Lab Report: Bike-Sharing Data Analysis
MA 575 Fall 2021 - C3 Team #2
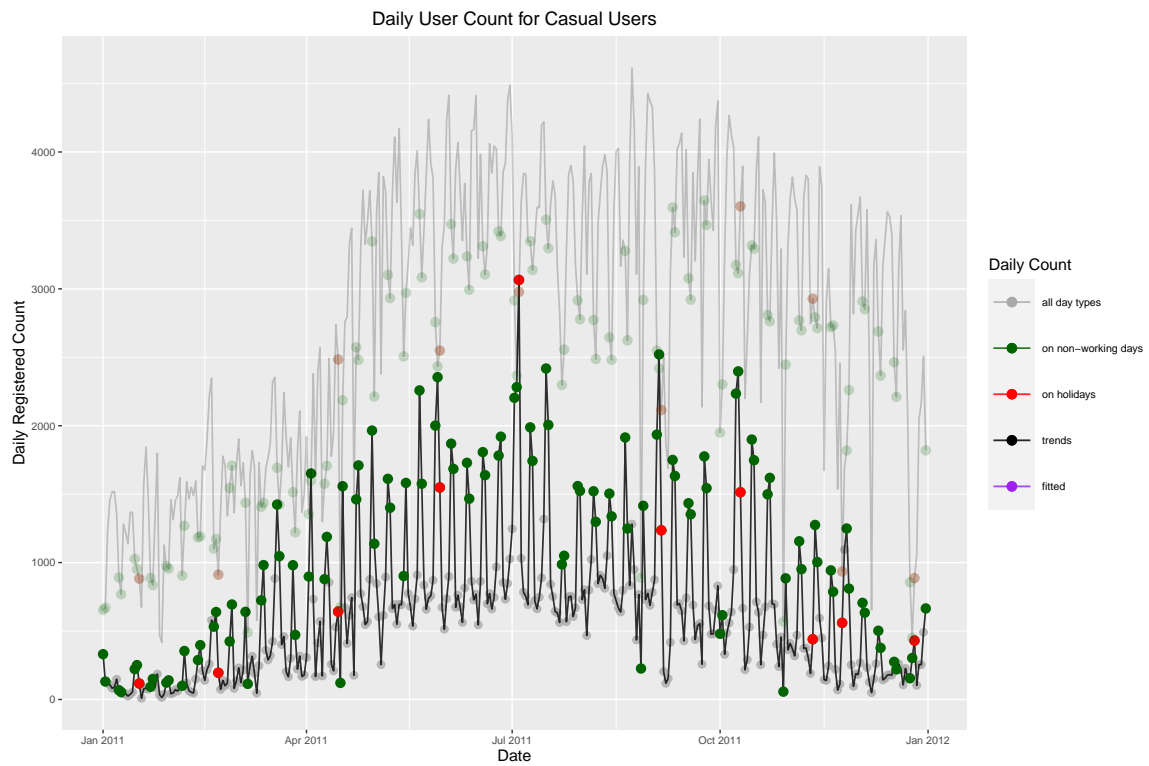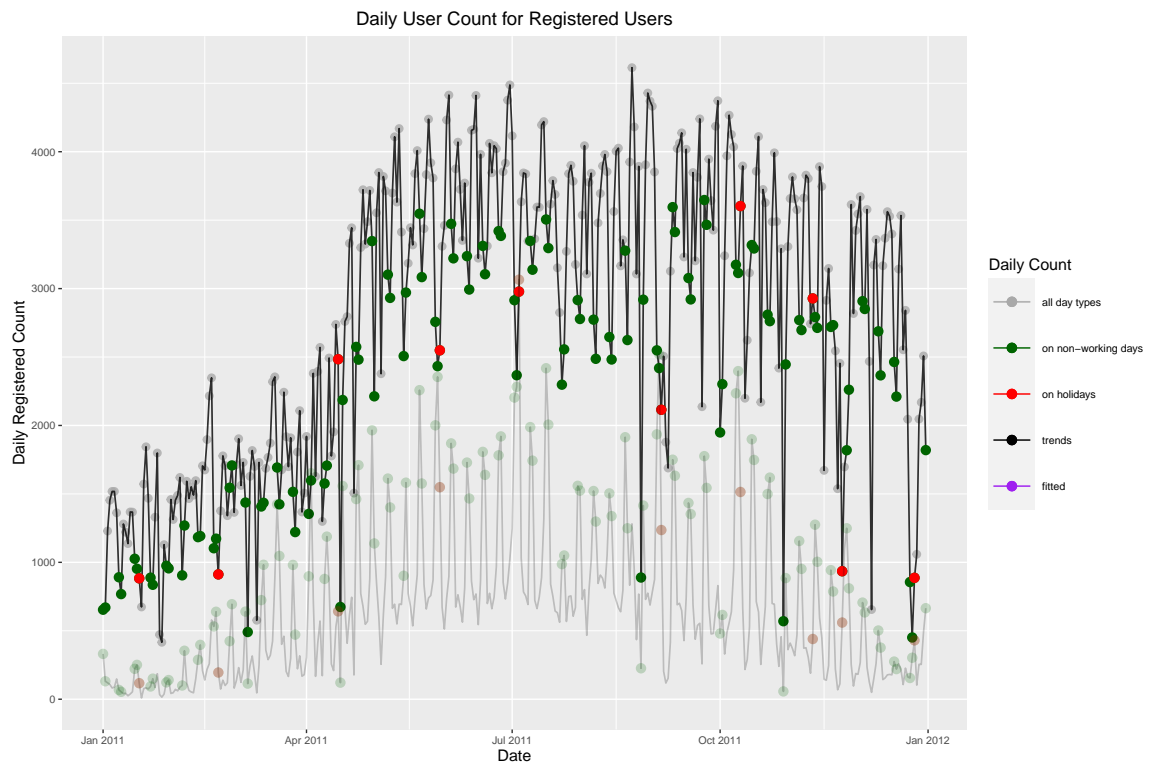
Ali Taqi, Hsin-Chang Lin, Huiru Yang, Ryan Mahoney, Yulin Li

