Time Series Analysis for San Diego Fire and EMS Dispatchment

Cole Bailey, Kevin Stewart, Lina Nguyen

Shiley-Marcos School of Engineering, University of San Diego

Abstract:

The Emergency Command and Data Center are all in charge of Fire and EMS dispatching in San Diego. Whether it's medical, fire, or rescue, they oversee emergency communication to ensure that Fire or Medical staff and the public are able to easily communicate with each other during dire situations. This project aims to analyze time series dispatchment data in San Diego, in order to investigate trends. Due to the large data set, only instances with fire burns and explosions will be analyzed in our project using the ARIMA package in R-Studio. The results showed that seasonality and incidents are intertwined, and suggest that trends can be analyzed to mitigate problem count and severity by taking proactive measures .

Keywords: ARIMA, San Diego, EMS dispatchment, fire dispatchment, time series, burns

Table of Contents

Background:

The San Diego Emergency and Data Control Center handles all calls for EMS and Fire dispatching, and annually handles over 130,000 emergency calls (Fire-Rescue Department). Based on where the 9-1-1 call is made, calls are routed to a Public-Safety Answering Point. They are then either transferred to respective Police Departments for law enforcement emergencies, or the Metro Zone Emergency Command and Data Center for medical or fire related emergencies (Fire-Rescue Department). Addresses are displayed to dispatchers when calls are made from landlines, while calls made from cell phones only display the cell tower location (Fire-Rescue Department).

Project Objective:

The objective of the project is to identify potential trends over time for incidents that require the dispatchment of Fire and EMS by the San Diego Fire Communications Center. In identifying potential trends for specific incidents (fire, medical, accident, etc....), various fire departments and civil service groups may allocate resources accordingly. For example, a history of increased drunk-driving incidents around the holidays stirs Fire and EMS departments to take necessary precautions to limit the damage.

By tracking the potential trends available, it is possible to establish causes for these incidents. In the case of drunk driving, a potential cause could be holiday celebrations. By determining the cause, future incidents and countermeasures can be properly applied. For drunk driving, an increased number of checkpoints and DUI stops discourage individuals from drinking and driving. Increasing accessibility to public transportation, as well as reducing prices for ride sharing services such as Uber during the holidays can also help reduce frequency of Fire and EMS dispatching.

As a result, the motivation behind this project is to reduce Fire and EMS dispatched incidents both numerically and severity. Only fire burns will be assessed but the process can be repeated over different incidents.

Data Description:

The data set includes 25,000-line items for incidents related to Fire and EMS dispatchment provided by the San Diego Fire Communications center from 2006 - 2021.

Literature Review:

There are several studies and initiatives relevant to the project. Each of these analyze fires and fire-related incidents from a variety of angles and perspectives. Information surrounding the causes, intensities and geographic involvement have all been documented and published. A review of these sources is instrumental in analyzing the San Diego records.

The first source focuses on fires and related instances in the country of Jordan. (Swies, 2006) discusses the “types of accidents and the causes of fires.” The analysis of the records showed that half of the records occurred in the wildlands with an additional twenty percent occurring in residential homes. This information is useful for the San Diego study by demonstrating the importance of location. If the San Diego data contained geographical information, it would be possible to include this information in any trend analysis and identification. The study also found that “children's carelessness [was] identified to be the highest cause of fire incidents.” This information is relevant in determining a driver in fire causation. For the study, causes are associated with trends and incident occurrence. The identification of the drivers in San Diego would allow for further sensitivity studies to be utilized in making strategic decisions.

An additional fire related study was conducted in Dubai. In this analysis, (Alqassim, 2014), recorded 5,000 fire related incidents from the years 2006-2013 with “more than one third of the total number of incidents involving motor vehicles.” These accidents were discovered to account for more than half of the recorded incidents in Dubai. Additionally, “A further one third of the incidents reviewed were in residential units,” providing a stark increase to the 20% of residential fires in the Jordan study. Lastly, Dubai accredited electrical failures as their leading cause of incidents as opposed to children carelessness.

There are plenty of studies determined to identify the risk factors associated with fire related instances such as house-fires, injuries, and automobile accidents. One study (Turner & Johnson, 2017), evaluated on the “characteristics commonly associated with increased risk of house fire incidents, injuries and fatalities included: higher numbers of residents, male, children under the age of 5 years … and buildings in poor condition.” This report is like the previous two. This study closely determined themes and trends for the causation of incidents rather than the incidents themselves. As a result, it provides insight on preventative measures centered on the build-up rather than the actual cause. This study shows the importance of not only determining the highest causes – but the causes to the cause of the fire.

Lastly, a study conducted on incidents and their respective causes was carried out in Toronto. This study took aspects of all the studies above. It identified the number of different causes and the characteristics for the leading causes. The mission to "determine the extent to which existing data can be used as a baseline to improve fire prevention and response activities at local levels" (Asgary and Levy, 2009). The study found that significant "differences exist with respect to fire causation over time and space". A time-series approach is imperative for this study.

In conclusion, the evaluation of previous studies has indicated several things. Fire incidents are unique based on time and geographics. This is indicated by different leading causes at different locations in the world. This indicates that causes for San Diego need to be analyzed and determined independently of other studies. Leading causes in Jordan, Dubai and Toronto may not be leading causes in San Diego. Additionally, the incidents needed to be isolated by date. The leading causes for fires are different at different points in the year. After the identification of the leading incidents and their date of occurrence, causations and characteristics of the causations may be determined. In turn, it is imperative then that preventive fire tasks and resources are allocated accordingly for the city of San Diego.

Project Overview:

This San Diego Fire and EMS dispachment data set was very large, plagued with a lot of dirty data. Due to limitations of time and manpower, our team decided to perform exploratory data and time series analysis only on dispatchements that were related to fire burn and explosion problems in San Diego. The fire burn and explosion incidences were plotted as a time series and seasonal plots. ACF and PACF were used to check for autocorrelation, and the best ARIMA model was found to fit the data set.

Data Preparation:

The data set includes problem counts for the number of occurrences that existed for a specific issue for each month-year, meaning that the 25,000 incidents occurring is a total of over 2,000,000 incidents in the data set. The dataset also contained a lot of misspellings, as well as a couple year response outliers most likely due to human error, resulting in a dataset that requires a lot of time and manpower to clean and process the entire set. Due to the limitations of our class, exploratory data and time series analysis was run on the highest occurring feature, fire burns and explosions. Subsetting the data only into fire burns or explosions resulted in a cleaner dataset with no outliers.

The number of fire burns and explosions were plotted on a barplot to view the number of incidents per month.

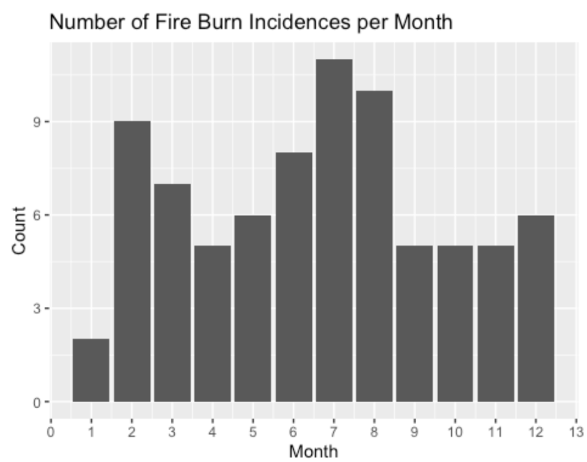


Figure 1. Number of Incidences of Fire Burns and Explosions per Month

This barchart shows that peaks in fire burns and explosion incidents occur in July, August and February. The peak in July can be due to Independence Day firework celebrations.

The number of fire burns and explosions per year were also plotted on a barchart.

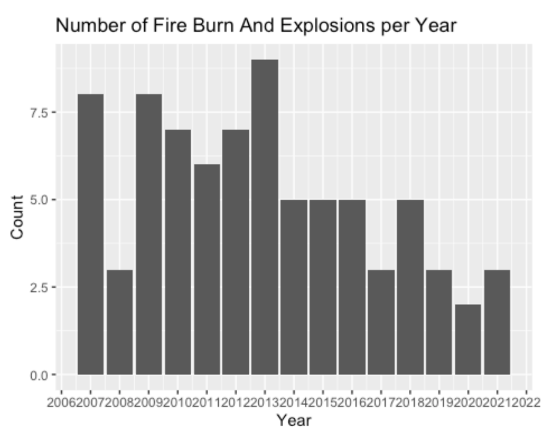


Figure 2. Number of Incidences of Fire Burns and Explosions per Year

Peaks in fire burns and explosions peaked in 2013, 2007, and 2009, and decreased from 2019-2021 possibly due to the COVID-19 pandemic.

The burn data month responses, and the logarithm burn data month responses were plotted on a time series plot.

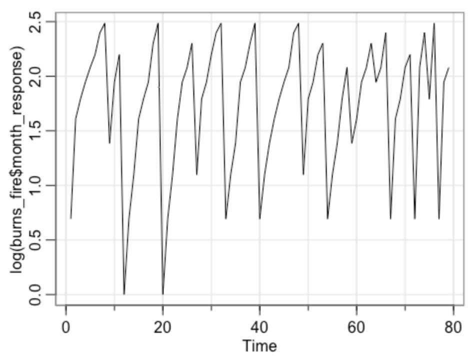


Figure 3. Time Series plot of logarithm of Fire Burns and Explosions per Month

The data shows that there is seasonality that needs to be removed. This was done using a seasonality plot.

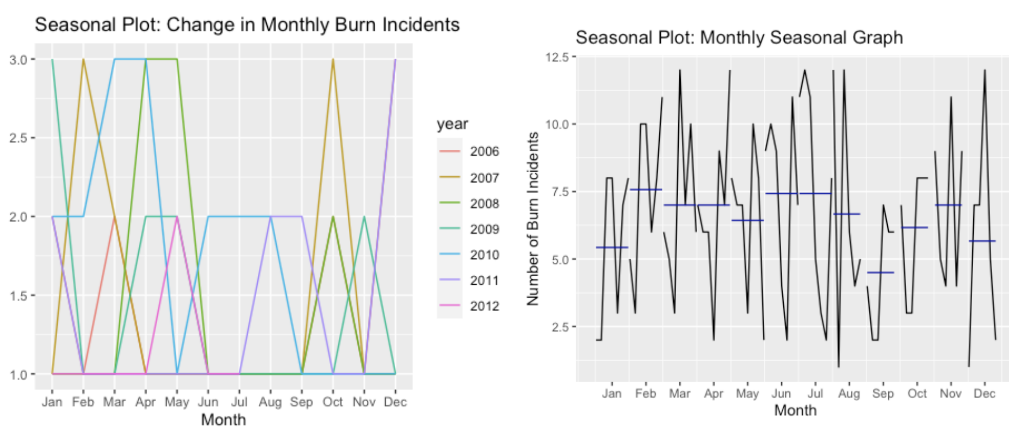


Figure 4. Change in Monthly Burn and Explosion Incidents Seasonality Plot (left) and Seasonal Plot (right)

The monthly seasonal plot shows that there is a spike in the winter to spring months and drastic decrease at the start of summer.

ACF and PACF were plotted to check for autocorrelation.

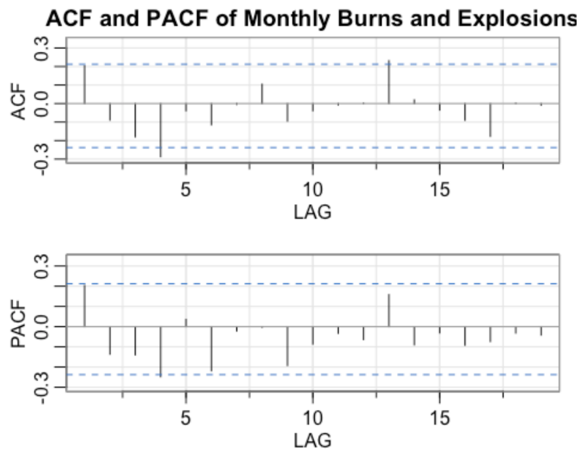


Figure 5. ACF and PACF of Fire Burns and Explosions

The ACF and PACF plot show that the data set needs to be differenced to remove trend and seasonality due to the significant amount of correlation.

