Visualizing TED Talks: A Suite Of Interactive Dashboards

Abstract

This paper describes the interactive dashboards created as a part of the HCDE 511 final project. It introduces the TED talks dataset that was used to create the visualizations and then goes on too describe the design process and then present the final results. The project led to the creation of three interactive visualizations which can be used to explore TED talks.

Introduction

TED Talks were conceived in 1984 as a one-time conference highlighting the convergence of Technology, Entertainment, and Design. This concept gained traction, and these talks grew to become a forum to express thoughts across a wide variety of topics including politics, technology, science, sociology, and entertainment. With over 1,700 talks freely available online, the TED videos have been viewed by over 1 billion users as of 2012. Given the breadth of content available and the popularity of videos with the general public, there exists a challenge in narrowing down content for each user based on their unique interests. We used a dataset made available through the Idiap Research Institute from Martigny, Switzerland to develop a general audience visualization tool for browsing TED talks. This dataset consists of metadata collected from the official TED website and includes information such as the talk themes, publication date, number of views, talk transcripts, and user comments. Though viewers are able to search for talks by filtering across topic, event, language, and rating on the current TED website, this is a limited perspective on the richness available within TED talks. We extend on this work by developing a visualization that allows users to:

- Explore TED talks by thematic content
- Identify talks across dimensions of interest
- Navigate talk by word concordance within transcript

The target audience for this visualization tool includes a primarily general audience of consumers interested in identifying TED talks. Given that this audience is diverse across technological skill levels and familiarity with TED content, there are distinct design challenges towards developing an appropriate visualization tool. In particular, we balanced the needs of an expert user in quickly identifying a talk of interest with the needs of a more novice user interested in browsing through the selection available and fortuitously finding a video of interest.

We present Explore TED, a visualization tool consisting of a Discover TED module, WordView module and Dogeify TED module. Discover TED allows users to narrow down a talk based on varying criteria while also providing a comparison of talks based on the number views, favorites, comments, and words within the talk. WordView provides a resource for users to explore the content within a talk based on keywords identified in the transcript. Users are able to search for words within talk, identify where and how often words occur during the talk, and go to segments of the video where these excerpts are spoken. Dogeify TED creates a humorous Doge meme (http://knowyourmeme.com/memes/doge) which is dynamic based on transcript key words and thematic content identified within a TED talk.

Design Process

1. Related work

As preliminary work, we conducted a review of existing visualization approaches towards the exploration of TED videos and similar collections of clips. The initial curators of this data set, Pappas and Popescu-Belis, applied a nearest neighbor algorithm incorporating user comments to provide recommendations on related TED talk videos. They also performed a sentiment analysis to categorize content within the videos (Pappas and Popescu-Belis, 2013). These applications were developed from an academic perspective to extract information from the metadata of TED talks; however this was not translated into a consumer focused resource for exploration of TED talks. Below we provide examples of more relevant information visualization based approaches:

Example: Tracking Web Video Topics

Shao, Ma, Lu, and Zhuang (2012) describe a unified framework to analyze a collection of web videos for topic discovery and visualization. The authors generate a k-partite cluster of features grouped by similarity of topic. The nodes within the graph network are ranked by popularity and visualized to allow users to browse topics at multi-level scales. In similar work, Cao, Ngo, Zhang, and Li (2011) develop a system for discovery, visualization, and monitoring of web videos. Applying an algorithm based on salient trajectory extraction, the authors generate a topic evolution link graph. This is used for discovery of videos by topic. The authors define a measure of saliency for topics based on social popularity and life-span evolution. They then provide a visualization using a 2D trajectory of how topics trend over time (Figure.1).

The work provided by this research is an interesting approach towards categorizing videos based on similarity of content, though for the purposes of our project, the algorithmic approach may not apply since TED Talks have a manually curated list of related videos. Instead, we want to extend on the visualization of video content. Showing trajectories and trends over time would be of practical use to users; however it would also be valuable to provide comparisons of different topics simultaneously. In addition, linking the trends with a topic exploration network would be a useful component of the visualization, allowing the viewer to simultaneously identify trending.

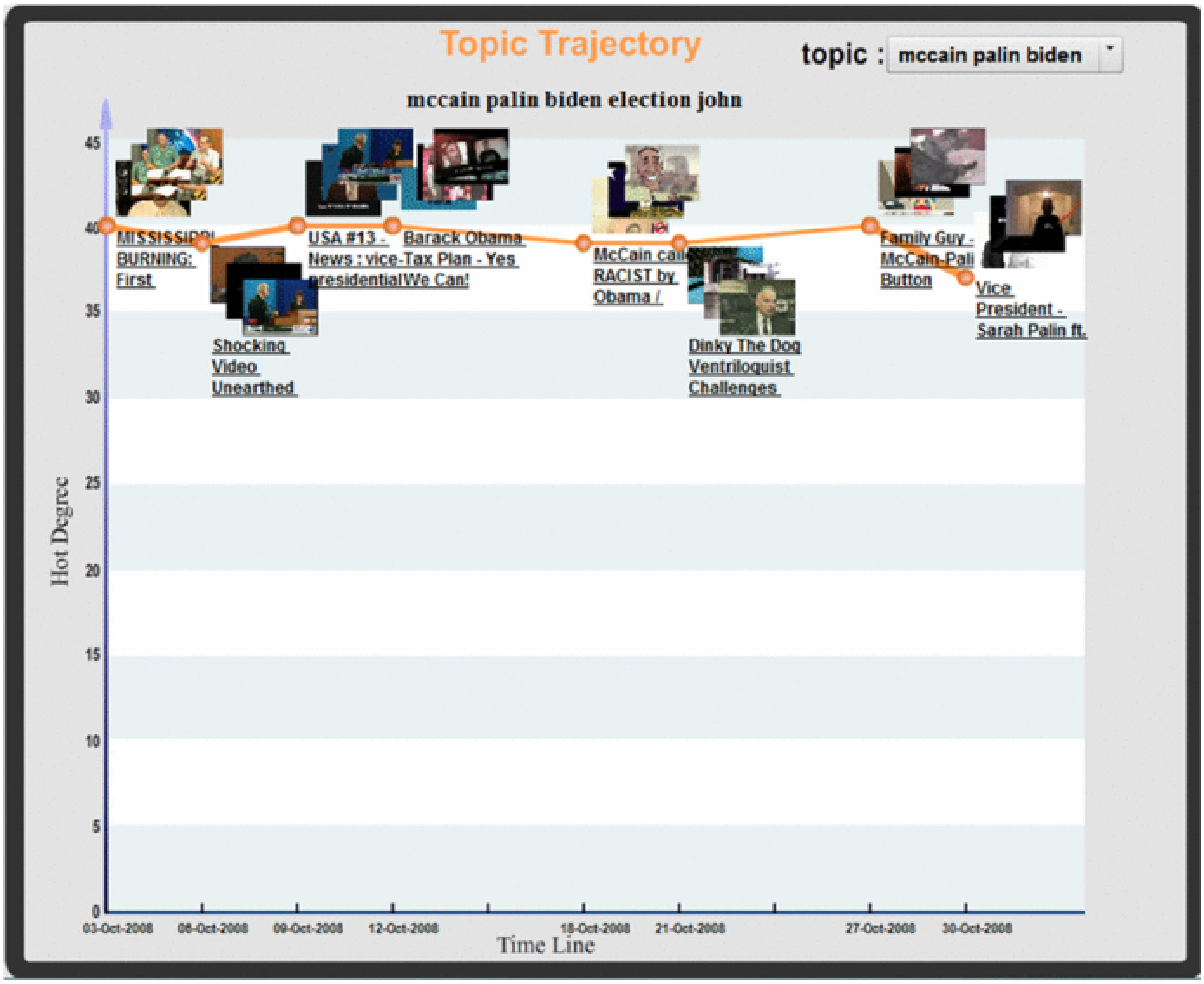

Topic Trajectory
topic : mccain palin biden
mccain palin biden election john
Hot Degree
45
40
35
30
25
20
15
10
5
0
MISSISSIPPI BURNING: First
Shocking Video Unearthed
USA #13 - News : vice-presidential
Barack Obama Tax Plan - Yes We Can!
McCain call RACIST by Obama /
Dinky The Dog Ventriloquist Challenges
Family Guy - McCain-Pali Button
Vice President - Sarah Palin ft.
03-Oct-2008
06-Oct-2008
09-Oct-2008
12-Oct-2008
18-Oct-2008
21-Oct-2008
27-Oct-2008
30-Oct-2008
Time Line

Figure 1. Trajectory of video topics over time

Example: Videosphere
http://www.bestiario.org/research/videosphere/
Videosphere is a project created by Bestiario, which visualizes the related TED talks in the form of a sphere. The visualization (Figure 2 & 3) relates similar TED talks with node-link diagram; links (lines) will be highlighted once relevant videos are selected. Users can select videos from tags lists on the left side, and each video provides a short description, related tags and its url to the video on TED.com.

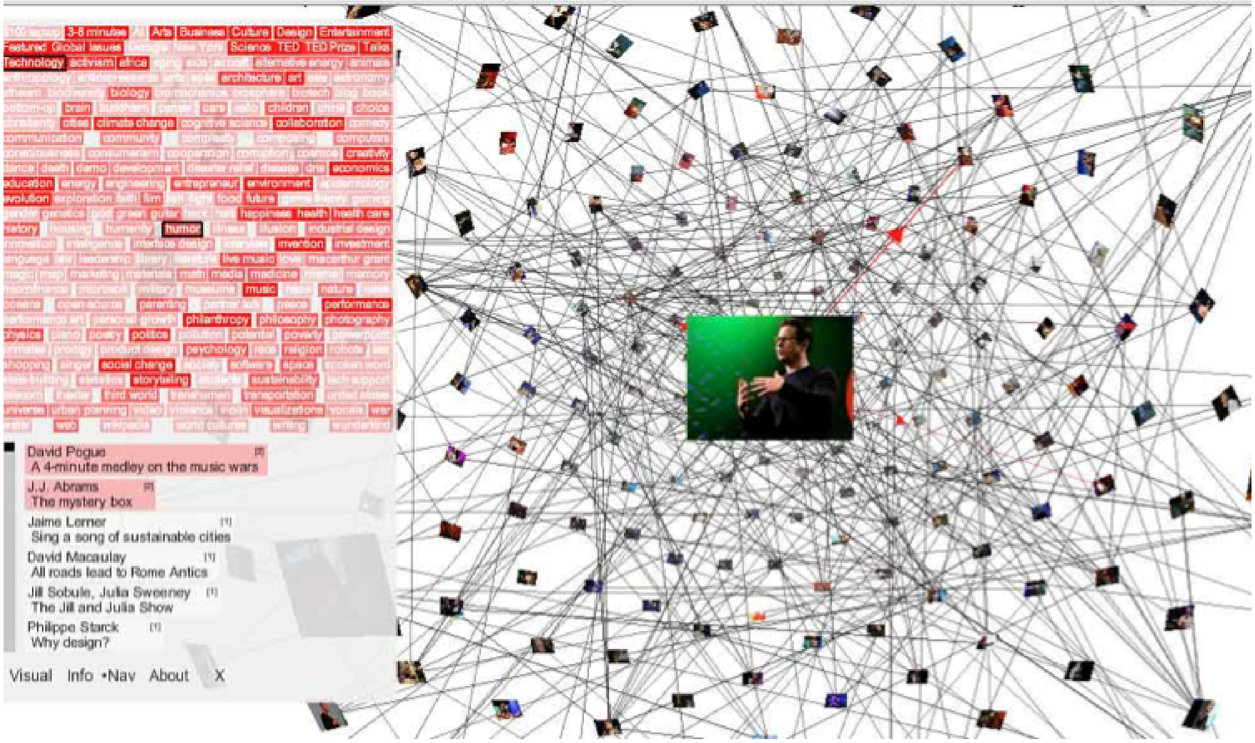


Figure 2. Tag list of Videosphere

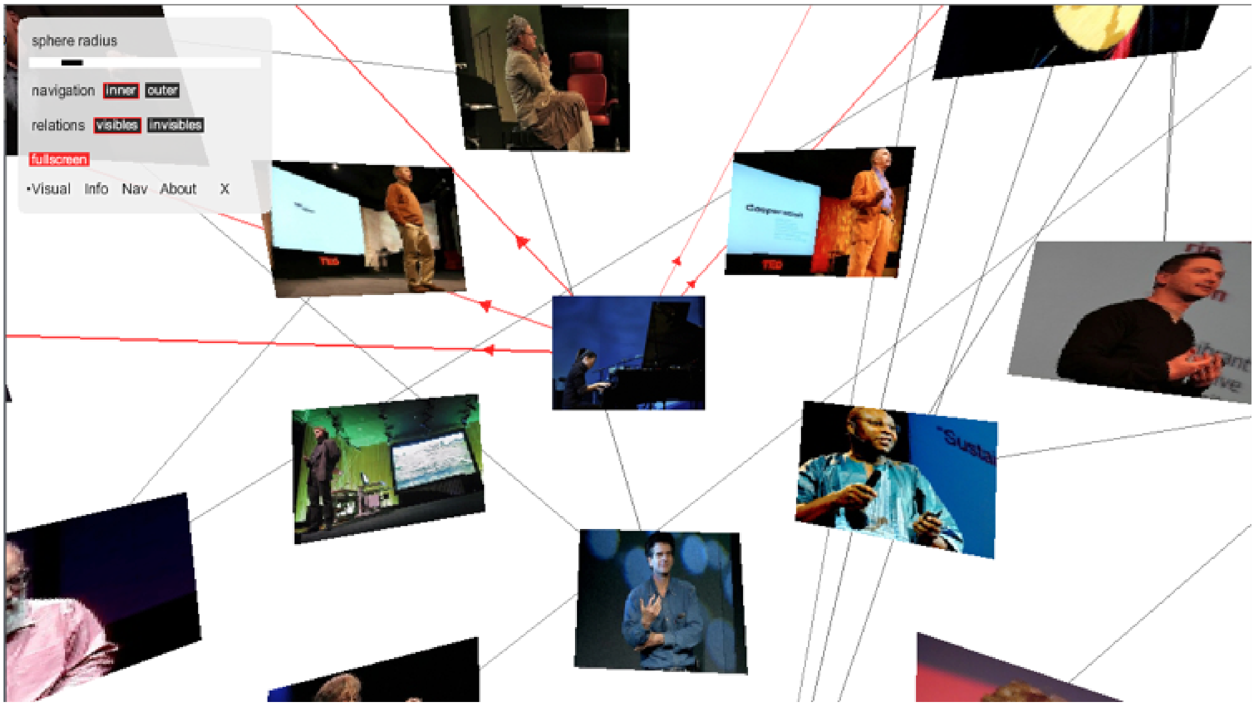
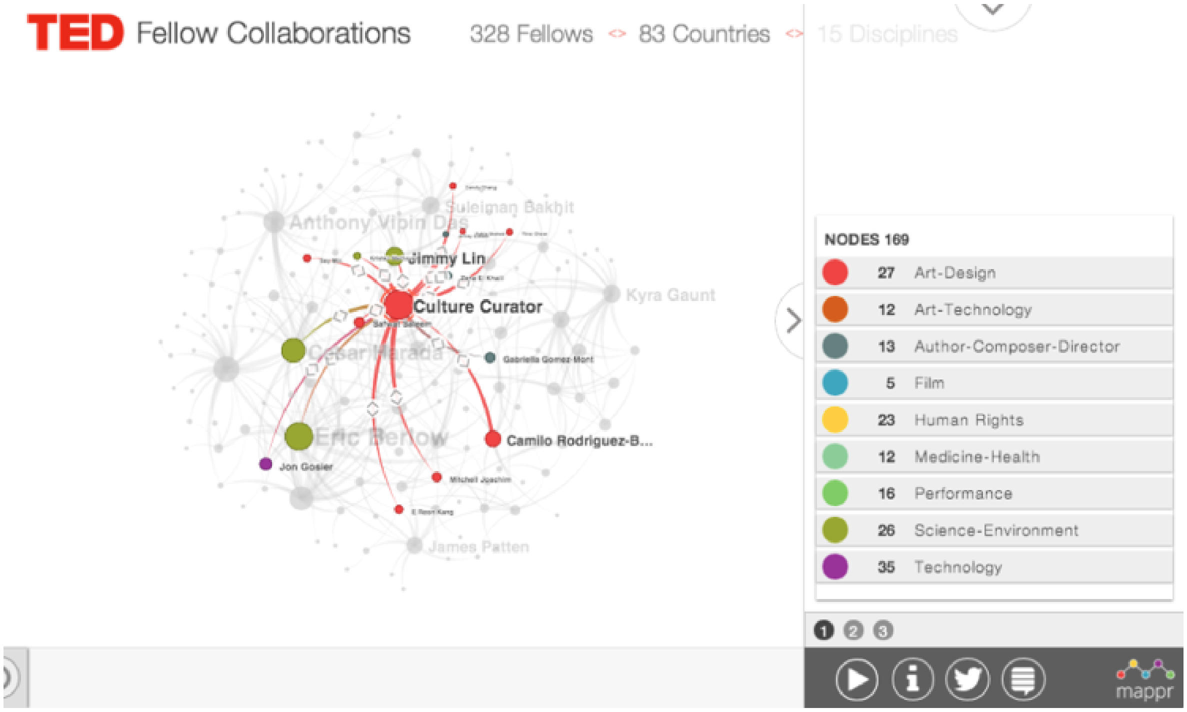


Figure 3. Highlighted links in Videosphere

Example: TED Fellow Collaboration

http://mappr.io/p/tedfellowcollaborations

Created by Vibrant Data, the visualization (Figure 4) represents a network map of TED Fellows' collaboration. Its nodes represent TED Fellows and the link represents a significant collaboration. Users can filter the nodes by disciplines, roles and geographical area of the fellows. TED Fellows are doers, makers, inventors, advocates, filmmakers and photographers, musicians, scientists and etc.