12/01/2021
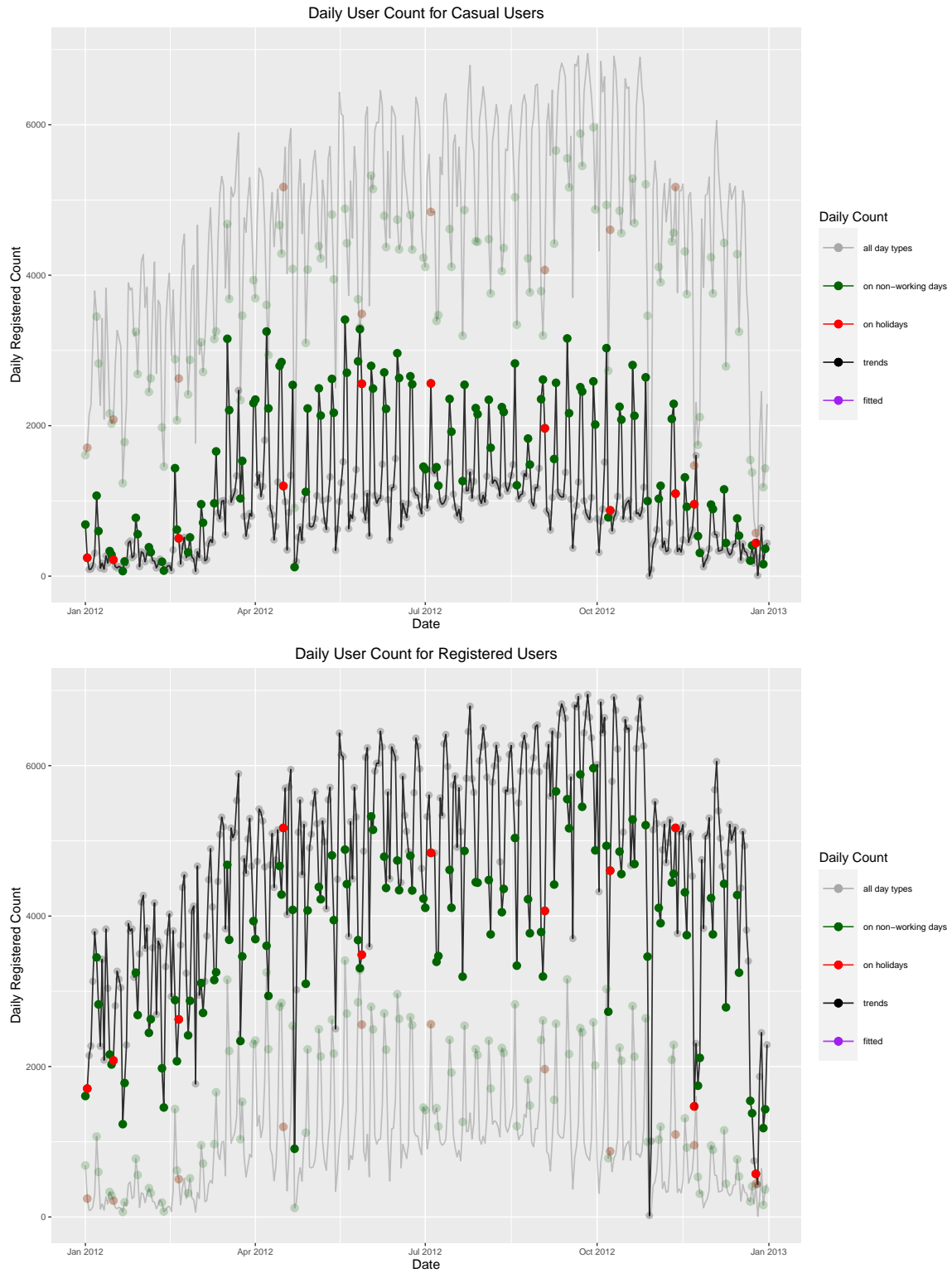
1 Data Visualization

1.1 Plots

1.1.1 2011 Data

Daily User Count for Registered Users

Daily User Count for Casual Users

Daily User Count for Casual Users

Daily User Count for Registered Users

- 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.

2.1 Casual Users

```
mod.cas.3 <- lm(casual ~ season:workingday:atemp + weathersit, data = data2011)
```

```
## # A tibble: 11 x 7
##    term                  estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                    <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept                -79.0      61.4     -1.29   0.199    -200.     41.8
##  2 weathersit: 2           -145.       29.8     -4.88   0        -204.    -86.7
##  3 weathersit: 3           -415.       71.7     -5.79   0        -556.   -274.
##  4 season: 1:workingdayF~    37.5       5.00      7.5    0         27.7     47.3
##  5 season: 2:workingdayF~    62.1       2.95     21.1    0         56.3     67.9
##  6 season: 3:workingdayF~    53.9       2.36     22.9    0         49.3     58.5
##  7 season: 4:workingdayF~    61.9       3.77     16.4    0         54.5     69.3
##  8 season: 1:workingdayT~    23.0       4.62      4.98   0         13.9     32.1
##  9 season: 2:workingdayT~    28.2       2.54     11.1    0         23.3     33.2
## 10 season: 3:workingdayT~    26.4       2.07     12.8    0         22.3     30.4
## 11 season: 4:workingdayT~    26.2       3.08      8.53   0         20.2     32.3

## # A tibble: 1 x 9
##   r_squared adj_r_squared    mse  rmse sigma statistic p_value    df  nobs
##       <dbl>         <dbl>  <dbl> <dbl> <dbl>     <dbl>   <dbl> <dbl> <dbl>
## 1     0.783         0.777 66838.  259.  263.      128.       0    10   365
```

