Massachusetts Institute of Technology
Department of Urban Studies and Planning

11.520: A Workshop on Geographic Information Systems

11.188: Urban Planning and Social Science Laboratory

ANSWERS to In-Lab TEST - November 22, 2004

Test Instructions

- This test starts at 2:05 PM; You have until 4:05 AM to finish.
- This is an open-book, open-note test. However, do not contact any "live" non-staff person via electronic or other means while you take this test.
- Please create your answers in electronic format - just as you have for lab assignments - and save your work, as usual, in your personal locker (i.e., the H:\ drive). For your convenience we have provided an answer sheet for your *textual* answers. (On this answer sheet, we've provided some extra space where you can explain how you obtained certain numbers and county names. These explanations don't earn extra points but *allow you to get partial credit* if some of your numerical answers and names are incorrect but your method had merit.)
- Turn in your test. *Remember to turn in your test before leaving the room. Please confirm with an instructor that your exam has been received. We strongly recommend that you retain the file containing your answers until we return the graded exams to you.*
- Make sure to include your name and the MIT server user ID near the beginning of each document that you turn in. Failure to include your name and user ID on both your text file and PDFs *can cost you* up to 5 points (in total)!
- *Finally, don't spend all your time on one or two questions.* Start by looking over all the questions to get a general sense of what, and how much, is being asked, and the grade points associated with each question--then start work on your answers. Move on to the next question if you've spent more than 5 minutes on any short answer question; mapping questions will likely take longer.

Good luck!

Datasets for the Test

For this test, we will use US county-level data from the recent November 2, 2004, presidential election. Mount the class locker as drive M, and look for an ArcMap document called **`11.520test04.mxd`** that has been saved in the class locker as: **`M:\test04data\11.520test04.mxd`**. This ArcMap document uses a single shapefile, **`election04_county.shp`**, with the boundaries of the continental US counties expressed in latitude/longitude (using the Geographic Coordinate Systems (GCS) NAD 1983). The corresponding attribute table has several columns of data indicating the votes for Bush, Kerry, and Nader both as raw counts and as percents. *Before opening the ArcMap document,copy save it to your own workspace and rename to something like* ***`<your_username>_test04.mxd`***.

Another data table, **`ohiodata.dbf`**, in DBF format is included in the test04data directory. This table contains a number of additional demographic, socio-economic, and environmental data for Ohio counties. This DBF table can be joined to the US county shapefile using the FIPS code column that uniquely identifies counties within each state. We have also provided a data dictionary, **`2004_Election_Counties_ObMeta.csv`**, that is readable in Excel if you wish to know additional information about the fields in the attribute table or the Ohio data table. *[This tabulation of election data by county was prepared by the GeoVISTA Center at Penn State.]* We have also supplied shapefiles of the US States and the US interstate highways (also saved in lat/lon using NAD 1983). These shapefiles are available in: **`M:\test04data\USinterstates.shp`** and **`M:\test04data\USstates.shp`**. Finally, we provide a grid coverage that rasterized Ohio into 10 km grid cells (using the Ohio North State Plane coordinate system, NAD 1983). This grid coverage is available in: **`M:\test04data\ohio_gd.`**

When you open the ArcMap document, you will see, in the Data Frame, two thematic maps using the **`election04_county.shp`** shapefile. The top thematic map shades the number of Bush votes received in each county using a red-to-blue color scale with 5 categories of 'natural break' classification. The bottom thematic map shades the same attribute field (Bush votes) using quantile classification.

For some of the maps you create, we request a color scale varying from red to blue (as in the ArcMap document that we provide). Note that ArcMap provides tools that allow you to 'flip' the color scale in the event that your red-to-blue scale has red at the wrong (Kerry) end rather than at the Republican (Bush) end. For example, in the 'symbology' window for 'graduated symbols', you can click the 'Symbol' column heading and it will provide an option to 'flip symbols'. You may find these features helpful to get the shading that you want.

