Motivation (incl. Literature Review)

“Do online searches on topics that raise privacy concerns decrease after exogenous online state surveillance shocks?” More specifically, we wish to explore the relationship between the NSA scandal in two countries (USA, Canada and UK) and changes in Google Trends data on certain privacy-sensitive keywords before and after June 2013.

A fellow TU Graz student from a non-EU country witnessed police misconduct at Vienna airport last year when three officers failed to provide assistance to a collapsed passenger. Despite Marko's* urging and a delayed ambulance arrival, the passenger died from a sudden cardiac arrest.[1] Shaken by the event, Marko started drafting a twitter thread but decided it would be ‘smarter’ to not publish it due to his pending job-seeker visa status in Austria.

This small incident exemplifies an everyday occurance of so-called “chilling effects”. In a legal context, a chilling effect is the inhibition or discouragement of the legitimate exercise of natural and legal rights by the threat of legal sanction.[2] This example is particularly fitting because a) tweeting about gross police misconduct is not only legal, but even desirable behavior in a deliberative democracy; b) the seemingly mundane behavior of a bystander tweeting is juxtaposed by a disproportionately grave consequence (not getting a visa renewed) and, lastly, c) the “chilling effect” is based on a vague assumption and unclear link between tweeting and the consequences, which is likely born from a diffused feeling of being somehow monitored.

The "chilling effects" concept originally arose in the 1970s due to concerns over legal prosecution.[3] After the 9/11 attacks, the concept became relevant again with regards to anti-terrorism surveillance.[4] However, in the current age of digital communication and datafication, the issue has intensified in relation to expressing opinions and searching for information online. The term "dataveillance" has been proposed to describe this phenomenon.[5] While historically, the debate has been mostly conceptual, Jonathan Penney's 2017 study was the first empirical analysis linking government online surveillance to chilling effects on online user activities.[6] This finding contradicts the "privacy paradox" which suggests users are willing to trade privacy for convenience.[7] Penney's study was further supported by a similar study by Marthews and Tucker on Google Search behavior.[8]

Just as it was recognised half a century ago, we were motivated to empirically look into a phenomenon that inhibits people from exercising their fundamental rights and “constitutes a subtle, cumulative risk for individual autonomy, well-being, and democratic participation in digital societies”.[9]

Data retrieval:

- Data was gathered using the pytrends.request library in Python from Google Trends for two categories: keywords related to terrorism (48 keywords) and keywords related to domestic security (25 keywords).

Authors: Elias Gatternig, Imelda Lizeth Chavez Martinez, Suraj Yathinatti, Adam Urosevic

- The domestic security category was chosen as a quasi-control group to compare against the terrorism category.[1]
- To create a (fully) unrelated control group, a list of 48 random keywords was generated using the random-word library in Python (they can be found in our GitHub repository).
- For the main analysis, the same keywords were used as in the paper of Penney (2016).
- Data was collected from Canada, USA, and Great Britain in weekly intervals from 01.01.2012 - 31.12.2014 for a total of 157 weeks
- For each country, three csv files were created to store data for the terrorism keywords, domestic security keywords, and random keywords.
- Keywords were sourced from the "Department of Homeland Security National Operations Center (NOC) Media Monitoring Capability Desktop Reference Binder 2011". The list of search terms formed a part of the basis for their monitoring activities.
- As we struggled to extract some words from the NOC document, the Python module pytesseract was used to extract information from an image in the original keyword source, with some keywords cleaned by hand.
- After some trial & error, we decided to use the more selective and relevant list of keywords from the original paper (instead of the extensive list of terms found in the NOC document).

Data processing:

Since the keywords from Penney (2016) were embedded in a file with data containing search volume of Wikipedia articles, the keywords had to be selected in a way that there are no duplicates in the end. This was done using the *pandas* library of Python.

