# History of science in Europe 1970 – 2010

*a.k.a.*
*Does urban scaling law apply to university publications?*

Štefan[1] & Balazs[2] & Miroslav[1]

[1] University of Economics in Bratislava

[2] Centre for Economic and Regional Studies in Budapest, Corvinus University of Budapest

# Intro

- Scientific production and innovation are mostly **urban phenomenom** (Audretsch and Feldman, 1996, Glaeser et al. 1992)
- Complex system research suggests that the magnitude of urban innovative output is the function of population size characterized by the "**urban scaling law**" (Bettencourt, Lobo, Helbing et al. 2007, Bettencourt, 2013)
  - Most of the research is using **patents** as output indicator (Fritsch, Wyrwich 2020; Bettencourt, Lobo 2016)
- Scientific performance of universities is influenced by research funding (Auranen and Nieminen, 2010), by the **history** and size of the university (Frenken et al. 2017) that do not necessarily follow the urban scaling mechanisms.
  - Other universities have been established more recently in small cities for regional development purposes (Barbera and Fassero, 2013).

# Urban scaling law

$$Y(N,t) = Y_0(t)N(t)^{\beta}e^{\xi(t)}$$

$Y(N,t)$ – city property/output (urbanised area, GDP, patents...)

$N(t)$ – population at time t

$Y_0(t)$ – baseline constatnt, function of time capturing contry wide development/decline

$\beta$ - scaling exponent (elasticity), constant, time independent

$\beta > 1$ superlinear scaling, $\beta < 1$ sublinear scaling

$\xi(t)$ – statistical fluctuation

$$\beta_{socio-economic} \cong 1{,}17$$

# Data

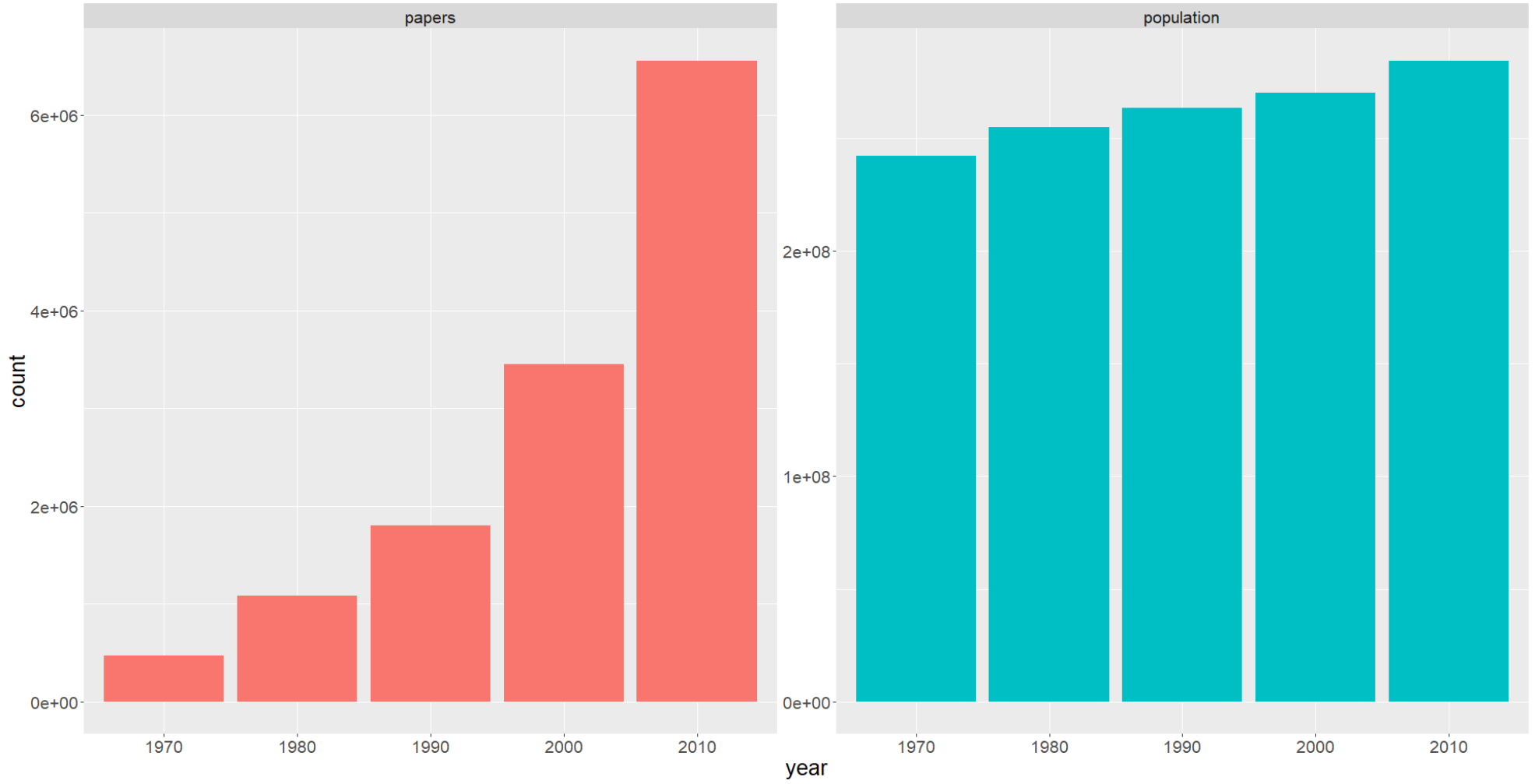
- List of European universities - **European Tertiary Education Register** (ETER, 2021).
- Matching with **Scopus Elsevier Database**:
  - 1488 European universities (ongoing data cleaning progess), time coverage 1788 and 2020
  - journal articles only and exclude books, conference papers, reviews, or editorials
- Cities - **functional urban areas** (FUA) - densely inhabited cities and their surrounding areas (commuting zones) such that they constitute an integrated labor market (OECD 2019).
- Population **- population density grids** of the Center for International Earth Science Information Network (CIESIN, 2021) we estimate the number of inhabitants of FUAs in 1970, 1980, 1990, 2000 and 2010.
- The final dataset consists of population and university publications of **451 cities from 31 European countries covering 5 time periods**. The first period (1970) uses population data from 1970 and publication data in a 10 year window (1966 – 1975). Other four time periods (1980, 1990, 2000, and 2010) follow the same rule.
- We also use **four size categories of cities**: small (up to 200 thousand inhabitants), smaller medium (up to 500 thousand), larger medium (up to 1.5 million) and large (more than 1.5 million)
- Cities are part of **different urban systems** - **centered data** for each country

