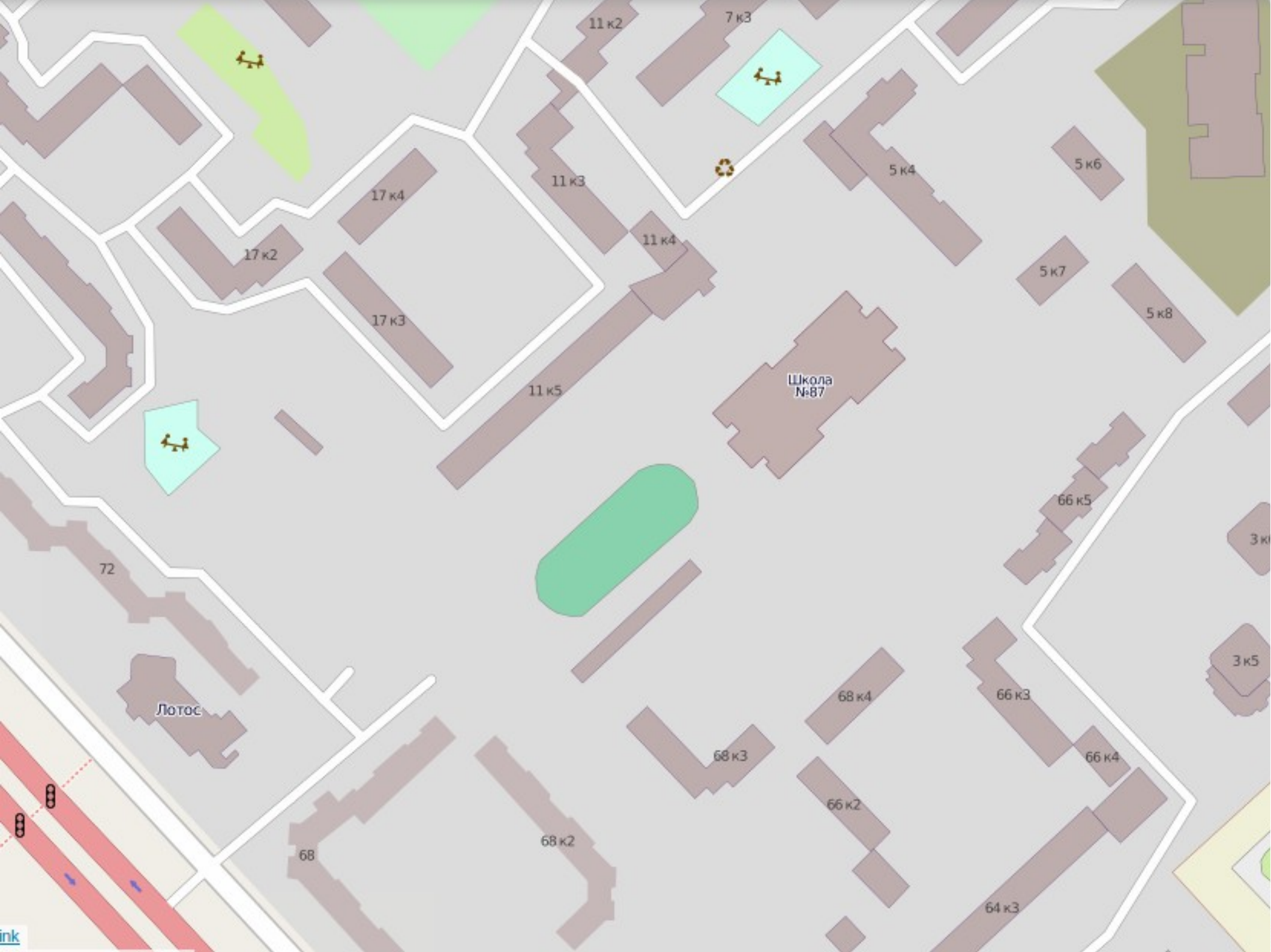


OpenStreetMap address base: ready for prime time?

Maxim Dubinin
sim@gis-lab.info

SotM Baltics 2013



The Question

Can OpenStreetMap address database be used to create complete geographic datasets?

Setup

- Creation of data layers for different features
- Large areas (Russia)
- Thousands objects
- Practical applications

You will find answers here for...

1. How good is OSM address database and fully automatic geocoding?
2. How much does postprocessing help?
3. How is completeness distributed across Russia?
4. How good is the quality compared to other geocoders?

...but, nothing about...

- What are the mistakes of geocoding and how it can be improved?
- What is the right scheme for addressing?
- When will OSM take over the world?

