

Measuring Geographic Distributions



This guide was created by the staff of the GIS/Data Center at Rice University and is to be used for individual educational purposes only.

The steps outlined in this guide require access to ArcGIS Pro software and data that is available both online and at Fondren Library.

- Obtaining the Tutorial Data
 - OPTION 1: Accessing tutorial data from Fondren Library using the gistrain profile
 - OPTION 2: Accessing tutorial data online using a personal computer
- Measuring Geographic Distributions in ArcGIS Pro
 - Opening an Existing Project
 - Opening the Geoprocessing Pane
 - Central Feature
 - 1790
 - 1870
 - 2010
 - Mean Center
 - 1790
 - Standard Distance
 - 1790
 - 2010
 - Directional Distribution (Standard Deviation Ellipse)
 - 1790-1SD
 - 1790-2SD
 - 2010-1SD
 - 2010-2SD
 - Linear Directional Mean
 - Roads
 - Freeways



The following text styles are used throughout the guide:

Explanatory text appears in a regular font.

1. Instruction text is numbered.
2. Required actions are underlined.
3. **Objects of the actions are in bold.**

Folder and file names are in italics.

Names of Programs, Windows, Panes, Views, or Buttons are Capitalized.

'Names of windows or entry fields are in single quotation marks.'

"Text to be typed appears in double quotation marks."



The following step-by-step instructions and screenshots are based on the Windows 7 operating system with the Windows Classic desktop theme and ArcGIS Pro 2.1.3 software. If your personal system configuration varies, you may experience minor differences from the instructions and screenshots.

Obtaining the Tutorial Data

Before beginning the tutorial, you will copy all of the required tutorial data onto your *Desktop*. Option 1 is best if you are completing this tutorial in one of our short courses or from the GIS/Data Center and Option 2 is best if you are completing the tutorial from your own computer.

OPTION 1: Accessing tutorial data from Fondren Library using the *gistrain* profile

If you are completing this tutorial from a public computer in Fondren Library and are logged on using the *gistrain* profile, follow the instructions below:

1. On the *Desktop*, double-click the **Computer** icon > **GISData (\\smb.rdf.rice.edu\\research\\FondrenGDC) (O:)** > **GDCTraining** > **1_Short_Courses** > **Measuring_Geographic_Distributions**.
2. To create a personal copy of the tutorial data, drag the **GeographicDistribution** folder onto the *Desktop*.
3. Close all windows.

OPTION 2: Accessing tutorial data online using a personal computer

If you are completing this tutorial from a personal computer, you will need to download the tutorial data online by following the instructions below:



Tutorial Data Download

- [GeographicDistribution.zip](#)

1. Click **GeographicDistribution.zip** above to download the tutorial data.
2. Open the **Downloads** folder.
3. Right-click **GeographicDistribution.zip** and select **Extract All...**
4. In the 'Extract Compressed (Zipped) Folders' window, accept the default location into the **Downloads** folder and click **Extract**.
5. Drag the unzipped **XY** folder onto your **Desktop**.
6. Close all windows.

Measuring Geographic Distributions in ArcGIS Pro

Opening an Existing Project

1. On the **Desktop**, double-click the **GeographicDistribution** folder.
2. Double-click the **GeographicDistribution.aprx** project file to open the project in ArcGIS Pro.
3. In the Catalog pane on the right, expand the **Databases** folder.
4. Expand the **GeographicDistribution.gdb** geodatabase.
5. Right-click the **USStatesPopTimeline** feature class and select **Add to New Map**.

You may wish to zoom into the continental United States.

1. In the Contents pane to the left, right-click the **USStatesPopTimeline** layer name and select **Attribute Table**.

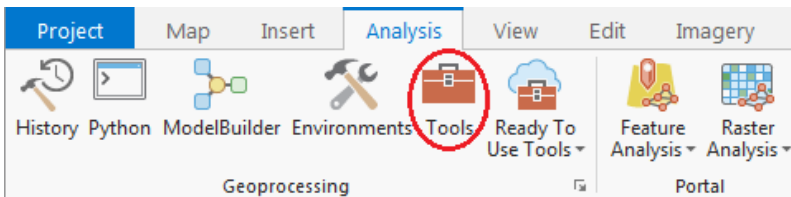
Notice that the attribute table contains population data every decade between 1790 and 2010 for each of the 50 states and the District of Columbia. These population counts only include people counted in the official US Census.

