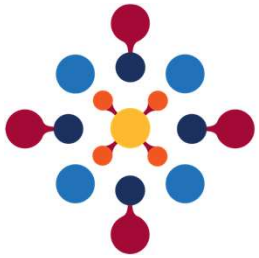


VIS 2022

NLP4Vis: Natural Language Processing for Information Visualization

Half-day Tutorial

Enamul Hoque, York University
Shafiq Joty, Salesforce AI Research and Nanyang Technological University

https://nlp4vis.github.io/

Presenters

Enamul Hoque,
Assistant Professor
York University
enamulh@yorku.ca

Shafiq Joty,
Research Director, Salesforce AI Research
Associate Professor, NTU, Singapore
sjoty@salesforce.com

Tutorial Overview

- Part 1: Introduction
- Part 2: Natural language processing
  - Basics of NLP
  - Deep Learning for NLP
- Part 3: NLP4Vis applications
- Part 4: Future challenges and research opportunities

Tutorial Overview

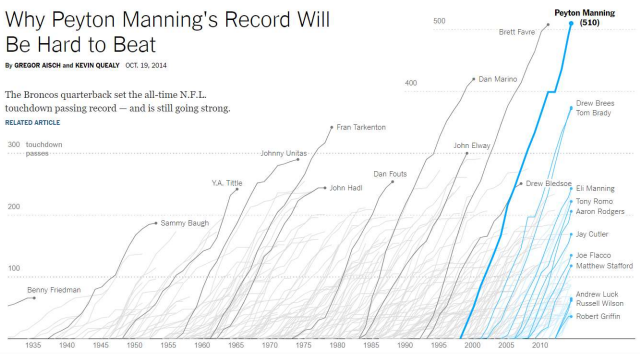
- **Part 1: Introduction**

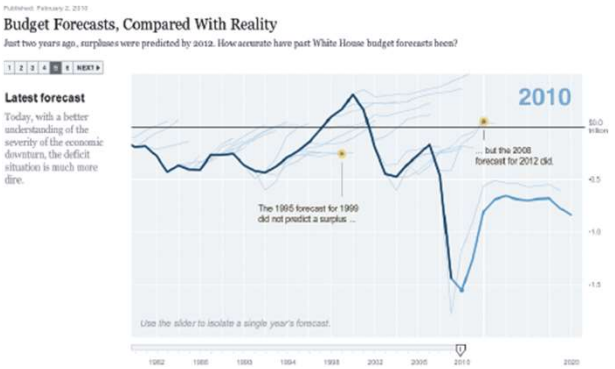
- Part 2: Natural language processing
  - Basics of NLP
  - Deep Learning for NLP

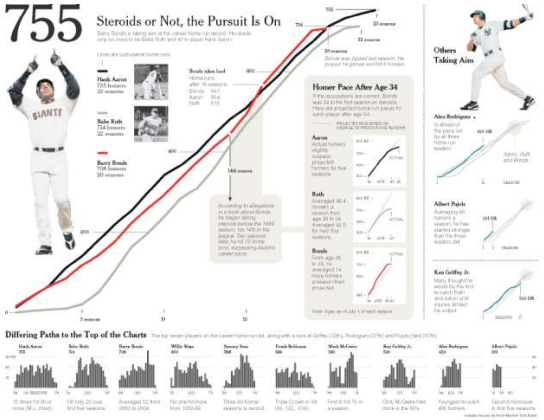
- Part 3: NLP4Vis applications

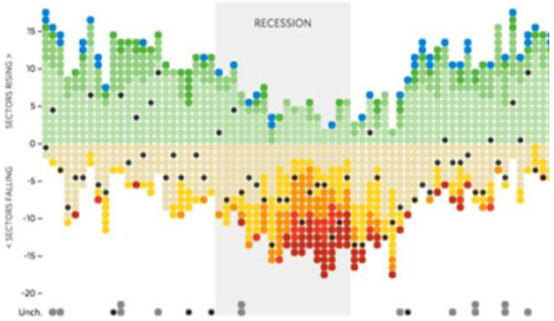
- Part 4: Future challenges and research opportunities

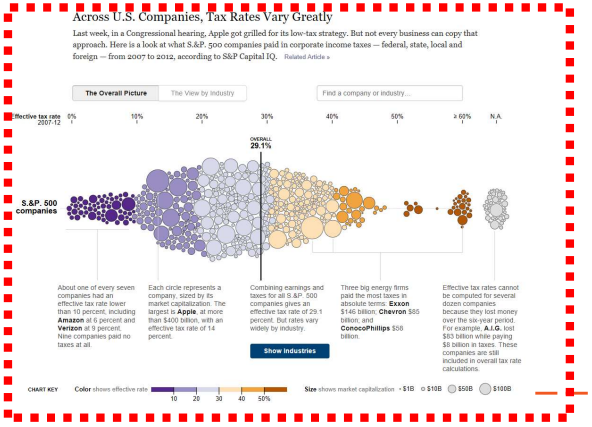
Complementary Roles of Language and Visualization

