ThulasiRam_RuppaKrishnan_HW6

```
library(ggplot2)
library(reshape2)
library(stringr)
library(scales)
library(plyr)
library(lubridate)
```

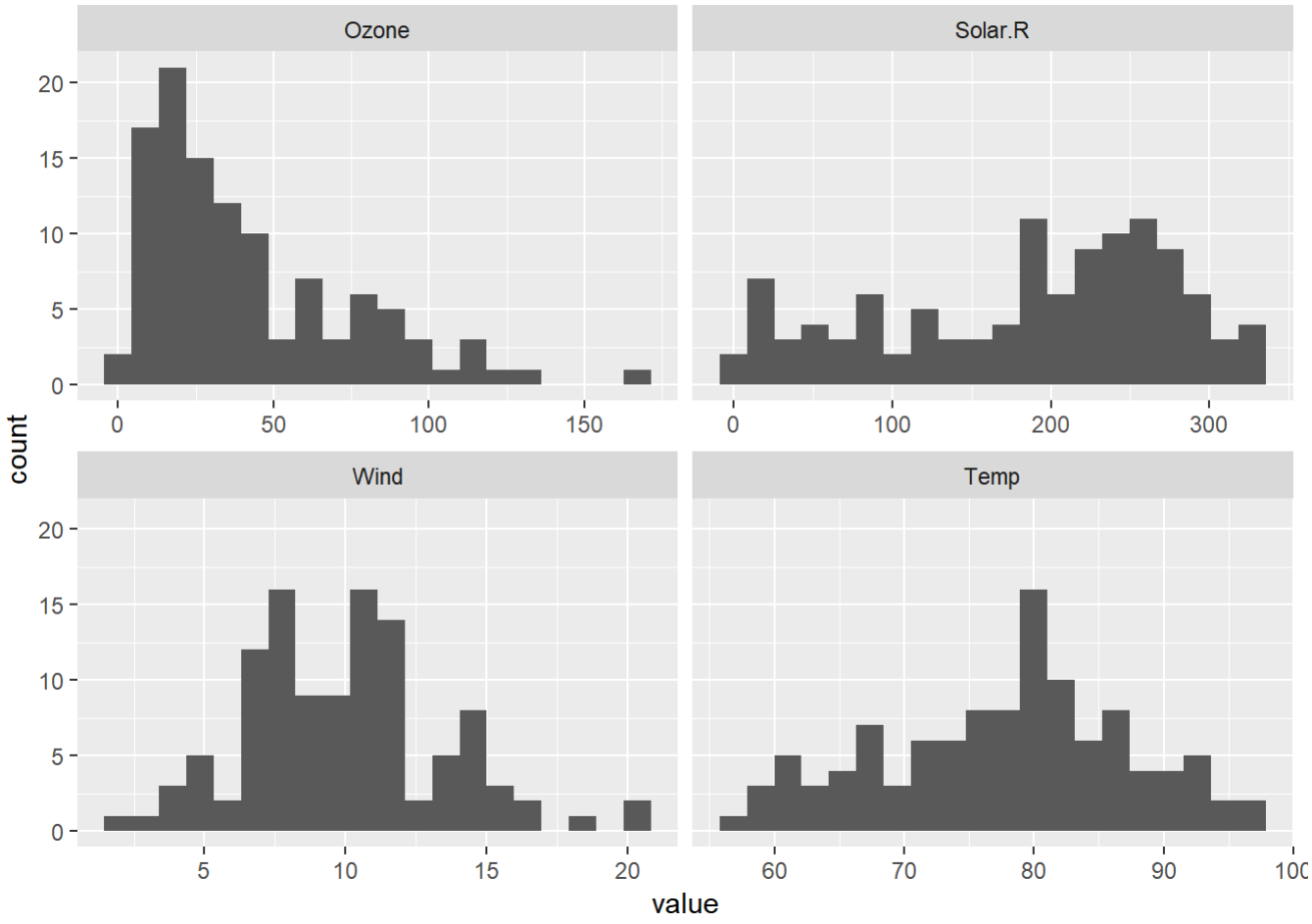
```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:plyr':
##
##     here
```

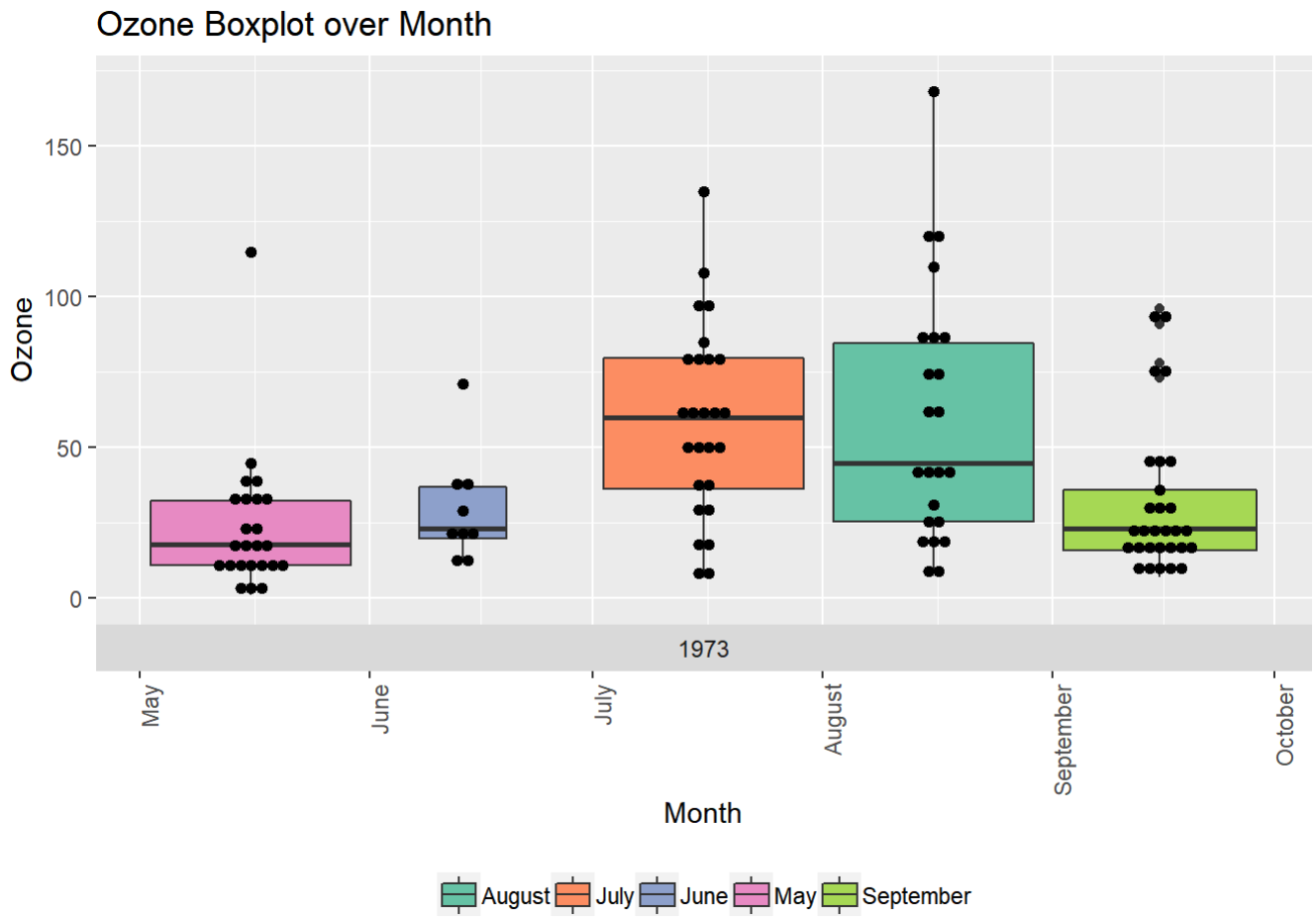
```
## The following object is masked from 'package:base':
##
##     date
```

```
# Step  1:  Load    the data
my_aq<- airquality

# Step  2:  Clean   the data
my_aq<-na.omit(my_aq)
my_aq$Month<-as.factor(my_aq$Month)
my_aq$Day<-as.factor(my_aq$Day)
my_aq<-cbind.data.frame(my_aq,"date"=as.Date(gsub(" ","",paste(str_pad(my_aq$Month,2,side="left"
,pad = "0"),"-",str_pad(my_aq$Day,2,side="left",pad = "0"),"-1973")),"%m-%d-%Y"))
my_aq.m <- melt(my_aq,id.vars = "date", measure.vars = c("Ozone", "Solar.R","Wind","Temp"))
my_aq.m <- ddply(my_aq.m, .(variable), transform, rescale = rescale(value))
```

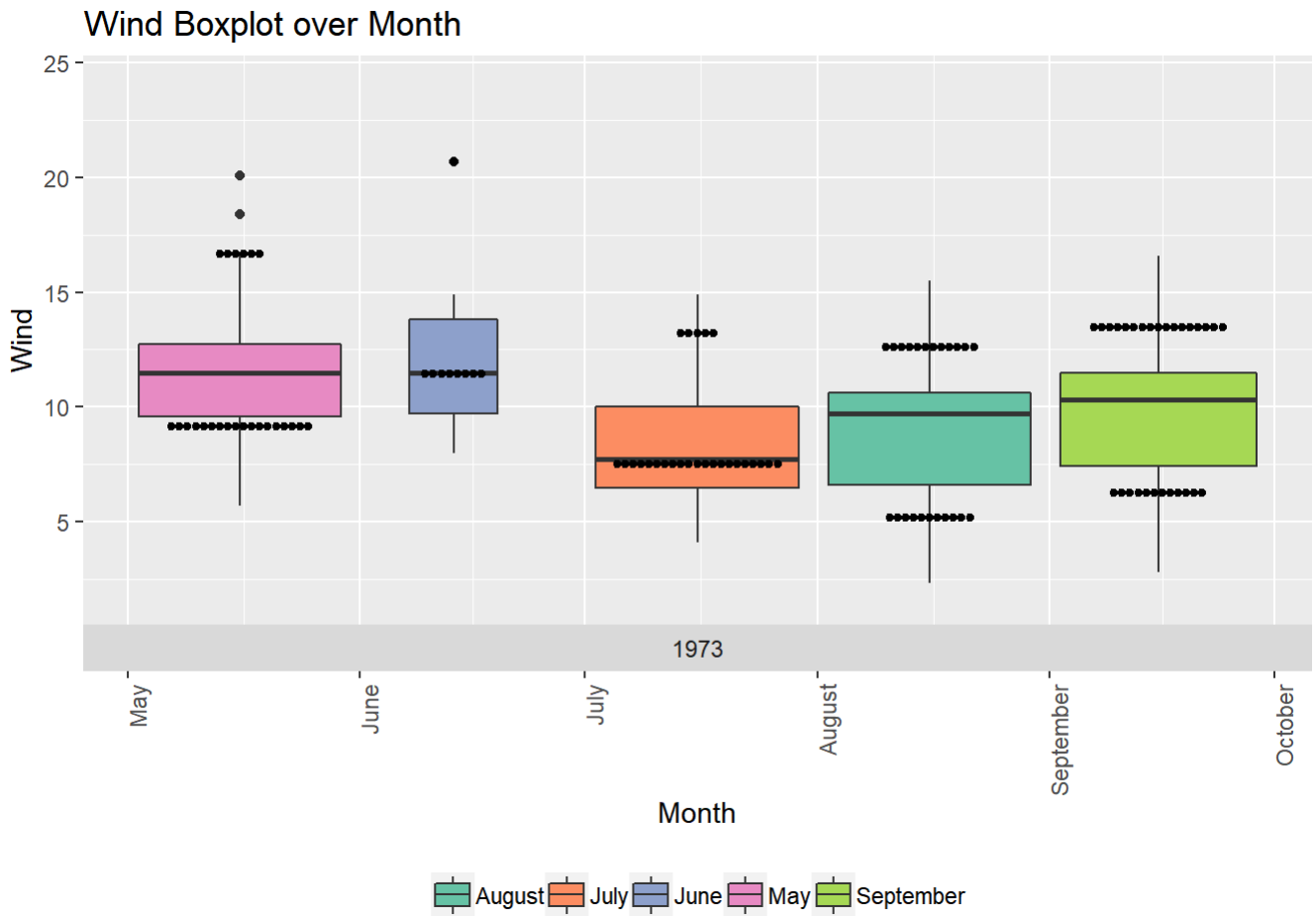
```
# Step  3:  Understand  the data    distribution

# Histograms    for each    of  the variables

ggplot(data = melt(my_aq,measure.vars = c("Ozone", "Solar.R","Wind","Temp")), mapping = aes(x =
 value)) + geom_histogram(bins = 20) + facet_wrap(~variable, scales = 'free_x')
```



