**Exercise 4 – Part 1**

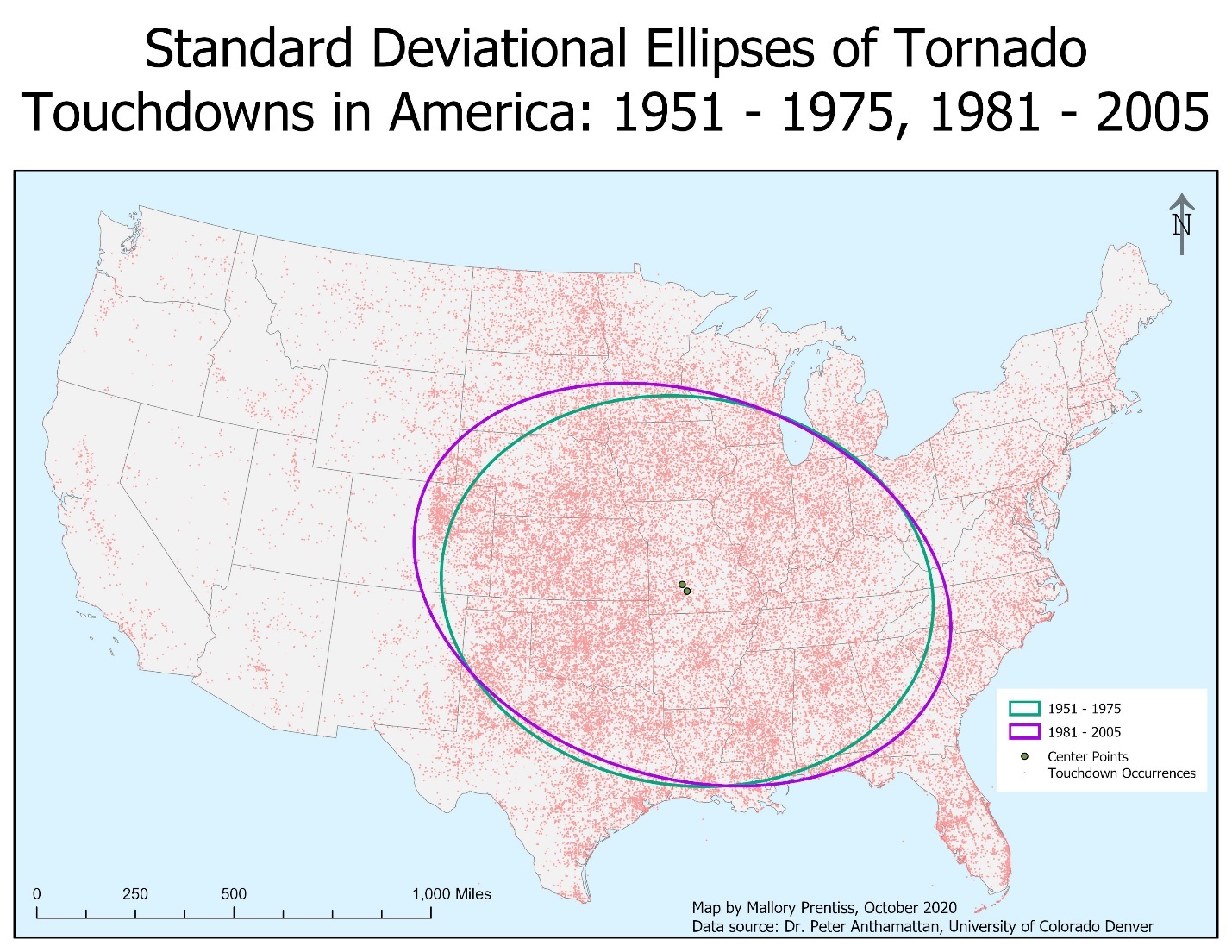
*Nearest Neighbor (Euclidean)*

|  |  |  |
| --- | --- | --- |
|  | **Group A** | **Group B** |
| **Expected NN Distance** | 13,663.2565 meters | 10,628.6282 meters |
| **Observed NN Distance** | 7,644.1691 meters | 5,956.1504 meters |
| **NN Ratio** | 0.559469 | 0.560387 |
| **Z-Score** | -108.677684 | -138.483399 |
| **P-Value** | 0.000000 | 0.000000 |

*Directional Mean*

|  |  |  |
| --- | --- | --- |
|  | **Group A** | **Group B** |
| **Average Length** | 7,041.686515 | 4,008.045854 |
| **Directional Mean** | 41.939648 | 41.71113 |

