Times series forecasting

#load the packages

```
library(forecast)
```

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

```
library(ggplot2)
```

#set Working Directory

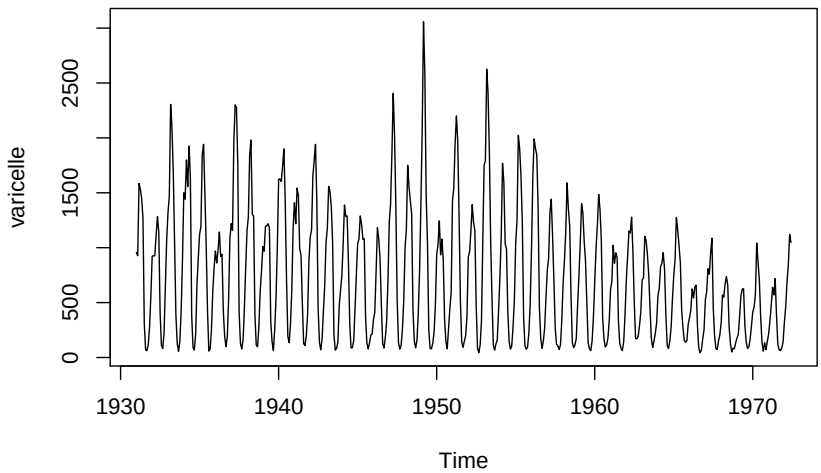
```
getwd()
```

```
## [1] "/Users/dabasn/Desktop/Nitesh/Studies/DSTI/TimeSeries/TSLabs/Day1/Varicella_files"
```

```
setwd("/Users/dabasn/Desktop/Nitesh/Studies/DSTI/TimeSeries/TSLabs/Day1/Varicella_files")
```

#load the data and plot them

```
data=read.csv('varicelle.csv', header = TRUE)
varicelle<-ts(data$x,start=c(1931,1),end=c(1972,6),freq=12)
plot(varicelle)
```



It seems to be a seasonal pattern. We can check this with the seasonplot

The mean is given by

```
mean(varicelle)
```

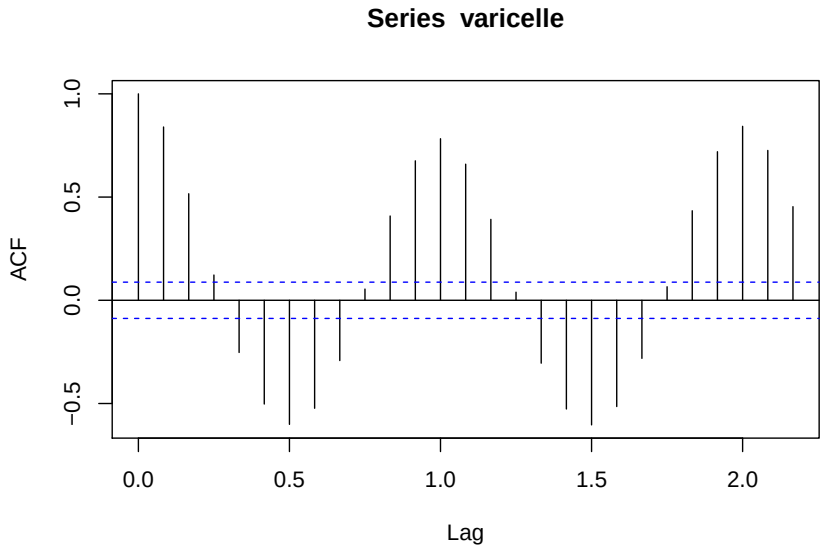
```
## [1] 732.4076
```

And the auto-correlation mean that there is a seasonal pattern in the data

```
tmp=acf(varicelle,type="cor",plot = FALSE)
tmp$acf[1:3,1,1]
```

