5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16482

Spatial distribution of health care facilities in City of Cape Town, South Africa

Sebnem Er[1]
[1]Statistical Sciences Department, University of Cape Town, South Africa.

Abstract

Health care in South Africa is the fourth largest item of government expenditure. Most South Africans do not have medical insurance and therefore seek medical care from public health care facilities such as public hospitals, clinics and community day centres. In City of Cape Town (CoCT), in 2016 there were 149 health care facilities and in 2023 this number increased to 160 facilities with an annual growth rate of 1.19%. The population in 2016 was 4 million and in 2021 it is estimated as approximately 4.76 million which indicates an annual growth rate of 3.5% between 2016 and 2021. The annual growth rate of health care facilities (primarily representing data applicable to the City of Cape Town's Environmental Health Department) is clearly not matching the annual growth rate of the population which is potentially a big problem in a fast growing region like CoCT. The main aim of this paper is to analyse the distribution of health care facilities (clinics, hospitals) within the City of Cape Town, South Africa using spatial kernel density estimation methods and explore what factors affect the number of health care facilities within each ward using generalized linear models.

Keywords: *Health care facilities; spatial distribution; geographically weighted poisson regression.*

1. Introduction

City of Cape Town (CoCT) is the only metropolitan municipality in the Western Cape province of South Africa that is the home for the Mother City, Cape Town. The municipality is made up of 111 wards and as of the latest available 2011 census data, the total population of the municipality is 3,740,026. According to city's measures the population in 2016 was 4,005,016 and in 2021 it is estimated as approximately 4,758,433. The very recent census results are not published yet however it is in no doubt that the city is growing fast with an annual growth rate of 1.38% from 2011 to 2016 and 3.5% from 2016 to 2021. One of the main challenges that South Africa faces is the access to health care services, most people do not have a medical insurance and they seek the services from public health care facilities. Health care facility point pattern distribution in 2021 is provided in Figure 1, where the point pattern plotted over the population measures within wards is provided on the left and on the right pane, the pattern is plotted over the total number of health care facilities within each ward. It is clear that highly dense areas have fewer number of health care facilities compared to less populated areas. Considering that most South Africans take public transport to get to work and to access the services provided within the municipality, it is crucial to explore the distribution of health care facilities within the CoCT.

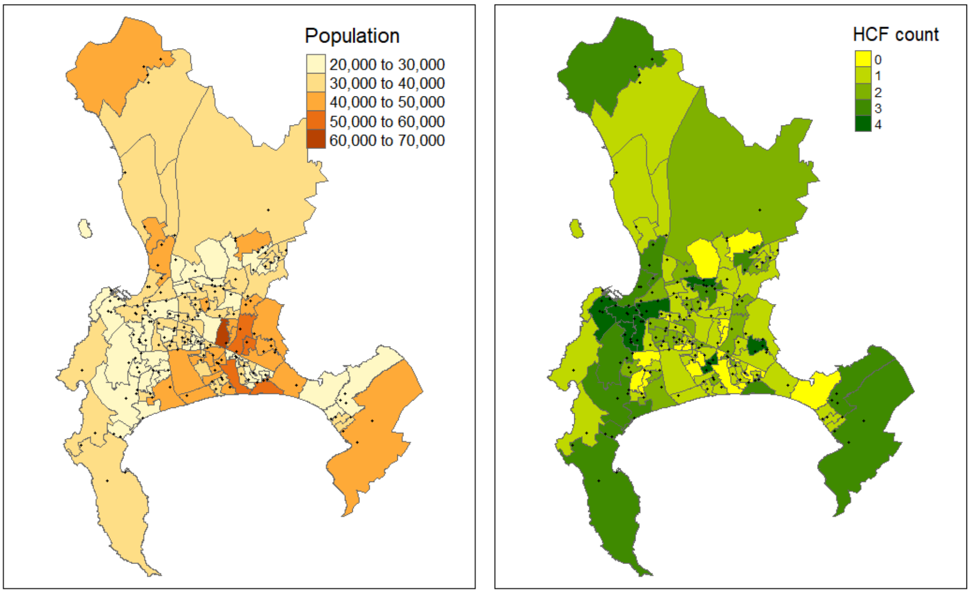


Figure 1. CoCT health care facility distribution overlaid on Left: CoCT population distribution in 2011 (census data), Right: CoCT ward level health care facility count.

