**Clustering scenarios for Summit Paper 07-06-16**

**Scenario 1: Economic variables only (modified)**

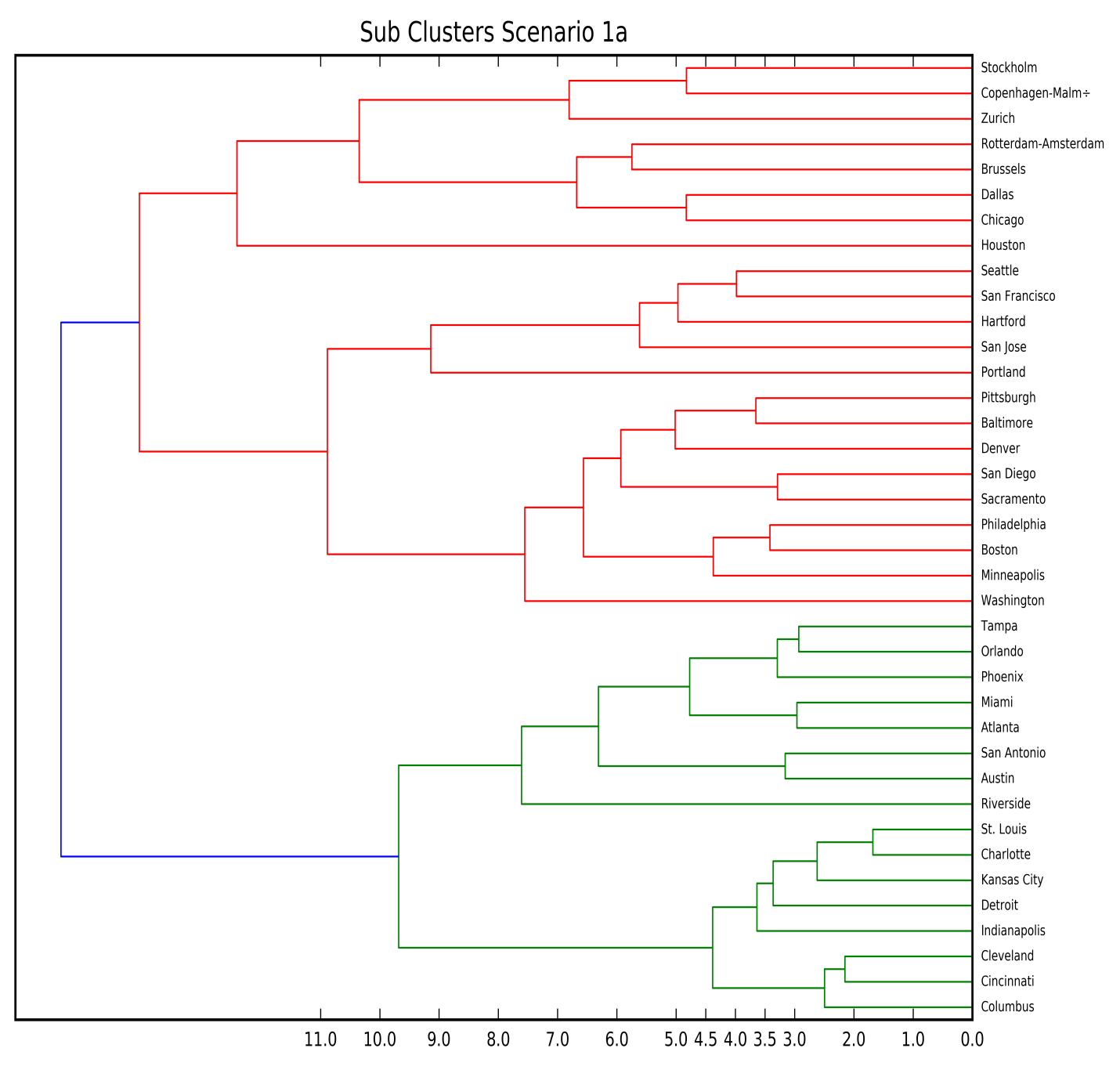
**22 Variables included:**

* Share of industry in overall GDP for 2015 (8 industries), and productivity by sector (8 industries)
* Population, nominal GDP, Real GDP, Real GDP per capita, and share of population in labor force.
* Seven principal components use in the model that explains 86 percent of overall variation
* Five clusters obtained through kmeans, further subdivided to obtain a total of 7 clusters.

