Scaling: lost in the smog

- **M. Barthelemy**
- **2014, Environment and Planning B: Planning and Design**
- **Link to paper**

General:
This paper is a commentary on the use of power-law distributions to learn about the science of cities.

Major Takeaway:
Scaling relationships, and more generally data analysis, have an important role to play in the rising new science of cities. However, it is dangerous to interpret empirical results without any mechanistic insight. **Conclusions cannot safely be draw from data analysis alone.**

Details:

- The growing availability of data allows the science of cities to take an empirically grounded footing.
- For example, there has been a recent growth of "scaling laws" which present as power-law relationships between socioeconomic (GDP, # of patents) or structural quantities *Y (i.e. length of roads or cables)*, and the size of the population *P in the city:*
    - $Y \sim P^{\beta}$

    Where the exponent β can be different from 1.

- As these scaling laws grow in number, it is unclear what we are really learning from them.
- Mechanistic insights about where these scaling laws are coming from are often nonexistent – which leads to misguided interpretations.
- Barthelemy then points to a few papers which try to answer the question:
    - "*Are larger cities greener – in the sense that there are fewer emissions per capita for larger cities – or smoggier?*"
    - These papers all provided contradictory evidence and Barthelemy hopes to show why this might be.

Two Main Sources of Error:

- 1. Being unable to measure the true quantity of CO_2 emissions, they all use some sort of proxy.
    - Using the wrong proxy (for example, the average *distance* traveled to work versus the average *amount of time in traffic* [the latter is correct btw]) is going to affect the estimation and negative affect the "observed" measurements
- 2. How a research defines a city itself is a major source of error.
    - **There is confusion within the field, and this is crucial as scaling exponents are very sensitive to the definition of the city**

An Example of How this Problem May Present Itself

Barthelemy then collects his own data to answer the above CO^2 question and illustrates how the definition of the city can lead to contradictory findings.

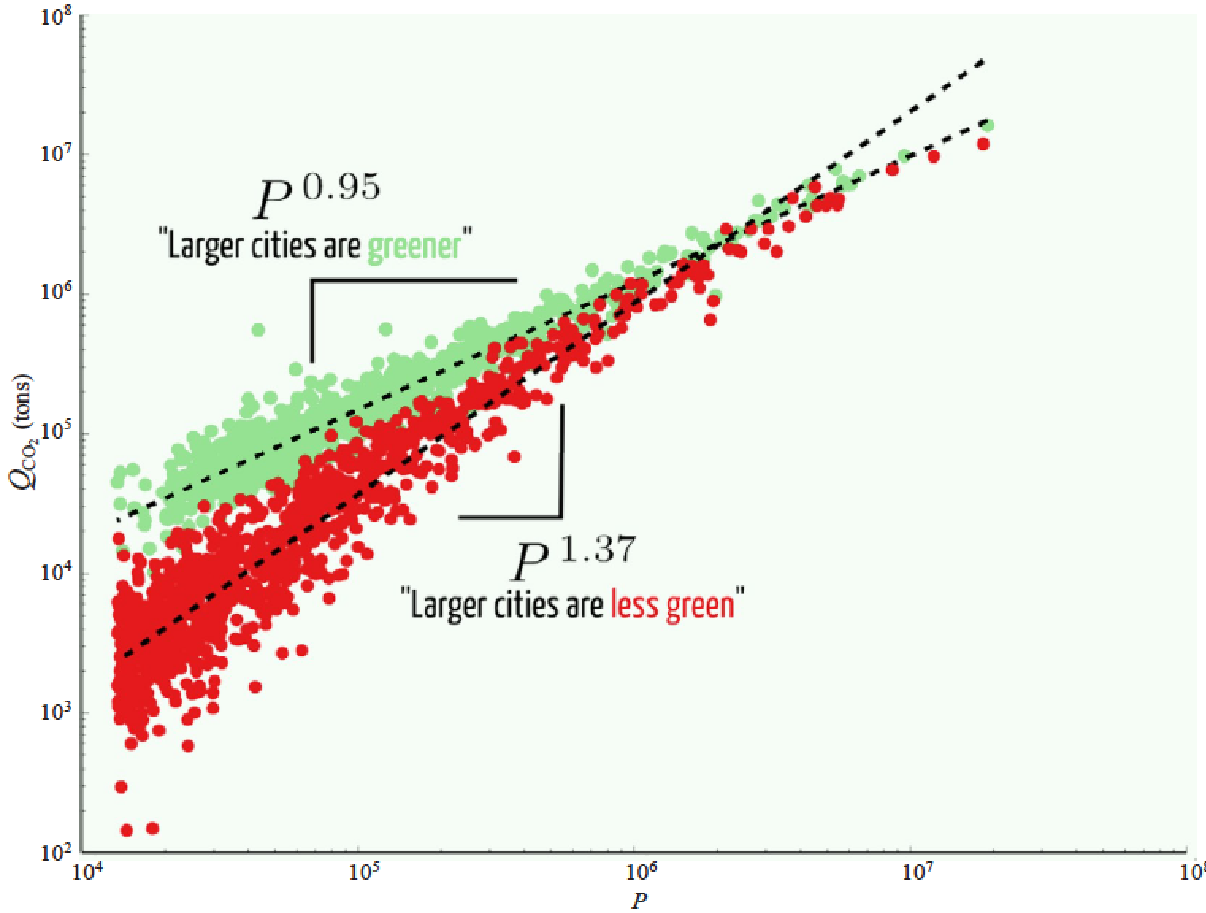

Figure 1. [In color online.] Are larger cities greener or smoggier? Scaling of transport-related CO_2 emissions with the population size for US cities from the same dataset but at different aggregation levels. In red, the aggregation is done at the level of urban areas and in green for combined statistical areas. Depending on the definition of the city, the scaling exponents are qualitatively different, leading to two opposite conclusions. Data on CO_2 emissions were obtained from the Vulcan Project (http://vulcan.project.asu.ed) (see Fragkias, 2013; Oliveira, 2014). Data on the population of urban areas and metropolitan statistical areas were obtained from the Census Bureau (http://www.census.org).

Faced with these two opposite results, what should one conclude? Our point is that, in the absence of a convincing model that accounts for these differences and how they arise, nothing. Scaling relationships, and more generally data analysis, have an important role to play in the rising new science of cities. But, as the previous discussion illustrates, it is dangerous to interpret empirical results without any mechanistic insight. Conclusions cannot safely be drawn from data analysis alone.