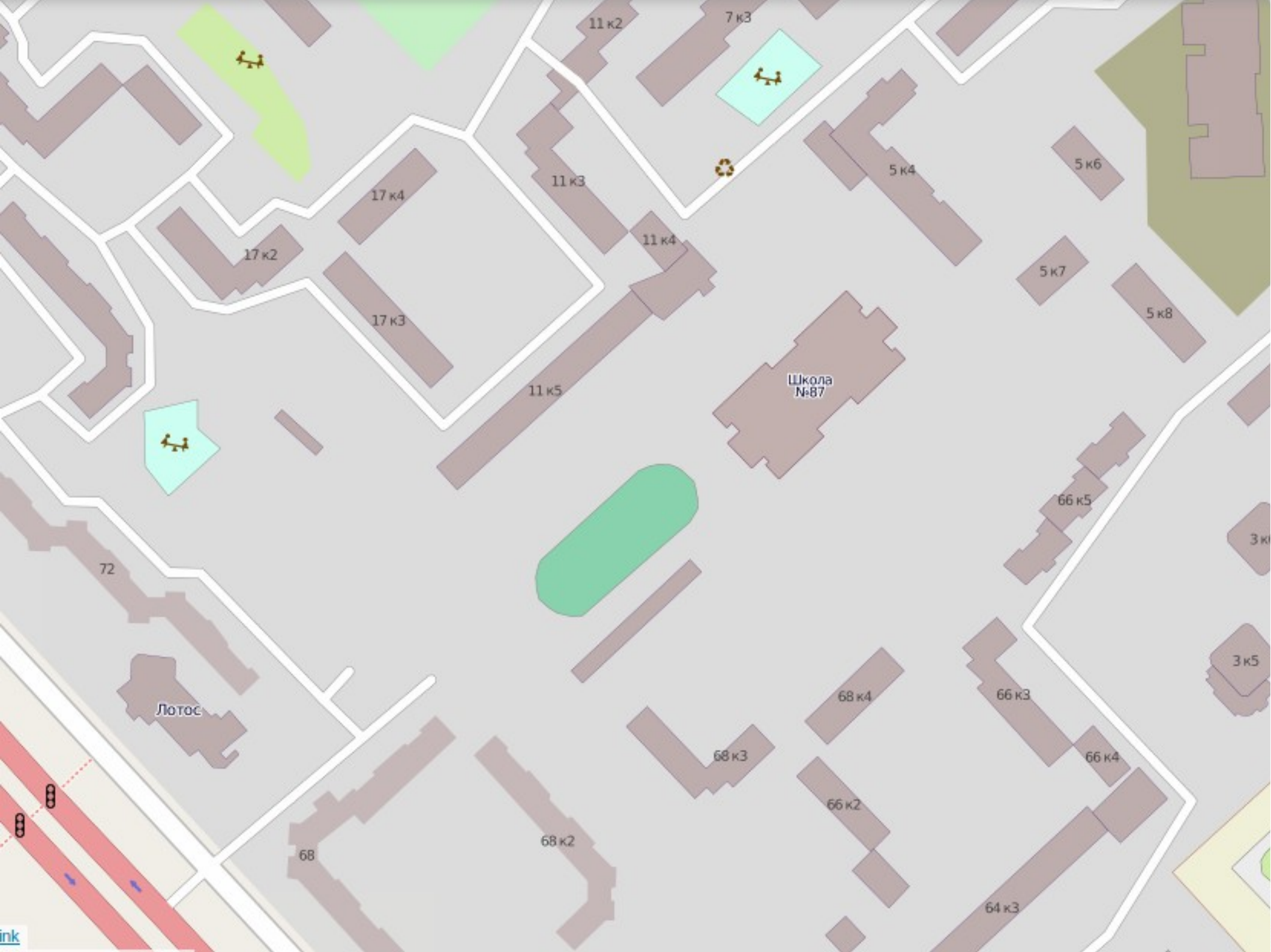


# **OpenStreetMap address base: ready for prime time?**

Maxim Dubinin  
[sim@gis-lab.info](mailto: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](https://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?**

