02.221 – Lab 10: Spatial Autocorrelation in ArcMap

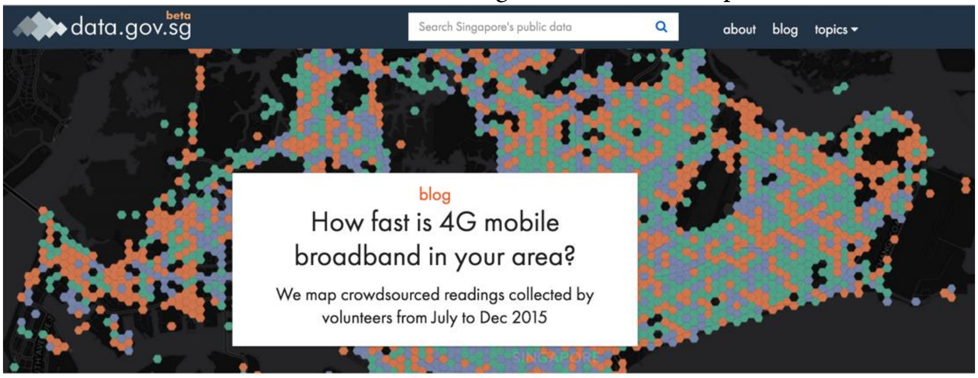
There is not additional data needed for this lab. You will find all the data sources in the explanation below.

Goals

The primary goals of this lab are to get acquainted with ArcMap as well as learn how to perform analyses of both global and local spatial-autocorrelation.

MyConnection SG Initiative & Data Conversion

While browsing the data.gov.sg website, you find a new blogpost on the data coming from the MyConnection SG initiative. The Infocomm Development Authority of Singapore (IDA) launched an app in 2014 that collects mobile broadband speed readings from users that have the app installed. The blogpost (https://data.gov.sg/blog/singapore-mobile-broadband) details an analysis and visualization of six months' worth of readings (1.5 million data points).



By datagovsg | 01 April, 2016 |

Singapore leads the world in smartphone usage, with at least nine in ten people using one last year. Increasingly, most people do without computers at home, preferring to use smartphones and tablets due to their ease of use and portability.

With smartphones becoming such an important part of our lives, any drop in mobile broadband speeds can be a disruption to our daily routines.

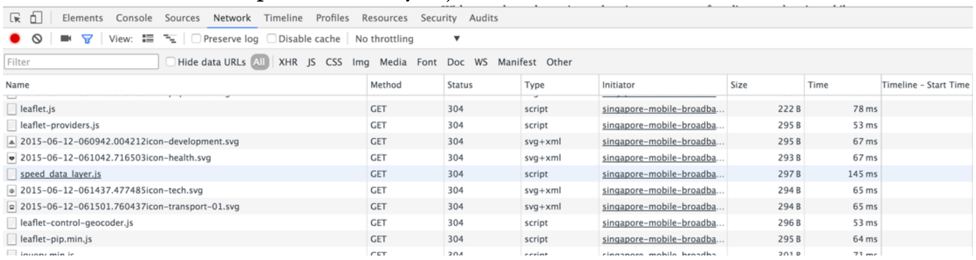
The MyConnection SG initiative

To identify places in need of improvement, the Infocomm Development Authority of Singapore (IDA) launched the MyConnection SG app in October 2014. The app, available on Android and iOS devices, collects crowdsourced data from volunteers from across the island, which is then aggregated to give a picture of how good (or bad) mobile broadband speeds are.

The map below shows the median speeds in different areas in Singapore on 4G networks, based on the 1,582,521 anonymised readings collected by IDA last year from July 1 to Dec 31.

Although the maps available in the blog are interesting, you would like to embark on a slightly more detailed analysis. Is the pattern shown by the mobile broadband speed readings indeed as random as the maps and accompanying text indicate? Or can we find statistically significant clusters of high and low speeds? Unfortunately, the open

data platform does not actually publish the underlying data. You will have to do some data-sleuthing to find the source data.
As the maps in the post are not static images but dynamically drawn, you use your browser's developer tools to look at all the files being requested when the page is loading. After some trial-and-error, you establish that the data you are interested in is contained within the speed_data_layer.js file.



You note down the url (https://data.gov.sg/media/uploads/blog/2016/04/myconnection/data/speed_data_layer.js) but also realize the data is currently not in a format that is easy to read into a desktop GIS program like QGIS or ArcMap. That's why you open a new R session to do some initial data wrangling and save the data into a format that you can read more easily.

