basic_regression2

Don Li

03/06/2020

Libraries

```
library( data.table )
load( "regression1.RData" )
do_rmse = function( yhat, test_set ){
    rmse = sqrt( mean( (yhat - test_set$timediff)^2 ) )
    rmse
}
```

Join data

E has computed distances for each trip. These distances were computed by summing over the distance between points for each journey. The Haversine distance was used. Documentation/code is in the `real-distances.py`. The units are in km.

We will join this to the existing summary dataset.

```
new_dist = data.table( read.csv("../distances_km.csv" ) )
new_dist[ , trj_id := as.character(ID) ]
summary_data[ new_dist, new_dist := i.km, on = "trj_id" ]
```

Updating the model

Previously, I used speed information. But, in our model inputs, we won't have speed, since we are predicting with the origin, the destination, and the time. I will remove the speed variables for the regression model.

```
# Standard remove trip ID and month
summary_data[ , trj_id := NULL ]
summary_data[ , month_ := NULL ]
# Speed removed
summary_data[ , speed_avg := NULL ]
summary_data[ , speed_var := NULL ]
```

Add some other variables

```
summary_data[ , is_weekend := weekday_ %in% c("Sat", "Sun") ]
summary_data[ , is_rushhour_morning := hour_ > 7 & hour_ < 10 ]
summary_data[ , is_rushhour_night := hour_ > 17 & hour_ < 20 ]
```

```
set.seed(1)
n = nrow( summary_data )
training_set_id = sample( 1:n, n * 0.75 )
```

```
training_set = summary_data[ training_set_id ]
test_set = summary_data[ -training_set_id ]
```

Fit the models

Without speed, our basic model isn't too good. Adjusted R^2 is about 25%. With a proper distance, the regression model is a lot better. Adjusted R^2 is 55%. We can also add both distance measures and see if that improves anything. Summary table below:

```
model1_vars = setdiff( names(training_set), "new_dist" )
model1 = lm( timediff ~ .*., training_set[,mget(model1_vars)] )
model1_summary = summary(model1)
model1_rmse = do_rmse( predict( model1, test_set ), test_set )

model2_vars = setdiff( names(training_set), "dist_" )
model2 = lm( timediff ~ .*., training_set[,mget(model2_vars)] )
model2_summary = summary(model2)
model2_rmse = do_rmse( predict( model2, test_set ), test_set )

model3 = lm( timediff ~ .*., training_set )
model3_summary = summary(model3)
model3_rmse = do_rmse( predict( model3, test_set ), test_set )
```

```
rmse_list = round( c( model1_rmse, model2_rmse, model3_rmse ), 3 )
adj_r2_list = round( c( model1_summary$adj.r.squared, model2_summary$adj.r.squared,
    model3_summary$adj.r.squared ), 3 )
model_names = c("Total dist", "New dist", "Both dist")
data.table( model_names, rmse_list, adj_r2_list )
```

```
##    model_names rmse_list adj_r2_list
## 1:  Total dist   277.073       0.254
## 2:    New dist   220.360       0.553
## 3:   Both dist   198.414       0.665
```

Further questions:

What if we also add the L1 distance?

```
load( "regression1.RData" )
summary_data[ new_dist, new_dist := i.km, on = "trj_id" ]
summary_data[ , is_weekend := weekday_ %in% c("Sat", "Sun") ]
summary_data[ , is_rushhour_morning := hour_ > 7 & hour_ < 10 ]
summary_data[ , is_rushhour_night := hour_ > 17 & hour_ < 20 ]

L1_dist = dataset[ , {
    indices = c(1,.N)
    lat_diff = abs( diff( rawlat[indices] ) )
    lng_diff = abs( diff( rawlng[indices] ) )
    list( L1 = lat_diff + lng_diff )
}, by = "trj_id" ]
```

```
summary_data[ L1_dist, L1_dist := i.L1, on = "trj_id" ]
summary_data[ , eval( c("trj_id", "month_", "speed_avg", "speed_var")) := NULL ]

set.seed(1)
n = nrow( summary_data )
training_set_id = sample( 1:n, n * 0.75 )
training_set = summary_data[ training_set_id ]
test_set = summary_data[ -training_set_id ]
```

```
model4 = lm( timediff ~ .*., training_set )
model4_summary = summary(model4)
model4_rmse = do_rmse( predict( model4, test_set ), test_set )
```

```
## Warning in predict.lm(model4, test_set): prediction from a rank-deficient fit
## may be misleading
```

```
model4_rmse
```

```
## [1] 195.5132
```

```
summary_data[ , L1_dist := NULL ]
training_set[ , L1_dist := NULL ]
```

If we add the L1 distance with everything, the model gets a bit worse.

