Descriptive Finding

War and mobility: Using Yandex web searches to characterize intentions to leave Russia after its invasion of Ukraine

Athina Anastasiadou
Artem Volgin
Douglas R. Leasure

© 2024 Athina Anastasiadou, Artem Volgin & Douglas R. Leasure.

This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.
See https://creativecommons.org/licenses/by/3.0/de/legalcode.
## Contents

1 Introduction 206

2 Data 206
   2.1 Yandex Wordstat 206
   2.2 Data triangulation 209

3 Drivers of mobility intentions 212
   3.1 Geographic preferences reflected in the search trends 213
   3.2 Socioeconomic correlations with the search trends 215

4 Discussion and conclusion 216

5 Acknowledgements 217

   Bibliography 219
War and mobility: Using Yandex web searches to characterize intentions to leave Russia after its invasion of Ukraine

Athina Anastasiadou¹
Artem Volgin²
Douglas R. Leasure³

Abstract

BACKGROUND
Following the Russian invasion of Ukraine, many citizens left Russia due to increasing government repression, the fear of mobilization, or to escape the economic downturn. As yet, reliable statistical data on those who left are not available. Hence, much remains unknown about the characteristics and scope of this population. In the digital age, people prepare their journeys by searching online and these digital traces provide clues about their intentions and the scale of mobility events.

OBJECTIVE
Here we aim to leverage this resource of search engine data to study geographic, temporal, and demographic trends in the international mobility intentions of Russian residents.

METHODS
Our analysis combines search queries provided by the search engine Yandex with city-level data on sociodemographic and geographic characteristics. We examine two mobility events, the first after the start of the invasion and the second after mobilization started. We explore the relationships between cities’ sociodemographic characteristics and the number of mobility-related searches during those events using a set of generalized linear models.

¹ Max Planck Institute for Demographic Research, Rostock, Germany; Netherlands Interdisciplinary Demographic Institute – KNAW and University of Groningen, the Netherlands. Email: anastasiadou@demogr.mpg.de.
² University of Manchester, UK.
³ University of Oxford, UK.
CONCLUSIONS
We find that regional indicators of wage levels were less important in the second wave and the proximity of a country of interest increased fourfold in importance. These and other results support dominant narratives in the media about a potential Russian brain drain, particularly during the first wave, and about the scale of the outflow.

CONTRIBUTION
This work combines spatial data on sociodemographic characteristics with search engine data to understand the mobility patterns of Russian nationals from Russia after the invasion of Ukraine. Our findings confirm the potential of Yandex search engine data as a novel data source for informing migration and mobility research.

1. Introduction

The Russian invasion of Ukraine not only caused a mass outflow of refugees from Ukraine but also triggered a notable outflow of its own citizens from Russia. Our empirical knowledge about the so-called ‘Russian exodus’ reported in the media is fragmented, as official data and reliable estimates are missing. Estimates of the number of people leaving Russia due to increasing repression, sanctions, and military mobilization amount to several hundred thousand and vary across sources (Kasyanchuk 2022). The Russians that left the country immediately after the start of the invasion are assumed to be from the most highly educated and highly skilled section of the country’s population (Demytrie 2022). Meanwhile, the outflow experienced after mobilization is believed to consist of less affluent and educated men, many intent on escaping military service (Light 2022). Amidst these recent developments, we aim to corroborate or refute common media narratives by employing computational methods and by using digital traces of international mobility intentions from Yandex web searches across Russian cities.

2. Data

2.1 Yandex Wordstat

Our choice of data is rooted in the theory of planned behavior, which proposes that the best predictor of a specific behavior is the intention to perform that behavior (Willekens 2021). We observed mobility intentions through online search activity, and we used this to model relative mobility behaviors, hinging on the assumption that actual mobility is a
combination of mobility intentions and the capability to move (Carling and Schewel 2018). Although online searches can only reflect intentions to move, migration aspirations are necessary for actual international migration flows (Tjaden et al. 2019).

