

THE ROLE OF GEOGRAPHY IN DATA INTEGRATION AND PREDICTIVE ANALYTICS

SEPTEMBER 2018

Tony Lea, Ph.D.

Senior VP and Chief Methodologist

Environics Analytics, Toronto

Tony.Lea@environicsanalytics.com

ENVIRONICS
ANALYTICS



WHO AM I?

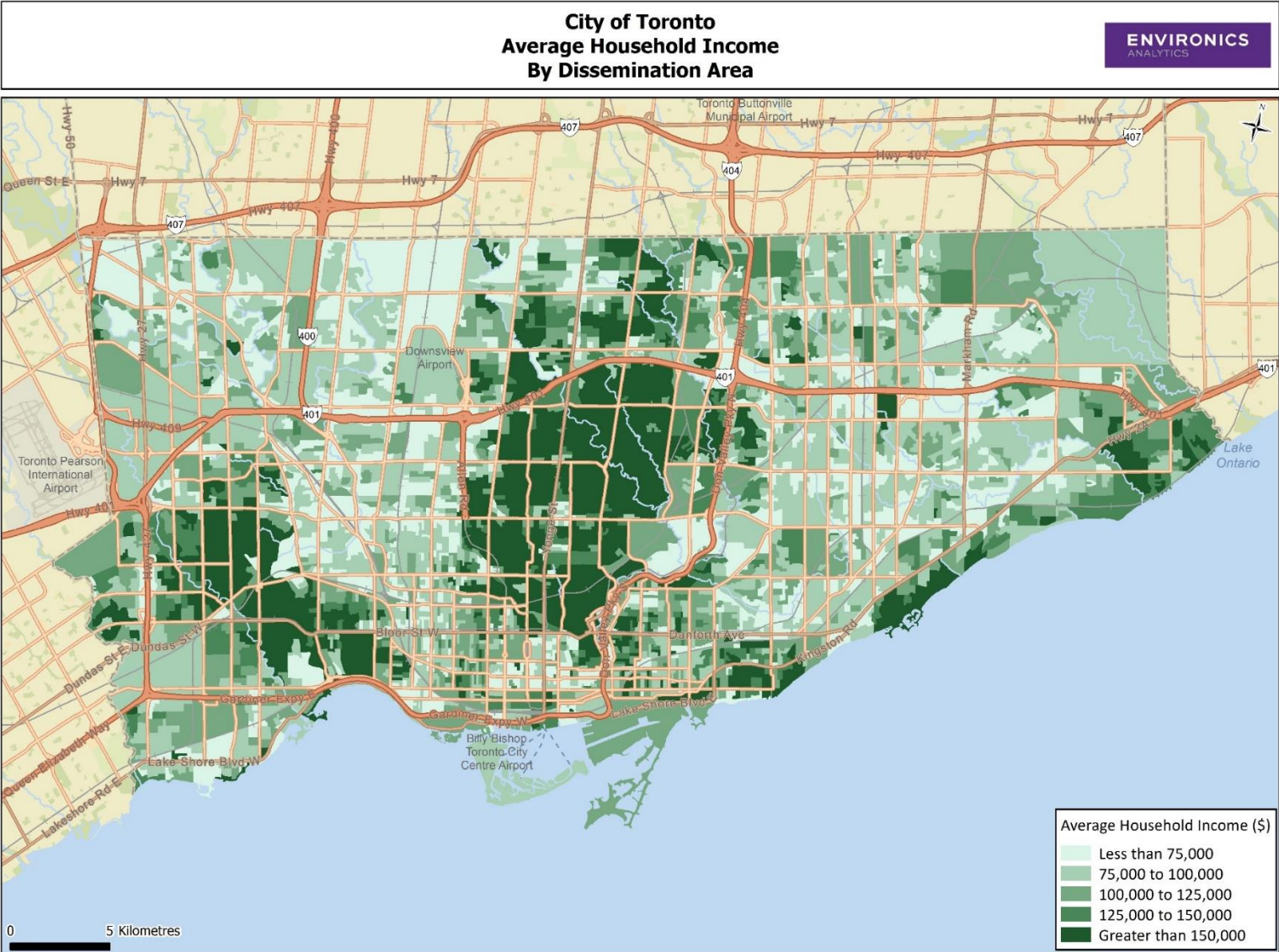


- Chief Methodologist and SVP at Environics Analytics
- Been there for its history – since 2003
- Involved in similar predecessor marketing data and analytics firms (e.g. Compusearch) since 1979
- BA, MA, Ph.D. - University of Toronto
- Taught at the University of Minnesota, University of Toronto, Queen's, Ryerson University
- Currently teach a graduate course on Geodemography at Ryerson University
- Presented about 400 papers at marketing and geography conferences

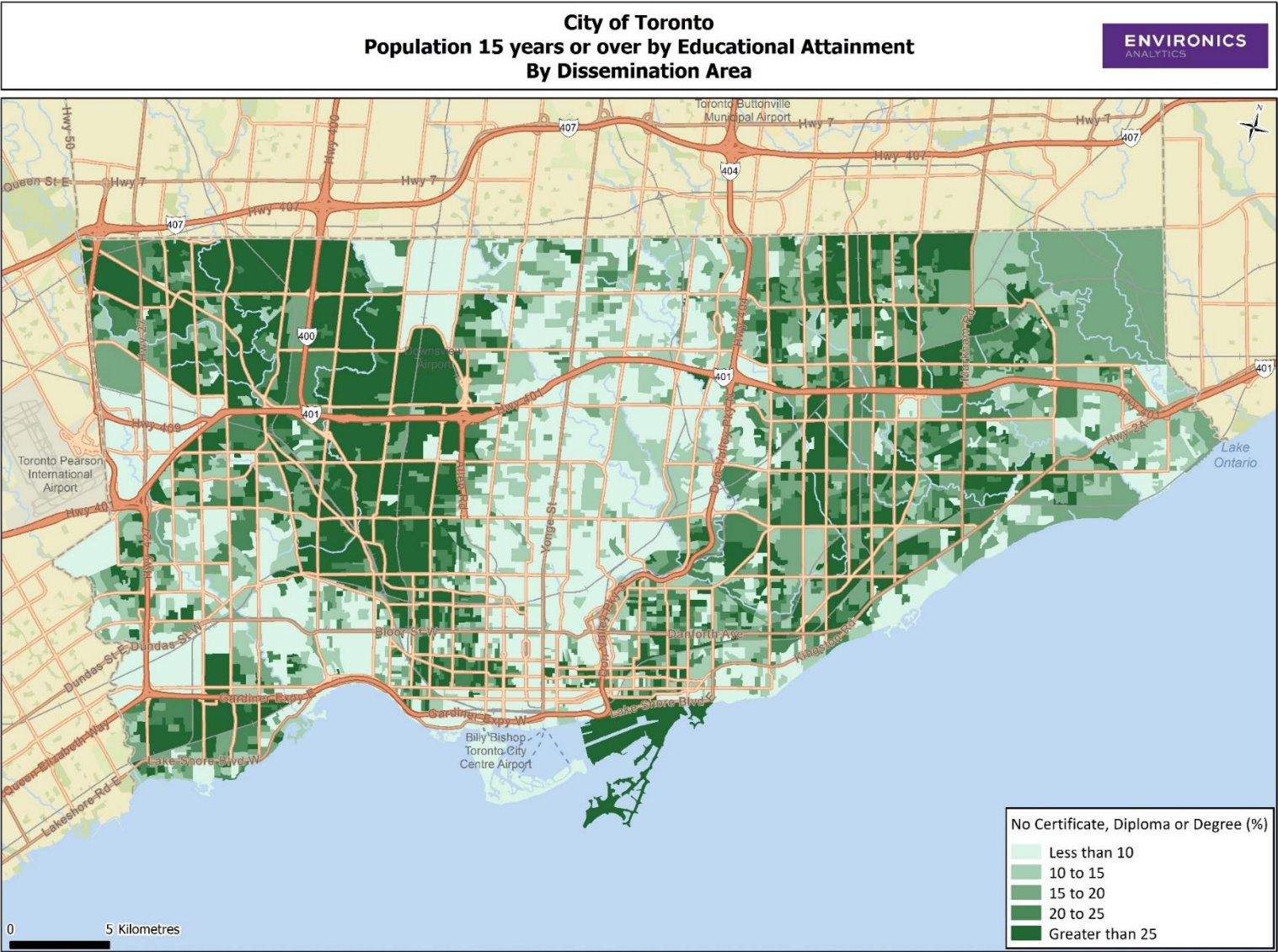
THIS TALK RELATES TO...

- How geography and spatial data can help to enhance predictive modelling practice
- Essentially - applied geography in terms data analytics
- Especially for marketing but not necessarily for marketing – as we will see
- Interesting that in North America this is not taught to geographers in university

AVERAGE HOUSEHOLD INCOME IN TORONTO

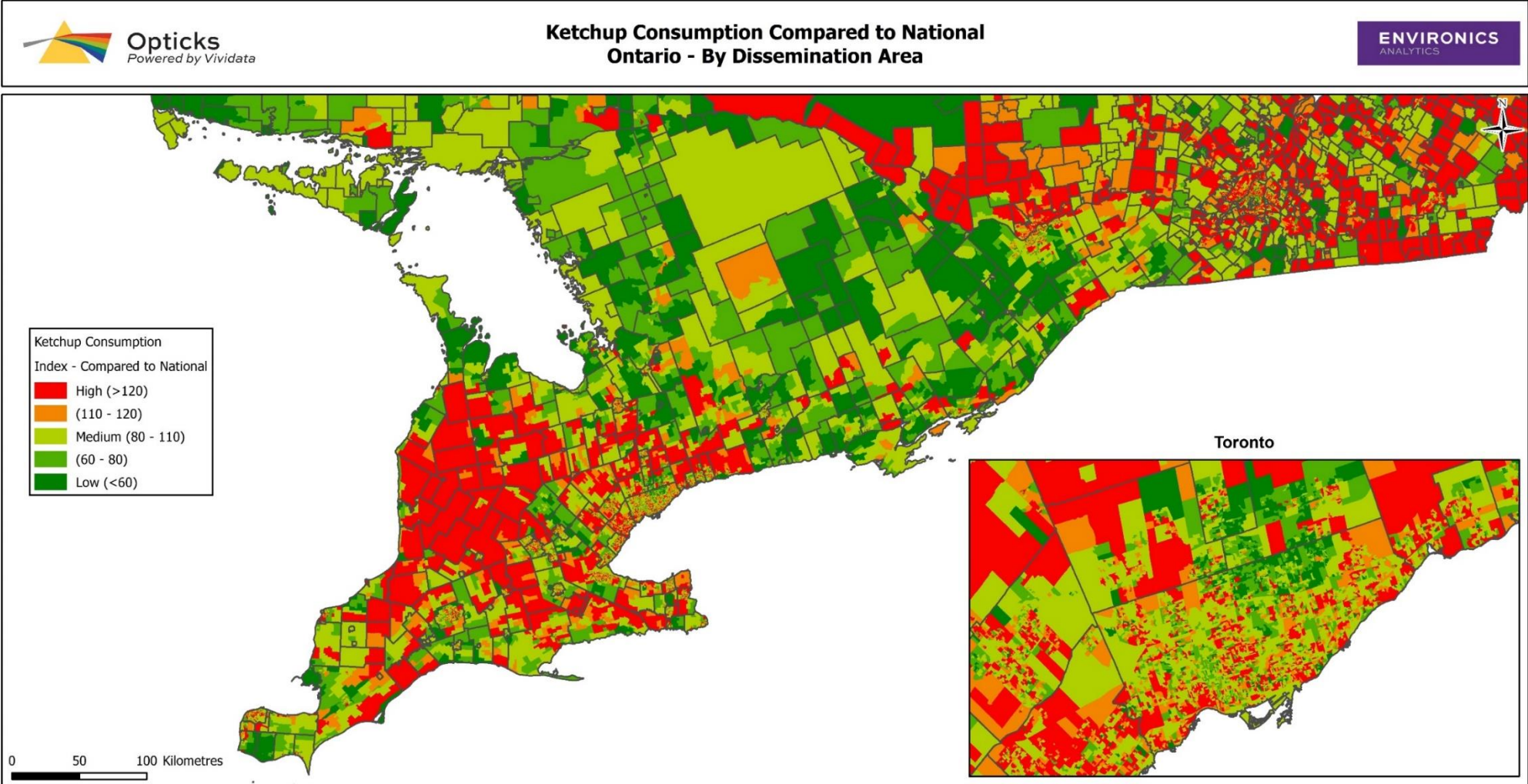


EDUCATION ATTAINMENT IN TORONTO



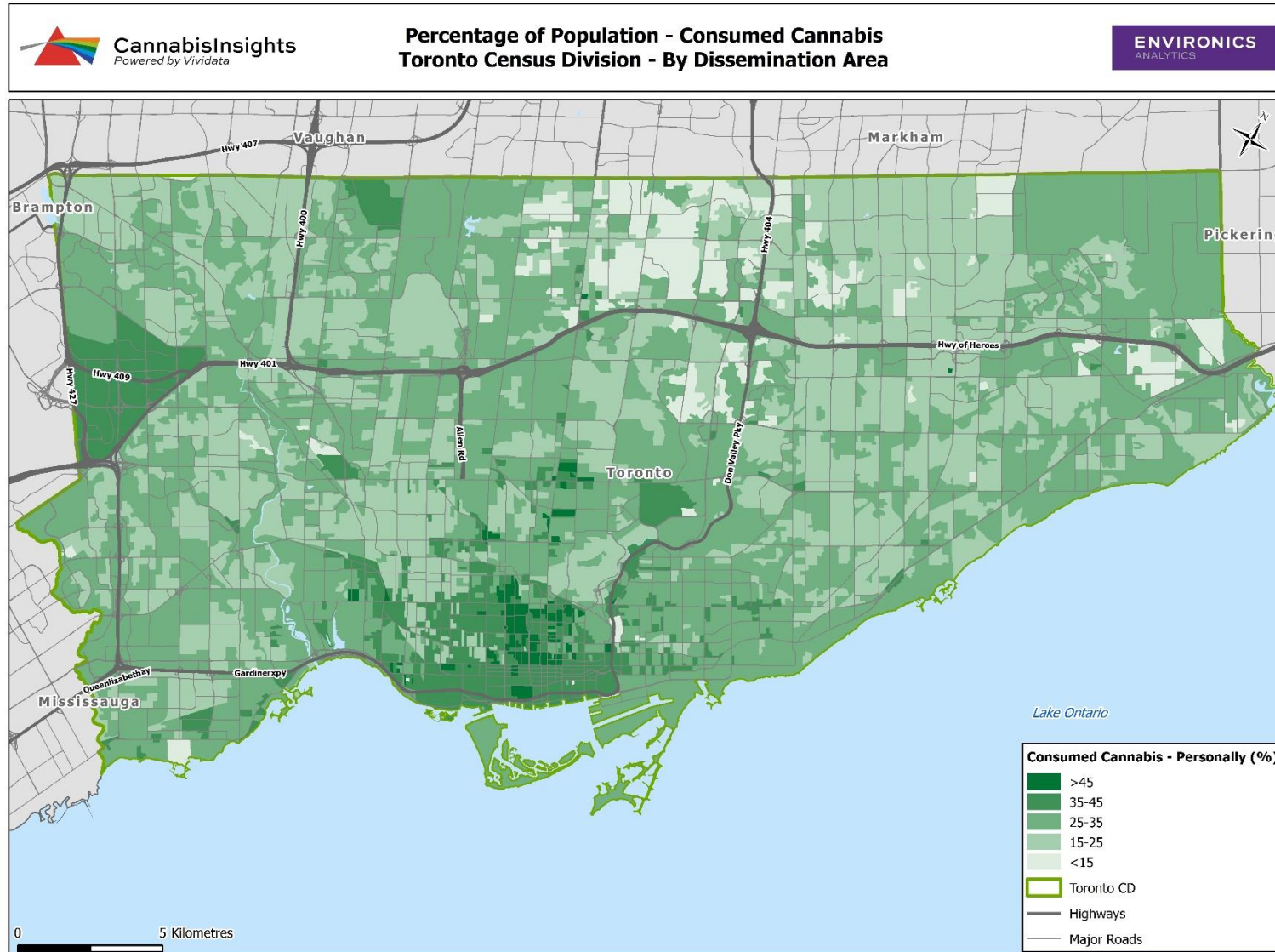
Copyright ©2016 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM's nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

