

R Notebook

The following is your first chunk to start with. Remember, you can add chunks using the menu above (Insert -> R) or using the keyboard shortcut Ctrl+Alt+I. A good practice is to use different code chunks to answer different questions. You can delete this comment if you like.

Other useful keyboard shortcuts include Alt- for the assignment operator, and Ctrl+Shift+M for the pipe operator. You can delete these reminders if you don't want them in your report.

```
#setwd("../")

library("tidyverse")

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.2.1      v purrr  0.3.3
## v tibble  2.1.3      v dplyr  0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library("tidymodels")

## Registered S3 method overwritten by 'xts':
##   method      from
##   as.zoo.xts  zoo

## -- Attaching packages ----- tidymodels 0.0.3 --

## v broom      0.5.3      v recipes  0.1.9
## v dials      0.0.4      v rsample  0.0.5
## v infer      0.5.1      v yardstick 0.0.4
## v parsnip    0.0.5

## -- Conflicts ----- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()       masks stats::lag()
## x dials::margin()   masks ggplot2::margin()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## x recipes::yj_trans() masks scales::yj_trans()
```

```

library("plotly")

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout

library("skimr")

```

Data preparation Create the additional variables: Create the COUNT variable and add it to the data frame. Extract MONTH from the DATE variable and add it to the data frame. This time, do NOT use lubridate. Use the base months() function instead. Scale the data (and save it as dfbStd): Start by standardizing the four variables, TEMP, ATEMP, HUMIDITY, WINDSPEED. If you don't remember what it means to standardize a variable, see the link. Surely, you don't need to do this manually!

```

dfbOrg <-
  read_csv("assignment2BikeShare.csv")

## Parsed with column specification:
## cols(
##   DATE = col_date(format = ""),
##   HOLIDAY = col_character(),
##   WEEKDAY = col_character(),
##   WEATHERSIT = col_double(),
##   TEMP = col_double(),
##   ATEMP = col_double(),
##   HUMIDITY = col_double(),
##   WINDSPEED = col_double(),
##   CASUAL = col_double(),
##   REGISTERED = col_double()
## )

Sys.setlocale("LC_TIME", "C")

## [1] "C"

```

Data preparation Create the additional variables: Create the COUNT variable and add it to the data frame. Extract MONTH from the DATE variable and add it to the data frame. This time, do NOT use lubridate. Use the base months() function instead. Scale the data (and save it as dfbStd): Start by standardizing the four variables, TEMP, ATEMP, HUMIDITY,

WINDSPEED. If you don't remember what it means to standardize a variable, see the link. Surely, you don't need to do this manually!

```
skim(dfbOrg)
```

Data summary

Name	dfbOrg
Number of rows	731
Number of columns	10

Column type frequency:

character	2
Date	1
numeric	7

Group variables	None
-----------------	------

Variable type: character





skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
HOLIDAY	0	1	2	3	0	2	0
WEEKDAY	0	1	2	3	0	2	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
DATE	0	1	2011-01-01	2012-12-31	2012-01-01	731

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
WEATHERSIT	0	1	1.40	0.54	1	1.0	1	2.00	3.00	
TEMP	0	1	15.87	8.83	1	8.0	16	23.15	34.00	
ATEMP	0	1	16.00	9.67	1	6.6	16	23.95	41.00	
HUMIDITY	0	1	63.17	15.47	17	51.0	62	74.00	100.00	
WINDSPEED	0	1	12.82	5.54	0	9.0	12	16.00	40.16	

ED											
CASUAL	0	1	848.1	686.6	2	315.	71	1096.	3410.		
			8	2		5	3	00	00		
REGISTER	0	1	3656.	1560.	2	2497	36	4776.	6946.		
ED			17	26	0	.0	62	50	00		

Q1)a)i)

```
dfbOrg<-dfbOrg%>%mutate(Count = CASUAL+REGISTERED)
dfbOrg%>%arrange(desc(Count))

## # A tibble: 731 x 11
##   DATE          HOLIDAY WEEKDAY WEATHERSIT  TEMP ATEMP HUMIDITY WINDSPEED
CASUAL
##   <date>      <chr>   <chr>          <dbl> <dbl> <dbl>    <dbl>    <dbl>
<dbl>
## 1 2012-12-31 NO      YES            2    3    2      59      11
439
## 2 2012-12-30 NO      NO             1    4    1      51      24
364
## 3 2012-12-29 NO      NO             2    3    2      75       6
159
## 4 2012-12-28 NO      YES            2    3    1      60       9
644
## 5 2012-12-27 NO      YES            2    3    1      60      21
247
## 6 2012-12-26 NO      YES            3    2.5  3      87      22
9
## 7 2012-12-25 YES     YES            2    5.5  3.5    73.5     11
440
## 8 2012-12-24 NO      YES            2    3    1      80       6
174
## 9 2012-12-23 NO      NO             1    3    3      51       9
408
## 10 2012-12-22 NO      NO             1    4    1      45      28
205
## # ... with 721 more rows, and 2 more variables: REGISTERED <dbl>, Count
<dbl>
```

Q2)a)

```
dfbOrg<-dfbOrg%>%mutate(month=months(Count))

dfbOrg

## # A tibble: 731 x 12
##   DATE          HOLIDAY WEEKDAY WEATHERSIT  TEMP ATEMP HUMIDITY WINDSPEED
CASUAL
##   <date>      <chr>   <chr>          <dbl> <dbl> <dbl>    <dbl>    <dbl>
<dbl>
```

```
## 1 2011-01-01 NO NO 2 11 11 81 17
331
## 2 2011-01-02 NO NO 2 9 6.5 71.5 17
131
## 3 2011-01-03 NO YES 1 1 4 44 18
120
## 4 2011-01-04 NO YES 1 2 2.5 64 9
108
## 5 2011-01-05 NO YES 1 2.5 1 42.5 13
82
## 6 2011-01-06 NO YES 1 2 2 52 6
88
## 7 2011-01-07 NO YES 2 1 3 47.5 11
148
## 8 2011-01-08 NO NO 2 1 5 51 17
68
## 9 2011-01-09 NO NO 1 2 8.5 46 25
54
## 10 2011-01-10 NO YES 1 2 6 50 15
41
## # ... with 721 more rows, and 3 more variables: REGISTERED <dbl>, Count
<dbl>,
## # month <chr>
```