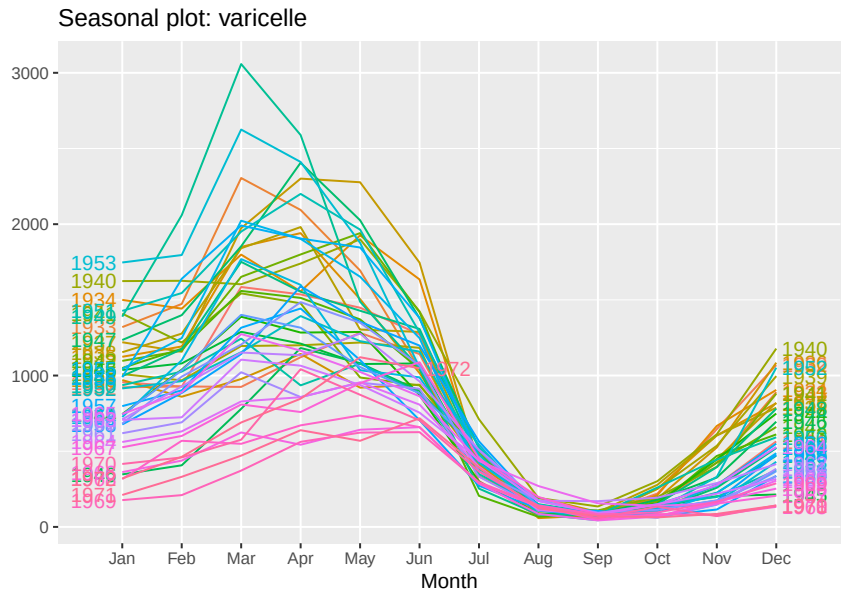
```
## [1] 1.0000000 0.8394105 0.5160841
```

```
plot(tmp)
```



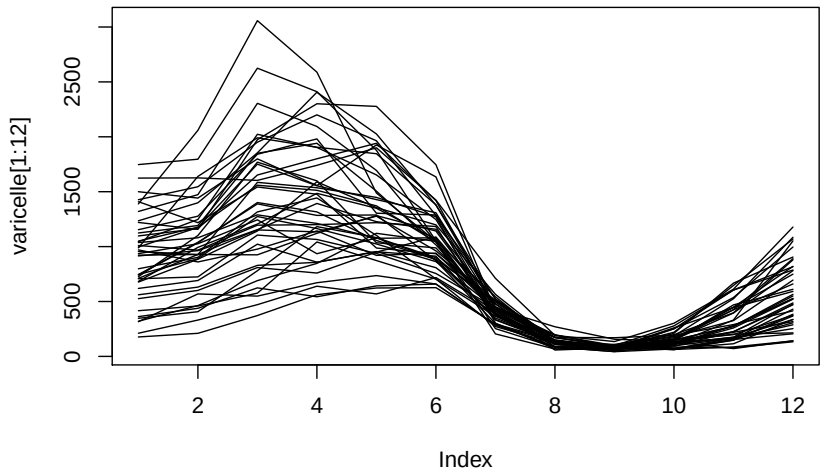
What is confirmed by the seasonal plot

```
ggseasonplot(varicelle,year.labels= TRUE,year.labels.left=TRUE)
```



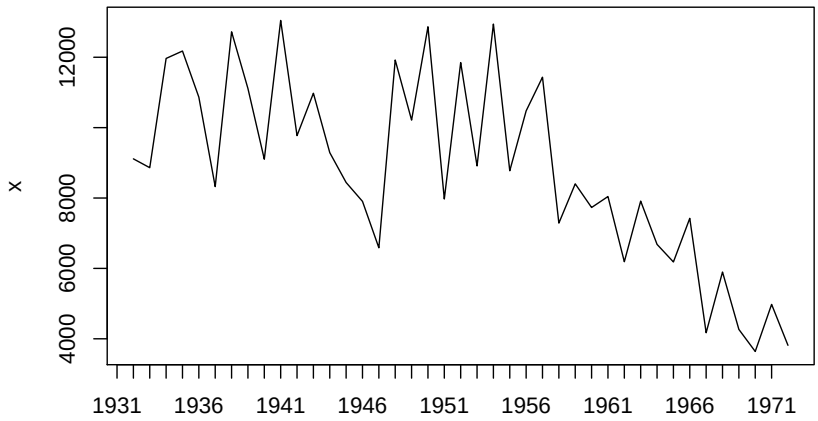
We can also plot manually the seasonal plot

```
plot(varicelle[1:12],type="l",ylim=c(min(varicelle),max(varicelle)))
for (i in 1:41) lines(varicelle[(1+12*i):(12*(i+1))])
```