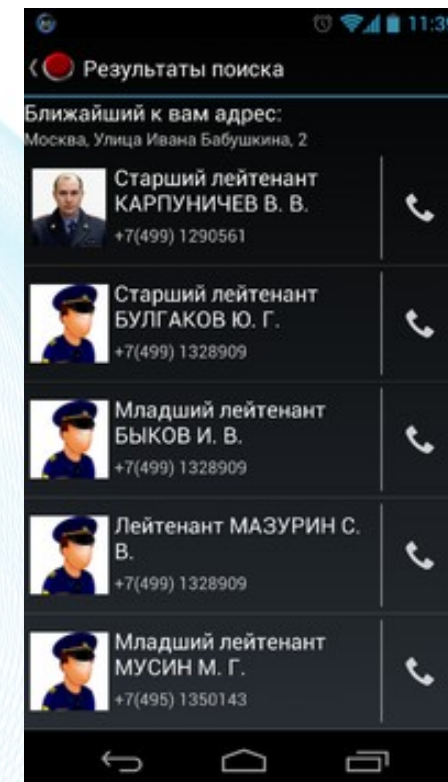
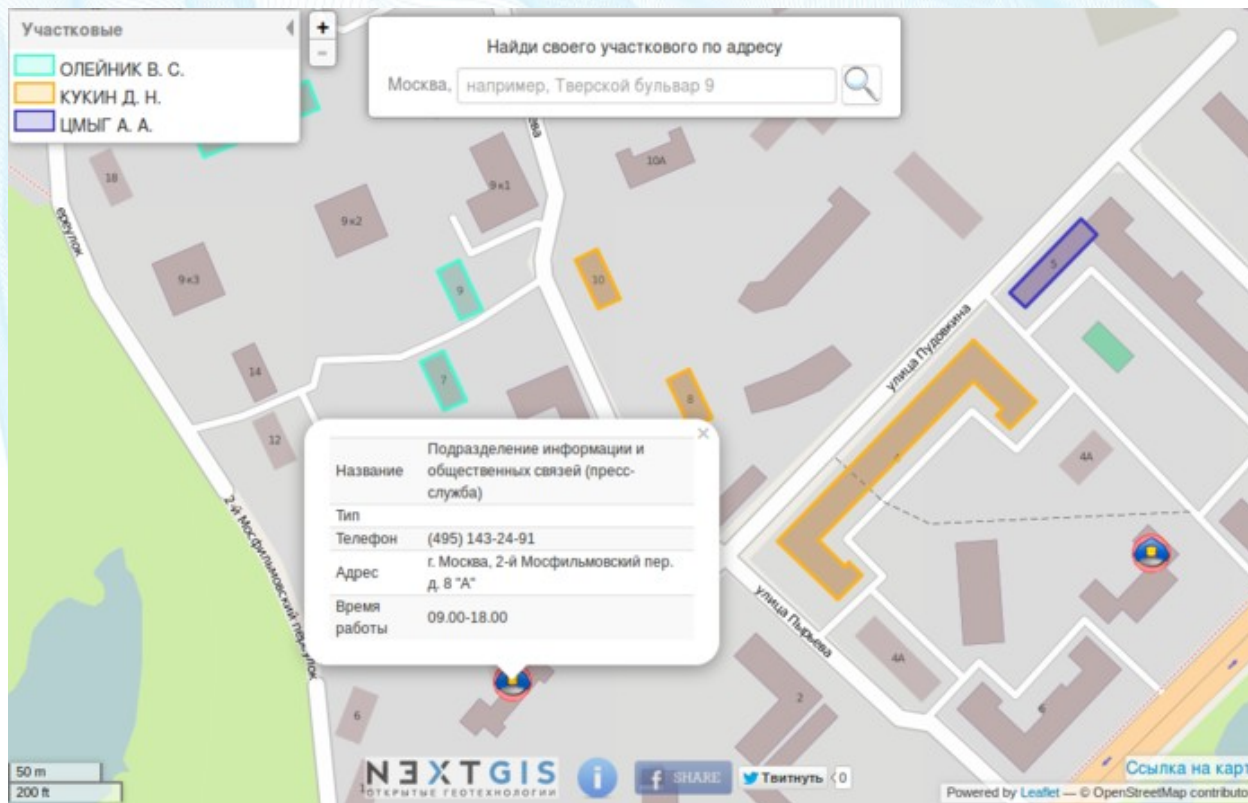
2.How much does postprocessing help?

