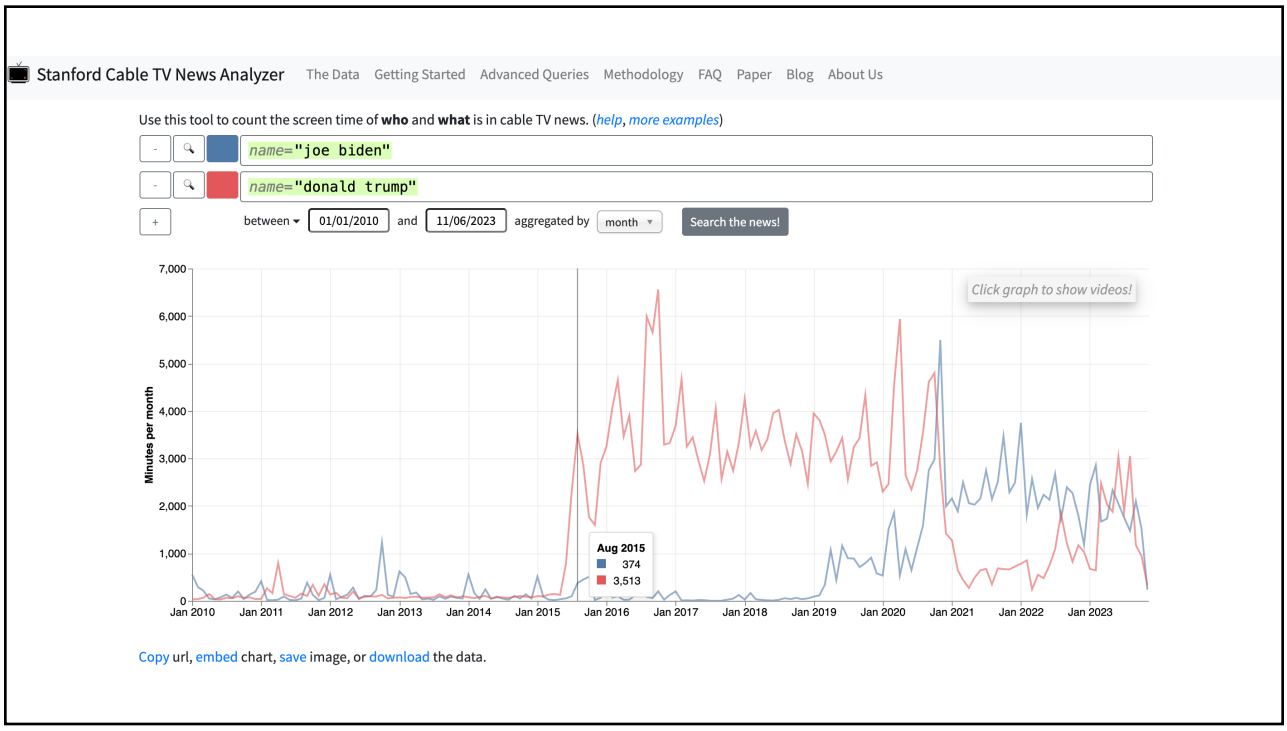
VISUALIZATION AND NATURAL LANGUAGE PROCESSING

CS 448B | Fall 2023

MANEESH AGRAWALA

1



2

“Words and pictures belong together.”

Tufte 1983

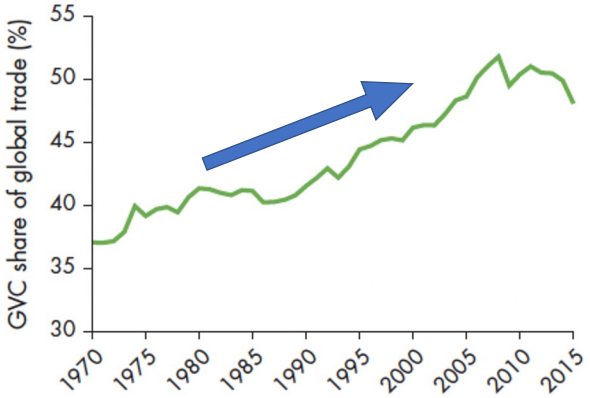
3

TODAY

Learning Objectives

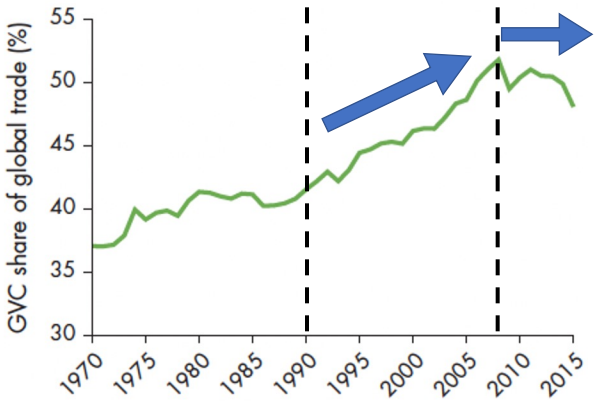
1. Understand how people read charts and text together.
2. Visual QA with explanations for charts and graphs.

4



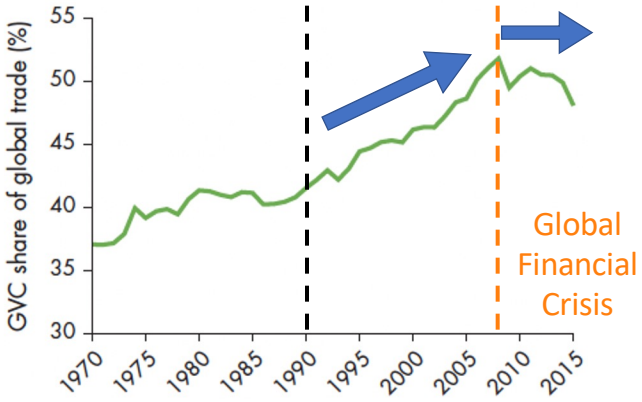
5

Figure 1.2 GVC trade grew rapidly in the 1990s but stagnated after the 2008 global financial crisis



6

Figure 1.2 GVC trade grew rapidly in the 1990s but stagnated after the 2008 global financial crisis



7

The evolution of GVC participation

The overall share of GVC trade in total world trade—encompassing both forward and backward linkages—grew significantly in the 1990s and early 2000s, but it appears to have stagnated or even declined in the last 10 years (figure 1.2). Still, about half of world trade appears to be related to GVCs.

What explains the remarkable rise in GVC participation in the 1990s and 2000s? And why has this process stalled since the financial crisis?

The global wave of fragmentation of production in the 1990s and 2000s was driven by a combination of factors. The information and communication technology (ICT) revolution brought forth cheaper and more reliable telecommunications, new information management software, and increasingly powerful personal computers (figure 1.3, panel a). Manufacturing firms then found it easier to outsource and coordinate complex activities at a distance and ensure the quality of their inputs. In addition, firms were able to disperse production across the world because transport costs fell significantly (figure 1.3, panel b). Declining air and sea freight costs boosted the trade in goods, while services benefited from cheaper communication costs.

Successive rounds of trade liberalization have resulted in rapidly falling barriers to trade and investment for both developed and developing countries. Tariffs have declined, especially for manufactured goods, and the gradual, although still insufficient, lowering of nontariff barriers has facilitated the international trade of goods and services (figure 1.4). Finally, the creation of the European single market—together with the integration of China, India, and the Soviet Union into the global economy—created huge new product and labor markets, and so firms could sell the same goods to more people and take advantage of economies of scale leading to the further deepening of GVCs. The new supply of cheap labor encouraged profit-seeking companies to either reallocate their production facilities or find local suppliers in low-wage countries.[3]

Since the global financial crisis in 2008, the dynamics of GVC expansion have changed. Trade has bounced back from its deep crisis level, but it has grown only marginally faster than output. Trade in parts and components also stalled after the financial crisis and even fell between 2011 and 2014, with a modest increase since then.

The factors behind the trade and GVC slowdown are both cyclical and structural in nature. On the one hand, trade growth is lower because global output growth is lower in economies that account for large shares of global trade and global output, such as Europe and China. Trade has also grown at a slower pace because the trade-to-income elasticity—defined as the amount of trade generated as output rises—has decreased. This is especially true in large trading countries, including China and the United States. China is producing more at home, thereby becoming less reliant on imported components for its exports. The share of intermediate imports in exports of Chinese goods dropped from about 50 percent in the 1990s to a little over 30 percent in 2015. In the United States, a booming shale sector reduced oil imports by one-fourth between 2010 and 2015.[4]

Figure 1.2 GVC trade grew rapidly in the 1990s but stagnated after the 2008 global financial crisis

Sources: WDR 2020 team, using data from Eora26 database; Borin and Mancini (2015, 2019); and Johnson and Noguera (2017). See appendix A for a description of the databases used in this Report.

Note: Unless otherwise specified, GVC participation measures used in this and subsequent figures throughout the Report follow the methodology from Borin and Mancini (2015, 2019). The Eora26 database is used because it offers the largest country coverage: 190 countries between 1990 and 2015. GVC participation corresponds to the share of world exports that flow through at least two borders. For 1990–2015, the GVC participation measure is computed as the share of GVC exports in total international exports using the Borin and Mancini methodology. GVC exports include transactions in which a country's exports embody value added that it previously imported

9

READING CHARTS AND TEXT

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts Kim, Setlur and Agrawala, CHI 2021

12

13

14

The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

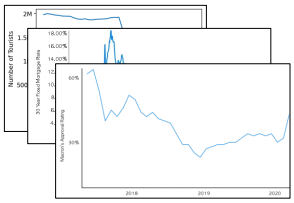
15

The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

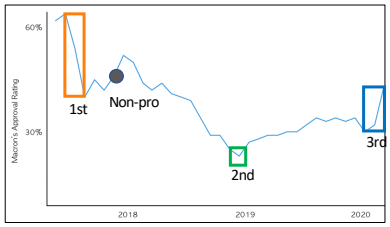
16

Do readers rely more on the chart or captions for their takeaways?