NODES 169		
●	27	Art-Design
●	12	Art-Technology
●	13	Author-Composer-Director
●	5	Film
●	23	Human Rights
●	12	Medicine-Health
●	16	Performance
●	26	Science-Environment
●	35	Technology

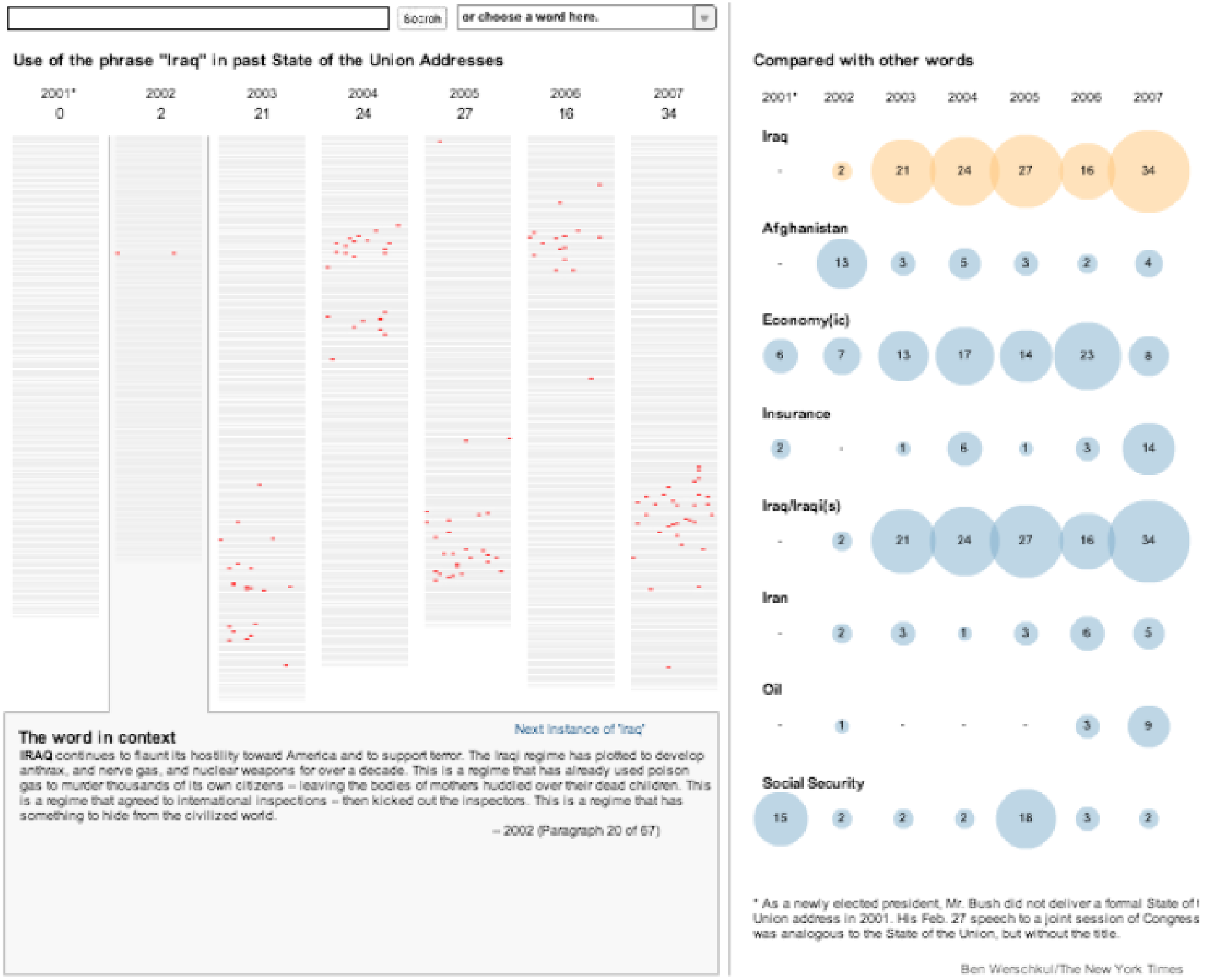
Figure 4. Network map of TED Fellows' collaborations

Example: The 2007 State of Union Address

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

This visualization (Figure 5) was once used by the New York Times as a tool to explore the words used in the State of the Union Addresses over 7 years, from 2001 to 2007. Users can either search or choose a keyword they are interested in, and then the dashboard will show the frequency and distribution of that word over years, its context (e.g. how the word is used in the certain paragraph in transcript) and the word's comparison with other words of high frequency. This visualization provides inspiration in ways to visualize the frequencies of usage of the words across year, as well as present words within context and a different way to compare frequencies.

Although this work does not directly incorporate TED talks, it provides an inspiration for how specific words from a corpus, such as as TED talk transcriptions, can be visualized.



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

2001*	2002	2003	2004	2005	2006	2007
0	2	21	24	27	16	34

The word in context

IRAQ continues to flaunt its hostility toward America and to support terror. The Iraqi regime has plotted to develop anthrax, and nerve gas, and nuclear weapons for over a decade. This is a regime that has already used poison gas to murder thousands of its own citizens – leaving the bodies of mothers huddled over their dead children. This is a regime that agreed to international inspections – then kicked out the inspectors. This is a regime that has something to hide from the civilized world.

– 2002 (Paragraph 20 of 67)

Compared with other words

	2001*	2002	2003	2004	2005	2006	2007
Iraq	-	2	21	24	27	16	34
Afghanistan	-	13	3	5	3	2	4
Economy(ic)	6	7	13	17	14	23	8
Insurance	2	-	1	6	1	3	14
Iraq/Iraqi(s)	-	2	21	24	27	16	34
Iran	-	2	3	1	3	6	5
Oil	-	1	-	-	-	3	9
Social Security	15	2	2	2	18	3	2

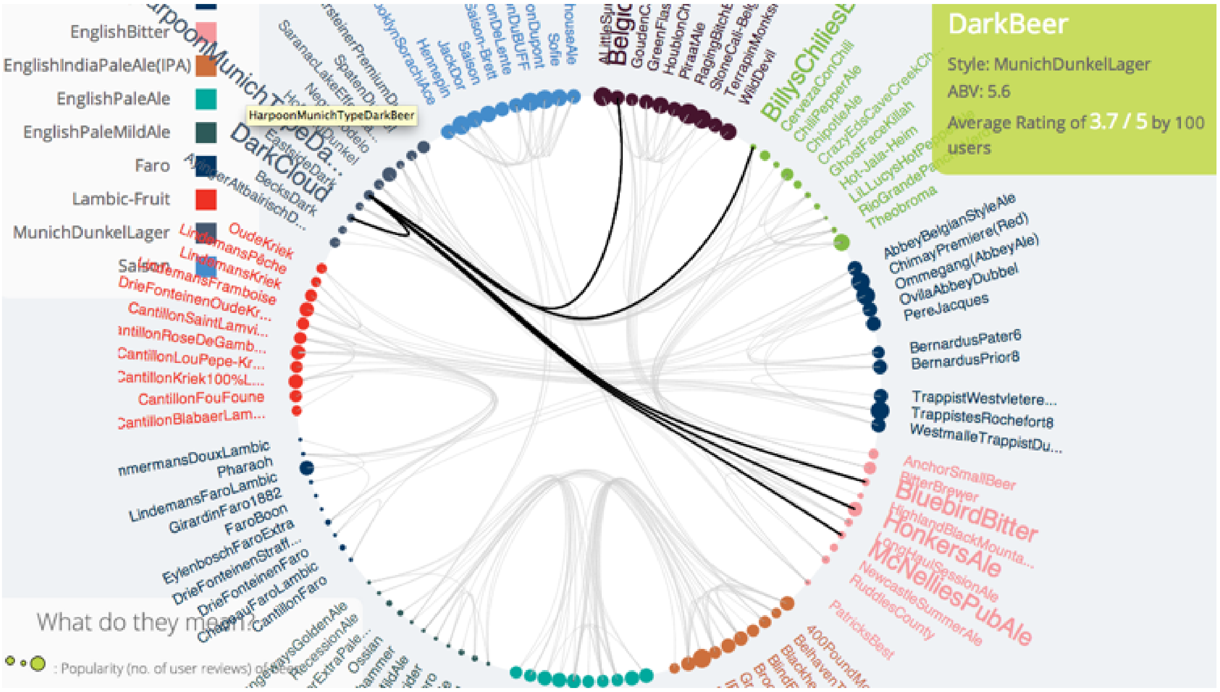
* As a newly elected president, Mr. Bush did not deliver a formal State of | Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

Ben Werschkul/The New York Times

Figure 5. The 2007 State of Union Address

Example: BEERVIZ

Live site: http://seekshreyas.com/beerviz/
Documentation: http://blog.yhathq.com/posts/recommender-system-in-r.html
Dataset: http://snap.stanford.edu/data/web-BeerAdvocate.html
BEERVIZ is a visualization of beer recommendation system, where users can compare, select and filter beer by style, appearance, taste, aroma and overview information (Figure.6). It presents an alternative way to visualize similarity between elements within different categories, which provides inspirations in visualizing related video across themes in our case.



DarkBeer

Style: MunichDunkelLager

ABV: 5.6

Average Rating of 3.7 / 5 by 100 users

Figure 6. Similarity between light beers, BEERVIZ

Example: Bibly

Live site: http://stanford.edu/~garylee/bibly/

Bibly is a visualization created by Gary Lee and Anirudh Venkatesh to visualize word distribution within the King James Version Bible. Users search for words which are then displayed in concordance with their location in the Bible. This allows users to quickly identify frequency of occurrence of words along with their location. We applied a similar approach as Bibly to visualize transcripts of TED talks within the WordView module.

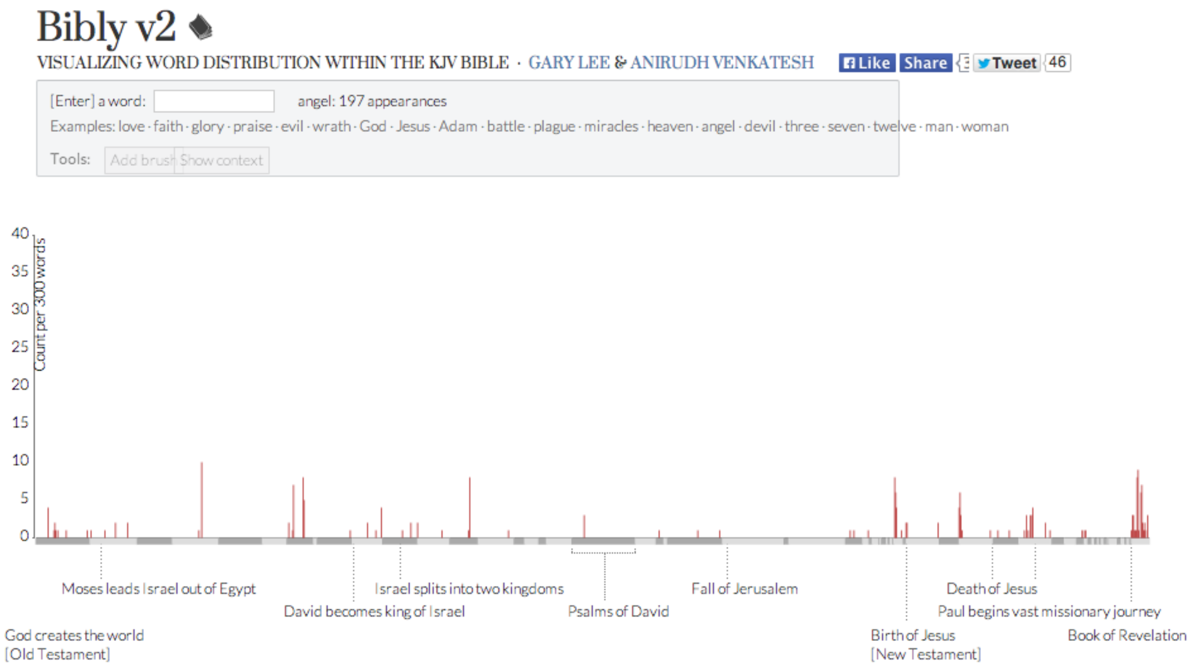

Bibly v2
VISUALIZING WORD DISTRIBUTION WITHIN THE KJV BIBLE · GARY LEE & ANIRUDH VENKATESH
Like
Share
Tweet
46
[Enter] a word:
angel: 197 appearances
Examples: love · faith · glory · praise · evil · wrath · God · Jesus · Adam · battle · plague · miracles · heaven · angel · devil · three · seven · twelve · man · woman
Tools:
Add brush
Show context
Count per 300 words
40
35
30
25
20
15
10
5
0
Moses leads Israel out of Egypt
David becomes king of Israel
Israel splits into two kingdoms
Psalms of David
Fall of Jerusalem
Death of Jesus
Paul begins vast missionary journey
God creates the world
[Old Testament]
Birth of Jesus
[New Testament]
Book of Revelation

Figure 7. Visualizing the word distribution within the KJV Bible, Bibly

2. Data Preprocessing

The dataset we used was originally created by Pappas & Popescu-Belis (2013) to perform a sentiment analysis of user comments and creating an algorithm for generating user recommendations (Pappas & Popescu-Belis, 2013a; Pappas & Popescu-Belis, 2013b). They applied a nearest-neighbor algorithm incorporating user comments to provide recommendations on related TED Talk videos. Although their dataset has been publicly released, there has been no work done to create a publicly viewable visualization.

The original dataset was formatted as several independent JSON files containing measures of TED video metadata retrieved from the web. The fields present in the dataset included:

- Talk Title
- Talk Description (limited to the first 255 characters)
- Talk Transcript
- Speaker
- TED Event
- Film Year
- Tags (determined by administrators of the official TED website)
- Themes (determined by administrators of the official TED website)
- Related Videos (determined by administrators of the official TED website)
- Number of Views

The authors provided a Python script packaged with the dataset to demonstrate how to interact with the data. For this project, the script was modified and expanded on in several ways to format the data in ways pertaining to the visualization.

In order to be compatible with the Tableau data visualization software, it was required that the data be transformed from JSON format to CSV format. During the automated transformation, the script computed several measures which were not present in the original dataset.These computations included:

- Number of words, computed from the talk transcript using the Natural Language Toolkit (Bird et al, 2009)
- Number of times a talk had been favorited, computed from data encoding the talks which individual users had favorited
- Number of comments, computed by traversing and tallying the associated user comment tree for each talk
- Sentiment analysis of comments, computed using Python's textblob library (retrieved from http://textblob.readthedocs.org/en/dev/).
- Transformation of the TED talk URL to the associated URL for the embedded video

Speaker gender was coded by hand and added to the resulting CSV data file.

