FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

Pattern recognition in election data

Similarity, visualization and forensic analysis – what statistics can and cannot tell us about an election

Mgr. Michal Škop, Ph.D.

DataTimes.cz

Czech local elections 2026 · Most

“It was tight, *but we are in the assembly.*”

Pirates, Greens and Mayors, the night after the vote. They finished at 5.001 % – the threshold is 5 %.



Elections run on *trust.*

What can an algorithm check? Three steps – see, compare, check. Most is our test case: we come back to it in step 3.

PART 1 OF 3

See ⟶

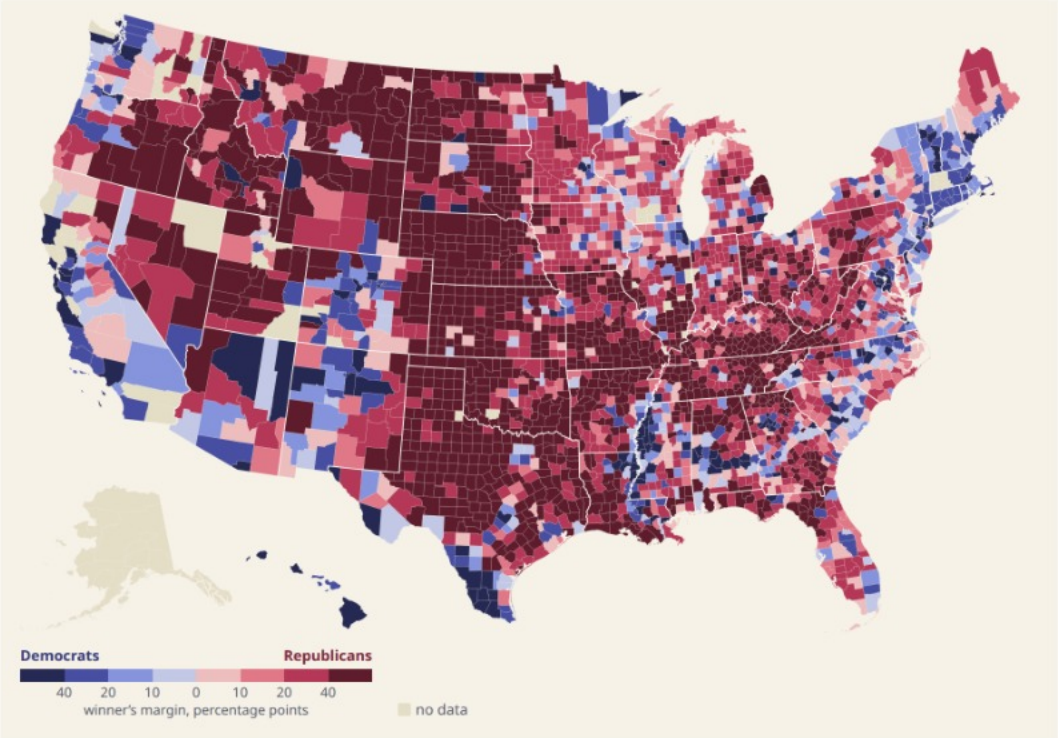
A map is the first statistic people see – and it often lies



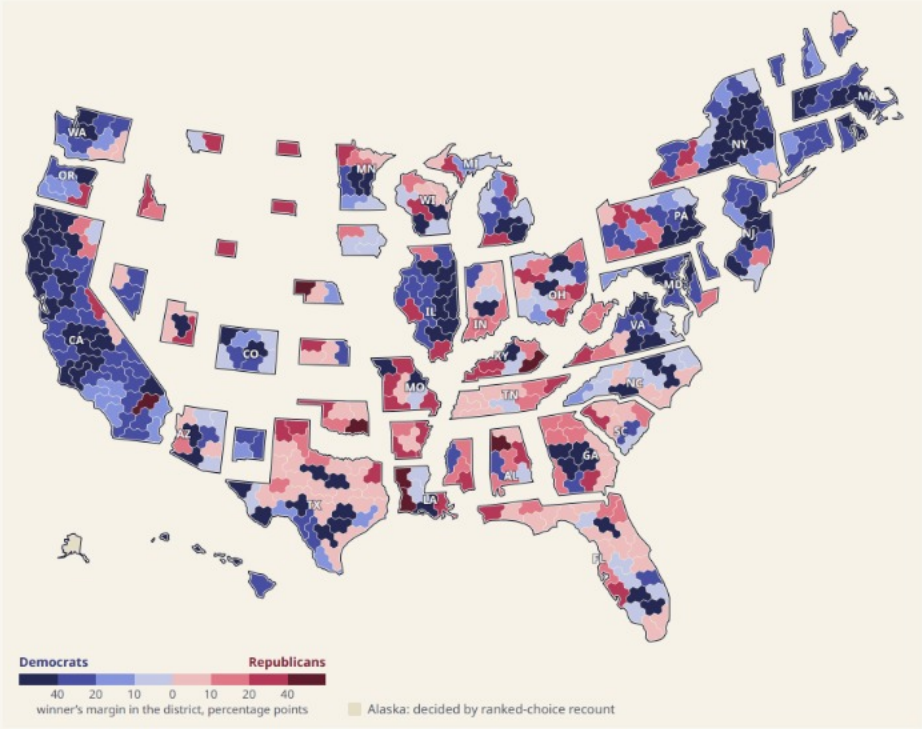
SEE

Same House election, two maps, two stories

US House 2026 · left: by county, right: every one of 435 districts the same size · colour = winner's margin



• authors: Kateřina Mahdalová & Michal Škop
• data: NYT, Civic API (as of 4 Nov 2026), US Census Bureau, DataTimes.cz

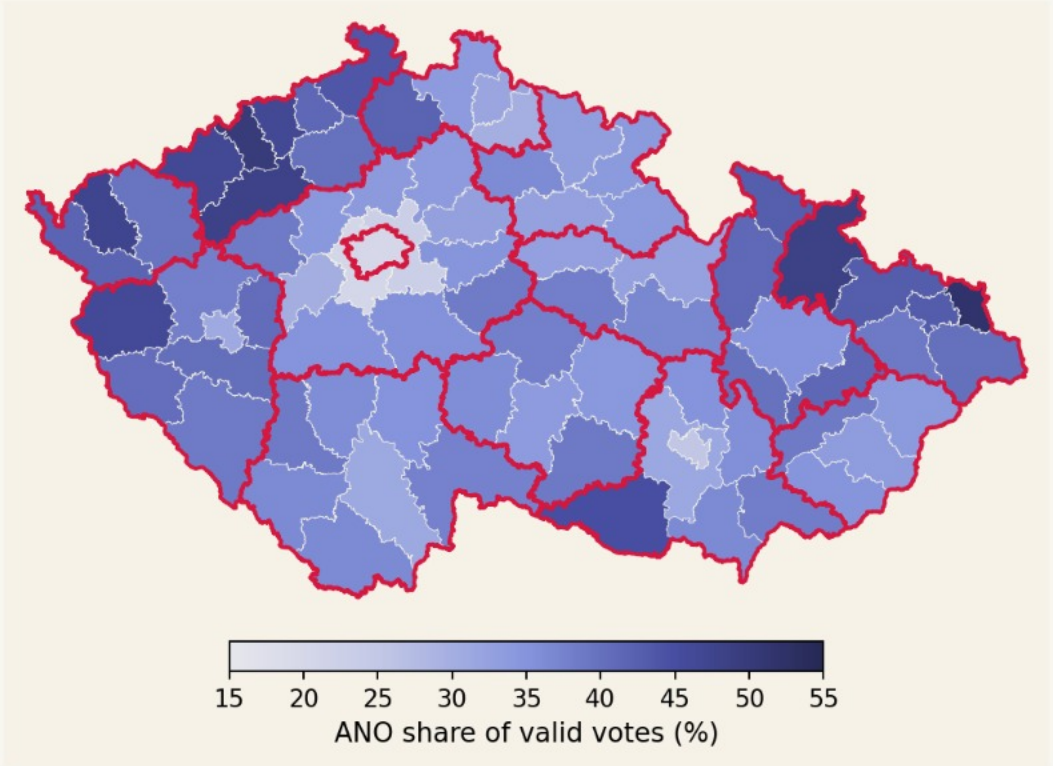


• authors: Kateřina Mahdalová & Michal Škop
• data: NYT, Civic API (4 Nov 2026), Daniel Donner / The Downballot (CC BY 4.0), DataTimes.cz

