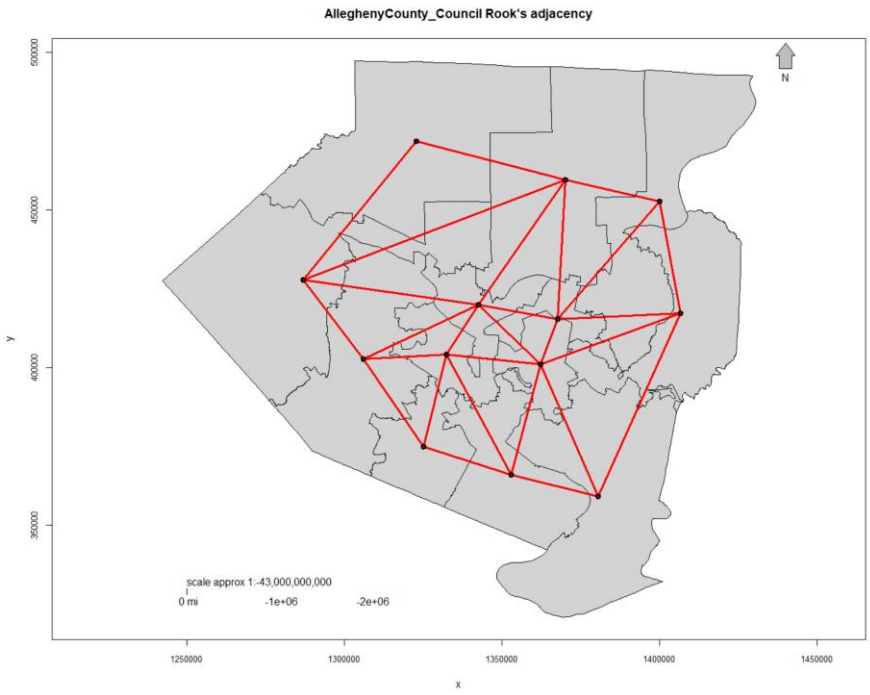
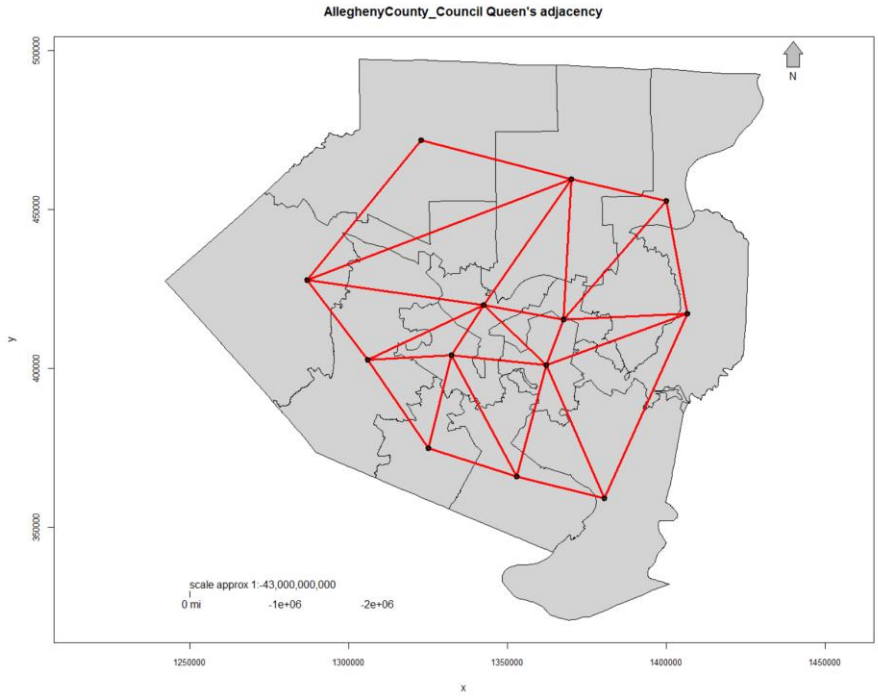
INFSCI 2809: Spatial Data Analytics