17



Collect Time-Series Charts



Identify Visually Prominent Features

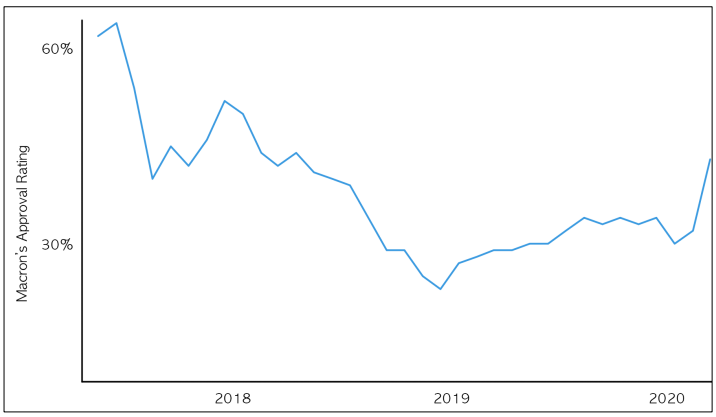
Macron's approval rating steeply dropped between June and August of 2017.

Generate Captions

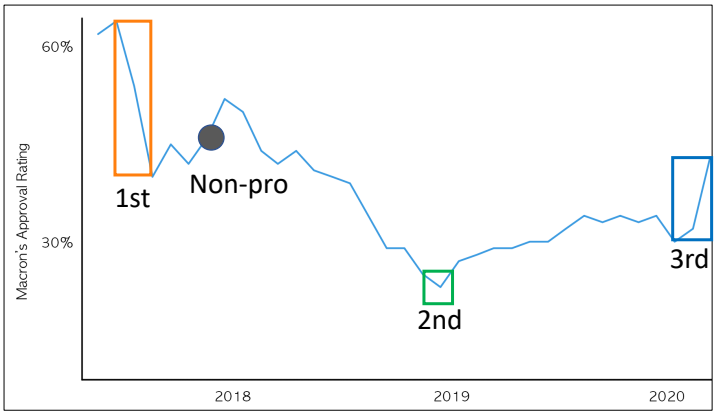
Subjects Viewed Chart-Caption Pairs

Reported Takeaways

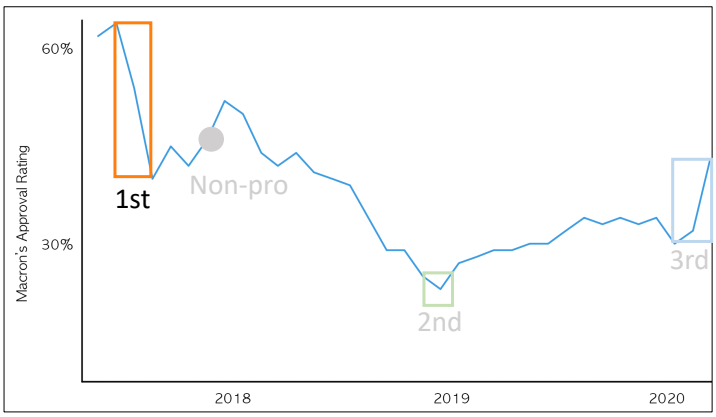
19



20



21



[DOMAIN] [FEATURE] between [START DATE] and [END DATE].

↓

Macron's approval rating steeply dropped between June and August of 2017.

22

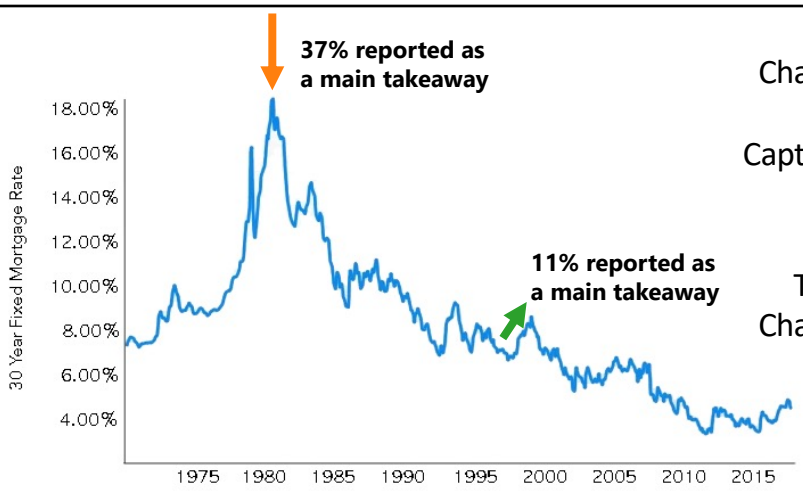
Study Results

23

Chart Emphasis
≠
Caption Emphasis
↓
Takeaway:
Chart Emphasis

The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

24



The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

25

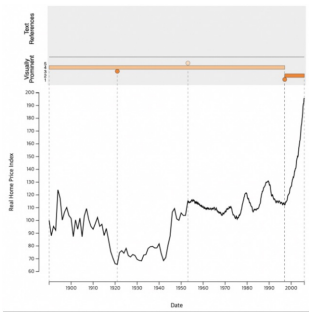
The 30-year fixed mortgage rate reached its peak of 18.5% in 1981.

26

Do readers rely more on the chart or captions for their takeaways?

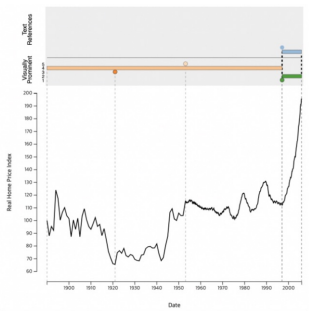
When text and visualization emphasis mismatch, readers rely more on the chart and can miss information in the caption.

27



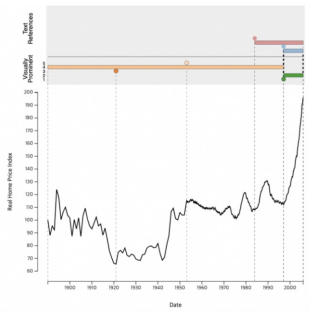
This chart shows the real home price index between 1890 and 2006.

(a) Prominent features & Basic caption



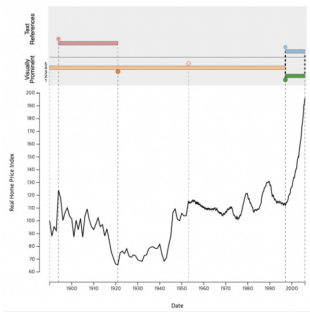
This chart shows the real home price index between 1890 and 2006. The housing prices have skyrocketed starting around 1997 and we need to act.

(b) Caption text about prominent feature



This chart shows the real home price index between 1890 and 2006. The housing prices have skyrocketed starting around 1997 and we need to act. Looking back, they declined since 1984 with an increased housing supply as manufactured homes became available to the public.

(c) Caption including false information



This chart shows the real home price index between 1890 and 2006. The housing prices have skyrocketed starting around 1997 and we need to act. Looking back, they declined since 1894 with an increased housing supply as manufactured homes became available to the public. A similar supply-side solution is what we need.

(d) Caption about less prominent feature

EmphasisChecker: A Tool for Guiding Chart and Caption Emphasis
Kim, Choi, Kim, Setlur and Agrawala, IEEE Visualization 2023

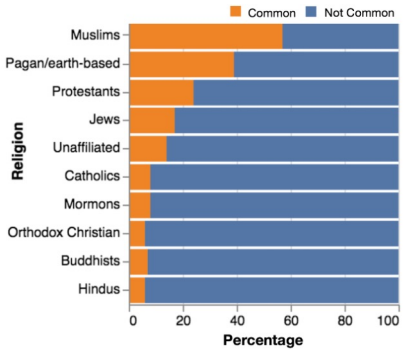
28

CHART QUESTION ANSWERING WITH VISUAL EXPLANATIONS

Answering Questions about Charts and Generating Visual Explanations Kim, Hoque and Agrawala, CHI 2020

29

Chart



Question

For which religion did the fewest chaplains think that religious extremism is common?

33

Chart

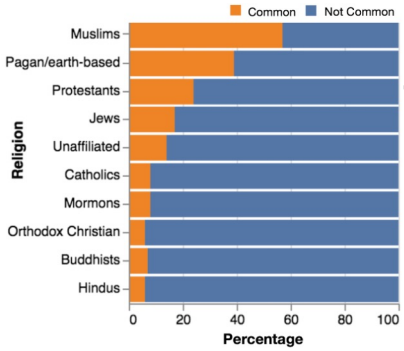


Chart QA

Answer

Orthodox Christians, Hindus.

Question

For which religion did the fewest chaplains think that religious extremism is common?

34

Chart

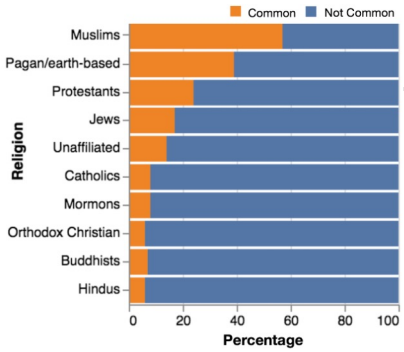


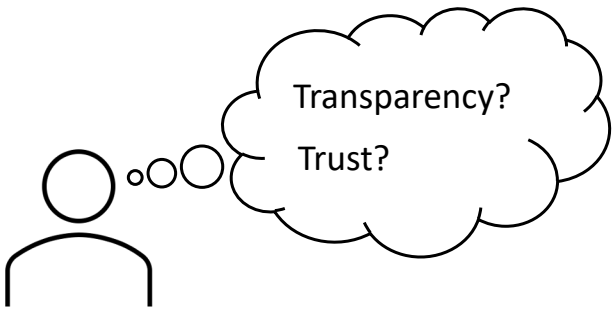
Chart QA

Answer

Orthodox Christians, Hindus.

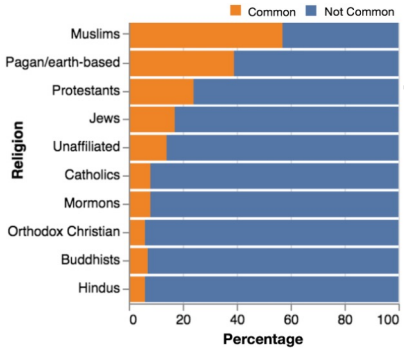
Question

For which religion did the fewest chaplains think that religious extremism is common?



35

Chart



Question

For which religion did the fewest chaplains think that religious extremism is common?

