Crowdsourcing Visual Analysis

Maneesh Agrawala
Jessica Hullman

CS 294-10: Visualization
Fall 2014

Last Time: Collaborative Visual Analysis

A graveyard of "YouTubes for Data"

Many Eyes – circa 4/2012

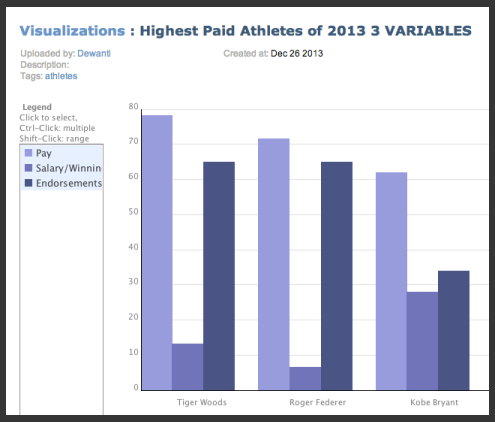
128,478 Visualizations

17,340 Comments

only ~11% of comments provided a plausible hypothesis or explanation for the data in the chart

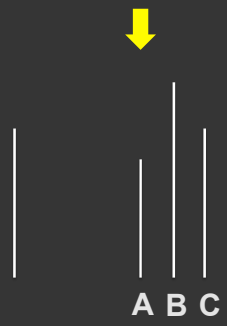
Can having access to social information negatively influence analysis?

The Impact of Social Information on Visual Judgments



Does information on others' graphical judgments impact the accuracy of a new person's judgment?

Social Influence



Asch 1951

Social Influence

Normative versus Informational
(fitting in) *(being correct)*

Using social signals as evidence when we're choosing:

- Restaurants
- Music (Salganik et al. 2009)
- Tags (Golder et al. 2007)

Controlled Experiment

Does information on others' graphical judgments impact the accuracy of a new person's judgment?

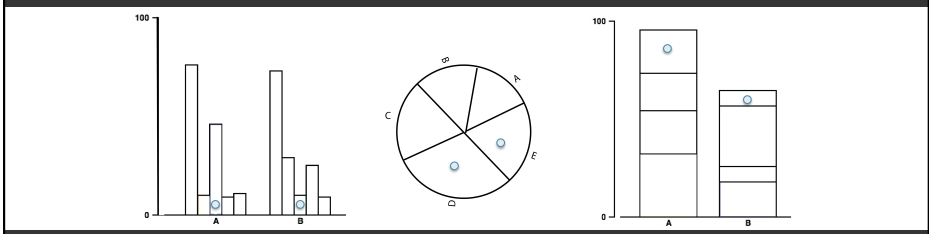
Two conditions:

Non-social: Visual judgment task with no social information

Social: Same task, presented with summary of judgments from non-social group

Participants are incentivized ($$) to be correct.

Visual judgment task

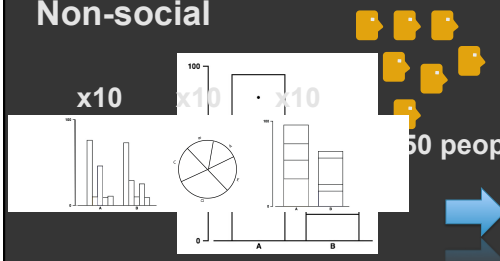


Cleveland & McGill 1984

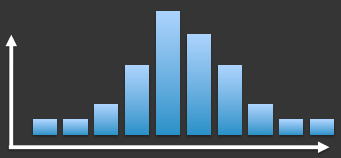
1. Which of the two marked sections is larger?
2. Make a quick visual judgment of what proportion the smaller section is of the larger.

AMT study: Heer & Bostock 2010

Non-social



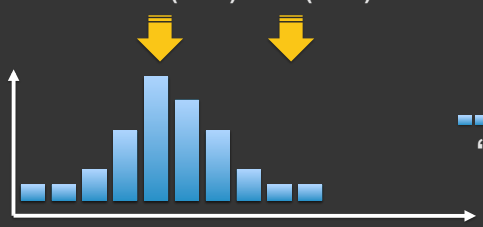
50 people



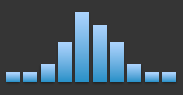
proportion estimate

Social

Mean (56%) 1SD (78%)



proportion estimate

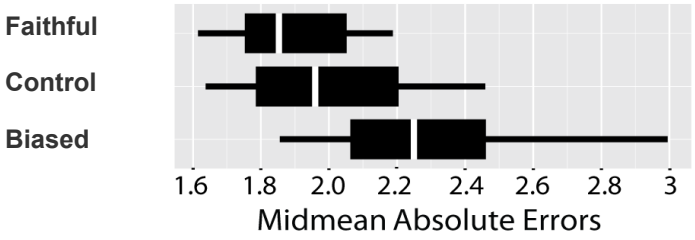


"Faithful"

"Biased"

Results

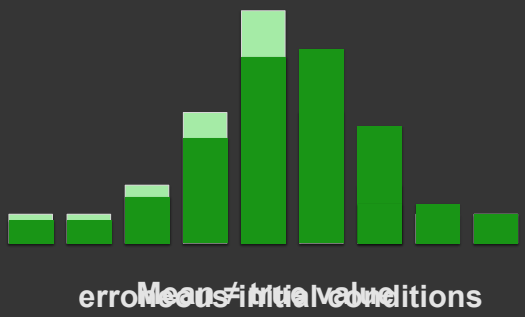
Does information on others' graphical judgments impact the accuracy of a new person's judgment?



Does the amount of social information matter?

Implications

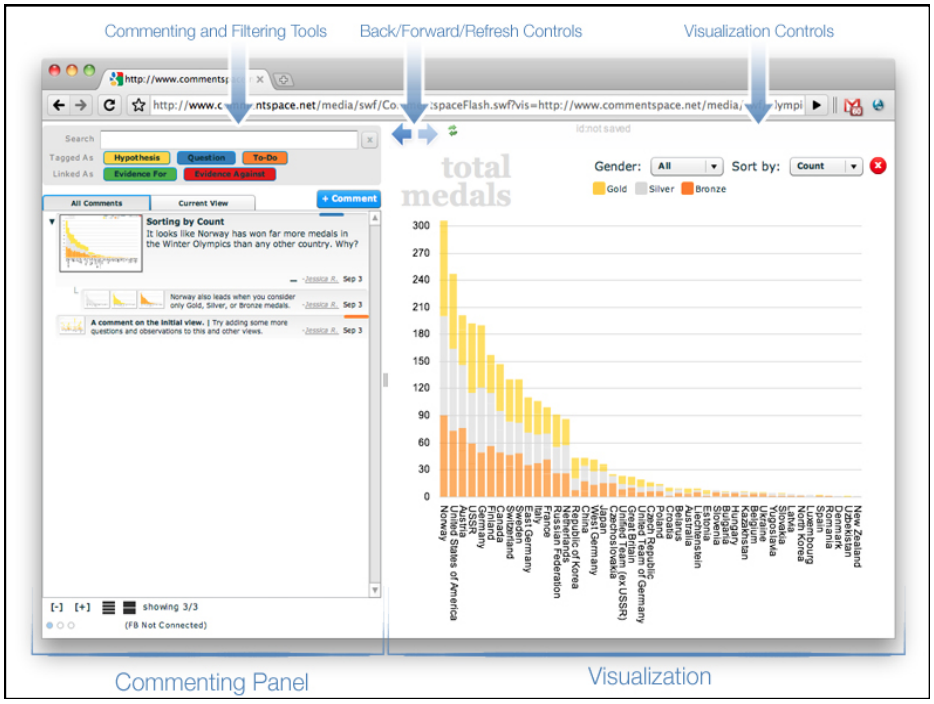
No social information?



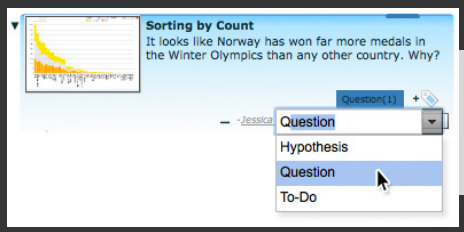
Multiple issues when collaborative analysis remains shallow.

Can we augment social data analysis to support deeper analysis and synthesis?

