Geog 5050: Introduction to Applied Spatial Statistics in Geography
Exercise 04: Point Pattern Analysis in GIS and R

Value: 60 points
Due: Wednesday, 14 October, by 5:00 PM

Overview: The purpose of this exercise is to get you to use some of the GIS and R skills you are learning in the class to a basic, applied problem in geography. I have prepared the data from public data sources for this work.

You should be able to work through this exercise from what you learned in the readings and in class. In the .mxd file "G5050_Ex04", you will find several layers, which will get you started; I have downloaded, projected, clipped, and organized these for your use (*these are techniques that you can learn in other GIS classes—I have performed this work for the current assignment so that you can focus on the topic at hand*).

The Problem:
You are interested in exploring the distribution of tornadoes around the country; this sort of task is often used to guide policy and funding (such as sending funding to mitigation efforts or setting up an office to administer such efforts).

Objectives (A, C, D, E,G):
-Practice producing and comparing basic point pattern descriptions
-Practice producing a linear directional mean for a large body of data
-Transform and standardize data by aerial units from point data
-Import GIS data into R
-Construct a kernel density map from point data
-Perform a K-function, G-function, and L-function test on point data
-Transform some complex point pattern analysis into a readable report
-Keeping pushing on your projects!

Sources:
Tornado segment data were downloaded from the National Oceanographic and Atmospheric Administration's (NOAA) website (http://www.nws.noaa.gov/geodata/). I also took state and county boundary data from this URL, though these are originally from the US National Atlas.

Getting Started:

Download open the file Geog5050_ex04.mapx in *ArcPro*. You should find all the files you'll need for this work.

Some Notes about the Data Layers:

Frame: This is a simple polygon file that will serve as the extent boundary of your data. I also like to use frames like this to help the display (it serves as the source of the "back" layer as well).

StudyFrame: is a clipping frame for part of the exercise.

States: State boundaries for reference.

Counties: County boundaries for reference and analysis.

torn_tchdwn: These are points built from the NOAA database on tornadoes. Notice that the starting and ending points of the tornadoes are included in the line layer noted just below.

torn_vector: This shows a line feature class of the paths of tornadoes along with all their attribute information.

Part 01: Examine Changes in Tornado Distribution across Two Time Periods

[20 points]

Overview

You are interested in exploring the idea that the spatial patterns in tornadoes have changed over time. Perhaps climate change has affected patterns in tornados? You will build some spatial summaries of patterns in two periods: (A) 1951 to 1975 and (B) 1981 to 2005. The explicit research questions we're asking here is "have tornado patterns changed between 1951 and 2005. If so, how have they changed?"

Your job is to produce the **expected nearest neighbor (NN) distance**, **observed nearest neighbor distance**, the **NN ratio**, **z-score**, and **p-value** for both **groups A and B** so that you can compare them. You should also analyze the directional vectors of the tornados and report them as the **average length and directional distance** for each group. Finally, report the **distance between the mean centers of the two groups**.

I plan on going over the GIS component in class and / or making a video.

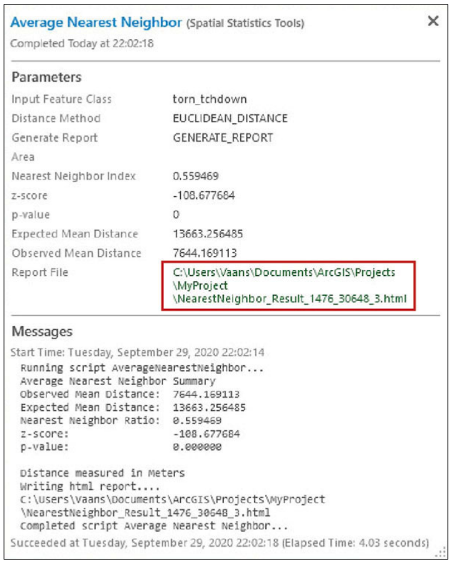
Steps

1. Start by making a new attribute in the torn_tchdwn and torn_vector layers. You want to separate them into your study periods with a new attribute, drawing from the attribute data on the year. Name the grouping field something like "group" and make the field a "text" field with a width of 1. Give the two groups the attribute "A" for 1951 to 1975 and "B" for 1981 to 2005 (you should be able to select those groups out with a basic select-by-attribute command; make sure you chose the range carefully to avoid including an extra year not in the range. Remember to use quotes in your field calculation to indicate text and avoid confusing the software).

