Dmitry Kobak is a research scientist at Tübingen University, Germany.

Suspect peaks in Russia's "referendum" results

Dmitry Kobak, Sergey Shpilkin and **Maxim S. Pshenichnikov**

find suspicious patterns in data from Russia's nationwide vote on a package of constitutional changes

Evgeny Feldman

Left: One of many makeshift polling stations set up throughout Russia to allow people to vote in the July 2020 plebiscite, which took place during the Covid-19 pandemic.

On 1 July 2020, a plebiscite was held in Russia concerning 206 constitutional amendments. Voters could only vote yes to all amendments or no to all amendments. Yes won. One amendment allows Vladimir Putin to run for president two more times, potentially remaining leader until 2036.

This was not a normal vote. It did not technically qualify as a referendum under Russian constitutional law, so it was not held according to referendum laws. Instead, it was conducted under a set of *ad hoc* rules, ostensibly to reduce the risk of Covid-19 infection. The rules meant that the vote lasted an entire week rather than a single day, with many makeshift polling stations set up in the streets and within the grounds of private and public organisations (bit.ly/2CLY8bA). Experts warned that in such unusual circumstances, the integrity of the vote was at risk (bit.ly/31lL3iV). Falsifications of voting records would be particularly easy to perform and difficult to prevent, they said, as independent observers would find it challenging to monitor the conduct of the vote.

Such warnings appear to have been justified. Our analysis of returns from polling stations finds an unusually high number of neat, round percentages voting in favour of Putin's amendments, which we interpret as evidence of vote manipulation.

We, and others, have previously argued that if integer percentages, such as 85.0%, are more frequent across polling stations than non-integer percentages, such as 85.3%, then this can only plausibly be explained by fraud – with multiple polling stations having forged the results in order to achieve a certain percentage.[1–4]

Importantly, in Russia, it is the number of ballots, and not the percentage, that is reported by polling stations. So, a polling station with 1,020 collected ballots aiming to report an 85.0% yes vote would need to report precisely $1{,}020 \times 0.85 = 867$ yes ballots. The number 867 is not remarkable in itself; it is only after division by 1,020 that it yields an integer percentage. If many polling stations aim at the same integer percentage, this will produce an excess of integer percentages across the country, leading to noticeable spikes in the number of stations reporting these integer percentages when graphed.

Such spikes were indeed very prominent in the histograms of voter turnout and yes vote percentages across all 96,765 polling stations (Figures 1a and 1b). The spikes occurred at almost every integer percentage above ~70% turnout and ~75% voting yes. In a two-dimensional scatter plot of polling stations, with turnout percentage plotted against yes vote percentage, a square grid-like pattern appears at integer percentages, particularly prominent at multiples of 5% (Figure 1c). The two-dimensional distribution forms two clusters: one centred at ~43% turnout and ~65% yes, and another at 70% or more turnout and 75% or more yes. It is only the latter cluster that exhibits the grid-like pattern, suggesting that this cluster has a malignant origin. Because of the pronounced bimodality of the distribution, the official vote outcome (67.9% turnout, 78.6% yes) corresponds to a point in a minimum of the density between the two clusters (the black cross in Figure 1c).

As in previous Russian elections, many integer spikes can be traced down to individual

Sergey Shpilkin is a technical translator and writer, and data journalist and election analyst in Russia.

Maxim S. Pshenichnikov is a professor of physics at the University of Groningen, the Netherlands.

