

Making SA Pollution Tangible

This project aims to find artistic means of making pollution in South Africa understood .

The code to all graphics produced can be found in this [Colab notebook](#), and the code to the midi files used in the musical output can be found in this [Colab notebook](#).

This is in the spirit of making data analysis more transparent.

Explanation of Data Set

We made use of 2 datasets to inspire our output.

StatsSA Annual General Household Survey

The first source of data came from [StatsSA's annual general household survey](#). The data we considered came out of the question, “ *Which of the following environmental problems do you experience in your community/on your, and neighbouring farms?*”

The statistic considered, *% of population that experiences air, litter, noise and water pollution*, reflects the % of households in the survey that answered “yes” to experiencing that form of pollution. It was left up to the participant to interpret what experiencing pollution means.

Google Trends Data

The second source of data came from [Google Trends](#). Using the Python module, [pytrends](#), an unofficial API (Application Programming Interface - in this case a means of obtaining data from the website) for the Google Trends site, the interest over time for *pollution-related* keywords was obtained.

The interest of a keyword is related to the number of times it was searched using Google. The interest is scaled relative to the highest level of interest (the ‘peak’) over the considered period of time. This peak is assigned a value of 100. For future work, it would be interesting to see whether we can translate this scale of 0 to 100 to an accurate estimation of the number of searches - for instance by including a currently ‘trending’ keyword where actual numbers are available.

Visualisations of StatsSA dataset

Location

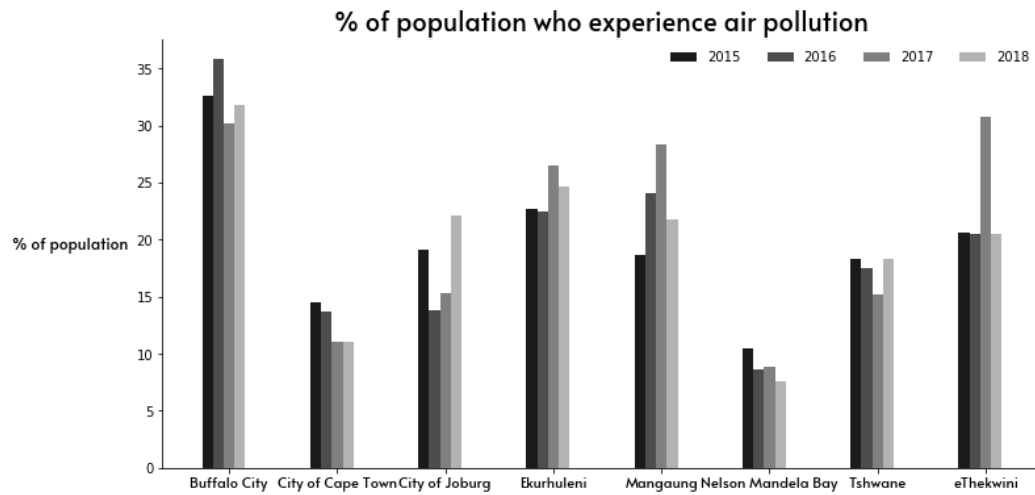
We start by showing the locations of the metropolitans considered in the dataset.



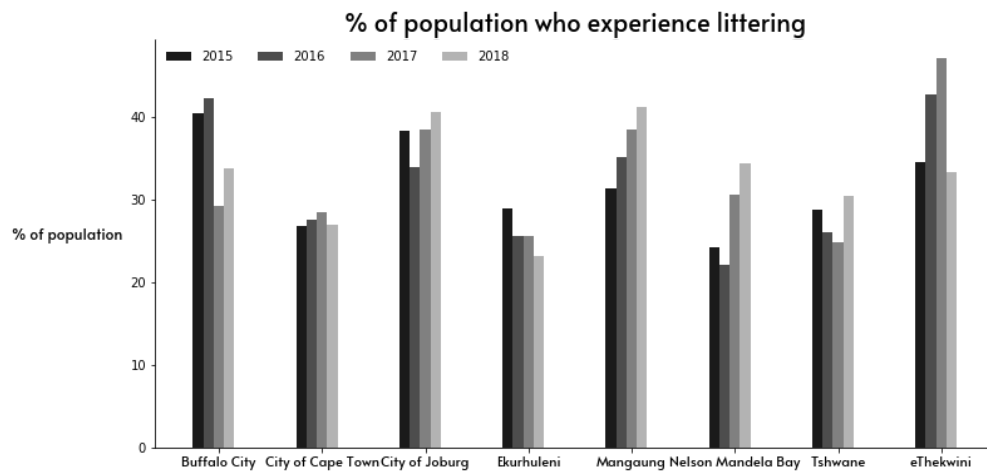
Bar-plots of % of respondents who experience different types of pollution

Below, we visualise the *percentage of the households surveyed (% population)* in each of these metropolitans who experience *air, littering, noise* and *water pollution* over the period of 2015-2018. We plot individual bar-plots for each metropolitan, and look at changes over the 4 years considered.