2. Run separate nearest neighbor analyses on both study groups in the torn_tchdown layer. You should be able to find and run the tool easily yourself, but I've included step-by-step instructions in case you have trouble:

   A. Open the attribute table and select study group "A," which you just built. Keep the group selected – this ensures that the analysis will only occur on group A.
   B. Click on the analysis tab
   C. Click on the "tools" icon
   D. Search for "average nearest neighbor" and open the tool.
   E. Specify **torn_touchdown** as your input feature class
   F. Specify **Euclidean** as the distance method
   G. Toggle **generate report** on
   H. Click "Run"
   I. Once it runs, you should see a dialog box at the bottom of your screen, click on "view details"
   J. You should see a report window appear with the information from the analysis. You can cut and paste from this dialog box to get the information you need.

K. Click on “Report File” (see the image below) and it should open an html file in a browser with a full report.



Make a table in that shows the expected NN distance, the observed NN distance, the ratio, z-score, and p-value for both groups. (**Ex04_Answers.docx**)

3. Calculate the linear directional mean of the two study periods (label it “torn_ln_dir”). The interface is very similar as the tool described above, but it yields a new feature class, rather than a report. In this case, all you have to do is select “stdy_grp” under the case field option. Include a table in your responses with the average length and directional mean. (**Ex04_Answers.docx**).

4. Calculate standard deviational ellipses (use one standard deviation) for the touchdown points for the two groups (label it “tornpt_se”). You can find the tool by searching for “standard deviational ellipse.” Using the tool should be fairly self-evident from there. Don’t forget to specify the case field.

5. Add a point for the center of each ellipse. There are tools for calculating this, but it’s helpful to know how to work with output that includes coordinates. Just in case you can’t remember how to transform coordinate pairs into points, here are the steps:
(5A): Export the attribute table for tornpt_se
(5B): Add the exported table to your map when the software asks.
(5C): Right-click on the table and go to “display XY data.” This should open an “XY to Point” tool.
(5D): Specify the X and Y fields as “CenterX” and “CenterY,” respectively. Run the tool.
(5E): The export events should show up in your table of contents. Right click on that layer and select data→export features.
(5F): Specify where you would like the data saved and add it to your map. (call it “centerpts”)
(5G): Add the new layer to your map and remove the “events” you made in step 5E from your table of contents.

6.

Calculate or report the following in your table:
-The distance between the two points
-The standard distance for the x-axis for each group
-The standard distance for the y-axis for each group

7. Produce a basic map showing the standard deviational ellipses and center points of the two study groups. Make sure you include labels to clearly distinguish between the two. I recommend including touchdown data in way that does not interfere with the main idea of the map (the two ellipses and the center point) (**Ex04_TornMap.jpg**). Include map information (legends, title, etc.) so that it's legible to someone who has not written this exercise!

8. Write a paragraph (no more than a few sentences, please) summarizing your observations about the changes in the patterns across the two periods. (This is not meant to call for formal writing – I just want to see evidence that you're thinking about the output). Include any thoughts or ideas you have about the meaning of what the changes you've observed.

Save your GIS work after you complete part 01. I will draw from this work in the following parts and may come back to these data in future exercises.

Part 02: Produce a Kernel Density Map from Tornadoes in R *[10 points]*

Overview
For this part, I would like you to use what you've learned in R to perform some kernel density work on tornados, building from the work you did in the GIS in part 01 of this exercise. Your goal is to produce a basic isopleth map of tornadoes in R that shows state boundaries and is appropriately clipped. I'm also stepping through one of several ways to move between R and GIS.

