Workshop on Crime Travel Demand Model

Ned Levine, PhD
Ned Levine & Associates
Houston

Ninth Crime Mapping Conference
MAPS
March 29, 2007

Need for Metropolitan Offender Travel Model

- **Current crime travel models too simple**
  Journey to crime models are only descriptive. Little theory behind them.
- **Many offenders travel over sizeable distances**
- **Temporal variation is considerable**
- **Increasing mobility of American society**
- **Transportation links crime generators & attractors**

Travel Demand Forecasting in Transportation

- Travel demand forecasting developed since 1950s
- FHWA funded development of four stage model
- This model is used in every metropolitan area in the U.S. and many cities worldwide
- While newer approaches are emerging, we wanted to apply model to crime travel behavior
- Don’t ‘reinvent the wheel’

A Crime Travel Demand Model

- Adapts travel demand theory to crime analysis
- Statistical approach to offender travel
- Model is part of *CrimeStat III*

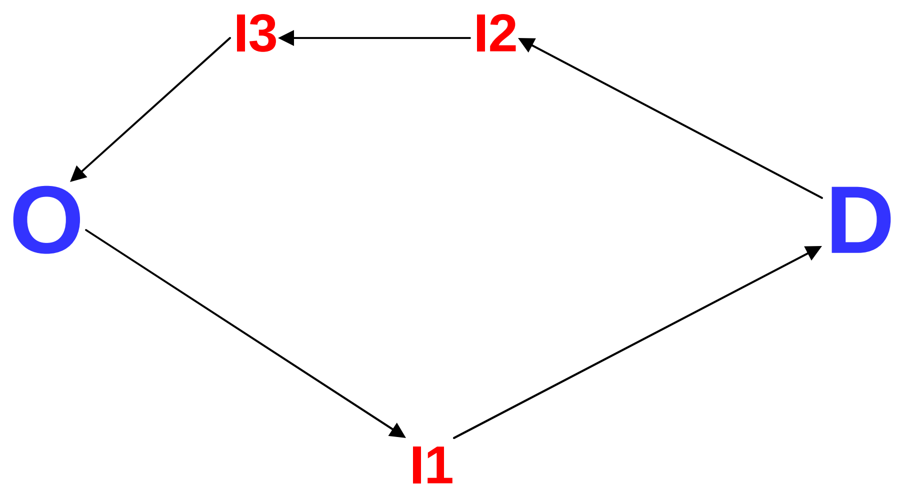
Crime Trip: Definition

A complete *round-trip* journey from an offender's residence (origin) that includes a committed crime at a specified location (destination)

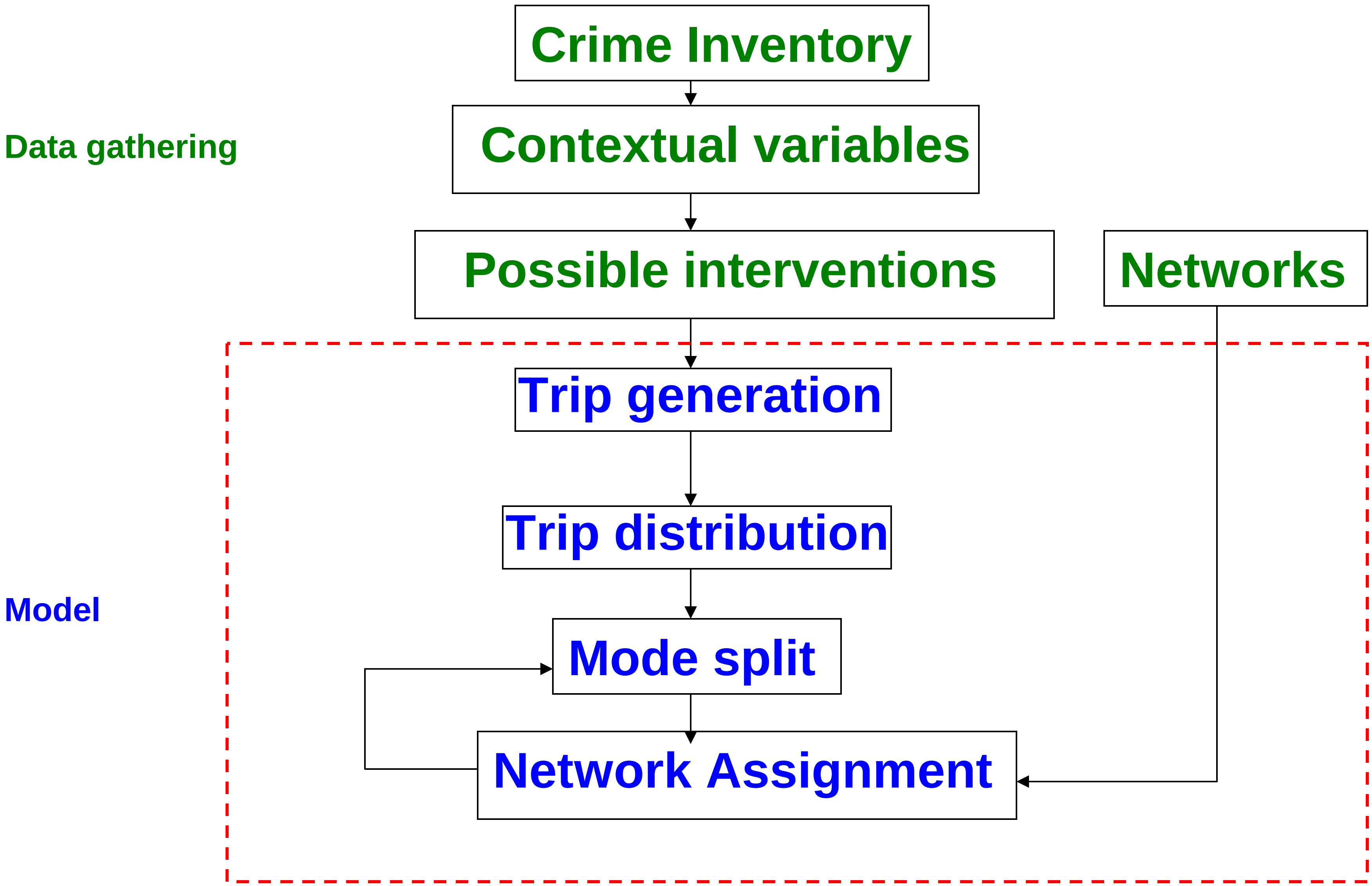
There is an *origin* (residence)

There is a *destination* (crime location)

There may be *intermediate* links



Crime Travel Demand Forecasting



Crime Travel Demand is NOT Journey to Crime Modeling

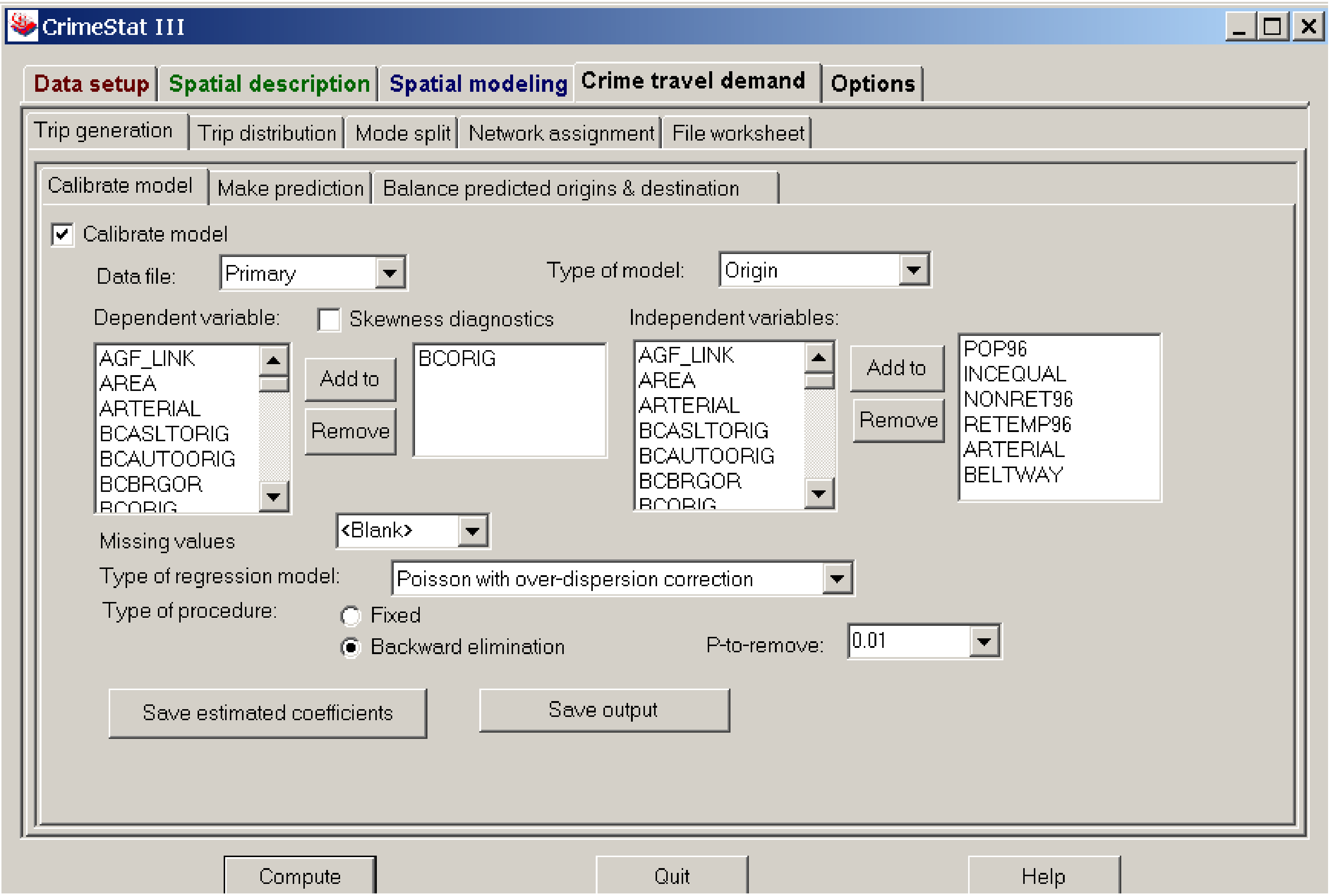
Journey to Crime is a description, not an explanation

- **Descriptive framework**
  - ***Crime trips = f(distance, crime type, & ‘buffer zone’)***
- **Single-stage model**
- **Non-adjustable**

Travel Demand is a predictive model ***(with some explanations)***

- **Predictive framework**
  - ***Crime trips = f(productions, attractions, impedance)***
  - ***Productions = g(predictive variables)***
  - ***Attractions = h(predictive variables)***
  - ***Impedance = I(cost & availability variables)***
- **Multi-stage model**
- **Can manipulate variables to produce sensitivity analysis**
- **Can use it to make predictions**

Trip Generation Model



Data Gathering

Three Types of Data Needed

Needed Data: I

Crime data

- **Crime data by location (“destination”)**
- **Crime data by residence location (“origin”)**
- **By crime type**
- **By time of day**

Where to find?

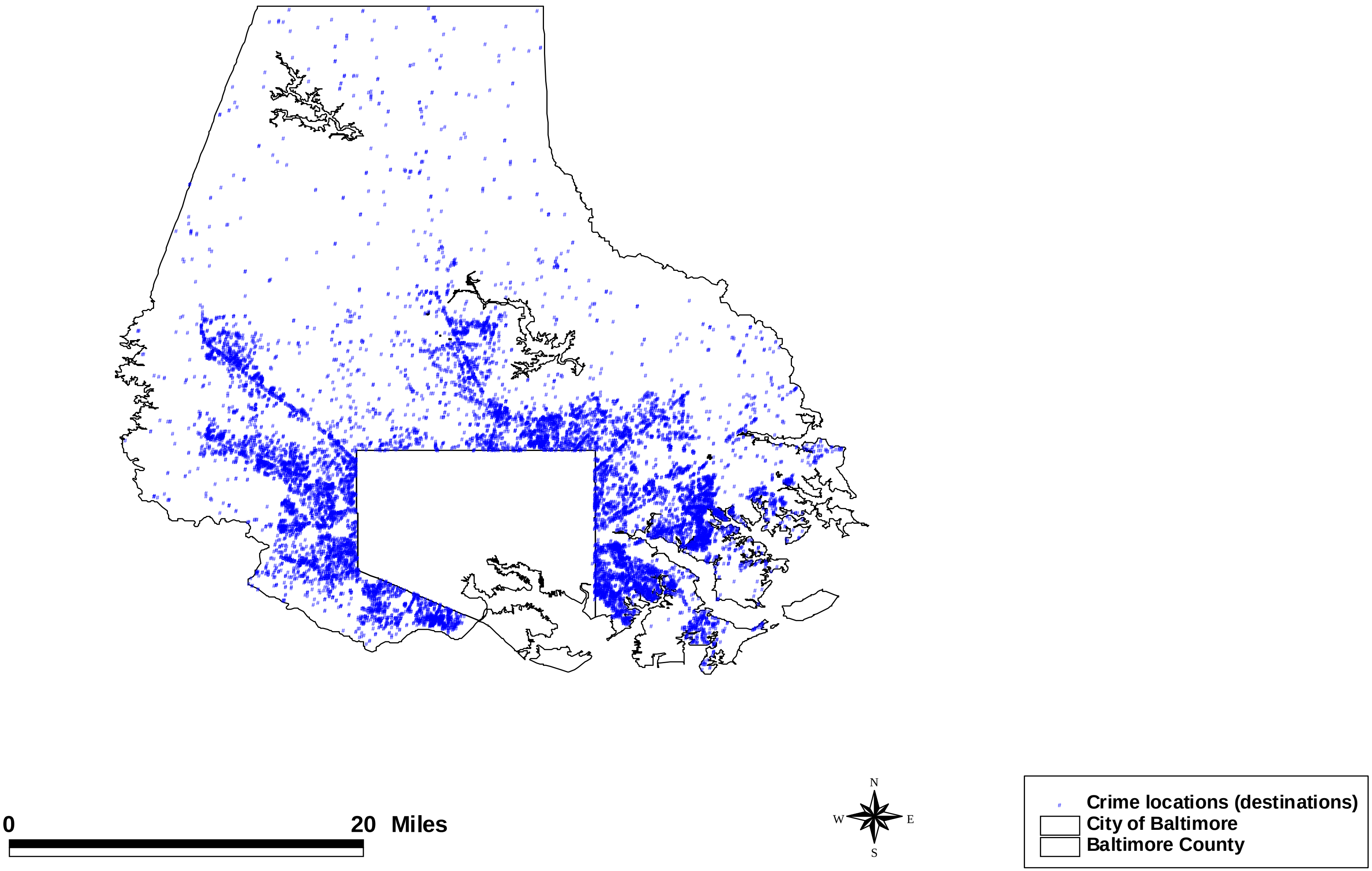
Police departments

Crime Data Requirements

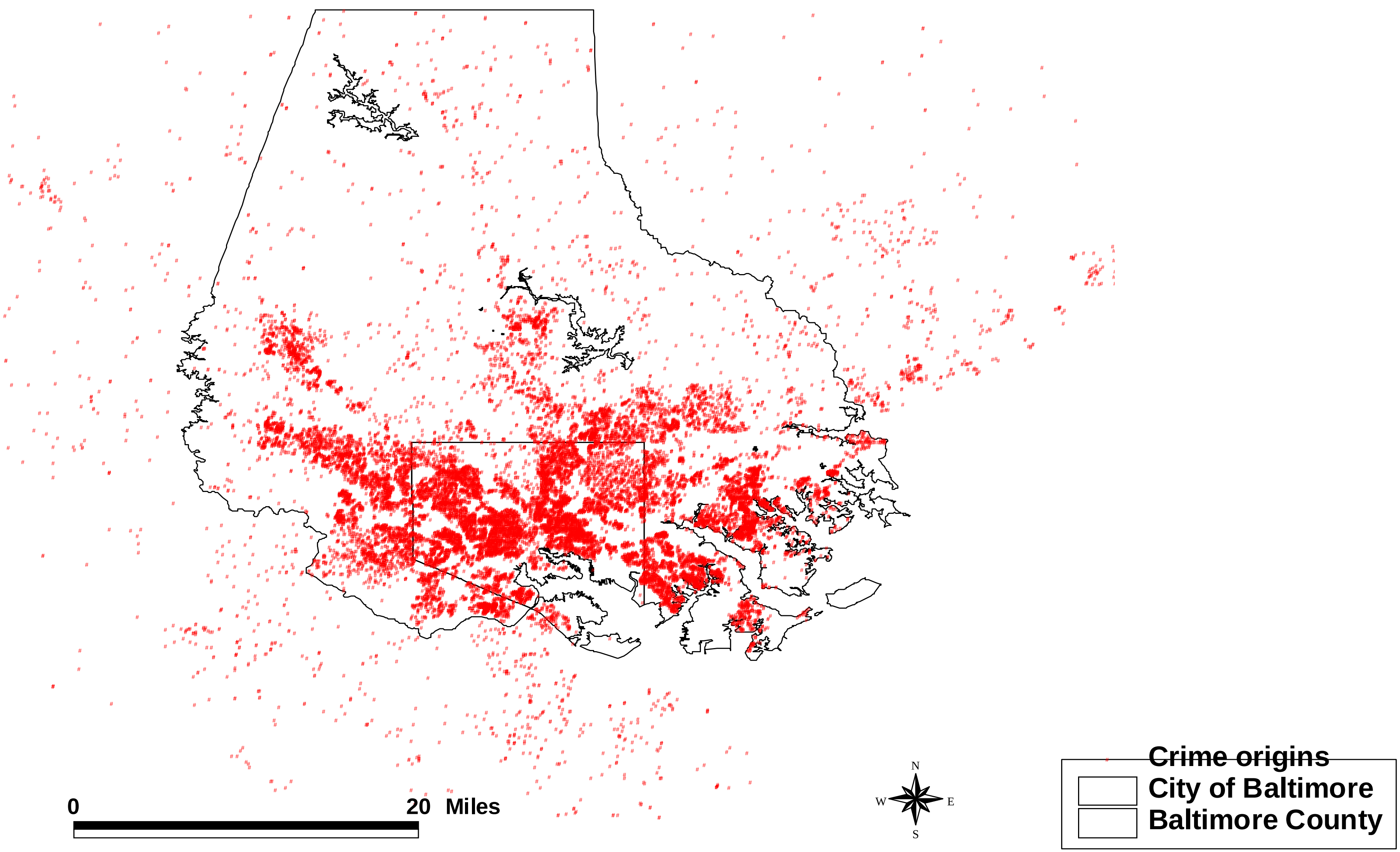
Minimum data requires origin and destination location

UCR	DATE	INCIDX	INCIDY	HOMEX	HOMEY
430	1/5/97	-76.8131	39.3822	-76.8131	39.3822
440	5/17/95	-76.4490	39.3355	-76.4489	39.3355
210		-76.4068	39.3388	-76.5281	39.3085
210		-76.4142	39.2801	-76.4142	39.2801
430		-76.5527	39.3908	-76.4410	39.3080
440		-76.7581	39.3131	-76.7709	39.3105
440	3/29/94	-76.5095	39.2735	-76.5095	39.2735
440	1/22/96	-76.7344	39.3212	-76.6899	39.3364
690	7/13/93	-76.4525	39.3012	-76.6050	39.3020
690	10/8/94	-76.5278	39.2584	-76.5051	39.3970
690	8/10/97	-76.7384	39.3275	-76.7384	39.3275
690	3/10/96	-76.7325	39.3018	-76.7325	39.3018

Baltimore County Crime Locations: 1993-1997
Location of Crimes Committed by Offenders (N=41,974)



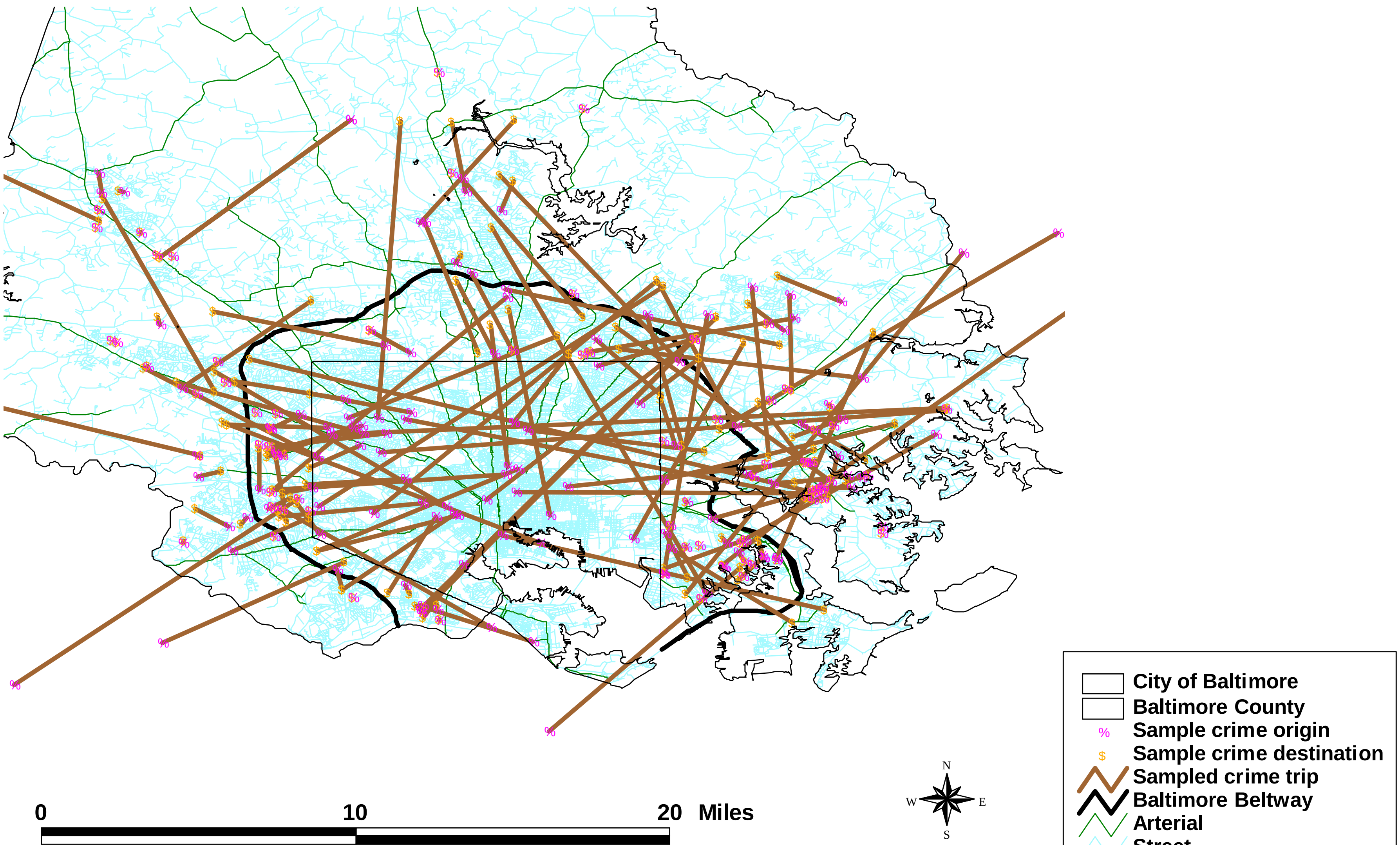
Baltimore County Offender Residences: 1993-1997
Location of Baltimore County Offenders When Arrested (N=41,974)



Implicit is a “Trip” from the Origin to the Destination

- Don’t know intermediate trips or what follows
- Also, uncertainty as to true origin
- Nevertheless, can construct a consistent estimate
on the assumption of an origin-to-destination crime trip

Baltimore County Crime Trips: 1993-1997
Origins and Destinations
Sample of 200 Crime Trips



Crimes are assigned to zones

- By crime location (destination)
- By offender residence location (origin)

Choice of Zones Must Balance:

Ability to obtain data

Small enough to capture zone-to-zone trips

But not so small as to make too many empty zones

Minimize statistical distortion due to size and shape

Commonly-used Zone Models

Census geography

- Available population data for small areas
- Lack of information on employment
- Zone size increases with distance from center
- Irregular shapes may create distortion

Traffic analysis zones (TAZ)

- Available population and employment estimates
- Zone size increases with distance from center
- Irregular shapes may create distortion
- May be too big for crime trips

Grid cells

- Uniform cell sizes eliminate size distortion
- Standardize shape distortion
- Usually, lack of population and employment data

Problems with any type of zone model:

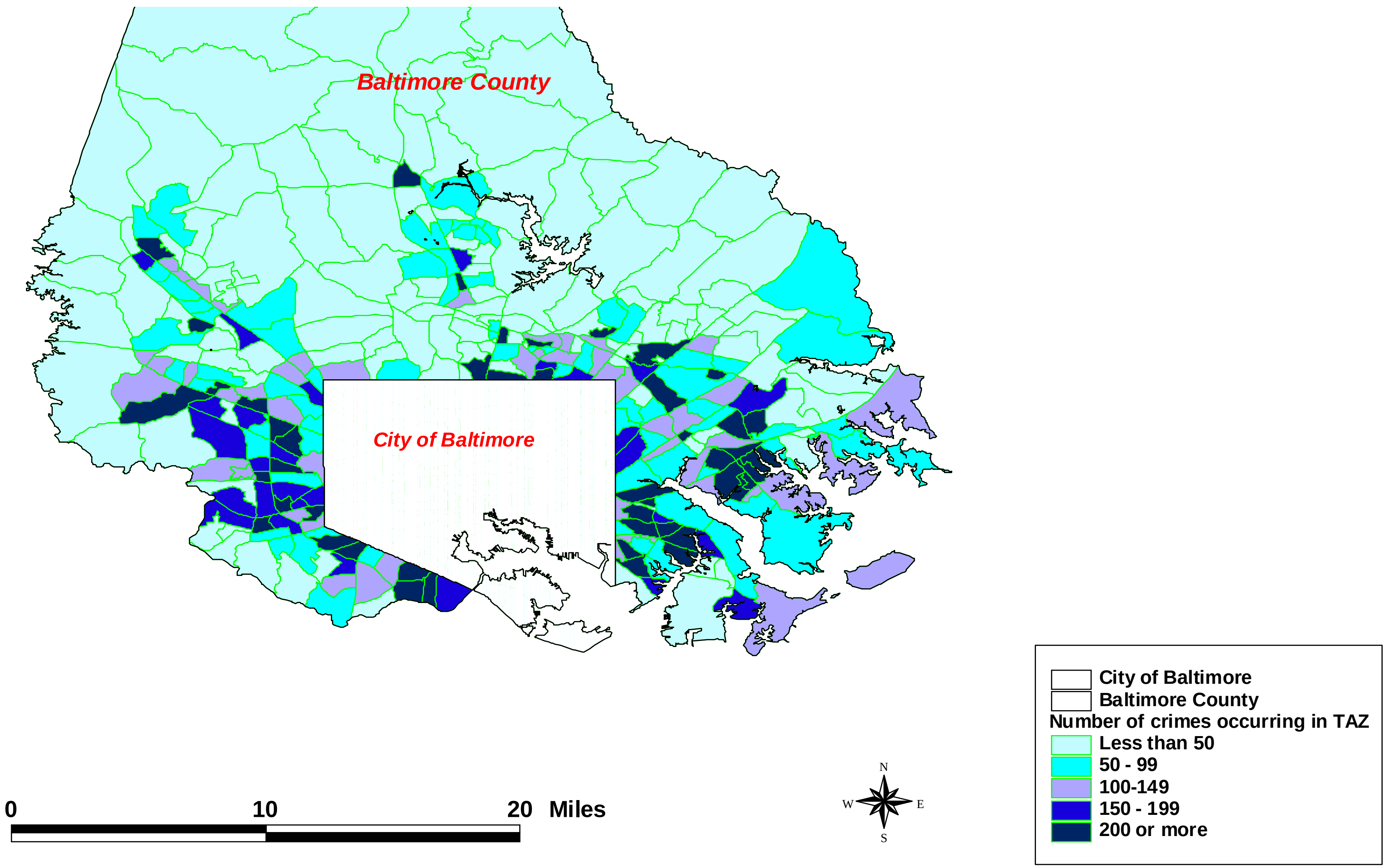
Allocating other data to it (e.g., land use, policing)

Intra-zonal (or local) trips

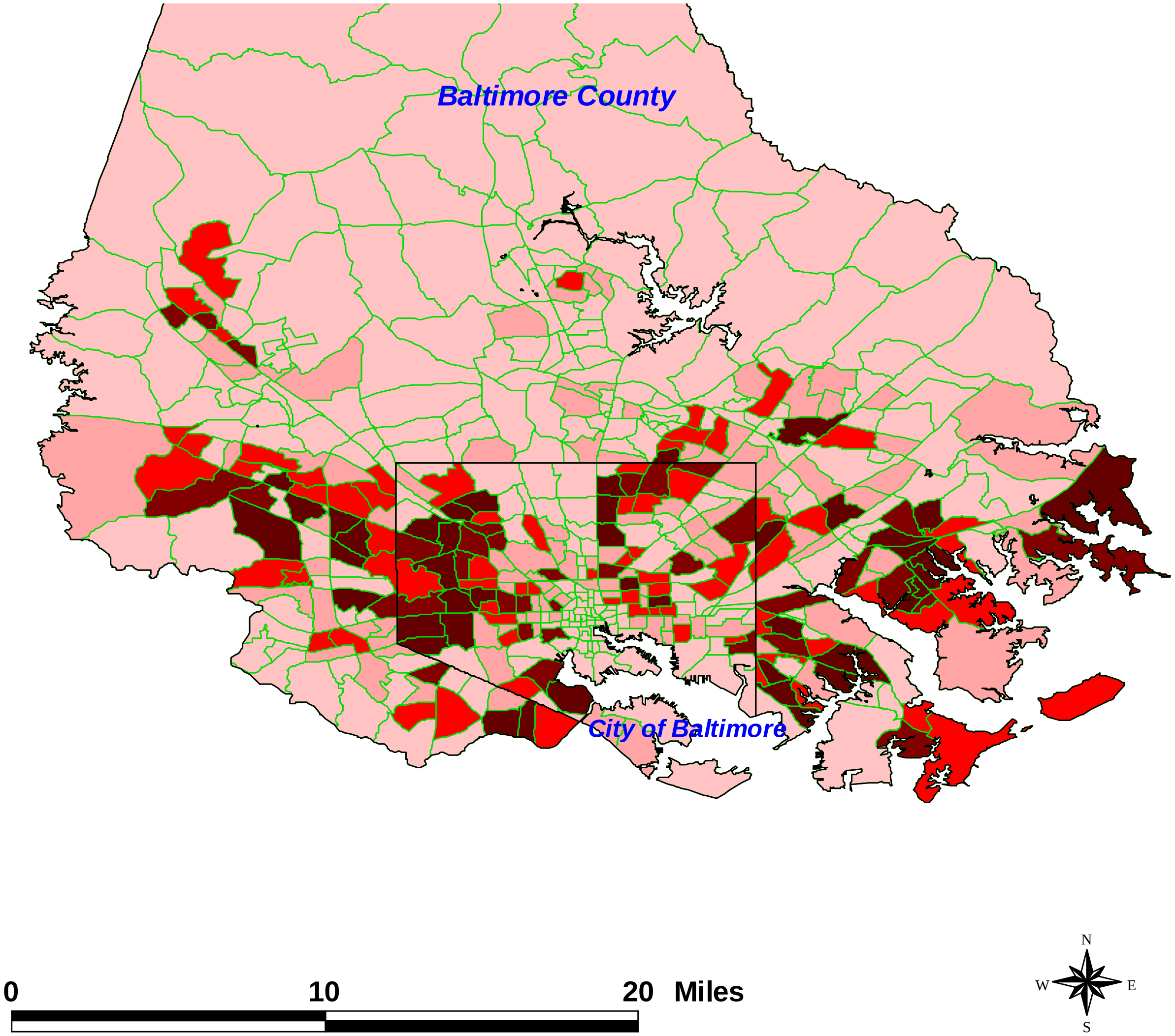
Trips from outside the study area ("external trips")

Metropolitan Baltimore Traffic Analysis Zones: 1998

Crimes Destinations by TAZ
Number of Crimes Occurring in TAZ
Baltimore County: 1993-1997



Crimes Origins by TAZ
Number of Crimes Originating in TAZ
Baltimore County: 1993-1997



City of Baltimore
Baltimore County
Number of crimes occurring in TAZ
Less than 50
50 - 99
100 - 149
150 - 199
200 or more

Needed Data: II-A

Socio-economic ("contextual") data for zones

Population
Employment
Income/poverty
Land use data

Where to find?

Census
City/County governments
Metropolitan Planning Organizations
http://www.ampo.org
Council of Governments
http://www.narc.org

Needed Data: II-B

Policy-related data for zones

- **Strategic data *(deployment, facilities)***
- **Planned interventions**

Where to find?

- **??**
- **Media**
- **Local government**
- **MPO/COG**
- **Police**

Needed Data: III

Network data

- **Road network**
- **Bus network**
- **Rail network**

Where to find?

- **TIGER (and variants)**
- **Private companies**
- **Transit agencies**
- **Metropolitan Planning Organizations**
- **State/local Dept. of Transportation**

Crime Trip Generation

Modeling Crime Generation:

Separate models of crime by zone are developed:

- **By crime origins (productions)**
- **By crime destinations (attractions)**

Crime production model

Crime productions (origins) for zone are a function of:

Population/households
Income
Poverty
Vehicle ownership
Particular land uses
(e.g., housing projects)
Spatial location
(centrality & spatial autocorrelation)

Crime attraction model

Crime attractions (destinations) for zone are a function of:

Population/households
Retail trade
Income
Particular land uses
(e.g., bars, parking lots, convenience stores, banks, pawn shops)
Special generators
(e.g., parks, stadiums, large shopping malls)
Spatial location
(centrality & spatial autocorrelation)

Count (Volume) Model

Estimate number of crimes in a zone (volume)

- **Not rates**

Population size is base variable

- **Should include it or proxy in model**

Generally, should not mix up rates with volumes

- **Rates are volumes divided by base populations**

- **Including rates may add multicolinearity**
  (e.g., vehicles per household co-mingles income and population)

Two Classic Approaches To Trip Generation

- Trip tables
- Regression models

Use of Trip Tables

- Cross-classify predictor variables
- Estimate mean number of trips per class

Median Income

Vehicle Ownership	*Low*	*Medium*	*High*
0-1	3.2	4.6	6.7
2+	5.4	7.8	8.1

Problems with
Trip Table Approach
For Crime Trip Generation

- Requires individual-level data to estimate rates

- Could be used with aggregate (zonal) data but..

    Requires interpretation

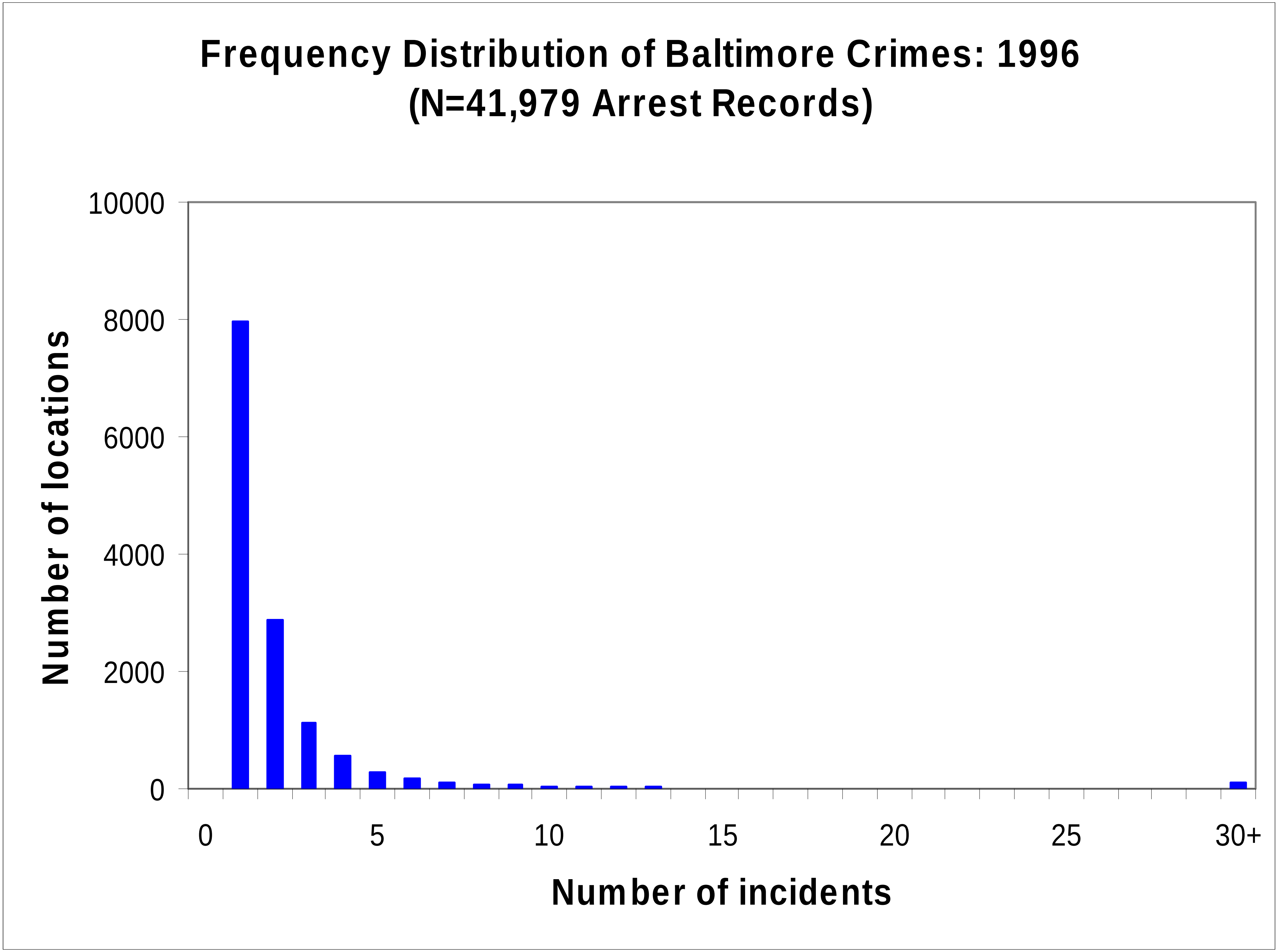
    Requires large sample to obtain estimates

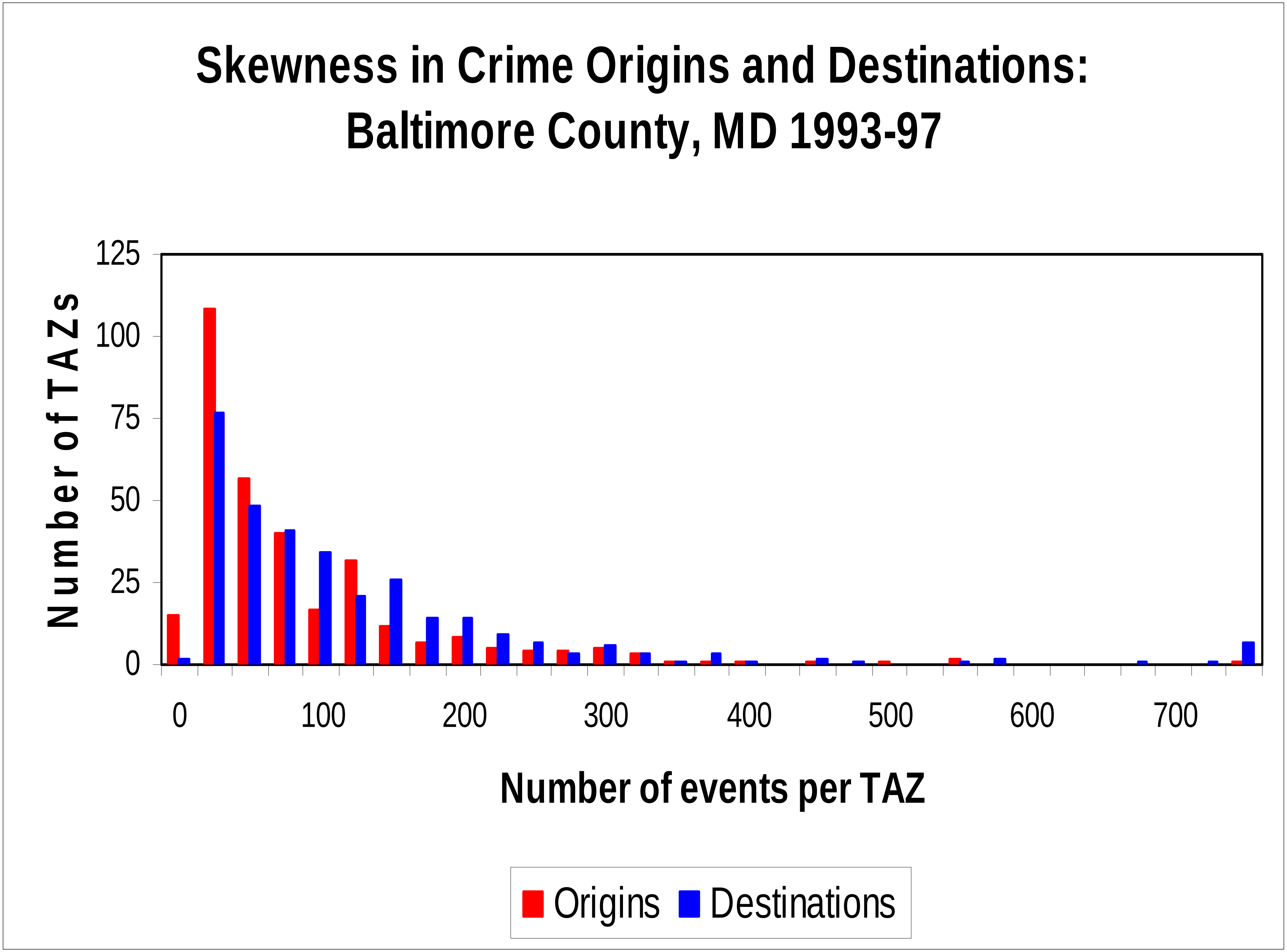
    Assumes homogeneity over study area

Use of Regression Models

Relates mean expected number of crimes to independent predictors

$$Y_i = f(X_1, X_2, X_3, \ldots X_k) + \varepsilon$$


Frequency Distribution of Baltimore Crimes: 1996
(N=41,979 Arrest Records)
10000
8000
6000
4000
2000
0
Number of locations
0
5
10
15
20
25
30+
Number of incidents


Skewness in Crime Origins and Destinations:
Baltimore County, MD 1993-97
Number of TAZs
125
100
75
50
25
0
0
100
200
300
400
500
600
700
Number of events per TAZ
Origins
Destinations

Classical Choice:
Ordinary Least Squares Regression

Number of crimes is a normally distributed variable, Y_i, which is function of a set of predictors and an error term

$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon$$

where

Y_i	*Number of events in zone i*
X1, X2, ….XK	*Independent variables*
α, β	*Intercept and coefficients*
ε	*Error term*

Problems with
Ordinary Least Squares Regression

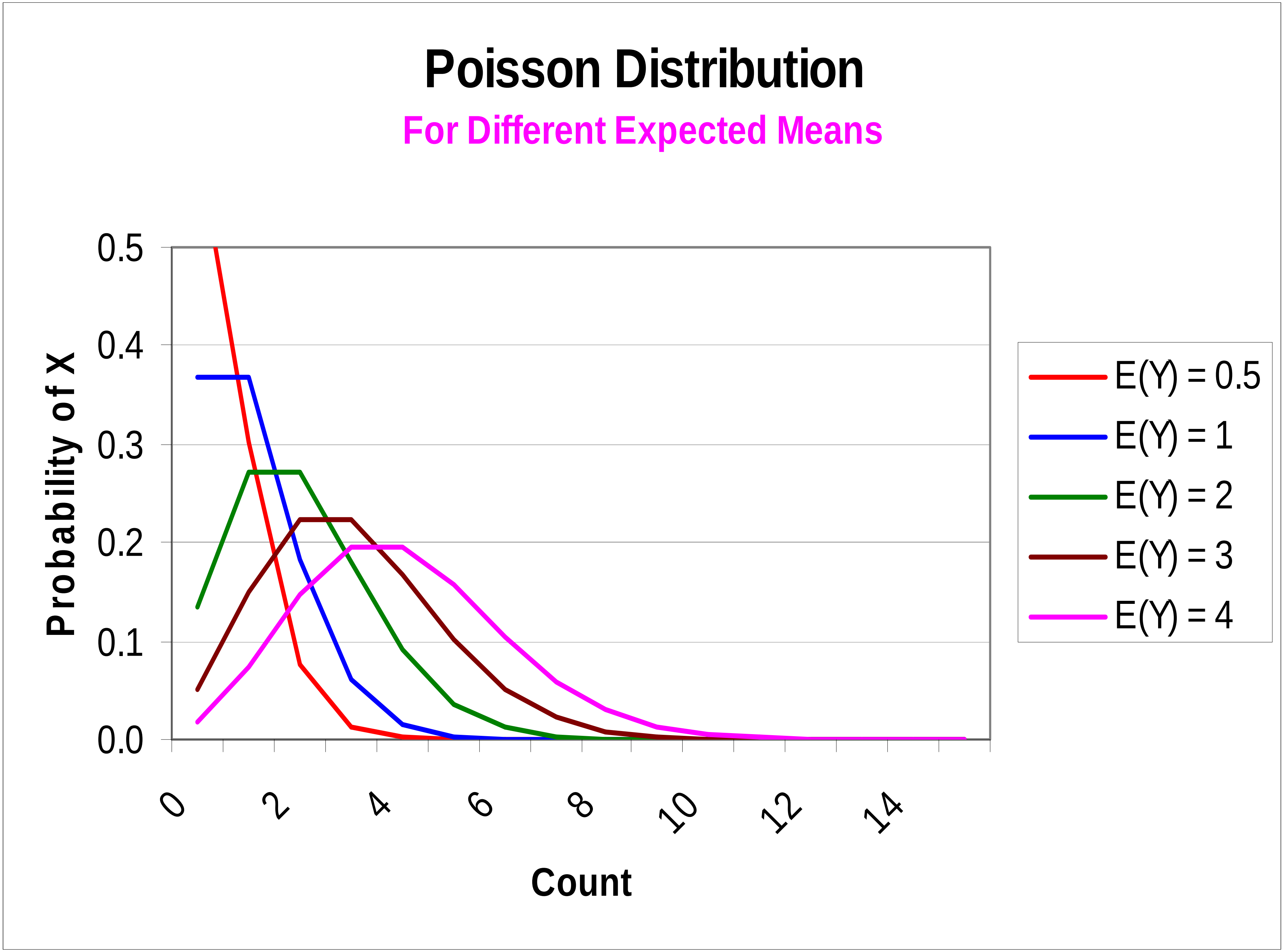
- **Negative predictions** ***(i.e., minimum not 0)***
- **Sum of predicted values does not equal sum of input values**
- **Assumes independent effects are linear**
- **Greater residual errors:**
  - **Overestimates crimes for most zones**
  - **Underestimates crimes for high crime zones**