While most papers establishing a link between search results and international mobility focus on Google Trends (Avramescu and Wiśniowski 2021; Böhme, Gröger, and Stöhr 2020), the context of our analysis allows for the use of Yandex Wordstat. Yandex is the most popular search engine in Russia (62.1% market share) followed by Google (36.7%) (Statista n.d.). It provides weekly as well as monthly absolute and relative search counts at the sub-national level in Russia and some other countries (Yandex Wordstat n.d.). We collected counts of mobility-related words for 48 of the most likely destination countries and their capital cities, which include most European Free Trade Association members and neighboring countries to Russia (see complete list in the supplementary materials (Anastasiadou, Volgin, and Leasure 2023)). Many of these countries are popular Russian tourist destinations and some have relatively relaxed entry regulations for Russian citizens. For the analysis we obtained weekly data for the year starting December 20, 2021 and ending December 18, 2022, and for comparison with other data sources we used monthly data from December 2020 to November 2022. Monthly data were obtained in order to extrapolate the weekly time series to a period of two years. We compiled a list of mobility-related queries (see lists below) based on previous work that uses search results for migration prediction (Avramescu and Wiśniowski 2021; Böhme, Gröger, and Stöhr 2020). Only the terms that exhibited an obvious spike around the start of the invasion were included in the sample. This list is not exhaustive, but additional mobility-related keywords did not influence our results due to their low search incidence. Although these terms were designed to reflect intentions to change residence, we cannot definitively rule out tourism as a motive.

**Query keywords:** move, work, residence permit, permanent residence, rent, flight, accommodation, (to) rent, flat, citizenship, real estate, visa, car (see query keywords in Russian language in the supplementary materials (Anastasiadou, Volgin, and Leasure 2023)).

**Query destinations:** Armenia, Azerbaijan, Egypt, Georgia, Israel, Kazakhstan, Kyrgyzstan, Mongolia, Montenegro, Serbia, Tajikistan, Thailand, Turkey, United Arab Emirates, Uzbekistan, and the majority of European Free Trade Association (EFTA) countries. 4

To collect the data, each mobility-related word was paired with each potential destination country and the number of web searches from each region was recorded. Data were available for all 83 Russian regions and for 388 cities and towns at the municipal

---

4 All countries of the EFTA are included in the analysis except for Denmark, Iceland, Liechtenstein, Luxembourg, and Malta (for details please refer to the lists in the supplementary materials).
level (Volgin and Anastasiadou 2023). We assumed that Russian citizens prepared for their move within their home country and performed searches in the Russian language.

Figure 1 shows the time trend of the share of mobility-related queries in regard to all the countries in our data set. Clear spikes in relevant searches are visible shortly after the start of the invasion and right after mobilization efforts began, indicating increased interest in leaving the country.

**Figure 1: Mobility-related internet searches across Russia**

![Graph showing the time trend of mobility-related internet searches across Russia.](image)

*Note: Mobility-related searches (share of total weekly searches) performed at the city level (grey lines) in Russia and their average (black line). Red dots indicate February 24, 2022 (beginning of the full-scale invasion) and September 21 (beginning of mobilization).*

The same pattern can be observed spatially across cities and towns. Figure 2 depicts the difference in search behavior between cities and towns during the three-week periods after the start of the invasion and after the announcement of mobilization. The map focuses on the western part of the country since urban centers are mainly located there. It is also notable that the increase in mobility-related search behavior is greater after the announcement of military mobilization than in the three-week period following the invasion.
Figure 2: Geographic variation in mobility-related web searches for various destinations

Note: Geographic variation in the number of mobility-related web searches and links between origin cities and potential destination countries for the three-week periods following the Russian invasion of Ukraine (February 2022) and the announcement of military mobilization (September 2022). The size of the red circles indicates the overall increase in the number of mobility-related searches for each city, and the darkness of the blue lines indicates the increase in the number of searches from a city that specified each potential destination country. The increase is calculated as the difference between the number of queries three weeks after and three weeks before invasion or mobilization.

2.2 Data triangulation

As Yandex searches are not yet an established data source for international mobility intentions, we compared them to other innovative and official data sources. We found high positive correlations and similar patterns, allowing us to conclude that relying on Yandex searches for the purpose of this study was justified.

It is important to note that Yandex search data come with a number of limitations, similar to Google Trends. Like any other digital trace data type, online search data do not represent the behaviors and preferences of the entire Russian society or of the people performing the searches. Nor do they indicate absolute levels of mobility, as queries do not necessarily translate into moves and a person can undertake several searches without ultimately migrating. Even though our approach benefits from the subnational availability of Yandex search data, the sparsity and null-inflation of the search counts across regions pose a threat to the analysis. Additionally, the collection and production of the data is not fully transparent and thus we cannot make any claims about the quality of the data generating processes.
In Figure 3, the plot on the right-side shows the correlation of the Yandex data with the more established Google Trends Index. This index has already been used for scientific purposes and has shown to be a reasonable predictor of international and internal migration flows (Avramescu and Wiśniowski 2021; Böhme, Gröger, and Stöhr 2020). By cross-referencing Yandex searches with the Google Trends Index, we found a correlation coefficient of 0.61 (Google Trends n.d.). While Google Trends offered a longer search history, spanning several years compared to Yandex’s two years (Yandex Wordstat n.d.), Yandex provided finer data at the sub-regional level. Our preference for Yandex stems from our focus on Russia-based, Russian-language searches, despite Google Trends’ broader international scope. Yandex’s advantage lies in offering raw absolute search counts, unlike Google Trends’ indexed data. This choice prioritized data precision and spatial granularity, albeit with some trade-offs in time and coverage aspects. Even though Google Trends is more established for migration research, it has similar measurement, representativeness, and coverage limitations as Yandex data.

