

Lab 6: Raster Modeling with ArcView's Spatial Analyst

Overview

The purpose of this lab exercise is to introduce spatial analysis methods using raster models of geospatial phenomena. Thus far this semester we have represented spatial phenomena as discrete features modeled in the GIS as points, lines, or polygons—i.e., so-called 'vector' models of geospatial features. Sometimes it is useful to think of spatial phenomena as 'fields' such as temperature, wind velocity, or elevation. The spatial variation of these 'fields' can be modeled in various various ways including contour lines and raster grid cells. In this lab exercise, we shall focus on **raster** models and examine ArcView's '**Spatial Analyst**' extension. We shall use raster models to create a housing value 'surface' for a portion of Boston. To create the 'surface' we will explore ArcView's tools for converting vector data sets into raster data sets—in particular, we will 'rasterize' the **1986-1995 housing sales data** for Boston and the **1990 Census data** for Boston block groups. (For those of you who took 11.520 in the Fall, this lab is very similar to what we did to interpolate housing values for Cambridge. The main difference is that, this time, the sales data are in Oracle and they involve multiple-sale and other complications that require data manipulation before being pulled into ArcView.)

A housing value 'surface' for Boston would show the high and low value neighborhoods much like an elevation map shows height. The computation needed to do such interpolations involve lots of proximity-dependent calculations that are much easier using a so-called 'raster' data model instead of the vector model that we have been using. Thus far, we have represented spatial features—such as Boston block group polygons—by the sequence of boundary points that need to be connected to enclose the border of each spatial object—for example, the contiguous collection of city blocks that make up each Census block group. A raster model would overlay a grid (of fixed cell size) over all of Boston and then assign a numeric value (such as the block group median housing value or the block group ID) to each grid cell depending upon, say, which block group contained the center of the grid cell. Depending upon the grid cell size that is chosen, such a raster model can be convenient but coarse-grained with jagged boundaries, or fine-grained but overwhelming in the number of cells that must be encoded.

In this exercise, we only have time for a few of the many types of spatial analyses that are possible using rasterized datasets. Remember that our immediate goal is to use the Boston sales data **to generate a housing-value 'surface' for Boston**. We'll do this by 'rasterizing' the sales data and then taking advantage of the regular grid structure in the raster model so that we can easily do the computations that let us smooth out and interpolate the housing values reported in the housing sales datasets.

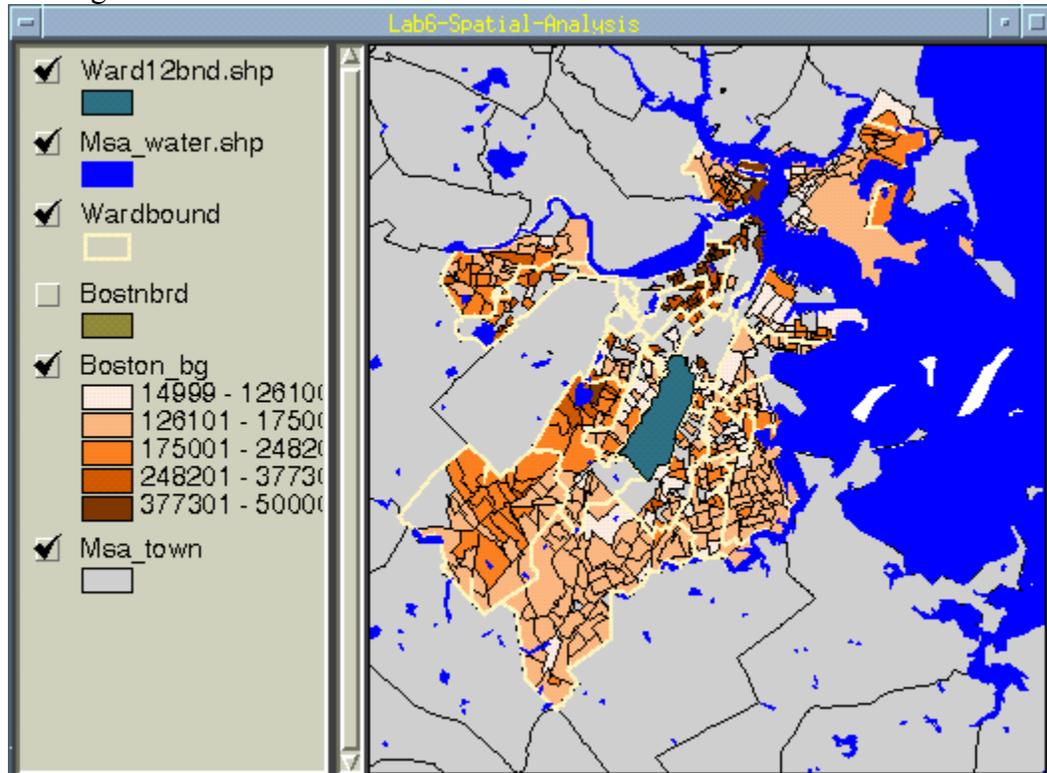
(1) Spatial Analyst Setup

Start with a brand new ArcView project, create a new view, and add several Boston themes (stored in the **/mit/crlmaps/meters83/massmaps/boston** directory) as indicated below. If you have a CRLspace directory, you should also set your working directory to be a subdirectory of your CRL space. Hence, you should:

- Activate the **Project** window
- Select '**Properties**' from the **Project** menu
- Set the working directory to **/mit/crlspace1/YourAthenaID/lab6**
(*Note: you may need to 'attach' the locker and/or create the sub-directory first.*)
- Activate a View window or open a new one
- Add the following themes to the view (all from the **/mit/crlmaps/meters83/massmaps/boston** directory which, on the PC is also know as: **\\crypt\crlmaps\meters83\massmaps\boston**). Note that some of these are shapefiles and some are ArcInfo 'coverages' where you can click the coverage icon in the dialogue box to get the expanded list of coverage types to be sure that you load the 'polygon' layer.
 - **bostnbrd (polygon)**
 - **msa_town (polygon)**
 - **msa_water.shp**
 - **wardbound (polygon)**
- Also add a polygon and a point theme for Boston's Ward 12 parcels and Boston's 1990 Census block-groups:
 - **/mit/crlclass/11.521/data/boston_bg (polygon)**
 - **/mit/crlclass/11.521/data/ward12bnd.shp**
- Order the themes, top to bottom, as follows:
 - **Ward12bnd.shp**
 - **Msa_water.shp**
 - **Wardbound**
 - **Bostnbrd**
 - **Boston_bg**
 - **Msa_town**

Highlight the **Wardbound** theme and press the 'Zoom to Active Theme(s)' button  so that the map shows all of Boston. Adjust the colors and transparency of their symbols so that the ocean is blue with no border, the town boundaries around Boston are visible and you can see Boston's Ward 12 as the area from Franklin part up through Lower Roxbury to Dudley Square. Create a thematic map of median housing value by shading the **Boston_bg** theme based on the 1990 **medhhval** values. (See [the data dictionary for this](#)

[layer](#) on the MassGIS site for the meaning of all the variables in the **Boston_bg** attribute table—but you will only need **medhhval**). Be sure to exclude block groups with $\text{medhhval} = 0$ by setting '0' to be the null value. At this point, your ArcView window should look something like this:



Setting 'Spatial Analyst' Properties

ArcView's raster manipulation tools are bundled with its **Spatial Analyst** extension. It's a big bundle so let's open ArcView's help system first to find out more about the tools.

Follow this sequence of clicks to find the relevant spatial analyst topics: **Help Topics > Contents > Extensions > Spatial Analyst**. During the exercise, you'll find these online help pages helpful in clarifying the choices and reasoning behind a number of the steps that we will explore. Be sure, at some point, to take a look at the **overview** and **performing analysis** sections.

The Spatial Analyst module is an extension, so it must be loaded into ArcView separately. (Note: this is because raster modeling capabilities do not come standard with ArcView. The Spatial Analyst extension must be purchased separately, and actually costs more than ArcView alone.)

- Select '**Extensions**' from the '**File**' menu
- Checkmark '**Spatial Analyst**' and **OK** the selection (Note: if you only highlight 'Spatial Analyst' without checking the box, then clicking '**OK**' will **not** load the extension.)

Once the Spatial Analyst extension is loaded, a new main-menu heading called *Analysis* will be available whenever a **View** window is active. Make sure that you go to **View > Properties** and set the map units to meters and the distance units to miles. If the map units are not set, Spatial Analyst will not know what units to use when performing an analysis.

Before building and using raster datasets, we should set the grid cell sizes and the spatial extent of our grid, and we need to identify the 'no data' regions outside of Ward 12 that we wish to 'mask' off. Let's begin by specifying a grid cell size of 100 meters and an analysis extent covering all of Boston's Ward 12. To do this, select 'Properties' from the 'Analysis' menu and specify:

| Item | Value |
|--------------------|------------------------------|
| Analysis Extent | Same As Ward12bnd.shp |
| Analysis Cell Size | As Specified Below |
| Analysis Cell Size | 100 |
| Number of Rows | 40 |
| Number of Columns | 25 |
| Analysis Mask | No Mask Set |

Note: Hitting 'Enter' (or 'Return') after typing 100 for the cell size will cause the row/column numbers to be recomputed. When all the values are properly set, click '**OK**'.

