

What electoral behaviors do voters complain about? A new dataset on electoral malfeasance in Brazil

Jessie Trudeau¹

¹Department of Political Science, Syracuse University

The data

This dataset contains user-generated **electoral complaints** in Rio de Janeiro, Brazil. The dataset covers the **2018** federal and state and the **2020** local elections. During the campaign period, individuals can submit anonymous tips to the **State Electoral Authority** about candidates, parties, or anyone else they believe is breaking electoral laws. There are four channels through which individuals can communicate their tips:

- The **State Public Prosecutor's office** (MPRJ) has an online form to submit tips
- The **Federal Electoral Authority** (TSE) has an app, **Pardal**, to file tips across Brazil
- Users can submit **written tips** via e-mail, a WhatsApp channel, or Facebook Messenger
- Users can **call** a telephone hotline

98% of complaints are submitted through the first two electronic channels. These channels protect user anonymity. Figure 1 shows the sample intake forms for the State Public Prosecutor's office (left) and three screenshots from the Federal Electoral Authority's **Pardal** app (right).

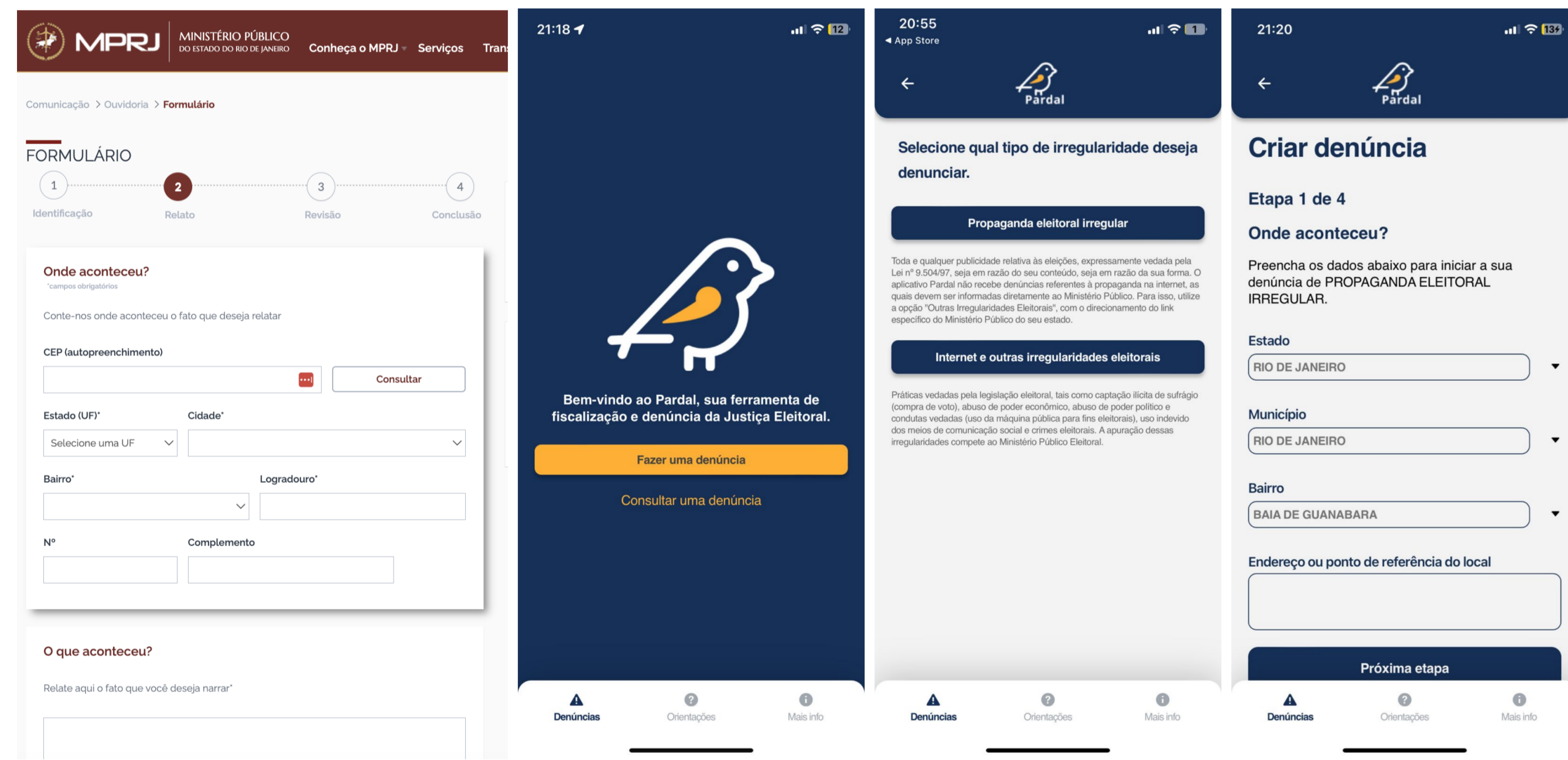


Figure 1. Electoral complaint intake forms: MPRJ (panel 1) and TSE's **Pardal** (panels 2-4).

