PRACTICAL TIME SERIES

Student: Huong, LE Thu

ADEO2 – EISTI 2018-2019

PROFESSOR: Marietta Manolessou

Date: 19/ 12/ 2018

Abstract

This report aims to focus on applying time series to explore the data, using the mathematical models needed to analyze time series data then start forecasting. There are 4 key characteristic movements of time series: Trend, Seasonal, Cycle and Irregular. Trend is a general long term movement. Seasonal tells the up down movements with respect to the trend. Cycle is a fluctuations recurring to seasonal that have duration of several years. Irregular is all the remaining parts which are could not be describe into either seasonal or Cycle components. This report is designed to dig deeper on analyze and explore the data in 4 methods, which are (1) average to trend method, (2) percentage to trend method, (3) percentage moving average method, (4) ARMA and ARIMA. The reader is able to find application in three datasets "The US Housing 1990 – 1995", "The US Death 1973 - 1978", "The Air Passengers 1946 - 1960". For each dataset, we a complete solution of 4 methods, forecast for 12 coming months, then compare with its actual's data to concluded the best methods for each problem.

Keywords: *average to trend method, percentage to trend method, percentage moving average method, ARMA, ARIMA, seasonal index, deseasonalize.*

ARMA Properties

ARMA(p,q) process is to just bring together an *MA(q)* and an *AR(p)*

Xt = Noise + Autoregressive Part + Moving Average Part

Xt =

Autoregressive with backwards difference:

AR(p) term:

Moving average with backwards difference:

MA(q) term:

Thus, for a ARMA process we write:

ARMA(p,q) expressed as MA( ):

ARMA(p,q) expressed as AR( ):

Best way to understand it by example:

In our example, it quite simple cause we have given theta and phi. However, our problem in reality does not tell us these value. Our goal is looking for the good estimation all these coefficients in ARMA process. We usually rely on software for estimating coefficients in models, but we will have a term to measure the quality of time series model (here is ARMA model).We used AIC which is Akaike Informatin Criterion, its general formula is:

Using AIC:

The AIC tries to help you assess the relative quality of several competing models, just like adjusted R in the linear regression; by giving credit for models which reduce the error sum of squares and at the same time by building in a penalty for models which bring in too many parameters. Once again, since we don't know the best order of our model, we do generate many candidate models. In the end, we are looking to compare the AIC for a variety of candidate models. We prefer a model with a lower AIC.

Table of contents:

THE US HOUSING 1990 - 1995

OBSERVATION DATA

Based on observation data, we plot our first graph of observation data

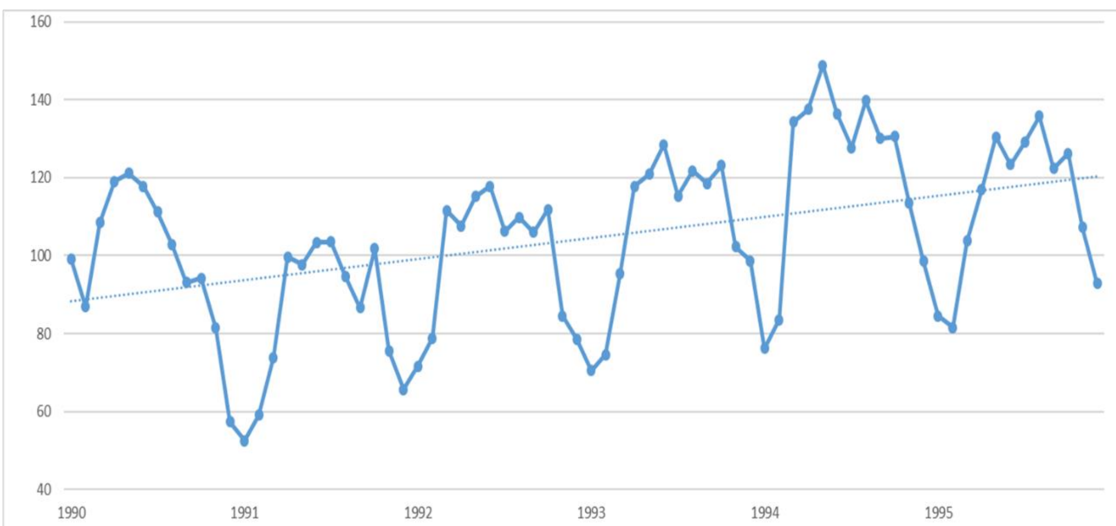
Table 0:

[*Observation data*]

Years	1990	1991	1992	1993	1994	1995
Jan	99.2	52.5	71.6	70.5	76.2	84.5
Feb	86.9	59.1	78.8	74.6	83.5	81.6
Mar	108.5	73.8	111.6	95.5	134.3	103.8
Apr	119	99.7	107.6	117.8	137.6	116.9
May	121.1	97.7	115.2	120.9	148.8	130.5
Jun	117.8	103.4	117.8	128.5	136.4	123.4
July	111.2	103.5	106.2	115.3	127.8	129.1
Aug	102.8	94.7	109.9	121.8	139.8	135.8
Sep	93.1	86.6	106	118.5	130.1	122.4
Oct	94.2	101.8	111.8	123.2	130.6	126.2
Nov	81.4	75.6	84.5	102.3	113.4	107.2
Dec	57.4	65.6	78.6	98.7	98.5	92.8

Graph 0:

[*Initial data plot*]



AVERAGE PERCENTAGE METHOD

Step1: From the initial data, calculate the total sum and yearly averages by corresponding the years. Construct table 1

Table 1

[*yearly total sum and yearly average of the initial data*]

Years	1990	1991	1992	1993	1994	1995
Total	1192.6	1014.0	1199.6	1287.6	1457.0	1354.2
Average	99.4	84.5	100.0	107.3	121.4	112.9

Step 2: Each data is constructed by dividing original monthly data to the yearly average in table 1. Then, we calculate the monthly average by its corresponding years by mean and by median. We also do adjusted mean and adjusted median to obtain precise data. We compute table 2 as below

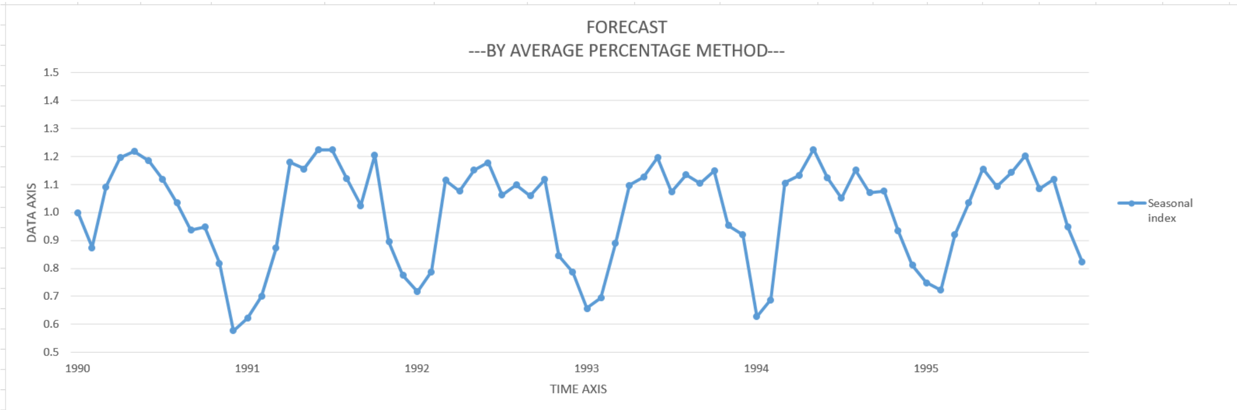
Table 2

[*percentage of original data to yearly average data*]

Years	1990	1991	1992	1993	1994	1995	Mean	adjusted mean	Median	adjusted median
Jan	1.00	0.62	0.72	0.66	0.63	0.75	0.73	0.73	0.69	0.69
Feb	0.87	0.70	0.79	0.70	0.69	0.72	0.74	0.74	0.71	0.71
Mar	1.09	0.87	1.12	0.89	1.11	0.92	1.00	1.00	1.01	1.01
Apr	1.20	1.18	1.08	1.10	1.13	1.04	1.12	1.12	1.12	1.12
May	1.22	1.16	1.15	1.13	1.23	1.16	1.17	1.17	1.16	1.16
Jun	1.19	1.22	1.18	1.20	1.12	1.09	1.17	1.17	1.18	1.18
July	1.12	1.22	1.06	1.07	1.05	1.14	1.11	1.11	1.10	1.10
Aug	1.03	1.12	1.10	1.14	1.15	1.20	1.12	1.12	1.13	1.13
Sep	0.94	1.02	1.06	1.10	1.07	1.08	1.05	1.05	1.07	1.07
Oct	0.95	1.20	1.12	1.15	1.08	1.12	1.10	1.10	1.12	1.12
Nov	0.82	0.89	0.85	0.95	0.93	0.95	0.90	0.90	0.91	0.92
Dec	0.58	0.78	0.79	0.92	0.81	0.82	0.78	0.78	0.80	0.80
							12.00	12.00	11.98	12.00

Graph 1:

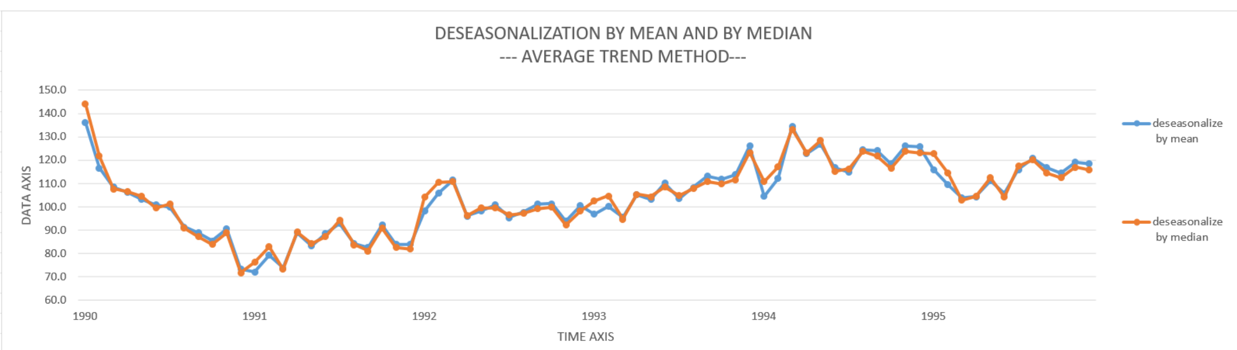
[Seasonal index]



Step 3: De-seasonalization and visualize it. We divide the monthly original data to the seasonal index by mean and by median as calculated in table 2 to compute the deseasonalize data by mean and by median.

Graph 2:

[*Deseasonal by mean and by median*]



Step 4: Apply linear regression to the deseasonal index by mean and by median to get a linear function: (x is the index 1-60, y is the deseasonal data by mean and by median)

By mean: $y = 0.425*X + 88.7176$

By median: $y = 0.4049*X + 89.70971$

The, we apply the function to get 72 new predicted data by each function.

<u>Table 3:</u>

[New data predicted by linear regression mean and median]

Table 1.3: new predicted data by mean and by median

	By mean							By median						
Years	1990	1991	1992	1993	1994	1995	1996	1990	1991	1992	1993	1994	1995	1996
Jan	89.1	94.2	99.4	104.5	109.6	114.7	119.8	90.1	95.0	99.8	104.7	109.6	114.4	119.3
Feb	89.6	94.7	99.8	104.9	110.0	115.1	120.2	90.5	95.4	100.2	105.1	110.0	114.8	119.7
Mar	90.0	95.1	100.2	105.3	110.4	115.5	120.6	90.9	95.8	100.6	105.5	110.4	115.2	120.1
Apr	90.4	95.5	100.6	105.7	110.8	115.9	121.1	91.3	96.2	101.0	105.9	110.8	115.6	120.5
May	90.8	95.9	101.1	106.2	111.3	116.4	121.5	91.7	96.6	101.5	106.3	111.2	116.0	120.9
Jun	91.3	96.4	101.5	106.6	111.7	116.8	121.9	92.1	97.0	101.9	106.7	111.6	116.4	121.3
July	91.7	96.8	101.9	107.0	112.1	117.2	122.3	92.5	97.4	102.3	107.1	112.0	116.8	121.7
Aug	92.1	97.2	102.3	107.4	112.5	117.6	122.8	92.9	97.8	102.7	107.5	112.4	117.2	122.1
Sep	92.5	97.7	102.8	107.9	113.0	118.1	123.2	93.4	98.2	103.1	107.9	112.8	117.6	122.5
Oct	93.0	98.1	103.2	108.3	113.4	118.5	123.6	93.8	98.6	103.5	108.3	113.2	118.1	122.9
Nov	93.4	98.5	103.6	108.7	113.8	118.9	124.0	94.2	99.0	103.9	108.7	113.6	118.5	123.3
Dec	93.8	98.9	104.0	109.1	114.2	119.3	124.5	94.6	99.4	104.3	109.1	114.0	118.9	123.7

Step 5: Forecasting data Y~ = T~*S~

T~ comes from table 1.2 and S~ comes from adjusted mean and adjusted median in table 1.2

We have table 1.4. We could do forecasting to the whole data and visualized it to evaluate our forecasting

<u>Table 4:</u>

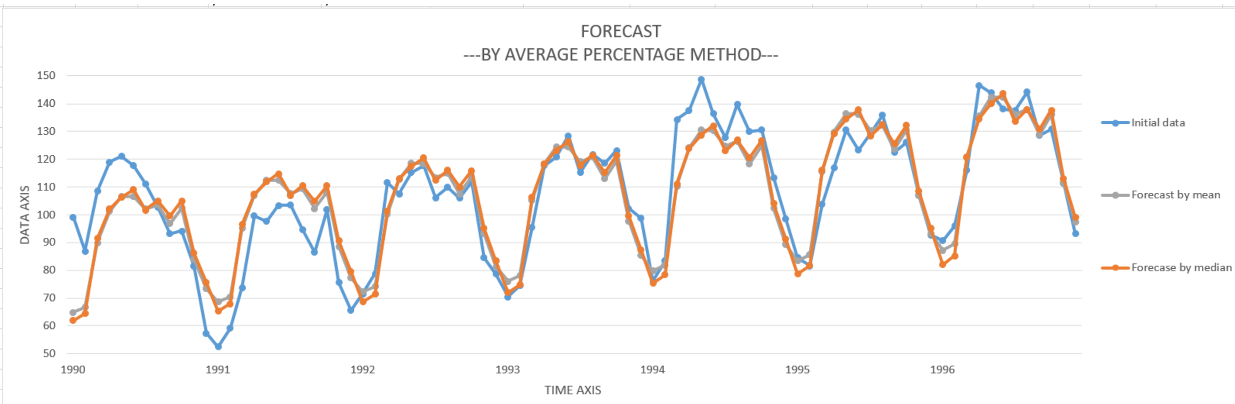
[*Forecasting data Y~ = T~*S~]*

Table 1.4: forcasting data by time series

	By mean							By median						
Years	1990	1991	1992	1993	1994	1995	1996	1990	1991	1992	1993	1994	1995	1996
Jan	64.9	68.6	72.3	76.1	79.8	83.5	87.2	62.0	65.3	68.7	72.0	75.4	78.7	82.0
Feb	66.7	70.5	74.3	78.1	81.9	85.7	89.5	64.5	68.0	71.4	74.9	78.3	81.8	85.3
Mar	90.0	95.1	100.2	105.3	110.4	115.5	120.6	91.6	96.5	101.4	106.3	111.2	116.1	121.0
Apr	101.3	107.0	112.7	118.4	124.2	129.9	135.6	102.1	107.5	112.9	118.3	123.8	129.2	134.6
May	106.5	112.5	118.5	124.5	130.5	136.5	142.4	106.3	111.9	117.5	123.1	128.8	134.4	140.0
Jun	106.5	112.5	118.4	124.4	130.3	136.3	142.3	109.1	114.8	120.6	126.3	132.1	137.8	143.6
July	102.0	107.7	113.4	119.1	124.8	130.5	136.1	101.7	107.0	112.3	117.7	123.0	128.4	133.7
Aug	103.5	109.3	115.0	120.8	126.5	132.2	138.0	105.0	110.5	116.0	121.5	127.0	132.5	138.0
Sep	96.9	102.2	107.6	112.9	118.3	123.6	129.0	99.7	104.9	110.1	115.2	120.4	125.6	130.8
Oct	102.5	108.1	113.7	119.4	125.0	130.6	136.2	105.0	110.5	115.9	121.4	126.8	132.3	137.7
Nov	84.0	88.6	93.2	97.8	102.4	107.0	111.5	86.2	90.7	95.1	99.6	104.0	108.5	112.9
Dec	73.4	77.4	81.4	85.4	89.4	93.4	97.4	75.7	79.6	83.4	87.3	91.2	95.1	99.0

Graph 3:

[*Forecasting graph*]



Conclusion 1:

Based on the results of average to percentage method, we will consider the forecasting data by mean because it is closer and give a better results compare to the actual data of housing in US for the year 1996. Please refer to the conclusion part of this report to observe the actual data of 1996.

PERCENTAGE TO TREND METHOD

Step1: From the initial data, we apply the monthly averages by corresponding the years, by the corresponding median

Table 5

[*Monthly averages of observation data*]

Table 2.1						
Time	6.5	18.5	30.5	42.5	54.5	66.5
Monthly average	99.4	84.5	100.0	107.3	121.4	112.9
Year	1990	1991	1992	1993	1994	1995
Monthly average	99.4	84.5	100.0	107.3	121.4	112.9

Step 2: Apply the linear regression on the data of the table 5

Table 6

[*Table of new data Y, calculated by linear regression equation* $y = 0.44*X + 88.1$]

Table 2.2 : new data by linear regression						
Years	1990	1991	1992	1993	1994	1995
Jan	88.6	93.9	99.2	104.5	109.8	115.1
Feb	89.0	94.3	99.6	104.9	110.2	115.5
Mar	89.4	94.7	100.0	105.3	110.6	115.9
Apr	89.9	95.2	100.5	105.8	111.1	116.4
May	90.3	95.6	100.9	106.2	111.5	116.8
Jun	90.8	96.1	101.4	106.7	112.0	117.3
July	91.2	96.5	101.8	107.1	112.4	117.7
Aug	91.7	97.0	102.2	107.5	112.8	118.1
Sep	92.1	97.4	102.7	108.0	113.3	118.6
Oct	92.5	97.8	103.1	108.4	113.7	119.0
Nov	93.0	98.3	103.6	108.9	114.2	119.5
Dec	93.4	98.7	104.0	109.3	114.6	119.9

Step 3: Divide the given monthly values of the initial data by the corresponding trend values of table 6. We express results as percentages and create a new table. Then, we calculate the monthly average by its corresponding years by mean and by median. We also do adjusted mean and adjusted median to obtain precise data.

Table 7
[*Table of percentage results, in order to calculate seasonal index*]

Table 2.3 : Percentage results							Mean	Adjusted mean	Median	Adjusted Median
Years	1990	1991	1992	1993	1994	1995				
Jan	112.0	55.9	72.2	67.5	69.4	73.4	75.09	75.09	70.82	71.78
Feb	97.6	62.7	79.1	71.1	75.8	70.7	76.16	76.16	73.45	74.45
Mar	121.3	77.9	111.6	90.7	121.4	89.5	102.05	102.05	101.11	102.49
Apr	132.4	104.7	107.1	111.4	123.9	100.4	113.32	113.32	109.22	110.71
May	134.1	102.2	114.1	113.8	133.4	111.7	118.22	118.22	113.98	115.54
Jun	129.8	107.6	116.2	120.5	121.8	105.2	116.86	116.86	118.34	119.96
July	121.9	107.2	104.3	107.7	113.7	109.7	110.75	110.75	108.67	110.15
Aug	112.2	97.7	107.5	113.3	123.9	114.9	111.57	111.57	112.71	114.24
Sep	101.1	88.9	103.2	109.7	114.8	103.2	103.50	103.50	103.22	104.63
Oct	101.8	104.1	108.4	113.6	114.8	106.0	108.12	108.12	107.22	108.68
Nov	87.5	76.9	81.6	94.0	99.3	89.7	88.18	88.18	88.64	89.85
Dec	61.4	66.5	75.6	90.3	85.9	77.4	76.18	76.18	76.48	77.52
							1200.0	1200.0	1183.8	1200.0

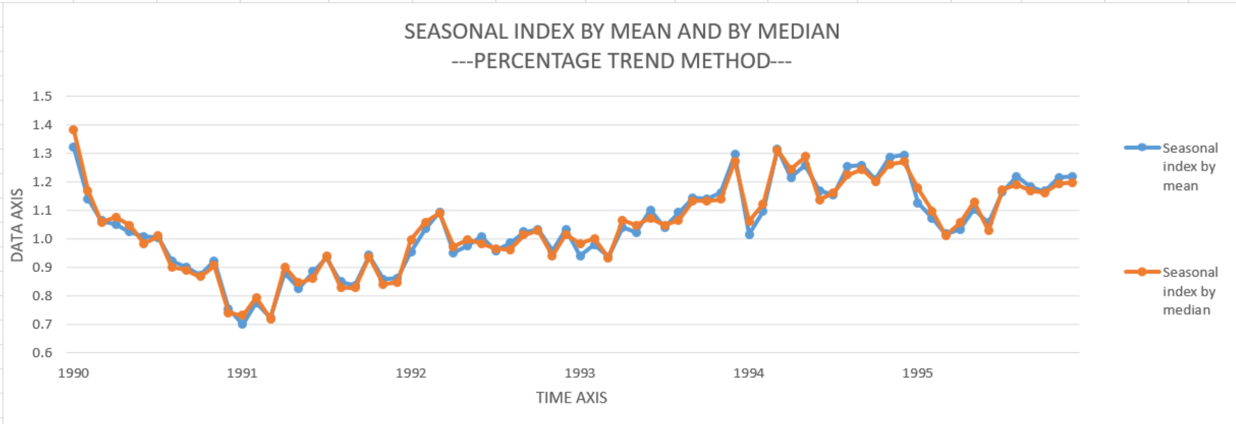
Step 4: Divide the given monthly values of the initial data by the adjusted mean and adjusted median of table 7. We create a new table.