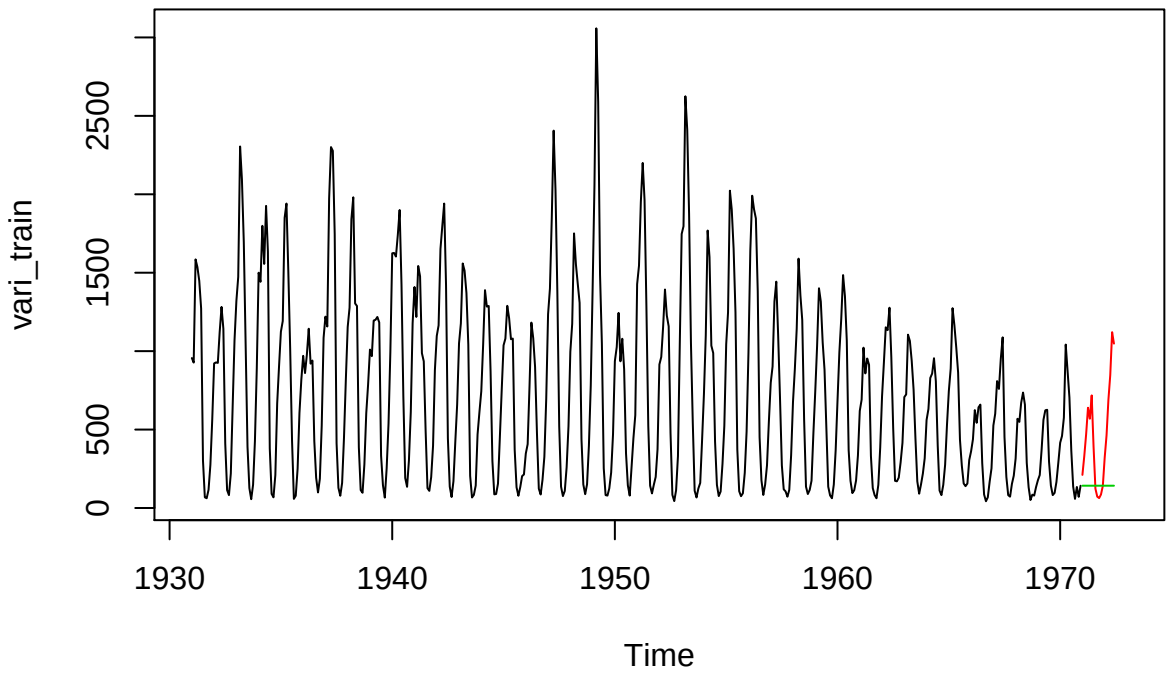
We now compute and plot the annual evolution - To removed the seasonality (Only trend plot)

```
x=rep(0,41)
for (i in 0:40) x[i+1]<-sum(varicelle[(1+12*i):(12*(i+1))])
plot(x,type='l',xaxt='n',xlab='')
axis(1,at = 0:40,labels = 1931:1971)
```



Split dataset into training and testing sets and forcast. Red - Actual data Green - Predicted data

```
vari_train <- ts(data$x[1:480],start=c(1931,1),end=c(1970,12),freq=12)
vari_test <- ts(data$x[481:498],start=c(1971,1),end=c(1972,6),freq=12)
plot(vari_train,xlim=c(1931,1973))
lines(vari_test,col=2)
SES=HoltWinters(vari_train,alpha=NULL,beta=FALSE,gamma=FALSE)
p1<-predict(SES,n.ahead=18)
lines(p1,col=3)
```

