Retail Anatomy

An Analysis on Commercial Activity in Lisbon

1995-2010

Rui Manuel Cordeiro Colaço

Dissertation to obtain the degree of Master in

Urbanism and Spatial Planning

Supervisor: Prof. João António de Abreu e Silva

Examination Committee

Chairperson: Prof. Jorge Manuel Gonçalves

Supervisor: Prof. João António de Abreu e Silva

Members of the Committee: Prof. Herculano Alberto Pinto Cachinho

October 2017

A far as the laws of mathematics refer to reality, they are not certain;
and as far as they are certain, they do not refer to reality.

Albert Einstein

ABSTRACT

The location of commercial activity, and its role within the urban system, has been subject to study since at least the first half of the last century, with prior location models dating even earlier. Those first models are still, however, mandatory references in research, considering the formidable task of understanding the processes that underlie at the foundations of a city. The city is a complex system, where the presence of commercial activity presents itself as a desirable characteristic, whose presence is a both a measure of centrality and liveability: commercial establishments are places of consumption, but also socialization. Economic and social characteristics, accessibility and spatial syntax, along with many other factors, have been used when accounting for their presence, with 200 years of research providing for some relevant insight on the subject. The present dissertation aims to contribute to it, by analysing three commercial censuses carried out in three distinct decades (1995, 2002 and 2010) in the city of Lisbon. Retail location and agglomeration, relocation or extinction, among other phenomena are identified and interpreted, while trying to explain for spatial patterns and establish causal relations between these and other available information (census data from 1991, 2001 and 2011). A better understanding of the processes thus identified can help in developing measures that will contribute for preserving liveability while maintaining relevance, since they can inform urban planners in areas as diverse as urban renewal, mobility issues, or even social intervention.

***Keywords*:** Spatial analysis; location models; urban economics.

RESUMO

A localização de atividade comercial, e a sua inserção no sistema urbano, são alvo de estudo desde a primeira metade do século passado, sendo que os primeiros modelos de localização são ainda anteriores. Esses modelos, contudo, permanecem como referências obrigatórias em trabalhos onde seja proposto analisar a dinâmica da cidade, considerando a complexidade que lhe é inata, e que sempre impossibilitará que todos os processos subjacentes ao seu funcionamento possam ser decifrados. A cidade é, afinal, um sistema complexo, onde a presença de atividade comercial se apresenta como característica desejável, cuja presença constitui medida de centralidade e (con)vivência: os estabelecimentos comerciais são, desde sempre, locais de consumo, mas também de socialização. A sua presença tem sido explicada por fatores económicos e sociais, e características de acessibilidade e sintaxe espacial, entre muitos outros. A presente dissertação tem como objetivo contribuir para essa pesquisa, utilizando como sujeito de experimentação a cidade de Lisboa, e como chave de desencriptação três recenseamentos comerciais (CML) efetuados em três décadas distintas (1995, 2002 e 2010). São analisados fenómenos observados para o comércio a retalho (localização e aglomeração, relocalização ou extinção, entre outros), procurando identificar padrões espaciais e estabelecer relações de causalidade entre estes e outra informação disponível (Censos 1991, 2001 e 2011). A interpretação dos processos identificados permite ganhar conhecimento sobre os mesmos, podendo essa informação contribuir para o planeamento urbano em áreas tão diversas como renovação urbana, mobilidade, ou intervenção social.

Palavras-chave: análise espacial; modelos de localização; economia urbana.

ACKNOWLEDGEMENTS

I am forever in debt to my research supervisor, Professor João de Abreu e Silva, for having accepted to guide me through this process. It started with close to nothing and if it amounted to something, it was certainly because of his immense patience, his willingness to keep pushing me forward, and his undefeatable optimism.

Dr. Luis Inostroza offered me very insightful feedback, for which I'm especially thankful.

I must thank my friends, colleagues, and family. Please know that your support is treasured. It'd be impossible to name everyone, but unfair not to mention Júlia. Her very inceptive remarks were precious, and greatly appreciated.

To my mother. For everything.

LIST OF ABBREVIATIONS

AIC – Akaike Information Criterion

CAE – Classificação Portuguesa de Atividades Económicas (Portuguese Classification of Economic Activities)

CBD – Central Business District

CML – Câmara Municipal de Lisboa (Lisbon's City Council)

CPT – Central Place Theory

GIS – Geographic Information System

INE – Instituto Nacional de Estatística (The National Institute of Statistics)

LM – Lagrange Multiplier

MAUP – Modifiable Areal Unit Problem

OLS – Ordinary Least Squares

PNRB – Primarily Non-Residential Buildings

SAR – Spatial-Autoregressive Lag Model

SEM – Spatial-Autoregressive Error Model

SC – Schwarz Criterion

TABLE OF CONTENTS

ABSTRACT I

RESUMO II

ACKNOWLEDGEMENTS III

LIST OF ABBREVIATIONS IV

TABLE OF CONTENTS V

FIGURE INDEX VII

TABLE INDEX IX

1. Introduction 1

1.1. Aim And Motivation 1

1.2. Objectives 1

1.3. Structure Of The Work 2

2. State Of The Art 3

2.1. On Location 3

2.1.1. A Note On Von Thünen 3

2.1.2. Central Place Theory 4

2.1.3. Bid Rent Theory 7

2.2. On Agglomeration 8

2.2.1. Competitive Clustering 9

2.2.2. Multipurpose Shopping 10

2.2.3. Information Spillover Effects 10

2.3. On Size 11

2.4. On Classification 13

2.5. Recent Developments 14

2.6. Summary 16

2.7. Retail Location In Lisbon – A Brief Literature Review 17

3. An Exploratory Analysis Of Commercial Data From Lisbon 21

3.1. Data 21

3.2. Preliminary Analysis Of Data 21

3.3. An Analysis On Clustering – Commercial Centers 23

3.4. An Analysis On Change – Commercial Density 27

3.5. An Analysis On Mix - Commercial Diversity 38

3.6. Exploratory Analysis – Summary 40

4. Spatial Regression Analysis 42

4.1. Preparing A Spatial Regression Analysis 42

4.2. Modelling A Spatial Regression Analysis 46

4.3. Discussion/Results 57

5. Conclusions And Implications For Planning Practice 61

5.1. Conclusions 61

5.2. Possible Implications For Planning Practice 63

Annex I 71

FIGURE INDEX

Figure 1 – A System of Central Places: from Circle to Hexagon 5

Figure 2 – Conceptualization of a rent gradient 7

Figure 3 – Establishments (Total): 1995, 2002 and 2010 (left to right) 21

Figure 4 – Commercial Locations within Lisbon 22

Figure 5 – Total Establishments. Optimized Hotspot analysis results, 1995, 2002 and 2010 (left to right) 24

Figure 6 – Retail (left) and Restaurant Sector (right). Optimized Hotspot analysis results, 1995 24

Figure 7 – Foods. Optimized Hotspot analysis results, 1995 25

Figure 8 – Personal Use Items, Household Articles, Health and Hygiene. Optimized Hotspot Analysis, 1995 25

Figure 9 – Leisure items, Other Items, Repairs. Optimized Hotspot Analysis, 1995 26

Figure 10 – Restaurants, Cafes, Bars. Optimized Hotspot Analysis, 1995 26

Figure 11 – Total Establishment Density (2010) using 100m, 300m and 600m radii (left to right) 28

Figure 12 – Total Establishments Density Variation, 2002 to 2010. Real Density (left) and Smoothed Density (right) 30

Figure 13 – A possible legend Density Variation Maps 30

Figure 14 – Total Establishments Density Variation: 1995 to 2002, and 2002 to 2010 (left to right) 31

Figure 15 – Retail Sector (as defined) – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right) 32

Figure 16 – Restaurant Sector (as defined) – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right) 33

Figure 17 – Foods (as defined) – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right) 35

Figure 18 – Foods (left) and Population (right) – Density Variation: 2002 to 2010 (Foods) and 2001 to 2011 (Population) 36

Figure 19 – Household Articles – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right) 37

Figure 20 – Commercial Diversity: 1995, 2002 and 2010 (clockwise) 39

Figure 21 – Establishment Diversity (2010) and Population Density (2011) (left to right) 40

Figure 22 – Population (left), Population and Foods Hotspots (right), 2010-11..... 43

Figure 23 – Population and Leisure Hotspots (left), Population and Restaurants Hotspots (right), 2010-11..... 44

Figure 24 – PNRB (left) PNRB and Foods (middle) PNRB and Restaurants (right), 2010-11 44

Figure 25 – Population Density, 2010-11: INE's grid (left) and the 150mx150m grid used in the study 45

Figure 26 – Spatial Lag (left) and Spatial Error (right) 47

Figure 27 – Foods, Observed and Predicted Values, 2010-11. Using Separate Types as Predictors (left) or as a Whole (right) 60

Figure 28 – Foods, Statistics of the Previous Scatters Plots. Using Separate Types as Predictors (above) or as a Whole (below)..... 60

TABLE INDEX

Table 1 – Generational Cycles of Retail Location Theory 17

Table 2 – Correlation Matrix of Rasters for Radii 100m to 600m using Band Collection Statistics 28

Table 3 – Correlation of Observed Density and Raster Densities using CORREL (Microsoft Excel) ... 29

Table 4 – Total Number of Establishments and Variation 1995-2002 and 2002-2010 33

Table 5 – Establishments located outside Malls – Total Number and Variation 1995-2002 and 2002-2010 34

Table 6 – Establishments located in Malls - Total Number, Percentage and Variation, 1995, 2002 and 2010 34

Table 7 – Results for OLS Regression, Foods 1991-95 48

Table 8 – Diagnostics for Spatial Dependence, OLS Regression, Foods 1991-95 49

Table 9 – Spatial Lag and Spatial Error for Foods, 1991-95 49

Table 10 – Spatial Error for Foods, 1991-95 50

Table 11 – Diagnostics for Spatial Dependence, OLS Regression, Foods 2010-11 51

Table 12 – Model Fit Statistics, Foods 2010-11 51

Table 13 – Spatial Error for Foods, 1991-95 and 2010-11 52

Table 14 – Spatial Error model for Restaurants, Household Articles and Leisure 2010-11 53

Table 15 – Spatial Error for Other Items, Repairs and Cafes 2010-11 54

Table 16 – Spatial Error for Foods, using aggregated commercial activity as a predictor, 2010-11 59

1. INTRODUCTION

1.1. AIM AND MOTIVATION

First theories on retail location date back to, at least, the late 1920's (Haig, 1926, 1927; Hotelling. 1929), but the 1960's were especially productive in the academic field, with Alonso's bid-rent theory (1960; 1964) and Huff's gravitational model (1963; 1964) still explaining for the better part of retail location to this day. A cause for an apparent lack of interest when compared to other areas might be explained by the private sector taking lead when pertaining research on where to locate an establishment, and its apparently effective success.

But if from a private sector perspective, research on retail location is an investment, from a public perspective, it allows for planning decisions to be taken from a more informed perspective. Retail location can impact on matters as diverse as sustainable mobility, urban regeneration, social interaction or urban logistics. A better understanding on how commercial activity locates can contribute to achieve better planning solutions in as many areas, with some specific questions being pointed out in 1.2.

Contributing to research demanded a case study; Lisbon turned out to be a possibility when a database of commercial establishments was made available by the municipality. Even though a great deal is being written about "big data", data collected in a more traditional manner (like door-to-door enquiries) is far from being obsolete. A mix of sources has been used in some very innovative models (Batty *et al*, 2017), but having reliable data at block level is still difficult to obtain, and therefore, relevant for analysis. Data from commercial establishments was available for the 1995-2010 period (CML, 2016), as was data about population and housing for the 1991-2011 period (INE, 2014), and that amounted to a good starting point.

1.2. OBJECTIVES

The aim of the research was, primarily, to identify retail location patterns in Lisbon in 1995, 2002 and 2010, focusing on cross sectional patterns and their longitudinal changes, and relating them with population and housing data.

First analyses were exploratory, namely, what kind of information could be obtained by using that data and GIS software. Off-the-shelf software poses the risk of any user finding himself in a position in which he feels all data is accessible and all results reliable. "The user needs to be a skilful spatial modeller" (Clarke, 1998) to obtain proper results. Exploratory analyses allowed, if not to become *skilful*, at least

for the user to become aware of what he was dealing with, and what could effectively be extracted from the dataset.

Eventually, the process led to modelling a spatial regression model, relating *external factors* (population and housing characteristics, with primarily non-residential buildings used as a possible source for employment location, and both standing as proxies for *demand*) with *endogenous retail location decisions* (with the relationships between different types accounting for composition of the commercial structure at block level – and eventually, within the city).

The main objective of the dissertation was, therefore, to gain knowledge about retail activity location within Lisbon. A better understanding on commercial location might contribute to planning: by understanding why and how activity is located, informed decisions can be taken, ranging from parking policies to urban regeneration, from sustainable mobility to urban logistics or even social intervention.
Some specific examples are given:

- If commercial activity is increasing within a neighbourhood (like what's happening in Príncipe Real), should the City Council add to the *momentum* by (e.g.) regenerating public buildings?
- If the opposite is happening, could the same model be reproduced on other parts of town (and should it?)? Can a museum (e.g.) produce the same effect ("anchor")?
- If shopping malls, which have a competitive edge on traditional street level activity by offering parking facilities (e.g.) are allowed to locate within city, and parking offer is not only limited, but even conditioned within those neighbourhoods, how can the later offer resistance to the first?

It is, therefore, obvious, that a better understanding on commercial location can help in developing measures that will contribute for preserving liveability while maintaining relevance, both for a city or for a neighbourhood within it. Better knowledge on an activity as relevant as retail will allow for better planning and better management of the city.

1.3. STRUCTURE OF THE WORK

The dissertation starts with a literature review, which leads to state-of-the-art on the subject of retail location. This is followed by an overview of commercial establishments' spatial location in Lisbon, for the period 1995-2010, using different spatial analysis techniques. Finally, an OLS-type regression is modelled, using different types of activities as predictors for the location of a specific one, along with census variables. This allows for an analysis on the role of both endogenous decisions and exogenous factors on the composition of the city's commercial structure.

Conclusions pertain the processes of retail location and agglomeration, which are relevant for urban planning and urban management.

2. STATE OF THE ART

This chapter is perhaps built in an unorthodox manner: from 2014 (Sevtsuk, 2014) to 1826 (von Thünen, 1826/1966) and even earlier, but eventually leading back to present day (Batty *et al* 2017; Sevtsuk & Kalvo, 2017). That, though, was the way literature was approached; for if an assessment on current state-of-the-art was going to be made, it was only logical to start by reading what is being done now, and then go back to the landmark publications that led the field into its current state of development.

A coherent timeline is created for the benefit of the reader, even though the process itself involved leaping back and forward into the many seminal works that have allowed for "a substantial conceptual edifice" (Brown, 1993) to be constructed for retail location theory, especially during the 20th century

Literature review thus identifies what questions have been acknowledged and, to some extent, answered, eventually allowing for an interpretation of the results that will be presented later on, starting with von Thünen, whose work "*The Isolated State in Relation to Agriculture and Political Economy*" (1826/1966) has influenced geographers, economists, and other urban planners for what is now almost two centuries.

2.1. ON LOCATION

2.1.1. A NOTE ON VON THÜNEN

"*Imagine a very large town, at the centre of a fertile plain which is crossed by no navigable river or canal. Throughout the plain the soil is capable of cultivation and of the same fertility. Far from the town, the plain turns into an uncultivated wilderness which cuts off all communication between this state and the outside world. There are no other towns on the plain. The central town must therefore supply the rural areas with all manufactured products, and in return it will obtain all its provisions from the surrounding countryside*". (von Thünen, J.H., 1966: 7).

Location models' date, consensually, to von Thünen. Before him, Adam Smith, Thomas Malthus and David Ricardo had all been dealing with concepts of rent and land use, with the first having concluded the following.

"*The rent of land, therefore, considered as the price paid for the use of the land, is naturally a monopoly price. It is not at all proportioned to what the landlord may have laid out upon the improvement of the land, or to what he can afford to take; but to what the farmer can afford to give."* (Smith, 1776/1976: 162)

Considering what “the farmer can afford to give”, von Thünen introduced *transportation costs*; and with those, *distance* to the market, and therefore, *location*: the location in relation to the market in which a farmer should locate so he could pay the highest rent, as defined by the landowner.

This allows for spatially zoning the use of land: the locations closer to the market must be occupied by the most distance-sensitive crops. These will be perishable goods, and the price of location will be compensated by very small transportation costs (or not at all – and the farmers will move, according to Ricardo). When distance exceeds the benefit obtained from proximity, another crop will succeed, one which is less distance sensitive, and so on, until “the wilderness” is reached (the model establishes one center/town, surrounded by 4 areas at varying distances, accordingly to estimated viable crops for each of the 4; beyond the 4th area, no land use is economically feasible due to distance, and thus “the wilderness” encloses the model). Location is, therefore, determined by linear functions between *rent* and *distance* for each of the 4 “rings”.

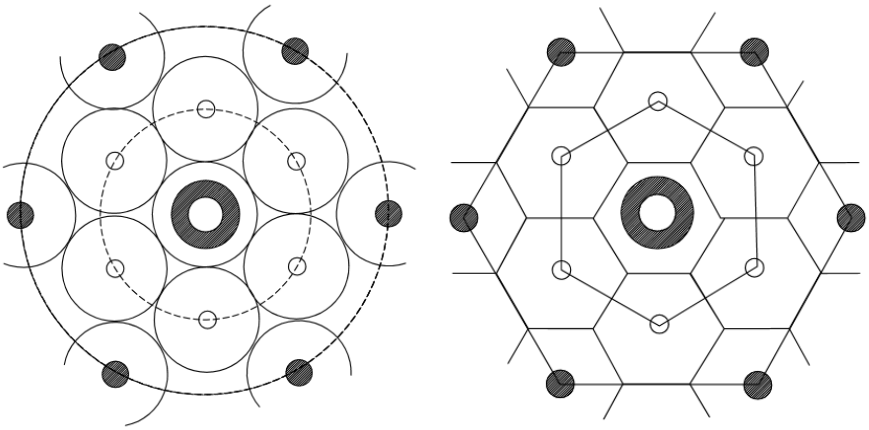
Location and distance (and therefore, *transportation*) were already paired up in von Thünen´s model. As will be demonstrated, these two factors are at the base of almost every attempt at modelling the location of economic activities. His work on location and *rent* (as cause-effect) didn’t find a successor until the 1960’s (Alonso, 1960); for more than a century, it evolved in two separate lines of thought, one about location, and another one about rent.

2.1.2. Central Place Theory

In 1933, Walter Christäller presented a theory which has since become an inevitable reference in location studies: Central Place Theory (CPT) (Christäller, 1933/1966) introduced the concept of *importance* of a place, coined as its *centrality*, related with the functions associated with that place (by definition, *central functions*: goods and services being offered at a central place; *central professions* are also mentioned). Each place is central to a region, and a hierarchy of places is built upon the types of functions associated with the centres. Higher order centres offer higher order types of central goods and services, which are therefore more numerous and more diverse than the ones offered at centres located at a lower level.

The concept of *range* is introduced, related with the distance a consumer is willing to travel to obtain a certain good, which is shorter for lower order goods. A “lower limit” is also introduced, as the minimum demand necessary for a good to be made available.

Considering the cost of transport to be equal in any direction, a scheme, similar to the one presented in Figure 1 was produced. The well-known hexagonal grid comes in response to the need to model “a perfectly uniform net of places (…) so that there is no unsupplied part.” (Christäller, 1966).



Source: Adapted from Christäller (1966) and Lösch (1954)

Figure 1 – A System of Central Places: from Circle to Hexagon

Christäller acknowledged the strictness of the model, designed for optimizing market profit, and provided adaptations that would allow for an efficient transportation network or an eventual administrative grid to be superimposed.

Even though Lösch's work is usually paired with the one of fellow countryman Christäller, *Die räumliche Ordnung der Wirtschaft* (1940, translated into "The Economics of Location", 1954) incorporated and reviewed concepts not only by Christäller, but also by Reilly (1929, 1931) and Hotelling (1929). Considering the importance their theories would be given later, it is reductive to consider Lösch's work as complimentary only to Christäller's: it is a *magnum opus* on economic geography, and an heir to the work of von Thünen (with Thünen's rings being thoroughly studied). The hexagon shaped market is, after all, a derivation of those: "any kind of Thünen rings meet" (Lösch, 1954: 14). The hexagon shaped market is seen an advantageous way to promote that encounter, "because hexagons, like triangles but unlike squares, can cover the earth without a remainder" (Lösch, 1954: 114).

But although the hexagon can't help but draw comparison with Christäller's, his alternative to the positioning of centres includes gravity into the model; a feature closer to Reilly's "Law of Retail Gravitation" (Reilly, 1931), which he deems "a serious attempt to express quantitatively the co-operation of the most important factors" (Lösch, 1954: 412), with distance and size of demand in retail trade being those factors.

In the Löschian system, "centres of similar size do not necessarily perform the same functions; centres need not necessarily be larger, the greater the size of their market areas; centres if ranked by size may under certain circumstances appear to be a continuum; and market areas may vary sufficiently in size as to preclude a nesting pattern characteristic of Christäller's system" (Scott, 1967: 13). Lösch's inclusion of *heterogeneity* is of foremost importance.

An admirable feature of “The Economics of Location” is its acknowledgement of Hotelling’s work (Hotelling, 1929): Lösch objects to Hotelling’s theory and presented his own theories on “Agglomeration of Different Enterprises” (1954: 76, 77):

“*Advantages in Numbers: There are common advantages to producers in the agglomeration at one place of a moderately large output, no matter of what kind. One of the most important is the advantage of having a railway station, not to mention better streets and drainage, cheaper water and electricity, and a larger labour market*” (Lösch 1954: 76).

In 2010, Sevtsuk refers to the importance of other exogenous factors like “spatial accessibility and land use attractions” (Sevtsuk, 2010: 6) in the location of retail establishments and heavily explores those factors.

“*Advantages of Association: (…) under any given market situation, the preference of consumers for combining small purchases or comparing various qualities of differentiated products is hardly less important for the formation of towns than for the existence of special business districts within a town and of department stores in these districts. The mere fact of their proximity not only lowers the cost of production, especially general costs, but at the same time increases the share of the demand*” (Lösch 1954: 76).

In 1982, Eaton and Lipsey acknowledge this as “multipurpose and comparison shopping” (1982: 58), a feature that explains for much of retail agglomeration.

“Advantages of Proximity: It is advantageous in every economic region to have the market networks for individual goods coincide at one point, as we shall see later. This metropolis, to be sure, fulfils functions for the whole area, but its functions could be fulfilled also if the market networks were set down indiscriminately. Aside from the reasons previously given they will meet in one place because industries largely determined by consumption mutually create in this way the advantages of a large local demand for each other. These advantages consist partly in the feasibility of more or larger firms, and partly in the fact that only thus does there arise a sufficiently large demand for many goods” (Lösch 1954: 77).

And since the 1980’s, agglomeration economies have been described in a manner not dissimilar to this: “(…) one of the most remarkable things about the United States is that in a generally sparse populated country (…) the bulk of the population resides in a few clusters of metropolitan areas” (Krugman, 1991: 483).

In 1958, Berry and Garrison (1958, 1958a) set out to verify to what extent CPT could be applied to in an actual urban scenario: “52 types of retail and service activities present in 33 small urban centers” (1958a: 305), picking up the notion of *range* from Christäller and Lösch, and introducing *threshold* as the lower limit of minimum demand necessary for a good to be made available: *threshold population*, necessary to assure the presence of a different type of function (either good or service) is determined. A first article

(Berry and Garrison, 1958) classifies centres in accordance to the number and type of functions that are present in each. An observed hierarchy of centres stems from this observation, and doesn't rely on a hexagonal pattern – or any spatial pattern, by that matter: just population. A second article models a function that relates number of stores with population, thus allowing for additional considerations on range of a centre.

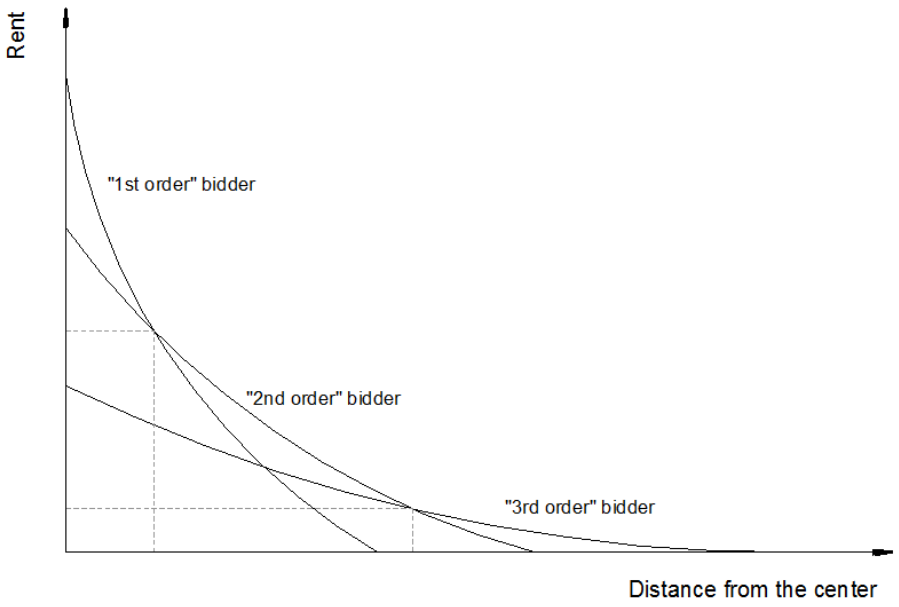
Empirical studies relating central places with specific functions are thus reinforced by observed data: smaller communities will present stores selling daily necessity goods (like food stores or filling stations), while a hospital will only be found at the larger ones.