# Methodology

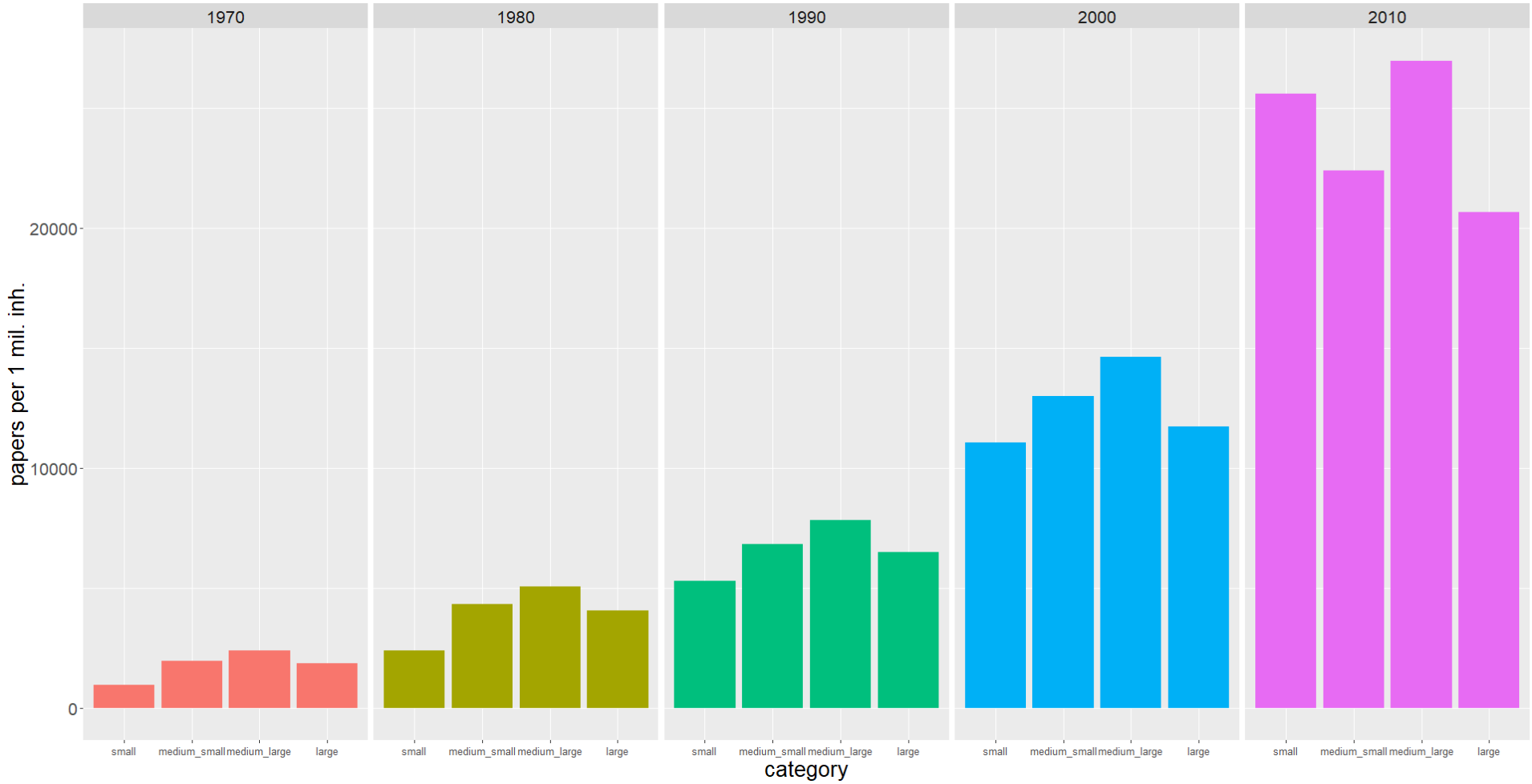
- ***RQ1: What is the evolution of urban scientific production in European cities?***
  - EDA - Exploratory data analysis

