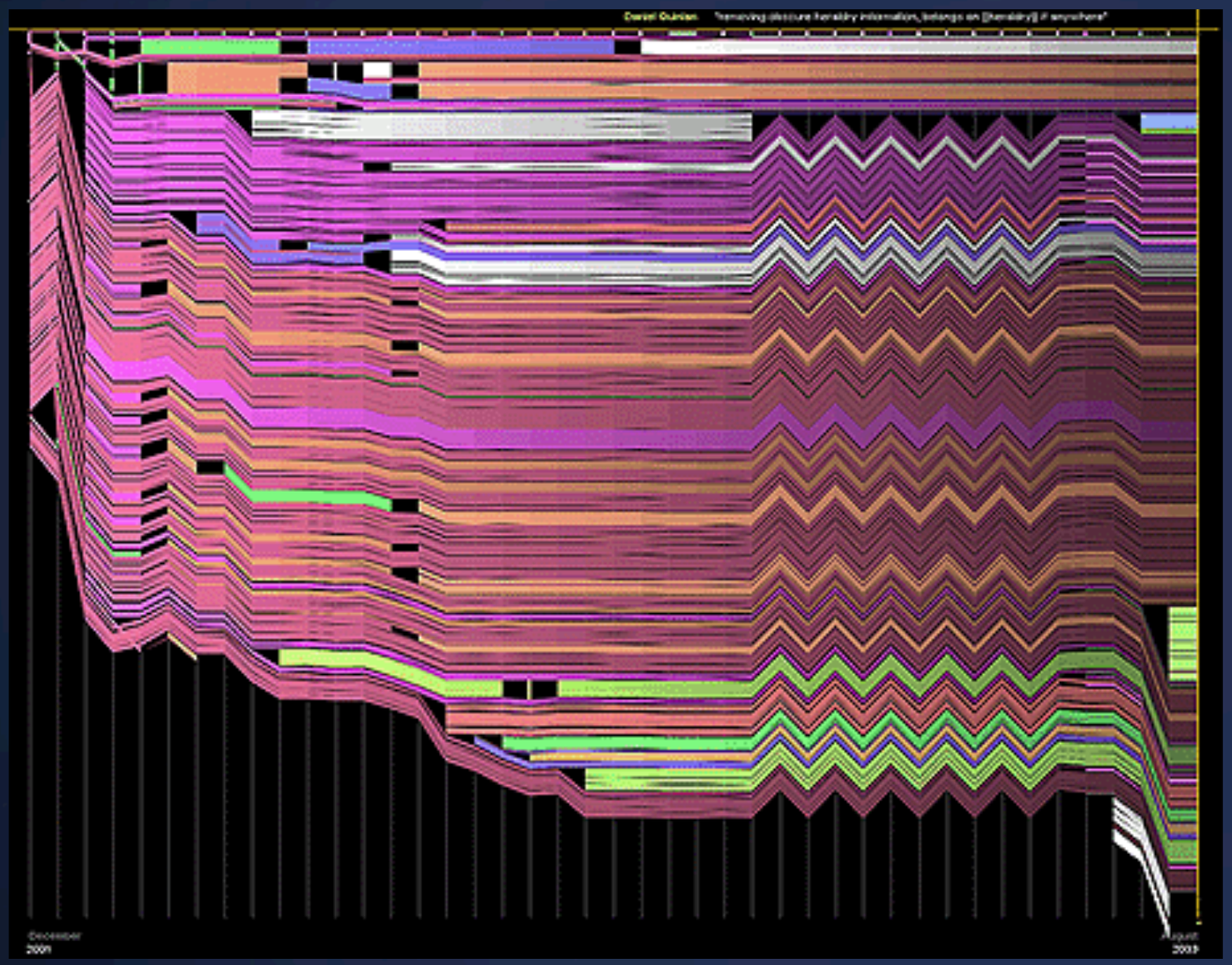
Evolution of a document

Visualizing additions / deletions in Wikipedia articles