2.2 Registered Users

```
mod.reg.3 <- lm(registered ~ season:workingday:atemp + season:I(atemp^2) + weathersit, data = data2011)
```

```
## # A tibble: 15 x 7
##    term               estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                 <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept             611.      282.      2.16   0.031    55.4   1.17e+3
##  2 weathersit: 2        -373.       50.5    -7.39   0        -473.  -2.74e+2
##  3 weathersit: 3       -1716.      120.    -14.3    0       -1952.  -1.48e+3
##  4 season: 1:I(atemp^2)   -1.14      1.41    -0.813  0.417    -3.91  1.62e+0
##  5 season: 2:I(atemp^2)   -0.59      0.576   -1.02   0.306    -1.72  5.42e-1
##  6 season: 3:I(atemp^2)   -4.49      0.487   -9.20   0        -5.44 -3.53e+0
##  7 season: 4:I(atemp^2)   -3.21      0.826   -3.89   0        -4.84 -1.59e+0
##  8 season: 1:workingdayF~  59.5     40.5     1.47   0.143    -20.2  1.39e+2
##  9 season: 2:workingdayF~  99.7     25.6     3.90   0         49.4  1.50e+2
## 10 season: 3:workingdayF~ 220.      22.2     9.88   0        176.   2.63e+2
## 11 season: 4:workingdayF~ 179.      30.3     5.89   0        119.   2.38e+2
## 12 season: 1:workingdayT~ 105.      41.0     2.55   0.011     23.9  1.85e+2
## 13 season: 2:workingdayT~ 121.      25.6     4.74   0         71.1  1.72e+2
## 14 season: 3:workingdayT~ 248.      22.2    11.2    0        205.   2.92e+2
## 15 season: 4:workingdayT~ 211.      30.2     6.98   0        152.   2.70e+2
```

```
## # A tibble: 1 x 9
##   r_squared adj_r_squared     mse  rmse sigma statistic p_value    df  nobs
##       <dbl>         <dbl>   <dbl> <dbl> <dbl>     <dbl>   <dbl> <dbl> <dbl>
## 1     0.838         0.832 181008.  425.  434.      130.       0    14   365
```
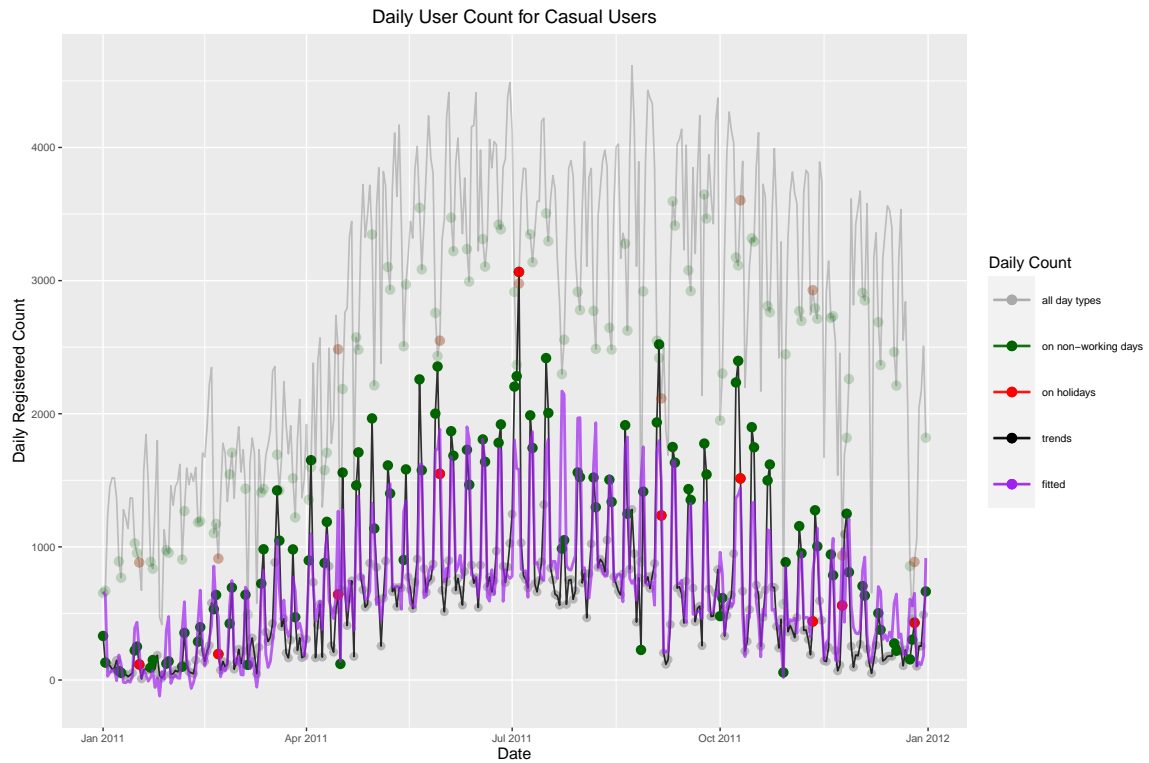
3 Diagnostics: 2011 Data

