Presentation Linkedin data 4 competitive cities

⚔

with xaringan

Luisa M. Mimmi

OUTLINE

1. Conceptual Framework

2. Data exploration

3. Further data exploration by city

4. Bivariate measures of association - categorical variables across cites

5. Zooming in on domestic migration

6. Relevant policy questions and potential follow-up

7. Reference

Conceptual Framework

- A city's **productivity** can be simply defined as the net result of contrasting forces

$$Productivity = Agglomeration(positive) - Congestion(negative)$$

- Both $Agglomeration$ and $Congestion$ could be modelled as a function of various dimensions $= f(Skills, Amenities, Form, Access)$

- Dimensions
  - **Skills**, a city's aggregate stock of human capital
  - **Amenities** - *attracting 'skills'* - job opportunities, housing values, cultural attractions
  - **Form**, the size and spatial configuration of a city (*density* vs. *sprawl, wider metropolitan areas*)
  - **Access**, a city's connectedness (or barriers) to other cities, both at home and abroad, through the transportation network

[1] This comparative analysis will focus on the first dimension **Skills**, but keeping in mind also the other key dimensions that can explain the competitiveness of a city

Data Exploration

Metadata on the LinkedIn dataset

- Data contain *347 origin metro areas* and only *282 destinations*
- All Linkedin members (in data) have some tertiary education degree
- Linkedin members (in data) are distributed across *10 industry sectors* (plus 5 members with *unknown* level)
  - About 50% of them have an *Entry level* job -> likely, active members are also people who look for jobs
- Linkedin members (in data) are distributed across *14 industry sectors*
  - Highest % in *Government/Education/Non-profit*, followed by *Technology - Software*, *Healthcare & Pharmaceutical*, and *Professional Services*
- Linkedin members (in data) are distributed across *26 job positions*

Distribution of LinkedIn Members' *HIGHEST DEGREE* across countries of origin - table

both[, 7]	both[, 2]	n	freq	rel.freq
Australia	associate	255	0.0286452	2.86%
Australia	bachelor	5632	0.6326668	63.27%
Australia	doctor	611	0.0686363	6.86%
Australia	master	2404	0.2700517	27.01%
United Kingdom	associate	1312	0.0358323	3.58%
United Kingdom	bachelor	22107	0.6037689	60.38%
United Kingdom	doctor	3533	0.0964905	9.65%
United Kingdom	master	9663	0.2639082	26.39%
United States	associate	79505	0.1849818	18.5%
United States	bachelor	225213	0.5239961	52.4%
United States	doctor	36358	0.0845930	8.46%
United States	master	88723	0.2064291	20.64%

Distribution of LinkedIn Members' *HIGHEST DEGREE* across countries of origin - Main findings

- The US has a significantly higher percentage of *Associates* migrants (18% vs 3% UK and 4% AUS)

  - **Question:** Is it due to a different education system in America?

- Conversely, the proportions across country of *SENIORITY / SECTOR / POSITION* appear remarkably similar across coountries

  - **Question:** Are this similarities due to a certain selection bias of the sample? Or, possibly mean that the three English speaking counties have homogenous job market and working populations?

Top 10 *DESTINATION* and *ORIGIN* cities (mostly US)
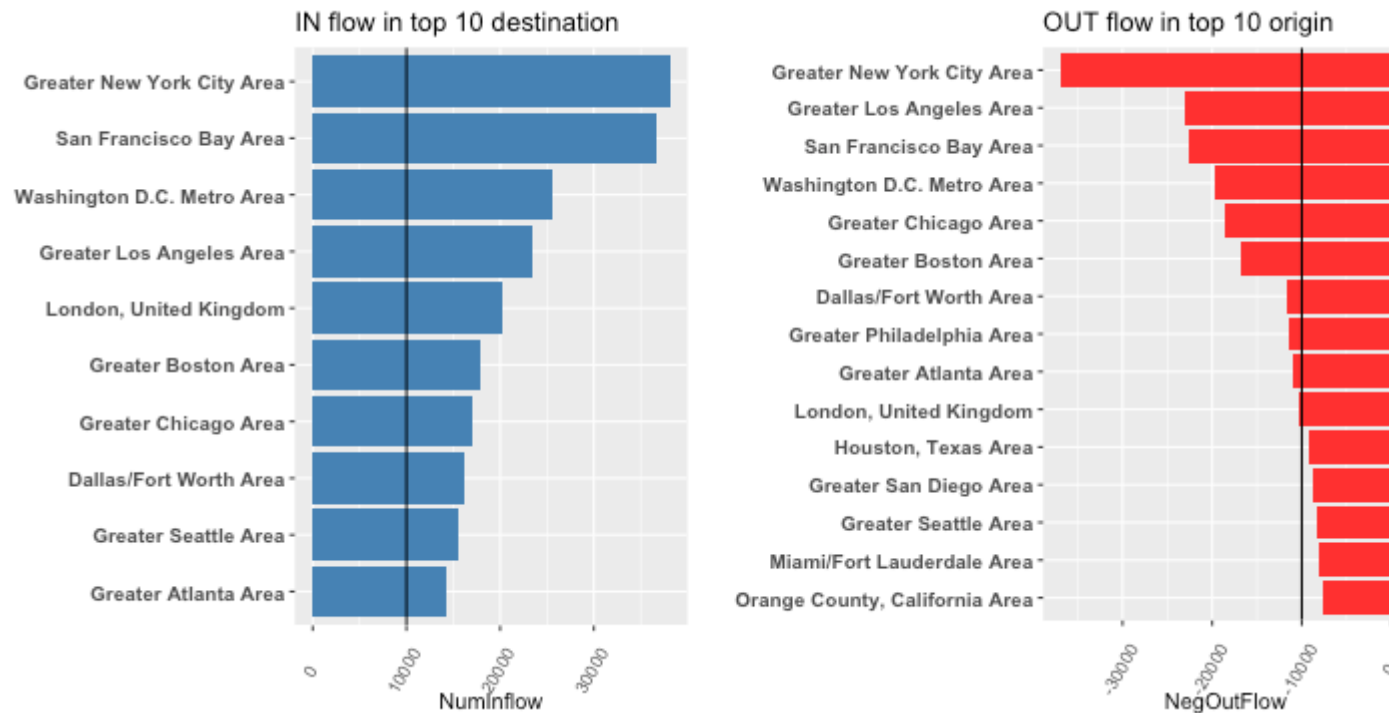
- Understandably, these are mostly American cities, although London is # 5 Destination and # 10 Origin for migrants
- Interestingly, 9/10 of top *destinations* are the same as top *origin* cities
  - There may be high inter-city *mobility*, but not necessarily other dimensions (*Amenities, Access*) of competitiveness, such that people decide to stay
 - Or, it may just be the American way of moving to and from college!

