

Article

# An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil

Silvana Philippi Camboim \*, João Vitor Meza Bravo † and Claudia Robbi Sluter †

Geodetic Science Graduate Program, Department of Geomatics, Federal University of Parana, Curitiba 81531970, Brazil; E-Mails: jvmbravo@gmail.com (J.V.M.B.); robbisluter@gmail.com (C.R.S.)

† These authors contributed equally to this work.

\* Author to whom correspondence should be addressed; E-Mail: silvanacamboim@gmail.com; Tel.: +55-41-3061-3636.

Academic Editor: Wolfgang Kainz

Received: 2 April 2015 / Accepted: 20 July 2015 / Published: 12 August 2015

---

**Abstract:** The integration of user-generated content made in a collaborative environment is being increasingly considered a valuable input to reference maps, even from official map agencies such as USGS and Ordnance Survey. In Brazil, decades of lack of investment has resulted in a topographic map coverage that is both outdated and unequally distributed throughout the territory. This paper aims to analyze the spatial distribution of updates of OpenStreetMap in rural and urban areas in the country to understand the patterns of user updates and its correlation with other economic and developmental variables. This analysis will contribute to generating the knowledge needed in order to consider the use of this data as part of a reference layer of the National Spatial Database Infrastructure as well to design strategies to encourage user action in specific areas.

**Keywords:** geoinformation quality; assessment; VGI; Brazil; integration.

---

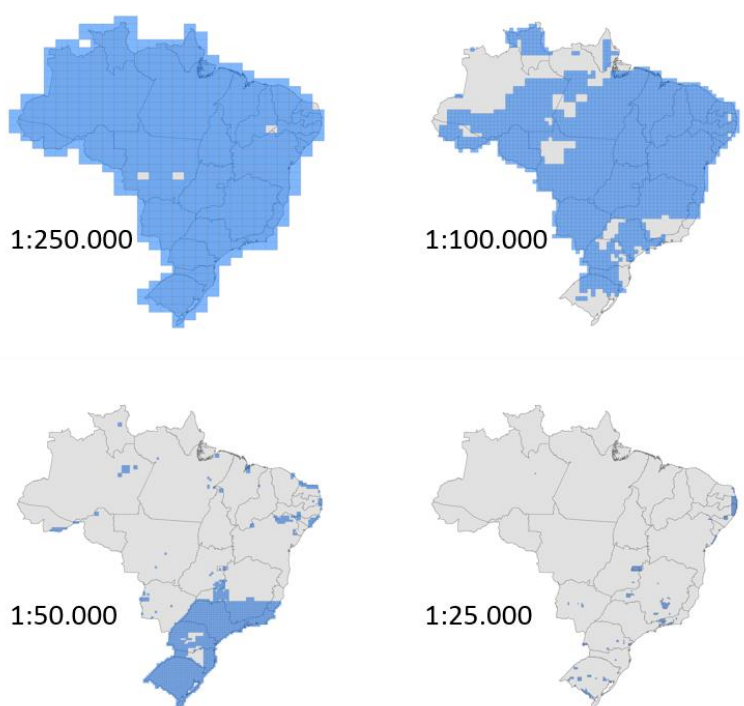
## 1. Introduction

Web technologies enable subjects without education in map design/production to become potential cartographers or “producers” [1–9]. Because many types of individuals are involved in the use and production of geoinformation, the Volunteered Geographic Information (VGI) has increased in importance

due to two main factors. The first one comprises the emergent natural interest of individuals in the use of web 2.0 media such as Facebook or OpenStreetMap generating content and disseminating their own information [3,6,7,10,11]. The second is the interest of official mapping agencies in updating their geodatabases with this rich crowd-sourced content [9–17]. It is this second factor which is our research motivation. In this case, the concerns in adopting VGI content for official purposes are related to the lack of methods that allow us to measure the reliability of this kind of data [18,19].

It is noticeable that the integration of user-generated content made in a collaborative environment is being increasingly considered a valuable input to reference maps, even from official map agencies such as USGS and Ordnance Survey [16,20,21]. While systems like OpenStreetMap, Wikimapia or Google Maps have triggered a powerful revolution—transforming “map users” into “map makers”—researchers and mapping agencies have observed VGI aiming to comprehend how these VGI systems could provide information for the official databases [16,17,21]. VGI systems could be a viable alternative to increase the speed of national information updating processes in developing countries, such as Brazil, where decades of a lack of investment has resulted in topographic map coverage that is both outdated and unequally distributed throughout the territory [22,23].

In 1994, Estes and Mooneyhan [22] presented an interesting critique of the situation of the national mapping coverage in developing countries. They used such impactful words when describing the non-existence of official geoinformation about these territories: “in many developing countries, even the most basic information related to resources and the environment does not exist” [22]. Within Brazil, the scenario is still the same: scarce investment in mapping agencies results in an outdated and unequally distributed map coverage for the Brazilian territory [23]. Figure 1 shows the topographic mapping coverage in Brazil by different scales. At a 1:250,000 scale, almost the whole territory is mapped. In contrast, at a 1:25,000 scale there are few maps available.



**Figure 1.** Map coverage in Brazil by scales. Adapted from Picanço Jr. and Delazari [24].

In Brazil, topographic mapping is a shared responsibility of both the Geographic Service of the Brazilian Army—DSG—and the Brazilian Institute of Geography and Statistics—IBGE. Brazil is the biggest country in South America with over 8.5 million·km<sup>2</sup>, producing an expensive scenario for mapping projects, especially for government funded ones.

Nevertheless, a connection between official map agencies and VGI systems will still depend on several factors such as quality tests and standardization [18,21]. In order to establish a first perspective about VGI quality, Haklay [21] has examined the positional accuracy of VGI content in the OpenStreetMap. He has found that the volunteered information is accurate to about 6 m, on average, of the position recorded by the Ordinance Survey in the UK territory [21]. After Haklay [21], several researchers have tested the quality of crowd-sourcing—in terms of positional as well as semantic accuracy—with similar findings [20,25–29]. However, what really matters in this case is that researchers and mapping agencies have repeatedly tested the Volunteered Geographic Information aiming to use this rich source [18] to gather both VGI and official initiatives [15].

The critical issue in VGI seems to be the evaluation of quality [18,19]. Moreover, there are such comprehensive efforts into investigating how reliable this content is because the reliability is considered a concern of quality control [16,18,19,30]. By reliability, we mean “the correctness or accuracy of the information” as stated by Comber *et al.* [19]. Goodchild and Li [18] argue that there are three alternatives to assure the quality of VGI content, in contrast with those well-known from the classical approach—Guptill and Morrison [31]—and employed by traditional mapping agencies such as Ordinance Survey and USGS. However, we have focused our attention on the second and the third of these alternatives, because they are relative to “the ability of a group to validate and correct the error that an individual might make”—a Linus’s Law approach—as well as possibly being the key element for the reliability in understanding [32].

