Stanway Liau

804181960

cs35L week 10

3/11/2014

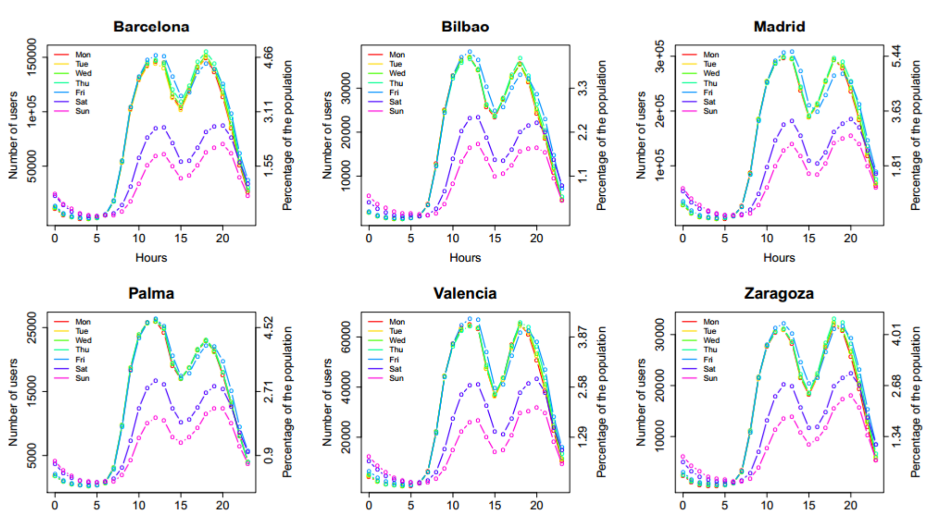
From Mobile Phone Data to the Spatial Structure of Cities

With the ever increasing amount of sensor data available to use in the world today also comes an increasing amount of potential for harnessing and analyzing this data. From our mobile phones to personal GPS, RFID cards, and social media, one asks the question: “Can this data be analyzed to learn about something and this knowledge applied to solve some problem?” A research team decided to do just that.

The research team analyzed anonymous mobile phone traces from cell towers for 55 days in 31 Spanish cities of varying geography – central, coastal, or island. Their goal in mind was to be able to detect individual mobility patterns, the location of communities, urban land use, and be able to characterize a city by its space consumption, landscape density, activity centers, and degree of polycentrism – to be able to define the spatial structure of a city quantitatively. The dynamic data that the phone traces provide was integral to this. Static data from a census would not have sufficed.

All of this information was used to develop a function p(i, t) – the density of users at location i at time t. This function provides a lot of information and the team took two approaches to analyzing it. One was a more “local” perspective that aimed to find the minima and maxima of the function to find hotspots. The second was a more “global” view that defined some city-wide parameter that takes a value dependent on data points throughout the city.

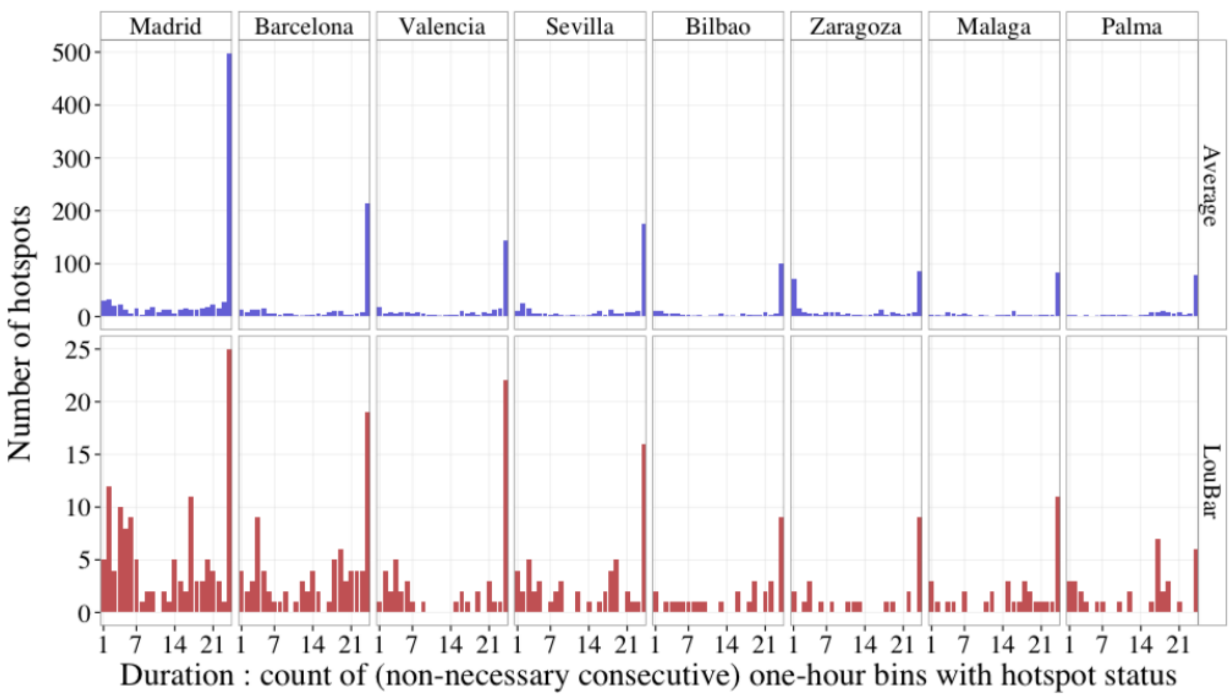
To see how busy a city is at different times of the day, the number of cell phone users was plotted against time. The results show that there was a “common rhythm” to cities. For all 31 cities, there were two peaks in activity at 12pm and 6pm. This was an example of one of the global indicators mentioned before. Here are the plots for six of these cities:



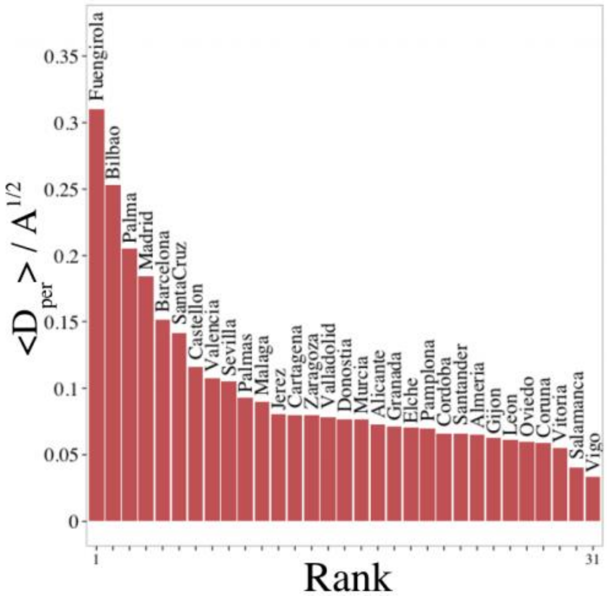
Another plot that was made was the physical size of the city vs. population. There was no correlation to this which surprised me. Just because a city is larger doesn't mean it must be dense. However, the percentage of the population using a cell phone vs. population size did have a linear fit, indicating that a constant ratio of a city's population owns a cell phone. This made some intuitive sense to me because of the law or large numbers in statistics.

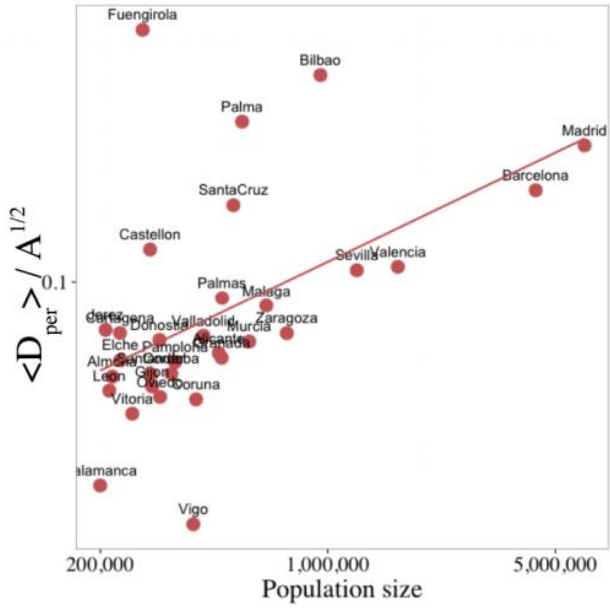
The average weighted distance D between cell phone users in the city can tell us about the city's organization. A dilation index was defined as D_max/D_min. An example of what the dilation index can tell us is a city with a lower dilation index would have more residential areas mixed with businesses as people would stay within the city to do most of their daily activities as opposed to a high dilation index indicating some city with a central business district.

