From mobile phone data to the spatial structures of cities

BY STANWAY LIAU

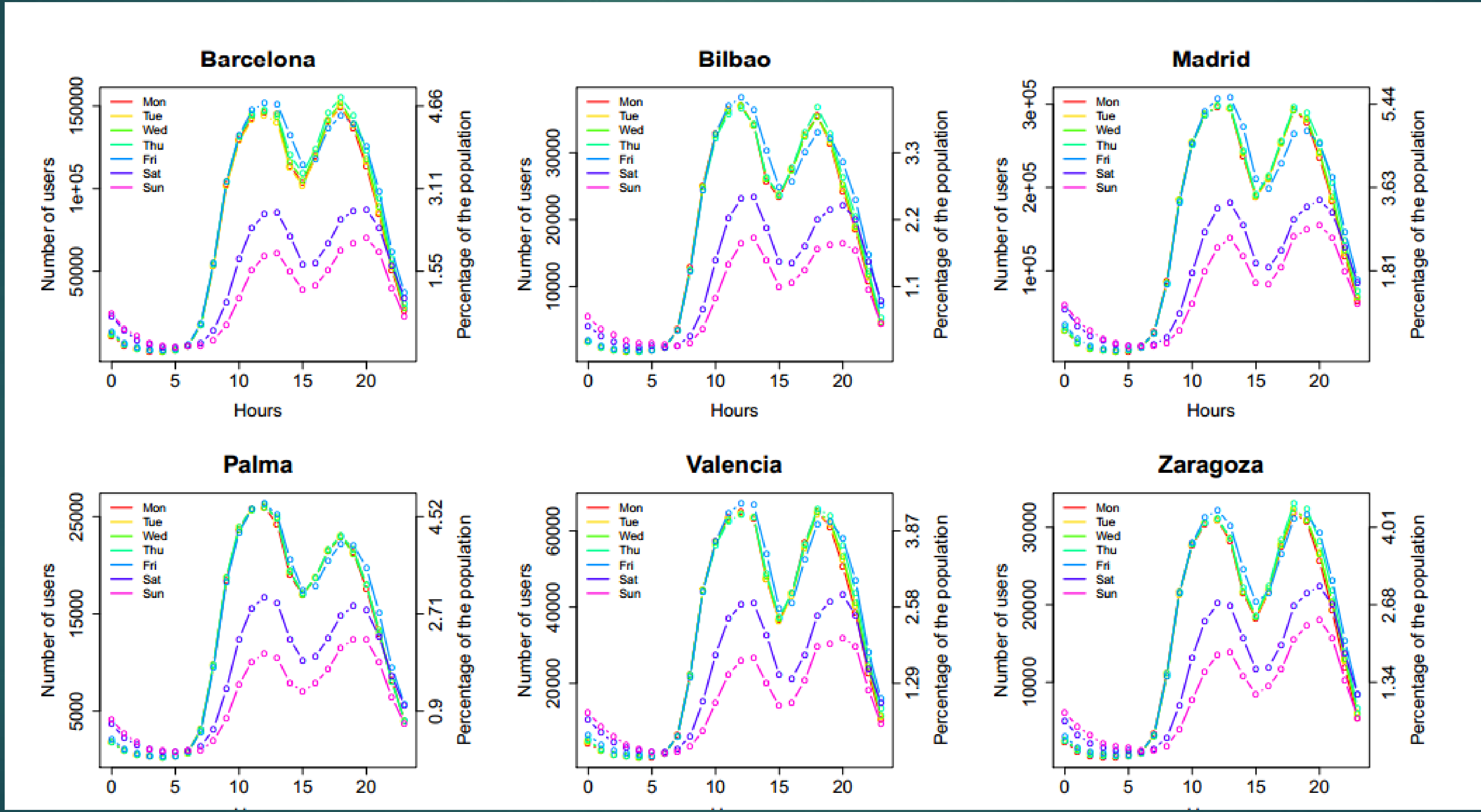
Introduction

- Analyzed anonymous mobile phone traces in 31 Spanish cities during weekdays for 55 days
- Detect individual mobility patterns, urban land use, and locations of communities
- Characterize city by landscape density, space consumption, activity centers, and degree of polycentrism
- Gives a dynamic analysis as opposed to using census data
  - E.g. changes throughout the day
- Goal is to characterize the spatial structure of cities quantitatively