What about the *NET* migration flow to & from Top 20 cities?

NET migration flow in first 20 destination
(ordered by destination)
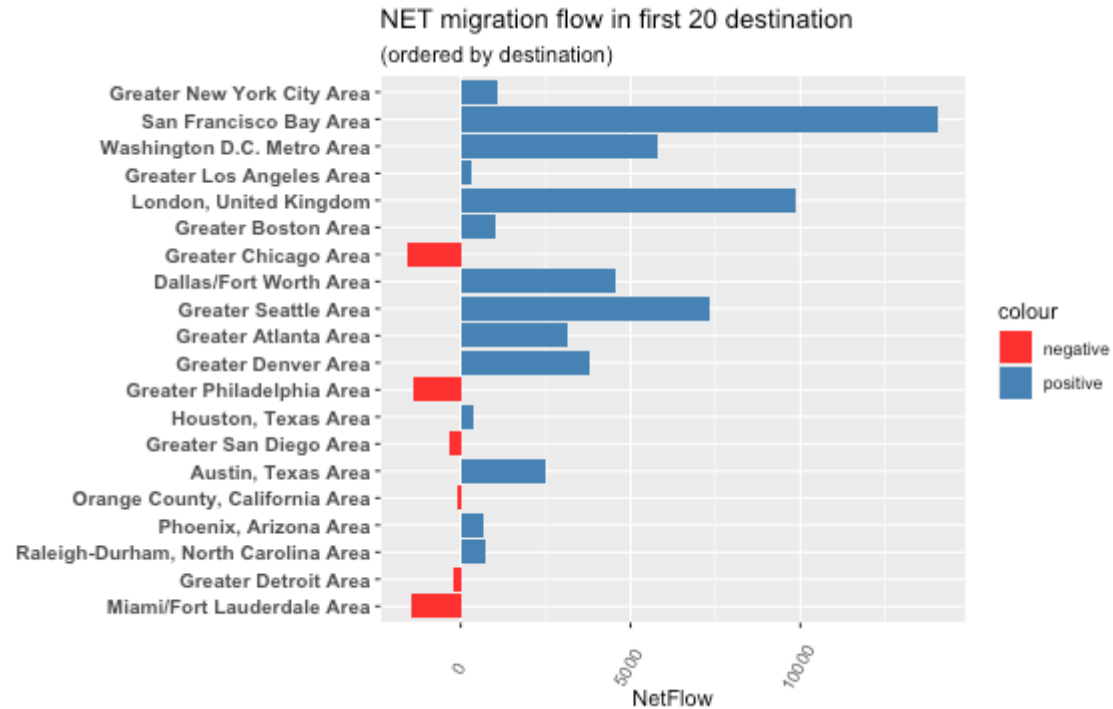
- Overall, most of the top American destination cities maintain a strong positive net flow of migrants.
- Chicago, Philadelphia and Miami had a negative net flow over 2016
  - Given Miami's position as a main entry port for Latin American migrants, this is a bit hard to believe. Possibly suggests non-representativity of Linkedin dataset for capturing the whole migrant population
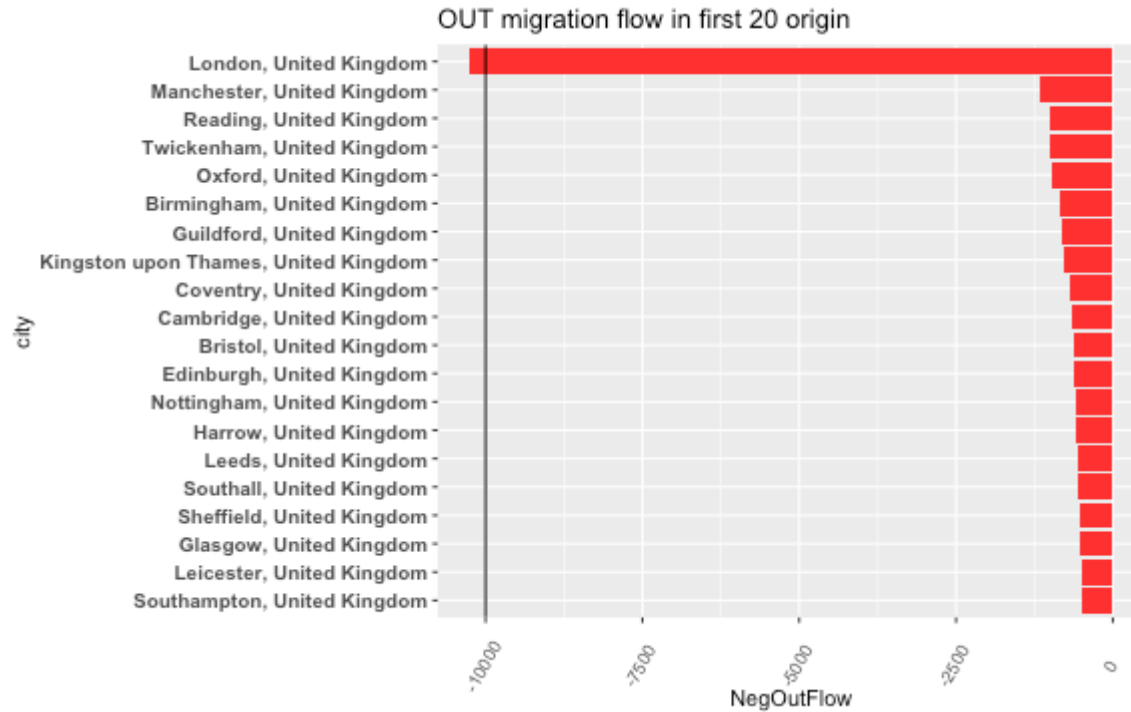
Top 20 DESTINATION cities for migrants in UK

