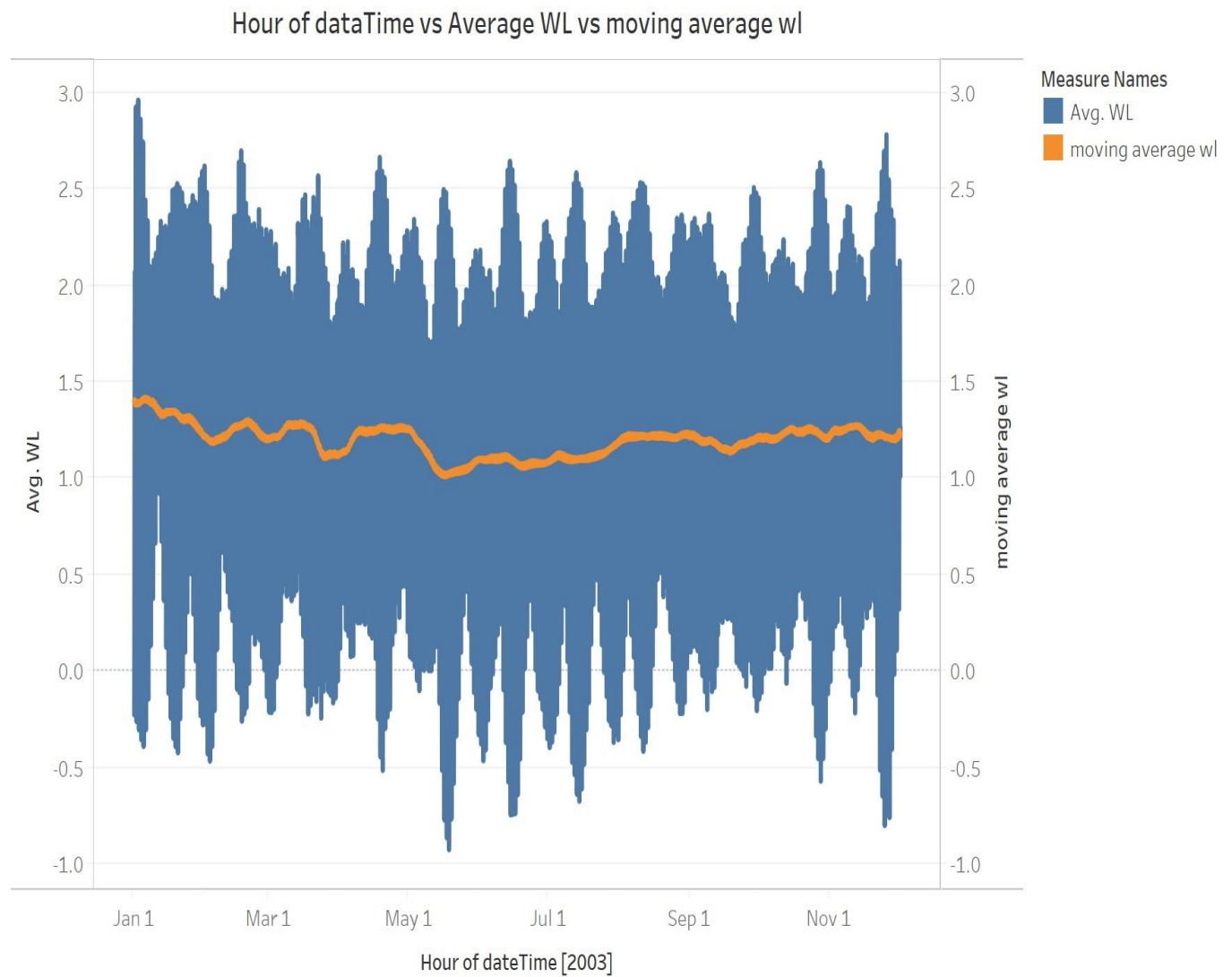


Comparison of Tableau and R for analysis of the Portland Water Level dataset

Figure 1 Tableau graph



The trends of Avg. WL and moving average wl for dateTime Hour. Color shows details about Avg. WL and moving average wl.

