

## BU Statistics

This project processes the `bus-%UF.comp` files (or a directory of ".bu.json.gz" files) and generates multiple statistics of the BU data.

- BU stands for "Boletim de Urna" (Ballot Box).
- See the `../tse_public_data/` directory to generate the `bu-files/` directory and the `.bu.json.gz` files.
- See the `bu_compressor.dart` tool to generate the compressed `bus-%UF.comp` files.

## Dart SDK

To download and install the Dart SDK see: <https://dart.dev/get-dart>

- You need at least Dart 2.17.0 to run this project.

After install the Dart SDK you need to get the project dependencies:

```
cd path/to/bu_statistics
dart pub get
```

## Usage

To generate statistics of a state in Brazil you need 2 parameters:

- The `bu-files` directory:
  - With ".bu.json.gz" files in the structure:
    - `bu-files/$uf/$cityCode/$uf-$cityCode-$zoneCode-$sectionCode.bu.json.gz`
    - OR: `bu-files/$uf/any/path/any-file-name.bu.json.gz`
- OR a compressed BUs files: `bu-files/bus-%UF.comp`
- The output directory of the generated statistics (CSV files).

Generating the statistics for a state in Brazil:

```
cd ../bu_statistics/
dart bin/bu_statistics.dart ../tse_public_data/bu-files/sp /tmp/bu-statistics-output
```

Using the compressed BUs file:

```
dart bin/bu_statistics.dart ../tse_public_data/bu-files/bus-sp.comp /tmp/bu-statistics-output
```

- UF stands for a state in Brazil.

## Decimal delimiter

To change the decimal delimiter of the CSV files pass the parameter `commadecimal` :

```
dart bin/bu_statistics.dart ../tse_public_data/bu-files/sp /tmp/bu-statistics-output commadecimal
```

- NOTES:
  - In US the decimal delimiter is . (point) and in Brazil the decimal delimiter is , (comma).
  - The default decimal delimiter in Dart for double values is . .
  - Some software in Brazil need the , (comma) delimiter to work correctly (like MS Excel and Numbers - Apple ).

## BU Metrics

The metrics will analyse the BU files using a BU variable to order them and separate them in blocks. Then it will generate the voting statistics for each block (non-cumulative from previous blocks), to see the voting evolution based on some metric.

- `closeDate` :
  - BUs ordered by `close date`.
    - The date of the LAST vote int the Voting Machine.
  - Default block time window: 30s
  - Generates the voting statistics of BUs blocks for each 30s, ordered by `closeDate` .
  - Checks if the `closeDate` (last vote) of a Voting Machine influences the voting ratio for a candidate.
- `generationDate` :
  - BUs ordered by `generation date`.
    - The date that the original BU file was generated.
  - Default block time window: 30s
  - Generates the voting statistics of BUs blocks for each 30s, ordered by `generationDate` .
  - Checks if the `generationDate` of a Voting Machine influences the voting ratio for a candidate.
- `emissionDate` :
  - BUs ordered by `emission date`.
    - The date that the original BU file was emitted.
  - Default block time window: 30s
  - Generates the voting statistics of BUs blocks for each 30s, ordered by `emissionDate` .
  - Checks if the `emissionDate` of a Voting Machine influences the voting ratio for a candidate.
- `loadDate` :
  - BUs ordered by `load date`.
    - The date that the Voting Machine system was loaded/installed.
  - Default block time window: 30s
  - Generates the voting statistics of BUs blocks for each 30s, ordered by `loadDate` .
  - Checks if the loaded/installed system at some date influences the voting ratio for a candidate.
- `votersReleasedByCodeRatio` :
  - BUs ordered by `votersReleasedByCodeRatio` .
    - The ratio of votes released by code (without biometric input: finger scan).
  - Default block window: 1% (0.01)
  - Generates the voting statistics of BUs blocks for each 1%, ordered by `votersReleasedByCodeRatio` .
  - Checks if Voting Machines with more votes without biometric input have some anomaly/tendency for a candidate.
- `onlyPresidentVotesRatio` :
  - BUs ordered by `onlyPresidentOfficeVotesRatio` .
    - The ratio of votes only for president (without vote for governor) in the same election day.
    - It can have a ratio > then 1.0, since it's possible to have votes for governor and not for president too.
  - Default block window: 1% (0.01)
  - Generates the voting statistics of BUs blocks for each 1%, ordered by `onlyPresidentOfficeVotesRatio` .
  - Checks if Voting Machines with more votes only for president have some anomaly/tendency for a candidate.
- `presidentAbstentionRatio` :
  - BUs ordered by `presidentOfficeAbstentionRatio` .
    - The ratio of abstentions for president in the same election day.
  - Default block window: 1% (0.01)
  - Generates the voting statistics of BUs blocks for each 1%, ordered by `presidentOfficeAbstentionRatio` .
  - Checks if Voting Machines with more abstentions for president have some anomaly/tendency for a candidate.