Poisson Regression Model

Number of crimes is a function of a Poisson random variable with mean λ

$$\mathrm{Prob}(Y_i) = \frac{e^{-\lambda}\,\lambda_i^{Y_i}}{Y_i!}$$

$$\mathrm{Ln}(\lambda_i) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k$$



Poisson Distribution
For Different Expected Means
Probability of X
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
12
14
Count
E(Y) = 0.5
E(Y) = 1
E(Y) = 2
E(Y) = 3
E(Y) = 4

Use of Poisson Regression

Overcomes many of the problems with OLS

- **Minimum value is zero**
- **Fundamentally skewed model**
- **Sum of predicted values equals sum of input values**
- **Estimates more accurately number of crimes**

Modeling Crime Trip Generation

Two Separate Models

1. $$Ln[\ E(O_i)\ ] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots .. + \beta_k X_k$$

2. $$Ln[\ E(D_j)] = \phi + \acute{\eta}_1 X_1 + \acute{\eta}_2 X_2 + \ldots .. + \acute{\eta}_k X_k$$

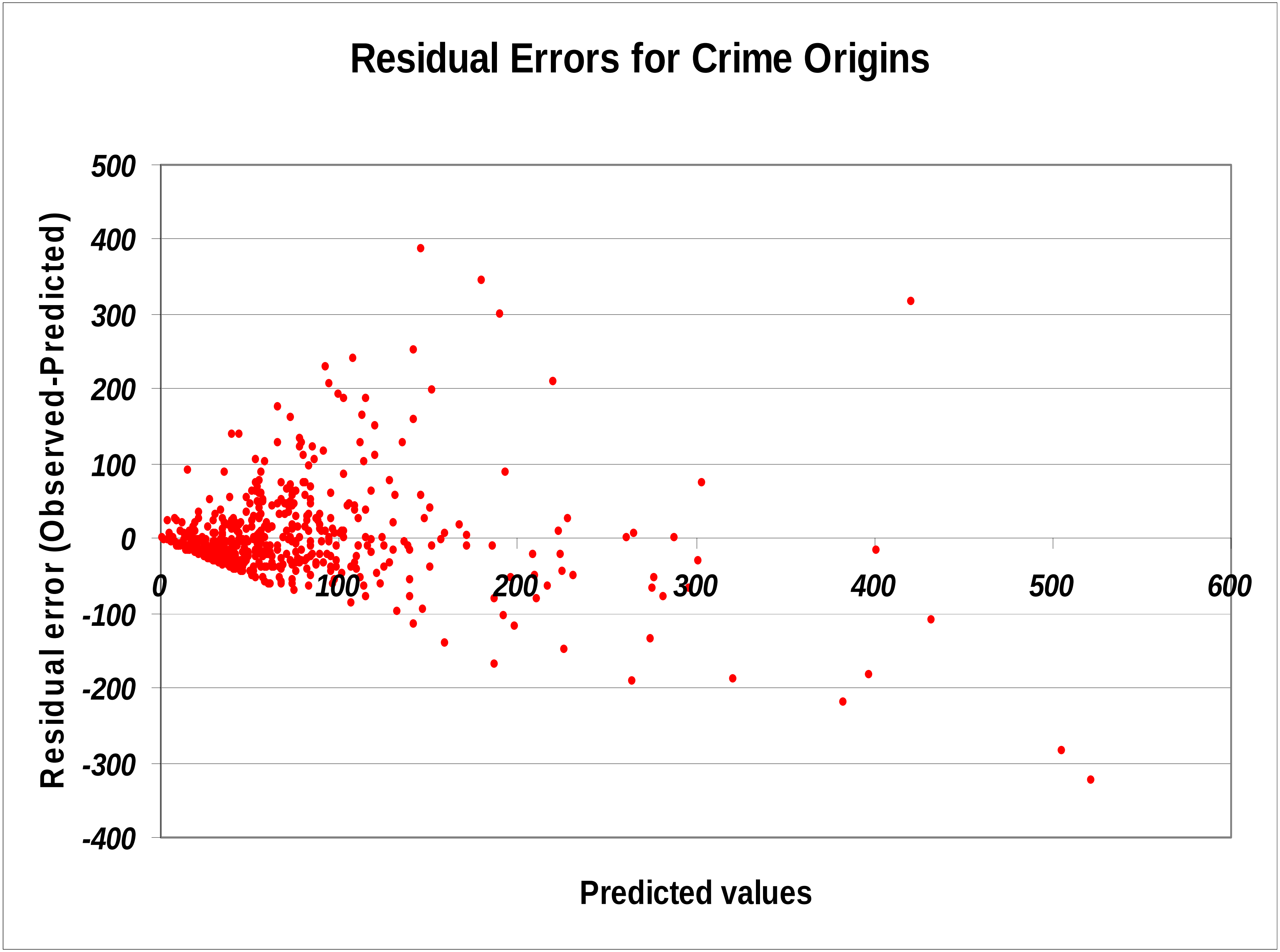
where

O(i)	Number of trips originating in zone i
D(j)	Number of trips attracted to zone j
X1, X2, ….XK	Independent variables
α, ϕ, β , $\acute{\eta}$	Constants and coefficients
ε	prediction error (residual)

Modeling Crime Origins

DepVar: Origins
N: 532
Df: 525
Type of regression model: Poisson with over-dispersion correction
R-square: 0.46

Predictor	Coefficient	Std. Error	Tolerance	z-value	p-value
CONSTANT	2.286699	0.39	.	58.13	0.001
INCOME EQUALITY	-0.018525	0.00	0.85	-18.05	0.001
NON-RETAIL EMPLOYMENT	-0.000186	0.00	0.87	-6.14	0.001
RETAIL EMPLOYMENT	-0.000353	0.00	0.96	-2.82	0.01
POPULATION	0.000284	0.00	0.94	22.47	0.001
BELTWAY	0.123109	0.04	0.97	3.25	0.01
MILES OF ARTERIAL	-0.085070	0.03	0.94	-3.15	0.01



Residual Errors for Crime Origins
500
400
300
200
100
0
-100
-200
-300
-400
Residual error (Observed-Predicted)
0
100
200
300
400
500
600
Predicted values

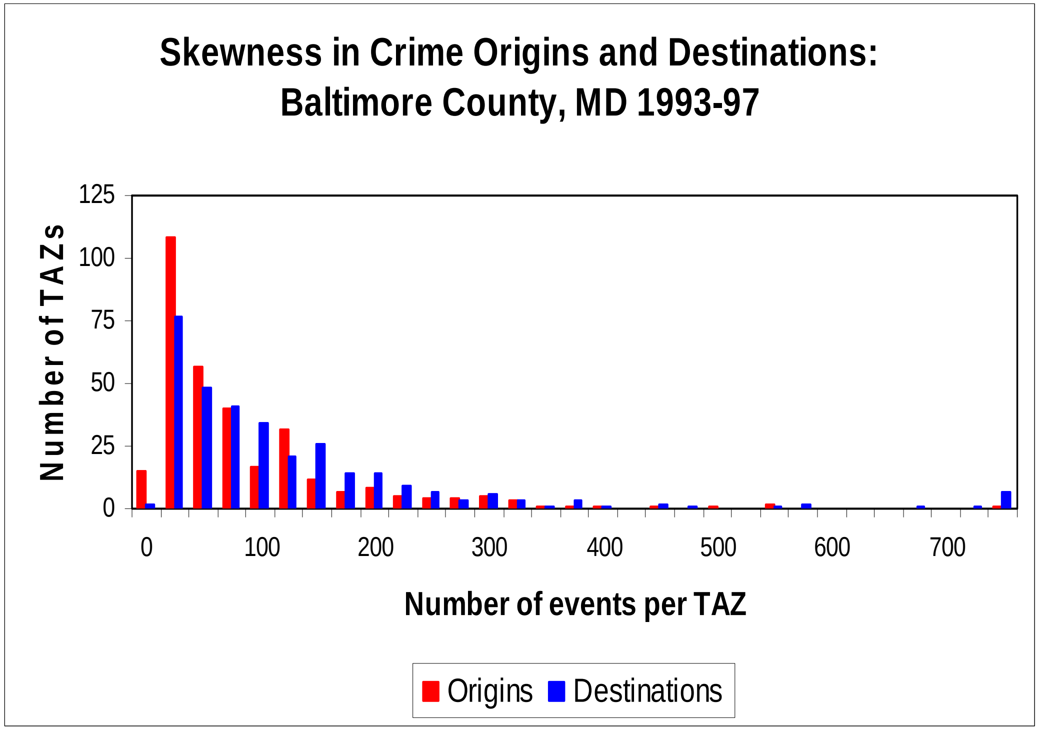
Output Will Have These Fields

Zone	PREDICTED
0401	107.462172
0402	127.595436
0403	78.436282
0404	56.770169
0405	85.439448
0406	58.045936
0407	67.910819
0408	44.514433
0409	56.670835
0410	38.383160
0411	45.080447
0412	43.655245
0413	39.926116
0414	77.641052
etc.	etc.

Problems with Poisson Regression

- In practice, doesn't produce zero estimates
- Assumes conditional variance = conditional mean. However, most real data is over-dispersed
  (conditional variance > conditional mean)
- Residual errors will be greater than expected
- Underestimates standard errors
- Therefore, overestimates significance
- Limited incorporation of spatial effects
  (deferred to the trip distribution stage)

Over-dispersion



Origins:

Mean = 75.8

Variance = 7848.8

Ratio of variance to mean = 14.7

Destinations:

Mean = 129.1

Variance = 51,849.1

Ratio of variance to mean = 401.5

Dispersion Parameter

Adjusts Standard Error for Skewness

$$\text{Dispersion} = \frac{1}{N-K-1} \sum \left\{ \frac{(Y_i - P_i)^2}{P_i} \right\}$$

where

Y(i)	*Observed number of trips from zone i (or to zone j)*
P(i)	*Predicted number of trips from zone i (or to zone j)*
N	*Sample size*
K	*Number of independent variables*

Over-dispersion Correction

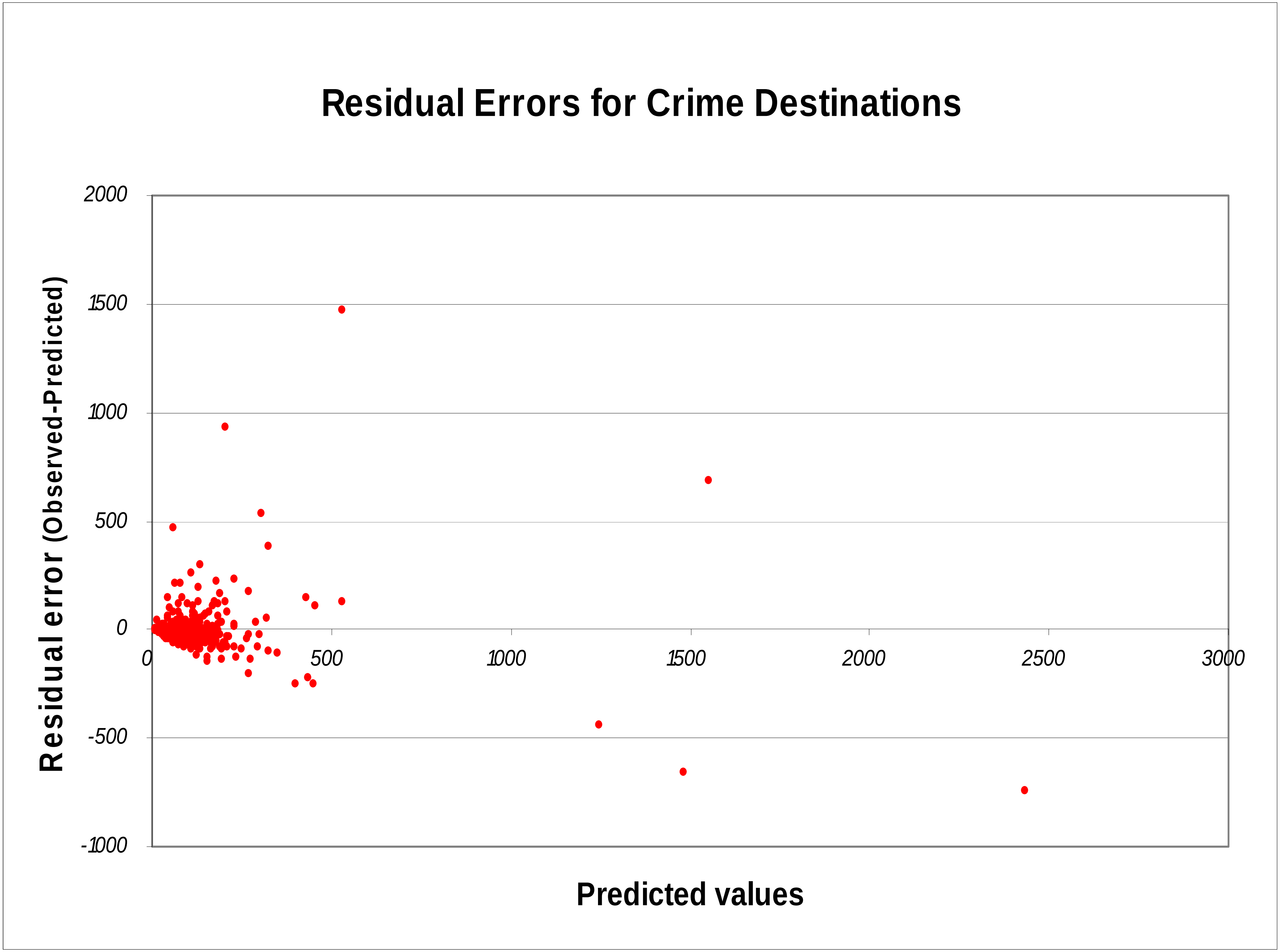
Standard errors are corrected by dispersion parameter

Adjusted Standard Error	**=**	**Poisson Standard Error**	**x**	**Square root of Dispersion Parameter**

Modeling Crime Destinations

DepVar:	Destinations
N:	325
Df:	319
Type of regression model:	Poisson with over-dispersion correction
R-square: 0.60	

Predictor	Coefficient	Std. Error	Tolerance	z-value	p-value
CONSTANT	5.485851	0.22	.	25.05	0.001
INCOME EQUALITY	-0.017176	0.01	0.90	-3.14	0.01
RETAIL EMPLOYMENT	0.001018	0.00	0.72	16.30	0.001
VERY LARGE MALL ACREAGE	0.006446	0.00	0.74	6.61	0.001
POPULATION	0.000190	0.00	0.93	6.93	0.001
DISTANCE FROM CBD	-0.115709	0.02	0.88	-6.78	0.001



Residual Errors for Crime Destinations
Residual error (Observed-Predicted)
2000
1500
1000
500
0
-500
-1000
0
500
1000
1500
2000
2500
3000
Predicted values

Models for Specific Crime Types

Origin Model

	All Crimes	Robbery	Burglary	Vehicle Theft
CONSTANT	2.286699	-0.652291	1.621546	-0.800759
INCOME EQUALITY	-0.018525	-0.023964	-	-0.019620
NON-RETAIL EMPLOYMENT	-0.000186	-0.000237	-0.000239	-0.000188
RETAIL EMPLOYMENT	-0.000353	-	-	
POPULATION	0.000284	0.000297	0.000242	0.000342
BELTWAY	0.123109	-	-	
MILES OF ARTERIAL	-0.085070	-	-	-0.180966

Models for Specific Crime Types

Destination Model

	All Crimes	Robbery	Burglary	Vehicle Theft
CONSTANT	5.485851	3.284488	3.246183	2.610299
INCOME EQUALITY	-0.017176	-0.027946	-0.034598	-0.012910
RETAIL EMPLOYMENT	0.001018	0.000844	-	0.000507
VERY LARGE MALL ACREAGE	0.006446	0.004332	-	-
POPULATION	0.000190	0.000223	0.000309	0.000247
DISTANCE FROM CBD	-0.115709	-0.096330	-0.038715	-0.096088

Adding Special Generators

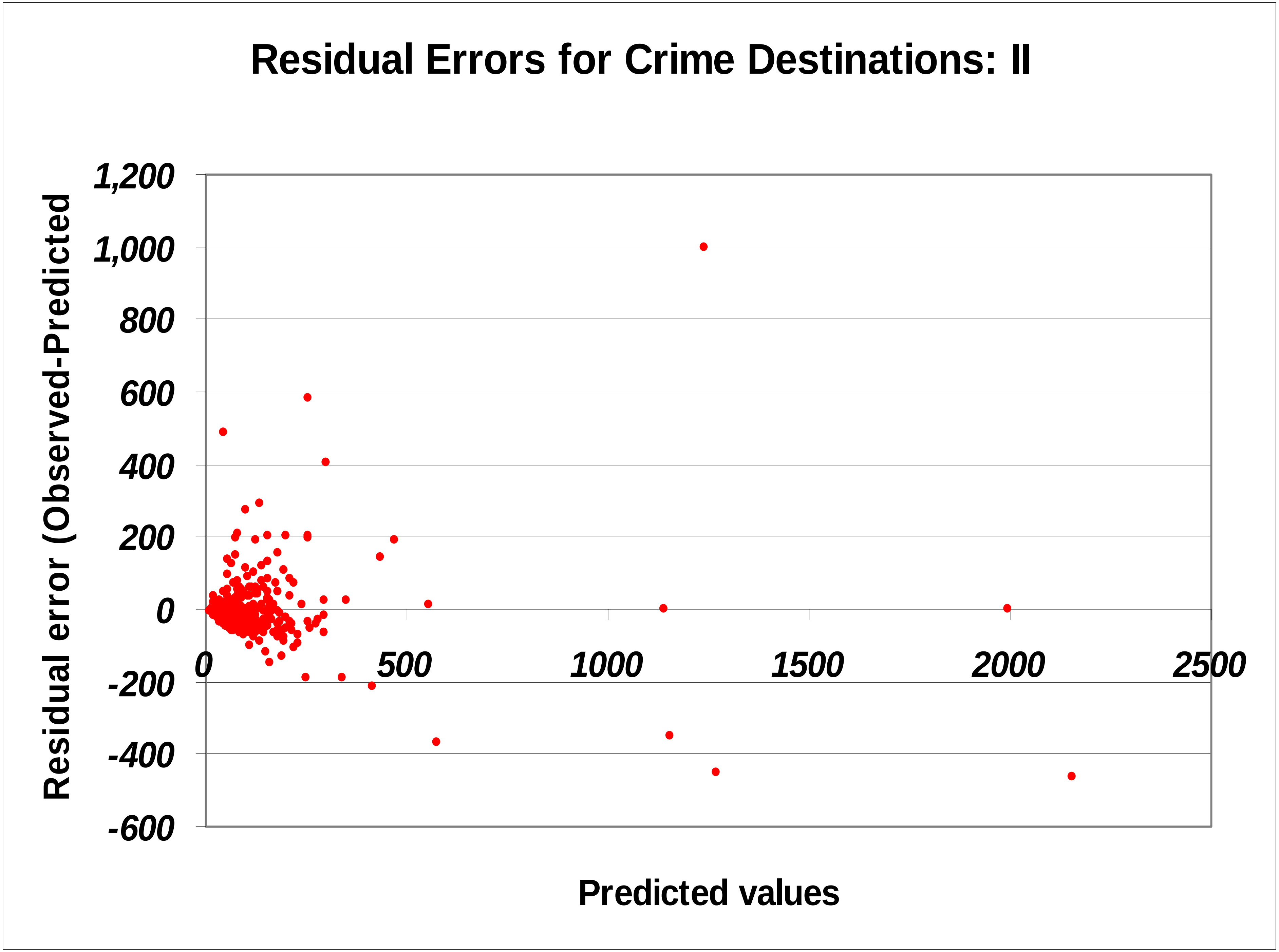
Create Two Dummy Variables

- **TAZ 0675** **(Golden Ring mall)**
- **TAZ 0716** **(Eastpoint mall)**

Modeling Crime Destinations: II

DepVar: Destinations
N: 325
Df: 317
Type of regression model: Poisson with over-dispersion correction
R-square: 0.784194

Predictor	Coefficient	Std. Error	Tolerance	z-value	p-value
CONSTANT	5.182117	0.07	.	76.36	0.001
INCOME EQUALITY	-0.020797	0.00	0.90	-5.28	0.001
RETAIL EMPLOYMENT	0.000995	0.00	0.70	16.34	0.001
VERY LARGE MALL ACREAGE	0.006590	0.00	0.72	7.58	0.001
POPULATION	0.000238	0.00	0.92	12.16	0.001
DISTANCE FROM CBD	-0.087826	0.01	0.87	-7.05	0.001
TAZ 0675	1.933321	0.07	0.97	27.76	0.001
TAZ 0716	1.602000	0.07	0.94	23.58	0.001



Residual Errors for Crime Destinations: II
Residual error (Observed-Predicted
1,200
1,000
800
600
400
200
0
-200
-400
-600
0
500
1000
1500
2000
2500
Predicted values

Balancing Predicted Origins and Destinations

For 'trips':
Number of origins = Number of destinations

However, some trips come from outside the study area:

Number of crimes ending in 325 Baltimore County zones:	**41,969**
Number of crime originating in 532 Baltimore County/City zones:	**40,342**
Crimes from outside the study area: (external trips)	**1,627** (3.9%)

External Trips by Crime Type

% of Crimes by Destination

Vehicle Theft:	**39 (1.4%)**
Robbery:	**155 (4.0%)**
Burglary:	**215 (4.5%)**

Balancing Calculations

Number of crime origins + Number of crimes from outside area
= Number of crime destinations

This is necessary for the distribution stage

CrimeStat allows external crime 'trips' to be added and *ensures* that origins and destinations are balanced

Two options:
- Hold destinations constant
- Hold origins constant

Adjusted Data Should Have These Fields

Zone	PREDICTED	ADJORIGIN
0001	225.818482	225.850955
0002	187.527819	187.554785
0003	320.877458	320.923600
0004	75.096631	75.107430
0005	44.981775	44.988243
0006	32.574758	32.579442
0007	107.334835	107.350270
0008	74.683931	74.694671
0009	76.425236	76.436226
0010	34.183846	34.188762
0011	66.975803	66.985434
etc	etc	etc

Models are NOT Behavioral

Predictive, not explanatory

Measure conditions associated with crime

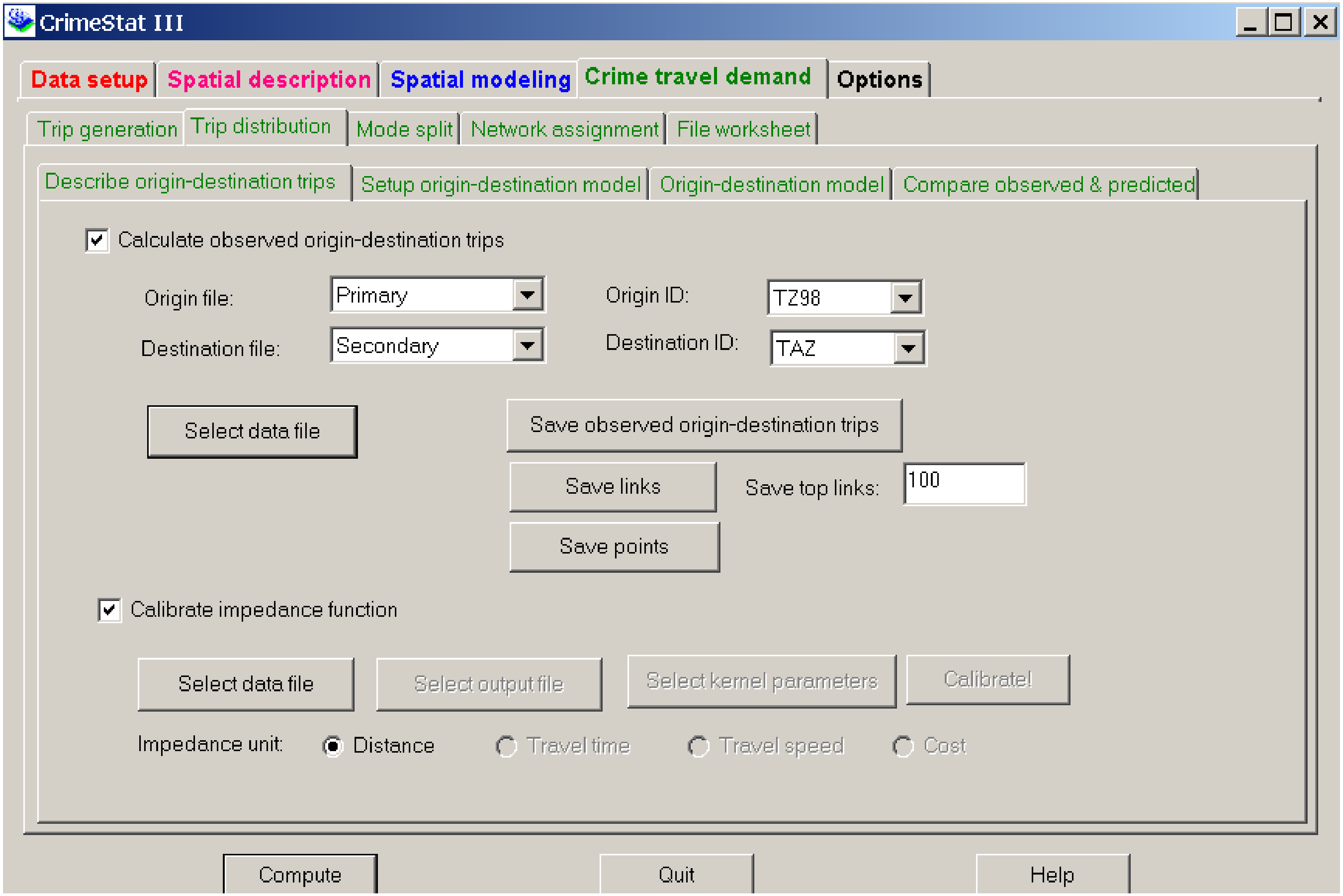
The conditions should be *real* in that they produce an estimate that is stable over time

Poisson is Not the Only Regression Model That Could be Used for Zonal Predictions

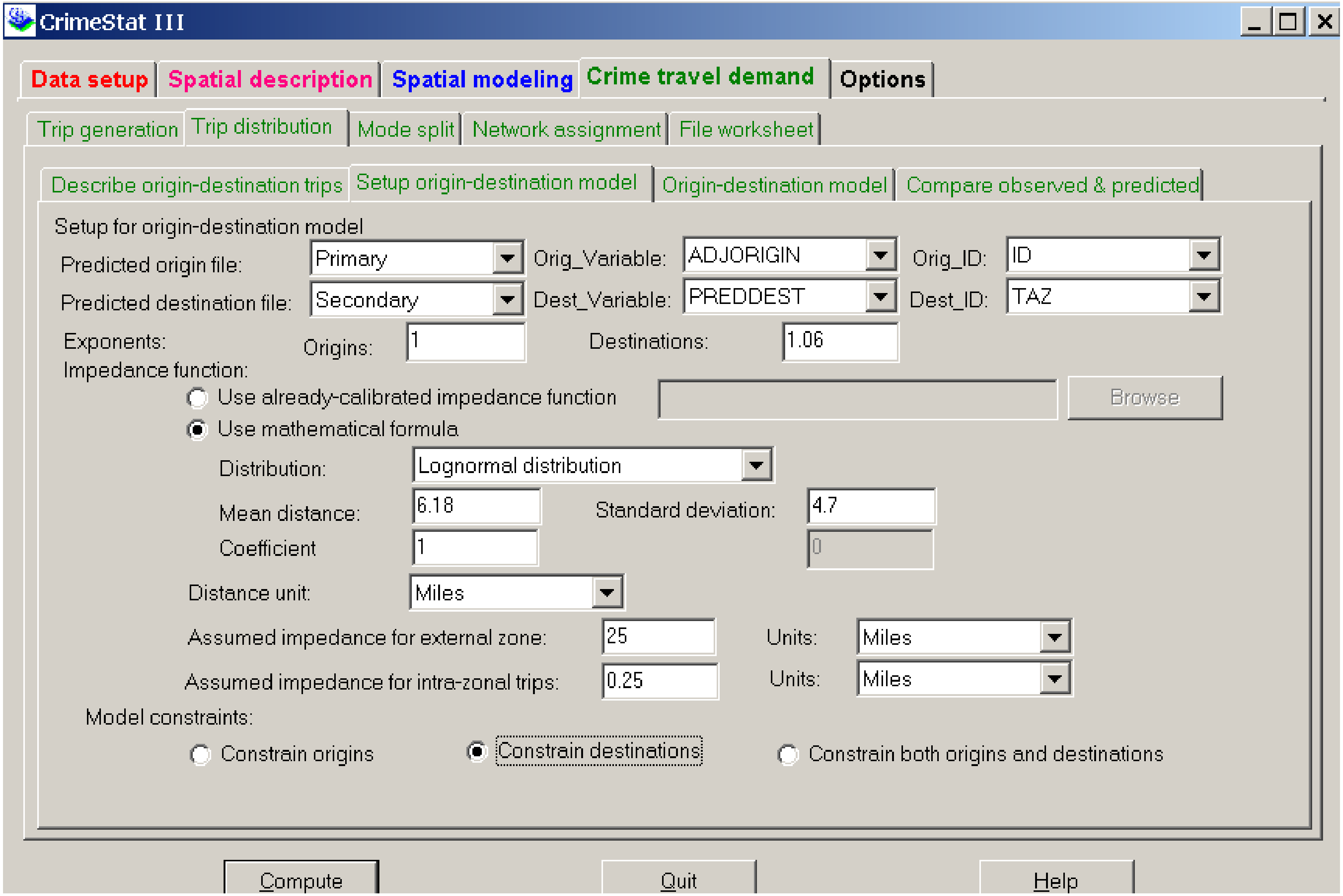
- Negative binomial and other compound Poisson models
- Zero-inflated Poisson
- Geographically-weighted regression using Poisson
- Weibull, lognormal, gamma and other skewed distributions
- Hierarchical (empirical) Bayes
- Markov Chain Monte Carlo
- Discrete choice, disaggregate models

Crime Trip Distribution

Trip Distribution Module



Trip Distribution Model Setup



Crime trip distribution analysis:

- From each origin zone

to

- Each destination zone

Crime Origin-Destination Matrix

Crime destination zone

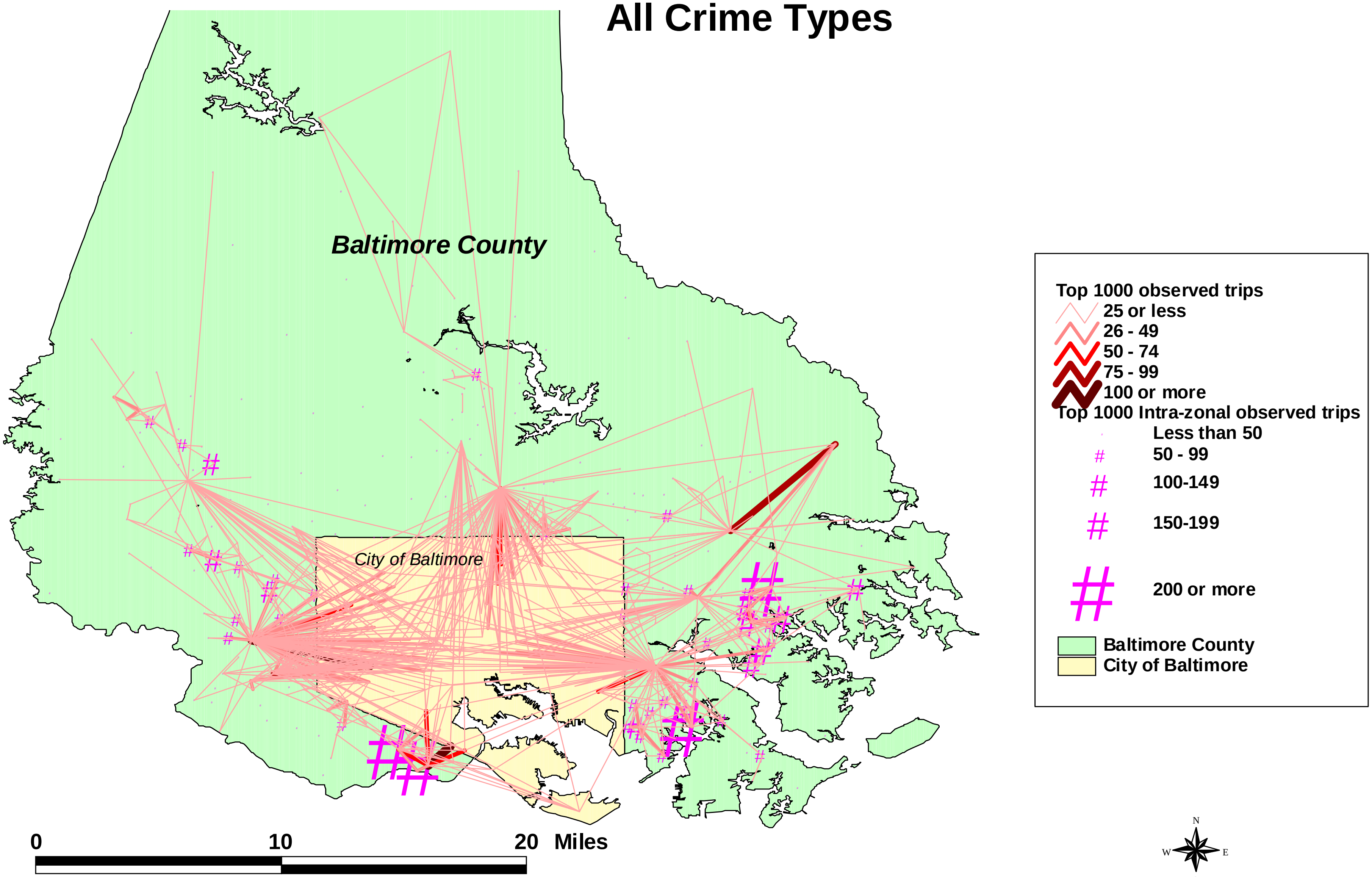
Crime origin zone

	1	2	3	4	5		N	Σ
1	**37**	15	21	4	3	. . . . . .	12	346
2	7	**53**	14	0	4	. . . . . .	15	1050
3	12	9	**81**	7	6	. . . . . .	33	711
4	4	10	6	**12**	1	. . . . . .	0	84
5	8	7	28	2	**24**	. . . . . .	14	178
	...	...	...	...	...		...	
M	12	5	43	3	10	. . . . . .	**92**	1466
Σ	153	276	1245	99	110		812	43,240

Two Different Distribution Matrices

- Actual (Observed)
- Modeled (Predicted)

Observed Baltimore County Crime Trips: 1993-1997
Top 1000 Links
All Crime Types



Classic Gravity Model

$$A(ij) = g \frac{M(i)\,M(j)}{d^2}$$

where

A(ij)	*Attraction of object i for object j*
M(i)	*Mass of object i*
M(j)	*Mass of object j*
d	*Distance between objects (cost or impedance)*
g	*Constant (gravitational)*

Spatial Interaction Model

$$A(ij) = \alpha \, P(i)\, P(j)\, I(ij)$$

where

A(ij)	*Attraction of object i for object j*
P(i)	*Size of area i*
P(j)	*Size of area j*
I(ij)	*Utility between objects (inverse of ‘cost’)*
α	*Constant*

Modeling Crime Trip Distribution

$$T(ij) = \alpha * P(i)^{\lambda} * \beta * A(j)^{\rho} * I(ij) + e$$

where

T(ij)	***Predicted trips from zone i to zone j***
P(i)	***Predicted trips produced by zone I (origins)***
A(j)	***Predicted trips attracted to zone j (destinations)***
I(ij)	***Impedance function for travel between zone i and j***
α, β	***Constants*** *(designed to ensure that total productions equal total attractions)*
λ, ρ	***Exponents*** *('fine tuning' adjustments)*
e	***Error term***

Estimating the Cost of Crime Travel

Estimating travel cost

- **By travel distance**
- **By travel time**
- **By generalized costs**
  (e.g., fuel, parking, waiting time, transfer time)

Varies by type of crime, time of day, and other factors

Estimating the Equation

Solution is iterative

- Constrain destinations
- Constrain origins
- Double-constraint

Constrain origins

$$T(ij) = \alpha_i * P(i)^{\lambda} * A(j)^{\rho} * I(ij)$$

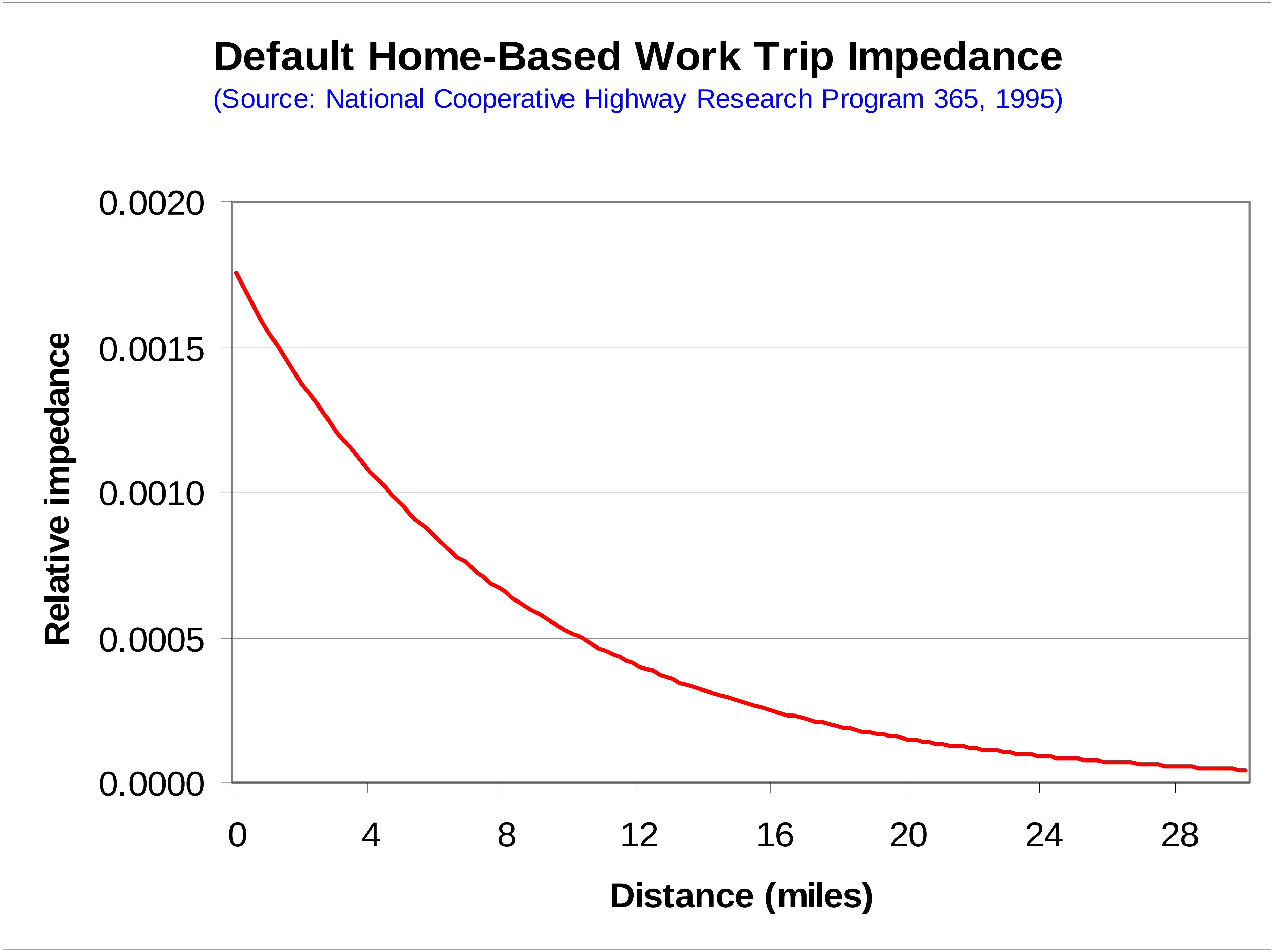
Constrain destinations

$$T(ij) = P(i)^{\lambda} * \beta_j * A(j)^{\rho} * I(ij)$$

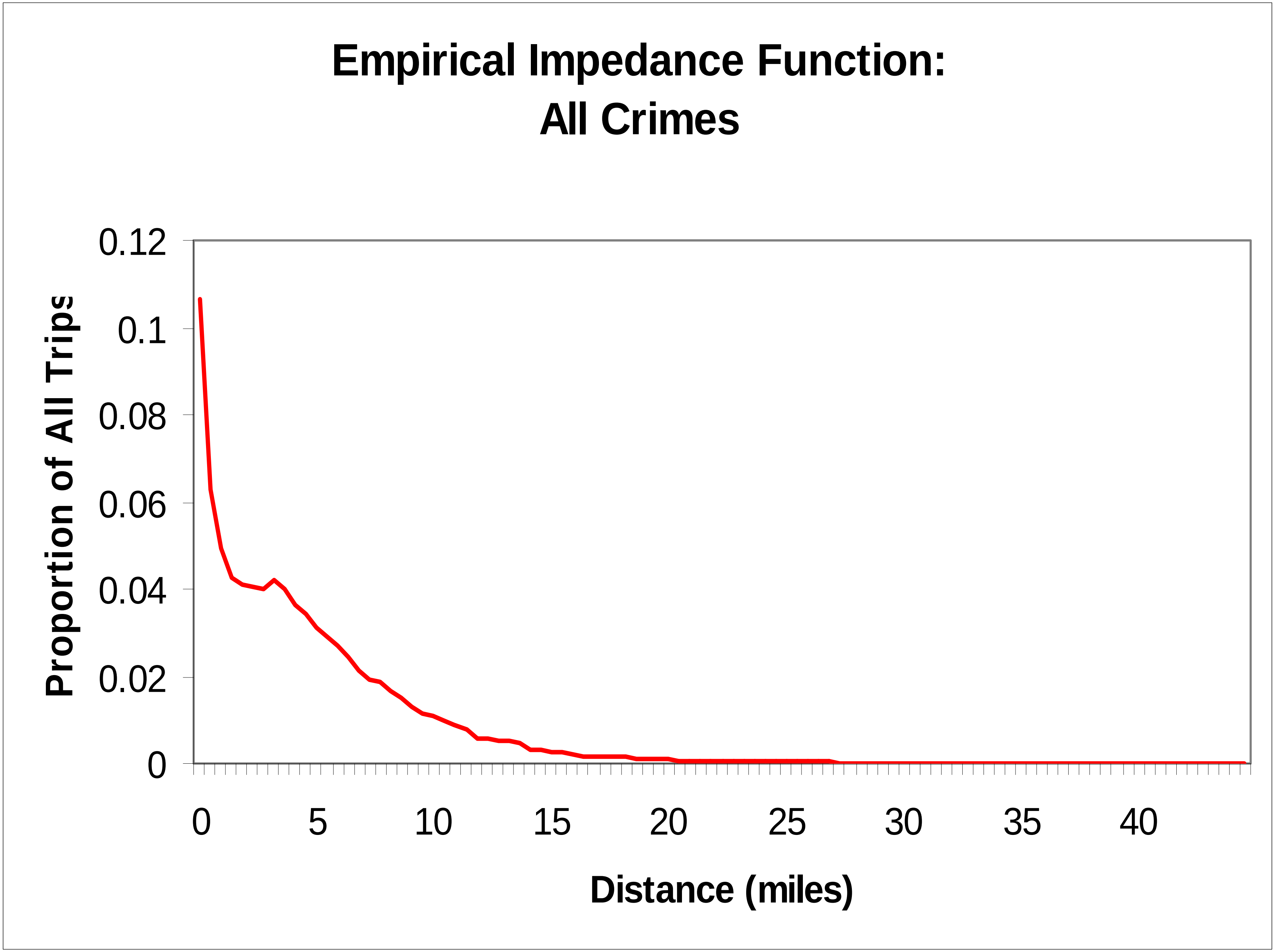
Double-constraint

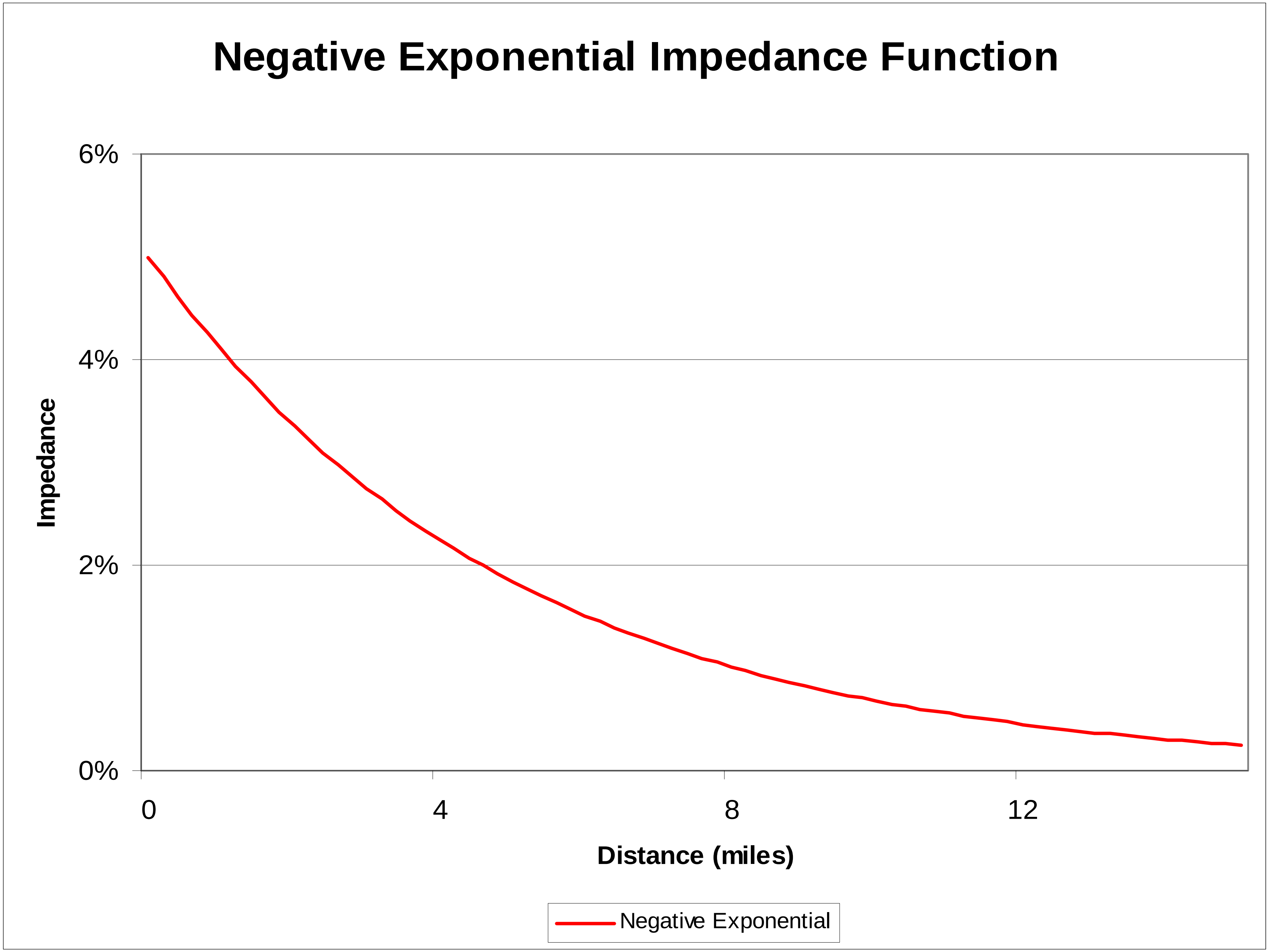
$$T(ij) = \alpha_i * P(i)^{\lambda} * \beta_j * A(j)^{\rho} * I(ij)$$