Tableau:

dateTime

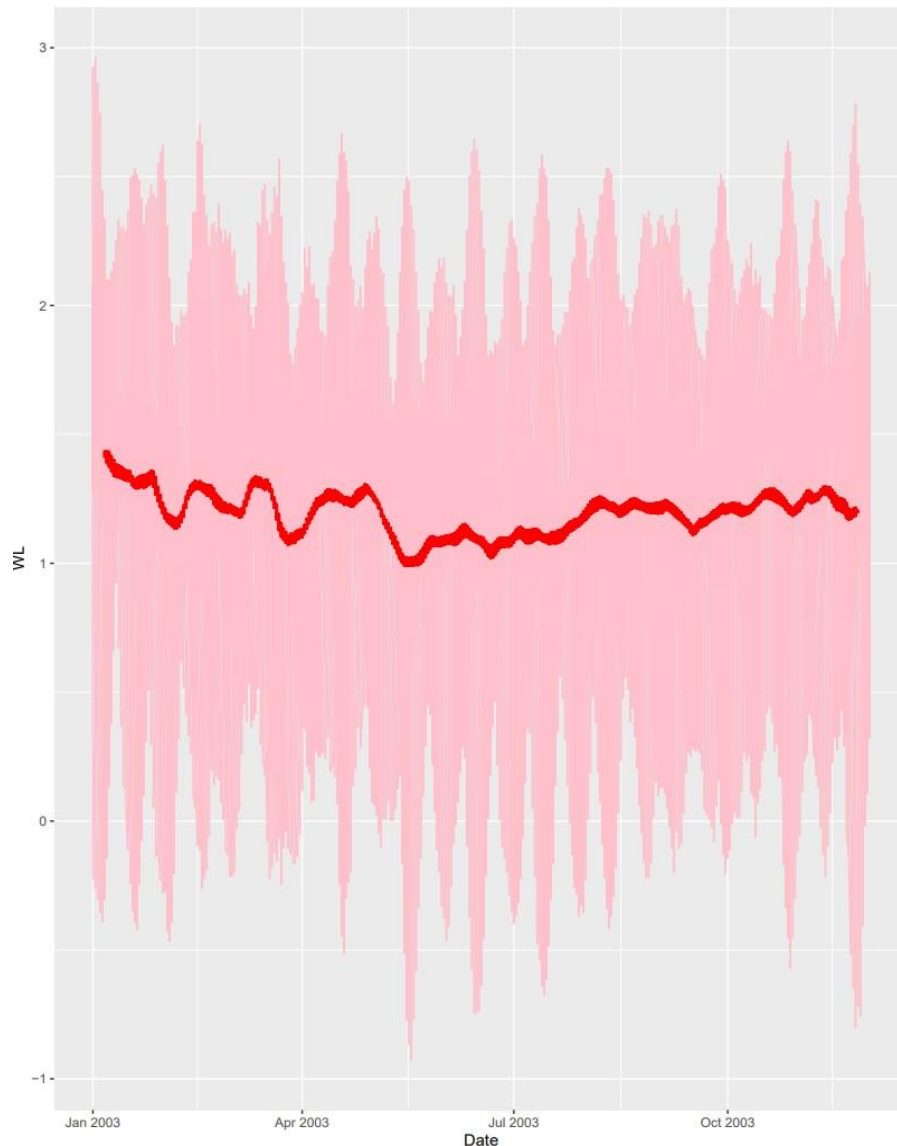
e:

```
DATETIME(STR([Date]) + " "  
+ STR(DATEPART('hour', [Time])) + ":"  
+ STR(DATEPART('minute', [Time])) + ":"  
+ STR(DATEPART('second', [Time]))
```

moving average wl:

`WINDOW_AVG(SUM([WL]), -190,190)`

Figure 2 R window average graph



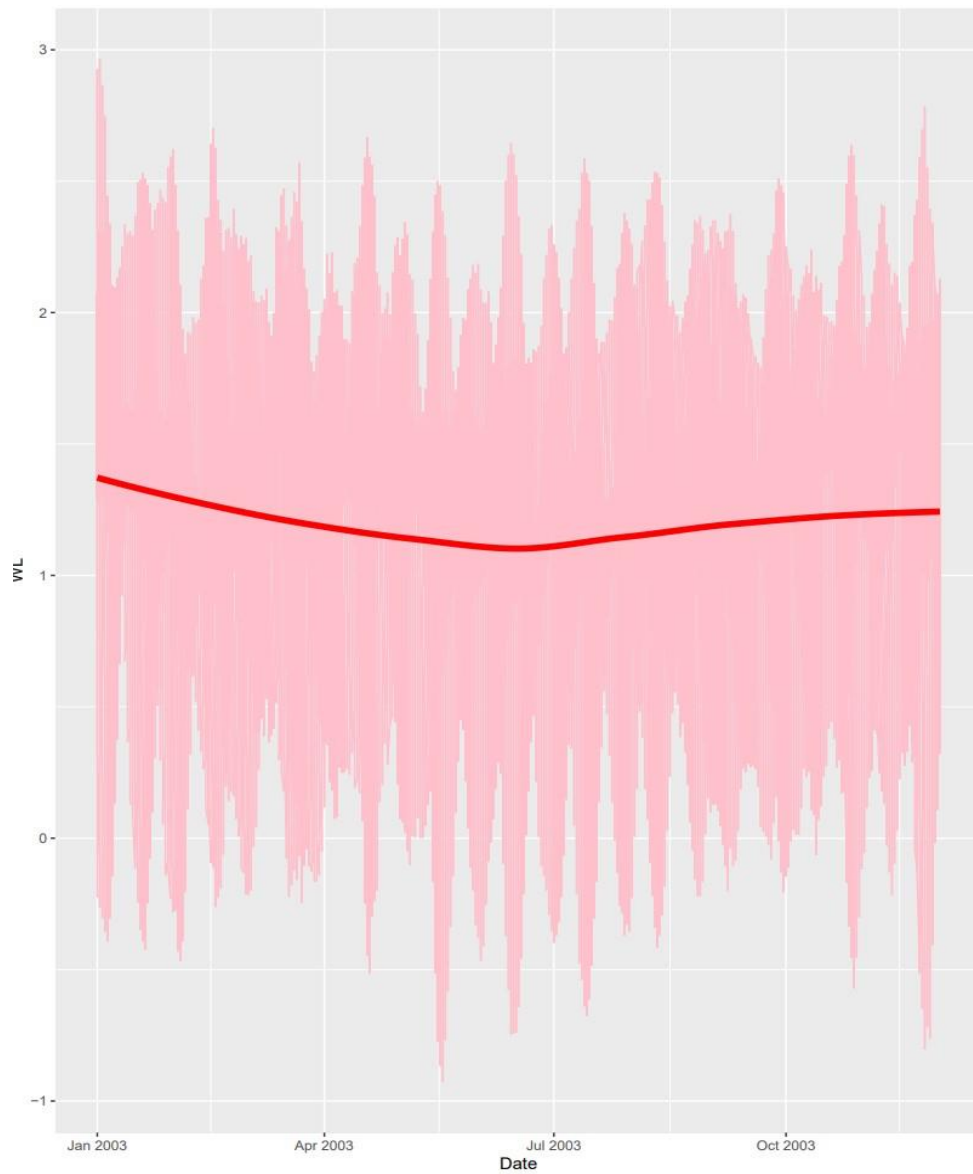
Both graphs created by Tableau (Figure 1) and created by R window average

(Figure 2) have the same shapes of the smoothing average WL and average WL.

The figures show that January had the highest average water level and May had the lowest average water level. The average water level fluctuated between -1 and +3

through the year period. The moving average of WL fluctuated between around 1 and 1.5. Figure 3, below created by R loess has a different smoothing moving average shape compared to those of Figure 1 and Figure 2. The R loess smoothing curve in Figure 3 does not show the detailed fluctuation as the two figures above. Instead, it is much more smoothing than them. But it shows the same changes trend of the WL including decreasing from January with the highest WL to May with the lowest WL and then increasing until the end of the year. Also, Figure 3 shows the same change range of the smoothing average WL as 1 to 1.5.

