# Advanced Models

## Hw2 Part 2

a)

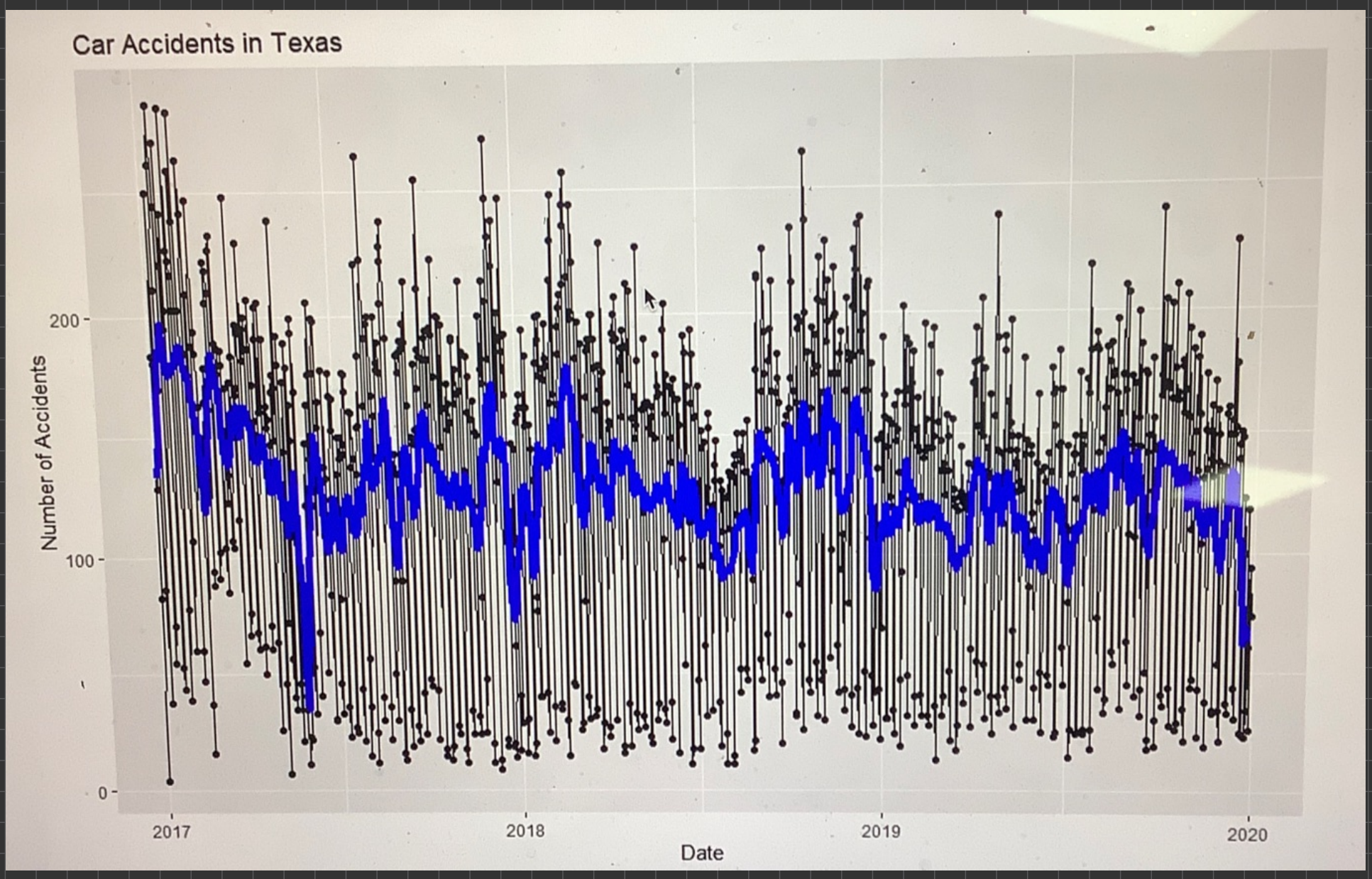
a. Consider the following plots which show the application of a Moving Average Filter (in blue) to the Texas car crashes data. Match the filters described by $a_1$ and $a_2$ to their corresponding plots. Explain why the two filtered time series differ in terms of their smoothness.