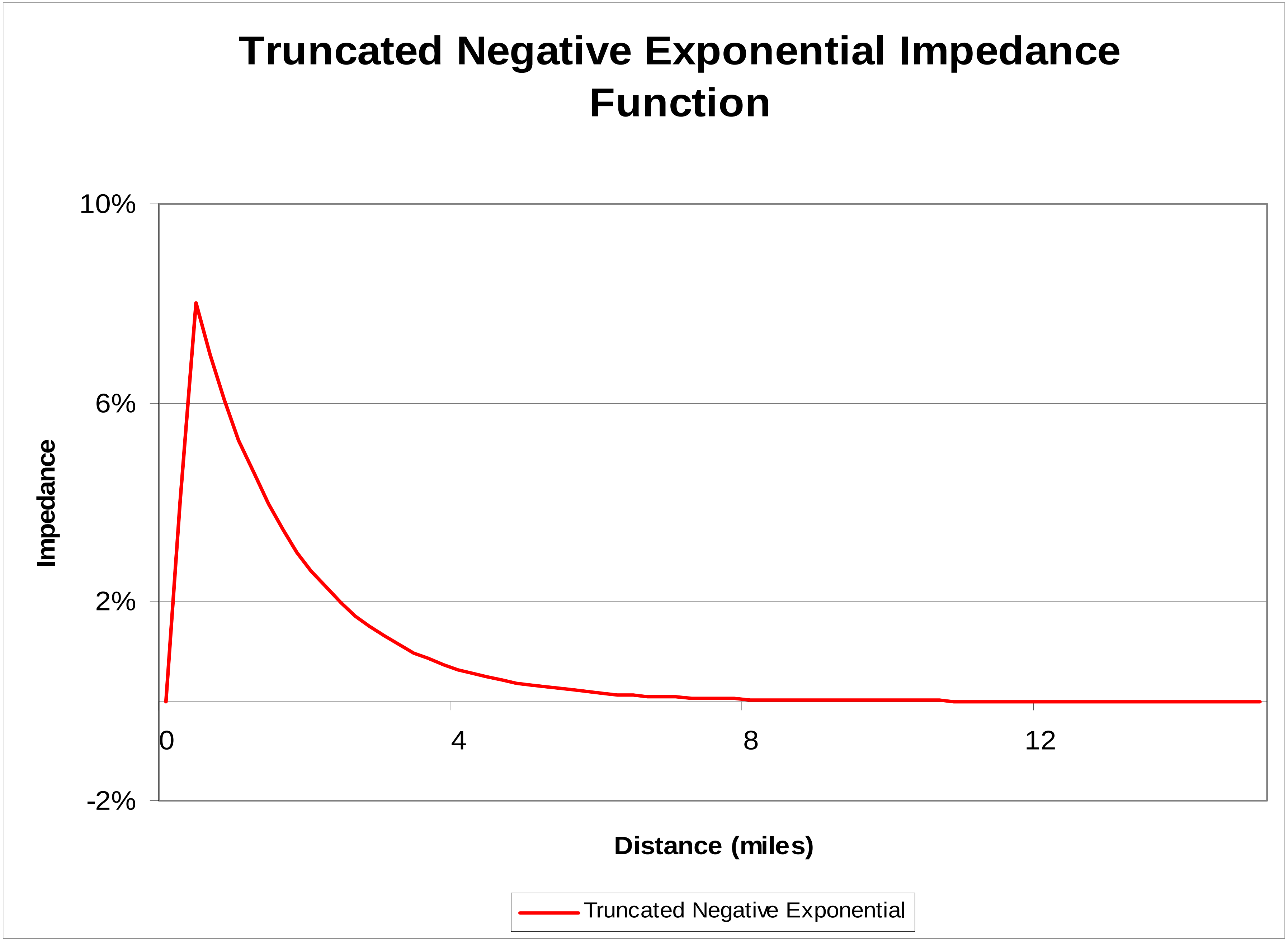
Measuring Impedance

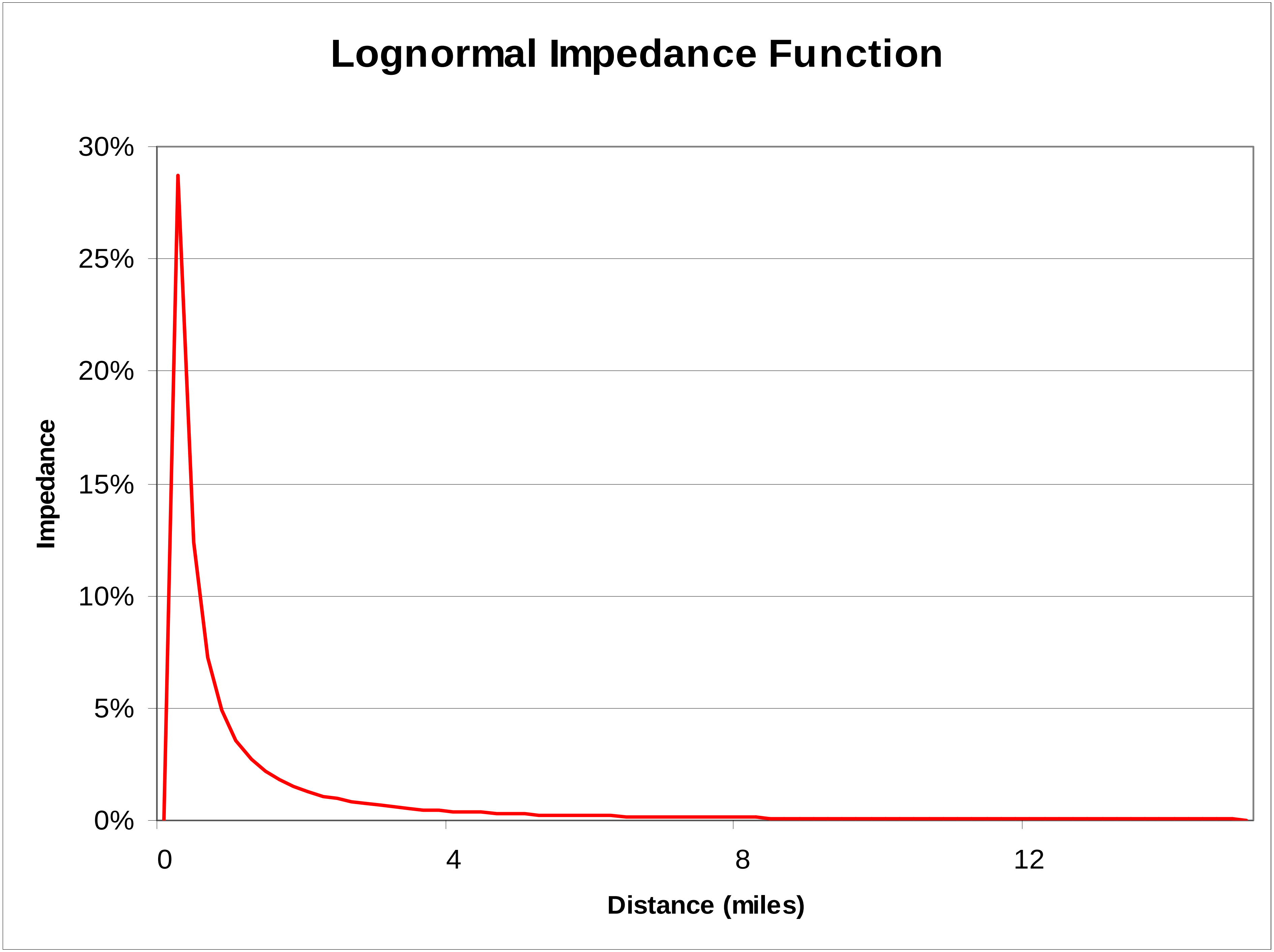


Default Home-Based Work Trip Impedance
(Source: National Cooperative Highway Research Program 365, 1995)
Relative impedance
0.0020
0.0015
0.0010
0.0005
0.0000
0
4
8
12
16
20
24
28
Distance (miles)

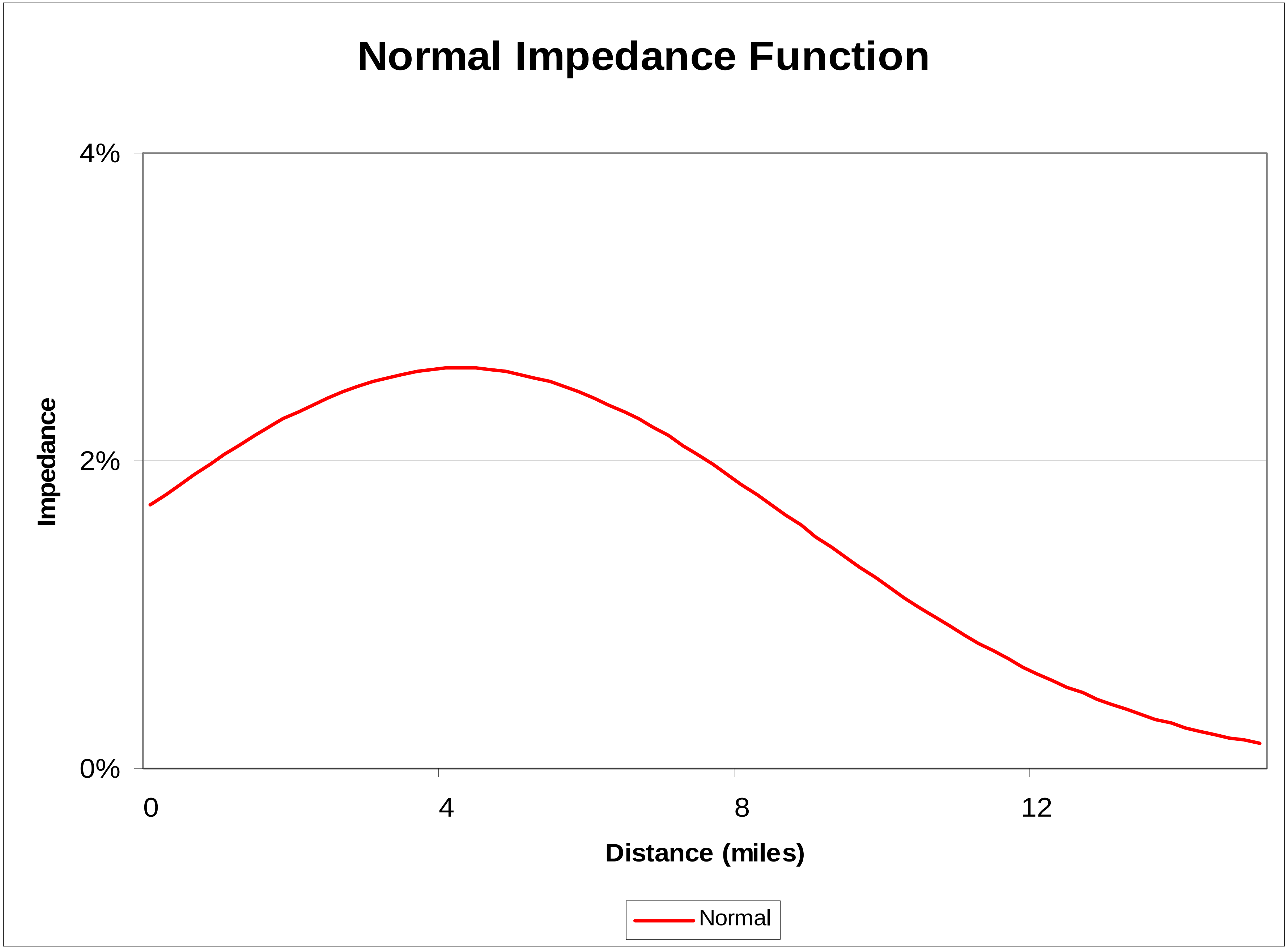

Empirical Impedance Function:
All Crimes
Proportion of All Trips
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40
Distance (miles)


Negative Exponential Impedance Function
6%
4%
2%
0%
Impedance
0
4
8
12
Distance (miles)
Negative Exponential

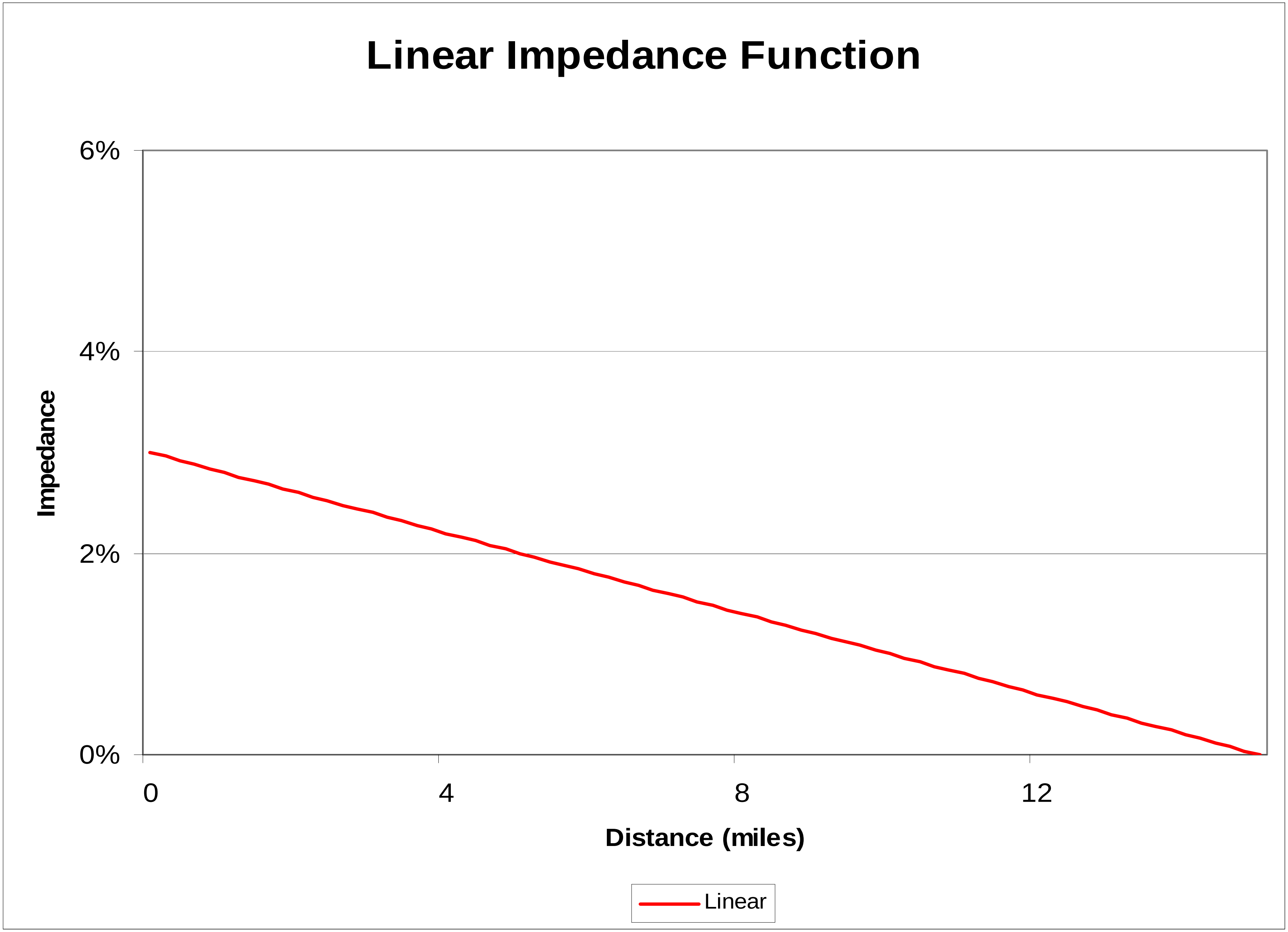

Truncated Negative Exponential Impedance Function
10%
6%
2%
-2%
Impedance
0
4
8
12
Distance (miles)
Truncated Negative Exponential



Lognormal Impedance Function
30%
25%
20%
15%
10%
5%
0%
Impedance
0
4
8
12
Distance (miles)



Normal Impedance Function
4%
2%
0%
Impedance
0
4
8
12
Distance (miles)
Normal


Linear Impedance Function
6%
4%
2%
0%
Impedance
0
4
8
12
Distance (miles)
Linear

Impedance Function is Fit Iteratively

- **Mathematical function selected**
  - **e.g., with Lognormal there are 3 parameters**
    - ***Mean distance***
    - ***Standard deviation of distance***
    - ***Coefficient***
  - **Exponents of origins & destinations (all models)**
- **Best mean distance is found**
- **Best standard deviation of distance is found**
- **Best coefficient is found**
- **Exponents are adjusted ("fine tuning")**

Empirical Tests of Fit

- **Cell comparisons**

  Can't use chi-square; too many zero cells

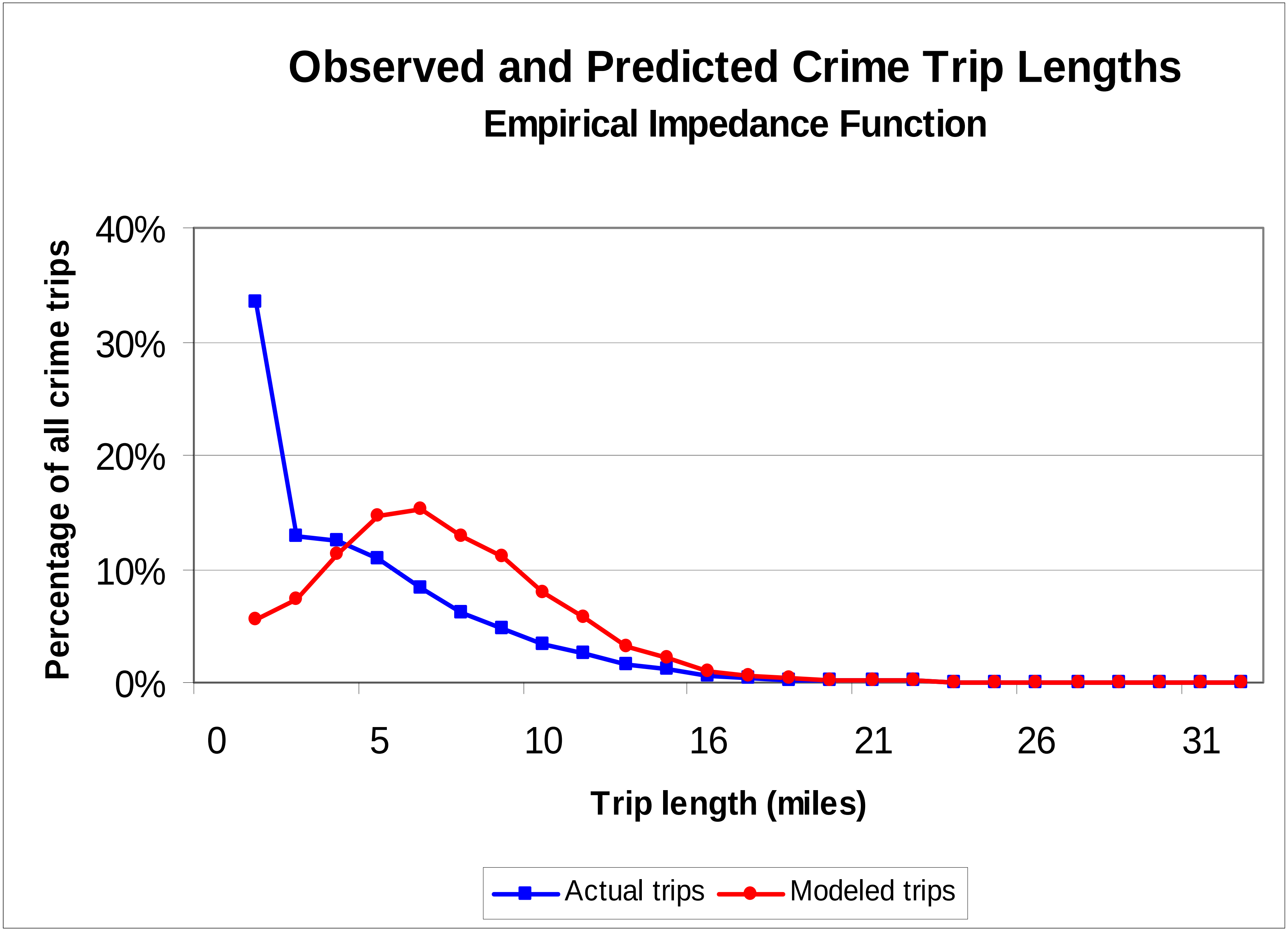
- **Inter-zonal and Intra-zonal trip distribution**

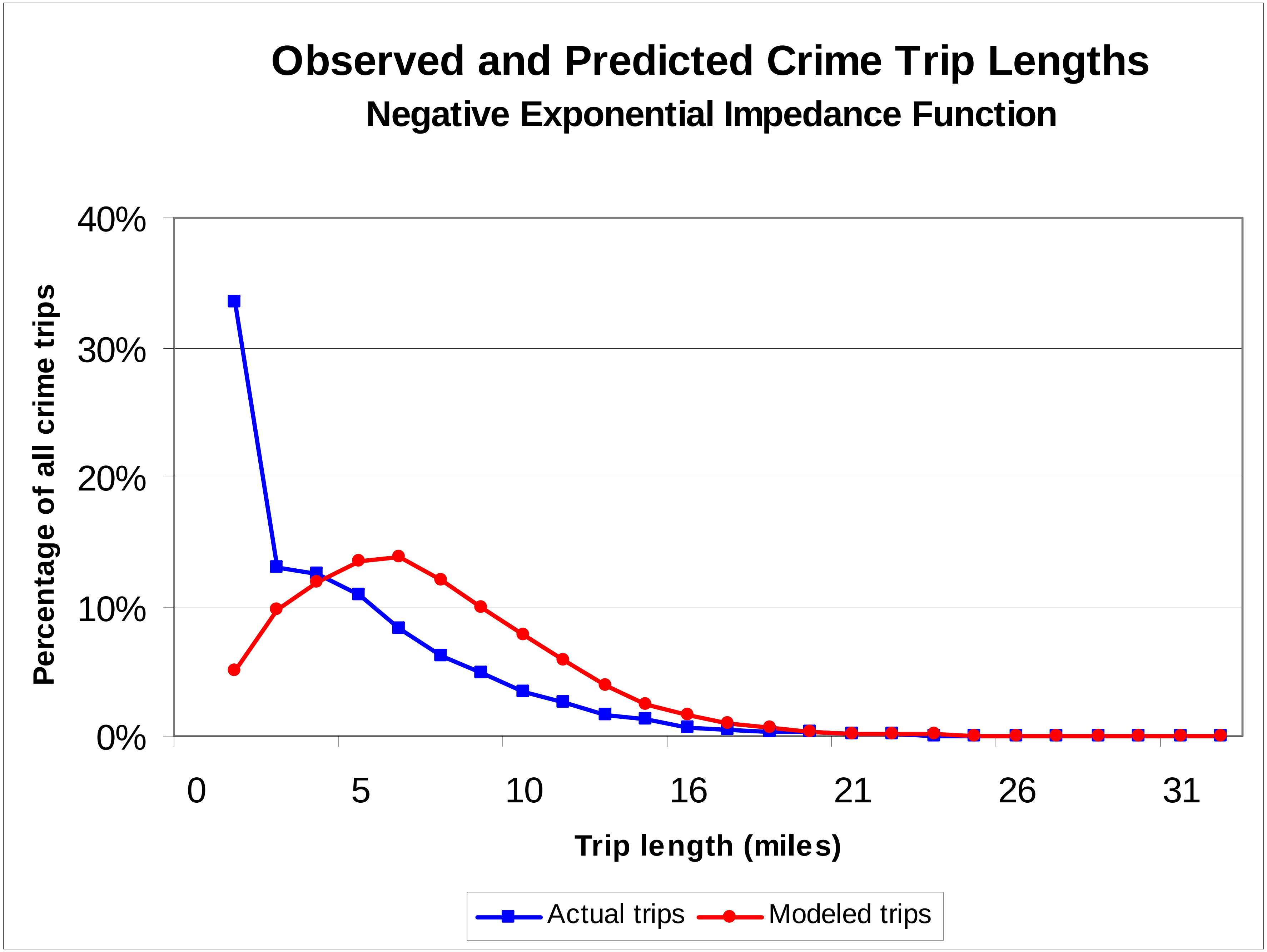
- **Trip length distributions**

  Coincidence ratio (0 to 1 with 0.9+ being ideal)
  Komolgorov-Smirnov two-sample test

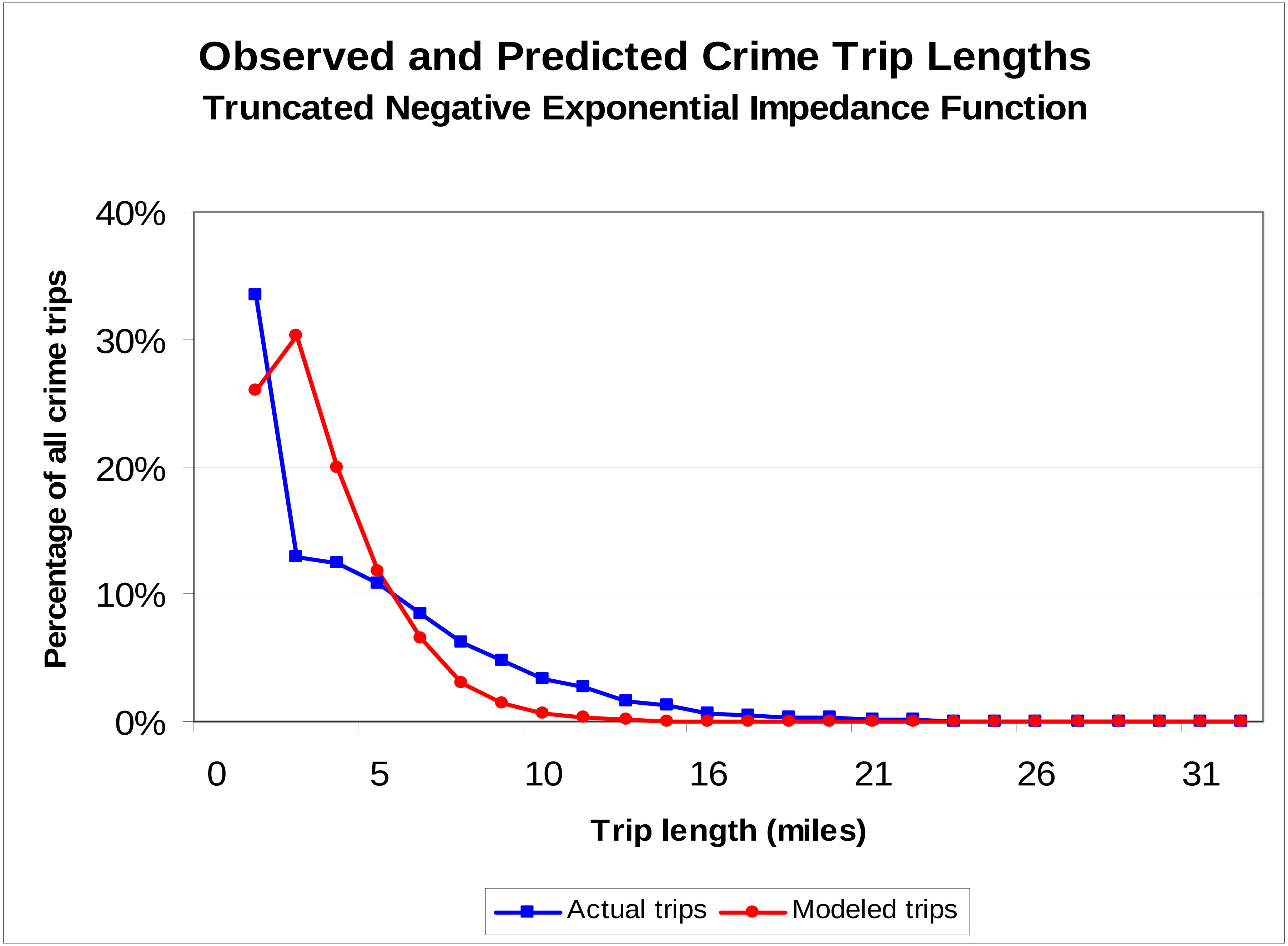
- **Predict major links**

  Replicate highest volume trip links

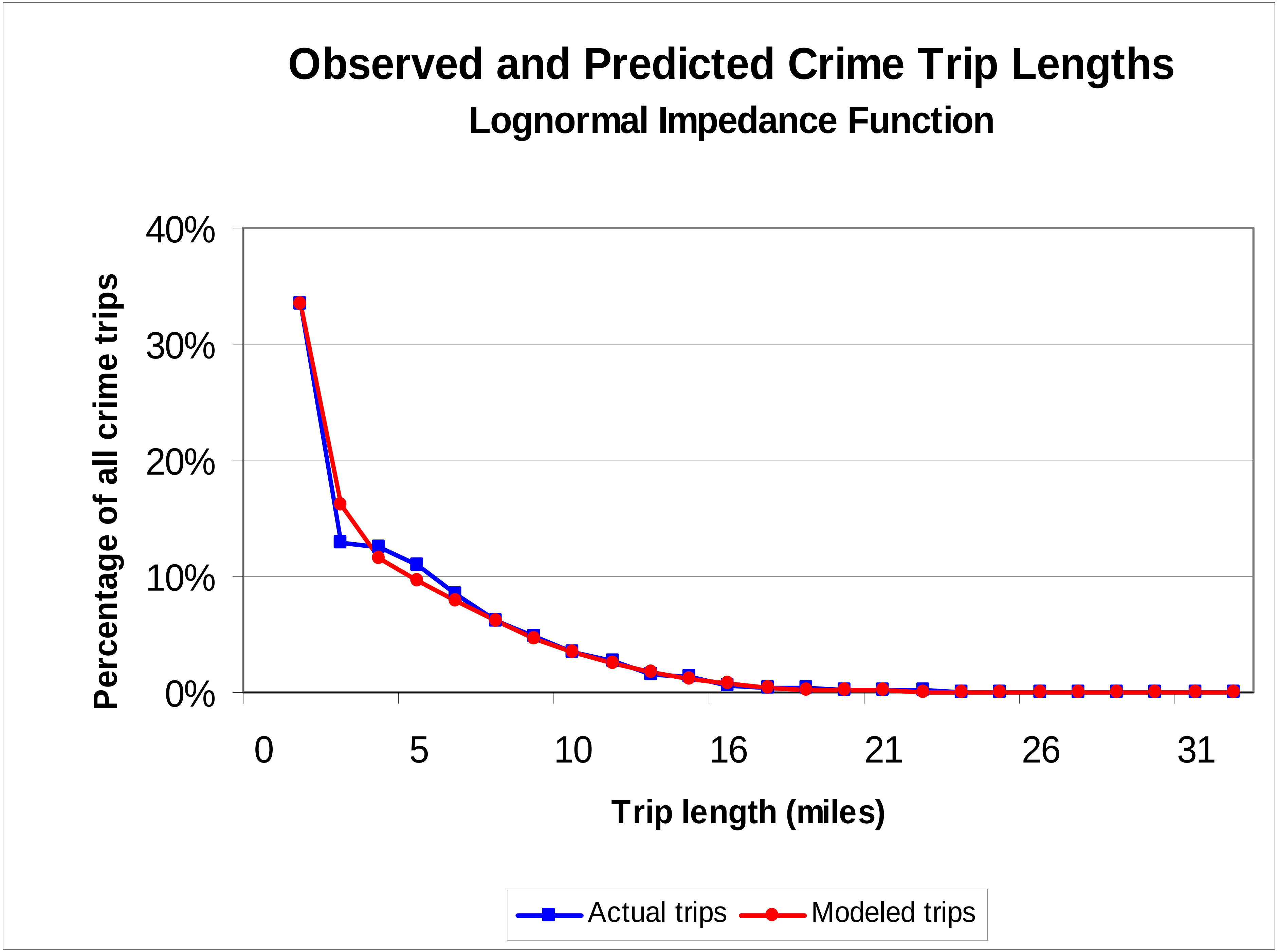

Observed and Predicted Crime Trip Lengths
Empirical Impedance Function
Percentage of all crime trips
40%
30%
20%
10%
0%
0
5
10
16
21
26
31
Trip length (miles)
Actual trips
Modeled trips



Observed and Predicted Crime Trip Lengths
Negative Exponential Impedance Function
Percentage of all crime trips
40%
30%
20%
10%
0%
0
5
10
16
21
26
31
Trip length (miles)
Actual trips
Modeled trips



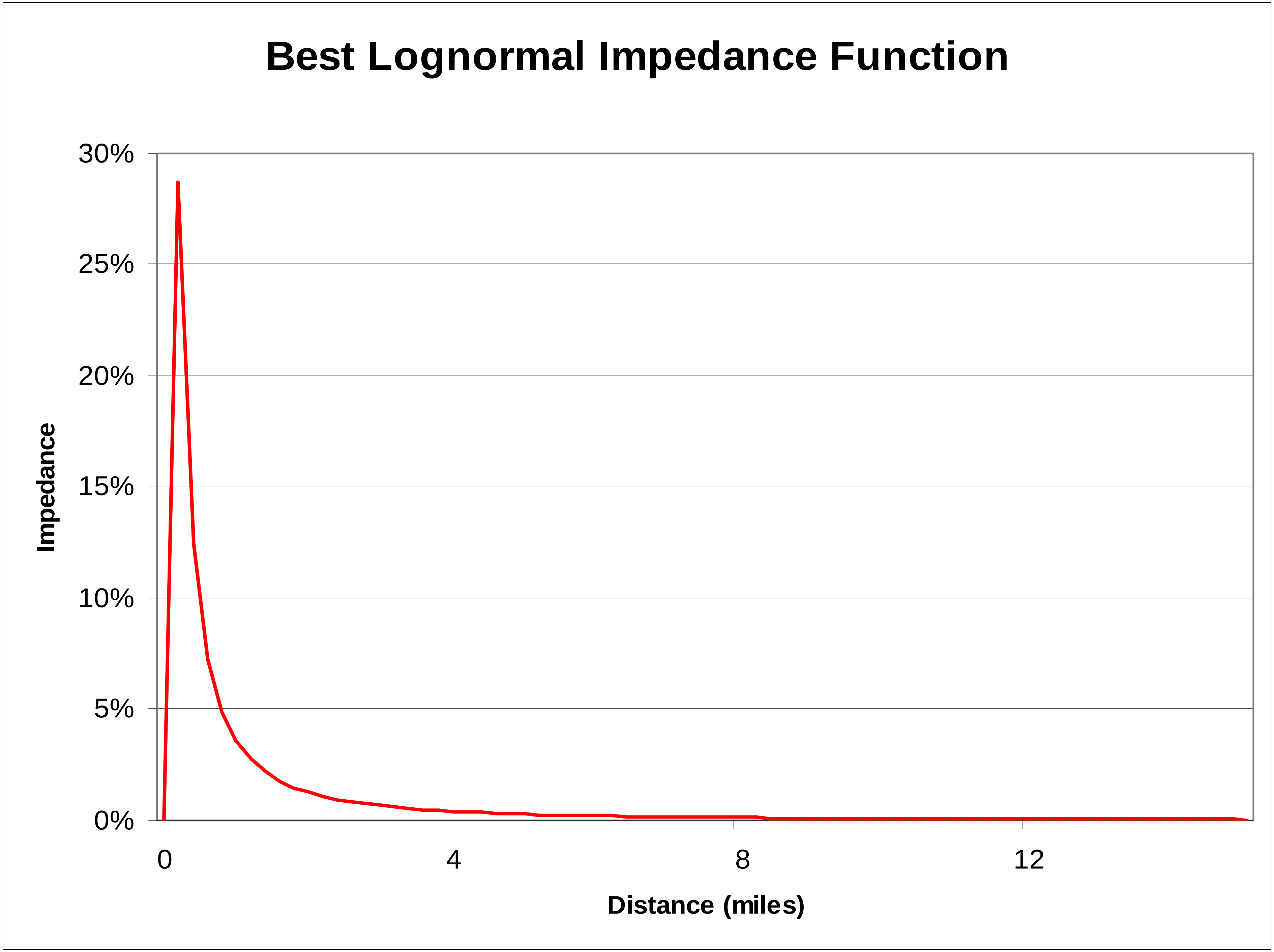
Observed and Predicted Crime Trip Lengths
Truncated Negative Exponential Impedance Function
Percentage of all crime trips
40%
30%
20%
10%
0%
0
5
10
16
21
26
31
Trip length (miles)
Actual trips
Modeled trips


Observed and Predicted Crime Trip Lengths
Lognormal Impedance Function
Percentage of all crime trips
40%
30%
20%
10%
0%
0
5
10
16
21
26
31
Trip length (miles)
Actual trips
Modeled trips

Best Model: All Crime Types

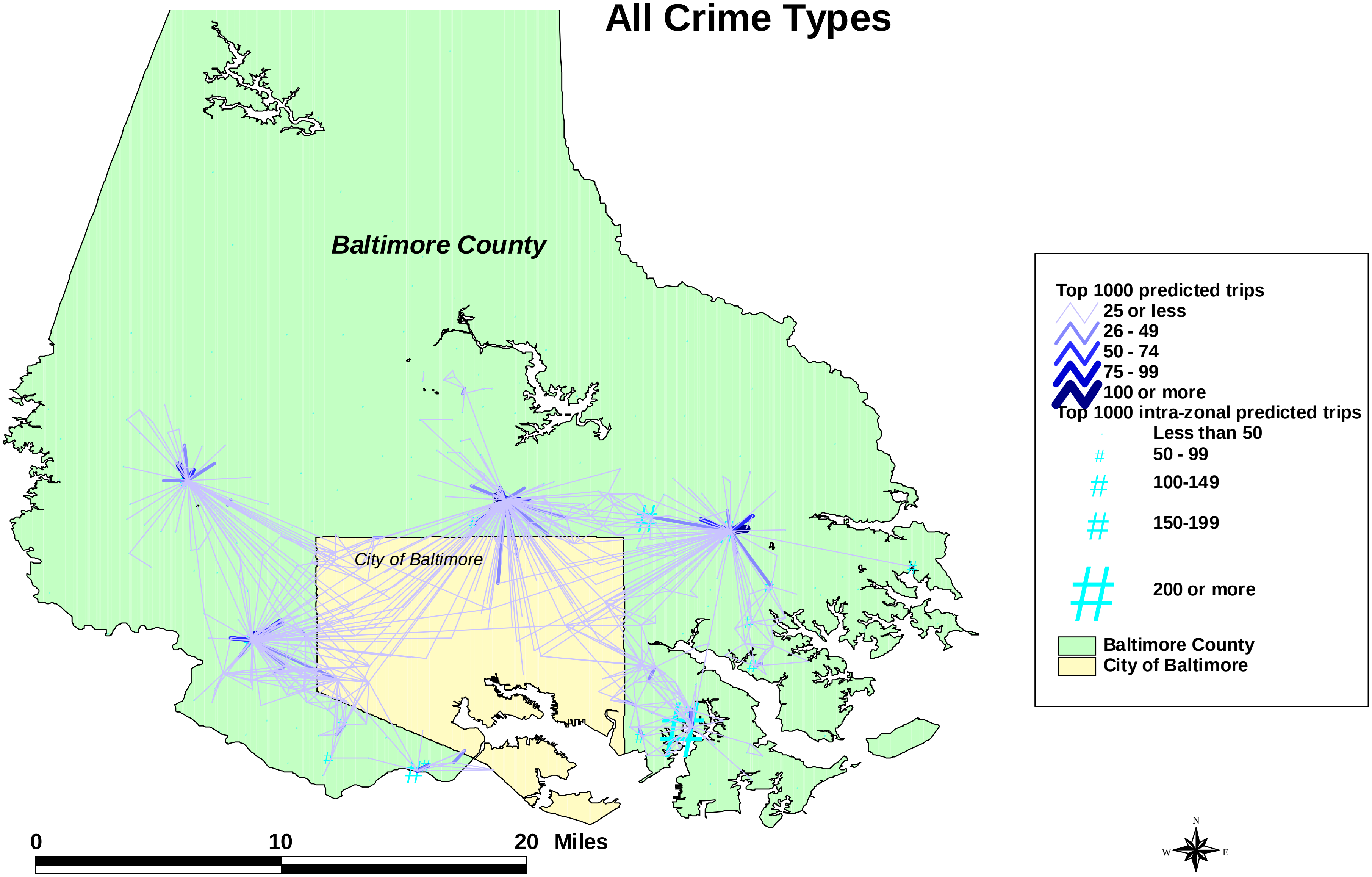
Log-normal function:

- **Mean distance =** ***6.18 miles***
- **Standard deviation =** ***4.7 miles***
- **Coefficient =** ***1***
- **Exponents:** ***Origins = 1.0 Destinations = 1.06***
- **Congruity tests:**
  Average observed trip length = 4.76 miles
  Average Predicted trip length = 4.62 miles
  Median observed trip length = 1.70
  Median predicted trip length = 1.36
  Coincidence ratio = 0.91

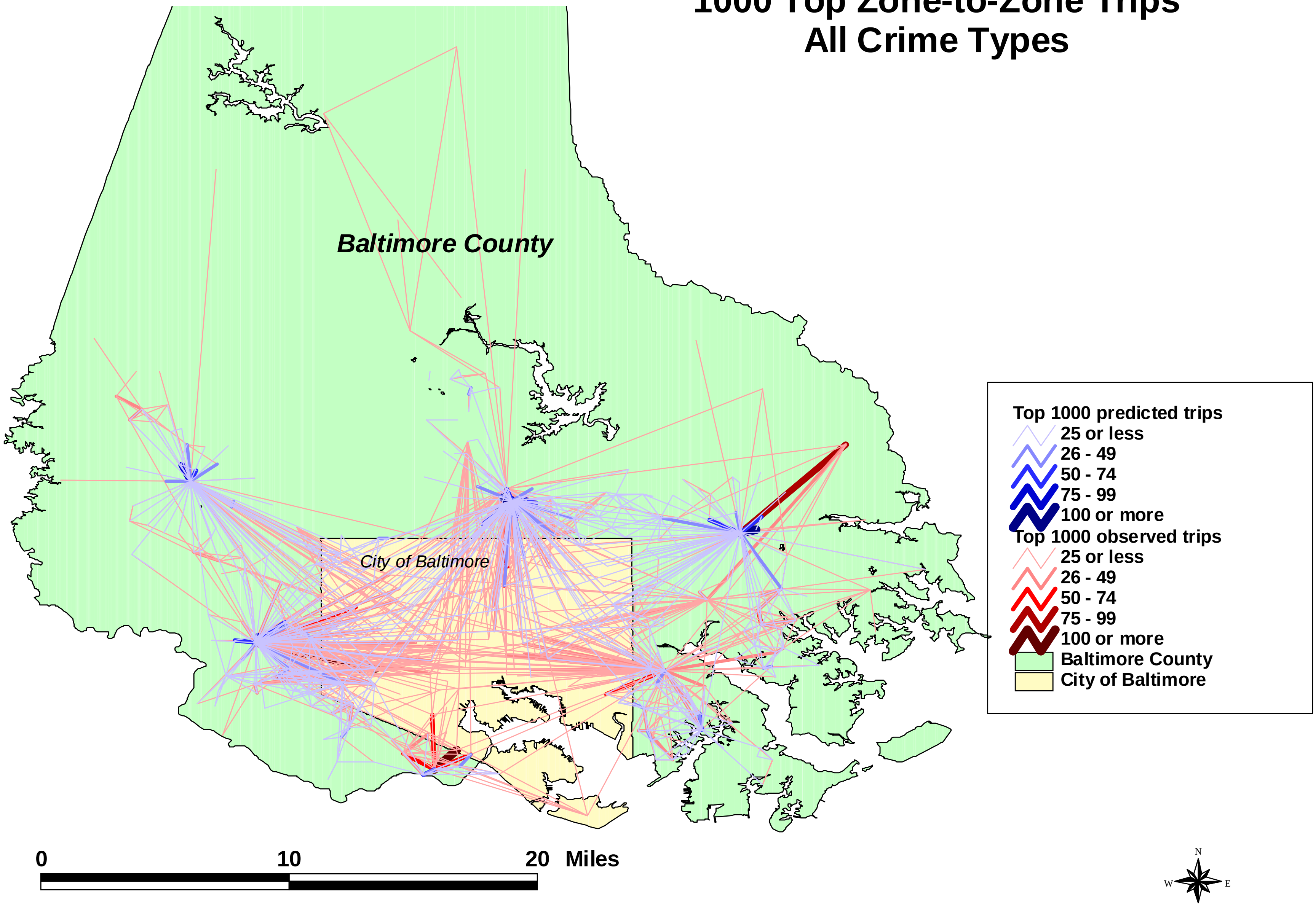


Best Lognormal Impedance Function
30%
25%
20%
15%
10%
5%
0%
Impedance
0
4
8
12
Distance (miles)