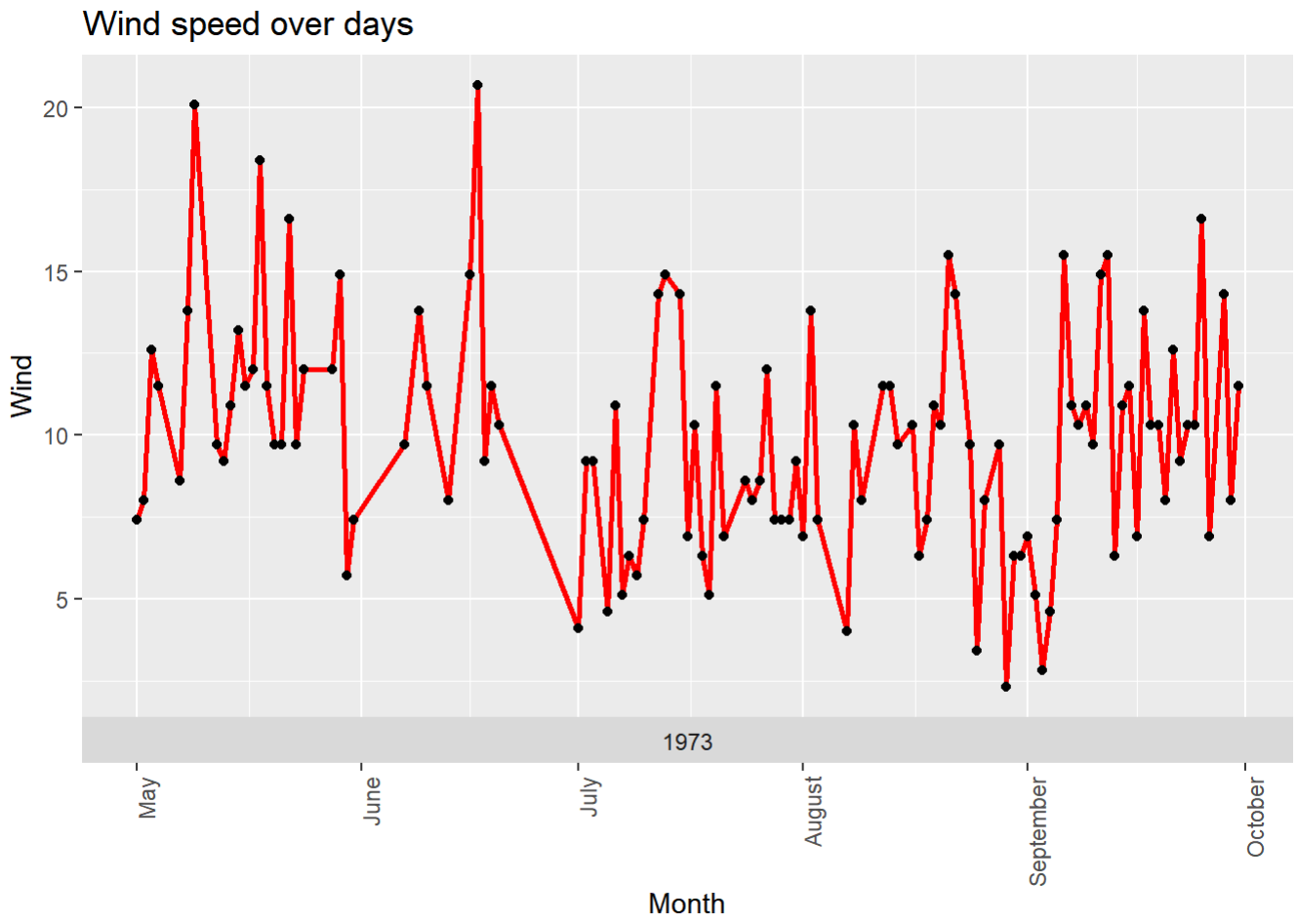

```
# Boxplot   for Ozone
ggplot(my_aq,aes(x=date , y=Ozone, group=Month ,fill=format.Date(date,"%B"))) + geom_boxplot() +
 geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, binwidth = 7,fill="Black") +scale_fil
l_brewer(palette="Set2") +scale_x_date(labels = date_format("%B")) +  facet_grid(~ year(date), s
pace="free_x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = ele
ment_text(angle = 90, hjust = 1)) +
   labs(x = "Month",title = "Ozone Boxplot over Month")  + theme(legend.title=element_blank())
```



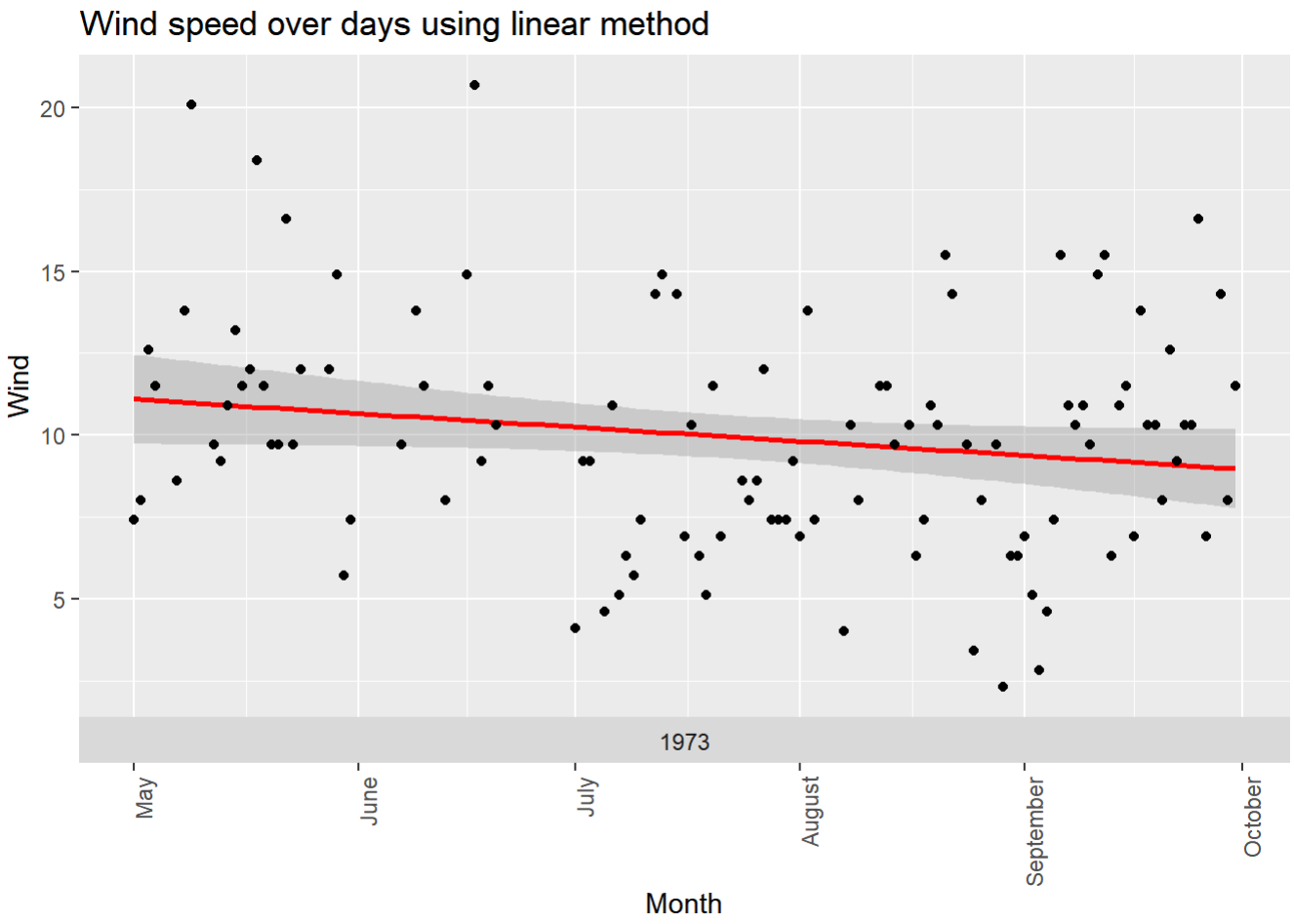

```
# Boxplot for   wind    speed
ggplot(my_aq,aes(x=date , y=Wind, group=Month ,fill=format.Date(date,"%B"))) + geom_boxplot() +
 geom_dotplot(binaxis='y', stackdir='center', dotsize=0.05, binwidth = 7,fill="Black") +scale_fi
ll_brewer(palette="Set2") +scale_x_date(labels = date_format("%B")) +  facet_grid(~ year(date),
 space="free_x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = e
lement_text(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Wind Boxplot over Month")  + theme(legend.title=element_blank())
```



