Capstone Project

Reginald Mordi

8/2/2021

A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short-term basis i.e., one day or more for a price, and can also be free in some cases. Many bike share systems allow people to borrow a bike from a dock which is usually computer-controlled where the user enters the payment information, and the system unlocks the bike. This bike can then be returned to another dock belonging to the same system. I decided to choose bike share for my project based on my experience during my vacation time in South Caroline. My objective is to combine historical usage patterns with weather data to forecast bike rental demand, and the goal is to predict the total count of bikes rented during each hour covered by the test set, using only information available prior to the rental period. I would be using my dataset for (<https://archive.ics.uci.edu/ml/datasets/Apartment+for+rent+classified>) Reading the csv files

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.3

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

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v stringr 1.4.0  
## v tidyr 1.1.2 v forcats 0.5.0  
## v readr 1.3.1

## -- Conflicts ---------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(broom)

## Warning: package 'broom' was built under R version 4.0.3

library(caTools)

## Warning: package 'caTools' was built under R version 4.0.3

library(car)

## Loading required package: carData

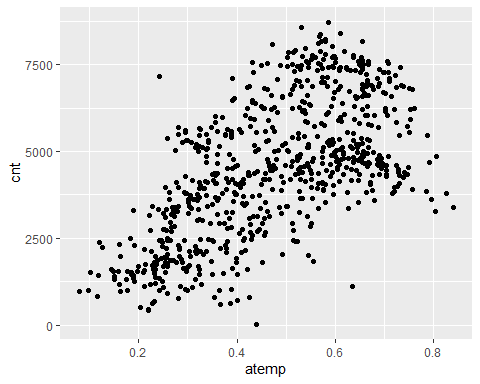
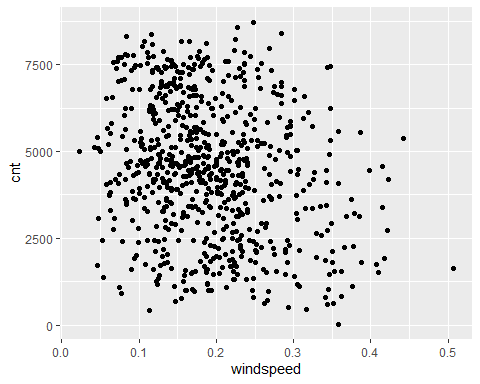
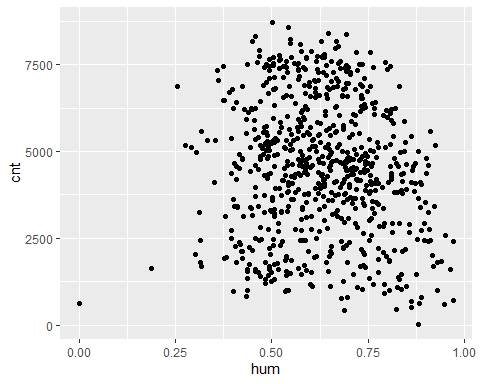
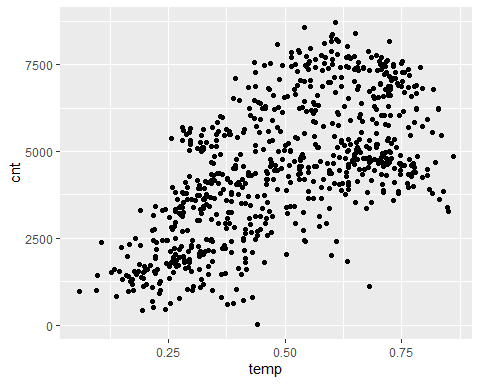
##   
## Attaching package: 'car'

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

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

library(ggplot2)

