NetGen\_USA

# This project is to analyze the electricity generation from different sectors in USA.

## To input data

library(readr)  
netgen <- read\_csv("Data\_TS/Net\_generation\_United\_States\_all\_sectors\_monthly.csv",   
 col\_types = cols(Month = col\_character()),   
 skip = 4)  
  
# inspect the data set  
head(netgen)

## # A tibble: 6 x 8  
## Month `all fuels (utility-sca~ `coal thousand meg~ `natural gas thousa~  
## <chr> <dbl> <dbl> <dbl>  
## 1 Aug 2~ 410485. 115218. 164954.  
## 2 Jul 2~ 412383. 115472. 167066.  
## 3 Jun 2~ 372386. 101508. 130827.  
## 4 May 2~ 339671. 85311. 116110.  
## 5 Apr 2~ 301791. 73383. 99447.  
## 6 Mar 2~ 321015. 80613. 104939.  
## # ... with 4 more variables: `nuclear thousand megawatthours` <dbl>,  
## # `conventional hydroelectric thousand megawatthours` <dbl>, `wind  
## # thousand megawatthours` <dbl>, `all solar thousand  
## # megawatthours` <dbl>

names(netgen)

## [1] "Month"   
## [2] "all fuels (utility-scale) thousand megawatthours"   
## [3] "coal thousand megawatthours"   
## [4] "natural gas thousand megawatthours"   
## [5] "nuclear thousand megawatthours"   
## [6] "conventional hydroelectric thousand megawatthours"  
## [7] "wind thousand megawatthours"   
## [8] "all solar thousand megawatthours"

colnames(netgen) <- c("Month", "all\_fuels", "coal", "natural\_gas", "nuclear", "hydro", "wind", "solar")  
  
head(netgen)

## # A tibble: 6 x 8  
## Month all\_fuels coal natural\_gas nuclear hydro wind solar  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Aug 2018 410485. 115218. 164954. 72282. 21398. 19507. 10000.  
## 2 Jul 2018 412383. 115472. 167066. 72456. 24013. 16022. 10087.  
## 3 Jun 2018 372386. 101508. 130827. 69688. 27953. 24340. 10869.  
## 4 May 2018 339671. 85311. 116110. 67320. 30433. 23542. 10088.  
## 5 Apr 2018 301791. 73383. 99447. 59133. 27488. 26803. 8993.  
## 6 Mar 2018 321015. 80613. 104939. 67033. 25950. 27287. 7636.

# change the 'month' to date type   
# try as.Date first  
# netgen$Month <- as.Date.character(netgen$Month)  
library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

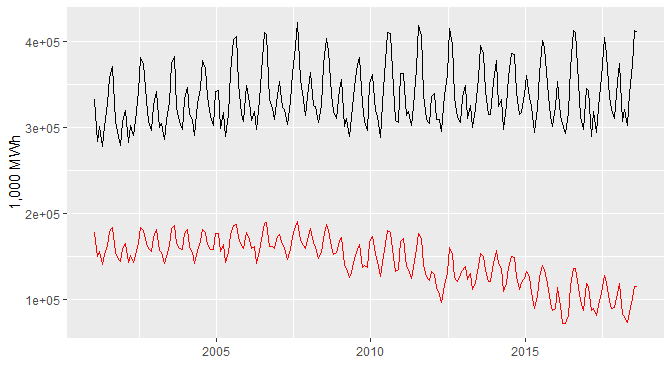
netgen$Month <- mdy(netgen$Month)  
head(netgen)

## # A tibble: 6 x 8  
## Month all\_fuels coal natural\_gas nuclear hydro wind solar  
## <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2018-08-20 410485. 115218. 164954. 72282. 21398. 19507. 10000.  
## 2 2018-07-20 412383. 115472. 167066. 72456. 24013. 16022. 10087.  
## 3 2018-06-20 372386. 101508. 130827. 69688. 27953. 24340. 10869.  
## 4 2018-05-20 339671. 85311. 116110. 67320. 30433. 23542. 10088.  
## 5 2018-04-20 301791. 73383. 99447. 59133. 27488. 26803. 8993.  
## 6 2018-03-20 321015. 80613. 104939. 67033. 25950. 27287. 7636.

# plot all the trend on the same graph

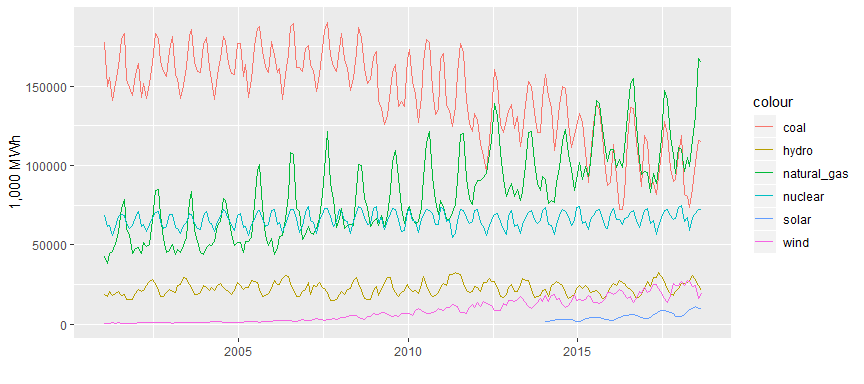
<https://stackoverflow.com/questions/10349206/add-legend-to-ggplot2-line-plot>

library(dplyr)  
library(ggplot2)  
netgen %>% ggplot (aes(x=Month, y=all\_fuels)) +  
 geom\_line() +  
 labs (x="", y="1,000 MWh")+  
 geom\_line(aes(x=Month, y=coal), color="red")



my.cols <- heat.colors(8, alpha=1)  
my.names <- colnames(netgen)  
names(my.cols) <- my.names  
  
  
data<- netgen  
p <- ggplot(data, aes(x = Month))  
  
for (i in 2:7){  
 p <- p + geom\_line(aes\_(y = as.name(names(data[i+1])), colour =   
colnames(data[i+1])))#as.character(my.names[i])))  
}  
p + labs (x="", y="1,000 MWh")

## Warning: Removed 156 rows containing missing values (geom\_path).

 # Observations : \* It shows that ‘solar’ only picks up the momentumn from around year 2014 in the US. \* It also shows that after 2015, electricity from natural gas started exceeding coal based electricity !