IN migration flow in first 20 destination
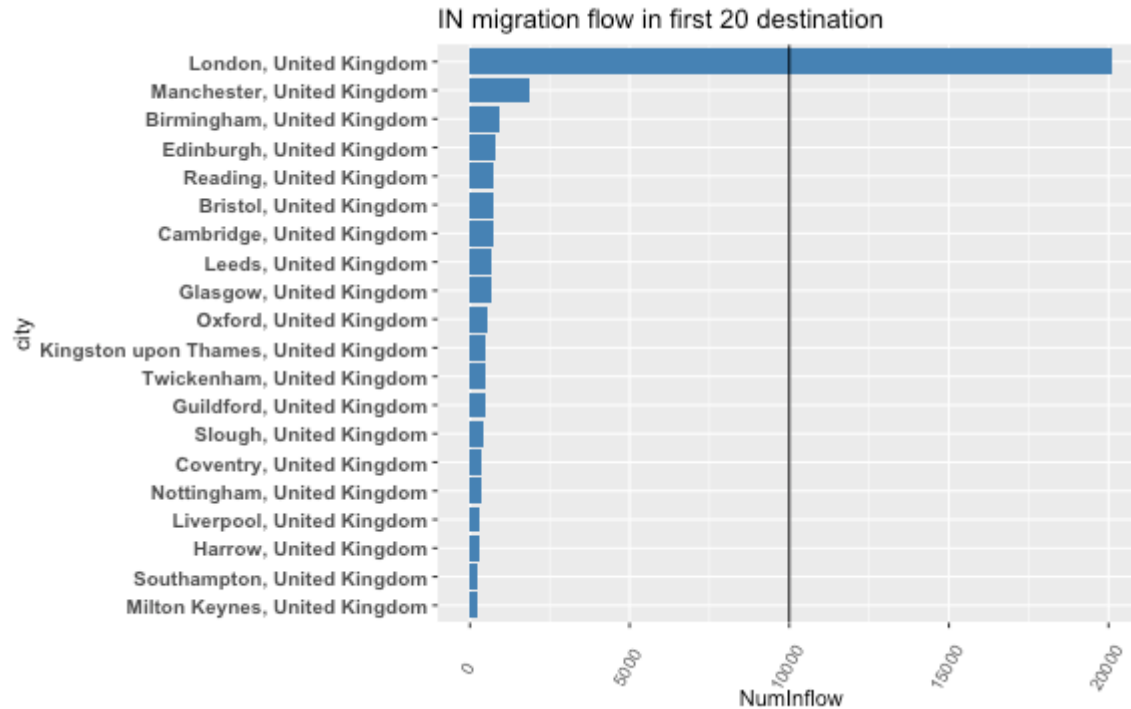
- Contrary to the USA's generalized high mobility across the whole country, in the UK, London is definitively an outlier
- This makes sense in light of London special role as European *hub*, at least for financial and other sectors
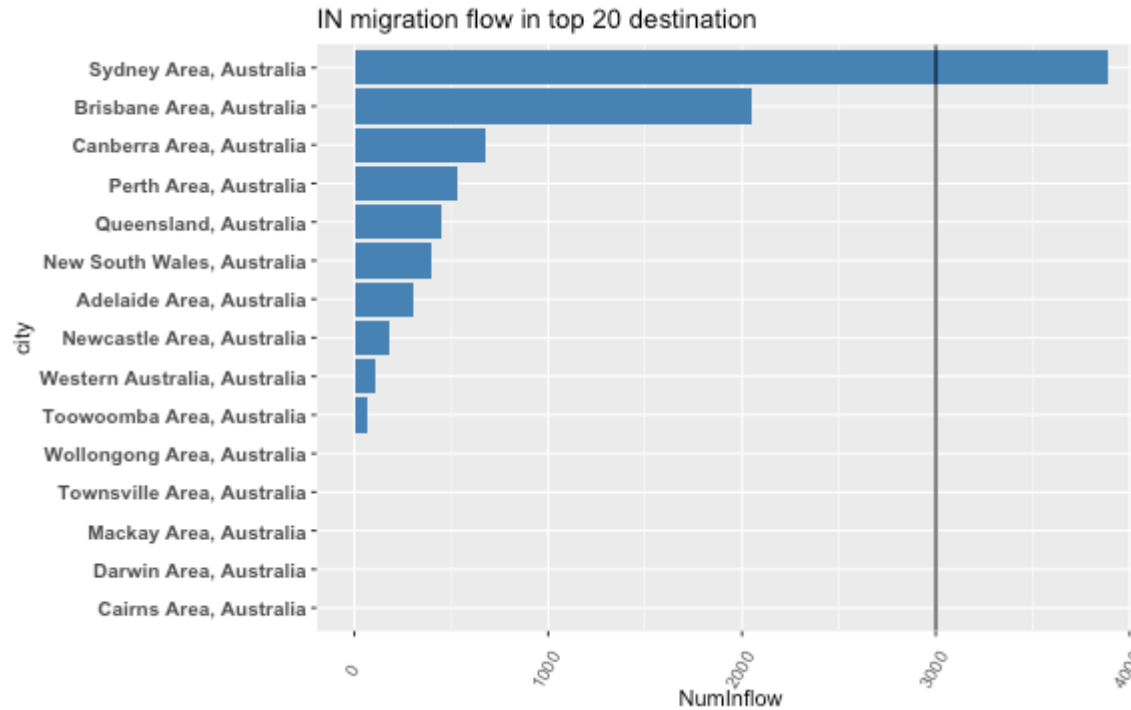
Top 20 ORIGIN cities for migrants in UK

