Crash Analysis Toolbox

ArcGIS Python Tool Suite

User Manual v 2.2

Table of Contents

1. Introduction

This is the user manual for the custom tools in the “Crash Analysis Toolbox” that development team, Bus Drivers, has developed for use by Dr. Ghazan Khan. This manual depicts the most up to date state of the ArcGIS Toolbox called “Crash Analysis Toolbox”.

The primary user of the Hotspot Analysis Tool Suite is Dr. Ghazan Khan, an assistant professor of civil engineering, specializing in transportation engineering, at California State University, Sacramento. Dr. Khan has expert-level knowledge of the ArcGIS platform, and he is a subject matter expert in the crash analysis field.

Dr. Khan lectures about ArcGIS and road safety, and his civil engineering students may utilize the developed tools in their research (particularly graduate students). All users of the Hotspot Analysis Tool Suite are expected to be proficient in ArcGIS.

The Crash Analysis toolbox has 7 tools: *Crash Radius Density, Crash Network Density, Create Random Points on a Network, Cross K Function, Global K Function, Network Dataset Length, Random ODCM Permutations* . Each one of these tools automates various processes in Dr. Khan’s analysis processes in his Crash Data Analysis in answering the following questions:

- “Where do crashes occur or cluster?”
- “Is there a systematic trend to the occurrence of crashes or are crashes just randomly distributed?”

- “If there is a systematic trend, what factors could be contributing to the trend?“

The motivation for Dr. Khan’s analysis was the need for location-specific results as compared to geographically homogeneous results without the consideration of location or the proximity of data points was critical to identifying hotspots with regards to road safety. The detection of spatial patterns in spatial data would identify clusters or hotspots of crashes.

These tools offer a more streamlined approach to analyzing traffic crash data, compared to ArcGIS’ model builder. The tools are also modular, meaning that python tools can be removed or added to each project without much effort.

This manual includes:

- A description of the ArcGIS version needed to use of the Crash Analysis Toolbox and how to install the Crash Analysis Toolbox(*Section 2*).

- A brief tutorial of how to import a set of tabular crash data into ArcGIS (*Section 3*). This is a required step before using any one of the Crash Analysis Tools.

- descriptions of the 3 major tools, inputs and outputs (Section 4).

- Known Bugs which the Bus Drivers development team encountered in developing the Crash Analysis Tool box (Section 5).

Note: this manual is written with the expectation that the users of the Crash Analysis Toolbox have at least a basic technical understanding in the following topics in ArcGIS:

- What is ArcGIS - Using ArcMap and ArcCatalog
- What are Shapefiles, Feature Classes, and Geodatabases as well as when and how to use them
- What toolboxes are in ArcCatalog

There is significant amount of documentation on Esri’s website on these topics as well as many tutorials on the web.

2. ArcGIS Environment

a. ArcGIS Version Used

The Crash Analysis toolbox are intended for use on ArcGIS 10.3 and above. ArcGIS 10.1 has numerous issues with these tools. Obtain a license for ArcGIS for ArcGIS 10.2 or above, then install the correspond ArcGIS for Desktop, install the additional packages required, then import the Crash Analysis Toolbox into your ArcGIS project.

b. Additional Packages Required

The ArcGIS Editor for OSM (OpenStreetMap) is an ArcGIS extension required for the *Crash Network Density* and *Network K Analysis* Tools, specifically to be able use open-source network data of streets within the toolbox.

For installation, follow the link below (urls subject to change):
The version to download and install corresponds to the version of your installed ArcGIS application.

10.2.x - http://www.arcgis.com/home/item.html?id=16970017f81349548d0a9eead0ebba39
10.3.x - http://www.arcgis.com/home/item.html?id=75716d933f1c40a784243198e0dc11a1

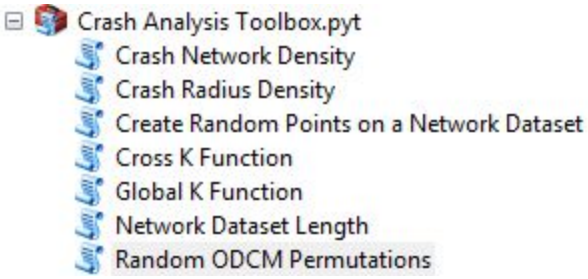
Python 2.7 is also required for using the crash analysis toolbox.