Goodchild and Li [18] have suggested that this approach—the Linus’s Law [32]—can be applied to quality assurance of VGI. A similar idea is provided by Flanagan and Metzger [30] whose statement is relevant for the understanding of how to manage VGI content and assure its reliability, or, as they prefer, its “credibility”. They indicate that the more visited a place is, the more accurate the information about it will be. This means that, by taking into account these two points of view [18,30] we believe that the reliability of VGI can be observed and measured by the level of completeness and how often the database content is updated. In this case, areas with a high level of urbanization might be more visited in VGI systems, because individuals prefer to describe the geographic region in which they live or know something about, digitizing their personal experiences [33]—this sense is supported by the Topophilia concept [34]. Additionally, Haklay [21] pointed out some relevant results of a type of segregation phenomenon—the difference between the existence of geoinformation in areas with high urbanization level in contrast with those in rural areas—observing few users posting data in the OpenStreetMap. He indicates, “the centers of big cities in England (such as London, Manchester, Birmingham, Newcastle, and Liverpool) are well mapped” while suburban areas—especially the boundary between the city and rural areas—are not. Thus, Haklay [21] stated “it is important to know which areas are well covered and which are not—otherwise, the data can be assumed to be unusable” when someone is thinking about the integration of VGI and official databases. If these kind of discrepancies exist in countries with a tradition of reliable maps such as the United Kingdom, it is reasonable to expect that this heterogeneity will be a important issue in developing countries around the world such as Brazil. Figure 2 shows a comparative scenario between areas with high and low levels of urbanization in Brazil, as part of the OpenStreetMap

system. The visual contrast of features mapped in both situations (“a” and “b”) is relevant to us because it indicates similar findings to Haklay [21].



**Figure 2.** The first figure (a) shows an area with high urbanization level represented on OpenStreetMap, comprising the cities of Curitiba (South region of Brazil), S ão Paulo (Southwest region of Brazil), and Rio de Janeiro (Southwest region of Brazil). This first window (a) comprises over 25% of the total population of Brazil. In contrast, the second one (b) shows an area with a low level of urbanization—at the same scale as (a). This second and less populated area comprises two state capitals (Cuiab á and Goi ãnia) as well as the federal capital, Bras ília. Looking at the pictures one can see the large difference between the amounts of features mapped in both cases. This is likely the result of the “segregation phenomenon” described by Haklay [21]. Source: Adapted from OpenStreetMap, 2015.

While Brazil has quite recently reached over 200 million inhabitants [35] the distribution of this population over the large territory is extremely unequal as described by Carvalho [36]. The majority of Brazilians live in urban centers such as São Paulo, Rio de Janeiro, Curitiba, and other state capitals. Although the state capitals remain the largest cities, medium-sized cities have been the focus of demographic investigations because they have attracted internal migrants for reasons such as diversity in economic activities, decentralization of industry and better quality of life indicators [37–39]. Municipalities for the lowest level of administrative units after the federation and the states. The concept of “city” in Brazil, as defined by law, comprises the most representative urban agglomeration. As such, every municipality must incorporate at least one city and, in most cases, adjacent rural areas [40].

Therefore, an interesting research subject would be understanding which attributes of a geographic region with heterogeneous level of urbanization could have an association with the reliability of VGI content concerning these areas. Moreover, there are several countries, as we pointed out previously [22], which could benefit substantially from that understanding, once the reliability problem starts to be solved and VGI content can fill in the mapping coverage gaps.

Accordingly, the following sentence determines what we have addressed as the research problem in this paper: Do demographic and economic characteristics of a geographic region have a relationship with the level of completeness of, and the frequency of updates to, the VGI content? The hypothesis which we argue here is that the more editors working in a single region, the more likely it is to keep the mapping accurate dynamically over time. It follows that areas with a high level of urbanization are going to have the best reliability levels—in terms of completeness and how updated the data is. Moreover, a second premise we argue is that the level of completeness and how updated the content is in a VGI system might be measured by the representation of roads and buildings. We advocate this because when there are no maps at a suitable scale in a region, one of the first geographic features that individuals represent on a VGI system are roads and buildings—especially in the emergency situations given by Zook *et al.* [41] and Liu and Palen [14]. In addition, we argue that the method proposed is suited to work in areas where no other database is available to make comparisons, a situation commonly occurring in many developing countries.

Thus, this paper aims to understand the patterns of user updates and its correlation with other economic and developmental variables provided by the census data from IBGE. This type of research can lead to strategies to address the use of VGI considering local characteristics and needs, and use open data, standards and software to achieve the best spatial reference data in order to support much needed optimal decision-making processes. Besides this, such an analysis can contribute to generate the knowledge needed to consider the use of this data as part of a reference layer of the National Spatial Database Infrastructure as well to design strategies to encourage user action in specific areas. This paper describes initial efforts into investigating a way to assess the reliability of VGI content in Brazil—or other developing countries.

## 2. The Case Study

Considering how diverse and challenging the Brazilian territory is—a case of heterogeneity which is repeated in many developing countries [22]—we have selected a study field as part of a first attempt to understand how VGI data could provide benefits to the official map coverage [22]. The selected study

field comprises the Metropolitan Mesoregion of Curitiba (Figure 3) a good example of how diverse Brazil, or any other developing country, might be. This study field has high heterogeneous characteristics as there are large and small cities—urban areas of municipalities, areas with a high level of industrialization as well as mainly agricultural municipalities. There are also areas dependent on tourism, and areas protected due to their environmental importance have no urban occupation of any kind. We considered that this heterogeneity is a good first challenge to observe the usefulness of our hypothesis.

Furthermore, we have established this region because its municipalities have particular interest for the Federal University of Paraná community as our university is involved in developing local strategies that benefit the population in the surrounding areas.



**Figure 3.** The Metropolitan Mesoregion of Curitiba within the Brazil and Paraná State context.

The Metropolitan Mesoregion of Curitiba is a unit of Brazilian territory, although not an administrative area. The Brazilian Institute of Geography and Statistics—IBGE—has created mesoregions in Brazil grouping municipalities with the same characteristics in terms of proximity, population and economy, for statistical purposes [42]. The selected region comprises 37 municipalities (Table 1) all inside the State of Paraná (Figure 3).

It is necessary to highlight one more time (see Table 1.) that the area we have selected has high degree of heterogeneity in terms of population demographic, financial power distribution as well as the distribution of urban and rural population—and these are the characteristics which compose what we meant by heterogeneous indicators. The main city (Curitiba) accounts for over 50% of the total population of the whole Mesoregion and the 10 biggest municipalities account for almost 90% of the total inhabitants. There are cities achieving high or medium Human Development Index (HDI) levels (e.g., Curitiba, Pinhais, Paranaguá, São José dos Pinhais) and cities scoring low levels (e.g., Doutor Ulysses, Itaperuçu, Tunas do Paraná, Tijucas do Sul). Cities with good indicators, such as great GDP per capita and HDI, are also those ones with a higher urban population as well as being closer to the state capital of Curitiba. In addition, these cities also attend to have a variety of industries (e.g., Campo Largo, Araucária, Pinhais, and São José dos Pinhais).



The selected area also comprises a diverse range of natural resources projects with cities extracting minerals and petroleum (Guaratuba, Matinhos, Pontal do Paraná), coastal tourism (Guaratuba, Matinhos, Paranaguá, Pontal do Paraná), agricultural practices (Piên, Balsa Nova, Quitandinha), and industry activities (Curitiba, Campo Largo, Araucária, Pinhais). Table 1 provides more details about the presented scenario. Furthermore, Figure 4 shows the municipalities comprising the Metropolitan Mesoregion of Curitiba.



