GGR372 Lab 2 Report

Blaise Calaycay

Feb 15 2017

Question 1

I chose to explore the rate of prostate cancer among males (variable name "PROM_RATE") across the Southeast US. Visually, it seems that in the eastern part of the region (i.e. Florida, Georgia, North Carolina, South Carolina), there is a higher rate of prostate cancer relative to the 20 kinds of cancer in the US Geological Survey dataset. However as we move further west, the rate of prostate cancer goes down while the rate of the 20 kinds of cancer goes up.

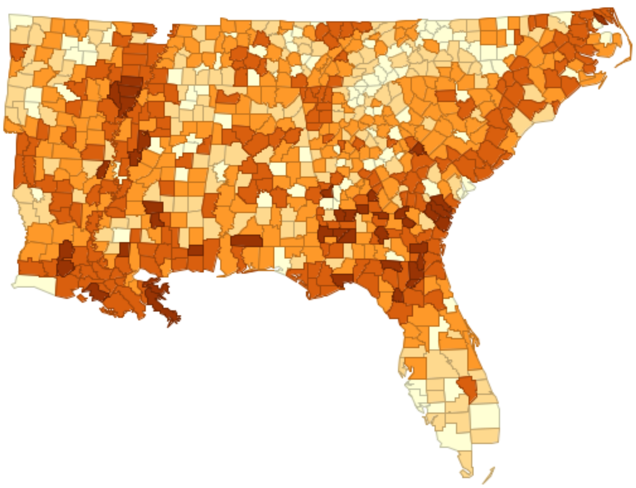

Figure 2. Male cancer rates in Southeast US

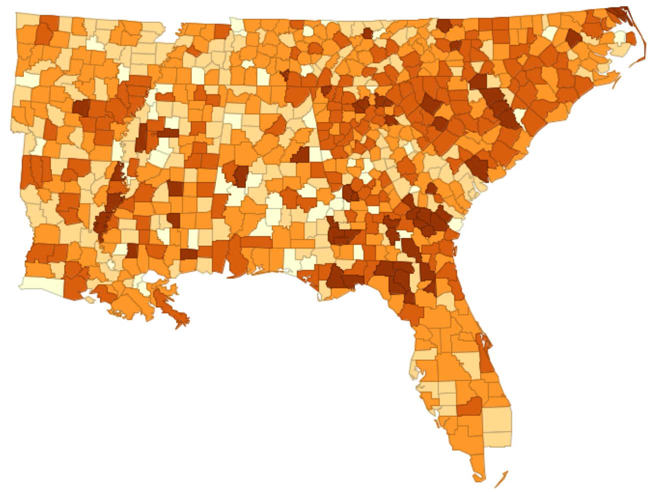

Figure 1. Prostate cancer rates in Southeast US

Question 2

Cartograms are appropriate to emphasize extreme values on a map. Rates can be used to size geographic areas in a cartogram if the values are extreme enough. Hence, even though the female lung cancer rates in our data could have a possible range of 0 to 100,000, the actual values are too close together (low-10s to mid-20s). On the other hand, the values for count range from as little as 5 and as large as 6,000 (i.e. the extremes are further apart).

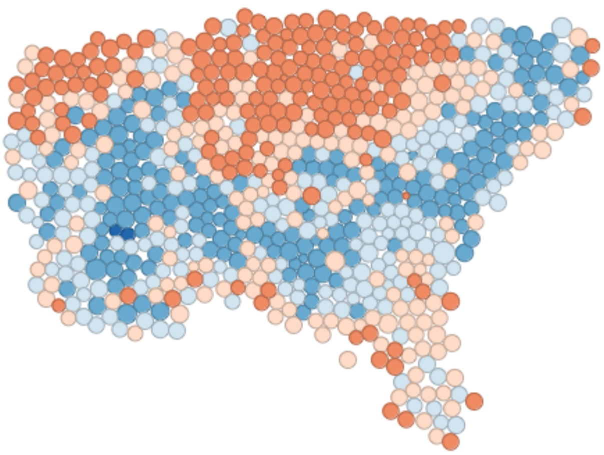

Figure 3. Cartogram, Southeast US; size: female lung cancer rate, colour: percentage of white population

