Zooming in on ZIP codes

Factor based models for US longevity

12.30pm ET

9 September 2019

@ClubVita #longevityZIP

linkedin.com/company/club-vita

Club Vita US, LLC

Club Vita

Proper noun, [kluhb vee-*tuh*], \ˈkləb vē-tə\


2008


2015


2019

1. Center of excellence for improving understanding of human longevity.
2. Community of organizations with a shared interest in longevity and belief that the ‘bigger’ the data, the lower the (statistical) noise.
3. Provider of longevity risk informatics to support pension funds’ risk management strategies and enable market innovation.

Club Vita is an independent data utility,
supporting pension funds, advisors, insurers & asset managers

CLUB VITA

Introducing today's panel

Douglas Anderson
Founder
Club Vita

Bruce Cadenhead
Global Chief Actuary - Wealth
Mercer

Matt Forrest
Director of Spatial Data Science
Carto

Steven Baxter
Head of Innovation & Development
Club Vita

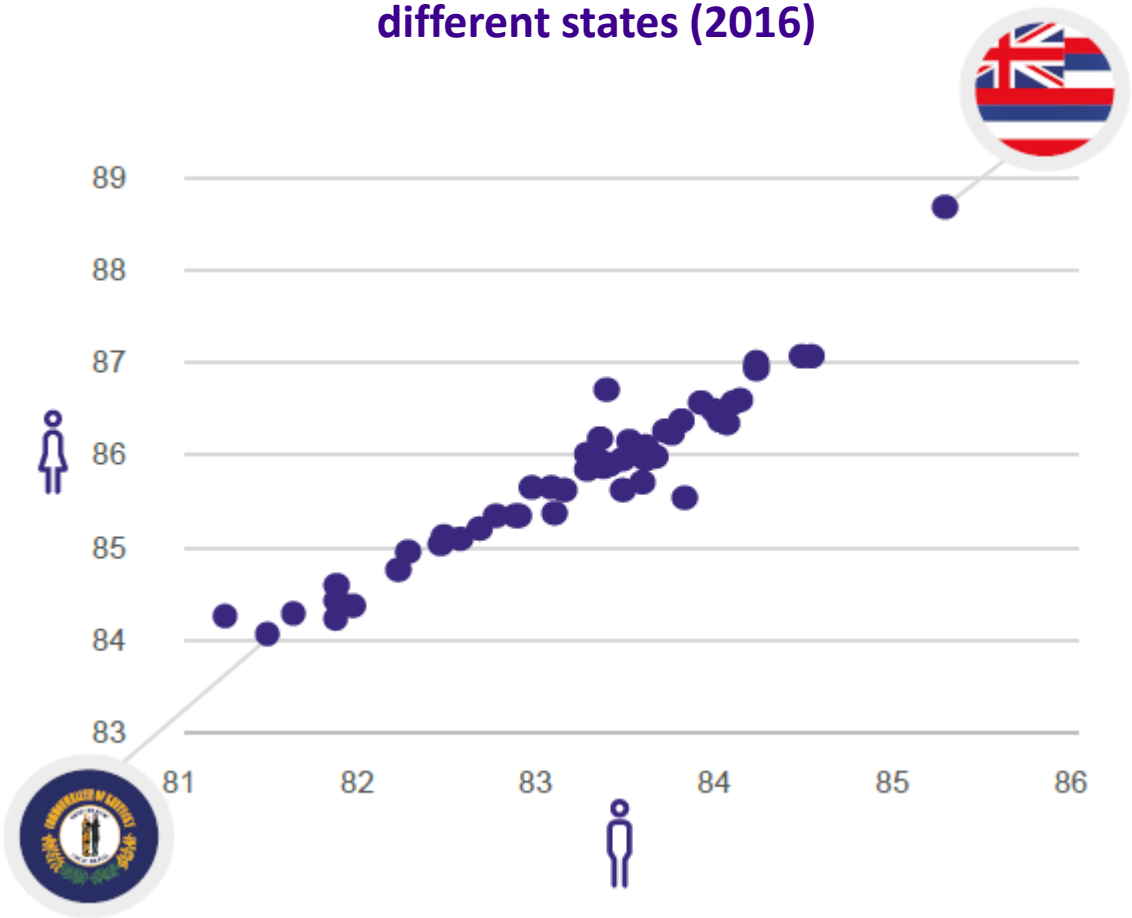
Join the discussion…
LinkedIn: Friends of Club Vita group
Twitter: @ClubVita #longevityZIP

Why look at ZIP codes for longevity?

Steven Baxter
Head of Innovation & Development
Club Vita

4+ year life expectancy gap between states

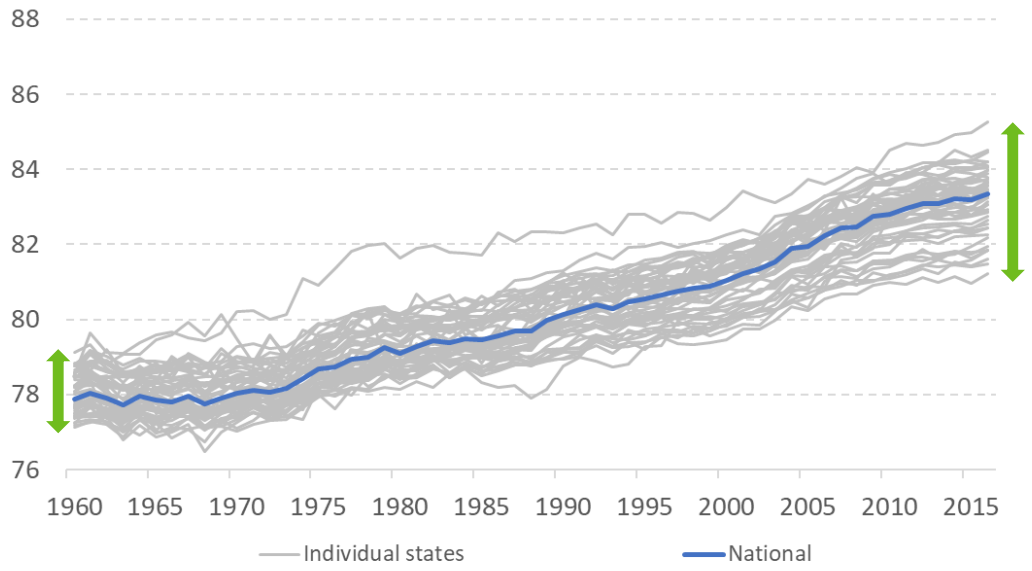
Life expectancy from age 65 for different states (2016)



Source: Club Vita analysis of United States Mortality Database (https://usa.mortality.org/)

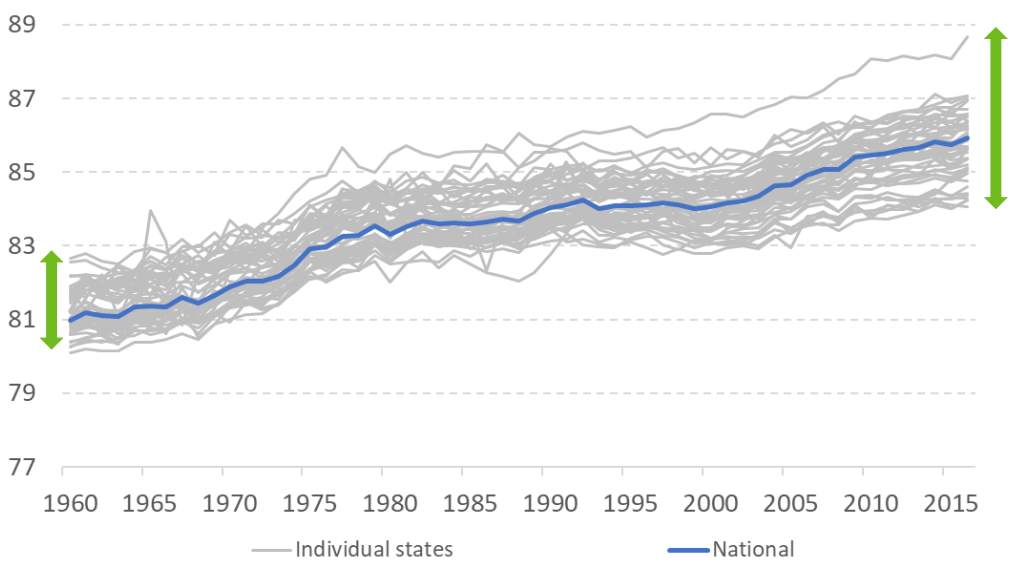
Longer lifespans, but widening gaps

Life expectancy from 65



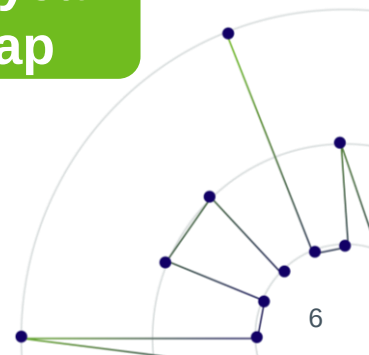
2 year gap → **4 year gap**

Life expectancy from 65



2.5 year gap → **4.6 year gap**

What is driving this diversity? And how can we describe it?



CLUB VITA

What affects how long people live?

Genetics

Is it not all written in our genes?

Paper	Genes	Environment	Group studied
Herskind et al., 1996	26%	74%	Danish twins
Ljungquist et al., 1998	33%	67%	Swedish twins
Gavrilova et al., 1998	18%	82%	Royal families
Mitchell et al., 2001	25%	75%	Amish
Skytthe et al, 2003	25%	75%	Danish twins
Joshi et al (in prep)	16%	84%	Scottish nuclear families

- **Research suggests lifespan is only 20% genetic**
- **Lifespan driven more by nurture than nature**

What else affects how long people live?

Lifestyle

Affluence

Occupation

Health

Gender

Something else?

Why 9-digit ZIP codes?

