

In [110...]

```
# KHANG G MACH
library(ggplot2)
library(lubridate)
library(reshape2)
library(MASS)
utility<- read.table(file = "~/Documents/SP2021/CS424/utilitydata2020.tsv", sep = "\t")
#utility
utility$newDate<-paste(utility$Year,utility$Month,01,sep="/ ")
names(utility)[names(utility) == "E_kWh_per_Day"] <- "eDay"
names(utility)[names(utility) == "Gas_Th_per_Day"] <- "gDay"
names(utility)[names(utility) == "Water_Gals_per_Day"] <- "wDay"
utility$newDate<-paste(utility$Year,utility$Month,01,sep="/ ")

#say e/Day similar to E_kWh_per_Day
#say g/Day similar to Gas_Th_per_Day
#say w/Day similar to Water_Gals_per_Day
```

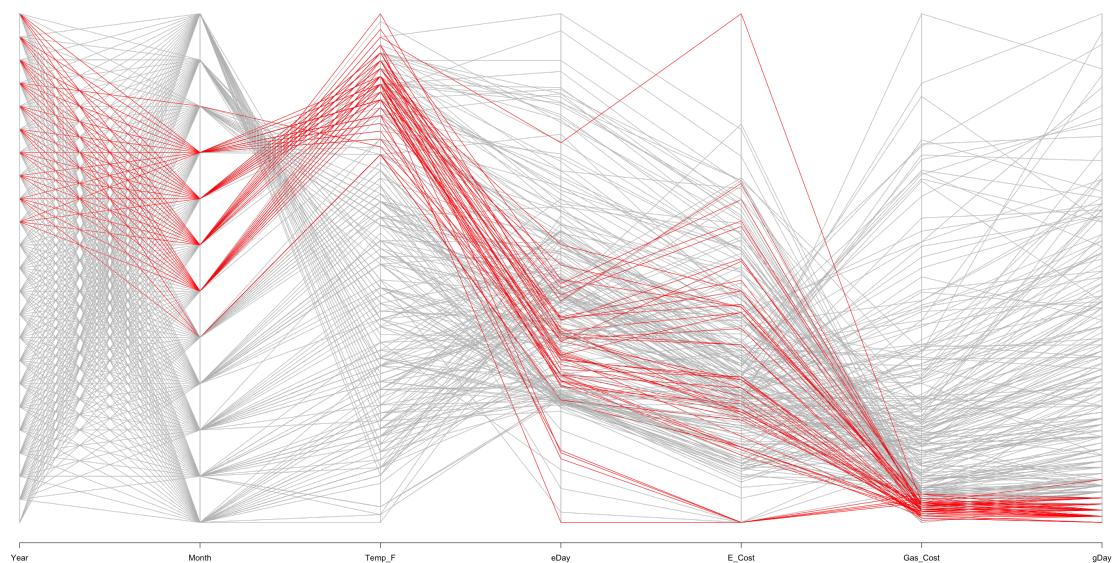
In [100...]

```
options(repr.plot.width=25, repr.plot.height=13)
z<-rbind(utility[,1],utility[,2],utility[,3])
uti <-utility[,c("Year","Month","Temp_F","eDay","E_Cost","Gas_Cost","gDay")]

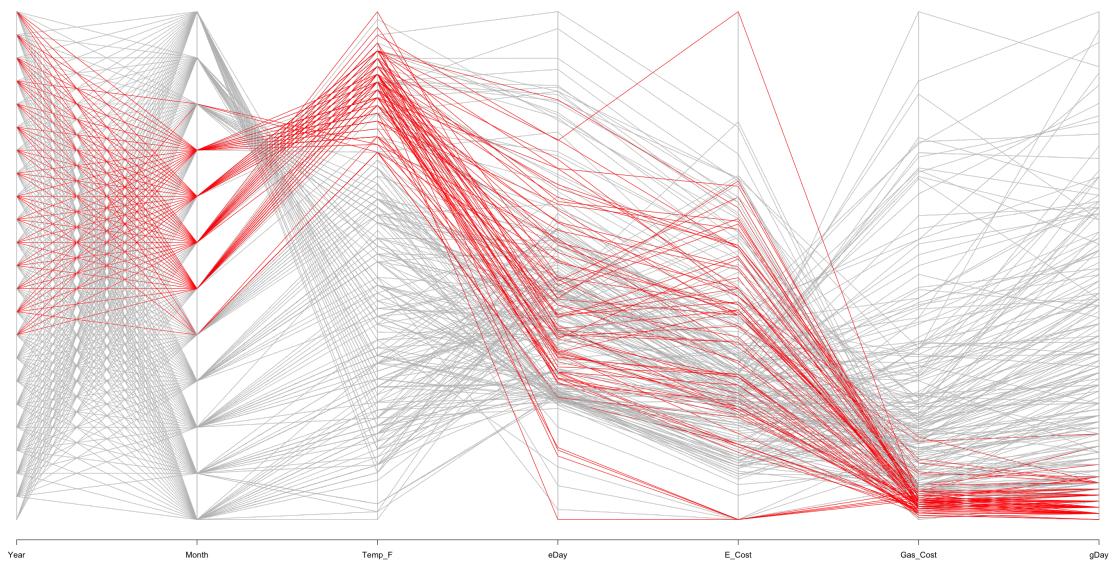
parcoord(uti,col =ifelse(uti$Temp_F> 60 & uti$Year>2010,"red","grey"),main="utisYear>2010 & uti$Temp_F> 60")
parcoord(uti,col =ifelse(uti$Temp_F> 60 & uti$Year>1998 & uti$Year>2005,"red","grey"))
parcoord(uti,col =ifelse(uti$Temp_F <40 & uti$Year>2008 , "red","grey"), main="uti$Temp_F <40 & uti$Year>2008")
parcoord(uti,col =ifelse(uti$Temp_F <40 & uti$Year>1998 & uti$Year>2005 , "red","grey"))

parcoord(uti,col =ifelse(uti$E_Cost >= 25 & uti$E_Cost <= 45 , "red","grey"),main="uti$E_Cost >= 25 & uti$E_Cost <= 45")
parcoord(uti,col =ifelse(uti$E_Cost >= 25 & uti$E_Cost <= 45 & uti$Gas_Cost >= 50 , "red","grey"))
parcoord(uti,col =ifelse(uti$Gas_Cost <=35 , "#BDB76B",ifelse(uti$Gas_Cost >50 , "red","grey")))
parcoord(uti,col =ifelse(uti$E_Cost > 50 & uti$E_Cost < 80 , "#5F9EA0", ifelse(uti$E_Cost > 80 , "red","grey")))
```