Some definitions

Result — correct lat/long for an address

Result ~ data preparation + geocoding + postprocessing

- Data prep — make well structured address
- Geocoding — find lat/long for it with osm.org.ru
- Postprocessing — fix it manually if wrong

Examples

- OpenPolice — where are the local cops in Moscow
- Elections — where are the voting stations in Moscow
- Orphanages — where are the children orphanages in Russia

Question 1

1.How good is OSM address database and fully automatic geocoding?

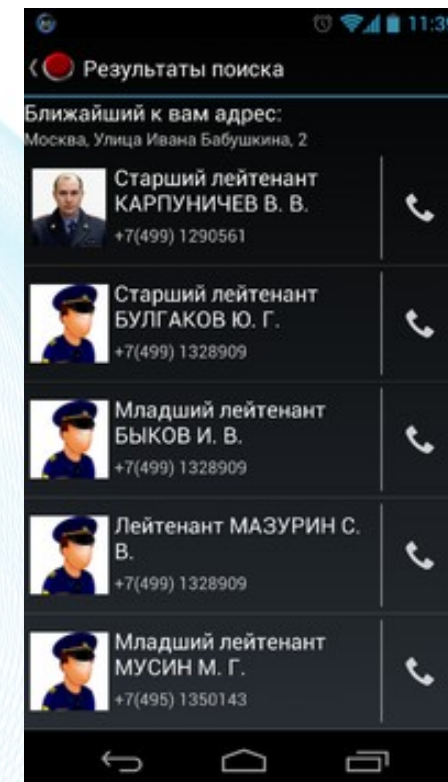
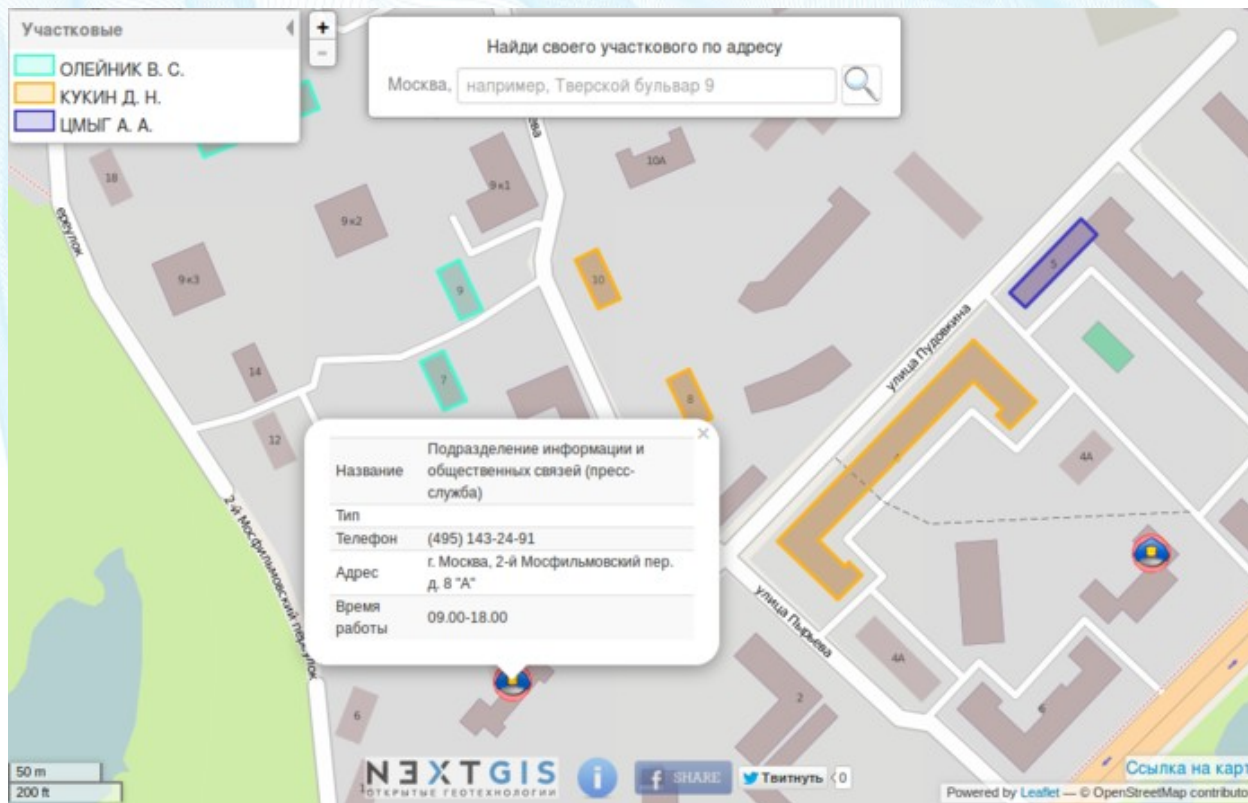
2.How much does postprocessing help?