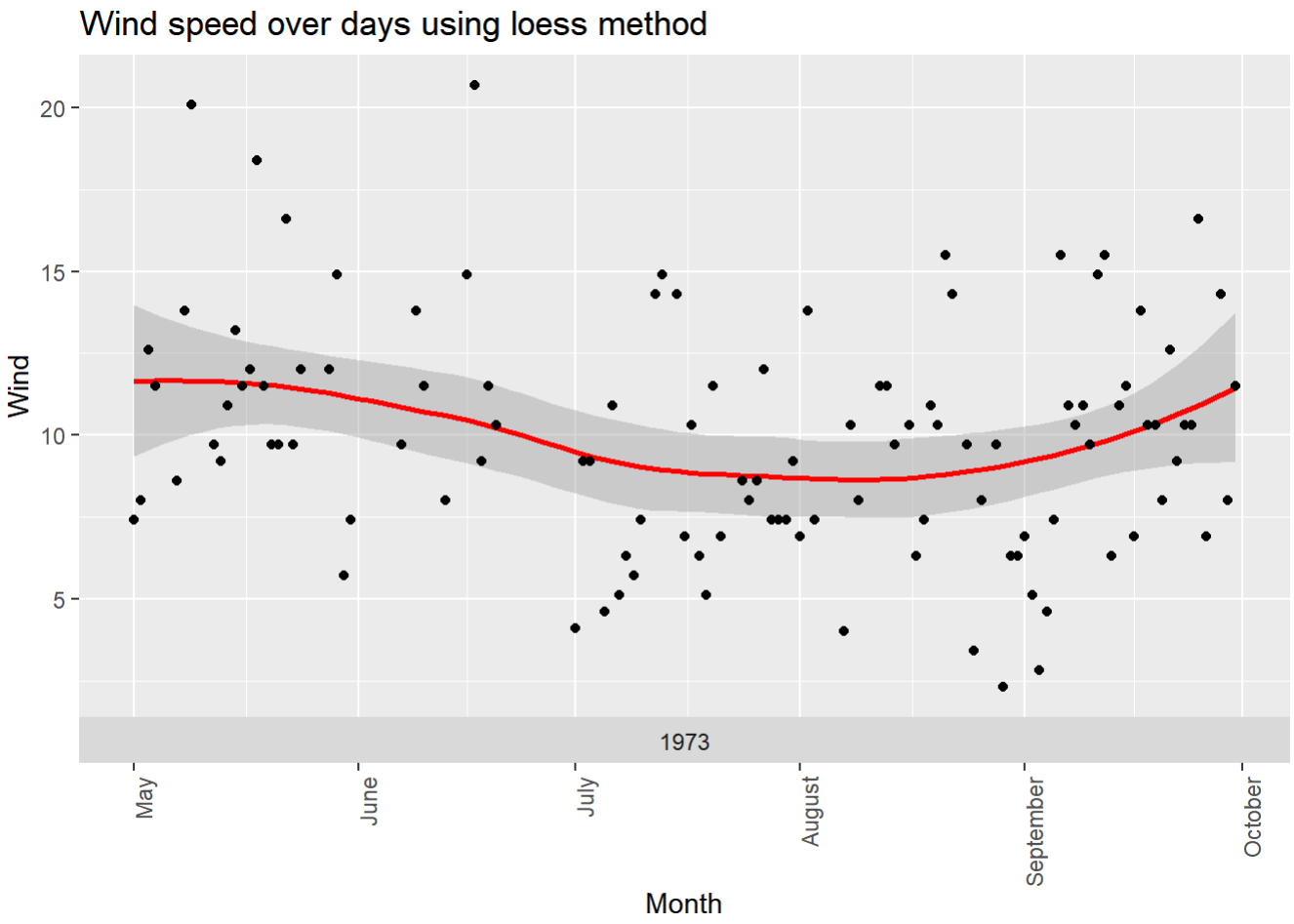

```
# Step  3:  Explore how the data    changes over    time

# Wind data over time
ggplot(my_aq,aes(x=date , y=Wind ,group=1)) +   geom_line(color="red",size=1) +scale_x_date(labe
ls = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_x", scales="free_
x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(angle = 90, hjust
 = 1)) +
  labs(x = "Month",title = "Wind speed over days")
```



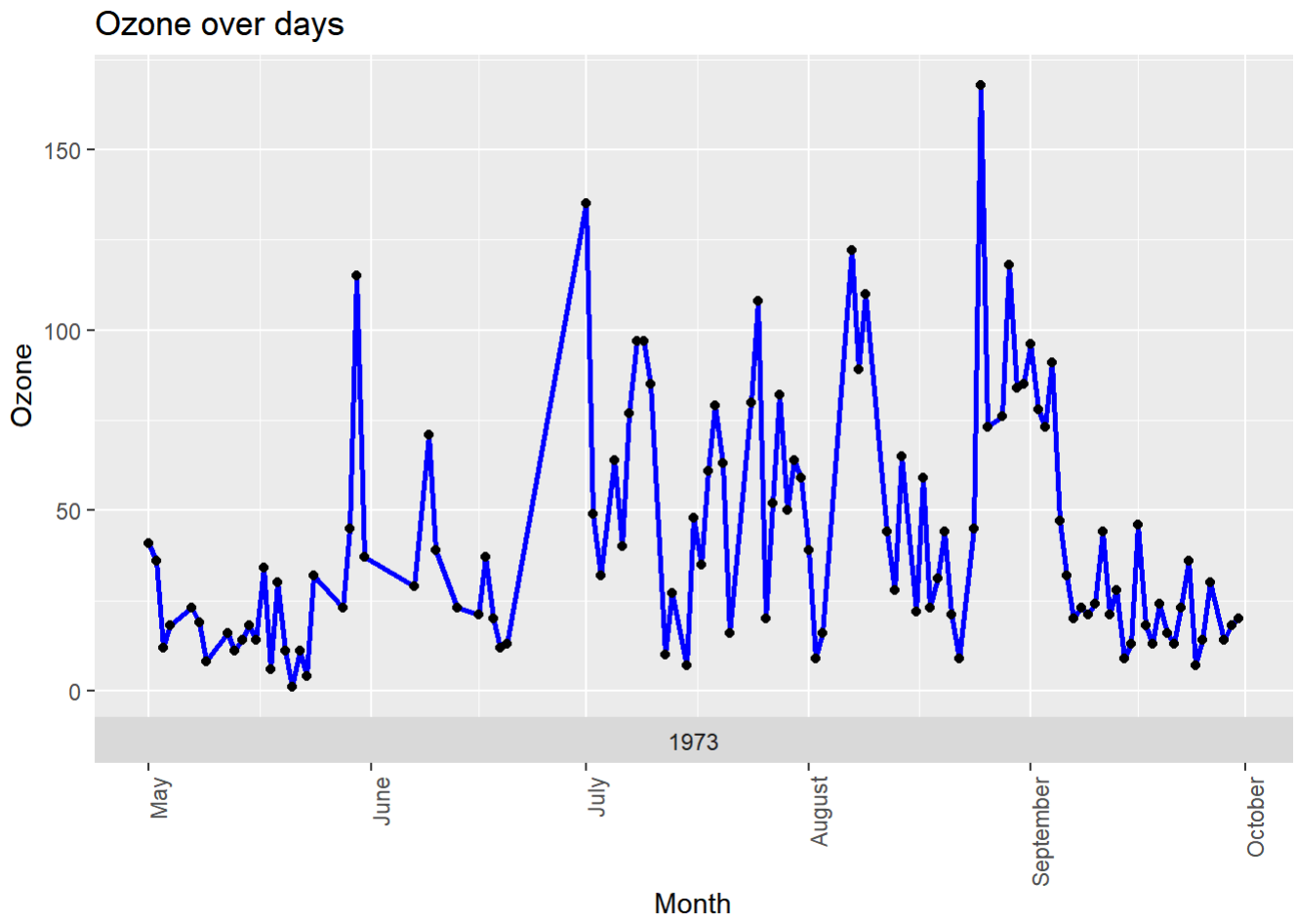

```
ggplot(my_aq,aes(x=date , y=Wind ,group=1)) +   geom_smooth(method="lm",color="red",size=1) +sca
le_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_x",
 scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(angl
e = 90, hjust = 1)) +
  labs(x = "Month",title = "Wind speed over days using linear method")
```



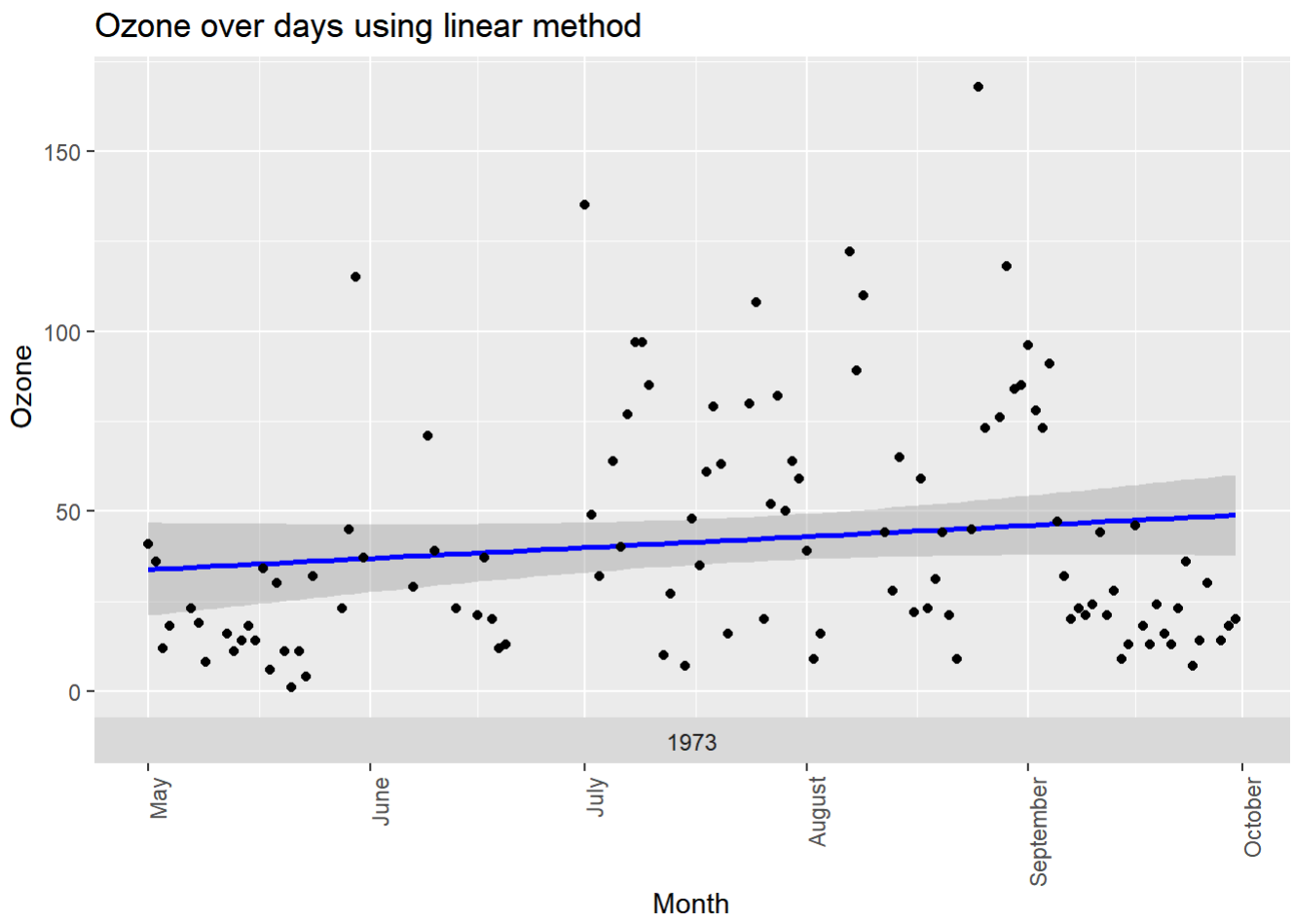

```
ggplot(my_aq,aes(x=date , y=Wind ,group=1)) +   geom_smooth(method="loess",color="red",size=1) +
scale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_
x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(a
ngle = 90, hjust = 1)) +
  labs(x = "Month",title = "Wind speed over days using loess method")
```