About the complaints

Each complaint contains the following information:

- Date
- Location
- Candidate name
- Candidate's party
- Description of electoral offense (*open text entry*)

The description of electoral offenses ranges from benign...

There is a campaign sticker somewhere it shouldn't be.

...to disturbing.

The candidate is giving gasoline from his gas stations to drug traffickers so he can campaign in that neighborhood.

Electoral complaints in Rio de Janeiro

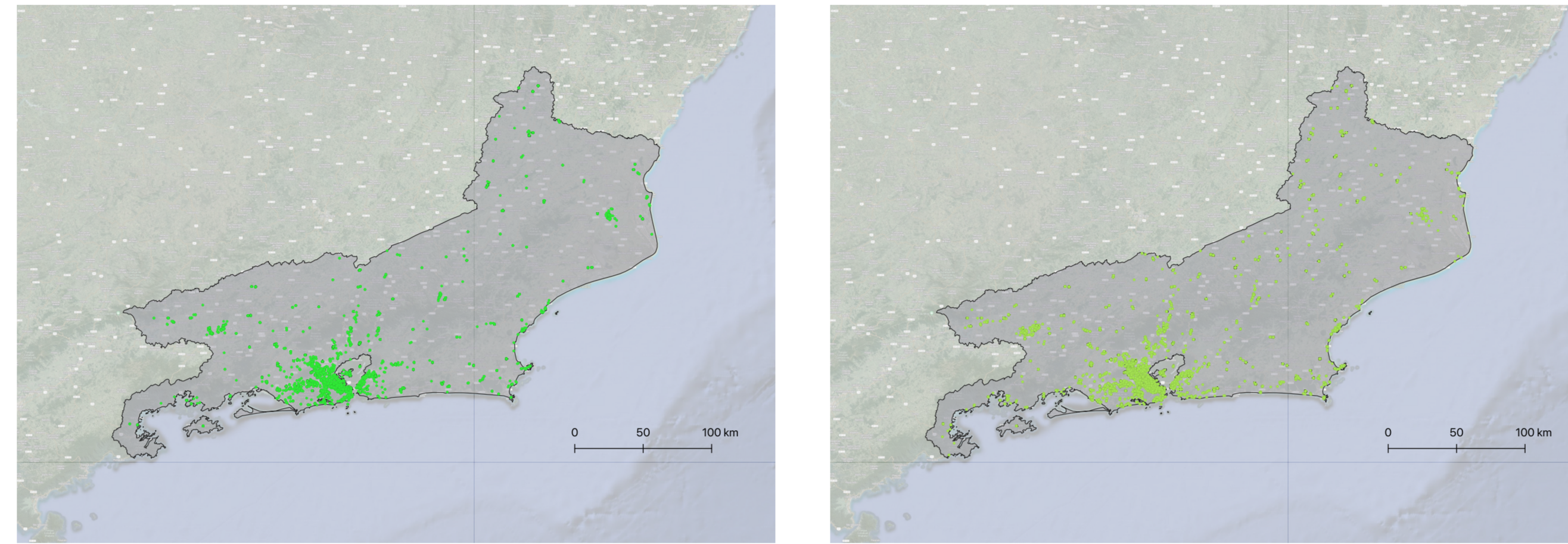


Figure 2. Electoral complaints in the State of Rio de Janeiro in 2018 (left, $n = 9,815$) and 2020 (right, $n = 19,927$).

Methods

Prior to analysis, I cleaned and standardized the data¹:

- **Candidate names:** The electronic forms (online and the **Pardal** app) have an open text-entry field for the candidate name. This generates noise because of spelling errors, compound phrases (e.g., a candidate's name plus their party), or because of inconsistencies between a candidate's legal name and their "Ballot box name," a nickname or a shorthand devised for the elections. To match each complaint with a candidate in the electoral court's list, I used the following sequential string matching protocol:
 1. Clean user-entered names to remove punctuation and stop words, including party and coalition names.
 2. Match user-entered names to candidates if they are exact matches to candidate legal names or candidate ballot box names. Verify by confirming the party and municipality are a match.
 3. For user-entered names with no exact match, calculate the Jaro-Winkler distance (Wang et al. 2017) to determine the best possible match. The Jaro-Winkler distance is a multidimensional measurement of the similarity of two strings. The output suggests the candidate name that has the highest match score with what the user wrote.
 4. Hand-code the results of the Jaro-Winkler matches through a close read of the complaint text, replacing the incorrect matches with correct ones. **Upon completion, 96% of complaints were either exactly matched to a candidate name or verified through the Jaro-Winkler and hand-coding process.**
- **Complaint location:** Following some cleaning and processing of the user-entered location information, I used Google Maps and ArcGIS Pro to geolocate each complaint. For complaints without a specific address, I located them either a) in the geometric centroid of the neighborhood or b) in the geometric centroid of the municipality mentioned.

