# Theory and Data for Land Use Forecasting

Webinar 2 of an 8-part TMIP Webinar series on land use forecasting methods.

Paul Waddell, 2011

## Land Use Forecasting Webinar Series

- The Evolving State of the Practice
- Land Use Theory and Data
- Sketch Planning and Visioning (e.g. I-PLACE3S)
- Spatial Input-Output Frameworks (e.g. Pecas)
- Dynamic Microsimulation (e.g. UrbanSim)
- Modeling Real Estate Demand
- Modeling Real Estate Supply
- Scenario Planning and Visualization

2

# Webinar 2: Theory and Data for Land Use Forecasting

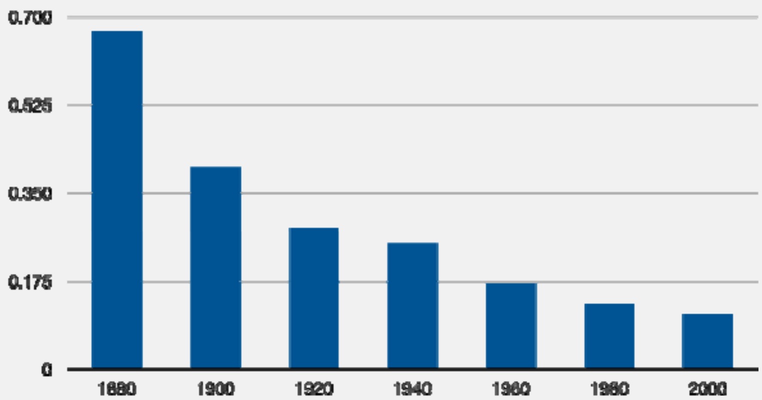
1. How transportation influences land use
   - The role of accessibility
   - Measuring accessibility
2. Understanding real estate markets
   - Demand factors
   - Supply factors
   - Demand - supply interactions
3. Data for Land Use Forecasting
   - Data Requirements
   - Data Problems

# Changes in Transportation Technology

| Technology | Introduced | Travel Time per Mile (minutes) |
|---|---|---|
| Walking | ? | 20 |
| Horse-Drawn Streetcar | 1835 | 12 |
| Cable Car | 1875 | 7.5 |
| Electric Streetcar | 1890 | 6 |
| Rail Rapid Transit | 1910 | 4 |
| Motor Bus | 1915 | 3 |
| Automobile | 1920 | 2 |

Source: Pickrell

## Changes in Urban Population Density: Chicago




## How Transportation Affects Land Use

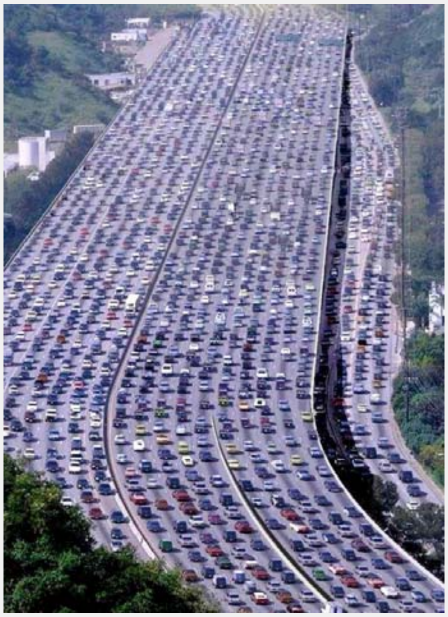
- Accessibility has long been known to influence the location of firms and households activities
  - Firms reduce transport costs to labor, suppliers and consumers
  - Households trade off transportation costs to work and shop against housing costs and other amenities
- Accessibility strongly influences land prices and density: bid rent theory
- Auto use supports low-density, dispersed urban form, which reinforces auto use
- Transit ridership increases with high density, mixed use development, nodal urban form, and high walk access – and supports Transit Oriented Development
- Walk access to shopping increases pedestrian activity, which increases pedestrian oriented retail

## How Transportation Affects Land Use

- Transport effects moderated by durable real estate stock
  - Older cities have dense core built in response to slow travel
  - Newer cities built mostly after highways lack this dense core
- The current debates:
  - What will be the effect on land use of increasing telecommunications as a ‘mode of transportation’?
  - With so much existing transportation capacity, does a marginal addition significantly affect land use?
  - Is transportation becoming less of a factor in location choices?
    - Multiple worker households
    - High costs of housing relocation
    - Growing importance of non-work travel
    - Growing importance of location amenities
- How much do neo-traditional neighborhood design, transit and non-motorized access affect location, trip rates, distances and modes?

## Mobility vs. Accessibility

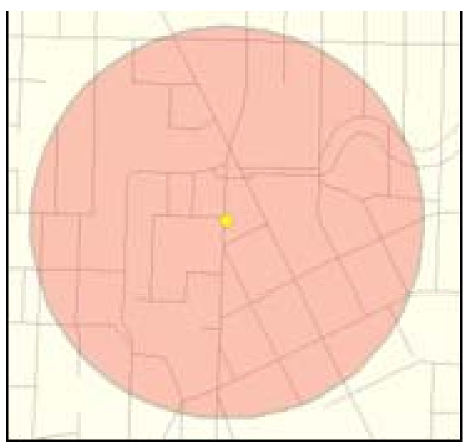
- Traditionally, transportation planning focused on mobility
  - How congested are roads, how fast is traffic moving on them?
- More recently, increasing attention has focused on the concept of accessibility
  - How easy is it to access the activities you need or want to engage in?
- These concepts focus planning efforts very differently
  - Accessibility adds proximity, not just speed of travel. Slow modes like walking can be accessible if destinations are at walking scale.



Mobility vs. Accessibility

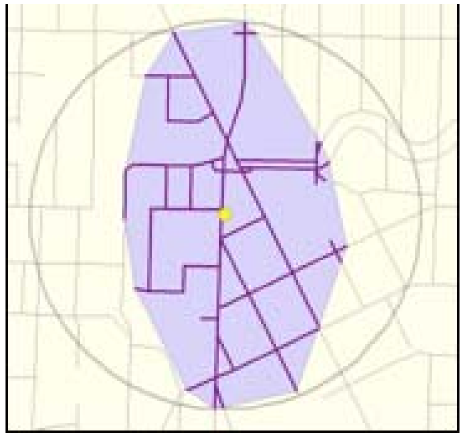
## Measuring Accessibility: Euclidian Distance

- Advantages:
  - Easy to compute
  - Easy to Interpret
- Disadvantages
  - No network information
  - No congestion effects



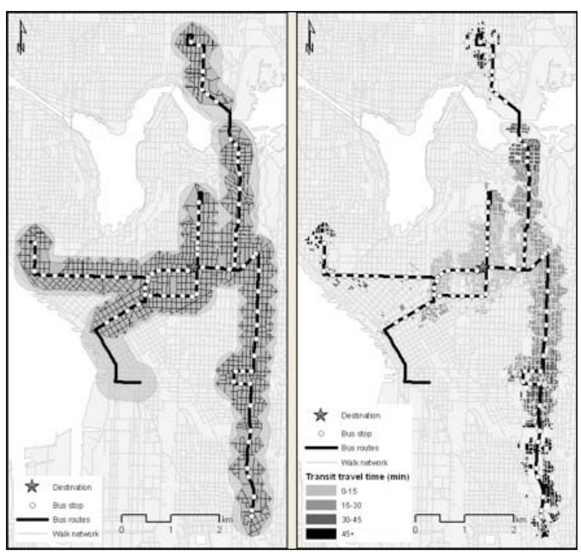
## Measuring Accessibility: Network Distance