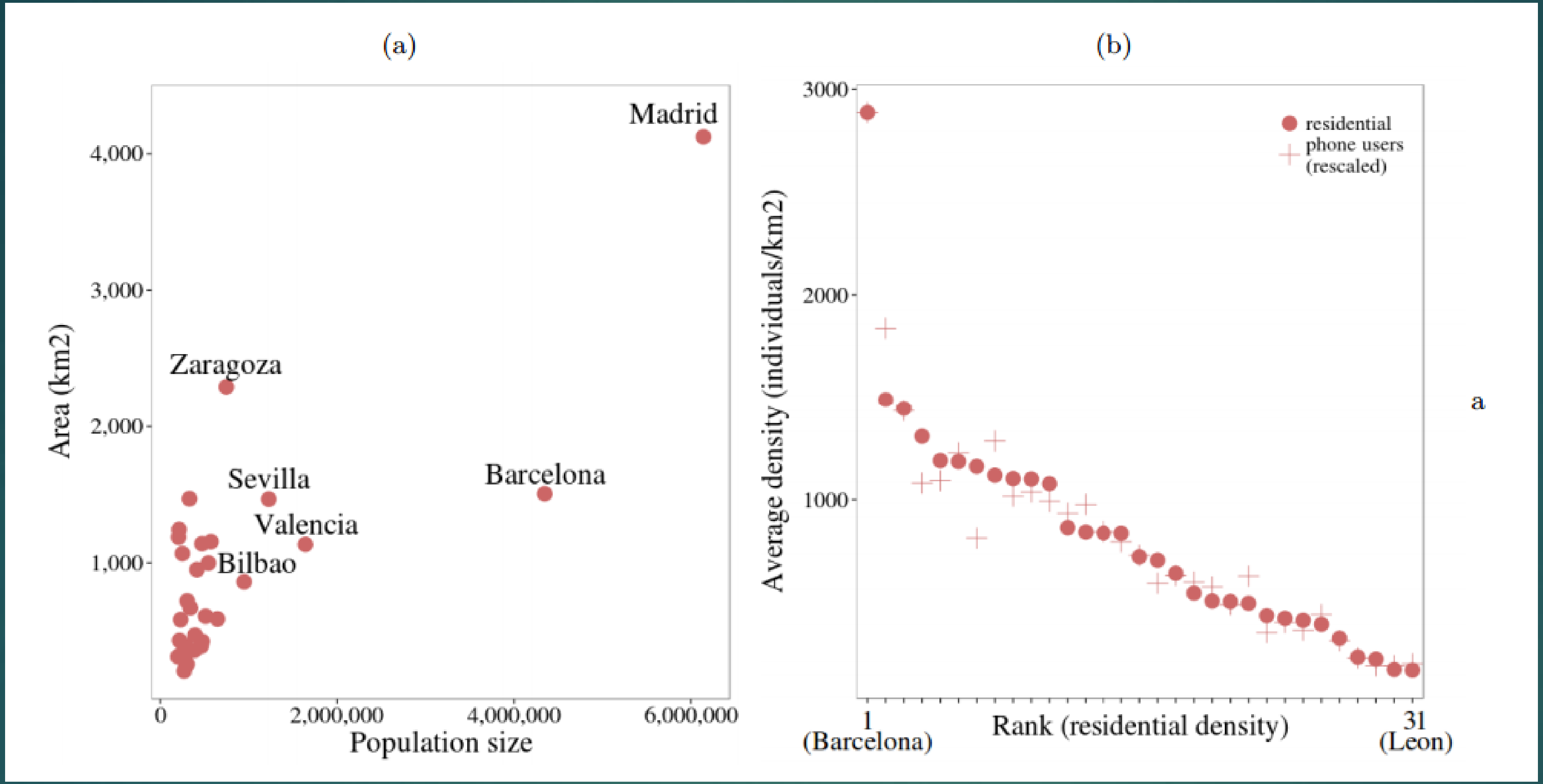
Collecting data

- Cities diverse in geography
  - E.g. port, island, or central cities
- Plot the time evolution of number of users along the day
  - Peaks at 12pm and 6pm in all cities
- Scale results by number of cell phone users to compare different cities
  - Results show a common thread