NOTE:

- All the metrics above should NOT have ANY influence over candidates voting ratio.
- Voting ratio statistics are generated for the 2 top winners.

## Generated CSV

The output directory will be used to generate the statistics in the CSV format.

Here's an example of generated statistics:

- **Metric:** `closeDate`
- **UF:** `sp`
- **Parameter:** `commadecimal` (using `,` as decimal separator)
- **CSV File:** `bu--statistics--closeDate--sp.csv`

```
closeDate,ratio:22,ratio:13,votes:22,votes:13,votes.mean:22,votes.mean:13,votes.stdv:22,votes.stdv:13,votes.mean.ratio:22,votes.mean.ratio:13,votes.stdv.ra
17:00:00,0.5319,0.3602,545029,369012,123.8139,83.8092,33.5488,27.9830,0,0,0,0,1024592,232.7032,4403,1081960,312508,0.2241,1394468,245.7325,70.9762,316.7086,
17:00:30,0.5331,0.3607,935715,633197,124.6125,84.3251,33.0360,27.8231,0,0,0,0,1755296,233.7590,7509,1853732,533225,0.2234,2386957,246.8680,71.0115,317.8795,
17:01:00,0.5277,0.3645,978305,675654,125.5042,86.6668,33.0531,27.7923,0,0,0,0,1853807,237.7895,7796,1958942,556199,0.2211,2515141,251.2753,71.3442,322.6194,
17:01:30,0.5205,0.3694,875198,621159,124.7609,88.5473,32.5571,28.7578,0,0,0,0,1681477,239.6974,7015,1776777,500792,0.2199,2277569,253.2825,71.3887,324.6713,
17:02:00,0.5124,0.3760,801159,587892,124.1337,91.0754,33.1706,29.0581,0,0,0,0,1563479,242.2121,6455,1652913,464824,0.2195,2117737,256.0671,72.0099,328.0770,
17:02:30,0.5046,0.3823,687946,521232,123.0893,93.2603,32.8592,29.4477,0,0,0,0,1363313,243.9279,5589,1442363,402836,0.2183,1845199,258.0717,72.0766,330.1483,
17:03:00,0.4993,0.3870,617503,478658,122.3263,94.8213,33.3557,30.4531,0,0,0,0,1236732,244.9945,5048,1308617,363639,0.2175,1672256,259.2347,72.0363,331.2710,
17:03:30,0.4934,0.3912,537492,426138,122.3241,96.9597,33.6995,31.1891,0,0,0,0,1089377,247.8673,4395,1153053,315410,0.2148,1468463,262.3556,71.7656,334.1213,
17:04:00,0.4882,0.3955,467084,378421,120.9749,98.0111,33.3388,31.7255,0,0,0,0,956838,247.8213,3861,1014121,278180,0.2153,1292301,262.6576,72.0487,334.7063,
17:04:30,0.4791,0.4033,422384,355564,119.7913,100.8406,33.6010,31.6719,0,0,0,0,881624,250.0352,3526,934212,256274,0.2153,1190486,264.9495,72.6812,337.6307,
17:05:00,0.4742,0.4069,378934,325173,119.1242,102.2235,33.2467,32.3289,0,0,0,0,799152,251.2267,3181,847335,233998,0.2164,1081333,266.3738,73.5611,339.9349,
17:05:30,0.4645,0.4165,318909,285990,116.5603,104.5285,32.9519,31.7874,0,0,0,0,686615,250.9558,2736,727772,200512,0.2160,928284,265.9985,73.2865,339.2851,
17:06:00,0.4654,0.4174,279008,250226,118.1236,105.9382,32.9189,32.6771,0,0,0,0,599499,253.8099,2362,636380,174808,0.2155,811188,269.4242,74.0085,343.4327,
17:06:30,0.4553,0.4265,239231,224091,116.5843,109.2061,33.1938,33.4634,0,0,0,0,525386,256.0361,2052,557622,150824,0.2129,708446,271.7456,73.5010,345.2466,
17:07:00,0.4589,0.4211,226740,208061,118.0938,108.3651,32.1293,32.3798,0,0,0,0,494048,257.3167,1920,524576,142979,0.2142,667555,273.2167,74.4682,347.6849,
17:07:30,0.4502,0.4299,196948,188068,116.3995,111.1513,31.4052,32.8595,0,0,0,0,437509,258.5751,1692,464587,126995,0.2147,591582,274.5786,75.0561,349.6348,
17:08:00,0.4470,0.4324,177433,171646,116.5788,112.7766,32.9938,32.8857,0,0,0,0,396935,260.7983,1522,421668,113090,0.2115,534758,277.0486,74.3035,351.3522,
17:08:30,0.4469,0.4317,157426,152061,117.1324,113.1406,32.4949,32.2370,0,0,0,0,352276,262.1101,1344,374234,102765,0.2154,476999,278.4479,76.4621,354.9100,
17:09:00,0.4391,0.4394,141773,141868,115.1690,115.2461,32.2342,32.7763,0,0,0,0,322882,262.2924,1231,343516,93084,0.2132,436600,279.0544,75.6166,354.6710,
17:09:30,0.4393,0.4391,131108,131057,116.4369,116.3917,32.0260,31.2394,0,0,0,0,298462,265.0639,1126,317010,85285,0.2120,402295,281.5364,75.7416,357.2780,
17:10:00,0.4437,0.4447,114929,117562,115.7392,118.3907,32.5335,31.5703,0,0,0,0,264365,266.2286,993,280689,76326,0.2138,357015,282.6677,76.8640,359.5317,0.21
```