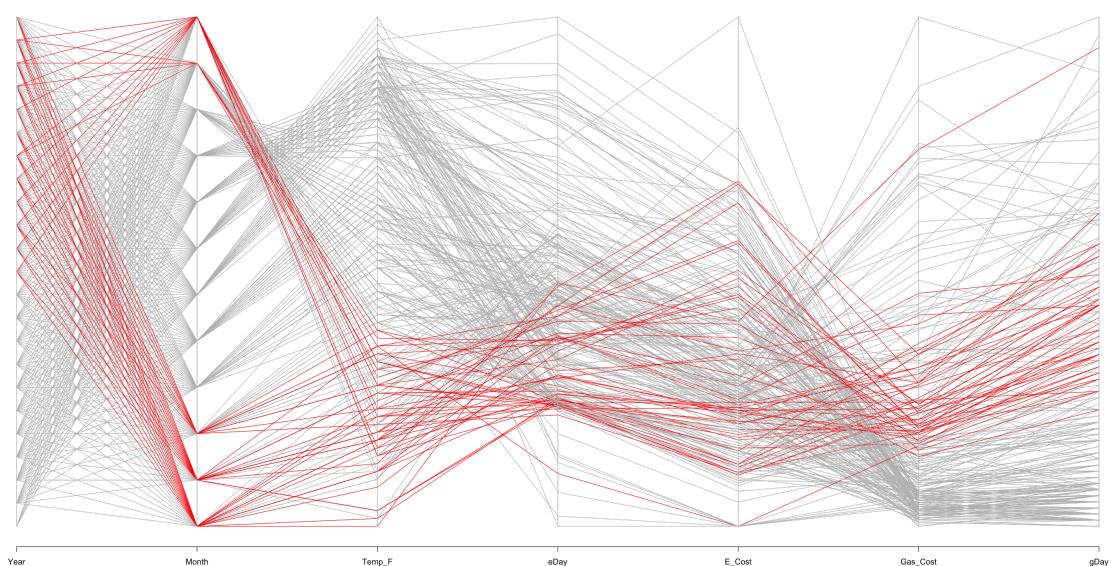
uti\$Year>2010 & uti\$Temp_F> 60



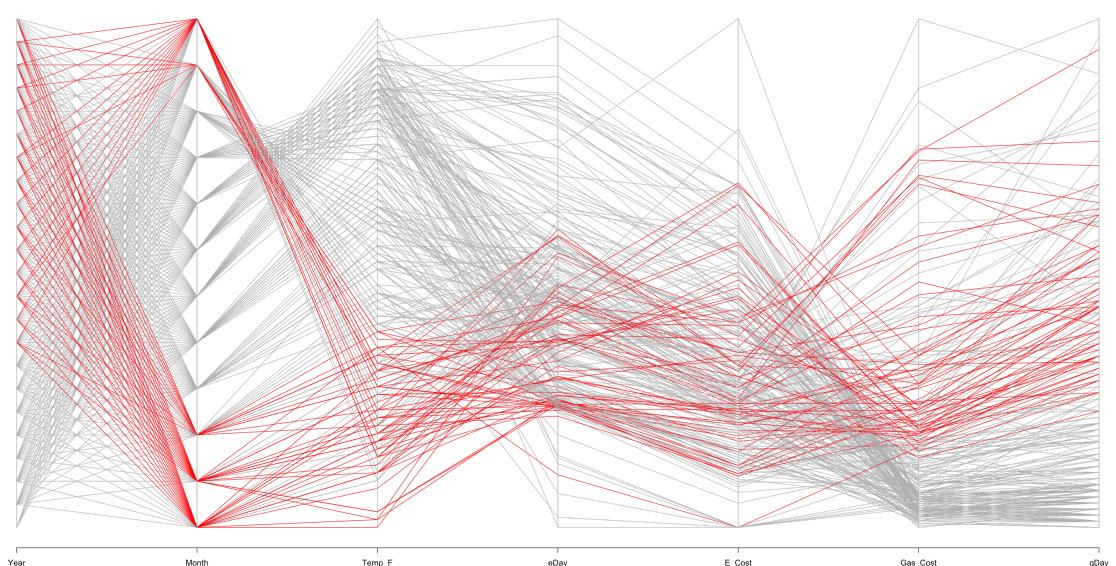
uti\$Year>2005 & uti\$Temp_F> 60 & uti\$Year>1998