79936 El Paso, TX
Population 115k

90011 Los Angeles, CA
Population 106k

60629 Chicago, IL
Population 105k

90650 Norwalk, CA
Population 105k

90201 Bell Gardens, CA
Population 101k

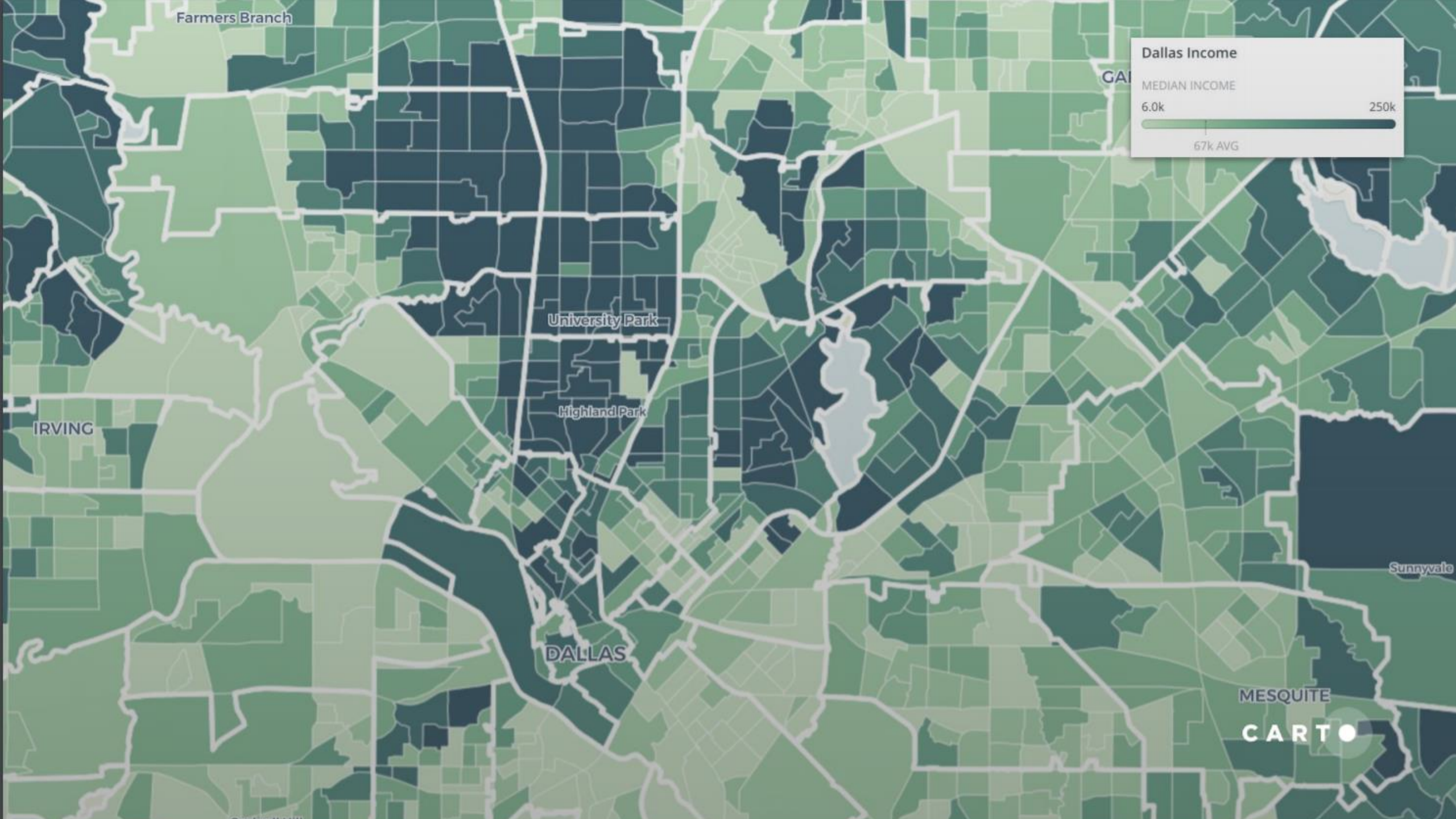
On average, a 5 digit ZIP code covers 7,800 people – but in many cases covers more than 100,000 residents

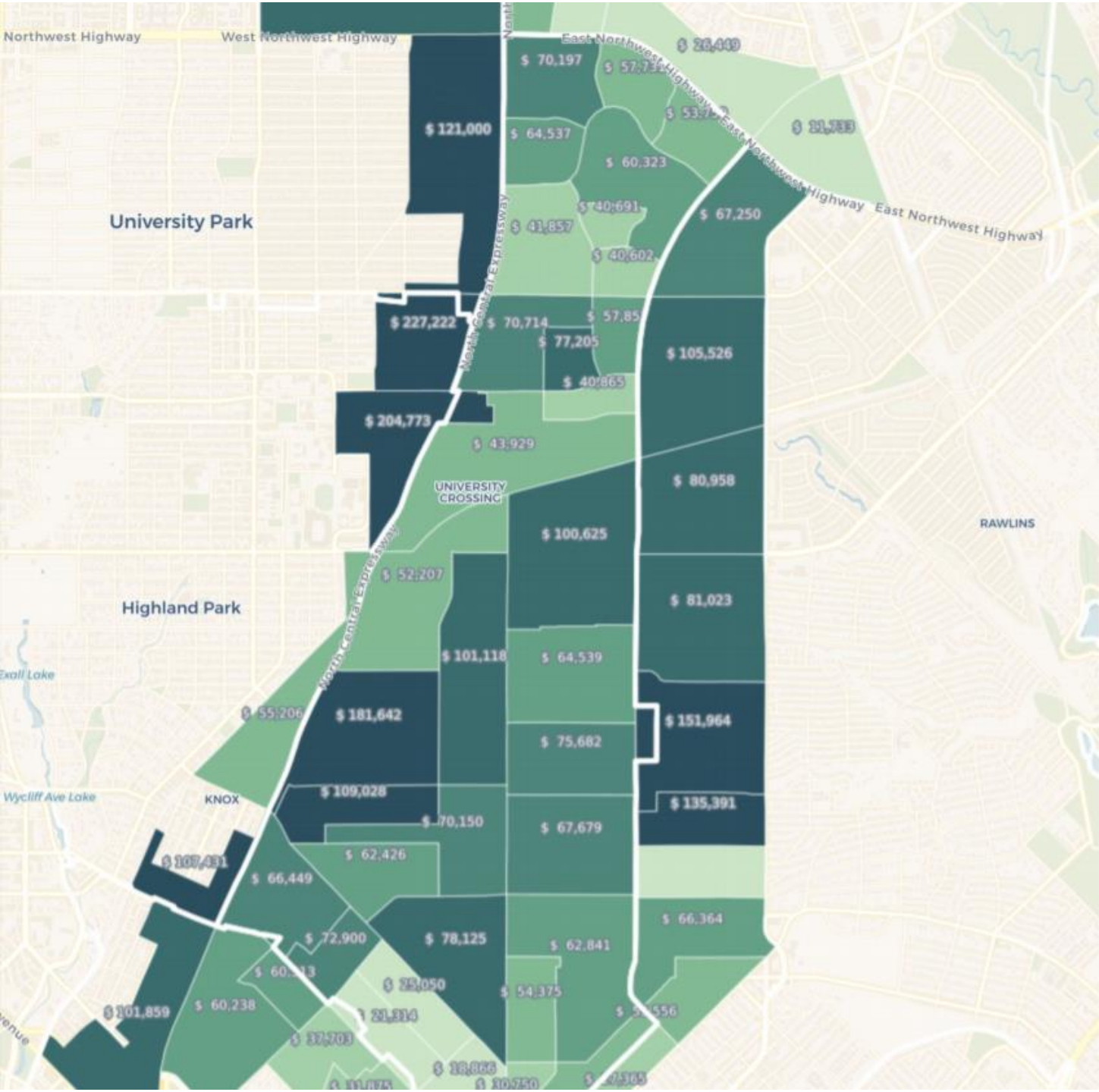
Source: localistica, as at 4 April 2019

What can ZIP(+4) codes tell us about human behaviour?

Matt Forrest
Director of Spatial Data Science
CARTO


Spatial Modeling with CARTO
The difference between knowing where and knowing why
CARTO


Farmers Branch
University Park
Highland Park
IRVING
DALLAS
MESQUITE
Sunnyvale
Dallas Income
MEDIAN INCOME
6.0k
250k
67k AVG
CARTO



Median Income

B0 Source dallas_for_deck

0 NULL ROWS 0 MIN 69K AVG 0 MAX

59 SELECTED

CARTO

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

Tobler's First Law of Geography

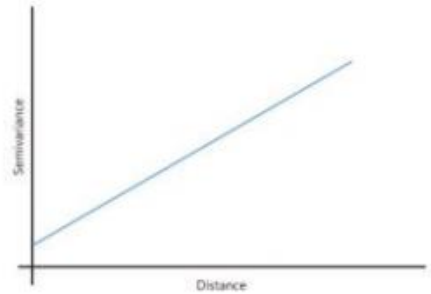
CARTO

The phenomenon external to an area of interest affects what goes on inside.

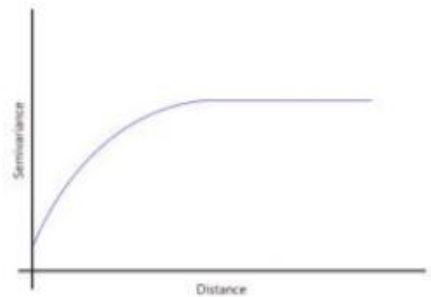
Tobler's Second Law of Geography

Human behavior is related spatially, and external factors have an impact on that behavior.