## Linear regression model



## Group.1 x  
## 1 2011-01-01 0.3441670  
## 2 2011-01-02 0.3634780  
## 3 2011-01-03 0.1963640  
## 4 2011-01-04 0.2000000  
## 5 2011-01-05 0.2269570  
## 6 2011-01-06 0.2043480  
## 7 2011-01-07 0.1965220  
## 8 2011-01-08 0.1650000  
## 9 2011-01-09 0.1383330  
## 10 2011-01-10 0.1508330  
## 11 2011-01-11 0.1690910  
## 12 2011-01-12 0.1727270  
## 13 2011-01-13 0.1650000  
## 14 2011-01-14 0.1608700  
## 15 2011-01-15 0.2333330  
## 16 2011-01-16 0.2316670  
## 17 2011-01-17 0.1758330  
## 18 2011-01-18 0.2166670  
## 19 2011-01-19 0.2921740  
## 20 2011-01-20 0.2616670  
## 21 2011-01-21 0.1775000  
## 22 2011-01-22 0.0591304  
## 23 2011-01-23 0.0965217  
## 24 2011-01-24 0.0973913  
## 25 2011-01-25 0.2234780  
## 26 2011-01-26 0.2175000  
## 27 2011-01-27 0.1950000  
## 28 2011-01-28 0.2034780  
## 29 2011-01-29 0.1965220  
## 30 2011-01-30 0.2165220  
## 31 2011-01-31 0.1808330  
## 32 2011-02-01 0.1921740  
## 33 2011-02-02 0.2600000  
## 34 2011-02-03 0.1869570  
## 35 2011-02-04 0.2113040  
## 36 2011-02-05 0.2333330  
## 37 2011-02-06 0.2858330  
## 38 2011-02-07 0.2716670  
## 39 2011-02-08 0.2208330  
## 40 2011-02-09 0.1347830  
## 41 2011-02-10 0.1443480  
## 42 2011-02-11 0.1890910  
## 43 2011-02-12 0.2225000  
## 44 2011-02-13 0.3165220  
## 45 2011-02-14 0.4150000  
## 46 2011-02-15 0.2660870  
## 47 2011-02-16 0.3182610  
## 48 2011-02-17 0.4358330  
## 49 2011-02-18 0.5216670  
## 50 2011-02-19 0.3991670  
## 51 2011-02-20 0.2852170  
## 52 2011-02-21 0.3033330  
## 53 2011-02-22 0.1822220  
## 54 2011-02-23 0.2217390  
## 55 2011-02-24 0.2956520  
## 56 2011-02-25 0.3643480  
## 57 2011-02-26 0.2825000  
## 58 2011-02-27 0.3434780  
## 59 2011-02-28 0.4072730  
## 60 2011-03-01 0.2666670  
## 61 2011-03-02 0.3350000  
## 62 2011-03-03 0.1983330  
## 63 2011-03-04 0.2616670  
## 64 2011-03-05 0.3841670  
## 65 2011-03-06 0.3765220  
## 66 2011-03-07 0.2617390  
## 67 2011-03-08 0.2925000  
## 68 2011-03-09 0.2958330  
## 69 2011-03-10 0.3890910  
## 70 2011-03-11 0.3165220  
## 71 2011-03-12 0.3291670  
## 72 2011-03-13 0.3843480  
## 73 2011-03-14 0.3252170  
## 74 2011-03-15 0.3173910  
## 75 2011-03-16 0.3652170  
## 76 2011-03-17 0.4150000  
## 77 2011-03-18 0.5400000  
## 78 2011-03-19 0.4725000  
## 79 2011-03-20 0.3325000  
## 80 2011-03-21 0.4304350  
## 81 2011-03-22 0.4416670  
## 82 2011-03-23 0.3469570  
## 83 2011-03-24 0.2850000  
## 84 2011-03-25 0.2641670  
## 85 2011-03-26 0.2658330  
## 86 2011-03-27 0.2530430  
## 87 2011-03-28 0.2643480  
## 88 2011-03-29 0.3025000  
## 89 2011-03-30 0.3000000  
## 90 2011-03-31 0.2683330  
## 91 2011-04-01 0.3000000  
## 92 2011-04-02 0.3150000  
## 93 2011-04-03 0.3783330  
## 94 2011-04-04 0.5733330  
## 95 2011-04-05 0.4141670  
## 96 2011-04-06 0.3908330  
## 97 2011-04-07 0.4375000  
## 98 2011-04-08 0.3358330  
## 99 2011-04-09 0.3425000  
## 100 2011-04-10 0.4266670  
## 101 2011-04-11 0.5956520  
## 102 2011-04-12 0.5025000  
## 103 2011-04-13 0.4125000  
## 104 2011-04-14 0.4675000  
## 105 2011-04-15 0.4466670  
## 106 2011-04-16 0.4308330  
## 107 2011-04-17 0.4566670  
## 108 2011-04-18 0.5125000  
## 109 2011-04-19 0.5058330  
## 110 2011-04-20 0.5950000  
## 111 2011-04-21 0.4591670  
## 112 2011-04-22 0.3366670  
## 113 2011-04-23 0.4600000  
## 114 2011-04-24 0.5816670  
## 115 2011-04-25 0.6066670  
## 116 2011-04-26 0.6316670  
## 117 2011-04-27 0.6200000  
## 118 2011-04-28 0.6175000  
## 119 2011-04-29 0.5100000  
## 120 2011-04-30 0.4725000  
## 121 2011-05-01 0.4516670  
## 122 2011-05-02 0.5491670  
## 123 2011-05-03 0.6166670  
## 124 2011-05-04 0.4141670  
## 125 2011-05-05 0.4591670  
## 126 2011-05-06 0.4791670  
## 127 2011-05-07 0.5200000  
## 128 2011-05-08 0.5283330  
## 129 2011-05-09 0.5325000  
## 130 2011-05-10 0.5325000  
## 131 2011-05-11 0.5425000  
## 132 2011-05-12 0.5350000  
## 133 2011-05-13 0.5125000  
## 134 2011-05-14 0.5208330  
## 135 2011-05-15 0.5625000  
## 136 2011-05-16 0.5775000  
## 137 2011-05-17 0.5616670  
## 138 2011-05-18 0.5500000  
## 139 2011-05-19 0.5308330  
## 140 2011-05-20 0.5366670  
## 141 2011-05-21 0.6025000  
## 142 2011-05-22 0.6041670  
## 143 2011-05-23 0.6316670  
## 144 2011-05-24 0.6600000  
## 145 2011-05-25 0.6608330  
## 146 2011-05-26 0.7083330  
## 147 2011-05-27 0.6816670  
## 148 2011-05-28 0.6558330  
## 149 2011-05-29 0.6675000  
## 150 2011-05-30 0.7333330  
## 151 2011-05-31 0.7750000  
## 152 2011-06-01 0.7641670  
## 153 2011-06-02 0.7150000  
## 154 2011-06-03 0.6200000  
## 155 2011-06-04 0.6350000  
## 156 2011-06-05 0.6483330  
## 157 2011-06-06 0.6783330  
## 158 2011-06-07 0.7075000  
## 159 2011-06-08 0.7758330  
## 160 2011-06-09 0.8083330  
## 161 2011-06-10 0.7550000  
## 162 2011-06-11 0.7250000  
## 163 2011-06-12 0.6925000  
## 164 2011-06-13 0.6350000  
## 165 2011-06-14 0.6041670  
## 166 2011-06-15 0.6266670  
## 167 2011-06-16 0.6283330  
## 168 2011-06-17 0.6491670  
## 169 2011-06-18 0.6966670  
## 170 2011-06-19 0.6991670  
## 171 2011-06-20 0.6350000  
## 172 2011-06-21 0.6808330  
## 173 2011-06-22 0.7333330  
## 174 2011-06-23 0.7283330  
## 175 2011-06-24 0.7241670  
## 176 2011-06-25 0.6950000  
## 177 2011-06-26 0.6800000  
## 178 2011-06-27 0.6825000  
## 179 2011-06-28 0.7441670  
## 180 2011-06-29 0.7283330  
## 181 2011-06-30 0.6966670  
## 182 2011-07-01 0.7225000  
## 183 2011-07-02 0.7383330  
## 184 2011-07-03 0.7166670  
## 185 