Project 4 Report

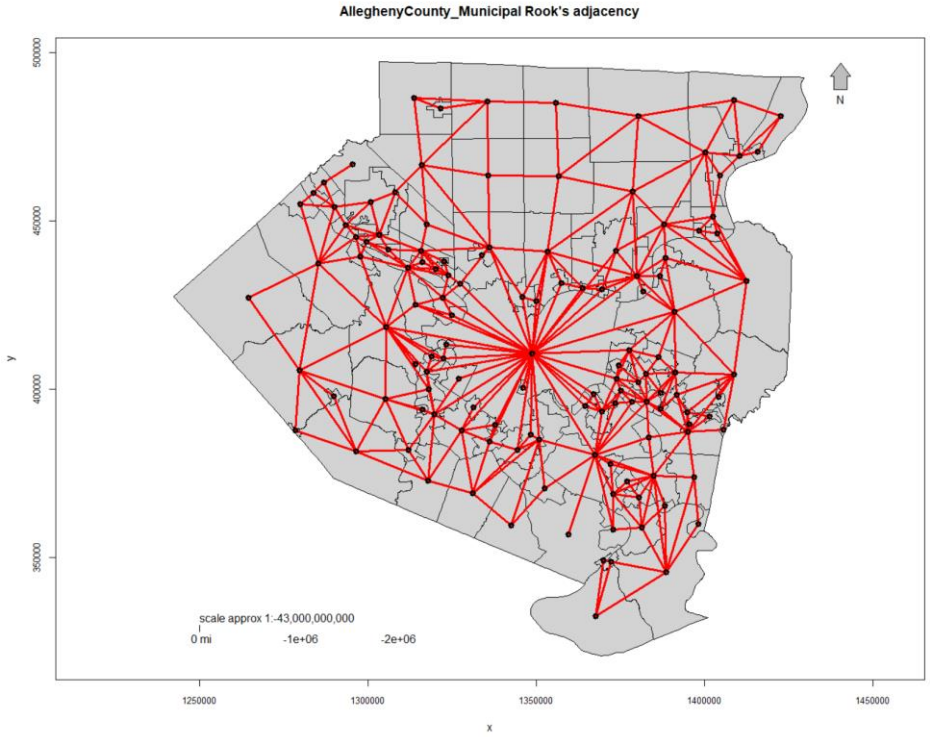
Student: Zechen Wang (zew20)