Additionally, a trimmed JSON file was created for use in D3. This file excluded the nearly 130 megabytes of raw user comment data present in the original file, and added the data computation fields above. In addition, two more fields were added:

- An array of key-value pairs, where each key was a unique word in the talk and the value encoded the number of times used. Additionally, any word present in a list of function words derived from this resource was excluded: http://myweb.tiscali.co.uk/wordscape/museum/funcword.html
- An array containing the 10 most frequently-used words for the talk, drawn from the above array.

It is important to note that the dataset contained several incomplete entries. Entries which were missing a TED event field, or who had no clear speaker (as in the case of a band playing music) were excluded from the visualization. In addition to the entries containing missing data fields, all entries contained only the first 255-characters of the talk's description. This was the weakest part of an otherwise strong dataset, limiting the polish on the final dashboard.

3. Initial Sketches & First exploration with Tableau

We did initial sketches at the same time when we were dealing with the dataset. Inspired by the examples mentioned above, we came up with several sketches based on the dimensions of the TED datasets (e.g. transcript, theme, views, comments, publish_date, events, favorite videos and etc.), representing some trends over years and the interrelationships between these dimensions. Following are the pictures of the sketches:

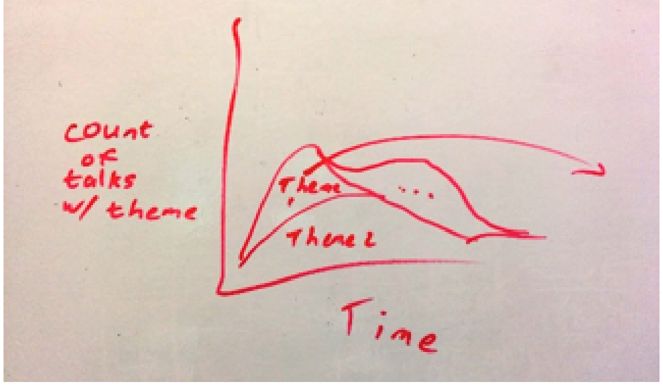


Figure 8. Plotting view counts per theme to time

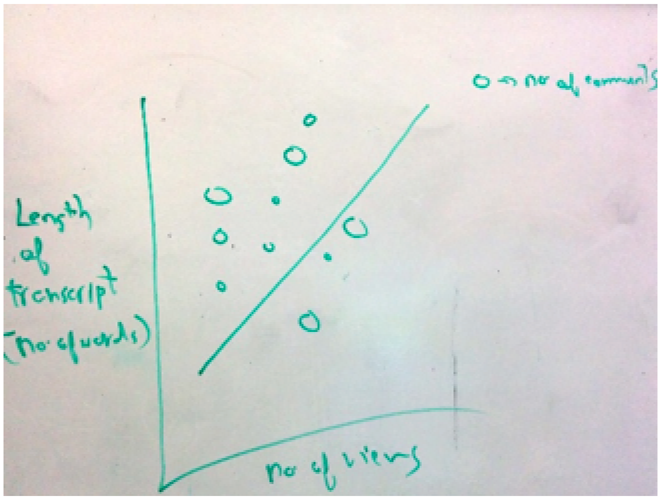


Figure 9. Plotting length of transcript to number of views

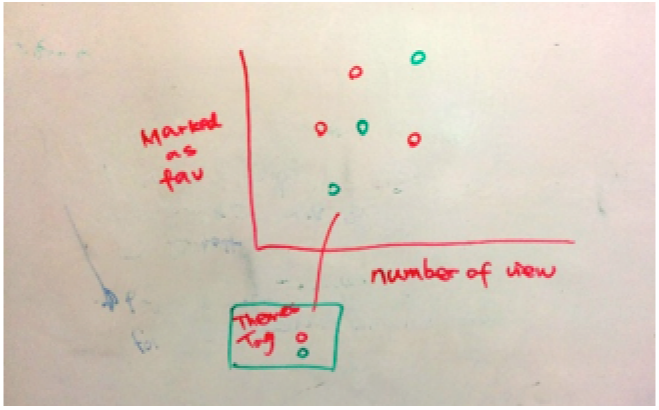


Figure 10. Plotting #favorited to #views

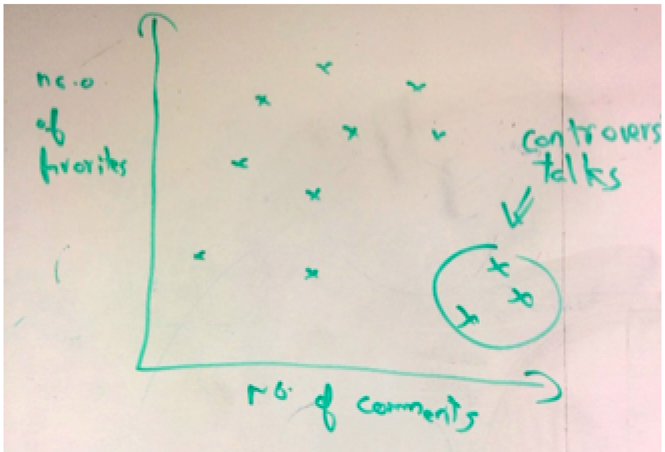


Figure 11. Plotting #favorited to #comments

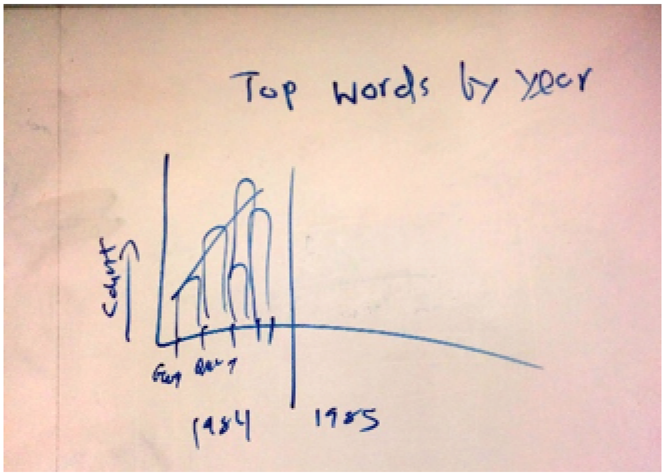


Figure 12. Plotting most frequent words in talks by year

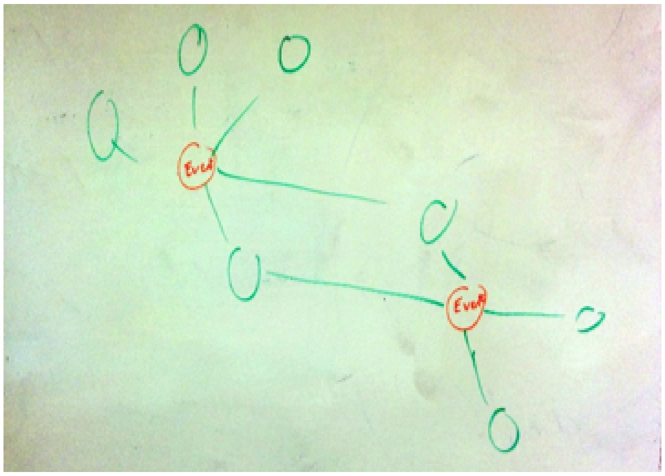
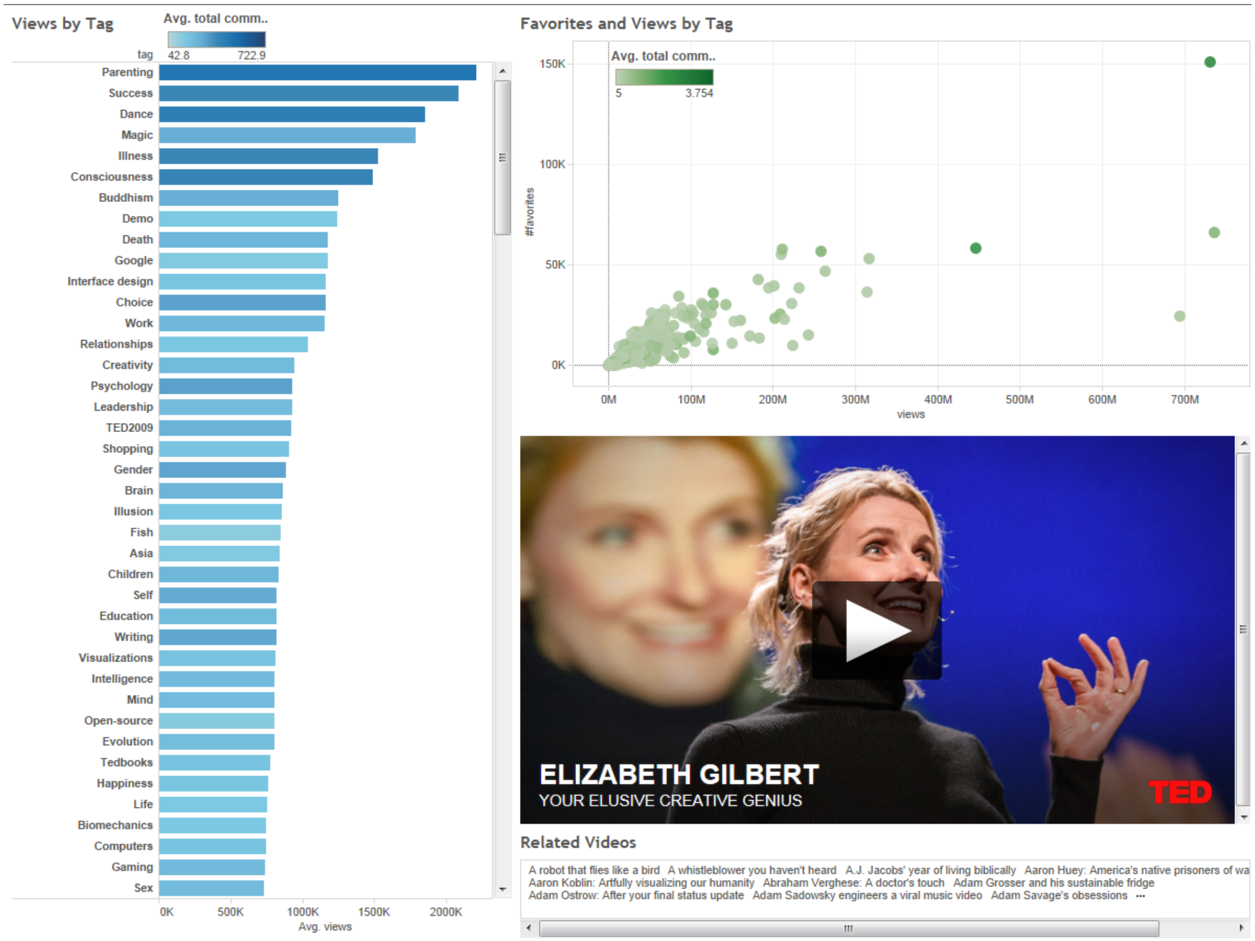


Figure 13. Creating relationship network between videos and events

We explored the different views in Tableau in order to test the technical feasibility of our initial ideas. We created three initial dashboard prototypes.

Dashboard - Topics and talks

This dashboard allows viewers to view talks' popularity by topics, as well as explore the relation between number of views and number of favorites for selected talks in a scatterplot. Clicking a bar on the left will filter talks according to the selected topic, and clicking a point on the scatterplot will load the actual talk in the embedded video.



A robot that flies like a bird   A whistleblower you haven't heard   A.J. Jacobs' year of living biblically   Aaron Huey: America's native prisoners of wa
Aaron Koblin: Artfully visualizing our humanity   Abraham Verghese: A doctor's touch   Adam Grosser and his sustainable fridge
Adam Ostrow: After your final status update   Adam Sadowsky engineers a viral music video   Adam Savage's obsessions   ···

Figure 14. Screenshot of Dashboard "Topics and Talks"

Dashboard - Themes by years

This dashboard visualizes the TED talks themes by years with a themeriver overview. It allows viewers to find the most popular talks and popular speakers for each theme (in terms of number of views and number of favorites).

