# Forecasting Competition

Biz Yoder & Ryan McCord

2023-04-24

**Link to GitHub repository:** https://github.com/biz-yoder/biz-yoder-MCCORDYODER_ENV790_TSA_Competition_S2023.git

```
#Import load data
load <- read_excel(path="./Data/load.xlsx")
```

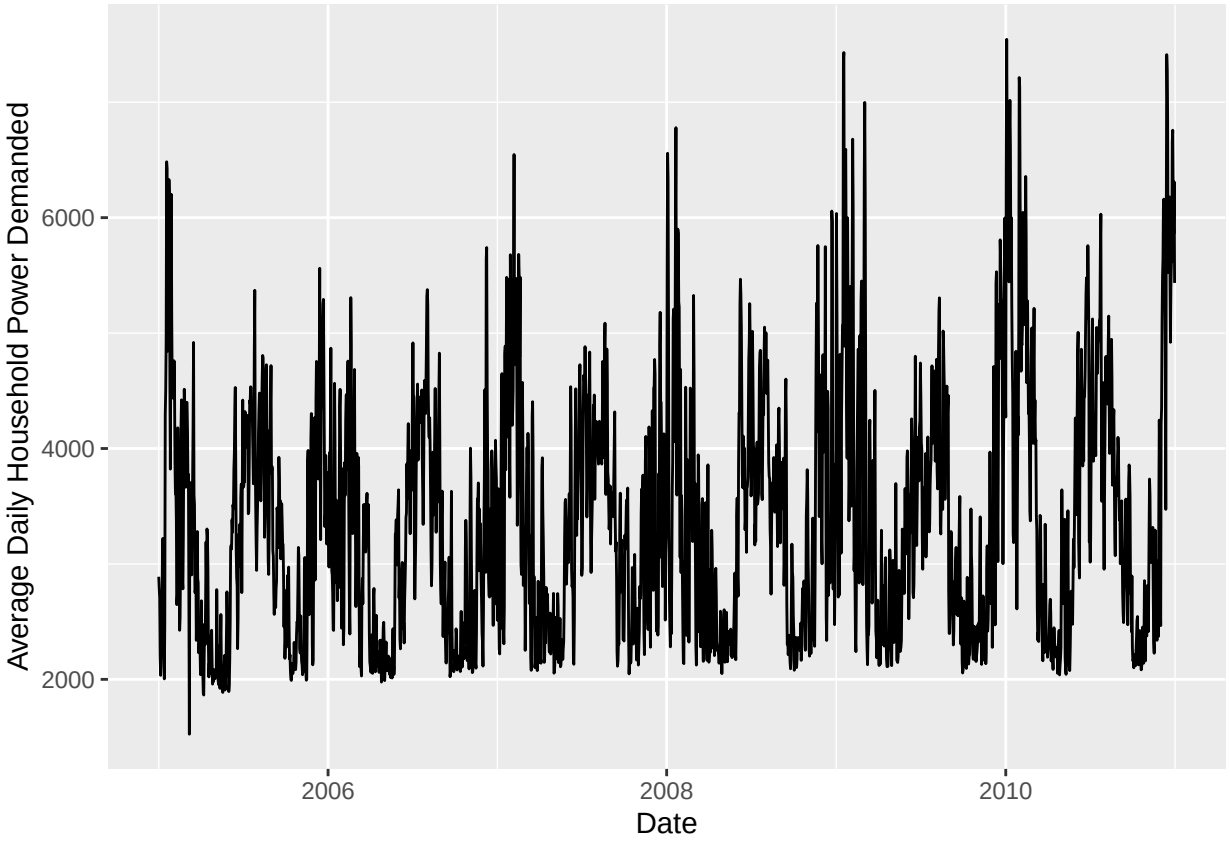
```
#Find daily load (average of hourly values)
load_daily <- load %>%
  mutate(daily_load = rowMeans(select(load, starts_with('h')), na.rm = TRUE))
```

```
#Construct time series object
ts_daily <- msts(load_daily$daily_load, start = c(2005, 01, 01), end = c(2010, 12, 31),seasonal.periods
```

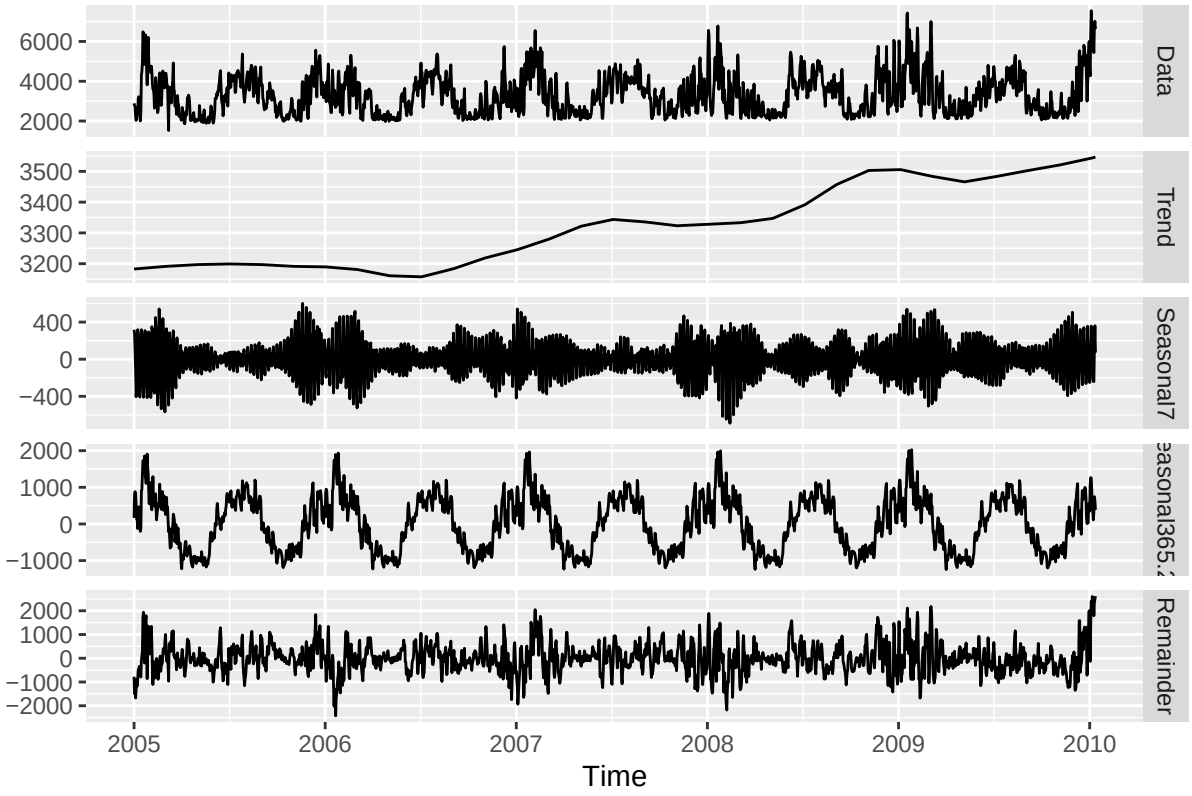
```
#Visualize data
summary(load_daily$daily_load)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1525    2453    3220    3382    4046    7545
```

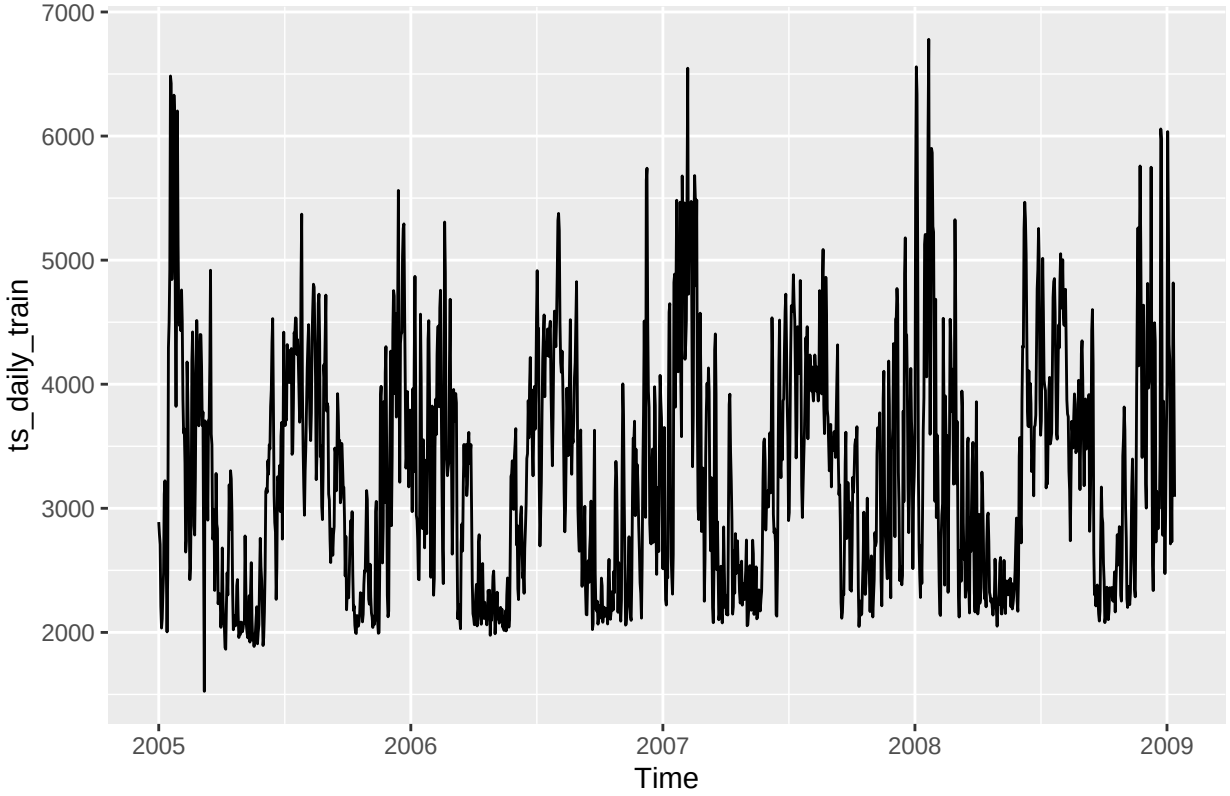
```
ggplot(load_daily, aes(x=load_daily$date,y=load_daily$daily_load)) +
  geom_line() +
  xlab("Date") +
  ylab("Average Daily Household Power Demanded")
```



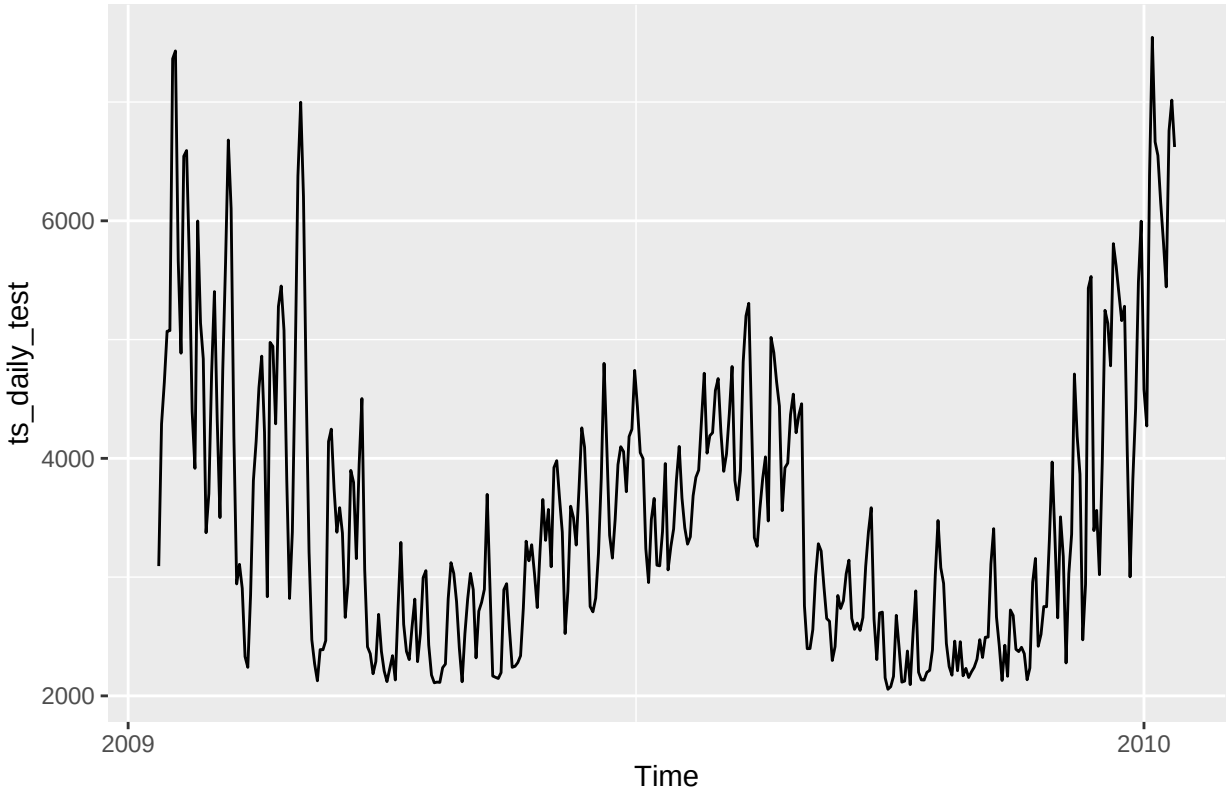

```
ts_daily %>% mstl() %>%
  autoplot()
```



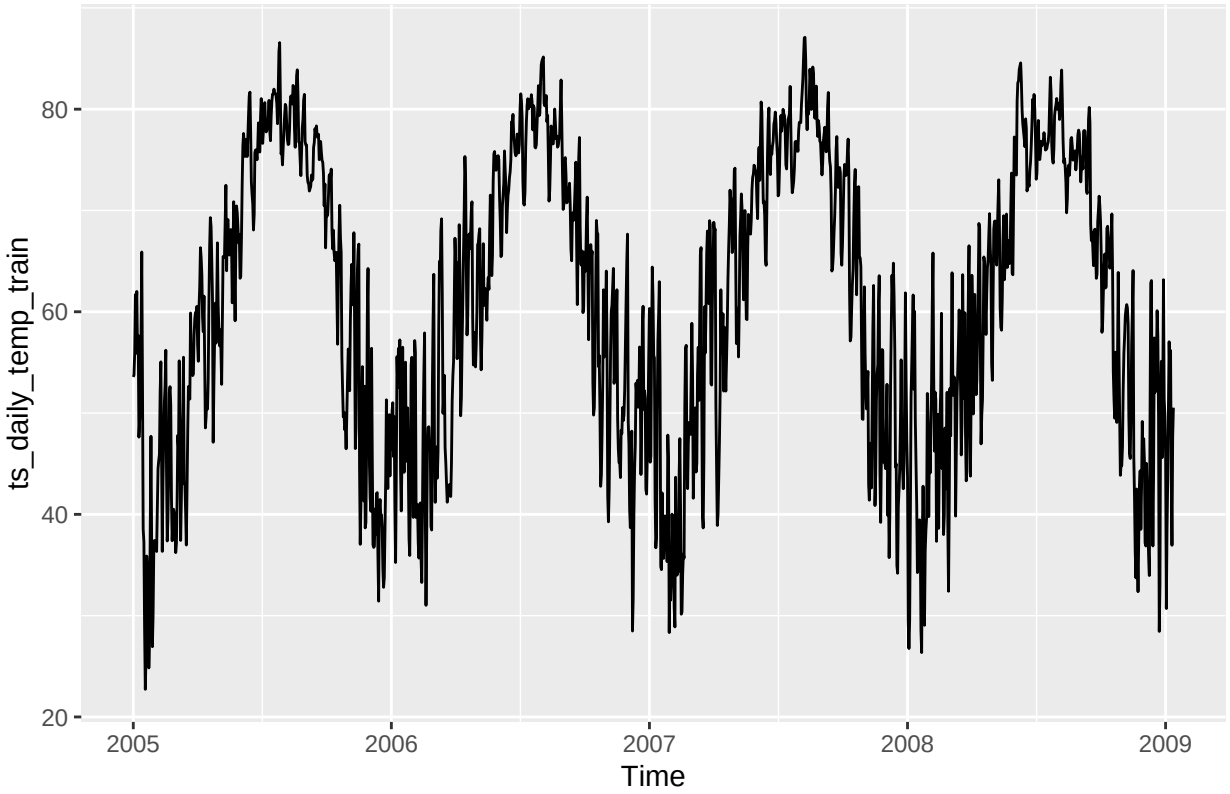

```
#Create train and test sets
one_year = 365 #leave out one year of data for testing
##Create train ts
ts_daily_train <- subset(ts_daily, end = length(ts_daily)-one_year)


##Create test ts
ts_daily_test <- subset(ts_daily, start = length(ts_daily)-one_year)

autoplot(ts_daily_train)
```



