Atop Canada's Urban Hierarchy. Toronto, October 2005 (Elvin Wyly). In 2006, Canada's most populous urban region, the Toronto Census Metropolitan Area (CMA), had a population of 4.75 million. Montreal was ranked second, at 3.31 million, and Vancouver was ranked third, at 1.95 million.

Theories of Urban System Development

Geography 350, *Introduction to Urban Geography*
September 26, 2012
Elvin Wyly

In 1871, Montreal was the largest city in Canada; Montreal held this position for the next century, before Toronto assumed the highest rank. In the United States, New York City was the nation's largest city in 1790; it has occupied this dominant position ever since. *Why?*

In 1881, Ontario had a single city with more than 100,000 people, but four centers with populations between 25,000 and 99,999, and 98 towns with populations between 1,000 and 2,499.[1] More than a century later the same pattern held for Ontario, and for the rest of Canada as well: "The number of cities in each size category is inversely related to city size, with 50 or more cities in each of the two smallest categories and only four with populations over

[1] George A. Nader (1975). *Cities of Canada: Theoretical, Historical, and Planning Perspectives.* Toronto: Macmillan, p. 226.

1,000,000."[2] Meanwhile, the United States has nine cities with populations exceeding one million, and 16,683 incorporated places with populations less than 10,000. *Why?*

Today, more than a third of all Canadians live in just three metropolitan regions: Toronto, Montreal, and Vancouver. The fraction rises to one half if we add Ottawa-Gatineau, Calgary, Edmonton, Quebec City, and Winnipeg.[3] As the prominent urban geographers Larry Bourne and Jim Simmons conclude, "The average Canadian, if there is such a person, now lives in a large metropolitan environment," and these enormous agglomerations have in recent years accounted for more than two-thirds of the nation's net population growth.[4] *So what?*

Three metropolitan areas account for one third of Canada's total population: Toronto, Montreal, and Vancouver.

These kinds of questions have figured prominently in urban studies for several generations. The subfield of urban geography concerned with these questions is the "urban systems" literature, and for many years its theories and explanations provided a key organizing framework for understanding cities and for planning certain aspects of urban life.[5] The approach remains an important area of theoretical, empirical, and methodological debate today; but it can no longer claim the distinction of providing the grand organizational scheme for the study of cities.

Cities are systems. Sets of cities are also systems. An ***urban system*** *is a network of interdependent urban places.*

Urban Systems

In the most general terms, an *urban system* may be defined as any network of interdependent urban places. Any significant change in one city will have consequences for other cities in the system. The phrase was introduced by Brian J.L. Berry in an article published in the early 1960s, under the curious title, "Cities as Systems within Systems of Cities."[6] In an ambitious attempt to summarize a decade of new kinds of research

---

[2] Larry D. McCann and Jim Simmons (2000). "The Core-Periphery Structure of Canada's Urban System." In Trudi Bunting and Pierre Filion, eds., *Canadian Cities in Transition*, Second Edition. Don Mills, ON: Oxford University Press, 76-120, quote from p. 98.

[3] See Trudi Bunting and Pierre Filion (2010). "Epochs of Canadian Urban Development." In Trudi Bunting, Pierre Filion, and Ryan Walker, eds., *Canadian Cities in Transition*, Fourth Edition. Don Mills, ON: Oxford University Press, 19-38, city statistics on p. 23.

[4] Larry S. Bourne and Jim Simmons (2003). "New Fault Lines? Recent Trends in the Canadian Urban System and Their Implications." *Canadian Journal of Urban Research* 12(1), 22-47, quote from p. 25.

[5] Indeed, at one point the prominence of urban geography combined with the popularity of urban-systems thinking meant that some observers saw the framework as providing a coherent framework in which to understand *all* of human geography; some even hoped for a grand synthesis of the theories and methods of both human and physical geography. Richard J. Chorley and Peter Haggett, eds. (1967), *Models in Geograp*hy. London: Methuen.

[6] Brian J.L. Berry (1964), "Cities as Systems Within Systems of Cities." *Papers of the Regional Science Association* 13, 147-163.

undertaken by a variety of scholars, Berry argued that new computing technologies and new sources of data were making it possible for the first time for urbanists to develop genuinely scientific theories comprising two distinct parts: "(a) simple inductive generalizations drawn from observable facts about the world, and (b) abstract logical constructs. ... Ten years ago, urban studies were in an either/or situation -- either inductive generalizations or logical constructs existed, the former as likely as not produced by urban geographers and the latter by urban economists. ... The importance of the last decade has been that the twain have met through the medium of regional science. Moreover, the meeting came just when quantitative methods of analysis, facilitated by rapid developments in computer technology, began a technological revolution which has wrought havoc throughout the sciences."[7]

Berry sought to understand urbanization by borrowing from general systems theory, an approach that emerged in the natural sciences to address superficially different processes that could be ultimately traced to similar causes; Berry argued that "cities and sets of cities are *systems* susceptible of the same kinds of analysis as other systems and characterized by the same generalizations, constructs, and models."[8] Although this all sounds quite abstract, the approach actually began with a very simple puzzle that had been the topic of considerable curiosity for many years. In the 1940s, George Zipf[9] identified a curious regularity in the distribution of cities of varying sizes; if cities were ranked from largest to smallest populations, then

$$P_r = \frac{P_1}{r^q}$$

which simply says that the r-ranked city (P_r) will be expected to have a population equal to the top-ranked city divided by the rank, r. Usually, q remained very close to 1, such that it would be a simple matter to, say, predict the size of the second-ranked Canadian city in 1871 (Quebec City) by dividing the population of the largest city (Montreal, 115,000 people) by the rank of the second-largest city (2). This theoretical guide, the rank-size rule, predicts a population of 57,500 for Quebec City, not far off from the actual figure recorded at the time (60,000). If we estimate this rule for all cities in a region or country, and transform everything to a linear form, the relationship can be expressed as

The rank-size rule:

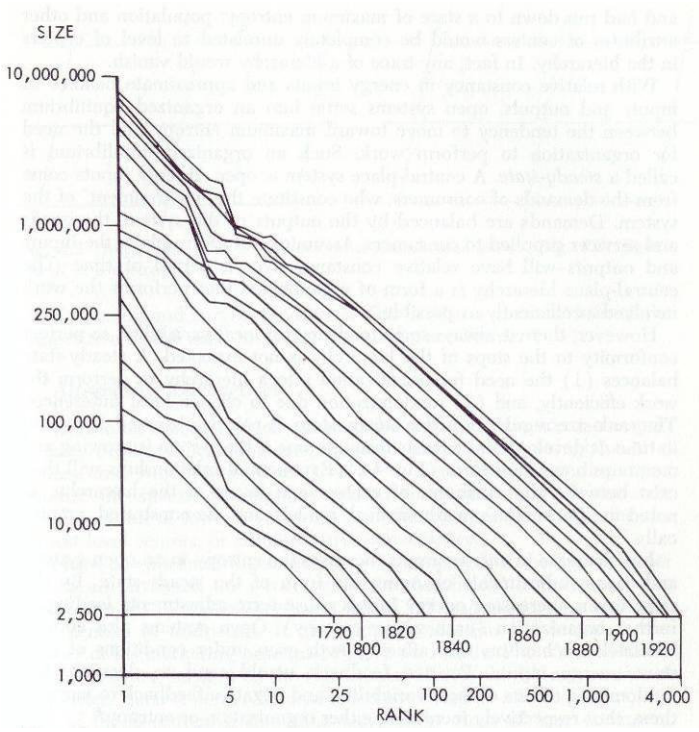
$$P_r = \frac{P_1}{r^q}$$

$$\log P_r = \log P_1 - q \log r$$

To the degree that this kind of model fits the observed data reasonably well, it means that the populations of smaller cities in any given system follow a log-linear relationship with the most populous city in the system; and the relationship can be quantified by the slope coefficient, q. In other words, if you rank the cities in a system according to their population, and you plot the

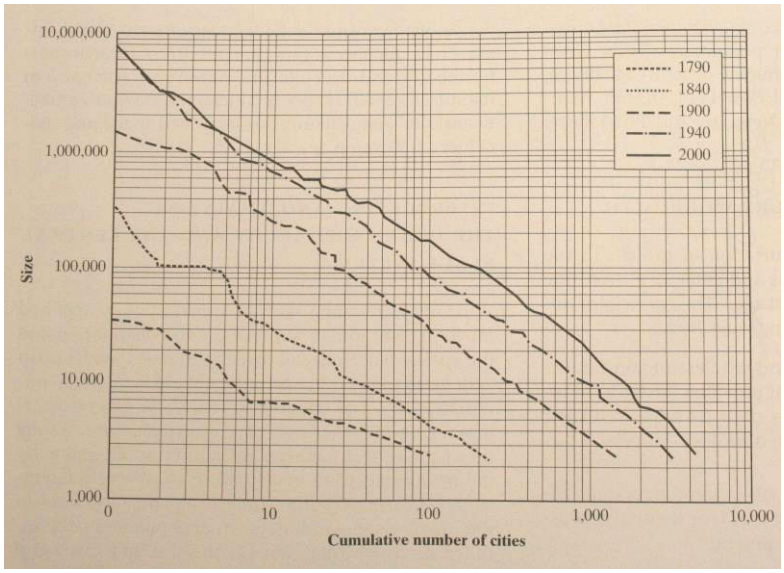
[7] Berry, "Cities as Systems," p. 147.
[8] Berry, "Cities as Systems," p. 158.
[9] George K. Zipf (1941), *National Unity and Disunity*. Bloomington, IN: Principia Press.

results on log-log graph paper, you will get something close to a straight line. Berry performed the laborious calculations to calibrate this equation for every census of the population in the United States between 1790 and 1950, and this was the result:



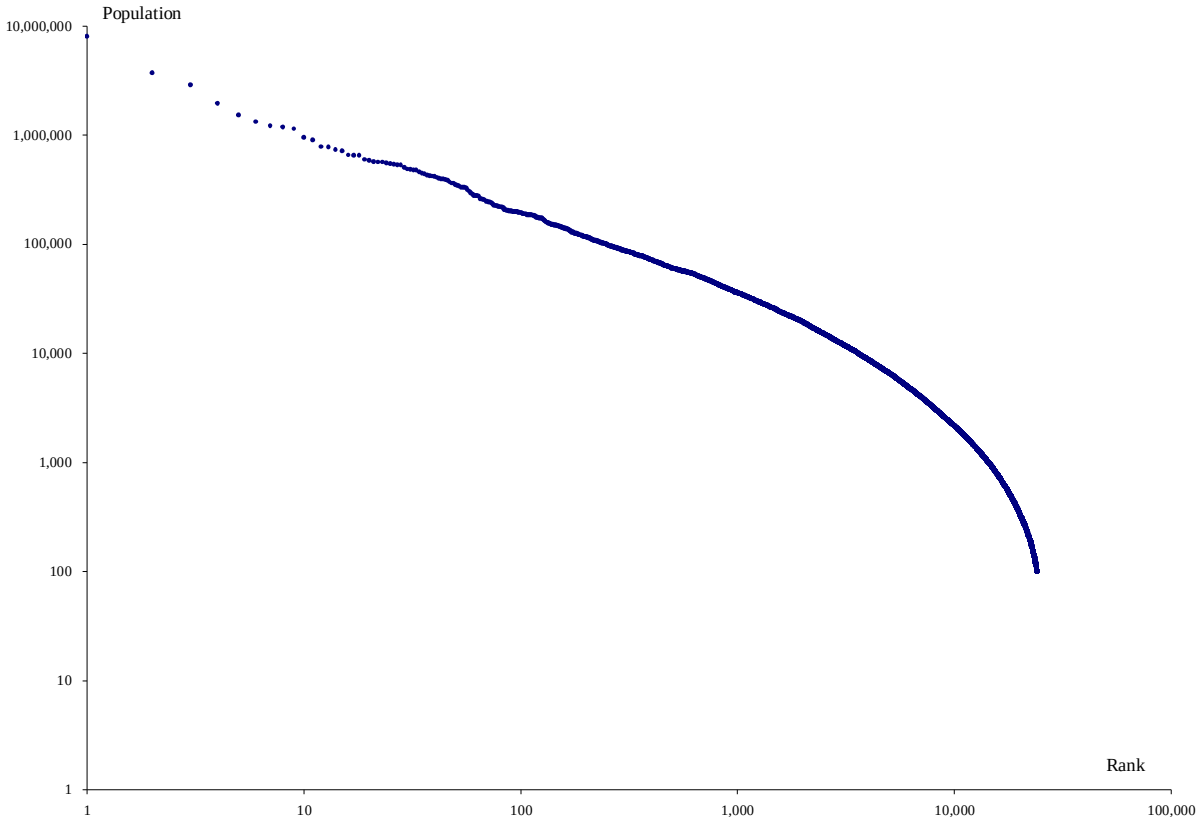
Rank-size Distribution of U.S. City Populations, 1790-1950. Because of distinctive features in the way the U.S. Census Bureau collects and reports data, Berry restricted his attention to places with populations over 2,500. *Source:* Brian J.L. Berry (1967), *Geography of Market Centers and Retail Distribution.* Englewood Cliffs, NJ: Prentice-Hall, p. 77. Reproduced pursuant to Sections 29 ("Fair dealing for the purpose of research, private study, education, parody, or satire") and 30.04 ("work available through Internet") provisions of Canada Bill C-11.

Paul Knox and Linda McCarthy updated this procedure with data for the year 2000, and the pattern persists -- another line appears a bit farther out from the cluster of lines portraying the mid-twentieth century pattern.



