GIS CAPACITY BUILDING EXPERT FOR UNION OF MUNICIPALITIES IN AKKAR AND NORTH LEBANON

ADVANCED COURSE MANUAL

FEBRUARY 2021

GIS Center, Faculty of Engineering

University of Balamand

Catalan Agency
for Development
Cooperation

UNIVERSITY OF
BALAMAND

TABLE OF CONTENT

Mapping Tabular Data – Display XY points from csv

Materials needed: *AussiePublicToilets.csv*.

1   Open and examine the data:

Open ArcMap and use the **Add Data** ![add data icon] button to add the table *AussiePublicToilets.csv* to the map. **Right-click > Open** to open the table. Take a moment to inspect the field (column) names and the values beneath them. At the bottom of the table, note the number of records, each representing one public restroom, though none show up on the map yet.

Notice that there are columns for Latitude and Longitude, and that the values in these columns look a lot like decimal degrees. (Metadata confirms that the data were recorded using GCS_WGS84.)

2   Set the dataframe coordinate system to WGS84:

Right-click on the word Layers, then **Properties**, then choose the tab for **Coordinate System**. Search for **WGS 1984**, and choose the plain version that lives under **Geographic Coordinate Systems > World**. Notice that the units in the lower right-hand corner are now decimal degrees.

3   Plot the coordinates as points:

Right-click on table in the Table of Contents. Choose **Display XY Data ..** A dialog box opens, asking which fields to use as X and Y coordinates. Choose Longitude for X, Latitude for Y.
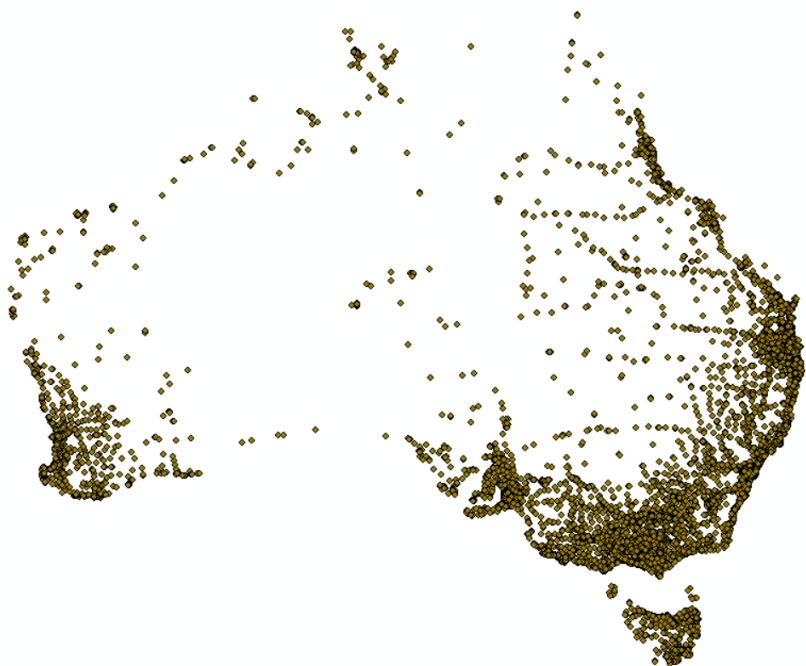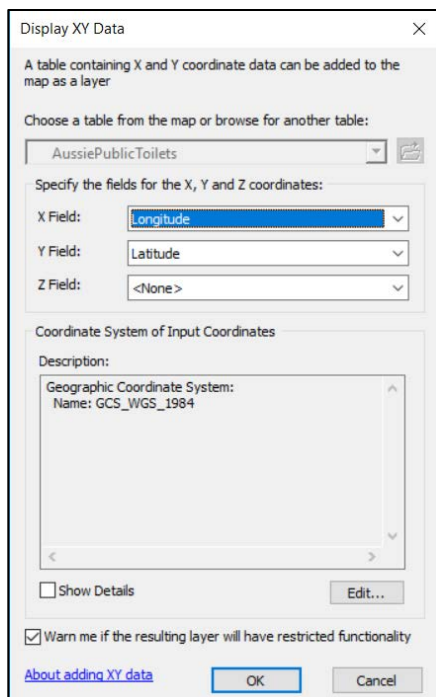
4   Remember that this layer of points is <u>temporary</u>.

To make it permanent, **right-click** on *AussiePublicToilets Events*. Choose **Data > Export**. Name the new file *AussiePublicToilets.shp* and save it somewhere you will be able to find it again. Add that saved file to the map and remove both the events layer and the .csv (**right-click > Remove**).

Mapping Tabular Data – Select a Subset by Attributes

1   Click the **Select** icon. Practice click selecting on the Australian points. **Shift-click** will allow you to select several points. You can also hold down the left mouse button and drag a box to select a large area. When you are done, **Clear Selected Features**. At any time, you can return to a regular cursor by clicking on the black arrow.

2   From the **Selection** menu at the top, choose **Select by Attribute**. A calculator opens, where you will concoct a logical expression. Make sure the Layer says *AussiePublicToilets* and the Method is "Create a new selection." We will select only the bathrooms that are free of charge.

Double-click the field name "PaymentReq." It will appear in the expression box. Then click the = and that will appear as well.

To see what PaymentReq could possibly be equal to, click **[Get Unique Values]**. Choose "False." Your expression should now read: "PaymentReq = 'False'

Click **OK**.

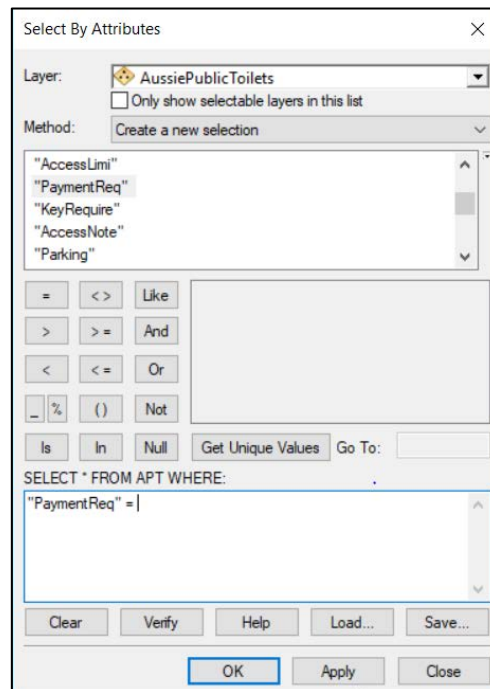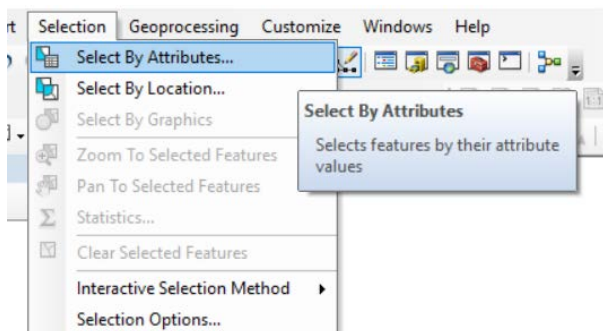3   Inspect the map and the Attribute Table. 162 public toilets were excluded from the selection. Play with the **Switch Selection** button at the top of the Attribute Table, and **Show Selected Records** at the bottom.

4   Narrow the selection by launching **Select By Attributes** again. This time, change **Method** to "Select from Current Selection" and build the expression "KeyRequire" = 'False' and inspect the updated selection. These are all of the public bathrooms on the entire continent that are open access, no key or payment required.

5   Data selections are temporary. To make a new layer that is only the selected portion of the old, **right-click** on *AussiePublicToilets* in the Table of Contents and choose **Data > Export** again. This time, make sure the top box says "Selcted Features." Give it a name and storage destination and click **OK**.

Close ArcMap for now.

Spatial Join

Materials needed: *AussiePublicToilets.shp, Municipalities.shp*

1   Open and examine the data:

Open ArcMap and use the **Add Data** button to add both shapefiles to the map.
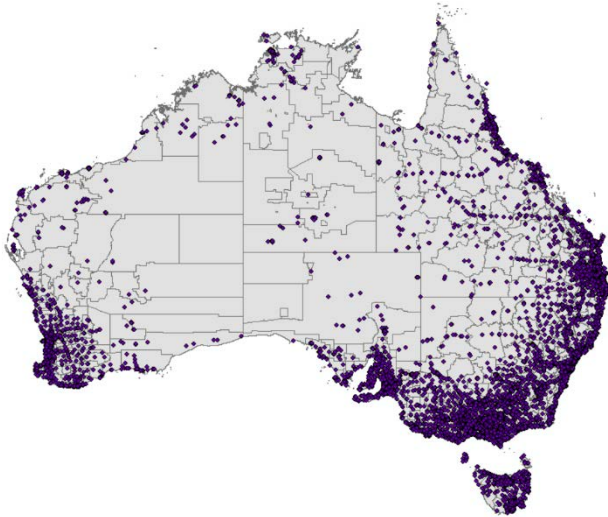
Spatial Join is a technique that associates one layer with the features of another based on their location in space.

2   First, let's join the municipalities to the toilets, so that each toilet will have information about which municipality it's in. Click the ArcToolbox icon to open the Toolbox.

Then navigate to Analysis > Overlay > Spatial Join.

Double-click to run the tool.

The first time we join, the Target Layer will be the toilets and the Join Layer will be the municipalities. Join one-to-one. At the bottom of the dialog, you will see a choice for Match Option, set to INTERSECT. For now, leave that choice, but inspect the alternatives in the drop-down menu. Run the tool and inspect the

output. The output is a set of points, and new columns have been added to the right of the attribute table containing all of the information from the Municipalities shapefile.

3   Now let's do the opposite: tell each municipality about the toilets it contains. We could spatially join one-to-many, but that would make many copies of each municipality, one for each toilet it contains.

Instead, let's get a summary for each municipality, using one-to-one. Open the Spatial Join tool, make the municipalities the Target Layer and the toilets the Join Layer.
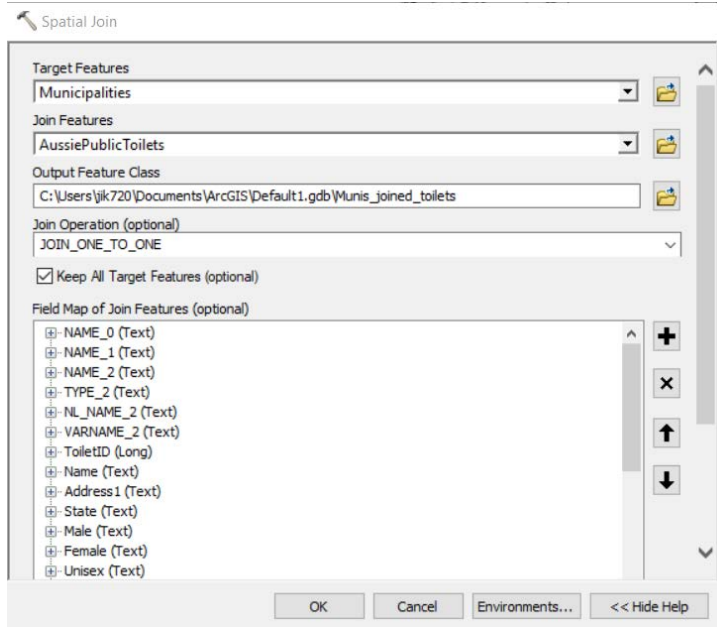
Now, as it says in the help window, we need to set a Merge Rule.