2.1.3. Bid Rent Theory

Alonso (1960) starts "A Theory of Urban Land Market" by referring the work of Haig (1927), but crediting that "the foundations of the formal spatial analysis of agricultural rent and location are found in the work of J. von Thünen." (1960: 149).

Departing from similar premises to the ones of von Thünen, Alonso incorporates business and residential use of the land, apart from agricultural, into his model.

A bid-rent function is established to fit that relationship, and different curves are produced for different land uses. The curves can be superimposed, with the organization of the space deriving from the steepness of the curves. "If the curves of business are steeper than those of residences, and the residential curves steeper than the agricultural ones, there will be business at the centre of the city, surrounded by residences, and these will be surrounded by agriculture" (Alonso, 1960: 166).



Source: Adapted from Alonso (1960, 1964)

Figure 2 – Conceptualization of a rent gradient

In Figure 2, the "1st order" bidder would therefore be business, the "2nd order" residential use and the "3rd order" agricultural use. Within each use, "families" of bid-price curves can be constructed: using business as an example, different activities will be able to pay different rents, and therefore, bid at a different price for the same location. If the existing structure of land prices is superimposed, it is possible to predict how activities will locate.

"*Jewellers, for instance, are able to pay on average higher rents than furniture retailers because they can exploit more fully those sites where pedestrian flows are heaviest. Similarly, variety stores can normally outbid women's clothing shops which in turn can outbid grocery stores. It follows that each type of retail establishment will have a 'rent gradient' or downward progression of maximum rents with decreasing accessibility to the core and that the steepness of the slope will depend upon the sensitivity of the retailer's output and costs to changes in accessibility*." (Scott, 1967: 21).

Therefore, and focusing again on Figure 2, one could expect variety stores to be the "1st order bidder", women's clothing the "2nd order" and grocery stores the "3rd order". Different groceries would have different curves, and the structure would have to be more complex to accommodate all possible activities, but a general idea on how bid-rent would explain the pattern of retail is attained.

Location decisions, though, are related with the characteristics each retailer values the most: Scott (1967) uses pedestrian flows as an external factor that jewellers can exploit better than furniture retailers, but gives no reason to why clothing stores can outbid grocery stores. The *higher order* nature of a good (explored by Lösch and Christäller) might explain it: jewellers possess a *higher range* than a furniture retailer (while demanding a smaller area to operate). Berry and Garrison's *threshold population* (1958, 1958a) might also help understand this, with stores selling frequently purchased goods ("food stores") depending on a smaller population and having, thus, limited interest in bidding for space at central locations (where eventually, population isn't even present: "(…) there will be business at the centre of the city, surrounded by residences (…)" (Alonso, 1960)).

2.2. ON AGGLOMERATION

While the previous section was about the foundation of location theories, for economic activities in general, this one will pertain the foundations for individual location choices within the city. Alonso's work helped understand that transition: from locations around the city to locations within the city, rent plays a relevant part in the decision process that underlies the choice to locate at a specific point around the centre.

Rent, though, is merely a constraint. The impossibility to bid for the best location may prevent from locating at a desired spot. The desire, though, exists, and it's what lies underneath that explains for

agglomeration: the desire to locate at the best place, where the best result will be achieved, given the purpose of the bidder.

2.2.1. Competitive Clustering

Harold Hotelling called it “cupidity” (Hotelling, 1929: 49): the desire to maximize profit, thus breaking any price-fixed agreement (and with that, any location-fixed arrangement).

This introduces *competition* into location, for von Thünen’s rings work only if the landowner is the same for the entire isolated state, renting to farmers according to optimal distances to the market depending on the nature of the crops, so that they can maximize profits and, consequently, maximize the sum of the rent due – to him. And Christäller and Lösch’s models are optimal arrangements, which therefore can only be implemented ideally (or by a central authority that replaces that of the landowner).

The basis, though, is the same: Hotelling sets his entrepreneurs at a linear market, along which demand is evenly distributed, and each buyer values only the price of the commodity plus transportation (distance). An optimal arrangement is that of two businesses locating at ¼ from each end: each will have half of the market or, from a buyer’s perspective, each will locate at the ideal distance of ¼ total or less. Hotelling finds that this arrangement, though optimal, is hardly relatable with an actual scenario: two competitors will hardly locate at the same time; and if a store is already established, at the centre, there will be no advantageous position for the second competitor: the most advantageous strategy will be to locate as close as possible to the first establishment, and compete for the same market, by selling the same good at a lower price, or by introducing a difference that, though minimal, can be perceived as interesting by the consumer. This is valid for a 3rd competitor, but only for very similar products.

If instead of a line the model is applied to a plane, it’s easy to relate it with the circles drawn around each centre by Christäller and Lösch, and their natural evolution into the well-known honeycomb shaped grid (Figure 1). This was attempted by Eaton & Lipsey (1975) who designed a model to optimally accommodate 19 firms. At first, firms were arranged within a hexagonal grid. But after testing, what was found was that it’d be more likely for firms to cluster than to maintain that perfect, löschian equilibrium, which led them to enunciate the “Principle of Local Clustering”:

“*When a new firm enters a market, or when an existing firm relocates, there is a strong tendency for that firm to locate as close as possible to another firm. This behaviour tends to create local clusters of firms in many equilibrium and disequilibrium situations. The principle of minimum differentiation is a special case of the principle of local clustering when the number of firms in the market is restricted to two*.” (Eaton & Lipsey, 1975: 46).

Taking the example of food courts agglomerating at regional shopping malls, Church and Bell (1990: 109) state that "patterns of retail sites selected by entrepreneurs distributing the same good or service tend to confirm the importance of co-location as a viable locational strategy" thus transporting Hotelling's theory into shopping malls of the 1990's, whose very specific function within retail structure will be noticed further on this dissertation.

2.2.2. Multipurpose Shopping

Lösch acknowledge the possibility of multipurpose shopping extensively, on his notes on "Agglomeration of Different Enterprises" (1954: 76, 77)), as was mentioned before. And so did Christäller. They just didn't fit it into their models, which were, after all, *optimal solutions* based on strict economic principles, not on empirical observation. But *transportation costs* have been included in all location models, dating back to von Thünen. Bid-rent is about location. Huff (1963) makes a case for the *perceived utility* of a trip being greater when a shopping centre has a larger size (the work of Huff will be explored in the following section). Therefore, it is only expectable that a centre that offers the possibility of multipurpose shopping will be interesting both for the consumers and the "producers" (in this case, retailers). Consumers will value the possibility of acquiring different goods in the same trip, thus saving in transportation costs, which will also allow them to patronize more distant centres (Sevtsuk, 2010; Larsson & Öner, 2014). And retailers will benefit from locating in proximity to one another, thus creating *spillover effects* (as will be described next) from which all will be able to take advantage. The location itself might possess features that make it interesting for all retailers (*externalities*, like Lösch's "train station"). All this makes a case for multipurpose shopping, with multipurpose shopping, in return, explaining for a large part of retail agglomeration.

"*(…) multi-purpose shopping behaviour provides, for the first time, a rationale for the existence of a hierarchy of centres; for differences in the functional composition of centres at the same hierarchical level; for the existence of several outlets selling the commodity within a single centre; and, not least, for the very fact that dissimilar retail outlets cluster together into recognizable shopping centres, both planned and unplanned*." (Brown, 1993: 190).

2.2.3. Information Spillover Effects

Spillover effects account for the advantages one activity might obtain by locating near another. It's not different from what Church and Bell (1990) or, by that matter, Hotelling (1929) stated about the "second" store to locate: it'll be advantageous to locate near the pre-existing one and compete for the same market: the location is already known as a commercial location, and a better chance of getting a share of the market will arise from co-locating. But while those authors were mentioning the same type of good, a spillover effect can be applied to any store locating nearby a pre-existing one: it benefits from the fact consumers are informed about that location as one where commercial activity occurs. When

making a trip they're already used to, they might notice that second store: it'll be advantageous for that store, and costless.

A study on information spillover effects is made by Caplin and Leahy (1998): a store locates at a place at considerable risk; if it succeeds, other stores may locate because the first one made the location *noticeable*. It attracts costumers that wouldn't have come other way. A piece from the New York Times is quoted by the authors: "Let's face it, on a sunny afternoon Fifth Avenue is bustling and Sixth is deserted. Who knows to what extent bargain hunters will seek bargains that cost a $20 round-trip cab ride to get to.". After one department store makes that "cab ride" worthwhile *per se*, others can co-locate at a much smaller risk. This too is "price of the commodity plus transportation" as was mentioned when asserting about Hotelling's theory.

A similar type of spillover effect occurs within shopping malls: building on bid rent theory, but applying it to shopping malls, Carter and Vandell (2005) introduce anchor stores within shopping malls as having a desirable spillover effect, based on which their location inside the mall is defined: "anchors should be at opposite ends of the mall and, in the case of more than two anchors, should be equidistant from the center, the idea being that they draw shopper traffic through the mall and past non-anchor tenants" (Carter & Vandell, 2005: 242). The effect is the same, and since the shopping mall is centrally managed, the risk that the anchor has to take is limited.

"*Classical models of perfect and imperfect competition are generally based on the assumption that consumers have complete information about the offerings of different sellers. However, in reality, the transmission of such information is usually not costless, and the degree of competition between firms will therefore depend on what firms do to facilitate price comparison by consumers*." (Dudey, 1990: 1103).

2.3. ON SIZE

Until now, commercial activity has been described based on its *location*; and possible explanation for *agglomeration* has been presented. *Size*, though, is yet to be explained: the circular form of von Thünen or the hexagonal of Christäller and Lösch derives from planned locations. Hotelling suggested perfect competition based on circular markets, but made a case for clustering (a case that was later built upon, as was described). But what exactly is the *optimal size* of that cluster?

The answer to that question may lie on *spatial interaction*, a concept first introduced by Reilly (1931), who postulated that "two cities draw trade from any intermediate town (or city) approximately in direct proportion to the populations of the two cities and in inverse proportion to the square of the distances from these two cities to the intermediate town" (Reilly, 1931: 9), a concept introduced as the Law of

Retail Gravitation (in analogy with that of Isaac Newton), which Huff (1963) came to develop 30 years later, with still considerable influence in the present day.

Huff's model introduces *probability* into Reilly's Law: *distance* decreases the probability for a shopping centre to attract a costumer, while *size* increases it (size altering the perceived *utility* of the trip – and thus relating it with what was written about retail agglomeration).

A function is modelled, with the hypothetical trade that derives from it being relatable with households, via census tracts. A parameter is introduced to calibrate for different types of retail activity (defined by Huff as "the parameter estimate appropriate to the type of product class, e.g., 3.191 for clothing and 2.723 for furniture", and therefore introducing a notion similar to that of *range*).

$$P(C_{ij}) = \frac{\frac{S_j}{{T_{ij}}^{\lambda}}}{\sum_{j=1}^{n}(\frac{S_j}{{T_{ij}}^{\lambda}})} \tag{1}$$

"Where:

$P(C_{ij})$ = the probability of a consumer at a given point of origin *i* traveling to a given shopping center *j*;

S_j = the square footage of selling space devoted to the sale of a particular class of goods by shopping center *j*;

T_{ij} = the travel time involved in getting from a consumer's travel base *i* to shopping center *j*;

λ = a parameter which is to be estimated empirically to reflect the effect of travel time on various kinds of shopping trips". (Huff, 1963).

The impact of Huff's model on current research is still visible (Klaesson & Öner 2014; Dolega et al, 2016; Sevtsuk & Kalvo 2017), with modern technologies and more abundant information allowing to model around limitations that have been pointed to Huff's model since its original publication.

Some of the limitations of the model were pointed by Batty (1971), who demonstrated how the sole calibration of the parameter in Huff's model could lead to an infinitude of possibilities (and error), testing for *linear regressions* and *model goodness of fit* in order to verify its applicability.

With that in consideration, a note on recent developments in spatial analysis will be done, later on this chapter.

2.4. ON CLASSIFICATION

Attempts on delimitation of commercial centres have been subject to proposals on classification since at least Proudfoot (1937) who first attempted to systematically describe the organization of commercial clusters within a city. Central Business District would be characterized by “a greater volume of business per area”, with diverse types of stores (jewellery, clothing items) and, with “subordinate importance”, stores selling convenience goods. Also, “the central business district draws customers from all parts of the city” (Proudfoot, 1937: 425).

This is consistent with notions of range and threshold, but also with bid-rent theory: only higher order goods are able to bid for a spot at the centre; the offered density and diversity allow for both comparison and multipurpose shopping.

And at the opposite end: “*the isolated store cluster is the final, and individually the least significant, type of retail structure. These clusters usually comprise two or more complementary rather than competitive convenience goods stores. Thus, there may be (…) a grocery store, a meat market, a fruit and vegetable store, a delicatessen and possibly a small lunch room grouped together at a minor street intersection*.” (Proudfoot, 1937: 427).

In between those, *outlying business districts* are similar to CBD, but smaller, while the *business thoroughfares and streets connecting these clusters*, rely on transient traffic (pedestrians and automobiles, depending on the extension and location).

All notions that would later be modelled are fit within Proudfoot’s classification, but in his case, come from observation: they are supported by what would later be modelled by Berry and Garrison (1958, 1958a), Alonso (1960, 1964) or Huff (1963). And they are also present in Lisbon, as will be seen.

“*The principal cities of the United States, for the most part, possess five types of retail structure. These five types have been named: (1) the central business district; (2) the outlying business center; (3) the principal business thoroughfare; (4) the neighborhood business street; and (5) the isolated store cluster*.” (Proudfoot, 1937).

Jakle and Mattson (1981) make an interesting study on cluster classification: by observing how retail spreads from the business centre, they theorize on *how* these different clusters take form. Studying Champaign and Urbana, Illinois in the period between 1919 and 1979, at 10 year intervals, and focusing on automobile-oriented businesses (gasoline stations and auto repairs), they concluded that several stages occur: the first is when car oriented businesses locate at a symmetrical distance from the CBD, to provide for transient traffic and taking advantage of lower land prices at residential areas; then, other similar activities will locate nearby, which are followed by more diverse activities, spreading along the street that connects to the CBD. Residential use decreases and excess competition eventually drives business to move into other more advantageous parts of town, where the process will start again.

Competitive clustering, multipurpose shopping and even spillover effect (the car oriented services posing as the initial "anchors" in this case study) are therefore verified and described, even though these concepts are not used. Proudfoot's CBD, outlying business districts and business streets and thoroughfares are visible, while in the process of formation (Proudfoot was focused on Chicago, Philadelphia and "the principal cities of the United States" (Proudfoot 1937: 425) where the process had become stable).

On Chicago, a most complete study had been done by Berry (1963, 1967), with Davies (1972) using Newcastle upon Tyne to arrive to his classification:

"(…) *there are three prevailing forms most commonly alluded to: the need to be central to a surrounding consumer body (described as a condition of general accessibility); the need to be aligned alongside a major access route (arterial accessibility), and the need to occupy areas exhibiting prestige, historical affiliations or some particular environmental attribute (special accessibility). These result in three different conformations of activities, called nucleated facilities, ribbon developments and specialised functional areas. The three types of conformation, however, are not necessarily mutually exclusive or spatially discrete."* (Davies, 1972: 278).

Brown (1991) proposes an approach based on form and function, while Guy (1998), though acknowledging's Brown classification's appeal, proposes yet another classification, based on a relation with location (land use in the vicinity). "Classification is essential as a means of understanding and analysing relationships in the world of retailing." (Guy, 1998: 255).

2.5. RECENT DEVELOPMENTS

The economic basis for the better part of location, agglomeration and size were laid more than 100 years ago. Multiple attempts on classification, based on each or several of these attributes have been made since. The 1960's were especially prolific, not only for economic theories, but also social and socio-economic theories: Jane Jacobs' "The Death and Life of Great American Cities", published 1961, introduced retail as an activity who added to the city fabric's richness:

"Not only do public characters spread the news and learn the news at retail, so to speak. They connect with each other and thus spread word wholesale, in effect." (Jacobs, 1961: 70).

The city was presented as more than just a market centre – it was an organic system.

Since the 1960's, much as been written in detraction of the abstract economic models, and as much in their defense. Research has focused, though, in improving shortcomings that couldn't be dealt with at the time they were presented. Some still can't.

Since 1990, new and more sophisticated techniques for analyses started, not only being available, but being available for any user. GIS packages allow for spatial interactions models (Huff, 1963) to be modelled with great ease. But as Clarke (1998: 295) noticed, data must be dealt carefully.

"*What are the drawbacks with spatial interaction models? It has been argued that they are more powerful predictors of store turnovers and spatial interactions than any of the other methods so far introduced. They are however difficult to calibrate (data intensive) and it is likely that they must be disaggregated to fit the case-study under investigation. Given the complexities of consumer behaviour for different types of goods and services it is unlikely that a model which fits the car market, for example, can be taken off the shelf to operate in the bingo market. Thus the user needs to be a skilful spatial modeller. For these reasons, the availability of off-the-shelf- models in some GIS packages now might actually be a rather dangerous development*".

In academia, though, the use off increasingly sophisticated software allows data to be approached in ways that show an improvement on identifying patterns of phenomena that occur within the fabric of a city. Thurstain-Goodwin and Unwin (2000) proposed to eliminate the *granularity* of pre-defined enumeration districts by using continuous density surfaces. Batty (Batty et al, 2004) elaborated on that concept, applying *density* to different variables, thus demonstrating the advantages that arise for visually relating them by eliminating *discontinuities*.

New methodologies allow for innovative approaches: measures of consumption can be used to relate social interaction and city size through a spatial regression model (Borck, 2007); on externalities other than demand, Porta *et al* (2015) use correlation to establish relationships between street centrality and commercial activities; centralities may also be identified by combining diversity with density through convolution (Batty et al, 2015); demand location can be obtained via Twitter, and retail location via Foursquare, therefore introducing real time demand when modelling consumer location choices (Batty et al, 2017).

But finally, and pertaining relevant literature when defining state of the art for this dissertation, the work of Sevtsuk (2010, 2012, 2014) has to be acknowledged as crucial, and has and will be heavily quoted. Recent work (Sevtsuk & Kalvo, 2017) incorporates the Huff model into this extensive contribution for retail location theory.

These authors bring art to its current state (pertaining this dissertation). A serious attempt on reviewing literature led to this point, but it is impossible to sustain that there aren't other researchers dealing with the same subject.

"*It is the current state-of-the-art of knowledge in combination with the open questions in the field, which results in different people thinking along the same lines. Then suddenly new concepts pop up and are being picked up by several groups simultaneously*" (Survival Blog for Scientists, 2010).

2.6. SUMMARY

On retail location, the cornerstone was set with by von Thünen, in relating rent and location. That relation was established by means of *distance* to the centre. In 1826, though, in an agricultural society, cities were "unimportant in the landscape and viewed as parasitic on the honest toil of agriculture" (Alonso, 1960: 2). The introduction of *central function* and *range of a good*, and a consequent hierarchy of places was vital for location theory (the "isolated state" finds finite application in urban studies, especially in a context in which cities started growing *in continuum*). Differences between Thünen's model, and the ones by Christäller (1933/1966) and Lösch (1940/1954) represent a *shift in society*: from one in which the focus was on production (the city was important because convenient, as a trading centre) and the area around the city, to one in which the focus is on goods and services offered at the city, with its *position within a hierarchy of places* depending on the *order of the goods and services* being offered.

On location, Christäller inscribed the honeycomb pattern into planning, with the focus being at the *central place* and not the surrounding plain. Lösch further elaborated that different centres could rely on *different goods* and still maintain its hierarchical position. Higher order goods will be found at higher order centres: their *range* is the sum of the areas that compose the network of influence of all the centres underneath it. Without the constraint of *shape*, Berry and Garrison (1958, 1958a) tested for the inferior limit of that good, introducing the concept of *threshold,* with *range and threshold*, along with *population and distance*, still explaining for the greater part of the location patterns of economic activities. Since the 1960's, and though its many limitations have been recognized, CPT remains subject of discussion (Fujita & Krugman, 1999), even though the focus has shifted to *economies of scale* arising from *concentration of activity* and *spillover effects* (Fujita & Thisse, 2002).

But within the centre, what defines the best use of the land? Alonso (1960) elaborated on Haig's (1927) work and stipulated that the best located areas are occupied by uses that can *bid for the rent* at that specific location. More than a century over von Thünen, the concept is not that different from one owner defining what use is best in order for his farmers/tenants to obtain maximum profit and therefore pay maximum rent; but the owners are multiple and specific crops are replaced by multiple uses, all location sensitive and aiming at maximizing their own profits. Land is at its highest and best use when that use maximizes its value: a concept originating from the work of Fisher (1930) and still being used by Fujita (Fujita et al, 1988) or, for that matter, to the present day.

To the agglomeration of retail activity, specifically, *competitive clustering* (Hotelling, 1927) may explain for a good part, while *multipurpose shopping* and *information spillover* effects all agree on one point: all theories must incorporate the notion that a consumer will most likely make a trip if he's *informed* about what goods he might expect to find, since that trip will have a cost that ads to the price of the good. *Cost of transportation* ads to the overall cost, so Huff (1963) defined that a larger retail area will attract more customers (for they'll expect a more diverse offer). A *dense and diverse* commercial centre is therefore more likely to attract more trips/costumers, and an implementation of Huff's model was published only

last July at Harvard (Sevtsuk & Kalvo, 2017), while diverse spatial data is used to quantify retail agglomeration at UCL's CASA (Batty *et al*, 2017).

A review on state of the art therefore confirms what Brown called a "conceptual cycle" that, more or less every 30 years, leads to the production of relevant insight about retail location. After the surge in the 1930's, with major developments in the 1960's and an apparent renewed interest in the 1990's, research continues as 2020 approaches.

Table 1 – Generational Cycles of Retail Location Theory

Retail Location Theory	Generational Cycle		
	c.1930	c.1960	c.1990
Central Place Theory	Christaller	Berry	Storbeck?
Spatial Interaction Theory	Reilly	Huff	Guy?
Bid Rent Theory	Haig	Alonso	Haila?
Principle of Minimum Differentiation	Hotelling	Nelson	Thisse?

Source: Adapted from Brown (1993)

2.7. RETAIL LOCATION IN LISBON – A BRIEF LITERATURE REVIEW

An overview of the functional structure of the city of Lisbon was presented by Gaspar (1985, based on a study originally made in 1970, and published as a book in 1976). In that, not only the functions (location of goods and services) are analysed, but also their evolution since the 1500's. It's a study on form, function and urban dynamics that therefore allows to understand how, and most importantly, why does activity locate at a certain place (since the original reason that led retailers to locate might no longer exist, while the activity still does).

The importance of this "historic inertia" is acknowledged by researchers.

"Finally, our methodology also falls short in representing retailers' moving costs and historic inertia in location patterns. Locations that have become suitable for retailing for various reasons in history might still remain active today, despite their disadvantages in accessibility compared to alternative locations" (Sevtsuk, 2010: 119)

"*Urban economists argue that the relationship between institutions and the property market within an urban area is not always indicative of a straightforward "best economic use of space" (McNamara, 1979). Other critics base their opposition in a somewhat "historical accident" idea that suggests that urban space is influenced by the past. This implies that the concentration in the urban core may be a reflection*

of a regular urban growth that radiates from a predefined central market place to newly emerged peripheries (…)." (Öner, 2014: 19).

On historic factors that explain for the structure of Lisbon: in 1565, Madalena, S. Julião and Santa Justa were already the parishes with the biggest number of commercial establishments (Gaspar, 1985). These parishes occupied what is now Baixa; and as an additional note, after 1755's earthquake, it was designated that the *algibebes* (clothing items sellers) would locate at S. Julião Street ("A Rua de S. Julião, n.d.). During the reconstruction that took place after the earthquake, two parishes offered safe haven to commercial establishments: Santa Isabel and Santos-o Velho (Gaspar, 1985). Santa Isabel was the name of the parish that since 2012 is named Campo de Ourique.

In Lisbon, for the period before 1976, Gaspar (1985) defined the following areas as homogenous:

1) Baixa-Chiado is the historical center of the city, where higher order central functions locate: banking, public administration and personal use items of the highest order ("comércio de retalho de produtos de consumo ocasional de nível hierárquico mais elevado, particularmente no campo de objectos de uso pessoal");
2) Cais do Sodré is associated with port-area activities (the study is from 1970: sea trade, especially with african countries, then portuguese colonies, was still very much relevant, and based on that area);
3) Avenida-Marquês is the result of the first expansion from the centre to North, along the Avenida and around the Marquês de Pombal roundabout, with commercial activity focused on services (e.g. hotels, medical doctors' offices);
4) Avenidas Novas, though still a high-end residential area, was starting to attract central type functions, with accessibility (by car) eventually explaining it;
5) Almirante Reis (which Gaspar calls "ribbon", a classification also used by Berry (1967) and Davies (1972)) is a special case within Lisbon: since it cannot compete with either the centre (Baixa) or its expansion area up North along Avenida/Marquês, it has become a ground place for activities that demand large leasable areas but can't afford to bid at a high price, like furniture and home appliances (later on this dissertation, the sum of those will be called "household articles"). Its accessibility to people from outside the city explained for its commercial success.
6) And finally, the "Áreas Anexas ao Centro" ("Areas Adjacent to the Center") present multiple activities, and some specialized clusters (like an antiques stores' cluster in Rua da Escola Politécnica – Rua D. Pedro V, that still exist to this date).