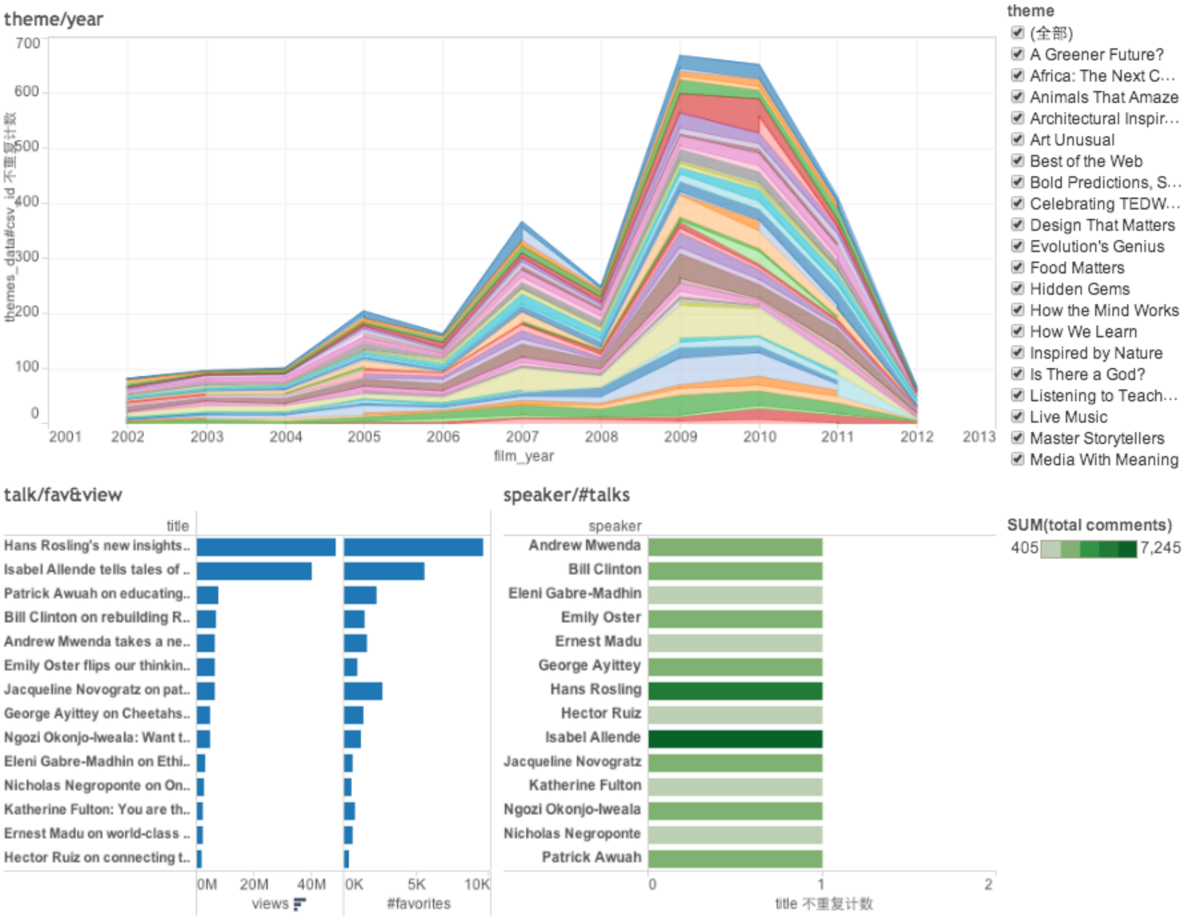


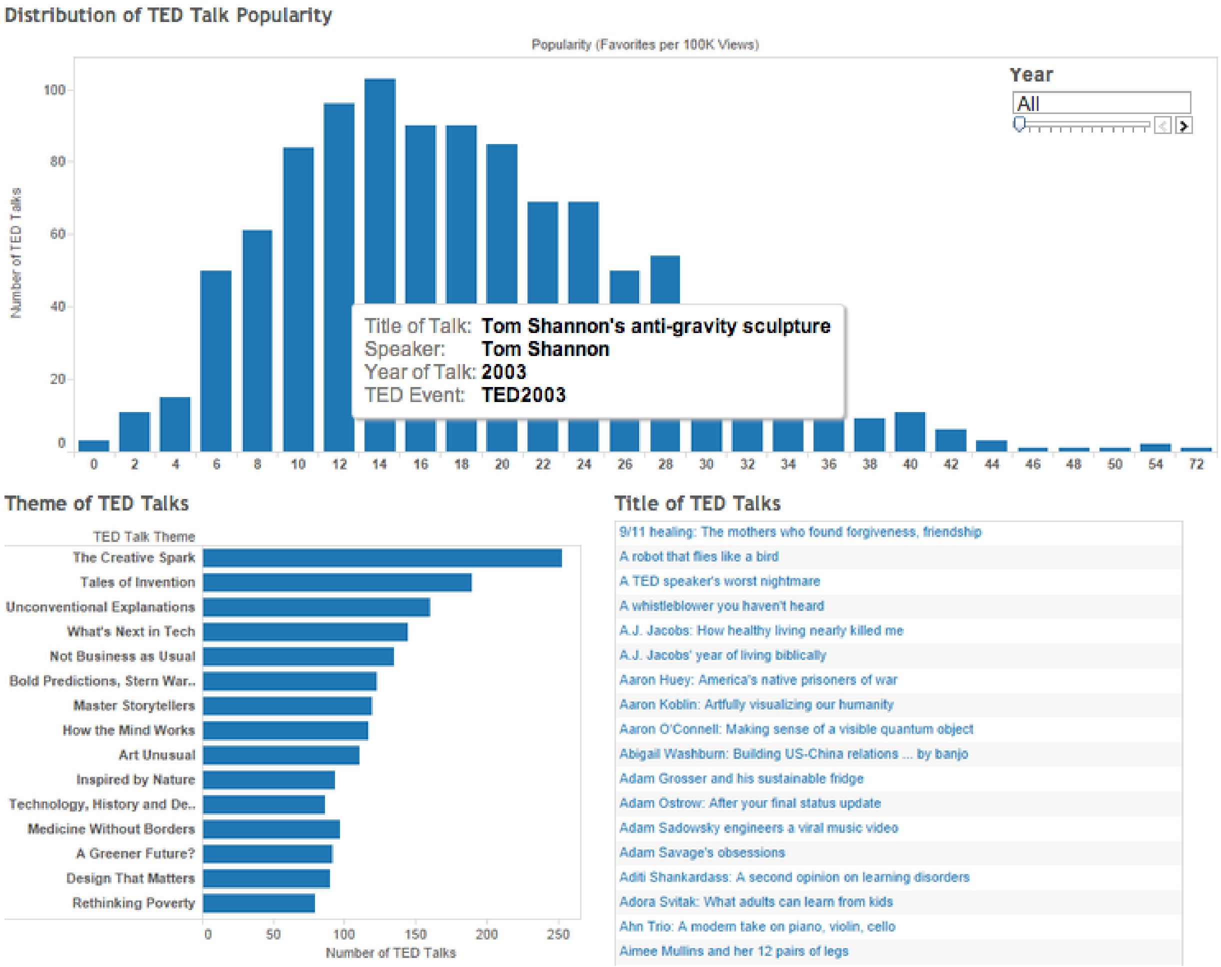
Figure 15. Screenshot of Dashboard “Themes by years”
http://public.tableausoftware.com/profile/#!/vizhome/Prototype_V1/Themetalkspeaker

Select a theme will filter the related talks, which are encoded with bars according to the measure of their number of views and number of favorites. Select a specific talk on the left will highlight its speaker on the right, and click on the speaker’s name will highlight the others talks given by that speaker within that themes on the left.

Dashboard - Topics by popularity

This dashboard visualizes the distribution of talks by a field we defined as popularity, which is calculated as the number of favorites per 100k views. In this dashboard, users can also identify talks by themes and the year published.

Selecting a talk on the distribution chart will show its theme and title. Viewers can also use a slider to see the number of popular talks in a selected year, which will be highlighted on the distribution chart. Selecting a theme will show the distribution of popular talks within that theme by years.



Title of Talk: **Tom Shannon's anti-gravity sculpture**
Speaker: **Tom Shannon**
Year of Talk: **2003**
TED Event: **TED2003**

9/11 healing: The mothers who found forgiveness, friendship
A robot that flies like a bird
A TED speaker's worst nightmare
A whistleblower you haven't heard
A.J. Jacobs: How healthy living nearly killed me
A.J. Jacobs' year of living biblically
Aaron Huey: America's native prisoners of war
Aaron Koblin: Artfully visualizing our humanity
Aaron O'Connell: Making sense of a visible quantum object
Abigail Washburn: Building US-China relations ... by banjo
Adam Grosser and his sustainable fridge
Adam Ostrow: After your final status update
Adam Sadowsky engineers a viral music video
Adam Savage's obsessions
Aditi Shankardass: A second opinion on learning disorders
Adora Svitak: What adults can learn from kids
Ahn Trio: A modern take on piano, violin, cello
Aimee Mullins and her 12 pairs of legs

Figure 16. Screenshot of Dashboard "Topics by popularity"
http://public.tableausoftware.com/profile/#!/vizhome/TEDTalkv1/PopularityofTEDTalks

4. Prototype Evaluation (Usability testing)

In order to gauge user interest, a short user test was conducted on each dashboard (Dashboards 1, 2 and 3). A total of four participants were asked to think aloud while exploring each prototype. After all dashboards had been explored, a short unstructured interview was conducted to prompt users for their overall thoughts on each visualization.

Users enjoyed the embedded videos and were also interested in some of the interactivity in general.

However, users reported being overwhelmed by the amount of information on each dashboard. Additionally, it was apparent that users were not interested in exploring the dataset simply to discover relationships present in the data. Accordingly, there was no interest expressed regarding the changes through time or the scatterplot present in the prototypes. Rather, users appeared interested in the “tags,” or topics, which described the subject matter present in each talk.

Users were also confused regarding the interaction with the dashboards themselves. Users did not realize that it was possible to highlight and filter the dashboards by selecting the associated data points. We expect this was due in part to the lack of textual cues hinting at the possible interactions afforded by the dashboard.

The scatterplot in particular seemed confusing to interact with. Besides the lack of user engagement with the quantitative metrics which were plotted, each point in the scatterplot provided only small click targets for users to interact with. Additionally, in almost every talk category there was inevitably an outlier which caused the remaining talks present in the scatterplot to be clumped together. Selecting a talk in this region was especially tedious due to Tableau’s requirement that the mouse cursor be moved far away from a given target to “clear” the tool tip information before the cursor could return to display details for another nearby data point.

Milestone 1 - Major changes we made

Create a single dashboard to select TED talks

The feedback from the user evaluation and the mid-term peer review showed a large interest for a single dashboard which could be used to select a TED talk to watch. In addition, we decided not to pursue visualizing the themes and dates of the talks, partially because users seemed less interested in these two dimensions. In addition, the dataset predominantly contains talks from 2006 to 2012, which may limit the complete picture of TED talks.

Although it had not been present in any of the dashboard prototypes so far, we decided not to pursue sentiment analysis for several reasons. Since participants in our tests did not express an interest for visualizing quantitative metrics of the tals, we reasoned that potential users would likely not be interested in exploring the sentiment analysis data. The large dataset of user comments would have caused speed issues in Tableau, especially given the row duplication which occurs on table joins. Finally, visualizing sentiment analysis may have required an entirely separate dashboard due to its difference with our other data types.

5. User studies

To better understand what people really care about the TED talks as well as differentiate us from the TED talks website, we conducted another round of gathering data on potential users. 4 user interviews and an online survey were used to understand users’ needs and their preferences for exploring TED talks. These results helped shape refined design requirements for TED talk visualization.

Interview and task analysis

We conducted 4 interviews, asking participants about their preferences of exploring TED talks and their priorities of different dimensions of interests (See Appendix 1 for interview questions).

In general, interviewees reported a preference to browse videos by topics ("tags" in our dataset) and may follow the related videos in order to find additional talks that they may enjoy. Length of the video also factored in to interviewees' decisions in selecting a talk. When asked about the possibility of a "guided tour" of the visualization, interviewees reported that they would rather explore the visualization by themselves.

Online survey

In order to sample a wide breadth of TED talk watcher interests, we created a short online survey. The main goal of the survey was to see which fields potential users would want to use to sort and filter the available talks. The survey was posted to reddit, Facebook, and UW mailing lists and received 157 responses. (See appendix 2 for online survey question).

Among all the respondents, 45 (29%) had watched more than 10 talks in the past year, 19 (12%) watched 7-9 times, 35 (22%) watched 4-6 times, 50 (32%) watched only 1-3 times, while 8 did not watch any talks.

125 respondents selected that browsing videos by topics as one way in which they would want to be able to browse videos, with 99 respondents selecting this as the most important way they would prefer to find videos. 75 respondents selected that they would hope to be able to browse by the number of times a video had been viewed, but only 16 selected that this would be the most important way they would hope fo browse videos. The number of times a talk had been favorited was also important to some people, with 50 selecting that this would be one way which they would hope to browse, 19 of whom selecting that this would be the most important way they would want to see the videos.

People were much less interested in browsing videos by the number of comments (13 interested, 1 most important), the TED event that the video was filmed at (37 interested, 4 most important), or by specific words used in the talks (25 interested, 5 most important). However, for those who were more frequent watchers of TED talks, there was a higher interest in browsing by event type or talks given by a specific speaker. For reponsonders who watched more than 10 talks in the past year, 11 of them indicated they would like to select talks by event and 2 of them selected this as the most important way for decision making.

In general, participants seemed neutral regarding the usefulness of a guided tour which would describe the possibilities afforded by the visualization (Mean=2.93 out of 1-5 Likert scale, SD=1.23)

Based on the interview and survey results, we stepped back reviewing our previous design and decided to address the user needs by building two dashboards. Survey questions are listed in Appendix 2.

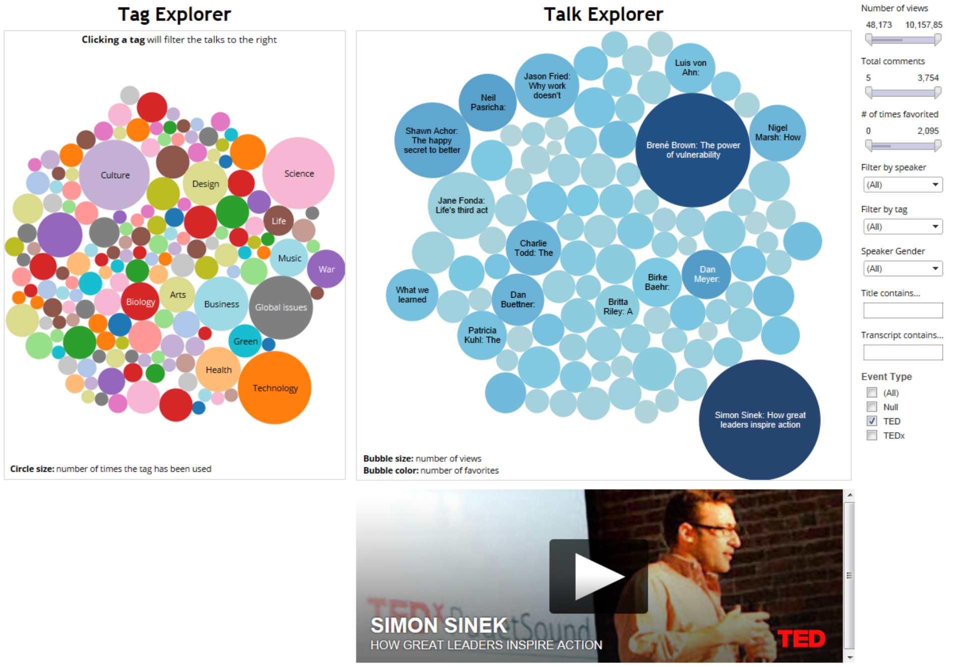
Milestone 2 - Refine Dashboards

Dashboard 1 - Discover TED

Discover TED is an interactive Tableau dashboard, which provides prominent way to view talks by topics. This dashboard was based off of the “Topics and Talks” dashboard (Figure 14) above, but with a much higher focus on finding talks rather than exploring potential relationships between quantitative metrics.

Users in our survey reported that viewing talks by topics was the most important to them. In order to cater to this need, we decided that our visualization would contain a prominent way to view talks by topic. Packed bubble plots contain several drawbacks such as the lack precise comparison of values, general lack of labels, and difficulty selecting the smaller bubbles. However, the benefits of being able to view a wide range of data in a small space were thought to outweigh these drawbacks. Additionally, a packed bubble plot seemed to invite more interactivity than a bar chart.

Based on the results of our user testing, we removed the scatterplot on Dashboard 1 and replaced bar charts with bubbles to encode topics and talks (Figure 17).


Tag Explorer
Clicking a tag will filter the talks to the right
Culture
Design
Science
Life
Music
War
Biology
Arts
Business
Global issues
Green
Health
Technology
Circle size: number of times the tag has been used
Talk Explorer
Jason Fried: Why work doesn't
Luis von Ahn:
Neil Pasricha:
Shawn Achor: The happy secret to better
Brené Brown: The power of vulnerability
Nigel Marsh: How
Jane Fonda: Life's third act
Charlie Todd: The
Birke Baehr:
Dan Meyer:
What we learned
Dan Buettner:
Britta Riley: A
Patricia Kuhl: The
Simon Sinek: How great leaders inspire action
Bubble size: number of views
Bubble color: number of favorites
SIMON SINEK
HOW GREAT LEADERS INSPIRE ACTION
TED

Number of views
48,173 10,157,85

Total comments
5 3,754

# of times favorited
0 2,095

Filter by speaker
(All)

Filter by tag
(All)

Speaker Gender
(All)

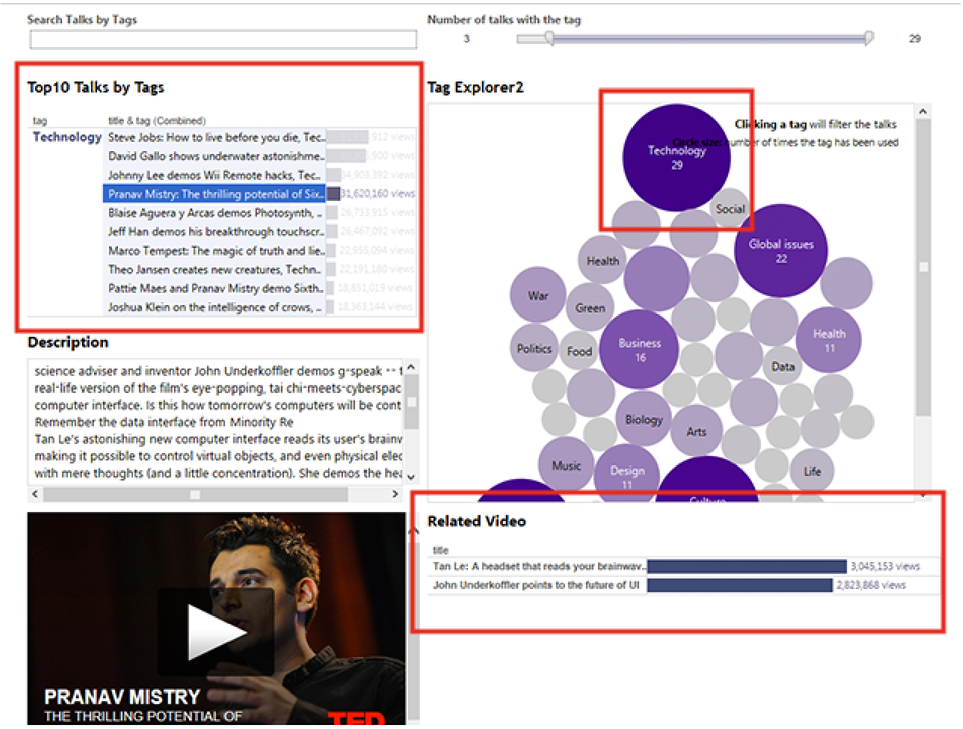
Title contains...

Transcript contains...

Event Type
☐ (All)
☐ Null
☑ TED
☐ TEDx

Figure 17. Screenshot of 1st iteration of Dashboard 1

However, using bubbles makes it difficult to see all of the talks and topics since most bubbles do not contain a text label. This means users must hover over the bubbles one by one to get information, which becomes particularly tedious when the titles of the talks themselves are encoded as bubbles. In response, we switched the encoding of talks to a bar plot and limited the number to just the top 10 (by number of views) for each topic. Additionally, we added related talks to the dashboard as users indicated this need in the interviews. We ended up with the dashboard shown in Figure 18.