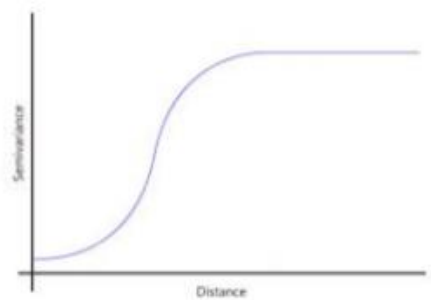
The influence “decays” as you move away from your target



Linear



Circular

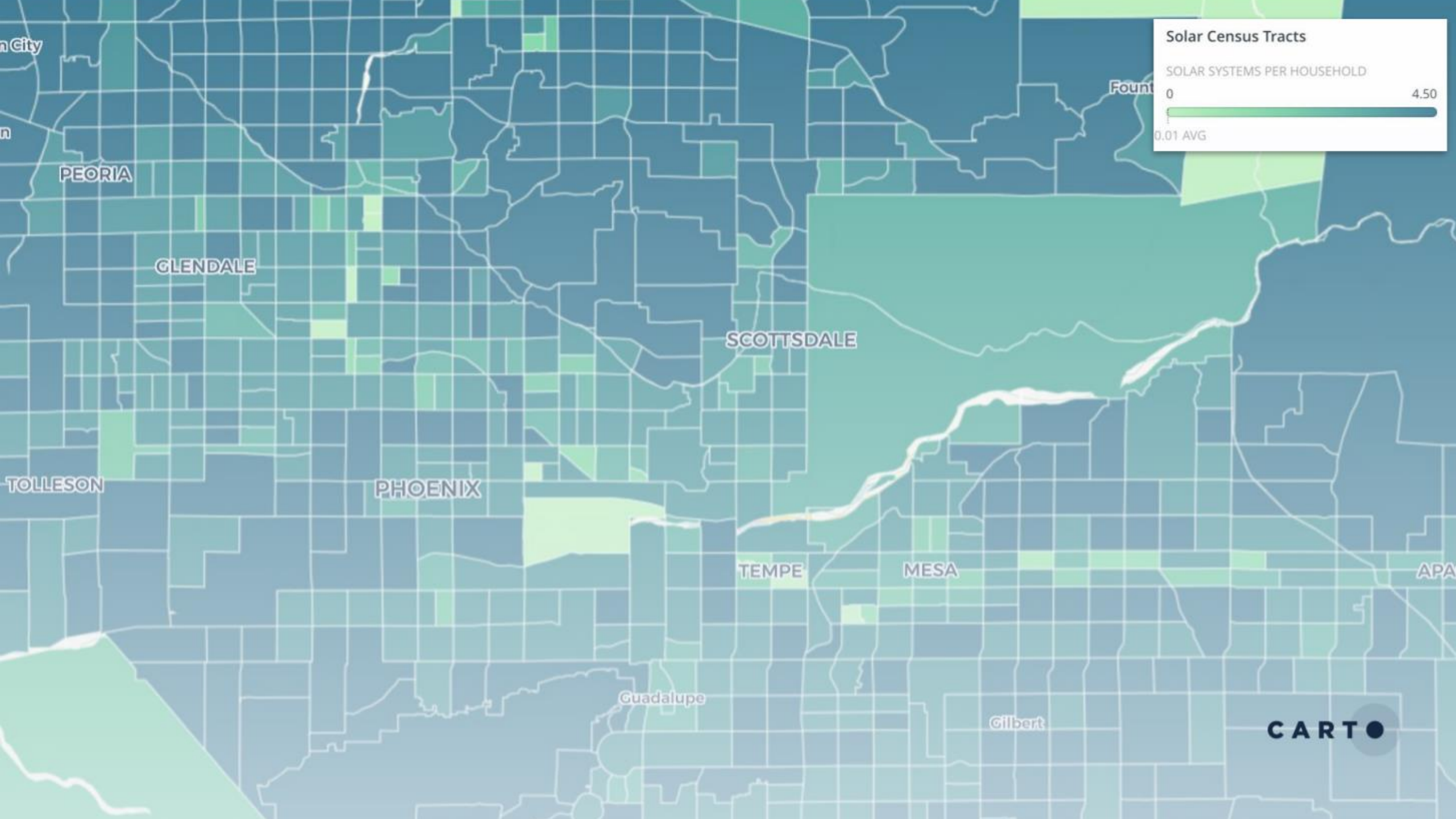


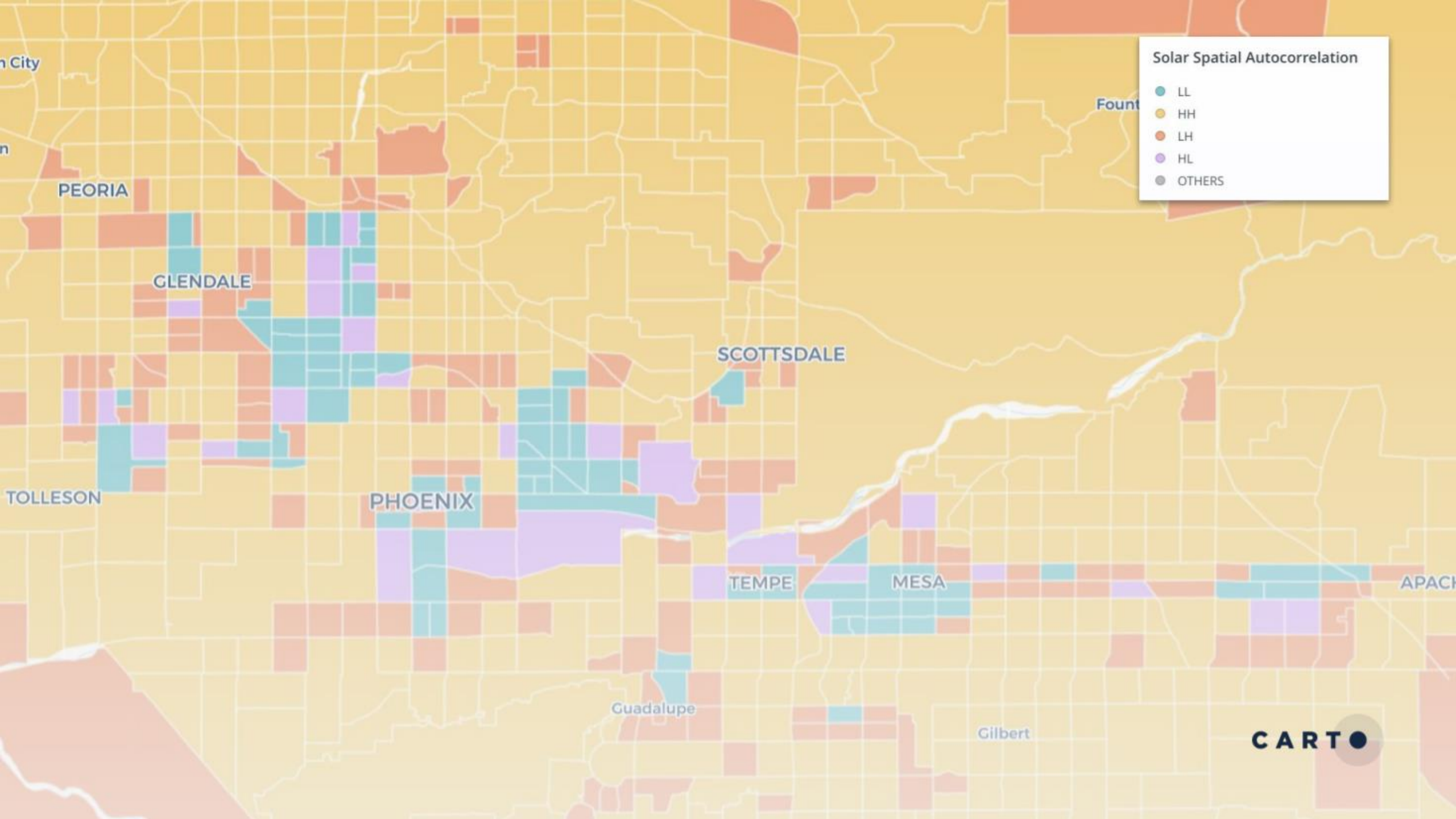
Gaussian (Radius)

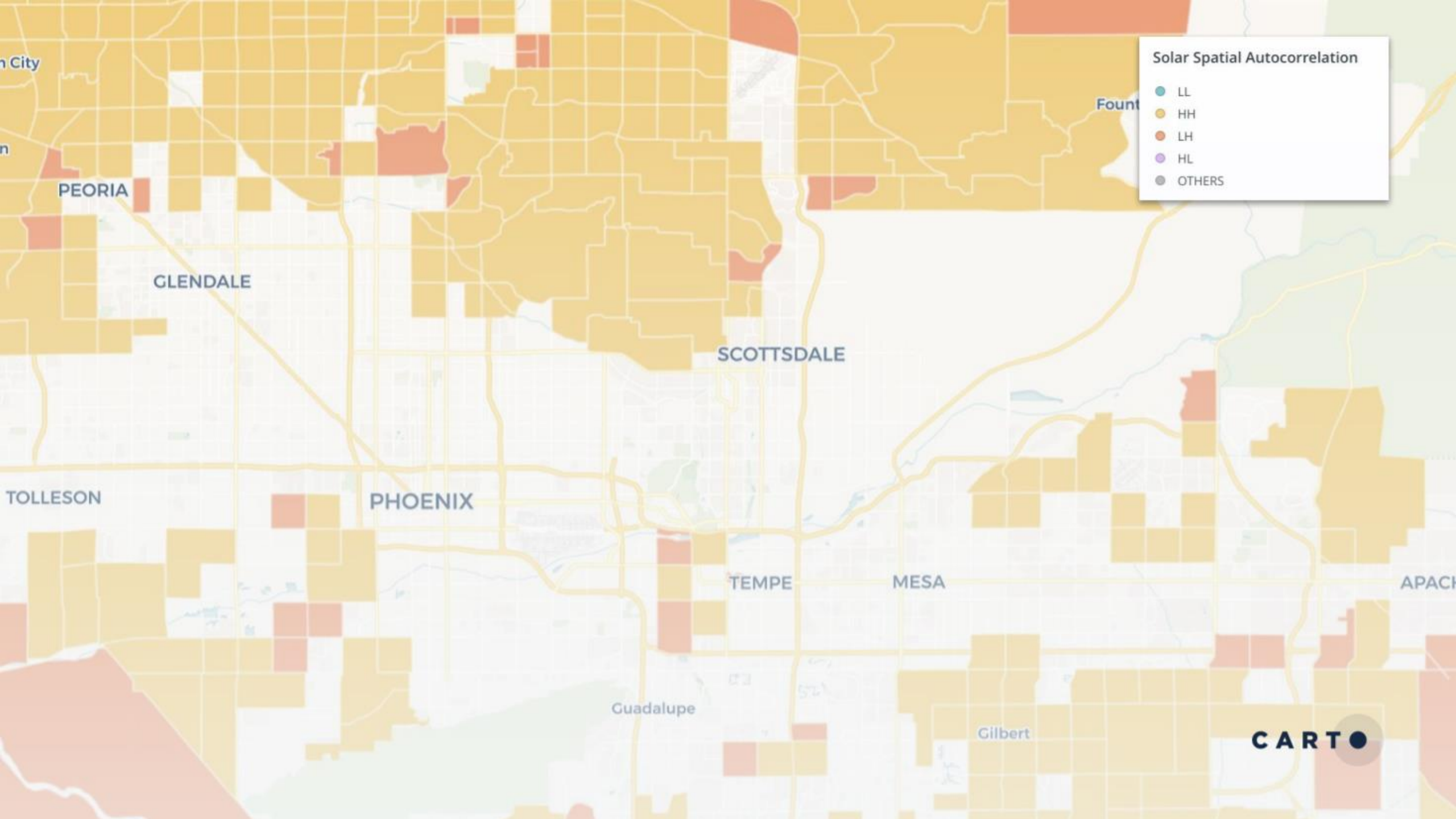
Features can be spatially correlated, yet arranged differently