- Advantages
  - Easy to interpret
  - Relatively easy to compute
- Disadvantages
  - Usually excludes local streets
  - Excludes congestion
  - Not multi-modal

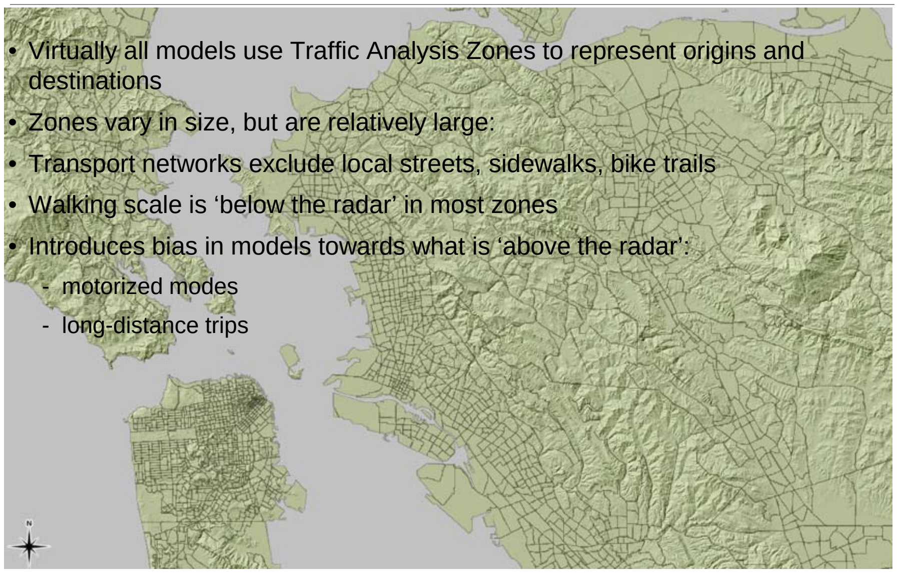


## Measuring Accessibility: Parcel + Network

- a) network distance from transit stops
- b) parcels classified by total travel time by transit to destinations (e.g. grocery)
- Brian Lee, Access to Destinations



## Accessibility in Transportation and Land Use Models



- Virtually all models use Traffic Analysis Zones to represent origins and destinations
- Zones vary in size, but are relatively large:
- Transport networks exclude local streets, sidewalks, bike trails
- Walking scale is ‘below the radar’ in most zones
- Introduces bias in models towards what is ‘above the radar’:
  - motorized modes
  - long-distance trips

## Travel Model ‘Accessibility’ Outputs

- Standard practice (4-step) travel models, and even more advanced activity-based models, predict for each zone-to-zone O-D pair, by time period, mode and purpose:
  - Travel time
  - Travel cost (tolls) or Generalized cost (time + the time-equivalent of tolls)
  - Trips
  - Most compute a ‘composite utility’ or logsum from the mode-choice model

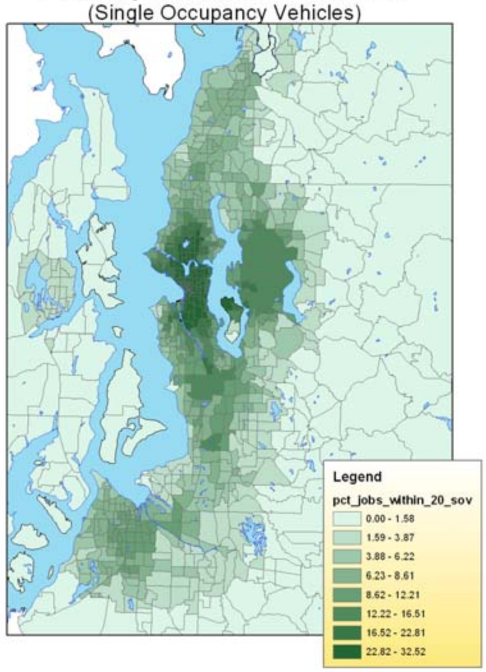
## Alternative Measures of Accessibility

- Access to activity within time threshold
  - How many jobs can you reach within 30 minutes by car, during the AM peak period?

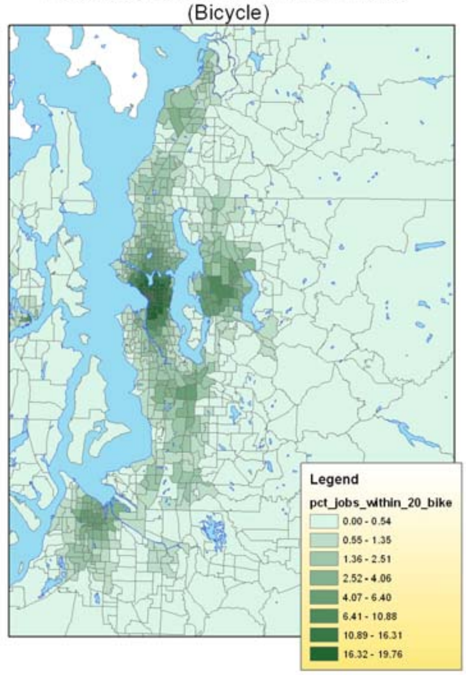
$$Access_i = \sum_j Jobs_j \forall Time_{ij} < X$$

- Advantages
  - Easy to interpret
  - Straightforward to compute
- Disadvantages
  - Arbitrariness of threshold
  - Non-time factors ignored
  - Can it represent influence of multiple modes?

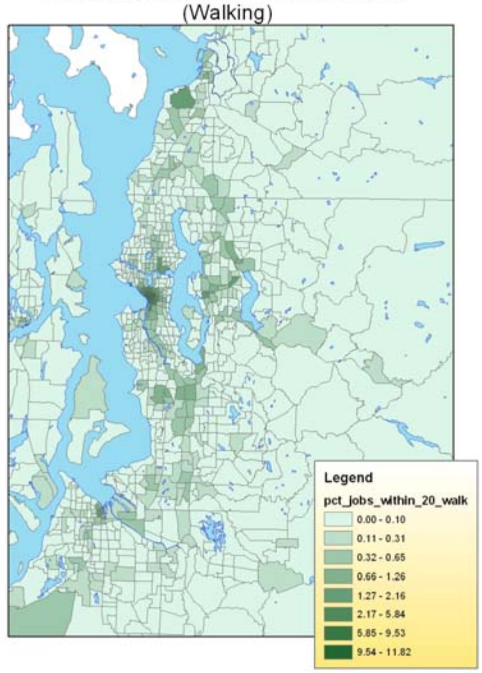
Percentage of Jobs Within 20 Minutes
(Single Occupancy Vehicles)



Percentage of Jobs Within 20 Minutes
(Bicycle)




Percentage of Jobs Within 20 Minutes
(Walking)



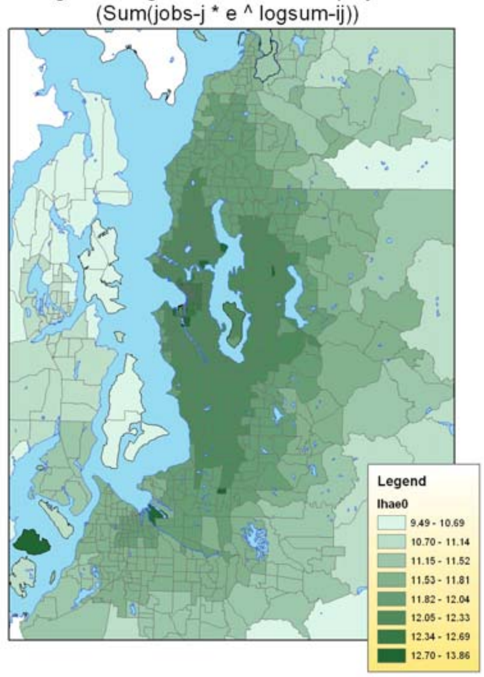
## Alternative Measures of Accessibility