3.How is quality distributed across Russia?

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

# OpenPolice

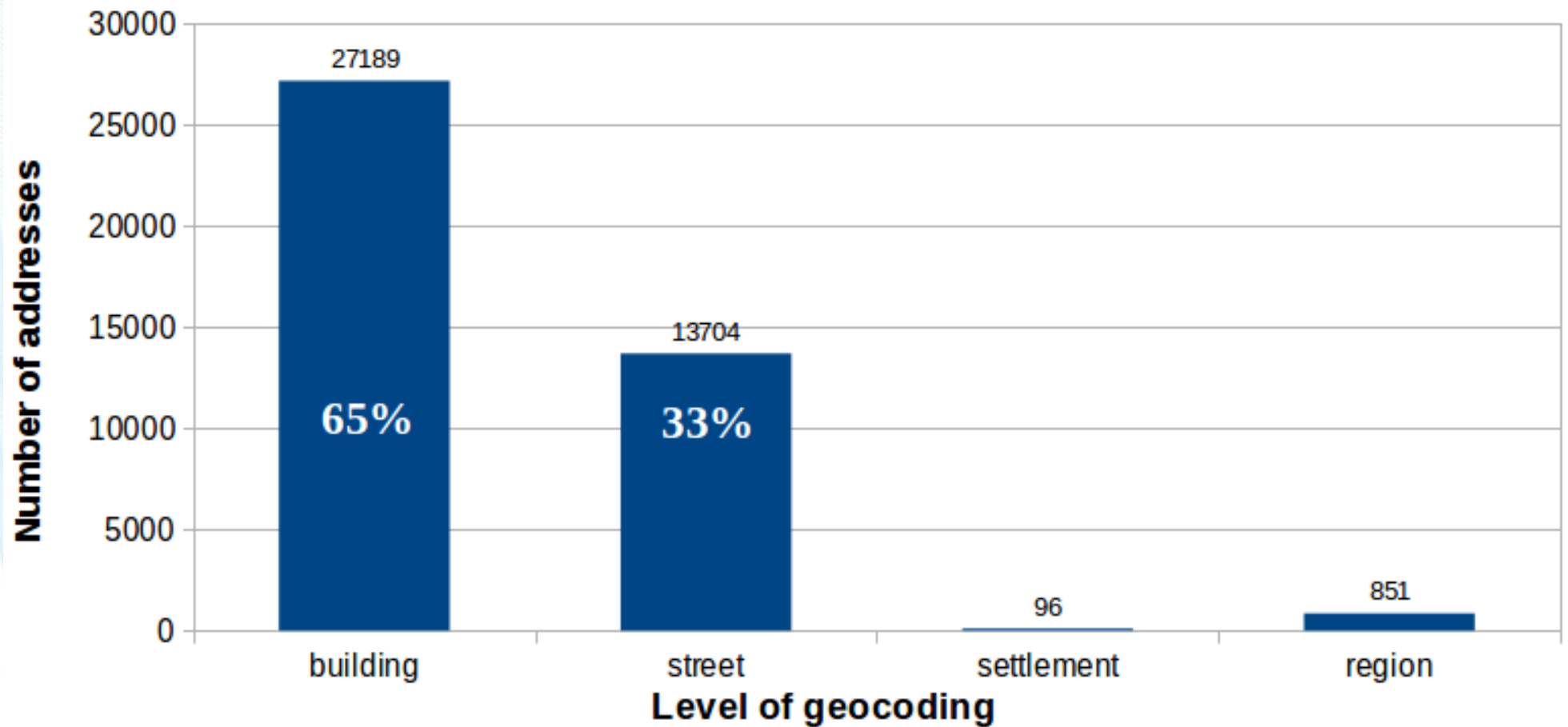
- Extract all addresses from 112.ru
- Geocode them
- Relate them to buildings in Moscow to get areas of responsibility