U.S. Rank-Size Distribution, to 2000. Source: Paul Knox and Linda McCarthy (2005). *Urbanization.* Upper Saddle River, NJ: Prentice-Hall, p. 66. Reproduced pursuant to Sections 29 ("Fair dealing for the purpose of research, private study, education, parody, or satire") and 30.04 ("work available through Internet") provisions of Canada Bill C-11.

If we update these graphs for the 2010 Census, the pattern seems to look a bit different:



Rank Size Distributions of U.S. Cities and Towns, 2010. *Data Source:* U.S. Bureau of the Census (2011). *2010 Census of Population and Housing, Summary File 1*. Washington, DC: U.S. Department of Commerce.

On closer inspection, however, it is clear that the rank-size stability holds quite well for cities and towns above the 1,000 population threshold. The slope of the curve shifts notably below this point, indicating perhaps a fundamental break between the smallest towns and the rest of the national urban system.

Three features of these graphs stand out.

First, notice that over time, the entire system -- in this case, the entire network of urban places in the U.S. -- expands, with the lines moving up and to the right. Berry observed, "Much analysis in the quarter-century since Zipf did his work leads to the conclusion that the regularity will result wherever the rate of relative population growth of centers at any level of the hierarchy is, on the average, a constant fraction of the rate of relative population growth of the entire hierarchy of centers."[10] In other words, the nearly-straight lines on this graph can be understood as the "equilibrium" or steady-state when the growth rate of each city-size category stays the same in relation to national trends. If small cities suddenly accelerate over time, the lines on this chart will have a more gentle slope. Conversely, if the largest cities suddenly grow faster than the smaller ones, then the lines will show a steeper slope.

In graphs of rank-size city size distributions:

1. Straight lines suggest a growth equilibrium.
2. Deviations or interruptions suggest "primacy" or exogenous shocks.
3. Individual cities can rise or fall, but what endures are the ***relations*** *of the* ***system****.*

Second, note that the lines are never perfectly straight: when this procedure is applied to different regions or countries, the results include "interruptions" in the rank-order patterns (as shown above), as well as widely varying slopes. Urban systems in which the most populous city is much larger than predicted (e.g., far more than twice as populous as the second-ranked city) are said to show a bias of "primacy," with the large, primate city often the result of "exogenous" forces such as colonial dependency that led to disproportionate concentration of economic activity and population into a single center. Primacy has a non-linear relationship to measures of national wealth and economic development. Concentrating population and economic activity in cities brings great efficiency gains; but as the size of the largest city increases, diminishing returns may set in. The efficiency benefits may be cancelled out by the costs of congestion, pollution, and all the other side effects of size and

Urban Primacy

Countries with the most extreme measures of urban primacy. Primacy is here calculated as: the ratio of the population of the country's largest city to the combined population of the next two largest cities. For the example of a country with no unusual primacy (i.e., where the rank-size rule fits perfectly) and a population of the largest city of 1 million, the ratio would be 1.20 (1,000,000 divided by [500,000 + 333,333]).

Highest	
Thailand	9.48
Suriname	8.24
Togo	7.92
Uruguay	7.37
Chile	5.98
Uganda	5.94
Ethiopia	5.82
Mongolia	5.67
Peru	5.43
Guinea	5.27

Lowest	
Benin	0.58
South Africa	0.59
Venezuela	0.65
The Netherlands	0.70
Egypt	0.72
Australia	0.76
China	0.78
United States	0.84
Bolivia	0.84
India	0.86
Canada	0.89

Source: John Rennie Short and Luis Mauricio Pinet-Peralta (2009). "Urban Primacy: Reopening the Debate." Geography Compass 3(3), 1245-1266.

[10] Berry, *Geography of Market Centers*, p. 76.

density. Still, there do seem to be some general patterns. The most extreme cases of urban primacy tend to be in low-income countries. The top-ranked cases of primacy include no high-income countries, but the list of countries with unusually low primacy does include both poor and wealthy countries.

Third, observe that there are no labels identifying particular cities, and for good reason. Individual cities can rise or fall in the rankings: *what endures are the relations, hierarchies, and the system*. Consider once again where we began. In 1871, Montreal was the largest city in Canada; a century later, Montreal remained at the top of the list. But shortly thereafter Montreal was displaced by Toronto. In fact, if we examine the ten most populous urban places in Canada in the century or so since 1871, we can see a number of cities changing positions[11]:

Changes in the Rank-Order of Canada's Largest Urban Places, 1871-1981.

1871		1921		1971		1981	
Montreal	115,000	Montreal	724,000	Montreal	2,743,000	Toronto	2,999,000
Quebec City	60,000	Toronto	611,000	Toronto	2,628,000	Montreal	2,828,000
Toronto	59,000	Winnipeg	228,000	Vancouver	1,082,000	Vancouver	1,268,000
St. John	41,000	Vancouver	223,000	Ottawa-Hull	603,000	Ottawa-Hull	718,000
Halifax	30,000	Hamilton	114,000	Winnipeg	540,000	Edmonton	657,000
Hamilton	27,000	Ottawa	108,000	Hamilton	498,000	Calgary	593,000
Ottawa	24,000	Quebec City	95,000	Edmonton	496,000	Winnipeg	585,000
St. John's	23,000	Calgary	63,000	Quebec City	481,000	Quebec City	576,000
London	18,000	London	61,000	Calgary	403,000	Hamilton	542,000
Kingston	12,000	Edmonton	59,000	Windsor	259,000	St. Catharines-Niagara	304,000

In 1871, the top of the urban hierarchy conforms quite well to the rank-size rule: the second-largest city is about half the size of the largest. But for almost a century, the top of the list has been occupied by two dominant national metropoli, and so all of Zipf and Berry's theoretical elegance and *q*-coefficients are rather useless for understanding this part of the urban system; rather, all we need are a few elementary reminders of the historical and functional roles of Montreal and Toronto in Canada's provincial and federal relations. Nevertheless, the urban systems approach is not entirely useless. In 1871, the ratio of the top-ranked city (Montreal) to the tenth-largest city (Kingston) was 9.58, quite close to the rank-predicted value of 10.00. More than a century later, despite the enormous historical-geographical transformation of all of Canada as well as its individual cities, this ratio, now contrasting Toronto to St. Catherine's-Niagara, stands at 9.86. *Why?*

Explanations for Urban System Stability

We are presented with an empirical regularity that has been observed and debated for more than half a century. Some urbanists are fascinated by the question and its implications; others are consumed by boredom at yet another graph and system of equations. In the end, explanations for urban systems can be divided into two broad alliances. On one side are theorists who generally emphasize the order, regularity, symmetry, and predictability of the empirical relationships. On

[11] Table source: Maurice W. Yeates (1990), *The North American City*, Fourth Edition. New York: Harper & Row, Table 2.5, p. 61.

the other side are those who emphasize unique circumstances, historical context, and regional contingency.

