

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

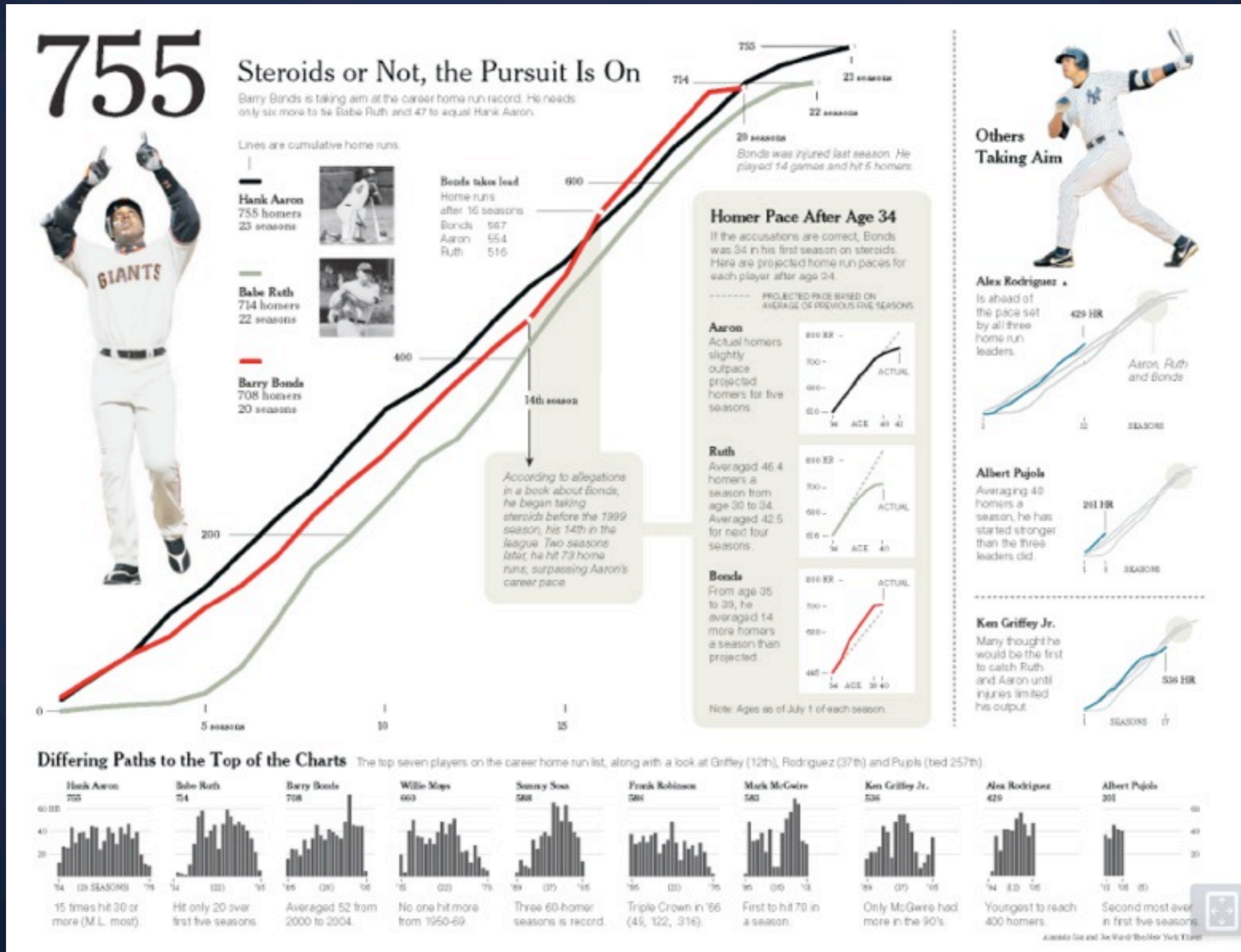
tell the story

Interactivity

engage the story

Visual Structure

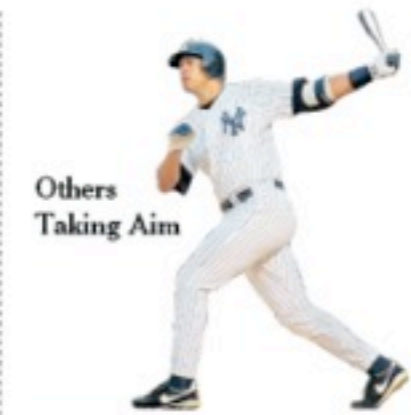
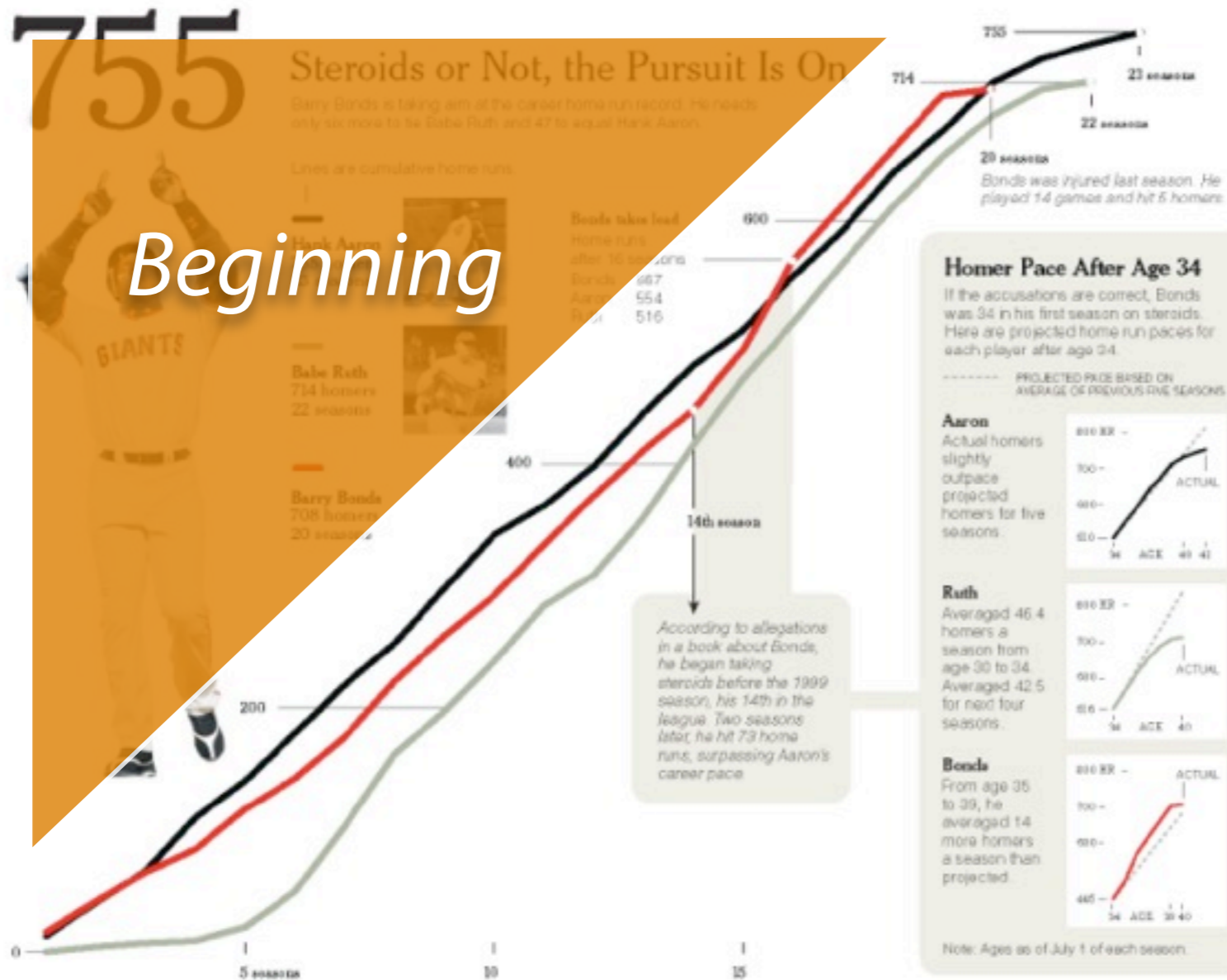
NY Times: Steroids or Not, the Pursuit Is On



Visual Structure

NY Times: Steroids or Not, the Pursuit Is On

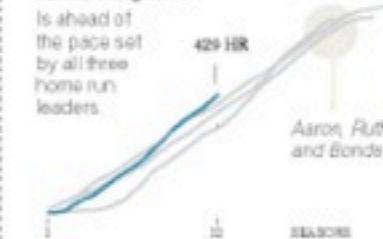
Beginning



Others Taking Aim

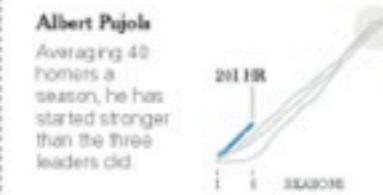
Alex Rodriguez

Is ahead of the pace set by all three home run leaders.



Albert Pujols

Averaging 40 homers a season, he has started stronger than the three leaders did.

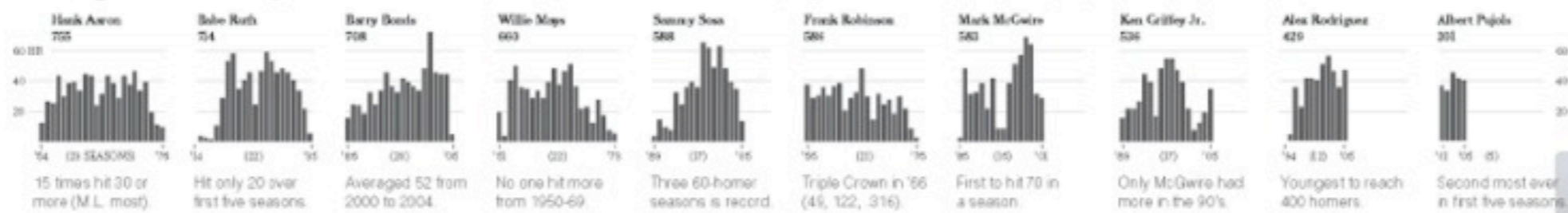


Ken Griffey Jr.

Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

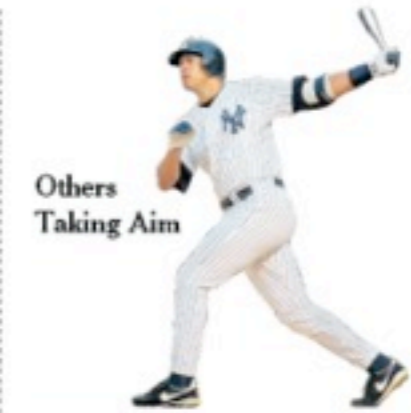
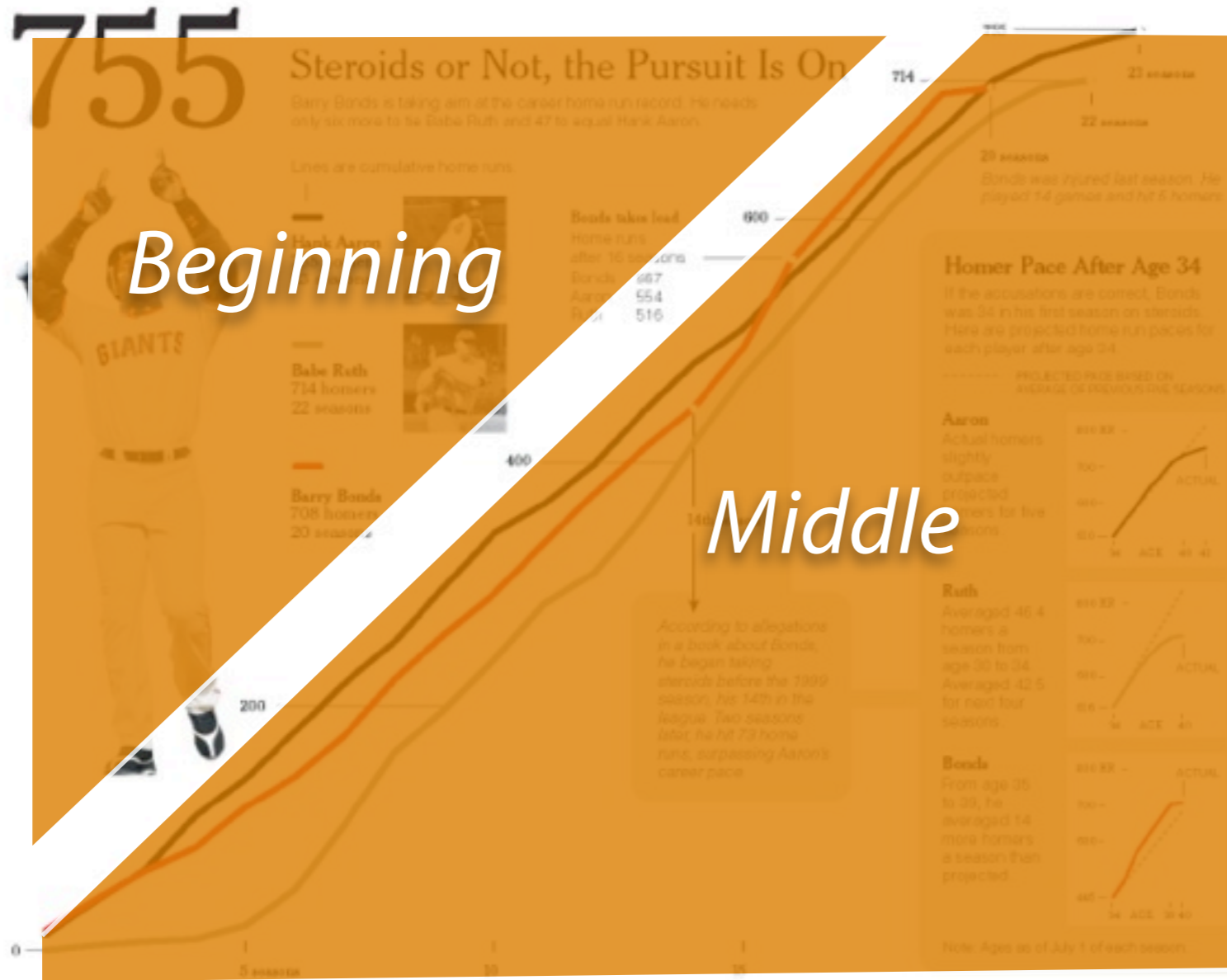


Differing Paths to the Top of the Charts The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (ted 257th).

