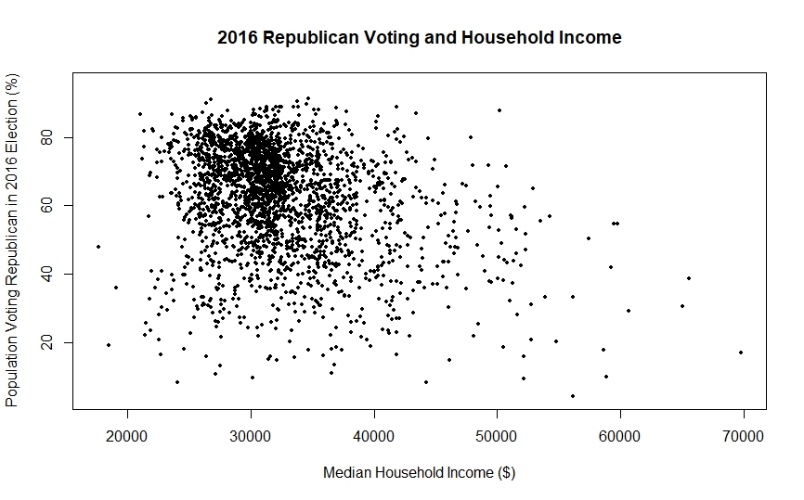
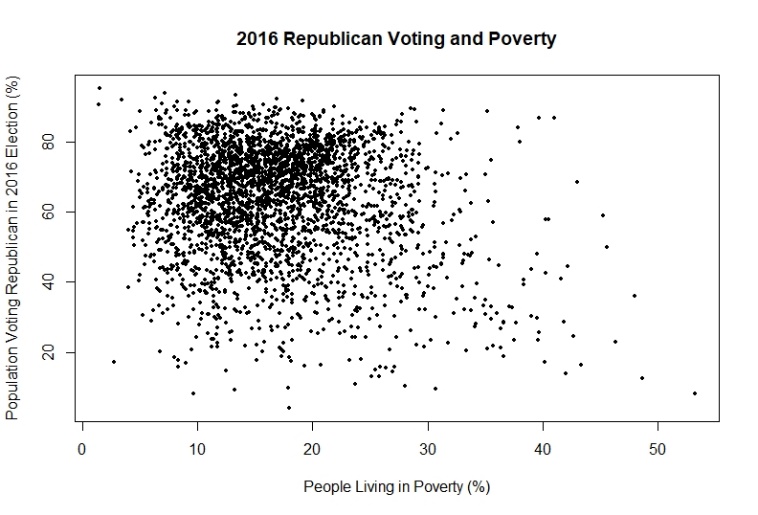
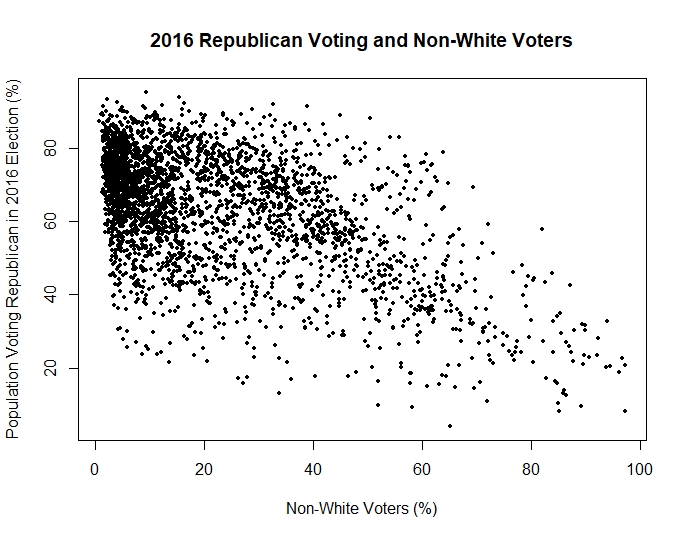
**Exercise 6**

*Multivariate Voting Analysis*

In examining some of the variables in the provided 2016 voting behavior table, I first looked at how Republican voting behavior (E2016\_RPct2) was associated with the median income for 2016 (S\_MedEarn2015). I was surprised to see in the plot that there generally looked to be a trend that the lower the median household income, the higher the percentage that that census area would vote Republican. I believe one reason the data might’ve shown this trend is due to Donald Trump’s policies that were aimed at promoting the livelihoods of blue-collar workers. To further look into how the two variables were correlated, I ran a summary of the linear regression that was used to plot the graph and came up with an r-squared value of 0.077. I found this to mean that 7.7% of the variation in Republican voting percentage can be attributed to the variation in the median household income amount. This didn’t seem to be very statistically significant or as though the two variables were overly correlated with one another.