- a. Close the **USStatesPopTimeline** table view.

You will now investigate how the distribution of the population of the United States has changed over time using geographic distribution tools, which calculate the spatial equivalent of summary statistics.

Opening the Geoprocessing Pane

1. On the ribbon, click the **Analysis** tab.
2. On the Analysis tab, within the first Geoprocessing group, click the **Tools** button.



Notice that the Geoprocessing pane has opened on the right as a new tab on top of the Catalog pane. Typically, you would use the 'Find Tools' search box at the top of the Geoprocessing pane to search for the name of the tool you'd like to use, but, at times, especially when learning the software, it can be helpful to view the full hierarchy of all the tools available, because you will often discover related and helpful tools that you didn't know existed and wouldn't know to search for. You might also completely forget the name of a tool, but be able to locate it based on the hierarchy. For these reasons, we will be manually navigating the toolboxes throughout this tutorial. The more typical workflow of searching directly for a specific tool is covered at the end of the [Introduction to Geoprocessing](#) tutorial.

1. At the top of the Geoprocessing pane, click the **Toolboxes** tab.

Central Feature

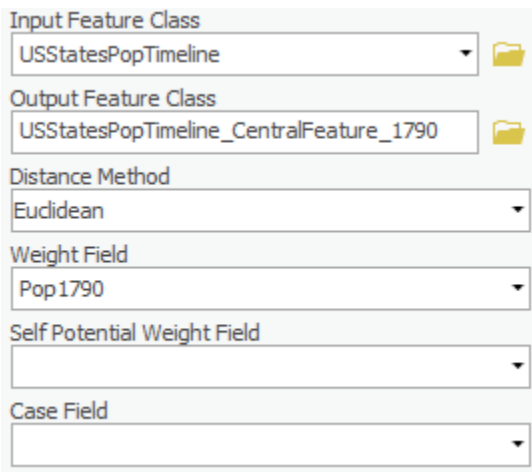
The first tool you will learn is the Central Feature tool, which will identify the most centrally located feature in a point, line, or polygon feature class. You will run the tool three times, weighting the central feature by the population in 1790, 1870, and 2010 to see how the country expanded to the west.

1790

You will start by determining which state is most centrally located out of all the states admitted as of 1790 when weighted by the 1790 population.

1. In the Geoprocessing pane, click the **Spatial Statistics Tools** toolbox > **Measuring Geographic Distributions** toolset > **Central Feature** tool.

2. For 'Input Feature Class', use the drop-down menu, select the ***USStatesPopTimeline*** layer.
3. For 'Output Feature Class', rename the feature class "***USStatesPopTimeline_CentralFeature_1790***".
4. For 'Weight Field,' use the drop-down menu to select the ***Pop1790*** field.
5. Ensure your Central Feature tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Feature Class
USStatesPopTimeline_CentralFeature_1790

Distance Method
Euclidean

Weight Field
Pop1790

Self Potential Weight Field

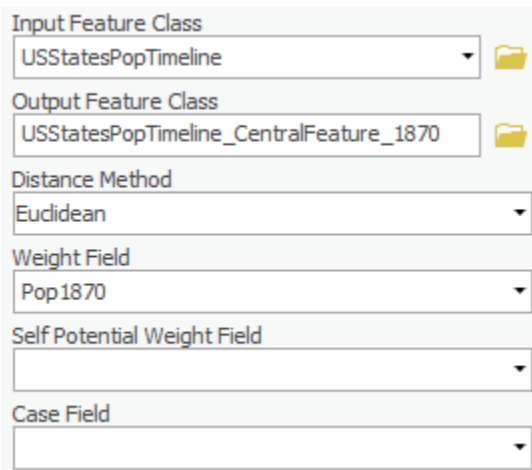
Case Field

Notice that the output of this tool is the polygon state of Maryland, which was one of the existing features from the original *USStatesPopTimeline* layer. This state selection makes intuitive sense, since Maryland is located near the center of the original US colonies and is adjacent to the nation's capital.

1870

As the Geoprocessing pane doesn't reset after a tool has finished running, it is easy to rerun tools with slightly modified settings. In future versions of ArcGIS Pro, batch processing is also supported, which facilitates multiple runs of the same tool within a single interface.