Evolution of a document

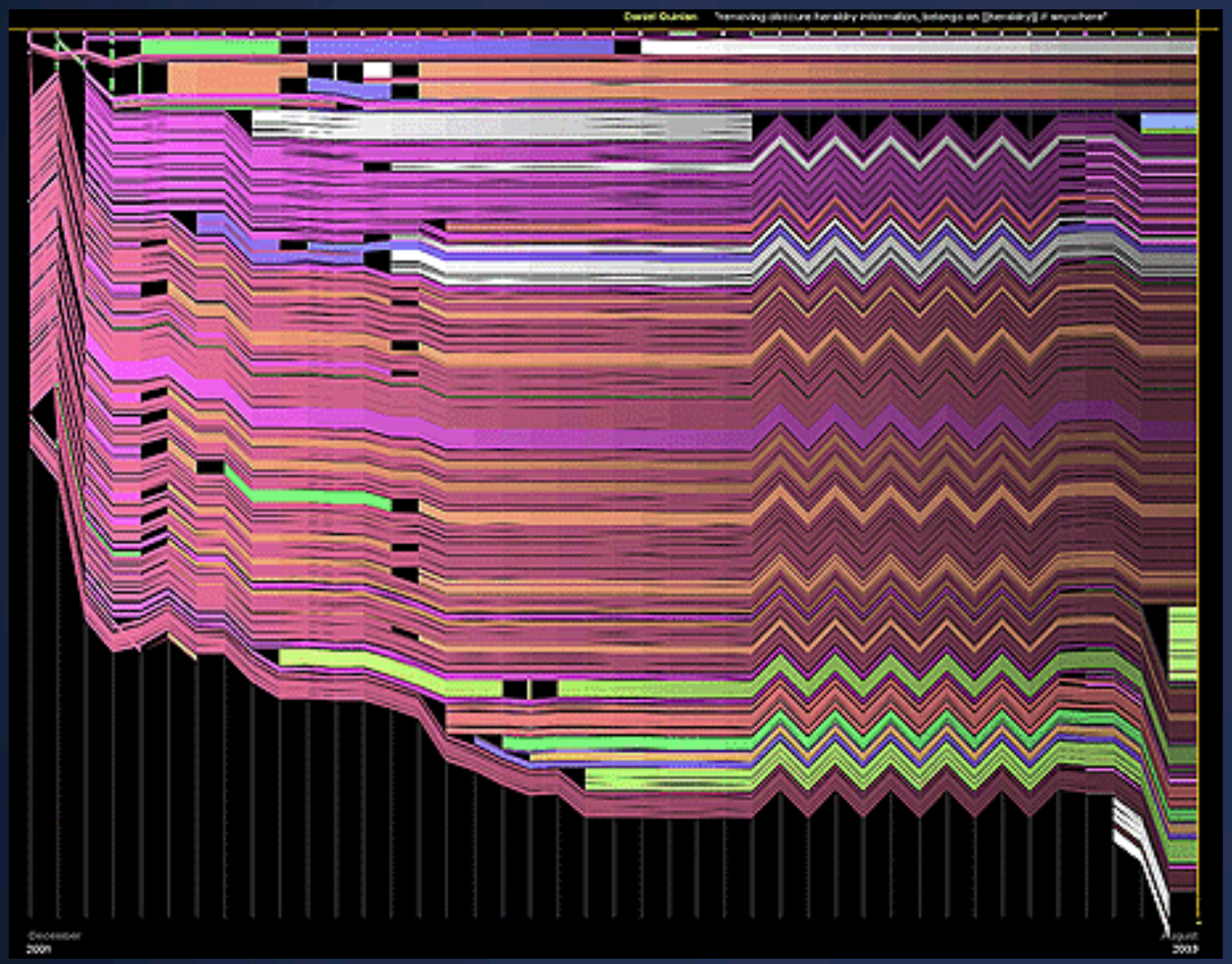
Visualizing additions / deletions in Wikipedia articles