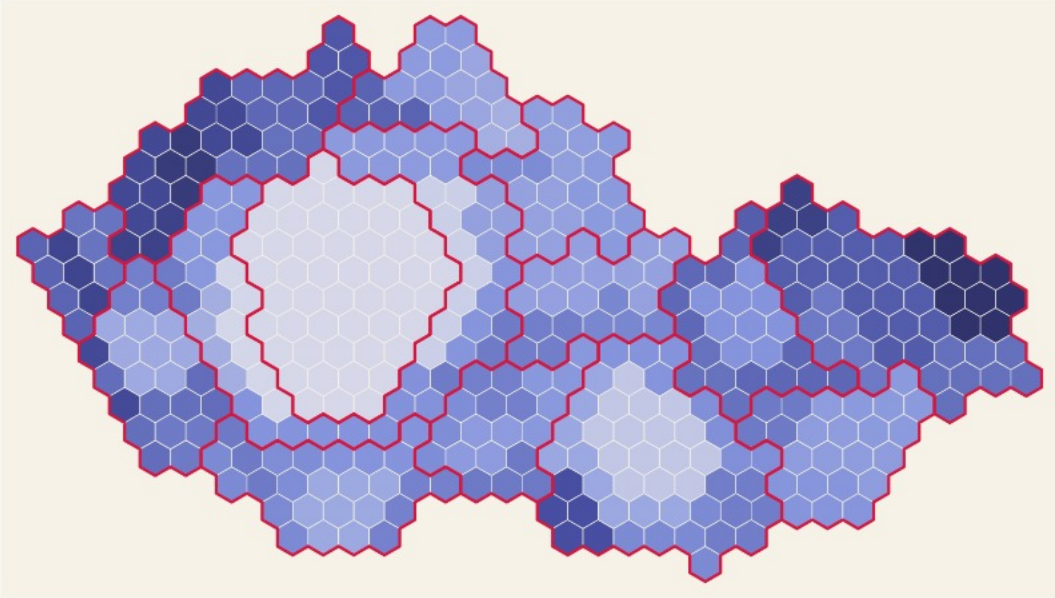
SEE

Prague is a dot on one map, a region on the other

ANO share of valid votes by okres · Chamber of Deputies 2025 · red lines = kraje · right: one hexagon = 25,000 people



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation; okres boundaries ČÚZK/ČSÚ



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation; tilegram by DataTimes.cz

PART 2 OF 3

Compare ⟶

Which parties are voted for similarly – and which places vote alike?



COMPARE

A polling station is a fingerprint of a few numbers

Share of valid votes per party · real polling stations, Chamber of Deputies 2025

Input.

Czechia: 14,711 polling stations × 8 shares (7 parties + others), weighted by valid votes.

Idea.

Dimensionality reduction: 8 dimensions → 2 that keep most of the information. The same trick as embeddings in ML.

44 %

of all variation sits on one single axis (PC1).

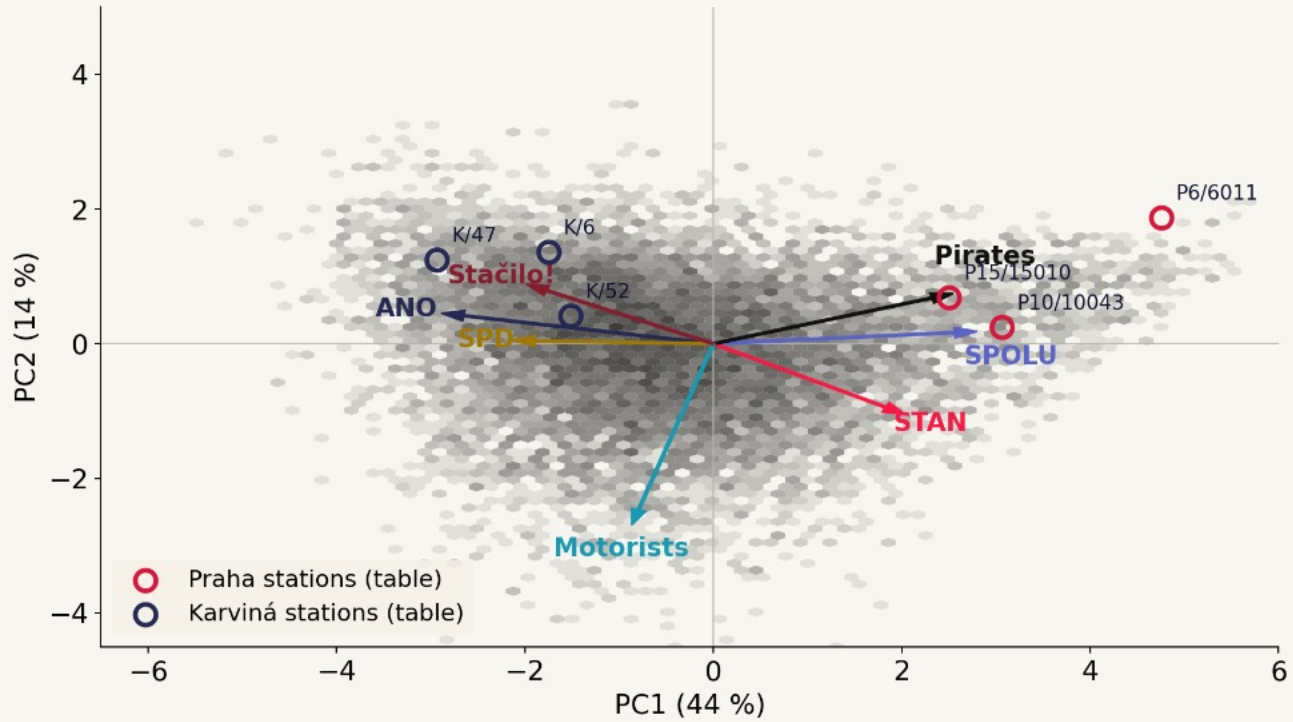
Polling station	valid votes	ANO	SPD	SPOLU	STAN
Praha 6 · no. 6011	669	12.6	4.2	37.5	13.0
Praha 10 · no. 10043	612	19.0	4.2	34.6	14.7
Praha 15 · no. 15010	612	27.9	3.4	35.0	13.1
Karviná · no. 52	696	43.2	8.3	16.2	9.6
Karviná · no. 6	715	52.6	6.6	14.0	7.7
Karviná · no. 47	796	62.4	7.3	7.0	6.7

Share of valid votes, %. Colour = higher share within the party's column. Praha on top, Karviná below. Chamber of Deputies 2025, Czech Statistical Office.

COMPARE

One dominant axis splits the country

PCA of 14,711 polling stations · Chamber of Deputies 2025 · rings = the six stations from the table



- authors: Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

PC1 · 44 %.

ANO · SPD · Stačilo! on the left, SPOLU · Pirates · STAN on the right.

PC2 · 14 %.

Motorists.