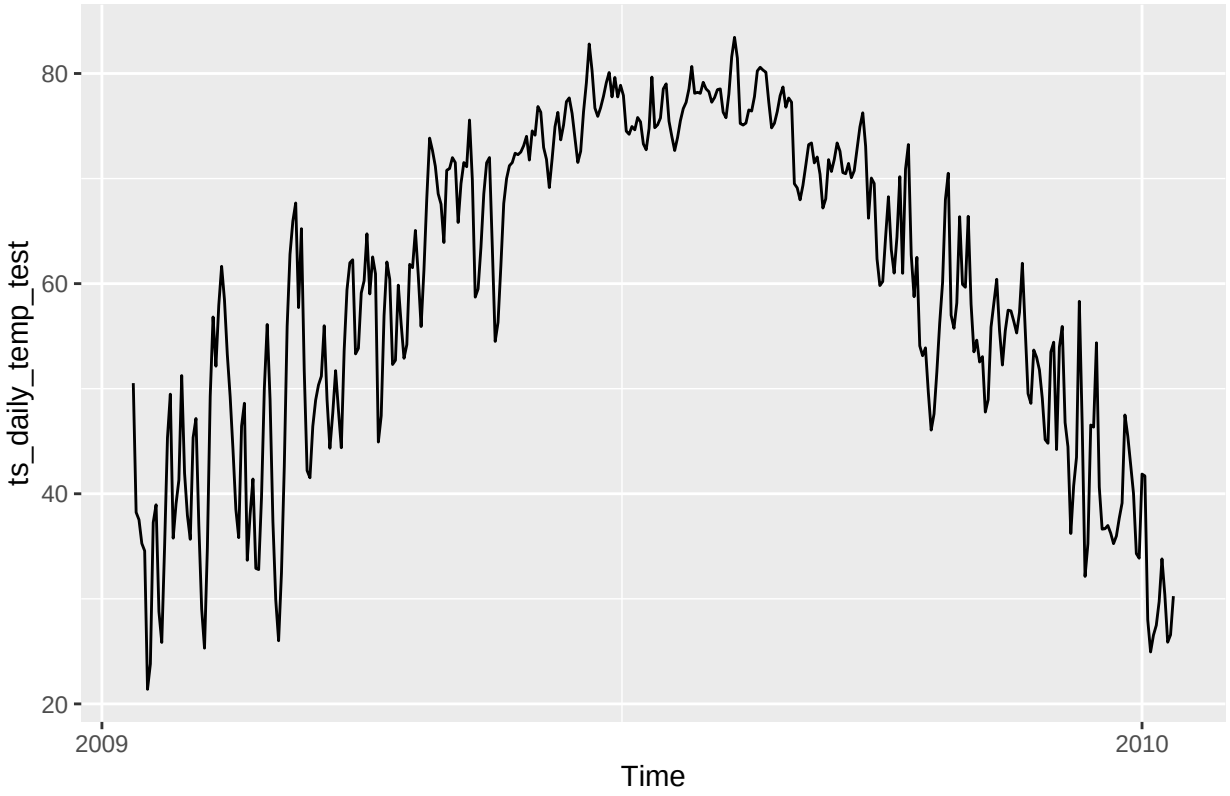

```
autoplot(ts_daily_test)
```



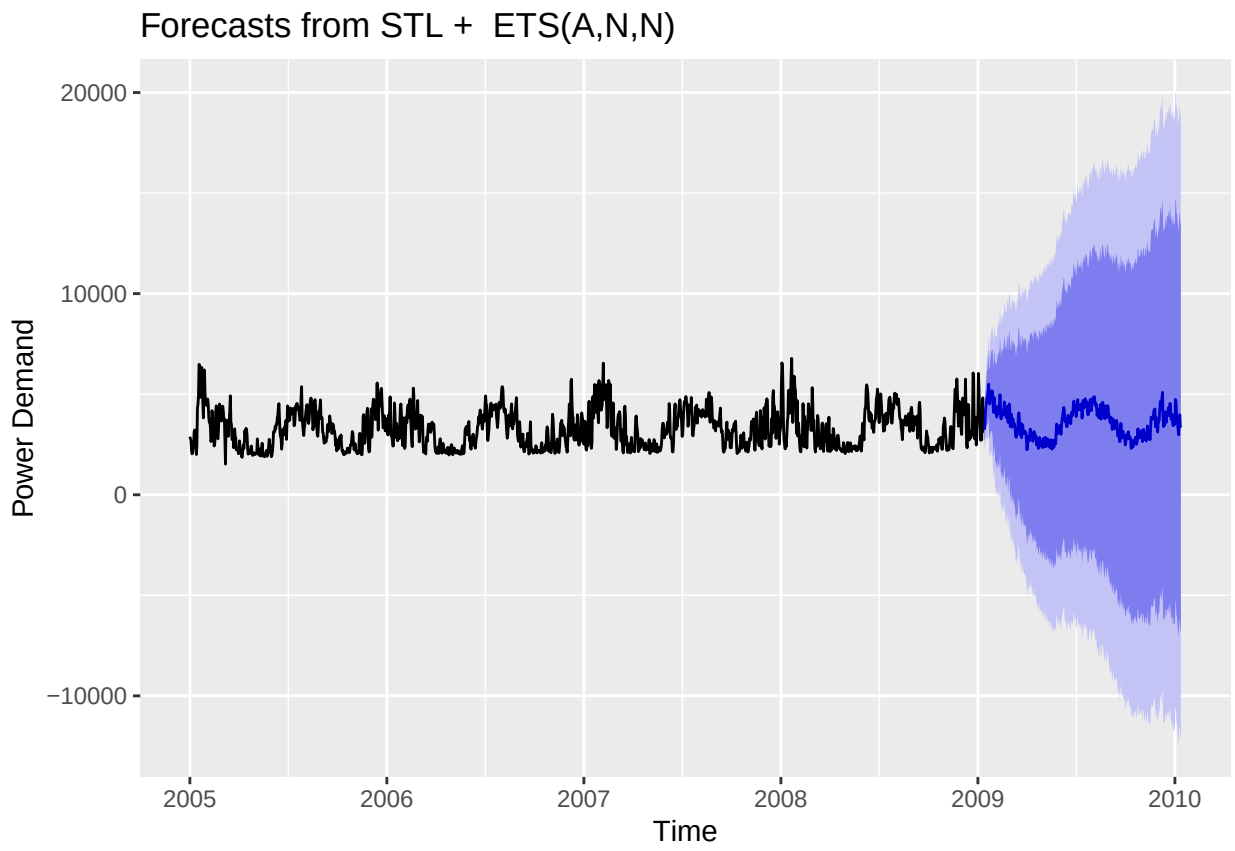

```
#Create time series for external regressors - TEMPERATURE
temp <- read_excel(path="./Data/temperature.xlsx")
temp_daily <- temp %>%
  mutate(temp_daily = rowMeans(select(temp, starts_with('t')), na.rm = TRUE))

temp_daily <- cbind.data.frame(temp_daily$date, temp_daily$hr, temp_daily$temp_daily)

temp_daily_wide <- reshape(temp_daily, idvar = "temp_daily$date", timevar = "temp_daily$hr", direction =

temp_daily_avg <- temp_daily_wide %>%
  mutate(temp_daily_avg = rowMeans(select(temp_daily_wide, starts_with('temp_daily$temp')), na.rm = TRU

temp_daily_avg <- cbind.data.frame(temp_daily_avg$`temp_daily$date`, temp_daily_avg$temp_daily_avg)
names(temp_daily_avg) <- c("date", "temp_daily")

#Construct time series object
ts_daily_temp <- msts(temp_daily_avg$temp_daily, start = c(2005, 01, 01), end = c(2010, 12, 31),seasonal

#Create train and test sets
ts_daily_temp_train <- subset(ts_daily_temp, end = length(ts_daily_temp)-one_year)


##Create test ts
ts_daily_temp_test <- subset(ts_daily_temp, start = length(ts_daily_temp)-one_year)

autoplot(ts_daily_temp_train)
```



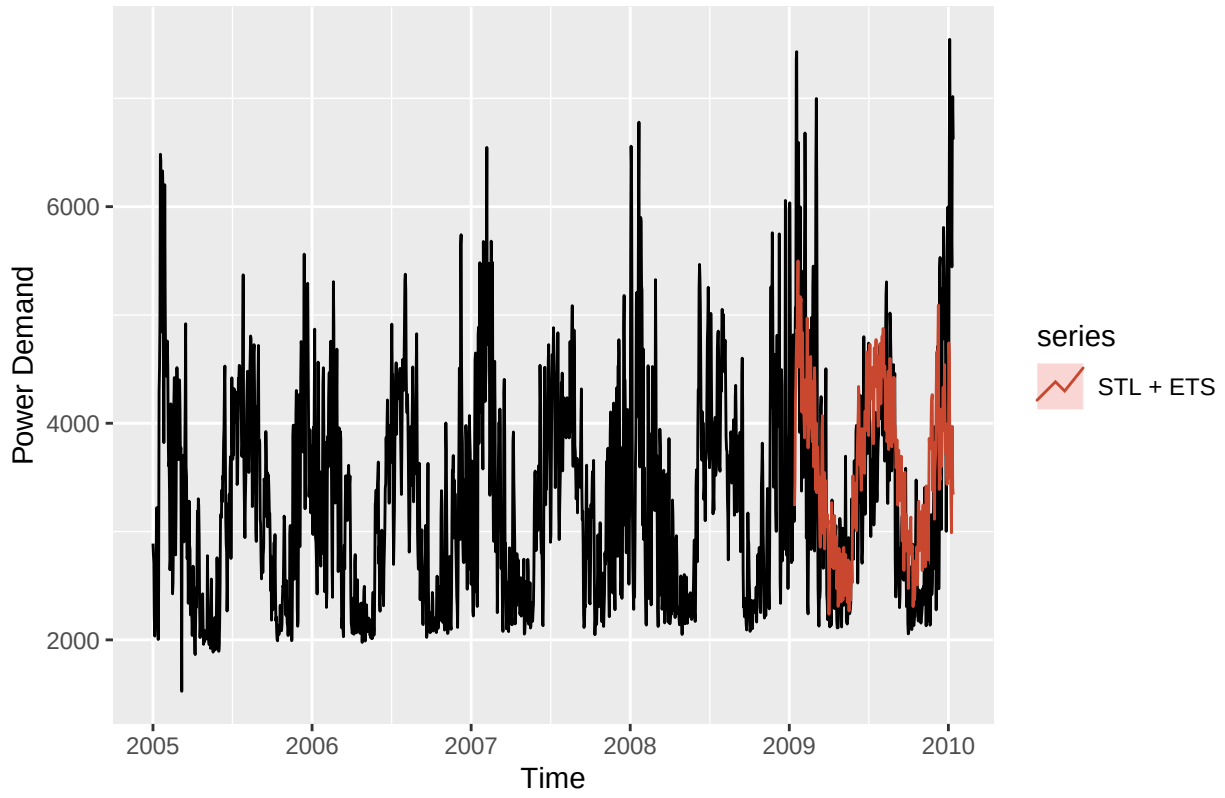

```
autoplot(ts_daily_temp_test)
```




```
#Model 1 - STL + ETS model
ETS_fit <-  stlf(ts_daily_train,h=365)

#Plot foresting results
autoplot(ETS_fit) + ylab("Power Demand")
```

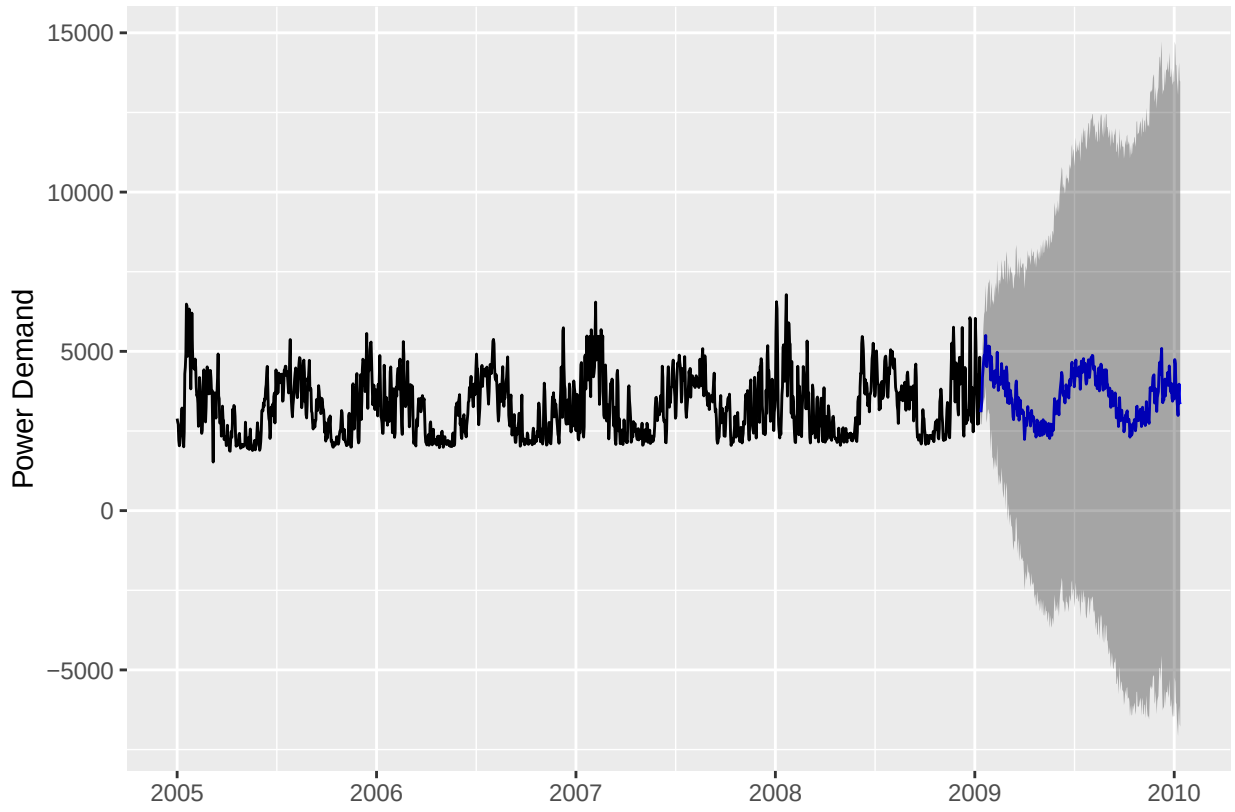



```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(ETS_fit, series="STL + ETS",PI=FALSE) +
  ylab("Power Demand")
```

