**COIS 3510H**

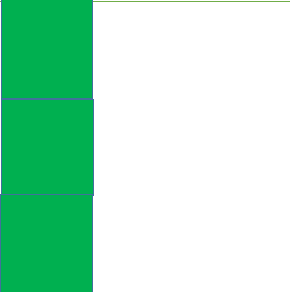
**STUDENT NAME**

**DATE**

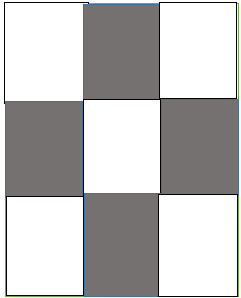
**Introduction:**

In Geography, spatial autocorrelation determines how similar one data point is to other data points around it. According to Morans *I* formula, spatial correlation can be categorized as either positive, negative or no spatial correlation. From these, we can draw three types of autocorrelation that exists.

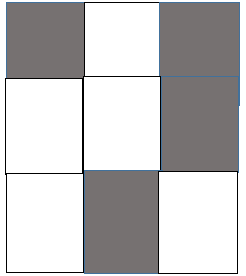
**Positive autocorrelation**; is characterise by perfect similarity of objects surrounding each other, which gives perfect clustering of the objects and subjects in question. This would look like something below:



Positive correlation gives birth to perfect clustered data points with closely related data sticking nearly close to each other.



Negative spatial correlation leads to the dispersal of data, within data points distributed all over across the data points



A spatial correlation of 0 indicates randomness and that the distribution of data points are random and are not related in any way across the data points.

**Question One**

**Case study:**

Health inequalities revealed from analysis of spatial dependence in US Counties.

In this study, an analysis of the spatial autocorrelation was done on the basis of health rankings of some of the counties in the USA and the frequency or pattern of Covid 19 cases and Covid 19 related deaths across the counties in the United States. Using some of the publicly available records and data, analysis of the spread of the disease was done based on some of the variables in USA county health rankings like the intensive care unit beds, the county socio-demographic details and finally the health care resources found/available per county. The map distributions of these variables are as indicated below:

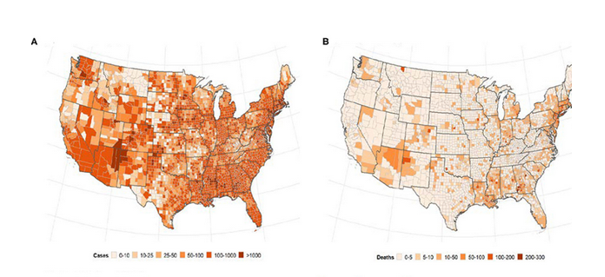
**Discussion:**

According to Saffary et al (2021) the results from this study indicate quite a number of geographical variations from the occurring distribution of covid cases across the counties. A positive correlation was spotted when the data was subjected into the Moran I formula that indicted a relationship of =0.174 and 0.264, p<0.0001 among black Americans and the Covid virus. What this means is that there were more black Americans suffering from Covid 19 related deaths according to the research done.

The study also further reveals that there was a positive partial autocorrelation between the black American deaths as a result of covid 19 across the counties (Moran *I* = 0.264, p< 0.0001). Moreover, there wasn’t any significant detected spatial correlation between the race of Hispanic and the covid 19 cases and related deaths. None Hispanic whites all the same showed negatively spatial correlation in cases (Moran I = -0.203, p < 0.001 and then deaths at (Moran *I* = -0.137, p < 0.0001) as a result of the covid 19. The study further reveals a very weak spatial autocorrelation between the quantity of intensive care units and the number of Covid 19 spikes shown at (Moran I = 0.08, p <0.001 and then deaths at (Moran I = 0.15, p <0.0001).In summary therefore, blacks and Hispanic races across the US, suffered more from Covid across the counties. Further there was a weak correlation between the number of intensive care units and the number of Covid 19 cases.

**Map summary and descriptions:**

The summary shows the distribution of the diseases across the 3108 contiguous counties as per the study. Few cases in Map A as a result of the separation brought by the central counties made significant shows case that there was high disease infection rates between the Eastern and western sides per 100,000 of people. The number of deaths per 100,000 people as shown in Map B also revealed a strong correlation between the Covid 19 infections and the counties in the Western and Eastern side. In terms of correlation, since most of the Hispanic and blacks live in the East and Western sides of the states, (regions most dominated by the blacks and Hispanic), these regions elicited higher covid 19 prevalence, indicating a stronger correlation between race and locations.