**Results**

|  |  |  |
| --- | --- | --- |
| **Number of observations** | **Metros** | **Notes** |
| **19** | Ankara, Beijing, Cape, Town, Delhi, East, Rand, Guangzhou, Hong, Kong, Istanbul, Johannesburg, Mexico, City, Moscow, Pretoria, Rio, de, Janeiro, Saint, Petersburg, Santiago, Sao, Paulo, Seoul-Incheon, Shanghai, Shenzhen |  |
| **25** | Athens, Barcelona, Berlin, Birmingham (UK), Brasilia, Frankfurt, am, Main, Hamburg, Karlsruhe, Kitakyushu-Fukuoka, Koln- Dusseldorf, Madrid, Melbourne, Milan, Montreal, Munich, Nagoya, Rome, Singapore, Stuttgart, Sydney, Tel, Aviv, Toronto, Vancouver, Vienna-Bratislava, Warsaw |  |
| **16** | Atlanta, Austin, Charlotte, Cincinnati, Cleveland, Columbus, Detroit, Indianapolis, Kansas, City, Miami, Orlando, Phoenix, Riverside, San, Antonio, St., Louis, Tampa | **This cluster was originally composed of 38 metros** |
| **22** | Baltimore, Boston, Brussels, Chicago, Copenhagen-Malmo, Dallas, Denver, Hartford, Houston, Minneapolis, Philadelphia, Pittsburgh, Portland, Rotterdam-Amsterdam, Sacramento, San, Diego, San, Francisco, San, Jose, Seattle, Stockholm, Washington, Zurich | **This cluster was originally composed of 38 metros** |
| **12** | Changchun, Changsha, Dongguan, Haerbin, Nantong, Perth, Qingdao, Shijiazhuang, Suzhou, Tangshan, Yantai, Zibo |  |
| **6** | London, Los, Angeles, New, York, Osaka-Kobe, Paris, Tokyo |  |
| **23** | Busan-Ulsan, Changzhou, Chengdu, Chongqing, Dalian, Foshan, Fuzhou, Hangzhou, Hefei, Jinan, Katowice-Ostrava, Monterrey, Mumbai, Nanjing, Ningbo, Shenyang, Tianjin, Wenzhou, Wuhan, Wuxi, Xi'an, Xuzhou, Zhengzhou |  |

**Figure 1: Subdivision of cluster 3 and 4**

****

**Summary statistics**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **label** | **count** | **gmm14 2015 pop** | **tot emp2015** | **gmm14 2015 realgdp** | **gdp pc real2015** | **Product.2015** | **share15** | **share rgva2015 business** | **share rgva2015 commodities** | **share rgva2015 local** | **share rgva2015 manufact.** | **sector prod2015 business** |
| 1 | 19 | 12,171.7 | 5,579.1 | 213,311.3 | 13,445.1 | 27.6 | 45.8% | 30.4% | 0.7% | 15.7% | 13.9% | 53.7 |
| 2 | 25 | 4,313.5 | 2,146.5 | 194,432.9 | 46,134.7 | 89.3 | 51.3% | 32.1% | 0.6% | 25.0% | 13.0% | 157.3 |
| 3 | 16 | 2,406.8 | 1,145.5 | 116,479.1 | 48,099.5 | 100.4 | 47.5% | 36.6% | 0.5% | 27.1% | 9.3% | 209.6 |
| 4 | 22 | 3,412.1 | 1,759.0 | 201,733.0 | 60,751.6 | 125.4 | 50.1% | 39.4% | 0.4% | 25.3% | 11.5% | 238.4 |
| 5 | 12 | 7,701.8 | 1,695.7 | 90,655.0 | 12,642.0 | 50.9 | 24.5% | 10.4% | 14.9% | 10.5% | 38.4% | 68.8 |
| 6 | 6 | 16,747.3 | 8,828.6 | 878,890.7 | 60,692.2 | 117.7 | 49.9% | 44.8% | 0.3% | 23.6% | 8.6% | 212.6 |
| 7 | 23 | 8,244.7 | 3,068.4 | 105,532.3 | 13,017.5 | 30.7 | 42.0% | 13.7% | 6.6% | 10.0% | 38.5% | 42.2 |