- ***RQ 2: How does the urban scaling law characterize the scientific production across European cities?***
  - Linear OLS $logY_i = logY_0 + \beta logN_i + \xi_i$
    - Comparison with theoretical value of $\beta$
  - Nonlinear OLS $logY_i = logY_0 + \beta logN_i + \beta logN_i^2 + \xi_i$

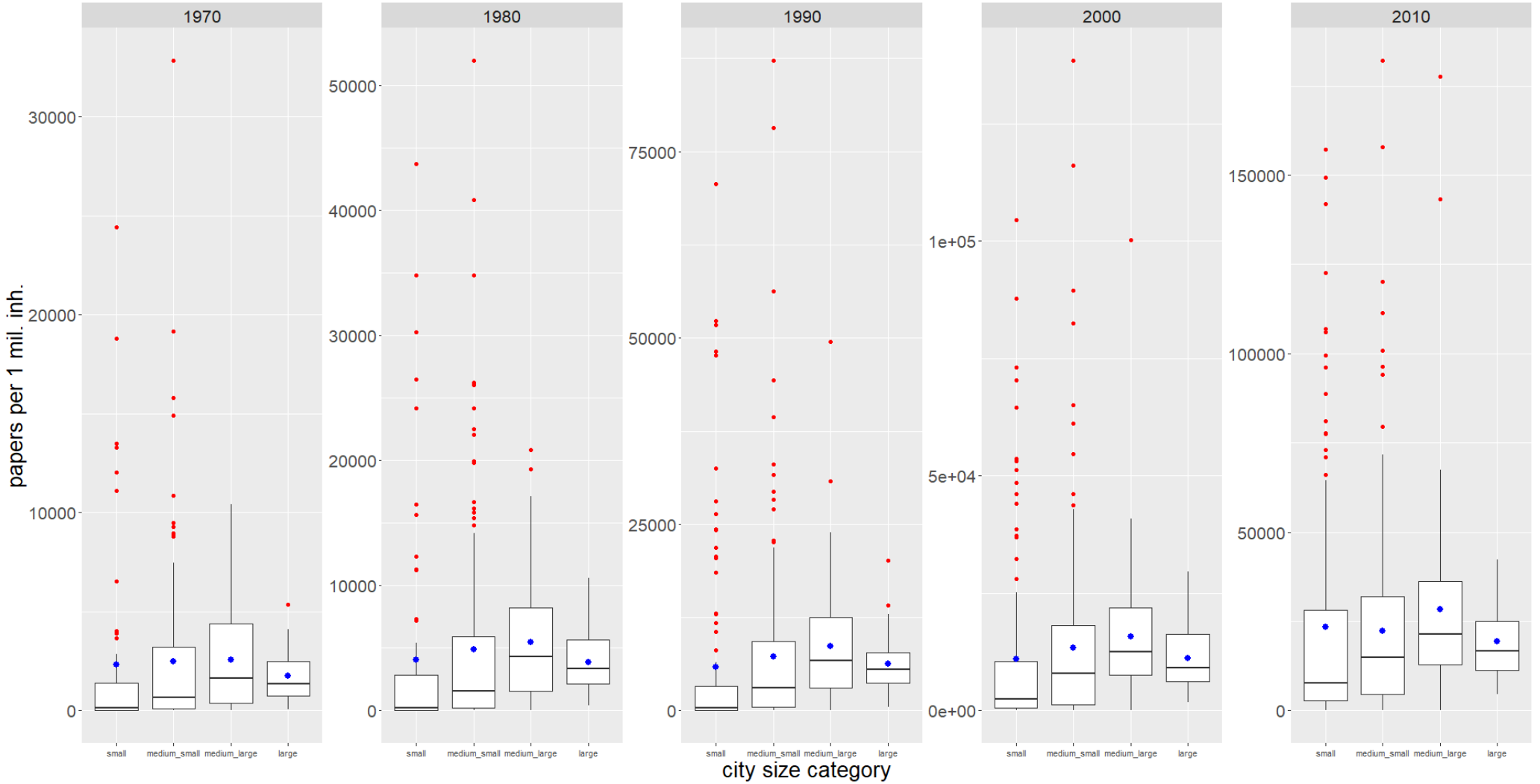
# Overview 1 – growth of sum of papers and population