The best ARIMA model was then determined to be ARIMA(4,0,0) with zero mean for the fire burns and explosions because it was the lowest AIC score, using the auto.arima function. This was also confirmed through forecasting and checking residuals.

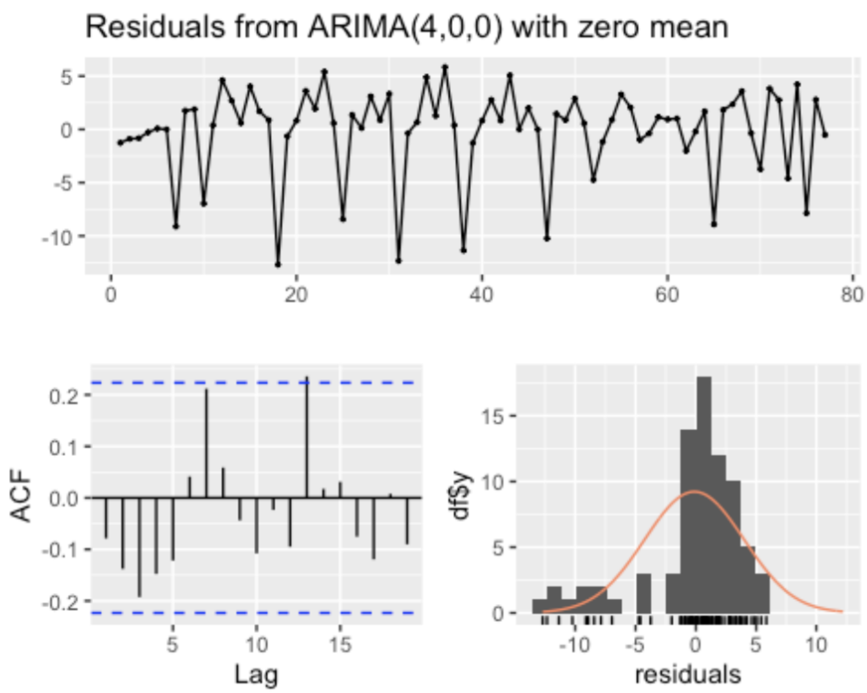


Figure 6. Residuals of Fire Burns and Explosions using Forecast and Lag

Most of the autocorrelation is removed from the model as shown in the ACF plot.

The dickey fuller test was used to check for stationarity, and was found that the p-value was 0.01, so we can reject the null hypothesis because there is stationarity.

Lastly, the ARIMA(4,0,0) with zero mean was then used to forecast monthly trends of fire burns and explosion incidents in San Diego of old and recent data.

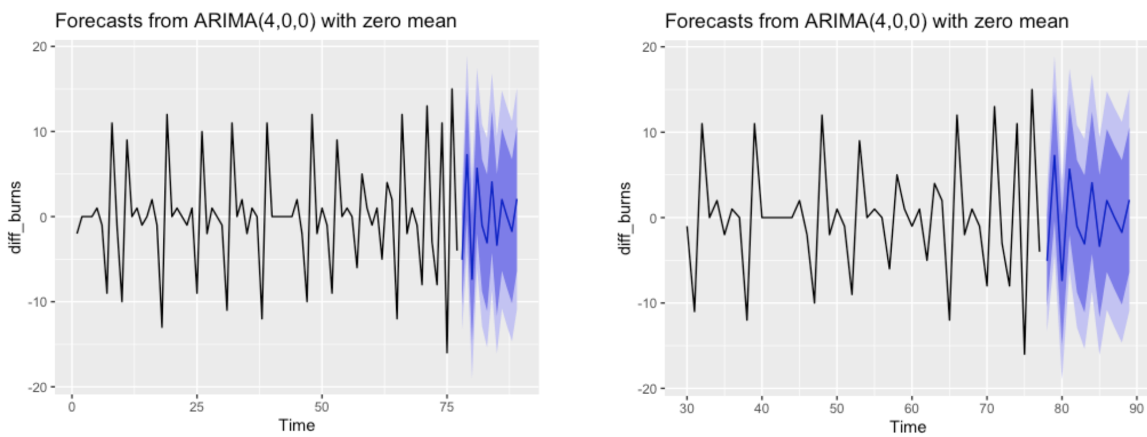


Figure 7. Forecasts of ARIMA(4,0,0) with zero mean of old (left) and recent (right) data

Discussion:

The project covered several topics and the distribution of fire-related events. By examining key characteristics of specific and timely events, the project was able to detect seasonality higher in specific events more so than others. We found that the most frequent occurring incidents stayed relatively constant throughout the timelines with local minima and maxima over the timeline. Additionally, months towards the end of the year consistently had higher averages of events. With the majority of the major incidents happening in September - December, fire staff can take precautionary measures such as maximizing volunteer work, increasing on-site staff, and making equipment immediately available.

This success could be extended to other emergency service incidents. The police or highway patrol could adopt the study unto their own incidents. The documentation and analysis of police related incidents (domestic abuse, gang violence, etc….) could be related with the project process to incite their own precautionary steps. Additionally, CHP could utilize the process to evaluate the most likely risk of highway/freeway incidents.

In collaboration with other emergency services and first responders, the project on the fire department could serve as a template for the city of San Diego to justify additional allocations to specific departments and time-specific intervals.

There are also additional steps that could be taken with the data.

Conclusion:

The project in its entirety was designed to improve San Diego readiness and availability in regards to fire burn and explosion related incidents. By evaluating specific incidents and problem counts as a whole for events that occured in San Diego from 1990 - 2020, future actions could be justified and decided as a precautionary measure. There were plenty of opportunities to study and evaluate key incidents.

By first evaluating all fire burn and explosion related incidents, the analysis showed that most events, unmindful of any specific case, occur in the latter months of the year. Namely, September - January there are on average thirty more incidents than Spring and Summer months. This information in its solidarity can serve to mitigate future incidents by inspiring look aheads and decision making. San Diego can utilize the analysis to increase staff and methods to lower the risk. Proactive decision making such as DUI checkpoints and fire retardant spreads can isolate and prevent some of the incidents before they happen.

There were additional noteworthy bits from the incidents as a whole. Namely, the massive spike of roughly four hundred percent around the year 1984. The incident count this year jumped to over 700. While this can be considered an anomaly with only the late 2000’s even close to reaching similar numbers (roughly 500), the possibility of a repeat 1984 is possible. As a result, staff and resources must be available to combat this possibility at any given time.

Additionally, the project evaluated specific incidents for analysis as opposed to general problem counts. Primarily, actual fires were the main incident targeted for evaluation. On average, there stands to be around six burning incidents per month in the city of San Diego. Most noteworthy is that these incidents are above average in the Summer months of June - August. The duration of these burn incidents also is dramatically higher than in the winter months.

Also, the project utilized both autocorrelation and Arima processes to evaluate the seasonality and stationarity of the data. By differencing and utilizing processes to forecast monthly trends of burn incidents in the city of San Diego, the problem count and severity may be reduced. This is shown in the final ARIMA forecasting model.

In conclusion, the project shows that seasonality is indeed a factor in fire related incidents for the city of San Diego. The project primarily focused on incidents as a whole and fires themselves but can be expanded to evaluate additional incidents with the highest cardinality. This will lead to additional forecasting across the board and ultimately make the difference in limiting both the number and severity of these incidents.

References

Alqassim, M. A., & Daeid, N. N. (2014). Fires and related incidents in Dubai, United Arab Emirates (2006–2013). *Case Studies in Fire Safety*, *2*, 28–36. https://doi.org/10.1016/j.csfs.2014.10.001

Asgary, A., Ghaffari, A., & Levy, J. (2010). Spatial and temporal analyses of structural fire incidents and their causes: A case of Toronto, Canada. *Fire Safety Journal*, *45*(1), 44–57. https://doi.org/10.1016/j.firesaf.2009.10.002

Fire-Rescue Department. (n.d.). *Dispatch Process*. City of San Diego Official Website. Retrieved December 5, 2021, from https://www.sandiego.gov/fire/about/organization/commcenter/dispatch

Sweis, F. K. (2006). Fires and related incidents in Jordan (1996–2004). *Fire Safety Journal*, *41*(5), 370–376. https://doi.org/10.1016/j.firesaf.2006.02.002

Turner, S. L., Johnson, R. D., Weightman, A. L., Rodgers, S. E., Arthur, G., Bailey, R., & Lyons, R. A. (2017). Risk factors associated with unintentional house fire incidents, injuries and deaths in high-income countries: A systematic review. *Injury Prevention*, *23*(2), 131–137. https://doi.org/10.1136/injuryprev-2016-042174

Predicting Fire Burn Occurrences

Lina Nguyen,Cole Bailey,Kevin Stewart

12/3/2021

Importing packages

```
library(tidyverse);library(tidyr);library(zoo);library(zoo);library(xts);libr
ary(tseries);library(astsa);library(lubridate);library(ggplot2);library(dplyr
);library(fpp2);library(dplyr);library(tsibble);library(tseries);library(fore
cast);library(sarima);library(fpp2);library(ggplot2)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.1 ──

## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.3     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   2.0.0     ✓ forcats 0.5.1

## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

## ── Attaching packages ───────────────────────────────────────────── fpp2
2.4 ──

## ✓ forecast  8.15     ✓ expsmooth 2.3
## ✓ fma       2.4

##

##
## Attaching package: 'fpp2'

## The following object is masked from 'package:astsa':
##
##     oil

##
## Attaching package: 'tsibble'

## The following object is masked from 'package:lubridate':
##
##     interval

## The following object is masked from 'package:zoo':
##
##     index

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, union

## Loading required package: stats4

##
## Attaching package: 'sarima'

## The following object is masked from 'package:astsa':
##
##     sarima
```

#loading the data file and seperating string values

```
df_fire <-
read.csv("/Users/kevinstewart/Desktop/business_a/san_diego_fire_incident.csv"
, sep = ",")
```

#viewing data

```
head(df_fire)
```

```
##   fd_problem_nature_agg_datasd_v1            X
X.1
## 1                     agency_type address_city
problem
## 2                            Fire     SAN DIEGO                  Ringing
Alarm
## 3                            Fire     SAN DIEGO   Cardiac / Respiratory
Arrest
## 4                            Fire     SAN DIEGO              Assault/Rape
(L4)
## 5                            Fire     SAN DIEGO        Assist PD - Ladder
Bldg
## 6                            Fire     SAN DIEGO Back Pain (Non Traumatic)
(L4)
##             X.2            X.3           X.4
## 1 problem_count month_response year_response
## 2             1              1          1900
## 3             1              9          2001
## 4             2              9          2006
## 5             4              9          2006
## 6            42              9          2006
```

correct colnames & and subsetting data based on year and most occurring problem

```r
names(df_fire) <- df_fire %>% slice(1) %>%
unlist()
fire <- df_fire %>% slice(-1)
```

