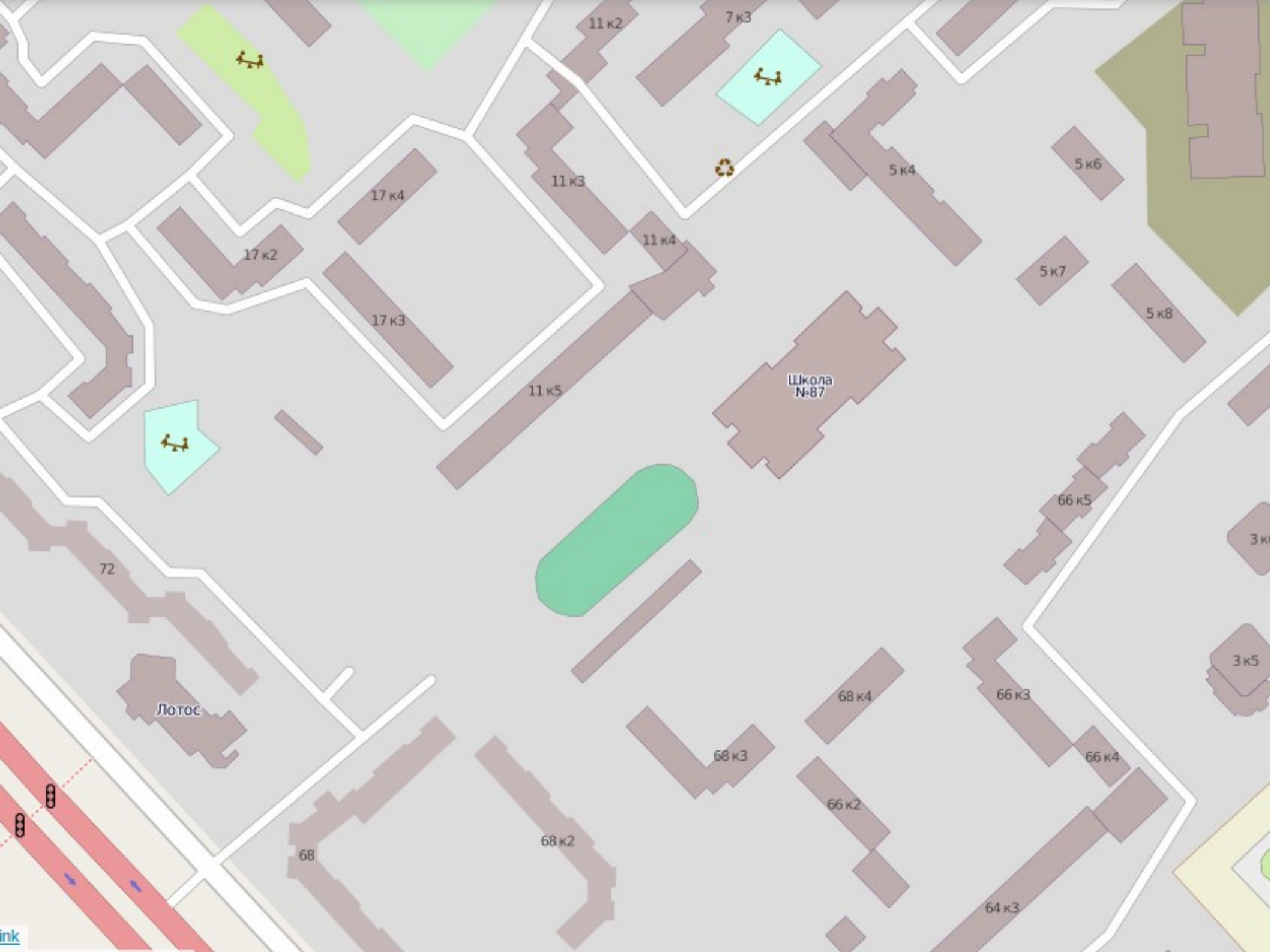
OpenStreetMap address base: ready for prime time?