Positive Spatial Autocorrelation

Negative Spatial Autocorrelation

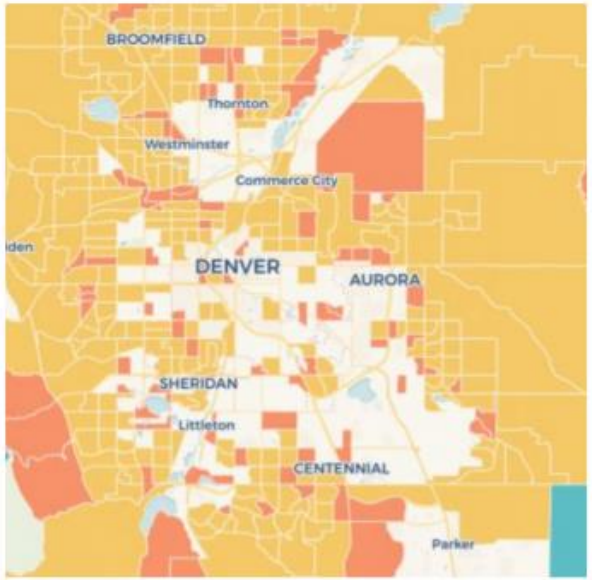

Solar Census Tracts
SOLAR SYSTEMS PER HOUSEHOLD
0
4.50
0.01 AVG
PEORIA
GLENDALE
SCOTTSDALE
TOLLESON
PHOENIX
TEMPE
MESA
Guadalupe
Gilbert
CARTO

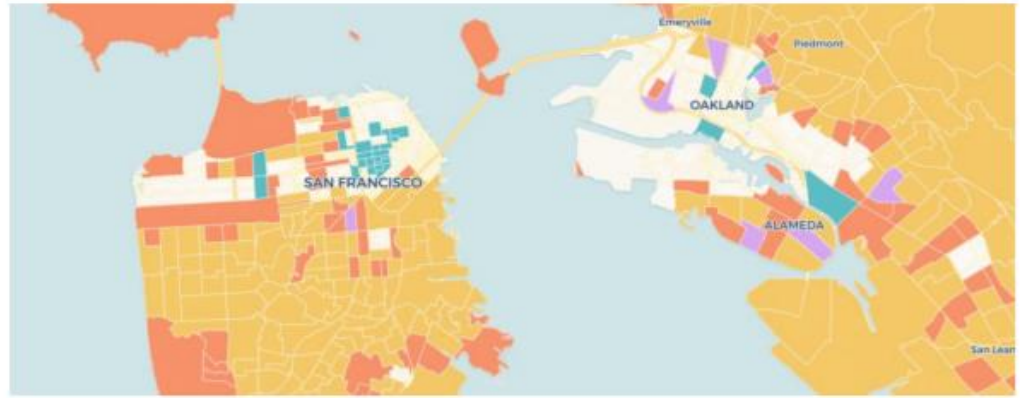

Solar Spatial Autocorrelation
LL
HH
LH
HL
OTHERS
PEORIA
GLENDALE
SCOTTSDALE
TOLLESON
PHOENIX
TEMPE
MESA
Guadalupe
Gilbert
CARTO

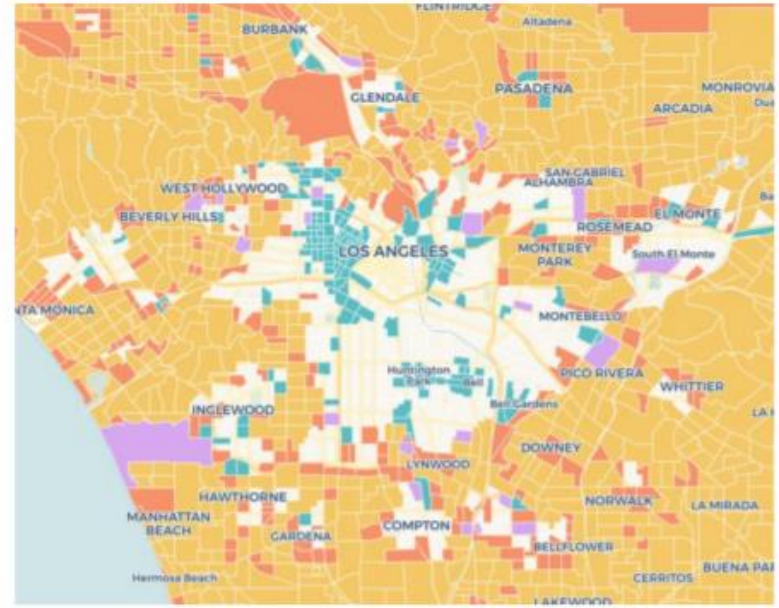

Solar Spatial Autocorrelation
LL
HH
LH
HL
OTHERS
PEORIA
GLENDALE
SCOTTSDALE
TOLLESON
PHOENIX
TEMPE
MESA
Guadalupe
Gilbert
CARTO

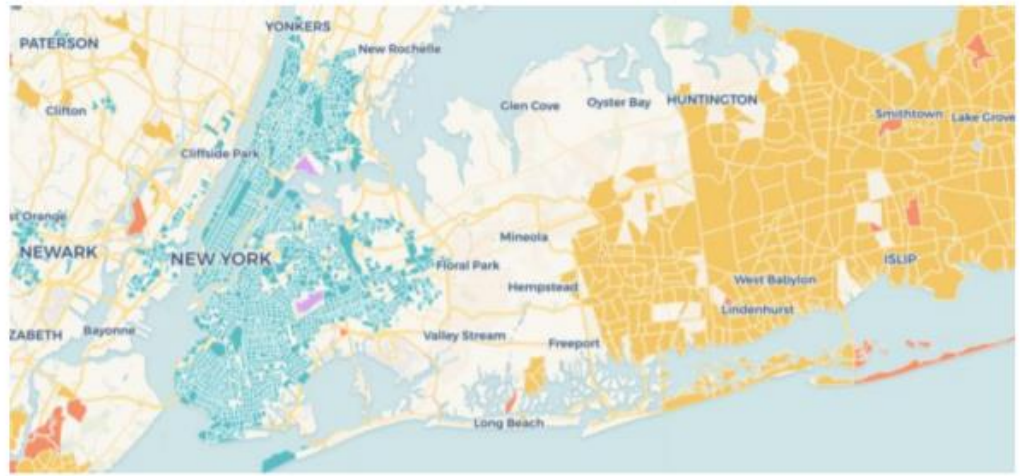
The impact can also be non stationary, it can change from place to place

CARTO

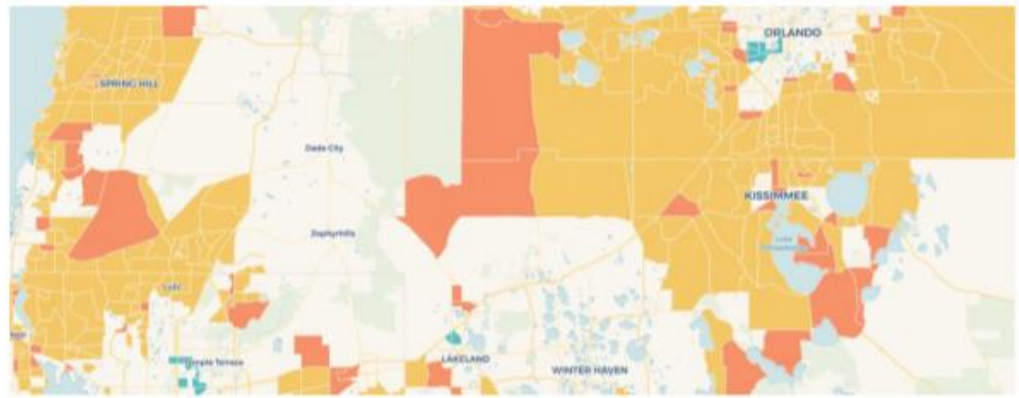

BROOMFIELD
Thornton
Westminster
Commerce City
DENVER
AURORA
SHERIDAN
Littleton
CENTENNIAL
Parker


Emeryville
Piedmont
OAKLAND
SAN FRANCISCO
ALAMEDA


BURBANK
Altadena
GLENDALE
PASADENA
ARCADIA
WEST HOLLYWOOD
SAN GABRIEL
ALHAMBRA
BEVERLY HILLS
ROSEMEAD
EL MONTE
LOS ANGELES
MONTEREY PARK
South El Monte
MONTEBELLO
PICO RIVERA
WHITTIER
INGLEWOOD
Bell Gardens
DOWNEY
LYNWOOD
HAWTHORNE
NORWALK
LA MIRADA
MANHATTAN BEACH
GARDENA
COMPTON
BELLFLOWER
Hermosa Beach
CERRITOS