In Figure 3, the left plot displays the correlation between Yandex searches for destination countries and Russian departures to these countries from Federal Security Service of Russia data (Federal Security Service of the Russian Federation n.d.). Quarterly data may obscure short-term spikes, and popular transit destinations could skew destination preferences. Even after removing countries without direct travel options, the correlation remains high at 0.74. For more insightful comparisons, monthly border crossing data would be valuable if available, particularly when disaggregated by subnational origin locations. The monthly visitor statistics of Georgia, Israel, and Turkey for 2022 (Ministry of Culture and Tourism of Turkey 2022; Ministry of Internal Affairs of Georgia 2022; Ministry of Tourism of Israel 2022) yield a 0.9 correlation with mobility-related searches (middle plot in Figure 3). However, these comparisons should be treated with caution because visitor data is only an indicator of movement and provides no information regarding duration of stay or underlying intentions.
Figure 3: Correlations between Yandex web searches and other mobility indicators

Note: Correlations between total mobility-related searches on Yandex and departures of Russians abroad, border crossings into selected destination countries, and respective Google Trends Indices indicate that Yandex web searches are a reasonable proxy for measuring migration intentions, similar to Google Trends. The left plot shows the number of departures of Russians abroad in each quarter from 2020 Q4 to 2022 Q3 for each potential destination country in our sample and the corresponding quarterly number of Yandex searches for those countries. The middle plot shows the monthly number of Russian visitors to Turkey, Georgia, and Serbia from January 2021 to November 2022 and the corresponding monthly number of Yandex searches for those countries. The right plot depicts the monthly value of the Google Trends Index for mobility-related queries in each country in the sample from December 2020 to November 2022 and the corresponding monthly number of Yandex searches for those countries normalized within the country.

To assess if increased search activity within Russia corresponded to actual movement, we analyzed search patterns in the potential destination countries. Yandex provided data for Russia and for many destinations in our sample. When search behavior in Russia and these destinations was compared, both showed spikes around the key dates. The autocorrelation plot in Figure 4 demonstrates that Russian migration-related searches align with similar queries in all destination countries in our sample, with a lag of several weeks. This implies that such queries can effectively mirror mobility behavior, as recent Russian movers may continue searching for mobility-related information upon arrival in destination countries.
Figure 4: Correlations between mobility-related searches performed within and outside of Russia

Note: The top two panels show that searches performed within Russia and in all potential destination countries in our sample peaked in the days following the Russian invasion of Ukraine and the announcement of military mobilization. Vertical black dashed lines indicate the invasion in February 2022 and the announcement of military mobilization in September 2022. The bottom panel provides temporal autocorrelations measured with various weekly lags between the two time series, showing that peaks in initial mobility-related web searches within Russia were followed by peaks in destination countries.

3. Drivers of mobility intentions

In order to quantify the picture drawn in the media with the help of digital trace data, we employed a negative binomial regression that regresses the search results for each potential destination country on several sociodemographic indicators and regional characteristics. We refrained from adding interaction terms for each variable of interest as this would make interpretation harder. Instead we compared the two mobility waves by constructing one model for the period following the initial full-scale invasion of Ukraine and a second model for the period following the Russian announcement of mobilization. The unit of analysis was pairwise combinations of potential Russian origin cities and potential destination countries for a total sample size of 17,146 in each model.
For the analysis we also used public data sources. The number of Russian migrants in the destination countries was taken from the United Nations Population Division’s international migrant stock data for 2020 (United Nations Department of Economic and Social Affairs, Population Division 2020). Other socioeconomic indicators for the destination countries were derived from the World Development Indicators database (World Development Indicators n.d.). Municipal-level indicators for 2020 were provided by the Russian statistics office (Federal State Statistics Service of the Russian Federation 2020).