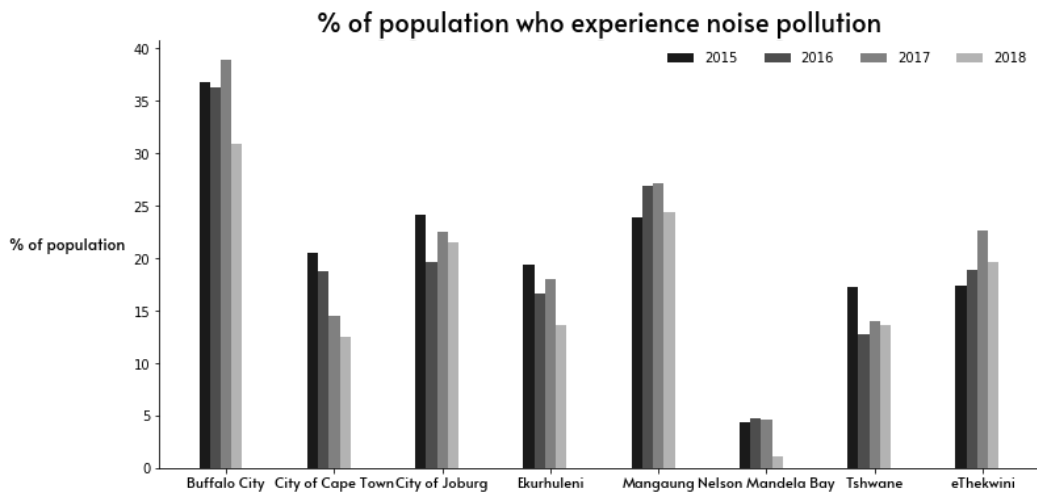
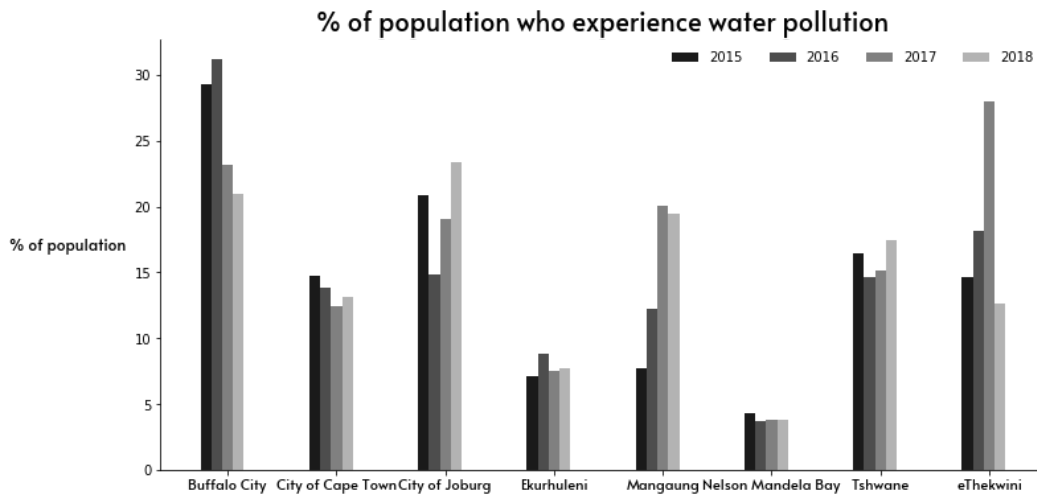
Our main query here is to what extent does the % of respondents who say they do (or do not) experience pollution actually experience it. Our feeling was this is more a reflection of perceptions of pollution than levels of pollution.



Buffalo city seems to have the highest percentage of the population who say they experience air pollution. On the other hand Nelson Mandela Bay seems to have the lowest. There does not seem to be any trends in the perceptions of pollution over time.



In this, all of the metropolitans seem to have 'more similar' percentages of the population who say they experience littering (in comparison to the more varying rates of other types of pollution). Notice the steady increase in the experience of littering in Mangaung over the time period. Does this reflect a real increase in pollution, or is this just random variation? So far, these bar-charts seem to have more evidence for random variation.



It would be interesting to experience noise pollution in Buffalo City further. Perhaps we could compare the most polluted metropolitan to the least polluted metropolitan. This exploration could be through collecting resources (images, personal accounts, news clippings?) as well as through visualising data.

Virtual Perspective

Using the same time period (2015-2018), we track interest in the search terms, “*air pollution*”, “*noise pollution*”, “*litter*” and “*water pollution*”. This is purely a reflection of the amount of searches on Google related to these terms, not the reality of pollution. Although a breakdown for the municipalities was unavailable, we were able to consider interest for each province, which have some link to metropolitans.