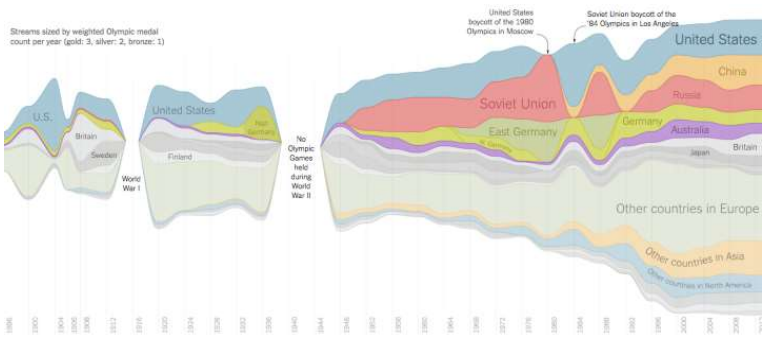










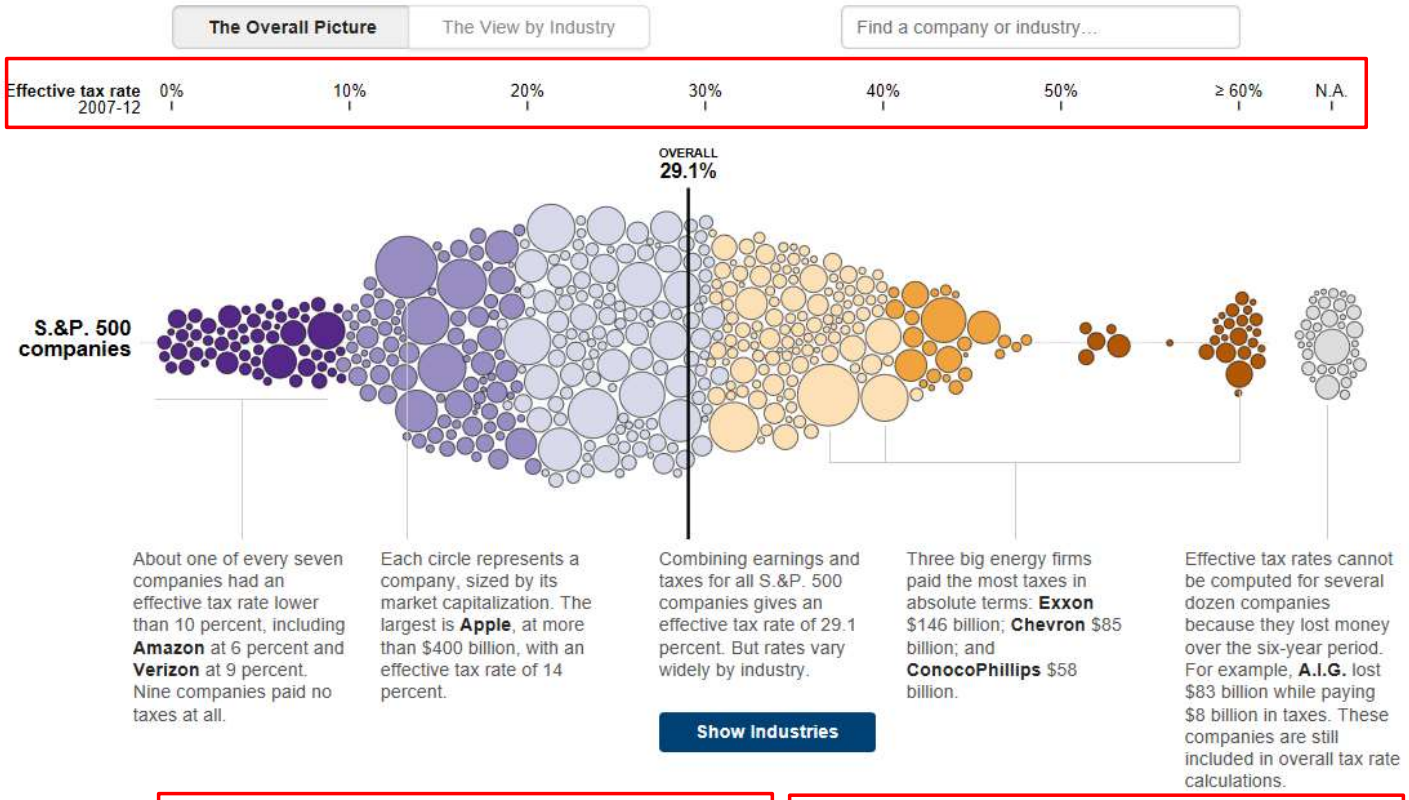


VIS 2022

capitalization	earnings	taxes	sector	name	symbol	rate
73048000000	3.3992E+10	9379000000	Industrials	3M Company	MMM	0.27592
57408000000	3.4694E+10	6317000000	Health care	Abbott Laboratories	ABT	0.18208
3867000000	2551000000	960000000	Retailers	Abercrombie & Fitch Co.	ANF	0.37632
49763000000	1.8735E+10	5025000000	Information technology	Accenture plc	ACN	0.26821
30619000000	1.6453E+10	2834000000	Insurance	ACE Limited	ACE	0.17225
12565000000	1897000000	1218000000	Pharma	Actavis Inc.	ACT	0.64207
21980000000	5824000000	1035000000	Information technology	Adobe Systems Inc.	ADBE	0.17771
1850000000	-4934000000	81000000	Information technology	Advanced Micro Devices Inc.	AMD	
17797000000	1.5139E+10	4637000000	Health care	Aetna Inc.	AET	0.30629
23183000000	1.7461E+10	4441000000	Insurance	AFLAC Inc.	AFL	0.25434
14463000000	4259000000	565000000	Health care	Agilent Technologies Inc.	A	0.13266
5019000000	2272000000	228000000	Utilities	AGL Resources Inc.	GAS	0.10035
17783000000	7775000000	1217000000	Materials	Air Products & Chemicals Inc.	APD	0.15653
7355000000	2270000000	374000000	Materials	Airgas Inc.	ARG	0.16476
6155000000	1531000000	203000000	Information technology	Akamai Technologies Inc.	AKAM	0.13259