General Systems Theory Explanations

For those inclined to search for regularity and organization in spatial patterns, rank-size stability is compelling evidence of some kind of systemic process. **Innovation diffusion theory** has provided one persuasive explanation. Building on the observation that smaller cities tend to have much more volatile rates of growth and decline when compared to the more stable rates in larger cities, many scholars hypothesize that information and competition maintain an overall status quo:

General systems *explanations for urban systems stability:*

1. Innovation diffusion theory.

2. Central place theory.

"Over the medium to long term, urban population growth is a function of the number of innovations originating within a city's economy or successfully adopted from outside sources. Larger cities, with larger economies, are likely to produce more innovations, thus attracting migrants to the city and helping to ensure that residents do not leave for other cities. Therefore, the sheer size of large cities ensures a strong and steady stream of innovations, which in turn contributes to a steady source of population growth. Smaller cities must depend more on adopting innovations that diffuse from other, usually larger, centers."[12]

The logic of innovation diffusion theory might seem quaint and antiquated. Indeed, the landmark research on the role of innovation diffusion in creating and sustaining a national urban system was done by Allan Pred in the 1970s, working with nineteenth-century records on railroad and telegraph networks and bank records to trace the paths of financial "panics" as a way of inferring lines of regular communication.[13] And yet even today, in a climate where the Internet seems to have conquered spatial constraints, innovation is not entirely freed from the confines (and possibilities!) of place. What viable city does not want to have its own version of 'Silicon Valley,' the suburban office complexes south of San Francisco that became a hotbed of computer innovation beginning in the 1950s, and achieved globalized notoriety in the dot-commodified rage of recent years? A broad stream of recent research documents the continuing importance, even in today's digitally-lubricated global economy, of agglomeration economies: benefits that companies gain by locating near functionally-related companies. Another stream of research is now demonstrating how these effects matter not just for firms, but for workers. In a recent and valuable comparative study of Vancouver, London, Cologne, San Francisco, Seattle, and Singapore, Tom Hutton emphasizes that economic agglomeration benefits "are

[12] Paul Knox and Linda McCarthy (2005). *Urbanization*. Upper Saddle River, NJ: Prentice-Hall, p. 66.
[13] Allan Pred (1977), *City-Systems in Advanced Economies: Past Growth, Present Processes, and Future Development Options*. New York: Wiley.

complemented by what we can broadly describe as the unique environment of certain inner city districts."[14] This environment is especially important to "new economy" and "high-tech" creative workers, who in any event are often defined by the blurred boundaries between work, play, education, recreation. Entrepreneurial innovation is the shared element, and for this principle, certain inner city environments have precisely the right environment, "a rich, complex, and interdependent set of attributes, including the *built environment*, particularly heritage buildings and architecturally distinctive structures which comprise the essential infrastructure and working spaces for New Economy firms in many cities; *cultural amenity* including galleries, museums, exhibition spaces, and ... the 'historical legacy' of certain inner city spaces ...."[15]

But if innovation diffusion theory explains one question, it remains silent on another. Entrepreneurialism and size imply that cities that get to be large early on will, all other things being equal, stay large because of their rich environment for new innovations. But why are there only a few spaces for the largest cities? And why do maps often show a regular lattice-work of many small settlements, interrupted by a few larger cities, all of them in the shadow of one dominant metropolis? What explains *this* regularity? The second main general-systems explanation, **central-place theory**, emerged from the work of a German geographer, Walter Christaller, who sought to explain what he saw as a striking pattern of settlements in Southern Germany in the 1930s. Christaller explained the regular size and spacing in terms of the roles that settlements play in serving trade areas, in which people have needs for very different kinds of goods and services. These different needs arise because each good or service has a particular "range" and "threshold." The *range* of a particular good or service is the maximum distance that an average consumer is willing to travel for it under typical circumstances; "higher-order" goods and services like specialized medical care have ranges of hundreds or thousands of kilometers; "lower-order" goods include everyday needs (groceries, convenience stores, fast-food restaurants, bars) for which people are only willing to travel short distances. If we view these principles not from the consumer's perspective but from the merchant's or service-provider's, the question becomes: how large must the market be to make my business viable? This minimum market size is referred to as the *threshold*. From these first principles, Christaller deduced a network of small villages emerging to serve market areas, which could be defined as circles each with a radius equivalent to the range, and (under circumstances of an evenly distributed population) each including a threshold large enough to support village-like trade activities. Circles overlap and leave some areas unserved, however, and so over time,

Central place theory: an explanation of the distribution of large and small settlements as the result of the economics of serving trade areas. A hierarchy emerges that reflects differences among goods and services. Consumers are willing to travel very far for some goods and services, for example, but not for others.

[14] Thomas A. Hutton (2004), "The New Economy of the Inner City." *Cities* 21(2), 89-108, quote from p. 93.
[15] Hutton, "The New Economy," p. 93.

competition between trade centers creates a network of hexagons. Some of these hexagons win out in the competition to provide higher-order goods and services, and the result is a nested hierarchy of numerous small villages, fewer but larger towns, and only a very few large cities. Relax a few of the assumptions involving uninterrupted topography and an even distribution of population, and the hexagons squish around a bit to resemble Christaller's classic analysis of the Bavarian landscape (see the map on the next page).

The technical procedures required to work out central place theory, and to test its ability to predict the size, spacing, and location of settlements, obsessed an entire generation of geographers. Those who came to see the world through hexagonal-shaped glasses contributed a great deal of creativity and insight. Eventually, however, it became clear that the approach a) required major adjustments in order to account for the variations of historical and geographical context, b) accounted only for the final stage of retail distribution, and c) worked best in those places where production and consumption obeyed a neat hierarchy from local to regional to national to international. But as Knox and McCarthy observe,

> "Many central place functions, even in the early nineteenth century, depended on a long chain of production and exchange that involved numerous steps, from harvesting or mining through processing, packaging, shipment, and storage to retail distribution. ... Today, the internationalization of the economy means that almost every purchase we make is the result of a complex geography of international specialization and trade involving vast quantities of products whose origins range from makeshift kitchen workshops in the least developed countries to the giant electronics factories of South Korea and the perfumeries of Grasse (in Provence, France)."[16]

McCann and Simmons, in a chapter in *Canadian Cities in Transition*, make a similar point:

> "No city today functions free of the unprecedented forces currently restructuring the world system. As the organizational units of the world economy, all cities are interconnected through an intricate geography of core and periphery shaped by various economic, demographic, and political processes."[17]

[16] Knox and McCarthy, *Urbanization*, p. 68.
[17] McCann and Simmons, "The Core-Periphery Structure," p. 76.