Relative interest in pollution as defined by Google trends, over
2015-2018

	<i>air pollution</i>	<i>noise pollution</i>	<i>litter</i>	<i>water pollution</i>
<i>Eastern Cape</i>	29	0	28	43
<i>Free State</i>	34	0	24	42
<i>Gauteng</i>	33	5	23	39
<i>KwaZulu-Natal</i>	33	4	23	40
<i>Limpopo</i>	31	6	15	48
<i>Mpumalanga</i>	26	0	32	42
<i>North West</i>	41	0	21	38
<i>Northern Cape</i>	33	0	27	40
<i>Western Cape</i>	25	7	36	32

There seems to be relatively little interest in noise pollution based on Google searches compared to the other types of pollution. You should notice that each row adds up to 100, so it is a measure of relative interest - how much were people in, say Limpopo, interested in *water pollution* compared to *litter*.

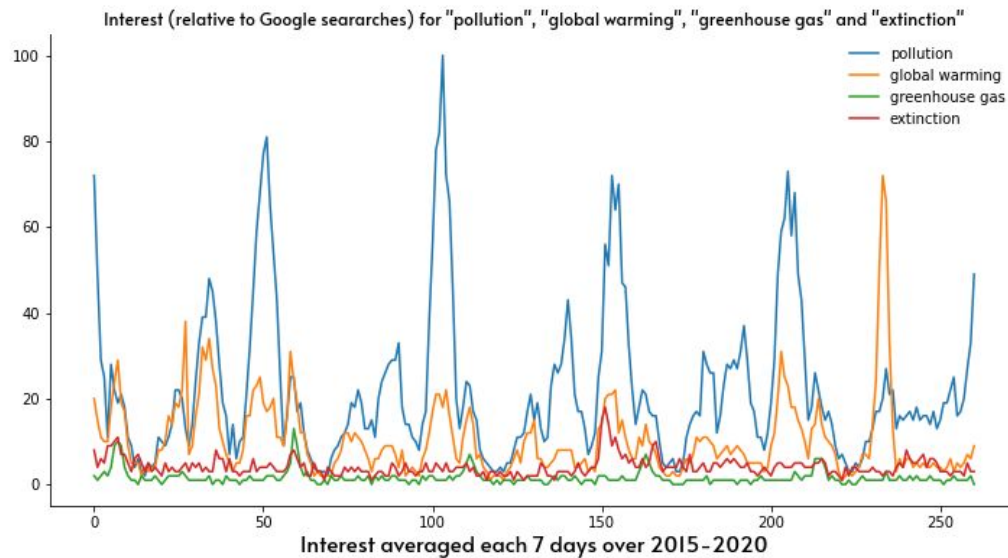
Artistic Outputs

Mapping Pollution Search Trends to Midi (and then Music)

As a more abstract representation of data, we considered a way of mapping the interest in keywords associated with pollution, “pollution”, “global warming”, “greenhouse gas” and “extinction”, to musical notes (no extensive thought went into choosing these keywords).

For each keyword, interest, defined relative to the number of searches on Google in South Africa, was tracked over the past 5 years.

A time-series plot of the interest data is presented below:



An interesting thing to observe is how peaks in interest in pollution coincide with peaks in interest in global warming. It may be worth trying to find out which events were associated with these peaks.

A fairly predefined structure was defined for mapping the interest levels (ranging from 0-100) to note values (an A minor scale over 4 octaves). We will briefly explain this here, but for exact implementation details, see the [Colab notebook](#).

For the interest data associated with “pollution”, which had the most extreme variations, the interest value was scaled from [0-100] to [0-6], with each number corresponding to a note in [A, B, C, D, E, F, G].

For the interest data associated with the other three keywords, these were mapped to the chords A minor, C major and E minor (cycling between them in this order). This mapping was done using the modulo operator, which returns the remainder of a value after division. In this context, the interest value was divided by 3, and the remainder determined whether it would be assigned the root, the third or the fifth of the note.

[Watch the video/music output here](#)

Installation Concept

- mark municipality's in chalk, sized to scale on the floor
- translate the percentage of people experiencing litter in each province to cubic measurement relative to the surface area of each province.
- scatter the litter for each municipality in its circle
- Viewers are invited to walk through the municipalities

Extensions of the above:

- Have sets of municipalities, 1 for each year, with the relative percentages of litter experienced, so that viewers can experience the change in litter experiences over years.
- Representations of relative sound pollution experiences could be produced as audio tracks in each 'municipality year'.
- Bearing in mind that the conceptual experience for viewers would be different if the litter/province ratio used data on actual litter measurements (hard to gather) and a critical framing for the participatory installation here would be that viewers are walking through a representation of the numbers of people *experiencing* litter, rather than the litter itself. This is a bricolage to offer an experiential engagement with how big the communal sense of litter is in each municipality over time.

For presentation on 4 Sept: stop animation of sketch of South Africa with growing litter.