First, you read the entire .js file into an R object.

```
s <-
readLines("https://data.gov.sg/media/uploads/blog/2016/04/mycon
nection/data/speed_data_layer.js")
```

Have a look at content of your object:

```
head(s)
[1] "var speed_data_layer = {"
[2] "\"type\": \"FeatureCollection\","
[3] "\"crs\": { \"type\": \"name\", \"properties\": { \"name\":
\"urn:ogc:def:crs:OGC:1.3:CRS84\" } },"
[4] "
[5] "\"features\": ["
[6] "{ \"type\": \"Feature\", \"properties\": { \"grid_id\": 2552, \"perc_disc_50\":
16.6416 }, \"geometry\": { \"type\": \"Polygon\", \"coordinates\": [ [ [ 103.616,
1.252272733872298 ], [ 103.615, 1.250540683064729 ], [ 103.613
```

You recognize the format as GeoJSON (http://geojson.org/). GeoJSON is a spatial vector format just as KML or SHP. It is predominantly used for online data visualizations. However, the first line of the .js file also contains some Javascript that assigns the data to a variable called speed_data_layer. Since we are not working inside

Javascript, we need to get rid of that assignment to turn our data into pure GeoJSON (cf. the geojson definition at geojson.org). We can use the familiar gsub() command to search for the extraneous variable assignment and then get rid of it.

```
s <- gsub(s, pattern = "var speed_data_layer = ", replacement = "")
```

Perfect! Now we have all our data in a clean format, we can read it into R as a spatial object. We save the intermediate file and then read it in with our trusty readOGR function.

```
writeLines(s, "speed.geojson")
s.sp <- readOGR("speed.geojson", "OGRGeoJSON")
```

Import package rgdal.

Let's plot the data and have a look at the attribute table to make sure everything is OK.

```
plot(s.sp)
head(s.sp@data)
```

You look back at the blog post and deduce that the perc_disc_50 column represents the median speed in each grid cell. Can you use R to calculate the mean and median speed over all grid cells? *Write down both the command you used and the answer.*

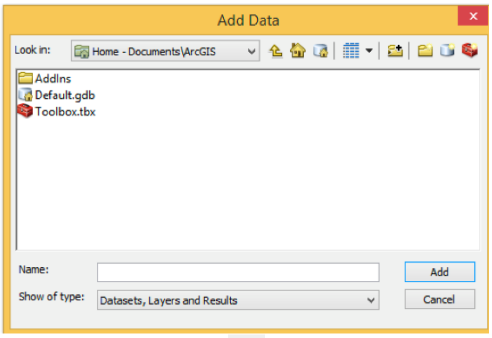
Now that we have the data correctly loaded, we can use R to save it to a data format that is easier to read into a desktop GIS program.

```
writeOGR(s.sp, ".", "4g-hex", driver="ESRI Shapefile")
```

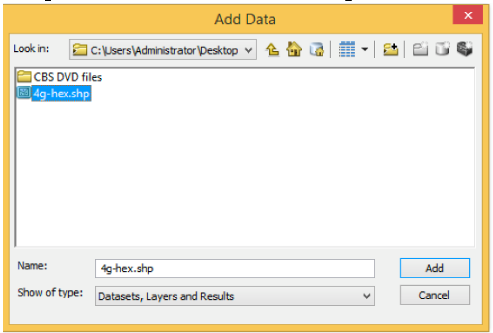
ArcMap

Open ArcMap (on the lab computers, you should be able to find it in the 'Start' menu). ArcMap has the same basic structure as QGIS so should feel fairly familiar. On the left, you have a table of contents that shows all the active layers. On the right, you have the map window. The top section of the screen has all the tools and menu items that you use to interact with your data and perform analysis.

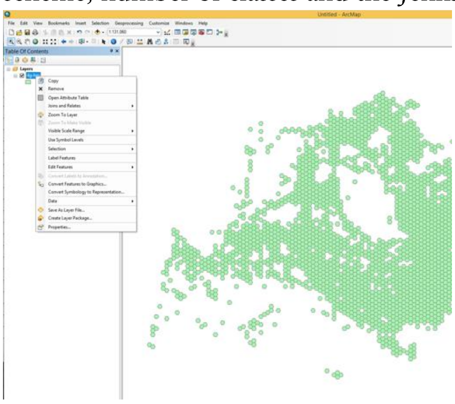
Before we do anything else, we need to add our newly created Shapefile to the table of contents. To do so, click on the 'Add Data'  icon. Arc is a little awkward when it comes to navigating the file system. You cannot navigate the file system like you are used to in other programs but need to create a 'Folder Connection' to your working directory first.