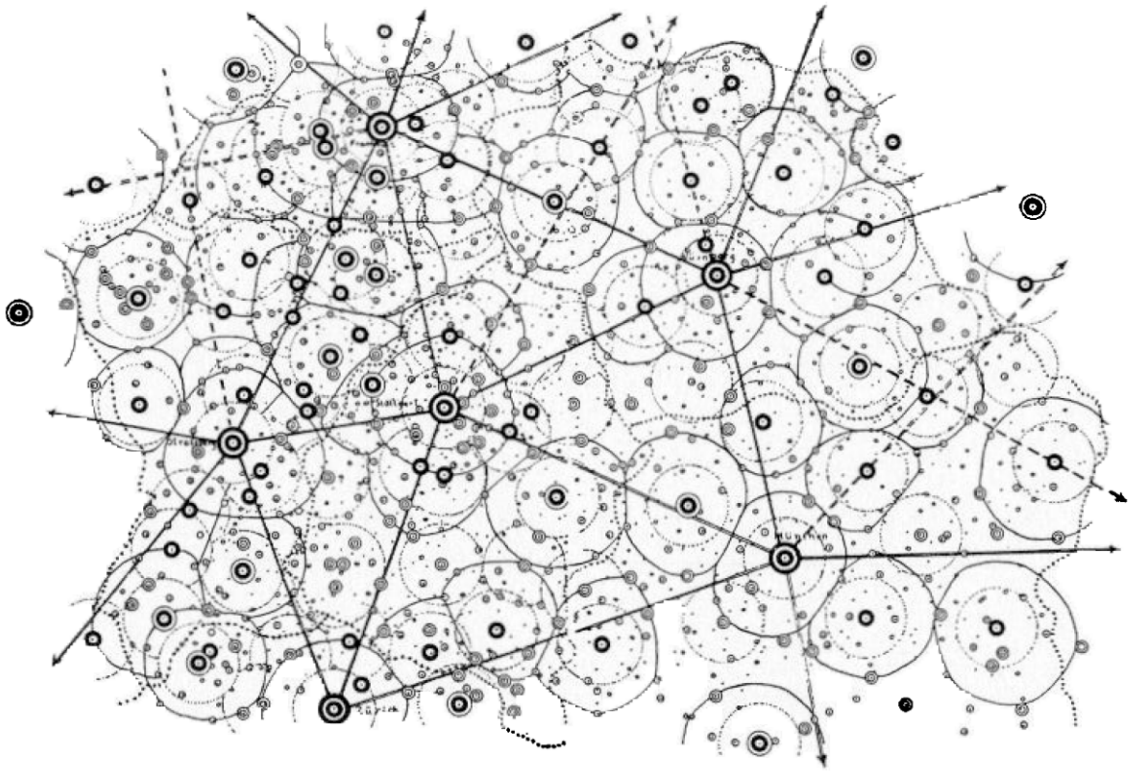


Figure 4. Christaller's Map of Central Places in Southern Germany, early 1930s. Source: reproduced in Peter Gould (1986), *The Geographer at Work*. London: Routledge and Kegan Paul, p. 92. Reproduced pursuant to Sections 29 ("Fair dealing for the purpose of research, private study, education, parody, or satire") and 30.04 ("work available through Internet") provisions of Canada Bill C-11.

Regional-Historical Explanations

While general-systems theorists emphasize order, symmetry, and predictability, many geographers working in the regional tradition give more attention to the unique circumstances and rich histories behind the settlement and growth of villages, towns, and cities. Two main regional-historical theories have been developed to explain urban system growth and change.

The first comes from **staples theory**, an influential account of economic growth and change developed by Harold Innes (1894-1952), a Canadian economic historian. Innes suggested that change and growth in any regional economy can be understood in terms of the timing and location of staples discovery and exploitation: places where valuable natural resources are discovered and utilized for export earnings enjoy, for a time, a windfall that can be invested in other economic activities to build a solid foundation for diversification, industrialization, and sustained urban and regional growth. As Trevor Barnes writes, however, this developmental scenario can also have a pessimistic twist: places can become trapped in a dependent resource economy, as

> "diversification is trapped for reasons such as an export mentality among producers, the domination of the economy by a few, and often foreign-owned

multinational corporations, and a truncated industrial branch-plant structure that minimizes the development of higher-order control and research functions."[18]

Places that come to dominate staples markets are quite vulnerable: the markets for natural resources more closely resemble the 'perfect competition' of neoclassical economic theory than the markets for manufactured goods, creating considerable volatility in earnings; the reliance on foreign exports usually entails the dominance of foreign multinational firms; and long-run technological changes and efficiencies usually reduce the terms of trade for staples. As a consequence, staples-based settlements may enjoy an "accumulation of cyclonic frenzy" but "such intense accumulation ... never lasts, and because of the very instabilities of staples production, sooner rather than later investment shifts to yet other staples regions, leaving in its wake abandoned resource sites and communities."[19]

Regional-historical *explanations for urban systems stability:*

1. Staples theory.

2. Economic base theory.

A variation on the theme of staples comes from **economic base theory**, which retains the notion of a leading, specialized sector driving diversification and urban growth -- but considers all kinds of industries. In economic base theory, the *basic sector* is any economic activity in which a particular place specializes, and exports to other regions or cities; the *non-basic* sector consists of economic activities oriented towards local consumption (for instance, the mundane assortment of lower-order retail and service functions that all cities have). A city's basic sector is regarded as the engine that drives local growth, since it is this sector that attracts "new" earnings from other places; non-basic activities, by contrast, simply reallocate money that is already in the local economy. It is also important to recognize that a city's exports include not only the physical commodities actually sent to other places, but also unique services or amenities (Victoria's exports, for instance, include provincial governmental 'services,' educational services, and tourist amenities). Usually, analysts measure the basic sector by identifying the type of industry that is disproportionately concentrated in a particular city or region; then a battery of econometric tools can be used to estimate the economic linkages between the basic and non-basic sectors of the local and regional economy.[20] But the technical details should not obscure the fundamental simplicity. Any map that classifies cities according to different economic functions is implicitly based on the ideas of economic base theory: McCann and Simmons, for instance, present a map of the economic specialization of Canadian cities, which shows a broad scattering of agriculture-based cities (primarily in the prairie provinces, on the Canadian Shield across central Ontario and Quebec, and across the Maritimes), resource-based cities (disproportionately concentrated in the Maritimes and B.C.), commerce-based cities,

[18] Trevor J. Barnes (2000), "Staples Theory." In R.J. Johnston, Derek Gregory, Geraldine Pratt, and Michael Watts, eds., *The Dictionary of Human Geography*. Oxford: Blackwell, 786-788, quote from pp. 786-787.
[19] Barnes, "Staples Theory," p. 788.
[20] Norman J. Glickman (1977), *Econometric Analysis of Regional Systems: Explorations in Model Building and Policy Analysis.* New York and London: Academic Press.