**Figure 4.** The selected municipalities comprising the Metropolitan Mesoregion of Curitiba.

**Table 1.** The selected municipalities comprising the Metropolitan Mesoregion of Curitiba.

Municipality	HDI	Distance to Curitiba (km)	Area-Urban (km <sup>2</sup> )	Area-Rural (km <sup>2</sup> )	Average Monthly Income Per Capita-Total (R\$)	GDP Per Capita-Year (R\$)	Pop.-Urban	Pop.-Rural
ADRIANÓPOLIS	0.44	127	1.32	1348.0	828.33	16,061.03	2060	4316
AGUDOS DO SUL	0.41	70	4.16	188.0	854.42	9024.28	2822	5448
ALMIRANTE TAMANDARÉ	0.55	15	81.48	113.0	1077.71	7486.27	98,892	4312
ANTONINA	0.51	77	13.26	869.0	1029.55	12,181.12	16,063	2828
ARAUCÁRIA	0.70	29	92.01	377.0	1355.90	109,142.87	110,205	8918
BALSA NOVA	0.54	52	9.65	339.0	1075.73	28,120.94	6870	4430
BOCAIUVA DO SUL	0.34	40	2.83	824.0	901.96	10,557.00	5128	5859
CAMPINA GRANDE DO SUL	0.62	33	51.56	488.0	1121.73	17,592.71	31,961	6808
CAMPO DO TENENTE	0.51	92	4.25	300.0	883.97	15,367.86	4194	2931
CAMPO LARGO	0.72	32	170.42	1079.0	1194.21	15,674.14	94,171	18,206
CAMPO MAGRO	0.56	20	12.97	262.0	927.01	8633.26	19,547	5296
CERRO AZUL	0.10	86	4.34	1337.0	623.26	14,239.09	4808	12,130
COLOMBO	0.68	20	76.58	121.0	1106.26	10,917.15	203,203	9764
CONTENDA	0.49	46	10.06	289.0	959.95	10,411.76	9231	6660
CURITIBA	1.00	0	435.04	*	2359.23	32,916.44	1,751,907	*
DOUTOR ULYSSES	0.00	163	3.12	83.0	549.43	15,613.28	929	4798
FAZENDA RIO GRANDE	0.63	26	33.81	2008.0	1158.41	8464.88	75,928	5747
GUARAQUEÇABA	0.15	181	11.80	1283.0	546.59	8810.23	2683	5188
GUARATUBA	0.62	143	43.35	302.0	1209.71	10,553.58	28,805	3290
ITAPERUÇU	0.33	29	12.86	2071.0	841.32	9621.06	19,956	3931
LAPA	0.58	69	22.84	372.0	992.92	19,840.67	27,222	17,710
MANDIRITUBA	0.39	43	6.86	75.0	885.71	12,546.80	7414	14,806
MATINHOS	0.71	115	42.50	672.0	1285.11	11,782.16	29,279	149
MORRETES	0.51	69	12.82	723.0	1146.32	8754.75	7178	8540



Table 1. Cont.

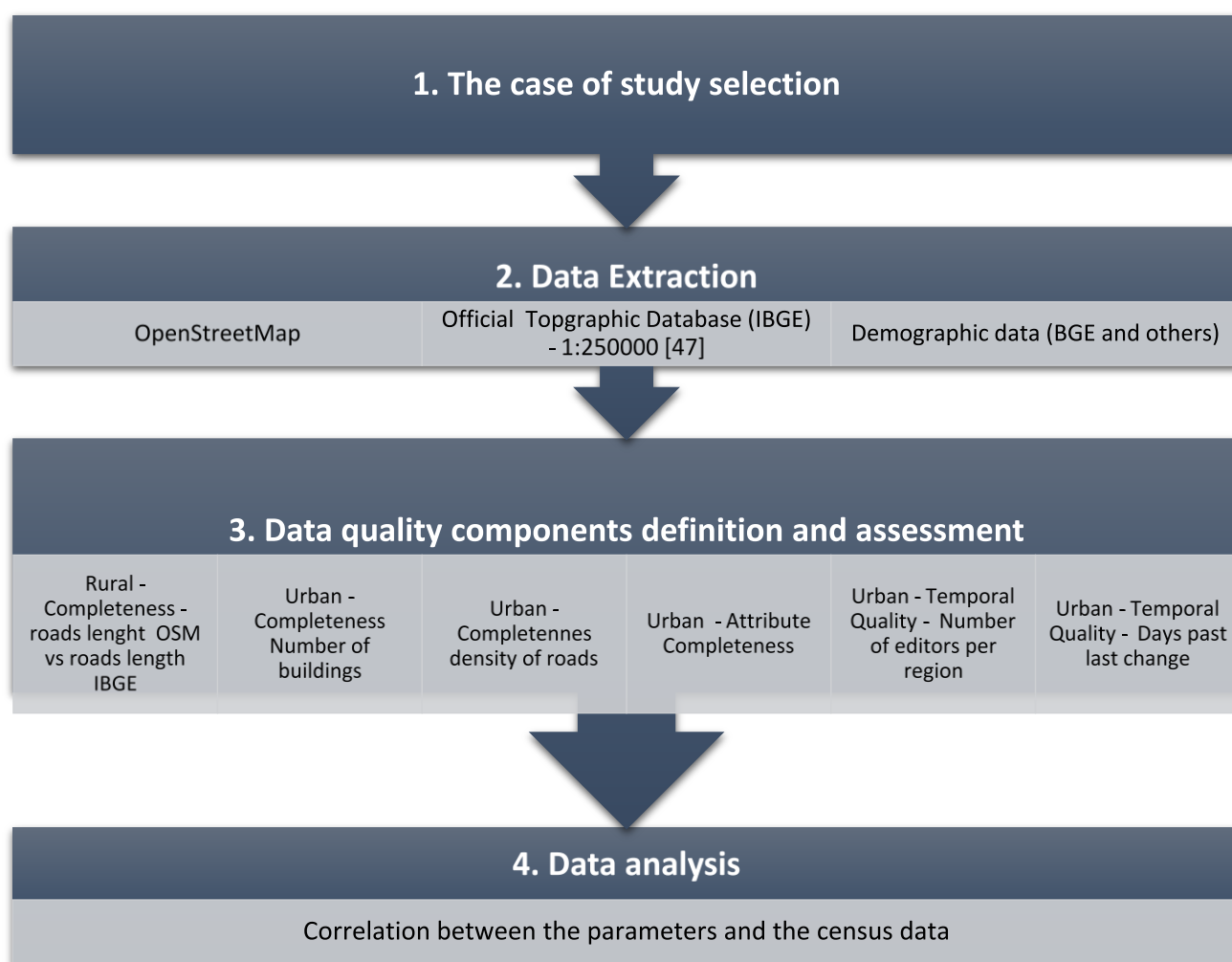
Municipality	HDI	Distance to Curitiba (km)	Area-Urban (km <sup>2</sup> )	Area-Rural (km <sup>2</sup> )	Average Monthly Income Per Capita-Total (R\$)	GDP Per Capita-Year (R\$)	Pop.-Urban	Pop.-Rural
PARANAGUÁ	0.74	91	103.21	248.0	1387.71	63,280.82	135,386	5083
PIÊN	0.53	86	7.18	186.0	859.58	25,531.37	4523	6713
PINHAIS	0.74	9	60.87	*	1441.60	26,054.72	117,008	*
PIRAQUARA	0.56	22	40.66	54.0	992.84	6878.75	45,738	47,469
PONTAL DO PARANÁ	0.69	100	146.23	181.0	1264.57	10,948.67	20,743	177
PORTO AMAZONAS	0.56	79	5.31	152.0	1088.70	11,653.75	2948	1566
QUATRO BARRAS	0.71	25	28.07	439.0	1331.65	35,003.47	17,941	1910
QUITANDINHA	0.48	67	8.03	793.0	774.43	9111.03	4887	12,202
RIO BRANCO DO SUL	0.48	31	19.33	562.0	967.88	19,231.09	22,045	8605
RIO NEGRO	0.77	117	42.29	845.0	1144.08	18,999.40	25,710	5564
SÃO JOSÉ DOS PINHAIS	0.77	14	101.40	670.0	1370.73	54,784.67	236,895	27,315
TIJUCAS DO SUL	0.32	64	1.75	667.0	906.86	15,086.31	2285	12,252
TUNAS DO PARANÁ	0.23	79	1.24	778.0	884.95	8179.20	2792	3464
Average	0.53	66	47	603	1062.98	19,704.23	86,469	8411

\* Curitiba and Pinhais do not have a rural area.