- JOIN_ONE_TO_ONE—If multiple join features are found that have the same spatial relationship with a single target feature, the attributes from the multiple join features will be aggregated using a field map merge rule. For example, if a point target feature is found within two separate polygon join features, the attributes from the two polygons will be aggregated before being transferred to the output point feature class. If one polygon has an attribute value of 3 and the other has a value of 7, and a Sum merge rule is specified, the aggregated value in the output feature class will be 10. This is the default.

Decide which field you would like summary information about, and right-click it in the list.

Let's get the count of unique toilet ID numbers per municipality.

Right-click on ToiletID, select Merge Rule, and then Count.

Leave the Match Option as INTERSECT and run the tool.

Inspect the output. This time, the output is a set of polygons, and a new field has been added which gives the count of toilets per municipality.

We can change the symbology to display that quantity.
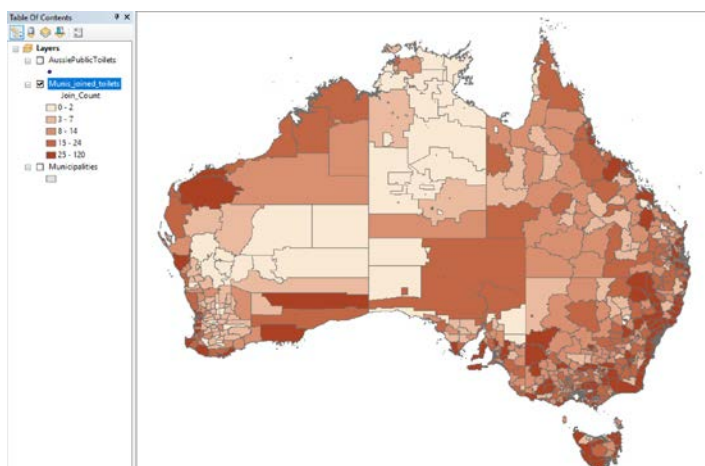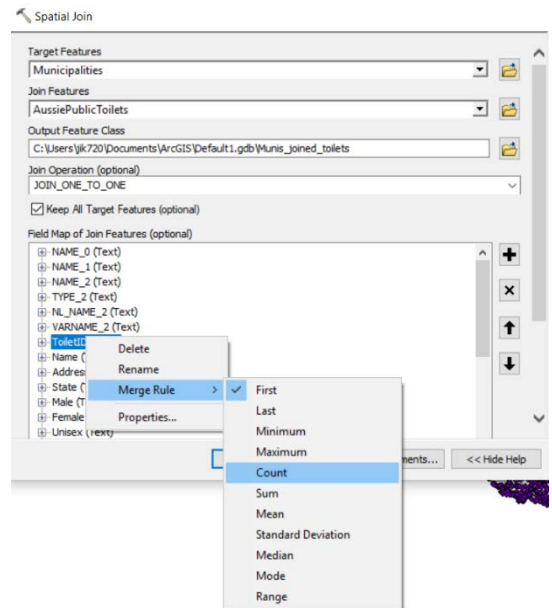
Spatial Analysis: Vector – Clip and Select

Materials needed: *massachusetts.shp, US_census_tracts.shp*

1   Open and examine the data:

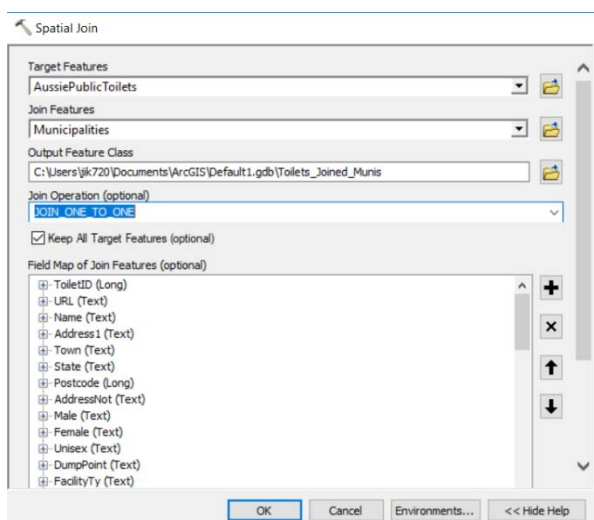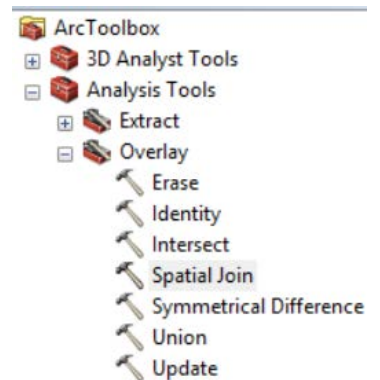Open ArcMap and use the **Add Data** ✛ button to add both shapefiles to the map. **Right-click > Open** on the census shapefile to open the table. Take a moment to inspect the field (column) names and the values beneath them. At the bottom of the table, note that there are 73,682 records, each representing one census tract. Now do the same for the outline of Massachusetts. There is not much information there, but the number we need is the State FIPS number: 25. In the full census shapefile, only the tracts with FIPS 25 are in Massachusetts.

2   Our goal is to create a new shapefile containing only the Massachusetts tracts. As we did in the Mapping Tabular Data Lab, use Select by Attribute to select only the census tracts with State_FIPS = 25.

Remember to fill in the calculator formula by clicking on choices, rather than typing them in, to get the proper syntax. As expected, the Massachusetts tracts are highlighted.
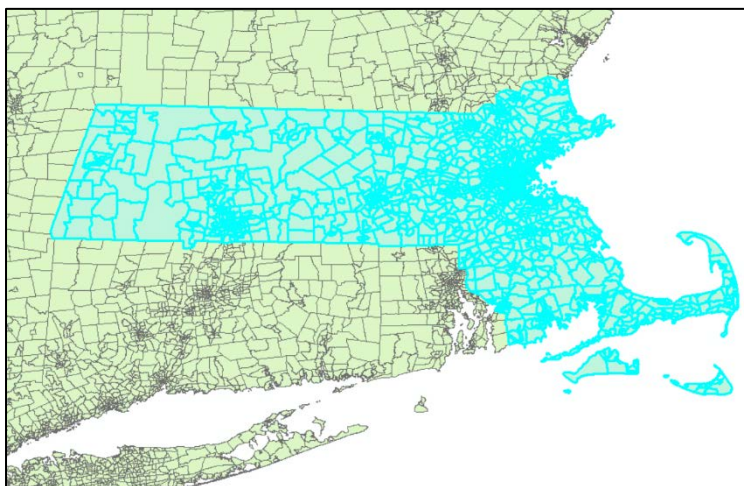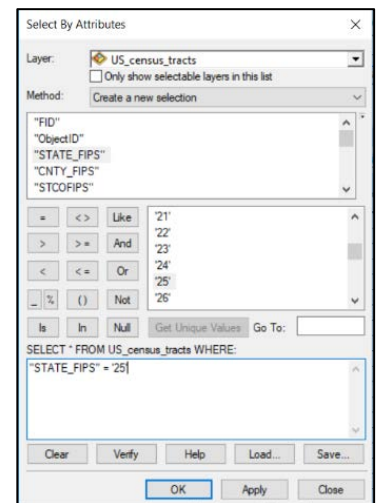
Right-click > Data > Export, with the top drop-down set to "Selected Features" is one way to create the new shapefile we want.

This lab exercise will teach you another method.

3   Use the Clear Selected Features ⬚ button to clear your selection. Then, use the ArcToolbox button 🟥 to open the toolbox, if it is not already open. Navigate to Analysis > Extract > Clip.

The Clip tool works like a cookie cutter. It needs two inputs: the larger Input layer that needs to be cut down, and a Clip layer to do the cutting.
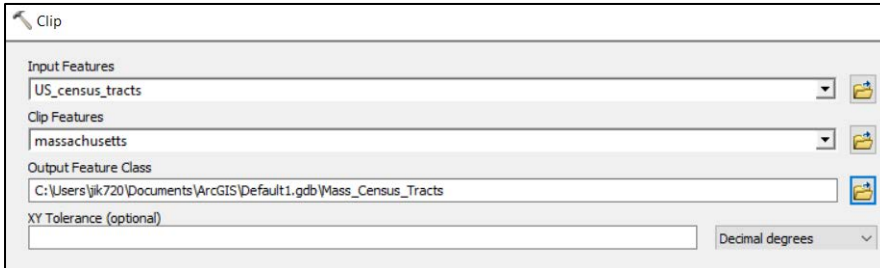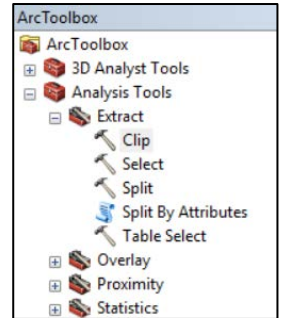
Name your output Mass_Census_Tracts and run the tool. The output will be the same as the earlier selection.
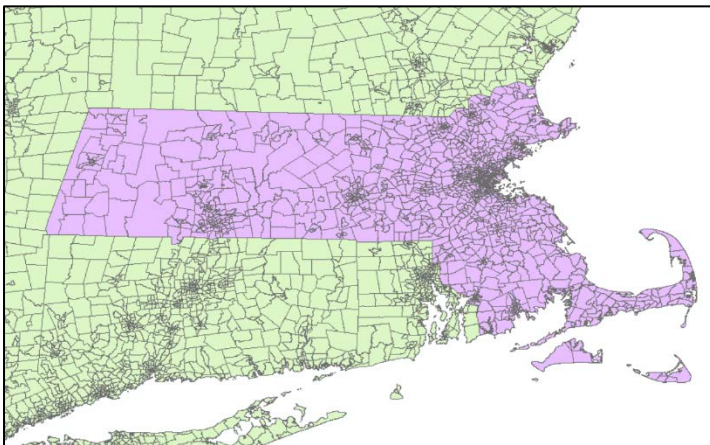
Clipping is a good technique to use when there is no common attribute to select by or when you need a purely spatial subset of a layer. Note: the Clip layer has to be one or more polygons, but the Input layer can be any vector layer. There is a separate Clip tool for rasters.

Notice that the Attribute Table for the output layer retains no information from the Clip layer.