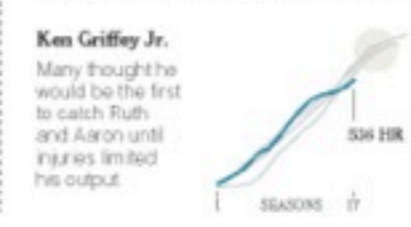
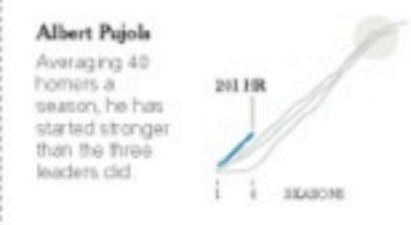
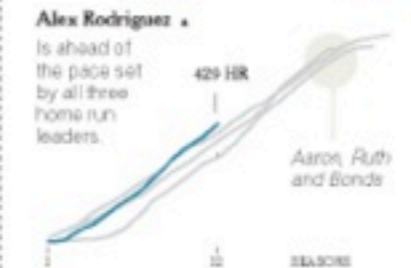


Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



Others Taking Aim



Differing Paths to the Top of the Charts The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (ted 257th).

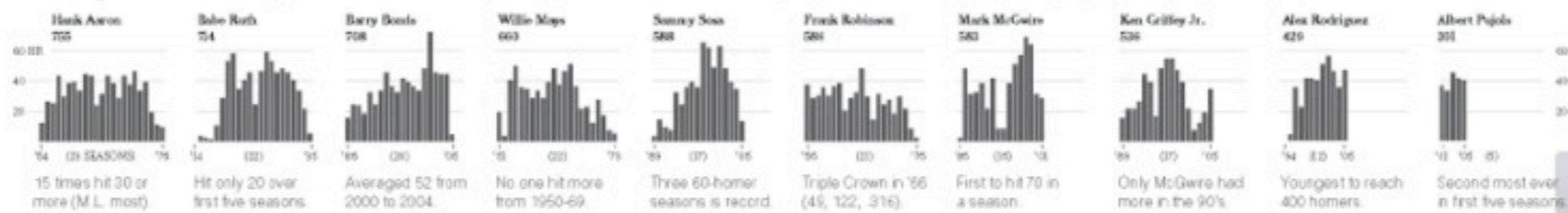
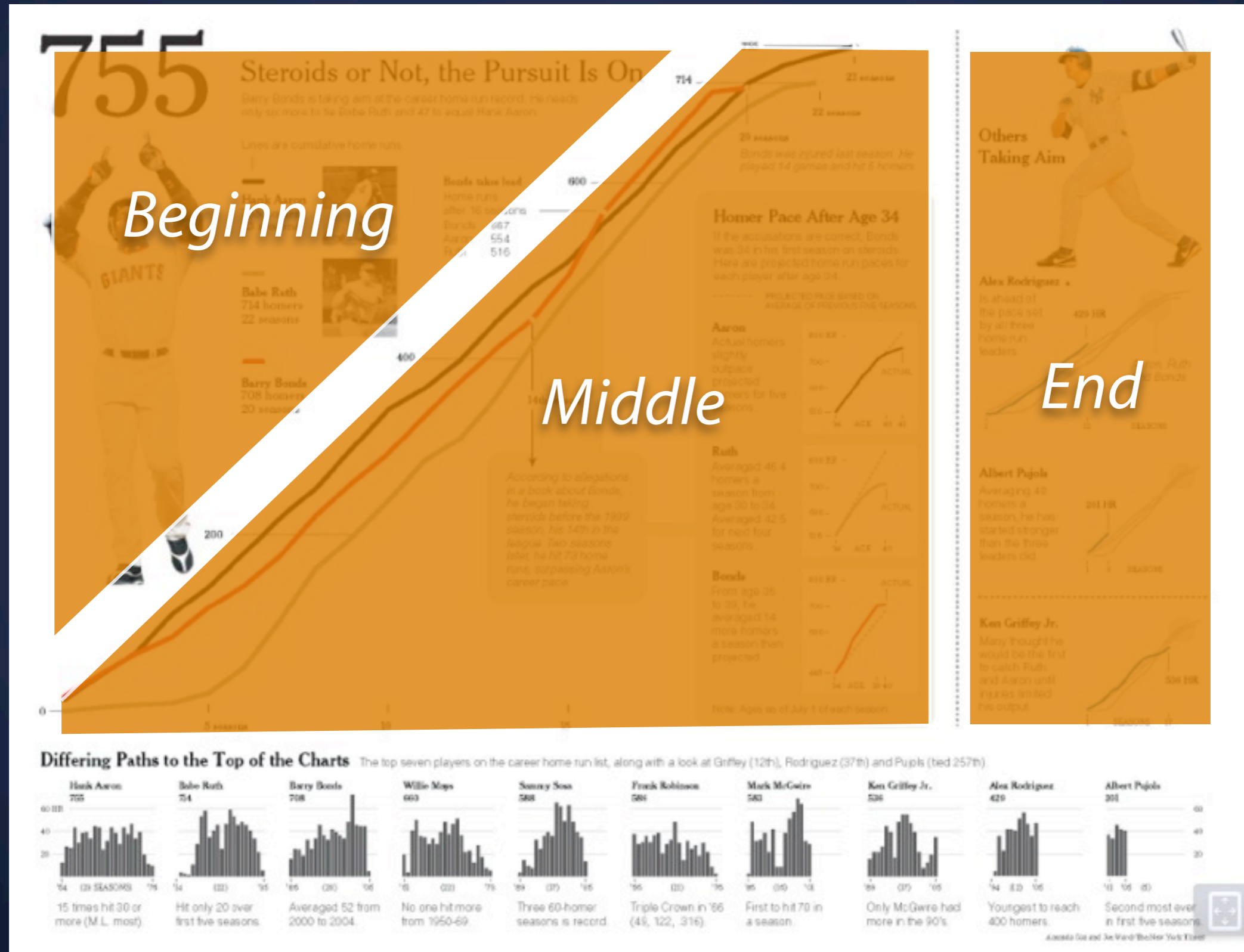