A classification is proposed: Baixa is the CBD. Alcântara, Campo de Ourique, Benfica, Alvalade, Praça de Londres – Av. Roma are "regional centres" ((Gaspar notes that this is "american terminology" in reference to Hoyt (1957)). And smaller "community centres" are distributed alongside town.

A remark on how a closer look might reveal Praça de Londres – Av. Roma to be more similar to Campo de Ourique that it might appear just from looking at *types* of activities (and not at specific, individual activities) is especially interesting: it implies that the disaggregation of data might indeed lead to different conclusions. It's not a nod to Openshaw's Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984), but it is about the importance of *scale* and *zoning* in any analyses.

The study presents several tables and maps, zoning functional and social areas (the later by quality of the activity: "rich" in Baixa-Chiado, "poorer" on Martim-Moniz-Almirante Reis) and producing insightful views on how the city may develop: the area around the Gulbenkian Foundation might become relevant (and this is something that won't be studied on the scope of this dissertation but, hopefully, later: an equipment/museum as an anchor for commercial activity, and a "driving force" for liveability in general), public administration might move to Marquês (it has been moving away from Baixa in recent years), leaving the area of Baixa for higher order commercial activity and cultural activity. "Modest" retail might move to Martim Moniz and up north along Almirante Reis, amongst many others.

In 1985 the book was republished, and Gaspar added an introduction, recognizing that while some updating was due, what had been written previously was still very much valid and what was predicted had, in general, come to happen. Port-area activities had spread from Cais do Sodré along the riverfront; Baixa hadn't lost relevance, but that was mostly because it had become a central point in commuting trips, both from residents from Tejo south bank commuting by ferryboat, and by residents coming from North by train. The axis Baixa–Almirante Reis–Alvalade had revealed some dynamics, attributed, partially, to the way the subway network was being developed (as was some of the dispersion: Estrada da Luz, Estrada de Benfica). Specific areas revealed specialization (e.g. restaurants in Bairro Alto). A final note was made on urban planning: it was being made thinking on form, not on function. A contradiction was noted: in one side, urban planners and architects trying to "revive the street"; and in the other, the number of shopping malls in which the city was "burying itself" (Gaspar, 1985: 12).

For the period after 1985, pertinent information was found in the work of Barata-Salgueiro (2001). In general, it's noted that Baixa has lost its place as the CBD, with the migration of businesses and offices to Marquês de Pombal, City Council services to Campo Grande and financial activities to Berna-Forças Armadas. On the commercial structure, specifically, it is mentioned that after 1985, it has changed. Commercial supply became more diverse, and new ways of doing business were introduced (franchise stores, hypermarkets). High-end commercial activity ("rich" in quality, as defined by Gaspar) is definitely installed in Baixa-Chiado, but the axis Av. Guerra Junqueiro–Praça de Londres–Av. de Roma is presenting itself as a valid alternative. Other category of centres have appeared: shopping malls. Amoreiras, closer to the ring that encloses the city centre, and Colombo and Vasco da Gama, closer to the ring that encloses the city itself. These centres compete directly with Baixa as higher-order centres. Specialization noted by Gaspar in Bairro Alto by the presence of restaurants has become more accentuated by the co-location of bars and nightclubs. Other areas also show this tendency: Almirante Reis (household articles) and Av. da Liberdade (personal use items – luxury segment). Expo98 impacted

on port-area activities on the northern part of the city, which has been subject to extensive renewal and now presents a wide number of activities (commercial, business, housing, etc.).

In 2011, Cachinho (2011) proposed a classification of retail location based on Brown's (1991), and therefore, on form and function. Function is divided into general, specialist and ancillary, while form is separated into cluster (unplanned), cluster (planned), linear and isolated. Some examples are given:

1) Baixa is presented as a general (function) unplanned cluster (form) while Estrada de Benfica is a general linear cluster. Corner shops and supermarkets are isolated in form and general in function.
2) Bairro Alto is a specialist unplanned cluster, and Rua do Benformoso a specialist cluster of linear form. Furniture stores are isolated in form, and specialist in function.
3) Snack bars are ancillary in function, while in form, they present themselves both as unplanned clusters (within Business Districts) or of isolated form (in health clubs and cinemas).
4) Shopping centres are planned clusters: the larger ones are of general function (Colombo), and the smaller ones (like "Shops in the Airport"), ancillary.

Hierarchy is not evident (as is expected "there is no doubt that Brown's innovation resides in the way he dispenses with the notion of hierarchy in business clusters" (Cachinho, 2011)), but underlies the classification: "isolated" or "ancillary" reveal a smaller importance than "cluster" and "general". This lack of apparent hierarchy, though, creates a "fracture" that doesn't allow for an immediate comparison with previous analyses of Lisbon's retail structure, as described. The larger number of categories (3 functions and 4 forms: 12 categories) also allows for a finer analysis than that of "regional" and "community centres".

As will be noted in Chapters 3 and 4, this classification was of use when interpreting the structure of retail in Lisbon, while also considering the following: *"besides the city centre, it is important to distinguish two types of centres: (i) those of convenience or proximity, oriented to the satisfaction of the basic necessities of the population, and therefore with a close connection to the supply of consumer goods of frequent acquisition; and (ii) centres of regional influence, oriented to offer a great diversity of goods and services, most of them occasional acquisition"* (Cachinho & Salgueiro, 2016: 98).

A summary of the added work of Gaspar, Salgueiro and Cachinho on retail in Lisbon, as presented, thus allows for a brief characterization of retail activity in Lisbon, up to the present day, thus closing state-of-the-art (much else has been written by these authors on this subject; choice was made considering the purpose of a short literature review).

3. AN EXPLORATORY ANALYSIS OF COMMERCIAL DATA FROM LISBON

3.1. DATA

Commercial activity information was available as point features, within a database containing 14.673 locations for 1995, 16.092 locations for 2002 and 17.035 locations for 2010, which included both retail and restaurant establishments (CML, 2016). Each feature shape had an associated feature table with information pertaining Economic Activity Description and Economic Activity Code (based on the official Classification of Economic Activities – "CAE"), amongst others (namely address and designation of the establishment).

Population and housing data was available as polygon features. INE uses a zoning system of statistical subsections ("subsecção estatística"), with each polygon corresponding to a block in urban areas (INE, 2014, 2014a). This database contained 3700 polygons for 1991, 4390 for 2002 and 3623 for 2011, and numerous information (it's the geographical support for the data collected during the national population and housing census ("Censos")). Variables pertaining Population and Primarily Non-Residential Buildings were chosen, to account for demand: the first being related with residents and the second with workers/employees. There was no other available source to account for employment, and these two variables were available for all years (even though INE's database contains numerous information, not all variables are available for all years).

3.2. PRELIMINARY ANALYSIS OF DATA

Data was handled in several ways during this phase. Considering the dataset and examples from literature review, focus was made on clustering and diversity. Change was also analysed.

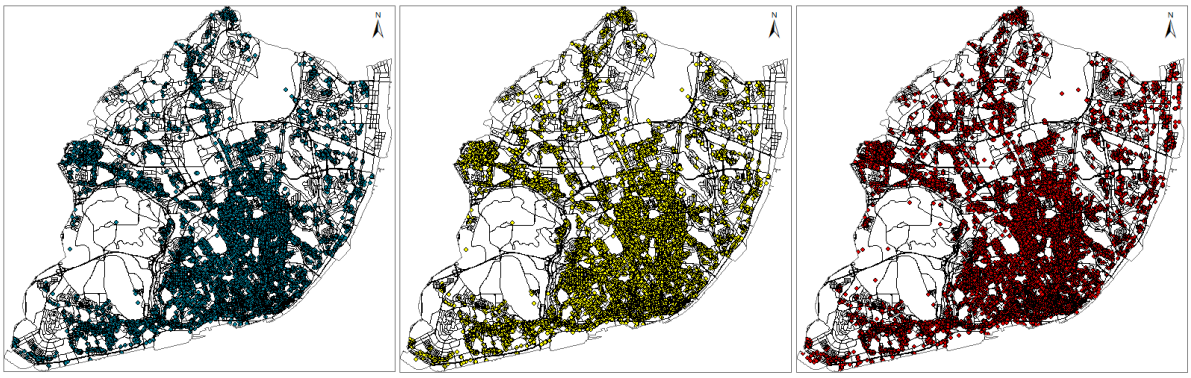


Figure 3 – Establishments (Total): 1995, 2002 and 2010 (left to right)

A first impression was that focusing on the entire dataset, presented as point features, would produce very limited results. Establishments appeared clustered at Baixa–Chiado and Bairro Alto, and along the Almirante Reis–Av. de Roma axis, up to Alvalade, with the area between Marquês de Pombal and Saldanha also showing relevant presence of commercial activity, as did Campo de Ourique and Estrada de Benfica (Figure 3). This was expected, and in line with what had been stated by Gaspar in 1985: Baixa as the higher order centre, complimented by “regional” and “community” centres and some disperse lower order commercial clusters.

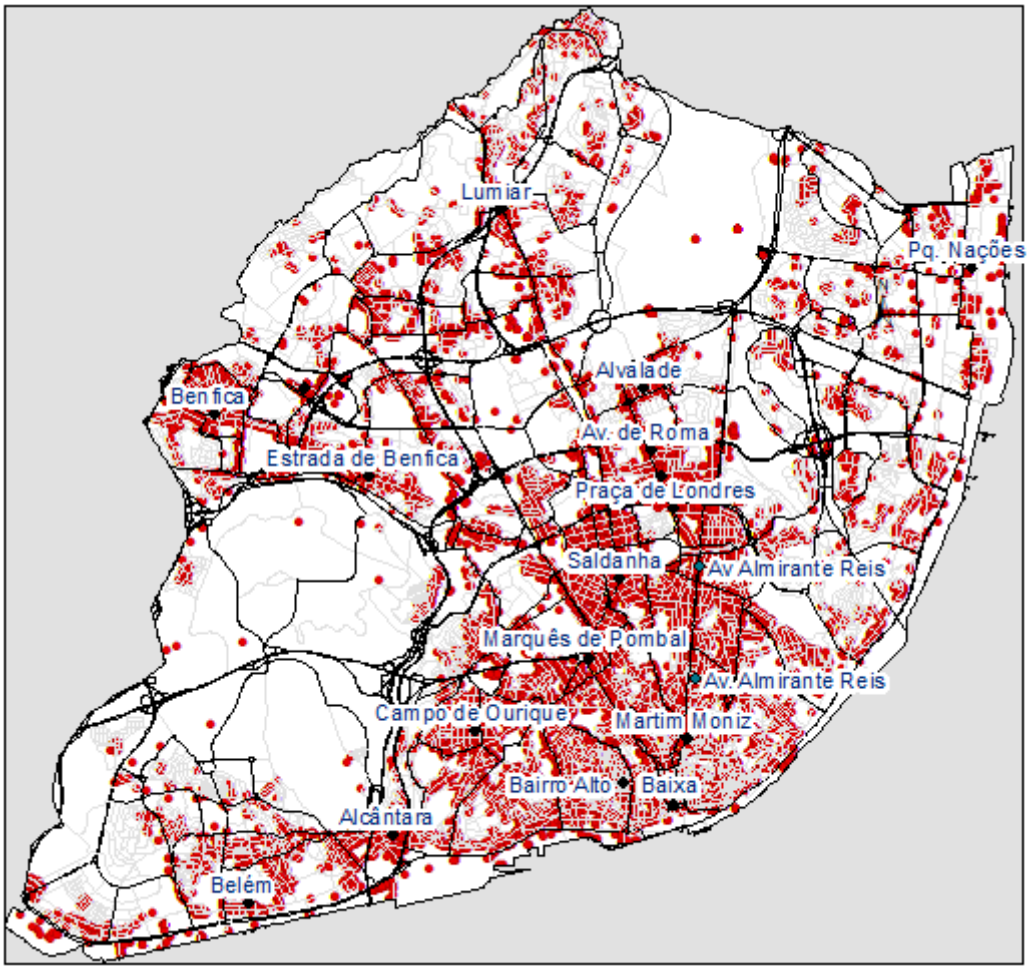


Figure 4 – Commercial Locations within Lisbon

The dataset was then divided into 10 groups (types) of activities:

1) Foods (supermarkets, bakeries, groceries and similar establishments);
2) Personal Use Items (mainly clothing and clothing accessories);
3) Household Articles (mostly furniture and home appliances);
4) Health and Hygiene items (essentially pharmacies and optical shops);
5) Leisure items (sporting goods, bookshops, music stores, etc.);
6) Repairs (mostly auto-repair shops);

7) Other items (all other items non-included in the remaining types);
8) Restaurants and similar establishments;
9) Cafes and similar establishments;
10) Bars and similar establishments.

Types 1 to 7 represent the retail sector, and the remaining, the restaurant sector. This partition was made considering the classification used by the City Council (CML/DMEI, 2009), which itself is an adaptation of the Portuguese Classification for Economic Activities ("CAE"). It's presented in Annex I, in portuguese language (attempts at translation didn't produce a satisfactory result: some activities are very specific). Some adjustments were made: Cafes, Restaurants and Bars were considered as a whole by City Council, while here are analysed as separate types; Foods is divided into "Specialized" and "Non-Specialized" by the City Council, and are here considered as a whole. There was, in general, an attempt to establish types considering the different *ranges* and *thresholds* of the goods being sold, and based on the *frequent* or *occasional* nature of the acquisition, as well as in the goods themselves.

3.3. AN ANALYSIS ON CLUSTERING – COMMERCIAL CENTERS

Considering what was stated on Chapter 2 about distinct centres, both hierarchically and pertaining specialization, and the visually apparent clustering of activity, analysis of the data proceeded with an analysis on clustering patterns. A similar dataset had been analysed for that purpose (Alho & de Abreu e Silva, 2014), but the associated feature table had measurable attributes (establishment areas and number of employees). Although with a different scope (urban freight), that analysis revealed establishments to be clustered both globally and locally (with Global Statistics of Spatial Association used "for validation of the clusters beyond what was visually apparent" (Alho & de Abreu e Silva, 2014:3) and Local Statistics of Spatial Association to identify ""hot" spots, which can be considered commerce islands, e.g., commercial centres or retail parks" (Alho & de Abreu e Silva, 2014: 3)).

The dataset being analysed here had one relevant attribute for the analysis: location (and was therefore *incident data*, defined as "points representing events or objects where your focus is on presence or absence rather than some measured attribute associated with each point" (ESRI, 2016)). For this type of data, Optimized Hot Spot Analysis is appropriate ("Given incident points or weighted features (points or polygons), creates a map of statistically significant hot and cold spots using the Getis-Ord Gi* statistic. This tool "(…) automatically aggregates incident data, identifies an appropriate scale of analysis, and corrects for both multiple testing and spatial dependence" (ESRI (2016)). More on this tool at ESRI (2016)).

Optimized hotspot analysis was therefore used for a visual overview of data, disaggregated into different sectors and types.

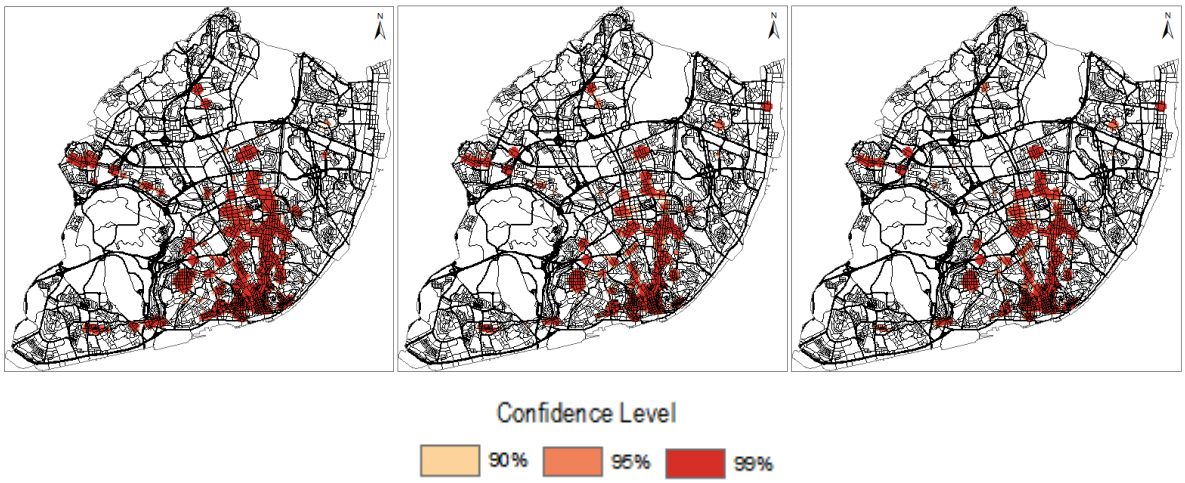


Figure 5 – Total Establishments. Optimized Hotspot analysis results, 1995, 2002 and 2010 (left to right)

An analysis on the results showed that this tool produced data that was visually easier to interpret than point data. The conclusions were the same, but the visual results more satisfactory – it better complements the previous analysis pertaining commercial centres. Considering this, optimized hotspot analysis was done for the 2 different sectors, and all 10 different types of activity.

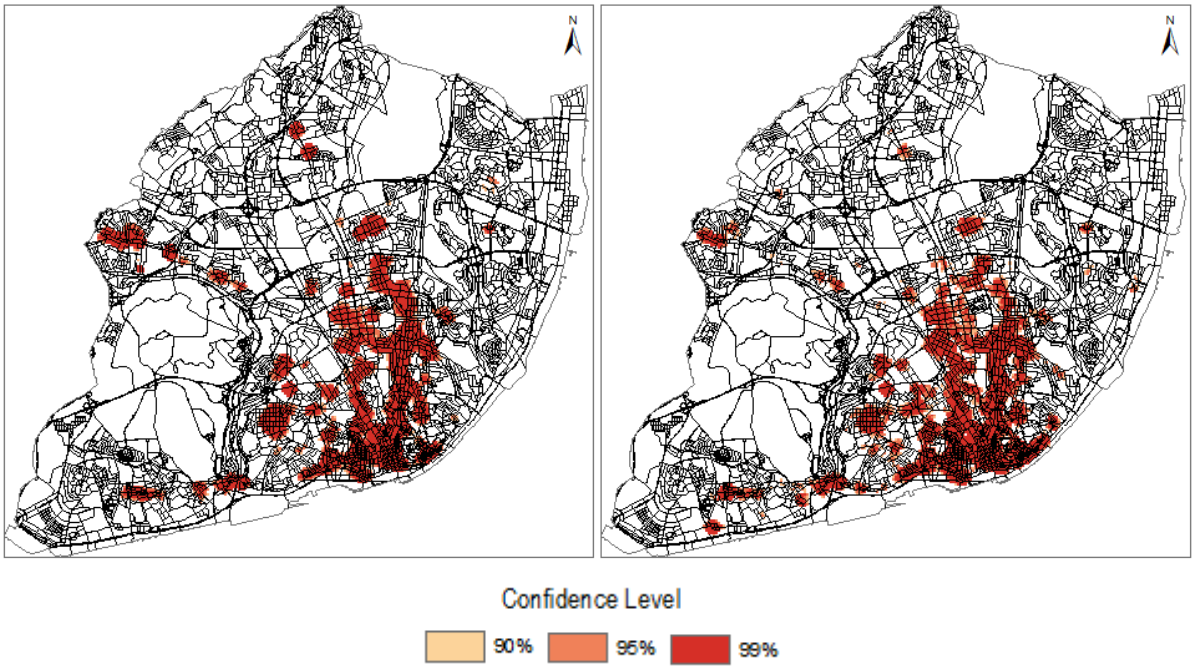


Figure 6 – Retail (left) and Restaurant Sector (right). Optimized Hotspot analysis results, 1995

Comparing Retail with Restaurant, retail presents a pattern which is similar to the one in total activity. That is expectable, since retail amounts for a larger percentage within total establishments. But the restaurant sector also presents a similar pattern, revealing how pertinent it is to consider both sectors, and not only retail, when proceeding with an analysis on commercial activity.

Proceeding to separate types, analysis starts with Foods, 1995.

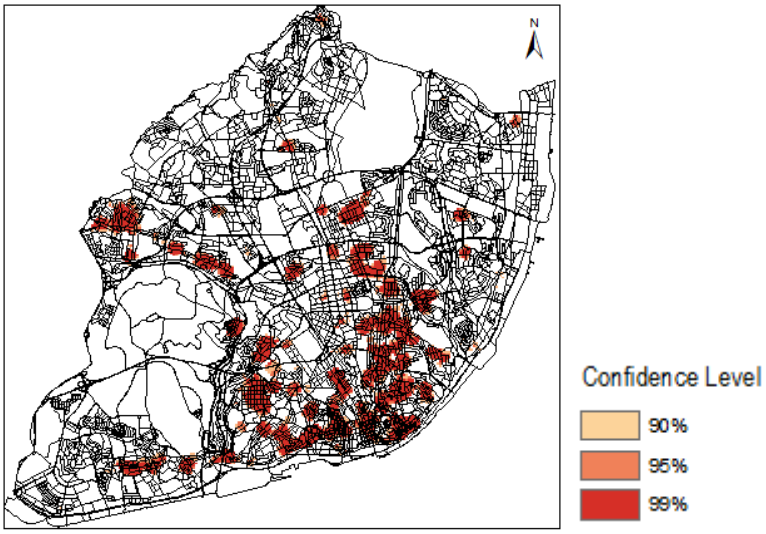


Figure 7 – Foods. Optimized Hotspot analysis results, 1995

The clustering pattern of Foods is compatible with the previously presented, and to some extent, similar: one might be misled into thinking all types are similar at this level of disaggregation. They aren't, as will be shown in the following figures.

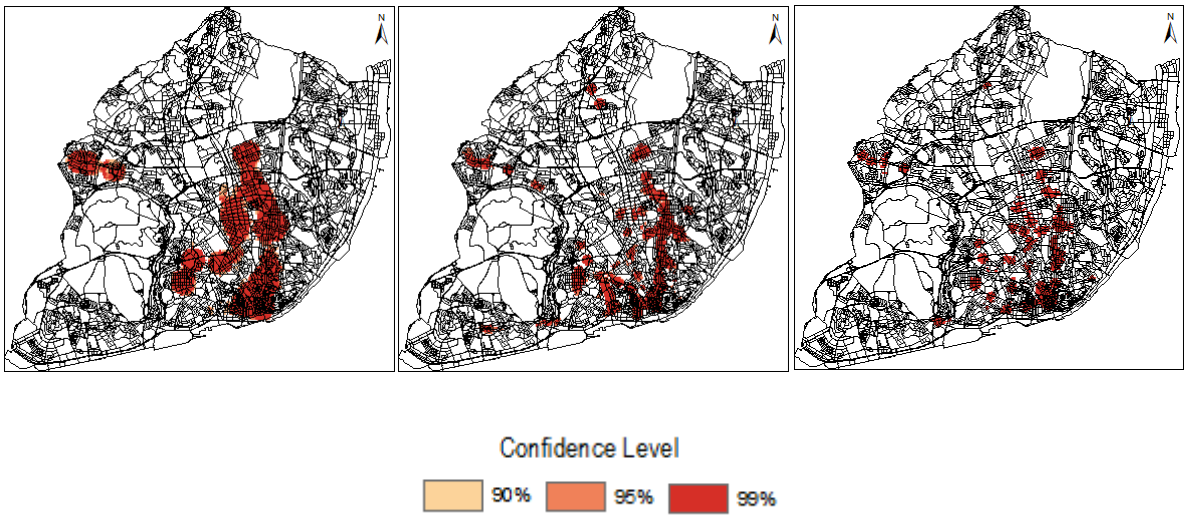


Figure 8 – Personal Use Items, Household Articles, Health and Hygiene. Optimized Hotspot Analysis, 1995

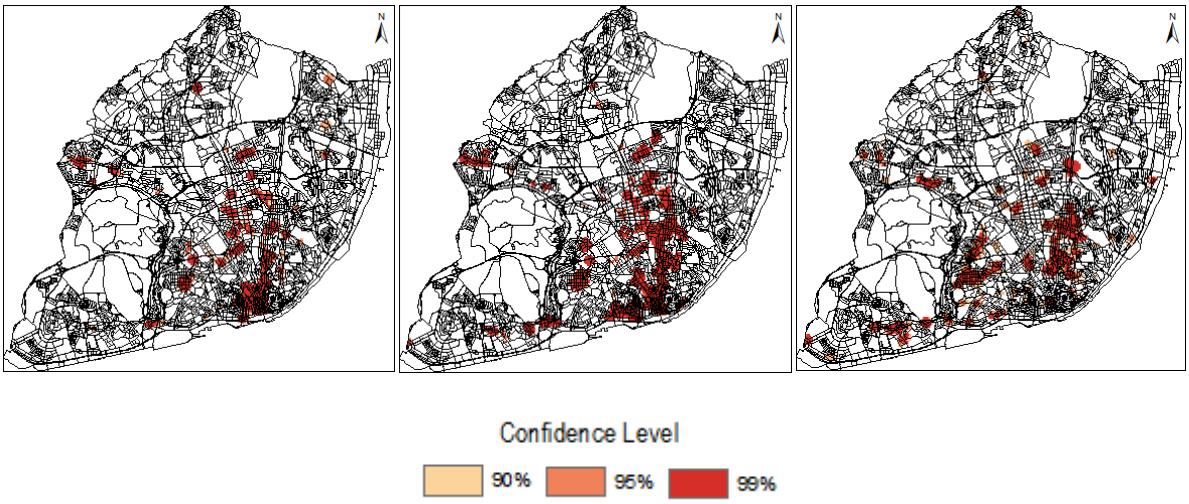


Figure 9 – Leisure items, Other Items, Repairs. Optimized Hotspot Analysis, 1995

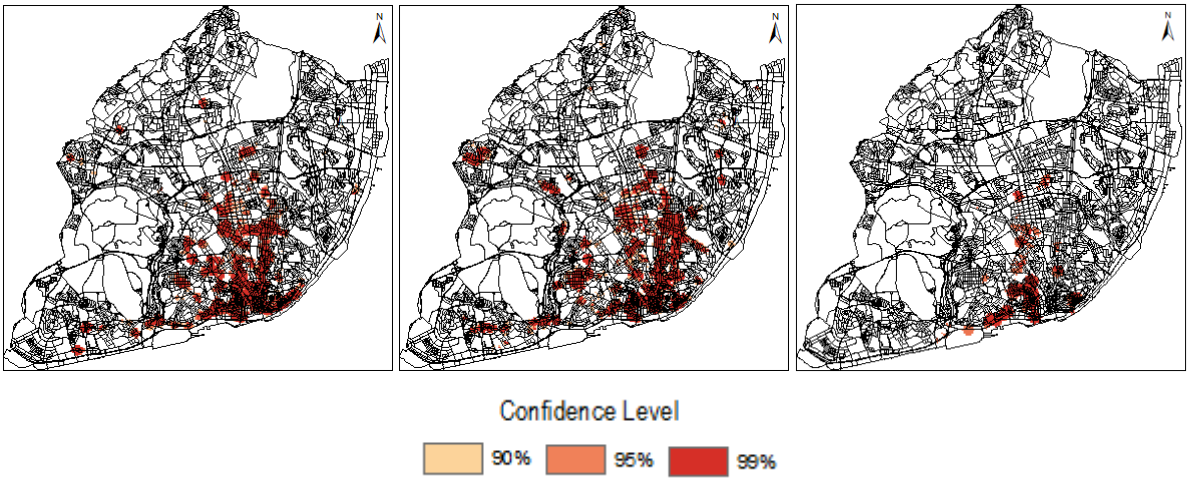


Figure 10 – Restaurants, Cafes, Bars. Optimized Hotspot Analysis, 1995

As shown, Foods presents a homogenous distribution amongst all commercial centres. It seems to be matched only by Cafes, and, eventually, Other Items (frequent acquisition – *low threshold* – goods (Berry & Garrison, 1958). Leisure Items or Health and Hygiene seem to concentrate on Baixa, Saldanha-Av. Novas and Av. Roma-Alvalade, with Repairs being the only type that avoids locating at Baixa (one can risk saying it's a *bid-rent* classical example of an activity that, not being particularly location sensitive, finds optimal locations (considering price/area) outside the higher order cluster (Alonso 1960; 1964). Personal Use Items displays the most concentrated pattern, suggesting the effect of *competitive clustering*. (Hotelling 1929, Eaton & Lipsey, 1975). Household Articles displays a pattern similar to the one of frequent acquisition goods. This might be related to the fact that, although Household Articles have a higher threshold (higher order activity) their characteristics (store area, accessibility) aren't compatible with the city centre/higher order cluster. An optimal location is found in a "commercial strip" leading to the central business district, where the rent is lower, but the *visibility* is still high (Proudfoot, 1937; Jakle & Mattson, 1981).