Polynomial terms for the distance/time variables

I manually tuned the interactions. In general, it's not good to have x * X^2 * x^3 interaction. Cubic interactions are also not great either.

```
# Remove is_weekend because weekday_ is already a factor. Model is over-specified.
training_set[ , is_weekend := NULL ]
model5 = lm( timediff ~ .*. +
        weekday_ * I(dist_^2) * I(hour_^2) * I(new_dist^2) +
        I(new_dist^3) + I(hour_^3) + I(dist_^3),
    training_set )
model5_summary = summary(model5)
model5_rmse = do_rmse( predict( model5, test_set ), test_set )
```

```
## Warning in predict.lm(model5, test_set): prediction from a rank-deficient fit
## may be misleading
```

```
model5_rmse
```

```
## [1] 191.257
```

Decent improvement over the non-polynomial model.

Trip distance percentiles

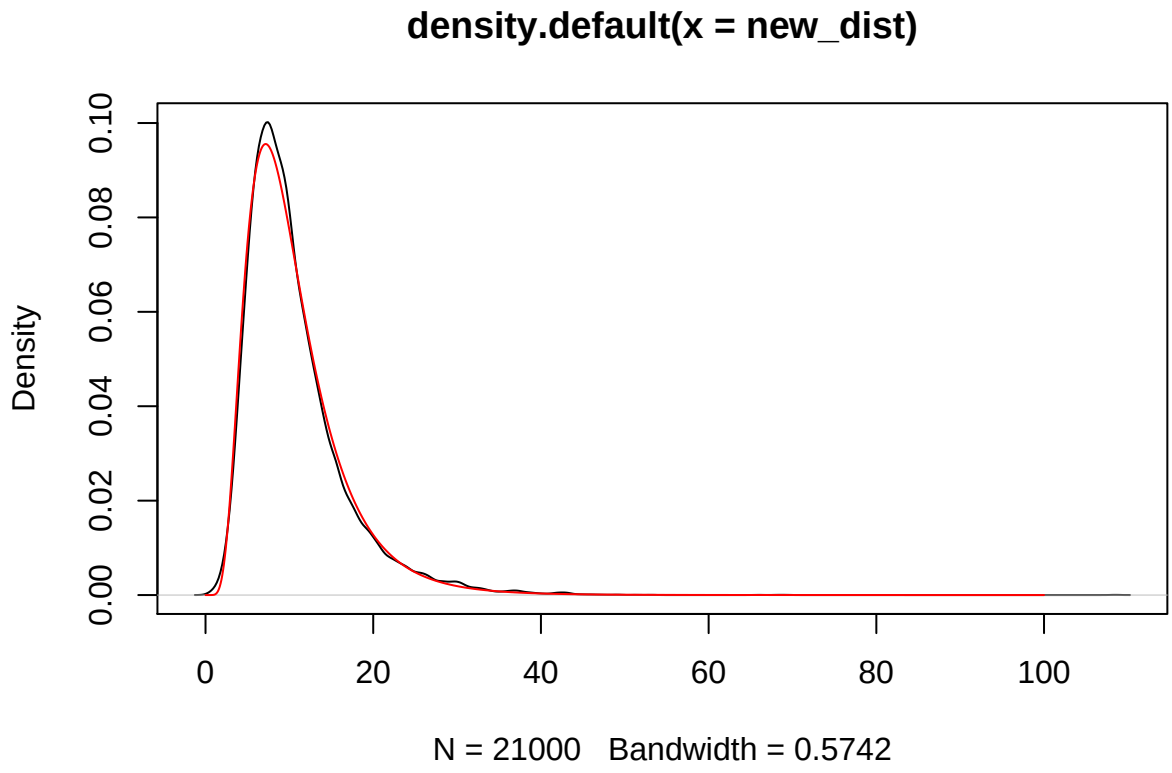
Use a model-based percentile. Log-normal was good from my testing.

```
dist_train = training_set$new_dist

lognorm_loglik = function( theta ){
    if ( any( theta < 0 ) ) return( 1e5 )
```

```
    -sum( dlnorm( training_set$new_dist, theta[1], theta[2], log = T ) )
}
lognorm_params = optim( c(1,1), lognorm_loglik )

training_set[ , plot(density(new_dist)) ]
```

```
## NULL
```

```
xrange = seq(0,100,by=0.1)
lines(
    xrange,
    dlnorm( xrange, lognorm_params$par[1], lognorm_params$par[2] ),
    col = "red"
)
```




```
training_set[ , dist_quant := {
    plnorm( new_dist, lognorm_params$par[1], lognorm_params$par[2] )
} ]
# Very important to use training quantiles to categorise the test set
test_set[ , dist_quant := {
    plnorm( new_dist, lognorm_params$par[1], lognorm_params$par[2] )
} ]

model6 = lm( timediff ~ .*. +
        weekday_ * I(dist_^2) * I(hour_^2) * I(new_dist^2)+
        I(new_dist^3) + I(hour_^3) + I(dist_^3)+
        dist_quant,
```

```
    training_set )
model6_summary = summary(model6)
model6_rmse = do_rmse( predict( model6, test_set ), test_set )
```

```
## Warning in predict.lm(model6, test_set): prediction from a rank-deficient fit
## may be misleading
```

```
model6_rmse
```

```
## [1] 191.5042
```

Not much of a change, but we could find a use for this variable somewhere.

