# Weather Prediction

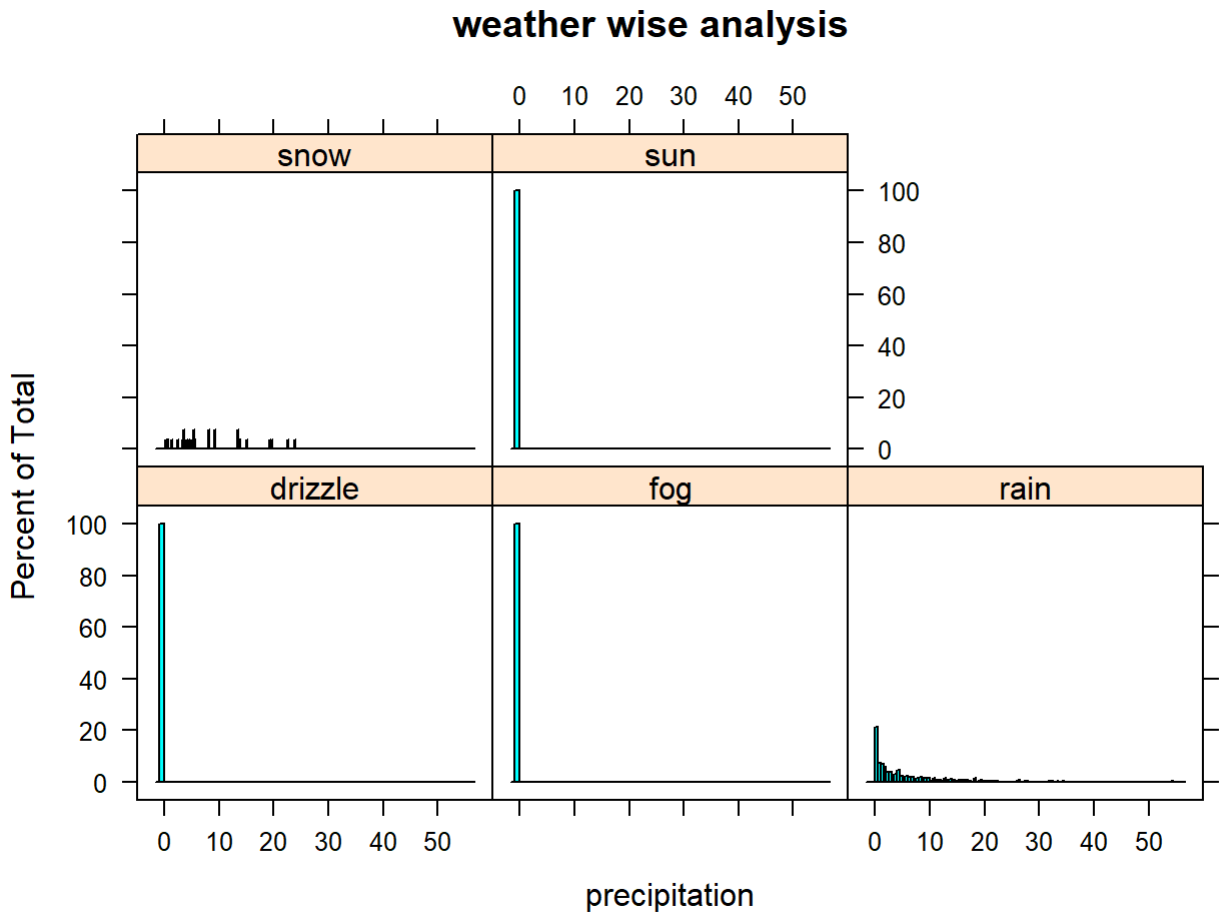
Shalini B

2023-02-19