8972000000	6031000000	2812000000	Materials	Alcoa Inc.	AA	0.46626
19166000000	767000000	41000000	Pharma	Alexion Pharmaceuticals Inc.	ALXN	0.05346
3214000000	2796000000	489000000	Materials	Allegheny Technologies Inc.	ATI	0.17489
33550000000	5279000000	2049000000	Pharma	Allergan Inc.	AGN	0.38814
10627000000	3371000000	132000000	Information technology	Altera Corp.	ALTR	0.03916
70631000000	3.2126E+10	1.3771E+10	Consumer products	Altria Group Inc.	MO	0.42866
1.17706E+11	5522000000	345000000	Retailers	Amazon.com Inc.	AMZN	0.06248
8512000000	2571000000	246000000	Utilities	Ameren Corporation	AEE	0.09568
23915000000	1.1736E+10	243000000	Utilities	American Electric Power Co. Inc.	AEP	0.02071
72290000000	3.1487E+10	7600000000	Financials	American Express Company	AXP	0.24137
59187000000	-8.2917E+10	8171000000	Insurance	American International Group Inc.	AIG	
31253000000	2711000000	206000000	Financials	American Tower Corporation	AMT	0.07599
14662000000	5659000000	986000000	Financials	Ameriprise Financial Inc.	AMP	0.17424
12191000000	5633000000	1484000000	Health care	AmerisourceBergen Corporation	ABC	0.26345
78660000000	2.8488E+10	4506000000	Pharma	Amgen Inc.	AMGN	0.15817

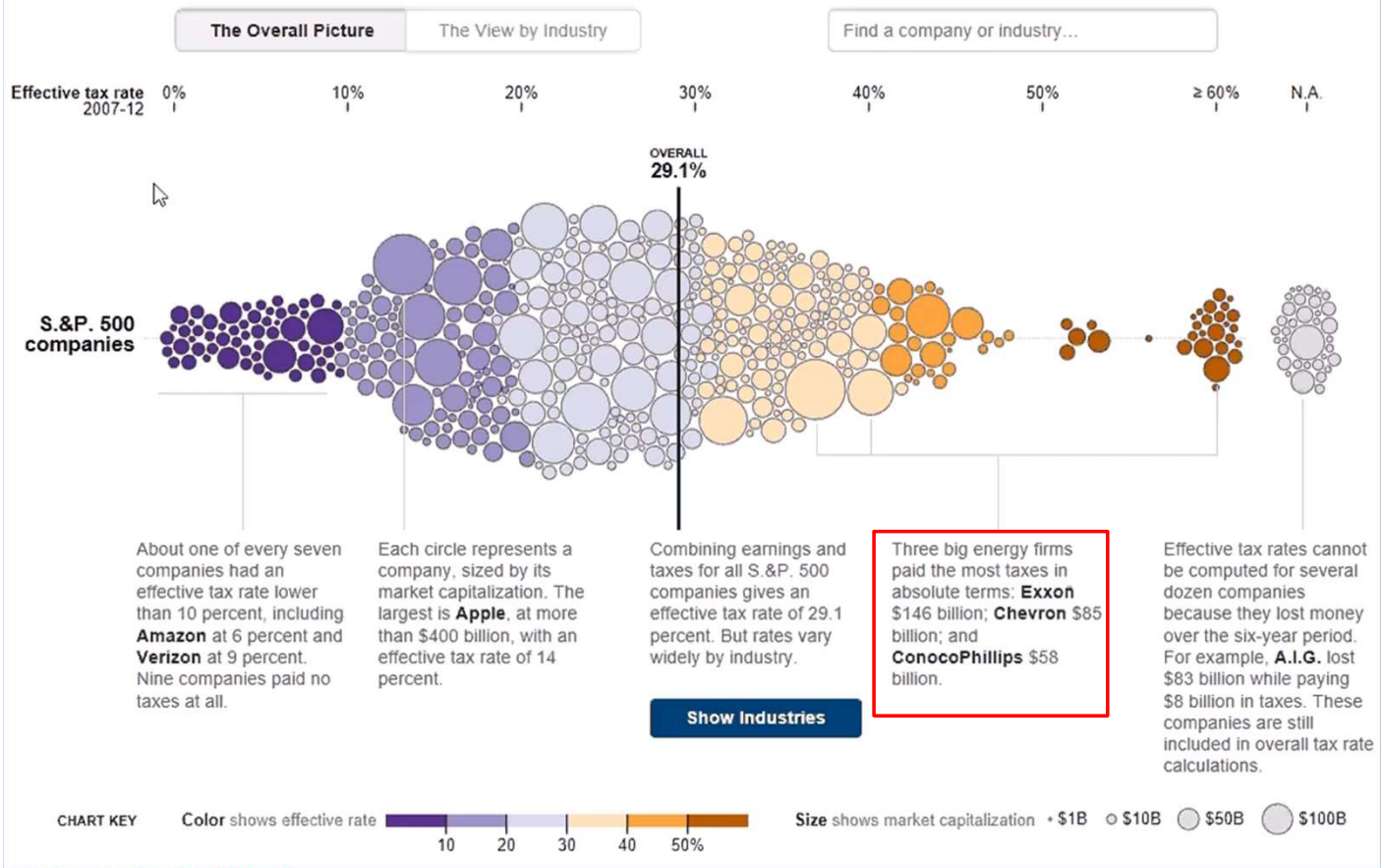
Across U.S. Companies, Tax Rates Vary Greatly