Illustration by David J. Phillip for The New York Times

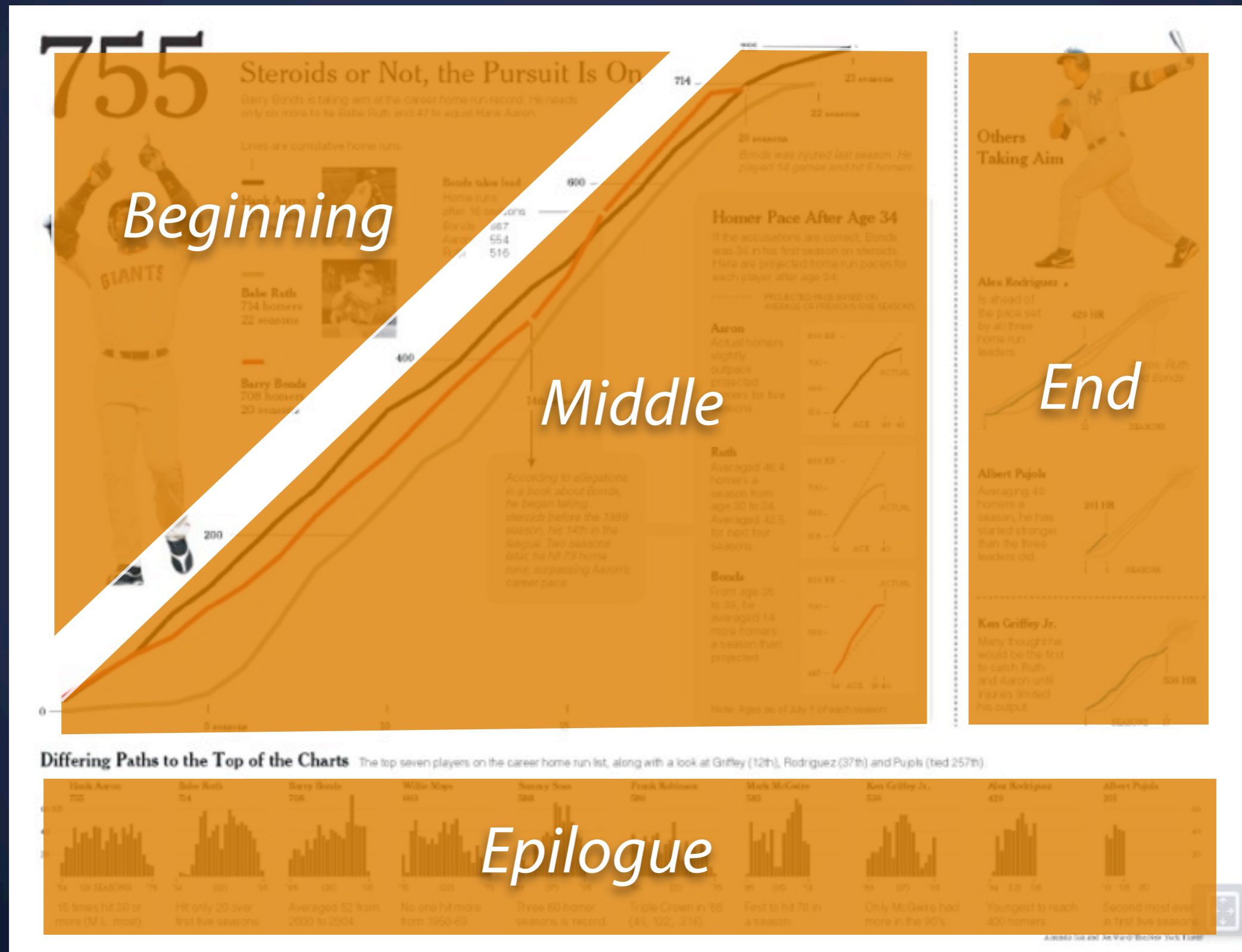
Visual Structure

NY Times: Steroids or Not, the Pursuit Is On



Visual Structure

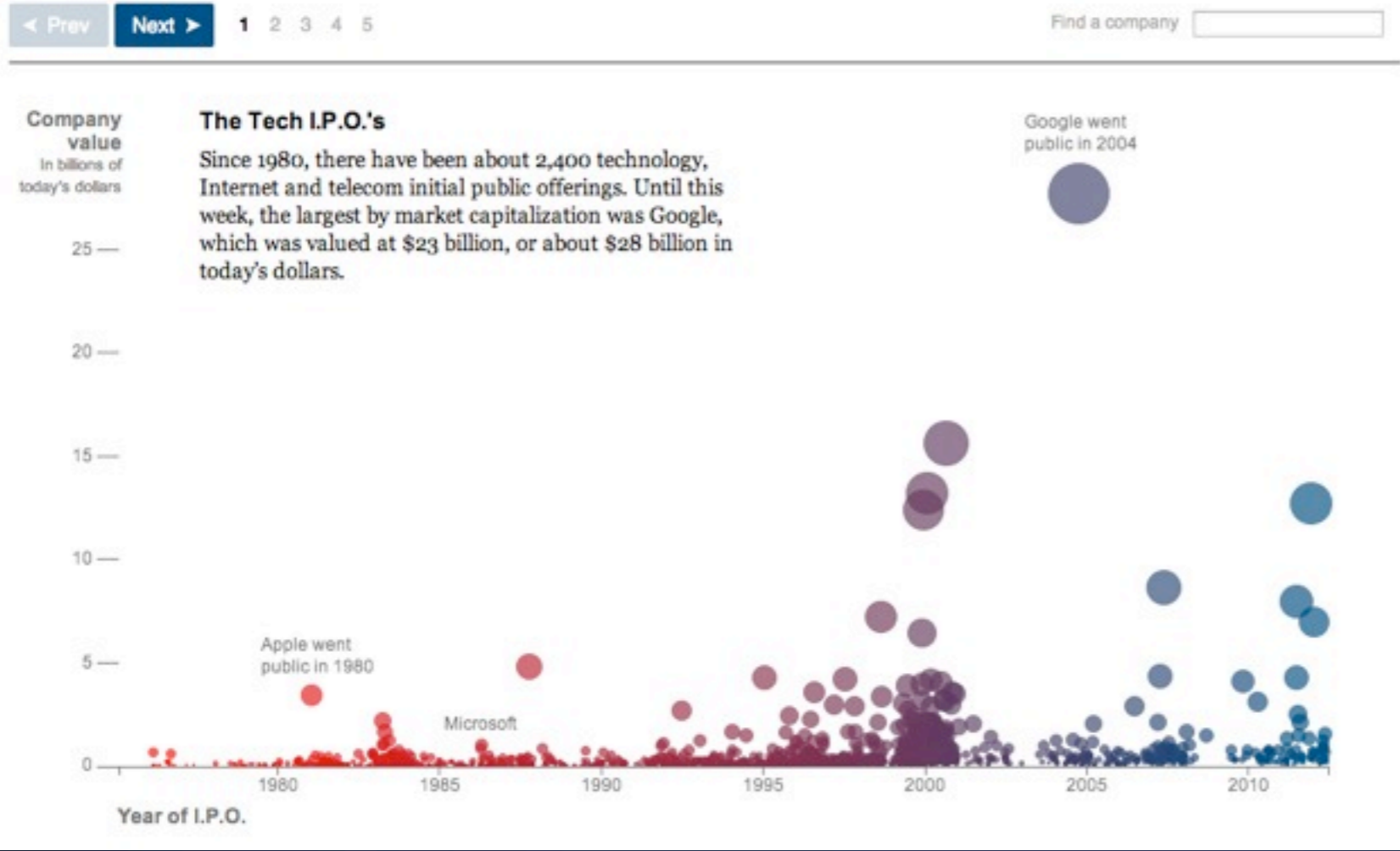
NY Times: Steroids or Not, the Pursuit Is On



Messaging

Tell the story; not JUST with visuals

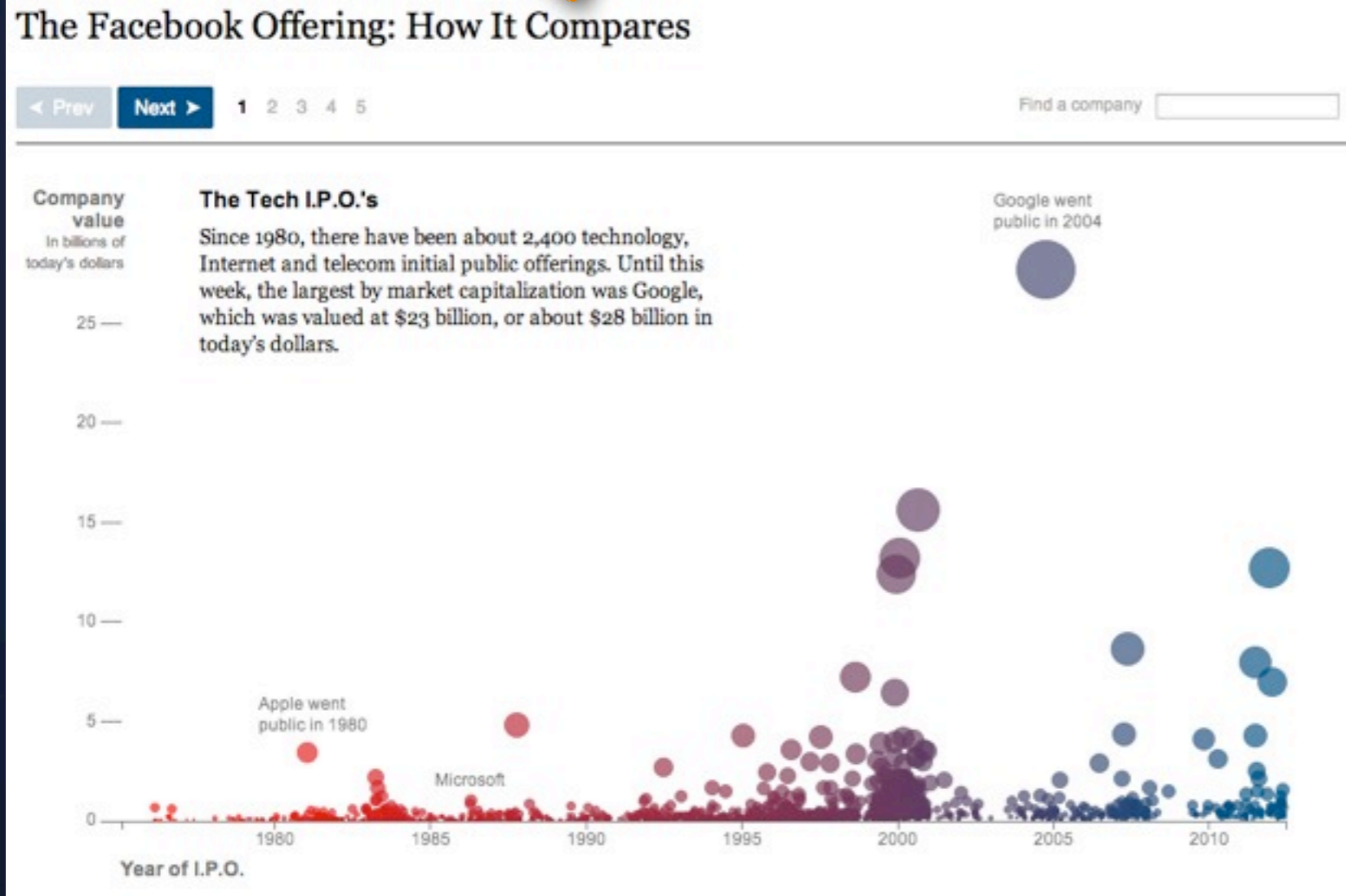
The Facebook Offering: How It Compares



Messaging

Tell the story; not JUST with visuals

Headline

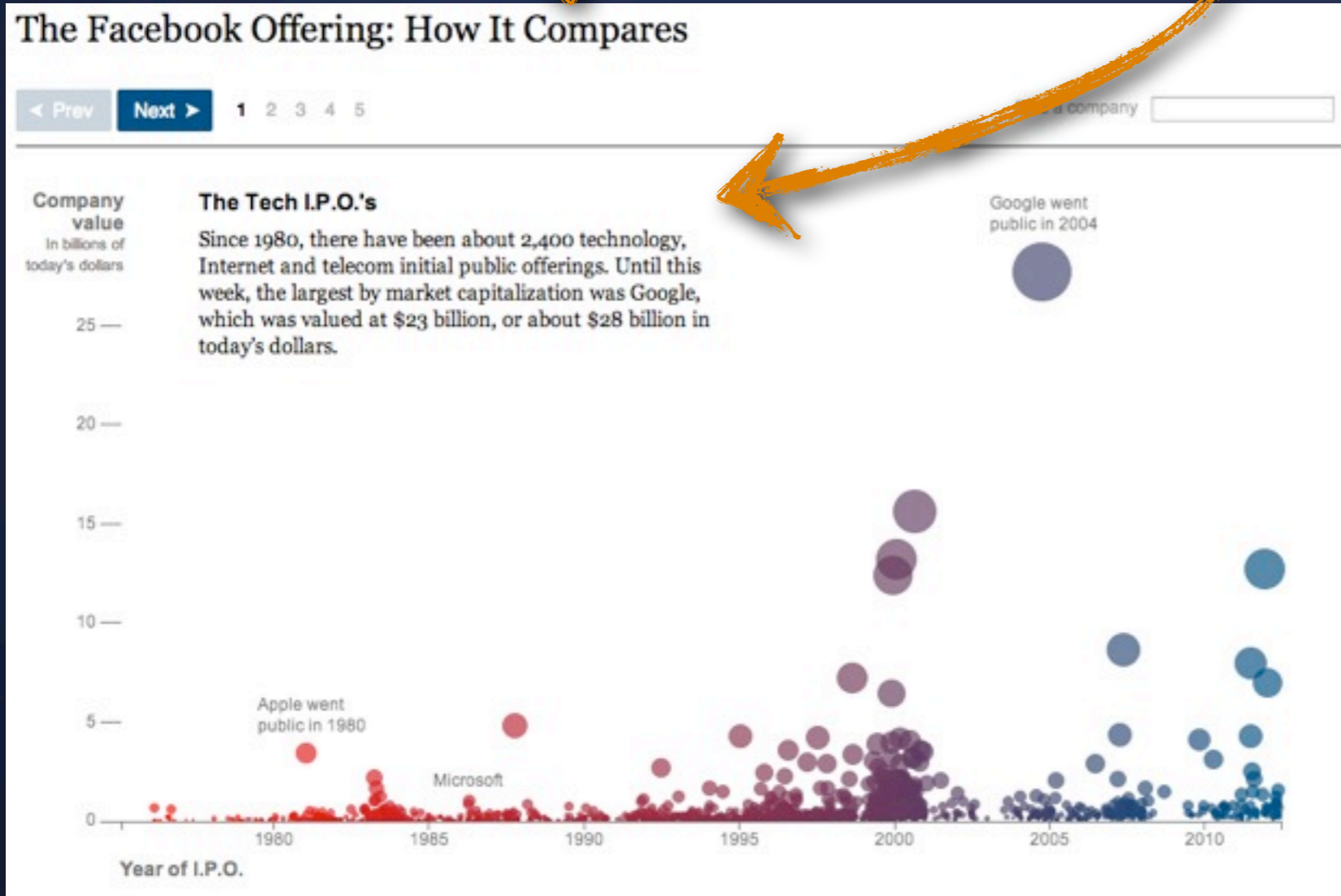


Messaging

Tell the story; not JUST with visuals

Headline

Caption



Messaging

Tell the story; not JUST with visuals

Headline

Caption

The Facebook Offering: How It Compares

< Prev Next >

1 2 3 4 5

Company value
In billions of
today's dollars

The Tech I.P.O.'s

Since 1980, there have been about 2,400 technology, Internet and telecom initial public offerings. Until this week, the largest by market capitalization was Google, which was valued at \$23 billion, or about \$28 billion in today's dollars.

Google went public in 2004

Apple went public in 1980

25 —
20 —
15 —
10 —
5 —
0 —

Year of I.P.O.

1980 1985 1990

2005 2010

Annotation



Messaging

Tell the story; not JUST with visuals

Headline

Caption

The Facebook Offering: How It Compares

< Prev Next > 1 2 3 4 5

Company value
In billions of
today's dollars

The Tech I.P.O.'s

Since 1980, there have been about 2,400 technology, Internet and telecom initial public offerings. Until this week, the largest by market capitalization was Google, which was valued at \$23 billion, or about \$28 billion in today's dollars.

Google went
public in 2004

Apple went
public in 1980

Year of I.P.O.

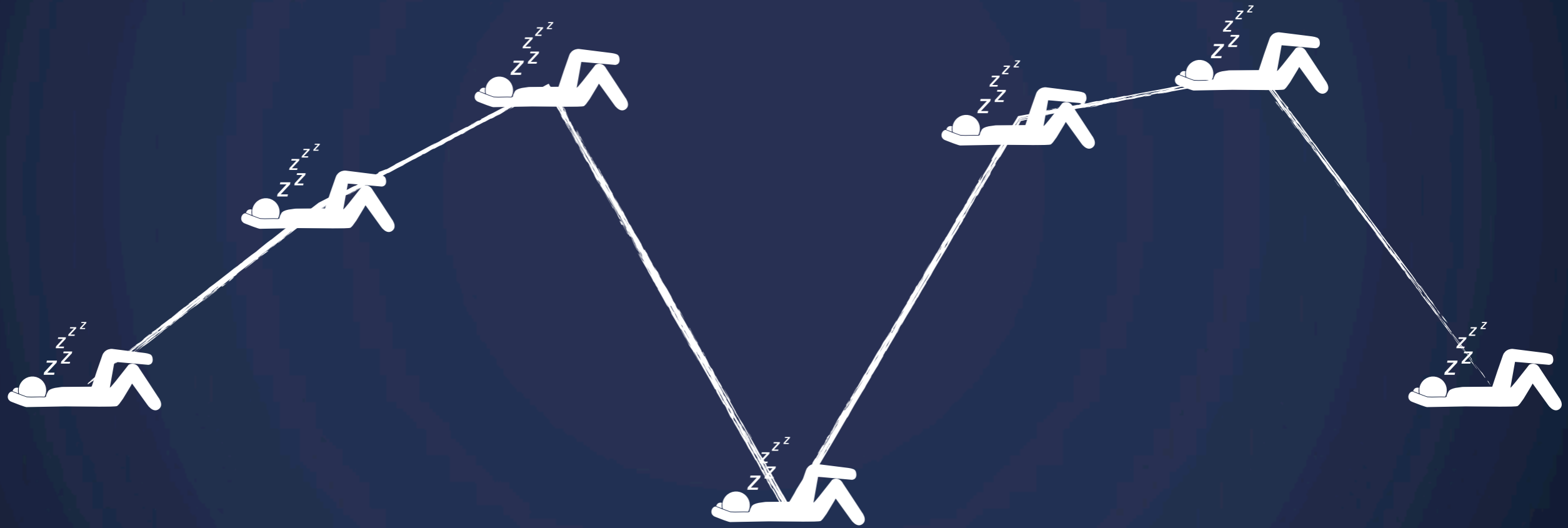
Annotation

weave text *into*
the graphic!