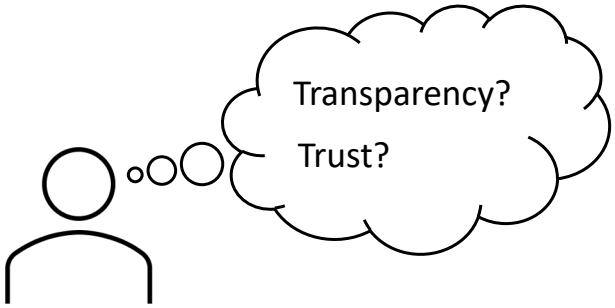
Chart QA

Answer

Orthodox Christians, Hindus.

Explanation

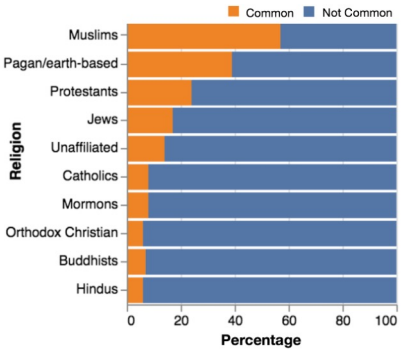
I looked up 'Religion' for the shortest orange bars.



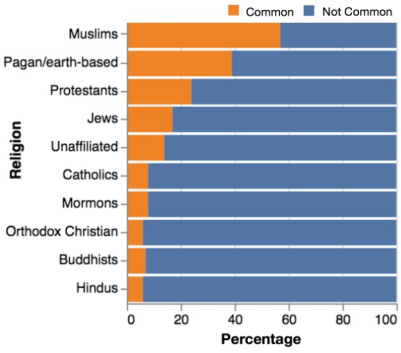
36

Formative Study

37

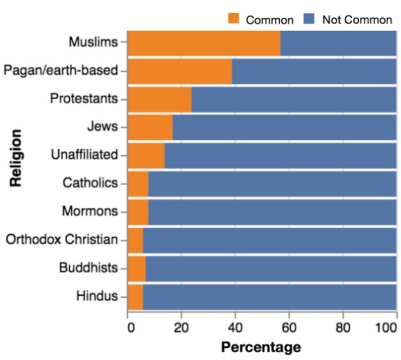


38



Question Which religion has the greatest value for Common?

39



Question Which religion has the greatest value for Common?

Answer Muslims

Explanation I picked religions with the greatest orange percentage.

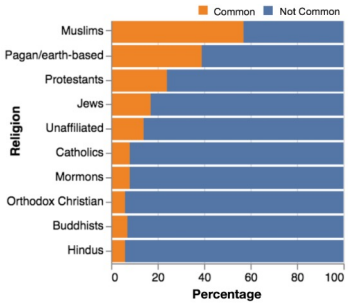
40

Explanations describe procedure for computing answer

41

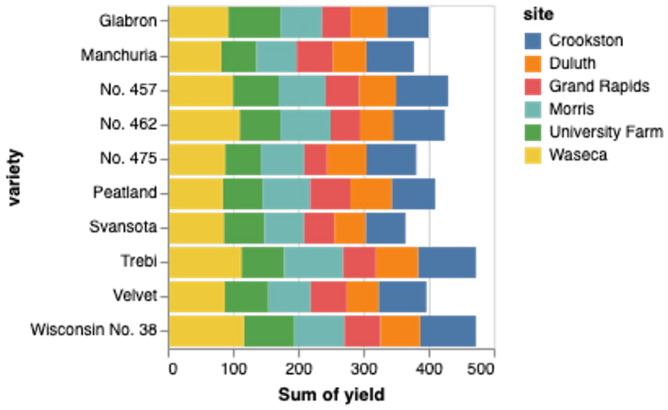
Explanations describe procedure for computing answer

Half of the explanations referred to visual features of chart

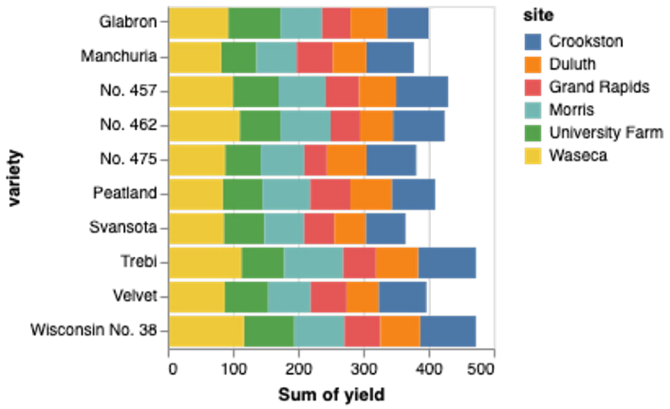


*"I picked religions with the greatest **orange** percentage."*

42

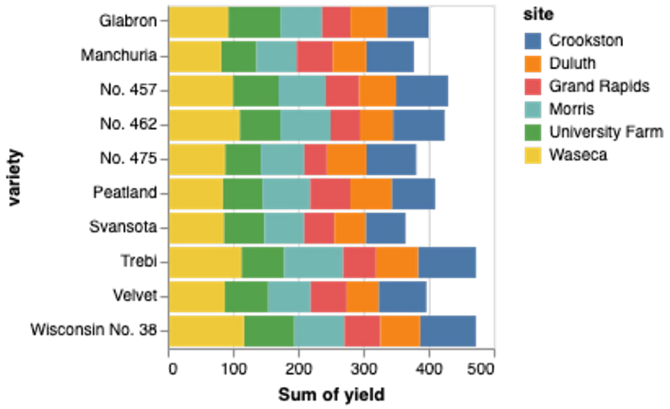


43



Green component in the top bar

44



Green component in the top bar

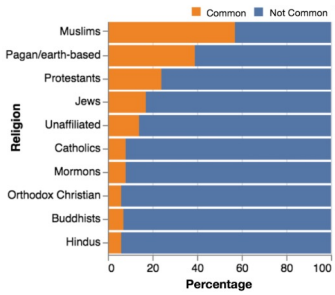
Glabron at University Farm

45

Chart QA Pipeline

46

Chart

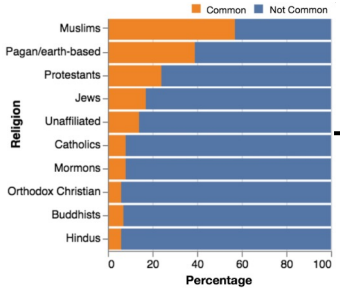


Question about Chart

Which religion has the shortest orange component?

55

Chart



Data Table

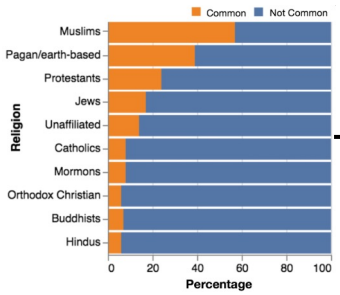
Religion	Common	Not common
Muslims	57	43
Pagan/earth-based	39	61
Protestants	24	76
Jews	17	83
⋮	⋮	⋮
Buddhists	7	93
Hindus	6	94

Question about Chart

Which religion has the shortest orange component?

56

Chart



Data Table

Religion	Common	Not common
Muslims	57	43
Pagan/earth-based	39	61
Protestants	24	76
Jews	17	83
⋮	⋮	⋮
Buddhists	7	93
Hindus	6	94

Question about Chart

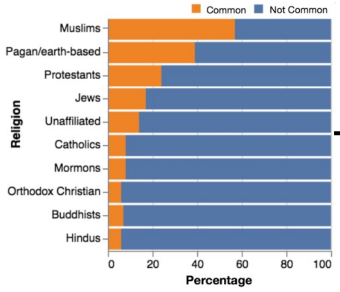
Which religion has the shortest orange component?

Question about Table

Which religion has the least 'Common' Percentage?

57

Chart



Data Table

Religion	Common	Not common
Muslims	57	43
Pagan/earth-based	39	61
Protestants	24	76
Jews	17	83
⋮	⋮	⋮
Buddhists	7	93
Hindus	6	94

Question about Chart

Which religion has the shortest orange component?

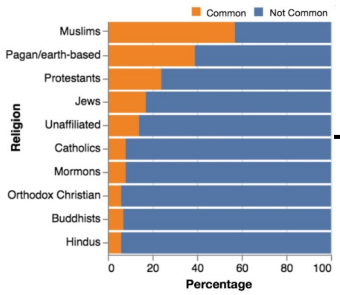
Question about Table

Which religion has the least 'Common' Percentage?

Sempre
[Pasupat 2015]

58

Chart



Data Table

Religion	Common	Not common
Muslims	57	43
Pagan/earth-based	39	61
Protestants	24	76
Jews	17	83
⋮	⋮	⋮
Buddhists	7	93
Hindus	6	94

Question about Chart

Which religion has the shortest orange component?

Question about Table

Which religion has the least 'Common' Percentage?

Sempre
[Pasupat 2015]

Answer

Orthodox Christians, Hindus.