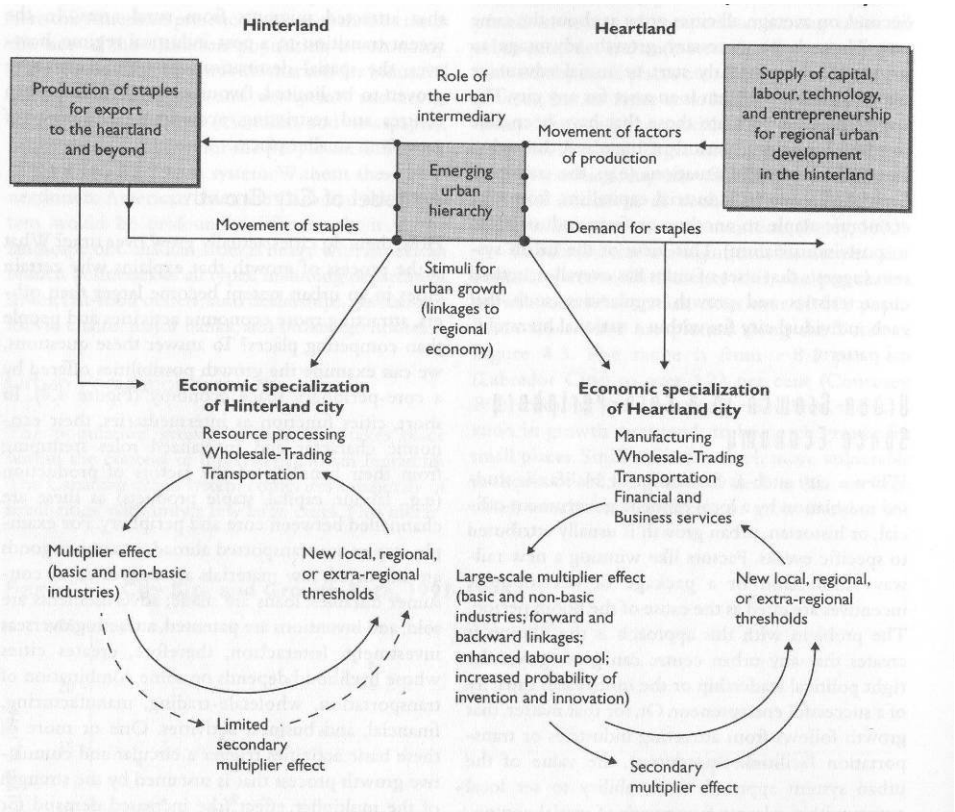
public-sector cities, and a dense cluster of manufacturing-oriented cities in southern Quebec and southern Ontario.[21]

Beyond the Dichotomy

The simple division between order/symmetry versus history/context is helpful in understanding some of the differences in emphasis of alternative theories. In reality, however, many urban geographers have sought to transcend such dualisms, integrating elements of abstract, deductive theory with inductive, empirical generalization. James E. Vance, Jr., for instance, developed a synthesis of elements of staples theory, economic base theory, innovation diffusion theory, and central-place dynamics to account for the settlement history of North America. Vance's **mercantile model** portrays settlement growth in a series of distinct phases. A first stage of exploration and search for knowledge by European colonial powers, Vance argued, initiated a subsequent wave of resource extraction (fish from Newfoundland, timber and furs from New England); later colonization created a network of local staples economies (exporting tobacco, cotton, indigo, etc.) that also served as markets for the nascent manufacturing output of Europe. Over time, "continuing demand for staple exports and increasing colonization combine to draw settlement further into the interior" of North America, "requiring the development of long-distance routes and the emergence of towns serving as 'depots of staple collection' at strategic locations along these routes."[22] Finally, a phase of economic maturity fills in the central place network, with the "gateways" -- the 'depots of staple collection' enjoying early advantages in the growing networks of long-distance trade; in between the largest centers, agricultural settlements evolve into a fairly orderly hierarchical system.

It is possible to integrate the deductive, general-systems theories and the inductive regional-historical narratives.

[21] Simmons and McCann, "The Core-Periphery Structure," p. 79.
[22] Knox and McCarthy, *Urbanization*, p. 59.



The Canadian Heartland-Hinterland Model. *Source:* Larry D. McCann and Jim Simmons (2000). "The Core-Periphery Structure of Canada's Urban System." In Trudi Bunting and Pierre Filion, eds., *Canadian Cities in Transition*, Second Edition. Don Mills, ON: Oxford University Press, 76-120, figure from p. 86. Reproduced pursuant to Sections 29 ("Fair dealing for the purpose of research, private study, education, parody, or satire") and 30.04 ("work available through Internet") provisions of Canada Bill C-11.

Similarly, McCann and Simmons develop a **heartland-hinterland** model of city growth that fuses key aspects of staples theory, innovation diffusion theory, and economic base theory. Focusing on Canada, McCann and Simmons distinguish between the nation's a) dominant core area -- the densely settled and highly urbanized corridor between Quebec City and Windsor; b) resource ecumene (the Maritimes, the Canadian Shield, most of Western Canada), and c) the sparsely settled Far North. All of these broad regions have cities, of course, but the roles available to each settlement are constrained by regional and national context. Within these constraints,

> "cities function as intermediaries, their economic character and specialized roles stemming from their handling of the factors of production (e.g., labour, capital, staples products) as these are channeled between core and periphery. ... staples are transported abroad; consumer goods are imported; raw materials are converted to consumer durables; loans are made; advertisements are sold; and inventions are patented, attracting overseas investment. Interaction, therefore, creates cities whose livelihood depends on some combination of transportation, wholesale-trading, manufacturing, financial, and business activities. One or more of these basic activities trigger a circular and cumulative growth process that is sustained by the strength of the multiplier effect (the increased demand for services and,

hence, for more service workers that is engendered by the entry of more employees into the economy) and the threshold size of critical markets."[23]

Paul Knox and Linda McCarthy take a slightly different route to a similar explanation, synthesizing many of the same theoretical literatures in a **self-propelling model of urbanization** that we considered in our evaluation of the rise of the North American "manufacturing belt." Any initial growth of an industrial establishment generates a variety of localized consequences, including a) backward linkages (as new establishments spring up to provide components, supplies, or services) b) forward linkages ("downstream" firms that buy the finished products of one firm and use it as an input to their own operations), c) the development of a skilled and experienced local work force with ties to the distinctive local industries, and d) a balance of competition and cooperation amongst local companies, such that there is growing support for research and development activities that will enhance local innovation and regional advantage. Taken together, these processes create powerful multiplier effects, provide an expanded tax base which can be invested in utilities, education, and other collective needs that will strengthen urban competitive position, and drive a process of "cumulative causation."