Chocolate Wikipedia page

Evolution of a document

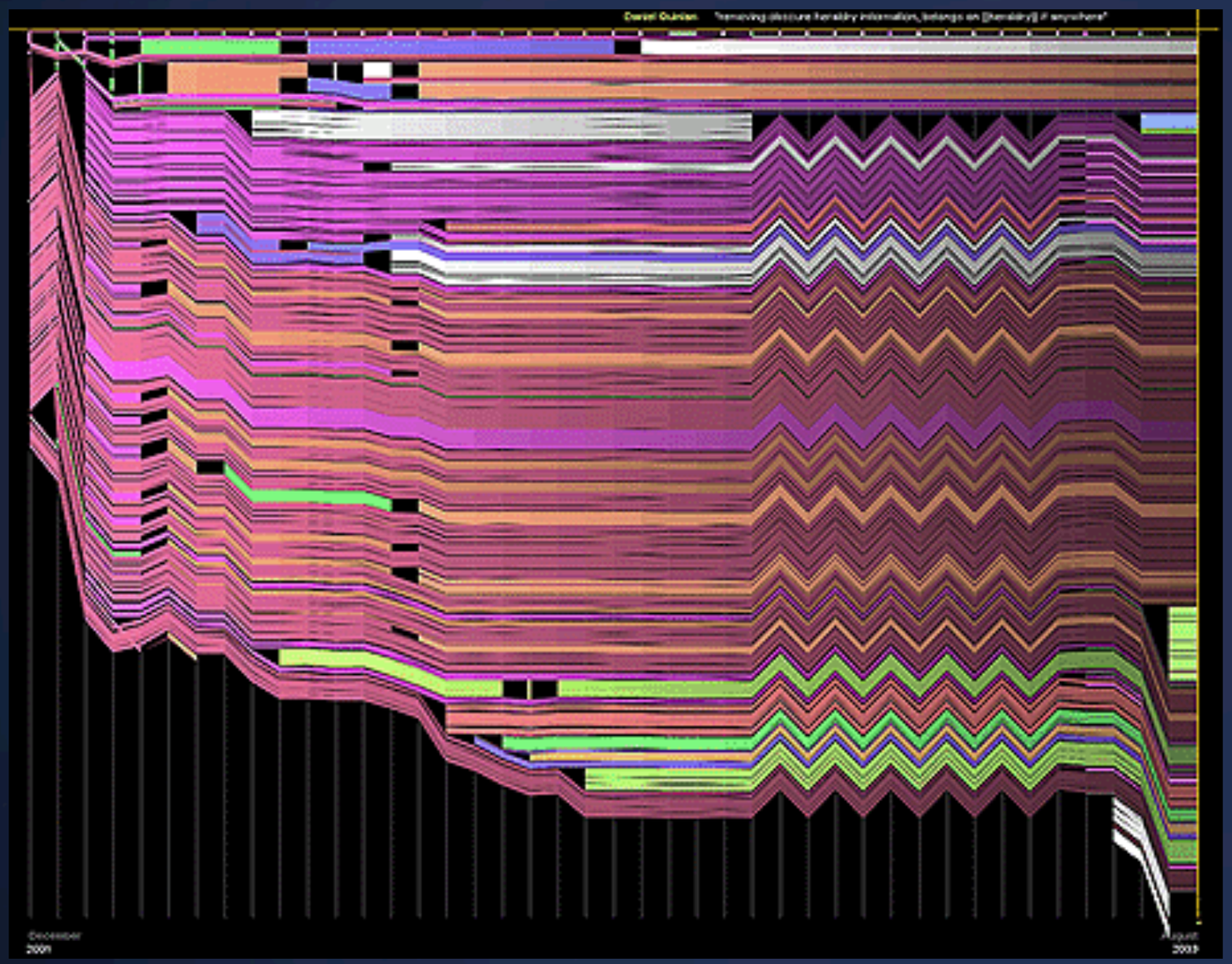
Visualizing additions / deletions in Wikipedia articles