2011-07-04 0.7266670  
## 186 2011-07-05 0.7466670  
## 187 2011-07-06 0.7200000  
## 188 2011-07-07 0.7500000  
## 189 2011-07-08 0.7091670  
## 190 2011-07-09 0.7333330  
## 191 2011-07-10 0.7475000  
## 192 2011-07-11 0.7625000  
## 193 2011-07-12 0.7941670  
## 194 2011-07-13 0.7466670  
## 195 2011-07-14 0.6808330  
## 196 2011-07-15 0.6633330  
## 197 2011-07-16 0.6866670  
## 198 2011-07-17 0.7191670  
## 199 2011-07-18 0.7466670  
## 200 2011-07-19 0.7766670  
## 201 2011-07-20 0.7683330  
## 202 2011-07-21 0.8150000  
## 203 2011-07-22 0.8483330  
## 204 2011-07-23 0.8491670  
## 205 2011-07-24 0.8300000  
## 206 2011-07-25 0.7433330  
## 207 2011-07-26 0.7716670  
## 208 2011-07-27 0.7750000  
## 209 2011-07-28 0.7791670  
## 210 2011-07-29 0.8383330  
## 211 2011-07-30 0.8041670  
## 212 2011-07-31 0.8058330  
## 213 2011-08-01 0.7716670  
## 214 2011-08-02 0.7833330  
## 215 2011-08-03 0.7316670  
## 216 2011-08-04 0.7100000  
## 217 2011-08-05 0.7108330  
## 218 2011-08-06 0.7166670  
## 219 2011-08-07 0.7425000  
## 220 2011-08-08 0.7650000  
## 221 2011-08-09 0.7750000  
## 222 2011-08-10 0.7666670  
## 223 2011-08-11 0.7175000  
## 224 2011-08-12 0.7083330  
## 225 2011-08-13 0.6858330  
## 226 2011-08-14 0.6766670  
## 227 2011-08-15 0.6658330  
## 228 2011-08-16 0.7008330  
## 229 2011-08-17 0.7233330  
## 230 2011-08-18 0.7116670  
## 231 2011-08-19 0.6850000  
## 232 2011-08-20 0.6975000  
## 233 2011-08-21 0.7108330  
## 234 2011-08-22 0.6916670  
## 235 2011-08-23 0.6408330  
## 236 2011-08-24 0.6733330  
## 237 2011-08-25 0.6841670  
## 238 2011-08-26 0.7000000  
## 239 2011-08-27 0.6800000  
## 240 2011-08-28 0.7070590  
## 241 2011-08-29 0.6366670  
## 242 2011-08-30 0.6391670  
## 243 2011-08-31 0.6566670  
## 244 2011-09-01 0.6550000  
## 245 2011-09-02 0.6433330  
## 246 2011-09-03 0.6691670  
## 247 2011-09-04 0.7091670  
## 248 2011-09-05 0.6733330  
## 249 2011-09-06 0.5400000  
## 250 2011-09-07 0.5991670  
## 251 2011-09-08 0.6339130  
## 252 2011-09-09 0.6500000  
## 253 2011-09-10 0.6600000  
## 254 2011-09-11 0.6533330  
## 255 2011-09-12 0.6443480  
## 256 2011-09-13 0.6508330  
## 257 2011-09-14 0.6733330  
## 258 2011-09-15 0.5775000  
## 259 2011-09-16 0.4691670  
## 260 2011-09-17 0.4916670  
## 261 2011-09-18 0.5075000  
## 262 2011-09-19 0.5491670  
## 263 2011-09-20 0.5616670  
## 264 2011-09-21 0.5950000  
## 265 2011-09-22 0.6283330  
## 266 2011-09-23 0.6091670  
## 267 2011-09-24 0.6066670  
## 268 2011-09-25 0.6341670  
## 269 2011-09-26 0.6491670  
## 270 2011-09-27 0.6366670  
## 271 2011-09-28 0.6350000  
## 272 2011-09-29 0.6166670  
## 273 2011-09-30 0.5641670  
## 274 2011-10-01 0.4100000  
## 275 2011-10-02 0.3566670  
## 276 2011-10-03 0.3841670  
## 277 2011-10-04 0.4841670  
## 278 2011-10-05 0.5383330  
## 279 2011-10-06 0.4941670  
## 280 2011-10-07 0.5108330  
## 281 2011-10-08 0.5216670  
## 282 2011-10-09 0.5408330  
## 283 2011-10-10 0.5708330  
## 284 2011-10-11 0.5666670  
## 285 2011-10-12 0.5433330  
## 286 2011-10-13 0.5891670  
## 287 2011-10-14 0.5508330  
## 288 2011-10-15 0.5066670  
## 289 2011-10-16 0.5116670  
## 290 2011-10-17 0.5341670  
## 291 2011-10-18 0.5325000  
## 292 2011-10-19 0.5417390  
## 293 2011-10-20 0.4758330  
## 294 2011-10-21 0.4275000  
## 295 2011-10-22 0.4225000  
## 296 2011-10-23 0.4216670  
## 297 2011-10-24 0.4633330  
## 298 2011-10-25 0.4716670  
## 299 2011-10-26 0.4841670  
## 300 2011-10-27 0.4700000  
## 301 2011-10-28 0.3308330  
## 302 2011-10-29 0.2541670  
## 303 2011-10-30 0.3191670  
## 304 2011-10-31 0.3400000  
## 305 2011-11-01 0.4008330  
## 306 2011-11-02 0.3775000  
## 307 2011-11-03 0.4083330  
## 308 2011-11-04 0.4033330  
## 309 2011-11-05 0.3266670  
## 310 2011-11-06 0.3483330  
## 311 2011-11-07 0.3950000  
## 312 2011-11-08 0.4083330  
## 313 2011-11-09 0.4000000  
## 314 2011-11-10 0.3800000  
## 315 2011-11-11 0.3241670  
## 316 2011-11-12 0.3566670  
## 317 2011-11-13 0.4408330  
## 318 2011-11-14 0.5300000  
## 319 2011-11-15 0.5300000  
## 320 2011-11-16 0.4566670  
## 321 2011-11-17 0.3416670  
## 322 2011-11-18 0.2741670  
## 323 2011-11-19 0.3291670  
## 324 2011-11-20 0.4633330  
## 325 2011-11-21 0.4475000  
## 326 2011-11-22 0.4166670  
## 327 2011-11-23 0.4408330  
## 328 2011-11-24 0.3733330  
## 329 2011-11-25 0.3750000  
## 330 2011-11-26 0.3758330  
## 331 2011-11-27 0.4591670  
## 332 2011-11-28 0.5034780  
## 333 2011-11-29 0.4583330  
## 334 2011-11-30 0.3250000  
## 335 2011-12-01 0.3125000  
## 336 2011-12-02 0.3141670  
## 337 2011-12-03 0.2991670  
## 338 2011-12-04 0.3308330  
## 339 2011-12-05 0.3858330  
## 340 2011-12-06 0.4625000  
## 341 2011-12-07 0.4100000  
## 342 2011-12-08 0.2658330  
## 343 2011-12-09 0.2908330  
## 344 2011-12-10 0.2750000  
## 345 2011-12-11 0.2208330  
## 346 2011-12-12 0.2383330  
## 347 2011-12-13 0.2825000  
## 348 2011-12-14 0.3175000  
## 349 2011-12-15 0.4225000  
## 350 2011-12-16 0.3750000  
## 351 2011-12-17 0.2583330  
## 352 2011-12-18 0.2383330  
## 353 2011-12-19 0.2766670  
## 354 2011-12-20 0.3858330  
## 355 2011-12-21 0.4283330  
## 356 2011-12-22 0.4233330  
## 357 2011-12-23 0.3733330  
## 358 2011-12-24 0.3025000  
## 359 2011-12-25 0.2747830  
## 360 2011-12-26 0.3217390  
## 361 2011-12-27 0.3250000  
## 362 2011-12-28 0.2991300  
## 363 2011-12-29 0.2483330  
## 364 2011-12-30 0.3116670  
## 365 2011-12-31 0.4100000  
## 366 2012-01-01 0.3700000  
## 367 2012-01-02 