Spatial Analysis: Vector – Buffer

Materials needed: *Mass_Census_Tracts.shp, hospitals.shp*

1   Add the hospitals shapefile to the map and turn off or remove the full US_census_tracts layer for simplicity. In Drawing Order, drag the hospitals on top of the census tracts so you can see them.
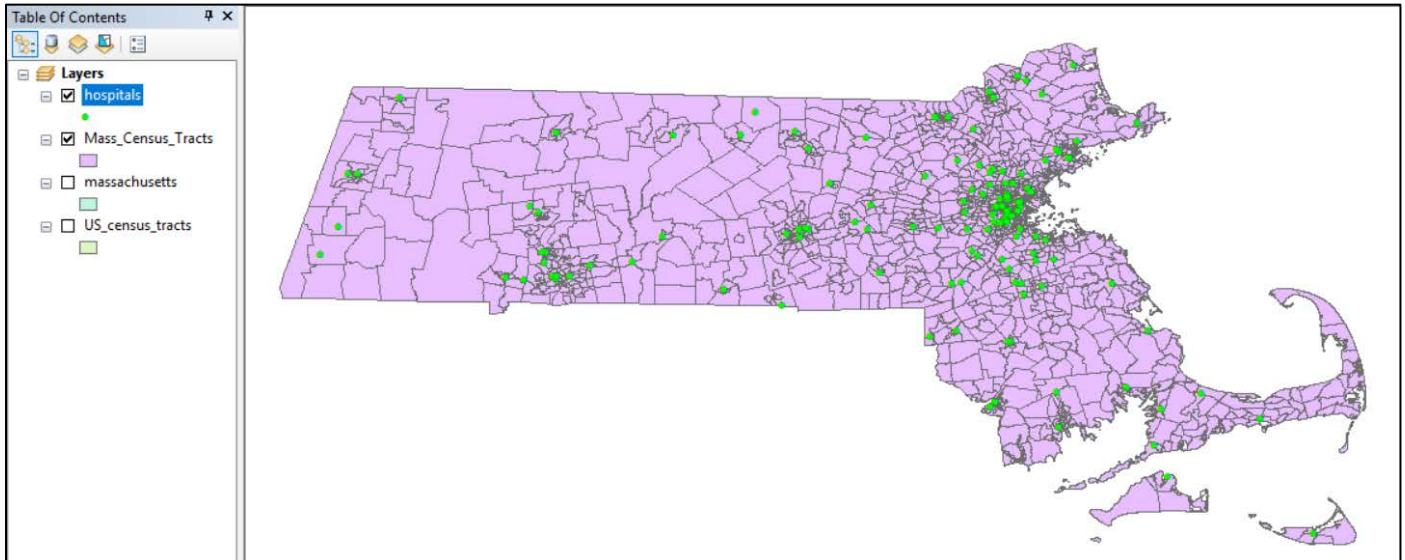
2   Suppose we would like to know what parts of Massachusetts are within 5 miles of a hospital. The Buffer tool will draw regions that mark off a distance from each, or from the nearest, hospital.

In the toolbox, navigate to Analysis > Proximity > Buffer and launch the Buffer tool.
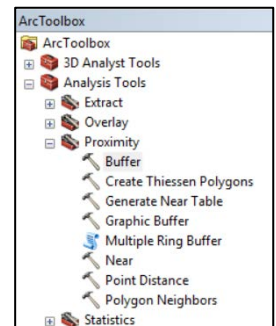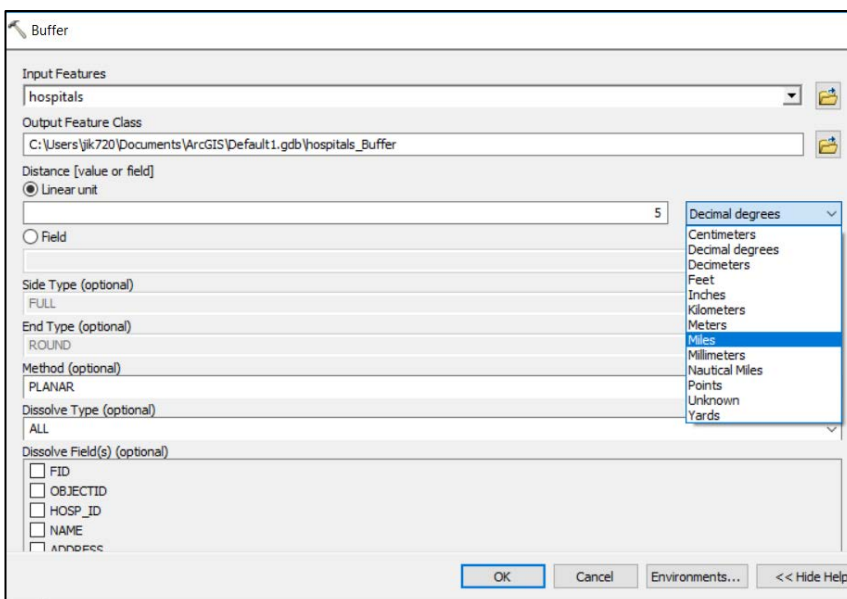
The input layer is the Hospitals. The output name can be hospitals_buffer. Type 5 in the Linear Unit box.

You will notice that the default units for Distance are Decimal Degrees. That is because these layers are in the WGS 84 coordinate system, and because they are all we added to the DataFrame, that also defaults to decimal degrees. Best practice is to use distance-based tools, such as buffer, in a distance-based projection such as Massachusetts State Plane. But, ArcMap will make the adjustment for us, for the purposes of this exercise. Just change the dropdown to "Miles."

Also set the Dissolve Type to ALL. If Dissolve is set to NONE, we will get individual discs around each hospital point. All will blend those into irregular blobs indicating land that is within 5 miles of any hospital, not a particular one. Click [OK] to run the tool.

It is also possible to create buffers of different distances, based on the values in some field in the Attribute Table. To do so, you would use the Field option instead of Linear Unit.
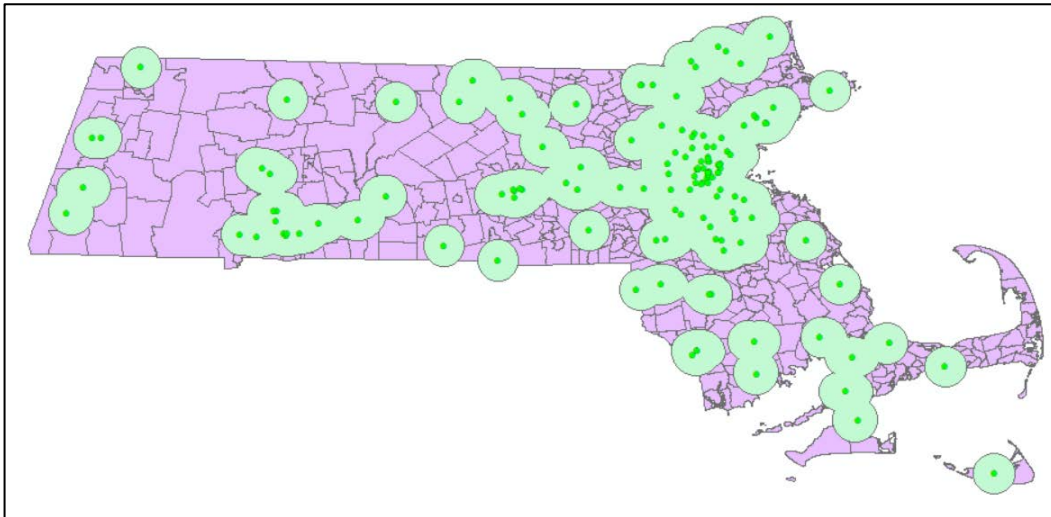
Notice that the Attribute Table for this buffer layer has only one entry. All of these blobs are combined to make one multi-part polygon. You could split it up with a tool called Multipart-to-Singlepart if needed.

Spatial Analysis: Vector – Intersection

Materials needed: *Mass_Census_Tracts.shp, hospitals_buffer.shp*

1 Suppose we were conducting an analysis of how many Massachusetts residents live within 5 miles of a hospital. We have the distance information in the buffer layer and the population information in the census tracts. The Intersect tool will extract only the areas where the two overlap.

2 Navigate to Analysis > Overlay > Intersect and launch the Intersect tool. Add both the hospital buffer and the census tracts to the inputs. The order matters because the projection of the first input will be used.
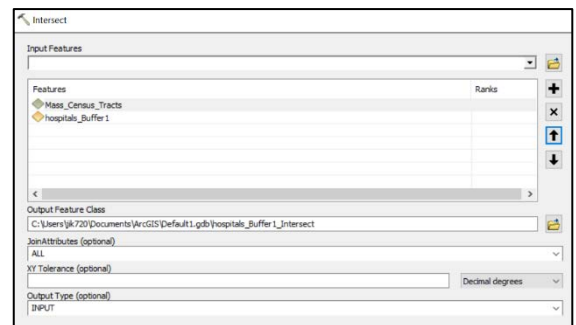
[3] Examine the Attribute Table. There is a field for population, but that field has just been copied from the original. *It has not been adjusted to represent the partial census tracts in the intersection!*

FIPS	POP2000	POP00_SQMI	POP2010	POP10_SQMI
25001011801	2456	654.9	2403	801
25001011802	4254	719.8	4255	851
25001012001	5794	1350.6	5732	1910.7
25001012002	3335	1407.2	3007	1503.5
25001012101	5484	982.8	5339	1067.8
25001012102	3484	862.4	3057	1019
25001012200	5041	271	5046	280.3
25001012502	3726	1527	3080	1540
25001012601	2755	2056	2877	2877
25001012602	4991	3809.9	4915	4915
25001012700	5173	1286.8	4562	1520.7
25001012800	3980	963.7	3923	1307.7
25001013100	6146	640.9	5935	741.9
25001013200	5021	654.6	4783	683.3
25001013300	3311	654.3	3218	804.5
25001013400	3943	458	4163	520.4
25001013500	6588	677.8	6940	771.1
25001013600	6207	816.7	6334	904.9
25001013700	4158	764.3	4911	982.2
25001013800	4658	950.6	4840	1210
25001013900	4381	693.2	4831	805.2
25001014002	3936	515.9	4053	579
25001014100	1599	50.2	1141	36.8
25001014300	4729	744.7	4384	730.7
25001014402	6469	386.4	6947	463.1
25001014500	5664	824.5	5395	899.2

Extra Practice: To make the adjustment would require adding a new field and calculating (using Field Calculator) the ratio of the areas of the partial census tracts to the areas of their originals. Then multiply that fraction by the whole-tract population. Consider trying this exercise later on your own.

Spatial Analysis: Vector – Union

Materials needed: *Mass_Census_Tracts.shp, hospitals_buffer.shp*

The Union tool combines two or more layers in a way that keeps everything, rather than excluding anything. Navigate to Analysis > Overlay > Union and launch the Union tool. As before, add the input layers in the order shown.
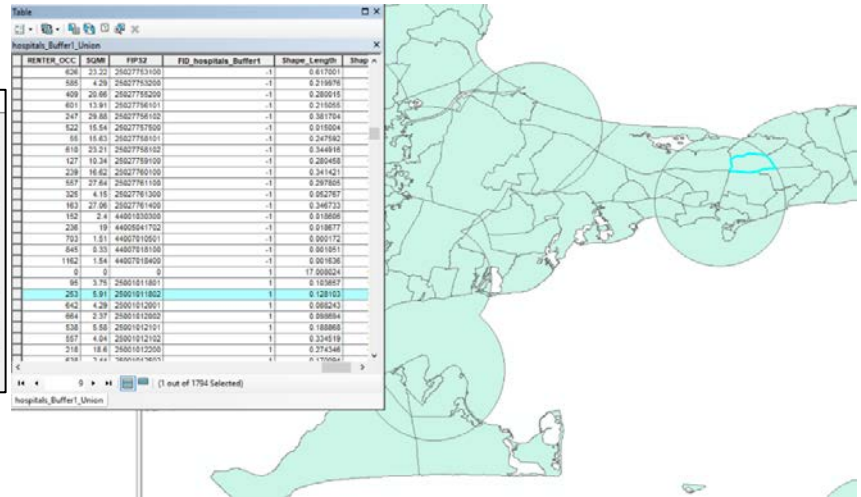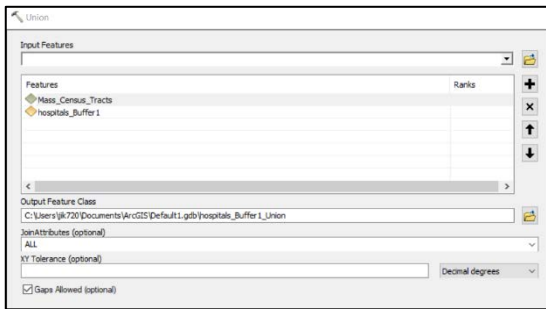
The output of Union can be a little hard to parse. There are now more small pieces of the original census tracts, as well as pieces and the entirely of the buffer multi-polygon.
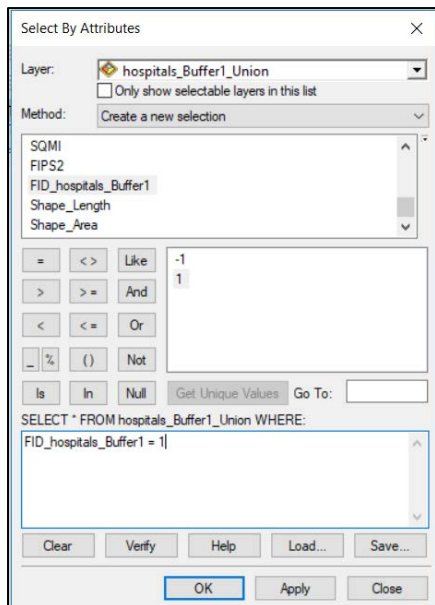
One useful field in the output is a column of 1's and -1's which indicates whether a polygon originally fell in the buffer layer or not.