CommentSpace: Structured Support for Social Data Analysis



Tags and links for organization

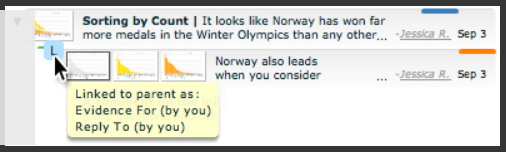


Tags

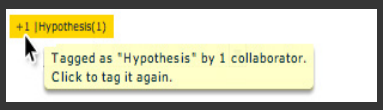
- Hypothesis
- Question
- To-Do

Links

- Evidence For
- Evidence Against
- (Related)

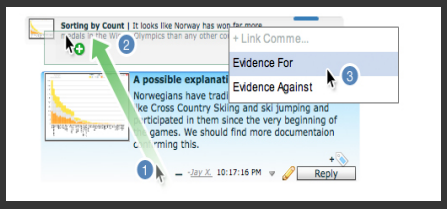


Hypothesis generation/evidence gathering

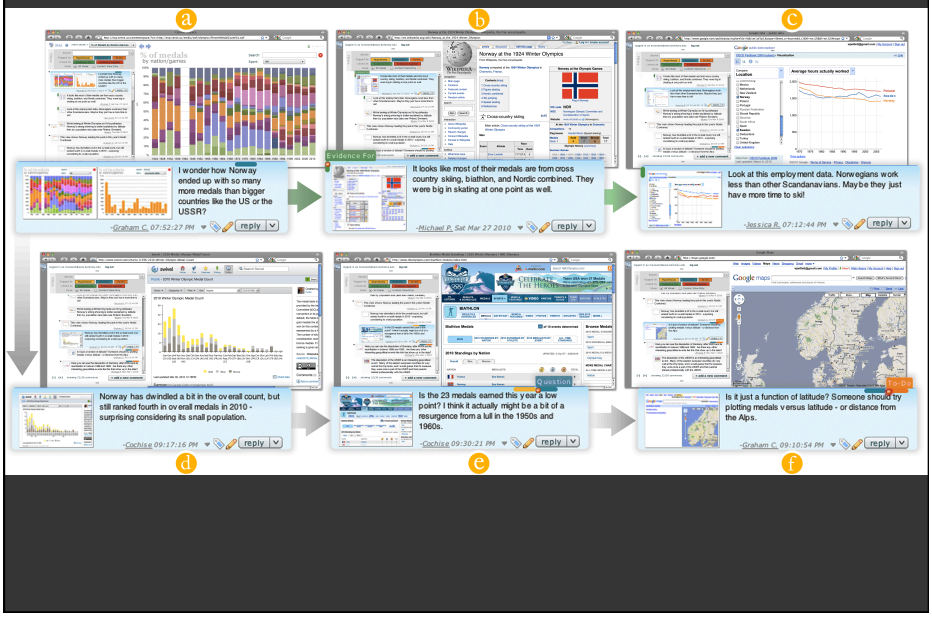


Contributors highlight important items with tags

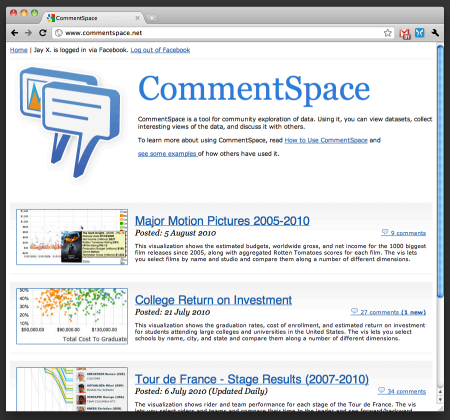
Tags help late-joiners identify and build on important comments



Contributors use links to organize contributions and build narrative



Studies and Deployments



Controlled lab studies to test core analysis subtasks

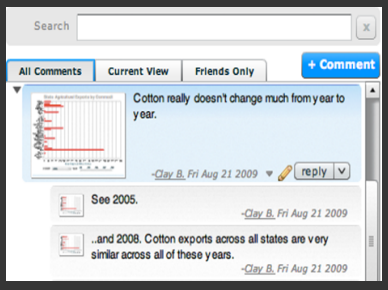
Live deployments (www.commentspace.net)

Study: Use of Tags and Links

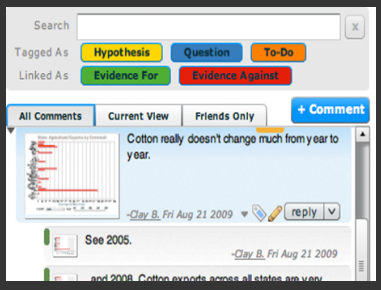
Hypothesis: Tags and links can provide common ground and encourage continued discussion.

A between subjects study (n=24) with 2 conditions.

"No-Tag" Condition

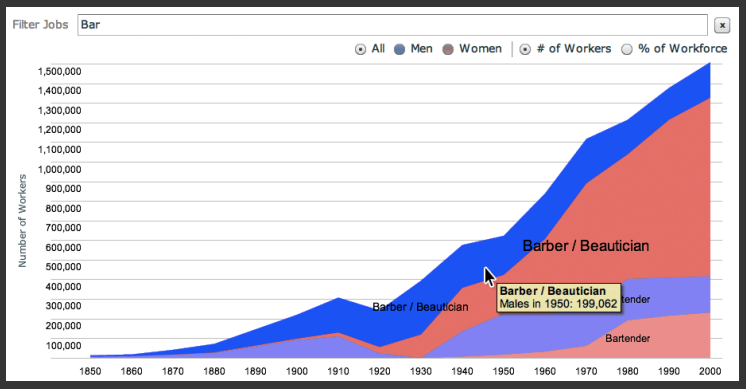


"Tag" Condition



Part 1: Sensemaking from existing comments (classify given hypothesis)
Part 2: Find views and generate comments.

Study: Prompt



Hypothesis: Stereotypically male jobs have remained almost entirely male even as women have joined the work force.

Study: Results

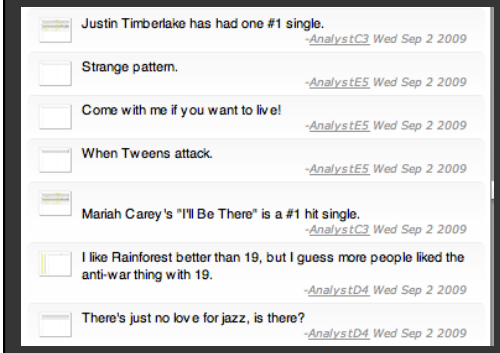
Participants who used tags and links classified comments more consistently and accurately than those who didn't

(greater in-group agreement) (greater agreement with experts)

Participants using tags and links generated significantly more replies to existing comments

Tag (Median=7)
No-Tag (Median=2)

But ….



In open-ended tasks, participants still engaged mostly in superficial, exploratory analysis

We saw very little use of tags or links

Can we use paid crowds to perform small pieces of analysis tasks?

Announcements

Final project

Design new visualization method

- Pose problem, Implement creative solution

Deliverables

- Implementation of solution
- 8-12 page paper in format of conference paper submission
- 1 or 2 design discussion presentations

Schedule

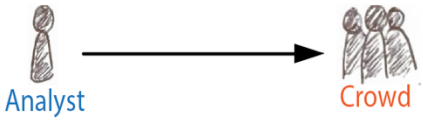
- Project proposal: 10/28
- Project presentation: 11/13-11/20
- Final paper and presentation: 12/2-12/6

Grading

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

Crowdsourcing Visual Analysis

Can we use paid crowds to perform small pieces of analysis tasks?



Open ended crowd tasks - problems

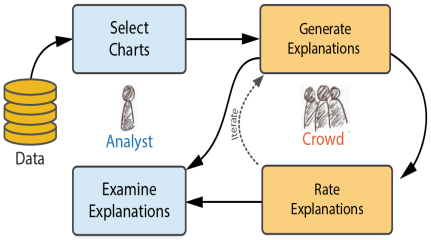
Expectations may be unclear to workers.

Workers may explain irrelevant features.

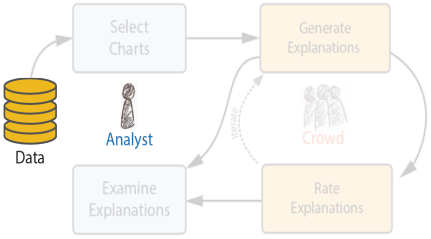
Workers' may give speculative explanations.

Workers may not attend to chart details.

How can we organize crowdsourced visual analysis tasks more effectively?