Number of users per hour per day



Users vs. population size



-No correlation between population size and physical size of city (left graph)

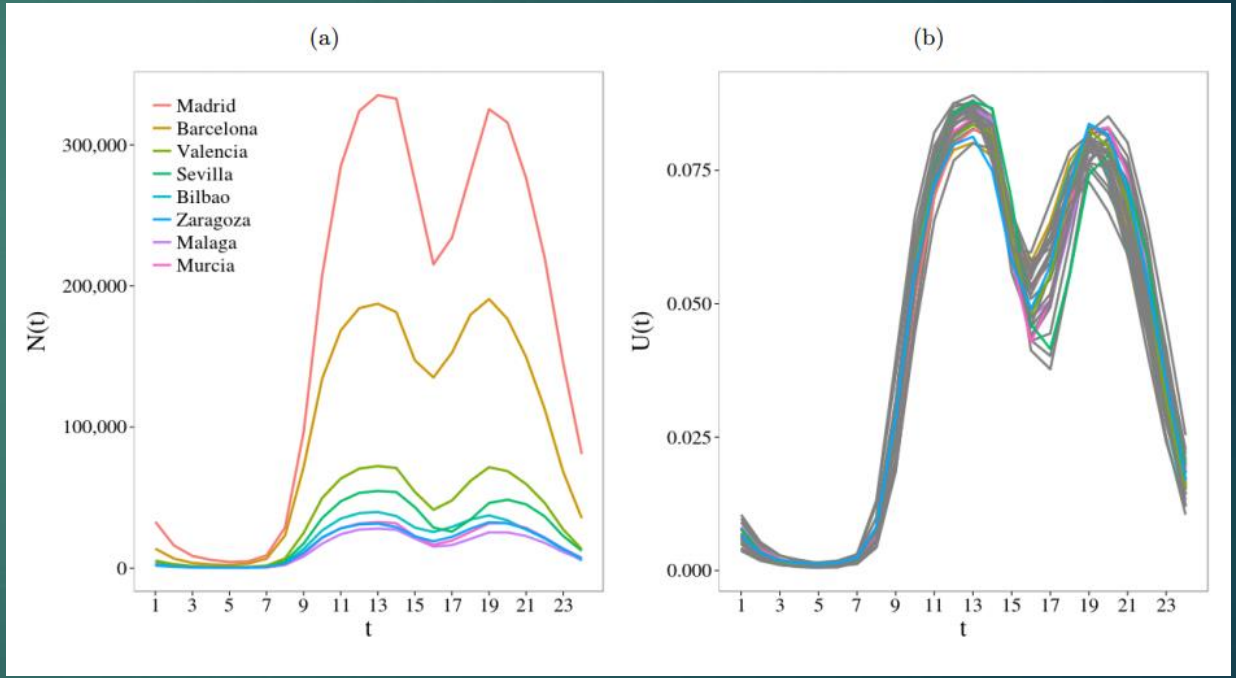
-However, the percentage of population using a phone compared to population size holds at a constant ratio (right graph)