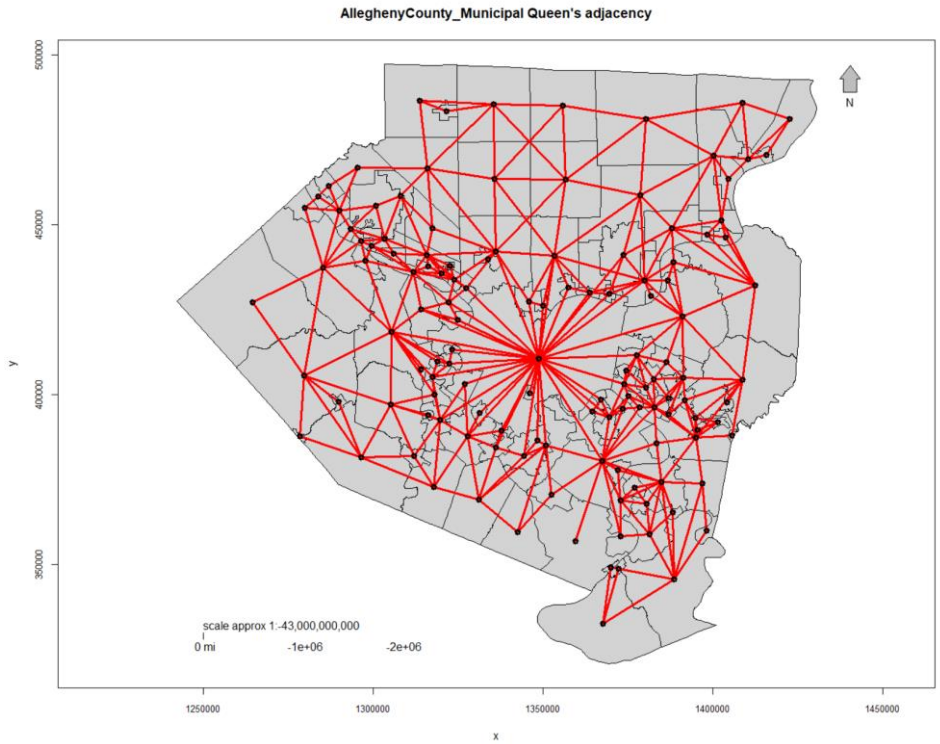
I. Part A





AlleghenyCounty_Council	Moran's I	Geary's C
Rook's adjacency	0.04388988	1.004298
Queen's adjacency	0.04388988	1.004298