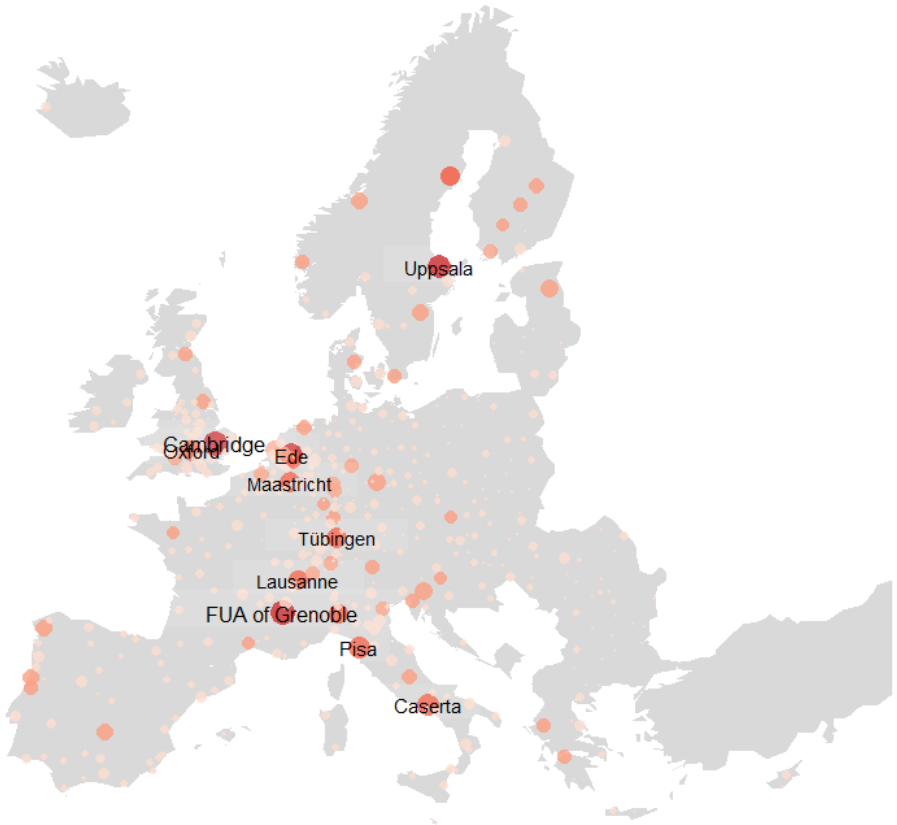
# Overview 1 – growth of scientific productivity across city size categories

