

# R통계분석 10주차 과제 보고서

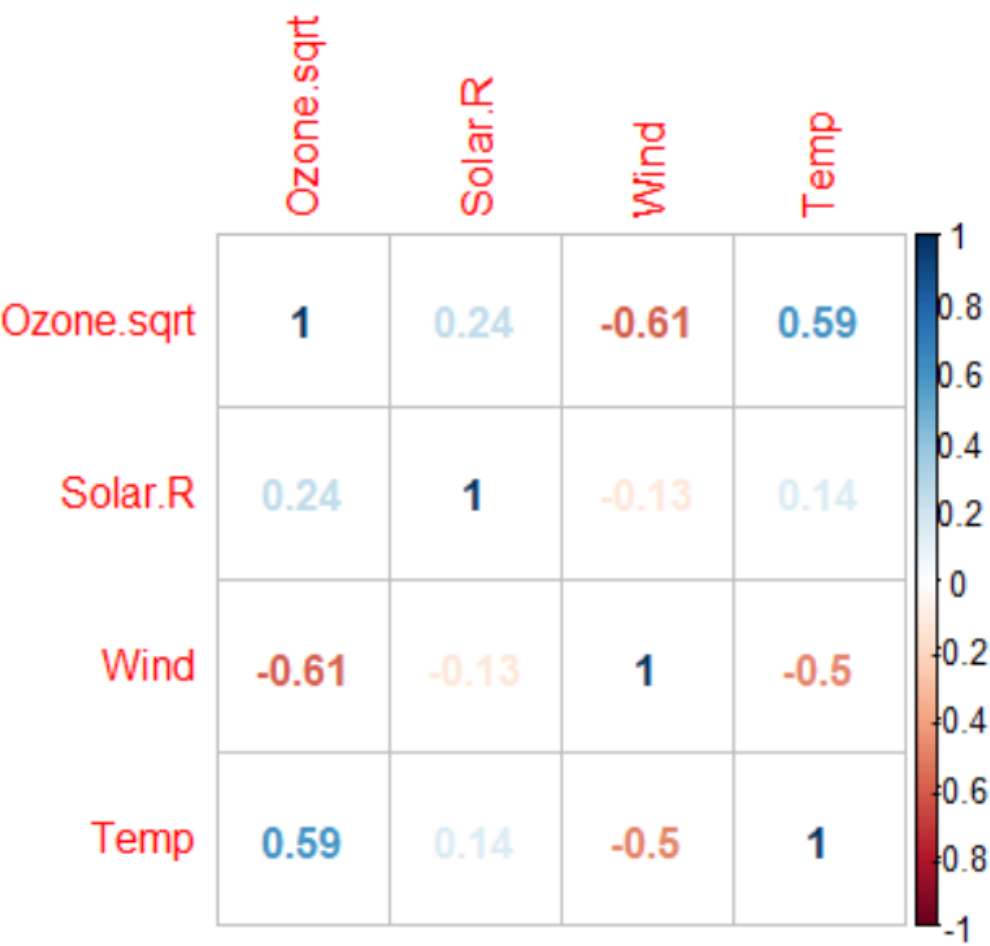
60191697 최 솔

## I. 과제 R 코드

```
1 library(corrgram)
2 library(corrplot)
3 library(MASS)
4
5 air = airquality
6
7 air1 <- air[complete.cases(air),]
8 #결측치 제거
9
10 air1$Day <- as.factor(air1$Day)
11 air1$Month <- as.factor(air1$Month)
12 #factor 설정
13
14 air1$Ozone <- (air1$Ozone)^(1/3)
15 names(air1)[1]<-c("Ozone.sqrt")
16 #Ozone의 세제곱근을 Ozone으로 바꿈
17
18 apply(air1[,1:4],2,shapiro.test)
19 #Solar만 정규 분포 만족 안함
20
21 library(rcompanion)
22 t.solar <- transformTukey(air1$Solar.R)
23 #transformTukey를 이용해
24 #Solar.R이 정규분포를 만족하지 않음을 확인
25
26 x = 1
27 y = 1
28 ans = c()
29 while(x < 5) {
30   while(y < 5) {
31     c.p = cor(air1[,x],air1[,y],method="pearson")
32     c.s = cor(air1[,x],air1[,y],method="spearman")
33     c.k = cor(air1[,x],air1[,y],method="kendall")
34     ans <- append(ans,min(c.p,c.s,c.k))
35     y = y+1
36   }
37   y=1
38   x= x+ 1
39 }
40 ans <- matrix(ans,nrow=4,dimnames = list(c("Ozone.sqrt","Solar.R","Wind",
41                                           "Temp"),c("Ozone.sqrt","Solar.R",
42                                           "Wind","Temp")))
43 ans
44
45 #세 가지 방법으로 상관 계수를 구한 뒤, 가장 작은 상관 계수를 택해,
46 #이를 matrix로 만든다.
47
48 air2 = round(ans,2)
49 corrplot(air2,method="number")
50
51 ggplot(data=air1,aes(x=Ozone.sqrt,y=Solar.R,colour=Month))+geom_point()
52
53 ggplot(data=air1,aes(x=Ozone.sqrt,y=Wind,colour=Month))+geom_point()
54
55 ggplot(data=air1,aes(x=Ozone.sqrt,y=Temp,colour=Month))+geom_point()
56
57 ggplot(data=air1,aes(x=Solar.R,y=Wind,colour=Month))+geom_point()
58
59 ggplot(data=air1,aes(x=Solar.R,y=Temp,colour=Month))+geom_point()
60
61 ggplot(data=air1,aes(x=Temp,y=Wind,colour=Month))+geom_point()
62
```