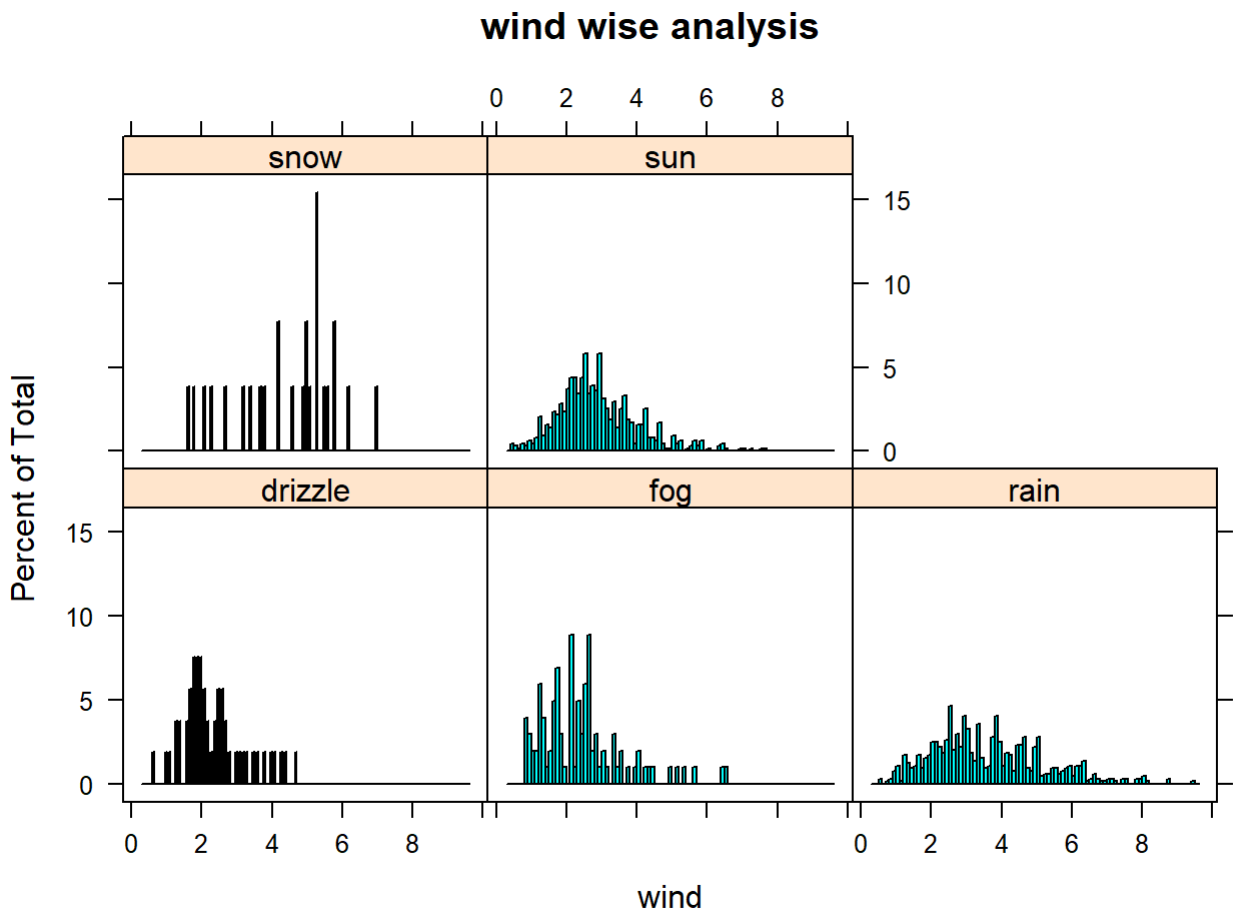
```
library(MASS)
x=file.choose()
weather=read.csv(x)
str(weather)