59

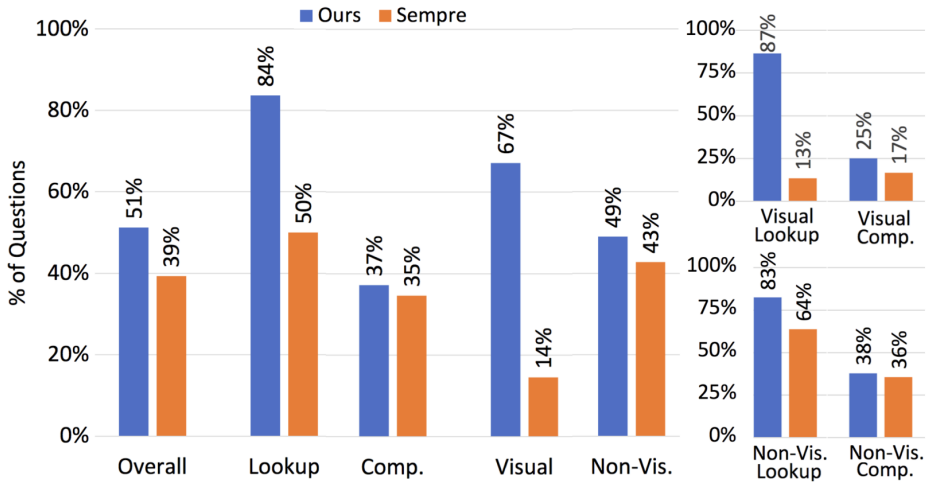
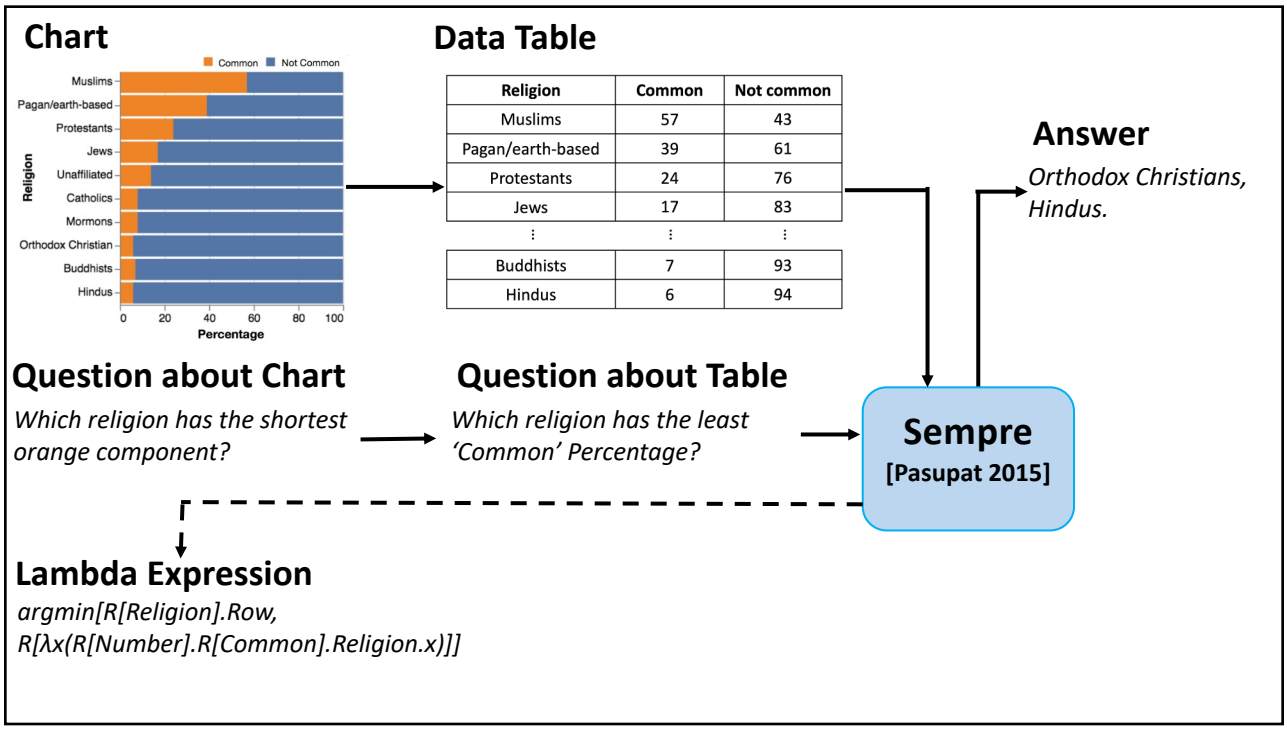


Figure 10. Accuracy of our pipeline (blue) compared to a baseline version of Sempre (orange) for questions of each type (visual/non-visual and lookup/compositional).

60



61

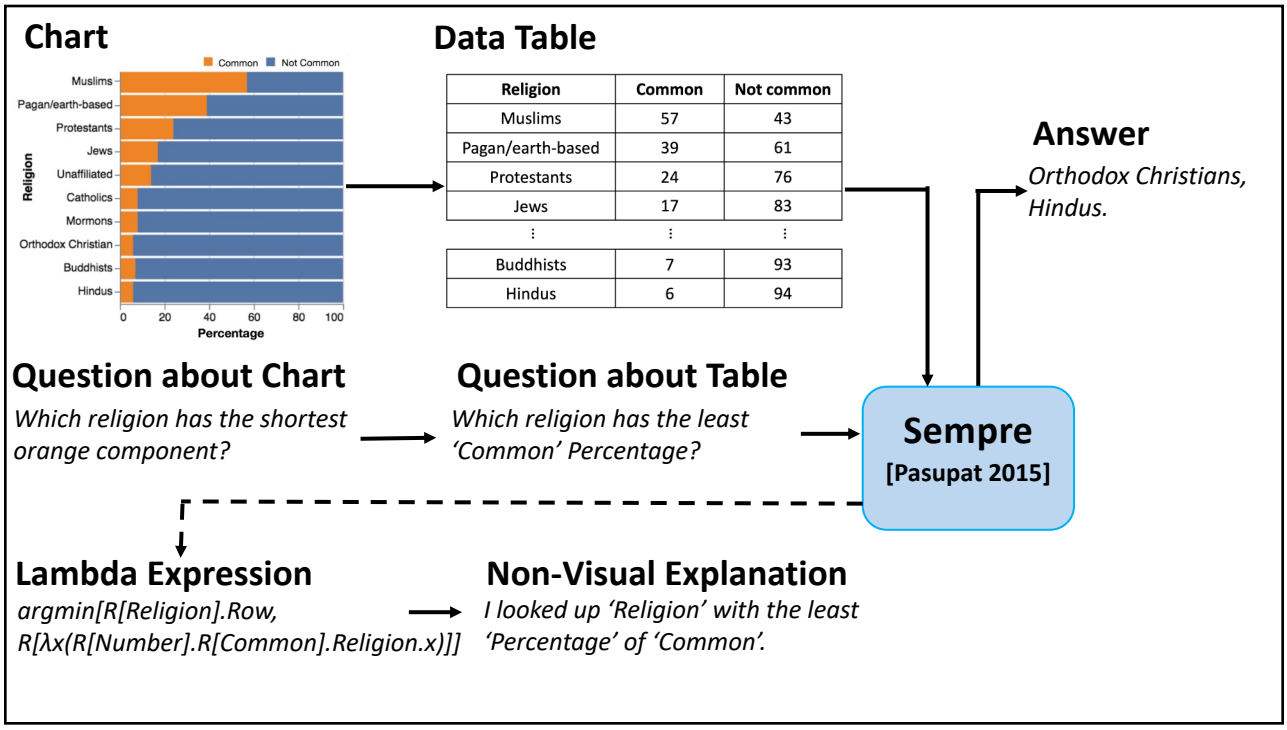


Chart
Common
Not Common
Muslims
Pagan/earth-based
Protestants
Jews
Unaffiliated
Catholics
Mormons
Orthodox Christian
Buddhists
Hindus
Religion
0
20
40
60
80
100
Percentage
Data Table
Religion | Common | Not common
Muslims | 57 | 43
Pagan/earth-based | 39 | 61
Protestants | 24 | 76
Jews | 17 | 83
Buddhists | 7 | 93
Hindus | 6 | 94
Answer
Orthodox Christians, Hindus.
Question about Chart
Which religion has the shortest orange component?
Question about Table
Which religion has the least 'Common' Percentage?
Sempre
[Pasupat 2015]
Lambda Expression
argmin[R[Religion].Row, R[λx(R[Number].R[Common].Religion.x)]]
Non-Visual Explanation
I looked up 'Religion' with the least 'Percentage' of 'Common'.

62

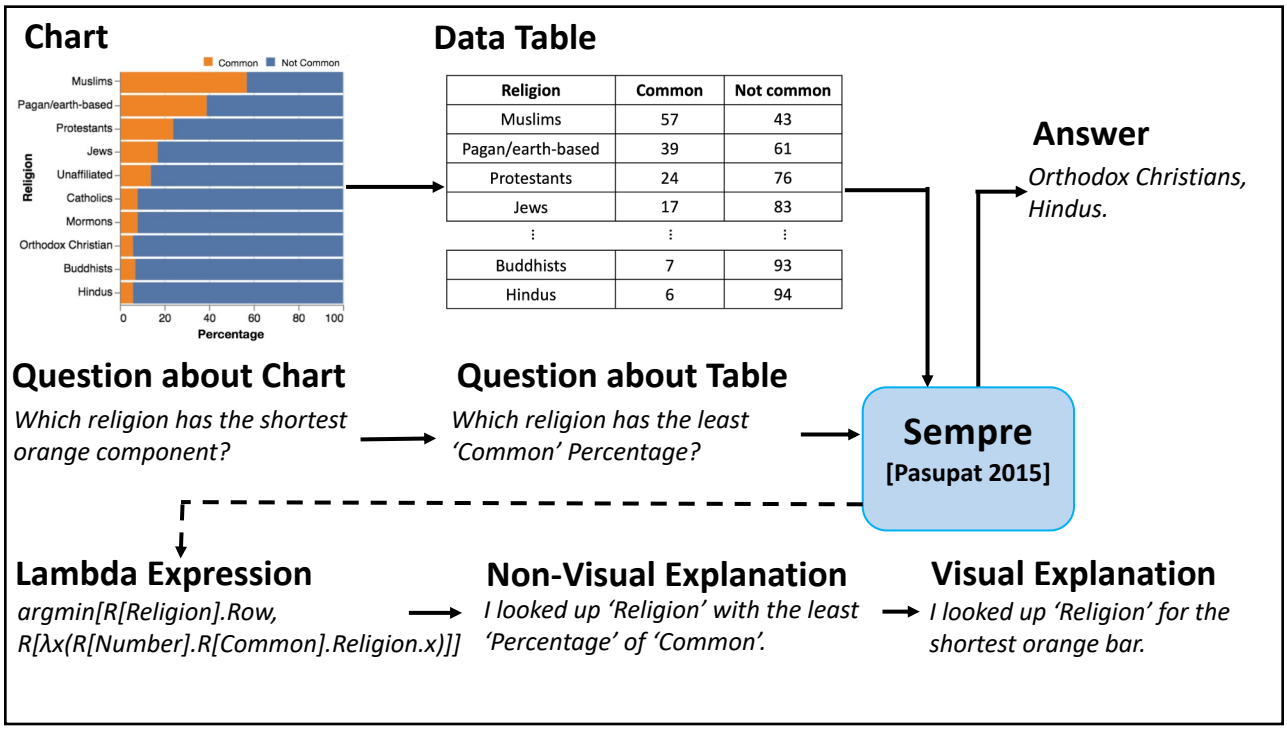
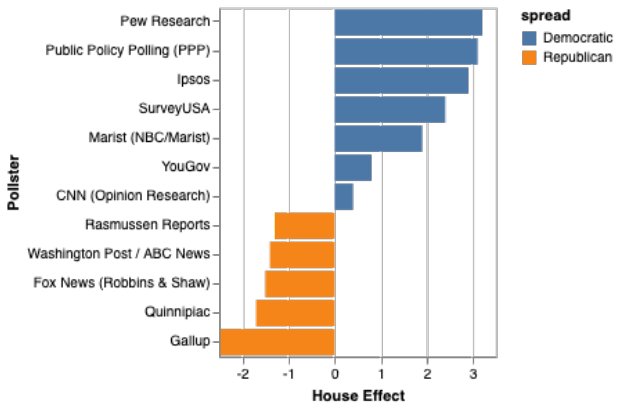


