Data Science – Live Session Unit 11 R Markdown File

Nagesh Padiyar

Contents

Introduction 1

Data Analysis for ukcars 2

Plot of Time Series 3

Decomposition of Time Series 4

Plot of Seasonally Adjusted Data 6

Introduction

The following document describes the analysis of ukcars data for seasonal variations and trends using Time Series analysis. It also shows how classical decomposition can be used for calculating trend-cycle and seasonal indices.

The data analysis is done using the fpp data

Code to load data:

library(fpp) #fpp package must be installed first

data(ukcars)

data(visitors)

data(hsales)

I have used ukcars data to perform the analysis.

> data(ukcars)

> head(ukcars)

Qtr1 Qtr2 Qtr3 Qtr4

1977 330.371 371.051 270.670 343.880

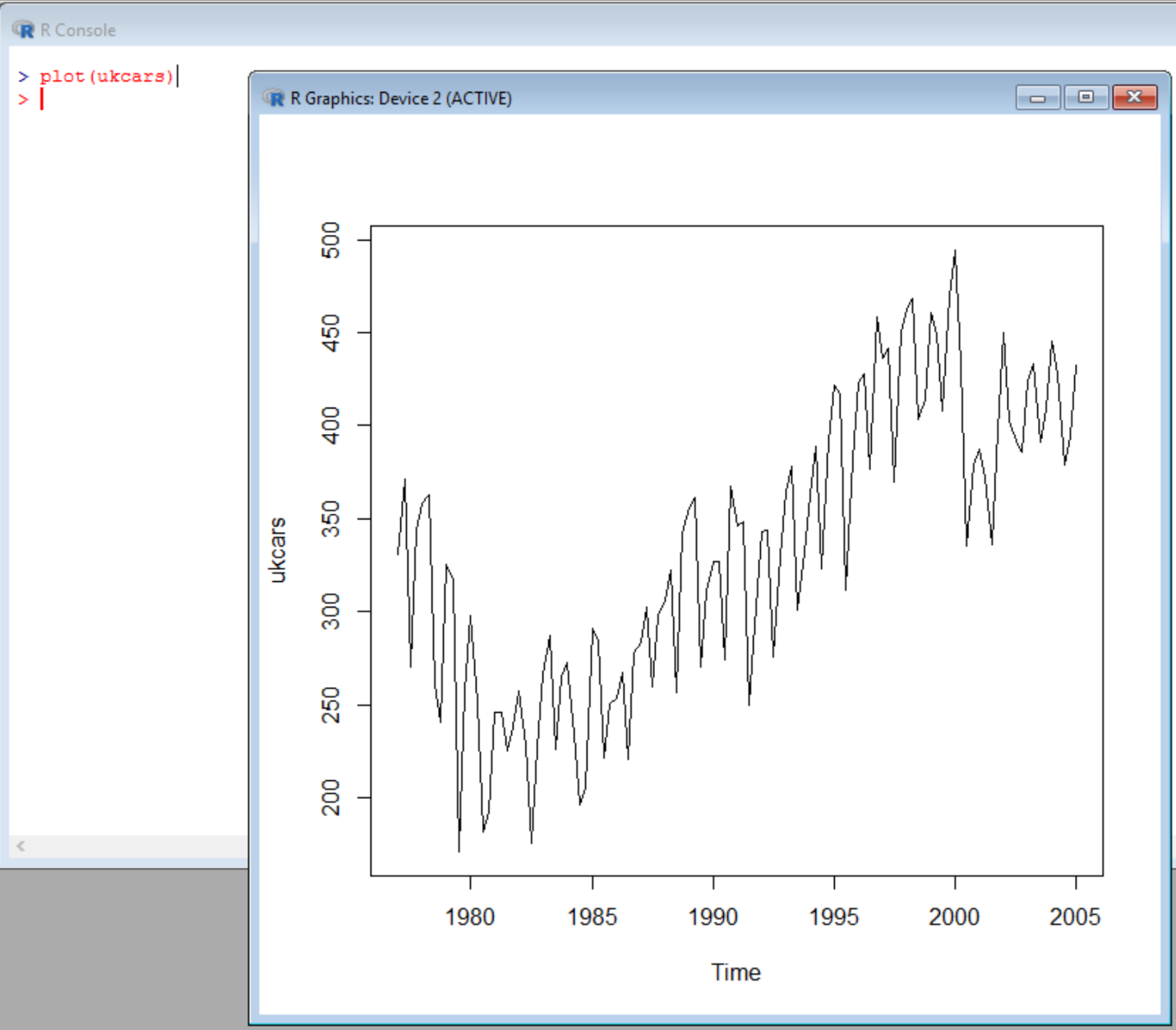
1978 358.491 362.822

> summary(ukcars)