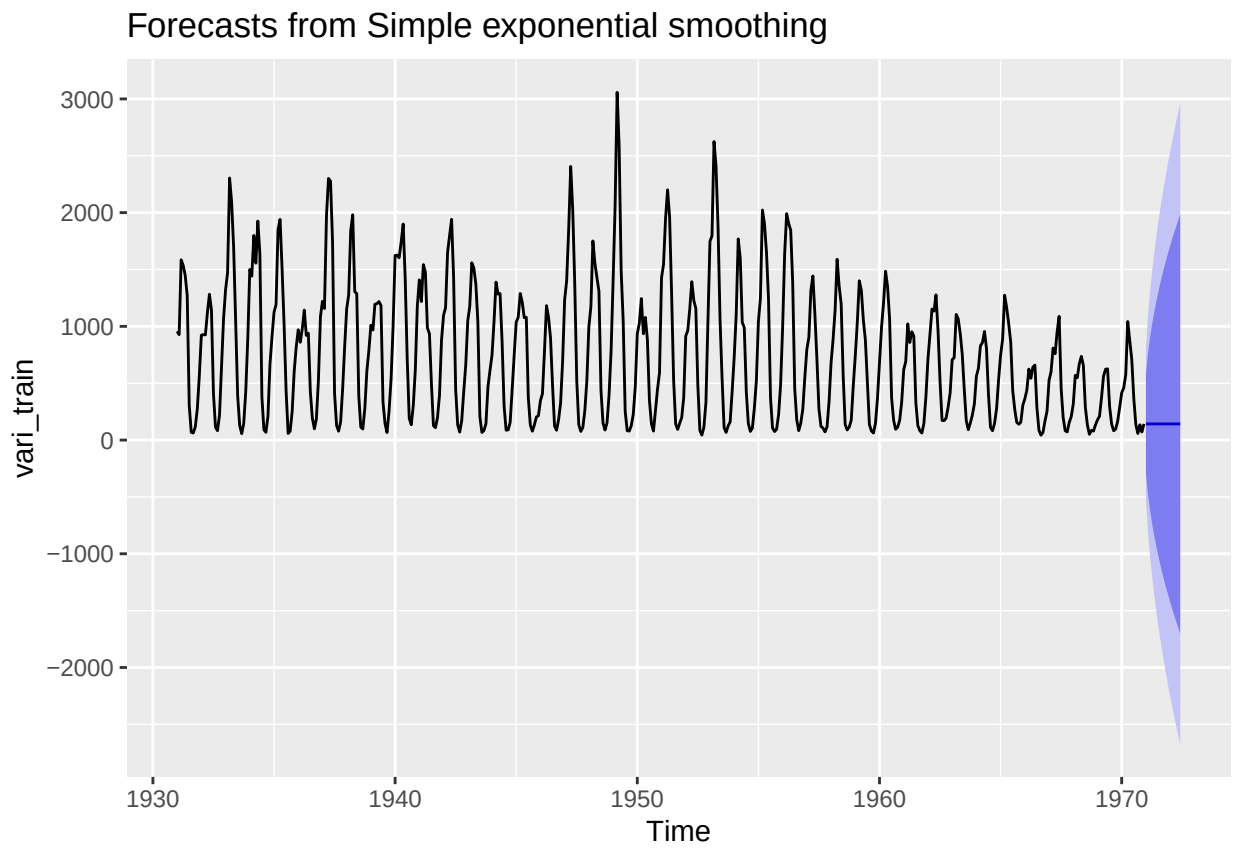


Forecasting with SES

```
SES=ses(vari_train,h=18)
round(accuracy(SES),2)
```

```
##                    ME    RMSE     MAE    MPE MAPE MASE ACF1
## Training set -1.6 338.15 251.29 -24.53   61 1.04 0.51
```

```
autoplot(SES)
```

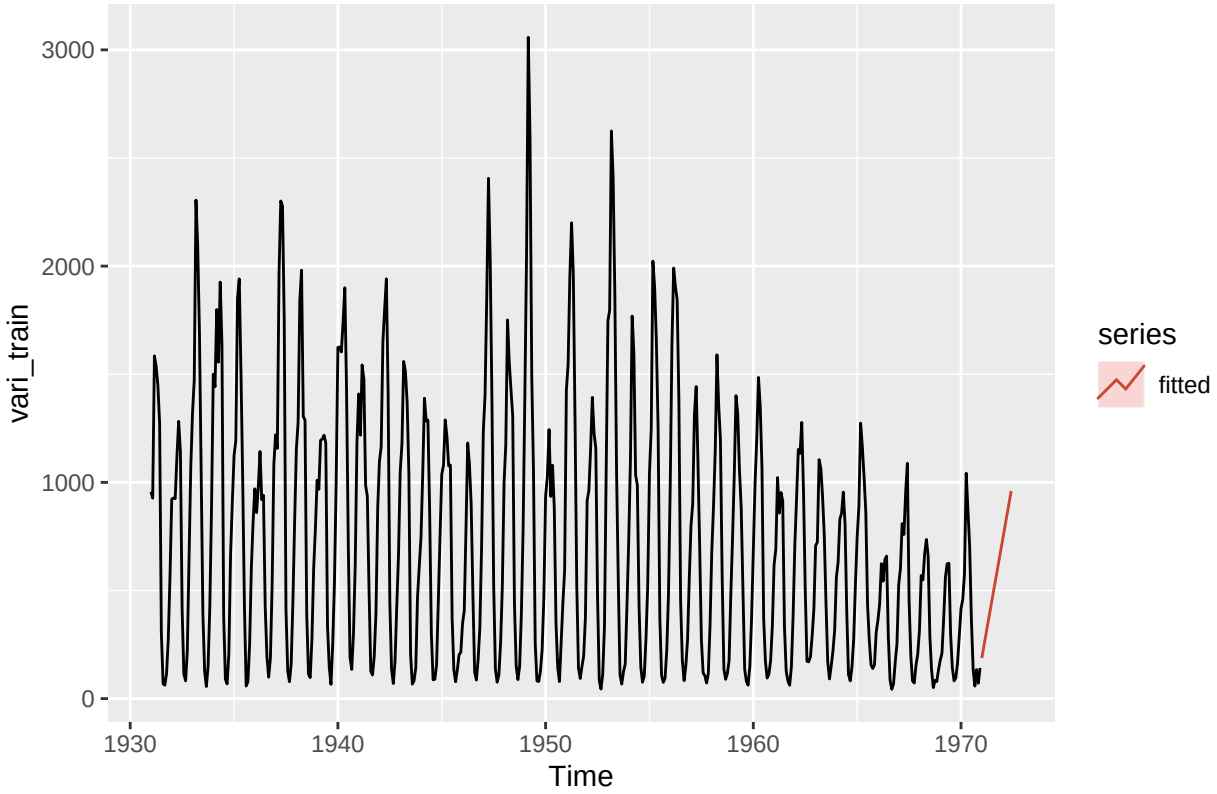


#Varicella forecasting with non seasonal HW smoothing

```
HOLT=holt(vari_train,h=18)
round(accuracy(HOLT),2)
```

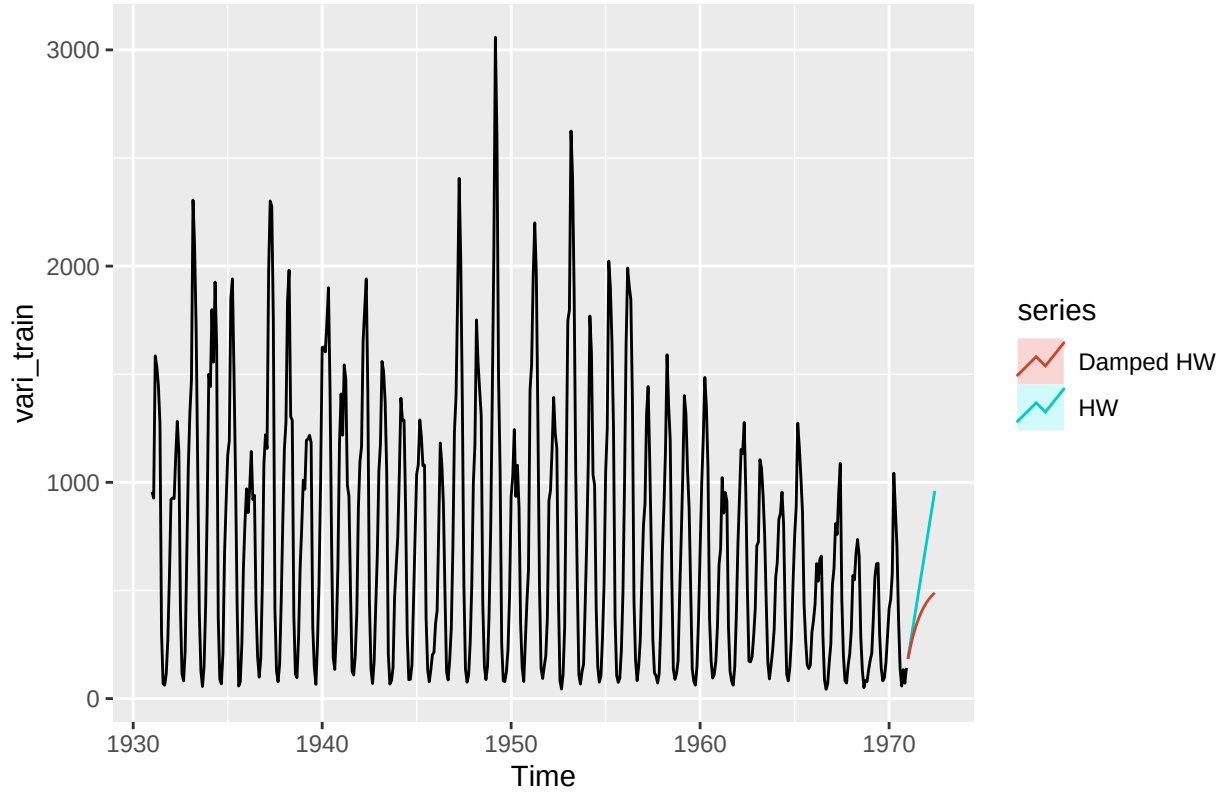
```
##                    ME   RMSE    MAE   MPE  MAPE MASE ACF1
## Training set 0.22 330.42 259.65 54.15 89.34 1.08 0.01
```

```
autoplot(vari_train) + autolayer(HOLT,series='fitted',PI=FALSE) #PI=FALSE remove the prediction interva
```



