Cities through the prism of people spending behavior

Stanislav Sobolevsky[1,*], Izabela Sitko[2], Remi Tachet des Combes[1], Bartosz Hawelka[2], Juan Murillo Arias[3], Carlo Ratti[2]
1 Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA, United States of America
2 Department of Geoinformatics - Z_GIS, University of Salzburg, Salzburg, Austria
3 New Technologies, BBVA, Madrid, Spain
∗ **E-mail: stanly@mit.edu**

Abstract

Scientific studies of laws and regularities governing human behavior increasingly rely on digital traces produced by various aspects of human activity. In this paper, we introduce a new source of large-scale ubiquitous data set: the anonymized countrywide records of bank card transactions performed by the customers of one of the largest Spanish banks. We employ this data to study how individual spending behavior is qualitatively and quantitatively affected by various factors such as customer age, gender as well as the home city size. Normalizing for the discovered dependencies, we quantify the impact of each particular city on it's residents spending behavior. Based on this impact we propose a novel scale-independent classification of cities across Spain, which appears to be generally consistent for different ways of city definition and could be given a meaningful socio-economical explanation.

Introduction

Laws and regularities in human behavior have been the subject of intense research for decades. In the age of ubiquitous digital media, different aspects of human activity are being increasingly analyzed by means of their digital footprints, such as mobile call records [1–5], social media posts [6–8] or smart cards usage [9,10]. The wide popularity of debit and credit cards, which are intensively replacing cash spending, fosters the appearance of yet another source of information. The extensive data set of transactions collected by banks and other providers of payment infrastructures offers direct evidences on the economic activity of individuals. While the spatio-temporal granularity of this data may be sparser compared to the previously explored sources, incorporated contextual information such as category, magnitude or place of spending, enable to analyze not only patterns but also content of human activity. In addition to that known demographic profiles of bank customers give an additional important layer of information able to explain and estimate an impact of important factors behind such activity. In the urban context aggregated form bank card data may further provide a novel mean of describing and comparing economic dimension of cities, similarly to other types of digital records that have already proved its potential for a discovery of urban structure [11], land use [12], mobility [13, 14] or well-being [10].

Previous studies of individual economic activity were mostly based on dedicated field studies [15], questionnaires [16] and surveys [17], or retailer information [18]. Shopping patterns have been put in relation with demographic factors such as gender, age, education, occupation or income [17,19,20]. They were also found to depend on shopping context such as retail characteristics [21], season of a year [22] or, in the context of e-commerce, service functionality [16]. The important impact of demographics on bank card usage was demonstrated in studies comparing propensity for different payment methods among groups of customers [23–26]. Regarding gender, some results indicate an increased likelihood of bank cards usage among women [24, 26], while others point to their preference for checks over cash or cards [25]. Age is reported to either lower the probability of card usage [26] or have no significant effect [25]. Since those findings were based upon survey results, we see their verification using actual measures of human spending as an interesting scientific challenge.

Direct analysis of individual bank card records has not been extensive so far. Card transaction data are highly sensitive and include a lot of private information, their access is of course restricted. Hitherto, applications were mostly focused on issues regarding the card system itself [27, 28], rather than on the associated human behavior. More recently, Krumme et al. [29] employed this new type of data to uncover the predictability of spending choices, and their relation to wealth. Initial analysis of bank card data was also carried out in the field of regional delineation [30] and analyzing city attractiveness for different types of customers [31].

Another angle to analyze the factors governing economic activity of humans is through its geographic location. In the case of urban customers, this notably means a city of residence and its size, considered in terms of population. Due to agglomeration effects and intensified human interactions, a variety of urban processes were shown to vary with the number of inhabitants in the form of scaling laws [32–35]. While urban infrastructure dimensions, e.g. total road surface, reveal a sublinear relation to city size, socioeconomic quantities e.g. gross metropolitan product, crime rate, patenting and human interactions usually increase in a superlinear manner [36]. One can expect that human spending in urban areas holds similar property, which is one of the hypothesis being verified by the authors.

In this paper, individual purchase activity via bank card records is explored in order to discover collective patterns of economic behavior. We analyze the data set of customer spending across Spain during the year 2011 and ask a broad range of questions regarding the fundamental factors that impact economic conduct. At first, we focus on the influence of age and gender, wondering about their repercussion on five representative quantities characterizing said conduct. Beyond the impact of demography, we further investigate if people from different places tend to spend money in a different way. Therefore we seek for the impact of a city size on the economic activity of its inhabitants, in line with the previous studies on scaling laws governing urban quantities.

However the variation of customer spending behavior from one specific city to another goes far beyond just the impact of city size. Such individual variations in collective spending behavior are used to built an index of a city performance. With respect to each city's size and demographic composition, we compute the relative spending profile of its inhabitants. Following the general approach of [37], we treat the appropriately quantified deviations from the general trends as a numerical characterization of the observed discrepancies and the actual signature of a city scale-independent performance. Those signatures are used to build a novel scale-independent classification of Spanish cities.

Materials and Methods

Data set of bank card transactions

Our study relies on the complete set of bank card transactions (both debit and credit) performed by the Spanish customers of Banco Bilbao Vizcaya Argentaria (BBVA) within the country during 2011. The total number of active customers reaches around 4.5 M, who executed more than 178 M transactions, with a cumulative spending exceeding 10.3 billion euro. Due to the sensitive nature of bank records they were anonymized by BBVA prior to sharing, in accordance to all local privacy protection laws and regulations. Randomly generated IDs of customers are connected with certain demographic characteristics and an indication of a residence - at the level of zip code, further aggregated into coarser spatial units. Each transaction is denoted with its value, a time stamp, a location of a retail point where it was performed, and the business category it belonged to. The business classification includes 76 categories such as e.g. Restaurants, Gas Stations, Supermarkets or Travels. In order to compensate for the inhomogeneous penetration of BBVA on the individual banking market in Spain we normalize the activity of customers by the BBVA market share in the respective residence location (provided by the bank). The raw data set is protected by the appropriate non-disclosure agreement and is not publicly available. However, the researchers and the data provider may share certain aggregated data upon request and for the purpose of findings validation.