**Scenario 2a: Economic variables and Global Profiles variables without logarithmic transformation**

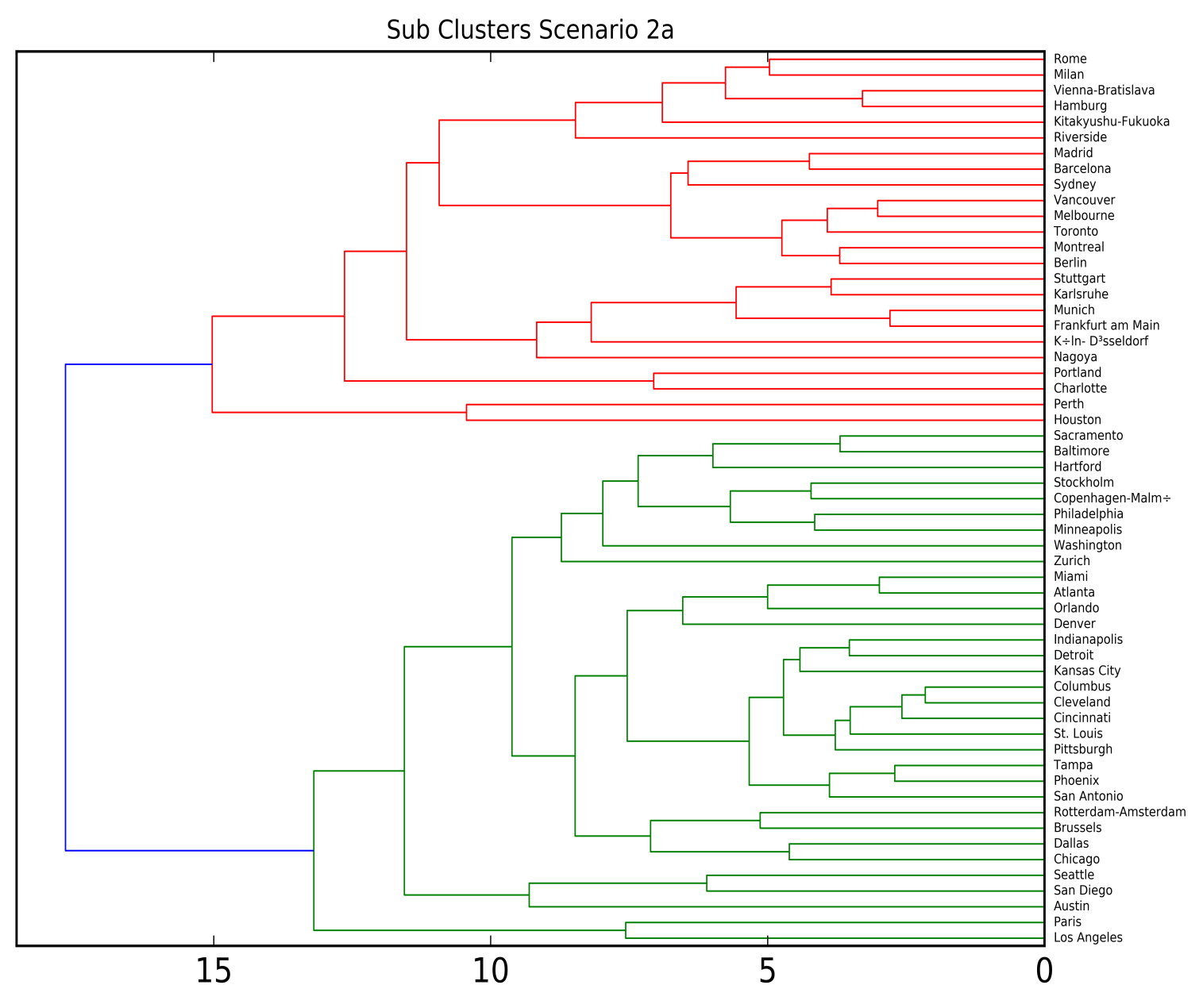
35 Variables included:

* Share of industry in overall GDP for 2015 (8 industries), and productivity by sector (8 industries)
* Population, nominal GDP, Real GDP, Real GDP per capita, and share of population in labor force.
* Fdi per capita 2009-2015, total stock of FDI 2009-2015, total stock of jobs created by FDI 2009-2015, Share of publications in top 10 percent journals between 2010-2013, mean citation score between 2010-2013, patens between 2008-2012, patents per capita 2008-2012, stock of VC 2006-2015, stock of VC per capita 2006-2015, passengers 2014, passengers per capita 20014, download speed in 2015, and share of population with tertiary education
* Nine principal components use in the model that explains 86 percent of overall variation

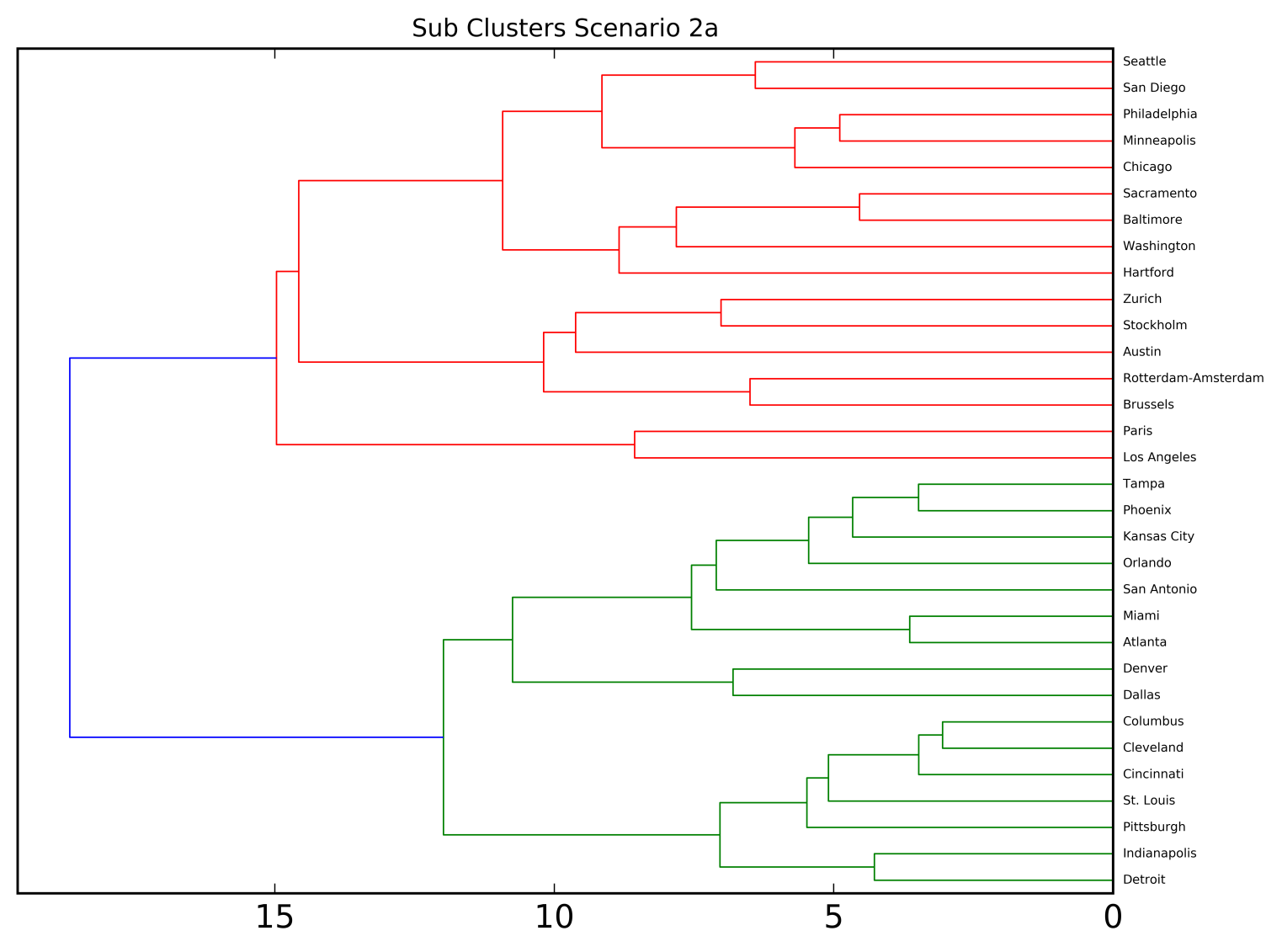
**Results**

|  |  |  |
| --- | --- | --- |
| **Number of observations** | **Metros** | **Notes** |
| **25** | Ankara Athens Birmingham (UK) Brasilia Busan-Ulsan Cape Town Delhi East Rand Guangzhou Hangzhou Istanbul Johannesburg Mexico City Monterrey Moscow Mumbai Pretoria Rio de Janeiro Saint Petersburg Santiago Sao Paulo Shenzhen Tel Aviv Tianjin Warsaw |  |