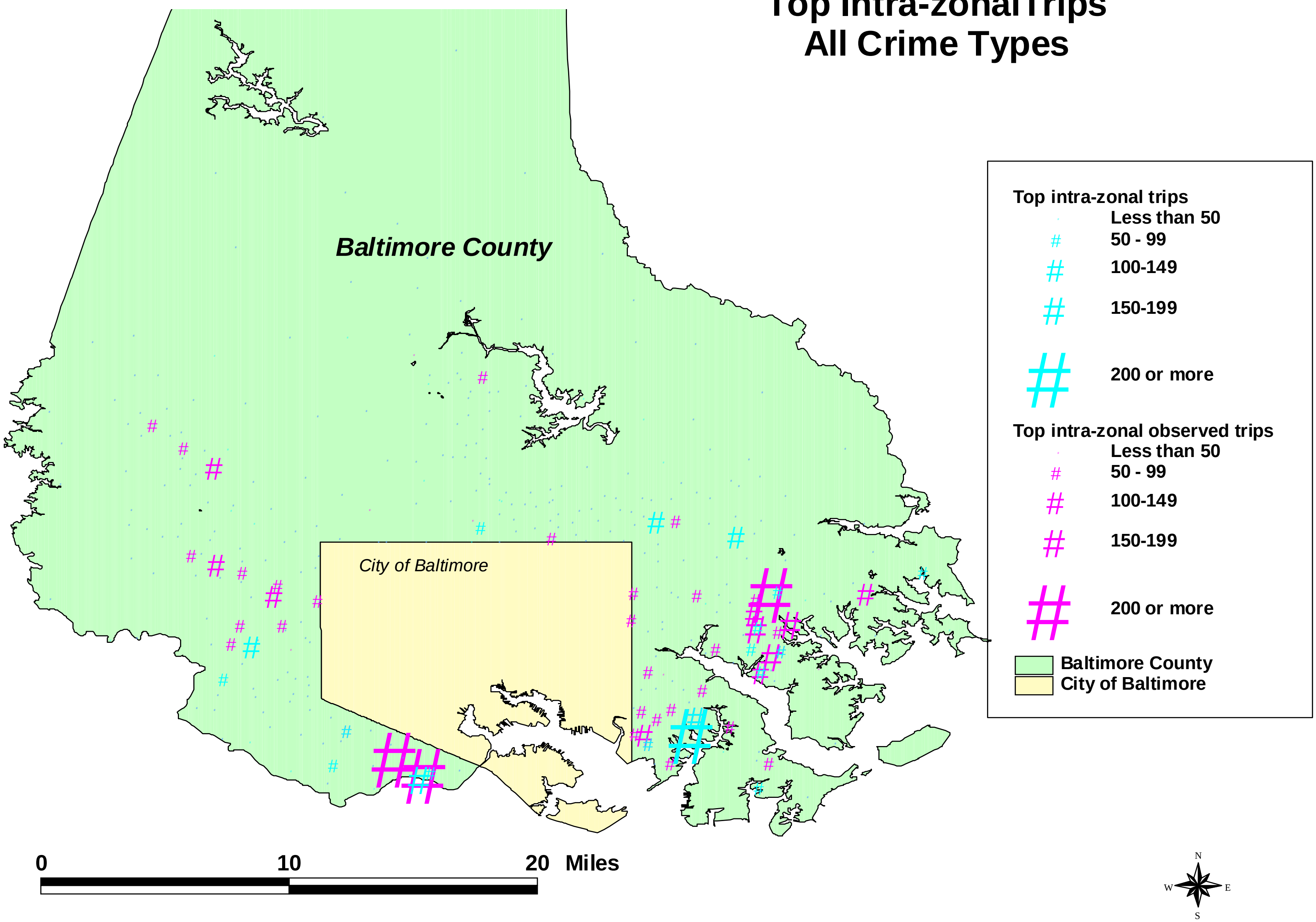
Predicted Baltimore County Crime Trips: 1993-1997
Top 1000 Links
All Crime Types



Comparison of Predicted and Observed Crime Trips
1000 Top Zone-to-Zone Trips
All Crime Types



Comparison of Predicted and Observed Crime Trips
Top Intra-zonalTrips
All Crime Types



Need for Multiple Optimization Criteria

- **Trip length congruence is not sufficient**
  Trips = f(attractions, productions, impedance)
  Multiple combinations can produce same congruence

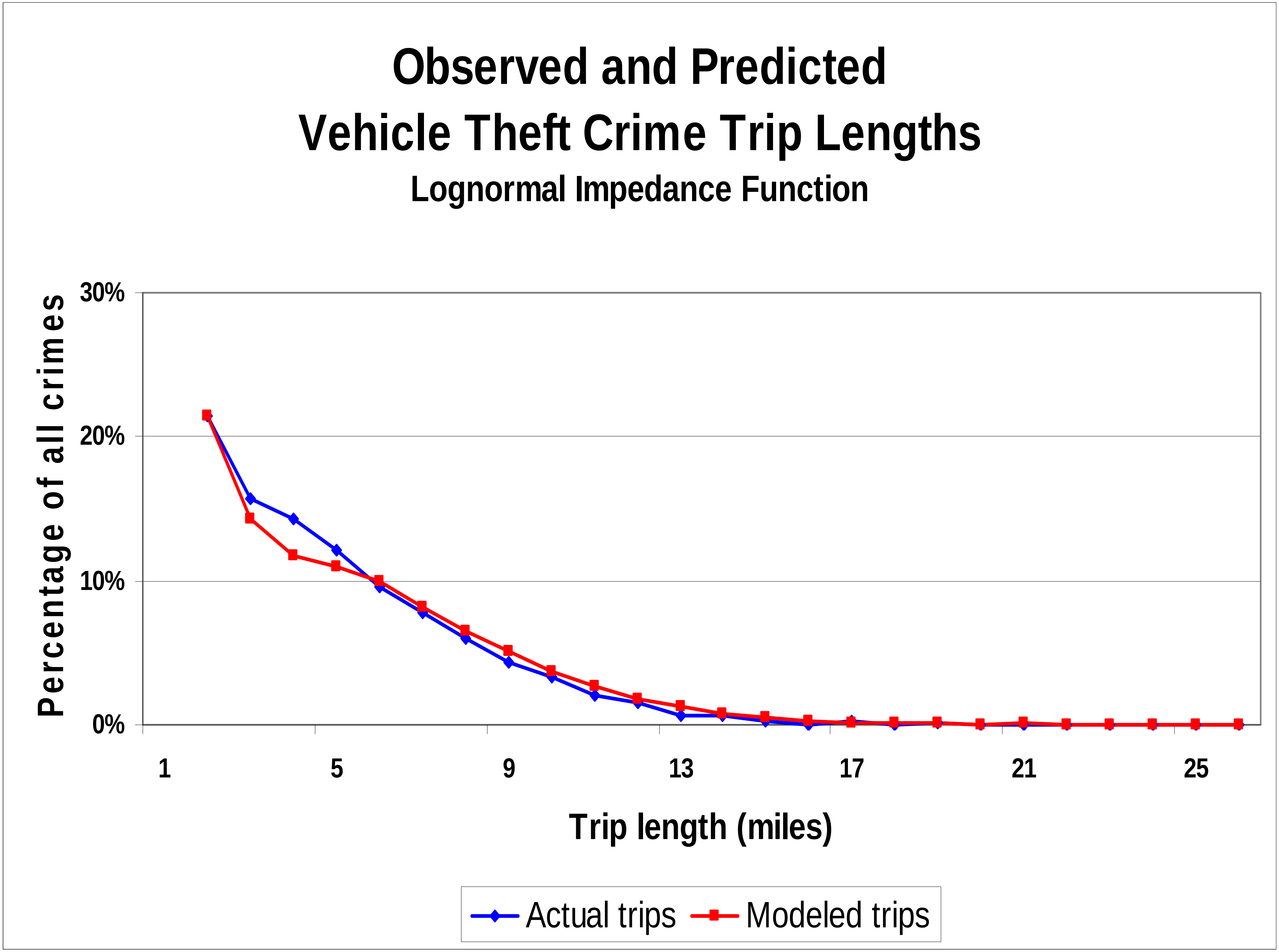
- **Must predict major links**
  Replicate highest volume trip links

- **Best solution <u>balances</u> trip length congruence with prediction of major links**

Vehicle Theft Model
With Highest Coincidence Ratio

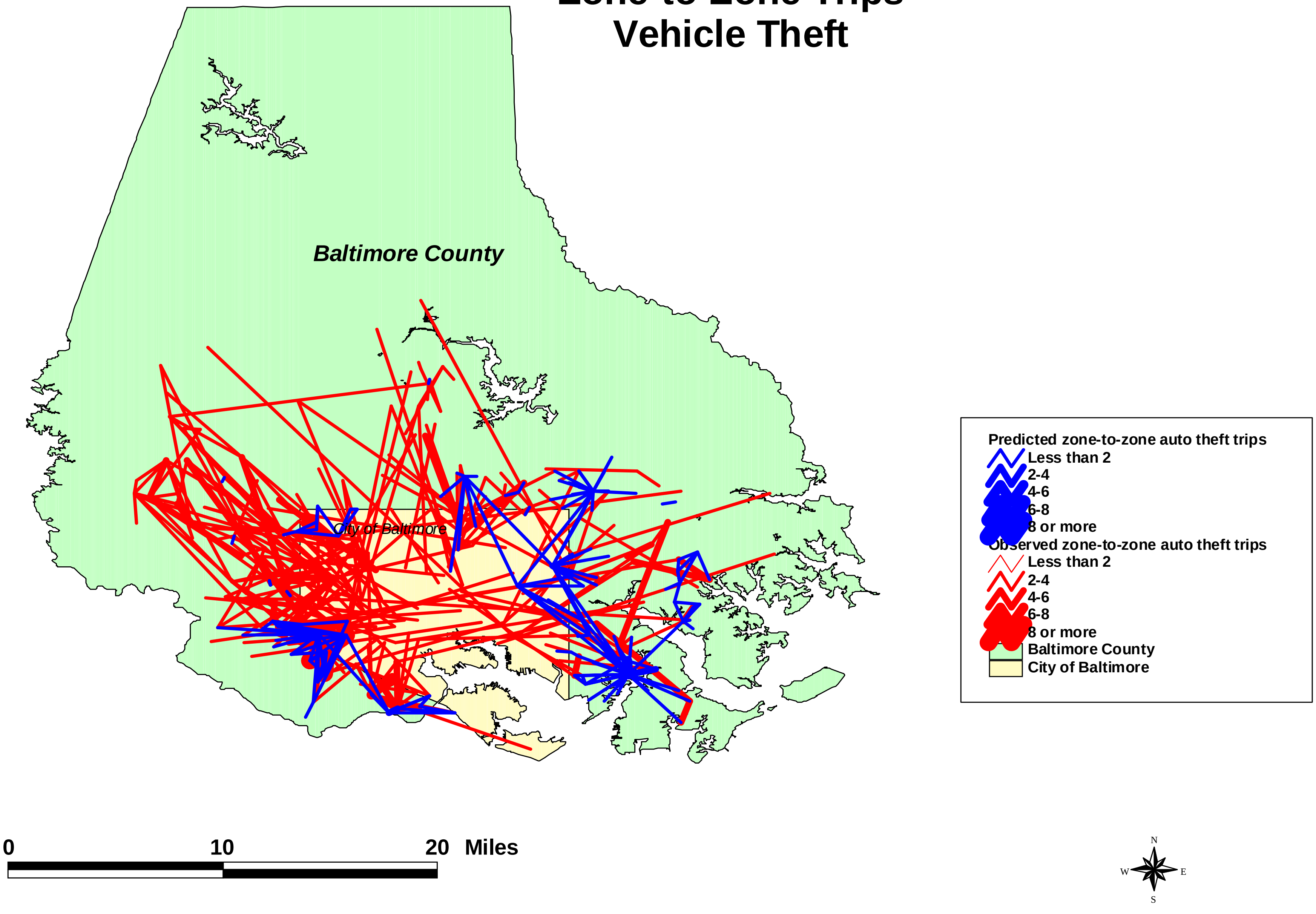
Lognormal function:

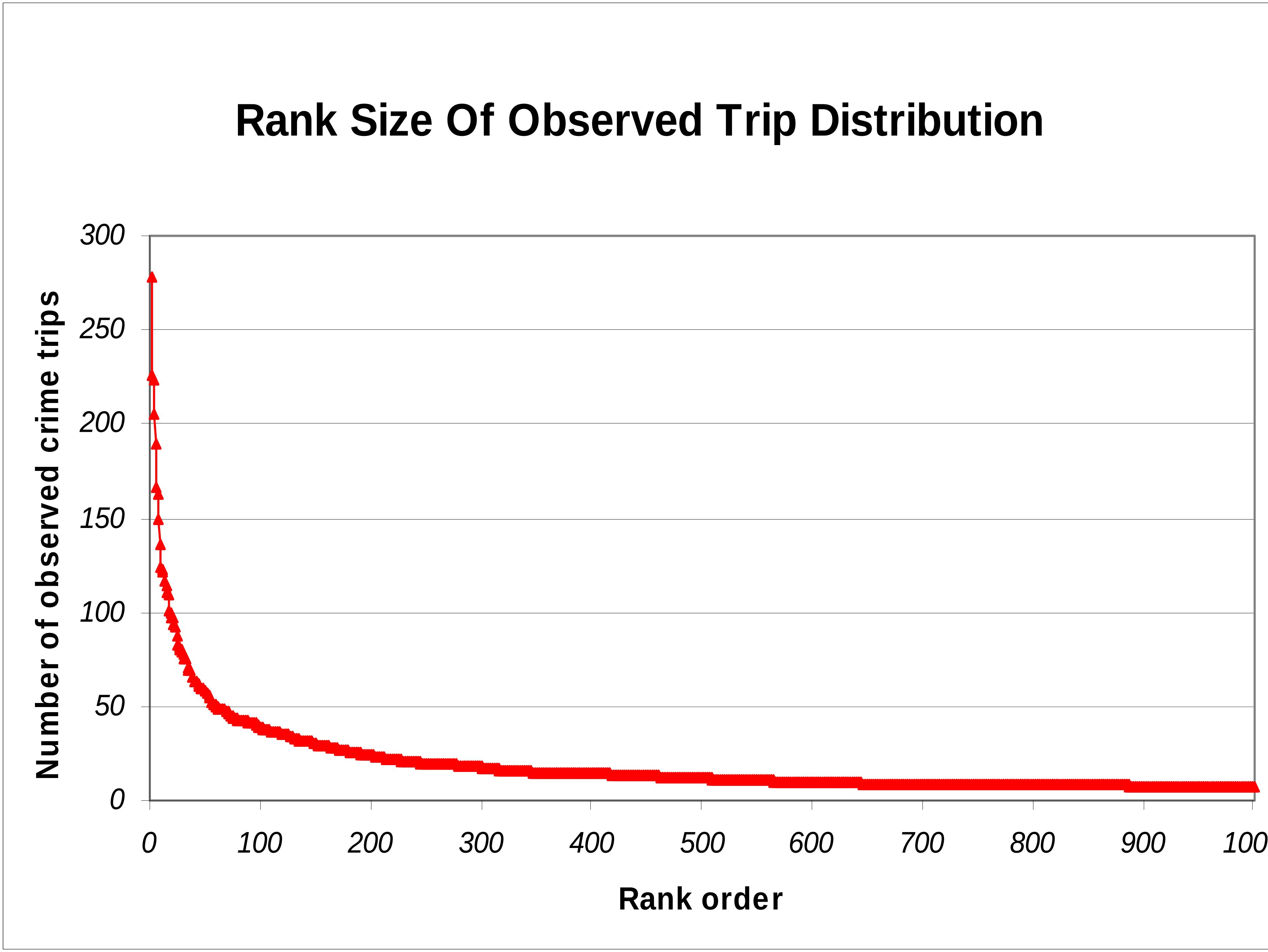
- Mean distance = *8.83 miles*
- Standard deviation = *4.6 miles*
- Coefficient = *1*
- Exponents: *Origins = 1.02 Destinations = 3.09*
- Congruity tests:
  Average observed trip length = 5.37 miles
  Average Predicted trip length = 5.82 miles
  Median observed trip length = 2.49
  Median predicted trip length = 2.93
  Coincidence ratio = 0.903765



Observed and Predicted
Vehicle Theft Crime Trip Lengths
Lognormal Impedance Function
Percentage of all crimes
30%
20%
10%
0%
1
5
9
13
17
21
25
Trip length (miles)
Actual trips
Modeled trips

Comparison of Actual and Predicted Crime Trips
Zone-to-Zone Trips
Vehicle Theft



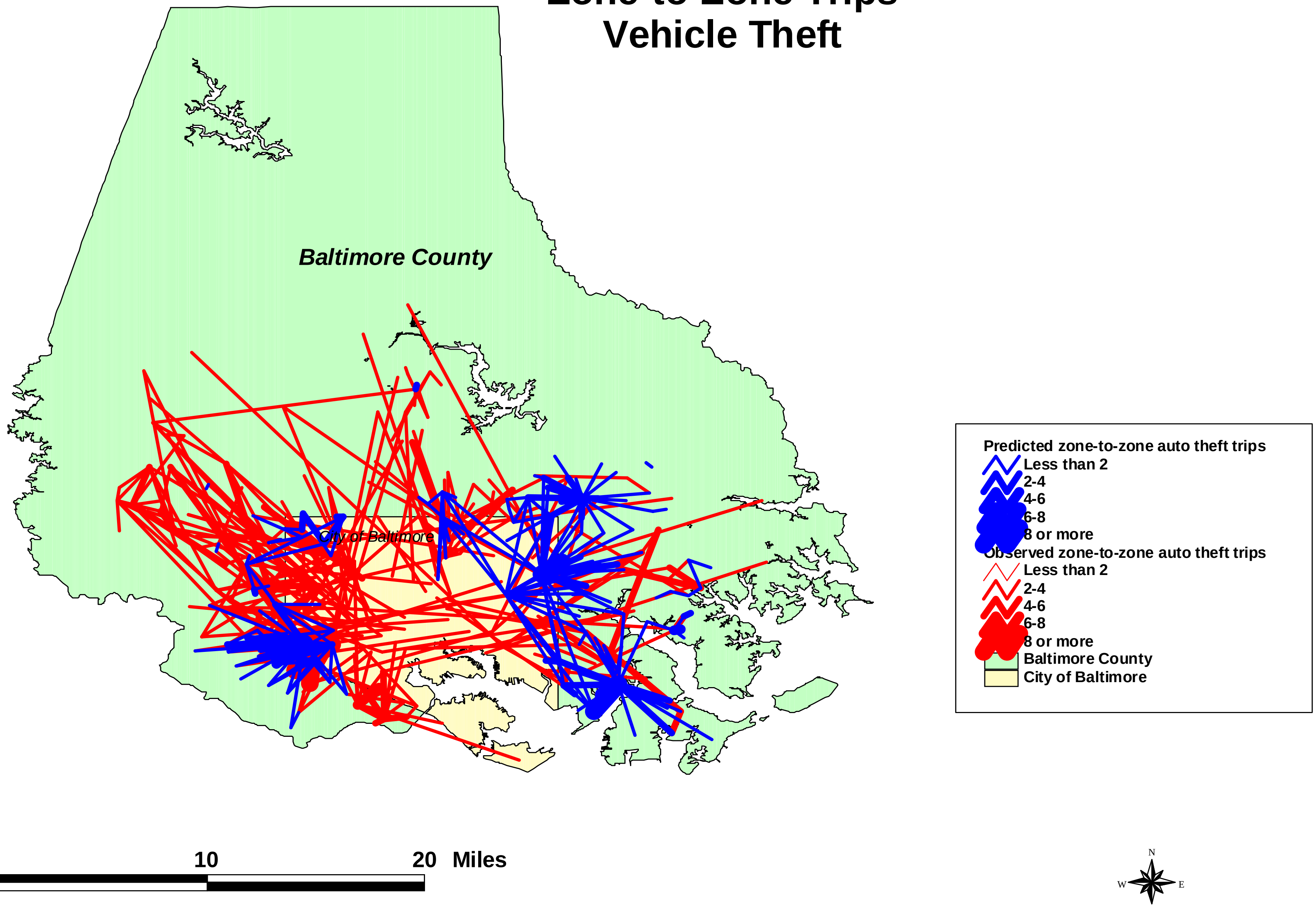

Rank Size Of Observed Trip Distribution
300
250
200
150
100
50
0
Number of observed crime trips
0
100
200
300
400
500
600
700
800
900
1000
Rank order

Alternative Vehicle Theft Model

Lognormal function:

- **Mean distance =** ***5.3 miles***
- **Standard deviation =** ***4.6 miles***
- **Coefficient =** ***1***
- **Exponents:** ***Origins = 1.5 Destinations = 5.75***
- **Congruity tests:**
  Average observed trip length = 5.37 miles
  Average Predicted trip length = 5.45 miles
  Median observed trip length = 2.49
  Median predicted trip length = 2.37
  Coincidence ratio = 0.903422

Comparison of Actual and Predicted Crime Trips
Zone-to-Zone Trips
Vehicle Theft



Problems with Using Gravity Model For Trip Distribution Modeling

- **Does only a fair job of replicating major links**
  Good average, but not good for major links

- **Assigns 'trips' to each cell because of the arithmetic**
  However, crime trips are very skewed

- **Need to develop skewed version of gravity model**

Crime Mode Split

Crime Mode Split

- **Estimates travel mode used (e.g., walk, drive, bus)**
- **Lack of information about offender travel modes**
- **Because of lack of data, we estimated general accessibility functions based on travel surveys**

Estimating the Utility of Crime Travel

Utility = f(benefits, costs)

For travel between two zones, benefits are assumed to be equal for different modes

Therefore, differences in utility between modes represent differences in costs

Relative Utility Function: Multinomial Logit

Probability of using mode k to travel from zone i to zone j is the utility of mode k relative to all utilities

$$P_{ijk} = \frac{e^{(-\beta C_{ijk})}}{\Sigma[\, e^{(-\beta C_{ijk})} \,]}$$

Where

P_{ijk}	*Probability of travel from i to j by mode k*
c_{ijk}	*Cost of travel from i to j by mode k*
β	*Coefficient*
e	*Base of natural logarithm*

Generalized Relative Utility Function

$$P_{ijk} = \frac{F(\beta C_{ijk})}{\Sigma[\, F(\beta C_{ijk})\,]}$$

Where

F *Some non-linear decay function of cost*

What kind of costs?

Difficulty in accessing destination

- Distance
- Travel time
- Anciliary costs ***(e.g., parking, gasoline)***
- Perceived costs ***(e.g., risk of arrest)***

Aggregate modeling assumes average costs

Disaggregate 'utility' theory measures costs for individuals with surveys

Total cost = travel time + parking + convenience + comfort + privacy + ….other things

With Crime, Can’t Measure True Costs

- Lack of individual-level data
- Must use approximations
- Thus, use accessibility index

Relative travel patterns by mode

Can Make Assumptions About Travel Based on Research

- **Identify logical mathematical function**
- **Create target mode split proportions**
- **Iteratively estimate parameters**

Can Make Assumptions About Travel Based on Research

Trip Distribution by Mode

	Houston	Portland
Drive	98.9%	88.6%
Bus	1.1%	5.8%
Walk	0.9%	4.6%
Bike	-	1.1%

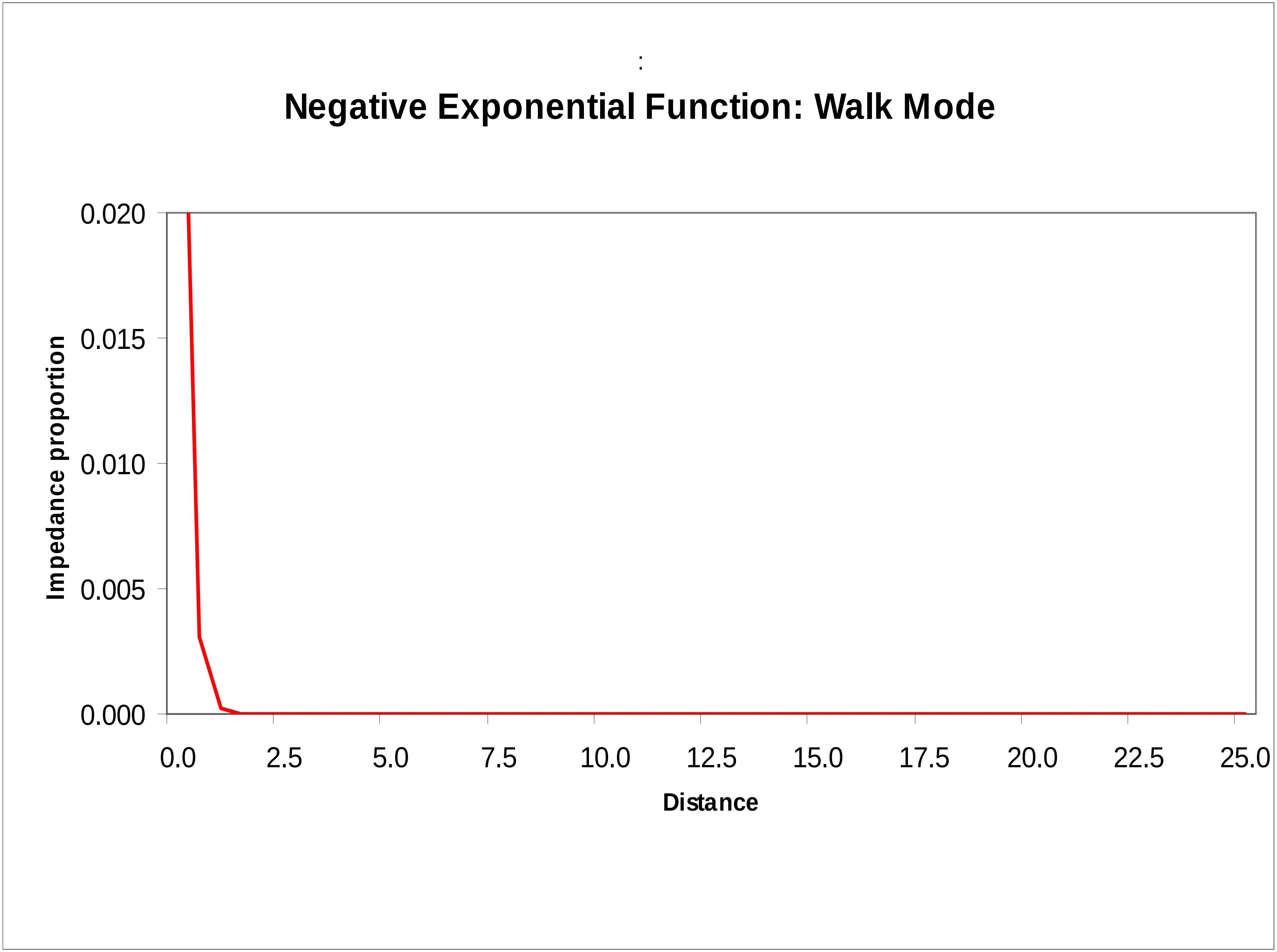
Created Plausible Default Parameter Estimates

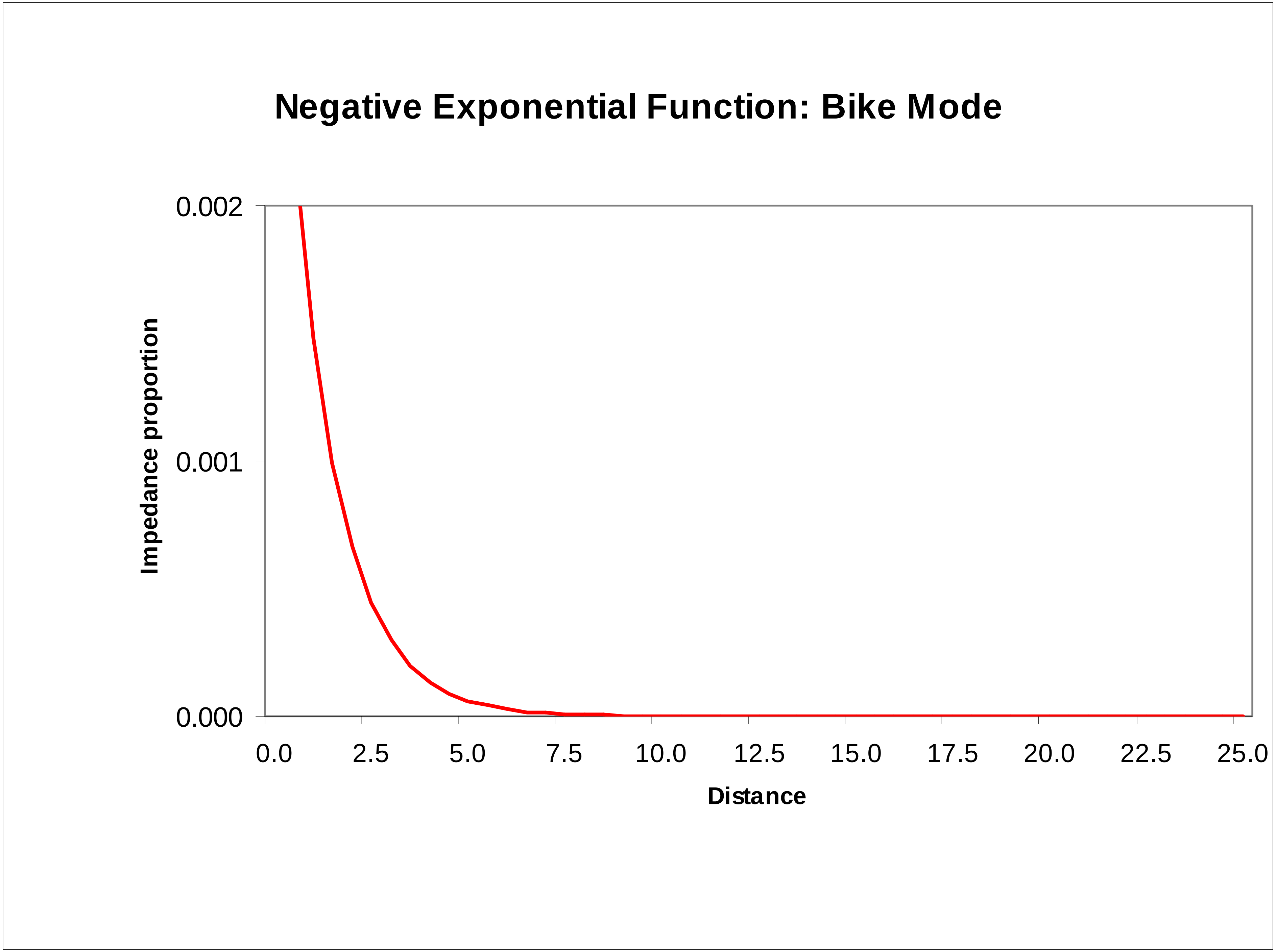
- **Walking trips** **(mean=0.5 mi /4% of all trips)**
- **Bicycle trips** **(mean=2 mi /1% of all trips)**
- **Driving trips** **(mean=6 mi /90% of all trips)**
- **Bus trips** **(mean = 4 mi /4% of all trips)**
- **Train trips** **(mean = 6 mi /1% of all trips)**

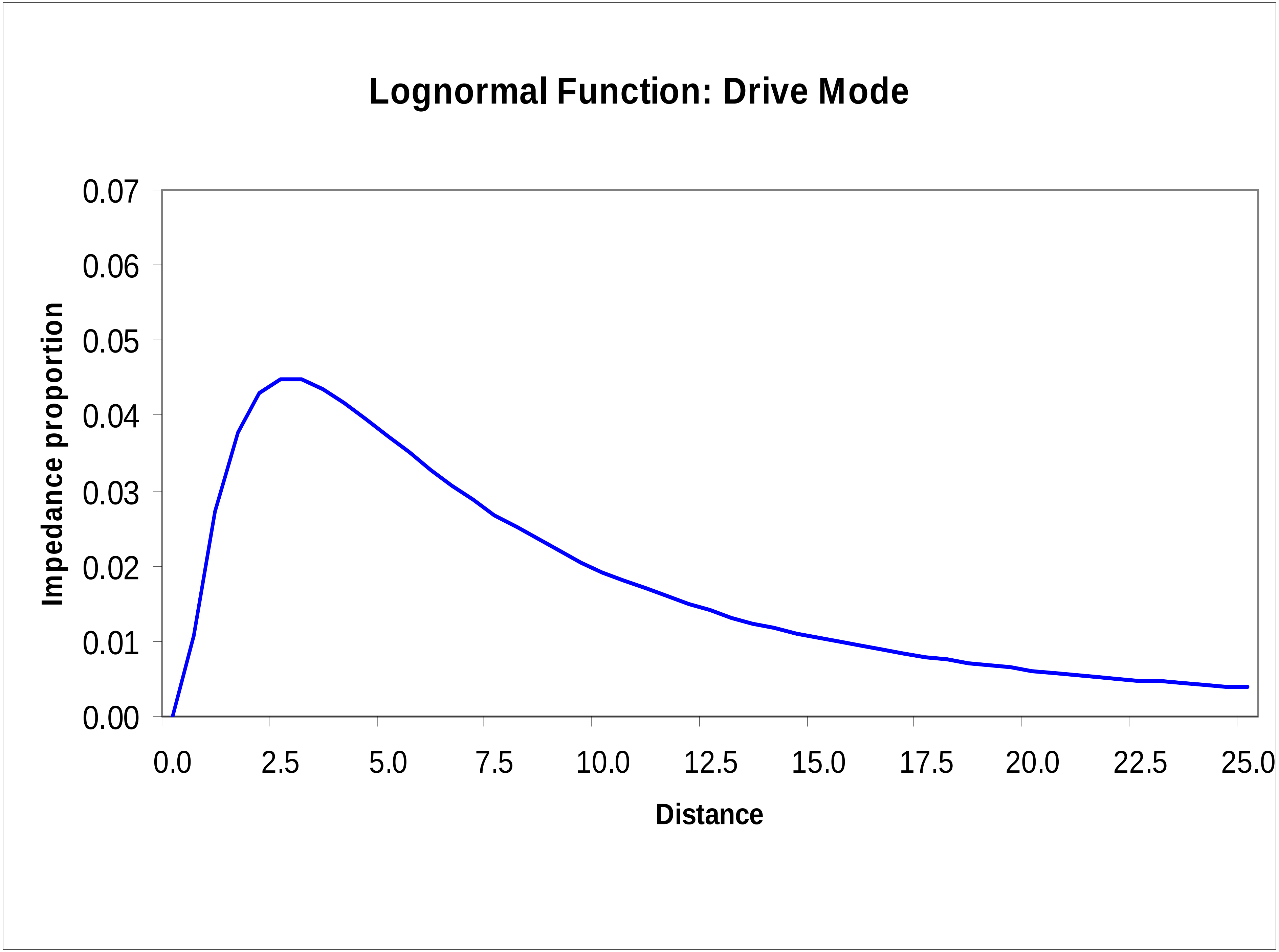
Can Estimate Likely Travel Functions

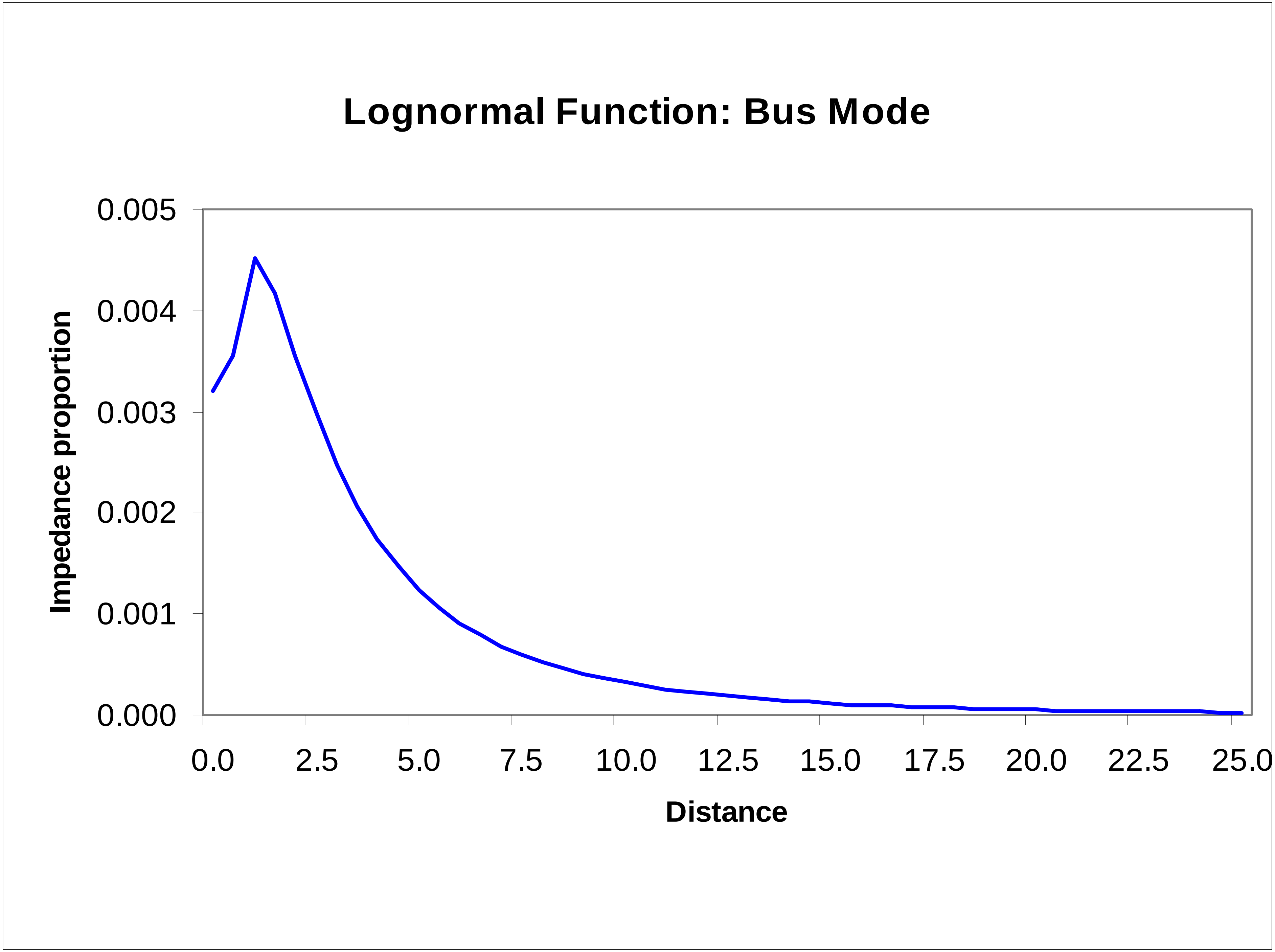
- Must capture mean distance assumptions
- Must produce overall proportional split between modes
- Solution is iterative

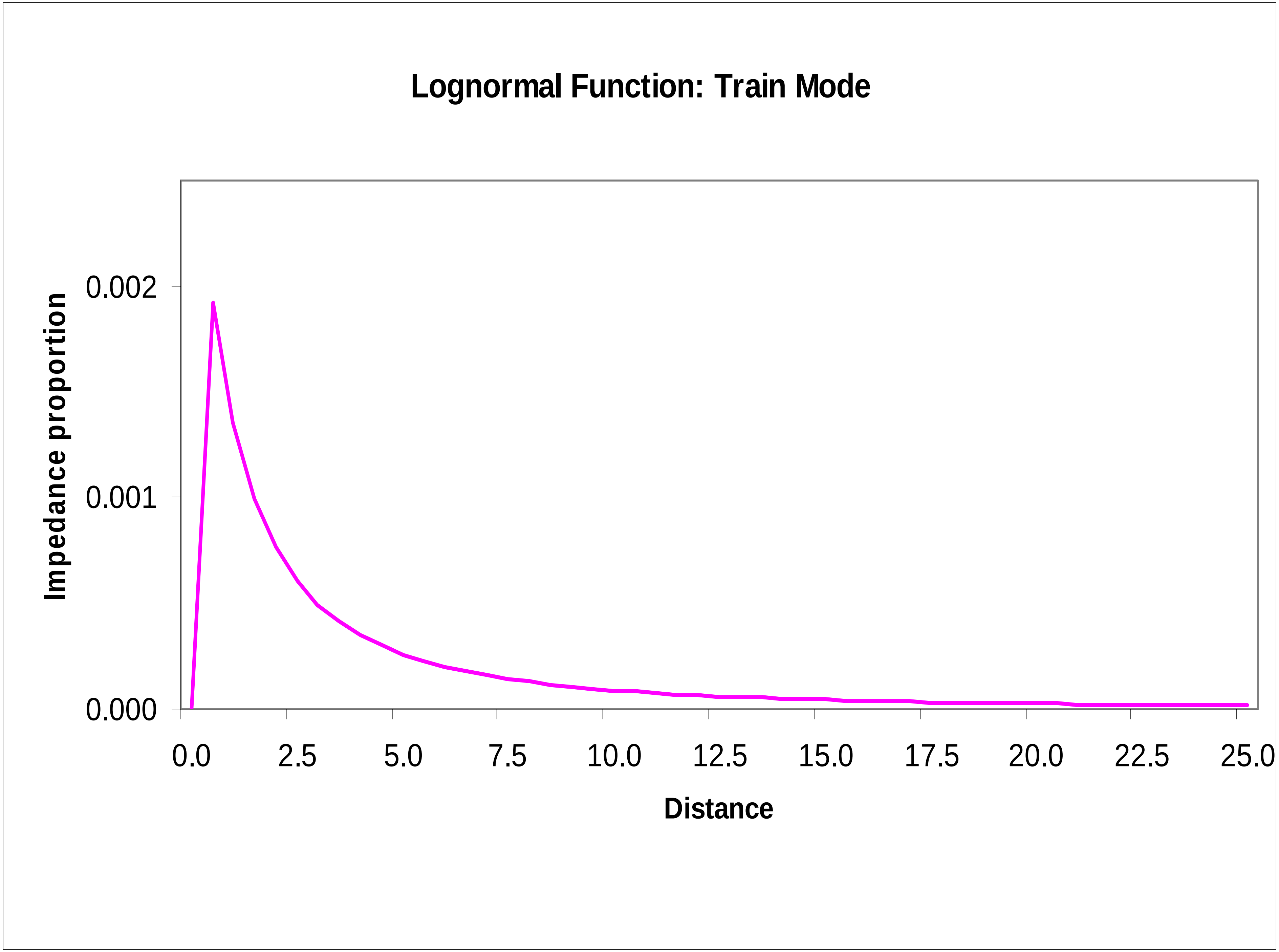
:

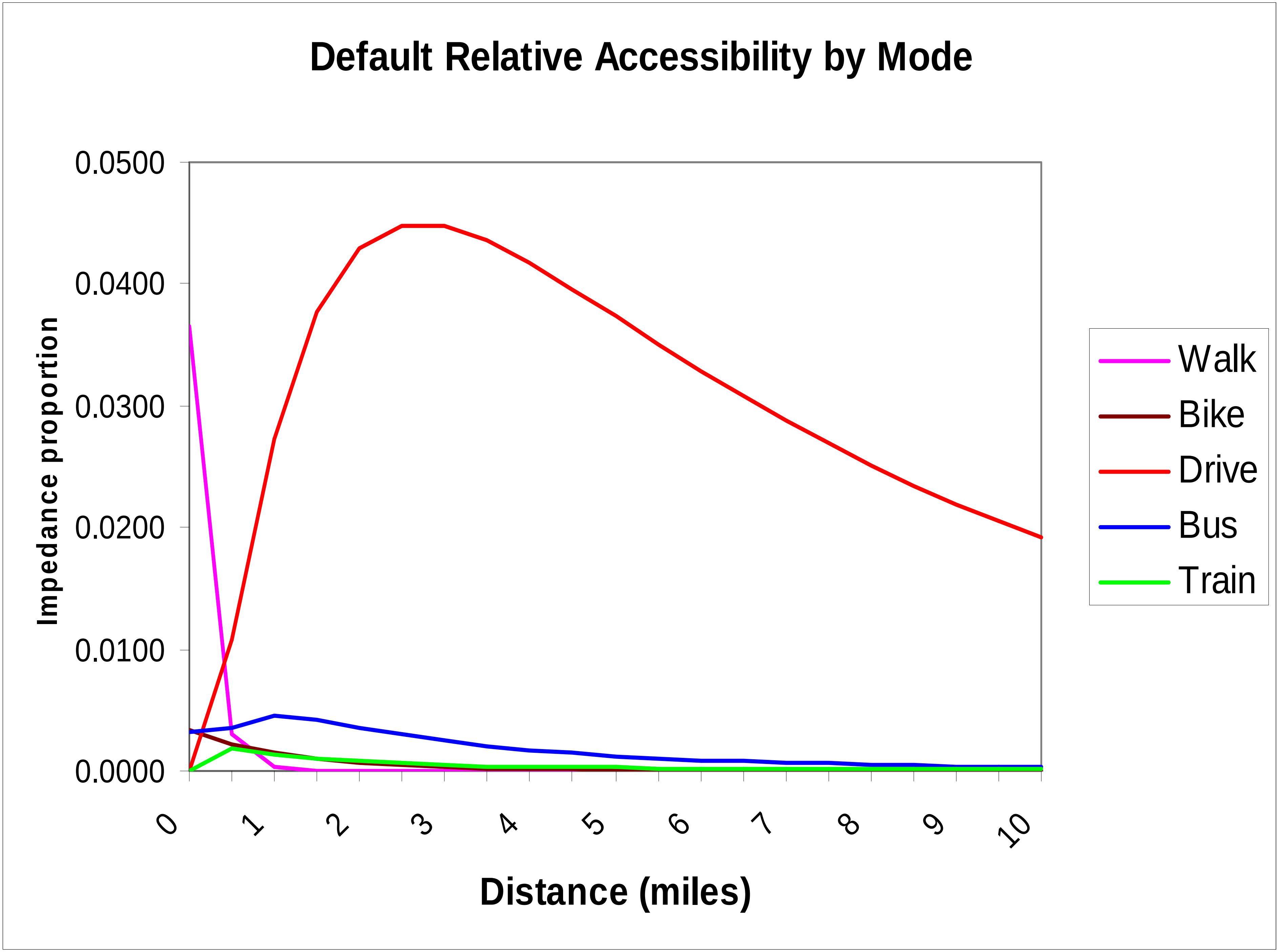

Negative Exponential Function: Walk Mode
Impedance proportion
0.020
0.015
0.010
0.005
0.000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Distance