## 'data.frame':    1461 obs. of  6 variables:
##  $ date         : chr  "2012-01-01" "2012-01-02" "2012-01-03" "2012-01-04" ...
##  $ precipitation: num  0 10.9 0.8 20.3 1.3 2.5 0 0 4.3 1 ...
##  $ temp_max     : num  12.8 10.6 11.7 12.2 8.9 4.4 7.2 10 9.4 6.1 ...
##  $ temp_min     : num  5 2.8 7.2 5.6 2.8 2.2 2.8 2.8 5 0.6 ...
##  $ wind         : num  4.7 4.5 2.3 4.7 6.1 2.2 2.3 2 3.4 3.4 ...
##  $ weather      : chr  "drizzle" "rain" "rain" "rain" ...
#summary
summary(weather)
##      date           precipitation       temp_max        temp_min           wind
##  Length:1461        Min.   : 0.000   Min.   :-1.60   Min.   :-7.100   Min.   :0.
400
##  Class :character   1st Qu.: 0.000   1st Qu.:10.60   1st Qu.: 4.400   1st Qu.:2.
200
##  Mode  :character   Median : 0.000   Median :15.60   Median : 8.300   Median :3.
000
##                     Mean   : 3.029   Mean   :16.44   Mean   : 8.235   Mean   :3.
241
##                     3rd Qu.: 2.800   3rd Qu.:22.20   3rd Qu.:12.200   3rd Qu.:4.
000
##                     Max.   :55.900   Max.   :35.60   Max.   :18.300   Max.   :9.
500
