1. Problems encountered in the map

I downloaded OpenStreetMap data about Tartu County (Tartumaa in Estonian), a region in Estonia. After reading it into MongoDB database I found that this data contained several errors and inconsistencies. Here is the list of found problems:

- Inconsistent street name abbreviations (maantee, mnt, MNT or puiestee, pst)
- Inclusion of several different house numbers into one element in different tags (as housenumber:, housenumber2:, housenumber3 etc.)
- Inclusion of apartment numbers into house number tags (as Tammiku 70/12)
- Some errors in data entry (for example a cafe name in the street name field)
- Incompatibility of given tag names and predefined data model (Some elements contained a field named “addr:address”)

Inconsistent street name abbreviations:
There was two street name endings that were abbreviated in more than one way. All the different substrings were updated to match the most common form of abbreviations.

- Maantee (Road) was abbreviated as “mnt” (1016 cases), “MNT” (1 case), or left unabbreviated “maantee” (1 case).
- Puiestee (Avenue) was abbreviated as “pst” (784 cases) or left unabbreviated “puiestee” (12 cases).

Inclusion of several different house numbers into one element in different tags
Some elements contained more than one housenumber tags (housenumber, housenumber2, housenumber3, etc). All the house numbers were included into one list. Here is an example of address tags in one element:

[{'addr:city': 'Tartu'}, {'addr:country': 'EE'}, {'addr:housenumber': '17'}, {'addr:housenumber2': '19'}, {'addr:housenumber3': '17a'}, {'addr:street': 'Uus'}]

And here is the corresponding address data after shaping the element:
{'city': 'Tartu', 'street': 'Uus', 'housenumber': ['17', '19', '17a'], 'country': 'EE'}

Inclusion of apartment numbers into house number tags
Apartment numbers were added into some housenumber tags, while other housenumber tags contained only the number of respective building. I changed all the housenumber fields to contain only the number of the building and collected info about apartments into the Apartments field. Here is an example address data before and after that change:

Before:
{'city': 'Tartu', 'street': 'Ilmatsalu', 'housenumber': ['18c/1', '18c/2'], 'country': 'EE'}
After:
{'apartments': {'18c': ['1', '2']}, 'city': 'Tartu', 'street': 'Ilmatsalu', 'housenumber': ['18c'], 'country': 'EE'}

Errors in data entry
One street name field contained a name of a cafe (Soolo kohvik). It was changed to contain the actual street name (Oja).

Incompatibility of given tag names and predefined data model
Some elements contained a tag named “address” that did not fit into the current data model. This tag contained postal address of a given building so I moved it under address data with a “postal address” keyword.

Here is an example node that contained “address” tag:
<node id="687057027" lat="58.3735448" lon="26.7221638" version="6" timestamp="2014-06-24T17:00:29Z" changeset="23126805" uid="357111" user="enedaniel">
<tag k="addr:city" v="Tartu"/>
<tag k="address" v="Riia 12, 51013 Tartu"/>
<tag k="muuseum_id" v="194"/>
<tag k="name" v="Kaitseväe Ühendatud Õppeasutuste Muuseum"/>
<tag k="name:en" v="Museum of the Estonian National Defence College"/>
<tag k="opening_hours" v="Opened only when requested"/>
<tag k="phone" v="731 4161"/>
<tag k="source" v="http://www.muuseum.ee/et/muuseumid"/>
<tag k="tourism" v="museum"/>
<tag k="url" v="http://www.ksk.edu.ee/est/muuseum"/>
</node>