Q1)b)

```
dfbStd<-
dfbOrg%>%mutate(TEMP=scale(TEMP),ATEMP=scale(ATEMP),HUMIDITY=scale(HUMIDITY),
WINDSPEED=scale(WINDSPEED))
dfbStd

## # A tibble: 731 x 12
## DATE HOLIDAY WEEKDAY WEATHERSIT TEMP[,1] ATEMP[,1] HUMIDITY[,1]
## <date> <chr> <chr> <dbl> <dbl> <dbl> <dbl>
## 1 2011-01-01 NO NO 2 -0.552 -0.517 1.15
## 2 2011-01-02 NO NO 2 -0.779 -0.982 0.538
## 3 2011-01-03 NO YES 1 -1.68 -1.24 -1.24
## 4 2011-01-04 NO YES 1 -1.57 -1.40 0.0536
## 5 2011-01-05 NO YES 1 -1.51 -1.55 -1.34
## 6 2011-01-06 NO YES 1 -1.57 -1.45 -0.722
## 7 2011-01-07 NO YES 2 -1.68 -1.34 -1.01
## 8 2011-01-08 NO NO 2 -1.68 -1.14 -0.787
## 9 2011-01-09 NO NO 1 -1.57 -0.775 -1.11
## 10 2011-01-10 NO YES 1 -1.57 -1.03 -0.852
## # ... with 721 more rows, and 5 more variables: WINDSPEED[,1] <dbl>,
## # CASUAL <dbl>, REGISTERED <dbl>, Count <dbl>, month <chr>
```

Q2)

```
fitAll<-lm(formula=Count ~
DATE+HOLIDAY+WEEKDAY+WEATHERSIT+WINDSPEED+REGISTERED+CASUAL+REGISTERED+TEMP+A
```

```
TEMP+HUMIDITY+month, data=dfbOrg)
```

```
summary(fitAll)
```

```
## Warning in summary.lm(fitAll): essentially perfect fit: summary may be
## unreliable
```

```
##
```

```
## Call:
```

```
## lm(formula = Count ~ DATE + HOLIDAY + WEEKDAY + WEATHERSIT +
##     WINDSPEED + REGISTERED + CASUAL + REGISTERED + TEMP + ATEMP +
##     HUMIDITY + month, data = dfbOrg)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1.742e-11 -1.674e-13 -1.900e-14  1.341e-13  2.761e-11
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  -9.277e-11  7.340e-12 -1.264e+01 < 2e-16 ***
## DATE          6.319e-15  4.960e-16  1.274e+01 < 2e-16 ***
## HOLIDAYYES    -1.475e-12  3.658e-13 -4.033e+00 6.11e-05 ***
## WEEKDAYYES     1.514e-13  2.065e-13  7.330e-01  0.4638
## WEATHERSIT    -1.126e-13  1.406e-13 -8.010e-01  0.4236
## WINDSPEED     -2.449e-14  1.147e-14 -2.134e+00  0.0332 *
## REGISTERED     1.000e+00  8.452e-17  1.183e+16 < 2e-16 ***
## CASUAL         1.000e+00  1.566e-16  6.384e+15 < 2e-16 ***
## TEMP          -3.086e-14  4.761e-14 -6.480e-01  0.5170
## ATEMP          3.581e-14  4.070e-14  8.800e-01  0.3792
## HUMIDITY       2.631e-16  5.250e-15  5.000e-02  0.9600
## monthAugust   -1.221e-13  3.268e-13 -3.740e-01  0.7089
## monthDecember  1.467e-13  3.342e-13  4.390e-01  0.6608
## monthFebruary  1.171e-13  3.112e-13  3.760e-01  0.7067
## monthJanuary   3.313e-13  3.314e-13  1.000e+00  0.3177
## monthJuly      -1.054e-13  3.541e-13 -2.980e-01  0.7660
## monthJune      -1.389e-13  3.191e-13 -4.350e-01  0.6633
## monthMarch     4.200e-14  2.759e-13  1.520e-01  0.8790
## monthMay       -2.738e-14  2.870e-13 -9.500e-02  0.9240
## monthNovember  1.227e-13  3.068e-13  4.000e-01  0.6894
## monthOctober   3.624e-14  2.818e-13  1.290e-01  0.8977
## monthSeptember -8.313e-15  3.001e-13 -2.800e-02  0.9779
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 1.477e-12 on 709 degrees of freedom
```

```
## Multiple R-squared:  1, Adjusted R-squared:  1
```

```
## F-statistic: 5.98e+31 on 21 and 709 DF, p-value: < 2.2e-16
```

```
##?scale
```

Q3)a)

#Q3)a)

```
#dfbOrg5<-dfbOrg
dfbOrg<-dfbOrg%>%
```

```
mutate(BADWEATHER = ifelse(WEATHERSIT==3|WEATHERSIT==4, "YES", "NO"))
```

dfbOrg

```
## # A tibble: 731 x 13
```

```
##   DATE          HOLIDAY WEEKDAY WEATHERSIT  TEMP  ATEMP  HUMIDITY  WINDSPEED
CASUAL
```

```
##   <date>      <chr>   <chr>         <dbl> <dbl> <dbl>    <dbl>    <dbl>
<dbl>
```

```
## 1 2011-01-01 NO      NO              2  11    11      81        17
```

331

```
## 2 2011-01-02 NO      NO              2   9     6.5    71.5       17
```

131

```
## 3 2011-01-03 NO      YES             1   1     4      44        18
```

120

```
## 4 2011-01-04 NO      YES             1   2     2.5    64         9
```

108

```
## 5 2011-01-05 NO      YES             1  2.5     1    42.5       13
```

82

```
## 6 2011-01-06 NO      YES             1   2     2      52         6
```

88

```
## 7 2011-01-07 NO      YES             2   1     3     47.5      11
```

148

```
## 8 2011-01-08 NO      NO              2   1     5      51        17
```

68

```
## 9 2011-01-09 NO      NO              1   2     8.5    46         25
```

54

```
## 10 2011-01-10 NO     YES             1   2     6      50        15
```

41

```
## # ... with 721 more rows, and 4 more variables: REGISTERED <dbl>, Count
<dbl>,
```

```
## #   month <chr>, BADWEATHER <chr>
```