- $a_1 = (\frac{1}{14}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{14})$
- $a_2 = (\frac{1}{60}, \frac{1}{30}, \frac{1}{30}, \ldots, \frac{1}{30}, \frac{1}{60})$, where $a_2$ is length 31

Answer:

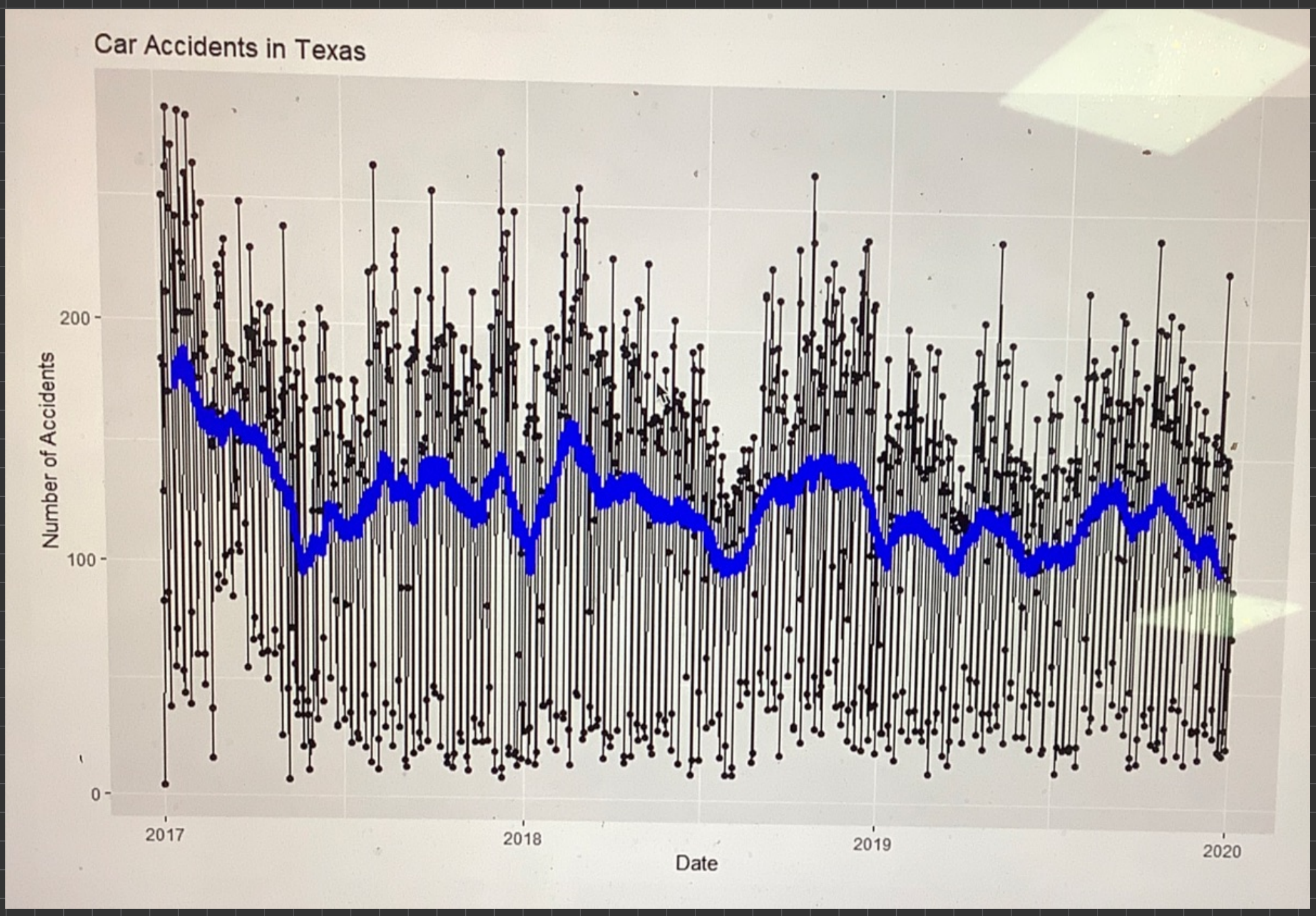
$a_1 = (\frac{1}{14}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{14})$