Here is this node after shaping the element:
{'tourism': 'museum', 'name:en': 'Museum of the Estonian National Defence College', 'name': u'Kaitsev\xe4e \xdchendatud \xd5ppeasutuste Muuseum', 'created': {'changeset': '23126805', 'user': 'enedaniel', 'version': '6', 'uid': '357111', 'timestamp': '2014-06-24T17:00:29Z'}, 'url': 'http://www.ksk.edu.ee/est/muuseum', 'muuseum_id': '194', 'pos': ['58.3735448', '26.7221638'], 'phone': '731 4161', 'source': 'http://www.muuseum.ee/et/muuseumid', **'address': {'city': 'Tartu', 'postal address': 'Riia 12, 51013 Tartu'}**, 'opening_hours': 'Opened only when requested', 'type': 'node', 'id': '687057027'}

2. Overview of the Data

This section contains basic overview of the dataset.
File sizes:
Tartumaa.osm 124,5 MB
Tartumaa.json 164,2 MB

Number of documents:

```
> db.tartumaa.find().count()
601319
```

```
Number of nodes:
> db.tartumaa.find({"type":"node"}).count()
529081
```

```
Number of ways:
> db.tartumaa.find({"type":"way"}).count()
72231
```

```
Number of distinct contributing users:
> db.tartumaa.distinct("created.user").length
217
```

```
Number one contributing user:
> db.tartumaa.aggregate([{"$group":{"_id":"$created.user",
"count":{"$sum":1}}},{"$sort":{"count":-1}},{"$limit":1}])
{ "_id" : "SviMik_import", "count" : 210289 }
```

3. Additional Ideas

It would be interesting to use address and geotag data of nodes to map density of buildings in different parts of the region. Such data could be especially useful for urban planning but also for planning of the land use in the countryside. However, this purpose would greatly benefit from some additional data like for example data about daily movements of people. The data about movements could be collected in collaboration with mobile operators and it would considerably increase the usefulness of the OpenStreetMap data for the mentioned task. Without some additional data, the type and usage of many building remains unknown, which complicates meaningful analysis.

Nevertheless, for the start I did some querying with mongoDB to find the streets in the region that have the most buildings.

```
 [{"$match" : {"address.street" : {"$exists" : 1}}},
             {"$unwind" : "$address.housenumber"},
             {"$group" : {"_id" : {"city": "$address.city",
                                   "street" : "$address.street"},
                                   "buildings" : {"$addToSet" : "$address.housenumber"}}},
              {"$unwind" : "$buildings"},
              {"$group" : {"_id":{"city" : "$_id.city",
```

```
                        "street" : "$_id.street"},
                    "nr_of_buildings" : {"$sum" : 1}}},
            {"$sort" : {"nr_of_buildings" : -1}},
            {"$limit" : 3}
    ]
```

```
[{u'_id': {u'city': u'Tartu', u'street': u'Võru'}, u'nr_of_buildings': 232},
 {u'_id': {u'city': u'Tartu', u'street': u'Riia'}, u'nr_of_buildings': 213},
 {u'_id': {u'city': u'Tartu', u'street': u'Tähe'}, u'nr_of_buildings': 177}]
```

And second I asked what cities (they are actually towns and villages) have the highest average number of buildings per street.

```
[{"$match":{"address.street":{"$exists":1}}},
            {"$unwind" : "$address.housenumber"},
            {"$group" : {"_id" : {"city": "$address.city",
                        "street": "$address.street"},
                        "buildings" : {"$addToSet" : "$address.housenumber"}}},
            {"$unwind" : "$buildings"},
            {"$group" : {"_id" : {"city" : "$_id.city",
                         "street" : "$_id.street"}er,
                         "nr_of_buildings" : {"$sum" : 1}}},
            {"$group" : {"_id" : "$_id.city",
                          "buildings_on_street" : {"$avg" : "$nr_of_buildings"}}},
            {"$sort" : {"buildings_on_street" : -1}},
            {"$limit" : 3}
]
```

```
[{u'_id': u'Tartu', u'buildings_on_street': 24.84510250569476},
 {u'_id': u'Ülenurme', u'buildings_on_street': 15.863636363636363},
 {u'_id': u'Elva', u'buildings_on_street': 14.661157024793388}]
```