- Logsum-based Measures of Accessibility
  - Accessibility to employment

$$Access_i = \sum_j Jobs_j e^{Logsum_{ij}}$$

- Advantages
  - Non-time factors incorporated
  - Can represent influence of multiple modes
  - Reflects spatial distribution of activities
  - Consistency with mode-choice model
  - Relatively straightforward to compute
- Disadvantages
  - Difficult to interpret
  - Does not reflect information from Trip Distribution

Logsum-weighted Access to Employment
(Sum(jobs-j * e ^ logsum-ij))



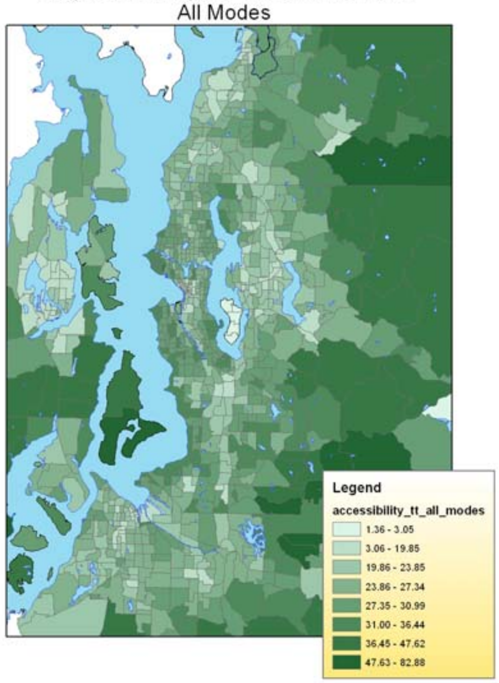
## Alternative Measures of Accessibility

- Trip-Weighted Measures of Accessibility
  - Accessibility for Home-based Work Purpose

$$Access_i = \frac{\sum_j (Trips_{ij} Logsum_{ij})}{\sum_j Trips_{ij}}$$

- Advantages
  - Non-time factors incorporated
  - Can represent influence of multiple modes
  - Reflects spatial distribution of activities
  - Consistency with mode-choice and trip distribution models
  - Relatively straightforward to interpret
- Disadvantages
  - Shows less spatial variation

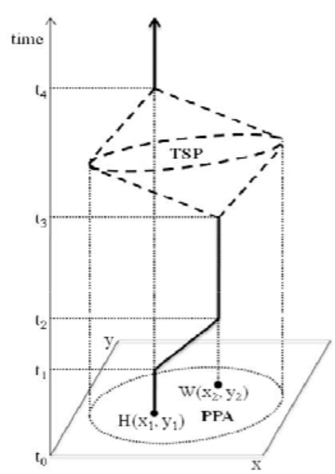
Weighted Average of Travel Time, HBW
All Modes



## Alternative Measures of Accessibility

- Measures of transit accessibility
  - Combine into an overall measure or keep separate?
  - Importance to different consumer types
- Measures of non-motorized accessibility
  - Travel model networks too coarse
  - Zones too large
  - Pedestrian-friendliness measures
  - Data?

## Time-Space Prism (TSP) Accessibility Approach (Lee et al, 2009)




At individual worker level

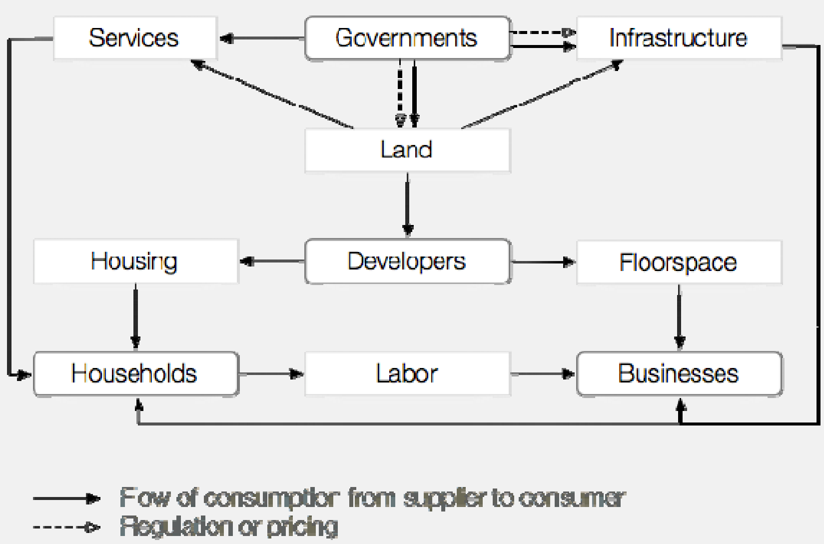
Constraints:

- Spatial: home & workplace
- Temporal: home & work schedules
- Travel: transportation network & travel speeds

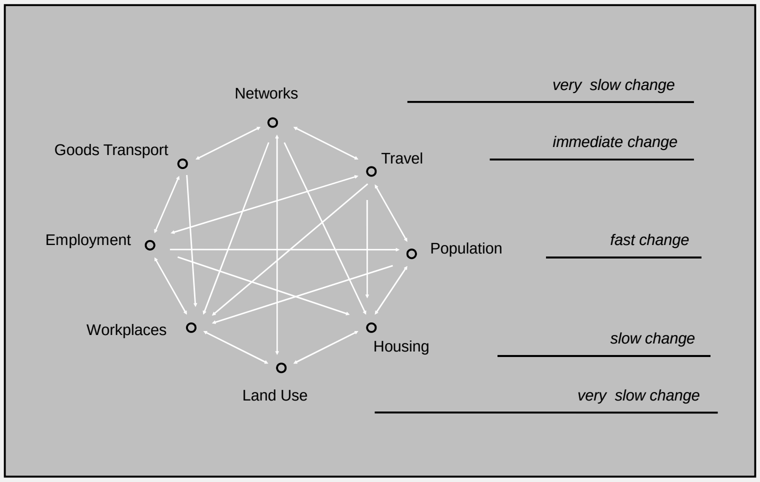
## Webinar 2: Theory and Data for Land Use Forecasting

1. How transportation influences land use
   - The role of accessibility
   - Measuring accessibility
2. Understanding real estate markets
   - Demand factors
   - Supply factors
   - Demand - supply interactions
3. Data for Land Use Forecasting
   - Data Requirements
   - Data Problems

## Role of Urban Real Estate Markets



## Interacting Drivers at Different Scales




Source: Wegener, 1995

## Real Estate Demand: Theoretical Framework

- Macroeconomic Growth
- Product Differentiation
- Market Segmentation
- Market Conditions
- Location Factors
- Disaggregate Demand
- Submarkets
- Demand and Supply Interaction
- Policy Influence

## Demand: Macroeconomic Growth

- Primary driver of demand for urban land and real estate
- Starting point for analysis of demand:

$$D = AR$$

Where:

*D* is the demand for real estate

*A* is the aggregate number of households or employment

*R* is the space utilization ratio (units/household; square feet per employee)

## Demand: Product Differentiation

- Demand varies by type of real estate
- Introduces complexity of choice:

$$D_U = AP_U R_U$$

Where:
$D_U$ is the demand for real estate
*A* is the aggregate number of households or employment
$P_U$ is the probability that a household will choose to occupy housing type *U*
$R_U$ is the space utilization ratio for housing type *U*

