

airquality correleation analysis

```
#install.packages("PerformanceAnalytics")
#install.packages("corrplot")
library(PerformanceAnalytics)
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
##
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
##
##      legend
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(knitr)
```

airquality data : 1973년 뉴욕의 공기의 질을 측정한 데이터로 Ozon, Solar.R, Wind, Temp 4개의 변수로 구성되어 있다.

```
head(airquality)
```

```
##      Ozone Solar.R Wind Temp Month Day
## 1      41      190  7.4   67     5    1
## 2      36      118  8.0   72     5    2
## 3      12      149 12.6   74     5    3
## 4      18      313 11.5   62     5    4
## 5      NA       NA 14.3   56     5    5
## 6      28       NA 14.9   66     5    6
```

month와 day를 제외한 변수들을 aq1에 대입

```
aq1 <- airquality[,c(1:4)]
kable(aq1)
```

Ozone	Solar.R	Wind	Temp
41	190	7.4	67
36	118	8.0	72
12	149	12.6	74
18	313	11.5	62
NA	NA	14.3	56
28	NA	14.9	66
23	299	8.6	65
19	99	13.8	59
8	19	20.1	61
NA	194	8.6	69
7	NA	6.9	74
16	256	9.7	69
11	290	9.2	66
14	274	10.9	68
18	65	13.2	58
14	334	11.5	64
34	307	12.0	66
6	78	18.4	57
30	322	11.5	68
11	44	9.7	62
1	8	9.7	59
11	320	16.6	73
4	25	9.7	61
32	92	12.0	61
NA	66	16.6	57
NA	266	14.9	58
NA	NA	8.0	57
23	13	12.0	67
45	252	14.9	81
115	223	5.7	79
37	279	7.4	76
NA	286	8.6	78

Ozone	Solar.R	Wind	Temp
NA	287	9.7	74
NA	242	16.1	67
NA	186	9.2	84
NA	220	8.6	85
NA	264	14.3	79
29	127	9.7	82
NA	273	6.9	87
71	291	13.8	90
39	323	11.5	87
NA	259	10.9	93
NA	250	9.2	92
23	148	8.0	82
NA	332	13.8	80
NA	322	11.5	79
21	191	14.9	77
37	284	20.7	72
20	37	9.2	65
12	120	11.5	73
13	137	10.3	76
NA	150	6.3	77
NA	59	1.7	76
NA	91	4.6	76
NA	250	6.3	76
NA	135	8.0	75
NA	127	8.0	78
NA	47	10.3	73
NA	98	11.5	80
NA	31	14.9	77
NA	138	8.0	83
135	269	4.1	84
49	248	9.2	85
32	236	9.2	81

Ozone	Solar.R	Wind	Temp
NA	101	10.9	84
64	175	4.6	83
40	314	10.9	83
77	276	5.1	88
97	267	6.3	92
97	272	5.7	92
85	175	7.4	89
NA	139	8.6	82
10	264	14.3	73
27	175	14.9	81
NA	291	14.9	91
7	48	14.3	80
48	260	6.9	81
35	274	10.3	82
61	285	6.3	84
79	187	5.1	87
63	220	11.5	85
16	7	6.9	74
NA	258	9.7	81
NA	295	11.5	82
80	294	8.6	86
108	223	8.0	85
20	81	8.6	82
52	82	12.0	86
82	213	7.4	88
50	275	7.4	86
64	253	7.4	83
59	254	9.2	81
39	83	6.9	81
9	24	13.8	81
16	77	7.4	82
78	NA	6.9	86

Ozone	Solar.R	Wind	Temp
35	NA	7.4	85
66	NA	4.6	87
122	255	4.0	89
89	229	10.3	90
110	207	8.0	90
NA	222	8.6	92
NA	137	11.5	86
44	192	11.5	86
28	273	11.5	82
65	157	9.7	80
NA	64	11.5	79
22	71	10.3	77
59	51	6.3	79
23	115	7.4	76
31	244	10.9	78
44	190	10.3	78
21	259	15.5	77
9	36	14.3	72
NA	255	12.6	75
45	212	9.7	79
168	238	3.4	81
73	215	8.0	86
NA	153	5.7	88
76	203	9.7	97
118	225	2.3	94
84	237	6.3	96
85	188	6.3	94
96	167	6.9	91
78	197	5.1	92
73	183	2.8	93
91	189	4.6	93
47	95	7.4	87