Part I: Short Answers (40 Points)

Question I-1 (14 points total, 2 points each)

- What is the maximum number of votes received by George Bush in any one **county**: ____954,764_____
- What is the state, county name, and FIPS code for that **county**? ___California, Los Angeles County, 06037___
- How many votes did John Kerry receive in that **county**? __1,670,341____
- How many counties are there in California? _____58_____
- What is the smallest number of votes received by George Bush in any one of the **48 contiguous US states**:? _120,710__ (Don't include the District of Columbia as a 'state' even though it voted in the election.)
- What is the name of that **state**? __Vermont_____
- How many votes did John Kerry receive in that **state**? ____183,621_____

Most of these answers are obtained by sorting the attribute table or by using the Field/statistics and Field/summarize tools in ArcMap

Question I-2 (9 points total, 3 points for 'a' and 6 for 'b')

This election map is displayed using a lat/lon coordinate system (GCS - NAD 1983). Local planning agencies typically use projected coordinate systems. For example, the MassGIS mapfiles that we have used are generally saved in the Massachusetts (mainland) State Plane coordinated system. Explain briefly (a) why local agencies display maps in state plane coordinates rather than lat/lon, and (b) ***two*** noticeable changes in the visual appearance of the US election map if it were displayed in a projected coordinate system (such as the US Contiguous Albers Equal Area Conic projection).

I-2-A: Latitude/Longitude measurement allows one consistent coordinate system to be used for measurements anywhere on Earth. However, plotting lat/lon as coordinates on a flat two-dimensional surface leads to distortions. Local agencies use projected coordinates with a projection method that keeps North straight up and does a better job of preserving distance measurements or areal measurements in the local area that is their focus of attention.

I-2-B: One change is that the northern border of the US would no longer be horizontal but would curve up on one or another side (or both, depending upon the projection). Another change is that the map would be taller and skinner because, for the lattitude of Ohio (like Mass), one degree of latitude is a much longer distance on the ground than one degree of longitude.

Question I-3 (4 points)

Overall, Bush earned 51.3% of the Ohio votes vs. 48.7% for Kerry. However, if you compute the arithmetic average of the county-by-county percentage of Ohio votes for Bush, that average is 58.6%. Briefly explain the reason for the difference.

Bush did relatively better than Kerry in the Counties with smaller populations. As a result, the average of the county percents is greater than the fraction of all votes that were for Bush.

Question I-4 (13 points total)

Part I-4A (3 points): In the ArcMap document, the 'Bad Election Map A' map looks strange. It shades votes for Bush using a natural-break classification. We all know that news reports tend to use red colors where Republicans won and blue colors where Democrats won. This map does use red for the counties where Bush wins the most votes and blue where he receives the least. Yet the map is mostly blue even though Bush won the election. Explain briefly why this is the case.

The map shading is based on the ***raw vote count*** in each county. Counties are shaded red only if Bush receives a very high number of votes *compared to his votes in all other counties*. The number of votes in each county is highly skewed and Bush did very well in counties with relatively small populations. So most of the blue counties in Bad_Election_Map_A are small-population counties that Bush won. The distribution of votes is so skewed that the 'natural break' classification puts most counties at the low-vote end of the spectrum - and, therefore, shades them blue.

Part I-4B (4 points): The Bush_Votes_Quantile map isn't so blue, but it still looks strange. For example, many counties along the NorthEast and West coast are red even though Bush lost the states along these coasts. Explain briefly why there are lots of red counties in these areas even though Bust lost those states.

The quantile classification in Bad_Election_Map_B isn't all blue because it puts about the same number of counties (one-fifth) in each shading category. Nevertheless, the counties where Bush got the ***most votes*** (i.e., the most red category) are the biggest counties on the east and west coast. In these big counties, Kerry received ***even more votes***.

Part I-4C (6 points): Explain briefly your choice of attribute field, symbology, and classification choice in order to display a map that presents a better indication of the geographic pattern of the voting results. (You do not need to turn in a map at this point - just explain what you would do and why.)