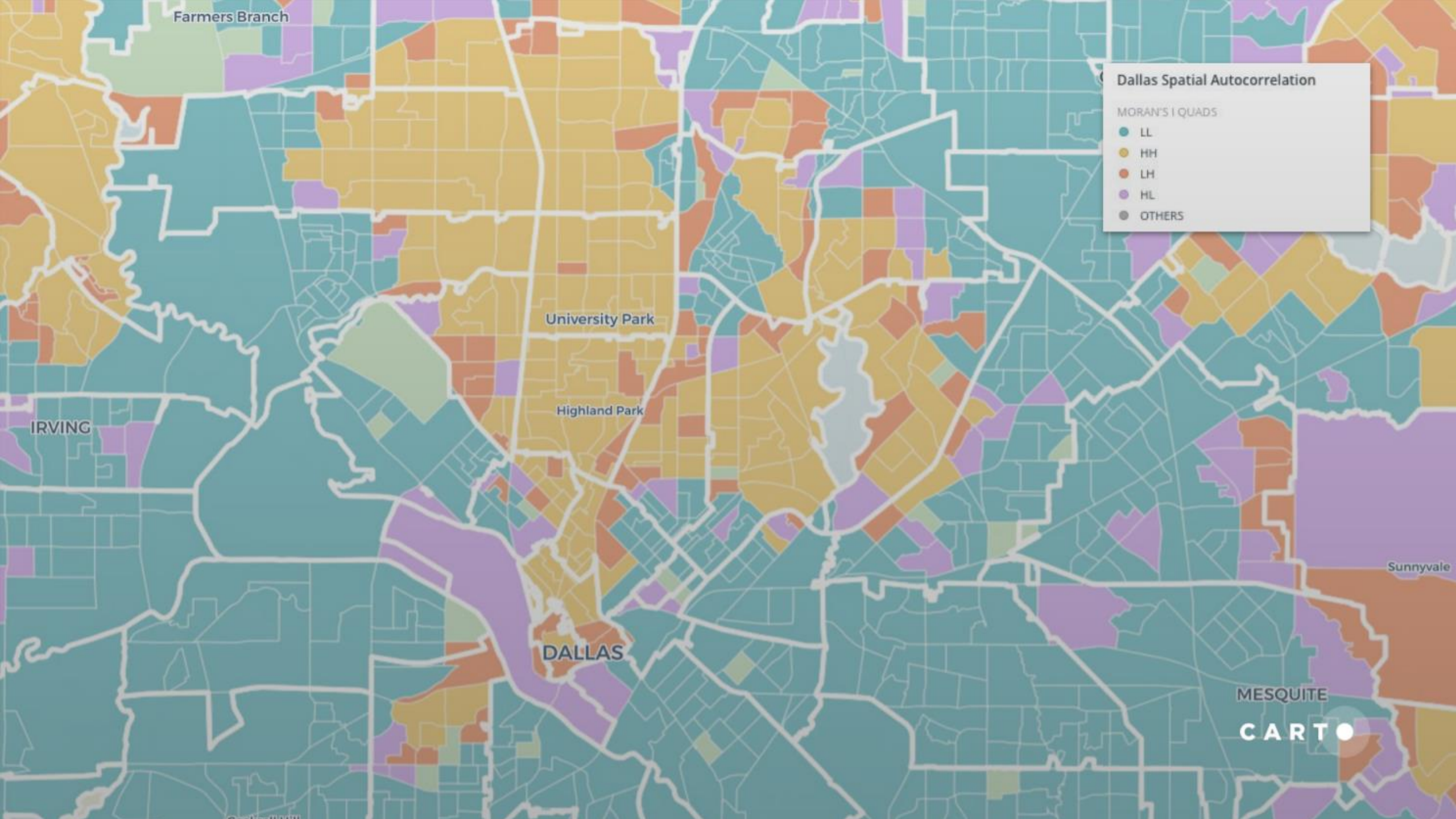

PATERSON
YONKERS
New Rochelle
Clifton
Glen Cove
Oyster Bay
HUNTINGTON
Smithtown
Lake Grove
Cliffside Park
NEWARK
NEW YORK
Mineola
Floral Park
Hempstead
West Babylon
ISLIP
Lindenhurst
Bayonne
Valley Stream
Freeport
Long Beach

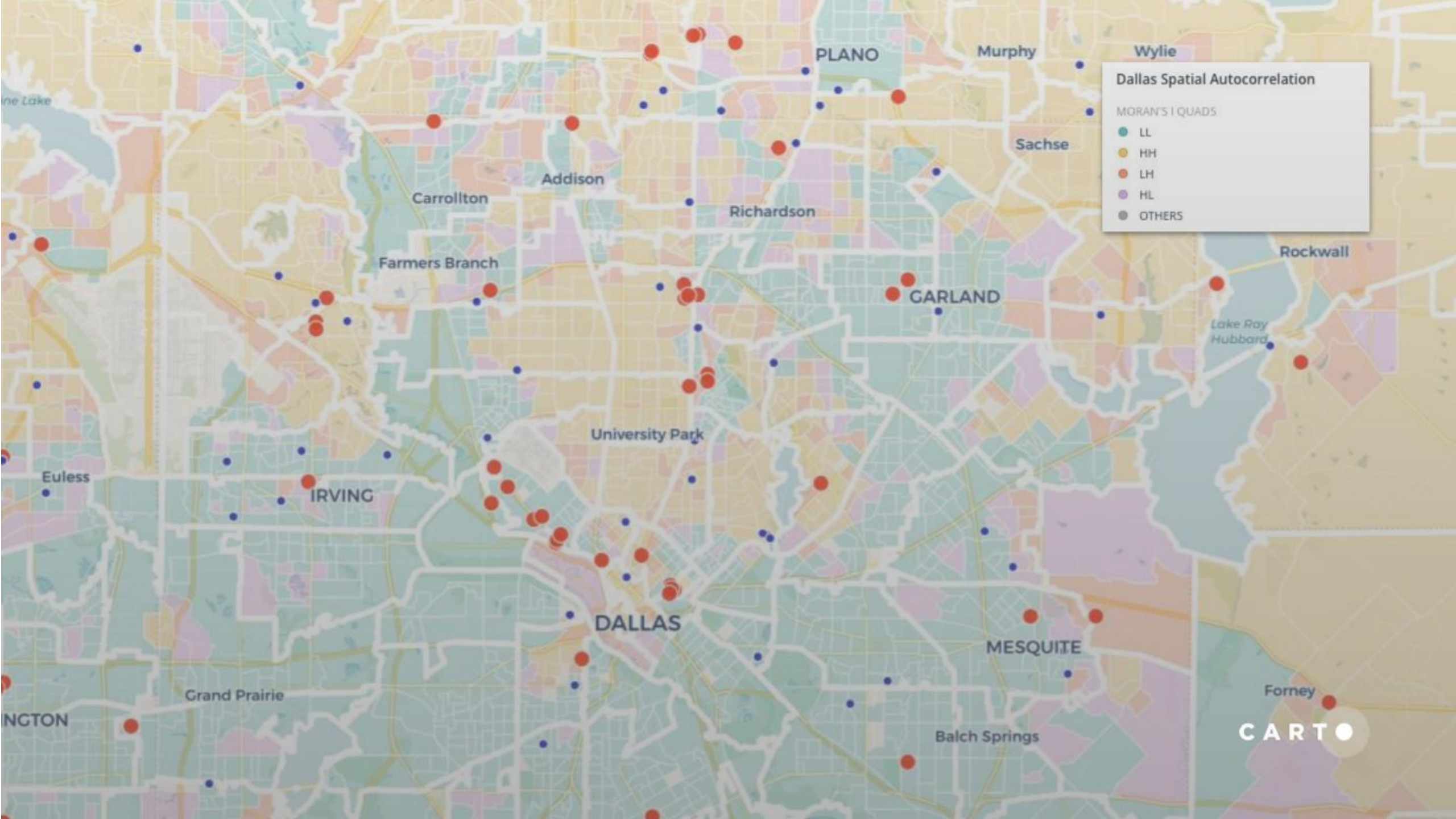

Redland
Olney
Laurel
Potomac
SILVER SPRING
COLLEGE PARK
Chillum
McLean
Landover
Bowie
WASHINGTON D.C.
Kettering
Suitland
Annandale
ALEXANDRIA
Camp Springs
Rosaryville
Fort Washington
Accokeek
Waldorf


ORLANDO
SPRING HILL
Dade City
Zephyrhills
KISSIMMEE
LAKELAND
WINTER HAVEN


EVANSTON
CHICAGO
Cicero
Burbank
Oak Lawn


NORTH LAS VEGAS
LAS VEGAS
Winchester
Paradise
CARTO


Farmers Branch
University Park
Highland Park
IRVING
DALLAS
Sunnyvale
MESQUITE
Dallas Spatial Autocorrelation
MORAN'S I QUADS
LL
HH
LH
HL
OTHERS
CARTO


Dallas Spatial Autocorrelation
MORAN'S I QUADS
LL
HH
LH
HL
OTHERS
PLANO
Murphy
Wylie
Sachse
Addison
Carrollton
Richardson
Farmers Branch
GARLAND
Rockwall
Lake Ray Hubbard
University Park
Euless
IRVING
DALLAS
MESQUITE
Grand Prairie
Forney
Balch Springs
CARTO

How can we reflect this diversity in longevity models?

