Final Research Report
Agreements T4118 Task 87, A72 196
VMT Reductions

SIDEWALK DATA IN KING COUNTY'S URBAN GROWTH BOUNDARY

by
Anne Vernez Moudon, Professor
Bumjoon Kang and Jason Scully, Doctoral Research Assistants
Orion Stewart, MUP, Research Scientist
URBAN FORM LAB (UFL)
University of Washington, Box 354802
Seattle, Washington 98195

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
Seattle, Washington 98105-4631

Washington State Department of Transportation Technical Monitors
Paula Reeves, Transportation Planning Supervisor
Ed Spilker, Transportation Planning Specialist

Prepared for
The State of Washington
Department of Transportation
Lynn A. Peterson, Secretary

June 2013

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 806.2		

4. TITLE AND SUBTITLE	5. REPORT DATE
SIDEWALK DATA IN KING COUNTY'S URBAN GROWTH BOUNDARY	June 2013
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	**8. PERFORMING ORGANIZATION REPORT NO.**
Anne Vernez Moudon, Bumjoon Kang, Jason Scully, Orion Stewart	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	**10. WORK UNIT NO.**
Washington State Transportation Center (TRAC) University of Washington, Box 354802 University District Building; 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631	11. CONTRACT OR GRANT NO. Agreement T4118 Task 87
12. SPONSORING AGENCY NAME AND ADDRESS	**13. TYPE OF REPORT AND PERIOD COVERED**
Research Office Washington State Department of Transportation Transportation Building, MS 47372 Olympia, Washington 98504-7372 Project Manager: Kathy Lindquist, 360-705-7976	Research Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT:

This report describes the development of geospatial sidewalk data for the King County Urban Growth Area. Prior to the development of this data set, sidewalk data in King County were limited to select jurisdictions and existed in multiple, sometimes incompatible, formats.

Existing sidewalk data were collected from 30 of 40 jurisdictions and standardized to a geographic information system (GIS) data format that stores sidewalk coverage as attributes of King County street network centerlines. For each street segment, each sides was coded as full, no, or partial sidewalk coverage. An automated coding method was developed to standardize existing data when possible. Sidewalk coverage for jurisdictions with no existing sidewalk data or with data formats incompatible with automated coding were coded manually with the aid of internet mapping resources. A total of 27 jurisdictions, including unincorporated King County, were manually coded.

Overall rates of agreement between automated and manual coding were 0.95, with higher rates of agreement for street segments with full and no (0.94 to 0.97) compared to partial sidewalk coverage (0.70 to 0.72).

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Sidewalks, Pedestrian travel, Walking, Pedestrian infrastructure, GIS data	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None		

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Department of Transportation or Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Table of Contents

Figures

Tables

Executive Summary

This report describes the development of the Urban Form Lab King County Sidewalk Data set (kcn_skw_20120507.csv). This data set contains information about sidewalks within King County's Urban Growth Boundary.

The sidewalk data are provided in an attribute table that can be appended to the King County Street Network data set. Two fields (swk_left and swk_right) store information about the presence or absence of a sidewalk on either the left or right side of each street segment. The following codes were used:

0 = absence of a sidewalk;
1 = presence of a sidewalk; and
2 = partial presence of a sidewalk on the segment.

The data on sidewalks were assembled from two sources:

1) Data in GIS, tabular or description formats supplied by local city governments and jurisdictions in King County; and
2) Internet mapping websites (Google Maps, Google Street View and Bing Maps).

Sidewalk data were obtained from 30 of the 40 jurisdictions in King County, usually in the form of geographic information system (GIS) shape files. However, these GIS data varied greatly between cities. The lack of compatible sidewalk data formats, along with the lack of any sidewalk data for a sizable portion of the King County Urban Growth Area, necessitated the development of a complete and consistent sidewalk data set. This was done by coding sidewalk attributes to the King County Street Network data set.

The field "swk_method" refers to the method used to code the street network data. When possible, sidewalk data provided by city governments were automatically converted to the 0, 1, 2 coding schema using an algorithm executed in PostGIS 2.0 and R. The word "automatic" in the "swk_method" field refers to street network segments that were coded in this manner. The remaining street network segments—those where spatial sidewalk data were unavailable or incompatible with the automated coding process—were manually coded by a GIS technician using visual reviews of online mapping services. These are identified by a value of "manual" in the "swk_method" field.

This report is divided into two parts. The first part presents a basic introduction to the challenges and our solutions in creating the sidewalk data set. Section 1 discusses the sidewalk data sources used. Section 2 reviews the data structure chosen to represent the sidewalk data and presents some of the strengths and weaknesses of this approach. Section 3 discusses the tradeoffs in using the King County Transportation Network data set as base data to which sidewalk data were linked. Section 4 discusses the automatic and manual coding procedures.

The second part of the report discusses the coding methods in greater detail. Section 5 presents the definition of a sidewalk that guided the manual coding process. Section 6 discusses the manual coding process. Section 7 explains the method used to automatically convert the data provided by the cities into the chosen format. Section 8 presents the method used to validate the automatic coding process. Finally, Appendix A summarizes the metadata for the sidewalk data set and presents ways that the data could be visualized. Algorithms using ArcGIS, PostGIS, and R, which served to transfer the data from the individual jurisdictions to the standardized data set generated in this project, can be made available upon request.

Part I: Main Report

Walking is becoming increasingly recognized as a carbon neutral mode of transportation that not only reduces congestion and pollution, but also increases positive health outcomes. As research on this topic grows, so does the need for data about sidewalks. Nationwide, the lack of complete and accurate sidewalk data has impeded research as to the actual effect of sidewalks on walking. In King County, and the entire Puget Sound region, different agencies, such as the University of Washington (UW), the Puget Sound Regional Council (PSRC), the Washington State Department of Transportation (WSDOT), and Public Health—Seattle & King County (PHSKC), have needed sidewalk data for many years.

Previous efforts at creating sidewalk databases have been limited in scope. Both PSRC and WSDOT have assembled data that cover state routes and major arterials. Neither includes residential areas where people are likely to walk. Also, available data are attached to different street network systems that are not readily compatible—the lengths of street segments in each system are different, the segments are labeled differently, and there is some spatial mismatch between the different systems (Section 3 discusses the differences between street network systems). Perhaps even more problematic is that many available sidewalk data sets do not distinguish between the absence of sidewalks and the absence of data. It is therefore impossible to tell if a street segment is lacking a sidewalk or lacking data about the presence of a sidewalk.

To meet local data needs, many jurisdictions have assembled their own sidewalk data. These data sets were designed to meet the individual requirements of their jurisdictions and vary widely in format. Section 1 reviews the different formats used by jurisdictions with sidewalk data.

The University of Washington Urban Form Lab (UFL) was commissioned by WSDOT to complete a sidewalk data set for the area within the King County Urban Growth Boundary (UGB) in December 2011. Some sidewalk source data had already been collected by UFL, but work to aggregate them into a comprehensive database was stopped due to lack of funding. This report documents the process of developing a complete sidewalk data set for all of King County within the UGB that is based on the King County Street Network data set.

Section 1: Sidewalk Data Sources

UFL previously assembled data from jurisdictions in King County. Between 2008 and 2009 each of the 39 cities in King County's urban growth boundary, as well as King County itself, was contacted to obtain sidewalk data. Of the 40 jurisdictions, we were able to obtain sidewalk data from 21.

In 2011 we received additional sidewalk data for nine cities from the PSRC. The PSRC data for five of these cities were not used because we already had received data directly from those cities. In addition, data for five cities were obtained from former University of Washington doctoral students Gina Lovasi and Amber Pearson, who had previously assembled sidewalk data from various King County jurisdictions.

Overall, our data collection efforts resulted in sidewalk data sets for 30 of the 40 jurisdictions within the King County urban growth boundary (Table 1.1). These data covered about 82% of the land area within the UGB. However their usability varied greatly. Data obtained from the jurisdictions came in four different structures (Table 1.1 column 3) (Figure 1.1):

1). Descriptive Data: data in the form of written descriptions regarding the location and quality of sidewalks in relation to each street within the jurisdiction (3 cities).

2). Double-Line Data: GIS data in the form of polylines flanking street segments on either side wherever a sidewalk was present (17 cities).

3). Single-Line Data: GIS data in the form of a single polyline matching the city's own street segment line, which included tabular data indicating whether the sidewalk was on the right or left side of the street (5 cities); and

4). Polygon Data: GIS data in which sidewalks, walkways, paths, and impervious surfaces were depicted as polygons that showed their size and shape (5 cities).

Figures 1.2 and 1.3 summarize the data available from the jurisdictions and their format. The different data formats meant that finding a way to merge all the data sets into one sidewalk network would be a challenge.

(A) (B) (C)

Figure 1.1: Three GIS Formats . (A) Single-line data depicted in Bellevue; (B) Double-line data as depicted in Seattle; and (C) Polygon data as depicted in Federal Way.