1. For 'Output Feature Class', rename the feature class "***USStatesPopTimeline_CentralFeature_1870***".
2. For 'Weight Field,' use the drop-down menu to select the ***Pop1870*** field.
3. Ensure your Central Feature tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Feature Class
USStatesPopTimeline_CentralFeature_1870

Distance Method
Euclidean

Weight Field
Pop1870

Self Potential Weight Field

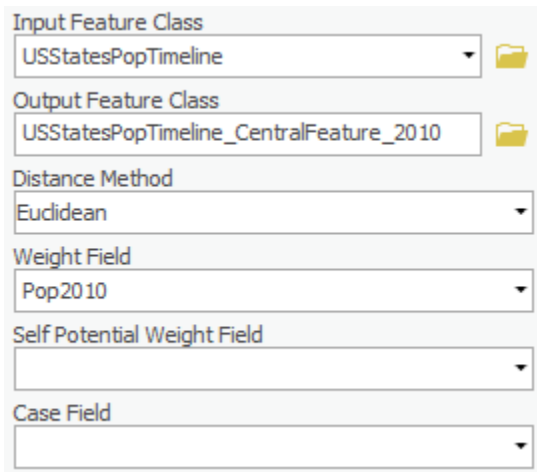
Case Field

The resulting feature is Ohio, which is west of Maryland, as we would expect.

2010

1. For 'Output Feature Class', rename the feature class "***USStatesPopTimeline_CentralFeature_2010***".
2. For 'Weight Field,' use the drop-down menu to select the ***Pop2010*** field.

3. Ensure your Central Feature tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Feature Class
USStatesPopTimeline_CentralFeature_2010

Distance Method
Euclidean

Weight Field
Pop2010

Self Potential Weight Field

Case Field

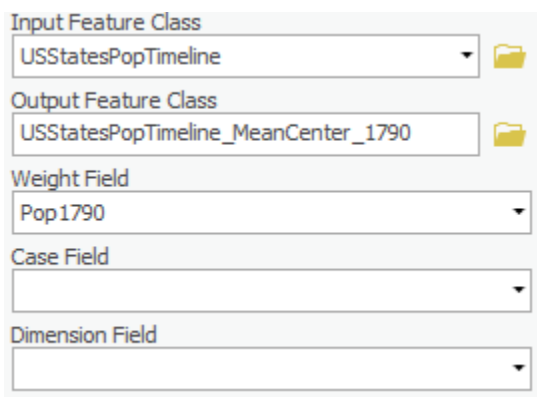
The resulting feature is Illinois. Think about what a different picture the weighted central feature paints compared to simply mapping which states were admitted in every decade. While we would expect the unweighted central feature to be a state such as Kansas, the weighted central feature, Illinois, is significantly northeast of Kansas, despite Alaska and Hawaii to the far west being included in the analysis. This discrepancy is due to the dense population on the East Coast compared to most of the rest of the country.

Mean Center

The mean center is similar to the central feature, but the tool will plot the average x and y coordinate of all the states in the study area as a point, rather than selecting an existing state polygon. The mean center is useful for tracking changes in the distribution over time or for comparing the distributions of different types of features.

1790

1. At the top of the Geoprocessing pane, click the **Back** arrow button.
2. Within the Measuring Geographic Distributions toolset, click the **Mean Center** tool.
3. For 'Input Feature Class', use the drop-down menu to select the **USStatesPopTimeline** layer.
4. In 'Output Feature Class,' rename the feature class "**USStatesPopTimeline_MeanCenter_1790**".
5. For 'Weight Field,' use the drop-down menu to select the **Pop1790** field.
6. Ensure your Mean Center tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Feature Class
USStatesPopTimeline_MeanCenter_1790

Weight Field
Pop1790

Case Field

Dimension Field

In this case, the mean center is located within the central feature for the same year, but that does not always happen.