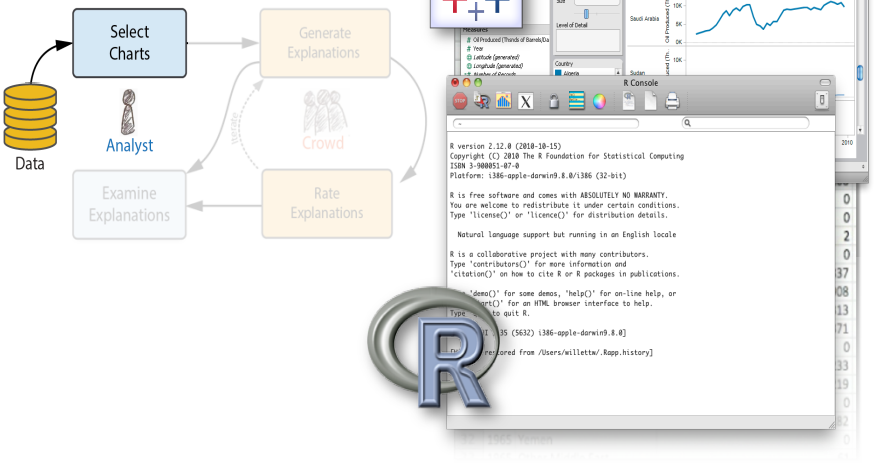
Data



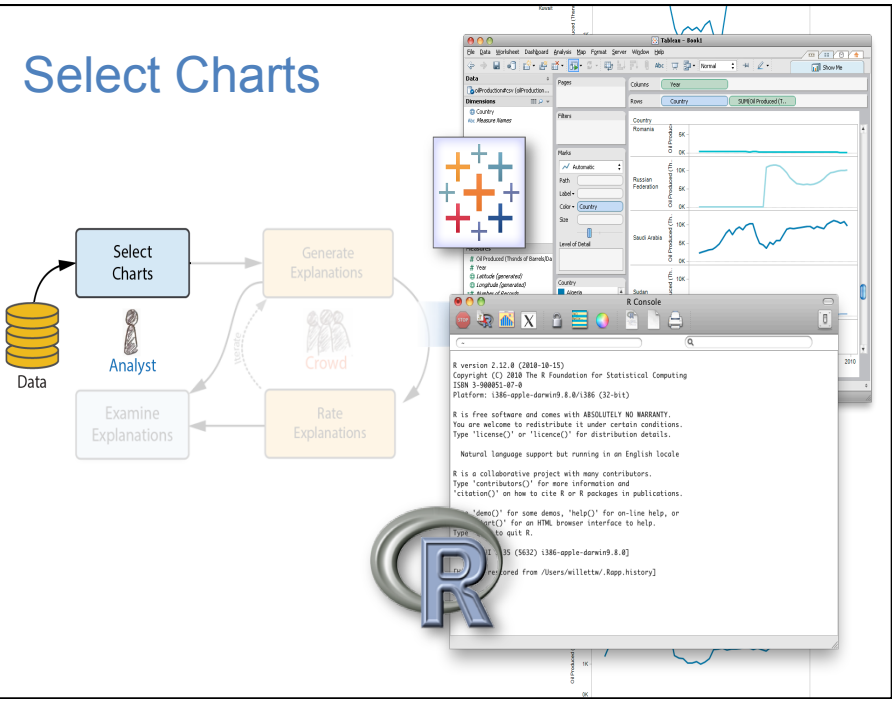
World Oil Production 1965-2010

	Year	Country	Oil Produced (Thsnds. of Barrels/Day)
2	1965	United States	9014
3	1965	Canada	920
4	1965	Mexico	362
5	1965	Argentina	276
6	1965	Brazil	96
7	1965	Colombia	203
8	1965	Ecuador	8
9	1965	Peru	66
10	1965	Trinidad & Tobago	135
11	1965	Venezuela	3503
12	1965	Other S. & Cent. America	47
13	1965	Azerbaijan	0
14	1965	Denmark	0
15	1965	Italy	48
16	1965	Kazakhstan	0
17	1965	Norway	0
18	1965	Romania	266
19	1965	Russian Federation	0
20	1965	Turkmenistan	0
21	1965	United Kingdom	2
22	1965	Uzbekistan	0
23	1965	Other Europe & Eurasia	5337
24	1965	Iran	1908
25	1965	Iraq	1313
26	1965	Kuwait	2371
27	1965	Oman	0
28	1965	Qatar	233
29	1965	Saudi Arabia	2219
30	1965	Syria	0
31	1965	United Arab Emirates	282
32	1965	Yemen	0