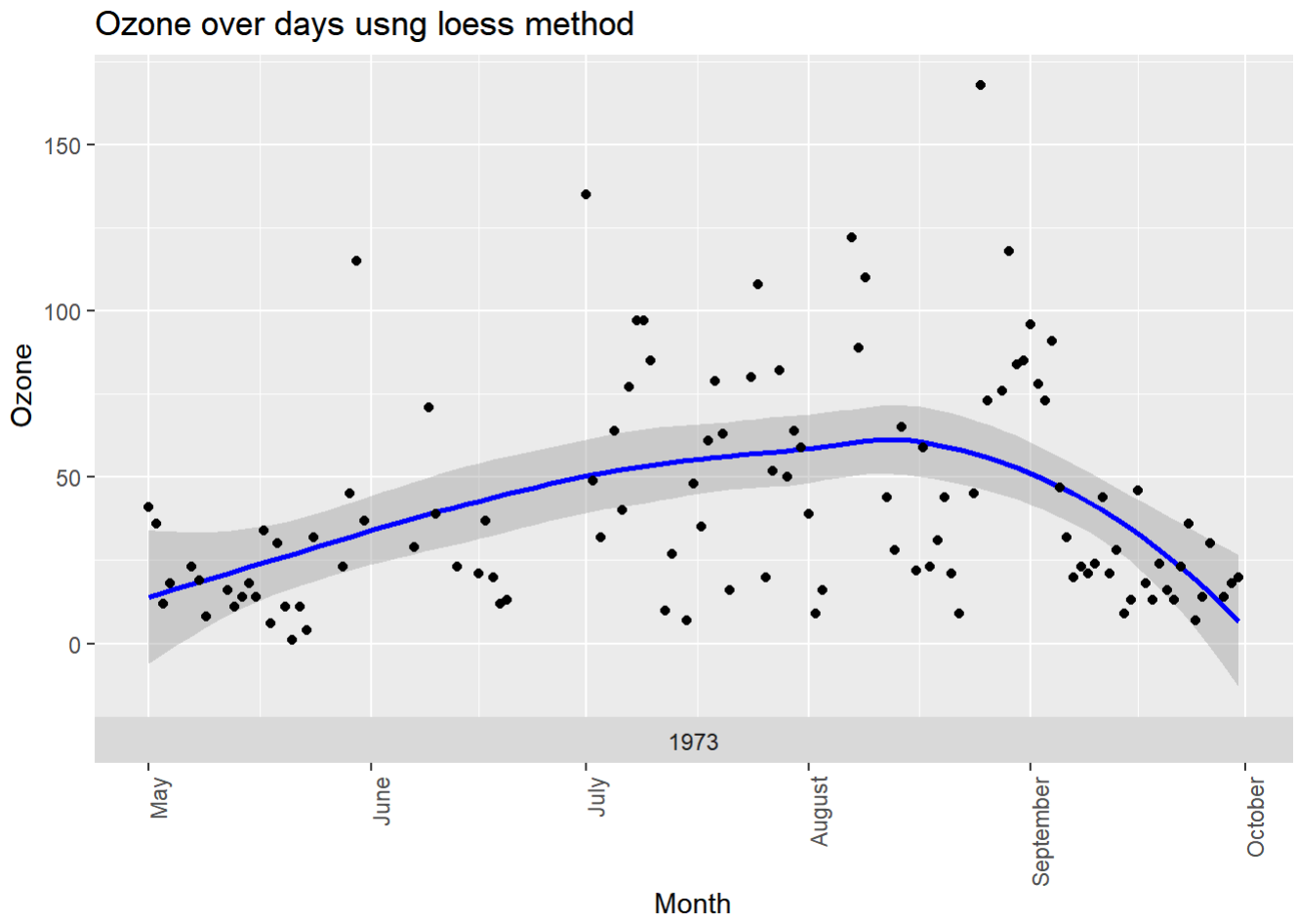

```
# Ozone data over time
ggplot(my_aq,aes(x=date , y=Ozone ,group=1)) +   geom_line(color="blue",size=1) +scale_x_date(la
bels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_x", scales="fre
e_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(angle = 90, hjus
t = 1)) +
  labs(x = "Month",title = "Ozone over days")
```



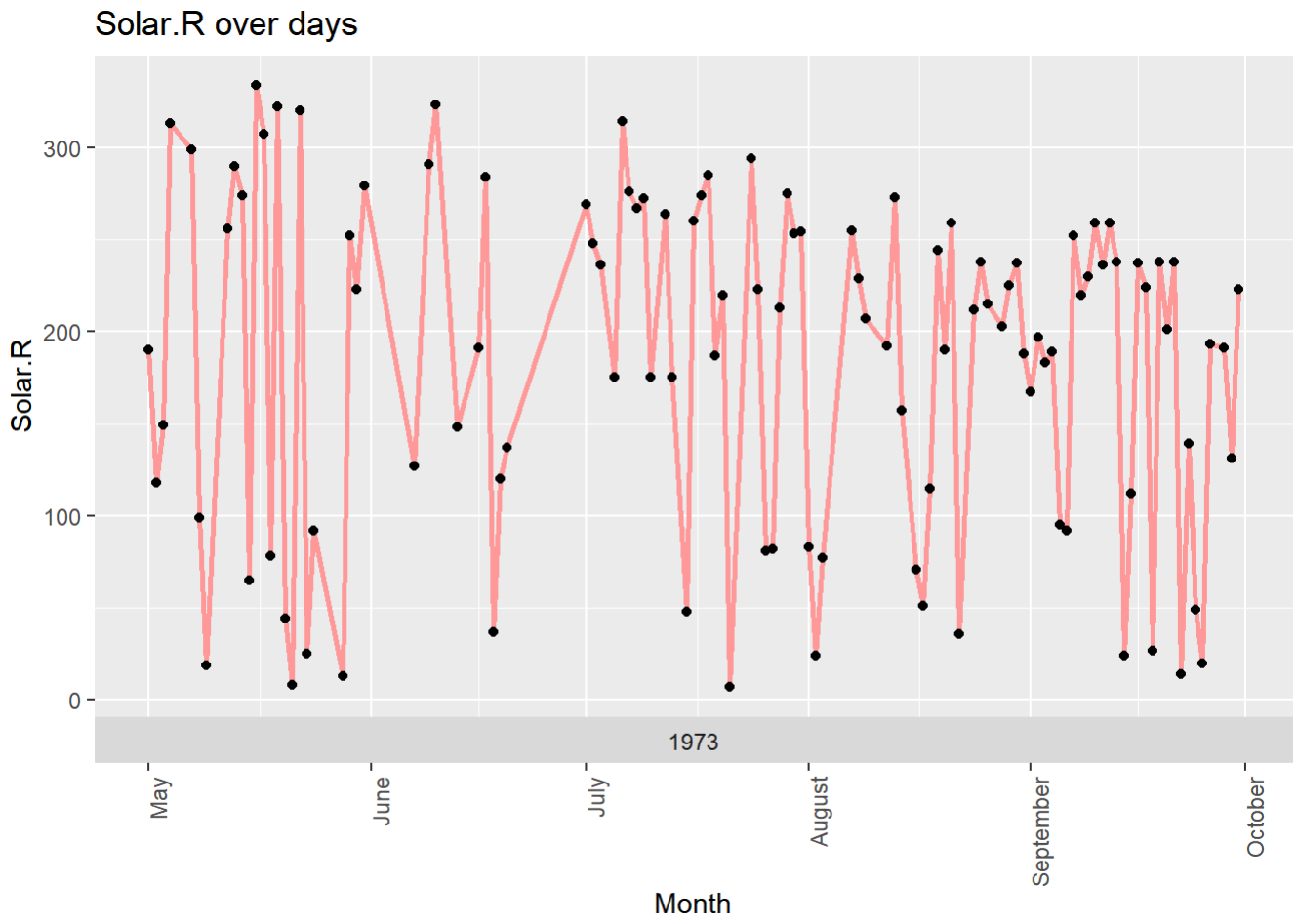

```
ggplot(my_aq,aes(x=date , y=Ozone ,group=1)) +   geom_smooth(method="lm",color="blue",size=1) +s
cale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_
x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(a
ngle = 90, hjust = 1)) +
  labs(x = "Month",title = "Ozone over days using linear method")
```



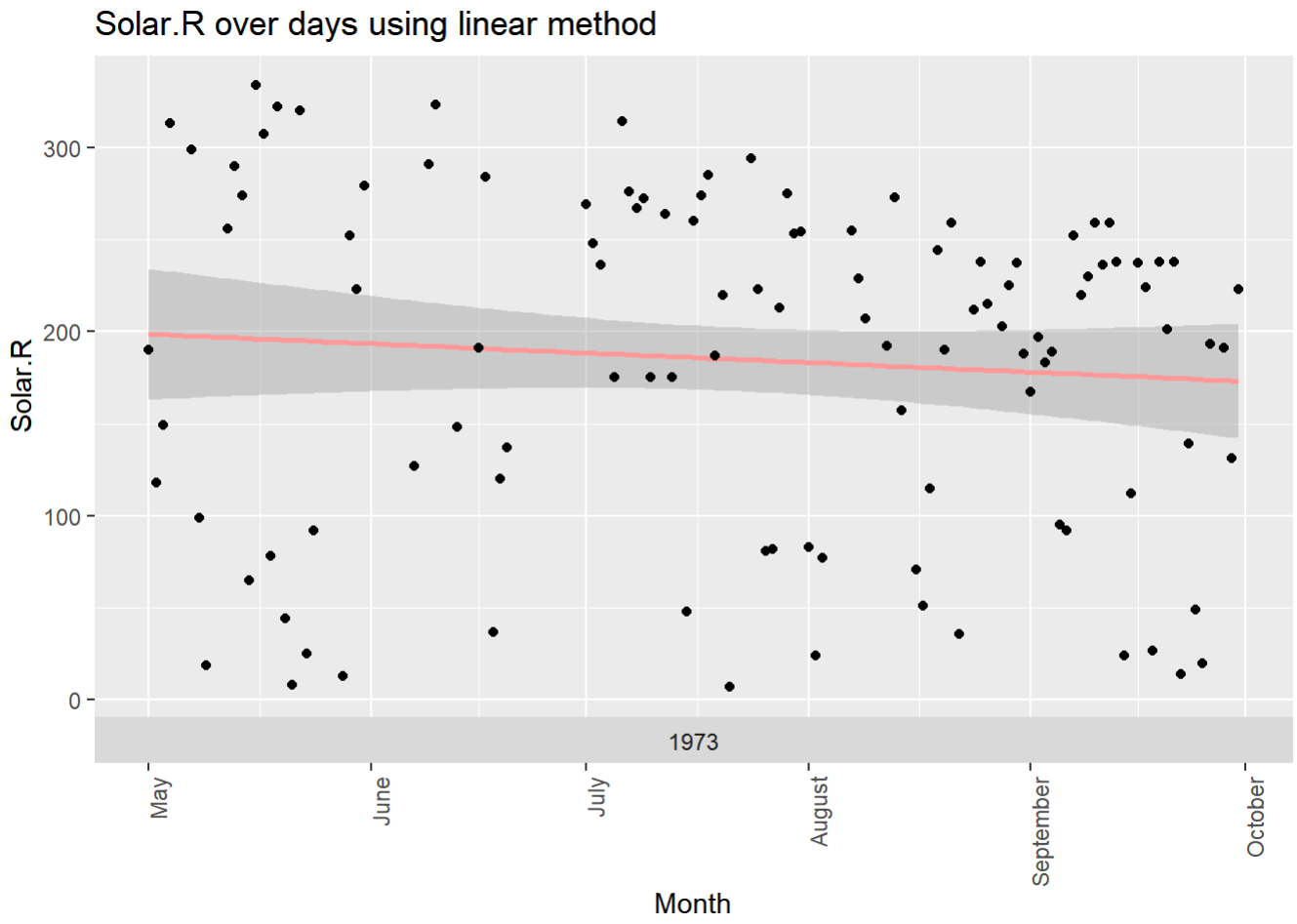

```
ggplot(my_aq,aes(x=date , y=Ozone ,group=1)) +   geom_smooth(method="loess",color="blue",size=1)
 +scale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="fre
e_x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text
(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Ozone over days usng loess method")
```