The files containing the Crash Analysis Toolbox are required to use the tools.

c. Setting up the Crash Analysis Toolbox for Use, Importing a Toolbox

In order to use the Crash Analysis Toolbox in a project, we need to import the toolbox like we would import any other custom toolbox in ArcGIS

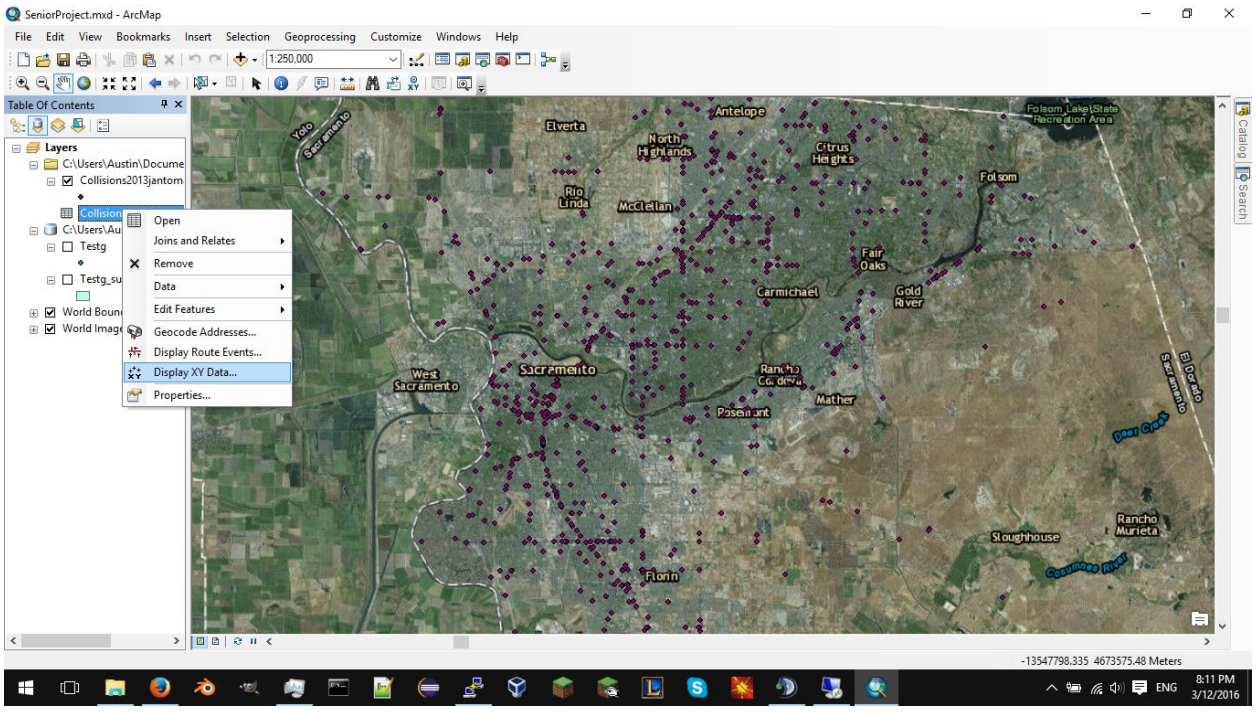
1. Find a location where the Crash Analysis Toolbox is saved
2. In ArcMap, go to the Top right and click the Catalog tab.
3. Right click Folder Connections and select Connect to Folder.
4. In the submenu, go to where you saved the Crash Analysis Toolbox and select that folder.
5. Click the + to the left of the Folder Connections icon: This should show the folder you have connected, containing the python tool suite files.