KETCHUP CONSUMPTION



Copyright ©2018 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

TORONTONIANS WHO HAVE CONSUMED CANNABIS



ABOUT ENVIRONICS ANALYTICS



Founded in
November 2003



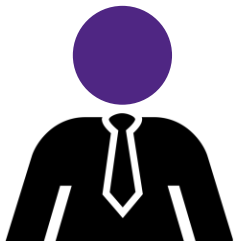
Partnership with Environics
Research Group



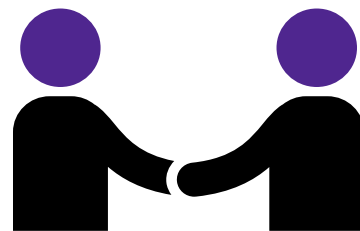
Specialists in data-driven
insights and target
marketing



Canada's largest marketing
services provider



200+ employees strong



3,000+ customers across
North America



Customer Advisory Board
of 16 organizations



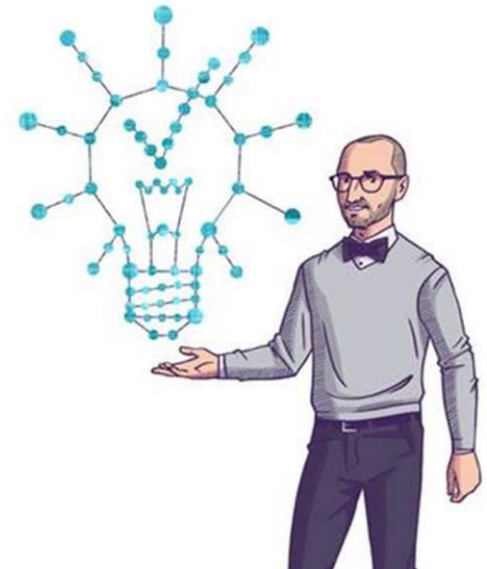
PRIZM5

Developers of PRIZM for
Canada

OUR EXPERIENCE

We help clients achieve their business objectives by delivering actionable customer and market insight grounded in data and analytics.

- We have over 3,000 clients in every industry sector in North America including retail, finance, health, not for profit and government.



WHAT DATA IS TRADITIONALLY USED IN PREDICTIVE MODELLING?

- Extremely large data sets relating to human behaviour and interactions.
- Lots of data available today from many sources; the age of “BIG DATA”
- The data could be customer data, credit card data, and large syndicated survey data, twitter and other social media data etc.
- The focus in this talk is on individual or household data relating to demographics and behaviour: age, income, kids, shopping, travel, leisure, purchasing, donating, subscribing

THE CHALLENGE WITH THESE DATA



Because of privacy or small sample size, or both, a very large portion of these data cannot be “really used”



Privacy means inability to share or display names, phone numbers and other proprietary details



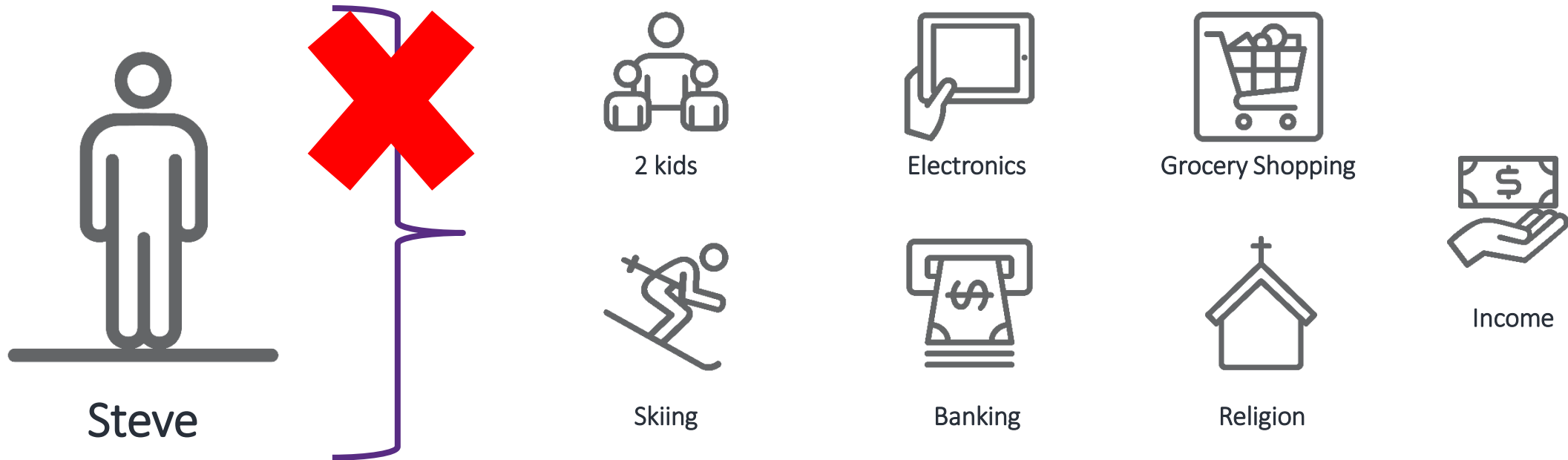
“Really used” means entered into an integrated data base and usable along with other data to draw conclusions



Unlike the past - much of these data now have a spatial reference (an address or postal code) – which can be used to aggregate data geographically

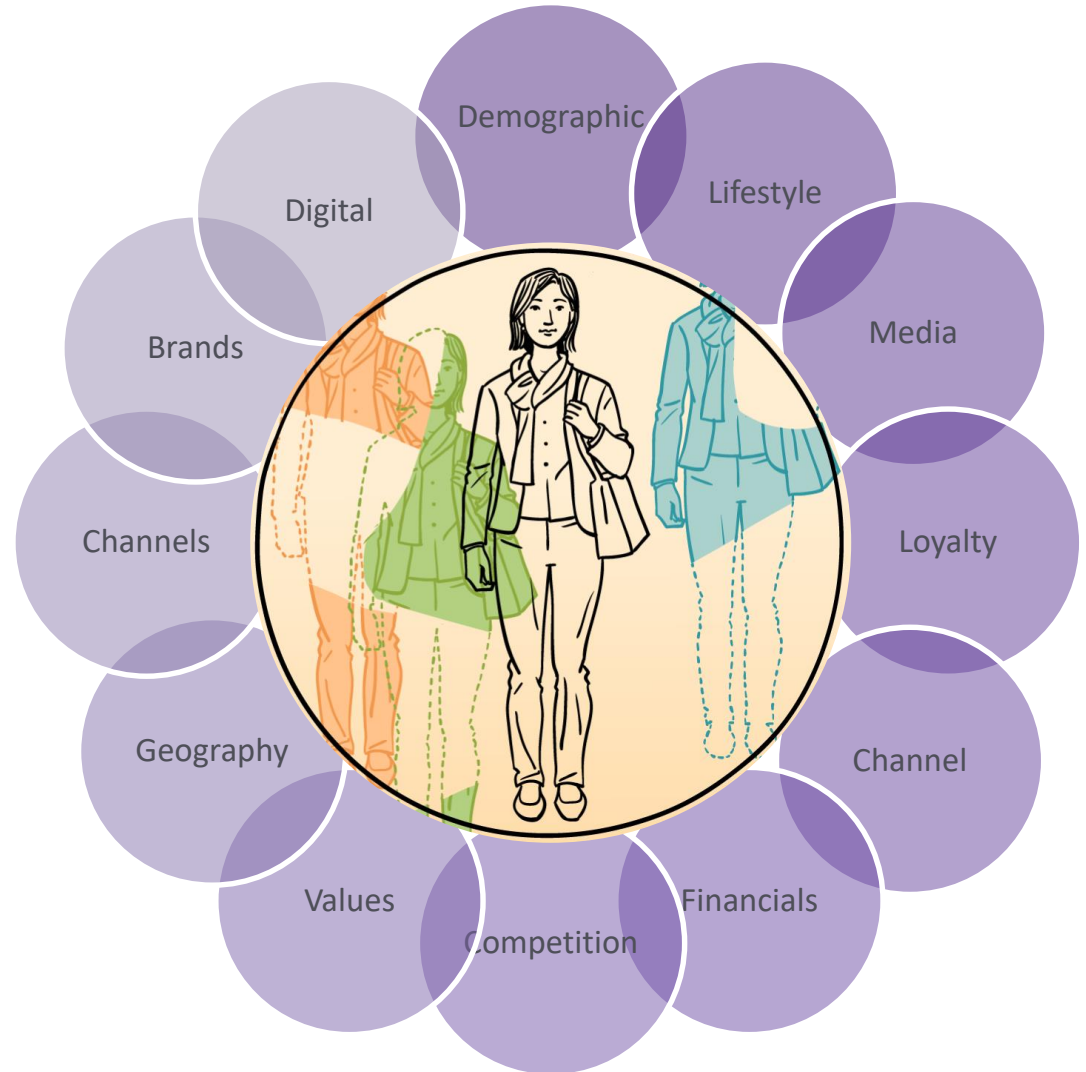
WHY ARE MUCH OF THESE DATA NOT USEFUL IN ANALYSIS?

- They cannot be linked together to tell an integrated story
- An integrated database requires that all the data relates to the same records, or observations (essential part of data integration)
- This permits statistical analysis and ability to act on the observations based on the results



CHALLENGES TO ONE VIEW OF THE CUSTOMER

- Legacy organizational silos that prevent enterprise-wide analytics
- Absence of analytical roadmaps linked to business strategy
- Methodology oversight is lacking or in the wrong hands



WHY ARE THESE DATA NOT USEFUL IN ANALYSIS?

- Need to find a way to assign these rich data (incomes, downhill ski rates etc.) **indirectly** to persons or households or to actual addresses for marketing somehow ... without breaking the privacy rules
- Since preparing and using **these individual data is not legal in Canada** under most circumstances - the challenge is clear.
- But there is a good approximation ...

ENTER GEODEMOGRAPHICS

- Since about 1977 an industry has developed in Canada, and the UK, USA, etc. called geodemographics – also called micromarketing, customer analytics
- Here, demographics, socioeconomic, behavioural, and attitudinal data are integrated in datasets and analyzed, all done “geographically” - **Here geography becomes a tool for data integration**
- The major application areas are marketing of all kinds and retail site selection
- But there are also many other areas in public policy analysis, urban and regional planning, health care analysis, policing, tourism analysis, etc.

GEODEMOGRAPHY

- The geodemographic approach is based on **3 ideas**:
 1. The first law of geography; the attributes of persons/households that are close together are more similar than those that are far away (*positive spatial autocorrelation*)
 2. “Birds of a feather flock together”; so knowing a person’s neighbourhood (at a small spatial scale) helps make good inferences about the person
 3. When privacy is critical, one can make use of small-area attributes as reasonable estimates of the attributes of residents of the small area

1. THE FIRST LAW OF GEOGRAPHY

The attributes of persons/households that are close together are more similar than those that are far away (positive spatial autocorrelation, Waldo Tobler)

Everything is related to everything else, but near things are more related than distant things.

-Waldo Tobler



2. BIRDS OF A FEATHER FLOCK TOGETHER

- This old adage, has considerable truth, especially in human geography
- Knowing a person's neighbourhood (at a small spatial scale) helps make good inferences about the person - in almost all contexts



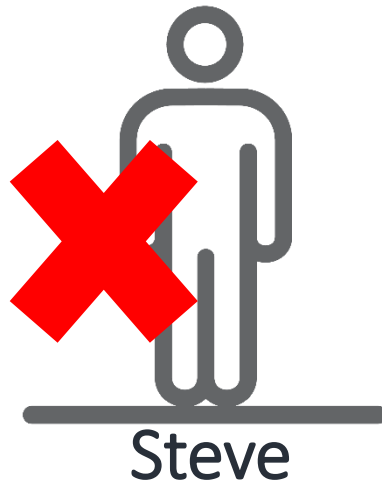
3. PRESERVING PRIVACY

When privacy is critical one can make use of small-area attributes as reasonable estimates of the attributes of residents



GEODEMOGRAPHICS & PRIVACY

- Under most circumstances **data on persons/households can be shared in the form of aggregate data** - when assigned to small geographical areas; this is what the census does
- This is done for all small and larger census areas: Dissemination Areas (DA), Census Tracts (CT) and postal areas like Forward Sortation Areas (FSAs), and also FSALDU's – 6 character postal codes.



QUESTIONS ADDRESSED BY GEODEMOGRAPHICS

The basic questions of geodemography are:

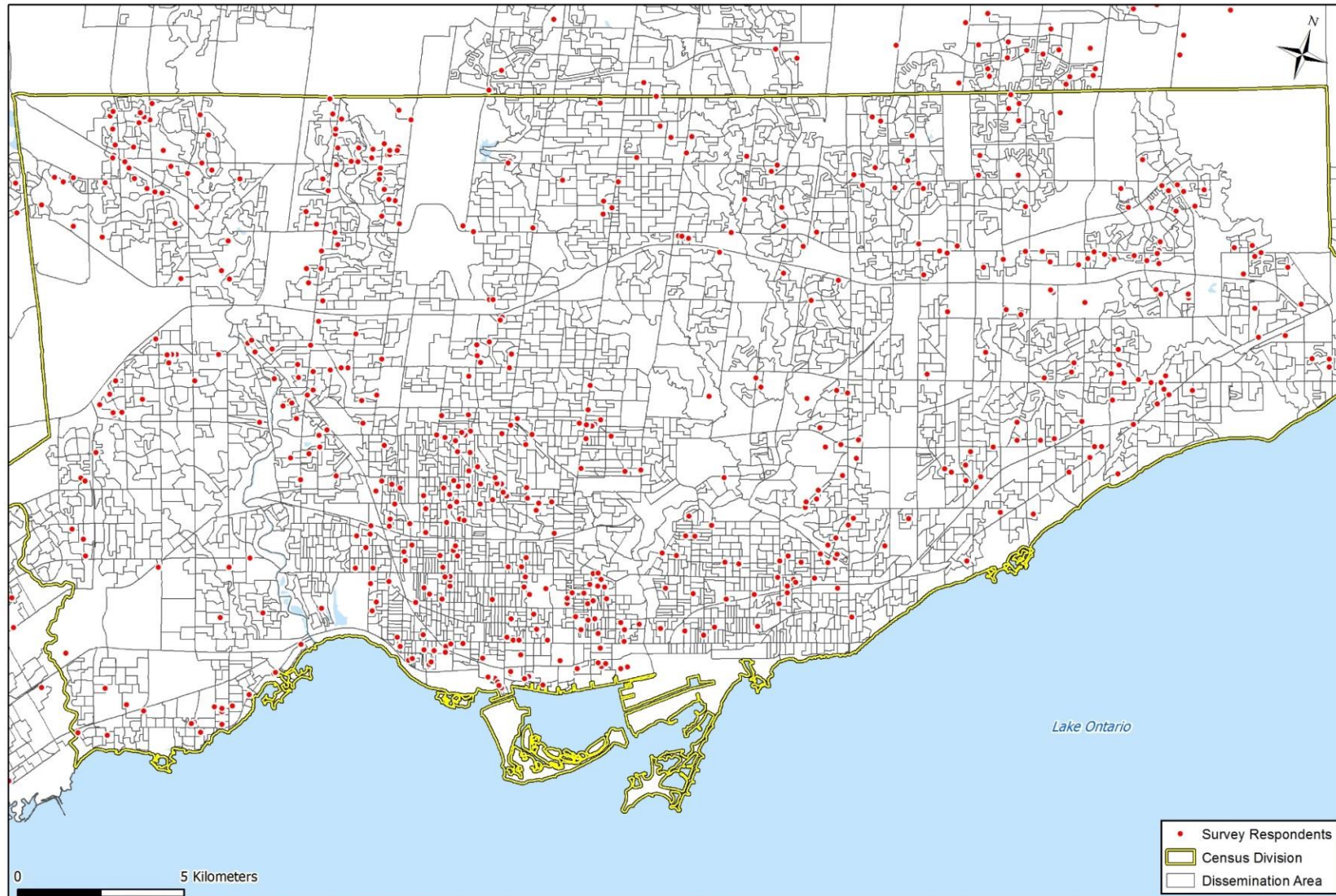
1. What are the attributes of or the segment for this particular small area (e.g. census area)?
2. What is the profile of a specific set of customers/subscribers/donors/patients in a database?
3. Where can I find more of these types of persons and households?
4. What is the profile of a particular market or trade area?
5. What is the profile of surveyed persons who buy X or prefer Y (yoga, shop at Walmart, eat donuts every day, vacation in Barbados, buy Tide detergent, go canoeing)?