To portray the outcome of the election we should look not at the Bush votes in each county, but at the percentage of votes for Bush in each county. (If there were significant third party candidates, the choice of measure for Bush's 'support' could be more ambiguous.) In this case, I would choose to shade the map based on the Bush-percent of

votes with quantile classification so we see the upper, middle, and lower ranges across counties. Even so, many sparsely populated counties (in the central US) are large, and that fact leads to an areal distortion.that exaggerates that impact on the vote count of these larger, low-population counties. One way around this is to plot the vote as a pie chart for each county with the size (i.e. area) of the pie proportional to the county population and the shading of the pie showing the Bush/Kerry. One might argue for another choice of classification scheme or a slightly different measure. For example, you could look at the difference in votes between Bush and Kerry and use standard deviations for the classification - this map would look less red/blue since most counties are closer than 60/40. We can accept more than one choice if reasonably argued. [If you would like to see further exploration of state and county level voting patterns, see (a) the website of the GeoVISTA Center staff member, Anthony Robinson, who packaged and posted the Election 2004 data that we used: http://www.personal.psu.edu/users/a/c/acr181/election.html, (b) the U. of Michigan site of Michael Gastner, who has developed cartogram algorithms that strectch or shrink polygons in various ways that produce density-equalizing maps: http://www.santafe.edu/~mgastner, and (c) the 'Purple Map' of Robert Vanderbei on his Princeton site: http://www.princeton.edu/~rvdb/JAVA/election2004].

Part II: Spatial Analysis and Mapping Using ArcGIS (60 Points)

For this portion of the test, you will use ArcMap to prepare your answers starting with the **`11.520test04.mxd`** document. Remember to copy this ArcMap document to your own workspace as **`<your_username>_test04.mxd`** before opening it. Our questions will focus on Ohio, one of the large 'battleground' states that was heavily contested in the recent election.

Question II-1 (12 points)

Create a 'Definition Query' for the US County map layer so that only those counties in Ohio are included. (For the rest of the test we will focus only on Ohio and this restriction will speed up the processing.) Also change the coordinate system of the Data Frame so that, instead of lat/lon, you use 'NAD 1983 State Plane Ohio North FIPS 3401'. (Hint: Set the properties of the Date Frame to be the appropriate pre-defined coordinate system.) Zoom in on Ohio and shade the counties based on the percentage of the votes that Bush received. (Note that Nader was not on the ballot in Ohio so the total votes = Bush votes + Kerry votes.) Use quantiles with 5 categories and a red-to-blue color range with red indicating more support for Bush. Add the US Interstate Highway shapefile (**USinterstates.shp**) and the US State boundaries (**USstates.shp**) to your Data Frame.

Turn in a PDF file showing a layout view of the Ohio thematic map that you create. Be sure to have your name and Athena userid on the map. Also be sure to project the map to Ohio North state plane NAD-83 and have the highways on top with symbology that

makes them clearly visible. Include a North Arrow and legend as well. The data source is: 2004 US Presidential Election, Penn State GeoVISTA Center. See Map 1.

Question II-2 (23 points total)

The newspapers have made a lot of the urban/rural dichotomy - Keery was strong in the cities and Bush was strong outside the cities. Let's look at the vote in counties that are close to, and far from, the major highways.

Part II-2A (8 points): Highlight those Ohio counties that have at least one US Interstate highway cross through. How many Ohio counties have interstates running through them? ___47____ ***Turn in a PDF file*** showing a layout view of Ohio counties and interstate highways with those counties clearly highlighted that have an interstate running through them.

See Map 2a: Use 'select by location' to find the counties that intersect with US Interstates.

Part II-2B (11 points): Examine the attribute table of election results in order to fill in the eleven blanks in the following table

Ohio	Number of Counties	Bush Votes	Kerry Votes	Difference (Bush-Kerry)	Total Votes	Bush Percent	Kerry Percent
With interstates	47	2,269,304	2,319,126	(49,822)	4,588,430	49.5	50.5
Without interstates	41	526,843	340,538	186,305	867,381	60.7	39.3
All counties	88	2,796,147	2,659,664	136,483	5,455,811	51.3	48.7

Part II-2C (4 points): Briefly interpret your results. Is Kerry a lot stronger in those counties with interstates?