In sum, in 1995, the higher order cluster of Baixa concentrates occasional acquisition goods (higher-order), complemented by the areas of Saldanha-Av. Novas and Av. Roma-Alvalade, which is consistent with what was stated by Gaspar, a decade earlier (Gaspar, 1985).

As for the restaurant sector, Restaurants and Bars seem to concentrate at Bairro Alto, but also Marquês de Pombal – Saldanha. Cafes present a different pattern. This can be used to explain why the restaurant sector wasn't analysed as a whole: the goods sold at Cafes are within a different scope from the ones sold at Bars and (some) Restaurants: that of frequent acquisition goods. The relationship between Bars and Restaurants also seems to imply that the later cater to different customers *depending on where they are located*.

At this point, though, an analysis on change is due: location of commercial centres in 1995 was analysed and revealed compatible with literature. But how did the city's commercial structure change between 1995 and 2010? For an analysis on change, a different approach was devised.

3.4. AN ANALYSIS ON CHANGE – COMMERCIAL DENSITY

3.4.1. A NOTE ON METHODOLOGY

A preliminary idea on how to analyse change came from the work of Thurstain-Goodwin and Unwin (2000), who elaborated on using "continuous surface representations" (densities) for identifying regularities within urban areas. This idea was further developed by Batty (Batty et al., 2004: 327): by "smoothing the data to iron out the inevitable discontinuities that take place from data that are originally represented by land parcels and/or the fine scale postal geography" it is possible to better to better identify a pattern.

Commercial locations were available as point features. Therefore, point density was used to create continuous surfaces, by calculating "a magnitude-per-unit area from point features that falls within a neighbourhood around each cell" (ESRI, 2016a). Thus, for commercial establishments, density represents the number of point features found at a fixed distance (radius) around each cell, per unit area (km2).

Creating density surfaces also allowed to introduce map algebra into the analysis: by subtracting density raster maps, it is possible to observe where variation occurred, linking data and space.

This implied, though, the choice of a radius. A range of radii was tested, with radii smaller than 100m producing surfaces that were visually very similar to the actual points, while radii larger than 600m represented areas larger than 1.3km2, with the connection with city centres being lost, consequently (the larger cluster – Baixa – occupies an area smaller than 0.5km2). Smaller radii thus *added to the*

resolution, but increased the difficulty in discerning patterns, while larger radii *implied significant data loss* (though revealing patterns that could be relevant at regional scale).

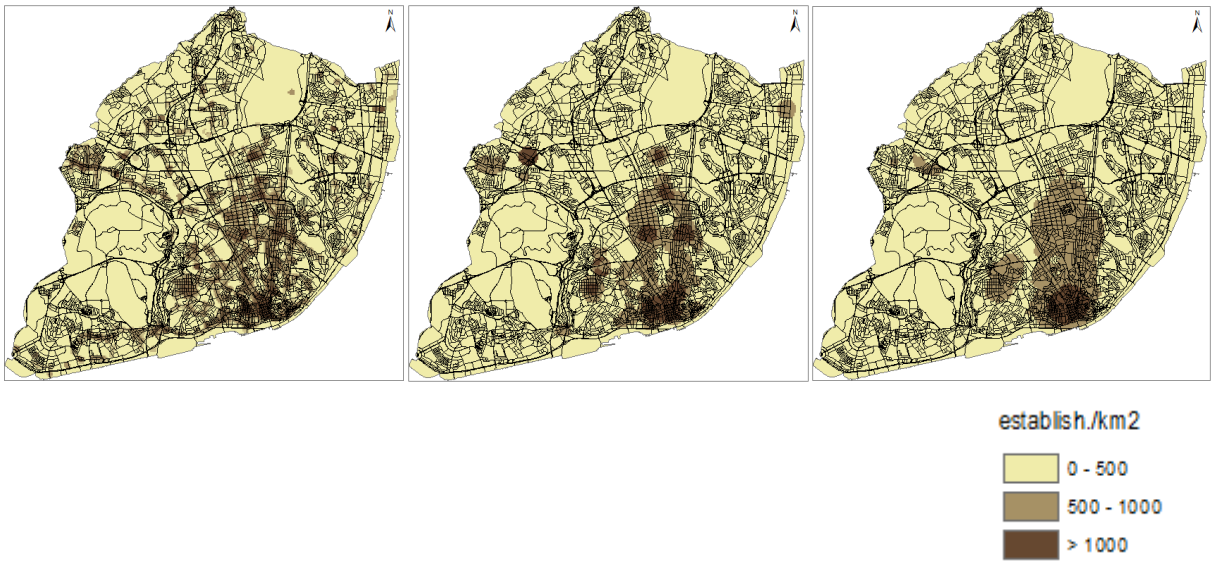


Figure 11 – Total Establishment Density (2010) using 100m, 300m and 600m radii (left to right)

Eventually "a range of resolutions is necessary to adequately describe the fit of models with reality" (Costanza, 1989: 201).

That, though, was not the scope of the research, so one specific radius had to be chosen. To choose the one that presented a better relationship between data presentation and data accuracy, a testing methodology was drafted . First, for raster's comparisons, the Band Collection Statistics tool was used, and Table 1 was produced.

Table 2 – Correlation Matrix of Rasters for Radii 100m to 600m using Band Collection Statistics

Radius	100	200	300	400	500	600
100	1	0.99	0.96	0.90	0.80	0.61
200	0.99	1	0.98	0.92	0.82	0.63
300	0.96	0.98	1	0.96	0.87	0.67
400	0.90	0.92	0.96	1	0.93	0.73
500	0.80	0.82	0.87	0.93	1	0.83
600	0.61	0.63	0.67	0.73	0.83	1

"The correlation matrix shows the values of the correlation coefficients that depict the relationship between two datasets. It is the ratio of the covariance between the two layers divided by the product of their standard deviations". More on Band Collection Statistics can be found at ESRI (2016b).

The correlation matrix allowed revealed a high correlation between radii below 300m; above that value, the correlation was quickly lost. This indicated that using radii up to 300m would not present significant difference in terms of data accuracy.

To support this conclusion, a fishnet of 150m x 150m polygons was created (meaning an area of 22.500m2, similar to the average size area of INE's statistical subsections, that would later be used). This allowed to calculate the real density of establishments (the sum of all establishments for each cell divided by its area) and compare it to raster densities (the Zonal Statistics as Table tool "summarizes the values of a raster within the zones of another dataset and reports the results to a table"; more on this tool can be found at ESRI (2016c).

An extract of that table is shown on Table 3 (the actual Table has 3642 rows, one for each cell). Using that table, it was possible to attain the correlation between real density at cell size, and the densities produced by each of the different radii, using the CORREL function within Excel (Pearson Product-Moment Correlation Coefficient). More on this tool can be found at Microsoft (2017).

Table 3 – Correlation of Observed Density and Raster Densities using CORREL (Microsoft Excel)

	Real	Different Radii					
		100	200	300	400	500	600
Density (Row 4 (e.g.))	83,48	89,12	87,81	77,21	48,40	29,17	22,22
Correlation*	1	0,97	0,94	0,90	0,84	0,79	0,76

*The Correlation presented is total correlation, considering all rows.

The result is *compatible* with the one obtained in Band Statistics (even though it's not the same comparison: on Band Statistics Collection, there wasn't a raster with the real density of establishments to be compared).

Finally, the radius was set at 300m, which was considered appropriate as a compromise between data fit and data visualization, and analysis started, beginning with Total Establishments Density Variation.

A final note must be made before the analysis: the density is, as explained, not a real density. It's the number of points in a 300m radius of a cell, per square kilometre. So, firstly: *if a cell is 22.500m2 and 1 establishment is gained (or loss), this won't appear as 1 establishment/cell, but as 44.4 (roughly) establishment/km2* (this is how density is calculated, always, but establishment density isn't a measure as usual as population density (e.g.)). And secondly: *since the density is being calculated considering a 300m radius around the cell, a variation of plus (or less) 1 establishment (e.g.) within a 300m radius will appear as a variation of plus (or less) 44.4 establishments/km2 at cell-size*. The map should be analysed with this in mind: the values presented area those of "smoothed" densities, considering a 300m radius, per km2. They're not "real" densities. A compromise between accuracy and visualization is made, in

order to better identify patterns. Figure 12 shows how real density would appear, and how it will be shown.

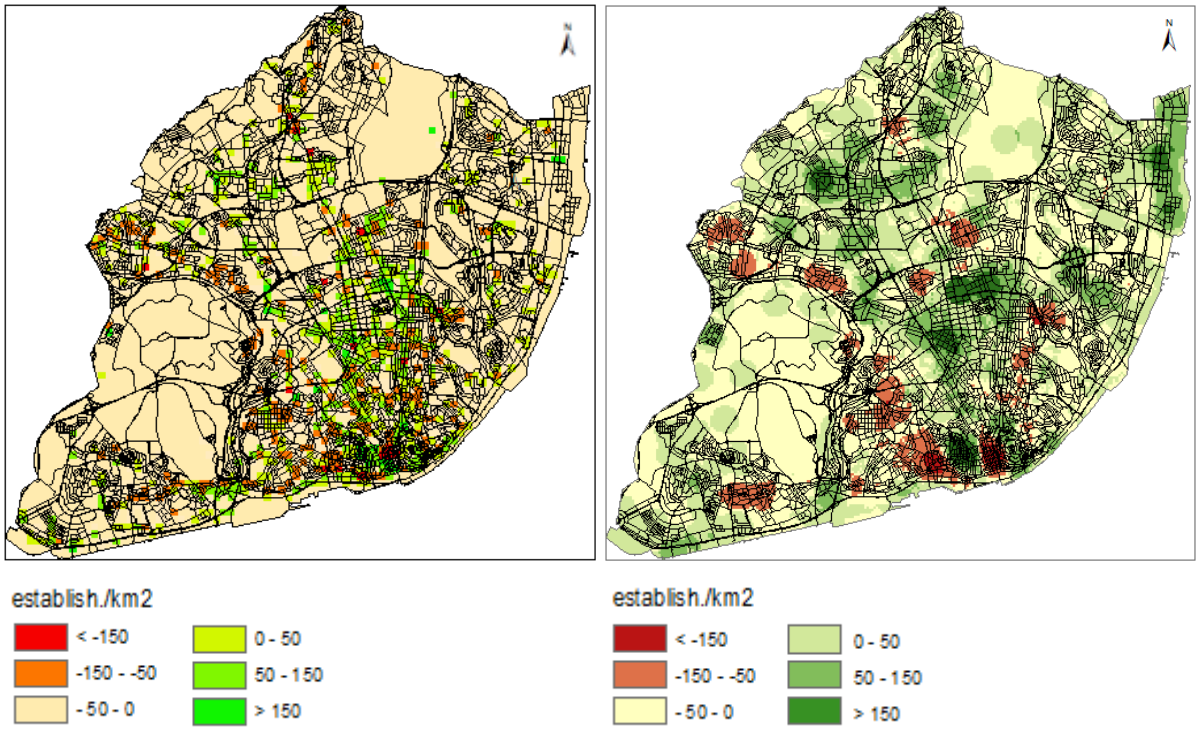


Figure 12 – Total Establishments Density Variation, 2002 to 2010. Real Density (left) and Smoothed Density (right)

Another way to interpret the maps would be by introducing a legend that was only based on a colour-scale, from red, representing “significant loss”, to dark green, representing “significant gain”, like shown below (Figure 13).

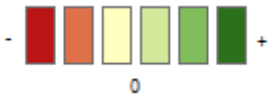


Figure 13 – A possible legend for Density Variation Maps

For an exclusively visual analysis, maps can be interpreted in this manner. For the sake of keeping a relation with ground truth, all figures are presented with numeric intervals, as is more usual.

3.4.2. Analysing Change

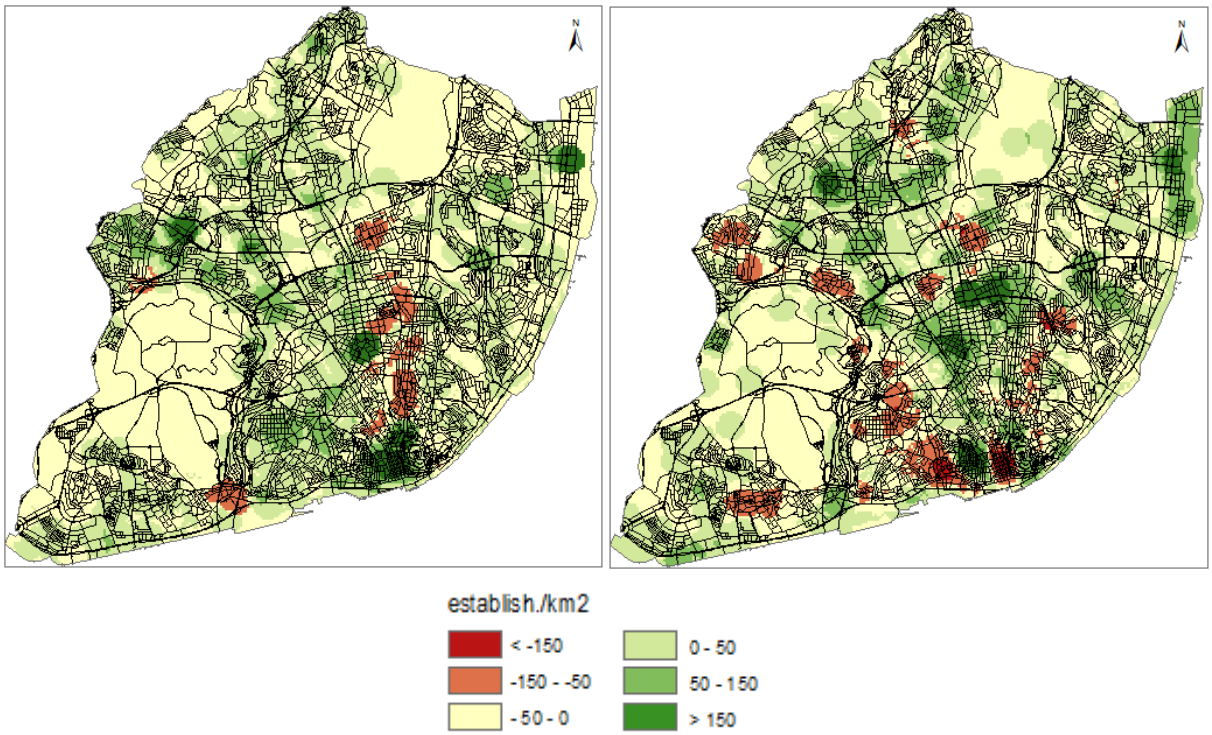


Figure 14 – Total Establishments Density Variation: 1995 to 2002, and 2002 to 2010 (left to right)

The period between 1995 and 2002 can be directly related to a stage in the evolution of shopping malls characterized by the opening of few but (much) larger units when compared with previous decades (Cachinho, 2002). Apart from Bairro Alto-Baixa-Martim Moniz, they amount for all other significant, positive variation that is visible: Colombo can be seen northwest, Vasco da Gama northeast, and close to the centre, in the Saldanha neighbourhood, the opening of Atrium Saldanha, Saldanha Residence and Picoas Plaza is also perceptible, since these 3 smaller malls have a significant commercial area when considered as a whole.

The impact of the shopping malls is evident on Avenida Almirante Reis, Avenida de Roma and Alvalade. This period thus represents (roughly) a decade where commercial activity (both retail and restaurant) was still growing, but malls were presenting themselves as desirable location alternatives.

In the following period, this location effect was extended to other areas traditionally associated with commercial activity, including Baixa, but also Campo de Ourique and Estrada de Benfica. Two additional shopping malls opened (and are visible) in a central area: Campo Pequeno, less than 1 km north of Saldanha, and Acqua Roma, east of it.

In this second period, Vasco da Gama became central in a new neighbourhood of Lisbon: Parque das Nações (the former Expo98 grounds). Up north, the residential areas of Telheiras and Lumiar / Alta de Lisbon also revealed some growth dynamics.

The city was therefore presented with alternative centres to traditional commercial centres: shopping malls, offering a multitude of shopping opportunities in a controlled environment: an “arranged scenery” destined for consumption (Cachinho, 2006). Vasco da Gama and Colombo, with a vast number of stores, good accessibility and parking possibilities, posed themselves as alternatives to Baixa itself (as stated by Salgueiro (2001)).

Figures 15 and 16 present the variation for retail and restaurant sectors, as previously defined.

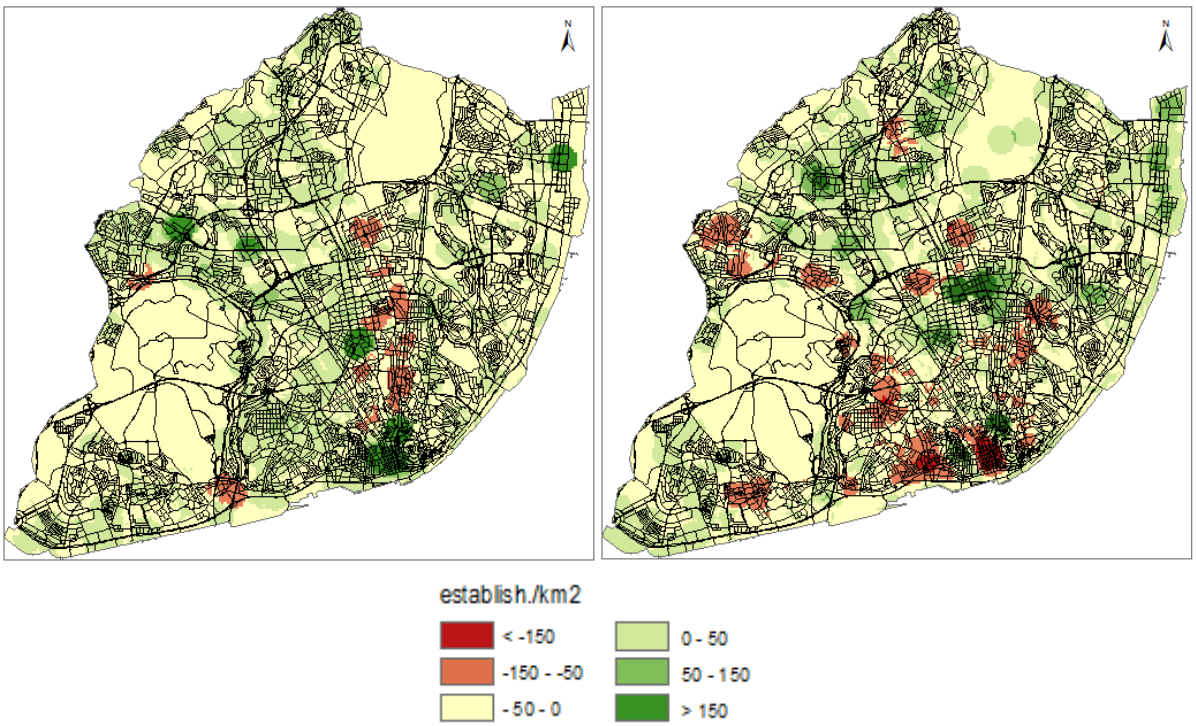


Figure 15 – Retail Sector (as defined) – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right)

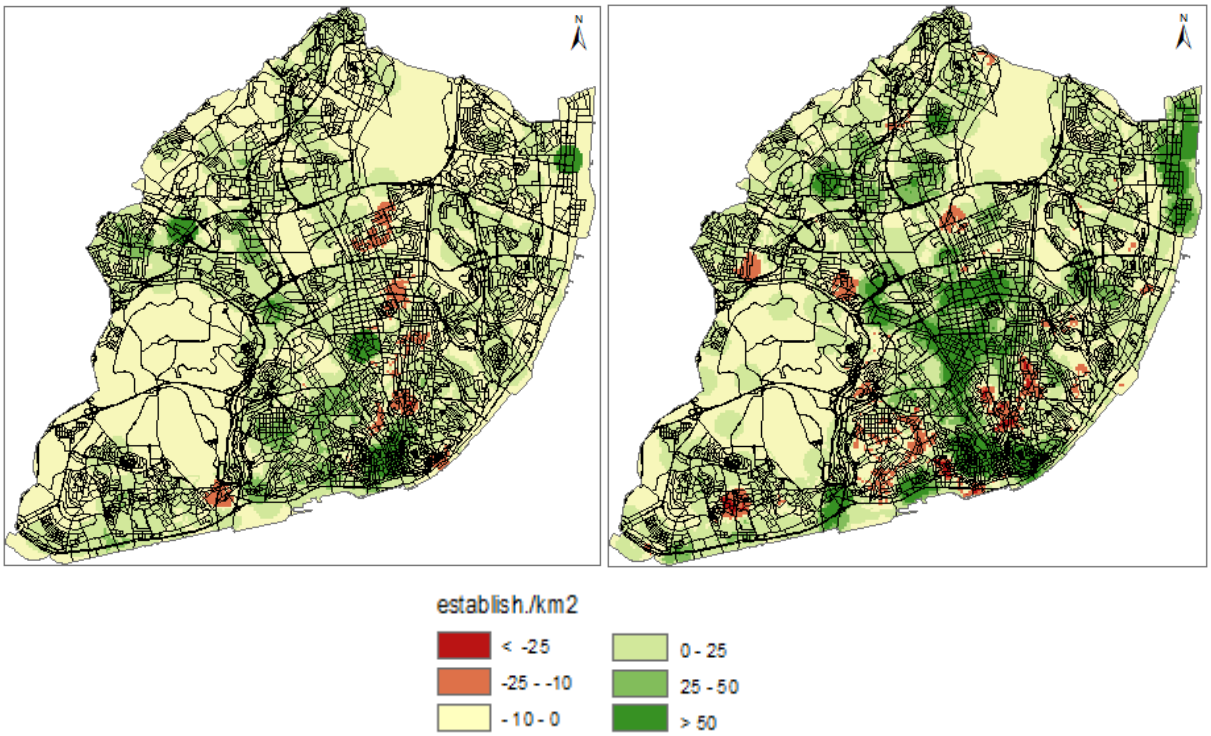


Figure 16 – Restaurant Sector (as defined) – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right)

Analysed separately, the two sectors show different trajectories.