Major characteristics of customer's spending behavior

In order to characterize the spending behavior of customers we consider five basic parameters of bank card usage. Three of them are related to the economic dimension of transactions:

- the activity of each customer, defined as the total number of transactions performed during a year
- the average value of a single transaction
- the spending diversity, measured by the number of distinct business subcategories visited by a customer during 2011

Additionally, we introduce two characteristics of customers' mobility:

- distant mobility, measured as the percentage of transactions executed over 200 km from home
- local mobility, measured as the average distance between the customer's home location and the retail points (calculated based on transactions made within 100 km from home)

Correct computation of four of the aforementioned quantities (all but activity) requires the consumer to be using his bank card frequently enough, e.g. there is no point in measuring the spending diversity of someone who use a card only occasionally. In the further analysis, when it is necessary, we thus only consider customers who performed at least 50 transactions in 2011 (which gives an average close to one transaction per week). Moreover, for the sake of consistency, particularly in the cases of activity and spending diversity, we restrict the analysis to customers active during the entire year. In practice, we identify that set of customers as those who performed at least one transaction during both the first and the last month of the year. All five characteristics of spending behavior are further considered in the scale of a city - as an average value across the activity of residents.

Three levels of a city definition

For a spatial definition of Spanish cities we test three different types of units. The coarser city level consists of 24 Large Urban Zones (LUZs) as defined by the European Urban Audit Survey [38]. For the intermediate city level we aggregate Administrative Cities of Urban Audit into 40 Functional Urban Areas (FUAs), so as to reflect metropolitan regions in agreement with the Study on Urban Functions of the European Spatial Planning Observation Network (ESPON) [39]. The finer level of city definition concerns 211 conurbations (CONs) identified within the AUDES project (Aŕeas Urbanas de Espanã [40]). Population for LUZ and FUA city levels were obtained from Eurostat [41] and National Statistics Institute of Spain [42], and for the CON level from the AUDES project.

Impact of demography on customers behavior

Among the primary factors one could think about to affect human economic behavior are age and gender [17, 19, 20]. Similarly in our case, we can expect men and women, as well as younger or older people, to exhibit different spending habits. In order to verify this hypothesis, we explore how the distributions of five parameters of bank card usage - namely activity, average value of a transaction, spending diversity, distant and local mobility - changes with customer age for both genders. We present results for respective parameters on Figures 1-4. From a global perspective, one can observe that even though trends for both genders are substantially different quantitatively, in most cases they exhibit remarkable similarities in their shape from a qualitative viewpoint. For instance, the number of transactions is usually higher

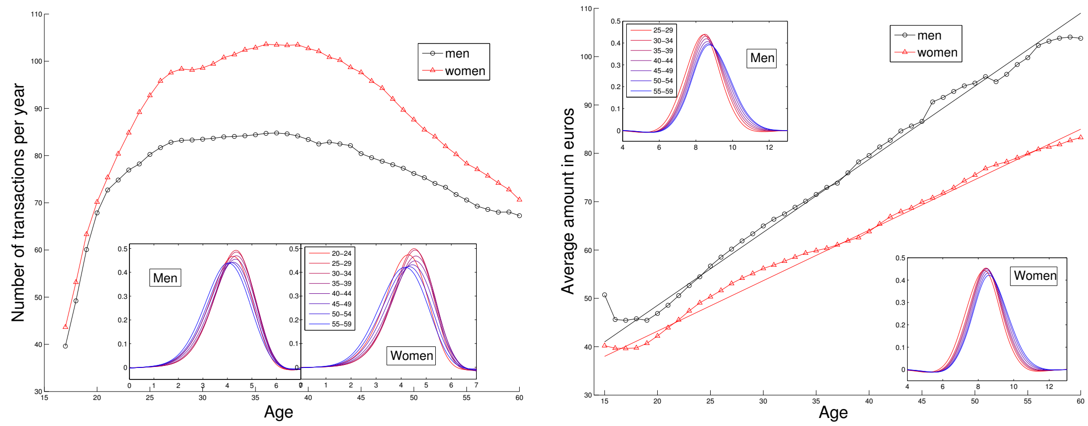

Figure 1: Impact of age and gender on (A) the average number of transactions per year and on (B) the average amount of a single purchase

for women, while the average value per transaction is higher for men, who seem to concentrate their economic activity more than women. Also, while the spending diversity of women customers is higher, their mobility is substantially lower on average. Nevertheless, the tendency for both men and women, as well as the important age thresholds where these tendencies change, appear to be strikingly similar. Let us now take a closer look at the respective parameters.

Customer activity and average amount per transaction

On Figure 1A, the average number of transactions per year is plotted against age, for both men (in black) and women (in red). We also aggregated the data into age groups and plotted the distribution of the transactions number (in log) for five-year brackets (between 20 and 24, 25 and 29, 30 and 34, and so on). One can see that customer activity increases rapidly between 18 and 30 years old, as expected with the entry in active life. It then reaches its peak and remains more or less constant for both genders till 40 years, before starting to steadily decrease. From an economic point of view, it thus appears that people are the most active during their thirties. Moreover, comparing the two curves shows that women make every year in average, 20 extra transactions.

After the number of transactions, let us focus on the average amount of each one of them presented on Figure 1B. It is quite remarkable to notice that quantity's average grows with the customers age in a quasi linear way, doubling between the youngest (18 years old) and the oldest (60 years old) customers. This effect could be intuitively explained by the ability of older people to spend more or to buy more expensive goods, as they potentially have more money, nevertheless it does not go in line with the pattern displayed by the total amount of spending (a quick increase till 40 years old and then no more variations, the actual graph can be found in SI). Thus it looks more natural to explain the steady increase of the average purchase amount by a habit of concentrating purchases, making fewer transactions but buying more every time.