We can use that column to explore whether the population living within 5 miles of a hospital is different from the population living further away. We will explore if they differ in age.

Use Select by Attribute to select the polygons within the buffer zones. In the Attribute Table, right-click on the MED_AGE column and choose Statistics. Then switch the selection to those outside the buffer zones and repeat the Statistics. The average median age is younger in the zones close to a hospital.
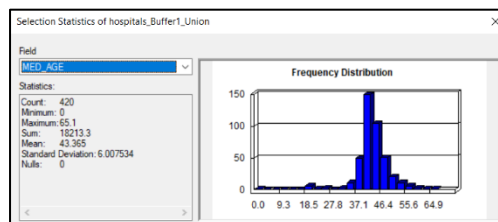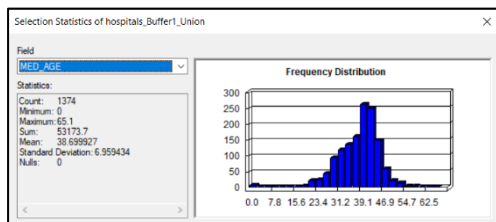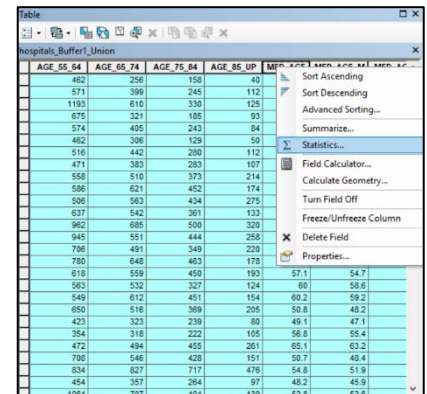
Spatial Analysis: Raster - Slope

Materials needed: *frack.mxd* and *frack.gdb*

1 Open and examine the data:

Open ArcMap, and in the file menu, find Open. Navigate to *frack.mxd* and open it. If the layers appear with small red exclamation marks, [**!**] repair the links by clicking on one such mark and navigating to that layer in *frack.gdb*.

The following layers appear in the Table of Contents: (explore by turning them on and off and viewing Attribute Tables)

Wells – a point layer of existing fracking wells

Roads – a line layer of the major roads

Counties – the outlines of eight counties in Pennsylvania

NLCD11 – National Land Cover Data from 2011, each color is a category of land cover

NLCD_AG – the same Land Cover data, where every category has been set to 0 except Agriculture (1)

DEMs – "Digital Elevation Model," a layer showing elevation

Turn off the County labels by right-clicking on the layer name and unchecking "Label Features." Turn off or re-arrange layers so that you can see the DEM. Zoom in until you can see the pixels and use the Identify Cursor to view some of the pixel values. Zoom back out to the full layer.
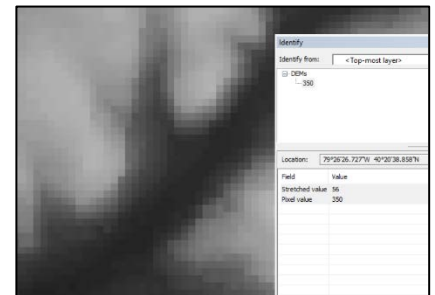
2 For every cell, the Slope Tool will calculate the greatest "steepest" difference between the cell's elevation and that of its neighbors. Open ArcToolbox using the toolbox icon. Then go to Spatial Analyst > Surface > Slope and launch the Slope tool.

ArcToolbox
- Geostatistical Analyst Tools
- Linear Referencing Tools
- Multidimension Tools
- Network Analyst Tools
- Parcel Fabric Tools
- Schematics Tools
- Server Tools
- Spatial Analyst Tools
  - Conditional
  - Density
  - Distance
  - Extraction
  - Generalization
  - Groundwater
  - Hydrology
  - Interpolation
  - Local
  - Map Algebra
  - Math
  - Multivariate
  - Neighborhood
  - Overlay
  - Raster Creation
  - Reclass
  - Segmentation and Classification
  - Solar Radiation
  - Surface
    - Aspect
    - Contour
    - Contour List
    - Contour with Barriers
    - Curvature
    - Cut Fill
    - Hillshade
    - Observer Points
    - Slope
    - Viewshed
    - Viewshed 2
    - Visibility

Slope

Input raster
DEMs

Output raster
C:\Users\jjk720\Documents\ArcGIS\Default1.gdb\Slope_DEMs1

Output measurement (optional)
DEGREE

Method (optional)
PLANAR

Z factor (optional)
1

Z unit (optional)
METER

Choose DEMs as the input layer. You may set the output directory to anywhere you choose. Name the output Slope_DEMs and run the tool.

Change the symbology to Classified, Quantile with 7 classes.

The units in the Table of Contents are degrees. You will notice that the surface of the rivers is very level (green) while their banks are steep. Mountainsides are also rather steep.

Table Of Contents

layers
- Wells
- Roads
- Counties
- Slope_DEMs
  <VALUE>
  - 0 - 2.728093944
  - 2.728093945 - 4.774164402
  - 4.774164403 - 6.82023486
  - 6.820234861 - 9.093646648
  - 9.093646481 - 11.82174042
  - 11.82174043 - 15.68654018
  - 15.68654019 - 57.97199631
- NLCD11
- NLCD_AG
  - 0
  - 1
- DEMs
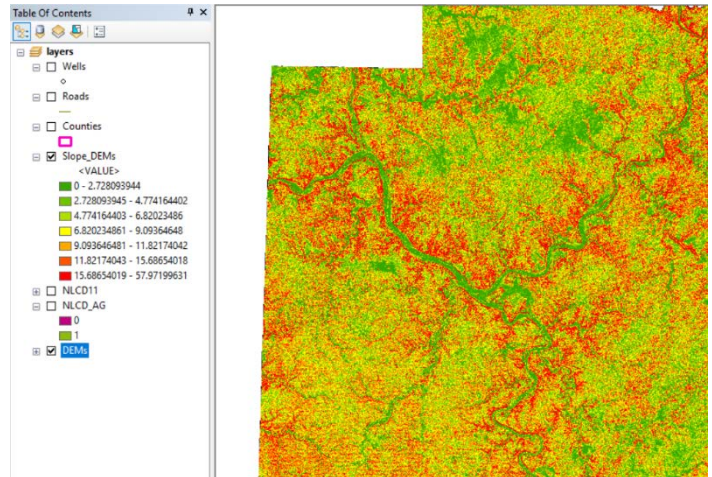
We can probably assume fracking companies would want to place wells on more level terrain.

Spatial Analysis: Raster – Distance

Turn off the Slope_DEM layer and turn on the Wells layer. The Euclidean Distance Tool will compute, for each raster cell, how far it is from the nearest well. In the Toolbox, go to Spatial Analyst > Distance and launch the Euclidean Distance Tool. The input layer is Wells. This input layer can be a vector layer or a raster layer, but special considerations must be taken when it is a raster layer.
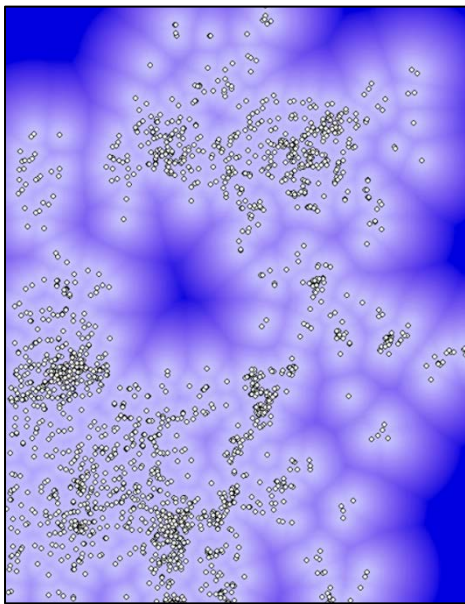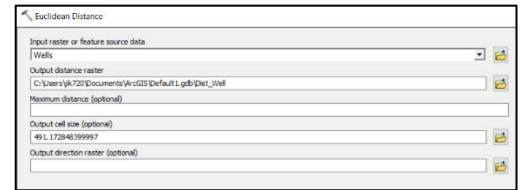
Set the output name to Dist_Wells and run the tool. Change the symbology to Stretched for a nicer image.

In the color scheme displayed here, the dark blue areas are furthest from any well. The Inspector Cursor will reveal how many meters any location is from its nearest well.

We can assume that fracking companies would locate a new well close to others, since that is where the gas is.

Repeat this exercise to create a layer showing distance to the Roads.

We can assume that new wells would be located close to major roads, to make it easy to transport the gas they extract.

The final exercise in this lab will attempt to predict likely locations for new fracking wells to understand which agricultural lands might be at risk from fracking. We will develop a formula to compute an index based on the slope of the terrain and the distance to existing roads and wells. The highest values of the computed index will be the most at-risk for fracking development.

