**STAT455 23S2: Time Series Methods**

**Assignment 1**

**Due Date: 6th Aug. 2023**

**STAT455 23S2 - Assignment 1**

**Yuanyuan Qin (43865058) Huizhe Zhao （ ? ）**

**Question 1:**

We are forecasting, or predicting all the time. In the following questions, the number in brackets applies to (single/pair) submission.

1. We want you to describe (two/four) forecasts [predictions] that you make routinely, and probably informally, in your daily life (e.g. when do I need to do my laundry next, can I work part-time and pass this year). One should be predicting in the short-term (7 days or less ahead), and one longer-term (more than a month in advance).

2. What decisions are aided by your (two/four) forecasts?

3. How might you measure the “success” or “goodness” of your (two/four) forecasts?

4. Can you hand-sketch a time series plot of past data (with why or why not)?

**Answer:**

1. The routine short-term forecast in my daily life is the weekly grocery spending. This involves estimating how much I would likely spend on groceries for the upcoming week based on my typical consumption, planned meals, and any special occasions or events.

Another prediction I make regularly is the real estate price around my living community. This prediction means the estimation of how property prices may change over the next year based on factors such as market trends, economic conditions, local developments, and selling prices of properties in the community area in the past 12~24 months.

1. The first short-term forecast gives me the proper balance numbers I should maintain in my bank account.

The real estate market prediction gives me a reasonable price when I have in mind when I plan to buy new or sell the properties.

1. The success measure of my weekly grocery spending is that my cost of living is on track.

If the forecasted amount is close to the actual spending, the forecast can be considered accurate.

The success of this forecast can be evaluated by comparing the predicted property prices with the actual prices at the end of the forecast period. If the forecasted trends align with the actual price movements, the forecast can be considered successful.

1. Yes. I think I can hand-sketch time-serial data because of its inherent time ordering. It is a type of data that is collected and recorded over time. Each data point is associated with a specific time or date stamp, which allows the data to be ordered chronologically. This time order is what makes time series data unique and suitable for plotting.

**Question 2:**

1. From the Statistics New Zealand website (stats.govt.nz) download one of the following series – the choice is yours – and import it into R. Assign a date variable to the points in the time series.

• Quarterly Total merchandise trade exports

• Quarterly Total number of visitor arrivals

• Quarterly Total number of residential building consents

• Quarterly Gross Domestic Product

Warning: Some have monthly and annual equivalents so choose the right one.

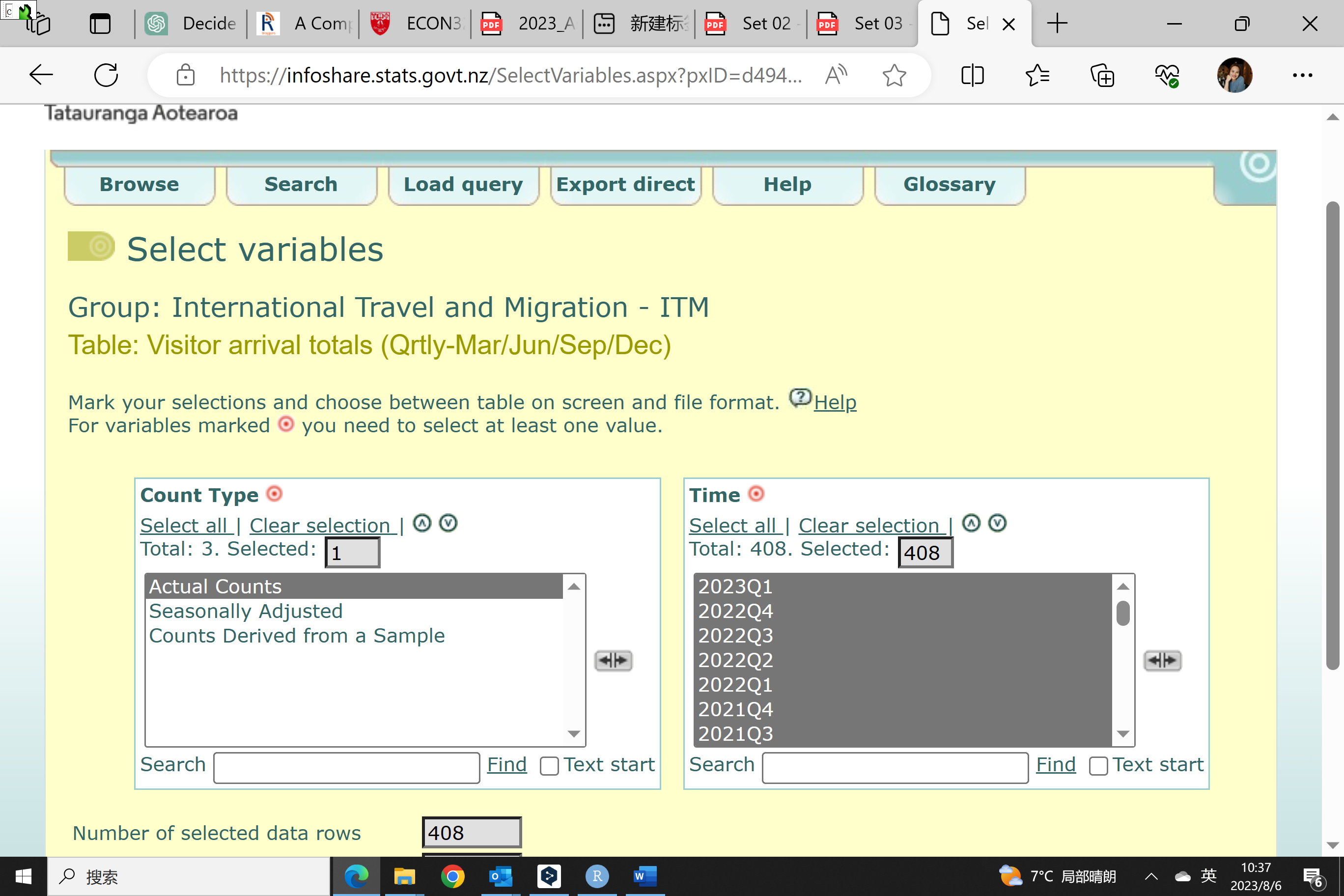
2. Write a paragraph explaining what the series measures and why people may want to analyze it.

3. Graph the time series and describe the main features of the time series you have plotted. 4. In no more than a page indicate some possible explanations for the features you have described above. You will be assessed on the plausibility of the explanations, not their correctness.

5. If were going to do a time series analysis would you use the whole series? Give your reasons for your decision.

**Answer:**

1. From the Info Share webpage from Stats NZ ([Browse - Infoshare - Statistics New Zealand (stats.govt.nz)](https://infoshare.stats.govt.nz/?_ga=2.165361917.31530558.1691274937-1461365771.1684461768)), I downloaded the quarterly total of visitors’ arrivals in NZ. The original data is submitted in the separate Visitor\_arrival\_Total\_NZ\_2023Q1.csv file.



1. This serial measures the total inbound tourism traffic to New Zealand and captures the volume of people visiting the country for various purposes, such as leisure, business, education, or visiting friends and family.

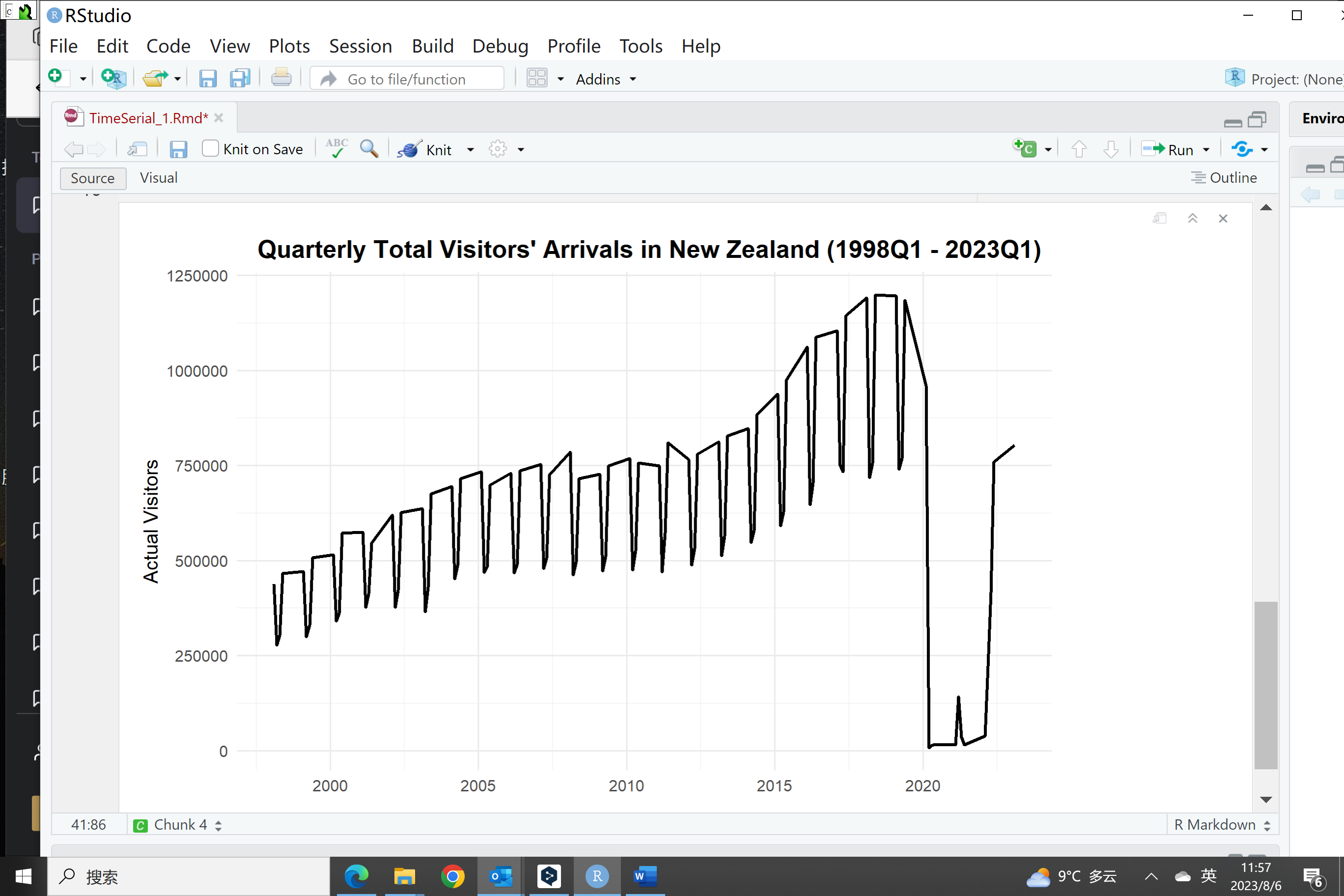
Tracking the quarterly total visits in New Zealand statistics is crucial for several reasons, considering the importance of the tourism industry to the country:

• Economic Impact: Tourism is a significant contributor to New Zealand's economy, generating revenue, creating jobs, and supporting local businesses. This data allows the government to assess the revenue generated from tourism, identify trends in tourist spending patterns, and understand the contribution of the tourism sector to the overall GDP.

• Identify seasonal Patterns and adjust the marketing price and strategy: By analyzing these patterns, businesses in the tourism industry can adjust their marketing strategies, pricing, and capacity planning to effectively cater to peak and off-peak periods.

• Policy and Infrastructure Planning: The quarterly data on visitor arrivals provides valuable insights for policymakers and government agencies to allocate resources for tourism infrastructure development, transportation, and other initiatives to enhance the overall attractiveness of New Zealand as a tourist destination.

1. Graph the data (1998Q1 – 2023Q1) in R



Here are the description of the key main measures:

• Average and Trend: There is a clear increasing trend in the number of visitors to New Zealand from 1998 to 2019. The trend indicates positive growth in tourism over the years.

• Seasonality: The data exhibits strong seasonality, with regular fluctuations occurring within each year.

• Outliers: Th 2020 to 2022 (3 Pandemic years) shows in a sharp drop in visitor numbers, creating an outlier in the time series.

1. The quarterly visitor data in New Zealand from 1998 to 2020 exhibits seasonality with fluctuations and increase trends throughout the years. Regular peaks and falls are likely to correspond to seasonal travel patterns, such as holidays, vacation periods, and events. Besides the seasonal fluctuation, there is also a noticeable increasing trend from 1998 to 2019, prior to the COVID-19 pandemic. The trend indicates a positive growth in tourism over the years, likely driven by factors such as improved infrastructure, marketing efforts, and global interest in New Zealand as a tourist destination.

The COVID-19 pandemic and lockdown since 2020 make the number of visits amount drop dramatically and the visitor just recover in recent quarters, such as 2022 Q4 and 2023 Q1.

There is a very small peak in 2021Q1 and 2022 Q1 because of the limited open-gate policy for the Christmas holidays, however, the number of visitors amount is still quite low because of medical concerns at that time.

1. The decision to use the entire time series or a subset of it will depend on the research questions, the nature of the data, and the specific analysis objectives.

I would use a subset rather than whole data set under the following considerations.

• Data Quality: Stats NZ makes the statements that the total number of visitors in the early has errors and missing data, which might not be so reliable. I would rather select the subsets without missing or incomplete data points.

• Relativity: Recent data may be more relevant for understanding the current state of the phenomenon, especially if there have been significant changes or tourism developments in recent years.

• Identified Outliers: Pandemic years from 2020 to 2022 include a lot of unusual events or global policy disruptions that lead to a sharp drop in visitor numbers. Those data sets might not be very useful for future prediction unless some policy and structural changes would have long-term impacts on the New Zealand tourism markets.

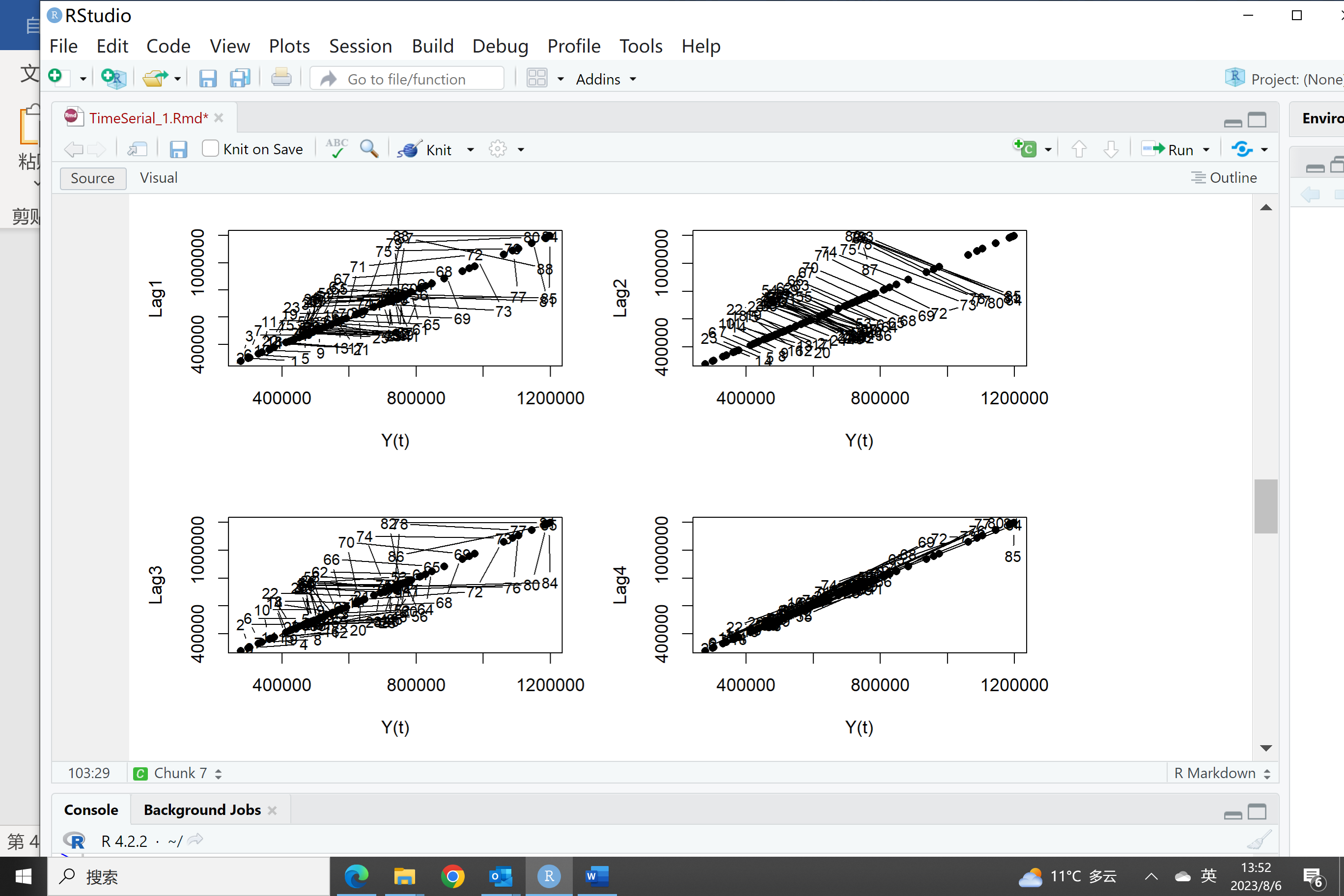
**Question 3:**

1. Make a scatterplot of the values at time t against their lag 1 values (at time t − 1). Repeat for the values at time t against their lag 2, lag 3, and lag 4 values. What does this show in terms of the correlations between a value and its lagged (past) values?
2. Create the acf for the series. Describe and explain what it shows.
3. Create a time series of the first differences of the series and plot it. Describe and explain it in relation to the original undifferenced series.
4. Would it be better to use the original time series or the time series of the differenced data for time series analysis? Provide your reasoning.

**Answer:**

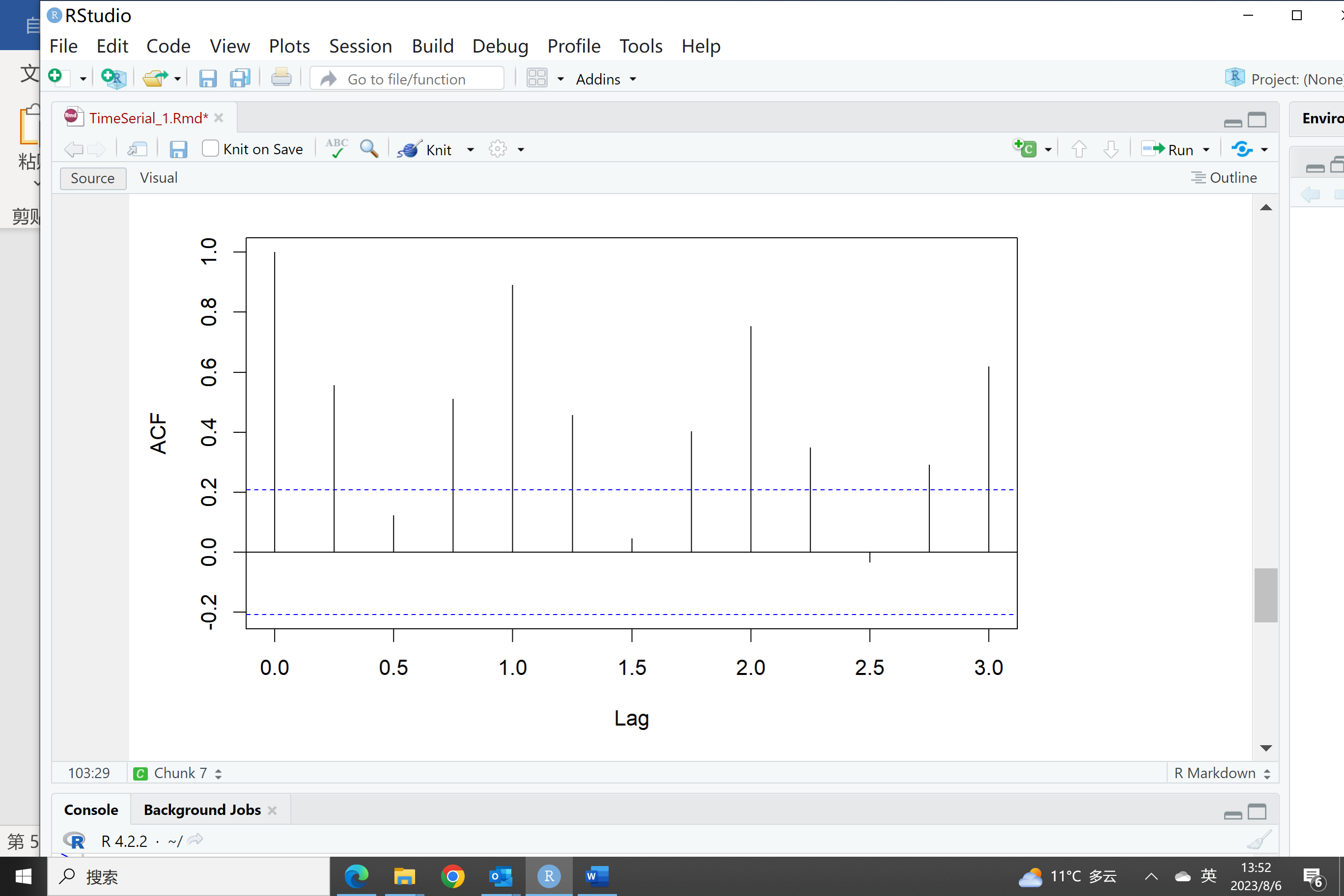
1. In R, the first step is to create the lag data and then plot the lag data vs the original ts\_data.

The 2\*2 plots are put below.



These scatter plots help to understand the correlation between a value and its lagged (past) values in the time series. A positive correlation showed between y(t) with lag1 and lag2, lag3 and Lag4 values. Compared to the other lagged scatterplots (Y(t) against Lag1, Lag2, and Lag3), the lag4 plot suggests that there is a noticeably stronger relationship between the current value Y(t) and the value four periods ago (Y(t-4)). This suggests that the time series exhibits a seasonal or periodic behavior with a periodicity of four quarters.

1. ACF of the original time serial data:

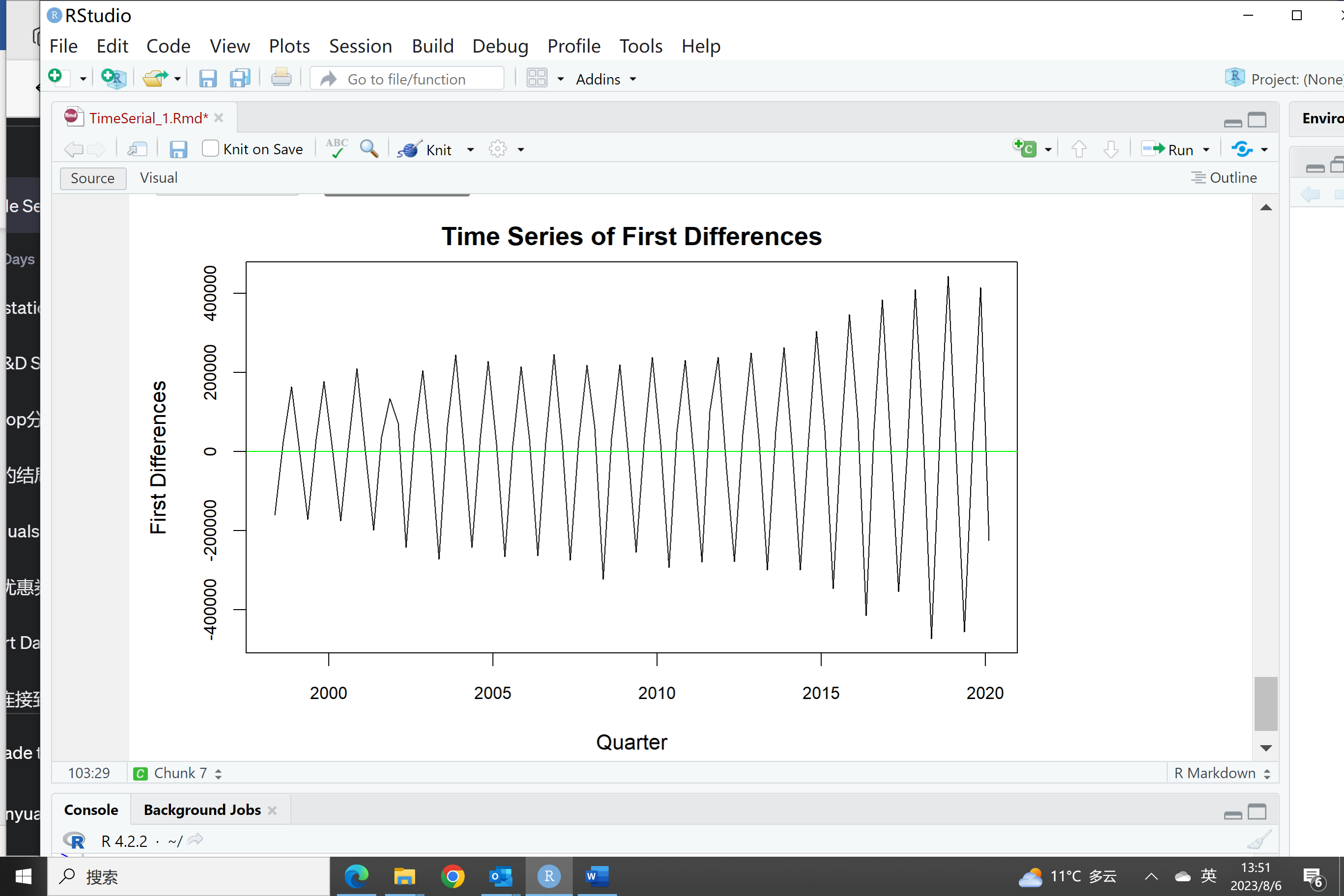


The acf() function calculates the autocorrelation at different lags and produces the ACF plot. The lag.max argument is 12 and it means a 3-year ACF-lag plot.

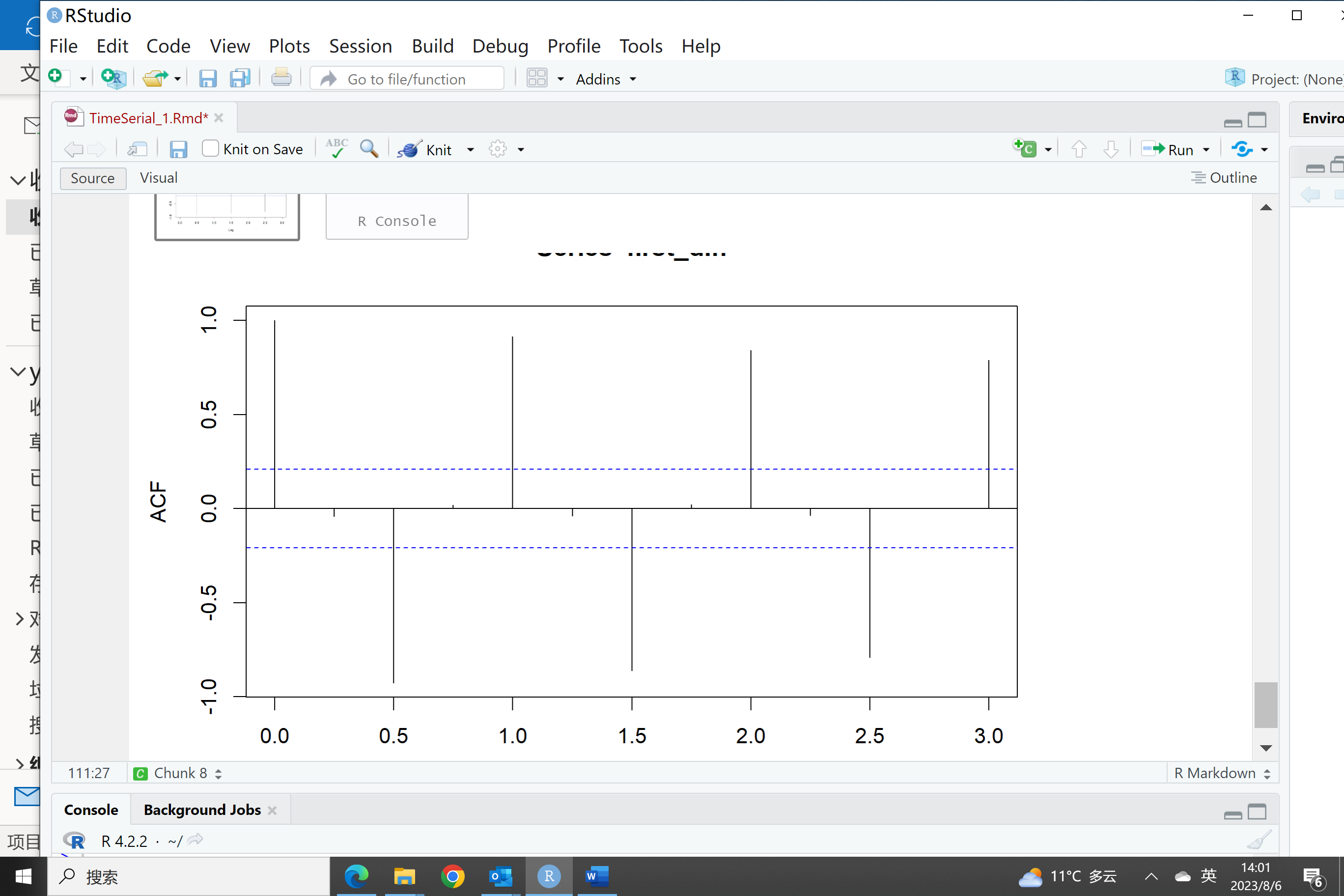
The ACF plot shows a bar chart with lags (time intervals) on the x-axis and the autocorrelation values on the y-axis. The lag 0 (correlation of the time series with itself) is always 1. The other bars represent the autocorrelation values at different lags. The significance of the bars is assessed based on their heights and whether they fall within the blue confidence interval bands. If a bar extends beyond the blue bands, it suggests that the autocorrelation at that lag is statistically significant.

ACF plot shows the original time serial set is not a stationary process and needs to be transformed.

1. Here is the first difference time serial plot



And ACF of the first difference time serial should be checked.



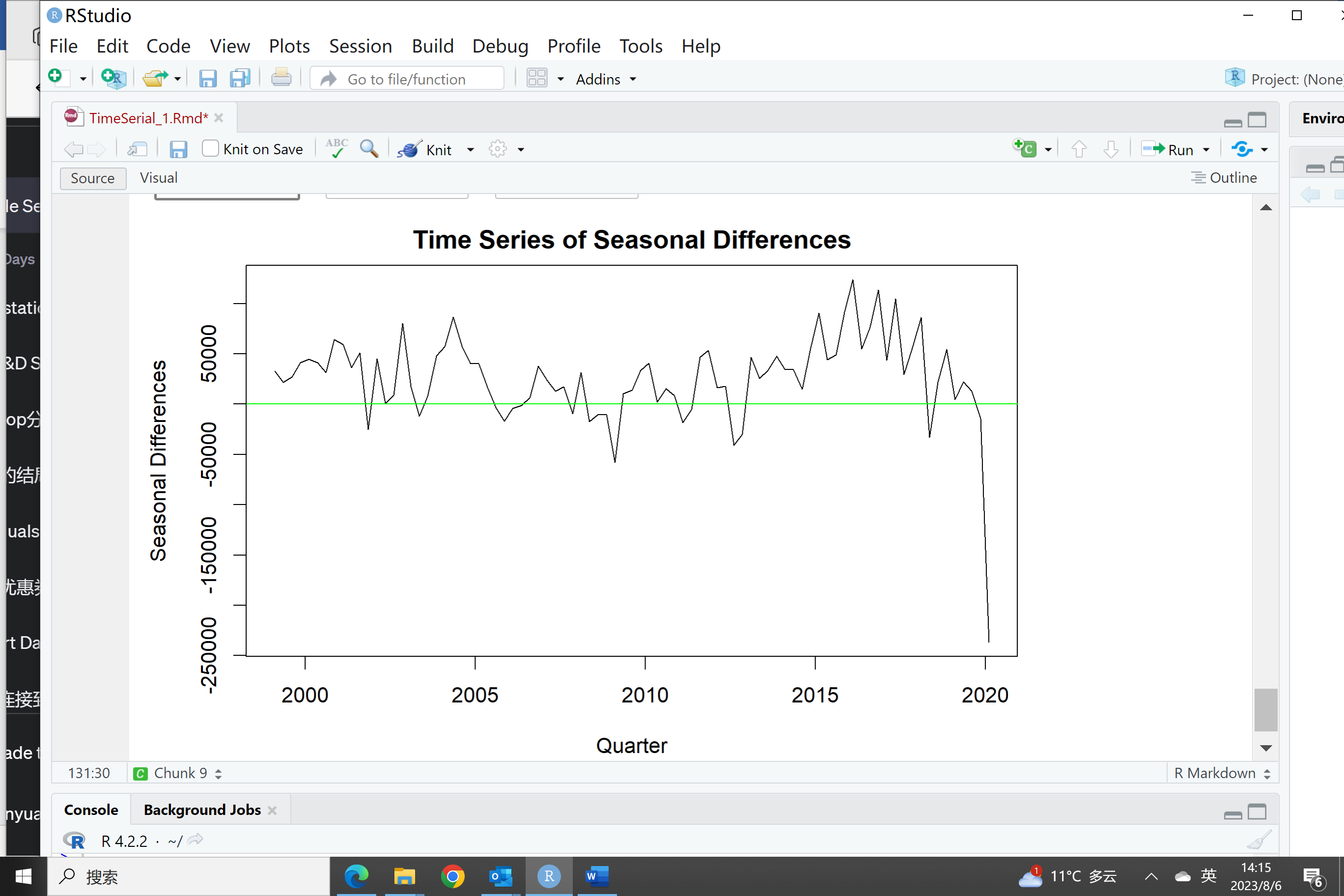
The first difference time series is obtained by taking the difference between consecutive observations of the original undifferenced series. It represents the rate of change or the increment between consecutive data points. The first differences remove the trend component from the original series and focus on the short-term fluctuations or changes between observations.

However, it’s also obvious that the first differences do not completely remove the seasonal component in the time series data. The reason for this is that the seasonal component is not necessarily present in the differences between consecutive observations. It may still persist across multiple seasonal periods, causing the differences to exhibit some remaining seasonal patterns. In this case, the first different serial is still not stationary because of the high correlation between Y(t) and Y(t-2) and Y(t-4).

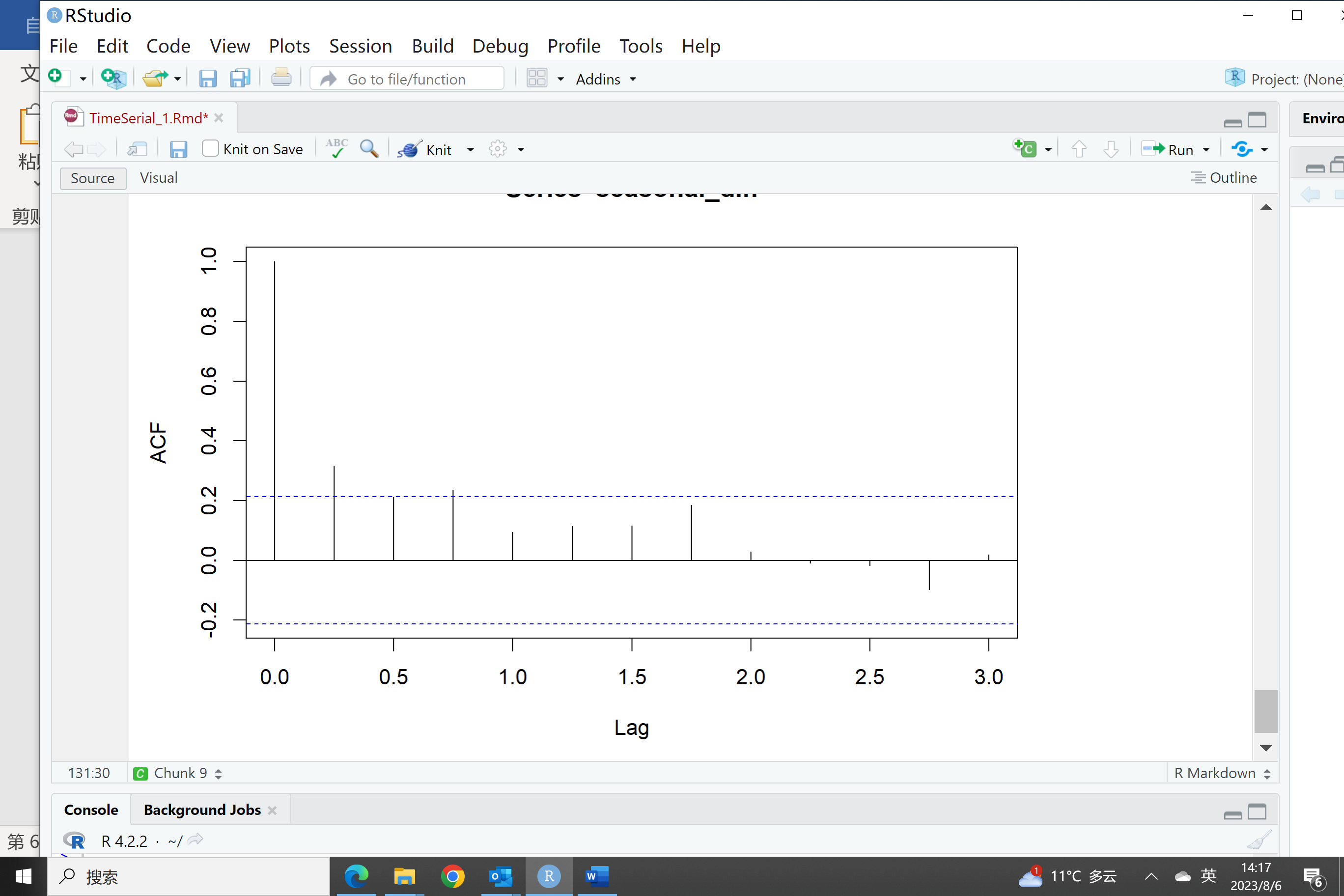
1. In this case, I think it would not be a good solution to choose the original time series or

the time series of the differenced data because both of them are not stationary time series. So, I would like to perform the seasonal difference on the original time series.

Here are the data plot and ACF plot after seasonal difference transformation.



ACF of the seasonal difference time serial set



When checking the autocorrelations from the ACF plot of seasonal difference, the autocorrelations tend to decrease rapidly, and most of the autocorrelations will be close to zero for lags. It shows, after seasonal difference transformation, the time serial could comply with weak stationary criteria.

In conclusion, when dealing with quarterly total visitors’ time serial, seasonal difference is preferred because the time serial could approximately comply with the Weak Stationary principle after such transformation.