In this paper, the intensity of the health care facilities will be estimated using kernel smoothing estimation method and thereafter the total number of health care facilities within 111 wards will be modeled using generalized linear models. There are in total 160 health care facilities, of which 70 of them are clinics, and 48 of them are hospitals (district, regional, psychiatric, tertiary and private hospitals). The data is obtained from City of Cape Town's open data portal which provides publicly accessible data. "*Access to City information helps to increase transparency, as well as benefit the wider community and other stakeholders*" (CoCT Open Data Portal). There are very few studies making use of the data available and analysing with appropriate methods. This research aims to use the health facility data to provide insight on the disparity of the distribution of the facilities within the city.

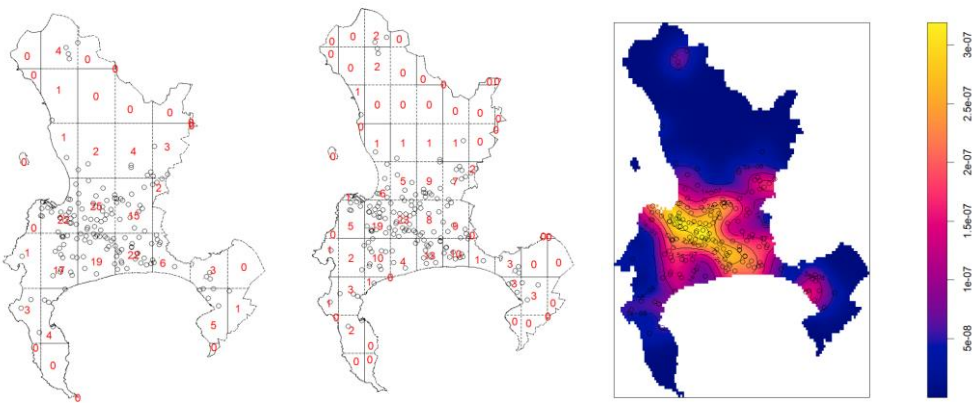
2. Spatial Kernel Density Estimation

Spatial point patterns can be completely random, clustered and regular within a bounding region (A). A completely spatial random (CSR) point process, often associated as homogeneous poisson process (HPP), asserts that the number of events in a quadrat a_j with area $|a_j|$ follows a Poisson distribution with mean $\lambda|a_j| = \frac{n(X)}{|A|}|a_j|$ where λ is the expected number of points within region A and $\bar{\lambda} = \frac{n(X)}{|A|}$ is the unbiased estimate of the true intensity. The second assumption asserts that given n events x_i in a region a_j, the x_i are an independent random sample from the uniform distribution on a_j (Cressie, 1991, pp.586).

There are many different measures that can be calculated to test for completely spatial random processes. Quadrat counting and the associated chi-square test, distance methods and second order statistics are a few of many ways to explore divergence from CSR. In most cases, such as the distribution of the health care facilities, it would be naïve to assume CSR and therefore best practice would be to estimate a varying intensity rather than a constant one. The estimation of the intensity that varies over the location ($\hat{\lambda}(x)$) is known as the inhomogeneous poisson process (IPP) which is a generalisation of CSR. For IPP, the estimation of the intensity can be done by means of non-parametric kernel smoothing or by means of a parametric function for the intensity whose parameters are estimated by maximising the likelihood of the point process. Non-parametric kernel smoothing estimator is provided in the following equation (Bivand et al., 2008, pp.165):

$$\hat{\lambda}(x) = \frac{1}{h^2}\sum_{i=1}^{n} \kappa\left(\frac{\|x - x_i\|}{h}\right) / q(\|x\|)$$