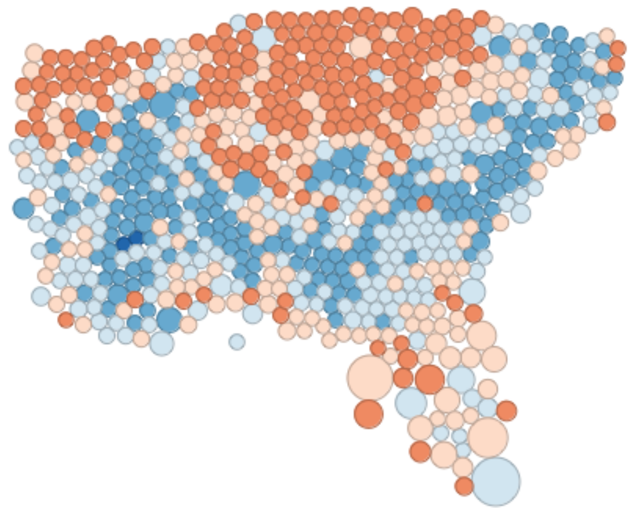

Figure 4. Cartogram, Southeast US; size: female lung cancer count; colour: percentage of white population

Question 3

I decided to categorize the colon cancer rate using 5 quantiles. Figure 5 shows the resulting choropleth map and Figure 6 shows the LISA cluster map for reference (see Appendix for Moran's I scatter plot and cluster significance).

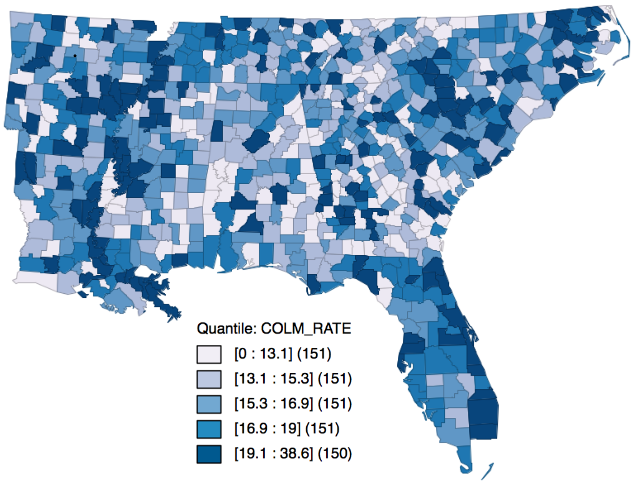

Figure 5. Male colon cancer rates in Southeast US

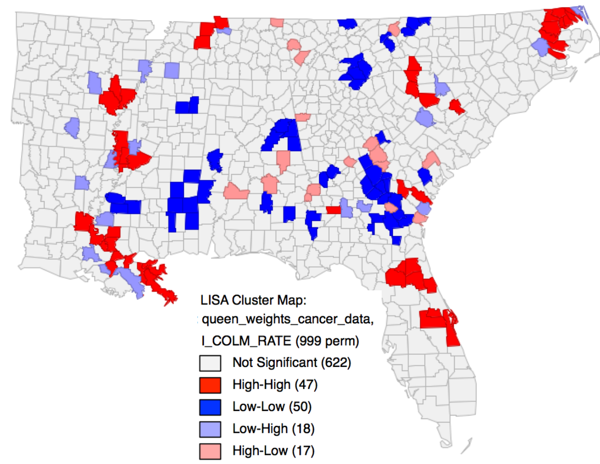

Figure 6. Clusters of male colon cancer occurrences in Southeast US

The choropleth map generally corresponds well to the cluster map. For example, the map on the right tells us that we should see clusters in the area around New Orleans, with Lafourche Parish (an outlier) having a low-high relationship with its neighbours. Exploring the choropleth map on the left confirms this relationship (see Figure 7; Lafourche Parish is highlighted). It is important to show what clusters are significant because we see many "clusters" in the choropleth map with the naked eye; however, a lot of these "clusters" do not have an important spatial correlation with the rate of colon cancer