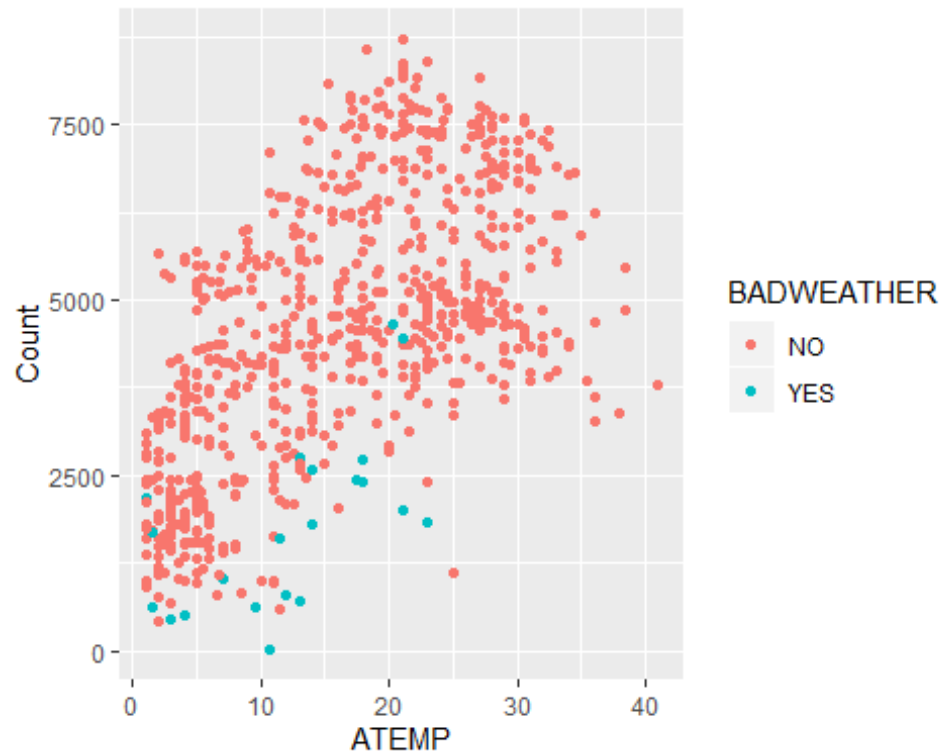
Q3)b)

```
SCATTER <-
```

```
dfbOrg %>%
```

```
ggplot(aes(x =ATEMP , y = Count, color = BADWEATHER)) + geom_point()
```

SCATTER

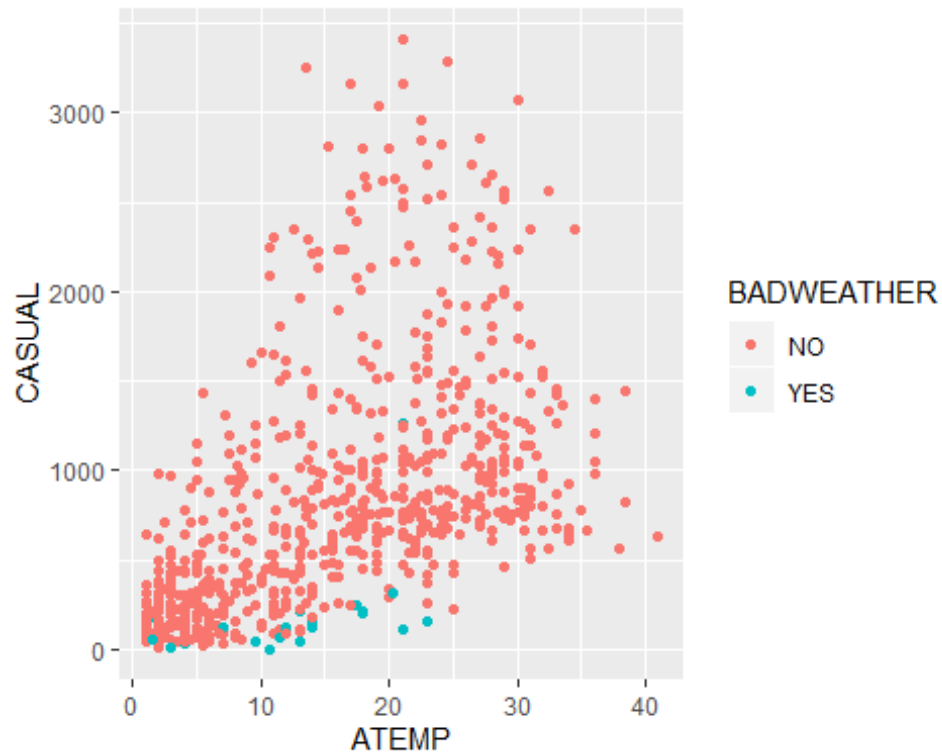


```
ggplotly(SCATTER)
```

Q3)c)

```
CASUAL <-  
dfbOrg %>%  
ggplot(aes(x =ATEMP , y = CASUAL, color = BADWEATHER)) + geom_point()
```