As to gender differences, the average amount per transaction is smaller for women, who appear to spend more often, but for smaller values. These two statements might be related to the following demographic fact: between the age of 20 and 60 years old, 75% of women and 89% of men are active. And even though the situation is evolving (the percentage of working women was only 65% in 2005), one cannot help but think that women remain more involved in domestic tasks, in particular the essential shopping. Consequently, they would be bound to use their bank card more frequently and for smaller amounts.

Spending diversity

Given the data at hand, we are also able to analyze a diversity of human economic behavior based on a variety of places where people spend their money. In Figure 2, we correlate the average number of business categories visited over a year with customers' age and gender. After a rapid increase until the age of 27-28, the diversity of spending declines steadily in a linear way, showing that, while people spend more money growing old, they also tend to spend it in fewer types of businesses. One can only wonder if the abrupt change in the trend occurring in the late twenties is related to the foundation of a family.

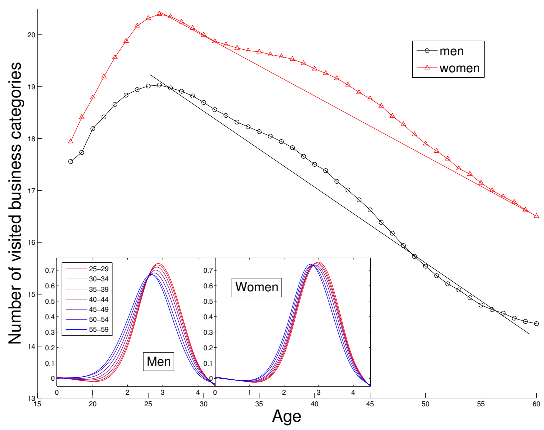

Figure 2: Average spending diversity against age for men and women

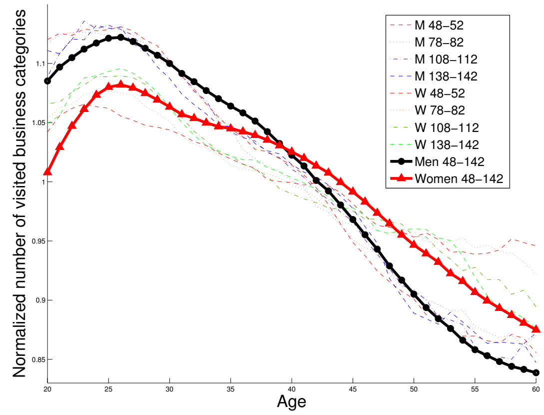

Figure 3: Normalized number of visited business categories for different activity levels

As far as gender is concerned, we notice that women tend to visit a larger number of business categories. This raises a question however. In Figure 2, the average was taken on every customer of a given age regardless of their total number of transactions, while we have seen that age greatly impacts the transaction activity which in turn could impact the observed diversity perhaps. To ensure that this observed impact on diversity is not biased by different activity of the customers of different age, we group customers according to their level of activity (from 50 +/- 2 to 140 +/- 2) and plot on Figure 3 the normalized number of visited business categories for each group. We also plot the same quantity for the entire set of customers (the thick black and red lines). Here too, the graphs exhibit the pattern already seen, which confirms our previous conclusions.

Customer's mobility

While home location of a customer is irrelevant to the aforementioned considerations, it becomes essential when studying human mobility. In the data, for each customer is given an address zip code. However, the exactness of this zip code is questionable: students registered at their parents home, people moving and not informing their bank... To get rid of that bias, we compute the fraction of transactions that took place in the theoretical home zip code and discard all customers for which said value is smaller than 0.6. They represent around 18 % of customers. On Figure 4, we plot the two parameters of customers' mobility (percentage of transactions performed more than 200km away from home and average distance traveled to businesses less than 100km from home) against age for both genders.

The distant mobility, shown in 4A is the first quantity showing a big difference in trends between men and women under 40 years old. Whereas for men a fraction of distant purchases increases in a roughly linear way for all ages, distant mobility of women firsts stagnates, then decreases until 40 years old, and finally starts to increase similarly to the curve observed for men. In a parallel way to the analysis of

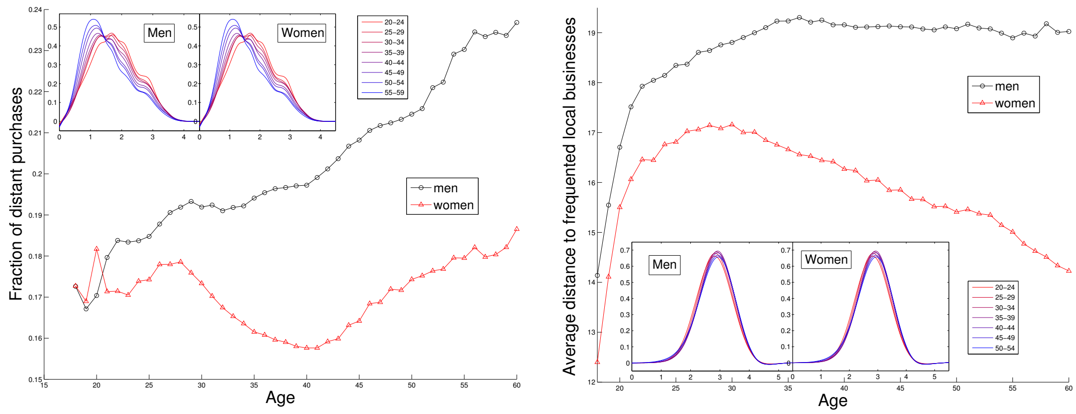

Figure 4: Impact of age and gender on the customer mobility (A) frequency of distant travels and (B) average distance to home of local transactions.

activity, one can think of societal explanations to cast light on such a difference. The average age for childbearing in Spain is 29.8 years old , which strikingly corresponds to the change in the curve evolution. Less active from a professional point of view and more tied to home, the old cliches remain partially true. Regarding local mobility, an interesting pattern can be identified in Figure 4B. While local mobility of men remains nearly stable after 25-30 years old, with only a very slight tendency to decrease with age, women exhibits significant and stable decrease of the average distance to visited local retailers. They tend to shop closer to their home when growing old.