Table 8
[*Calculating seasonal index by mean and by median*]

Table 2.4: Seasonal index by mean and by median											
	By mean						By median				
Years	1990	1991	1992	1993	1994	1995	1990	1991	1992	1993	1994
Jan	1.3	0.7	1.0	0.9	1.0	1.1	1.4	0.7	1.0	1.0	1.1
Feb	1.2	1.2	1.0	1.0	1.1	1.1	1.2	0.8	1.1	1.0	1.1
Mar	0.9	0.9	1.1	0.9	1.3	1.0	1.1	0.7	1.1	0.9	1.3
Apr	0.8	0.8	0.9	1.0	1.2	1.0	1.1	0.9	1.0	1.1	1.2
May	0.8	0.8	1.0	1.0	1.3	1.1	1.0	0.8	1.0	1.0	1.3
Jun	0.8	0.8	1.0	1.1	1.2	1.1	1.0	0.9	1.0	1.1	1.1
July	0.8	0.9	1.0	1.0	1.2	1.2	1.0	0.9	1.0	1.0	1.2
Aug	0.8	0.9	1.0	1.1	1.3	1.2	0.9	0.8	1.0	1.1	1.2
Sep	0.9	0.9	1.0	1.1	1.3	1.2	0.9	0.8	1.0	1.1	1.2
Oct	0.9	0.9	1.0	1.1	1.2	1.2	0.9	0.9	1.0	1.1	1.2
Nov	1.1	1.1	1.0	1.2	1.3	1.2	0.9	0.8	0.9	1.1	1.3
Dec	1.2	1.3	1.0	1.3	1.3	1.2	0.7	0.8	1.0	1.3	1.3

<u>Graph 3</u>

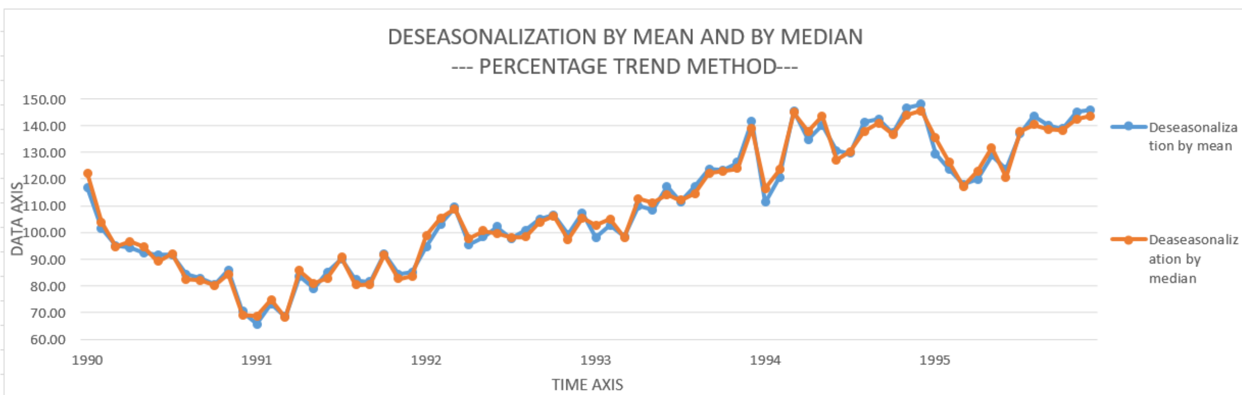
[*Seasonal index*]



Step 5: De-seasonalization and visualize it

<u>Graph 4</u>

[*De-seasonalization*]



Step 6: Forecasting data Y~ = T~*S~

T~ comes from table 6 and S~ comes from adjusted mean and adjusted median.

We have table 9. We could do forecasting to the whole data and visualized it to evaluate our forecasting

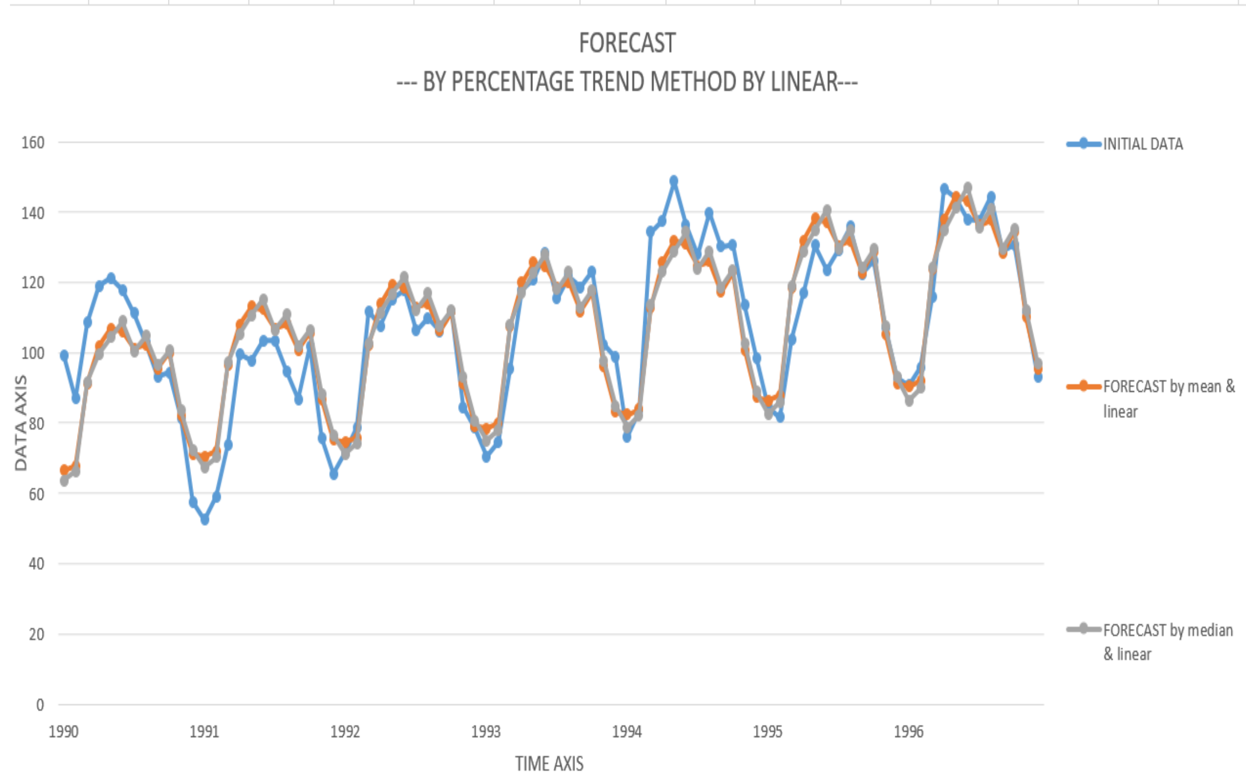
Table 9

[*Forecasting data by percentage to trend method, apply linear regression, seasonal index by mean and by median*]

Table 2.5 : Forecasting data		
1996	By mean	By meadian
Jan	90.37	86.39
Feb	92.00	89.93
Mar	123.72	124.25
Apr	137.88	134.71
May	144.37	141.09
Jun	143.22	147.02
July	136.22	135.48
Aug	137.72	141.02
Sep	128.22	129.62
Oct	134.42	135.11
Nov	110.02	112.10
Dec	95.38	97.06

Graph 5

[Fore*casting by percentage to trend method by mean and by median*]



BONUS: Quadratic and Linear, which one is the best? The answer is we have to try many methods to our observation data to see which one gives the best explanation and better forecasting results.

Here, I had done all these above steps to quadratic form. The result equation is: y = 92.95 + 0.029*x + 0.005* x^2. However, the quadratic was fail this time because it has not given a better explanation neither a better forecasting results, compared to linear regression.

Conclusion 2:

Based on the results of percentage to trend method, we will consider the forecasting data by mean with liner regression because it is closer and give a better results compare to the actual data of housing in US for the year 1996. Please refer to the conclusion part of this report to observe the actual data of 1996.

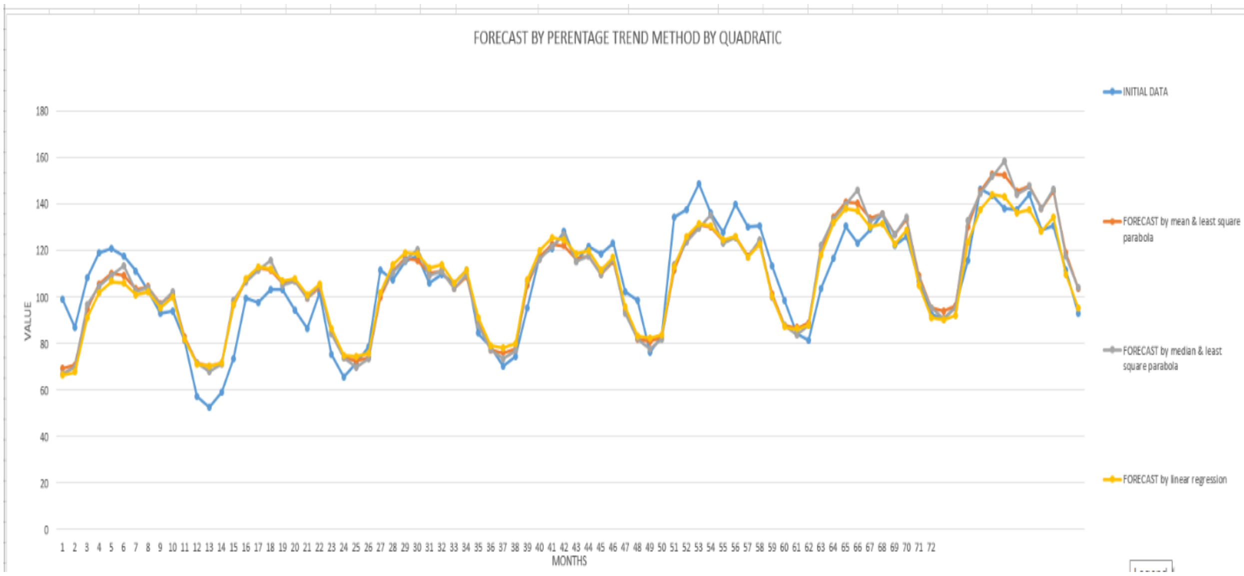
Table 10

[*Forecasting data of the year 1996, by quadratic trend equation*]

Table 2. : Forecasting data by quadratic		
1996	By mean	By meadian
Jan	94.07	90.68
Feb	96.37	95.44
Mar	130.28	132.89
Apr	145.72	144.62
May	153.17	152.00
Jun	152.62	158.49
July	145.60	144.28
Aug	147.77	147.67
Sep	138.12	137.99
Oct	145.36	146.46
Nov	119.26	117.73
Dec	103.82	104.00

Graph 6

[*Compared percentage trend methods by linear regression and by quadratic equation*]


FORECAST BY PERENTAGE TREND METHOD BY QUADRATIC
INITIAL DATA
FORECAST by mean & least square parabola
FORECAST by median & least square parabola
FORECAST by linear regression
VALUE
MONTHS

PERCENTAGE MOVING AVERAGE METHOD

Step1: From the initial data, calculate 12 months moving average. Construct table 11

Table 11
[*12 months moving average*]

Table 3.1:	data of 12 months moving average					
Years	1990	1991	1992	1993	1994	1995
Jan		85.5	94.2	100.4	116.4	115.1
Feb		84.9	94.4	101.2	117.4	115.2
Mar		84.2	95.7	102.2	118.9	114.9
Apr		83.7	97.3	103.2	119.9	114.2
May		84.3	98.1	104.1	120.5	113.8
Jun		83.8	98.9	105.6	121.4	113.3
July	99.4	84.5	100.0	107.3	121.4	112.9
Aug	95.5	86.1	99.9	107.8	122.1	
Sep	93.2	87.7	99.5	108.5	122.0	
Oct	90.3	90.9	98.2	111.8	119.4	
Nov	88.7	91.5	99.0	113.4	117.7	
Dec	86.7	93.0	99.5	115.7	116.2	

Step 2: From table 11, we compute the 12 months centered moving average

Table 12
[12 months centered moving average]

Table 3.1:	data of 12 months centered moving average					
Years	1990	1991	1992	1993	1994	1995
Jan		85.2	94.3	100.8	116.9	115.1
Feb		84.5	95.1	101.7	118.2	115.0
Mar		83.9	96.5	102.7	119.4	114.5
Apr		84.0	97.7	103.7	120.2	114.0
May		84.1	98.5	104.9	121.0	113.6
Jun		84.2	99.4	106.5	121.4	113.1
July	97.4	85.3	99.9	107.5	121.8	
Aug	94.3	86.9	99.7	108.1	122.0	
Sep	91.7	89.3	98.9	110.1	120.7	
Oct	89.5	91.2	98.6	112.6	118.5	
Nov	87.7	92.3	99.3	114.6	116.9	
Dec	86.1	93.6	100.0	116.1	115.6	

Step 3: Compute the ratio

Divide the initial data to the 12 months centered moving average of table 12. We compute the percentage as listed in table 13. Then, we calculate the average by month of the correspond year to find the seasonal index.

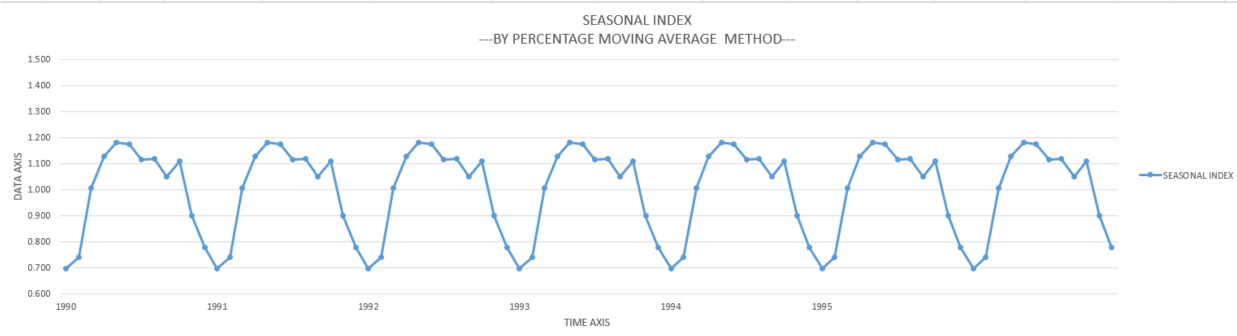
<u>Table 13:</u>

[Ratio of initial data / 12 months centered moving average]

Table 3.3:	Percentage to 12 months centered moving average						
Years	1990	1991	1992	1993	1994	1995	by mean	adjusted mean
Jan	0.0	0.6	0.8	0.7	0.7	0.7	0.7	0.7
Feb	0.0	0.7	0.8	0.7	0.7	0.7	0.7	0.7
Mar	0.0	0.9	1.2	0.9	1.1	0.9	1.0	1.0
Apr	0.0	1.2	1.1	1.1	1.1	1.0	1.1	1.1
May	0.0	1.2	1.2	1.2	1.2	1.1	1.2	1.2
Jun	0.0	1.2	1.2	1.2	1.1	1.1	1.2	1.2
July	1.1	1.2	1.1	1.1	1.0	0.0	1.1	1.1
Aug	1.1	1.1	1.1	1.1	1.1	0.0	1.1	1.1
Sep	1.0	1.0	1.1	1.1	1.1	0.0	1.0	1.1
Oct	1.1	1.1	1.1	1.1	1.1	0.0	1.1	1.1
Nov	0.9	0.8	0.9	0.9	1.0	0.0	0.9	0.9
Dec	0.7	0.7	0.8	0.9	0.9	0.0	0.8	0.8

<u>Graph 7:</u>

[*Seasonal index*]



Step 4: De-seasonalization and visualize it. We divide the monthly original data to the seasonal index by adjusted mean as calculated in table 13 to compute the deseasonalize data by mean

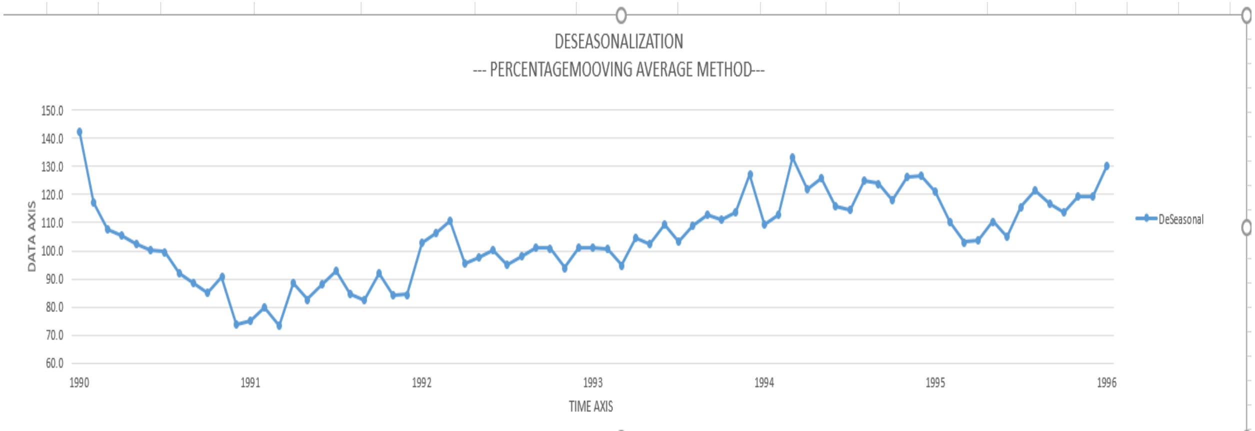
Table 14:

[Deseasional data]

Table 3.4	Deseasonal data					
Years	1990	1991	1992	1993	1994	1995
Jan	142.2	75.3	102.7	101.1	109.3	121.2
Feb	117.2	79.7	106.3	100.7	112.7	110.1
Mar	107.8	73.3	110.8	94.8	133.4	103.1
Apr	105.6	88.4	95.4	104.5	122.1	103.7
May	102.5	82.7	97.5	102.3	125.9	110.4
Jun	100.2	87.9	100.2	109.3	116.0	104.9
July	99.6	92.7	95.1	103.3	114.5	115.7
Aug	91.9	84.6	98.2	108.8	124.9	121.3
Sep	88.7	82.5	100.9	112.8	123.9	116.6
Oct	85.0	91.9	100.9	111.2	117.9	113.9
Nov	90.5	84.1	94.0	113.8	126.1	119.2
Dec	73.9	84.4	101.1	127.0	126.8	119.4

<u>Graph 8:</u>

[*Deseasonal by mean*]



Step 5: Apply linear regression to the deseasonal data and by median to get a linear function: (x is the index 1-72, y is the deseasonal data). We get the function y = 0.42*x+88.9

Then, we apply the function to get 84 new predicted data by the function.

<u>Table 15:</u>

[New data predicted by linear regression]

Table 3.5	New y by liner regression						
Years	1990	1991	1992	1993	1994	1995	1996
Jan	89.4	94.5	99.7	104.8	109.9	115.0	120.2
Feb	89.8	95.0	100.1	105.2	110.3	115.5	120.6
Mar	90.3	95.4	100.5	105.6	110.8	115.9	121.0
Apr	90.7	95.8	100.9	106.1	111.2	116.3	121.5
May	91.1	96.2	101.4	106.5	111.6	116.8	121.9
Jun	91.5	96.7	101.8	106.9	112.1	117.2	122.3
July	92.0	97.1	102.2	107.4	112.5	117.6	122.7
Aug	92.4	97.5	102.7	107.8	112.9	118.0	123.2
Sep	92.8	97.9	103.1	108.2	113.3	118.5	123.6
Oct	93.2	98.4	103.5	108.6	113.8	118.9	124.0
Nov	93.7	98.8	103.9	109.1	114.2	119.3	124.5
Dec	94.1	99.2	104.4	109.5	114.6	119.8	124.9

Step 6: Forecasting data Y~ = T~*S~

T~ comes from table 12 and S~ comes from adjusted mean and adjusted median in table 12

We have table 14. We could do forecasting to the whole data and visualized it to evaluate our forecasting

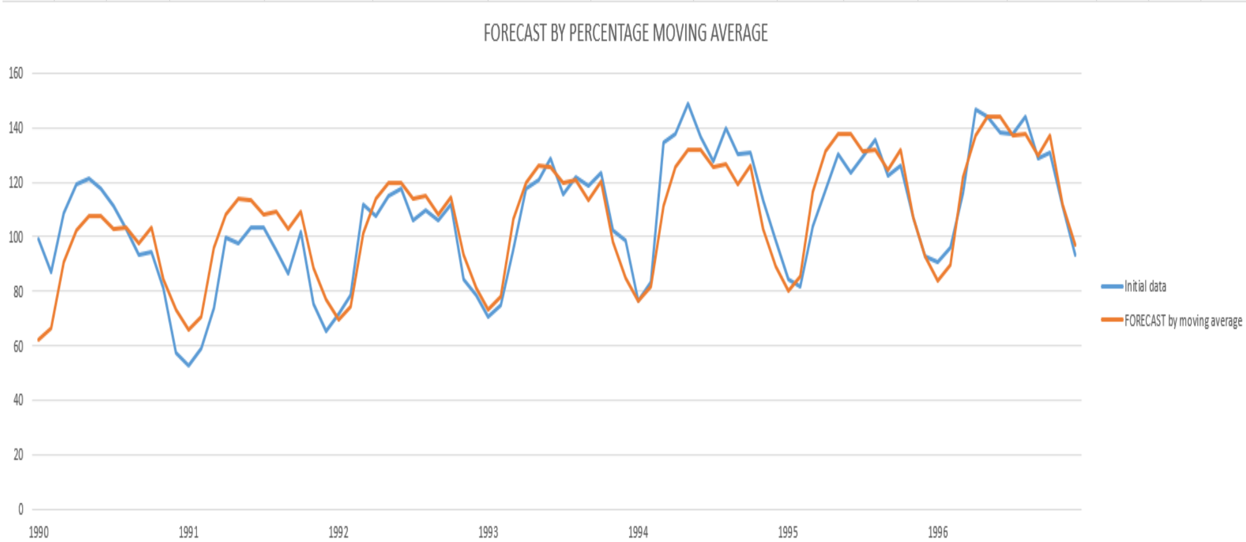
<u>Table 16:</u>

[*Forecasting data Y~ = T~*S~]*

Table 3.6	Forecasting						
Years	1990	1991	1992	1993	1994	1995	1996
Jan	62.3	65.9	69.5	73.1	76.7	80.2	83.8
Feb	66.6	70.4	74.2	78.0	81.8	85.6	89.4
Mar	90.9	96.0	101.2	106.4	111.5	116.7	121.9
Apr	102.2	108.0	113.8	119.6	125.4	131.2	136.9
May	107.6	113.7	119.8	125.8	131.9	138.0	144.0
Jun	107.6	113.7	119.7	125.7	131.8	137.8	143.8
July	102.7	108.4	114.1	119.8	125.6	131.3	137.0
Aug	103.4	109.1	114.9	120.6	126.4	132.1	137.8
Sep	97.5	102.9	108.2	113.6	119.0	124.4	129.8
Oct	103.3	109.0	114.7	120.4	126.1	131.7	137.4
Nov	84.2	88.8	93.4	98.1	102.7	107.3	111.9
Dec	73.1	77.1	81.1	85.1	89.1	93.1	97.0

