Health Opportunity Index Methodology

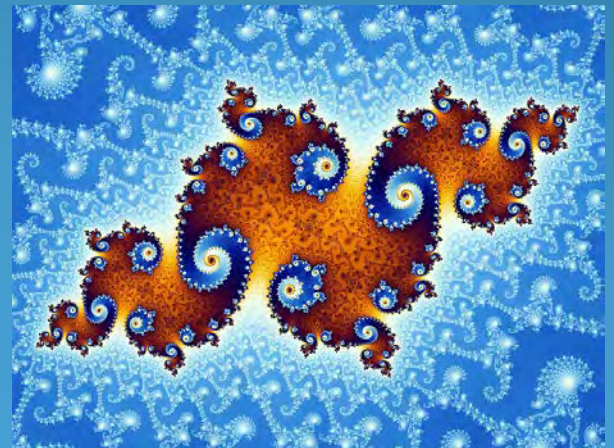**The Health Opportunity Index; Version 2 (2015)—A Methodological, Analytical and Policy Perspective**

Rexford Anson-Dwamena, MPH
Social Epidemiologist & GIS Analyst

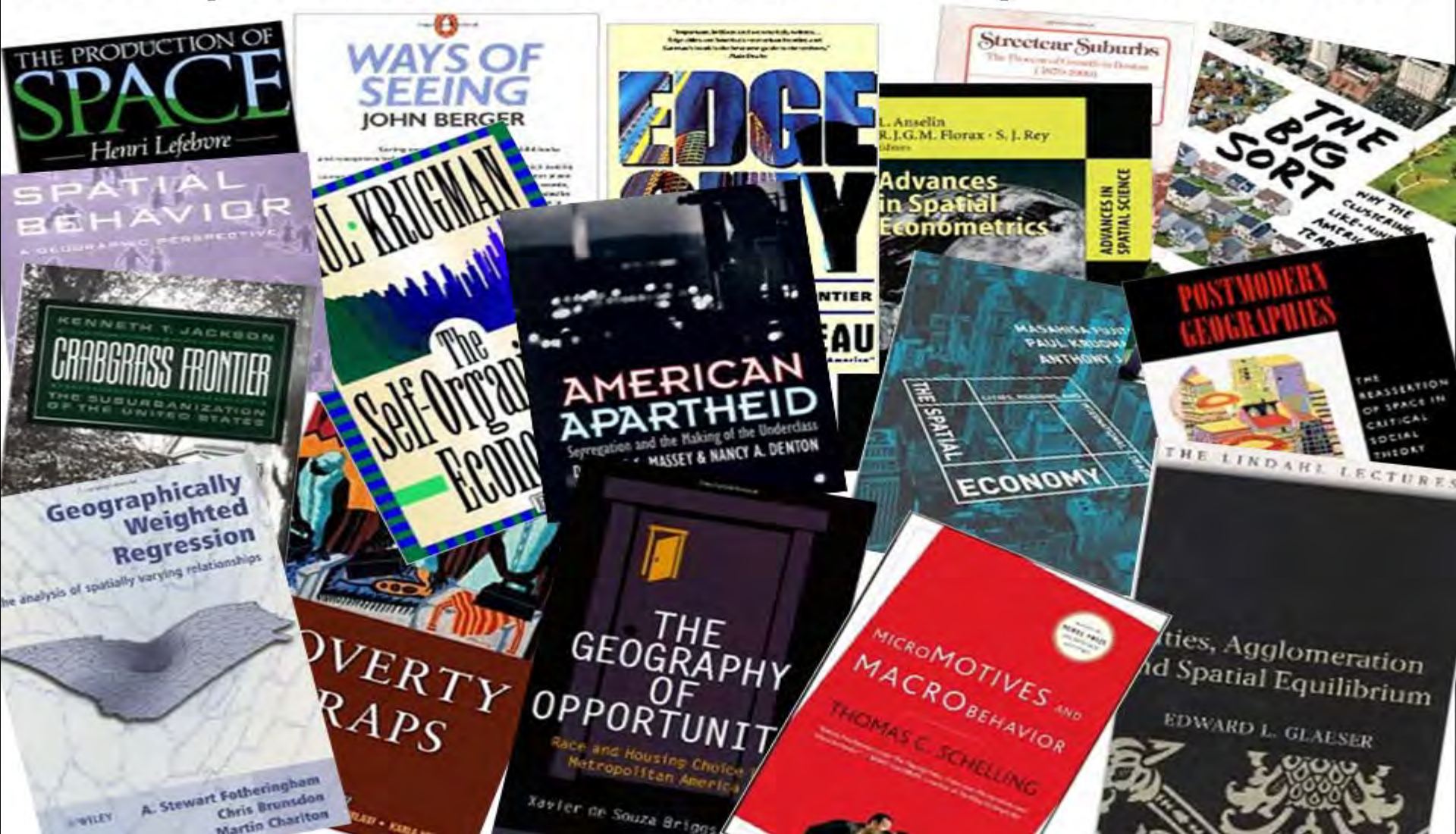Justin Crow, MPA
Designation & GIS Manager

Anna Riggan, MPH
Research Analyst,

Wednesday, 16th, 2015

Office of Minority Health & Health Equity
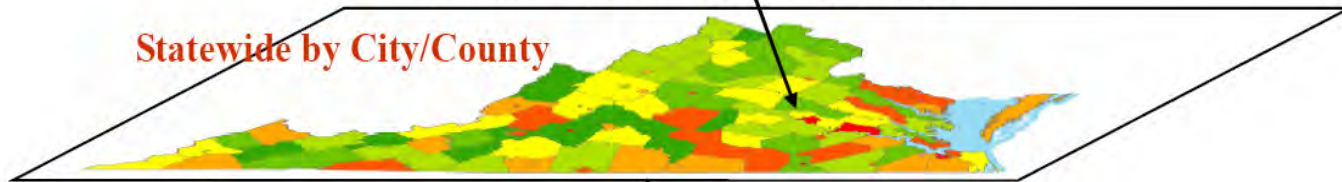*Advancing Health Equity For All Virginians*

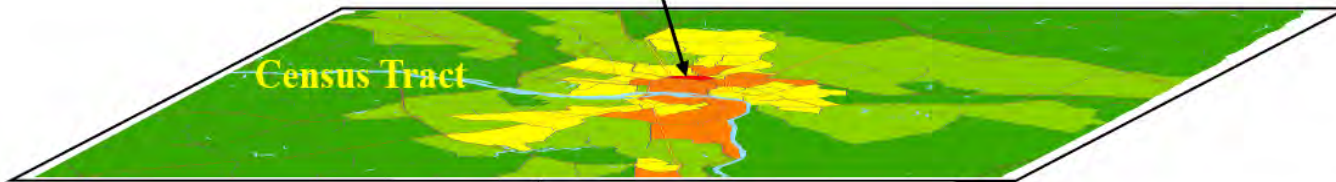…it is space not time that hides consequences from us.

Multilevel Spatial Analysis of Fundamental Causes
&
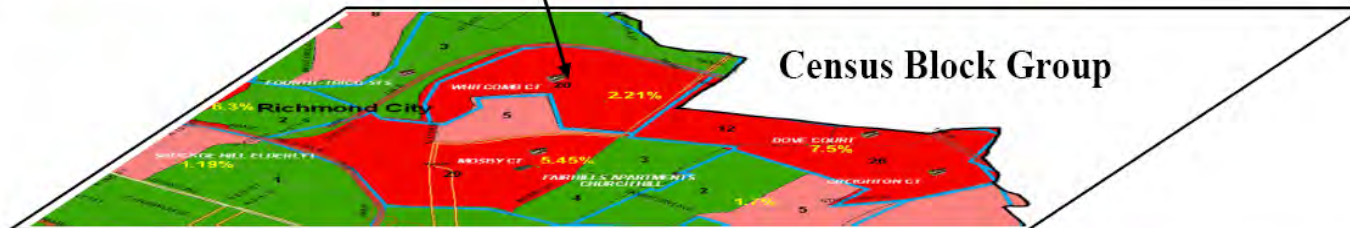the Social Determinants of Health

Statewide by City/County

Census Tract

Census Block Group

Social Profiles, Social Networks & Social Capital
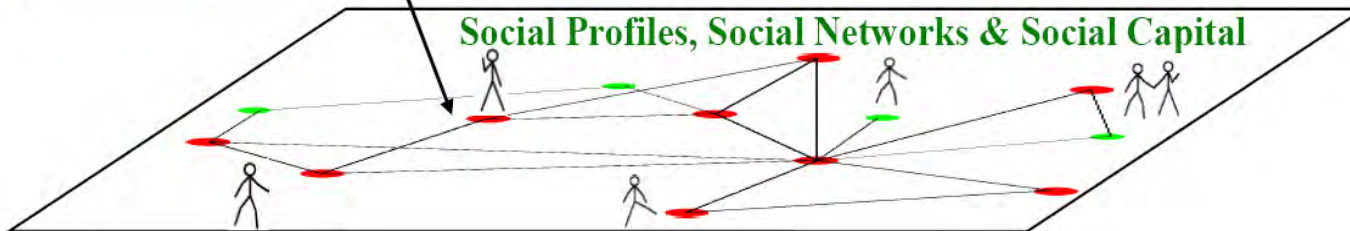
The Methodology

Process Variables (Actionable)