Figure 3 R loess

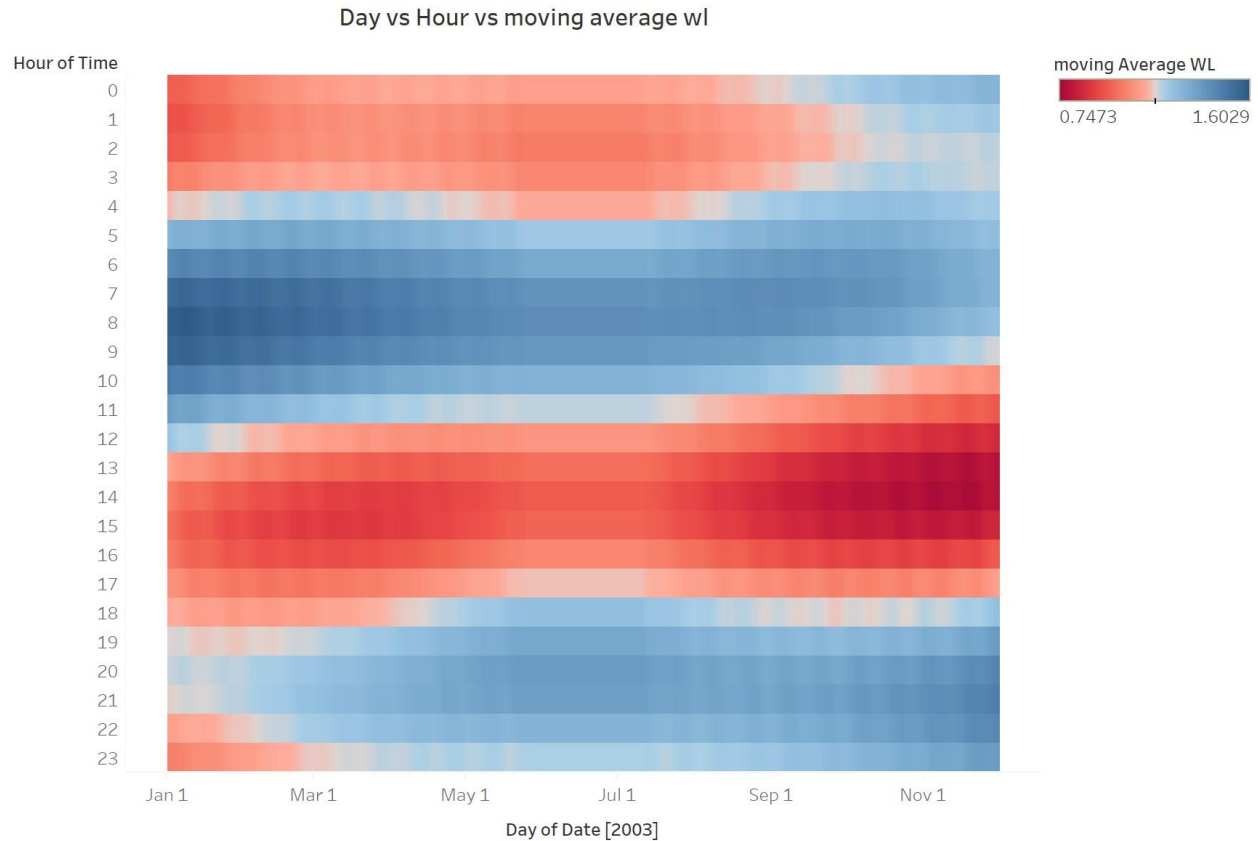


R code:

```
setwd("C:/Users/Yingping Li/Documents/csc465/hw3")
ds = read.table("PortlandWaterLevel2003.csv", sep=";",
header=T) head(ds)
#ds[order(as.Date(ds$Date, format="%m/%d/%Y")),]
#ds$dateTime = paste(ds$Date, ds$Time, sep=" ")
#ds$dateTime = strptime(ds$dateTime, format="%m/%d/%Y
%H:%M") ds$Date = as.Date(ds$Date, "%m/%d/%Y")
```

```
# 3a-R weights = rep(1/280,280) ds$weightedAverage =  
filter(ds$WL, weights, sides=2) p = ggplot(ds, aes(Date,  
WL)) p + geom_line(color="pink") +  
geom_line(size=1, color="red", aes(x=Date,  
y=weightedAverage))
```

```
# 3a-R(lfit) lFit = loess(ds$WL ~  
as.numeric(ds$Date)) p =  
ggplot(ds, aes(Date, WL)) p +  
geom_line(color="pink") +  
geom_line(size=2, color="red", aes(x=ds$Date, y=predict(lFit)))
```



Date Day for each Time Hour. Color shows moving Average WL.

The figure above graphed by Tableau shows that the lowest average water level is between September and December at time interval around 12 to 16, which shows in the dark red area in the figure. The hour period from 10 to 17 over the year and hour of the time interval from 0 to 4 over January to August have a relatively lower water level, which shows in the red area of the figure. And Hour from 5 to 10 over the year and hour from 19 to 24 over March to December have a relative higher moving average WL, which shows in the blue area in the figure. The highest moving average WL is from 7 to 9 over January to March, which shows in the dark blue area in the figure.

Company.csv dataset visulization

Figure 4 budget vs company vs division