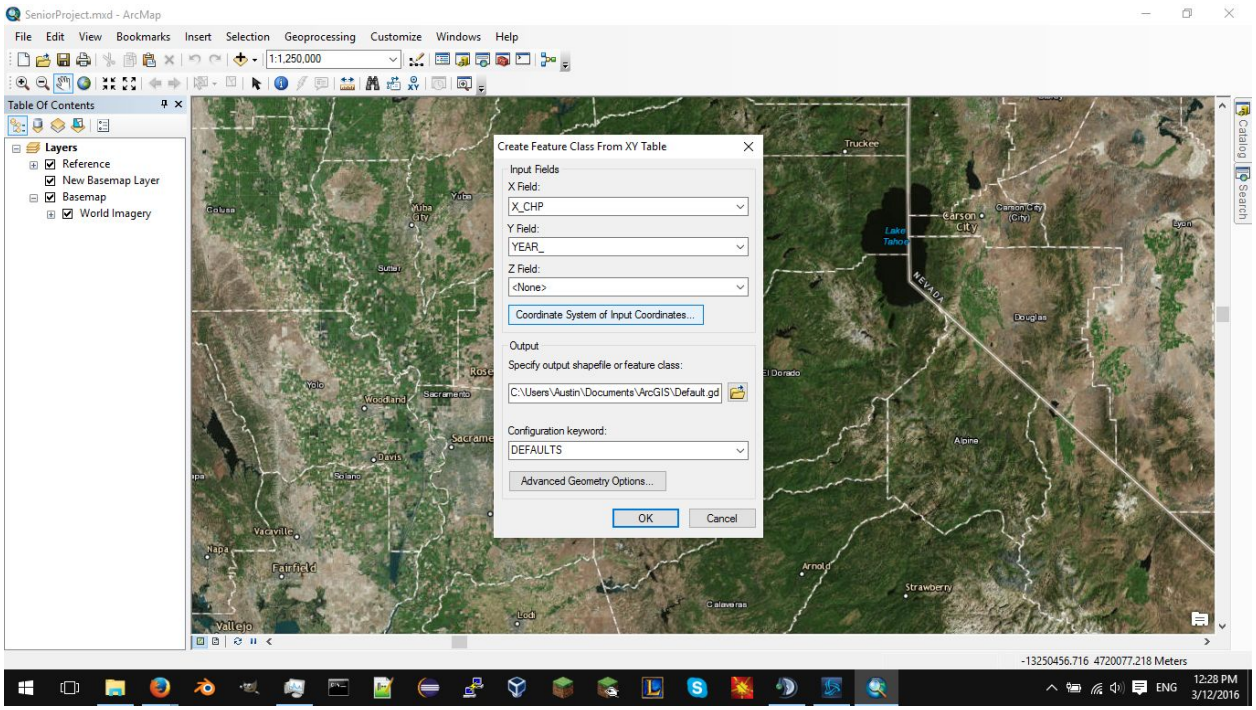


3. Importing Crash Data into ArcGIS

1. Get the crash data in a shapefile (.csv file). TIMS is a great resource for traffic crash .csv data.
2. Once you have a .csv file for crash data, use ArcCatalog (it is located on the right side of the ArcMap window) to connect to the folder containing the crash data file.
3. Drag the .csv file to the Table of Contents, to the left.
4. Right click on
5. Select your X and Y values (the columns on the shapefile that correlate to x positions and y positions of the traffic crash, respectively). Also, Specify the coordinate system you are using by pressing the 'edit' button and finding your specific Coordinate system(for example, WGS 1984 is located in the 'Geographic Coordinate System-> World' Folder).
6.

Alternatively, you can follow the instructions in this video (video url subject to change): https://www.youtube.com/watch?v=XkixWXu5jJE