SMALL SAMPLE (SURVEY) DATA

- What if the set of observations (customers, surveyed persons) is less than the number of small neighbourhoods? (.68 respondents per area)
 - *The small sample problem*
- Small census areas have a certain average household income, a certain distribution of ages, household sizes, types of dwellings, ethnicities etc. These data are published by Stats Canada routinely and are widely used
- If we are interested in hip replacement patients, ketchup consumption, blue box use, or Lexus buyers? What if – as usual - we have just survey data? In these cases the number of data points is usually small?
- How do we proceed with analysis and the making of helpful inferences?

Small Sample Size of Survey Toronto Census Division by Dissemination Area

ENVIRONICS
ANALYTICS



Sample survey locations of respondents - a relatively small sample of points

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

HOW CAN WE HANDLE THESE DATA ?

- This map is difficult to generalize and to analyze (we could cartographically “smooth the map”)
- The number of Dissemination Areas (DAs) in Ontario is just over 20,000 so the number of survey respondents in Ontario almost always be much smaller
 - So less than one respondent per DA
- So we cannot create a meaningful DA-level database to analyze statistically, and we cannot map it as a thematic or choropleth map

This is a problem. We have good data, it is geographically referenced, but a good deal of analysis is impossible.

SOLUTIONS

Solution 1 (this is what was done before about 1983)

- Run a **regression** (or similar model) on the density per household of the behaviour using the small areas (such as dissemination areas) as observations:
 - Downhill ski rate = $0.068 + 0.0025 * (\text{Income}) - 0.027 * (\% \text{ age } 55 \text{ plus}) + \dots$

Solution 2 (use a segmentation system - like Environics Analytics PRIZM5)

- EAG has developed a segmentation system (PRIZM5) for small areas based on demographics, socioeconomic, urbanity, density and some behaviour
- PRIZM5 classifies all Canadian neighbourhoods into 68 unique segments
- In this approach we create an **analysis database** of these 68 segments
- It includes rates of buying or preferring X (or penetration rates)
- We then analyze and map the segments at a small geographical scale - not at the persons/households level

GEODEMOGRAPHIC SEGMENTATION

DEMOGRAPHICS



\$175,000



DEMOGRAPHICS




\$82,000



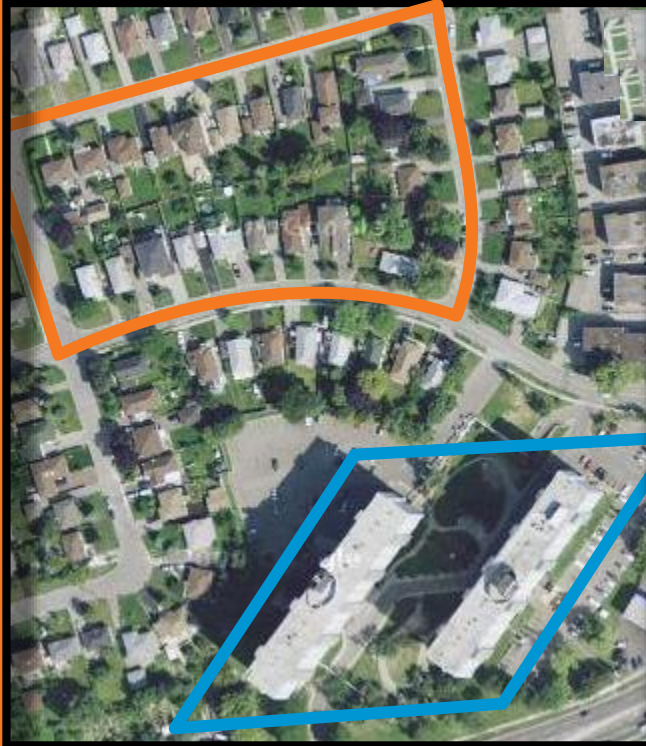
WE CAN THEN ASSIGN LIFESTYLES TO NEIGHBOURHOODS

04 SUBURBAN SUCCESS



S1 SUBURBAN ELITE PROSPEROUS PARENTS **F8**

Wealthy, middle-aged and older homeowners



19 GREY PRIDE



U5 URBAN OLDER LATER YEARS **M3**

Middle-income seniors in urban apartments

...AND FIND THESE LIFESTYLES ALL OVER CANADA





PRIZM5

01 COSMOPOLITAN ELITE
U1 VERY WEALTHY MIDDLE-AGED AND OLDER FAMILIES AND COUPLES F8

02 URBANE VILLAGERS
U1 WEALTHY MIDDLE-AGED AND OLDER CITY RESIDENTS F8

03 ARTS & AFFLUENCE
U1 WEALTHY ESTABLISHED URBAN FAMILIES AND COUPLES F5

04 SUBURBAN SUCCESS
S1 WEALTHY MIDDLE-AGED AND OLDER HOMES F8

05 ASIAN SOPHISTICATES
U2 UPSCALE URBAN ASIAN FAMILIES F9

06 KIDS & CAREERS
S1 WELL-OFF MIDDLE-AGED SUBURBAN FAMILIES F8

07 NOUVEAUX RICHES
S1 WELL-OFF SUBURBAN QUEBEC FAMILIES AND COUPLES F6

08 ROOMERANG CITY
U1 UPSCALE MULTI-GENERATIONAL URBAN HOUSEHOLDS F9

09 SATELLITE BURBS
E1 OLDER SPECIALTIES URBAN COUPLES AND FAMILIES F8

10 EMPTYING NESTS
S1 OLDER UPPER-MIDDLE-INCOME SUBURBAN COUPLES M1

11 URBAN DIGERATI
U3 YOUNGER WELL-EDUCATED CITY SINGLES Y1

12 STREET SCENES
U3 YOUNGER UPPER-MIDDLE-INCOME SINGLES AND FAMILIES Y2

13 ASIAN AVENUES
U2 SUCCESSFUL MIDDLE-AGED AND OLDER ASIAN FAMILIES F9

14 DIVERSITY HEIGHTS
S3 DIVERSE MIDDLE-AGED AND OLDER SUBURBAN FAMILIES F9

15 HERITAGE HUBS
S3 DIVERSE MIDDLE-AGED MULTI-ETHNIC SUBURBAN FAMILIES F3

16 PETS & PCS
S2 YOUNGER SPECIALTIES SUBURBAN FAMILIES F1

17 EXURBAN WONDERLAND
E1 MIDDLE-AGED SPECIALTIES EXURBAN FAMILIES F2

18 MANAGEMENT MATERIAL
S2 YOUNGER UPPER-MIDDLE-INCOME SUBURBAN FAMILIES F1

19 GREY PRIDE
U6 MIDDLE-INCOME SENIORS URBAN APARTMENTS M3

20 SOUTH ASIAN ACHIEVERS
S3 DIVERSE UPPER-MIDDLE-INCOME SOUTH ASIAN FAMILIES F1

21 BEAU MONDE
U5 MIDDLE MIDDLE-INCOME QUEBEC SENIORS F6

22 AGING IN SUBURBIA
S4 OLDER UPPER-MIDDLE-INCOME SUBURBAN COUPLES AND FAMILIES F5

23 ASIAN NEW WAVE
U4 MIDDLE WELL-EDUCATED ASIAN SINGLES AND FAMILIES F7

24 FRESH AIR FAMILIES
E2 MIDDLE-AGED MIDDLE-INCOME URBANITES F3

25 SOUTH ASIAN SOCIETY
S3 MIDDLE-AGED MIDDLE-INCOME SOUTH ASIAN FAMILIES F7

26 SECOND CITY RETIREES
S4 OLDER AND MATURE MIDDLE-INCOME HOUSEHOLDERS M1

27 DIVERSE CITY
U2 MIDDLE UPPER MIDDLE-INCOME CITY DESIRES F9

28 METRO MULTICULTURALS
U2 DIVERSE MIDDLE-AGED AND OLDER HOUSEHOLDS M1

29 SILVER LIVINGS
U5 URBAN SENIORS IN-HIGH-RISE APARTMENTS M3

30 LA VIE EST BELLE
E3 MIDDLE-AGED MIDDLE-INCOME QUEBEC FAMILIES AND COUPLES F6

31 NEW WORLD SYMPHONY
U4 DIVERSE BOOM MIDDLE-INCOME CITY DESIRES F1

32 MINI VAN & VIN ROUGE
E3 YOUNGER AND MIDDLE-AGED QUEBEC HOMES F2

33 HEARTLAND RETIREES
R2 RURAL OLDER AND MATURE LOWER-MIDDLE-INCOME COUPLES M2

34 ROOMS WITH A VIEW
U4 MIDDLE UPPER MIDDLE-INCOME YOUNG UPPER CITY DESIRES Y1

35 COUNTRY ACRES
R1 MIDDLE-AGED AND OLDER RURAL COUPLES AND FAMILIES F5

36 EXURBAN HOMESTEADERS
E2 EXURBAN MIDDLE-AGED AND MIDDLE-INCOME HOUSEHOLDERS F5

37 TRUCKS & TRAILERS
S2 YOUNGER AND MIDDLE-AGED UPPER-MIDDLE-INCOME FAMILIES F2

38 BRADS & PADS
U3 YOUNG SINGLE URBAN HERNERS Y1

39 OUR TIME
S4 OLDER AND MATURE LOWER-MIDDLE-INCOME SUBURBANITES M1

40 WIDE OPEN SPACES
R1 MIDDLE-AGED MIDDLE-INCOME FARMERS AND BLUE-COLOR WORKERS F3

41 VIEILLE ECOLE
E3 MIDDLE-AGED AND OLDER QUEBEC URBANITES F6

42 HOME SWEET ROWNS
S3 DIVERSE MIDDLE-AGED MIDDLE-INCOME SUBURBANITES F7

43 NEWCOMERS RISING
U4 YOUNGER DOWNSIDE CITY IMMIGRANTS F1

44 JEUNES ET ACTIFS
U7 YOUNGER DOWNSIDE QUEBEC SINGLES Y1

45 JEUNES D'ESPRIT
R3 OLDER DOWNSIDE RURAL COUPLES F6

46 VILLES TRANQUILLES
R3 MIDDLE-AGED RURAL QUEBEC COUPLES AND FAMILIES F3

47 TRADITIONAL TOWN LIVING
T1 MIDDLE-AGED AND OLDER MIDDLE-INCOME HOUSEHOLDERS F4

48 VARIÉTÉ SUBURBAINE
S5 LOWER-MIDDLE-INCOME QUEBEC SUBURBANITES F3

49 ENCLAVES MULTICULTIQUES
U7 DIVERSE LOW-INCOME YOUNGER CITY DESIRES Y2

50 SUBURBAN SCRAMBLE
S2 YOUNGER LOWER-MIDDLE-INCOME SUBURBANITES Y2

51 AGING & ACTIVE
T1 OLDER AND MATURE LOWER-MIDDLE-INCOME BOOM HOUSEHOLDS F4

52 STRIVING STARTUPS
U6 YOUNGER UPPER LOWER-MIDDLE-INCOME SINGLES AND FAMILIES Y2

53 OUTDOOR ORIGINALS
E2 MIDDLE-AGED AND OLDER LOWER-MIDDLE-INCOME URBANITES F4

54 SERENITY SPRINGS
T1 MATURE LOWER-MIDDLE-INCOME TEEN SINGLES AND COUPLES M3

55 LA VIE RUCOLIQUE
R3 RURAL MIDDLE-AGED AND OLDER COUPLES AND FAMILIES F4

56 SINGLE CITY JAZZ
U6 MIDDLE MIDDLE-AGED CITY SINGLES IN APARTMENTS Y1

57 FIRST NATIONS FAMILIES
T1 YOUNGER LOWER-MIDDLE-INCOME INTERCULTURAL FAMILIES F2

58 RUSTIC ROADS
R2 RURAL DOWNSIDE OLDER COUPLES M2

59 LOCATAIRES EN BANLIEUES
S5 YOUNGER DOWNSIDE SUBURBAN RENTERS Y2

60 BONS VIVANTS
S5 OLDER DOWNSIDE SUBURBAN SINGLES AND COUPLES M3

61 LES SENIORS
U5 OLDER LOW-INCOME QUEBEC SENIORS M3

62 TERRE À TERRE
R3 DOWNSIDE MIDDLE-AGED AND OLDER RURAL HOUSEHOLDS F6

63 LUNCH AT TIM'S
U6 URBAN DOWNSIDE SINGLES AND FAMILIES F5

64 FÊTE AU VILLAGE
R3 RURAL DOWNSIDE QUEBEC SENIORS M2

65 YOUNG & CONNECTED
U6 YOUNGER LOW-INCOME MULTI-ETHNIC URBAN SENIORS Y2

66 SUNSET TOWERS
U5 LOWER LOW-INCOME SENIORS IN URBAN APARTMENTS M3

67 SURVIVRE EN VILLE
U7 YOUNGER LOW-INCOME QUEBEC URBAN RENTERS Y1

68 LOW-RISE RENTERS
U6 YOUNG LOW-INCOME CITY SINGLES AND FAMILIES Y2

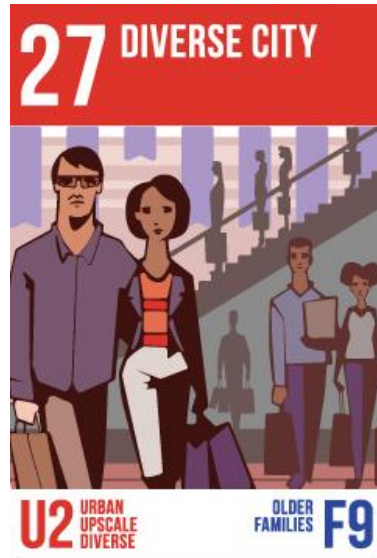
Copyright ©2015
Environics Analytics
PRIZM is a trademark of
The Nielsen Company (US),
and is used with permission.
Sources of data shown
include Environics
Analytics, Environics
Research Group, Statistics
Canada, NAACRS (PM),
Numeris and
AskingCrestions™ (all used
with permission).
Icon illustrations by Scott
Brooke.