II. airQuality 변수들 간의 관계 분석

1. 상관 계수



<그림 1>

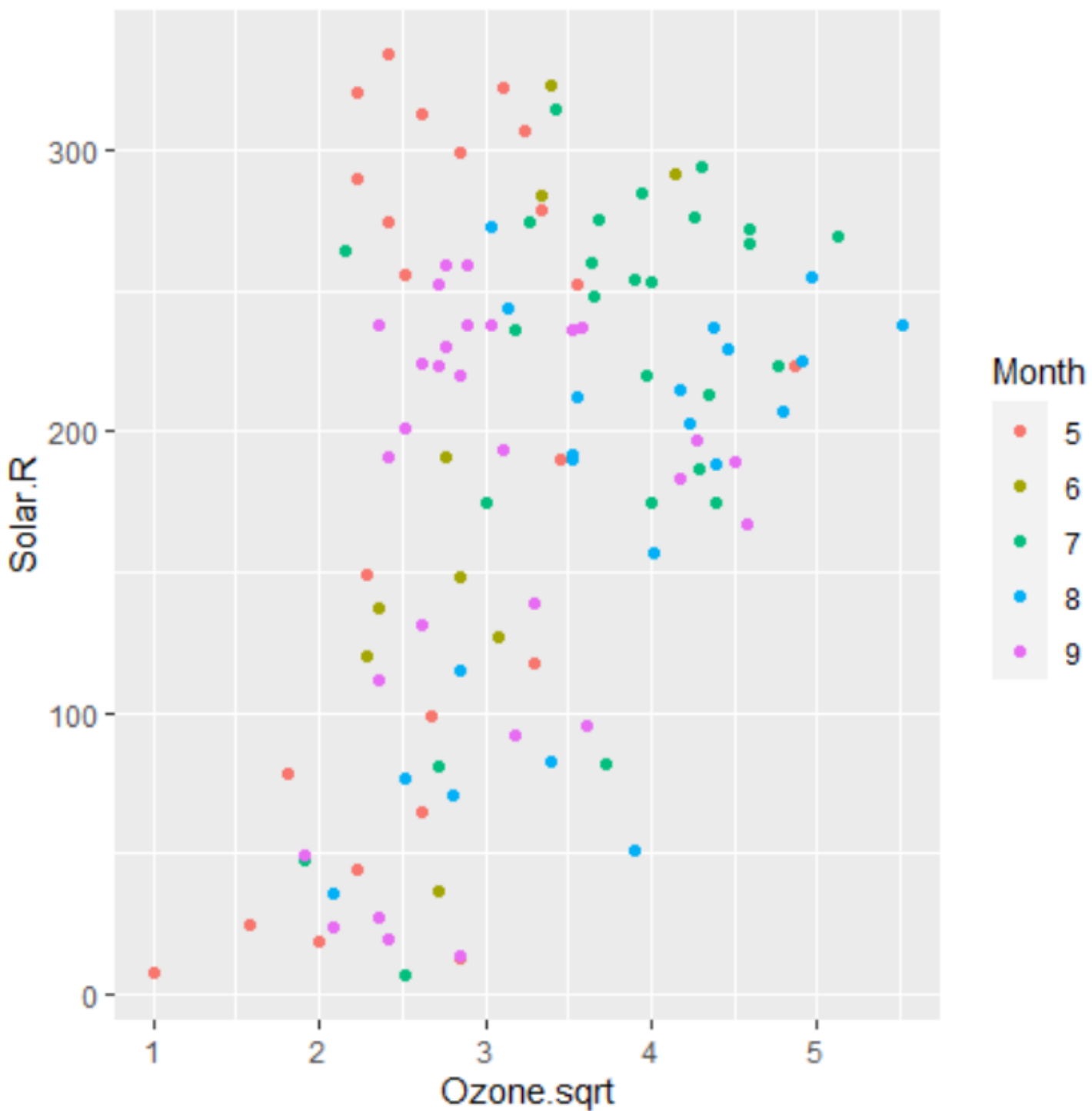
corrplot을 이용해 그린 그림1을 통해, 각 변수들 간의 상관 계수를 알 수 있다.