where $\kappa(u)$ is a bivariate and symmetrical kernel function (such as quartic - spherical kernel) and $x_1, x_2,\ldots, x_n$ are the data points. $q(\|x\|)$ is the border correction to compensate for the missing observations that occur when x is close to the border of the region A. In this paper, border correction is not a concern since the entire region (CoCT metropolitan municipality) is under study. Finally, the bandwidth h measures the level of smoothing where small values will produce very peaky estimates, and large values will produce very smooth functions. CoCT health care facility intensity is explored in **R** (R Core Team, 2022) with **spatstat** package (Baddeley, et al., 2015) using 7x7 and 10x10 quadrats and corresponding chi-square tests. The results from quadrat tests are provided in Figure 2 left and middle panes. According to the results, it is evident that the intensity is not constant across the study region and therefore it is important to estimate the intensity that varies over location. The estimation of health care facility intensity varying over the study region has been done using kernel density estimation with a bandwidth optimized using cross validation of the minimization of mean-square error criterion formulated in Diggle (1985).



$\chi^2 = 222.12$
$(p - value \ll 0.001)$

$\chi^2 = 294.76$
$(p - value \ll 0.001)$

Figure 2. CoCT health care facility quadrat counts and chi-square test results for Left: 7x7 quadrats, Middle: 10x10 quadrats. Right: CoCT health care facility kernel density estimation.

3. Geographically Weighted Poisson Regression

To model the health care facility counts within wards, Poisson regression specific to count data will be utilised. Since the number of health care facilities are different locally, geographically weighted regression models are investigated. Geographically weighted poisson regression (GWPR) is the locally predicted version of a Poisson regression model which involves the selection of a bandwidth chosen by LOOCV or manually for count data. Once a decision is made on the bandwidth, the model is fit with a local kernel function which is usually a Gaussian kernel function. The results simply are the estimated geographically varying coefficients within the chosen bandwidth provided as follows (Fotheringham et al., 2003):

$$y_i = e^{\beta_0(u_i,v_i)+\sum_k \beta_k(u_i,v_i)x_{k,i}}$$

where $\beta_0(u_i, v_i)$ is the intercept parameter specific to location i and $\beta_k(u_i, v_i)$ are the coefficients of independent variables at location i. Estimation procedure is conducted in **R** (R Core Team, 2022) using **GWmodel** package (Gollini et al., 2015, Lu et al., 2014).

Table 1 provides the single variable GWPR results for predicting health care facility counts in a ward using several variables such as (i) white population ratio, (ii) black African population ratio, (iii) population density of each ward, (iv) ratio of employed population to unemployed population, (v) unemployed population ratio, and (vi) employed population ratio.

It has to be noted that all models need to be assessed with extensive caution since the pseudo R-square values obtained between the predicted and observed number of health care facilities are very low. Keeping in mind the very low measures, it can be seen that Models 1 and 2 indicate that black African population has access to fewer health care facilities compared to white population. When population density per ward is considered (Model 3), it is observed that denser areas have fewer number of health care facilities compared to less dense areas outlining a negative relationship. Looking at Models 4, 5 and 6, employment plays a positive role in the number of health care facilities. The single variable results point out a possible imbalance of the access to health care facilities at ward level. Several other factors such as closeness to a major road or the specialisation of the health care facilities should be considered for a full picture. The model with the minimum AIC measure is Model 5 where the ratio of unemployed to the population is considered. The parameter estimate is negative which indicates that the less the unemployed ratio is the more there is health care facilities.

Table 1: Single variable generalized poisson regression model results

	Model0	Model1	Model2	Model3	Model4	Model5	Model6
Intercept	0.3656 [4.625] (0.000)	0.2148 [2.123] (0.034)	0.5852 [5.334] (0.000)	0.6585 [5.852] (0.000)	0.1957 [1.772] (0.0764)	0.9012 [6.145] (0.000)	-0.4645 [-1.339] (0.181)
White_ratio		0.7395 [2.737] (0.006)					
Black_Afri_ratio			-0.6276 [-2.588] (0.009)				
popdensity				-0.00005 [-3.194] (0.00141)			
Employed/unemployed					0.0245 [2.420] (0.0155)		
Unemployedratio						-5.6113 [-3.924] (0.000)	
Employedratio							2.3221 [2.515] (0.012)
bw	25	37	105	38	32	102	97
AIC	108.27	103.22	103.03	98.38	104.91	93.96	104.09

bw: bandwidth (# of nearest neighbours), AIC: Akaike information criterion

Table 2: Single variable GWPR model results with minimum, maximum, median and 1st and 3rd quartile values for the specific variables