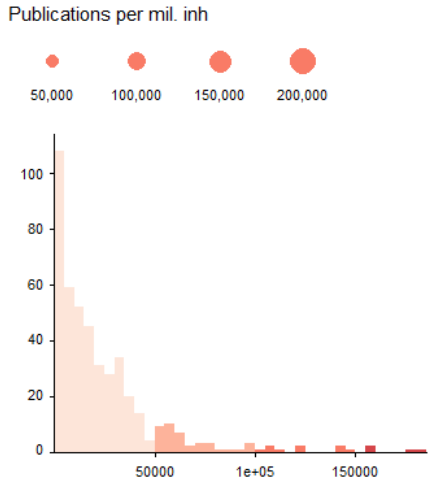


# Overview 3 – variation of productivity of cities across size categories



# Overview 3 - Most productive cities (2010)

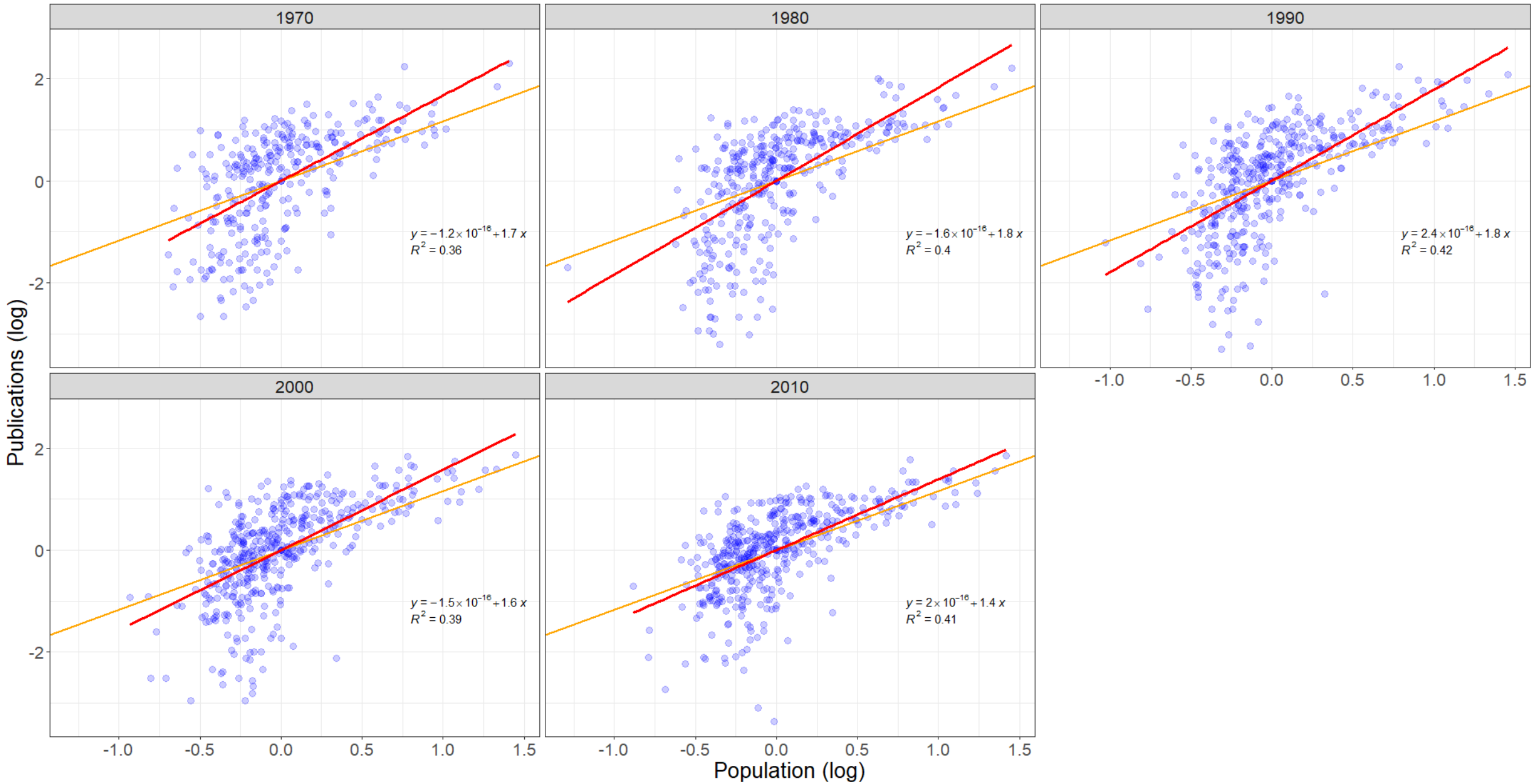



| | | | | | |
|---|---|---|---|---|---|
| Cambridge | UK | 65 490 | 359 659 | medium_small | 182 089 |
| Grenoble | FR | 115 521 | 650 689 | medium_large | 177 536 |
| Uppsala | SE | 42 907 | 271 784 | medium_small | 157 872 |
| Ede | NL | 22 933 | 145 808 | small | 157 282 |
| Pisa | IT | 28 299 | 189 470 | small | 149 359 |
| Oxford | UK | 73 404 | 512 675 | medium_large | 143 178 |
| Caserta | IT | 17 258 | 121 563 | small | 141 968 |
| Maastricht | NL | 21 954 | 179 103 | small | 122 578 |
| Tübingen | DE | 25 517 | 212 268 | medium_small | 120 211 |
| Lausanne | CH | 42 408 | 380 551 | medium_small | 111 438 |