3.How is quality distributed across Russia?

4.How good is the quality compared to other geocoders?

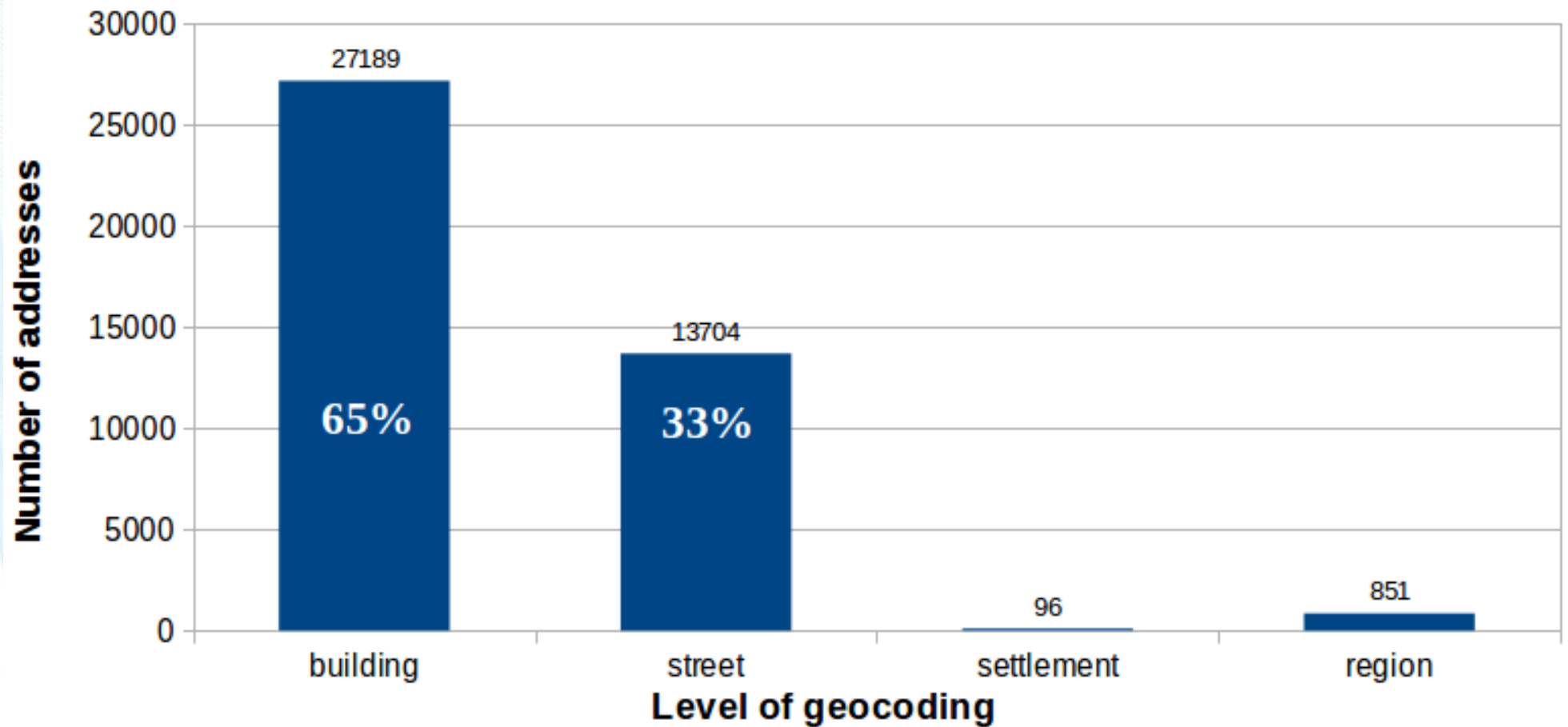
OpenPolice

- Extract all addresses from 112.ru
- Geocode them
- Relate them to buildings in Moscow to get areas of responsibility



Results

- Total: ~41000 addresses in Moscow



Question 2

1. How good is OSM address database and fully automatic geocoding?
- 2. How much does postprocessing help?**
3. How is quality distributed across Russia?
4. How good is the quality compared to other geocoders?