Table 1.1: Data Availability and Format by Jurisdiction

Jurisdiction Name	Jurisdiction Data Available?	Data Source	Data Format*
Algona	No	NA	NA
Auburn	Yes	City of Auburn	Polygon
Beaux Arts	Yes	City of Beaux Arts	Description
Bellevue	Yes	City of Bellevue	Double-Line
Black Diamond	No	NA	NA
Bothell	Yes	City of Bothell	Polygon
Burien	Yes	Lovasi & Pearson	Double-Line
Carnation	Yes	City of Carnation	Description
Clyde Hill	Yes	Lovasi and Pearson	Double-Line
Covington	Yes	City of Covington	Description
Des Moines	Yes	PSRC	Single-Line
Duvall	Yes	Lovasi & Pearson	Double-Line
Enumclaw	No	NA	NA
Federal Way	Yes	PSRC	Single-Line
Hunts Point	No	NA	NA
Issaquah	Yes	City of Issaquah	Polygon
Kenmore	Yes	Lovasi & Pearson	Double-Line
Kent	Yes	City of Kent	Double-Line
Kirkland	Yes	City of Kirkland	Double-Line
Lake Forest Park	Yes	Lovasi & Pearson	Double-Line
Maple Valley	Yes	City of Maple Valley	Double-Line
Medina	No	NA	NA
Mercer Island	Yes	City of Mercer Island	Polygon
Milton	Yes	Lovasi & Pearson	Double-Line
Newcastle	Yes	City of Newcastle	Double-Line
Normandy Park	Yes	Lovasi & Pearson	Double-Line
North Bend	Yes	City of North Bend	Polygon
Pacific	No	NA	NA
Redmond	Yes	City of Redmond	Double-Line
Renton	Yes	City of Renton	Single-Line
Sammamish	No	NA	NA
SeaTac	Yes	PSRC	Single-Line
Seattle	Yes	City of Seattle	Double-Line
Shoreline	Yes	City of Shoreline	Double-Line
Skykomish	No	NA	NA
Snoqualmie	Yes	PSRC	Single-Line
Tukwila	Yes	City of Tukwila	Double-Line
Woodinville	Yes	City of Woodinville	Double-Line
Yarrow Point	No	NA	NA
Unincorporated	No	NA	NA

** All data, except descriptions, are in the form of GIS shapefi*

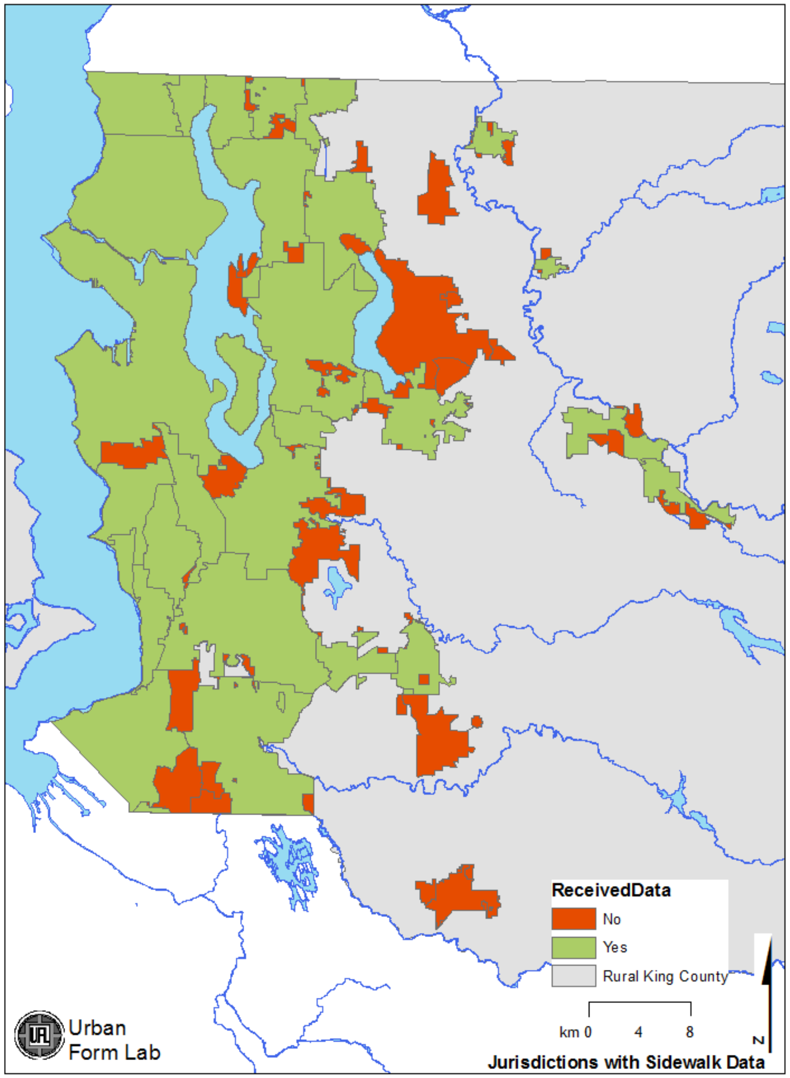


Figure 1.2: Jurisdictions with Sidewalk Data

City of Skykomish outside of map extent had no data.

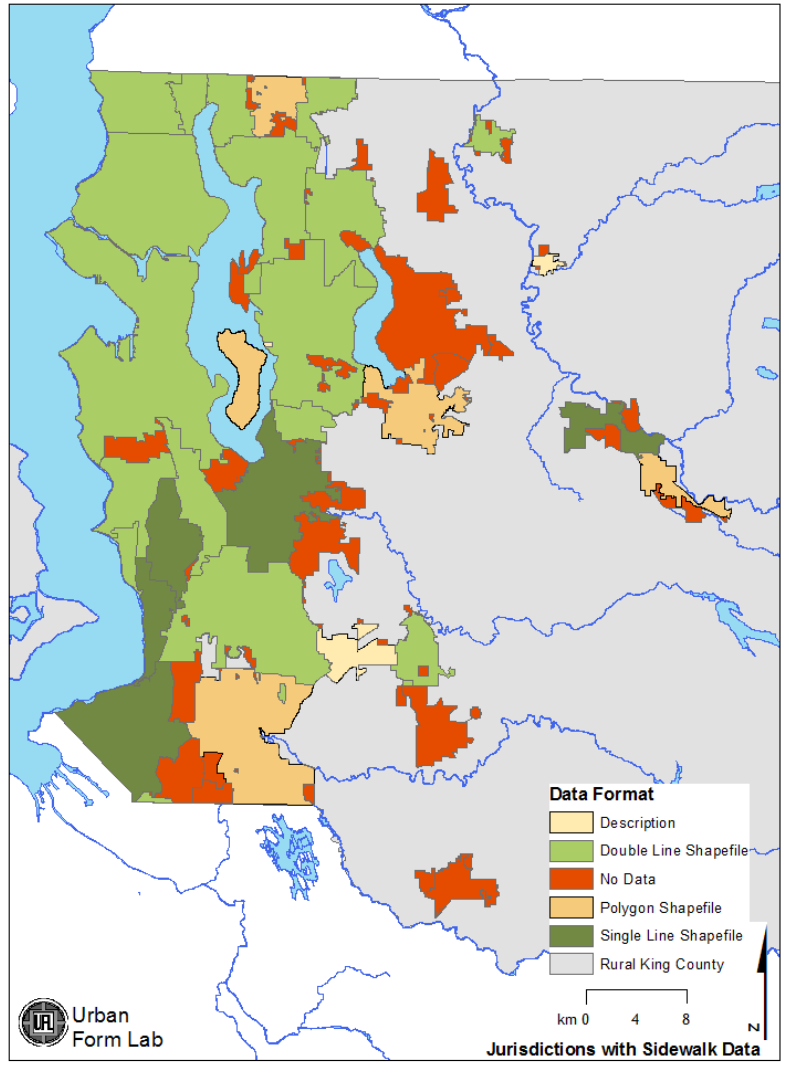


Figure 1.3: Sidewalk Data Format by Jurisdiction

City of Skykomish outside of map extent had no data.

Section 2: UFL Data Structure

The multiple formats and data types we received did not allow for easy comparison of street network characteristics with sidewalk characteristics. They also made comparisons between sidewalk characteristics in jurisdictions with different formats very difficult, if not impossible. It was therefore necessary to convert all the sidewalk data into one format and preferably in one data set using a data structure that would meet our analytic needs. Our goal in choosing a data structure was to maximize the completeness of the data relative to all individual street segments in the county and to insure the quality of the data. We also wanted to make sure the data were accessible to a wide range of users.

It might have been possible to merge the single-line, double-line, and polygon data into one integrated data set with a line structure. However, this method would have created a number of additional challenges. First, analyzing sidewalk coverage in relation to street characteristics would be limited because the sidewalk data would not be directly associated with the segments of street networks. Second, there would be discontinuities in the sidewalk network at jurisdictional boundaries, which would have to be checked manually and lines would have to be knitted together—a time-consuming enterprise.

The UFL decided to code the sidewalks in a tabular format linked to the individual street segments of an existing street network database in GIS. Coding sidewalks in this fashion allows for consistency and continuity in the sidewalk network segment IDs, as well as for ready analytic integration with the street network itself. Such a sidewalk network data set allows for detailed analyses of sidewalk "completeness," which calculates the ratio of streets lined by sidewalks. Sidewalk completeness is an important aspect of research on walking as a travel mode because high levels of walking are associated with motorized street travel and transit use.

The decision to employ a tabular data format raised the question of which street network data set would be best to attach to sidewalk data to. Our options were King County's Transportation network data set, the PSRC's street network data set, or ESRI's commercial NAVTEQ data. We chose to use King County's street network data set. Our rationale for doing so is discussed in Section 3.