To be sure, this entire model is only a generalization, and viewing it as a definitive explanation brings us down the slippery slope of tautology. Modifying processes are crucial. Being near a city going through a "virtuous circle" of cumulative causation can provide an unexpected shower of trickle-down demand (spread effects), but under other circumstances proximity can draw away local vitality (backwash effects: Why would a company locate in Oakland, California when it could be in San Francisco? Why choose Newark, New Jersey, when New York City is just across the river?) And the entire virtuous circle of growth can also work the other way. The costs, congestion, and pollution of cumulative causation can create powerful agglomeration diseconomies; and external shocks to the system can also lead to a downward spiral.

Ultimately, the self-propelling model of urbanization must be seen in historical-geographical context. Just as staples theory can be understood as distinctively suited to the urban history of Canada in the nineteenth and early twentieth centuries, the cumulative causation model is particularly well-suited to what is often regarded as the post-World War II "Golden Age" of American capitalism, running from the mid-1940s to the (first) oil crises of 1973-1974.

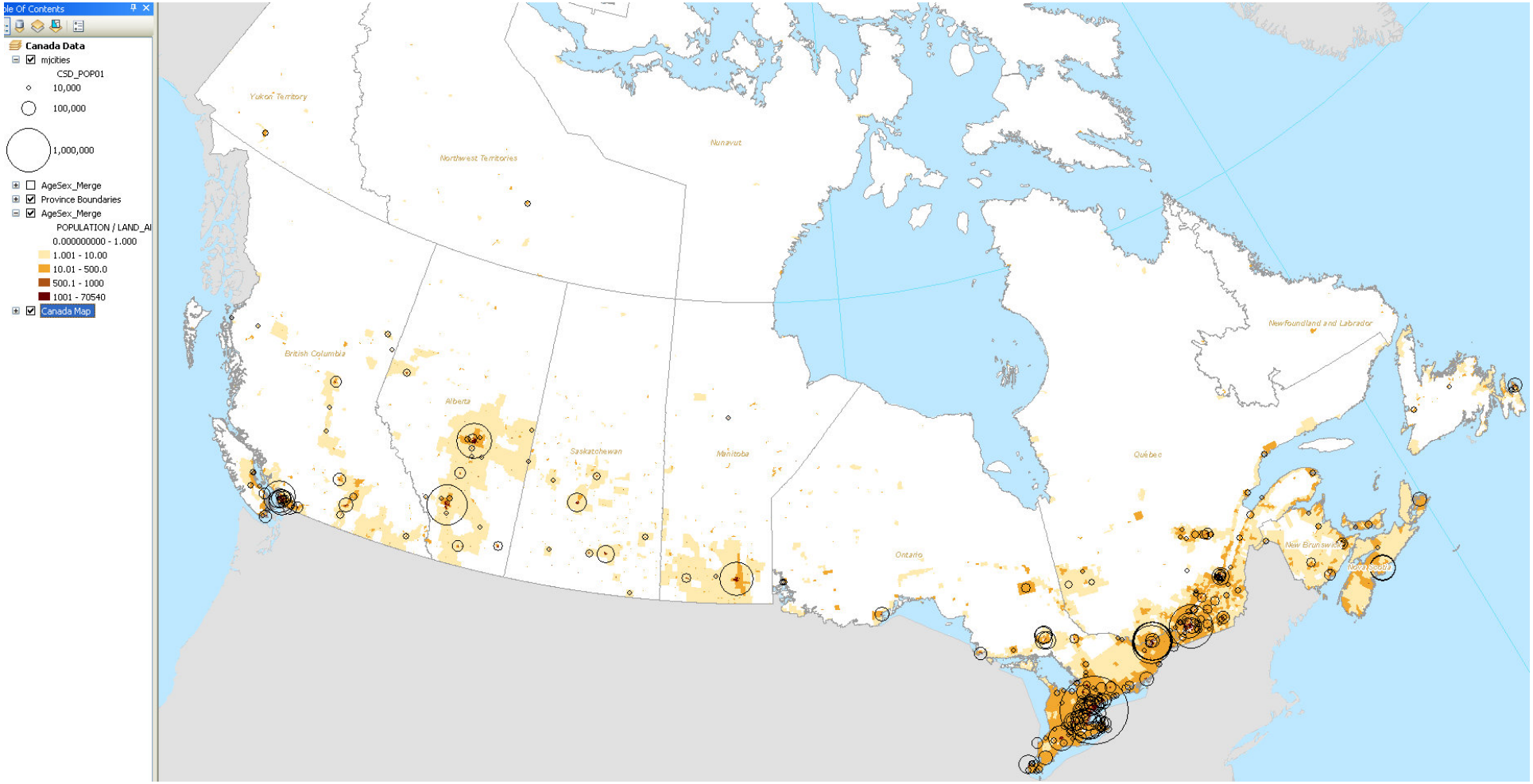
Even so, key facets of this model have been revised and updated in a new generation of research by urban economists, some of whom are working to revive the kind of thinking pioneered by Zipf and Berry half a century ago. Duncan Black and Vernon Henderson, for instance, model the evolution of the urban system in the United States as a process of *endogenous* urban growth -- predicting the relative growth of different urban areas as a function of their endogenous, essential internal characteristics.[24] Black and Henderson place special emphasis on production specialization and the competition of innovation rediscovering some of the historical insights

[23] McCann and Simmons, "The Core-Periphery Structure," p. 85.
[24] Duncan Black and Vernon Henderson (2003). "Urban Evolution in the USA." *Journal of Economic Geography* 3, 343-372.

developed by Allan Pred.[25] Echoing the ideas of cumulative causation, Black and Henderson marshal evidence to show that continued urban growth is driven by technological innovation and the accumulation of human capital -- the skills and creativity of firms and workers in each regional economy. "Urban evolution, as predicted by an urban endogenous growth model, involves on-going increases in absolute metro area sizes," thanks to "technological change represented by local human capital accumulation."[26] But specialization also matters: "Cities of different types, or industrial compositions, tend to have different absolute sizes, as well as educational attainments, with relative size changing as industrial composition changes."[27] They also find that with the entry of new cities into the system, it is possible for some places to move up the rankings very quickly; but cities fall from the top tier much more slowly. Most of those older, larger cities at the top of the hierarchy are the old industrial centers of the Northeast and Midwest; Black and Henderson attribute their survival to the rich market potential associated with the close-knit pattern of settlement in this part of the country. And so all cities in the system are increasingly subjected to increasing competitive pressures to innovate as trade flows and production complexes cross national boundaries.