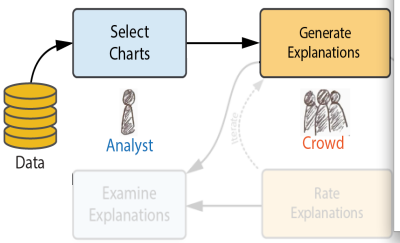
Select Charts



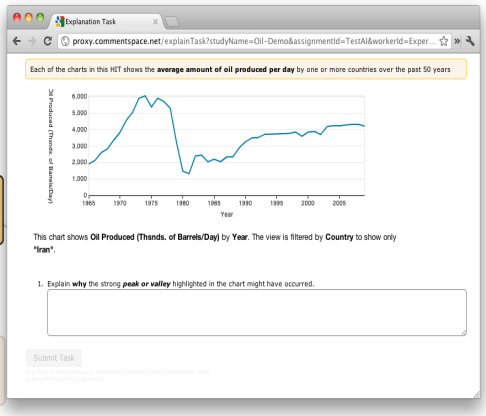
Select Charts



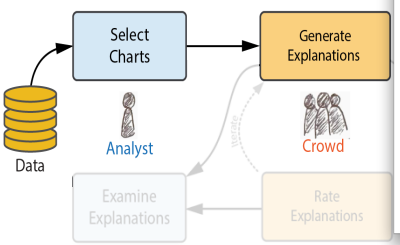
Generate Explanations



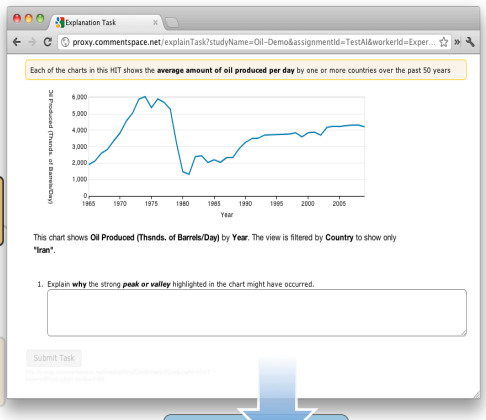
Explanation Microtask

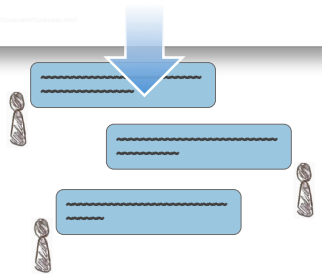


Generate Explanations

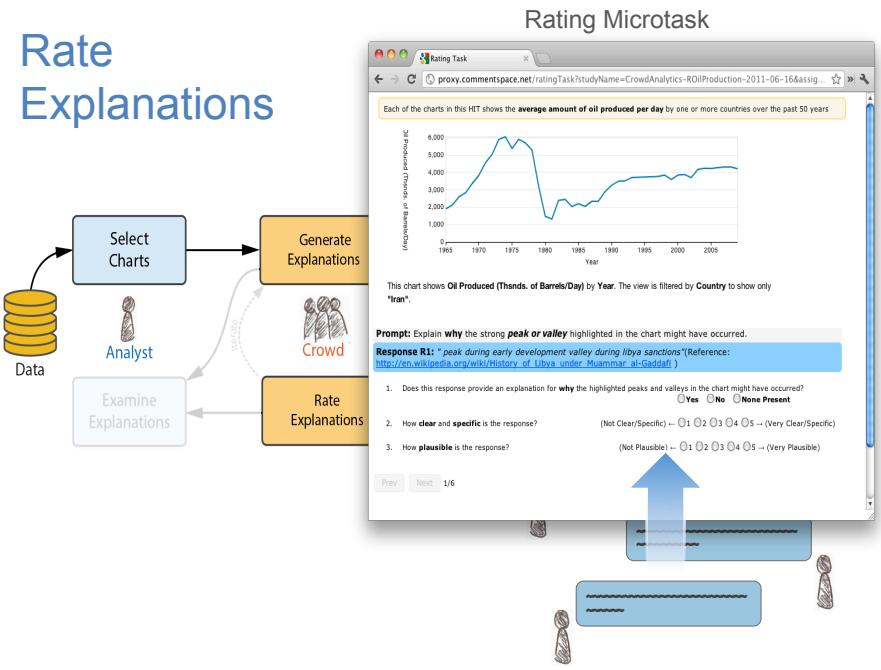


Explanation Microtask

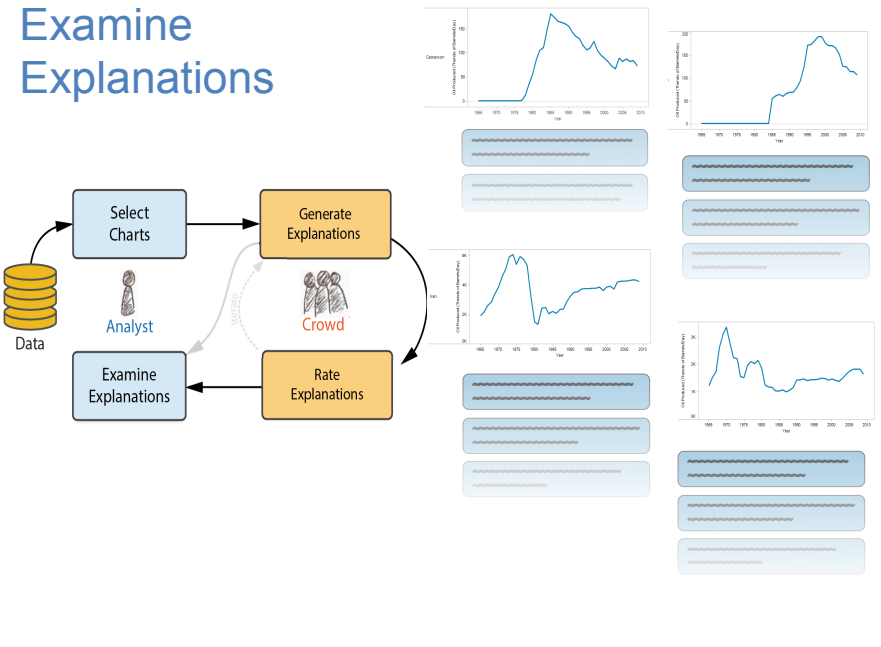




Rate Explanations

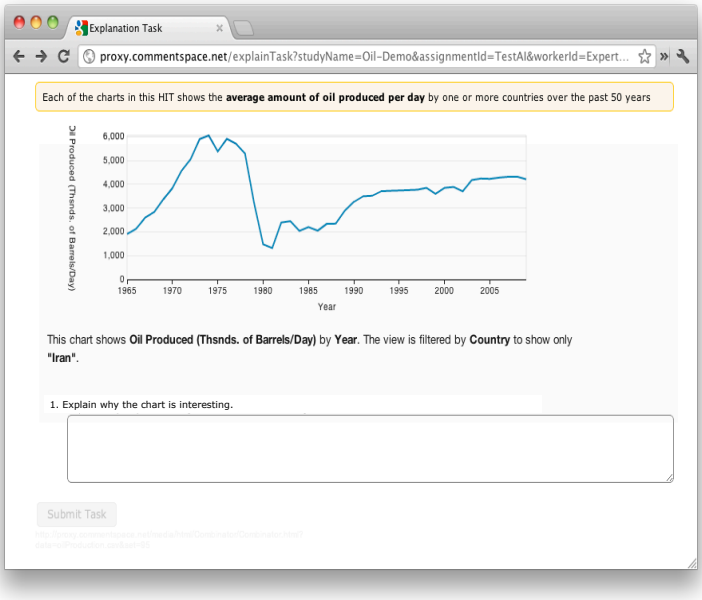


Examine Explanations



Seven Strategies for Crowdsourcing Social Data Analysis

(Typical) Explanation Microtask



Problems

Expectations may be unclear to workers.

Workers may explain irrelevant features.

Workers' may give speculative explanations.

Workers may not attend to chart details.

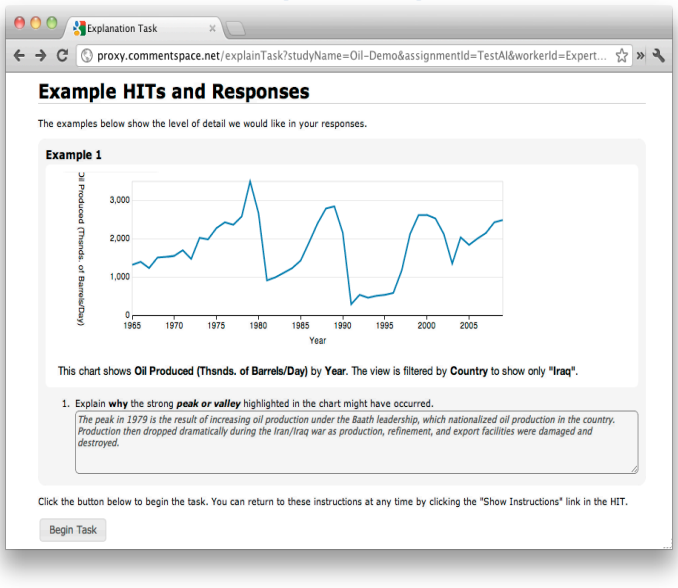
S1. Use feature-oriented prompts.

~~"Explain why the chart is interesting."~~

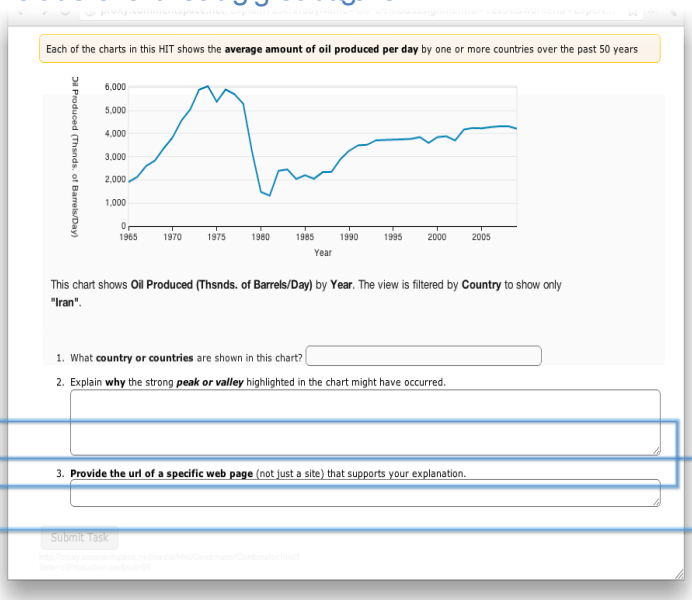
"Explain the **long term trend** in the chart."

"Explain the **peaks and/or valleys** in the chart (if any exist)."

