

Big data and Housing Demography: Differences among Apartment Announcement Dynamics in a medium size Brazilian inner-city

Gustavo Pedroso de Lima Brusse¹

Leandro Becceneri²

Abstract

Alternative data sources can potentially provide new approaches applications, enrich traditional analysis or fill blanks in conventional data sources, such as Census and cross-sectional surveys. In Household and Family Demography, although important family events can be well understood by conventional data sources, they cannot provide information on household dynamics during the time and limited information is available on housing. An innovative data source alternative emerges from mobile and web device applications (apps) that collect household data for the purposes of selling and renting apartment announcements. It can be seen as a follow-up study that provides unique information on housing dynamics and the household spatial distribution inside metropolitan urban areas. Under the hypothesis of high socioeconomic population segregation in a medium size Brazilian city (Campinas-SP), an Event History Analysis using Discrete Time Proportional Hazard Model (PH) will be applied to explore whether there are statistical significant differences between apartment announcement dynamics among different inner-city regions.

Keywords: Housing Market, Urban spatial distribution, Big Data, Proportional Hazard Model, Apartment Announcement Dynamics.

¹ Universidade Estadual de Campinas (NEPO/IFCH - UNICAMP); e-mail: gustavo.brusse@gmail.com

² Universidade Estadual de Campinas (NEPO/IFCH - UNICAMP); e-mail: leandrobecc@hotmail.com

Introduction

Nowadays, alternative data sources can potentially provide new approaches applications, enrich traditional analysis or fill blanks in conventional data availability. It can also be applied regarding Household and Family Demography studies. Usually, conventional data sources, such as census and household surveys can only state about what happens inside the group that share the same household at the moment of the survey in a cross-sectional approach (CAVENAGHI and ALVES, 2011). Important family events that affect household formation and dissolution can be well understood by these traditional data sources, including the increasing time spent in the parents' home, the decreasing number of children, changes in marital status pattern and the increasing longevity. However, cross-section approaches cannot provide information on household dynamics during the time and limited information is available on housing. Mulder and Lauster (2010) explicit the close relation between family and housing dynamic as "family is as much a context for understanding housing needs and residential outcomes as housing is a context for understanding family events" (MULDER and LAUSTER, 2010).

An innovative data source alternative emerges from mobile and web device applications (apps) that collect household data for the purposes of selling and renting apartment announcements. Individuals who want to sell their apartments are able to login in the apps, fill in information about the dwelling characteristics and post the announcements online. Once having exposure, time, entry and exit information, and geographic location, it can be seen as a follow-up study that provides unique information on housing dynamics and the household spatial distribution inside metropolitan urban areas. Here, it will be used to better understand the inner-city household spatial distribution differences.

The observed household spatial distribution in the metropolitan region depends not only on demographic issues, mostly covered by conventional data sources, but also on a complex context of intra-urban dynamics, that includes housing market, individual choices, economics and environmental constraints. In Brazilian society, the urban space production is also strongly associated with spatial, social-economic and race segregation. In that context, the housing dynamics is part of a larger process of urban expansion and can produce other forms of space segregation or perpetuate traditional forms of socio-spatial distributions.

This uneven socio-spatial distribution in the urban area can be clearly seen in Campinas city, Brazil. The third most populous municipality in São Paulo state and the

fourteenth in the entire country, it has an estimated population of 1.204.073 inhabitants in 2019 (IBGE, 2020). It has a large urban area formed by a very contrasting and dynamic areas mosaic, as highlighted by Cunha and Jakob (2010), Dota (2015) and Cunha (2016). Geography and physical barriers, such as the Anhanguera state highway, not only contributed to shape the Campinas urban landscape, but also influenced the socioeconomic population distribution. As an example, the regions on the south and southeast of the city, which are also in the south of the Anhanguera highway, have a younger population, with higher average households' size and with lower socioeconomic conditions.

Using an alternative data source from an apartment announcement app, considered as “Big data”, the aim of this study is to explore whether there are statistical significant differences between apartment announcement dynamics among different inner-city regions in Campinas, under the hypothesis of socioeconomic population segregation.

Materials and methods

Data was collected from DataZAP³ Company which followed thousands⁴ of selling and renting apartment announcements in Campinas city from January 2014 to April 2019, during 64 months (Figure 1). The announcements' neighborhoods were grouped by regions with similar socio-demographic characteristics. An Event History Analysis using Discrete Time Proportional Hazard Model (PH) will be applied to estimate the announcement's probability of surviving according to dwelling characteristics variables and socio-demographic characteristics variables by geographic location: South and North Anhanguera. The model can be described as follows:

$$h(t) = h_0(t) * e^{X'\beta}$$

Where, $h_0(t)$ is the is the renting apartment hazard based on parametric distribution, X is the time-varying covariates matrix and β is the vector of estimated parameters of the model.