<u>Graph 9:</u>

[*Forecasting graph*]



ARMA

We are using RStudio in the following step:

Step 1: Install the libraries (if you already installed it, just need to call and use)

Step 2: Import the data set

Step 3: Split into training set and test set

Step 4: Converting training set into time series

Step 5: Plot the training set

Step 6: Visualize all components of time series (trend, seasonal index…)

Step 7: Plot ACF and PACF to find good p and q

Step 8: Decompose the trend, seasonal and plot ACF again to visualize the good p and q

Step 9: Run auto.arima()

Step 10: Visualize the arima(p,d,q) process and its predictions

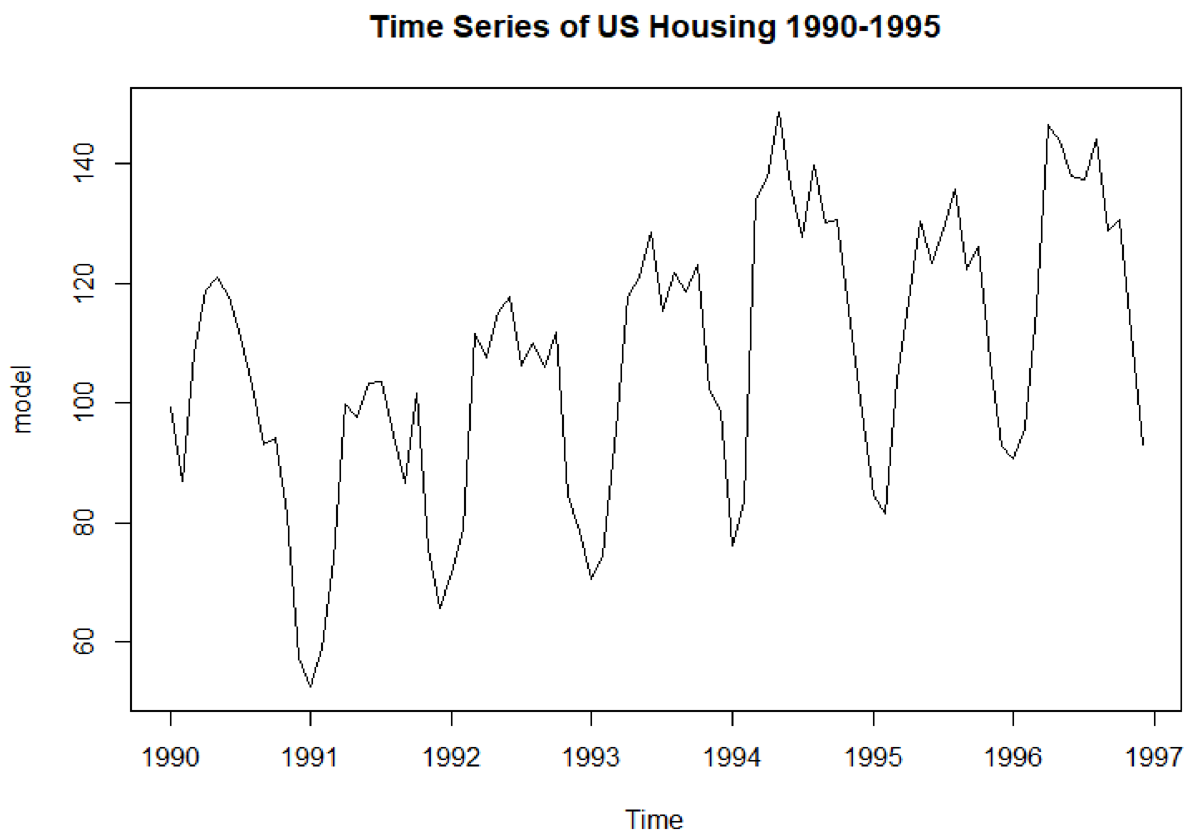
ARIMA() required 3 the three integer components (p, d, q) are the AR order, the degree of differencing, and the MA order. ARMA = ARIMA() with parameter d = 0 or degree of diff = 0

**All the step have written clearly in the script 'myUSHousing.r'.

Step 5: Plot the training set

<u>Graph 10:</u>

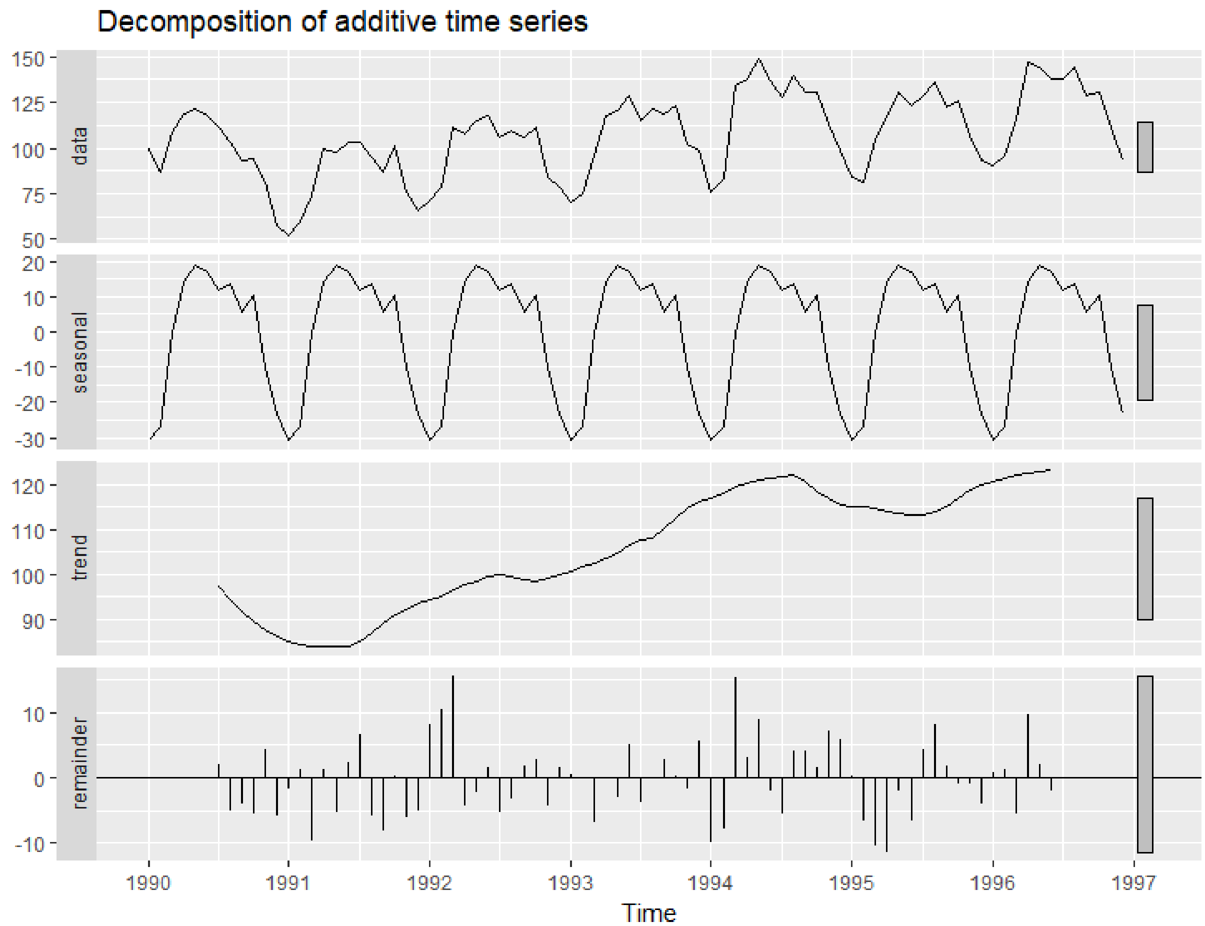
[*Initial data plot in RStudio*]



Step 6: Visualize all components of time series (trend, seasonal index…)

<u>Graph 11:</u>

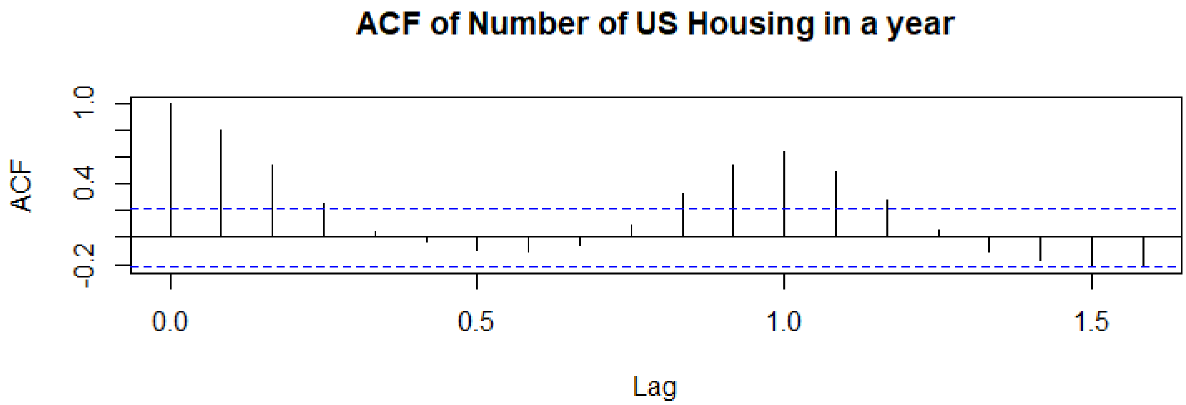
[trend, seasonal index plot in RStudio]

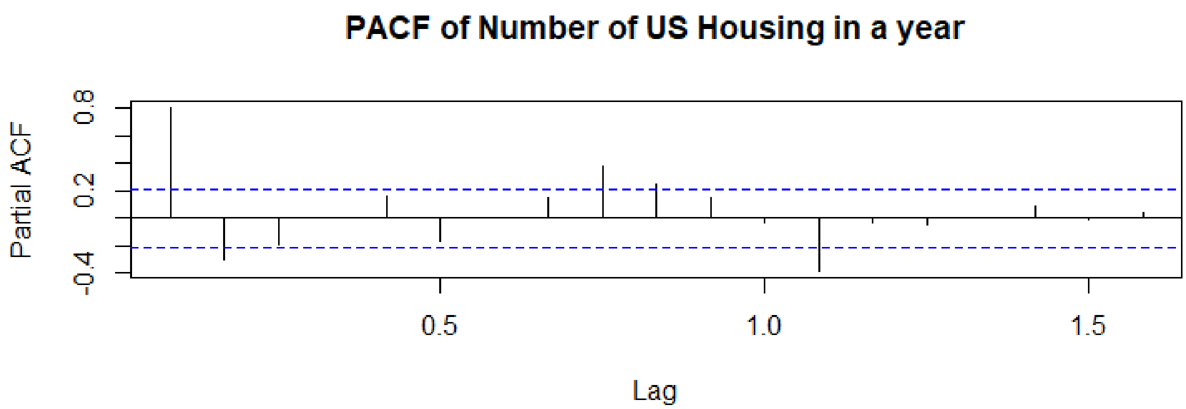


Step 7: Plot ACF and PACF to find good p and q

Graph 12:

[ACF and PACF plot]

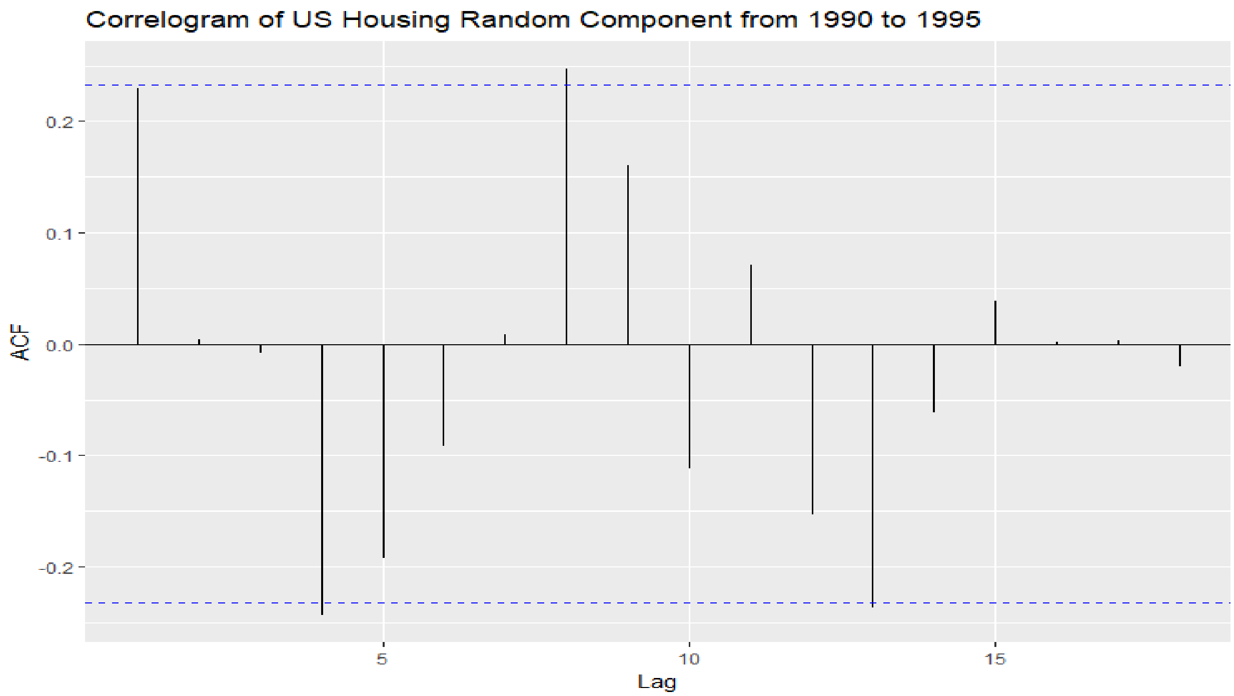




Step 8: Decompose the trend, seasonal and plot ACF again to visualize the good p and q

<u>Graph 13:</u>

[second ACF plot in RStudio]



Step 9: Run auto.arima()

Results obtain 3th order AR(p = 3) and AIC = 527.28

Phi 1 = 0.4636; phi 2 = 0.1956, phi 3 = 0.2482 and residual = -0.5915

```
> arima <- auto.arima(model)
> arima
Series: model
ARIMA(3,0,0)(1,1,0)[12]

Coefficients:
         ar1     ar2     ar3     sar1
      0.4636  0.1956  0.2482  -0.5915
s.e.  0.1213  0.1314  0.1301   0.1063

sigma^2 estimated as 75.06:  log likelihood=-258.64
AIC=527.28   AICc=528.18   BIC=538.66
```

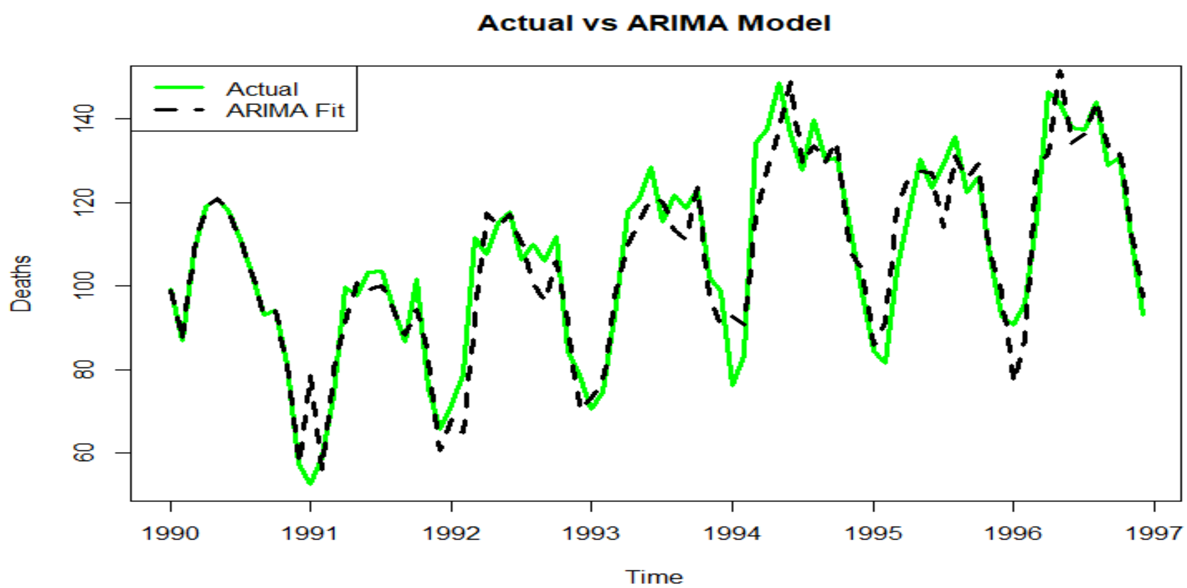
Try different ARMA but AIC are not as good at optimal arima()

ARMA process	ACI
000	765.52
100	680.46
300	671.8
400	673.79

Step 10: Visualize the arima(p,d,q) process and its predictions

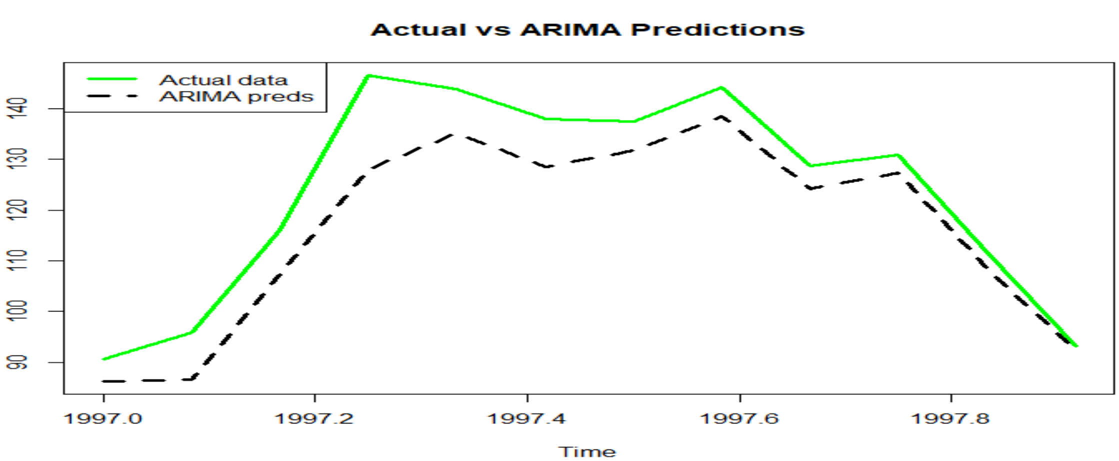
<u>Graph 14:</u>

[Actual data and ARIMA model fitted]



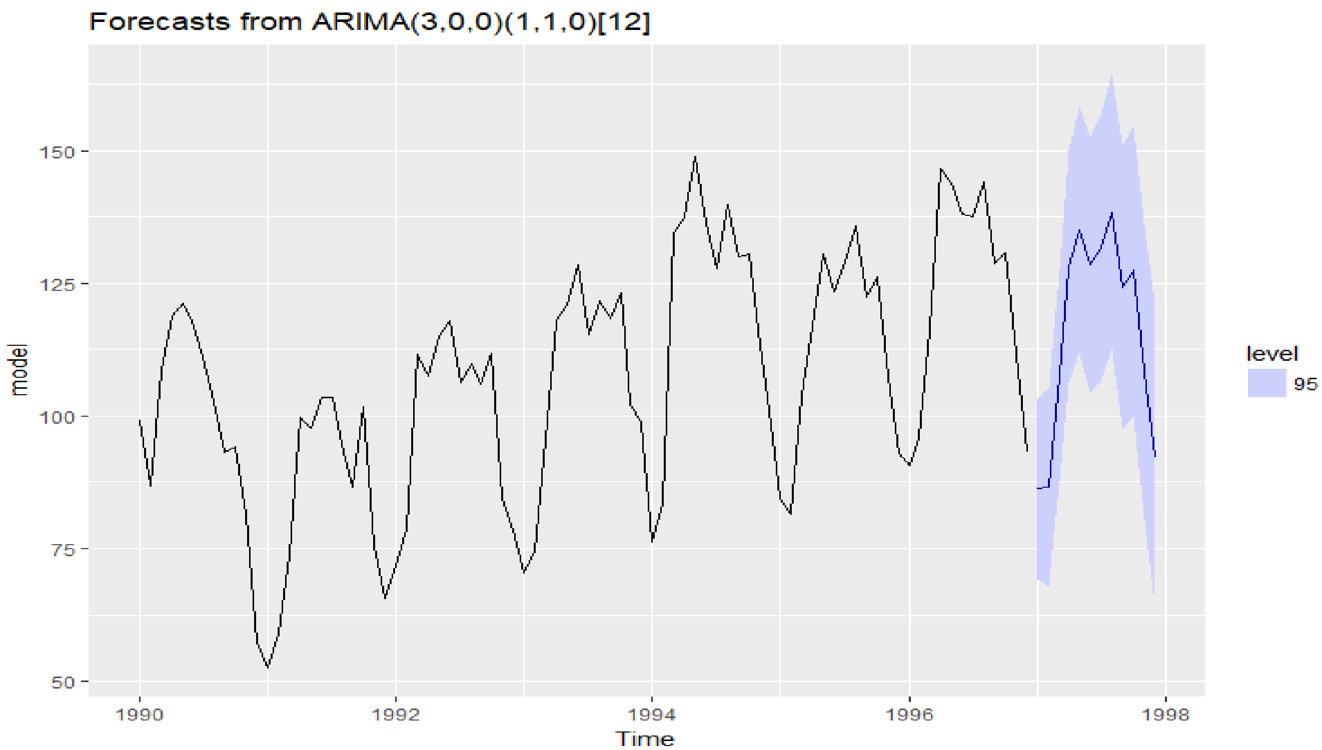
<u>Graph 15:</u>

[Actual data of 1996 and ARIMA predictions]



Graph 16:

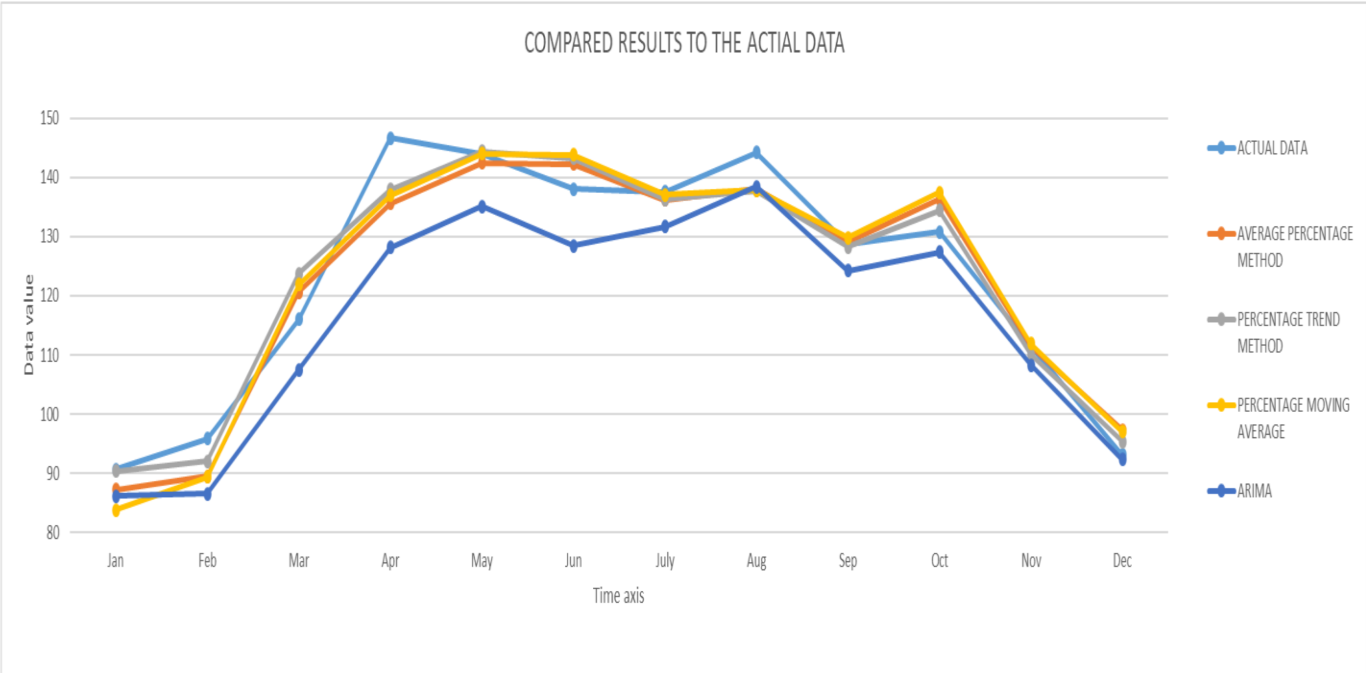
[Actual data and predictions with highest and lowest bounds]



EVALUATED RESULTS

Table 1: Compare obtained results to the actual data of 1960

1996	Actual data	Average percentage methods	Percentage to trend methods	Percentage moving average	ARIMA
Jan	90.7	87.2	90.4	83.8	86.2
Feb	95.9	89.5	92.0	89.4	86.6
Mar	116	120.6	123.7	121.9	107.5
Apr	146.6	135.6	137.9	136.9	128.1
May	143.9	142.4	144.4	144.0	135.1
Jun	138	142.3	143.2	143.8	128.4
July	137.5	136.1	136.2	137.0	131.7
Aug	144.2	138.0	137.7	137.8	138.4
Sep	128.7	129.0	128.2	129.8	124.2
Oct	130.8	136.2	134.4	137.4	127.4
Nov	111.5	111.5	110.0	111.9	108.3
Dec	93.1	97.4	95.4	97.0	92.3



Based on the three methods applied in this data set: US Housing from 1990 to 1995

We had observed that the forecasting data computed by the percentage trend method is the most precise and the closest to the actual data of 1996. The forecast by percentage to trend method is the line gray in the graph, which go closely to the initial data. The second best winner to this challenge is the percentage moving average, which is in yellow. It comes almost along to the gray line (the percentage to trend method).

DATA SET 2: DEATH 1973 -1978

OBSERVATION DATA

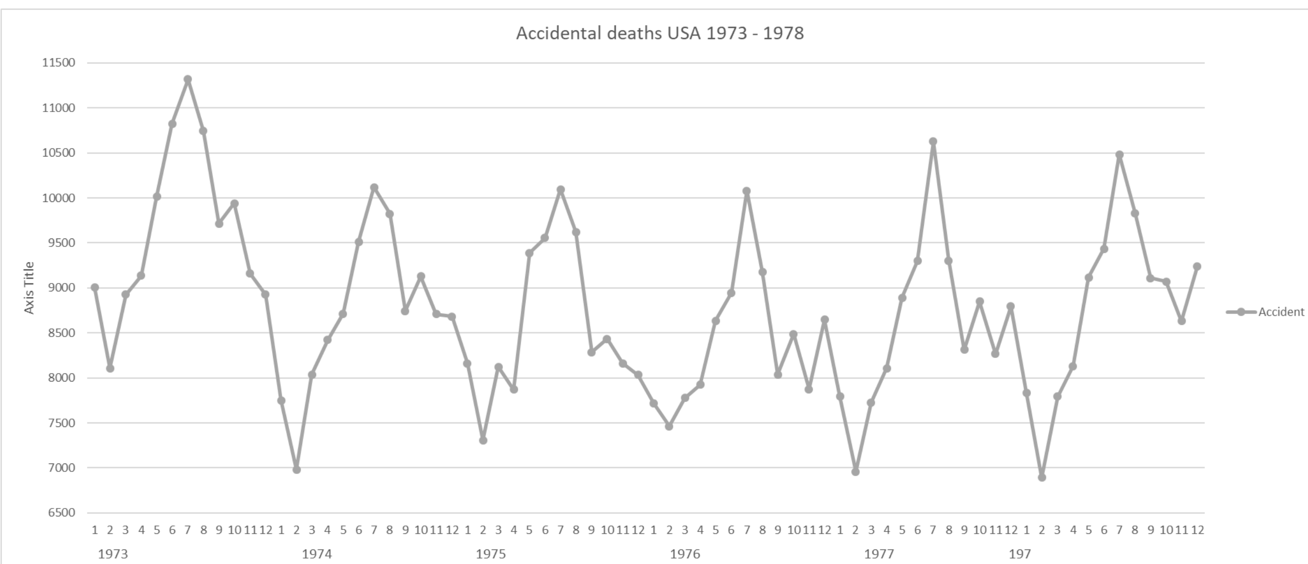
Based on observation data, we plot our first graph of observation data. We split data into training set and test set to easily to evaluate each and every model. Based on its prediction, we will choose the best candidate.

Training set: data from 1973 – 1977

Test set: data in 1978

Graph 17:

[*Initial data plot*]



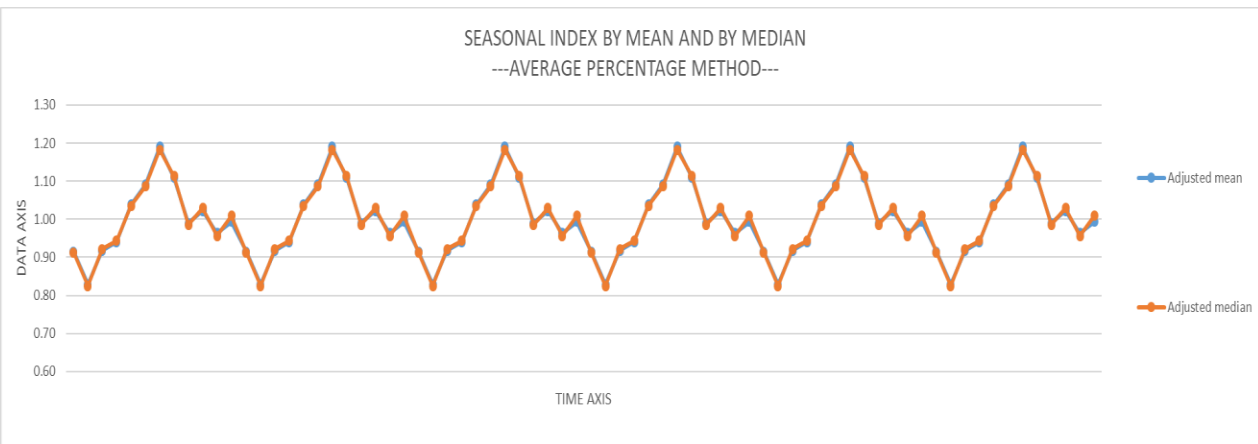
AVERAGE PERCENTAGE METHOD

Apply all the step as explained in page 7 – 10. Here, we will summarize quickly and jump to the analyze results and prediction. All calculation and graph in detail could be able to find in excel find under the sheet ‘AVERAGE PERCENTAGE METHOD’)

Step 1: Calculate the total sum by year and its average
(refer to column E and F in sheet ‘AVERAGE PERCENTAGE METHOD’)

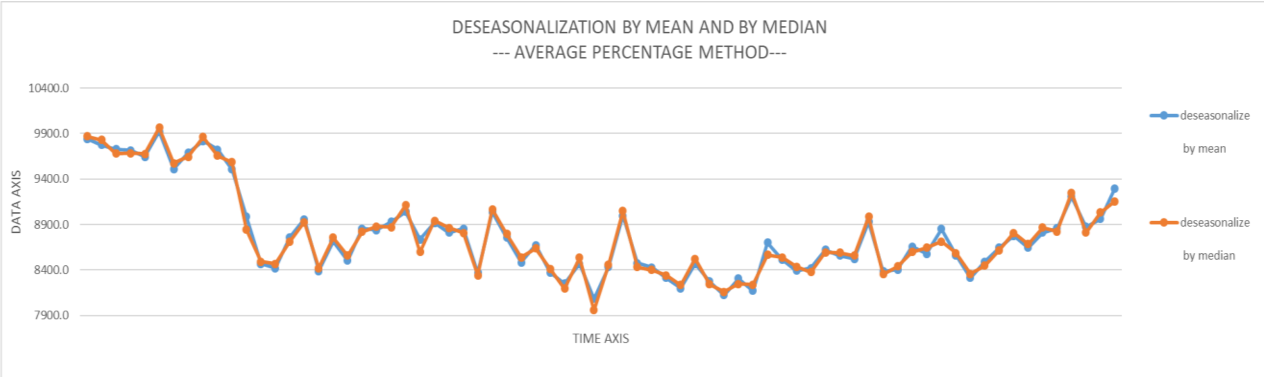
Step 2: Calculate the ratio by dividing the monthly data to its yearly average (refer to column G)
Then, we calculate the monthly average by its corresponding years by mean (column H) and by median (column J
We also do adjusted mean (column I) and adjusted median (column K) to obtain precise data.

Graph 18:
[Seasonal index]



Step 3: De-seasonalization (by mean in column L and by median in column M) and visualize it.
We divide the monthly original data to the seasonal
index by mean and by median

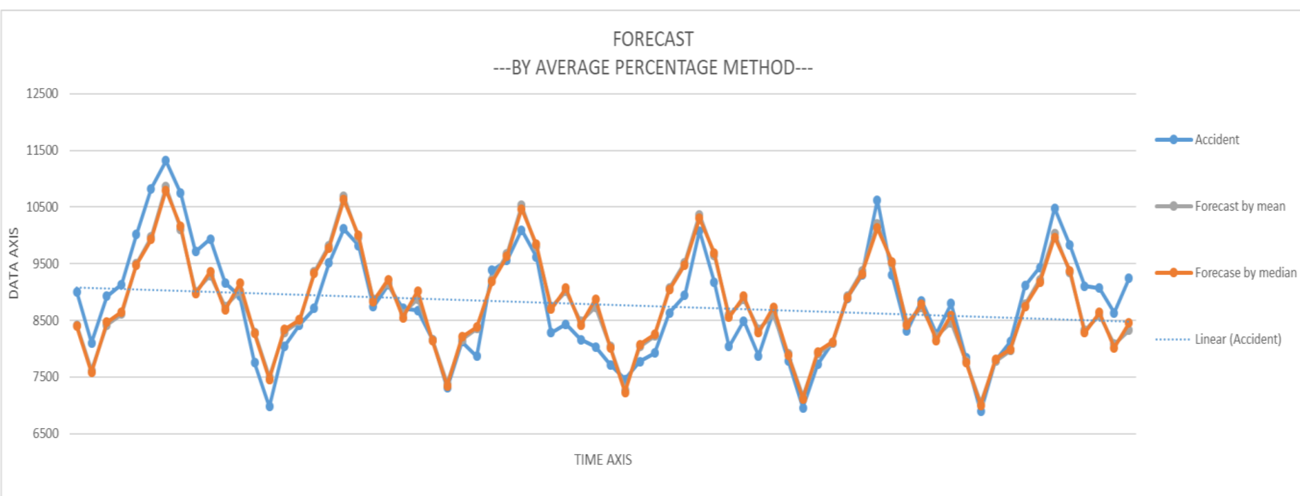
Graph 19:
[de-seasonal by mean and median]



Step 4: Apply linear regression to the deseasonal index by mean and by median to get a linear function: (x is the index 1-72, y is the deseasonal data by mean and by median)
By mean: Y = -19.85*X + 9391.51
By median: Y = -19.94*X + 9393
Then, calculate new y as showed in trend by mean (column N) and trend by median (column O)

Step 5: Forecasting data Y~ = T~*S~
T~ comes from step 4 and S~ comes from adjusted mean and adjusted median in step 2
We could do forecasting to the whole data and visualized it to evaluate our prediction

Graph 20:
[*Forecasting graph*]



Conclusion 1:
Based on the results of average to percentage method, the best candidate is the forecasting data by mean because it is fitted better to the initial data and give a better result.

QUADRATIC PERCENTAGE TO TREND METHOD

Apply all the step as explained in page 11 – 16. Here, we will summarize quickly and jump to the analyze results and prediction. All calculation and graph in detail could be able to find in excel find under the sheet 'PERCENTAGE TO TREND METHOD' AND 'QUADRATIC'). We apply both linear regression and quadratic regression to find out which will do a better job.
Please refer to the 'QUADRATIC sheet'

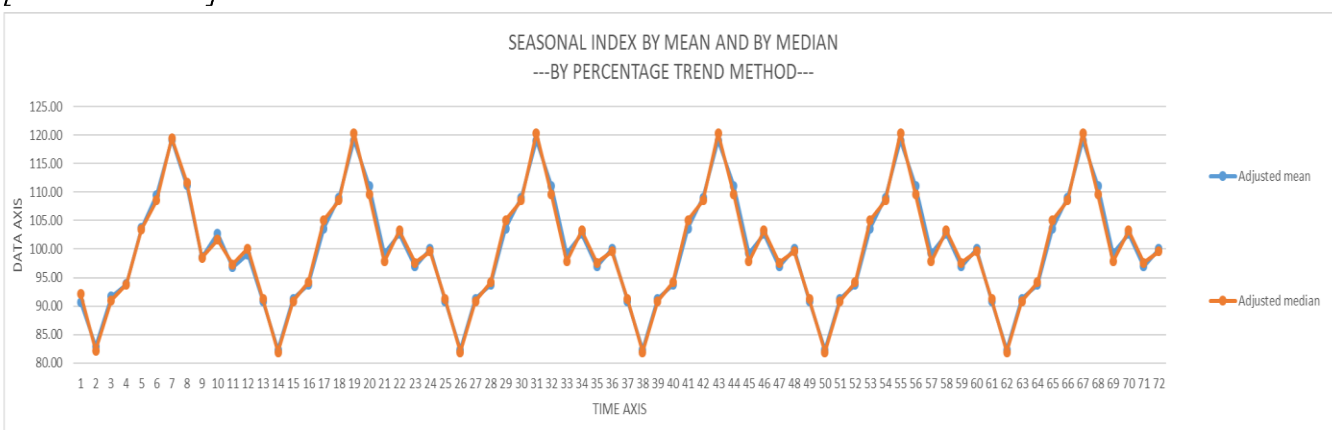
Step 1: Insert new X^2 by multiple month index* month index (refer to column B in sheet 'QUADRATIC'). Then calculate the yearly average of month (column C), x^2 (column D) and actual data (column G)

Step 2: Apply regression to obtain a quadratic regression. Calculate new trend as column J

$$Y = 10092.69 - 86.26*x + 1.07*x^2$$

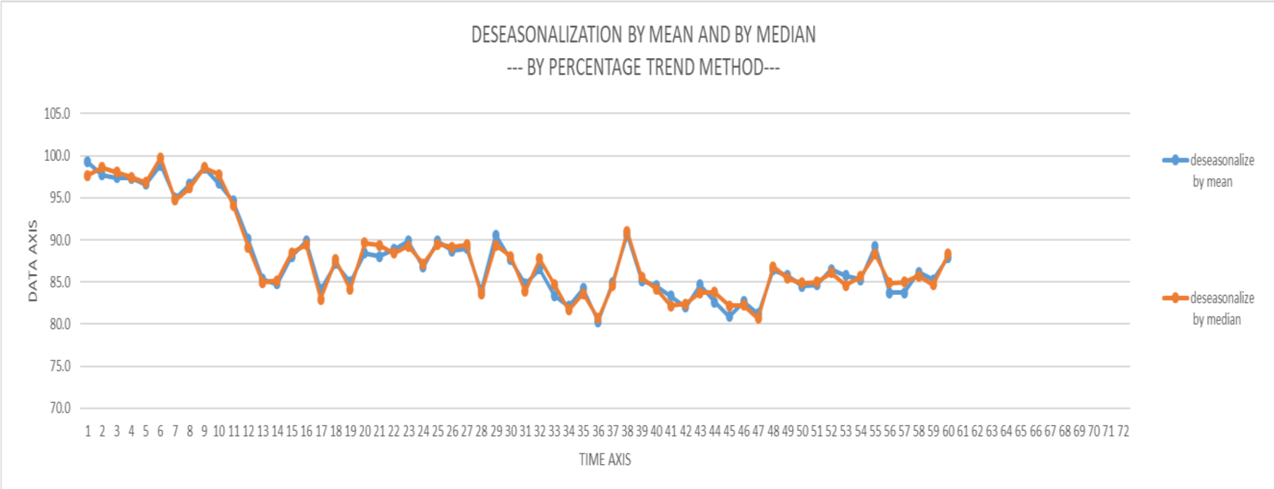
Step 3: Divide the given monthly values of the initial data by the corresponding trend values in step 2 (column K). Then, we calculate the monthly average by its corresponding years by mean (column L) and by median (column N). We also do adjusted mean (column M) and adjusted median (column O) to obtain precise data.

Graph 21:
[Seasonal index]



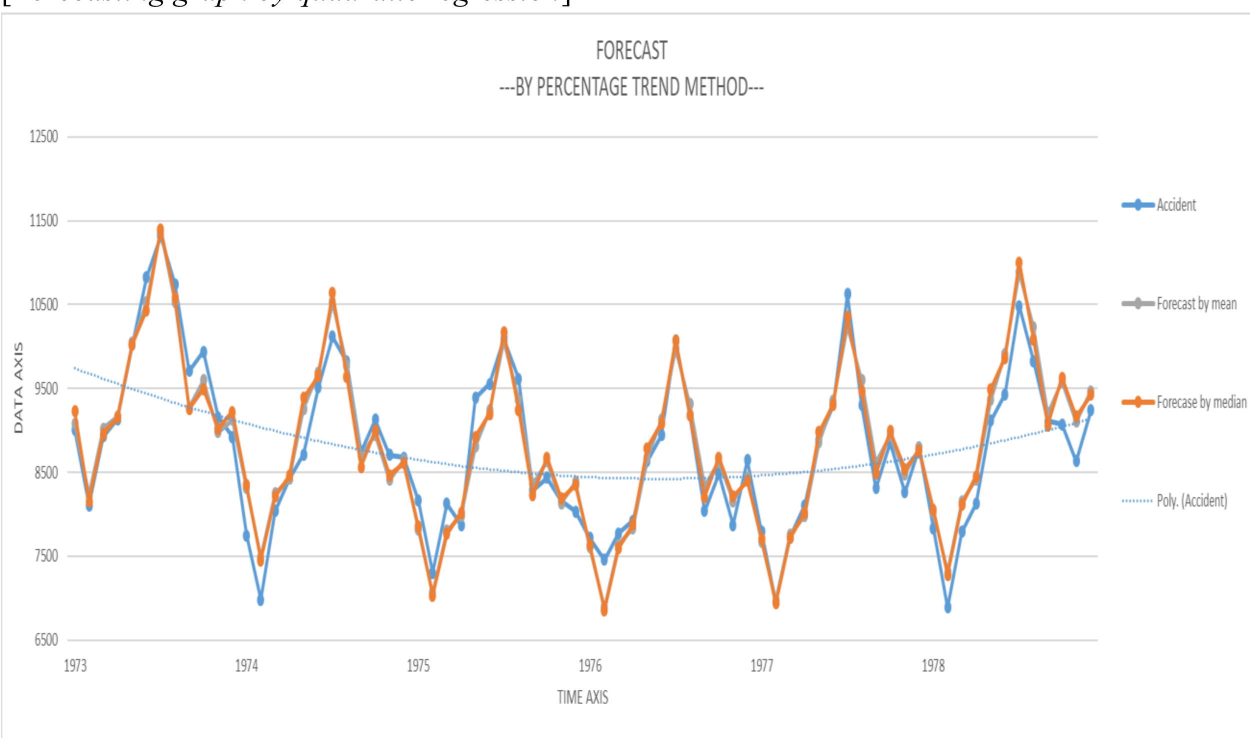
Step 4: De-seasonalization (by mean in column P and by median in column Q) and visualize it. In order to obtain it, we divide the monthly original data to the seasonal index by mean and by median

Graph 22:
[de-seasonal by mean and median]

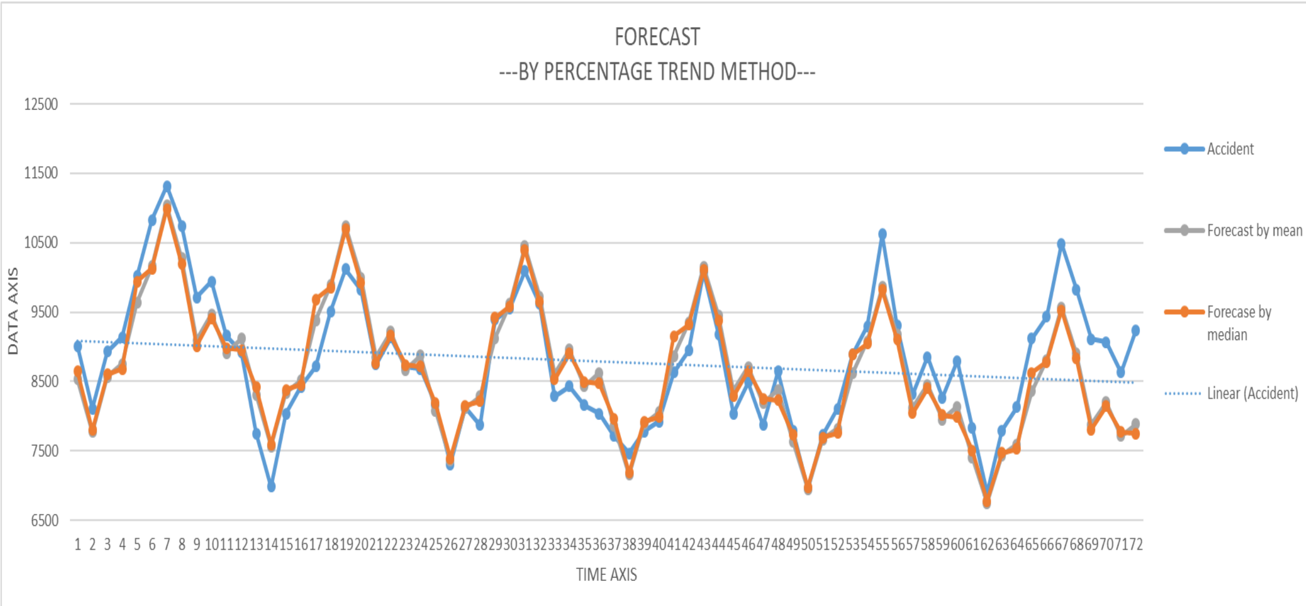


Step 6: Forecasting data Y~ = T~*S~
T~ comes from step 2 and S~ comes from adjusted mean and adjusted median in step 3
We could do forecasting to the whole data and visualized it to evaluate our prediction

Graph 23:
[*Forecasting graph by quadratic regression*]



Graph 24:
[*Forecasting graph by linear regression*]



Conclusion 2: Quadratic and Linear, which one is the best?
Based on the results, the best candidate is the quadratic regression, the quadratic above to identified a slightly increasing in the trend of the year 1978. Thus, it provides a better trend to better predictions. Thus, the best candidate is the quadratic by mean because it is fitted better to the initial data and give a better result.