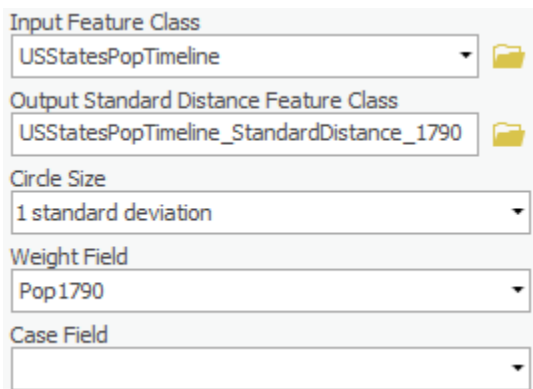
Standard Distance

The standard distance displays the spatial dispersion of data locations on a map, centered on the mean center, similar to how the standard deviation of numeric data describes the dispersion of data values centered on the mean value.

1790

1. At the top of the Geoprocessing pane, click the **Back** arrow button.
2. Within the Measuring Geographic Distributions toolset, click the **Standard Distance** tool.
3. For 'Input Feature Class', use the drop-down menu to select the **USStatesPopTimeline** layer.

4. In 'Output Standard Distance Feature Class', rename the feature class "**USStatesPopTimeline_StandardDistance_1790**".
5. For 'Circle Size', use the drop-down menu to select **1 standard deviation**.
6. For 'Weight Field,' use the drop-down menu to select the **Pop1790** field.
7. Ensure your Standard Distance tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Standard Distance Feature Class
USStatesPopTimeline_StandardDistance_1790

Circle Size
1 standard deviation

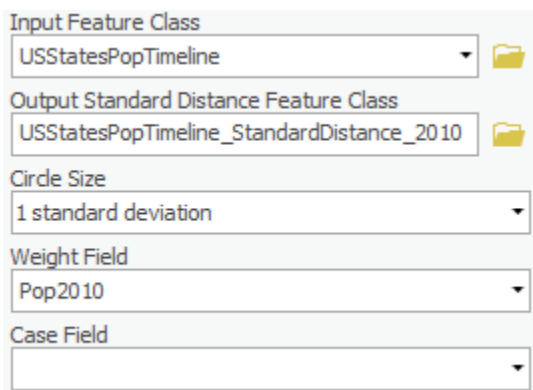
Weight Field
Pop1790

Case Field

The result is a circle centered on Maryland. Because this circle represents one standard distance of the 1790 population, we estimate that 68% of the US population lives within that circle in 1790.

2010

1. For 'Output Standard Distance Feature Class', rename the feature class "**USStatesPopTimeline_StandardDistance_2010**".
2. For 'Weight Field,' use the drop-down menu to select the **Pop2010** field.
3. Ensure your Standard Distance tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Standard Distance Feature Class
USStatesPopTimeline_StandardDistance_2010

Circle Size
1 standard deviation

Weight Field
Pop2010

Case Field

In comparing the 2010 standard distance to the 1790 standard distance, you can see how much more widely distributed the population has become.

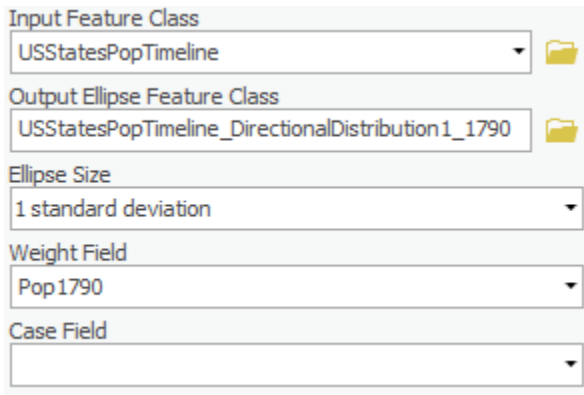
Directional Distribution (Standard Deviatonal Ellipse)

In addition to the distribution of the population, we can choose to find the directional trend, which calculates the standard distance separately in the x and y directions. These two measures define the axes of an ellipse encompassing the distribution of features, which is referred to as the standard deviational ellipse. The ellipse allows you to see if the distribution of features is elongated and hence has a particular orientation.

1790-1SD

1. At the top of the Geoprocessing pane, click the **Back** arrow button.
2. Within the Measuring Geographic Distributions toolset, click the **Directional Distribution (Standard Deviatonal Ellipse)** tool.
3. For 'Input Feature Class', use the drop-down menu to select the **USStatesPopTimeline** layer.
4. For 'Output Ellipse Feature Class', rename the feature class "**USStatesPopTimeline_DirectionalDistribution1_1790**".
5. For 'Ellipse Size', use the drop-down menu to select **1 standard deviation**.
6. For 'Weight Field,' use the drop-down menu to select the **Pop1790** field.