Bigger cities boost up spending activity

It is well established that living in a bigger city boosts up many aspects of human life: intensity of interactions [35], creativity [37], economic efficiency e.g. measured in GDP [36], as well as certain negative aspects: crime [37], infectious diseases [33]. In the following section, we examine if this property holds true for the individual economic activity of city residents. To do so, average values of our five bank card usage characteristics are quantified and their dependance on city size, expressed in terms of population, analyzed. As the urban scaling laws were found sensitive to the selection of city boundaries [43], we test and compare three levels of city definition, namely 24 Large Urban Zones (LUZs), 40 Functional Urban Areas (FUAs) and 211 conurbations (CONs).

In the previous sections we prove individual economic behavior to depend on customers demography. It thus appears necessary to take into account possible variations of demographic profiles between the different cities considered. As a matter of fact, age and gender vary quite significantly from one city to another. Among the 24 LUZs for instance, the fraction of male customers lies between 47.5 and 51%, the average customer age goes from 41 to 48 years old and the respective fractions of different age groups change up to a factor of 1.7. In order to correct for that demographic heterogeneity, we normalize each of the city characteristics by their expected value (computed using the demographic composition of the city and the average parameters estimated on the entire set of customers, for more details, see SI).

On Figures 5 and 6, the total activity of a given city is log-log plotted against its size for the three levels of city definition. In accordance with existing studies, we observe a slightly superlinear scaling with exponents 5.2 %, 4.4 % and 4.8% for LUZs, FUAs and CONs respectively. Importantly, the exponents for all city levels are of the same order, pointing to a discovered scaling as a distinctive feature of urban areas, regardless of the considered boundaries. Satisfactory confidence of the observed trends are further proved by the corresponding p-values. The results are obtained considering all types of businesses. With the data at our disposal, we are also able to distinguish different categories of businesses. And because people have common needs, regardless of their place of residence (e.g. spending on food or gas), studying the scaling for different consumption types might bring interesting results.

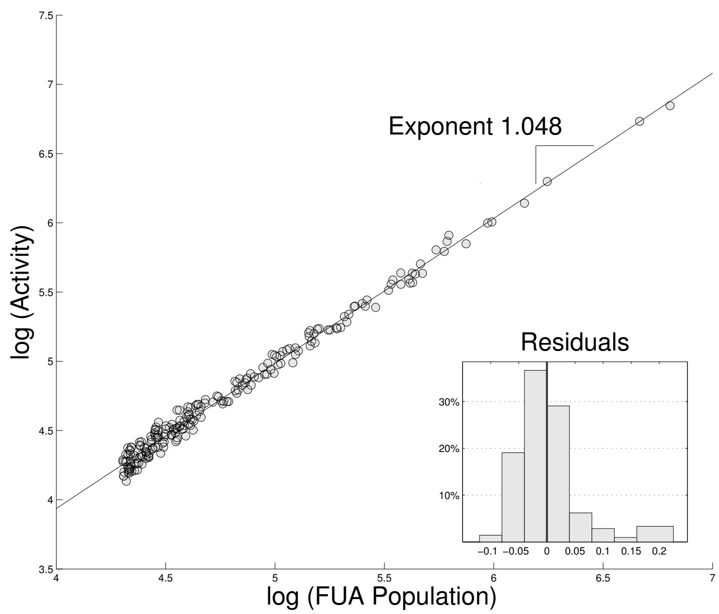

Figure 5: Superlinear scaling of total spending activity (cumulative number of transactions made by city residents) with CON size. Exponent: 4.79%, CI: [0.03,0.06], p-value: $3e$-9%, $R^2 = 15.6\%$.

In the Table 1, we give the scaling exponents for a few business categories. To confirm validity of trends we also provide confidence intervals, p-values, along with coefficients of determination. The fraction of activity represented by the corresponding business category is in the last column. As can be seen, entertaining activities like traveling, going out for a drink or dinner or for a party is boosted by city size, with a scaling exponent of 13.1 %. In bigger cities, people seem to engage more easily in social activities, which confirms [35]. Similarly, health institutions are more frequented in big cities, a fact that can be put in relation with results from [33] showing the superlinear scaling of diseases outbreaks with population, the proximity of a larger number of individuals generates a higher need for medical care.

On the contrary, and as one could have foreseen, fundamental needs are less impacted by city size, with exponent values almost equal to 1. Living in a big or a small town does not affect your attendance to grocery stores, supermarkets or gas stations. A last category that caught our eyes is the business of wellness beauty and fashion. It seems, surprisingly, uncorrelated to the number of inhabitants, while perhaps one could have expected a larger consumption of that kind of offers in larger cities. Visually, the differences in slope for a few meaningful categories can be compared in Figure 7.

The other transactional parameters considered in this paper demonstrate various behaviors in terms of scaling. Let us go over a few quantities (all others, together with the respective graphs, can be found in SI). The average amount per transaction seems to be generally independent from city size. Conversely, a small scaling trend appears for spending diversity, suggesting that customers from larger cities have a slightly broader variety of purchases. The exponent of 3.27 % proves that phenomenon with a 95.8% confidence (in the case of LUZs). The most meaningful case in terms of mobility is the one observed for