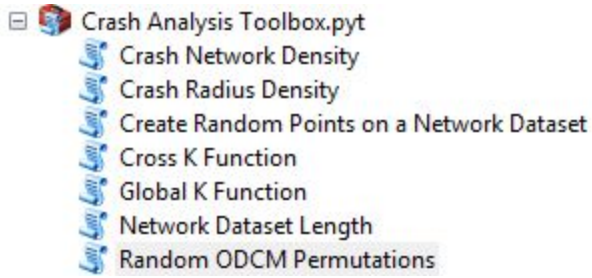


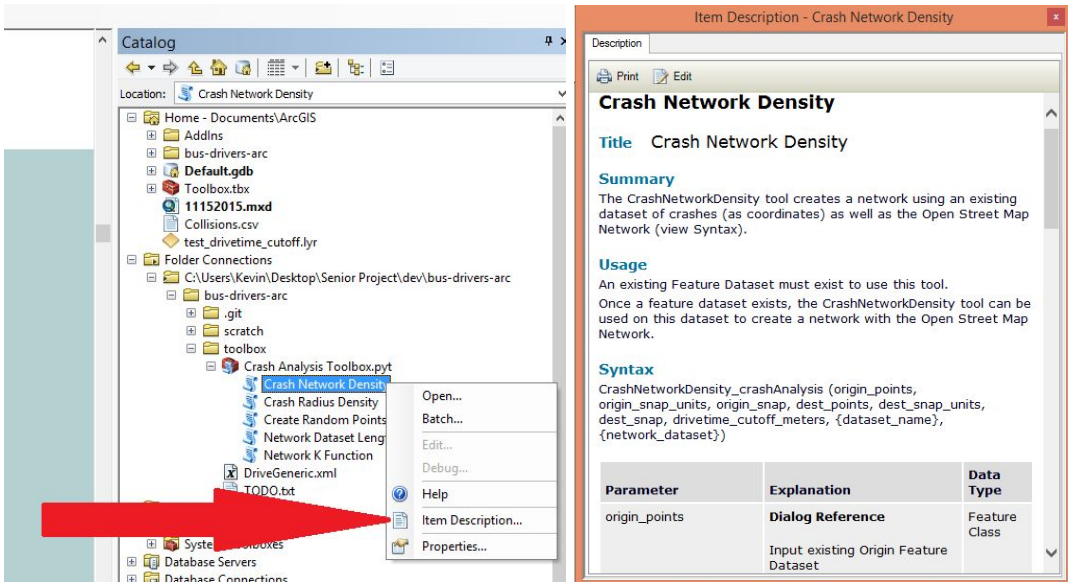


4. Tools

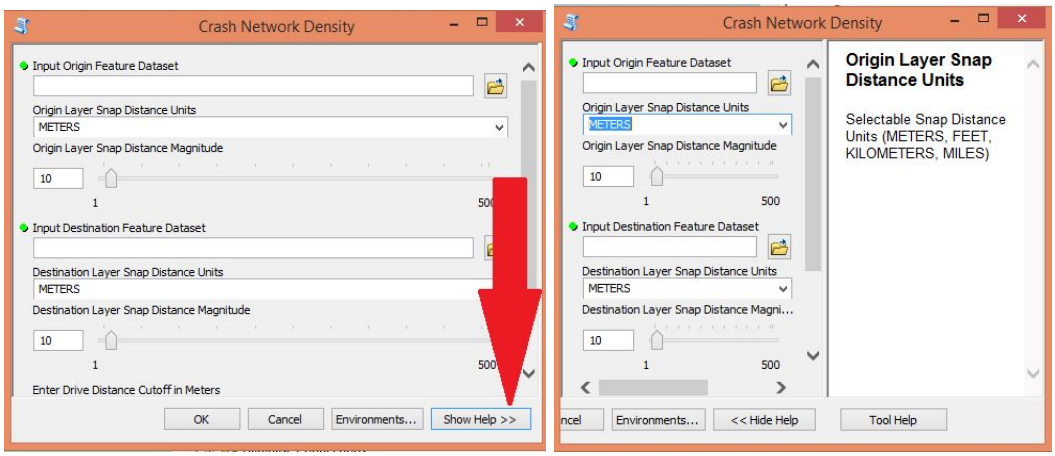
The Crash Analysis toolbox has 3 major tools: *Crash Radius Density, Crash Network Density,* and *Network K Analysis*. Each one of these tools acts takes an existing crash data and takes in various inputs (defined by each tool) and outputs analysis by each type of tool.

Each of the tools support have documentation built into them. To see them, go to the ArcCatalog, then right click on a python tool and select “Item Description”. A new window will open, describing the tool.