Structure

30+ Variables

13 Indicators

4 Clusters

1 Health Opportunity Index
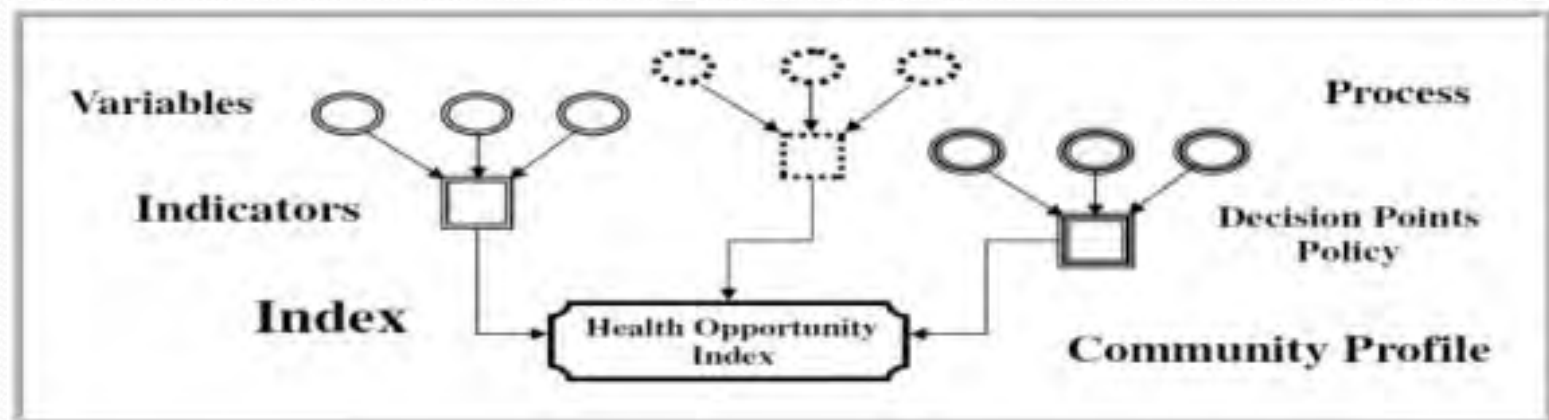
Affordability Index

- The affordability index is developed to measure the proportion of income spent on housing and transportation. The index of affordability is calculated by combining housing and transportation costs in a neighborhood and dividing that number by income.

$$\text{Affordability Index} = \frac{\text{Housing Cost} + \text{Transportation Cost}}{\text{Total Income}}$$

Income Inequality (GINI) Index

- the GINI coefficient (also known as the index of income concentration).
- Measures inequality of income.
- Measures how homogeneous or diversity of actual earned income by neighborhood (0 is homogeneous and 10 means more diversity in income earning positions.

$$(1) \qquad GINI = \frac{1}{\mu N (N-1)} \sum_{i>j} \sum_{j} \left| y_i - y_j \right|$$

Where,
GINI is the Gini index;
μ is the mean of the variable (income, e.g.);
N is the total number of observations;
For income Gini, y_i and y_j are dollar values of income of individuals;

Townsend Dep. Index *

- Townsend deprivation index is a measure of material deprivation. According to Townsend, "Material deprivation entails the lack of goods, services, resources, amenities and physical environment which are customary, or at least widely approved in the society under consideration.

1. *Unemployment* – % of economically active residents aged 16-59/64 who are unemployed.

2. *Car ownership* – % of private households who do not posses a car.

3. *Home ownership* – % of private households not owner occupied.

4. *Overcrowding* – % of private households with more than one person per room.

* The four variables were combined into one indicator using principal component analysis (using the natural log of the variables)

Job Participation Index

- Job Participation Rate is the percentage of individuals 16-64 years of age in the active labor force. The job participation rate is often used by economics as an indicator for economic development and growth.

$$\text{Job Participation rate} = \frac{\text{Civilian Employed} + \text{Civilian Unemployed}}{\text{Civilian Population (16-64 yrs)}}$$

Employment Access Index

- Poor job access leads to difficulties in job search or job retention and, consequently, to poverty and socioeconomic disadvantages

- Employment accessibility index: you may have a workforce but how accessible are they to the potential jobs --- how far are you (distance) from a potential job. In other words, the index is based on jobs and distance decay function

- Ownership of a vehicle plays a function

$$A_i = \sum_{j=1}^{n} \frac{J_j d_{ij}^{-\beta}}{V_j}, \text{ where } V_j = \sum_{k=1}^{m} W_k d_{kj}^{-\beta}$$

A_i is the job accessibility at location i,

J_j is the number of jobs in location j,

d_{ij} is now the travel time between them,

β is the friction coefficient and n is the total number of job locations.

V_j is measured by this job location's proximity to all workers

Education Index

- The majority of the empirical literature uses average years of education as a proxy of the human capital stock

- higher level of human capital stock leads to faster technological development and ultimately higher growth rates

- Preschool through doctorate (this index is weighted based upon how far you have advanced in education)

- Higher the number the higher average number of schooling in the area

The average years of schooling (AYS) can be calculated as follows.

$$\mu = AYS = \sum_{i=1}^{n} p_i y_i$$

μ is the average years of schooling for the concerned population;

p_i and p_j stand for the proportions of population with certain levels of schooling;

y_i and y_j are the years of schooling at different education attainment levels;

Years of Schooling Assigned for Data Beginning in 1992[6]

Educational Attainment Level	Years of Schooling, y
Elementary School	
None	0
1 to 4 years	2.5
5 to 6 years	5.5
7 to 8 years	7.5
High School	
9th Grade	9
10th Grade	10
11th Grade	11
Graduate	12
College	
Some College, No Degree	13
Associate's Degree, Occupational	14
Associate's Degree, Academic	14
Bachelor's Degree	16
Master's Degree	18
Professional Degree	19
Doctoral Degree	20

Population Churning Index

- Population churning rates relate the combined inflow and outflow for an area to the resident population. The rates can provide a useful measure of the potential disruption to local services caused by migration into and out of the Census tract.

$$\text{Population Churning Rate} = \frac{\text{In Migration} + \text{out Migration}}{\text{Total population}}$$

Population-Weighted Density Index

- Weighted density is to capture the density at which the average person lives (aka Dysmetric).

- It is necessary to differential between urban and rural in the model

FIPS	CT	POP2013	SQMI	POP13_SQMI	BG_PopWt	BG_PopWtedSQIM
510010901001	51001090100	901	3.75	240.3	0.311333794	74.81351071
510010901002	51001090100	790	0.75	1053.3	0.272978576	287.5283345
510010901003	51001090100	599	1.11	539.6	0.206979959	111.6863856
510010901004	51001090100	604	1.77	341.2	0.208707671	71.21105736
	Std	2894	7.38		PopWeighted	545.2

Std Pop-Density - 392.1 per SQMI

Segregation Index (Spatial Dissimilarity)

- Calculates a segregation or dissimilarity index that measures how different the population composition of a subarea (CT) is from the population composition of the State as a whole
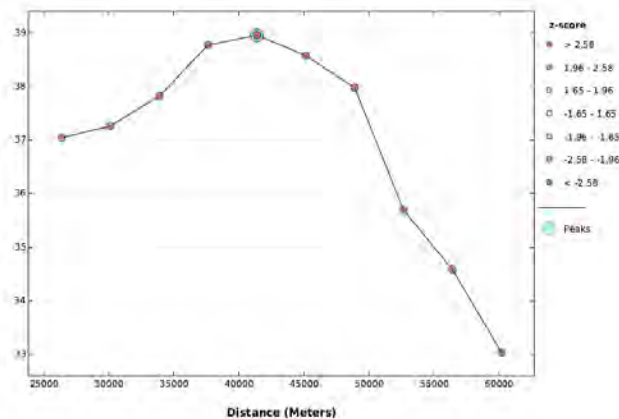
Spatial Dissimilarity Index (SD)

$$SD = (1/2)(\sum_{i=1}^{n}\sum_{j=1}^{m}|CN_{ij} - CE_{ij}| / \sum_{j=1}^{m} CN \times CP_j(1-CP_j) \text{ with}$$

$$CE_{ij} = (CN_{i.} - CN_{j})/CN \text{ and } CN_{ij} = \sum_{k=1}^{n} d(N_{kj})$$

CN_{ij} : Composite population count of ethnic group j in spatial unit i

$d()$: Function defining surrounding spatial units i and k

CN_i : Total composite population count in spatial unit i

CN_j : Total composite population count of ethnic group j

CN : Total population in the city

CP_i : Proportion of population in ethnic group j
     M

Food Accessibility Index

- Limited access to supermarkets, supercenters, grocery stores, or other sources of healthy and affordable food may make it harder for some Americans to eat a healthy diet

Low access was measured as living far from a supermarket, where 1 mile was used in urban areas and 10 miles was used in rural areas to demarcate those who are far from a supermarket.

The index was calculated based on the share (proportion of the target population) of the population that fit the USDA criteria rather than using the binary indicator.