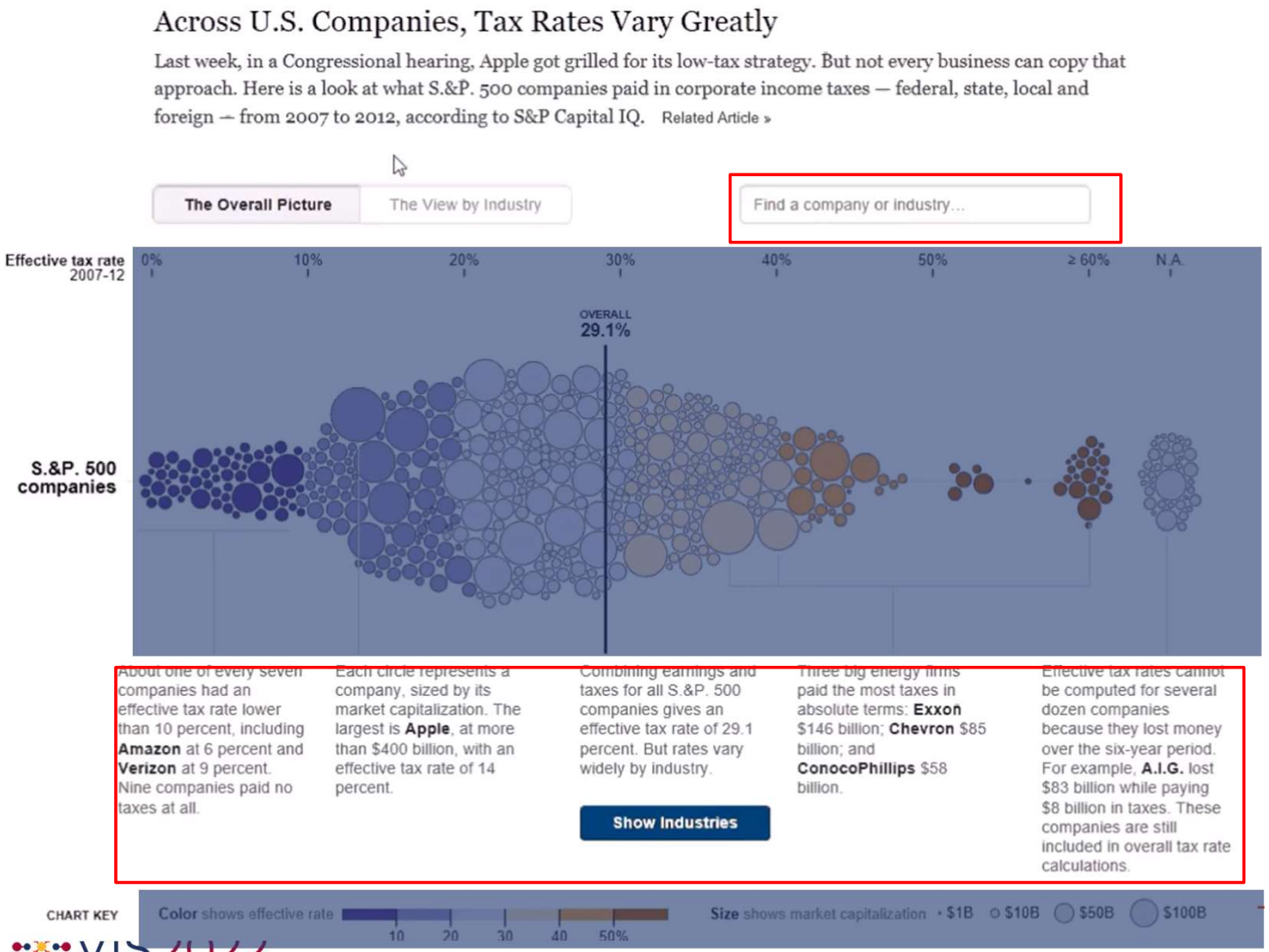
Last week, in a Congressional hearing, Apple got grilled for its low-tax strategy. But not every business can copy that approach. Here is a look at what S.&P. 500 companies paid in corporate income taxes — federal, state, local and foreign — from 2007 to 2012, according to S&P Capital IQ. Related Article »