Steps

9. **In your GIS**, export the following layers to shapefiles into ***<u>your R workspace</u>*** for the exercise. I've provided the correct names for the files in the code below so that you can ensure that the R code I've provided works correctly. The tool for exporting data into shapefiles from ArcGIS Pro is called "feature class to shapefile." Be careful with file management here…
   -States
   -StudyFrame
   -torn_tchdown

10. BEFORE YOU START working in R for this this section, make sure you've installed and loaded the following packages:
    sf
    tmap
    GISTools
    spatstats

I recommend that you work with an R script file, rather than typing directly into the command prompt. The first step here is getting the spatial data from the GIS into a form that R can work with. As the book describes, sf makes working with spatial data easy once you learn a bit of syntax. The first step is to load the layers into R with the following:

```
tornus.sf<-st_read("torn_tchdown.shp")
states.sf<-st_read("States.shp")
studyframe.sf<-st_read("StudyFrame.shp")
```

11. You can check the type of object you are working with the typeof() and class() commands

```
typeof(tornus.sf)
class(tornus.sf)
```

12. The book has some deprecated code (already!) and so I'll provide you with a little help to sort through this. Your goal is to produce a kernel density map of tornados, and a good way to do that is through the GISTools package, which primarily works through the sp format. So… we have convert our tornadoes and states layers from sf to sp objects (as described in the book).

```
tornus.sp<-as(tornus.sf,"Spatial")
states.sp<-as(states.sf,"Spatial")
frame.sp<-as(studyframe.sf,"Spatial")
```

13. You can check that everything is working by plotting it. (You might give your computer a little while to sort through this – it's a lot of processing!)

```
plot(tornus.sp,pch=16,cex=.05)
```

14. Now that you've got the data in R in sp, you produce a basic isopleth map of tornados in R. A good function for this comes from the GISTools package: kde.points(). You can read about the function here:

    https://www.rdocumentation.org/packages/GISTools/versions/0.7-4/topics/Kernel%20Density%20Estimates%20From%20Points

15. I recommend assigning an object and then plotting it as is done just below (I've provided the general form of the kde.points function – your job is to fill in the details!).

```
torn_kde<-kde.points(tornus.sp,30000,n=200,lims=NULL)
level.plot(torn_kde)
```

Experiment a bit with the bandwidth and resolution to get a good map. Once it's in your plot, you can throw down a map of states on top to give some important context:

```
plot(states.sp,alpha=0,add=T)
```

16. Indicate which bandwidths you used in your answers document (**Ex04_Answers.docx**) and briefly explain how you chose those parameters. Name your image file **Ex04_tornUS.jpg** and include it with your work.

Part 03: Run some K-Function Tests on Tornados in R *[25 points]*

Overview
For this part, I would like you to use what you've learned in R to perform some K-function tests on tornados, building off the work you did in the GIS. Once again, you'll be working with ***sf*** and a few other packages covered in the book. Your goal is to evaluate the degree of clustering of tornados in Iowa, as well as use the K-, G-, and L-functions to determine the distances at which clustering is significant. In order to make the processing manageable, I've directed you to clip the tornados to the *study frame* layer from the GIS.

Steps

17. Once again, make sure that you have loaded the right libraries for this: **sf**, **tmap**, **GISTools**, and **spatstat**. Note that this part continues from Part 02, so you should have the appropriate layers already loaded as sf objects in your workspace ("tornus.sf", "states.sf", and "studyframe.sf").