Mutate the data to month response

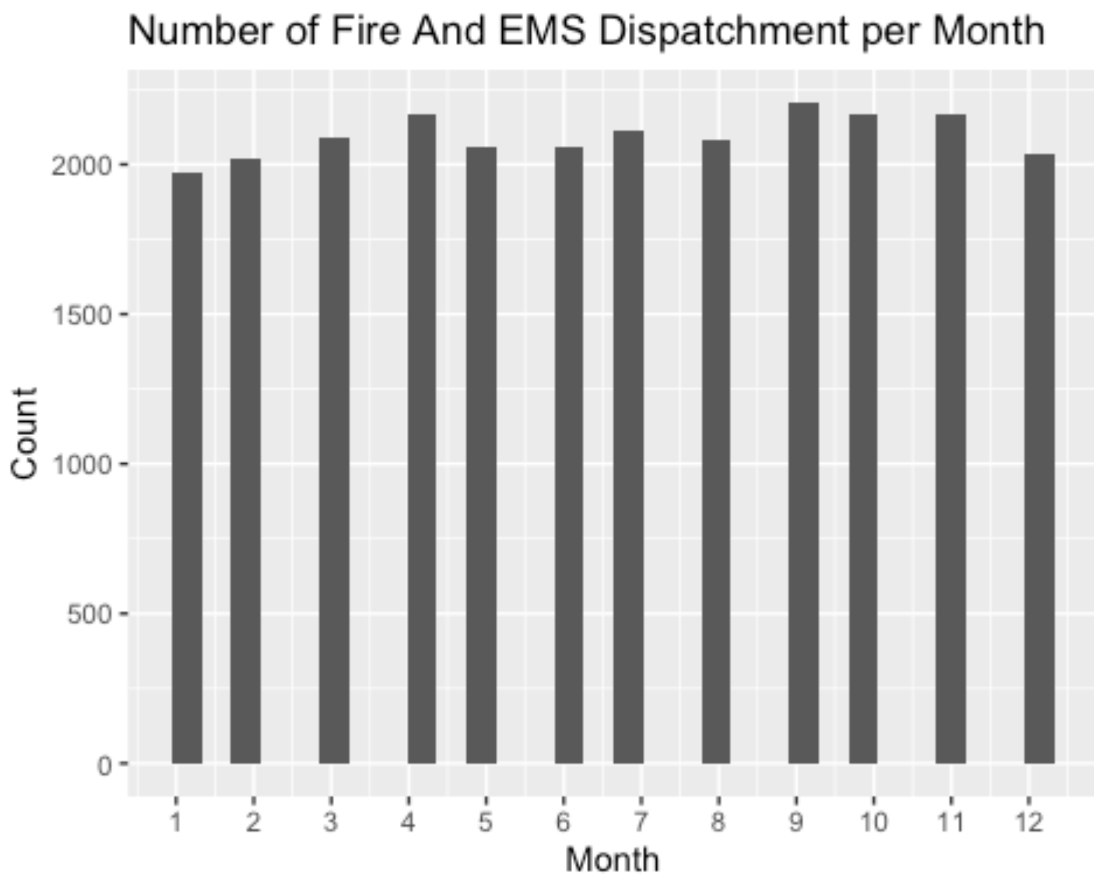
```r
fire$problem_count <- as.numeric(fire$problem_count)
fire$month_response <- as.numeric(fire$month_response)
fire$year_response <- as.numeric(fire$year_response)
fire$problem <- as.factor(fire$problem)

########################################
# verify structure of data
str(fire)

## 'data.frame':    25138 obs. of  6 variables:
##  $ agency_type   : chr  "Fire" "Fire" "Fire" "Fire" ...
##  $ address_city  : chr  "SAN DIEGO" "SAN DIEGO" "SAN DIEGO" "SAN DIEGO"
...
##  $ problem       : Factor w/ 485 levels ".Confined Space/Trench
Rescue",..: 329 111 70 77 84 88 97 100 103 119 ...
##  $ problem_count : num  1 1 2 4 42 2 9 1 1 7 ...
##  $ month_response: num  1 9 9 9 9 9 9 9 9 9 ...
##  $ year_response : num  1900 2001 2006 2006 2006 ...
```

Verify the length and dimensions of the data

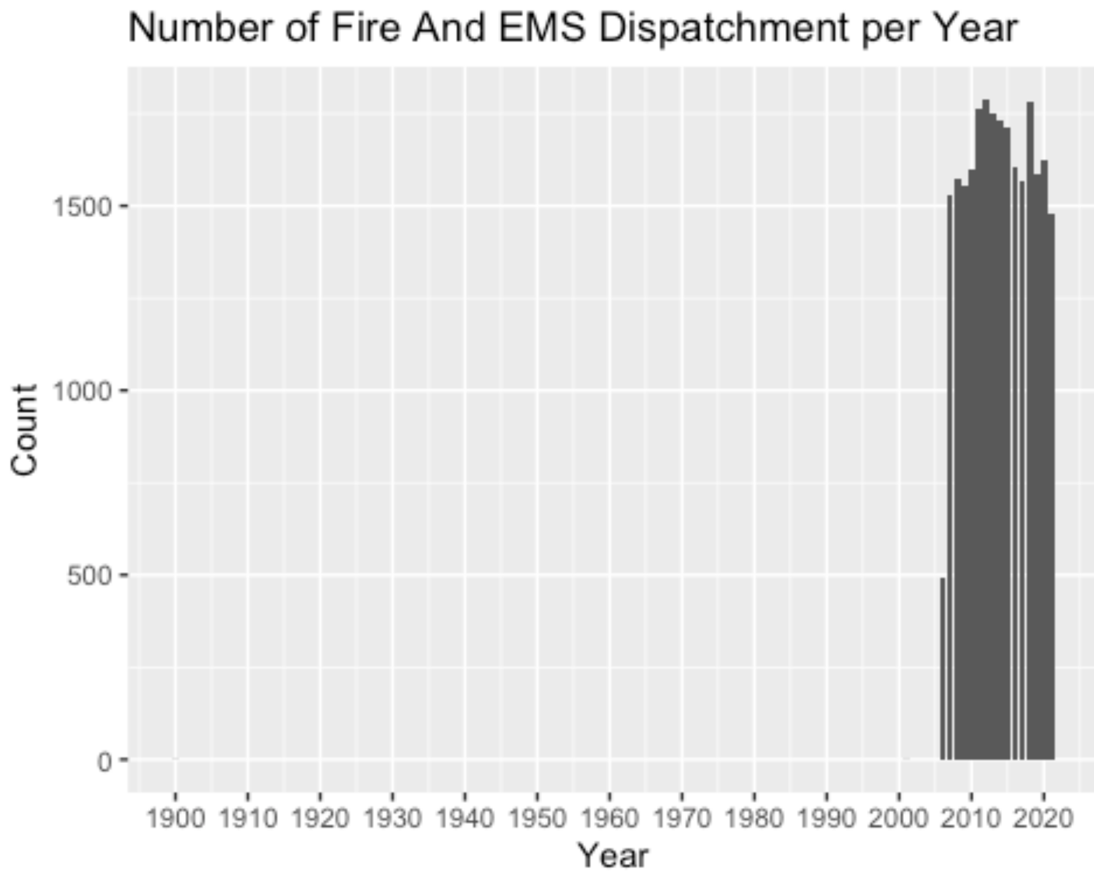
```
################################
# Instantiate the variable
burns_fire <- fire %>%
  filter(problem == "Burns / Explosion (L3)")
```

Viewing frequency of data

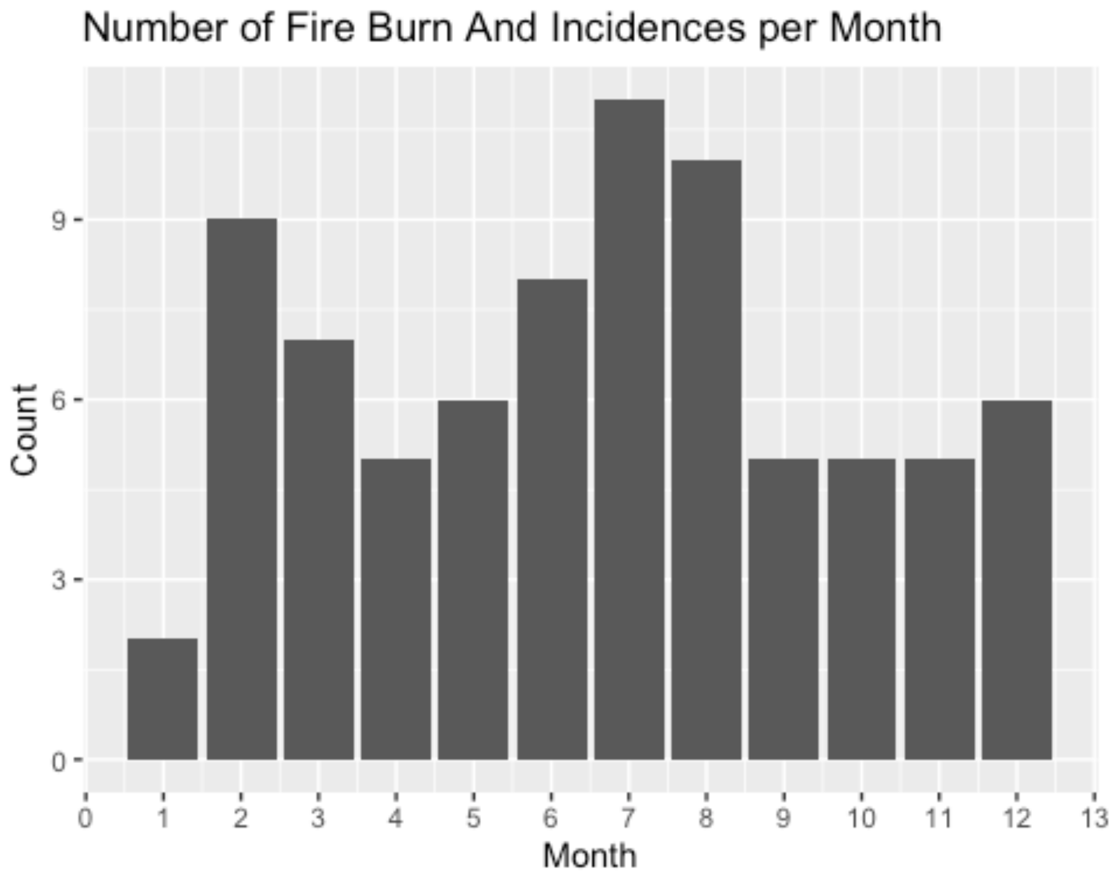
```
#############################################
ggplot(data = fire) + geom_histogram(mapping = aes(x = month_response)) +
scale_x_continuous(breaks = scales::pretty_breaks(n = 12)) + ggtitle('Number
of Fire And EMS Dispatchment per Month') + xlab('Month') + ylab('Count')

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



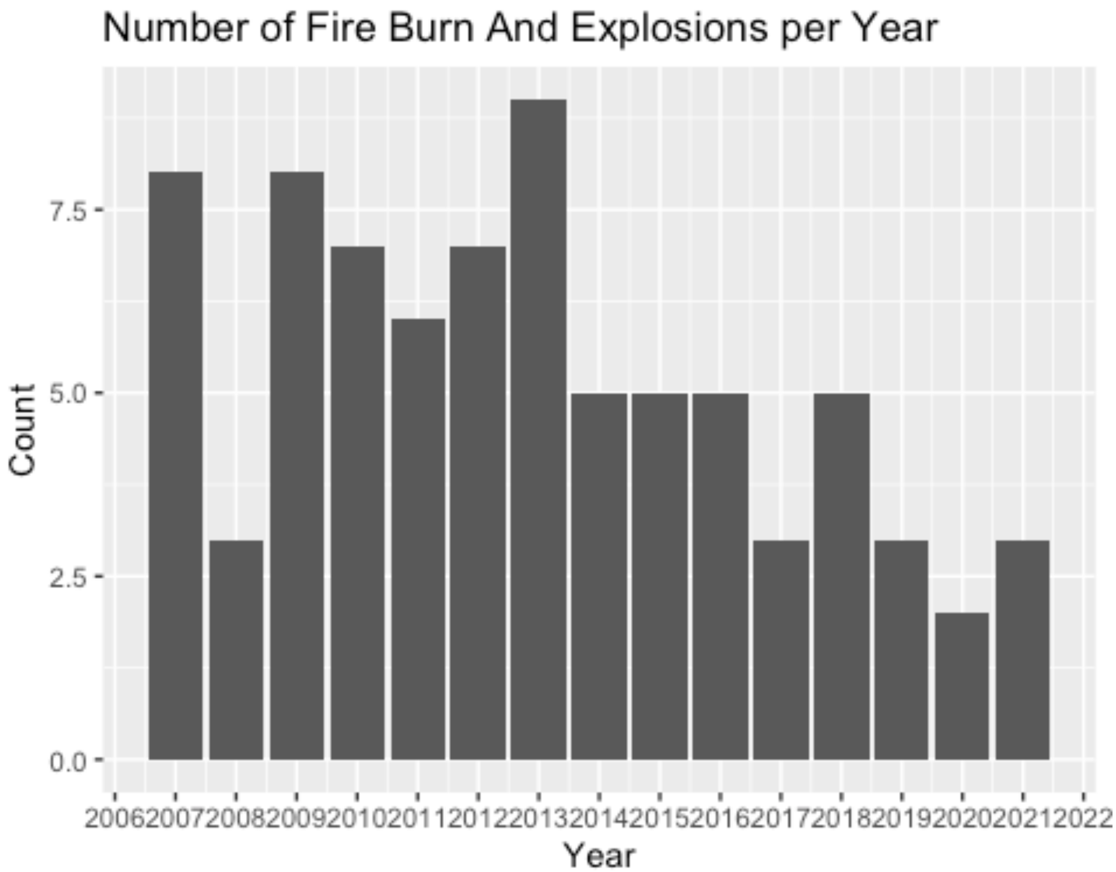

