**Title suggestions:**

**1) What are popular policy sections of open data policy amongst residents?**

**2) What kind of open data policy do the residents want?**

**3) Evidence-based approach to finding popular sections of open data policy**

**4) Governance, utility, and privacy:why are these open data policy sections popular?**

**5) Comments on open data policies focus on governance*,* utilityand privacy**

*Audience: users of open data*

[Co-authored by Greg and Faraz]

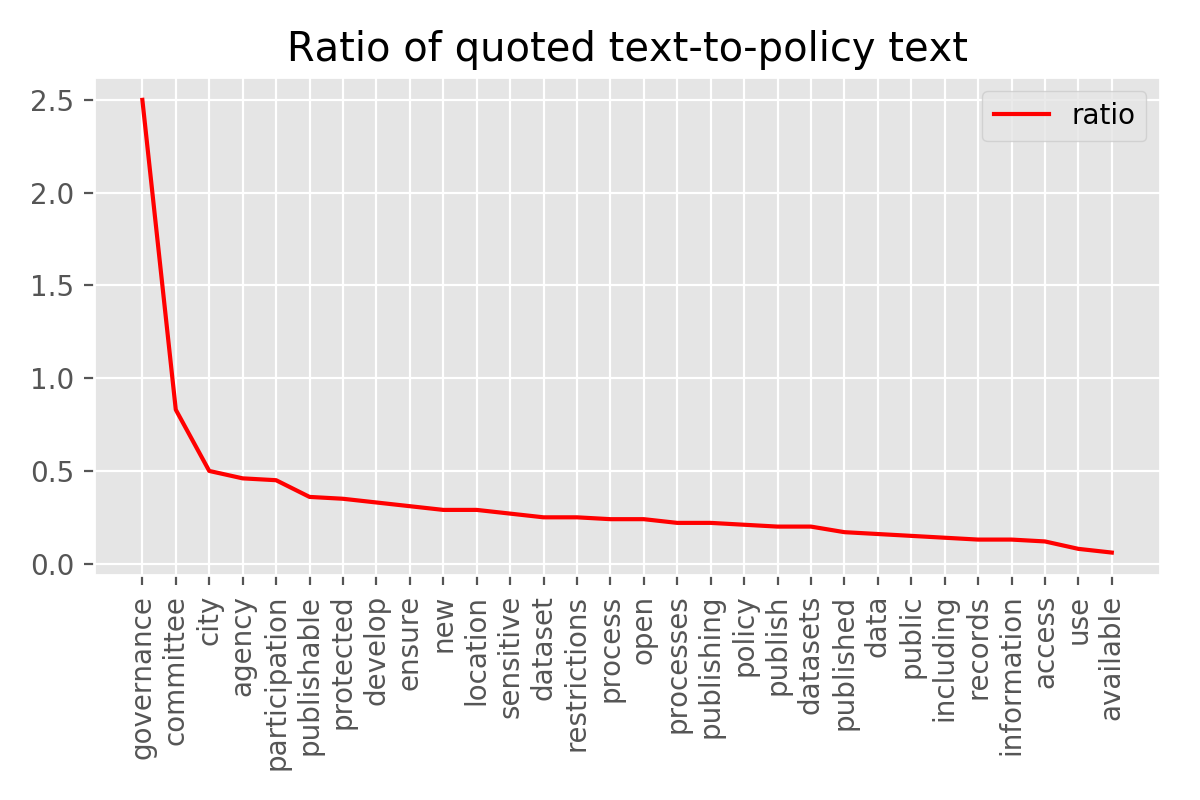
Numerous cities have posted draft open data policies online for public comment. But what issues do commenters usually raise? They tend to focus on topics like governance committees, privacy and protected/sensitive data, participation, frequency of data updates, and restrictions on use. Cities would do well to pay a lot of attention to these hot topics.

These findings come from our examination of comments on open data policies that cities have posted on the [OpenGov Foundation](https://opengovfoundation.org/)’s [Madison platform](https://mymadison.io/). Look out for two more posts in this series showcasing our research.

The Madison platform allows users to quote a specific section of a policy. One way of seeing what topics are popular is to pool together all the words in quoted policy text and sort the words by how often they appear. The his will result in a high ranking for words like “open” and “data” that appear many times in policies.