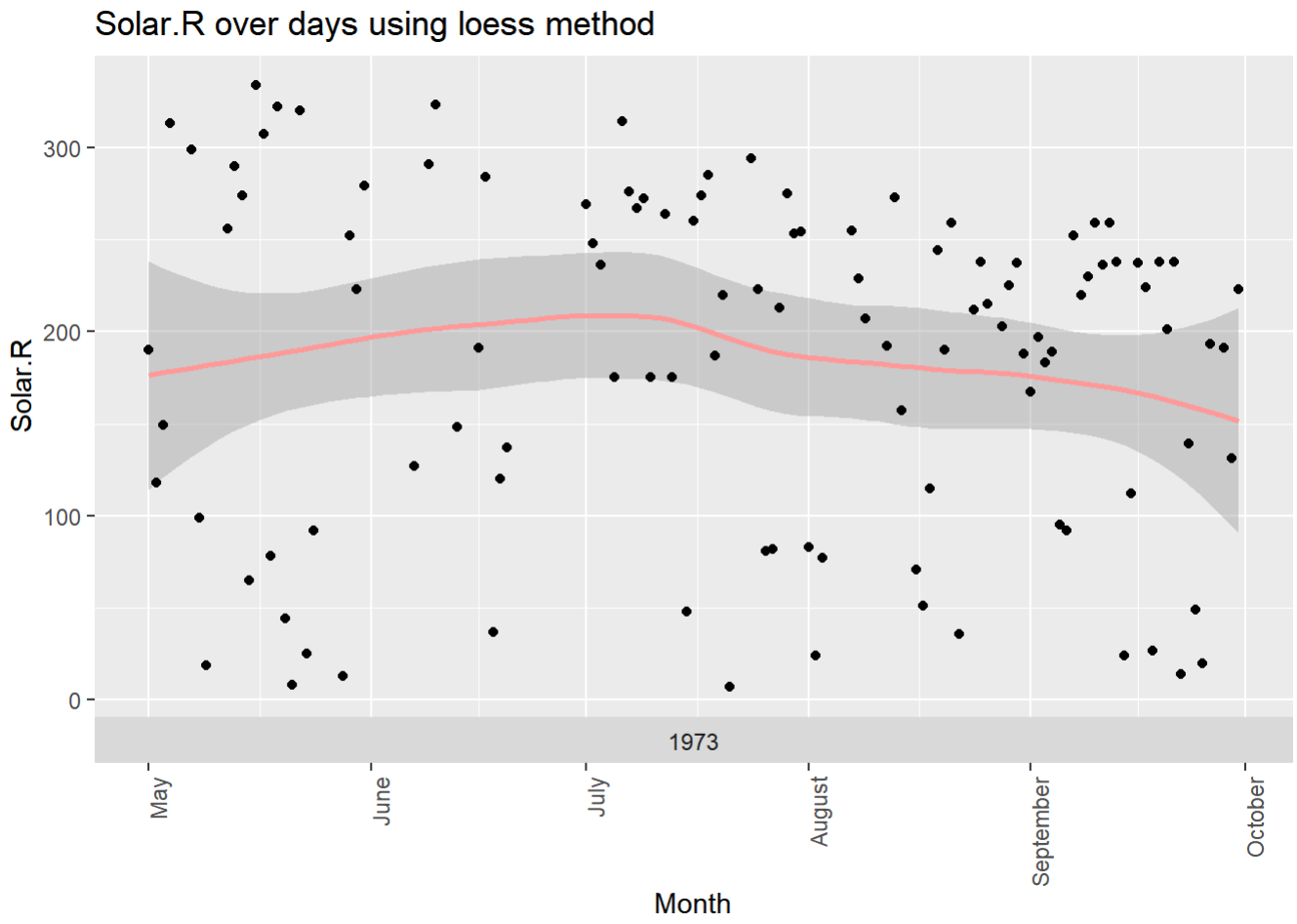

```
# Solar data over time
ggplot(my_aq,aes(x=date , y=Solar.R ,group=1)) +   geom_line(color="#FF9999",size=1) +scale_x_da
te(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_x", scales
="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(angle = 90,
 hjust = 1)) +
  labs(x = "Month",title = "Solar.R over days")
```



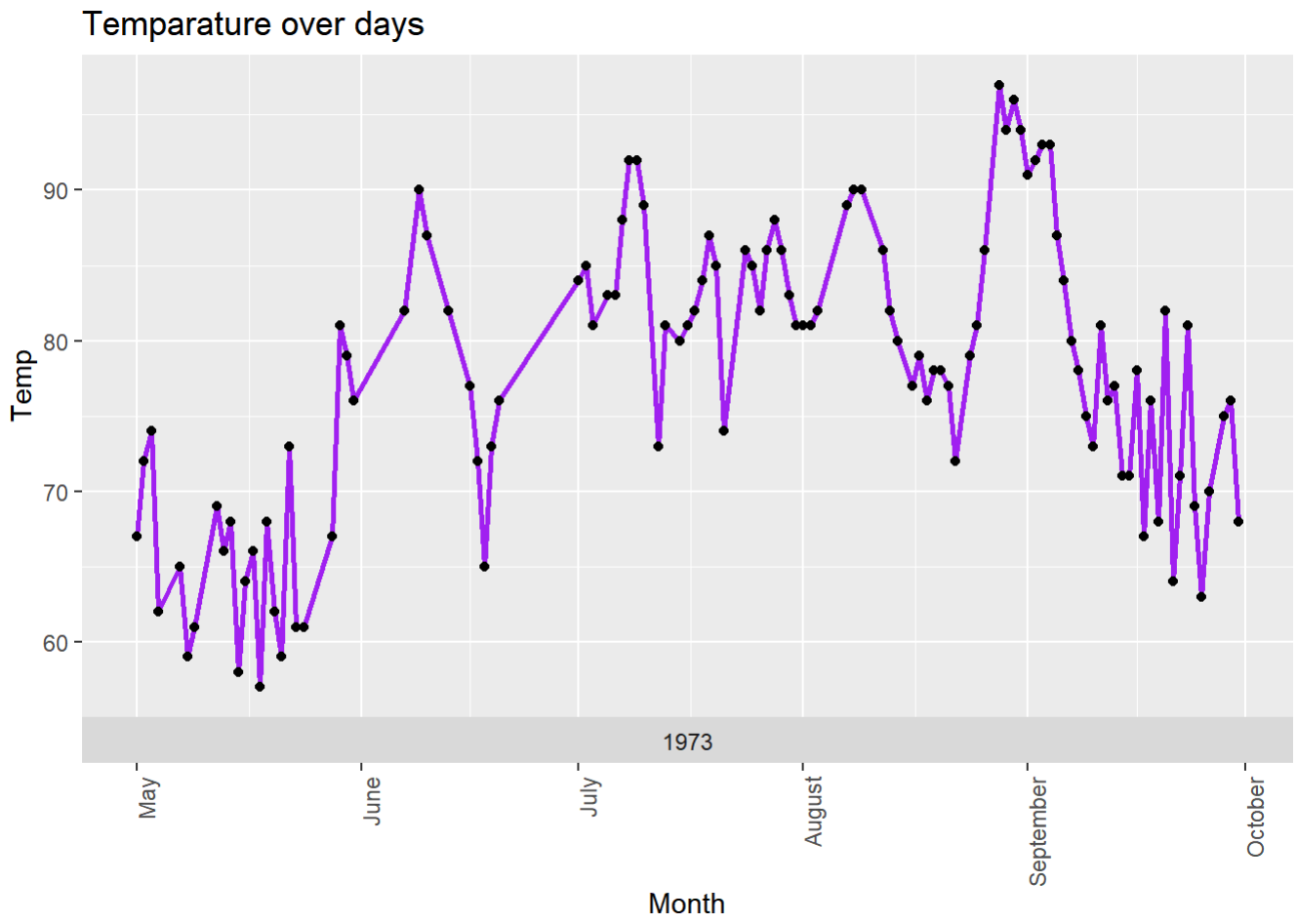

```
ggplot(my_aq,aes(x=date , y=Solar.R ,group=1)) +   geom_smooth(method="lm",color="#FF9999",size=
1) +scale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="f
ree_x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_te
xt(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Solar.R over days using linear method")
```



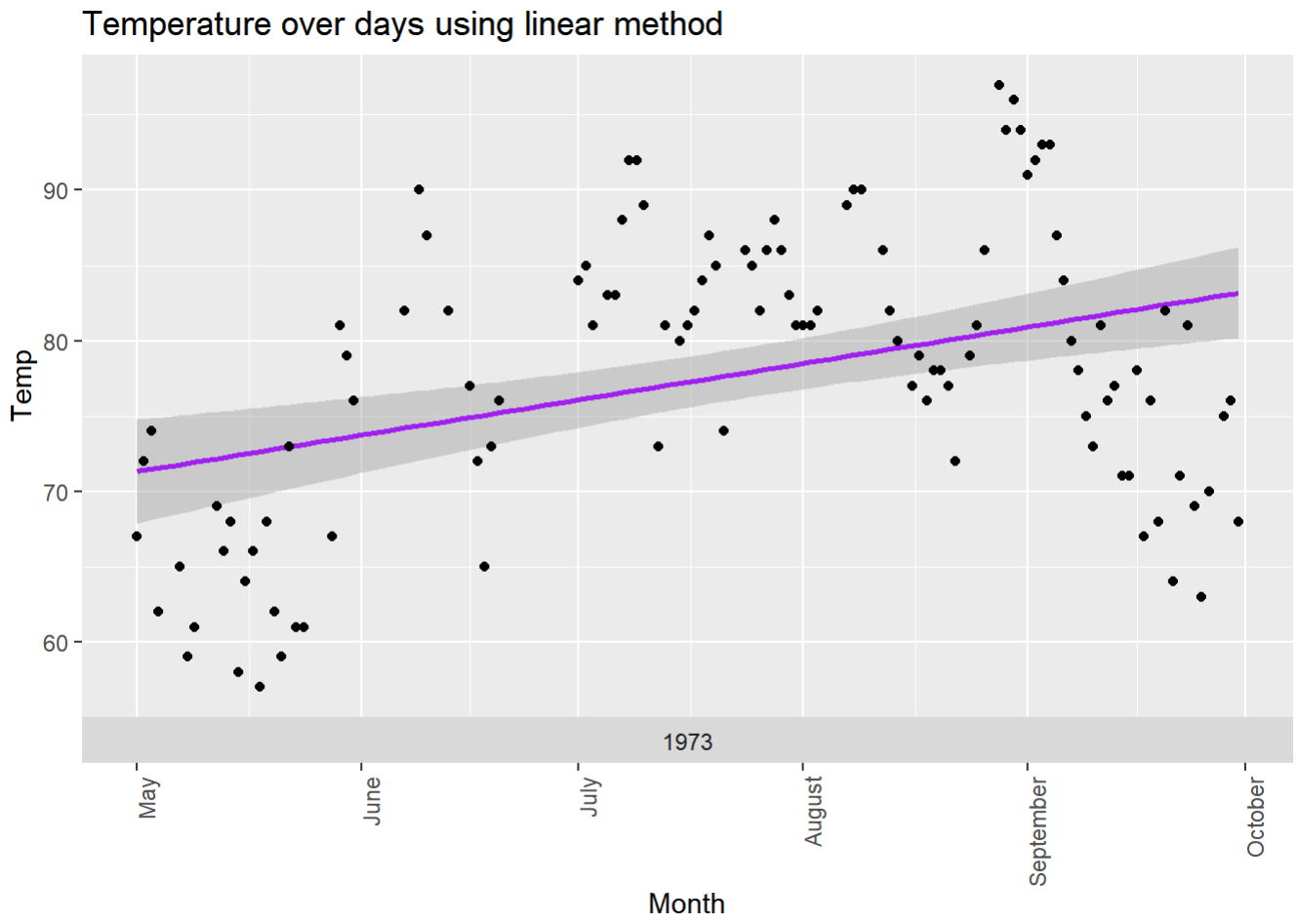

```
ggplot(my_aq,aes(x=date , y=Solar.R ,group=1)) +   geom_smooth(method="loess",color="#FF9999",si
ze=1) +scale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space
="free_x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element
_text(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Solar.R over days using loess method")
```