The samples for the models were restricted to the three-week period after the onset of the invasion and after mobilization as this captured the largest jumps in the search incidence. The dependent variable was the sum of mobility-related queries performed in the three weeks after the key dates within each city. The independent variables used were the distance between the respective city and the potential destination country, the population of the origin city, the population of the potential destination country, its gross domestic product per capita, the number of Russian migrants in the country, the average wage in each city, the share of inhabitants with higher education, the share of ethnic Russian population, the share of young (20–29 years old) male inhabitants, binary variables indicating entry regulations for Russian passport-holders, and whether the country bordered Russia. We also controlled for the number of queries three weeks before the respective key date.

### 3.1 Geographic preferences reflected in the search trends

In both models the coefficient for log distance was negative, as expected. However, it was much larger in absolute size in the after-invasion model than in the after-mobilization model, indicating more searches for closer destinations right after the invasion. The number of Russian migrants in a destination country was positively related to the search behavior. Historically, potential movers tend to consider destinations with a higher share of their fellow nationals, as diaspora presence is associated with lower migration costs and increased motivation to migrate (Beine, Docquier, and Özden 2010).

The results in Table 1 show that in both models the population size of the origin city contributed significantly to mobility-related search behavior. This is unsurprising, since large urban agglomerations are likely to be close to international airports and their inhabitants are likely to have greater financial means. On the other hand, the coefficient for population size of the countries specified in the queries differed across models. It was associated with a slight increase in searches in the after-invasion model and a slight decrease in searches in the mobilization model.
Table 1: Regression coefficients from a generalized linear model (GLM) showing relationship between explanatory variables and mobility-related web searches after the large-scale invasion versus after military mobilization in Russia

<table>
<thead>
<tr>
<th>Dependent variables:</th>
<th>log(mobility-rel. queries before war)</th>
<th>log(mobility-rel. queries before mob.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>log(mobility-rel. queries before war)</td>
<td>0.750</td>
<td>0.734</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>log(distance)</td>
<td>–0.285</td>
<td>–0.018</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>log(pop. of origin)</td>
<td>0.254</td>
<td>0.262</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>log(pop. of destination)</td>
<td>0.046</td>
<td>–0.012</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>log(ave. wage in origin)</td>
<td>0.136</td>
<td>–0.017</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>% of ethnic Russians in origin</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>% of higher educated in origin</td>
<td>0.013</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>% of young men in origin</td>
<td>0.005</td>
<td>–0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>log(GDP of destination)</td>
<td>0.055</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>log(Russians in destination)</td>
<td>0.078</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>visa-free and internal passport</td>
<td>0.760</td>
<td>1.164</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>visa-free and international passport</td>
<td>0.638</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>bordering Russia</td>
<td>0.092</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Intercept</td>
<td>–1.955</td>
<td>–3.325</td>
</tr>
<tr>
<td></td>
<td>(0.305)</td>
<td>(0.335)</td>
</tr>
</tbody>
</table>

| Observations | 18,700 | 18,700 |
| Log Likelihood | –82,414.820 | –82,368.160 |
| theta         | 2.523 (0.032) | 2.041 (0.026) |
| Akaike Inf. Crit. | 164,857.600 | 164,764.300 |

Note: Standard errors in parentheses.

The binary variables for entry regulations and vicinity were both positively associated with mobility-related search behavior. The transformed coefficient for countries bordering Russia increased more than fourfold in the mobilization model. Moreover, the possibility of entering a country with only a national identification
Such differences in the coefficients of the variables of the two models might be an indication of differences in the composition and preparations of the two outflows from Russia, the first involving individuals from higher socioeconomic backgrounds and likely with visas for Western countries, the second more sudden and involving many individuals without visas or international passports.

3.2 Socioeconomic correlations with the search trends

Comparing the coefficients for the socioeconomic city-level indicators in the two models hints at differences in the search preferences of the second and first waves. The gross domestic product of the destination country does not seem to have played an important role in the decision of which country to search for. The coefficients were fairly small for the model after the start of the invasion and negative for the model after the start of mobilization.

The coefficient for average wage in the search location is almost four times larger in the after-invasion model than in the after-mobilization model. In particular, there were notably more mobility-related searches in high-wage origin cities after the invasion, but this effect was three times smaller after mobilization. The shares of ethnic Russian and highly educated individuals contributed positively to the search incidence while the coefficients remained small in size.

After the Russian mobilization it became apparent that some ethnic minority regions were more affected by recruitment and war casualties than regions populated by ethnic Russians (Bessudnov 2023). However, an analogous relationship at the city level is not visible in our model results. This is most likely due to the importance of rurality and average income in the regions more affected by recruitment. The relationship between ethnicity and search behavior could even be reversed, as regions with higher shares of ethnic minorities are often poorer and therefore their citizens have less means to prepare for a journey. A spatial analysis including the larger regions would be more suitable to capture such a pattern.