S2. Provide sample explanations

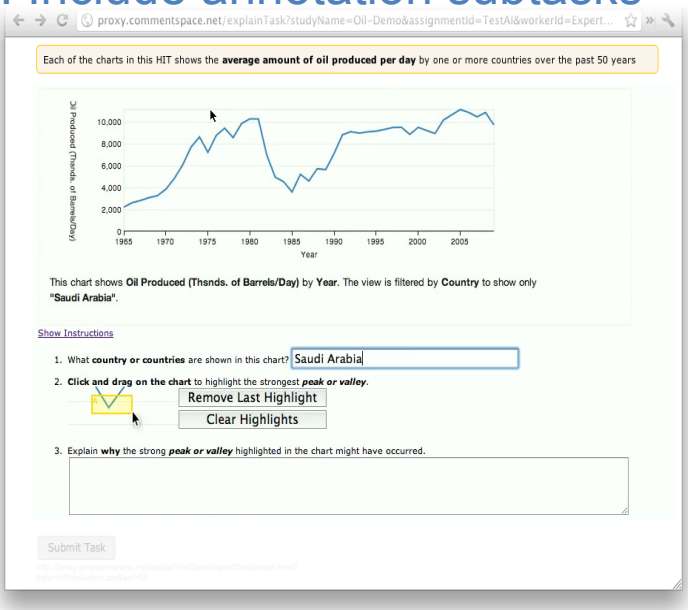


S3. Include chart reading subtasks



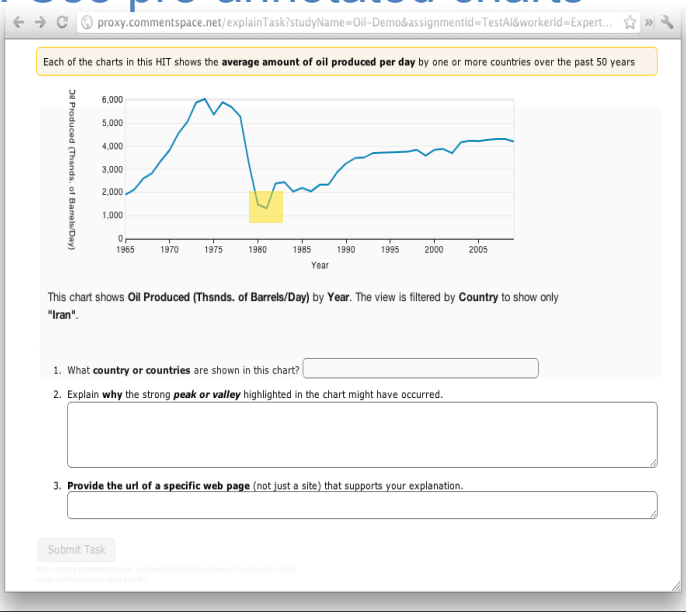
Explanation Microtask

S5. Include annotation subtasks

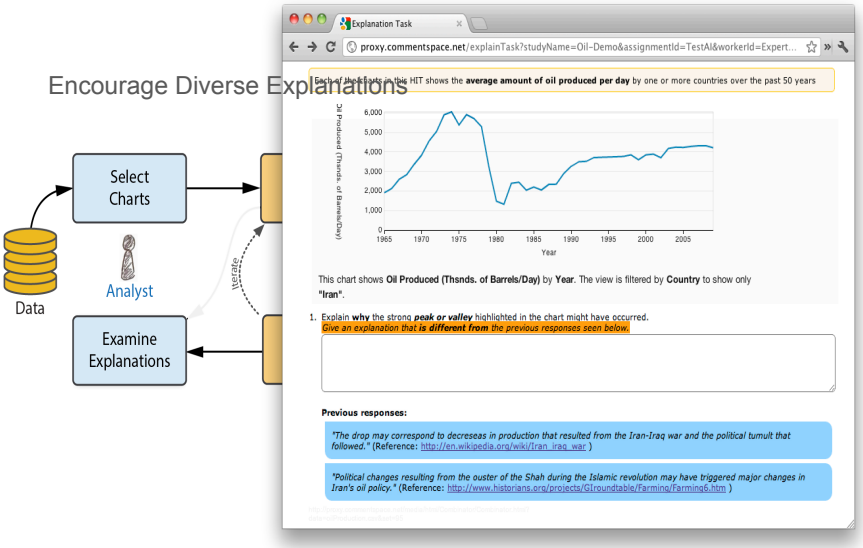


Explanation Microtask

S6. Use pre-annotated charts



S7. Elicit explanations iteratively



Deployment

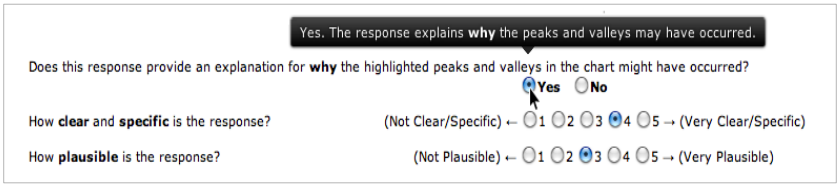
Deployment via Mechanical Turk

910 Responses for
64 Charts from
16 Datasets

Quantifying Response Quality

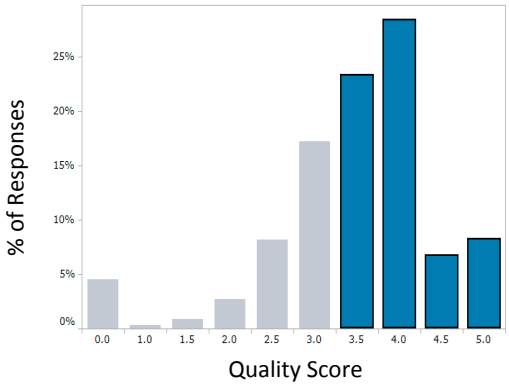
relevance (0-1) Does the response explain the requested feature?
clarity (1-5) How clear and specific is the explanation?
plausibility (1-5) How likely is the explanation to be valid?

$$quality = relevance * (clarity + plausibility) / 2$$

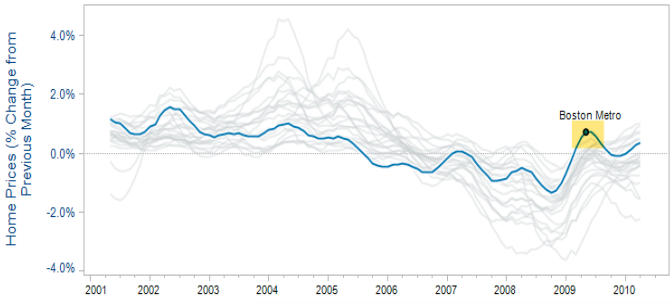


Experts scored response quality this way.

Based on our ratings, the majority (63%) of responses were of high quality (*quality* ≥ *3.5*).



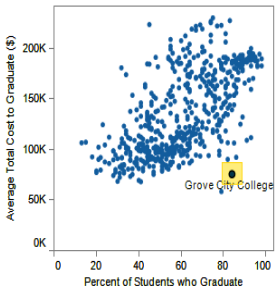
Dataset: US Metro Home Prices (% Change / Month) 2001-2010



"The primary reason for the peak is …
… annual variation where home prices rise in the spring.
… the $8,000 home buyer tax credit.

url: http://articles.boston.com/2011…indices-housing-prices

Dataset: US College Graduation Rates vs. Cost To Graduate



"Grove City College is Christian **based and** strives to keep tuition rates affordable **for most students.**
… located in a rural community …
… selective enrollment …"

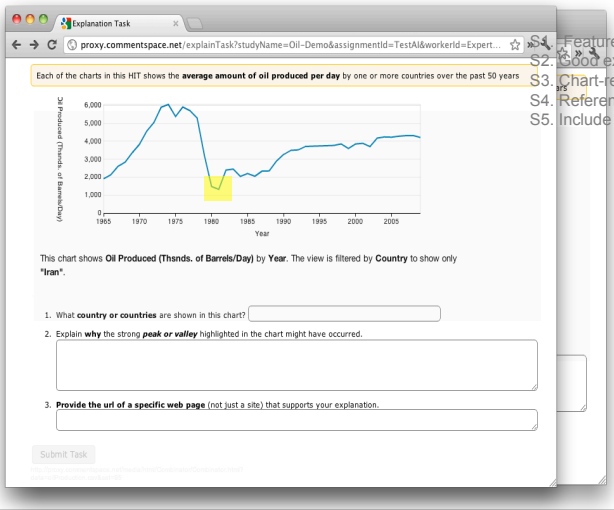
url: http://stateuniversity.com/…college.html

Do our strategies improve explanation quality?

Experiment 1 – Strategies S1-S5

No Strategies vs. Strategies (S1-S5)

US



S1. Feature-oriented prompts.
S2. Good examples.
S3. Chart-reading subtasks.
S4. Reference-gathering subtasks.
S5. Include annotation subtasks.

Experiment 1 - Strategies S1-S5

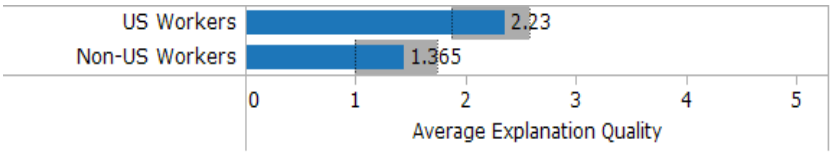
Multiple trials per condition, 4 conditions
(2 interfaces x 2 worker pools)

H1: Results from US workers would be of higher quality than results from non-US workers, but

H2: Employing strategies S1-S5 would increase the quality for both groups.

Experiment 1

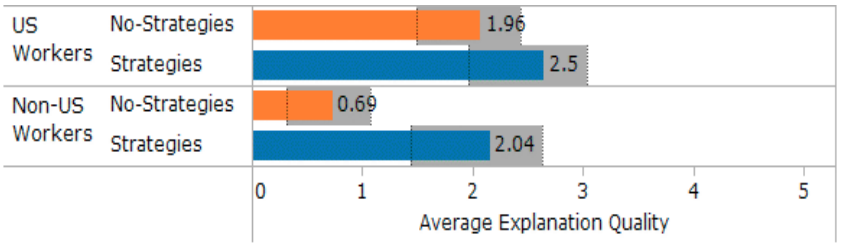
US Workers performed better than Non-US.



Experiment 1

Using S1-S5 increased explanation quality:

196% for Non-US Workers
28% for US Workers



Significant main effects for { **Worker Pool ($F_{1,198} = 12.2$, $p < 0.01$)**
Strategies ($F_{1,198} = 14.5$, $p < 0.01$)

Experiments

Reference Gathering (S4)

Produced useful references, but slowed workers, increased attrition, and lowered quality.

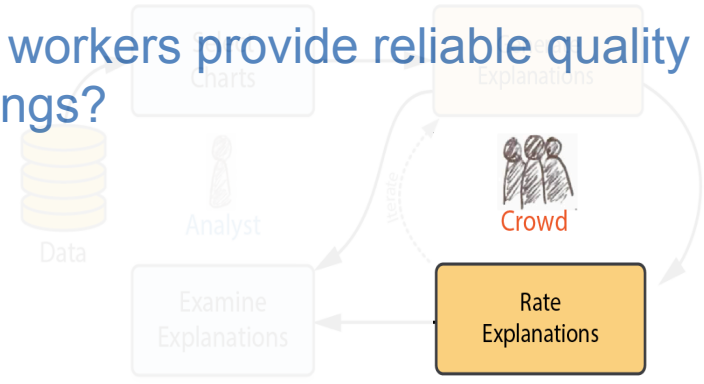
Annotation Subtasks (S5) and Pre-Annotation (S6)

Both improve attention to detail but are useful in different situations.

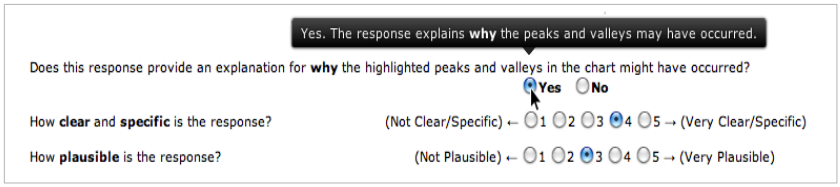
Eliciting Explanations Iteratively (S7)

An iterative round produced 71% new explanations, increasing diversity.

Do workers provide reliable quality ratings?



Experiment 2 - Rating



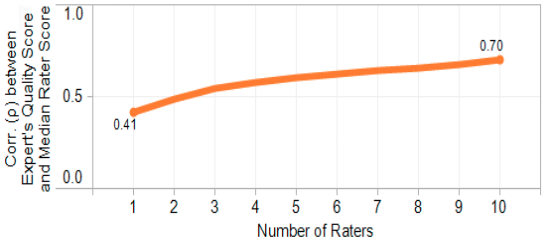
Elicited 1,334 ratings for 192 responses.

Compared against expert quality ratings.