Chart
Common
Not Common
Muslims
Pagan/earth-based
Protestants
Jews
Unaffiliated
Catholics
Mormons
Orthodox Christian
Buddhists
Hindus
Religion
0
20
40
60
80
100
Percentage
Data Table
Religion | Common | Not common
Muslims | 57 | 43
Pagan/earth-based | 39 | 61
Protestants | 24 | 76
Jews | 17 | 83
Buddhists | 7 | 93
Hindus | 6 | 94
Answer
Orthodox Christians, Hindus.
Question about Chart
Which religion has the shortest orange component?
Question about Table
Which religion has the least 'Common' Percentage?
Sempre
[Pasupat 2015]
Lambda Expression
argmin[R[Religion].Row, R[λx(R[Number].R[Common].Religion.x)]]
Non-Visual Explanation
I looked up 'Religion' with the least 'Percentage' of 'Common'.
Visual Explanation
I looked up 'Religion' for the shortest orange bar.

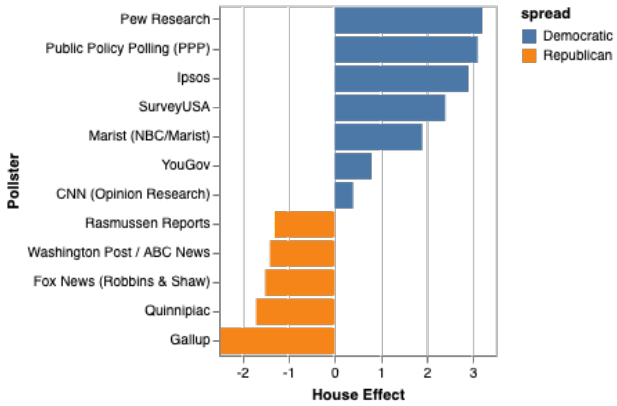
63

Example Explanations

64



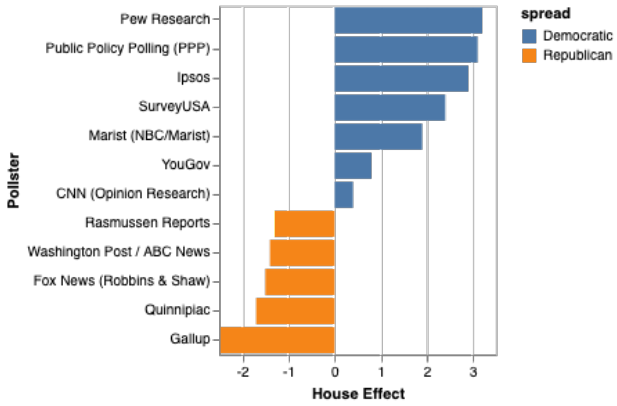
65



Question

What is the difference between the value of Gallup and Quinnipiac?

66



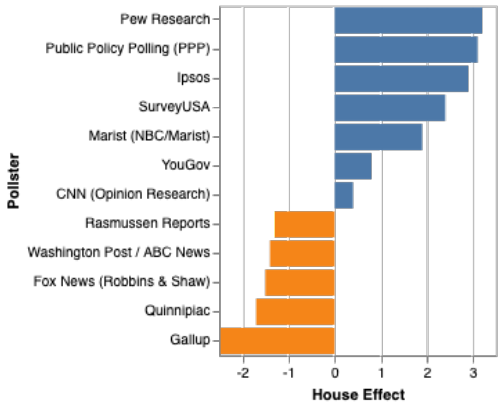
Question

What is the difference between the value of Gallup and Quinnipiac?

Answer

0.8

67



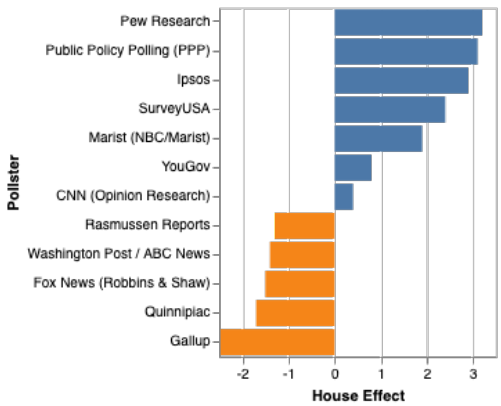
Question

What is the difference between the value of Gallup and Quinnipiac?

Answer

0.8 ✓

68



Question

What is the difference between the value of Gallup and Quinnipiac?

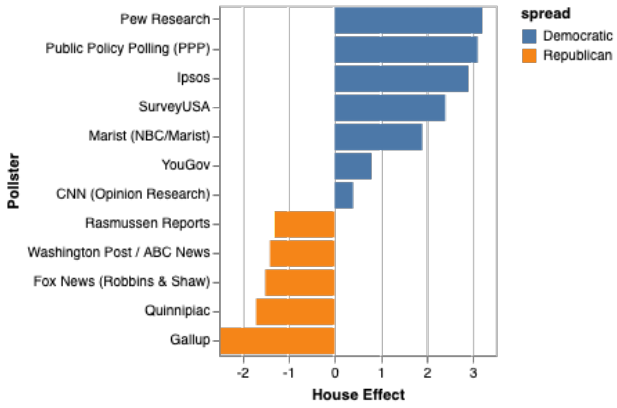
Answer

0.8 ✓

Explanation

I computed the difference between the length of the bar for 'Gallup' and 'Quinnipiac'.

69



Question

What is the difference between the value of Gallup and Quinnipiac?

Answer

0.8 ✓

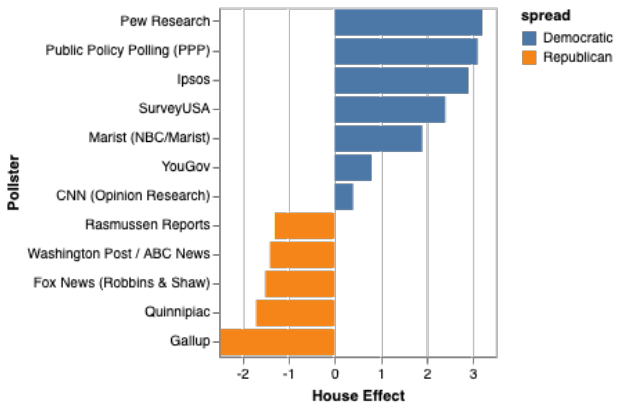
Explanation

I computed the difference between the length of the bar for 'Gallup' and 'Quinnipiac'.

Question

What position is Ipsos in?

70



Question

What is the difference between the value of Gallup and Quinnipiac?

Answer

0.8 ✓

Explanation

I computed the difference between the length of the bar for 'Gallup' and 'Quinnipiac'.

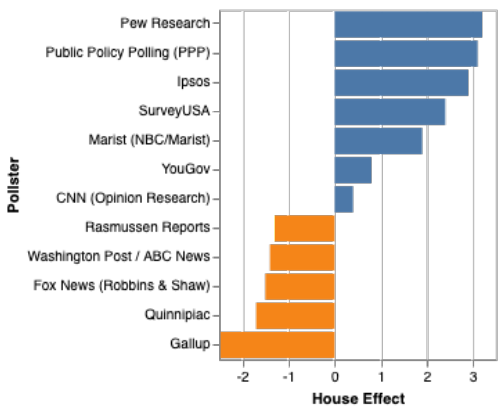
Question

What position is Ipsos in?

Answer

2.9

71



Question

What is the difference between the value of Gallup and Quinnipiac?

Answer

0.8 ✓

Explanation

I computed the difference between the length of the bar for 'Gallup' and 'Quinnipiac'.

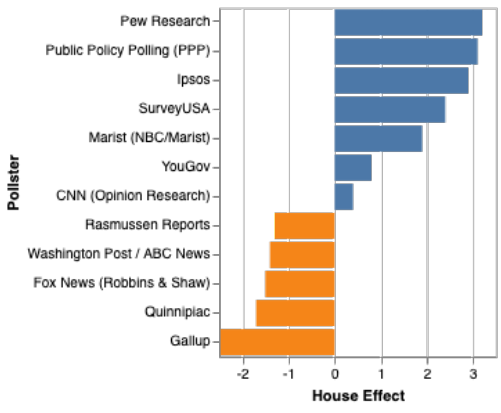
Question

What position is Ipsos in?

Answer

2.9 ✗

72



Question

What is the difference between the value of Gallup and Quinnipiac?

Answer

0.8 ✓

Explanation

I computed the difference between the length of the bar for 'Gallup' and 'Quinnipiac'.

Question

What position is Ipsos in?

Answer

2.9 ✗

Explanation

I looked up the length of the bar for 'Ipsos'.