3.1 Our Wrapper Functions

```r
plot_fitted_cnt_2011 <- function(data, user, model, compare = FALSE){
  plot_usertype_cnt(data, user, compare) +
    geom_line(mapping = aes(x = dteday, y = predict(model, data2011), color = 'fitted'),
              size = 0.5, alpha = 0.7)
}
```
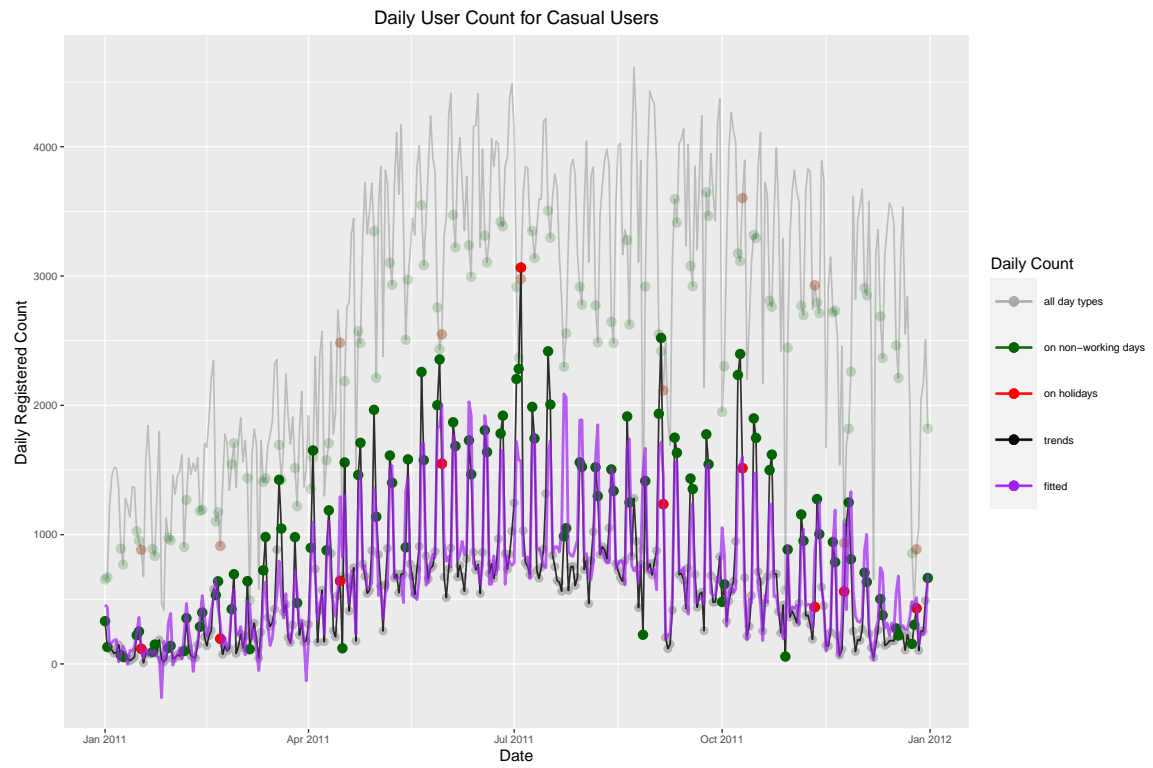
3.2 Casual Users