## Sample Classification of Real Estate

| Residential | Commercial | Industrial |
|---|---|---|
| Single-family detached | Retail | Light manufacturing |
| Rural density | Strip center | Heavy manufacturing |
| Low-density | Neighborhood center | Industrial Incubator |
| Mid-density | Power center | Warehouse |
| High-density | Community center | High-tech/flex space |
| Duplex/Townhouse | Regional mall | Research and development |
| Condominium | Office | |
| Low-density | Low-rise | |
| High-density | Mid-rise | |
| Apartments | High-rise | |
| Low-density | Campus | |
| Mid-density | | |
| High-density | | |

## Demand: Market Segmentation

- Demand also varies by consumer characteristics such as:
  - Income, Age, Household Size, Life Cycle

$$D_U^C = \sum_C \left( A^C P_U^C R_U \right)$$

Where:
$D^C_U$ is the demand for real estate of type U by consumer type C
$A^C$ is the aggregate number of households or employment of type C
$P^C_U$ is the probability that a household will choose to occupy housing type U
$R_U$ is the space utilization ratio for housing type U

## Sample Household Classification

| Income | Age of head | Persons | Workers | Children |
|---|---|---|---|---|
| Less than $5 | 15 to 24 | 1 | 0 | 0 |
| $5,000 to $9,999 | 25 to 34 | 2 | 1 | 1 or more |
| $10,000 to $14,999 | 35 to 44 | 3 | 2 or more | |
| $15,000 to $24,999 | 45 to 54 | 4 | | |
| $25,000 to $34,999 | 55 to 64 | 5 or more | | |
| $35,000 to $49,999 | 65 to 74 | | | |
| $50,000 to $74,999 | 75 or over | | | |
| $75,000 to $99,999 | | | | |
| $100,000 or more | | | | |

## Sample Employment Sector Classification

| Standard industrial classification codes | Sector description |
|---|---|
| 99 - 999 | Agriculture |
| 2400 - 2499 | Lumber and wood |
| 2500 - 2599, 3200 - 3999 | Other durable |
| 2000 - 2099 | Food products |
| 2100 - 2399, 2600 - 3199 | Other nondurable |
| 1500 - 1799 | Construction |
| 1000 - 1499 | Mining |
| 4000 - 4999 | Transportation |
| 5000 - 5199 | Wholesale trade |
| 5200 - 5999 | Retail trade |
| 6000 - 6999 | Services |
| 7000 - 7999 | Education |
| 8200 - 8299 | Government |
| 9000 - 9999 | |

## Demand: Market Conditions

- Demand is sensitive to market conditions: price elasticity

$$D_U^C = \sum_C \left( A^C P_U^C \left( p \right) R_U \left( p \right) \right)$$

Where:
$D^C_U$ is the demand for real estate of type U by consumer type C
$A^C$ is the aggregate number of households or employment of type C
$P^C_U(p)$ is the probability that a household will choose to occupy housing type U
$R_U(p)$ is the space utilization ratio for housing type U, as a function of prices

## Demand: Location Factors

- Demand for real estate is also demand for location
- Ignoring locational factors will seriously bias demand and supply estimates

$$D_{UL}^C = \sum_C \left( A^C P_{UL}^C \left( p_L \right) R_U \left( p_L \right) \right)$$

Where:
$D^C_{UL}$ is the demand for real estate of type U at location L by consumer type C
$A^C$ is the aggregate number households or employment of type C
$P^C_{UL}(p_L)$ is the probability that a consumer will choose type U at location L
$R_{UL}(p_L)$ is the utilization ratio for type U at location L, as a function of prices at L

## Demand: Disaggregate Choice

- Examine demand at a disaggregate level of the individual consumers
- The problem becomes one of predicting choice from available real estate
- Preferences will depend on:
  - Household income or firm profit
  - Other household or firm characteristics
  - Price of real estate
  - Structural characteristics
  - Location characteristics

## Demand: Disaggregate Choice

- Most common technique for analyzing disaggregate consumer choice is the multinomial logit model:

$$P_{UL}^{C} = \frac{e^{\alpha(Y-P_U)+\beta S_U+\delta L_U}}{\sum_U e^{\alpha(Y-P_V)+\beta S_V+\delta L_V}}$$

Where
$P^{C}_{UL}$ is the probability of a consumer of type C choosing site u from the set U
$Y$ is household income or firm profit
$P$ is he price of the building
$S$ is a set of structural characteristics of the building
$L$ is a set of location characteristics at the site are sets of estimated demand parameters

## Real Estate Supply: Inelastic in Short Run

- Market signals (prices, vacancy rates)
- Identification of site; option to buy
- Permitting and review
- Financing
- Land acquisition
- Subdivision
- Site preparation
- Construction
- Sale or lease

## Supply: Constraints

- Site Availability
- Physical Constraints
  - Topography
  - Water
  - Slope
  - Access/egress
- Zoning
- Infrastructure Availability
- Environmental Regulations
- Urban Growth Boundary

## Supply of Urban Development

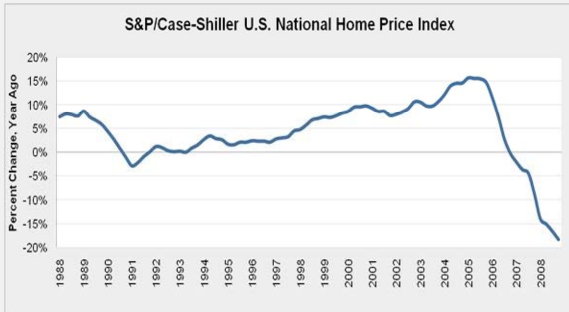
- Development Decisions Based on Expected Rate of Return on Investment:
  - Expected sale price; net present value of rents
  - Cost of capital: Interest rate
  - Opportunity cost of next best investment alternative
  - Tax treatment
  - Costs of development: land, labor, material
  - Infrastructure extension; concurrency
  - Impact fees or other development fees
  - Holding costs due to construction delay

## Supply of Urban Development

- Developer Industry
  - Economies of scale and ease of financing favor large-scale development; standard products
  - Specialization in specific real estate products
  - Policy interest in stimulating infill and redevelopment; smaller lot residential development
  - Policy interest in development ‘paying its way’
  - Effects of policies on development not well understood

## Interaction of Demand and Supply

- Demand is elastic in the short run
- Supply is inelastic in the short run
- Change in demand signals supply response, but time lags and constraints may create sustained disequilibrium
- For those that like the concept of equilibrium, can you find it in this graph?




## Policy Influence on Urban Development

- Local government policies may affect:
  - Supply of real estate
    - UGBs
    - Land use plans, zoning
    - Permit limitations
    - Impact fees
    - Concurrency requirements
  - Demand for real estate
    - Infrastructure and public services
    - Taxes and fees