Corresponds to the first plot (this plot) ✓



$$a_2 = \left(\frac{1}{60}, \frac{1}{30}, \frac{1}{30}, \dots, \frac{1}{30}, \frac{1}{60}\right)$$

Corresponds to the second plot (this plot) ☺

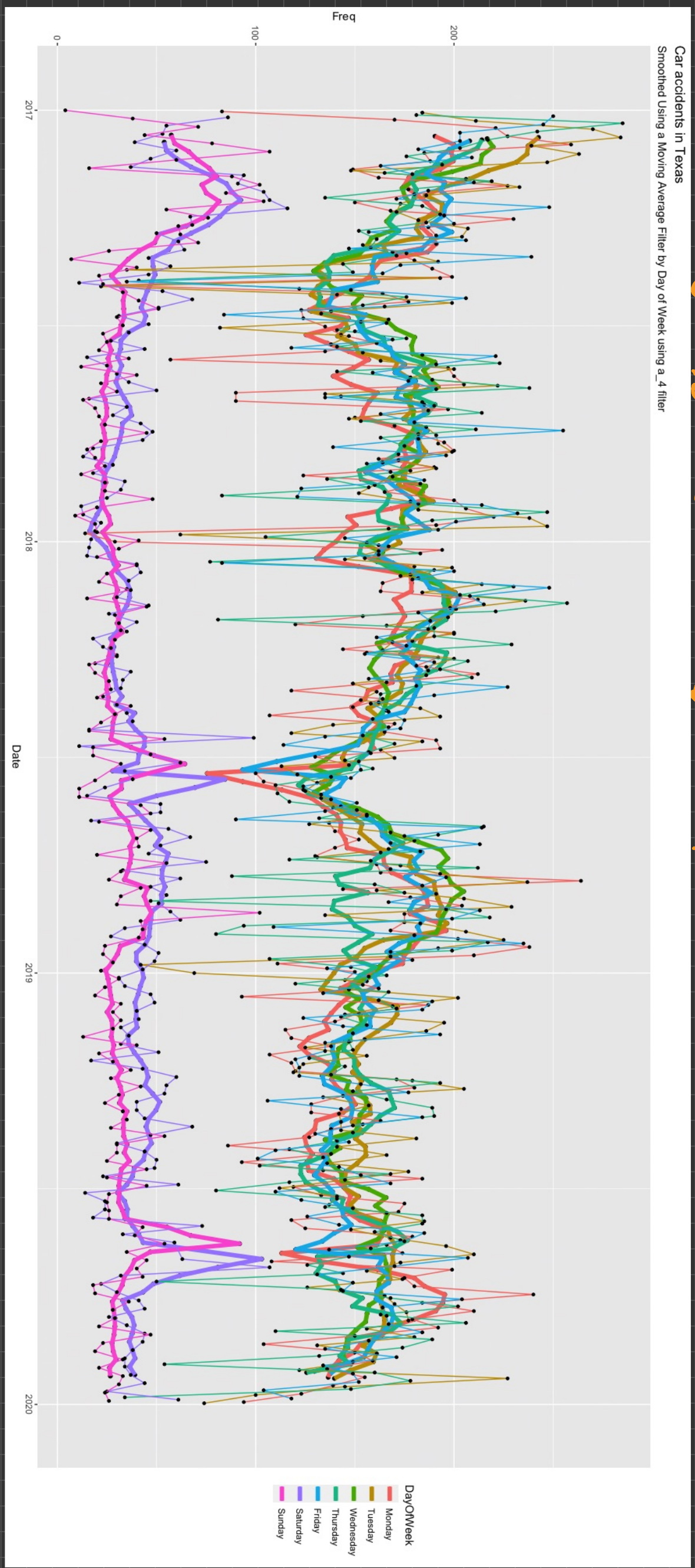



$a_2$ filter will have a smoother effect than $a_1$ because each neighbor weight of $a_2$ is smaller than any of the $a_1$ filter neighbors weights. If a neighbor value $X'$ differs a lot to the current value $X_t$ at time $t$ or any of the other neighbors considered in the filter window in the original time series; that $X'$ value will be normilized by $\frac{1}{30}$ or $\frac{1}{60}$ in the $a_2$ filter; making that this does not affect abruptly the trend of the filtered time series.

On the other hand the $a_1$ weights for any neighbor are either $\frac{1}{7}$ or $\frac{1}{14}$; so if a value abroptly grows or decreases, it will be multiplied by a number greater than any of $a_2$ weights, hence making that the trend of the filtered time series $a_1$ be affected more drastically by abrupt changes in neighbor values.

b)

b. Consider the plot shown in Figure 3 which shows the Moving Average Filter specified by $a_3 = (\frac{1}{3}, 0, 0, 0, 0, 0, 0, \frac{1}{3}, 0, 0, 0, 0, 0, 0, \frac{1}{3})$. Create an analogous plot for the filter $a_4 = (\frac{1}{7}, 0, 0, 0, 0, 0, 0, \frac{1}{7}, 0, 0, \ldots, 0, \frac{1}{7})$ where $a_4$ is of length 43. Explain why the two filtered time series (applying filters $a_3$ and $a_4$) differ in terms of their smoothness.




Result:

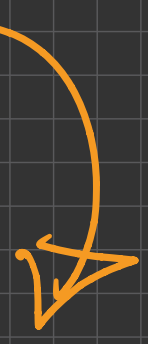
$a_4$ filter will have a smoother effect than $a_3$. There are many more "0" value weights for neighbors in $a_4$ than $a_3$. Hence, many more neighbors that could abruptly change their value, they are not considered for the filtered $a_4$ times series trend. Also, any considered neighbor for the $a_4$ filter will be multiplied by a smaller value than any considered neighbor for the $a_3$ filter. So, $a_4$ filter normalizes more any noise with abrupt behaviour, making the general trend to be smoother.

c)