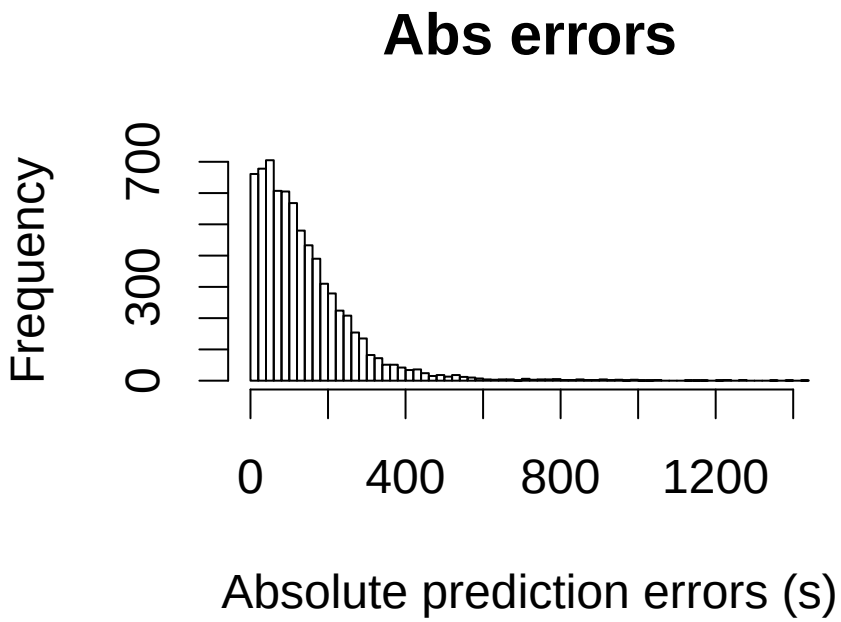
Conclusions

The model so far is a linear model with some polynomial things. I will put the summary of the model at the very end if anyone wants to know what the values of the coefficients were.

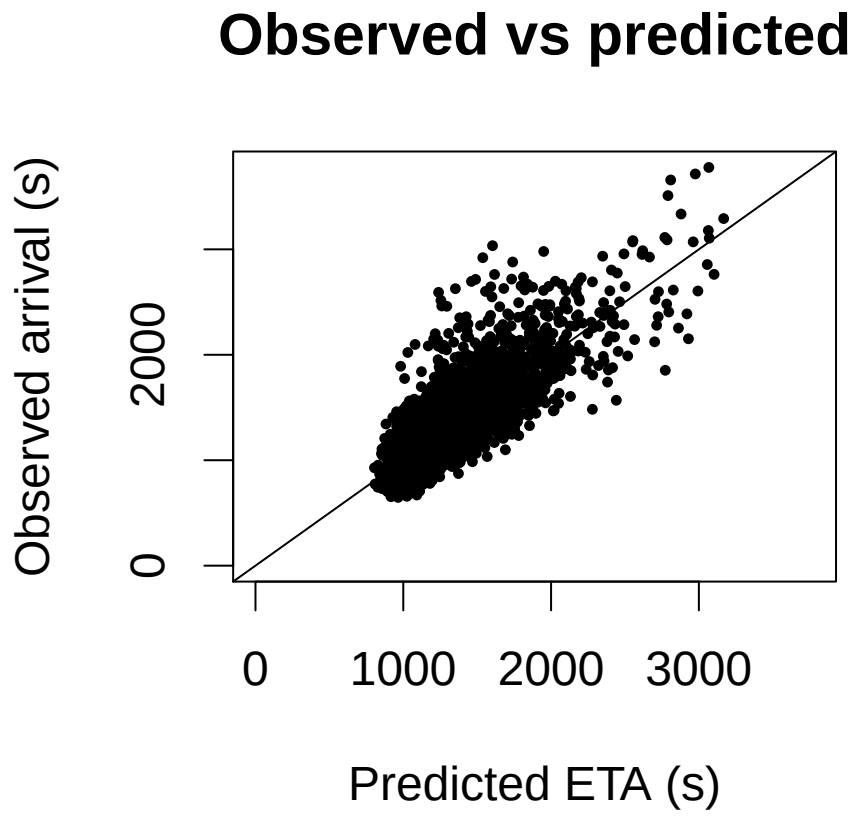
```
model5 = lm( timediff ~ .*. +
        weekday_ * I(dist_^2) * I(hour_^2) * I(new_dist^2) +
        I(new_dist^3) + I(hour_^3) + I(dist_^3),
    training_set )
```