But before that, we should re-scale the slope and distance factors so that more desirable pixels have higher values. We can "reclassifly" the slope and distance rasters to a five-point scale.
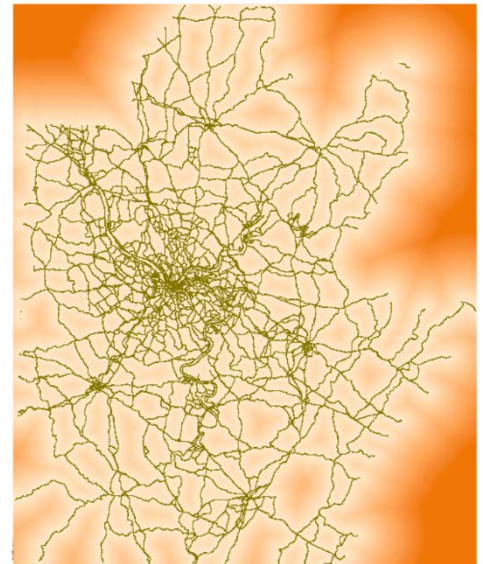
Spatial Analysis: Raster – Reclassify
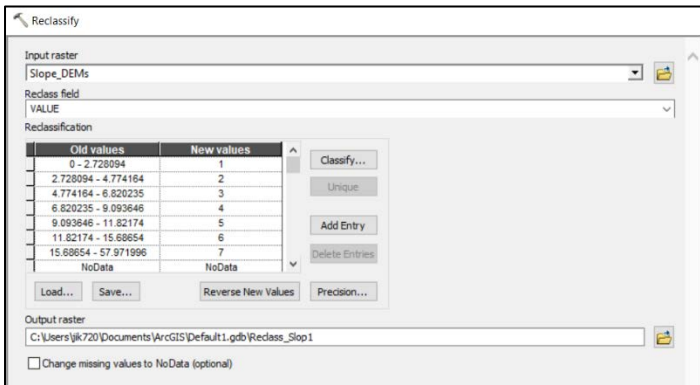
First, reclassify the slope layer by navigating to Spatial Analyst > Reclass > Reclassify. When you choose Slope_DEM as the input layer, the Reclassify Tool automatically takes a guess at how you might want to change the raster's values – from degrees to a 7-point scale.

We would prefer a 5-point scale, and the new values run backwards – low slopes (flat) are given low scores.

To change the value ranges, click Classify. The same window that allows you to classify raster symbology appears. Change the classification to Quantile with 5 classes. Click ok, and then **hand-type** new values in descending order in the Reclassification box. Save the output as Slope_reclass and run the tool.

Repeat this exercise with the roads. Change the classification to quintiles and give lower distances higher scores.

When you are done, you will have three reclassified layers. We will use those and the NLCD_AG layer to determine which agricultural lands are at risk from fracking.

Spatial Analysis: Raster – Map Algebra

Now we will create the formula that combines all four layers. Go to Spatial Analyst > Map Algebra and launch the Raster Calculator. We will multiply the four layers together to get a score that indicates fracking risk, based on these four variables. Enter the following formula by clicking the layer names from the list (rather than typing them in).
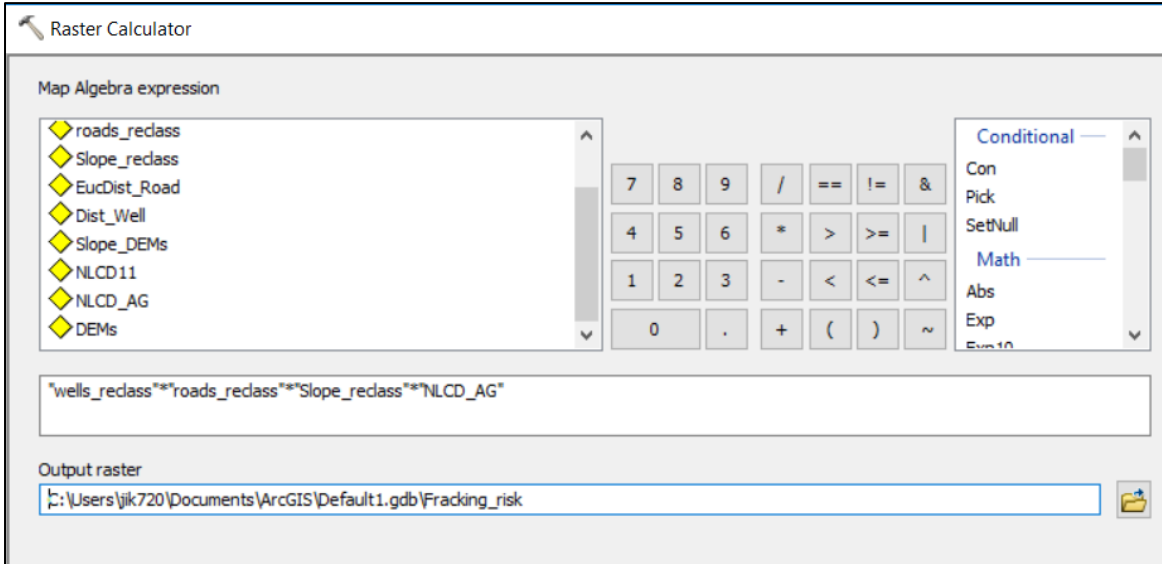
Notice that all of the non-agricultural lands will have the value zero in the output because of the multiplication.

Here, the symbology has been changed so that the areas most at risk are colored red.
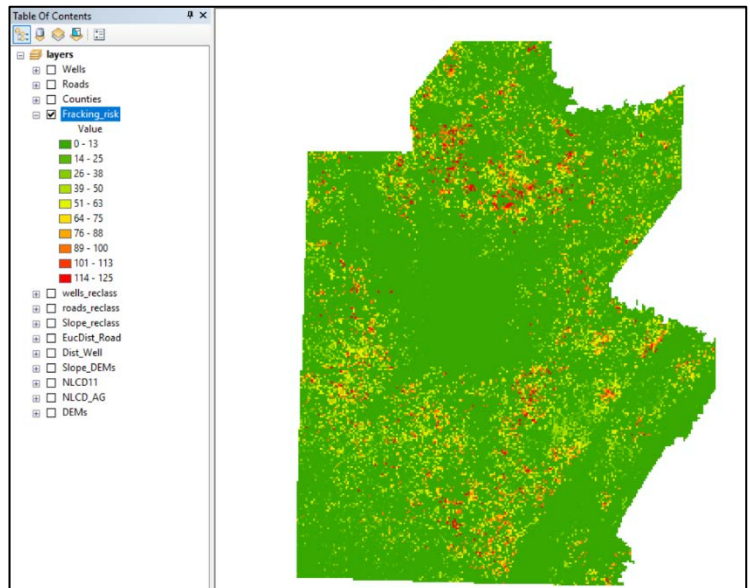
Zonal Statistics

Now turn back on the counties layer. The Zonal Statistics Tool can create a statistical summary of the risk per county.

Go to Spatial Statistics > Zonal and launch Zonal Statistics as Table.

Zonal Statistics needs two inputs: a raster layer to be summarized and a layer of "zones" – in this case counties – giving the units to summarize over. The Zone Field can be any field that has a unique value for each zone.
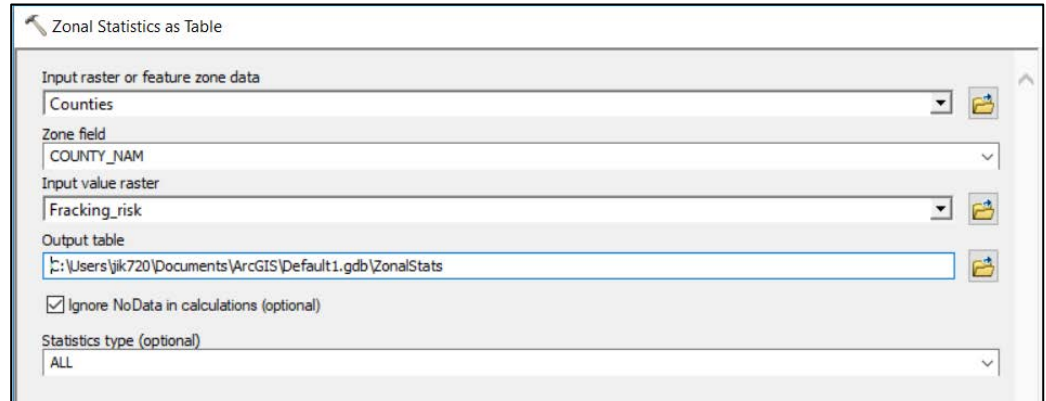
The output of Zonal Statistics as Table is . . . a table! But that table can then be joined, using County Name as a key, to the counties layer.

OBJECTID *	COUNTY_NA	ZONE_CODE	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM	VARIETY	MAJORITY	MINORITY	MEDIAN
1	Butler	1	7116	2010425768.85452	0	125	125	16.42664	32.68879	11689	26	0	6	0
2	Westmoreland	2	9191	2596658690.49211	0	125	125	11.73180	25.78035	10782	29	0	3	0
3	Fayette	3	7318	2067495190.62357	0	125	125	10.98565	25.94536	80393	27	0	3	0
4	Greene	4	5278	1491150535.13408	0	125	125	7.011936	19.76860	37009	26	0	27	0
5	Allegheny	5	6822	1927364333.21044	0	125	125	2.035767	10.6273	13888	26	0	6	0
6	Armstrong	6	6064	1713212740.63150	0	125	125	10.83113	24.83853	65680	28	0	2	0
7	Beaver	7	4076	1151559223.4192	0	125	125	8.805447	23.57207	35891	26	0	6	0
8	Washington	8	7896	2230792843.01226	0	125	125	14.67920	27.83749	11590	26	0	27	0

The newly joined layer can then be colored according to the mean value of the risk index for each county.
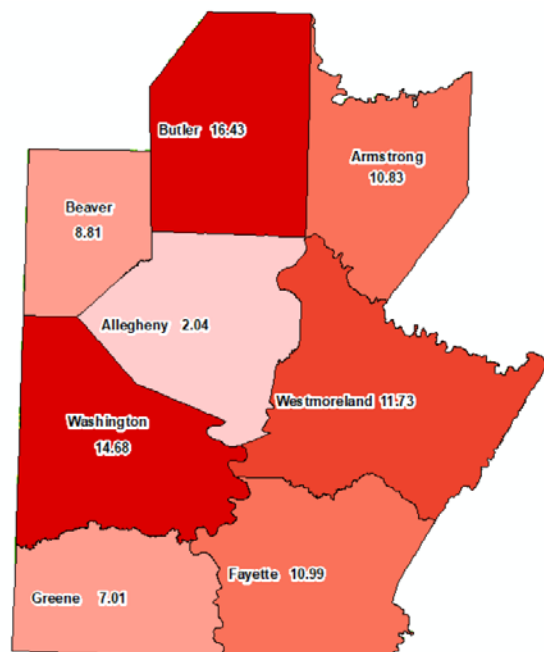
Network Analyst

Network Analyst is a GIS extension that allows you to analytically interface networks (systems of lines and points. The tool's features can help design efficient routes, determine service areas around a specific site (in terms of time and distance), and calculate the nearest facilities or vehicles.

For this lab assignment, you will explore the four major ways to analyze a transportation network: by route, closest facility, service area, and origin-destination cost matrix.

In this lab you will use:

Lab8_PreparedFiles: SFRoad_speed shapefile, SF_groc_two_locs, SF_groc_safeway, San Francisco US Bank branches geocodes

Part I: Creating a Shapefile Based Network Dataset

1   To create a network dataset, we will use a more detailed road dataset that contains information not contained in our Tiger/Line road dataset, such as average speed. **Download Lab8_PreparedFiles** and find the following road shapefile obtained from DataSF and modified so as to include travel time on all road segments.

To do this, each street segment has the speed limit in the attribute table. Time was calculated from that by using the formula (Distance = velocity*time; Time = distance/velocity)

2   Open ArcCatalog. On the Customize menu, click **Extensions** and in the Extensions dialog, check the box next to Network Analyst. Close the Extensions dialog.

3   **Navigate** to your **lab8** folder. Expand your lab folder to show its contents. Right-click on "SFRoads_speed" and choose **New Network Dataset**. This shapefile contains street data for the City of San Francisco.