0.2730430  
## 368 2012-01-03 0.1500000  
## 369 2012-01-04 0.1075000  
## 370 2012-01-05 0.2658330  
## 371 2012-01-06 0.3341670  
## 372 2012-01-07 0.3933330  
## 373 2012-01-08 0.3375000  
## 374 2012-01-09 0.2241670  
## 375 2012-01-10 0.3086960  
## 376 2012-01-11 0.2741670  
## 377 2012-01-12 0.3825000  
## 378 2012-01-13 0.2741670  
## 379 2012-01-14 0.1800000  
## 380 2012-01-15 0.1666670  
## 381 2012-01-16 0.1900000  
## 382 2012-01-17 0.3730430  
## 383 2012-01-18 0.3033330  
## 384 2012-01-19 0.1900000  
## 385 2012-01-20 0.2175000  
## 386 2012-01-21 0.1733330  
## 387 2012-01-22 0.1625000  
## 388 2012-01-23 0.2183330  
## 389 2012-01-24 0.3425000  
## 390 2012-01-25 0.2941670  
## 391 2012-01-26 0.3416670  
## 392 2012-01-27 0.4250000  
## 393 2012-01-28 0.3158330  
## 394 2012-01-29 0.2825000  
## 395 2012-01-30 0.2691670  
## 396 2012-01-31 0.3900000  
## 397 2012-02-01 0.4691670  
## 398 2012-02-02 0.3991670  
## 399 2012-02-03 0.3133330  
## 400 2012-02-04 0.2641670  
## 401 2012-02-05 0.2658330  
## 402 2012-02-06 0.2826090  
## 403 2012-02-07 0.3541670  
## 404 2012-02-08 0.2566670  
## 405 2012-02-09 0.2650000  
## 406 2012-02-10 0.2808330  
## 407 2012-02-11 0.2241670  
## 408 2012-02-12 0.1275000  
## 409 2012-02-13 0.2225000  
## 410 2012-02-14 0.3191670  
## 411 2012-02-15 0.3483330  
## 412 2012-02-16 0.3166670  
## 413 2012-02-17 0.3433330  
## 414 2012-02-18 0.3466670  
## 415 2012-02-19 0.2800000  
## 416 2012-02-20 0.2800000  
## 417 2012-02-21 0.2878260  
## 418 2012-02-22 0.3958330  
## 419 2012-02-23 0.4541670  
## 420 2012-02-24 0.4075000  
## 421 2012-02-25 0.2908330  
## 422 2012-02-26 0.2791670  
## 423 2012-02-27 0.3666670  
## 424 2012-02-28 0.3591670  
## 425 2012-02-29 0.3443480  
## 426 2012-03-01 0.4858330  
## 427 2012-03-02 0.3533330  
## 428 2012-03-03 0.4141670  
## 429 2012-03-04 0.3258330  
## 430 2012-03-05 0.2433330  
## 431 2012-03-06 0.2583330  
## 432 2012-03-07 0.4041670  
## 433 2012-03-08 0.5275000  
## 434 2012-03-09 0.4108330  
## 435 2012-03-10 0.2875000  
## 436 2012-03-11 0.3617390  
## 437 2012-03-12 0.4666670  
## 438 2012-03-13 0.5650000  
## 439 2012-03-14 0.5725000  
## 440 2012-03-15 0.5575000  
## 441 2012-03-16 0.4358330  
## 442 2012-03-17 0.5141670  
## 443 2012-03-18 0.4725000  
## 444 2012-03-19 0.5450000  
## 445 2012-03-20 0.5608330  
## 446 2012-03-21 0.5316670  
## 447 2012-03-22 0.5541670  
## 448 2012-03-23 0.6016670  
## 449 2012-03-24 0.5025000  
## 450 2012-03-25 0.4375000  
## 451 2012-03-26 0.4458330  
## 452 2012-03-27 0.3233330  
## 453 2012-03-28 0.4841670  
## 454 2012-03-29 0.4941670  
## 455 2012-03-30 0.3700000  
## 456 2012-03-31 0.4241670  
## 457 2012-04-01 0.4258330  
## 458 2012-04-02 0.4339130  
## 459 2012-04-03 0.4666670  
## 460 2012-04-04 0.5416670  
## 461 2012-04-05 0.4350000  
## 462 2012-04-06 0.4033330  
## 463 2012-04-07 0.4375000  
## 464 2012-04-08 0.5000000  
## 465 2012-04-09 0.4891670  
## 466 2012-04-10 0.4466670  
## 467 2012-04-11 0.3486960  
## 468 2012-04-12 0.3975000  
## 469 2012-04-13 0.4425000  
## 470 2012-04-14 0.4950000  
## 471 2012-04-15 0.6066670  
## 472 2012-04-16 0.6641670  
## 473 2012-04-17 0.6083330  
## 474 2012-04-18 0.4633330  
## 475 2012-04-19 0.4983330  
## 476 2012-04-20 0.5266670  
## 477 2012-04-21 0.5700000  
## 478 2012-04-22 0.3966670  
## 479 2012-04-23 0.3216670  
## 480 2012-04-24 0.4133330  
## 481 2012-04-25 0.4766670  
## 482 2012-04-26 0.4983330  
## 483 2012-04-27 0.4575000  
## 484 2012-04-28 0.3766670  
## 485 2012-04-29 0.4583330  
## 486 2012-04-30 0.4641670  
## 487 2012-05-01 0.6133330  
## 488 2012-05-02 0.5641670  
## 489 2012-05-03 0.5600000  
## 490 2012-05-04 0.6275000  
## 491 2012-05-05 0.6216670  
## 492 2012-05-06 0.5625000  
## 493 2012-05-07 0.5375000  
## 494 2012-05-08 0.5816670  
## 495 2012-05-09 0.5750000  
## 496 2012-05-10 0.5058330  
## 497 2012-05-11 0.5333330  
## 498 2012-05-12 0.5641670  
## 499 2012-05-13 0.6125000  
## 500 2012-05-14 0.5733330  
## 501 2012-05-15 0.6116670  
## 502 2012-05-16 0.6366670  
## 503 2012-05-17 0.5933330  
## 504 2012-05-18 0.5641670  
## 505 2012-05-19 0.6000000  
## 506 2012-05-20 0.6208330  
## 507 2012-05-21 0.5983330  
## 508 2012-05-22 0.6150000  
## 509 2012-05-23 0.6216670  
## 510 2012-05-24 0.6550000  
## 511 2012-05-25 0.6800000  
## 512 2012-05-26 0.6925000  
## 513 2012-05-27 0.6900000  
## 514 2012-05-28 0.7125000  
## 515 2012-05-29 0.7225000  
## 516 2012-05-30 0.6566670  
## 517 2012-05-31 0.6800000  
## 518 2012-06-01 0.6541670  
## 519 2012-06-02 0.5833330  
## 520 2012-06-03 0.6025000  
## 521 2012-06-04 0.5975000  
## 522 2012-06-05 0.5408330  
## 523 2012-06-06 0.5541670  
## 524 2012-06-07 0.6025000  
## 525 2012-06-08 0.6491670  
## 526 2012-06-09 0.7108330  
## 527 2012-06-10 0.7266670  
## 528 2012-06-11 0.7208330  
## 529 2012-06-12 0.6533330  
## 530 2012-06-13 0.6558330  
## 531 2012-06-14 0.6483330  
## 532 2012-06-15 0.6391670  
## 533 2012-06-16 0.6316670  
## 534 2012-06-17 0.5925000  
## 535 2012-06-18 0.5683330  
## 536 2012-06-19 0.6883330  
## 537 2012-06-20 0.7825000  
## 538 2012-06-21 0.8058330  
## 539 2012-06-22 0.7775000  
## 540 2012-06-23 0.7316670  
## 541 2012-06-24 0.7433330  
## 542 2012-06-25 0.7158330  
## 543 2012-06-26 0.6308330  
## 544 2012-06-27 0.6975000  
## 545 2012-06-28 0.7491670  
## 546 2012-06-29 0.8341670  
## 547 2012-06-30 0.7650000  
## 548 2012-07-01 0.8158330  
## 549 2012-07-02 