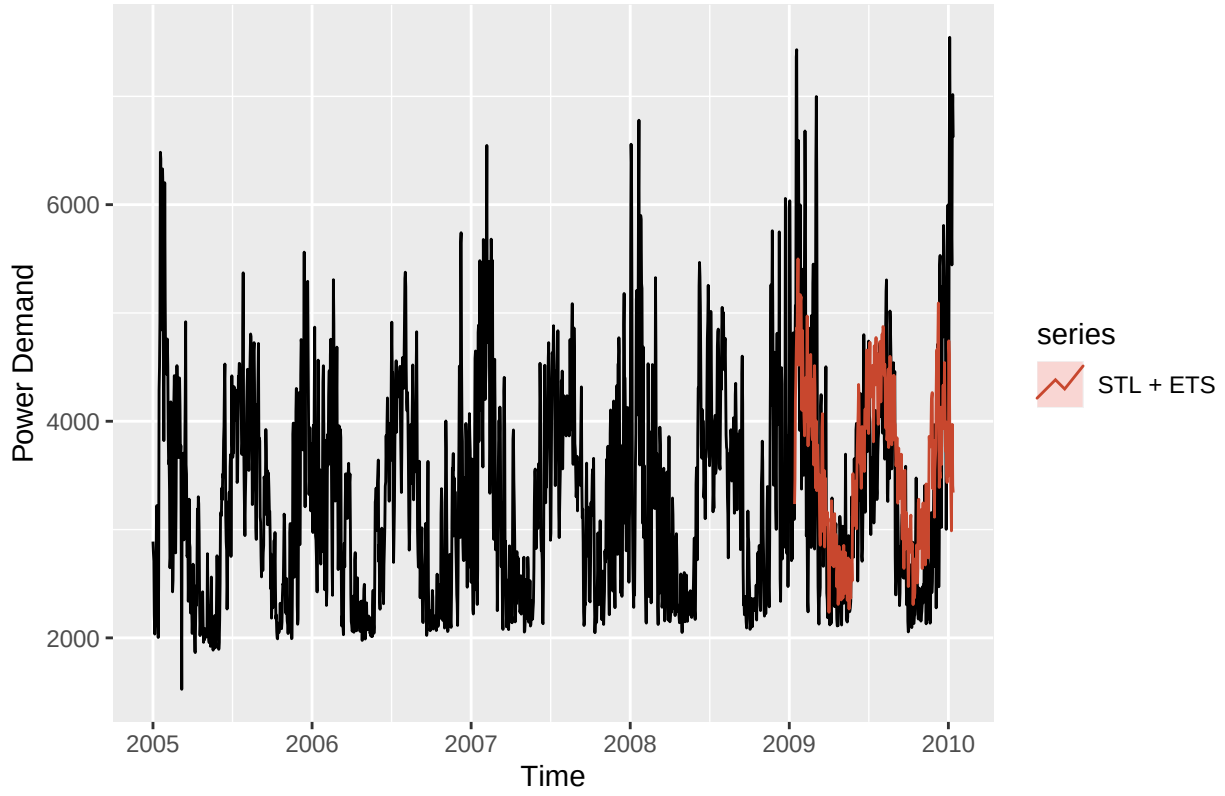



```
#Run on full dataset to upload to Kaggle
ETS_fit_full <- stlf(ts_daily, h=59)
library(ggfortify)
model1_ets <- fortify(ETS_fit_full)
model1_ets_data <- model1_ets[1838:1896,4]
date <- seq(as.Date("2011-01-01"), as.Date("2011-02-28"), by = "days")

submission1 <- cbind.data.frame(date, model1_ets_data)
names(submission1) <- c("date", "load")

write.csv(submission1, "./Submissions/submission1.csv", row.names = FALSE)
```

```
#Model 1b - STL + ETS model
ETS_fit2 <-  stlf(ts_daily_train, allow.multiplicative.trend=TRUE, h=365)

#Plot foresting results
autoplot(ETS_fit2) + ylab("Power Demand")
```



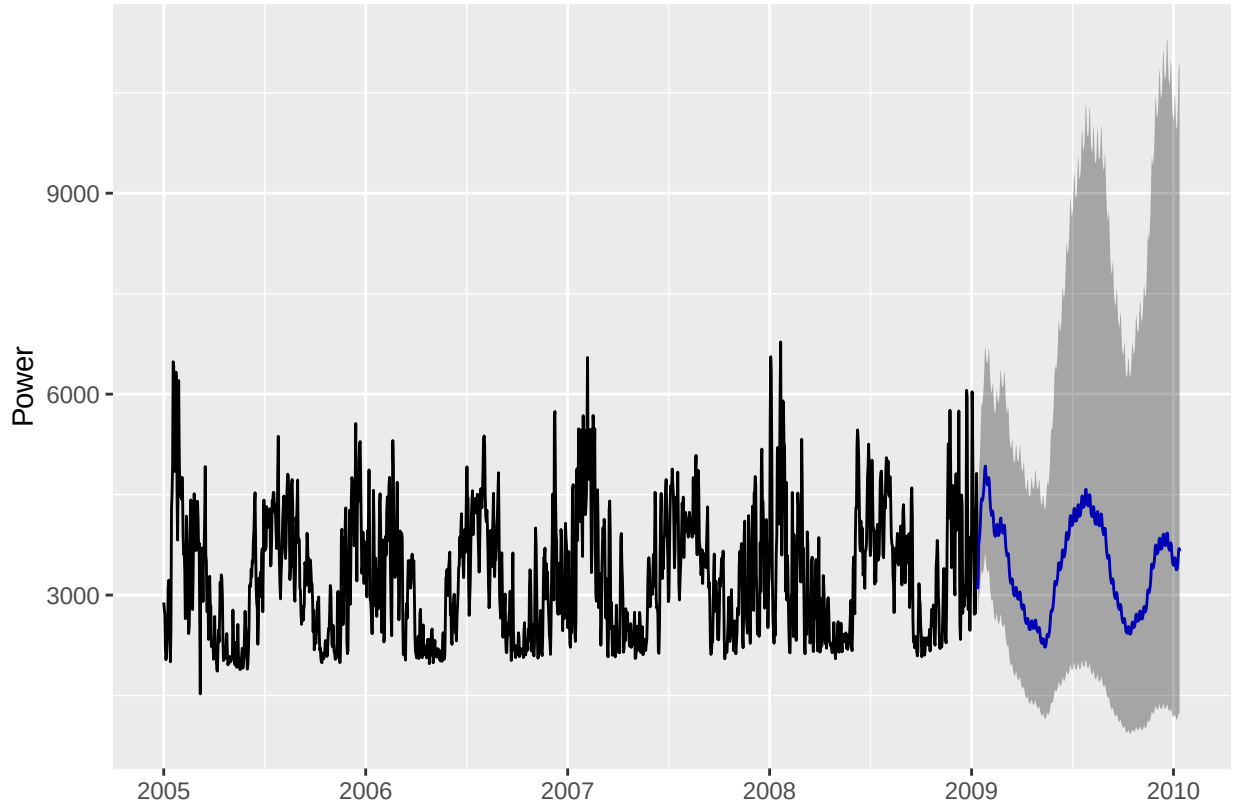

```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(ETS_fit2, series="STL + ETS",PI=FALSE) +
  ylab("Power Demand")
```



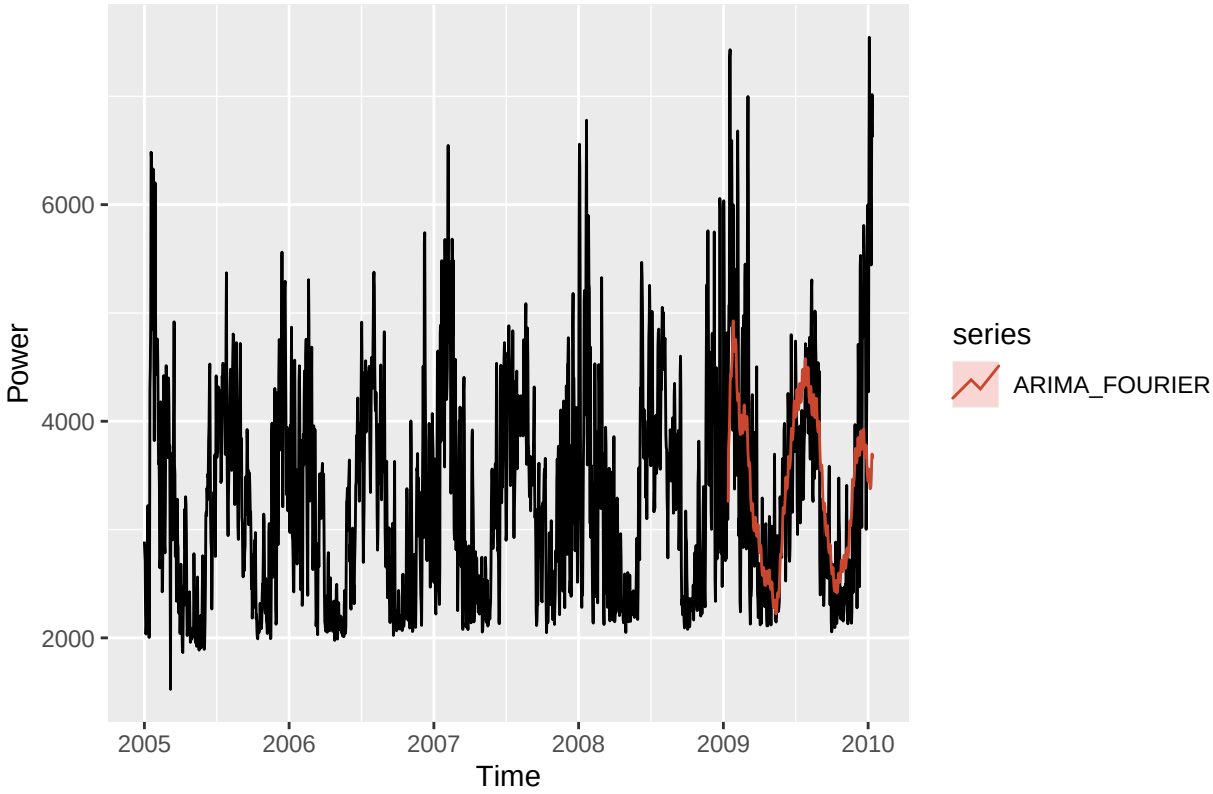

```
#Run on full dataset to upload to Kaggle
ETS_fit_full2 <- stlf(ts_daily, h=59)
model1b_ets <- fortify(ETS_fit_full2)
model1b_ets_data <- model1b_ets[1838:1896,4]

submission1b <- cbind.data.frame(date, model1b_ets_data)
names(submission1b) <- c("date", "load")

write.csv(submission1b, "./Submissions/submission1b.csv", row.names = FALSE)
```

```
#Model 2 - ARIMA with fourier terms
ARIMA_four_fit <- auto.arima(ts_daily_train,
                             seasonal=FALSE,
                             lambda=0,
                             xreg=fourier(ts_daily_train,
                                          K=c(2,12))
                             )

ARIMA_four_for <- forecast(ARIMA_four_fit,
                           xreg=fourier(ts_daily_train,
                                        K=c(2,12),
                                        h=365),
                           h=365
                           )

#Plot foresting results
```

```
autoplot(ARIMA_four_for) + ylab("Power")
```




```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(ARIMA_four_for, series="ARIMA_FOURIER",PI=FALSE) +
  ylab("Power")
```

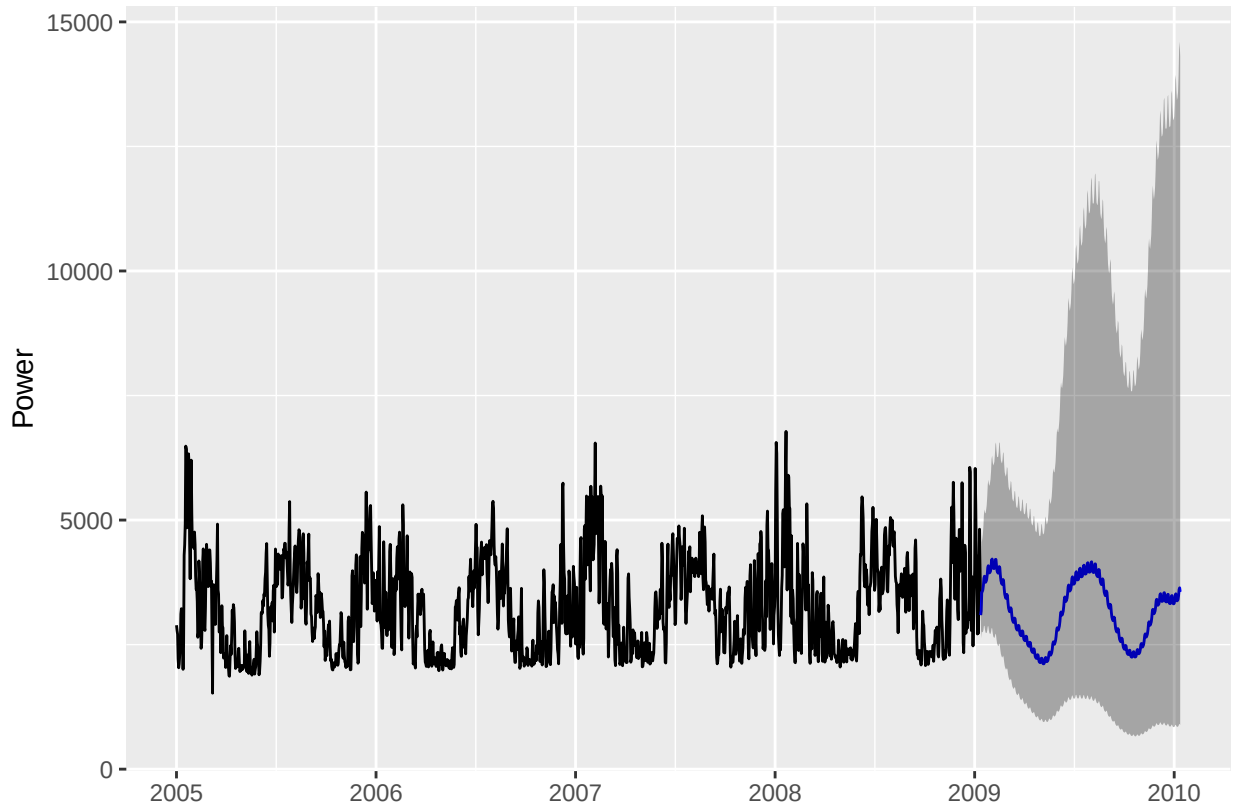



```
ARIMA_four_full <- auto.arima(ts_daily,
                              seasonal=FALSE,
                              lambda=0,
                              xreg=fourier(ts_daily,
                                           K=c(2,12))
                              )

ARIMA_forecast_full <- forecast(ARIMA_four_full,
                            xreg=fourier(ts_daily,
                                         K=c(2,12),
                                         h=59),
                            h=59
                            )
```

