

Lab Report: Bike-Sharing Data Analysis

MA 575 Fall 2021 - C3 Team #2

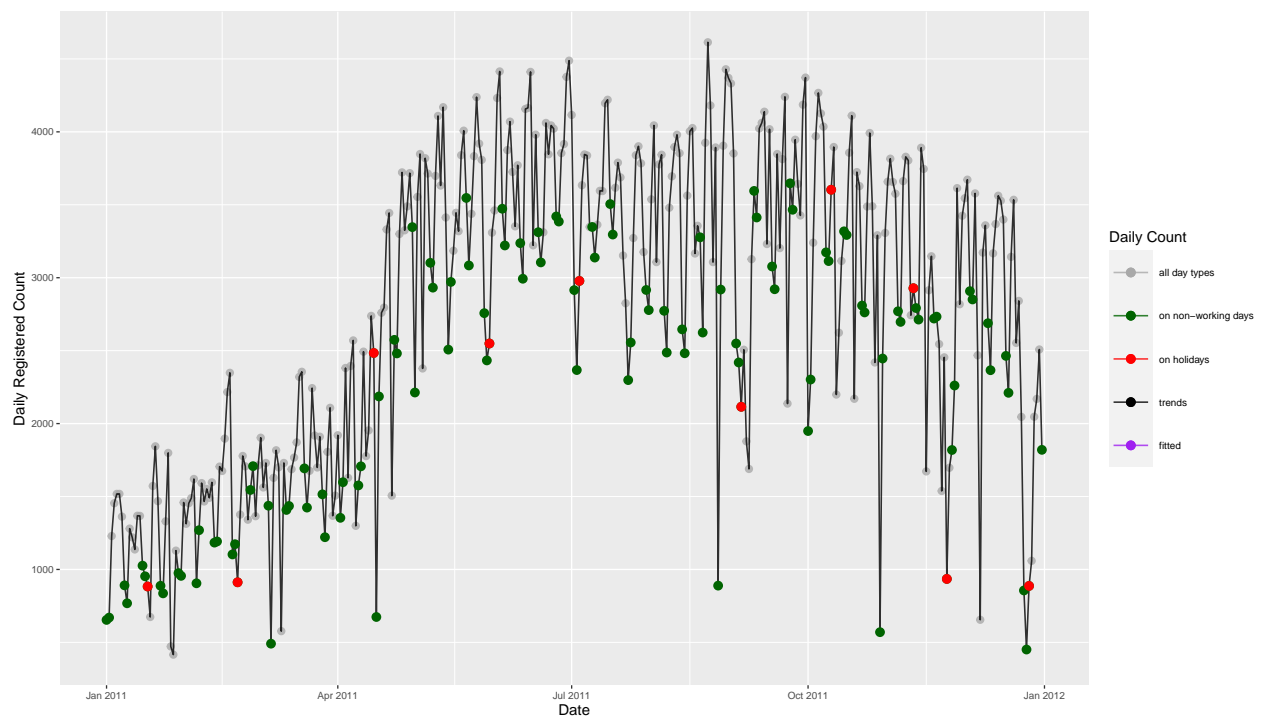
Ali Taqi, Hsin-Chang Lin, Huiru Yang, Ryan Mahoney, Yulin Li