CASUAL

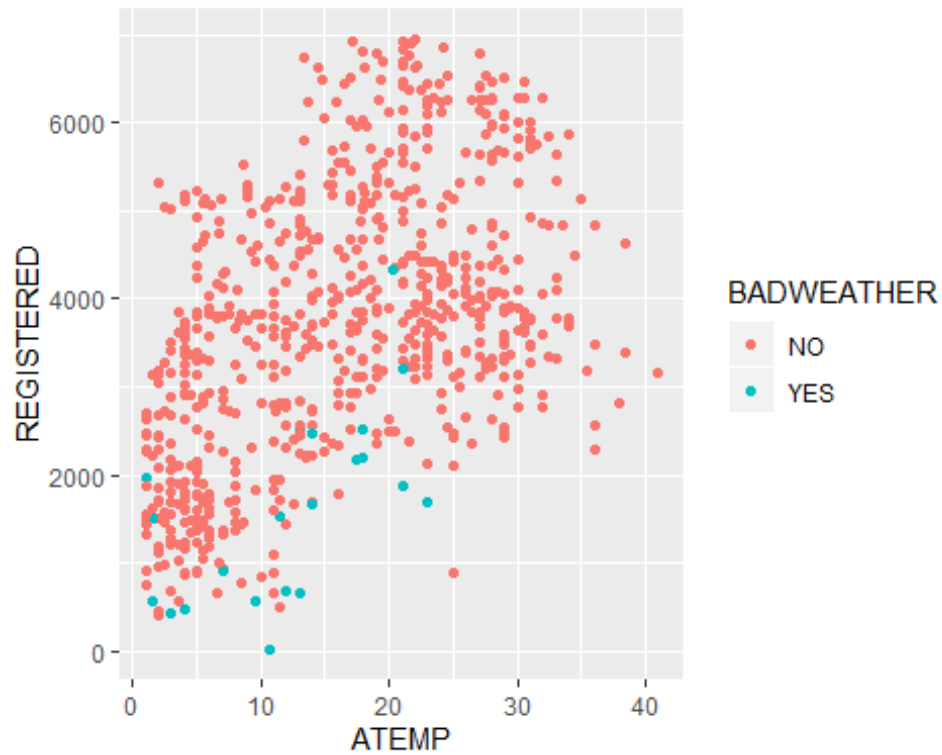


```
ggplotly(CASUAL)
```

Q3)c)

```
REGISTERED <-  
dfbOrg %>%  
ggplot(aes(x =ATEMP , y = REGISTERED, color = BADWEATHER)) + geom_point()
```

```
REGISTERED
```



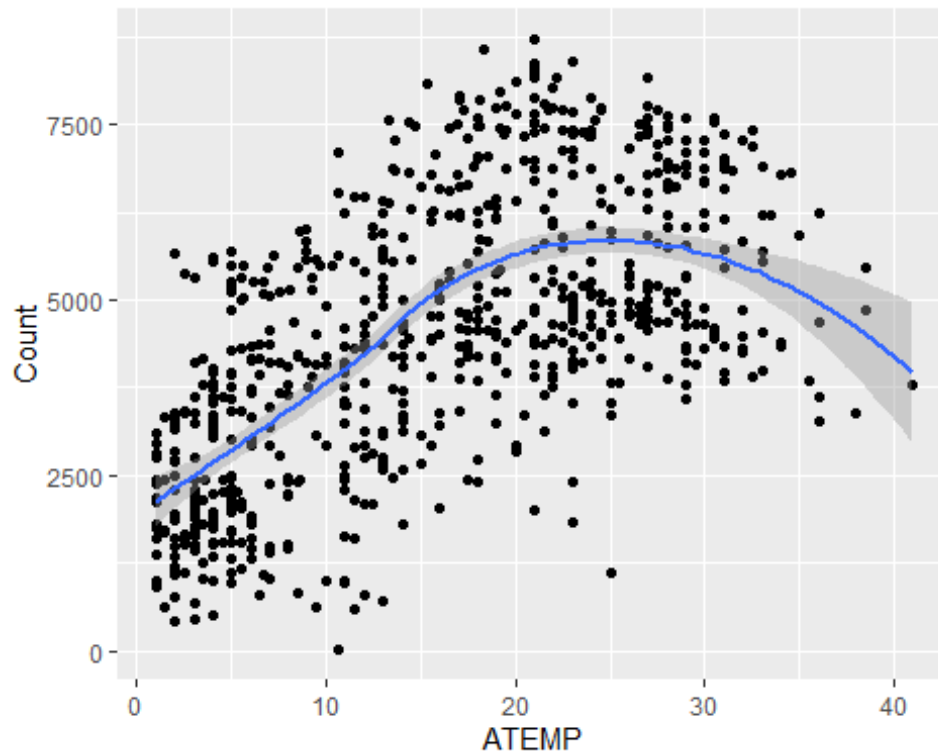
```
ggplotly(REGISTERED)
```

Q3)c)iv)

```
plot<-dfbOrg %>%  
  ggplot(aes(x =ATEMP , y = Count)) + geom_point()+geom_smooth()
```

plot

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
ggplotly(plot)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Q4)

```
dfbOrg
```

```
## # A tibble: 731 x 13
```

```
##   DATE      HOLIDAY WEEKDAY WEATHERSIT  TEMP ATEMP HUMIDITY WINDSPEED
CASUAL
```

```
##   <date>     <chr>   <chr>         <dbl> <dbl> <dbl>    <dbl>    <dbl>
<dbl>
```

```
## 1 2011-01-01 NO      NO              2  11    11      81        17
331
```

```
## 2 2011-01-02 NO      NO              2   9     6.5    71.5       17
131
```

```
## 3 2011-01-03 NO      YES             1   1     4      44        18
120
```

```
## 4 2011-01-04 NO      YES             1   2     2.5    64         9
108
```

```
## 5 2011-01-05 NO      YES             1  2.5    1     42.5       13
82
```

```
## 6 2011-01-06 NO      YES             1   2     2      52         6
88
```

```
## 7 2011-01-07 NO      YES             2   1     3     47.5       11
148
```

```

## 8 2011-01-08 NO      NO      2      1      5      51      17
68
## 9 2011-01-09 NO      NO      1      2      8.5    46      25
54
## 10 2011-01-10 NO      YES     1      2      6      50      15
41
## # ... with 721 more rows, and 4 more variables: REGISTERED <dbl>, Count
<dbl>,
## #   month <chr>, BADWEATHER <chr>

fit1<-lm(formula=Count ~ WEEKDAY+BADWEATHER+TEMP+ATEMP+HUMIDITY+month,
data=dfbOrg)
summary(fit1)

##
## Call:
## lm(formula = Count ~ WEEKDAY + BADWEATHER + TEMP + ATEMP + HUMIDITY +
##     month, data = dfbOrg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3729.0 -1005.1  -190.3   1115.0   3750.1
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3967.981    335.628   11.823 < 2e-16 ***
## WEEKDAYYES         69.745    110.118    0.633  0.52670
## BADWEATHERYES  -1954.835    316.601   -6.174 1.11e-09 ***
## TEMP             184.596     42.011    4.394 1.28e-05 ***
## ATEMP            -48.640     36.621   -1.328  0.18454
## HUMIDITY         -25.341      3.623   -6.995 6.09e-12 ***
## monthAugust     -209.660    291.004   -0.720  0.47147
## monthDecember    105.664    265.660    0.398  0.69094
## monthFebruary   -802.319    273.000   -2.939  0.00340 **
## monthJanuary    -858.334    293.371   -2.926  0.00355 **
## monthJuly       -676.644    312.956   -2.162  0.03094 *
## monthJune       -189.229    286.067   -0.661  0.50851
## monthMarch      -242.020    249.333   -0.971  0.33204
## monthMay         279.730    259.634    1.077  0.28166
## monthNovember    651.966    257.460    2.532  0.01154 *
## monthOctober    1072.312    246.970    4.342 1.62e-05 ***
## monthSeptember   742.473    267.293    2.778  0.00562 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1341 on 714 degrees of freedom
## Multiple R-squared:  0.5315, Adjusted R-squared:  0.521
## F-statistic: 50.64 on 16 and 714 DF, p-value: < 2.2e-16

?months