*Standard Deviational Ellipses*



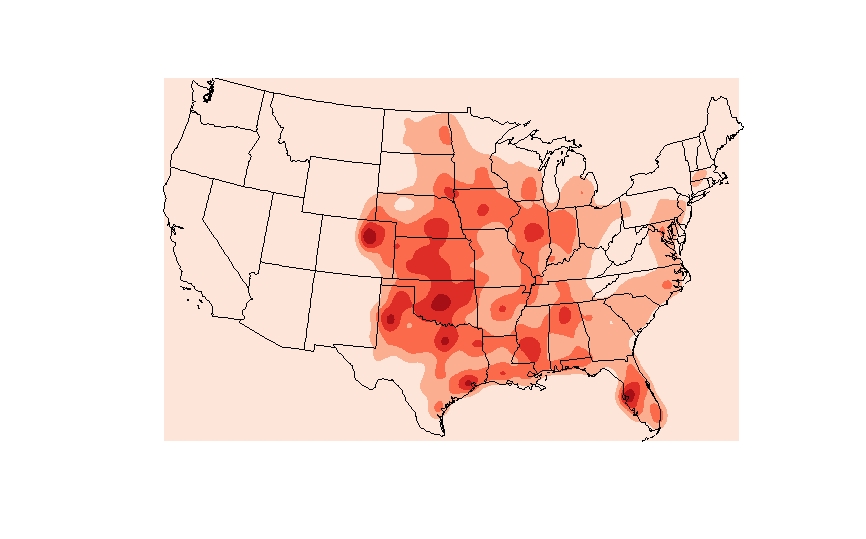
*Center Point Distances*

|  |  |  |
| --- | --- | --- |
|  | **Group A** | **Group B** |
| **Distance Between Points (m)** | 33,016.26 | 33,016.26 |
| **Standard Dist. – X-Axis (m)** | 1,008,146.417809 | 1,123,483.643352 |
| **Standard Dist. – Y-Axis (m)** | 792,503.453845 | 783,688.707917 |

*Observations*

The change in the ellipses and spread of the tornado touchdowns over the two time periods can be seen in how the range of group A (1951 – 1975) is smaller than group B (1981 – 2005). The ellipse for group B is elongated and almost entirely encompasses that of group A. This means that there are more tornado touchdowns in group B and that the dispersion is larger.

**Exercise 4 – Part 2**

The bandwidth that I chose for this part of the exercise is 200,000 with n=500. I chose this because I wanted the data to be fairly granular so that the differences between hotspots and lighter touchdown areas were easier to identify. The more that I increased the bandwidth, the greater clarity was brought to the plot, and as such I concluded with the image below.

**Exercise 4 – Part 3**

25. I ran a MAD Test on the K-Test and received a p-value of 0.01. This means that we can assume that tornadoes are clustered and that this isn’t due to random error. The MAD test stands for the Maximum Envelope Distance and the purpose is to verify the results of the k test. For the k test, it looks to find the distance at which the expected values is the furthest from the observed line. This value clues us into whether or not we can attribute any randomness to this particular result.

26. In response to the research question asking about whether there is spatial clustering of tornadoes in and around Iowa, my answer would be yes. I drew this conclusion from the multiple tests that I ran and the output values that I received, which I will explain using the graphics produced. Starting with the K-test, this test examines clustering of point data by looking at the distances between points. From the graphic we can see that clustering increases as the search radius of the function increases—meaning that as we look closer, there doesn’t appear to be any clustering, but as we expand our range of scope the clustering increases. On the graphic, the greatest distance between values (which shows clustering) is at 250km and the smallest distance is at 0-50km. Conversely, the G-test compares the probability of finding a data points nearest neighbor, or closest point, within a given search radius. Looking at the G-test graphic, the data we observed is plotted above the data we would’ve expected if there were random spatial patterns. This means that the data is clustered within the given search distance. Between the K and G tests, we can see similar trends of clustering with tornadoes in the study area. Finally, the MAD test looks to verify the result of the K test. With an output p-value of 0.01, this means that the pattern of clustering that we observed in the previous tests is not due to random error.

**Exercise 4 - Part 4**

The research question I will be looking at for my project is: how are the year that a building was built in Lower Downtown and the size of the building associated? The concept that I will be analyzing initially is if there is an association between how old a building is and its size. If I am able to identify a connection between these two variables, then I will also look at to see if there are any clustering patterns. To test the association, I will run a regression and use Pearson’s R. For the cluster analysis, I will look at spatial autocorrelation to tell if the size of buildings and/or year built show any clustering trends. The test for this is Moran’s I.

The data that I will be using for this project I have been working with on my MA thesis. I have parcel shapefiles that I pulled from the Denver Open Data Catalog that I will be using to pull the attributes of interest. To double check the square footage, I’ll compare the attribute data from the parcel shapefiles with current information from the Denver Property Taxation and Assessment System. This will help me answer my research question by allowing me to look at the output values from each test to verify whether or not there is an association between the two variables.