Maxim Dubinin
sim@gis-lab.info

SotM Baltics 2013


11 к2
7 к3
11 к3
17 к4
17 к2
17 к3
11 к4
11 к5
5 к4
5 к6
5 к7
5 к8
Школа
№87
66 к5
72
Лотос
68 к4
68 к3
66 к3
66 к4
66 к2
68
68 к2
64 к3
3 к5

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

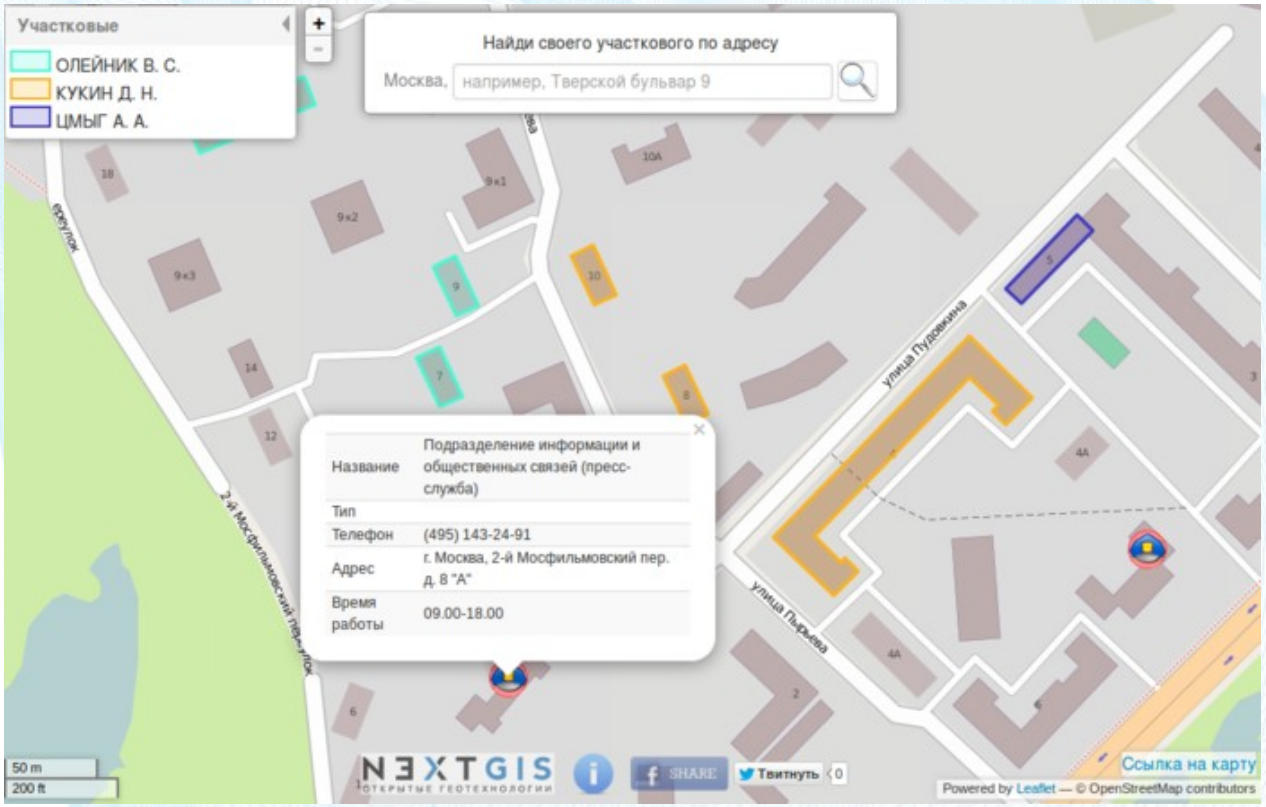
- OpenPolice — where are the local cops in Moscow
- Elections — where are the voting stations in Moscow
- Orphanages — where are the children orphanages in Russia