EXAMPLE: LOCATING 3 SEGMENTS IN THE GTA



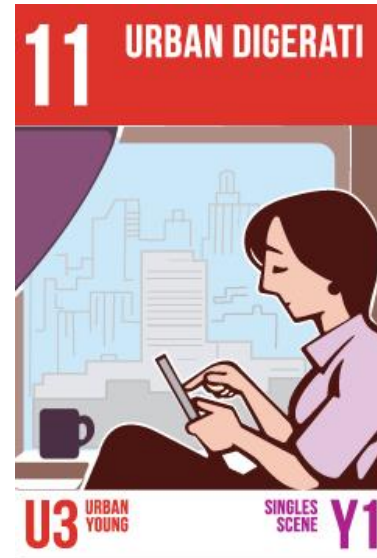
Successful, middle-aged and older Asian households

45-64
\$88,609
Mixed Education
Majority Chinese
Yoga
Asian Restaurants



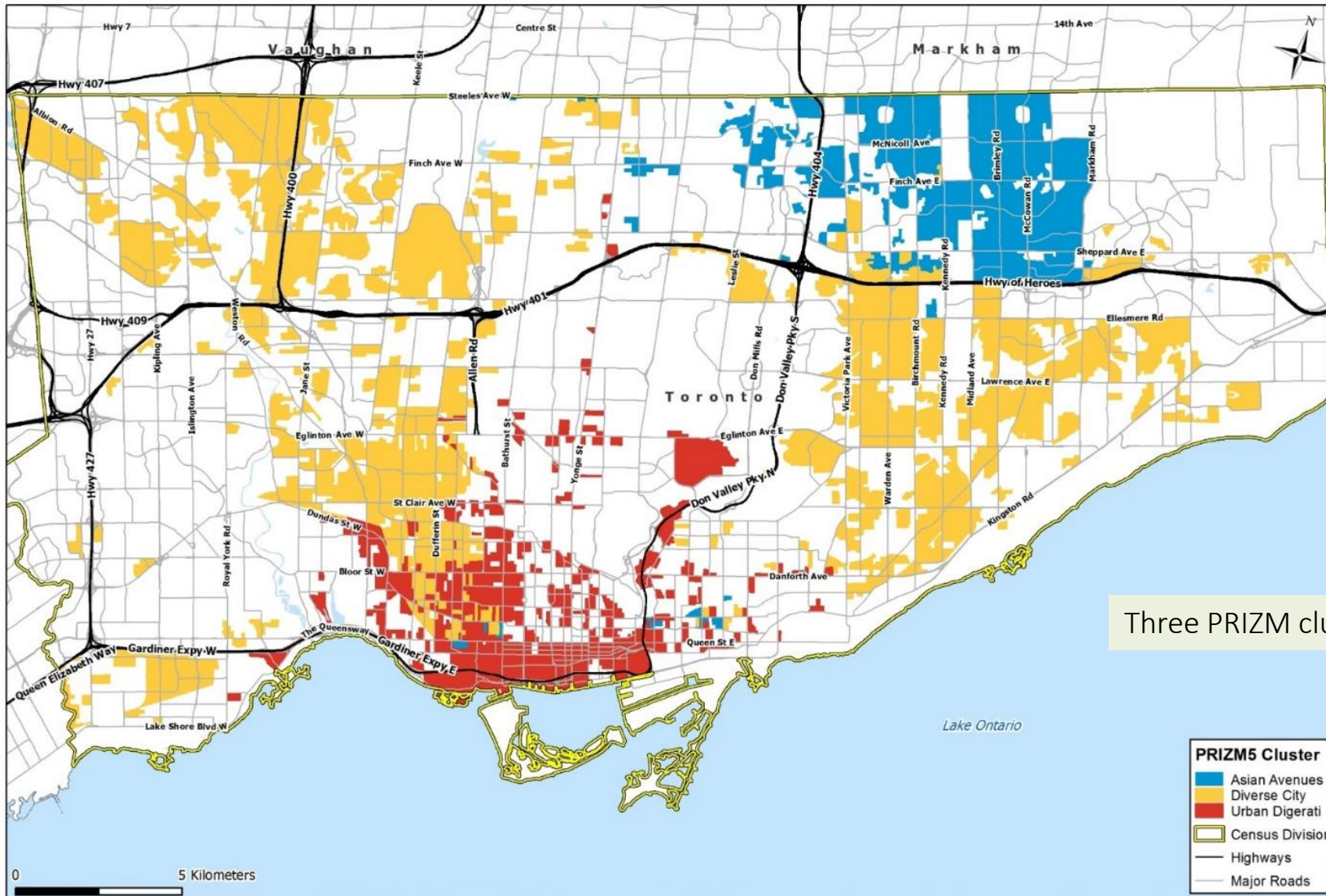
Middle-income, diverse city dwellers

Over 55
\$85,204
High School
Multicultural
Soccer Games
Pizza Pizza



Younger, well-educated city singles

25-45
\$105,803
University
Multicultural
Online Dating
David's Tea



Three PRIZM clusters mapped in Toronto

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

ILLUSTRATION OF APPROACH 2

- The following is a very simple example that illustrates the idea
- It is a worked example of participation in downhill skiing using a small database – shown here as an Excel file
- The survey data come from a well known Canadian large syndicated survey

Glimpse of a Survey Database								
Respondents	PRCDDA	PRCDCSD	CMA CA	Downhill Skiing in the last year	Own Ski Equipment	Watch Skiing Television	Attend Outdoor Shows/Exhibitions	Travelled on Skiing Holiday
M1M2L4	35204207	3520005	535	0	0	0	0	0
M4N2T7	35202348	3520005	535	1	1	0	1	1
M4X1S5	35200796	3520005	535	0	0	0	0	0
M1V2N5	35200128	3520005	535	0	0	0	0	0
M3B3J9	35202764	3520005	535	1	1	0	1	1
M2H1V2	35202462	3520005	535	1	1	0	1	1
M6S1P5	35203153	3520005	535	0	0	0	0	0
M2P1C3	35202614	3520005	535	1	1	1	1	1
M6N2Z1	35201168	3520005	535	0	0	0	0	0
M8X2K5	35201501	3520005	535	1	1	0	0	1
M4G2Z9	35202769	3520005	535	1	1	1	0	1
M9B4N2	35201649	3520005	535	0	0	0	0	0
M6N4S8	35201242	3520005	535	0	0	0	0	0
M4G1N7	35202685	3520005	535	0	0	0	0	0
M2POA1	35202614	3520005	535	1	1	1	0	1
M4L3R7	35200667	3520005	535	0	0	0	0	0
M4W1W6	35202879	3520005	535	1	0	1	0	1
M2H2M7	35202477	3520005	535	1	1	1	1	0
M2P1G1	35202616	3520005	535	1	1	0	0	1
M4J2P2	35203864	3520005	535	0	0	0	0	0
.
.
.

A small piece of a customer database or a large (N=5000) syndicated survey

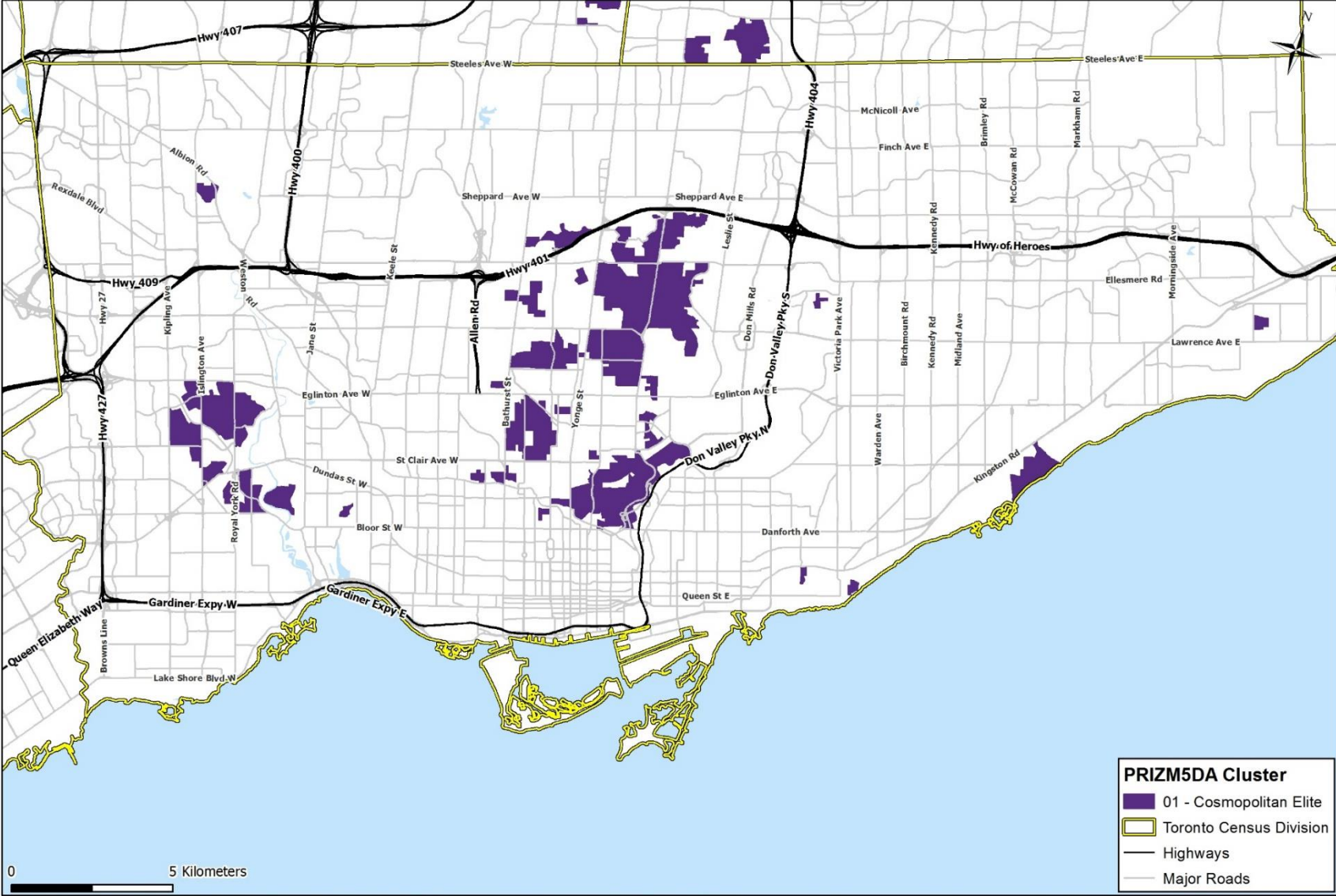
Survey Database - PRIZM5 Cluster Field Appended										
Respondents	PRCDDA	PRCDCSD	CMA CA	Downhill Skiing in the last year	Own Ski Equipment	Watch Skiing Television	Attend Outdoor Shows/Exhibitions	Travelled on Skiing Holiday	PRIZM5	NAME
M1M2L4	35204207	3520005	535	0	0	0	0	0	28	Metro Multiculturals
M4N2T7	35202348	3520005	535	1	1	0	1	1	01	Cosmopolitan Elite
M4X1S5	35200796	3520005	535	0	0	0	0	0	03	Arts & Affluence
M1V2N5	35200128	3520005	535	0	0	0	0	0	13	Asian Avenues
M3B3J9	35202764	3520005	535	1	1	0	1	1	01	Cosmopolitan Elite
M2H1V2	35202462	3520005	535	1	1	0	1	1	13	Asian Avenues
M6S1P5	35203153	3520005	535	0	0	0	0	0	12	Street Scenes
M2P1C3	35202614	3520005	535	1	1	1	1	1	01	Cosmopolitan Elite
M6N2Z1	35201168	3520005	535	0	0	0	0	0	27	Diverse City
M8X2K5	35201501	3520005	535	1	1	0	0	1	01	Cosmopolitan Elite
M4G2Z9	35202769	3520005	535	1	1	1	0	1	11	Urban Digerati
M9B4N2	35201649	3520005	535	0	0	0	0	0	28	Metro Multiculturals
M6N4S8	35201242	3520005	535	0	0	0	0	0	27	Diverse City
M4G1N7	35202685	3520005	535	0	0	0	0	0	01	Cosmopolitan Elite
M2POA1	35202614	3520005	535	1	1	1	0	1	01	Cosmopolitan Elite
M4L3R7	35200667	3520005	535	0	0	0	0	0	03	Arts & Affluence
M4W1W6	35202879	3520005	535	1	0	1	0	1	01	Cosmopolitan Elite
M2H2M7	35202477	3520005	535	1	1	1	1	0	13	Asian Avenues
M2P1G1	35202616	3520005	535	1	1	0	0	1	01	Cosmopolitan Elite
M4J2P2	35203864	3520005	535	0	0	0	0	0	11	Urban Digerati
.
.
.

Each DA in Canada has an assigned PRIZM5 code

File Sorted: Cluster 01 Rate Creation										
Respondents	PRCDDA	PRCDCSD	CMA CA	Downhill Skiing in the last year	Own Ski Equipment	Watch Skiing Television	Attend Outdoor Shows/Exhibitions	Travelled on Skiing Holiday	PRIZM5	NAME
M4N2T7	35202348	3520005	535	1	1	0	1	1	01	Cosmopolitan Elite
M3B3J9	35202764	3520005	535	1	1	0	1	1	01	Cosmopolitan Elite
M2P1C3	35202614	3520005	535	1	1	1	1	1	01	Cosmopolitan Elite
M8X2K5	35201501	3520005	535	1	1	0	1	1	01	Cosmopolitan Elite
M4G1N7	35202685	3520005	535	1	0	0	0	0	01	Cosmopolitan Elite
M2P0A1	35202614	3520005	535	1	1	1	0	1	01	Cosmopolitan Elite
M4W1W6	35202879	3520005	535	1	1	1	0	1	01	Cosmopolitan Elite
M2P1G1	35202616	3520005	535	1	1	0	0	1	01	Cosmopolitan Elite
M8X2C7	35201504	3520005	535	0	0	0	0	0	01	Cosmopolitan Elite
M5P3A6	35202896	3520005	535	1	0	0	0	0	01	Cosmopolitan Elite
M4R1X5	35202341	3520005	535	1	1	1	1	1	01	Cosmopolitan Elite
M5M2J6	35202363	3520005	535	1	1	1	1	1	01	Cosmopolitan Elite
M2L2G6	35202622	3520005	535	1	0	0	0	1	01	Cosmopolitan Elite
M2L1W2	35200371	3520005	535	1	1	0	0	1	01	Cosmopolitan Elite
M4W2Z5	35203848	3520005	535	1	0	0	0	1	01	Cosmopolitan Elite
M5P1G8	35202905	3520005	535	0	0	0	0	0	01	Cosmopolitan Elite
M4T1J6	35202776	3520005	535	0	0	0	0	0	01	Cosmopolitan Elite
M4N3J6	35204201	3520005	535	1	0	0	0	1	01	Cosmopolitan Elite
.
.
.
		Rate >		0.733	0.500	0.333	0.400	0.633		

Sort DB by cluster code and focus on one cluster as an example.
 All the records here are those classed in cluster 1