³ Authorized by Grupo ZAP© in 02/2020

⁴ The total number of announcements cannot be exposed due to confidentiality terms.

Results

Preliminary descriptive statistics points out great differences among regions (South Anhanguera and North Anhanguera), such as price per square meter and average household size distribution (Figure 2).

Figure 1 - Spatial distribution of apartment announcements in Campinas, Brazil (2014-2019)

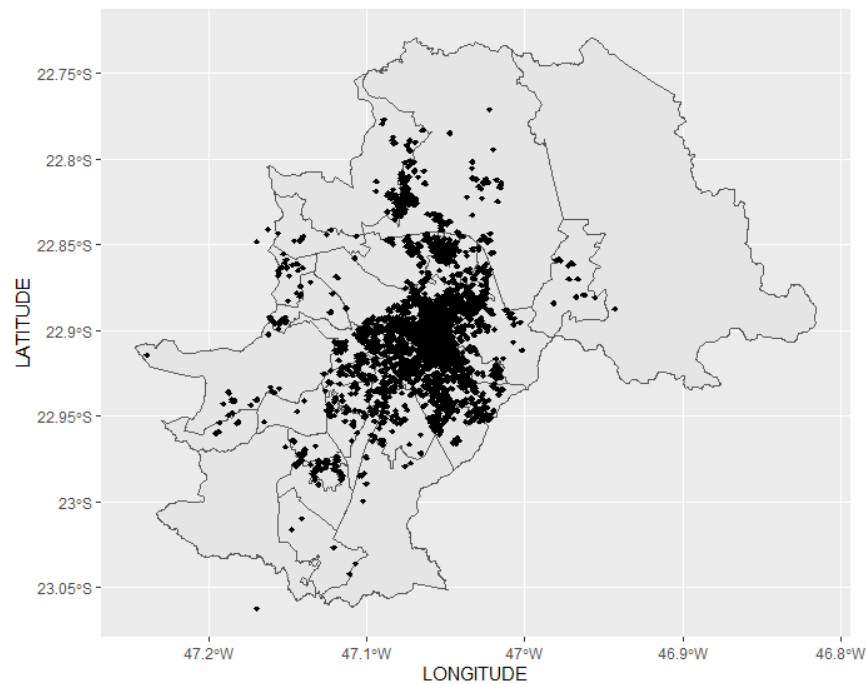
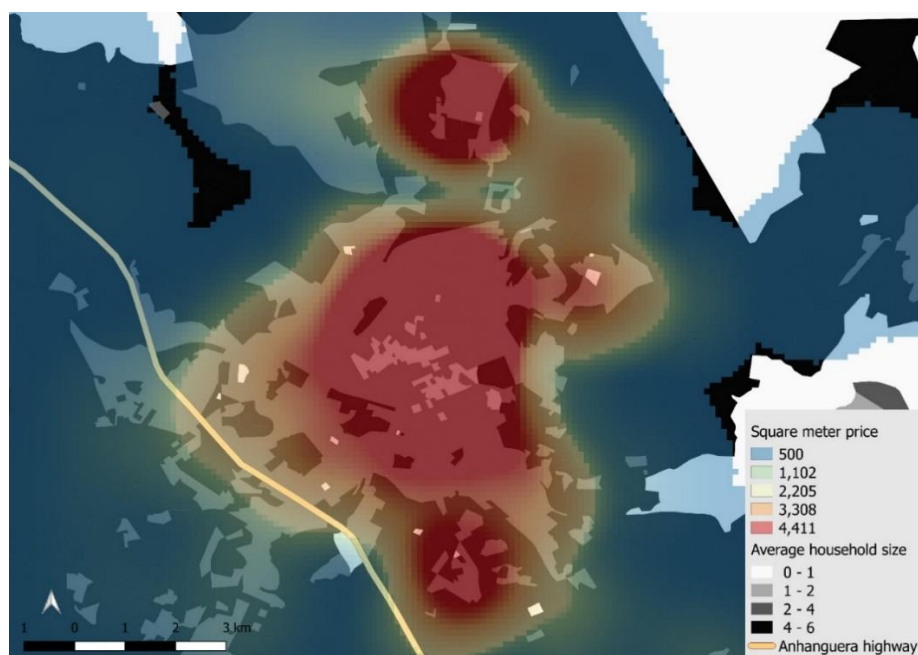