| **32** | Atlanta Austin Baltimore Brussels Chicago Cincinnati Cleveland Columbus Dallas Denver Detroit Hartford Indianapolis Kansas City Los Angeles Miami Minneapolis Orlando Paris Philadelphia Phoenix Pittsburgh Rotterdam-Amsterdam Sacramento San Antonio San Diego Seattle St. Louis Stockholm Tampa Washington Zurich | **This cluster was originally composed of 57 metros**  **We could further subdivide this cluster using ward method to obtain 16 clusters** |
| **23** | Barcelona Berlin Charlotte Frankfurt am Main Hamburg Houston Karlsruhe Kitakyushu-Fukuoka Madrid Melbourne Milan Montreal Munich Nagoya Perth Portland Riverside Rome Stuttgart Sydney Toronto Vancouver Vienna-Bratislava | **This cluster was originally composed of 57 metros** |
| **9** | Beijing Hong Kong London New York Osaka-Kobe Seoul-Incheon Shanghai Singapore Tokyo |  |
| **29** | Changchun Changsha Changzhou Chengdu Chongqing Dalian Dongguan Foshan Fuzhou Haerbin Hefei Jinan Katowice-Ostrava Nanjing Nantong Ningbo Qingdao Shenyang Shijiazhuang Suzhou Tangshan Wenzhou Wuhan Wuxi Xi'an Xuzhou Yantai Zhengzhou Zibo |  |
| **3** | Boston, San Francisco and San Jose |  |