12/01/2021

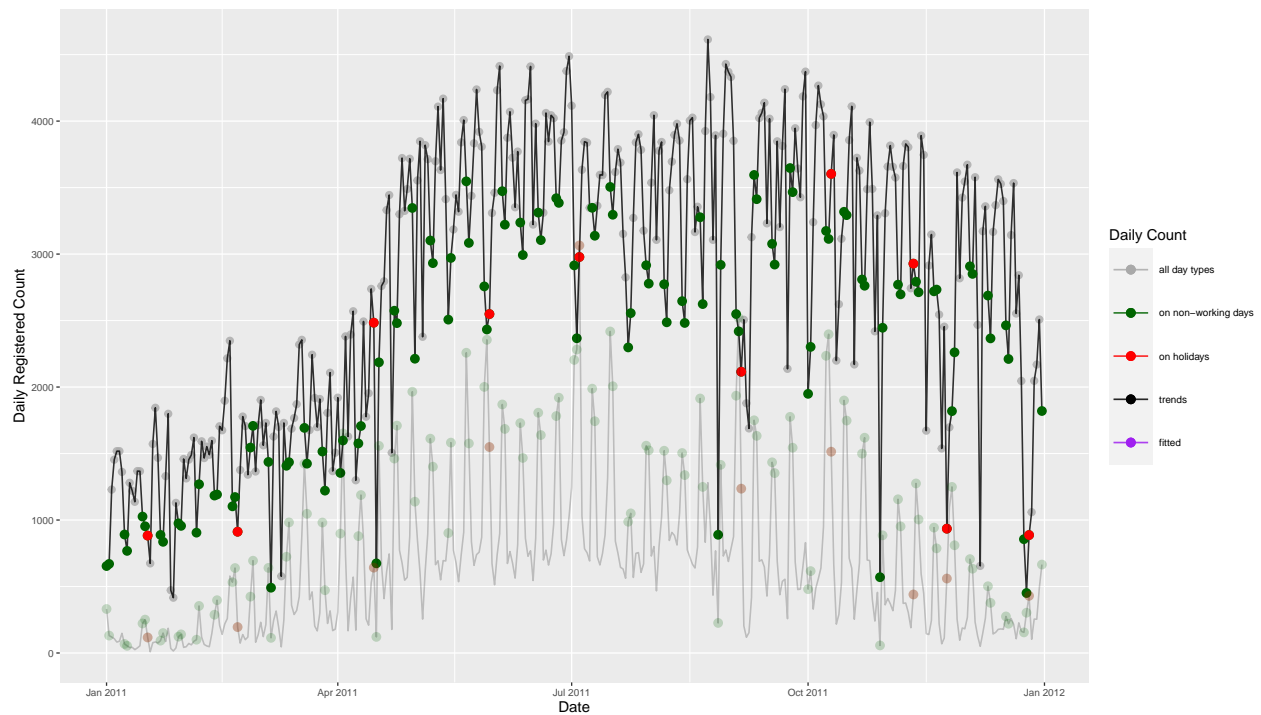
1 Data Visualization

1.1 Plots

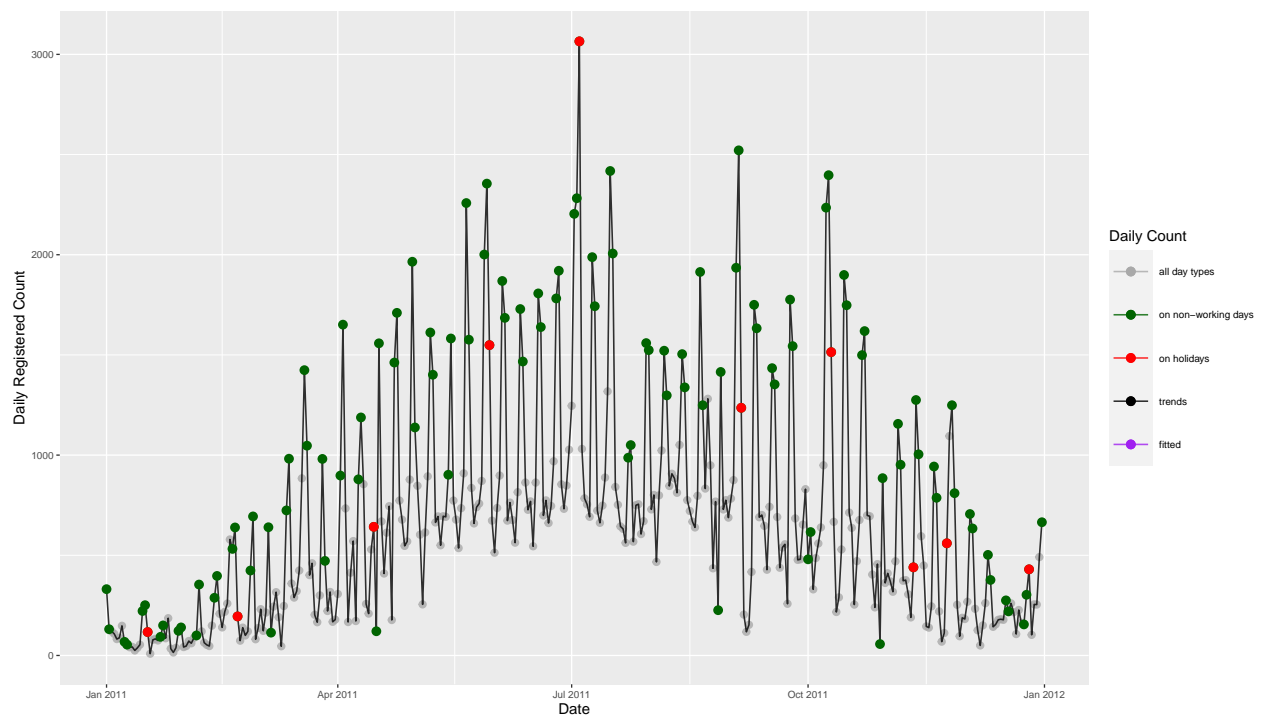
```
# Plot 1: 2011, Registered  
plot_usertype_cnt(data2011, user = "registered")
```



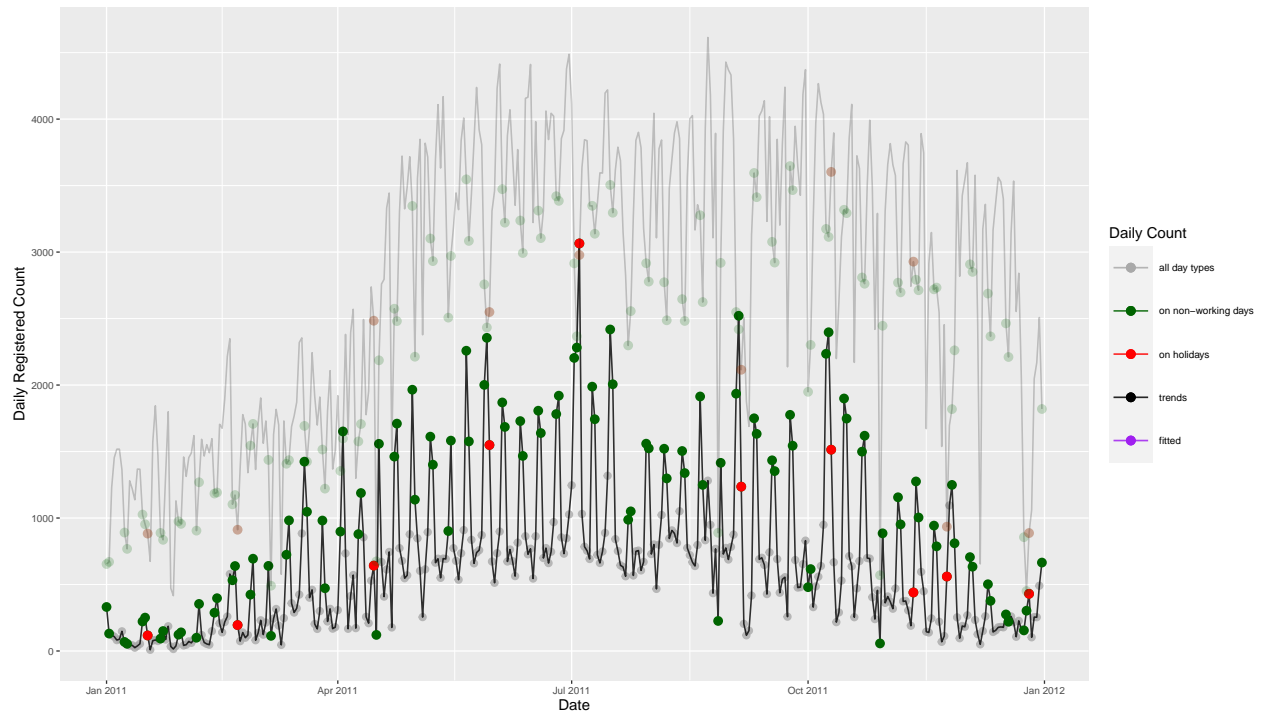
```
p1 <- plot_usertype_cnt(data2011, user = "registered", compare = T)  
p1
```