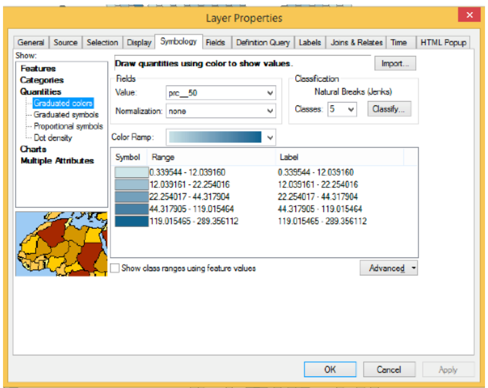


Use the little + icon  to create a folder connection to the folder where your Shapefile resides. Subsequently, you can use the folder connection to navigate to your Shapefile and add it to the map.



Success! Similar to QGIS, we can now use the layer properties to visualize our data in different ways. Right-click on the layer, select Properties. In the Layer Properties dialog window, select Symbology. This is similar to the 'Style' tab in QGIS. You can now use *Graduated Colors* to symbolize the prc_50 field. Use an appropriate color scheme, number of classes and the Jenks classification method to do so.

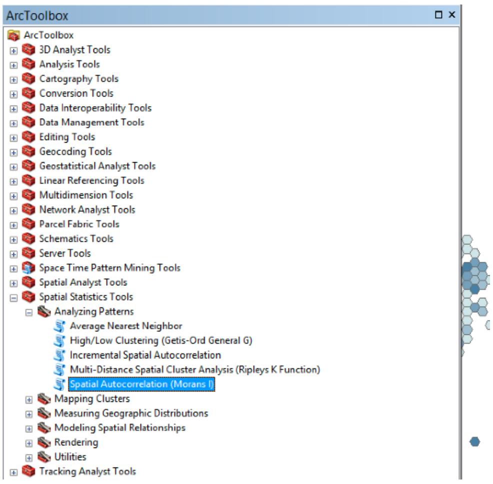




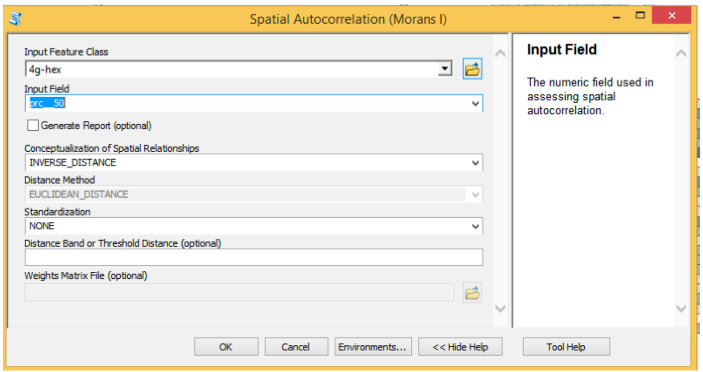
Can you find any discernible patterns in the resulting map? To enhance the contrast in the map further, you could, for example, create a diverging color scheme with a middle point based on the median speed you calculated in R earlier. This would more clearly show areas with high versus low speeds.

However, we can also use ArcMap's analytical tools to determine if there is a *statistically* significant pattern present without relying on human sight and interpretation.

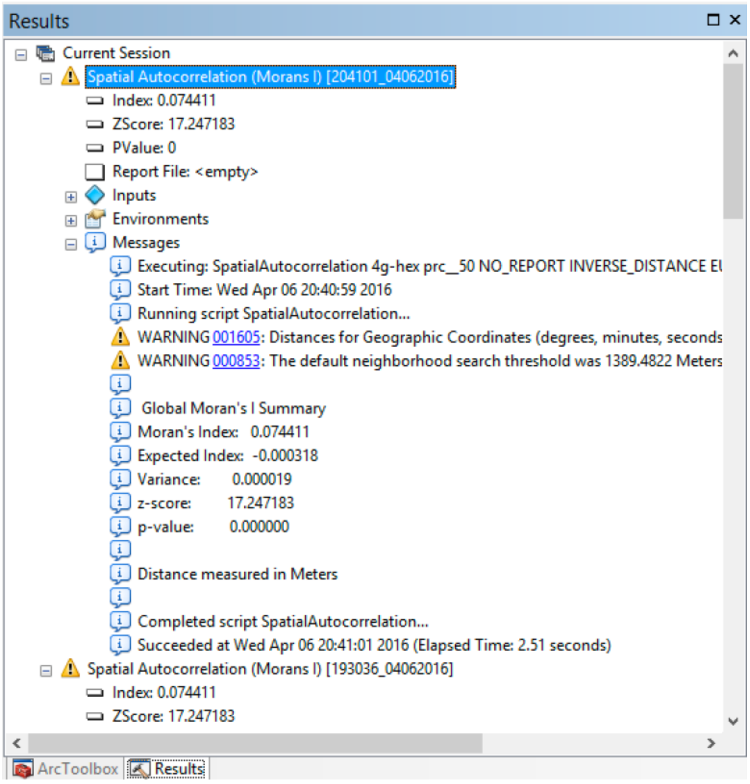
To do so, open up the ArcToolbox (it's the little red box) and navigate to Spatial Statistics Tools | Analyzing Patterns | Spatial Autocorrelation. This will calculate the global Moran's I. Moran's I has value between -1 and +1. -1 meaning negative correlation (dispersion), 0 meaning completely random and +1 meaning positive correlation (clustering).