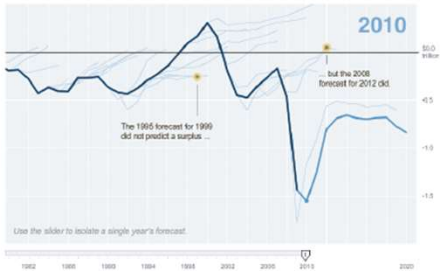


Across U.S. Companies, Tax Rates Vary Greatly

Last week, in a Congressional hearing, Apple got grilled for its low-tax strategy. But not every business can copy that approach. Here is a look at what S.&P. 500 companies paid in corporate income taxes — federal, state, local and foreign — from 2007 to 2012, according to S&P Capital IQ. Related Article »




Across U.S. Companies, Tax Rates Vary Greatly
Last week, in a Congressional hearing, Apple got grilled for its low-tax strategy. But not every business can copy that approach. Here is a look at what S.&P. 500 companies paid in corporate income taxes — federal, state, local and foreign — from 2007 to 2012, according to S&P Capital IQ. Related Article »
The Overall Picture
The View by Industry
Find a company or industry...
Effective tax rate 2007-12
0%
10%
20%
30%
40%
50%
≥ 60%
N.A.
OVERALL 29.1%
S.&P. 500 companies
About one of every seven companies had an effective tax rate lower than 10 percent, including Amazon at 6 percent and Verizon at 9 percent. Nine companies paid no taxes at all.
Each circle represents a company, sized by its market capitalization. The largest is Apple, at more than $400 billion, with an effective tax rate of 14 percent.
Combining earnings and taxes for all S.&P. 500 companies gives an effective tax rate of 29.1 percent. But rates vary widely by industry.
Show Industries
Three big energy firms paid the most taxes in absolute terms: Exxon $146 billion; Chevron $85 billion; and ConocoPhillips $58 billion.
Effective tax rates cannot be computed for several dozen companies because they lost money over the six-year period. For example, A.I.G. lost $83 billion while paying $8 billion in taxes. These companies are still included in overall tax rate calculations.
CHART KEY
Color shows effective rate
10
20
30
40
50%
Size shows market capitalization
$1B
$10B
$50B
$100B

	Natural language processing (NLP)	Information visualization (InfoVis)
Output modality: **Which information is shown?**	Topic modeling Sentiment analysis Text summarization Text generation *Mining and summarizing text*	 *Creating visual representation of data*
Input modality: **How to search for information?**	Find me a good Italian restaurant *Natural language question answering*	 *Mouse and touch interactions*

NLP4Vis

Applying NLP for various visualization-related tasks:

- Visual text analytics
- Natural language interfaces for visualizations
- Text generations for visualizations
- Automatic visual story generation
- Visualization retrieval and recommendation etc.

NLP4Vis

Applying NLP for various visualization-related tasks:

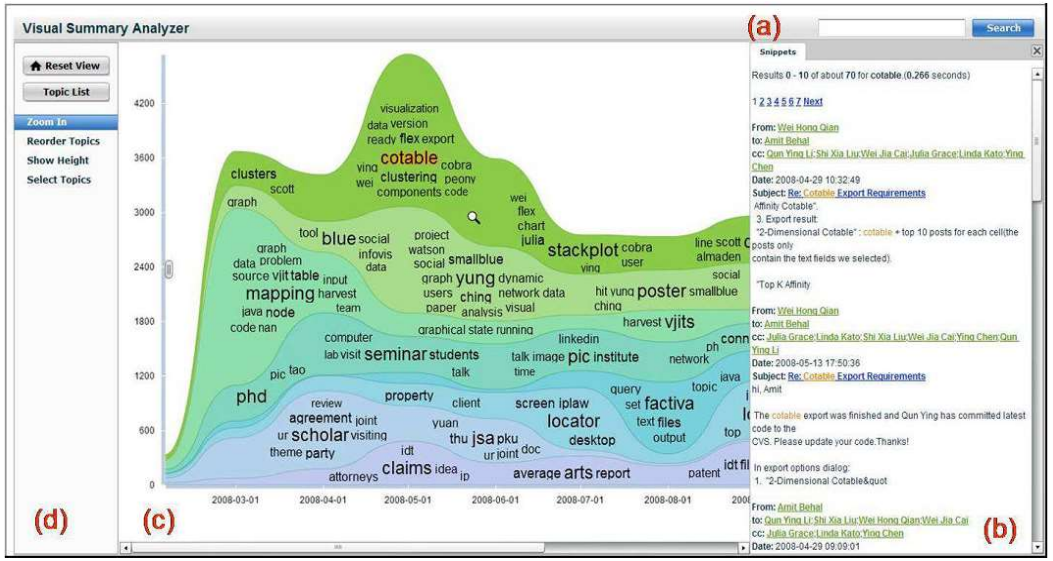
- **Visual text analytics**
- Natural language interfaces for visualizations
- Text generations for visualizations
- Automatic visual story generation
- Visualization retrieval and recommendation etc.