Negative Exponential Function: Bike Mode
Impedance proportion
0.002
0.001
0.000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Distance

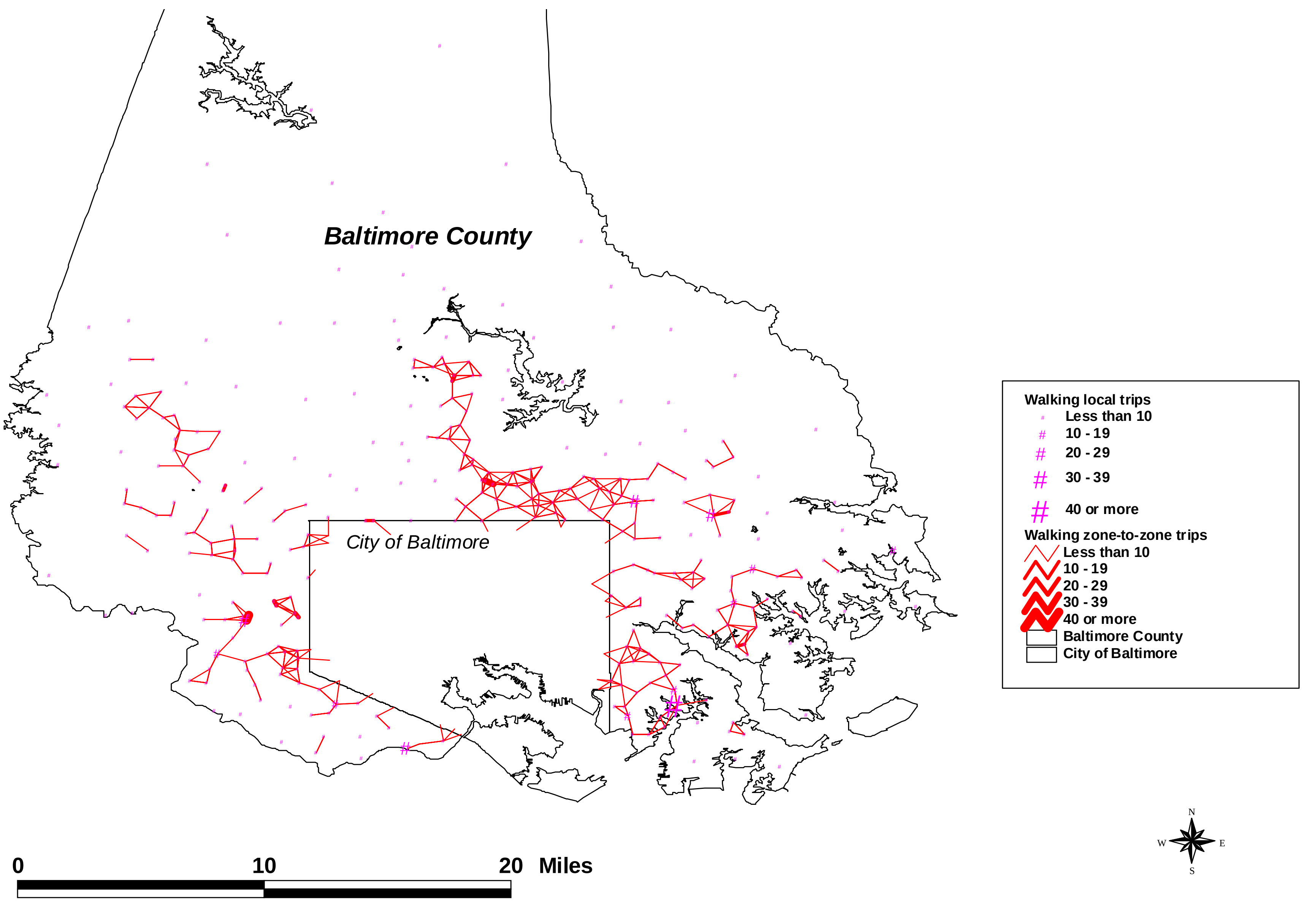

Lognormal Function: Drive Mode
Impedance proportion
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Distance


Lognormal Function: Bus Mode
Impedance proportion
0.005
0.004
0.003
0.002
0.001
0.000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Distance

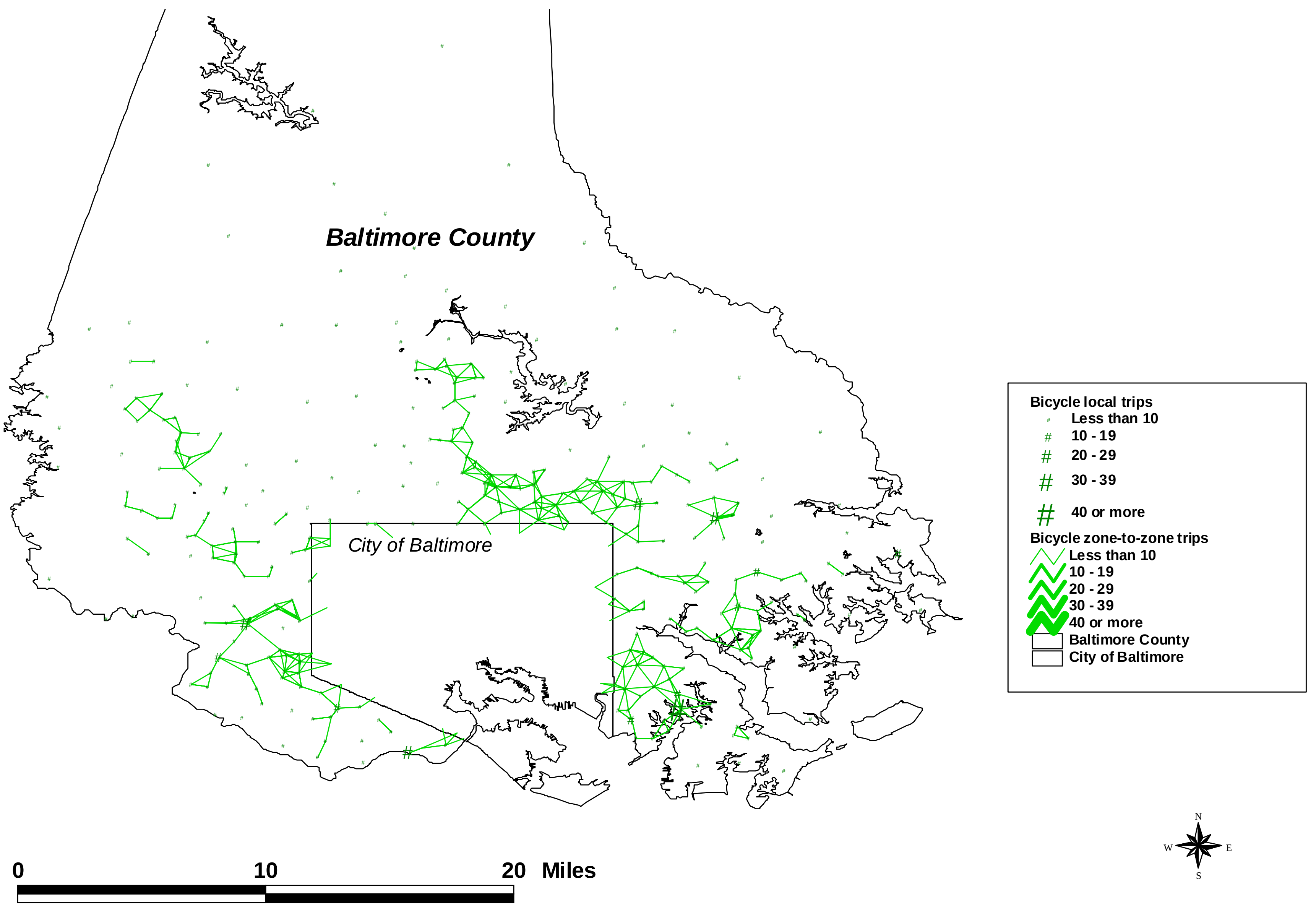

Lognormal Function: Train Mode
Impedance proportion
0.002
0.001
0.000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Distance


Default Relative Accessibility by Mode
Impedance proportion
0.0500
0.0400
0.0300
0.0200
0.0100
0.0000
0
1
2
3
4
5
6
7
8
9
10
Distance (miles)
Walk
Bike
Drive
Bus
Train

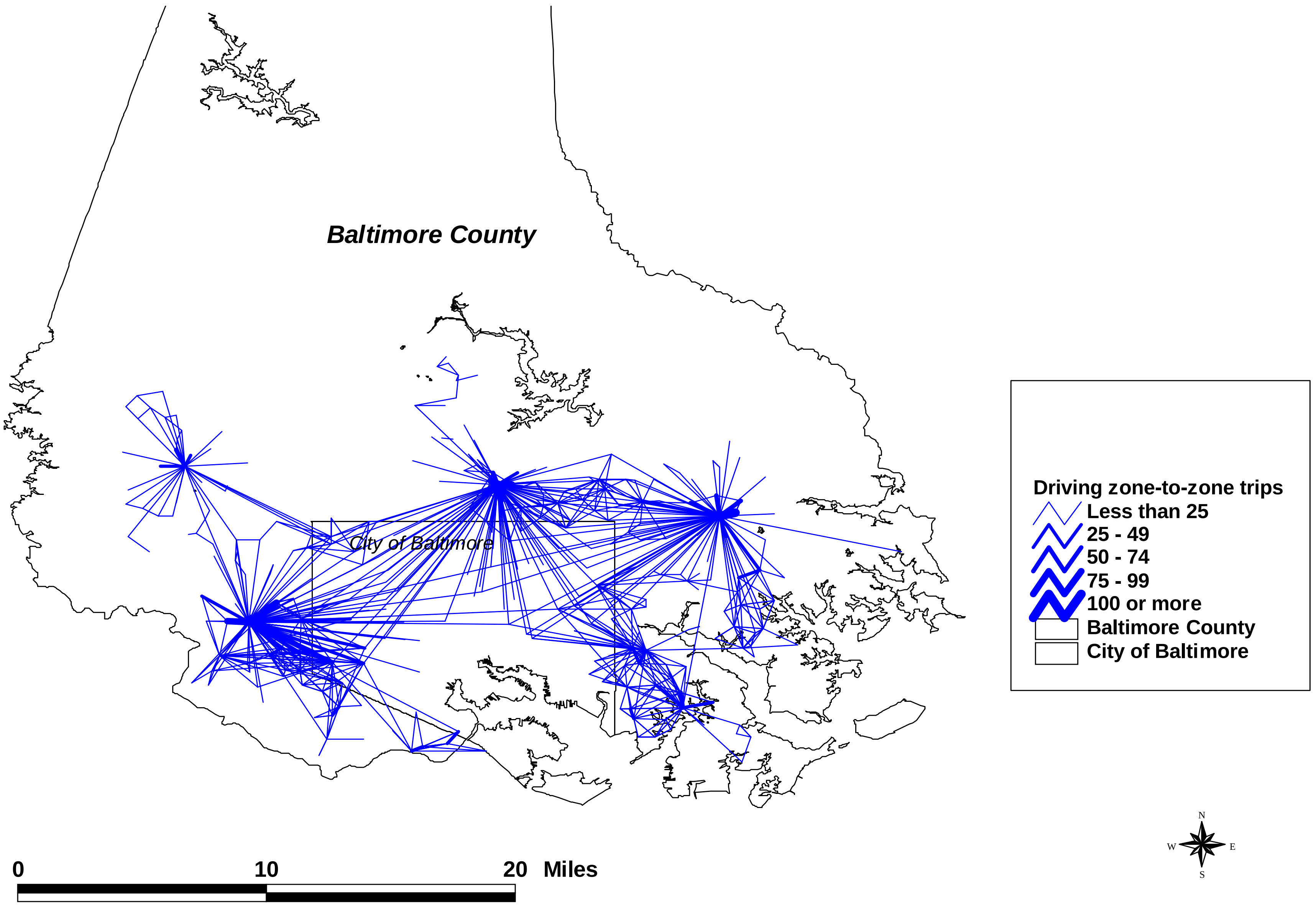
Mode Split: Walking Crime Trips



Mode Split: Bicycle Crime Trips



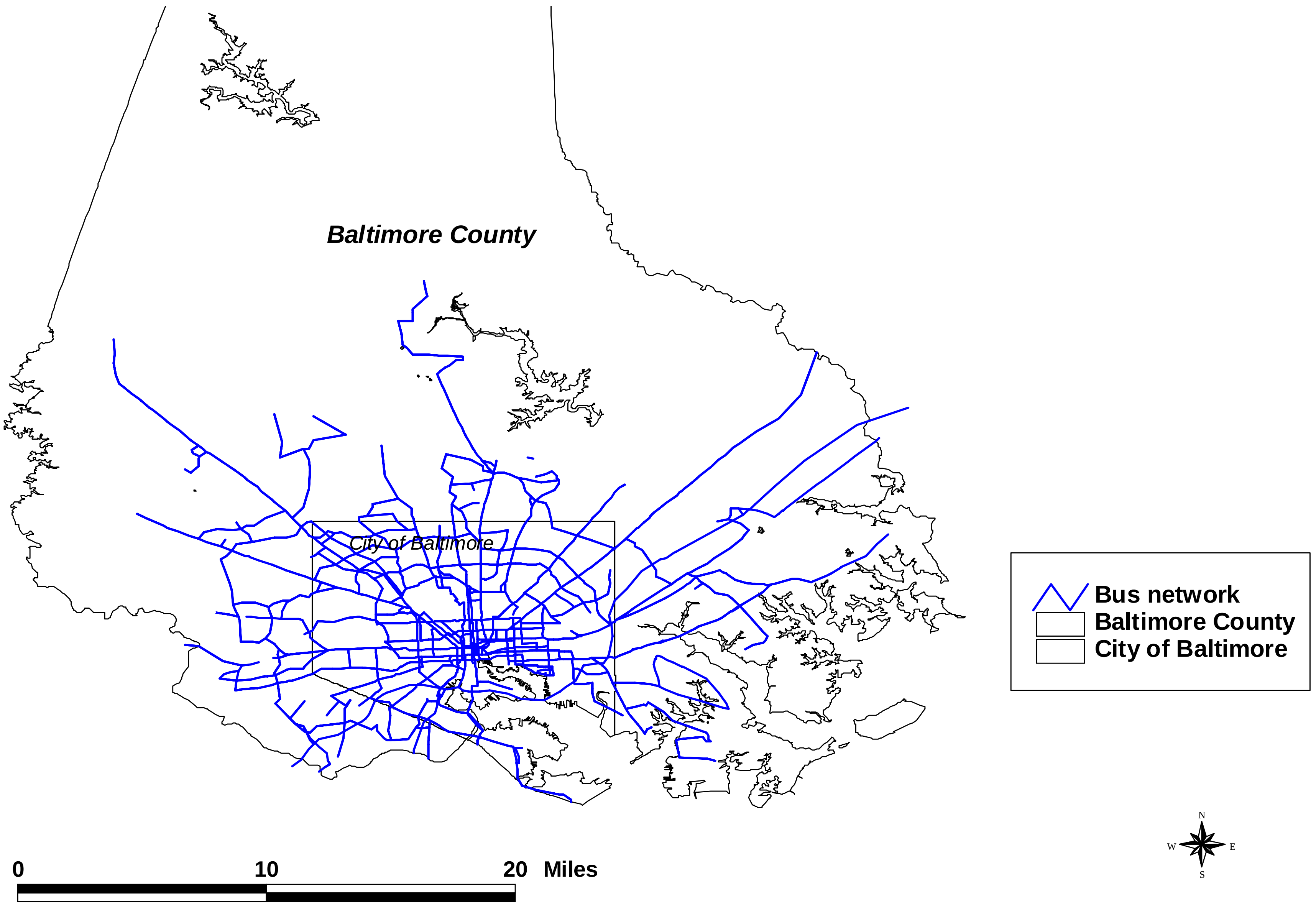
Mode Split: Driving Crime Trips



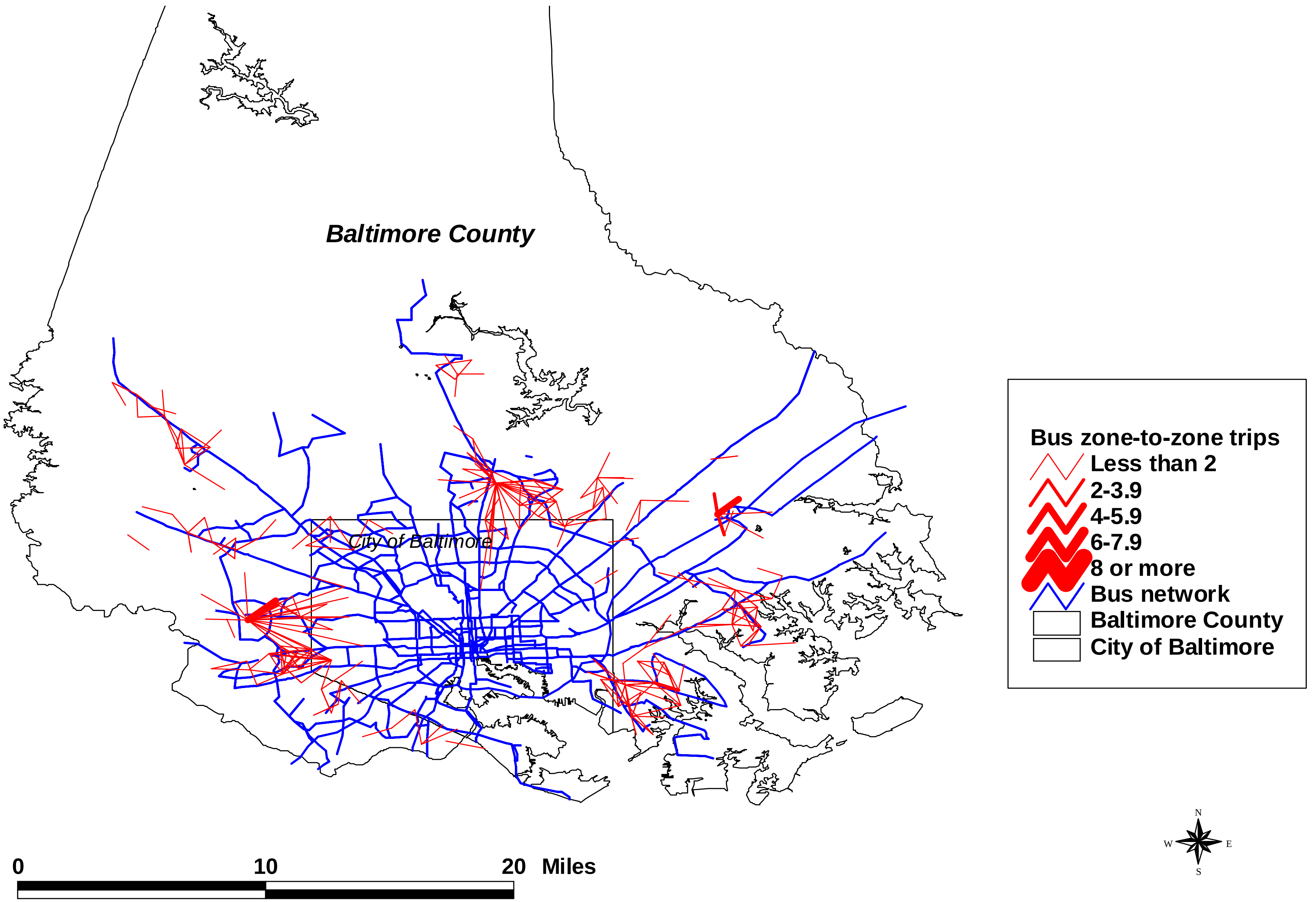
Transit Modes

- Must constrain to a network
  to avoid illogical results
- Need bus routes
- Need intra-urban train routes
- Access to transit complex
  (walk, park & ride, drop off)
- We assume straight distance access

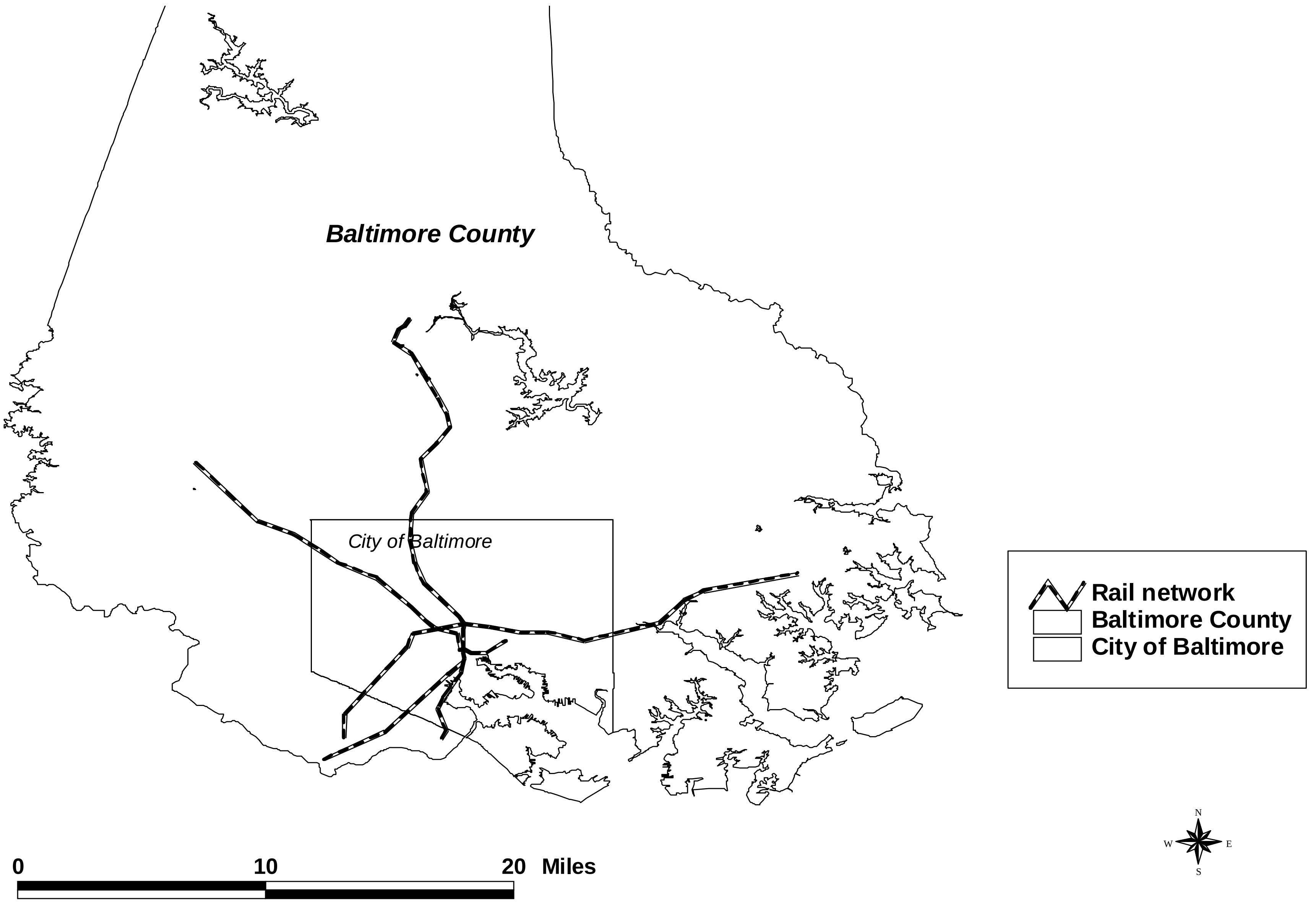
Baltimore Bus Routes



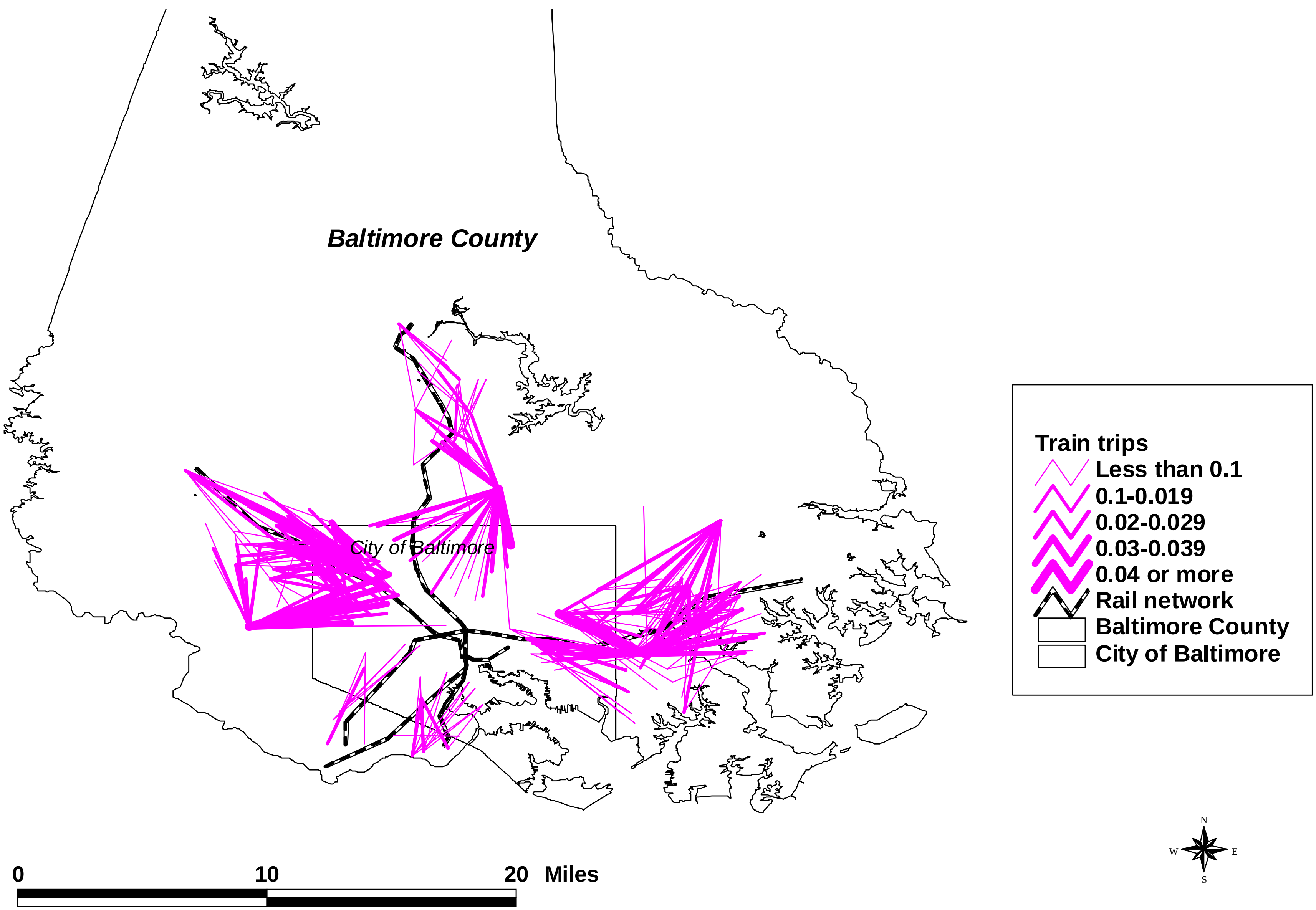
Mode Split: Bus Crime Trips



Baltimore Intra-Urban Rail Network



Mode Split: Train Crime Trips



Limitations

Problems with aggregate approach

- Does not consider characteristics of individuals
- Can’t explain within zone travel
- Can’t explain individual variations for people in same zone
- Can’t explain linked trips
- Does not incorporate time of day well

Newer approaches use utility theory to predict

- Multinomial logit
- Hierarchical logit

But, must have individual-level data.

Crime Trip
Network Assignment

Network Assignment

- Finally, the predicted trips are assigned to a likely route
- The shortest *cost* path on a network is calculated
- The trip by mode is assigned to the path
- Different modes use different networks

Common Types of Networks: I

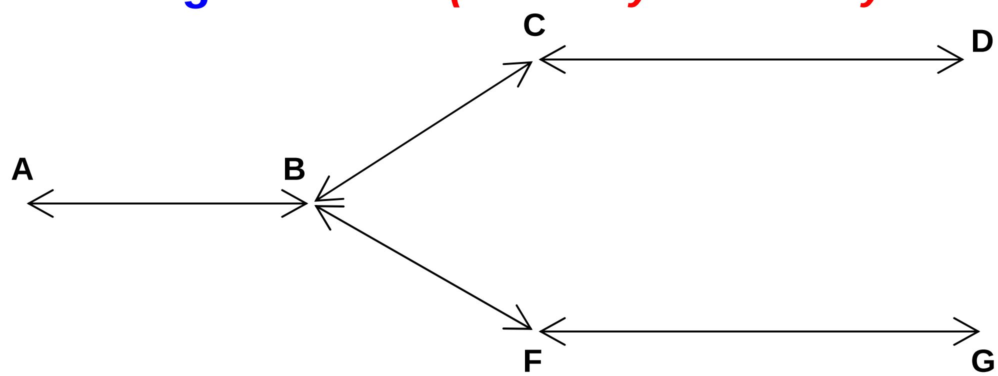
Linear (or bidirectional) referencing system (e.g., TIGER)

Advantages

- Widely available
- Represents all roads
- Spatially more accurate *(multiple intermediate nodes)*

Disadvantages

- Connectivity often not tested
- Distance-based
- Does not distinguish directions (one way streets)
- Large network *(too many 'blind alleys' in search)*



Common Types of Networks: II

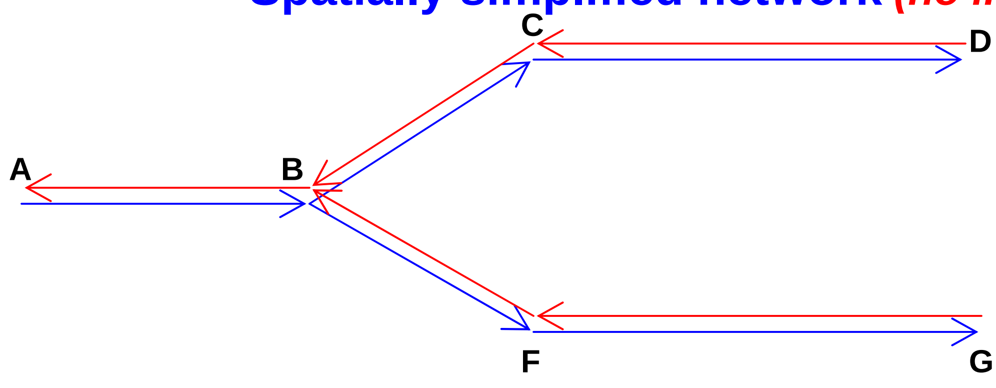
Modeling network (unidirectional or duel referencing)

Advantages

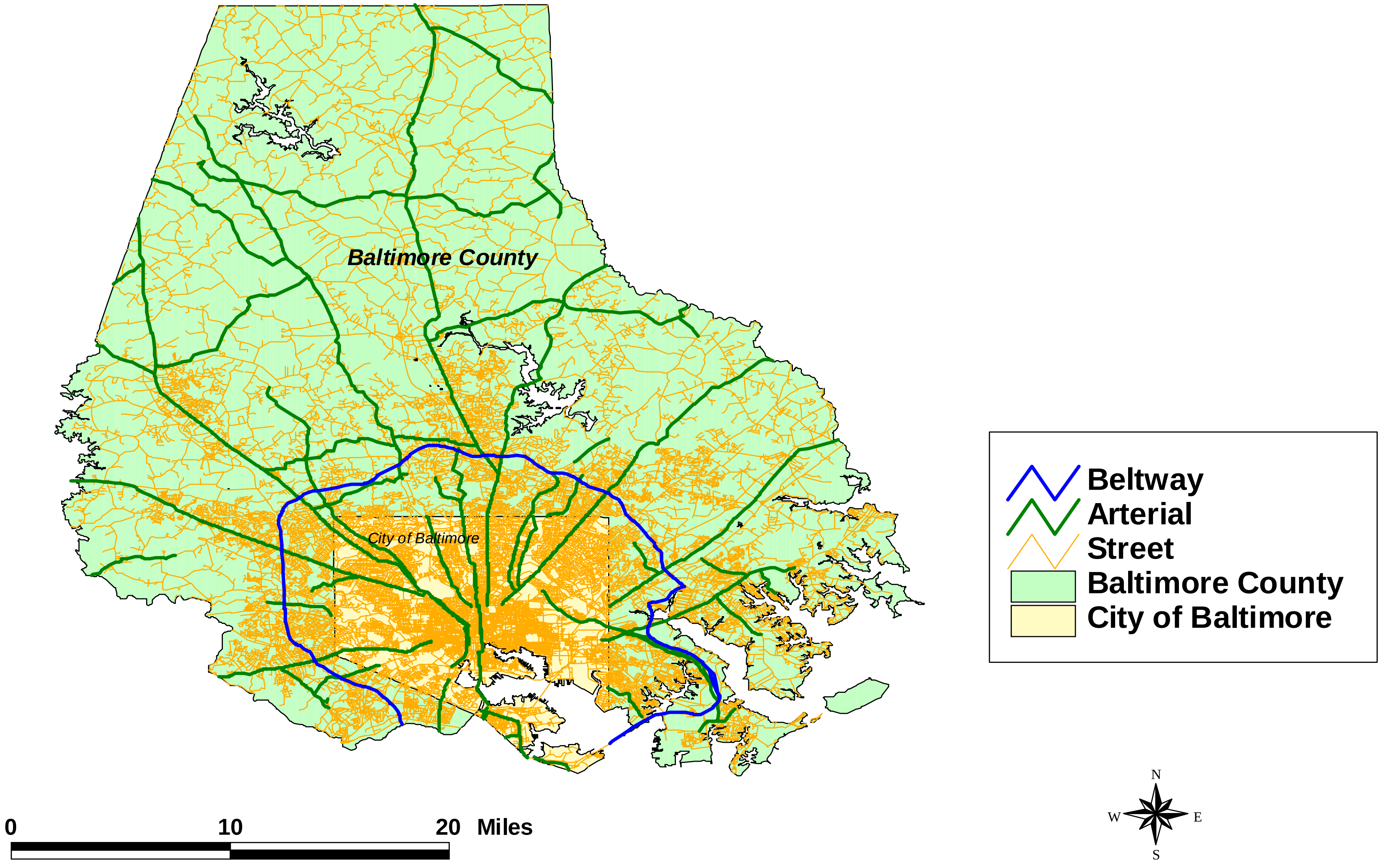
- Connectivity is tested
- Allows modeling by travel time, speed, time of day
- Distinguishes directions
- Efficient for calculations

Disadvantages

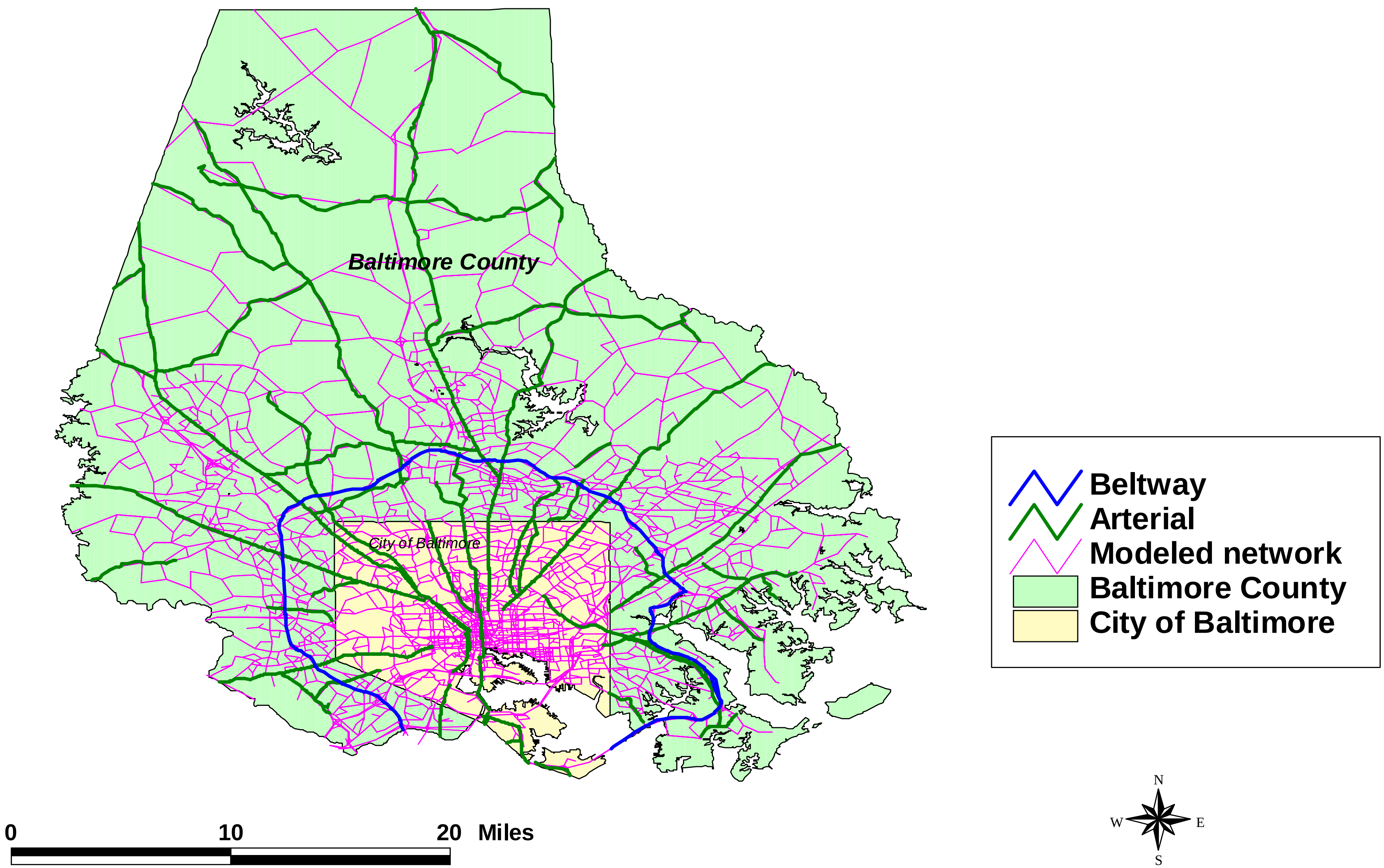
- Does not include all roads
- Spatially simplified network *(no intermediate nodes)*



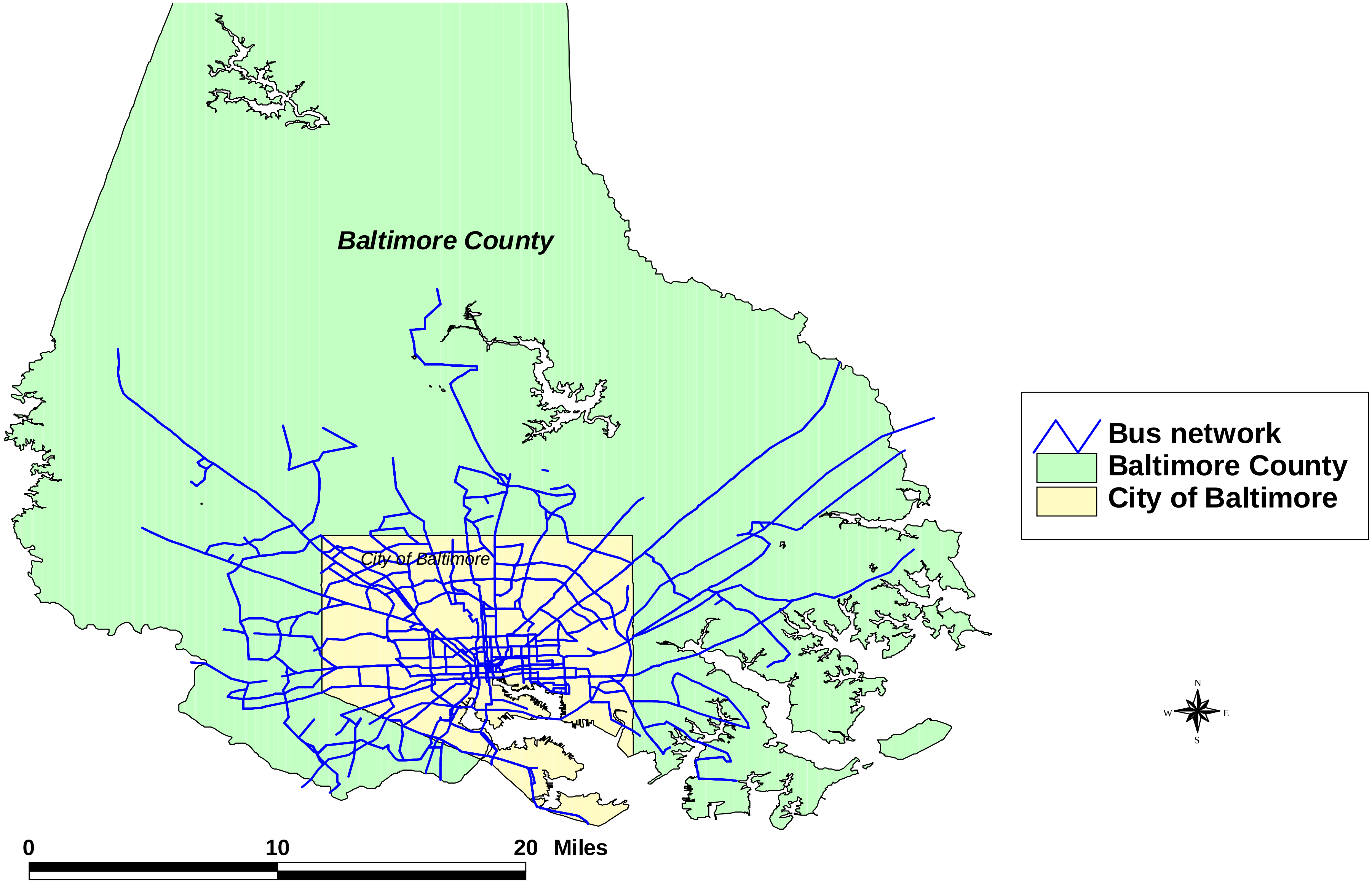
TIGER Street Network
49,015 Road Segments



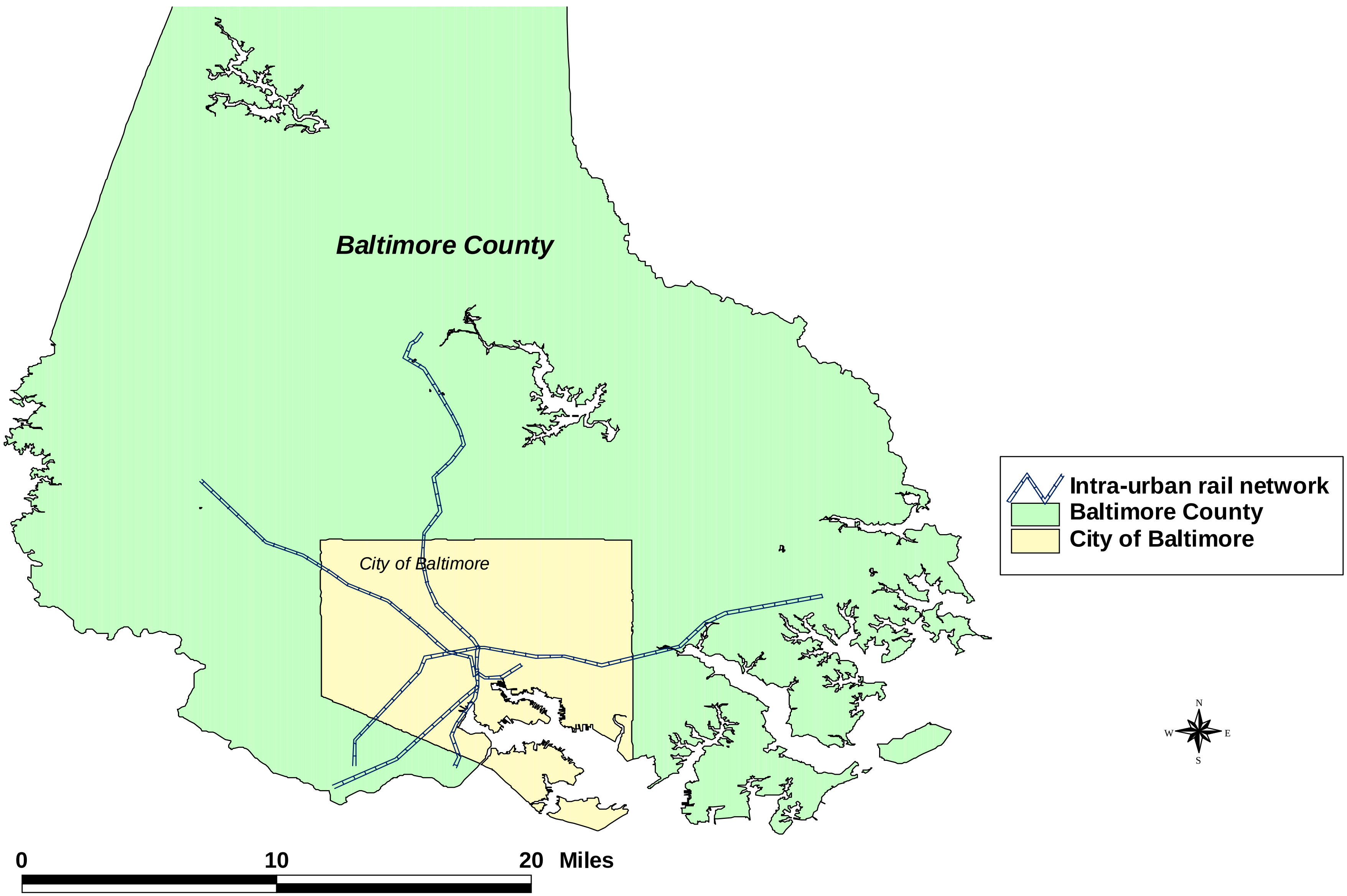
Modeled Street Network
11,045 Road Segments