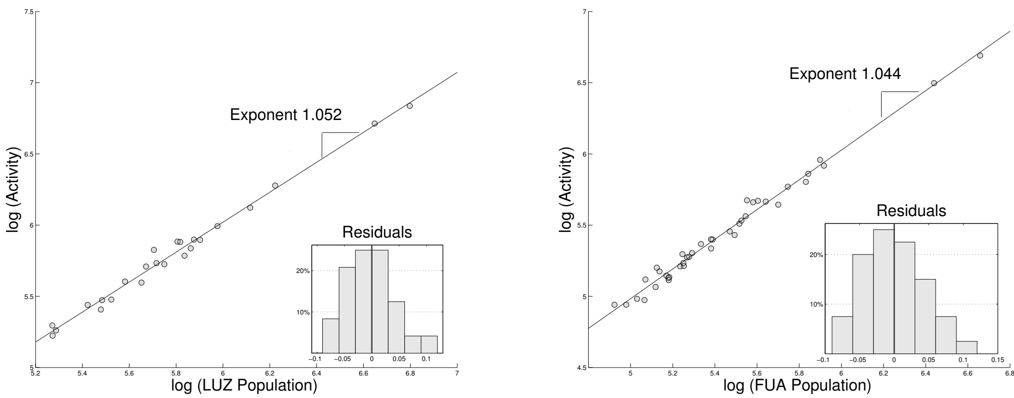

Figure 6: Superlinear scaling of total spending activity (cumulative number of transactions made by city residents) with city size for (A) LUZ and (B) FUA. LUZ exponent: 5.17.%, CI: [0.0,0.10], p-value: 4.51%, $R^2 = 17.0\%$. FUA exponent: 4.36%, CI: [0.0,0.08], p-value: 3.22%, $R^2 = 11.51\%$.

Table 1: Scaling of customers' activity with LUZ size for different business categories.

Business category	Exponent	Confidence intervals	p-value, R^2	Fraction of activity
Everyone	1.044	$[0.0, 0.09]$	$6.8\%, 14.4\%$	100%
Bars, restaurants and clubs	1.131	$[0.05, 0.21]$	$0.3\%, 33.1\%$	7.66%
Travels	1.051	$[-0.05, 0.15]$	$33.1\%, 4.3\%$	2.02%
Health institutions	1.049	$[-0.03, 0.12]$	$19.6\%, 7.5\%$	4.33%
Entertainment	1.024	$[-0.08, 0.13]$	$64\%, 0.9\%$	0.41%
Gas and Supermarkets	1.007	[-0.03, 0.04]	$70\%, 0.6\%$	51.99%
Wellness, beauty and fashion	0.992	[-0.08, 0.13]	$64\%, 0.9\%$	17.92%
Others	1.071	$[0.0, 0.14]$	$4.4\%, 17.1\%$	15.67%

CON city level: local mobility exhibits a downward trend, with a scaling exponent of -3.1%, while distant travels show a 15.8% positive scaling. It appears that in larger cities, people find what they need to buy closer to home, thus sparing them from longer journeys to perform local transactions. The uncovered trend for distant mobility (strongly increasing with population) indicates that inhabitants of large cities explore the country on a wider dimension. Mobility trends however depend on the city definition as of course mobility at the scale of a large urban zone could be different from the one at the scale on a core city.

All together those observations lead us to a new scale-free classification of Spanish cities based on the deviation of their particular individual impact on the spending behavior of city residents from the corresponding trend predictions.

Classification of Spanish cities beyond the impact of demography and scale

Scaling laws described in the previous section explain how economic behavior of residents is expected to change with city size in general. However, the actual values of the spending parameters always

Figure 7: Superlinear scaling of total spending activity (cumulative number of transactions made by city residents) with city size for different types of businesses

deviate from the trend estimations due to the additional influence of local factors. For example from the figures 5, 6 one can see that although p-values point out statistical significance of the trend, the R^2 values in the range between 10 and 20% show that only a small part of the observed value variation could be actually explained by the trend. A higher value can be treated as an indication of an over-performance in a given domain. On the contrary, a value lower than indicated by a scaling trend may be interpreted as under-performance. Such deviations are assessed in relation to the size-specific estimates, therefore they can be used for the qualitative comparison of different cities, regardless of their population. Similarly to [37], we quantify those deviations as log-scale residuals, i.e., the decimal logarithm of the actual city characteristic minus the decimal logarithm of the corresponding trend estimation. Since the five parameters were normalized beforehand for age and gender variability, the residuals are free from the impact of both scale and demography. This allows to define a novel classification of Spanish cities revealing the impact of local circumstances on residents' spending behavior.

The set of five log-scale residuals, corresponding to the scaling trends of the five spending parameters, represents a distinctive signature for each city. Depending on the strength of the trend, the residuals of different parameters exhibit different ranges of deviations. In order to bring the values on a common scale we apply a standard z-score transformation (subtracting the mean and normalizing with standard deviation). We consider each city definition level separately, which results in three separate sets of city signatures. Similarity between the signatures of cities within one level is assessed based on the k-means clustering algorithm [44], optimized with the majority voting across several dozens iterations in order to receive stable separation of clusters. According to the silhouette metric [45], the most optimal approach is to divide cities into three clusters in cases of CONs and FUAs, and two clusters for LUZs (detailed values are provided in SI). However, at all levels the division into three clusters introduce an additional pattern, meaningful for the qualitative interpretation of results. We also observe a significantly consistent hierarchy between the two- and three-cluster cases, quantified as 89% agreement for CONs, 90% for FUAs and 90% for LUZs (quantified by the number of pairs of cities being in the same cluster for clustering in three that also appear to be in the same cluster in case of two clusters). Therefore the distinction into three categories of cities is retained as the basic one for the presentation of results.

Spatial distribution of the clusters obtained for three city levels is presented in Figure 8. The most distinctive category, especially for CONs and FUAs, is the red cluster covering cities located along the most visited part of Mediterranean coast and on the Islands. This pattern clearly refers to the most touristic parts of Spain, and is further supported by the incorporation of Toledo - the World Heritage Site by UNESCO and important visiting hot-spot. Separation of this "touristic" cluster is interesting, as our procedure relies exclusively on the economic activity of city residents. It may indicate that this type of a city profile importantly affects the behavior of its inhabitants. Blue and green clusters originate from the split of one group from the two-categories case. The core of blue cluster includes cities from Basque and Navarra regions and Santiago the Compostela. In case of LUZs it covers also Madrid and Barcelona. In case of CONs it expands into much wider area in the North and partly also in the Spanish interior. Green cluster concentrates around southern cities and the remaining ones from the North. Spatial alignment of clusters exhibits a high robustness across three city definition levels, with a certain level of variations that can be to a large degree attributed to the changes of a sample size, although of course some specific issues related to the city spacial scale (such as mobility patterns within LUZs and CONs) could also matter. In general this serves as an additional evidence towards consistency and stability of the approach.