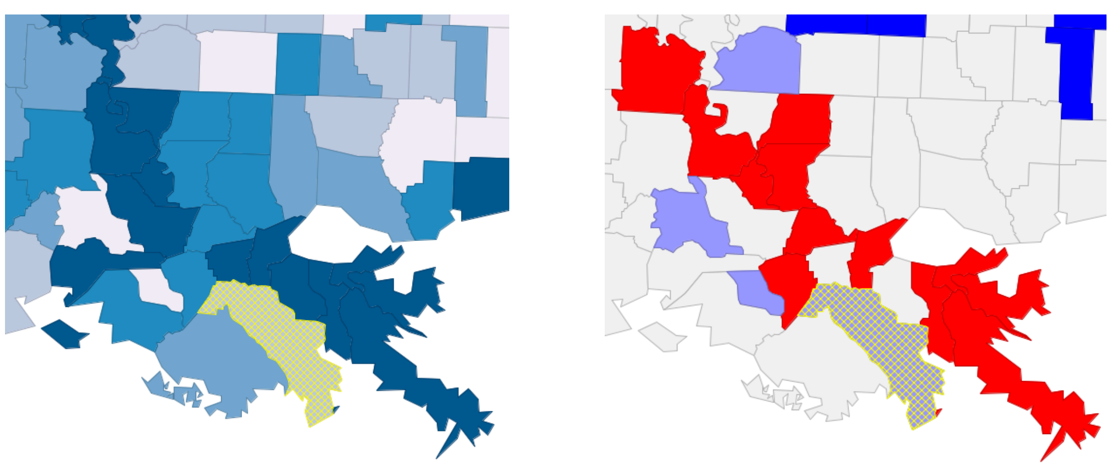

Figure 7. Lafourche Parish highlighted on choropleth map and cluster map

Independent Work: Exploring Diabetes Clusters in Toronto

I created a diabetes rate variable using `totalNbrDb` (diabetes cases per neighbourhood) as the event variable and `totalPop20` (population of residents aged 20 and up per neighbourhood) as the base variable. Below is a choropleth map using five quantiles for classification of neighbourhood diabetes rates in Toronto.

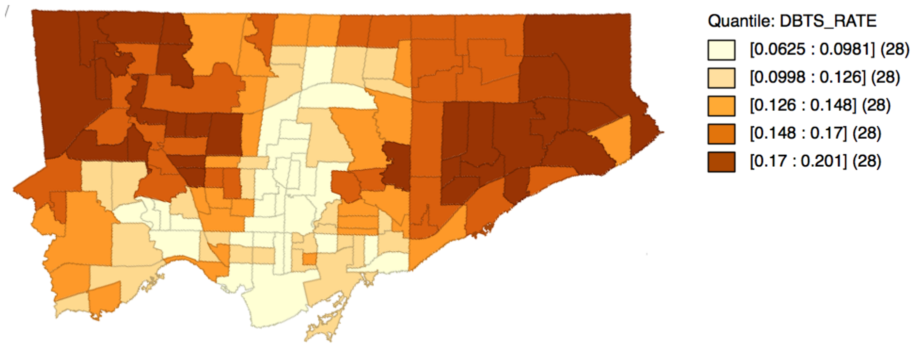

Figure 8. Diabetes rates in Toronto

We see groups of the darkest neighbourhoods on the upper east and west sides of the city and a group of the lightest neighbourhoods in the middle of the city. Now we want to explore whether these groups and others create spatially significant clusters of diabetes cases. Let's look at Moran's I to find out the global spatial autocorrelation in our dataset.