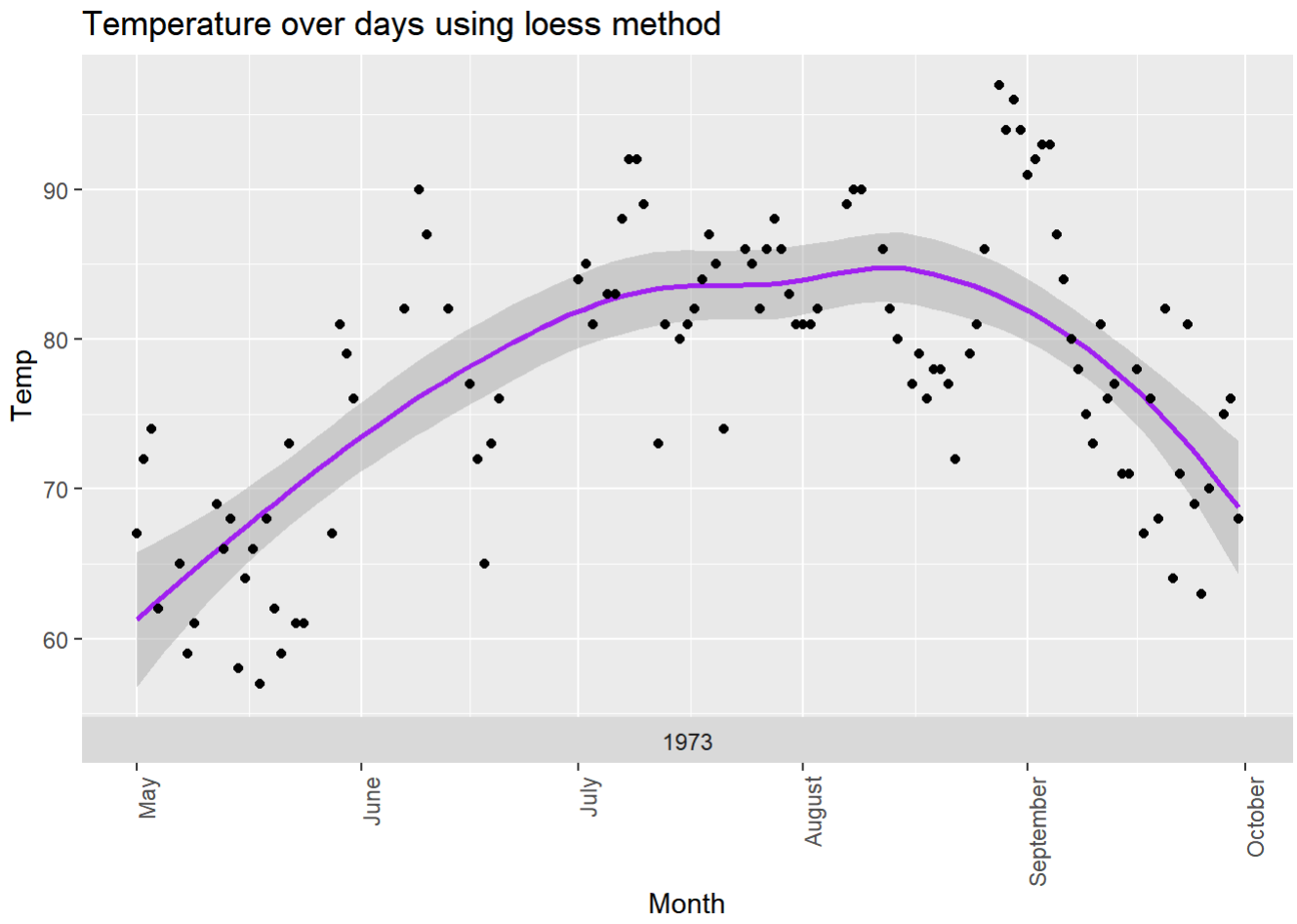

```
# Temp data over time
ggplot(my_aq,aes(x=date , y=Temp ,group=1)) +   geom_line(color="purple",size=1) +scale_x_date(l
abels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_x", scales="fr
ee_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(angle = 90, hju
st = 1)) +
  labs(x = "Month",title = "Temparature over days")
```




```
ggplot(my_aq,aes(x=date , y=Temp ,group=1)) +   geom_smooth(method="lm",color="purple",size=1) +
scale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="free_
x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_text(a
ngle = 90, hjust = 1)) +
  labs(x = "Month",title = "Temperature over days using linear method")
```




```
ggplot(my_aq,aes(x=date , y=Temp ,group=1)) +   geom_smooth(method="loess",color="purple",size=1
) +scale_x_date(labels = date_format("%B")) + geom_point() +  facet_grid(~ year(date), space="fr
ee_x", scales="free_x", switch="x") + theme(legend.position = "bottom",axis.text.x = element_tex
t(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Temperature over days using loess method")
```




```
# Descriptive statistics on different Variable
summary(my_aq$Ozone)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0    18.0    31.0    42.1    62.0   168.0
```

```
summary(my_aq$Wind)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.30    7.40    9.70    9.94   11.50   20.70
```

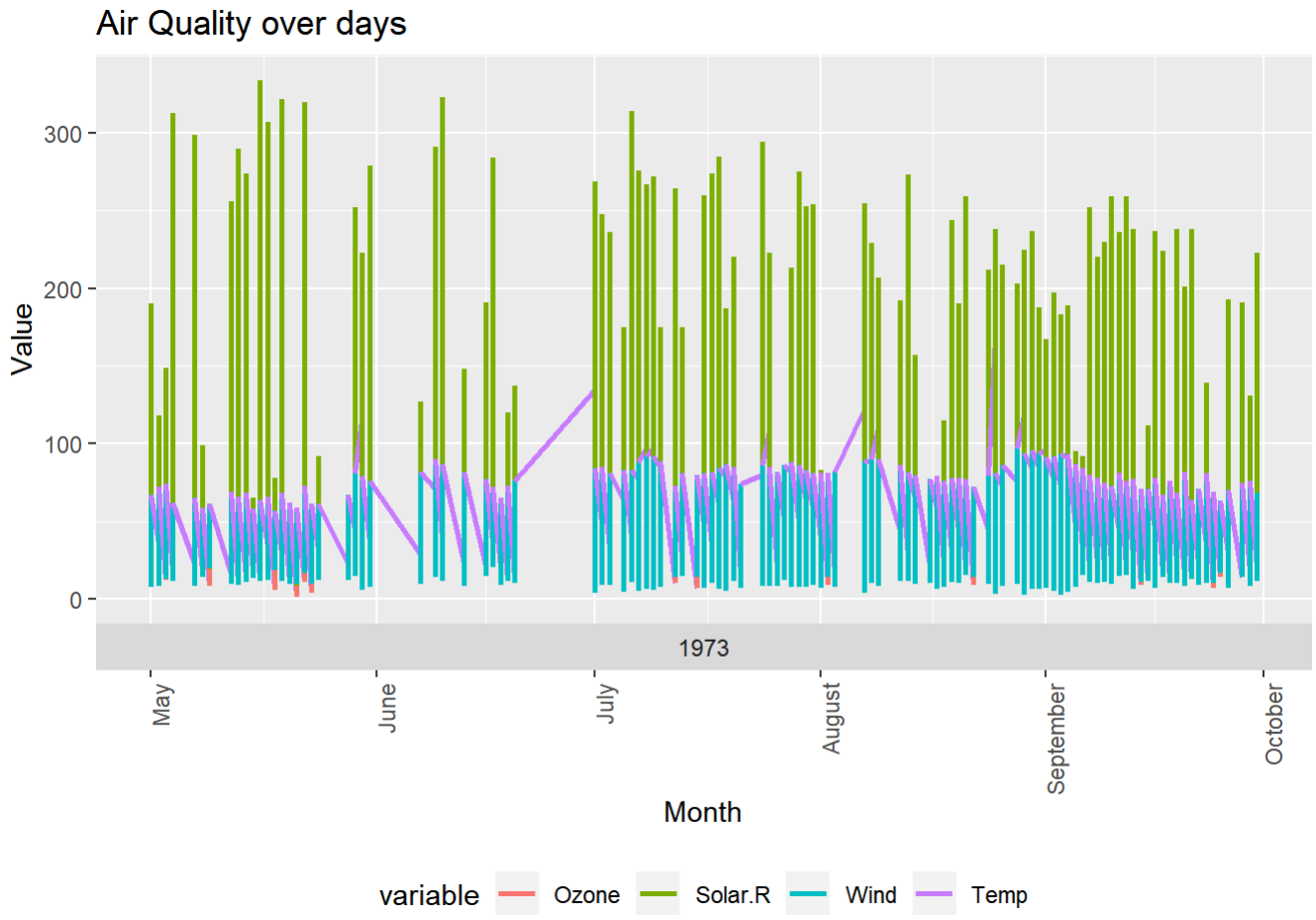
```
summary(my_aq$Temp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   57.00   71.00   79.00   77.79   84.50   97.00
```

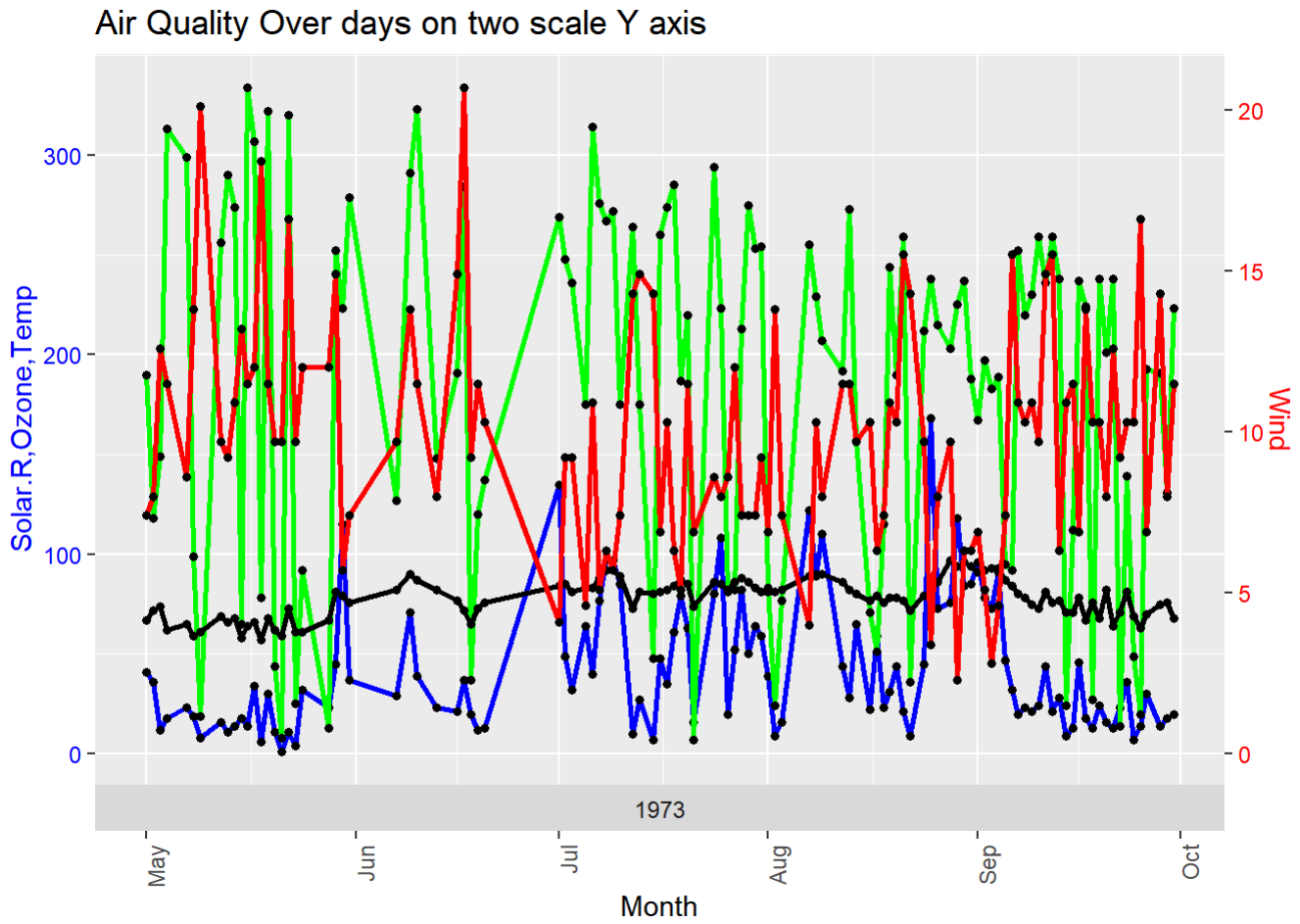
```
summary(my_aq$Solar.R)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     7.0   113.5   207.0   184.8   255.5   334.0
```