18. You should be able to run through an example in your book (check Chapter 03) and figure out how to clip the tornado data to the study area and plot a nice little map that shows the state boundaries. Perhaps the only tricky bit is how to get the intersection to work. To clip tornadoes with the study area, try the following:

```
tornus_clip <- tornus[studyframe,]
```

19. That will pull out only the points within the frame. Because it can't just select the states within the
study area and you want to perform a true intersection using the following:

```
states_clip <- st_intersection(frame,states)
```

20. You can check that everything is in good shape by running through some of the scripting to map things (from Chapter 3, once again).

```
tm_shape(studyframe.sf)+
  tm_borders(col="black") +
  tm_layout(frame=F) +
  tm_shape(tornus_clip) +
  tm_dots(col = NA, size = 0.02, shape = 16, title = NA,
  legend.show = TRUE, legend.is.portrait = TRUE) +
  tm_shape(states_clip) +
  tm_borders(col="grey70",lw=2) +
  tm_layout (frame=F)
```

21. The problem is that it's still not a ppp object and the multi-distance clustering functions using the spatstat package uses ppp objects. As I've mentioned a few times now, this is the sort of problem you'll encounter quite a lot; there's often no clear solution and you will probably have to hack away at solutions until one works. With a few minutes of internet sleuthing, you will find out that presently the only way to convert an sf object to a ppp object is to **(a)** first convert it to first convert it to an sp object, and then **(b)** turn that into a ppp.

```
# Transform the clipped, sf layers into sp objects
tornus.sp <- as(tornus_clip, "Spatial")

# Transform the sp object into a ppp object
tornus.ppp <- as(tornus.sp, "ppp")

# Check it…
class(tornus.ppp)
```

22. Now that your data are in the right format, you can follow the book closely (look in section 6.5.1). Produce a Ripley's K function plot <u>with an envelope </u>and submit it with your assignment. You can try improving the labels a bit (such as giving it a sensible title) to make it easier to read. Submit a copy of the graphic with your assignment. (**Ex04_KPlot.jpg**).

23. Also produce the L-function equivalent and submit the graphic. (**Ex04_LPlot.jpg**).

24. Finally, mix up your assumptions a bit by producing a G-plot and submit the graphic (**Ex04_GPlot.jpg**).

25. Run a test to see if there is statistically significant clustering, based on the K-function model (use one of the tests described in Chapter 06). Applying the "boss-afraid-of-math scenario" (e.g., use English that an educated person not trained in statistics can understand), briefly describe the test, what it does, and what you found.

26. Use the output to write a very brief report evaluating the research question: "Do tornados in the plains region of the Midwest around Iowa exhibit spatial clustering?" Once again, you're writing this for your boss, so you need to explain to him or her what you're doing in language she or he can understand. Explain what the graphics mean and make a conclusion. Based on your work, what distances (these are in meters) does there appear to

be the most clustering? Be sure to explain the differences in the results between the K-function and the G-function, if you detect one. *Limit your response to 500 words*. Add your response to **Ex04_Answers.docx**.

Part 04: Painless Project Progress [5 points]

Record your research question as it currently stands; I know that I ask this frequently but appreciate that this is an evolving process!

Briefly describe your analytical plan – namely how do you answer the research question? Explain what basic concept you are examining (for example, whether an association, a test for difference, clustering or pattern analysis) and how you will use statistical or spatial statistical tools to answer the question. Please limit your response to half a page.

Summary of What to Turn In

Part 01:
-A basic map of the directional ellipses for each study group. Label the ellipses well and include the raw point data in a way that does not interfere with the main messages of the maps.
Ex04_TornMap.jpg.
-A word document containing a paragraph with your tables, observations, and question responses:
Ex04_Answers.docx. *Please number your responses in your answers document to match the assignment item to which it corresponds.*

Part 02:
Ex04_tornUS.jpg
Include question responses in **Ex04_Answers.docx.** *(I'm only asking for a couple of sentences for this part).*

Part 03:
Ex04_KPlot.jpg
Ex04_LPlot.jpg
Ex04_GPlot.jpg
Include responses and your tornado mini report in **Ex04_Answers.docx.**

Part 04:
Include responses in **Ex04_Answers.docx.**