Complaint topics

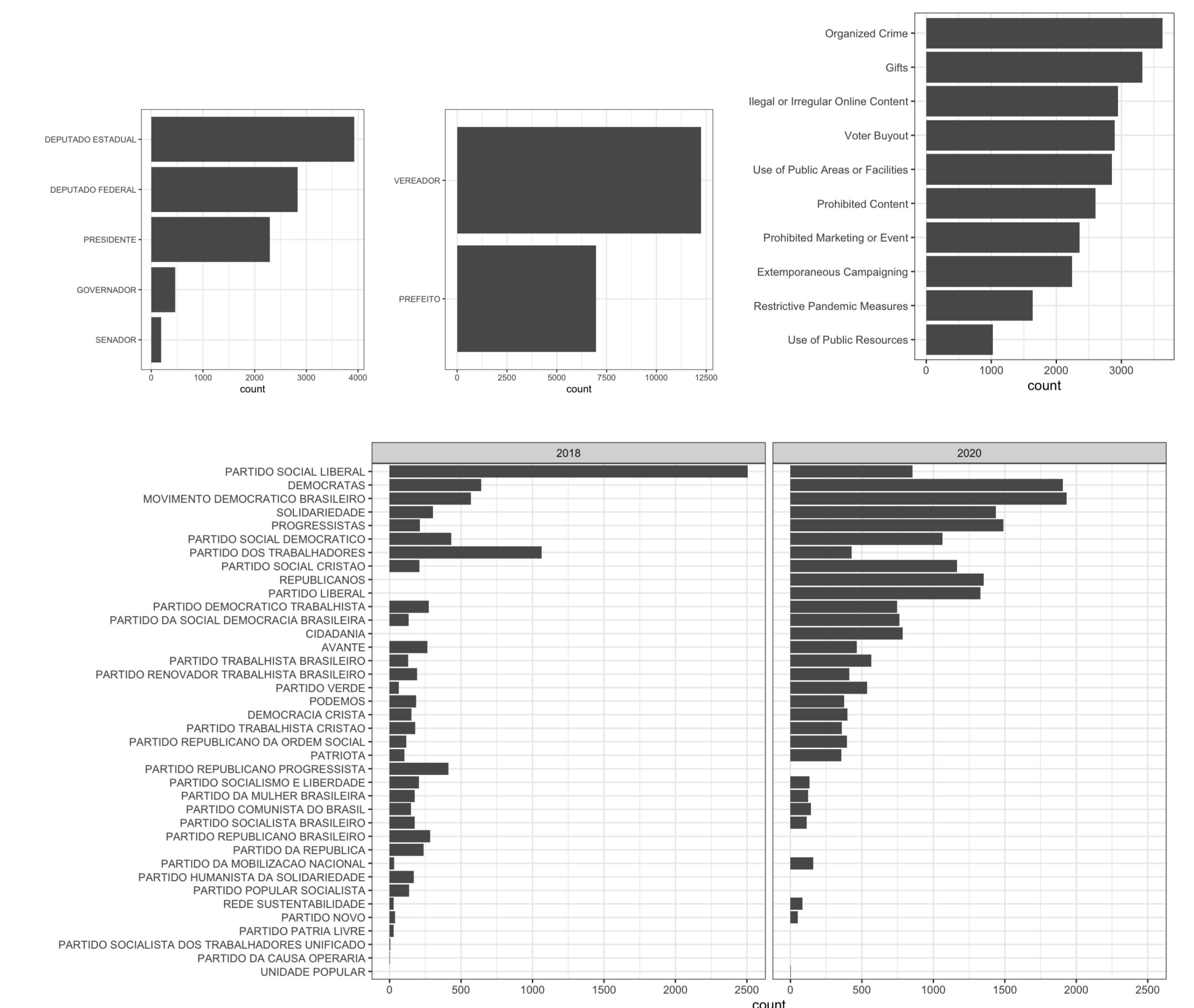
I constructed dictionaries of common words to classify the text description of the complaint by topic. I calculate the **probability the complaint is about each of the below topics**:

- Extemporaneous campaigning
- Gifts
- Illegal online content
- Organized crime
- Prohibited content
- Prohibited marketing or event
- Restrictive pandemic measures
- Use of public areas
- Use of public resources
- Vote buying

Descriptive findings

- **Voters are more likely to complain about downballot candidates:** both State Deputies (2018) and City Council (2020) candidates
- **Voters are most likely to complain about:** 1) organized crime, 2) provision of gifts to voters, 3) online campaigning, and 4) vote buying.
- **All major parties** are complained about. In **2018**, the parties of the two leading presidential candidates (PSL and PT) were complained about most. Right-wing and center-right parties were complained about most in the local election of **2020**.

Descriptive statistics



References

Wang, Yaoshu, Jianbin Qin, and Wei Wang (2017). "Efficient Approximate Entity Matching Using Jaro-Winkler Distance". *Web Information Systems Engineering – WISE 2017*. Cham: Springer International Publishing, pp. 231–239.

¹This research was supported by the Electoral Integrity Project and could not have been done without the help of research assistants Rodrigo Abreu Carvalho, João Oliveira, and Almila Basak.