73

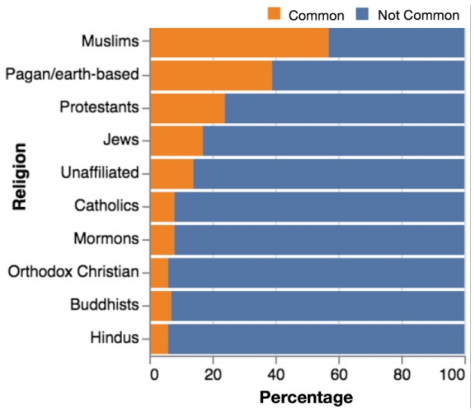
User Study

74

Hypothesis: Visual explanations increase transparency and trust

75

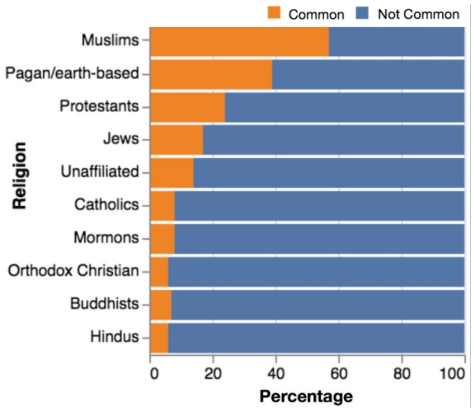
Hypothesis: Visual explanations increase transparency and trust



Q: *For which religion did the fewest chaplains think that religious extremism is common?*

76

Hypothesis: Visual explanations increase transparency and trust

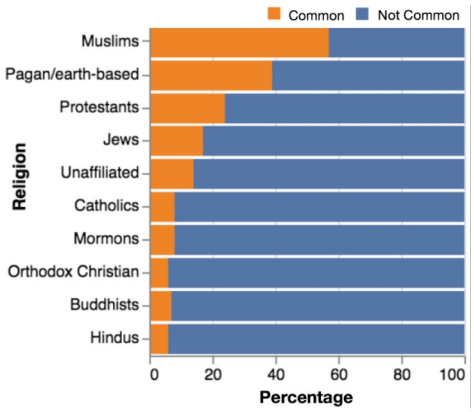


Q: *For which religion did the fewest chaplains think that religious extremism is common?*

A1 (vis): *Orthodox Christians, Hindus. I looked up 'Religion' for the shortest orange bar.*

77

Hypothesis: Visual explanations increase transparency and trust



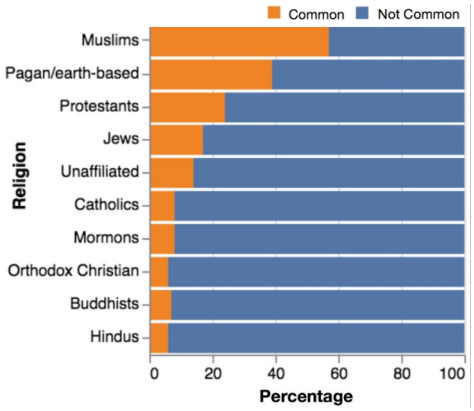
Q: *For which religion did the fewest chaplains think that religious extremism is common?*

A1 (vis): *Orthodox Christians, Hindus. I looked up 'Religion' for the shortest orange bar.*

A2 (no-exp): *Orthodox Christians, Hindus.*

78

Hypothesis: Visual explanations increase transparency and trust



Q: *For which religion did the fewest chaplains think that religious extremism is common?*

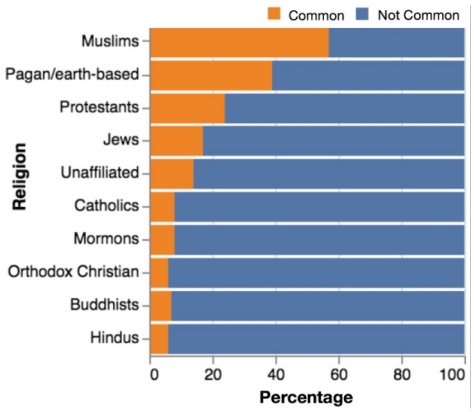
A1 (vis): *Orthodox Christians, Hindus. I looked up 'Religion' for the shortest orange bar.*

A2 (no-exp): *Orthodox Christians, Hindus.*

A3 (non-vis): *Orthodox Christians, Hindus. I looked up 'Religion' with the lowest value for 'Common'.*

79

Hypothesis: Visual explanations increase transparency and trust



Q: *For which religion did the fewest chaplains think that religious extremism is common?*

A1 (vis): *Orthodox Christians, Hindus. I looked up 'Religion' for the shortest orange bar.*

A2 (no-exp): *Orthodox Christians, Hindus.*

A3 (non-vis): *Orthodox Christians, Hindus. I looked up 'Religion' with the lowest value for 'Common'.*

A4 (human): *Orthodox Christians, Hindus. They have lowest values for 'Common'.*

80

16 participants

81

16 participants

20 Chart-Question-Answer-Explanation tuples (5 per explanation type)

82

16 participants

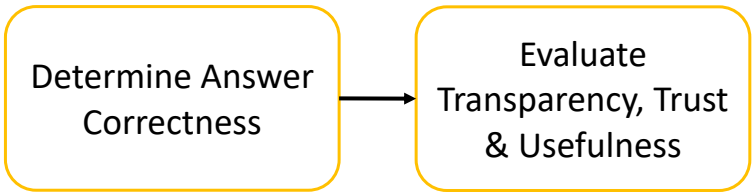
20 Chart-Question-Answer-Explanation tuples (5 per explanation type)

Determine Answer Correctness

83

16 participants

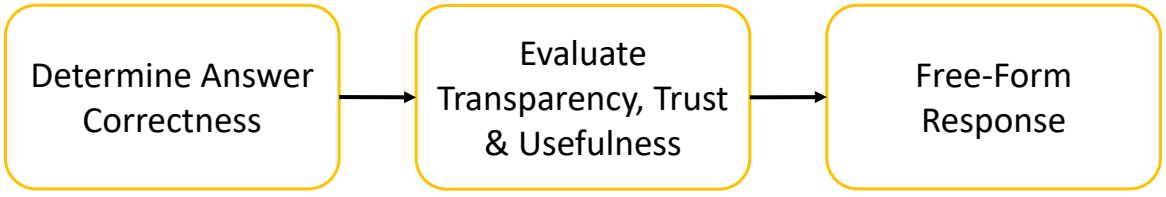
20 Chart-Question-Answer-Explanation tuples (5 per explanation type)



84

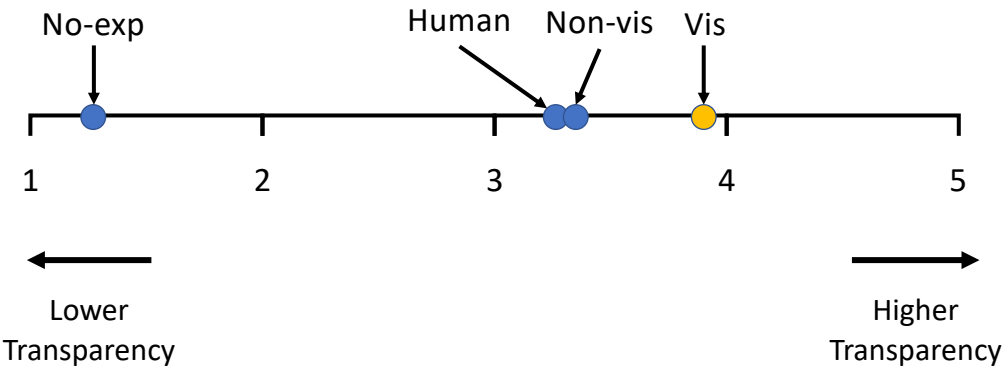
16 participants

20 Chart-Question-Answer-Explanation tuples (5 per explanation type)



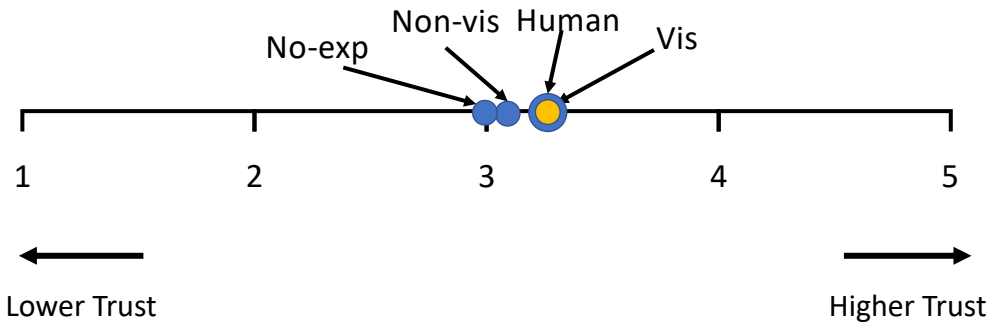
85

TRANSPARENCY



89

TRUST



94

Understanding how text and charts are related is difficult.

Tools that clarify connections between text and charts can help guide people towards the intended messages.

98

ANNOUNCEMENTS

99

FINAL PROJECT

Final Code and Video due Sun Dec 10, 8pm

Data analysis/explainer

Analyze dataset in depth & make a visual explainer

Deliverables

An article with multiple different interactive visualizations
Short video (2 min) demoing and explaining the project

Schedule

Project proposal: Mon 11/6
Design Review and Feedback: 9th week of quarter, 11/27 and 11/29
Final code and video: Sun 12/10 8pm

Grading

Groups of up to 3 people, graded individually
Clearly report responsibilities of each member

100

COURSE EVALUATION

Official course evaluation, due by 12/18

Your opinion is valued!