Visual Text Analytics

- Tightly integrates text analytics with Visualization

Obama's Health Care Speech to Congress, 2009

economix.blogs.nytimes.com/2009/09/09/obama-in-09-vs-clinton-in-93



*F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In Proc. ACM Conf. on **KDD**, pages 153–162, 2010*

Visual Text Analytics

- Platforms and Domains

Conversations

Meetings

Phone conversation

Social media conversations

Interview

Beyond conversations

Lecture

Broadcast news

Radio news

Scientific documents

Websites

News articles

Literature

Reviews

Visual Text Analytics

Various tasks:

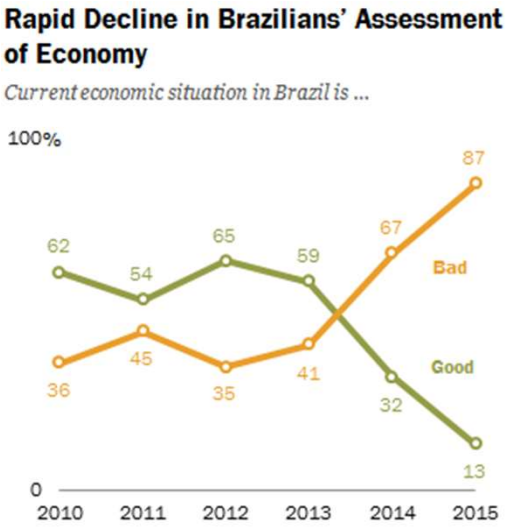
- Understand
- Summarize
- Sentiment analysis
- Argument analysis
- Fact-check
- Rumor detection
- Plagiarism detection
- Abusive comments detection
- Topic controversy detection
- etc.

NLP4Vis

Applying NLP for various visualization-related tasks:

- Visual text analytics
- **Natural language interfaces for visualizations**
- Text generations for visualizations
- Automatic visual story generation
- Visualization retrieval and recommendation etc.

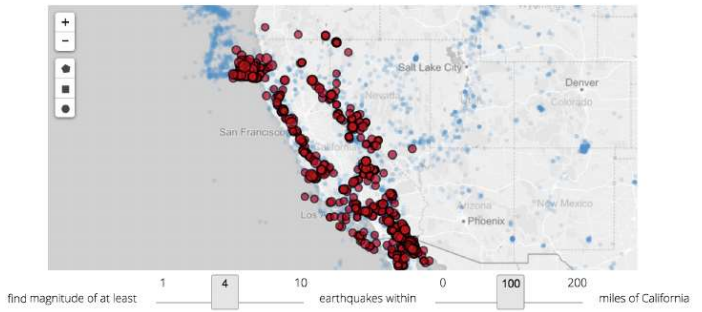
Natural Language Interfaces for Visualizations



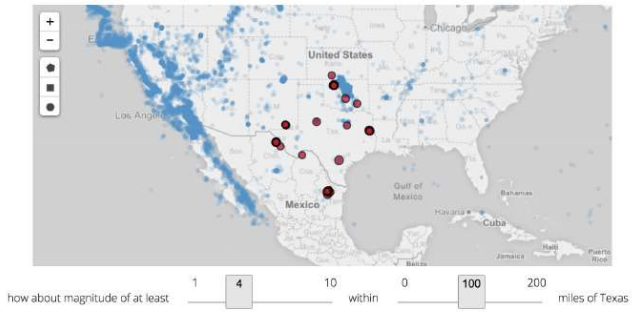
Question: Which year has the most divergent opinions about Brazil's economy?
Answer: 2015

Question: What is the peak of the orange line?
Answer: 87

ChartQA (Masry et al., ACL 2022)



(a) Previous query: "Large earthquakes near California"



(b) Subsequent query: "how about near Texas?"

Eviza (Setlur et al., UIST 2016)

Masry A., Long d. X., Tan j. Q., Joty S. R., Hoque E.: ChartQA: A benchmark for question answering about charts with visual and logical reasoning, ACL 2022.

Natural Language Interfaces for Visualizations

Use natural language as an input modality to facilitate data analysis.