Steven Baxter
Head of Innovation & Development
Club Vita

Grouping ZIP codes

Marketing principles:

1. People living in similar neighborhoods have similar characteristics *(large diversity within ZIP codes mean ZIP+4 is necessary)*

2. Neighborhoods can be characterized by types of people living there

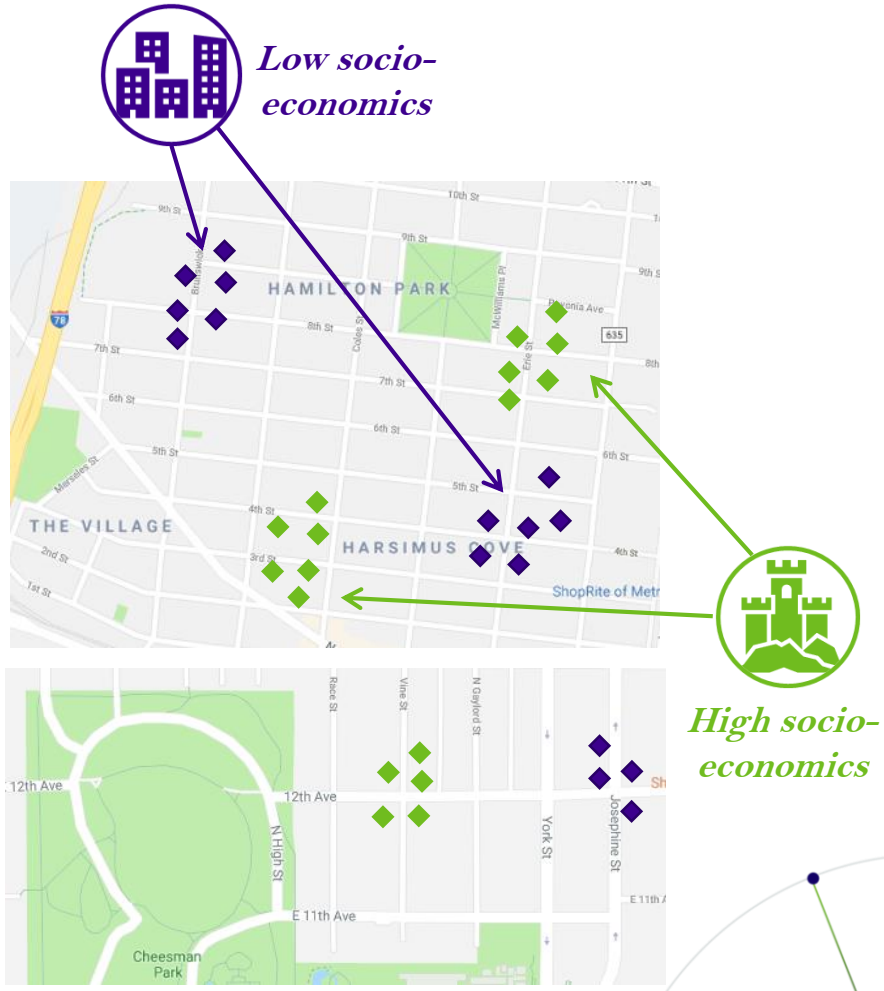
3. Neighborhoods with same characterization appear all over the country

46 million + US ZIP+4 codes => 58 marketing groups

Longevity modeling principle:

4. People with similar characteristics have similar longevity

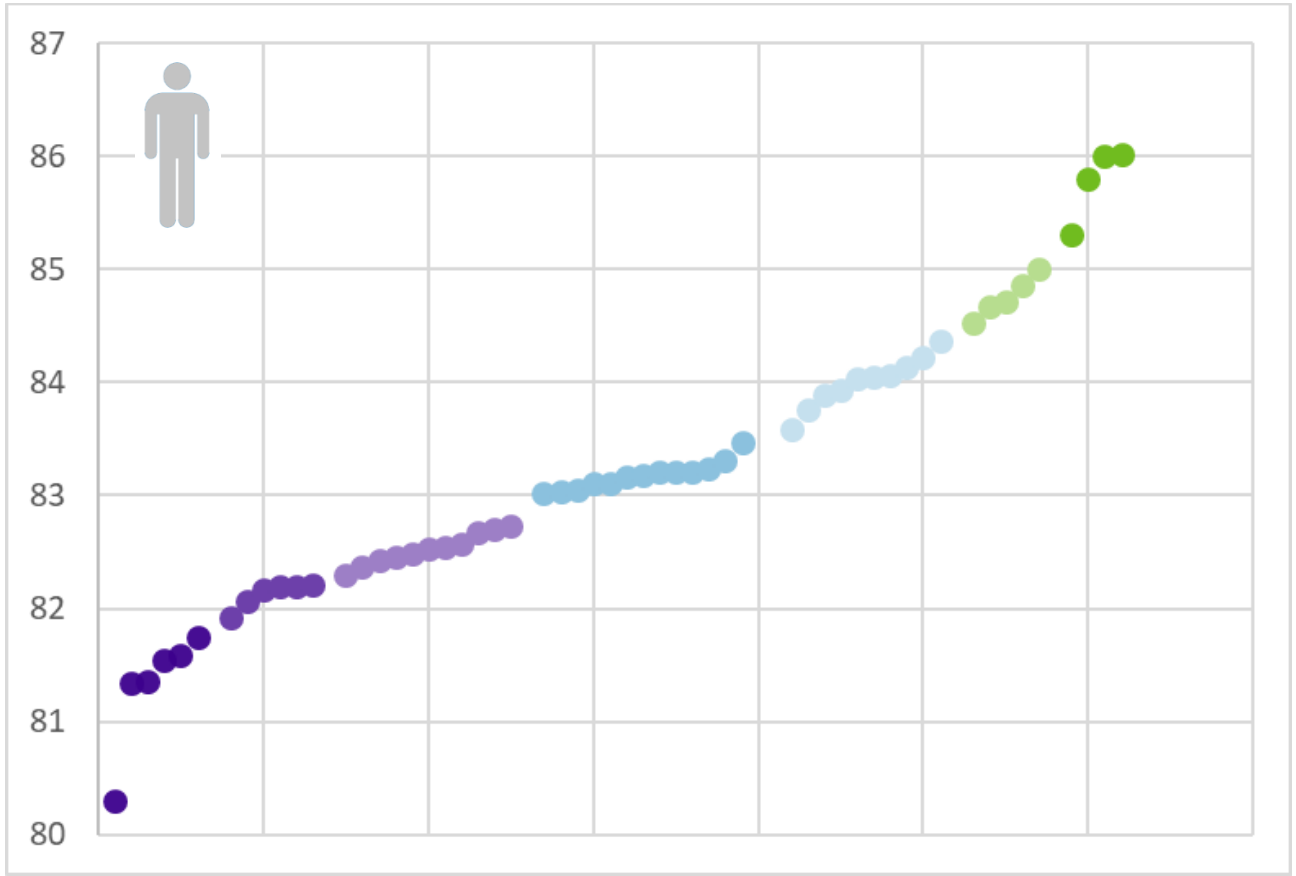
58 marketing groups => 7 longevity groups men (6 for women)



*Neighborhood characteristics for illustration only

Clustering of Zip+4 groups

Crude life expectancy at age 65, men,
split by marketing groups, colored by longevity group