We tried different approaches to obtain the data from Google Trends. To have data which is more comparable, our first approach was to get the batches of five keywords, where each one had the same reference keyword in order to normalize all the data on one keyword. However, we ultimately decided to discard this method since Google data is relative/normalized anyways and our initial approach caused issues with the Google API (“TooManyRequests”). Instead, data was obtained word for word. Once we saved the csv files, we:

- A pandas dataframe was created, and search volume per calendar week was summed up for all words of the given category (e.g. all search volumes of the first week of the timeframe of the 43 terrorism keywords were summed up).
- The timeframe was split into two sub-timeframes (week 0 to week 76 and week 77 to 156) based on the exogenous event of the Edward Snowden revelations published on June 6, 2013.
- This method of splitting the timeframe is called interrupted time series (ITS) and allows for the assessment of the impact of an intervening event by comparing patterns in the time series data before and after the event.

[1] This category was chosen, because people interested in terrorism searches might also be attracted to searches relating to domestic security. In theory searching for search terms related to terrorism is far more likely to raise privacy concerns for users concerned about government surveillance than viewing information about domestic security agencies (Penney, 2016, p.156).

Analysis:

To visualize the chilling effect, we plotted the data of both time series for the three countries: Canada, Great Britain and the United States. In these plots, we added a linear regression for each sub-timeframe, for each country. Furthermore, we performed a paired t-test for each country, comparing the average search volume per week before and after the exogenous event, with this, we obtained a p-value, which tells us if the increase or decrease of this average is significant or not.

If the exogenous event had a chilling effect, this would be visualized in the plots in the form of a smaller search volume after such an event, compared to before. Additionally, a short-term chilling effect can be inferred if the linear regression after the event presents a drop compared to the first linear regression. Furthermore, we can infer a long term chilling effect if the line stays flat or points down after several weeks. A quantitative way for finding a long-term chilling effect is having a p-value smaller than 0.05 for the paired t-test, and a drop in the average search volume. This would indicate that after the event, the search volume decreased, and that this decrease is not by chance, but caused by an event.

We performed these statistical analysis for the terrorism-related keywords, expecting to see all the results mentioned above. We also did this for domestic-violence keywords, which, whilst also related to security and monitored by the NOC, are not words that would cause a chilling effect, as they do not represent private sensitive searches (assumption based on Penney, 2016, p.156). For this reason, the “domestic” category is our quasi-control group, and we expected not to see any of the behavior previously described for any of the statistics.

Results:

In Figure 1, we can observe the results of these analyses for our data of interest. We can clearly see an immediate drop in the linear regression for the three countries, indicating a short term chilling effect. We can also see that the average search volume of Canada is significantly lower after the event, however, as this is not true for the other two countries, and the linear regression shows an increase in the search volume after a few weeks, we can not show a long-term chilling effect.

For comparison, in Figure 2, we plotted the analysis of the domestic-related keywords, and we can see that there is neither a clear drop in the search volume immediately after the event, nor weeks later. The average search volume significantly decreases for the United States, however, this is again not supported for the other two countries, so we can conclude that, as expected, there is neither a short-term nor a long-term chilling effect for domestic security searches.