Data analysis

- Mobile phone data gives the function $p(i, t)$ – the density of users at location i at time t
- Two ways of analyzing data:
  - Identify the local maxima and minima of $p(i, t)$ to find hotspots
    - Gives a more “local” look at different parts of the city
  - Define global indicators that consider all points and weigh them by user density
    - Informs us about the “global” properties of the city as a whole

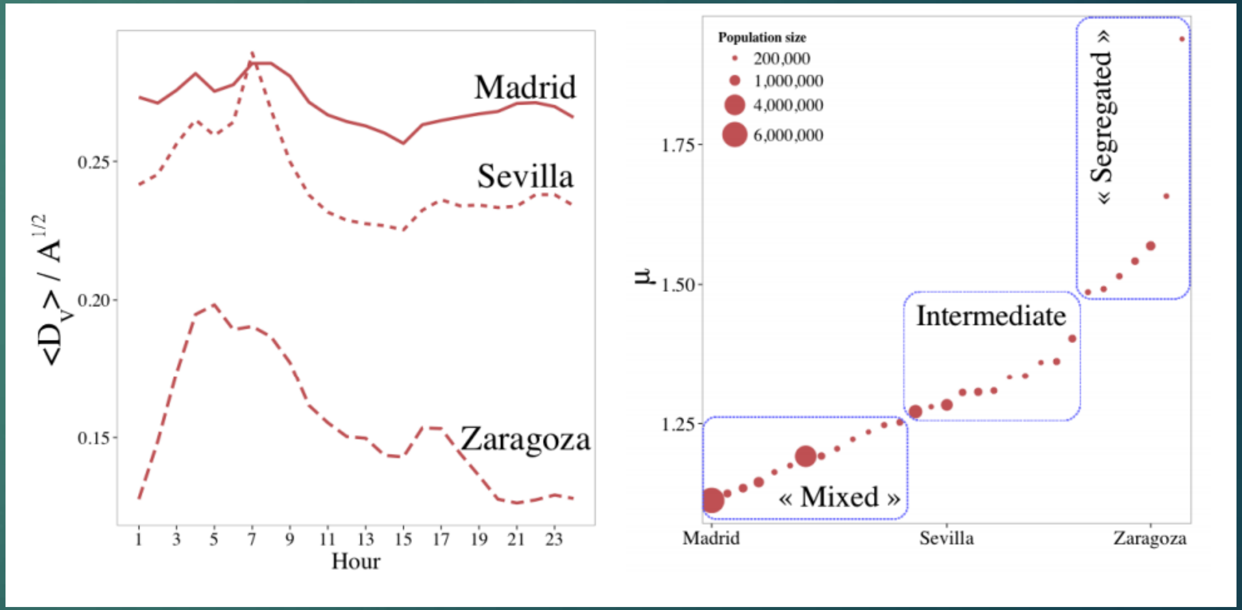
Global analysis

- Average cell phone users per hour during an average day for the 8 biggest cities (left graph)
- Same data but rescaled for all 31 cities (right graph)
  - Rescaled by dividing the number of users in city i at time t by the total number of users in city i during the entire day
  - There’s a common “rhythm” to all cities