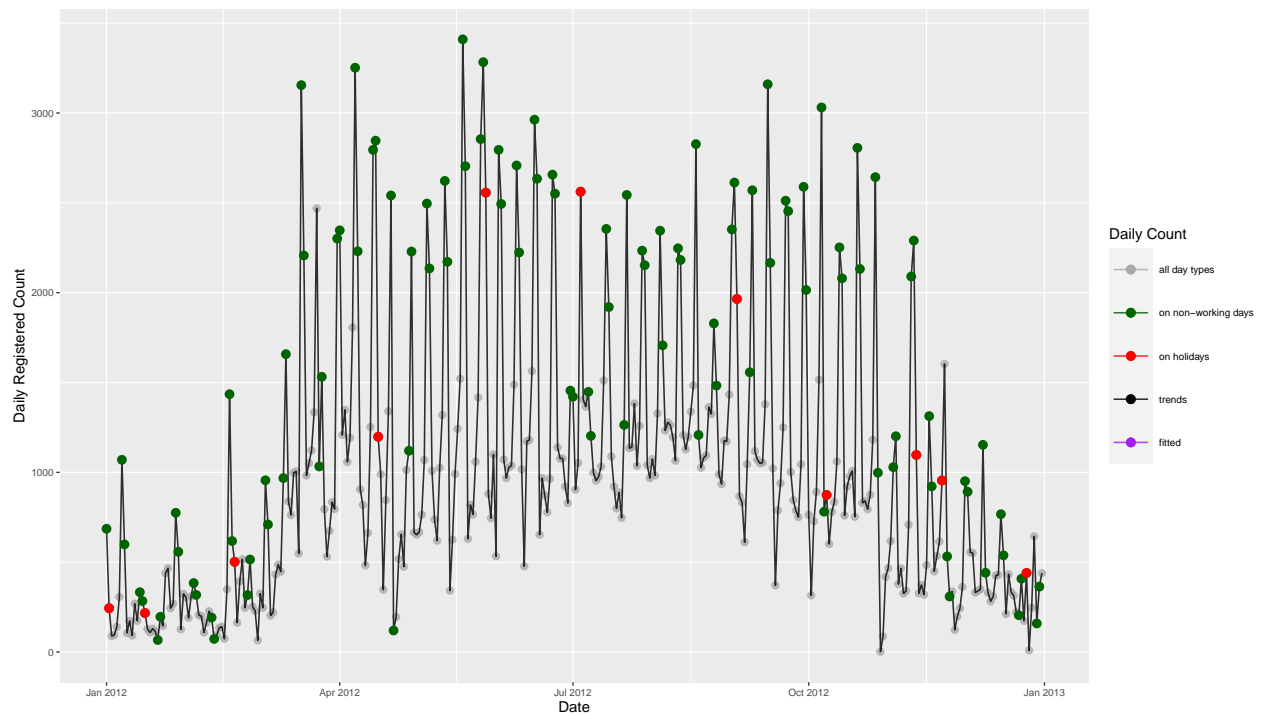
```
# Plot 2: 2011, Casual
plot_usertype_cnt(data2011, user = "casual")
```



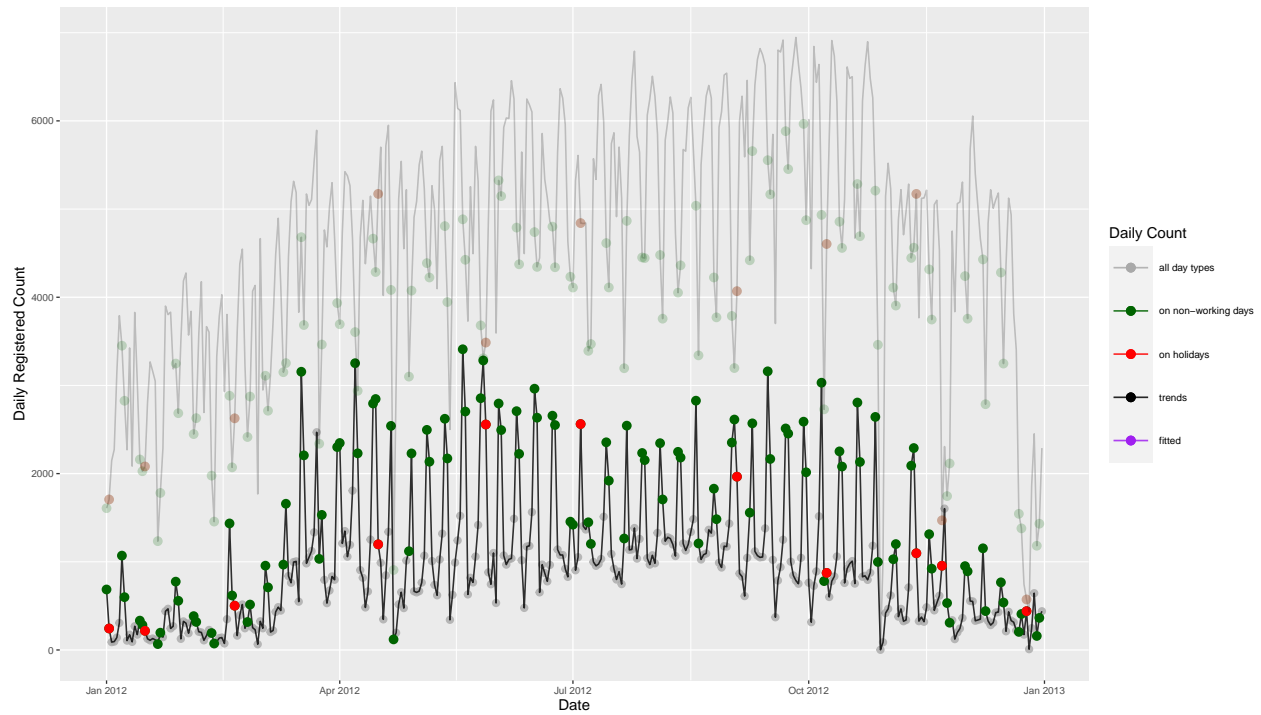
```
p2 <- plot_usertype_cnt(data2011, user = "casual", compare = T)
p2
```