### 3. Methodology and Results

As a first attempt to explain how we have conducted this research work, Figure 5 demonstrates the workflow we have adopted. Accordingly, in this section we are going to describe all procedures illustrated in the Figure below.

In the first step in our research, we selected the case study with the aim of testing our procedures. The second step comprised a data extraction operation from the sources OpenStreetMap, the official Brazilian topographic database (from IBGE [43]), and demographic data (from IBGE [34,44] and others [45]). The OSM data used was an extract from Planet.osm [46] in 25 January 2015, along with the change sets available at JOSM on the same date. Geofabrik extracts large portions of OSM database and makes it available for download, while downloading directly from OSM database has a limit on download size. Demographic data, such as that for population and average household income, is from the 2010 census [35]. Human Development Index (HDI) is calculated by UNPD, also using 2010 data [45]. The municipal GDP is 2011 data from IBGE [44]. Census spatial data information, also from the IBGE [35] census, provided the basis for area calculations and for the division between official rural and urban areas, along with the official municipality boundaries. Topographic mapping only at 1:250,000 scale is available as a geodatabase from IBGE [43].



**Figure 5.** Workflow adopted.

As a third step for the research project, we defined the data quality components, which were going to compose the assessment and support the discussions, during the fourth step (data analysis). In this case, without detailed reference data, there is no knowledge supporting a specific methodology for assessing crowd-sourced geoinformation, so we have adopted two parameters of comparison from ISO: 19157 [47], regarding the main goal of this research work: completeness and temporal quality. For completeness, we have used four groups of evaluations: one case for rural areas and three cases for urban ones. For temporal quality, we have considered two topics. Table 2 shows the parameters and their definitions.

**Table 2.** Parameters and definitions.

Area	Data Quality Component	Measure
Rural	Completeness 1	Length of rural roads in OSM <i>versus</i> length of rural roads in official topographic database (1:250.000)
Urban	Completeness 2	Density of urban roads per square kilometre
Urban	Completeness 3	Number of buildings in each urban area
Urban	Completeness 4	Percentage of classified roads (with name or ref attribute)
Urban	Temporal quality 1	Number of days since last edition
Urban	Temporal quality 2	Number of editors

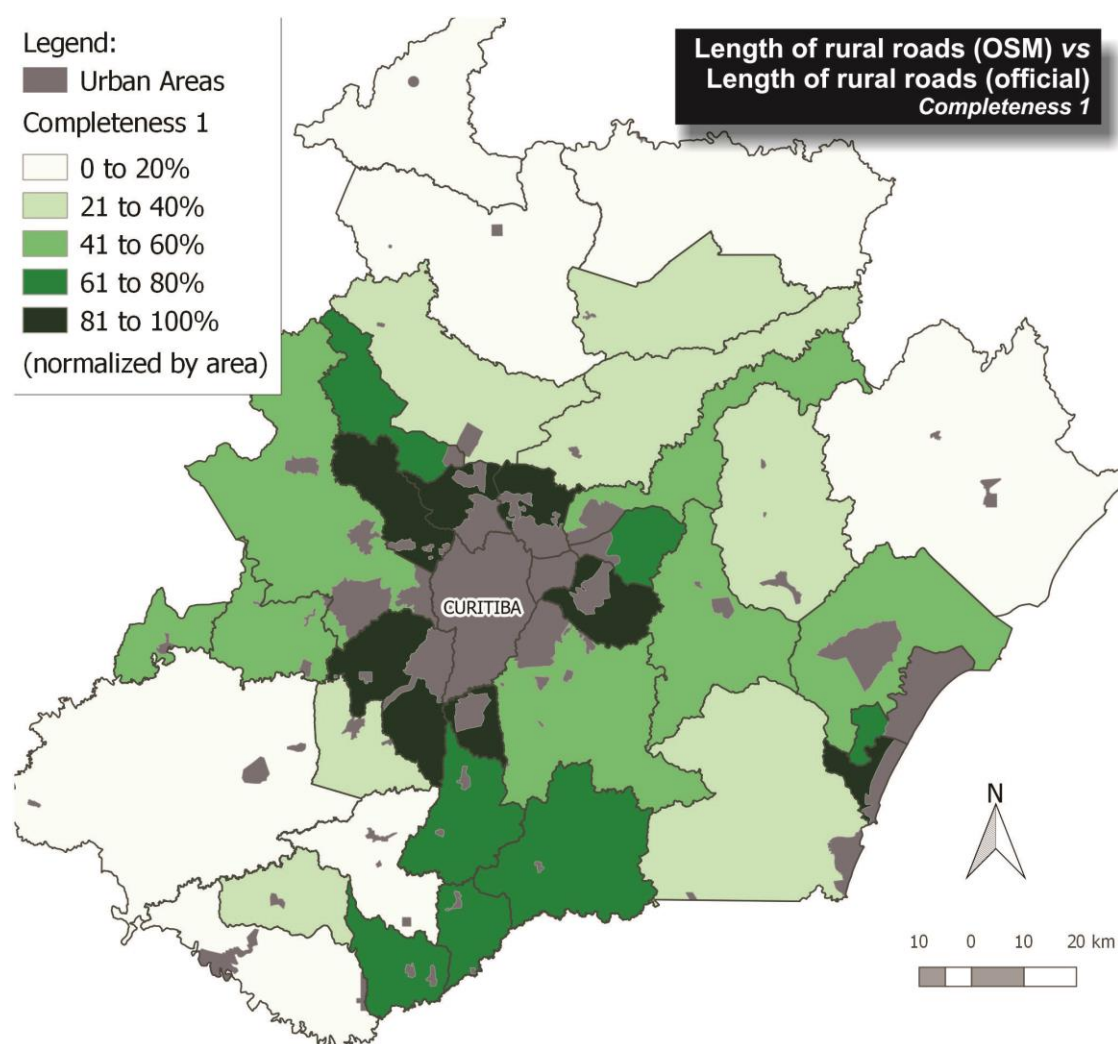
All parameters were designed to allow for a comparison between the municipalities, not as absolute quality measures. Therefore, as Table 2 shows, the first element studied was completeness in rural areas (completeness 1, at Table 2). This parameter was calculated dividing the total length of roads in official 1:250,000 topographic database road layers on the IBGE database by the total length of roads on OSM. This last parameter excluded footpaths and railroads that are stored in different layers on the IBGE database. Additionally, the major roads represented by double lines in OSM had their length divided by two, because the official database, at this scale, represents them as single lines. Figure 6 shows the results we have found applying this systematic approach; Table 3 shows the values of total length of roads in rural areas by municipality. Both representations (Figure 6 and Table 3) demonstrate that municipalities close to the capital (Curitiba) are more detailed, and this is the focus of our attention on the data analysis item for this topic.