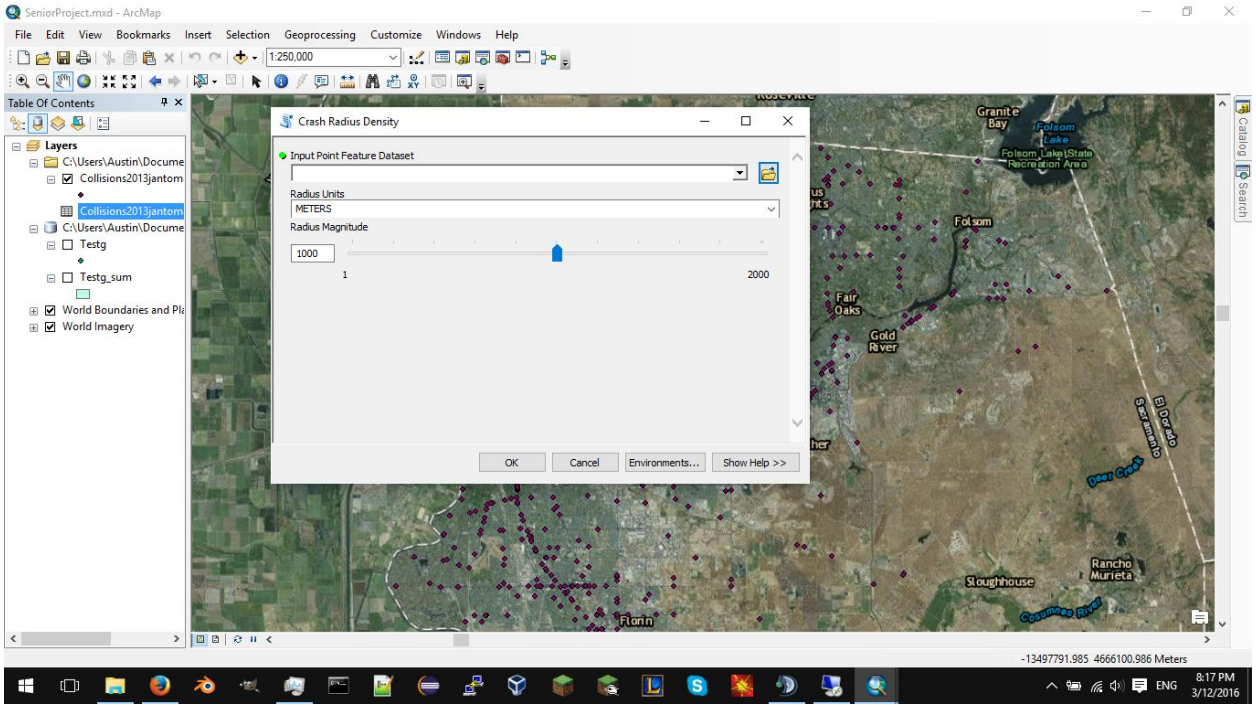
Also, when opening each tool for use, there is supporting documentation in the “Show Help” button for each of the tools and the tool’s parameters.