Ozone	Solar.R	Wind	Temp
32	92	15.5	84
20	252	10.9	80
23	220	10.3	78
21	230	10.9	75
24	259	9.7	73
44	236	14.9	81
21	259	15.5	76
28	238	6.3	77
9	24	10.9	71
13	112	11.5	71
46	237	6.9	78
18	224	13.8	67
13	27	10.3	76
24	238	10.3	68
16	201	8.0	82
13	238	12.6	64
23	14	9.2	71
36	139	10.3	81
7	49	10.3	69
14	20	16.6	63
30	193	6.9	70
NA	145	13.2	77
14	191	14.3	75
18	131	8.0	76
20	223	11.5	68

na값들로 인해 상관분석의 결과가 유의미하지 않다.

```
cor(aq1)
```

```
##           Ozone Solar.R      Wind      Temp
## Ozone         1      NA         NA         NA
## Solar.R       NA         1         NA         NA
## Wind         NA         NA  1.0000000 -0.4579879
## Temp         NA         NA -0.4579879  1.0000000
```

aq1의 결측치들을 제거한 뒤 aq2에 대입

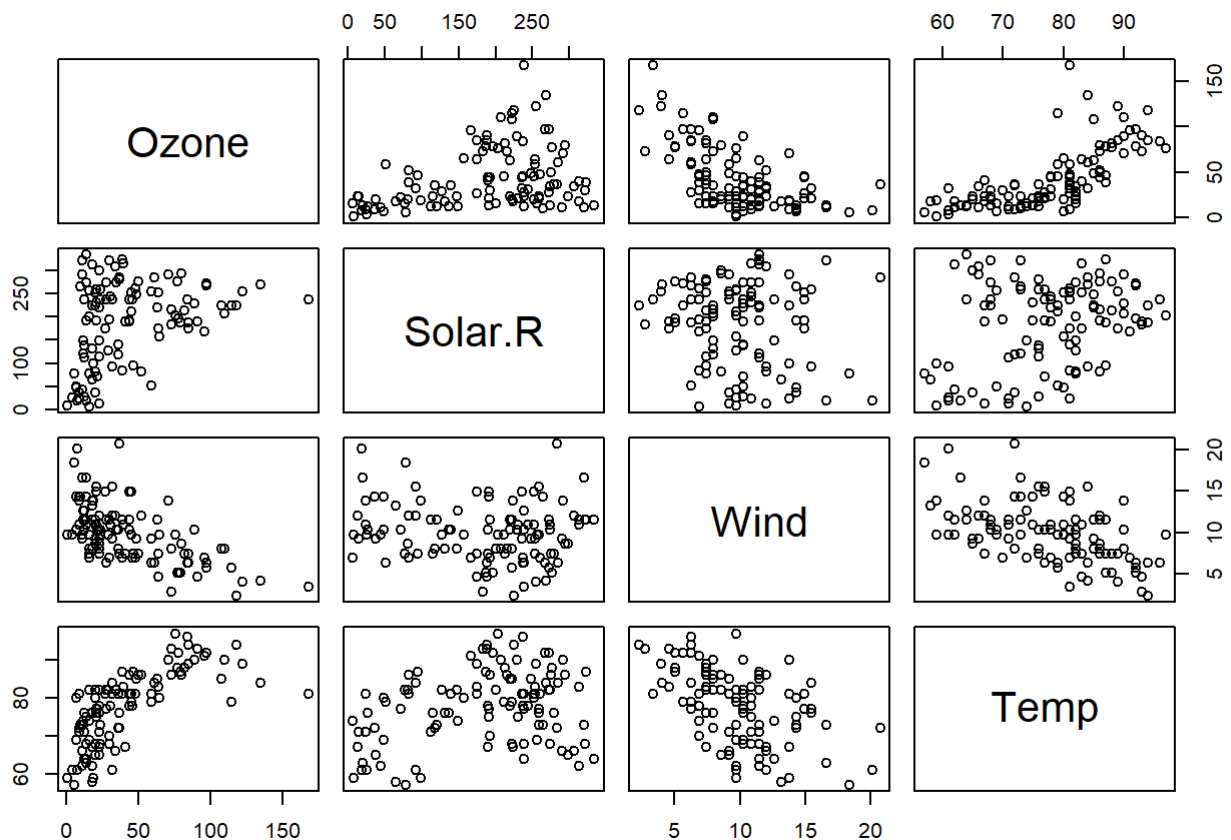
```
aq2 <- na.omit(aq1)
aq.cor <- cor(aq2)
aq.cor
```

```
##           Ozone   Solar.R     Wind     Temp
## Ozone      1.000000  0.3483417 -0.6124966  0.6985414
## Solar.R    0.3483417  1.0000000 -0.1271835  0.2940876
## Wind      -0.6124966 -0.1271835  1.0000000 -0.4971897
## Temp       0.6985414  0.2940876 -0.4971897  1.0000000
```