4   The name of the network dataset is set to **SfRoads_speed_ND** by default. Click **Next** to continue.

5   The dialog box will ask you whether you want to model turns in the network. By default, Global Turns are selected. Global Turns are rules that can be set such as all left turns have a delay of 15 seconds, thus giving an advantage to right turning movements. **Click Yes** to model turns in the network.

6   You will accept the default <u>connectivity</u>; therefore click **Next** to continue.

7   The next dialog box will ask whether you want to specify elevation fields to establish connectivity. This information is used to determine whether the endpoint of line features have the same Z-elevation value. If the values are different, then the two endpoints do not connect, e.g. in the case of bridges or tunnels. ArcGIS Network Analyst automatically searches for the relevant fields. As our dataset does not contain elevation fields, the Network Analyst automatically chooses the No radio button and assigns no fields. Click Next to accept.

8   The next dialog box will ask you to specify the attributes in the network dataset. Network attributes are properties of the network that are used to control navigation, i.e. costs that function as impedances over the entire network.

Network Analyst automatically searches for and assigns relevant fields. Here, "Minutes" is found and assigned.

9   In order to add another impedance Click on **Add**.

10   In the "Name" field, type **Length.**

11   The "Usage Type" field select **Cost**.

12   For "Units", select **Miles** (as this is the unit of measure associated with our shapefile).

13   For "Data Type" select **Double**.

14   Click **OK.**

Let's also add **Speed** as an impedance.

15   Click **Add** again.

16   In the "Name" field, type **Speed**.

17   In the "Usage Type" field select **Descriptor**. (There are no units associated with this attribute).

18   For "Data Type", select **Integer**.

19   Click **OK.**

When finished, you should have three attributes listed. **Click Next** to continue.
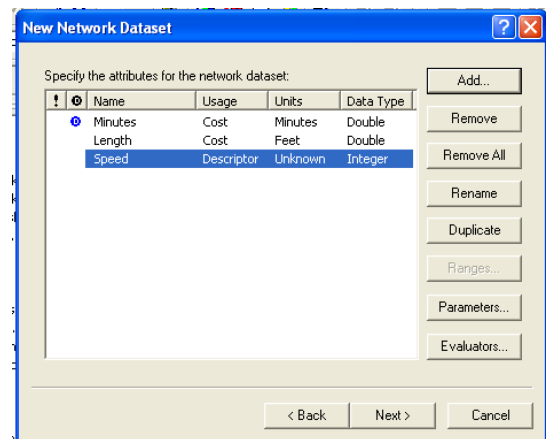
20   The next dialog box asks whether you want to establish driving directions. **Select Yes**. Then click on the "Directions" button.

21   Under the "General" tab, click on **Streetname** under the "Name" field (see red circle—it may have a value there by default, you should also widen the field to see the full name)

22   Change the value of the "Name" field from "Streetname" to **Street**.

23   Under "Suffix", select **ST_Type**.

24   When finished, click **OK** and then click **Next** to continue.
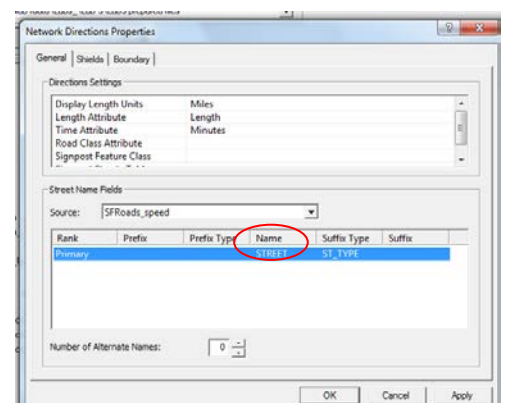
25   A summary of all the setting that have been chosen are displayed in the dialog box. **Click Finish** to create the new shapefile network dataset.

Once created, the system prompts for the network to be built. **Click Yes.**

26   **Open Arcmap and close Arc Catalog.**

Part II: Finding the Best Route Using a Network Dataset

On Saturdays you visit Golden Gate Park. Every week you visit friends, go shopping and make other trips. As you know GIS, you choose to use Network Analysis to find the best route.

Preliminaries

1. **Save** your document.
2. **Turn on the network analyst extension, just the same as in ArcCatalog.**
3. On the <u>Main menu</u>, click **Customize**, point to **Toolbars**, and click **Network Analyst**.
4. Add the SFRoads_speed_ND.nd shapefile you created in ArcCatalog. Add all feature classes that are associated with this network dataset.
5. If the Network Analyst Window is not already open (if it is open, you should see a blank box to the right of your table of contents), click the Network Analyst Window button (blue and white icon with blue flag) on the Network Analyst toolbar.

Button to open Network Analyst window

Network Analyst Window
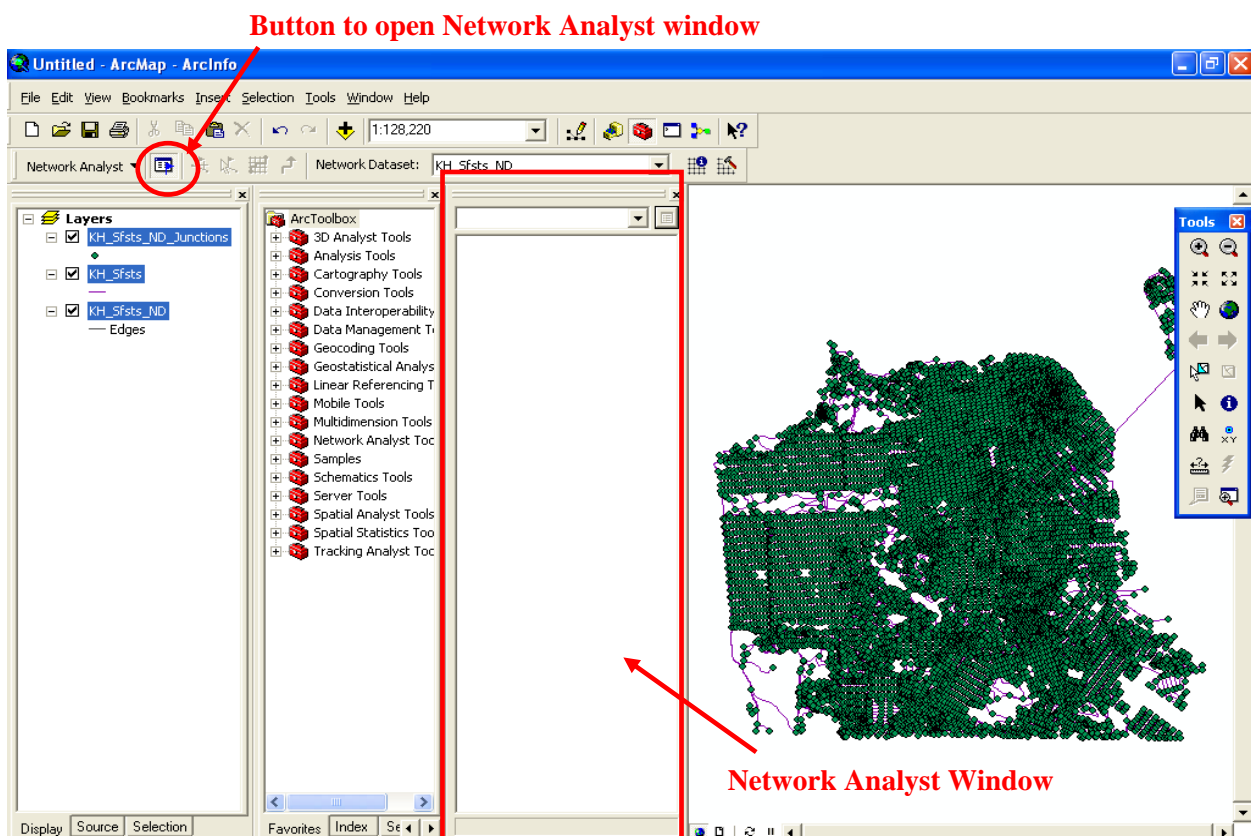
Creating the Route analysis layer and Adding stops

1. On the Network Analyst toolbar, click the Network Analyst drop-down menu and click **New Route**.

2. Right-click **Stops(0)** on the Network Analyst Window and then click on **Find Address**. Use the ArcGIS 10.0 online geocode service. Enter "6101 Fulton St, San Francisco, CA" in the text box. Click Find.

3. **Right-click** on the first object found and click **Add as Network Analyst Object**. This adds the located address as a stop on the Network Analyst Window.
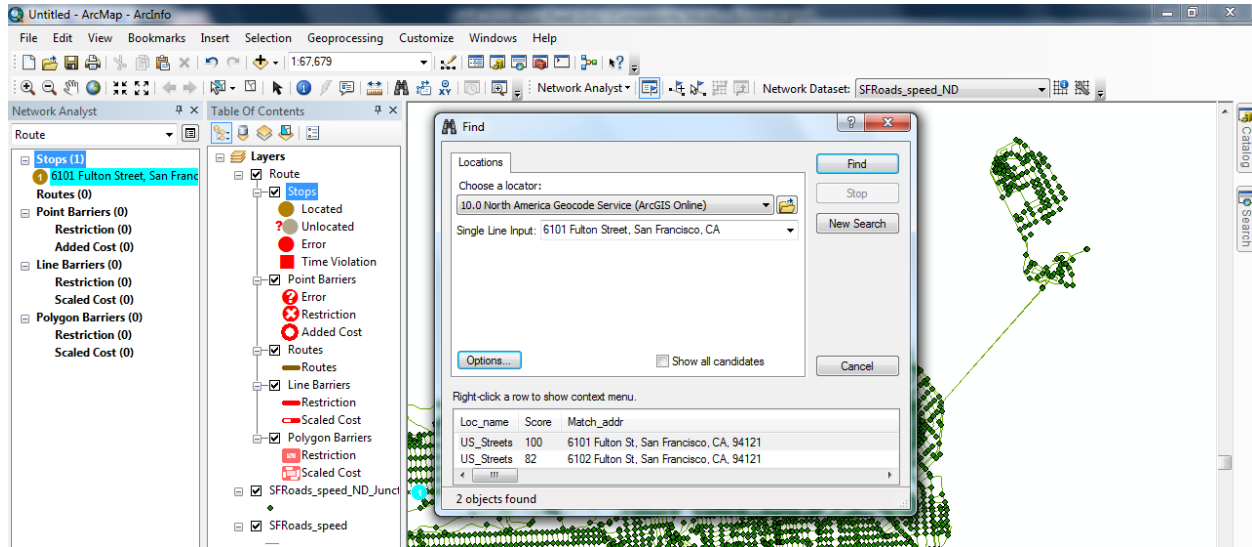
4. Click **New Search** and repeat these steps to add the following addresses as stops in your Network Analyst Window, being sure to use the **US_Streets** shapefile if given a choice.

No.	Street or Intersection
2.	1021 Sanchez St, San Francisco
3.	500 Raymond Ave, San Francisco
4.	1948 Ocean Ave, San Francisco
5.	147 Townsend St, San Francisco
6.	6101 Fulton St, San Francisco

These are all the places you are going to stop, beginning at the Golden Gate Park (6101 Fulton St) and finishing there.

5. **Close** the Find dialog box window.

Setting up the parameters for the analysis

Next, you will specify that your route will be calculated based on time (minutes) and that U-turns are allowed only at dead ends.