Walkability Index (4Ds)

- Methodology was adopted from International Physical Activity and the Environment Network (IPEN)

$$WAI = (2 * con) + ent + far + hdens$$

* Design of the built environment *Land use diversity* *Distance to transit* Residential and employment *density*

- The 4Ds include concepts such as:
- *Density – Residential and employment*
  - Indicator: Total activity units per acre of land
  - Measures the concentration of activity types within a walkable area
- *Diversity – Land use and destinations*
  - Indicator: Range of land uses by census tract
  - Measures the mix of activities available within a walkable area
- *Design – Built environment and safety features*
  - Indicator: Number of street crossings by census tract
  - Measures the degree of connectivity to support safe pedestrian travel
- *Distance – Transit accessibility*
  - Indicator: Aggregate frequency of transit service per square mile
  - Measures level of accessibility for pedestrians to reach a transit stop

* Variable was weighted twice

Access to Care (2SFCA)

- It first assesses "physician availability" at the physicians' (supply) locations as the ratio of physicians to their surrounding population (i.e., within a threshold travel time from the physicians)
- It sums up the ratios (i.e., physician availability derived in the first step) around (i.e., within the same threshold travel time from) each residential (demand) location. Based on Population-weighted centriod
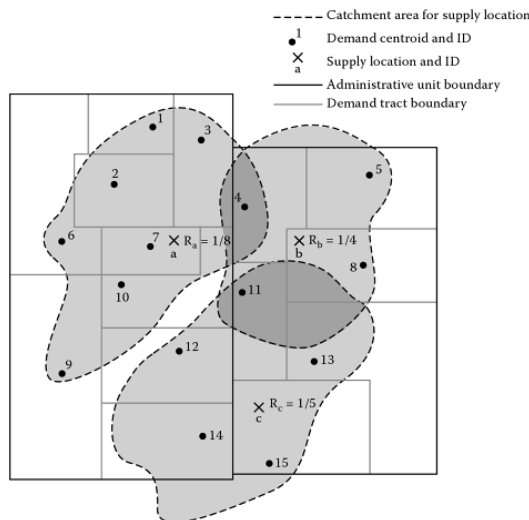

FIGURE 5.2 The 2SFCA method.

% of Healthcare Uninsured Population was also combined with the FTEs to form the indicator

Environmental Quality Index (EPA) *

- **National-Scale Air Toxics Assessment** (NATA) provides estimates of the risk of cancer and other serious health effects from breathing (inhaling) air toxics in order to inform both national and more localized efforts to identify and prioritize air toxics, emission source types and locations which are of greatest potential concern in terms of contributing to population risk

Indicator	Details
Neurological Risk	provides the estimated neurological risk (hazard quotient) for each census tract, compiled by sector and contributing chemical.
Cancer Risk	The probability of contracting cancer over the course of a lifetime
Respiration Risk	Air toxics respiratory hazard index
On-road Pollution	Vehicles found on roads and highways (e.g., cars, trucks, buses).
Non-road Pollution	Mobile sources not found on roads and highways (e.g., airplanes, trains, lawn mowers, construction vehicles, farm machinery)
Non-Point Pollution	NPS pollution is caused by rainfall or snowmelt moving over and through the ground. As the runoff moves, it picks up and carries away natural and human-made pollutants, finally depositing them into lakes, rivers, wetlands, coastal waters and ground waters

* The six databases were transformed into one indicator using principal component analysis

Data Reduction Technique

Reducing the Dimensions of the Data

Correlation Matrix

		HealthCare Access	Employment Access	Affordability Index	AirQualityIndex	PopChurning	AverageYrsof Schooling	FoodAccess	Income InqualityGINI	Job Participation	Pop_Weighted dDensity	Spatial Segragation	Townsend Index	Walkability Index
Correlation	HealthCareAccess	1.000	.207	.116	-.230	-.056	.179	-.081	.060	.221	.124	-.013	.008	.194
	EmploymentAccess	.207	1.000	.191	-.112	-.108	.209	-.076	.334	.514	.199	.004	.128	.305
	AffordabilityIndex	.116	.191	1.000	.103	.206	.587	-.261	.352	.343	.089	.056	.600	.076
	AirQualityIndex	-.230	-.112	.103	1.000	.265	-.122	-.074	.224	-.132	-.285	.041	.328	-.409
	PopChurning	-.056	-.108	.206	.265	1.000	-.119	-.119	.126	-.026	-.358	.002	.368	-.391
	AverageYrsofSchooling	.179	.209	.587	-.122	-.119	1.000	-.216	.128	.494	.245	.026	.318	.314
	FoodAccess	-.081	-.076	-.261	-.074	-.119	-.216	1.000	-.114	-.152	-.032	-.042	-.183	-.049
	IncomeInqualityGINI	.060	.334	.352	.224	.126	.128	-.114	1.000	.384	.027	.058	.398	-.020
	JobParticipation	.221	.514	.343	-.132	-.026	.494	-.152	.384	1.000	.301	.046	.021	.338
	Pop_WeightedDensity	.124	.199	.089	-.285	-.358	.245	-.032	.027	.301	1.000	.041	-.251	.604
	SpatialSegragation	-.013	.004	.056	.041	.002	.026	-.042	.058	.046	.041	1.000	.042	.001
	TownsendIndex	.008	.128	.600	.328	.368	.318	-.183	.398	.021	-.251	.042	1.000	-.250
	WalkabilityIndex	.194	.305	.076	-.409	-.391	.314	-.049	-.020	.338	.604	.001	-.250	1.000

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.689
Bartlett's Test of Sphericity	Approx. Chi-Square	7171.823
	df	78
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.068	23.599	23.599
2	2.649	20.378	43.977
3	1.123	8.638	52.615
4	1.044	8.032	60.647
5	.951	7.313	67.960
6	.865	6.657	74.617
7	.707	5.437	80.054
8	.642	4.939	84.993
9	.586	4.505	89.499
10	.524	4.034	93.533
11	.362	2.781	96.313
12	.280	2.156	98.469
13	.199	1.531	100.000

Extraction Method: Principal Component Analysis.

Scree Plot
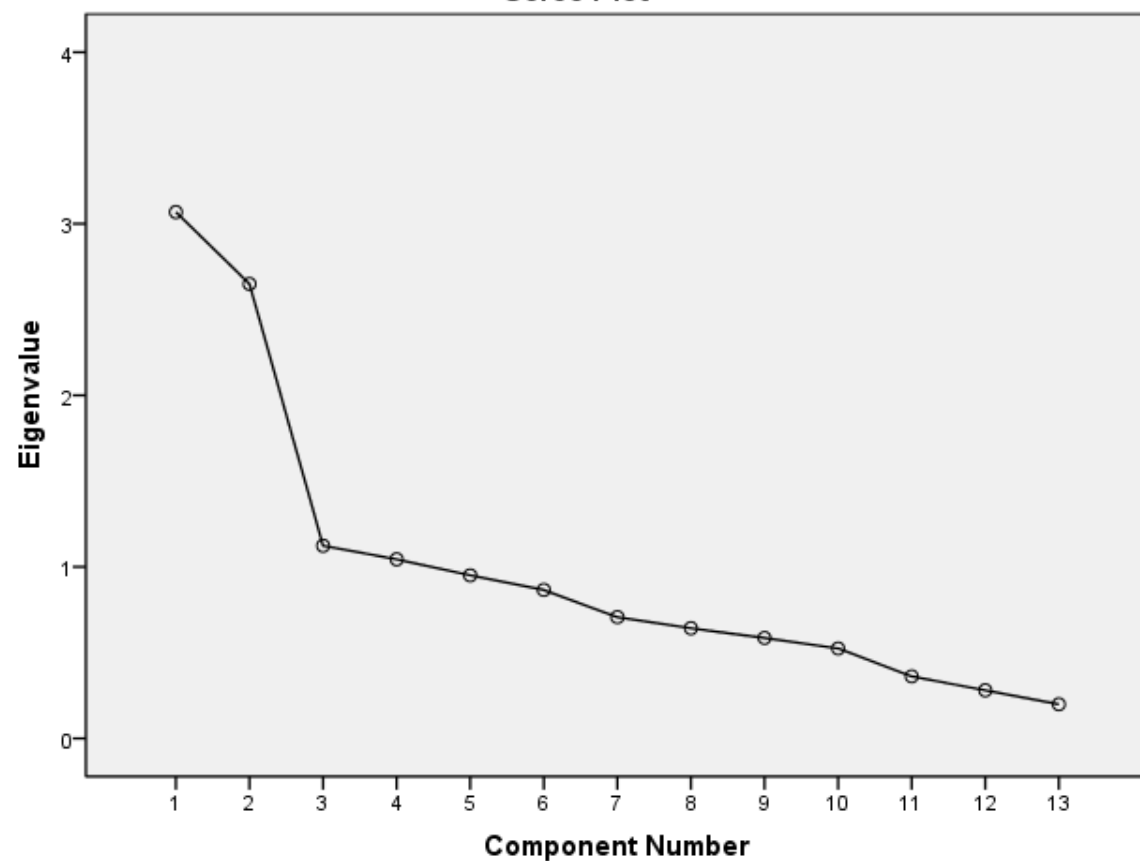
Method: Principal components

Analyze
◉ Correlation matrix
◯ Covariance matrix

Display
☐ Unrotated factor solution
☑ Scree plot

Extract
◯ Based on Eigenvalue
   Eigenvalues greater than: 1
◉ Fixed number of factors
   Factors to extract: 4

Maximum Iterations for Convergence: 25

[Continue] [Cancel] [Help]

Rotated Component Matrix[a]

	Component			
	1	2	3	4
WalkabilityIndex	.791	.142	.192	-.093
Pop_WeightedDensity	.747	.122	.175	.088
PopChurning	-.667	.191	.018	-.071
AirQualityIndex	-.601	.029	.114	.404
AffordabilityIndex	-.117	.822	.273	.010
AverageYrsofSchooling	.297	.774	.170	-.086
TownsendIndex	-.545	.599	.235	.061
FoodAccess	.021	-.554	.040	-.008
EmploymentAccess	.191	.014	.785	-.161
IncomeInqualityGINI	-.251	.142	.770	.148
JobParticipation	.304	.292	.686	-.105
SpatialSegragation	.142	.112	.057	.734
HealthCareAccess	.174	.164	.196	-.579