In the period between 1995 and 2002, the impact of the shopping malls is noticeable in both sectors, but the restaurant sector is, in general, growing steadily, while retail, outside shopping malls, shows significant growth only in Bairro Alto-Baixa-Martim Moniz, and therefore contributes for most of the losses in total commercial activity observed in Figure 14 in this period.

As for the following period, the restaurant sector shows significant growth (mostly) all over Lisbon, while in retail, the street-level commercial areas that hadn't been affected between 1995 and 2002 suffered losses in this period. Alvalade shows persistent loss throughout both periods. As for Baixa, it lost all that had gained on the previous period - and more.

Tables 4 to 6 sum up this information, referring to number of establishments and their variation in the periods between 1995-2002 and 2002-2010, thus presenting a general overview of the studied period.

Table 4 – Total Number of Establishments and Variation 1995-2002 and 2002-2010

	Year/Variation	1995	2002	Variation	2010	Variation
Type	Description	Total Number	Total Number	1995-2002	Total Number	2002-2010
1	Foods	2259	2166	-4.12%	1868	-13.76%
2	Personal Use Items	2387	2678	12.19%	2566	-4.18%

3	Household Articles	1356	1503	10.84%	1265	-15.83%
4	Health and Hygiene	586	642	9.56%	740	15.26%
5	Leisure	1519	1657	9.08%	1616	-2.47%
6	Repairs	598	683	14.21%	944	38.21%
7	Other Items	2129	2306	8.31%	2386	3.47%
Total (Retail Sector)		10834	11635	--	11385	--
8	Restaurants (and similar)	2139	2515	17.58%	3202	27.32%
9	Cafes (and similar)	1597	1833	14.78%	2212	20.68%
10	Bars (and similar)	103	109	5.83%	236	116.51%
Total (Restaurant Sector)		3839	4457	--	5650	--
Total		14673	16092	--	17035	--

Table 5 – Establishments outside Shopping Malls – Total Number and Variation 1995-2002 and 2002-2010

Year/Variation		1995	2002	Var 1995-2002	2010	Var 2002-2010
Type	Description					
1	Foods	2240	2097	-6.82%	1770	-18.47%
2	Pers. Use It.	2239	2213	-1.17%	1854	-19.36%
3	House. Art.	1314	1362	3.52%	1146	-18.85%
4	Health &Hyg.	563	569	1.05%	621	8.37%
5	Leisure It.	1441	1452	0.76%	1401	-3.64%
6	Repairs	2084	2137	2.48%	2152	0.70%
7	Other Items	593	665	10.83%	863	22.94%
Total (Retail Sector)		10474	10495	0.20%	9807	-7.02%
8	Restaurants	2101	2348	10.52%	2886	18.64%
9	Cafes	1563	1728	9.55%	2065	16.32%
10	Bars	103	109	5.50%	236	53.81%
Total (Restaurant Sector)		3767	4185	9.99%	5187	19.32%
Total		14241	14680	2.99%	14994	2.09%

Table 6 – Establishments inside Shopping Malls - Total Number, Percentage and Variation, 1995, 2002 and 2010

Year/Variation		1995		2002		Var.	2010		Var.
Type	Description	Number	% of Total	Number	% of Total	1995-2002	Number	% of Total	2002-2010
1	Foods	19	0.84%	69	3.19%	263.16%	98	5.25%	42.03%
2	Pers. Use It.	148	6.20%	465	17.36%	214.19%	712	27.75%	53.12%
3	House. Art.	42	3.10%	141	9.38%	235.71%	119	9.41%	-15.60%
4	Health &Hyg.	23	3.92%	73	11.37%	217.39%	119	16.08%	63.01%
5	Leisure It.	78	5.13%	205	12.37%	162.82%	215	13.30%	4.88%
6	Repairs	5	0.84%	18	2.64%	260.00%	81	8.58%	350.00%
7	Other Items	45	2.11%	169	7.33%	275.56%	234	9.81%	38.46%
Total (Retail Sec.)		360	3.32%	1140	9.80%	--	1578	13.86%	--

8	Restaurants	38	1.78%	167	6.64%	339.47%	316	9.87%	89.22%
9	Cafes	34	2.13%	105	5.73%	208.82%	147	6.65%	40.00%
10	Bars	0	0.00%	0	0.00%	--	0	0.00%	--
Total (Rest.Sector)		72	1.88%	272	6.10%	--	463	8.19%	--

Foods presents a very low percentage of activity inside malls. Since establishments are presented as points, part of this is explained by the fact that Foods inside malls (mainly supermarkets), usually the *anchors* of the mall (Carter & Vandell, 2005) are large in area, but small in number. Foods at street level (smaller supermarkets, and all other subtypes, like groceries, butchers, bakeries, etc.) thus suffered heavily from competition presented by shopping malls. Visually, the effect is as displayed on Figure 17.

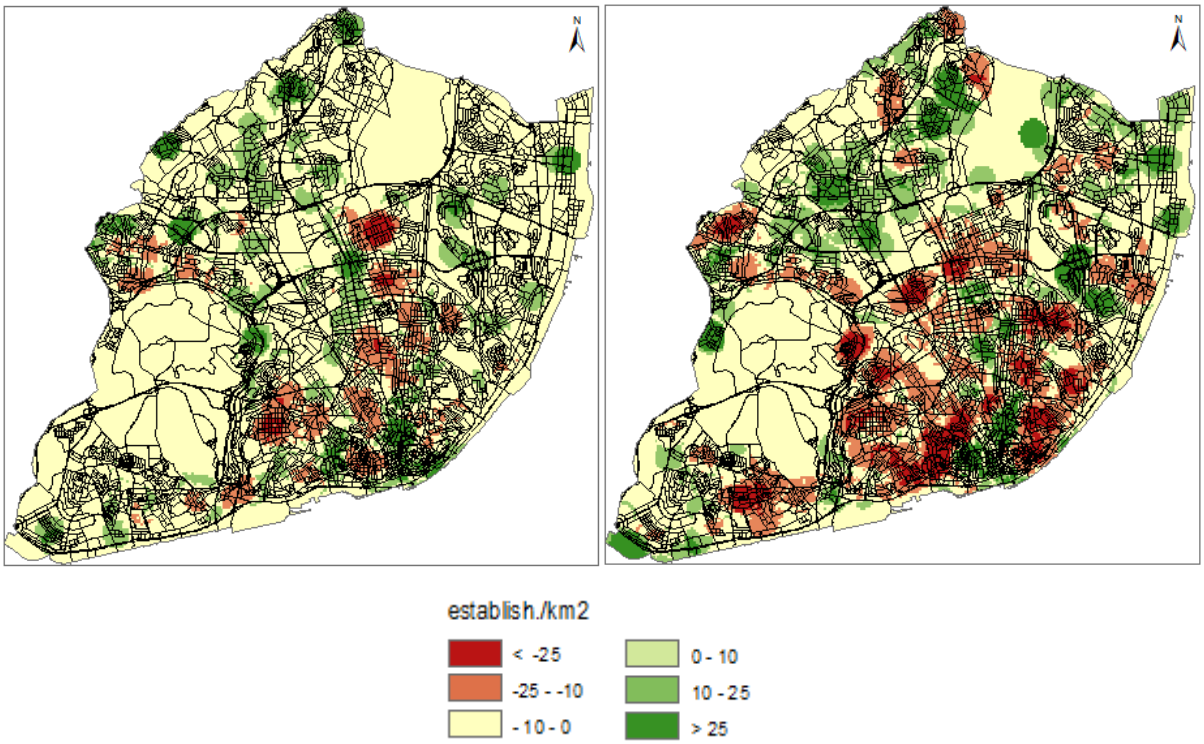


Figure 17 – Foods (as defined) – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right)

Vasco da Gama and Colombo are visible on the first period (the malls in Saldanha don't rely on supermarkets as anchors). Another phenomenon, though, becomes noticeable in the second period: older neighbourhoods of town are losing this type of activity while recent ones are gaining it (with Bairro Alto–Baixa–Martim Moniz appearing as the exceptions, south in green).

A preliminary conclusion about Foods can be made: even though, at street level, they suffer from competition from (larger) establishments located in shopping malls, they are needed in residential neighbourhoods, as is noticeable in the ones that are still growing. This would suggest a direct relation between number of establishments and number of residents, and therefore between Foods and Population, dependent of the location of Foods establishments, and not necessarily their sales area.

Foods are needed at street level, even if only to a certain number (eventually, the number that is needed to satisfy daily buys, for a certain population located at “reachable distance” – *threshold population*).

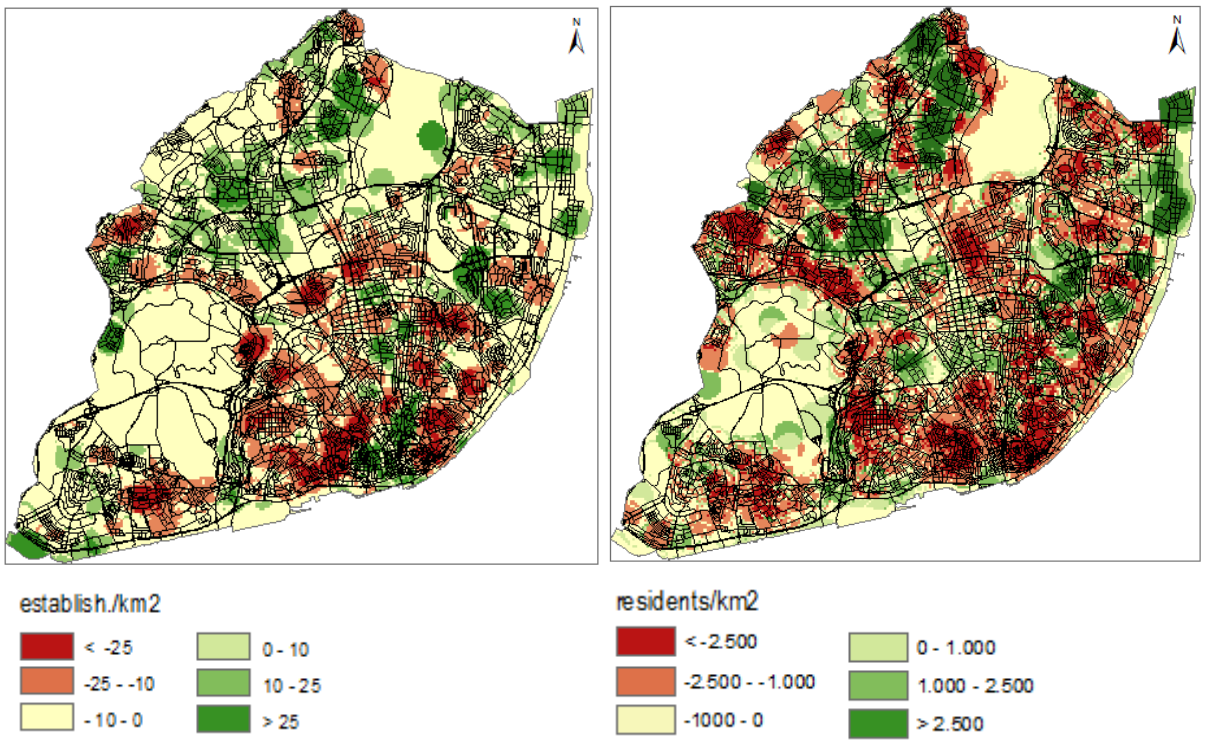


Figure 18 – Foods (left) and Population (right) – Density Variation: 2002 to 2010 (Foods) and 2001 to 2011 (Population)

Figure 18 shows the variation in Foods for the 2002-2010 period, and the variation in Population for the correspondent 2001-2011 period. The relationship is evident, and suggests a *ratio* between population and commercial establishments, compatible with the notion of *threshold population*. For food stores, Berry and Garrison (1958) set threshold population at 254. Figure 18 suggests that a similar relationship could be established in Lisbon. A correlation of 0.47 between these variables was found using CORREL (Microsoft, 2017), which is significant given the number of observations (more than 3000, both in 2002 and in 2010).

Household Articles presents an interesting case-study for occasional acquisition goods: in 1995 to 2002, the impact of malls is visible. But the period between 2002 and 2010 reveals a phenomenon that’s even more interesting: not only the areas that suffered from the competition from malls display persistent loss, but also the shopping malls themselves.

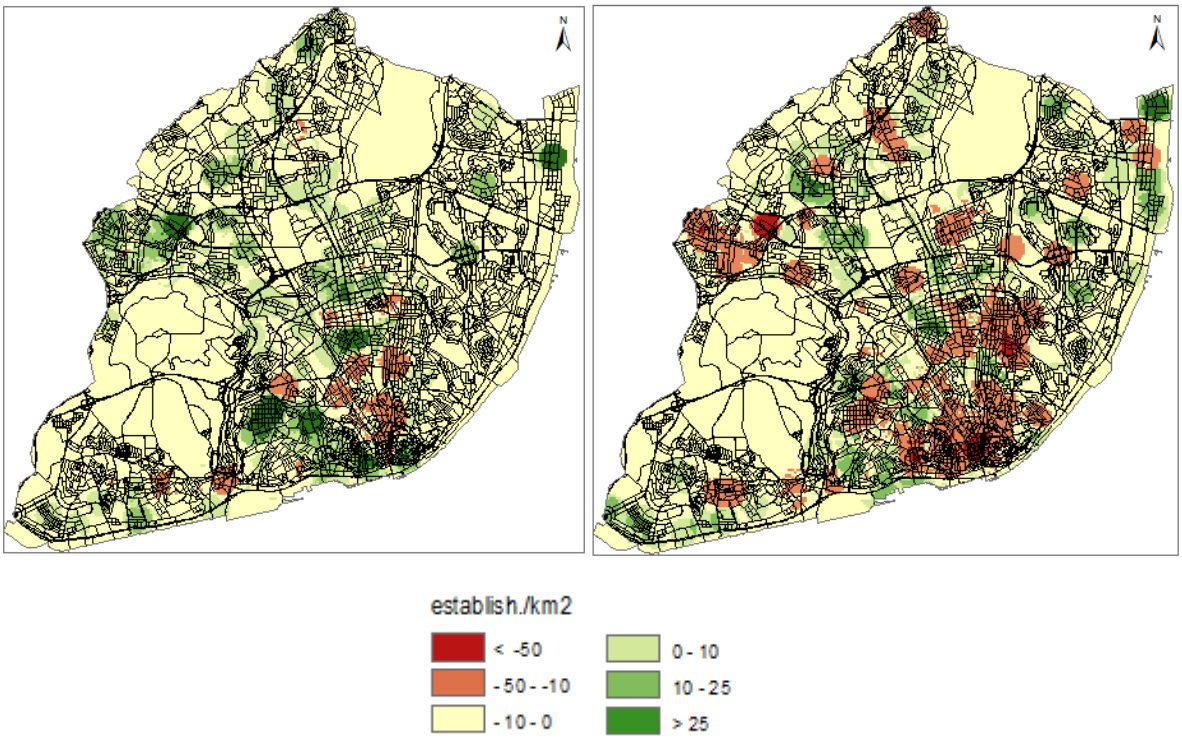


Figure 19 – Household Articles – Density Variation: 1995 to 2002, and 2002 to 2010 (left to right)

This can perhaps be attributed with the opening of several shopping malls *outside* Lisbon. This was already noticed by Salgueiro in 2001 (Salgueiro, 2001), with Cachinho (2006) attributing the success of modern shopping malls to the fact that they can "easily be replicated and built anywhere in the metropolis to meet the needs and desires of middle-class consumers". These replicas (Odivelas Parque, opened in 2003, Loures Shopping, opened in 2005, among several others in the outskirts of Lisbon) offered *competition to the competition*: what Colombo and Vasco da Gama had done to Baixa, dislocating commercial activity from the city centre, was now done to these shopping malls – and the city itself, with other locations around the metropolitan area presenting themselves as alternatives to satisfy consumption demands. Specifically, pertaining Household Articles, the opening of the first IKEA store in Portugal (inaugurated in 2004 in Alfragide) is certainly no stranger this effect.

This type of analysis (on change) was made to every type within the 10 defined, and was considered an excellent complement to the information that can be extracted from tables and literature: it locates change in space, and when used alongside pattern analysis (commercial centres – "hotspots"), it provides an array of information about commercial location patterns and its temporal change on Lisbon. An example of a frequent purchase good was presented, and another of an occasional purchase good: it wouldn't be reasonable to analyse all 10 types within the scope of this dissertation.

Foods are present within neighbourhoods, and although having lost establishments, remain relevant throughout the city. The nature of the goods sold (frequent acquisition) explains its distribution and resilience. As for goods of occasional purchase, they may be dislocated to malls, both inside and outside

the city, as centrality, which was once the basis of commercial activity, is succeed by conveniences offered by malls, such as parking and accessibility, “leading to the decentralization of commercial activity and de-hierarchization of the system of centres” (Cachinho & Salgueiro, 2016).

Frequent acquisition/lower threshold goods thus present a different behaviour from occasional purchase goods; and as the structure of commercial centres changes, one thing is revealed: smaller clusters of frequent acquisition goods seem resilient, based on Foods, and presenting a direct relation with (resident) Population (population density). In proceeding with the analysis, focus will therefore be made on Foods.

3.5. AN ANALYSIS ON MIX - COMMERCIAL DIVERSITY

To examine the diversity of commercial activity, all data pertaining commercial activity was related to a grid (“fishnet”) of 150m x 150m polygons, using Zonal Statistics as Table (ESRI, 2016c)). That grid had already been used to validate radii (and, as said, has polygons with an area of 22.500m2, which is similar to the average size area of INE’s statistical subsections).

For each grid cell, an index based on Shannon’s entropy was used (Song and Knaap, 2004).

$$\text{Entropy} = -\sum \frac{[Pj \times \ln(Pj)]}{\ln(J)} \qquad (2)$$

with *Pj* being the proportion of establishments of type *j*, and *J* the total number of establishment types, for each grid cell. This index varies between 1 and 0, with 1 representing a perfect balance between all J types found at cell level.

Entropy was calculated using the disaggregated data for the 10 establishment types. Results were as follows.

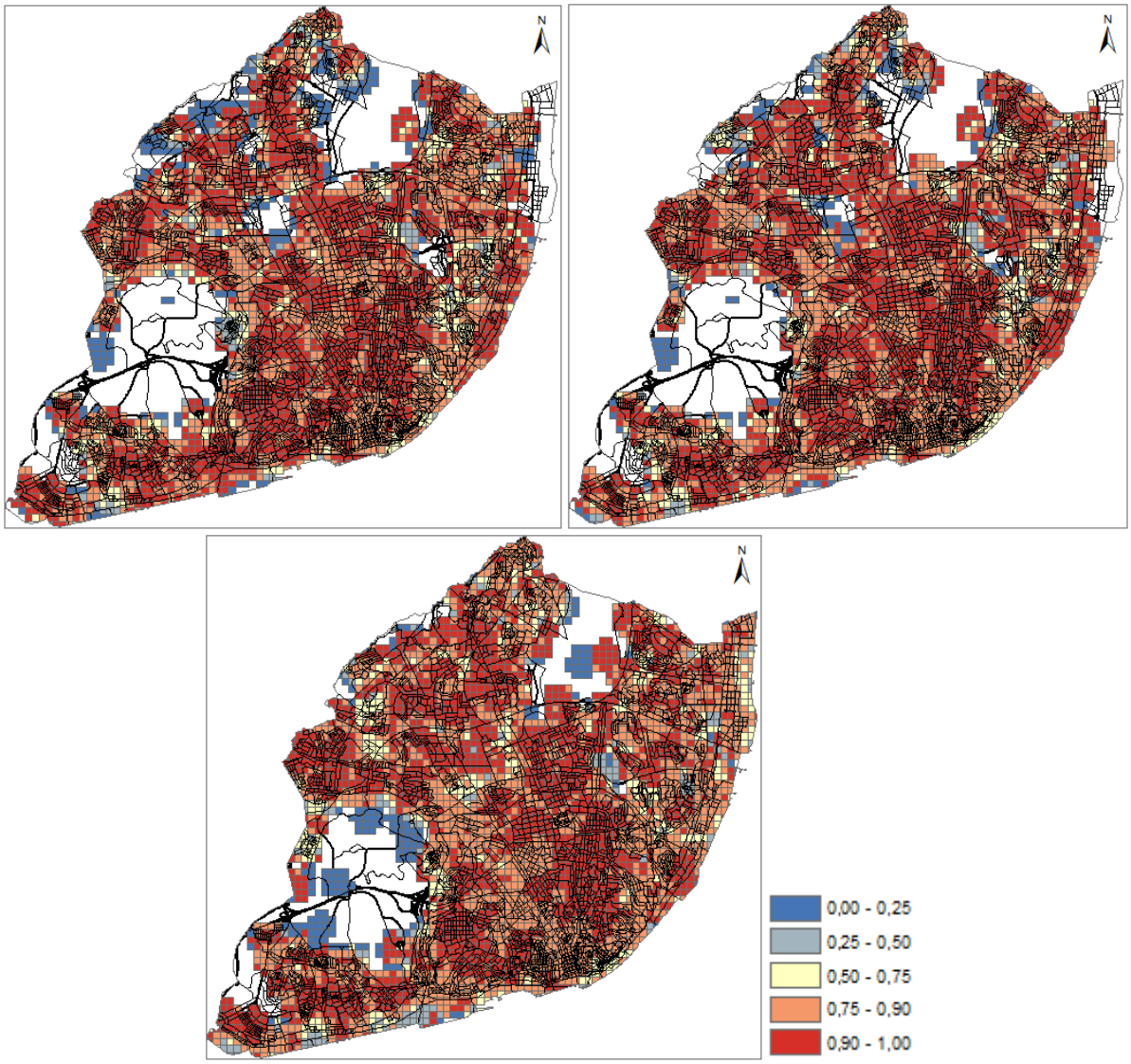


Figure 20 – Commercial Diversity: 1995, 2002 and 2010 (clockwise)

Commercial diversity, here measured by entropy, is high at cell-size level, with some small variations between 1995, 2002 and 2010. Smoothing of the data might lead into error: entropy was calculated using densities: what is being measured is the *mix of densities,* that should therefore be interpreted in the following way: for a 150mx150m cell, entropy is calculated considering the densities of types that occur within a 300m radius. Since cell-size is similar to block-size, and a 300m radius encompasses the adjacent cells (2nd level), high values of entropy thus indicate a good balance of commercial types at, roughly, the scale of a neighborhood.

Therefore, within Lisbon, diversity of establishment types is high at neighborhood-scale.

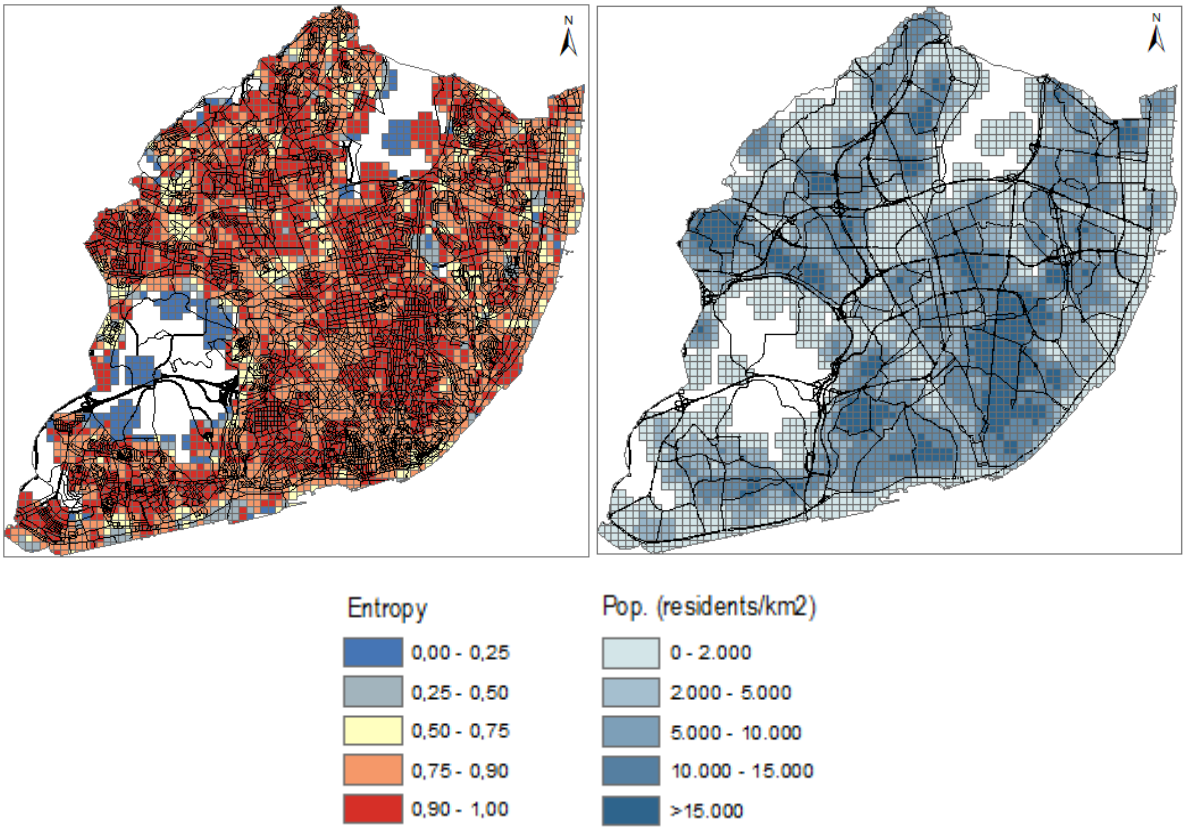


Figure 21 – Establishment Diversity (2010) and Population Density (2011) (left to right)

When comparing Entropy with Population Density (population, as was said, is available within INE's database), a relation between diversity and density also becomes apparent, and is too compatible with literature: activities will tend to mix in order to attract more customers. Either by way of multipurpose trips or information spillovers, or both (as have been discussed on Chapter 2) locating at a diverse commercial environment is interesting for retailers. An apparent relationship between commercial diversity and population density implies that it's not only interesting to retailers, but eventually, "demanded" by consumers, that will favour the convenience of having their shopping needs satisfied within their neighborhood.