PERCENTAGE MOVING AVERAGE METHOD

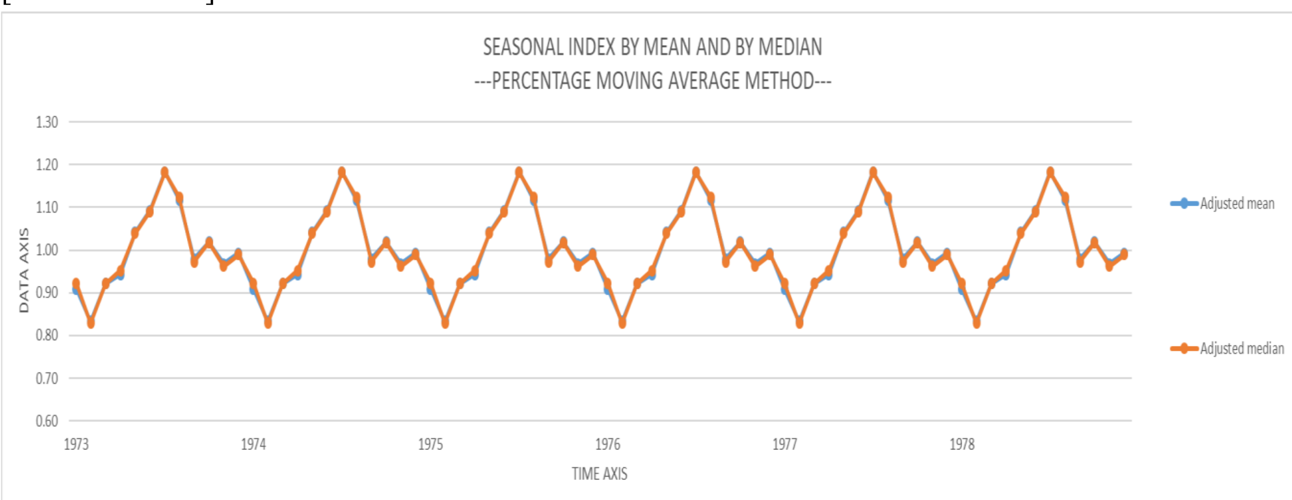
Apply all the step as explained in page 17 – 22. Here, we will summarize quickly and jump to the analyze results and prediction. All calculation and graph in detail could be able to find in excel find under the sheet 'PERCENTAGE MOVING AVERAGE'.

Step1: From the initial data, calculate 12 months moving average (refer to column D in 'PERCENTAGE MOVING AVERAGE')

Step 2: We compute the 12 months centered moving average (column E)

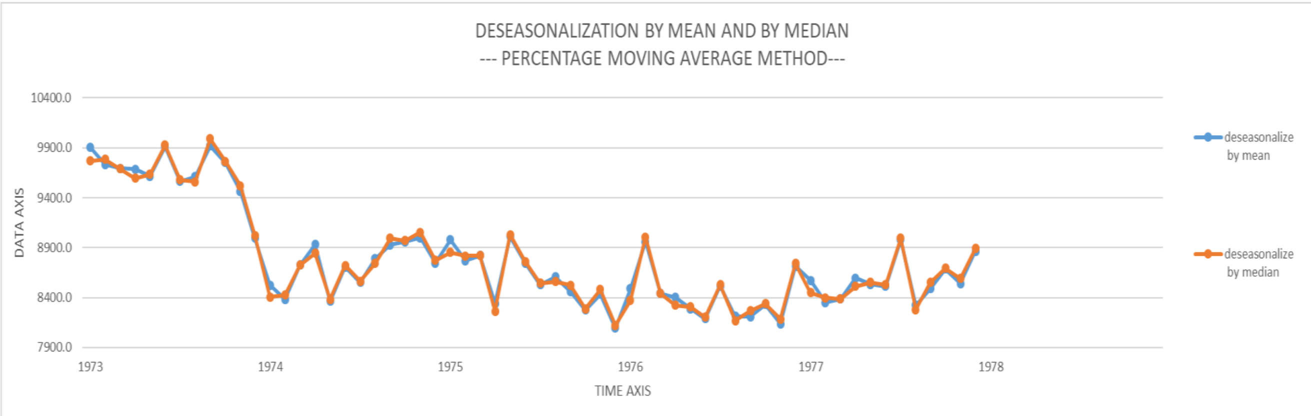
Step 3: Compute the ratio
Divide the initial data to the 12 months centered moving average of step 2. We compute the percentage as column F.
Then, we calculate seasonal index by take monthly average of 5 years by mean (column G) and by median (Column I) and do adjusted mean (column H) and adjusted median (column J)

Graph 25:
[Seasonal index]



Step 4: De-seasonalization and visualize it. We divide the monthly original data to the seasonal index in step 3 to compute the deseasonalize data by mean (column K) and by median (column L)

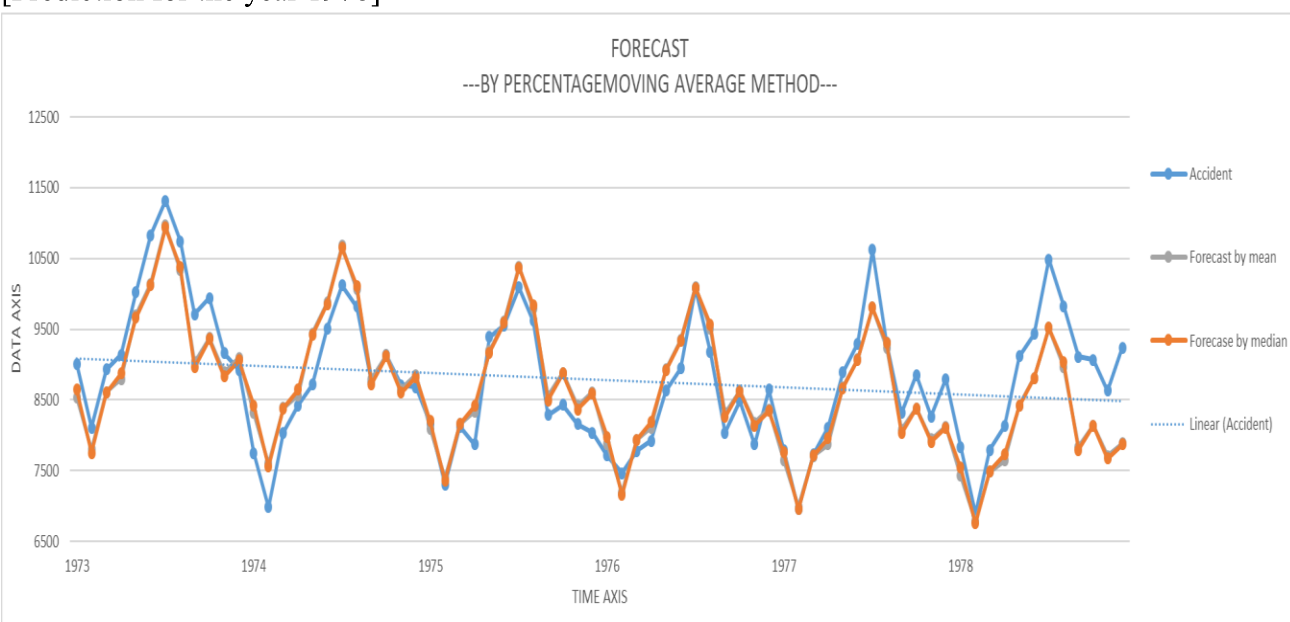
Graph 26:
[De - seasonal index]



Step 5: Apply linear regression to the deseasonal data by mean and by median to get a linear function. Then, we apply the function to get 84 new predicted data by mean (column M) and by median (column N)

Step 6: Forecasting data Y~ = T~*S~
T~ comes from step 5 and S~ comes from adjusted mean and adjusted median in step 3
We could do forecasting to the whole data and visualized it to evaluate our forecasting

Graph 27:
[Prediction for the year 1978]



Conclusion 1:
Based on the results, the best candidate is the forecasting by percentage moving average by median because it is fitted better to the initial data and give a better result.

ARMA

We are using RStudio in the following step:

Step 1: Install the libraries (if you already installed it, just need to call and use)

Step 2: Import the data set

Step 3: Split into training set and test set

Step 4: Converting training set into time series

Step 5: Plot the training set

Step 6: Visualize all components of time series (trend, seasonal index…)

Step 7: Plot ACF and PACF to find good p and q

Step 8: Decompose the trend, seasonal and plot ACF again to visualize the good p and q

Step 9: Run auto.arima()

Step 10: Visualize the arima(p,d,q) process and its predictions

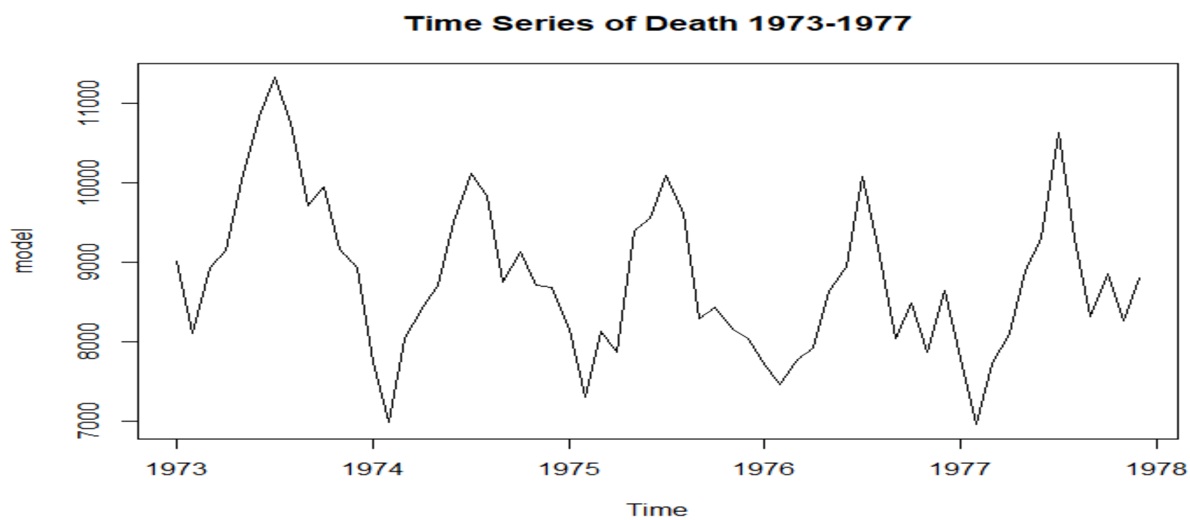
ARIMA() required 3 the three integer components (p, d, q) are the AR order, the degree of differencing, and the MA order. ARMA = ARIMA() with parameter d = 0 or degree of diff = 0

**All the step have written clearly in the script 'myDeath.r'.

Step 5: Plot the training set

Graph 28:

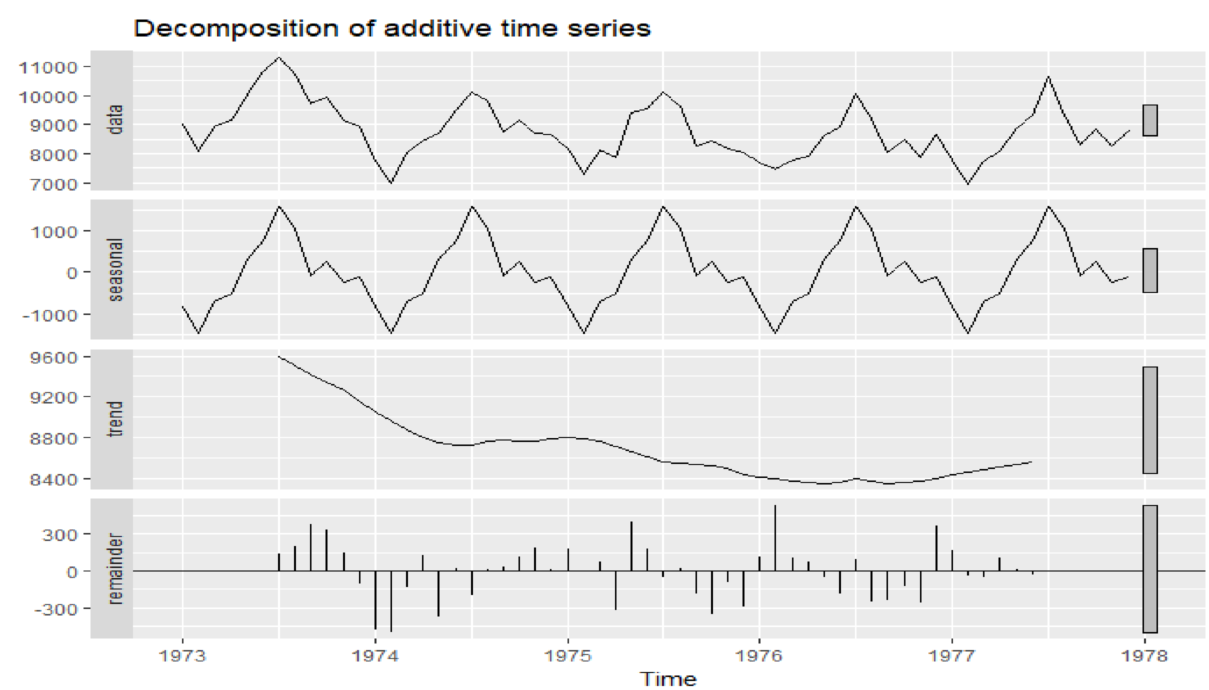
[*Initial data plot in RStudio*]



Step 6: Visualize all components of time series (trend, seasonal index…)

Graph 29:

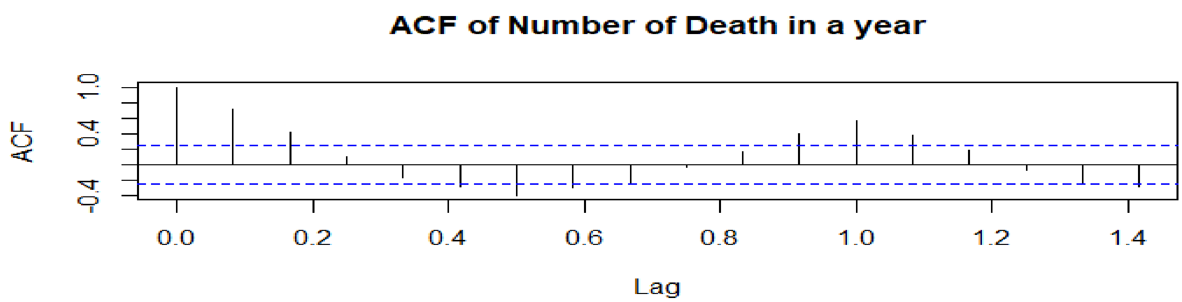
[trend, seasonal index plot in RStudio]

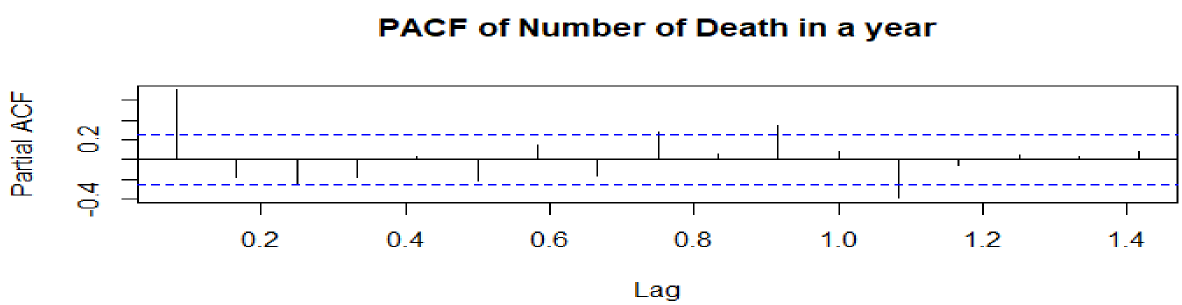


Step 7: Plot ACF and PACF to find good p and q

<u>Graph 30:</u>

[ACF and PACF plot]

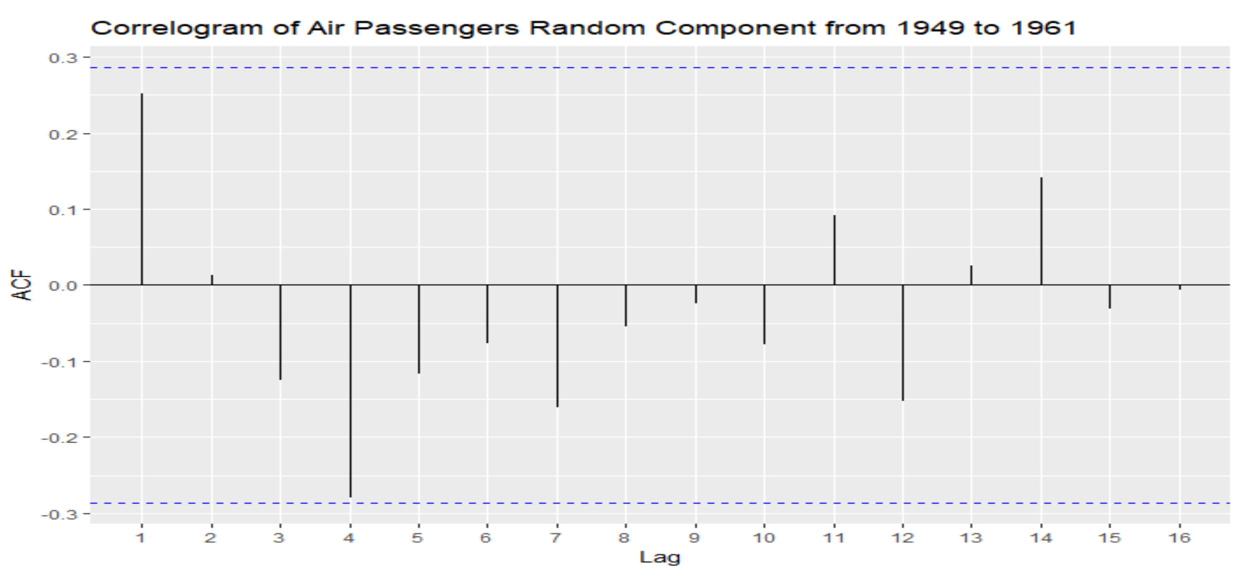




Step 8: Decompose the trend, seasonal and plot ACF again to visualize the good p and q

<u>Graph 31:</u>

[second ACF plot in RStudio]



Step 9: Run auto.arima()

Results obtain ARIMA(0,1,1) and AIC = 689.78

theta1 = -0.4303 and residual = -0.4503

```
Component from 1949 to 1961")
> arima <- auto.arima(model)
> arima
Series: model
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.4303  -0.4503
s.e.   0.1367   0.1831

sigma^2 estimated as 119659:  log likelihood=-341.89
AIC=689.78   AICc=690.34   BIC=695.34
>
```

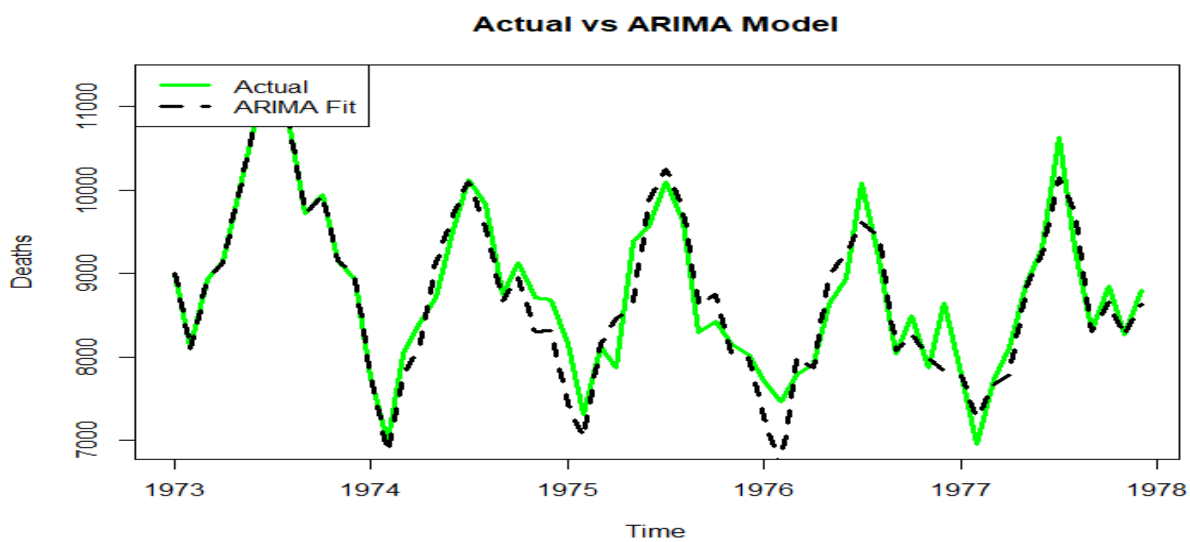
Try different ARMA but AIC are not as good at optimal arima()