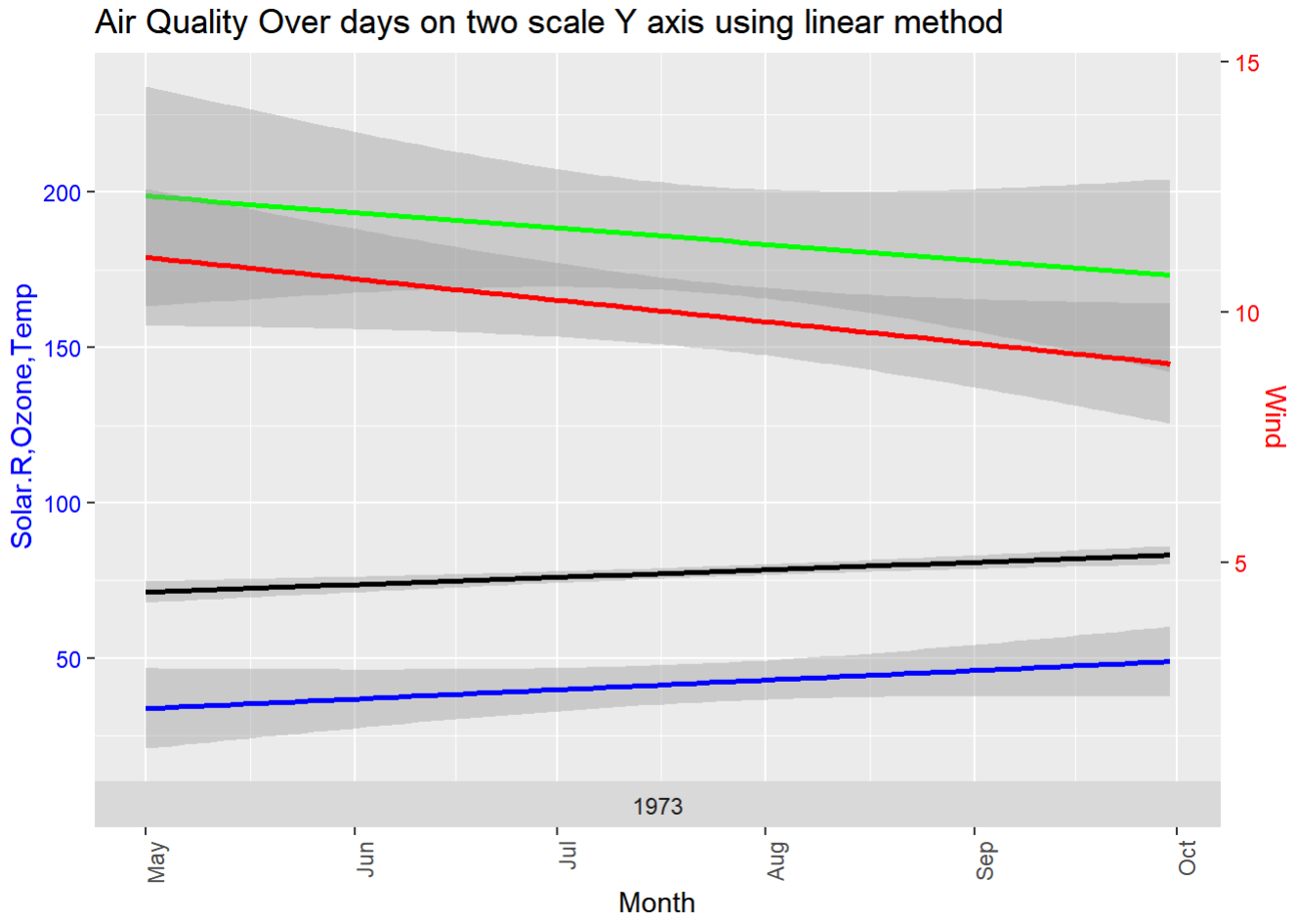
```
# Plotting all four variables without changing scale
ggplot(data = melt(my_aq,id.vars = "date", measure.vars = c("Ozone", "Solar.R","Wind","Temp")),
 aes(x=date,y=value,color=variable,group=1)) + geom_line(size=1) +scale_x_date(labels = date_for
mat("%B"))+  facet_grid(~ year(date), space="free_x", scales="free_x", switch="x") + theme(legen
d.position = "bottom",axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Month", y="Value",title = "Air Quality over days")
```



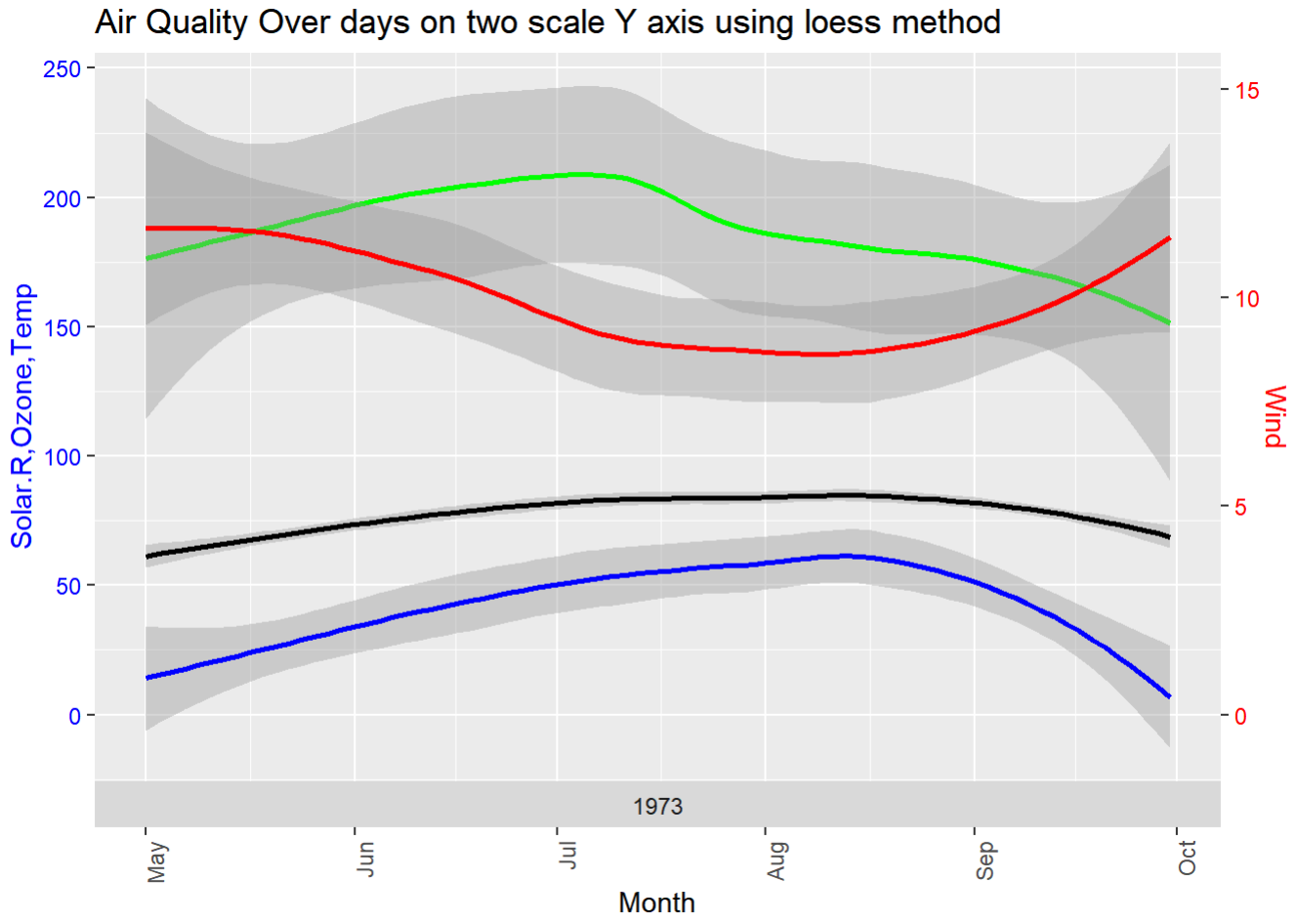
```
# Change the scale for Wind alone to get good visualization on 4 variables together
scaleFactor <- max(my_aq$Solar.R) / max(my_aq$Wind)

# using  geom_line
ggplot(my_aq, aes(x=date)) +
  geom_line(aes(y=Ozone),size=1, col="blue") + geom_point(aes(y=Ozone),size=1.25)+
  geom_line(aes(y=Solar.R), size=1, col="green")+ geom_point(aes(y=Solar.R),size=1.25) +
  geom_line(aes(y=Temp), size=1, col="black") + geom_point(aes(y=Temp),size=1.25)+
  geom_line(aes(y=Wind * scaleFactor), size=1, col="red")+ geom_point(aes(y=Wind * scaleFactor),
size=1.25) +
  scale_y_continuous(name="Solar.R,Ozone,Temp", sec.axis=sec_axis(~./scaleFactor, name="Wind"))
 +
  theme(
    axis.title.y.left=element_text(color="blue"),
    axis.text.y.left=element_text(color="blue"),
    axis.title.y.right=element_text(color="red"),
    axis.text.y.right=element_text(color="red")
  ) +  facet_grid(~ year(date), space="free_x", scales="free_x", switch="x") + theme(legend.posi
tion = "right",axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Air Quality Over days on two scale Y axis") +
    scale_color_discrete(name = "Variable", labels = c("Solar.R", "Ozone","Temp","Wind"))
```