## Webinar 2: Theory and Data for Land Use Forecasting

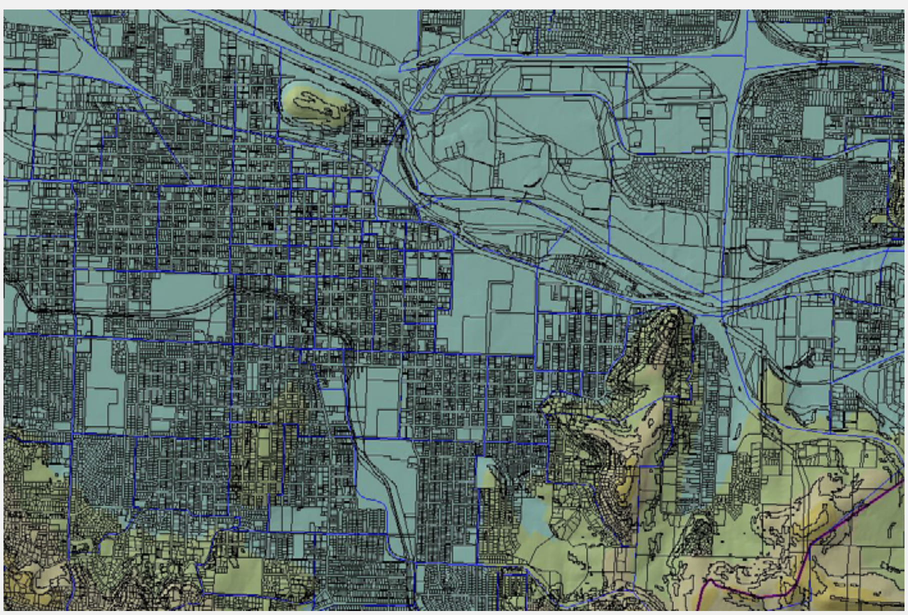
1. How transportation influences land use
   - The role of accessibility
   - Measuring accessibility
2. Understanding real estate markets
   - Demand factors
   - Supply factors
   - Demand - supply interactions
3. Data for Land Use Forecasting
   - Data Requirements
   - Data Problems

## Data Requirements: Macroeconomy

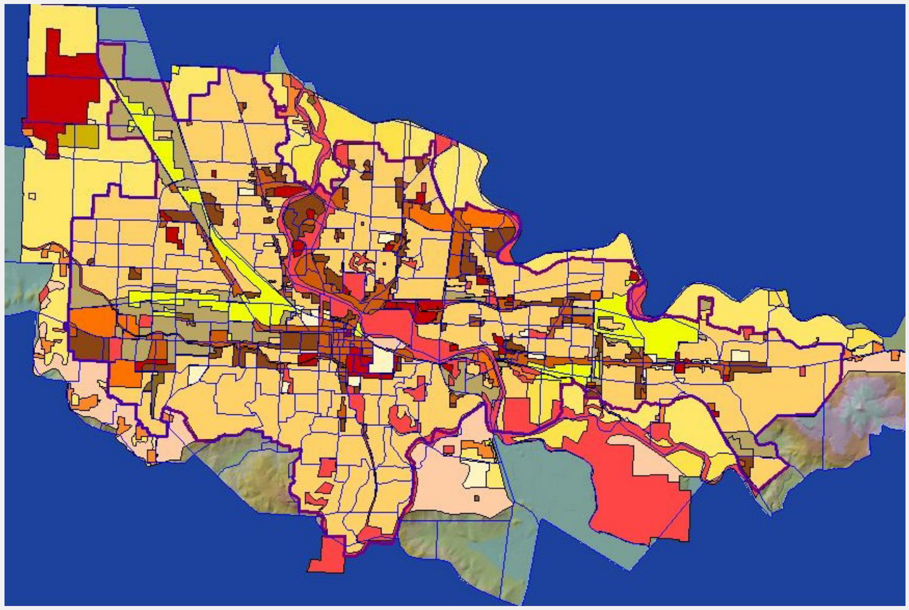
- Macroeconomic and demographic forecasts for a metropolitan area used to ‘drive’ land use forecasts
- Principal sources:
  - State economic forecasts
  - Regional planning agency forecasts
  - DRI or other commercial sources
- Methods
  - Generally use econometric or hybrid Input-Output/Econometric techniques
  - Link to national forecasts of output by sector
  - Derive population through migration and ‘cohort survival’
- Feedback of local policies on macro growth ignored

## Data Requirements: Land Use

- Existing land use
  - Aerial photo interpretation
  - Parcel databases (assessor records)
  - Building square footage, value, age
- Land use plans, zoning
- Development projects ‘in the pipeline’
- GIS is an efficient way to integrate, organize, review and manage these data
- Problems
  - Data are often messy and incomplete
  - Land use coding not standardized



## Input Data: Parcels



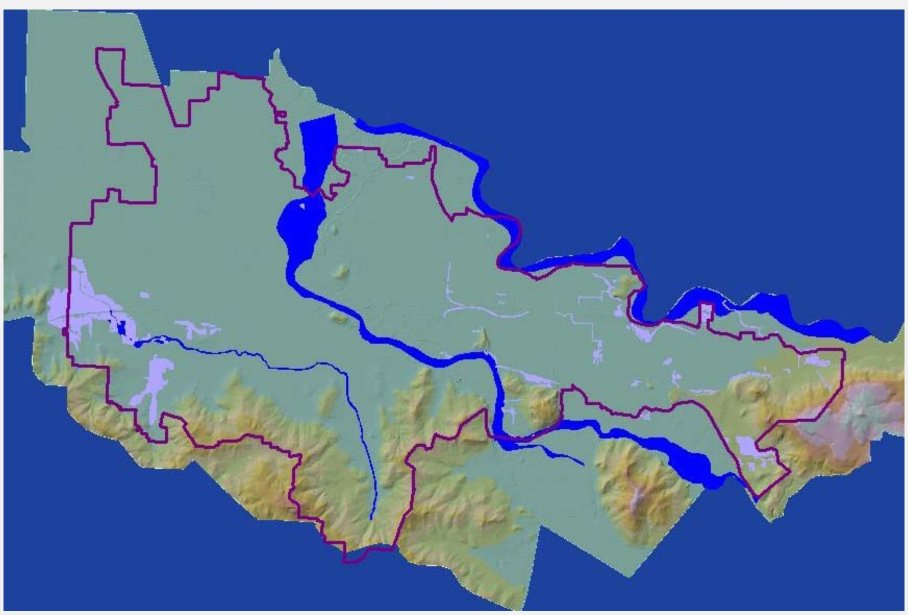
Input Data: Land Use Plans

## Data Requirements: Market

- Some methods use real estate market information
  - Housing units and commercial square footage
  - Land and improvement values
- Assessed values dubious indicators of real prices
  - Especially in California! (Prop 13)
  - Sales prices are better
  - Often available from boards of realtors, or from assessor offices
  - Sales still messy (not always arms-length transactions)
  - Sales of certain types of properties may be sparse

## Data Requirements: Environmental

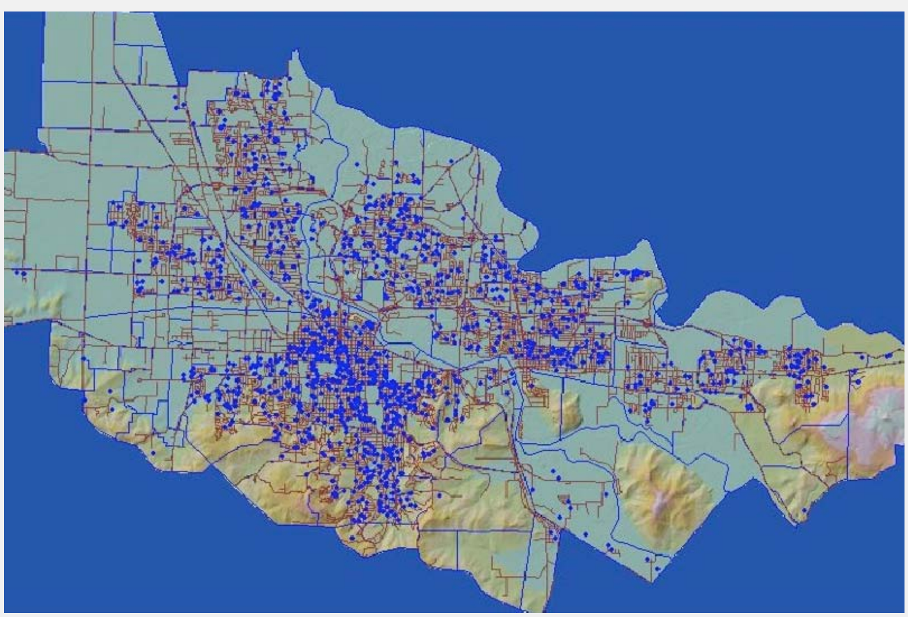
- Some methods use environmental data, usually as GIS layers
  - Wetlands
  - Floodplains
  - Steep slopes
  - Stream buffers
  - Special habitat areas
  - Agricultural, forest or other resource lands
  - Slide areas
  - Other environmentally sensitive lands