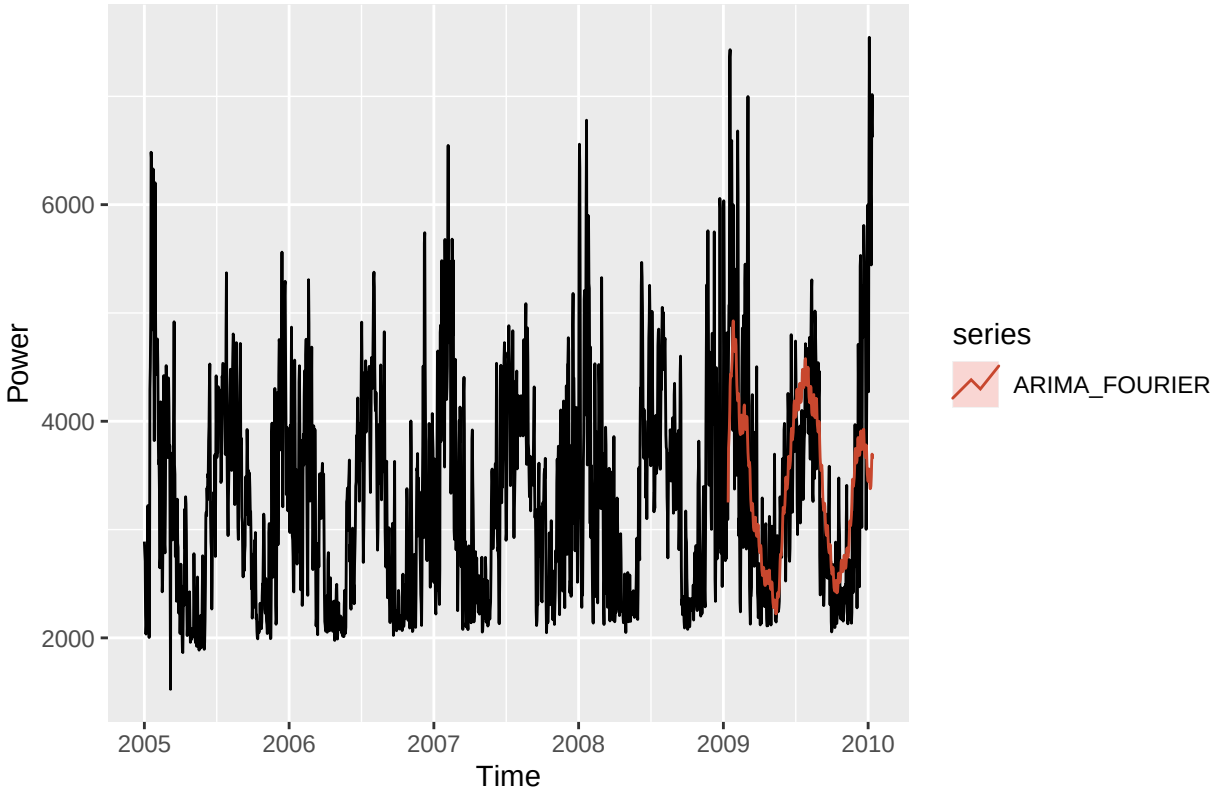
```
model2 <- fortify(ARIMA_forecast_full)
model2_data <- model2[1838:1896,4]

submission2 <- cbind.data.frame(date, model2_data)
names(submission2) <- c("date", "load")

write.csv(submission2, "./Submissions/submission2.csv", row.names = FALSE)
```

```
#Model 2 - ARIMA with fourier terms
ARIMA_four_fit_2b <- auto.arima(ts_daily_train,
                              seasonal=FALSE,
                              lambda=0,
                              xreg=fourier(ts_daily_train,
```

```
                                        K=c(2,6))
                            )

ARIMA_four_for_2b <- forecast(ARIMA_four_fit_2b,
                          xreg=fourier(ts_daily_train,
                                       K=c(2,6),
                                       h=365), #telling it you also need a forecast for the fourier ex
                          h=365
                          )

#Plot foresting results
autoplot(ARIMA_four_for_2b) + ylab("Power")
```



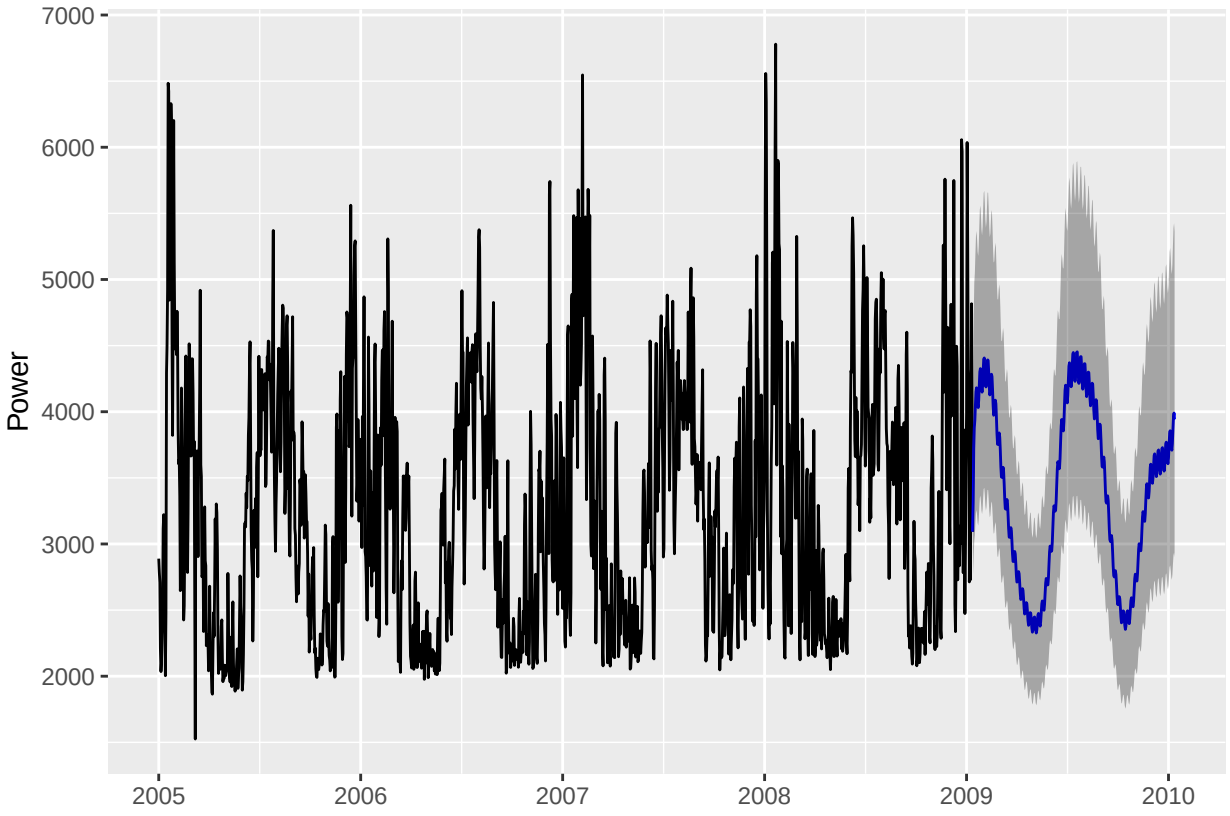

```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(ARIMA_four_for, series="ARIMA_FOURIER",PI=FALSE) +
  ylab("Power")
```



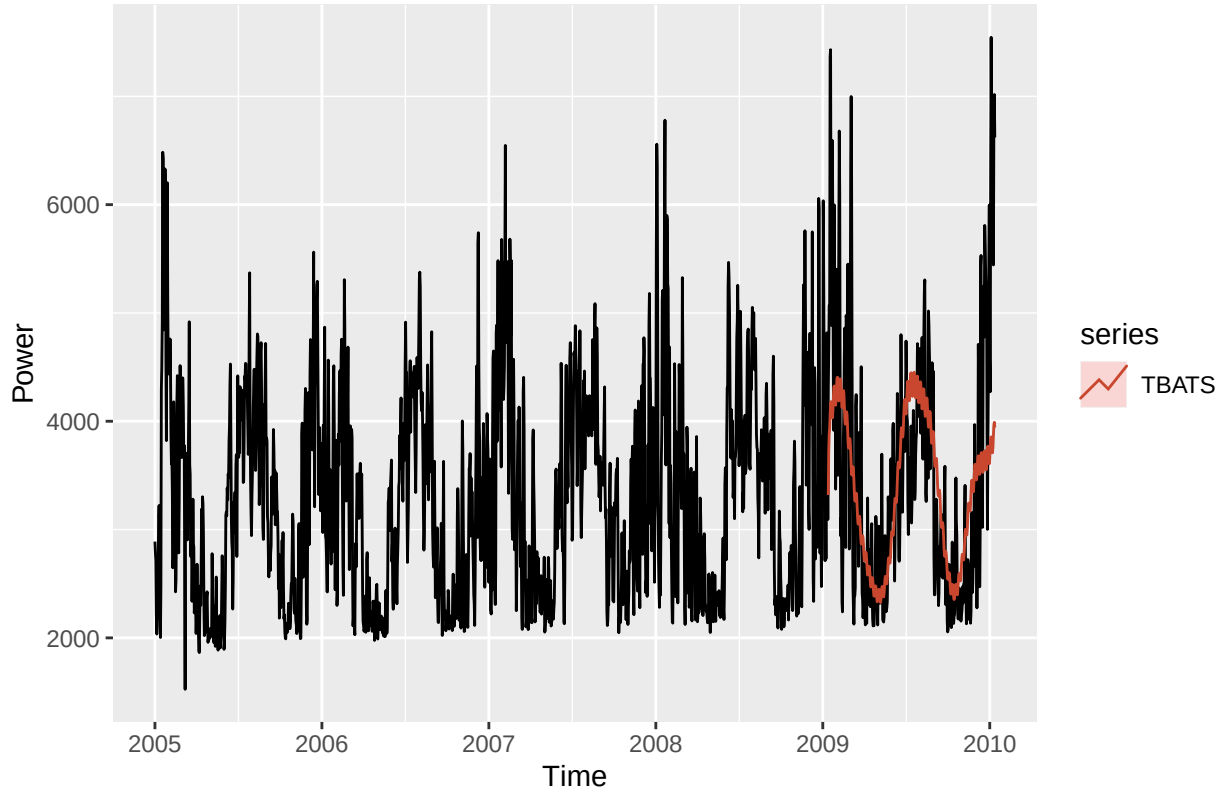

```
ARIMA_four_fit_full_2b <- auto.arima(ts_daily,
                             seasonal=FALSE,
                             lambda=0,
                             xreg=fourier(ts_daily,
                                          K=c(2,6))
                             )

ARIMA_forecast_full_2b <- forecast(ARIMA_four_fit_full_2b,
                            xreg=fourier(ts_daily,
                                         K=c(2,6),
                                         h=59), #telling it you also need a forecast for the fourier ext
                            h=59
                            )

model2b <- fortify(ARIMA_forecast_full_2b)
model2b_data <- model2b[1838:1896,4]

submission2b <- cbind.data.frame(date, model2b_data)
names(submission2b) <- c("date", "load")

write.csv(submission2b, "./Submissions/submission2b.csv", row.names = FALSE)
```

```
# TBATS train
TBATS_fit <- tbats(ts_daily_train)

TBATS_for <- forecast(TBATS_fit, h=365)
```

```
#Plot foresting results
autoplot(TBATS_for) + ylab("Power")
```

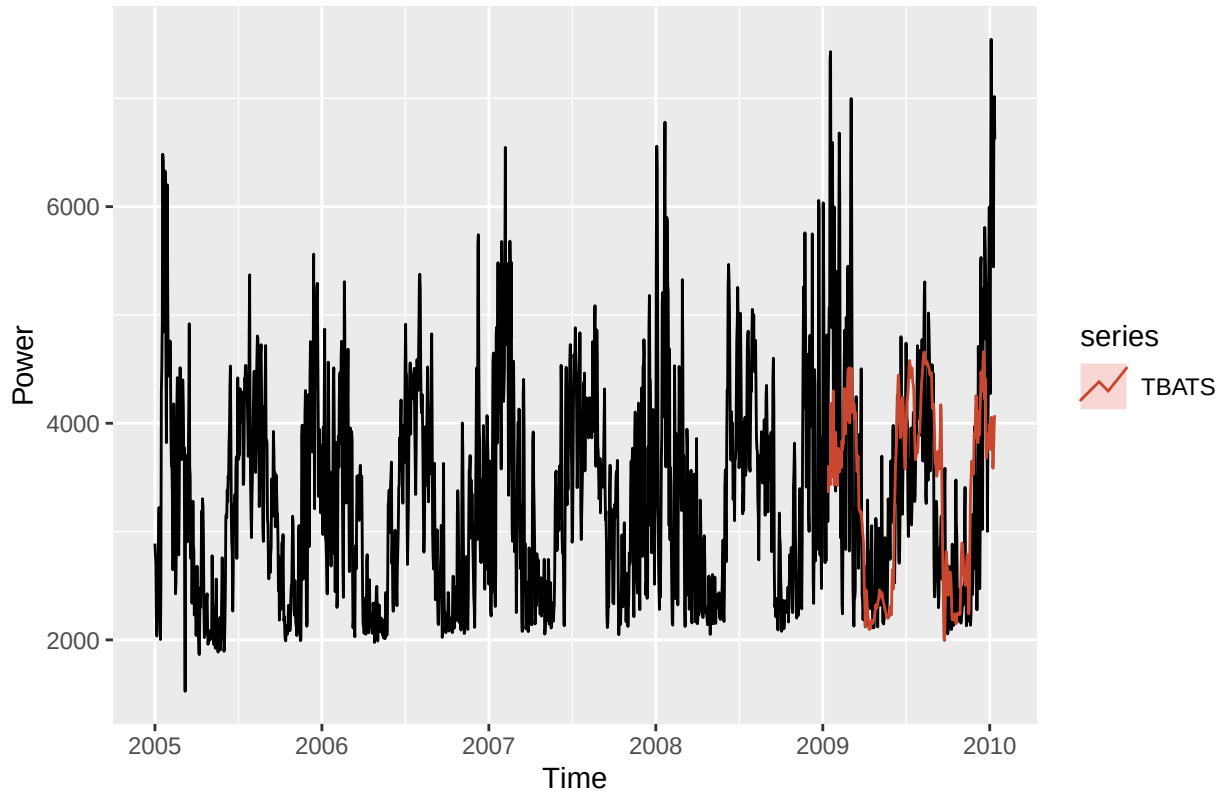



```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(TBATS_for, series="TBATS",PI=FALSE) +
  ylab("Power")
```