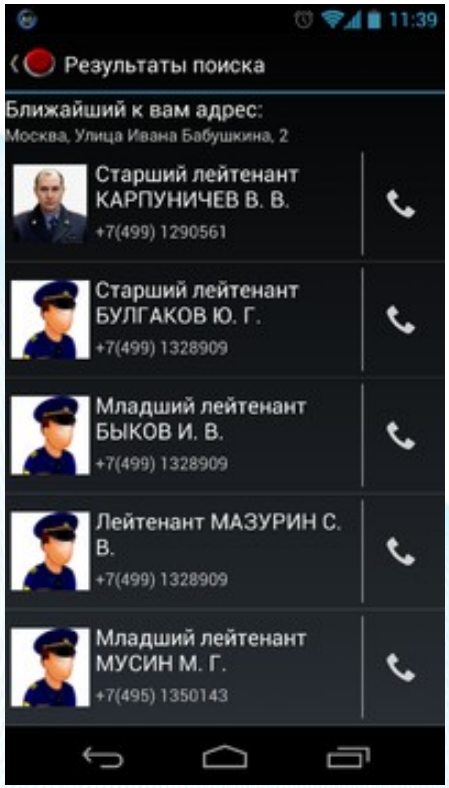
Question 1

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

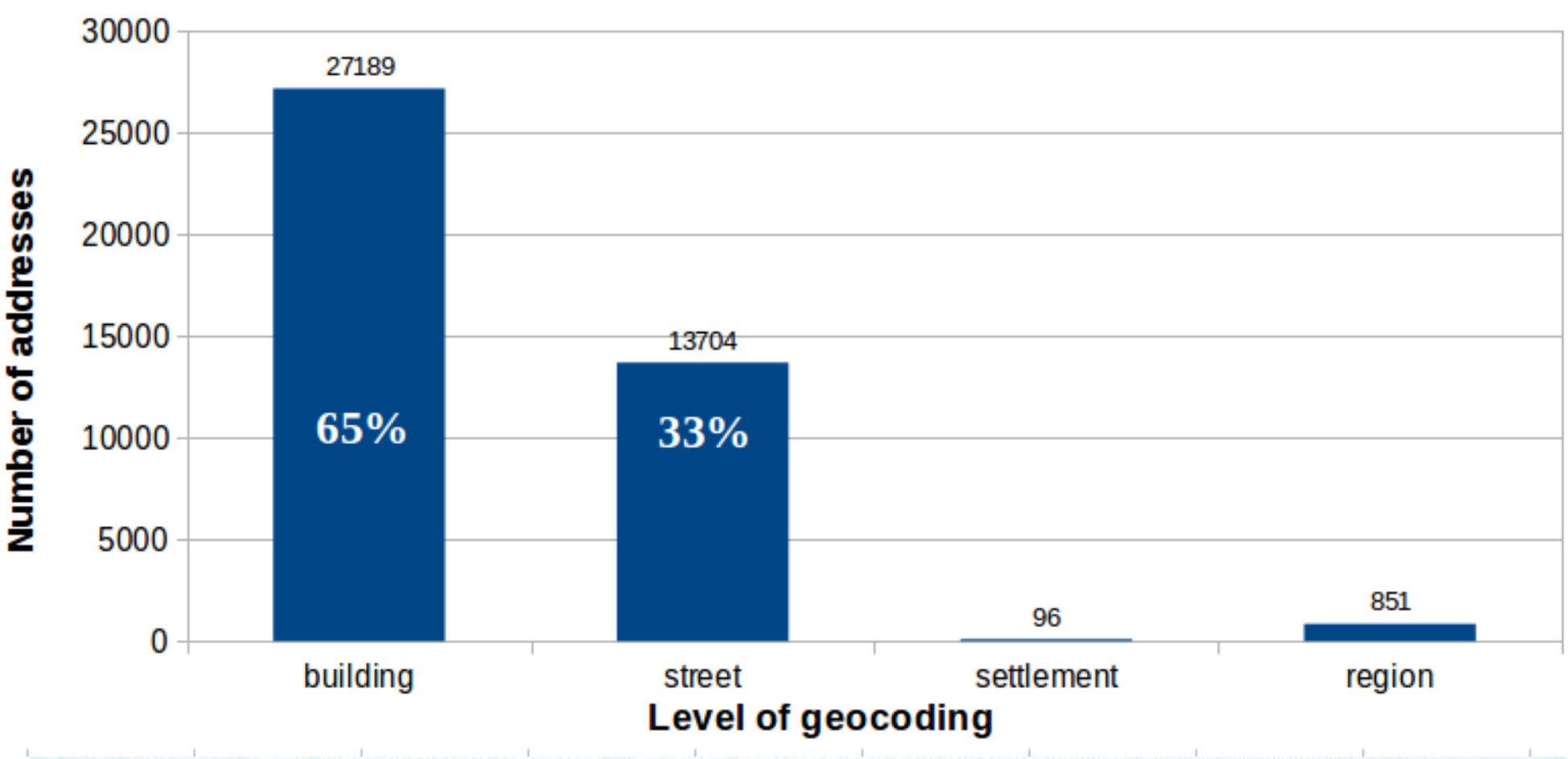




http://gis-lab.info/qa/openpolice.html

Results

- Total: ~41000 addresses in Moscow

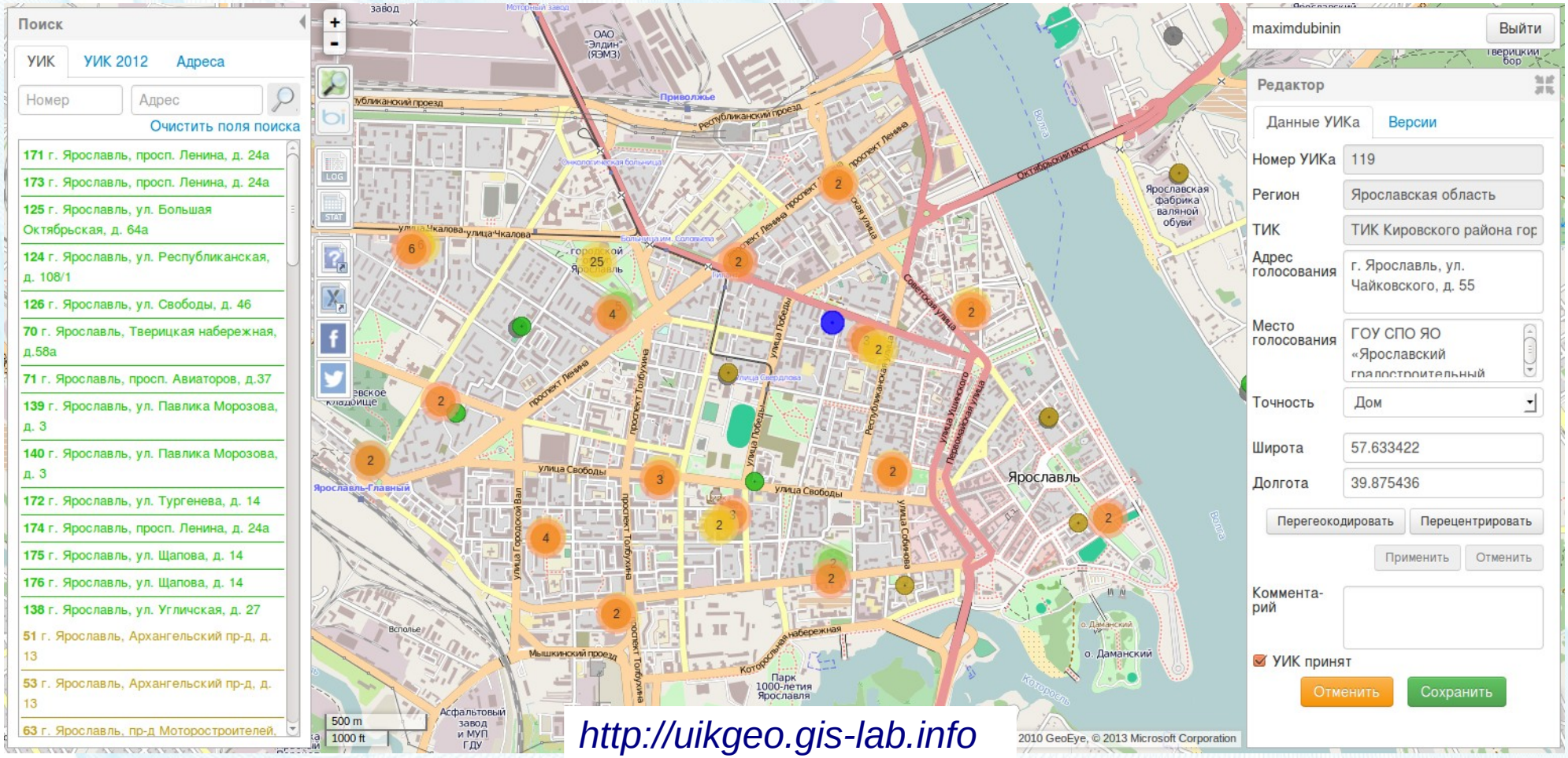