##    weather
##  Length:1461
##  Class :character
##  Mode  :character

#dimension
dim(weather)
## [1] 1461    6
#installing packages
library(lattice)
```

```
#data manipulation
weather$year <- format(as.Date(weather$date, format="%d%m%y"),"%y")
#View(weather)

#subsetting the data by year
#df1(2012)
df1=subset(weather,year=="12")
df1
## [1] date          precipitation temp_max      temp_min      wind          weathe
r
## [7] year
## <0 rows> (or 0-length row.names)
#df2(2013)
df2=subset(weather,year=="13")
df2
## [1] date          precipitation temp_max      temp_min      wind          weathe
r
## [7] year
## <0 rows> (or 0-length row.names)
#df3(2014)
df3=subset(weather,year=="14")
df3
## [1] date          precipitation temp_max      temp_min      wind          weathe
r
## [7] year
## <0 rows> (or 0-length row.names)
#df4(2015)
df4=subset(weather,year=="15")
df4
## [1] date          precipitation temp_max      temp_min      wind          weathe
r
## [7] year
## <0 rows> (or 0-length row.names)
#subsetting the weather of 2012
sn1=subset(weather,weather=="snow")
head(sn1)
##            date precipitation temp_max temp_min wind weather year
## 14   2012-01-14           4.1      4.4      0.6  5.3    snow <NA>
## 15   2012-01-15           5.3      1.1     -3.3  3.2    snow <NA>
## 16   2012-01-16           2.5      1.7     -2.8  5.0    snow <NA>
```

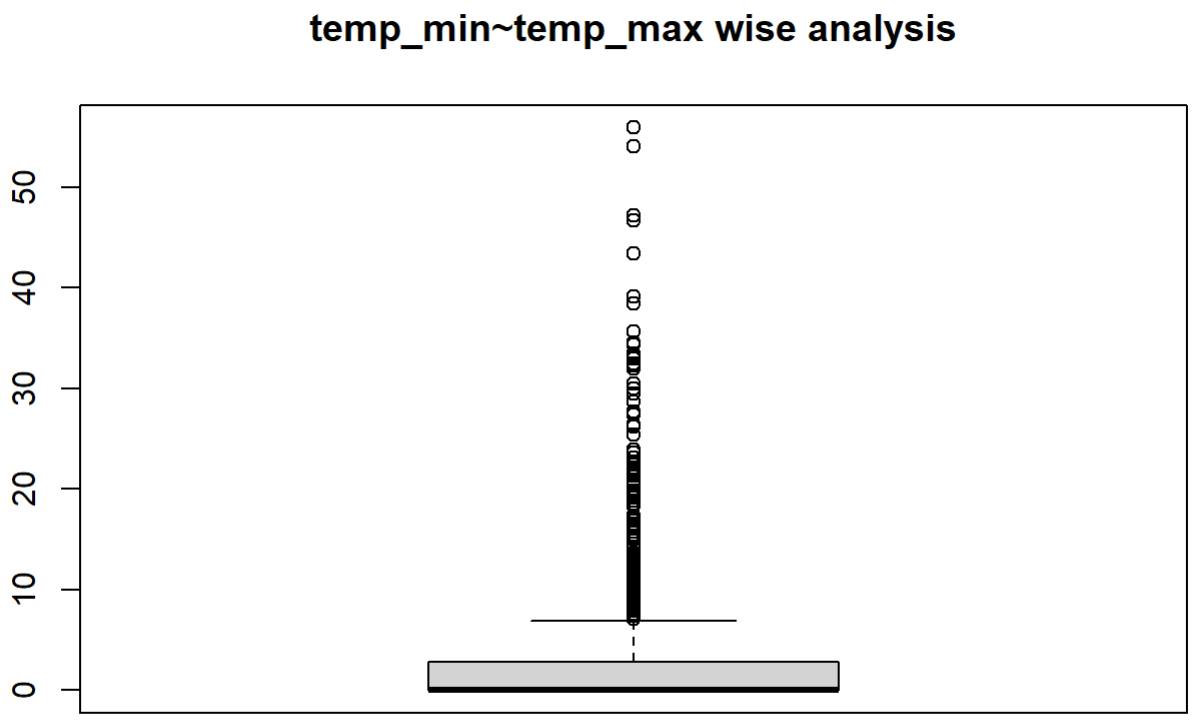
```
## 17   2012-01-17           8.1      3.3      0.0  5.6    snow <NA>
## 18   2012-01-18          19.8      0.0     -2.8  5.0    snow <NA>
sn2=subset(weather,weather=="rain")
head(sn2)
##           date precipitation temp_max temp_min wind weather year
## 2   2012-01-02          10.9     10.6      2.8  4.5    rain <NA>
## 3   2012-01-03           0.8     11.7      7.2  2.3    rain <NA>
## 4   2012-01-04          20.3     12.2      5.6  4.7    rain <NA>
## 5   2012-01-05           1.3      8.9      2.8  6.1    rain <NA>
## 6   2012-01-06           2.5      4.4      2.2  2.2    rain <NA
sn3=subset(weather,weather=="drizzle")
head(sn3)
##            date precipitation temp_max temp_min wind weather year
## 1    2012-01-01             0     12.8      5.0  4.7 drizzle <NA>
## 27   2012-01-27             0      6.7     -2.2  1.4 drizzle <NA>
## 46   