Thinking about the completeness, although, for urban areas, we have considered three more cases (see Table 2, completeness 2–4). In this case, the approach chosen was due to the lack of an official database for urban areas in all municipalities. In fact, open data access for urban areas for the region comprising our study field has just been released for the state capital of Curitiba. For this reason, completeness in urban areas was evaluated in tandem with other methods, as there was not official geoinformation available. Therefore, the first criterion adopted was the density of urban roads per square kilometre (completeness 2, Table 2). This is maybe not the optimal solution, as urban areas are legally defined (in Brazil), and show an uneven distribution of urbanization between municipalities. In future studies, satellite images could provide a better assessment of urban patterns. A second criterion taken into consideration was the total number of features in the building layer of OSM in each urban area (completeness 3, Table 2). The third evaluation was the attribute completeness in urban areas (completeness 4, Table 2). This considered the percentage of roads with neither name attribute nor detailed description. This situation can be caused by the use of purely remotely sensed data, without

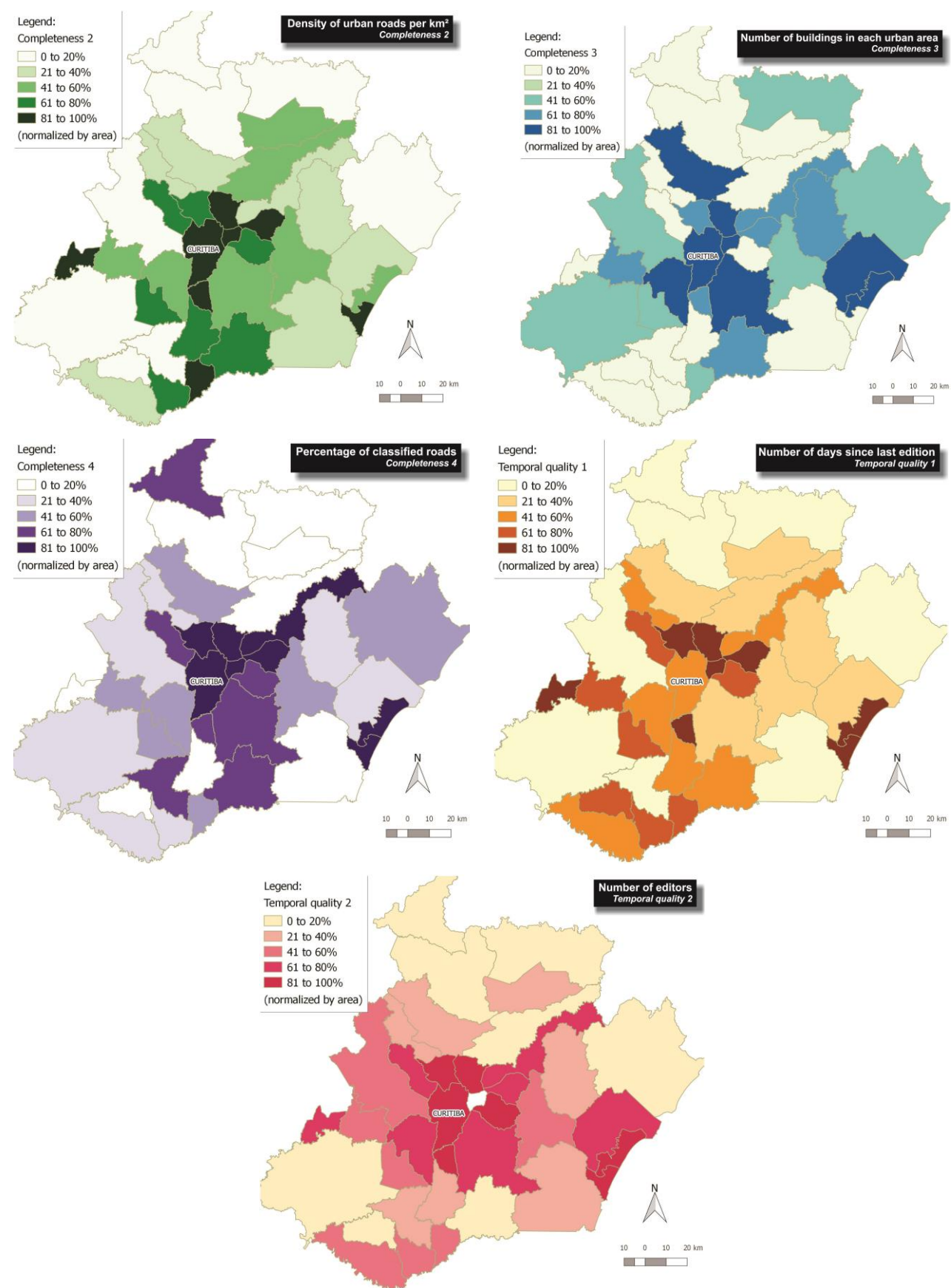
field knowledge of the feature attributes. Therefore, this parameter was calculated as the percentage of unclassified features in total features.

The next two elements refer to temporal quality in urban areas. This assessment involved the premise that the more editors working in a single region, the higher the likelihood of accurate mapping dynamically over time. The number was calculated by summarizing the editors that worked in change sets with data in urban areas supplied in the JOSM. The date of the last edition was also recorded to compute the number of days since the last change. All results in urban areas are described in Table 4 and Figure 7.

The fifth and final step in the research comprised the data analysis. In this case, we explored the data connecting the parameters and the demographic information aiming to comprehend the dynamics of user updates inside OSM and its correlation with other economic and demographic variables such as GDP per capita and HDI.



**Figure 6.** Completeness 1: Length of rural roads from OSM base vs. Length of rural roads from the official topographic mapping (at 1:250,000 scale).



**Figure 7.** Five data quality elements in urban areas.

**Table 3.** Total length of roads in rural areas by municipality.

Municipality	Rural Area Roads-Official (km)-(B)	Rural Area Roads- OSM (km)-(A)	(A)/(B)
ADRIANÓPOLIS	139,832	56,588	40%
AGUDOS DO SUL	30,436	59,549	196%
ALMIRANTE TAMANDARÉ	4327	117,513	2716%
ANTONINA	62,338	107,428	172%
ARAUCÁRIA	20,473	185,675	907%
BALSA NOVA	64,045	129,246	202%
BOCAIUVA DO SUL	77,211	158,321	205%
CAMPINA GRANDE DO SUL	90,159	233,983	260%
CAMPO DO TENENTE	21,722	16,033	74%
CAMPO LARGO	80,248	407,524	508%
CAMPO MAGRO	33,299	248924	748%
CERRO AZUL	145,910	82,494	57%
COLOMBO	10,007	107,824	1077%
CONTENDA	57,355	63,856	111%
CURITIBA	*	*	*
DOUTOR ULYSSES	57,355	43,294	75%
FAZENDA RIO GRANDE	4910	84,909	1729%
GUARAQUEÇABA	72,043	94,504	131%
GUARATUBA	139,077	213,542	154%
ITAPERUÇU	19,179	55,336	289%
LAPA	282,865	226,156	80%
MANDIRITUBA	72,900	286,504	393%
MATINHOS	8776	19,469	222%
MORRETES	74,794	242,709	325%
PARANAGUÁ	15,738	59,303	377%
PIÊN	12,618	30792	244%
PINHAIS	*	*	*
PIRAQUARA	38,358	233,070	608%
PONTAL DO PARANÁ	4161	4289	103%
PORTO AMAZONAS	41,116	59,728	145%
QUATRO BARRAS	31,339	63,968	204%
QUITANDINHA	67,950	31,776	47%
RIO BRANCO DO SUL	121,619	187,041	154%
RIO NEGRO	39,510	13,302	34%
SÃO JOSÉ DOS PINHAIS	127,582	780,866	612%
TIJUCAS DO SUL	89,952	1,008,408	1121%
TUNAS DO PARANÁ	158,677	310,261	196%
TOTAL	2,317,881	6,024,184	260%

\* Curitiba and Pinhais do not have rural areas.