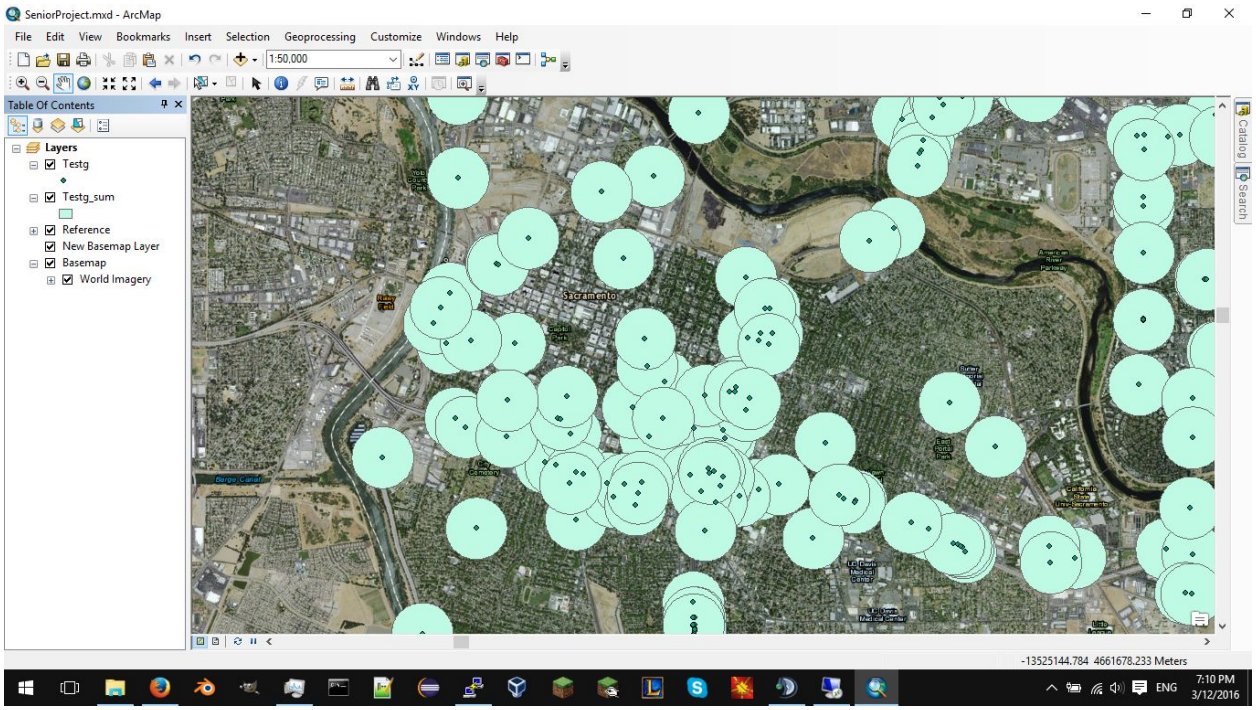
a. Crash Radial Density Tool

This tool generates a circle around every crash point, and determines the number of collisions within that circle.

Refer to Crash Radial Density tool Item Description for how to use the tool



Outputs: Graphical display, showing a polygon circle. The center of the circles are where vehicle crashes occurred.



b. Crash Network Density Tool

Finds the distance between origins and destinations using a network dataset. The network dataset can optionally be generated automatically.
Refer to Crash Network Density tool Item Description for how to use the tool

c. Create Random Points on Network Dataset Tool

In Dr. Khan's analysis, he needed to solve the problem of generating a confidence interval for randomly occurring crashes on a network. In order to do so, he needed to be able to generate randomly placed points on a network. This is what this tool does.

Refer to the Item Description for how to use the tool

d. Random ODCM Permutations

This tool generates Origin-Destination Cost Matrix.
An OD cost matrix finds and measures the paths on a network from multiple origins to multiple destinations. You can specify the number of destinations to find and a maximum distance to search. The currently supported Analysis Types include:

- Global Analysis - Source and Destination points are the same and is used for clustering for crashing.
- Cross K-Function - Used when analyzing 2 sets crash point sets.

The Cross K Function and Global K Function both utilize this tool.

e. Cross K Function / Global K Function

The Cross K Function and Global K Function tools that allows the user to find crash clustering and dispersion at different distance bands using Dr. Khan's modified version of the Ripley's *K* Function. Dr. Khan's version of the algorithm is known as a Network *K* Function. For a given set of crash points, this function determines deviations from spatial homogeneity at different distances along a road network. The "global" type calculated the global statistic to identify the presence or absence of a systematic process, the magnitude, and the extent of spatial patterns in crash data. The second type of analysis, Cross-K analysis, is comprised of local statistics at individual crash locations to identify individual crashes as part of a statistically significant clustered (hotspot), dispersed (coldspot), or random point pattern. The methods under the crash cluster analysis category were developed based on the principles of the K-Function method in network space. Outputs a table of raw data which is used to generate the Network K Function.