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.068	23.599	23.599	2.634	20.261	20.261
2	2.649	20.378	43.977	2.158	16.603	36.864
3	1.123	8.638	52.615	1.963	15.102	51.966
4	1.044	8.032	60.647	1.129	8.681	60.647
5	.951	7.313	67.960			
6	.865	6.657	74.617			
7	.707	5.437	80.054			
8	.642	4.939	84.993			
9	.586	4.505	89.499			
10	.524	4.034	93.533			
11	.362	2.781	96.313			
12	.280	2.156	98.469			
13	.199	1.531	100.000			

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
WalkabilityIndex	.791			
Pop_WeightedDensity	.747			
PopChurning	-.667			
AirQualityIndex	-.601			.404
AffordabilityIndex		.822		
AverageYrsofSchooling		.774		
TownsendIndex	-.545	.599		
FoodAccess		-.554		
EmploymentAccess			.785	
IncomeInqualityGINI			.770	
JobParticipation	.304		.686	
SpatialSegragation				.734
HealthCareAccess				-.579

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Community Environmental Profile

Air Quality Index (EPA)
- Neurological Risk
- Cancer Risk
- Respiration Risk
- On-road Pollution
- Non-road
- Non-point

Population Churning Index
- Inflow Mobility
- Outflow Mobility

Population-Weighted Density

Workability Index
- Density
- Diversity (Land-use)
- Design (Connectivity)
- Distance to Transit
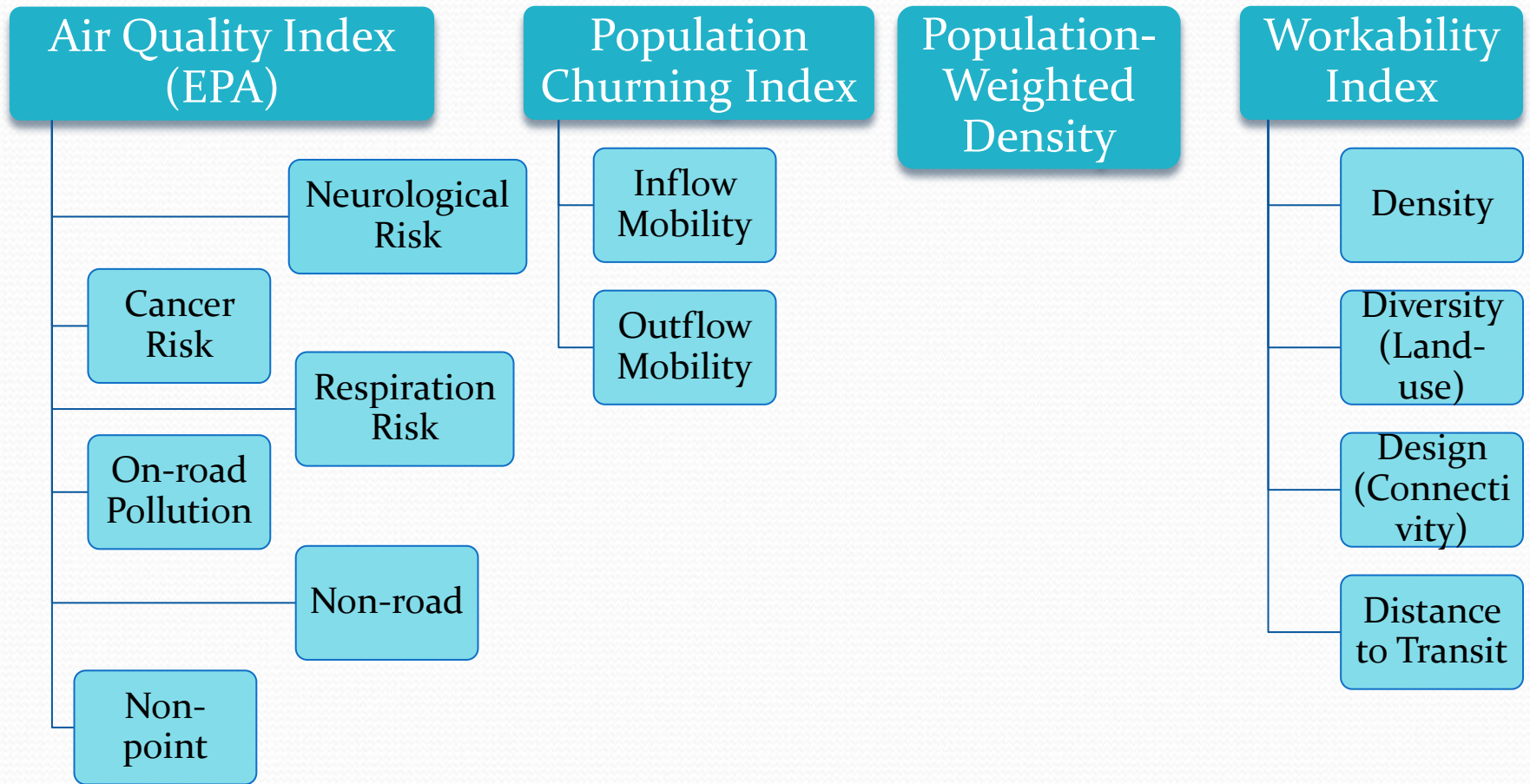
Consumer Opportunity Profile

Affordability Index
- Housing Cost
- Transportation Costs
- Average Income

Education Index
- Ave. Years of Schooling

Food Accessibility Index
- % Low Income
- % Low Access to Major Grocery Store
- Distance to Grocery Store

Material Deprivation Index
- Unemployment
- Autoless Homes
- Home Ownership
- Overcrowding

Economic Opportunity Profile

Employment Access Index	Income Inequality Index	Job Participation Index
Number of Jobs	GINI Coefficient	% of Working Age Population in the Labor Force
Distance to Jobs		

Wellness Disparity Profile

Access to Care Index
- % Uninsured
- Primary Care Physician FTEs within 30 miles

Segregation Index
- Race/Ethnicity
- Population
- Spatial Influence Weighting

HOI Spatial Modeling (Analytical)

Place Matters

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	34.580391	0.892283	38.754946	0.000000*	1.029163	33.600506	0.000000*	--------
ENVIRONMEN	3.368709	1.227647	2.744038	0.006128*	1.523337	2.211401	0.027115*	1.002880
OPPORTUNIT	33.515151	0.826917	40.530226	0.000000*	0.847554	39.543388	0.000000*	1.007598
ECONOMIC_P	27.932742	1.859343	15.022906	0.000000*	2.421745	11.534136	0.000000*	1.014627
WELLNESS_P	0.282575	0.858888	0.329001	0.742205	0.867857	0.325601	0.744775	1.011853