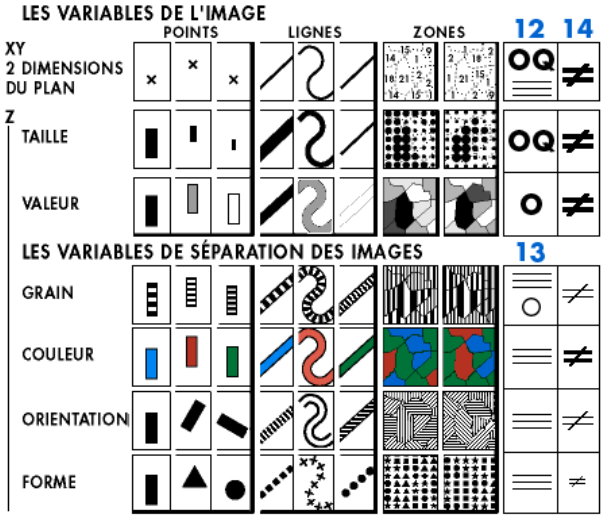
http://course-evaluations.stanford.edu

101

COURSE SUMMARY

102

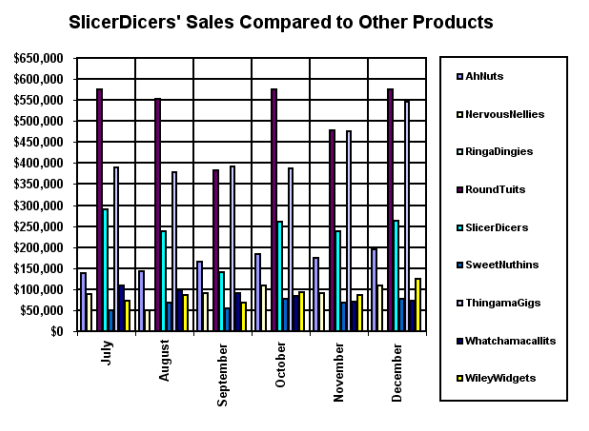
DATA AND IMAGE MODELS



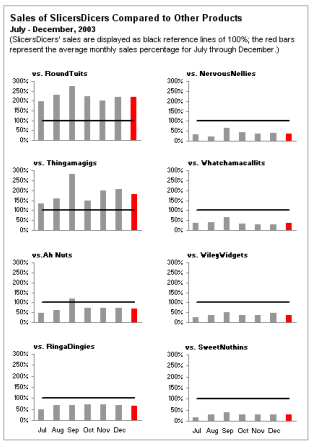
Graphics and Graphic Information Processing [Bertin 1981]

103

VISUALIZATION DESIGN AND REDESIGN



Problematic design



Redesign

104

EXPLORATORY DATA ANALYSIS

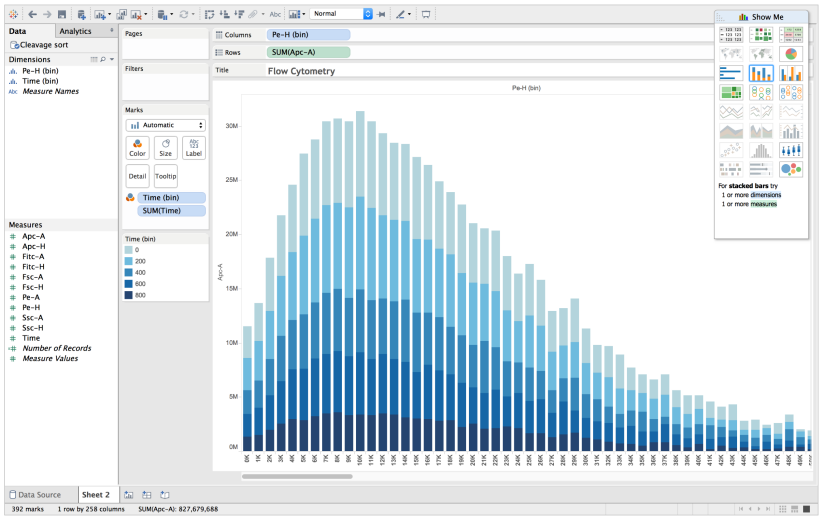
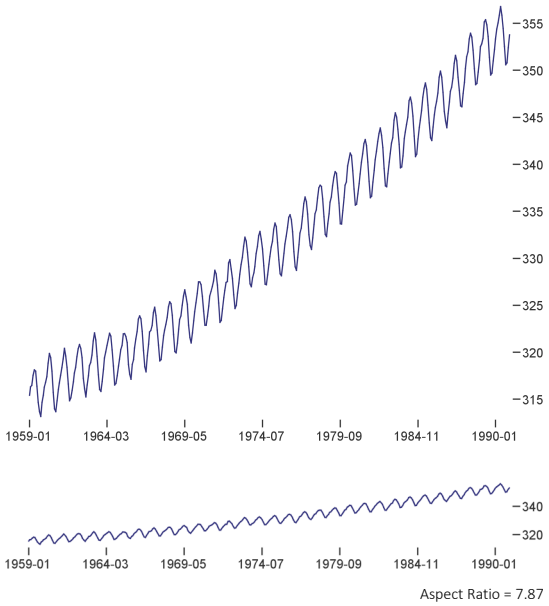


Tableau: based on Polaris [Stolte 2002]

105

PRINCIPLES OF CHART DESIGN

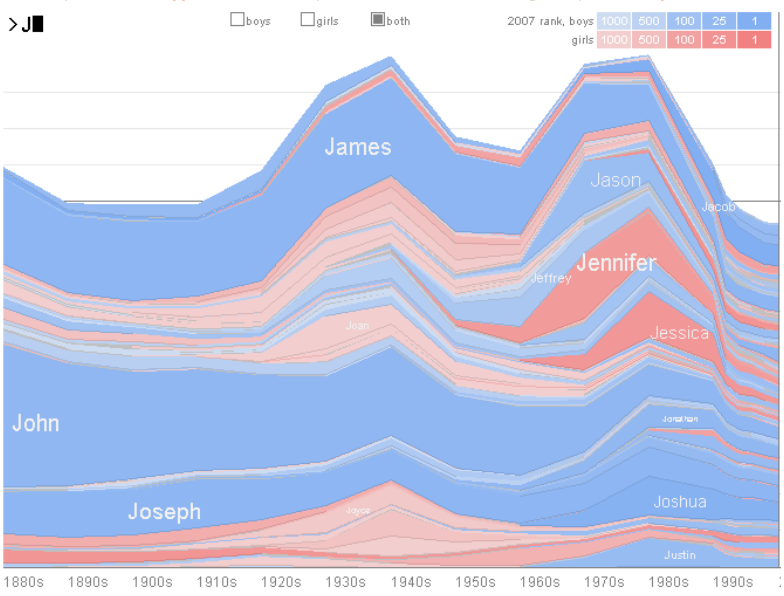


CO_2 Measurements
from William S. Cleveland
Visualizing Data, 1993

Trends may occur at different scales!
Apply banking to the original data or to fitted trend lines. [Heer & Agrawala 2006]

106

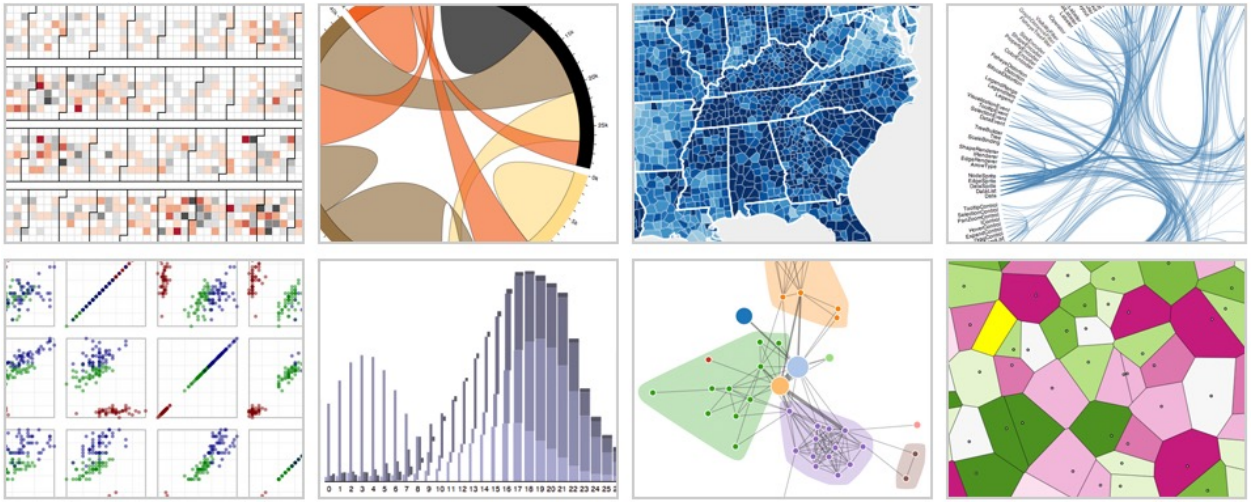
INTERACTION



Baby name voyager
[Wattenberg 2005]

107

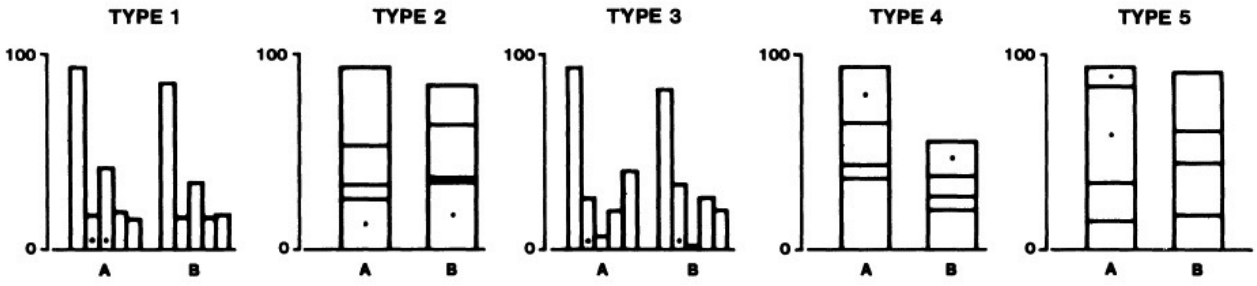
INTRODUCTION TO D3



D3: Data Driven Documents [Bostock 2011]

108

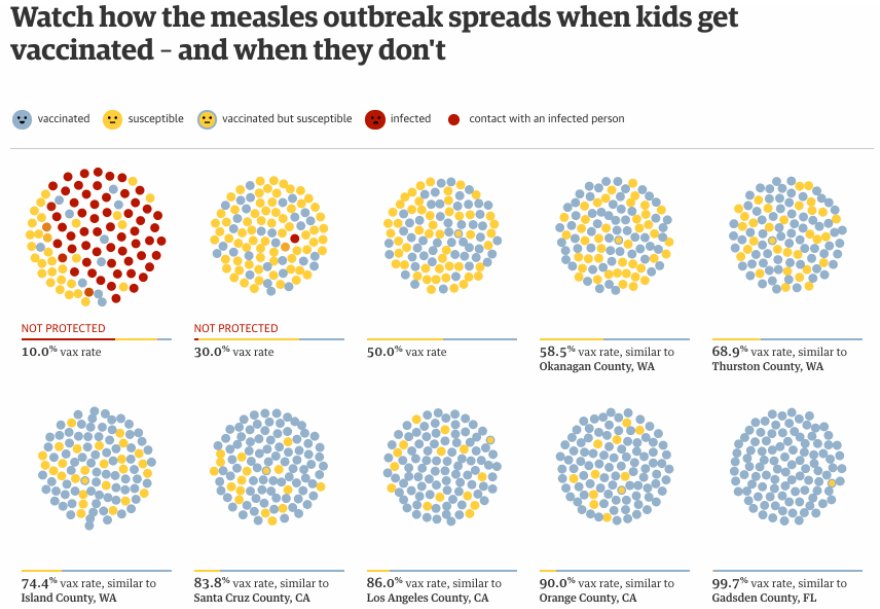
PERCEPTION



Graphical Perception [Cleveland and McGill 1984]