Interpretation of the received city categories is possible based on the variations of residuals of spending parameters across the Spanish cities (Figure 9). Distinction between the cluster is given mainly by the combination of spending activity and diversity, as well as distant mobility. First two parameters, which are in any case correlated as higher number of transactions fosters higher diversity, explain the separation of the red cluster. Deviations of distant mobility justify the split of two further categories - the green and the blue ones. Provided distinctions are consistent across the city levels and can be well observed on the scatter-plots provided in the SI.

The red cluster is characterized by the intensified spending activity and diversity, which are accompanied by the under-performance in terms of average purchase. This indicates that residents use their bank cards more often than just for occasional shopping, covering all types of small, everyday purchases. Negative deviations recorded for the distant mobility parameter are also well understandable, given the red cluster contains a group of naturally separated cities located on islands. Cities grouped within the green and blue clusters share common patterns of the negative residuals for the activity and diversity parameters, but the deviations of the remaining three parameters are of a different character. In the blue cluster the residuals of average purchase are overwhelmingly positive, which can point to the situation opposite than the one observed for the red cluster, i.e. bank card payments being relatively sparse, utilized for the sporadic bigger purchases. Furthermore, share of transactions executed far from the resident city is substantially higher than the expectation given by the scaling trend, which may also confirm the occasional character of a bank card usage. In case of the green cluster the average purchase values are close to the trend expectation. At the same time distant mobility records mainly negative deviations, especially at the level of CONs. Accompanied by the increased local mobility for CONs, this may indicate that the economic activity of residents is concentrated around their cities, which are satisfying most of their needs. Card purchases are less frequent than the baseline provided by a general trend, however at the price level well agreed with the expectations for a given city size.

Conclusions

In the present study, we explored the impact of different factors such as age, gender and place of residence on customer spending behavior, quantified by means of 5 different parameters: customer activity, average transaction amount, spending diversity and local/distant mobility.

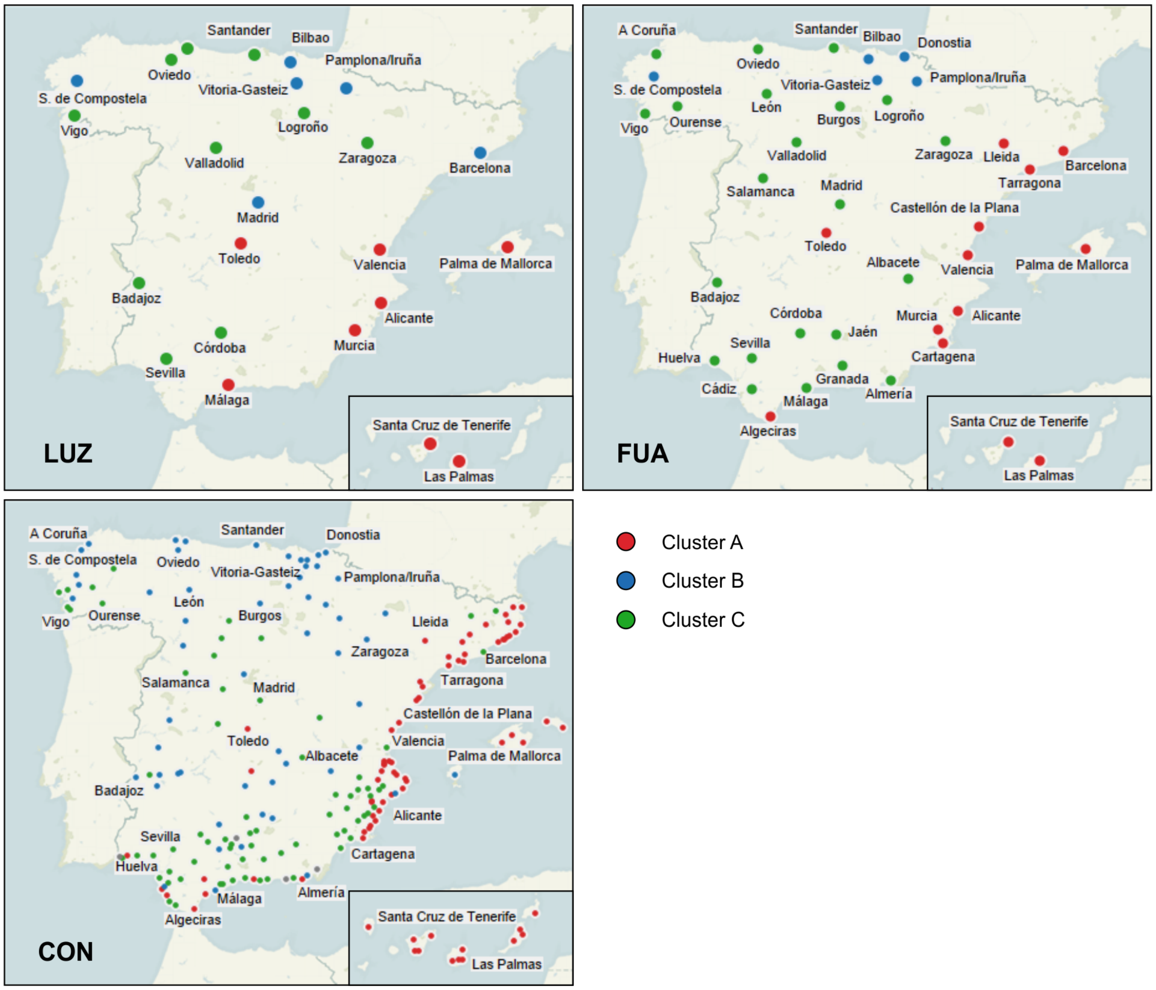

Figure 8: Classification of Spanish cities for all three city definitions - CONs (top left), LUZs (top right), FUAs (bottom right), based on the spending behavior of their residents into three categories. Bottom right map presents the correspondance between clusters obtained for different city levels.