Our RMSE on the test set was 191.2569535. So, about 3mins and 12 seconds average error.

```
yhat = predict( model5, test_set )
par( cex = 1.5 )
hist( abs(yhat - test_set$timediff), main = "Abs errors",
    xlab = "Absolute prediction errors (s)",
    breaks = 100 )
```



```
par( cex = 1.5 )
total_range = range( c( 0, yhat, test_set$timediff ) )
plot( yhat, test_set$timediff, pch = 16, cex = 0.5,
    main = "Observed vs predicted",
    xlab = "Predicted ETA (s)",
    ylab = "Observed arrival (s)",
    xlim = total_range, ylim = total_range
    )
abline( 0, 1 )
```




```
model5_summary
```

```
##
## Call:
## lm(formula = timediff ~ . * . + weekday_ * I(dist_^2) * I(hour_^2) *
##     I(new_dist^2) + I(new_dist^3) + I(hour_^3) + I(dist_^3),
##     data = training_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1193.5  -119.2   -23.4    93.1  4552.2
##
## Coefficients: (1 not defined because of singularities)
```

```
##                                        Estimate Std. Error t value
## (Intercept)                           1.090e+03  4.767e+01  22.864
## dist_                                -5.445e+03  7.497e+02  -7.263
## weekday_Mon                          -1.447e+01  5.210e+01  -0.278
## weekday_Tue                           9.188e+01  6.151e+01   1.494
## weekday_Wed                           6.890e+01  6.188e+01   1.113
## weekday_Thu                          -1.099e+02  5.079e+01  -2.164
## weekday_Fri                          -1.653e+02  5.068e+01  -3.262
## weekday_Sat                          -4.156e+01  5.280e+01  -0.787
## hour_                                -1.175e+01  3.989e+00  -2.945
## min_                                 -9.384e-02  4.074e-01  -0.230
## new_dist                              9.564e+01  4.644e+00  20.592
## is_rushhour_morningTRUE               2.713e+02  1.637e+02   1.657
## is_rushhour_nightTRUE                 2.874e+02  1.519e+02   1.892
## I(dist_^2)                            1.029e+04  4.445e+03   2.314
## I(hour_^2)                            4.667e-02  2.674e-01   0.175
## I(new_dist^2)                         3.135e-02  2.039e-01   0.154
## I(new_dist^3)                         2.969e-03  1.519e-03   1.954
## I(hour_^3)                            1.705e-02  6.501e-03   2.623
## I(dist_^3)                            2.886e+04  9.272e+03   3.112
## dist_:weekday_Mon                     2.921e+02  7.288e+02   0.401
## dist_:weekday_Tue                    -5.259e+02  9.184e+02  -0.573
## dist_:weekday_Wed                    -2.412e+02  9.497e+02  -0.254
## dist_:weekday_Thu                    -1.041e+03  7.281e+02  -1.429
## dist_:weekday_Fri                     2.075e+03  6.836e+02   3.035
## dist_:weekday_Sat                     5.063e+02  7.153e+02   0.708
## dist_:hour_                          -6.779e+01  2.058e+01  -3.294
## dist_:min_                            2.970e+00  3.286e+00   0.904
## dist_:new_dist                       -2.369e+02  1.709e+01 -13.862
## dist_:is_rushhour_morningTRUE         1.045e+02  3.452e+02   0.303
## dist_:is_rushhour_nightTRUE           2.538e+02  2.074e+02   1.224
## weekday_Mon:hour_                     1.464e+01  2.824e+00   5.183
## weekday_Tue:hour_                     1.775e+01  2.943e+00   6.032
## weekday_Wed:hour_                     1.755e+01  2.831e+00   6.198
## weekday_Thu:hour_                     7.471e+00  3.006e+00   2.485
## weekday_Fri:hour_                     1.263e+01  2.762e+00   4.573
## weekday_Sat:hour_                    -1.940e-01  2.836e+00  -0.068
## weekday_Mon:min_                     -3.680e-01  2.923e-01  -1.259
## weekday_Tue:min_                     -1.656e-01  3.043e-01  -0.544
## weekday_Wed:min_                     -1.655e-01  3.004e-01  -0.551
## weekday_Thu:min_                      1.802e-01  3.024e-01   0.596
## weekday_Fri:min_                      2.774e-01  2.951e-01   0.940
## weekday_Sat:min_                      4.134e-01  2.952e-01   1.400
## weekday_Mon:new_dist                 -1.347e+01  5.173e+00  -2.603
## weekday_Tue:new_dist                 -4.137e+01  9.421e+00  -4.391
## weekday_Wed:new_dist                 -4.314e+01  9.020e+00  -4.782
## weekday_Thu:new_dist                  3.880e+01  4.220e+00   9.195
## weekday_Fri:new_dist                 -5.374e+00  4.685e+00  -1.147
## weekday_Sat:new_dist                  8.321e+00  4.171e+00   1.995
## weekday_Mon:is_rushhour_morningTRUE  -7.023e+01  3.418e+01  -2.055
## weekday_Tue:is_rushhour_morningTRUE  -6.300e+01  3.564e+01  -1.767
## weekday_Wed:is_rushhour_morningTRUE  -4.360e+01  3.512e+01  -1.241
## weekday_Thu:is_rushhour_morningTRUE  -5.572e+01  3.533e+01  -1.577
## weekday_Fri:is_rushhour_morningTRUE   4.261e+01  3.889e+01   1.095
```

```