7. Ensure your Directional Distribution tool parameters are configured as shown below and [click Run](#).



Input Feature Class
USStatesPopTimeline

Output Ellipse Feature Class
USStatesPopTimeline_DirectionalDistribution1_1790

Ellipse Size
1 standard deviation

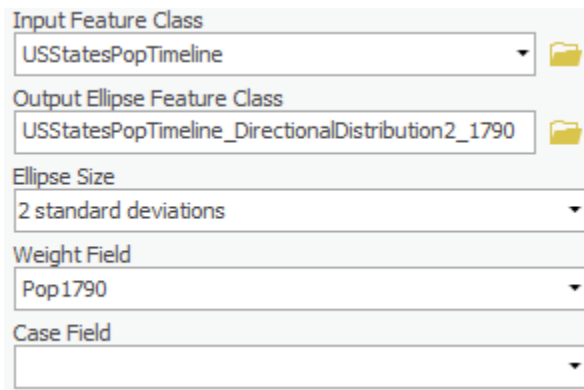
Weight Field
Pop1790

Case Field

Notice how closely the ellipse follows the shape of the northeastern states and how little of the ellipse intersects with the ocean, where no one would be living, compared to the standard distance for the same year.

1790-2SD

1. For 'Output Ellipse Feature Class', [rename](#) the feature class "**USStatesPopTimeline_DirectionalDistribution2_1790**".
2. For 'Ellipse Size', use the drop-down menu to [select 2 standard deviations](#).
3. Ensure your Directional Distribution tool parameters are configured as shown below and [click Run](#).



Input Feature Class
USStatesPopTimeline

Output Ellipse Feature Class
USStatesPopTimeline_DirectionalDistribution2_1790

Ellipse Size
2 standard deviations

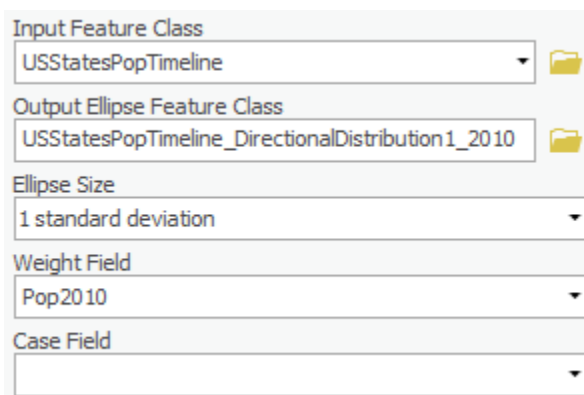
Weight Field
Pop1790

Case Field

Again, notice how closely the ellipse follows the shape of the East Coast. Because this ellipse represents 2 standard deviations, we estimate that 95% of the US population lived within this ellipse in 1790.

2010-1SD

1. For 'Output Ellipse Feature Class', [rename](#) the feature class "**USStatesPopTimeline_DirectionalDistribution1_2010**".
2. For 'Ellipse Size', use the drop-down menu to [select 1 standard deviation](#).
3. For 'Weight Field,' use the drop-down menu to [select the Pop2010](#) field.
4. Ensure your Directional Distribution tool parameters are configured as shown below and [click Run](#).



Input Feature Class
USStatesPopTimeline

Output Ellipse Feature Class
USStatesPopTimeline_DirectionalDistribution1_2010

Ellipse Size
1 standard deviation

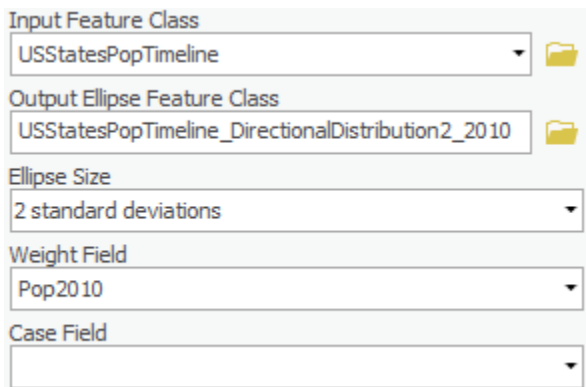
Weight Field
Pop2010

Case Field

Again, in 2010, the directional distribution ellipse is much better fit to the shape of the US than the standard distance circle, which extends far into Canada and the Gulf of Mexico.