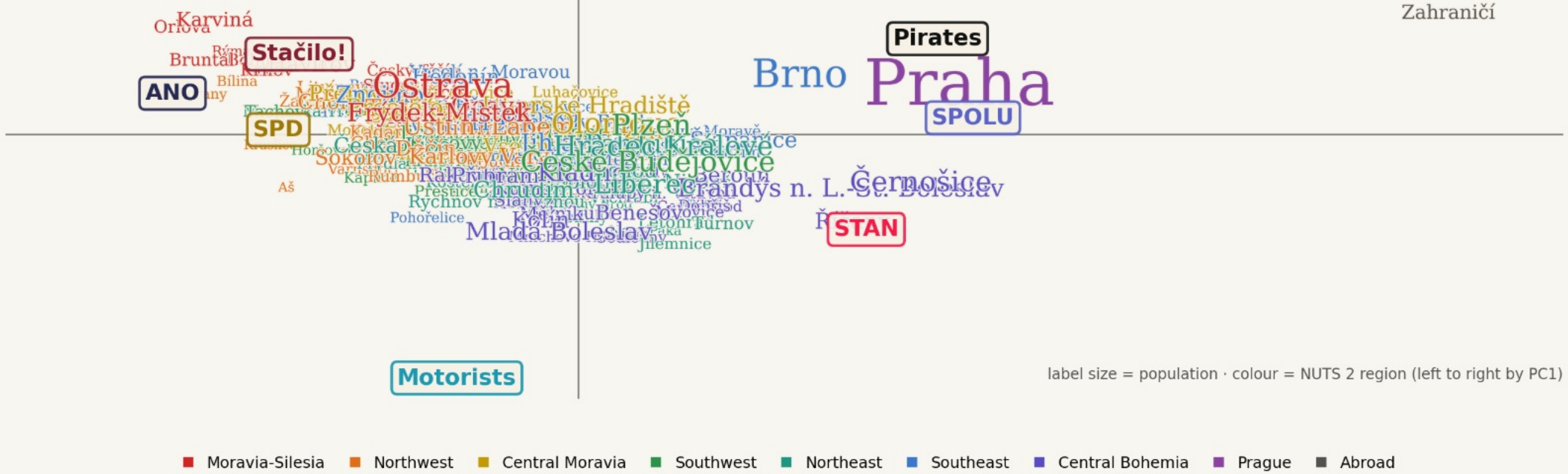
Karviná left, Prague right

– as the table suggested.

COMPARE

Prague and the Ostrava region at the ends

206 ORPs + votes abroad in the same PCA axes · label size = population · colour = NUTS 2 region · 2025

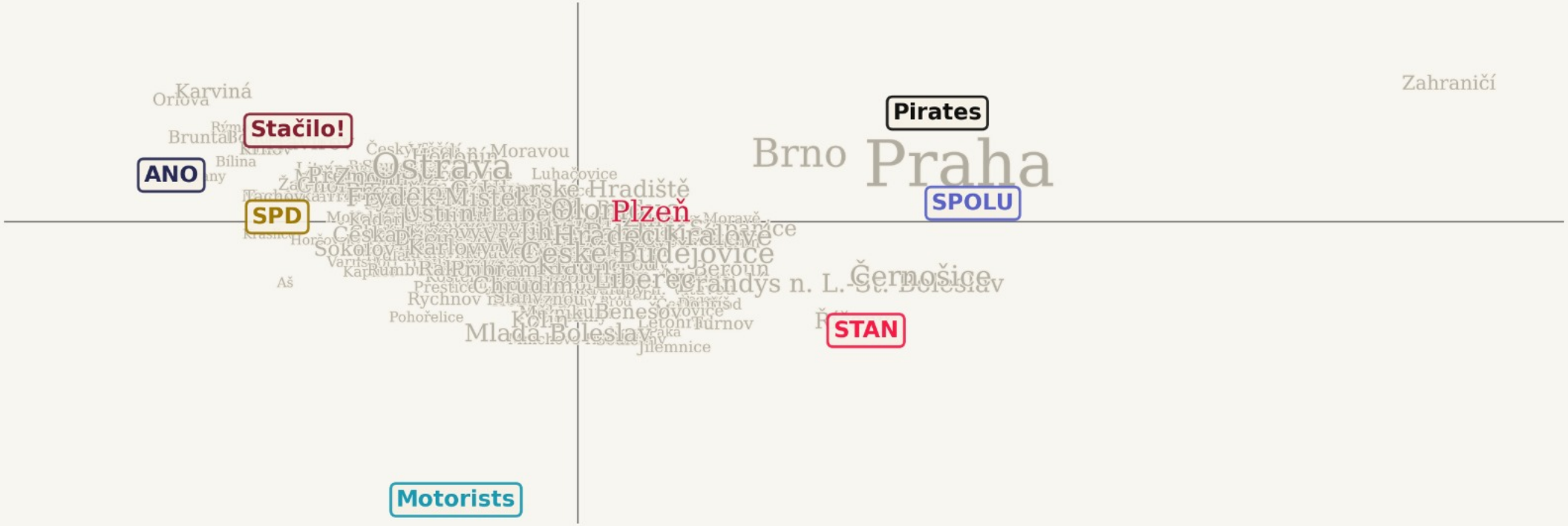


- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

QUIZ

Plzeň is one label. Where will its 10 city districts land?

Same axes · 206 ORPs in grey, Plzeň ORP in red · raise your hand: left, centre, right?