**Table 4.** Urban data quality elements.

<b>Municipality</b>	<b>Completeness (Roads-km/km<sup>2</sup>)</b>	<b>Completeness (Number of Buildings/km<sup>2</sup>)</b>	<b>Completeness—Roads Attributes (% of Roads Unclassified)</b>	<b>Temporal Quality (Number of Editors)</b>	<b>Temporal Quality (Days since Last Edition)</b>
ADRIANÓPOLIS	6.10	0.0	68%	6	92
AGUDOS DO SUL	5.57	0.2	6%	5	100
ALMIRANTE TAMANDARÉ	3.80	0.0	37%	25	15
ANTONINA	4.81	1.7	7%	13	34
ARAUCÁRIA	3.57	0.9	28%	34	61
BALSA NOVA	5.34	1.1	20%	9	73
BOCAIUVA DO SUL	9.11	0.0	4%	6	92
CAMPINA GRANDE DO SUL	3.75	0.1	58%	17	28
CAMPO DO TENENTE	0.00	0.0	0%	0	
CAMPO LARGO	3.24	0.1	10%	33	8
CAMPO MAGRO	6.00	0.0	29%	16	66
CERRO AZUL	3.58	0.0	9%	3	9
COLOMBO	8.24	0.2	51%	41	32
CONTENDA	4.68	0.2	12%	7	55
CURITIBA	11.42	5.7	68%	201	3
DOUTOR ULYSSES	3.89	0.0	75%	3	673
FAZENDA RIO GRANDE	7.99	0.0	9%	13	11
GUARAQUEÇABA	2.49	0.1	84%	7	101
GUARATUBA	7.18	0.0	10%	20	17
ITAPERUÇU	1.65	0.0	3%	6	320
LAPA	3.48	0.7	20%	9	110
MANDIRITUBA	6.32	0.0	1%	7	103
MATINHOS	4.41	0.0	39%	25	15
MORRETES	5.25	0.2	29%	17	66
PARANAGUÁ	4.55	0.4	25%	26	49
PIÊN	4.41	0.0	2%	6	60

Table 4. Cont.

Municipality	Completeness (Roads-km/km <sup>2</sup> )	Completeness (Number of Buildings/km <sup>2</sup> )	Completeness—Roads Attributes (% of Roads Unclassified)	Temporal Quality (Number of Editors)	Temporal Quality (Days since Last Edition)
PINHAIS	7.35	0.3	55%	41	10
PIRAQUARA	5.05	0.0	22%	17	24
PONTAL DO PARANÁ	3.07	0.1	71%	22	15
PORTO AMAZONAS	4.33	0.0	0%	5	98
QUATRO BARRAS	4.77	0.2	31%	13	73
QUITANDINHA	0.70	0.0	33%	4	1685
RIO BRANCO DO SUL	5.27	5.8	55%	12	27
RIO NEGRO	3.16	0.0	5%	14	48
SÃO JOSÉ DOS PINHAIS	8.48	1.7	84%	59	18
TIJUCAS DO SUL	11.26	7.4	66%	6	8
TUNAS DO PARANÁ	10.31	0.0	1%	6	128
Average	5.26	1	30%	20	120

#### 4. Discussion

As stated earlier, the purpose of this research is to analyse the spatial distribution of updates of OpenStreetMap through rural and urban areas in Brazil to understand the patterns of user updates and its correlation with other economic and development variables. Up until this point, we have presented the results of the quality parameters we have investigated. However, the main focus is to compare these results with those demographic parameters shown in Table 1 in order to explore the possible existence of a correlation between these variables. In this case, we have calculated the Pearson coefficient of correlation Equation (1) as below. From this calculation, we have obtained the data presented in Table 5, which shows the correlation between the quality parameters and the demographic data. For the purposes of this analysis, we considered a strong correlation to be above 0.70, moderate between 0.40 and 0.69 and weak below 0.39.

$$\rho = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}} \quad (1)$$

Regarding the previous points about the strength of correlation, the first distinct correlation analysed was the completeness of the road layer in rural areas is with the total population (0.43) and with distance from Curitiba (−0.45). This suggests, although to a moderate degree, that the most populated areas, nearer the capital, have higher density of roads mapped in OSM as the representation (Figure 6) indicated. More specifically, cities adjacent to Curitiba with a large population such as Araucária, Almirante Tamandaré, Colombo, Fazenda Rio Grande and Tijucas do Sul, have almost 10 times (1000%) the length of roads the official (IBGE 1:250,000) database store. On the other hand, the smaller and more isolated municipalities such as Adrianópolis, Campo do Tenente and Doutor Ulysses, have fewer roads represented in OSM than on official topographic maps. Flanagan and Metzger [30] have suggested something similar: the more visited a place is, the more accurate the information about it is. In other words, we have proved that the best-mapped municipalities—in terms of roads in rural regions—are also the more populated municipalities.

The urban quality parameters did not exhibit any uniform behaviour from the data collected. However, in four of the five tests, the higher correlation shown suggests a weak association with population, which leads to a possible understanding that the more populated area, the more data is available and the more often it is updated, as stated before.

Analyzing the road completeness in urban areas, it is noticeable that some areas returned higher scores in highly dense urbanized areas, such as Curitiba (11.42) and São José dos Pinhais (8.48), with some outliers in small cities as Tunas do Paraná and Boca Úva do Sul. Even so, the higher correlation (0.69) is with population density, which is consistent with the nature of this parameter. The main issue here is that the legal designation of urban areas is not always consistent among distinct local authorities and, therefore, does not necessarily imply uniform urban patterns.

**Table 5.** Correlation between data quality observations and demographic data.

	V1	V2	V3	V4	V5	V6	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
<b>V1. Completeness (R)</b>	1																	
<b>V2. Completeness—Roads (U)</b>	0.25	1																
<b>V3. Completeness—Buildings (U)</b>	0.13	0.49	1															
<b>V4. Completeness—Roads attributes (U)</b>	0.07	0.2	0.39	1														
<b>V5. Temp. Editors (U)</b>	0.31	0.45	0.44	0.36	1													
<b>V6. Temp. Days (U)</b>	−0.2	−0.3	−0.1	0.06	−0.2	1												
<b>D1. HDI</b>	0.17	0.14	0.12	0.07	0.6	−0.3	1											
<b>D2. Distance</b>	−0.5	−0.2	−0.2	0.15	−0.4	0.14	−0.4	1										
<b>D3. Area (U)</b>	0.28	0.29	0.36	0.33	0.94	−0.2	0.64	−0.2	1									
<b>D4. Area (R)</b>	0	−0	−0	−0	−0.1	0.06	−0.2	0.04	−0.1	1								
<b>D5. Income</b>	0.19	0.39	0.34	0.19	0.85	−0.3	0.86	−0.4	0.84	−0.2	1							
<b>D6. GDP per capita</b>	0.05	0.01	0.15	0.15	0.28	−0.1	0.39	−0.2	0.3	−0.2	0.43	1						
<b>D7. Population (U)</b>	0.43	0.43	0.47	0.29	0.97	−0.1	0.48	−0.5	0.9	−0.1	0.77	0.21	1					
<b>D8. Population (R)</b>	0.11	0.11	0.11	0.08	0.3	0.01	0.09	−0.4	0.18	−0.1	0.02	0.07	0.33	1				
<b>D9. Population—total</b>	0.43	0.43	0.47	0.29	0.97	−0.1	0.48	−0.5	0.9	−0.1	0.77	0.21	1	0.46	1			
<b>D10. Population Density (U)</b>	0.37	0.69	0.37	0.14	0.69	−0.2	0.32	−0.5	0.53	0.26	0.53	0.16	0.69	0.19	0.69	1		
<b>D11. Population Density (R)</b>	0.09	0.04	−0.1	−0.1	0.05	−0.1	0.02	−0.2	0.02	−0.3	−0	−0.1	0.07	0.81	0.18	0.04	1	
<b>D12. Population Density—total</b>	0.2	0.44	0.38	0.31	0.91	−0.1	0.49	−0.4	0.81	−0.4	0.74	0.16	0.9	0.57	0.9	0.7	0.65	1