```
#############################################
ggplot(data = fire) + geom_bar(mapping = aes(x = year_response)) +
ggtitle("Number of Fire And EMS Dispatchment per Year") + xlab("Year") +
ylab("Count") + scale_x_continuous(breaks = scales::pretty_breaks(n = 16))
```




```
#############################################
ggplot(data = burns_fire) + geom_bar(mapping = aes(x = month_response)) +
scale_x_continuous(breaks = scales::pretty_breaks(n = 12)) + ggtitle('Number
of Fire Burn And Incidences per Month') + xlab('Month') + ylab('Count')
```

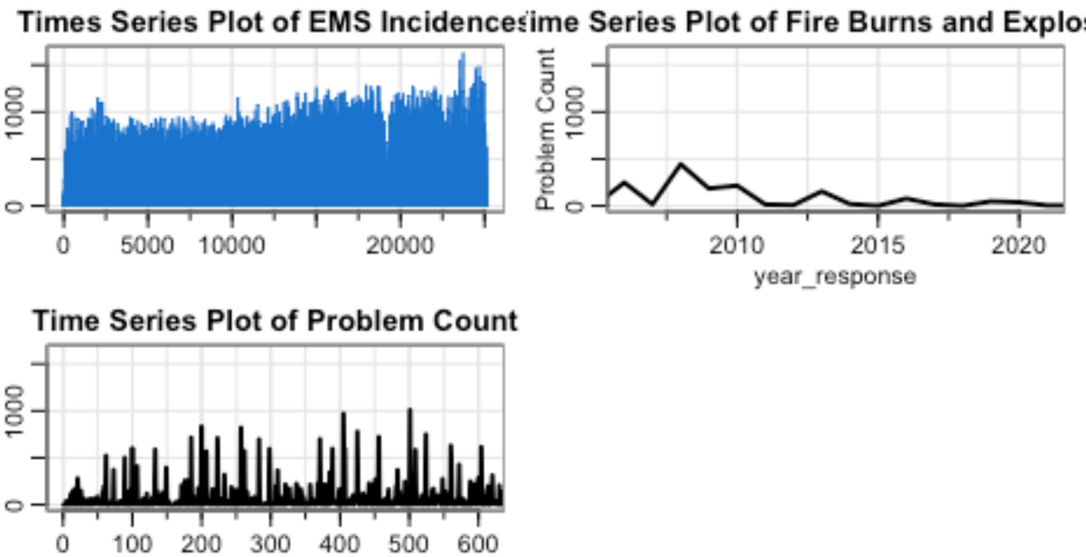



```
##################################################
ggplot(data = burns_fire) + geom_bar(mapping = aes(x = year_response)) +
scale_x_continuous(breaks = scales::pretty_breaks(n = 12)) + ggtitle('Number
of Fire Burn And Explosions per Year') + xlab('Year') + ylab('Count')
```



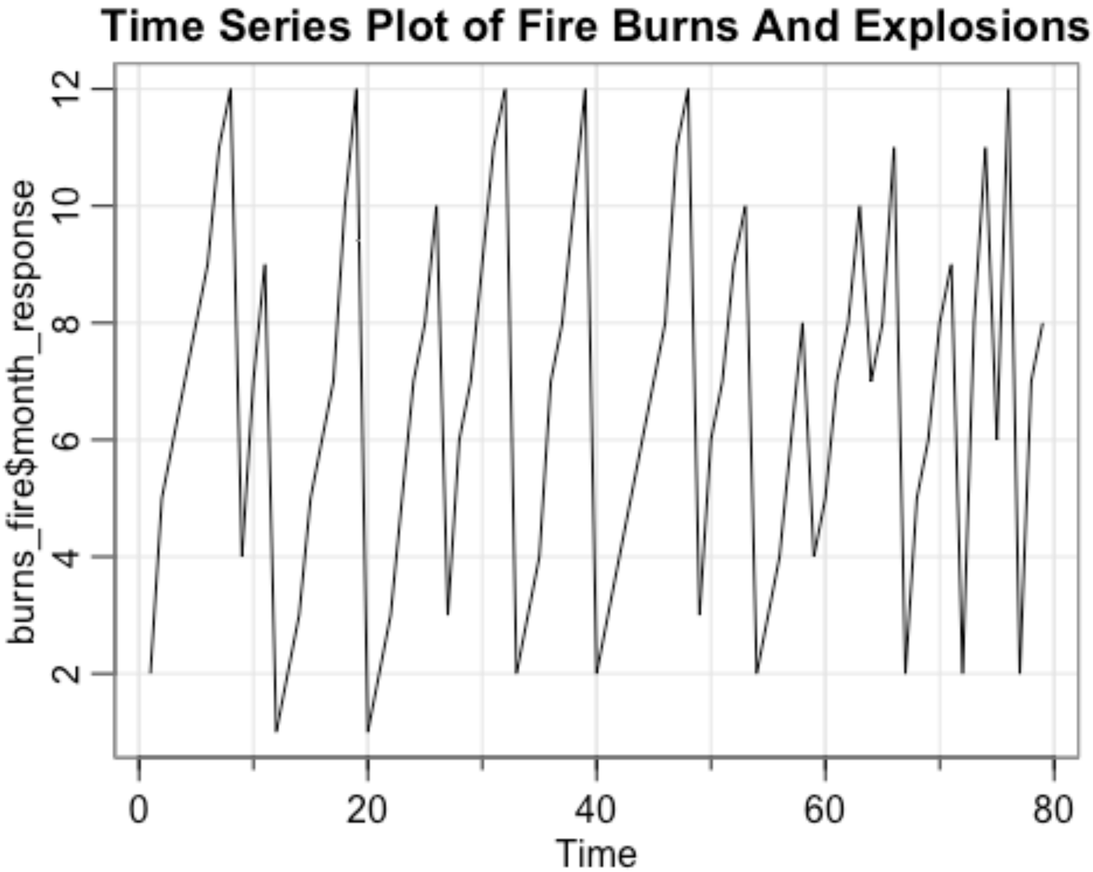
Looking at time series data

```
par(mfrow=c(3:2))
tsplot(fire[,4], ylab = "", xlab = "", type = "l", main = "Times Series Plot
of EMS Incidences", col = 4)
tsplot(fire[,4], xlab = "year_response", ylab = "Problem Count", main = "Time
Series Plot of Fire Burns and Explosions", type = "l", lwd=2, xlim = c(2006,
2021))
tsplot(fire[,4], xlab = "", ylab = "", main = "Time Series Plot of Problem
Count", type = "l", lwd=2, xlim = c(1,612))
```

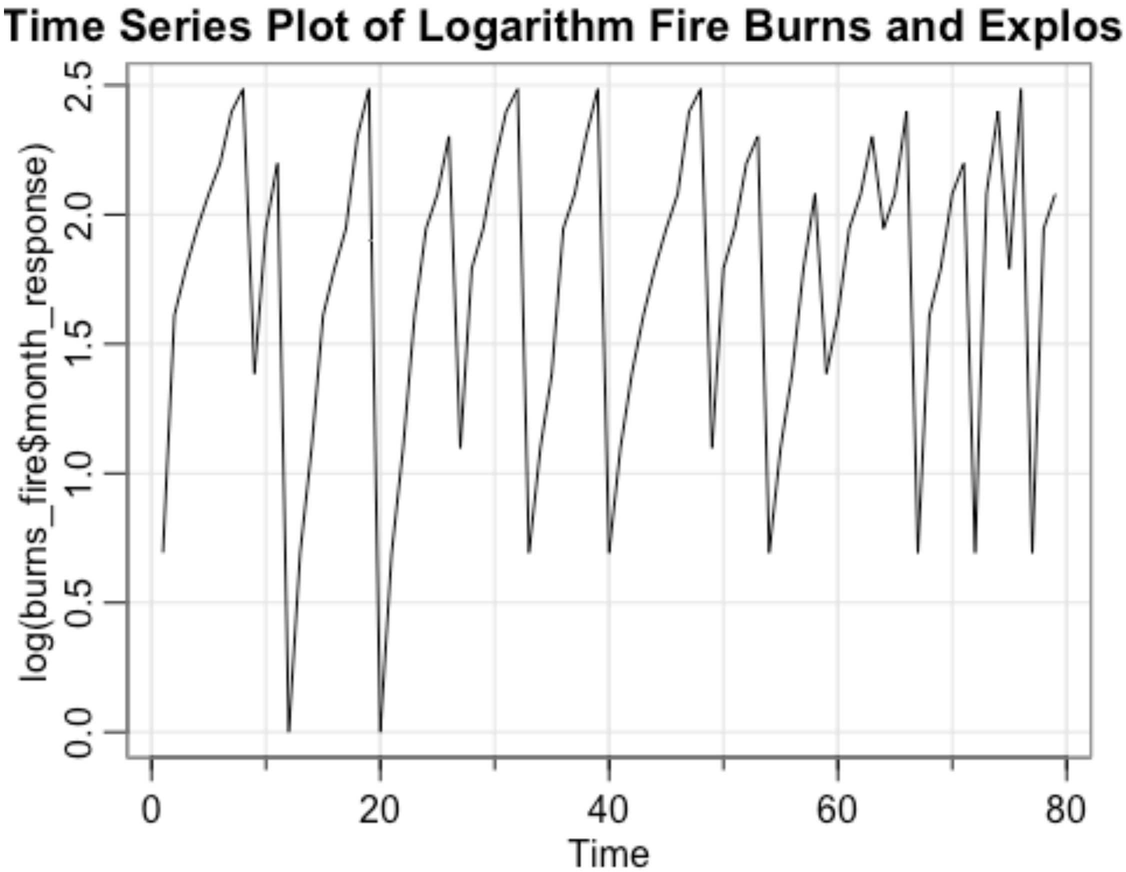


view data by individual fire incidents

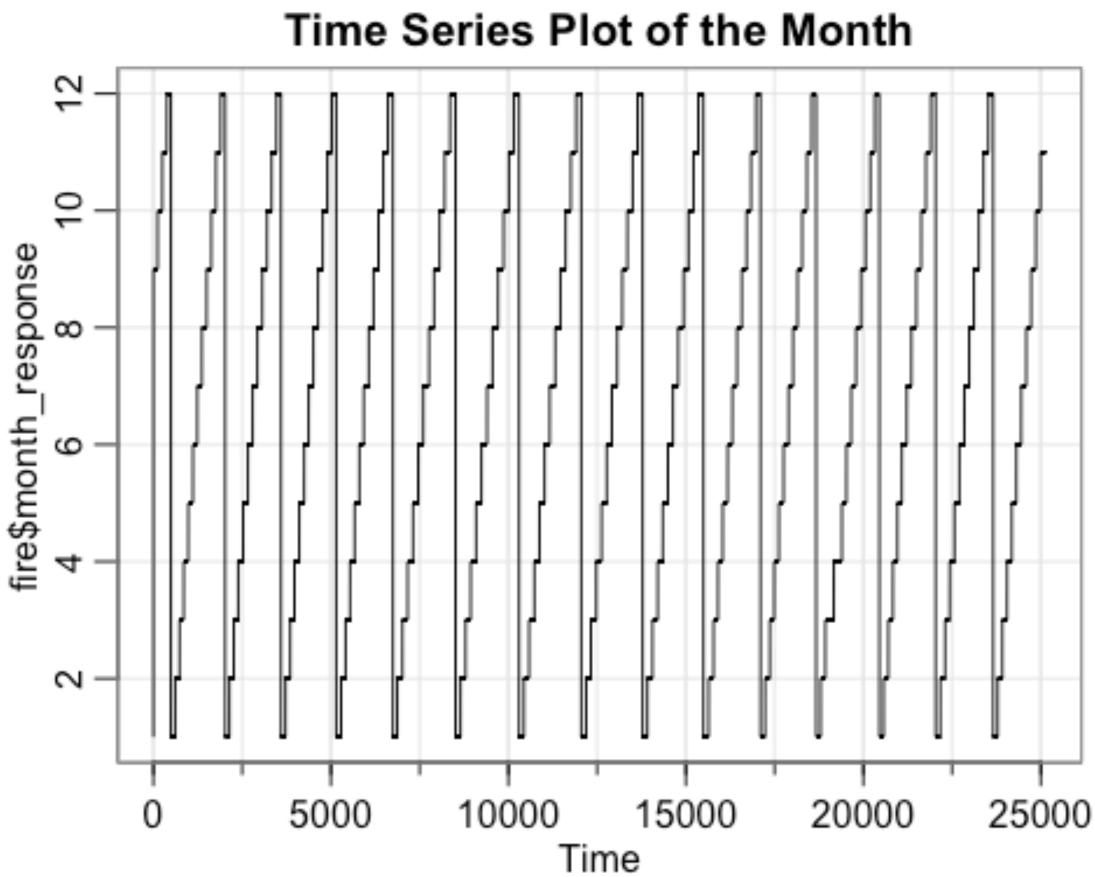
```
tsplot(burns_fire$month_response, main = "Time Series Plot of Fire Burns And
Explosions")
```



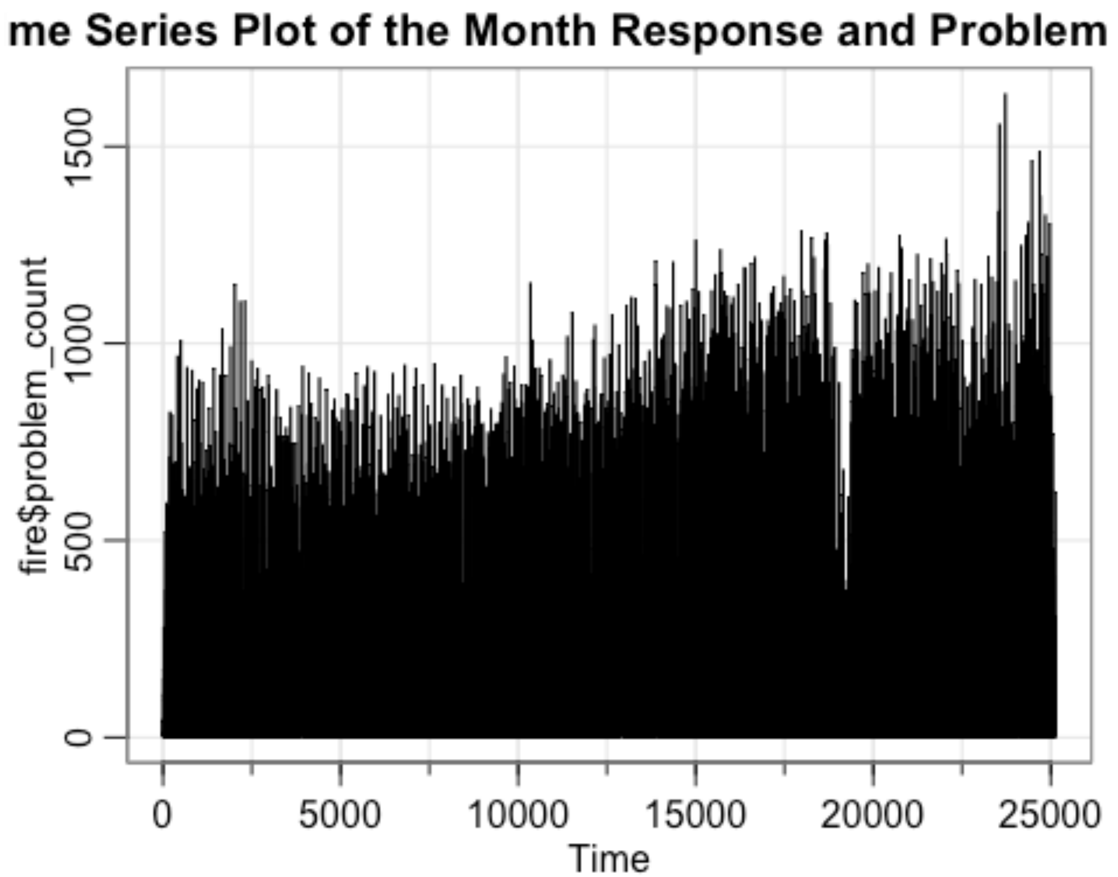