Refer to Network K Function tool Item Description for how to use the tool

5. Known Bugs

As with all software, These tools have their faults. These are a list of known bugs found. Note that there may be bugs that are not documented in this section.

ArcGIS 10.1: In general, many of these tools have issues when run under ArcGIS 10.1. 10.1 itself has many bugs, such as deleting a layer immediately after making the layer. As a result, we ask that developers and maintainers run ArcGIS 10.2 and above when using our tools. For example there is a bug in 10.1 where if you make a layer, that layer is immediately deleted after.

6. Frequently Asked Questions (FAQ)

How do I update ArcGIS?

Once you have an EVA number and license key, follow the instructions on this website: http://www.esri.com/landing-pages/software/arcgis/arcgis-desktop-student-trial

Why does ArcGIS crash when I try to make a new python tool, or connect to a folder containing python tools?

> One reason may be that python is not installed, or installed in a different directory than where ArcGIS is looking for. Make sure you have python installed in the proper directory.

Why do my points show up in a different spot than expected? For example, if I have Sacramento traffic crash data, it is instead showing in an ocean close to Africa?

> This is because you have incorrectly set up your WGS. Refer to the 'Importing Crash Data in ArcGIS' Section. For Sacramento data, use WGS 1984.

7. References

a. Dr. Ghazan Khan Dissertation "Integration of GIS and Spatial Statistics – A New Paradigm in Crash Data Analysis"

Dr. Ghazan Khan's Dissertation provides the background needed to understand the Network K Function. Request a copy of this dissertation from Dr. Khan and consult with him to gain a better understanding of the Network K Function.

8. Glossary

GIS: Geographical Information System. A system designed to store, visualize, and manipulate geographical information such as roads, buildings, waterways, and latitude-longitude points.

ArcGIS: The GIS platform of choice used in the Hotspot Analysis Tool Suite project. The tool suite described herein will run under this platform.

Network Dataset: Network datasets are databases consisting of edges (lines), junctions (edge connections), and turns (edge directions), and can be used in GISs to model roadways. The Hotspot Analysis Tool Suite will use network datasets to determine the driving distance and driving time between crash incidents.

Toolbox: A suite of ArcGIS tools with similar functionality bundled together.

Network Analyst: A toolbox available in ArcGIS that provide functionality related to network datasets. The functionality includes driving directions, driving time, and driving distance between origins and destinations.

OpenStreetMap (OSM): An openly-licenced map with functionality similar to Google Maps. OpenStreetMap provides a series of APIs that developers can use to programmatically download street data; the Network Crash Density tool downloads street data as needed and converts it into a network dataset suitable for use in ArcGIS.

Origin-Destination Cost Matrix: The Origin-Destination Cost Matrix tool is a tool provided by the ArcGIS Network Analyst toolbox that allows a user to find the least-cost distance and driving time between a series of origin points, and a series of destination points.

Ripley's K-Function: A statistical algorithm for determining if a series of geographical points are distributed in a homogeneous manner. That is, if points are randomly distributed about a geographical area in a uniform manner, the points are said to be homogeneous; otherwise, points may be clustered or dispersed.

Network K-Function: A statistical algorithm used for testing the hypothesis that points are uniformly and independently distributed over a network. Developed by Okabe and Yamada, the network K-function addresses the problems caused by the Ripley's K-function assumptions of a continuous plane with Euclidean distances.

Geodatabase (GDB): A database, similar to a relational database, that efficiently stores geographical data such as points, polygons, and lines. The database consists of one or more tables which can be joined spatially. For instance, two polygons are joinable if they overlap.

Feature Class: Stored in a geodatabase.

9. Additional Resources:

http://pro.arcgis.com/en/pro-app/arcpy/geoprocessing_and_python/defining-parameter-data-types-in-a-python-toolbox.htm

This website will show a list of available data types that ArcGIS for python can use.

https://www.arcgis.com/home/item.html?id=75716d933f1c40a784243198e0dc11a1

This is the website that contains a download for OpenStreetMaps.

http://tims.berkeley.edu/

This is the website where you can get your shapefiles (.csv files) to convert into geodatabases.