Next it was necessary to determine the level of detail used in coding the sidewalks. It was deemed important to record sidewalks by their relative location along each street segment (i.e. left side, right side, or both sides of the street). In the more recently developed areas of the county, many streets (including both minor streets and major arterials) only have a sidewalk on one side. It was also important to record the continuity of the sidewalk as many gaps still exist in the current infrastructure. Hence the UFL decided to code three conditions of sidewalk coverage for each street segment: full coverage, partial coverage, and no coverage. Partial sidewalk coverage was defined as between 10 and 90 percent of the street segment having a sidewalk. Anything over was considered full coverage, anything under was considered no coverage. Thus each street segment was coded for full, partial, or no coverage on either or both its right and left sides (Fig.2.1)

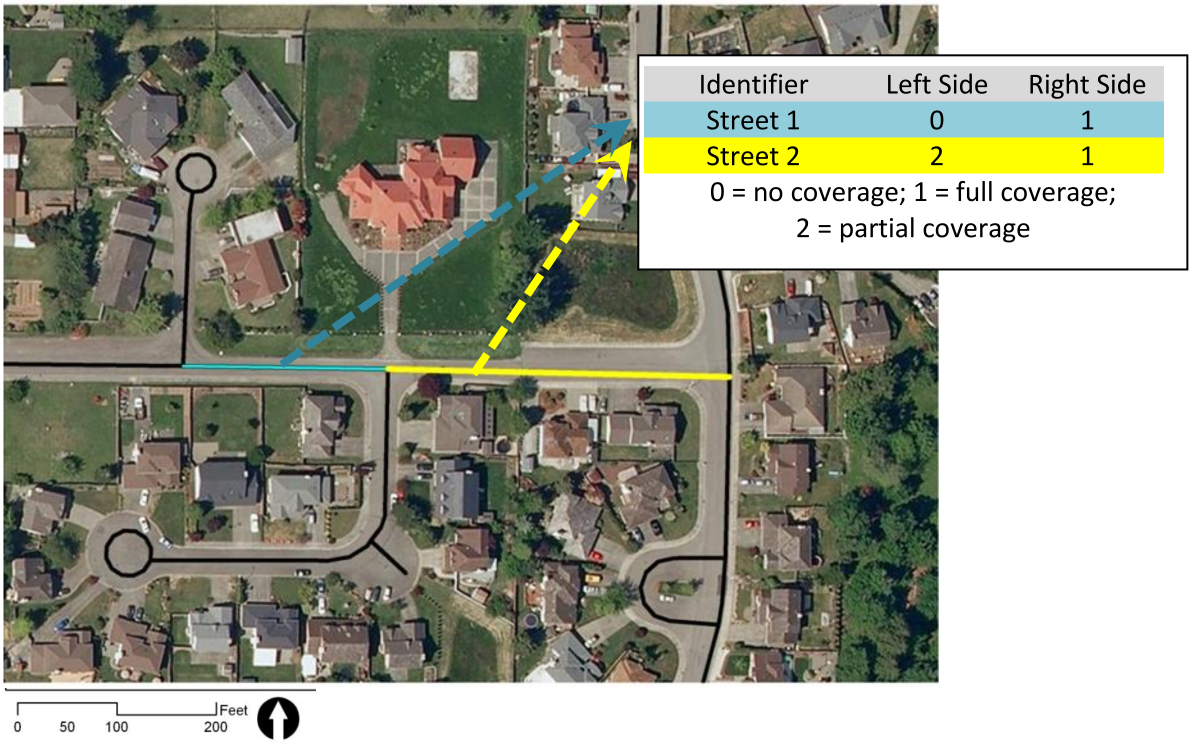


Figure 2.0.1: Example of Data Schema. Street centerlines are represented by polylines (shown in black). The attribute table contains fields indicating sidewalk coverage on the left and right side of the street. The yellow segment has complete sidewalks along the right (south) side of the street but only has sidewalks along about half the left (north) side. The blue segment has no sidewalks on its left side and full coverage on its right side.

Additional sidewalk attributes (e.g., condition, planting strips, ADA wheelchair ramps, presence of street furniture, number of intersecting driveways, etc.) would be desirable for a robust database. They are beyond the scope of the current project, but could be added at a later date. By coding sidewalk data to a street network, the main limitation of UFL data structure is that graphic displays will not readily show whether sidewalks exist on one or both sides of streets. However, color coding of the polylines can be used to indicate the different sidewalk conditions and types (Figs. 1 and 2, Appendix 1).

Table 2.1 summarizes the data available. A full description of the metadata is provided in Appendix 1.

Table 2.1: Data Dictionary

Field	Values	Description
tlink_id	Numerical ID	Identifier linking each segment to the King County Street Network Data set.
Swk_left	0, 1, 2	Sidewalk coverage on the left side of street segment.
Swk_right	0, 1, 2	Sidewalk coverage on the right side of street segment.
swk_date	year-month-day	Date the street segment was coded with sidewalk data by UFL. (*Note:* this is not the data date.)
swk_method	"auto" "manual"	Denotes whether street segment was coded manually or using an algorithm.

Section 3: Selection of the Street Network Data Set

Three GIS databases of street networks covering King County were considered for use in this project:

- the Puget Sound Regional Council (PSRC) street network data set;
- the King County (KC) Street Network Data set; and
- the NAVTEQ commercial database available through ESRI StreetMap Premium.

All three street network data sets are polylines representing street centerlines and they all have an extent at least as great as King County. They differed based on their primary purposes. The PSRC data were developed for traffic modeling. They offered a somewhat abstract representation of the street network but contained good information on traffic characteristics. The KC data were designed for management and trip planning. They were the most detailed but were limited in their attributes on traffic conditions. The NAVTEQ data were designed for routing purposes and offered a simple yet accurate representation of the street network with attributes that travelers could use for planning a trip.

The three street network data sets were reviewed for the following characteristics:

- spatial accuracy;
- links to other street network data sets;
- street and traffic conditions attributes that identify the type of segments that could be included or excluded in future analyses; and
- street and traffic conditions attributes that complement sidewalk attributes.

Tables 3.1 and 3.2 summarize the characteristics of the data sets. The PSRC data offer a limited and imprecise spatial representation of the street network. The KC data offer a detailed representation of the street network, including semi-public streets and small cul-de-sacs. The NAVTEQ data are spatially precise and appear to represent newer developments well. They are not particularly detailed, but were considered adequate for our purposes. No data set was readily linkable to any other data set.

The NAVTEQ data offer the best (most extensive and clear) set of attributes for use in excluding data and identifying traffic conditions. The KC data have less extensive attributes, but are still adequate for data development and analytical purposes. The PSRC data have limited segment attributes, which could cause difficulties during coding and analysis.

The NAVTEQ data provide a precise and sufficiently detailed representation of the street network in King County. They contain attributes that allow for clear identification of street segments that relate only to vehicular traffic and therefore would not need to be coded, and of attributes that reflect the traffic conditions that sidewalks can help mitigate. However, within the King County UGB, it contains about 60 % fewer street segments than found in the KC data set. Additionally, NAVTEQ street segment lengths are of a more consistent length, suggesting that fewer ‘abnormal’ street segments, such intersection-only segments, exist.

Some of the data characteristics are shown for an area in the city of Auburn (Figure 3.3).

In the final selection of the appropriate data set, data access trumped technical advantages. As a commercial database, NAVTEQ can only be used under the terms of its licensing agreement, which prevents data sharing with other agencies. Data would also need to be purchased periodically for updating. As a result, the KC data were chosen for their accessibility, widespread use in various county agencies, and their higher degree of accuracy compared to the PSRC data.

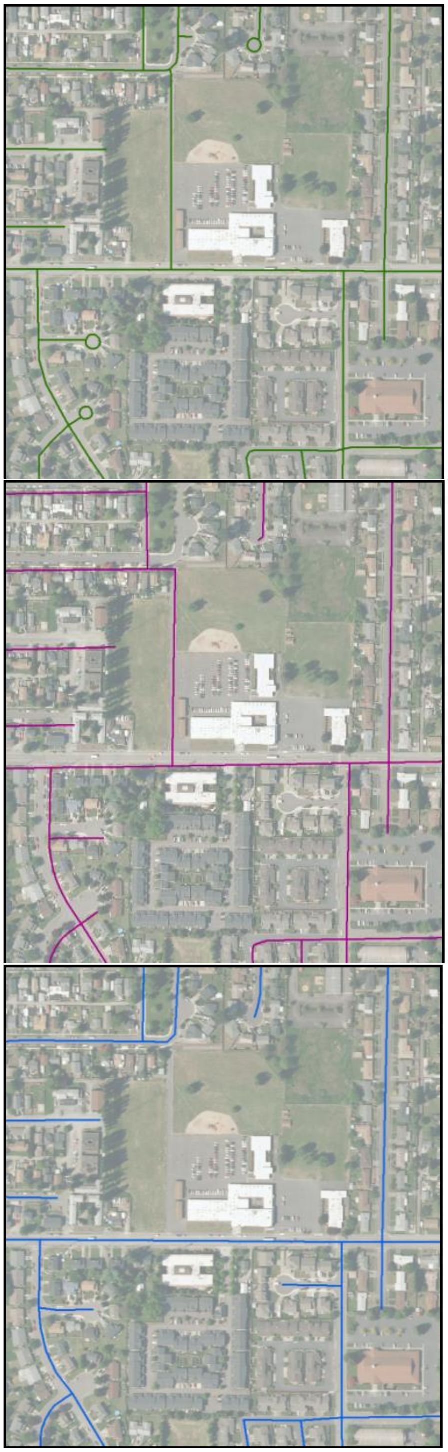
Table 3.1: Length Characteristics of Street Segments within the King County Urban Growth Area

	n (segments)	Mean segment length	Stand. dev.	Min.	Max.
KC	149,406	483.4 ft	1064.6 ft	1.3 ft	166,981.5 ft
PSRC	99,045	552.6 ft	960.7 ft	0.3 ft	59,713.4 ft
NAVTEQ	89,228	424.1 ft	454.5 ft	0.1 ft	25,386.6 ft