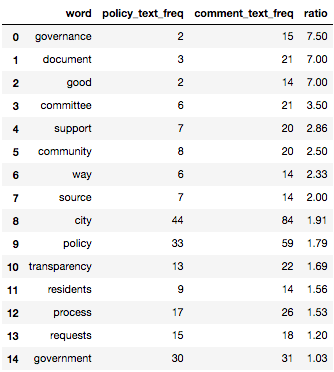
To get a better sense of what topics people were especially interested in discussing, we compared the frequency of quoted words to the frequency of all words that appear in policies. This lets us see what words are disproportionately quoted. If we remove [basic words](https://en.wikipedia.org/wiki/Stop_words) like “is” and “the,” we find the following words were used the most often in quotes compared to all policy text:

In a more graphical form, the data looks like this:

[Should we say any more about the implications of this?]

What about the actual substance of people’s comments, as opposed to the policy sections they’re quoting?

We compared the frequency of words in comments compared the frequency of words in policies and found these were the most disproportionately used:



Here we can see many similar themes to the last analysis (especially regarding interest in governance), but with the addition of words indicating commenters were expressing their support for open data policies — revealed by words like “good” and “support.”

This is not a perfect methodology — if there even is such a thing — but it provides a lot of useful insight about what topics generate the most discussion.

Our next post will look at what types of people comment on open data policies, followed by a post examining the sentiment of comment text.

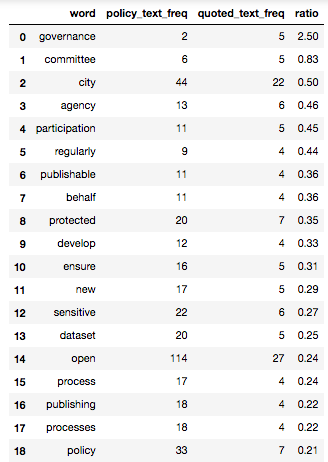
*Old Faraz text (could include additional parts of this):*

[Crowdlaw](https://sunlightfoundation.com/2016/03/11/crowdlaw-and-open-data-policy-a-perfect-match/) for open data is an online collaborative policy making approach between residents and governments. It allows residents to provide feedback on an open data policy, with the hope that the feedback will be considered to become part of the policy.

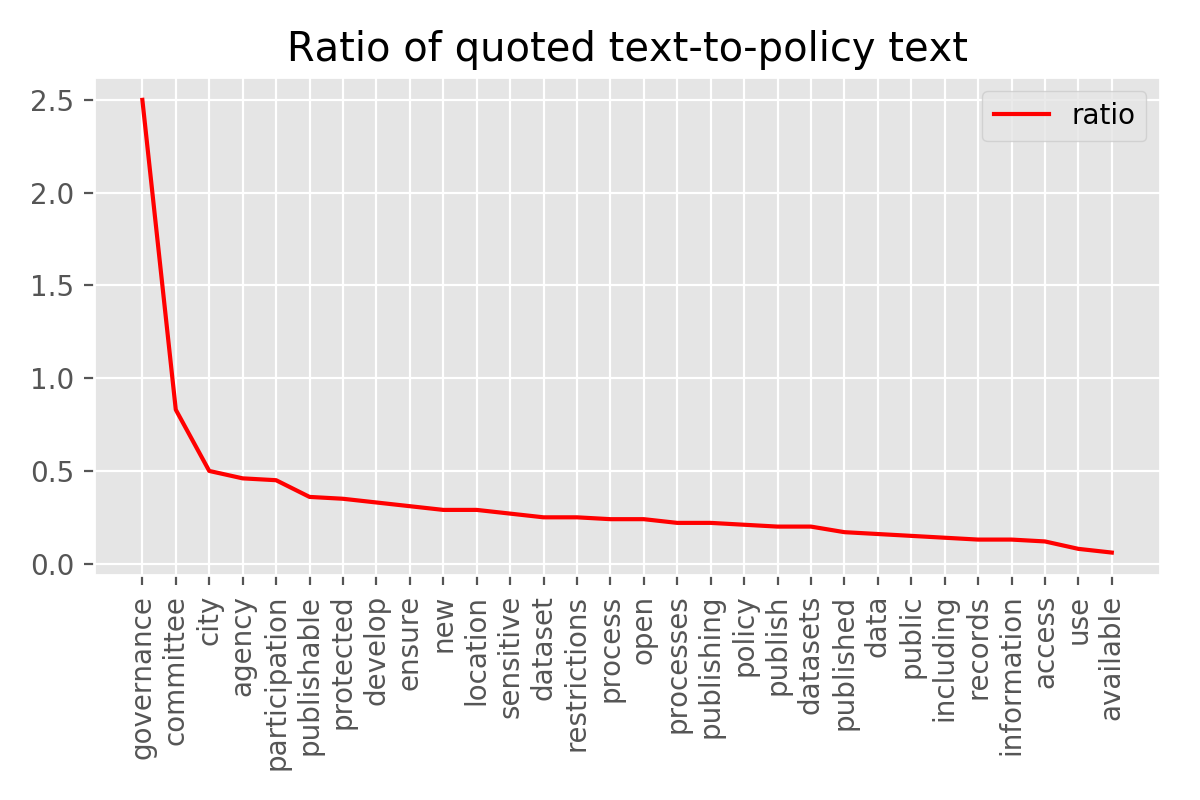
Sunlight has prepared a memo that highlights takeaways from crowdlaw such as the popularity of policy sections, intensity of “emotional” response to policy, and backgrounds of popular authors’. The cities that are part of this analysis are Buffalo, Durham, Tyler, Glendale, Nashville, Syracuse, Naperville, the BART system and San Francisco.