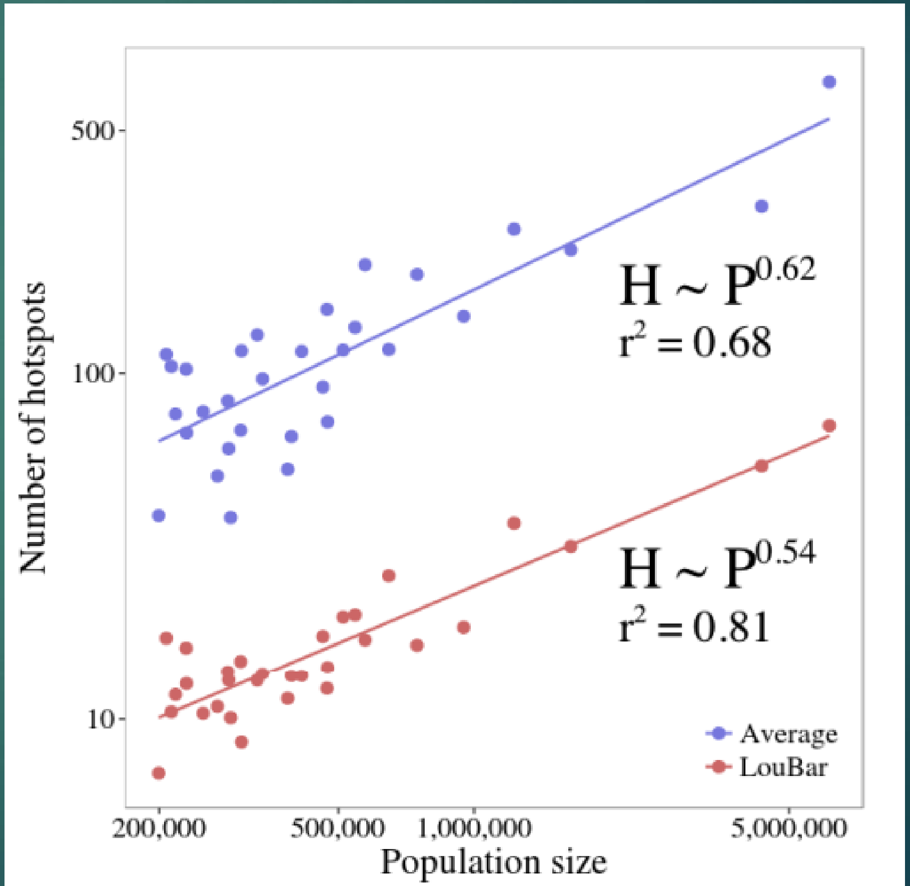
Density Analysis

- The average weighted distance between users, D, tells a lot about the organization of a city
  - Normalized for density of each cell
- Data normalized for the location of the scattered locations of the cell towers
- People cluster together at different times of the day (left graph)
- Dilation index u defined as D_max/D_min
  - Plotted on right graph
  - Tells about how "spatially segregated" a city is
  - Lower value corresponds to businesses being mixed with residential areas