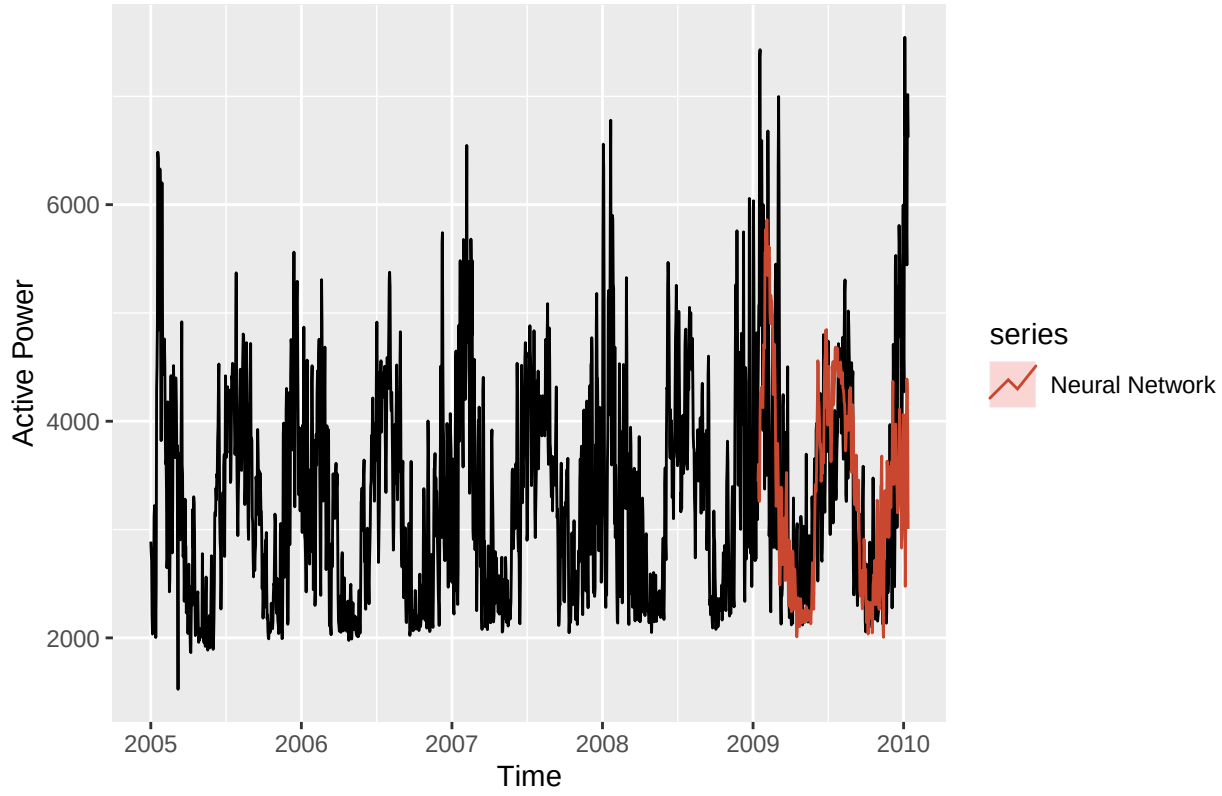

```
#Full dataset
TBATS_fit_full <- tbats(ts_daily)
TBATS_for_full <- forecast(TBATS_fit, h=59)

model3 <- fortify(TBATS_for_full)
model3_data <- model3[1838:1896,4]

submission3 <- cbind.data.frame(date, model3_data)
names(submission3) <- c("date", "load")

write.csv(submission3, "./Submissions/submission3.csv", row.names = FALSE)
```

```
#NN_fit <- nnetar(ts_act_power_daily_train,p=1,P=1)
NN_fit <- nnetar(ts_daily_train)
NN_for <- forecast(NN_fit, h=365)

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for, series="TBATS",PI=FALSE) +
  ylab("Power")
```



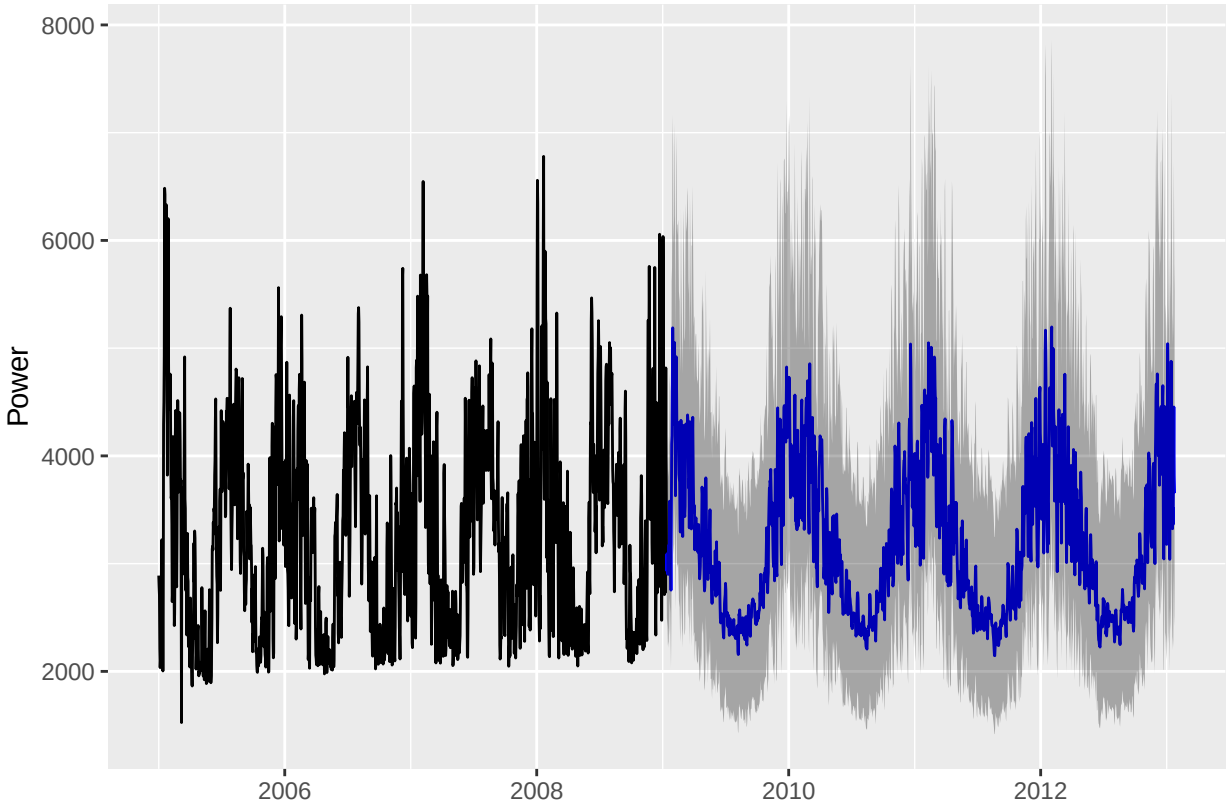

```
#Full dataset
NN_fit_full <- nnetar(ts_daily)
NN_for_full <- forecast(NN_fit_full, h=59)

model4 <- fortify(NN_for_full)
model4_data <- model4[1838:1896,4]

submission4 <- cbind.data.frame(date, model4_data)
names(submission4) <- c("date", "load")

write.csv(submission4, "./Submissions/submission4.csv", row.names = FALSE)
```

```
#Neural networks with fourier regressors -- use better fit from earlier models
NN_fit_4b <- nnetar(ts_daily_train,xreg=fourier(ts_daily_train, K=c(2,6)))

NN_for_4b <- forecast(NN_fit_4b, h=365,xreg=fourier(ts_daily_train, K=c(2,6),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_4b, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```

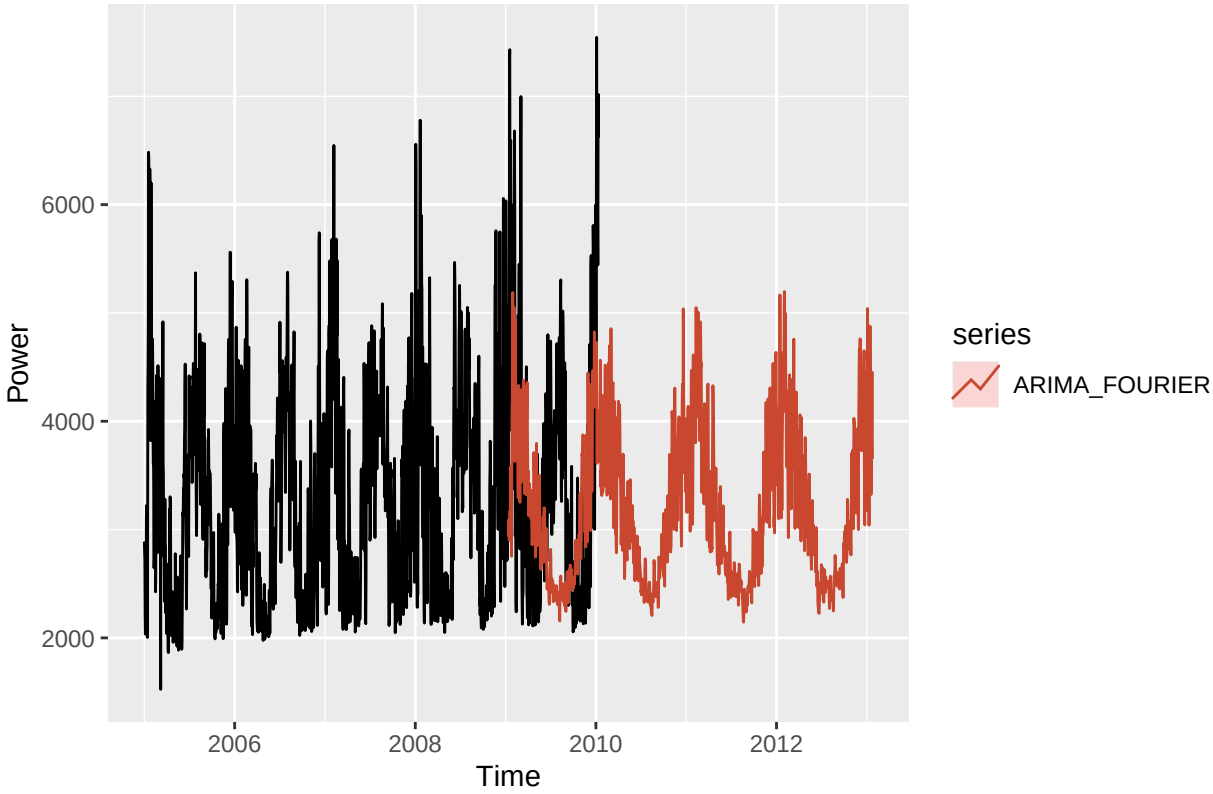



```
#Full dataset
NN_fit_full_4b <- nnetar(ts_daily,xreg=fourier(ts_daily, K=c(2,6)))
NN_for_full_4b <- forecast(NN_fit_full_4b, h=59,xreg=fourier(ts_daily, K=c(2,6),h=59))

model4b <- fortify(NN_for_full_4b)
model4b_data <- model4b[1838:1896,4]

submission4b <- cbind.data.frame(date, model4b_data)
names(submission4b) <- c("date", "load")

write.csv(submission4b, "./Submissions/submission4b.csv", row.names = FALSE)
```

```
#Model 5 - ARIMA with external regressors
ARIMA_four_fit_5 <- auto.arima(ts_daily_train,
                               seasonal=FALSE,
                               lambda=0,
                               xreg=(ts_daily_temp_train))

ARIMA_four_for_5 <- forecast(ARIMA_four_fit_5,
                             xreg=(ts_daily_temp_train), h=365)

#Plot foresting results
autoplot(ARIMA_four_for_5) + ylab("Power")
```



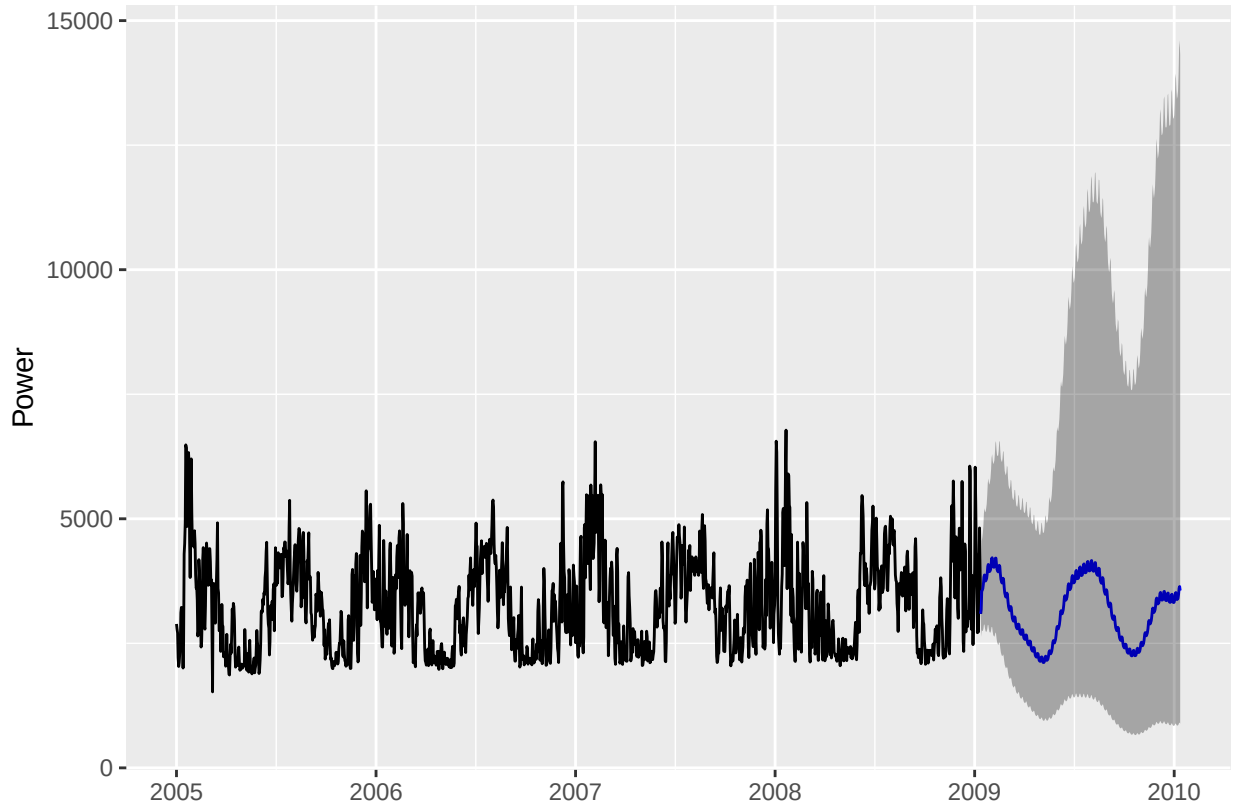

```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(ARIMA_four_for_5, series="ARIMA_FOURIER",PI=FALSE) +
  ylab("Power")
```



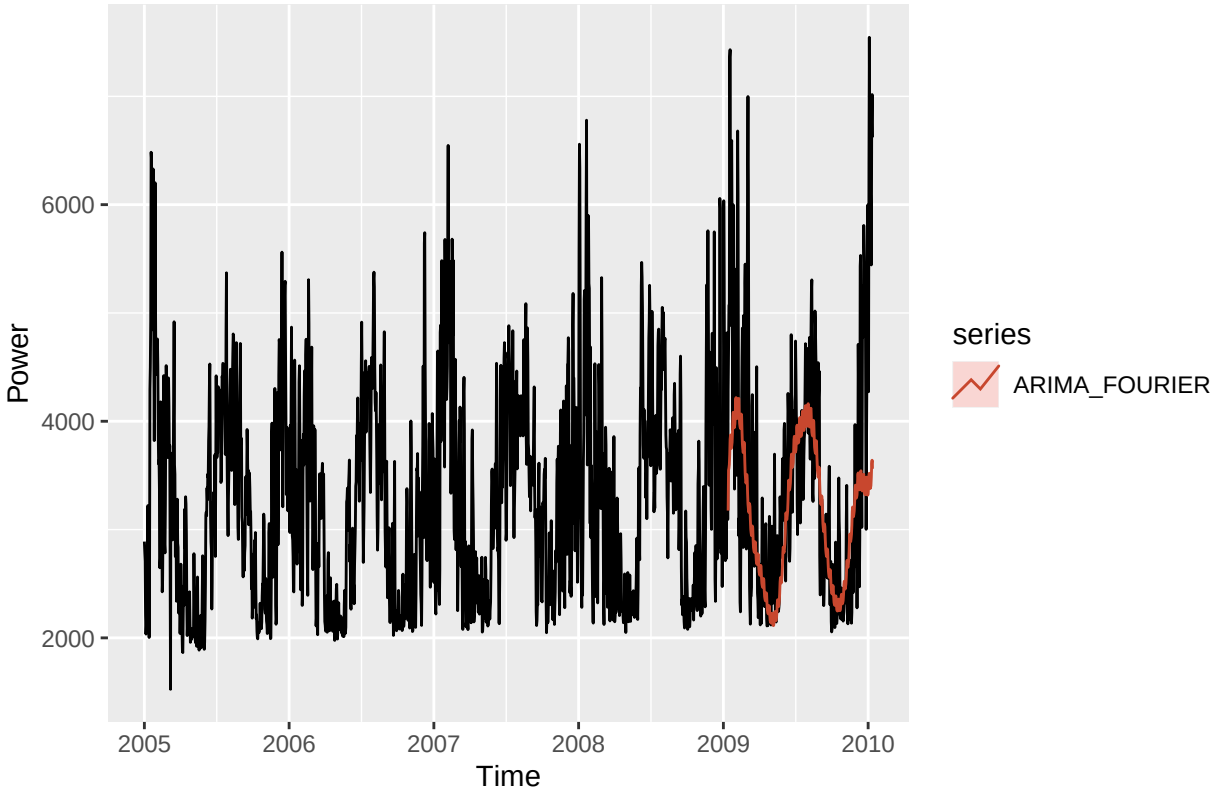