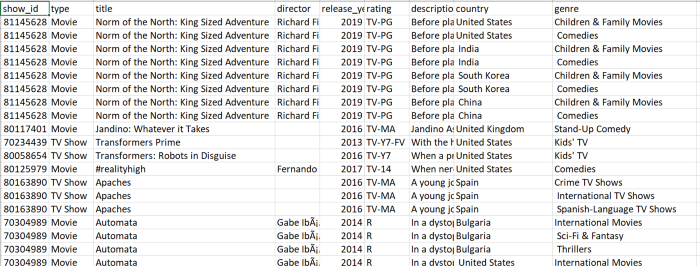
****

**Question Two:**

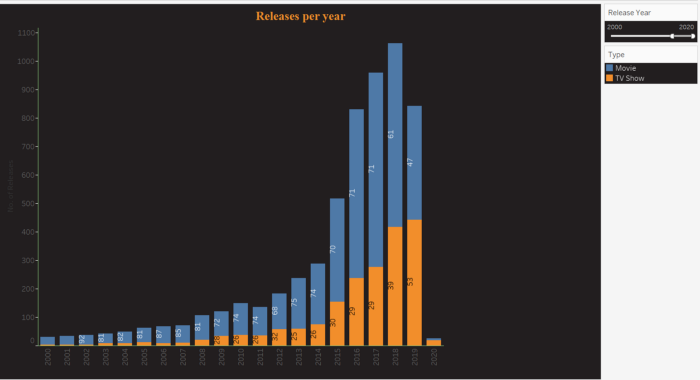
For this particular assignment, I chose data visualisation tool called Tableau. This is a tool that works well for both desktop and online data analysis Eaton et al (2018). The tool is rich in capabilities and functionalities that can be used to get a visual inferences of the data supplied. For this exercise, I chose the dataset found at the Tableau public repository and selected the Entertainment Netflix data found here: <https://public.tableau.com/s/sites/default/files/media/netflix_titles.xlsx> for analysis. My next step was to create a trial account within the Tableau Single sign on SSO centre as an admin. Thereafter I proceeded activate the link in order to get access to the online platform. The next step is to use the Alteryx tool that will help us transform the data and two things happen here;

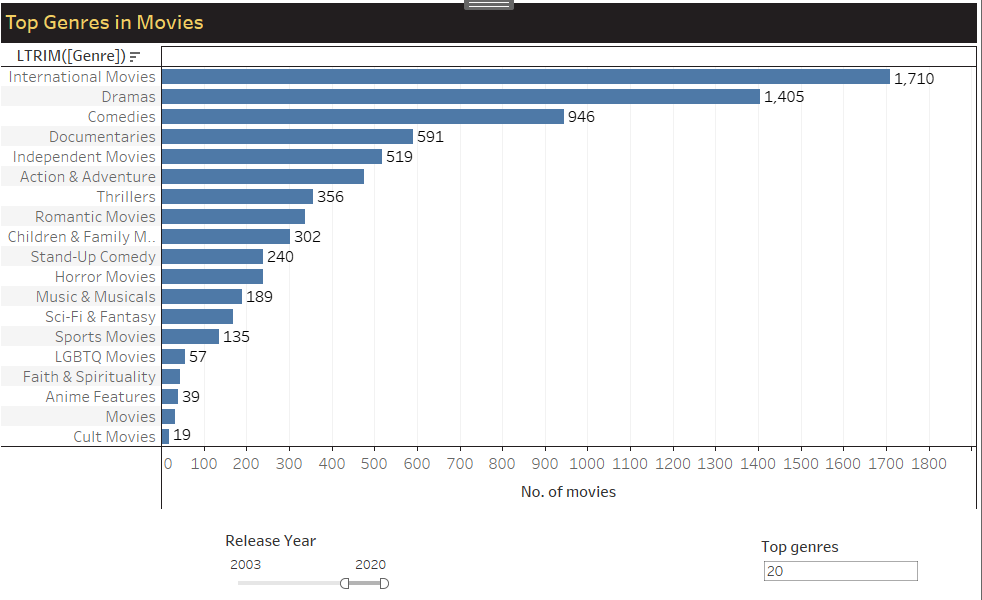
* The text in country and genre columns is converted into multiple columns
* The columns are then transposed to get are a record for each genre and country

The result looks something like below;



Next we can do an exploratory data analysis on the data set and for this, we chose the bar char option to view the movies versus the TV shows counts on Netflix. From the illustrations below, it can be observed that Netflix has lately started to shift more attention to producing more TV shows.





From the second analyses, we can also observe that top genres that lead at Netflix include international movies and drama.

**Experience and criticism;**

Overall, tableau is a powerful web data visualisation tool to use for primary and high level industrial analyses. I loved the overall tool pack within tableau that enables you to choose any type of visualisation that you want to do with your data just at the click of an option. Secondly, I loved the fact that tableau public has got an online data source for one to select whichever data set they want to use, especially if they are new to data analyses and are undecided on which dataset to use.

Critically, the platform may quite be confusing for newbies, especially with the inability to recognise that table’s online resources are separated from each other, for instance, the tableau public account URL at <https://public.tableau.com/> is different from the analysis resource site at <https://online.tableau.com/> that is used for data analyses, and this may be confusing to newbies. Secondly, using the online platform requires an admin user to perform activities on behalf of others, many analysts may not be aware of this and hence struggle to get to use the platform.