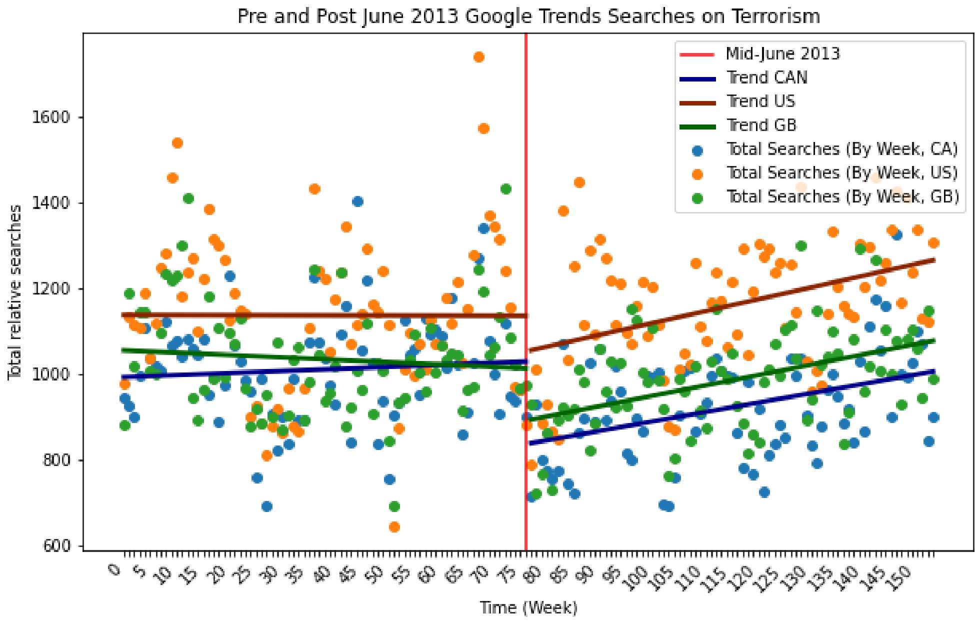

Fig 1. Statistical analysis for terrorism-related keywords, before/after exogenous event

```
Search volume per week:

CAN before: 1010.987012987013, after: 922.6753246753246
p-value CAN: 3.125161400129812e-05
US before: 1136.7792207792209, after: 1160.3766233766235
p-value US: 0.30036797339424715
GB before 1034.3766233766235, after: 985.7532467532468
p-value GB: 0.026616958429662396
```

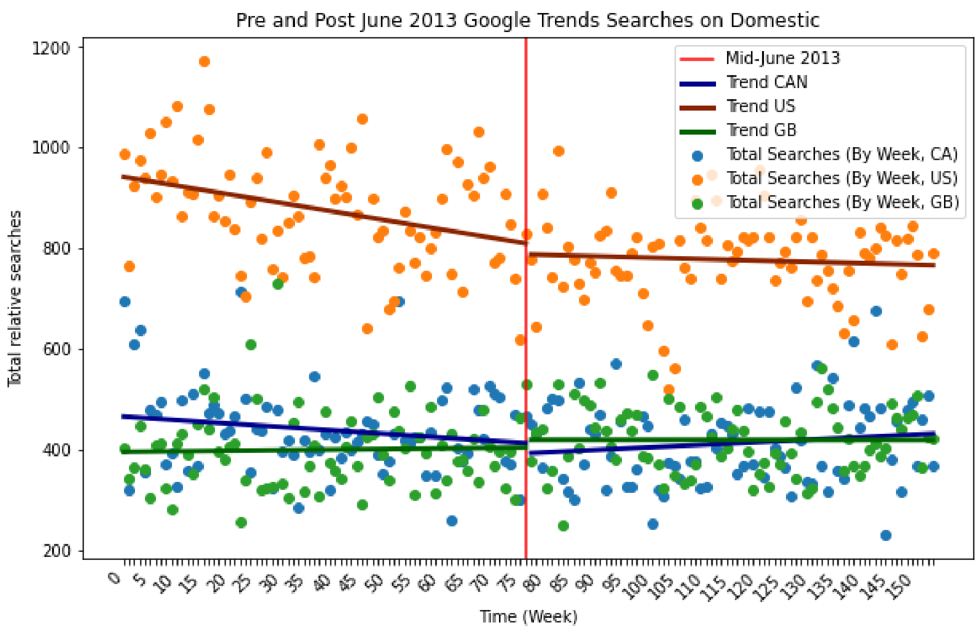

Figure 2. Statistical analysis for domestic-security keywords, before/after exogenous event

```
Search volume per week:

CAN before: 439.31168831168833, after: 412.4155844155844
p-value CAN: 0.050590663104865825
US before: 875.9090909090909, after: 777.1558441558442
p-value US: 1.5106014363483156e-09
GB before 399.83116883116884, after: 419.38961038961037
p-value GB: 0.11056628454893633
```

The results showed a short-term chilling effect. To quantify it, we calculated the average search volume before and after the event, but for a short-term period, of only 12 weeks, the results are shown below:

Canada drop: -194.33333333333337, p-value 0.011198847841780005
US drop: -177.91666666666674, p-value: 0.16981654264562848
GB drop: -218.5000000000001, p_value: 0.0010722347305603156

As we can observe, in the short term, the search volume drops in all three countries. For Canada and Great Britain, the drop is considered significant.