Question 2

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 comissions

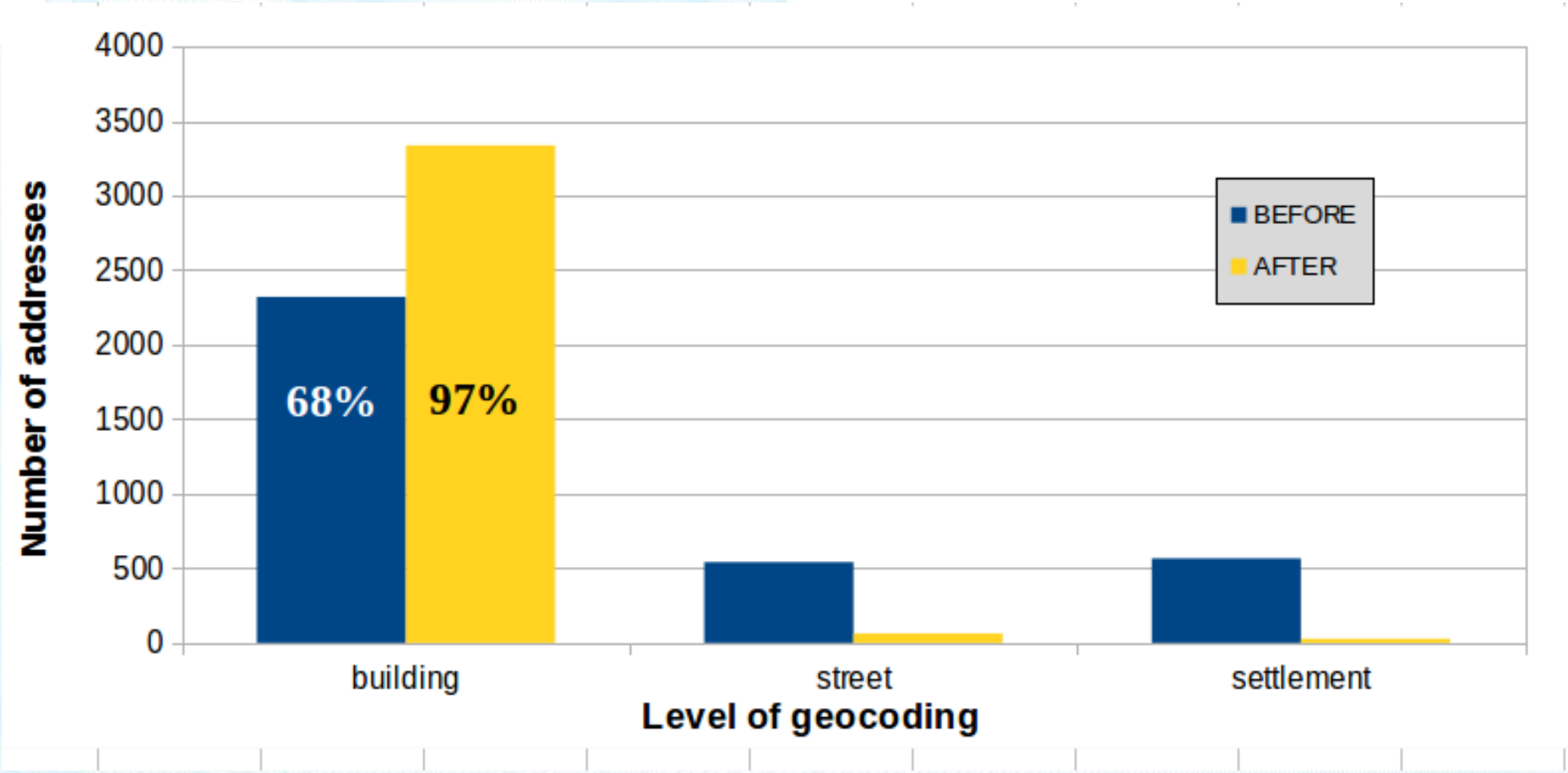
- Extract all addresses from public database
- Geocode them
- Crowdsource post-processing



http://uikgeo.gis-lab.info

Results

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