> "Under an urban product-cycle hypothesis, product development occurs in major metro areas with their diverse industrial-occupational environments. Once products and production processes are standardized, production is decentralized to smaller, perhaps more specialized cities with their lower wage and land costs .... In the US, this pattern has been altered in recent decades as is evident from the relative growth in imports and decline in manufacturing. Standardized production is now decentralized not so much to smaller US cities, but more to other countries. The overall composition of US production has shifted toward the service sector -- services in finance, research, management and consulting, law and education, engineering and architecture, and business (advertising, credit computer, personnel, etc.). ... a number of these activities tend to be disproportionately located in large metro areas ... enabling increased urban concentration."[28]

[25] Although it is worth noting that Black and Henderson do not, in fact, mention Allan Pred's earlier work. They do acknowledge Zipf and Berry, but draw much of their foundational theoretical inspiration from the economist Paul Krugman and his colleagues.
[26] Black and Henderson, "Urban Evolution," p. 366.
[27] Black and Henderson, "Urban Evolution," p. 366.
[28] Black and Henderson, "Urban Evolution," p. 344.



Heartland-Hinterland Contrasts in Canada. About 70 percent of Canada's population lives within 150 km of the border with the United States, and about 80 percent lives in towns or cities with at least 1,000 residents. Most of Canada's land area has a population density below 1 person per square kilometer. *Data Source:* Statistics Canada (2010). *Cumulative Profile for Census Subdivisions and Census Metropolitan Areas, 2006 Census.* Ottawa: Statistics Canada.

'No Convincing Explanation'?

We began with the simple observation of empirical regularities, such as the varying number of cities of different sizes: for McCann and Simmons, "This hierarchical city size distribution is an almost universal feature in all countries across time, even when urban systems are in transition...."[29] And yet, summarizing more than half a century's worth of research since George Zipf first penned the equation for the rank-size rule, Ron Johnston concludes, "No convincing explanation for the existence of the relationship has been developed, despite the frequency with which it is observed."[30]

This seems a bit harsh. All of the approaches we've reviewed *do* provide valuable insights, and they do provide persuasive -- if necessarily partial and contextual -- explanations for certain regularities in urban systems. The search for a unifying "general systems theory" explanation of everything was called off long ago; but that does not leave us in a world of complete unpredictability and chaos. It's a fair bet that no matter how fast Courtenay may grow (it posted the fastest growth of all Canadian cities between 1991 and 1996), it's not likely to climb its way to the very top of the urban system to compete with Toronto, Montreal, and Vancouver.[31]

*The **explanations** for urban systems stability remain imperfect and perhaps even mysterious; but in the realm of **policy**, urban systems analysis is used widely to plan cities, regional territories, and corporate strategies. It works, although it's not always clear <u>why</u>.*

But in the end, the competing *explanations* for the historic roots of today's urban systems may be giving way to urgent contemporary dilemmas of *policy*. Even if the ultimate reasons for urban systems stability remain unclear, the framework is used widely to understand existing urban patterns and to plan new ones. Consider a few examples:

1. Israeli settlements to the east of the Gaza Strip were based on a hierarchy of 'A' settlements, including protective border kibbutzim, used by settlers on a daily basis; 'B' settlements that function as 'rural community centres' used by most settlers only once or twice a week; and 'C' regional centres located "roughly at the geographical centre of their region, providing administrative, educational, medical, and cultural facilities, and with factories for crop-processing."[32] Settlements in the West Bank have been, for many years, one of the key points of struggle in the long-running Palestinian-Israeli conflict. Central-place theory and urban-systems analysis may not be mentioned specifically in the news headlines. But behind the scenes they are quite important.

---

[29] McCann and Simmons, "The Core-Periphery Structure," p. 84.
[30] R.J. Johnston (2000), "Rank-Size Rule." In R.J. Johnston et al., eds., *Dictionary*, p. 672.
[31] See McCann and Simmons, "The Core-Periphery Structure," p. 84.
[32] Michael Pacione (2005). *Urban Geography: A Global Perspective.* London: Routledge, p. 130.

2. Central-place theory was used extensively to plan new settlements in the lands reclaimed from the sea in the Dutch polderlands.[33]

3. Location analysts for large corporations routinely use city size and counts of competitors' locations to plan strategy for growth, expansion, and relocation. Many of the key insights of central place theory and urban systems analysis have been woven into software programs used to evaluate the market potential of alternative locations.[34]

4. In the Canadian context, Consider once again Larry Bourne's and Jim Simmons' analysis of the varying growth rates of different centers in the Canadian urban system. Between 1991 and 2001, the six largest metropolitan regions -- Toronto, Montreal, Vancouver, Ottawa-Gatineau, Calgary, and Edmonton -- accounted for more than two thirds of all of Canada's population growth. In the latter half of this decade, non-urban areas across Canada lost about 0.4 percent of their total population. Theoretical talk of rank-size distributions and equilibrium have important consequences for daily life and for public policy:

> "New patterns of difference -- new fault lines -- do indeed seem to be emerging, especially within urban Canada. The winners in this context will be those urban places and their immediate hinterlands (or zones of influence) that are larger, are plugged into the continental and global economy, are the destinations of immigration flows, have younger and more diverse populations, proximity to certain environmental amenities, and have reasonable access to high-order public goods and services."[35]

Unfortunately, even for those regions near the booming large metropolitan areas, the pressures of growth are outstripping the planning capacities of local government, and the provincial restrictions on municipalities and metropolitan governance offer few possibilities for coherent regional solutions.

And what of the vicious cycle of urban and regional decline? As Simmons and McCann portray the political subsystem that links different parts of the Canadian urban system, the trajectory of growth and decline presents very difficult choices:

> "The *political subsystem* as a network of city connections refers to the movement of political influence and the exchange of money both within and among political jurisdictions. Movement and exchange are illustrated by tax revenues that flow to governments from firms and individuals and by transfer payments and services that flow in other directions .... This subsystem is more closed than open, with connections moving through the urban hierarchy from Ottawa or the provincial capitals to other places in the national urban system: connections are seldom international. Different levels of government (especially federal and provincial)

---

[33] See Pacione, *Urban Geography*, pp. 130-131.
[34] See, for example, Kiran Karande and John R. Lombard (2005). "Location Strategies of Broad-Line Retailers: An Empirical Investigation." *Journal of Business Research* 58(5), 687-695, and Margaret J. Daniels (2007). "Central Place Theory and Sport Tourism Impacts." *Annals of Tourism Research* 34(2), 332-347.
[35] Bourne and Simmons, "New Fault Lines," p. 43.

redistribute funds in an attempt to mediate tensions between economic change and the resistant, more immobile demographic structure. Traditionally, this has been accomplished largely through federal policies that support declining regions by taxing areas of growth -- i.e., a kind of subsidization of geography. From one of point of view ... these fiscal measures actually slow down the response to economic innovation and prevent the national economy from achieving its full potential. From another, more liberal viewpoint ... these programs are the necessary relief systems for people suffering the effects of unpredictable economic changes through no fault of their own. They represent counter shifts in financial support that ultimately ... should benefit both regional and national economies."

Slave Lake, Alberta, August 2010 (Elvin Wyly). Small towns are particularly sensitive to the rhythms of particular industries, and so it's not uncommon to see visual clues to the local economic base -- in this case, a "nodding donkey" oil well portrayed prominently along the city's main street downtown.