OLS Diagnostics

Input Features:	Epi_Seminar	Dependent Variable:	DFLE
Number of Observations:	1839	Akaike's Information Criterion (AICc) [d]:	10808.431632
Multiple R-Squared [d]:	0.517291	Adjusted R-Squared [d]:	0.516238
Joint F-Statistic [e]:	491.347488	Prob(>F), (4,1834) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	1701.638771	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	54.124483	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	235.963270	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Histogram of Standardized Residuals
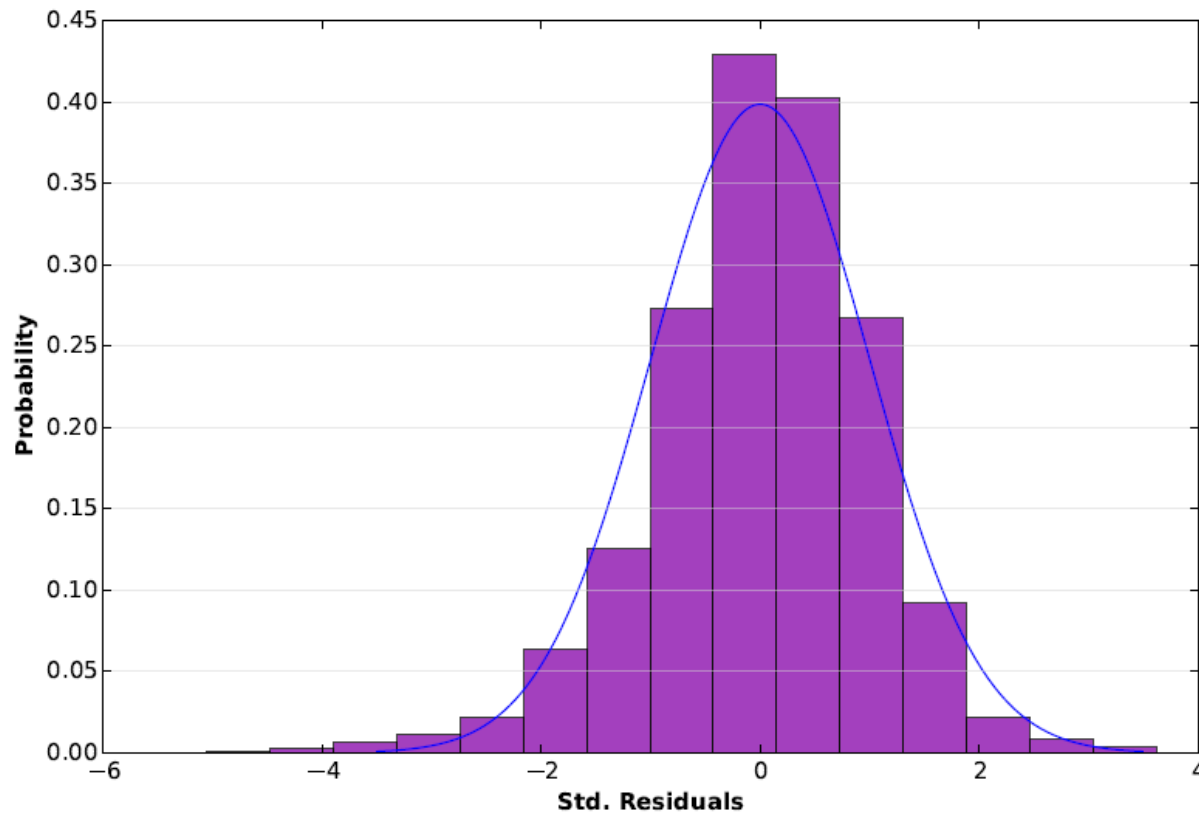
This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.
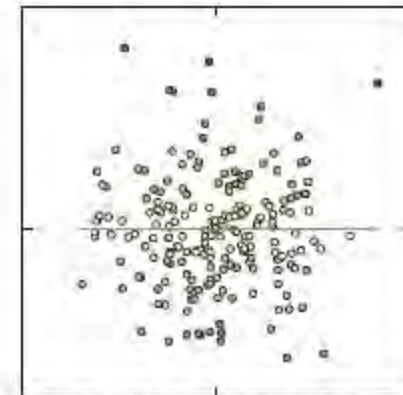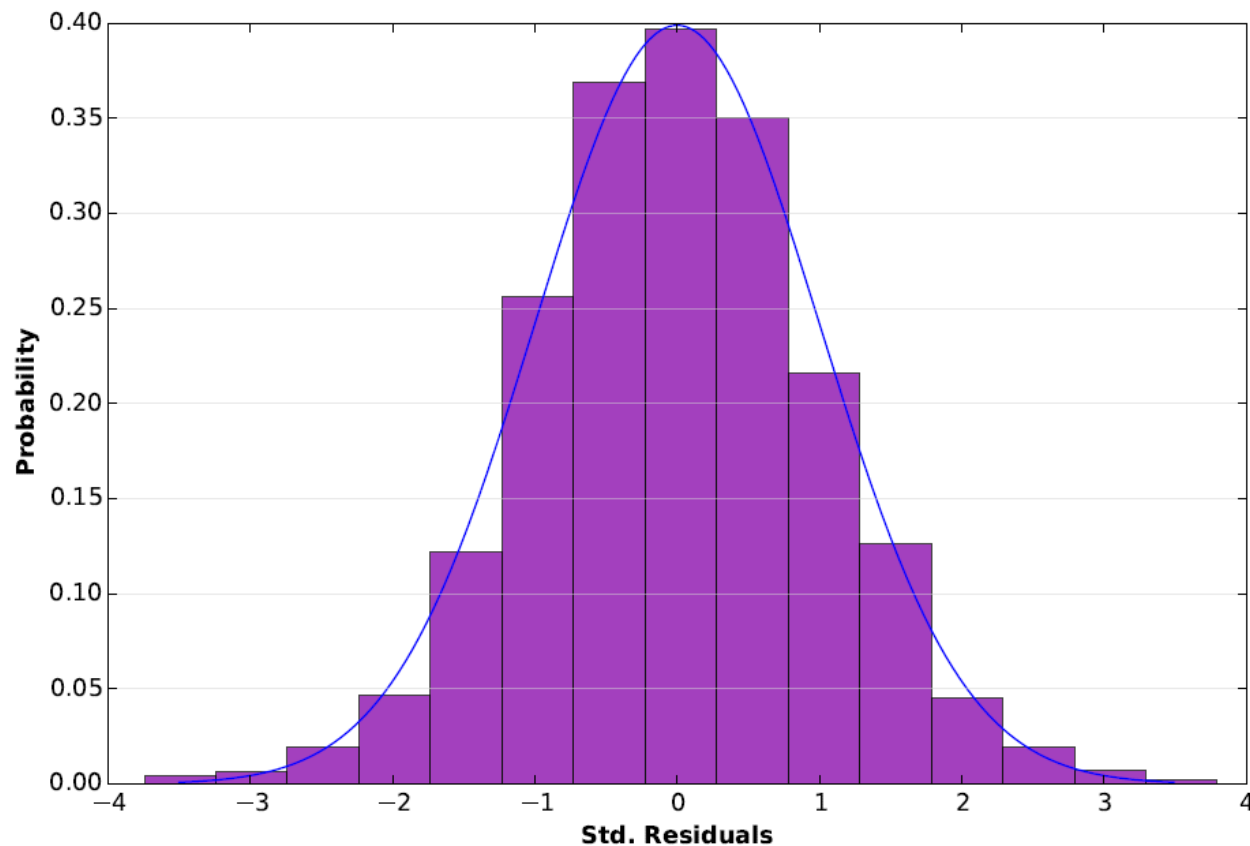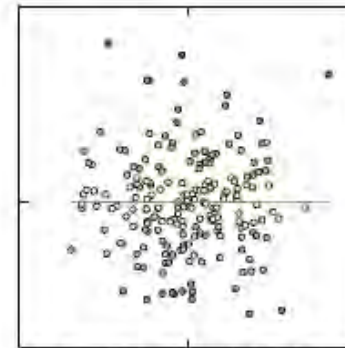

Random Residuals

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	50.494325	0.666501	75.760321	0.000000*	0.803969	62.806298	0.000000*
MAIN_HOI_S	34.218107	1.542373	22.185368	0.000000*	1.795923	19.053210	0.000000*

OLS Diagnostics

Input Features:	Epi_Seminar	Dependent Variable:	DFLE
Number of Observations:	1839	Akaike's Information Criterion (AICc) [d]:	11705.263538
Multiple R-Squared [d]:	0.211314	Adjusted R-Squared [d]:	0.210885
Joint F-Statistic [e]:	492.190539	Prob(>F), (1,1837) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	363.024821	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	62.727166	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	12.469836	Prob(>chi-squared), (2) degrees of freedom:	0.001960*

Histogram of Standardized Residuals

This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.

Random Residuals

VARNAME	VARIABLE	DEFINITION *
Neighbors	10.0	
ResidualSquares	29.3	
EffectiveNumber	462.5	
Sigma	0.30	
AICc	1102.7	
R2	0.83	
R2Adjusted	0.59	
Dependent Variable	0	DFLE
Explanatory Variable	1	HOI

Note: DFLE - Disability Free Life Expectancy (Healthy Life)

HOI - Health Opportunity Index (composite Index)
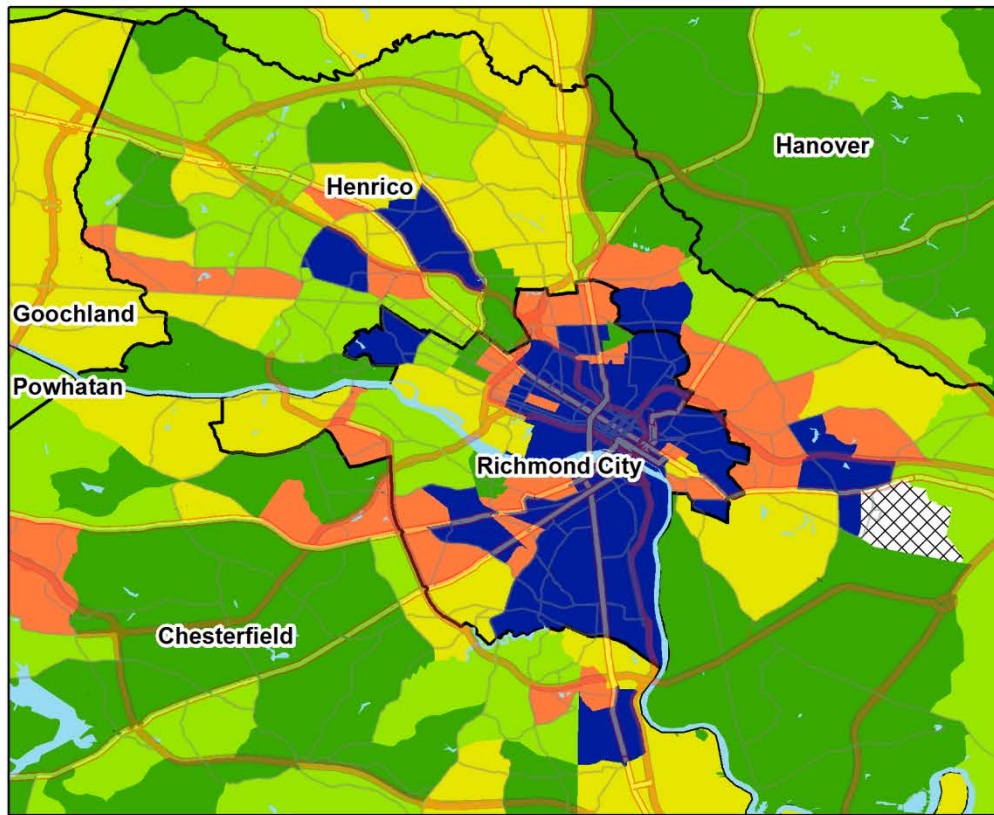
Policy Perspective

