GEOG 400/500 Project 5 (PCA and K-means clustering) – Courtesy of Mr. Carvalho

Deliverables:

- ☐ a copy of your lab report (this document), autographed;

- ☐ a table (Calc spreadsheet) representing the character of the clusters you set up, showing their mean scores on each of the three principal components, autographed;

- ☐ two graphs based on that table (in spreadsheet): a line graph and a stacked bar chart

- ☐ **Extra Credit:** a copy of your choropleth map showing each county colored by its cluster classification (don't forget to put a legend explaining your color codes!). You can save it as a .png, .jpg, or .pdf

Lab Report:

1. What is the **first principal component** picking up on? (it doesn't polarize, BTW)

2. What about the **second component**? What would a high negative score mean and what would a high positive score mean?

3. What about the **third component**? Again, characterize high negative and high positive component scores.

4. Come up with a pithy description of your **first cluster**, considering the mean PC scores it shows on the three components (whether they're high positive, high negative, or sort of middling, especially in comparison with the other four clusters).

5. Do the same with your **second cluster**.

6. Ditto with the **third cluster**.

7. And the **fourth cluster**.

8. And the **fifth cluster**.

9. Looking at the distribution of PC scores by cluster and the meaning of high/low and positive/negative scores on each component, which of your clusters was likely to vote GOP in November 2012 and what makes you think so?

10. Which cluster or clusters was likely to go Democratic and why?

11. Which cluster or clusters is likeliest to see voters swing either way from one election cycle to the next and, so, was likeliest to be at play in November 2012?

12. How would you critique your use of PCA and K-means clustering on these data? Are their assumptions of scalar data and roughly normally distributed data met for all original variables? (hint: do the means and medians for each variable and see if they are roughly similar)

13. **Optional:** If you have done the extra credit choropleth map of your clusters, please write a very generalized statement about the major regional concentrations of each of your clusters. Spatially, name a few states that look as though they are predominantly of your "swing vote" cluster and, thus, might be where elections are decided?

14. **Optional**: If you did the map in QGIS, how would you rate it in terms of ease of learning enough to do the map? Have you ever used a GIS before? If you tried QGIS but have also used ArcGIS, how would you compare them for ease of use?