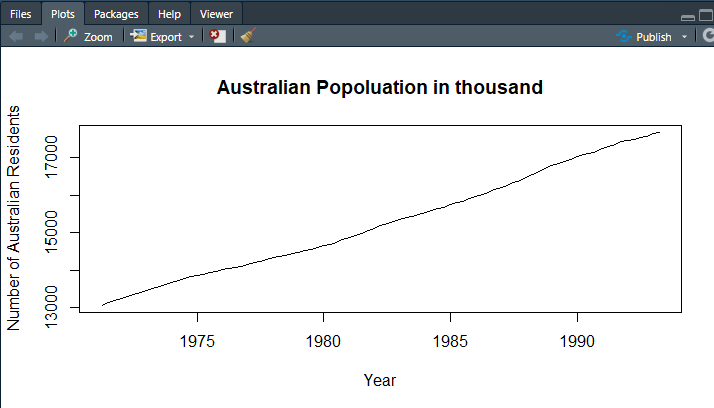
**Question 3**

> summary(austres)

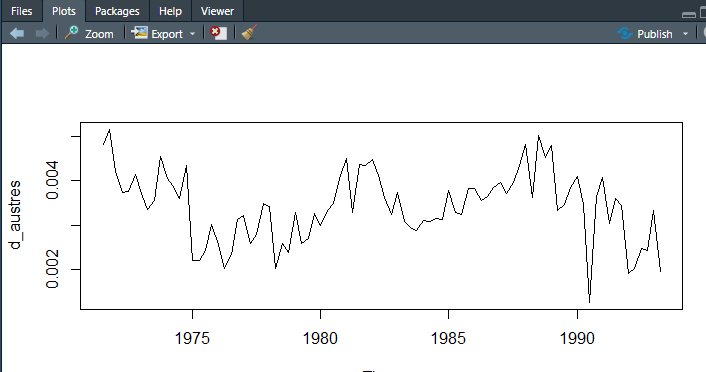
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

  13067   14110   15184   15273   16399   17662

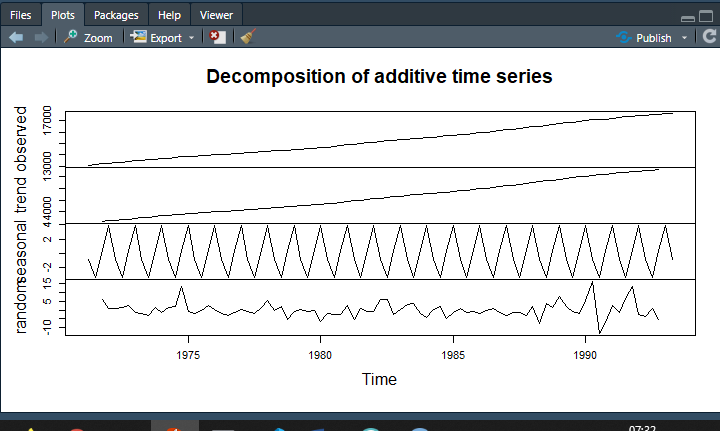
> plot(austres, xlab="Year",ylab="Number of Australian Residents",main="Australian Popoluation in thousand")

****

2. d\_austress <-diff(log(austress))

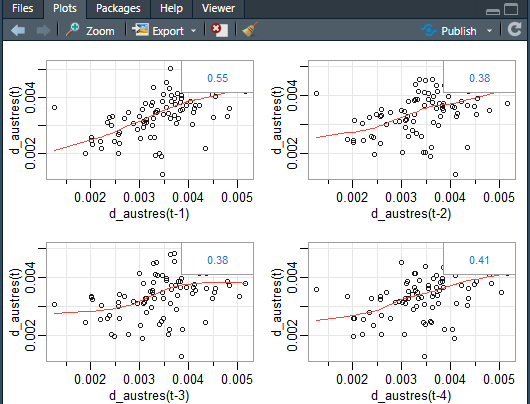


aust <- decompose(austres)



C.

> lag1.plot(d\_austres, 4)



The relationship is interpret as linear, there is strong correlation between the two variables

D.

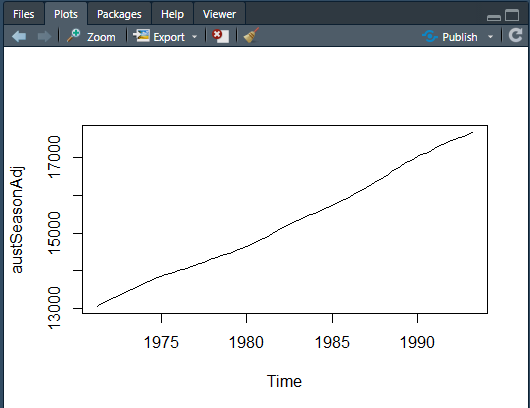
> plot(aust)

> austSeasonAdj <- austres -aust$seasonal

> plot.ts(austSeasonAd)

> plot.ts(austSeasonAdj)

> lag1.plot(d\_austres, 4)



**REFERENCES**

Saffary, T., et al. "Analysis of COVID-19 cases' spatial dependence in US Counties reveals health inequalities." Frontiers in public health 8 (2020).

Eaton, Tim V., and Michael Baader. "Data Visualization Software: An Introduction to Tableau for CPAs." The CPA Journal 88.6 (2018): 50-53.