Photo © Sylvia Germes

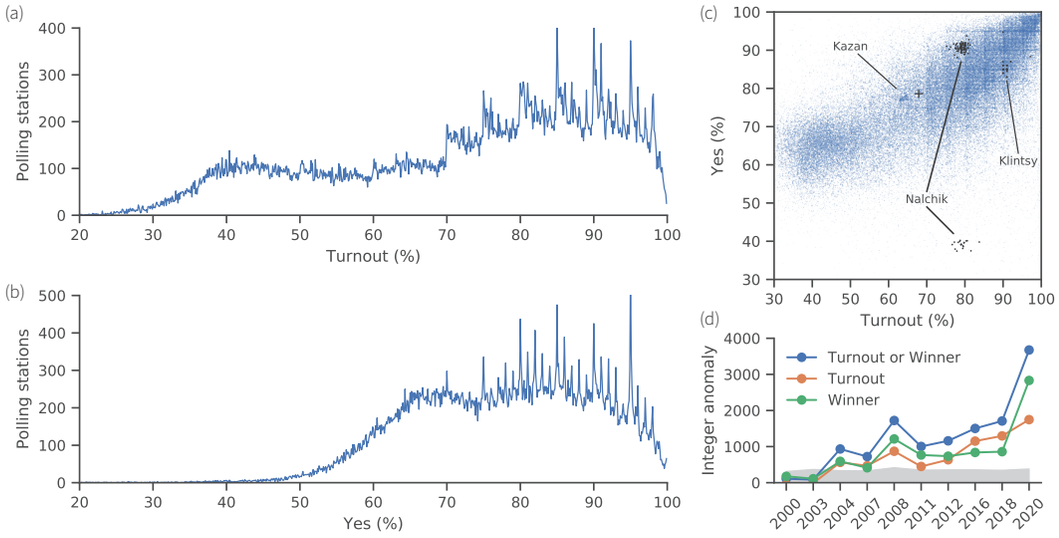

Figure 1: (a) Turnout histogram. We used 0.1% bins (centred in such a way that the value at, for example, 50.0% corresponds to 50.0 ± 0.05%). To avoid any artefacts due to integer division, we added a random number sampled from the uniform distribution $\boldsymbol{U}(-0.5, 0.5)$ to the numerator of each fraction. Polling stations with 100% turnout were excluded from this figure. (b) Yes vote share histogram, constructed as above. (c) Scatter plot of all polling stations. The black cross marks the official totals. Several regional oddities are highlighted. (d) Integer anomaly for all Russian federal elections from 2000 to 2020. The anomaly is defined as the excess of polling stations with integer turnout (orange), winner's result (green), or either of the two (blue) relative to the expected values from binomial Monte Carlo simulations. Grey shading shows the interval up to the 99.9th percentile of the Monte Carlo values (for the joint anomaly shown in blue), corresponding to $p < 0.001$ for all values above it.

federal regions, cities, or constituencies. Several such cases are highlighted in Figure 1c. The polling stations in the small town of Klintsy in western Russia reported almost exclusively either 90.0% or 91.0% turnout, with odd-numbered polling stations all reporting 90.0% and even-numbered ones reporting 91.0% (bit.ly/3lCbIQi). The city of Nalchik had most of its polling stations reporting ~80% turnout with ~90% yes votes, except for a small number of randomly located stations that had ~40% yes votes (which is in the lowest percentile country-wide) but the same ~80% turnout, resulting in a strongly bimodal distribution. The city of Kazan, meanwhile, formed its own conspicuous, remarkably tight cluster with ~65% turnout and ~77% yes votes, making it look as if just over one half of all registered voters in Kazan voted yes (0.65 × 0.77 = 0.50). Such a result was speculated to be the outcome desired by the Kremlin for the country as a whole (bit.ly/31ggc6Z).

After the plebiscite, Putin's press secretary said that this was "a *de facto* triumphant referendum on trust in Putin" (reut.rs/2QdNHAF). Our analysis suggests that the triumph was largely staged. We have previously defined the *integer anomaly* as the excess of polling stations reporting integer percentages relative to binomial Monte Carlo simulations.[1,3] The integer anomaly for this referendum totalled 3,670 stations, setting a new record among all Russian federal elections in the Putin era (Figure 1d) and providing a grim outlook for the future of the country's electoral system. ■

As in previous elections, many integer spikes can be traced down to individual federal regions, cities, or constituencies

Note

Raw data and analysis code are available at github.com/dkobak/elections.

References

1. Kobak, D., Shpilkin, S. and Pshenichnikov, M. S. (2016) Integer percentages as electoral falsification fingerprints. *Annals of Applied Statistics*, **10**(1) 54–73.
2. Kobak, D., Shpilkin, S. and Pshenichnikov, M. S. (2016) Statistical fingerprints of electoral fraud? *Significance*, **13**(4), 20–23.
3. Kobak, D., Shpilkin, S. and Pshenichnikov, M. S. (2018) Putin's peaks: Russian election data revisited. *Significance*, **15**(3), 8–9.
4. Rozenas, A. (2017) Detecting election fraud from irregularities in vote-share distributions. *Political Analysis*, **25**(1), 41–56.

Digital detectives

Picture the scene: a crime scene. There's a victim and a discarded cell phone. Who does the phone belong to? Perhaps the pattern of user events logged on the device can offer some clues.

That's the idea put forward by Christopher Galbraith and colleagues from the University of California, Irvine.[1] They used time-stamps of web-browsing events from a selection of students, split into Facebook and non-Facebook usage. Not only did they see different time patterns of usage for each student, they also found that general web usage and Facebook usage tended to coincide. This meant they were able to statistically distinguish pairs of time-stamps that came from the same students from those that came from different students. Tests on other data sets led to similar results.

What does this mean for law enforcement? Not much at the moment, says Galbraith. But if there are relationships between different types of digital events – web and social media use, or email, say – and there is a log of events from an unknown user on a recovered device, it may be possible to compare this record against a database of activity from known users in order to identify possible owners of the device. ■

Reference

1. Galbraith, C., Smyth, P. and Stern, H. S. (2020) Quantifying the association between discrete event time series with applications to digital forensics. *Journal of the Royal Statistical Society, Series A*, **183**, 1005–1027.