Voting commissions

- Extract all addresses from public database
- Geocode them
- Crowdsource post-processing

The screenshot displays the uikgeo.gis-lab.info web application. The central map shows Yaroslavl with various streets and landmarks. Orange circles with numbers (1, 2, 3, 4, 5, 6) are placed on the map, indicating the locations of voting commissions. The left sidebar contains a search bar and a list of addresses, including:

- 171 г. Ярославль, просп. Ленина, д. 24а
- 173 г. Ярославль, просп. Ленина, д. 24а
- 125 г. Ярославль, ул. Большая Октябрьская, д. 64а
- 124 г. Ярославль, ул. Республиканская, д. 108/1
- 126 г. Ярославль, ул. Свободы, д. 46
- 70 г. Ярославль, Тверицкая набережная, д. 58а
- 71 г. Ярославль, просп. Авиаторов, д. 37
- 139 г. Ярославль, ул. Павлика Морозова, д. 3
- 140 г. Ярославль, ул. Павлика Морозова, д. 3
- 172 г. Ярославль, ул. Тургенева, д. 14
- 174 г. Ярославль, просп. Ленина, д. 24а
- 175 г. Ярославль, ул. Шапова, д. 14
- 176 г. Ярославль, ул. Шапова, д. 14
- 138 г. Ярославль, ул. Угличская, д. 27
- 51 г. Ярославль, Архангельский пр-д, д. 13
- 53 г. Ярославль, Архангельский пр-д, д. 13
- 63 г. Ярославль, пр-д Моторостроителей

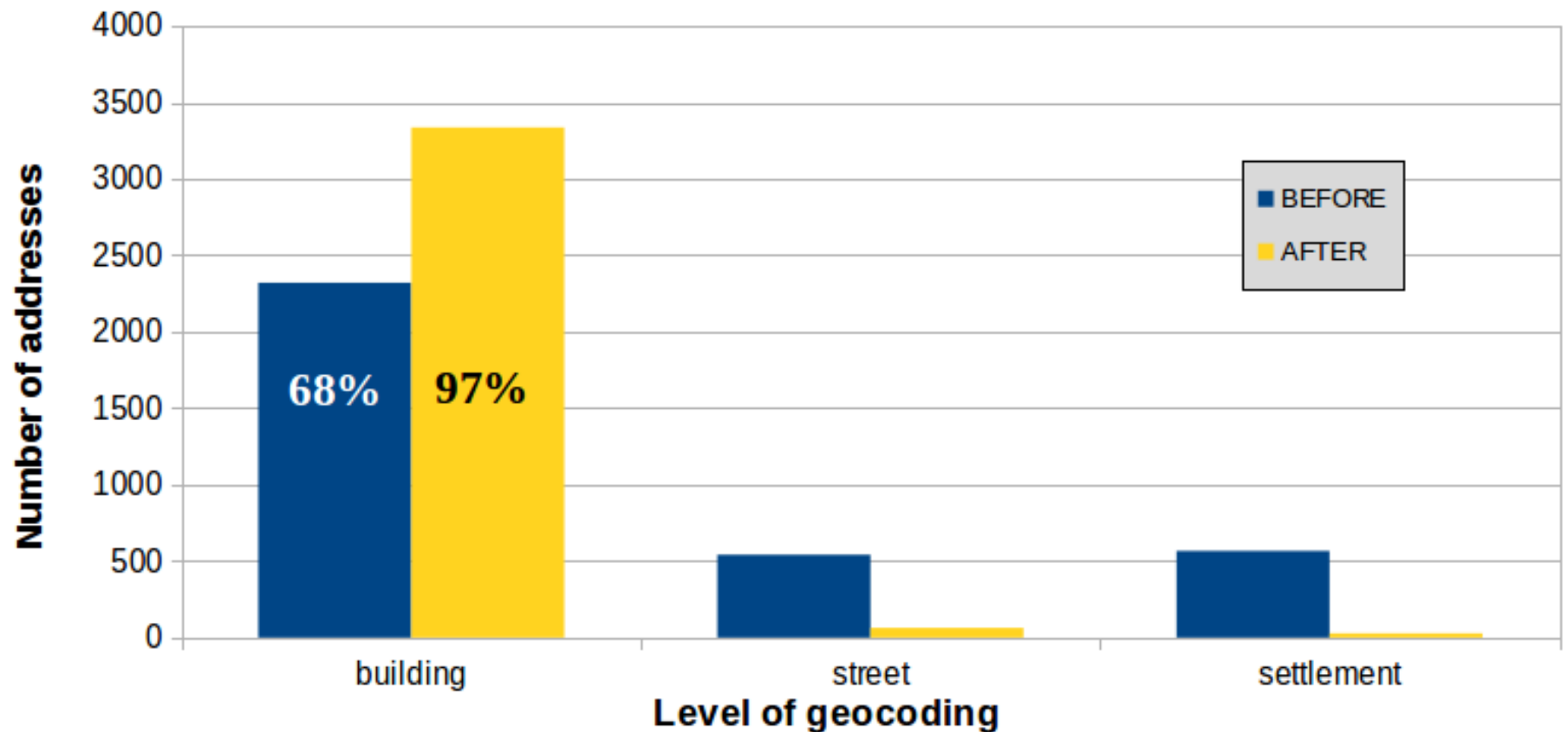
The right sidebar contains a form for editing commission data, including fields for:

- Номер УИКа: 119
- Регион: Ярославская область
- ТИК: ТИК Кировского района гор
- Адрес голосования: г. Ярославль, ул. Чайковского, д. 55
- Место голосования: ГОУ СПО ЯО «Ярославский гиапостроительный
- Точность: Дом
- Широта: 57.633422
- Долгота: 39.875436

Buttons for 'Перегеокодировать', 'Перецентрировать', 'Применить', and 'Отменить' are present. A checkbox for 'УИК принят' is checked. The bottom of the page shows the URL <http://uikgeo.gis-lab.info> and the copyright notice '2010 GeoEye, © 2013 Microsoft Corporation'.

Results

- Total: ~3500 addresses in Moscow
- Before post-processing VS after post-processing



Question 3

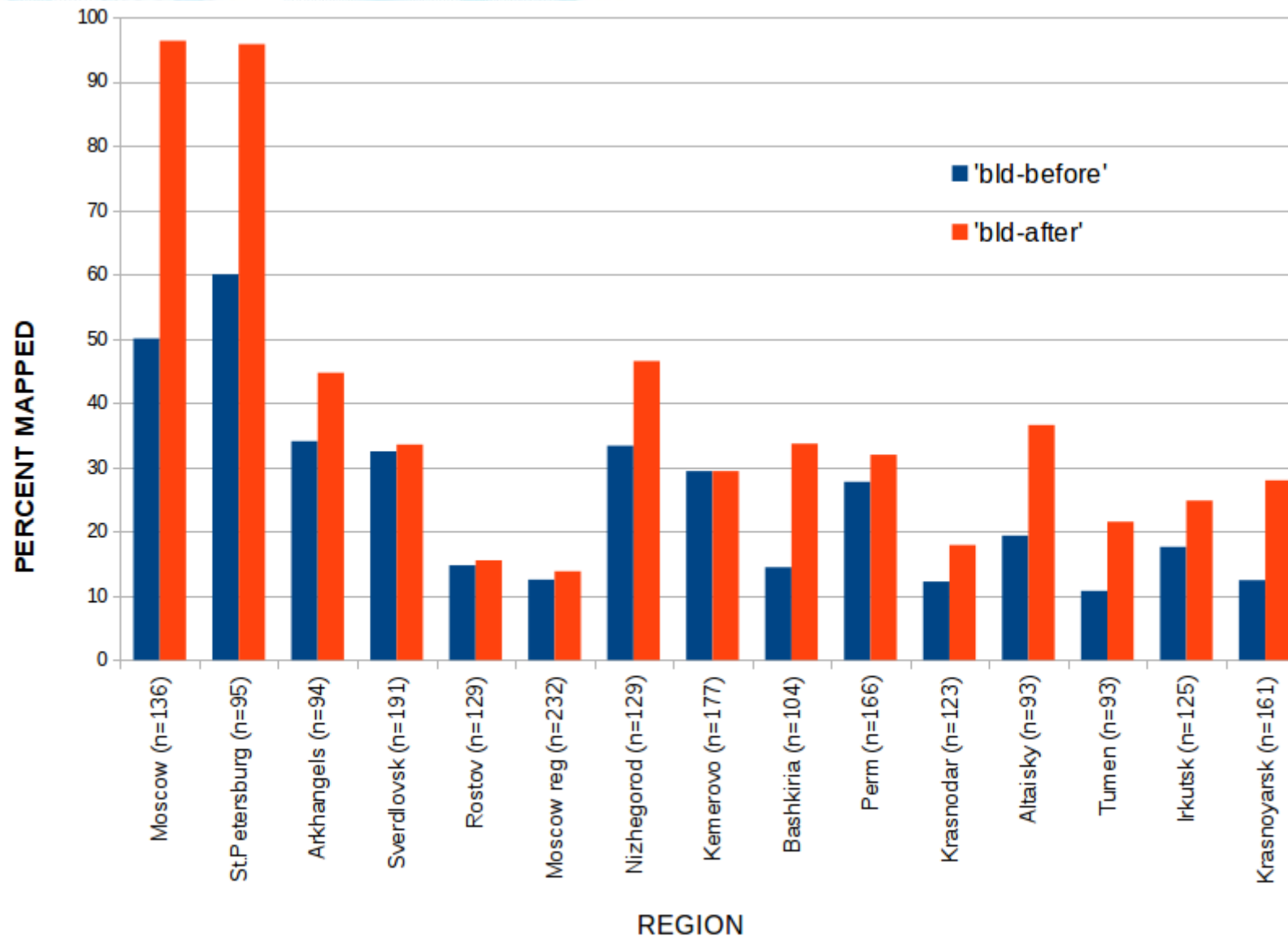
1. How good is OSM address database and fully automatic geocoding?
2. How much does postprocessing help?
- 3. How is completeness distributed across Russia?**
4. How good is quality compared to other geocoders?