Table 3.2: Summary of Street Network Review

	KC	PSRC	NAVTEQ
Spatial accuracy	best	inadequate	adequate
Alignment	√		√
Complex Intersections	√		√
Detailed Network*	√	√	
New Developments	√		√
Links	adequate	adequate	Adequate
Unique ID**			
Addresses***	√		√
Exclusion attributes	adequate	inadequate	Best
Allowed Travel Modes	√		√
Paved			√
Parking Lot	√		√
Complementary attributes	adequate	best	best
Traffic Volume			
Street Class	√	√	√
Speed Limit	√	√	√
Lanes		√	√
Slope			

* including minor and other streets that may be important to walking
** potential to link segments to those of the local jurisdictions
*** segments include address ranges



King County
Accurately aligned with aerial imagery
34 segments, 1.95 miles
Includes minor cul-de-sacs, but missing at least one cul-de-sac, which is likely a newer development
Includes limited access (gated) street through school ground

PSRC
Slightly misaligned with aerial imagery
27 segments, 1.85 miles
Includes some minor cul-de-sacs, but missing at least two cul-de-sacs, which are likely newer developments
Includes limited access (gated) street through school ground

NAVTEQ
Aligned with aerial imagery
27 segments, 1.85 miles
Excludes minor cul-de-sacs, but includes a cul-de-sac that is absent in KC and PSRC data, which is likely a newer development
Excludes limited access (gated) street through school ground

Figure 3.1: Sample of the Three Network Data sets in an Area of the City of Auburn

Section 4: Manual and Automatic Coding

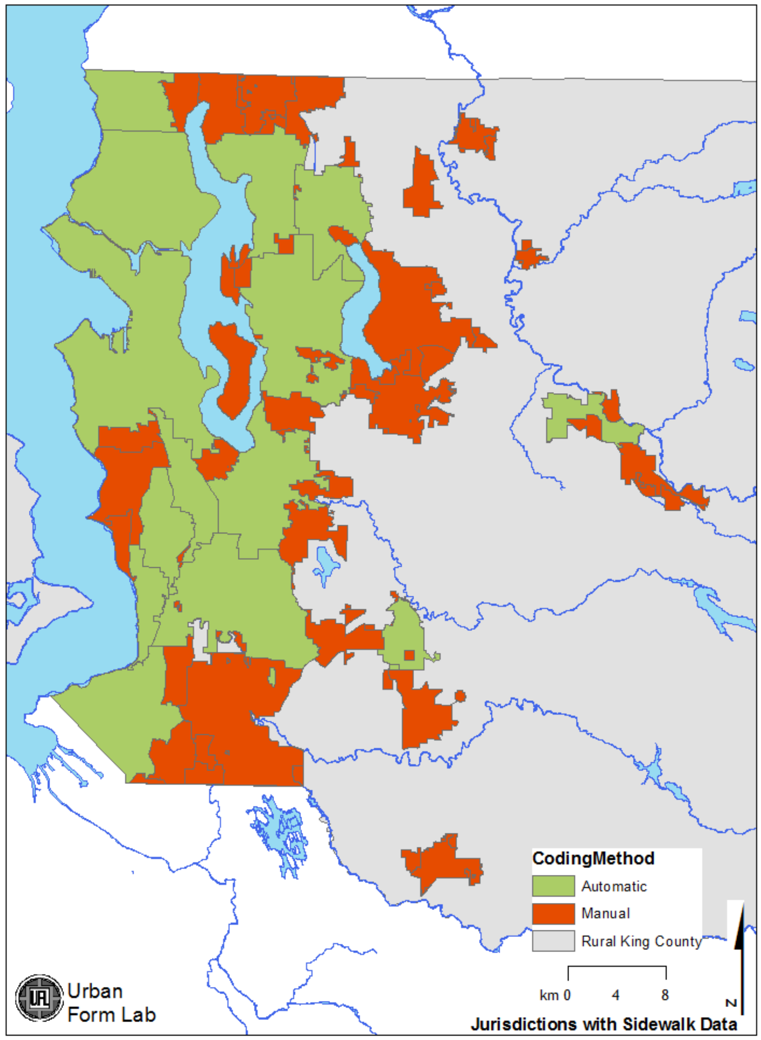


Figure 4.1: Coding Method by Jurisdiction

City of Skykomish not depicted

Sidewalk attributes were coded to street segments using two methods: manual coding of areas that did not have data or for which the data we received were unsuitable for conversion (described in detail in Section 6); and automatic coding using an algorithm to convert the different existing sidewalk data sets to our chosen structure and to consolidate them into one data set (described in detail in Section7).

Of the 30 sidewalk data sets received, 13 could be automatically coded. The remaining 17 cities were manually coded along with the other 10 jurisdictions that did not have data. The three sets of descriptive data could not be automatically coded because they were not in GIS formats. The five polygon data sets were excluded from automatic coding due to difficulties in calculating their lengths and establishing the continuity of the sidewalk network. Also excluded from the automatic coding were the seven sets of Lovasi and Pearson data, because of concerns about their accuracy and completeness. Finally, two sets of double-line data were excluded from automatic coding because they had been manually coded prior to the development of the algorithm.

Table 4.1 (columns 3 and 4) lists the coding method and data source used to code sidewalks in each jurisdiction. Figure 4.1 shows which jurisdictions were manually or automatically coded.

Of the 123,554 street segments within the King County UGB, 79,928 or about 64.7% were automatically coded. The remaining 35.3% were manually coded. Together, the jurisdictions that were automatically coded accounted for 60% of the land area within the King County UGB.

In the manual coding process, a GIS technician selected an individual street segment and then looked at one or more of many possible data sources to determine whether it had sidewalks associated with it. If jurisdiction sidewalk data were available, the technician would use that source, otherwise the technician used Internet sources (GoogleMaps, GoogleStreetView and Bing Maps). In many cases, a combination of data sources was necessary to code the same segment.

The manual coding process was time-consuming, at a pace of about 100 segments per hour, which varied with the quality of the source data (whether data came from the jurisdictions or from Internet sites such as GoogleMaps) and the complexity of the street network. The frustratingly slow pace of the manual coding process led one of the researchers to create the automated process for coding the data that is presented in this report.

The automatic coding process consists of an algorithm executed primarily in PostGIS. It relies on buffers created in GIS on either side of a street segment to capture the presence and attributes of sidewalks. GIS data recording double lines for the sidewalks could be directly used in the algorithm. GIS data with single lines had to be offset to the left and the right sides of street segments to be effectively converted into double lines. The algorithm creates a buffer on either side of every street segment. If a sidewalk is present in that buffer, the angles of the sidewalk and the street network are compared to each other. If the sidewalk's angle differs too much from that of the street segment, the sidewalk is not considered to be associated with that street. Next, the length of the sidewalk is compared to the length of its corresponding street. Some tolerance (extra length) is added to the total length of the sidewalk to account for intersections, curb cuts, and driveways. Figure 4.2 illustrates the elements of the algorithm. Panel A shows a typical street segment with an adjacent sidewalk. Panel B shows the elements of the algorithm used to automatically code a street segment as having a sidewalk.

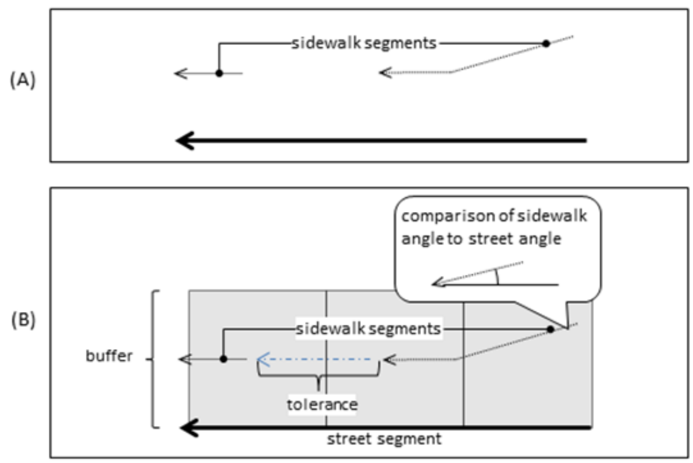


Figure 4.2: Elements of the Automated Coding Algorithm

A: Street segment with an adjacent sidewalk.

B: Elements of the algorithm used to automatically code a street segment as having a sidewalk.