```

```
## starting httpd help server ... done
```

Q5)

```
library(car)
```

```
## Loading required package: carData
```

```
## Registered S3 methods overwritten by 'car':
```

```
##   method                      from
```

```
##   influence.merMod            lme4
```

```
##   cooks.distance.influence.merMod lme4
```

```
##   dfbeta.influence.merMod      lme4
```

```
##   dfbetas.influence.merMod     lme4
```

```
##
```

```
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##   recode
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##   some
```

```
outlierTest(fit1) # Bonferroni p-value for most extreme obs
```

```
## No Studentized residuals with Bonferroni  $p < 0.05$ 
```

```
## Largest |rstudent|:
```

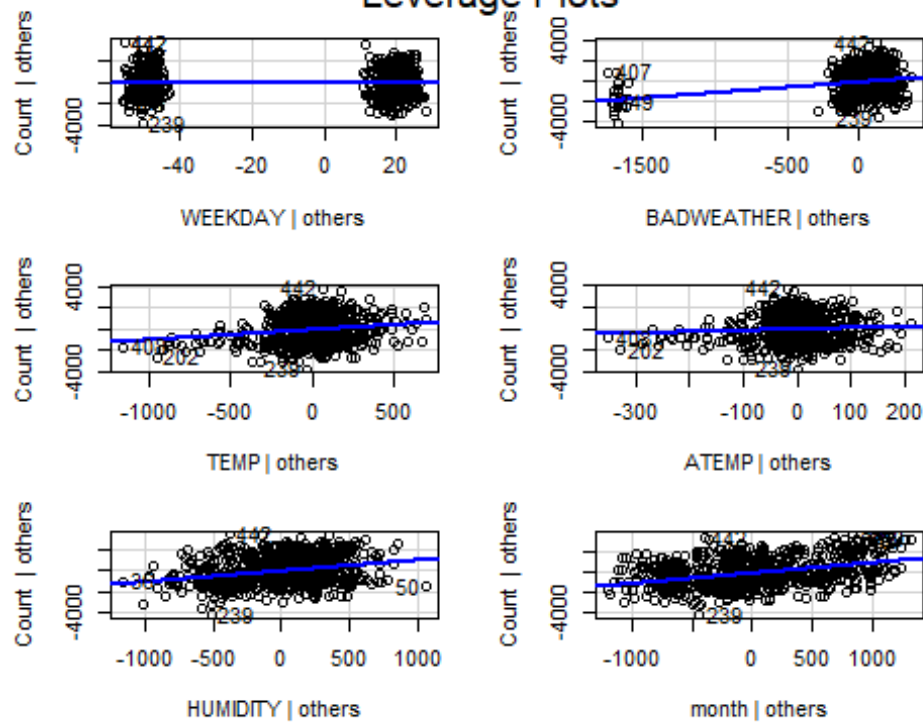
```
##   rstudent unadjusted p-value Bonferroni p
```

```
## 442  2.84885          0.0045142          NA
```

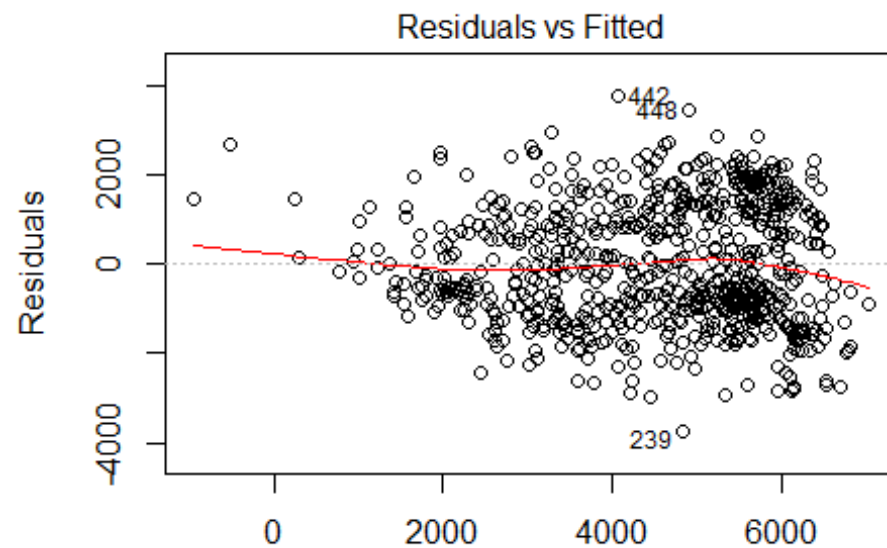
```
#qqPlot(fit1) #qq plot for studentized resid
```

```
leveragePlots(fit1)
```

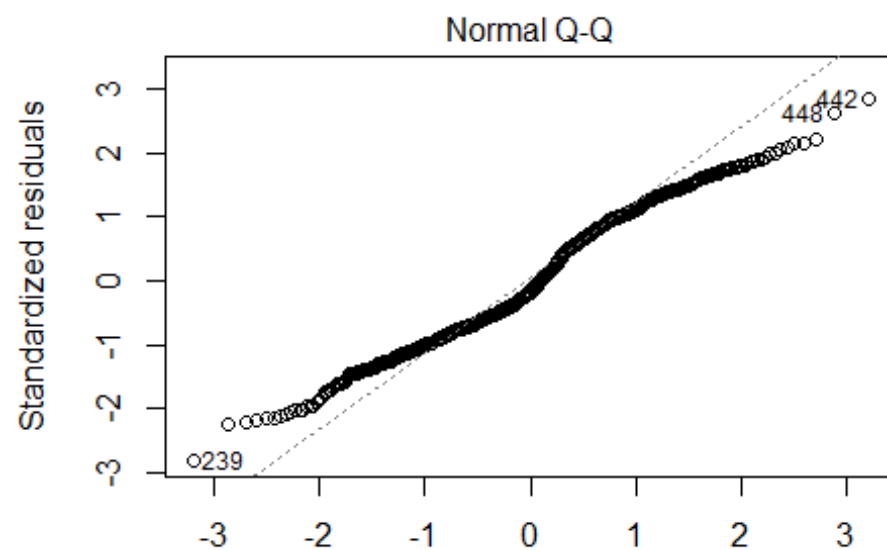
Leverage Plots



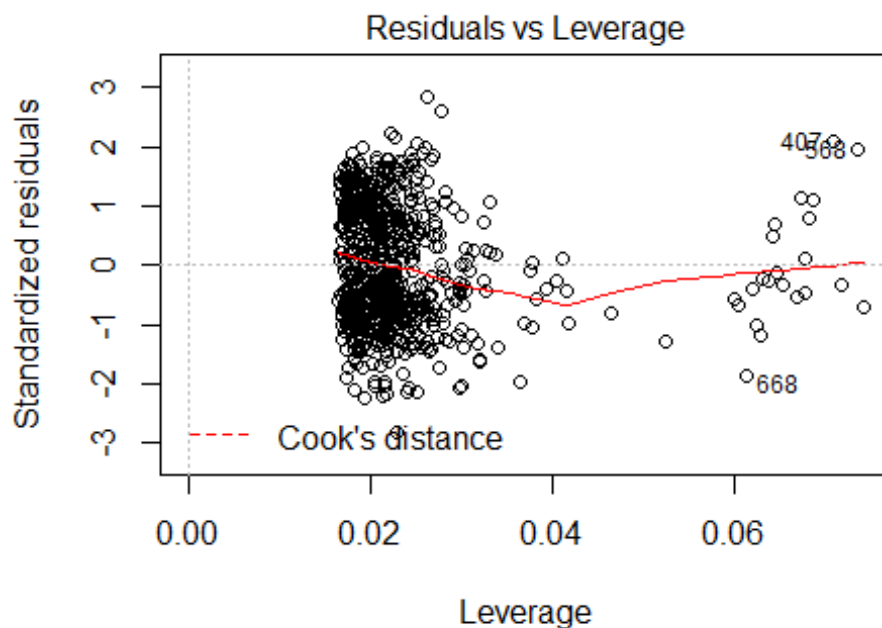
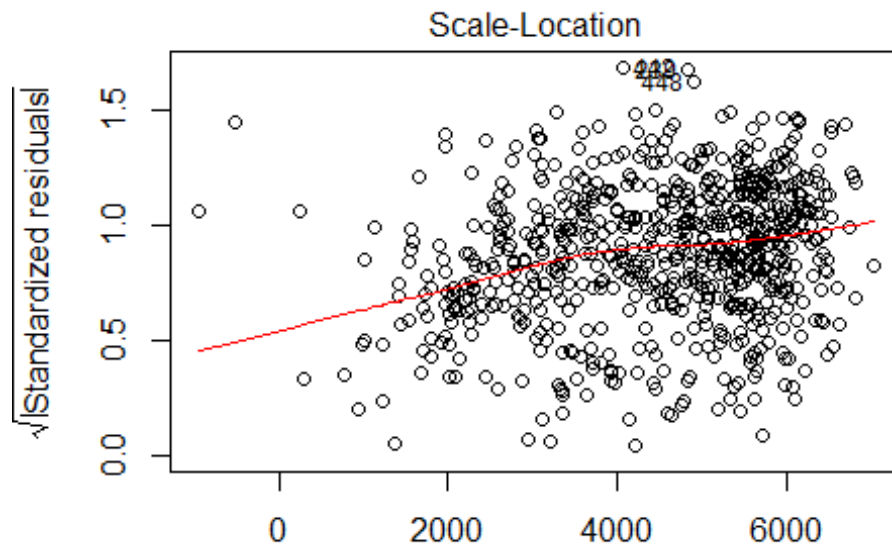
```
#plot(fit)
plot(fit1)
```



Fitted values
 $\text{Count} \sim \text{WEEKDAY} + \text{BADWEATHER} + \text{TEMP} + \text{ATEMP} + \text{HUMIDITY}$



Theoretical Quantiles
 $\text{Count} \sim \text{WEEKDAY} + \text{BADWEATHER} + \text{TEMP} + \text{ATEMP} + \text{HUMIDITY}$



```
vif(fit1)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## WEEKDAY    1.009743 1      1.004859
```



```
## BADWEATHER 1.137470 1 1.066522
## TEMP 55.856782 1 7.473739
## ATEMP 50.923158 1 7.136046
## HUMIDITY 1.275120 1 1.129212
## month 8.480466 11 1.102049
```

```
Afterdiagnosis<-lm(formula=Count ~ WEEKDAY+BADWEATHER+TEMP+HUMIDITY+month,
data=dfbOrg)
vif(Afterdiagnosis)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## WEEKDAY 1.009459 1 1.004718
## BADWEATHER 1.136911 1 1.066260
## TEMP 6.138377 1 2.477575
## HUMIDITY 1.272421 1 1.128017
## month 6.968140 11 1.092254
```

Q6)a)

```
fitbadweather<-lm(formula=Count ~ BADWEATHER, data=dfbOrg)
summary(fitbadweather)
```

```
##
## Call:
## lm(formula = Count ~ BADWEATHER, data = dfbOrg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4153.2 -1257.7      1.8  1404.8  4129.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4584.24      70.63  64.908 < 2e-16 ***
## BADWEATHERYES -2780.95     416.69  -6.674 4.93e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1882 on 729 degrees of freedom
## Multiple R-squared:  0.05758,    Adjusted R-squared:  0.05629
## F-statistic: 44.54 on 1 and 729 DF,  p-value: 4.934e-11
```

Q6)c)

```
fitweekend<-lm(formula=Count ~ BADWEATHER+WEEKDAY, data=dfbOrg)
summary(fitweekend)
```

```
##
## Call:
## lm(formula = Count ~ BADWEATHER + WEEKDAY, data = dfbOrg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -4205.2 -1263.2    -6.2  1406.8  4257.9
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4456.1      130.2   34.223 < 2e-16 ***
## BADWEATHERYES -2790.1      416.7   -6.696 4.27e-11 ***
## WEEKDAYYES    180.1       153.8    1.171  0.242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1881 on 728 degrees of freedom
## Multiple R-squared:  0.05935,    Adjusted R-squared:  0.05677
## F-statistic: 22.97 on 2 and 728 DF,  p-value: 2.126e-10
```

```
anova(fitbadweather,fitweekend)
```

```
## Analysis of Variance Table
##
## Model 1: Count ~ BADWEATHER
## Model 2: Count ~ BADWEATHER + WEEKDAY
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1      729 2581793230
## 2      728 2576938350  1   4854879 1.3715 0.2419
```

7)a)b)

```
set.seed(333)
dfbTrain <- dfbOrg %>% sample_frac(0.8)
dfbTest <- dplyr::setdiff(dfbOrg, dfbTrain)
```

```
dfbTest
```

```
## # A tibble: 146 x 13
##   DATE          HOLIDAY WEEKDAY WEATHERSIT  TEMP ATEMP HUMIDITY WINDSPEED
##   <date>      <chr>   <chr>      <dbl> <dbl> <dbl>    <dbl>    <dbl>
## 1 2011-01-10 NO      YES        1     2     6      50      15
## 2 2011-01-11 NO      YES        2     1    3.5     57      7
## 3 2011-01-13 NO      YES        1     2     7     48.5     20
## 4 2011-01-16 NO      NO         1    2.5    2     49.5     15
## 5 2011-01-19 NO      YES        2    5.5    2.5    71.5     10
## 6 2011-01-20 NO      YES        2     4     2     56      15
## 7 2011-01-23 NO      NO         1     4    10     42      15
```

```
## 8 2011-01-25 NO      YES      2      2      4      65      9
186
## 9 2011-02-13 NO      NO       1     9.5     6      36     20
397
## 10 2011-02-15 NO     YES      1      4     3.5     32     17
140
## # ... with 136 more rows, and 4 more variables: REGISTERED <dbl>, Count
<dbl>,
## #   month <chr>, BADWEATHER <chr>
```

Q7)c)

```
fitOrg<-lm(formula=Count ~ WEEKDAY+BADWEATHER+TEMP+HUMIDITY+month,
data=dfbTrain)
```

```
resultsOrg <-
```

```
  dfbTest %>%
```

```
  mutate(predictedCount = predict(fitOrg, dfbTest))
```

```
performance<-metric_set(rmse,mae)
summary(fitOrg)
```

```
##
## Call:
## lm(formula = Count ~ WEEKDAY + BADWEATHER + TEMP + HUMIDITY +
##     month, data = dfbTrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3694.4 -1026.4  -107.5   1135.6   3853.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4269.255    363.582   11.742 < 2e-16 ***
## WEEKDAYYES         69.967    123.314    0.567 0.570676
## BADWEATHERYES  -2136.339    364.147   -5.867 7.55e-09 ***
## TEMP             128.757     15.477    8.319 6.57e-16 ***
## HUMIDITY        -26.420      4.054   -6.517 1.59e-10 ***
## monthAugust     -330.744    323.715   -1.022 0.307350
## monthDecember    54.772    294.299    0.186 0.852425
## monthFebruary  -936.827    304.498   -3.077 0.002194 **
## monthJanuary  -1051.248    316.710   -3.319 0.000960 ***
## monthJuly       -658.969    341.332   -1.931 0.054032 .
## monthJune       -237.006    310.371   -0.764 0.445409
## monthMarch      -441.308    277.602   -1.590 0.112454
## monthMay         240.150    286.760    0.837 0.402687
## monthNovember   570.874    290.930    1.962 0.050222 .
## monthOctober   1002.527    278.826    3.596 0.000352 ***
```

```

## monthSeptember    573.889    300.255    1.911 0.056463 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1340 on 569 degrees of freedom
## Multiple R-squared:  0.5319, Adjusted R-squared:  0.5196
## F-statistic: 43.1 on 15 and 569 DF,  p-value: < 2.2e-16

Model<-performance(resultsOrg,truth=Count,estimate=predictedCount)

Model

## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 rmse    standard      1361.
## 2 mae     standard      1154.

fitNew<-lm(formula=Count ~ WEEKDAY+BADWEATHER+TEMP+HUMIDITY+month+WINDSPEED,
data=dfbTrain)

resultsNew <-

  dfbTest %>%

  mutate(predictedCount = predict(fitNew, dfbTest))

performance<-metric_set(rmse,mae)
summary(fitNew)

##
## Call:
## lm(formula = Count ~ WEEKDAY + BADWEATHER + TEMP + HUMIDITY +
##     month + WINDSPEED, data = dfbTrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3249.8 -1020.9  -160.1   1111.3   3505.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5573.353     427.492   13.037 < 2e-16 ***
## WEEKDAYYES         49.775     120.357    0.414 0.679352
## BADWEATHERYES  -1792.286     360.782   -4.968 8.97e-07 ***
## TEMP             126.868      15.102    8.400 3.57e-16 ***
## HUMIDITY        -32.619       4.115   -7.927 1.19e-14 ***
## monthAugust    -451.787     316.579   -1.427 0.154103
## monthDecember  -161.016     289.808   -0.556 0.578706
## monthFebruary -1042.941     297.690   -3.503 0.000495 ***
## monthJanuary  -1150.433     309.502   -3.717 0.000222 ***

```

```
## monthJuly      -886.846    335.590  -2.643 0.008453 **
## monthJune      -378.323    303.887  -1.245 0.213666
## monthMarch     -476.895    270.895  -1.760 0.078871 .
## monthMay       133.492    280.432   0.476 0.634240
## monthNovember  418.379    285.188   1.467 0.142922
## monthOctober   840.007    273.632   3.070 0.002244 **
## monthSeptember 401.891    294.603   1.364 0.173051
## WINDSPEED      -58.647     10.731  -5.465 6.93e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1307 on 568 degrees of freedom
## Multiple R-squared:  0.5553, Adjusted R-squared:  0.5428
## F-statistic: 44.33 on 16 and 568 DF, p-value: < 2.2e-16

Model2<-performance(resultsNew,truth=Count,estimate=predictedCount)

Model2

## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 rmse    standard      1317.
## 2 mae     standard      1136.

summary(fitNew)

##
## Call:
## lm(formula = Count ~ WEEKDAY + BADWEATHER + TEMP + HUMIDITY +
##     month + WINDSPEED, data = dfbTrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3249.8 -1020.9  -160.1   1111.3   3505.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5573.353    427.492   13.037 < 2e-16 ***
## WEEKDAYYES      49.775    120.357    0.414 0.679352
## BADWEATHERYES -1792.286    360.782  -4.968 8.97e-07 ***
## TEMP          126.868     15.102    8.400 3.57e-16 ***
## HUMIDITY       -32.619      4.115   -7.927 1.19e-14 ***
## monthAugust   -451.787    316.579  -1.427 0.154103
## monthDecember -161.016    289.808  -0.556 0.578706
## monthFebruary -1042.941    297.690  -3.503 0.000495 ***
## monthJanuary  -1150.433    309.502  -3.717 0.000222 ***
## monthJuly     -886.846    335.590  -2.643 0.008453 **
## monthJune     -378.323    303.887  -1.245 0.213666
## monthMarch    -476.895    270.895  -1.760 0.078871 .
## monthMay      133.492    280.432   0.476 0.634240
```

```
## monthNovember    418.379    285.188    1.467 0.142922
## monthOctober     840.007    273.632    3.070 0.002244 **
## monthSeptember   401.891    294.603    1.364 0.173051
## WINDSPEED        -58.647     10.731   -5.465 6.93e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1307 on 568 degrees of freedom
## Multiple R-squared:  0.5553, Adjusted R-squared:  0.5428
## F-statistic: 44.33 on 16 and 568 DF, p-value: < 2.2e-16
```

#Q8) #i)

```
library("lubridate")

##
## Attaching package: 'lubridate'

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

#detach('package:modelr', unload=TRUE)
set.seed(333)
dfbTrain2 <- dfbOrg %>% filter(year(Date)==2011)

dfbTest2 <- dfbOrg %>% filter(year(Date)==2012)

fitOrg2<-lm(formula=Count ~ WEEKDAY+BADWEATHER+TEMP+HUMIDITY+month,
data=dfbTrain2)

resultsOrg2 <-

  dfbTest2 %>%

  mutate(predictedCount = predict(fitOrg2, dfbTest2))

performance<-metric_set(rmse,mae)

Model3<-performance(resultsOrg2,truth=Count,estimate=predictedCount)

Model3

## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>      <dbl>
## 1 rmse    standard      2371.
## 2 mae     standard      2181.

summary(fitOrg2)
```

```
##
## Call:
## lm(formula = Count ~ WEEKDAY + BADWEATHER + TEMP + HUMIDITY +
##      month, data = dfbTrain2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2901.32  -287.54   40.61   366.05  2042.39
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3208.903     242.121   13.253 < 2e-16 ***
## WEEKDAYYES         7.973       74.352    0.107 0.914662
## BADWEATHERYES  -1420.313     184.521   -7.697 1.44e-13 ***
## TEMP             58.530       10.000    5.853 1.12e-08 ***
## HUMIDITY        -12.966        2.472   -5.245 2.72e-07 ***
## monthAugust      491.865     199.801    2.462 0.014308 *
## monthDecember    102.576     178.069    0.576 0.564956
## monthFebruary  -1100.065     189.951   -5.791 1.56e-08 ***
## monthJanuary   -1339.683     202.103   -6.629 1.29e-10 ***
## monthJuly       429.939     217.511    1.977 0.048871 *
## monthJune       837.684     199.663    4.195 3.46e-05 ***
## monthMarch     -744.946     177.380   -4.200 3.40e-05 ***
## monthMay        928.188     172.333    5.386 1.33e-07 ***
## monthNovember   588.880     169.362    3.477 0.000571 ***
## monthOctober   1009.738     164.474    6.139 2.26e-09 ***
## monthSeptember  933.138     180.209    5.178 3.80e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 638.7 on 349 degrees of freedom
## Multiple R-squared:  0.7942, Adjusted R-squared:  0.7854
## F-statistic: 89.8 on 15 and 349 DF, p-value: < 2.2e-16

#dfbOrg%>%mutate(date=as.year(DATE))
```

Q8)ii)

```
set.seed(333)
dfbTrain3 <- dfbOrg %>% filter(DATE>="2011-01-01" & DATE<="2012-06-30"))

dfbTest3 <- dfbOrg %>% filter(DATE>("2012-06-30") & year(DATE)<2013)

fitOrg3<-lm(formula=Count ~ WEEKDAY+BADWEATHER+TEMP+HUMIDITY+month,
data=dfbTrain3)

resultsOrg3 <-

dfbTest3 %>%
```

```
mutate(predictCount = predict(fitOrg3, dfbTest3))

performance<-metric_set(rmse,mae)

Model4<-performance(resultsOrg3,truth=Count,estimate=predictCount)

Model4

## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>      <dbl>
## 1 rmse    standard    2387.
## 2 mae     standard    2202.
```