```
ARIMA_four_fit_full_5 <- auto.arima(ts_daily,
                            seasonal=FALSE,
                            lambda=0,
                            xreg=(ts_daily_temp))

ARIMA_four_for_full_5 <- forecast(ARIMA_four_fit_full_5,
                           xreg=(ts_daily_temp_train), h=59)

model5 <- fortify(ARIMA_four_for_full_5)
model5_data <- model5[1838:1896,4]

submission5 <- cbind.data.frame(date, model5_data)
names(submission5) <- c("date", "load")

write.csv(submission5, "./Submissions/submission5.csv", row.names = FALSE)
```

```
ARIMA_four_fit_5b <- auto.arima(ts_daily_train,
                            seasonal=FALSE,
                            lambda=0,
                            xreg=fourier(ts_daily_temp_train,
                                         K=c(2,6))
                            )

ARIMA_four_for_5b <- forecast(ARIMA_four_fit_5b,
                           xreg=fourier(ts_daily_temp_train,
                                        K=c(2,6),
```

```
                                        h=365),
                          h=365
                          )

#Plot foresting results
autoplot(ARIMA_four_for_5b) + ylab("Power")
```



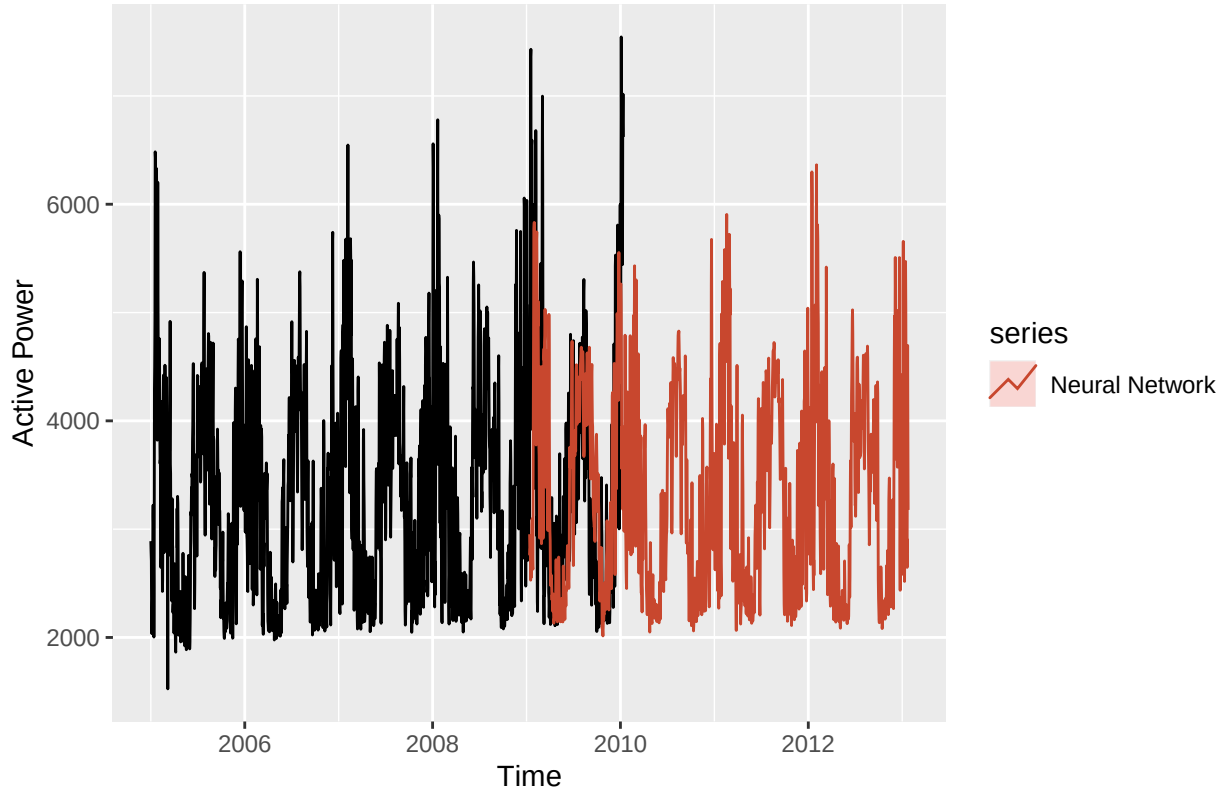

```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(ARIMA_four_for_5b, series="ARIMA_FOURIER",PI=FALSE) +
  ylab("Power")
```




```
ARIMA_four_fit_full_5b <- auto.arima(ts_daily,
                             seasonal=FALSE,
                             lambda=0,
                             xreg=fourier(ts_daily_temp,
                                          K=c(2,6))
                             )

ARIMA_four_for_full_5b <- forecast(ARIMA_four_fit_full_5b,
                            xreg=fourier(ts_daily_temp,
                                         K=c(2,6),
                                         h=59),
                            h=59
                            )

model5b <- fortify(ARIMA_four_for_full_5b)
model5b_data <- model5b[1838:1896,4]

submission5b <- cbind.data.frame(date, model5b_data)
names(submission5b) <- c("date", "load")

write.csv(submission5b, "./Submissions/submission5b.csv", row.names = FALSE)
```

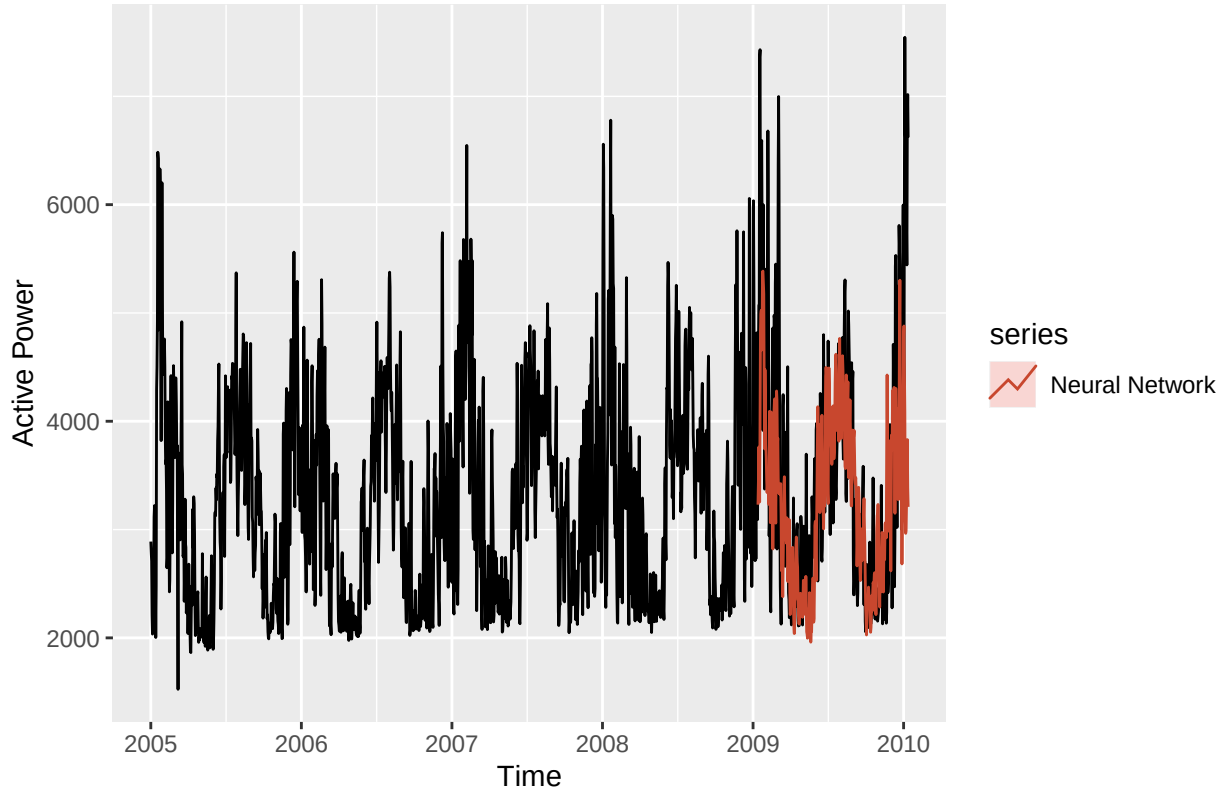
```
#Neural networks with external regressors
NN_fit_6 <- nnetar(ts_daily_train,xreg=ts_daily_temp_train)

NN_for_6 <- forecast(NN_fit_6, h=365,xreg=ts_daily_temp_train)
```

```
#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_6, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```



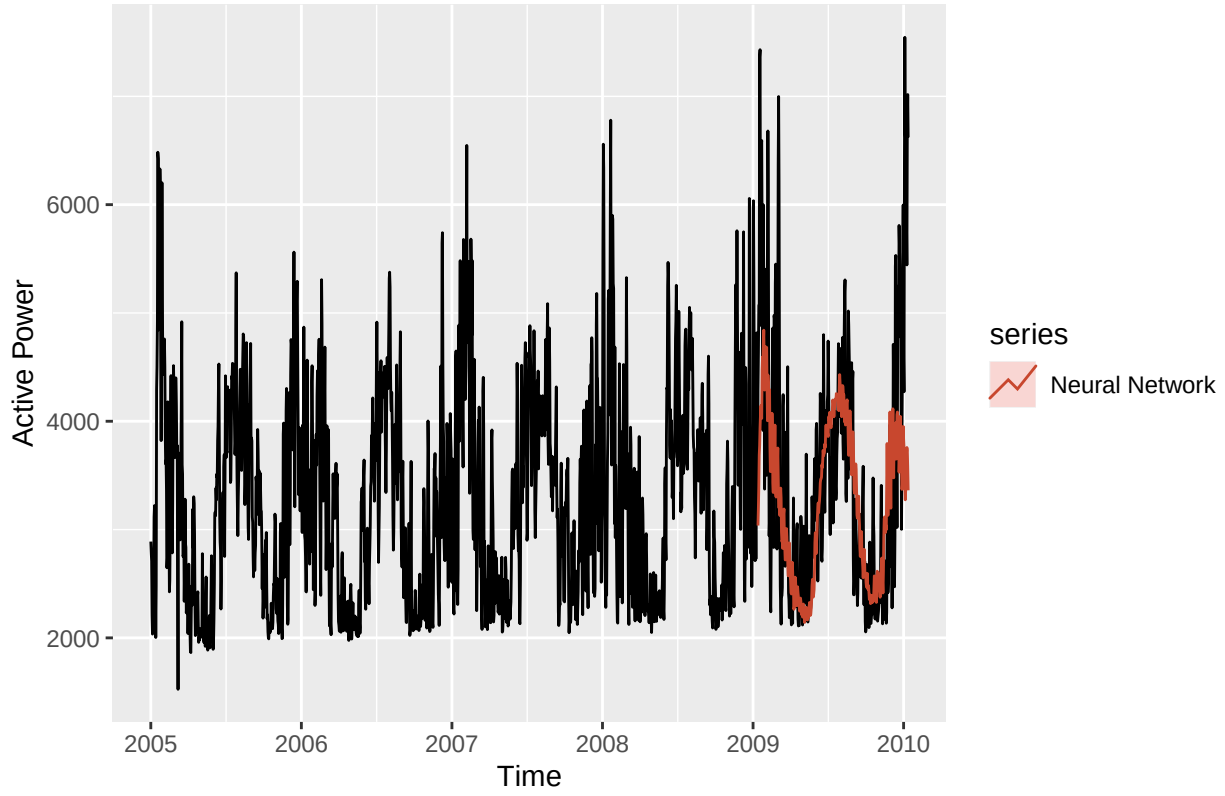

```
#Full dataset
NN_fit_full_6 <- nnetar(ts_daily,xreg=ts_daily_temp)
NN_for_full_6 <- forecast(NN_fit_full_6, h=59,xreg=ts_daily_temp)

model6 <- fortify(NN_for_full_6)
model6_data <- model6[1838:1896,4]

submission6 <- cbind.data.frame(date, model6_data)
names(submission6) <- c("date", "load")

write.csv(submission6, "./Submissions/submission6.csv", row.names = FALSE)
```

```
NN_fit_7 <- nnetar(ts_daily_train,p=1,P=0,xreg=fourier(ts_daily_train, K=c(2,12)))

NN_for_7 <- forecast(NN_fit_7, h=365,xreg=fourier(ts_daily_train,
                                        K=c(2,12),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_7, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```



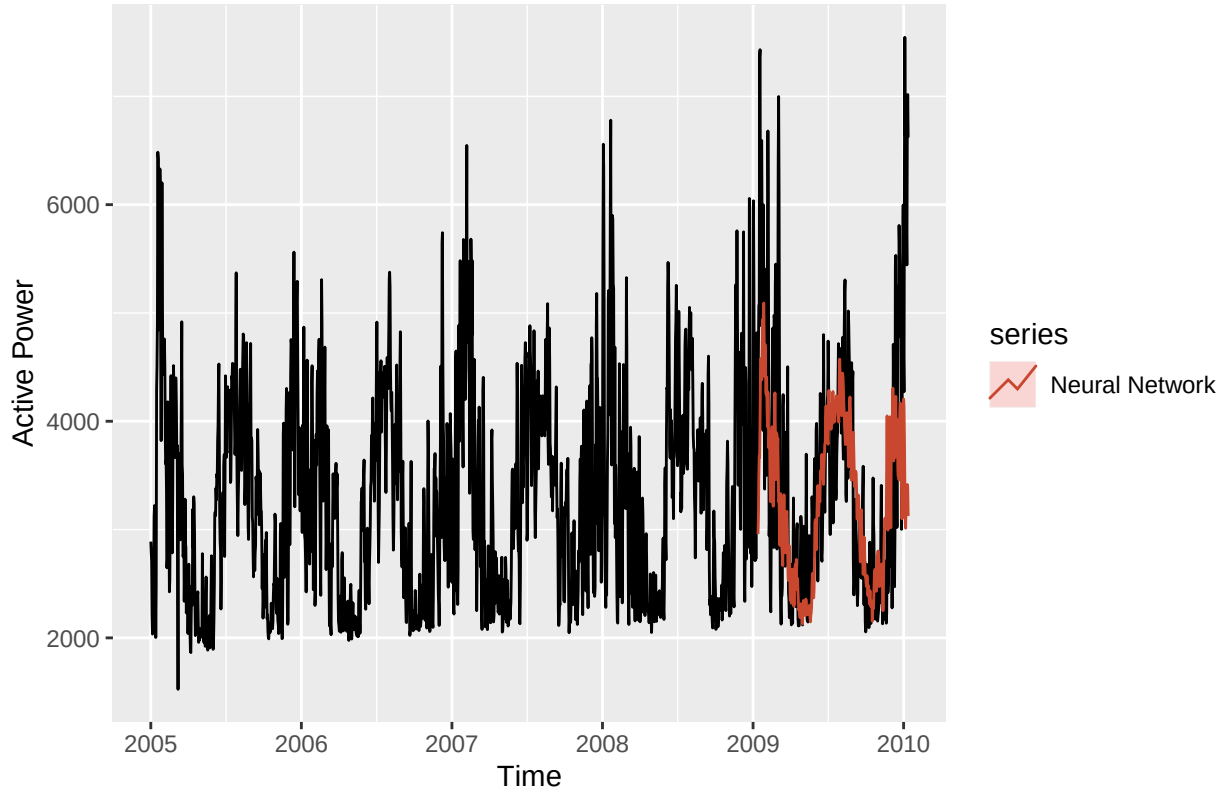