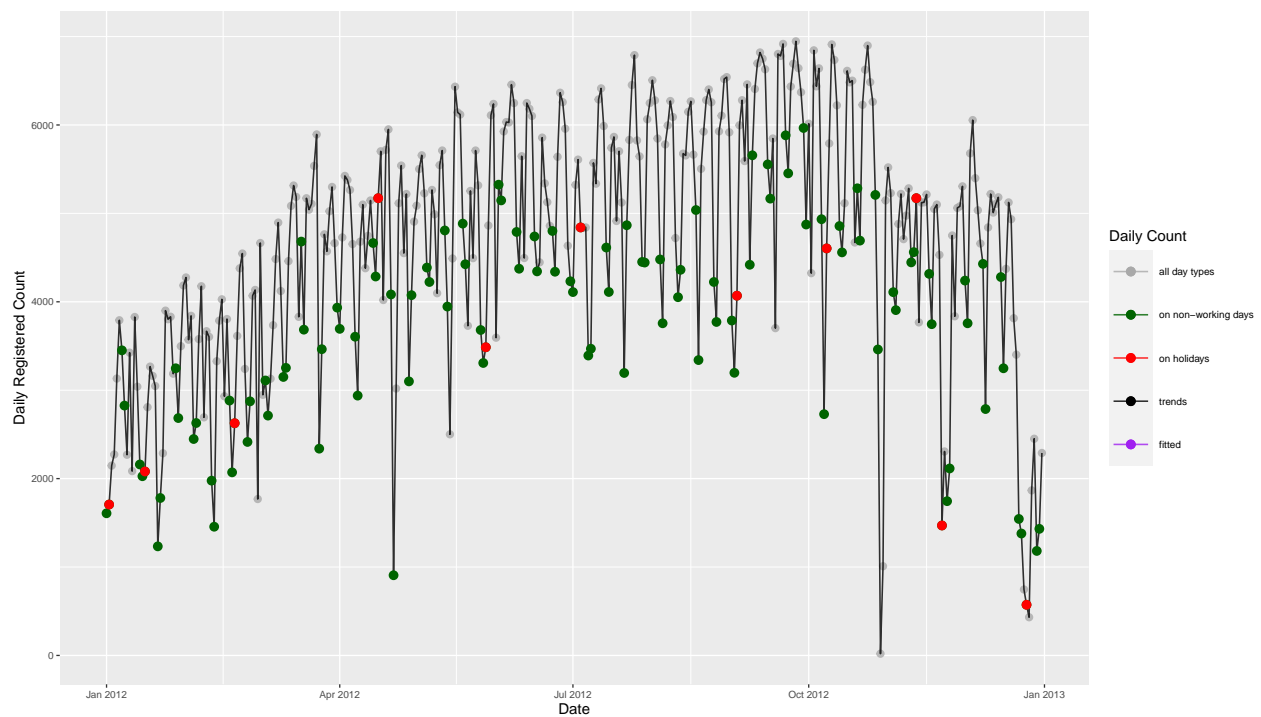
```
# Plot 3: 2012, Causal
plot_usertype_cnt(data2012, user = "casual")
```



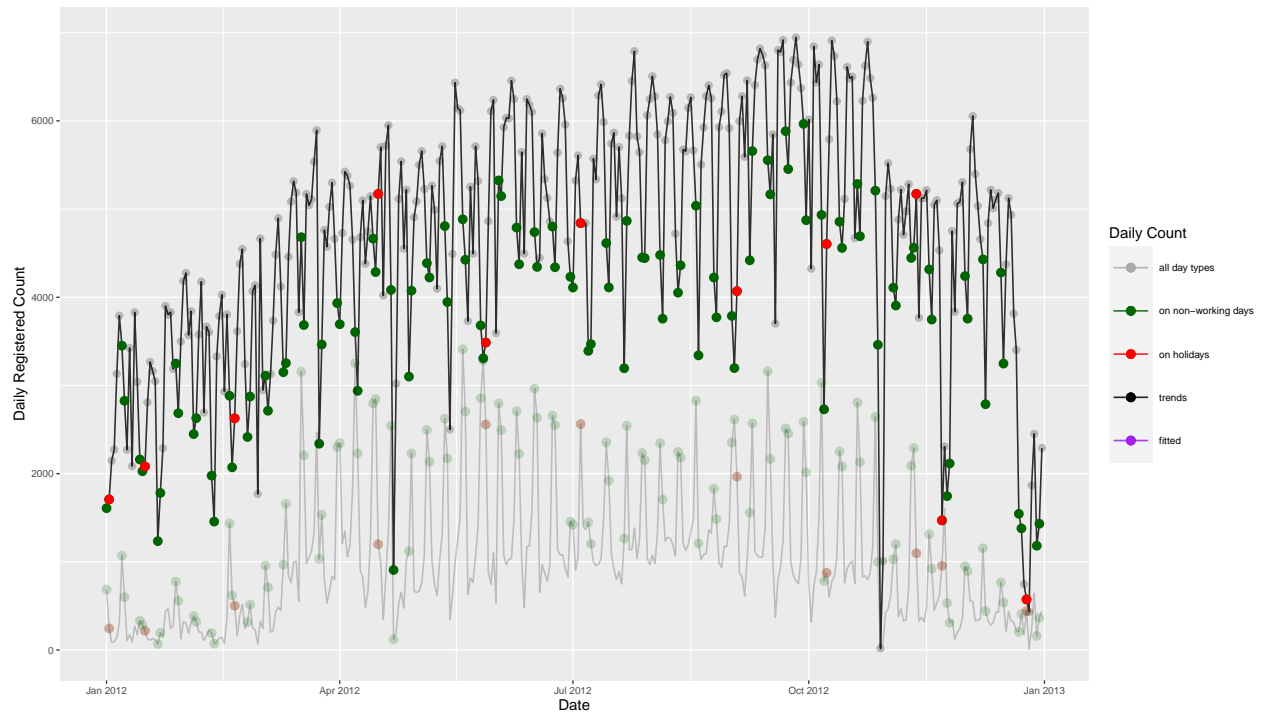
```
p3 <- plot_usertype_cnt(data2012, user = "casual", compare = T)
p3
```



```
# Plot 4: 2012, Registered
plot_usertype_cnt(data2012, user = "registered")
```



```
p4 <- plot_usertype_cnt(data2012, user = "registered", compare = T)
p4
```



- casual counts & registered counts: opposite behaviors - active on weekends vs active on working days.