2012-02-15             0      7.2      0.6  1.8 drizzle <NA>
## 86   2012-03-26             0     12.8      6.1  4.3 drizzle <NA>
## 104  2012-04-13             0     15.0      3.9  4.0 drizzle <NA>
sn4=subset(weather,weather=="sun")
head(sn4)
##           date precipitation temp_max temp_min wind weather year
## 8   2012-01-08             0     10.0      2.8  2.0     sun <NA>
## 11  2012-01-11             0      6.1     -1.1  5.1     sun <NA>
## 12  2012-01-12             0      6.1     -1.7  1.9     sun <NA>
## 13  2012-01-13             0      5.0     -2.8  1.3     sun <NA>
## 33  2012-02-02             0      8.3      1.7  2.6     sun <NA>
##  [ reached 'max' / getOption("max.print") -- omitted 498 rows ]
#data manipulation
#histogram between weather and precipitation
histogram(~precipitation|weather,data=weather,breaks=100,Col=c('skyblue','pink'),ma
in="weather wise analysis")
```



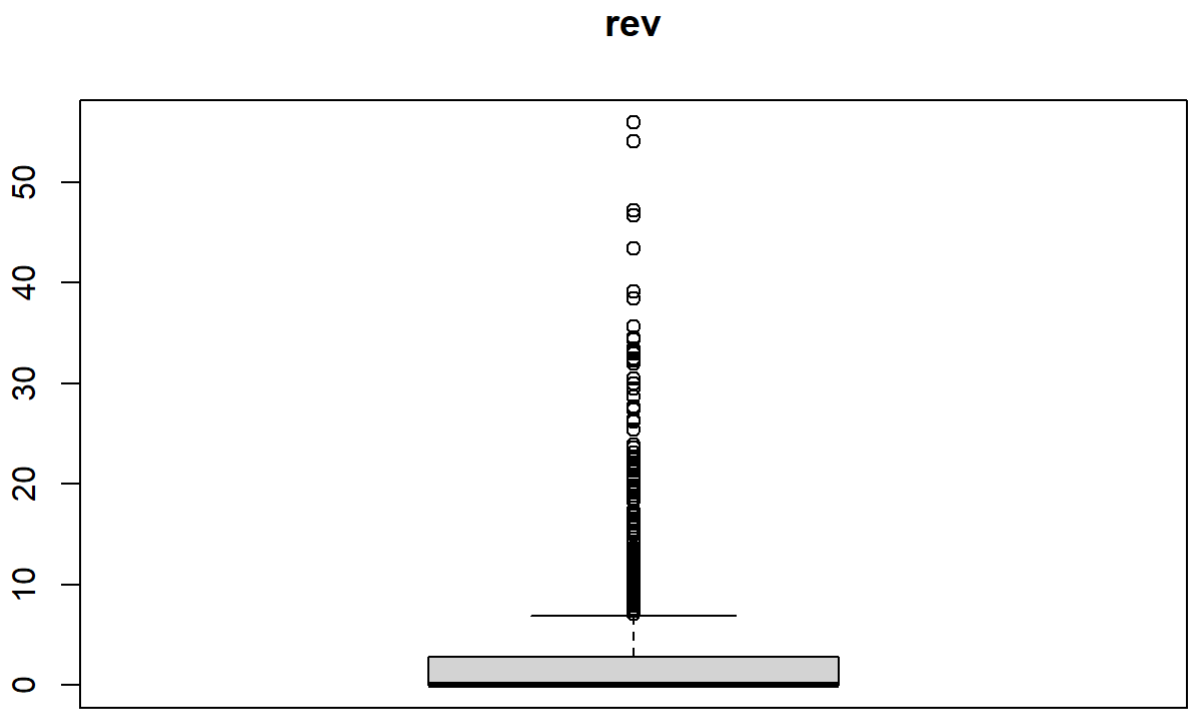

```
#histogram between weather and wind

histogram(~wind|weather,data=weather,breaks=80,Col=c('yellow','purple'),main="wind
wise analysis")
```




```
#boxplot