상관 계수  $|p|$  가  $|1.0|$ 과  $|0.7|$  사이이면 강한 상관 관계,  $|0.7|$ 과  $|0.3|$  사이이면 보통의 상관 관계,  $|0.3|$ 과  $|0.1|$  사이이면 약한 상관 관계이다. 또한, 상관계수가 음수, 양수인지에 따라 양적, 음적 관계를 알 수 있다.

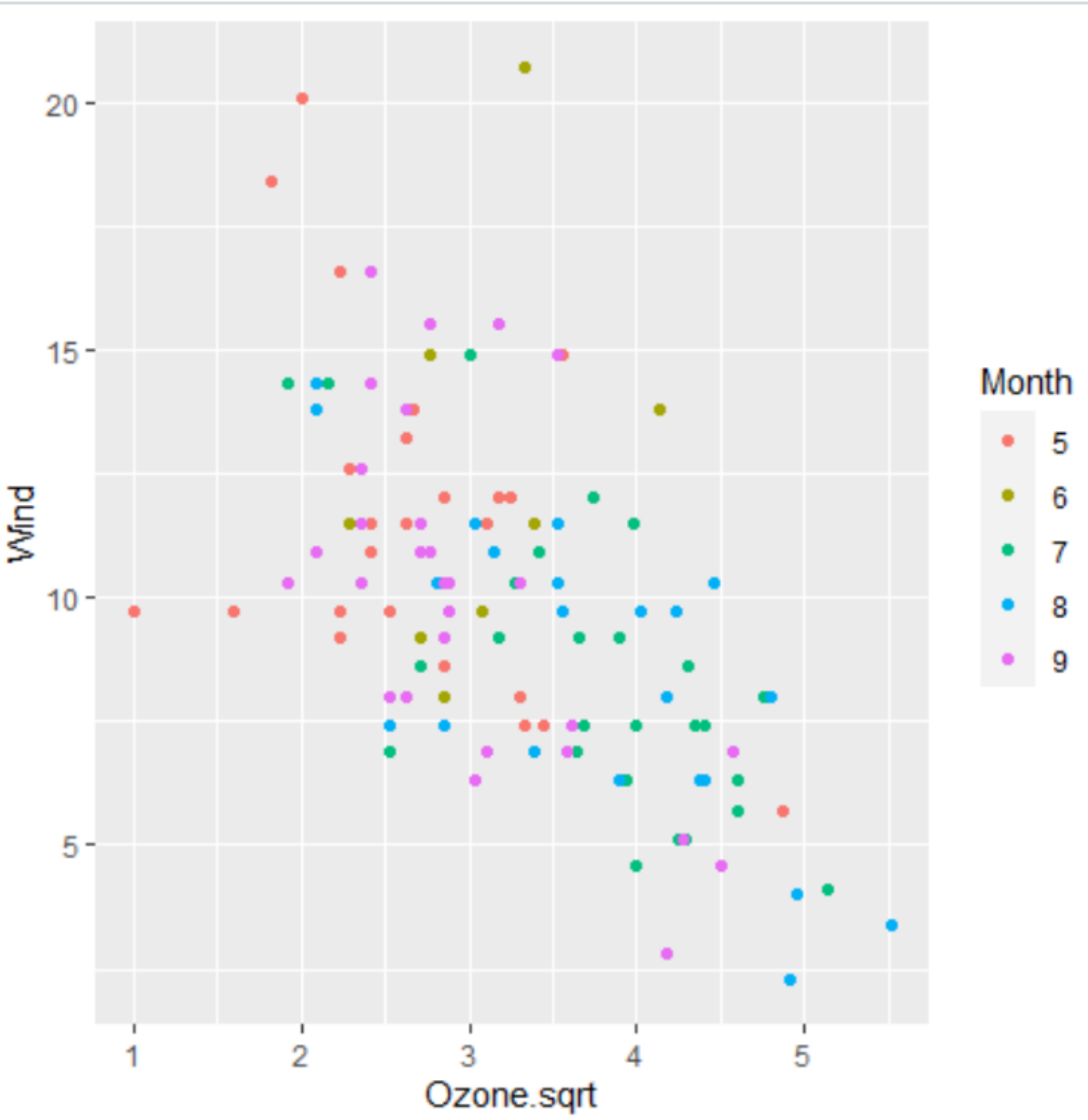
## 2. 분석 결과

그림1에서 구한 상관 계수와 앞서 언급한 방법을 이용해 구한 변수 간 관계와 이를 산점도 그래프로 표현한 것은 다음과 같다.

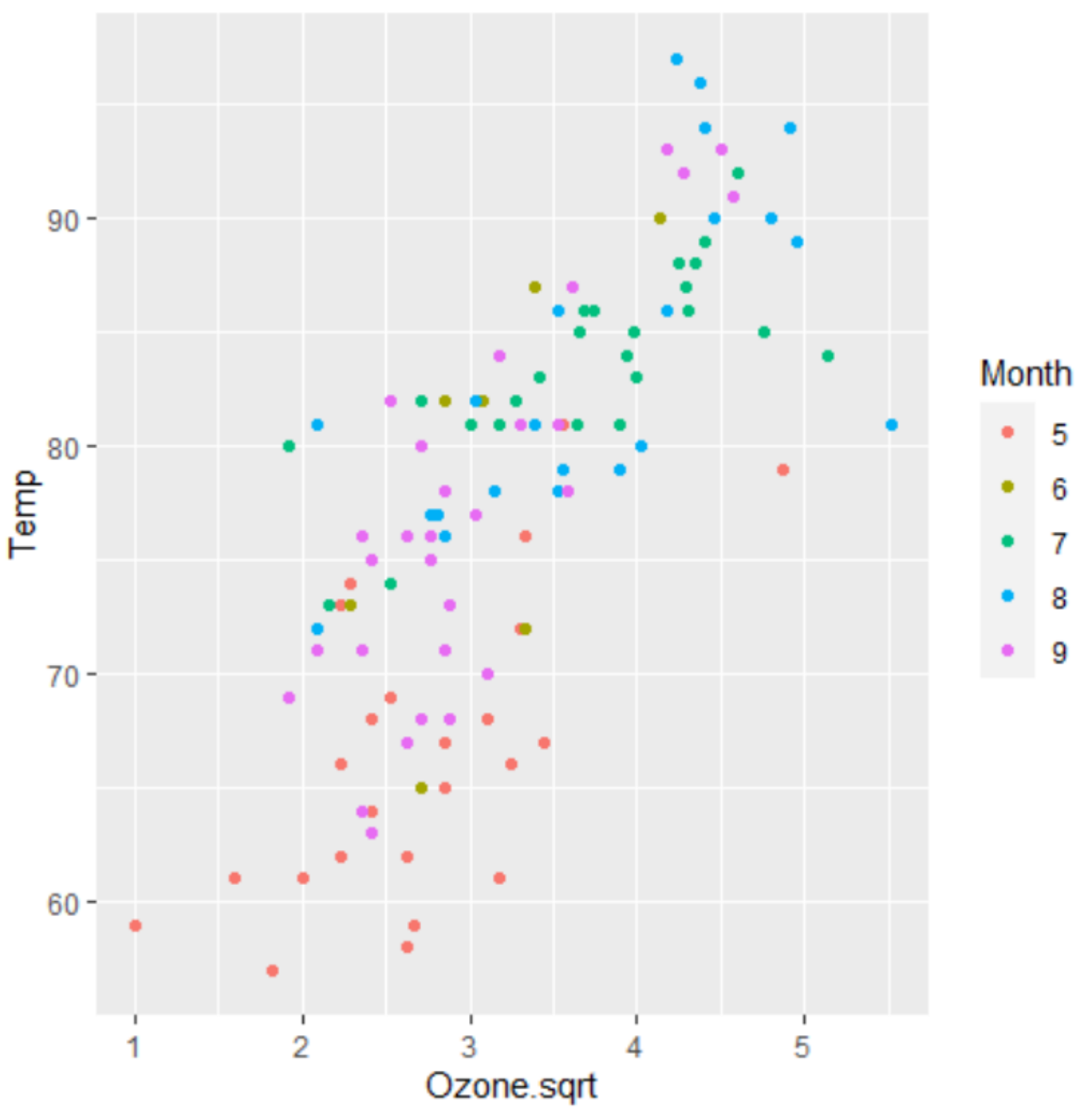
a) Ozone.sqrt와 Solar.R의 관계 : 약한 양적 상관 관계