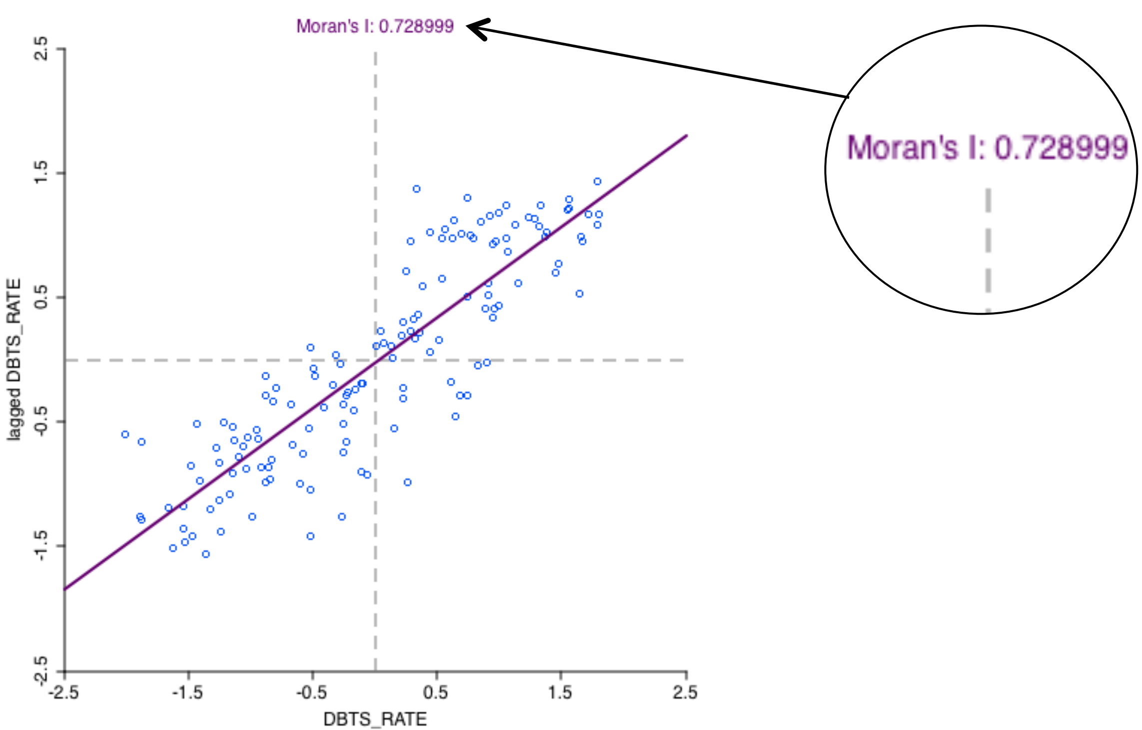

Figure 9. Moran's I scatter plot for diabetes rates in Toronto

We get a score of ~0.73, with few outliers, indicating a high positive spatial autocorrelation in our dataset. We can now comfortably say that there is <u>some</u> degree of dependency between geographic location and risk for diabetes (but we can only say that there is a correlational relationship and not a causal relationship). That is, we can say that there is some degree of clustering in our dataset.

Now, we want to know where these clusters are. Figures 10 and 11 show GeoDa-generated significance and cluster maps, respectively.