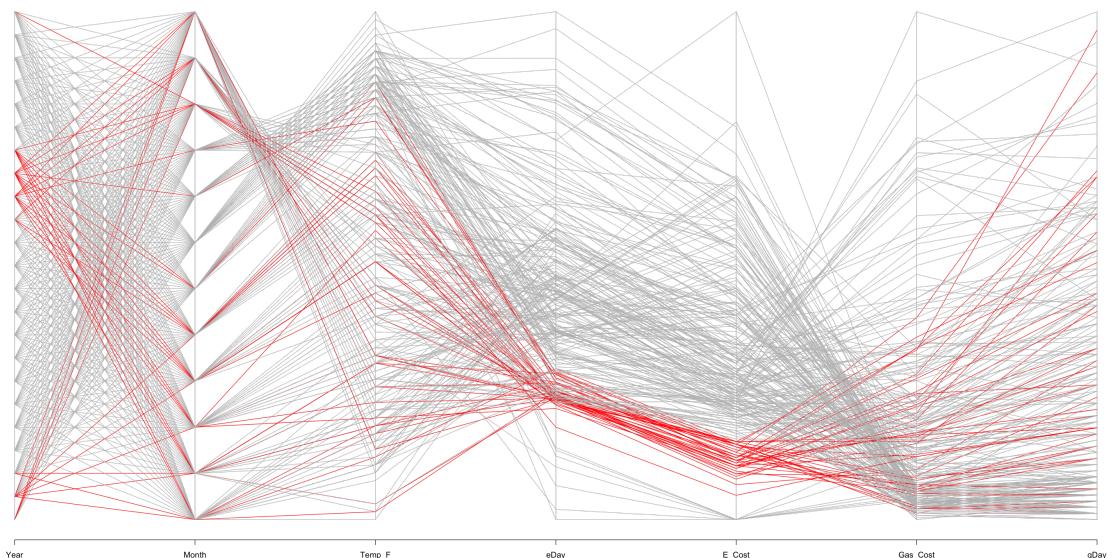
uti\$Year>2008 & uti\$Temp_F <40



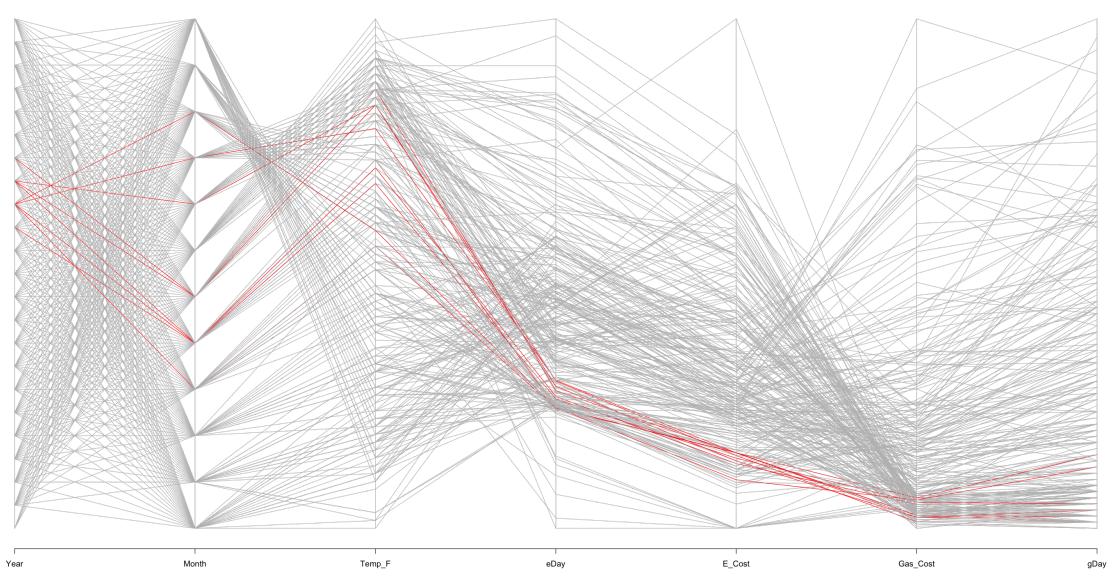
uti\$Year>1998 & uti\$Year>2005 & uti\$Temp_F <40