1. **Select** on **Route(0)** in the Network Analyst Window.

2. Then **click** the Analysis Layer Properties button (at the top of the window, the button that looks like a piece of paper with a bullet list next to the drop down list where Route has been selected) to bring up the Layer Properties dialog for Route.

3. In the Layer Properties dialog, click the **Analysis Settings tab**.

4. Make sure the impedance is in **Minutes**.

5. Leave the "Use Time Windows" box un-checked.

6. **Check** the "Reorder stops to find optimal route" box.

7. Change **U-Turns at Junctions** to "allowed only at dead ends."

8. **Check** the box labeled "Ignore Invalid Locations".

9. Make sure the Distance Units are set to **Miles** and the "Use Time Attribute" box is checked and set to **Minutes.**

10. **Click OK.**
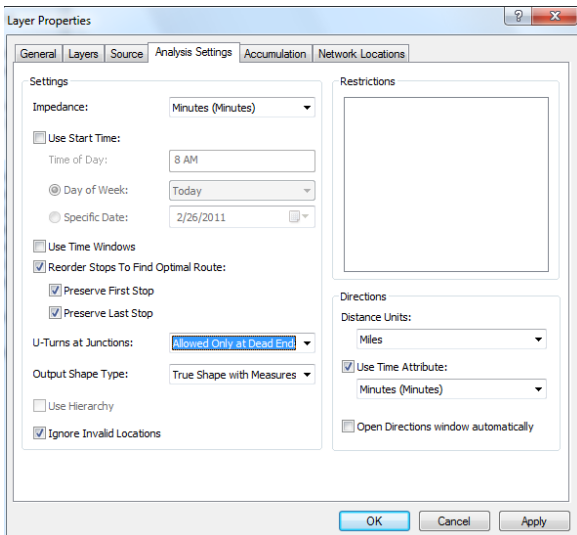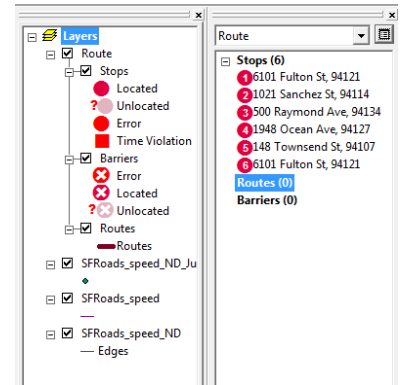
Run the process to find the best route

1. **Click** the **Solve** button (blue line over a grid, see picture) on the Network Analyst toolbar. A route polyline appears in the map and in the Route category of the Network Analyst Window.

2. Click the plus (+) sign next to Routes on the Network Analyst Window to show the Route.

3. Right-click on the new route and click **Directions Window** to display driving directions. **How do these directions compare to Google maps?** Close the Directions Window when finished.

Adding a barrier

Now we will add a barrier on the route that represents a road block due to a water main break. You will find an alternate route to the destination.

1. Right-click on **Point Barriers(0)** in the Network Analyst Window. Click on **Find Address**. Enter **Castro St & 24ᵗʰ Street, 94114**. Click find. **Right-click** on the first object found and **Add as a Network Analyst Object. Close** the Find dialog box.
2. On the Network Analyst toolbar, click **Solve**. An alternative route is computed.

Part III: Finding the Closest Facility

In this section, you will find the closest three grocery stores to your house.

Preliminaries – Add a new dataframe

1. **Add** a new dataframe to your current document.
2. **Add** your network dataset to the new dataframe. **Click Yes** to add all the features associated with the network dataset.
3. **Add** the prepared shapefile: "SF_grocs_two_locs" that you downloaded to your lab8 folder.

Creating the Closest Facility analysis layer and Adding facilities

1. On the Network Analyst toolbar, click the Network Analyst drop-down menu and click **New Closest Facility.**
2. **Right-click** on **Facilities(0)** in the Network Analyst window and click on **Load Locations.**
3. Select **SF_grocs_two_locs** from the Load From drop-down list. Keep the default settings.
4. Click **OK**.

Adding an Incident (your home location)

1. Right click on **Incidents(0)** on the Network Analyst Window and select Find Address from the menu that appears.
2. Enter **5268 Diamond Heights Blvd**, **94131** in the text box. Click Find. **Right-click** on the first object found and click **Add as a Network Analyst Object.** This adds the located address as the place of the incident.
3. **Close** the Find dialog.
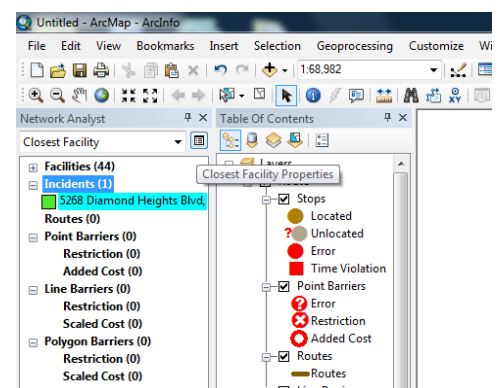
Setting the parameters for the analysis

1. Click on the **Closest Facility Properties Button** next to the Closest Facility on the Network Analyst Window to open the Layer Properties dialog.
2. In the Layer Properties dialog, click the **Analysis Settings** tab.
3. Click the Impedance drop-down arrow and click **Minutes.**
4. Set the "Default Cutoff Value" to **10**.

   ArcGIS will search for grocery stores that are within **10 Minutes** of your house. All other stores will be ignored.

5. Increase the Facilities to Find from 1 to **3.**

   ArcGIS will attempt to find the stores from your home, within the 10 Minute cutoff.

6. Under Travel From, keep **Incident to Facility.**
7. Select **Only at Dead Ends** from the U-Turns at junctions dropdown box.
8. Change **output shape type** to "**True Shape"**
9. **Check** the box labeled Ignore Invalid Locations.
10. **Click OK**

Run the process to identify the closest facility

1. Click the **Solve** button on the Network Analyst toolbar. The routes appear in the Map and in the Route Category on the Network Analyst Window. Note that three stores were found to be closest to your home within the 10 minute cutoff. You can optionally decrease the cutoff value, e.g. 5 minutes, and see how many stores are found. Note the distance of the routes and the paths taken compared to a straight line route.
2. Click the **Directions Window** button in the Network Analyst toolbar to generate directions for routes from each grocery store.

Part IV: Calculating Service Area and Creating An OD Matrix

In this section, you will create a series of polygons representing the distance that can be reached from a facility within a specified amount of time. These polygons are known as service area polygons. You will calculate 3-, 5-, and 10- minute service area polygons for the grocery stores in San Francisco. Additionally, you will create an Origin-Destination (OD) Cost Matrix for Safeway grocery stores to US Banks within a 5-minute drive time.

Preliminaries – Add a new dataframe

1. **Add** a new dataframe to your current document and name it **Service Area.**
2. **Add** your network dataset to the new dataframe. **Click Yes** to add all the features.
3. **Add** the USBank_Geocoded from the lab8 prepared files to the dataframe.
4. **Add** the SF_groc_safeway layer to the new dataframe.

Creating a Service Area analysis layer and Adding facilities

1. On the Network Analyst toolbar, click the Network Analyst drop-down menu and click **New Service Area.**
2. Right-click on **Facilities(0)** on the Network Analyst Window and select **Load Locations.** Select **SF_groc_safeway** from the Load From dropdown list. Click **OK.**

Setting up the parameters for the analysis

1. In the Network Analyst Window, click the **Service Area Properties** button to bring up the Layer Properties dialog box. Click on the **Analysis Settings** tab.

2. Click the Impedance dropdown list and select **Minutes**.
3. Type "**3 5 10**" in the Default breaks text box (Enter this as 3 5 10, the three numbers separated by a space, without quotes).
4. Under Direction, click **Away from facility.**
5. Click **Only at Dead Ends** from the U-Turns at junctions dropdown list.
6. **Check** the Ignore Invalid Locations checkbox.
7. In the same dialog box, **click** the **Polygon Generation** tab. Make sure that **Generate Polygons** is checked.
8. Click **Generalized** for Polygon Type. This results in faster analysis. Detailed polygons are more accurate but need more time to process.
9. **Un-check** the **Trim Polygons** option. This is a post-process that trims the outside polygon to remove spikes but takes longer to run.
10. Click **Not-overlapping** polygons per facility under Multiple Facilities Options. This results in individual polygons per facility that may not overlap. Since San Francisco is a small area, the polygons would all
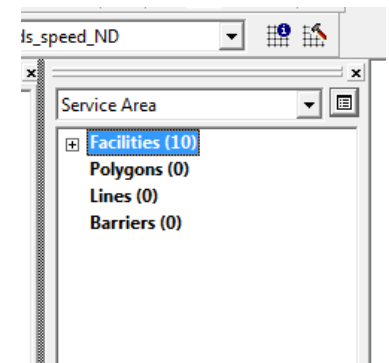
closely overlap. Therefore, in this case, not overlapping is selected. If the area is larger, you may want to try Overlapping first.

11. Click **Rings** for the Overlap type. This excludes areas of smaller breaks from the polygons of a bigger break.

12. In the same dialog box, **click** the **Line Generation** tab. Leave the box labeled Generate Lines **un-checked**. Click **OK** to save your settings.

Run the process to compute the Service Area

Now you will compute the service area based on the specified parameters, i.e. three service area polygons will be calculated for the Safeway stores, one at 3 minutes, one at 5 minutes, and another at 10 minutes. The direction of travel is set away from the facility, and U-turns are allowed at dead ends.

1. Click the **Solve** button on the Network Analyst toolbar. The service polygons appear on the map and on the Network Analyst Window (and are by default, set as transparent).

Creating the OD Cost Matrix analysis layer and Adding Origins/Destinations

Having limited resources but the use of GIS, you want to plan your grocery shopping around cashing your paycheck at the bank.

1. In the table of contents, **uncheck** the Service Area layer. On the Network Analyst toolbar, click **New OD Cost Matrix.**

2. Right-click on **Origins(0)** on the Network Analyst Window and click **Load Locations**.

3. Select **US_Bank_Geocoded** in the Load From drop-down box.

4. Click **OK.** The origins (banks) are now displayed on the map.

5. Right-click **Destinations(0)** on the Network Analyst Window and click **Load Locations.**

6. Choose **SF_grocs_safeway** from the drop-down list.

7. Under Location Analysis Properties, for the property "Name", click in the blank box in the field column and select "address" from the drop-down menu.

8. Click **OK.** The destinations (Safeways) are now displayed on the map.
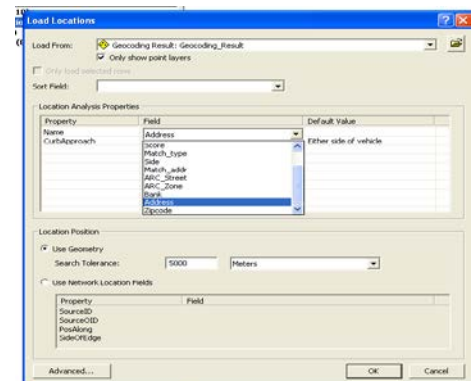
Setting up the parameters for the analysis

1. Click the **ODCostMatrix Properties** button next to ODCostMatrix on the Network Analyst Window.

2. Click the **Analysis Settings** tab. Click the Impedance drop-down list and click Minutes.

3. **Input 5** in the "Default Cutoff Value" drop-down list. This will create OD paths from the Banks to all Safeways that can be reached within 5 minutes.