Question 3

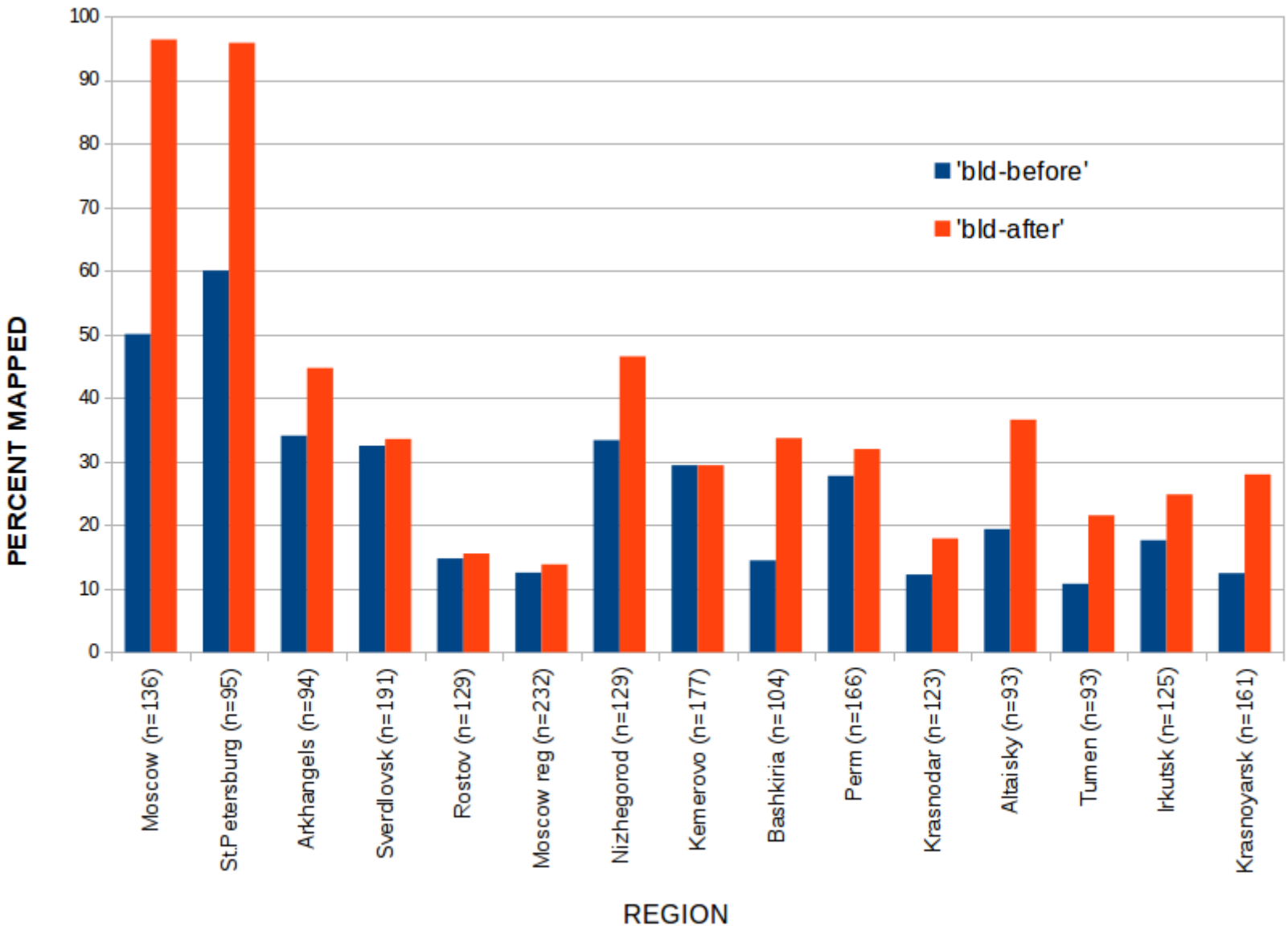
1. How good is OSM address database and fully automatic geocoding?
2. How much does postprocessing help?
3. **How is completeness distributed across Russia?**
4. How good is quality compared to other geocoders?