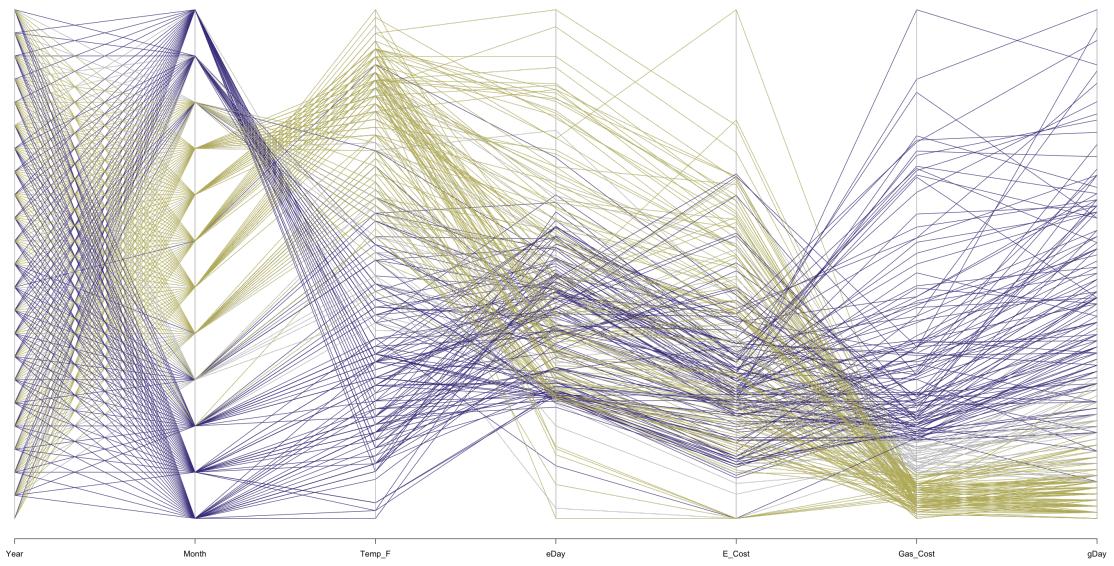
uti\$E_Cost >= 25 & uti\$E_Cost <= 45



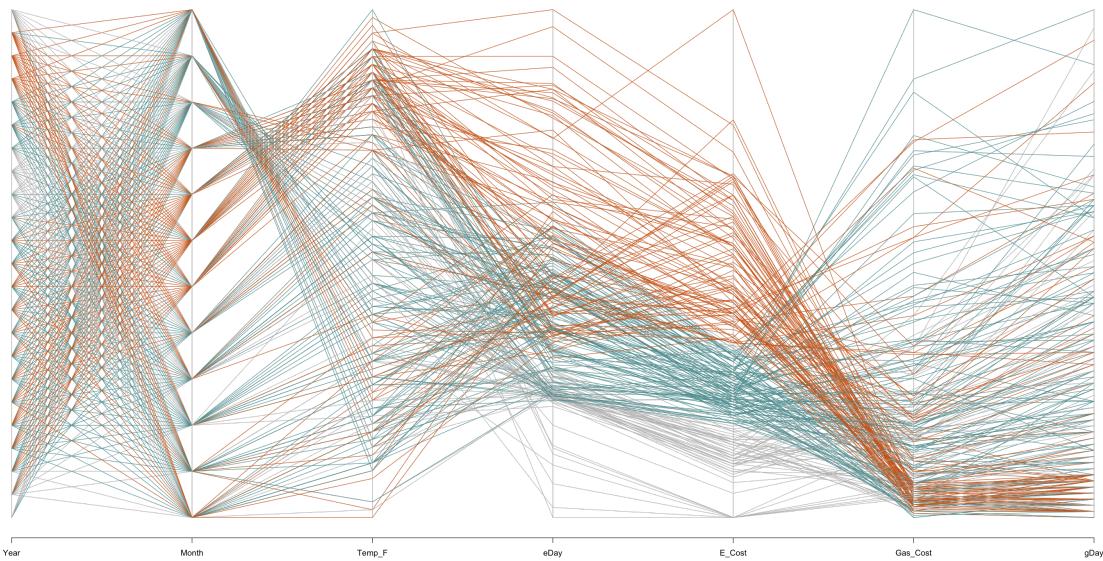