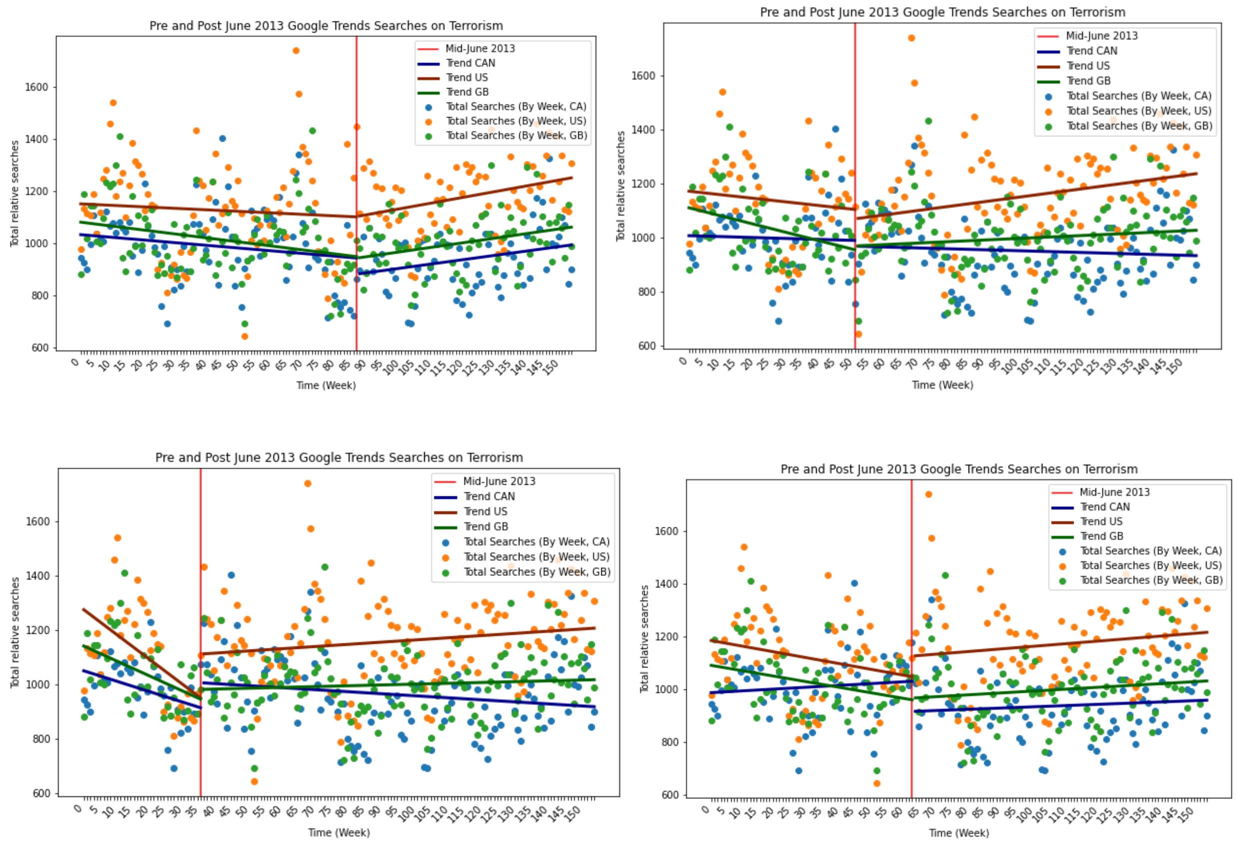

Figure 3. Random permutations of the week of the event

To visualize this, we performed several permutations on the “week of the event”, to see if different interruptions in the time series could also lead to a drop in the search volume. We plotted four samples from these permutations, as shown in Figure 3. In these graphs we can see small decreases in the search volume but also small increases, however, there aren’t big drops as in the plots using the real week of the event, which aligns with our hypothesis. To further test the relation of the event with a short-term chilling effect, we ran 1000 permutations and obtained a histogram showing the value of the difference between the average search volume before and after a random week, in the short term (12 weeks).