#Varicella forecasting with non seasonal HW smoothing; 0 < phi < 1

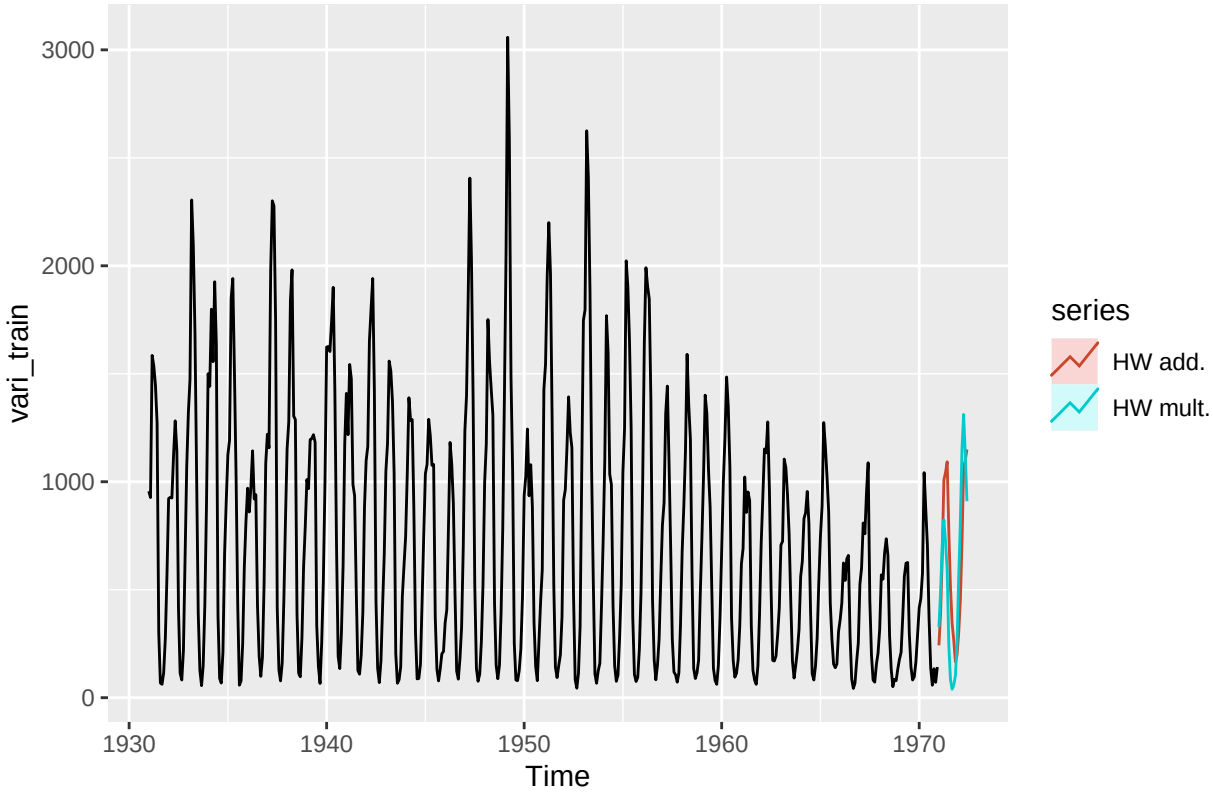
```
HOLT1=holt(vari_train,h=18)
HOLT2=holt(vari_train,damped=TRUE,phi=0.9,h=18)
autoplot(vari_train) + autolayer(HOLT1,series='HW',PI=FALSE) + autolayer(HOLT2,series='Damped HW',PI=FA
```