Search Talks by Tags

Number of talks with the tag 3 – 29

Top10 Talks by Tags

tag	title & tag (Combined)	
Technology	Steve Jobs: How to live before you die, Tec..	[illegible]912 views
	David Gallo shows underwater astonishme..	[illegible]900 views
	Johnny Lee demos Wii Remote hacks, Tec..	34,903,392 views
	Pranav Mistry: The thrilling potential of Six..	31,620,160 views
	Blaise Aguera y Arcas demos Photosynth, ..	26,733,915 views
	Jeff Han demos his breakthrough touchscr..	26,467,092 views
	Marco Tempest: The magic of truth and lie..	22,955,094 views
	Theo Jansen creates new creatures, Techn..	22,191,180 views
	Pattie Maes and Pranav Mistry demo Sixth..	18,851,019 views
	Joshua Klein on the intelligence of crows, ..	18,363,144 views

Description

science adviser and inventor John Underkoffler demos g-speak -- t
real-life version of the film's eye-popping, tai chi-meets-cyberspac
computer interface. Is this how tomorrow's computers will be cont
Remember the data interface from Minority Re
Tan Le's astonishing new computer interface reads its user's brainv
making it possible to control virtual objects, and even physical elec
with mere thoughts (and a little concentration). She demos the hea

Tag Explorer2

Clicking a tag will filter the talks
Circle size: number of times the tag has been used

Related Video

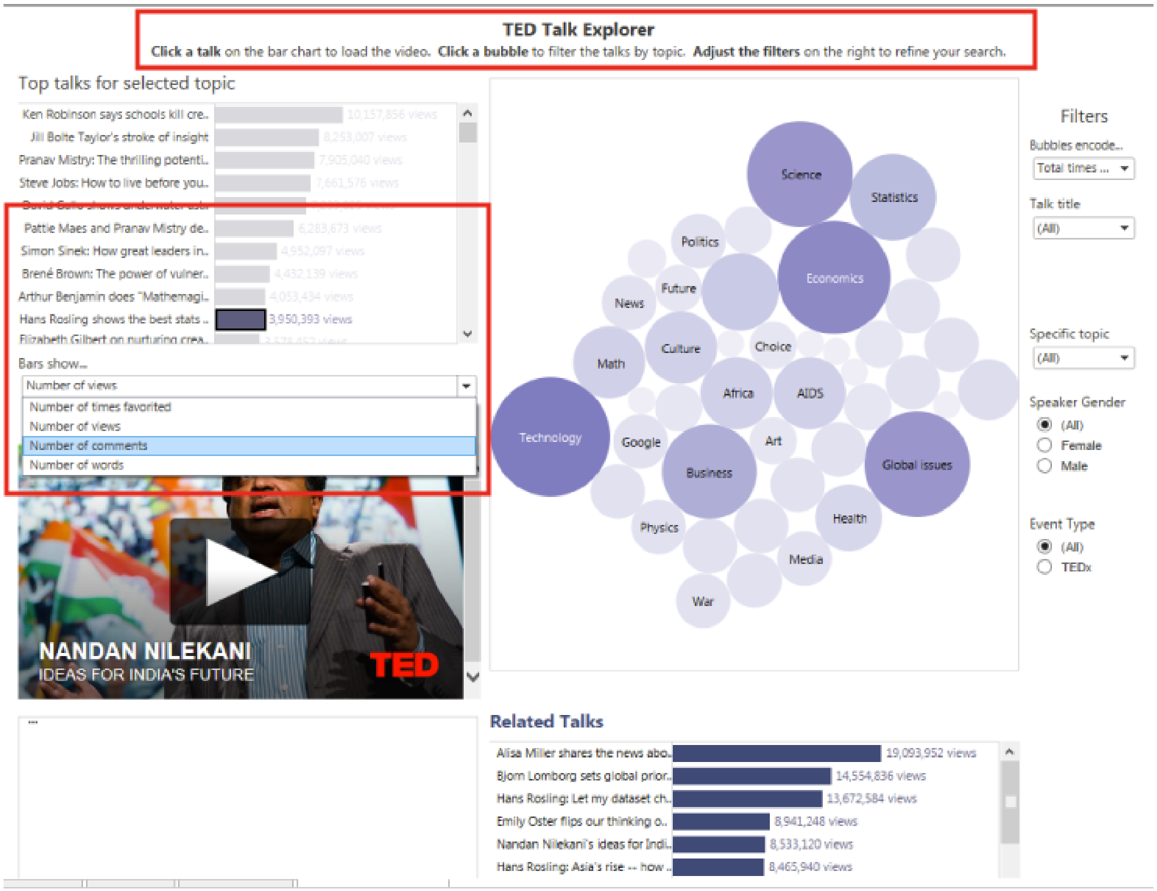
title	
Tan Le: A headset that reads your brainwav..	3,045,153 views
John Underkoffler points to the future of UI	2,823,868 views

Figure 18. Screenshot of 2nd iteration of Dashboard 1

Our user tests of this dashboard found:

- Overall, exploring talks by topics was a very useful feature for participants
- Watching video on the dashboard was appreciated by participants, although some of the links were broken
- The option to search by tags was useful for participants
- Participants would have liked to know more results instead of top 10 talks (e.g. “But what if I wanted to see more?”)
- The description did not seem to match with the video and on occasion included multiple descriptions
- The information on tooltip was confusing
- Several participants didn’t know exactly how to interact with the visualization (e.g. click a bar to load the video)

To address these problems, we removed the broken url and fixed the descriptions, as well as showed all talks ranked in descending order. In addition, we added more textual hints (both within the tooltip and right above the section) to indicate possible interactions. Considering users' needs in selecting talks by different metrics (the most important two being number of views and number of favorites), we added options to switch different metrics for bubbles and bars (Figure 19)



TED Talk Explorer

Click a talk on the bar chart to load the video. **Click a bubble** to filter the talks by topic. **Adjust the filters** on the right to refine your search.

Top talks for selected topic

Ken Robinson says schools kill cre..	10,157,856 views
Jill Bolte Taylor's stroke of insight	8,253,007 views
Pranav Mistry: The thrilling potenti..	7,905,040 views
Steve Jobs: How to live before you..	7,661,576 views
David Gallo shows underwater ast..	[illegible]
Pattie Maes and Pranav Mistry de..	6,283,673 views
Simon Sinek: How great leaders in..	4,952,097 views
Brené Brown: The power of vulner..	4,432,139 views
Arthur Benjamin does "Mathemagi..	4,053,434 views
Hans Rosling shows the best stats ..	3,950,393 views
Elizabeth Gilbert on nurturing crea..	[illegible]

Bars show...

Number of views

- Number of times favorited
- Number of views
- Number of comments
- Number of words

Filters

Bubbles encode...: Total times ...

Talk title: (All)

Specific topic: (All)

Speaker Gender

- ● (All)
- ○ Female
- ○ Male

Event Type

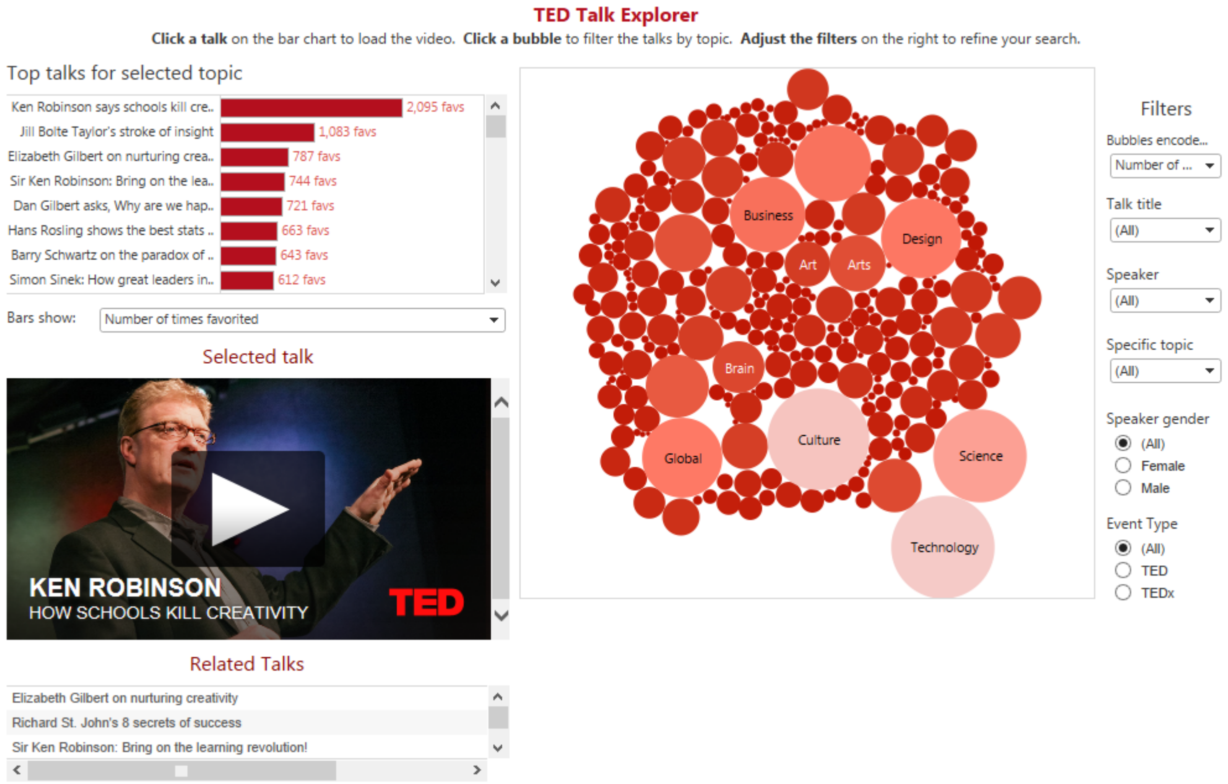
- ● (All)
- ○ TEDx

Related Talks

Alisa Miller shares the news abo..	19,093,952 views
Bjorn Lomborg sets global prior..	14,554,836 views
Hans Rosling: Let my dataset ch..	13,672,584 views
Emily Oster flips our thinking o..	8,941,248 views
Nandan Nilekani's ideas for Indi..	8,533,120 views
Hans Rosling: Asia's rise -- how ..	8,465,940 views

Figure 19. Screenshot of 3rd iteration of Dashboard 1

Considering the consistency with the branding of TED talks, we considered using the TED red as the theme color for our dashboards. To increase the usability for people with color vision impairment, we carefully selected different red colors with sufficient luminance contrast. However, the feedback from the participants were in general negative, pointing out the overall theme was too distracting and irritating for them to focus on the content. (Figure 20)



TED Talk Explorer

Click a talk on the bar chart to load the video. **Click a bubble** to filter the talks by topic. **Adjust the filters** on the right to refine your search.

Top talks for selected topic

Ken Robinson says schools kill cre..	2,095 favs
Jill Bolte Taylor's stroke of insight	1,083 favs
Elizabeth Gilbert on nurturing crea..	787 favs
Sir Ken Robinson: Bring on the lea..	744 favs
Dan Gilbert asks, Why are we hap..	721 favs
Hans Rosling shows the best stats ..	663 favs
Barry Schwartz on the paradox of ..	643 favs
Simon Sinek: How great leaders in..	612 favs

Bars show: Number of times favorited

Selected talk

Related Talks

Elizabeth Gilbert on nurturing creativity
Richard St. John's 8 secrets of success
Sir Ken Robinson: Bring on the learning revolution!

Filters

Bubbles encode...: Number of ...

Talk title: (All)

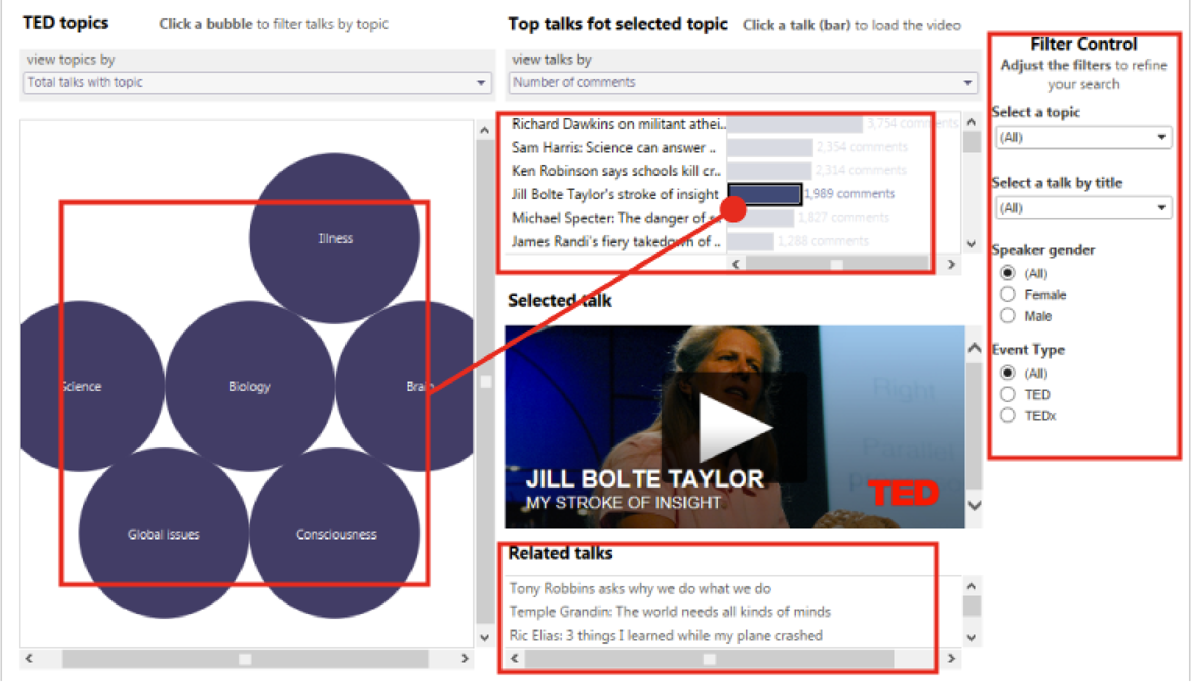
Speaker: (All)

Specific topic: (All)

Speaker gender
● (All)
○ Female
○ Male

Event Type
● (All)
○ TED
○ TEDx

Figure 20. Screenshot of 4th iteration of Dashboard 1



TED topics Click a bubble to filter talks by topic

view topics by: Total talks with topic

Top talks fot selected topic Click a talk (bar) to load the video

view talks by: Number of comments

Richard Dawkins on militant athei..	3,754 comments
Sam Harris: Science can answer ..	2,354 comments
Ken Robinson says schools kill cr..	2,314 comments
Jill Bolte Taylor's stroke of insight	1,989 comments
Michael Specter: The danger of s..	1,827 comments
James Randi's fiery takedown of ..	1,288 comments

Selected talk

Related talks

Tony Robbins asks why we do what we do
Temple Grandin: The world needs all kinds of minds
Ric Elias: 3 things I learned while my plane crashed

Filter Control
Adjust the filters to refine your search

Select a topic: (All)

Select a talk by title: (All)

Speaker gender
● (All)
○ Female
○ Male

Event Type
● (All)
○ TED
○ TEDx

Figure 21. Screenshot of 5th iteration of Dashboard 1

Replacing the TED red themes (Figure 21), we then did another round of evaluation with the dashboard, and found that:

- The textual hints were clear and helpful
- Participants were confused to see the topics “back filtered” when selecting a specific talk
- Participants would have liked to know how many total results were returned when searching topics (how many talks they have access to for a selected topic)
- Participants did not realize that the related talks were only related to the selected talk
- Participants prefered to use the filter control on the left
- Some of them did not know the difference between TED and TEDx (event type)

To fix these problems, we lowered the hierarchy of “Related Talks” by decreasing its title’s font size to indicate its relation with “Selected Talks”; besides, we added the number of accessible talks for a selected topic; we also moved the filter control to the left and added a short description of the event type. This resulted in our final design of Dashboard 1, which will be introduced in detail later in the next section.