boxplot(precipitation,data=weather,breaks=100,Col=c('green','orange'),main="temp_mi
n~temp_max wise analysis")
```

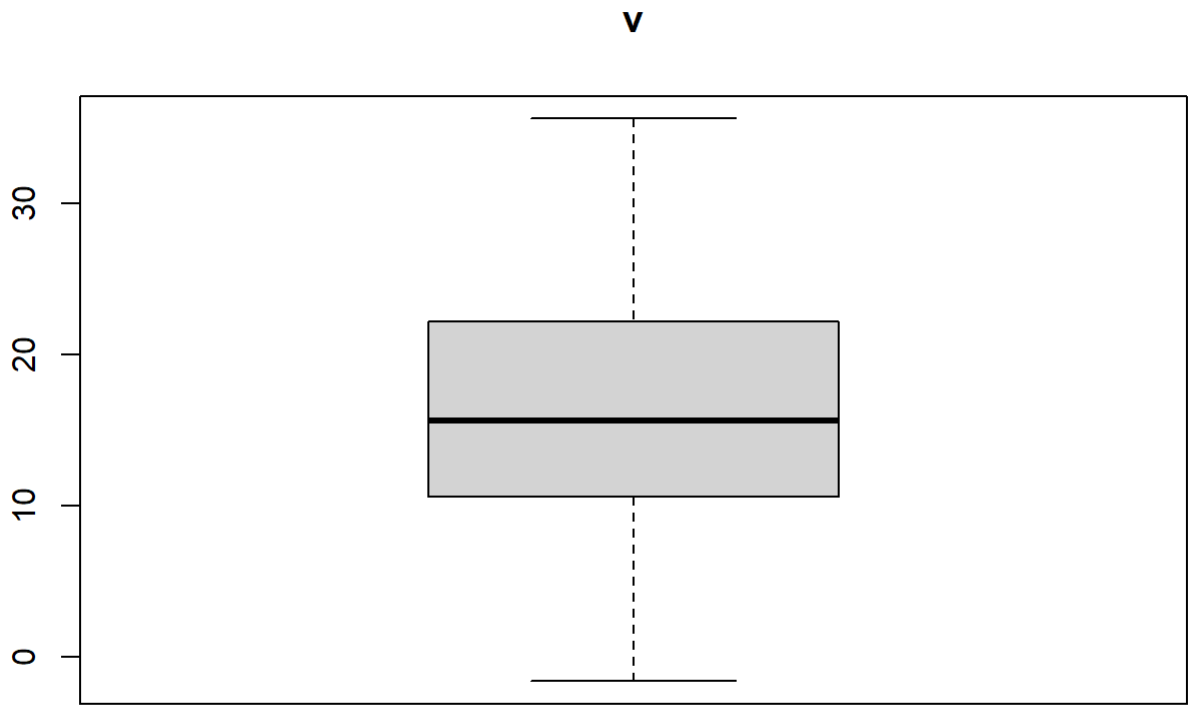



```
#outliers
#Q1 <- quantile(weather$precipitation, .25)
#Q3 <- quantile(weather$precipitation, .75)
#IQR <- IQR(weather$precipitation)
#weather <- subset(weather,weather[[precipitation]]>(Q1 - 1.5*IQR)&(weather[[precip
itation]] < (Q3 + 1.5*IQR)))


boxplot(precipitation,data=weather,breaks=100,Col=c('green','orange'),main="rev")
```

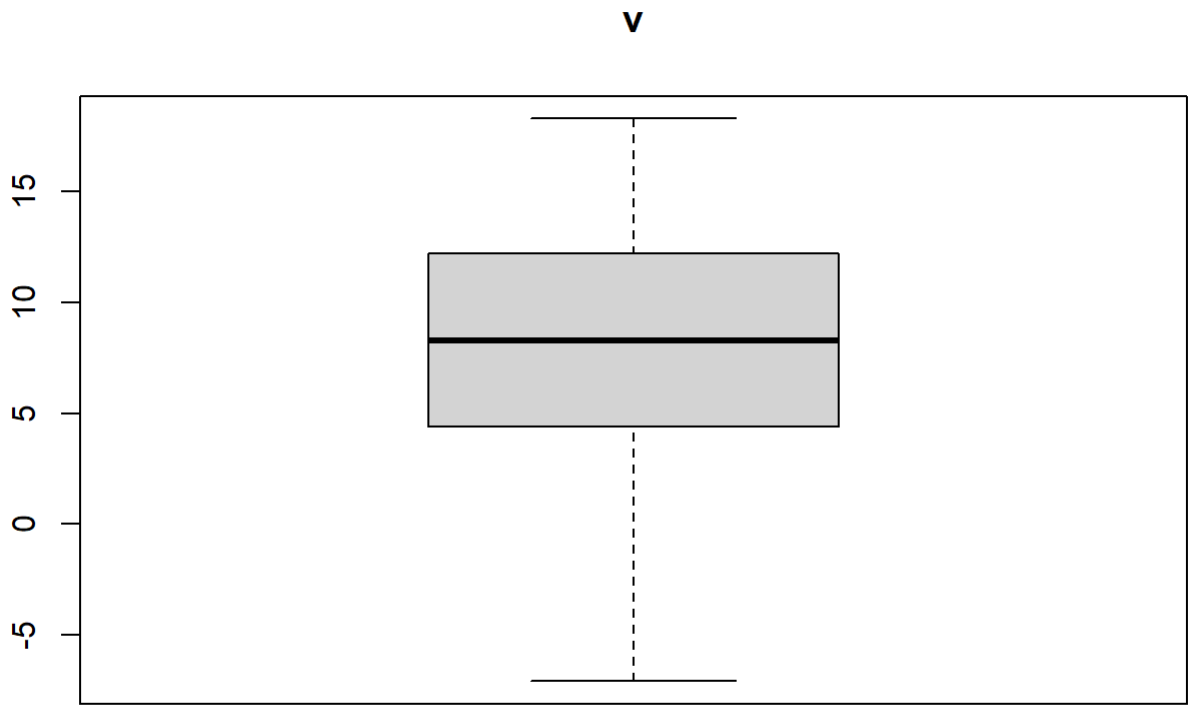



```
#sapply

sapply(weather,class)

##          date precipitation      temp_max      temp_min          wind       weat
her

##   "character"     "numeric"     "numeric"     "numeric"     "numeric"   "charact
er"

##          year

##   "character"
```

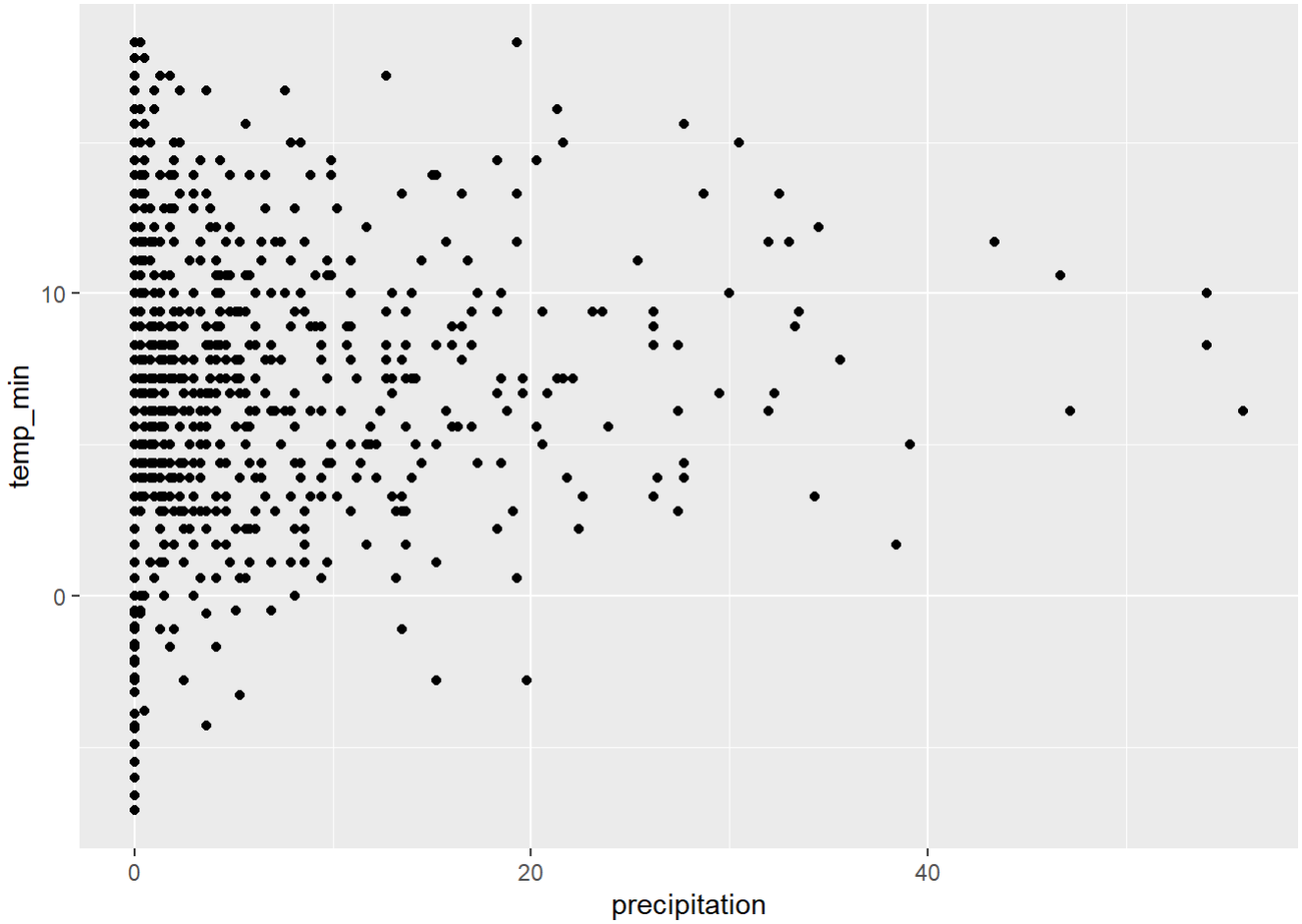
```
boxplot(temp_max,data=weather,breaks=100,Col=c('green','orange'),main="v")
```




```
boxplot(temp_min,data=weather,breaks=100,Col=c('yellow','pink'),main="v")
```




```
#scatter plot


library(ggplot2)
ggplot(weather, aes(x = precipitation, y =temp_min)) +
  geom_point()
```


temp_min
10
0
0
20
40
precipitation