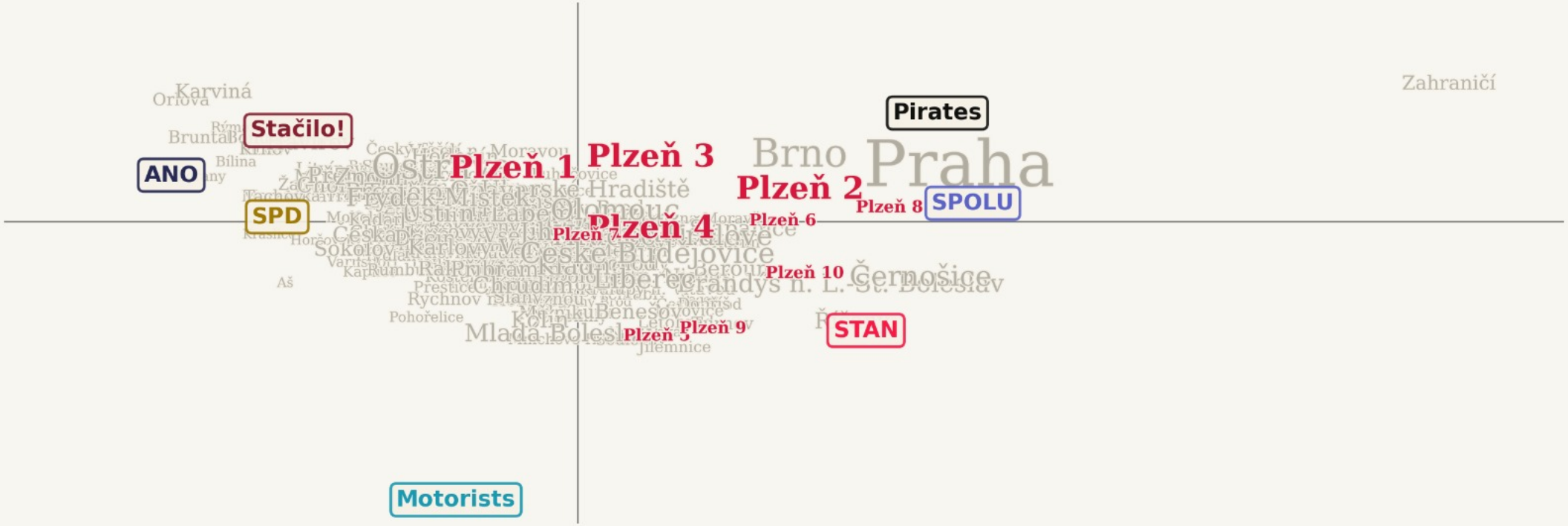


- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

QUIZ

The same city spreads from the centre towards SPOLU

Plzeň split into its 10 city districts · label size = valid votes



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

PART 3 OF 3

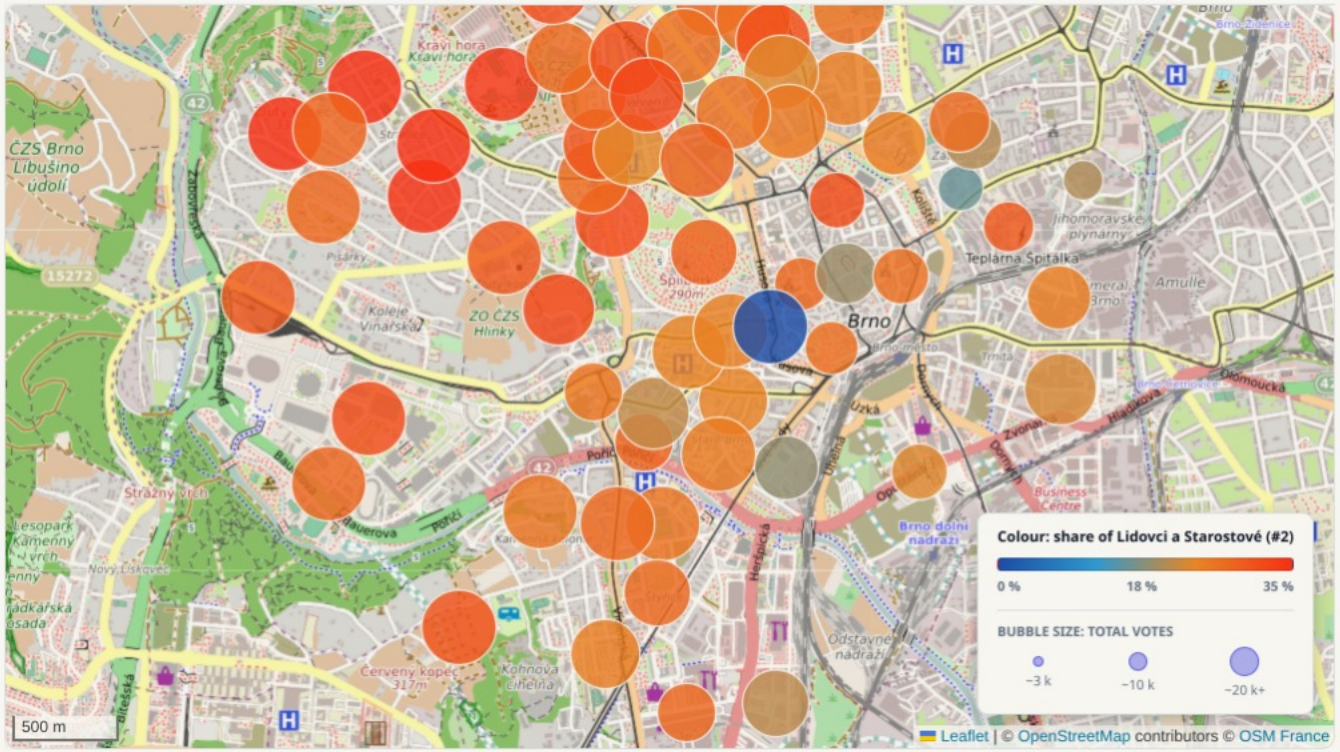
Check ⟶

Anomalies in a healthy system – and back to Most

CHECK

Zero is the simplest error: Brno-střed, station 23018

City council election 2026 · colour = share of list no. 2 (Lidovci a Starostové) · bubble size = total votes



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), OpenStreetMap contributors, own calculation

0 votes

recorded for the list that wins in the neighbouring stations (15–41 %, median 27 %). Model expects 2,700–3,300.

Special station

for people with an address at the town hall – no territory on the map, so different voters are possible.

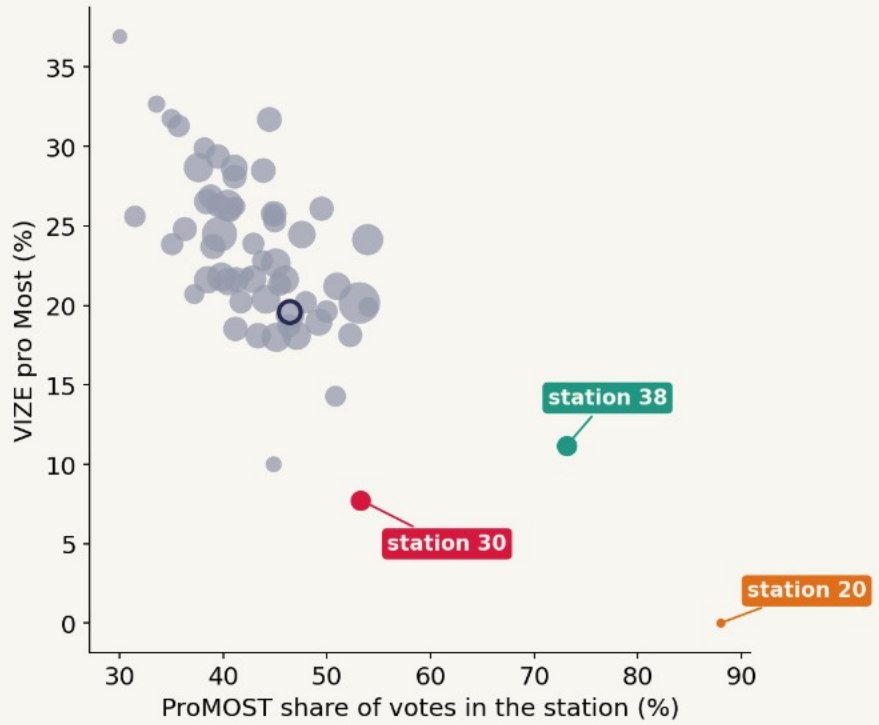
No seats change.

Still a number to check against the protocol.

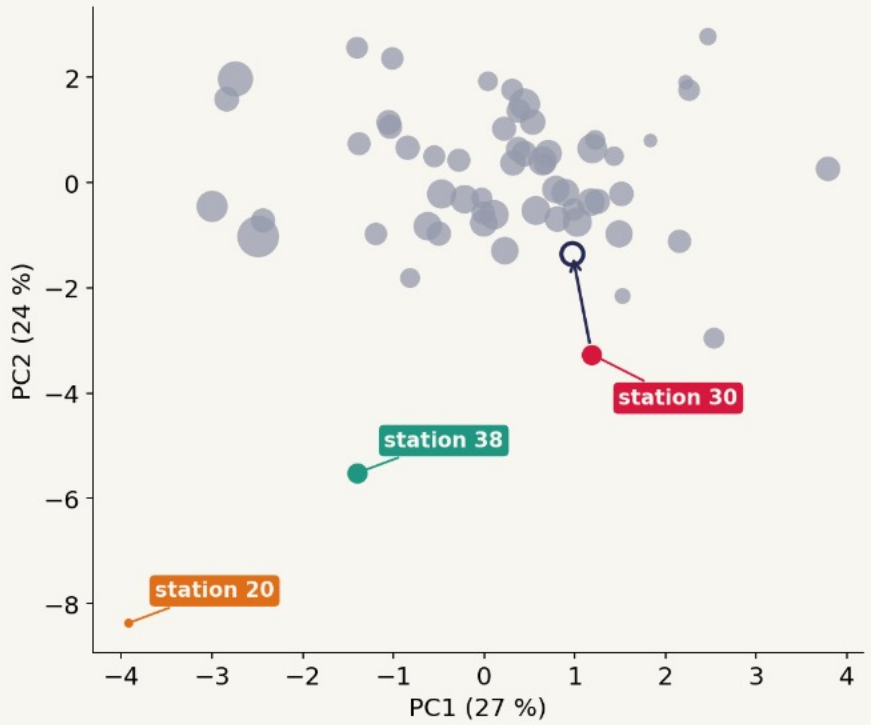
CHECK

Back to Most: plain PCA sees station 30 – but not alone

Most, 60 polling stations · red dot = station 30 as recorded, ring = with 1,786 votes · it ranks only 6th–9th by distance