**Figure 2: Subdivision of cluster 2 and 3 (average)**

****

**Figure 3: Additional Subdivision of Cluster 3 (ward)**

****

**Summary Statistics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| label cluster | count | gmm14 2015 pop | gmm14 2015 realgdp | gdp pc real2015 | productivity2015 | share rgva2015business | sector prod2015business | stockvalue15 |
| 1 | 25 | 7,299.8 | 122,519.8 | 14,811.6 | 32.8 | 28.7% | 54.7 | 9,261.8 |
| 2 | 33 | 2,975.2 | 181,882.2 | 54,896.0 | 112.8 | 38.0% | 217.8 | 2,528.9 |
| 3 | 24 | 4,475.2 | 198,292.0 | 48,387.0 | 96.9 | 31.7% | 170.7 | 7,300.5 |
| 4 | 9 | 20,182.3 | 546,031.4 | 41,279.9 | 73.9 | 30.4% | 93.3 | 37,856.2 |
| 5 | 29 | 7,802.5 | 92,030.4 | 12,062.1 | 36.9 | 12.7% | 48.6 | 4,649.1 |
| 6 | 3 | 4,656.1 | 337,288.3 | 72,991.7 | 149.1 | 47.3% | 292.1 | 7,164.2 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| label cluster | stockjobs15 | fdipc15 | share10pc13 | patentspc12 | stockvc15 | vcpc15 | passengerspc15 | f15 |
| 1 | 29,645.0 | 922.6 | 8.0% | 0.1 | 158.4 | 19.7 | 3,127.9 | 4.0 |
| 2 | 7,424.0 | 838.8 | 15.8% | 0.8 | 2,007.1 | 527.0 | 10,611.7 | 14.4 |
| 3 | 16,706.0 | 1,298.9 | 13.8% | 0.6 | 543.8 | 169.3 | 8,402.0 | 19.4 |
| 4 | 101,951.0 | 1,730.5 | 9.5% | 0.6 | 2,232.5 | 226.7 | 6,358.2 | 7.4 |
| 5 | 17,080.0 | 598.8 | 8.4% | 0.0 | 46.3 | 5.5 | 1,796.5 | 2.9 |
| 6 | 23,883.0 | 1,625.8 | 22.5% | 3.2 | 60,960.4 | 19,183.9 | 11,736.1 | 29.8 |

**Scenario 3: Economic variables and Global Profiles variables with logarithmic transformation**

35 Variables included:

* Share of industry in overall GDP for 2015 (8 industries), and productivity by sector (8 industries)
* Population, nominal GDP, Real GDP, Real GDP per capita, and share of population in labor force.
* FDI per capita 2009-2015, total stock of FDI 2009-2015, total stock of jobs created by FDI 2009-2015, Share of publications in top 10 percent journals between 2010-2013, mean citation score between 2010-2013, patens between 2008-2012, patents per capita 2008-2012, stock of VC 2006-2015, stock of VC per capita 2006-2015, passengers 2014, passengers per capita 20014, download speed in 2015, and share of population with tertiary education
* Nine principal components use in the model that explains 86 percent of overall variation