Baltimore Bus Network
Bus Routes



Baltimore Intra-Urban Rail Network



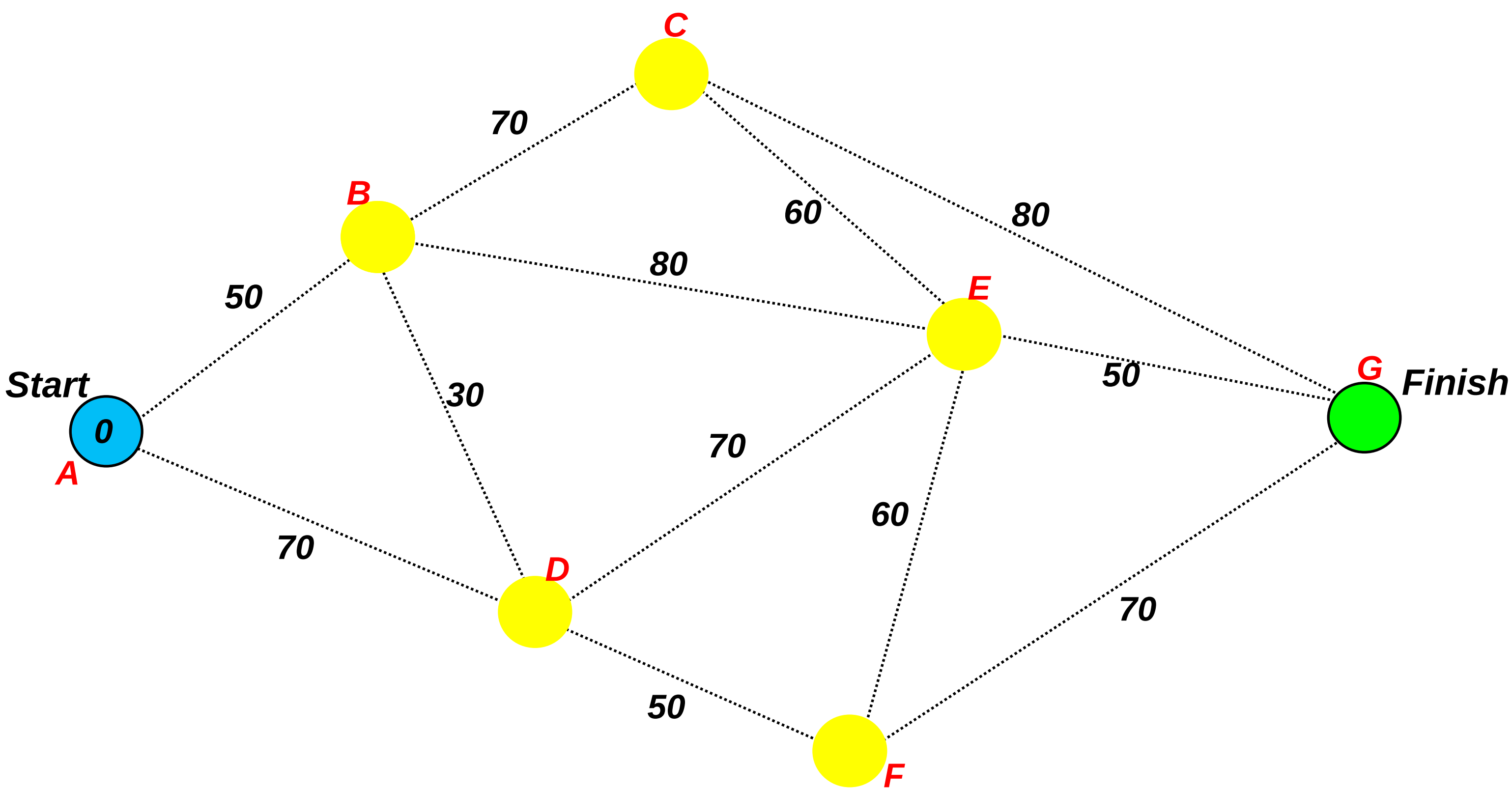
Dijkstra Shortest Path Algorithm

Shortest cost path from a single source to all other points

Iteratively finds the shortest path from the source to each point

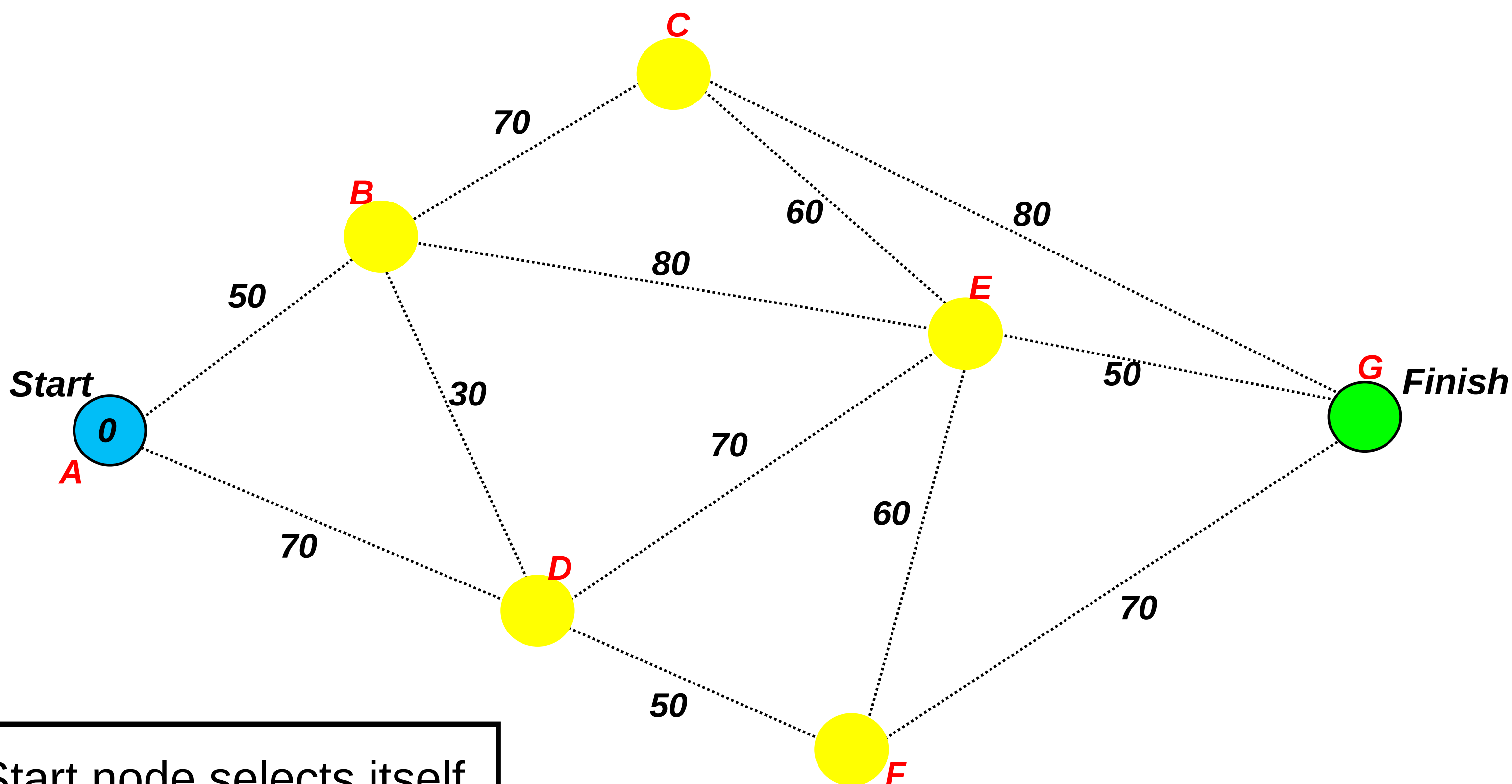
1. **Algorithm starts from an already-examined point**
2. **Examines the closest point not yet examined**
3. **Selects the point with the lowest impedance**
   (distance, travel time, cost)

Example of Dijkstra Algorithm



Example of Dijkstra Algorithm

Step 1

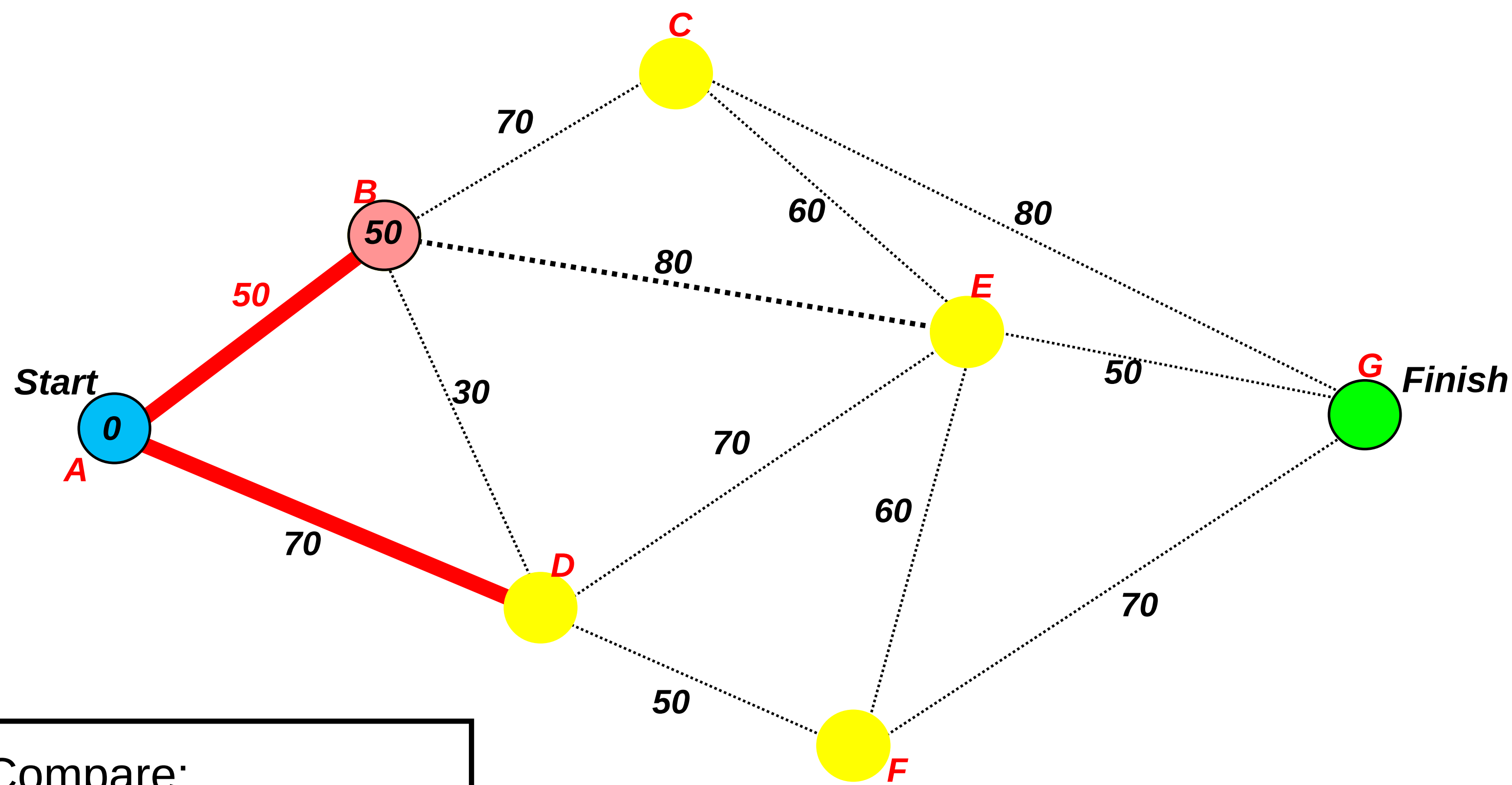


Start node selects itself

Path 1 = 0

Example of Dijkstra Algorithm

Step 2



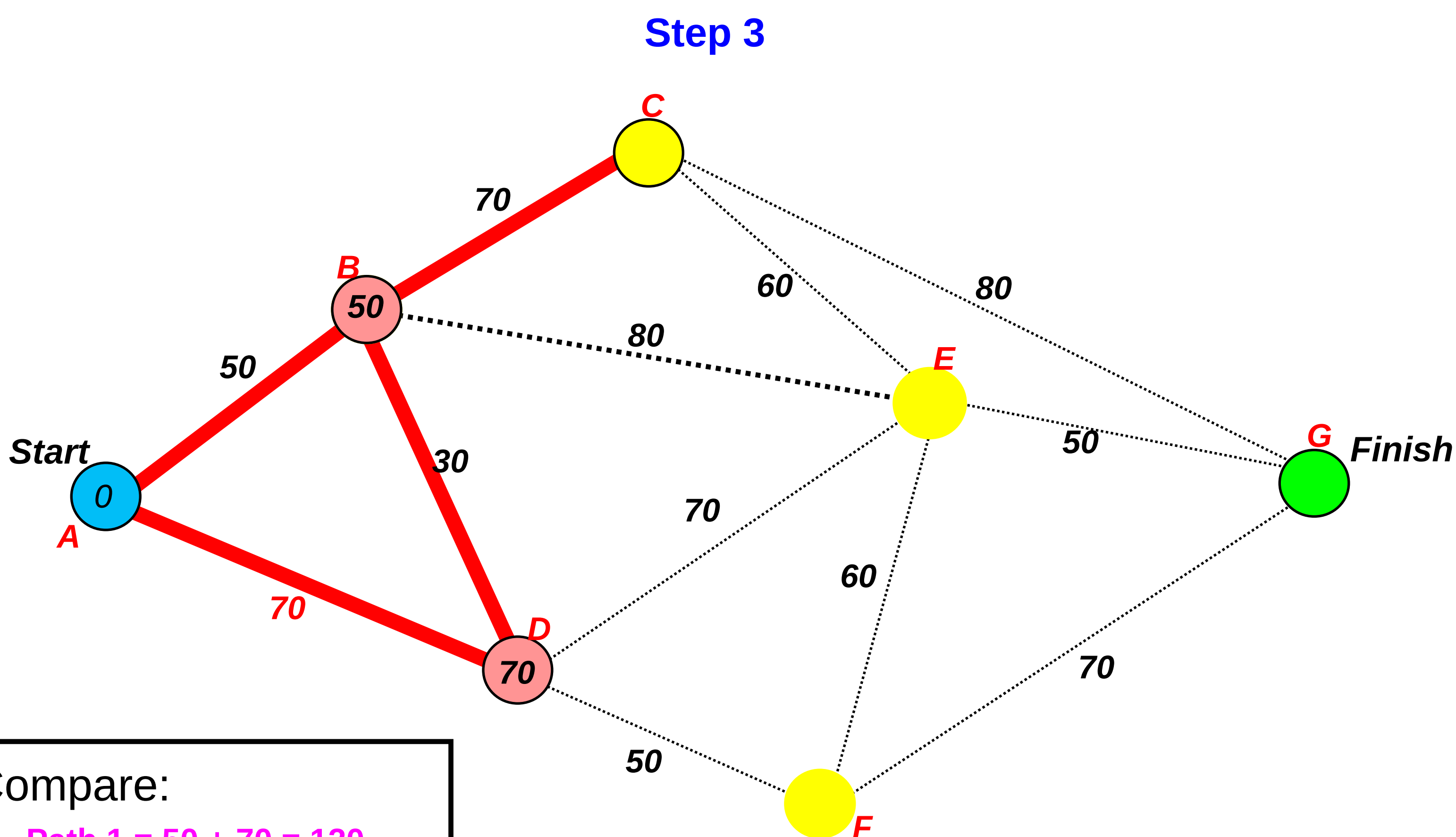
Compare:

Path 1 = 50

Path 2 = 70

Choose path 1

Example of Dijkstra Algorithm



Compare:

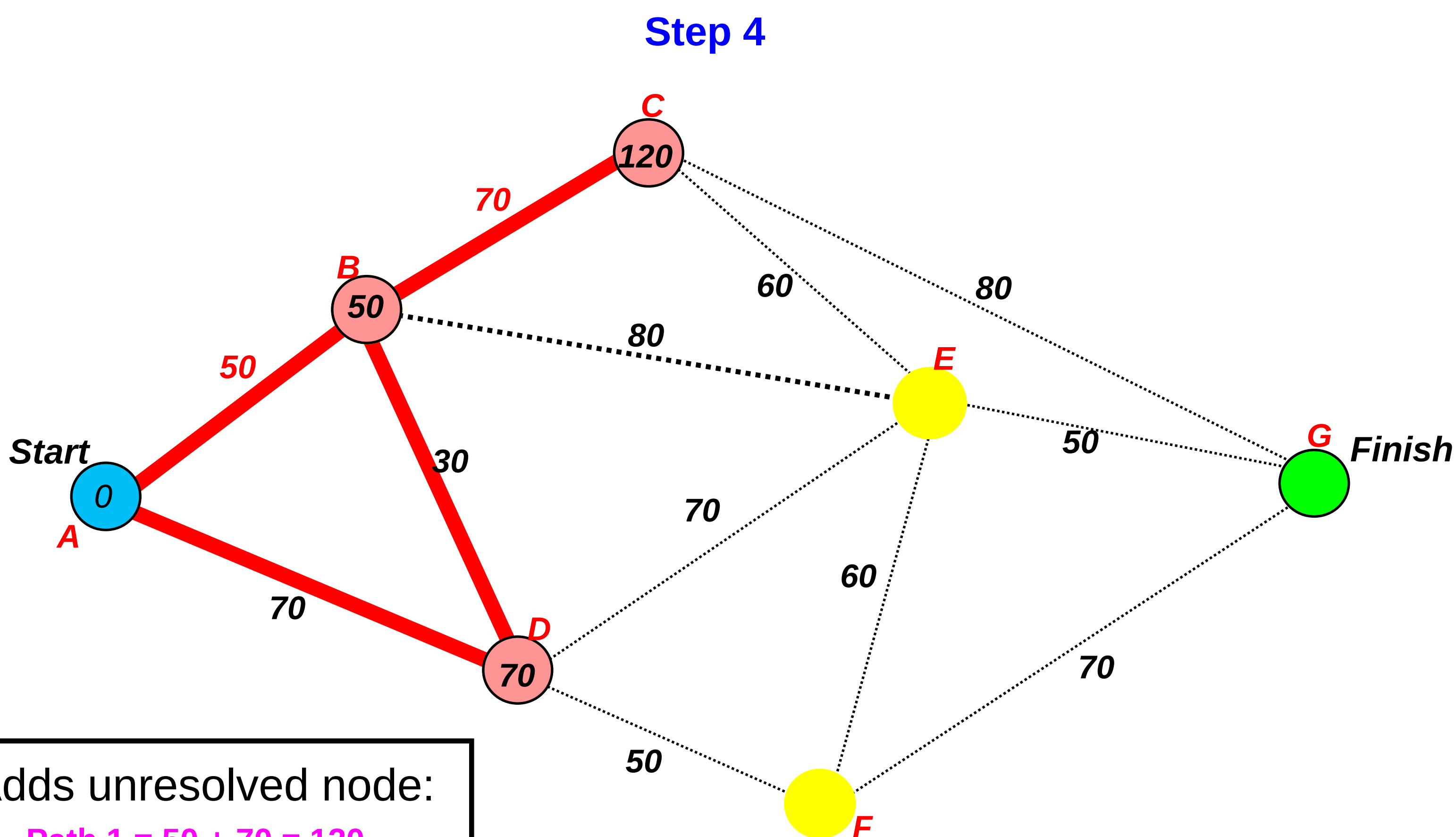
Path 1 = 50 + 70 = 120

Path 2 = 50 + 30 = 80

Path 3 = 70

Choose path 3

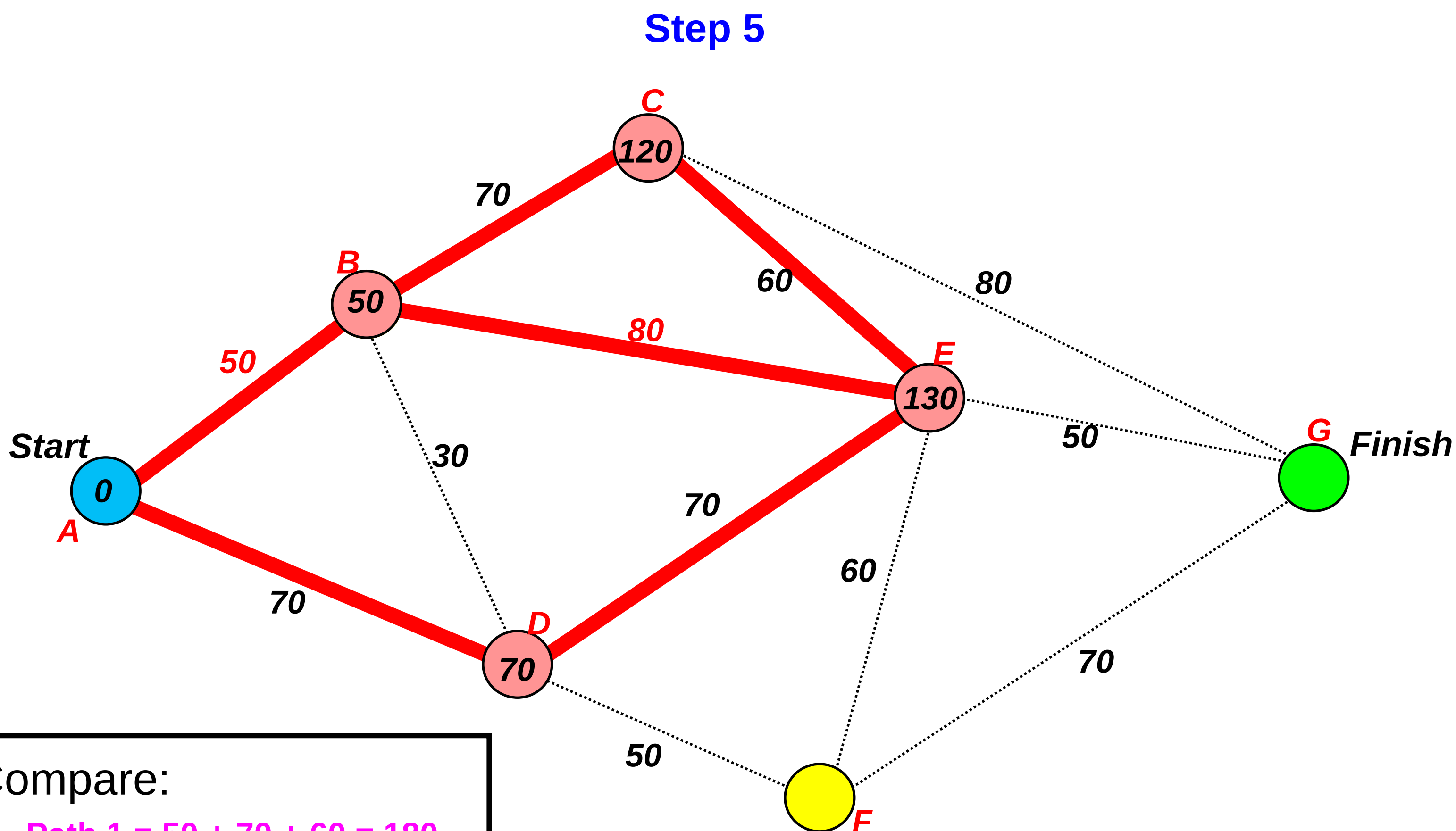
Example of Dijkstra Algorithm



Adds unresolved node:

Path 1 = 50 + 70 = 120

Example of Dijkstra Algorithm



Compare:

Path 1 = 50 + 70 + 60 = 180

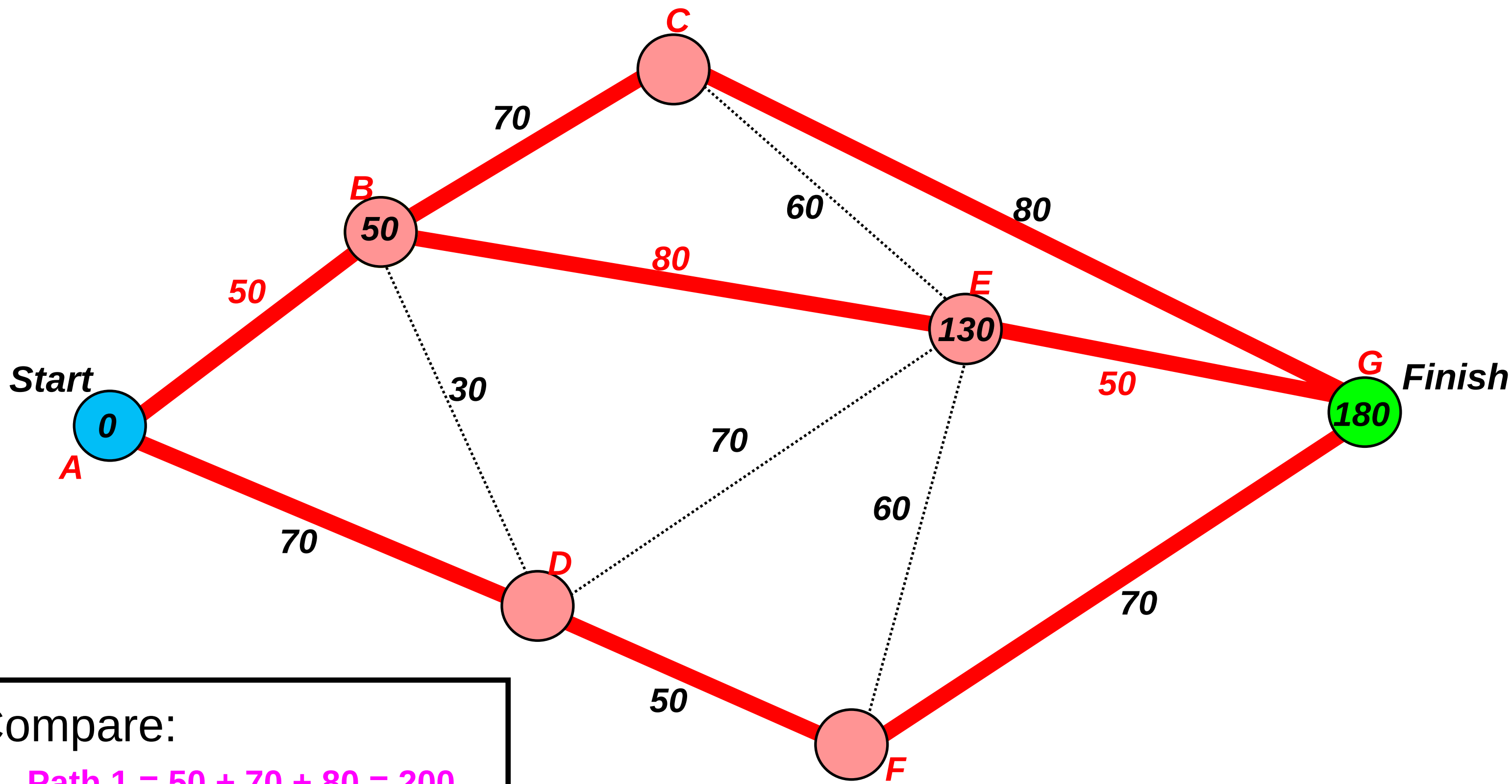
Path 2 = 50 + 80 = 130

Path 3 = 70 + 70 = 140

Choose path 2

Example of Dijkstra Algorithm

Step 6



Compare:

Path 1 = 50 + 70 + 80 = 200

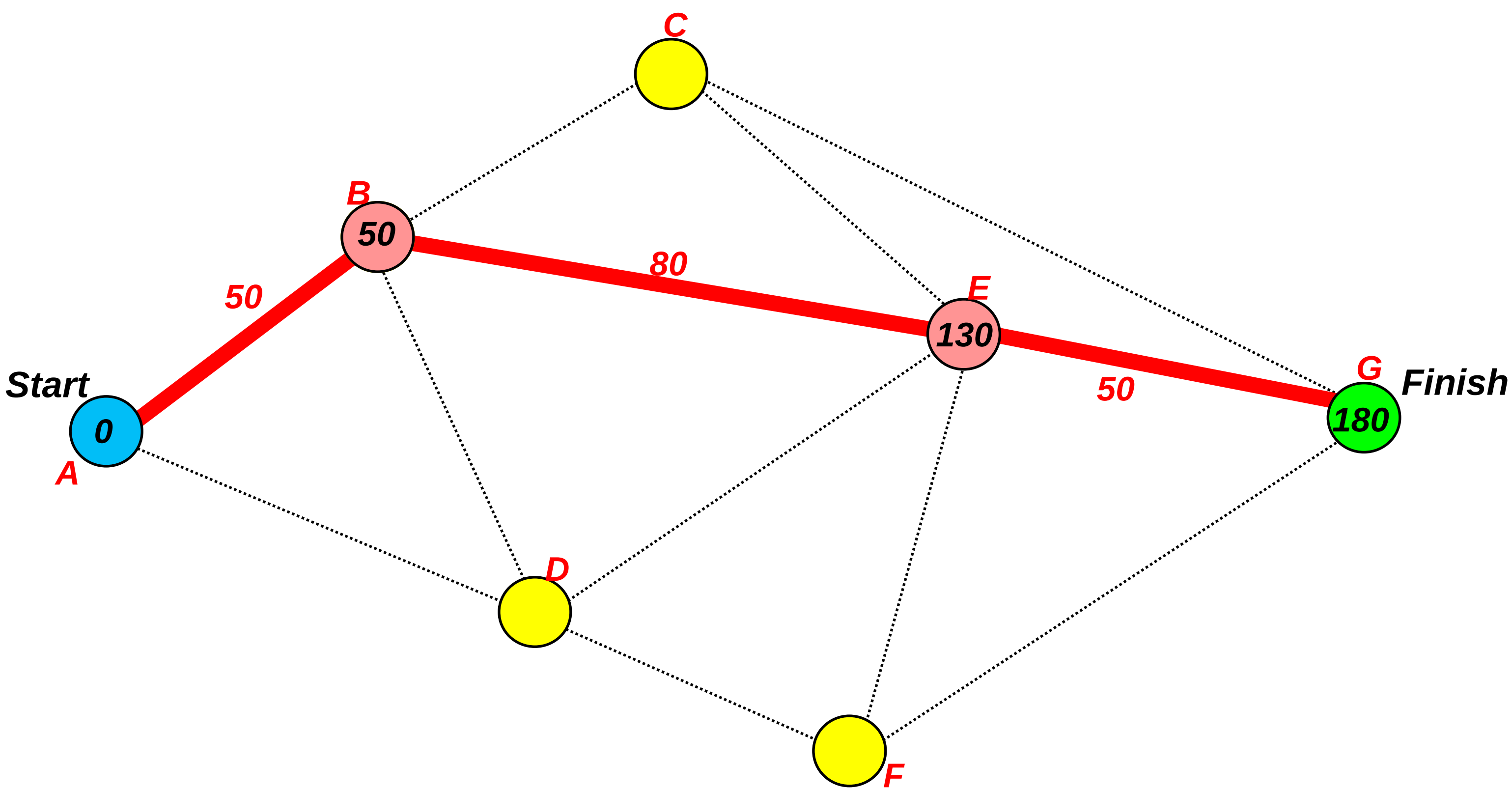
Path 2 = 50 + 80 + 50 = 180

Path 3 = 70 + 50 + 70 = 190

Choose path 2

Example of Dijkstra Algorithm

Shortest Path from Start to Finish



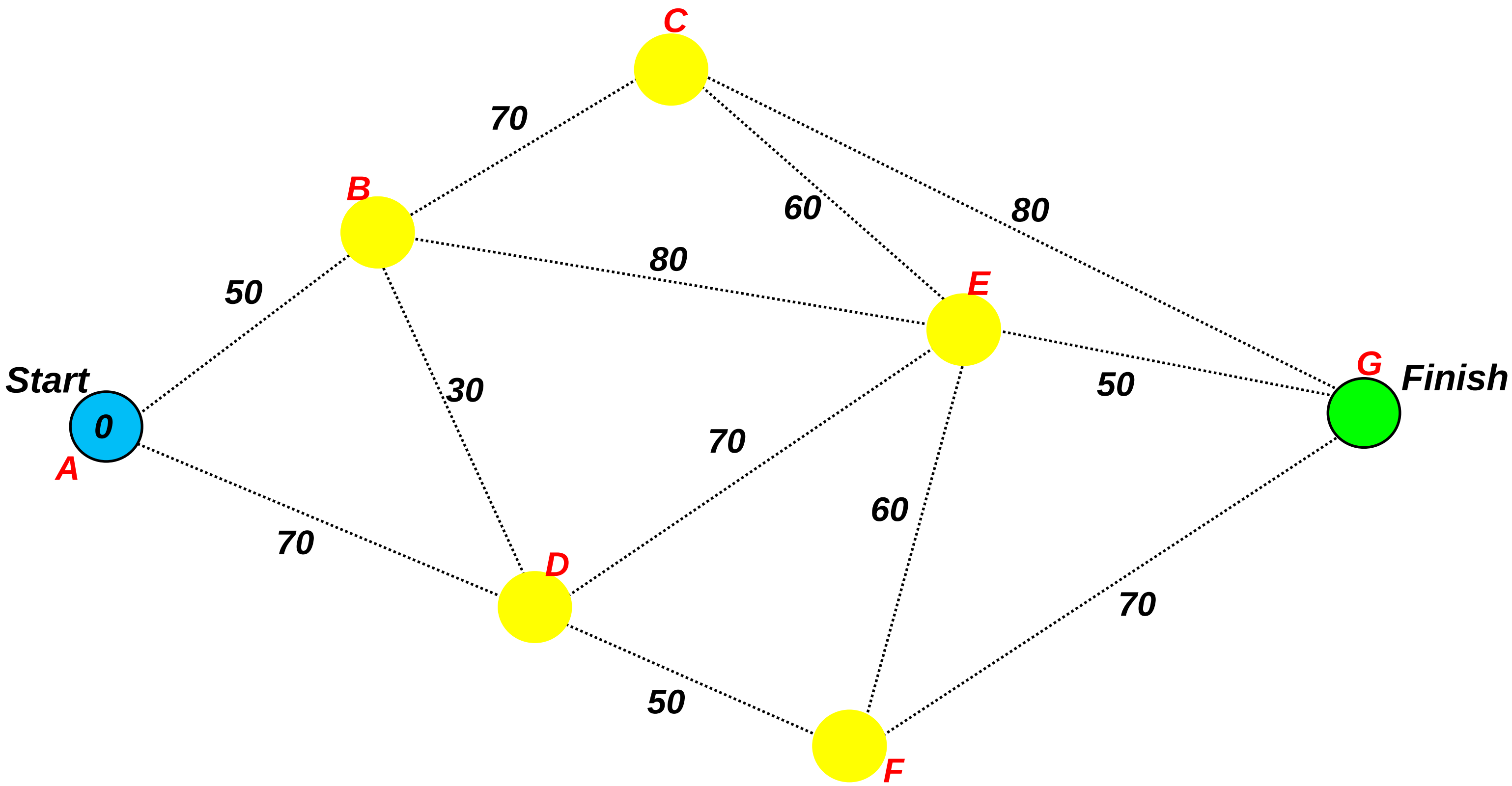
A* Shortest Path Algorithm

Combines Dijkstra Algorithm with Simplifying Assumption

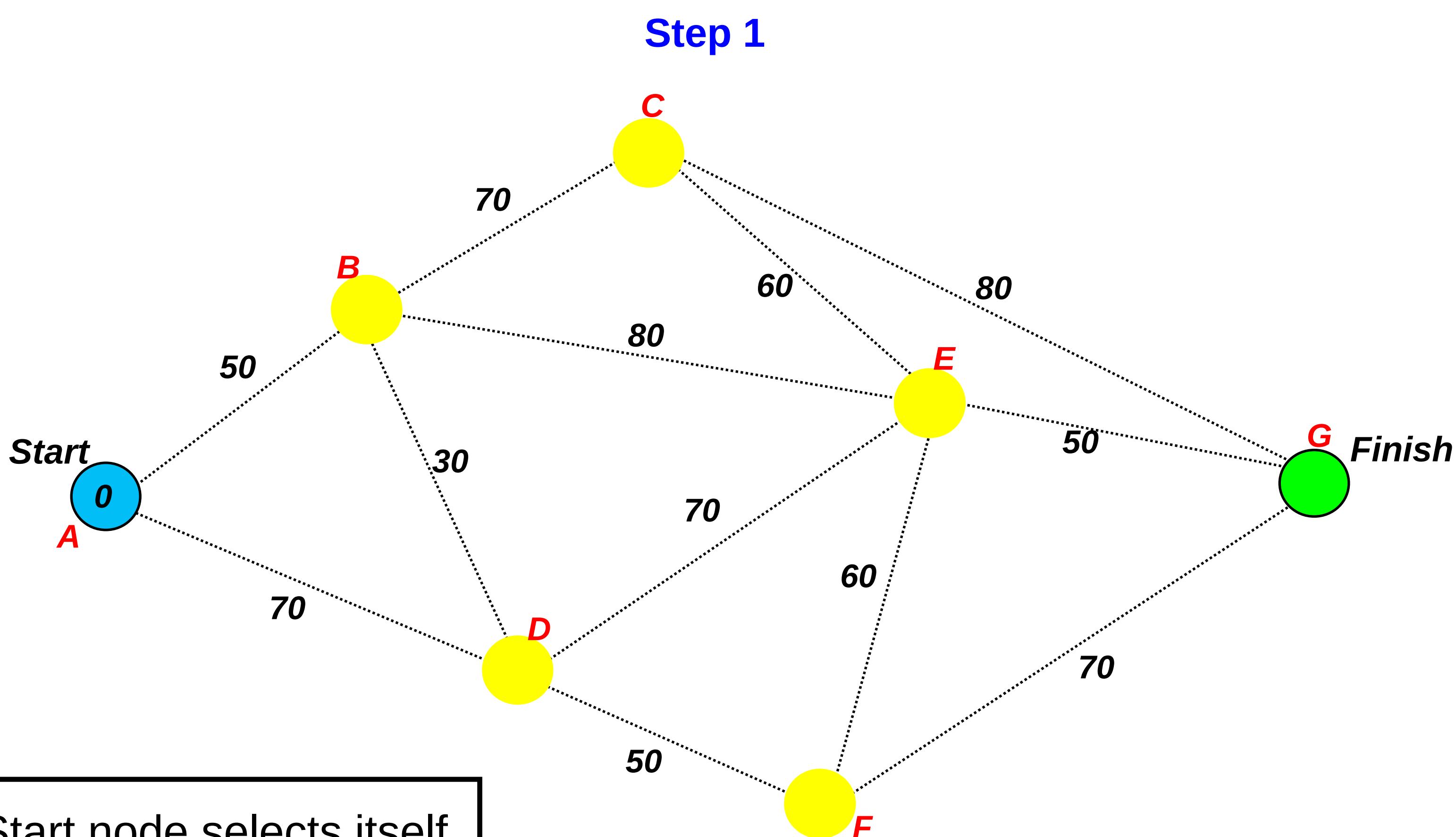
Path = Dijkstra distance + remaining distance

A* Modifies the Dijkstra Algorithm

Adding an Estimate of the Remaining Distance to the Dijkstra distance



A* Algorithm

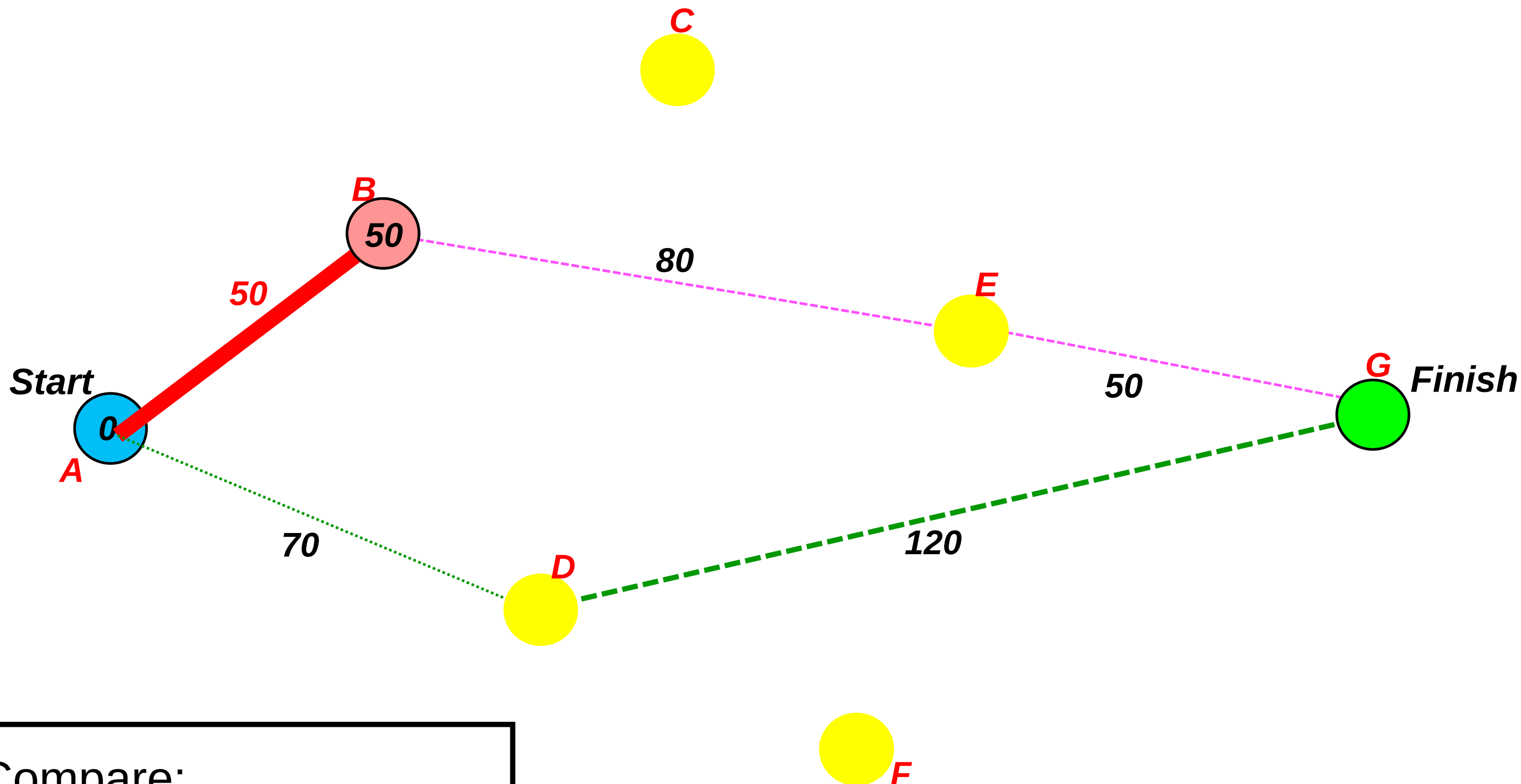


Start node selects itself

Path 1 = 0

A* Algorithm

Step 2



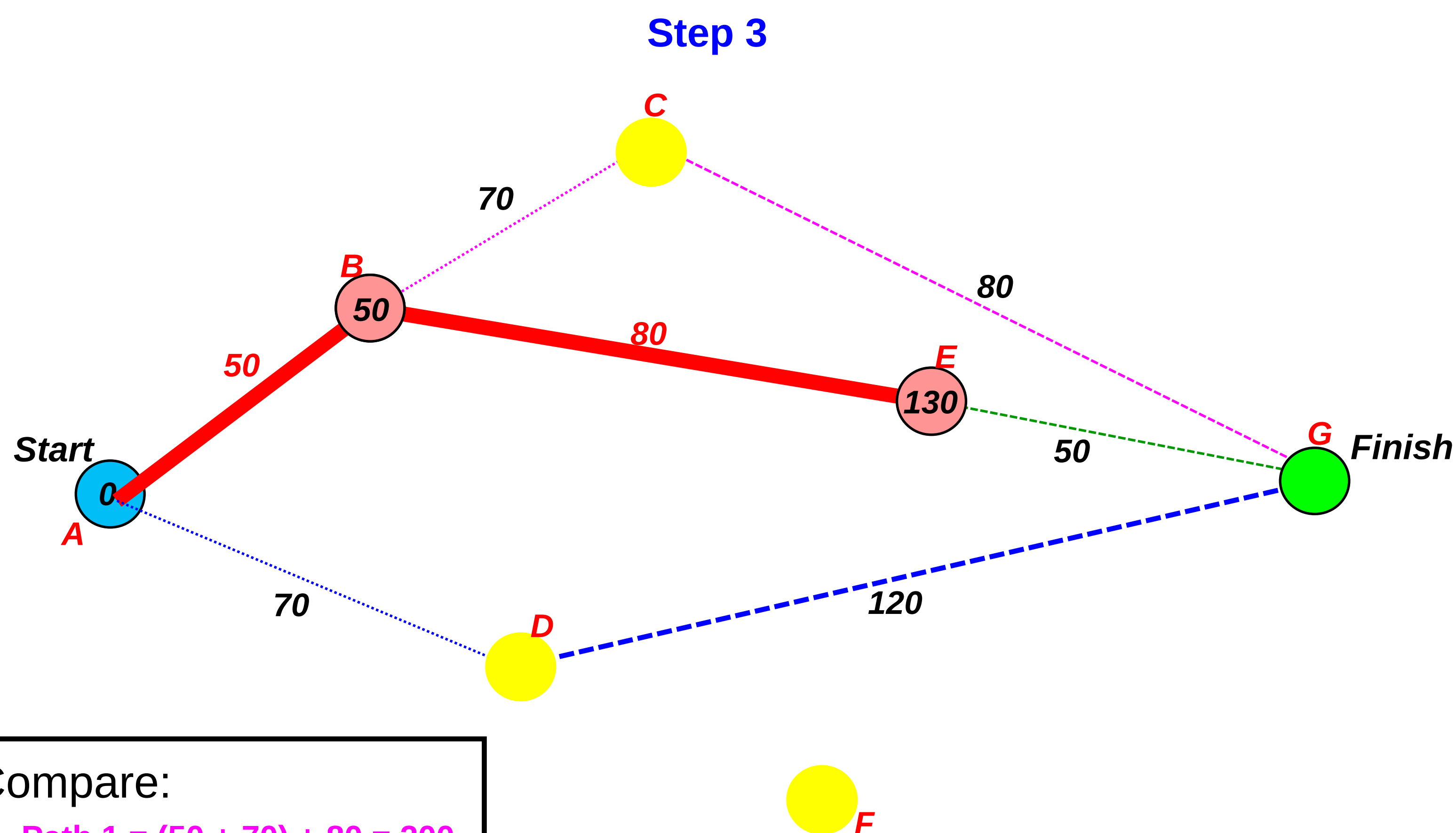
Compare:

Path 1 = 50 + 80 + 50 = 180

Path 2 = 70 + 120 = 190

Choose path 1

A* Algorithm



Compare:

Path 1 = (50 + 70) + 80 = 200

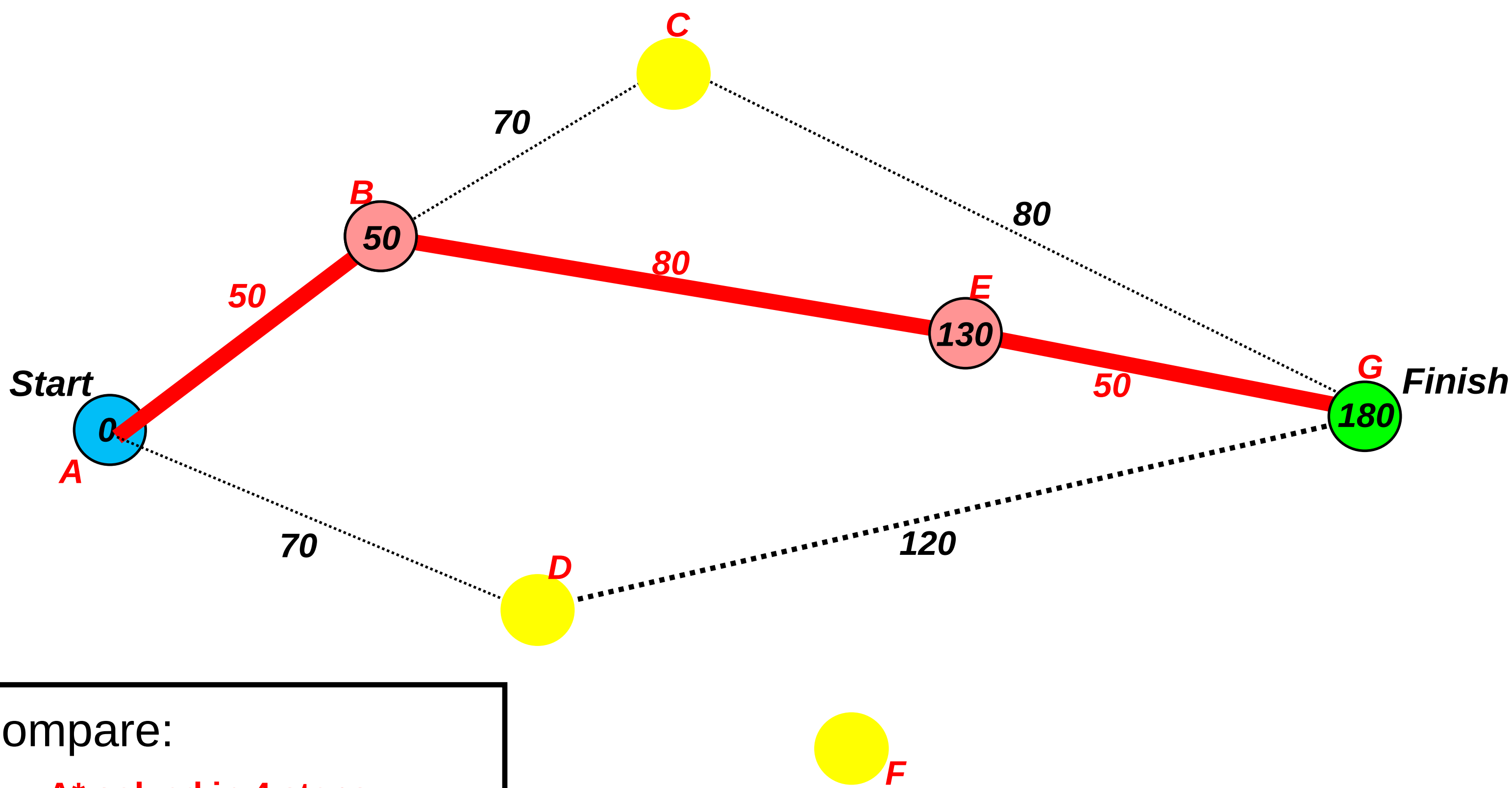
Path 2 = (50 + 80)+ 50 = 180

Path 3 = (70) + 120 = 190

Choose path 2

A* Algorithm

Finish



Compare:

A* solved in 4 steps

Dijkstra solved in 6 steps

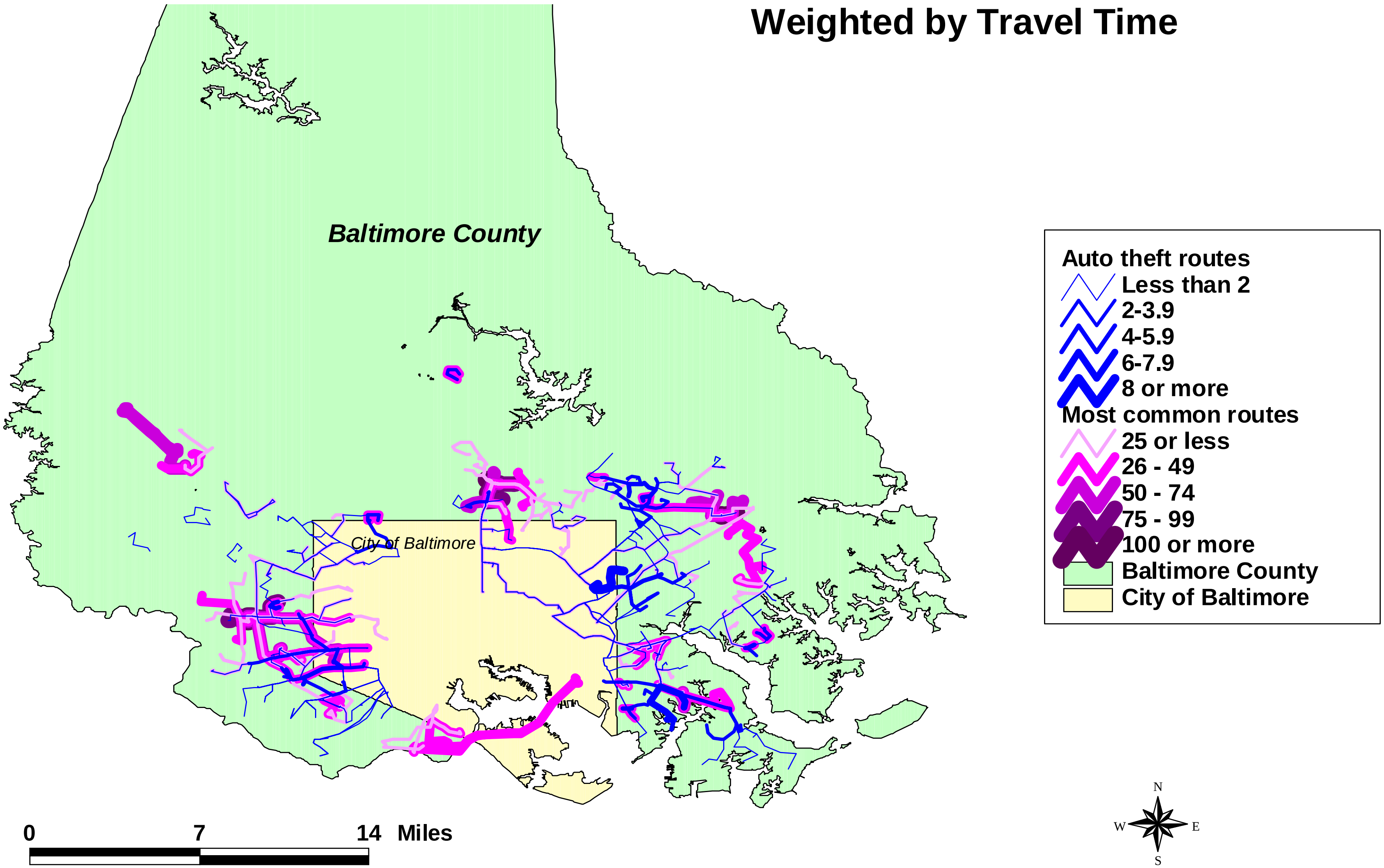
A* is More Efficient than Dijkstra

Calculating time proportional to:

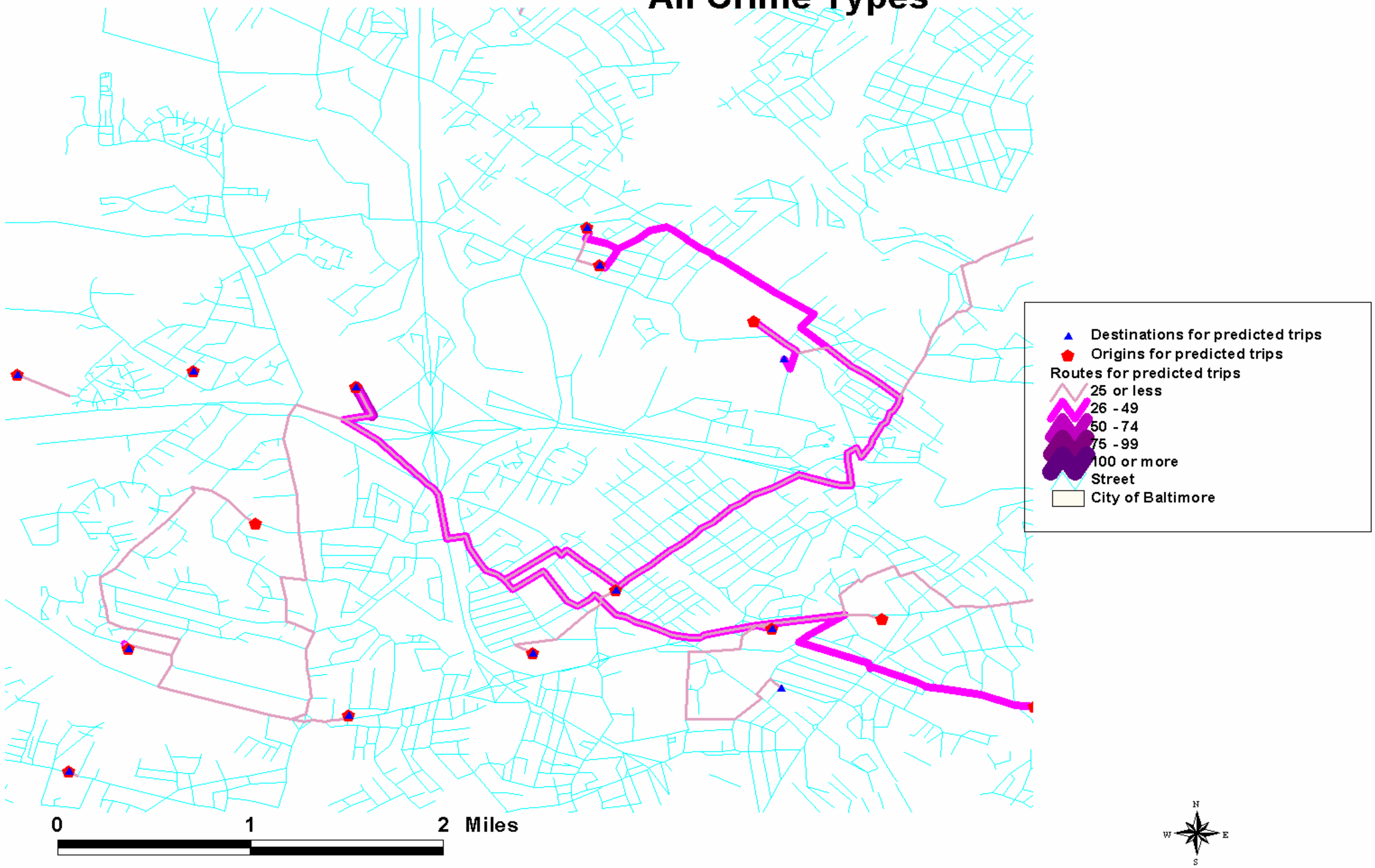
Dijkstra	V^2
A*	V

where V is the number of vertices

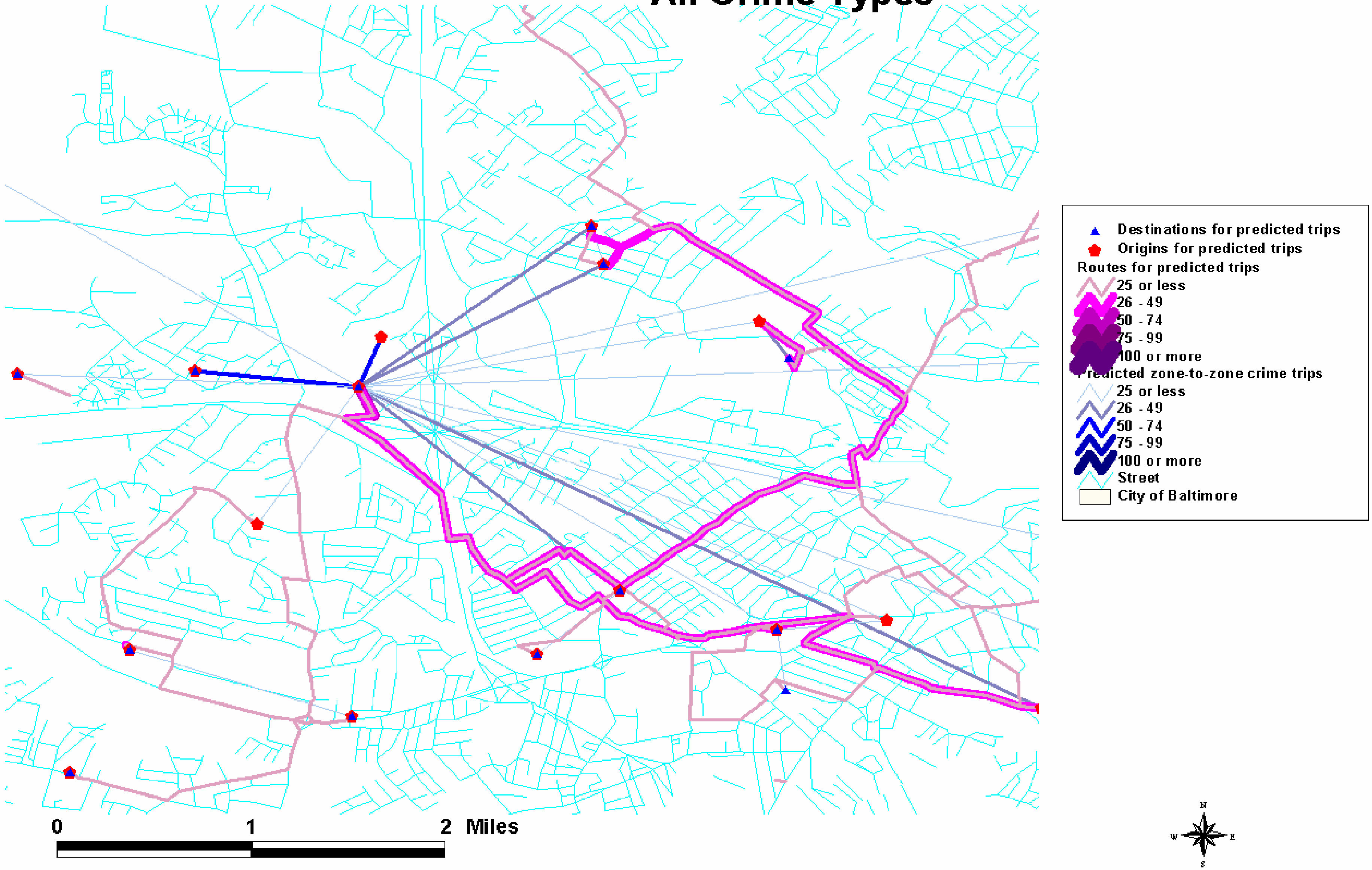
Predicted Routes by Crime Type: 1993-1997
All Crimes and Vehicle Thefts
Weighted by Travel Time



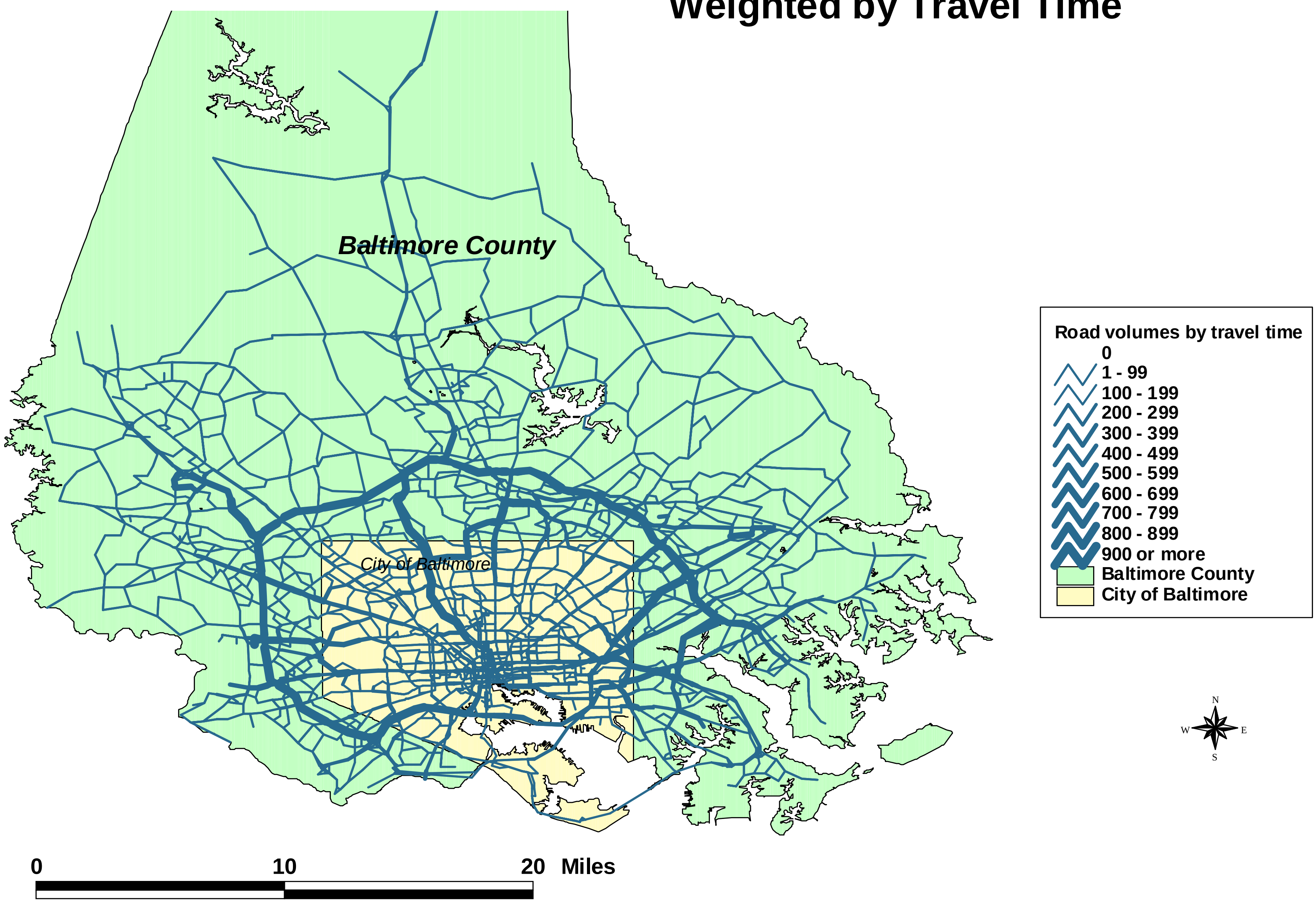
Predicted Baltimore County Crime Trips: 1993-1997
Links and Actual Routes for Zone-to-Zone Trips
All Crime Types



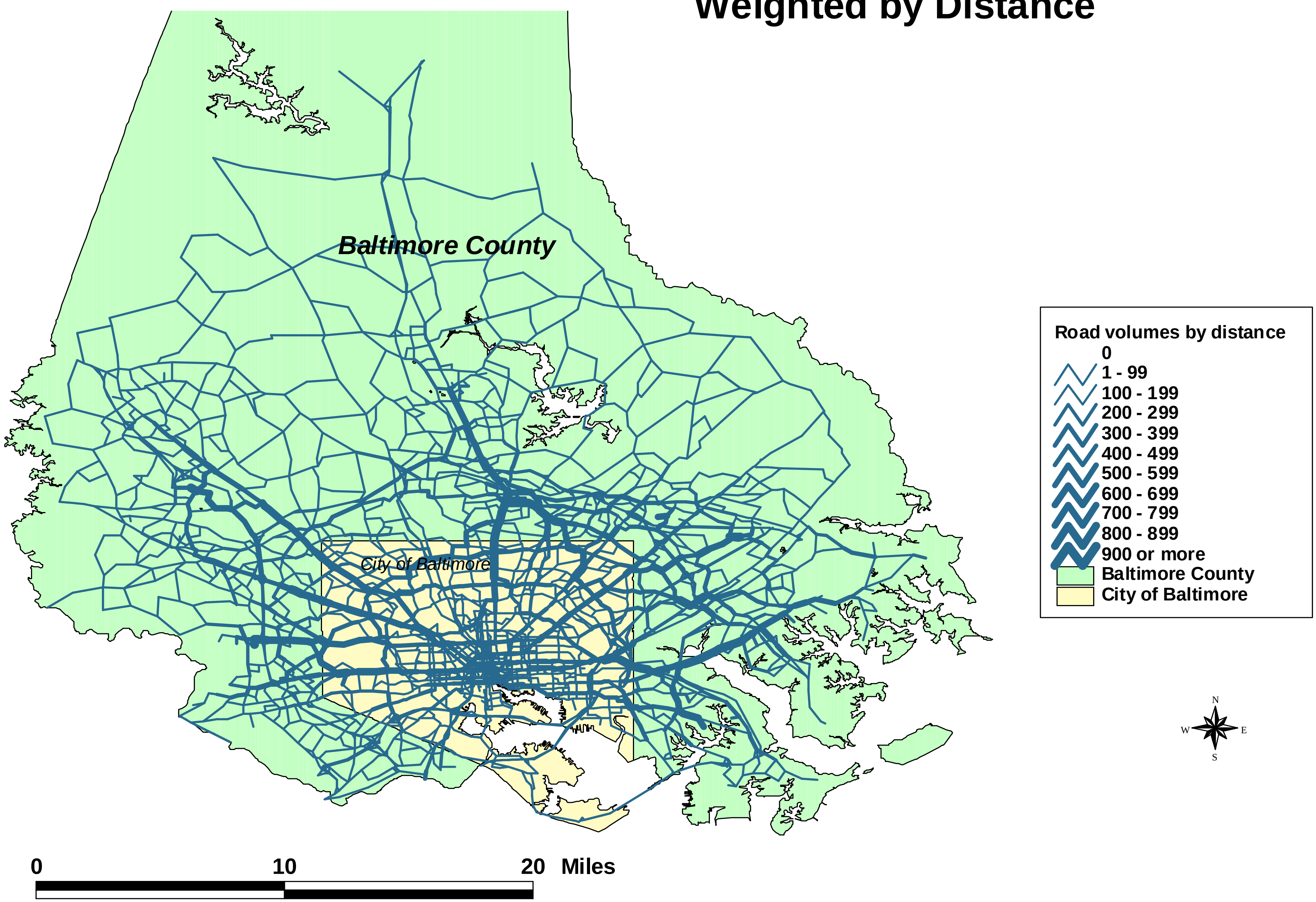
Predicted Baltimore County Crime Trips: 1993-1997
Links and Actual Routes for Zone-to-Zone Trips
All Crime Types



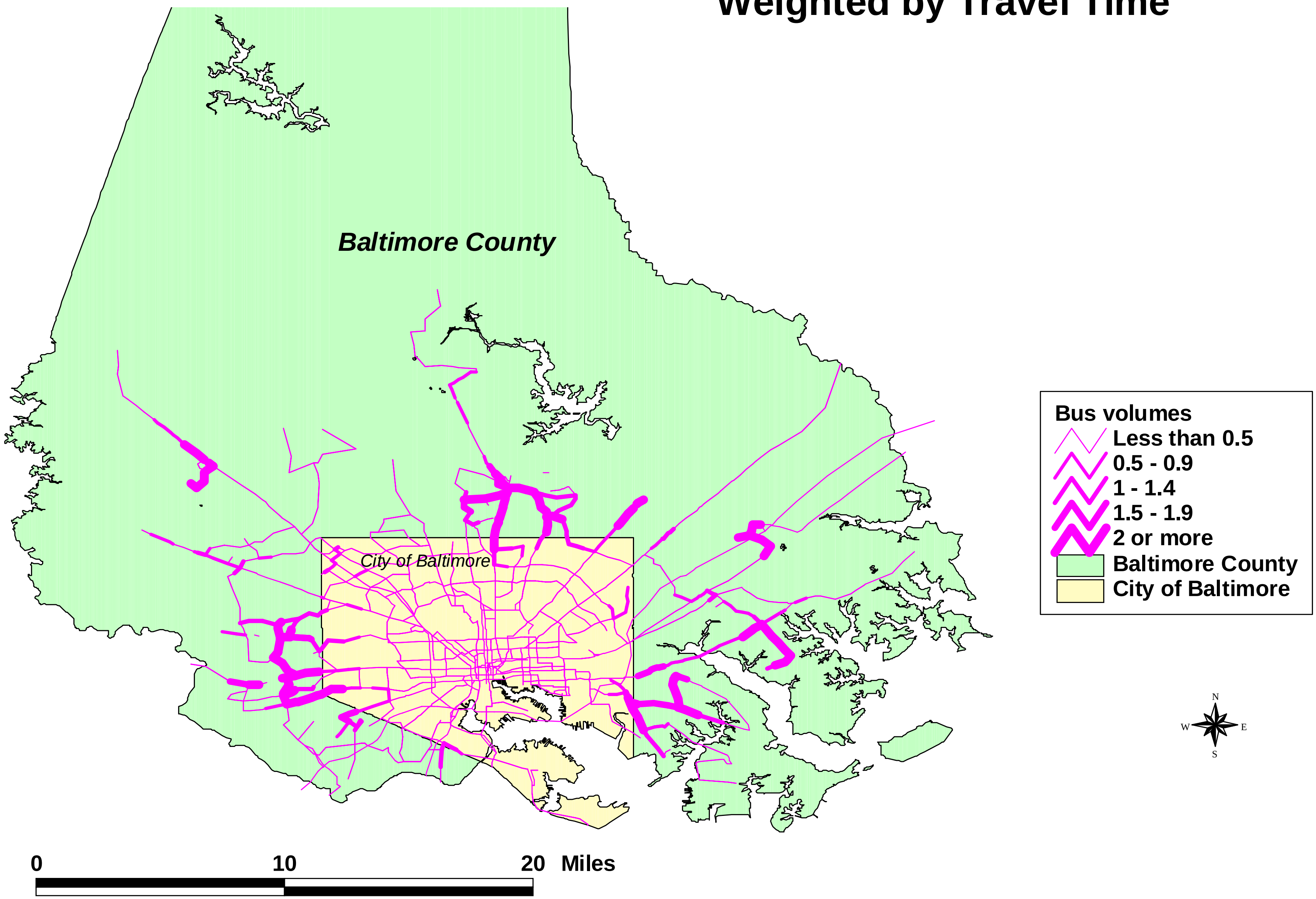
Crime Volume by Road Segment
Weighted by Travel Time



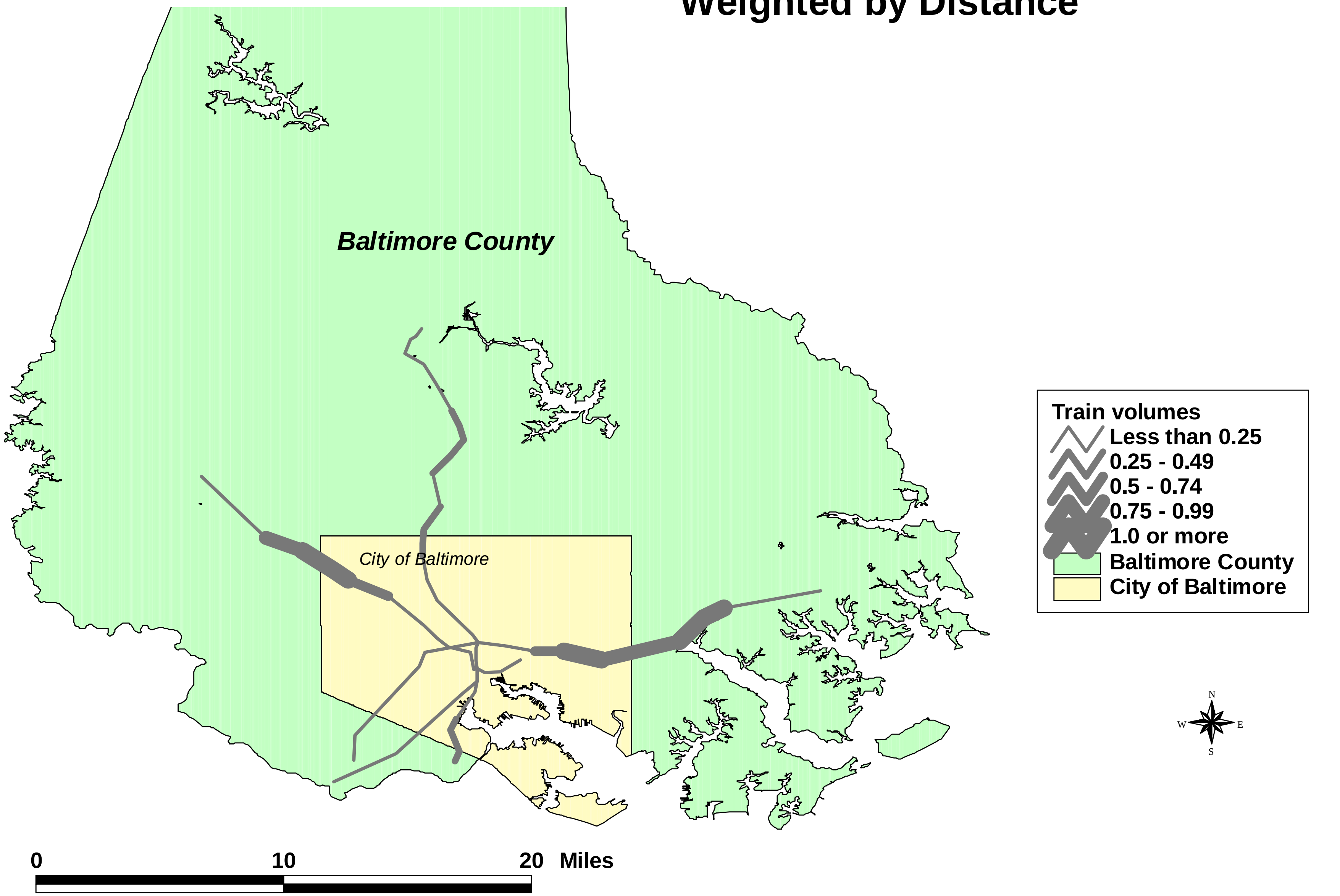
Crime Volume by Road Segment Weighted by Distance



Crime Volume by Bus Route Segment Weighted by Travel Time



Crime Volume by Rail Segment Weighted by Distance



Applications of
Crime Travel Demand Modeling

Modeling Policy Interventions

Some Ways to Model Interventions

- **Reduce the number of events at the origins**
  (e.g., selectively intervening in certain zones)

  Most fundamental, but probably most difficult to achieve.

- **Reduce the number of events at the destinations**
  (e.g., increasing deployment in certain zones)

  May lower levels but may displace crime to other zones

- **Increase impedance for mode split**
  (e.g., increasing surveillance on transit)

  Will reduce use of mode

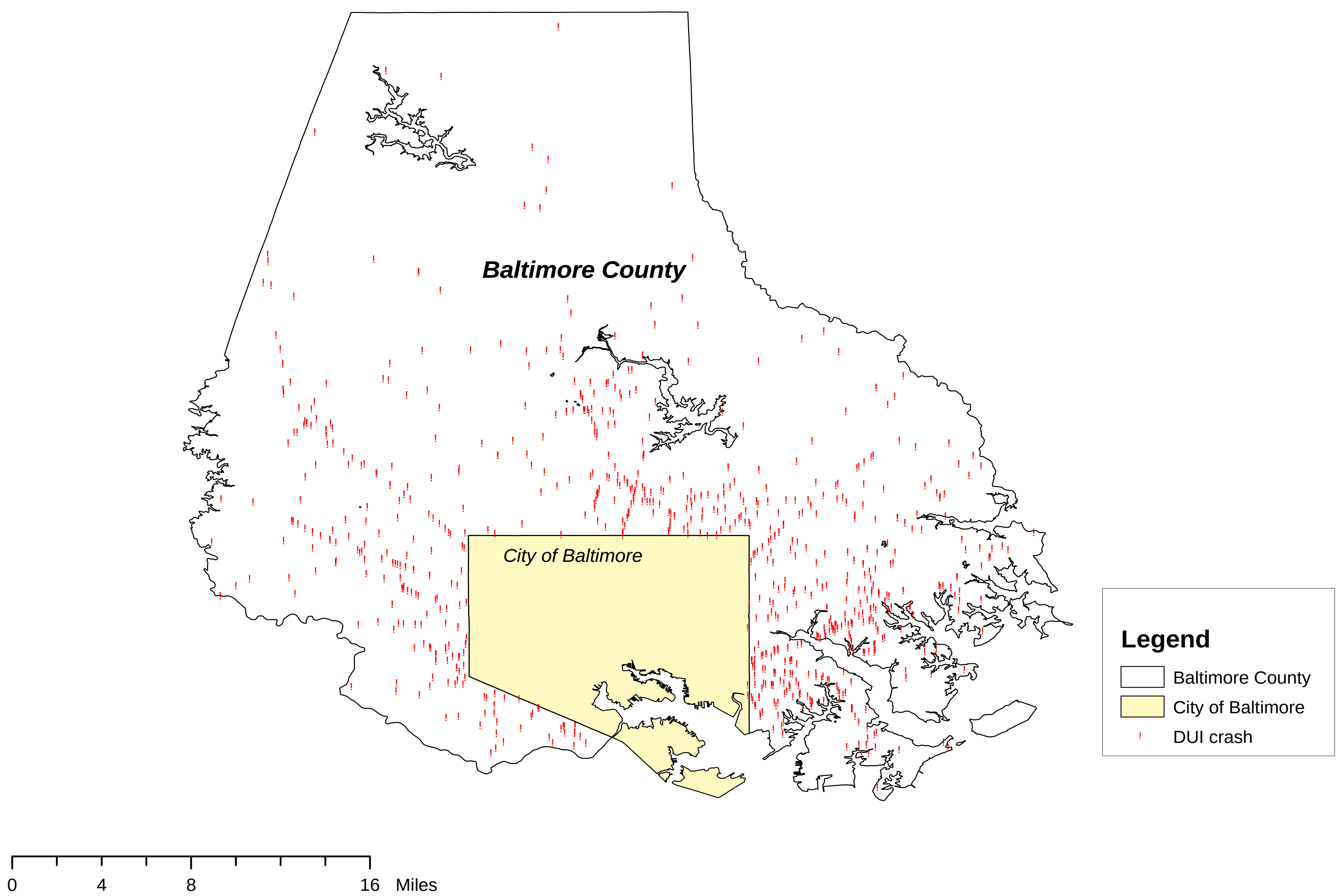
- **Increase the impedance value in network assignment**
  (e.g., increasing patrols along certain routes)

  Will probably displace routes, but not necessarily trips

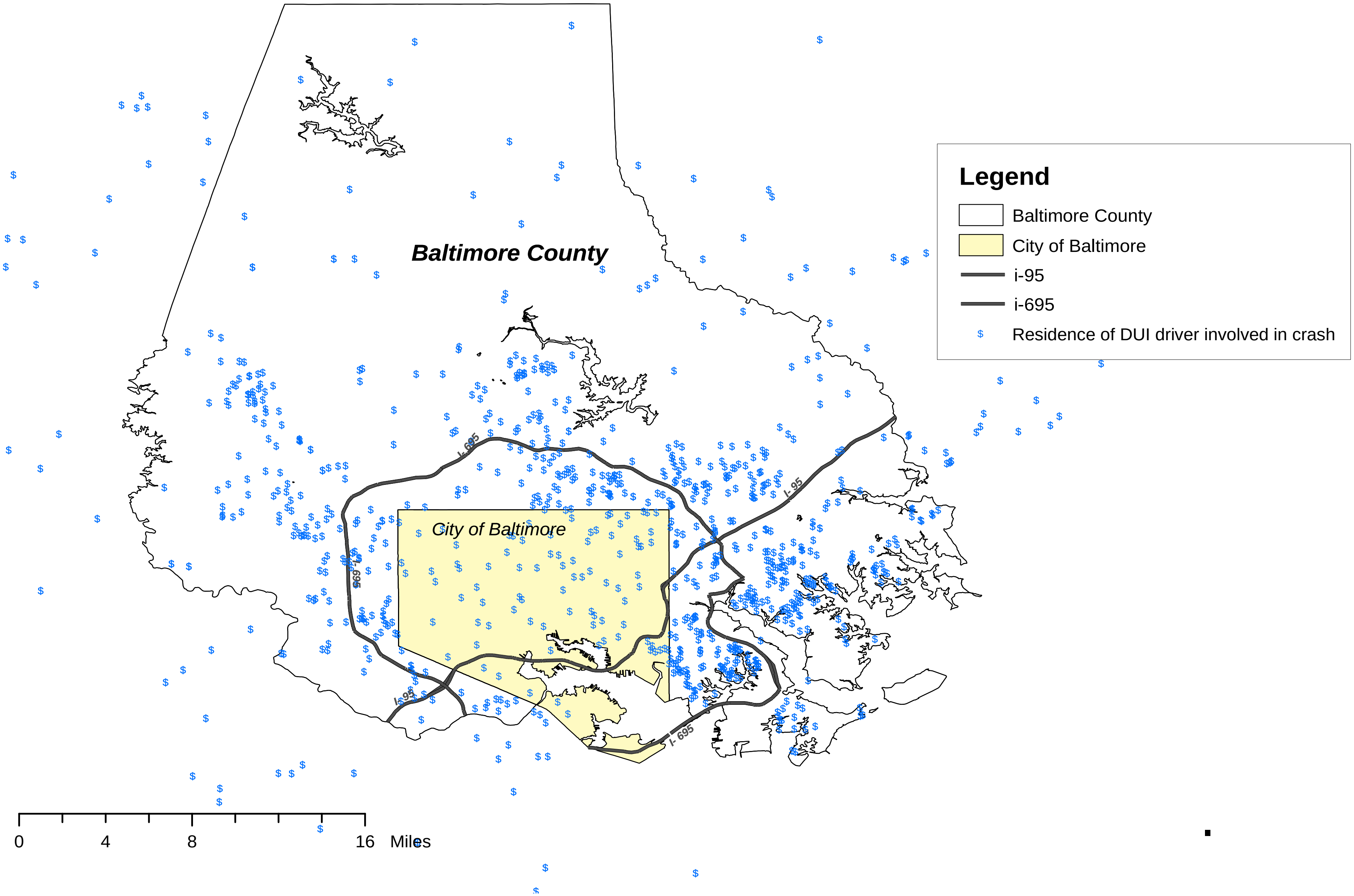
Purpose of Project

- **Examine *Driving-Under-The-Influence* (DUI) trips that end in motor vehicle crashes**
- **Model predictors of DWI crash trips with goal of intervening to reduce the number of crashes through:**
  - **DUI citations**
  - **Targeting high-risk individuals and communities**
  - **Making engineering improvements at selective crash hot spots**

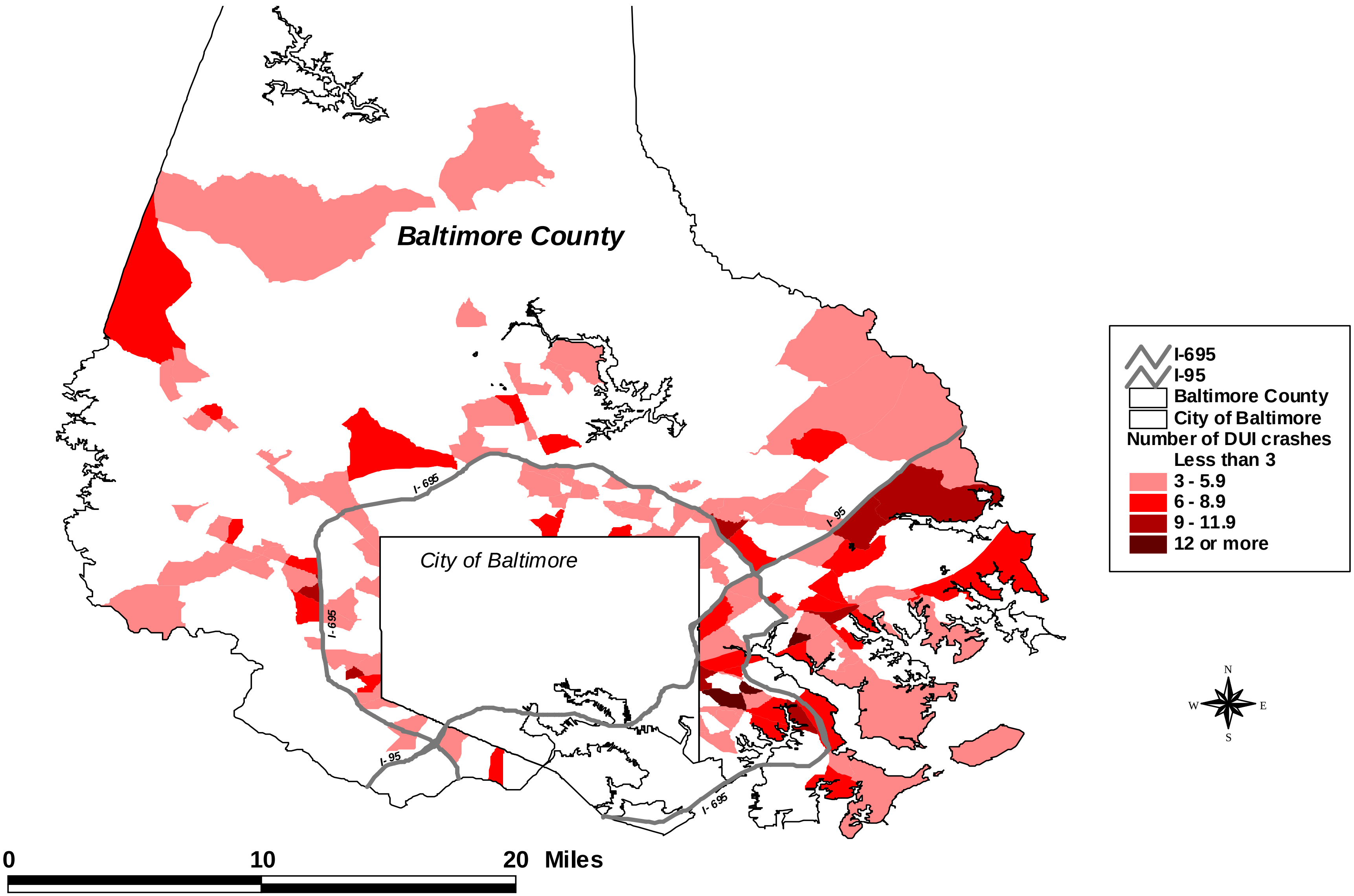
DUI Crashes: 1999-2001



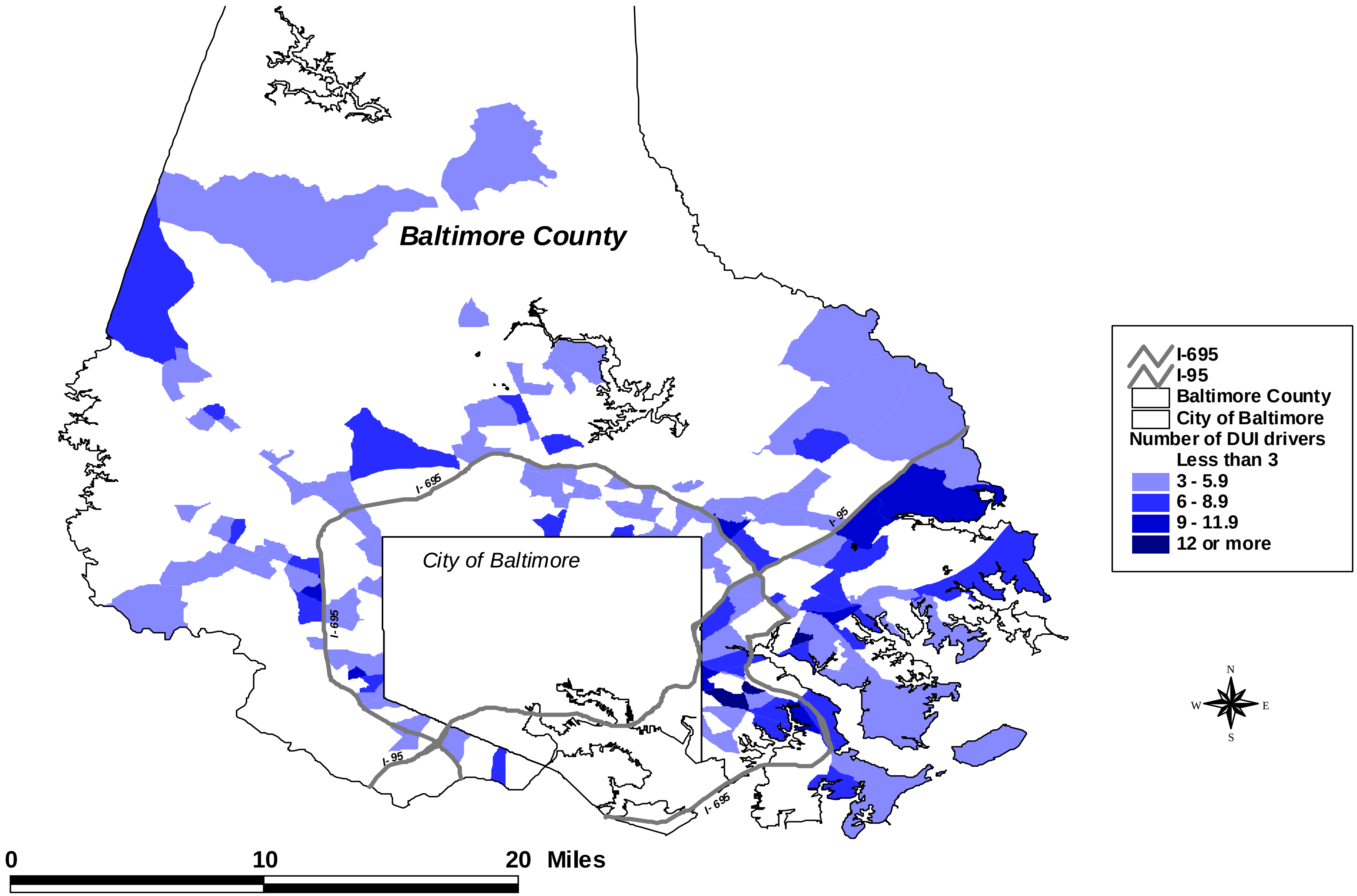
Residences of DUI Crash Offenders: 1999-2001



DUI Crashes by Traffic Analysis Zones: 1999-2001
Location of DUI Crashes



DUI Crashes by Traffic Analysis Zones: 1999-2001
Location of Residences of DUI Drivers Involved in Crashes



Predicting Annual DUI Crash Origins

```
Model result:
 Data file:                          DUI origins.dbf
 Type of model:                      Origin
 DepVar:                              ANNUAL DUI CRASH OFFENDERS BY RESIDENCE
 N:                                    534
 Df:                                   527
 Type of regression model:           Poisson with over-dispersion correction
 Log Likelihood:                     -437.538627
 Likelihood ratio(LR):               137.319678
 P-value of LR:                      0.0001
 AIC:                                889.077254
 SC:                                 919.040025
 Dispersion multiplier:              1.000000
 R-square:                           0.310621
 Deviance r-square:                  0.704103
```

Predictor	DF	Coefficient	Stand Error	Pseudo-Tolerance	z-value	p-value
CONSTANT	1	-2.254270	0.087658	.	-25.716687	0.001
POPULATION	1	0.000172	0.000020	0.871576	8.591754	0.001
PCT WHITE	1	0.011731	0.001267	0.762753	9.257638	0.001
RURAL	1	0.437638	0.085437	0.555970	5.122360	0.001
# LIQUOR STORES	1	0.374038	0.049077	0.829191	7.621515	0.001
# BARS	1	0.160486	0.027467	0.818204	5.842872	0.001
AREA OF ZONE	1	-0.007304	0.010644	0.556103	-0.686193	n.s.