Icons

Be **very** careful!



Icons

Be **very** careful!



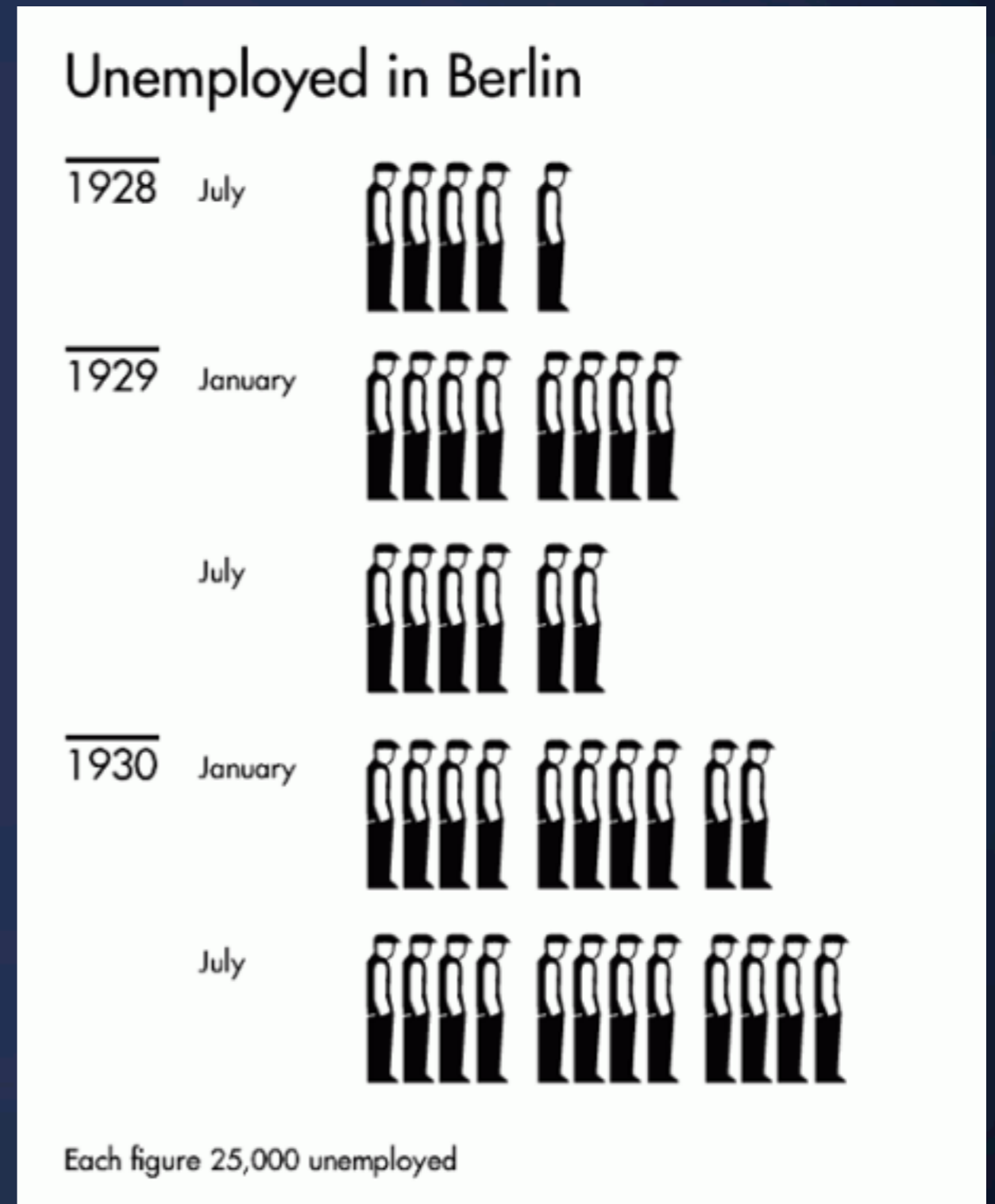
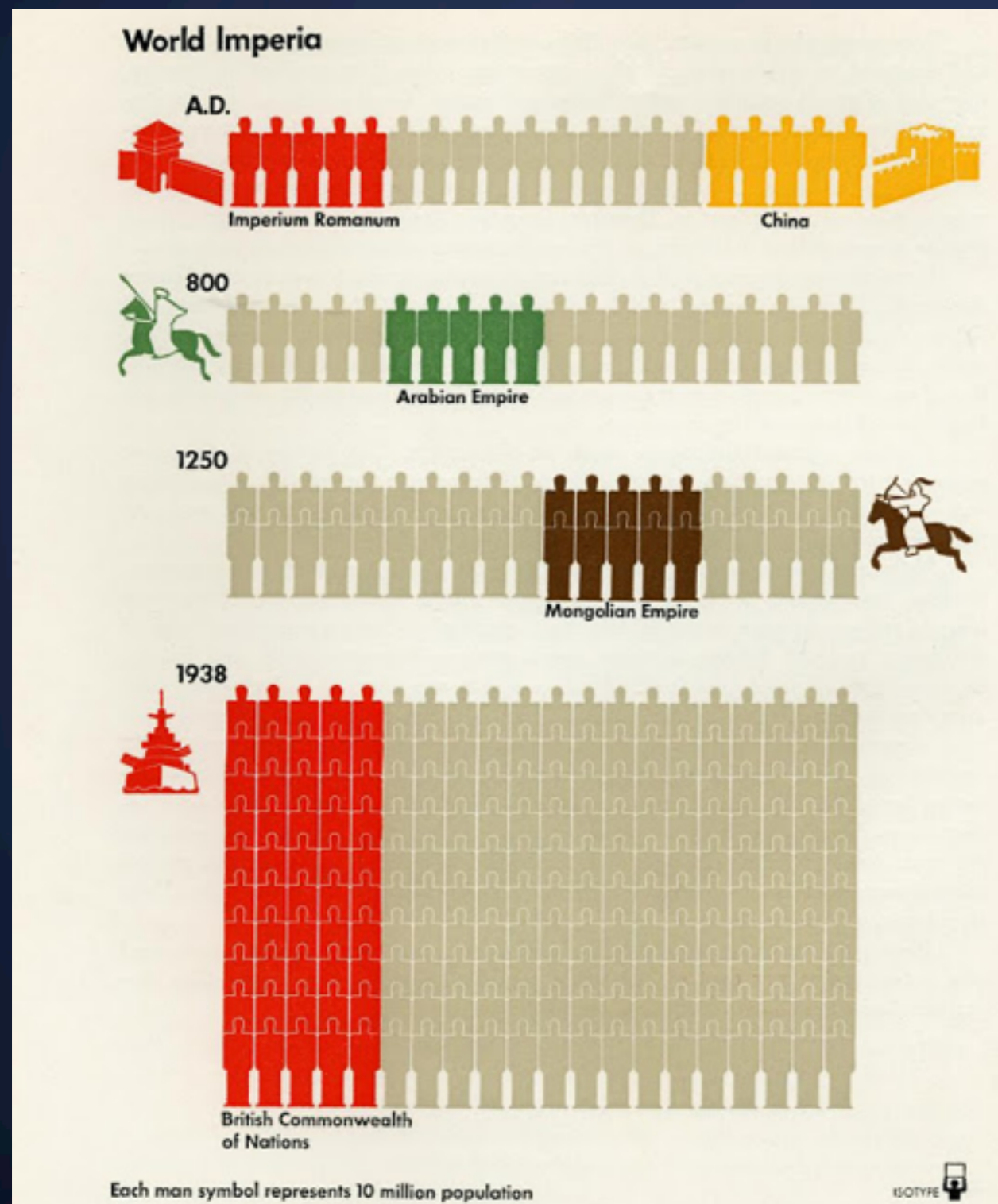
Sleep

hours per night



Icons

Pictograms / Unit Chart: Otto Neurath / Gerd Arntz



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

Wordle

Context



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?
#1247, @berkeleyischool
- Entities?
San Francisco, O'Connor, U.S.A.

Stemming

Group together different forms of a word

- Porter stemmer?
*visualization(s),
visualize(s), visually*
→ **visual**
- Lemmatization? *goes, went, gone*
→ **go**

Document-Level

Visualizations

Word Frequency Count

wordcount.org

WORDCOUNT

◀ PREVIOUS WORD

▶ NEXT WORD

the of and to ain that it is was i for on you he be with as by have re this or but in his es from the which us we d are in by e go on st in the that in d the y the e

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

CURRENT WORD

FIND WORD: ▶ BY RANK: ▶ REQUESTED WORD: **THE** RANK: 1

86800 WORDS IN ARCHIVE

[ABOUT WORDCOUNT](#)

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

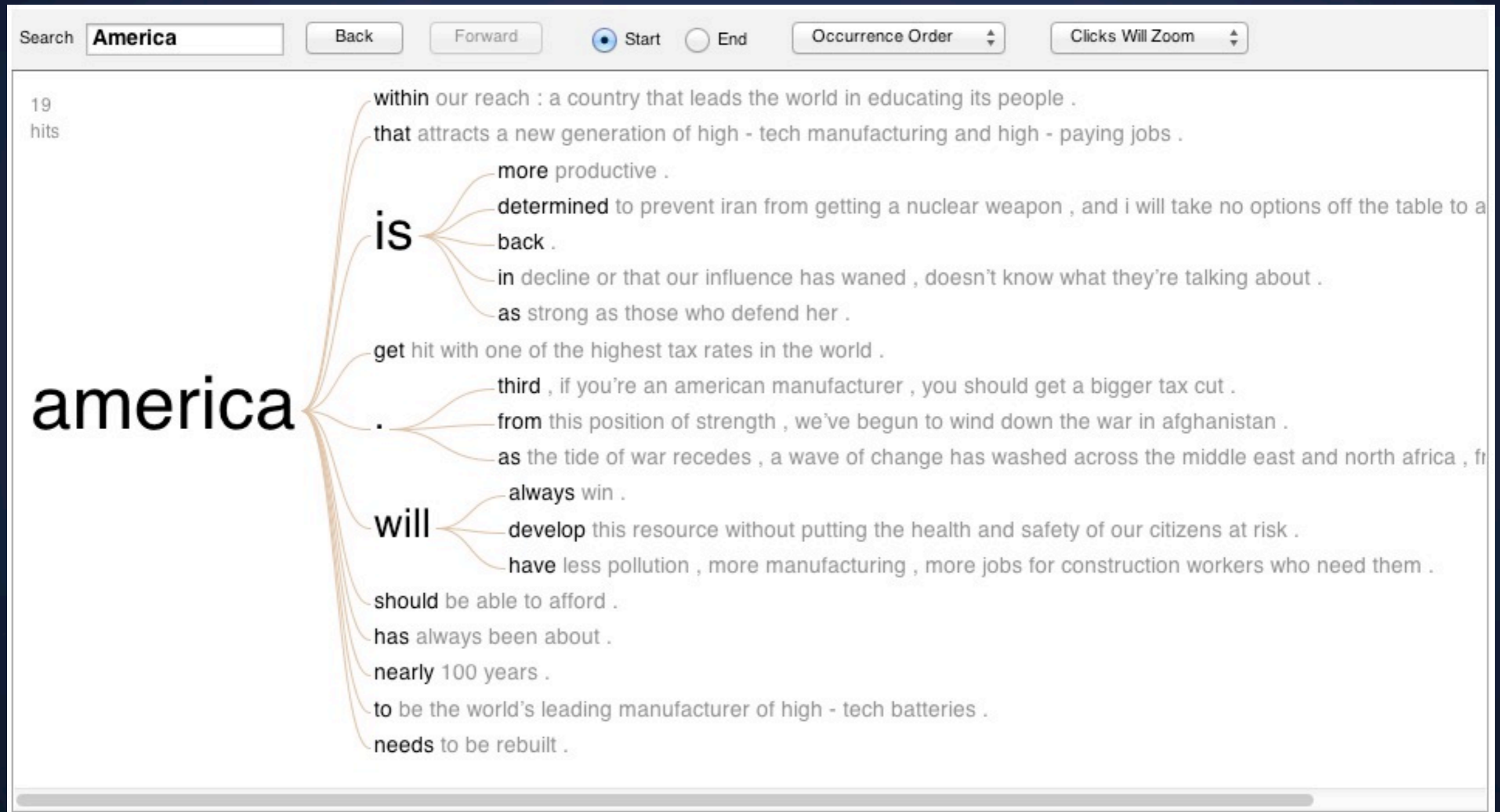
Often favors frequent (TF) or rare (IDF) terms