Orphanages

- Extract all addresses from public database
- Geocode and post-process them
- All regions of Russia, ~5000 orphanages total, mean 50 per region

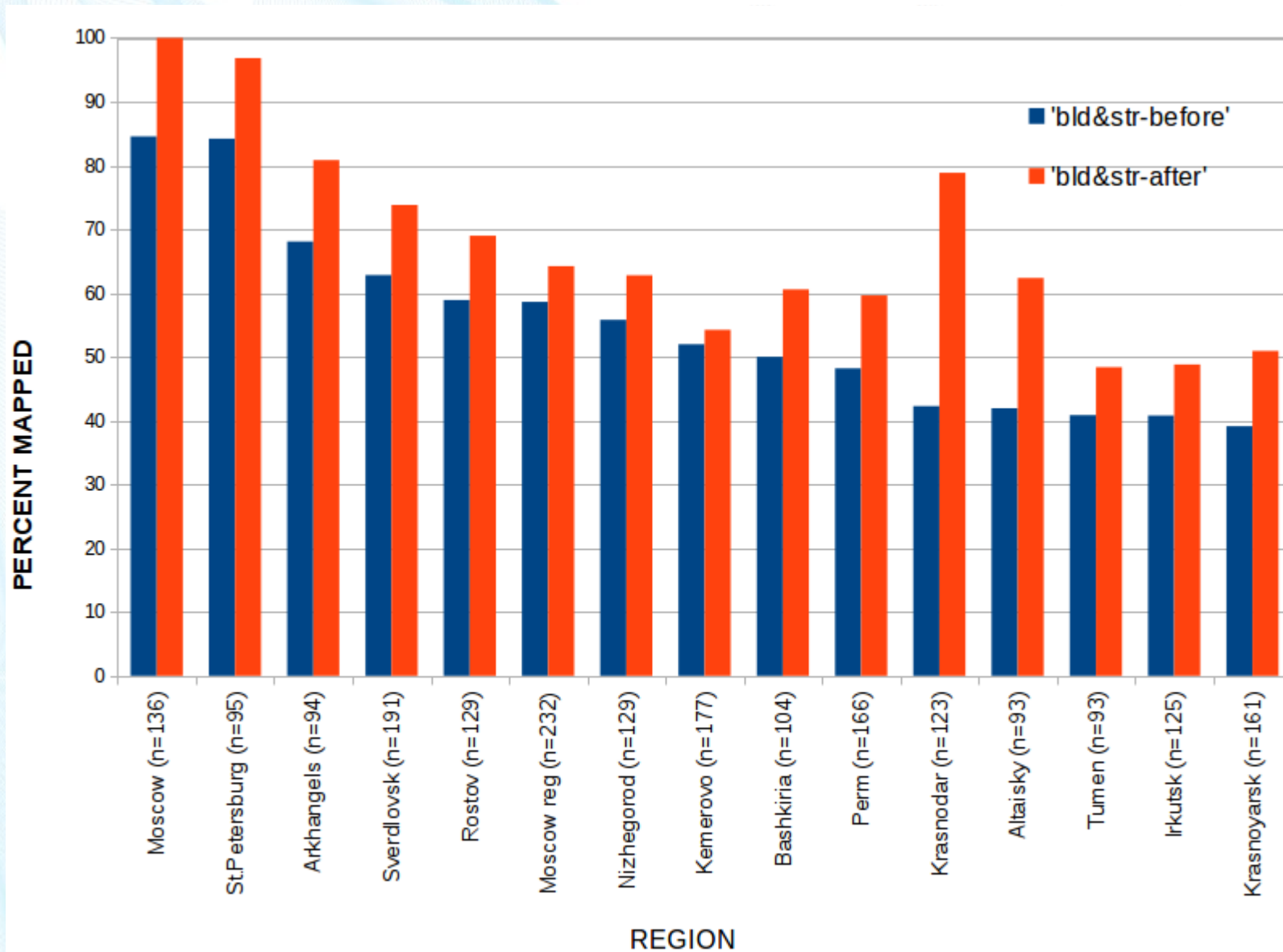
Orphanages

- Buildings before and after post-proc, % total



Orphanages