## weekday_Sat:is_rushhour_morningTRUE             -2.158e+01  3.479e+01  -0.620
## weekday_Mon:is_rushhour_nightTRUE               -7.372e+01  1.705e+01  -4.323
## weekday_Tue:is_rushhour_nightTRUE               -7.615e+01  1.726e+01  -4.412
## weekday_Wed:is_rushhour_nightTRUE               -7.667e+01  1.756e+01  -4.366
## weekday_Thu:is_rushhour_nightTRUE               -8.057e+01  1.749e+01  -4.607
## weekday_Fri:is_rushhour_nightTRUE               -6.266e+01  1.678e+01  -3.734
## weekday_Sat:is_rushhour_nightTRUE               -5.032e+01  1.588e+01  -3.169
## hour_:min_                                       1.111e-02  1.139e-02   0.976
## hour_:new_dist                                   3.786e-01  1.539e-01   2.459
## hour_:is_rushhour_morningTRUE                   -2.976e+01  1.836e+01  -1.621
## hour_:is_rushhour_nightTRUE                     -1.541e+01  8.147e+00  -1.892
## min_:new_dist                                   -3.662e-02  2.373e-02  -1.543
## min_:is_rushhour_morningTRUE                    -4.066e-02  5.472e-01  -0.074
## min_:is_rushhour_nightTRUE                       1.372e-01  2.548e-01   0.538
## new_dist:is_rushhour_morningTRUE                -4.954e+00  2.723e+00  -1.820
## new_dist:is_rushhour_nightTRUE                  -3.271e-01  1.517e+00  -0.216
## is_rushhour_morningTRUE:is_rushhour_nightTRUE           NA         NA      NA
## weekday_Mon:I(dist_^2)                          -3.255e+03  2.690e+03  -1.210
## weekday_Tue:I(dist_^2)                           8.269e+03  3.727e+03   2.219
## weekday_Wed:I(dist_^2)                           6.146e+03  4.016e+03   1.530
## weekday_Thu:I(dist_^2)                           1.644e+03  3.099e+03   0.531
## weekday_Fri:I(dist_^2)                          -9.799e+03  2.694e+03  -3.637
## weekday_Sat:I(dist_^2)                          -4.941e+03  2.803e+03  -1.763
## weekday_Mon:I(hour_^2)                          -4.707e-01  1.284e-01  -3.666
## weekday_Tue:I(hour_^2)                          -6.070e-01  1.311e-01  -4.631
## weekday_Wed:I(hour_^2)                          -6.252e-01  1.320e-01  -4.737
## weekday_Thu:I(hour_^2)                          -5.351e-02  1.364e-01  -0.392
## weekday_Fri:I(hour_^2)                          -2.573e-01  1.287e-01  -1.998
## weekday_Sat:I(hour_^2)                           1.262e-01  1.407e-01   0.897
## I(dist_^2):I(hour_^2)                            7.776e+00  3.527e+00   2.205
## weekday_Mon:I(new_dist^2)                        1.194e+00  1.938e-01   6.162
## weekday_Tue:I(new_dist^2)                        9.249e-01  5.481e-01   1.688
## weekday_Wed:I(new_dist^2)                        1.510e+00  5.187e-01   2.911
## weekday_Thu:I(new_dist^2)                       -3.872e+00  2.735e-01 -14.154
## weekday_Fri:I(new_dist^2)                        3.760e-01  1.976e-01   1.903
## weekday_Sat:I(new_dist^2)                       -5.767e-01  1.702e-01  -3.388
## I(dist_^2):I(new_dist^2)                        -2.725e+00  2.055e+00  -1.326
## I(hour_^2):I(new_dist^2)                         1.523e-04  3.376e-04   0.451
## weekday_Mon:I(dist_^2):I(hour_^2)                2.312e+01  4.664e+00   4.956
## weekday_Tue:I(dist_^2):I(hour_^2)                1.402e+01  5.725e+00   2.449
## weekday_Wed:I(dist_^2):I(hour_^2)                2.012e+01  5.573e+00   3.610
## weekday_Thu:I(dist_^2):I(hour_^2)               -9.647e+00  4.005e+00  -2.409
## weekday_Fri:I(dist_^2):I(hour_^2)                6.268e+00  3.561e+00   1.761
## weekday_Sat:I(dist_^2):I(hour_^2)                7.289e+00  3.723e+00   1.958
## weekday_Mon:I(dist_^2):I(new_dist^2)            -8.160e+00  2.280e+00  -3.579
## weekday_Tue:I(dist_^2):I(new_dist^2)            -6.880e+00  6.658e+00  -1.033
## weekday_Wed:I(dist_^2):I(new_dist^2)            -1.398e+01  6.941e+00  -2.015
## weekday_Thu:I(dist_^2):I(new_dist^2)             4.567e+01  4.364e+00  10.467
## weekday_Fri:I(dist_^2):I(new_dist^2)            -4.995e-01  2.338e+00  -0.214
## weekday_Sat:I(dist_^2):I(new_dist^2)             7.600e+00  2.240e+00   3.393
## weekday_Mon:I(hour_^2):I(new_dist^2)            -5.917e-03  6.920e-04  -8.550
## weekday_Tue:I(hour_^2):I(new_dist^2)            -2.862e-04  8.625e-04  -0.332
## weekday_Wed:I(hour_^2):I(new_dist^2)            -1.817e-03  8.557e-04  -2.124
## weekday_Thu:I(hour_^2):I(new_dist^2)             6.372e-03  5.256e-04  12.125
```

```
## weekday_Fri:I(hour_^2):I(new_dist^2)            -1.792e-03  3.530e-04  -5.077
## weekday_Sat:I(hour_^2):I(new_dist^2)             7.761e-04  3.783e-04   2.051
## I(dist_^2):I(hour_^2):I(new_dist^2)             -4.296e-03  4.936e-03  -0.870
## weekday_Mon:I(dist_^2):I(hour_^2):I(new_dist^2)  5.947e-02  1.369e-02   4.344
## weekday_Tue:I(dist_^2):I(hour_^2):I(new_dist^2) -2.684e-02  2.302e-02  -1.166
## weekday_Wed:I(dist_^2):I(hour_^2):I(new_dist^2) -3.139e-03  2.582e-02  -0.122
## weekday_Thu:I(dist_^2):I(hour_^2):I(new_dist^2) -8.407e-02  1.008e-02  -8.339
## weekday_Fri:I(dist_^2):I(hour_^2):I(new_dist^2)  1.765e-02  5.804e-03   3.041
## weekday_Sat:I(dist_^2):I(hour_^2):I(new_dist^2) -1.523e-02  6.977e-03  -2.182
##                                                 Pr(>|t|)
## (Intercept)                                      < 2e-16 ***
## dist_                                           3.91e-13 ***
## weekday_Mon                                     0.781281
## weekday_Tue                                     0.135265
## weekday_Wed                                     0.265510
## weekday_Thu                                     0.030451 *
## weekday_Fri                                     0.001107 **
## weekday_Sat                                     0.431282
## hour_                                           0.003234 **
## min_                                            0.817835
## new_dist                                         < 2e-16 ***
## is_rushhour_morningTRUE                         0.097490 .
## is_rushhour_nightTRUE                           0.058568 .
## I(dist_^2)                                      0.020671 *
## I(hour_^2)                                      0.861414
## I(new_dist^2)                                   0.877786
## I(new_dist^3)                                   0.050670 .
## I(hour_^3)                                      0.008723 **
## I(dist_^3)                                      0.001859 **
## dist_:weekday_Mon                               0.688537
## dist_:weekday_Tue                               0.566874
## dist_:weekday_Wed                               0.799505
## dist_:weekday_Thu                               0.152882
## dist_:weekday_Fri                               0.002409 **
## dist_:weekday_Sat                               0.479016
## dist_:hour_                                     0.000989 ***
## dist_:min_                                      0.366043
## dist_:new_dist                                   < 2e-16 ***
## dist_:is_rushhour_morningTRUE                   0.762037
## dist_:is_rushhour_nightTRUE                     0.221020
## weekday_Mon:hour_                               2.20e-07 ***
## weekday_Tue:hour_                               1.65e-09 ***
## weekday_Wed:hour_                               5.83e-10 ***
## weekday_Thu:hour_                               0.012948 *
## weekday_Fri:hour_                               4.84e-06 ***
## weekday_Sat:hour_                               0.945462
## weekday_Mon:min_                                0.208057
## weekday_Tue:min_                                0.586375
## weekday_Wed:min_                                0.581757
## weekday_Thu:min_                                0.551338
## weekday_Fri:min_                                0.347270
## weekday_Sat:min_                                0.161410
## weekday_Mon:new_dist                            0.009236 **
## weekday_Tue:new_dist                            1.13e-05 ***
```

```
## weekday_Wed:new_dist                                1.75e-06 ***
## weekday_Thu:new_dist                                 < 2e-16 ***
## weekday_Fri:new_dist                                0.251387
## weekday_Sat:new_dist                                0.046053 *
## weekday_Mon:is_rushhour_morningTRUE                 0.039937 *
## weekday_Tue:is_rushhour_morningTRUE                 0.077162 .
## weekday_Wed:is_rushhour_morningTRUE                 0.214445
## weekday_Thu:is_rushhour_morningTRUE                 0.114791
## weekday_Fri:is_rushhour_morningTRUE                 0.273319
## weekday_Sat:is_rushhour_morningTRUE                 0.535061
## weekday_Mon:is_rushhour_nightTRUE                   1.54e-05 ***
## weekday_Tue:is_rushhour_nightTRUE                   1.03e-05 ***
## weekday_Wed:is_rushhour_nightTRUE                   1.27e-05 ***
## weekday_Thu:is_rushhour_nightTRUE                   4.11e-06 ***
## weekday_Fri:is_rushhour_nightTRUE                   0.000189 ***
## weekday_Sat:is_rushhour_nightTRUE                   0.001534 **
## hour_:min_                                          0.329244
## hour_:new_dist                                      0.013935 *
## hour_:is_rushhour_morningTRUE                       0.104957
## hour_:is_rushhour_nightTRUE                         0.058519 .
## min_:new_dist                                       0.122781
## min_:is_rushhour_morningTRUE                        0.940769
## min_:is_rushhour_nightTRUE                          0.590350
## new_dist:is_rushhour_morningTRUE                    0.068828 .
## new_dist:is_rushhour_nightTRUE                      0.829293
## is_rushhour_morningTRUE:is_rushhour_nightTRUE             NA
## weekday_Mon:I(dist_^2)                              0.226301
## weekday_Tue:I(dist_^2)                              0.026499 *
## weekday_Wed:I(dist_^2)                              0.125952
## weekday_Thu:I(dist_^2)                              0.595734
## weekday_Fri:I(dist_^2)                              0.000276 ***
## weekday_Sat:I(dist_^2)                              0.077945 .