3.2.1 Model 2

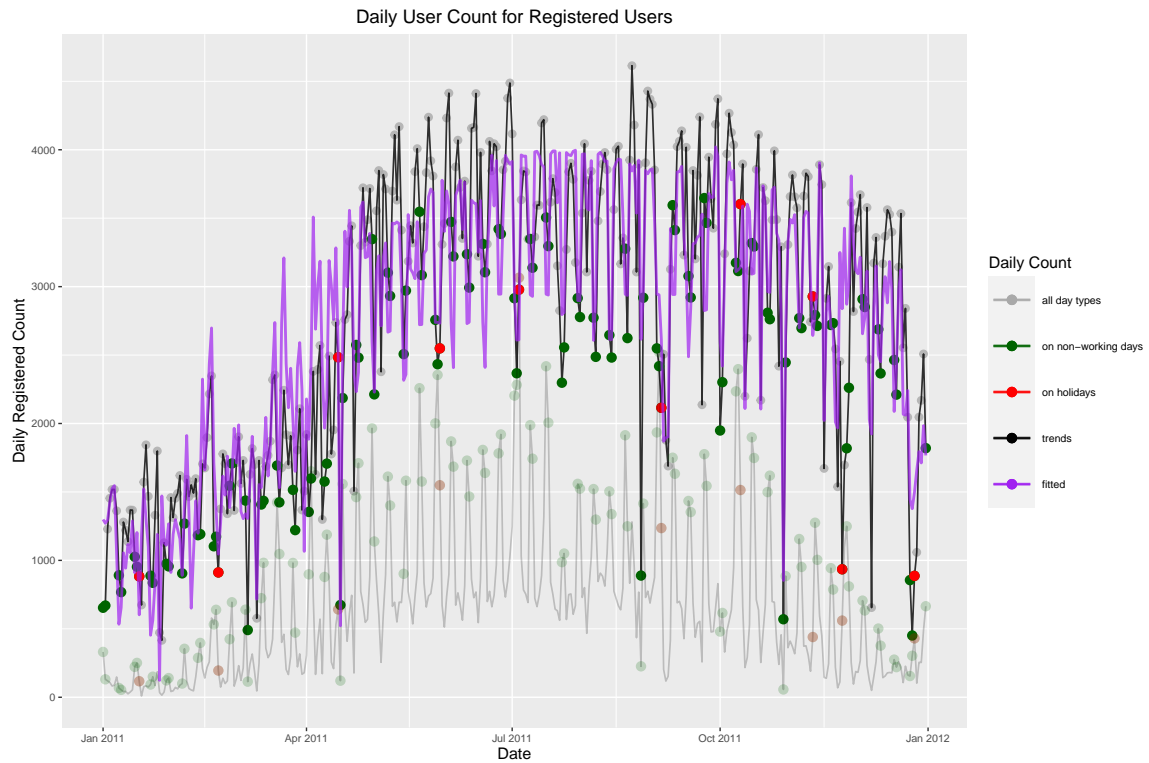Daily User Count for Casual Users

3.2.2 Model 3

```
plot_fitted_cnt_2011(data = data2011, user = "casual", model = mod.cas.3, compare = TRUE)
```
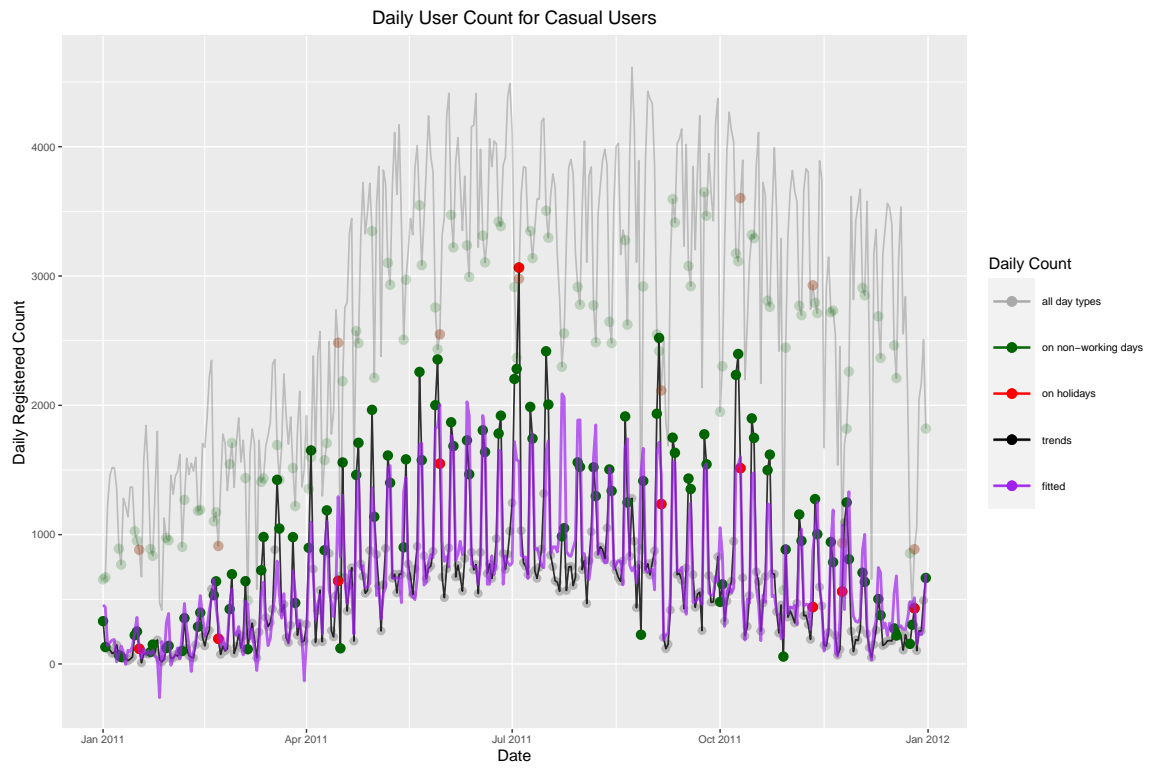
Daily User Count for Casual Users