3.6. EXPLORATORY ANALYSIS – SUMMARY

The aim of this chapter was to explore a dataset consisting of points, in terms of form and function, using point pattern evaluation techniques, density maps, and finally by calculating entropy, thus producing an overview on diversity.

These analyses revealed that within the city, commercial activity is distributed in a way that is compatible with proposed classifications (Gaspar 1985; Cachinho, 2011; Cachinho & Salgueiro, 2016), with a

general high order cluster of activity (Baixa), complemented by other commercial clusters (Av. Novas – Saldanha, Estrada de Benfica), eventually presenting some specialization (Bairro Alto, Almirante Reis). Smaller clusters of are present in a disperse way.

Activity has been partly transferred to shopping malls, that dispute importance with pre-existent “street level” clusters (Barata-Salgueiro, 2001). Frequent acquisition goods, though, remain relevant at street level. They’re not isolated, and while inside higher order clusters they seem to mix with all other types, a small “structure” seems to underlie at lower order clusters: a mix of Foods, Cafes, and Other types, directly related with Population (that eventually constitute the “centres of convenience or proximity” (Cachinho & Salgueiro, 2016)).

To follow these preliminary conclusions, an OLS-type regression is modelled in Chapter 4.

4. SPATIAL REGRESSION ANALYSIS

4.1. PREPARING A SPATIAL REGRESSION ANALYSIS

At the end of the previous chapter, a series of preliminary conclusions about location patterns of commercial establishments were attained. In order to further pursue them, an Ordinary Least Squares (OLS)-type regression analysis presented itself as a method that not only had been used with interesting results for similar subjects (Sevtsuk, 2010, 2014; Öner, 2014), but that would allow to further explore the relationships between the diverse types. In the quoted examples, regressions had used retail (as a whole) to predict for the presence of specific types. That is in line with retail location and agglomeration theory (retail attracts retail), but considering that the distribution of diverse types along town isn't homogenous, and that diverse centres/clusters present diverse mixes of activities, using different types to predict the presence of a specific one could would eventually allow for a better fit between these, and with other explanatory variables.

One problem when dealing with data with this level of disaggregation is an "intrinsic granularity" with variables presenting "extremely high variances, with a pre-dominance of zero values and hence extremely non-normal frequency distributions" (Thurstain-Goodwin & Unwin, 2000: 4).

By creating density surfaces for the usage of map algebra, that problem had been dealt with, and establishments densities could be used as variables.

Those densities had been transposed into a grid of a grid of 150m x 150m polygons, for the analysis on diversity. The choice of that grid (22.500m2 polygons, an area "similar to the average size area of INE's statistical subsections", as was said) was not without purpose: on the regression, variables collected during the censuses of 1991, 2001 and 2011 would be used. And since the limits of the statistical subsections had varied between these periods, this would allow using the same grid to relate INE's data with that of commercial establishments, for all different periods.

Polygon features that contained census data were converted to points, then into continuous surfaces, and finally transposed into the grid by using zonal statistics: a process similar to the one used for commercial establishments. In commercial establishments, density represents the number of point features found at a fixed distance (radius) around each cell, per unit area (km2). For population and housing variables, there are values associated with each point (e.g. population), which were used as *weight*. All data was related with the 150mx150m grid.

Finally, a 150mx150m grid containing all data related with establishments, population and housing variables was obtained in a *consistent* manner: the grid contains densities for each cell, using the same

radius and the same area unit (km2). The variables are, therefore, *comparable*, as they would be if the actual values were being used, but presenting a distribution that allows for a regression to be attempted.

As was stated in the beginning of the previous chapter, census variables pertain Population and Primarily Non-Residential Buildings (PNRB), to account for demand: the first is related with residents and the second can be related with employment/employees. Again, as said, there was no other available source to account for employment, and these two variables were available for all studied years.

Even before starting to work on the regression, and just by exploring the data associated with the grid, some interesting relationships may be expected.

Population (population *density*) was one of the available variables for all years. An analysis on that information when compared with data obtained at pattern analysis (optimized hotspot) for Foods shows the following.

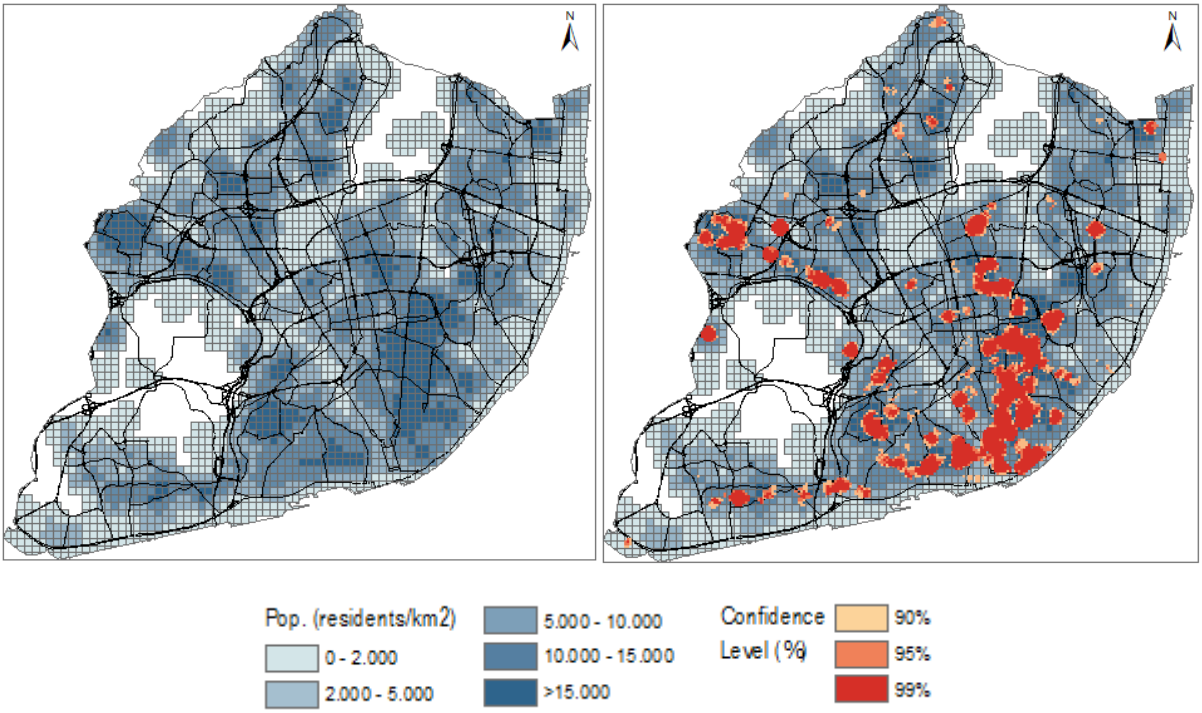


Figure 22 – Population (left), Population and Foods Hotspots (right), 2010-11

As Cadwallader (1996: 101) stated: “the two of the spatial distributions most closely associated with that of land values are the patterns of retailing and population density” (a reference also used by Öner (2014: 16). That can be applied to retail activity in general, but specifically to Foods: *Foods are present where Population is*. Because if another type is compared, the results aren’t quite the same. Using Leisure and Restaurants, results are as shown (Personal Use Items are so heavily clustered that using it to make a point would be biased).

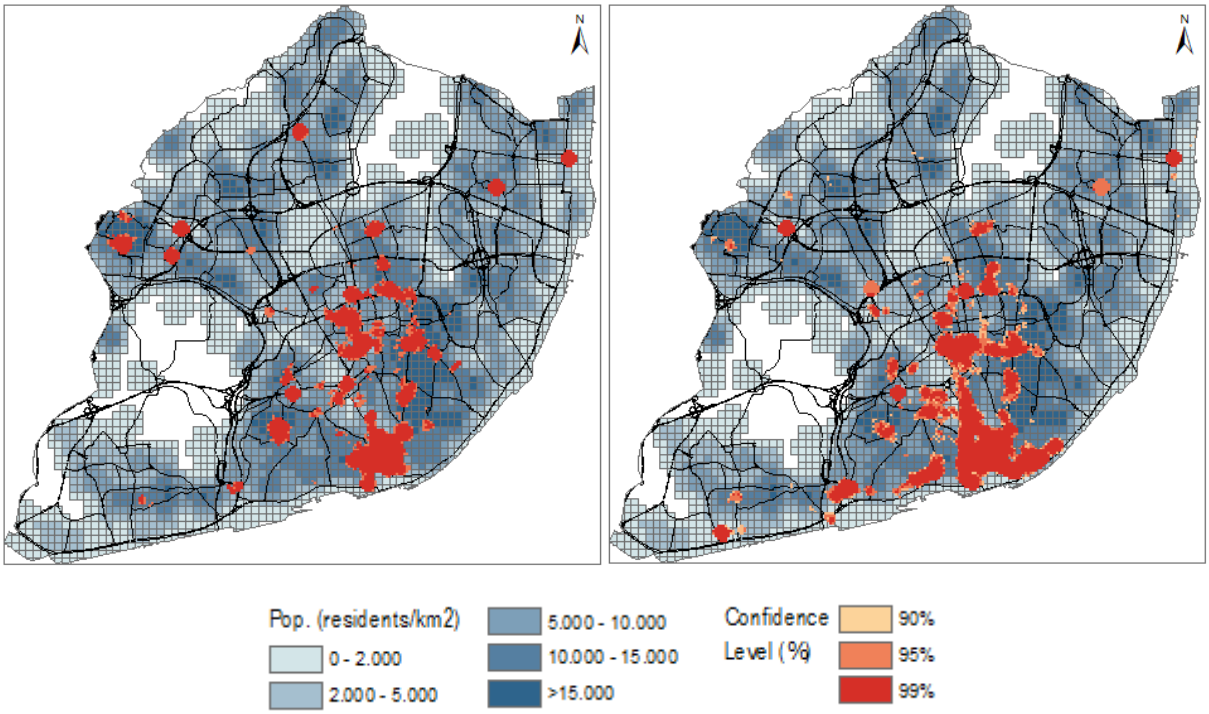


Figure 23 – Population and Leisure Hotspots (left), Population and Restaurants Hotspots (right), 2010-11

The relationship is very different that the one found between Population and Foods. A note must be made, which is that a regression will model relationships based on all values and all cells, not only high values and clustered cells, as are visible in Figures 22 and 23. They were just used to show what relationships *might* be expected, and to further help build the case for using diverse activities *separately* as variables. Finally, it's also interesting to present a similar example, but using Primarily Non-Residential Buildings.

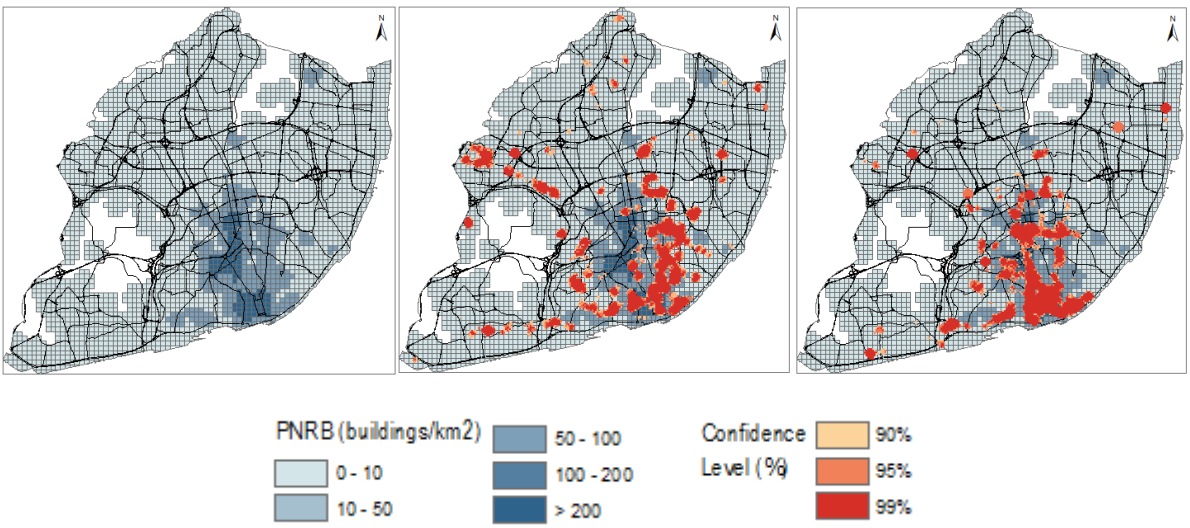


Figure 24 – PNRB (left) PNRB and Foods (middle) PNRB and Restaurants (right), 2010-11

The above figures also demonstrate the advantage of using densities / “smoothing the data” for both pattern detection and establishing relationships. The actual Population Density, calculated using INE’s statistical grid, is as shown below.

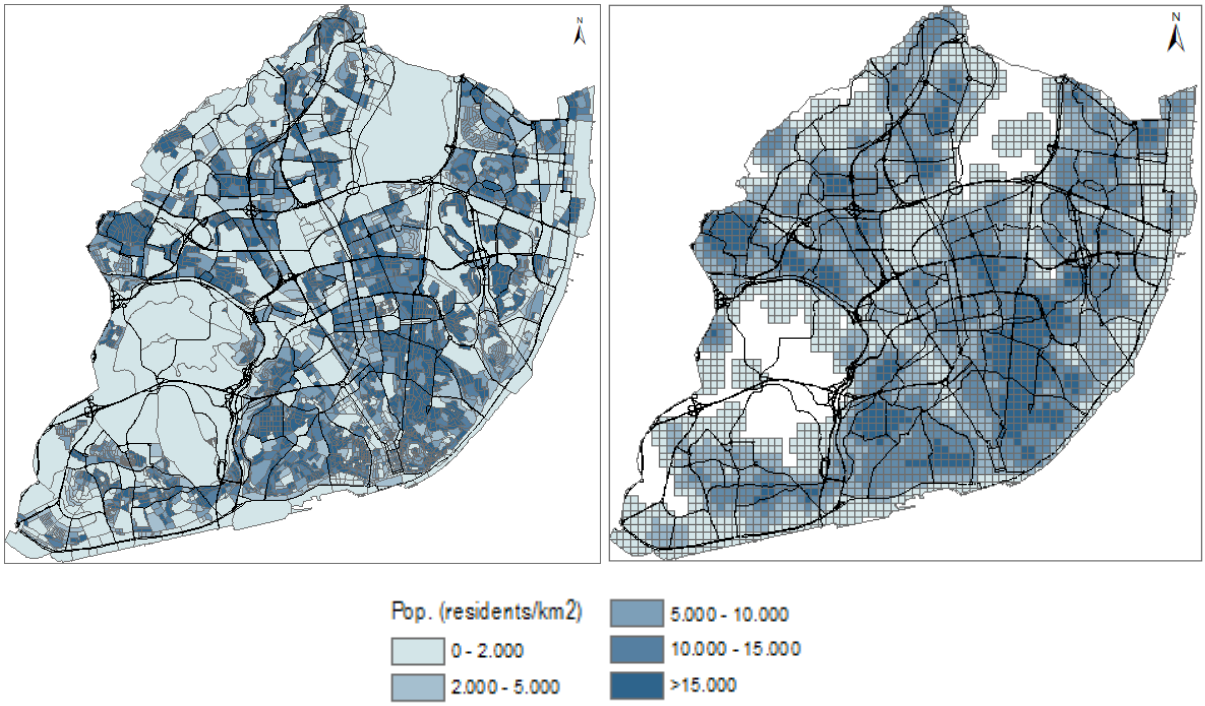


Figure 25 – Population Density, 2010-11: INE’s grid (left) and the 150mx150m grid used in the study

Both are calculated as residents by km2. The differences result from *zoning* and *scale* issues – coined within the “Modifiable Areal Unit Problem” (Openshaw, 1984). The problem of *scale* was acknowledged and contained by using a grid in which the polygons have the same average area as those of INE (different zoning, but same scale).

But scale and zoning problems eventually still arise when using a 300m radius. That’s why several radii were tested: to achieve the best fit between the real dataset and the one “smoothed” by density. The solution would be to use the actual dataset of points, but even cluster analysis wouldn’t have been possible: it implies a neighborhood of points (and that too implies creating a grid: an optimized grid, created by the Optimized Hot Spot Analysis tool (ESRI, 2016)). The size of that neighborhood also interferes with the apparent size of the cluster. And finally, it wouldn’t be possible to model a regression: the data presented an extremely non-normal distribution, as was mentioned.

Thurstain-Goodwin and Unwin (2000: 9) addressed this problem when modelling continuous surfaces:

“The major technical objection to area-based methods of urban analysis is essentially related to the uncertainty introduced by the well-known, if often badly understood, modifiable areal unit problem (Openshaw, 1984). It can be argued that any arbitrary areal units, such as the ED (Enumeration District)

of the census, must act as variable selective filters whose effects on how we `sample' the underlying real geography is almost impossible to determine".

But also stated (Thurstain-Goodwin and Unwin, 2000: 9) that:

"*It seems to us that surface representations of the type we have illustrated have two major advantages in urban analysis. First, they enable us to use very high spatial resolution data such as the UPC, or, for that matter, satellite derived land use in which the granularity of the urban fabric also becomes a major problem (see Donnay 1995 for a similar argument). Second, they seem to allow the development and calibration of models of urban dynamics in which space is represented continuously and hence in a potentially richer way than the discrete zone-based methods that are usually used.".*

MAUP was identified and assumed in the scope of this work. There is an attempt to control *scale* problems by using polygons that have the same area of the statistical sections to which commercial establishments point are related; and fit between data smoothed through the usage of a radius was also takes into consideration the necessity to control for *scale* and *zoning* issues. But when stating the MAUP, Openshaw also noted that "The areal units being studied should be meaningful in some way which is relevant to the purpose of the study" (Openshaw, 1984: 32). The purpose of the study, at this point, is to analyse commercial activity and relate it with housing and population variables. *Validation of scale and zoning lies within the apparent relationship found between Foods and Population, independently of using INE's grid or the 150mx150m grid with a 300m radius*. The problems that eventually might arise by altering zoning or scale are compensated by the attained possibility of modelling an OLS-regression.

4.2. MODELLING A SPATIAL REGRESSION ANALYSIS

The presence of retail, as said, has been used as a predictor for the presence of a specific type of retail, in research, and the analysis that is the scope of this work is in line with that hypothesis. But considering pattern distribution, it seems that modelling one type considering the whole of retail might not produce the most satisfying results, as was explained.

Regression analysis was therefore used to test if the location of *one type* of establishment could be related with the presence of other *specific* types. One establishment type was considered the dependent variable, with the remaining used as explanatory variables.

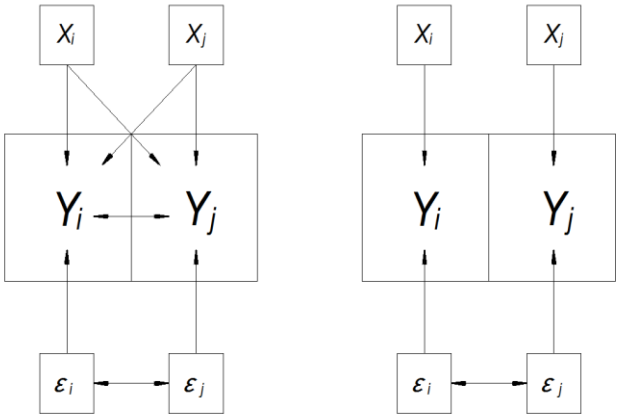
Population and Primarily Non-Residential Buildings (with the later posing as a possible source of information for job location, as stated previously) were retrieved from the census databases. These variables account for consumers, and therefore demand – with *demand* being the most important variable explaining the presence of retail (multipurpose shopping, competitive clustering, information

spillovers, all rely on *demand externalities* – that's why research relating retail with other kind of externalities, like space syntax, or accessibility indicators, is so relevant – because it's *rare*).

An additional variable, "Activity within Malls", was introduced. It was calculated by dividing the density of establishments located within malls by total establishment density. This variable was used since all retail was being analysed (both inside and outside malls) and a positive or negative relationship would help in determining what relationships were being considered in the model.

Once again: all variables represent densities or ratios of densities.

A classical OLS regression model was built, while testing for spatial dependence, due to the nature of the data. Before proceeding, a note on spatial lag and error, and the tests conducted.



Source: Adapted from Spatial Regression with GeoDa (n.d)

Figure 26 – Spatial Lag (left) and Spatial Error (right)

Spatial lag implies interdependence: "the dependent variable y in place i is affected by the independent variables in both place i and j" while spatial error reflects a problem of common exposure*:* "the error terms across different spatial units are correlated" (Spatial Regression with GeoDa (n.d)). In other words, lag indicates *interdependence between the considered variables*, while error signals for *omitted variables*. Both violate the assumptions of an OLS regression: lag violates the assumption of independent observations, while error violates the assumption of uncorrelated error terms (Spatial Regression with GeoDa (n.d)).

Diagnostics for spatial dependence and error include using the standard Lagrange Multiplier (LM) and the Robust LM. Since standard LM assumes that the other process (either lag or error) isn't present (which poses an obvious problem) analysis is focused on the Robust LM test (Franzese & Hays, 2014).

An OLS-type regression is performed by implementing the following function:

$$y = X\beta + \varepsilon \quad (3)$$

where y represents the dependent variable, X the explanatory variables with β estimated coefficients, and ε the term of error

If lag or error are detected in the diagnostics, a spatial-autoregressive lag model (SAR) or a spatial-autoregressive-error model (SEM) will have to be used, by running the following implementations of function 3:

$$y = \rho\, W y + X\beta + \varepsilon \quad (4)$$

where W is a spatial weights matrix applied to the variable y, and ρ is the spatial autoregressive coefficient (Anselin, 2001), is used in SAR, and

$$y = X\beta + \varepsilon \text{ with } \varepsilon = \lambda\, W \varepsilon + \mu \quad (5)$$

where W is a spatial weights matrix applied to the error term ε, λ is the spatial autoregressive coefficient and μ is the normally distributed error, is used in SEM.

Measures of fit of these models will be Log-Likelihood, Akaike info criterion (AIC) and Schwarz criterion (SC): "*While it is tempting to focus on traditional measures, such as the R2, this is not appropriate in a spatial regression model. The value listed in the spatial lag output is not a real R2, but a so-called pseudo-R2, which is not directly comparable with the measure given for OLS results. The proper measures of fit are the Log-Likelihood, AIC and SC.*" (Anselin, 2005: 207).

Regression started with Foods, 1995, with the results being presented in Table 7.

Table 7 – Results for OLS Regression, Foods 1991-95

Variable	OLS		
	coefficient	t-statistic	p-value
Constant	-2.2690	-6.633	0.000
Personal Use Items	-0.0185	-1.879	0.060
Household Articles	0.1438	8.501	0.000
Health and Hygiene	0.0786	1.702	0.088
Leisure	-0.1212	-5.686	0.000
Other Items	0.0060	0.391	0.696
Repairs	0.2196	7.818	0.000
Restaurants (and similar)	0.1762	13.962	0.000
Cafes (and similar)	0.5815	23.642	0.000
Bars (and similar)	0.1057	1.619	0.106
Population	0.0018	35.450	0.000

Primarily Non-Residential Building	-0.1275	-11.636	0.000
% Activity within Malls	-0.4900	-0.104	0.917
R-squared	0.882		
Log-likelihood	-13379.4*		
Akaike info criterion	26784.7*		
Schwarz criterion	26864.5*		

*The numbers represent values.

Table 8 – Diagnostics for Spatial Dependence, OLS Regression, Foods 1991-95

Test	Value	Probability
Lagrange Multiplier (lag)	3603.404	0.000
Robust LM (lag)	16.006	0.000
Lagrange Multiplier (error)	6148.095	0.000
Robust LM (error)	2560.697	0.000
Lagrange Multiplier (SARMA)	6164.101	0.000

Tests for spatial dependence (Table 8) indicate both the presence of spatial lag and spatial error.

"Spatial Regression with GeoDa" (n.d) discusses a similar case, offering the following interpretation: "*comparing the spatial lag and spatial error models, we can see both alternative models yield improvement to the original OLS model. Therefore, we should conclude that controlling spatial dependence will improve our model performance. Now the question is which of the two models is better? To some extent, this is an open question. The general advice is first to look for a theoretical basis to inform your choice. If there are strong substantive grounds for one model instead of the other, you should adopt it. When it is not so clear theoretically, you can compare the model performance parameters: the R-squared and Log likelihood"*.

Therefore, both a Lag Model and an Error model were tested, providing for the following results.