## weekday_Mon:I(hour_^2)                              0.000247 ***
## weekday_Tue:I(hour_^2)                              3.67e-06 ***
## weekday_Wed:I(hour_^2)                              2.18e-06 ***
## weekday_Thu:I(hour_^2)                              0.694729
## weekday_Fri:I(hour_^2)                              0.045696 *
## weekday_Sat:I(hour_^2)                              0.369704
## I(dist_^2):I(hour_^2)                               0.027499 *
## weekday_Mon:I(new_dist^2)                           7.33e-10 ***
## weekday_Tue:I(new_dist^2)                           0.091520 .
## weekday_Wed:I(new_dist^2)                           0.003609 **
## weekday_Thu:I(new_dist^2)                            < 2e-16 ***
## weekday_Fri:I(new_dist^2)                           0.057019 .
## weekday_Sat:I(new_dist^2)                           0.000705 ***
## I(dist_^2):I(new_dist^2)                            0.184840
## I(hour_^2):I(new_dist^2)                            0.651944
## weekday_Mon:I(dist_^2):I(hour_^2)                   7.25e-07 ***
## weekday_Tue:I(dist_^2):I(hour_^2)                   0.014343 *
## weekday_Wed:I(dist_^2):I(hour_^2)                   0.000306 ***
## weekday_Thu:I(dist_^2):I(hour_^2)                   0.016016 *
## weekday_Fri:I(dist_^2):I(hour_^2)                   0.078335 .
## weekday_Sat:I(dist_^2):I(hour_^2)                   0.050263 .
## weekday_Mon:I(dist_^2):I(new_dist^2)                0.000345 ***
```

```
## weekday_Tue:I(dist_^2):I(new_dist^2)            0.301464
## weekday_Wed:I(dist_^2):I(new_dist^2)            0.043943 *
## weekday_Thu:I(dist_^2):I(new_dist^2)             < 2e-16 ***
## weekday_Fri:I(dist_^2):I(new_dist^2)            0.830829
## weekday_Sat:I(dist_^2):I(new_dist^2)            0.000694 ***
## weekday_Mon:I(hour_^2):I(new_dist^2)             < 2e-16 ***
## weekday_Tue:I(hour_^2):I(new_dist^2)            0.740022
## weekday_Wed:I(hour_^2):I(new_dist^2)            0.033708 *
## weekday_Thu:I(hour_^2):I(new_dist^2)             < 2e-16 ***
## weekday_Fri:I(hour_^2):I(new_dist^2)            3.87e-07 ***
## weekday_Sat:I(hour_^2):I(new_dist^2)            0.040251 *
## I(dist_^2):I(hour_^2):I(new_dist^2)             0.384174
## weekday_Mon:I(dist_^2):I(hour_^2):I(new_dist^2) 1.40e-05 ***
## weekday_Tue:I(dist_^2):I(hour_^2):I(new_dist^2) 0.243685
## weekday_Wed:I(dist_^2):I(hour_^2):I(new_dist^2) 0.903221
## weekday_Thu:I(dist_^2):I(hour_^2):I(new_dist^2)  < 2e-16 ***
## weekday_Fri:I(dist_^2):I(hour_^2):I(new_dist^2) 0.002359 **
## weekday_Sat:I(dist_^2):I(hour_^2):I(new_dist^2) 0.029101 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 190.4 on 20885 degrees of freedom
## Multiple R-squared:  0.7035, Adjusted R-squared:  0.7019
## F-statistic: 434.7 on 114 and 20885 DF,  p-value: < 2.2e-16
```

```
anova(model5)
```

```
## Analysis of Variance Table
##
## Response: timediff
##                                                Df     Sum Sq    Mean Sq
## dist_                                           1  340582953  340582953
## weekday_                                        6  111468318   18578053
## hour_                                           1   17623558   17623558
## min_                                            1        450        450
## new_dist                                        1 1114660490 1114660490
## is_rushhour_morning                             1    2589540    2589540
## is_rushhour_night                               1    2976765    2976765
## I(dist_^2)                                      1    4536970    4536970
## I(hour_^2)                                      1   20060301   20060301
## I(new_dist^2)                                   1   11078580   11078580
## I(new_dist^3)                                   1   55037134   55037134
## I(hour_^3)                                      1    1501741    1501741
## I(dist_^3)                                      1    4342579    4342579
## dist_:weekday_                                  6    3175867     529311
## dist_:hour_                                     1    1391711    1391711
## dist_:min_                                      1       3356       3356
## dist_:new_dist                                  1   26764542   26764542
## dist_:is_rushhour_morning                       1     137990     137990
## dist_:is_rushhour_night                         1     244008     244008
## weekday_:hour_                                  6   25588860    4264810
## weekday_:min_                                   6     318454      53076
## weekday_:new_dist                               6   14287552    2381259
## weekday_:is_rushhour_morning                    6     414001      69000
## weekday_:is_rushhour_night                      6    2041484     340247
```

```