\*U—Urban Areas, R—Rural Areas.

The third urban data quality parameter was attribute completeness. This was the parameter which presented the weakest correlation with demographic variables. The strongest correlation was with population (0.29). The distribution has a much dispersed pattern with mainly small cities with almost no attributes on the streets at all (Campo do Tenente, 0%, Itaperuçu, 3%, Mandirituba, 1% and Porto Amazonas 0%). However, some of the less populated places have a very high rate, such as Doutor Ulysses with 75%, Guaraqueçaba with 84% and Pontal do Paraná with 71%. This could be due to individual efforts to actually provide attribute information in these places.

The next two observations aimed to compare the temporal aspects between areas. The individual contributors are highly concentrated in Curitiba, 200 (around 0.01% of the population). The next municipalities in number of contributors are São José dos Pinhais (54), Colombo and Pinhais (41 each), which are, respectively, the second, third and fifth largest cities. This is observed also in correlation of 0.97, a strong one, that shows that editors are more abundant in bigger cities, in proportion to an average 0.1% of the urban population.

The last urban analysis considered the number of days since the last edition. The idea here was that an active community would make more often additions to the database, keeping it updated and implying greater temporal quality. Fourteen cities had editions in the month prior to the study, such as Curitiba, São José dos Pinhais and Campo Largo. On the opposite side, Doutor Ulysses, Quitandinha and Itaperuçu did not have any updates at all in a year or more. These cities are among the poorest and more isolated areas. In fact, although the correlation of this parameter was not particularly strong, in this case, HDI and Average Income appears to have a stronger influence than the population.

In the maps in Figure 7, we can observe the distinct patterns of spatial distribution of the urban parameters proposed. Although, from a first look, they seem very different from each other, we can observe that Curitiba, São José dos Pinhais and Colombo, the more populated areas, are often in the higher class of each parameter. Doutor Ulysses, Tunas do Paraná and Porto Amazonas, and other smaller, poorer or isolated areas, with some exceptions, are mostly in the lower data quality classes. Instead of mathematically creating a formula of how these parameters behave, this study achieved its aim of analysing the specifics among the areas.

## 5. Conclusions

The aim of this paper was to observe the distribution of OpenStreetMap data in a significantly diverse region in Brazil. Comparing various parameters, it was observed that this distribution is uneven and concentrated mainly in areas with the largest population. One relevant point is that in these urban areas we could not actually measure absolute parameters of quality, as there were no available datasets to use as field truth. Instead, we had to compare a number of indicators of data quality observed with homogenous criteria in the 37 municipalities studied.

In order to consider the use of OSM data as inputs in official spatial databases, these quality issues must be addressed. When analyzing urban areas, the places where data is more abundant and more often updated are also the same cities with resources to invest in mapping initiatives. The places where these data are much needed and both financial and human resources are more scarce, the contributors might need some incentives or a specific call to concentrate these efforts as such spots seem not to naturally fall into the scope of volunteer map makers. This study showed that, even without official databases

available to assess absolute quality parameters, comparisons could be made between distinct areas that show that VGI alone cannot be the answer for providing data in poorer and isolated areas, the one area where a lack of official maps is more significant.

To expand this initial approach, a future research agenda could observe aspects as positional accuracy, correlation with other variables and studies in statistical significance of the correlation, including spatial statistics techniques. It is also important to define the thresholds of acceptable quality parameters in order to consider this data as part of official databases. An issue that could enhance the understanding the VGI updates is the peculiarities of tourist areas, such as municipalities in the coastal and Serra do Mar region. These areas seem to have an increased number of contributions, which could be due to visitors, not only the local population, but this effect was not part of the present study.

Open data in conjunction with open standards and software can help local authorities and the population to manage their space more efficiently considering social and economic factors. VGI information can play an important role in this process, once we understand its nature and the quality aspects related to it.

### Author Contributions

João Vitor Meza Bravo wrote the manuscript and was responsible for the context and the literature review. Silvana Philippi Camboim designed and performed the experiment. Claudia Robbi Sluter was responsible to structure and revise the manuscript.

### Acknowledgments

The authors thank the editor of the International Journal of Geo-Information and the anonymous reviewers for their valuable comments.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Fairbairn, D.J. The frontier of cartography: Mapping a changing discipline. *Photogramm. Record* **1994**, *14*, 903–915.
2. Peterson, M.P. Trends in internet map use. In Proceedings of the 18th International Cartographic Conference, Stockholm, Sweden, 23–27 June 1997.
3. Turner, A. Neogeography—Towards a Definition. Available online: <http://highearthorbit.com/neogeography-towards-a-definition/> (accessed on 10 August 2012).
4. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
5. Slocum, T.A.; McMaster, R.B.; Kessler, F.C.; Howard, H.H. *Thematic Cartography and Geovisualization*, 3rd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2009.
6. Heipke, C. Crowdsourcing geospatial data. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 550–557.