- authors: Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

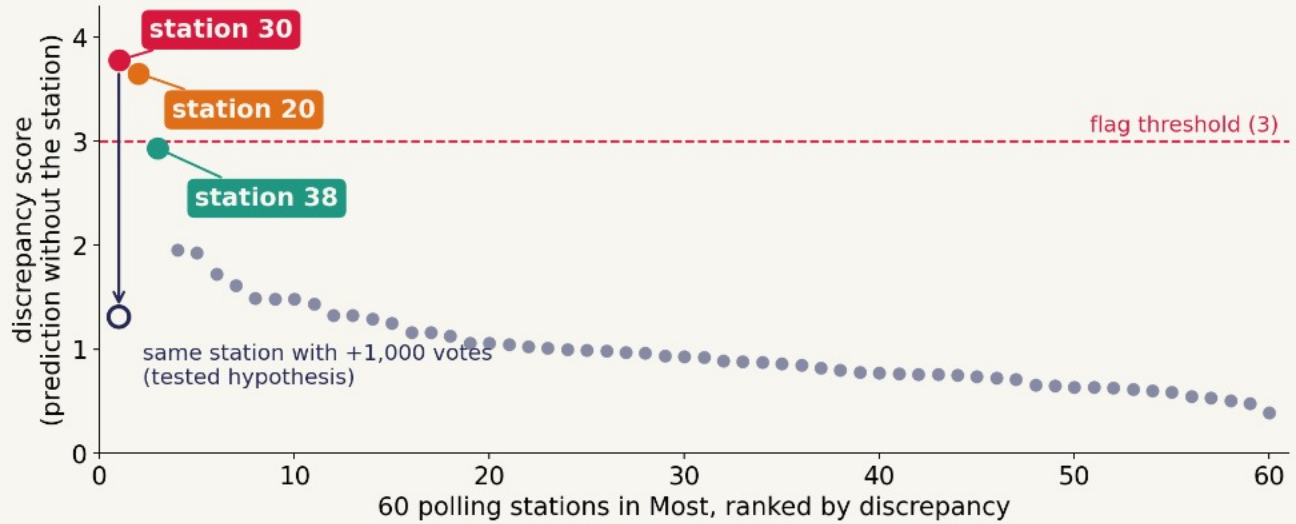


- authors: Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

CHECK

Predict each station without it, then compare

Most 2026 · discrepancy between the recorded numbers and a prediction made without the station itself



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), DataTimes.cz model, own calculation

Station 30:

highest of 60, all 4 model settings agree. With +1,000 votes: 1.3.

Station 20 = Chánov.

Turnout 8.6 % (median 31 %): an unusual place, not necessarily an error.

Station 38:

third, just under the threshold (2.9). ProMOST 73 %.

A flag is a hypothesis

to check against the protocol.

CHECK

Most: 616 recorded, 1,786 re-counted

VIZE pro Most · polling station 30 · Regional Court in Ústí nad Labem, no. 115 A 6/2026



• authors: Piráti, Zelení a Starostové (Instagram)
• data: Instagram @piratizelenistarostove (screenshot, 10 Nov 2026), court file 115 A 6/2026

Court recount.

The commission recorded 616 votes for VIZE pro Most; the court re-counted 1,786 – 1,170 missing.

5.001 % → 4.992 %.

Before: 7 votes above the threshold – less than one ballot (up to 45 votes). After: about 50 below – roughly **one full ballot**. Both seats lost.

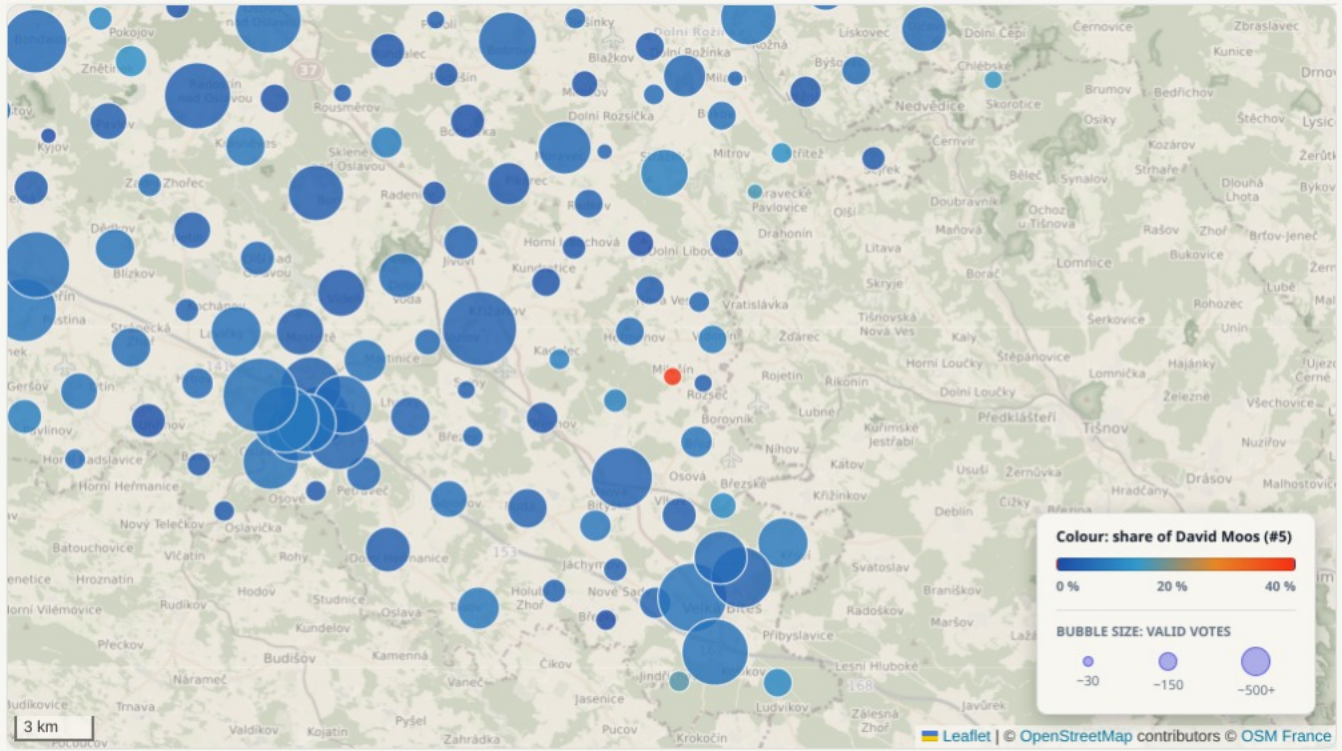
Status.

ČSÚ has not yet published corrected numbers.

CHECK

One red dot: Milešín, 13 votes for David Moos

Senate district 51 Žďár nad Sázavou · colour = share of David Moos in the polling station · bubble size = valid votes



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), OpenStreetMap contributors, own calculation

Milešín:

Moos 13 votes (39 %), neighbouring villages about 6 %. Everywhere else he is in the blue.

A possible swap 4 ↔ 13

between neighbouring ballot rows (Černý 4, Moos 13) would fit the neighbourhood and put Černý (11,372) ahead of Klement (11,373).