```r
# logging the data graphically
tsplot(log(burns_fire$month_response), main = "Time Series Plot of Logarithm
Fire Burns and Explosions")
```




```
# the data shows that there is a there is also seasonality in the data which
will need to be removed.
########################################
#viewing first variables of the burns month response data
head(burns_fire$month_response)

## [1] 2 5 6 7 8 9

#####################################
# graphically displaying month response and problem count
tsplot(fire$month_response, main = "Time Series Plot of the Month")
```

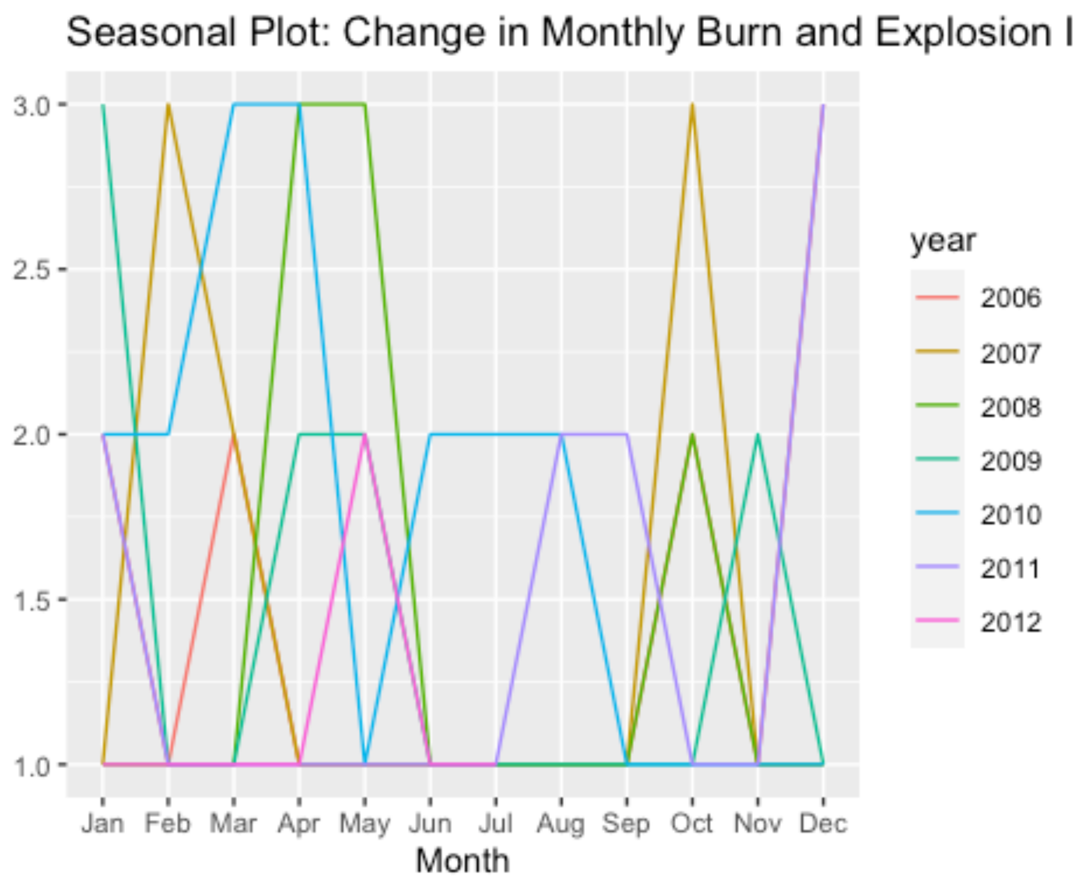



```
tsplot(fire$problem_count, main = "Time Series Plot of the Month Response and
Problem Count")
```

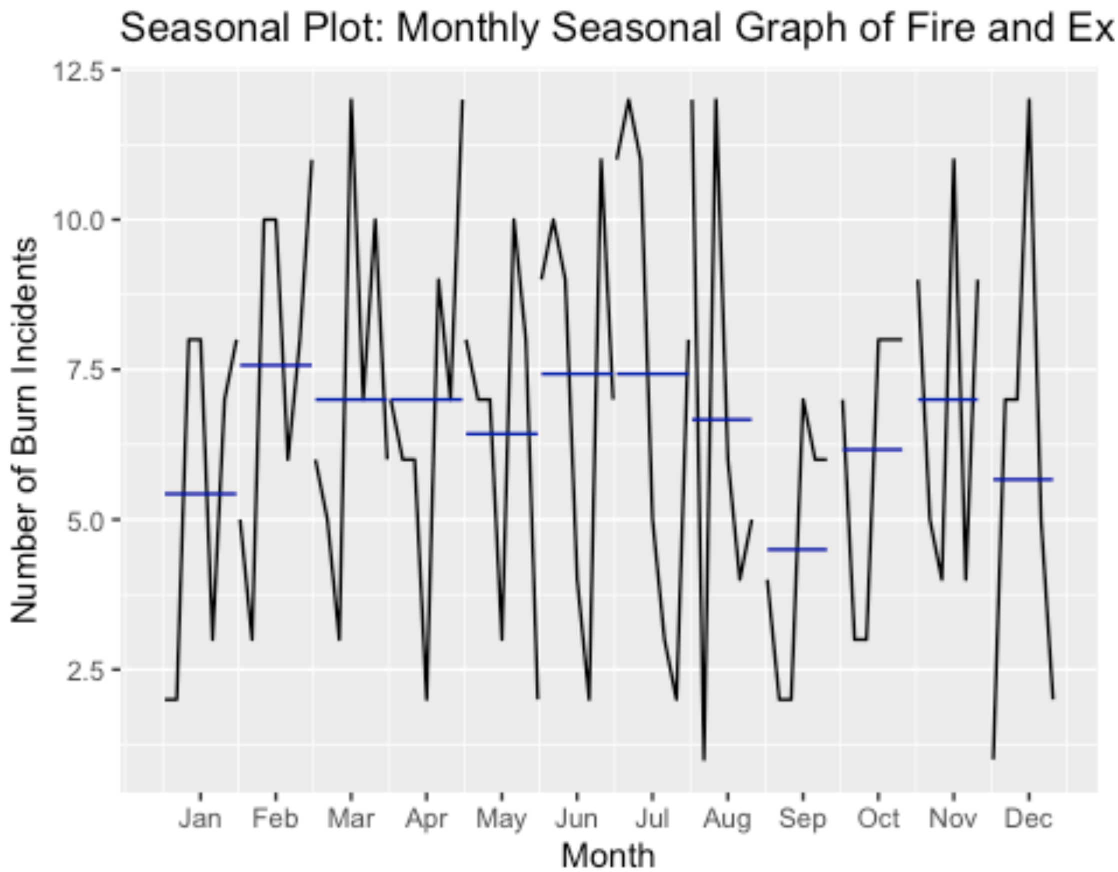


Declaring data as time series

```
# creating a time series variable
Y = ts(burns_fire[,4],start = c(2006,1),frequency = 12)
###############################################
# create a time series object to check for seasonality
ts_fire_b <- ts(burns_fire[,5],start = c(2006,1),frequency = 12)