Action Matters - HPTAs

Virginia
Health Opportunity Index (HOI)
by Census Tract *
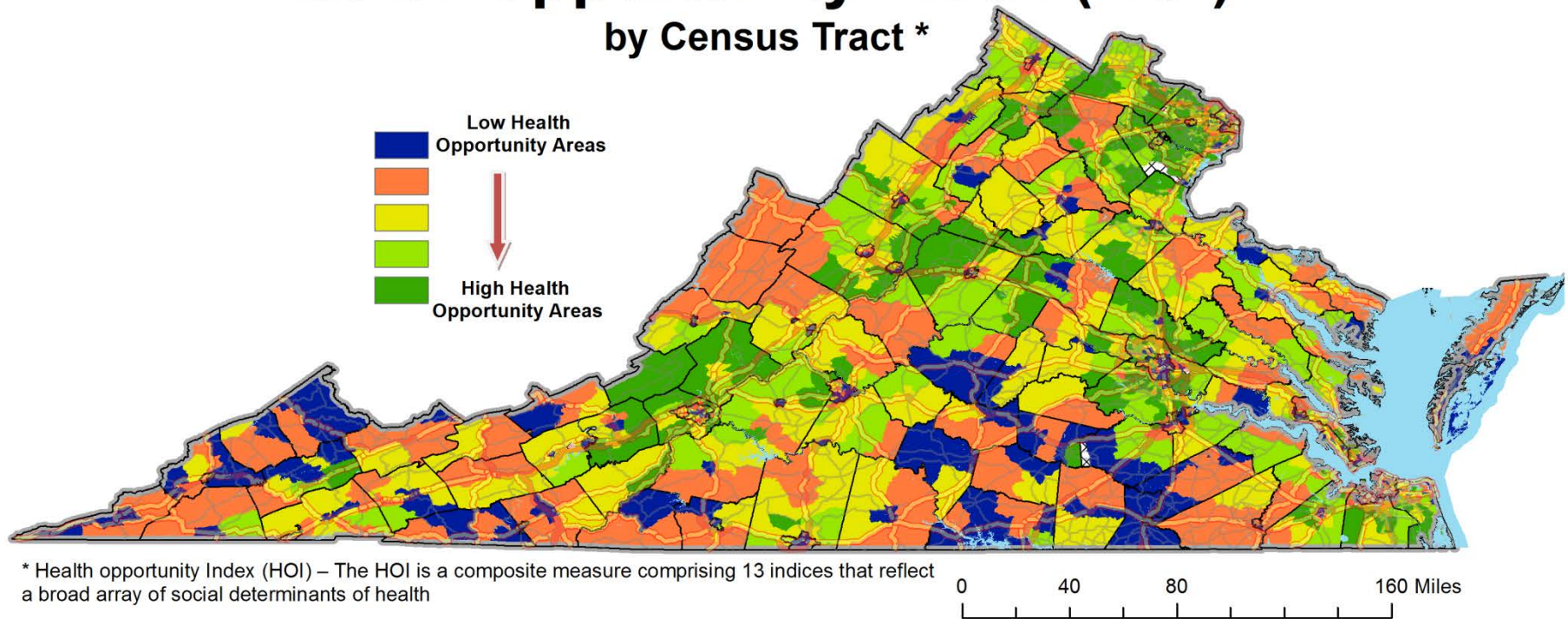
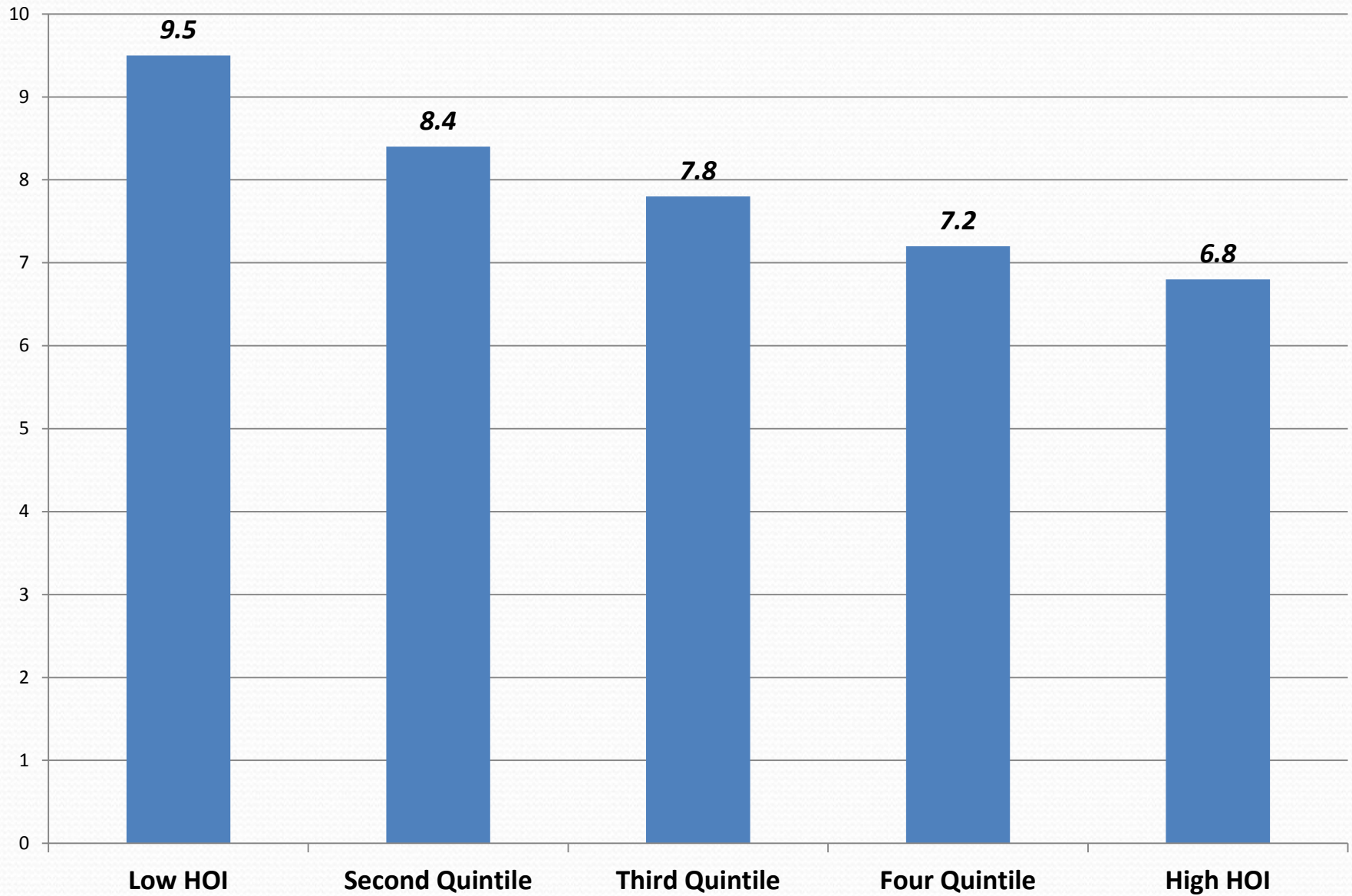Low Health
Opportunity Areas

High Health
Opportunity Areas

* Health opportunity Index (HOI) – The HOI is a composite measure comprising 13 indices that reflect a broad array of social determinants of health
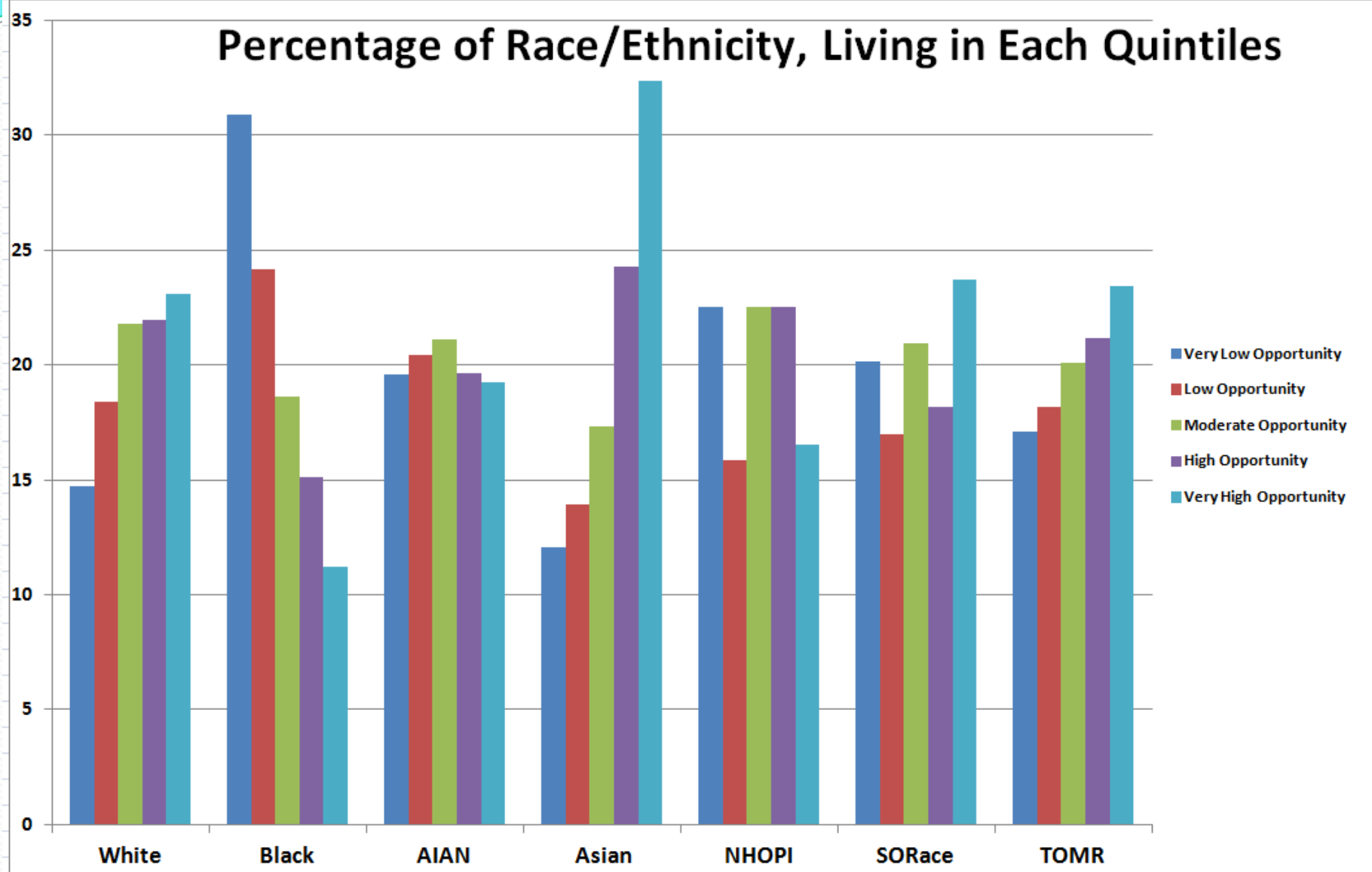
0   40   80   160 Miles