The metric generates the statistics for the 2 top winners.

CSV columns:

- `closeDate` : The time of BUs block (by `closeDate` and 30s window).
- `ratio:22` : Votes ratio for candidate 22 (non-cumulative).
- `ratio:13` : Votes ratio for candidate 13 (non-cumulative).
- `votes:22` : Votes for candidate 22 (non-cumulative).
- `votes:13` : Votes for candidate 13 (non-cumulative).
- `votes.mean:22` : The average of voters per BU for candidate 22 .
- `votes.stdv:22` : the standard deviation of `votes.mean:22` .
- `votes.mean:13` : The average of voters per BU for candidate 13 .
- `votes.stdv:13` : The standard deviation of `votes.mean:13` .
- `votes:*` : Votes for all candidates (non-cumulative).
- `votes*.mean` : The average of total valid votes per BU.
- `bus` : Number of BU files in the block.
- `turnout` : The total number of people voting.
- `abstentions` : The number of abstentions (people who did not attend the election).
- `abstentionsRatio` : The ratio of abstentions.
- `eligibleVoters` : The total number of eligible voters in the BU.
- `turnout.mean` , `abstentions.mean` , `eligibleVoters.mean` : The average by BU.
- `presidentAbstentionRatio` : Abstention ratio for the president votes.
- `presidentAbstentionRatio.stdv` : The standard deviation of `presidentAbstentionRatio` .
- `votersBiometric` : Number of votes with biometric identification.
- `votersByCode` : Number of votes with failed biometric identification, and released by code.
- `votersWithoutBiometrics` : Number of votes without biometric identification.
- `votersInTransit` : Number of voters in transit in the BU (can vote only for president).
- `votersByCode.mean` , `votersBiometric.mean` , `votersWithoutBiometrics.mean` , `votersInTransit.mean` : The average by BU.
- `total` : Total number of votes (cumulative).
- `totalRatio` : Ratio of `total` votes (cumulative).

## LICENSE

MIT License

## Authors

The authors of this analysis worked hard to give a fully reproducible work, what legitimates the data and the analysis of this Election data.

This work was verified by people at different major universities in Brazil, USA and Switzerland.

If necessary this address will be used to guarantee the authorship of the work:

1Gh5QtC7UpLt31Ma85HWZLduGwPFHFS2AH