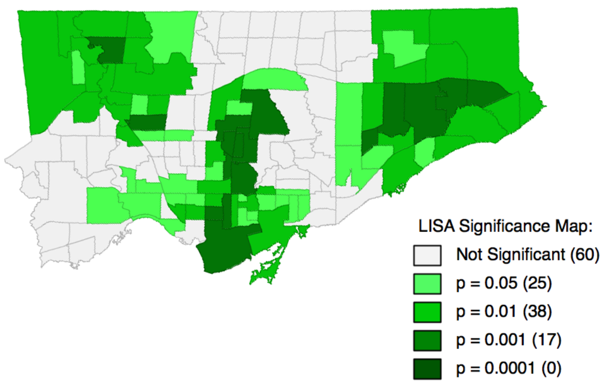

Figure 11. Significance of diabetes clusters in Toronto

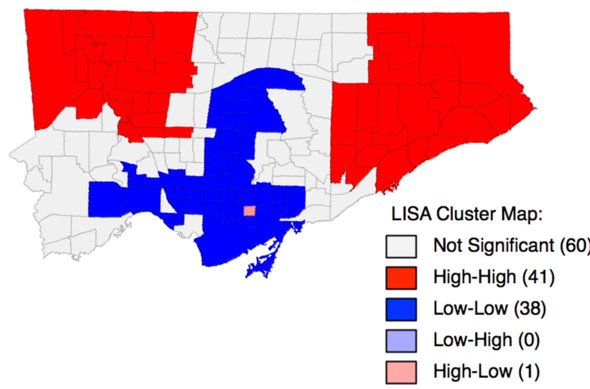

Figure 10. Cluster map of diabetes occurences in Toronto

Without looking at Figures 10 and 11, we might guess (by looking at the choropleth map) that there are clusters on the east, northwest, and central regions of Toronto. These regions in the map catch one's eye because they have the darkest and lightest colors packed tightly together. The significance and cluster map show that these regions in fact have clusters significant enough to our liking ($p \leq 0.05$).

In Lab 1, I discussed that the percentage of minorities in a neighbourhood has the strongest positive correlation with diabetes prevalence out of all the SES factors. In the map below, we see a high concentration of minorities in the northern part of Toronto and it gets higher the further from the center we go. In the cluster map, we see that these neighbourhoods

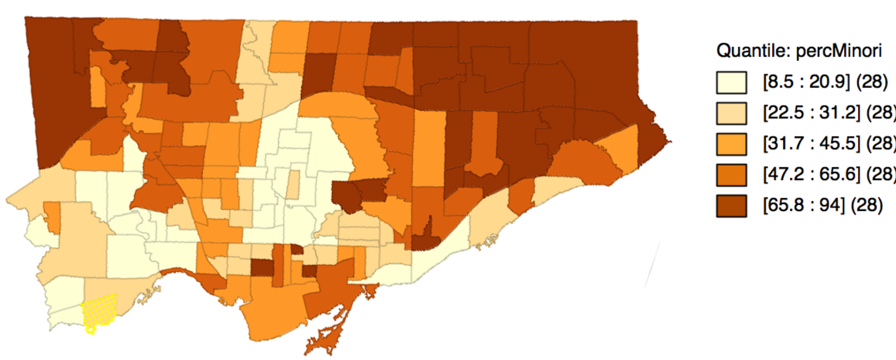

Figure 12. Percentage of minorities in Toronto

are significant clusters with high diabetes prevalence. On the other hand, the significant clusters with low diabetes prevalence span the downtown and midtown area. I discussed in Lab 1 that people with high income have a lower risk of developing diabetes. In addition to having a low percentage of minorities, the neighbourhoods in the midtown area (e.g. Yonge-Eglinton, Lawrence Park, Rosedale) are also wealthier than the rest of the city. This could explain why there is a low prevalence of diabetes in these areas. This reasoning becomes weaker as we move downtown (the percentage of minorities living downtown is high). However, let's recall that I used the percentage of university graduates as an indicator of "high income" in Lab 1. In the map below (Figure 13), we see a high percentage of university graduates in the downtown area. If we accept that a university graduate generally has a higher income than their non-university counterparts, this is a viable explanation of why the downtown area is part of the low-diabetes cluster. As a side note, we also notice that, in general, neighbourhoods with more minorities have fewer university graduates.