2 Modeling

- for casual counts: adding temp^2 makes the prediction even more unstable.
- for registered counts: adding temp^2 seems necessary.

```
##
## Call:
## lm(formula = casual ~ workingday + weathersit + atemp + I(atemp^2),
##     data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1130.52  -156.99   -19.31   149.46  1683.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -217.537    114.280  -1.904   0.0578 .
## workingdayTRUE -653.904     34.363 -19.029 < 2e-16 ***
## weathersit2    -159.217     34.505  -4.614 5.50e-06 ***
## weathersit3    -454.457     82.029  -5.540 5.85e-08 ***
## atemp          90.513     10.649   8.499 5.17e-16 ***
## I(atemp^2)     -1.132      0.230  -4.922 1.31e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 302.5 on 359 degrees of freedom
## Multiple R-squared:  0.7083, Adjusted R-squared:  0.7043
## F-statistic: 174.4 on 5 and 359 DF, p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = registered ~ workingday + weathersit + atemp + I(atemp^2),
##     data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1818.76  -374.53    21.61   445.23  1466.03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1170.6015    222.6652  -5.257 2.52e-07 ***
## workingdayTRUE    715.4036     66.9538  10.685 < 2e-16 ***
## weathersit2     -396.3328     67.2308  -5.895 8.64e-09 ***
## weathersit3    -1466.5450    159.8279  -9.176 < 2e-16 ***
## atemp          251.6810     20.7493   12.130 < 2e-16 ***
## I(atemp^2)      -3.6869      0.4481  -8.228 3.55e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 589.4 on 359 degrees of freedom
## Multiple R-squared:  0.6951, Adjusted R-squared:  0.6909
## F-statistic: 163.7 on 5 and 359 DF, p-value: < 2.2e-16
##
## Call:
## lm(formula = casual ~ weathersit:workingday:atemp + season, data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1285.45  -135.27    -6.85   125.70  1483.21
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -194.208     49.339  -3.936 9.96e-05 ***
## season2        277.629     52.804   5.258 2.52e-07 ***
## season3        208.365     68.030   3.063 0.00236 **
## season4        181.677     45.648   3.980 8.36e-05 ***
## weathersit1:workingdayFALSE:atemp  53.629      3.107  17.262 < 2e-16 ***
## weathersit2:workingdayFALSE:atemp  47.117      3.561  13.232 < 2e-16 ***
## weathersit3:workingdayFALSE:atemp   2.733     11.843   0.231 0.81763
## weathersit1:workingdayTRUE:atemp    23.475      2.959   7.934 2.81e-14 ***
## weathersit2:workingdayTRUE:atemp    18.109      3.147   5.754 1.88e-08 ***
## weathersit3:workingdayTRUE:atemp     