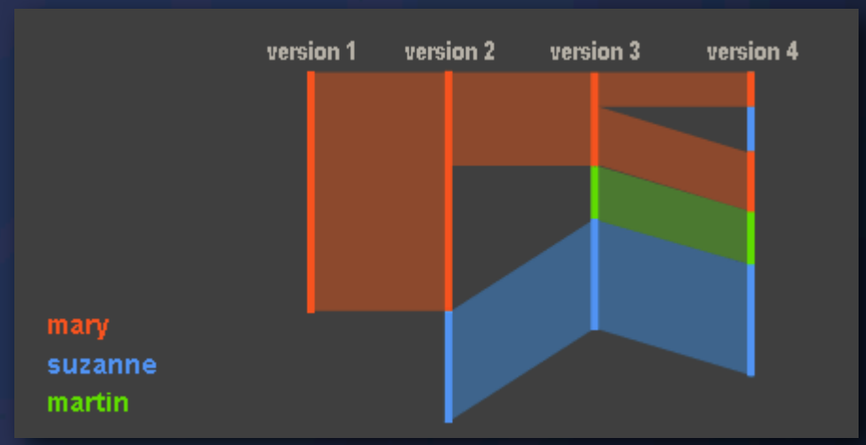
Chocolate Wikipedia page

Evolution of a document

Visualizing additions / deletions in Wikipedia articles



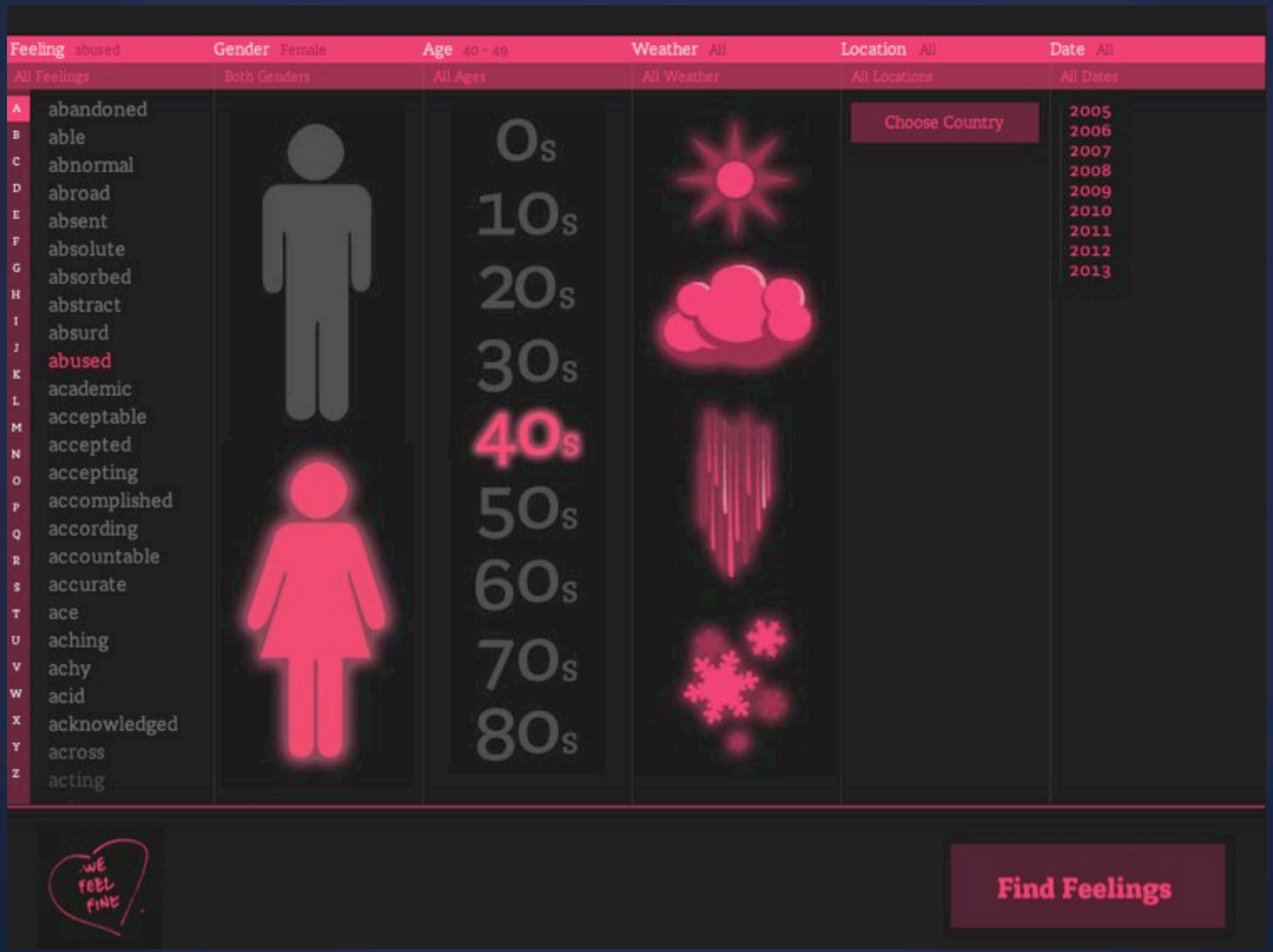
Chocolate Wikipedia page



Strategies