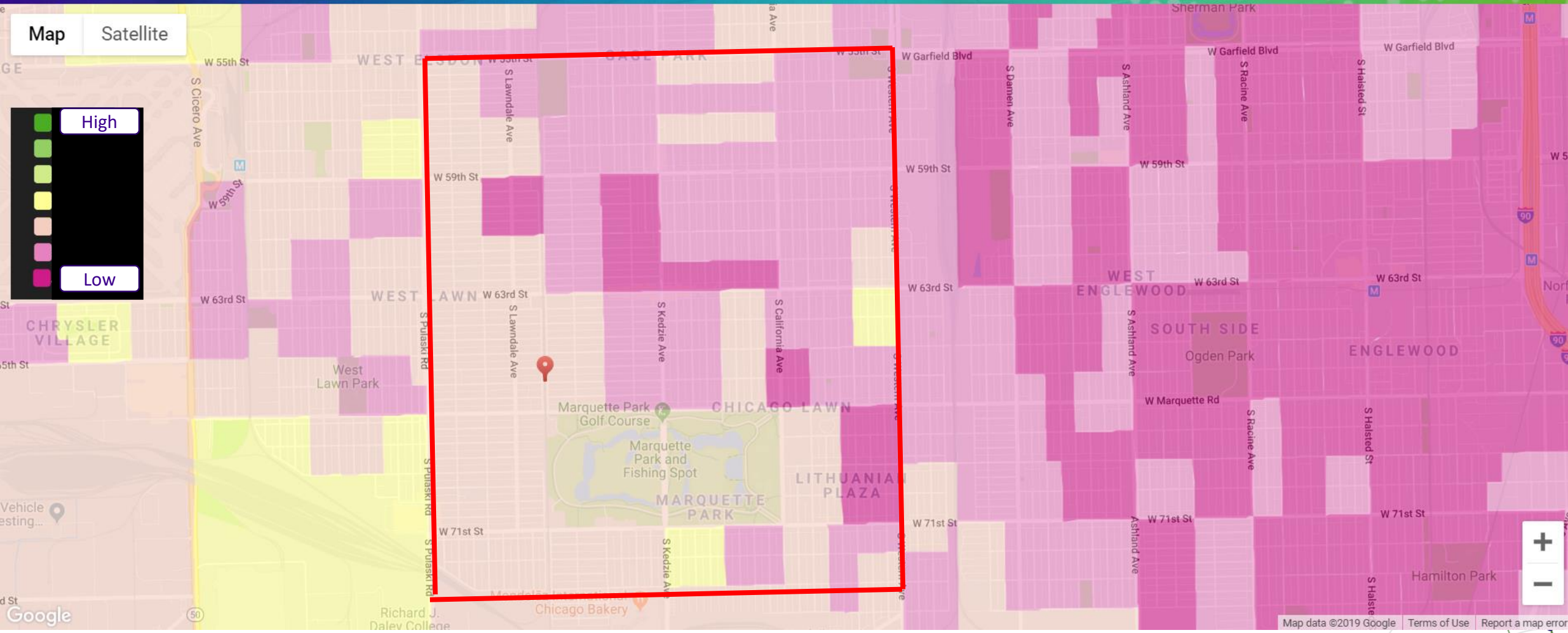


Algorithm gives 7 distinct groups for men, 6 for women

9 digit ZIP code based life expectancy at age 65

Chicago, IL 60629, USA

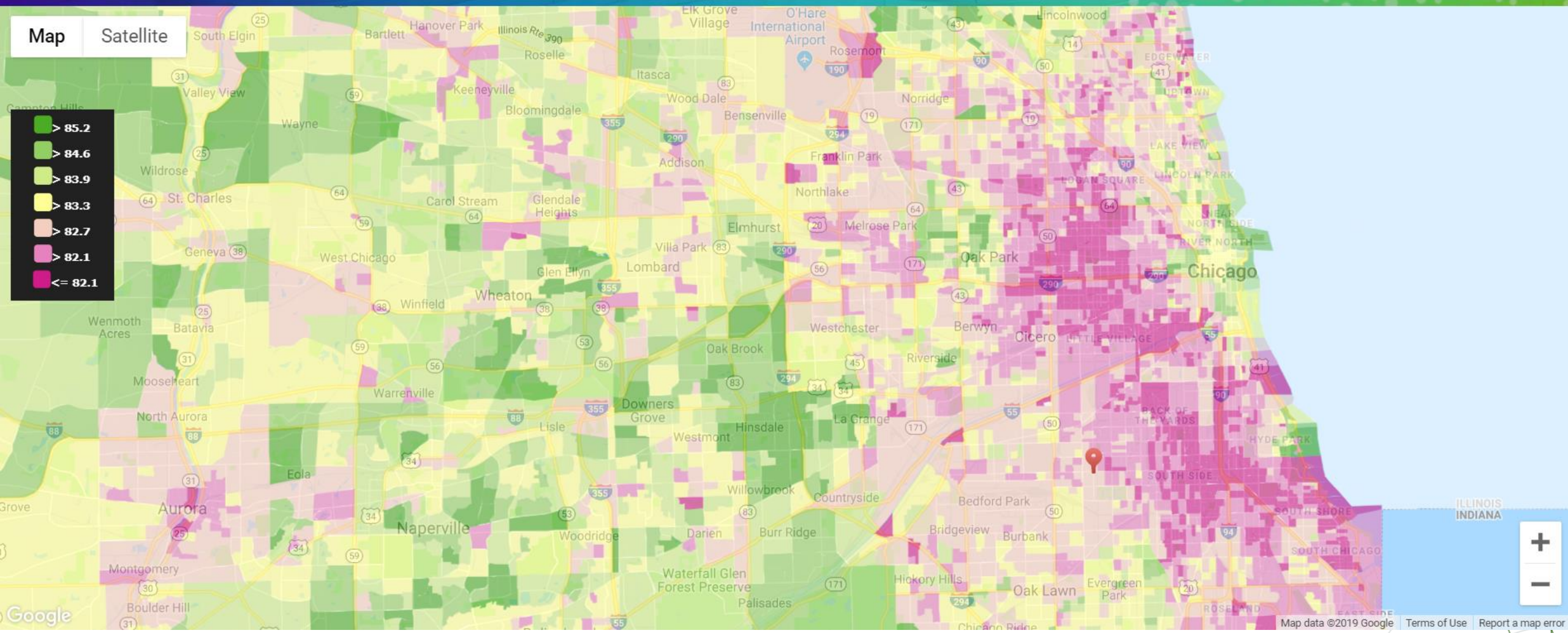
Zooming in on 60629



9 digit ZIP code based life expectancy at age 65

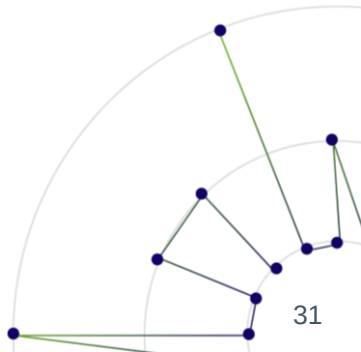
Chicago, IL 60629, USA

Zooming out…



1 in 3 of all 5-digit ZIP codes…

…contain **6** or more of our longevity groups



VitaCurves baseline model

ZIP+4

Pension

Collar

Disability

Gender

More in next gen model

"GLM" used to allow for the effect of factors *simultaneously*

What is “Generalised Linear Modelling”?

Fitting a wide range of internally consistent tables **simultaneously** across a range of variables

1 Split the data

Gender

Disability

2 Fit the data across lots of variables

ZIP+4

Annuity

Collar

Makes maximum use of the available data, improving confidence in the resulting tables while creating a model that captures the diversity of the underlying population.

Building a model for longevity

The predictors j are the longevity group (A to G as determined by ZIP+4), annuity amount and collar type

Main effect for each predictor: Additions depending on the value taken by each predictor j (can be negative)

Controls for mortality rate variations between calendar years, and is 0 for central year

$$logit(q_x|values\ of\ predictors, j) = \sum_i a_i x^i + \sum_j b_j + \sum_{i,j} c_{ij} x^{-i} + \text{YOE}$$

$$logit(q) = \ln\left(\frac{q}{1-q}\right)$$

Main age function: A polynomial in age, x, with a small number of terms (typically 3 or 4) where i takes values in range $[-4, -3, \ldots, 3, 4]$

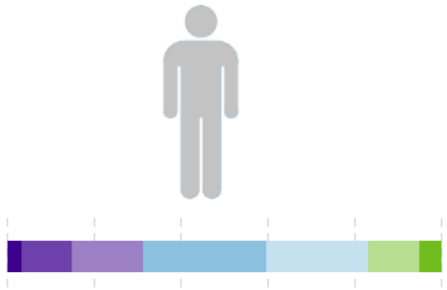
"Interaction" terms, whereby there is a small number of terms of the polynomial in age, x, which depend on the value taken by the predictor

Parsimony principle: A simpler model with few rather than many parameters is favored over comparatively complex ones, provided they fit the data about equally well.

For more information see our Modelling Technical Paper.

VitaCurves baseline model

ZIP+4

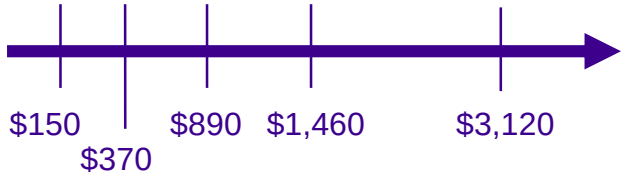


7 longevity groups
A - G

6 longevity groups
A - F

Pension

6 pension bands



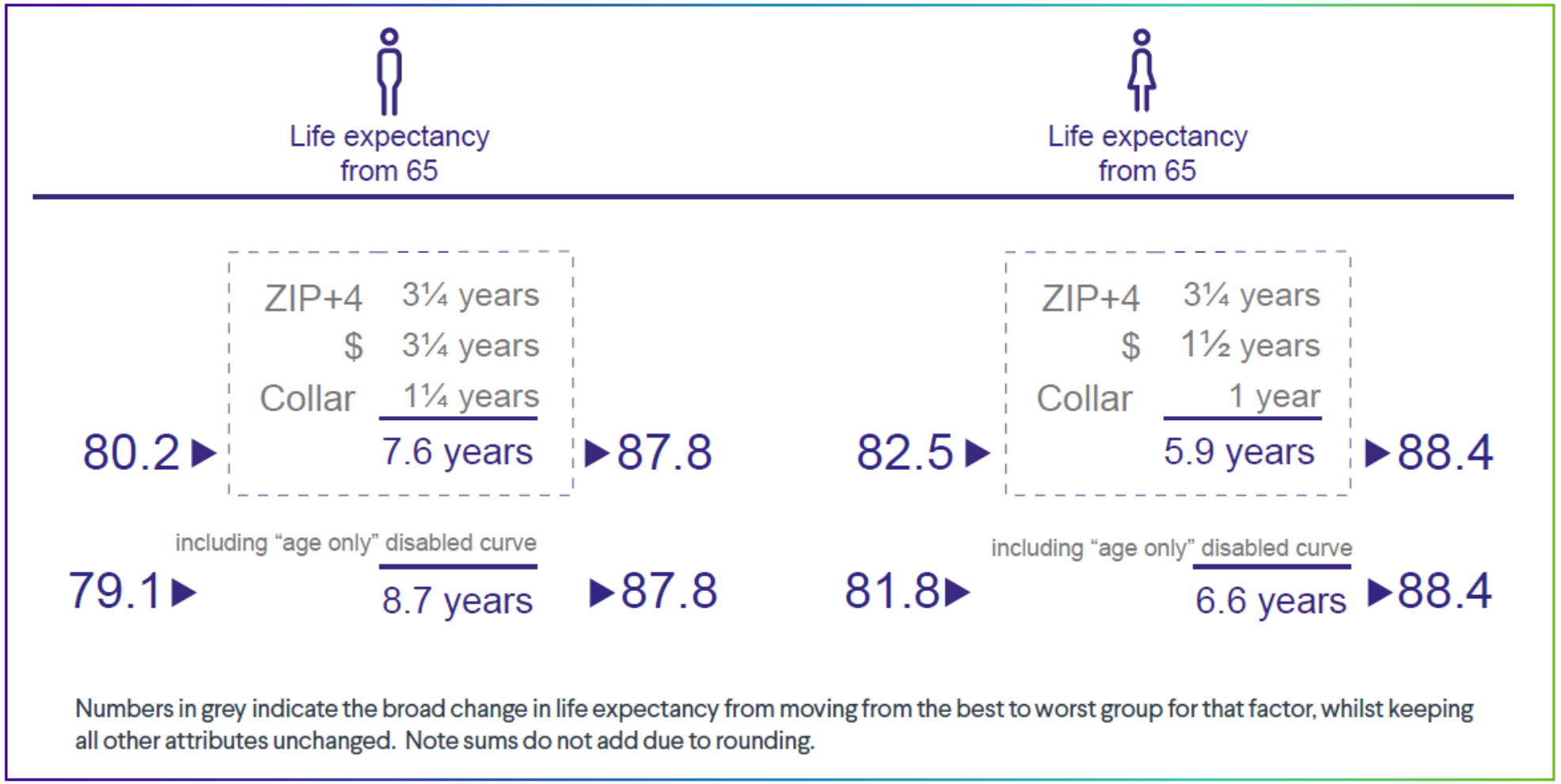
3 pension bands

Collar

2 collar groups
Blue/white

2 collar groups
Blue/white

VitaCurves baseline model



How do ZIP code techniques affect pension advice / the role of the actuary?