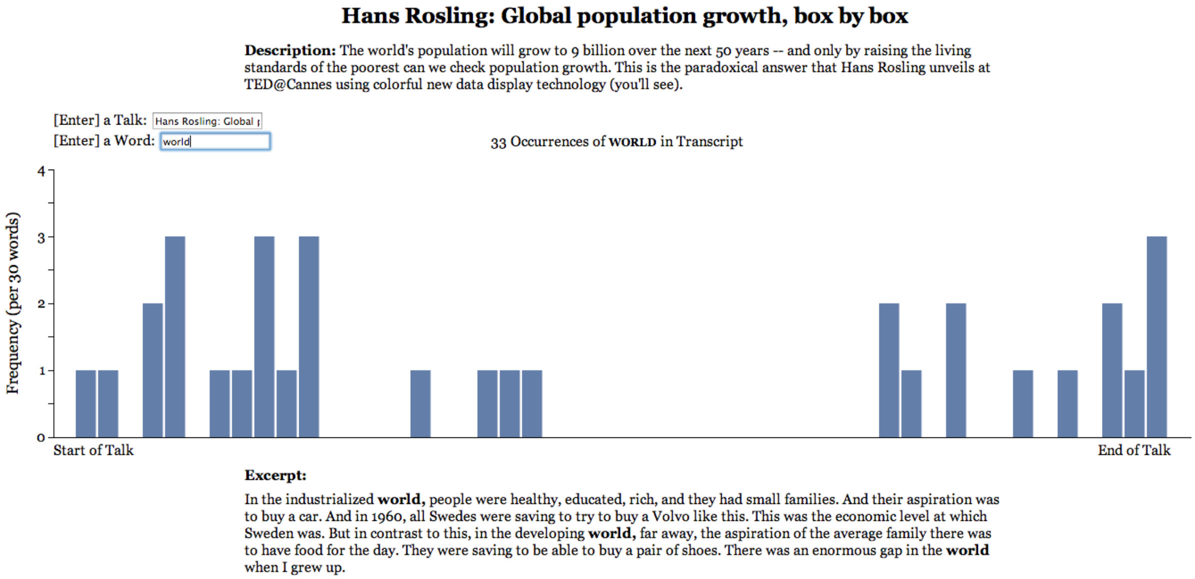
As potential users did not feel strongly about a introductory guided tour of the visualization (informed by both the interview and online survey), we decided that it was not important to create one. Instead, we hoped that placing contextual hints would give users the understanding they needed to explore the visualization.

Dashboard 2: WordView

Based on the prior task analysis and survey, we identified a need to allow users to explore content within a talk. Given that the TED dataset contained transcripts, we developed a word concordance visualization called WordView. At the overview level, WordView allows users to identify frequency of words within a talk. At the zoom-and-filter level, users are able to select specific words that they would like to see concordant with the TED Talk. At the details-on-demand level, users are able to view an excerpt of the transcript within the context of the TED Talk.

We selected D3 to build the WordView visualization due to the added expressiveness of a visualization grammar compared to the visual analysis language of Tableau (Heer, 2014). D3 also provided greater flexibility in dealing with the dynamic calculations required to work with a large transcript based dataset. As part of preprocessing, we calculated the word frequencies within each transcript (removing stop words such as “a”, “and”, “the”, “or”, etc). We stored this information, along with the full transcript, talk description, and talk title as JSON files for each TED Talk.

The WordView visualization is inspired by Lee and Venkatesh’s Bibly concordance visualization of the Bible. Users select a TED Talk by title and then search for words within the transcript of that given talk. Both the talk and word search bars are implemented as autocomplete dropdown menus using the jQuery UI. Once a word has been selected, the frequency occurrence of that word is shown on a bar plot. The transcript is divided into 30 word segments from which frequency of word occurrence is calculated. Hovering on the bars provides a transcript excerpt containing the word of interest emphasized (Figure 22).



Hans Rosling: Global population growth, box by box

Description: The world's population will grow to 9 billion over the next 50 years -- and only by raising the living standards of the poorest can we check population growth. This is the paradoxical answer that Hans Rosling unveils at TED@Cannes using colorful new data display technology (you'll see).

[Enter] a Talk: Hans Rosling: Global p
[Enter] a Word: world

33 Occurrences of **WORLD** in Transcript

Excerpt:

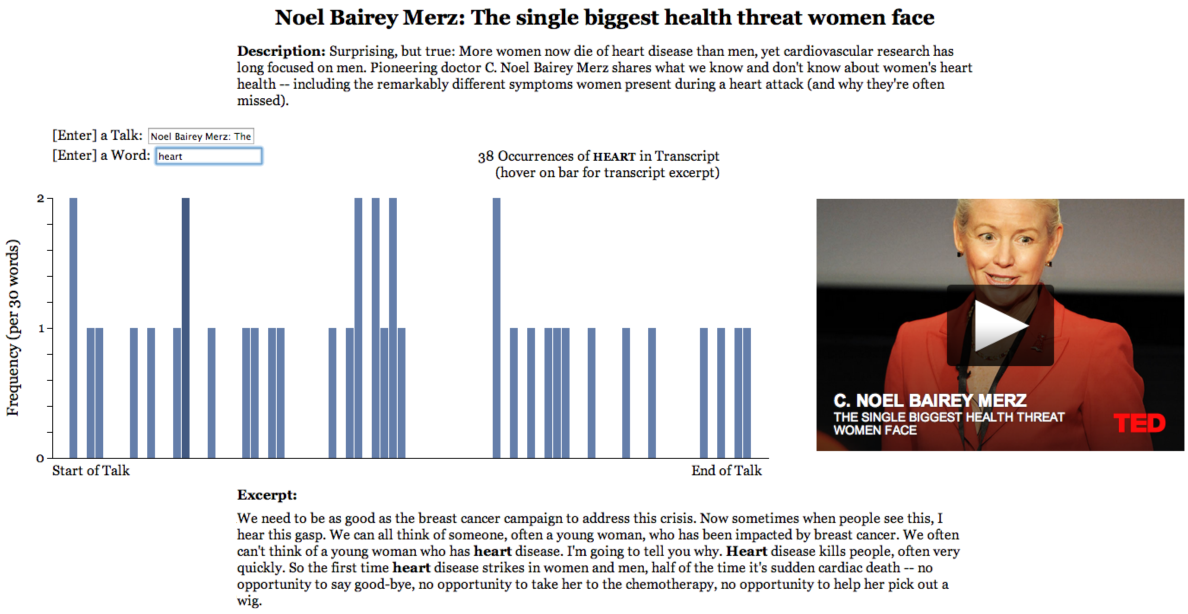
In the industrialized **world,** people were healthy, educated, rich, and they had small families. And their aspiration was to buy a car. And in 1960, all Swedes were saving to try to buy a Volvo like this. This was the economic level at which Sweden was. But in contrast to this, in the developing **world,** far away, the aspiration of the average family there was to have food for the day. They were saving to be able to buy a pair of shoes. There was an enormous gap in the **world** when I grew up.

Figure 22. Screenshot of the 1st iteration of Dashboard 2

As part of the iterative design process, we conducted user testing on the WordView visualization. The primary findings from the study included:

- Users did not initially discover that hovering on bars provided a transcript excerpt.
- When entering a TED Talk title into the search bar, users had to know the exact title of a talk in order to find a match.
- Though participants understood the graphical representation and found it a fun experience, they also noted the limited utility of the tool.

To address each of these challenges within the second iteration (Figure 23), we first embedded the TED Talk video alongside the word concordance graph. This allows users to interact with the graphical visualization while retaining a tangible connection to the TED Talk video. We also addressed the issues of recognition versus recall for the Talk Title search bar by changing the search to match by characters throughout the title string rather than require a first n-character match. We also provided instructions for the user to hover on the bars as a way to view the transcript excerpt.



Noel Bairey Merz: The single biggest health threat women face

Description: Surprising, but true: More women now die of heart disease than men, yet cardiovascular research has long focused on men. Pioneering doctor C. Noel Bairey Merz shares what we know and don't know about women's heart health -- including the remarkably different symptoms women present during a heart attack (and why they're often missed).

[Enter] a Talk: Noel Bairey Merz: The

[Enter] a Word: heart

38 Occurrences of HEART in Transcript
(hover on bar for transcript excerpt)

Excerpt:

We need to be as good as the breast cancer campaign to address this crisis. Now sometimes when people see this, I hear this gasp. We can all think of someone, often a young woman, who has been impacted by breast cancer. We often can't think of a young woman who has **heart** disease. I'm going to tell you why. **Heart** disease kills people, often very quickly. So the first time **heart** disease strikes in women and men, half of the time it's sudden cardiac death -- no opportunity to say good-bye, no opportunity to take her to the chemotherapy, no opportunity to help her pick out a wig.

Figure 23. Screenshot of the 3rd iteration of Dashboard 2

User testing on this iteration of WordView provided generally positive results with the interactions. However a problem consistently identified by users was the lack of connection between the concordance graph and the TED Talk video. Users expected to go to specific segments of the TED Talk where the transcript excerpts occurred. This would add overall utility to the visualization as a tool to not only identify frequently occurring words and their location within the talk, but also to jump to segments of interest within a TED Talk video.

We were limited in addressing the user needs identified within the second iteration of testing due to the currently available dataset, which does not provide timestamps associated with the transcript. However, within the TED Talk website, timestamps are provided for transcripts. We iterated on the WordView visualization with a demonstration of feasibility for integrating the concordance graph with the TED Talk video. This first involved manually indexing the timestamps found on the TED Talk website with their location in the transcript. The timestamps then provide a reference from which transcript excerpts are linked to location within the TED Video. This iteration of WordView allows users to go to segments of the TED Talk video identified within the transcript excerpt, which will be discussed in detail in next section.

Dashboard 3 - Dogeify TED talks

While the main goal of the project was to create something that was functional and useful when looking for a TED talk, we also wanted to add an element of humor and delight to the project. To accomplish this we created Dogeify TED. Dogeify TED aims to use the transcript data to create a Doge meme picture for every talk in the database dynamically.

With the trimmed JSON file generated using data preprocessing as a data source, we used JQuery, HTML, and CSS to generate the DOGE meme image. Text color, placement, and content are randomized each time the image is loaded, leading to mostly amusing results.

The very first iteration of the Dogeify TED (figure 24) was created using the raw transcripts as a source for the data. This meant that the source 130-megabyte JSON file was loaded to access the transcripts, making it impractical to host over the web. To solve this problem the JSON file was trimmed during pre-processing as mentioned above. We replaced the transcripts with word frequencies of words in the transcript and a list of the top ten words by frequency. The image was rendered on the canvas element in HTML 5.0. Text color and placement was randomised and 'Comic Sans MS' (as is the norm for doge memes) was used as the font. As you can see in the image below the first iteration was not styled.

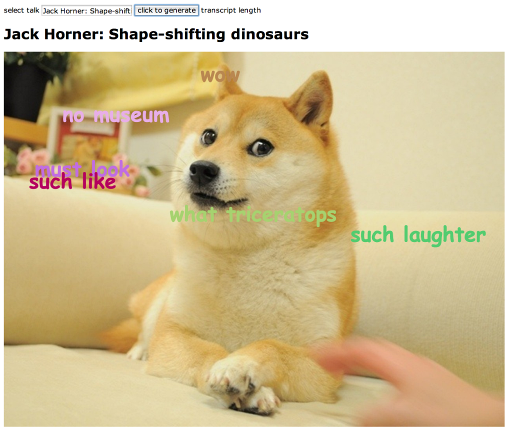


Figure 24. Iteration 1 of Dogeify TED

The second iteration (figure 25) used a more optimised algorithm to generate the picture. The page was also styled using CSS to better suit the theme of TED talks and also of doge. To add a little more value to the dashboard we also embedded the video like we have in the other dashboards (figure 26).

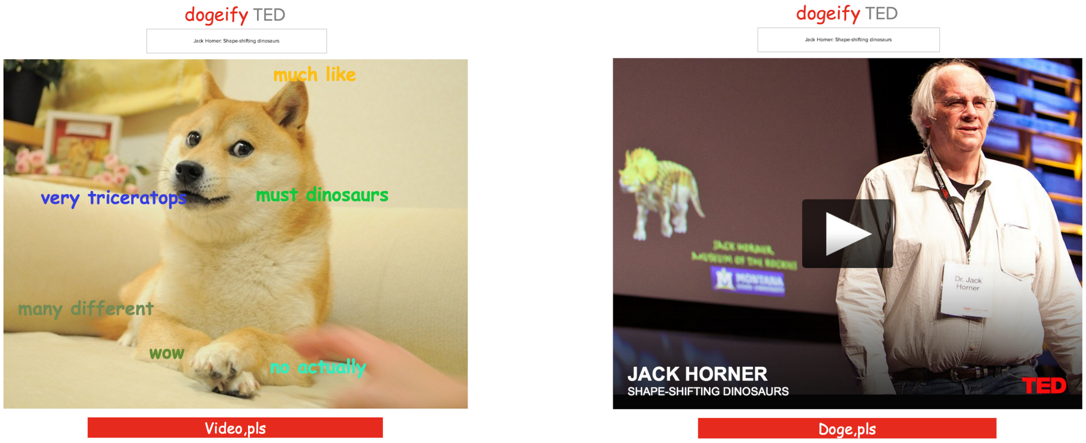


Figure 25 & 26. Iteration 2 of Dogeify TED.

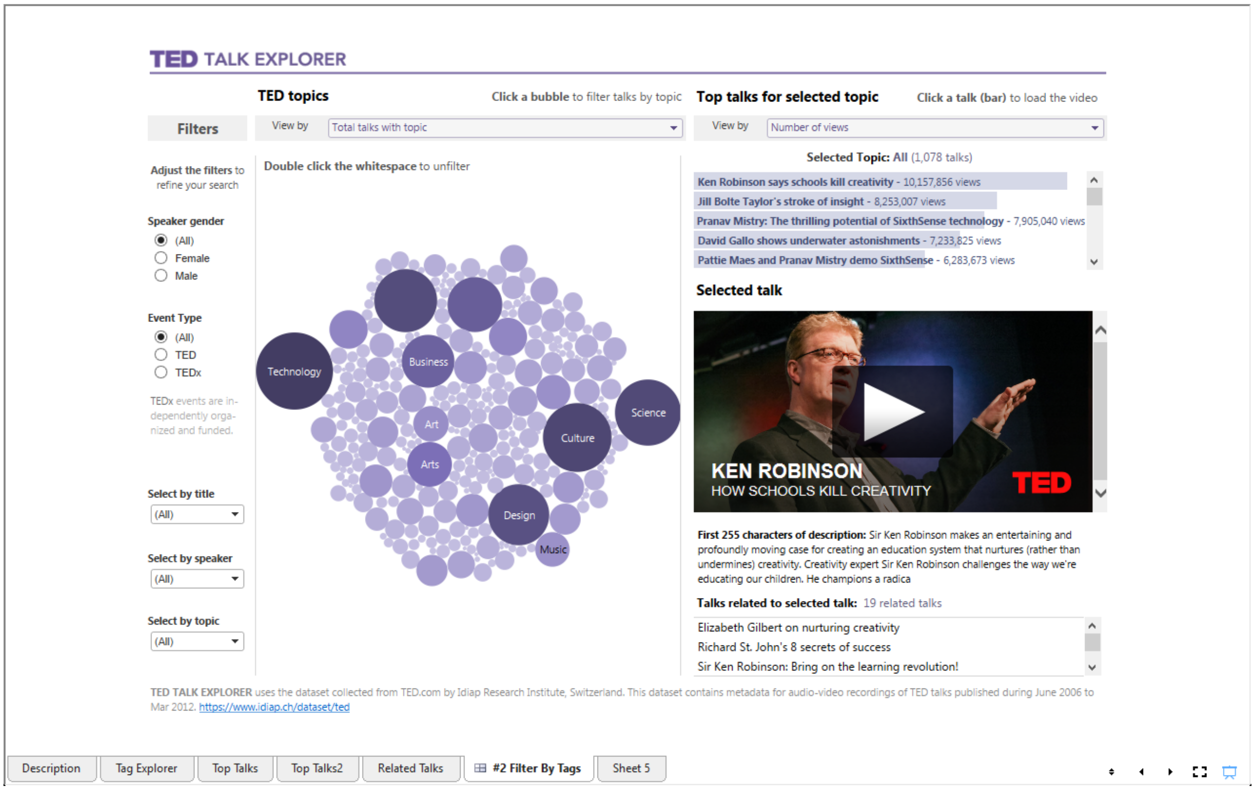
The text on the button to toggle between the photo and the video was kept deliberately trivial to stick with the theme. This iteration was used as the final one as we focused more on WordView and TED Explorer.