7.274      4.568   1.592 0.11221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 268 on 355 degrees of freedom
## Multiple R-squared:  0.7737, Adjusted R-squared:  0.7679
## F-statistic: 134.8 on 9 and 355 DF, p-value: < 2.2e-16
##
## Call:
## lm(formula = registered ~ weathersit:workingday:atemp + I(atemp^2) +
##     weathersit + season, data = data2011)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1903.56  -261.86    60.03   329.65   901.09
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -23.3105    198.3465  -0.118   0.9065
## I(atep^2)         -2.0322     0.4267  -4.763 2.79e-06 ***
## weathersit2       -359.9283    163.1466  -2.206   0.0280 *
## weathersit3       -915.1648    530.1124  -1.726   0.0852 .
## season2           731.3699     95.7513   7.638 2.10e-13 ***
## season3           936.4512    121.5230   7.706 1.34e-13 ***
## season4          1173.0231     86.2068  13.607 < 2e-16 ***
## weathersit1:workingdayFALSE:atep 128.5295     19.8658   6.470 3.29e-10 ***
## weathersit2:workingdayFALSE:atep 129.4648     20.4384   6.334 7.28e-10 ***
## weathersit3:workingdayFALSE:atep  77.5029     38.8128   1.997   0.0466 *
## weathersit1:workingdayTRUE:atep  158.3260     19.8731   7.967 2.28e-14 ***
## weathersit2:workingdayTRUE:atep  158.0100     20.3421   7.768 8.83e-14 ***
## weathersit3:workingdayTRUE:atep  125.0327     28.3313   4.413 1.36e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 476.4 on 352 degrees of freedom
## Multiple R-squared:  0.8047, Adjusted R-squared:  0.7981
## F-statistic: 120.9 on 12 and 352 DF,  p-value: < 2.2e-16
##
## Call:
## lm(formula = casual ~ workingday:atep + weathersit:season, data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1370.26  -137.01    -8.87   121.09  1391.97
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -372.763    115.352  -3.232 0.001348 **
## workingdayFALSE:atep    51.417     3.175  16.195 < 2e-16 ***
## workingdayTRUE:atep     22.041     3.054   7.217 3.30e-12 ***
## weathersit1:season1    213.426    104.653   2.039 0.042161 *
## weathersit2:season1     93.925    108.717   0.864 0.388212
## weathersit3:season1     88.066    213.461   0.413 0.680180
## weathersit1:season2    516.736    103.603   4.988 9.63e-07 ***
## weathersit2:season2    338.228    105.005   3.221 0.001397 **
## weathersit3:season2   -166.098    213.288  -0.779 0.436652
## weathersit1:season3    417.493    107.398   3.887 0.000121 ***
## weathersit2:season3    335.121    113.170   2.961 0.003273 **
## weathersit3:season3    -62.089    182.740  -0.340 0.734236
## weathersit1:season4    429.955    102.565   4.192 3.50e-05 ***
## weathersit2:season4    206.403    107.434   1.921 0.055516 .