# 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

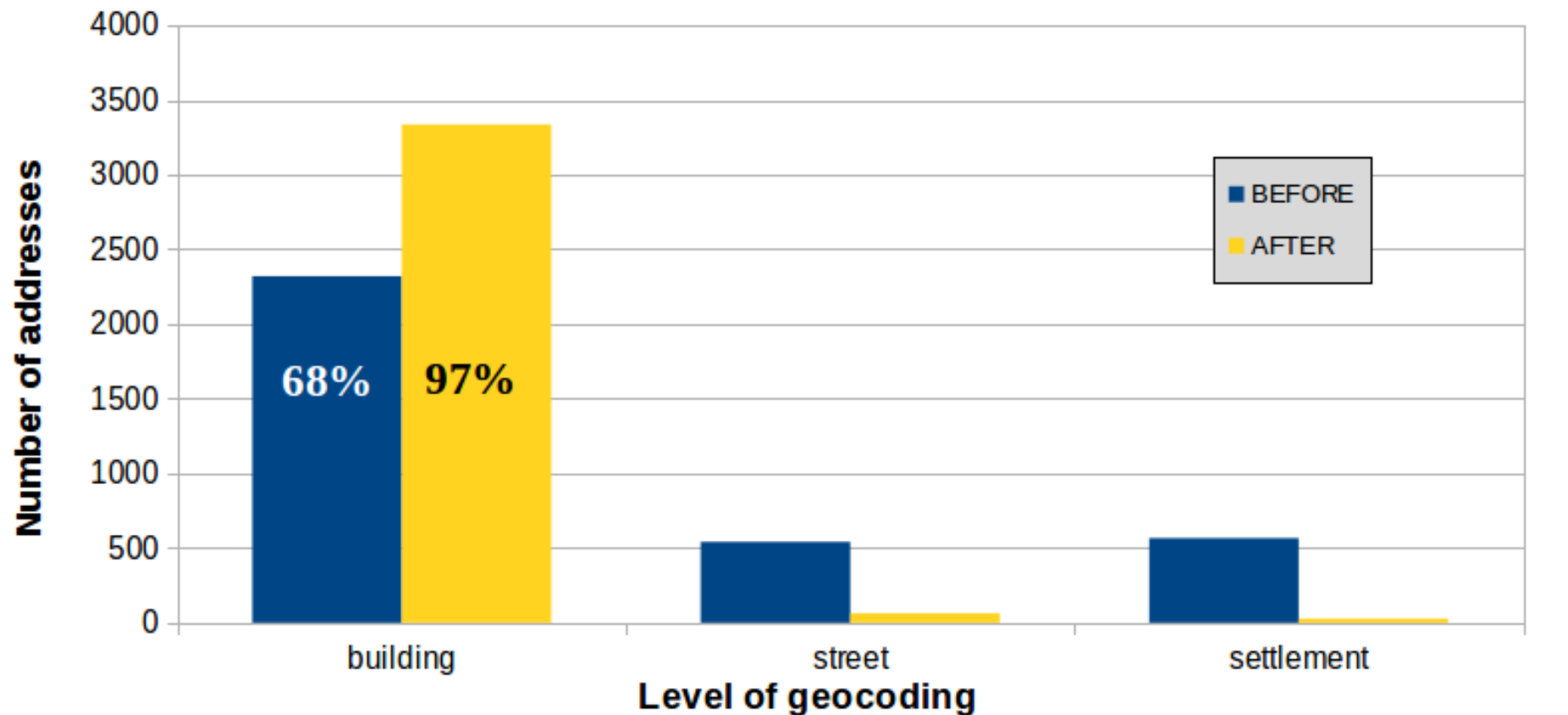
The screenshot displays the uikgeo.gis-lab.info web application. On the left, a search panel titled 'Поиск' (Search) contains fields for 'УИК' (Election District), 'УИК 2012', 'Адреса' (Addresses), 'Номер' (Number), and 'Адрес' (Address). Below these fields is a list of addresses in Yaroslavl, such as '171 г. Ярославль, просп. Ленина, д. 24а'. 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. On the right, a data entry panel titled 'maximdubinin' and 'Выйти' (Logout) contains a 'Редактор' (Editor) section. This section includes fields for 'Номер УИКа' (Election District Number), 'Регион' (Region), 'ТИК' (Territorial Electoral District), 'Адрес голосования' (Voting Address), 'Место голосования' (Voting Place), 'Точность' (Accuracy), 'Широта' (Latitude), and 'Долгота' (Longitude). Below these fields are buttons for 'Перегеокодировать' (Re-geocode), 'Пересцентрировать' (Re-center), 'Применить' (Apply), and 'Отменить' (Cancel). At the bottom right, there is a 'Комментарий' (Comment) field and a checkbox for 'УИК принят' (Election District accepted), with 'Отменить' (Cancel) and 'Сохранить' (Save) buttons.