Final evaluation of visualizations

Dashboard 1 - Discover TED

The final design of Dashboard 1 is shown in Figure 27. Its main goal is to let users explore talks by topics, which is highest demand according to the user studies. The dashboard contains four major areas, the TED topics, the talks for select topics, the video section to load a selected talk and the filter control. The tasks supported by the visualization include:

- Identifying the ranking of topics by different metrics
- Sorting the talks by different metrics
- Selecting and watching a talk video by topics, speakers and their gender, event types
- Reading description and finding related talks for a specific talk



TED TALK EXPLORER

TED topics

Click a bubble to filter talks by topic

Filters

View by: Total talks with topic

Adjust the filters to refine your search

Speaker gender: ● (All) ○ Female ○ Male

Event Type: ● (All) ○ TED ○ TEDx

TEDx events are independently organized and funded.

Select by title: (All)

Select by speaker: (All)

Select by topic: (All)

Double click the whitespace to unfilter

Top talks for selected topic

Click a talk (bar) to load the video

View by: Number of views

Selected Topic: All (1,078 talks)

Ken Robinson says schools kill creativity - 10,157,856 views
Jill Bolte Taylor's stroke of insight - 8,253,007 views
Pranav Mistry: The thrilling potential of SixthSense technology - 7,905,040 views
David Gallo shows underwater astonishments - 7,233,825 views
Pattie Maes and Pranav Mistry demo SixthSense - 6,283,673 views

Selected talk

First 255 characters of description: Sir Ken Robinson makes an entertaining and profoundly moving case for creating an education system that nurtures (rather than undermines) creativity. Creativity expert Sir Ken Robinson challenges the way we're educating our children. He champions a radica

Talks related to selected talk: 19 related talks

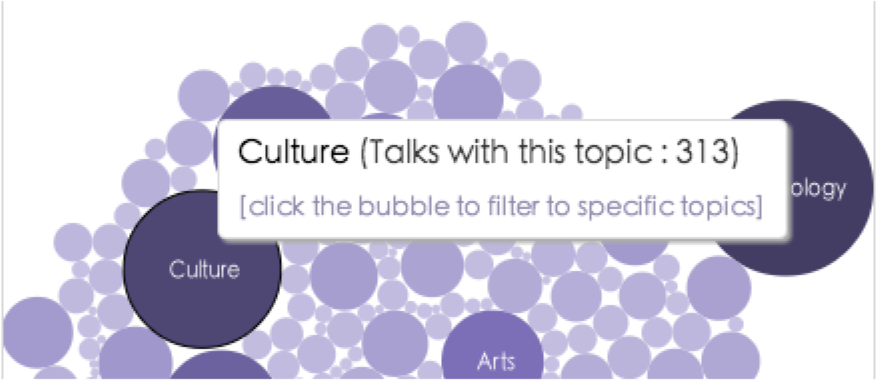
Elizabeth Gilbert on nurturing creativity
Richard St. John's 8 secrets of success
Sir Ken Robinson: Bring on the learning revolution!

TED TALK EXPLORER uses the dataset collected from TED.com by Idiap Research Institute, Switzerland. This dataset contains metadata for audio-video recordings of TED talks published during June 2006 to Mar 2012. https://www.idiap.ch/dataset/ted

Figure 27. Screenshot of Dashboard 1

Overview and Detail Views

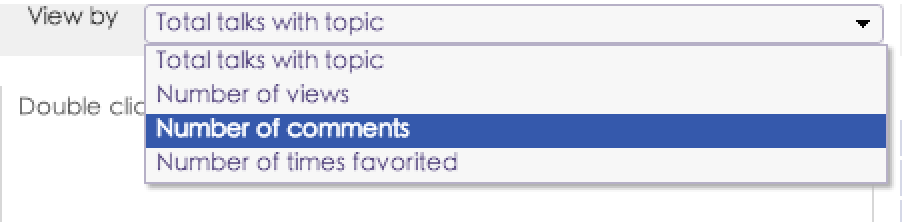
According to Schneiderman (1996), visualizations should start with an overview of the data. As an overview for the dashboard, the TED topics are represented by packed bubbles. As users care more about the topics with a higher rank and their relative popularity, using bubbles to encode quantitative values seem to be a fair choice. In terms of details-on-demand, hovering over the bubbles will show the number of talks associated with the topic, as well as a contextual hint indicating possible interaction (Figure 28). Instead of using different attributes to encode bubbles, we provide an option for users to select different metrics to visualize bubbles and double-encode them with both color and size (Figure 29). Therefore, the bubbles only convey a single dimension at a time, which attenuates the cognitive load for users, particularly those who with color vision deficiency (Few 2012, Chapter 6).



Culture (Talks with this topic : 313)

[click the bubble to filter to specific topics]

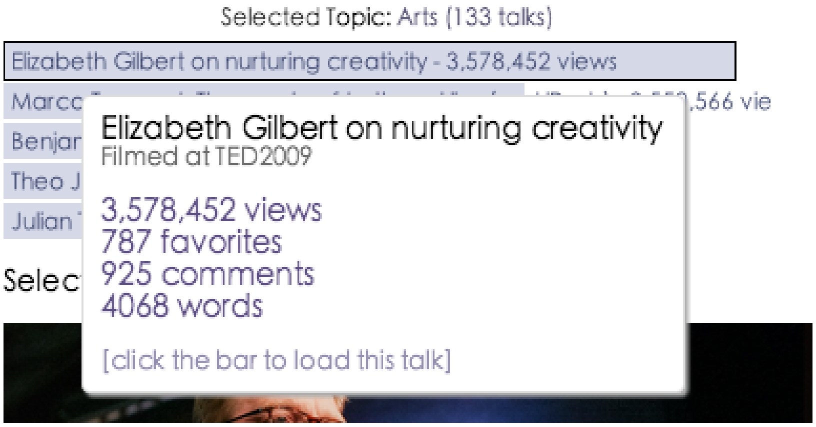
Figure 28. Details-on-demand & contextual hint



View by Total talks with topic

Figure 29. Option of switching between different metrics

For the list of talks by selected topic, we provide the option for users to select different metrics to rank the talks by. Users liked the idea of showing the top talks for a given metric and topic. Therefore, we rank the bars in a descending order according to the metrics selected. An additional benefit of using bar charts for this ranking is that bars are one of the best choices to encode quantitative data in ranking graphs (Few 2012, Chapter 6). In order to support details-on-demand, hovering over a bar will cause the tooltip to show information related to the talk (e.g. the title, date filmed, number of views, number of favorites, number of comments, and number of words), as well as a contextual hint indicating possible interaction with the bars (Figure 30).



Selected Topic: Arts (133 talks)

Elizabeth Gilbert on nurturing creativity - 3,578,452 views

Elizabeth Gilbert on nurturing creativity
Filmed at TED2009

3,578,452 views
787 favorites
925 comments
4068 words

[click the bar to load this talk]

Figure 30. Details-on-demand & contextual hints

The “Selected Talk” section contains three parts: the video, the description of the talk, and the other talks related to the selected talk. The description section only contains the first 255 characters of the description due to a limitation of the source dataset. The “Related Talks” section contains text, but no bars, so that it is visually differentiated from the “Talks for Selected Topic” section above.

Interactions

- Selecting a bubble will filter the talks associated with the topics on the right
- Selecting a specific talk from the bar chart will load the video, description, and related talks for the selected talks
- Filters are global, meaning they control both the topics and selected talks
- Topic, talk, and author filters allow users to search by typing keywords, or selecting from the dropdown list

Visual Design

Considering the users with color vision deficiency, we customize the color palette for large contrast in luminance (Stone, 2014). We removed redundant border lines of each section and used light gray to indicate functional areas (such as the filter control and metrics option box) to maximize the data-ink-ratio (Tufte, 1983).

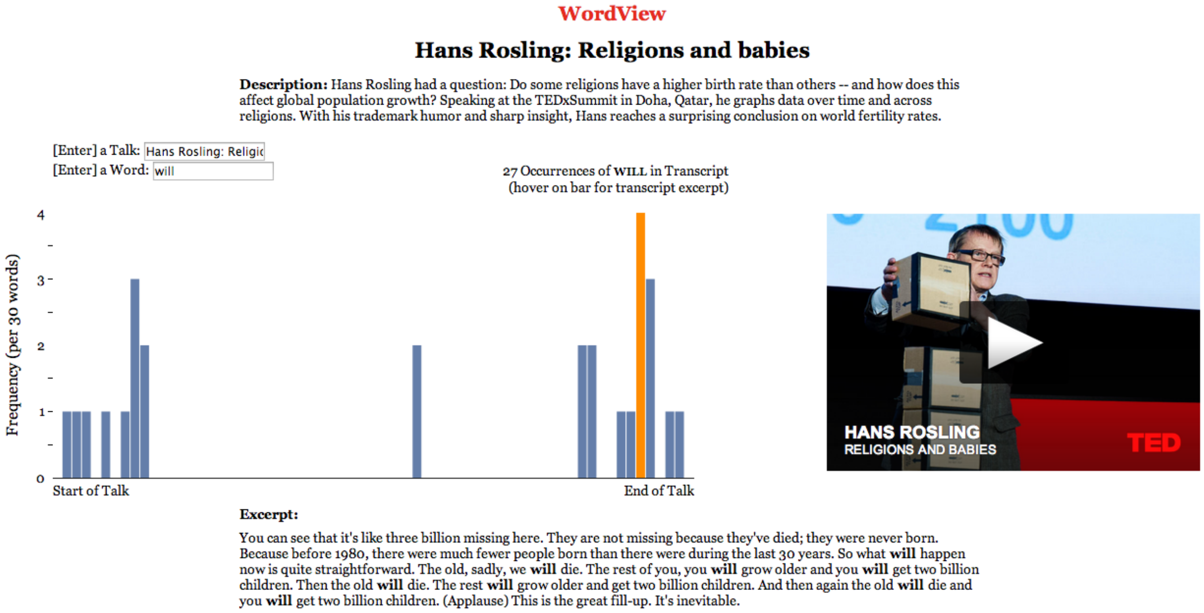
Limitation

There are several limitations of the dashboard. The first is that the description and the related videos can only be updated by clicking a particular talk. In addition, the visualization must load a particular talk by default, which may confuse users as to why a talk, description, and related videos appear without user action. The second limitation is the lack of detail in using packed bubble charts. A zoom feature would probably erase this problem but is not supported by Tableau. Other solutions include using D3 and JavaScript to integrate an updated section with the dashboard. The third limitation, which has accompanied us throughout the design process, is how to indicate possible interactions and provide more interactions within the dashboard. For example, the dashboard doesn't provide an easily understandable way for users to unfilter their selection to return to a data overview.

Dashboard 2 - Wordview

The WordView dashboard has undergone several iterations to improve its utility and usability (Figure 31). Based on the initial user studies, we designed a visualization that allows users to quickly identify content through a talk based on word concordance. The tasks supported in the visualization include:

- Searching for TED Talks by matching strings within the title
- Identifying which words frequently occur within a transcript and where in the transcript they occur
- Examining excerpts of the transcript containing keywords of interest
- Going to segments of the TED Talk video containing an excerpt



WordView

Hans Rosling: Religions and babies

Description: Hans Rosling had a question: Do some religions have a higher birth rate than others -- and how does this affect global population growth? Speaking at the TEDxSummit in Doha, Qatar, he graphs data over time and across religions. With his trademark humor and sharp insight, Hans reaches a surprising conclusion on world fertility rates.

[Enter] a Talk: Hans Rosling: Religic

[Enter] a Word: will

27 Occurrences of WILL in Transcript
(hover on bar for transcript excerpt)

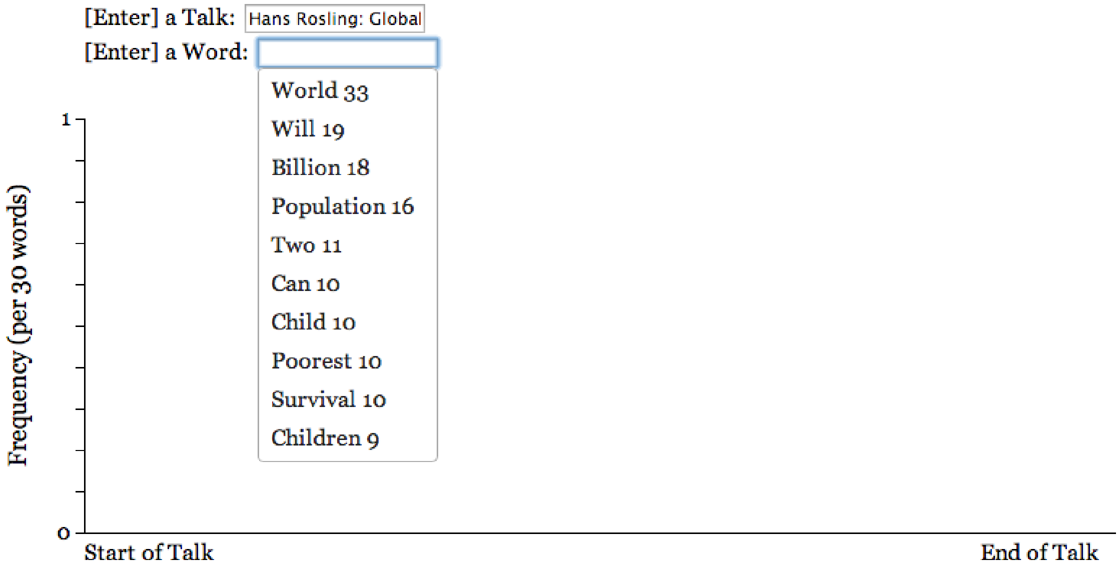
Excerpt:

You can see that it's like three billion missing here. They are not missing because they've died; they were never born. Because before 1980, there were much fewer people born than there were during the last 30 years. So what **will** happen now is quite straightforward. The old, sadly, we **will** die. The rest of you, you **will** grow older and you **will** get two billion children. Then the old **will** die. The rest **will** grow older and get two billion children. And then again the old **will** die and you **will** get two billion children. (Applause) This is the great fill-up. It's inevitable.

Figure 31. Screenshot of Dashboard 2

Overview and Detail Views

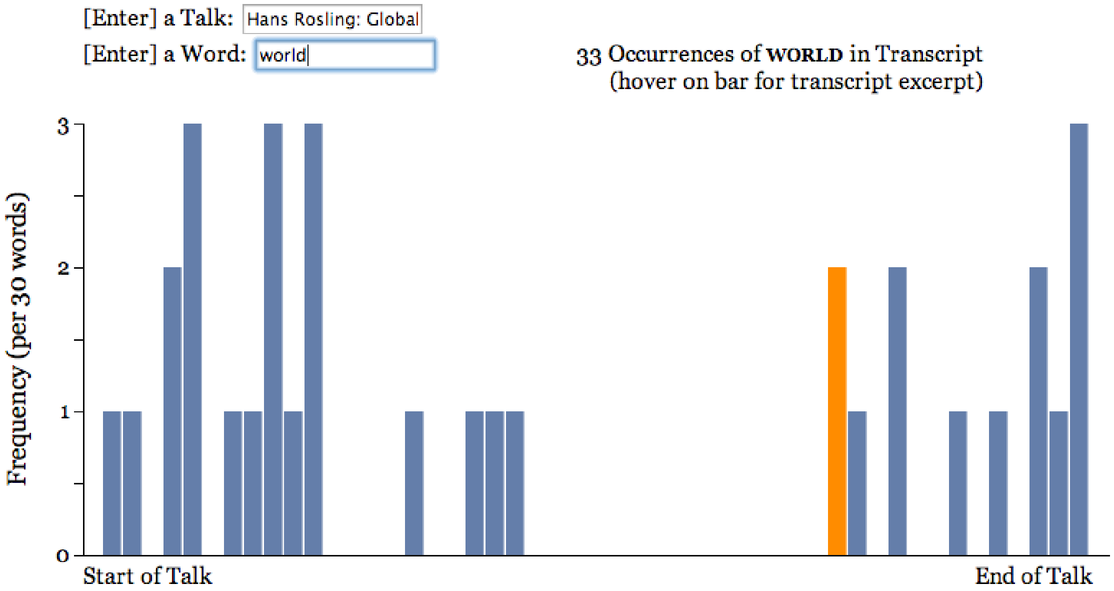
At an overview level, users are presented with a title and description of a talk. This visualization is developed for users who have a particular talk in mind to explore. A search bar prompts users to enter a word of interest. Alternatively, clearing out the word search bar will generate a dropdown list of the top ten most frequently occurring words in the transcript (Figure 32). When a word has been selected, the empty graph populates as a bar chart. We encode time on the transcript as length on the x-axis while frequency of occurrence for a word is encoded as length on the y-axis. Each bar represents a 30 word partition of the transcript, with the height of the bar encoding the number of times a word was spoken in that 30 word partition. This overview allows users to quickly visualize the landscape of the the TED Talk based on words concordant within the transcript (Figure 33). Within our evaluation, users are able to effectively interpret the concordance visualization. They are able to recognize that the bars represent locations within the talk where a keyword was spoken. Users were less clear on how to identify which words were most frequently spoken, only realizing the most frequent terms were displayed on the search bar after clearing out a previously searched word.