Virginia
Health Opportunity Index (HOI)
by Census Tract *

Low Health Opportunity Areas

High Health Opportunity Areas

Hanover

Henrico

Goochland

Powhatan

Richmond City

Chesterfield

* Health opportunity Index (HOI) – The HOI is a composite measure comprising 13 indices that reflect a broad array of social determinants of health

0   40   80   160 Miles
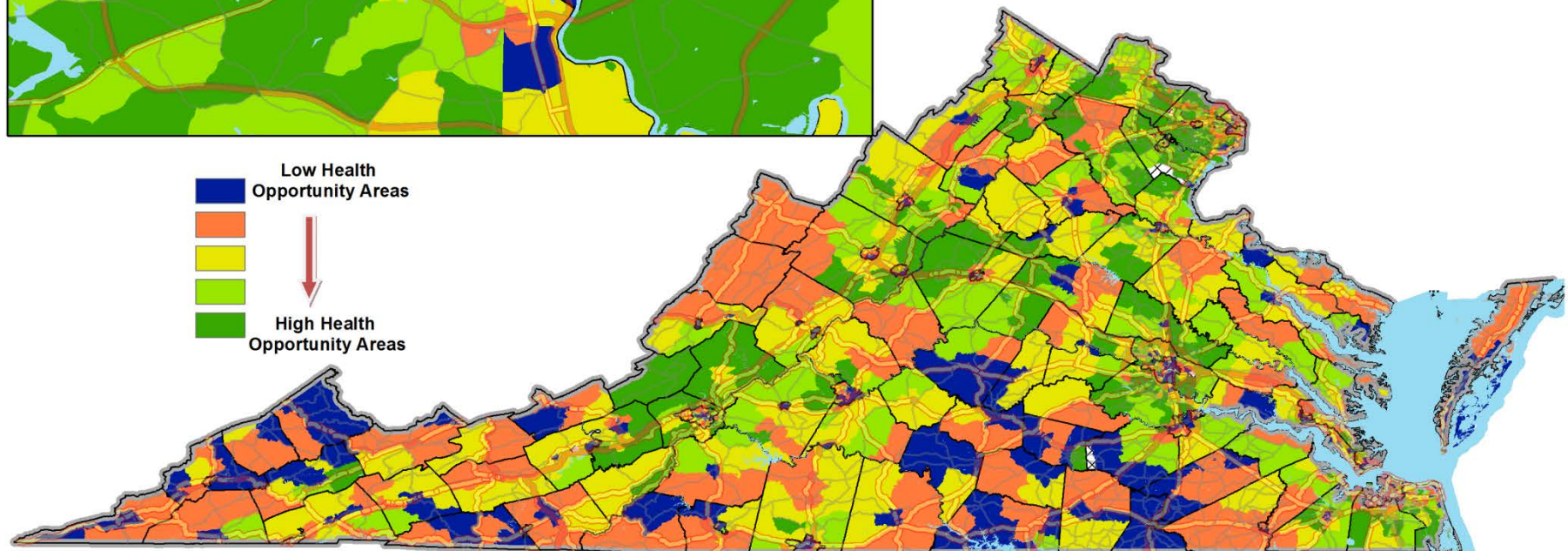
Infant Mortality per 1,000 Live Births

Low Birth Weight (%)

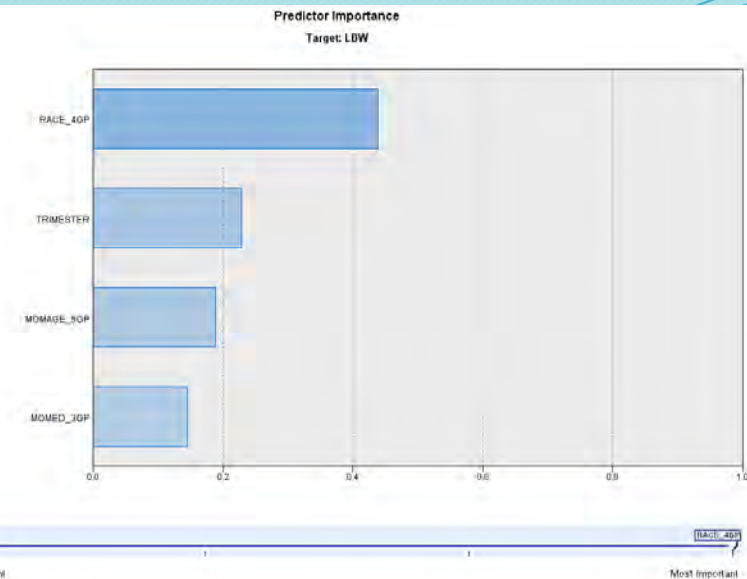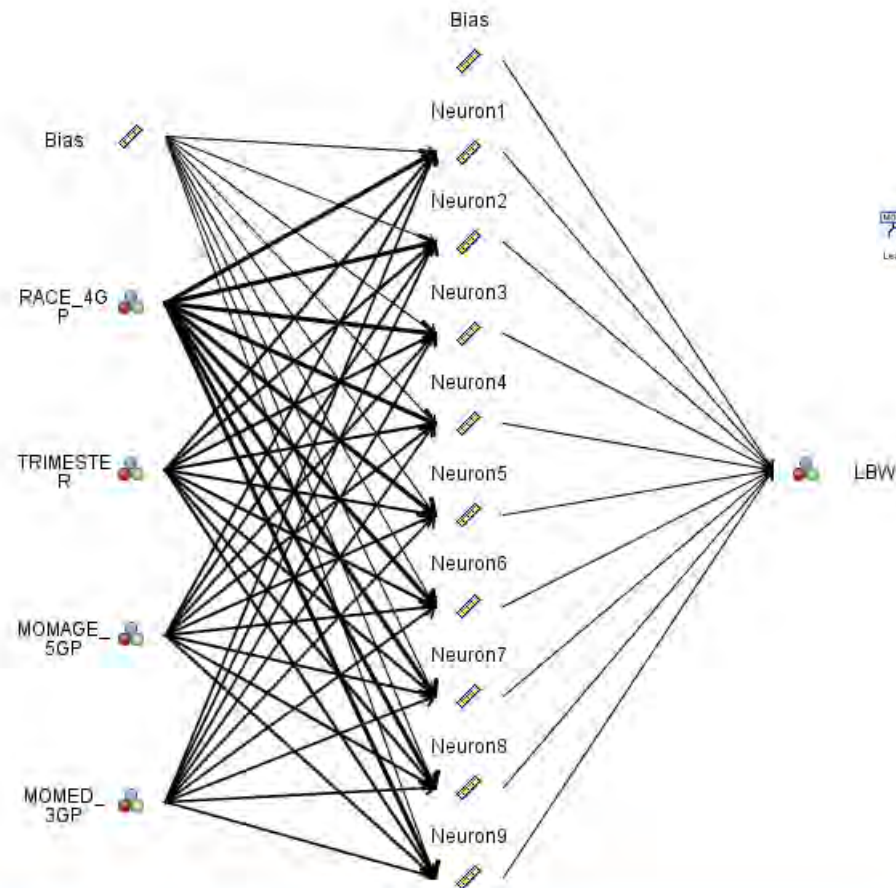| Category	Value
Low HOI	9.5
Second Quintile	8.4
Third Quintile	7.8
Four Quintile	7.2
High HOI	6.8
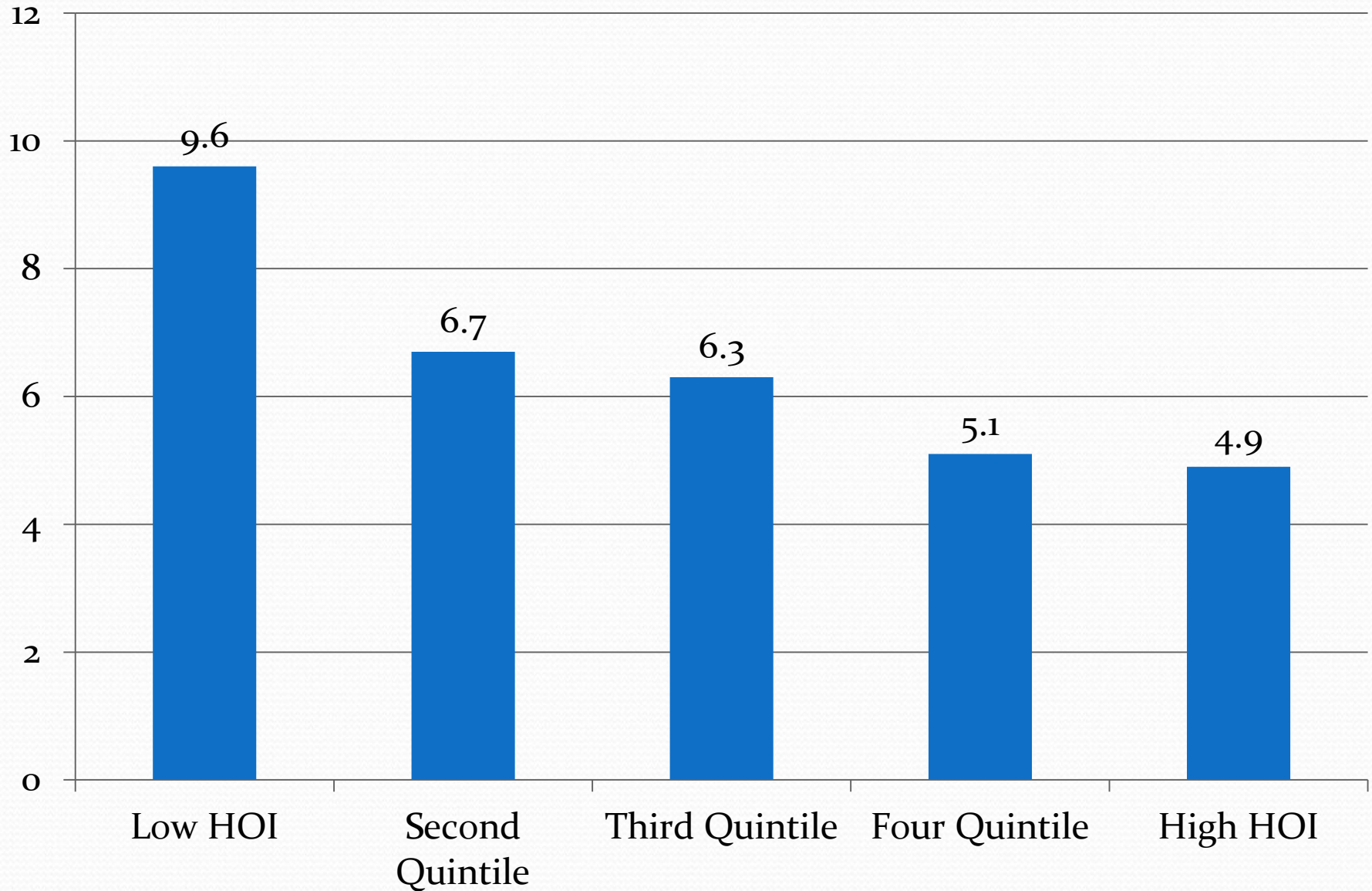
Percentage of Race/Ethnicity, Living in Each Quintiles