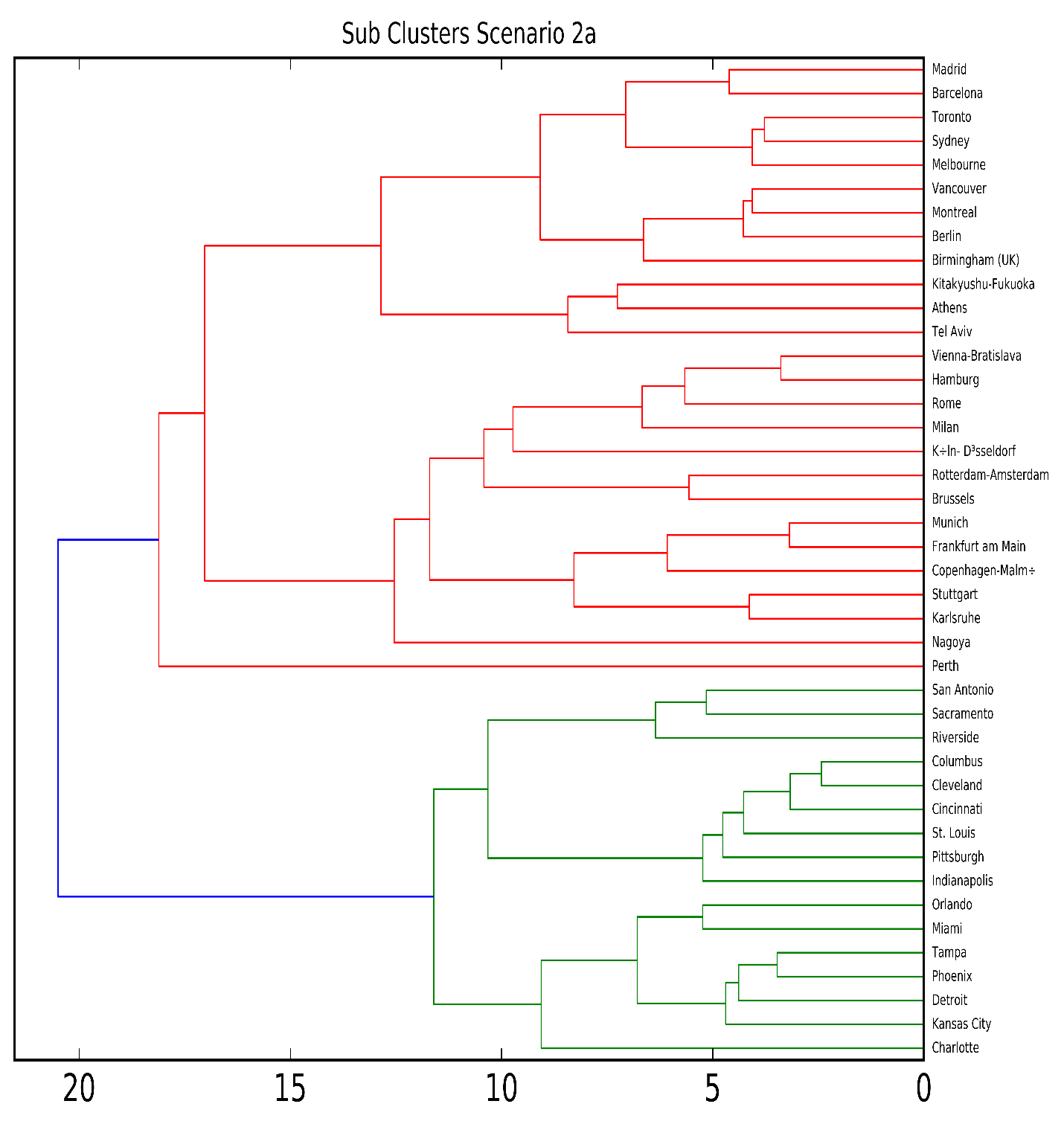
**Results**

|  |  |  |
| --- | --- | --- |
| **Number of observations** | **Metros** | **Notes** |
| **22** | Changchun Changsha Changzhou Chengdu Dalian Dongguan Foshan Fuzhou Haerbin Hefei Nantong Qingdao Shenyang Shijiazhuang Suzhou Tangshan Wenzhou Wuxi Xuzhou Yantai Zhengzhou Zibo |  |
| **19** | Atlanta Austin Baltimore Boston Chicago Dallas Denver Hartford Houston Minneapolis Philadelphia Portland San Diego San Francisco San Jose Seattle Stockholm Washington Zurich |  |
| **28** | Ankara Brasilia Busan-Ulsan Cape Town Chongqing Delhi East Rand Guangzhou Hangzhou Istanbul Jinan Johannesburg Katowice-Ostrava Mexico City Monterrey Mumbai Nanjing Ningbo Pretoria Rio de Janeiro Saint Petersburg Santiago Sao Paulo Shenzhen Tianjin Warsaw Wuhan Xi'an |  |
| **6** | Beijing Hong Kong Moscow Seoul-Incheon Shanghai Singapore |  |
| **6** | London Los Angeles New York Osaka-Kobe Paris Tokyo |  |
| **16** | Charlotte Cincinnati Cleveland Columbus Detroit Indianapolis Kansas City Miami Orlando Phoenix Pittsburgh Riverside Sacramento San Antonio St. Louis Tampa | **This cluster was originally composed of 42 metros** |
| **26** | Athens Barcelona Berlin Birmingham (UK) Brussels Copenhagen-Malmo Frankfurt am Main Hamburg Karlsruhe Kitakyushu-Fukuoka Koln- Dusseldorf Madrid Melbourne Milan Montreal Munich Nagoya Perth Rome Rotterdam-Amsterdam Stuttgart Sydney Tel Aviv Toronto Vancouver Vienna-Bratislava | **This cluster was originally composed of 42 metros** |

Notes:

8 principal components that account for 86 percent of the data variation. The test suggested an optimal number of clusters in the range of 6-8, I selected 6 to avoid having clusters with very few observations. This cluster did not use the per capita indicators for FDI and VC. Despite the elimination of two big sources of variation unstable clusters were still observed. Additional steps to merge sub clusters with each other might be required.

Figure 4



Summary Statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| label cluster | count | gmm14 2015 pop | gmm14 2015 realgdp | gdp pc real2015 | prod. 2015 | share rgva2015business | sector prod2015business | stock fdi value15 |
| 1 | 22 | 7,701.8 | 90,655.0 | 11,200.0 | 43.4 | 12.1% | 50.0 | 3,380.4 |
| 2 | 19 | 3,524.6 | 206,339.7 | 63,096.4 | 126.4 | 41.1% | 240.2 | 4,290.6 |
| 3 | 28 | 8,028.5 | 103,091.8 | 13,554.0 | 26.9 | 25.3% | 42.1 | 10,460.9 |
| 4 | 6 | 17,034.5 | 281,292.9 | 22,375.1 | 40.8 | 29.1% | 73.1 | 39,715.8 |
| 5 | 6 | 16,747.3 | 878,890.7 | 60,692.2 | 117.7 | 44.8% | 212.6 | 17,857.9 |
| 6 | 16 | 2,385.6 | 116,479.1 | 48,099.5 | 100.4 | 36.8% | 209.6 | 2,123.2 |
| 7 | 26 | 4,390.7 | 198,292.0 | 47,358.9 | 94.2 | 32.4% | 167.3 | 8,401.6 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| label cluster | stockjobs15 | fdipc15 | share10pc13 | patents pc12 | stock vc15 | vc pc15 | passengers pc 2015 | foreign 15 |
| 1 | 15,808.5 | 486.3 | 8.4% | 0.00 | 44.7 | 5.9 | 1,295.9 | na |
| 2 | 8,661.0 | 881.8 | 17.2% | 1.63 | 6,118.9 | 1,041.3 | 12,386.6 | 17.17 |
| 3 | 30,981.5 | 986.0 | 7.8% | 0.06 | 132.1 | 14.2 | 2,755.5 | 2.5 |
| 4 | 111,181.0 | 2,458.9 | 9.4% | 0.39 | 2,762.1 | 248.3 | 6,180.0 | 7.85 |
| 5 | 53,107.5 | 1,078.9 | 16.7% | 0.97 | 8,760.6 | 622.9 | 8,862.4 | 25.5 |
| 6 | 5,692.5 | 781.6 | 14.3% | 0.52 | 874.7 | 347.8 | 7,990.3 | 9.18 |
| 7 | 18,529.0 | 1,598.5 | 13.8% | 0.60 | 778.9 | 184.5 | 7,270.8 | 19.4 |