## weathersit3:season4         NA         NA         NA         NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 268.8 on 351 degrees of freedom
## Multiple R-squared:  0.7749, Adjusted R-squared:  0.7666
## F-statistic: 92.94 on 13 and 351 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = registered ~ workingday:atemp + I(atemp^2) + weathersit +
##     season, data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1906.63  -271.50    53.88   331.41   904.06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      19.0295    188.4535   0.101   0.92
## I(atemp^2)       -1.9535     0.4207  -4.644 4.82e-06 ***
## weathersit2      -357.9474     54.3461  -6.586 1.62e-10 ***
## weathersit3     -1651.5619    129.8118 -12.723 < 2e-16 ***
## season2          734.1729     94.8515   7.740 1.04e-13 ***
## season3          932.8039    121.0030   7.709 1.28e-13 ***
## season4         1175.3168     85.6419  13.724 < 2e-16 ***
## workingdayFALSE:atemp 124.8334     19.4143   6.430 4.11e-10 ***
## workingdayTRUE:atemp  154.2929     19.4210   7.945 2.59e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Residual standard error: 475.2 on 356 degrees of freedom
## Multiple R-squared:  0.8035, Adjusted R-squared:  0.799
## F-statistic: 181.9 on 8 and 356 DF,  p-value: < 2.2e-16

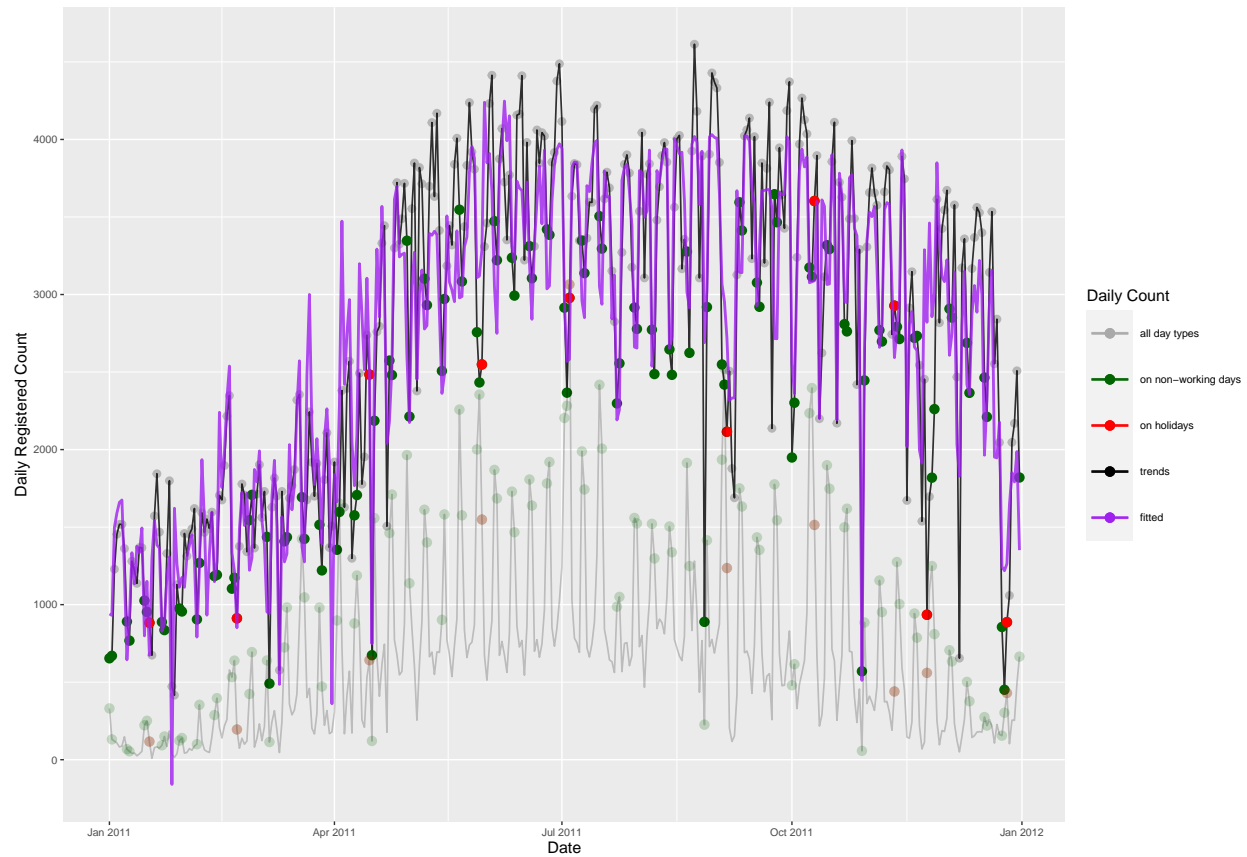
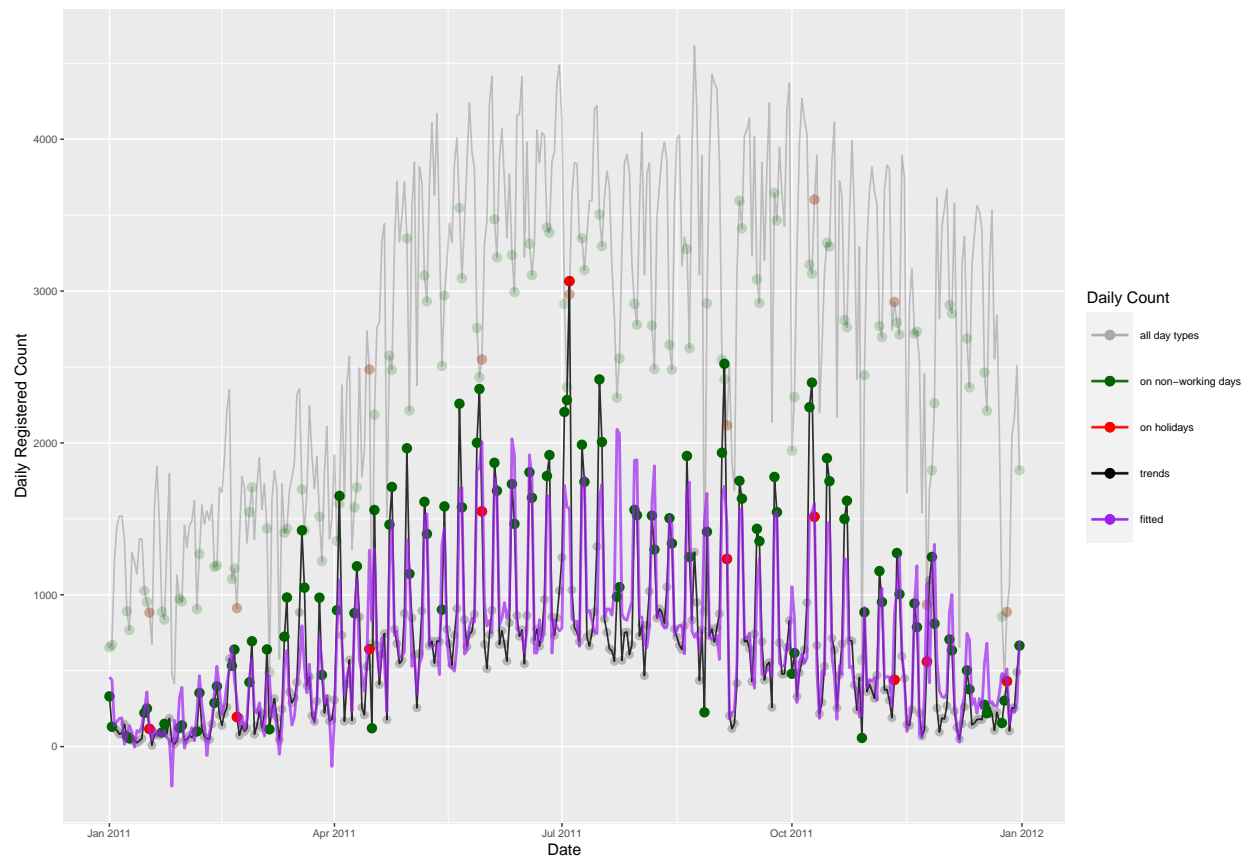
##
## Call:
## lm(formula = casual ~ season:workingday:atemp + weathersit, data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1263.04  -109.09   -25.13   120.42  1495.46
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -78.979     61.405  -1.286   0.199
## weathersit2     -145.169     29.754  -4.879 1.61e-06 ***
## weathersit3    -415.269     71.725  -5.790 1.56e-08 ***
## season1:workingdayFALSE:atemp  37.492      4.999   7.500 5.19e-13 ***
## season2:workingdayFALSE:atemp  62.133      2.949  21.066 < 2e-16 ***
## season3:workingdayFALSE:atemp  53.912      2.358  22.866 < 2e-16 ***
## season4:workingdayFALSE:atemp  61.919      3.770  16.424 < 2e-16 ***
## season1:workingdayTRUE:atemp   22.999      4.617   4.981 9.91e-07 ***
## season2:workingdayTRUE:atemp   28.246      2.536  11.139 < 2e-16 ***
## season3:workingdayTRUE:atemp   26.377      2.068  12.753 < 2e-16 ***
## season4:workingdayTRUE:atemp   26.244      3.077   8.528 4.38e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

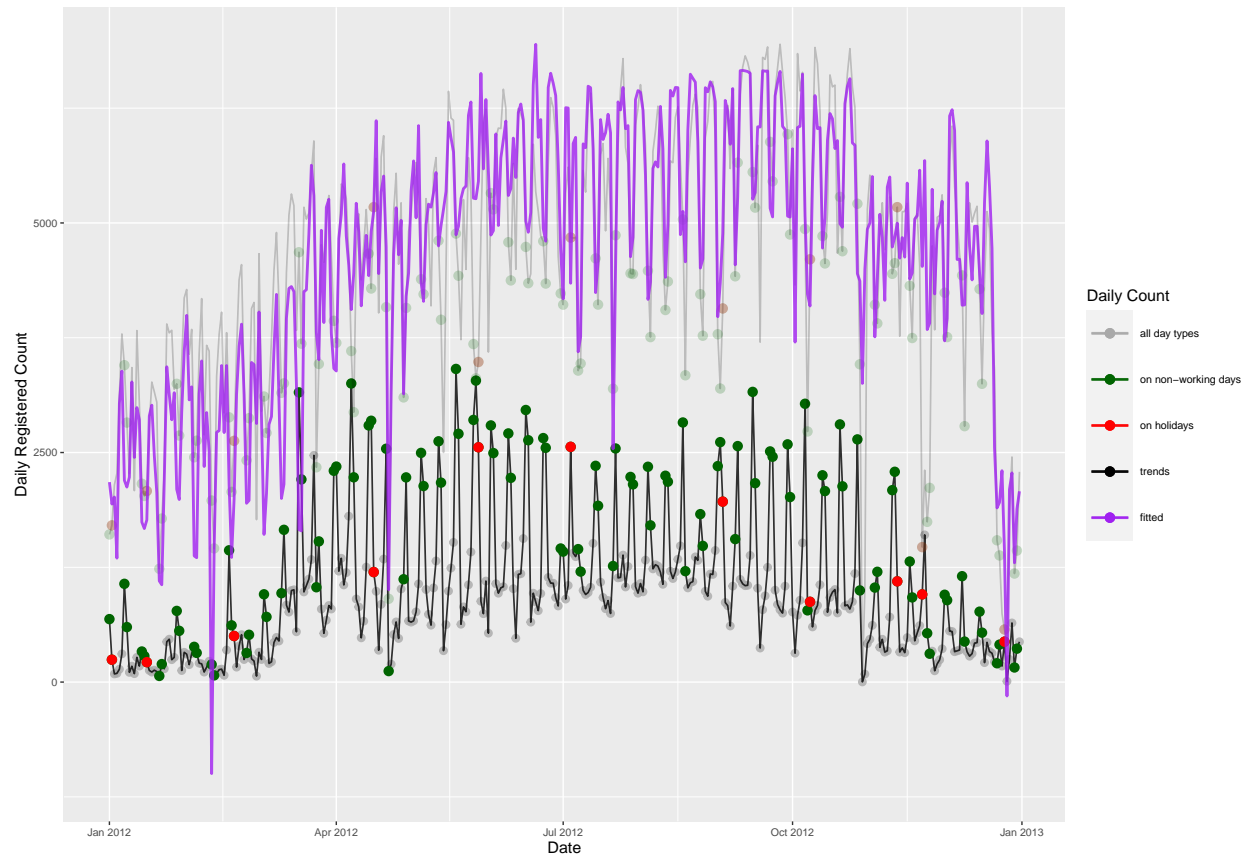
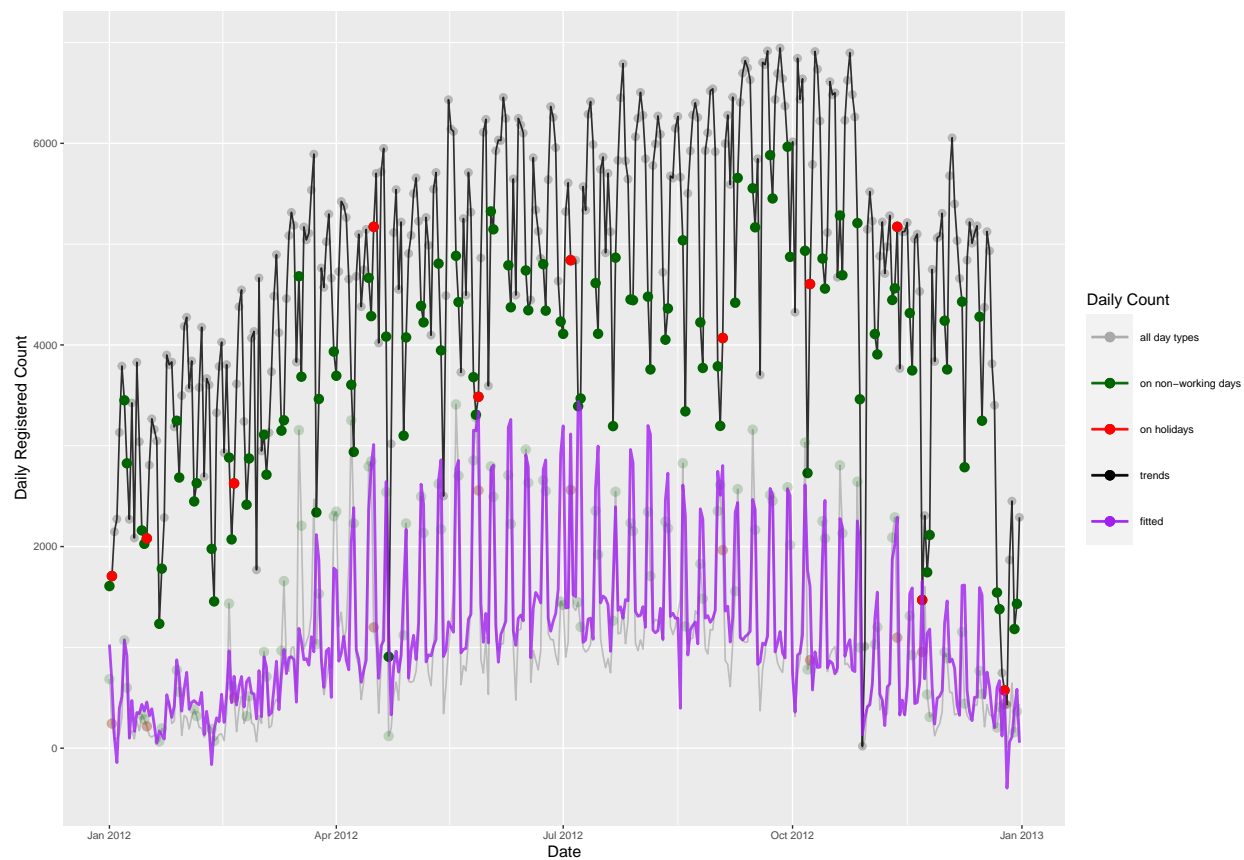