3.3 Registered Users

3.3.1 Model 2

Daily User Count for Registered Users

3.3.2 Model 3

```
plot_fitted_cnt_2011(data = data2011, user = "casual", model = mod.cas.3, compare = TRUE)
```

Daily User Count for Casual Users

3.3.3 Model 4

```
#plot_fitted_cnt_2011(data = data2011, user = "casual", model = mod.cas.4, compare = TRUE)
```
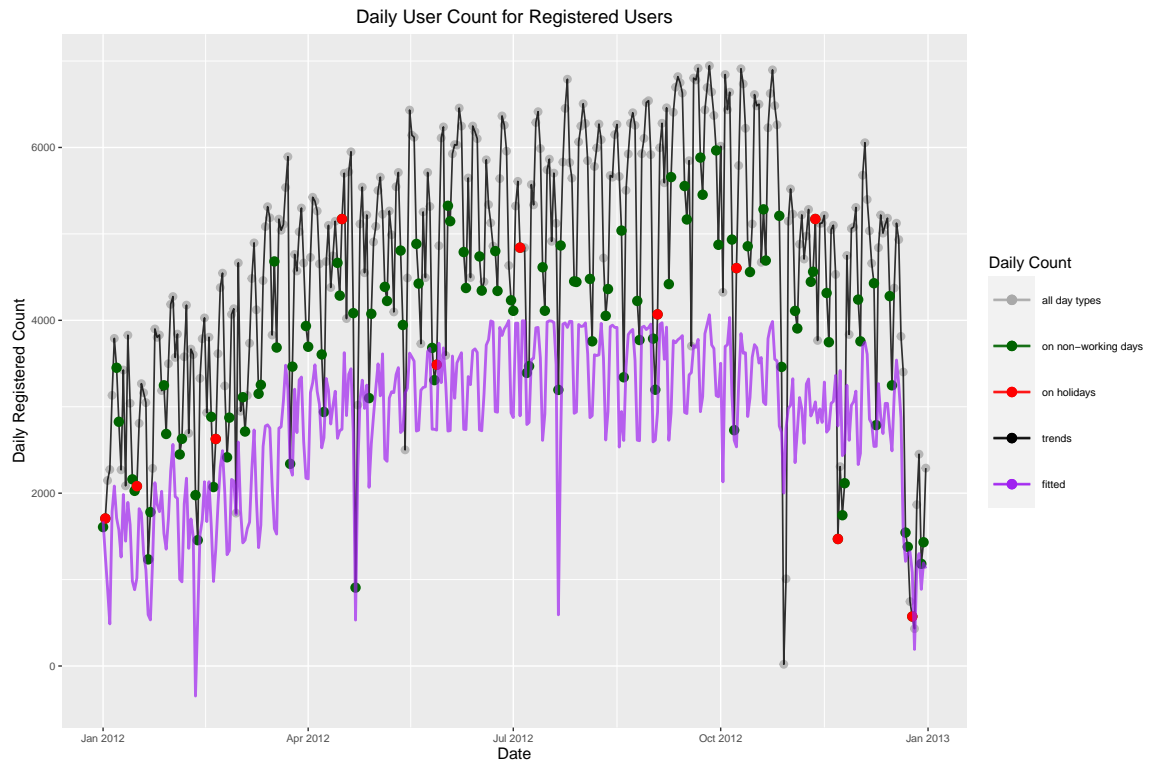
4 Diagnostics: 2012 Data