Predictive Analytics
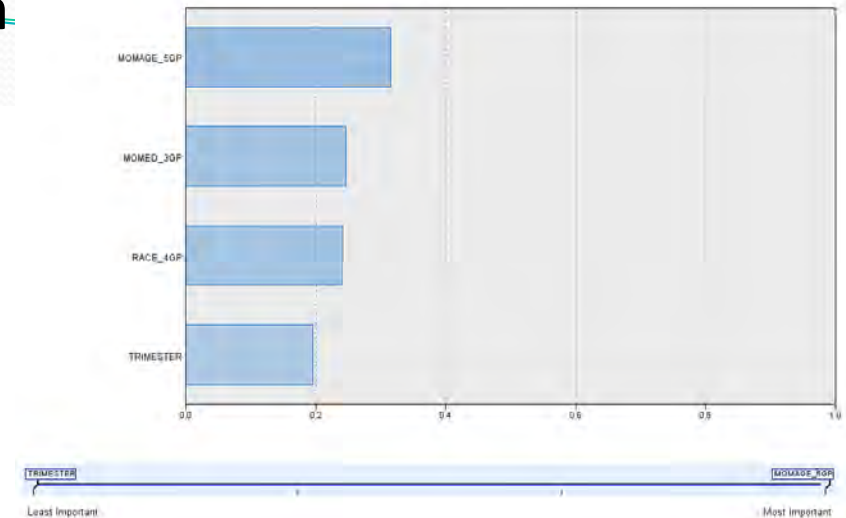
Predictive Analytics for Low Birth Weight (Low HOI)

Predictive Analytics for Low Birth Weight (High HOI)
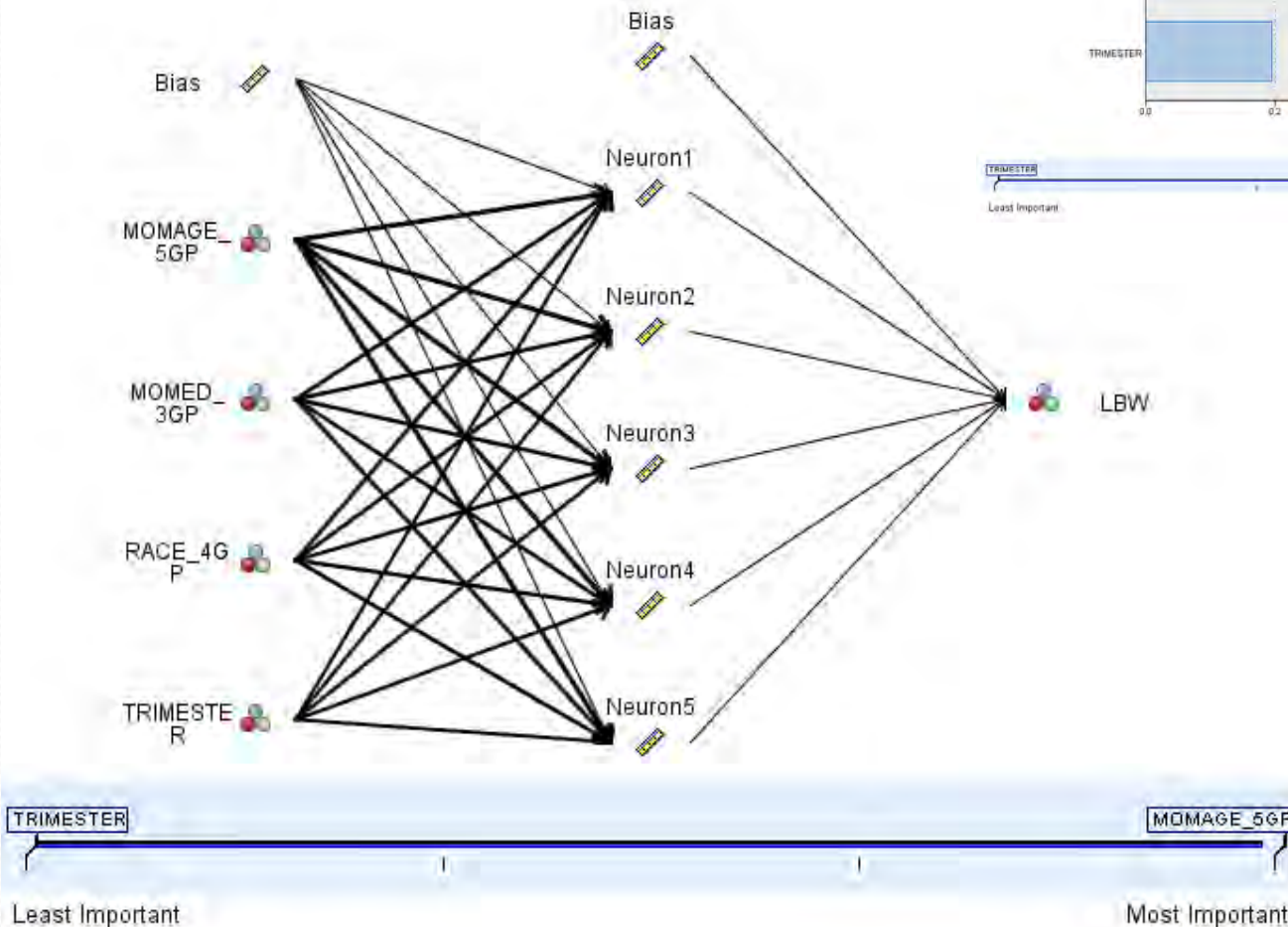
Thank you!!

Rexford Anson-Dwamena

Rexford.Dwamena@vdh.virginia.gov

Justin Crow, MPA

Justin.crow@vdh.virginia.gov

Anna Riggan

Anna.Riggan@vdh.virginia.gov

Brendan Noggle

Brendan.Noggle@vdh.virginia.gov

HOI Technical Workgroup

Graham Truelove

Joseph Grubbs

Brittany Cox

Dr. McFadden

Special Thanks to

Dr. Derek Chapman for his constructive suggestions