uti\$E_Cost >= 25 & uti\$E_Cost <= 45 & uti\$Gas_Cost < 30



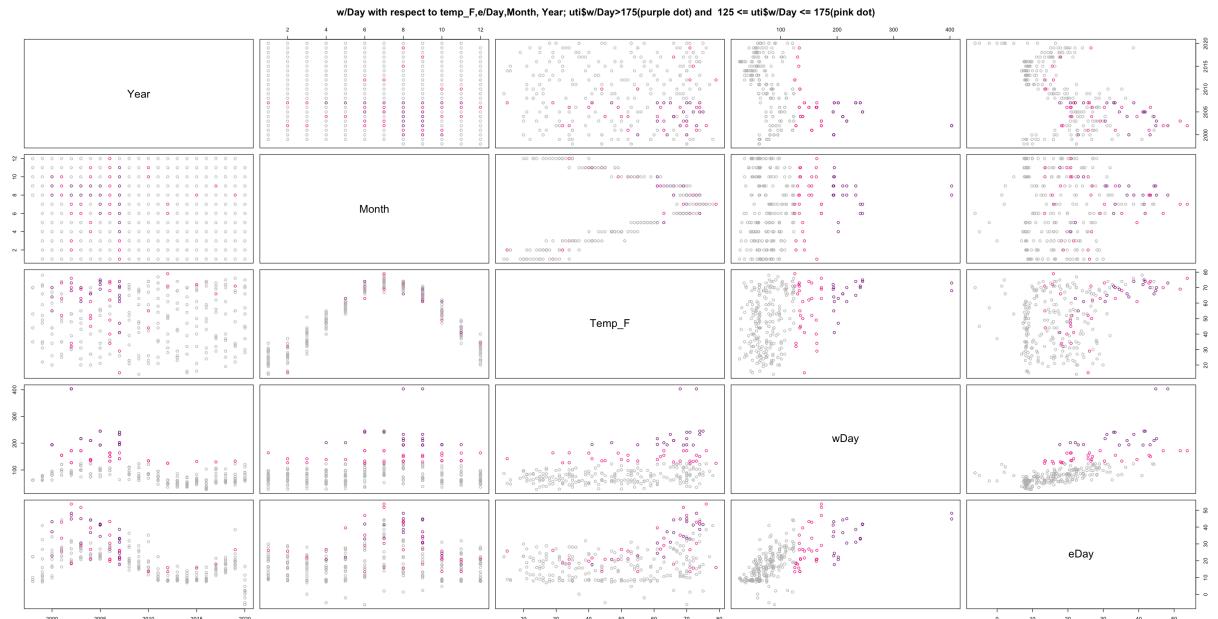
uti\$Gas_Cost <= 35 vs uti\$Gas_Cost > 50