4.1 Our Wrapper Functions

```
plot_fitted_cnt_2012 <- function(data, user, model, compare = FALSE){
  plot_usertype_cnt(data, user, compare) +
    geom_line(data = data2012, mapping = aes(x = dteday, y = predict(model, data2012), color = 'fitted')
              size = 0.5, alpha = 0.7)
}
```
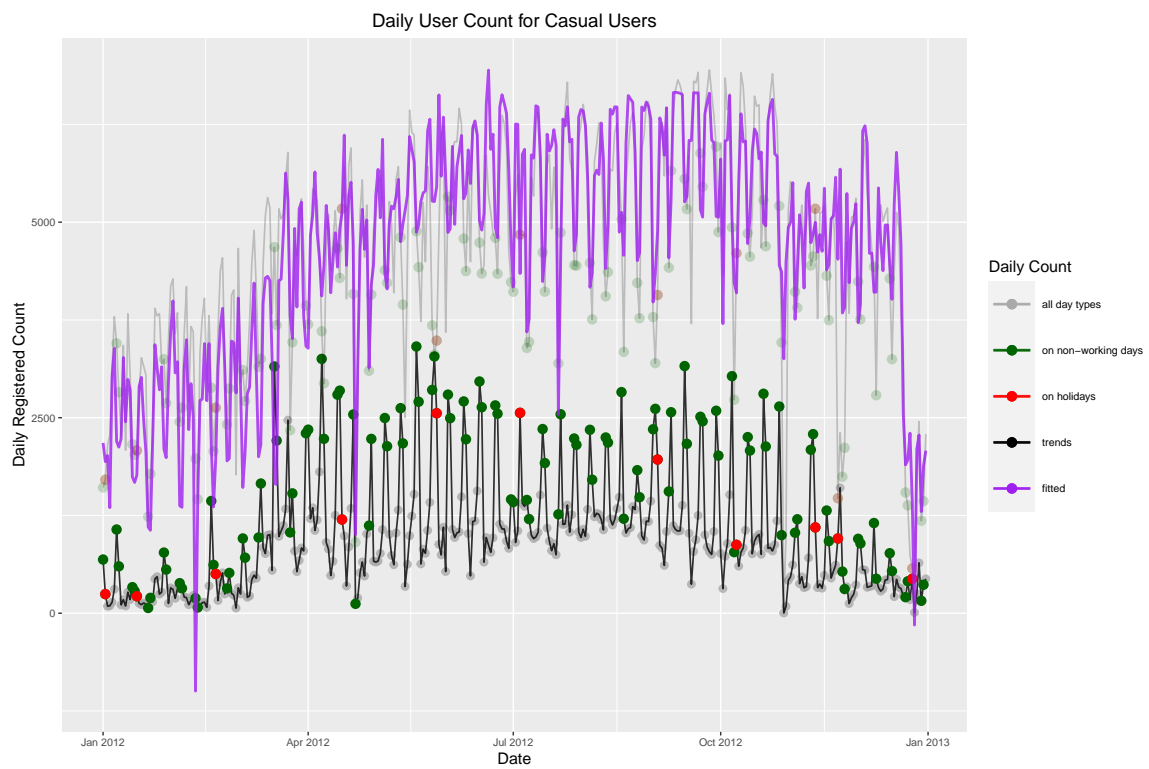