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

171.2 267.0 335.1 333.5 391.8 494.3

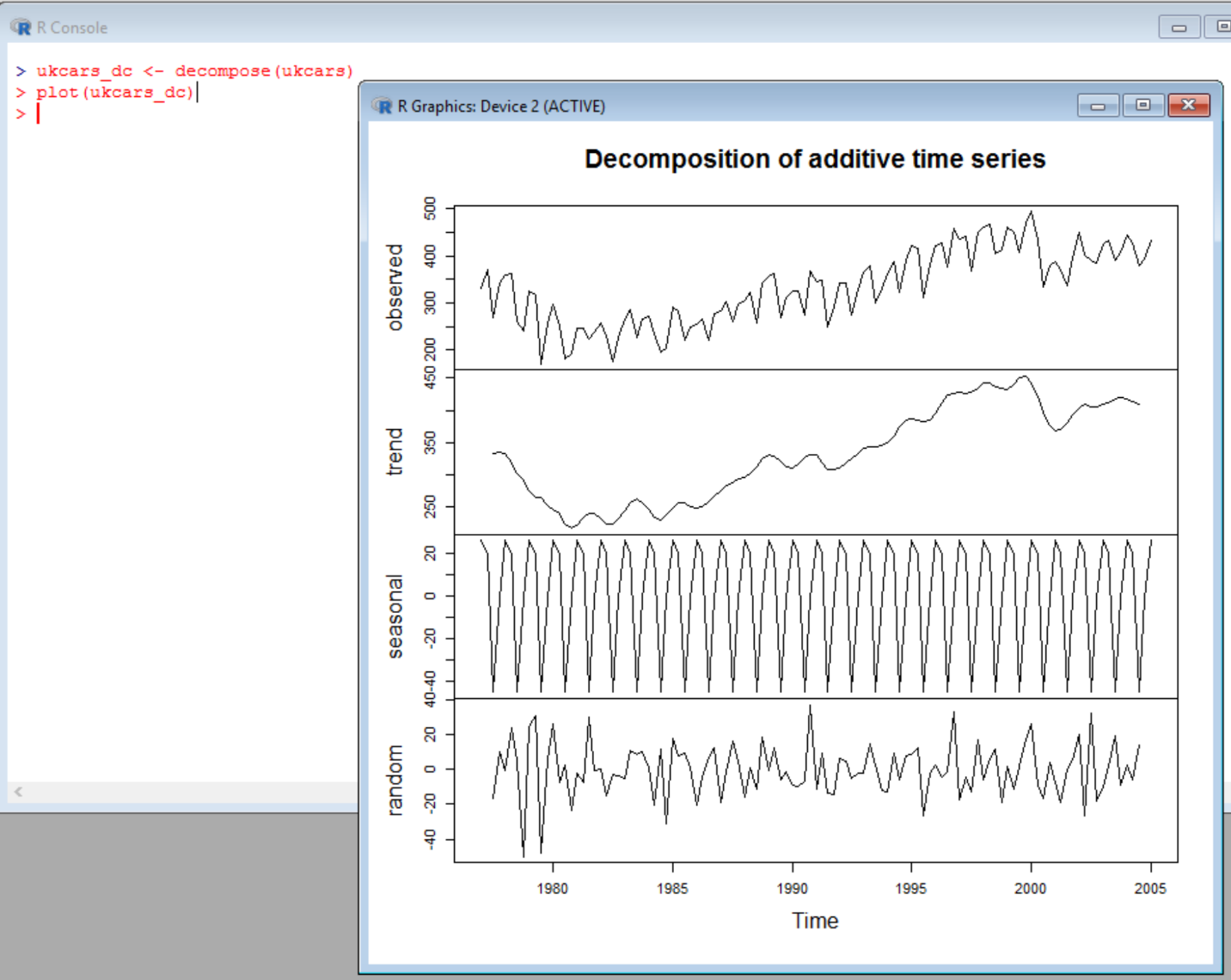
1. Plot the time series. Can you identify seasonal fluctuations and/or a trend?



We notice that there is a trend over the years. However, there are seasonal fluctuations for each quarter as well.

The sales were very low in the 1980 through 1985. Then picked up steadily up until 2000. And then there was a drop in 2000. It was more steady after 2000 onwards.

1. Use a classical decomposition to calculate the trend-cycle and seasonal indices.

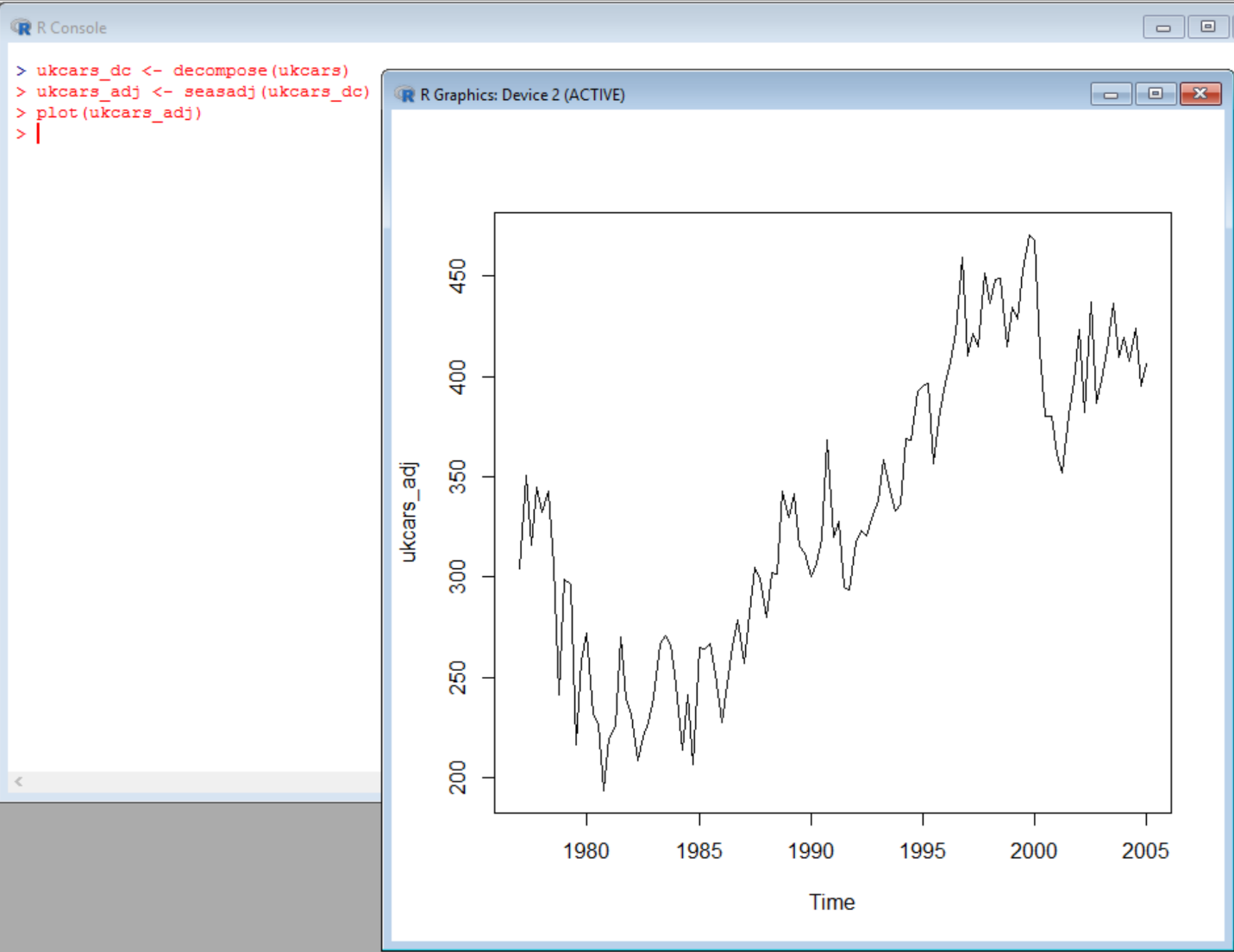


1. Do the results support the graphical interpretation from part (a)?

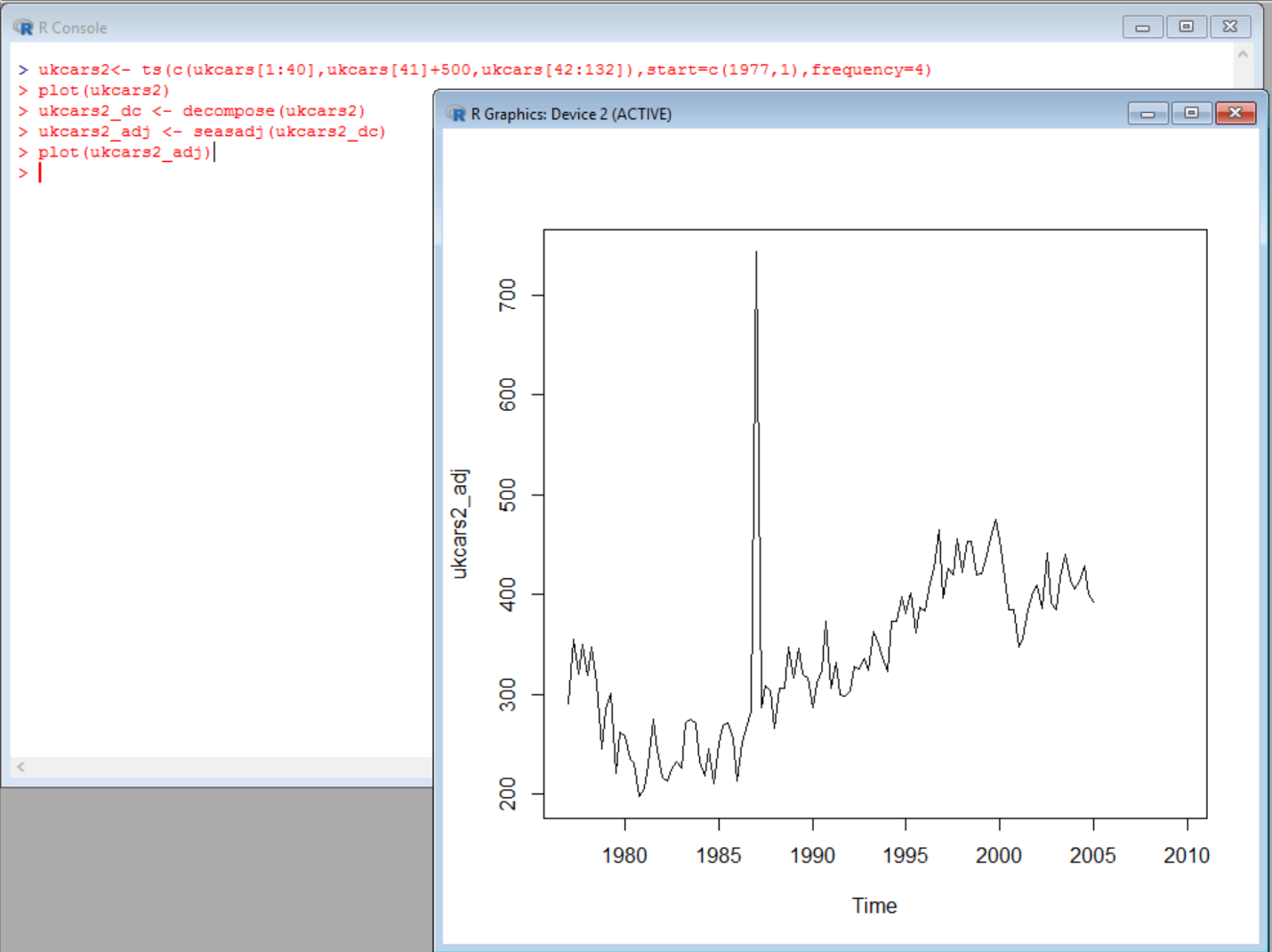
Yes for the Trend. The Trend shows that Car Sales dipped in year 1980 through 1985. However, picked up steadily thereafter until year 2000. In 2000 there was a drop in sales. It however picked up and stayed steady after 2000 to 2005.

However, the seasonal fluctuations seem to be steady over the years.

1. Compute and plot the seasonally adjusted data.

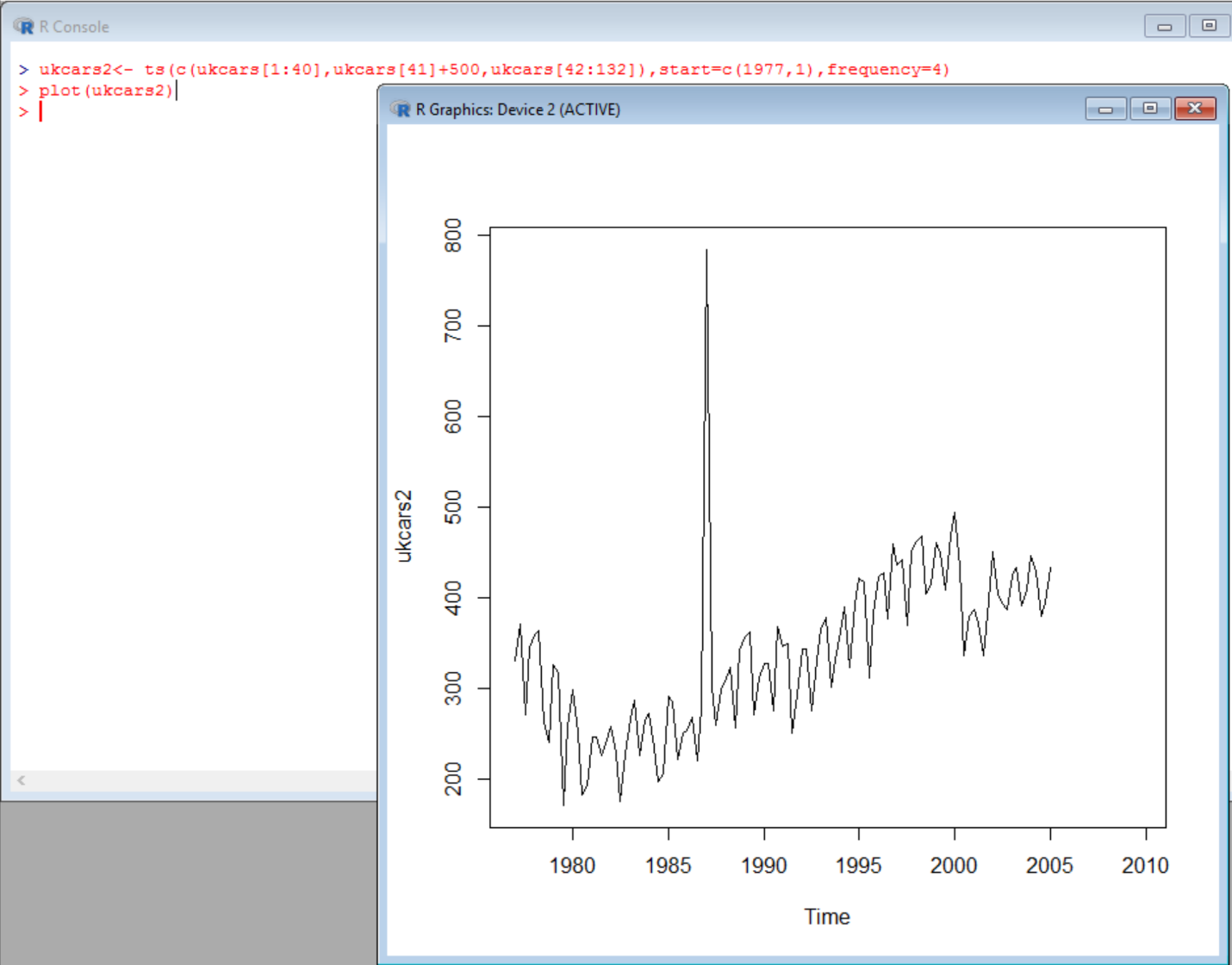


1. Change one observation to be an outlier (e.g., add 500 to one observation), and recompute the seasonally adjusted data. What is the effect of the outlier?



The outlier stands out in the graph regardless of the seasonally adjusted data.

ukcars2<- ts(c(ukcars[1:40],ukcars[41]+500,ukcars[42:132]),start=c(1977,1),frequency=4)



The seasonal adjusted data did not hide the outlier. The outlier stood out from the trend regardless.