Title: Regression results 1970 - 2010

| | Dependent variable: | | | | |
|---|---|---|---|---|---|
| | y70pap | y80pap | y90pap | y00pap | y10pap |
| | (1) | (2) | (3) | (4) | (5) |
| y70pop | 1.675*** | | | | |
| | (0.127) | | | | |
| y80pop | | 1.841*** | | | |
| | | (0.119) | | | |
| y90pop | | | 1.793*** | | |
| | | | (0.104) | | |
| y00pop | | | | 1.575*** | |
| | | | | (0.093) | |
| y10pop | | | | | 1.391*** |
| | | | | | (0.080) |
| Constant | -0.000 | -0.000 | 0.000 | -0.000 | 0.000 |
| | (0.046) | (0.044) | (0.039) | (0.035) | (0.030) |
| Observations | 307 | 359 | 414 | 443 | 448 |
| R2 | 0.364 | 0.402 | 0.417 | 0.394 | 0.406 |
| Adjusted R2 | 0.362 | 0.400 | 0.416 | 0.393 | 0.405 |
| Residual Std. Error | 0.810 (df = 305) | 0.841 (df = 357) | 0.794 (df = 412) | 0.729 (df = 441) | 0.626 (df = 446) |
| F Statistic | 174.398*** (df = 1; 305) | 239.587*** (df = 1; 357) | 295.093*** (df = 1; 412) | 287.215*** (df = 1; 441) | 305.404*** (df = 1; 446) |
| Note: | | | | | *p<0.1; **p<0.05; ***p<0.01 |


1970
1980
1990
2000
2010
Publications (log)
Population (log)
$y = -1.2 \times 10^{-16} + 1.7\,x$
$R^2 = 0.36$
$y = -1.6 \times 10^{-16} + 1.8\,x$
$R^2 = 0.4$
$y = 2.4 \times 10^{-16} + 1.8\,x$
$R^2 = 0.42$
$y = -1.5 \times 10^{-16} + 1.6\,x$
$R^2 = 0.39$
$y = 2 \times 10^{-16} + 1.4\,x$
$R^2 = 0.41$
2
0
-2
-1.0
-0.5
0.0
0.5
1.0
1.5

Title: Regression results 1970 - 2010

| | Dependent variable: | | | | |
|---|---|---|---|---|---|
| | y70pap | y80pap | y90pap | y00pap | y10pap |
| | (1) | (2) | (3) | (4) | (5) |
| y70pop | 1.841*** | | | | |
| | (0.151) | | | | |
| y70pop_sqr | -0.483** | | | | |
| | (0.243) | | | | |
| y80pop | | 2.028*** | | | |
| | | (0.136) | | | |
| y80pop_sqr | | -0.564*** | | | |
| | | (0.204) | | | |
| y90pop | | | 2.001*** | | |
| | | | (0.125) | | |
| y90pop_sqr | | | -0.541*** | | |
| | | | (0.184) | | |
| y00pop | | | | 1.755*** | |
| | | | | (0.112) | |
| y00pop_sqr | | | | -0.463*** | |
| | | | | (0.164) | |
| y10pop | | | | | 1.537*** |
| | | | | | (0.099) |
| y10pop_sqr | | | | | -0.358** |
| | | | | | (0.146) |
| Constant | 0.064 | 0.079 | 0.076 | 0.064 | 0.049 |
| | (0.056) | (0.052) | (0.046) | (0.041) | (0.036) |
| Observations | 307 | 359 | 414 | 443 | 448 |
| R2 | 0.372 | 0.414 | 0.429 | 0.405 | 0.414 |
| Adjusted R2 | 0.368 | 0.411 | 0.427 | 0.402 | 0.412 |
| Residual Std. Error | 0.806 (df = 304) | 0.833 (df = 356) | 0.787 (df = 411) | 0.723 (df = 440) | 0.623 (df = 445) |
| F Statistic | 90.029*** (df = 2; 304) | 125.856*** (df = 2; 356) | 154.634*** (df = 2; 411) | 149.817*** (df = 2; 440) | 157.457*** (df = 2; 445) |
| Note: | | | | | *p<0.1; **p<0.05; ***p<0.01 |

# Conclusions

- Substantial increase of university publication activity in cities during last 50 years
- Partly confirmed urban scaling law for scientific production measured by publications
  - Suggested non linear relation
- History matters in scientific production – established university cities
  - Large cities are uderperforming, medium sized cities perform better than expected by the urban scaling law
  - Small cities are catching up in recent periods