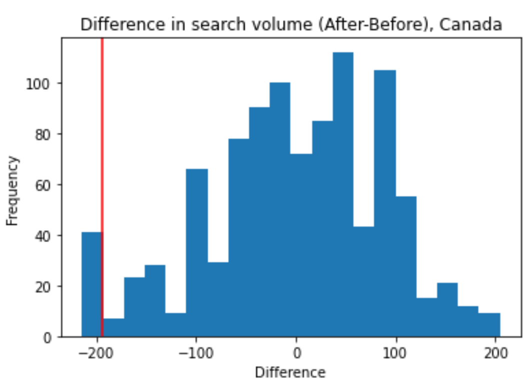

Fig 4. Histogram CAN

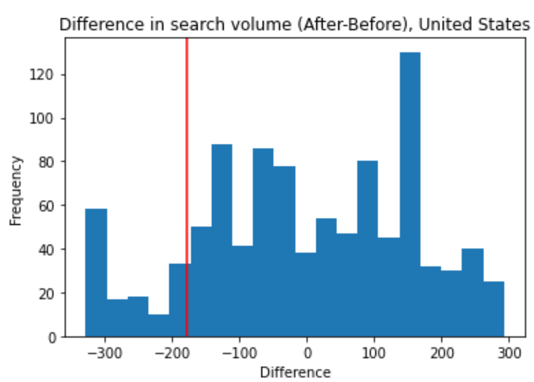

Fig 5. Histogram US

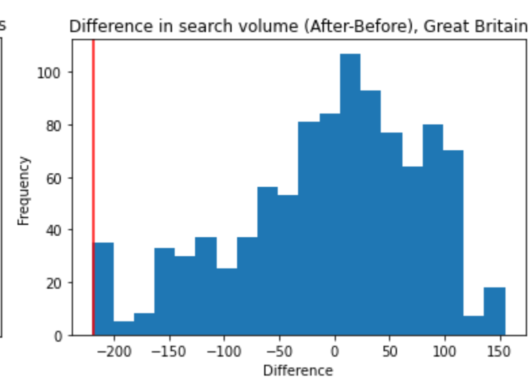

Fig 6. Histogram GB

In Figures 4, 5 and 6, we can observe that the drop in the search volume could also occur by dividing the time series in other random weeks. However, we believe that the drop is not as common (not the most observed case), and we also think that using the difference in the average of 12 weeks before and after the event might not be the optimal metric for the histogram. For these reasons, we still consider that the real event had a short-term chilling effect on the search of terrorism-related keywords.

Conclusion & Critique

We achieve to replicate the main results from the paper, namely a short-term chilling effect, as shown by a statistically significant short-term drop in the search of terrorism-related keywords in the three countries. By taking longer time series into account than the original paper, we additionally aimed to see whether there is a longer-term chilling effect by the same exogenous event. However, we were unable to statistically confirm longer-term effects. Upon further thought, we think that this is unsurprising: usually, a severe self-inhibition after an external shock (characterized by uncertainty in terms of one's affectedness) will quickly subside and a "back to normal" dynamic kicks in. [10] Looking forward, robust validation of chilling effects of dataveillance will require much more research, notably to close empirical and theoretical gaps regarding the scope, prevalence, process and governance of surveillance-induced chilling effects.

Bibliography

Key reference paper: Penney, J. W. (2016). Chilling Effects: Online Surveillance and Wikipedia Use. Berkeley Law. https://lawcat.berkeley.edu/record/1127413

1. The student's name was changed for privacy purposes.
2. http://law.yourdictionary.com/chilling-effect
3. Schauer F (1978) Fear, risk and the first amendment: Unraveling the chilling effect. Boston University Law Review 58: 685–732.
4. Solove DJ (2006) A taxonomy of privacy. University of Pennsylvania Law Review 154(3)
5. Penney JW (2017) Internet surveillance, regulation, and chilling effects online: A comparative case study. Internet Policy Review 6(2): 1–39.
6. Büchi M. et al.(2020) The chilling effects of algorithmic profiling: Mapping the issues, Computer Law & Security Review 36(2020), https://doi.org/10.1016/j.clsr.2019.105367
7. Barth S, De Jong MD. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. Telemat Inf 2017;34(7):1038–58.
8. Marthews, A., & Tucker, C. (2017). The impact of online surveillance on behavior. In D. 5869 & S. Henderson (Eds.), The 3497 Handbook of Surveillance Law (3497 Law Handbooks) 112, 437-454
9. Véliz C (2020) Privacy Is Power: Why and How You Should Take Back Control of Your Data. London, UK: Bantam Press.
10. Preibusch S (2015) Privacy behaviors after Snowden. Communications of the ACM 58(5): 48–55