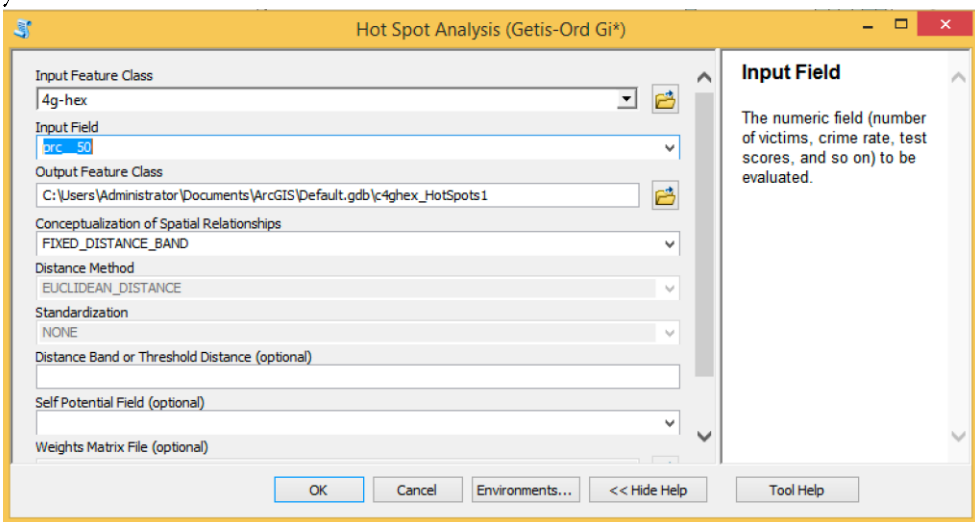
Set the appropriate input parameters according to the screenshot below and click 'OK'.



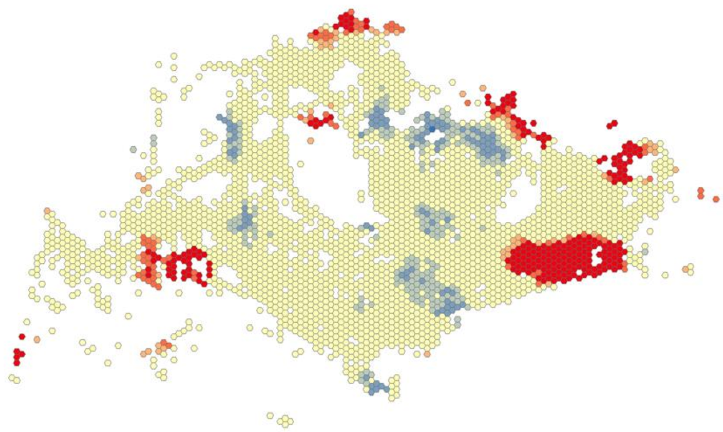
Once the calculation is finished, navigate to the Results tab to look at the results. As it turns out, the global Moran's I is highly significant but the degree of clustering is very small (0.07). This makes sense based on the visual pattern we saw earlier. However, this is a *global* calculation and does not mean there is no significant *local* clustering present. In other words, on the level of the entire country we do not find heavy clustering of high or low speeds in specific areas (e.g. west of Singapore much higher than east of Singapore), but that does not necessarily mean that smaller high or low speed clusters exist within the country.



To test for local clustering, we can run the Getis-Ord Gi* Hot Spot Analysis tool that you can find in the same toolbox.



The output of the tools is a new spatial layer that, for each hexagon, indicates whether the hexagon is part of a significant cold or hotspot.



Export the resulting map to PNG (File | Export Map). There is no need to worry about titles, legends or a finished cartographic product for now. *Include the map in your assignment and also include a short discussion of the results.* What are the significant cold and hotspots? If you were working for IDA, what would or could you do with this information?

Assignment

On the class website, you will find the assignment for this lab. It consists of the map of local hot and cold spots as well as the answers to the questions posed above. Please make sure you submit the assignment by **April 13.**