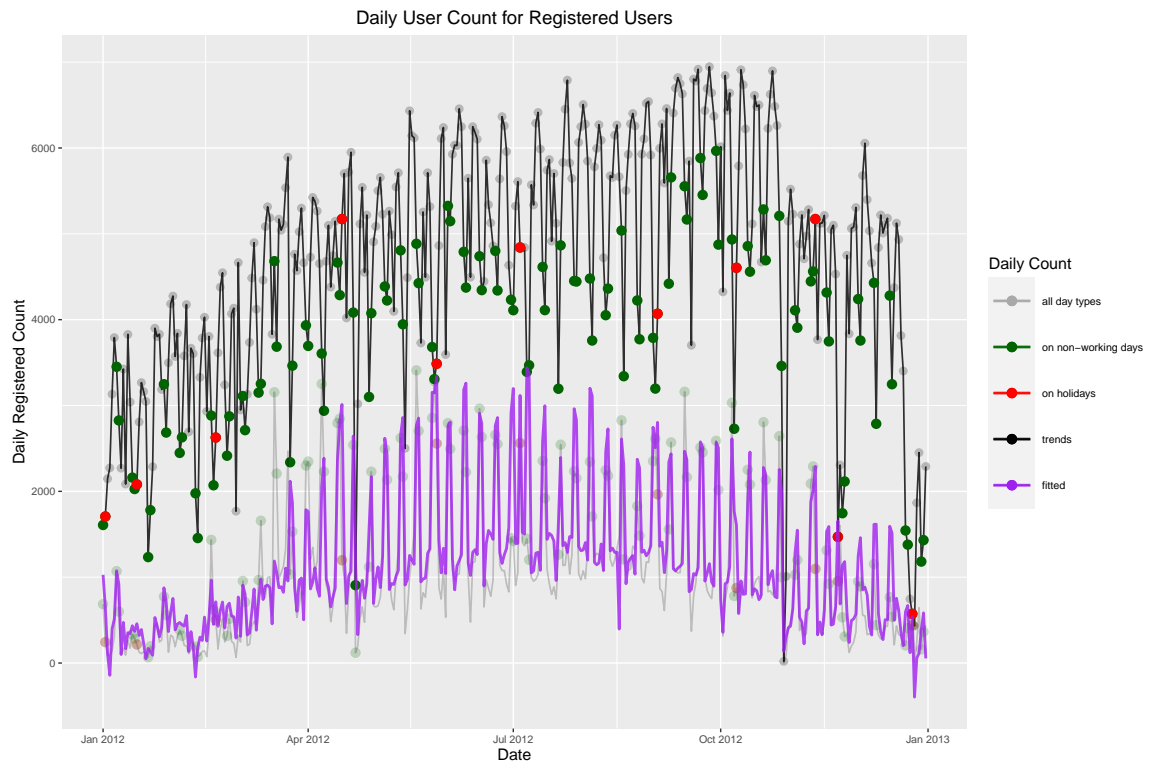## 4.2 Registered Users

4.2.1 Model 2

Daily User Count for Registered Users

4.2.2 Model 3

4.3 Casual Users

4.3.1 Model 2

4.3.2 Model 3

Daily User Count for Registered Users

Daily User Count for Casual Users

4.3.2.1 Model 2: current best

Daily User Count for Registered Users

Daily User Count for Casual Users