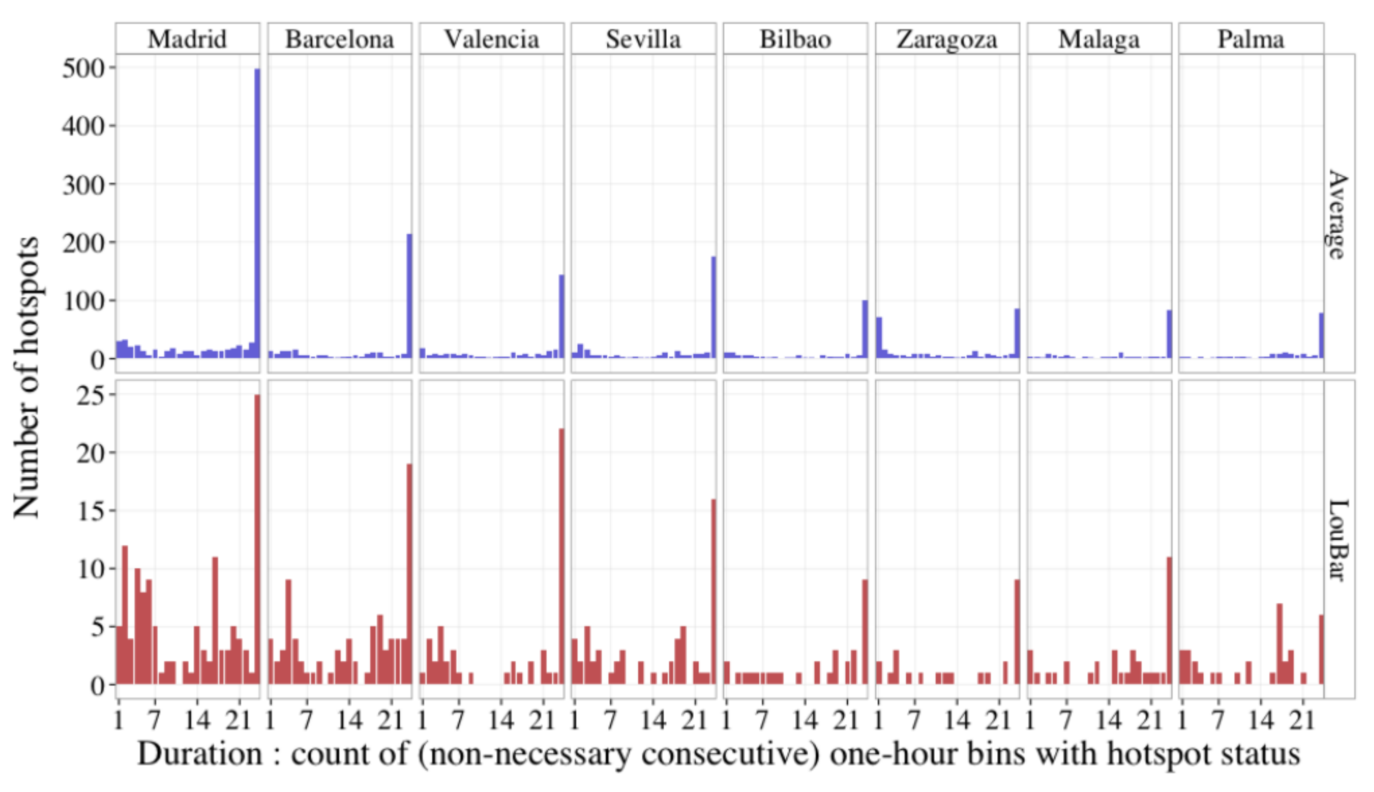
Hotspot analysis

- Determine some threshold value that when crossed, indicates above average traffic
  - d_min: average value of traffic
  - d_max: reasonable max value using the LouBar method
- Regardless of choice of threshold, results should be robust, indicating some intrinsic property of the city
- Shown that number of activity centers N_a in the US scales as N_a ~ P^b, where b = 0.64
  - Exponent is affected by choice of spatial boundaries
  - "Urban areas" should fit the pattern while official city boundaries may not