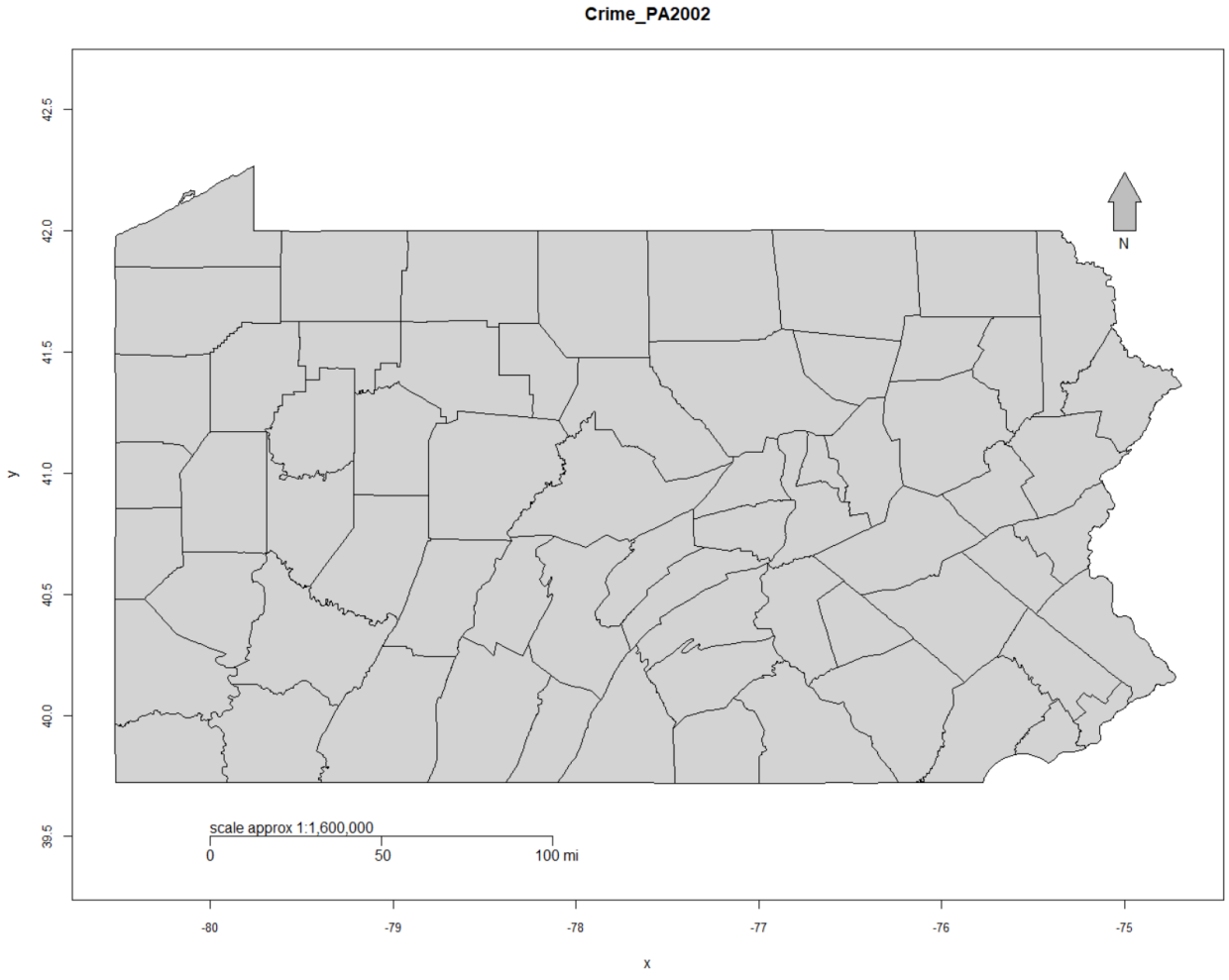
AlleghenyCounty_Municipal	Moran's I	Geary's C
Rook's adjacency	-0.00132743	2.142808
Queen's adjacency	0.001530271	2.099205