**Ignores additional
information**

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

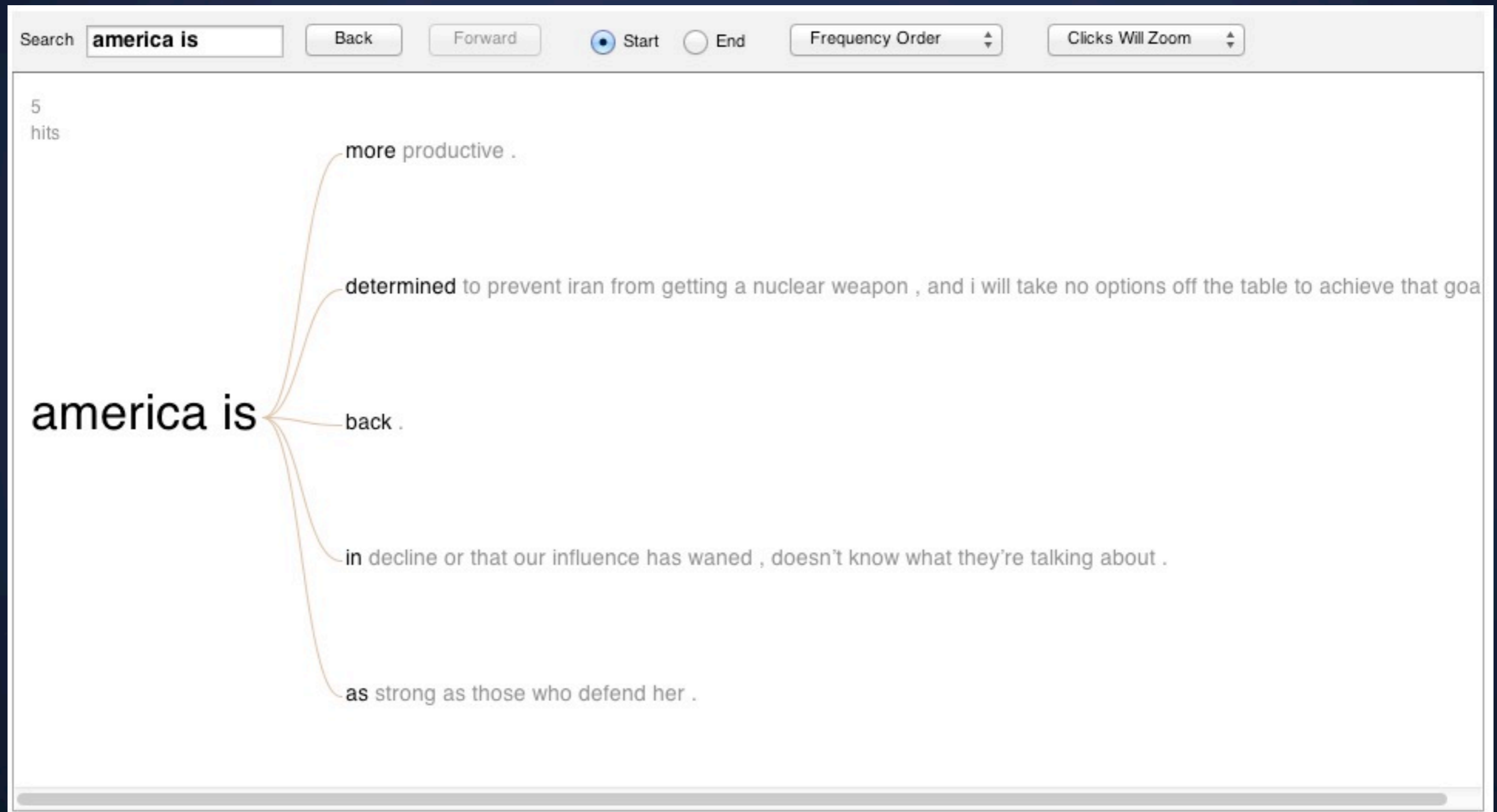
Word Tree

Considering n-grams (SOTU 2012)



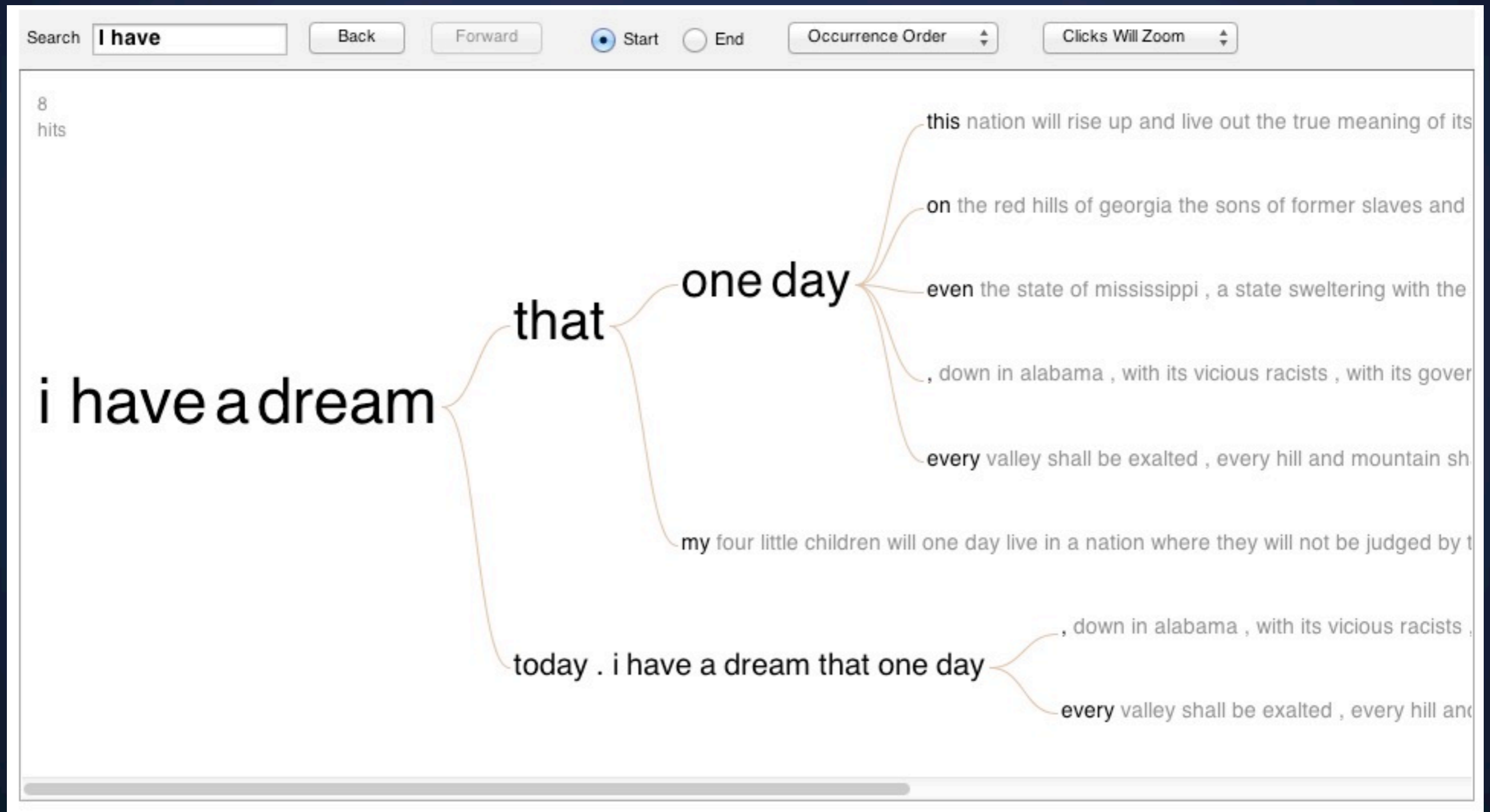
Word Tree

Considering n-grams (SOTU 2012)



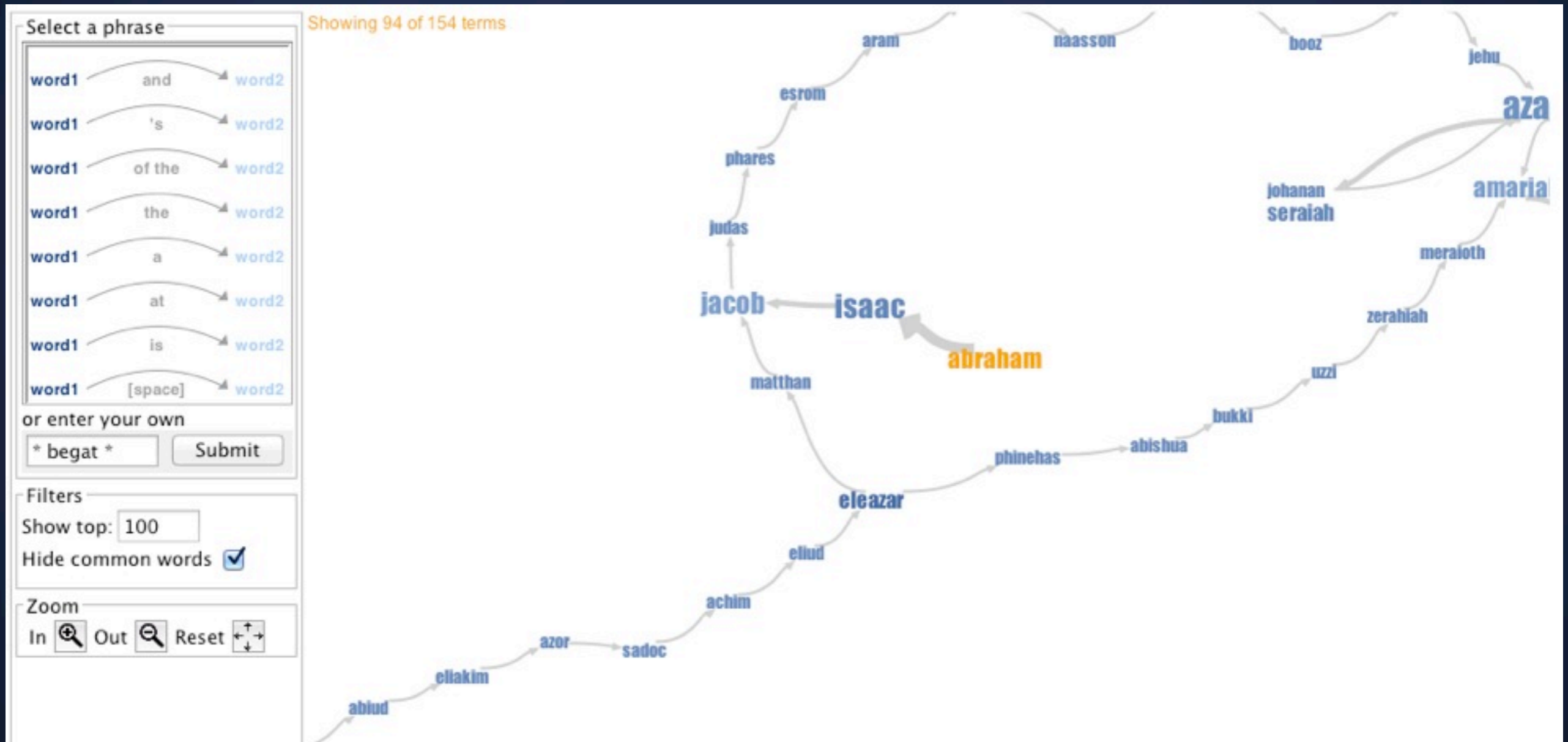
Word Tree

Visualizing rhetorical devices / repetition



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

Advertise 26 of 41


MARGARET THATCHER, 1925-2013
'Iron Lady' Who Set Britain on a New Course



Agence France-Presse — Getty Images

Margaret Thatcher, the first female prime minister of **Britain**, held the office for 11 years. [More Photos »](#)

By JOSEPH R. GREGORY
Published: April 8, 2013 | [880 Comments](#)