Hotspots stability

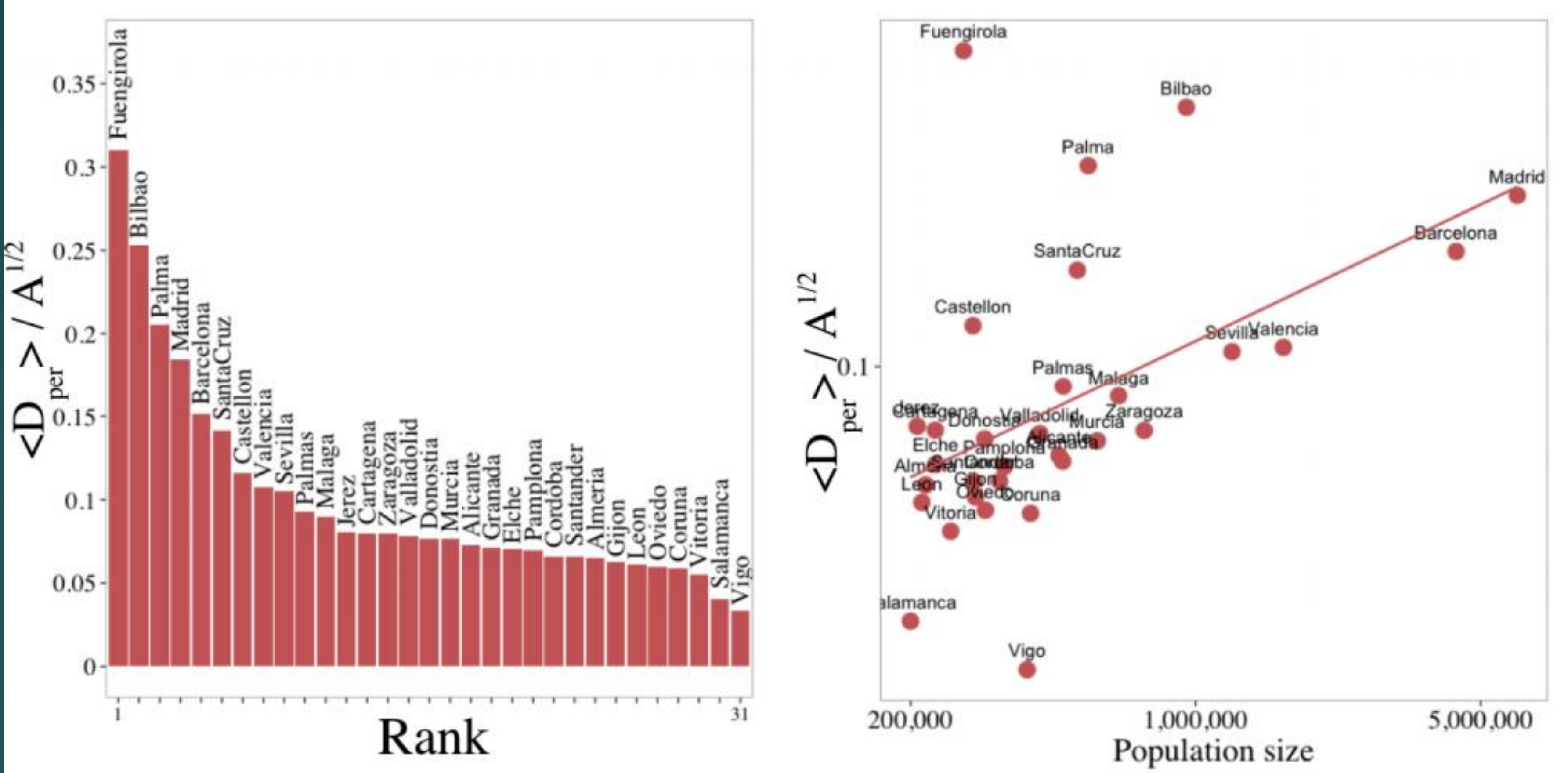
- Inspect the stability of the hotspots and their evolution throughout the day
- Divide a city into cells and count the number of one hour periods that that cell has been a hotspot
  - Permanent, intermediate, and intermittent hotspots
  - Results robust for whichever threshold value chosen
- Kendall tau value T(t) for permanent hotspots shows they are very stable throughout time and space



Spatial structure of hotspots

- Look at permanent hotspots from LouBar criteria and compare how distant they are from each other, D_per(i), compared to the city's area A_i
  - Compacity value $C(i) = \left(\frac{D_{per(i)}}{A_i^{0.5}}\right)$
- Shows how compact a city's core is by seeing how hotspots are distributed
  - Values close to 0 indicate a tight central core/hub
  - Values close to 1 indicate more sprawl
- Compacity increase with population size
  - Consistent with idea that a larger city is more polycentric