First we found that age and gender have a major impact on spending behavior affecting all of the above parameters. Consistent trends were obtained when correlating them with age, and the curves for genders were different from one another, similar in general shape but not in absolute values. For instance, the average purchase amount demonstrates a surprisingly clear steady linear increase with customer age. This phenomenon might be interpreted as a general tendency to group purchases when growing old. Also the spending diversity after a certain increase until a peak of around 30 years of age (consistent with the average age of first childbirth), starts a steady and nearly linear decrease afterwards. Trends for customer activity and local/distant mobility look more complicated but can still be interpreted in an intuitive way. Moreover, looking beyond the simple average value of each characteristic, i.e. analyzing their distributions, we again recognized the steady and continuous impact of customer age and gender on their shape meaning that those parameters demonstrate a homogeneous impact on the entire sample of customers varying by their individual performance.

Next, we found that the size of the city of residence has a noticeable impact on all parameters but

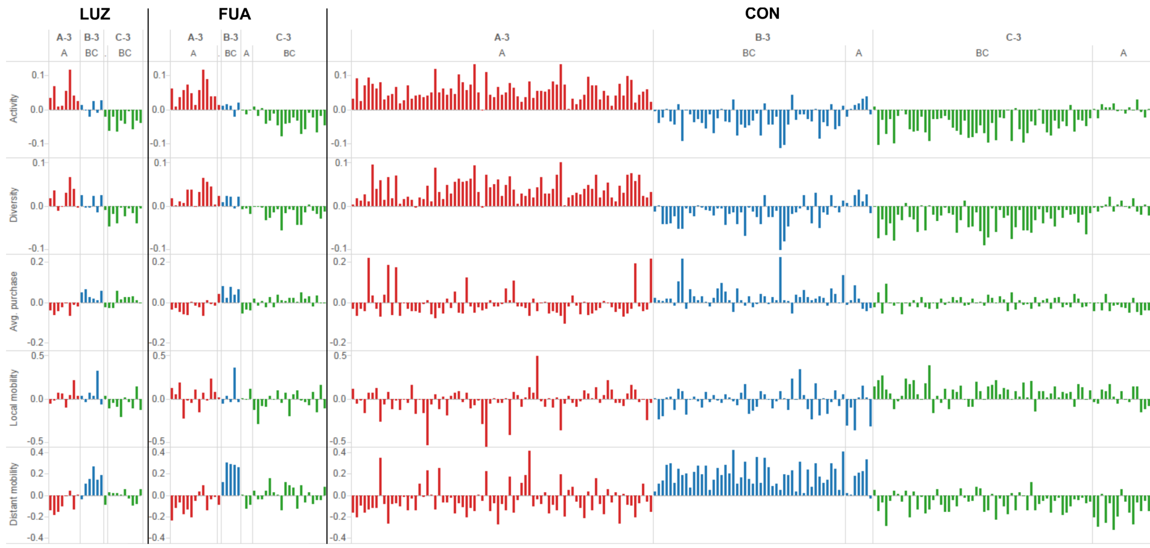

Figure 9: Deviations of the spending characteristics from their respective scaling trends with city size for the cities defined at the level of LUZ, FUA and CON. Colors indicate three clusters of cities obtained based on the clustering and presented in the figure 8).

the average amount of purchases in a way that often might be described as the statistically significant power-law scaling. However, even though a general trend exists when comparing parameters to city size, each city has its own, unique impact on the parameter values, beyond the general trend. Measuring this impact and regarding it as the city signature from the residents economic behavior viewpoint, we proposed a classification of the Spanish cities being defined at 3 different scales - Large Urban Zones, Functional Urban Areas and Conurbations. This purely data driven classification, independent from any spatial or topological consideration produced three clusters, that present a very clear geographical and socio-economic sense. Furthermore it remains stable to some extent under different ways of city definition which might serve as an additional evidence of results robustness. And although in the present paper the approach is demonstrated on the case of Spain we believe it is widely applicable to any other countries whenever appropriate data is available.

Acknowledgments

The authors would like to thank BBVA for providing the dataset for this research. Special thanks to Assaf Biderman, Marco Bressan, Elena Alfaro Martinez and María Hernández Rubio for organizational support of the project and stimulating discussions. We further thank the BBVA, MIT SMART Program, Center for Complex Engineering Systems (CCES) at KACST and MIT, the National Science Foundation,the MIT Portugal Program, the AT&T Foundation, The Coca Cola Company, Ericsson, Expo 2015, Ferrovial, The Regional Municipality of Wood Buffalo, Volkswagen Electronics Research Lab and all the members of the MIT Senseable City Lab Consortium for supporting the research. Part of this research was also funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience (DK W 1237-N23), Department of Geoinformatics - Z_GIS, University of Salzburg, Austria.

References

1. Girardin F, Calabrese F, Fiore FD, Ratti C, Blat J (2008) Digital footprinting: Uncovering tourists with user-generated content. Pervasive Computing, IEEE 7: 5276.

2. González M, Hidalgo C, Barabási AL (2008) Understanding individual human mobility patterns. Nature 453: 779–782.

3. Quercia D, Lathia N, Calabrese F, Di Lorenzo G, Crowcroft J (2010) Recommending social events from mobile phone location data. In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. pp. 971 -976. doi:10.1109/ICDM.2010.152.

4. Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, et al. (2013) Delineating geographical regions with networks of human interactions in an extensive set of countries. PloS ONE 8: e81707.