Orphanages

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

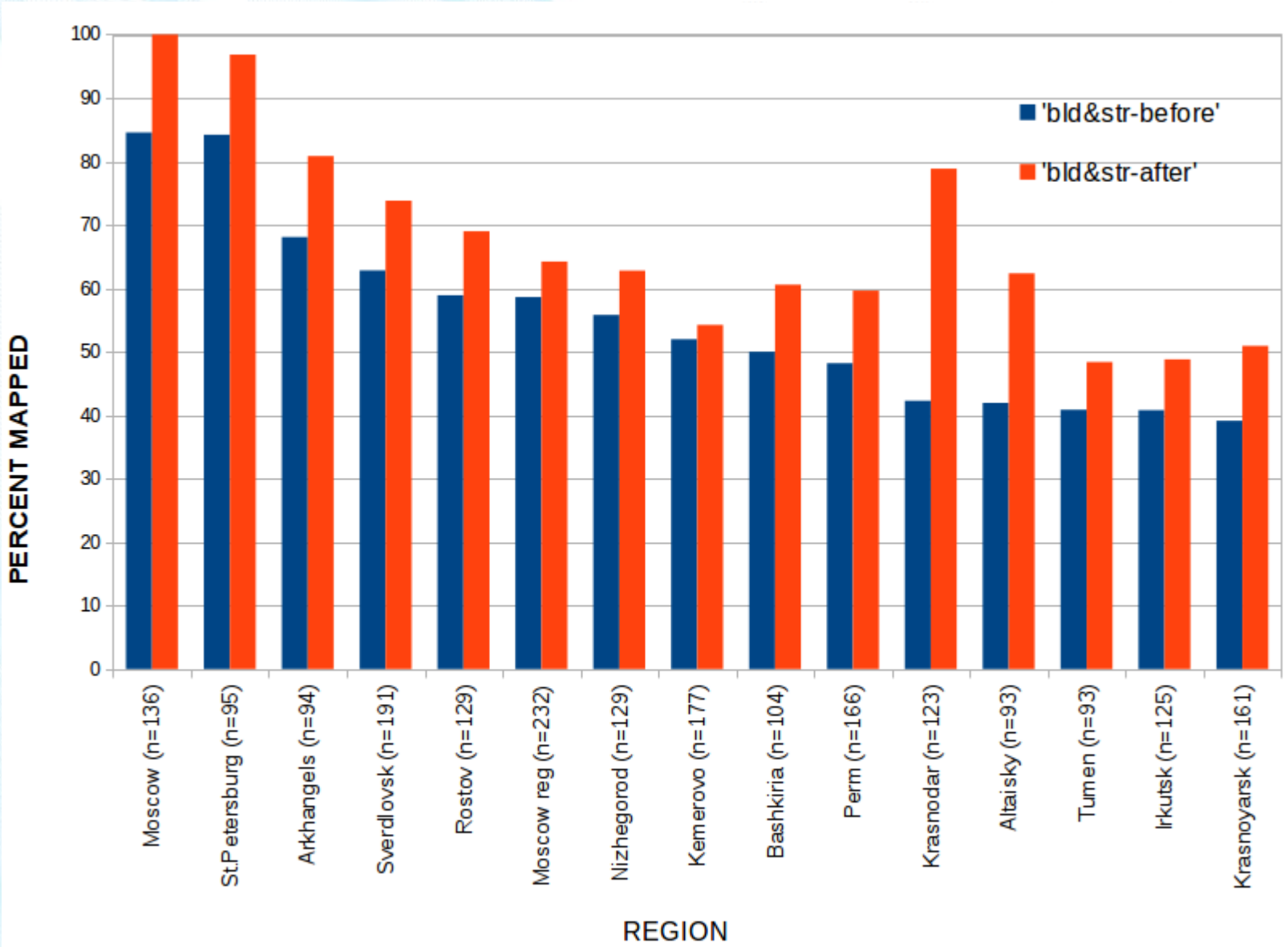
Orphanages

- Buildings before and after post-proc, % total



Orphanages

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



Question 4