Bruce Cadenhead
Global Chief Actuary - Wealth
Mercer

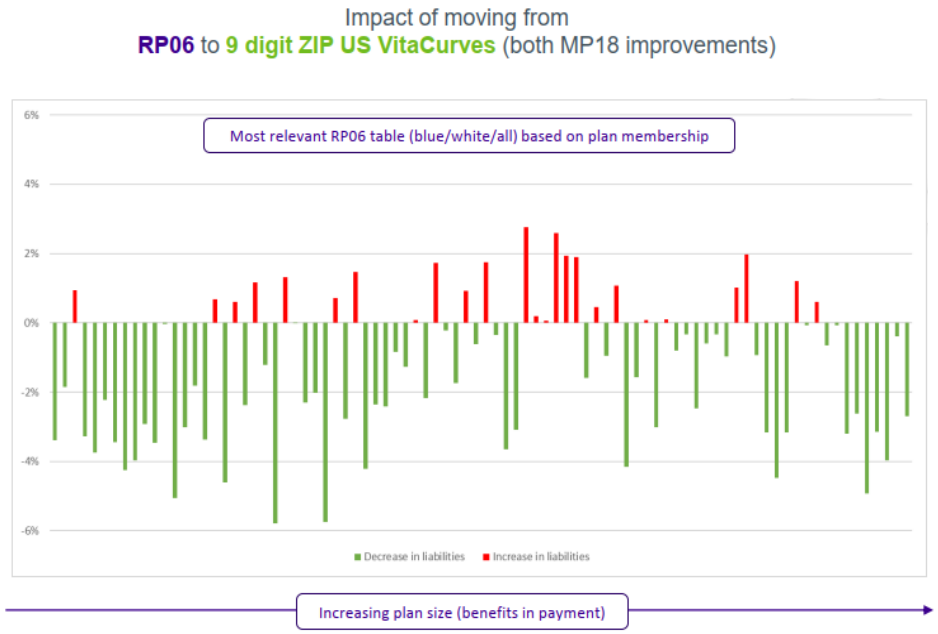
ZIP CODE MODELING TECHNIQUES
EXPANDING PENSION CONSULTING

ZIP-level modeling offers several advantages that expand the role of the actuary and the pension advice that can be provided:

- ✓ Rating factor modeling allows for **custom mortality adjustments** to apply to a wider range of plans
- ✓ An **adaptive approach to longevity** that captures the demographic characteristics of each plan and anticipates longevity changes that result from a changing workforce
- ✓ A better understanding of **pension risk transfer opportunities** and the impact of lump sums and annuity buyouts on funded status
- ✓ More frequent model updates **reduces liability volatility** associated with sporadically revised mortality standards
- ✓ **Potential improvement in pension balance sheet and P&L results** vs. standard tables

FINANCIAL STATEMENT IMPLICATIONS
POTENTIAL LIABILITY OVERVALUATION

Initial analysis shows average reduction in liabilities of over 1%



CLUB VITA

- Club Vita's initial analysis focused on data from a diverse collection of 108 large pension plans, across a range of different industries and locations
- This analysis revealed that pension plan liabilities may be either understated or overstated using the standard Society of Actuaries tables. Individual plan results varied as much as 6%, but with an **overall average reduction of over 1%* of liabilities.**
- Hidden "overvaluation" may be attributable to existing tables not staying as up-to-date – current tables are centered based on 2006 data (Pri-2012 study to update to 2012 centering), whereas VitaCurves model is based on data centered in 2015.

*Estimated plan liabilities using both VitaCurves and the Society of Actuaries RP-2006 base tables. Assumed MP-2018 mortality projection scale for future improvements, 4% discount rate, and excluded any surviving spouses' pensions that may commence after retirees die in the future.

Questions?

Douglas Anderson
Founder
Club Vita

Bruce Cadenhead
Global Chief Actuary - Wealth
Mercer

Matt Forrest
Director of Spatial Data Science
Carto

Steven Baxter
Head of Innovation & Development
Club Vita

Join the discussion…
LinkedIn: Friends of Club Vita group
Twitter: @ClubVita #longevityZIP

How similar is the US model to UK and Canada?

						
Total spread	12 years		10 years		9½ years	
Gender specific spread	10½	8½	7	7½	8½	6½
Retirement health	2½	3	½	2	1	1
"Normal health" spread	8	5½	6½	6	7½	6
Lifestyle	4¼	4½	2¾	3½	3¼	3½
Affluence	3½	½	2	2½	3¼	1½
Occupation	¼	½	1¼	<¼	1¼	1

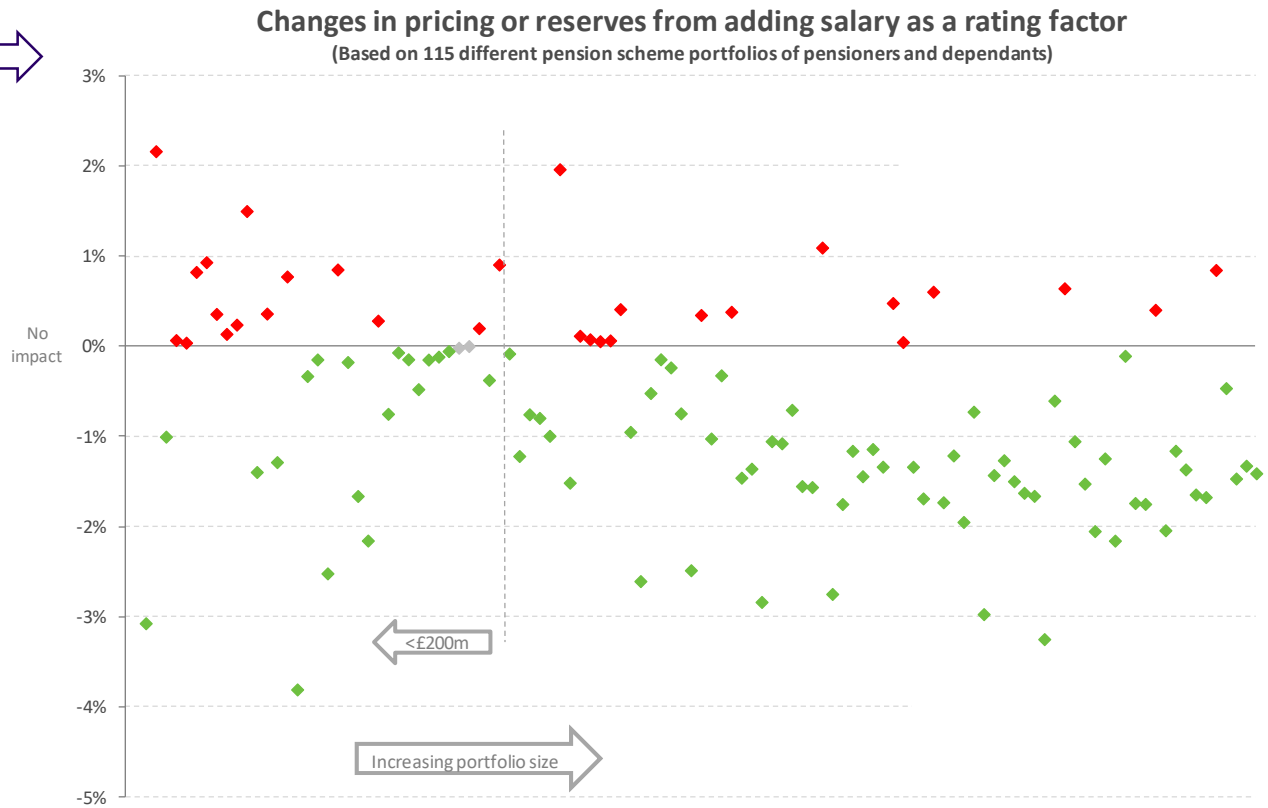
CLUB VITA

Effects shown are the impact of changing one rating factor in isolation. Precise impacts depend on order of changing variables and so above reflects broad quantum and therefore relative importance of each variable. Sums may not add due to rounding.

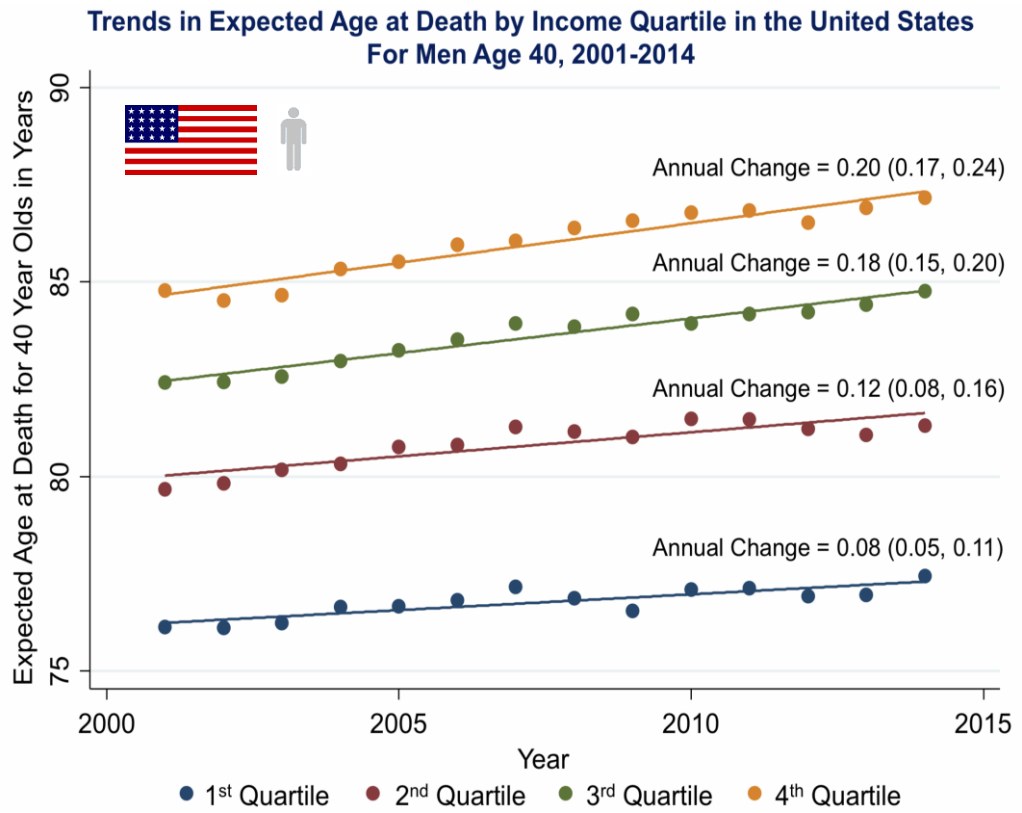
What next for US VitaCurves?

More in next gen model

1. Salaries
2. Exercised joint life option
3. Industry?
4. ?



Are the trends the same across socio-economic groups?



Evidence of “eras” for improvements

Higher improvements for “wealthiest/healthiest”

Club Vita needs your help!

Dan Reddy
CEO
Club Vita US
daniel.reddy@clubvita.net

Erik Pickett
CPO
Club Vita US
erik.pickett@clubvita.net

Thank you