<http://uikgeo.gis-lab.info>

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?

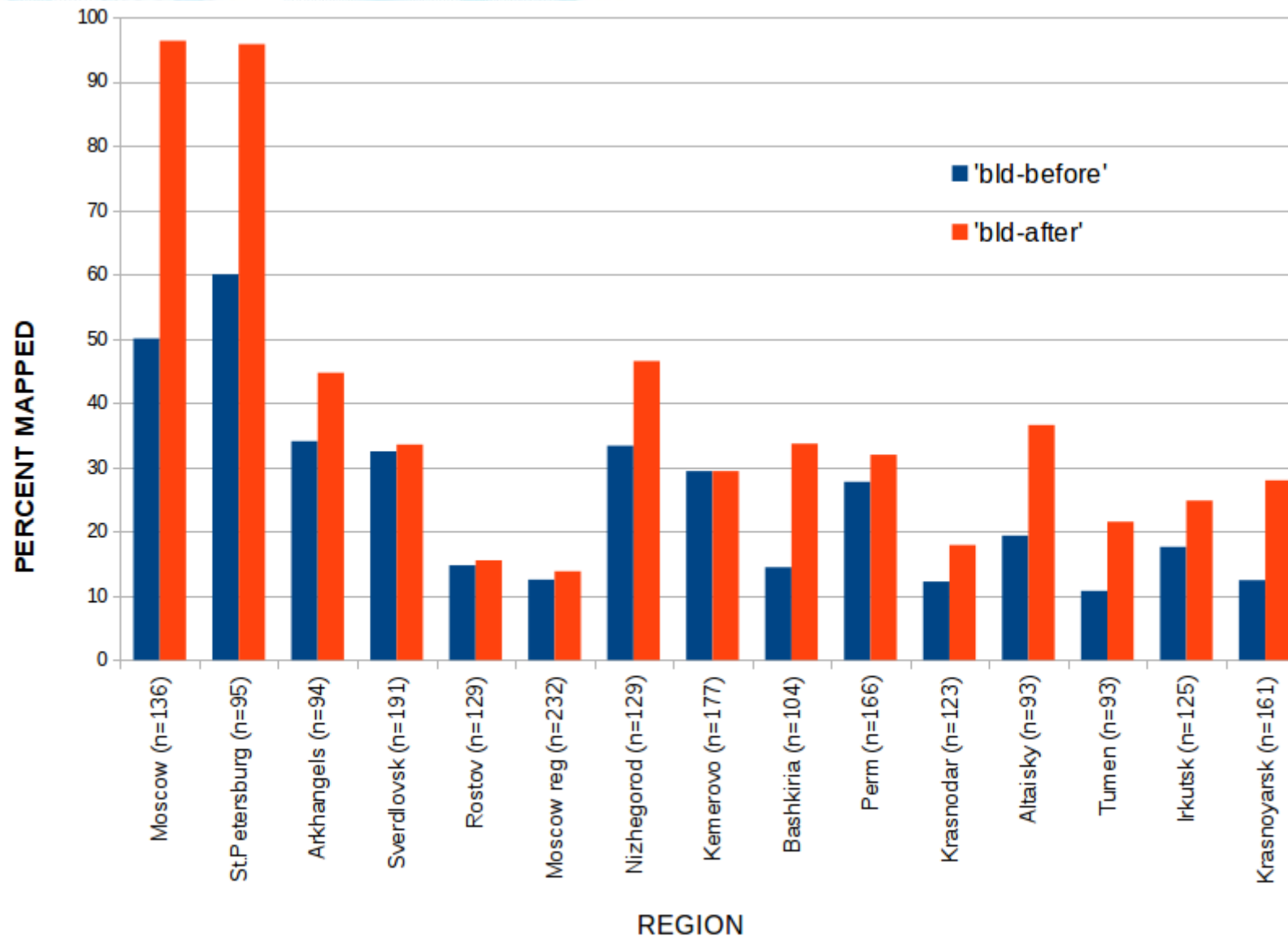
# Orphanages