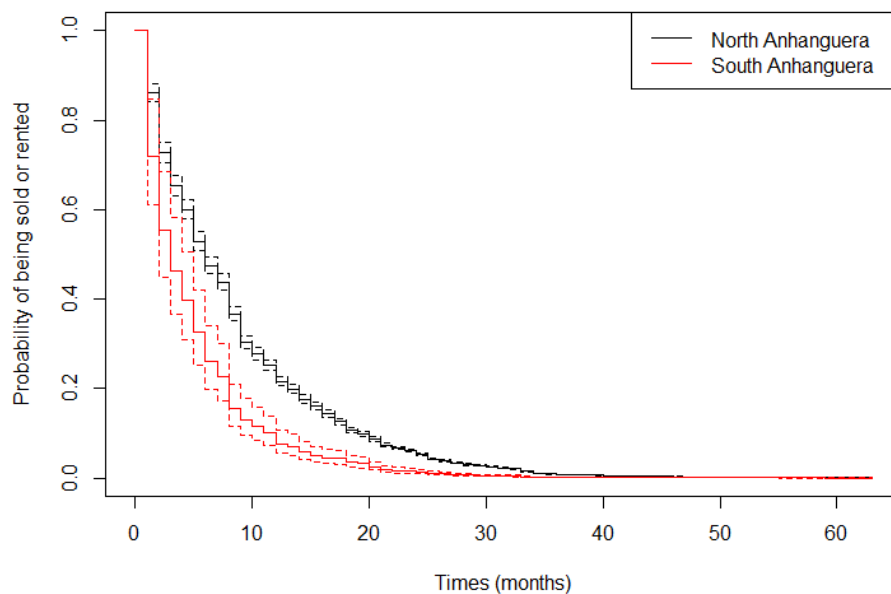
Figure 2 – Deflated square meter price (Brazilian Currency) in Campinas city, Brazil (March, 2019).



Source: Self-elabored using DataZAP data.

All the model coefficients associated with the discrete times are significant. apartment announcements which were located south of the Anhanguera highway have earlier experience of being sold or rented compared to apartment announcements which were located north of the Anhanguera highway indicating significant difference in housing dynamic in both regions (Figure 2). Association with demographic characteristics of both areas and housing dynamic will be tested in further models.

Figure 3 – Apartment announcement surviving probabilities among different inner-city areas, Campinas, Brazil.



References

- BEYER, M. A.; LANEY, D. The importance of big data: A definition. Stamford, CT: Gartner, 2012.
- BILLARI, F. C.; ZAGHENI, E. "Big data and population processes: A revolution?" (2017).
- CAVENAGHI, S.; ALVES, J. E. D.; Domicilios y familias en la experiencia censal del Brasil: cambios y propuesta para identificar arreglos familiares. Notas de población, CEPAL. N° 92. 2011
- CUNHA, J. M. P. da. Aglomerações urbanas e mobilidade populacional: o caso da Região Metropolitana de Campinas. Revista Brasileira de Estudos Populacionais, v. 33, n.1, p. 99-127. jan./abr, 2016.
- _____, J. M. P. da; JAKOB, A. A. E. Os efeitos da migração e da segregação socioespacial na modificação do espaço da Região Metropolitana de Campinas. In: R. bras. Est. Pop., Rio de Janeiro, v. 27, n. 1, p. 115-139, jan./jun. 2010.

DA FONSECA FEITOSA, Flávia et al. Big data e urban analytics à brasileira: questões inerentes a um país profundamente desigual. Revista de Morfologia Urbana, v. 8, n. 1, p. e00141-e00141, 2020.

DIEBOLD, F. X. On the Origin (s) and Development of the Term 'Big Data'. 2012.

DOTA, E. M. Mobilidade residencial intrametropolitana na RM de Campinas: uma abordagem a partir da distribuição espacial dos migrantes. Tese de doutorado. Unicamp-IFCH, Campinas, 2015.

GRAUNT, J. Natural and political observations made upon the bills of mortality. The Johns Hopkins Press, 1939.

IBGE. Instituto Brasileiro de Geografia e Estatística. IBGE Cidades, 2020. Acesso em 10/02/2020. Disponível em: <https://cidades.ibge.gov.br/brasil/sp/campinas>

MULDER, C. H.; LAUSTER, N. T. Housing and family: An introduction. Housing Studies, v. 25, n. 4, p. 433-440, 2010.

WARD, J. S.; BARKER, A. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821, 2013.