Now we are ready to build a 100 meter raster grid model for Ward 12. With your **View** window open, convert the **Ward12bnd.shp** to a grid theme using these steps and parameter settings:

- Highlight the **Ward12bnd.shp** theme and select '**Convert to Grid**' from the **Theme** menu
- Specify the subdirectory and grid file name. Use the name: **boswd12gd** (this is the name for the saved grid files and you should put them in your working directory that you setup earlier.)
- If you had not already set the Analysis Properties, you would have been prompted at this point to set them (as we did above) to:

| Item | Value |
|-----------------------|------------------------------|
| Output Grid Extent | Same As Ward12bnd.shp |
| Output Grid Cell Size | 100 |

| | |
|-------------------|----|
| Number of Rows | 40 |
| Number of Columns | 25 |

- A series of dialog boxes will now come up. Respond as follows:
 - *Pick 'conversion field' to set the cell values*
Use **Wd12bnd_id** (we want a single value for the entire ward at this point since we want to tag all the cells in Ward 12 for inclusion in the study)
 - *Join feature attributes to grid?*
No (we don't want to pull over any more data at this point)
 - *Add grid as theme to the View?*
Yes, do add **Boswd12gd** to the view

If successful, the **Boswd12gd** theme will be added to the View legend. Turn it on and notice that the shading covers all the grid cells whose center point falls inside of the spatial extent of the **Ward12bnd** theme. The cell value associated with the grid cells is 12 (if you picked wd12bnd_id for the cell values).

At this point we don't need the old **Ward12bnd.shp** theme coverage any longer. We used it to set the spatial extent for our grid work, but that setting is now retained. To reduce clutter in your View window, you can delete the **Ward12bnd.shp** theme from the View legend. (Later, you can also use **File > Manage Data Sources** to permanently erase some of the temporary shapefiles and grid themes that you create during this exercise.)

(2) Interpolating Housing Values Using SALES Data

This part of the lab will demonstrate some techniques for filling in missing values in your data using interpolation methods. In this case, we will explore different ways to estimate housing values for Boston's Ward 12. Keep in mind that there is no perfect way to determine the value of a property.

A city assessor's database of all properties in the city would generally be considered a good estimate of housing values because the data set is complete and maintained by an agency which has strong motivation to keep it accurate. This database does have drawbacks, though. It is updated sporadically, people lobby for the lowest assessment possible for their property, and it's values often lag behind market values by many years.

Recent sales are another way to get at the question. On the one hand, their numbers are believable because the price results out of an informed negotiation between a buyer and a seller that results in the true of the property being revealed (if you are a believer in the economic market-clearing model). However, the accuracy of such data sets are susceptible to short-lived boom or bust trends and, since individual houses (and lots) might be bigger or smaller than those typical of their neighborhood, individual sale prices may or may not be representative of housing prices in their neighborhood.

Alternatively, the census presents us with yet another estimate of housing value—the median housing values aggregated to the block group level. However, this data set is also vulnerable to criticism from many angles. The numbers are self-reported, only a sample of the population is asked to report, and the data are aggregated up to the block group level. The benefit of census data is that they are cheaply available and they cover the entire country.

(2.1) Pulling Ward 12 Data from the ASSESS_DATA.SALES Table

To begin exploring some of these ideas, we will use an Oracle table called SALES containing information about the sale price of residential Boston parcels that sold between 1986 and 1995. This table is owned by the user ASSESS_DATA. You need to specify the fully-qualified table name, ASSESS_DATA.SALES, when querying from this table, because you do not own this table and you already have a synonym called SALES that refers to the table PARCELS_DATA.SALES. From your Oracle window, take a look at the ASSESS_DATA.SALES table:

```
set pause '-- Press [Return] to continue--'
set pause on
set pagesize 24
set linesize 120
```

```
select * from assess_data.sales;
```

| SALE_ID | PARCEL_ID | SALEDATE | STATECLASS | PRICE | SELLER | BUYER |
|------------|---------------|-----------|------------|--------|----------------------|----------------------|
| W1 0100003 | 1 0100003000 | 01-JUL-87 | 105 | 135000 | DALY PAUL ET AL | PIAZZA DAVID |
| D3 0100003 | 2 0100003000 | 17-FEB-89 | 105 | 85000 | PIAZZA DAVID M | ONE HUNDRED REALTY T |
| W1 0100004 | 3 0100004000 | 09-JUN-86 | 105 | 125000 | CAPONE ROBERT ET AL | EMERSON JEFFREY |
| D3 0100004 | 4 0100004000 | 25-OCT-90 | 105 | 115000 | EMERSON JEFFREY F | DILLON KELLEY A |
| A1 0100005 | 5 0100005000 | 01-MAR-90 | 105 | 90000 | DIGIROLAMO JOSPH G | DIGIROLAMO JOHN F ET |
| D3 0100012 | 6 0100012000 | 03-OCT-90 | 105 | 115000 | STRAW BRUCE E JR ET | SCIACCA ANTONIO ET A |
| W1 0100013 | 7 0100013000 | 28-MAR-88 | 105 | 150000 | CARCO MARIE | MELIA GERARD ET AL |
| W1 0100014 | 8 0100014000 | 15-JUL-94 | 105 | 114000 | TRIPPI GIACOMO | TRIPPE VIRGINIA |
| D3 0100019 | 9 0100019000 | 01-NOV-89 | 105 | 125500 | TRICOMI JOSEPHINE | BROWN STEPHEN P |
| W2 0100022 | 10 0100022000 | 30-AUG-91 | 105 | 117000 | NIGRO ANTONIO J ET A | CAHILL STEPHEN F ET |
| A1 0100024 | 11 0100024000 | 20-NOV-87 | 105 | 60000 | IANNACCONE MARY | SHEA RAFFAELA |
| D2 0100028 | 12 0100028000 | 12-DEC-88 | 104 | 95000 | FRANCIS MARK S ET AL | MELE ANTHONY |
| D3 0100029 | 13 0100029000 | 01-OCT-91 | 105 | 80000 | BONO ANTHONY ET AL | STEWART ROBERT ET AL |
| A1 0100039 | 14 0100039000 | 13-APR-94 | 105 | 47800 | DAPOLITO ANTHONY M E | DAMATO FILIPPO |
| X1 0100043 | 15 0100043000 | 09-APR-90 | 105 | 165000 | DAVOLIO MICHAEL ET A | SCIORTINO SALVATORE |
| E1 0100043 | 16 0100043000 | 14-FEB-94 | 105 | 85000 | SCIORTINO SALVATORE | MASSACHUSETTS HOUSIN |
| L1 0100043 | 17 0100043000 | 23-NOV-94 | 105 | 49500 | MASSACHUSETTS HOUSIN | MORAN DESERRAE J |

```

      18 0100046000 28-FEB-90      104      47500 MATTHEWS ALICE      BICKFORD WILLIAM J E
D3 0100046 N
      19 0100048000 30-OCT-86      105      159000 CARBONE CHRIS ET AL  ROGERS ROBERT ET AL
W1 0100048 N
      20 0100048000 22-APR-91      105      35000 GARDEN MANAGEMENT CO TWO 06 PRINCETON RLT
L1 0100048 N
      21 0100048000 27-FEB-95      105      40000 TWO 06 PRINCETON RLT WARSHAW  GEORGE
0100048 N
-- Press [Return] to continue --
...
...
...
[ 68134 rows total ]