7. Griffin, A.L.; Fabrikant, S.I. More maps, more users, more devices means more cartographic challenges. *Cartogr. J.* **2012**, *49*, 298–301.
8. Budhathoki, N.R.; Bruce, B.C.; Nedovic-Budic, Z. Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal* **2008**, *72*, 149–160.
9. Haklay, M.; Singleton, A.; Parker, C. Web mapping 2.0: The neogeography of Geoweb. *Geogr. Compass* **2008**, *2*, 2011–2039.
10. Cormode, G.; Krishnamurthy, B. Key differences between Web 1.0 and Web 2.0. *First Monday* **2008**, *13*, doi:10.5210/fm.v13i6.2125.
11. Jackson, S.P.; Mullen, W.; Agouris, P. Crooks, A. Croitoru, A.; Stefanidis, A. Assessing completeness and spatial error of features in Volunteered Geographic Information. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 507–530.
12. O'Reilly, T. What is Web 2.0: Design patterns and business models for the next generation of software. *Commun. Strateg.* **2007**, *65*, 18–37.
13. Coleman, J.D.; Georgiadou, Y.; Labonte, J. Volunteered Geographic Information: The nature and motivation of producers. *Int. J. Spat. Data Infrastruct. Res.* **2009**, *4*, 332–358.
14. Liu, S.; Palen, L. The new cartographers: Crisis map mashups and the emergence of neogeographic practice. *Cartogr. Geogr. Inform. Sci.* **2010**, *37*, 69–90.
15. Johnson, P.A.; Sieber, R.E. Motivations driving government adoption of the Geoweb. *GeoJournal* **2012**, *77*, 667–680.
16. Bearden, M.J. The National Map Corps, 2007. Available online: [www.ncgia.ucsb.edu/projects/vgi/docs/position/Bearden paper.pdf](http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Bearden%20paper.pdf) (accessed on 15 October 2013).
17. Anand, S.; Morley, J.; Jiang, W.; Du, M.; Hart, G. Jackson, M. When worlds collide: combining Ordnance Survey and Open Street Map data. In Proceedings of 2010 AGI GeoCommunity, London, UK, 28–30 September 2010.
18. Goodchild, M.F.; Li, L. Assuring the quality of volunteered geographic information. *Spat. Stat.* **2012**, *1*, 110–120.
19. Comber, A.; See, L.; Fritz, S.; van der Velde, M.; Perger, C.; Foody, G.M. Using control data to determine the reliability of volunteered geographic information about land cover. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 37–48.
20. Brown, G. An empirical evaluation of the spatial accuracy of public participation GIS (PPGIS) data. *Appl. Geogr.* **2012**, *34*, 289–294.
21. Haklay, M. How good is a Volunteered Geographical Information? A comparative study of OpeenStreetMap and Ordnance Survey datasets. *Environ. Plan. B Plan. Des.* **2010**, *37*, 682–703.
22. Estes, J.E.; Mooneyhan, D.W. Of Maps and Myths. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 517–524.
23. Camboim, S.P.; Sluter, C.R. The National Topographic Mapping as an indispensable database for a Brazilian National Spatial Data Infrastructure (NSDI). In Proceedings of the 24th International Cartographic Conference, Santiago, Chile, 15–21 November 2009.
24. Picanço, P.L., Jr.; Delazari, L. Proposta de interfaces para um sistema VGI para usuários iletrados. In Proceedings of 2015 CPGCG-UFPR, Curitiba, Brazil, 23 June 2015.
25. Leeuw, J.; Said, M.; Ortegah, L.; Nagda, S.; Georgiadou, Y.; Debolis, M. An assessment of the accuracy of volunteered road map production in Westerns Kenya. *Remote Sens.* **2011**, *3*, 247–256.

26. Parker, C.J. A Human Factors Perspective on Volunteered Geographic Information. Ph.D. Thesis, Loughborough University, Leicestershire, UK, 2012.
27. Kessler, C.; Groot, R.T.A. Trust as proxy measure for the quality of Volunteered Geographic Information in the case of OpenStreetMap. In *Geographic Information Science at the Heart of Europe*; Vandenbroucke, D., Bucher, B., Cromptvoets, J., Eds.; Springer: Cham, Switzerland, 2013; pp. 21–37.
28. Forghani, M.; Delavar, M.R. A quality study of the OpenStreetMap dataset for Tehran. *ISPRS Int. J. Geo Inf.* **2014**, *3*, 750–763.
29. Brown, G.; Weber, D.; de Bie, K. Is PPGIS good enough? An empirical evaluation of the quality of PPGIScrowd-sourced spatial data for conservation planning. *Land Use Policy* **2015**, *43*, 228–238.
30. Flanagan, A.J.; Metzger, M.J. The credibility of volunteered geographic information. *GeoJournal* **2008**, *72*, 137–148.
31. Guphill, S.C.; Morrison, J.L. *Elements of Spatial Data Quality*; Elsevier: Amsterdam, The Netherlands, 1997; p. 201.
32. Wilkinson, D.M.; Huberman, B.A. Assessing the Value of Cooperation in Wikipedia. 2007. Available online: <http://arxiv.org/abs/cs/0702140> (accessed on 01 March 2015).
33. Kelley, M.J. The emergent urban imaginaries of geosocial media. *GeoJournal* **2011**, *78*, 181–203.
34. Tuan, Y-Fu. Language and the making of place: A narrative-descriptive approach. *Ann. Assoc. Am. Geogr.* **1991**, *81*, 684–696.
35. Instituto Brasileiro de Geografia e Estatística, IBGE. Censo Demográfico 2010. Available online: <http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm> (accessed on 1 March 2015).
36. Carvalho, J.A.M. Demographic dynamics in Brazil: Recent trends and perspectives. *Braz. J. Popul. Stud.* **1998**, *1*, 5–24.
37. Henderson, V. Medium size cities. *Reg. Sci. Urban Econ.* **1997**, *27*, 583–612.
38. Birdsall, N.; Sinding, S.W. How and why population matters: New findings, new issues. In *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*; Birdsall, N., Kelley, A.C., Sinding, S.W., Eds.; Oxford University Press: Oxford, UK, 2001; pp. 181–203.
39. Mccann, E.J. “Best places”: Interurban competition, quality of life and popular media discourse. *Urban Stud.* **2004**, *41*, 1909–1929.
40. Lima, M.H.P.; Rodrigues, C.M.; Silva, J.K.T.; Martins, P.C.; Terron, S.L.; Silva, R.L.S. Divisão Territorial Brasileira. Available online: [http://www.ipeadata.gov.br/doc/DivisaoTerritorialBrasileira\\_IBGE.pdf](http://www.ipeadata.gov.br/doc/DivisaoTerritorialBrasileira_IBGE.pdf) (accessed on 01 May 2015).
41. Zook, M.; Graham, M.; Shelton, T.; Gorman, S. Volunteered geographic information and crowdsourcing disaster relief: A Case study of the Haitian Earthquake. *World Med. Health Policy* **2010**, *2*, 7–33.
42. Instituto Brasileiro de Geografia e Estatística, IBGE. Divisão Regional do Brasil em Mesorregiões e Microrregiões Geográficas. Available online: [http://biblioteca.ibge.gov.br/visualizacao/monografias/GEBS%20-%20RJ/DRB/Divisao%20regional\\_v01.pdf](http://biblioteca.ibge.gov.br/visualizacao/monografias/GEBS%20-%20RJ/DRB/Divisao%20regional_v01.pdf) (accessed on 01 March 2015).

43. Instituto Brasileiro de Geografia e Estatística, IBGE. Base Cartográfica na Escala 1:250,000, 2013. Available online: [ftp://geofp.ibge.gov.br/mapeamento\\_sistematico/base\\_vetorial\\_continua\\_escala\\_250mil/Documentacao\\_bc250\\_v1.0.pdf](ftp://geofp.ibge.gov.br/mapeamento_sistematico/base_vetorial_continua_escala_250mil/Documentacao_bc250_v1.0.pdf) (accessed on 1 March 2015).
44. Instituto Brasileiro de Geografia e Estatística, IBGE. Produto Interno Bruto dos Municípios Brasileiros, 2011. Available online: <http://ibge.gov.br/home/estatistica/economia/pibmunicipios/2011/default.shtm> (accessed on 1 March 2015).
45. Programa das Nações Unidas para o Desenvolvimento, PNUD. Atlas do Desenvolvimento Humano do Brasil. 2013. Available online: [http://www.atlasbrasil.org.br/2013/pt/o\\_atlas/o\\_atlas/](http://www.atlasbrasil.org.br/2013/pt/o_atlas/o_atlas/) (accessed on 1 March 2015).
46. Geofabrik. Excerpts and Derived Data From the OpenStreetMap Dataset. Available online: <http://www.geofabrik.de/data/download.html> (accessed on 25 January 2015).
47. ISO TC-211. *ISO: 19157: 2013—Geographic Information—Data Quality*. International Organization for Standardization: Geneva, Switzerland, 2013; p. 146.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).