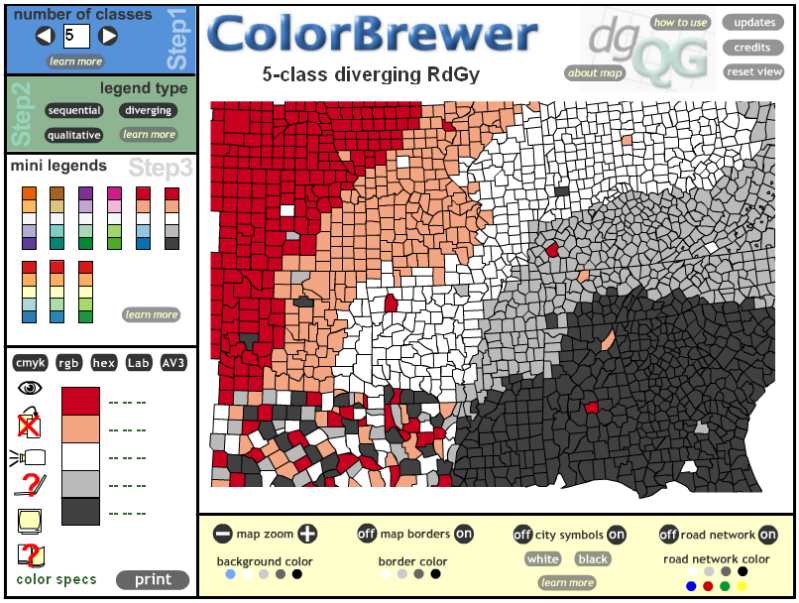
109

VISUAL EXPLAINERS



110

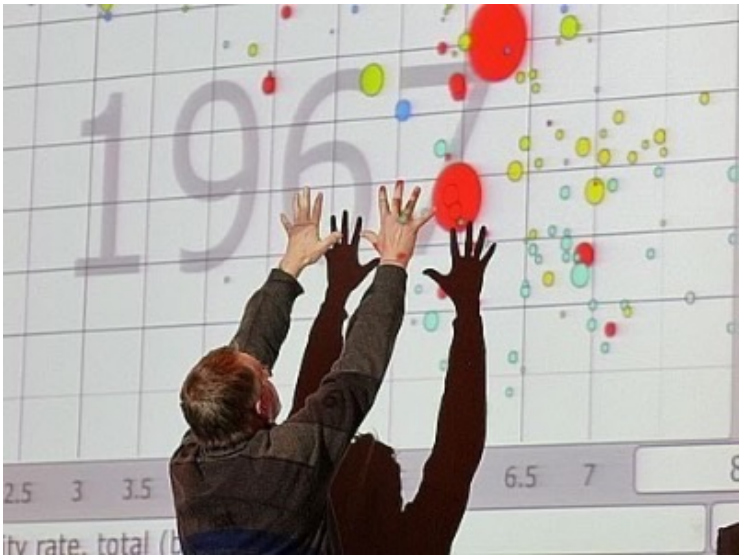
COLOR



Cynthia Brewer http://www.personal.psu.edu/faculty/c/a/cab38/

111

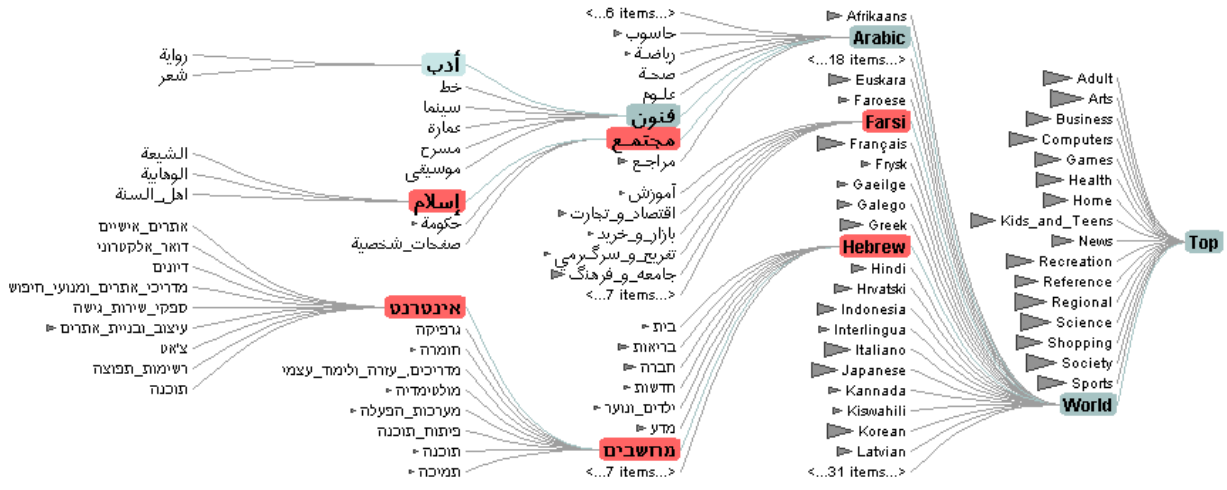
ANIMATION



Gapminder [Rosling 2006]

112

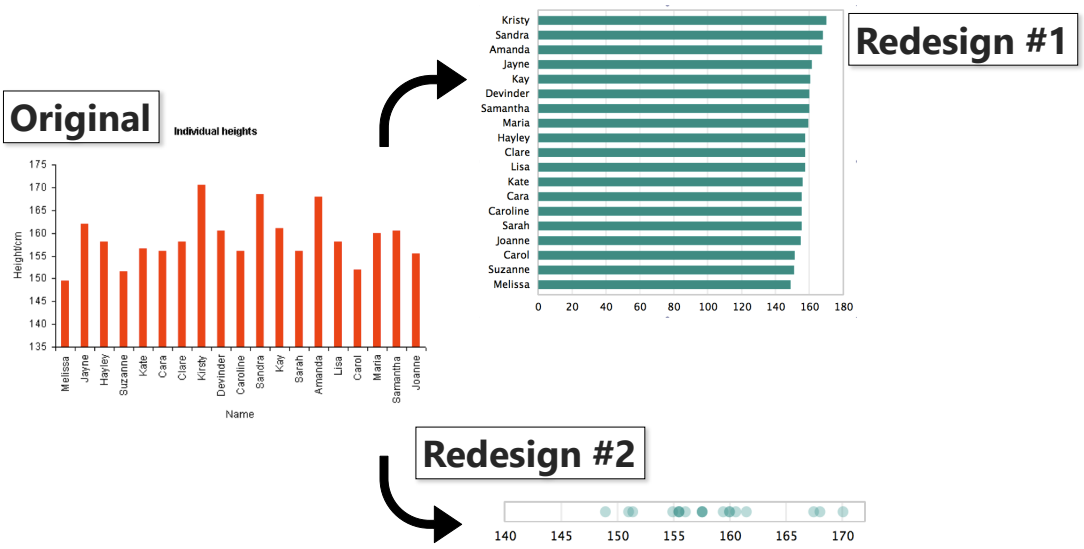
NETWORKS



Degree-of-Interest Trees [Heer 2004]

113

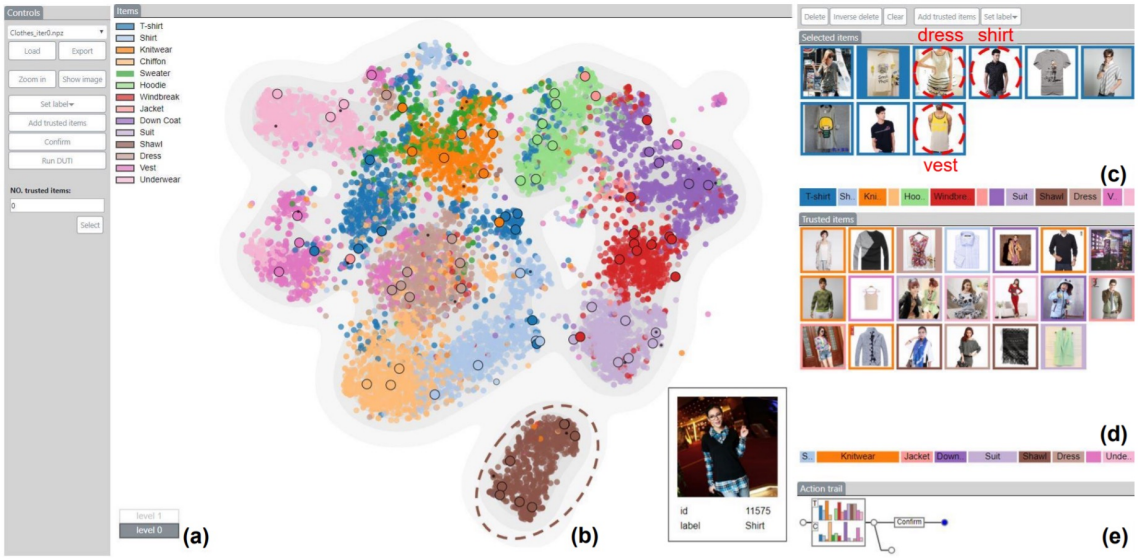
DECONSTRUCTING VISUALIZATIONS



ReVision [Savva 20011]

114

VISUALIZATION AND AI



115

THE FUTURE OF VISUALIZATION

Where is more work required?

What emerging technologies and societal trends will impact visualization design?

What did you find most difficult in creating visualizations and designing techniques?

116

Thank You!

117

118