Table 4.1: Type of Coding and Data Source by Jurisdiction

Jurisdiction Name	Type of Coding	Primary Data Source Used in Coding (Data Type)	Secondary Data Source Used in Coding (Data Type)	Comments
Algona	Manual	Internet	NA	
Auburn	Manual	Jurisdiction (polygon GIS data)	NA	
Beaux Arts	Manual	Jurisdiction (description)	NA	
Bellevue	Automatic	Jurisdiction (double-line GIS data)	NA	
Black Diamond	Manual	Internet	NA	
Bothell	Manual	Internet	Jurisdiction (polygon GIS data)	The secondary data were not accurate or complete, but were still helpful in manual coding.
Burien	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate or complete, but were still helpful in manual coding.
Carnation	Manual	Internet	Jurisdiction (description)	The secondary data were not accurate or complete, but were still helpful in manual coding.
Clyde Hill	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate or complete, but were still helpful in manual coding.
Covington	Manual	Internet	NA	
Des Moines	Automatic	PSRC (single-line GIS data)	NA	
Duvall	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate and complete, but helped in coding.
Enumclaw	Manual	Internet	NA	
Federal Way	Automatic	PSRC (single-line GIS data)	NA	
Hunts Point	Manual	Internet	NA	
Issaquah	Manual	Jurisdiction (polygon GIS data)	NA	
Kenmore	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate and complete, but helped in coding.
Kent	Automatic	Jurisdiction (double-line GIS data)	NA	
Kirkland	Automatic	Jurisdiction (double-line GIS data)	NA	
Lake Forest Park	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate and complete, but helped in coding.
Maple Valley	Automatic	Jurisdiction (double-line GIS data)	NA	
Medina	Manual	Internet	NA	
Mercer Island	Manual	Jurisdiction (polygon GIS data)	NA	

Milton	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate and complete, but helped in coding.
Newcastle	Manual	Jurisdiction (double-line GIS data)	NA	Manually coded before the decision to automate coding.
Normandy Park	Manual	Internet	Lovasi & Pearson (double-line GIS data)	The secondary data were not accurate and complete, but helped in coding.
North Bend	Manual	Jurisdiction (polygon GIS data)	NA	
Pacific	Manual	Internet	NA	
Redmond	Automatic	Jurisdiction (double-line GIS data)	NA	
Renton	Automatic	Jurisdiction (single-line GIS data)	NA	
Sammamish	Manual	Internet	NA	
SeaTac	Automatic	PSRC (single-line GIS data)	NA	
Seattle	Automatic	Jurisdiction (double-line GIS data)	NA	
Shoreline	Automatic	Jurisdiction (double-line GIS data)	NA	
Skykomish	Manual	Internet	NA	
Snoqualmie	Automatic	PSRC (single-line GIS data)	NA	
Tukwila	Automatic	Jurisdiction (double-line GIS data)	NA	
Woodinville	Manual	Jurisdiction (double-line GIS data)	NA	Manually coded before the decision to automate coding.
Yarrow Point	Manual	Internet	NA	
Unincorporated	Manual	Internet	NA	

Part II: Coding Methods in Detail

Both the manual and automatic coding processes required technical knowledge and assumptions. This part of this document discusses these technical aspects in greater detail.

Section 5: Sidewalk Definition

An important first step in mapping sidewalks is to establish the definition of a sidewalk. Unfortunately, we were unable to verify how the various jurisdictions that sent us data defined sidewalks. The UFL used the sidewalk standards described in Fig 2-002 (page 2-26) of the 2007 King County Road Design and Construction Standards document. (As of May 2012, this document is available online at: http://www.kingcounty.gov/transportation/kcdot/Roads/EngineeringServices/RoadStandards2007.aspx. Itis assumed that the sidewalk definitions used by the various jurisdictions are similar and compatible with the UFL definition.

According to King County Road Design and Construction Standards document, sidewalks should be elevated and separated from the roadway with curbstone. The standards call for the width of sidewalks to vary by road type. For example, sidewalks along arterials should be at least 6.5-feet wide. However, because most of the source data provided by the jurisdictions did not have information about sidewalk widths, we were unable to follow this specification. Additionally, it would have been prohibitively time consuming to estimate sidewalk widths from Internet imagery during the manual coding process.

The following rules were used to identify sidewalks. Sidewalks are:

- made out of concrete (gravel, dirt paths and wide shoulders are explicitly excluded from this definition);
- elevated;
- separated from the roadway by curbstone; and
- within approximately 20 feet of the street segment (as long as they are within 20 feet they can be next to parking, bike lines, or behind planting strips).

Some street segments have pedestrian walkways running alongside them. Pedestrian walkways within 20 feet of the street segment are considered sidewalks. Figure 5.1 left shows a street (in blue) with a pedestrian pathway (in red). For our purposes, this pedestrian walkway is considered a sidewalk due to its proximity to the street segment. In the photo on the right, 106th Lane NE (in black) is considered to have a sidewalk on its left side even though the sidewalk is separated from the street segment by plantings and parking.

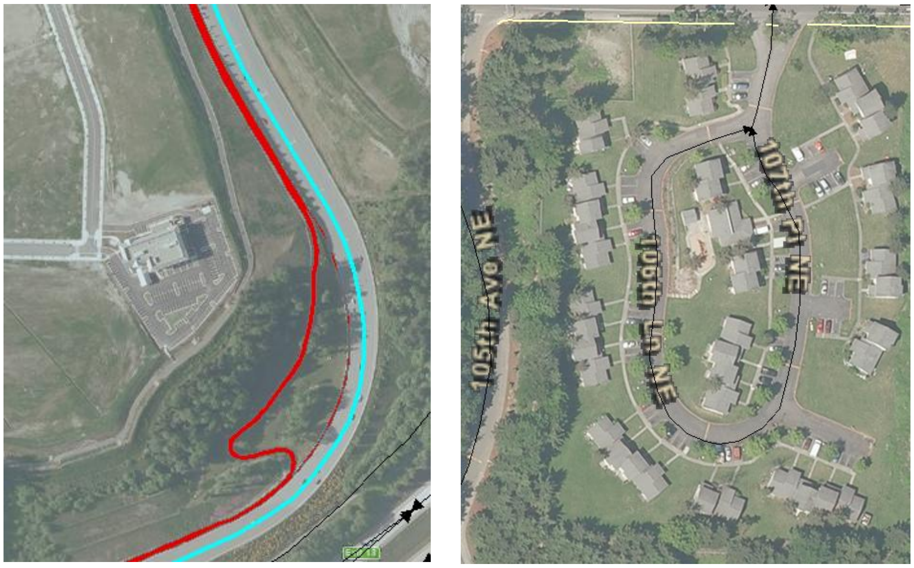


Figure 5.1: Sidewalk Examples. [Left] A pedestrian walkway (red line) is considered as sidewalk (street segment in blue). [Right] A pedestrian walkway behind planting (light grey in the photograph) is considered as sidewalk (street segment in black).

Section 6: Manual Coding

A total of 27 jurisdictions, including unincorporated King County, were manually coded. Data came from the Internet, Pearson and Lovasi, and cities with sidewalk data in GIS that could not be automatically coded .

Manual coding relied on the best judgment of the coders using the sidewalk definition presented in Section 5. The manual coding process using Internet-based data sources was straightforward. The street segment of interest was located in Google StreetView, Google Maps, or Bing Maps. The satellite image or the StreetView image was then used to determine if there were sidewalks on either side of the segment. (Because Internet mapping sites are constantly updating their images, the actual age of the source data for the manual coding process is unknown). For each street segment in the King County Transportation data set, the coder entered a 1 if there was a sidewalk, a 0 if there was no sidewalk or 2 if there was a partial sidewalk. The use of multiple sources of data provided different images of the same street segment. This was helpful in cases where one image was unclear due to obstructions by vegetation or lack of color contrast between sidewalks and adjacent areas..

Three types of errors could emerge in the manual coding process: data errors, coder errors, and errors resulting from ambiguous conditions. Data errors might occur because of out-of-date street network data, satellite imagery or StreetView photos. Coder errors could result from a number of factors. First, the King County Street Network data has many very short street segments, most of which are under 25 feet and located in intersections. Such segments would frequently go unseen by the coders. To save time during the manual coding phases of the project, these segments were coded as having no sidewalks on either side of the street regardless of the actual sidewalk conditions on the adjacent streets. Second, streets with a median strip were often represented by two parallel segments, which could lead to inconsistent coding of sidewalk to one or both parallel segments. Third, partial sidewalks around culs-de-sac were challenging to code. This difficulty was exacerbated by the fact that street segments in culs-de-sac were represented in a variety of forms, which could result in inconsistent coding (Figure 8.1).

Finally, confusing street layouts made assigning sidewalk codes difficult. For example, intersections involving more than two cross-streets oftentimes had sidewalks that could be associated with any number of street segments. In other cases, the streets represented by the GIS data did not match the streets depicted in the aerial imagery. In confusing situations, the coders used their best judgment in assigning the appropriate sidewalk codes to the adjacent street segment.

Section 7: Automatic Coding: Basic Principles Behind the Algorithm

Sidewalk data were automatically coded to our chosen data structure for 13 of the 40 cities and unincorporated King County (Table 4.1; Figure 4.1). The different geometries of the King Co street network and the various city data sets prevented our using a simple spatial join function in ArcGIS to establish the presence or absence of a sidewalk. We therefore needed to develop an algorithm which could calculate the geometrical relationships between streets and sidewalks.

Figure 7.1 shows the steps the algorithm used to automatically convert the data. The basic principle behind the algorithm was to spatially match streets segments in the King Co Street Network with those sidewalk segments in the jurisdictions' data. Expectedly, the matching process was simplest and the results most reliable when street and sidewalk segments were parallel to each other. Unfortunately, because this was not the prevailing situation, the algorithm had to process not only different street and segment lengths, but also different segment geometries. Figure 7.1-Step 1 depicts a street segment that has partial sidewalk coverage on its right side. The street and sidewalk polylines are not fully parallel (they are at different angles) and their segment lengths are different. The first step was to split the polylines representing both sidewalks and streets into smaller sub-segments (Step 2 of Figure 7.1), in order to simplify calculations for non-parallel and curvilinear segments and to minimize misclassification errors. Segments were split at vertices and at 10-foot intervals.

In Step 3, buffers were created that extended from each street's sub-segment for a distance of **B**. Next, sidewalk sub-segments within the buffers were evaluated to determine whether they were associated with the street sub-segment. This was done in Step 4 by comparing the angle of the sidewalk to the angle of the street. A maximum difference between the angles was set at **ϕ**degrees. If the difference in angles between these two sub-segments (**θ** degrees) was less than **ϕ**, then the sidewalk sub-segment was considered to be associated with the nearby street segment.