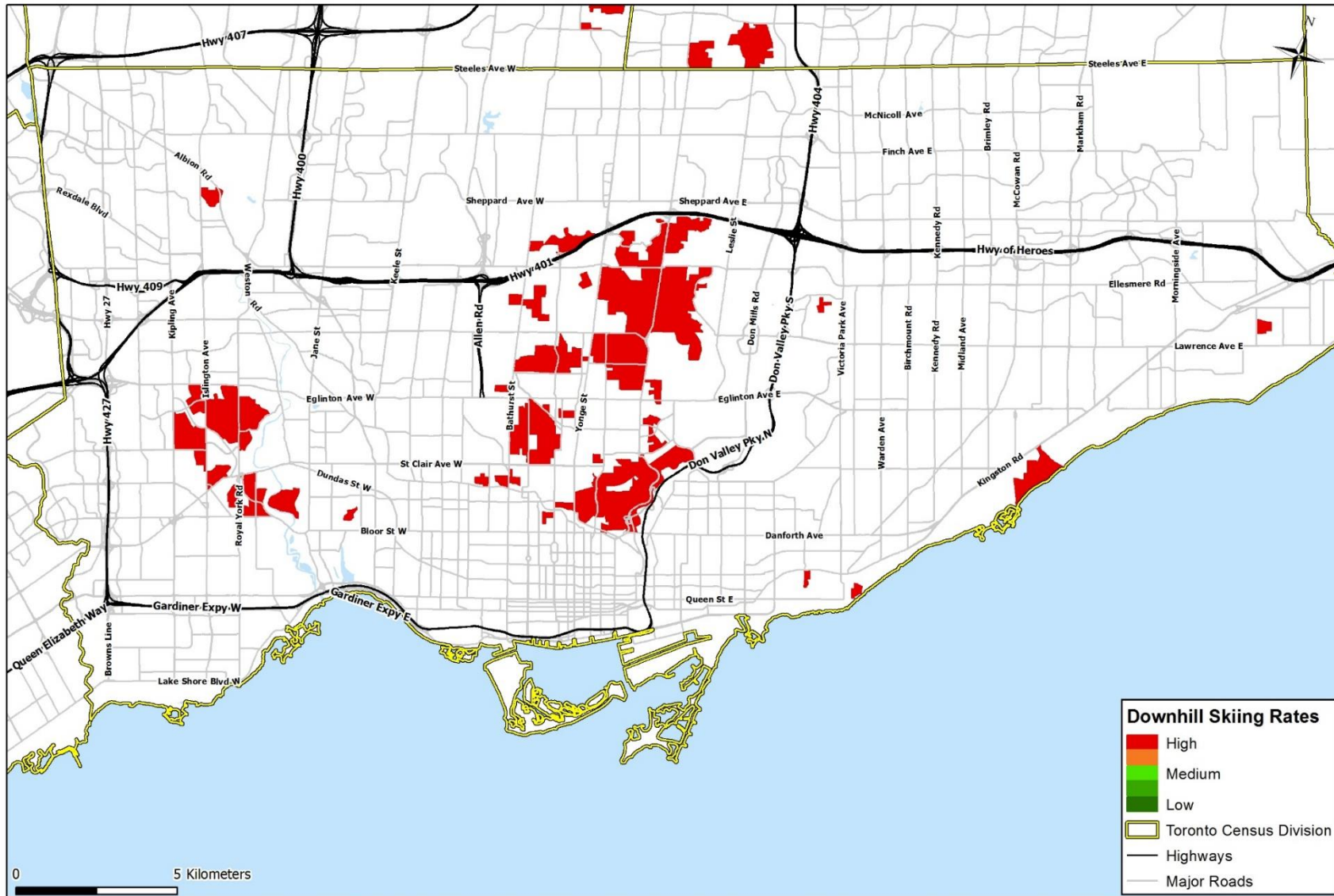
Cluster 01 - Cosmopolitan Elites Toronto Census Division by Dissemination Area



Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



Cluster 01 Cosmopolitan Elites - Downhill Skiing Rates Toronto Census Division by Dissemination Area



Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

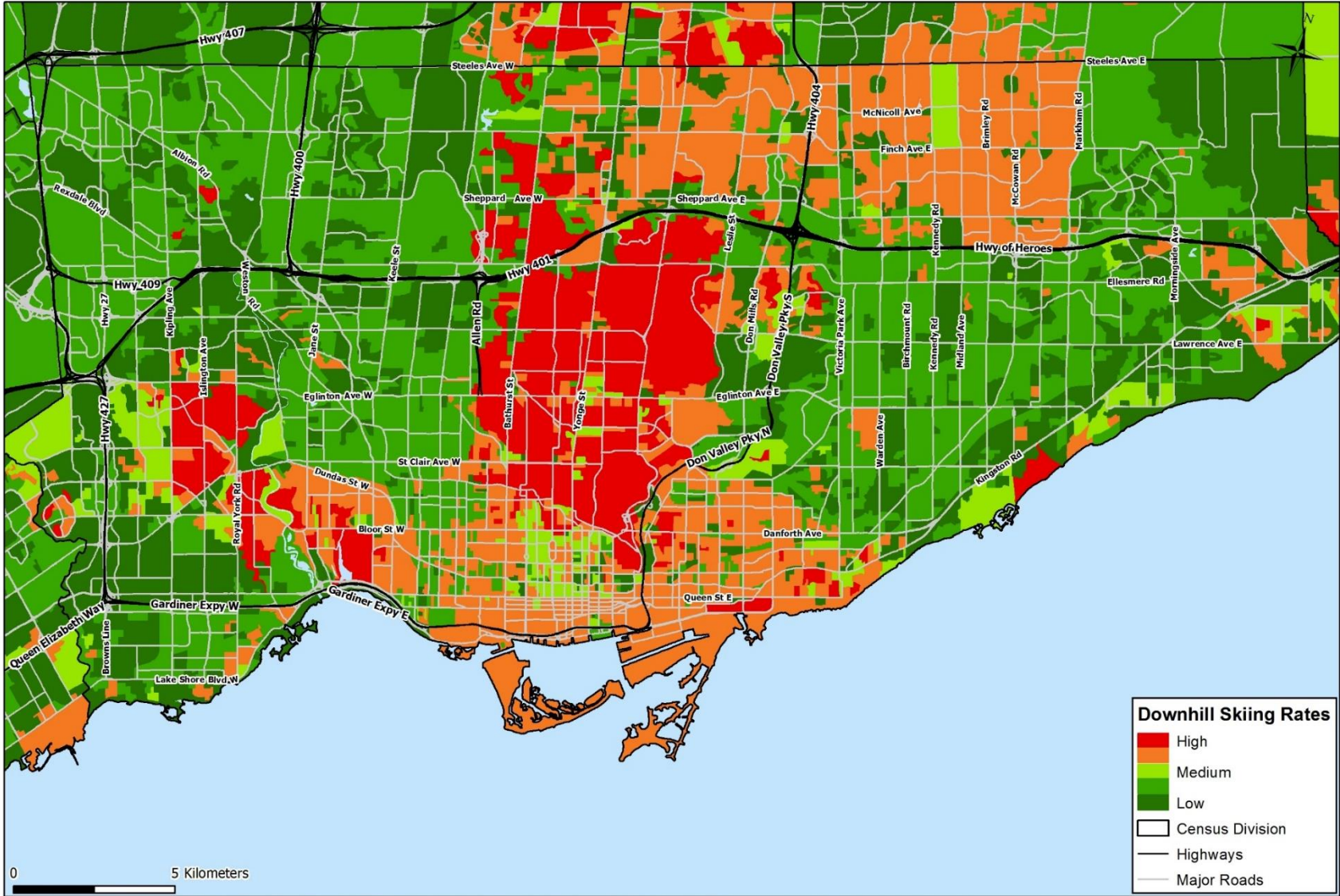
Final Cluster Rates						
PRIZM5	NAME	Downhill Skiing in the last year Rate	Own Ski Equipment Rate	Watch Skiing Television Rate	Attended Outdoor Shows/Exhibitions Rate	Travelled on Skiing Holiday Rate
01	Cosmopolitan Elite	0.733	0.500	0.333	0.400	0.633
02	Urbane Villagers	0.603	0.565	0.207	0.232	0.599
03	Arts & Affluence	0.655	0.456	0.051	0.344	0.466
04	Suburban Success	0.433	0.400	0.032	0.500	0.245
05	Asian Sophisticates	0.255	0.052	0.065	0.344	0.223
06	Kids & Careers	0.415	0.220	0.317	0.195	0.195
08	Boomerang City	0.134	0.040	0.033	0.033	0.055
10	Emptying Nests	0.010	0.000	0.065	0.075	0.025
11	Urban Digerati	0.480	0.249	0.249	0.249	0.220
12	Street Scenes	0.433	0.224	0.216	0.202	0.204
13	Asian Avenues	0.451	0.207	0.207	0.231	0.225
14	Diversity Heights	0.000	0.000	0.000	0.000	0.000
15	Heritage Hubs	0.000	0.000	0.000	0.000	0.000
16	Pets & PCs	0.355	0.299	0.120	0.125	0.150
18	Management Material	0.091	0.091	0.091	0.182	0.091
19	Grey Pride	0.167	0.000	0.000	0.167	0.000
20	South Asian Achievers	0.056	0.000	0.111	0.111	0.167
22	Aging in Suburbia	0.000	0.000	0.000	0.000	0.000
23	Asian New Wave	0.045	0.045	0.000	0.000	0.000
24	Fresh Air Families	0.000	0.000	0.000	0.000	0.000
.
.
.

High

Low

Rates then created for all 68 segments in exactly the way described above

Downhill Skiing Rates Toronto Census Division by Dissemination Area



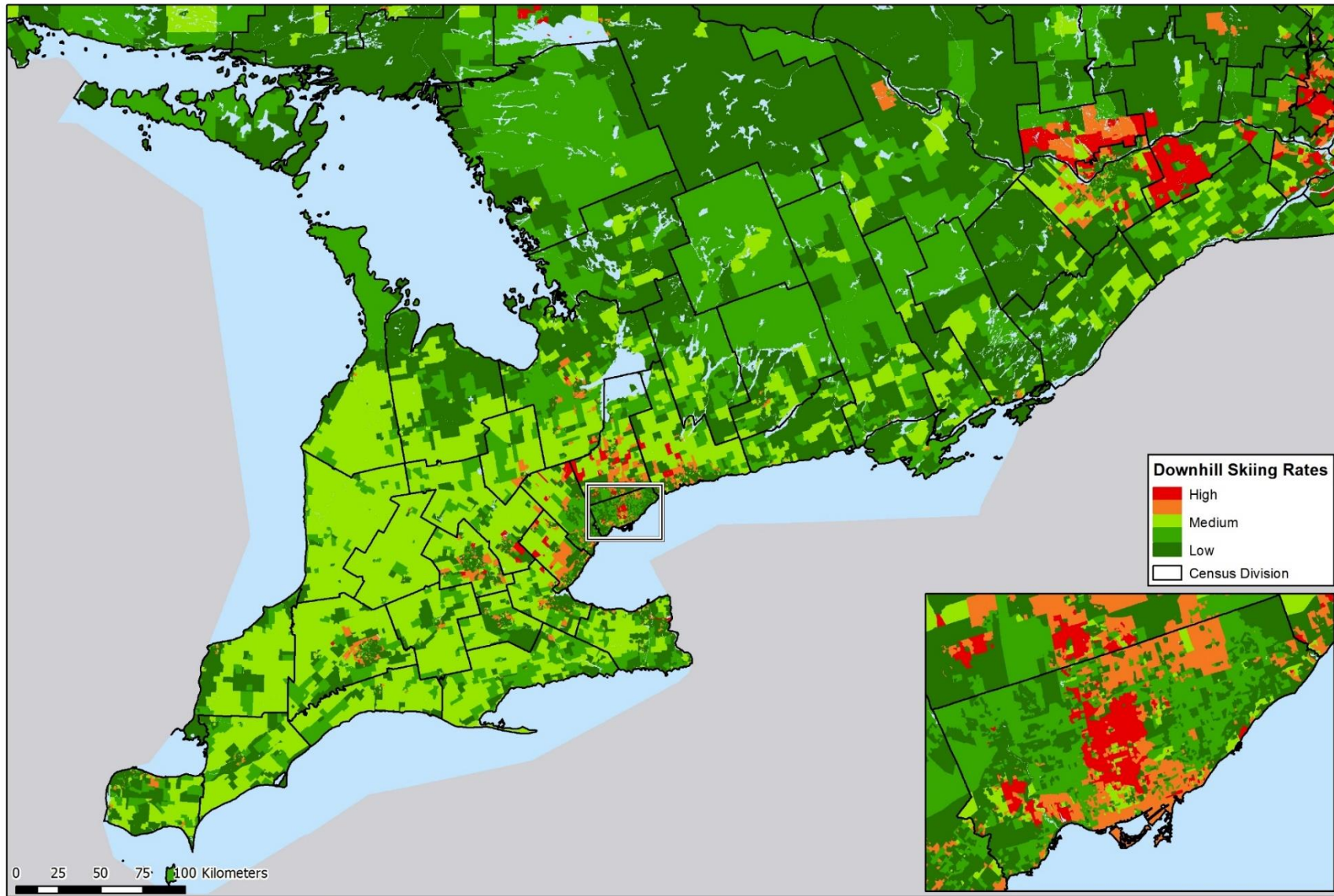
A choropleth map of all DAs downhill skiing participation rates based on their PRZM5 rates

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

Final Cluster Rates

PRIZM5	NAME	Downhill Skiing in the last year Rate	Own Ski Equipment Rate	Watch Skiing Television Rate	Attended Outdoor Shows/Exhibitions Rate	Travelled on Skiing Holiday Rate
01	Cosmopolitan Elite	0.733	0.500	0.333	0.400	0.633
02	Urbane Villagers	0.603	0.565	0.207	0.232	0.599
03	Arts & Affluence	0.655	0.456	0.051	0.344	0.466
04	Suburban Success	0.433	0.400	0.032	0.500	0.245
05	Asian Sophisticates	0.255	0.052	0.065	0.344	0.223
06	Kids & Careers	0.415	0.220	0.317	0.195	0.195
08	Boomerang City	0.134	0.040	0.033	0.033	0.055
10	Emptying Nests	0.010	0.000	0.065	0.075	0.025
11	Urban Digerati	0.480	0.249	0.249	0.249	0.220
12	Street Scenes	0.433	0.224	0.216	0.202	0.204
13	Asian Avenues	0.451	0.207	0.207	0.231	0.225
14	Diversity Heights	0.000	0.000	0.000	0.000	0.000
15	Heritage Hubs	0.000	0.000	0.000	0.000	0.000
16	Pets & PCs	0.355	0.299	0.120	0.125	0.150
18	Management Material	0.091	0.091	0.091	0.182	0.091
19	Grey Pride	0.167	0.000	0.000	0.167	0.000
20	South Asian Achievers	0.056	0.000	0.111	0.111	0.167
22	Aging in Suburbia	0.000	0.000	0.000	0.000	0.000
23	Asian New Wave	0.045	0.045	0.000	0.000	0.000
24	Fresh Air Families	0.000	0.000	0.000	0.000	0.000
.
.
.

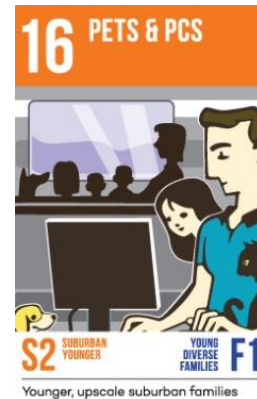
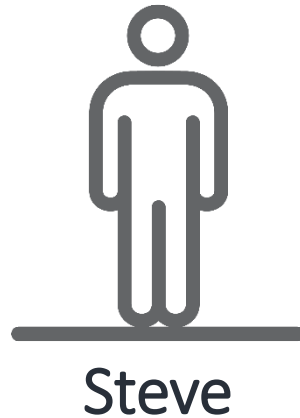
Downhill Skiing Rates Southern Ontario by Dissemination Area



Copyright © 2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

OTHER USES: RETAIL SITE SELECTION

- Lets say that a local retailer is looking to expand to a new location in Durham Region
- They analyze their data and find that Steve and others in similar neighbourhoods are the highest spending and most loyal patrons in their database
- To proceed they should “profile” Steve and colleagues using PRIZM5 and try to find neighbourhoods with people like Steve
- Then select a location that is nearby



People Like Steve
Durham Region



There are quite a few options actually

Variable Description: Target Sets
Copyright © 2016 Environics Analytics (EA). Source: PRIZMS 2016 - PRCDDA. PRIZM is a registered trademark of The Nielsen Company (U.S.).