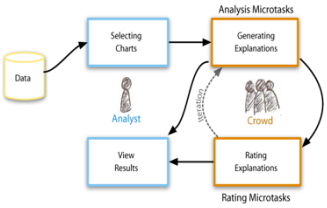
Experiment 2 - Rating

Individual workers' quality scores were moderately correlated with ours.
(ρ = 0.41)

The median quality from multiple workers correlates more strongly correlation with our scores.
(ρ = 0.70 w/ 10 raters)



Takeaways



A workflow for generating and assessing explanations for trends and outliers.

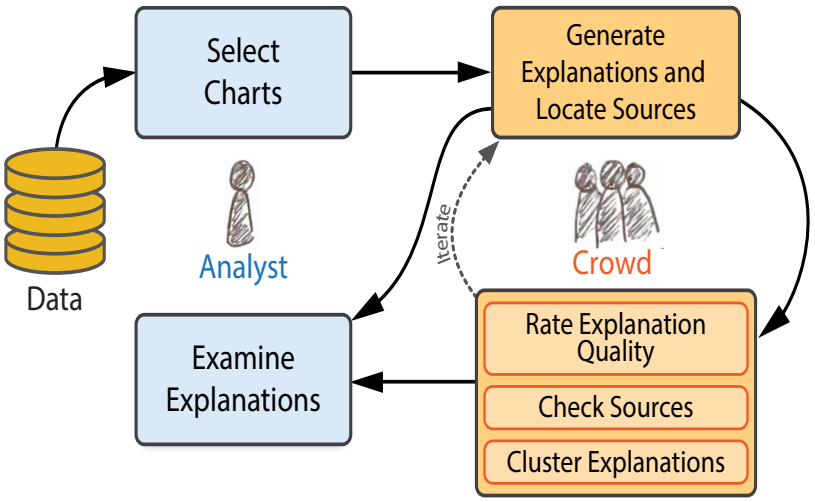
S1. Use feature-oriented prompts.
S2. Provide good examples.
S3. Include chart-reading subtasks.
S4. Include reference-gathering subtasks.
S5. Include annotation subtasks.
S6. Pre-annotate charts.
S7. Elicit Explanations Iteratively

Seven strategies for improving crowdsourced explanations.

Generates good explanations (63% were high-quality).

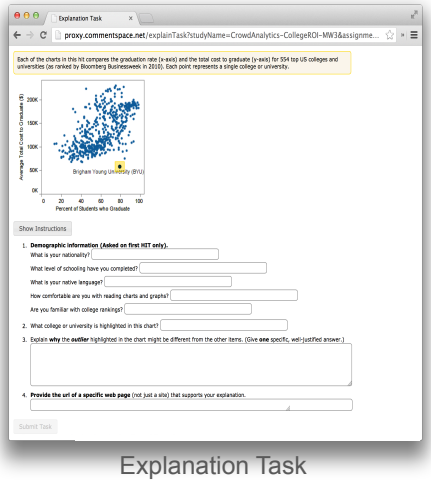
Exposing Provenance and Identifying Redundancy

Adding Additional Crowd Processing



Provenance

Provenance

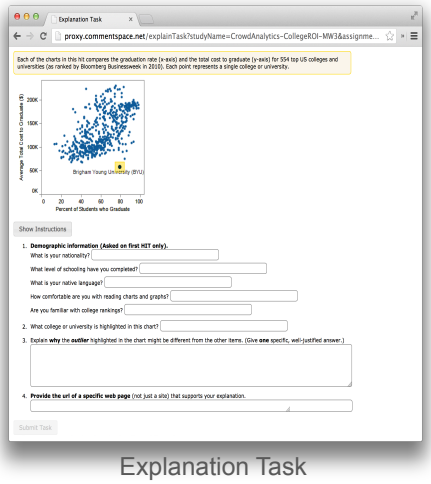


Explanation Task

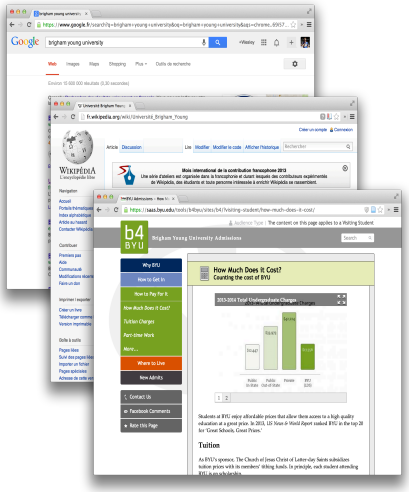
What are the

workers doing?

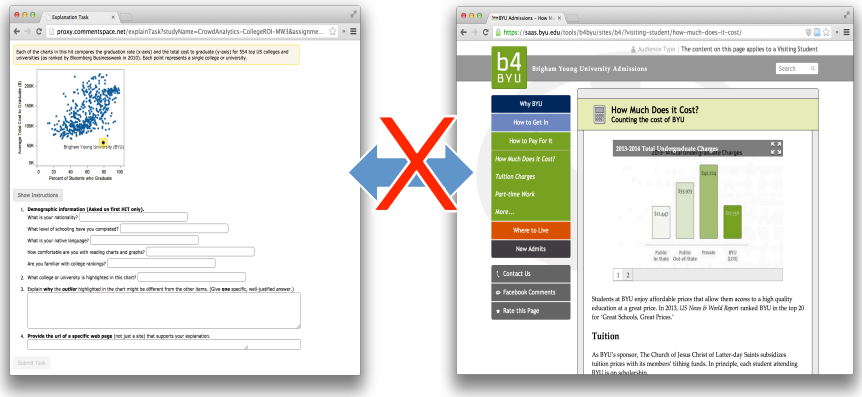
Provenance



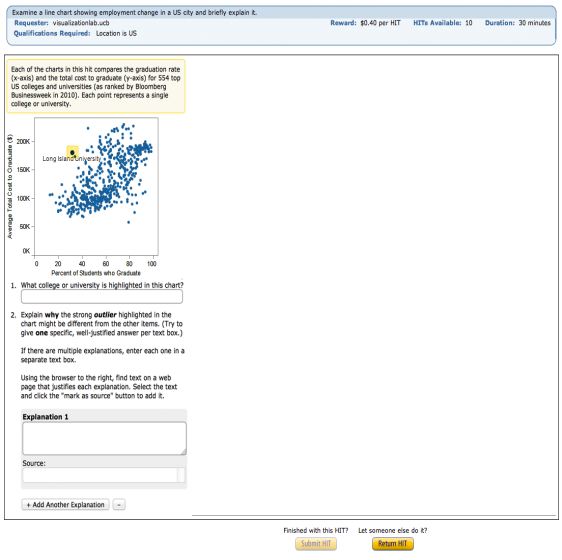
Explanation Task



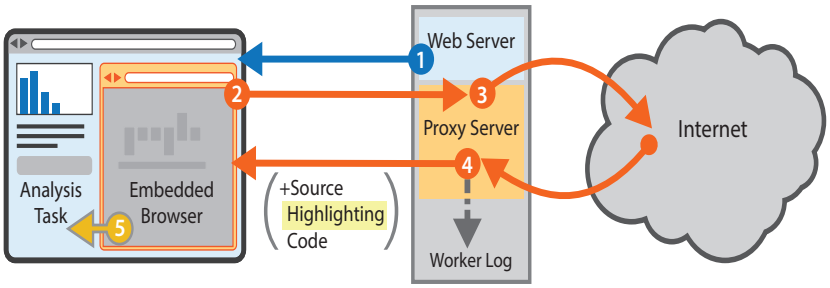
Cross-Domain Access is Limited



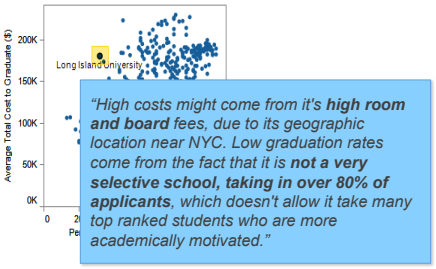
Instrumenting Explanation Tasks



Logging via Proxied Browser



Provenance



*"High costs might come from it's **high room and board** fees, due to its geographic location near NYC. Low graduation rates come from the fact that it is **not a very selective school, taking in over 80% of applicants**, which doesn't allow it take many top ranked students who are more academically motivated."*

Visitation logs

```
2011-12-11 09:22:04 google.com
2011-12-11 09:22:04 sgr:helo
2011-12-11 09:23:08 google.com/search?hl=en&source=hp
2011-12-11 09:23:11 google.com/search?hl=en&q=Long Is
2011-12-11 09:23:13 google.com/search?q=Long Island Un
2011-12-11 09:23:31 google.com/search?q=Long Island Un
2011-12-11 09:23:38 google.com/search?q=Long Island Un
2011-12-11 09:23:43 google.com/search?q=Long Island Un
2011-12-11 09:23:54 google.com/search?q=Long Island Un
2011-12-11 09:24:09 colleges.usnews.rankingsandreviews.c
```

Paragraph-level citations

#123 Regional Universities (North)

Summary

LIU Post is a private institution that was founded in 1954. It has a total undergraduate enrollment of 8,315, its setting is suburban, and the campus size is 308 acres. It utilizes a semester-based academic calendar. LIU Post's ranking in the 2014 edition of Best Colleges is Regional Universities (North), 123. Its tuition and fees are $34,070 (2013-14).