## Input Data: UGB and Environmental

## Data Requirements: Demographics

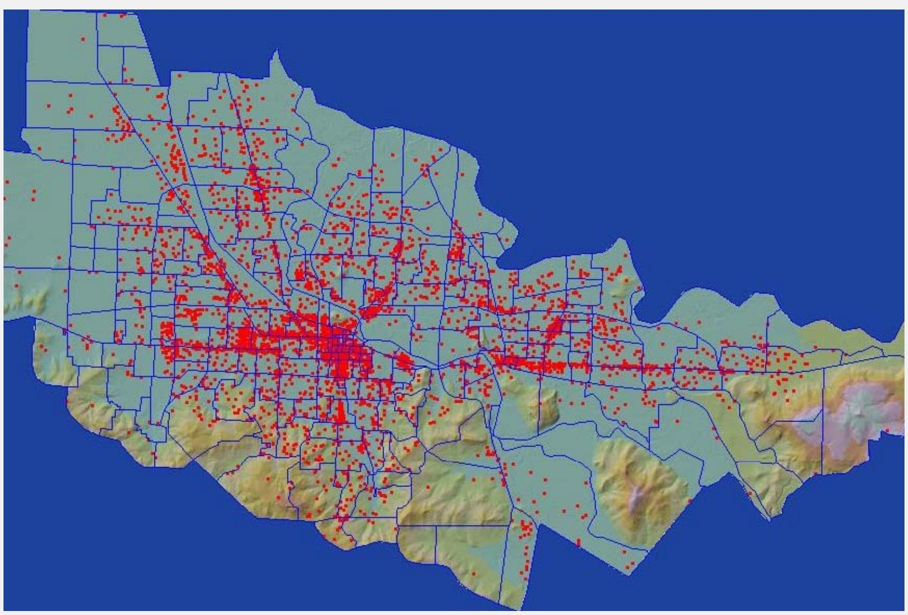
- Census
  - STF3 for socioeconomic data to the block group level
  - STF1 for limited data at the census block level
  - Public Use Microdata for metropolitan sample of individual households and persons
  - Household travel survey
  - Panel surveys ideal (but extremely rare)
  - Household synthesis methods combine census PUMS and STF3 to produce synthetic geocoded households for microsimulation approaches such as TRANSIMS, UrbanSim



## Input Data: Household Travel

## Data Requirements: Employment

- Most problematic of data inputs
- Generally insufficient investment in these data
- Sources:
  - State unemployment insurance records (ES202)
  - Dun & Bradstreet, InfoUSA, NETS, and other commercial sources
- Problems
  - Completeness (no self-employed, proprietors)
  - Headquarter reporting
  - Longitudinal matching
  - Classification confusion:
    - Industry
    - Land use
    - Occupation



## Input Data: Employment

## Problems of Data Integration

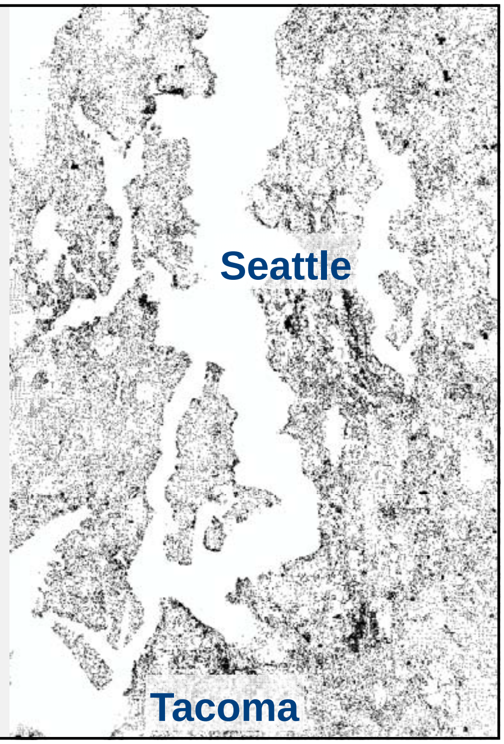
- Need consistent land use, employment and demographic data for modeling
- Numerous problems make this difficult:
  - Different sources, accuracies, time points, coding standards, levels of completeness and currency
  - Spatial mismatches due to geographies of data collection, geocoding procedures, cartographic inaccuracies
- The big question: is it easier to find and fix data inconsistencies in aggregate or disaggregate form?
- Tools improving for data cleaning, data integration

## Data Imputation using Data Mining

- Experimental software now in testing
- Uses Machine Learning and Data Mining algorithms to:
  - Detect errors in data
  - Impute missing values
- Works for both continuous and categorical variables

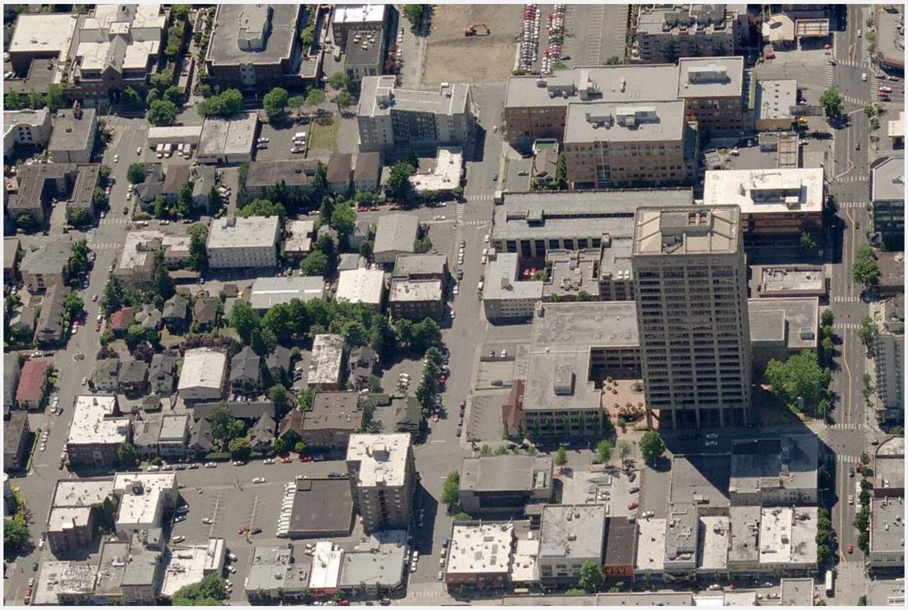
## The Magnitude of the Problem

- This map shows only buildings with missing values for “Building Type ID”, a description variable
- 195,501 out of ~1,200,000 parcels have Building Type ID = Null in King, Kitsap, Pierce & Snohomish counties





Data Imputation Tool



Data Imputation Tool

## Numeric Attributes: K-Nearest Neighbor

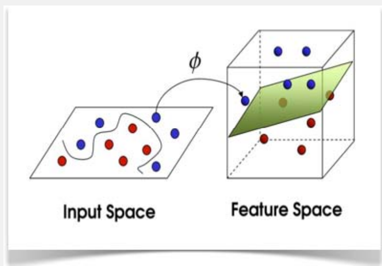
- Attributes:
  - Stories
  - Bldg SF
  - Improvement Value
- KNN Basics
  - Finds k closest neighbors in n dimensional space.
  - Uses k neighbors target values to make prediction.