Google Chrome

In-page search: Britain

MARGARET THATCHER, 1925-2013

'Iron Lady' Who Set Britain on a New Course



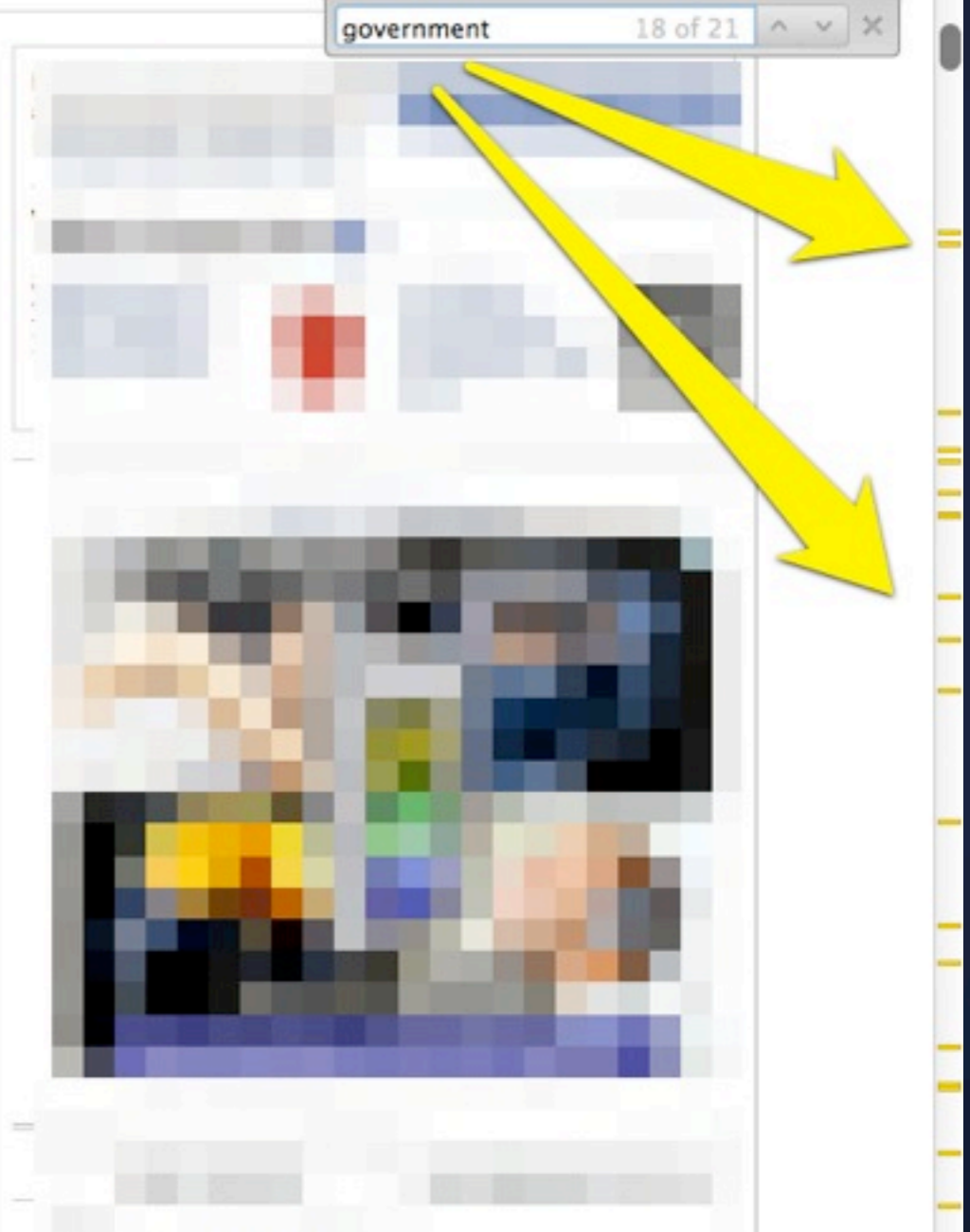
Agence France-Presse — Getty Images

Margaret Thatcher, the first female prime minister of Britain, held the office for 11 years. [More Photos »](#)

By JOSEPH R. GREGORY
Published: April 8, 2013 | [880 Comments](#)

[Margaret Thatcher](#), the "Iron Lady" of British politics, who set her

government 18 of 21



Collection-Level

Visualizations

George W. Bush's speeches

State of the Union / NY Times

The 2007 State of the Union Address

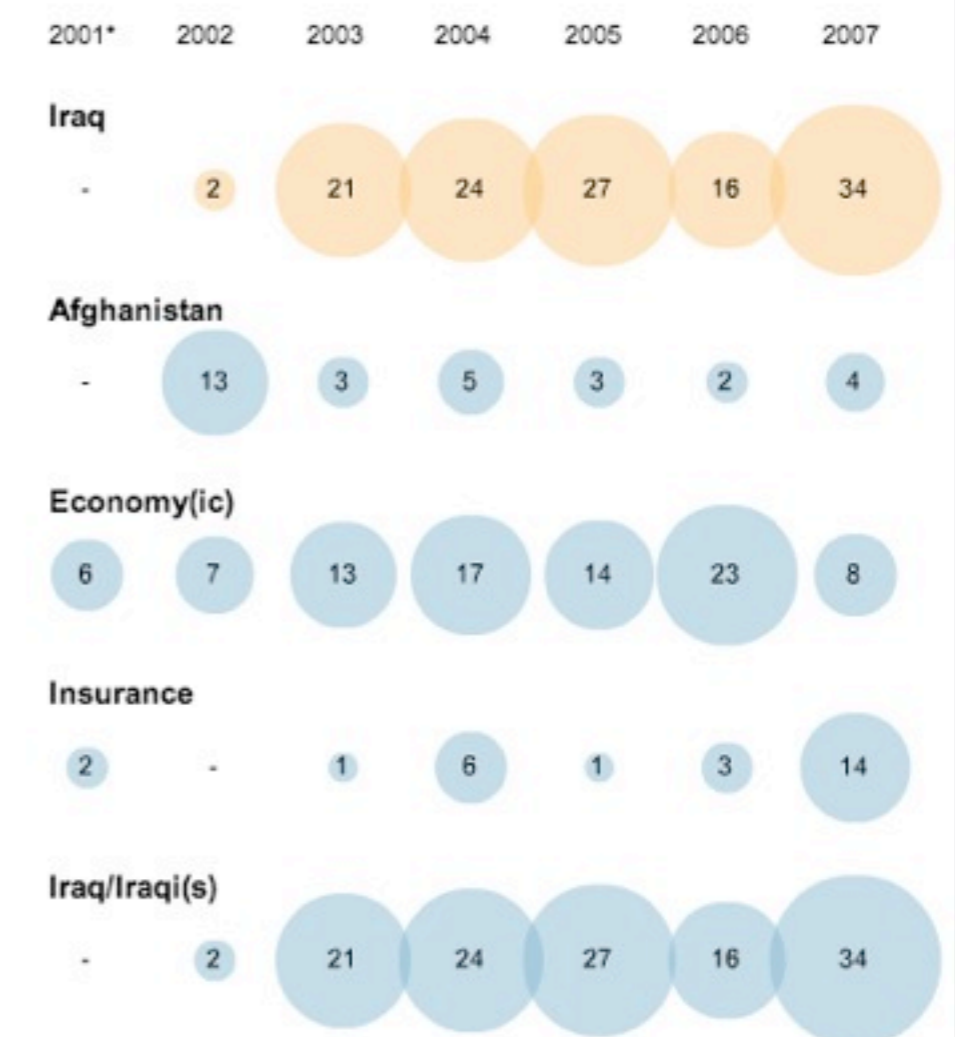
Over the years, President Bush's State of the Union address has averaged almost 5,000 words each, meaning the the President has delivered over 34,000 words. Some words appear frequently while others appear only sporadically. Use the tools below to analyze what Mr. Bush has said.

 or choose a word here.