5. Amini A, Kung K, Kang C, Sobolevsky S, Ratti C (2014) The impact of social segregation on human mobility in developing and industrialized regions. EPJ Data Science 3: 6.

6. Szell M, Grauwin S, Ratti C (2014) Contraction of online response to major events. PloS ONE 9: e89052.

7. Frank MR, Mitchell L, Dodds PS, Danforth CM (2013) Happiness and the patterns of life: A study of geolocated tweets. Scientific Reports : 2625.

8. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, et al. (2014) Geo-located twitter as proxy for global mobility pattern. Cartography and Geographic Information Science 41: 260–271.

9. Bagchi M, White P (2005) The potential of public transport smart card data. Transport Policy 12: 464–474.

10. Lathia N, Quercia D, Crowcroft J (2012) The hidden image of the city: Sensing community well-being from urban mobility. In: Kay J, Lukowicz P, Tokuda H, Olivier P, Krüger A, editors, Pervasive Computing, volume 7319 of *Lecture Notes in Computer Science*. pp. 91–98.

11. Louail T, Lenormand M, Cantu Ros OG, Picornell M, Herranz R, et al. (2014) From mobile phone data to the spatial structure of cities. Scientific Reports 4: 5276.

12. Grauwin S, Sobolevsky S, Moritz S, Gódor I, Ratti C (2014) Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong. arXiv:14064400 .

13. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of many cities: Universal patterns in human urban mobility. Plos ONE 7: e37027.

14. Kung K, Greco K, Sobolevsky S, Ratti C (2014) Exploring universal patterns in human home/work commuting from mobile phone data. PLoS ONE 9: e96180.

15. Lloyd R, Jennings D (1978) Shopping behavior and income: Comparisons in an urban environment. Economic Geography 54: 157–167.

16. Childers TL, Car CL, Peck J, Carson S (2001) Hedonic and utilitarian motivations for online retail shopping behavior. Journal of Retailing 77: 511–535.

17. Dholakia RR (1999) Going shopping: key determinants of shopping behaviors and motivations. International Journal of Retail & Distribution Management 27: 154–165.

18. Buckinx W, Van den Poel D (2005) Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research 164: 252–268.

19. Bhantagar A, Misra S, Rao HR (2000) On risk, convenience, and internet shopping behavior. Communications of the ACM 43: 98–105.

20. Hui TK, Wan D (2007) Factors affecting internet shopping behaviour in singapore: gender and educational issues. International Journal of Consumer Studies 31: 310–316.

21. Fox EJ, Montgomery AL, Lodish LM (2004) Consumer shopping and spending across retail formats. International Journal of Retail & Distribution Management 77: S25–S60.

22. Roslow S, Li T, Nicholls J (2000) Impact of situational variables and demographic attributes in two seasons on purchase behaviour. European Journal of Marketing 34: 1167–1180.

23. Boeschoten WC (1998) Cash management, payment patterns and the demand for money. The Economist 146: 117–142.

24. Hayhoe CR, Leach LJ, Turner PR, Bruin MJ, Lawrence FC (2008) Differences in spending habits and credit use of college students. Journal of Consumer Affairs 34: 113–133.

25. Bounie D, Francois A (2006) Cash, check or bank card? The effects of transaction characteristics on the use of payment instruments. SSRN Scholarly Paper .

26. Borzekowski R, Kiser EK, Ahmed S (2008) Consumers's use of debit cards: Patterns, preferences, and price response. Journal of Money, Credit and Banking 40: 149–172.

27. Chan PK, Fan W, Prodromidis AL, Stolfo SJ (1999) Distributed data mining in credit card fraud detection. Intelligent Systems and their ApplicationsL (IEEE) 14: 67–74.

28. Rysman M (2007) An empirical analysis of payment card usage. The Journal of Industrial Economics 55: 1–36.

29. Krumme C, Llorente A, Cebrian M, Pentland A, Moro E (2013) The predictability of consumer visitation patterns. Scientific Reports 3: 1645.

30. Sobolevsky S, Sitko I, Des Combes RT, Hawelka B, Murillo Arias J, et al. (2014) Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. the case of residents and foreign visitors in spain. In: 3rd IEEE International Congress on BigData, Anchorage, AK, USA. IEEE, pp. 136-143.

31. Sobolevsky S, Sitko I, Grauwin S, Des Combes RT, Hawelka B, et al. (2014) Mining urban performance: Scale-independent classification of cities based on individual economic transactions. In: 2nd ASE International Conference on Big Data Science and Computing, May 27-31, 2014, Stanford University, USA. ASE, p. 10.

32. Batty M (2008) The size, scale, and shape of cities. Science 319: 769–771.

33. Bettencourt LM, Lobo J, Helbing D, Kühnert C, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Sciences 104: 7301–7306.

34. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439: 462–465.

35. Schläpfer M, Bettencourt L, Grauwin S, Raschke M, Claxton R, et al. (2014) The scaling of human interactions with city size. Journal of the Royal Society Interface 11: 20130789.

36. Bettencourt LM (2013) The origins of scaling in cities. Science 340: 1438–1441.

37. Bettencourt LM, Lobo J, Strumsky D, West GB (2010) Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. PLoS ONE 5: e13541.

38. City Statistics - Urban Audit. URL http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/c

39. (2007). Espon project 1.4.3: Study on urban functions: Final report. URL http://www.espon.eu/main/Menu_Projects/Menu_ESPON2006Projects/Menu_StudiesScientificSupportProje

40. Aŕeas Urbanas de Espanã. URL alarcos.esi.uclm.es/per/fruiz/audes.

41. Eurostat. URL ec.europa.eu/eurostat.

42. Nacional Estadística I. URL www.ine.es.

43. Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, et al. (2013) City boundaries and the universality of scaling laws. arXiv:13011674 .

44. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1: Statistics: 281–297.

45. Rousseeuw P (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20: 53–65.