```

The first two characters of the **parcel_id** (and the **base_id**) are the Ward. Let's see how many Ward 12 sales are included in the **ASSESS_DATA.SALES** table:

```

select count(*), count(distinct parcel_id), count(distinct base_id)
  from assess_data.sales
 where substr(parcel_id, 1, 2) = '12';

```

```

COUNT(*) COUNT(DISTINCTPARCEL_ID) COUNT(DISTINCTBASE_ID)
-----
      853              557              433

```

Hmm, there are 853 sales involving 557 distinct parcel numbers and 433 unique **base_id** values. Condos within the same building/development share the same **base_id** (and parcel footprint) but have different parcel numbers (the last 3 digits differ).

Unfortunately, split parcels can also retain the same **base_id**—but they have different footprints as well as differing last three digits. For the purposes of this exercise, we will ignore complications involving split parcels but will address complications associated with condo sales and multiple sales of individual parcels.

To begin, let's construct a SQL query that pulls into ArcView the IDs and (average) sale price of all Ward 12 parcels:

```

create view holdsales1 as
  select base_id, avg(price) saleprice, count(*) sales
  from assess_data.sales
 where substr(base_id,1,2) = '12'
 group by base_id;

```

There are 433 rows in this table—including one **base_id** that showed up 72 times in the **SALES** table! From ArcView, we can connect to Oracle using the Properties/SQL-Connect menu option. (Remember to 'connect' to Oracle by making the *project window* active and then selecting **Project > SQL Connect**. The project window is the one labeled **Untitled** until you save your project, then **yourprojectfilename.apr** after you do save it.) In the **SQL Connect** dialog box, you will need to press the 'Connect' button, then enter your Oracle username and password in the format *YourAthenaID/password*, where *password* is your Oracle password). Make the connection and pull this **holdsales1** table into ArcView.

(2.2) Mapping Ward 12 Parcels

There are 135000+ parcels in Boston but we've saved a smaller shapefile for just the 2,592 in Ward 12. Add these Ward 12 parcels as a theme in your ArcView window:

/mit/crlclass/11.521/data/bospar96wd12.shp

You'll see that the attribute table contains both a **parcel_id** and a **base_id** column. We could use **base_id** to join the **holdsales1** table to 'Attributes of Bospar96wd12.shp' and then do a thematic map based on sales price. (Try it later if you have time.) But many of the parcels are so small and the number of sales so many that the resulting thematic map of housing prices won't be too readable. In addition, the unsold parcels won't be shaded at all.

We'd like to use the SALES data as an indication of housing value in the area surrounding the sale. One way to do such an interpolation is to compute a housing value for any particular location that is some type of average of nearby sales prices. By rasterizing the space within Ward 12, we can identify a finite and manageable number of places to consider (each grid cell) and then we can use the capacity of the GIS to know which sales are 'nearby' to help us compute a useful average. (We don't **have** to rasterize the Ward. We could consider every parcel and set its value to be a weighted average of its sales—if any—and its neighbors. But that requires a lot more geometry calculations so analysts often settle for a quicker raster analysis as illustrated next.)