Kerry is indeed stronger in those 47 counties that have interstates - he won there with a 50.5 to 49.5% split. But he was not enough stronger to make up for the 61/39 split for Bush in the other 41 counties - even though those small counties had far fewer total numbers of votes (867K vs. 4,588K). As an aside, note that Kerry's strongest counties appear to be those with at least two interstates that cross one another (i.e., the metropolitan areas). If we had a map layer showing the interstate exists, we might get a stronger correlation between interstates and Kerry strength. There may be some counties with interstates running through them but with no exit in the county. We could also look at the relationship between the vote and the number of exists or the number of interstate miles in each county.

Question II-3 (13 points total)

Add the **ohiodata.dbf** table to your ArcMap Data Frame and join it to the **election04_county.shp** shapefile using the FIPS code for the state+county.

Part II-3A (4 points): What is the average across all Ohio counties of the **pctpoor** field (measuring the percentage of person in each county who are determined in the 2000 census to be living in poverty)? _____13.0%_____

Part II-3B (3 points each): What is the mean of the **pctpoor** values across those counties that voted 60% or more for Bush? Indicate the number of counties: __44__ as well as the mean: __10.9%__ and standard deviation: _4.9%__.

Question II-4 (12 points total)

Instead of simply determining which counties have interstates, we decide to measure distance from interstates. Let's use ArcMap's spatial analyst for this purpose. To save time, we have already rasterized Ohio into 10 kilometer grid cells and saved this coverage for your use under the name: **ohio_gd**. Now use the Spatial-Analyst/Distance/Straight-line function to compute a new raster layer whose cell value is the straight-line distance (of the center of the grid cell) to the nearest Interstate. BEWARE: you will not be able to do this raster-cell distance computation **unless** the 'spatial analyst' extension is turned on in *Tools/Extensions **and*** the 'spatial analyst' toolbar is turned on in *View/Toolbars*, **and** the Data Frame is set to a projected coordinate system (we want Ohio North State Plane, NAD-83), **and** you have set all the appropriate properties in the Spatial-Analyst/Options dialogue box. (*So much for simple spatial analysis*!) Make use of the **ohio_gd** grid cells that we have provided when setting these properties.

We don't have time on this test to do much much raster exploration of the voting data. It will be enough to do the following:

Part II-4A (6 points): Determine that grid cell that is farthest from an interstate. What is the distance from that grid cell to the nearest interstate? __94.3 km____ What county is underneath the center of that grid cell? _Adams (39001)__ What percentage of votes in that county were votes for Bush? __64.1%___

Part II-4B (6 points): Shade the grid cells using a red-to-blue color scale with red for those cells farthest from interstates. ***Turn in a PDF file*** of your Ohio map showing the shaded grid cells underneath the interstates. Be sure to highlight that grid cell from Part II-4A that is farthest from an interstate.

See Map 4b: Because the Ohio grid was used to mask the raster operations when computing the distance to US Interstates, only the portions of interstates that fall within the Ohio grid were considered. In a few cases along the southern and western edge of Ohio, portions of US Interstates that are outside of Ohio are closer to the (center of) the

county than any interstate within Ohio. To include out-of-state interstates in the distance calculations, you need to set the mask and analysis units larger than Ohio. We did not expect you to do this even though we would accept answers computed either way.

Please note:

- Before creating the PDFs, be sure to include your name and the MIT server ID somewhere on the text file and the map layouts
- You should test your PDF files with Adobe Acrobat Reader before emailing them to us in order to be sure they are readable.
- You should keep a copy of your text output and PDFs file in your 11.520 user directory until your test is graded
- You should confirm with an instructor that your test files were received?
- You should turn in one text file with written answers to the questions plus three PDF files containing maps for Questions II-1, II-2A, and II-4A. For your convenience we have provided an answer sheet for your use in writing the textual answers.