OUT migration flow in first 20 origin

- Yet, similar to the top US destination cities, London is also a place from which many people leave
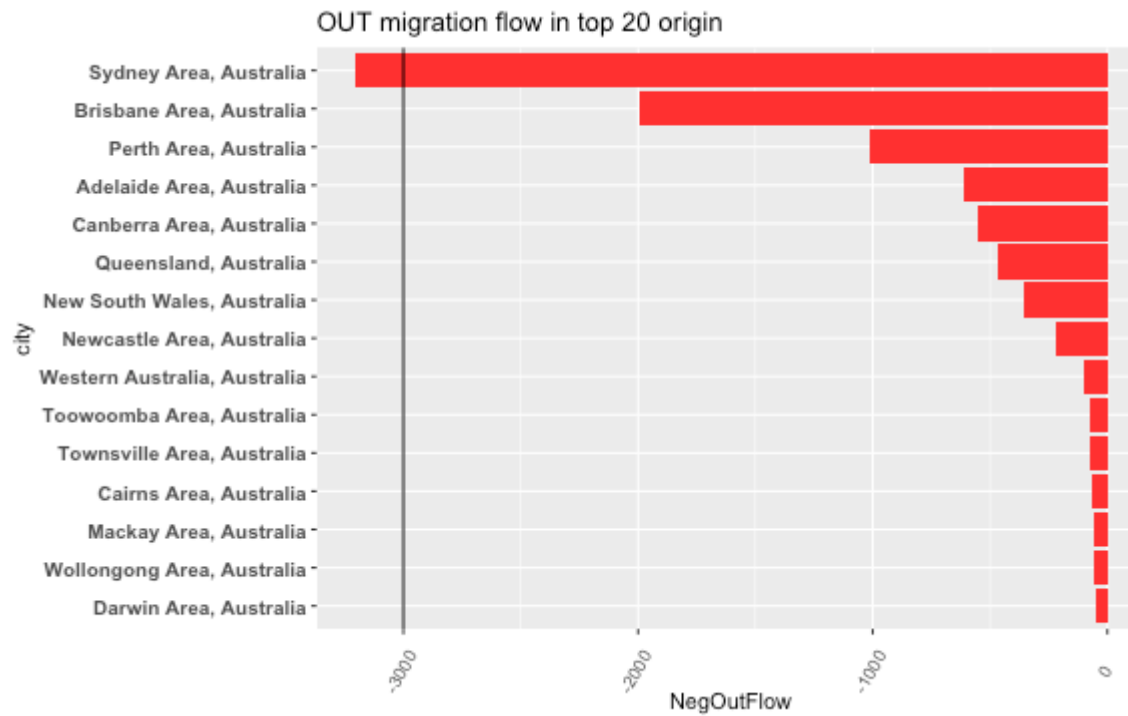
Net migrants flows in top 20 ORIGIN cities in UK

NET migration flow in first 20 destination
(ordered by destination)
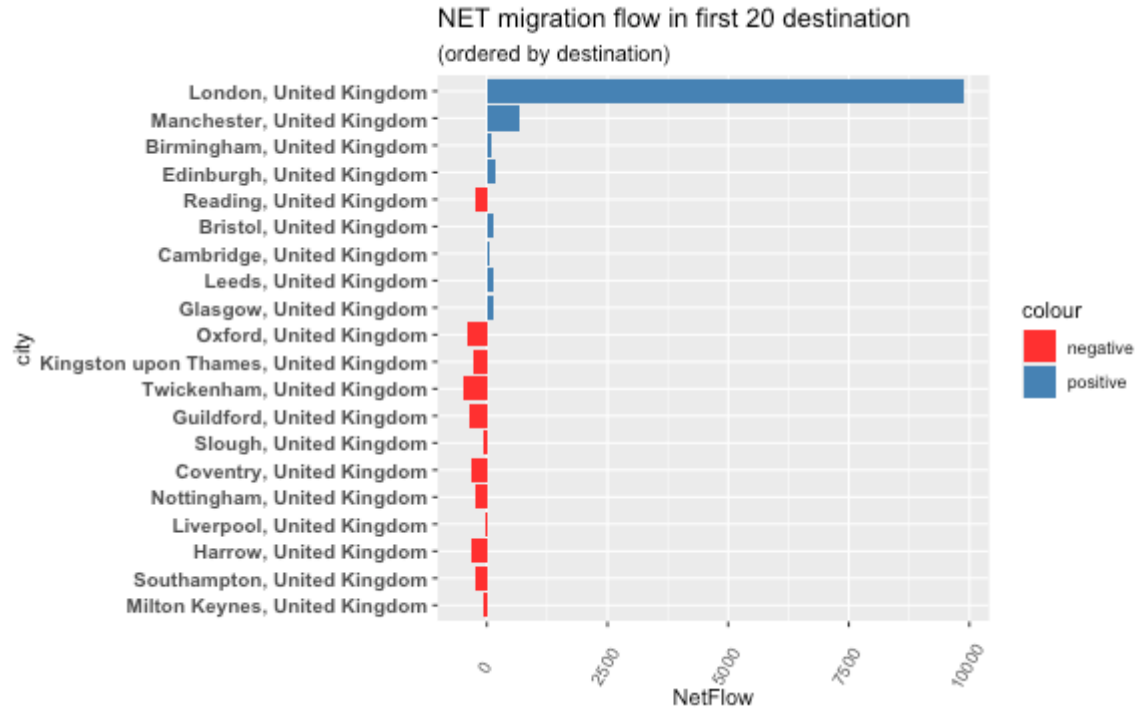
- In 2016, London had a net positive flow of migrants (followed at a distance by Manchester), while the other British cities where mostly loosing citizens

Top 20 DESTINATION cities for migrants in AUSTRALIA

IN migration flow in top 20 destination

- Evidently the magnitude of migration flows is much smaller in Australia
  - While I don't have the population sizes for comparison, it would seem Australian LinkedIn members move less than their counter-parts