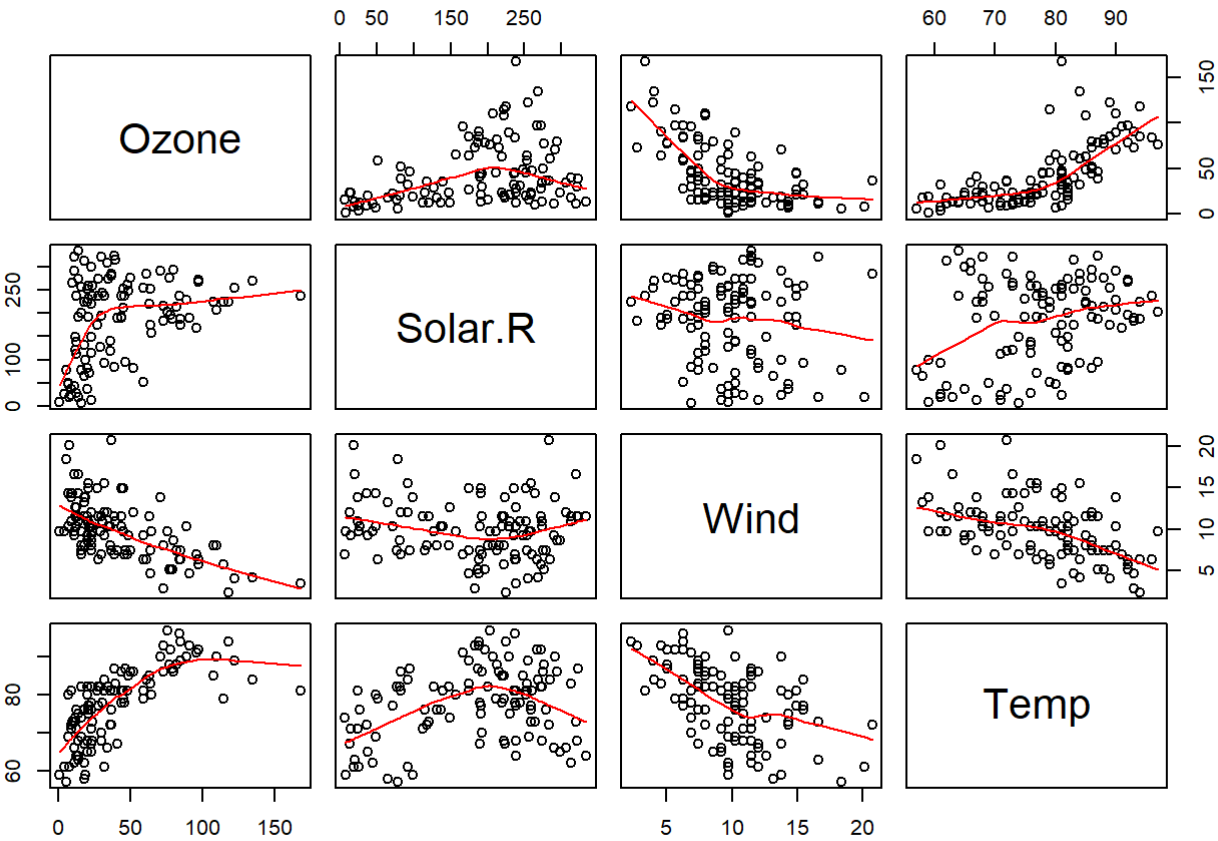
coefficient of determination : 결정계수

-1.0과 -0.7 사이이면, 강한 음적 선형관계
 -0.7과 -0.3 사이이면, 뚜렷한 음적 선형관계
 -0.3과 -0.1 사이이면, 약한 음적 선형관계
 -0.1과 +0.1 사이이면, 거의 무시될 수 있는 선형관계
 +0.1과 +0.3 사이이면, 약한 양적 선형관계
 +0.3과 +0.7 사이이면, 뚜렷한 양적 선형관계
 +0.7과 +1.0 사이이면, 강한 양적 선형관계