ARMA process	ACI
001	966.09
002	959.89
003	953.69
103	954.19

Step 10: Visualize the arima(p,d,q) process and its predictions

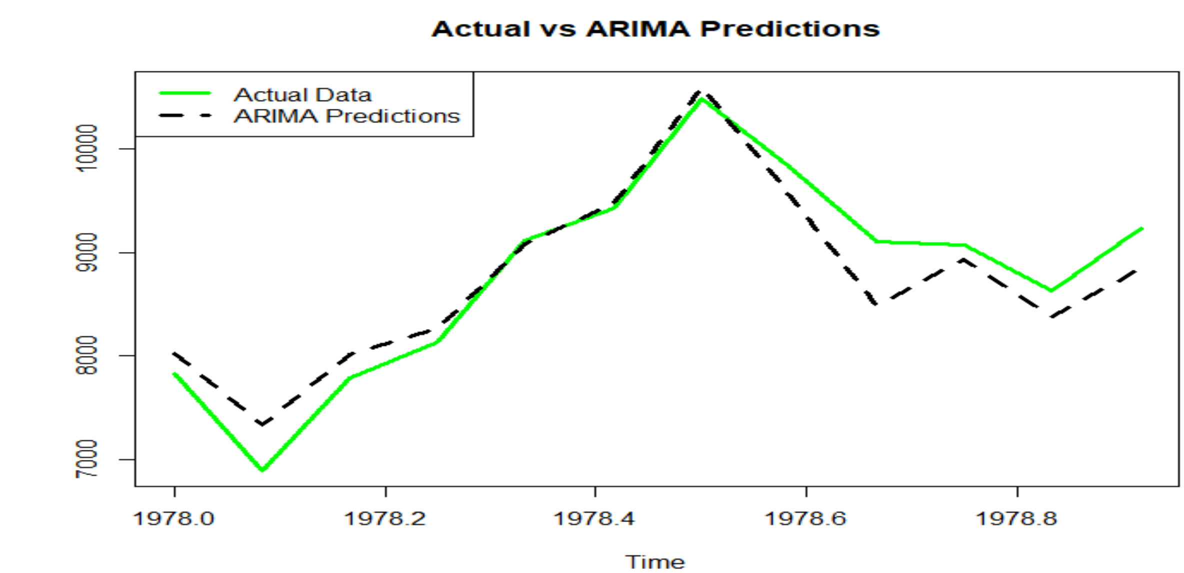
Graph 32:

[Actual data and ARIMA model fitted]



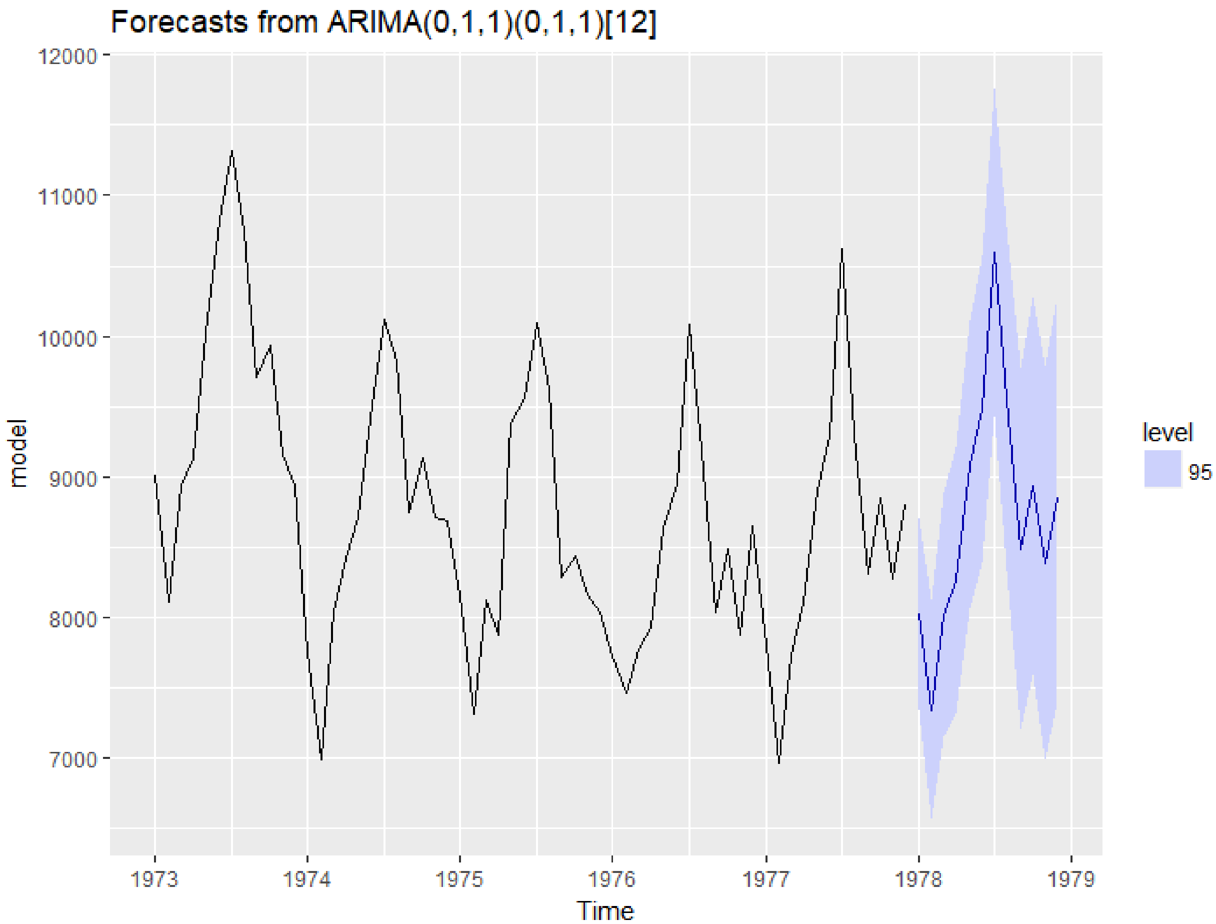
Graph 33:

[Actual data of 1996 and ARIMA predictions]



<u>Graph 34:</u>

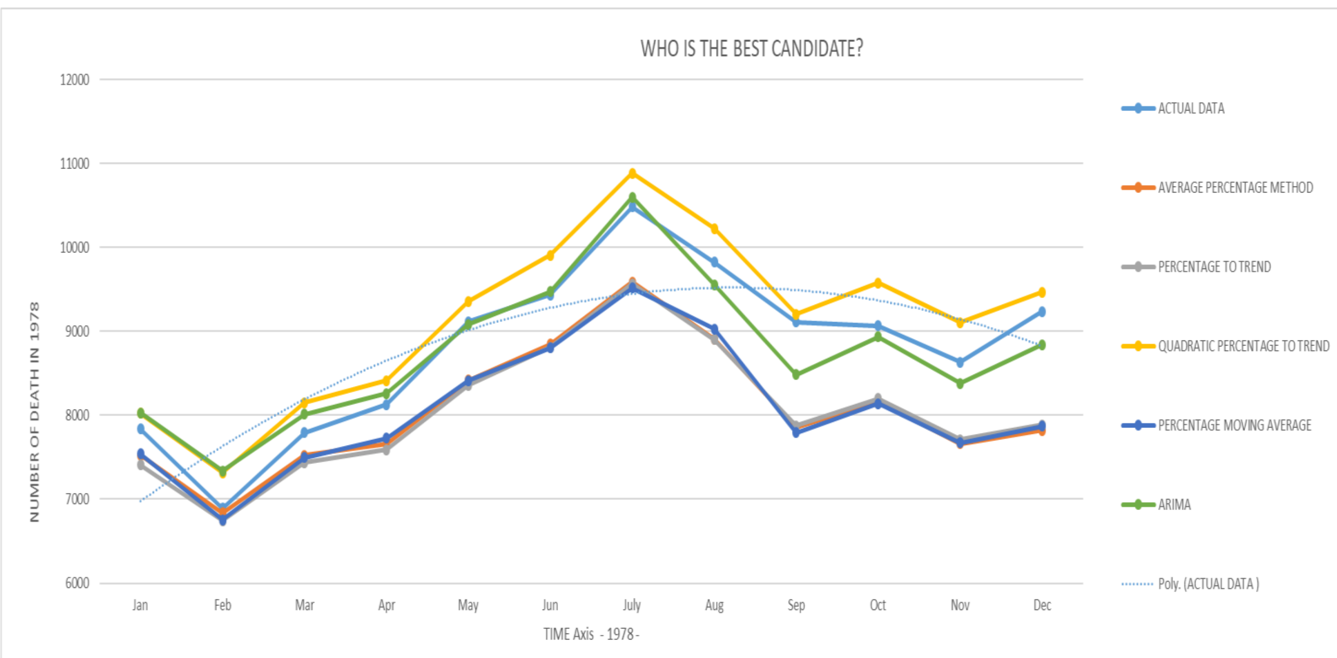
[Actual data and predictions with highest and lowest bounds]



EVALUATED RESULTS OF DEATH 1977

Table 1: Compare obtained results to the actual data of 1977

1977	Actual data	Average percentage methods	Percentage to trend methods	Quadratic Percentage to trend	Percentage moving average	ARIMA
Jan	7836	7526.3	7408.2	8026.0	7538.8	8027.4
Feb	6892	6840.4	6746.1	7314.5	6757.1	7335.4
Mar	7791	7522.0	7434.1	8152.1	7495.0	8014.6
Apr	8129	7663.0	7592.4	8415.3	7728.1	8259.8
May	9115	8417.5	8358.9	9361.6	8416.0	9084.6
Jun	9434	8849.0	8808.3	9911.3	8806.8	9476.4
July	10484	9589.3	9568.6	10887.6	9518.0	10598.1
Aug	9827	8906.6	8902.0	10228.2	9030.5	9557.8
Sep	9110	7862.0	7878.6	9202.8	7792.4	8485.7
Oct	9070	8164.3	8200.4	9580.6	8137.4	8938.1
Nov	8633	7662.4	7710.7	9105.1	7673.6	8384.1
Dec	9240	7821.1	7888.6	9467.8	7870.8	8844.8



Based on the three methods applied in this data set: Death in 1973 -1978
We had observed that the forecasting data computed by the ARIMA (0,1,1) is the most precise and the closest to the actual data of 1978. It is in green and approach closely to blue line, which is initial data in the graph. The second best winner to this challenge is the quadratic percentage moving average, which is in yellow.

DATA SET 3: AIR PASSENGER 1949 -1960

OBSERVATION DATA

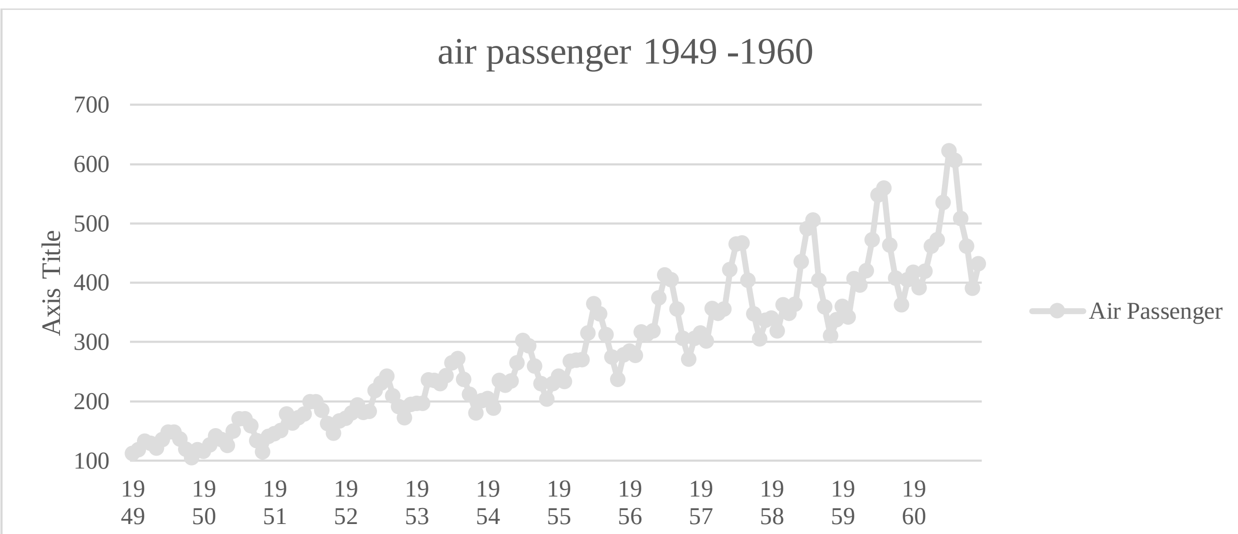
Based on observation data, we plot our first graph of observation data. We split data into training set and test set to easily to evaluate each and every model. Based on its prediction, we will choose the best candidate.

Training set: data from 1949 - 1959

Test set: data in 1960

<u>Graph 35:</u>

[*Initial data plot*]



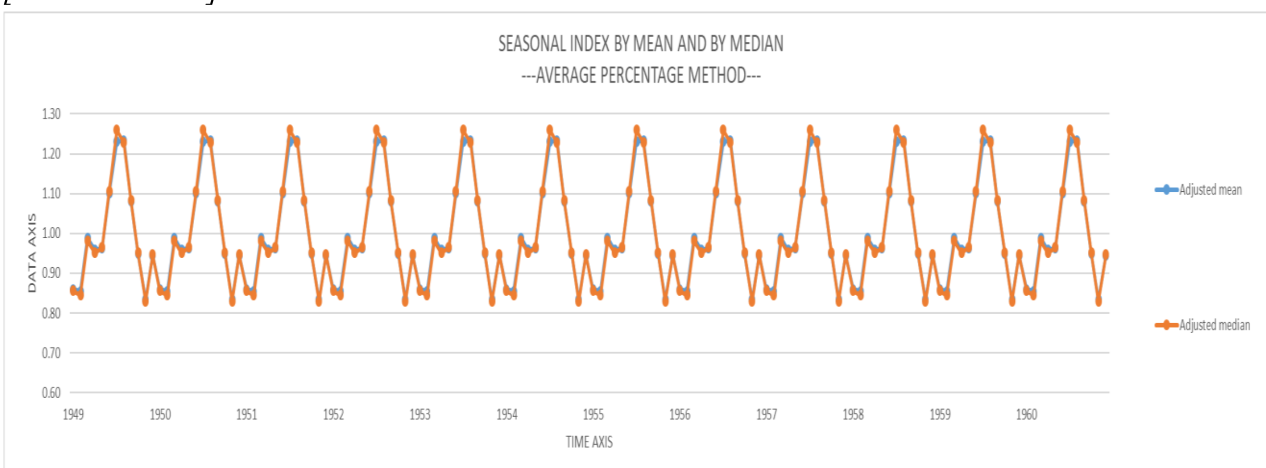
AVERAGE PERCENTAGE METHOD

Apply all the step as explained in page 7 – 10. Here, we will summarize quickly and jump to the analyze results and prediction. All calculation and graph in detail could be able to find in excel find under the sheet 'AVERAGE PERCENTAGE METHOD')

Step 1: Calculate the total sum by year and its average
(refer to column E and F in sheet 'AVERAGE PERCENTAGE METHOD')

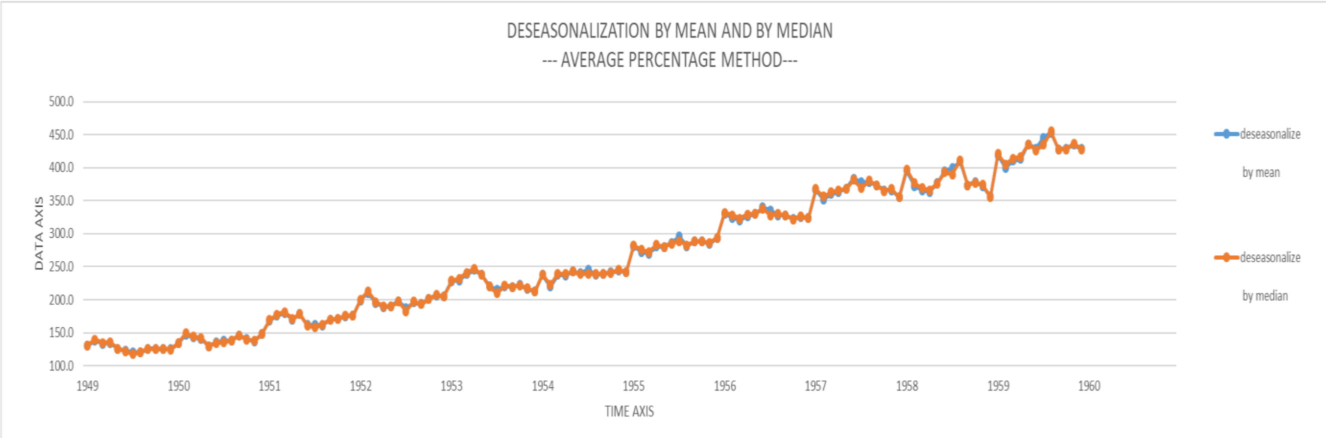
Step 2: Calculate the ratio by dividing the monthly data to its yearly average (refer to column G) Then, we calculate the monthly average by its corresponding years by mean (column H) and by median (column J
We also do adjusted mean (column I) and adjusted median (column K) to obtain precise data.

Graph 36:
[Seasonal index]



Step 3: De-seasonalization (by mean in column L and by median in column M) and visualize it. We divide the monthly original data to the seasonal
index by mean and by median

Graph 37:
[de-seasonal by mean and median]



Step 4: Apply linear regression to the deseasonal index by mean and by median to get a linear function: (x is the index 1-132, y is the deseasonal data by mean and by median)

By mean: Y = 2.53*X + 93.83

By median: Y = 2.53*X + 94

Then, calculate new y as showed in trend by mean (column N) and trend by median (column O)

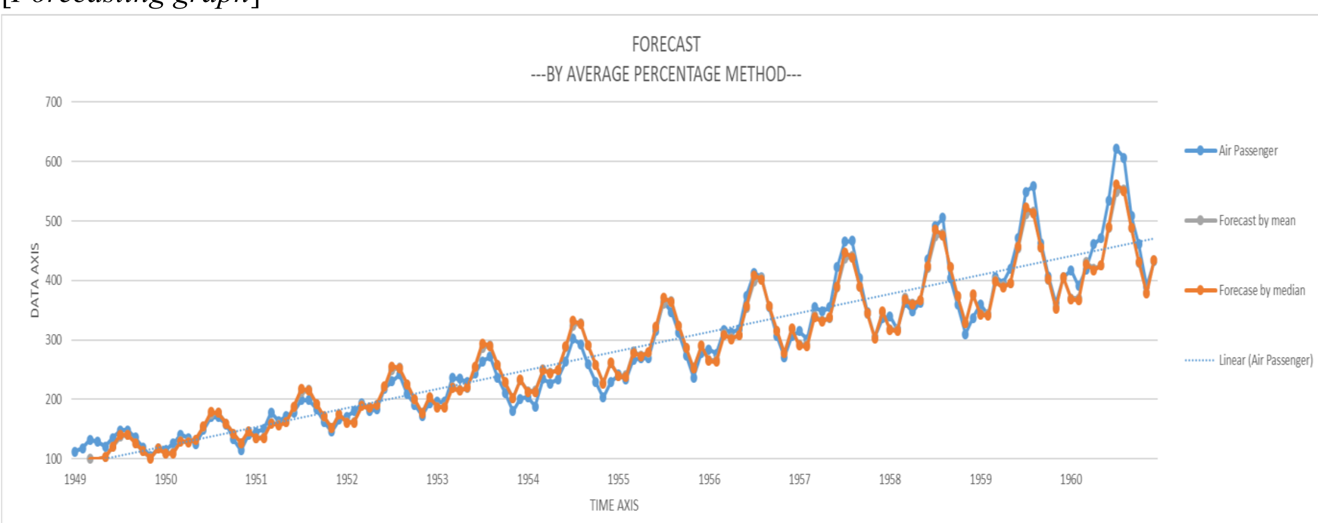
Step 5: Forecasting data Y~ = T~*S~

T~ comes from step 4 and S~ comes from adjusted mean and adjusted median in step 2

We could do forecasting to the whole data and visualized it to evaluate our prediction

Graph 38:

[*Forecasting graph*]



Conclusion 1:

Based on the results of average to percentage method, the best candidate is the forecasting data by mean because it is fitted better to the initial data and give a better result.