In contrast to our expectation, the share of young male inhabitants had little effect on search behavior in the after-mobilization model. However, the variation in shares of young men across cities was fairly small, making this variable a poor indicator of the search behavior of this population subgroup.
Both education and the average wage were associated with increased search behavior across both waves. We report consistent results for the search rates per population of origin city in the supplementary materials in the replicability materials. These results hint at a potential brain drain but cannot provide strong insights into the demographic composition of the actual outflow of leavers from Russia.

4. Discussion and conclusion

The Yandex search data employed in this analysis suggest that the socioeconomic and demographic composition of the two search waves might have differed significantly. We found that countries not requiring an international passport were more popular in searches after mobilization was announced than after the start of the invasion. The same holds for countries that share a land border with Russia. Higher average wages correlated positively with increased mobility-related search behavior after the invasion as well as after mobilization. However, the correlation was much weaker in the second wave. We observed a similar pattern for the importance of the potential destination country’s GDP, which even becomes negative in the second wave. These results indicate that the destination search preferences changed across waves, probably due to differences in the mobility motivations and social and economic capital of the searchers. Assuming that a certain share of the searches translated into actual moves out of Russia in the time following the war, we would expect those persons to be from the higher-earning urban middle class. Therefore, our results support common claims about Russia facing a brain drain (Wachs 2023).

Unfortunately, our data were not suitable for assessing the actual number and profiles of those who have left Russia. Searches can only indicate the intention to move, not actual mobility. Moreover, the absolute number of searches cannot be equated with the number of potential movers. We have no information on whether Yandex records the unique IPs of its users or if a small number of very active users drives up the search incidence. We only included sociodemographic indicators at the municipal level in our analysis and any conclusion made about individual search behavior is prone to ecological fallacy (i.e., erroneously inferring characteristics of individuals based on observations of the group to which they belong). Another problem induced by the data is that the provenance of geolocations provided by Yandex is a black box for us, and we cannot rule out the skewing presence of VPN use which could obscure the true locations of some Yandex users. Lastly, Yandex users are not representative of the Russian society as a whole.

Despite the limitations, the potential of Yandex searches for migration and mobility research is still untapped. Our work constitutes the first steps in the exploration of Yandex
Wordstat as an innovative data source. The potential of online searches clearly lies in their timeliness and public availability for predicting migration flows. The strength of Yandex compared to its competitor Google lies in its geographic granularity, its provision of raw search counts, and its dominant market share in Russia.

Against the backdrop of the ongoing Russian invasion of Ukraine, conflict-induced mobility can be expected to continue and rapid estimations of such sudden population changes become more crucial than ever. To facilitate this, future research should consider incorporating in the analysis more official data sources, like destination countries’ visitor statistics, and should explore ways of predicting the actual number of citizens who have left Russia due to the invasion and subsequent mobilization. Digital trace data like Yandex search trends can be a powerful ingredient in any timely estimation; however, its limitations and potential have to be thoroughly explored and considered. Therefore, future avenues of research should compare Yandex search trends with survey-based measures of migration aspirations (e.g., Gallup World Polls) to account for biases of such data types. However, in the special context of an authoritarian country like Russia, official data sources may not be reliable and digital traces can represent a valid alternative. With our work we encourage the exploration of Yandex data for characterizing and eventually quantifying outflows from Russia.

5. Acknowledgements

This project was initiated during the Summer Institute for Computational Social Science (SICSS) 2022 in Oxford. We thank the organizers for the opportunity to develop our project idea and for their helpful comments and encouragement to pursue this research. This work benefited from the support of Jisu Kim and Ebru Sanlitürk, who have reviewed the manuscript at various stages in its development and have provided advice and feedback, as well as other MPIDR colleagues who provided useful comments during an internal seminar talk. We are especially grateful to Domantas Jasilionis for sharing his expertise with us and to Sergey Timonin for providing us with the most recent data for Russia. Also, Ekaterina Melianova provided invaluable feedback on statistical modelling questions. We thank Egor Kotov for his suggestions. We gratefully acknowledge the resources provided by the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS). We also thank the reviewers and the editor for their contributions, which significantly improved the paper.

Data and code availability
The data and code for replication are publicly available on OSF (Anastasiadou, Volgin, and Leasure 2023).
Authors’ contributions
This project was initiated by AA and was developed in collaboration with AV and further elaborated with DRL. AV performed the Yandex data collection and analysis. AA wrote the manuscript and supported data collection, cleaning, and analysis. DRL has provided substantial feedback and guidance throughout the process. All authors contributed to the final version of the manuscript.
Bibliography