2014 Quick Stats

720 Northern Boulevard
Brookville, NY 11548-1300
[map]
Phone: (516) 299-2000

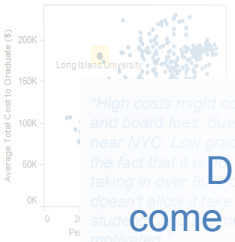
2013-2014 Tuition
$34,070 tuition and fees

Students
8,315 enrolled
25% male / 75% female

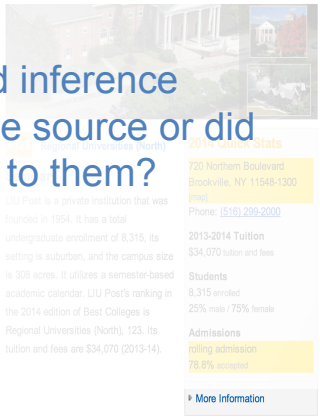
Admissions
rolling admission
78.8% accepted

More Information

Provenance

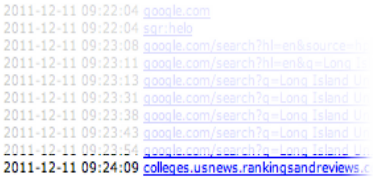


Paragraph-level citations

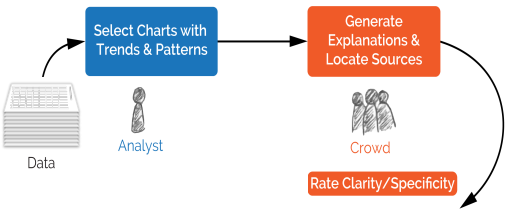


Did the facts and inference come directly from the source or did the worker add to them?

Visitation logs



Source-Checking Microtasks

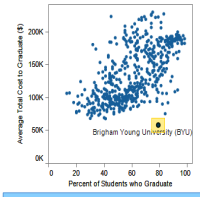


A second group of workers **verifies links** and attributes explanations to the **source** or the **worker**.
(75% accurate in our preliminary tests)

Redundancy

Redundancy

Many explanations provided by workers are redundant.



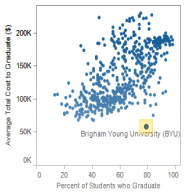
"The Church of Jesus Christ of Latter Day Saints pays a significant part of the tuition costs…"

"The cost of attendance at BYU is subsidized by the LDS church."

"98% of their students are members of LDS and they have lowered tuition…"

Redundancy

Many explanations provided by workers are redundant.



"The Church of Jesus Christ of Latter Day Saints pays a significant part of the tuition costs…"

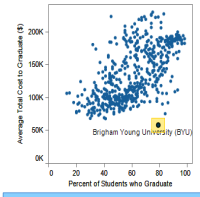
"The cost of attendance at BYU is subsidized by the LDS church."

"98% of their students are members of LDS and they have lowered tuition…"

- Duplicate results for analysts to examine.
- Redundancy can signal high support and corroborating sources.

Redundancy

Automated text similarity methods don't deal well with these kinds of content.

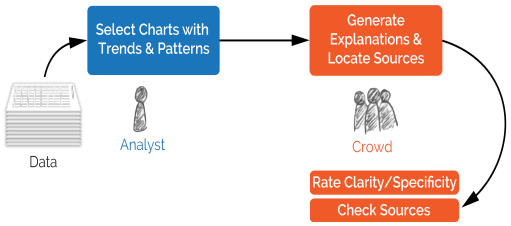


"The Church of Jesus Christ of Latter Day Saints pays a significant part of the tuition costs…"

"The cost of attendance at BYU is subsidized by the LDS church."

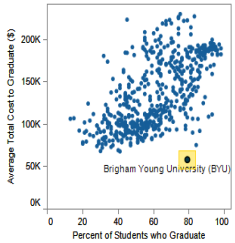
"98% of their students are members of LDS and they have lowered tuition…"

Redundancy



Can we crowdsource redundancy detection?

Clustering via Distributed Comparison

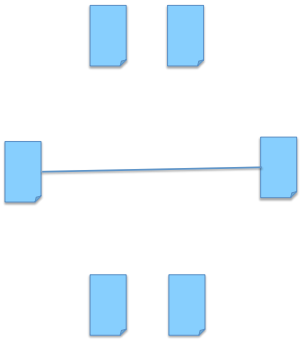


"98% of their students are members of LDS and they have lowered tuition…"

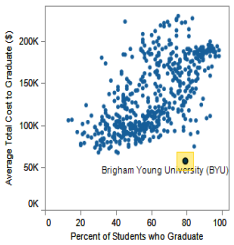
"The cost of attendance at BYU is subsidized by the LDS church."

"…students are mostly members of the church and bound by the honor code..."

"The Church of Jesus Christ of Latter Day



Clustering via Distributed Comparison

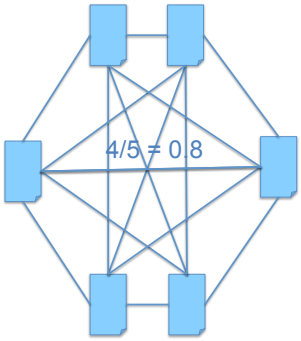


"98% of their students are members of LDS and they have lowered tuition…"

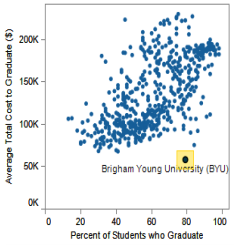
"The cost of attendance at BYU is subsidized by the LDS church."

Do these two responses give the same general explanation for the peaks and valleys in the chart?

- **Yes. Both responses give the same general explanation.**
- **No. The responses do not give the same explanation.**



Clustering via Distributed Comparison

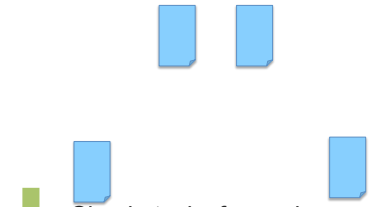


"98% of their students are members of LDS and they have lowered tuition…"

"The cost of attendance at BYU is subsidized by the LDS church."

Do these two responses give the same general explanation for the peaks and valleys in the chart?

- **Yes. Both responses give the same general explanation.**
- **No. The responses do not give the same explanation.**



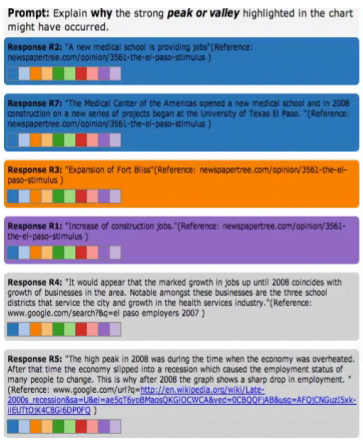
Simple tasks for workers

Scales p rly

Sensitive to **clustering method**

Workers have **little context**

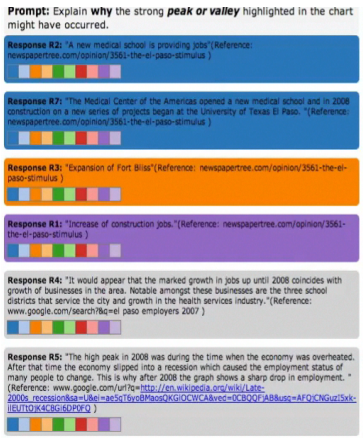
Clustering via Color-Coding



Individual workers cluster the whole set.

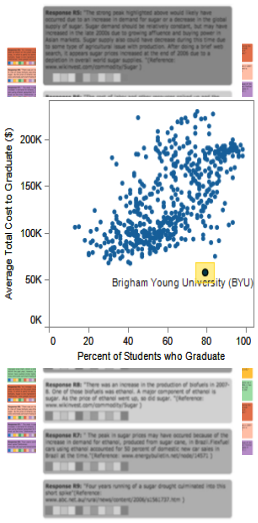
- ✚ Workers have complete context
- **Individual workers** can cluster badly
- ▬ Hard to integrate clusterings from multiple workers

How to Integrate Color-Clusterings?



1. A **single worker's clustering** is preferable to a combination of multiple clusterings.
2. **Clusters reproduced by multiple independent workers** are likely to reflect actual redundancy.
3. Errors tend to be either **noisy** or **easy to catch**.

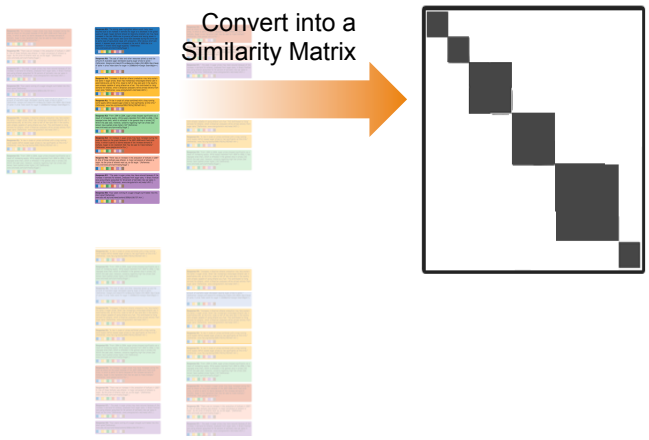
How to Integrate Color-Clusterings?