Benefits:

People can express information needs easily through language

Lower the threshold of required analytical skills

Reduce cognitive burden

Can improve chart accessibility

NLP4Vis

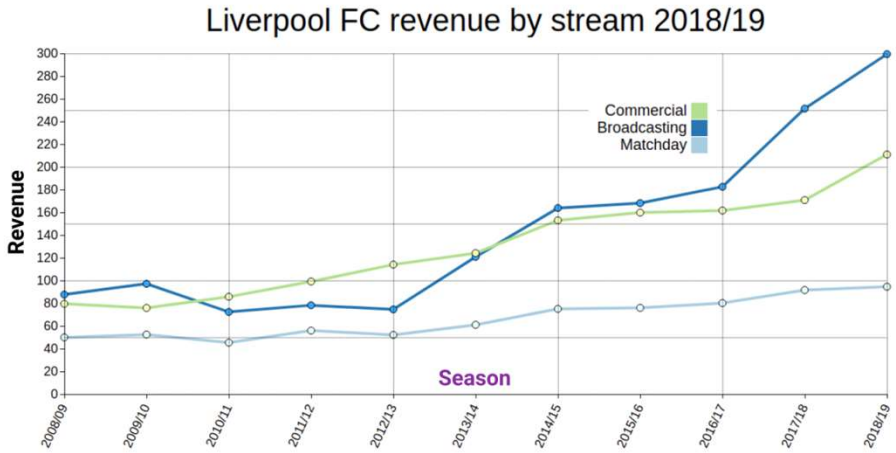
Applying NLP for various visualization-related tasks:

- Visual text analytics
- Natural language interfaces for visualizations
- **Text generations for visualizations**
- Automatic visual story generation
- Visualization retrieval and recommendation etc.

Text Generations for Visualizations

- Generate texts to explain key insights from a chart

Broadcasting is the largest source of revenue for Liverpool FC. In 2018/2019 , the club earned approximately 299.3 million euros from broadcasting, more than triple than in 2010/2011 . The second biggest revenue stream is the commercial one.



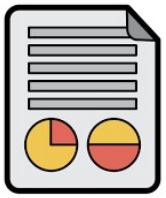
- Shankar K., Rixie tiffany ko l., Xiang l., Ahmed M., Megh T., Enamul H., Shafiq J.: Chart-to-text: A largescale benchmark for chart summarization. In proc ACL, 2022.
- Obeid J., Hoque E.: Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model., INLG 2020

Text Generations for Visualizations

Automatically summarize key data insights using texts.

Benefits:

Help readers in understanding data insights



Suggestions for writing data-driven articles

Can improve chart accessibility for people who are blind

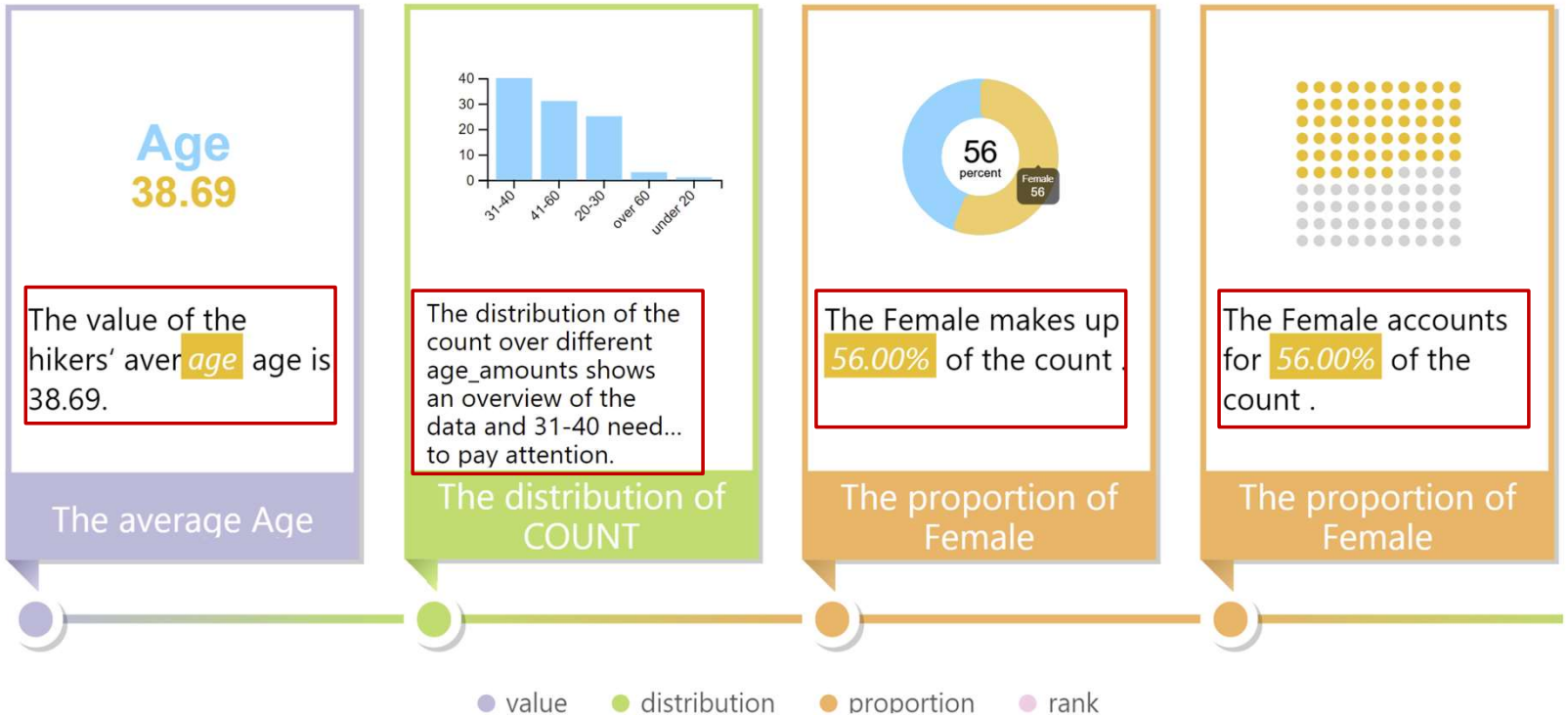
Improve visualization retrieval applications