Top 20 ORIGIN cities for migrants in AUSTRALIA

OUT migration flow in top 20 origin

- In 2016, migrants were also leaving from most cities in great numbers

Net migration flows in top 20 ORIGIN cities in AUSTRALIA

NET migration flow in top 20 destination
(ordered by destination)
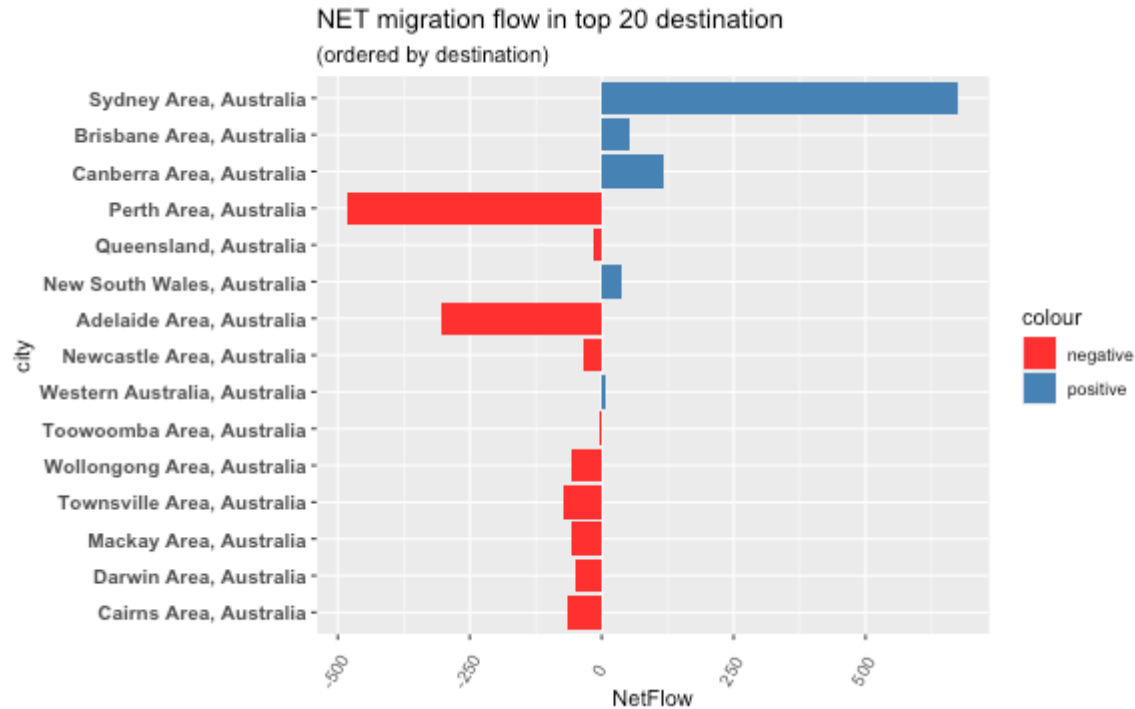
- The net flow figures, depicts a much more mixed situation than the UK.
  - 2 other cities (beyond the capital), Brisbane and Canberra, have been quite attractive in 2016
  - Conversely, several cities (notably Perth and Adelaide) have lost a lot of workers in 2016

Further data exploration by city

Bivariate measures of association

Upon studying the migration patterns in general, I now want to check if there is any relation between destination/origin of migration and the other variables in the dataset.

As an example, I select the top 3 cities of destination in the 3 countries:

- Greater New York City Area
- San Francisco Bay Area
- Washington D.C. Metro Area
- Manchester, United Kingdom
- London, United Kingdom
- Birmingham, United Kingdom
- Sydney Area, Australia
- Brisbane Area, Australia
- Canberra Area, Australia

Distribution of HIGHEST DEGREE obtained across top 3 destination cities in 3 countries
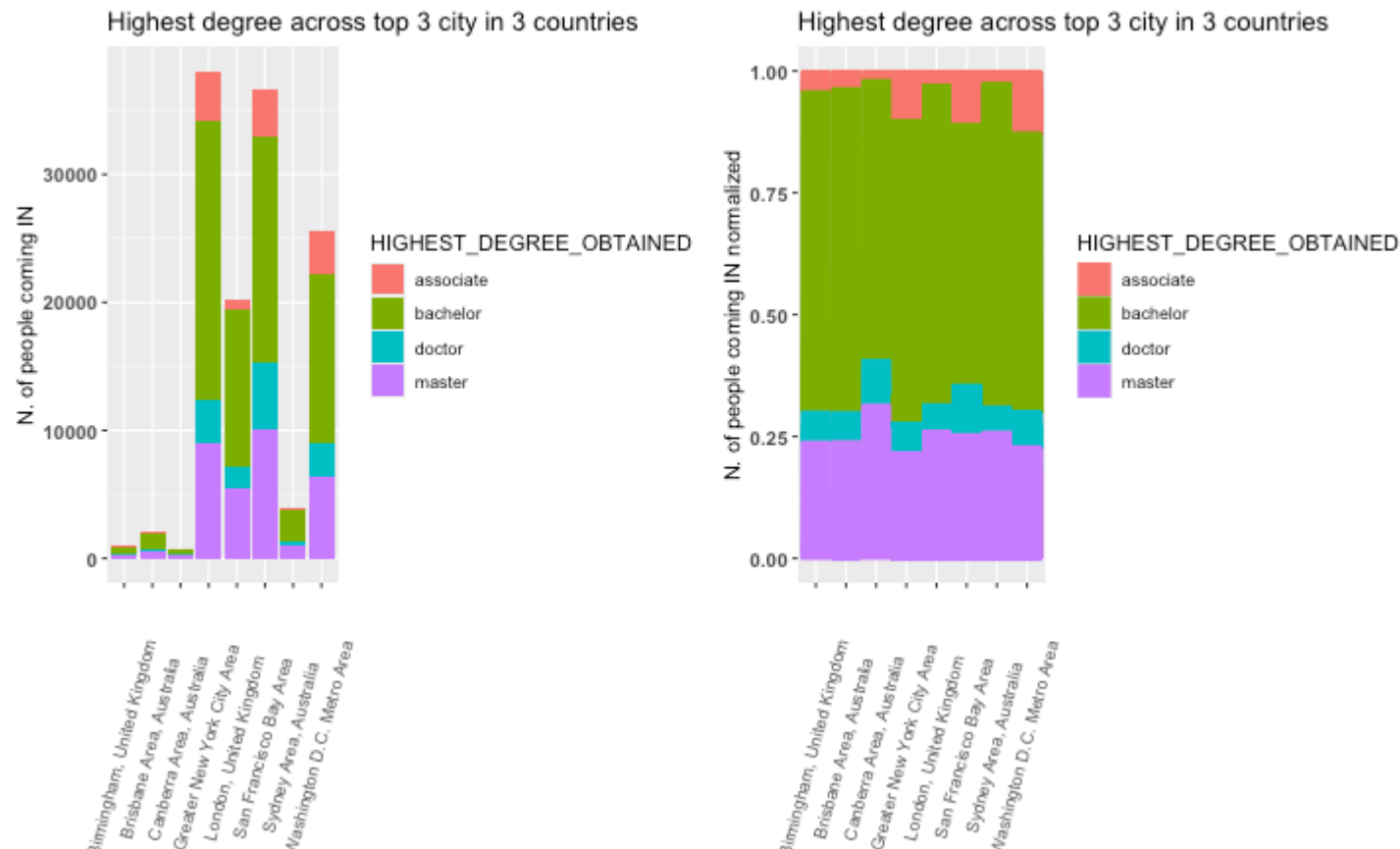
- San Francisco (followed by Canberra) seems to attract the highest amount of doctors: With Canberra receiving also the highest group of master holders
- NY, San Francisco and Washington DC seem to receive many with "Associate" level: possibly young people looking for their first job