This blog post is the first in our series of posts that discuss in depth about each of the crowdlaw takeaways. In this post, we’ll shed light on most and least popular policy sections of open data policies via a metric called *quoted text frequency to policy text frequency* ratio.

***Most popular policy sections***



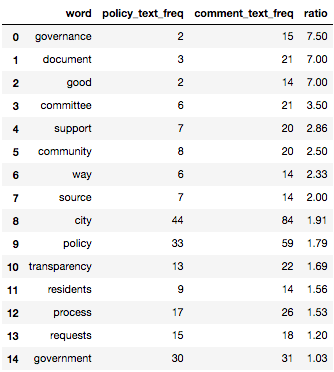
The *quoted text frequency to policy text frequency* ratio shows the number of times a word in a policy section has been *quoted* by a resident. The higher the ratio, the higher the popularity of that policy section. That is, a high ratio for a policy section demonstrates that residents engage more seriously with that section than the government that has drafted the open data policy. It highlights an opportunity for the government to collect feedback for that section and, wherever necessary, incorporate feedback in policy.

This ratio is also useful for the reason that it filters *noise* in the form of words such as *“data”* and *“open”* that otherwise have higher word frequency policy texts. Reducing noise ensures that a more accurate picture is presented of where and what kind of feedback residents provide.

The table (above) shows the ratio for all cities, sorted in descending order. Policy sections that are most talked about include words such as *“governance”* and *“committee”*. People who quoted this section commented on issues regarding how the Open Data Governance Committee will be formed, who’ll act as the liaison between the Committee and residents, and the need to make minutes of the Committee meetings available.

Words such as *“participation”, “regularly”, “protected”,* and *“sensitive”* have also been quoted frequently. This shows that residents want to have more opportunities to give feedback, a higher utility for interacting with timely and updated datasets such as by making datasets more visual-friendly, and mechanisms in place to protect residents’ private data.

***What policy sections do popular “comments” refer to?***



The ratio in the table on left compares the number of times a word in a comment made by a resident matches against the number of times that word is in a policy section. Again, words such as *“governance”* and *“committee”* appear at the top, confirming that this is the most discussed policy issue. A lot of people have also talked about *“transparency”*. They seem to have talk about this because the policy section on how the government can provide more transparency, for example, by making more government data sets available, is still not a fully developed section.

City governments can use this information to decide which policy sections are more important, what residents are talking about in those sections, and how can they [meaningfully](https://sunlightfoundation.com/2017/03/10/how-to-invite-feedback-on-an-open-data-policy/) engage with residents. Concurrently, governments should also make an effort to understand the residents’ or users of open data policy better, which is the focus of our [next blog post](https://docs.google.com/document/d/1ZFdfXRwFT0HQpbDrucfqHN3xBNmBbpVHdHY2cSKgvLU/edit).