QUADRATIC PERCENTAGE TO TREND METHOD

Apply all the step as explained in page 11 – 16. Here, we will summarize quickly and jump to the analyze results and prediction. All calculation and graph in detail could be able to find in excel find under the sheet ‘PERCENTAGE TO TREND METHOD’ AND ‘QUADRATIC’). We apply both linear regression and quadratic regression to find out which will do a better job.
Please refer to the ‘QUADRATIC sheet’

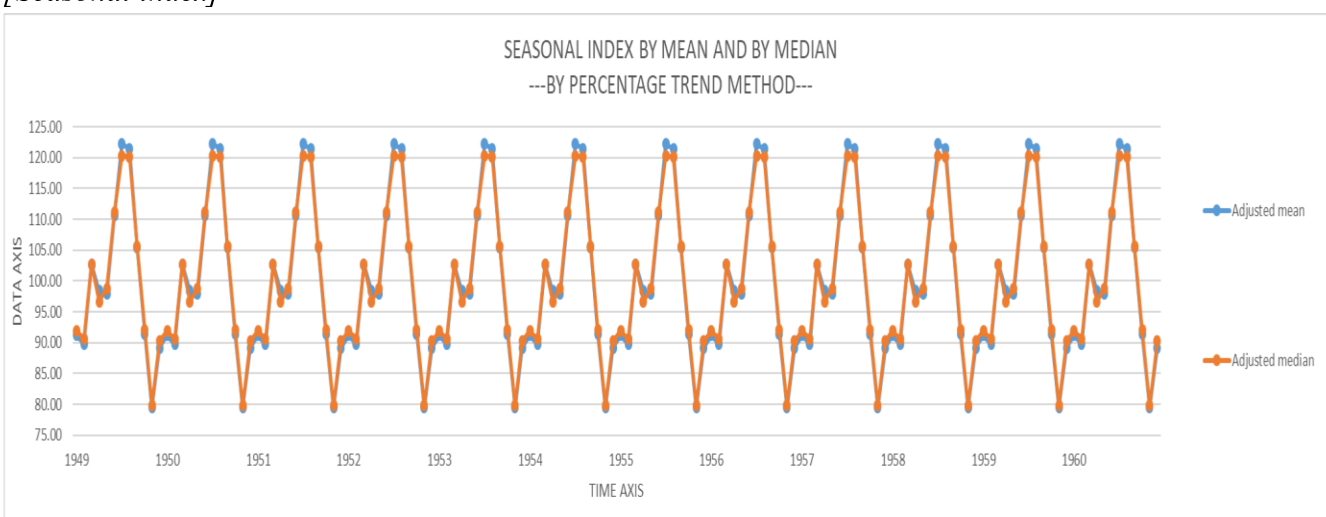
Step 1: Insert new X^2 by multiple month index* month index (refer to column B in sheet ‘QUADRATIC’). Then calculate the yearly average of month (column C), x^2 (column D) and actual data (column G)

Step 2: Apply regression to obtain a quadratic regression. Calculate new trend as column J

$$Y = 1.719*x + 0.006*x^2 + 110.675$$

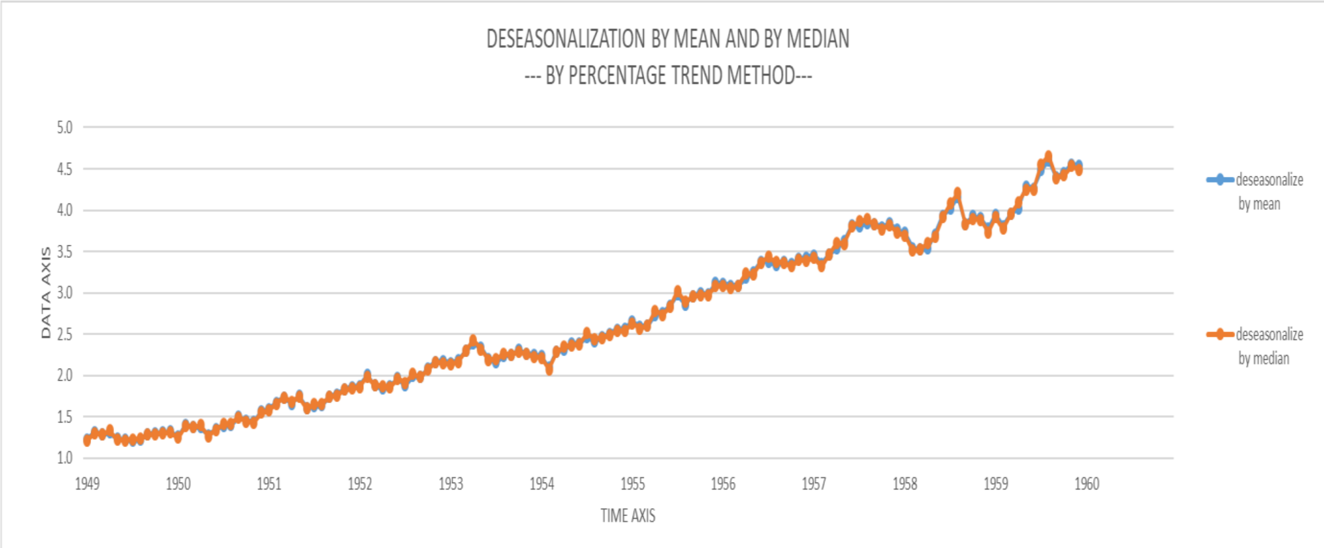
Step 3: Divide the given monthly values of the initial data by the corresponding trend values in step 2 (column K). Then, we calculate the monthly average by its corresponding years by mean (column L) and by median (column N). We also do adjusted mean (column M) and adjusted median (column O) to obtain precise data.

Graph 39:
[Seasonal index]



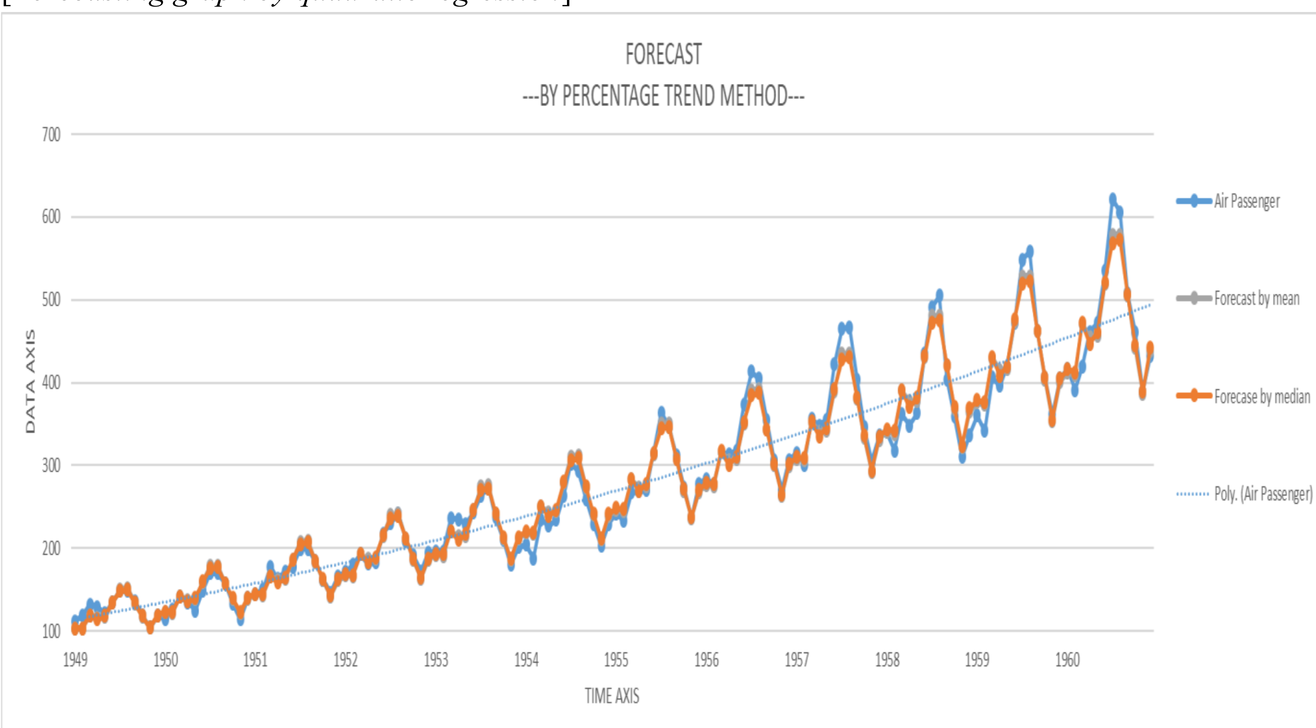
Step 4: De-seasonalization (by mean in column P and by median in column Q) and visualize it. In order to obtain it, we divide the monthly original data to the seasonal index by mean and by median

Graph 40:
[de-seasonal by mean and median]

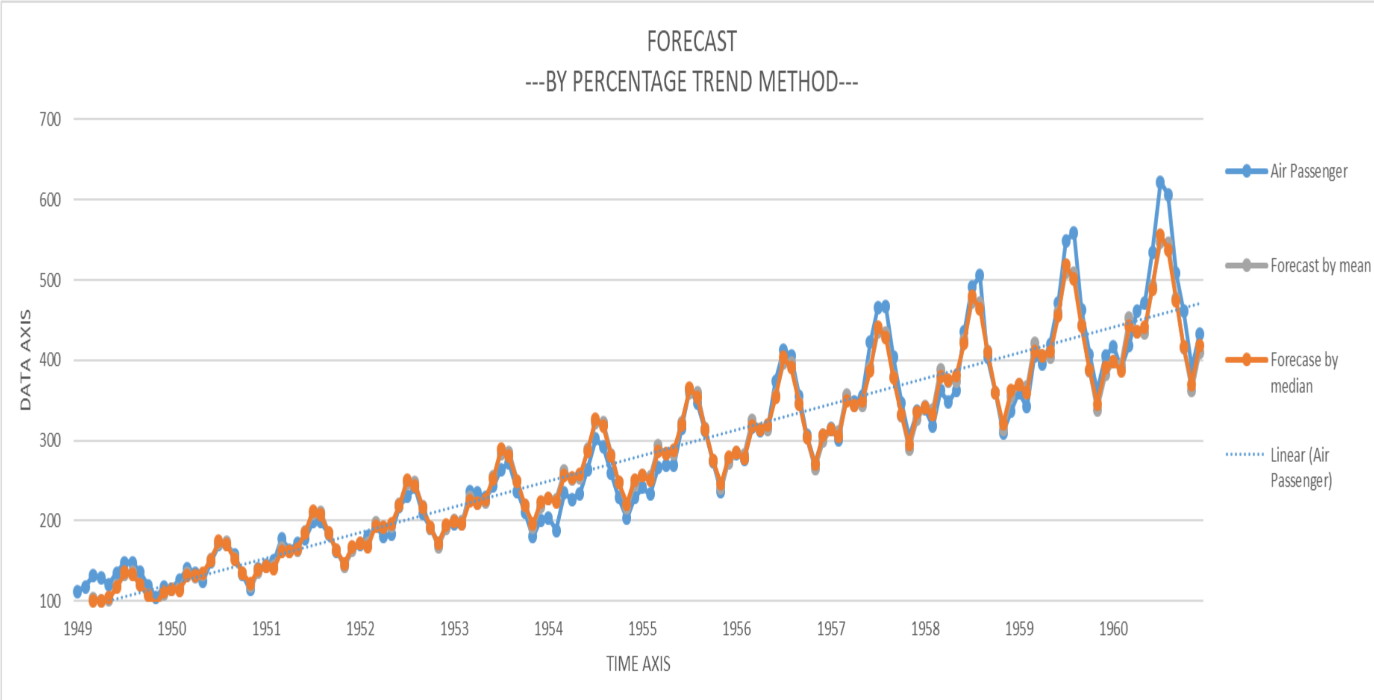


Step 6: Forecasting data Y~ = T~*S~
T~ comes from step 2 and S~ comes from adjusted mean and adjusted median in step 3
We could do forecasting to the whole data and visualized it to evaluate our prediction

Graph 41:
[*Forecasting graph by quadratic regression*]



Graph 42:
[*Forecasting graph by linear regression*]



Conclusion 2: Quadratic and Linear, which one is the best?
Based on the results, the best candidate is the quadratic regression, the quadratic above to identified a slightly increasing in the trend of the year 1960. Thus, it provides a better trend to better predictions. Thus, the best candidate is the quadratic by median because it is fitted better to the initial data and give a better result.

PERCENTAGE MOVING AVERAGE METHOD

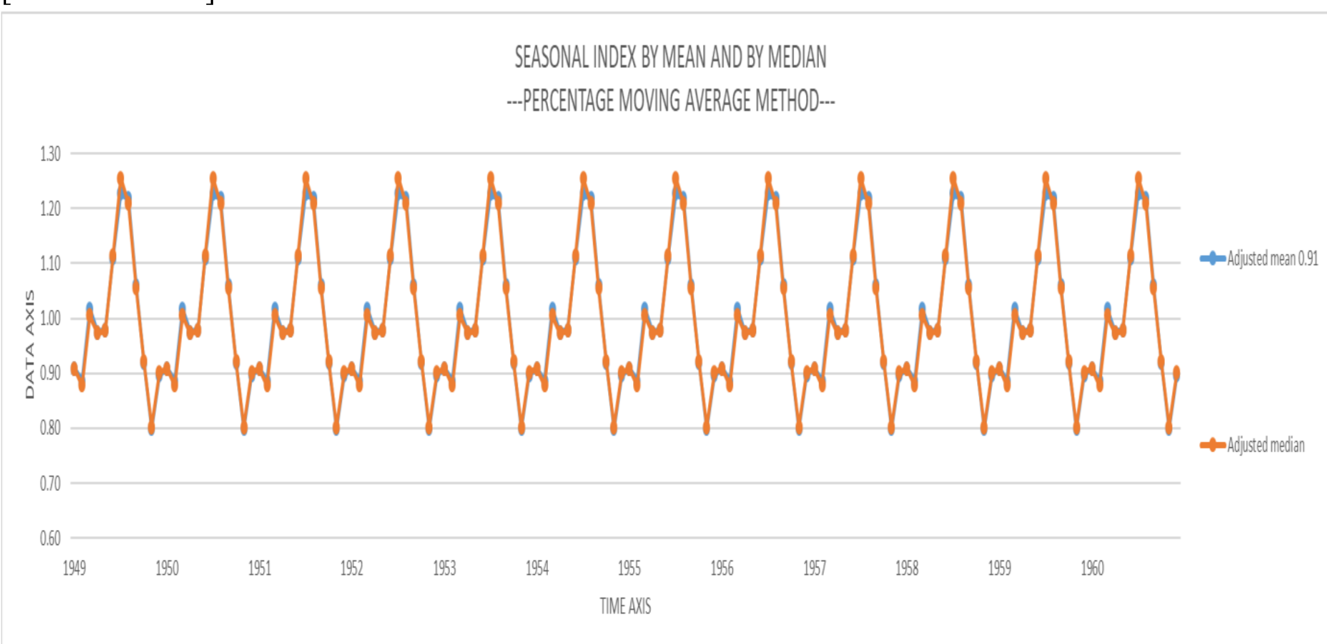
Apply all the step as explained in page 17 – 22. Here, we will summarize quickly and jump to the analyze results and prediction. All calculation and graph in detail could be able to find in excel find under the sheet 'PERCENTAGE MOVING AVERAGE'.

Step1: From the initial data, calculate 12 months moving average (refer to column D in 'PERCENTAGE MOVING AVERAGE')

Step 2: We compute the 12 months centered moving average (column E)

Step 3: Compute the ratio
Divide the initial data to the 12 months centered moving average of step 2. We compute the percentage as column F.
Then, we calculate seasonal index by take monthly average by mean (column G) and by median (Column I) and do adjusted mean (column H) and adjusted median (column J)

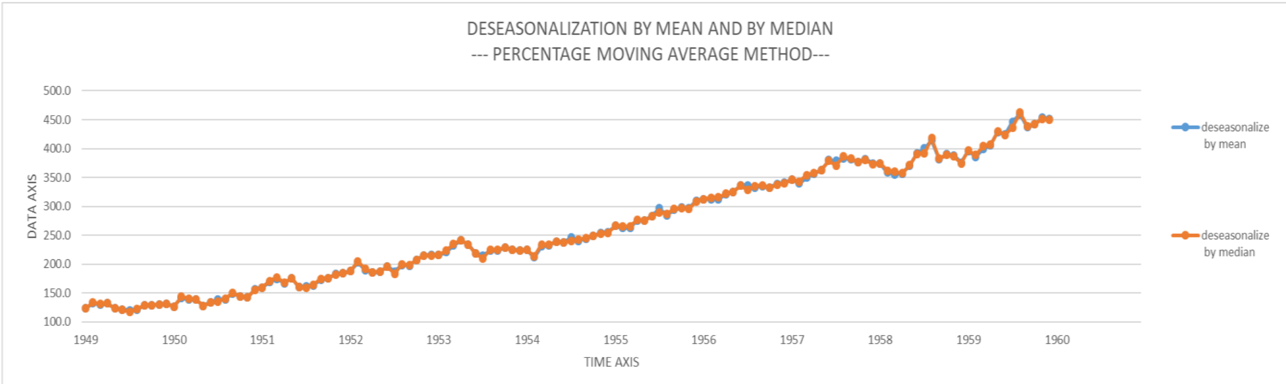
Graph 43:
[Seasonal index]



Step 4: De-seasonalization and visualize it. We divide the monthly original data to the seasonal index in step 3 to compute the deseasonalize data by mean (column K) and by median (column L)

Graph 44:

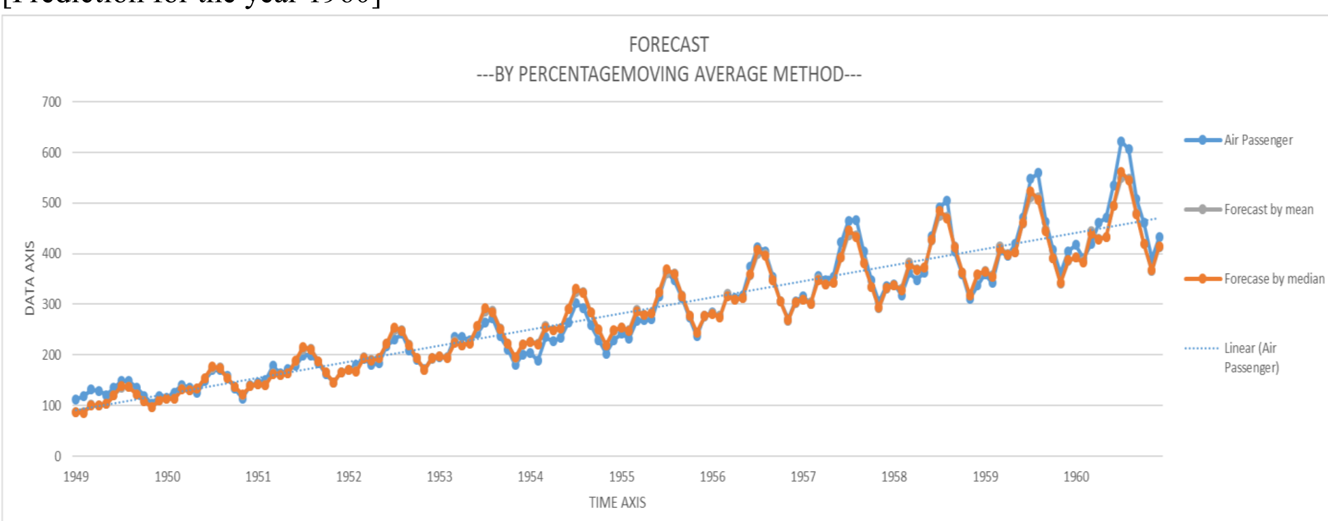
[De - seasonal index]



Step 5: Apply linear regression to the deseasonal data by mean and by median to get a linear function. Then, we apply the function to get 84 new predicted data by mean (column M) and by median (column N)
By mean : Y = 2.55*X + 92.52
By median: Y = 2.55*X +92.69

Step 6: Forecasting data Y~ = T~*S~
T~ comes from step 5 and S~ comes from adjusted mean and adjusted median in step 3
We could do forecasting to the whole data and visualized it to evaluate our forecasting

Graph 45:
[Prediction for the year 1960]



Conclusion 1:
Based on the results, the best candidate is the forecasting by percentage moving average by median because it is fitted better to the initial data and give a better result.

ARMA

We are using RStudio in the following step:

Step 1: Install the libraries (if you already installed it, just need to call and use)

Step 2: Import the data set

Step 3: Split into training set and test set

Step 4: Converting training set into time series

Step 5: Plot the training set

Step 6: Visualize all components of time series (trend, seasonal index…)

Step 7: Plot ACF and PACF to find good p and q

Step 8: Decompose the trend, seasonal and plot ACF again to visualize the good p and q

Step 9: Run auto.arima()

Step 10: Visualize the arima(p,d,q) process and its predictions

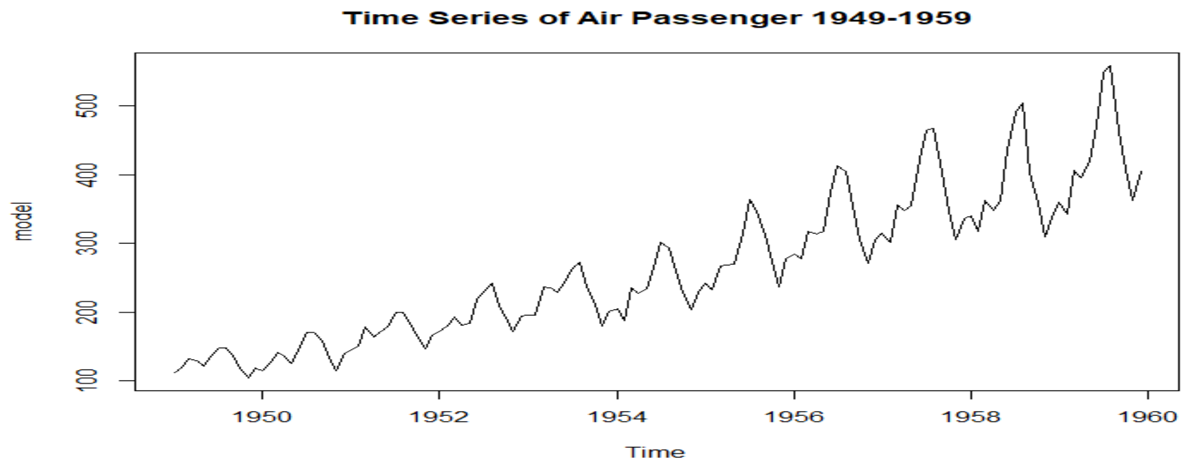
ARIMA() required 3 the three integer components (p, d, q) are the AR order, the degree of differencing, and the MA order. ARMA = ARIMA() with parameter d = 0 or degree of diff = 0

**All the step have written clearly in the script ‘myAirPassengers.r’.

Step 5: Plot the training set

Graph 46:

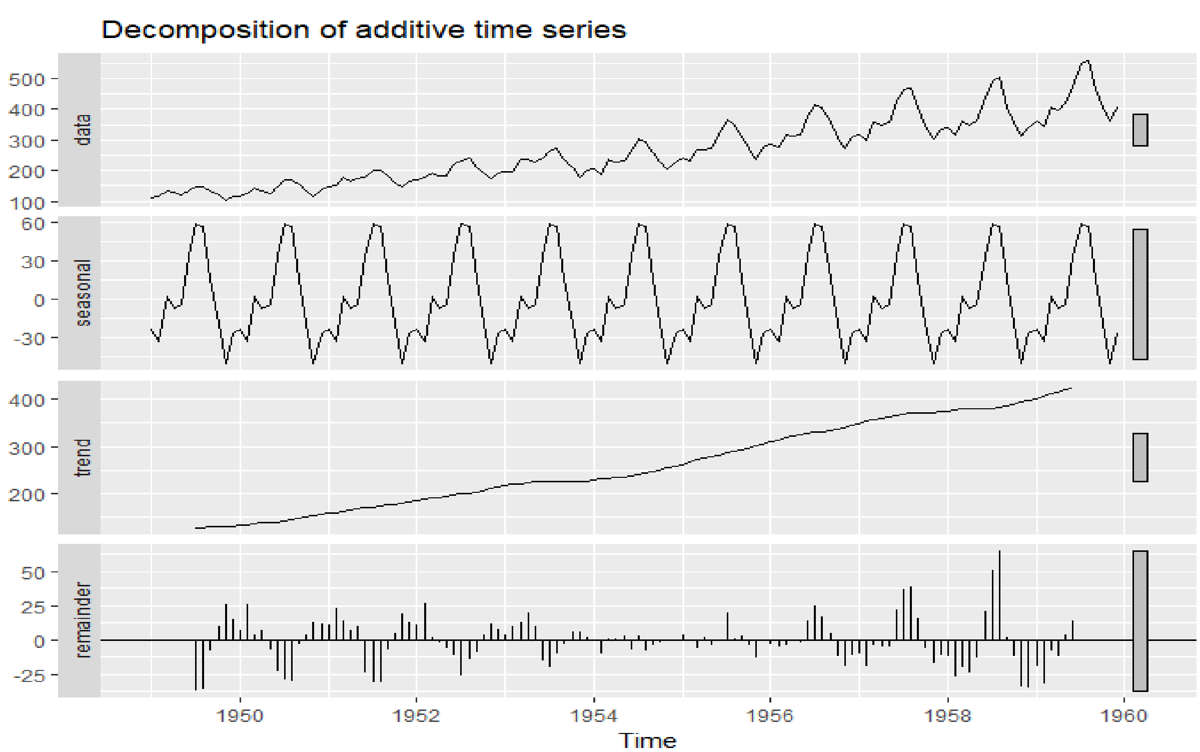
[*Initial data plot in RStudio*]