SECTOR distribution across top 3 destination cities in 3 countries

SECTOR across top 3 city in 3 countries

SECTOR across top 3 city in 3 countries
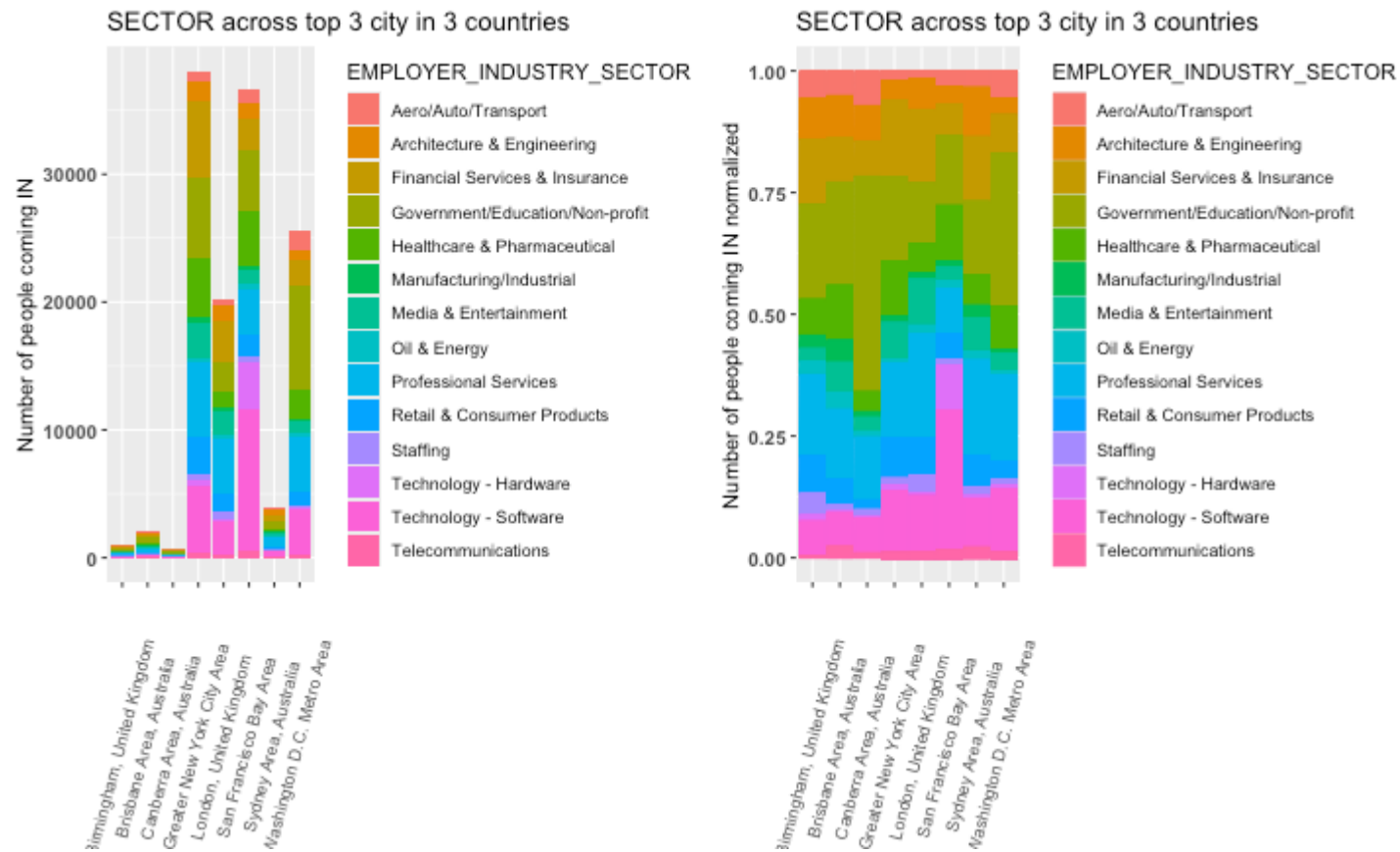
Interesting correlation between the two variables *SECTOR* & *city of destination*

- Not surprisingly, the share of immigrants by sector seems to vary a lot in a way that is correlated to the city of destination
  - Extremely high number in Govv/ Educ/ Non Profit in *Canberra* & *Washington*
  - Extremely high number in Software + Hardware Technology in *San Francisco*
  - Extremely high number in Govv/ Educ/ Non Profit + Financial Services also in *NY*

- This intuitive fact (especially knowing how American cities's sector specialization), is also validated by calculating `Cramer's V`
  - (Pearson chi-square statistic rescaled to have values between [0-1])

$$V = \sqrt{\frac{X^2}{[nobs * (min(ncols, nrows) - 1)]}}$$

- The Cramer's V statistic has the highest value for *Sector* and *Destination Region* = 0.1605205

Zooming in on domestic migration

It would be interesting to see how this noticeable association (between sector and destination city) plays out when looking at domestic migration only.

