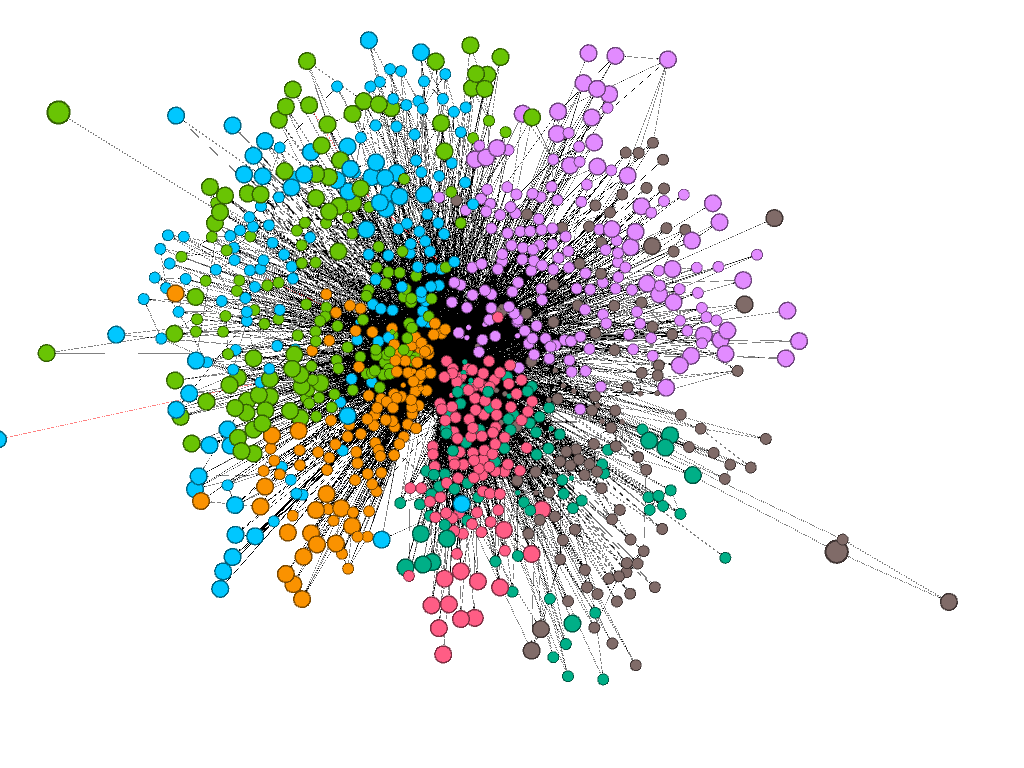
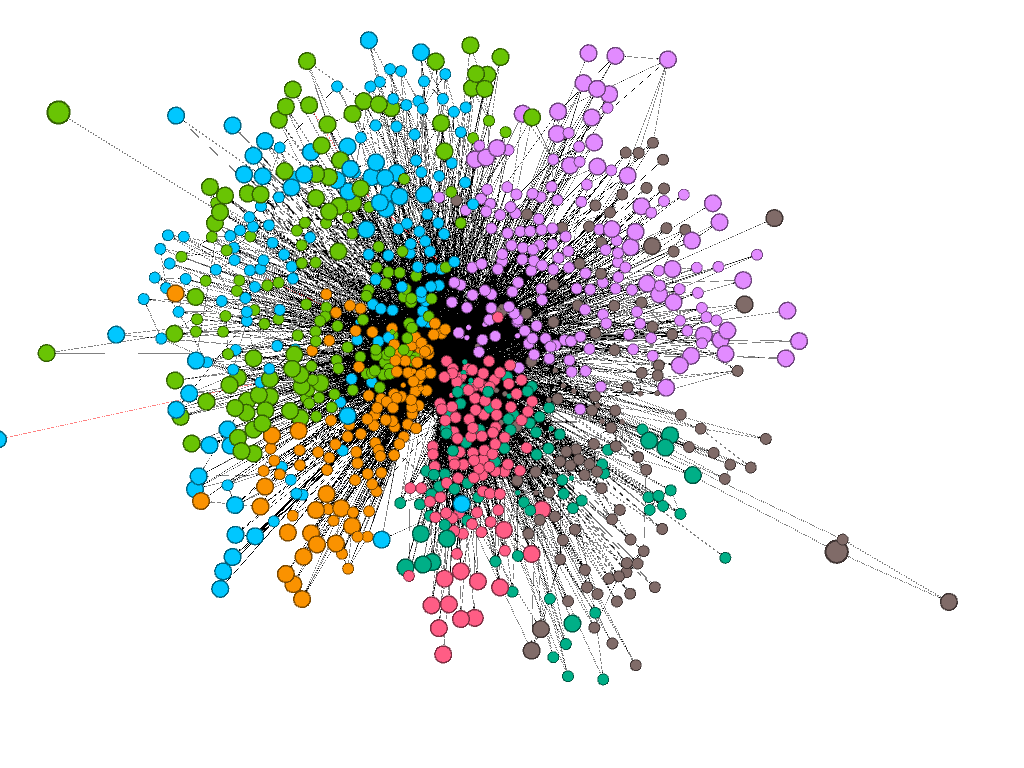
****