Step 6: Visualize all components of time series (trend, seasonal index…)

Graph 47:

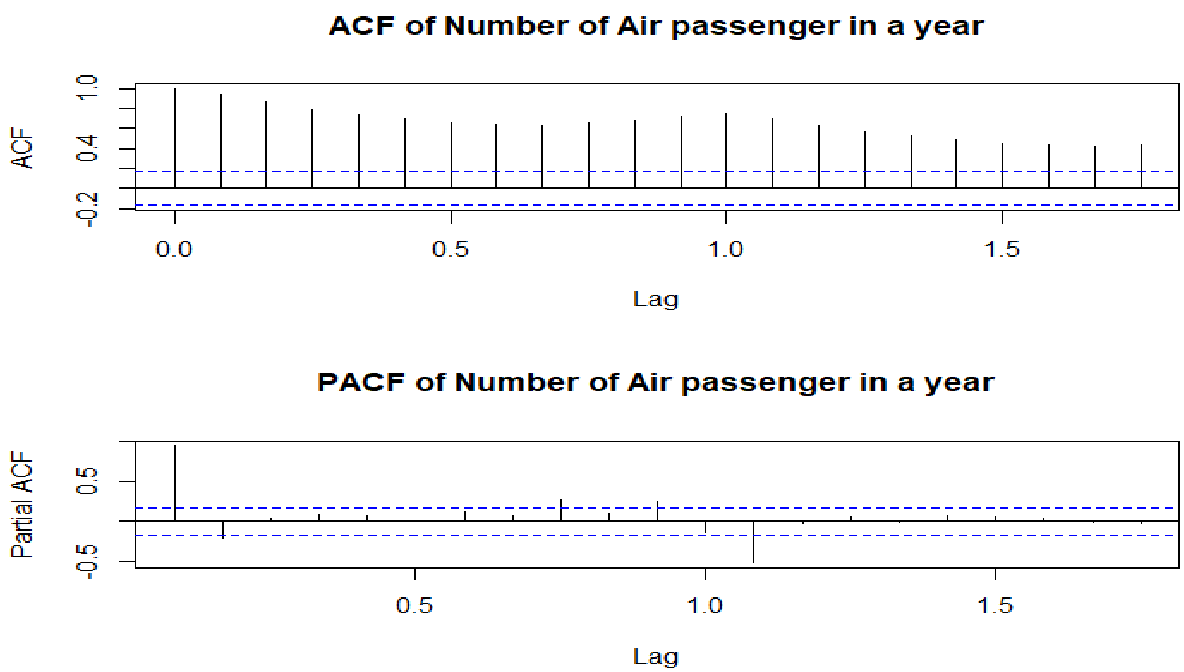
[trend, seasonal index plot in RStudio]



Step 7: Plot ACF and PACF to find good p and q

Graph 48:

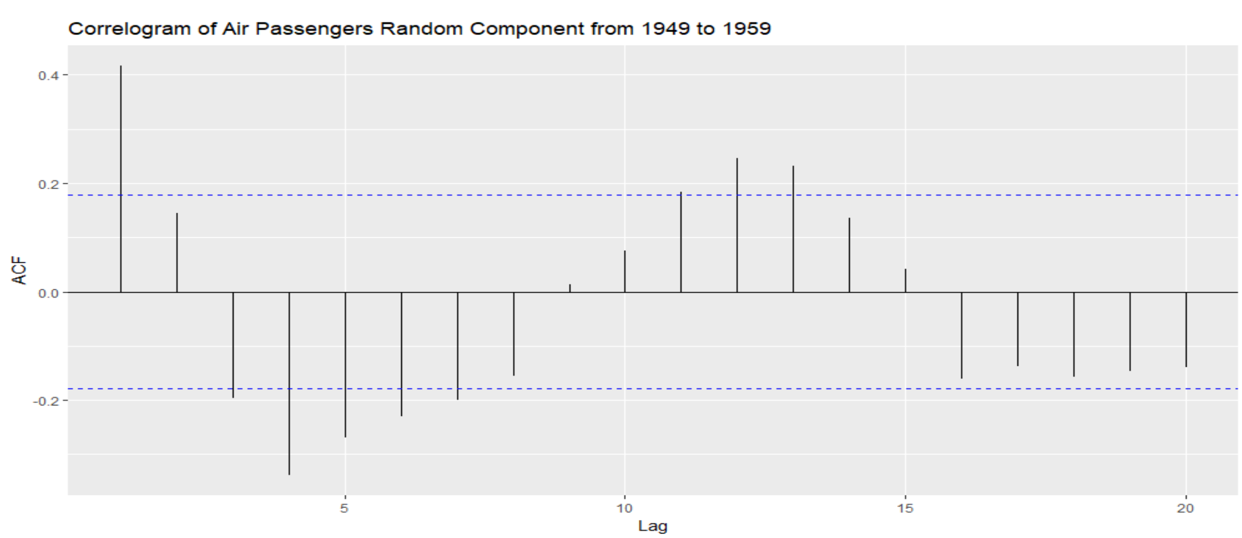
[ACF and PACF plot]



Step 8: Decompose the trend, seasonal and plot ACF again to visualize the good p and q

Graph 49:

[second ACF plot in RStudio]



Step 9: Run auto.arima()

Results obtain ARIMA(1,1,0) and AIC = 899.9

Phi1 = -0.2431

```
omponent from 1949 to 1961 )
> arima <- auto.arima(model)
> arima
Series: model
ARIMA(1,1,0)(0,1,0)[12]

Coefficients:
          ar1
      -0.2431
s.e.   0.0894

sigma^2 estimated as 109.8:  log likelihood=-447.95
AIC=899.9   AICc=900.01   BIC=905.46
```

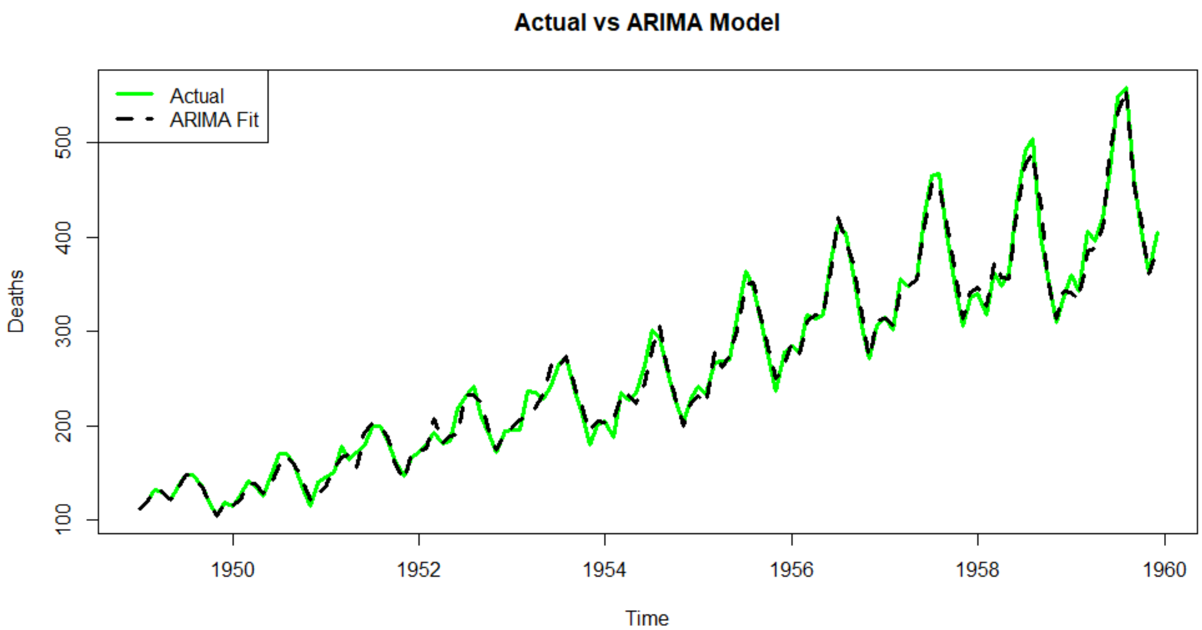
Try different ARMA but AIC are not as good at optimal arima()

ARMA process	ACI
100	1290.38
201	1270.55
007	1241.46
407	1229.81

Step 10: Visualize the arima(p,d,q) process and its predictions

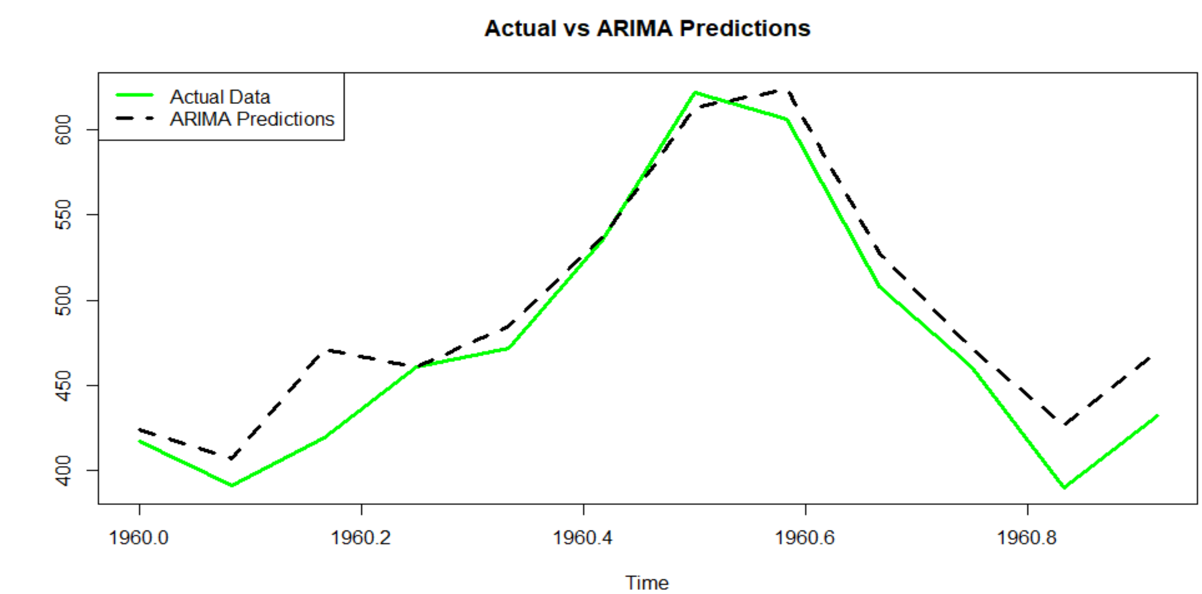
<u>Graph 50:</u>

[Actual data and ARIMA model fitted]



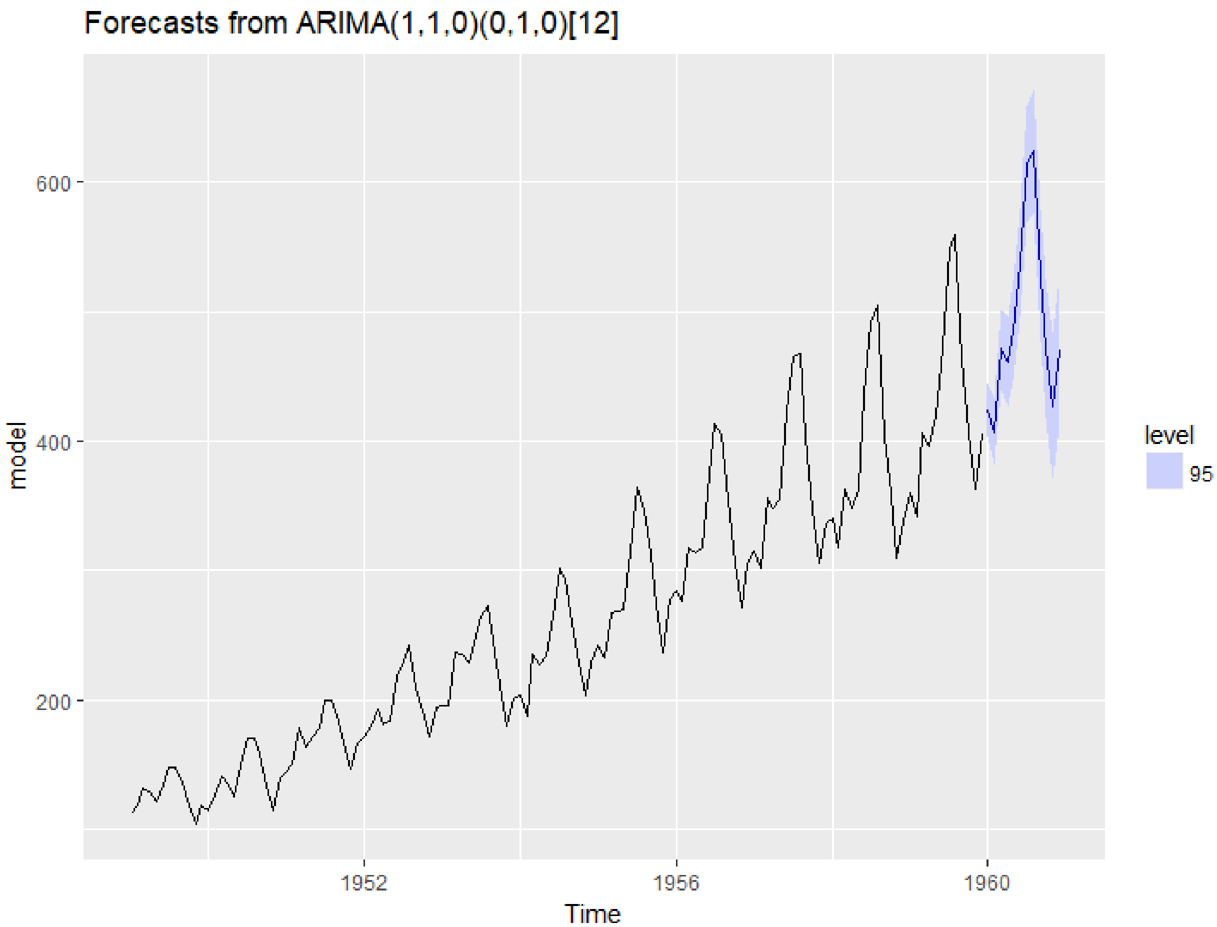
<u>Graph 51:</u>

[Actual data of 1996 and ARIMA predictions]



Graph 52:

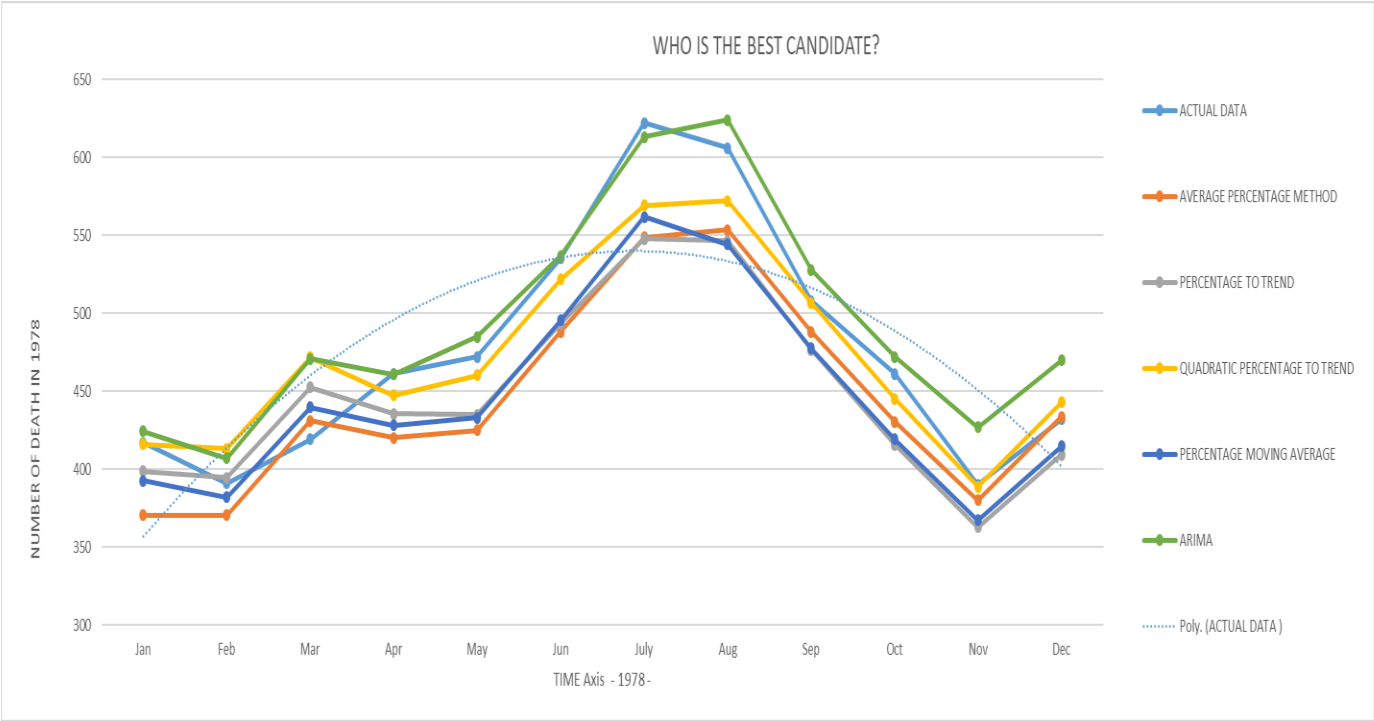
[Actual data and predictions with highest and lowest bounds]



EVALUATED RESULTS OF AIR PASSENGER 1960

Table 1: Compare obtained results to the actual data of 1960

1960	Actual data	Average percentage methods	Percentage to trend methods	Quadratic Percentage to trend	Percentage moving average	ARIMA
Jan	417	370.3	398.5	416.0	392.6	424.1
Feb	391	370.3	394.4	412.9	382.0	407.1
Mar	419	431.0	452.6	471.7	439.5	470.8
Apr	461	420.0	435.5	447.3	428.0	460.9
May	472	425.0	434.7	460.3	432.8	484.9
Jun	535	488.0	493.2	521.5	495.4	536.9
July	622	548.5	547.7	569.1	561.7	612.9
Aug	606	553.3	546.4	572.1	544.1	623.9
Sep	508	488.0	476.6	506.7	477.5	527.9
Oct	461	430.5	415.6	445.2	419.2	471.9
Nov	390	380.0	362.8	388.6	367.1	426.9
Dec	432	433.5	409.3	442.8	414.5	469.9



Based on the three methods applied in this data set: Death in 1946 - 1960
We had observed that the forecasting data computed by the ARIMA (1,1,0) is the most precise and the closest to the actual data of 1960. It is in green and approach closely to blue line, which is initial data in the graph. The second best winner to this challenge is the quadratic percentage moving average, which is in yellow.

References

D.Hamilton, (1994).*Time Series Analysis*. Princenton University Press

J.Brockwell and A.Davis, (2002). *Introduction to Time Series and Forecasting, Second Edition.* Springer.

Autoregressive Moving Average ARMA(p, q) Models for Time Series Analysis - Part 3

https://www.quantstart.com/articles/Autoregressive-Moving-Average-ARMA-p-q-Models-for-Time-Series-Analysis-Part-3

Forecasting: Principle and practice

https://otexts.org/fpp2/arima-r.html

Identifying the numbers of AR or MA terms in an ARIMA model

http://people.duke.edu/~rnau/411arim3.htm

Air Passenger R script

```
#Dataset 3: Air Passenger 1946 -1960
# Huong, LE

#IMPORT THE LIBRARY
library(ggfortify)
library(tseries)
library(forecast)

#IMPORT THE DATA
df = read.csv('AirPassenger_1949.csv',header=F)

#SPLIT INTO TRAINING SET AND TEST SET
trainingset = df[1:132,]
testset = df[133:144,]

#CONVERTING TRAINING SET INTO TIMESERIES
model = ts(trainingset, frequency = 12, start = c(1949,1))

#PLOT THE DATA
plot.ts(model, main = 'Time Series of Air Passenger 1949-1959')
#DECOMPOSE DATA: trend and seasonal
autoplot(decompose(model))

#EXAM ACF AND PACF TO FIND GOOD Q AND P
par(mfcol = c(2,1 ))
acf(model, main="ACF of Number of Air passenger in a year")
acf(model, type="partial", main="PACF of Number of Air passenger in a year")

#DECOMPOSE THE TREND AND SEASONAL
decompose<- decompose(model,"multiplicative")
autoplot(decompose)
x = decompose$random
autoplot(acf(decompose$random[7:126],plot=FALSE))+ labs(title="Correlogram of Air
Passengers Random Component from 1949 to 1959")

#RUN auto.arima()
arima <- auto.arima(model)
arima
ggtsdiag(arima)
```

```
#PREDICTION AND VISUALIZATION
forecast <- forecast(arima, level = c(95), h = 12)
forecast
autoplot(forecast)
preds <- predict(arima, n.ahead = 12)
preds

#VISUALIZE TESTSET AND PREDICTION
ts.plot(testset,preds$pred, main = 'Actual vs ARIMA Predictions',
        col = c('green','black'), lty = c(1,2), lwd = c(3,3))
legend('topleft',legend = c('Actual Data','ARIMA Predictions'),
       col = c('green','black'), lwd = c(3,3), lty = c(1,2))

#VISUALIZE THE MODEL AND ARIMA MODEL
plot(model, col = 'green', main = 'Actual vs ARIMA Model',
     ylab = 'Deaths', lwd = 3)
lines(arima$fitted, col = 'black', lwd = 3, lty = 2)
legend('topleft',legend = c('Actual','ARIMA Fit'),
       col = c('green','black'), lwd = c(3,3), lty = c(1,2))
```