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



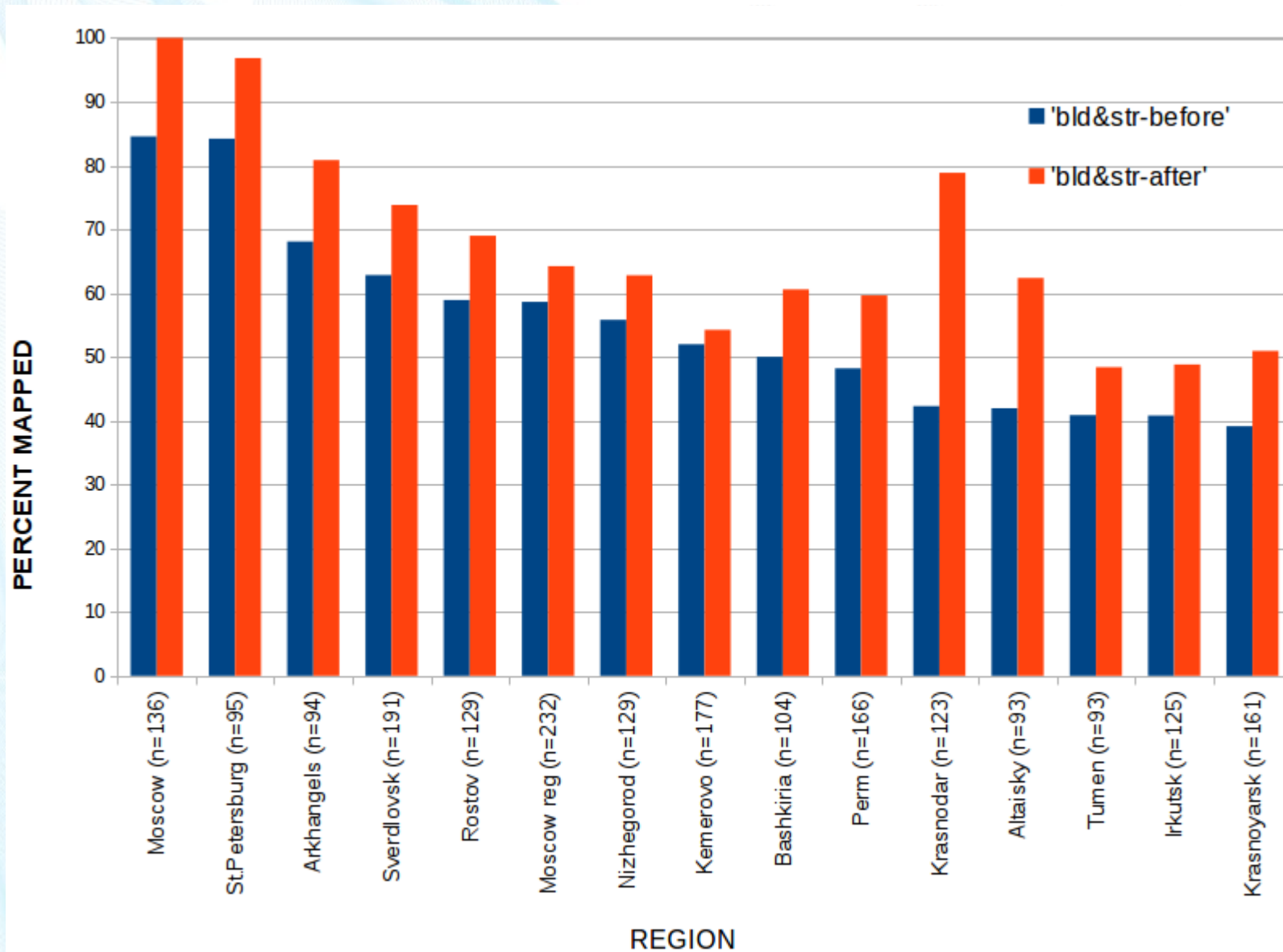
# Orphanages

- 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?