0.7816670  
## 550 2012-07-03 0.7808330  
## 551 2012-07-04 0.7891670  
## 552 2012-07-05 0.8275000  
## 553 2012-07-06 0.8283330  
## 554 2012-07-07 0.8616670  
## 555 2012-07-08 0.8225000  
## 556 2012-07-09 0.7108330  
## 557 2012-07-10 0.7208330  
## 558 2012-07-11 0.7166670  
## 559 2012-07-12 0.7158330  
## 560 2012-07-13 0.7316670  
## 561 2012-07-14 0.7033330  
## 562 2012-07-15 0.7458330  
## 563 2012-07-16 0.7633330  
## 564 2012-07-17 0.8183330  
## 565 2012-07-18 0.7933330  
## 566 2012-07-19 0.7700000  
## 567 2012-07-20 0.6658330  
## 568 2012-07-21 0.5958330  
## 569 2012-07-22 0.6675000  
## 570 2012-07-23 0.7416670  
## 571 2012-07-24 0.7508330  
## 572 2012-07-25 0.7241670  
## 573 2012-07-26 0.7766670  
## 574 2012-07-27 0.7816670  
## 575 2012-07-28 0.7558330  
## 576 2012-07-29 0.7216670  
## 577 2012-07-30 0.7308330  
## 578 2012-07-31 0.7133330  
## 579 2012-08-01 0.7175000  
## 580 2012-08-02 0.7525000  
## 581 2012-08-03 0.7658330  
## 582 2012-08-04 0.7933330  
## 583 2012-08-05 0.7691670  
## 584 2012-08-06 0.7525000  
## 585 2012-08-07 0.7358330  
## 586 2012-08-08 0.7500000  
## 587 2012-08-09 0.7558330  
## 588 2012-08-10 0.7158330  
## 589 2012-08-11 0.6925000  
## 590 2012-08-12 0.7008330  
## 591 2012-08-13 0.7208330  
## 592 2012-08-14 0.7266670  
## 593 2012-08-15 0.7066670  
## 594 2012-08-16 0.7191670  
## 595 2012-08-17 0.7233330  
## 596 2012-08-18 0.6783330  
## 597 2012-08-19 0.6358330  
## 598 2012-08-20 0.6358330  
## 599 2012-08-21 0.6491670  
## 600 2012-08-22 0.6675000  
## 601 2012-08-23 0.6958330  
## 602 2012-08-24 0.7025000  
## 603 2012-08-25 0.6616670  
## 604 2012-08-26 0.6533330  
## 605 2012-08-27 0.7033330  
## 606 2012-08-28 0.7283330  
## 607 2012-08-29 0.6850000  
## 608 2012-08-30 0.7066670  
## 609 2012-08-31 0.7641670  
## 610 2012-09-01 0.7533330  
## 611 2012-09-02 0.6966670  
## 612 2012-09-03 0.7075000  
## 613 2012-09-04 0.7258330  
## 614 2012-09-05 0.7366670  
## 615 2012-09-06 0.6966670  
## 616 2012-09-07 0.7033330  
## 617 2012-09-08 0.6591670  
## 618 2012-09-09 0.6100000  
## 619 2012-09-10 0.5833330  
## 620 2012-09-11 0.5775000  
## 621 2012-09-12 0.5991670  
## 622 2012-09-13 0.6125000  
## 623 2012-09-14 0.6333330  
## 624 2012-09-15 0.6083330  
## 625 2012-09-16 0.5800000  
## 626 2012-09-17 0.5808330  
## 627 2012-09-18 0.6233330  
## 628 2012-09-19 0.5525000  
## 629 2012-09-20 0.5466670  
## 630 2012-09-21 0.5991670  
## 631 2012-09-22 0.6500000  
## 632 2012-09-23 0.5291670  
## 633 2012-09-24 0.5141670  
## 634 2012-09-25 0.5500000  
## 635 2012-09-26 0.6350000  
## 636 2012-09-27 0.6500000  
## 637 2012-09-28 0.6191670  
## 638 2012-09-29 0.5425000  
## 639 2012-09-30 0.5266670  
## 640 2012-10-01 0.5208330  
## 641 2012-10-02 0.5908330  
## 642 2012-10-03 0.6575000  
## 643 2012-10-04 0.6575000  
## 644 2012-10-05 0.6150000  
## 645 2012-10-06 0.5541670  
## 646 2012-10-07 0.4158330  
## 647 2012-10-08 0.3833330  
## 648 2012-10-09 0.4466670  
## 649 2012-10-10 0.5141670  
## 650 2012-10-11 0.4350000  
## 651 2012-10-12 0.4375000  
## 652 2012-10-13 0.3933330  
## 653 2012-10-14 0.5216670  
## 654 2012-10-15 0.5616670  
## 655 2012-10-16 0.4683330  
## 656 2012-10-17 0.4558330  
## 657 2012-10-18 0.5225000  
## 658 2012-10-19 0.5633330  
## 659 2012-10-20 0.4841670  
## 660 2012-10-21 0.4641670  
## 661 2012-10-22 0.4875000  
## 662 2012-10-23 0.5441670  
## 663 2012-10-24 0.5875000  
## 664 2012-10-25 0.5500000  
## 665 2012-10-26 0.5458330  
## 666 2012-10-27 0.5300000  
## 667 2012-10-28 0.4775000  
## 668 2012-10-29 0.4400000  
## 669 2012-10-30 0.3181820  
## 670 2012-10-31 0.3575000  
## 671 2012-11-01 0.3658330  
## 672 2012-11-02 0.3550000  
## 673 2012-11-03 0.3433330  
## 674 2012-11-04 0.3258330  
## 675 2012-11-05 0.3191670  
## 676 2012-11-06 0.2808330  
## 677 2012-11-07 0.2958330  
## 678 2012-11-08 0.3521740  
## 679 2012-11-09 0.3616670  
## 680 2012-11-10 0.3891670  
## 681 2012-11-11 0.4208330  
## 682 2012-11-12 0.4850000  
## 683 2012-11-13 0.3433330  
## 684 2012-11-14 0.2891670  
## 685 2012-11-15 0.3216670  
## 686 2012-11-16 0.3450000  
## 687 2012-11-17 0.3250000  
## 688 2012-11-18 0.3425000  
## 689 2012-11-19 0.3808330  
## 690 2012-11-20 0.3741670  
## 691 2012-11-21 0.3533330  
## 692 2012-11-22 0.3400000  
## 693 2012-11-23 0.3683330  
## 694 2012-11-24 0.2783330  
## 695 2012-11-25 0.2458330  
## 696 2012-11-26 0.3133330  
## 697 2012-11-27 0.2916670  
## 698 2012-11-28 0.2966670  
## 699 2012-11-29 0.2808700  
## 700 2012-11-30 0.2983330  
## 701 2012-12-01 0.2983330  
## 702 2012-12-02 0.3475000  
## 703 2012-12-03 0.4525000  
## 704 2012-12-04 0.4758330  
## 705 2012-12-05 0.4383330  
## 706 2012-12-06 0.2558330  
## 707 2012-12-07 0.3208330  
## 708 2012-12-08 0.3816670  
## 709 2012-12-09 0.3841670  
## 710 2012-12-10 0.4358330  
## 711 2012-12-11 0.3533330  
## 712 2012-12-12 0.2975000  
## 713 2012-12-13 0.2958330  
## 714 2012-12-14 0.2816670  
## 715 2012-12-15 0.3241670  
## 716 2012-12-16 0.3625000  
## 717 2012-12-17 0.3933330  
## 718 2012-12-18 0.4108330  
## 719 2012-12-19 0.3325000  
## 720 2012-12-20 0.3300000  
## 721 2012-12-21 0.3266670  
## 722 2012-12-22 0.2658330  
## 723 2012-12-23 0.2458330  
## 724 2012-12-24 0.2313040  
## 725 2012-12-25 0.2913040  
## 726 2012-12-26 0.2433330  
## 727 2012-12-27 0.2541670  
## 728 2012-12-28 0.2533330  
## 729 2012-12-29 0.2533330  
## 730 2012-12-30 0.2558330  
## 731 2012-12-31 0.2158330