**Lab 5: Modularity, Migration & Regionalization**

NAME\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DUE \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Background**: New studies of geographic network partitioning (like modularity) look to create spatial regions from building blocks of areal units instead of study pre-existing administrative regions such as “Appalachia”.

Objective 1: Use modularity to find regions of migrants and map them in geographic space.

Objective 2: Use the gravity model in a very basic way to estimate the interaction between spatial entities (i.e. nodes) and find the interactions that highly over-perform and highly under-perform.

Migration data source: <https://www.irs.gov/statistics/soi-tax-stats-county-data>

**SECTION 1: MODULARITY**

1. In general, research on the regionalization (e.g. partitioning, delineation, clustering) of geographic space has a few different applications:

1. **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**
2. **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**
3. **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Part 1: Detect Migration Modules**

**Step 1:** Open Gephi and click new project. Go to File 🡪 Import Spreadsheet and bring in the file called “migrationEdgelist.csv” You should have about 900 nodes when you bring it in.

1. What is the minimum number of migrants between any two cities? \_\_\_\_\_\_\_\_\_\_\_\_\_.
2. (a) Where does this data come from? (b) What three values does it have associated with it? (c) What geographic level is it at ONLINE vs. what level are WE using it at? (d) What did we do to get the MSAs?

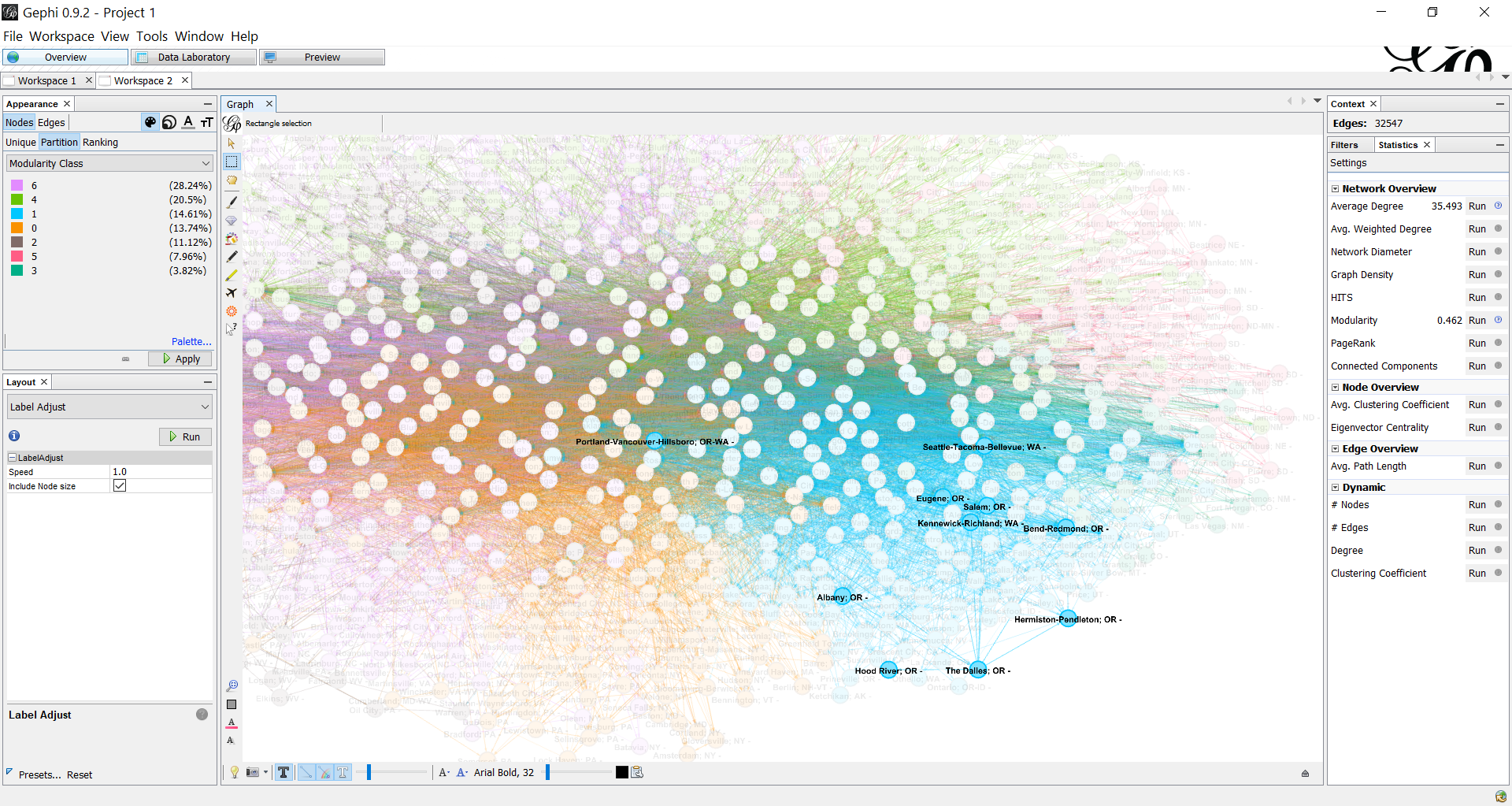
**Step 2:** On the right side, click the tab called Statistics. Click **Modularity**.

1. Report on the following statistics:
   1. How many modularity classes are there as a result (read the graph carefully!! Zero is a class!)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. What IS the modularity and what could its range be? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. How many nodes are in the largest class? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   4. How many nodes are in the smallest class? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 3:** Color your network based on the modules (under the appearance tab, go to partition, then select modularity class). Now go to your Layout tab and click a force-directed layout to use. See the different colors self-organize with each time step. Label the network and export the illustration or do a screen show with the little camera tool.

**Figure 1:** Label the nodes and create two screencaps of two very cities’ situations*.* Describe the difference in their spatial location in the network, their belongingness to groups (i.e. how many groups?), and what two independent variables could predict how many groups a node will belong to (does that play out in your examples?). *Below is an example of Baton Rouge LA, vs The Dalles OR. Use Label Adjust if you need to (under layout).*





*Think to yourself: What does the coloring say about how useful directed algorithms are?*

**Step 4:**  Go to your Data Laboratory. Make sure NODES is showing, and press Export Table. Call it *MigrationModules.csv*. Later on, you’ll bring this into ArcMap.

**Part 2: Mapping Regions**

**Step 1:** Open ArcMap. Bring in the shapefile called *cb\_2015\_us\_cbsa\_500k.shp* and *MigrationModules.csv*. Join *MigrationModules* to *cb\_2015\_us\_cbsa\_500k.shp* on the field of the NAME (*cb\_2015\_us\_cbsa\_500k* ) and ID (*MigrationModules)*.

**I highly recommend exporting this file as CBSA\_MOD.shp.**

**Step 2:** Symbolize each MSA by the color of the different module.

**Step 3:** Dissolve the data by module by going to Data Management 🡪 Generalization 🡪 Dissolve. Call this *ModRegions.shp*.

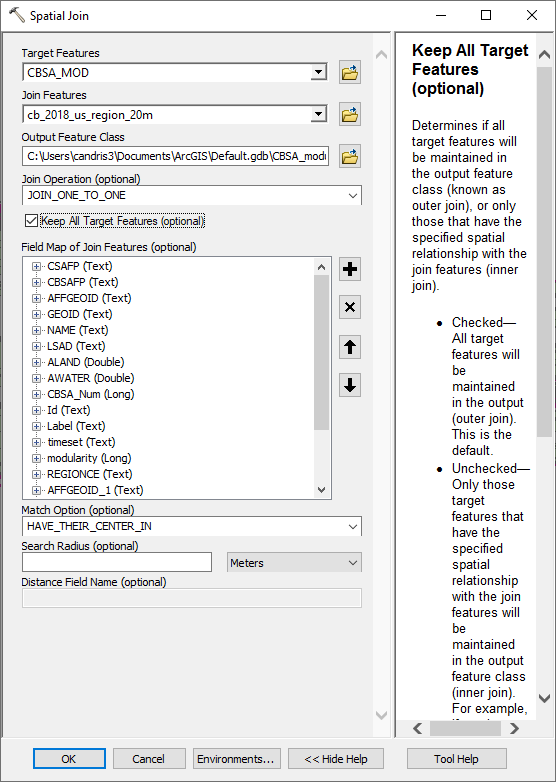
**Step 6:** Download Census **Region** Boundaries. (There are four Census Regions, and 8 divisions). <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>. Remember these? ☺ Add them in but only use the outlines so we can see the cities underneath.

**Figure 2:** Make a map and make resulting polygons clear, but give them different border colors. Put on the region boundaries. You don’t need to include Alaska and Hawaii but you can! Explain the regionalization patterns you see. Are the regions spatially contiguous? Are there areas that belong to NO region?

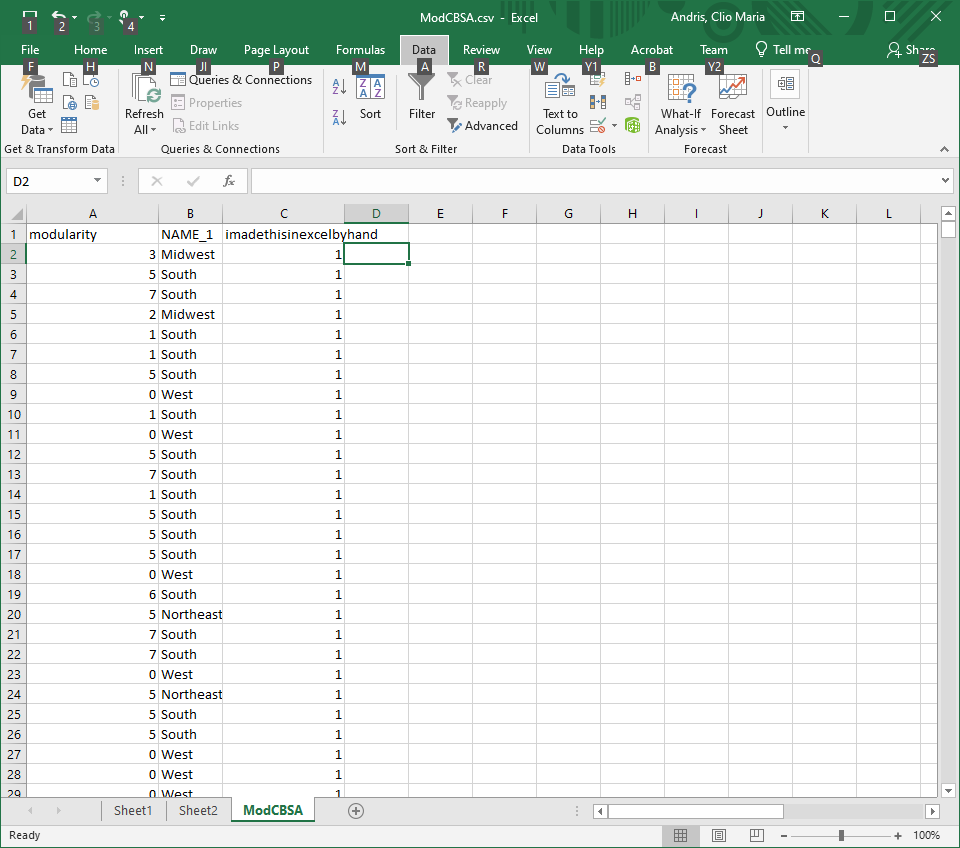
**Part 3: Evaluating Partitions vis-à-vis Administrative Boundaries**

Find the extent that your modules match up with these boundaries. There are many ways to do this.

Here is one way: do **a spatial join** with the **CBSA\_MOD.shp** and the **REGIONS**. Go to Analysis tools -> Overlay -> Spatial Join. Follow this window (below).



From the spatially joined shapefile, export a text file into Excel or R. Use {Pivot table in Excel or *Table* in R} to create a contingency table / confusion matrix with the **columns as the Name**, and the **rows as the modularity class**. Populate each with the count.



1. Put table 1 here and describe how well your modules align with the regions. You can use a statistic like chi squared, you can do percentage split, or entropy, whatever you want to help describe the phenomenon and answer the question: do US regions dictate migration?

Table 1: Confusion Matrix / Contingency Table: Migrants

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Region/Mod class | 1 | 2 | 3 | 4 | 5 | 6….? |
| Northeast |  |  |  |  |  |  |
| South |  |  |  |  |  |  |
| Midwest |  |  |  |  |  |  |
| West |  |  |  |  |  |  |

**Part 4: On your own. Repeat this process for commutes and divisions.**

**Commutes will have MANY modules. COMMUTERS ARE IN COUNTIES. FIPS CODES ARE COUNTIES.**

**To decrease the number of modules I will: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.**

1. What is the minimum number of commuters between any two cities? \_\_\_\_\_\_\_\_\_\_\_\_\_.
2. Report on the following statistics:
   1. How many modularity classes are there as a result (read the graph carefully!! Zero is a class!)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. What IS the modularity and what could its range be? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. How many nodes are in the largest class? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   4. How many nodes are in the smallest class? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Graphic 3: Reproduce graphic 2 but for commuters. (No need to reproduce graphic 1).

Download divisions here: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>.

1. Put table 2 here and describe how well your modules align with the regions. You can use a statistic like chi squared, you can do percentage split, or entropy, whatever you want to help describe the phenomenon and answer the question: do US regions dictate migration?

Table 2: Confusion Matrix / Contingency Table: Commutes (you can decide how many modules you want!)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Division/Mod class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8..? |
| New England |  |  |  |  |  |  |  |  |
| Middle Atlantic |  |  |  |  |  |  |  |  |
| East North Central |  |  |  |  |  |  |  |  |
| West North Central |  |  |  |  |  |  |  |  |
| South Atlantic |  |  |  |  |  |  |  |  |
| Mountain |  |  |  |  |  |  |  |  |
| Pacific |  |  |  |  |  |  |  |  |
| East South Central |  |  |  |  |  |  |  |  |

**SECTION 2**

**Part 5: Gravity Model and Computing Error (march 26)**

**In this part, we will look for flows that are performing above and below expectation in a very simple way.**

**Step 1:** Open the Excel file called *Migration\_2012\_GravityModel*. Or do it in R.

1. What is the formula for the gravity model? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What is the “2” called, and does it change? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 2:** Make a new column called **Gravity**. We will implement the gravity model, using population, population and distance. Watch the demonstration.

**Step 3:** Make a scatterplot of the Gravity vs. the Exemptions. (You don’t need to put the plot in the lab).

1. Find the linear trend. What is the trend equation? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What is the r^2 value on the trendline? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Is that good? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4:** Make a new column called ExpectedMigration. Predict the expected migration, given the trendline (model). Think to yourself: What is the formula you use to do this? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 5:** Sort the data ascending by the expected migration.

1. Lastly, create one final column called **Residuals**. What is the formula you use to compute the error? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(In R, you can use *LM* and *resid*, etc.)

**Step 6:** Sort the error ascending. Look at the biggest errors in both negative and positive directions (I.e. large absolute values).

1. Which city pairs exhibit this error and what does that say? Write a paragraph alerting a policy maker about proof for cities that may be in economic trouble!

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_