c. Suppose a legislator wants to know about the effectiveness of a traffic law that went into effect in Texas in September of 2017 that prohibited texting while driving. The legislator appreciates that the daily data on car crashes will be noisy and is willing to consider easy-to-understand summaries of the trend in the series, but is not interest in a fancy statistical analysis using causal inference methods. Provide a reason why assessing the impact of the law on traffic crashes by visual inspection of the filtered time series may not be a good idea.

With the plots obtained with the filters, the legislator could find visually certain trend in the filtered time series after September in 2017. This trend could lead to the conclusion that there is certain correlation in the data after September 2017 when the the traffic law went into effect. So, the legislator could be concluding that the law had effect. However, correlation does not imply causation. The fact that there could be certain correlation between the data after September 2017, does not imply that it was due to the traffic law that went into effect. There are many variables that could have NOT been taken into account. For example, it could have been the case that oil price increased after September 2017. Hence, many people in Texas decided to NOT use their cars, affecting the number of crashes.

APPENDIX (CODE)



# CODE PART A

```r
##################################################
#Importing the libraries to be used
library(ggplot2)
library(gridExtra)
library(tidyverse)
library(astsa)
library(xts)
library(lubridate)

#Importing the data from the csv file
data1 <- read.csv("/Users/rafa/Documents/Master Austin/MAESTRÍA_AUSTIN/Advanced Predictive Models/HW2/file2.xls", header=TRUE, stringsAsFactors=FALSE)
data1_2 <- data.frame(Freq = data1$Freq,Date =as.Date(data1$Date))

# Apply kernel smoothing to the imported data
a <- c(1/14, rep(1/7, 6), 1/14)

ma_smoothed_data <- data.frame(
  Freq = stats::filter(data1$Freq,
                       sides = 2,
                       filter = a),Date =as.Date(data1$Date))
#Plotting the data and the smoothed time series
gg_data <-    ggplot(ma_smoothed_data,aes(x = Date, y = Freq,group=1))+
  geom_line(size = 3.5, color = "blue") +
  geom_point(shape = 21, size = 1, fill = "white", color = "black") +geom_line(data=data1_2,size=0.5,color="black")


gg_data