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 22 3152 4548 4504 5956 8714

## [1] 1937.211

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 731 x 3  
## dteday avg sd  
## <chr> <dbl> <dbl>  
## 1 2011-01-01 0.344 NA  
## 2 2011-01-02 0.363 NA  
## 3 2011-01-03 0.196 NA  
## 4 2011-01-04 0.2 NA  
## 5 2011-01-05 0.227 NA  
## 6 2011-01-06 0.204 NA  
## 7 2011-01-07 0.197 NA  
## 8 2011-01-08 0.165 NA  
## 9 2011-01-09 0.138 NA  
## 10 2011-01-10 0.151 NA  
## # ... with 721 more rows

## instant dteday season yr   
## Min. : 1.0 Length:731 Min. :1.000 Min. :0.0000   
## 1st Qu.:183.5 Class :character 1st Qu.:2.000 1st Qu.:0.0000   
## Median :366.0 Mode :character Median :3.000 Median :1.0000   
## Mean :366.0 Mean :2.497 Mean :0.5007   
## 3rd Qu.:548.5 3rd Qu.:3.000 3rd Qu.:1.0000   
## Max. :731.0 Max. :4.000 Max. :1.0000   
## mnth holiday weekday workingday   
## Min. : 1.00 Min. :0.00000 Min. :0.000 Min. :0.000   
## 1st Qu.: 4.00 1st Qu.:0.00000 1st Qu.:1.000 1st Qu.:0.000   
## Median : 7.00 Median :0.00000 Median :3.000 Median :1.000   
## Mean : 6.52 Mean :0.02873 Mean :2.997 Mean :0.684   
## 3rd Qu.:10.00 3rd Qu.:0.00000 3rd Qu.:5.000 3rd Qu.:1.000   
## Max. :12.00 Max. :1.00000 Max. :6.000 Max. :1.000   
## weathersit temp atemp hum   
## Min. :1.000 Min. :0.05913 Min. :0.07907 Min. :0.0000   
## 1st Qu.:1.000 1st Qu.:0.33708 1st Qu.:0.33784 1st Qu.:0.5200   
## Median :1.000 Median :0.49833 Median :0.48673 Median :0.6267   
## Mean :1.395 Mean :0.49538 Mean :0.47435 Mean :0.6279   
## 3rd Qu.:2.000 3rd Qu.:0.65542 3rd Qu.:0.60860 3rd Qu.:0.7302   
## Max. :3.000 Max. :0.86167 Max. :0.84090 Max. :0.9725   
## windspeed casual registered cnt   
## Min. :0.02239 Min. : 2.0 Min. : 20 Min. : 22   
## 1st Qu.:0.13495 1st Qu.: 315.5 1st Qu.:2497 1st Qu.:3152   
## Median :0.18097 Median : 713.0 Median :3662 Median :4548   
## Mean :0.19049 Mean : 848.2 Mean :3656 Mean :4504   
## 3rd Qu.:0.23321 3rd Qu.:1096.0 3rd Qu.:4776 3rd Qu.:5956   
## Max. :0.50746 Max. :3410.0 Max. :6946 Max. :8714

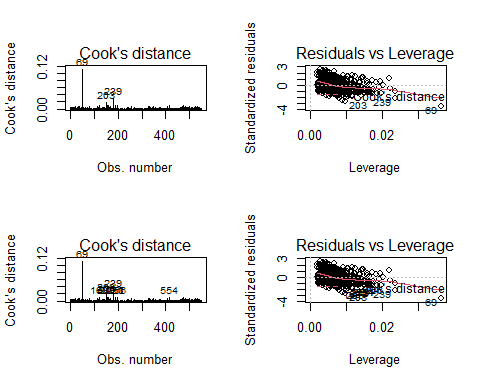
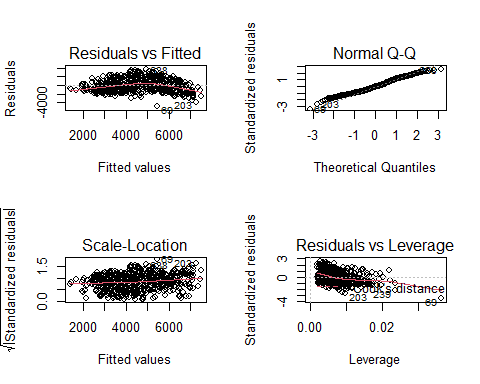
## temp atemp hum windspeed cnt  
## temp 1.0000000 0.9917016 0.1269629 -0.1579441 0.6274940  
## atemp 0.9917016 1.0000000 0.1399881 -0.1836430 0.6310657  
## hum 0.1269629 0.1399881 1.0000000 -0.2484891 -0.1006586  
## windspeed -0.1579441 -0.1836430 -0.2484891 1.0000000 -0.2345450  
## cnt 0.6274940 0.6310657 -0.1006586 -0.2345450 1.0000000