## hour_:min_                                        1      43944      43944
## hour_:new_dist                                    1     164656     164656
## hour_:is_rushhour_morning                         1      92977      92977
## hour_:is_rushhour_night                           1     116561     116561
## min_:new_dist                                     1     190508     190508
## min_:is_rushhour_morning                          1        197        197
## min_:is_rushhour_night                            1       3122       3122
## new_dist:is_rushhour_morning                      1      69866      69866
## new_dist:is_rushhour_night                        1      62427      62427
## weekday_:I(dist_^2)                               6    1500049     250008
## weekday_:I(hour_^2)                               6    5997975     999662
## I(dist_^2):I(hour_^2)                             1     386609     386609
## weekday_:I(new_dist^2)                            6    6441837    1073640
## I(dist_^2):I(new_dist^2)                          1     459573     459573
## I(hour_^2):I(new_dist^2)                          1       7097       7097
## weekday_:I(dist_^2):I(hour_^2)                    6    1877926     312988
## weekday_:I(dist_^2):I(new_dist^2)                 6    3140572     523429
## weekday_:I(hour_^2):I(new_dist^2)                 6    9441909    1573651
## I(dist_^2):I(hour_^2):I(new_dist^2)               1      12924      12924
## weekday_:I(dist_^2):I(hour_^2):I(new_dist^2)      6    5423470     903912
## Residuals                                     20885  757011132      36247
##                                                  F value    Pr(>F)
## dist_                                         9396.2621 < 2.2e-16 ***
## weekday_                                       512.5455 < 2.2e-16 ***
## hour_                                          486.2122 < 2.2e-16 ***
## min_                                             0.0124 0.9112962
## new_dist                                     30752.1031 < 2.2e-16 ***
## is_rushhour_morning                             71.4422 < 2.2e-16 ***
## is_rushhour_night                               82.1253 < 2.2e-16 ***
## I(dist_^2)                                     125.1694 < 2.2e-16 ***
## I(hour_^2)                                     553.4389 < 2.2e-16 ***
## I(new_dist^2)                                  305.6443 < 2.2e-16 ***
## I(new_dist^3)                                 1518.4064 < 2.2e-16 ***
## I(hour_^3)                                      41.4312 1.247e-10 ***
## I(dist_^3)                                     119.8064 < 2.2e-16 ***
## dist_:weekday_                                  14.6030 < 2.2e-16 ***
## dist_:hour_                                     38.3956 5.884e-10 ***
## dist_:min_                                       0.0926 0.7609156
## dist_:new_dist                                 738.4006 < 2.2e-16 ***
## dist_:is_rushhour_morning                        3.8070 0.0510527 .
## dist_:is_rushhour_night                          6.7319 0.0094771 **
## weekday_:hour_                                 117.6608 < 2.2e-16 ***
## weekday_:min_                                    1.4643 0.1860505
## weekday_:new_dist                               65.6960 < 2.2e-16 ***
## weekday_:is_rushhour_morning                     1.9036 0.0762380 .
## weekday_:is_rushhour_night                       9.3870 2.589e-10 ***
## hour_:min_                                       1.2124 0.2708764
## hour_:new_dist                                   4.5427 0.0330716 *
## hour_:is_rushhour_morning                        2.5651 0.1092598
## hour_:is_rushhour_night                          3.2158 0.0729467 .
## min_:new_dist                                    5.2559 0.0218827 *
## min_:is_rushhour_morning                         0.0054 0.9412491
## min_:is_rushhour_night                           0.0861 0.7691600
## new_dist:is_rushhour_morning                     1.9275 0.1650455
```

```
## new_dist:is_rushhour_night                        1.7223 0.1894133
## weekday_:I(dist_^2)                               6.8974 2.473e-07 ***
## weekday_:I(hour_^2)                              27.5794 < 2.2e-16 ***
## I(dist_^2):I(hour_^2)                            10.6661 0.0010929 **
## weekday_:I(new_dist^2)                           29.6204 < 2.2e-16 ***
## I(dist_^2):I(new_dist^2)                         12.6791 0.0003706 ***
## I(hour_^2):I(new_dist^2)                          0.1958 0.6581372
## weekday_:I(dist_^2):I(hour_^2)                    8.6349 2.093e-09 ***
## weekday_:I(dist_^2):I(new_dist^2)                14.4408 < 2.2e-16 ***
## weekday_:I(hour_^2):I(new_dist^2)                43.4151 < 2.2e-16 ***
## I(dist_^2):I(hour_^2):I(new_dist^2)               0.3566 0.5504332
## weekday_:I(dist_^2):I(hour_^2):I(new_dist^2)     24.9378 < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```