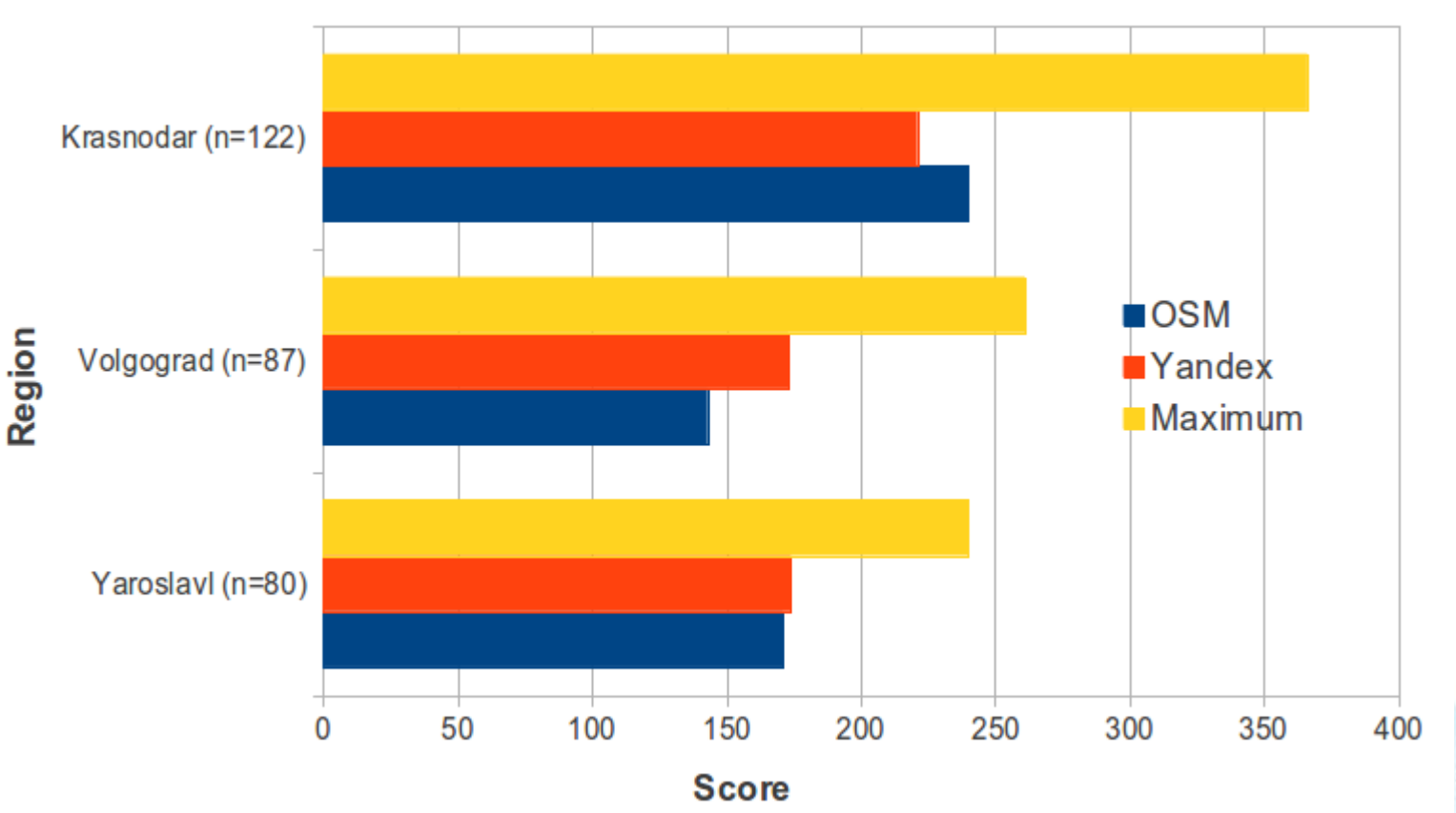
1. How good is OSM address database and fully automatic geocoding?
2. How much does postprocessing help?
3. How is quality distributed across Russia?
4. **How good is the quality compared to other geocoders?**