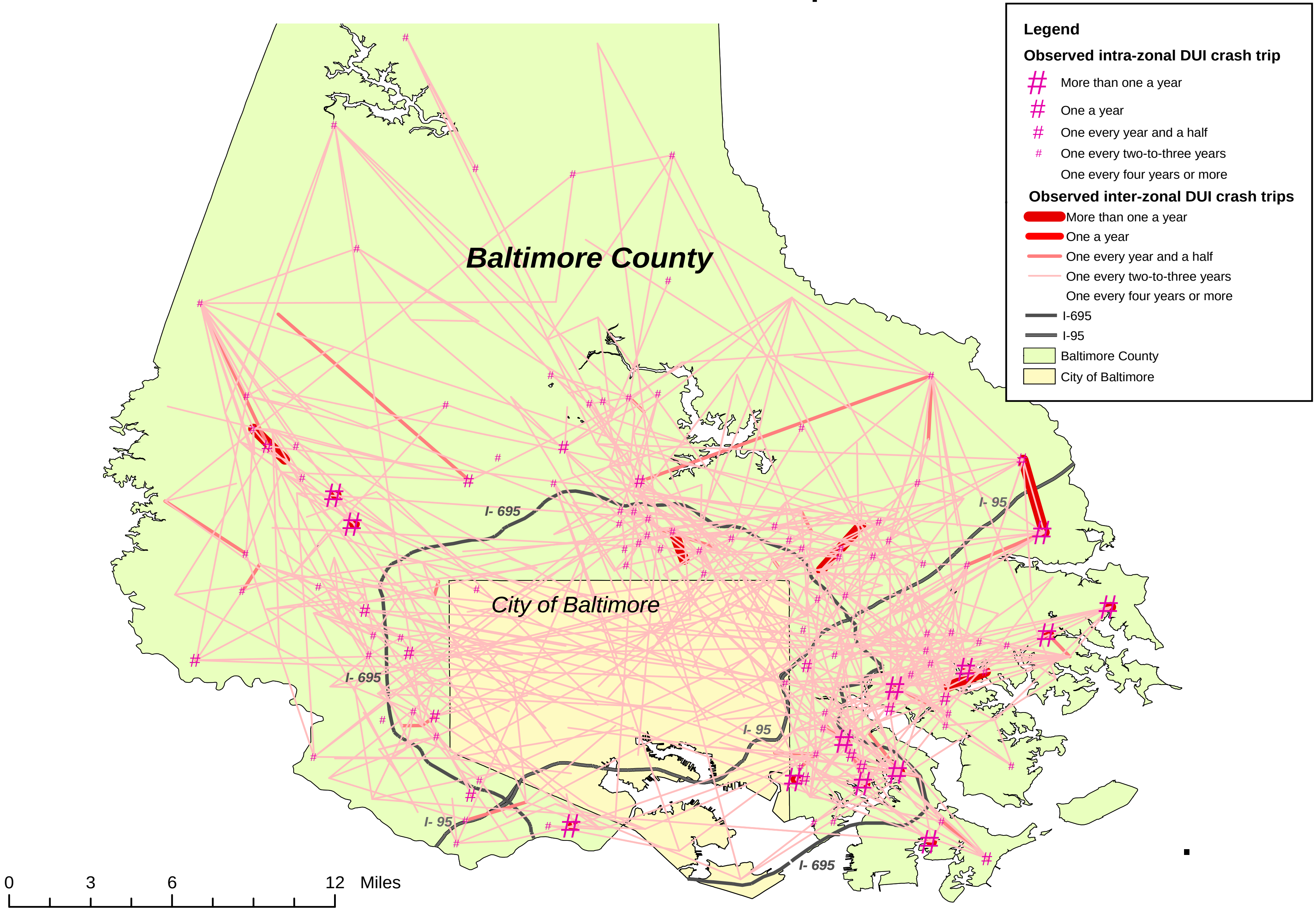
Predicting Annual DUI Crash Destinations

Model result:

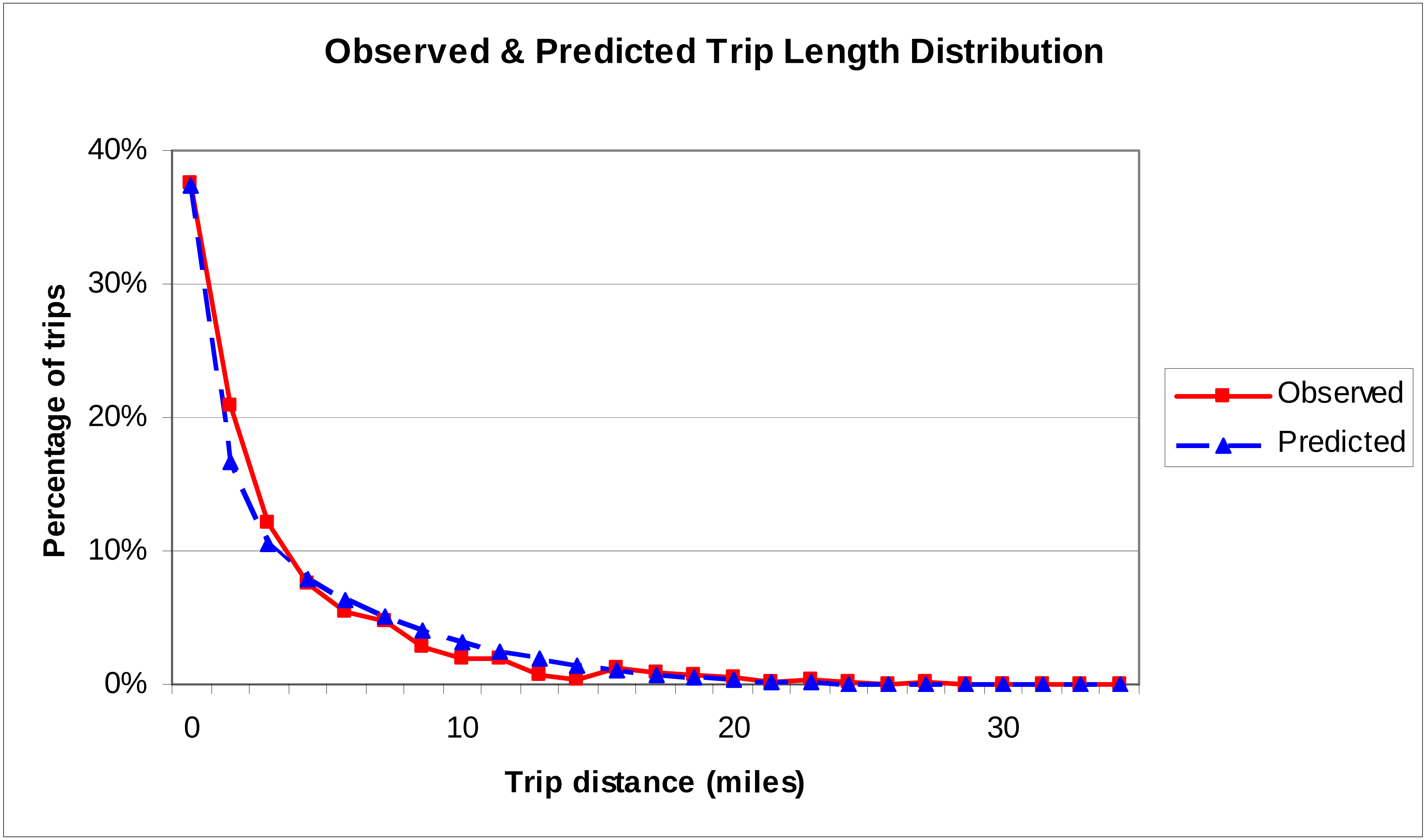
Data file:	DUI destinations.dbf
Type of model:	Destination
DepVar:	ANNUAL DUI CRASHES BY LOCATION
N:	325
Df:	320
Type of regression model:	Poisson with over-dispersion correction
Log Likelihood:	-337.989582
Likelihood ratio(LR):	59.034019
P-value of LR:	0.0001
AIC:	685.979165
SC:	704.898291
Dispersion multiplier:	1.000000
R-square:	0.251436
Deviance r-square:	0.763832

Predictor	DF	Coefficient	Stand Error	Pseudo-Tolerance	z-value	p-value
CONSTANT	1	-0.498485	0.078358	.	-6.361614	0.001
PCT OF ZONE IN RESIDENTIAL BLDG OF 10 OR MORE	1	0.004531	0.001539	0.950403	2.944569	0.010
# LIQUOR STORES	1	0.233486	0.050040	0.895238	4.665990	0.001
# BARS	1	0.192534	0.024990	0.893875	7.704392	0.001
AREA OF ZONE	1	-0.019994	0.014486	0.950081	-1.380286	n.s.

Observed DUI Crash Trips: 1999-2001
Annual Number of Trips

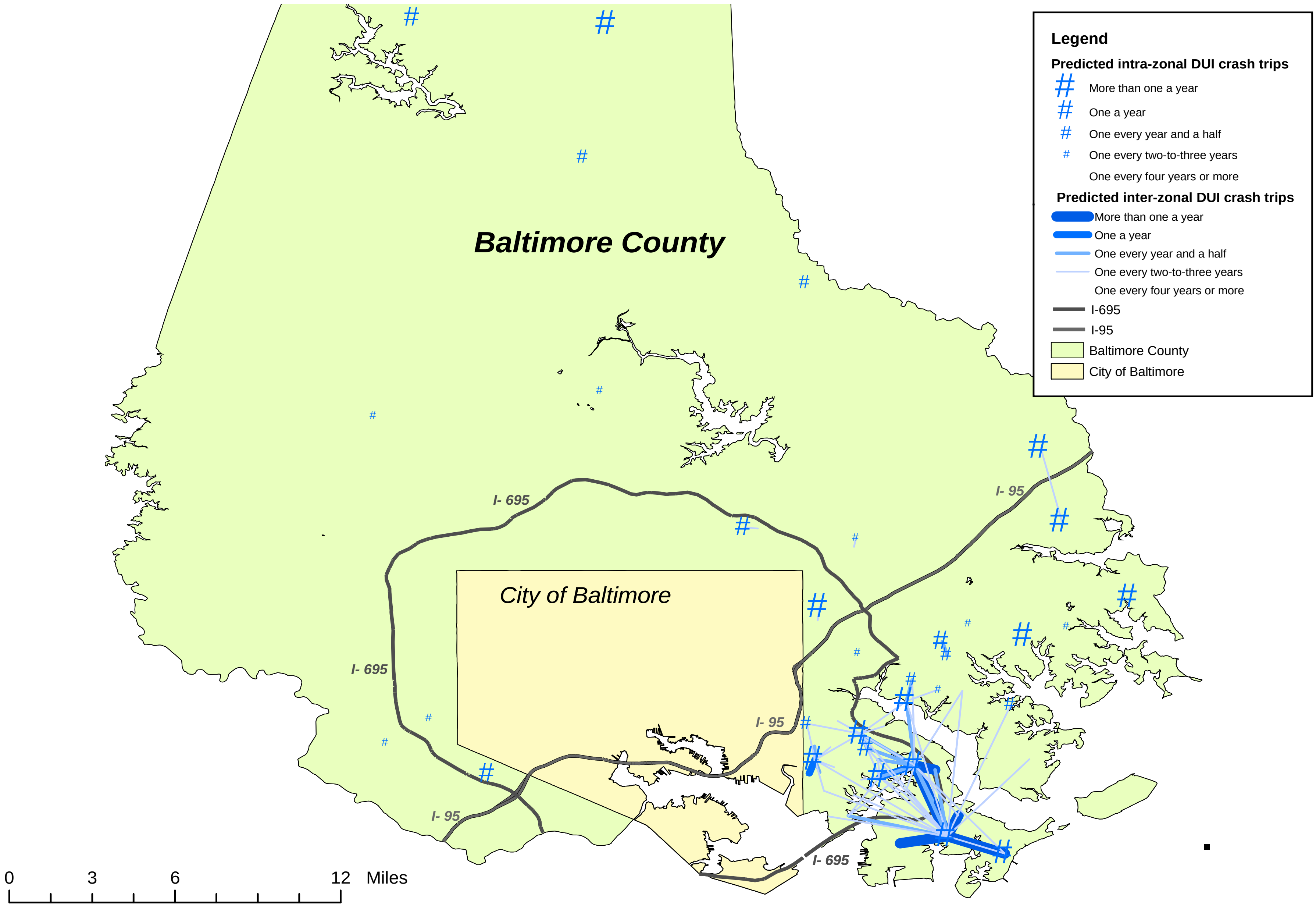


Best Impedance Function = Lognormal

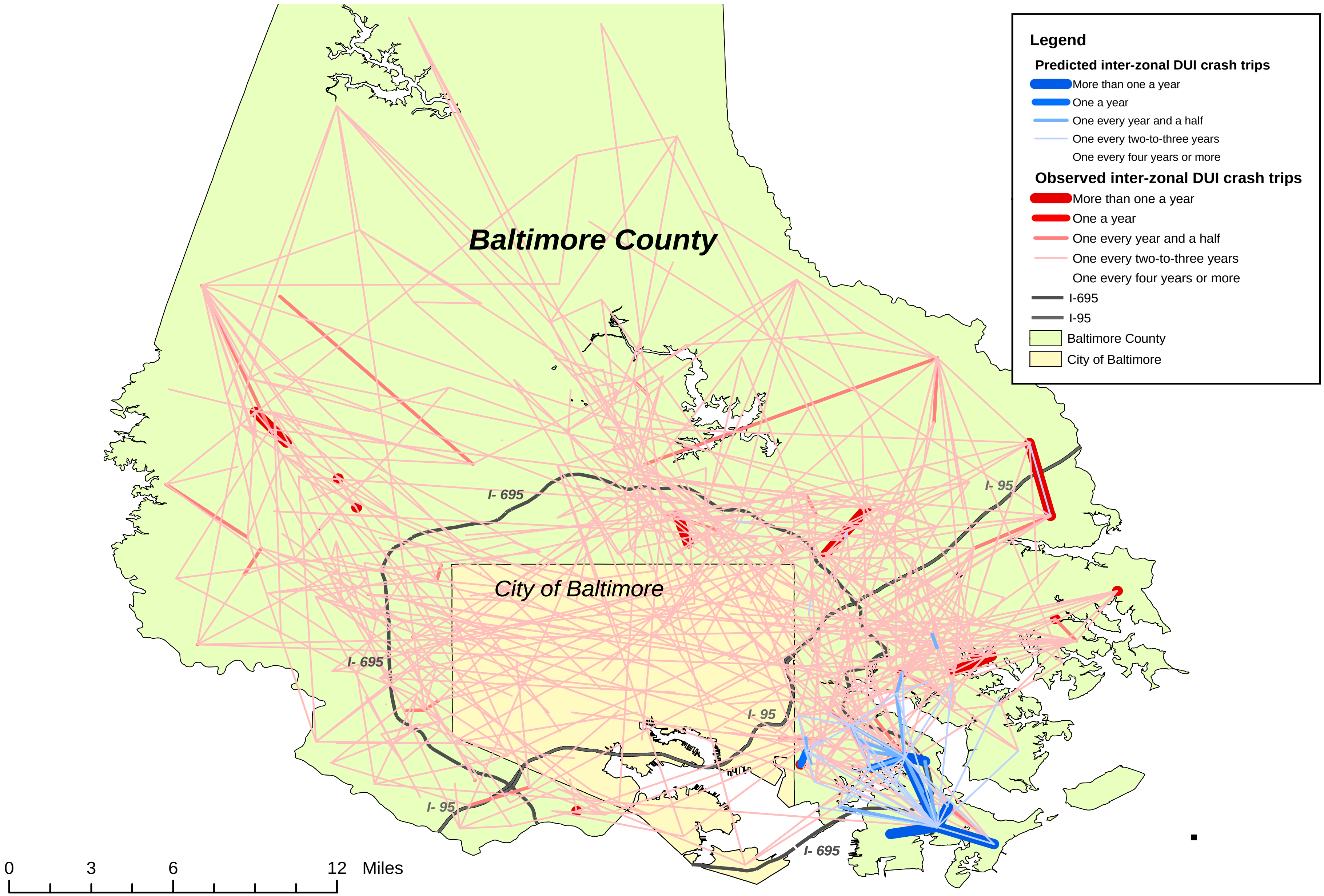


Observed & Predicted Trip Length Distribution
40%
30%
20%
10%
0%
Percentage of trips
0
10
20
30
Trip distance (miles)
Observed
Predicted

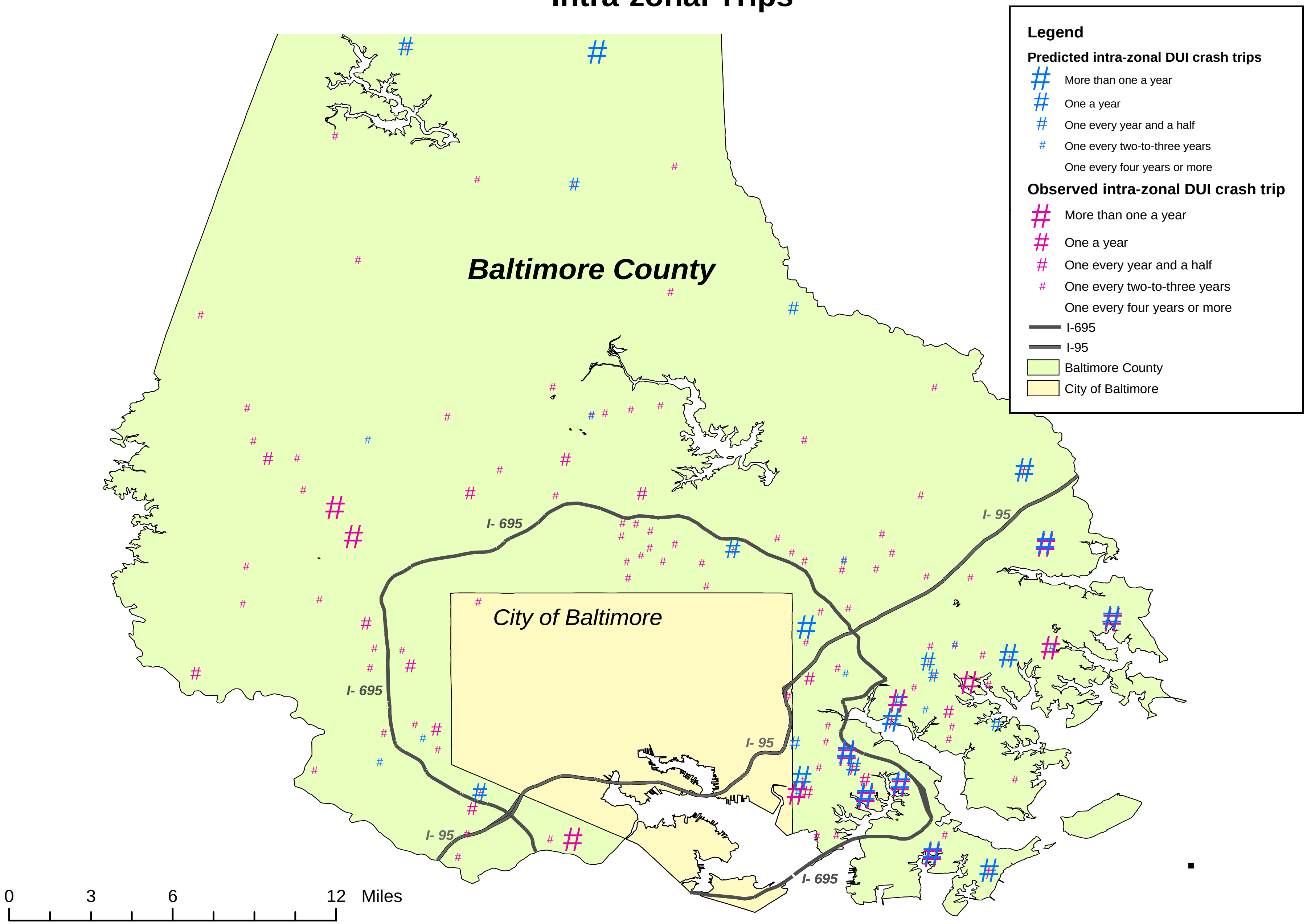
Predicted Annual DUI Crash Trips: 1999-2001
Inter- and Intra-zonal Trips



Comparing Observed and Predicted Annual DUI Crash Trips: 1999-2001
Inter-zonal Trips



Comparing Observed and Predicted Annual DUI Crash Trips: 1999-2001
Intra-zonal Trips



Possible DUI Policy Interventions

1. Enforcement – DUI citations

2. Target high risk persons/communities

3. Improve roadways where DUI crashes occur

The Effect of Citations

Testing Whether Citations Decrease Crashes

1999 Citations Predicting 2000 Crashes Holding 1999 Crashes Constant

```
Model result:
 Data file:                          DUI destinations.dbf
 Type of model:                      Destination
 DepVar:                             Number of Crashes in Zone: 2000
 N:                                    325
 Df:                                   322
 Type of regression model:           Poisson with over-dispersion correction
 Log Likelihood:                     -450.205872
 Likelihood ratio(LR):               90.312258
 P-value of LR:                      0.0001
 AIC:                                906.411743
 SC:                                 917.763219
 Dispersion multiplier:              1.000000
 R-square:                           0.174636
 Deviance r-square:                  0.834679
```

Predictor	DF	Coefficient	Stand Error	Pseudo-Tolerance	z-value	p-value
CONSTANT	1	-0.323192	0.086959	.	-3.716604	0.001
CRASHES 1999	1	0.130236	0.048271	0.733123	2.698049	0.010
CITATIONS 1999	1	0.040954	0.007370	0.733123	5.557177	0.001

Conclusion

DUI citations parallel the number of DUI crashes but don’t necessarily reduce them

Modeling the Effects of Targeting High Risk Individuals and DUI Crash Hot Spots

Concentration of DUI Offenders and Crash Locations

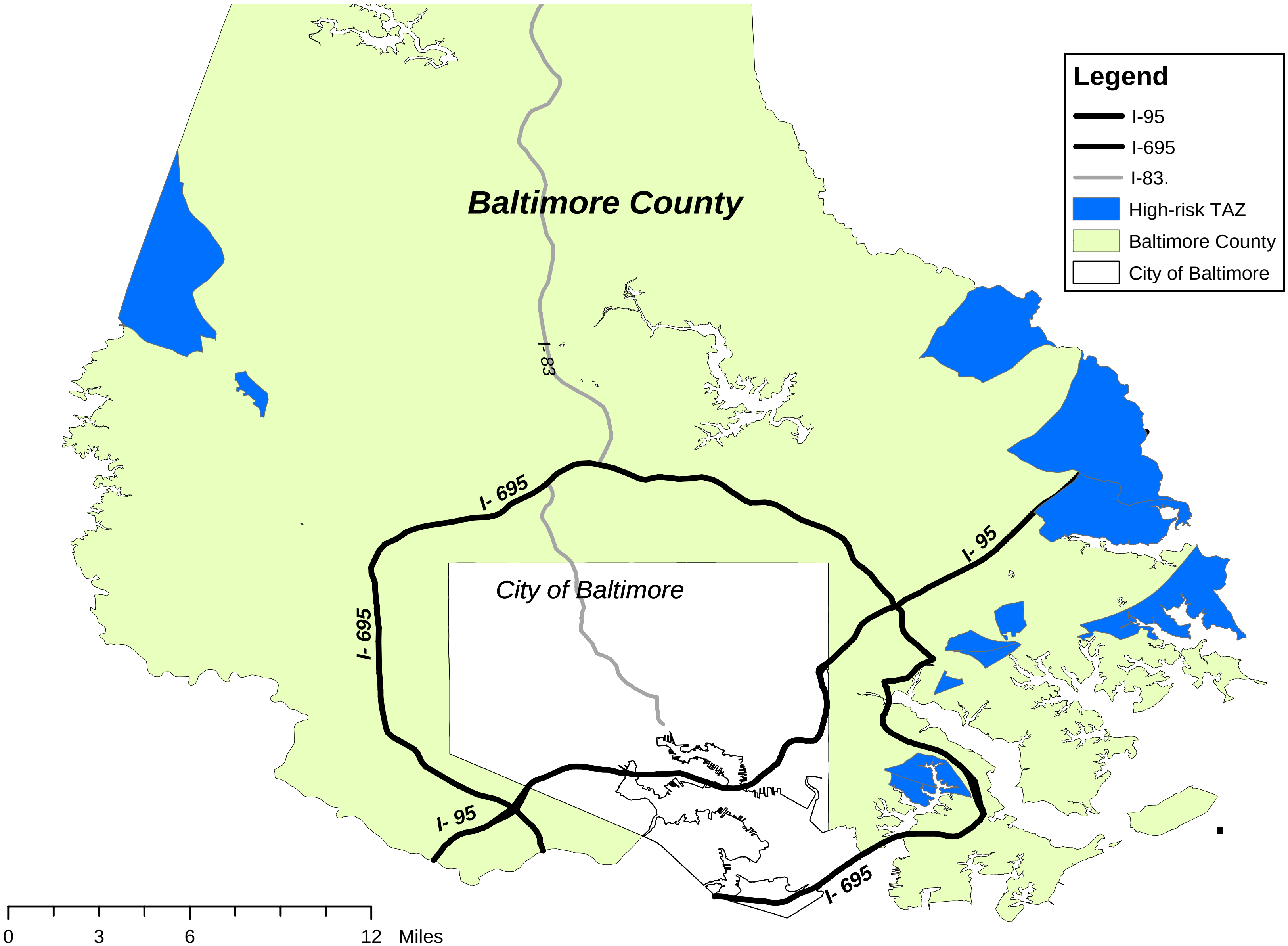
In 14 TAZ’s, 17% of DUI crash offenders live

In 11 TAZ’s, 13% of DUI crashes occur

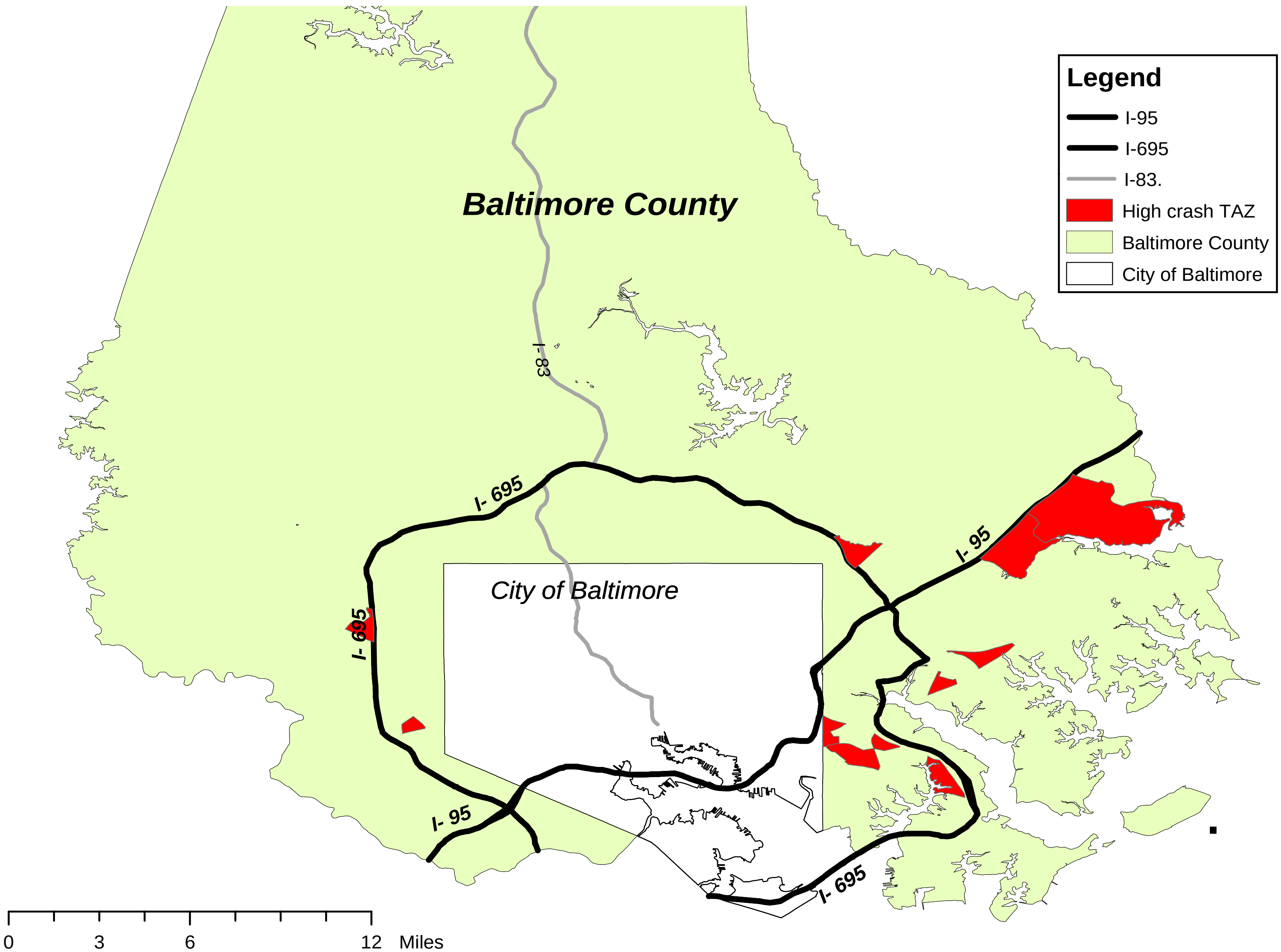
In model, identify these zones by dummy variable in origin and destination models

Model is realistic in that it addresses limited number of zones

DUI Crashes by Traffic Analysis Zones: 1999-2001
Location of Residences of DUI Drivers Involved in Crashes



DUI Crashes by Traffic Analysis Zones: 1999-2001
Location of Zones with 3 or More DUI Crashes



High-risk Communities & DUI Crash Origins

Model result:

Data file:	DUI origins.dbf
Type of model:	Origin
DepVar:	ANNUAL DUI CRASH OFFENDERS BY RESIDENCE
N:	534
Df:	526
Type of regression model:	Poisson with over-dispersion correction
Log Likelihood:	-421.359275
Likelihood ratio(LR):	169.678382
P-value of LR:	0.0001
AIC:	858.718550
SC:	892.961717
Dispersion multiplier:	1.000000
R-square:	0.474413
Deviance r-square:	0.634365

Predictor	DF	Coefficient	Stand Error	Pseudo-Tolerance	z-value	p-value
CONSTANT	1	-1.981253	0.080168	.	-24.713775	0.001
POPULATION	1	0.000120	0.000019	0.851100	6.406305	0.001
PCT WHITE	1	0.010071	0.001172	0.760218	8.593969	0.001
RURAL	1	0.137985	0.079311	0.545502	1.739785	0.100
# LIQUOR STORES	1	0.275987	0.047292	0.810189	5.835781	0.001
# BARS	1	0.107575	0.025758	0.785636	4.176404	0.001
AREA OF ZONE	1	0.007554	0.010195	0.554289	0.740924	n.s.
HIGH RISK TAZ	1	1.170567	0.075987	0.865070	15.404872	0.001

DUI Crash Hot Spots

Model result:

Data file:	DUI destinations.dbf
Type of model:	Destination
DepVar:	ANNUAL DUI CRASHES BY LOCATION
N:	325
Df:	319
Type of regression model:	Poisson with over-dispersion correction
Log Likelihood:	-328.554293
Likelihood ratio(LR):	77.904598
P-value of LR:	0.0001
AIC:	669.108586
SC:	691.811537
Dispersion multiplier:	1.000000
R-square:	0.398128
Deviance r-square:	0.688339

Predictor	DF	Coefficient	Stand Error	Pseudo-Tolerance	z-value	p-value
CONSTANT	1	-0.486630	0.072104	.	-6.748954	0.001
PCT OF ZONE IN RESIDENTIAL BLDG OF 10 OR MORE	1	0.004036	0.001415	0.949798	2.851752	0.010
# LIQUOR STORES	1	0.213080	0.046316	0.894716	4.600551	0.001
# BARS	1	0.116316	0.027832	0.764332	4.179273	0.001
AREA OF ZONE	1	-0.015696	0.013496	0.946482	-1.163054	n.s.
HOT SPOT	1	0.973275	0.149227	0.834926	6.522092	0.001

Targeting High Risk Zones/Individuals

"Intervention" involves two steps applied to 14 high risk zones:

1. **'Don't Drive while Drinking' campaign**
2. **Conduct interviews with drivers convicted of DUI driving**

To simulate, reduced DUI crashes in these zones by 20%

Result is that the total number of DUI crashes were reduced by 3.5%

Fixing Roadway Hot Spots

Process involves five steps applied to 11 hot spot zones:

1. **Identify hazardous location**
2. **Document crash pattern**
3. **Propose mitigation measures**
4. **Analyze benefit-cost of each measure**
5. **Implement measures with best benefit-cost ratio**

Targeting High Crash DUI Hot Spots

To simulate, reduced DUI crashes in 11 hot spot zones by 20%

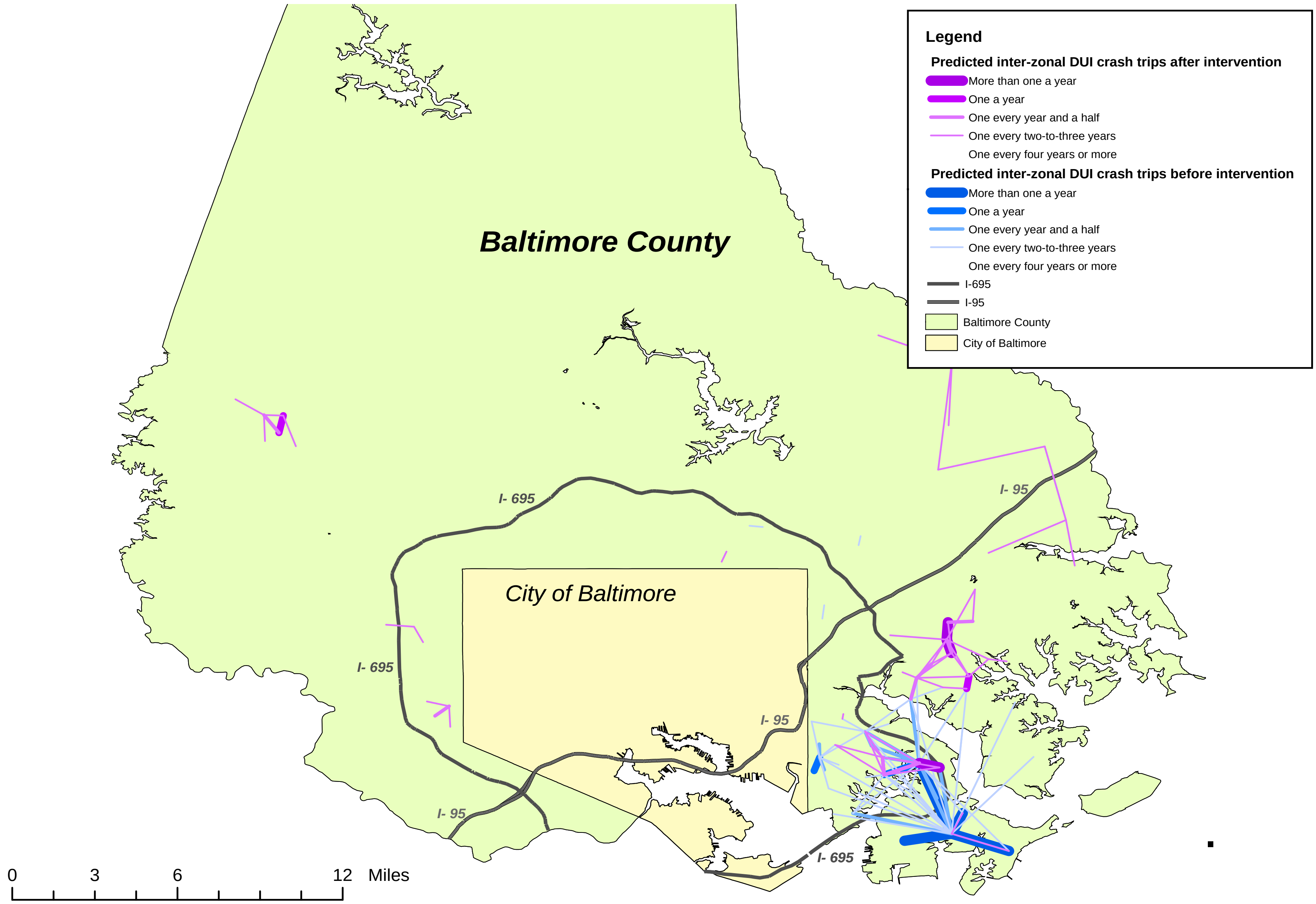
When combined with reduction in offenders, total number of DUI crashes decreased by 6%

Comparing DUI Crash Trips Before & After Intervention

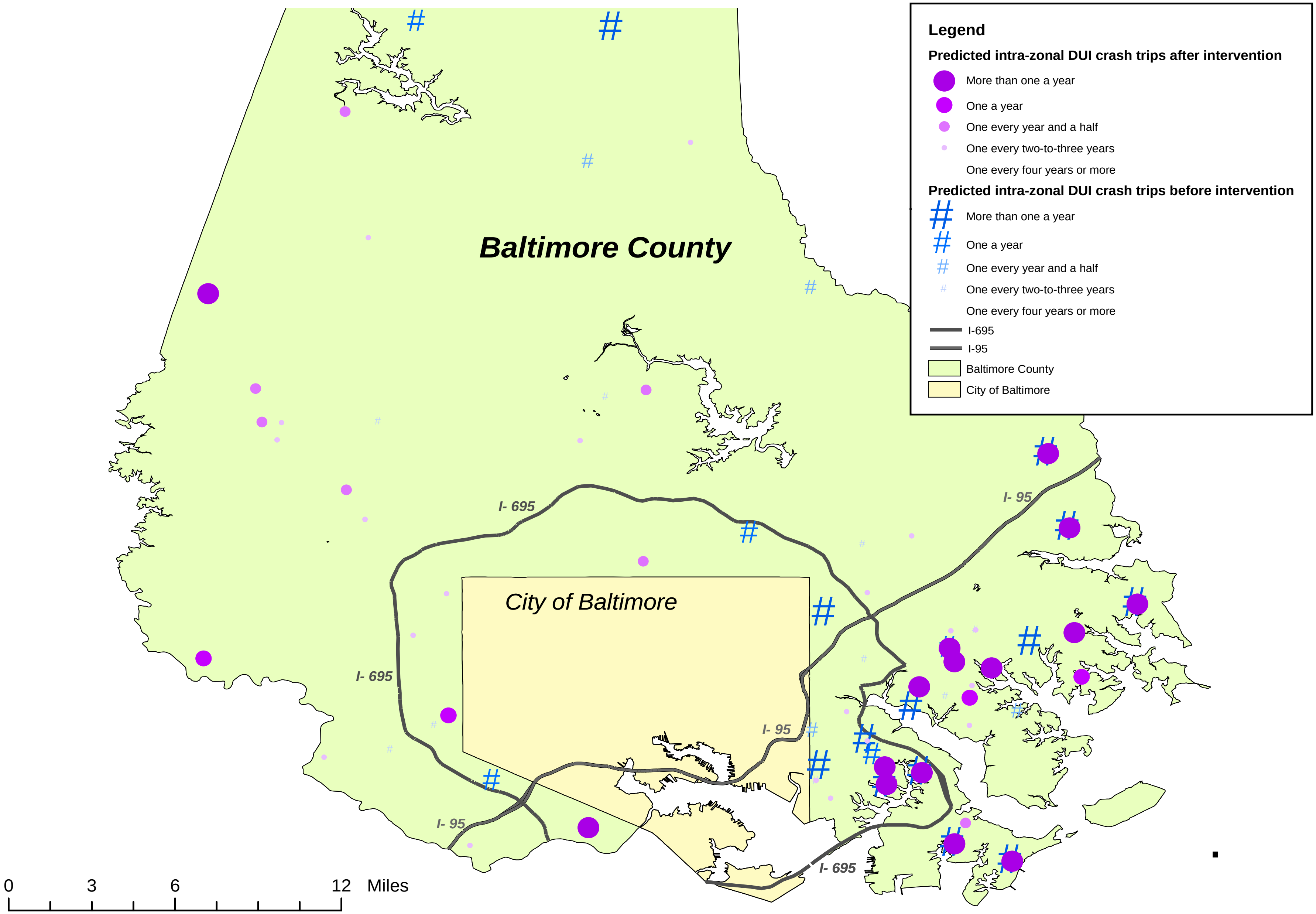
Annual Number of Expected Crashes

	Inter-zonal	Intra-zonal
Without intervention	236	53
With intervention	219	52
Expected Change	-17	-1

Predicted DUI Crash Trips Before & After Policy Intervention
Inter-zonal Trips



Predicted DUI Crash Trips Before & After Policy Intervention
Intra-zonal Trips



Two Conceptual Points

Travel Demand Modeling is
A Framework Rather than Specific Model

There are different:

- **Approaches**
- **Data sources**
- **Sequencing**
- **Modeling tools**

***CrimeStat* implemention is only one approach**

There is a Difference Between A Model and an Empirical Description

Model	**Empirical Description**
Few variables	Many variables
Simplified relationships	Complex relationships
Analogy	Literal
Ability to manipulate variables: *Prediction* *Scenarios ("What if?")* *Test theory, policy*	Limited manipulation of variables: *Description* *Theory-building* *Deployment*

Every model, no matter how detailed or how well conceived, designed, and implemented, is a vastly simplified representation of the world, with all of the intricacies we experience on a day-to-day basis

Alan Greenspan, August 2005

Uses of
Crime Travel Demand Modeling

Uses for Police

- **Developing policing strategies**
  (e.g., shifting patrol deployment)
- **Forecasting**
  (e.g., predicting crime levels five year later)
- **Modeling interventions**
  (e.g., add drug treatment center)
- **Anticipating changes**
  (e.g., building new shopping center)

Research Uses

- Organizes crime travel information systematically
- More realistic offender travel theory
- Dynamic analysis of crime travel patterns
- Comparisons between types of crime
- Comparisons between different cities

Future Directions in Offender Travel Modeling

- **Improved trip distribution tools**
- **Build utility functions for crime offenders**
- **Calibrate for many metropolitan areas for comparison**
- **Activity-based, disaggregate models (requires individual level data)**
- **Micro-simulation of individual characteristics**

But only after the aggregate model is developed and calibrated

More information at:

http://www.icpsr.umich.edu/crimestat