(50 < ult\$E_Cost < 80) vs ult_E_Cost >80



```
In [111]: pairs(data = utility,-Year+Month+Temp_F+wDay+eDay,col= ifelse(utility$wDay>=1
```

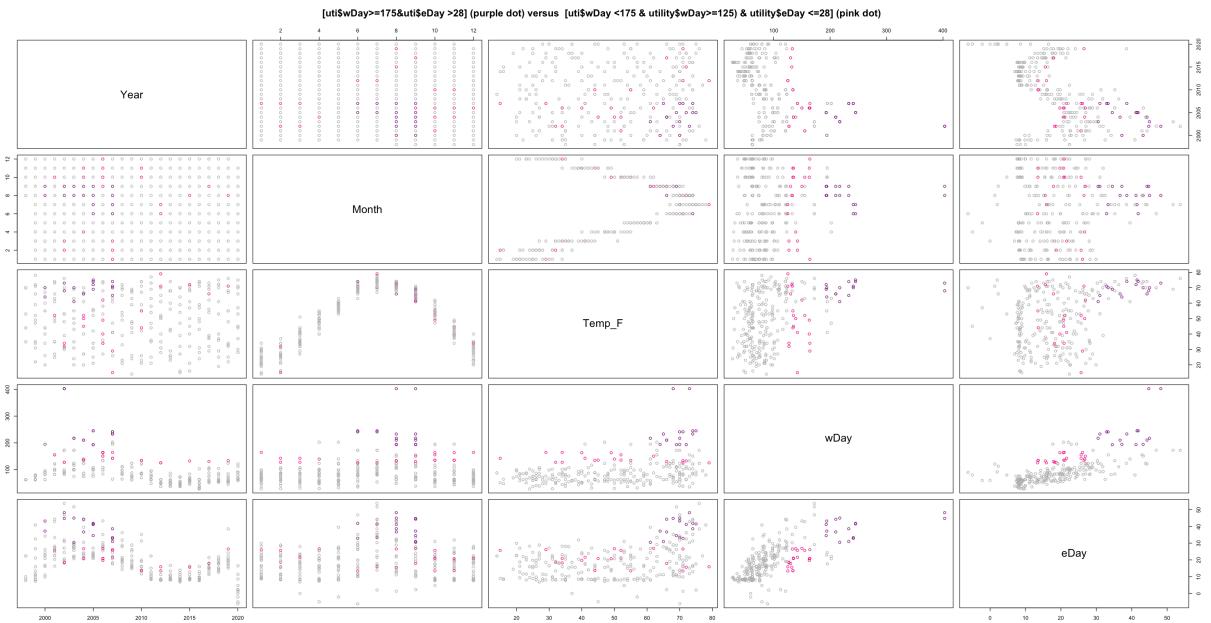


```
In [121]: pairs(data = utility,-Year+Month+Temp_F+wDay+eDay,col = ifelse( (utility$wDa
```

SOME OBSERVATION for scatter plot

If look at pair w/Day and Temp_F on [r=4,c=3], obviously high temp_F tend have more waters use. And w/Day vs e/Day, on [r =4,c =4], hard to see with any graphs, but with scatter plot, I can confirm that more e/Day relevant to use higher w/Day. And if you're not convince, let look at Temp_F vs e/Day [r= 5,c = 3], seem like those with high w/Day sit at high Temp_F and e/Day. And already know high Temp_F is high w/Day -> high w/Day might be roughly similar to high e/Day

.Also [r=1, c=2] say that more w/Day in month 7,8,9. which make sense if look back parcoord, higher temp_F tend to at those months, also more pink(125<= w/Day <=175 on month 6,10), so hotter on 7,8,9. Also at [r= 4, c=1] ,more w/Day before 2017 or 2018.



In []:

