Application

Refer to the Air Quality data described previously, and the analyses we have done with Ozone as the response variable, and the five explanatory variables (including the two engineered features).

1. Find and **report the median value for wind speed and temperature**

```
library(dplyr)
library(MASS)   # For ridge regression
library(glmnet) # For LASSO
source("Helper Functions.R")
data = na.omit(airquality[, 1:4])
data$TWcp = data$Temp*data$Wind
data$TWrat = data$Temp/data$Wind

#1. Find and report the median value for wind speed and temperature
wind.med = median(data$Wind)
temp.med = median(data$Temp)
```

```
> temp.med
[1] 79
> wind.med
[1] 9.7
```

2. Use this median value to create high and low regions on both variables. Show values for Temp, Wind, and the two high-low region factors for these variables.

```
#2. Use this median value to create high and low regions on both variables. Show values
#for Temp, Wind, and the two high-low region factors for these variables.
wind.hilo = (data$Wind < median(data$Wind))
temp.hilo = (data$Temp < median(data$Temp))

head(data.frame(data$Wind, wind.hilo, data$Temp, temp.hilo))
tail(data.frame(data$Wind, wind.hilo, data$Temp, temp.hilo))
```

	Wind	Temp	wind.split	temp.split
1	7.4	67	FALSE	FALSE
2	8.0	72	FALSE	FALSE
3	12.6	74	TRUE	FALSE
4	11.5	62	TRUE	FALSE
5	8.6	65	FALSE	FALSE
6	13.8	59	TRUE	FALSE
7	20.1	61	TRUE	FALSE
8	9.7	69	TRUE	FALSE
9	9.2	66	FALSE	FALSE
10	10.9	68	TRUE	FALSE
11	13.2	58	TRUE	FALSE
12	11.5	64	TRUE	FALSE
13	12.0	66	TRUE	FALSE
14	18.4	57	TRUE	FALSE
15	11.5	68	TRUE	FALSE
16	9.7	62	TRUE	FALSE
17	9.7	59	TRUE	FALSE
18	16.6	73	TRUE	FALSE
19	9.7	61	TRUE	FALSE
20	12.0	61	TRUE	FALSE
21	12.0	67	TRUE	FALSE
22	14.9	81	TRUE	TRUE
23	5.7	79	FALSE	TRUE

24	7.4	76	FALSE	FALSE
25	9.7	82	TRUE	TRUE
26	13.8	90	TRUE	TRUE
27	11.5	87	TRUE	TRUE
28	8.0	82	FALSE	TRUE
29	14.9	77	TRUE	FALSE
30	20.7	72	TRUE	FALSE
31	9.2	65	FALSE	FALSE
32	11.5	73	TRUE	FALSE
33	10.3	76	TRUE	FALSE
34	4.1	84	FALSE	TRUE
35	9.2	85	FALSE	TRUE
36	9.2	81	FALSE	TRUE
37	4.6	83	FALSE	TRUE
38	10.9	83	TRUE	TRUE
39	5.1	88	FALSE	TRUE
40	6.3	92	FALSE	TRUE
41	5.7	92	FALSE	TRUE
42	7.4	89	FALSE	TRUE
43	14.3	73	TRUE	FALSE
44	14.9	81	TRUE	TRUE
45	14.3	80	TRUE	TRUE
46	6.9	81	FALSE	TRUE
47	10.3	82	TRUE	TRUE
48	6.3	84	FALSE	TRUE
49	5.1	87	FALSE	TRUE
50	11.5	85	TRUE	TRUE
51	6.9	74	FALSE	FALSE
52	8.6	86	FALSE	TRUE
53	8.0	85	FALSE	TRUE
54	8.6	82	FALSE	TRUE
55	12.0	86	TRUE	TRUE
56	7.4	88	FALSE	TRUE
57	7.4	86	FALSE	TRUE
58	7.4	83	FALSE	TRUE
59	9.2	81	FALSE	TRUE
60	6.9	81	FALSE	TRUE
61	13.8	81	TRUE	TRUE
62	7.4	82	FALSE	TRUE
63	4.0	89	FALSE	TRUE
64	10.3	90	TRUE	TRUE
65	8.0	90	FALSE	TRUE
66	11.5	86	TRUE	TRUE
67	11.5	82	TRUE	TRUE
68	9.7	80	TRUE	TRUE
69	10.3	77	TRUE	FALSE
70	6.3	79	FALSE	TRUE
71	7.4	76	FALSE	FALSE
72	10.9	78	TRUE	FALSE
73	10.3	78	TRUE	FALSE
74	15.5	77	TRUE	FALSE
75	14.3	72	TRUE	FALSE
76	9.7	79	TRUE	TRUE
77	3.4	81	FALSE	TRUE
78	8.0	86	FALSE	TRUE
79	9.7	97	TRUE	TRUE
80	2.3	94	FALSE	TRUE
81	6.3	96	FALSE	TRUE
82	6.3	94	FALSE	TRUE
83	6.9	91	FALSE	TRUE
84	5.1	92	FALSE	TRUE
85	2.8	93	FALSE	TRUE
86	4.6	93	FALSE	TRUE
87	7.4	87	FALSE	TRUE
88	15.5	84	TRUE	TRUE
89	10.9	80	TRUE	TRUE
90	10.3	78	TRUE	FALSE
91	10.9	75	TRUE	FALSE
92	9.7	73	TRUE	FALSE
93	14.9	81	TRUE	TRUE

```
94   15.5   76        TRUE       FALSE
95    6.3   77       FALSE       FALSE
96   10.9   71        TRUE       FALSE
97   11.5   71        TRUE       FALSE
98    6.9   78       FALSE       FALSE
99   13.8   67        TRUE       FALSE
100 10.3    76        TRUE       FALSE
101 10.3    68        TRUE       FALSE
102  8.0    82       FALSE        TRUE
103 12.6    64        TRUE       FALSE
104  9.2    71       FALSE       FALSE
105 10.3    81        TRUE        TRUE
106 10.3    69        TRUE       FALSE
107 16.6    63        TRUE       FALSE
108  6.9    70       FALSE       FALSE
109 14.3    75        TRUE       FALSE
110  8.0    76       FALSE       FALSE
111 11.5    68        TRUE       FALSE
```

2. Fit a linear regression with the two region variables.

(a) **Report the results from summary().**

```
#3. Fit a linear regression with the two region variables.
mod.2step = lm(data$Ozone ~ wind.hilo + temp.hilo)

#(a) Report the results from summary().
summary(mod.2step)
```

```
> summary(mod.2step)

Call:
lm(formula = data$Ozone ~ wind.hilo + temp.hilo)

Residuals:
    Min      1Q  Median      3Q     Max
-55.394 -12.394  -1.063   9.210  96.606

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     48.848      4.338  11.260  < 2e-16 ***
wind.hiloTRUE   22.546      4.836   4.662  9.0e-06 ***
temp.hiloTRUE  -34.332      4.804  -7.146  1.1e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.26 on 108 degrees of freedom
Multiple R-squared:  0.5203,    Adjusted R-squared:  0.5114
F-statistic: 58.57 on 2 and 108 DF,  p-value: < 2.2e-16
```

(b) Do the two variables have statistically significant influence on the mean ozone level at the 5% Type 1 error rate? **Report their p-values and your conclusion.**
(No hypotheses needed.)

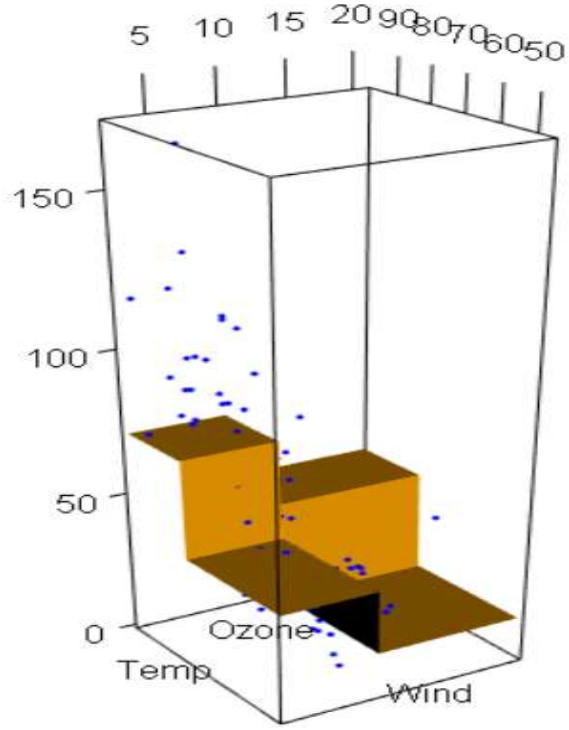
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     48.848      4.338  11.260  < 2e-16 ***
wind.hiloTRUE   22.546      4.836   4.662  9.0e-06 ***
temp.hiloTRUE  -34.332      4.804  -7.146  1.1e-10 ***
```

-> Both p-values < 0.05, so both are statistically significant

(c) Make a 3-D plot of the surface. **Report a screenshot from some angle that shows the whole surface and describe how the surface changes with**

1. **each variable (use one short sentence each).**

```
#(c) Make a 3-D plot of the surface. Report a screenshot from some angle that
#shows the whole surface and describe how the surface changes with
#each variable (use one short sentence each).
with(data, plot3d(Ozone ~ Wind + Temp))

#Wind range: 2.3~ 20.7
#Temp range: 47~ 97


open3d()
plot3d(data$Ozone ~ data$Wind + data$Temp, col="blue")

x1 <- seq(from=2.2, to=21, by=.05)
x2 = seq(from=46, to=98, by=.5)
xy1 <- data.frame(expand.grid(Wind=x1, Temp=x2))

xy1c = data.frame(wind.hilo = (xy1$Wind < median(data$Wind)),
                  temp.hilo = (xy1$Temp < median(data$Temp)))

pred2 <- predict(mod.2step ,newdata=xy1c)
surface2 = matrix(pred2, nrow=length(x1))


open3d()
persp3d(x = x1, y = x2,
        z = surface2, col = "orange", xlab="Wind", ylab="Temp",
        zlab="Ozone")
points3d(data$Ozone ~ data$Wind + data$, col="blue")
```



It looks like when Wind get larger, Ozone will get smaller.
It looks like Temp get larger, Ozone will also get larger

4. Add the interaction of the two region variables to the model

(a) **Report the results from summary().**

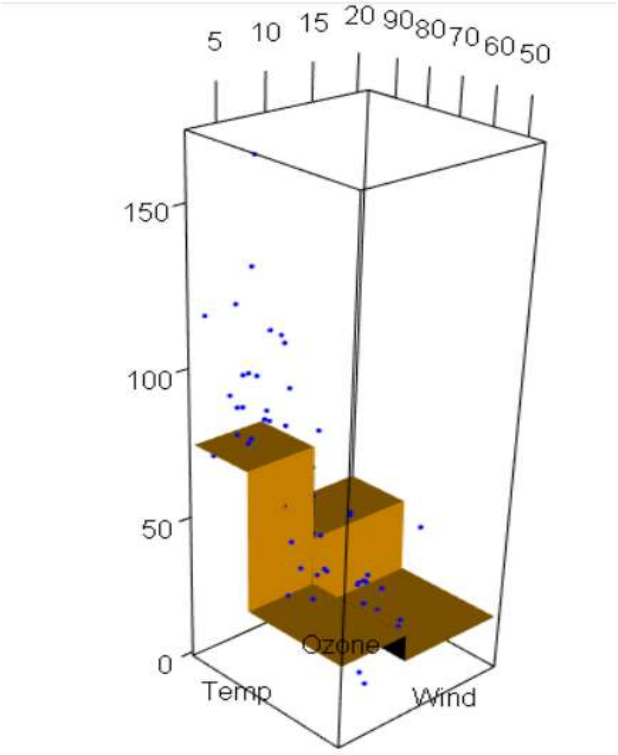
```
# 4. Add the interaction of the two region variables to the model
# (a) Report the results from summary().
mod.2step2 = lm(data$Ozone ~ wind.hilo + temp.hilo + wind.hilo * temp.hilo)
summary(mod.2step2)
> summary(mod.2step2)

Call:
lm(formula = data$Ozone ~ wind.hilo + temp.hilo + wind.hilo *
    temp.hilo)

Residuals:
    Min      1Q  Median      3Q     Max
-59.000 -10.175  -1.683   9.167  93.000

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                  42.667      4.968   8.589 7.78e-14 ***
wind.hiloTRUE                32.333      6.251   5.173 1.08e-06 ***
temp.hiloTRUE               -24.984      6.109  -4.090 8.38e-05 ***
wind.hiloTRUE:temp.hiloTRUE -22.939      9.570  -2.397   0.0183 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.76 on 107 degrees of freedom
Multiple R-squared:  0.5447,    Adjusted R-squared:  0.532
F-statistic: 42.68 on 3 and 107 DF,  p-value: < 2.2e-16
```

(b) Does the interaction have statistically significant influence on the mean ozone level at the 5% Type 1 error rate? **Report the p-values and your conclusion. (No hypotheses needed.)**

```
wind.hiloTRUE                32.333      6.251   5.173 1.08e-06 ***
temp.hiloTRUE               -24.984      6.109  -4.090 8.38e-05 ***
wind.hiloTRUE:temp.hiloTRUE -22.939      9.570  -2.397   0.0183 *
```

-> Yes, because all of the values are below 0.05

(c) Make a 3-D plot of the surface. **Report a screenshot from some angle that shows the whole surface and describe how the interaction affects the surface (use one sentence).**



Couldn't find that much difference.