## temp atemp hum windspeed   
## 152.468799 153.849442 1.078237 1.147322

##   
## Call:  
## lm(formula = cnt ~ temp + atemp + hum + windspeed, data = TrainingData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5086.1 -1093.8 -91.5 1094.4 3409.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3274.9 429.9 7.618 1.15e-13 \*\*\*  
## temp -9075.5 4118.8 -2.203 0.027985 \*   
## atemp 17849.8 4648.6 3.840 0.000138 \*\*\*  
## hum -3320.1 440.0 -7.546 1.91e-13 \*\*\*  
## windspeed -3508.2 848.8 -4.133 4.14e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1425 on 543 degrees of freedom  
## Multiple R-squared: 0.4709, Adjusted R-squared: 0.467   
## F-statistic: 120.8 on 4 and 543 DF, p-value: < 2.2e-16

## temp hum windspeed   
## 1.029784 1.064354 1.077038

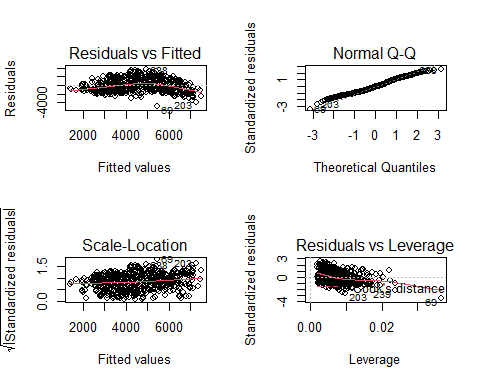
##   
## Call:  
## lm(formula = cnt ~ temp + hum + windspeed, data = TrainingData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4828.4 -1137.5 -73.3 1076.9 3628.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3979.6 393.6 10.111 < 2e-16 \*\*\*  
## temp 6686.6 342.7 19.509 < 2e-16 \*\*\*  
## hum -3128.4 442.7 -7.067 4.86e-12 \*\*\*  
## windspeed -4314.9 832.7 -5.182 3.10e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1443 on 544 degrees of freedom  
## Multiple R-squared: 0.4565, Adjusted R-squared: 0.4535   
## F-statistic: 152.3 on 3 and 544 DF, p-value: < 2.2e-16



## instant dteday season yr mnth holiday weekday workingday weathersit  
## 319 319 2011-11-15 4 0 11 0 2 1 2  
## temp atemp hum windspeed casual registered cnt  
## 319 0.53 0.507579 0.68875 0.199633 449 3746 4195

## Warning in instant != c(93, 271, 319): longer object length is not a multiple of  
## shorter object length

##   
## Call:  
## lm(formula = cnt ~ temp + hum + windspeed, data = x)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4817.4 -1147.2 -70.2 1072.6 3628.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3967.3 394.2 10.065 < 2e-16 \*\*\*  
## temp 6692.6 343.0 19.512 < 2e-16 \*\*\*  
## hum -3109.3 443.7 -7.008 7.18e-12 \*\*\*  
## windspeed -4318.4 833.1 -5.183 3.08e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1444 on 543 degrees of freedom  
## Multiple R-squared: 0.4569, Adjusted R-squared: 0.4539   
## F-statistic: 152.3 on 3 and 543 DF, p-value: < 2.2e-16

 ## Random forest approach

#Below is the random forest approach   
  
  
#Random Forest Approach  
  
  
  
  
#Look into if random forest is an option -- thought of as a logistic regression -- might not work in this situation unless binning of outcome  
#For Random Forest -- try binning the counts into discreet categories  
  
  
# Load the library  
library(randomForest)

## Warning: package 'randomForest' was built under R version 4.0.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

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

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

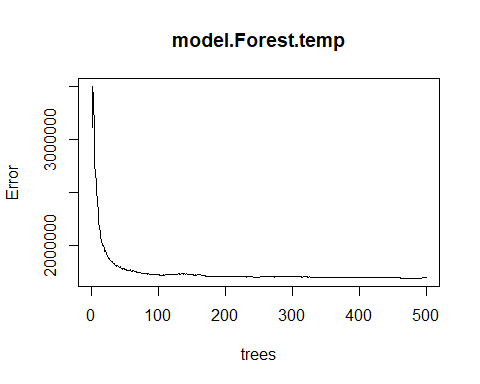
library(rsample)

## Warning: package 'rsample' was built under R version 4.0.5

set.seed(12309)  
  
model.Forest.temp <- randomForest(cnt ~ temp + hum + windspeed, data = TrainingData)  
  
print(model.Forest.temp)

##   
## Call:  
## randomForest(formula = cnt ~ temp + hum + windspeed, data = TrainingData)   
## Type of random forest: regression  
## Number of trees: 500  
## No. of variables tried at each split: 1  
##   
## Mean of squared residuals: 1699591  
## % Var explained: 55.31

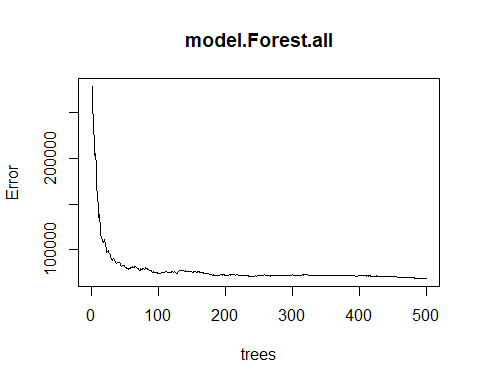
plot(model.Forest.temp)



model.Forest.all <- randomForest(  
 formula = cnt ~ .,  
 data = TrainingData)  
print(model.Forest.all)

##   
## Call:  
## randomForest(formula = cnt ~ ., data = TrainingData)   
## Type of random forest: regression  
## Number of trees: 500  
## No. of variables tried at each split: 5  
##   
## Mean of squared residuals: 69111.4  
## % Var explained: 98.18

#98.22% of variance is explained by the model (concern with overfitting)  
#500 trees is defaul; more trees means more models to average  
#During each split a subset of predictors are used to generate the model; in this case we are using the defualt  
#Default number of trees is determined by features/3 (15/3)  
#Mean of Squared residuals = 67841.89;   
plot(model.Forest.all)