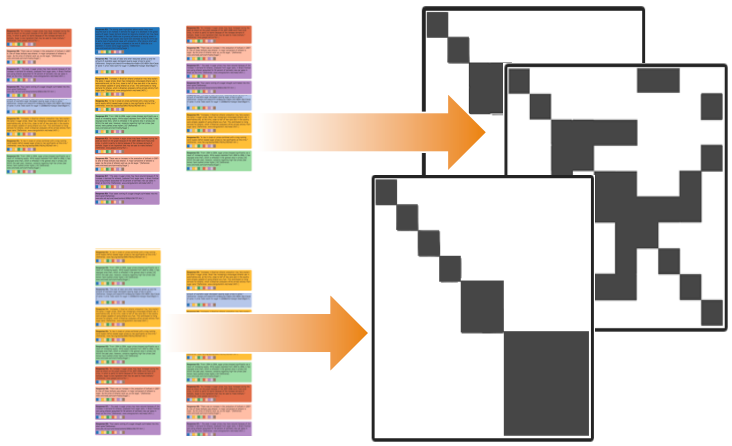
Selecting the **Most-Representative Clustering**

How to Integrate Color-Clusterings?

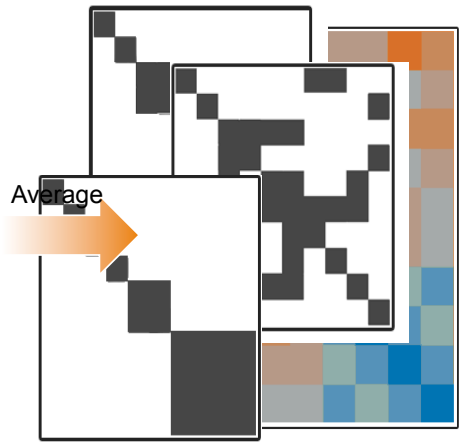
Selecting the Most-Representative Clustering



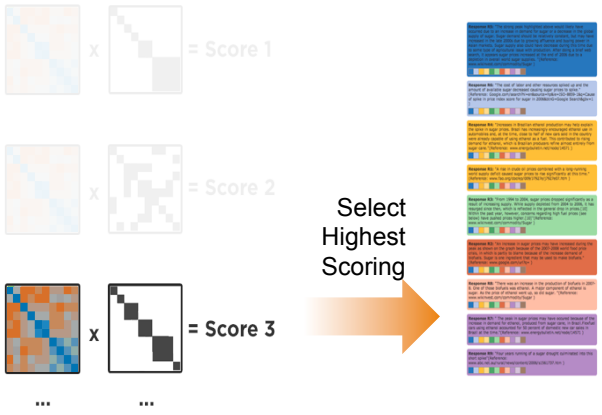
Selecting the Most-Representative Clustering



Selecting the Most-Representative Clustering

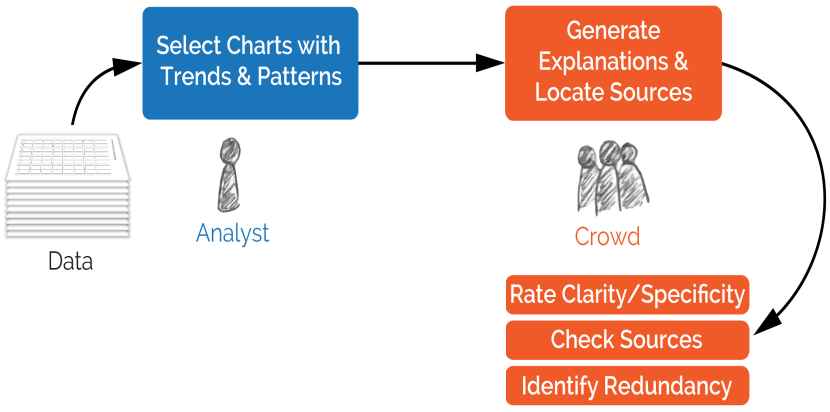


Selecting the Most-Representative Clustering

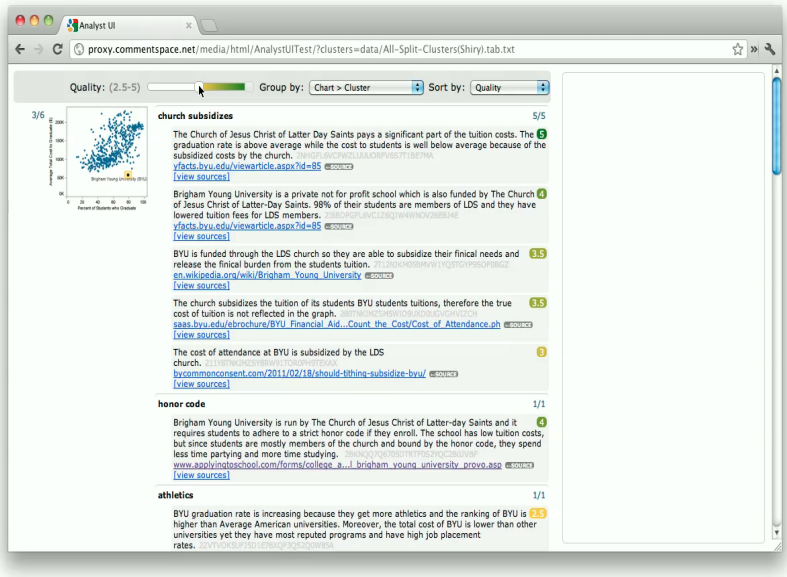


Managing the Crowd's Work

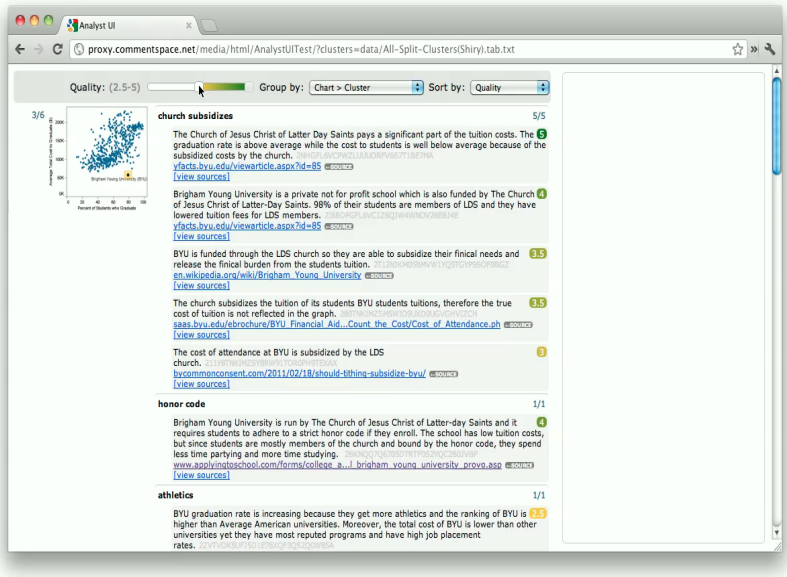
Managing the Crowd's Work



Explanation Management Interface



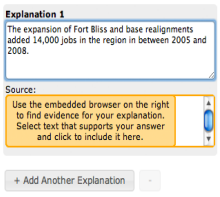
Explanation Management Interface



Open Questions

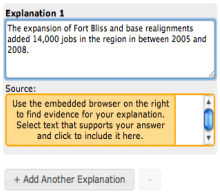
Open Questions

Segmenting Explanations

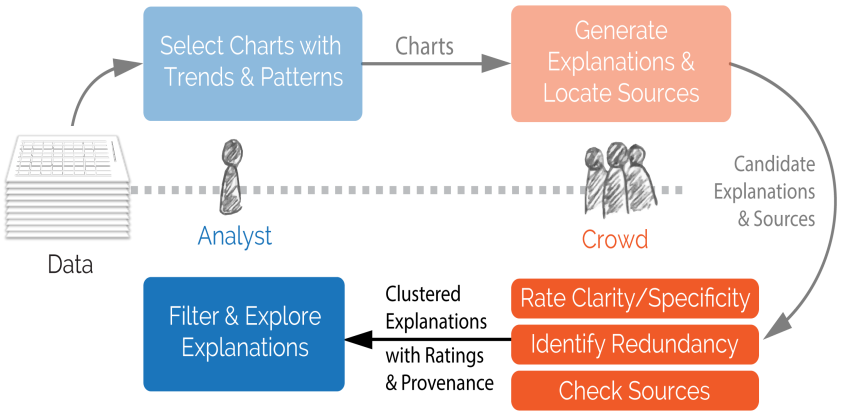


Open Questions

Segmenting Explanations



Crowd Composition

cheap low-skill crowds
vs.
more knowledgeable trusted ones

Conclusion