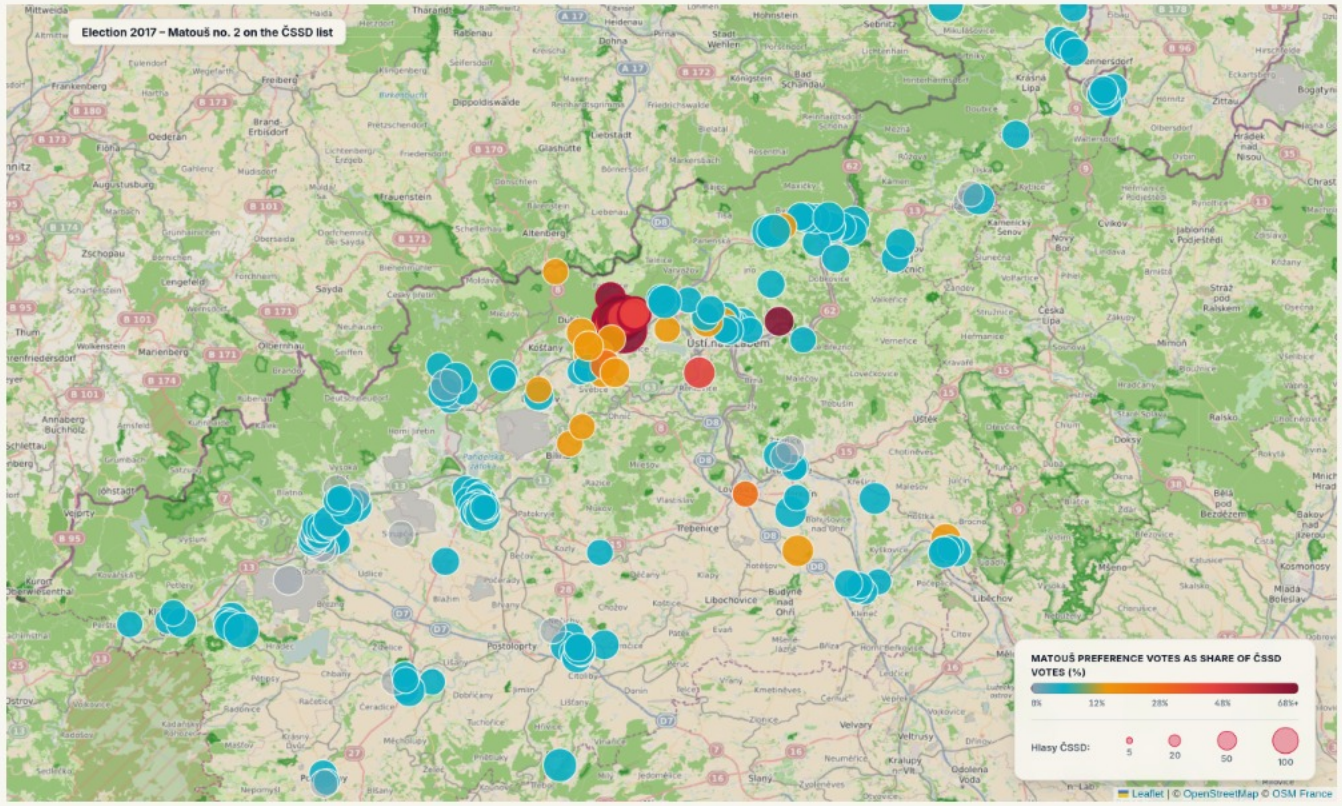
Hypothesis, not a finding.

Only a recount of the ballots from Milešín can decide.

POSSIBLE FRAUD, CZ

Ústí region 2017: where does one candidate's share spike?

ČSSD 2017 · colour = Zdeněk Matouš's preference votes as a share of ČSSD votes · size = ČSSD votes



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), OpenStreetMap contributors, DataTimes.cz (DataBites)



Zdeněk Matouš: Nebát se používat zdravý selský rozum

ČSSD

PhDr. Ing. Zdeněk Matouš, Ph.D. MBA, kandidát do Poslanecké sněmovny Parlamentu ČR za ČSSD

- authors: Homér
- data: Homér, 20 Oct 2017 (screenshot)

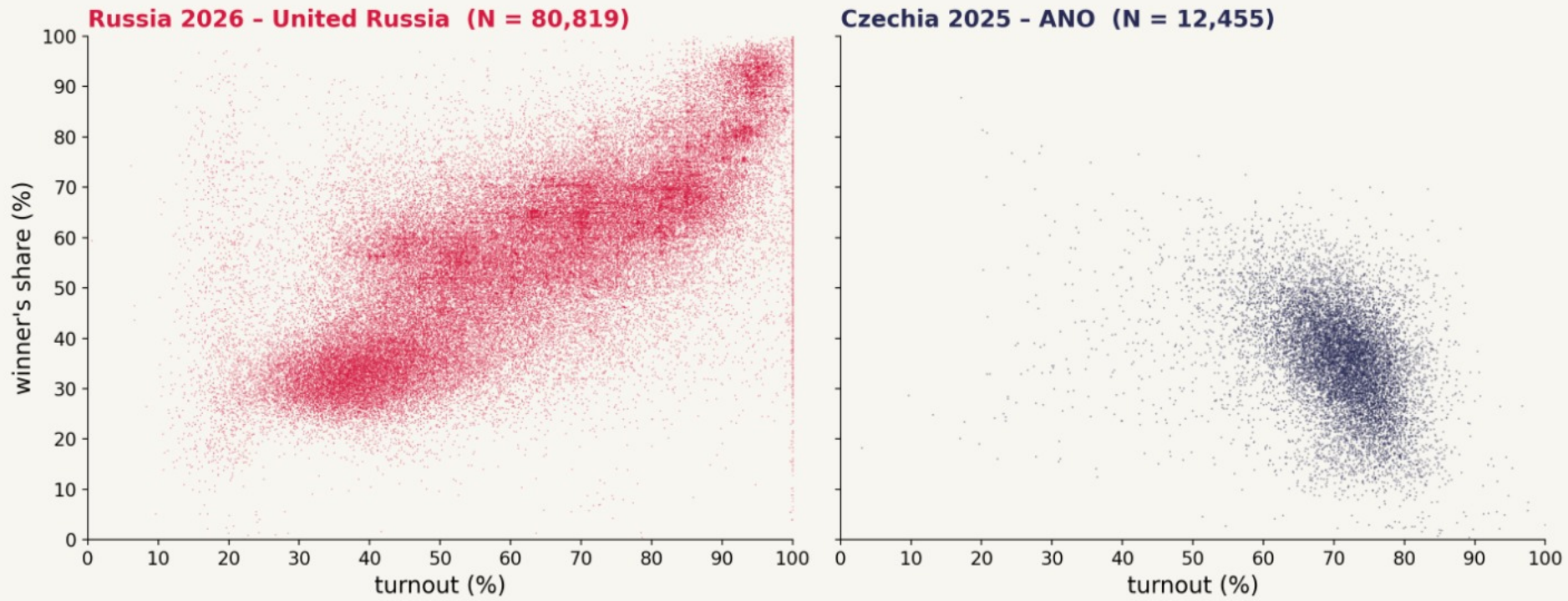
A pattern, not a proof.

Krupka is his own town (he is mayor) – look elsewhere.