# Checking seasonality with the seasonal graph
ggseasonplot(Y) + ggtitle("Seasonal Plot: Change in Monthly Burn and
Explosion Incidents")
```



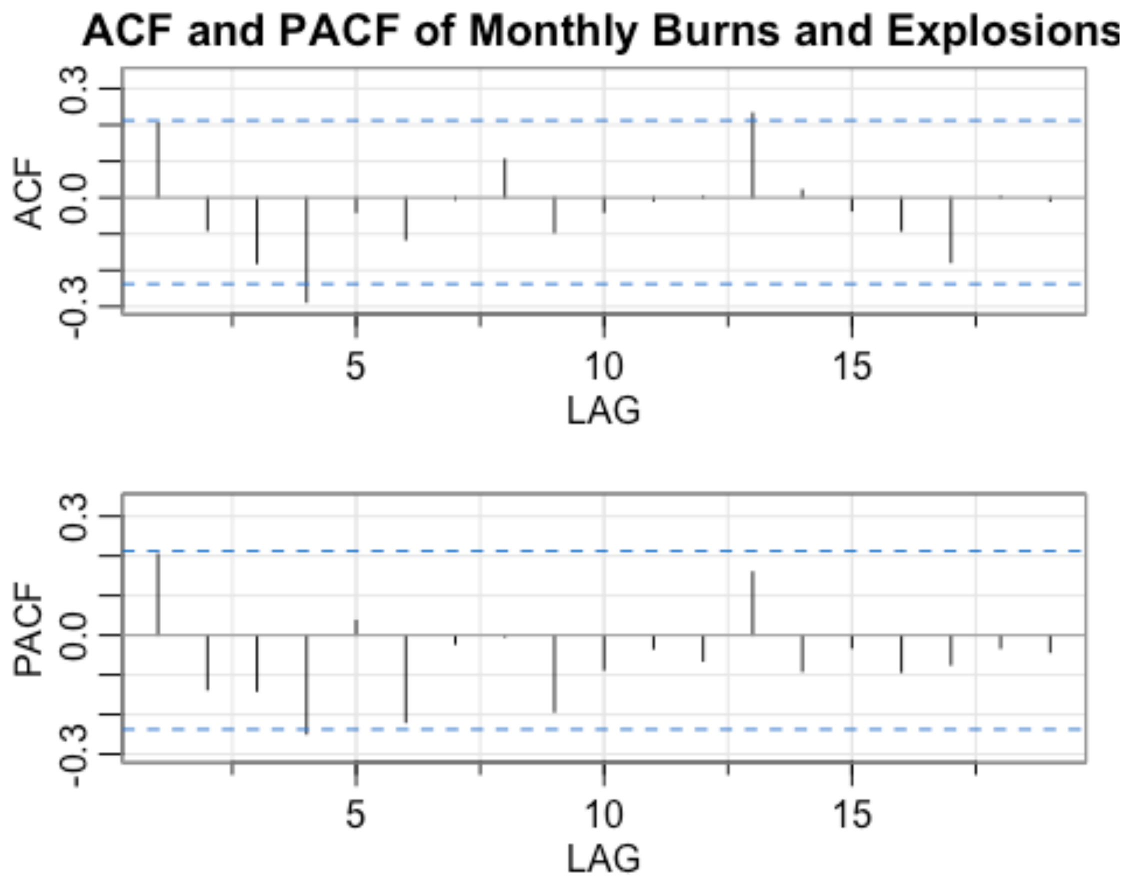

```
###############################################
ggsubseriesplot(ts_fire_b) +
  ggtitle("Seasonal Plot: Monthly Seasonal Graph of Fire and Explosion
Incidences")+ ylab("Number of Burn Incidents")
```




```
#There shows that there is a spike in the winter to spring months and drastic decrease at the start of summer.
```

Looking at acf and pacf plot to check autocorrelation

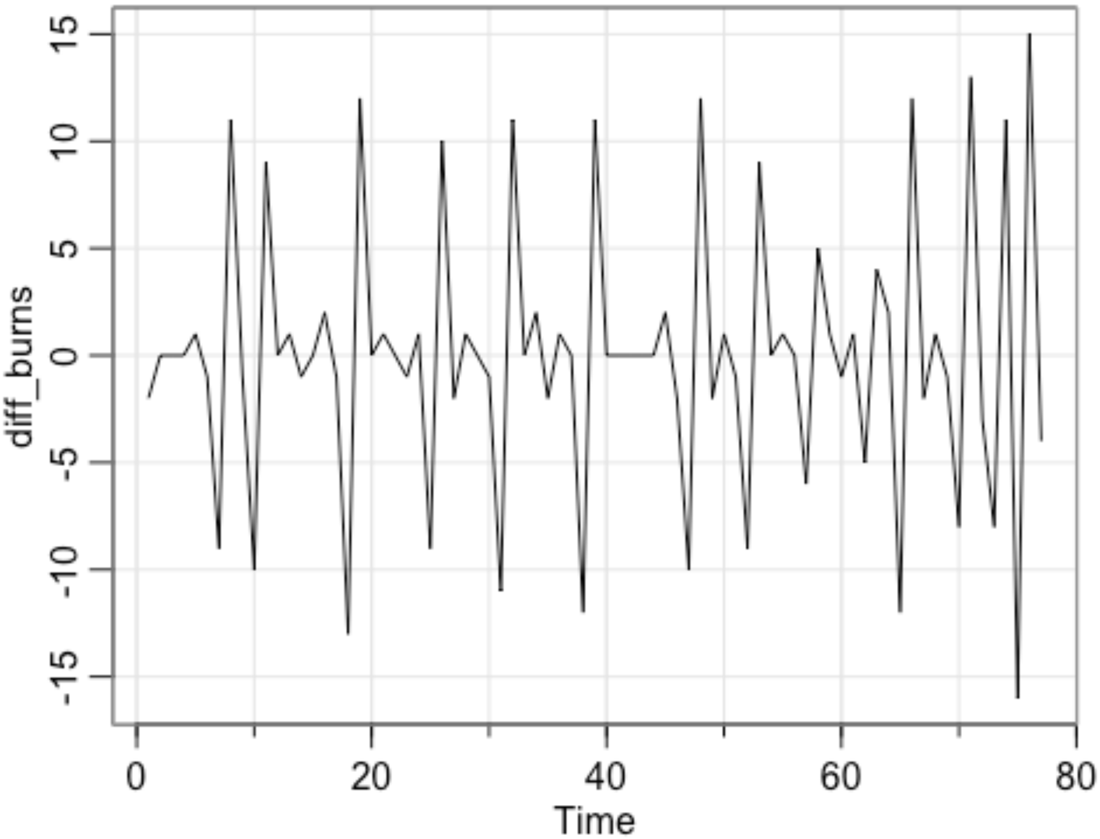
```
head(acf2(burns_fire$month_response,main =  "ACF and PACF of Monthly Burns and Explosions"))
```



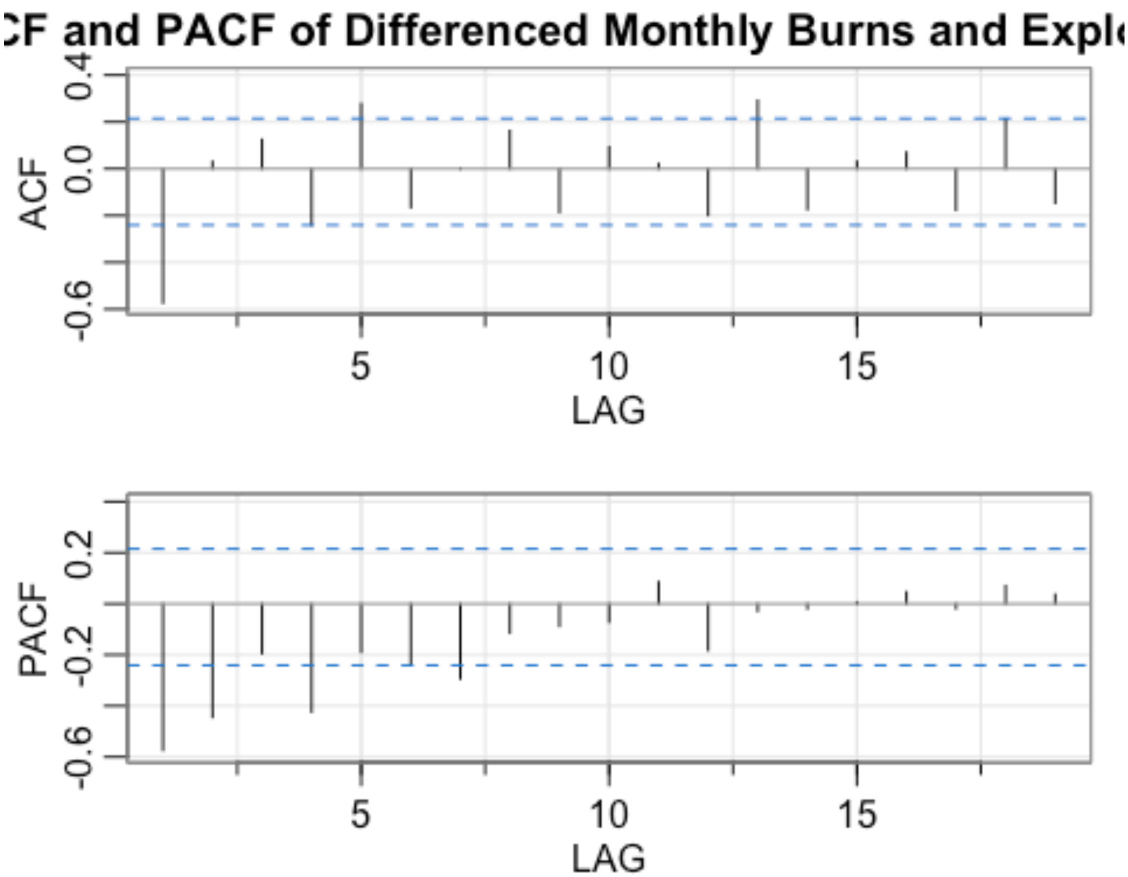

```
##       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7] [,8]  [,9] [,10] [,11] [,12]
## ACF    0.2 -0.09 -0.18 -0.29 -0.04 -0.12  0.00  0.1 -0.09 -0.04 -0.01  0.00
## PACF   0.2 -0.14 -0.14 -0.25  0.04 -0.22 -0.02  0.0 -0.19 -0.09 -0.03 -0.06
##      [,13] [,14] [,15] [,16] [,17] [,18] [,19]
## ACF   0.23  0.02 -0.04 -0.09 -0.18  0.00 -0.01
## PACF  0.16 -0.09 -0.03 -0.09 -0.07 -0.03 -0.04

#this show a significant amount of correlation in
```

#Differencing the data to obtain stationarity to remove trend and seasonality

```
diff_burns <- diff(diff(burns_fire$month_response))
tsplot(diff_burns)
```




```
# acf and pacf plots
acf2(diff_burns, main = "ACF and PACF of Differenced Monthly Burns and
Explosions")
```




```
##       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10] [,11]
[,12]
## ACF  -0.57  0.03  0.12 -0.24  0.28 -0.17  0.00  0.16 -0.19  0.09  0.02 -
0.20
## PACF -0.57 -0.44 -0.19 -0.42 -0.19 -0.24 -0.29 -0.11 -0.09 -0.07  0.09 -
0.18
##      [,13] [,14] [,15] [,16] [,17] [,18] [,19]
## ACF   0.29 -0.17  0.03  0.07 -0.18  0.21 -0.15
## PACF -0.03 -0.02  0.01  0.05 -0.02  0.07  0.04
```

Trying different AR models to find the best model

```
# Looking at ARIMA (1,1,1) model
ar1 <- arima(ts_fire_b, order = c(1,1,1))
ar1
```

```
##
## Call:
## arima(x = ts_fire_b, order = c(1, 1, 1))
##
## Coefficients:
##          ar1      ma1
##       0.2231  -1.0000
```