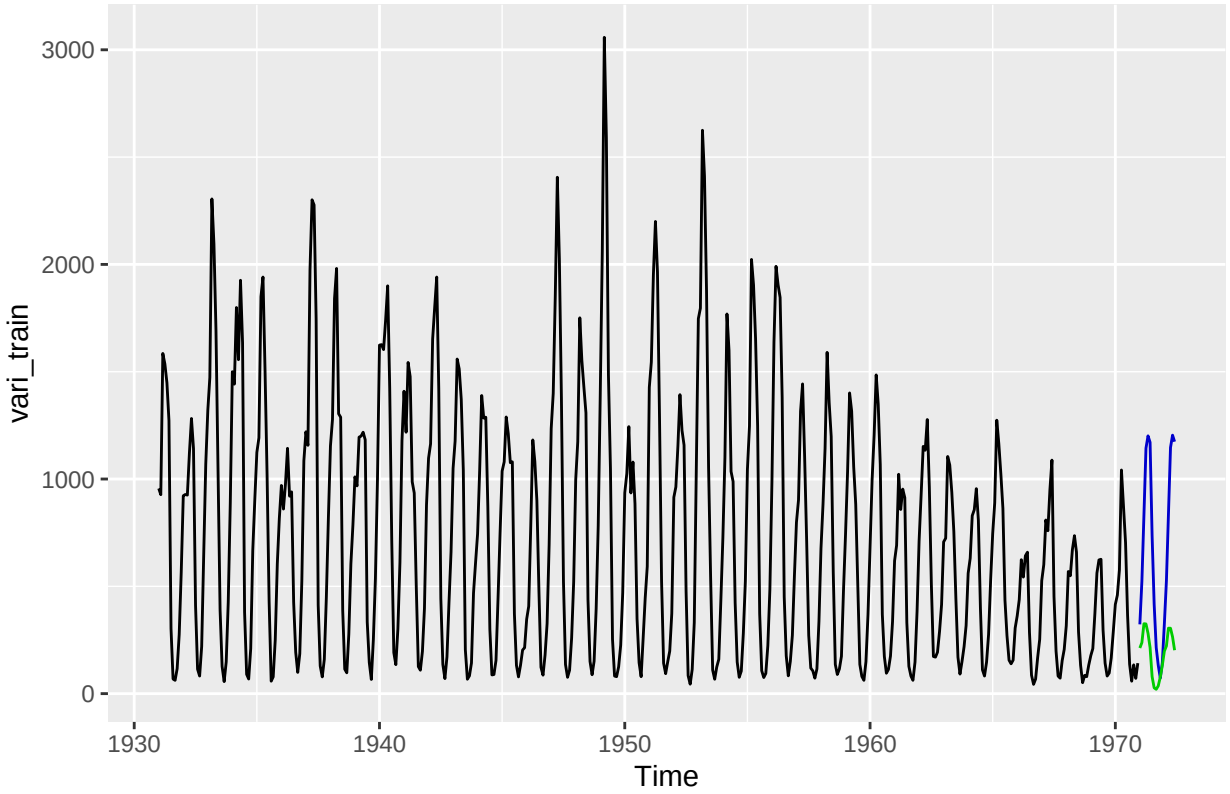
#Varicella forecasting with seasonal HW smoothing - Additive OR Multiplicative

```
fit1=hw(vari_train,seasonal='additive',h=18)
fit2=hw(vari_train,seasonal='multiplicative',h=18)
fit3=hw(vari_train,seasonal='additive',damped=TRUE,h=18)
fit4=hw(vari_train,seasonal='multiplicative',damped=TRUE,h=18)

autoplot(vari_train) +
  autolayer(fit1,series='HW add.',PI=FALSE) +
  autolayer(fit2,series='HW mult.',PI=FALSE)
```




```
autoplot(vari_train) +
  autolayer(fit3,series='HW mult.',PI=FALSE, col='blue') +
  autolayer(fit4,series='HW mult.',PI=FALSE, col='green')
```



#Additive/Multiplicative + Damped versions #damped version of the seasonal HW

```
#Model Comparision
print(sqrt(mean((fit1$mean-vari_test)^2)))
```

```
## [1] 238.2674
```

```
print(sqrt(mean((fit2$mean-vari_test)^2))) #best model
```

```
## [1] 214.7901
```

```
print(sqrt(mean((fit3$mean-vari_test)^2)))
```

```
## [1] 279.7424
```

```
print(sqrt(mean((fit4$mean-vari_test)^2)))
```

```
## [1] 375.6358
```