2010-2SD

1. For 'Output Ellipse Feature Class', rename the feature class "**USStatesPopTimeline_DirectionalDistribution2_2010**".
2. For 'Ellipse Size', use the drop-down menu to select **2 standard deviations**.
3. Ensure your Directional Distribution tool parameters are configured as shown below and click **Run**.



Input Feature Class
USStatesPopTimeline

Output Ellipse Feature Class
USStatesPopTimeline_DirectionalDistribution2_2010

Ellipse Size
2 standard deviations

Weight Field
Pop2010

Case Field

At two standard deviations, the ellipse no longer appears to represent the shape of the US very well. That is because Alaska and Hawaii are two geographic outliers far to the west. You can see what a dramatic effect these outliers have on the results, so you will need to carefully evaluate whether or not to exclude geographic outliers from your analysis when using this tool.

Linear Directional Mean

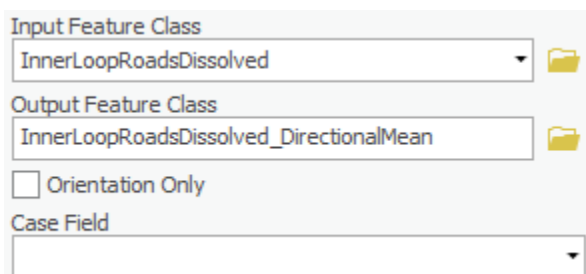
The linear directional mean calculates the average length and average direction, or angle, of a set of lines. A typical application of this tool might be to illustrate the average magnitude and direction of wind or ocean currents. Much like the other geographic distribution tools, the linear directional mean is most informative when compared over time or between variables.

Roads

1. At the bottom of the Geoprocessing pane, click the **Catalog** tab.
2. In the *GeographicDistributionData* geodatabase, right-click the **HoustonFreewaysDissolved** feature class and select **Add To New Map**.
3. Right-click the **InnerLoopRoadsDissolved** feature class and select **Add To Current Map**.

Road data is typically segmented such that the length of road between each pair of intersections is a separate feature. In this tutorial, we are asking what is the average length of roadway with a single name, so the road layer was previously dissolved on the road name field, such that all road segments with the same name were merged into a single feature. The roads layer was also clipped to approximately the extent of the Inner Loop.

1. At the bottom of the Catalog pane, click the **Geoprocessing** tab.
2. At the top of the Geoprocessing pane, click the **Back** arrow button.
3. Within the Measuring Geographic Distributions toolset, click the **Linear Directional Mean** tool.
4. For the 'Input Feature Class', use the drop-down menu to select the **InnerLoopRoadsDissolved** layer.
5. For 'Output Feature Class', rename the feature class "**InnerLoopRoadsDissolved_DirectionalMean**".
6. Ensure your Linear Directional Mean tool parameters are configured as shown below and click **Run**.



Input Feature Class
InnerLoopRoadsDissolved

Output Feature Class
InnerLoopRoadsDissolved_DirectionalMean

☐ Orientation Only

Case Field

7. In the Contents pane, right-click the **InnerLoopRoadsDissolved** layer name and select **Zoom To Layer**.

Notice that the result points in the same direction as the longest roads in downtown Houston. This result indicates that the majority of roads run in a true North-South and East-West direction, which would average out to 45 degrees.

Freeways

1. For the 'Input Feature Class', use the drop-down menu to select the **HoustonFreewaysDissolved** layer.
2. For 'Output Feature Class', rename the feature class "**HoustonFreewaysDissolved_DirectionalMean**".
3. Ensure your Linear Directional Mean tool parameters are configured as shown below and click **Run**.

Input Feature Class
HoustonFreewaysDissolved

Output Feature Class
HoustonFreewaysDissolved_DirectionalMean

☐ Orientation Only

Case Field

4. In the Contents pane, right-click the **HoustonFreewaysDissolved** layer name and select **Zoom To Layer**.

Notice how much longer the linear mean is for the freeways than for all roads. The length is approximately the diameter of Beltway 8, which is the middle of three beltways in Houston. Notice that the direction is also shifted, because the freeways do not run orthogonally, like the local streets.