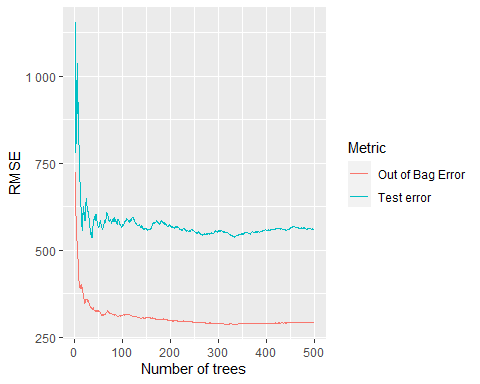
model.Forest.all.ER <- model.Forest.all$mse #Metric of model perofrmance; lower value is better goal is to minimize without overfitting  
which.min(model.Forest.all.ER) #Yield value of 496 -- number of trees with the lowest MSE

## [1] 490

sqrt(model.Forest.all.ER[which.min(model.Forest.all.ER)]) #Yields 257.826

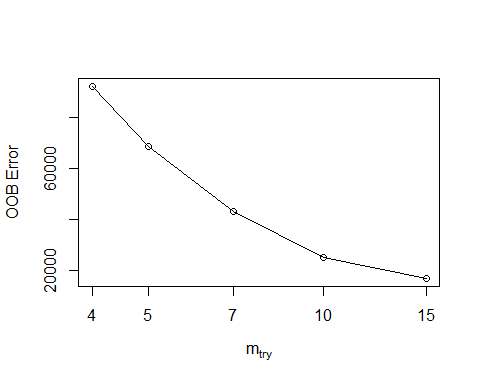
## [1] 262.6913

#So 496 trees provide an average count of 257.826 of bikes  
  
  
#Validation Set  
set.seed(1458)  
  
#For Splitting of TrainingData  
valid\_split <- initial\_split(TrainingData, 0.8)  
  
#Generating of two new data sets from TrainingDat  
TrainingData.V2 <- analysis(valid\_split)  
ValidationData <- assessment(valid\_split)  
  
#Separating out the predictor names from the outcome  
x\_test <- ValidationData[setdiff(names(ValidationData), "cnt")]  
y\_test <- ValidationData$cnt  
  
#Generate new model on validation set  
model.Forest.all.Validation <- randomForest(cnt ~ ., data = TrainingData.V2, x\_test, y\_test)  
  
#EXtract OOb (Out of Bag) and Validation Errors  
oob <- sqrt(model.Forest.all.Validation$mse) #Yields RMSE -- root mean squared residuals  
validation <- sqrt(model.Forest.all.Validation$test$mse)  
  
# compare error rates  
#OOB ERROR -- REMMEBER THAT DURINGH THE BOOTSTRAP ONLY A PORTION OF THE DATA IS USED, SO TEH OTHER PORTIONC AN BE USED TO VALIDATE EACH MODEL GENERATED -- TAHT SI WHAT IS BEING DONE HERE AND THEN PLOTTED WITH THE ACTUAL MODEL  
  
tibble::tibble(  
 `Out of Bag Error` = oob,  
 `Test error` = validation,  
 ntrees = 1:model.Forest.all.Validation$ntree  
) %>%  
 gather(Metric, RMSE, -ntrees) %>%  
 ggplot(aes(ntrees, RMSE, color = Metric)) +  
 geom\_line() +  
 scale\_y\_continuous(labels = scales::number) +  
 xlab("Number of trees")



#THE large difference between the two lines indicates a problem with overfitting -- This means we should try and tune the model  
features <- setdiff(names(ValidationData), "cnt")  
  
set.seed(132045)  
m2 <- randomForest::tuneRF(  
 x = TrainingData[features],  
 y = TrainingData$cnt,  
 ntreeTry = 500,  
 mtryStart = 5,  
 stepFactor = 1.5,  
 improve = .01,  
 trace = FALSE  
  
)

## -0.3404462 0.01   
## 0.3718082 0.01   
## 0.4131449 0.01   
## 0.3287689 0.01



model.Forest.all.Validation.mtry <- randomForest(cnt ~ ., data = TrainingData.V2, x\_test, y\_test, mtry = 15)  
  
print(model.Forest.all.Validation.mtry)

##   
## Call:  
## randomForest(formula = cnt ~ ., data = TrainingData.V2, x\_test, y\_test, mtry = 15)   
## Type of random forest: regression  
## Number of trees: 500  
## No. of variables tried at each split: 15  
##   
## Mean of squared residuals: 21278.35  
## % Var explained: 99.44  
## Test set MSE: 26017.2  
## % Var explained: 99.31

# #Some definition and EXPLANATIONS:  
# #OOB -- Out of Bag -- What is it?  
# ---  
# To answer we will examine what is a random forest. A random forest is a combination of two other methods:  
# \* Regression trees  
# \* Bagging  
#   
# Regression Trees  
# "Basic regression trees partition a data set into smaller groups and then fit a simple model (constant)  
# for each subgroup." --> meaning, a model is build after grouping similar observations; by doing so modelling becomes a simpler problem  
#   
# Problem is lack of stability and poor prediction (less predictors may translate to less explanation or capturing of unique cases)  
#   
# How is data partitioning done?   
# successive binary patitions; the constant to predict is based on the average response values for all observations that fall in that subgroup  
#   
# The partitions and genertaed in a top-down, greedy fahsion -- in other words correction to top level splits are not performed  
# So how was partitions made? How do we decide what predictor to split and how to split it?  
# The goal is to find the predictorsand split values that generate two partitioons with the lowest sums of squares error  
#   
# Pruning (cutting higher up) helps improve it amongs other tuning paramters  
# repeat  
#   
#   
#   
#   
# Bagging  
#   
# USed to remove problem of high variance from a simngle tree model. This is the building of multiple trees and averaging them  
# SEE REFERENCE FOR SPEICFIC STEPS  
#   
#   
# RANDOM FOREST IMPROVES ON ABOVE BY ADDRESSING HIGH CORRLEATION BETWEEN MULTIPLE MODELS PROBLEMS INTROIDUCED BY BAGGING  
# ACCOMPLISHED BY BOOTSTRAP -- EACH TREE IS GROW NO A BOOTSTRAP RESAMPLED DATA Seatbelts -- AND SPLIT-VARIABLE RNADOMIZATION --   
# NOT ALL VARIABLES ARE USED TO BUILD THE MODEL, NOT ALL USED FOR SPLITS A TYPICAL VALUE IS P/3 (5 FOR US). WHEN M = P THEN YOU ARE BVAGGGING  
  
   
   
  
  
#Run Some summary stats to find the best bins and define them. (i.e. high use, middle use, and low) <- alt third approach  
  
  
#Random Forest Regression   
#Random Forest Tutorial: https://www.geeksforgeeks.org/random-forest-approach-for-regression-in-r-programming/  
#https://uc-r.github.io/random\_forests  
#http://uc-r.github.io/regression\_trees