```
#Full dataset
NN_fit_full_7 <- nnetar(ts_daily,p=1,P=0,xreg=fourier(ts_daily, K=c(2,12)))
NN_for_full_7 <- forecast(NN_fit_full_7, h=59,xreg=fourier(ts_daily,
                                          K=c(2,12),h=59))

model7 <- fortify(NN_for_full_7)
model7_data <- model7[1838:1896,4]

submission7 <- cbind.data.frame(date, model7_data)
names(submission7) <- c("date", "load")

write.csv(submission7, "./Submissions/submission7.csv", row.names = FALSE)
```

```
#neural network with different fourier terms
NN_fit_7b <- nnetar(ts_daily_train,p=1,P=0,xreg=fourier(ts_daily_train, K=c(2,4)))

NN_for_7b <- forecast(NN_fit_7b, h=365,xreg=fourier(ts_daily_train,
                                          K=c(2,4),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_7b, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```



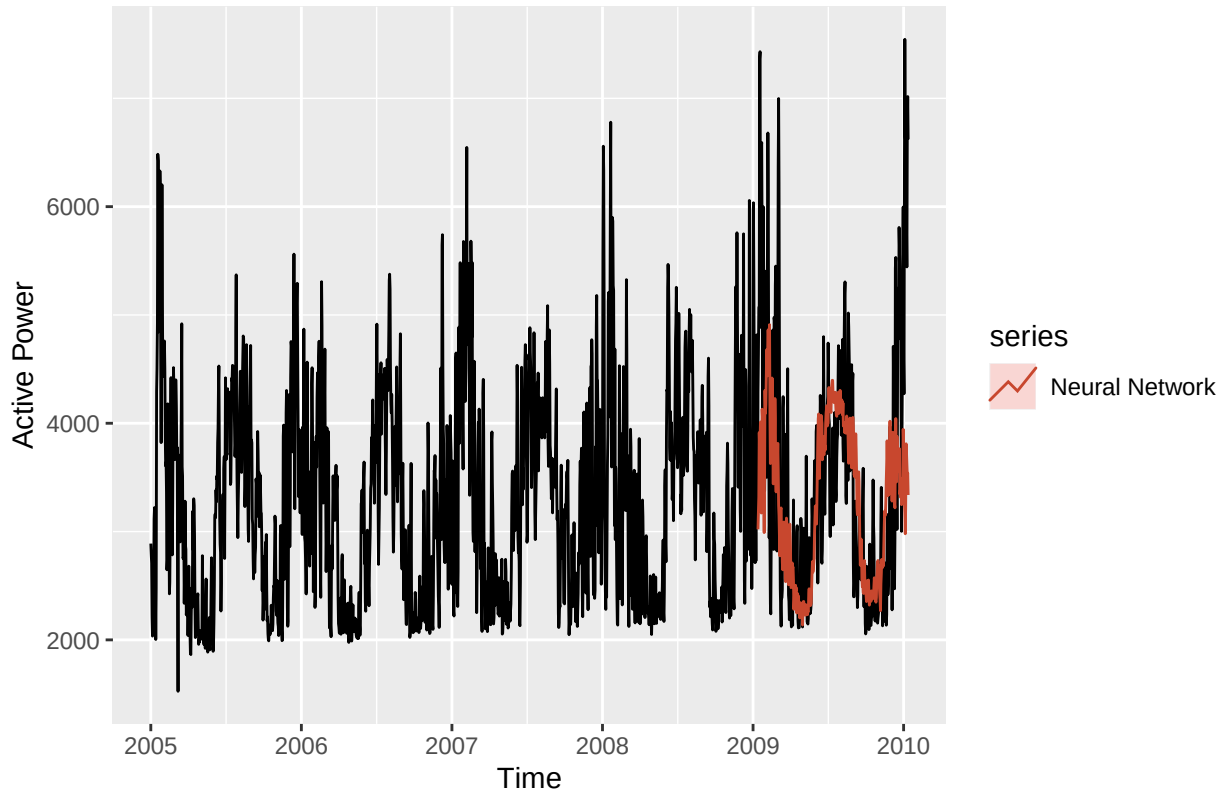

```
#Full dataset
NN_fit_full_7b <- nnetar(ts_daily,p=1,P=0,xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_7b <- forecast(NN_fit_full_7b, h=59,xreg=fourier(ts_daily,
                                          K=c(2,4),h=59))

model7b <- fortify(NN_for_full_7b)
model7b_data <- model7b[1838:1896,4]

submission7b <- cbind.data.frame(date, model7b_data)
names(submission7b) <- c("date", "load")

write.csv(submission7b, "./Submissions/submission7b.csv", row.names = FALSE)
```

```
#same model but with different fourier parameters
NN_fit_7c <- nnetar(ts_daily_train,p=1,P=0,xreg=fourier(ts_daily_train, K=c(2,6)))

NN_for_7c <- forecast(NN_fit_7c, h=365,xreg=fourier(ts_daily_train,
                                          K=c(2,6),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_7c, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```



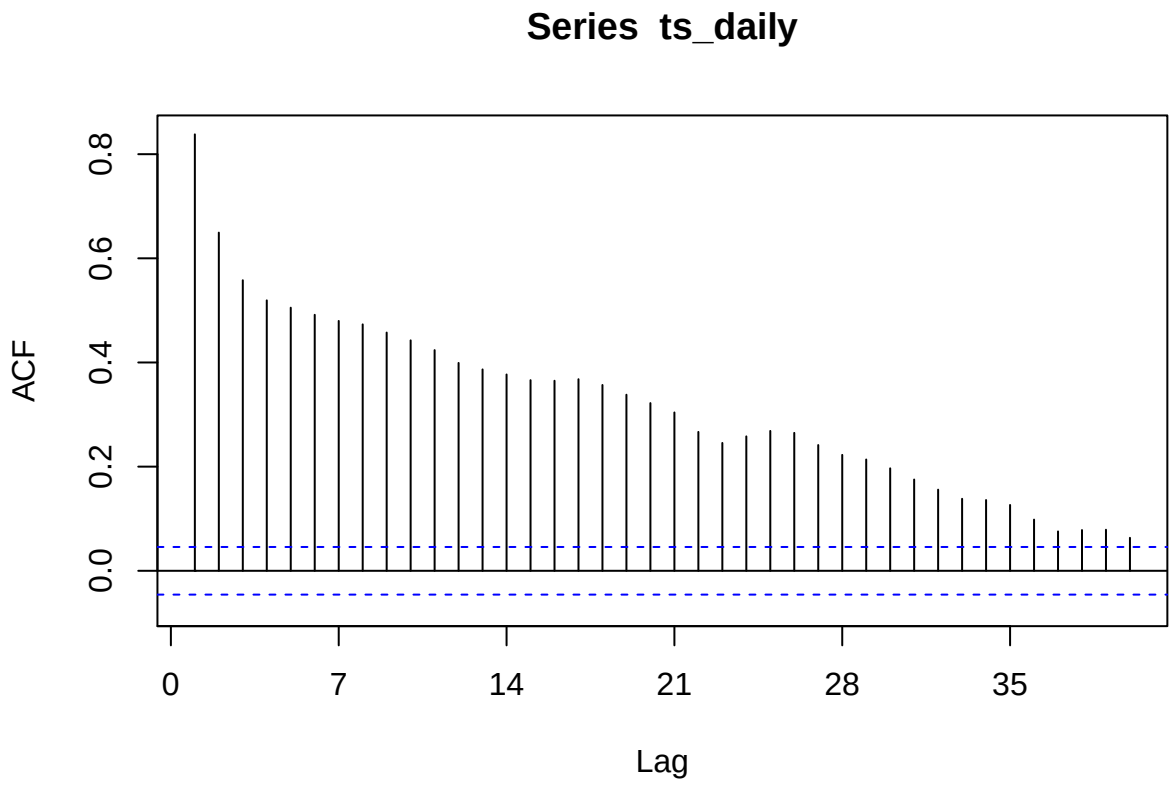

```
#Full dataset
NN_fit_full_7c <- nnetar(ts_daily,p=1,P=0,xreg=fourier(ts_daily, K=c(2,6)))
NN_for_full_7c <- forecast(NN_fit_full_7c, h=59,xreg=fourier(ts_daily,
                                          K=c(2,6),h=59))

model7c <- fortify(NN_for_full_7c)
model7c_data <- model7c[1838:1896,4]

submission7c <- cbind.data.frame(date, model7c_data)
names(submission7c) <- c("date", "load")

write.csv(submission7c, "./Submissions/submission7c.csv", row.names = FALSE)
```

```
#neural network with different fourier terms
NN_fit_7d <- nnetar(ts_daily_train,p=1,P=1,xreg=fourier(ts_daily_train, K=c(2,4)))

NN_for_7d <- forecast(NN_fit_7d, h=365,xreg=fourier(ts_daily_train,
                                          K=c(2,4),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_7d, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```

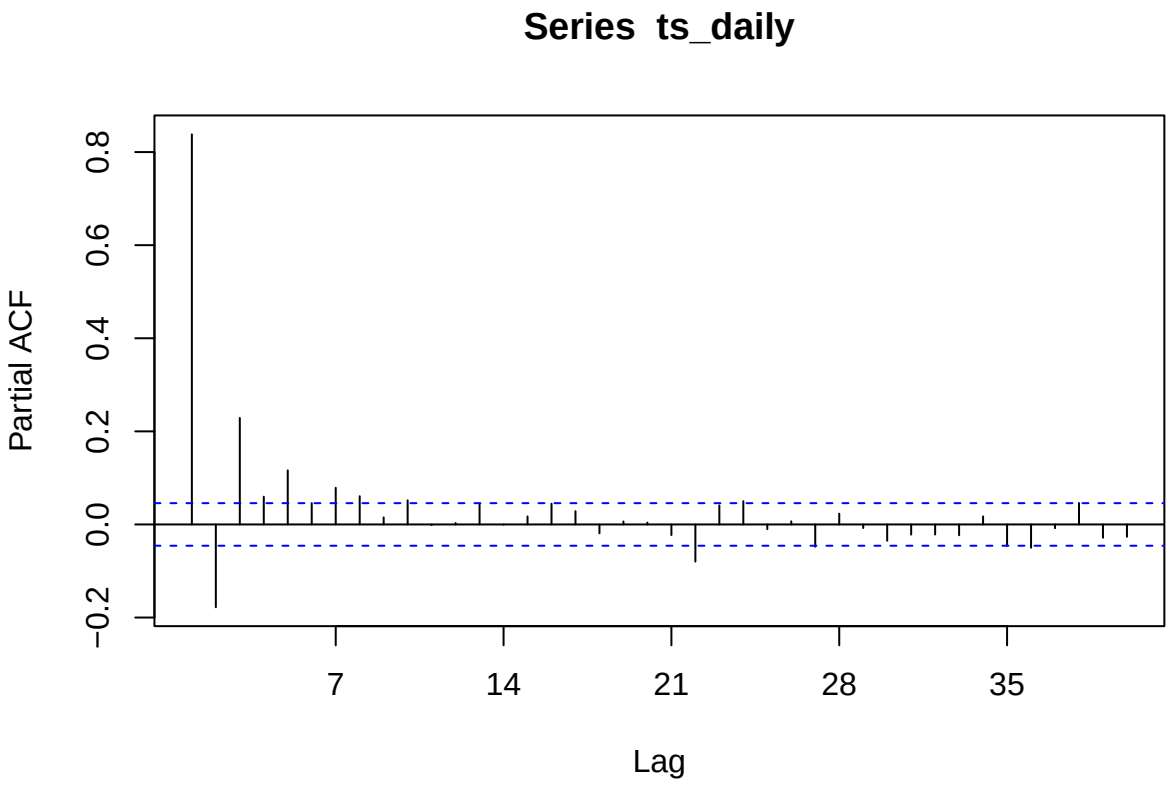



```
#Full dataset
NN_fit_full_7d <- nnetar(ts_daily,p=1,P=1,xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_7d <- forecast(NN_fit_full_7d, h=59,xreg=fourier(ts_daily,
                                          K=c(2,4),h=59))

model7d <- fortify(NN_for_full_7d)
model7d_data <- model7d[1838:1896,4]

submission7d <- cbind.data.frame(date, model7d_data)
names(submission7d) <- c("date", "load")

write.csv(submission7d, "./Submissions/submission7d.csv", row.names = FALSE)
```

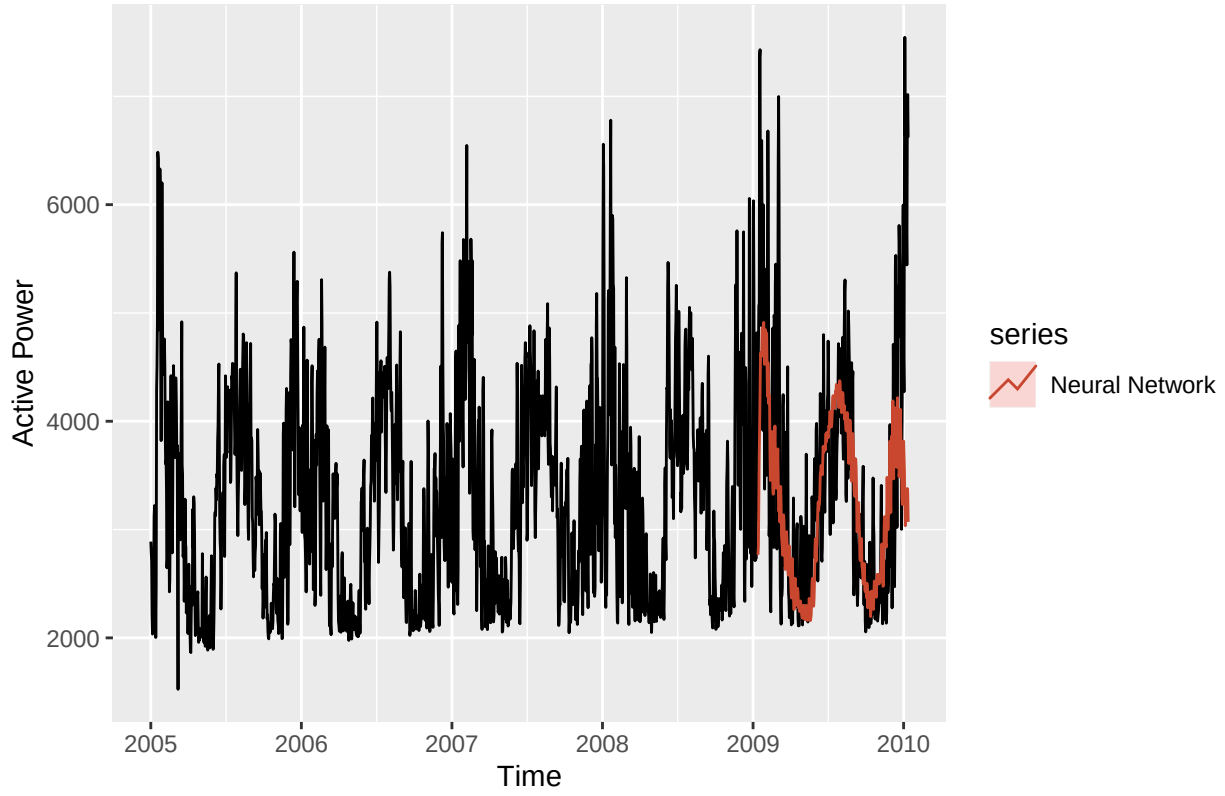
```
Acf(ts_daily, lag.max = 40)
```




```
Pacf(ts_daily, lag.max=40)
```