Summary:

For AlleghenyCounty_Council.shp file, the Rook's and Queen's adjacency are same with each other, therefore, the Moran's I and Geary's C are the same between them. The Moran's I is 0.04388988, which indicates a weak positive autocorrelation. The Geary's C is 1.004298, it is more than 1, so it indicates a negative autocorrelation.

For AlleghenyCounty_Municipal.shp file, the Moran's I of Rook's adjacency is -0.00132743, which indicates a weak negative autocorrelation, but the Moran's I of Queen's adjacency is 0.001530271, which indicates a weak positive autocorrelation. Both Geary's C of the Rook's and Queen's adjacency is larger than 1, which indicates negative autocorrelation.

II. Part B



Global G statistic is 0.1562118.

For Geographically Weighted Regression (GWR), I use a Gaussian kernel function to assign weights, the weights are:

0	1	2	3	4	5	6	7	8	9	10	11	12	13
0.016070996	0.010949067	0.037774863	0.008258608	0.152873444	0.090300392	0.026936110	0.261936657	0.048745458	0.174186070	0.061567734	0.027501429	0.238516855	0.010092030
14	15	16	17	18	19	20	21	22	23	24	25	26	27
0.167688273	0.231531802	0.071448529	0.029523644	0.051498097	0.255958015	0.066237278	0.133919145	0.009612911	0.027289278	0.070630534	0.366327252	0.134969057	0.302566117
28	29	30	31	32	33	34	35	36	37	38	39	40	41
0.009258519	0.216979579	0.009499606	0.314271766	0.068834758	0.024108276	0.280411910	0.226602088	0.278905008	0.120035284	0.011427871	0.039957697	0.010028780	0.016533626
42	43	44	45	46	47	48	49	50	51	52	53	54	55
0.093995776	0.029338439	0.015334070	0.056938947	0.133089569	0.026877542	0.036933484	0.108882573	0.190359435	0.073483541	0.201648892	0.123438578	0.328474296	0.256626524
56	57	58	59	60	61	62	63	64	65	66			
0.321420660	0.148837801	0.372647134	0.338857461	0.398942280	0.275894437	0.388628960	0.198422187	0.366487587	0.339543835	0.211662608			

Then, by using this formula,

$$\mathbf{b} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{W}\mathbf{X}^{\mathrm{T}}\mathbf{y} \qquad (8.11)$$

The regression coefficient estimates **b** is:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	1.302098e+00	4.320048e-01	6.689398e+00	-4.104531e-02	8.598358e+00	-1.499361e+01	3.582786e+00	-1.943861e+01	-1.460124e+01	7.405568e+00	1.382108e+01	7.271886e+02
[2,]	1.782905e-06	1.120139e-05	4.458957e-05	3.219597e-05	-2.860160e-05	-7.694436e-05	2.182309e-05	1.996226e-05	-5.770771e-04	-2.891284e-04	-5.527741e-04	7.403701e-05
[3,]	-5.538467e-01	-4.551551e-01	-3.055567e-01	-1.314029e-01	-9.826511e+00	6.029990e+00	-6.286093e-01	-3.262339e+00	-1.512703e-01	1.494293e+01	4.166714e+00	2.105565e+00
[4,]	6.237686e+01	-5.371485e+00	-2.042737e+02	7.006387e-01	-1.967904e+02	2.337966e+02	-2.296348e+01	1.922776e+03	4.816518e+01	-4.998264e+02	-1.385787e+03	-5.516108e+02

	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
[1,]	5.516822e+01	-3.823454e-01	0.0825971324	2.914423e+01	1.729034e+01	2.298535e+00	6.343417e+01	-8.019589e+01	2.333295e+01	25.437381272	2.926411e-01	-3.287982e-01
[2,]	-9.076501e-05	-1.062762e-06	-0.0001731948	-3.931576e-04	-5.489128e-05	2.196584e-05	5.168625e-04	1.873086e-05	-2.694992e-06	-0.000635417	-8.618042e-06	1.704605e-05
[3,]	-2.520206e-01	-7.601406e-03	-0.3512375671	-9.731664e-02	-3.429993e+00	-1.878041e+00	-1.316869e+00	-9.046963e+00	-6.297253e-02	0.197631235	3.174504e-01	-1.335153e-02
[4,]	-1.097824e+02	2.613249e+01	83.0135158935	-1.623086e+03	-6.612562e+01	6.583492e-01	-5.464062e+01	7.643415e+01	-2.381870e+00	-2.159631802	-7.321944e-02	2.549347e+01

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]	[,36]
[1,]	2.771290e+01	-4.596675e+02	3.1644645195	8.350894e+02	2.683847e+01	-1.456554e+01	1.084292e+00	36.7591144768	-1.547392e+02	-66.740084433	224.79891925	-9.163273e+01
[2,]	4.826466e-05	-1.212683e-04	0.0003527353	-3.797914e-04	4.160447e-05	-2.379528e-04	-1.651878e-05	0.0007351024	1.569463e-05	0.000323116	-0.00121549	6.509594e-05
[3,]	-1.729808e+00	-1.008642e+00	-6.1017871166	2.745580e+01	-8.835162e-01	4.655643e+00	5.670720e-01	-1.7874586197	-7.010975e-02	1.925680948	0.37331797	-1.636401e+00
[4,]	7.282112e+00	7.425612e+01	69.7220289678	-2.096862e+04	-7.741145e+00	-1.148024e+01	-4.940879e+00	25.7563247790	5.920617e+01	-17.735992938	-765.75191033	4.308557e+02

	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]	[,43]	[,44]	[,45]	[,46]	[,47]	[,48]
[1,]	1.064192e+02	-2.117289e+01	-1.666388e+01	1.627403e+00	3.379372e+00	-8.516465e-01	-2.767583e+01	-1.279385e+00	1.222563e+02	-65.766053734	-1.942978e+01	6.386980e+00
[2,]	-2.959652e-05	-9.180515e-06	-2.480468e-05	6.724028e-05	5.689440e-06	1.758971e-04	8.061068e-06	4.066631e-04	7.669376e-03	0.001335185	1.935458e-06	8.623303e-05
[3,]	-2.990719e-01	7.362744e-02	2.257810e-01	-1.833451e-01	-2.042931e-01	-1.076848e+00	4.153531e-01	-2.472091e-02	-4.420279e+02	0.159753125	1.010741e+00	1.637330e-01
[4,]	-4.525537e+02	2.696552e+02	2.672015e+00	1.225390e+01	-9.234749e+00	-4.502128e+02	4.304875e+01	-6.077112e+02	-2.650168e+02	-2.540630708	6.445868e+01	-1.788614e+02

	[,49]	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]	[,57]	[,58]	[,59]	[,60]
[1,]	3.546719e+00	-4.319434e+00	-3.630265e+02	2.136247e+02	-4.365130e+01	-9.735921e+01	-1.430315e+00	1.035045e+02	5.036046e+02	-1.493094e+01	-2.522876e+02	1.335319e+02
[2,]	5.383540e-06	-1.136723e-04	-6.831507e-05	-2.148981e-06	-5.390676e-07	-1.105366e-05	-6.155528e-06	-4.195768e-05	9.299484e-06	-2.264441e-04	1.478148e-03	1.239459e-05
[3,]	-9.043998e-02	-5.335033e-02	-3.086923e-01	-9.280561e-01	-2.831081e+00	7.461140e-01	-2.893118e+00	3.207578e-01	-1.093574e+00	5.097278e+00	-1.218806e+02	-2.815491e+00
[4,]	2.452646e+02	1.084715e+03	1.537009e+03	-3.156696e+02	4.312429e+02	1.519192e+02	3.504680e+03	-3.607399e+02	-6.714949e+02	1.284968e+02	1.056066e+03	-1.783726e+03

	[,61]	[,62]	[,63]	[,64]	[,65]	[,66]	[,67]
[1,]	4.735855e+02	6.365030e+02	2.081876e+03	2.272281e+01	1.506079e+02	2.416814e+01	-8.743844e+01
[2,]	-4.569864e-05	9.651197e-03	3.217327e-05	-5.846776e-05	4.079235e-05	-9.004448e-06	2.919452e-04
[3,]	-2.593245e+01	1.668518e+01	-3.848233e+00	3.567105e+00	-9.350466e+00	4.212223e+01	-1.116342e+01
[4,]	-4.311922e+03	-2.354295e+03	-1.683089e+02	-2.749280e+02	-5.260187e+02	-3.916931e+01	1.362110e+03

Then, we use the formula **y = Xb + e** to predict the total number of crimes in Mifflin County, and the prediction is 308.3213, which is close to the actual value 215.

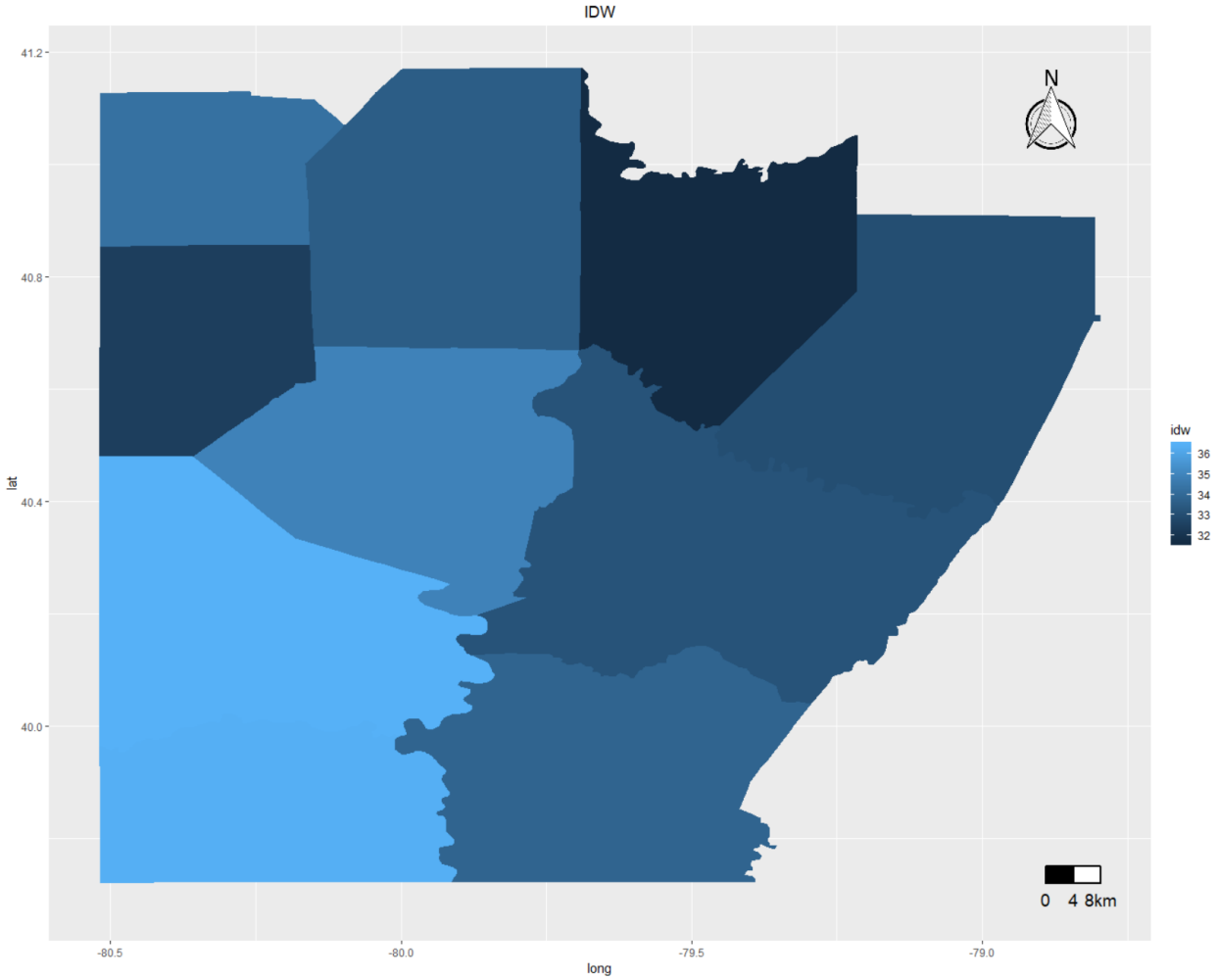
Summary:

The global G statistic is 0.1562118, it is not a high value, so it indicates the locations where low values are located near one another outweigh the locations where high values are located near one another.

For GWR, if I did not use a Gaussian kernel function to assign weights, the prediction of Mifflin County would be 3018, which is far more than the actual value 215, but after I use the Gaussian kernel function to assign weights, the prediction now is 308.3213, which is much closer to the actual value.

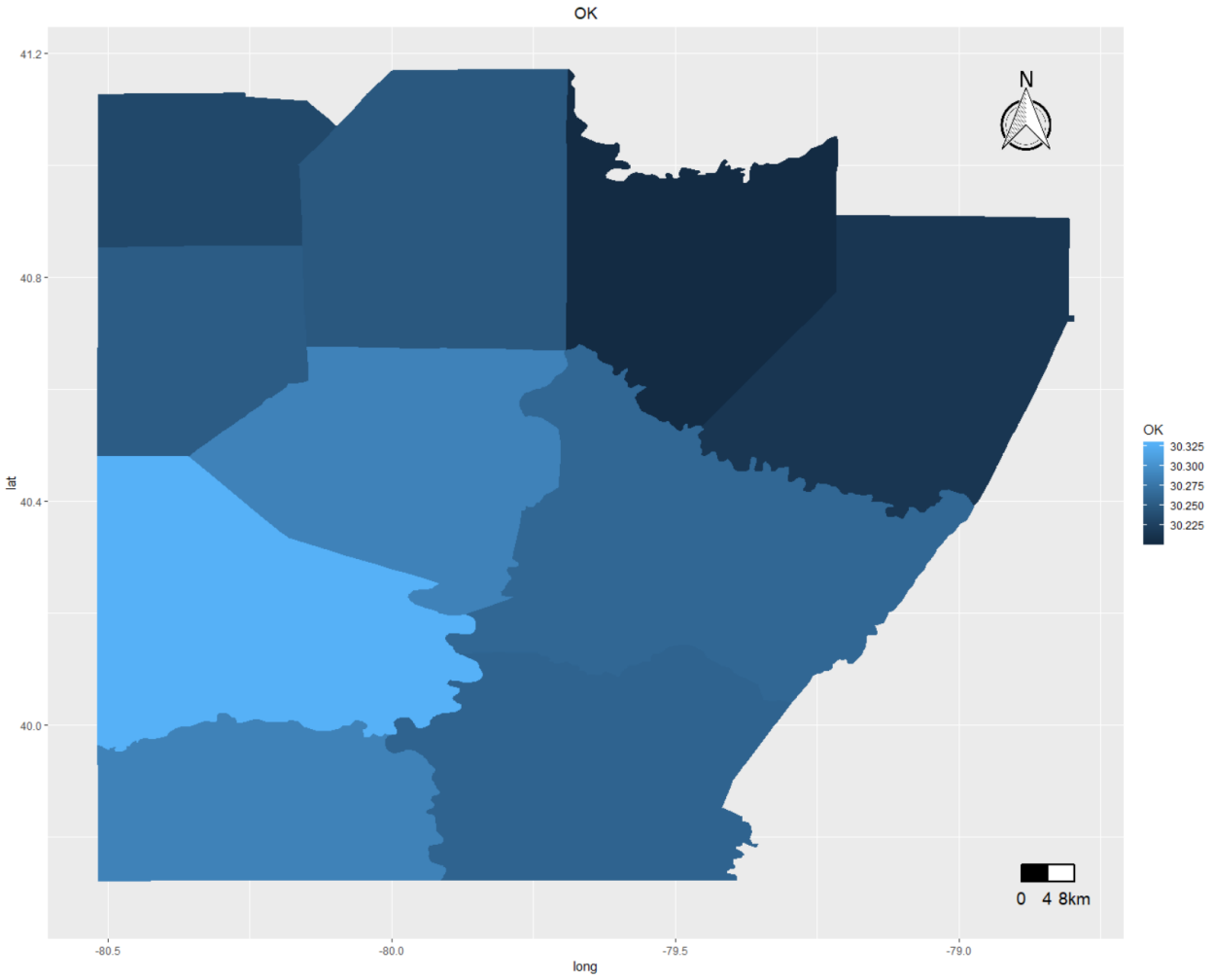
III. Part C

IDW:



	COUNTY	IDW
0	Butler County	33.53768
1	Armstrong County	31.62437
2	Indiana County	32.9769
3	Lawrence County	34.24218
4	Beaver County	32.39437
5	Westmoreland County	33.1684
6	Allegheny County	34.85891
7	Washington County	36.40584
8	Fayette County	33.78217
9	Greene County	36.32539

OK:



	COUNTY	OK
0	Butler County	30.24764
1	Armstrong County	30.20309
2	Indiana County	30.21493
3	Lawrence County	30.23216
4	Beaver County	30.25208
5	Westmoreland County	30.26212
6	Allegheny County	30.28668
7	Washington County	30.32653
8	Fayette County	30.25845
9	Greene County	30.28643

Summary:

	COUNTY	IDW	OK
0	Butler County	33.53768	30.24764
1	Armstrong County	31.62437	30.20309
2	Indiana County	32.9769	30.21493
3	Lawrence County	34.24218	30.23216
4	Beaver County	32.39437	30.25208
5	Westmoreland County	33.1684	30.26212
6	Allegheny County	34.85891	30.28668
7	Washington County	36.40584	30.32653
8	Fayette County	33.78217	30.25845
9	Greene County	36.32539	30.28643

We can see from the summary table, the data in OK column changes very little, but the data in IDW column changes a lot. And the interpolation surface maps are little bit different between each other.

When I do this part, I use the same γ(d) in the textbook, which is:

$$\gamma(d) = 0 + 60(d) \qquad (10.20)$$

As the book says, this is a very simple unbounded linear model in which the semivariance increases by 60 variance units for every unit increase in distance and there is no nugget. Therefore, my model might not be the best model for the Ordinary Kriging.

And from this part, I think it is hard to tell which method is better because the data set is very small (only 10 Counties), and the “bulls-eye” effects (the effect the textbook used to compare these two methods) do not appear in the IDW interpolation surface map here.