```
##
## Residual standard error: 262.5 on 354 degrees of freedom
## Multiple R-squared:  0.7834, Adjusted R-squared:  0.7773
## F-statistic: 128 on 10 and 354 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = registered ~ season:workingday:atemp + season:I(atemp^2) +
##     weathersit, data = data2011)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1886.05  -219.71    39.92   286.96  1144.98
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      610.6409    282.3255   2.163 0.031226 *
## weathersit2       -373.3875     50.5207  -7.391 1.08e-12 ***
## weathersit3      -1716.0539    120.0303 -14.297 < 2e-16 ***
## season1:I(atemp^2)      -1.1444     1.4073  -0.813 0.416634
## season2:I(atemp^2)      -0.5901     0.5757  -1.025 0.306088
## season3:I(atemp^2)      -4.4867     0.4874  -9.205 < 2e-16 ***
## season4:I(atemp^2)      -3.2135     0.8256  -3.892 0.000119 ***
## season1:workingdayFALSE:atemp    59.4714    40.5300   1.467 0.143181
## season2:workingdayFALSE:atemp    99.6691    25.5654   3.899 0.000116 ***
## season3:workingdayFALSE:atemp   219.7328    22.2470   9.877 < 2e-16 ***
## season4:workingdayFALSE:atemp   178.5360    30.2887   5.894 8.85e-09 ***
## season1:workingdayTRUE:atemp    104.6024    41.0241   2.550 0.011205 *
## season2:workingdayTRUE:atemp    121.4648    25.6036   4.744 3.06e-06 ***
## season3:workingdayTRUE:atemp    248.4427    22.1557  11.213 < 2e-16 ***
## season4:workingdayTRUE:atemp    210.9254    30.1975   6.985 1.44e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 434.5 on 350 degrees of freedom
## Multiple R-squared:  0.8385, Adjusted R-squared:  0.832
## F-statistic: 129.8 on 14 and 350 DF,  p-value: < 2.2e-16
```

2.0.0.1 Model 3





2.0.0.2 Model 2: current best