RUSSIA

Two clouds, a corner at 100 / 100, and lines at round numbers

Winner's share vs turnout · every dot a polling station · identical axes

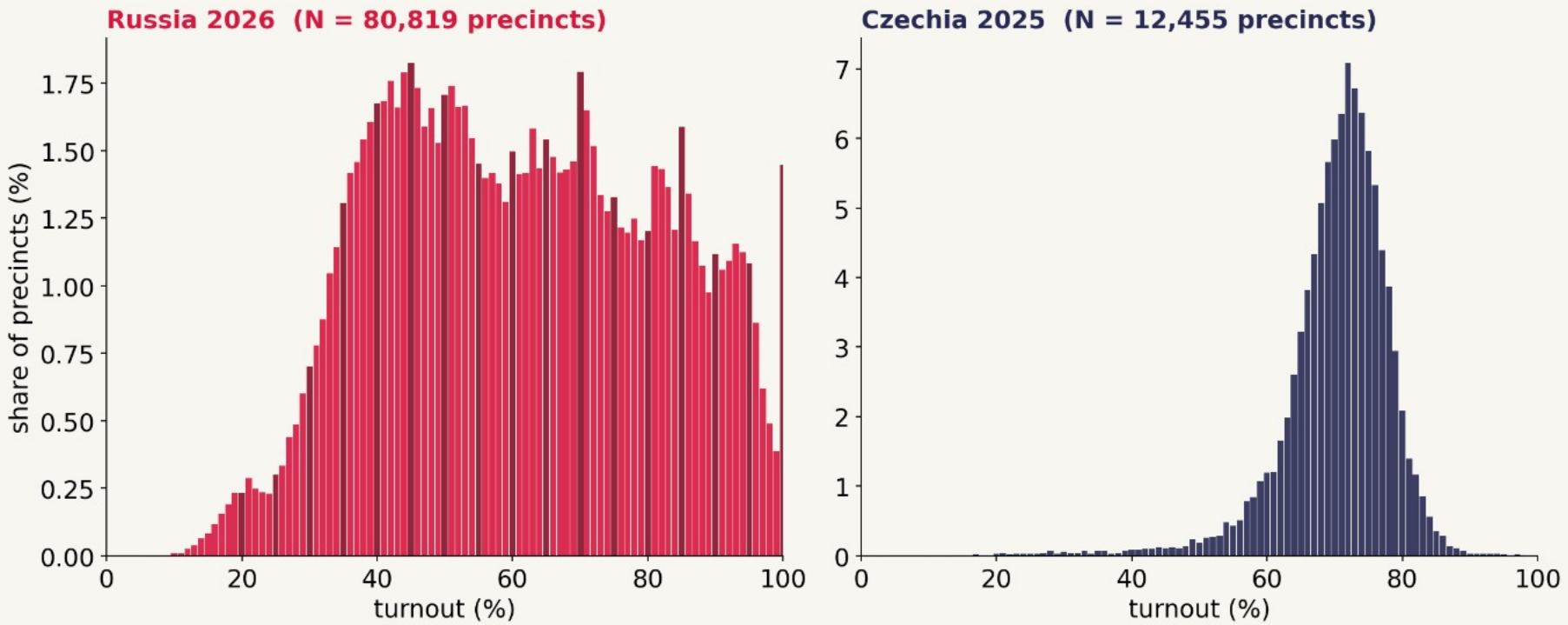


- authors: Kateřina Mahdalová & Michal Škop
- data: Russian CEC (2026 Duma), Czech Statistical Office, own calculation

RUSSIA

Round numbers leave fingerprints

Share of polling stations by turnout · 1 pp bins · dark bars = multiples of 5 (50, 55, … 100)



- authors: Kateřina Mahdalová & Michal Škop
- data: Russian CEC (2026 Duma), Czech Statistical Office, own calculation

VENEZUELA

No station-level results, nothing to test

Earlier elections came with polling-station detail – by 2023 the CNE publishes only totals



	Votos SI	Votos NO	Votos Válidos	Votos Nulos	Total
Pregunta 1	10,315,485 (98.00%)	210,329 (2.00%)	10,525,814 (99.72%)	29,278 (0.28%)	10,555,092
Pregunta 2	10,322,718 (98.27%)	181,725 (1.73%)	10,504,443 (99.52%)	50,649 (0.48%)	10,555,092
Pregunta 3	10,132,099 (96.35%)	384,321 (3.65%)	10,516,420 (99.63%)	38,672 (0.37%)	10,555,092
Pregunta 4	10,134,189 (96.37%)	381,770 (3.63%)	10,515,959 (99.63%)	39,133 (0.37%)	10,555,092

- authors: CNE Venezuela
- data: cne.gob.ve, Referendo Consultivo 2023 (screenshot, updated 6 Dec 2023)

Before.

Venezuelan results used to come with detail down to the polling table.

2023 referendum.

Filters for state … table are on the page, but the menu is empty – only totals: 98 % "yes", 51 % participation.

2024 presidential election.

A winner declared, no results by polling station (Carter Center). Only the opposition's tally sheets exist – validated by the Carter Center, not CNE data.

THE OPPOSITE: BRAZIL

A ballot box on a boat – and every box published

Brazil: results available per electronic ballot box, within hours, for everyone



- authors: TRE-AC (Brazilian electoral court, Acre)
- data: tre-ac.jus.br, Dia da Democracia, Sept 2023 (photo)

At the station.

Each ballot box prints a signed result sheet, posted on the spot.

Same evening.

Sheets appear on the TSE portal – by state, municipality, zone and section.

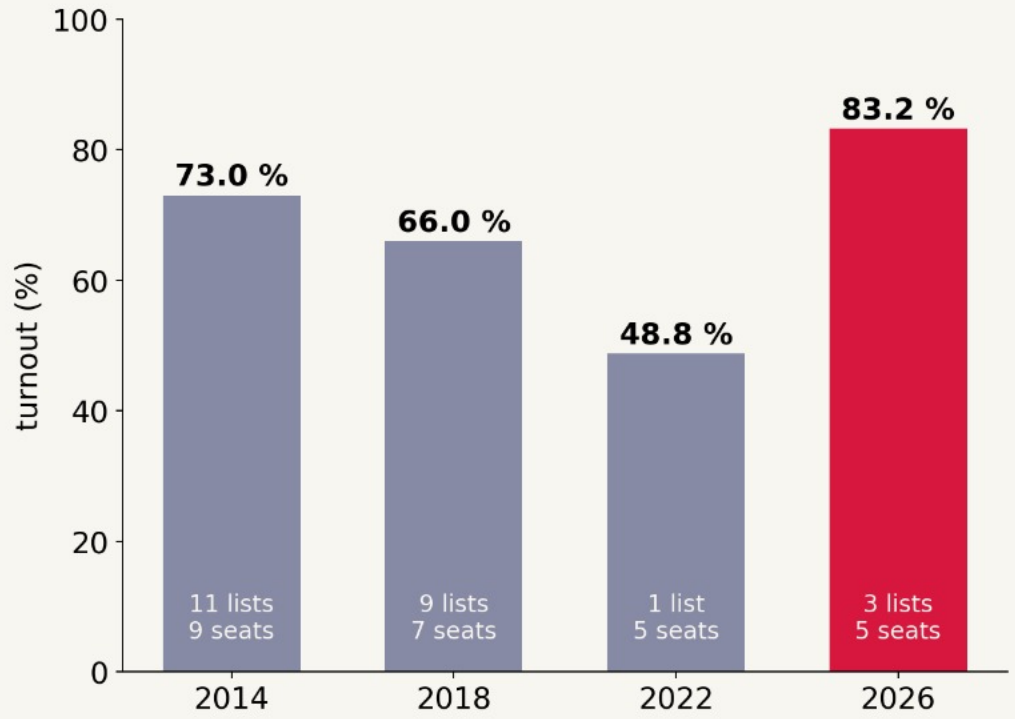
Open data.

All sheets downloadable – the raw material for every test here.

BLIND SPOT, CZ

Where statistics is blind: Vysoká u Příbramě

Municipal election 2026 · one polling station, 298 voters, 5 council seats · turnout by election



- authors: Kateřina Mahdalová & Michal Škop
- data: Czech Statistical Office (volby.cz), own calculation

Nothing to compare.

One precinct – our model had no other stations to predict from.

The numbers add up.

Sums, seat allocation, vote capacity: no arithmetic problem. Turnout 83 % (2022: 49 %) is unusual – and only descriptive.

The court decided.

Regional Court in Prague (37 A 65/2026) invalidated the voting on 95 sworn voter statements; it did not establish manipulation of the urn or ballots (ČTK, 10 Nov 2026).

SUMMARY

Every test has power against one kind of error

METHOD	CATCHES	BLIND TO
Leave-one-out prediction	Typos, missing digits, swaps	Coordinated shifts across many stations
Turnout × share scatter	Ballot stuffing, administrative pressure	Honest regional heterogeneity (false alarms)
Round-number histogram	Reported-to-quota protocols	Carefully randomised fraud
2nd-digit Benford	Fabricated numbers (sometimes)	Most manipulation – United Russia passes

TAKEAWAYS

Three things to remember

1. Anomaly ≠ fraud – a flag is a question for the protocol.
2. A test that passes proves nothing – know what it is blind to.
3. Detailed, public results are the first safeguard.