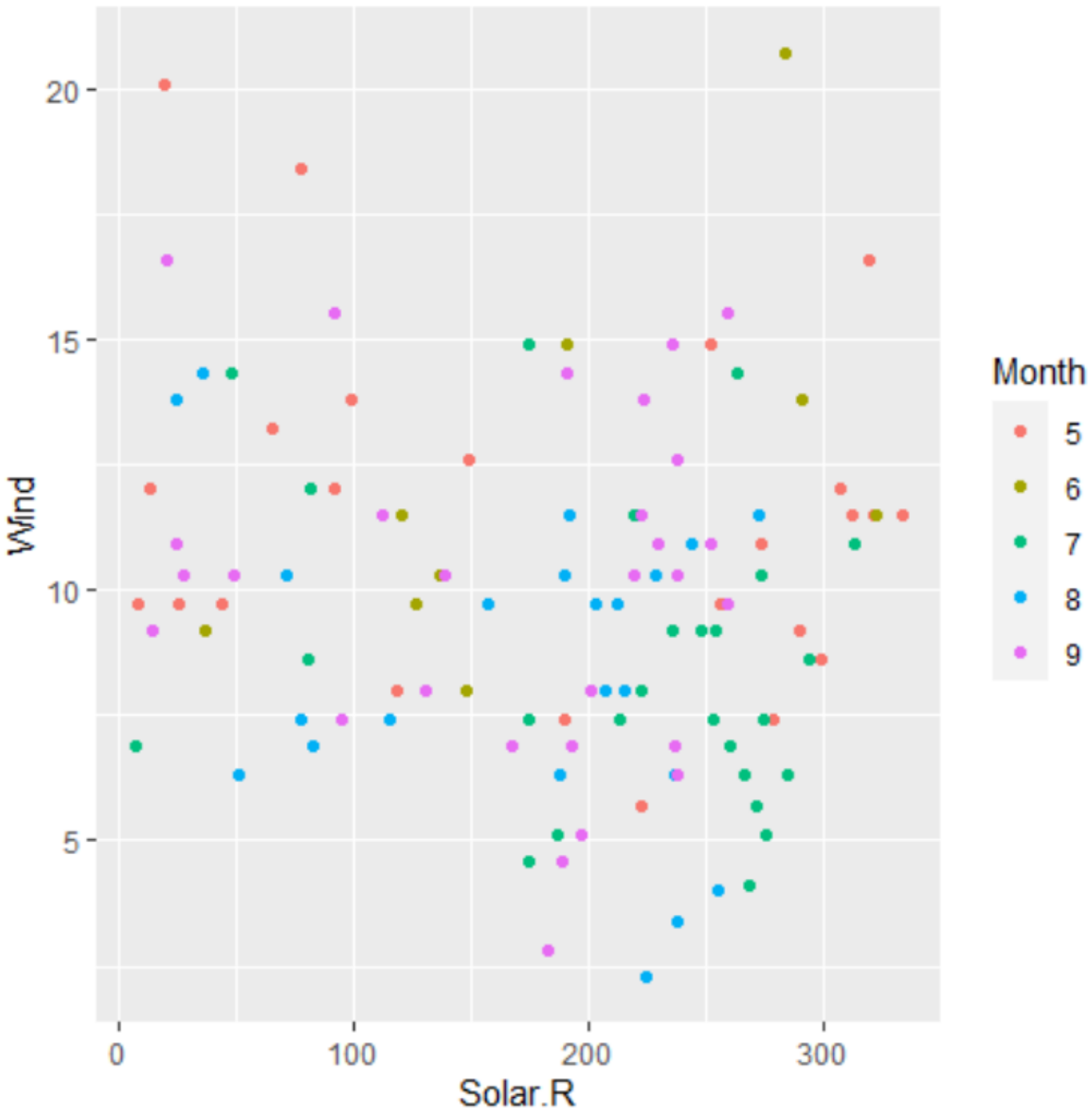
b) Ozone.sqrt와 Wind 의 관계 : 보통 음적 상관 관계



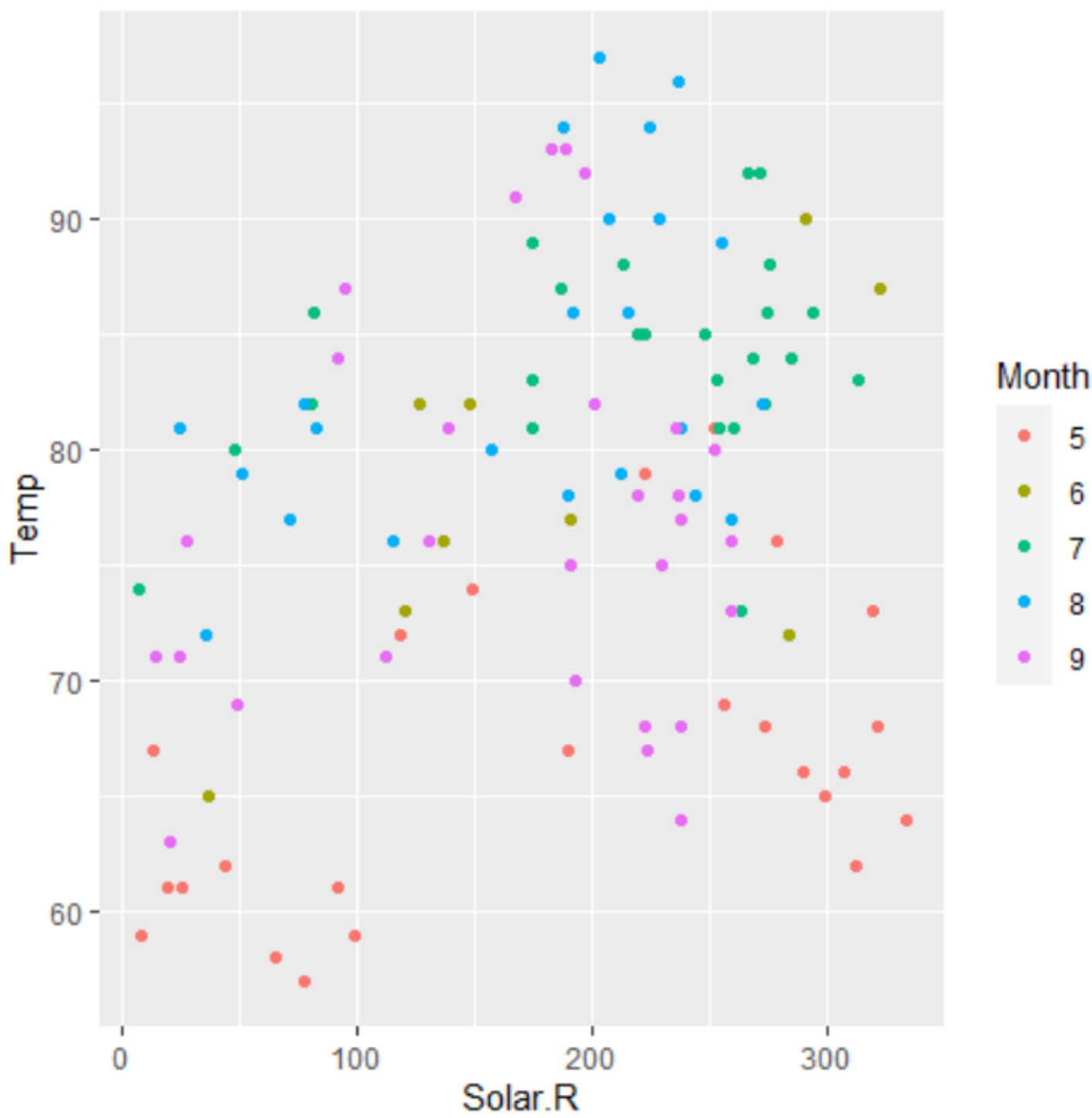
c) Ozone.sqrt와 Temp의 관계 : 보통 양적 상관 관계



d) Solar.R과 Wind의 관계 : 약한 음적 상관 관계



e) Solar.R과 Temp의 관계 : 약한 양적 상관 관계



f) Wind와 Temp의 관계 : 보통의 음적 상관 관계