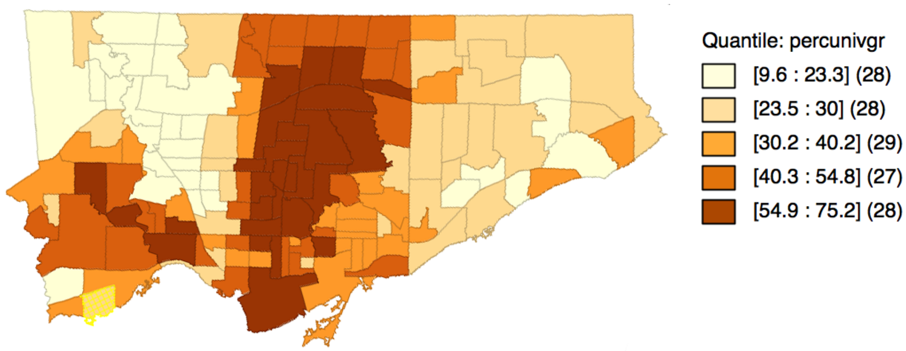

Figure 13. Percentage of university graduates in Toronto

Appendix

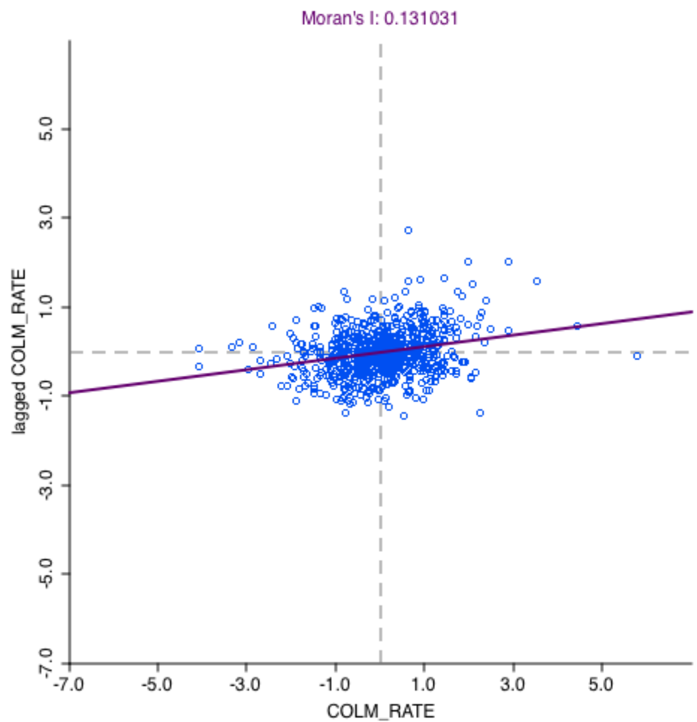

Figure 14. Moran's I scatter plot for male colon cancer rates in Southeast US

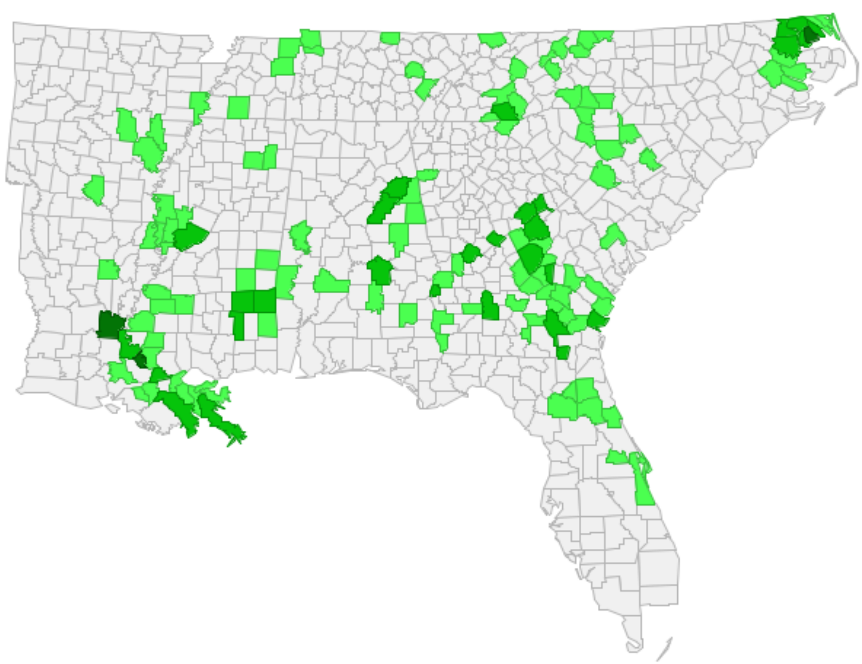

Figure 15. Significance of male colon cancer clusters in Southeast US