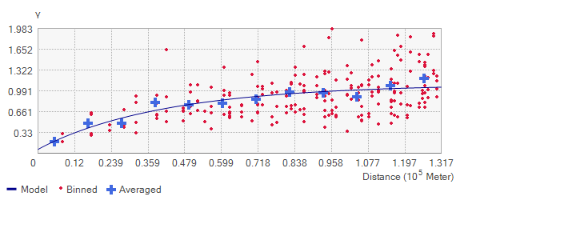
The next variable that I ran against the percent of people in the US voting Republican was the percent of people per census tract living in poverty. I found this to be more consistent with what I know about the common demographics of those voting Republican: according to the graphic, the less poverty there is within a census tract, the higher percentage of people there are voting Republican in that area. To confirm this, I ran the summary within R Studio between the two variables and found that the p-value was very close to zero (< 2.2e-16). I interpreted this to mean that there is an association between Republican voting behavior in 2016 and percent poverty and the null hypothesis could be rejected. The slope, or strength of effect, is -0.36874, meaning that the association is negative, meaning that as poverty increases, the percent voting Republican decreases.

The third variable that I decided to run against Republican voting percentage in 2016 is the percent of people that are non-white. This plot was very telling of the voting trend of this population. As seen in the graphic, there is a strong negative association between the two variables: as the percent of non-white voters in an area increases, the percentage of Republican votes decreases. The r-squared value was 0.293 in my summary, meaning that 29.3% of the variation in Republican voting trends can be attributed to the variation in the non-white voting population. This is extremely high and further contributes to the evidence that there is a very strong correlation between these two variables. In 2016, Trump’s primary supporting demographic was of a white voting population, meaning that this data follows the more general trends of the election.

The final thing I did with this data was run a summary of the linear regression of all variables added together. This reported back a large comparison of all three of my x-variables with the Republican voting behavior (y-variable) that confirmed the above analysis. The strongest apparent relationship appears to be between Republican voting behavior and percent non-white, followed by the percent of poverty, and finally the weakest relationship is with the median household income.

1. Slope: -0.24224  
   R-squared: 0.2376  
   Intercept: 73.70136  
   P-Value: 2.2e-16
2. The coefficient refers to how accurate the predicted value is in relation to the actual value. With this, the general relationship between percent voting Republican and percent urban is that in larger and more urban areas, the coefficient value becomes large and negative; conversely, the coefficient value becomes positive and is increased as the areas become more rural. This means that in places where the values are negative, the prediction is that the area will vote more Republican than the actual numbers report. The result is that many of the census tracts have a negative coefficient, meaning that there were less Republicans voting in these areas as was predicted given the percent of urban areas in each tract. The directionality of the data shows that the bigger cities in the US are more negative while the more rural areas (particularly in the Midwest) are more positive.  
   Some exceptions are Florida and the Seattle-area: Florida is typically pretty Republican and rural, but the data shows that the state as a whole is all pretty negative. With Seattle, as this is a large city, the prediction would be that Republican voting would be fairly low, meaning that this area should have a darker and lower coefficient. The result, though, is that the area is actually positive, meaning that there were more Republicans voting in this area than were predicted, given the urban area in its proximity.
3. The percent urban variable does the best job explaining Republican voting in central/northern Colorado, western Texas, and in the more rural parts of central Virginia.  
   The percent urban does the worst job in southwestern Oregon, near San Diego, northern Nebraska, northern Maine, and the coastal area from Providence, RI down to Philadelphia, PA. The best output for evaluating where the percent urban does the best and worst job of explaining Republican voting tendencies is using the standard error of the urban percent variable.
4. In the residuals there is a slight pattern seen in that many of the midwestern values are close to one while the rest of the country varies its residual values more greatly. This implies that the model predicted more accurate values in this region than in other regions. One reason this might have occurred is because there are less large urban areas in the Midwest compared to the coasts of the United States.
5. In summary, there is a negative correlation between percent of Americans that voted Republican in the 2016 election and percent urban area. This means that as the percent of urban area increases in a census tract, the percent of people voting Republican decreases. This is reflected in the slope that was observed in question one of this part. The intercept tells us that given 0% urban area, it can be expected that 73.7% of people in a census tract will have voted Republican in 2016. The r-squared value implies that about 24% of the variation in voting can be attributed to the variation in urban area percentage. In relation to the coefficient and error in predicted values versus actual values, the general trend in the data was that most of the larger urban areas were underpredicted in the level of Republican voting that would occur while the more rural areas were slightly underpredicted. The locations where the percent urban does the best and worst job explaining Republican voting trends in this election are varied throughout the country. Finally, the residual values produced by the GWR showed us that the test worked best at predicting and getting closer to the actual values in the Midwest region of the United States.

Part 3

1. The range of the semivariogram looks like it’s at about 75,000 meters or about 75 km.
2. 
3. I think the K-Bessel type creates the best line of fit for these data.
4. Lag size: 10,974.3762469786  
   Range: 131,692.514963743  
   Partial sill: 1.040399943115  
   Our range estimate was much different from the range reported in the method results.
5. The nugget and the sill would be closer in their values if not the same because zero spatial autocorrelation would mean that the point distribution on the graph are varied and all over the place. The model line would thus likely be flat, meaning that where it begins and where it ‘peaks’ at the location of the sill would be the same since they’re both based off the y-axis.