Here are the findings per country:

1. USA: `Cramer's V` = 0.1270301
2. Australia: `Cramer's V` = 0.1074171
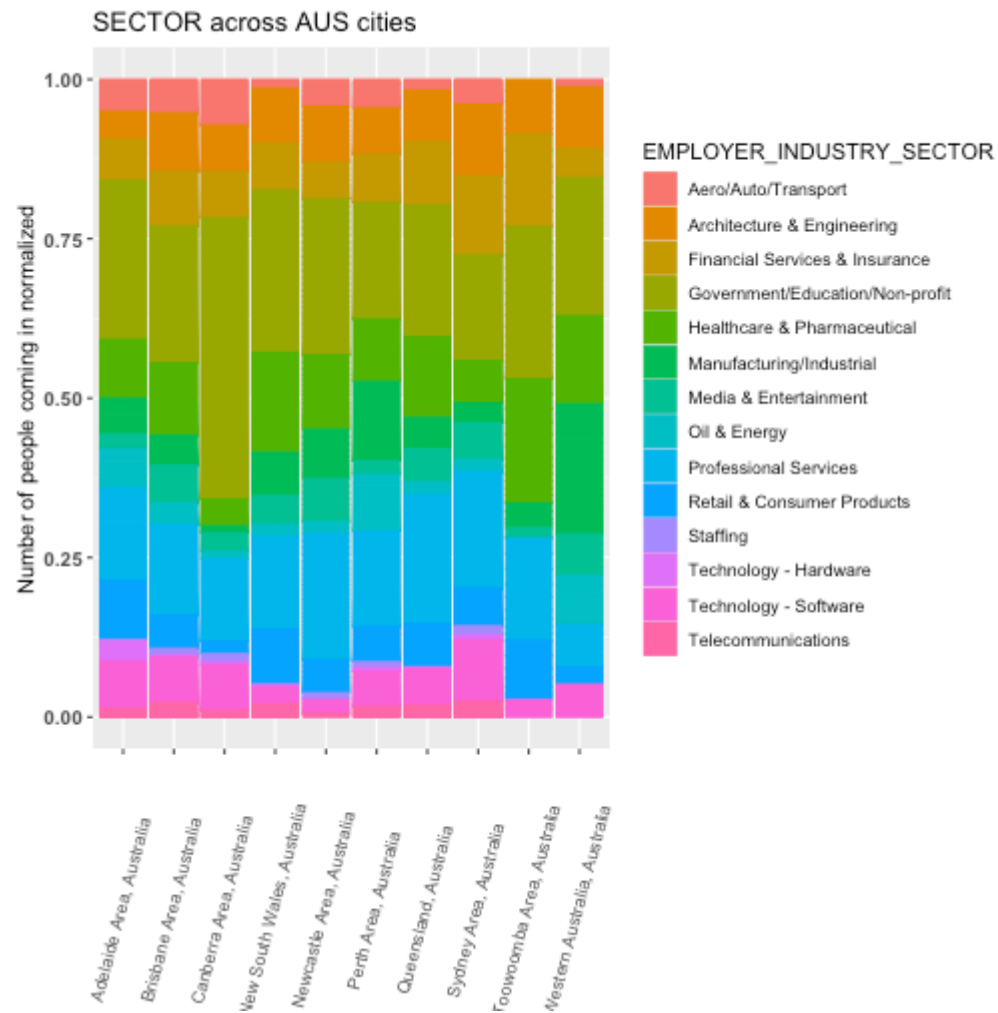3. UK: `Cramer's V` = 0.07060599

In the next 3 slides, the same association is illustrated graphically

[1] For plotting simplicity, I used random sub-samples of cities for US and UK
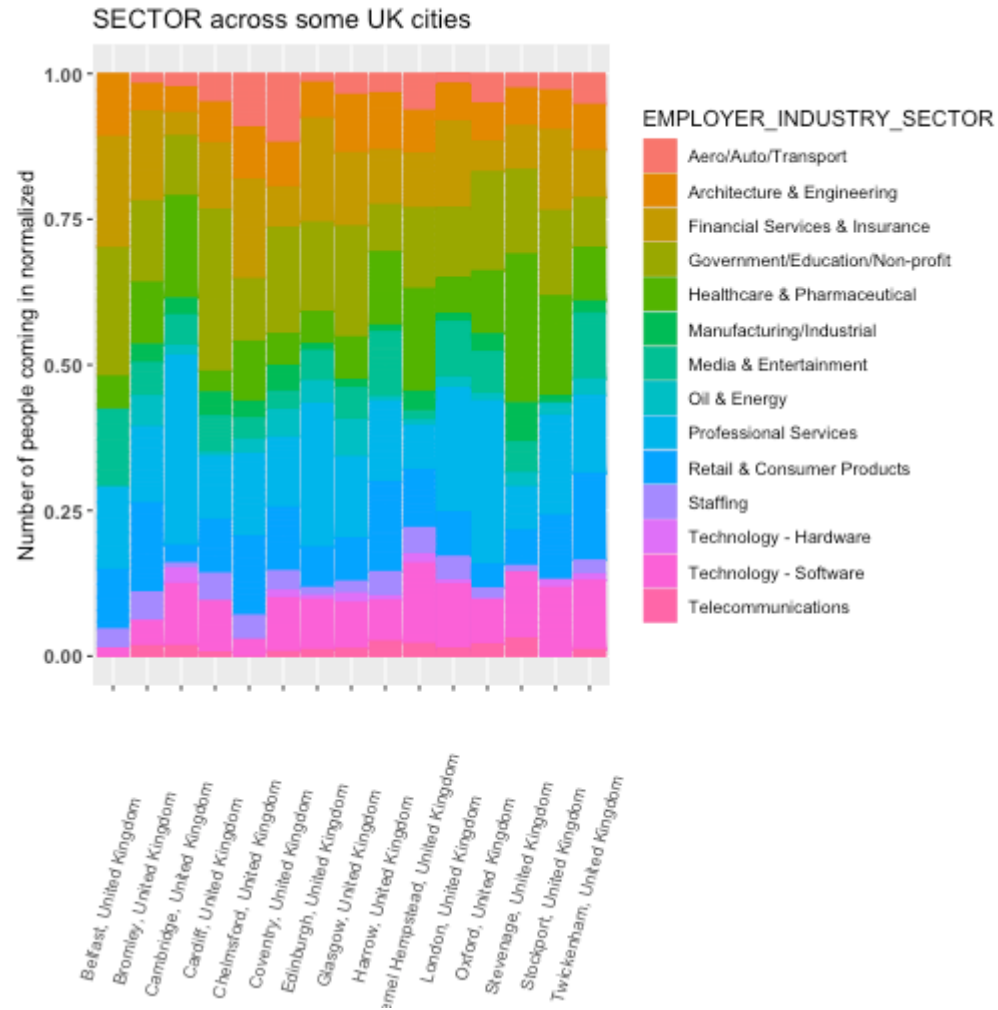
Correlation between *sector* & *city of destination* - USA domestic migration