THE ECOLOGICAL FALLACY – A POSSIBLE DOWNSIDE

- This geodemographic approach is becoming more important in a BIG DATA world focused on analytics and concerned with privacy
- **However**, the approach can yield erroneous inferences because of the celebrated problem called the “**ecological fallacy**” and this should be researched more
- The fallacy is: “because a small area, or a small group, behaves like X on the average, that all persons in the set behave like X”
- In the marketing realm researchers have found that the fallacy occurs, but it occurs infrequently ... and is weak
 - This means that if an area behaves like X on average *that most people or households* from this area generally behave like X
- **Leveraging the geodemographic approach is a critical tool for marketers and many other purposes**

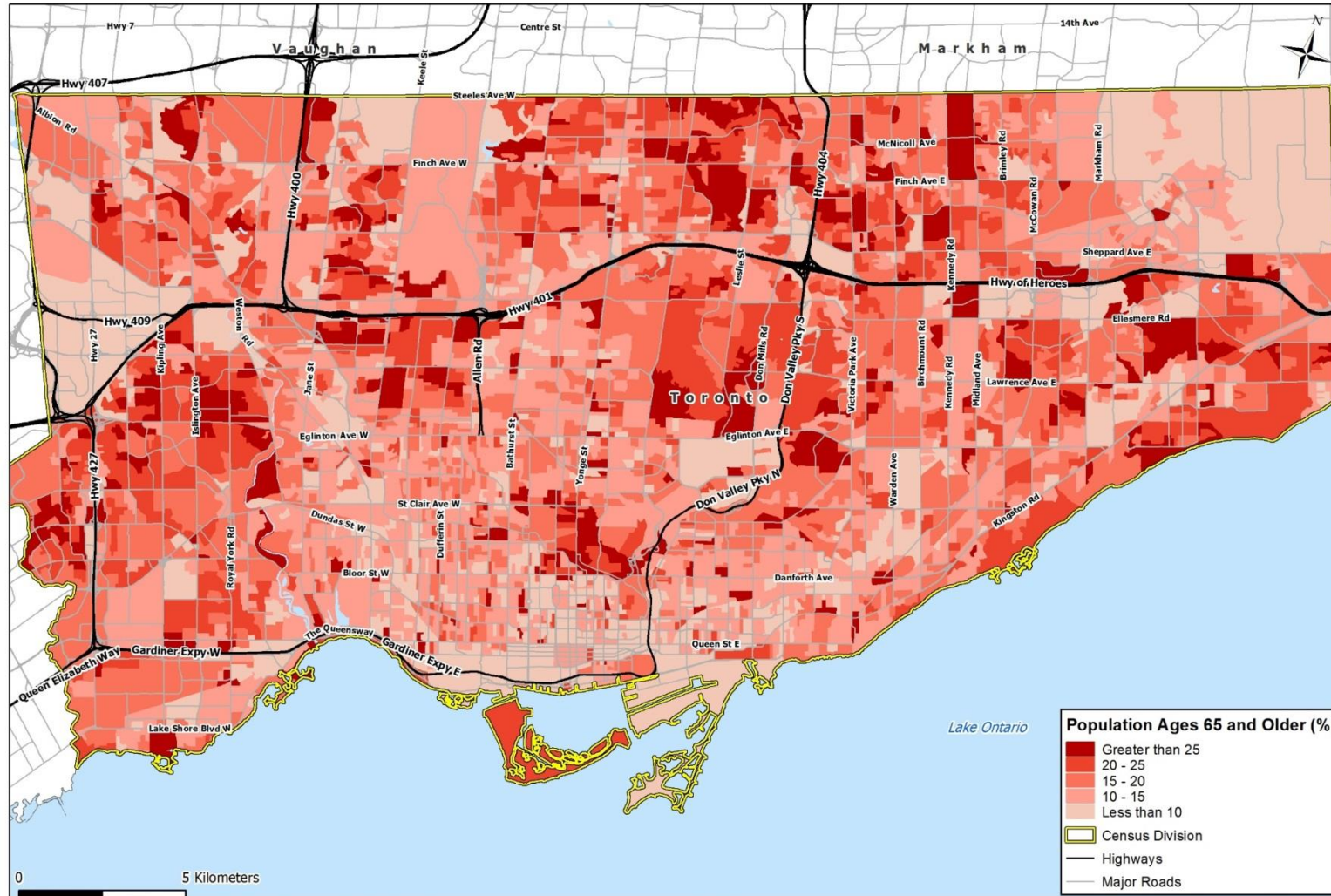
THE SEGMENTS NEED TO MAKE SENSE

- To feel comfortable about a simple cluster-based approach requires that users believe that the clusters/segments make sense as a classification system for the behaviour or purpose being studied
- The EA PRIZM5 segments were developed as general purpose market segments using: Census demographic and socioeconomic data, geographical context data, and some behavioural data and are updated annually based on new data
- Other countries have similar geodemographic segmentation systems



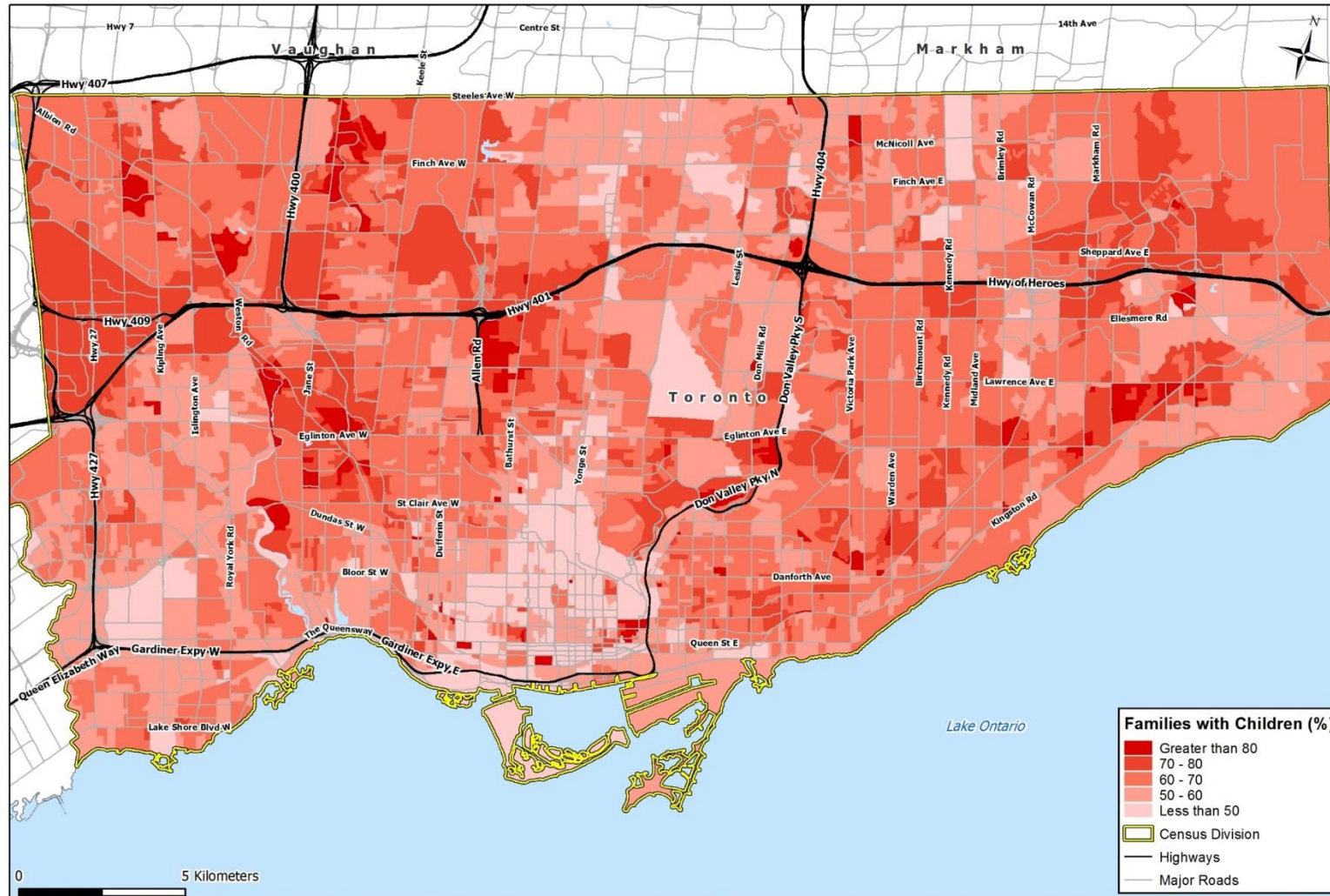
...NOW FOR SOME MORE FUN MAPS

- Let's quickly look at the kinds of maps that geodemographers and GIS analysts create and view each day
- First I show a few maps of interesting census-like socioeconomic and demographic variables
- Helpful in marketing research in understanding how these variables and trends can be used for estimating who likes certain products and services
- Then I show some derived maps of interesting behaviours and values



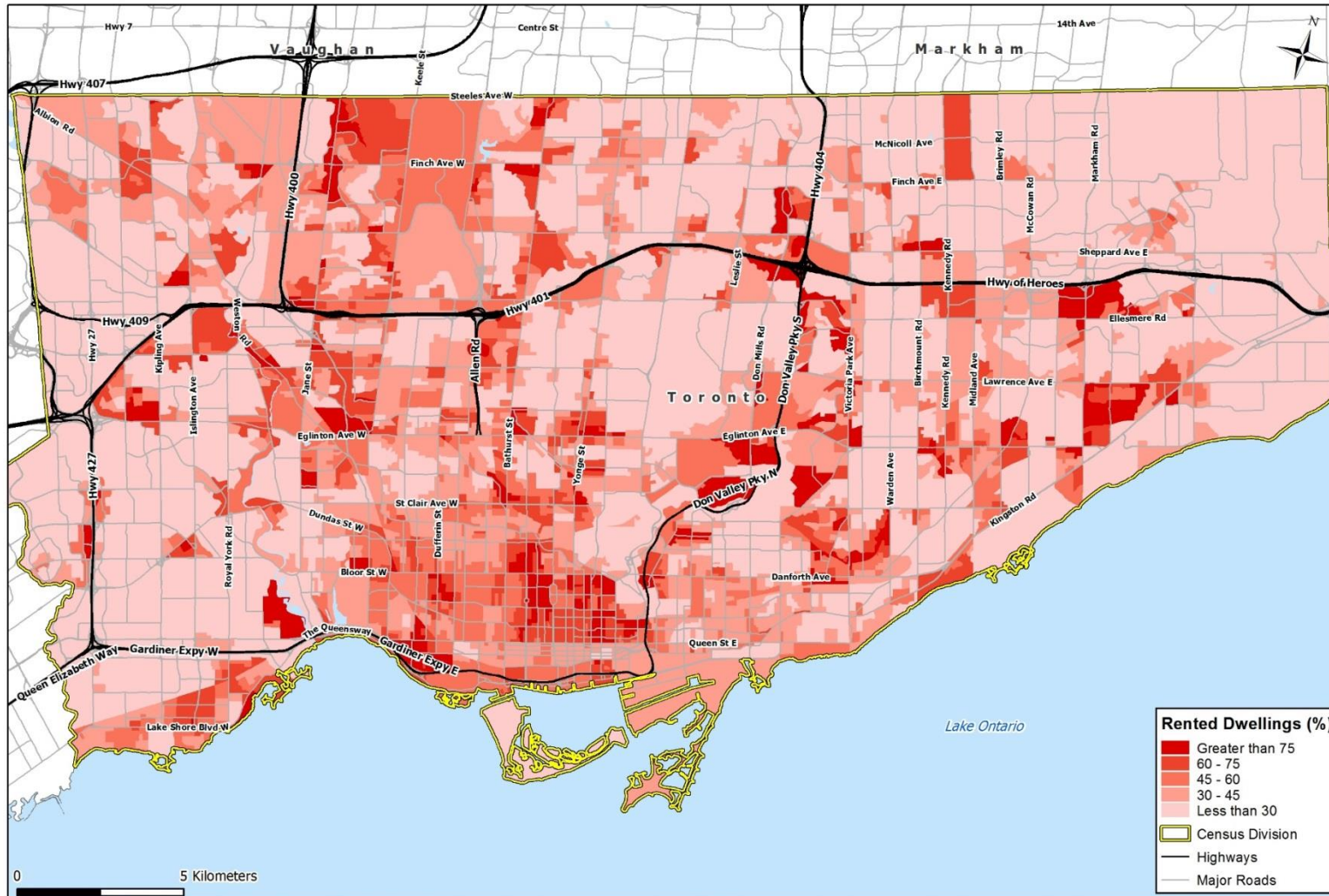
% age
65 plus

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



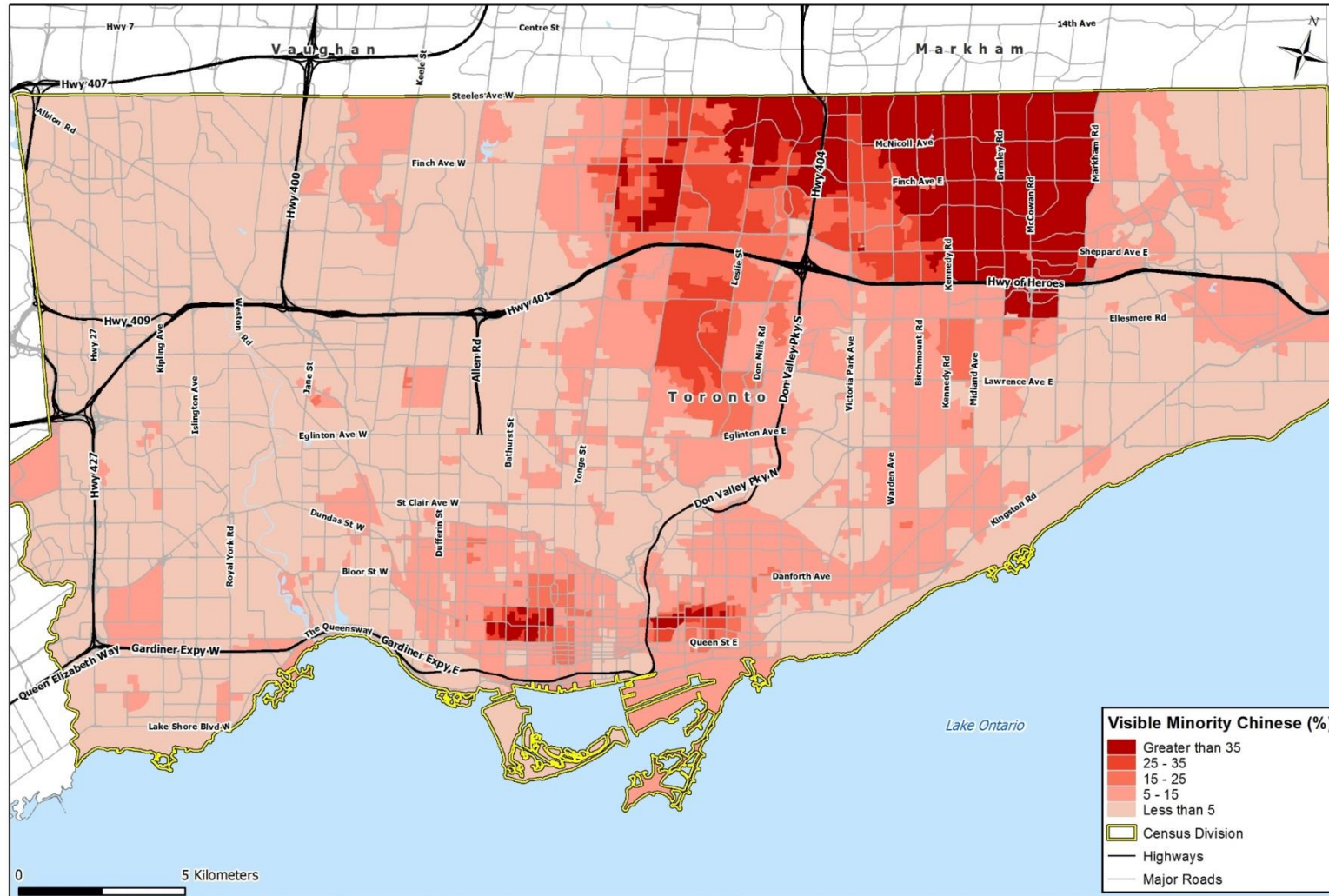
%
families
with
children

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM5 and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