```

# CODE PART B

```r
#-----------------------------------------------------------------Part b-----------------------------------------------------------------------------------

#imorting the data from the csv file
data1 <- read.csv("/Users/rafa/Documents/Master Austin/MAESTRÍA_AUSTIN/Advanced Predictive Models/HW2/file2.xls", header=TRUE, stringsAsFactors=FALSE)
data1_2<-data.frame(Freq = data1$Freq,Time =data1$Date,DayOfWeek=data1$DayOfWeek)
#the filtering weights
a<-c(rep(c(1/7,0,0,0,0,0,0),6),1/7)
#filtered data
ma_smoothed_data <- data.frame(
  Freq = stats::filter(data1$Freq,
                       sides = 2,
                       filter = a),Time =data1$Date,DayOfWeek=data1$DayOfWeek)
#Splitting the data by day
monday<-data1_2[data1_2$DayOfWeek=="Monday",]
tuesday<-data1_2[data1_2$DayOfWeek=="Tuesday",]
wednsday<-data1_2[data1_2$DayOfWeek=="Wednsday",]
thursday<-data1_2[data1_2$DayOfWeek=="Thursday",]
friday<-data1_2[data1_2$DayOfWeek=="Friday",]
saturday<-data1_2[data1_2$DayOfWeek=="Saturday",]
sunday<-data1_2[data1_2$DayOfWeek=="Sunday",]
# Splitting the smoothed data by day
monday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Monday",]
tuesday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Tuesday",]
wednsday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Wednsday",]
thursday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Thursday",]
friday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Friday",]
saturday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Saturday",]
sunday_s<-ma_smoothed_data[ma_smoothed_data$DayOfWeek=="Sunday",]

# Making that the levels appear in the ggplot in descendin order as Monday, Tuesday, Wednsday, Thursday , Friday, Saturday, Sunday
#instead of having alphabetical order
ma_smoothed_data$DayOfWeek <- factor(ma_smoothed_data$DayOfWeek,
                                     levels=c("Monday", "Tuesday", "Wednesday","Thursday","Friday","Saturday","Sunday"),
                                     labels=c("Monday", "Tuesday", "Wednesday","Thursday","Friday","Saturday","Sunday"))
#Making the plot
gg_all <-    ggplot(ma_smoothed_data,aes(x = Time, y = Freq,group=DayOfWeek),colour=DayOfWeek)+geom_point()+geom_line(lwd=2,aes(color=DayOfWeek)) +
  geom_line(data=monday,size=0.5,aes(color=DayOfWeek))+geom_point(data=monday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black")+
  geom_line(data=tuesday,size=0.5,aes(color=DayOfWeek))+geom_point(data=tuesday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black")+
  geom_line(data=wednsday,size=0.5,aes(color=DayOfWeek))+geom_point(data=wednsday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black")+
  geom_line(data=thursday,size=0.5,aes(color=DayOfWeek))+geom_point(data=thursday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black")+
  geom_line(data=friday,size=0.5,aes(color=DayOfWeek))+geom_point(data=friday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black")+
  geom_line(data=saturday,size=0.5,aes(color=DayOfWeek))+geom_point(data=saturday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black") +
  geom_line(data=sunday,size=0.5,aes(color=DayOfWeek))+geom_point(data=sunday,aes(x=Time,y=Freq),shape = 21, size = 1, fill = "black", color = "black")+
  scale_x_date()+
  labs(title = "Car accidents in Texas",subtitle = "Smoothed Using a Moving Average Filter by Day of Week using a_4 filter")

gg_all
```