Table 9 – Spatial Lag and Spatial Error for Foods, 1991-95

Variable	Spatial Lag			Spatial Error		
	coefficient	t-statistic	p-value	coefficient	z- statistic	p-value
W_Foods	0.8269	103.000	0.000	--	--	--
lambda	--	--	--	0.9802	328.015	0.000
Constant	-2.837	-13.564	0.000	-2.3695	-0.533	0.593
Personal Use Items	0.0085	1.435	0.151	-0.0360	-3.336	0.001
Household Articles	-0.0026	-0.249	0.803	0.1687	8.491	0.000
Health and Hygiene	0.0712	2.554	0.011	0.0349	0.834	0.404
Leisure	-0.0469	-3.642	0.000	0.0264	1.197	0.231
Other Items	-0.0140	-1.510	0.131	0.0313	1.832	0.067

Repairs	0.0490	2.866	0.004	0.4131	0.034	0.000
Restaurants (and similar)	0.0584	7.574	0.000	0.1637	11.046	0.000
Cafes (and similar)	0.1902	12.130	0.000	0.5585	22.726	0.000
Bars (and similar)	-0.0169	-0.429	0.668	-0.0423	-0.649	0.516
Population	0.0004	13.318	0.000	0.0013	20.463	0.000
Primarily Non-Residential Building	-0.0661	-9.928	0.000	-0.0554	-3.969	0.000
% Activity within Malls	-2.3101	-0.815	0.415	-9.8554	-2.257	0.024
Log-likelihood	-11904.1*			-10892.379*		
Akaike info criterion	23836.3*			21810.8*		
Schwarz criterion	23992.2*			21890.5*		

*The numbers represent values.

The spatial error model performs better as shown by Log-Likelihood, AIC and SC values. And, when testing for both LM and Robust LM, the values of the spatial error model were significantly higher. This suggests that, in the model, it'll be of greater importance to control for error than for lag (containing the bias introduced by missing variables will be more important than focusing on the one introduced by interaction amongst observations). It's likely that many variables that would explain the spatial location aren't being taken in consideration. Only demand is being considered, and even that, only by means of the available data on population and "job location".

Analysis is therefore focused on spatial error, considering the model results and fit indicators presented on Table 9.

Table 10 – Spatial Error for Foods, 1991-95

Variable	Spatial Error		
	coefficient	z- statistic	p-value
lambda	0.9802	328.015	0.000
Constant	-2.3695	-0.5331	0.593
Personal Use Items	-0.0360	-3.336	0.001
Household Articles	0.1687	8.491	0.000
Health and Hygiene	0.0349	0.834	0.404
Leisure	0.0264	1.197	0.231
Other Items	0.0313	1.832	0.067
Repairs	0.4131	0.034	0.000
Restaurants (and similar)	0.1637	11.046	0.000
Cafes (and similar)	0.5585	22.726	0.000
Bars (and similar)	-0.0423	-0.649	0.516
Population	0.0013	20.463	0.000
Primarily Non-Residential Building	-0.0554	-3.969	0.000

% Activity within Malls	-9.8554	-2.257	0.024
Log-likelihood	-10892.379*		
Akaike info criterion	21810.8		
Schwarz criterion	21890.5		

*The numbers represent values.

The model presents a variable (lambda) to control for spatial error. It's positive and significant; its coefficient of 0.9802 reflects the importance of missing information (variables) in the model. Analysis is done with that in mind, and spatial error is controlled by lambda: that means the relationships established amongst variables are *valid*. Therefore, variables relate; but *why they relate* can't be explained within the model (only by means of demand), since many of the variables that could eventually account for individual location choices (other *externalities* apart from demand) are missing.

Household articles, Repairs, Restaurants (and similar) and Cafes (and similar) present a positive and significant relation ($p<0.001$) with Foods. A negative and significant relation is found with Personal Use items.

Foods is significantly related with Population, even though with a small coefficient (which might be related with *magnitude* of the variables: values for Population are always much higher than those of other variables) and relationship with Primarily Non-Residential Building is significant and negative; both are consistent with preliminary conclusion about Foods and (threshold) Population.

After testing for spatial dependence and comparing model fit, the spatial error model was the used for 2010-11.

Table 11 – Diagnostics for Spatial Dependence, OLS Regression, Foods 2010-11

Test	Value	Probability
Lagrange Multiplier (lag)	4792.735	0.000
Robust LM (lag)	8.940	0.002
Lagrange Multiplier (error)	6834.650	0.000
Robust LM (error)	2050.855	0.000
Lagrange Multiplier (SARMA)	6843.591	0.000

Table 12 – Model Fit Statistics, Foods 2010-11

Values	Model		
	OLS	Lag	Error
Log-likelihood	-13799.5	-11916.7	-10973.707
Akaike info criterion	27624.9	23861.4	21973.4
Schwarz criterion	27705.5	23948.2	22054

Spatial dependence and model fit confirmed that using spatial error in 2010-11 would be appropriate, and analysis focused on that model, with the results being compared to the ones found in 1991-95.

Table 13 – Spatial Error for Foods, 1991-95 and 2010-11

Variable	1991-95			2010-11		
	coefficient	z- statistic	p-value	coefficient	z- statistic	p-value
lambda	0.9802	328.015	0.000	0.9828	353.96	0.000
Constant	-2.3695	-0.533	0.593	-3.9396	-0.948	0.343
Personal Use Items	-0.0360	-3.336	0.001	-0.0494	-7.173	0.000
Household Articles	0.1687	8.491	0.000	0.0908	5.361	0.000
Health and Hygiene	0.0349	0.834	0.404	-0.0446	-1.419	0.156
Leisure	0.0264	1.197	0.231	0.0846	4.427	0.000
Other Items	0.0313	1.832	0.067	0.1100	9.0842	0.000
Repairs	0.4131	0.034	0.000	0.2355	10.628	0.000
Restaurants (and similar)	0.1637	11.046	0.000	0.0952	9.108	0.000
Cafes (and similar)	0.5585	22.726	0.000	0.3613	19.247	0.000
Bars (and similar)	-0.0423	-0.649	0.516	0.0757	2.475	0.053
Population	0.0013	20.463	0.000	0.0010	15.627	0.000
Primarily Non-Residential Building	-0.0554	-3.969	0.000	-0.0609	-4.030	0.000
% Activity within Malls	-9.8554	-2.257	0.024	-1.8391	-1.449	0.147
Log-likelihood	-10892.379*			-10973.707*		
Akaike info criterion	21810.8*			21973.4*		
Schwarz criterion	21890.5*			22054*		

*The numbers represent values.

Foods maintained a positive and significant relation ($p<0.001$) with Restaurants (and similar), Cafes (and similar) Repairs and Household Articles. Leisure also became relevant. As for Personal Use items, one might risk concluding it was a clustered activity at street level in 1995, and has since become clustered at street level and inside shopping malls, maintaining a negative relationship with Foods.

Relationship with Other Items is especially interesting when considering additional information about the data: this type was created to include all types that couldn't be included in the other 9 categories. It includes, for example, "Bazares", which in 1995 were just 69 (about 3% of the total number of establishments) and in 2010 were already 311 (13% of the total number of establishments). "Bazares" are stores selling miscellaneous goods, from clothes to small household articles, usually managed by immigrants, with a large community settled in Intendente-Martim Moniz-Mouraria. That area, as was noted, amount for a large part of the positive variation on retail that can be observed east of Baixa in Figure 15, in the 2002-2010 period. It's a street level activity (only 10 out 311 are located inside malls) that almost didn't exist in 1995, but since then, was able to present itself as relevant.

As for relationship with Primarily Non-Residential Buildings, it's significant and negative, in both periods, which shows that Foods are frequently purchased goods by residents, not workers. That is compatible with the nature of the goods being sold.

To validate results, analysis proceeded with other types that showed significant relationship with Foods. Spatial dependence was tested, and so was model fit, with spatial error models presenting the best results for each individual type.

Table 14 – Spatial Error model for Restaurants, Household Articles and Leisure 2010-11

Variable	Restaurants		Household Articles		Leisure	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
lambda	0.9922	0.000	0.9927	0.000	0.9852	0.000
Constant	-1.425	0.920	3.201	0.734	1.4296	0.73295
Foods	0.2364	0.000	0.0850	0.000	0.0637	0.000
Personal Use Items	0.0072	0.509	0.0780	0.000	0.1079	0.000
Household Articles	-0.0348	0.192	--	--	0.0447	0.002
Health and Hygiene	0.4111	0.000	0.1911	0.000	0.1751	0.000
Leisure	0.3630	0.000	0.0563	0.003	--	--
Other Items	0.1728	0.000	0.0961	0.000	0.0462	0.000
Repairs	-0.0449	0.202	0.0760	0.001	0.0466	0.0165
Restaurants (and similar)	--	--	-0.0134	0.194	0.1123	0.000
Cafes (and similar)	0.3276	0.000	0.0947	0.000	0.1929	0.000
Bars (and similar)	1.0562	0.000	0.0934	0.002	-0.0495	0.062
Population	-0.0003	0.002	-0.0001	0.833	-0.0002	0.002
Primarily Non-Residential Buildings	0.1242	0.000	-0.0213	0.150	0.0562	0.0000
% Activity within Malls	9.8066	0.000	-3.5302	0.004	3.5627	0.0011
Log-likelihood	-12603.202		-10873.100		-10443.500	
Akaike info criterion	25232.4		21772.2		20913	
Schwarz criterion	25313		21852.8		20993.6	

Starting with the Restaurants model, it presents a positive and significant relationship with Health & Hygiene, and Bars, while the relationship with Household Articles and Repairs is negative (but not

significant). More important, relationship with Population is negative and significant, and positive and significant with Primarily Non-Residential Buildings.

This would imply Restaurants locate themselves in areas that are different from the ones where Foods are, but with a distribution that allows for that relationship to be significant when Foods are the dependent variable. A very expressive location of Restaurants inside higher order clusters, as was shown in previous analyses (Baixa, but also Colombo and Vasco da Gama), but also the "specialist cluster" of Bairro Alto might explain this.

Household articles presents as a positive and significant relationship with every type except for Restaurants. Relationships with Population and Primarily Non-Residential Buildings are not significant. This lack of relationship with both residents and workers (for which Non-Residential Buildings are considered as a proxy) suggests an activity that is relevant enough to attract customers, not having to locate immediately in their proximity. That is compatible with what has been stated about Household Articles: located mostly in the commercial ribbon/strip of Almirante Reis, and thus accessible and exposed to costumers, it might relate to all other types either by spillover effects, or considering that the same *external factor* – accessibility/ "exposure" – benefits all types of activities.

At this point in time, though, one might risk saying that the presence of commercial activity in Almirante Reis qualifies as *historical* (like Baixa): it was originally based on accessibility / "visibility" and rent (Gaspar, 1985) but has since become an *historic accident*: commercial establishments can currently locate on Almirante Reis due to (or especially because of) the added value of multipurpose shopping and information spillover. This conclusion is made based on analyses of data and literature up to this point, and open for discussion.

Leisure presents a significant and positive relationship with every type except Repairs and Bars. Leisure has a negative with Population and positive with Primarily Non-Residential Buildings. Since the % Activity within Malls is positive and significant, it's could indicate that a part of this activity occurs at street level (explaining for the relationship with Foods) and partly at Malls, (relating with Health and Hygiene and Personal Use Items; these three types have the biggest share of activity occurring inside malls, as was shown in Table 6: 13.30% for leisure, 16.08% for Health and Hygiene and 27.75% for Personal Use Items; no other type has more than 10%).

Table 15 – Spatial Error for Other Items, Repairs and Cafes 2010-11

Variable	Other Items		Repairs		Cafes	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
lambda	0.9892	0.000	0.9891	0.000	0.9772	0.000
Constant	1.9821	0.823	1.0711	0.823	2.7844	0.295
Foods	0.2002	0.000	0.1274	0.000	0.2552	0.000

Personal Use Items	0.1800	0.000	-0.0405	0.000	0.0157	0.007
Household Articles	0.1857	0.000	0.0442	0.000	0.0708	0.000
Health and Hygiene	0.1455	0.006	0.0802	0.001	0.1305	0.000
Leisure	0.1150	0.000	0.0341	0.016	0.1833	0.000
Other Items	--	--	0.0698	0.000	0.0413	0.000
Repairs	0.2358	0.000	--	--	0.0923	0.000
Restaurants (and similar)	0.1294	0.000	-0.0100	0.196	0.0944	0.000
Cafes (and similar)	0.1038	0.000	0.0697	0.000	--	0.000
Bars (and similar)	-0.2922	0.000	-0.0591	0.009	0.1468	0.000
Population	0.0001	0.149	0.0003	0.000	0.0008	0.000
Primarily Non-Residential Buildings	0.0630	0.002	-0.0066	0.554	0.0363	0.404
% Activity within Malls	-7.642	0.000	4.2106	0.000	-3.9330	0.000
Log-likelihood	-12073.921		-9857.3331		-10341.682	
Akaike info criterion	24173.8		19740.7		20709.4	
Schwarz criterion	24254.4		19821.3		20790	

Other Items presents a positive relationship with Primarily Non-Residential Buildings and negative and significant with % Activity within Malls. It's a street level activity, as expected, with the relationship with Primarily Non-Residential Buildings being consistent with a cluster in Baixa – Martim-Moniz (as can be seen in Figure 42 - Density of Primarily Non-Residential Buildings).

This could also explain for positive and significant relationship with almost every other type of activity. But the coefficients for Foods and Repairs (superior to 0.2000) imply a particular relationship with those kinds of activities, and the "smaller clusters" disperse alongside the city.

A negative relationship with Bars locates Other items mainly outside Bairro Alto, where the former is clustered, as has been shown previously, but also stated by Salgueiro (2001).

Repairs presents results that are, to some degree, compatible with the ones found in Foods: negative and significant relationship with Personal Use Items, positive and significant with Household Articles, Cafes and Other Items and Leisure. Negative relationship with Bars shows (location preferences mainly excluding Bairro Alto). The significant and positive relationship with Population, along with its specific needs (large stores) suggests that the Repair establishments that are not located inside shopping malls

locate in small commercial clusters at neighborhood level, alongside Foods (like Household Articles, Repairs needs larger areas than most shops; but since it doesn't need "exposure", an optimal location might be found outside higher order clusters, and inside residential neighbourhoods). The positive relationship with % Activity within Malls might be related with the positive relationship with Health & Hygiene.

As for Cafes, a positive and significant relationship with almost every other type effectively shows effective dispersion alongside the city. This is similar to Other Items, but with a relevant difference: it has a significant and positive relationship with Population. It relies on residents more than Other Items, and similarly to Foods.

Analysis on a "small cluster" based on Foods would end here: it'd be related with Population, Cafes and Repairs, and Other Items (when located outside Baixa - Martim Moniz). Relation with Household Articles would be due to the influence of Almirante Reis (where both activities occur, in a pattern significant enough to be signalled by the regression models), the same happening with Restaurants due to Baixa and Bairro Alto (Almirante Reis would be an outlier pertaining Household Articles, and Bairro Alto pertaining Restaurants; Baixa is an outlier, so significant, that it introduces a bias in any attempts at modelling a regression).

But having arrived at this point, an analysis on Personal Use Items, Health & Hygiene and Bars is due, method wise, even though one can expect the first to exhibit significant relationship with all other types and with % Activity Within Malls, as will the second (because of its expressive location at the highest order cluster of Baixa, and within shopping malls); and Bars to relate with Restaurants.

Variable	Personal Use Items		Health and Hygiene		Bars	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
lambda	0.98712	0.000	0.9768	0.000	0.9902	0.000
Constant	-8.7658	0.000	0.2788	0.865	0.3861	0.921
Foods	-0.2838	0.000	-0.0127	0.144	0.0222	0.013
Personal Use Items	--		0.0841	0.000	-0.0097	0.009
Household Articles	0.4799	0.000	0.0566	0.000	0.0292	0.022
Health and Hygiene	1.7265	0.000	--	--	-0.1671	0.000
Leisure	0.8225	0.000	0.0655	0.000	-0.0204	0.048
Other Items	0.5614	0.000	0.0228	0.000	-0.0465	0.000
Repairs	-0.4285	0.000	0.0416	0.000	-0.0318	0.008
Restaurants (and similar)	0.01375	0.5864	0.0471	0.000	0.1265	0.000

Cafes (and similar)	0.13082	0.0053	0.0513	0.000	0.0602	0.000
Bars (and similar)	-0.1922	0.0085	-0.1637	0.000	--	--
Population	-0.00126	0.000	-0.0000	0.059	-0.0001	0.000
Primarily Non-Residential Buildings	0.2111	0.000	0.0370	0.000	0.0135	0.010
% Activity within Malls	34.770	0.000	2.4762	0.000	-0.4169	0.542
Log-likelihood	-14143.298958		-8643.942		-8734.450	
Akaike info criterion	28312.6		17313.9		17494.9	
Schwarz criterion	28393.2		17394.5		17575.5	

Personal Use Items is more interesting than expected, “regression wise”. The negative relationship with Foods and Population does locate it at a high-level cluster. Lack of significance in the relationship with Cafes is also compatible with that. But the lack of significant relationship with Restaurants locates it outside Baixa: it’s a highly clustered activity, to which conclusions are similar to the ones made for Household Articles: the cluster at Alvalade – Roma – Av. Novas has created an effect, either by means of spillover or an externality that is interesting for all commercial activities, that at this point is historic: commercial establishments locate due to pre-existing commercial activity location patterns. No matter what the original externality was (an accessible alternative to Baixa, by car and by means of subway expansion, as described by Gaspar, 1985), it now relies on demand externalities: surely competitive clustering but, interestingly enough, multipurpose shopping.

Health and Hygiene presents the expected results, similar to the ones of Leisure, but presenting no relationship with Foods. Conclusions about both activities are therefore also similar: Health and Hygiene locates inside shopping malls or at high order clusters, but contrary to Leisure, it’s contained within those, presenting no relationship with Foods.

Bars present results compatible with those of a highly clustered activity, which it is, located at Bairro Alto alongside with Restaurants and Cafes.

4.3. DISCUSSION/RESULTS

Interpretation of the regression results, based on literature and previous results, indicates the following:

1) Foods, Cafes and Repairs have a positive and significant relation with Population and with each other. Other Items also presented similar results, eventually relating with those activities outside Baixa-Martim-Moniz, where an outlier creates a bias in the results.

Foods, Cafes and Other Items would therefore compose the small centres of "convenience and proximity" (Cachinho and Salgueiro (2016)). The relationship between these types and Population further backs up the idea of an underlying "population threshold" for this mix of activities.

The inclusion of Repairs is not related with the nature of the good, but with that of bid-rent: with a need for larger areas than most shops, but no special need for visibility/ "exposure", it finds an optimal location outside the higher order clusters, and inside residential neighbourhoods, where it relates with these small "centres of convenience and proximity".

2) Restaurants relate with the previously described "small centres", but also with higher order ones: eventually at Baixa, at shopping malls and at Bairro Alto "specialist cluster" (where their presence is significant, as was duly pointed when analysing change). This might be explained by different restaurants having different *ranges*: tourists, workers, residents. When modelling a regression, Population can only explain the location of the "lower range" types; a classification on range ("target clients") would be needed: the ones that cater for employees have higher ranges, and the ones that cater for tourists (or, in general, for "occasional purchases") have the highest range; only the lower range type could enter a regression where "convenience and proximity" goods are being explained by Population, mostly (because they would have *the same threshold* as those);
3) Household Articles is an interesting activity: relating with almost every other type, but lacking relevant relationships with both Population and Primarily Non-Residential Buildings (that is, residents or workers), it suggests it's an activity that is relevant enough to attract customers, not having to locate immediately in their proximity. Several things must be said about that: first, it's a high order activity: it relies on demand coming from all parts of city and also from outside of its borders. That's why they were originally located at the street leading to the higher order cluster (Gaspar, 1985). It also demands for more store area, so bid-rent would also explain its location outside Baixa. At this point in time, though, the presence of commercial activity in Almirante Reis qualifies as an *historic accident*: commercial activity originally located based on accessibility and rent, but now locates based on the presence of other commercial activity;
4) Personal Use Items is another interesting case: it's a highly clustered activity, on which conclusions are similar to the ones made for Household articles: the cluster at Alvalade – Roma – Av. Novas has created a spillover effect, either by means of demand externalities or other externalities, that is interesting for all commercial activities. No matter what the original externality was (it might be because it was an accessible alternative to Baixa, by car and by means of the subway expansion (Gaspar, 1985), it now relies on demand externalities: competitive clustering and, interestingly enough, multipurpose shopping. Like in Almirante Reis, at this point in time, commercial activity locates because of commercial activity;
5) Considering the introduction of historic accidents, the same point can be made for Baixa and Campo de Ourique: the first being the centre of commercial activity for hundreds of years, the second since 1755's earthquake (Gaspar, 1985): whatever made them attractive originally (centrality and proximity to the river in Baixa; being an alternative to devastated Baixa after the earthquake, in Campo de Ourique), has long lost importance: commercial activity locates due to the presence of other

commercial activities. Models will have to have to take that in consideration, whatever variables are considered.

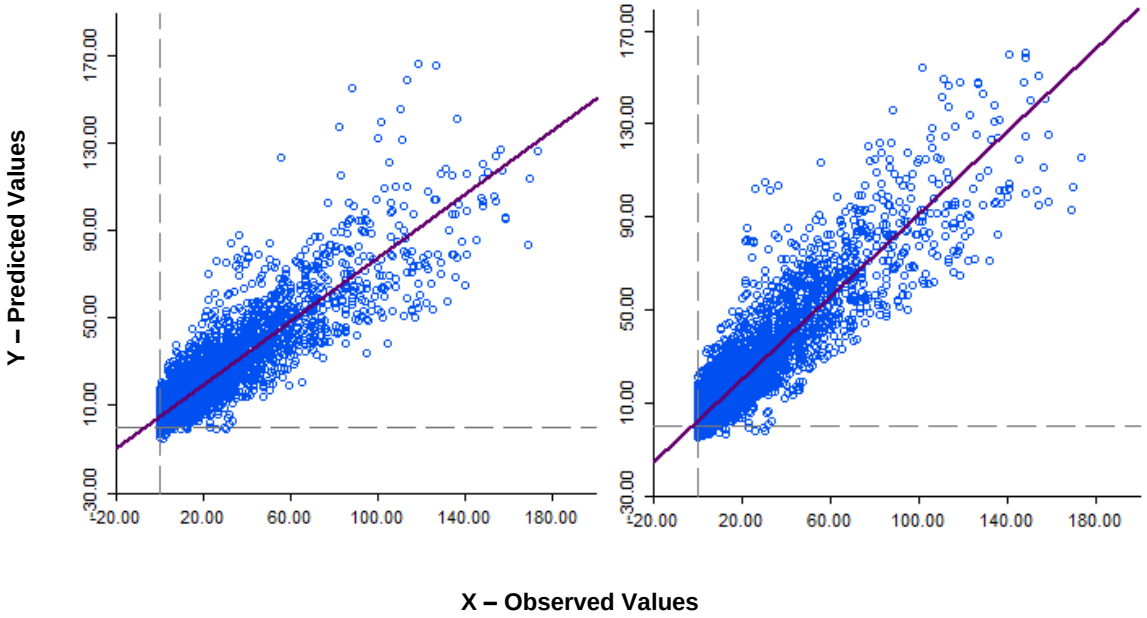
These conclusions further confirm what was stated at the end of chapter 3 (which was the purpose of the regression): commercial activity is distributed in a way that is still compatible with proposed classifications (Gaspar 1985; Cachinho, 2011; Cachinho & Salgueiro, 2016), with a general high order cluster of activity (Baixa), complimented by other clusters (Alvalade, Almirante Reis) which seem to use their *specialization as an anchor* (Personal Use Items in Alvalade-Roma-Av. Novas; Household Articles in Almirante Reis). Smaller clusters are present in a disperse way: eventually consisting of Foods, Cafes and Other Items. These smaller clusters are resilient (the verified relationship was already present in 1995, with Other Items adding to this “smaller clusters” in the period since), and rely directly on (resident) Population.

Reporting to the original question (can other types be used as predictors), it would seem so. But is it advantageous? Spatial Error was run for Foods, using Population, Primarily Non-Residential Buildings, % Activity within Malls, and all other types aggregated into “Other Commercial Activity”, a variable that consists of the density of the 9 types. Results were as follow.

Table 16 – Spatial Error for Foods, using aggregated commercial activity as a predictor, 2010-11

Variable	2010-11		
	coefficient	z- statistic	p-value
lambda	0.9813	335.494	0.000
Constant	-1.2660	-0.296	0.767
Other Commercial Activity	0.0660	39.770	0.000
Population	0.0019	28.109	0.000
Primarily Non-Residential Building	-0.1139	-6.875	0.000
% Activity within Malls	-7.7409	-5.621	0.000
Log-likelihood	-11363.147		
Akaike info criterion	22736.3		
Schwarz criterion	22767.3		

The model performs slightly worse as measured by Log-Likelihood, AIC and SC. A scatter plot between observed and predicted values shows the following.



Note: Points located around y = 170 are almost entirely located at Baixa.

Figure 27 – Foods, Observed and Predicted Values, 2010-11. Using Separate Types as Predictors (left) or as a Whole (right)

#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
3642	0.789	4.687	0.224	20.885	0.000	0.731	0.006	116.798	0.000
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3642	0.789	4.687	0.224	20.885	0.000	0.731	0.006	116.798	0.000

#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
3642	0.817	2.408	0.251	9.578	0.000	0.894	0.007	127.509	0.000
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3642	0.817	2.408	0.251	9.578	0.000	0.894	0.007	127.509	0.000

Figure 28 – Foods, Statistics of the Previous Scatters Plots. Using Separate Types as Predictors (above) or as a Whole (below).