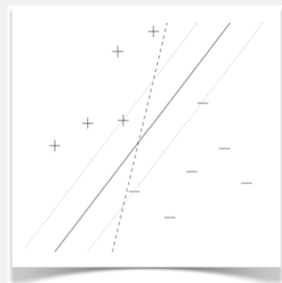
$$L^n(x_1, x_2) = \sqrt[n]{\sum_{i=1}^{\#\dim} |x_{1,i} - x_{2,i}|^n}$$

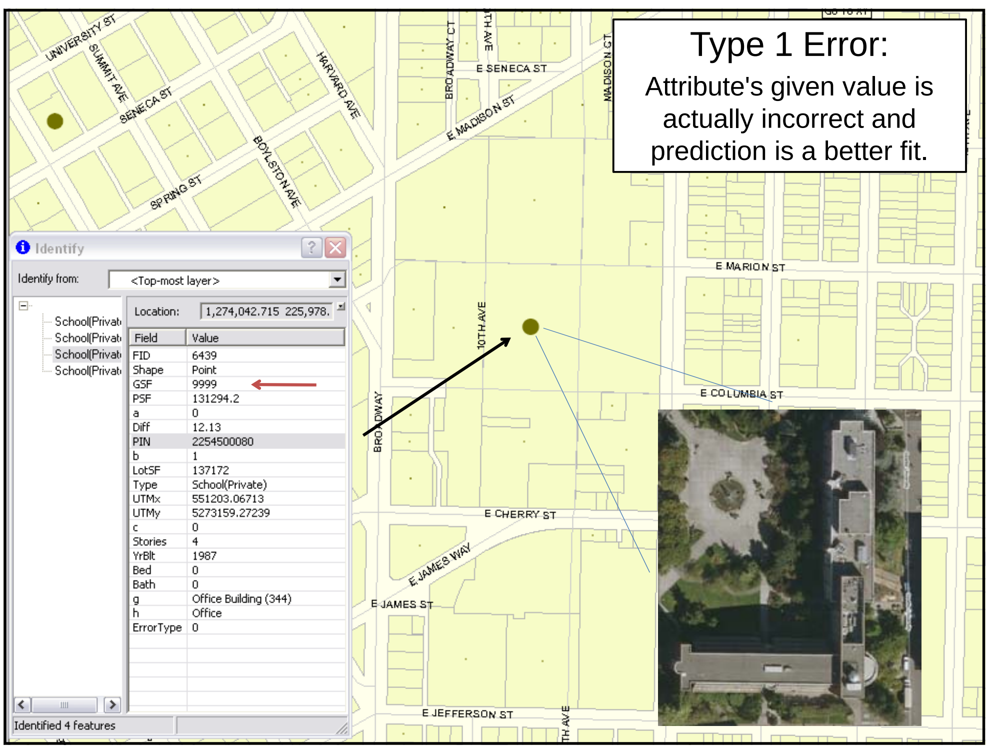
$$\hat{f}(x_q) \leftarrow \frac{1}{k}\sum_{i=1}^{k} f(x_i)$$

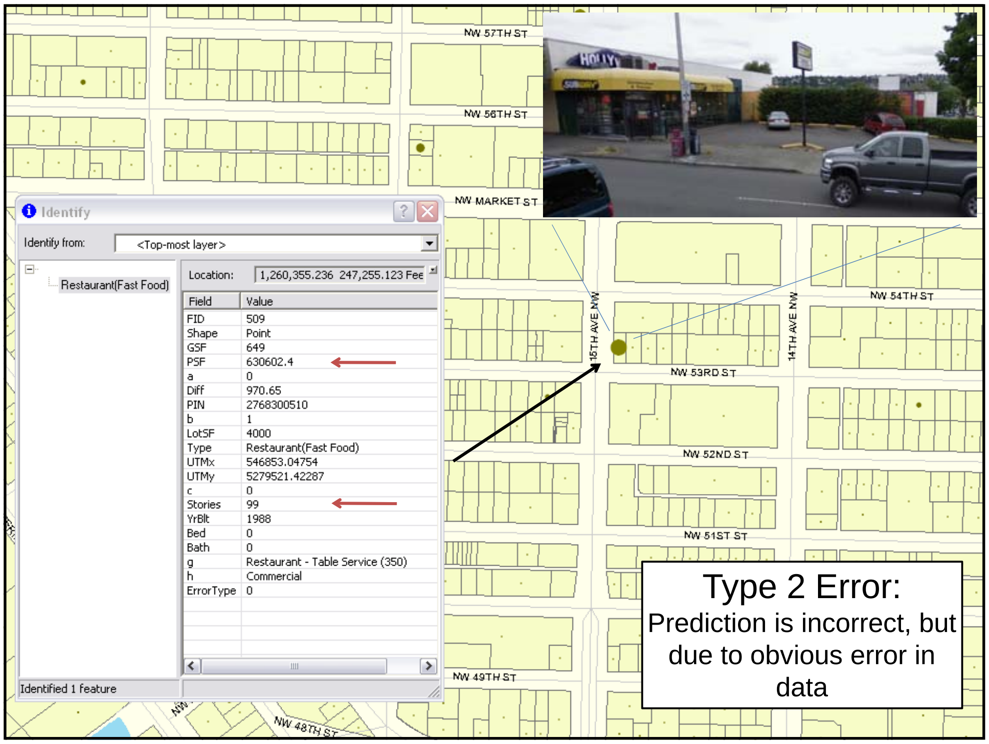
## Categorical Attributes: Support Vector Machines

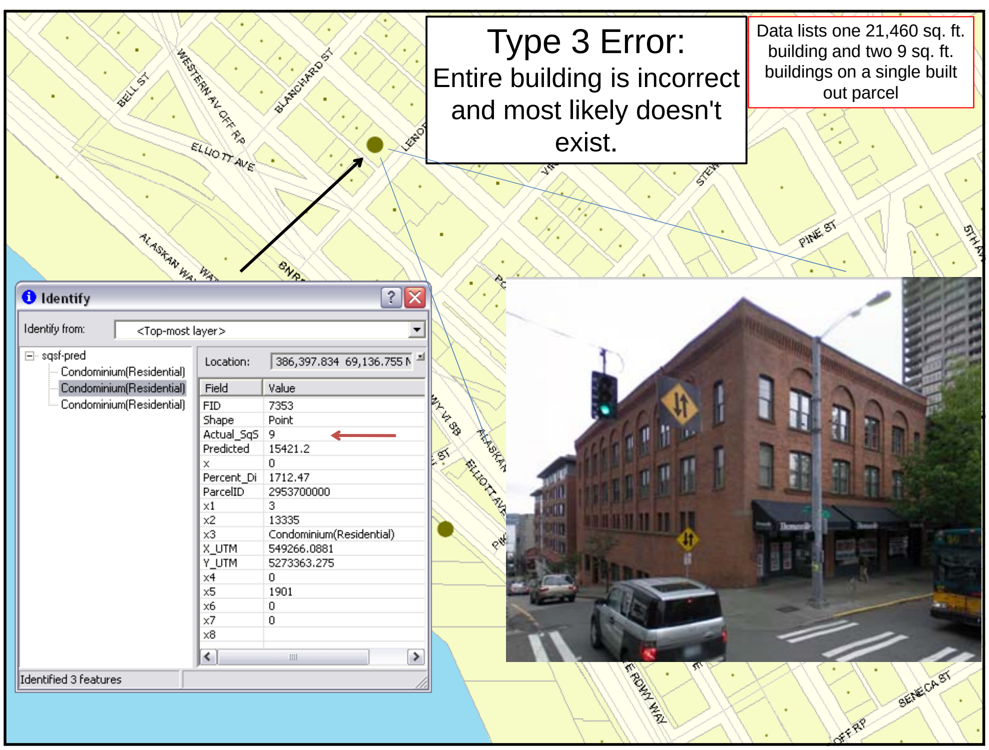
- Attributes: Building Use Code, Land Use Code, etc.
- SVM maps training instances into higher dimensional space.
- Creates hyper planes that have maximum distances from instances as category boundaries.