```
## s.e.  0.1128   0.0378
##
## sigma^2 estimated as 9.518:  log likelihood = -200.51,  aic = 407.03
```

```
# AIC of 262.73
###################################
# ARIMA (2,0,0)
ar2 <- arima(ts_fire_b, order = c(2,0,0))
ar2
```

```
##
## Call:
## arima(x = ts_fire_b, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       0.2350  -0.1365     6.5549
## s.e.  0.1123   0.1117     0.3792
##
## sigma^2 estimated as 9.218:  log likelihood = -199.87,  aic = 407.74
```

```
###################################
# In viewing the data it appears that the data is best with differncing
ar3 <- arima(ts_fire_b, order = c(3,0,0))
ar3
```

```
##
## Call:
## arima(x = ts_fire_b, order = c(3, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3  intercept
##       0.2153  -0.1030  -0.1446     6.5708
## s.e.  0.1121   0.1134   0.1125     0.3293
##
## sigma^2 estimated as 9.022:  log likelihood = -199.05,  aic = 408.11
```

```
###################################
ar4 <- arima(ts_fire_b, order =c(4,0,0))
ar4
```

```
##
## Call:
## arima(x = ts_fire_b, order = c(4, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3      ar4  intercept
##       0.1752  -0.1322  -0.0793  -0.2588     6.5735
## s.e.  0.1097   0.1110   0.1123   0.1125     0.2558
##
## sigma^2 estimated as 8.432:  log likelihood = -196.51,  aic = 405.03
```

```
####################################
ar5 <- arima(ts_fire_b, order = c(5,0,0))
ar5

##
## Call:
## arima(x = ts_fire_b, order = c(5, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3      ar4     ar5  intercept
##       0.1876  -0.1275  -0.0715  -0.2706  0.0498     6.5720
## s.e.  0.1132   0.1114   0.1136   0.1156  0.1162     0.2681
##
## sigma^2 estimated as 8.41:  log likelihood = -196.42,  aic = 406.85

print(ar1);ar2;ar3;ar4;ar5

##
## Call:
## arima(x = ts_fire_b, order = c(1, 1, 1))
##
## Coefficients:
##          ar1      ma1
##       0.2231  -1.0000
## s.e.  0.1128   0.0378
##
## sigma^2 estimated as 9.518:  log likelihood = -200.51,  aic = 407.03

##
## Call:
## arima(x = ts_fire_b, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       0.2350  -0.1365     6.5549
## s.e.  0.1123   0.1117     0.3792
##
## sigma^2 estimated as 9.218:  log likelihood = -199.87,  aic = 407.74

##
## Call:
## arima(x = ts_fire_b, order = c(3, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3  intercept
##       0.2153  -0.1030  -0.1446     6.5708
## s.e.  0.1121   0.1134   0.1125     0.3293
##
## sigma^2 estimated as 9.022:  log likelihood = -199.05,  aic = 408.11
```

```
##
## Call:
## arima(x = ts_fire_b, order = c(4, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3      ar4  intercept
##       0.1752  -0.1322  -0.0793  -0.2588     6.5735
## s.e.  0.1097   0.1110   0.1123   0.1125     0.2558
##
## sigma^2 estimated as 8.432:  log likelihood = -196.51,  aic = 405.03

##
## Call:
## arima(x = ts_fire_b, order = c(5, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3      ar4     ar5  intercept
##       0.1876  -0.1275  -0.0715  -0.2706  0.0498     6.5720
## s.e.  0.1132   0.1114   0.1136   0.1156  0.1162     0.2681
##
## sigma^2 estimated as 8.41:  log likelihood = -196.42,  aic = 406.85

####################################
# It is determined that an ARIMA model of (4,0,0) is the best model for
forecasting
```

Find the best arima model using a fit arima

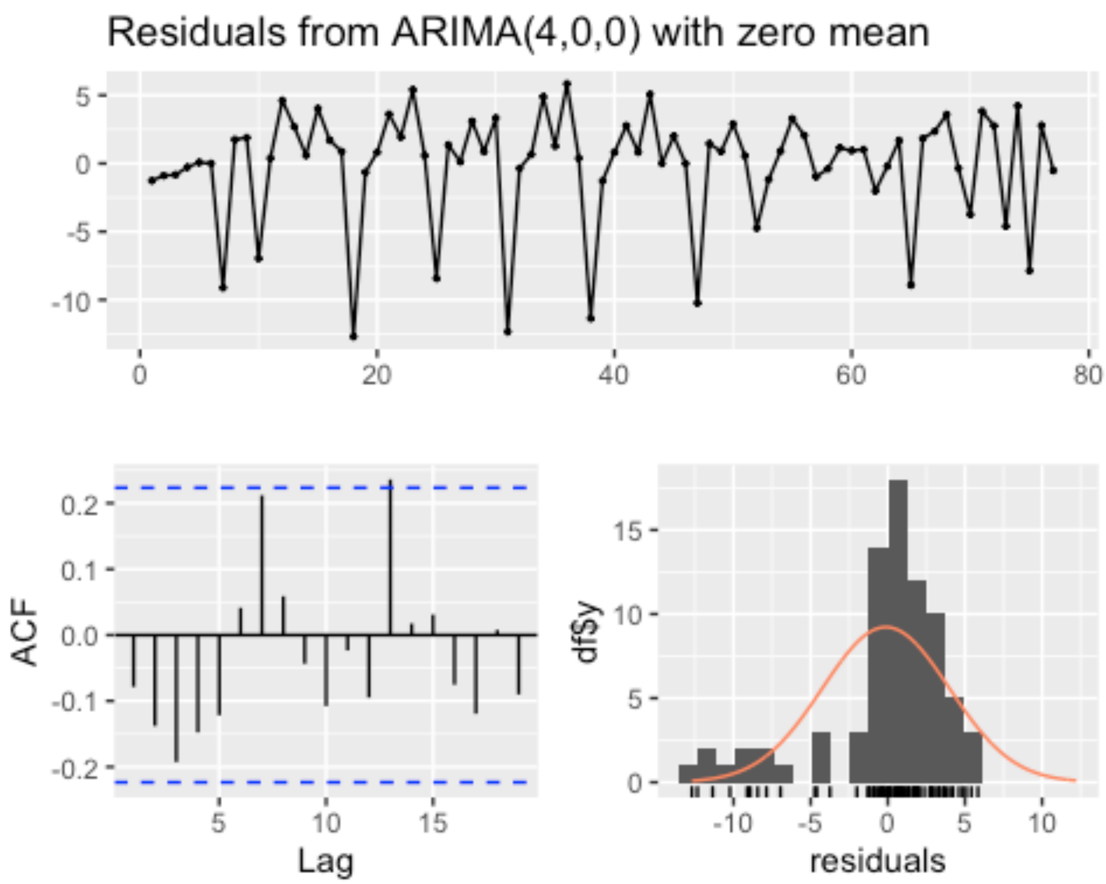
```
# Determining best model
fit_ar <- auto.arima(diff_burns, stepwise = FALSE, approximation = FALSE,
trace = TRUE)

##
##  ARIMA(0,0,0) with zero mean     : 506.6855
##  ARIMA(0,0,0) with non-zero mean : 508.793
##  ARIMA(0,0,1) with zero mean     : Inf
##  ARIMA(0,0,1) with non-zero mean : Inf
##  ARIMA(0,0,2) with zero mean     : Inf
##  ARIMA(0,0,2) with non-zero mean : Inf
##  ARIMA(0,0,3) with zero mean     : Inf
##  ARIMA(0,0,3) with non-zero mean : Inf
##  ARIMA(0,0,4) with zero mean     : Inf
##  ARIMA(0,0,4) with non-zero mean : Inf
##  ARIMA(0,0,5) with zero mean     : Inf
##  ARIMA(0,0,5) with non-zero mean : Inf
##  ARIMA(1,0,0) with zero mean     : 478.2234
##  ARIMA(1,0,0) with non-zero mean : 480.3899
##  ARIMA(1,0,1) with zero mean     : Inf
##  ARIMA(1,0,1) with non-zero mean : Inf
##  ARIMA(1,0,2) with zero mean     : Inf
```

```
##  ARIMA(1,0,2) with non-zero mean : Inf
##  ARIMA(1,0,3) with zero mean     : Inf
##  ARIMA(1,0,3) with non-zero mean : Inf
##  ARIMA(1,0,4) with zero mean     : Inf
##  ARIMA(1,0,4) with non-zero mean : Inf
##  ARIMA(2,0,0) with zero mean     : 463.2976
##  ARIMA(2,0,0) with non-zero mean : 465.516
##  ARIMA(2,0,1) with zero mean     : Inf
##  ARIMA(2,0,1) with non-zero mean : Inf
##  ARIMA(2,0,2) with zero mean     : Inf
##  ARIMA(2,0,2) with non-zero mean : Inf
##  ARIMA(2,0,3) with zero mean     : Inf
##  ARIMA(2,0,3) with non-zero mean : Inf
##  ARIMA(3,0,0) with zero mean     : 462.9781
##  ARIMA(3,0,0) with non-zero mean : 465.2571
##  ARIMA(3,0,1) with zero mean     : Inf
##  ARIMA(3,0,1) with non-zero mean : Inf
##  ARIMA(3,0,2) with zero mean     : Inf
##  ARIMA(3,0,2) with non-zero mean : Inf
##  ARIMA(4,0,0) with zero mean     : 448.2129
##  ARIMA(4,0,0) with non-zero mean : 450.521
##  ARIMA(4,0,1) with zero mean     : Inf
##  ARIMA(4,0,1) with non-zero mean : Inf
##  ARIMA(5,0,0) with zero mean     : 448.8007
##  ARIMA(5,0,0) with non-zero mean : 451.166
##
##
##
##  Best model: ARIMA(4,0,0) with zero mean
```

```
###################################
# Confirming that an ARIMA (4,0,0) is the best fit model
```

#check residuals for the best model