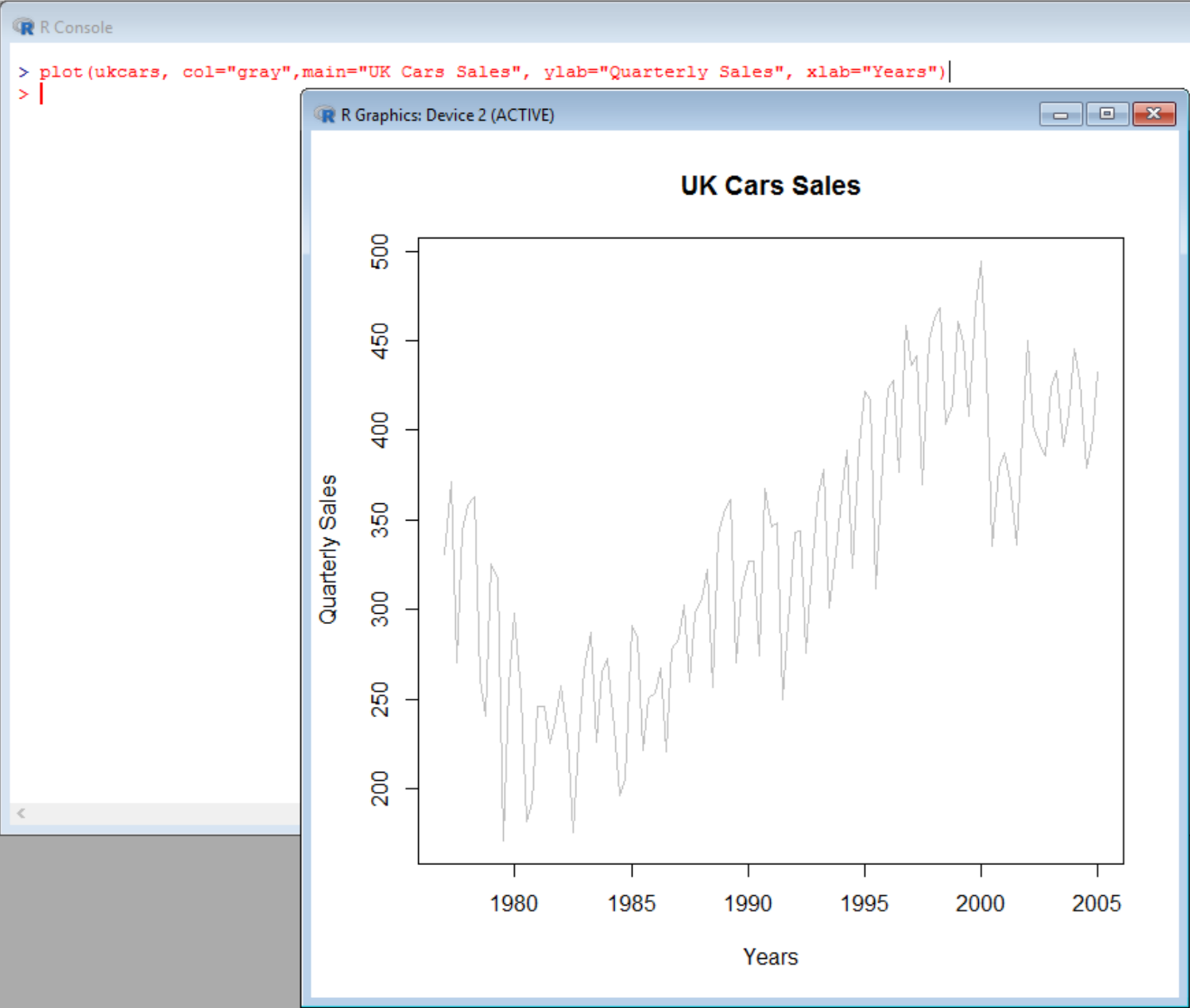
1. Does it make any difference if the outlier is near the end rather than in the middle of the time series?

No. I don’t think adding the outlier either in the beginning, end or in the middle would have made any difference in the trend.

1. Use STL to decompose the series.

plot(ukcars, col="gray",main="UK Cars", ylab="Sales Data", xlab="")

fit <- stl(ukcars, s.window=5)



lines(ukcars\_dc$time.series[,2],col="red",ylab="Trend")

1. Use STL to decompose the series.

lines(fit$time.series[,2],col="red",ylab="Trend")