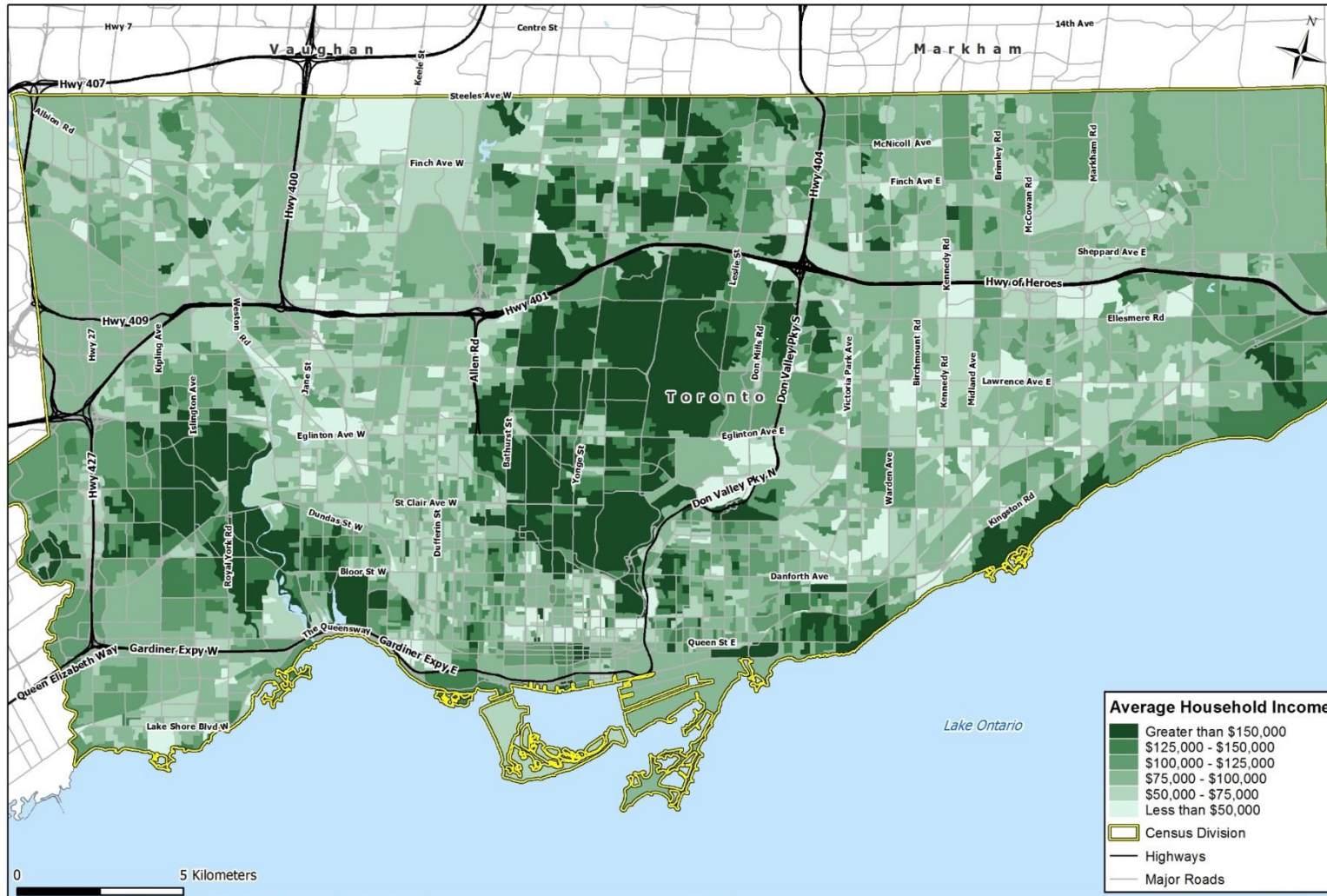
% occupied dwellings that are rented

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM5 and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



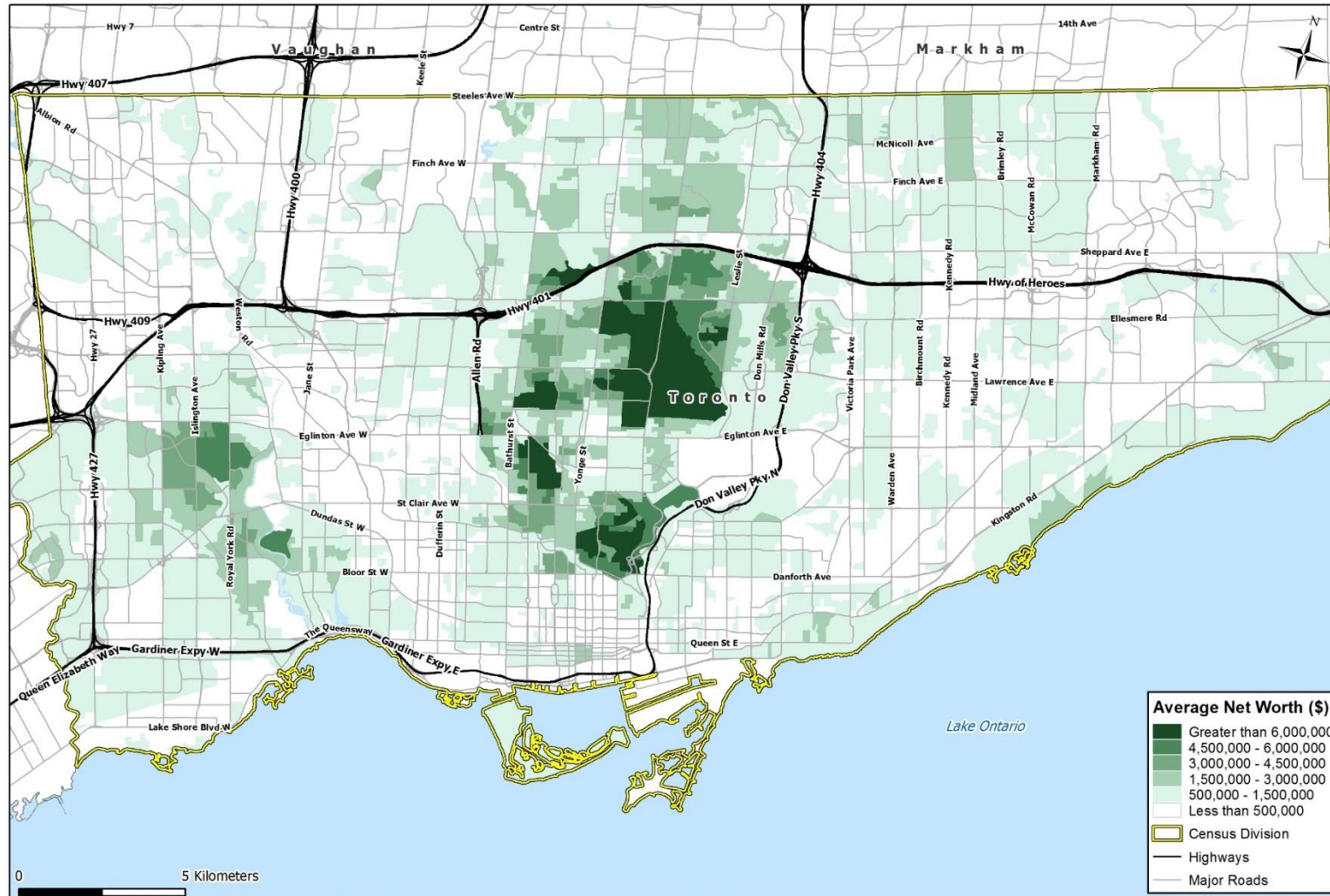
%
population
that is
visible
minority
Chinese

Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM5 and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



Average household income

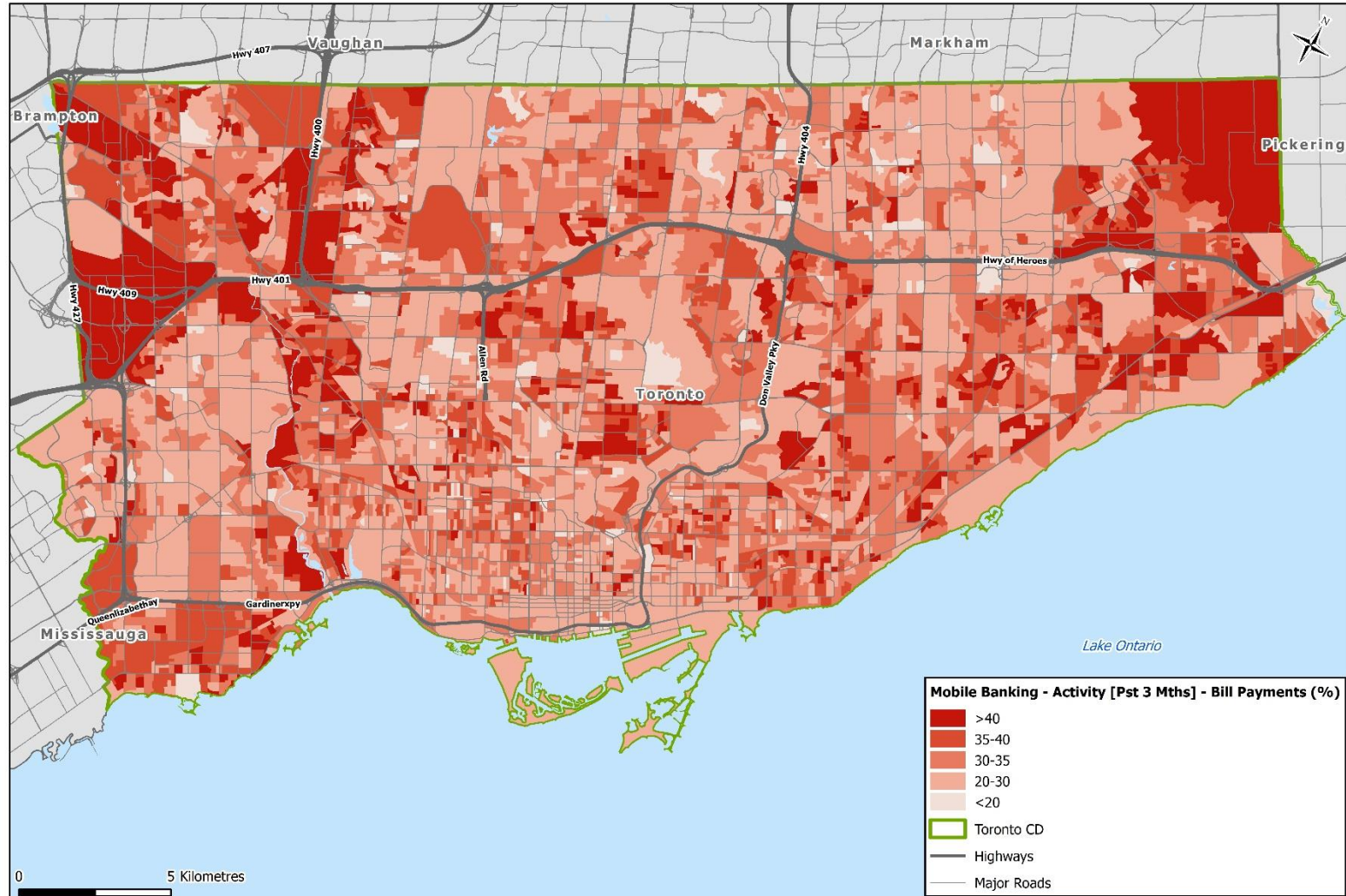
Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



Copyright ©2015 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM5 and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

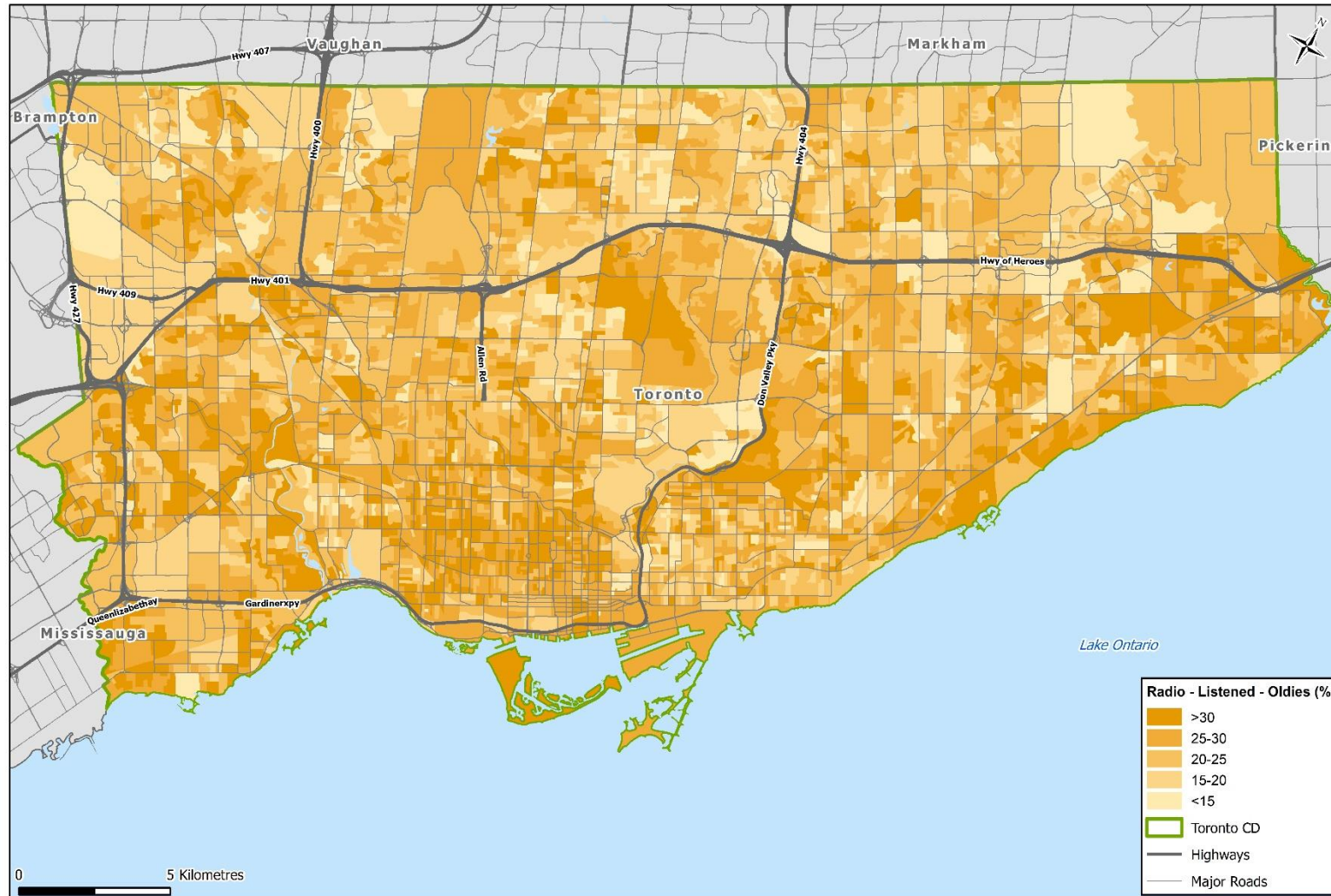
Net worth is created by EAG - not census-based

NOW BEHAVIOUR...