Introduce Faceted Search

Faceted search

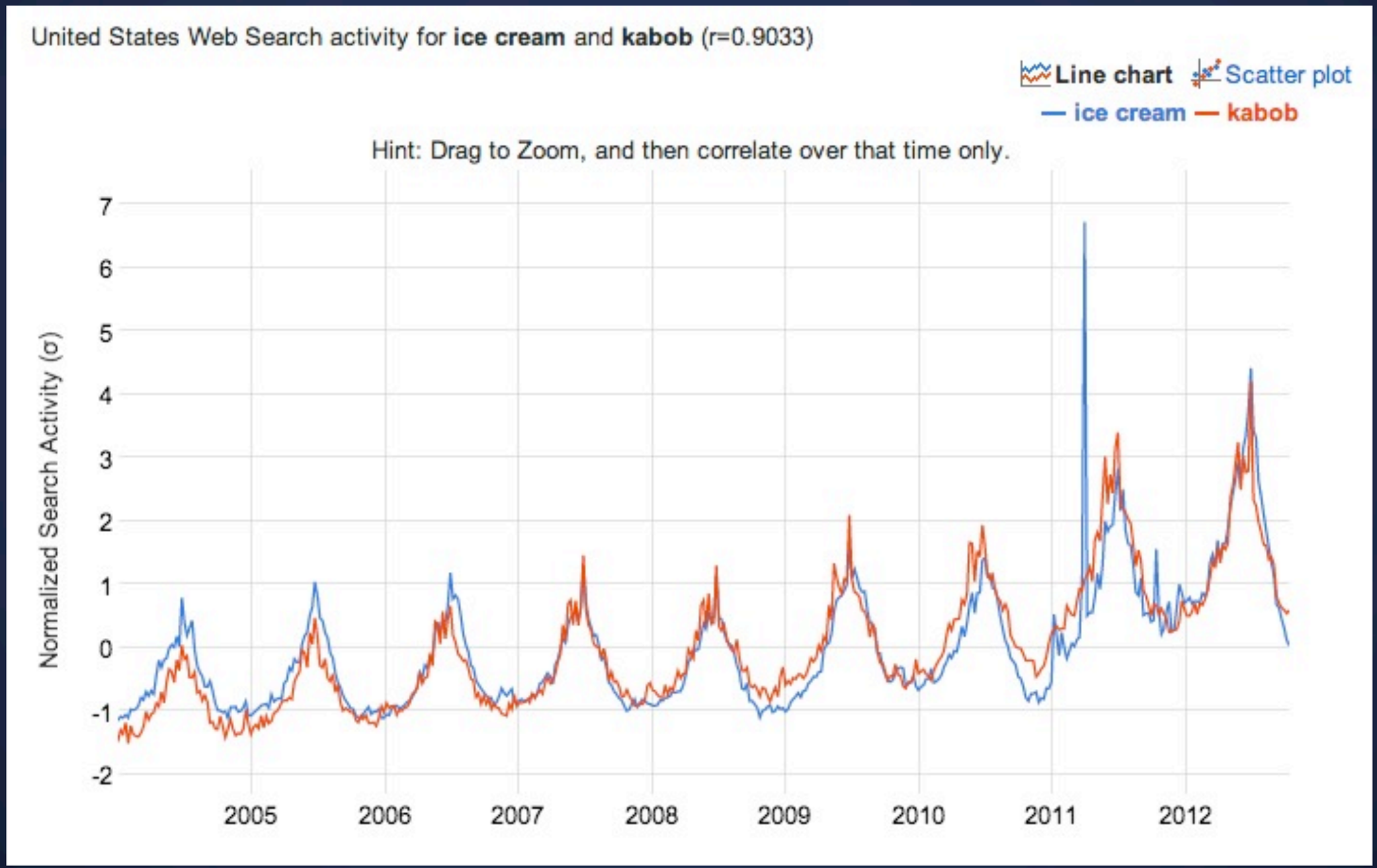


We Feel Fine

Faceted search

Add Time Dimension

Google Trends



Recap

Lab Thursday

Concept Critique

Next Lecture

Networks: Graphs and Hierarchy