4. Select **<All>** from the Destinations to Find drop-down list.

5. Select **Only at Dead Ends** from the Allow U-Turns dropdown list.

6. Select **Straight Line** from the Output Shape Type drop-down list.

7. **Check** Ignore Invalid Locations

8. Click **OK** to save the parameters**.**
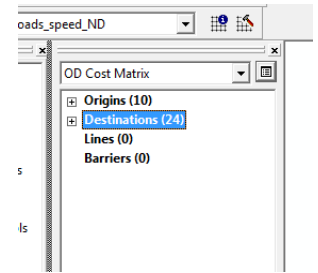
Run the process to create the OD cost matrix

1. Click the **Solve** button on the Network Analyst toolbar. The OD lines appear on the map. **Turn on** the Service Area Layer and display the Lines on top of the service areas. **Open** the attribute table of Lines to see the OD cost matrix. The OriginID column contains the IDs of the banks. The DestinationID features the IDs of the Safeways. The Destination Rank is a rank assigned to each destination (Safeways) that is served by the origin (Bank) based on total drive time. **What does this matrix tell you?**

Symbolizing the OD Matrix

1. Open the properties for the lines by double clicking on "Lines" in the table of contents

2. Go to the Symbology tab – quantities – graduated colors.

3. Change the value field to total minutes

4. Change the number of classes to 3 and click classify

5. In the manual classification screen, select 1.5 and 3 as the first two break values

6. Symbolize the newly created classes with Red, yellow and green to show the closest and furthest OD lines

Make a **3-part** layout to show the types of functions you can perform using Network Analyst. Recall, we

1. Found the best route and then recalculated the best route after adding a barrier

2. Found the closest facility to an incident

3. Calculated a service area and created an OD Cost Matrix

Add the bay area county from Lab 7.

Export your 3-part layout as a JPG.

Develop Mobile Data Collection Solutions

Mobile data collection refers to the utilization of existing information technology products such as phones, smartphones, and tablets (hardware), and a number of different possible programs (software), for compiling qualitative and quantitative information.

This approach has been proven to increase the speed and accuracy of data collection, service delivery effectiveness, and program staff performance.

Features of mobile data collection, such as decision support, form logic, and checklists, improve the quality of the data collected, while also ensuring adherence to data collection and care protocols.

Benefits of Mobile Data Collection

Speed and efficiency

Data can be ready for analysis in a significantly reduced amount of time since it is entered directly into an electronic form that can be exported into information management system. With the use of Internet technology, access may not only be real time but also global.

The speed with which actual interviews are conducted can be reduced, particularly for longer, more complex surveys. As the interviewer is automatically prompted with the correct questions to ask, they do not need to concentrate on skip patterns or spend time checking that each question has been asked and all required fields have been filled.

It allows better responsiveness to issues arising on the ground. This is especially important in cases where data collection is continuous, as rising or declining trends can be effectively identified and addressed while doing so is still possible.

Cost

Securing an acceptably small error rate with traditional survey methods often requires "double data entry," where two separate workers enter the same data and one supervisor identifies differences between the two. Eliminating this redundancy often adds up to significant cost reduction. The larger the size of the survey, the larger the savings from this activity become, as manual input requires a set amount of time per observation.

Additional savings can be made on printing costs, which can be significant as survey size and interview length increase (imagine a 30 page questionnaire administered to a sample of 1,000 respondents). As any and all

mistakes made in the survey can be corrected using a computer and updated in the collecting device used, the need for reprinting materials no longer applies.

Potential cost savings can also arise from the flexibility and customizability of mobile tools. For instance, imagine gathering data in an area where five different languages are spoken. With traditional pen and paper, estimates of language proportions (if such data is even available) must be taken into account while printing the questionnaires. To be on the safe side, some extra copies of each questionnaire will probably be necessary. With mobile tools, all languages can be downloaded on the collecting device and the data is input automatically in the same dataset category regardless of the language version.

From a program perspective, investment in a common platform that can be used in multiple projects can further spread costs and reduce the burden on a single study or round of data collection.

Data quality

The improvements to data quality from MDC tools stem from two distinct sources. Firstly, the survey development process allows for a large degree of customization (the extent depending on software, see section III), and the insertion of a number of quality checking tools. **Filter logic** can be utilized, which eliminates the possible confusion arising from questionnaire instructions. Secondly, as data does not need to be separately transcribed from paper to a computer, there is no possibility for data mis-entry during this phase.

Furthermore, MDC tools allow for validation of the data being entered during an interview. For example, questions requiring a numeric answer can be forced to only accept a numeric answer, which can also be forced within a range (such as between 1-99 for age) to reduce the chance of input errors. Answers to questions, such as date of birth and age, can be crossreferenced to validate each other and prompt the enumerator to clarify any inconsistencies. Forms that are not filled can be coded to not allow submission.

In addition to the questions programmed into the survey, some MDC tools automatically collect metadata such as timestamps. These allow for the survey supervisor to monitor interview start and end times to ensure that they match with expected interviewing behavior, and thus helps keep enumerators more accountable.

Customizability and added tools

Though the exact extent will depend on the software choice (see section III), MDC tools generally offer a wide range of customization possibilities. These include but are not limited to different types of questions (all that can be asked using pen and paper), programming in mathematical equations that can transform data according to pre-set functions, the utilization of GPS technology, and the addition of multimedia features such as graphics, audio, and video.

Examples of these added features include the ability to use audio, graphics or video to improve comprehension of survey topics, and using a camera function to capture a photo of a school while also capturing its location via GPS to create a map that allows users to see images as well as location of survey sites.

Control and flexibility of fieldwork

Faster access to data improves not only analysis and program responsiveness, but improves staff ability to react to unexpected changes. For example, since surveys in the MDC devices can be updated (given even limited Internet connection), it is possible to fix small issues in questionnaire design such as typos, badly phrased questions, or filter logic. For longer studies, further questions can even be added to investigate interesting findings. The costs for such responses are negligible, as lead staff or field supervisors can make any alterations on their computer.

Effective supervision of enumerators is also a common problem in fieldwork, and can be substantially mitigated by the use of MDC tools. In addition to the automatic metadata and timestamps, the additional feature of GPS (varies by hardware selection) allows tracking the enumerator's route to further validate that the supervisor's directions were followed. This enables staff to give informed feedback to the enumerator teams and improve their performance.

User friendly

Using the MDC technologies also empowers enumerators and builds their capacity for utilizing information technology. The possession of MDC devices can even enhance credibility of the data collection or research and staff in the eyes of the local community. As the programming guides the enumerator through the interview, they are able to spend less effort in making sure that all the correct questions are asked, and are thus able to better concentrate on other relevant aspects of interviewing and maintaining good rapport. The more complicated the conditions for asking a question (for example, only ask if three preceding questions have a certain answer), the more evident this benefit will be.

An enumerator collecting data with a mobile device does not need to carry around large numbers of survey material; all that is required is the device itself and perhaps a charger or backup battery. This reduces the hassle and burden of traditional pen and paper methods.

On the other hand, it must be ensured that the possession of the data collection device does not endanger the enumerator when operating in highly insecure areas. MDC has been conducted successfully in a number of potentially risky locations due to hardware choices appropriate for the local context and clear safety and equipment management protocols.

Planning for Data Collection

Consider these 3 questions before Collect data:

- ✓ What information do you actually need?
- ✓ What information is available?
- ✓ What information will be useful?

Mobile Data Collection Process

1. A paper survey is drafted
2. The paper survey is converted into a digital format
3. The digital survey is transmitted to a mobile device
4. Local enumerators conduct the survey on mobile devices
5. Survey data is uploaded
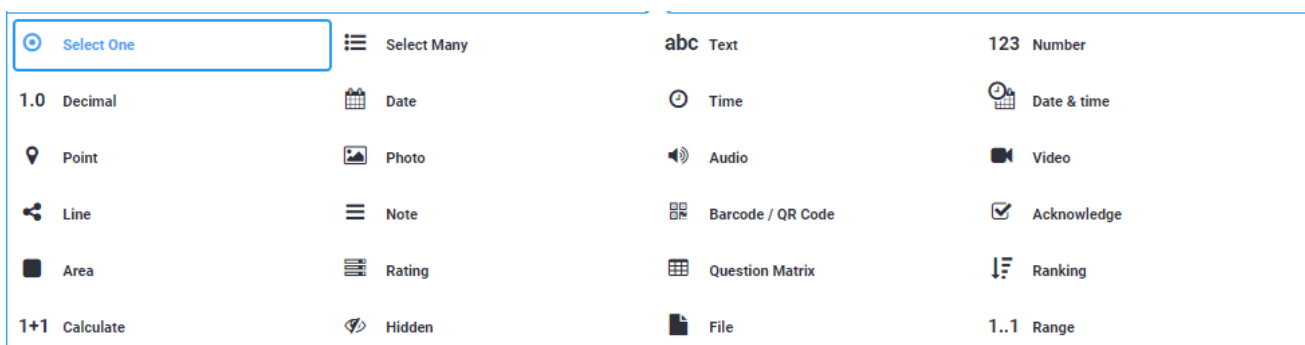6. The collected data is immediately available for analysis

What is an XLSForm?

XLSForm is a form standard created to help simplify the authoring of forms in Excel. Authoring is done in a human readable format using a familiar tool that almost everyone knows - **Excel**.

XLSForms are simple to get started with but allow for the authoring of complex forms by someone familiar with the syntax described below.

The **XLSForm** is then converted to an **ODK XForm**, a popular open form standard, that allows you to author a form with complex functionality like skip logic in a consistent way across a number of web and mobile data collection platforms.

Questions types

⊙ Select One	☰ Select Many	abc Text	123 Number
1.0 Decimal	📅 Date	⊙ Time	Date & time
● Point	🖼 Photo	🔊 Audio	🎥 Video
◁ Line	☰ Note	Barcode / QR Code	☑ Acknowledge
■ Area	Rating	▦ Question Matrix	↓F Ranking
1+1 Calculate	Hidden	📄 File	1..1 Range

select_one [choices] [or_other]	User can choose one of several choices
select_multiple [choices] [or_other]	User can choose one or more of several choices
select_one_from_file [file]	User can choose one of several choices from an external file
select_multiple_from_file [file]	User can choose one or more of several choices from an external file
select_one_external	
text	User can enter a text response
integer	User can enter an integer
decimal	User can enter a decimal number
date	User can enter a date
time	User can enter a time of day
datetime	User can enter date and time together
geopoint	User can record a GPS location
image	User can take or attach a picture
audio	User can record or attach audio
video	User can record or attach video
note	User is shown a note (no response possible)
barcode	User can enter a barcode by scanning it
acknowledge	User is asked to confirm or acknowledge something
calculate	A mathematical expression of existing values that can be used by other questions (no user input)
geotrace	User can record a line or polyline or multiple geo points
geoshape	User can record a polygon of multiple geo points - the last point is the same as the first point

Common Mobile-based Data Collection Platforms

Closed Source

Survey123 for ArcGIS (ESRI) (Android, iOS, Web)

Open Source

KoboToolbox (Android, Web)

Open Data Kit (Android, Web)

What is KoboToolbox?

KoBoToolbox is a suite of tools for field data collection for use in challenging environments.

The software is free and open source.

Most of the users are people working in humanitarian crises, as well as aid professionals and researchers working in developing countries such as students, journalists, government staff, humanitarian and development sectors.