- Buildings and streets before and after post-proc, % total



Question 4

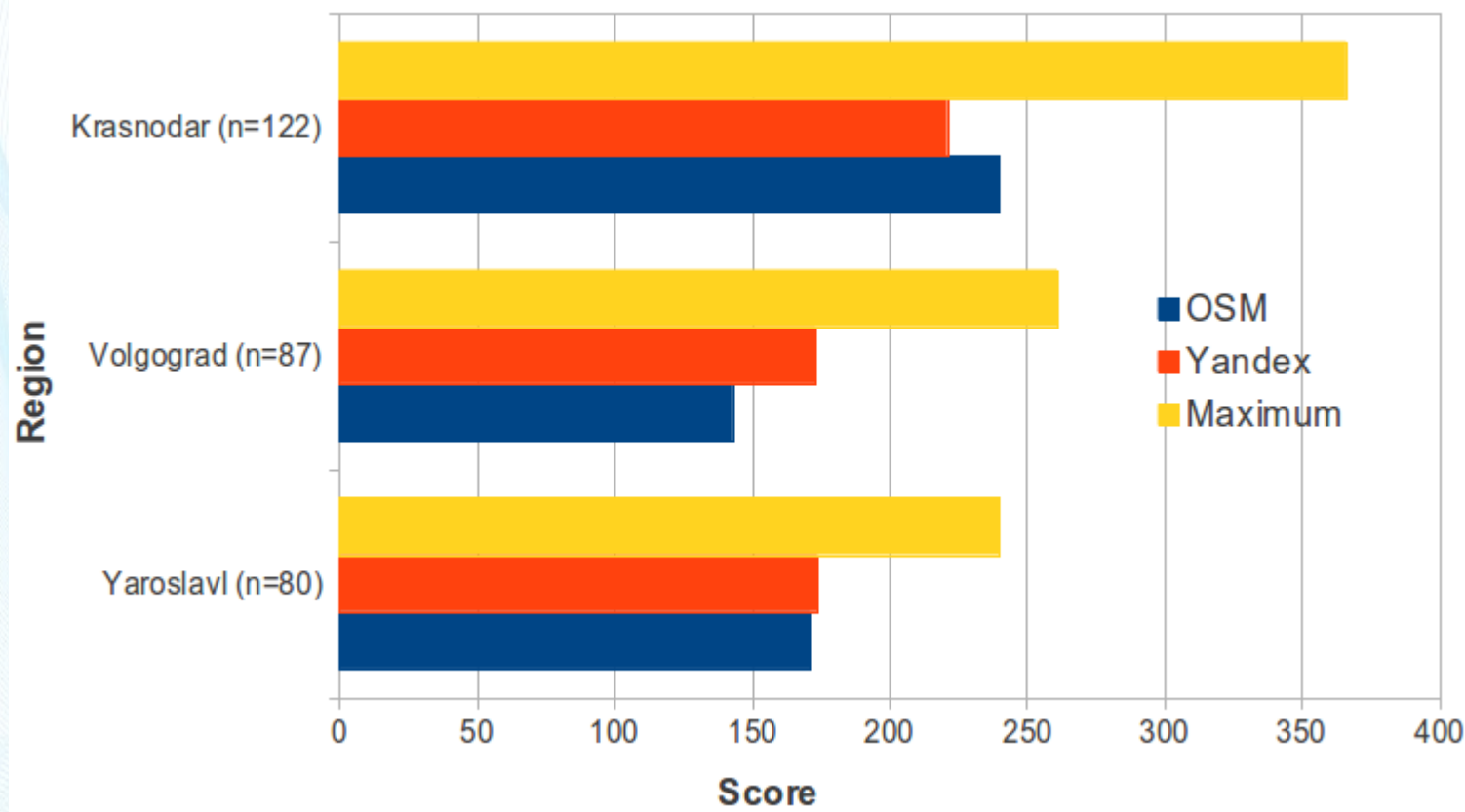
1. How good is OSM address database and fully automatic geocoding?
2. How much does postprocessing help?
3. How is quality distributed across Russia?
4. How good is the quality compared to other geocoders?