NLP4Vis

Applying NLP for various visualization-related tasks:

- Visual text analytics
- Natural language interfaces for visualizations
- Text generations for visualizations
- **Automatic visual story generation**
- Visualization retrieval and recommendation etc.

Automatic Visual Data-driven Story Generation



Calliope (Shi et al., IEEE Vis 2020)

Shi D, Xu X, Sun F, Shi Y, Cao N. Calliope: Automatic visual data story generation from a spreadsheet. IEEE TVCG. 2020 Oct 13;27(2):453-63.

NLP4Vis

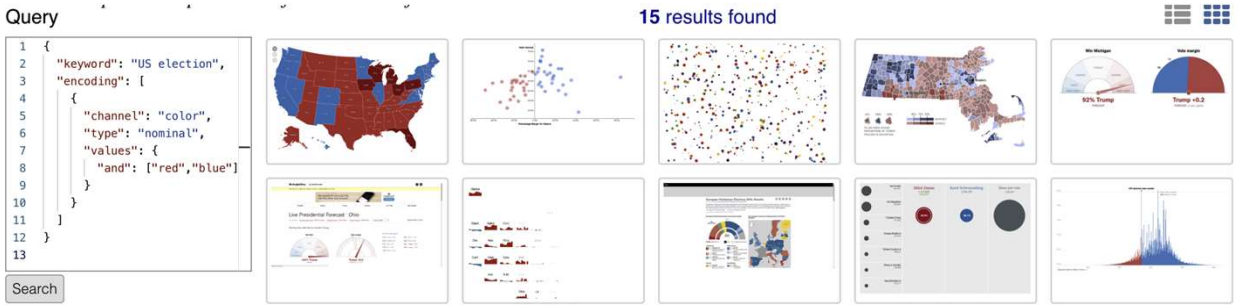
Applying NLP for various visualization-related tasks:

- Visual text analytics
- Natural language interfaces for visualizations
- Text generations for visualizations
- Automatic visual story generation
- **Visualization retrieval and recommendation** etc.

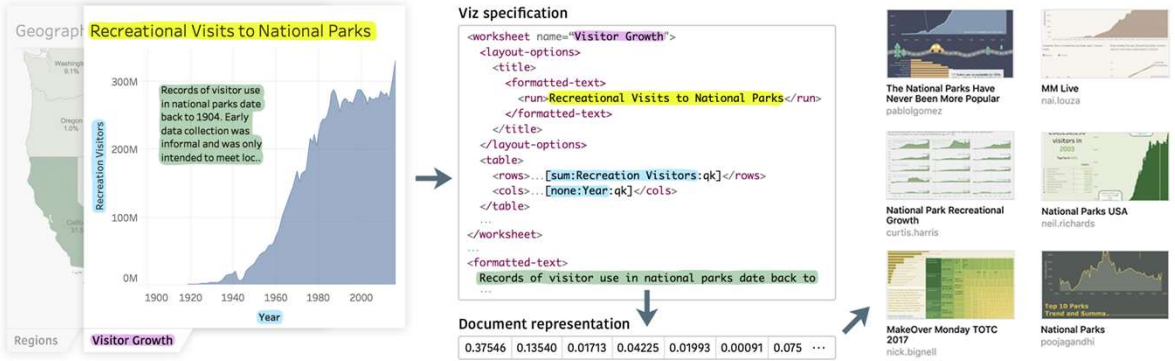
Visualization Retrieval and Recommendation

- Compute Text-based similarity for retrieval and recommendation

Enamul Hoque, Maneesh Agrawala
Searching the Visual Style and Structure of D3 Visualizations
IEEE Vis 2019.



Michael Oppermann, Robert Kincaid, and Tamara Munzner,
VizCommender: Computing Text-Based Similarity in Visualization Repositories for Content-Based Recommendations, *IEEE VIS Conf. (VAST), TVCG, 27(2): 495-505, 2020.*



Tutorial Overview

- **Part 1: Introduction**

- Part 2: Natural language processing
  - Basics of NLP
  - Deep Learning for NLP

- Part 3: NLP4Vis applications

- Part 4: Future challenges and research opportunities

Tutorial Overview

- Part 1: Introduction
- **Part 2: Natural language processing**
  - **Basics of NLP**
  - **Deep Learning for NLP**
- Part 3: NLP4Vis applications
- Part 4: Future challenges and research opportunities