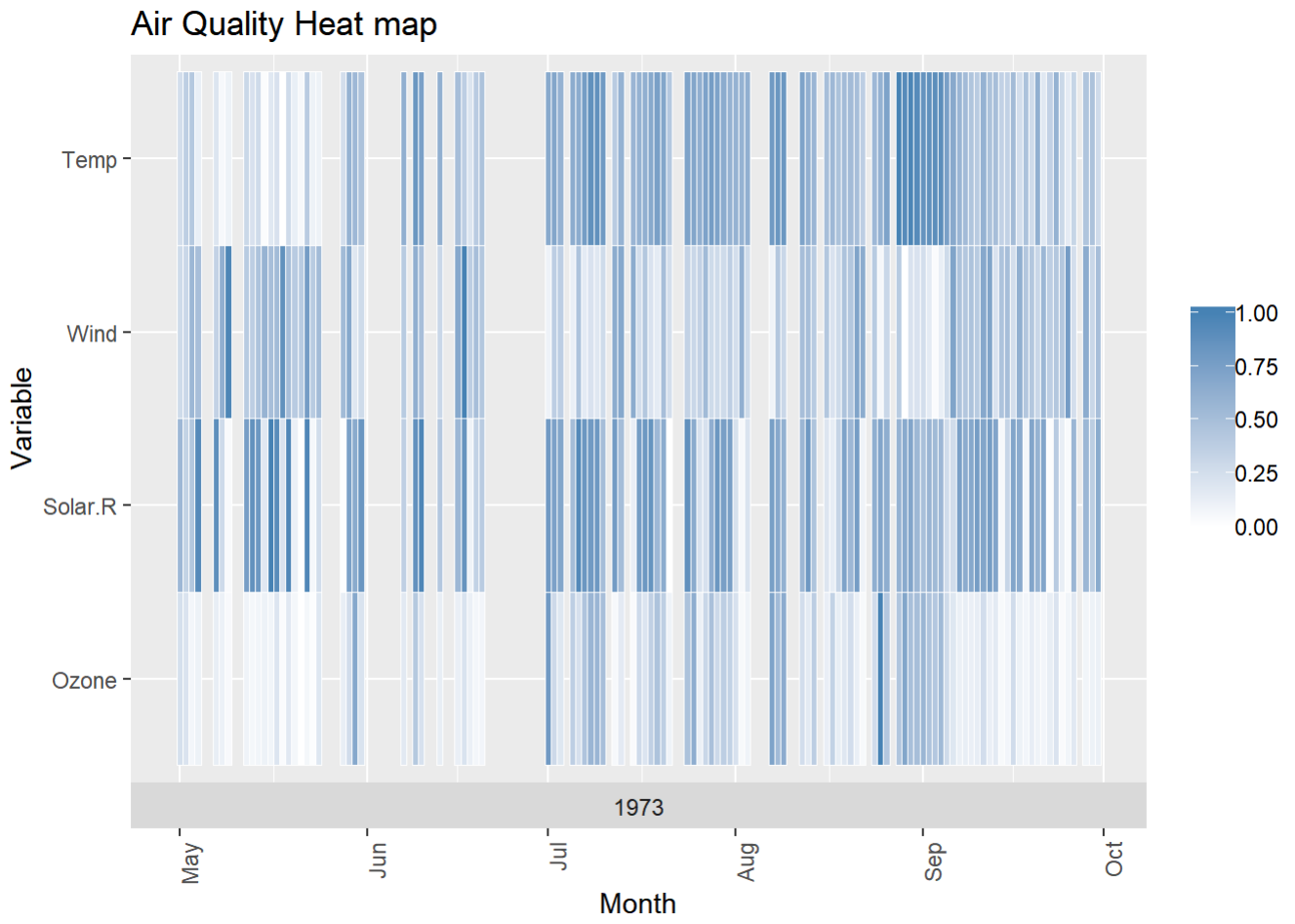
```
# After applying geom_smooth using linear method
ggplot(my_aq, aes(x=date)) +
  geom_smooth(aes(y=Ozone),method="lm",size=1, col="blue") +
  geom_smooth(aes(y=Solar.R),method="lm", size=1, col="green")+
  geom_smooth(aes(y=Temp),method="lm", size=1, col="black") +
  geom_smooth(aes(y=Wind * scaleFactor),method="lm", size=1, col="red")+
  scale_y_continuous(name="Solar.R,Ozone,Temp", sec.axis=sec_axis(~./scaleFactor, name="Wind"))
 +
  theme(
    axis.title.y.left=element_text(color="blue"),
    axis.text.y.left=element_text(color="blue"),
    axis.title.y.right=element_text(color="red"),
    axis.text.y.right=element_text(color="red")
  )+  facet_grid(~ year(date), space="free_x", scales="free_x", switch="x") + theme(legend.posit
ion = "right",axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Air Quality Over days on two scale Y axis using linear method") +
    scale_color_discrete(name = "Variable", labels = c("Solar.R", "Ozone","Temp","Wind"))
```



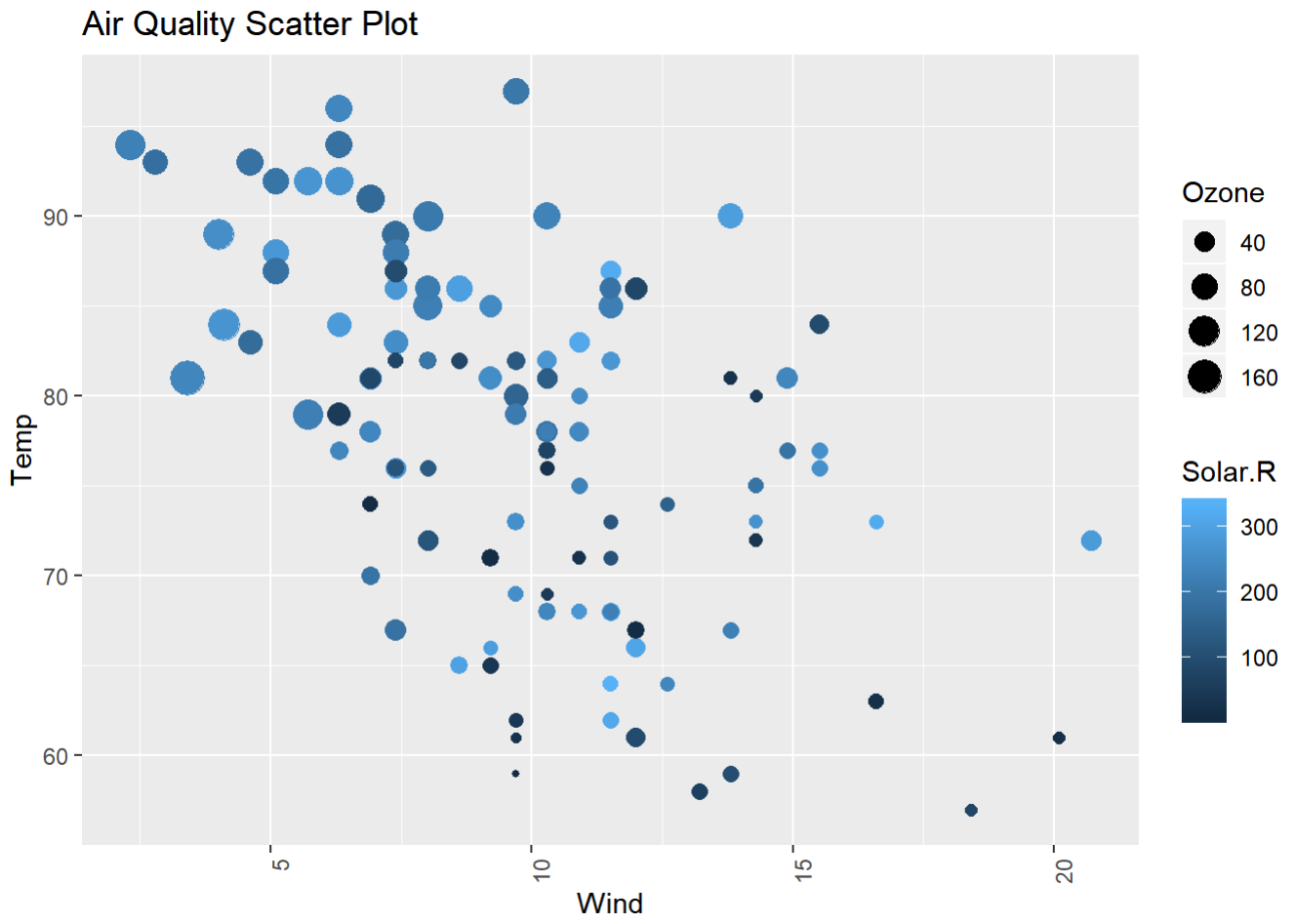
```
# After applying geom_smooth using loess method

ggplot(my_aq, aes(x=date)) +
  geom_smooth(aes(y=Ozone),method="loess",size=1, col="blue") +
  geom_smooth(aes(y=Solar.R),method="loess", size=1, col="green")+
  geom_smooth(aes(y=Temp),method="loess", size=1, col="black") +
  geom_smooth(aes(y=Wind * scaleFactor),method="loess", size=1, col="red")+
  scale_y_continuous(name="Solar.R,Ozone,Temp", sec.axis=sec_axis(~./scaleFactor, name="Wind"))
 +
  theme(
    axis.title.y.left=element_text(color="blue"),
    axis.text.y.left=element_text(color="blue"),
    axis.title.y.right=element_text(color="red"),
    axis.text.y.right=element_text(color="red")
  )+  facet_grid(~ year(date), space="free_x", scales="free_x", switch="x") + theme(legend.posit
ion = "right",axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Month",title = "Air Quality Over days on two scale Y axis using loess method") +
    scale_color_discrete(name = "Variable", labels = c("Solar.R", "Ozone","Temp","Wind"))
```



```
# Step  4:  Look    at  all the data    via a   Heatmap

ggplot(my_aq.m, aes(date, variable)) + geom_tile(aes(fill = rescale),colour = "white") + scale_f
ill_gradient(low = "white",high = "steelblue") +  facet_grid(~ year(date), space="free_x", scale
s="free_x", switch="x") + theme(legend.position = "right",axis.text.x = element_text(angle = 90,
 hjust = 1)) +
  labs(x = "Month", y="Variable",title = "Air Quality Heat map") + theme(legend.title=element_bl
ank())
```




```
#Step   5:  Look    at  all the data    via a   scatter chart

#scatter    chart (using    ggplot  geom_point),    with    the x-axis  representing    the win
d,   the
#y-axis representing    the temperature,    the size    of  each    dot representing    the ozon
e   and the
#color  representing    the solar.R

ggplot(my_aq) +
  geom_point(aes(x=Wind,y=Temp,size=Ozone,color=Solar.R)) + theme(legend.position = "right",axi
s.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Wind", y="Temp",title = "Air Quality Scatter Plot")
```




```
# After examining all of the visualization, it appears that different chart explores understandi
ng of data in multiple ways.

# 1. Boxlpot of Ozone and Wind over month gives us information on sample size and the distributi
on of the data from mean and the outliers
# 2. Line graph using loess method tells how the pattern of these variable changes over time a.
 Wind speed drops in June, July,August and rasising in September b. Ozone is in its peak in the
 month of August c. Solar.R shows variations throughout the month but on higher side in the mont
h of July d. Temperature raising from May untill August Mid and dropping thereafter
# 3. Line graph on all variables on single chart shows that when the temparature raises up the w
ind speed is getting reduced and the Ozone raising up over the same period of time
# Heat map shows that Wind and Solar.R are more concentrated in the month of Map and the pattern
s are darker
# Scattered plot is giving better visualization on the relationship of these variables where the
 light coloured bigger dot size on the top left whereas dark coloured smaller dot size are on th
e bottom right . This shows that When temparature is high the wind speed is less but the Solar.R
 and Ozone is on higher side. When the temparature goes down , Solar.R and Ozone is at the small
er but the wind is on the higher side. We got the similar inference from the line graph in 1 cha
rt overlaying all the variables using loess method.
```