Calgary Municipal Elections Data Report

Tin Le

Dr. Reardon

CPSC502.04

Updated 15/01/2022

Contents

Contents 2
General Notes 3
Mayor candidates 4
Number of HTTP Requests 4
Tracking/non-tracking external sites 5
Notes/Observations: 5
Tracking domains found across all browsers/candidates (Chrome, Firefox) 6
POST requests 7
Councillor candidates 8
Number of HTTP Requests 8
Tracking/non-tracking external sites 9
Tracking domains found across all browsers/candidates (Chrome, Firefox) 10
POST requests 11
Public school board trustee candidates 12
Number of HTTP Requests 12
Tracking/non-tracking external sites 13
Tracking domains found across all browsers/candidates (Chrome, Firefox) 14
POST requests 15
Separate school board trustee candidates 16
Number of HTTP Requests 16
Tracking/non-tracking external sites 17
Tracking domains found across all browsers/candidates (Chrome, Firefox) 18
POST requests 19
Overall observations and next steps 20

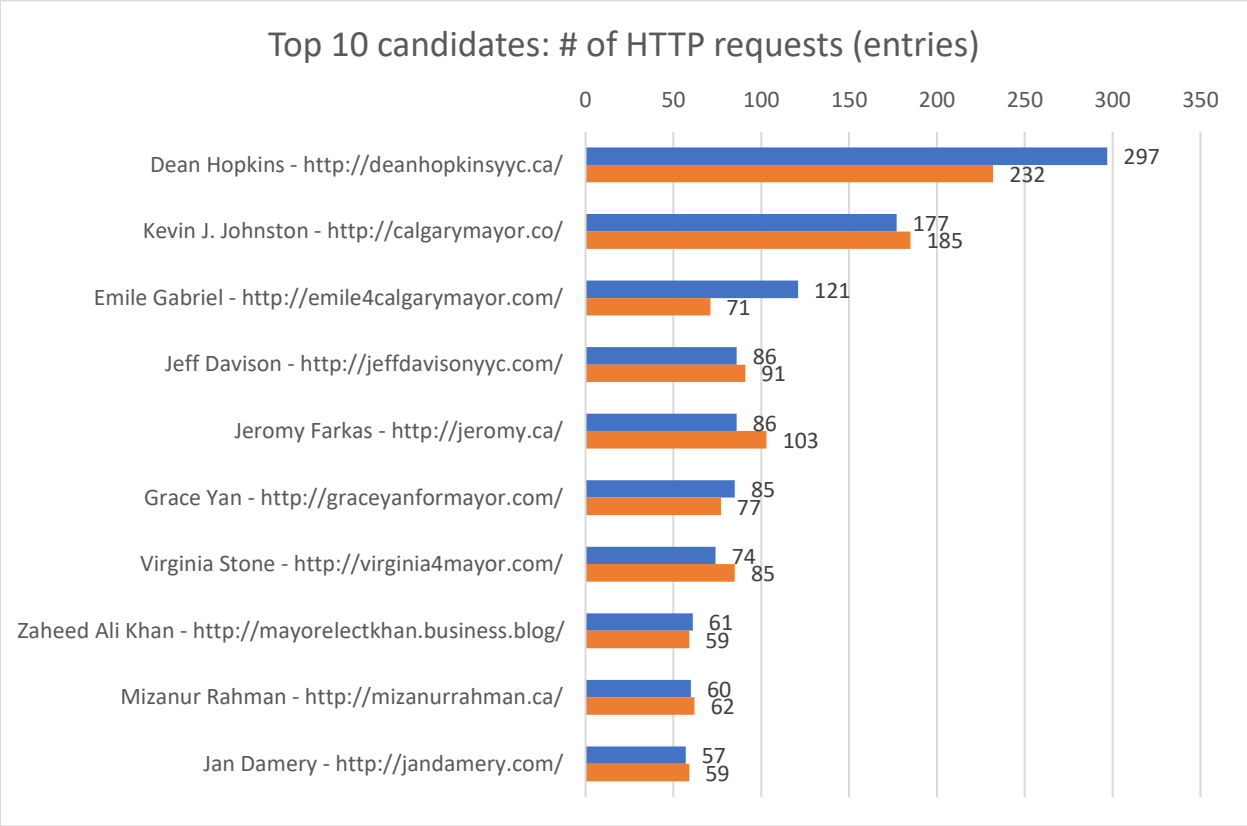
General Notes

- 4 positions analyzed: For each position, all candidates with campaign websites had their websites analyzed (list of sites can be found in RESULTS folder in GitHub)
    - *Mayor* – 24 candidates analyzed
    - *Councillor* – 91 candidates analyzed
    - *Public school board trustee* – 29 candidates analyzed
    - *Separate school board trustee* – 15 candidates analyzed
- Chrome/Firefox HAR files for Mayor and Councillor positions were collected on 17/10/2021, Chrome/Firefox HAR files for Public school board trustee and Separate school board trustee positions were collected on 18/10/2021.
- Relevant Python scripts created for this part of the project:
    - **getJSONsBUP.py**: Python script that automates a clean slate Firefox browser and visits a set of websites + downloads HAR files in JSON format for each website. Used for collecting data on federal elections. Referred to as the "Firefox scraper" in this report.
    - **getJSONsChrome.py**: Same as the above script but automates a clean slate Chrome browser instead. Was not developed in time for the federal elections, and therefore wasn't used in this portion of the project.
    - **analyzeJSON.py**: Data analysis program. Used to sift through collected HAR files in JSON format and returns important information in .txt files. Also has been recently updated to:
        - detect and mark requests to known tracker sites (based on https://github.com/duckduckgo/tracker-radar/tree/main/domains)
        - detect and mark potential tracking pixels
        - generate .txt files in csv format that contains important metadata for each HAR file analyzed. The data in these csv files is then used to create spreadsheets.
    - **generateDomainsList.py**: Generates a .txt file containing all the tracker domains in the DuckDuckGo database of trackers. The resulting .txt is used by **analyzeJSON.py** to efficiently detect and mark trackers.
- **Important Issues**
    - HAR files generated by scraper program seem to not include any tracking pixel entries (no tracking pixels detected by HAR analysis program). Still trying to figure out why this is the case.
        - Because of this, **there is no tracking pixel section in this report**. Tracking pixels were only detected in a few manually collected HAR files, which was done when there was an issue with scraper collection.

Mayor candidates

- 24 candidates/websites analyzed

Number of HTTP Requests

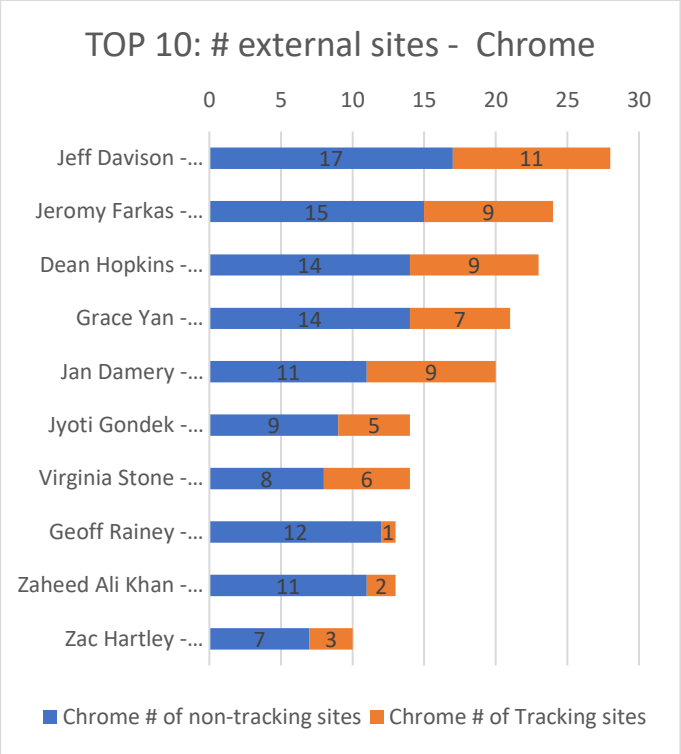


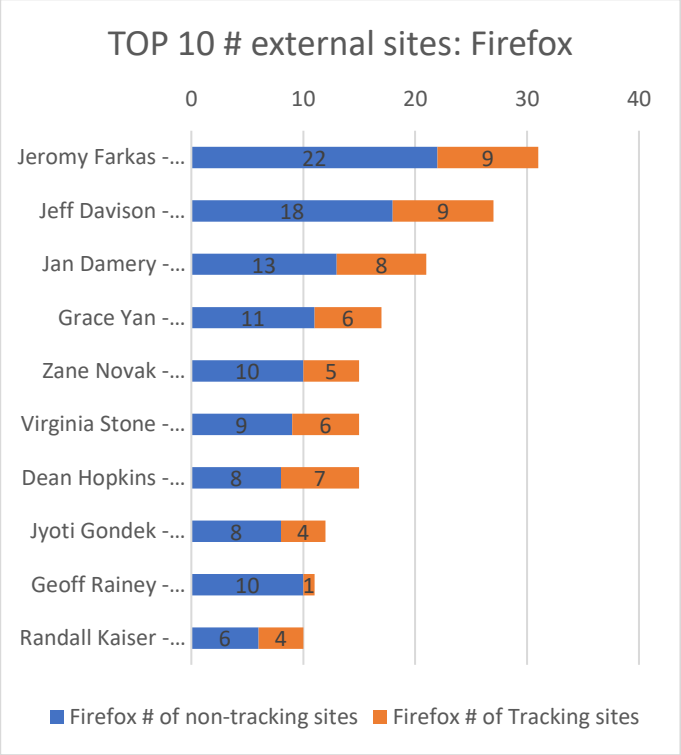
Tracking/non-tracking external sites

- An HTTP request's URL is considered an external site if it does not contain the party's website's hostname.
- An external site can be non-tracking or tracking. Tracking sites are based off https://github.com/duckduckgo/tracker-radar/tree/main/domains. Any external site that is found in this database is considered a tracking site.

Notes/Observations:

- Concern: some sites are not considered as a tracker.
    - Ex. stats.g.doubleclick.net



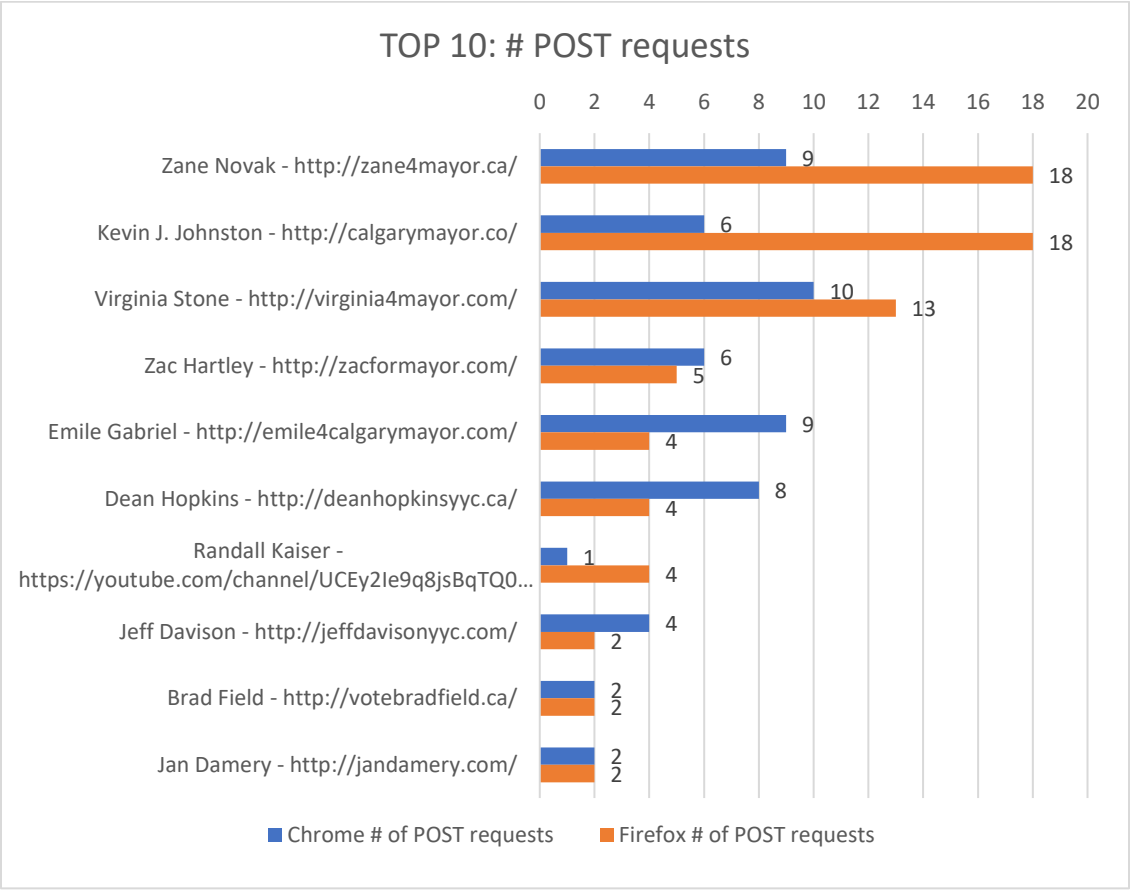


Tracking domains found across all browsers/candidates (Chrome, Firefox)

- google-analytics.com
- fonts.googleapis.com
- googletagmanager.com
- facebook.com
- google.ca
- youtube.com
- translate.googleapis.com
- gstatic.com
- google.com
- polyfill.io
- d3n8a8pro7vhmx.cloudfront.net
- ajax.googleapis.com
- unpkg.com
- statcounter.com
- shop.app
- paypal.com
- juicer.io
- sentry.io

Mayor candidates

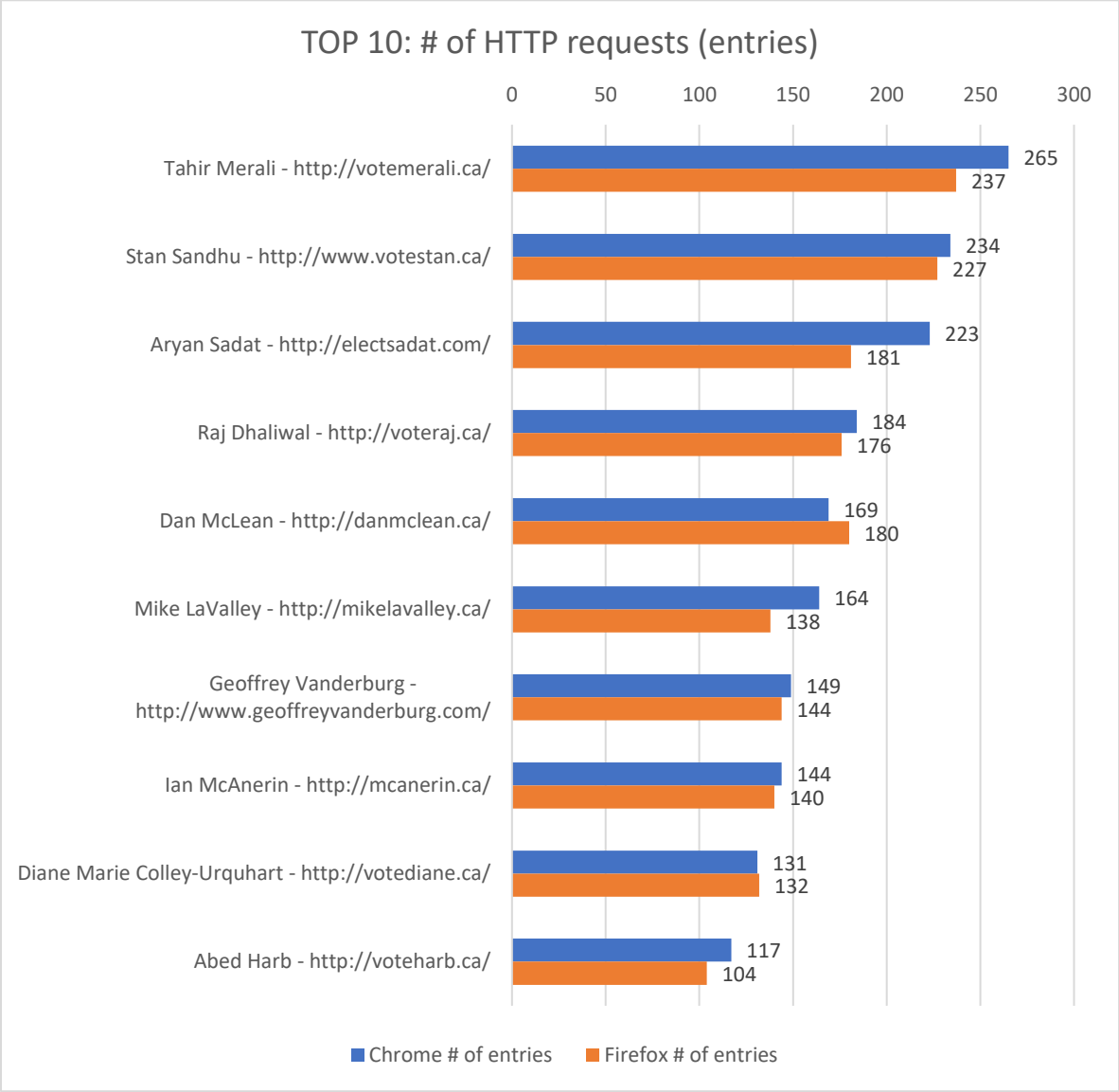
POST requests



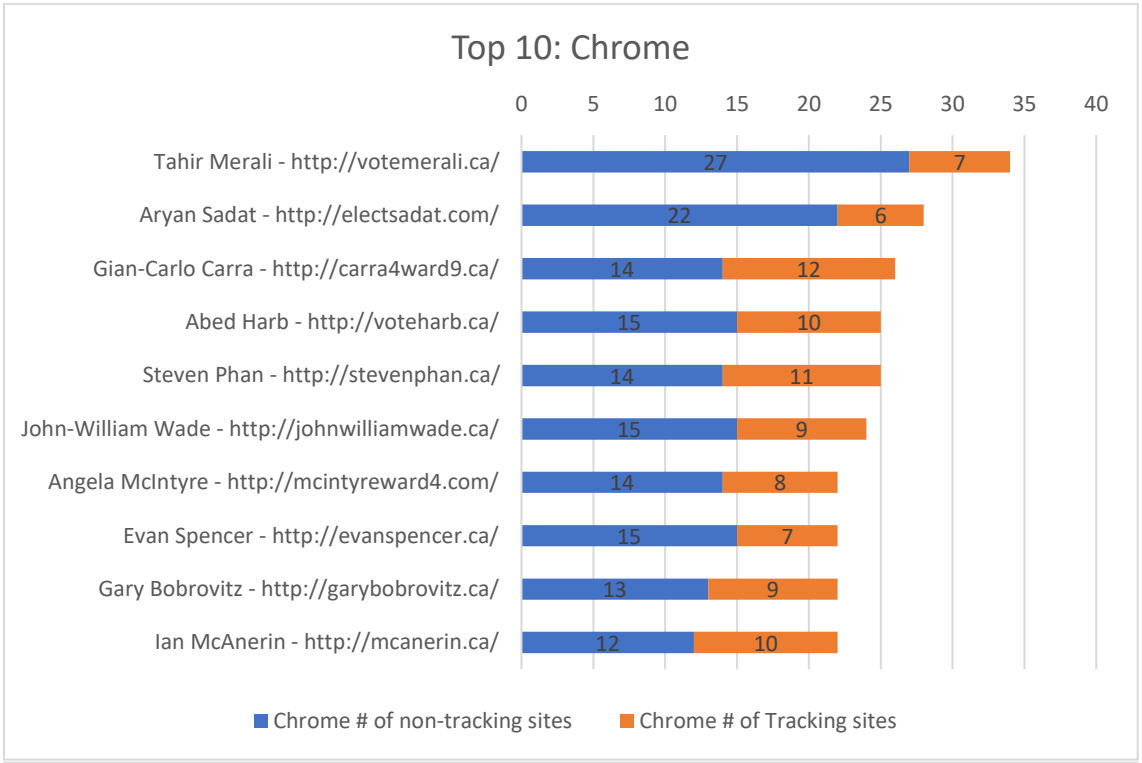
Councillor candidates

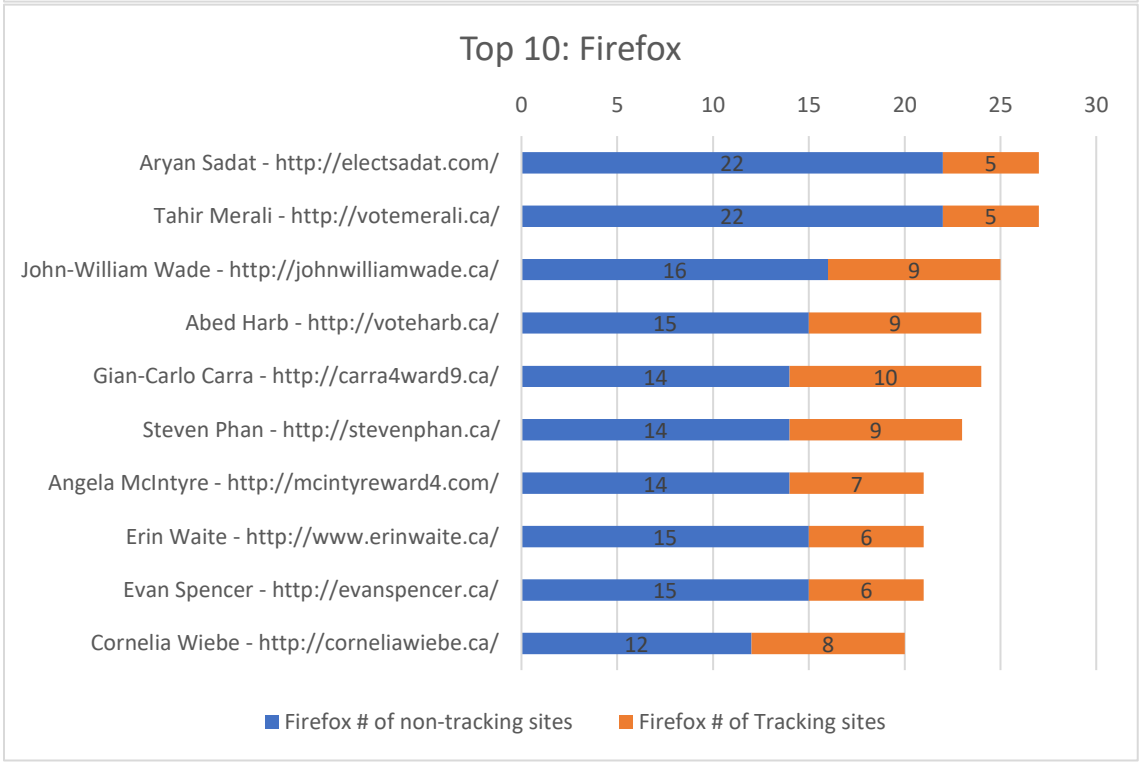
- 91 candidates/websites analyzed

Number of HTTP Requests



Tracking/non-tracking external sites

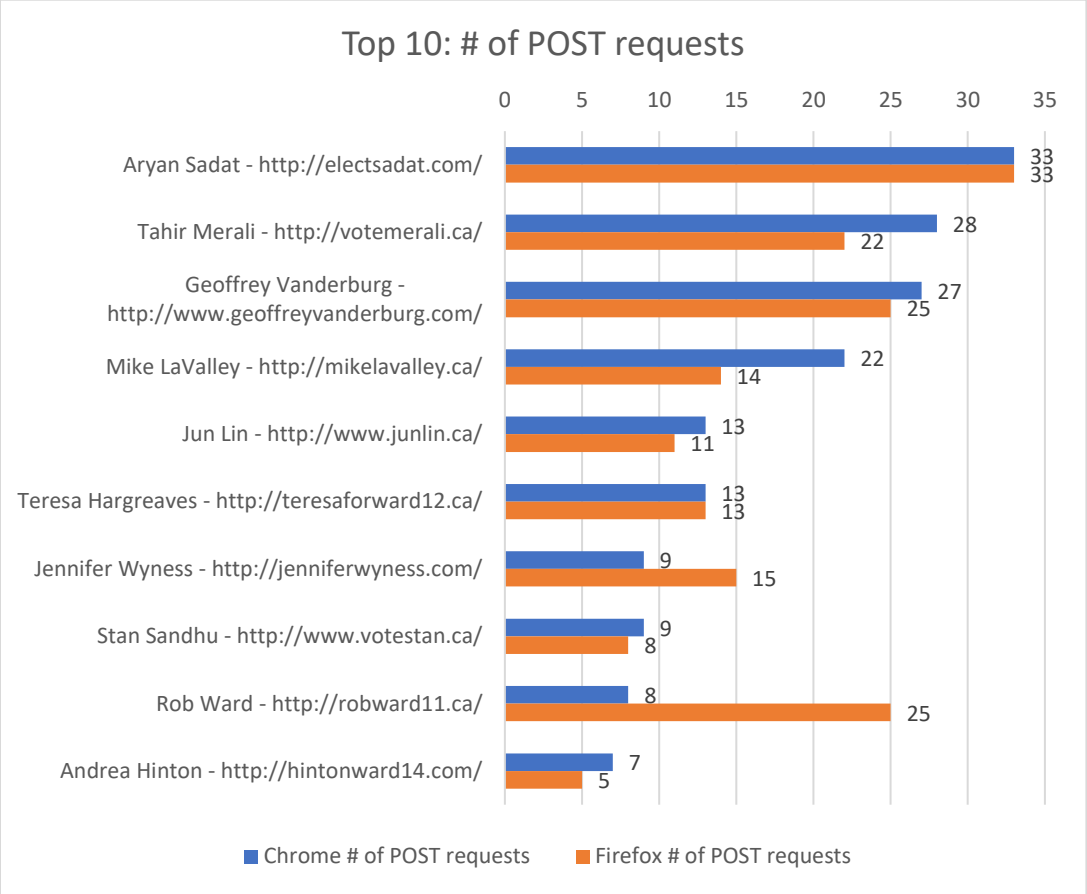




Tracking domains found across all browsers/candidates (Chrome, Firefox)

- fonts.googleapis.com
- ajax.googleapis.com
- d3n8a8pro7vhmx.cloudfront.net
- googletagmanager.com
- youtube.com
- google-analytics.com
- facebook.com
- google.com
- gstatic.com
- nationbuilder.s3.amazonaws.com
- polyfill.io
- unpkg.com
- paypalobjects.com
- cognitoforms.com
- googleadservices.com
- google.ca
- d3e54v103j8qbb.cloudfront.net
- weebly.com
- statcounter.com
- paypal.com
- secureservercdn.net
- webrtc.github.io
- vimeo.com

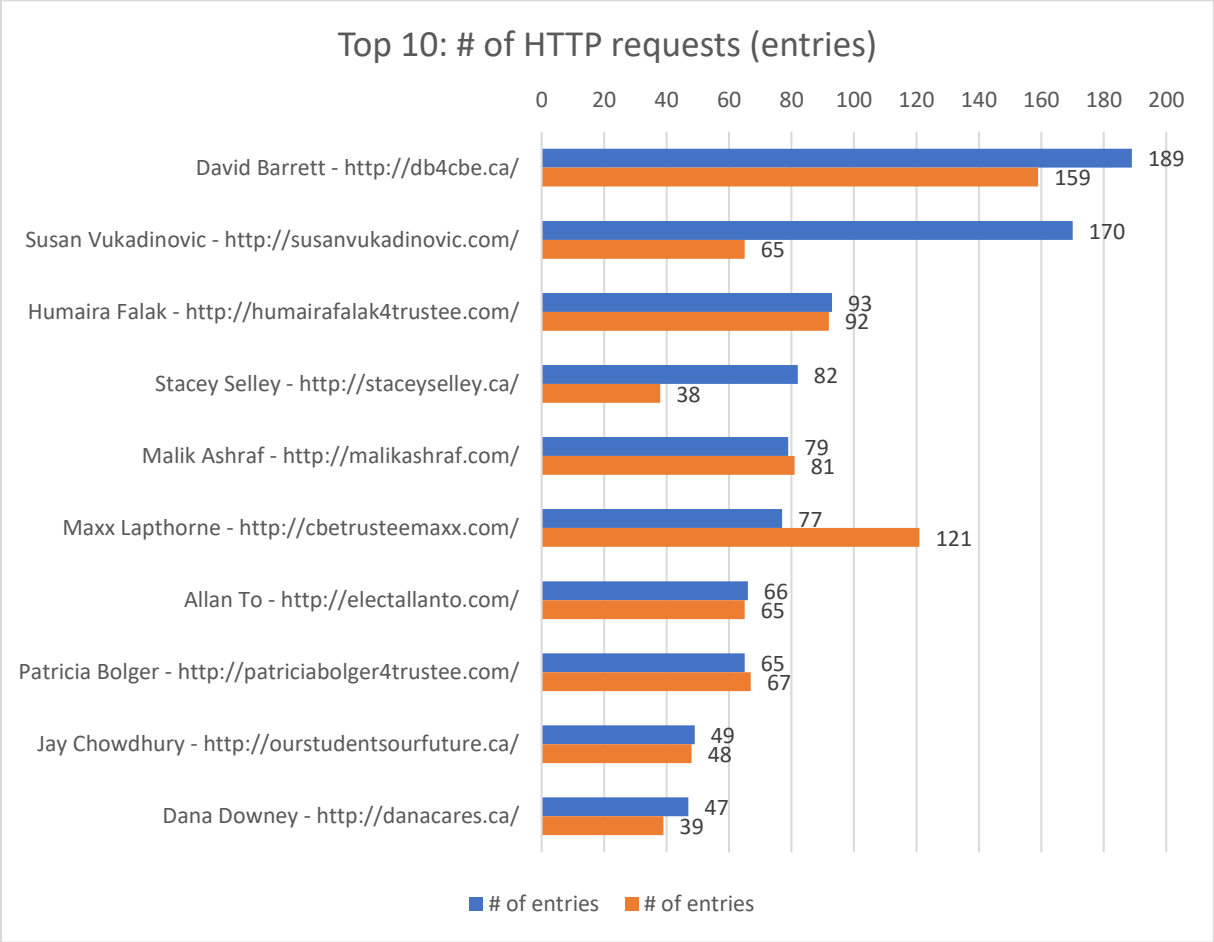
POST requests



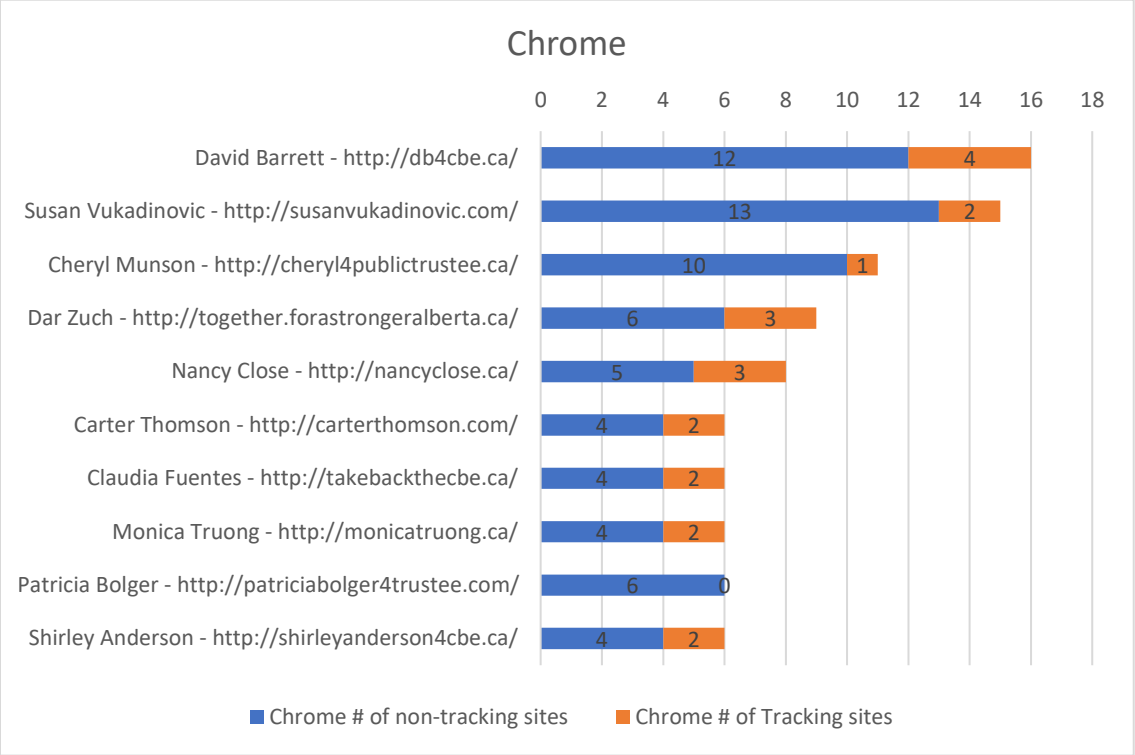
Public school board trustee candidates

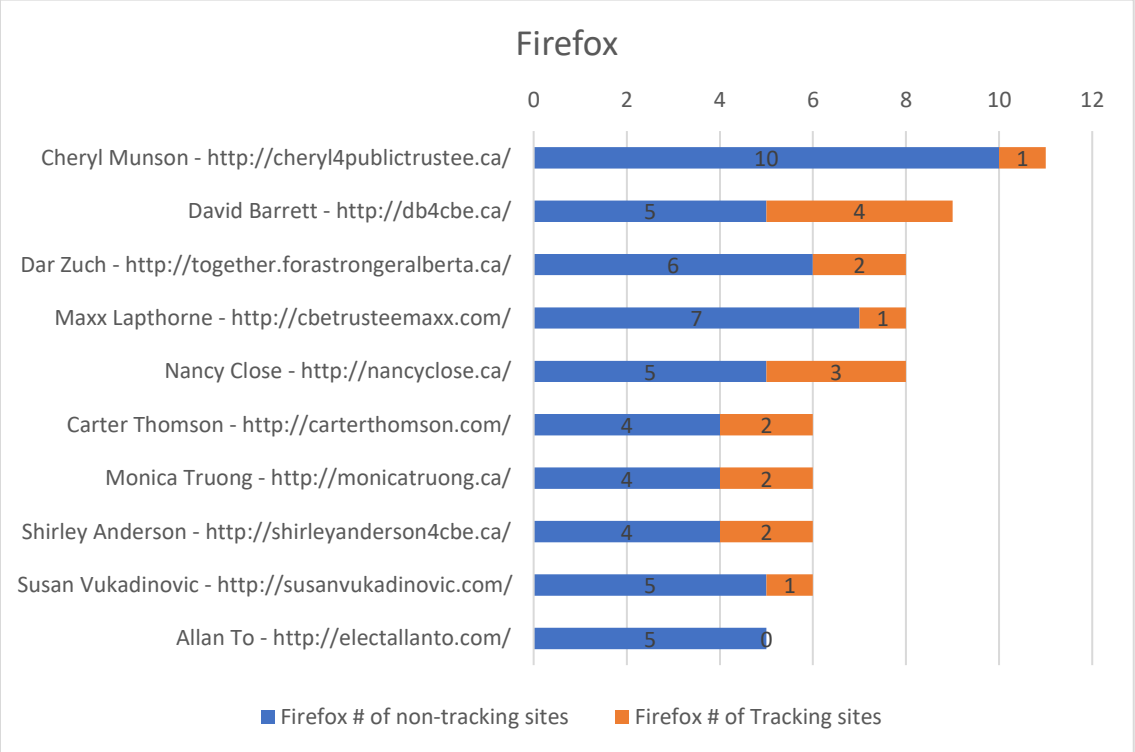
- 29 candidates/websites analyzed

Number of HTTP Requests



Tracking/non-tracking external sites

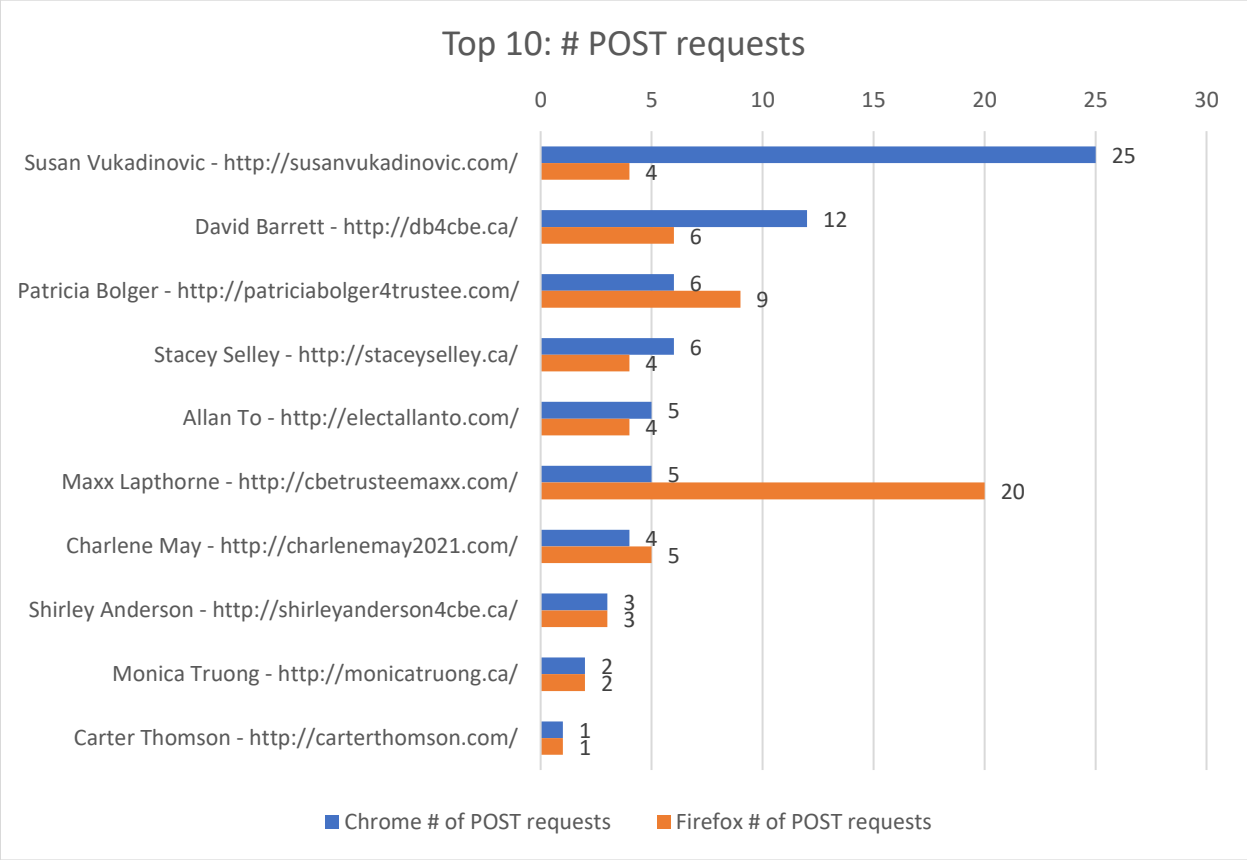




Tracking domains found across all browsers/candidates (Chrome, Firefox)

- google.com
- gstatic.com
- fonts.googleapis.com
- unpkg.com
- youtube.com
- ajax.googleapis.com
- d3n8a8pro7vhmx.cloudfront.net
- google-analytics.com
- wix.com
- paypalobjects.com

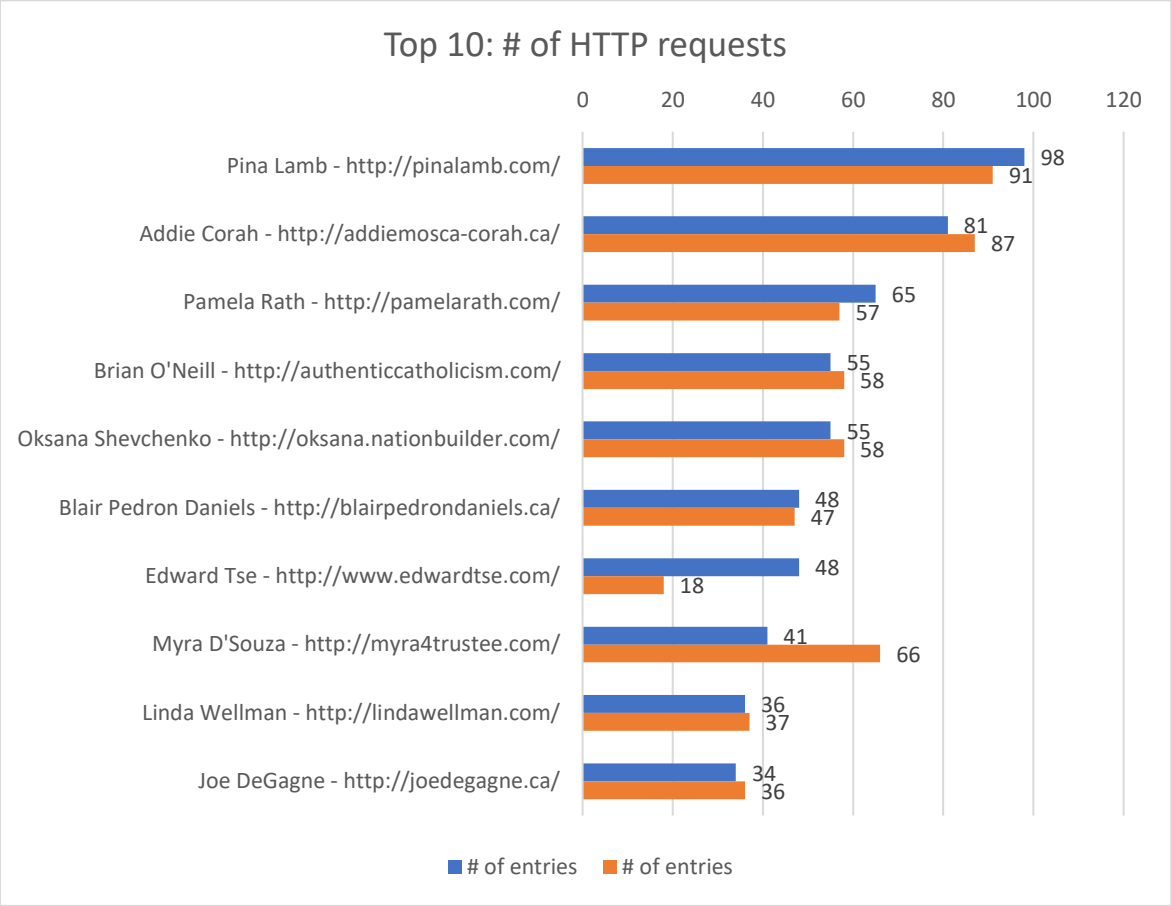
POST requests



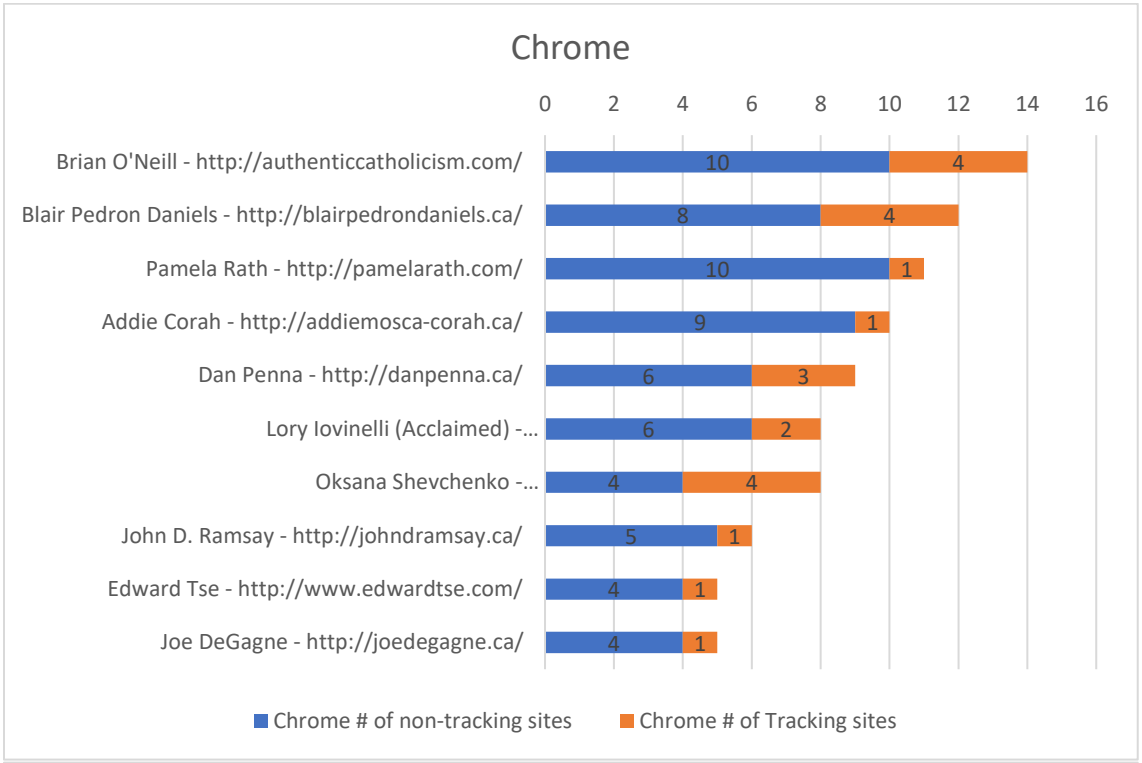
Separate school board trustee candidates

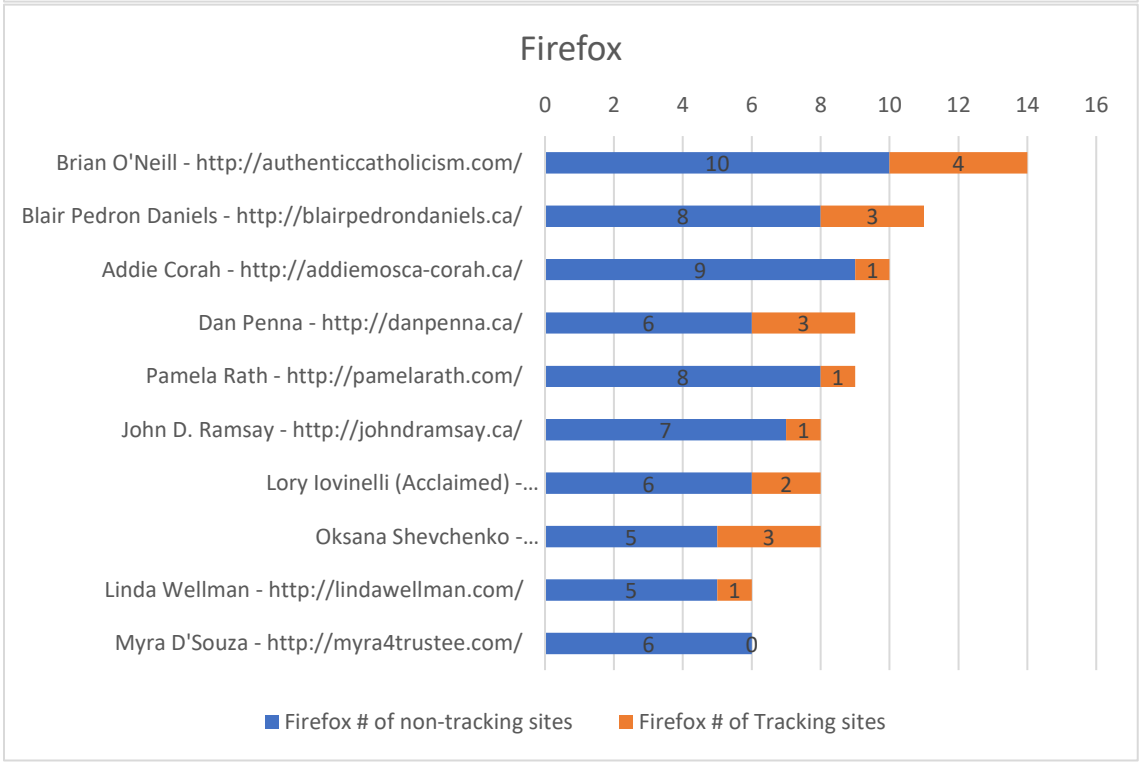
- 15 candidates/websites analyzed

Number of HTTP Requests



Tracking/non-tracking external sites



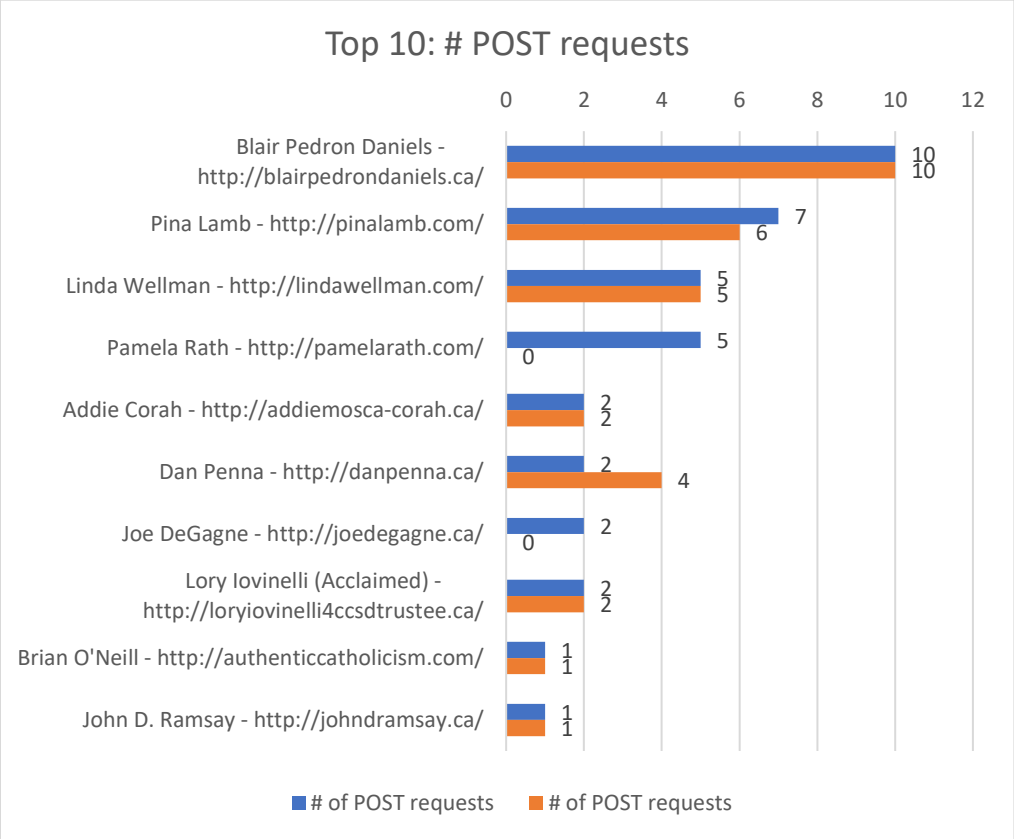


Separate school board trustee candidates

Tracking domains found across all browsers/candidates (Chrome, Firefox)

- fonts.googleapis.com
- youtube.com
- google.com
- gstatic.com
- googletagmanager.com
- google-analytics.com
- vimeo.com
- ajax.googleapis.com
- d3n8a8pro7vhmx.cloudfront.net
- facebook.com

POST requests



Overall observations and next steps

- Must investigate the issue with my programs not detecting any tracking pixels in the HAR files generated by the scrapers. It is possible that automatic collection with the scraper somehow affects the data: manually collected HAR files are much larger than the automatically collected ones. Will try to figure out the reason for this.
- Since there are too many candidates/JSON files for the Calgary elections, I will make notes of more specific observations (ex. requests to sketchy sites) in a separate document. This report is just an overview of the candidates that seem to do the most aggressive tracking.