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

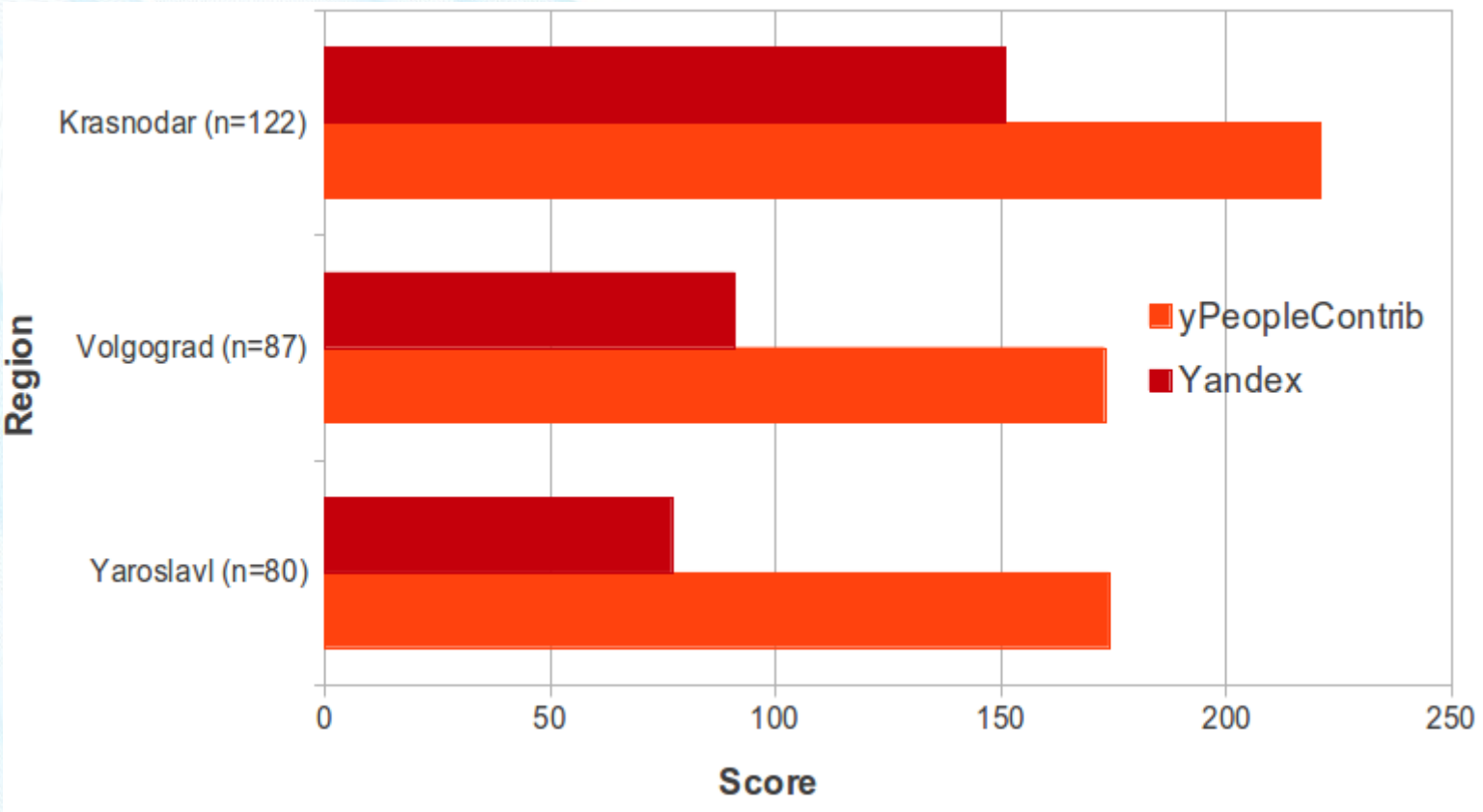
OSM vs Yandex

- 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