No project, just comparison

- Take few hundreds of addresses in different parts of Russia
- Geocode them with OSM and Yandex
- For each point, assign score: Building = 3, street = 2, settlement = 1
- Sum the scores up
- Compare

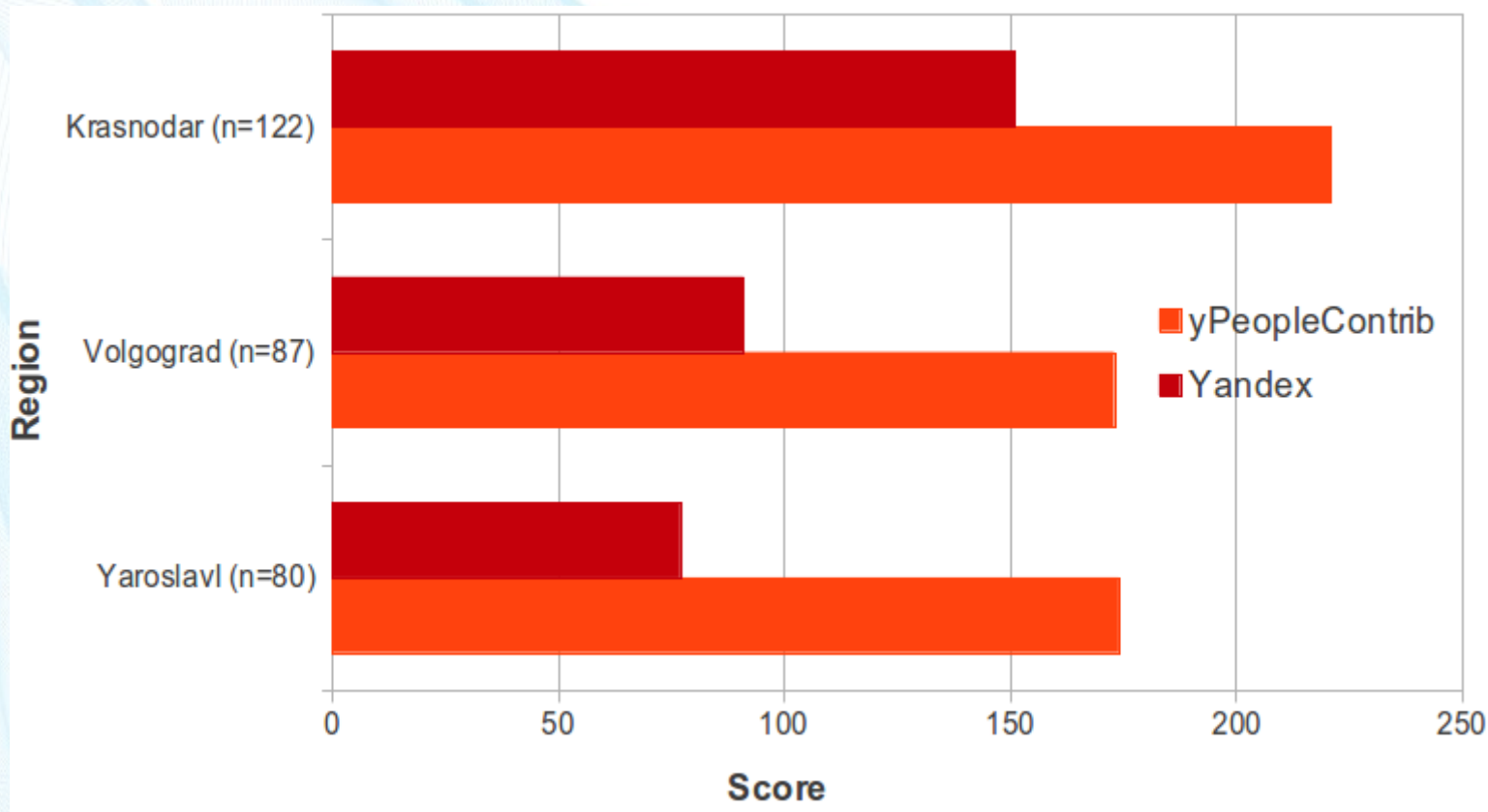
OSM vs Yandex

- Summed scores for geocoding accuracy



Yandex

- Yandex People's map contribution to total score



How to get better?

- Map more ;)
- Improve automatic geocoding
- Create positive feedback loop with geocoding projects

Sources for this presentation

<https://github.com/simgislab/osmaddress-sotmbaltics13>