```
#check residuals using forecast and a lag
forecast::checkresiduals(fit_ar, lag=12)
```

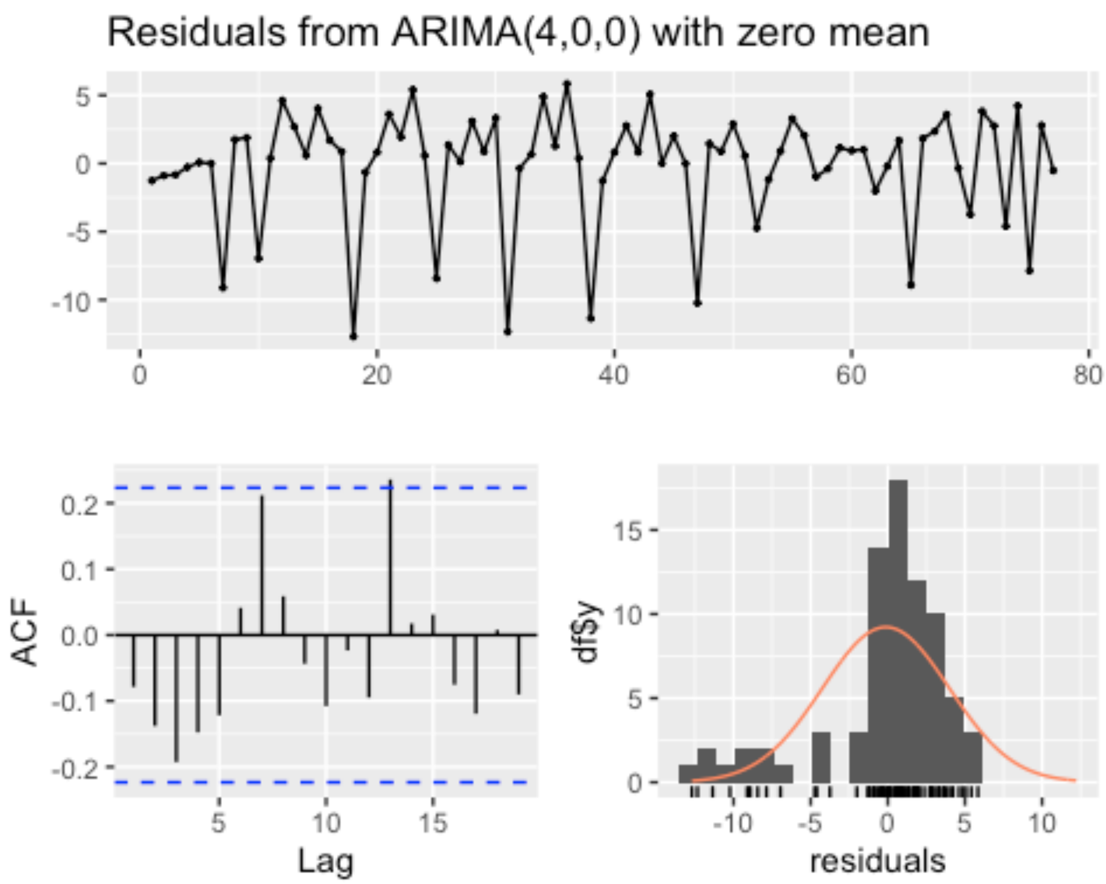



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(4,0,0) with zero mean
## Q* = 14.637, df = 8, p-value = 0.06659
##
## Model df: 4.   Total lags used: 12
```

```
# the best model based on forecasting and checking residuals
arima(burns_fire$month_response, order = c(4,0,0))
```

```
##
## Call:
## arima(x = burns_fire$month_response, order = c(4, 0, 0))
##
## Coefficients:
##          ar1      ar2      ar3      ar4  intercept
##       0.1752  -0.1322  -0.0793  -0.2588     6.5735
## s.e.  0.1097   0.1110   0.1123   0.1125     0.2558
##
## sigma^2 estimated as 8.432:  log likelihood = -196.51,  aic = 405.03
```

```
# print summary statistics
checkresiduals(fit_ar)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(4,0,0) with zero mean
## Q* = 13.746, df = 6, p-value = 0.03261
##
## Model df: 4.   Total lags used: 10
```

```
print(summary(fit_ar))
```

```
## Series: diff_burns
## ARIMA(4,0,0) with zero mean
##
## Coefficients:
##           ar1      ar2      ar3      ar4
##       -0.9945  -0.8816  -0.5776  -0.4603
## s.e.   0.1009   0.1372   0.1354   0.1035
##
## sigma^2 estimated as 17.64:  log likelihood=-218.68
## AIC=447.37   AICc=448.21   BIC=459.09
##
## Training set error measures:
```

```
##                          ME     RMSE      MAE MPE MAPE      MASE        ACF1
## Training set -0.1268859 4.089907 2.786639 NaN  Inf 0.3383141 -0.07872935

#Get the standard deviation
std <- sqrt(17.64)
std

## [1] 4.2
```

In viewing the acf most of all the autocorrelation is removed from the model.

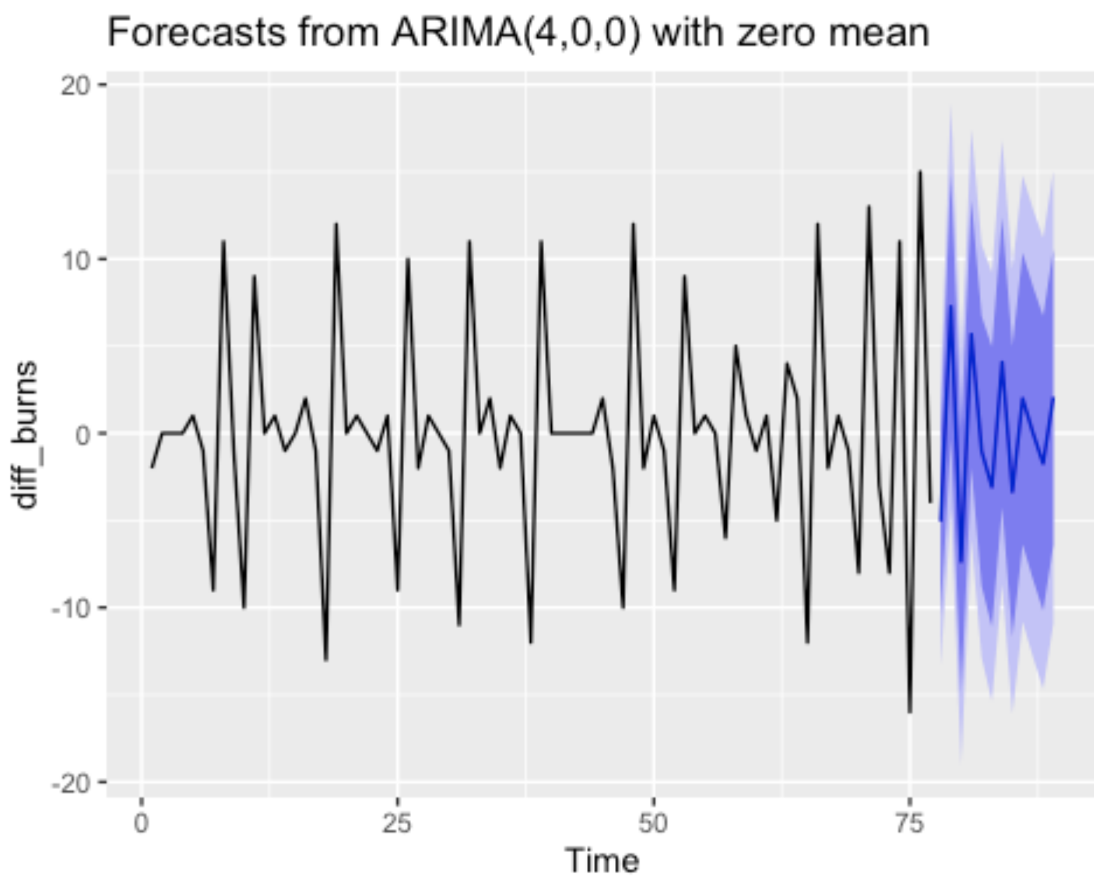
Dickey fuller Test to check for stationarity

```
adf.test(diff_burns)

## Warning in adf.test(diff_burns): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  diff_burns
## Dickey-Fuller = -7.0975, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

#The stationarity is removed according to the dickey fuller test
```

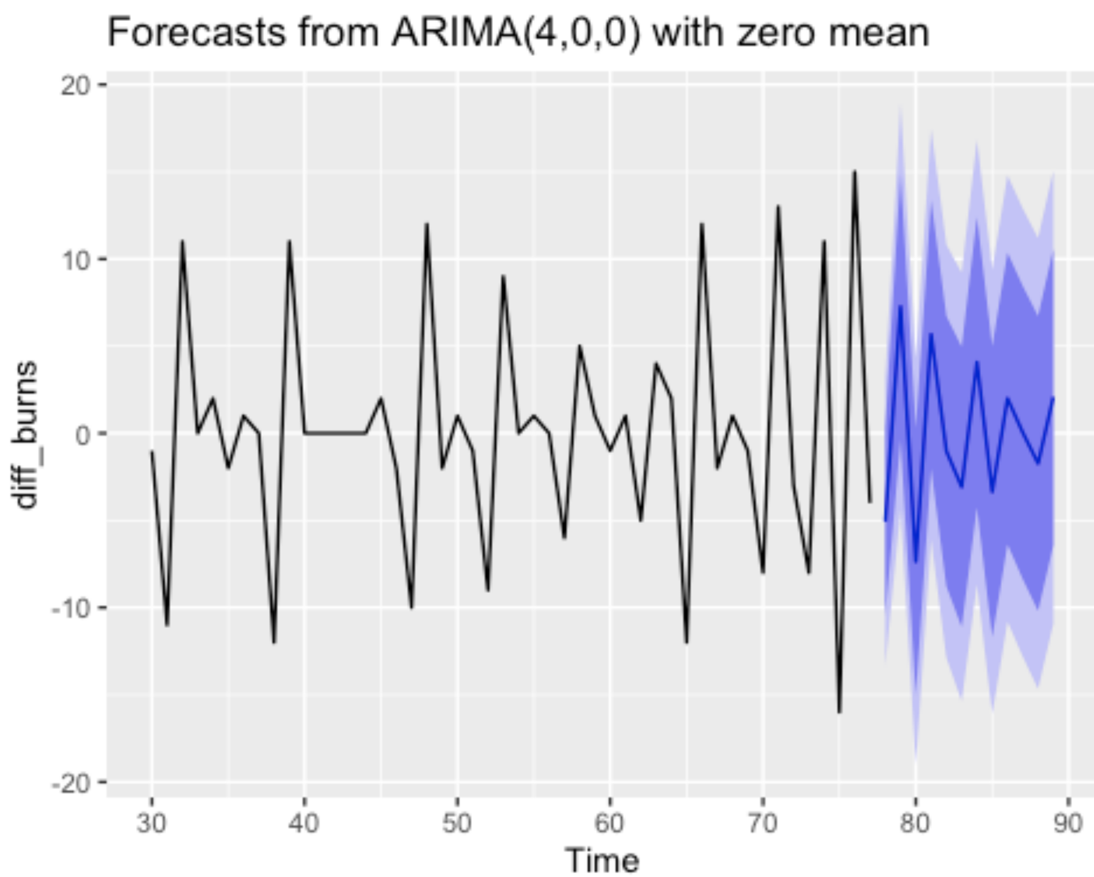
We would reject the null, shows that we have stationarity.

Using arima to forecast monthly trends of fire incidents in San Diego

```
# Forecasting 12 months into the future
f_cast <- forecast(fit_ar, h=12)
####################################
autoplot(f_cast)
```




```
# forecasting to look at most recent data
autoplot(f_cast, include = 48)
```




```
head(print(summary(f_cast)))

##
## Forecast method: ARIMA(4,0,0) with zero mean
##
## Model Information:
## Series: diff_burns
## ARIMA(4,0,0) with zero mean
##
## Coefficients:
##           ar1      ar2      ar3      ar4
##       -0.9945  -0.8816  -0.5776  -0.4603
## s.e.   0.1009   0.1372   0.1354   0.1035
##
## sigma^2 estimated as 17.64:  log likelihood=-218.68
## AIC=447.37   AICc=448.21   BIC=459.09
##
## Error measures:
##                       ME     RMSE      MAE MPE MAPE      MASE        ACF1
## Training set -0.1268859 4.089907 2.786639 NaN  Inf 0.3383141 -0.07872935
##
## Forecasts:
##    Point Forecast       Lo 80      Hi 80      Lo 95     Hi 95
```

```
## 78    -5.06772694 -10.4508388  0.3153849 -13.300486  3.165032
## 79     7.26692522  -0.3251188 14.8589693  -4.344104 18.877955
## 80    -7.35322034 -14.9672621  0.2608214 -18.997892  4.291452
## 81     5.67460117  -2.0095082 13.3587106  -6.077230 17.426432
## 82    -1.02554519  -8.7651319  6.7140415 -12.862222 10.811131
## 83    -3.08055512 -11.1098465  4.9487363 -15.360297  9.199186
## 84     4.07472804  -4.2437698 12.3932259  -8.647317 16.796773
## 85    -3.35611585 -11.6813903  4.9691586 -16.088524  9.376293
## 86     1.99676959  -6.3644246 10.3579638 -10.790573 14.784113
## 87     0.03735123  -8.3614488  8.4361512 -12.807505 12.882207
## 88    -1.73458611 -10.1829581  6.7137859 -14.655256 11.186084
## 89     2.08358759  -6.4035933 10.5707685 -10.896435 15.063611

##    Point Forecast       Lo 80      Hi 80      Lo 95     Hi 95
## 78      -5.067727 -10.4508388  0.3153849 -13.300486  3.165032
## 79       7.266925  -0.3251188 14.8589693  -4.344104 18.877955
## 80      -7.353220 -14.9672621  0.2608214 -18.997892  4.291452
## 81       5.674601  -2.0095082 13.3587106  -6.077230 17.426432
## 82      -1.025545  -8.7651319  6.7140415 -12.862222 10.811131
## 83      -3.080555 -11.1098465  4.9487363 -15.360297  9.199186
```