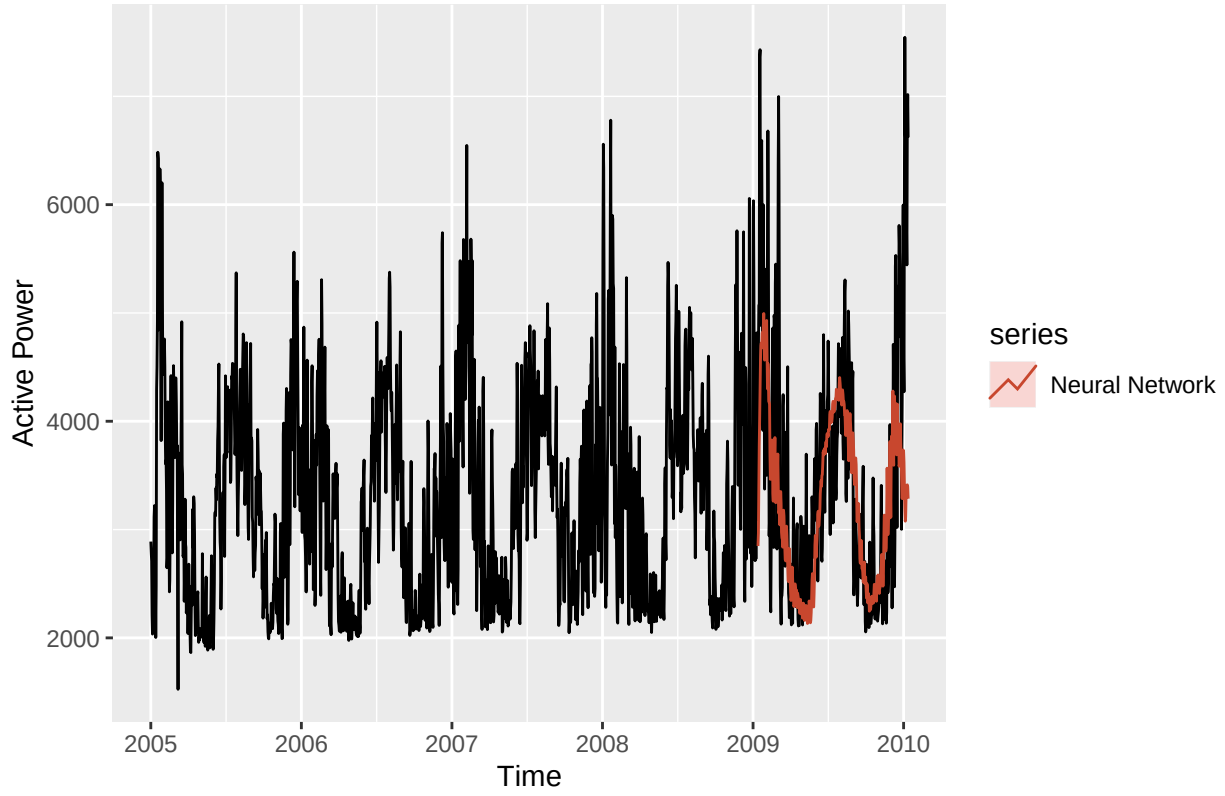



```
#neural network with lag 2
NN_fit_8 <- nnetar(ts_daily_train,p=2,P=0,xreg=fourier(ts_daily_train, K=c(2,4)))

NN_for_8 <- forecast(NN_fit_8, h=365,xreg=fourier(ts_daily_train,
                                          K=c(2,4),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_8, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```



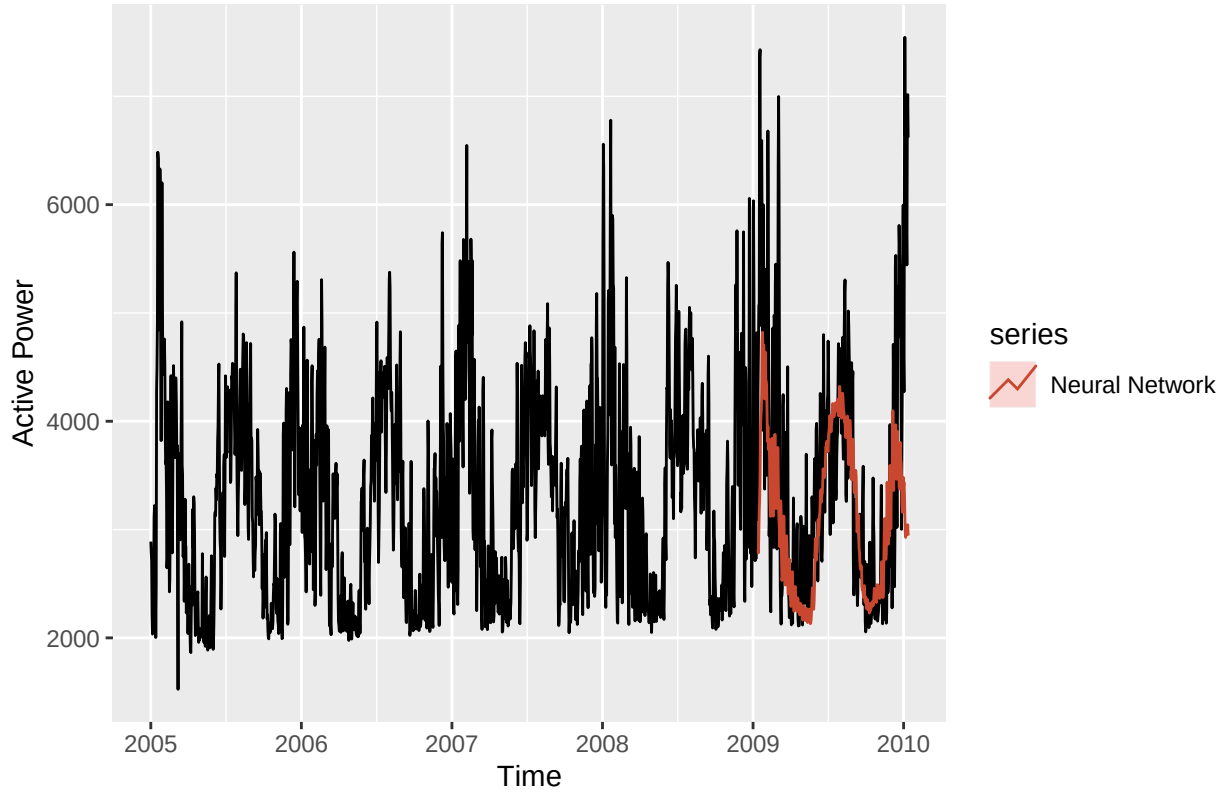

```
#Full dataset
NN_fit_full_8 <- nnetar(ts_daily,p=2,P=0,xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_8 <- forecast(NN_fit_full_8, h=59,xreg=fourier(ts_daily,
                                          K=c(2,4),h=59))

model8 <- fortify(NN_for_full_8)
model8_data <- model8[1838:1896,4]

submission8 <- cbind.data.frame(date, model8_data)
names(submission8) <- c("date", "load")

write.csv(submission8, "./Submissions/submission8.csv", row.names = FALSE)
```

```
#same as model 8, but with more repeats
NN_fit_8b <- nnetar(ts_daily_train,p=2,P=0, repeats=50, xreg=fourier(ts_daily_train, K=c(2,4)))

NN_for_8b <- forecast(NN_fit_8b, h=365,xreg=fourier(ts_daily_train,
                                          K=c(2,4),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_8b, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```



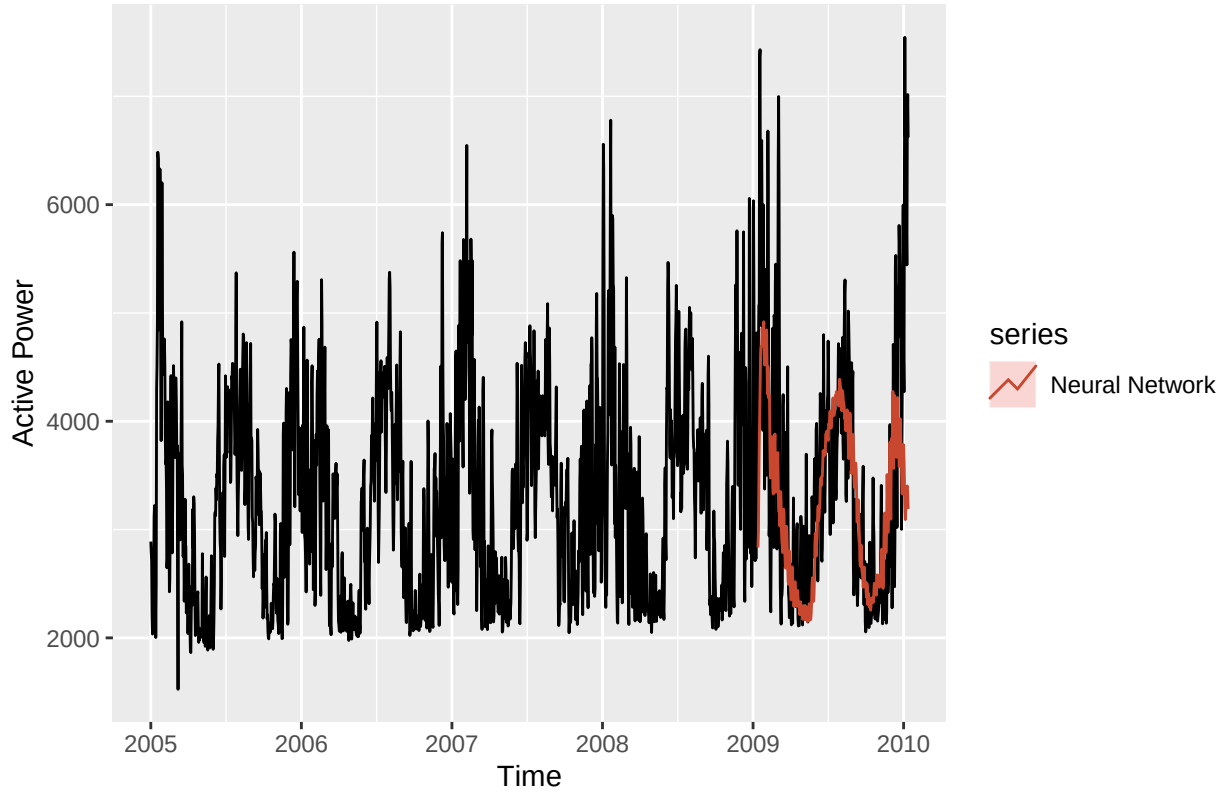

```
#Full dataset
NN_fit_full_8b <- nnetar(ts_daily,p=2,P=0, repeats=50, xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_8b <- forecast(NN_fit_full_8b, h=59,xreg=fourier(ts_daily,
                                          K=c(2,4),h=59))

model8b <- fortify(NN_for_full_8b)
model8b_data <- model8b[1838:1896,4]

submission8b <- cbind.data.frame(date, model8b_data)
names(submission8b) <- c("date", "load")

write.csv(submission8b, "./Submissions/submission8b.csv", row.names = FALSE)
```

```
#same as model 8, but with more repeats; lamda=auto
NN_fit_8c <- nnetar(ts_daily_train,p=2,P=0, repeats=50,lambda="auto", xreg=fourier(ts_daily_train, K=c(1

NN_for_8c <- forecast(NN_fit_8c, h=365,xreg=fourier(ts_daily_train,
                                          K=c(2,4),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_8c, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```




```
#Full dataset
NN_fit_full_8c <- nnetar(ts_daily,p=2,P=0, repeats=50, lambda="auto", xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_8c <- forecast(NN_fit_full_8c, h=59,xreg=fourier(ts_daily,
                                          K=c(2,4),h=59))

model8c <- fortify(NN_for_full_8c)
model8c_data <- model8c[1838:1896,4]

submission8c <- cbind.data.frame(date, model8c_data)
names(submission8c) <- c("date", "load")

write.csv(submission8c, "./Submissions/submission8c.csv", row.names = FALSE)
```

```
#same as model 8, but with more repeats
NN_fit_8d <- nnetar(ts_daily_train,p=2,P=0, repeats=500, xreg=fourier(ts_daily_train, K=c(2,4)))

NN_for_8d <- forecast(NN_fit_8d, times=10000, h=365,xreg=fourier(ts_daily_train,K=c(2,4),h=365))

#Plot model + observed data
autoplot(ts_daily) +
  autolayer(NN_for_8d, series="Neural Network",PI=FALSE)+
  ylab("Active Power")
```




```
#Full dataset
NN_fit_full_8d <- nnetar(ts_daily,p=2,P=0, repeats=500, xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_8d <- forecast(NN_fit_full_8d, times=10000, h=59,xreg=fourier(ts_daily, K=c(2,4),h=59))

model8d <- fortify(NN_for_full_8d)
model8d_data <- model8d[1838:1896,4]

submission8d <- cbind.data.frame(date, model8d_data)
names(submission8d) <- c("date", "load")

write.csv(submission8d, "./Submissions/submission8d.csv", row.names = FALSE)
```

```
#Full dataset
NN_fit_full_8e <- nnetar(ts_daily,p=2,P=0, repeats=50, xreg=fourier(ts_daily, K=c(2,4)))
NN_for_full_8e <- forecast(NN_fit_full_8e, h=59,xreg=fourier(ts_daily,
                                        K=c(2,4),h=59))

model8e <- fortify(NN_for_full_8e)
model8e_data <- model8e[1838:1896,4]

submission8e <- cbind.data.frame(date, model8e_data)
names(submission8e) <- c("date", "load")

write.csv(submission8e, "./Submissions/submission8e.csv", row.names = FALSE)
```

```
#Full dataset
NN_fit_full_8f <- nnetar(ts_daily,p=3,P=0, repeats=50, xreg=fourier(ts_daily, K=c(2,4)))
```

```
NN_for_full_8f <- forecast(NN_fit_full_8f, h=59,xreg=fourier(ts_daily,
                                          K=c(2,4),h=59))

model8f <- fortify(NN_for_full_8f)
model8f_data <- model8f[1838:1896,4]

submission8f <- cbind.data.frame(date, model8f_data)
names(submission8f) <- c("date", "load")

write.csv(submission8f, "./Submissions/submission8f.csv", row.names = FALSE)
```