# Annex 1

$$\beta_{socio-economic} = 1 + \delta$$
$$\beta_{employment} = 1$$
$$\beta_{build-infrastucture} = 1 - \delta$$

$$\delta = \frac{H}{2(H+2)}$$

*H – fractal dimension of movement*

$$2 \geq H \geq 0$$
$$H \to 0 \; only \; local \; movement, interactions$$
$$H \to 2 \; city \; widy \; movement, interactions$$
$$H \geq 1 \; mixing$$
$$H = 1 \; minimal \; movement \; costs$$

$$\beta_{socio-economic} = \frac{7}{6} = 1{,}17$$

$$\beta_{employment} = 1$$

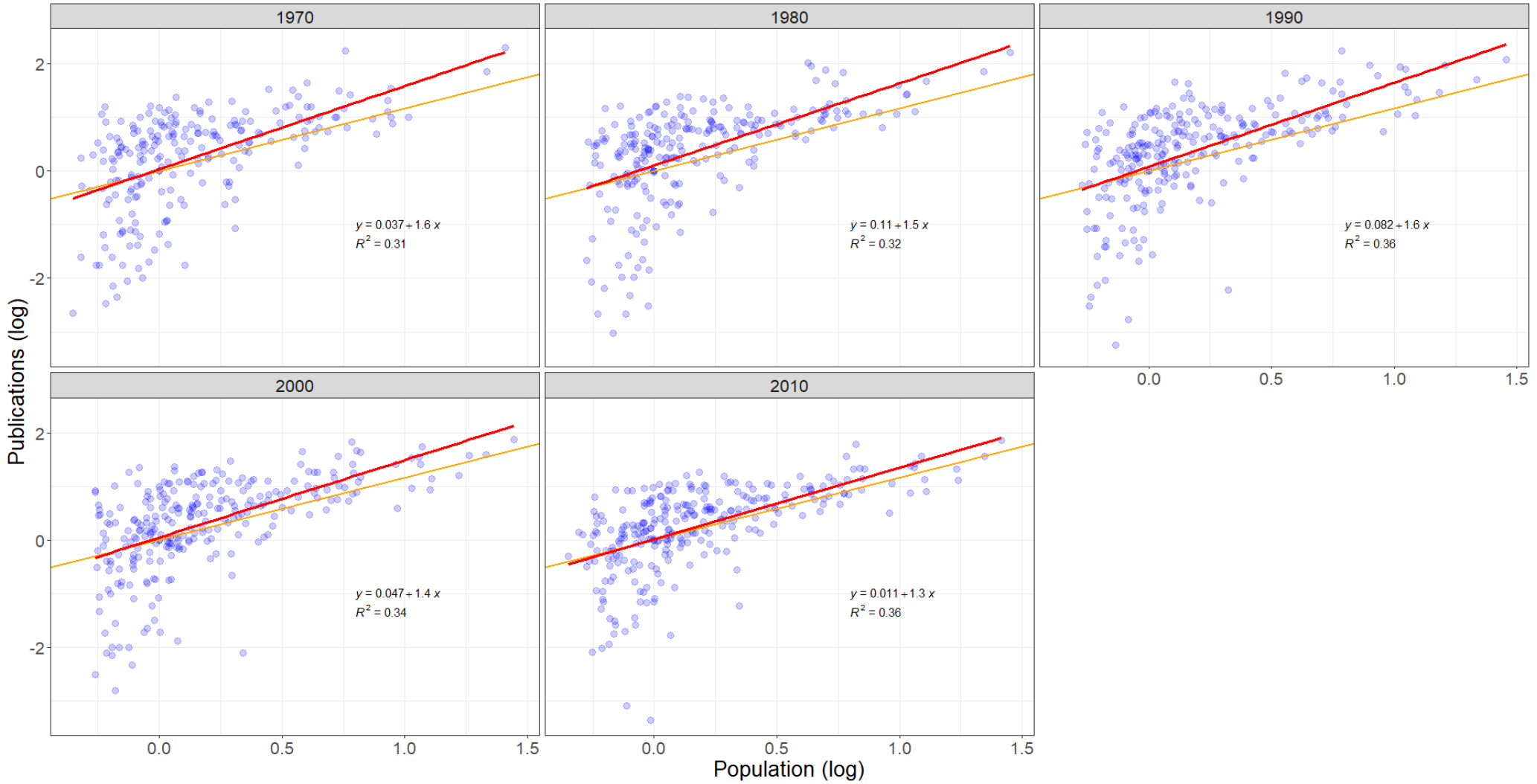
$$\beta_{build-infrastucture} = \frac{5}{6} = 0{,}83$$