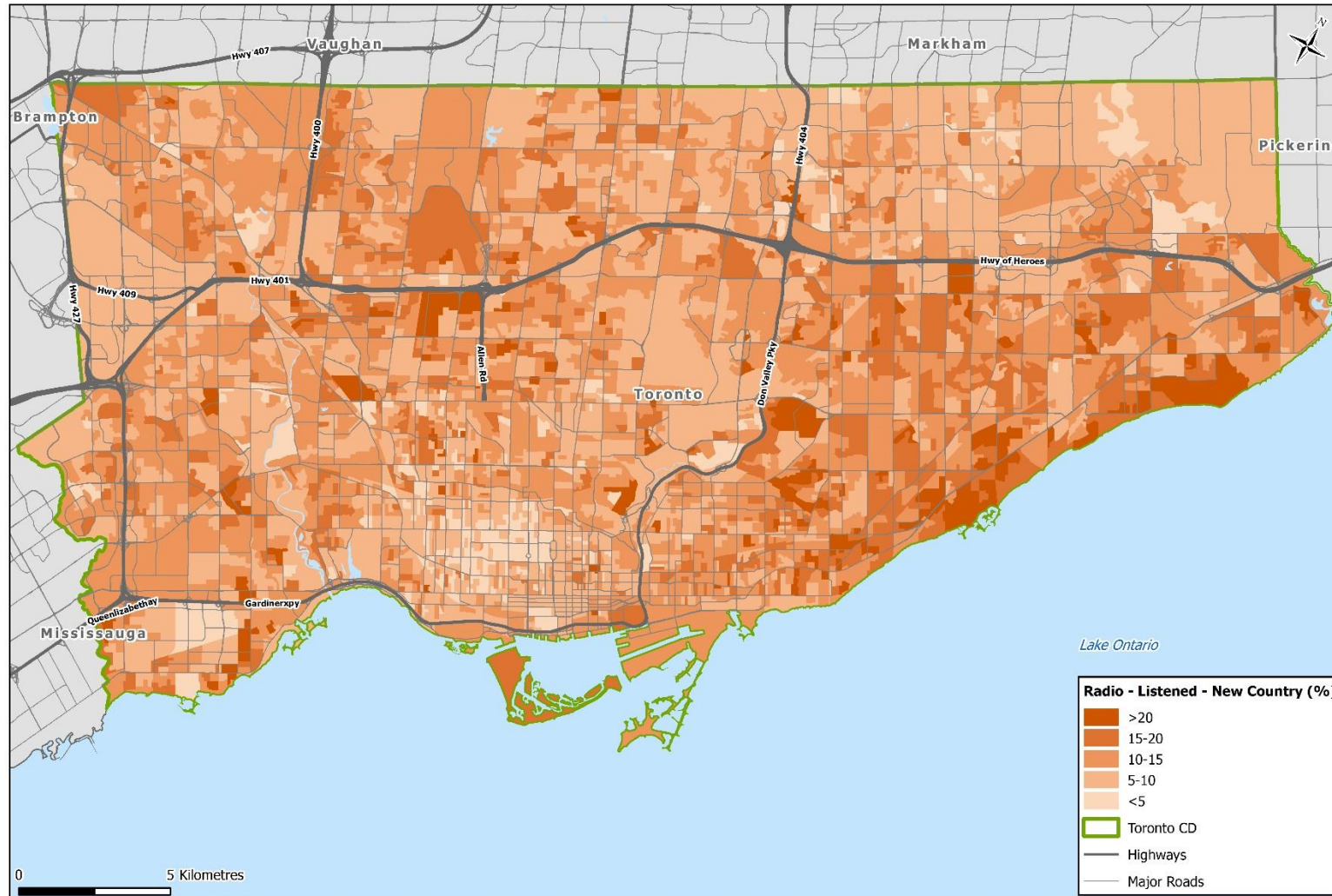
% that used mobile banking (past 3 months)

Copyright ©2018 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



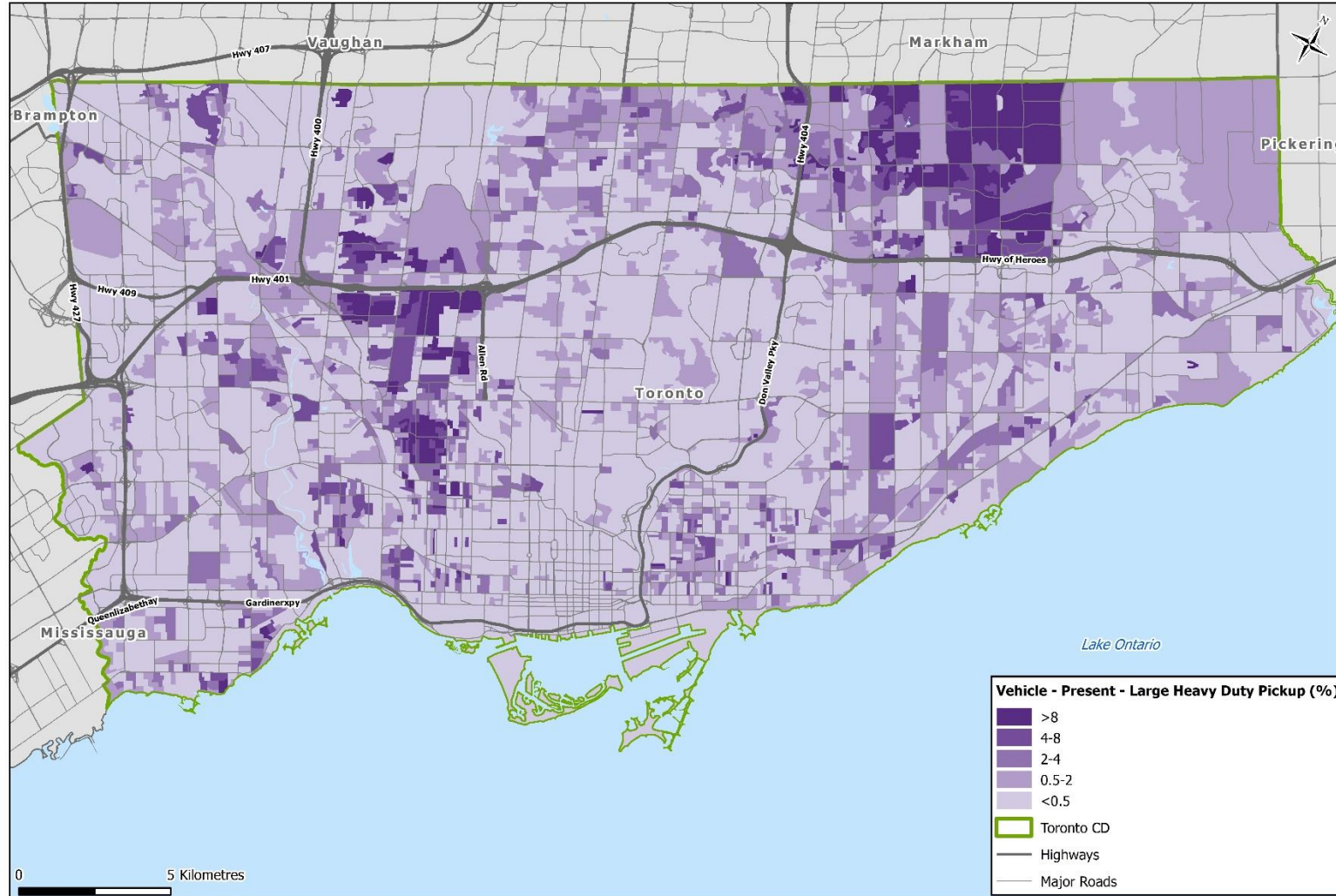
% that listen to oldies radio

Copyright ©2018 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM5 nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



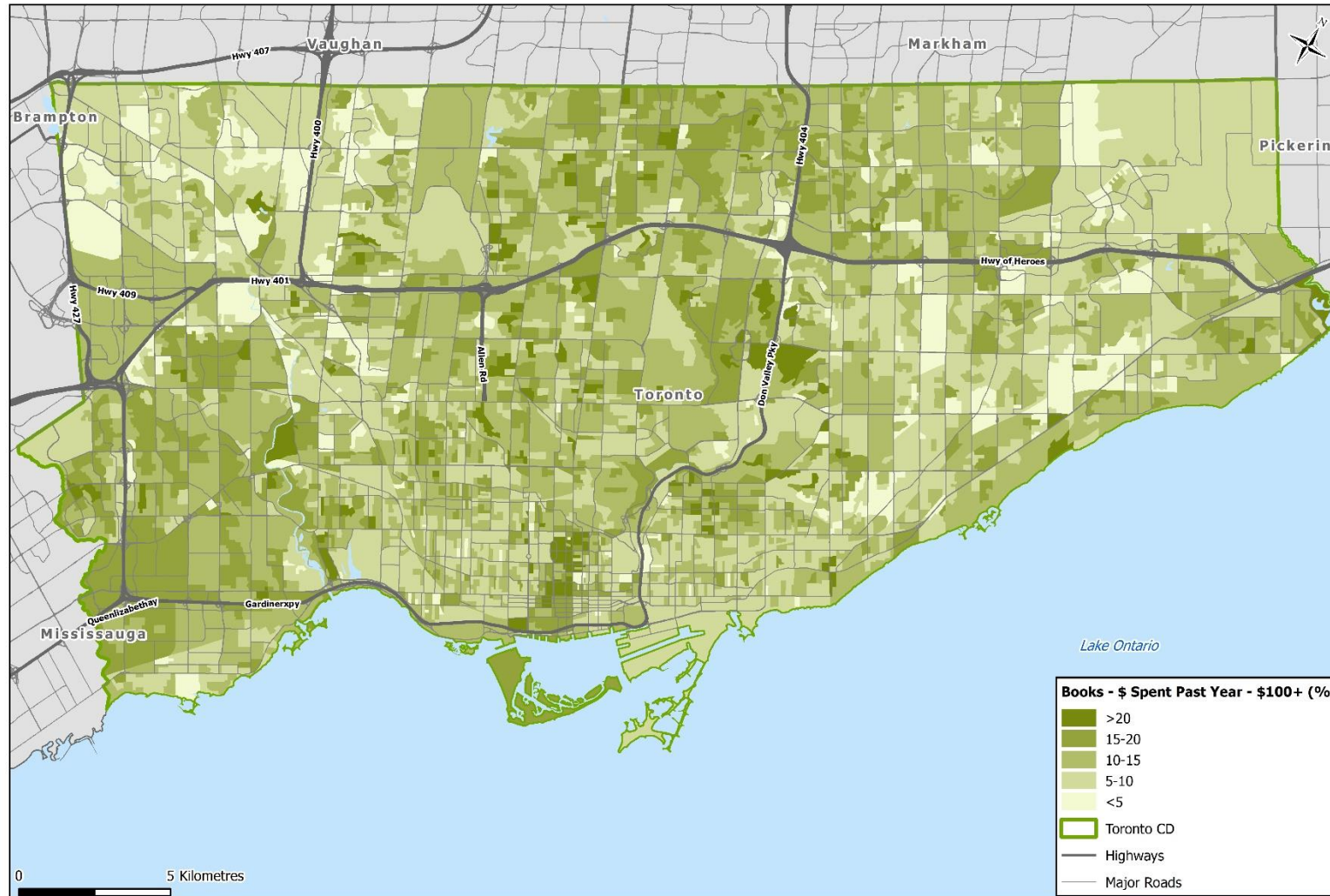
% that listen to new country radio

Copyright ©2018 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



% that own a large heavy pick-up truck

Copyright ©2018 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZM nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.



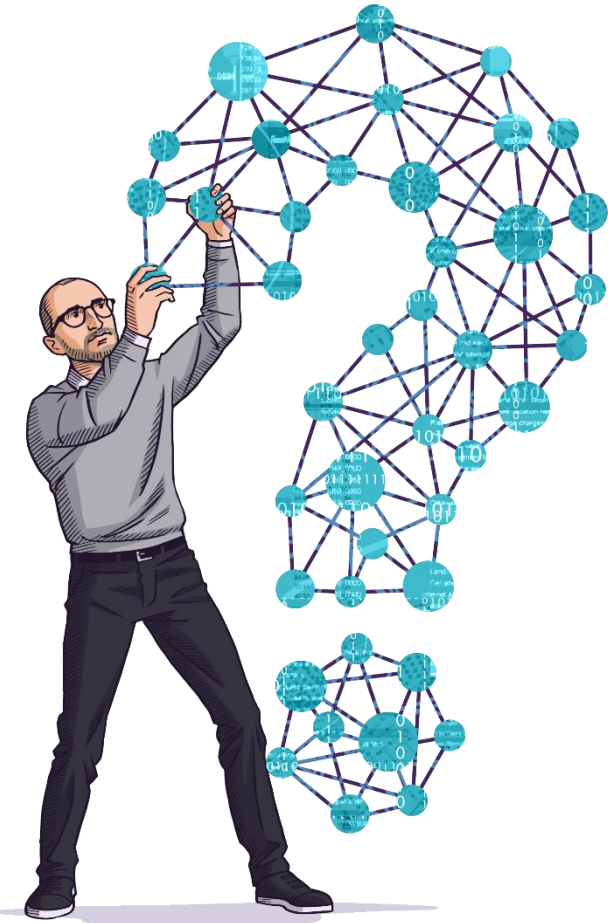
% that spent over \$100 on books in the last year

Copyright ©2018 Environics Analytics. Environics Analytics acquires and distributes Statistics Canada files in accordance with the Government of Canada's Open Data Policy. No information on any individual or household was made available to Environics Analytics by Statistics Canada. PRIZM and selected PRIZMS nicknames are registered trademarks of The Nielsen Company (U.S.) and are used with permission.

CONCLUSIONS

- Lots of good data in almost every area (customer data, big data, large surveys)
- These data are tied to persons one at a time and goal of integrated view of the customer is not possible
- Aggregation to small area geography allows all datasets with a spatial reference to be tied together to analyse and derive conclusions
- The geodemographic approach
- Socioeconomics and demographics at small area level can be clustered well to make good looking meaningful segments
- Segments simplify profiling of customers and also targeting of new customers

QUESTIONS?



THANK YOU

Tony Lea

Chief Methodologist and Senior Vice President



@EnvironicsA

ENVIRONICS
ANALYTICS