Use of the phrase "Iraq" in past State of the Union Addresses



Compared with other words

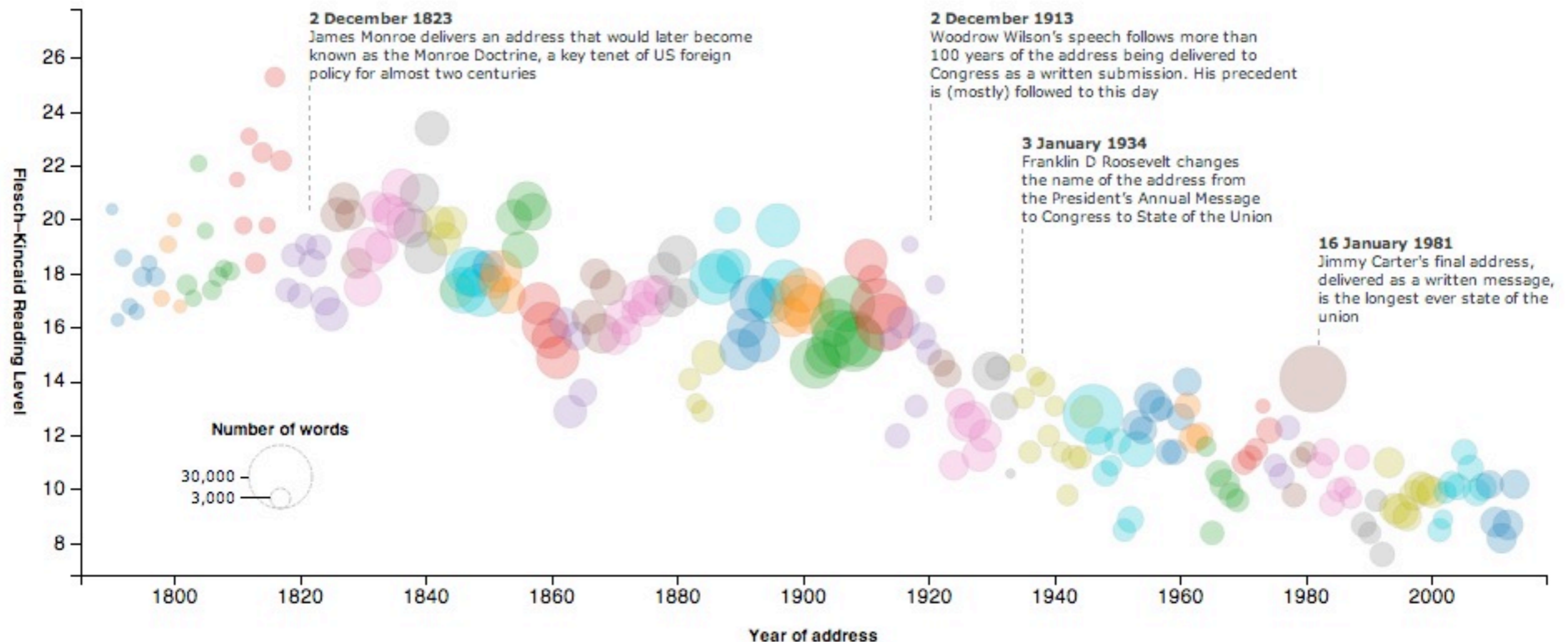


Reading Level

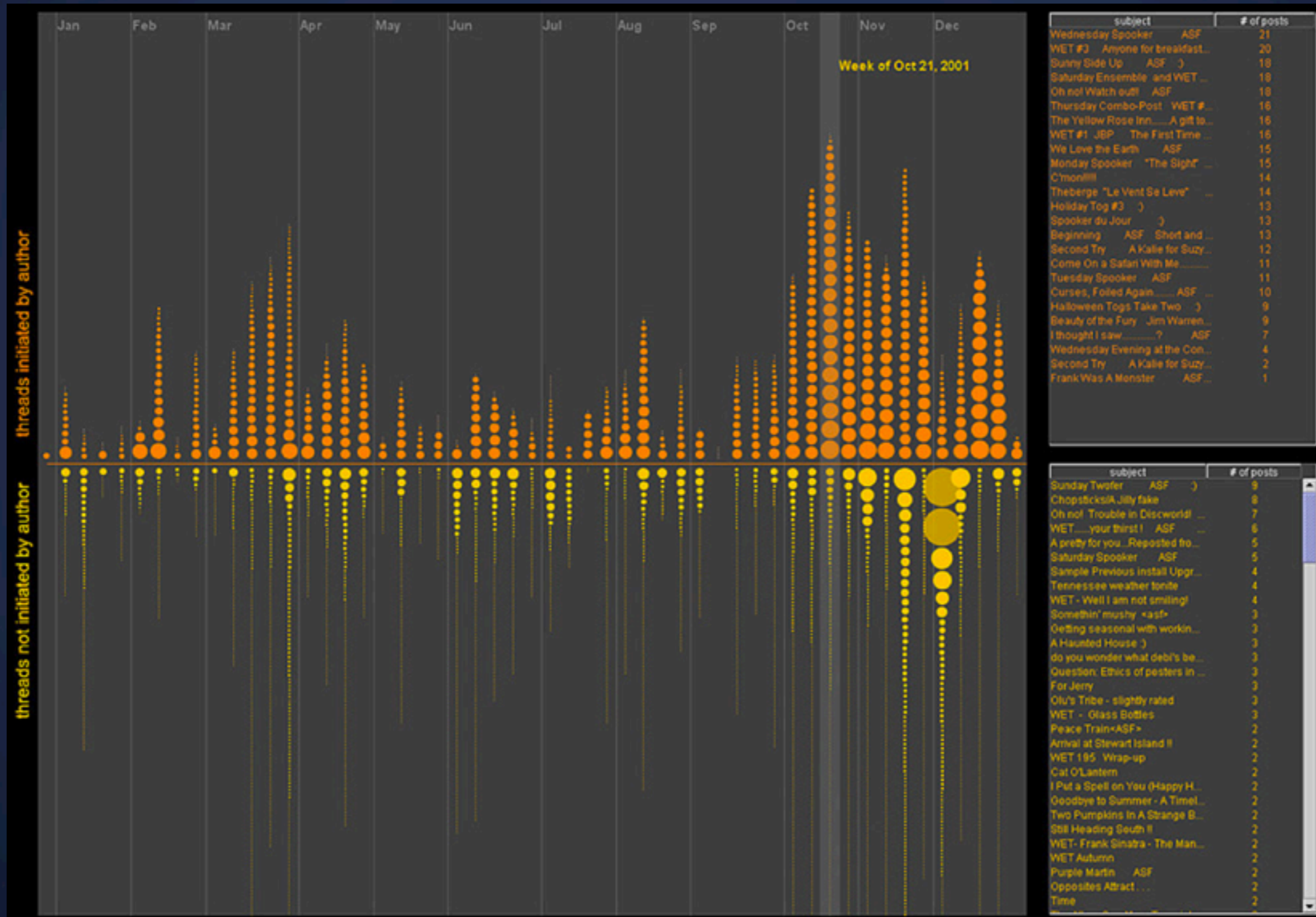
State of the union speeches over time

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



Evolving Documents

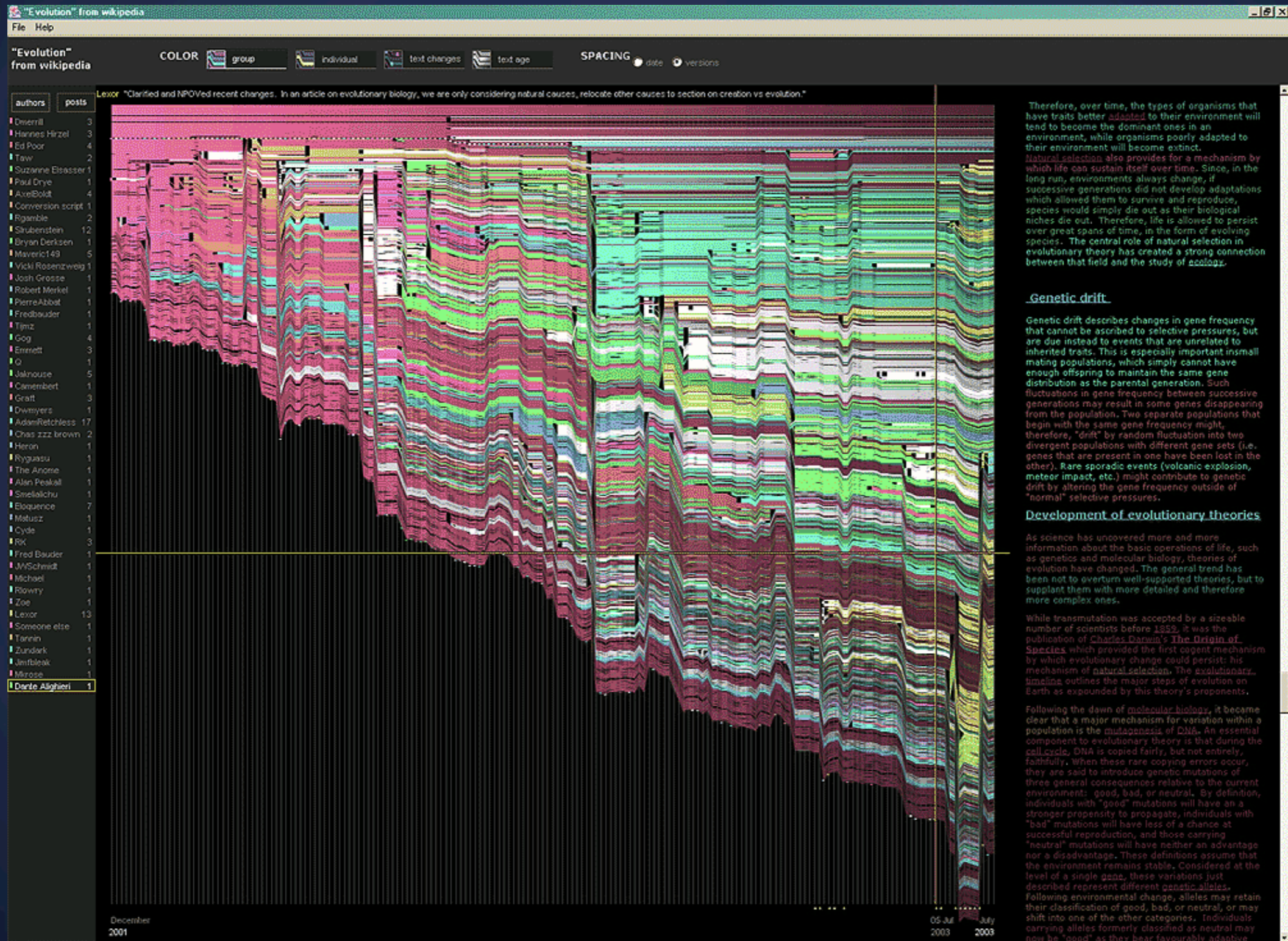
Document-Level Visualizations

On the Origin of Species: The Preservation of Favoured Traces



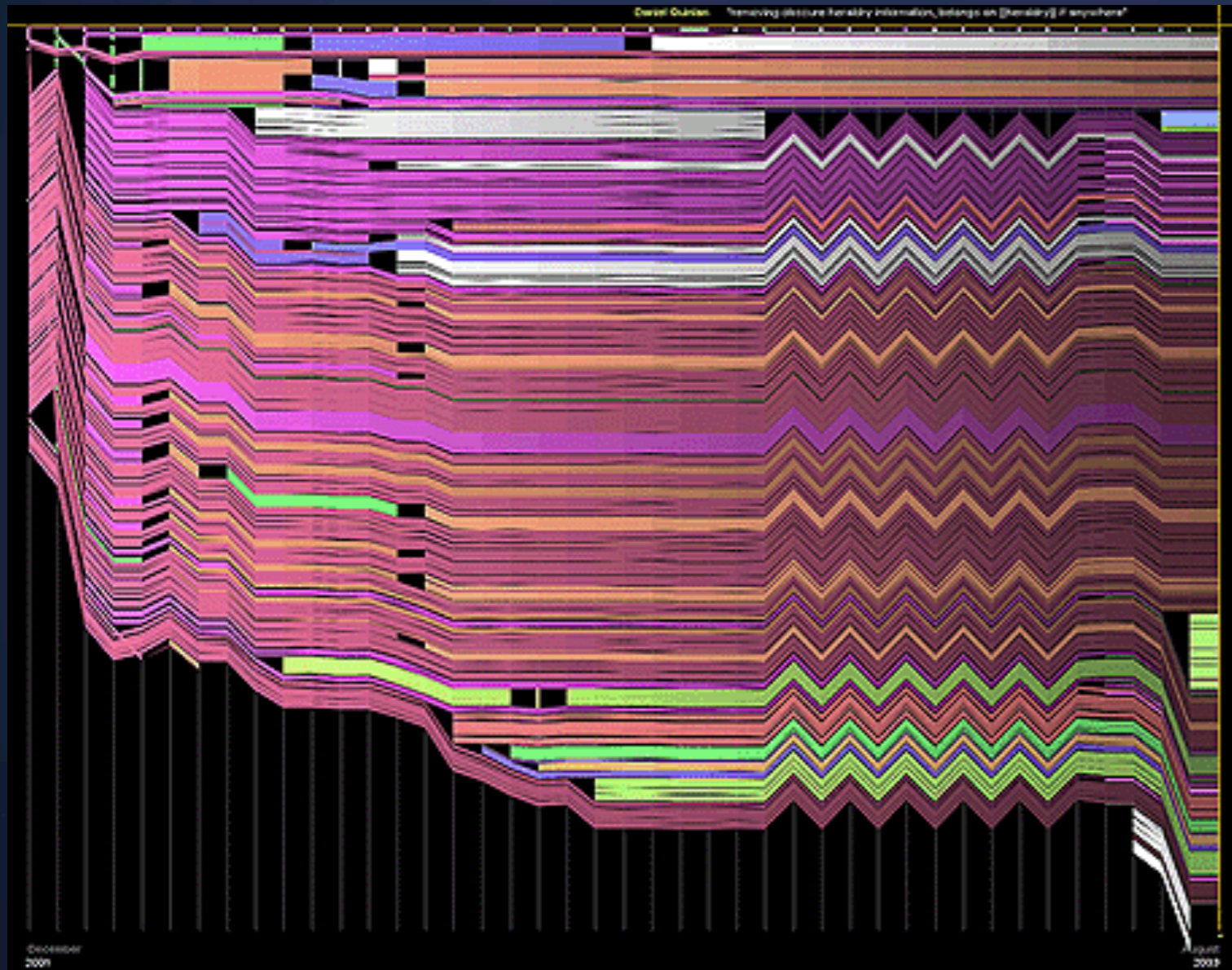
Evolution of a document

Visualizing additions / deletions in Wikipedia articles



Evolution of a document

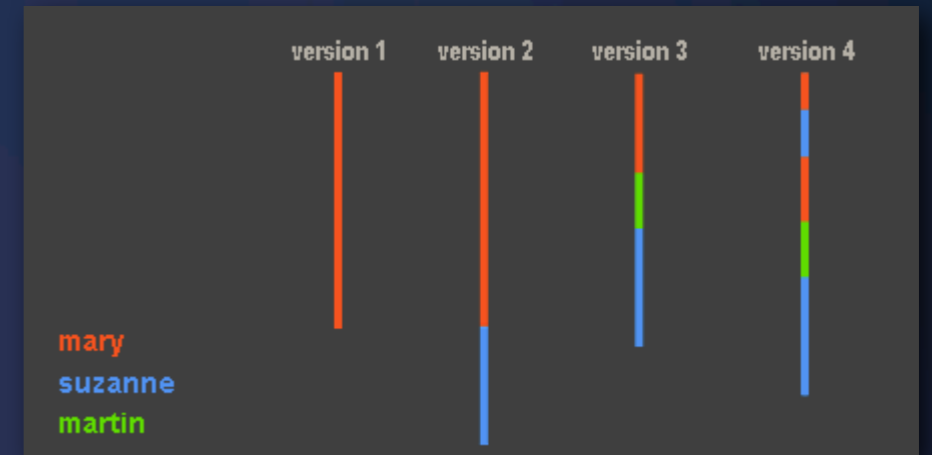
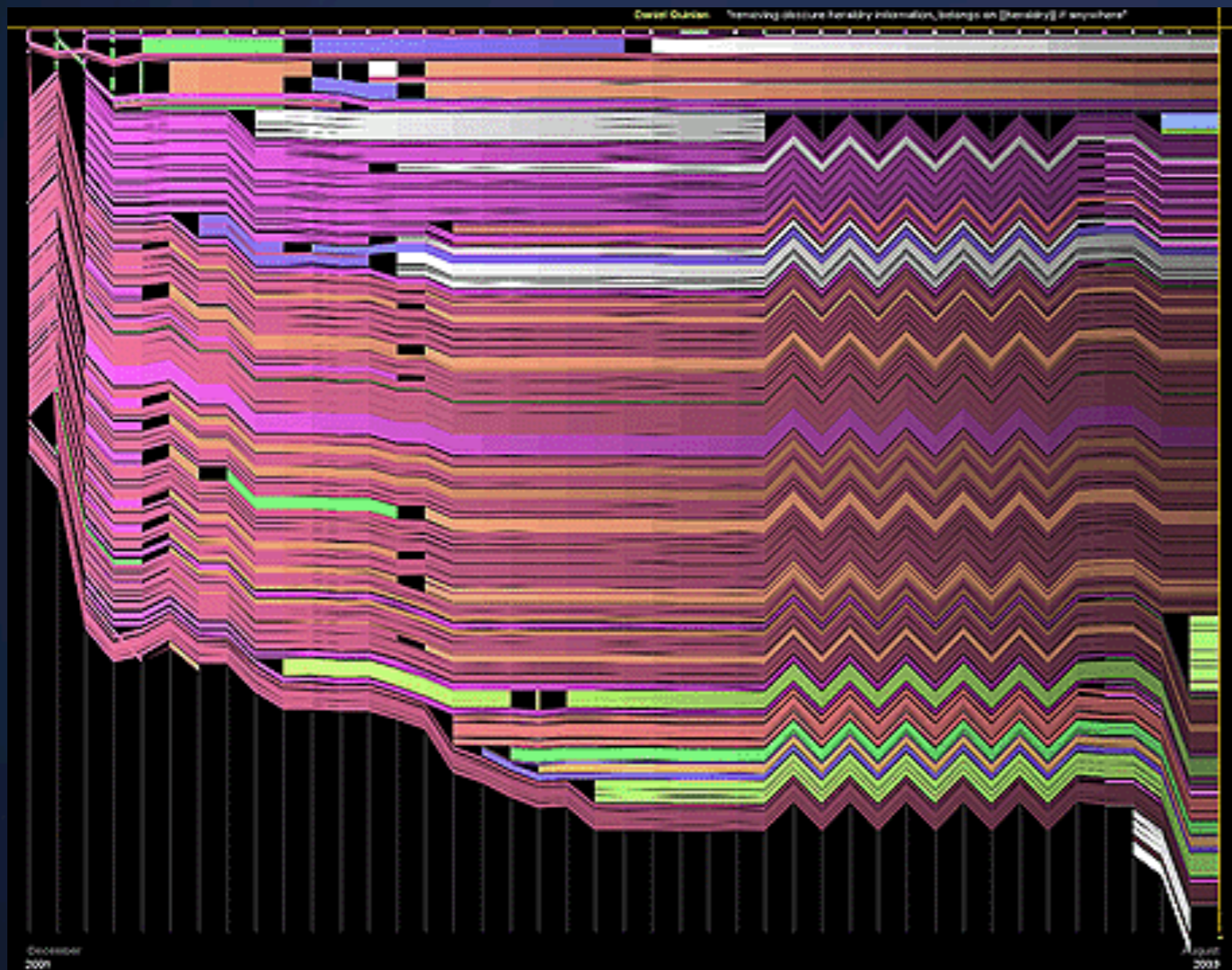
Visualizing additions / deletions in Wikipedia articles