Before doing this interpolation, we must join the sales data in **holdsales1** to the mappable parcel data. We could use **bospar96wd12.shp** for this purpose but it will be easier to do the analysis if we join the sales data to a 'point' coverage rather than to a polygon coverage. Since we'll only have time today to show one of the simpler interpolation methods, we'll do that. The staff has already created a 'point' coverage for Ward 12 parcels whereby each **parcel_id** (and **base_id**) is associated with a 'point' lying (more or less) in the middle of each parcel. Add this coverage to your ArcView window:

/mit/crlclass/11.521/data/bospar96wd12pt.shp

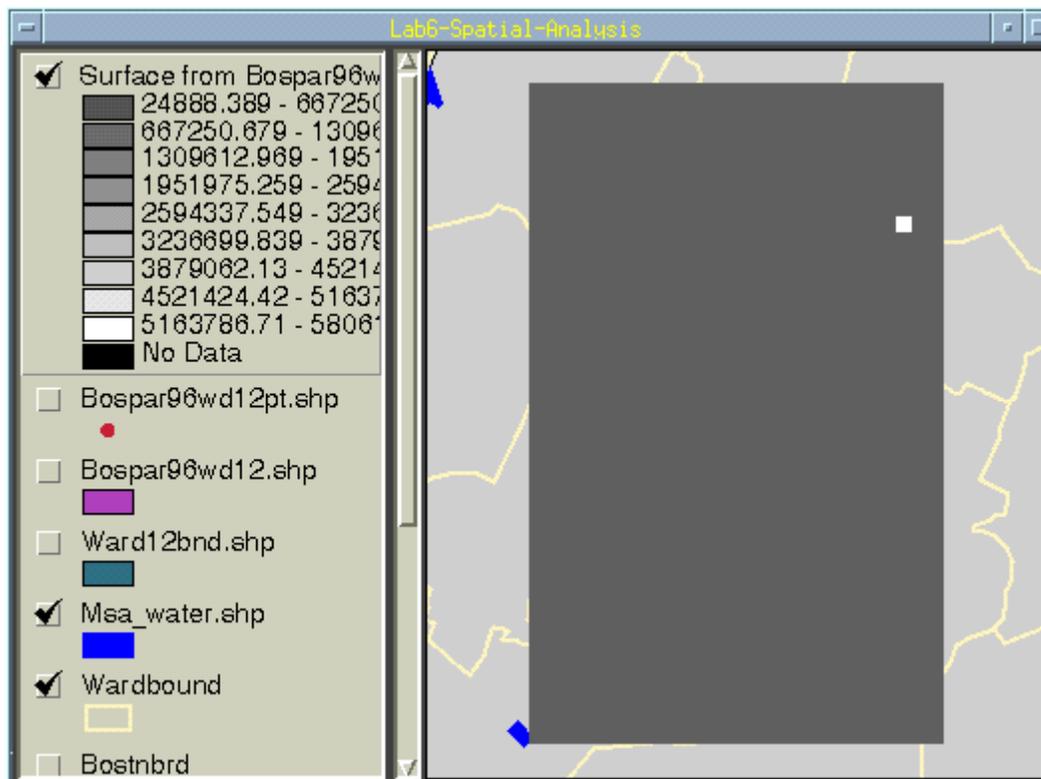
Now, join the **holdsales1** to the '**Attributes of Bospar96wd12pt.shp**' table using the **base_id** field. You should find 468 matches—a few dozen more than the 433 rows in **holdsales1** because of some split-parcels in Ward 12 (a problem that we will ignore for this exercise).

(2.3) Interpolating Housing Values from Sales Price Data

We are now ready to do the interpolation via the averaging approach that we discussed above.

- Highlight **Bospar96wd12pt.shp** and press the 'Zoom to Active Theme(s)' button  so Ward 12 fills your view window.
- Use the 'hammer' Query tool  to select all parcels that sold—that is, had a **saleprice > 0**. (If you don't, the interpolation will try to include all the parcels with null saleprices and probably produce errors as well as meaningless results.)

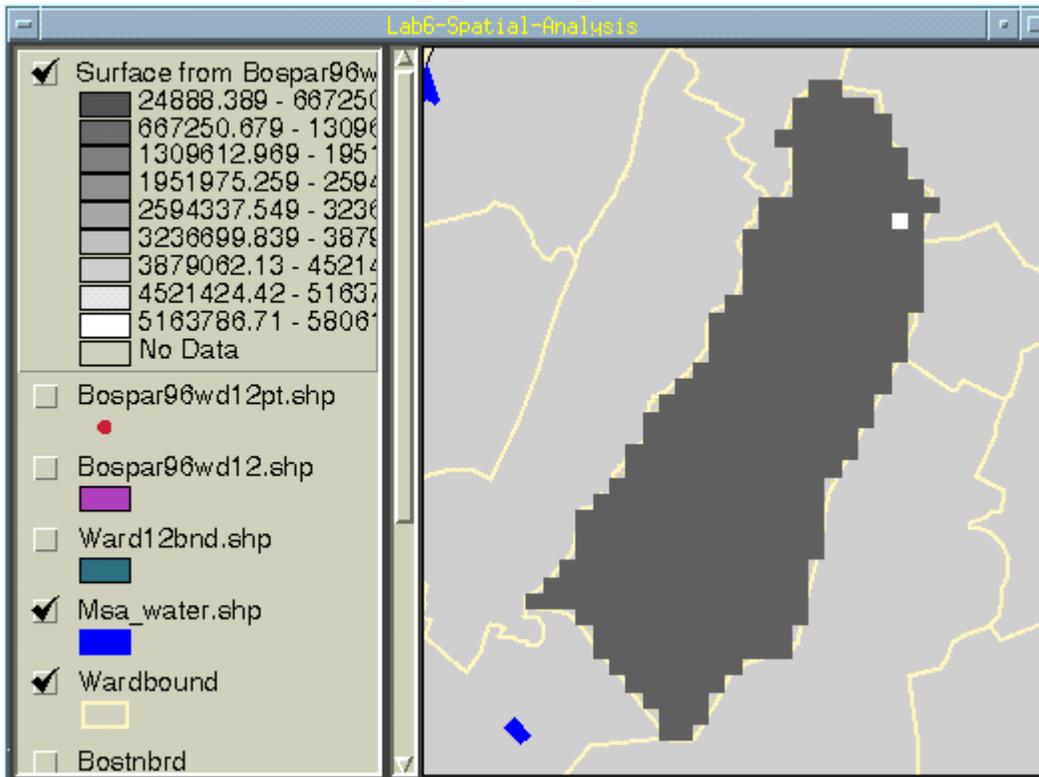
- Now, select '**Interpolate Grid**' from the '**Surface**' menu. Use the default 'Method' (which is a weighted inverse distance method, IDW).
- Use **Saleprice** as your **Z Value Field**, and the defaults for 'Number of Neighbors' (12), 'Power' (2), and 'No barriers'
- Click OK
- The grid theme that is created is quite strange—it fills the entire rectangle containing Ward 12 and looks something like this:



The interpolated surface is shown thematically by shading each cell dark or light depending upon whether that cell is estimated to have a lower housing value (darker shades) or higher housing value (lighter shades). Based on the parameters we set, the cell value is an inverse-distance weighted average of the 12 closest sales. Since the power factor was set to the default (2), the weights are proportional to the square of the distance. This interpolation heuristic seems reasonable, but the surface extends far beyond the Ward 12 borders (all the way to the rectangular bounding box that covers Ward 12). This problem is easily fixed by setting the 'mask' to be used when interpolating. We can prevent the interpolation from computing values outside of the Ward 12 boundary by 'masking' off those cells that fall outside. Do this by adding a mask to the Analysis Properties:

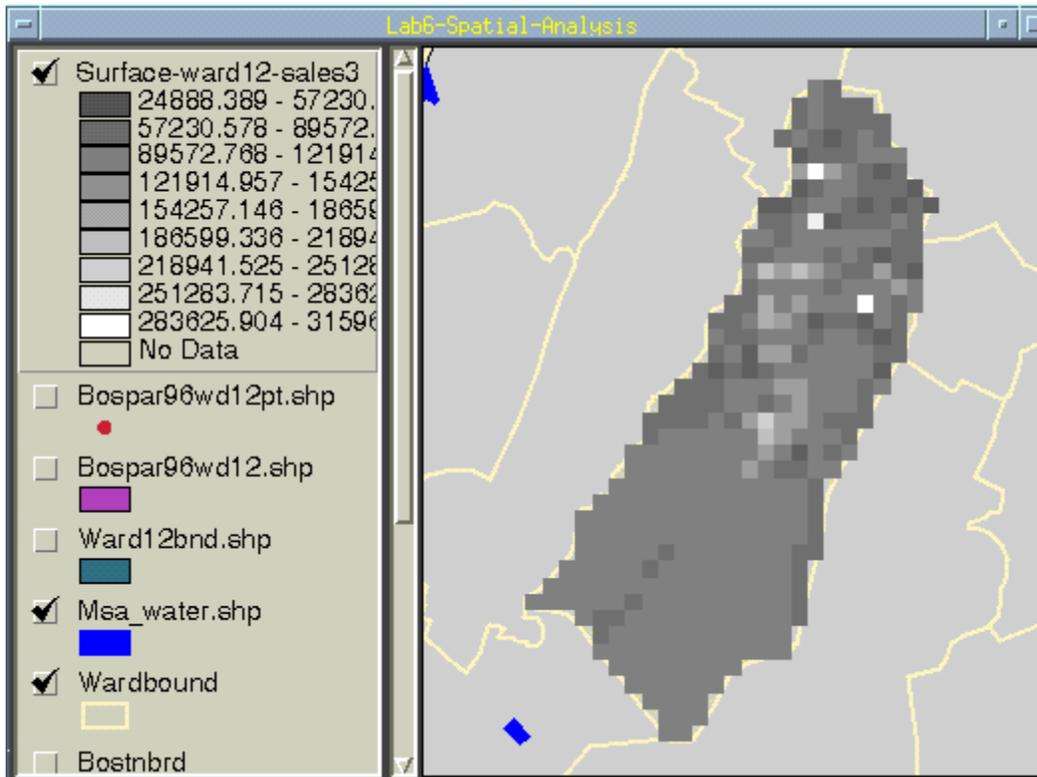
- Reopen the **Analysis > Properties** dialog box and set the **Analysis Mask** to be **Soswd12gd**—the grid that we computed earlier from the **Ward12bnd.shp** shapefile.

With this analysis mask set, interpolate the **Saleprice** values in **Bospar96wd12pt.shp** once again and you'll get a surface that looks like this (after making the 'no data' case transparent):



All the values inside Ward 12 are the same as before, but the cells outside Ward 12 are 'masked off' and set to be transparent. Delete the unmasked surface, and use **Theme > Properties** to rename the masked surface to be 'Surface-ward12-sales2'.

Close inspection indicates that the surface is severely distorted with almost all the cells in the lowest-priced shading category—which goes up to \$600,000! The one white cell contains the one parcel in Ward 12 that sold for more than \$675,000 and that parcel sold (once) for \$6,592,010. If we exclude just that one parcel before interpolating the value surface, the result looks like this:



This interpolated surface begins to look more reasonable, although it is still likely to be quite sensitive to outliers—especially since we have done nothing to control for the size of the housing units that are sold. Closer inspection of the Oracle data dictionaries associated with the Boston assessing data (see the bottom of the [CRL metadata page](#)) indicates that the SALES table contains a **stateclass** field that indicates the parcel's land use type as a three-digit code. A few lookup tables are also of interest. The USECAT table breaks the detailed stateclass code into 'categoryid' ranges such as RP = 'residential property' and RA = 'apartment property', and the STCLASS table relates each stateclass code to this two-character categoryid and to another abbreviated 'landuse' code. Also, a SALECODE table explains the codes that indicate which sales were 'arms length' and which were **not** a good indication of market value. (The tables USECAT, STCLASS, and SALECODE are also owned by the user ASSESS_DATA. As such, it is a good idea to fully qualify the table name with the owner name (e.g., ASSESS_DATA.SALECODE) as shown below. These tables have **public synonyms** defined for them, meaning that any user can query from them without qualifying the table name *provided* the user does not have a **private** object that conflicts. Actually, ASSESS_DATA.SALES has a public synonym defined for it too, but it conflicts with the **private** synonym already defined in your schema, so the private synonym takes precedence. You can write queries on the data dictionary 'view' that is called ALL_SYNONYMS in order to see what's going on.)

A quick look at the breakdown of sales codes indicates that about half of the Ward 12 sales were non-market value sales to/from a financial institution—probably due to foreclosure and only the SX, W1, W2, and X1 sales are likely to be representative of fair market value.

```

select s.salecode, substr(descript,1,55) description, count(*)
  from assess_data.sales s, assess_data.salecode c
  where substr(base_id, 1, 2) = '12'
     and s.salecode = c.salecode
group by s.salecode, substr(descript, 1, 55)
  having count(*) > 10
order by salecode, substr(descript, 1, 55);

```

| SA DESCRIPTION | COUNT (*) |
|--|-----------|
| A1 Sale among members of the same family or relatives | 52 |
| D2 Property significantly improved after assessment but be | 15 |
| D3 Sale of improved property at an atypical price; "specul | 53 |
| E1 Sale to or from federal, state, or local government | 17 |
| K1 Sale to or from a charity, educational, or religious or | 14 |
| L1 Sale to or from a financial institution not reflecting | 219 |
| MW Multi-parcel sale, not personally verified | 20 |
| N3 Multi-parcel sale conveyed in one deed, not fair market | 20 |
| N9 Sale price of unit reflects a tenant cash discount | 11 |
| SX Auction sale; verified as arm's length | 30 |
| W1 Price within 10% of most current assessment | 89 |
| W2 Consistent with inspection and market, sale not verifie | 20 |
| X1 Field inspected by staff, sale personally verified by s | 147 |

13 rows selected.

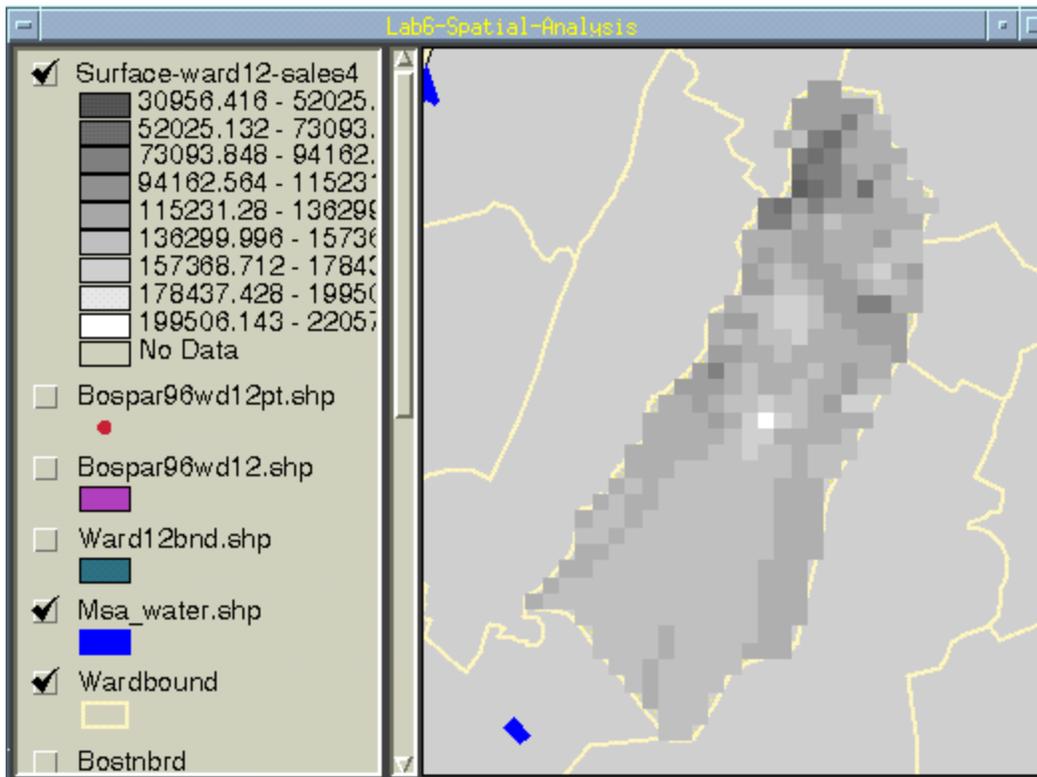
With this in mind, let's create a new view that restricts the selection to fair-market sales of condos and 1-4 family units:

```

create view holdsales2 as
  select base_id, landuse, avg(price) saleprice, count(*) sales
  from assess_data.sales s, assess_data.stclass c
  where substr(base_id,1,2) = '12'
     and s.stateclass = c.stateclass
     and (landuse in ('R1', 'R2', 'R3', 'R4') or
          s.stateclass = 102)
     and (salecode in ('SX', 'W1', 'W2', 'X1'))
group by base_id, landuse;

```

Let's bring this view, with 210 rows, over to ArcView as **holdsales2** and do another interpolation. This time, the interpolated surface looks like this:



This time, the interpolated values are much less skewed with the grid cells ranging from 31K—about 200K. The 'low' and 'high' value parts of Ward 12 are quite different than before. Do you have faith in these results? Remember the `block_id` that appeared 72 times in the SALES table. Our analysis counted it as one sale. Is it a split-parcel (with many units built on a sub-division) or does it represent multiple sales of condos in the same complex?

If you were doing this analysis for real, you would want to iron out your methodology on a small part of the City (such as Ward 12) before applying the analysis to the entire city. In such a case, you might want to adjust the grid cell size and, perhaps, adjust the mask so that it excluded parts of the city that weren't really residential neighborhoods. Picking appropriate cell sizes and masks can be time consuming and is as much an art as a science. There are also many more sophisticated statistic methods for interpolating and estimating value surfaces. (S*Plus is one package that runs on Athena machines and supports spatial regression, kriging and other techniques that explicitly account for spatial interactions).

We don't have time to refine the analysis further (for example to examine multiple sale issues, split parcels, and price inflation during the 1986-1995 period), or to compare alternative interpolation models. But this exercise should be enough to give you a sense of when and why it is often important to have more data manipulation capabilities than ArcView provides, and how the ArcView/Oracle link matters as you begin to work with larger datasets. You may also wish to review last Fall's raster modeling exercise, lab7 of 11.520 to see more examples and discussion of interpolation and 'map algebra' modeling.

Also, think about this interpolation approach versus using a thematic map of block group data or a nearest-neighbor smoothed grid cell map of either block group data or individual parcel assessments.

Analyzing 'real world' datasets—especially disaggregated administrative datasets such as assessment and sales data—often requires the complexities and subtleties involved in this exercise. What should a planner know in order to get past the simple (and wrong) first or second round of interpolation? How can you tell if someone you hire knows how to do disaggregated spatial analyses that are meaningful? How can you institutionalize the data processing and modeling so that planners can routinely tap these administrative datasets for planning purposes with confidence and reliability? (Development of the 'salecodes' used to tag market-value and non-market-value sales is one example of how data quality can be distinguished. However, given Boston's parcel numbering schemes and limited cross-reference tables, the handling of condo sales and parcel subdivisions remains complex and prone to analysis errors.)

(3) Lab Assignment

Most of the questions suggested above are simply to stimulate your thinking. Only a few items need to be turned in for this lab assignment. Create another view **holdsales3** for the same landuse categories but for the 'L1' salecodes that reflect foreclosures. This view should have 125 rows. Redo the interpolated housing value surface using these 125 rows. Turn in the following:

- A. *Your SQL query* to create the view.
- B. A *printout of the interpolated surface* (equivalent to the interpolated images shown above—a screen dump is okay, but make sure it is readable, especially if you use grayscale output).
- C. A *brief few sentences* commenting on changed housing value patterns that you see in this final map, and one or two other analysis strategies that you might take to account for one or another of the complications that we haven't yet addressed.

You may turn in this assignment on paper or electronically. This lab assignment is intended to be done during today's lab but it is okay if you want to finish it later as long as it is turned in a day before **lecture #14**.