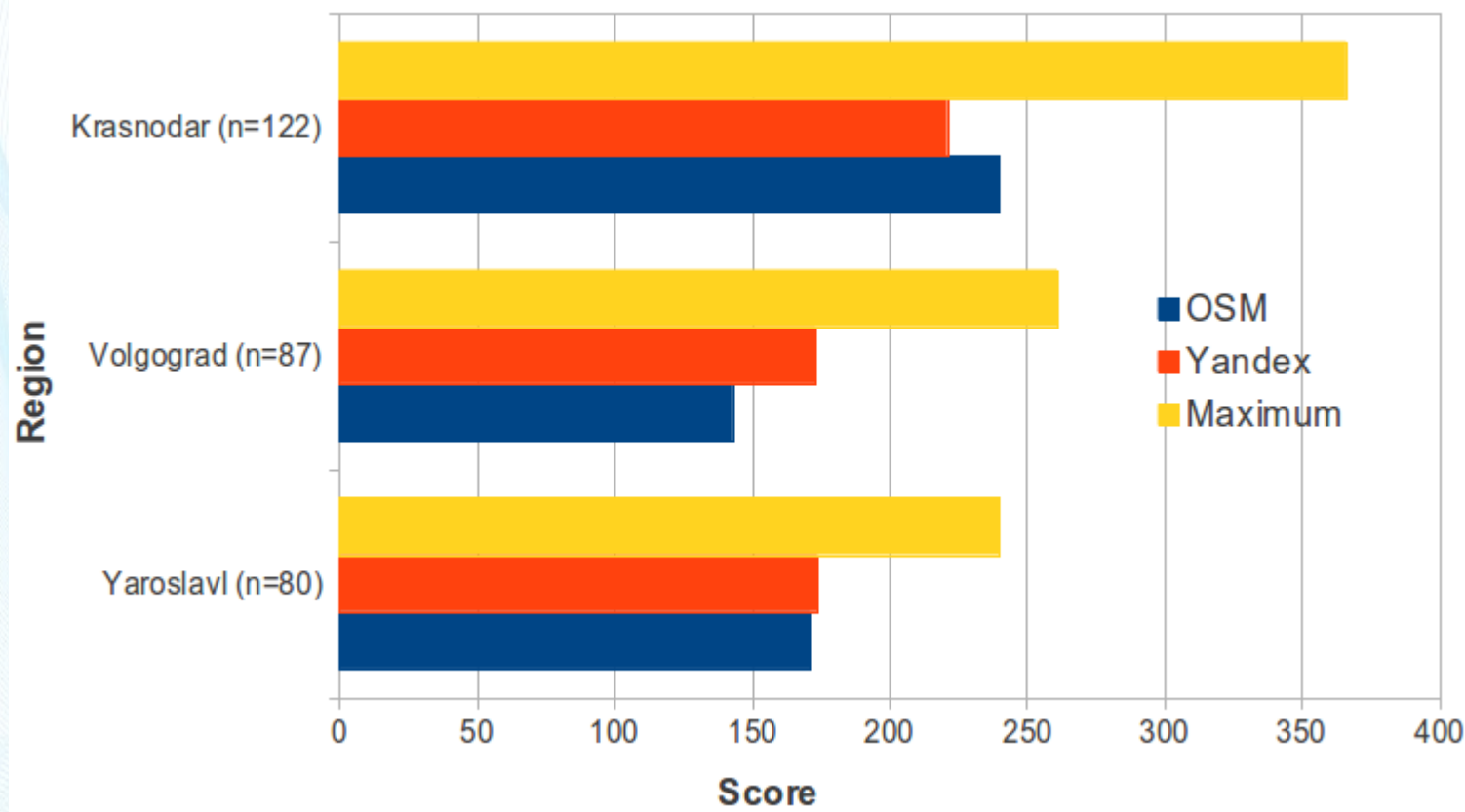
# No project, just comparison

- 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