Compacity for 31 cities

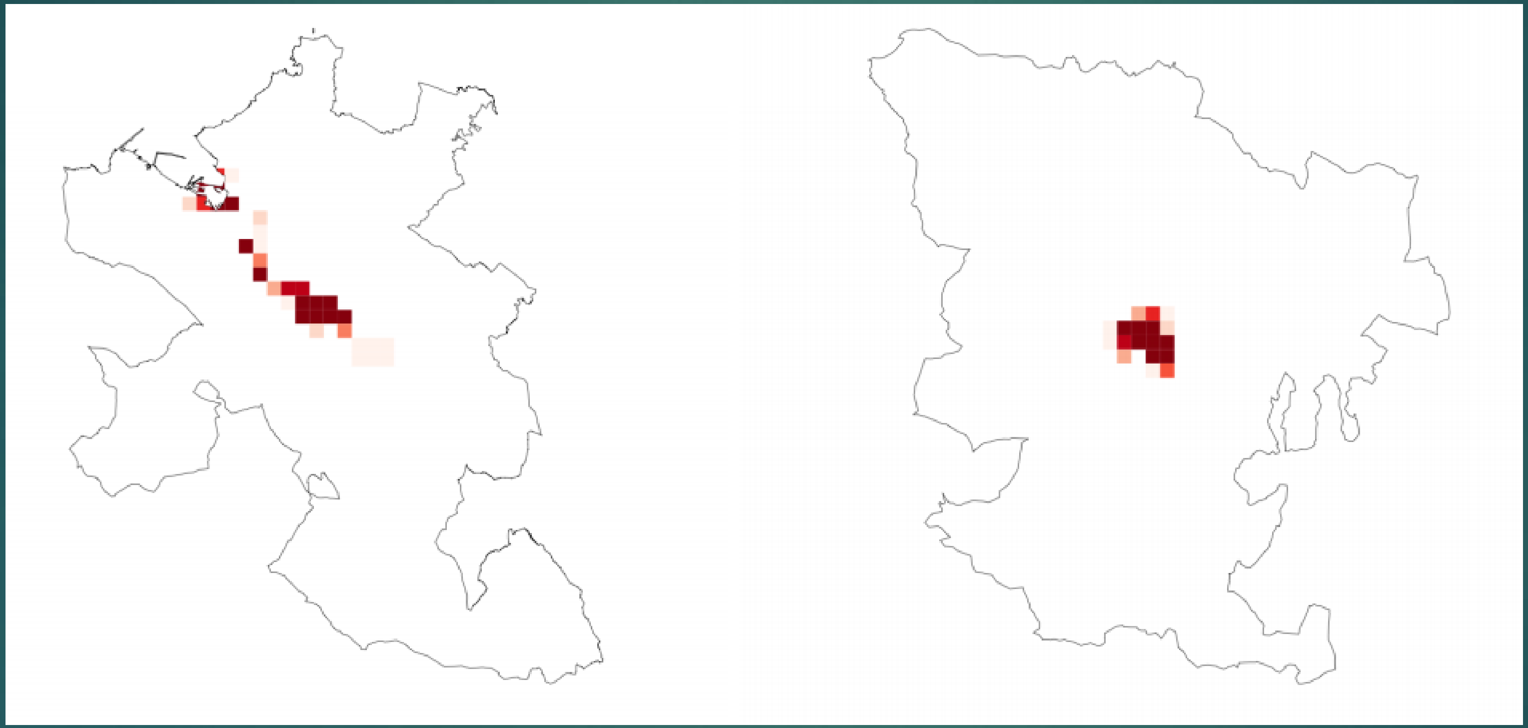


-Can see that compacity generally decreases with population

Hotspot distribution example

Bilbao (population 950,000)
-polycentric

Vitoria (population 250,000)
-monocentric



-Spatial distribution of 1 km^2 hotspots in Bilbao and Vitoria

Conclusion

- Researchers found that the intermittent and intermediate hotspots have a high compacity value, meaning they are pretty much spread out everywhere
  - So more significant to look at differences in spatial organization of permanent vs. nonpermanent hotspots
- Possible to characterize a city based on its dynamical properties using the indices in the study
  - Step towards a quantitative typology of cities
- Possible applications are where to focus efforts on infrastructure, traffic analysis
- Study shows a good application of big data

References

- 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/

Q&A