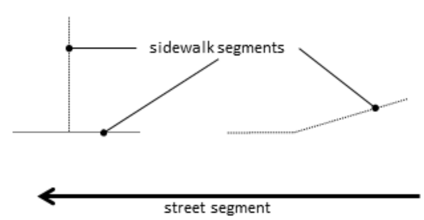
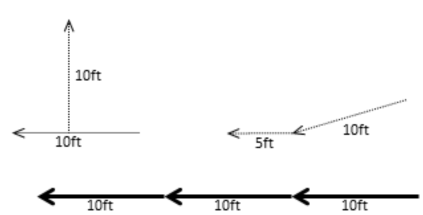
The level of sidewalk coverage was determined by calculating the lengths of the associated sidewalk sub-segments within each buffer (Step 5). These lengths were then summed to calculate the total length of the sidewalk associated with a corresponding street segment (Step 6). Because there are many breaks in sidewalk coverage due to curb cuts, driveways or intersections, a tolerance (**T**) (extra bit of length added to the sidewalk coverage along the street segment) was added to the total length of each sidewalk (Step 7). If the sidewalk length plus **T** was greater than 90 percent of the total length of street segment, then the street segment was considered to have full coverage. If it had less than 90 percent but more than 10 percent, it was considered as partial coverage. And if it had less than 10 percent of the street segment, it was considered as no coverage (Step 8). In Figure 7.1 Step 7, the calculated sidewalk length is 23 feet plus **T**, and the street length is 30 feet.

In summary, the algorithm includes the following processes:

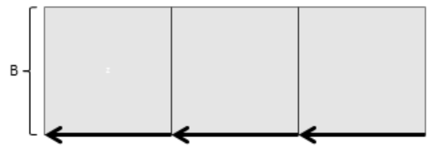
1. Select a street segment and its adjacent sidewalk polylines.
2. Split all street and sidewalk segments into sub-segments at every vertex and at 10-foot intervals.
3. Create two buffers along each side of a street sub-segment that extend from the right and left sides of the sub-segment (buffer size = **B**).
4. Exclude sidewalk sub-segments that have angle differences from the street sub-segment (**θ)** that are larger than **ϕ**degrees (**θ** > **ϕ**).
5. Within each buffer, calculate the lengths of the sidewalk sub-segments associated with each street sub-segment.
6. Sum the lengths of all sidewalk sub-segments within each street segment to obtain the total sidewalk length (SDW) for that street segment.
7. Compare the calculated sidewalk length (SDW) with a tolerance (**T**) to the length of street segment (STR).
8. Select the final sidewalk coding based on comparing the lengths of SDW_T with that of STR.
    a. Code as full coverage (1) if SDW + T ≥ 0.9 * STR;
    b. Code as partial coverage (2) if SDW ≥ 0.1* STR and SDW + T < 0.9 * STR; and
    c. Code as no coverage (0) for all other conditions.

The last step of the algorithm was to automatically assign a value of no sidewalk to every street segment classified as a freeway. This was equal to the entry of "F" in the KCC_FCC_ID field of the King County Street Network.

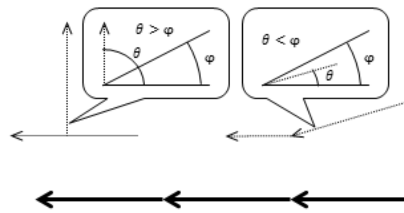
The coding algorithm presented here relies on GIS polyline files depicting sidewalks on both sides of the street. GIS data with double lines could be directly used in the algorithm. GIS data with single lines had to be offset to left and right of street segments to be effectively converted into double lines. The attributes representing each side of the street were offset from the other side by 20 feet using the "Copy Parallel" tool in ArcMap 10. Algorithms using ArcGIS can be made available upon request.

Step 1: Select a street segment and adjacent sidewalk polylines.	
Step 2: Split street and sidewalk segments at every vertex and at 10ft intervals.	

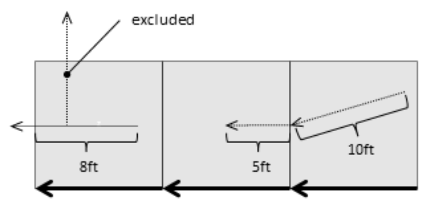
Step 3: Create buffer of length **B** along each street sub-segment (buffer shown on only one side of the street sub-segment).



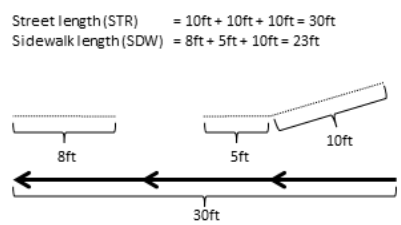
Step 4: Determine whether the sidewalk sub-segments in the buffer are associated with the street sub-segment by comparing the angle of the sidewalk sub-segment (**θ**) to the angle of the street sub-segment, and using maximum angle of **ϕ** as a cut off.



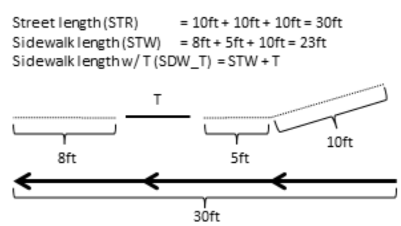
Step 5: Within each buffer, calculate the lengths of the sidewalk sub-segments that are associated with each street sub-segment.



Step 6: Sum the lengths of the sidewalk sub-segments to calculate the total length of the sidewalk (SDW). Similarly, sum the lengths of the street sub-segments to calculate the total length of the street segment (STR).



Step 7: Add Tolerance (**T**) to SDW and then compare it (SDW_T) to STR to determine the percentage of the street segment covered by a sidewalk.



Step 8: If SDW_T is ≥ 90% of STR, the street segment is fully covered by a sidewalk. If SDW_T is between ≥ 10% < 90% and SDW ≥ 10% of STR, the street segment is partially covered by a sidewalk. Otherwise, the street segment is defined as having no sidewalk.

Figure 7.1: Steps in the Automatic Coding Process

Section 8: Validating the Automatic Coding Method

The algorithm used three parameters that needed to be validated:

- Buffer (**B**)—the maximum possible distance between a street and a sidewalk segment;
- Angle (**ϕ**)—the maximum possible difference between the angle of the sidewalk segment and that of the street segment; and
- Tolerance (**T**)—an extra bit of length added to the estimated sidewalk segment length in order to account for differences in street and sidewalk geometries.

The algorithm was tested and validated using the following values for the above parameters:

B = {40 ft, 50 ft, 60 ft}
ϕ= {10 degrees, 15 degrees, 20 degrees, 25 degrees}
T = {50 ft, 60 ft, 70 ft, 80 ft, 90 ft, 100 ft}

The accuracy of results obtained from different combinations of parameters in the algorithm was tested and validated on a sample of about 5 percent (4,039) of the 79,928 street segments which could be automatically coded. The sample was selected using a random number generator that assigned all potential street segments a value ranging from 1 to 100. Segments assigned values one through five were then chosen for the sample. These 4,039 segments were coded manually to serve as the Gold Standard for comparing the sidewalk data generated from the automated process. Unlike the manual coding process described previously, the streets in the Gold Standard were manually coded using only the sidewalk data supplied by the jurisdictions—the same data used in the automatic coding process.

Every possible combination (72 in all) of the aforementioned parameter values was used to automatically code the sample and the results were compared to the Gold Standard. Success was based on the percentage of segments in which the Gold Standard and the combination of parameters used in the automatic coding algorithm were in agreement. Table 8.1 shows the results of agreements between five parameter combinations that best matched the Gold Standard.

Table 8.1: Top Five Parameter Combinations

	Parameter			**Percent Agreement Left Side**				**Percent Agreement Right Side**			
	B	**ϕ**	**T**	No Sidewalk	Full Sidewalk	Partial	**All Sidewalks**	No Sidewalk	Full Sidewalk	Partial	**All Sidewalks**
1	50	15	80	0.94	0.97	0.69	**0.95**	0.95	0.98	0.63	**0.95**
2	40	10	100	0.95	0.96	0.63	**0.95**	0.97	0.98	0.52	**0.96**
3	50	15	60	0.94	0.95	0.72	**0.94**	0.95	0.97	0.70	**0.95**
4	40	10	60	0.95	0.93	0.70	**0.94**	0.97	0.95	0.66	**0.95**
5	40	10	70	0.95	0.95	0.68	**0.94**	0.97	0.96	0.61	**0.95**

The algorithm was executed using a combination of PostGIS and R. Algorithms using PostGIS and R can be made available upon request.

The top five parameter combinations yielded a low of 0.52 to a high of 0.98 percent agreement between the algorithm and Gold Standard. Full and no sidewalk coverages showed high levels of agreement (> 0.93), while partial coverages ranged from 0.52 to 0.72. The third combination—which used a 50-foot buffer size, had sidewalk angles no greater than 15 degrees, and a tolerance of 60 feet—was selected because it yielded high agreements in the full and no sidewalk coverage categories (O.94 and 0.97) and had the highest rates of agreement for partial sidewalk coverage category (0.72 and 0.70). The results of combination three as compared to the Gold Standard are shown in Table 8.2.