	Model0 Intercept	Model1 White_ratio	Model2 B_A_ratio	Model3 popdensity	Model4 Emp/Une	Model5 Unempr	Model6 Emplr
Min	0.0864	0.4767	-0.6290	-0.00006	0.0111	-6.0797	2.1592
1st Quartile	0.1770	0.6939	-0.6159	-0.00005	0.0238	-5.9119	2.5107
Median	0.3061	0.9653	-0.6125	-0.00004	0.0392	-5.7813	2.6779
3rd Quartile	0.3911	1.3554	-0.6084	-0.00003	0.0528	-5.6277	2.7594
Max	0.5162	1.6092	-0.6004	0.00000	0.0616	-5.2767	2.8350
Correlation	0.3910	0.4236	0.2662	0.4362	0.4279	0.4183	0.2732
Pseudo-R square	0.1529	0.1797	0.0709	0.1903	0.1831	0.1750	0.0746
AIC	103.75	99.388	103.11	97.64	99.87	93.87	103.61

4. Conclusion and Future Work

The aim of this research is to provide insight on the distribution of the health care facilities within the City of Cape Town using publicly available data from the city's data portal and to draw attention to the fact that there is a disparity in the distribution of health care services. With this aim in mind, the health care facilities in the city have been initially analysed in an exploratory manner using quadrat tests and it has been concluded that the intensity of the health care facilities is not completely spatial random. The varying intensity is modeled with a kernel density function and it has been observed that the health care facilities are localized in certain areas in CoCT. In order to model the health care facility counts within each ward and to examine the effects of potential variables, different geographically weighted poisson regression models were estimated. It has been seen that there might be a possible imbalance in the distribution of the facilities within the metropolitan city especially where black African population has access to fewer health care facilities that are nearby compared to white population. Previous studies have also found similar imbalances in the locations of health care facilities (Lee, 2013). Lee (2013) analysed the distribution of health care facilities located in the metropolitan city of Daejeon, South Korea and found that there is a disparity. In City of Cape Town, given that the number of nearby health care facilities are limited to certain areas, the findings of the research could be used by policy makers to improve the distribution of the health care facilities, especially considering that most South Africans who live in the densely populated areas take public transport to get to work and to access the services provided within the municipality. This could be achieved by for example locating more mobile clinics or health care facilities that can provide services for those diseases that are more prevalent within the densely populated areas so that people who need these services will not need to travel far distances to get health care services.

The analysis can further be extended with hot-spot analysis and inclusion of other variables such as the specialty of the health care facility, convenient road access of the health care facility, the distance to city center or a major road, and poverty and socioeconomic status of each ward. More insight from several variables will extend the findings and increase the predictive power.

References

Baddeley A, Rubak E, Turner R (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London. https://www.routledge.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/.

CoCT Open Data Portal. (2023). URL: https://odp-cctegis.opendata.arcgis.com/datasets/cctegis::health-care-facilities-clinics-hospitals/explore?location=-33.867164%2C18.629850%2C10.45

Diggle, P.J. (1985). A kernel method for smoothing point process data. *Applied Statistics Journal of the Royal Statistical Society*, Series C, 34, 138–147.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Gollini I, Lu B, Charlton M, Brunsdon C, Harris P (2015). “GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models.” *Journal of Statistical Software*, 63(17), 1–50. doi:10.18637/jss.v063.i17.

Lee, Kwang-Soo (2013). “Disparity in the spatial distribution of clinics within a metropolitan city”. *Geospatial Health*, 7(2), 199-207.

Lu B, Harris P, Charlton M, Brunsdon C (2014). “The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models.” *Geo-spatial Information Science*, 17(2), 85–101. doi:10.1080/10095020.2014.917453

Municipalities of South Africa, CoCT demographic data, (2023). URL: https://municipalities.co.za/demographic/6/city-of-cape-town-metropolitan-municipality

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Western Cape Provincial, Socio-economic profile for City of Cape Town, (2021). URL: https://www.westerncape.gov.za/provincial-treasury/files/atoms/files/SEP-LG%202021%20-%20City%20of%20Cape%20Town.pdf