[Enter] a Talk: Hans Rosling: Global

[Enter] a Word:

Figure 32. Dropdown list of the top ten most frequently occurring words

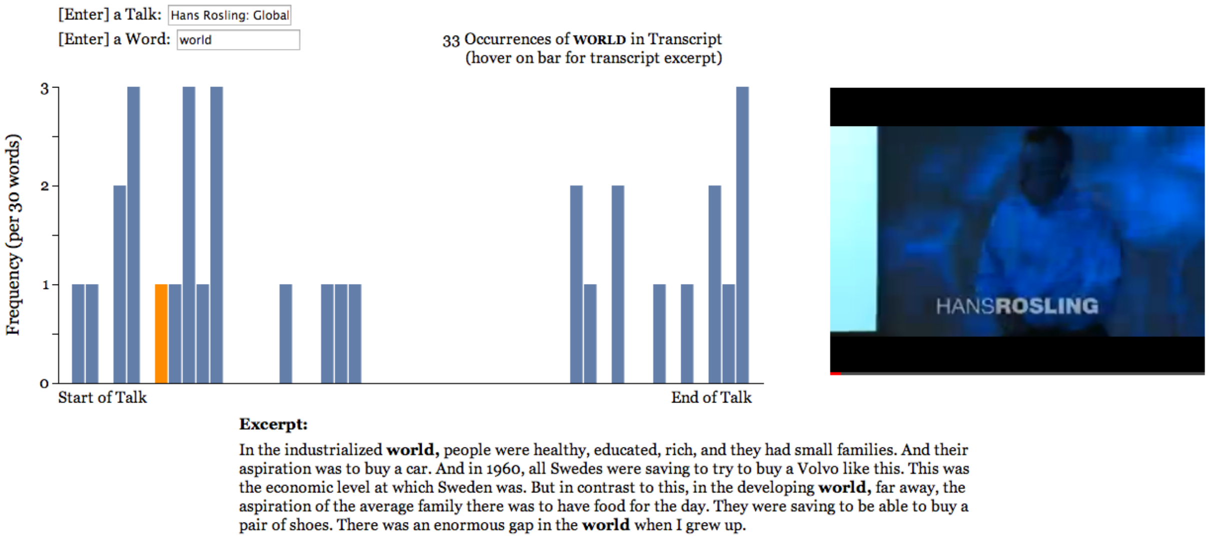


[Enter] a Talk: Hans Rosling: Global

[Enter] a Word: world

33 Occurrences of WORLD in Transcript
(hover on bar for transcript excerpt)

Figure 33. Overview of words concordant


[Enter] a Talk: Hans Rosling: Global

[Enter] a Word: world

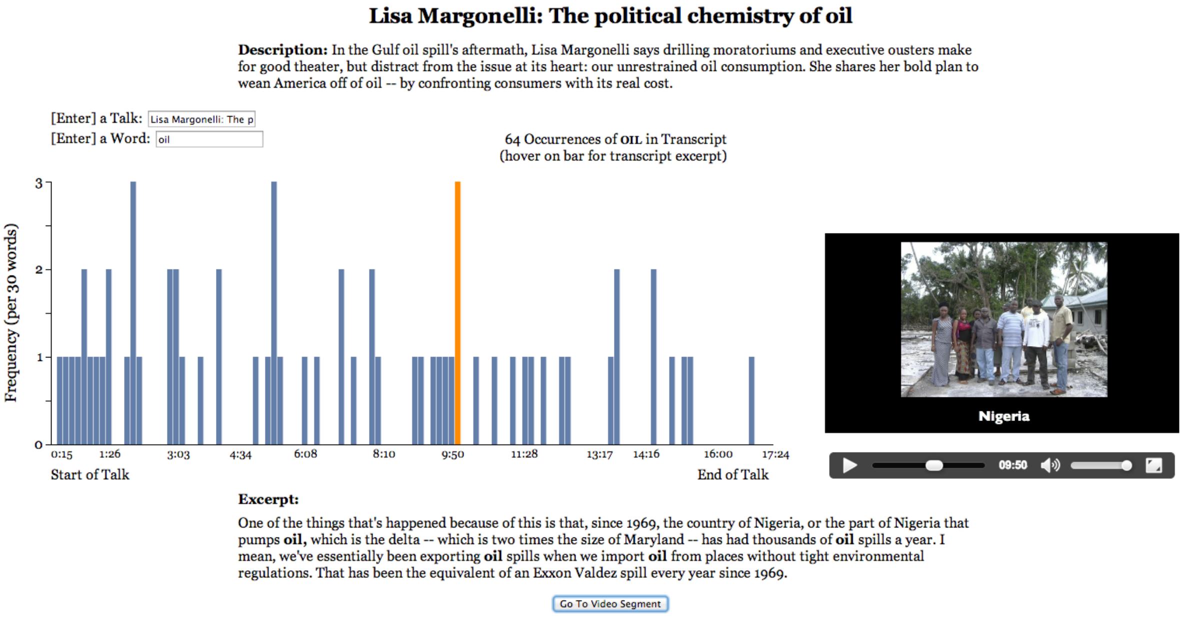
33 Occurrences of WORLD in Transcript
(hover on bar for transcript excerpt)
Frequency (per 30 words)
0
1
2
3
Start of Talk
End of Talk
HANSROSLING

Excerpt:

In the industrialized **world,** people were healthy, educated, rich, and they had small families. And their aspiration was to buy a car. And in 1960, all Swedes were saving to try to buy a Volvo like this. This was the economic level at which Sweden was. But in contrast to this, in the developing **world,** far away, the aspiration of the average family there was to have food for the day. They were saving to be able to buy a pair of shoes. There was an enormous gap in the **world** when I grew up.

Figure 34. Hovering over bars for excerpt of transcript

At a details-on-demand level, users can further view the context in which words are spoken by hovering over the frequency bars. This generates an excerpt of the transcript below the graph (Figure 33). The words are bolded for visual prominence and we encoded the hover interaction with color (orange). We made the hover interaction persistent so that the transcript excerpt is linked to the location of the talk on the graph. As a result, the colored orange encoding remains after the user has moused out of the bar of interest (unless they mouse over another bar). We also relate the concordance graph with the TED Talk video by providing an option for users to go to segments of the video containing the transcript excerpt. Evaluations with users found that the detailed view was a helpful resource to relate word frequency (a more abstract metric) with contextual information such as the transcript excerpt or TED Talk video. Users enjoyed the ability to navigate to different segments of the TED Talk, making it a valuable resource for quickly viewing a segment of interest within the talk. Users did express initial confusion however when the transcript excerpt was displayed and the number of words highlighted in the excerpt did not match the frequency on the bar chart. This is due to the 30 word partition used to define word frequency. The excerpts contain more than thirty words and as result may show highlighted words from adjacent bar segments.



Lisa Margonelli: The political chemistry of oil

Description: In the Gulf oil spill's aftermath, Lisa Margonelli says drilling moratoriums and executive ousters make for good theater, but distract from the issue at its heart: our unrestrained oil consumption. She shares her bold plan to wean America off of oil -- by confronting consumers with its real cost.

[Enter] a Talk: Lisa Margonelli: The p

[Enter] a Word: oil

64 Occurrences of OIL in Transcript
(hover on bar for transcript excerpt)

Excerpt:

One of the things that's happened because of this is that, since 1969, the country of Nigeria, or the part of Nigeria that pumps **oil,** which is the delta -- which is two times the size of Maryland -- has had thousands of **oil** spills a year. I mean, we've essentially been exporting **oil** spills when we import **oil** from places without tight environmental regulations. That has been the equivalent of an Exxon Valdez spill every year since 1969.

Figure 35. Using transcript excerpt to direct to the associated segment of the video

Limitation

Future steps for WordView involve generalizing the third iteration across all TED Talk videos. To do this we must address two challenges: 1) automating the indexing of timestamps with location in the transcript, and 2) connecting directly with the TED Talk video (currently the video is downloaded and hosted on the local web server to allow for the seeking function). Both these challenges can be addressed in the future by working directly with the TED Talk API. A

Dashboard 3 - Dogeify TED

Dogeify TED wasn't something that we planned to extensively test as it was just a value addition and not a key for the project to succeed. But we did note that other people who were shown this had a positive reaction to this. We were able to induce a smile, a chuckle and in some cases even a laugh from the people who used this. Thus we feel that Dogeify TED served its purpose.

Plans for the future

Word Frequency Data

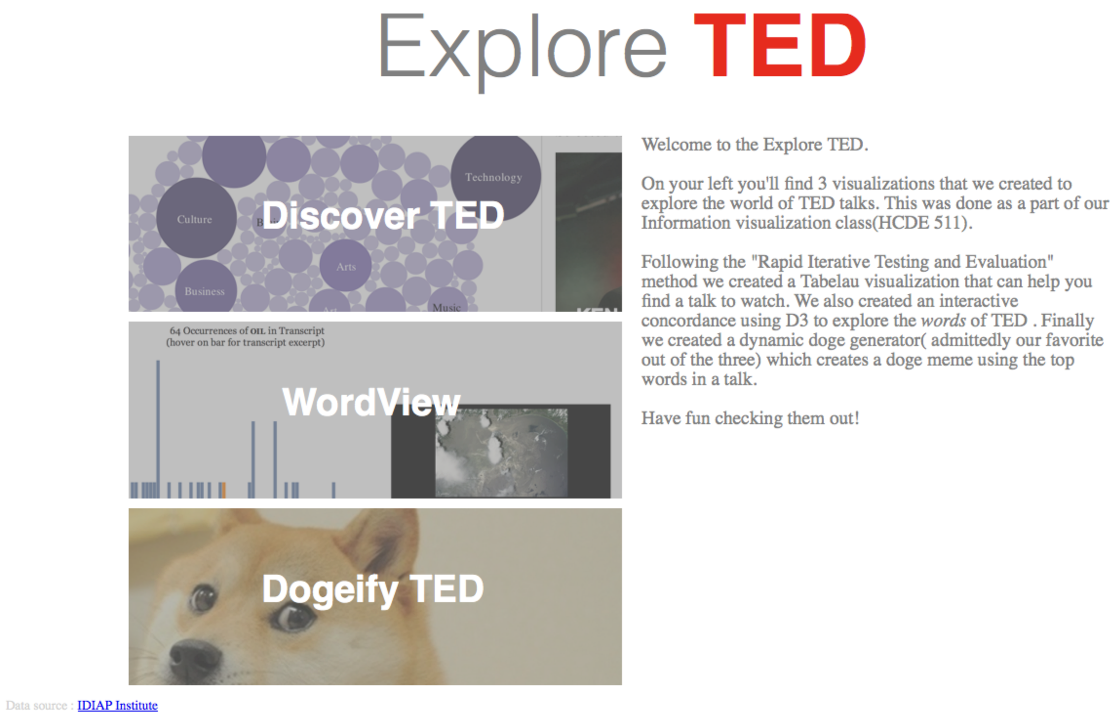
Due to our inability to wrangle with the large amount of data that was generated using the word frequency we were unable to import into Tableau in a meaningful way. We plan to create another dashboard which can use that data to allow people to search for talks by the actual words spoken in them. This will open up a new form of exploration. On the analysis side, we can use this data to find the popularity of words over time broken down by themes and topics. This can help us create some compelling textual visualizations.

Time Data

The dataset we used had details on the date talks were filmed and published. This data was not utilized right now as user testing revealed that this was not something that most people used browse talks by. We hope that we can find a way to incorporate this data in a meaningful way into the visualization so that instead of remaining unused it can add value to the visualization.

Publish Online

We plan to publish this project online. Before we can do this we need to refine the landing page that we created (Figure 36). Code written to generate the doge also needs to be optimized and cleaned up.



Welcome to the Explore TED.

On your left you'll find 3 visualizations that we created to explore the world of TED talks. This was done as a part of our Information visualization class(HCDE 511).

Following the "Rapid Iterative Testing and Evaluation" method we created a Tabelau visualization that can help you find a talk to watch. We also created an interactive concordance using D3 to explore the *words* of TED . Finally we created a dynamic doge generator(admittedly our favorite out of the three) which creates a doge meme using the top words in a talk.

Have fun checking them out!

Data source : IDIAP Institute

Figure 36. Current landing page of TED Talk Explorer

Use the TED API

The dataset we used indexed talks published up to 2012. Using the TED api that is publicly accessible it is possible to make the dataset current and make it update automatically. This is essential to have a comprehensive visualization. It will also help us solve some of the issues we have had with missing urls and incomplete description data.

Implement the Final WordView dashboard

The video segment seeking was done manually to show that it was feasible. Using the API we want to have this data available for all of the videos.

Use the Tableau Javascript API

We were unable to toggle between different chart types in a dashboard. As per Polle's advice we plan to use Tableau's javascript API to make this switch possible and also enable some advanced interactions which will solve the problems we face with selecting bubbles in our final TED Explorer dashboard.

Acknowledgments

This project would not have been possible without the valuable guidance provided by Professor Aragon and Zellweger. The advice given by them helped make a good project into a great one. The quick replies to our emails by the professors and the TA Daniel made sure that we were able to complete this project within a tight schedule.

We also appreciate that the IDIAP institute allowed us to use this dataset to create our project, and that Lee and Venkatesh made their code for Bibly available to us.

References

Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc.

Cao, Juan, et al. "Tracking web video topics: Discovery, visualization, and monitoring." Circuits and Systems for Video Technology, IEEE Transactions on21.12 (2011): 1835-1846.

Few, S. (2012). *Show me the numbers: Designing Tables and Graphs to Enlighten* (Vol. 2). Oakland, CA: Analytics Press.

Heer, J. (2014). *Visualization Tools*. Lecture: HCDE 511.

Marti A. Hearst. *Search User Interfaces*, Cambridge University Press, 2009. Chapter 11: Information Visualization for Text Analysis

Pappas, N., & Popescu-Belis, A. (2013, June). Combining content with user preferences for TED lecture recommendation. In *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on* (pp. 47-52). IEEE.

Pappas, N., & Popescu-Belis, A. (2013, July). Sentiment analysis of user comments for one-class collaborative filtering over TED talks. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 773-776). ACM.

Shao, Jian, et al. "A unified framework for web video topic discovery and visualization." Pattern Recognition Letters 33.4 (2012): 410-419.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (pp. 336-343). IEEE.

Stone, M (2014). *Expert Color Choices.* Lecture: HCDE 511.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics press.

Appendices

Appendix 1: Interview questions
Part 1:
How often do you watch TED Talks?
How do you currently go about selecting a TED talk?
What makes a TED talk interesting to you?
How relevant are these factors in helping you pick a TED talk?

- Tags of the talks
- Popularity of the talks
- Number of times a talk has been viewed
- Number of times a talk has been favorited
- Event of the TED talk
- Review of comments on the talk

What do you typically do when you want to watch a TED talk?

How frequently do you follow the related videos?
Would you enjoy a guided tour of the visualization?
Do you read the comments of the talks?
What else information would you like to know about TED talks?

Part 2:
Use the following dashboard to find a TED talk you may be interested in viewing. Think aloud as you explore the visualization.

Appendix 2: Survey Questions

1. How many times have you watched a TED talk in the past year?

- None
- 1-3
- 4-6
- 7-9
- 10 or more

2. If you were exploring a dataset of TED talks to find interesting videos to watch, on which criteria would you want to filter them?

- # of views
- # of comments
- Video Tags (e.g. Technology, Music)
- # of favorites
- TED event
- Count of specific words used in the talks
- Other: _________

3. Which of the below ways to filter TED talks would be THE MOST IMPORTANT to you?

- # of views
- # of comments
- Video Tags (e.g. Technology, Music)
- # of favorites
- TED event
- Count of specific words used in the talks
- Other: _________

4. How useful would a guided tour/tutorial of an interactive visualization be for you the viewer?

1 2 3 4 5

Not useful (1) <-------------------> Very Useful (5)