$$\delta = 1/6 = 0{,}17$$

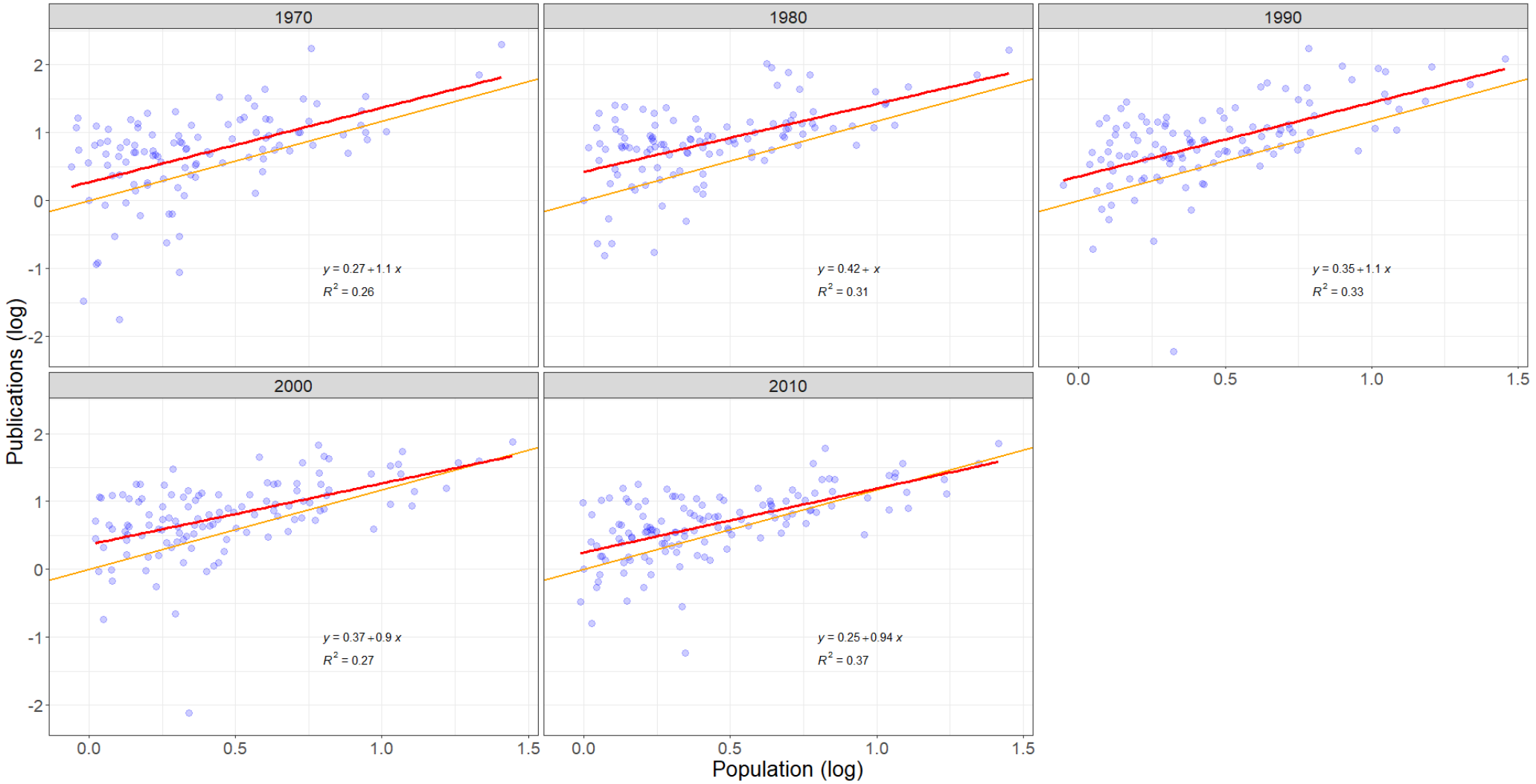
# Annex 2 Data problems

- Data problems
  - Clearing problems (matching)
  - Double counting of publications
- Definition of a city
  - Zero values problem (Finance, Cottineau 2018)
- Measure of scientific output
  - quality of papers is assumed to be same

# 250+


1970
1980
1990
2000
2010
y = 0.037 + 1.6 x
R² = 0.31
y = 0.11 + 1.5 x
R² = 0.32
y = 0.082 + 1.6 x
R² = 0.36
y = 0.047 + 1.4 x
R² = 0.34
y = 0.011 + 1.3 x
R² = 0.36
Publications (log)
Population (log)
2
0
-2
0.0
0.5
1.0
1.5

# 500+


1970
1980
1990
2000
2010
y = 0.27 + 1.1 x
R² = 0.26
y = 0.42 + x
R² = 0.31
y = 0.35 + 1.1 x
R² = 0.33
y = 0.37 + 0.9 x
R² = 0.27
y = 0.25 + 0.94 x
R² = 0.37
Publications (log)
Population (log)
2
1
0
-1
-2
0.0
0.5
1.0
1.5