SECTOR across various US cities
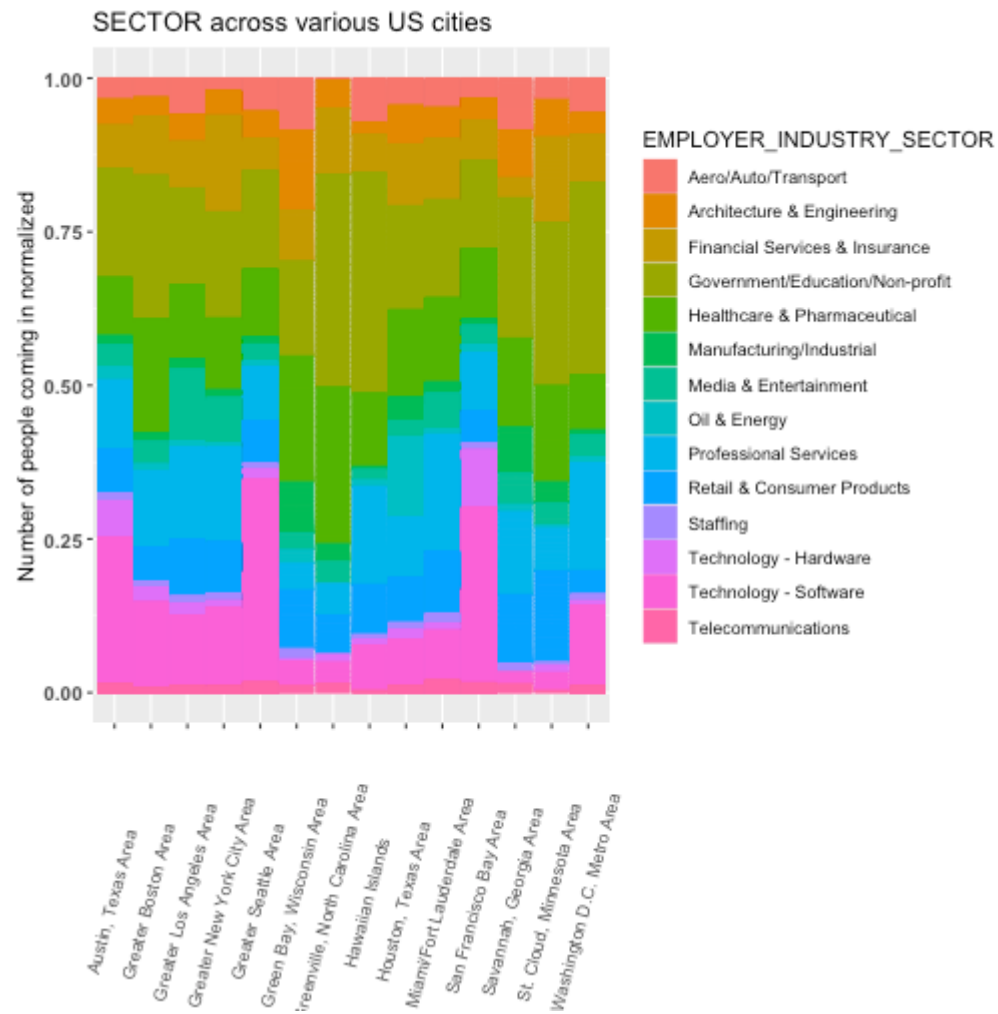
Correlation between *sector* & *city of destination* - AUS domestic migration

SECTOR across AUS cities

Correlation between *sector* & *city of destination* - UK domestic migration

SECTOR across some UK cities

EMPLOYER_INDUSTRY_SECTOR

- Aero/Auto/Transport
- Architecture & Engineering
- Financial Services & Insurance
- Government/Education/Non-profit
- Healthcare & Pharmaceutical
- Manufacturing/Industrial
- Media & Entertainment
- Oil & Energy
- Professional Services
- Retail & Consumer Products
- Staffing
- Technology - Hardware
- Technology - Software
- Telecommunications

Relevant policy questions and potential follow-up

The data - describing certain characteristics of 2016 migrants from 3 countries - is limited to 1 year and there is likely some selection bias that challenges the representativity of this sample. The sample could be biased towards highly educated, (obviously) internet-savvy, and likely skewed towards a young-ish stratum of the actual working/migrant population.

Nonetheless, the analysis of this dataset offered some very interesting insights. Some of the findings are not surprising, but some raise some crucial questions that policymakers should carefully consider if they intend to foster the competitiveness of cities.

To answer some of these questions, it would be extremely interesting to combine the Linkedin data with data from other sources that can make the information more complete and relevant.

Follow-up question # 1 and relevant external data sources

The most attractive cities have also lost many people over the past year. It would be advisable to combine the information about in-demand sectors and position with other **Amenities** that determine people's choices to relocate. It is plausible that *rent* in NY, DC or San Francisco is a strong deterrent, particularly for workers in "noncutting-edge" sectors if their *salary* is not growing at the same rate.

DATA SOURCE: glassdoor could be an interesting source because it has granular and comparable data on salaries that can be analyzed by *SECTOR*, *POSITION*, *CITY*

Follow-up question # 2 and relevant external data sources

Another important aspect for promoting cities' competitiveness is that they are actually not isolated entities, but rather *systems of cities* in that a healthy level of mobility can foster the exchange of **Skills** (intended as the individual city's collective human capital).

Clearly, the US shows a much higher and much more widespread mobility - at least in major towns all over the states. It could seem odd that in seemingly smaller countries the all mobility affect 2/3 top cities. My guess is that infrastructure and cultural factors are jointly at play.

Not so long ago, the World Bank hosted a very stimulating conference on "Space and Productivity" and will soon hold its "2nd Geospatial Day": just a couple of examples of how crucial it is to design place-based policies that acknowledge the fundamental importance of the **Form** and **Access** explanatory dimensions of the city/cities productivity function.

DATA SOURCES: the Space & Productivity conference and Geospatial Blog are good places to start.

Thank You!

Reference

- The conceptual framework is borrowed from a very interesting forthcoming report by some colleagues at the World Bank:

  - Ferreyra M.M, Roberts M. Lozano Gracia N. Restrepo Cadavid M., Selod H,. 2017 (forthcoming). *Falling Short - Cities and Productivity in Latin America and the Caribbean*. World Bank Group. Washington D.C.

- The slides were created via the R package xaringan created by '*Presentation Ninja*' Yihui Xie