Table 8.2: The Selected Parameter Combination Compared to the Gold Standard

		Gold Standard					
		Left Side of Street			Right Side of Street		
		No Sidewalk	Full Sidewalk	Partial	No Sidewalk	Full Sidewalk	Partial
Algorithm	No Sidewalk	2023	42	14	2046	34	14
	Full Sidewalk	101	1688	20	63	1710	21
	Partial	24	39	88	27	39	85
	Total	2148	1769	122	2136	1783	120
	Percent Agreement	**0.94**	**0.95**	**0.71**	**0.96**	**0.96**	**0.70**

There were 438 disagreements in the results between the automatic and manual processes. On the left sides of segments, there were 240 disagreements and 198 on the right (71 had disagreements on both sides). Overall, the total number of segments with disagreements was 367 (9.1% of the segments tested). All disagreements were reviewed by a GIS technician who examined each segment with a disagreement. Using his best judgment to determine which coding method best reflected reality, he compared the coded data to the sidewalk data to determine which coding method best reflected reality for each segment. The technician then took note of the possible primary reason a mistake was made in the coding process. These reasons were then categorized into three main types of errors: errors caused by the data; errors in the Gold Standard caused by the manual coder; and errors caused by the automatic coding process.

Four sources of data errors were identified: non-streets; ambiguity; disagreement between the street network and the sidewalk network data sets; and short segments at intersections. As part of the manual coding process all segments that did not depict streets (such as trails, railroads, or the Bus Tunnel) were coded as having no sidewalks. The automated coding process made no such distinction and treated all segments in the King County Street Network as streets and assigned sidewalks to many non-streets (Figure 8.2, panel A).

Ambiguity occurred when the GIS technician determined that the conflicting results generated by the automatic and manual coding processes both had equally plausible interpretations. In many of these cases the ambiguity was caused by a confusing street layout. For example, intersections oftentimes had sidewalks that could be associated with any number of street segments (Figure 8.3, panel B).

Culs-de-sac were very difficult to code for both the manual coder and the automated coding process. Street segments representing culs-de-sac varied greatly (Figure 8.1) and the manual and automatic coding processes differed in how they coded the sidewalks (Figure 8.2, panel C).

Data errors also occurred when the sidewalk and street network data were in disagreement (Figure 8.2, panel D). Sometimes this happened when the sidewalk segment overlapped or crossed paths with the street segment. Other times the street and sidewalk networks would be misaligned so that a street segment was adjacent to a pair of sidewalk segments rather than in between the segments. Such misalignments were determined using aerial imagery. In other cases, either the street or sidewalk network reflected a layout that did not match the imagery provided by GoogleMaps, GoogleStreetView or Bing Maps.

Table 8.3: Coding Errors in the Validation Process

	Count	Percent of Error	Percent of Gold Standard
Data Errors			
Ambiguous	4	1.09%	0.10%
Not a Street	7	1.91%	0.17%
Cul-de-Sac	78	21.25%	1.93%
Data Disagreement	39	10.63%	0.97%
Short Segment	39	10.63%	0.97%
Automatic Coder Errors			
Partial Coverage Error	53	14.44%	1.31%
Wrong Sidewalk	38	10.35%	0.94%
Cause Uncertain	13	3.54%	0.32%
Manual Coder Errors			
Simple Mistake	42	11.44%	1.04%
Partial Coverage Error	10	2.72%	0.25%
Median Errors	43	11.72%	1.06%
Total Errors	367	100.00%	9.09%
Total Number of Segments	4039		100.00%

Both the automatic and the manual coding processes were inconsistent in how short segments in intersections were coded. The King County Street Network has many short street segments (usually under 25 feet in length). Due to their size and location, they would frequently go unseen by the coders. To save time during the manual coding process, many of these segments were assigned a value of no sidewalk to both sides regardless of whether sidewalks were actually present in the segments behind or in front of the segment in the intersection. However if a short segment was spotted, the manual coder would the code segment based on the presence or absence of parallel sidewalks on either side of the intersection (Figure 8.2, panel E). The automatic process was also inconsistent; sometimes segments in intersections would be coded as having sidewalks if there were sidewalks on the joining street segments that were not in the intersection. Other times, the automatic process coded segments as having no sidewalks regardless of the segments in front or behind it.

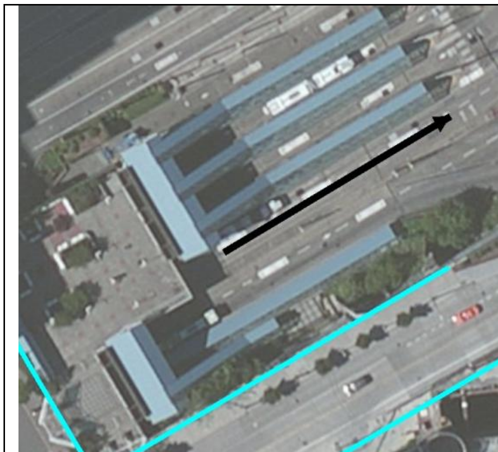
Three sources of error were identified in the automatic coding process: assignment of partial coverage; assignment of wrong sidewalk; and source unclear. Determining which streets had full or partial sidewalk coverage was tricky for both manual coder and the automated process. The automatic process would frequently code a street as having partial coverage if the front or end of the segment extended into the intersection (Figure 8.3, panel A). Another common source of error was for a segment to be coded based on a sidewalk not associated with that segment (Figure 8.3, panel B). In a small number of cases, the GIS technician reviewing the coding processes was unable to determine why the automatic process coded the sidewalk data incorrectly (Figure 8.3, panel C).

Figure 8.1: Types of Cul-de-Sac Street Segments. Culs-de-sac were commonly represented by one (left), two (middle), or three (right) street segments.

Manual coder errors were the result of human error. Three types of manual errors were identified: simple mistakes; assignment of partial coverage; and incorrect coding of streets comprising two parallel segments. Simple mistakes were many times the result of a typo when the coder pressed the wrong key on the keyboard while entering in the sidewalk code. Sometimes the coder would confuse the direction of the street and enter in sidewalk data for the left side of the street on the right side or vice versa (Figure 8.4, panel A). Other times the coder simply made an incorrect judgment about sidewalk coding.

As with the automated process, determining partial coverage was a source of errors for the manual coders. The GIS technician found a few examples in which the automatic coding process did a better job of coding partial coverage than the manual coder (Figure 8.4, panel B).

The last set of manual coder errors covers a very specific misunderstanding regarding the coder's incorrect assumption about how to code pairs of parallel segments (which are many times separated by a median). The coder assumed that if a street comprised two parallel segments, the lack of a sidewalk in between the two segments was what needed to be coded—rather than the sidewalk on the opposite side of the street.



(A): Not a street

	L	R
Auto	0	1
Gold Standard	0	0

A few segments that are not streets were included in the Gold Standard. As part of the manual coding process all non-streets were assumed to have no sidewalks. The automatic coding made no distinction. In the picture a line segment at the entrance to the Bus Tunnel was assigned a sidewalk incorrectly.

ID 89623

(B): Ambiguous

	L	R
Auto	2	1
Gold Standard	1	1

This segment merges into another street with a sidewalk. The source of ambiguity is whether the other street's sidewalk is also the sidewalk for the segment of interest.

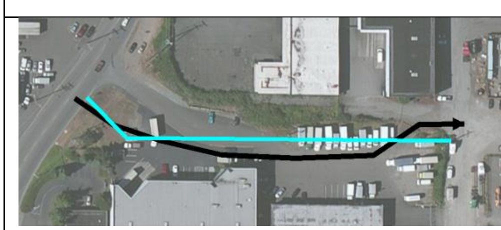
ID 197753

(C): Culs-de-Sac

	L	R
Auto	0	2
Gold Standard	0	1

There were a number of ways in which sidewalks around culs-de-sac could be coded. Street segments representing culs-de-sac could also be represented in a number of different ways (Figure 8.1).

ID 12401



(D): Data Disagreement

	L	R
Auto	2	0
Gold Standard	2	2

Sometimes the data sources were in disagreement. In the adjacent picture the sidewalk (in blue) overlaps the street segment (in black). However neither street nor sidewalk matches what is depicted in the aerial image.

ID 66474

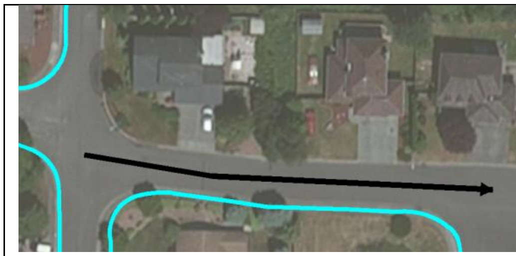
(E): Short segment in intersection

	L	R
Auto	1	1
Gold Standard	0	0

Usually located in intersections, short segments were easy to overlook in the manual coding process. To save time, the manual coder would assume there were no nearby sidewalks regardless of whether sidewalks were actually present on the adjacent streets.

ID 34787

Figure 8.2: Examples of Data Errors



(A): Automated Partial Error

	L	R
Auto	0	2
Gold Standard	0	1

The automated process and the Gold Standard had many disagreements about whether a street had partial or full sidewalk coverage.

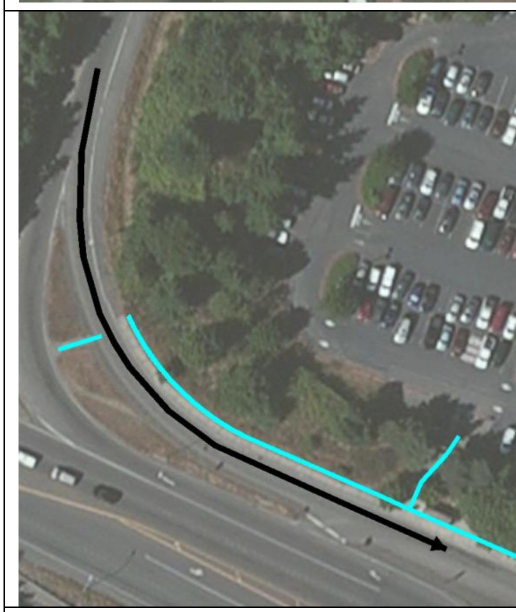
ID 129814

(B): Wrong Sidewalk

	L	R
Auto	1	0
Gold Standard	0	0

Sometimes the automated process would assign the wrong sidewalk to a street segment.