Chocolate Wikipedia page

Evolution of a document

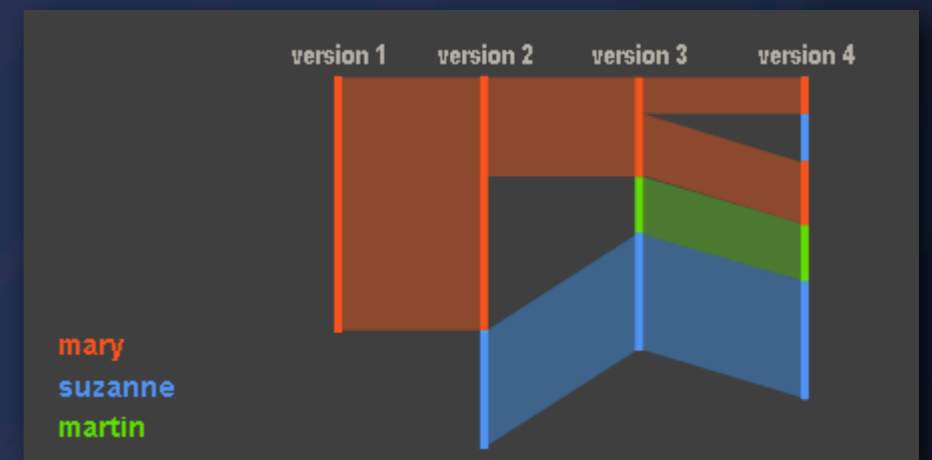
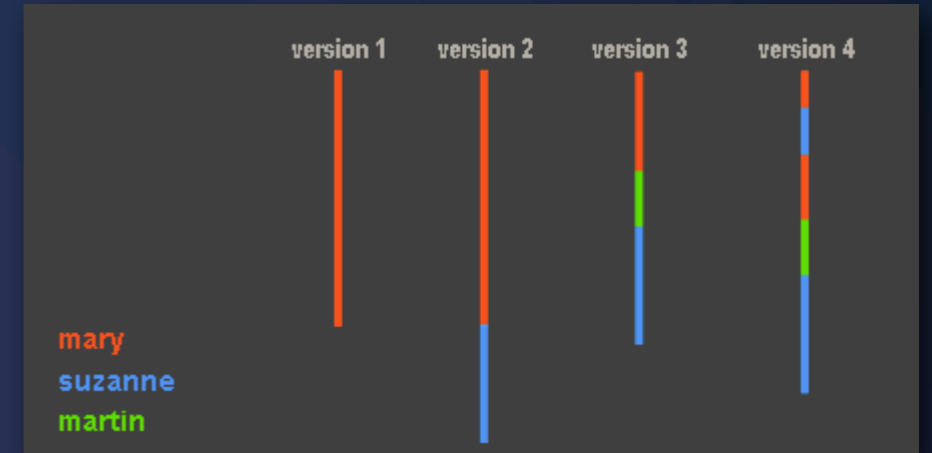
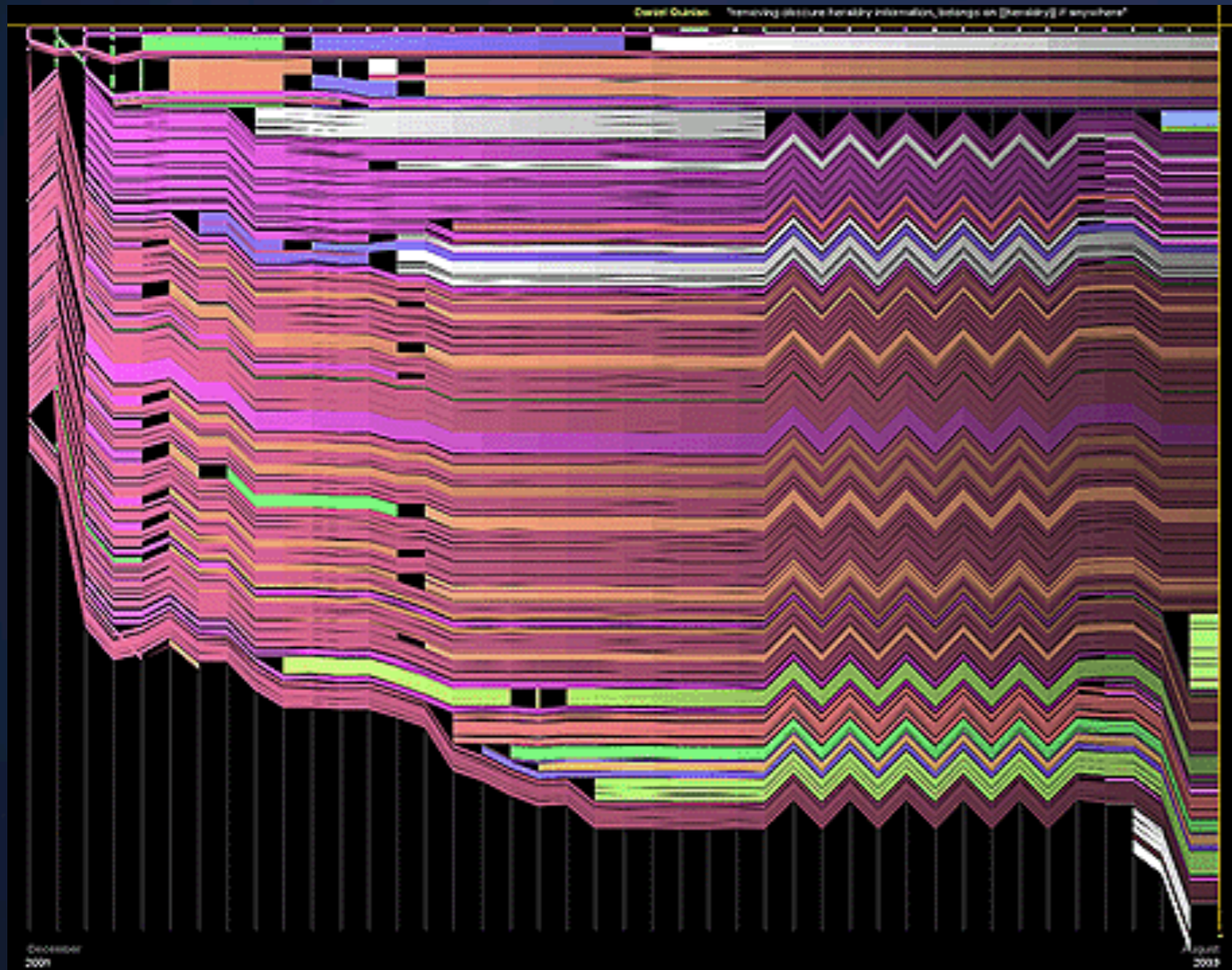
Visualizing additions / deletions in Wikipedia articles



Chocolate Wikipedia page

Evolution of a document

Visualizing additions / deletions in Wikipedia articles

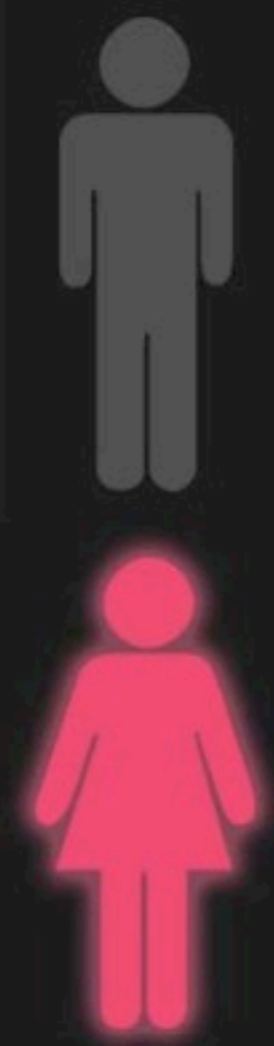




Chocolate Wikipedia page

Strategies

Introduce Faceted Search

Faceted search

Feeling	Gender	Age	Weather	Location	Date
abused	Female	40 - 49	All	All	All
All Feelings	Both Genders	All Ages	All Weather	All Locations	All Dates
A abandoned		0s 10s 20s 30s 40s 50s 60s 70s 80s		Choose Country	2005 2006 2007 2008 2009 2010 2011 2012 2013
B able					
C abnormal					
D abroad					
E absent					
F absolute					
G absorbed					
H abstract					
I absurd					
J abused					
K academic					
L acceptable					
M accepted					
N accepting					
O accomplished					
P according					
Q accountable					
R accurate					
T ace					
U aching					
V achy					
W acid					
X acknowledged					
Y across					
Z acting					

 [Find Feelings](#)

We Feel Fine

Faceted search



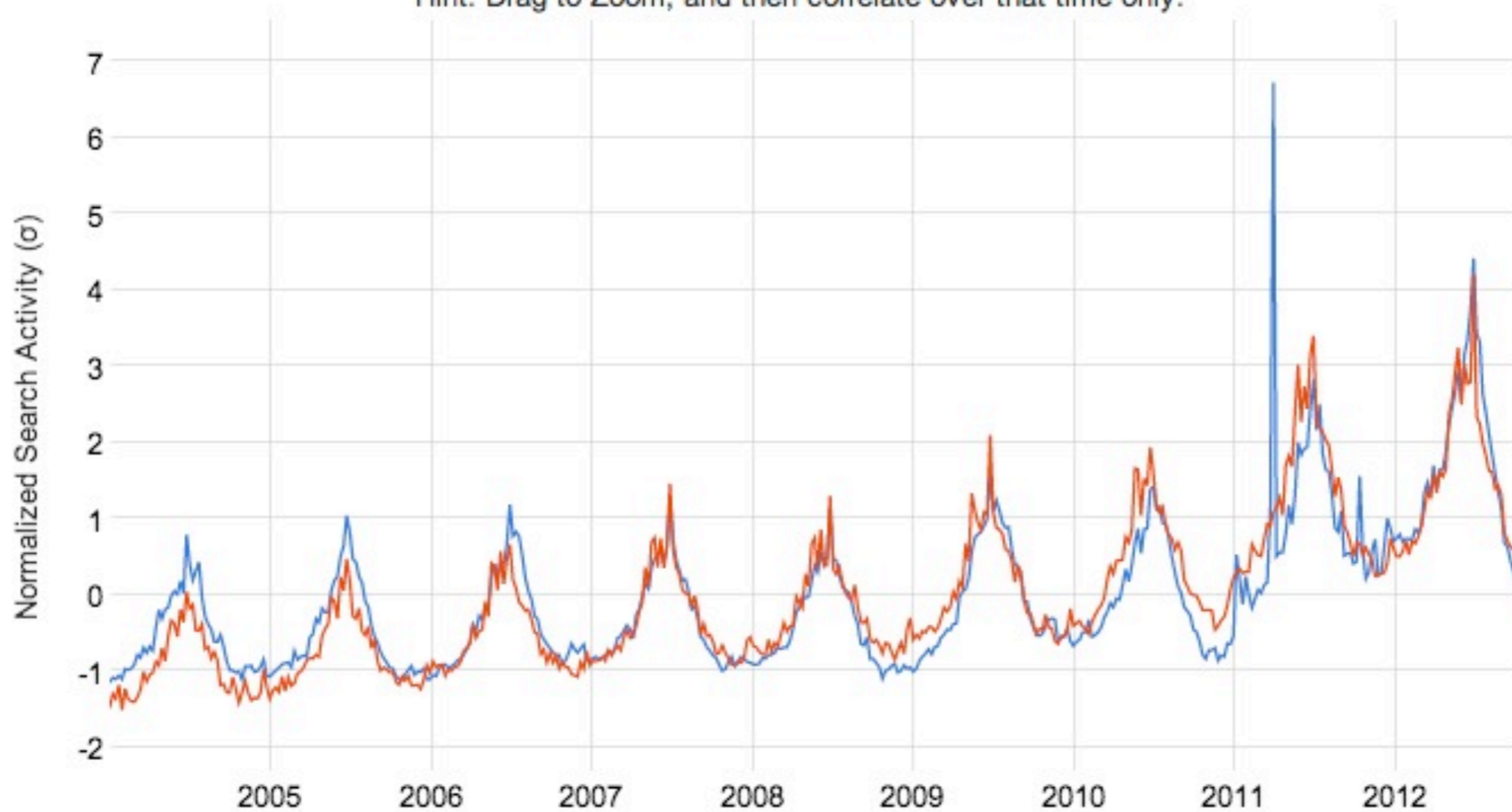
Add Time Dimension

Google Trends

United States Web Search activity for **ice cream** and **kabob** ($r=0.9033$)

Line chart Scatter plot
— ice cream — kabob

Hint: Drag to Zoom, and then correlate over that time only.



Recap

Lab Thursday

Concept Critique

Next Lecture

Networks: Graphs and Hierarchy