The "local" data analysis mainly aimed to look at a city's hotspots. The city was divided into a grid of cells. If the population of a cell exceeds some set threshold limit, the city is said to be hot in activity. One question tackled was "how stable are the hotspots?" Each cell was monitored and given a tally for each 1-hour block it exceeded the threshold value. A cell with a higher score is more "stably" a hotspot. Here is a graph of the number of hotspots vs. the stability of each for 8 Spanish cities. Some hotspots are hot almost all day – permanent hotspots, the hearts of the city. Statistical methods showed that these permanent hotspots are very stable throughout time and space.

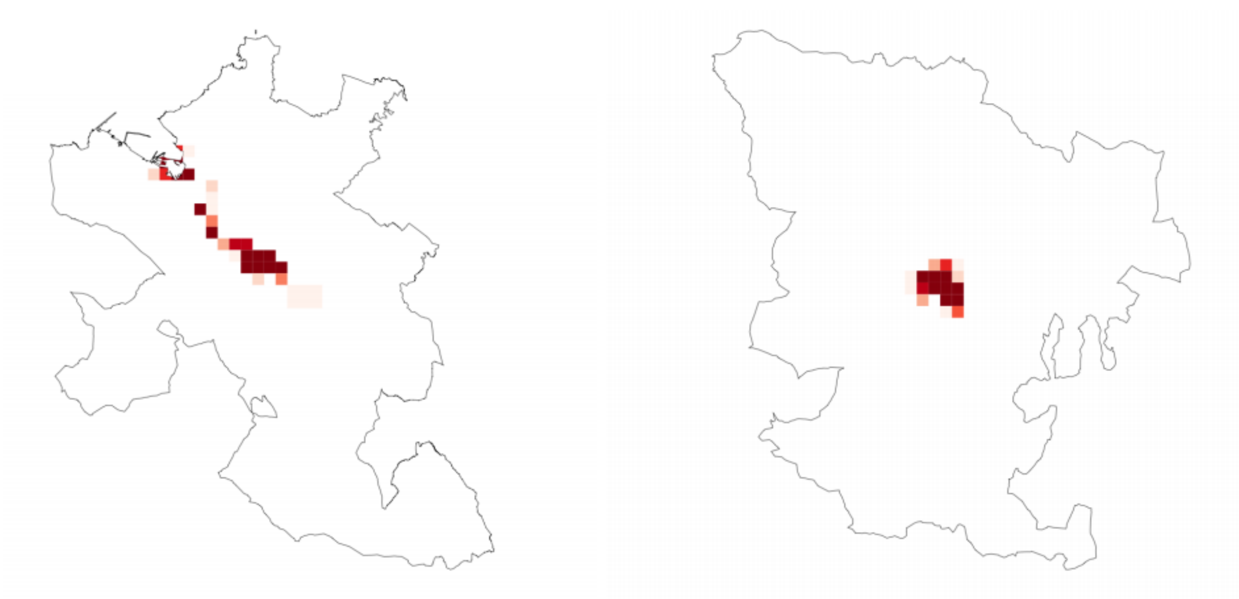


Finally, a metric was used to determine how spaced apart these hotspots were. A compacity value was defined as the average distance between hotspots divided by the square root of the physical area of the city. A value closer to 0 shows a tightly centered city (monocentric) while a value closer to 1 shows more sprawl (polycentric). The data backed up the idea that larger cities are more polycentric. In the real world, this makes sense as well because as there are more and more people, it is logistically difficult to cram them into small spaces and hence more hotspots are formed spread throughout the city. The following plots show that compacity generally increases with population size.





The following picture shows visually what a polycentric city (Bilbao on the left) and an monocentric city (Vitoria on the right) would look like depending on their hotspot distribution.



In conclusion, the researchers were able to peer into the spatial organization of a city using this wealth of data. It was found that one of the more meaningful metrics to define a city by is its organization of permanent vs. nonpermanent hotspots. By characterizing a city by these dynamic properties is a step towards a quantitative typology of cities.

I thought that this study was a good application of Big Data. We must find ways of parsing and analyzing increasingly large and formidable sets of data but the results can be satisfying and rewarding, if not practical. Examples of this in everyday life are when one shops on a website and personalized recommendations are given depending on browsing history. Google's search tries to find the most relevant information and suggest related search queries. Of course not all applications need be commercial. Some ways this "science of cities" might be applied is to see where a city is most busy and concentrate more resources towards infrastructure and maintenance in those areas. Also, analysis of traffic could be done on mobility patterns.

Sources:

http://arxiv.org/pdf/1401.4540v1.pdf

http://www.technologyreview.com/view/523926/how-a-new-science-of-cities-is-emerging-from-mobile-phone-data-analysis/