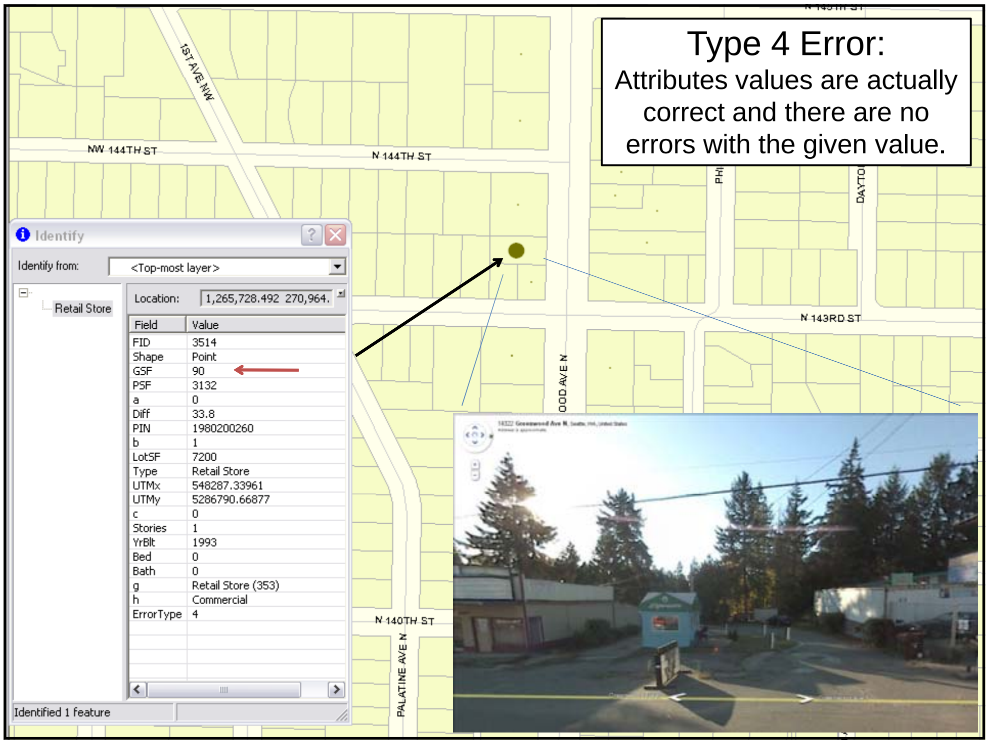


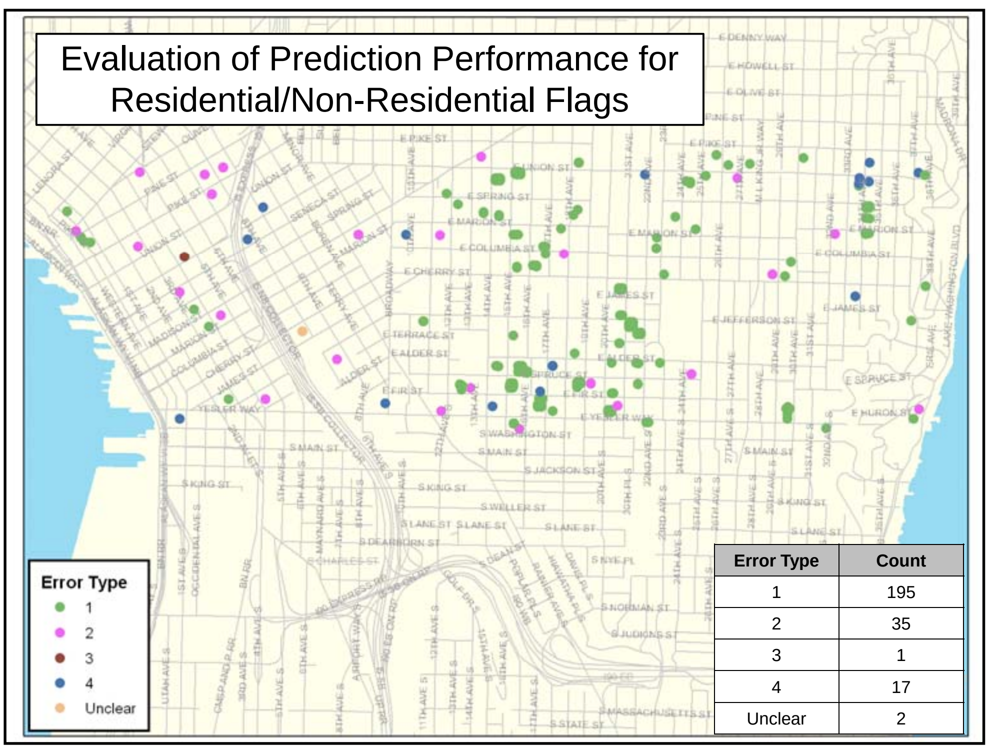



NW 57TH ST
NW 56TH ST
NW MARKET ST
NW 54TH ST
15TH AVE NW
14TH AVE NW
NW 53RD ST
NW 52ND ST
NW 51ST ST
NW 49TH ST
NW 48TH ST
Identify
Identify from: <Top-most layer>
Restaurant(Fast Food)
Location: 1,260,355.236 247,255.123 Fee
Field Value
FID 509
Shape Point
GSF 649
PSF 630602.4
a 0
Diff 970.65
PIN 2768300510
b 1
LotSF 4000
Type Restaurant(Fast Food)
UTMx 546853.04754
UTMy 5279521.42287
c 0
Stories 99
YrBlt 1988
Bed 0
Bath 0
g Restaurant - Table Service (350)
h Commercial
ErrorType 0
Identified 1 feature
Type 2 Error:
Prediction is incorrect, but due to obvious error in data


Type 3 Error:
Entire building is incorrect and most likely doesn't exist.
Data lists one 21,460 sq. ft. building and two 9 sq. ft. buildings on a single built out parcel
BELL ST
WESTERN AV OFF RP
BLANCHARD ST
ELLIOTT AVE
ALASKAN WAY
PINE ST
Identify
Identify from: <Top-most layer>
sqsf-pred
Condominium(Residential)
Condominium(Residential)
Condominium(Residential)
Location: 386,397.834 69,136.755 N
Field Value
FID 7353
Shape Point
Actual_SqS 9
Predicted 15421.2
x 0
Percent_Di 1712.47
ParcelID 2953700000
x1 3
x2 13335
x3 Condominium(Residential)
X_UTM 549266.0881
Y_UTM 5273363.275
x4 0
x5 1901
x6 0
x7 0
x8
Identified 3 features
SENECA ST




## Evaluation of Prediction Performance for Residential/Non-Residential Flags




| Error Type | Count |
|---|---|
| 1 | 195 |
| 2 | 35 |
| 3 | 1 |
| 4 | 17 |
| Unclear | 2 |