Not only is R-squared better, but the slope is especially meaningful: it's much closer to 1 when using all types (with standard error being only slightly worse). Considering that, and also a better fit, as noticed on Table 16, one can expect to obtain a better *prediction* when using separate types (even though, not necessarily, a better *explanation*, considering the model limitations that have been stated). The explanation comes from what was said: commercial activity is not distributed in a homogenous way, and using different types allows that only spatially related types are considered significant when predicting for the presence of a specific one.

5. CONCLUSIONS AND POSSIBLE IMPLICATIONS FOR PLANNING

5.1. CONCLUSIONS

This dissertation intended to identify retail location patterns in Lisbon in 1995, 2002 and 2010, focusing on cross sectional patterns and their longitudinal changes, and relating them with population and housing data. The analyses did allow for the identification of expected changes (the role attained by shopping malls among higher order clusters), but also for some that weren't quite so obvious (the suggested composition of smaller, "lower order" disperse clusters, and its apparent resilience).

The first part of this work was mainly exploratory: GIS software was used to interpret spatial phenomena, while in the process getting acquainted with the data. This process was preceded by a literature review, which was essential for interpreting the results.

A general conclusion is that retail maintains its relevance, but with a significant part of the commercial activity now occurring at shopping malls. The restaurant sector showed persistent growth, both inside shopping malls and at street level.

A first conclusion is that between 1995 and 2010 there was *a transference of preferred location for retail*, from street to shopping mall, and *a transference of activity at street level,* with the growth in restaurants more than doubling the losses in retail.

The second part was more innovative: an attempt to establish relationships between different types of activities, and between those and demand, via a spatial error regression model, allowed to further elaborate on Lisbon's retail structure composition. Conclusions were as follows.

Conclusions about commercial activity location:

1) A higher order cluster of activity is found at Baixa, presenting a *density* of activities that's noticeable when modelling a regression, since Baixa presents itself is a systematic *outlier*. That density would put Colombo and Vasco da Gama at the same level, pertaining commercial centre order, but they don't present the same *diversity* of subtypes of activities. That gives Baixa the highest *range* within centres;
2) Specialized clusters like Almirante Reis and Alvalade create *information spillovers* in which different stores benefit from the presence of a specific type (Household Articles and Personal Use Items); they are a viable alternative to Baixa, immediately located below in a hierarchy of centres;
3) Smaller clusters, as said, are present in a disperse way and consist of *convenience* goods: Foods, Cafes and Other Items, therefore relying on *proximity* with Population. They are apparently resilient, as demonstrated not only by the regression results, but also by the apparent *population threshold* to

which they seem related as was discussed when analysing change (Figure 18): it seems expectable that residents will always favour having a Cafe + Other Items + Foods (+ some "lower range" Restaurants) near their residence (an apparent "proof" on this theory comes from supermarkets that used to locate only within shopping malls (or in large stores, outside city limits) and that are now adopting a strategy of locating smaller units inside residential neighbourhoods);

4) Other clusters ("community centres") like Estrada de Benfica might rely on a mix of these: some specialization; an interesting externality (e.g. accessibility in Estrada de Benfica); an historic accident (Campo de Ourique). Their composition will probably be site-specific and regression models should take this into consideration.

Conclusions about the developed models:

1) Modelling a regression presented itself as an interesting way of relating all types of activities; but it wouldn't have led to reliable conclusions if it wasn't for knowledge of the data that came from other analyses within the dissertation, and literature review;
2) A better fit in the regression would be found if only *same threshold activities were considered.* Different range activities have different markets: the *demand* is different, because the *distance* people are willing to travel to purchase them is different. Since only local demand / resident population was available as a predictor, the results obtained for the smaller, disperse clusters were the most convincing. If the only predictor (besides other types) is demand, *a regression model will perform better if similar range activities* are considered, *and only if the demand for all different ranges* can be estimated (Öner and Klaesson came to a similar conclusion (2015));
3) These considerations are made *for a model that has the restraint of only using demand as a predictor for commercial activity*. A better model can be built considering other variables, especially the ones than can account for commercial clustering: accessibility indicators (e.g. explaining Household Articles), rent (e.g. explaining Repairs), space syntax (e.g. not necessarily *explaining* Baixa, but *predicting* it, since it has a very specific space syntax – and for the sake of model fit improvement, explaining or predicting will be the same) and others, because they *could reduce the effect of the outliers* and clustering in general, thus improving model fit, and therefore, the model (*access to data* will be paramount).
4) At this point, one can risk saying that a *panel data model*, considering *both spatial lag and error*, and *including predictors other than demand*, while considering the conclusions about demand and the relationships established between different commercial types, could produce very satisfactory results. Further developments / implementation of the presented models will be focused on this.

The main objective of the dissertation was to gain knowledge about retail activity location within Lisbon. That objective was, to some point, accomplished. It was stated at Introduction that a better understanding on commercial location might contribute to planning. The following and concluding section speaks of that potential.

5.2. POSSIBLE IMPLICATIONS FOR PLANNING PRACTICE

This dissertation created an opportunity for demonstrating the importance of retail related issues when it comes to urban planning practice and urban management. Some of these are presented in this section, and will further be analysed in future studies about retail activity in Lisbon.

1) Baixa possesses a range that might explain its resilience, to some extent, considering the diversity of activities present within: only 10 were analysed, but Baixa, especially when considered alongside Chiado, possesses more than 100. There are multipurpose trips that can only be satisfied at Baixa-Chiado – some activity will never be transferred to shopping malls – it's too costly and it possesses a quality of time (in ecology, it'd be "age of the stand") that can't be replicated. On that quality of time and the possibilities it offers for planning: "King of Georgetown" Anthony Lanier seems to be applying that concept to Príncipe Real: by carefully adding high-end retail activities to the ones that already existed and whose *pathos* can't be replicated inside a shopping mall (antique stores, art galleries, historic bars), Eastbanc is creating a spillover effect that will have housing prices spiralling up, by coordinating an uncoordinated commercial cluster. If the private sector sees this as an investment, the public sector should focus on its importance on the liveability of a city – in this case, it can help in the urban regeneration of several blocks and numerous derelict buildings – the information spillover created by Eastbanc will create a positive effect for other buildings on the area. *Príncipe Real can be a case study for public planning: retail as an anchor for other land-uses. On a downside, this effect might also act as a catalyst, and at the same time a symptom of gentrification; this symptom, though, might be controlled by public planning, if identified*.
2) An interesting attempt has been made at Intendente (City Council services as an anchor), but not for enough time. A "permanent attraction" should be created. At this moment, "A Vida Portuguesa" is the biggest anchor at Intendente. This could be seen as an opportunity for planning: creating partnerships with "anchor stores" or creating attractions that have a permanent spillover effects (if Santos was to be a Design District for years, why wasn't the Museum of Design and Fashion (MUDE) located over there?) or both. It could have a combined effect for producing desirable spillovers. *Combining the potential of retail with that of a museum (like the one Joe Berardo wants to open a stone's throw from Lx Factory) is within the sphere of public planning.*
3) Other pertinent case studies could be found, perhaps, at Almirante Reis and Alvalade: at some point, the joint area of all units located inside those clusters amounted to that of a shopping mall. Considering Huff's model, that would probably explain why Alvalade and Almirante Reis were able to keep part of their trademark activities: they didn't compete store-to-shopping-mall, but cluster-to-shopping mall with Colombo and Vasco da Gama. That could, to some point, explain the resilience of activity at street level, inside specialized clusters, against the effect of shopping malls. And taking into consideration what was said to this point: if a shopping mall opens with a very disturbing effect on street level commercial activity, why not make them generate a counterpart by creating an underground car park at those areas? It has an impact on traffic congestion: but the effect of people

going from city centre to the shopping malls also has an impact. *A balance could be found, within the sphere of public planning.*

4) The fact that the shopping malls in Saldanha have, in recent years, lost most of its activity, reveals the importance of an anchor inside a shopping mall, but also how interesting it'd be to apply a Huff-model to the 3 major shopping mall areas (Colombo, Vasco da Gama and the ones in Saldanha) and analyse if Saldanha ever had any chance at competing with them. Vasco da Gama and Colombo were able to dislocate activity from street level all around Lisbon, but they also depended on the population located north (V. Gama) and west of Lisbon (Colombo). Calibrating a model considering distance/time and population served by those 2 shopping malls (and distance between them), and Saldanha, would provide an idea of the chances of Saldanha in succeeding. If it never had a chance, this should be considered in planning: it could help prevent having semi-empty spaces inside the city. *Competition can't be avoided in a liberal market, but negative effects on the city should be prevented: what will happen if Vasco da Gama (almost 50.000m2 of leasable area) becomes obsolete? Public planning should take this in consideration.*

Commercial activity is not only a measure of centrality or liveability: its location has greater implications than the ones that are usually considered in public planning. Further research, and further knowledge about commercial activity location decisions, can be of significant use, and shouldn't be overlooked by the public sector or by academia.

REFERENCES

Alonso, W. (1960). A theory of the urban land market. *Papers and Proceedings of the Regional Science Association, 6*: 149-57.

Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent. Cambridge, MA: Harvard University Press.

Alho, A., de Abreu e Silva, J. (2014) Analysing the relation between land-use/urban freight operations and the need for dedicated infrastructure/enforcement — Application to the city of Lisbon. *Research in Transportation Business & Management, 11:* 85-97

Anselin, L. (2005). Exploring Spatial Data with GeoDaTM: A Workbook. Spatial Analysis Laboratory, Department of Geography, University of Illinois, Urbana-Champaign

Barata-Salgueiro, T. (2001). Lisboa. Periferia e Centralidades. Oeiras. Celta.

Barata-Salgueiro T, Erkip F. (2014). Retail planning and urban resilience – an introduction to the special issue. *Cities, 36:*107–11.

Batty, M. (1971) Exploratory Calibration of a Retail Location Model Using Search by Golden Section. *Environment and Planning, 3*: 411 -432

Batty, M., Besussi, E., Maat, K., Harts, J. (2004). Representing multifunctional cities: density and diversity in space and time, *Built Environment, 30:* 324–337.

Batty, M. Zhong, C. Schläpfer, M, Müller, S., Ratti, C., Schmitt, G. (2015) Revealing Centrality in the Spatial Structure of Cities from Human Activity Patterns. *Urban Studies 54 (2),* 437-455

Batty, M., Piovani, D., Zachariadis, V. (2017). Quantifying Retail Agglomeration using Diverse Spatial Data. *Scientific Reports 7*, 5451

Berry, B., Garrison, W. (1958) The Functional Bases of the Central Place. *Economic Geography, 34(2)*: 145-154

Berry, B., Garrison, W. (1958a) A Note on Central Place Theory and the Range of a Good. *Economic Geography, 34(4)*: 304-311

Berry, B. (1963). Commercial Structure and Commercial Blight: Retail Patterns and Progresses in the City of Chicago. University of Chicago. Research paper n. 85.

Berry, B. (1967). Geography of Market Centres and Retail Distribution. Englewood Cliffs, N.J: Prentice-Hall.

Borck, R. (2007). Consumption and social life in cities: Evidence from Germany. *Urban Studies, 44(11):* 2105-2121.

Brown, S. (1991). Retail Location: The Post Hierarchical Challenge, *International Review of Retail, Distribution & Consumer Research, 1:* 367–381.

Brown, S. (1993). Retail Location Theory: Evolution and Evaluation. *International Review of Retail, Distribution & Consumer Research, 3(2)*: 185-229

Cachinho, H. (2002). O Comércio Retalhista Português. Pós-modernidade, consumidores e espaço. Lisboa: Gabinete de Estudos e Prospectiva Económica do Ministério da Economia (GEPE).

Cachinho, H. (2006). Consumactor – da condição do indivíduo na cidade pós-moderna. *Finisterra, 81:* 33-56

Cachinho, H. (2011). Urban retail dynamics: from shopping spaces to consumer places, in T. Barata-Salgueiro, H. Cachinho (orgs.), Retail planning for the resilient city. Consumption and urban regeneration, Lisboa: Centro de Estudos Geográficos

Cachinho, H., Barata-Salgueiro, T. (2016). Os sistemas comerciais urbanos em tempos de turbulência: vulnerabilidades e níveis de resiliência. *Finisterra. 51*: 89-109

Cadwallader, M. (1996). Urban geography: An analytical approach. Upper Saddle River, NJ: Prentice Hall.

Clarke, G. (1998) Changing methods of location planning for retail companies. *GeoJournal 45(4):* 289-298

CML. (2016). Geodados, Recenseamento Comercial. Retrieved from
(Accessed December 2016)

CML/DMEI (2009). Número de estabelecimentos de comércio a retalho e restauração da cidade de Lisboa, por sector de atividade económica. Recenseamento dos estabelecimentos de comércio a retalho e restauração e bebidas da cidade de Lisboa.

Costanza, R. (1989). Model goodness of fit: a multiple resolution procedure. Ecological Modelling, 47, 199-215.

Caplin, A. and J. Leahy (1998). Miracle of Sixth Avenue: Information Externalities and Search. *The Economic Journal 108(446):* 60-74.

Carter, C., Vandell, K. (2005). Store Location in Shopping Centers: Theory and Estimates. Journal *of Real Estate Research 27(3).* 237-266

Church, R., Bell, T. (1990) Unpacking Central Place Geometry. Single Level Theoretical Systems. *Geographical Analysis, 22*: 95–115

Christaller, W. (1933) Central Places in Southern Germany, tr.by C. Baskin, 1966. Englewood Cliffs, NJ: Prentice-Hall.

Clark, C. (1973) The Value of Agricultural Land. Oxford, New York: Pergamon Press

Reilly, W.J. (1931) The Law of Retail Gravitation, New York: W.J. Reilly.

Davies, R., Bennison, D. (1978). Retailing in The City Centre: The Characters of Shopping Streets. *Journal of Economic and Social Geography, 69:* 270–285

Dolega, L., Pavlis, M., Singleton, A. (2016) Estimating Attractiveness, Hierarchy and Catchment Area Extents for a National Set of Retail Centre Agglomerations. *Journal of Retailing and Consumer Services*, 28:78-90

Dudey, M. (1990). Competition by Choice: The Effect of Consumer Search on Firm Location Decisions. *The American Economic Review, 80(5):* 1092-1104.

Eaton, B. and Lipsey, R. (1975). The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition. *Review of Economic Studies 42:* 27-49.

Eaton, B. and Lipsey, R. (1982). An Economic Theory of Central Places. *Economic Journal, 92*: 56-72.

Einstein, A. (1923) Sidelights on relativity, New York: E.P. Dutton&Company

ESRI (2016). Optimized Hot Spot Analysis. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm (Accessed December 2016)

ESRI. (2016a). Point Density. Retrieved from http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/point-density.htm (Accessed December 2016)

ESRI. (2016c). Band Collection Statistics. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/band-collection-statistics.htm (Accessed December 2016)

ESRI. (2016c). Zonal Statistics as a Table. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/zonal-statistics-as-table.htm (Accessed December 2016)

Fisher, I. (1930) The Theory of Interest, as Determined by Impatience to Spend Income and Opportunity to Invest It. New York: Macmillan.

Franzese, R, and Hays, J. (2014). Testing for spatial-autoregressive lag versus (unobserved) spatially correlated error-components." Benjamin F. Shambaugh conference: New frontiers in the study of policy diffusion. University of Iowa. Iowa City.

Fujita, M., Gabszewicz, J. J., Thisse, J. F., Schweizer, U. (1988). Location Theory. London: Routledge

Fujita, M., Krugman, P., Mori, T. (1999). On the evolution of hierarchical urban systems. *European Economic Review, 43(2):* 209-251.

Fujita, M., Thisse, J.-F. (2002). Economics of agglomeration: Cities, industrial location, and regional growth. Cambridge: Cambridge University Press.

Gaspar, J. (1985). A dinâmica funcional do centro de Lisboa. Lisboa: Livros Horizonte.

Guy, C. (1998) Classifications of Retail Stores and Shopping Centres: Some Methodological Issues. *GeoJournal, 45 (4)*: 255-264

Haig, R.M. (1926). Toward an understanding of the metropolis: some speculations regarding the economic basis of urban concentration. *Quarterly Journal of Economics, 40*: 179-208.´

Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*: 41-57.

Hoyt, Homer. (1958). A Re-Examination of the Shopping Center Market. Washington, D.C.: Urban Land Institute.

Huff, D. (1963). A probabilistic analysis of shopping centre trade areas. *Land Economics, 39*: 81-90.

Huff, D. (1964). Defining and estimating a trading area. *Journal of Marketing, 28(3):* 34-38.

INE. (2014). Base Cartográfica. Subseccao Estatistica. Retrieved from http://www.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos_base_cartogr (Accessed December 2016)

INE. (2014a). Censos. Retrieved from http://mapas.ine.pt/download/index2011.phtml (Accessed December 2016)

Jakle, J., Mattson, R (1981). The Evolution of a Commercial Strip. Journal of Cultural Geography, 1(2): 12-25

Jacobs, J. (1961). The Death and Life of Great American Cities. New York: Random House.

Klaesson, J., Öner, Ö. (2015). Market Reach for Retail Services. *The Review of Regional Studies*, *44(2):* 153-176

Krugman, P. (1991). Increasing Returns and Economic Geography, *Journal of Political Economy, 99*, 483–499.

Larsson, J., Öner, Ö. (2014). Location and co-location in retail: a probabilistic approach using geo-coded data for metropolitan retail markets. *The Annals of Regional Science, 52(2)*: 385-408

Losch, A. (1940) The Economics of Location, tr. W.H. Woglam and W.F. Stolper, 1954. New Haven: Yale University Press

Microsoft. (2017). CORREL function. Retrieved from https://support.office.com/en-us/article/CORREL-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92 (accessed January 2017)

Öner, Ö. (2014). Retail Location (Unpublished doctoral thesis). Jönköping University, Småland, Sweden

Openshaw, S. (1975) Some Theoretical and Applied Aspects of Spatial Interaction Models, Norwich: Geobooks.

Openshaw, S. (1983) The Modifiable Area Unit Problem, Norwich: Geobooks.

Porta, S., Strano E., Iacoviello V., Messora R., Latora V., Cardillo A., Wang F., Scellato S. (2009). Street Centrality and Densities of Retail and Services in Bologna, Italy. *Environment and Planning B: Urban Analytics and City Science, 36(3)*: 450-465.

Scott, Peter. (1967) Geography and Retailing. London: Routledge.

Sevtsuk, A., 2010. Path and Place: A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA. PhD dissertation in Urban Design and Planning, Massachusetts Institute of Technology. Cambridge MA.

Sevtsuk, A. & Mekonnen, M. (2012). Urban network analysis. A new toolbox for ArcGIS. *Revue internationale de géomatique. 22*: 287-305.

Sevtsuk, A. (2014). Location and Agglomeration: The Distribution of Retail and Food Businesses in Dense Urban Environments*. Journal of Planning Education and Research, 34*: 374-393.

Sevtsuk, A., K. Raul. (2017). Patronage of urban commercial clusters: A network-based extension of the Huff model for balancing location and size. *Environment and Planning B: Urban Analytics and City Science*

Smith, A. (1976). An Enquiry into the Nature and Causes of the Wealth of Nations. The Glasgow Edition of the Works and Correspondence of Adam Smith. Oxford: Oxford University Press (Original work published 1776).

Song, Y. & Knaap, G. (2004). Measuring the Effects of Mixed Land Uses on Housing Values. *Regional Science and Urban Economics. 34:* 663-680.

Spatial Regression with GeoDa (n.d), Geographical Information Systems Institute, Center for Geographic Analysis, Harvard University Retrieved from https://cga-download.hmdc.harvard.edu/publish_web/GIS_Institute/2017_summer/OpenGeoDa3.doc Accessed March 2017

Stahl, K. (1987). Theories of Urban Business Location. *Handbook of Regional and Urban Economics. 2:* 760-820.

Survival Blog for Scientists. (2010, May 19). Retrieved from http://www.sciencesurvivalblog.com/phd_life/getting-scooped_2435 Accessed May 2017

Toponímia de Lisboa. A Rua de S. Julião (n.d.). Retrieved from https://toponimialisboa.wordpress.com/2015/11/16/a-rua-de-sao-juliao-e-dos-algibebes/ Accessed February 2017

Thurstain-Goodwin M., Unwin D. (2000). Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations*. Transactions in GIS, 4:* 305–317.

von Thünen, J. (1826). The Isolated State, tr. Wartenberg, C.M., 1966. Oxford, London: Pergamon Press.

ANNEX I

1.Foods
Bebidas
Talhos
Charcutarias
Peixarias (Frescos E Congelados)
Produtos Horto-Frutícolas
Frutarias
Padarias
Produtos Pastelaria E Confeitaria
Leite E Derivados
Queijos
Criação E Caça
Produtos Congelados
Conservas, Frutos Secos E Aperitivos
Prod. Naturais (Diet.,Medicinais,Espec.)
Bombons E Chocolates
Café, Chá E Sucedâneos
Pronto A Comer (Churrasq.,Comida A Peso)
Outros N.E (C. Alim. Especializado)
Produtos Biológicos
Hipermercados
Supermercados
Mini-Mercados/Auto-Serviços
Mercearias
Cooperativas De Consumo, Cantinas
Outros N.E (C. Alim. N Especializado)
Lojas De Conveniência
2. Personal Use Items
Artigos Vestuário Diverso (Misto)
Artigos Vestuário Homem
Artigos Vestuário Senhora (Boutique)
Artigos Vestuário Criança
Roupa Bébé
Roupa Interior

Artigos Malha
Lãs
Camisarias
Vestuário Pele
Tecidos Para Confecção Vestuário
Luvas
Chapéus E Guarda-Chuvas
Marroquinaria E Artigos De Viagem
Sapatarias
Retrosarias
Bijuterias
Outros N.E (C.N.Alim.Uso Pessoal)
Artigos De Puericultura
3. Household Articles
Molduras
Artigos Diversos Lar/Decoração
Carpetes E Alcatifas
Tecidos Para Decoração
Roupa/Lar
Móveis E Colchoaria (Interior)
Móveis Antigos/Antiguidades
Mobiliário Cozinha E Casa De Banho
Mobiliário Jardim/Campo/Praia (Exterior)
Candeeiros/Iluminação
Electrodomésticos
T.V./ Video / Alta Fidelidade
Utensílios Para O Lar (Cutel.,Quinq.)
Loiças/Vidros
Revestimentos Lar (Papel Parede, Etc.)
Artigos Sanitários
Outros N.E (C.N.Alim.Equipament. Lar)
Antiquários

4. Health And Hygiene
Farmácias
Perfumarias
Material Óptico (Oculistas)
Instrumentos E Art.Médicos-Cirúrg./Ortop
Outros N.E. (C.N.Alim. Saude E Higiene)
Outros Produtos De Saúde E Cosmética
5. Leisure
Livrarias
Papelarias (Art. Escrit.,Jornais ..)
Tabacaria
Alfarrabista/Filatelia/Numismática
Galerias De Arte
Artigos De Desporto
Artigos De Campismo
Artigos De Bricolage
Artigos Jardinagem/Flores/Plantas
Pequenos Animais,Peixes Aqua.
Discos/Cassetes Audio E Video
Material Fotográfico,Cinema
Instrumentos Musicais
Brinquedos E Jogos Para Crianças
Outros N.E. (C.N.Alim.Cultura/Lazer)
6. Others
Materiais Construção,Metais E Ferragens
Tintas,Vernizes E Afins
Drogarias
Veículos Ligeiros (Novos)
Veículos Pesados (Novos)
Tractores,Reboques E Semi-Reboques
Motociclos
Acessórios P/ Veículos Excepto Motociclos
Combustíveis Veículos Automóveis

Combustíveis Liq,Gas. Não Destinado A Auto
Tubos,Borrachas,Plásticos
Material Eléctrico E P/ Electrónica
Equipamento P/ Industria Hoteleira
Equipamento Informático
Equipamento E Mobiliário De Escritório
Alfaias Agrícolas
Solas E Cabedais
Artesanato
Artigos De 2ª Mão
Prendas
Relojoarias
Ourivesarias/Relojoarias
Instrumentos Profissionais E Cientificos
Armas,Munições E Artigos De Caça
Sementes Adubos E Rações
Outros N.E. (C.N.Alim. Diversos)
Telecomunicações
Veículos Ligeiros (Usados)
Veículos Pesados (Usados)
Bicicletas
Grandes Armazéns
Bazares
Outros N.E. (C.N.Alim.Não Especial.)
7. Repairs
Reparação De Veículos Automóveis
Reparação De Motocíclos
Reparação De Calçado E Art. Couro
Reparação Art. Eléct. De Uso Domest
Reparação De Relógios E Bijutaria
Outras N.E. (Reparações)
Reparação Tv/Video/Hi-Fi
Arranjos De Costura
Reparação De Computadores

8. Restaurants And Similar
Restaurante - Luxo
Restaurante - Luxo (Típico)
Restaurante
Restaurante (Típico)
Restaurante - Cozinha Internacional
Cervejaria / Marisqueira
Snack-Bar
Self-Service
Outros Estab. De Comidas E Refeições
Cervejaria / Marisqueira 3ª
9. Cafes And Similar
Café/Pastelaria
Casa De Chá
Gelataria
Estab. Bebidas Tradicional
Outros Estab. De Bebidas
Outros Estab. De Bebidas - Taberna
10. Bars And Similar
Bar/Pub
Sala De Dança Discoteca
Sala De Dança Com Espectáculo
Outras Salas De Dança