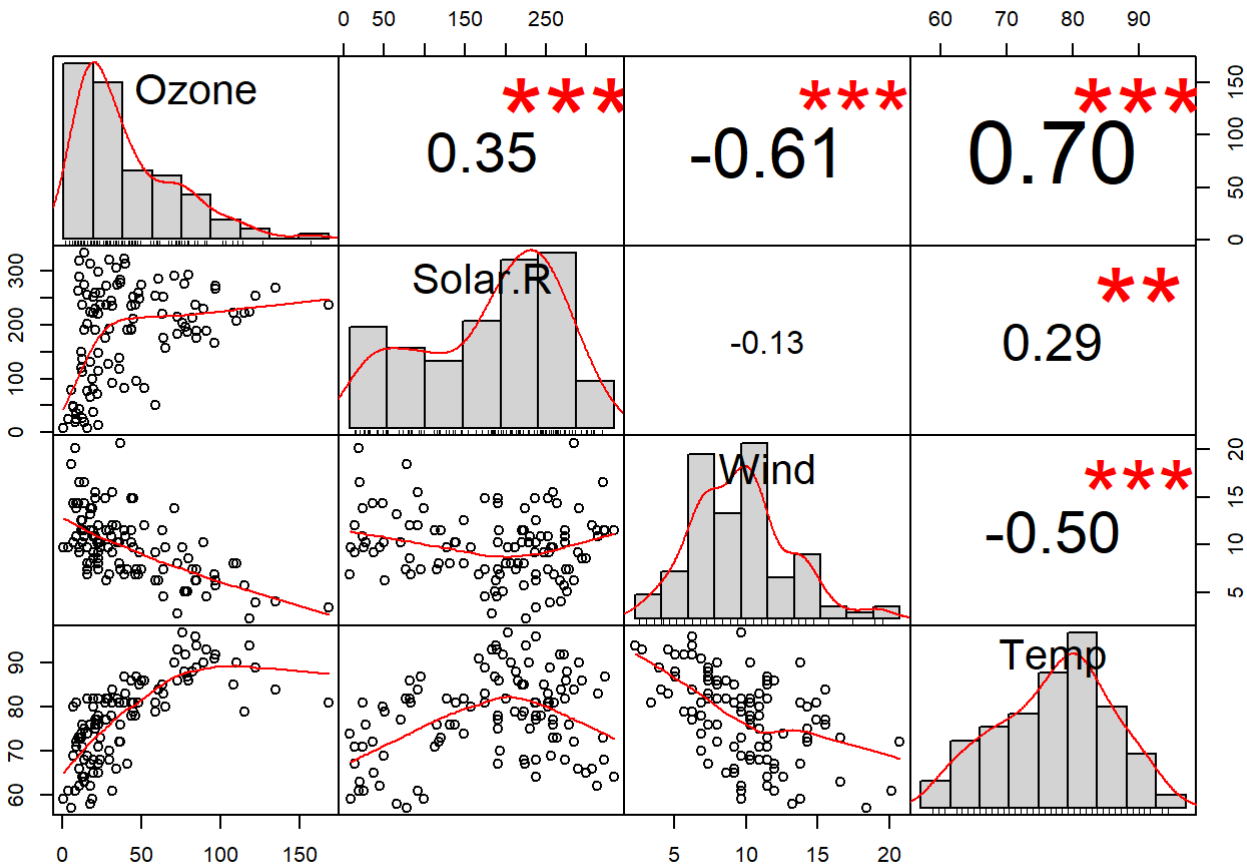
```
plot(aq2)
```



```
pairs(aq2, panel = panel.smooth)
```



```
chart.Correlation(aq2, histogram = TRUE, pch = 19)
```




```
corrplot(aq.cor, method = "number")
```

