University of Maryland

Department of Government and Politics

GVPT - 392 (849): Introduction to GIS for Social Science Research

---

Mid-Term Exam

Author:

Nick Thompson

Instructor:

Dr. James Gimpel

October 5-7, 2016

(extension granted to OCT 9 for illness)

The following coverages can be found in the PACD8 folder.

1. PA_CD8_Voterfile: all registered voters for Pennsylvania, Congressional District 8. This is north-suburban Philadelphia, including all of Bucks and part of Montgomery Counties.

2. PA_CD8_Boundary: the outline for CD 8.

3. PA_and_NJ_Counties: County boundaries for the two states.

4. Four_States: State boundaries for PA, DE, NJ, and MD.

5. CD8_PA_Pct_Data_2012: voter precinct data for 2012.

6. Mont_County_Recent_Movers_10_12

7. Bucks_County_Recent_Movers_10_12

8. CD8_Places

9. PA CD8_Tracts

Three files above contain points for voters at their residences. These are 1, 6, and 7. For these files, the following columns contain important information:

Age (and Year Born) = the age of the voter in 2012.

Rep_Party, Dem_Party, Ind_Unaf_Party = the party registration of the voter: Rep = Republican, Dem = Democratic, and Ind_Unaf = Independent/Unaffiliated.

And there are other items that will be less important for this exercise.

For the following questions, use whatever tools you deem appropriate form the ArcGIS package, but be sure to describe what you did to address the questions. Be resourceful, but you need not write more than one page in response to each question.

1. Aggregate the voter and mover data to the census tract level for PA CD8.

To aggregate the data, I used a three phase process with multiple steps in each phase. In Phase 1, I imported the data using the catalogue in ArcMap. To import the data I first created a geodatabase file named **exam.gdb**. Here I imported all exam shapefils included in the provided exam folder by right clicking on the **exam.gdb** and selecting import from multiple. Next, I systemtaically added four file layers to the ArcMap table of contents:

- PA_CD8_Voterfile (hereafter depicted as **voter**);

- Mont_County_Recent_Movers_10_12 (hereafter depicted as **MC**);

- Buck_County_Recent_Movers_10_12 (hereafter depicted as **BC**);

- PA_CD8_Tracts (hereafter depicted as **tracts**).

This was the end of Phase 1.

In Phase 2, I reviewed the data and deleted unnecessary fields. The number is too great to depict which were removed. I kept essential fields outlined in the instructions above, as well as some others that I anticipated would be necessary (including **MOVER** from the **voter** file, **ozipcode** and **dzipcode** from **MC** and **BC**, and **ORNIC**, **DRNIC**, and **RNIC** from **voter**, **BC**, and **MC**. The combination of fields chosen allowed me to manipulate the data to achieve the desired results. I removed fields by double-clicking on each layer in the table of contents and navigating to the **Fields** tab. After clearing all of the fields, I was able to check only the fields I wanted to keep. Next, I exported the data

into new layers within the geodatabase. This data management process ended Phase 2.

In Phase 3, I used the **Spatial Join** feature (hereafter known as **SJ**) to systemtaically join the layers. First I conducted a **SJ** of **vote_total**voters**vote_total** to **tracts** and created a new layer called **tracts01**. Next I created the following **SJ**s:

a. **tracts + BC = tracts02**

b. **tracts + MC = tracts04**

c. **tracts01 + tracts02 = tracts03**

d. **tracts03 + tracts04 = tracts07**

The last combination created a spatially joined dataset depicting the north-suburban part of Philadelphia.

- Then compute and calculate the Democratic % of total registered voters (10 points).

To compute and calculate the Democratic % of total registered voters I needed to create a new field in the **BC** and **MC** shapefiles. I completed these computations prior to merging all of the data to ensure that they were carried over in each of the **SJ**s. First, I created a new field called **vote_total**. Using the field calculator tool, I added the **Republican**, **Democratic**, and **Independent** fields together. This produced a one in each row of the **vote_total**. Next, I used the statistics tool to calculate the total sum of from the **Democratic** field and the the sum from the **vote_total** fields. I conducted statistical analysis before and after conducting the joins. The Table 1 below shows the outcomes. Note there is no significant difference in the percentages either pre- or post-join.

Field	Pre-Join	Post-Join
Democratic vote	$71,048$	$133,467$
_total	$540,451$	$1,019,887$
Percentages	53.23%	52.99%

Table 1: Percentage of Democratic Voters

- Compute and calculate the Democratic % of total movers in Bucks and Montgomery counties (10 points).

To calculate the percentage of democratic movers I created two fields in **BC** and **MC** once callde **zip_dif** and another called **move2**. The **zip_dif** field captured a difference between the originating zip code (**ozipcode**) of each voter in the respective counties and the destination zip code (**dzipcode**). Next, a phython code converted the **zip_dif** field into a 1 or a 0. This allowed me to total the number of people that moved from one zip code to another. Table 2 shows the results of this computation.

```
def is\_positive(x):
if (abs(x)>0):
return 1
elif (abs(x)==0):
return 0
```

Figure 1: Python Code

- Produce two maps of these percentages. (20 points)

To produce my maps I followed some formatting guidelines. First, I always included a legend, scale, and north seeking arrow. Also included was my name as the author and the date I finalized the map.

Field	Montgomery	Bucks	Sum
Democratic move	2,016	22,555	24,571
_total	4,620	42,040	46,660
Percentages	43.63 %	53.65 %	52.66 %

Table 2: Percentage of Democratic Movers

For this first set of maps I normalized the **Sum_Sum_Democratic** variable over the **Sum_Sum_vote_total** for the first map (as labeled in Figure 1). I normalized the **Sum_Sum_Democratic** variable over the **Sum_Sum_vote_total** variable for the second requirement.

This depicts a strong concentration of Democratic voters located in the southeastern portion of the country. Democratic voters show a propensity for migration with a large percentage moving in and around the southeastern portion of the county.

Note that this data does not reflect the **vote_total**MOVER**vote_total** field from the **MC** dataset. The **vote_total**MOVER**vote_total** variable was not used because it was only available in the **MC** dataset and only depicted a small swath of migration running from northwest to southeast along the southwestern third of the county. All map figures will be available in the Appendix.

2. How would you characterize the spatial distribution of Republicans, Democrats, and Independents in PA CD8? Write up two paragraphs based on what you have found, describing how you used ArcGIS to address the question. (20 points)

Democrats, by and large, tend to be located in the south eastern part of the map (which is northern Philadelphia). Republicans are a significantly lesser amount of the

population and tend to inhabit the norther part of the voting district, with a concentration in the center of Montgomery county in the southwest. Independents show similar patterns to the Republicans but on a much smaller scale.

I used a percentage of total voters for each party and compared them to one another, as depicted in Figure 2. This distribution by percentage was calculated using the **Sum_Sum_Democrat**, **Sum_Sum_Republican** and **Sum_Sum_Independent** variables normalized by **Sum_Sum_vote_total** variable. I utilized a five quantile break.

3. The data included also show two populations of recent movers from inside PA and from nearby states. How do the recent movers in Montogomery and Bucks counties compare by age and by party registration to the entire PA_CD8 voting populations? Explain your answer in no more than one page. (20 points????)

To answer this question I pulled the descriptive statistics for each dataset (**BC**, **MC**, **voter**) and collected the number of people that voted in each party, the party means, and the party standard deviations (Table 3). I also collected the mean, standard deviation, minimum, and maximum values of the voters from each of the three datasets (Table 4).

The Montgomery county movers are represented by a higher percentage of Republicans than Democrats (Independents are behing Republicans and Democrats in all three datasets. The Bucks county Democratic movers (46.10%) have a significantly higher percentage than Montgomery county movers (37.68%). This indicates that either Democrats in Bucks county are migrating at higher rates, or that the percentage of voters in both counties falls around those means.

One way to differentiate between the two possibilities is to compare the means of the total voter population from the district. If we assume a normal distribution of Republican, Democratic, and Independent voters throughout the district then the averages of

34.13% for Republicans, 52.63% for Democratic party voters, and 12.10% for Independent voters should represent similar means the two counties.

The Republican numbers in Bucks county are close to the overall voter mean, but the Montgomery mean is higher. This indicates that Republicans are migrating to Montgomery county. The mean of Democratic voters in both Bucks and Montgomery counties is significantly lower than in the **voter** database. This may indicate less migration for Democratic voters.

The age demographics do not tell us as much without a geographic dispersion. The mean and standard deviation of the three data sets is relatively similar, as depicted in Table 4. As depicted in Figures 3-5, the age dispersion.

In order to complete the maps I combined a choropleth map of each party (red for Republulican, blue for Democrat, green for Independent) with a centroid map of the overall voting population. This normalized the regiserted voters by each age group. While the ages look identical in each map, the scales of the circles are different. Note the low numbers of 18-29 year old voters in comparison to 30-49 year olds.

To create the centroid maps I followed this sequence. First I created XY coordinate for each dataset by adding a new X and Y field respectively. Next using the "Calculate Gemoetry" function for each X and Y respecitvely. Next, I joined the **tracts** map with the **BC**, **MC**, and **voter** data sets. Then I exported the joined tract datasets to a **vote_total**tractCentroids**vote_total** table. This table now had the centroids for each voting block.

To finalize the method I created a feature class from the XY **vote_total**tractCentroids**vote_total** table. Opening the cataloge in ArcMap, I right clicked on **vote_total**trackCentroids**vote_total** and clicked "Create Feature Class" > "From XY Table". I make the coordinate system

Dataset	Party	Vote Totals	Mean	Std_Dev
Montgomery	Republican	2230	0.4167	0.4930
Montgomery	Democratic	2016	0.3768	0.4845
Montgomery	Independent	1043	0.1949	0.3962
Bucks	Republican	16808	0.3436	0.4748
Bucks	Democratic	22555	0.461	0.4984
Bucks	Independent	8979	0.1835	0.3871
Voter	Republican	46085	0.3413	0.4742
Voter	Democratic	71048	0.5263	0.4993
Voter	Independent	16334	0.121	0.3261

Table 3: Descriptive Statistics by Party

the same as all of my other maps, the South Pennsylvania 1984. Finally, I saved the file as a File and Personal Geodatabase called **vote_total**voterTractCentroids**vote_total**.

Data Set	Mean	Standard Deviation	Min	Max
Montgomery	43.81	16.73	20	104
Bucks	44.83	15.99	20	104
Voter Data	49.57	17.50	0	112

Table 4: Descriptive Statistics by Age

4. Are movers more likely to relocate to particular locations than the general voter populations within PA CD8? Or are they geographically distributed about the same way? Explain what you find in no more than one page.

Based upon the analysis form the maps generated in Figures 3-5, the propensity for 18-29 year olds is to reside in large numbers in Philadeliphia across the spectrum of voting preferences. People in the 30-49 age blocks are more likely to live in the more rural areas of bucks county, with a strong showing in the more suburban areas (there are a large numer in the city, but less than the younger voting block). Finally, the oldest block of voters is more broadly dispersed and in generally smaller numbers overall.

The migration pattern appears to be that younger people migrate to the city, whil middle aged people have a propensity to migrate back to the suburbs. Seniors are likely to live in all of the areas in a similar pattern to middle aged voters.

5. How many Bucks County voters live within one half-mile of the border with Philadelphia? (10 points)

To answer questions 5, 6, and 7 I used the following processes and steps. Noted differences are discussed in each question, but generally a process or step for questions 6 and 7 follow described processes or steps described in the following discussion.

There are 10,318 voters in Bucks County within one half of a mile of the border with Philadelphia. I found this information through the following process:

1. Step 1: Identify data to use:

    (a) The **vote_total**voters**vote_total** dataset and the **vote_total**PA_and_NJ_counties**vote_total** dataset. I used the attribute tables from each. A single entry in the **vote_total**PA_and_NJ_cou dataset identifies the boundary of Philadelphia (object ID #72). To find this

quickly and select it I sorted ascending the **vote_total**City_Boundary**vote_total** field and scrolled to the one indicating Philadelphia.

2. Step 2: I needed to isolate the Bucks county voters within the **vote_total**voters**vote_total** dataset. I opened the attribute table and identified the field I needed to use (**vote_total**JURISNAME Selecting by attribute, from the selection menue, I chose the **voter** layer > Method: create new selection > double click **vote_total**JURISNAME**vote_total** > = > "Get Unique Values" > double click **vote_total**Bucks County**vote_total** > Apply > Ok.

   (a) Next I exported the selected features to a new **vote_total**File and Personal Geo-database Feature Class**vote_total** called **vote_total**Voter_BucksCty**vote_total** > Export Data > Selected Records > Output Table. Next I went to the catalog and put the new **vote_total**Voter_BucksCty**vote_total** into the Table of Contents.

3. Step 3: To identify the voters within a half of a mile (0.5 miles), I "Select by Location" from the "Selection" menu.

   (a) Select only **vote_total**Voter_BucksCty**vote_total** layer in the "Target Layers" dialogue box > Selection method: "Select from Features" > Source layer: **vote_total**PA_and_ (with Philadelphia selected as noted above) > Spatial selection method for target features: "are within a distance of the source layer feature" > distance = 0.5 > unit of measure: "miles".

4. Step 4: Open the **vote_total**Voter_BucksCty**vote_total** attribute table and "show selected records". At the bottom there are two numbers: 10,318 records selected of 425,956 total records.

6. How many Bucks County voters live within one mile of the border with Philadelphia? (10 points)

There are 17,206 voters in Bucks County within one mile of Philadelphia.

To answer this question I already had the base processes and layers compiled. To change the distance I returned to "Selection" > "Select by Location". I kept all of the information the same, but I changed the distance from 0.5 miles to (1.0) miles. Next, I returned to the **vote_total**Voter_BucksCty**vote_total** attribute table and recorded the new calculated numbers: 17,206 records selected of 425,956 total records.

7. What percentage are those two figures of the total Bucks County voter population?

To answer this question I have all of the data I need from the previous questions. Arithmatic reveals the answer:

- Voters residing within $\frac{1}{2}$ mile: $\frac{10,318}{425,956} = 0.0242$, or 2.42%

- Voters residing within 1 mile: $\frac{17,206}{425,956} = 0.0404$, or 4.04%

Appendix