SOURCES

Sources

Our analyses and data

- Czech Statistical Office: volby.cz (2025 Chamber of Deputies; 2026 municipal and Senate, XML feeds)
- DataTimes.cz / DataBites: databites.datatimes.cz (2026 anomaly dashboard; Matouš preference votes)
- Mahdalová & Škop: US 2026 House maps, mahdalova-skop.cz; NYT and Civic API results, Daniel Donner / The Downballot (CC BY 4.0), US Census Bureau
- Central Election Commission of Russia, 2026 Duma: precinct data (own database)
- OpenStreetMap contributors (base maps)

Cases and courts

- Most: Regional Court in Ústí n. L., 115 A 6/2026 (infodeska.gov.cz); ceska-justice.cz, 3 Nov 2026
- Vysoká u Příbramě: Regional Court in Prague, 37 A 65/2026; ČTK / České noviny, 10 Nov 2026
- Venezuela: Carter Center, final report 2025 and statement 30 Jul 2024; MOE Colombia, tally-sheet review, 3 Aug 2024; CNE, Referendo Consultivo 2023
- Brazil: TSE (tse.jus.br, boletim de urna; dadosabertos.tse.jus.br); photo TRE-AC, Sept 2023
- Homér, 20 Oct 2017; Instagram @piratizelenistarostove

Election forensics literature

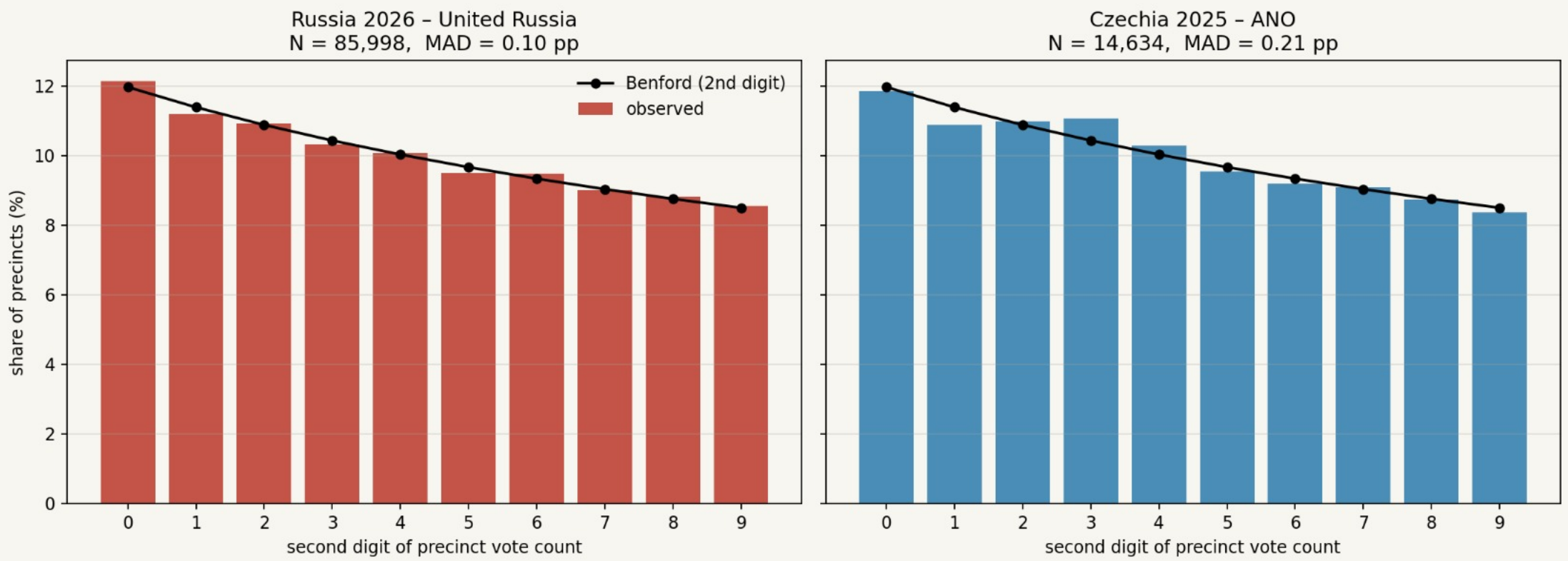
- Klimek, Yegorov, Hanel, Thurner (2012). Statistical detection of systematic election irregularities. PNAS 109(41)
- Kobak, Shpilkin, Pshenichnikov (2016). Integer percentages as electoral falsification fingerprints. Ann. Appl. Stat. 10(1)
- Myagkov, Ordeshook, Shakin (2009). The Forensics of Election Fraud: Russia and Ukraine. Cambridge UP
- Mebane (2008). Election forensics: the second-digit Benford's law test. In Alvarez, Hall, Hyde (eds.), Election Fraud. Brookings
- Deckert, Myagkov, Ordeshook (2011). Benford's law and the detection of election fraud. Political Analysis 19(3)
- Rozenas (2017). Detecting election fraud from irregularities in vote-share distributions. Political Analysis 25(1)

Tools and data

- github.com/dkobak/elections – cleaned Russian precinct data 2000–2024
- R packages eforensics, EFToolkit (Mebane, Kalinin et al.)
- Golos: golosinfo.org – independent observers

BACKUP

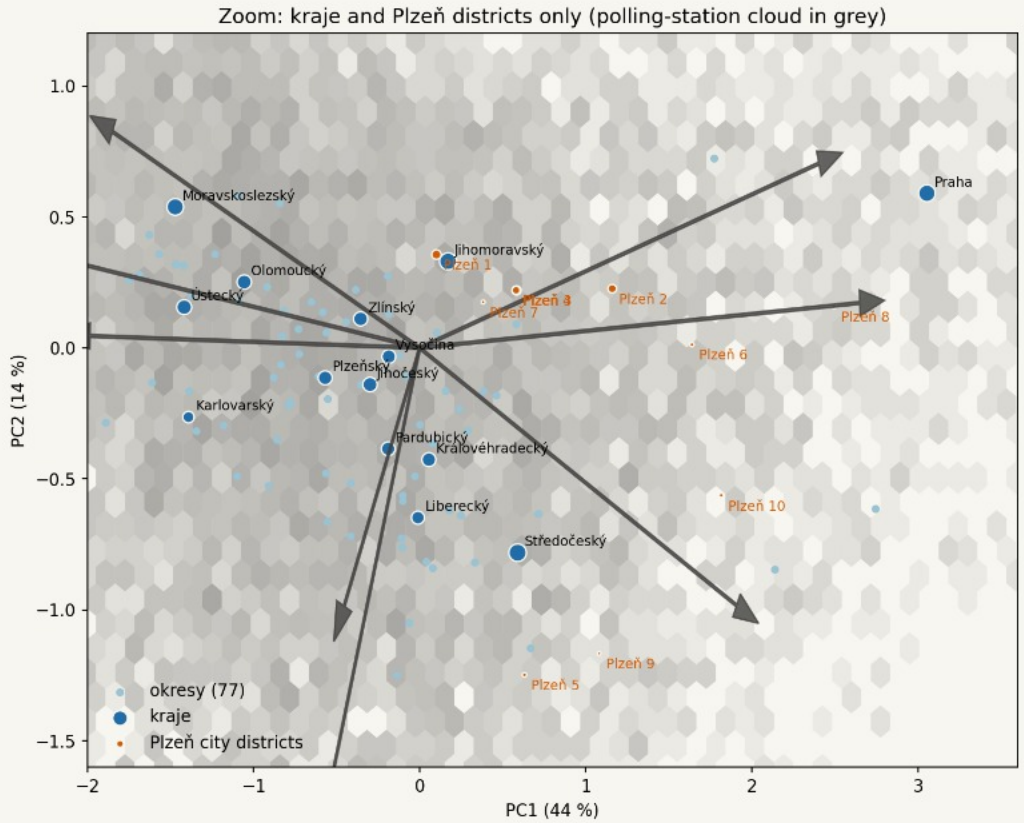
Second-digit Benford: a pass proves nothing



- authors: Michal Škop
- data: Russian CEC, Czech Statistical Office, own calculation

BACKUP

PCA zoom: kraje and okresy



- authors: Michal Škop
- data: Czech Statistical Office, own calculation