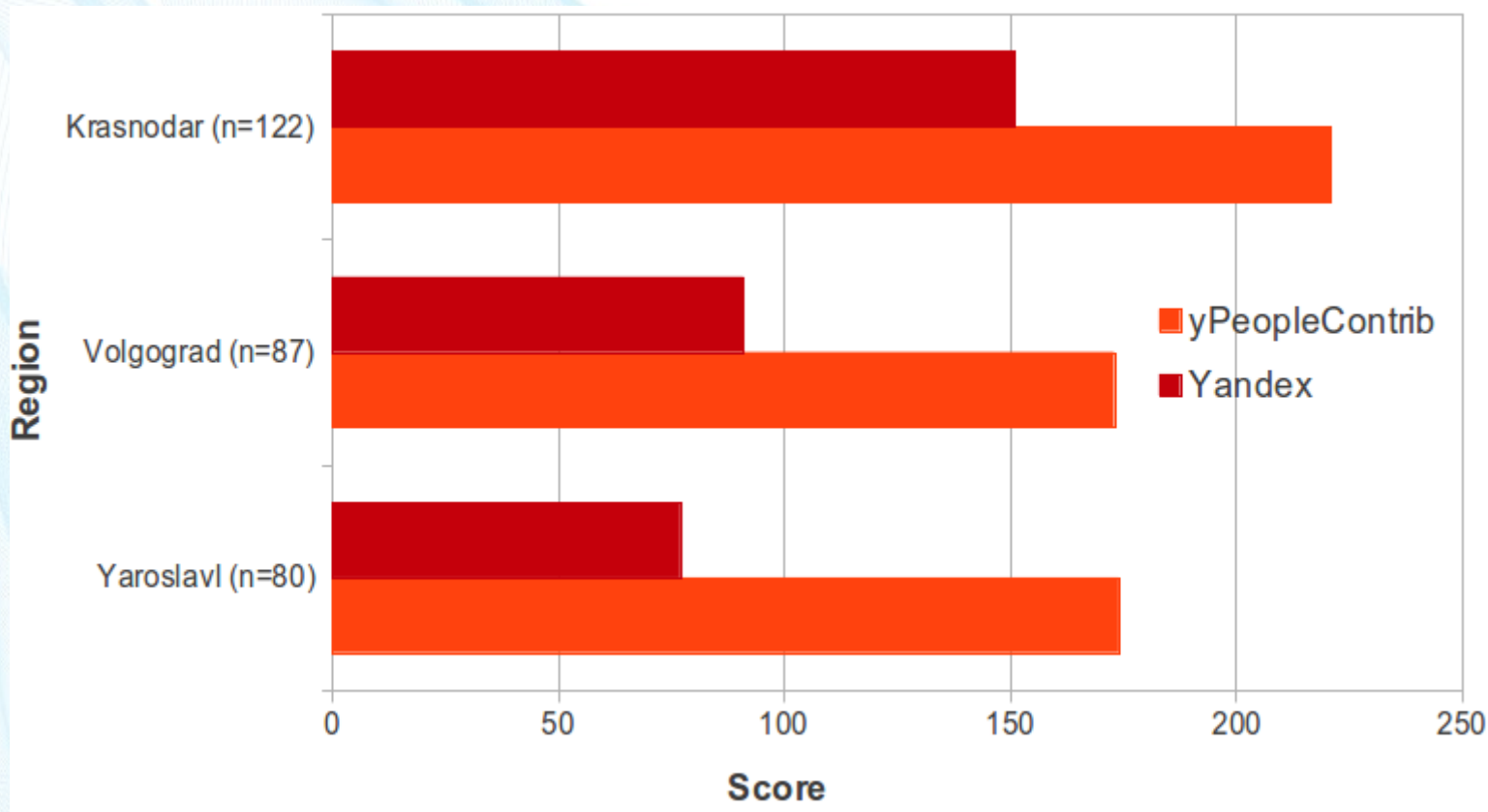
# OSM vs Yandex

- 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>

Check README.md for data sources