ID 131580



(C): Cause Uncertain

	L	R
Auto	0	0
Gold Standard	2	0

Sometimes the cause of a disagreement was unable to be determined.

ID 119632

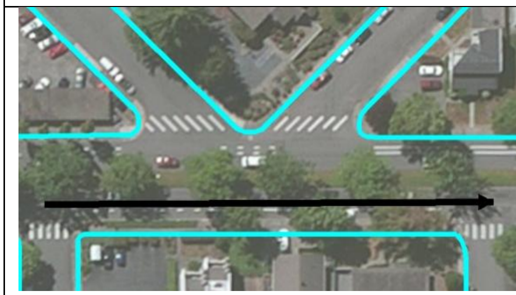
Figure 8.3: Examples of Automatic Errors

(A): Simple Mistake

	L	R
Auto	0	1
Gold Standard	1	0

Occasionally the manual coder made blatant mistakes. In this example, the coding is reversed in relation to the street direction.

ID 98657



(B): Manual Partial Error

	L	R
Auto	2	1
Gold Standard	1	1

Sometimes the manual coder's assessments of partial sidewalk coverage were incorrect.

ID 117098

(C): Parallel Segment Errors

	L	R
Top Segment		
Auto	1	1
Gold Standard	1	0
Bottom Segment		
Auto	1	1
Gold Standard	0	1

The manual coder miscoded many streets comprising parallel segments by coding for sidewalks in between the two segments rather than on the other side of the street.

ID 91401

Figure 8.4: Examples of Manual Coder Errors

APPENDIX A: METADATA

Summary
The Urban Form Lab King County Sidewalk Data set (kcn_skw_20120507.csv) is a tabular data set that contains information about the presence or absence of sidewalks for every street segment in the King County Transportation Network GIS database that lies within King County's urban growth boundary.

Data Sources
The data on sidewalks were assembled from two sources:

1) Local city governments and jurisdictions in King County that volunteered data to the Urban Form Lab; and
2) Internet mapping websites (Google Maps, Google Street View and Bing Maps).

The following GIS layers were used in the construction of this data set.

1. The King County Transportation Network, downloaded in December, 2010. Metadata available at: http://www5.kingcounty.gov/sdc/Metadata.aspx?Layer=trans_network
2. The King County Urban Growth Area shape file, downloaded in January, 2012. Metadata available at: http://www5.kingcounty.gov/sdc/Metadata.aspx?Layer=urban_growth
3. The Incorporated Areas of King County shape file, downloaded in January, 2012. Metadata available at: http://www5.kingcounty.gov/sdc/Metadata.aspx?Layer=citydst

The King County Transportation Network supplied the street segments that are linked to the UFL sidewalk data. The KC Urban Growth Area shape file was used to determine which of the KC Transportation Network street segments were within the urban growth boundary. The Incorporated Areas shape file was used to determine the location of cities within King County.

Table 1: Data Dictionary

Field	Values	Description
tlink_id	Numerical ID	Identifier linking each segment to the King County Street Network Data set
Swk_left	0, 1, 2	Sidewalk coverage on the left side of street segment
Swk_right	0, 1, 2	Sidewalk coverage on the right side of street segment
swk_date	year-month-day	Date the street segment was coded with sidewalk data by UFL Note: this is not data date.
swk_method	”auto” “manual”	Denotes whether street segment was coded manually or using an algorithm.

Data Description
The sidewalk data set consists of an attribute table with five fields that can be appended to the King County Street Network data set using the field "tlink_id". Two of the fields (swk_left and swk_right) store information about the presence or absence of a sidewalk on either the left or right side of each street segment. The following codes were used:

0 = absence of a sidewalk;
1 = presence of a sidewalk; and
2 = partial presence of a sidewalk on the segment (the sidewalk length is less than 90% of the street length).

The field "swk_date" refers to the date the data were entered into the UFL King County Sidewalk Data set. This date does not reflect the actual date the source data were created. Rather, it refers to the date the coding of that segment was completed. The field "swk_method" refers to the method used to code the data. Sidewalk data provided by city governments were automatically converted to the 0, 1, 2 coding scheming using an algorithm executed in PostGIS 2.0 and R. The word "automatic" in the "swk_method" field refers to segments that were coded in this manner.

Data Collection
Every city within in King County was contacted and asked to supply information about sidewalk coverage. Of the 40 jurisdictions, including unincorporated King County, the Urban Form Lab received data from 30.

Sidewalk Definition
Sidewalks were defined based on Fig 2-002 (page 2-26) of the 2007 King County Road Design and Construction Standards document. These guidelines state that sidewalks must be made out of concrete, elevated, separated from the roadway, and within 20 feet of the street. If source data from a city had sidewalk type identifiers (e.g., elevated sidewalks, shoulders, gravel pathways, and etc.), only the sidewalks satisfying our sidewalk definition were included in the sidewalk data.

Methods
Two methods were used to standardize the data into the format described in the Data Description: manual or automatic coding.

Manual Coding: A GIS technician selected an individual street segment and then looked at one of many possible data sources to determine whether it had a sidewalk associated with it. If jurisdiction sidewalk data were available the technician would use that source, otherwise the technician used Internet sources (GoogleMaps, GoogleStreetView and Bing Maps). In many cases, a combination of data sources was used to code the same segment.

The following jurisdictions were *manually coded*: Algona, Auburn, Beaux Arts, Black Diamond, Bothell, Burien, Carnation, Clyde Hill, Covington, Duvall, Enumclaw, Hunts Point, Issaquah, Kenmore, Lake Forest Park, Medina, Mercer Island, Milton, Newcastle, Normandy Park, North Bend, Pacific, Sammamish, Skykomish, unincorporated King County, Woodinville, and Yarrow Point.

Automatic Coding: Data sets from only 13 of the 30 jurisdictions that supplied data could be used for automatic coding. PostGIS, R, and ArcGIS served to process and automatically code these data. The first step in the automatic coding process was to split all of the street and sidewalk segments into smaller sub-segments. This helped to process curved segments. Segments were split at every vertex and at 10-foot increments. Then a buffer was created that extended from either side of every street segment. Next it was necessary to determine whether sidewalks inside the buffer were associated with the street segment. To do so, the angles of the sidewalks were compared to the angles of the streets. A maximum angle was set as a cutoff beyond which any sidewalk was considered to be unassociated with the street segment. To determine whether street segments had partial (the length of the sidewalk covered 10% to 90% of the length of street segment) or full coverage, the sub-segments were summed to get the full lengths of each street and sidewalk segment. An extra amount of length (tolerance) was added to the sidewalk length to account for curb cuts, driveways, intersections and other small breaks in the sidewalk.

Optimal buffer sizes, sidewalk cutoff angles, and sidewalk tolerances were identified by comparing 72 combinations of these parameters against a random sample of 5% of the manually coded streets. The combination that had the highest agreement with the manually coded data was a buffer size of 50 feet on either side of the street segment, a maximum sidewalk angle of 15 degrees, and a tolerance in the sidewalk length of 60 feet per street segment.

The following jurisdictions were *automatically coded*: Bellevue, Des Moines, Federal Way, Kent, Kirkland, Maple Valley, Redmond, Renton, SeaTac, Seattle, Shoreline, Snoqualmie, and Tukwila.

Of the 123,554 street segments within the King County urban growth boundary, 79,928 or about 64.7% were automatically coded. The remaining 35.3% were manually coded. Together, the jurisdictions that were automatically coded accounted for 60% of the land area within the King County urban growth boundary.

Visualizing the Sidewalk Data

Because the sidewalk data are in tabular format they do not lend themselves to ready visualization. One option is to create a new field in the attribute table that merges the swk_left and swk_right fields using a numeric code to represent all combinations of sidewalk coverage. This values in this code can then be given different colors and line styles in ArcMap's Symbology tab. Table 2 shows one possible way of recoding the data into a single field. Using this method will involve tradeoffs regarding readability of the map and amount of information depicted. For example, the recoding scheme in Table 2 does not include streets with partial sidewalks on one side and full sidewalks on the other. Nor does it indicate which sides of the street have partial coverage.

Table 2: Potential Coding Scheme for Visualizing Sidewalk Data

Coverage	Values to be Merged	Symbol
No sidewalk	[swk_left] = 0 and [swk_right] = 0	
Sidewalk on the left	[swk_left] = 1 and [swk_right] = 0	
Sidewalk on the right	[swk_left] = 0 and [swk_right] = 1	
Partial sidewalk on left or right	[swk_left] = 2 or [swk_right] = 2	
Full coverage on both sides	[swk_left] = 1 and [swk_right] = 1	

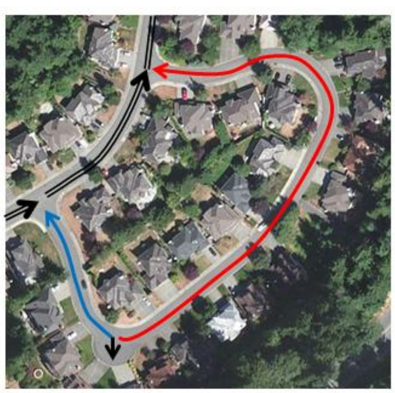


Figure 1: Examples of Sidewalk Visualization

Figure 2: Example of a Sidewalk Coverage Map