Assignment2.rmd

Jordan Whitaker

5/25/2020

options(tidyverse.quiet = TRUE)  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.6.3

## Warning: package 'ggplot2' was built under R version 3.6.3

## Warning: package 'tidyr' was built under R version 3.6.3

## Warning: package 'readr' was built under R version 3.6.3

## Warning: package 'dplyr' was built under R version 3.6.3

## Warning: package 'stringr' was built under R version 3.6.3

library(GGally)

## Warning: package 'GGally' was built under R version 3.6.3

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

##   
## Attaching package: 'GGally'

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

library(car)

## Warning: package 'car' was built under R version 3.6.3

## Loading required package: carData

##   
## Attaching package: 'car'

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

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

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.6.3

##   
## Attaching package: 'gridExtra'

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

library(MASS)

## Warning: package 'MASS' was built under R version 3.6.3

##   
## Attaching package: 'MASS'

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

bike = read\_csv("hour.csv")

## Parsed with column specification:  
## cols(  
## instant = col\_double(),  
## dteday = col\_date(format = ""),  
## season = col\_double(),  
## yr = col\_double(),  
## mnth = col\_double(),  
## hr = col\_double(),  
## holiday = col\_double(),  
## weekday = col\_double(),  
## workingday = col\_double(),  
## weathersit = col\_double(),  
## temp = col\_double(),  
## atemp = col\_double(),  
## hum = col\_double(),  
## windspeed = col\_double(),  
## casual = col\_double(),  
## registered = col\_double(),  
## count = col\_double()  
## )

View(bike)  
  
bike = bike %>% mutate(season = as\_factor(as.character(season))) %>%  
mutate(season = fct\_recode(season,  
"Spring" = "1",  
"Summer" = "2",  
"Fall" = "3",  
"Winter" = "4"))  
  
head(bike) #key reference to ensure code snippets below were successful

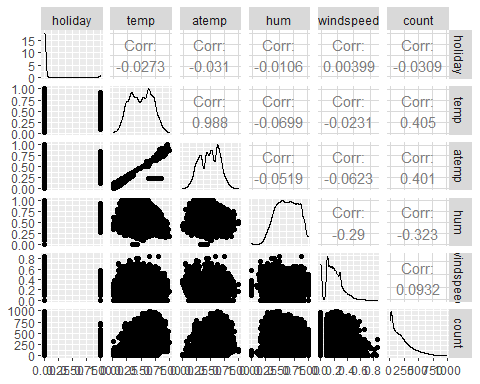
## # A tibble: 6 x 17  
## instant dteday season yr mnth hr holiday weekday workingday  
## <dbl> <date> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 2011-01-01 Spring 0 1 0 0 6 0  
## 2 2 2011-01-01 Spring 0 1 1 0 6 0  
## 3 3 2011-01-01 Spring 0 1 2 0 6 0  
## 4 4 2011-01-01 Spring 0 1 3 0 6 0  
## 5 5 2011-01-01 Spring 0 1 4 0 6 0  
## 6 6 2011-01-01 Spring 0 1 5 0 6 0  
## # ... with 8 more variables: weathersit <dbl>, temp <dbl>, atemp <dbl>,  
## # hum <dbl>, windspeed <dbl>, casual <dbl>, registered <dbl>, count <dbl>

bike = bike %>% mutate(yr = as\_factor(as.character(yr))) %>%  
mutate(yr = fct\_recode(yr))  
  
  
bike = bike %>% mutate(mnth = as\_factor(as.character(mnth))) %>%  
mutate(mnth = fct\_recode(mnth))  
  
bike = bike %>% mutate(hr = as\_factor(as.character(hr))) %>%  
mutate(hr = fct\_recode(hr))  
  
  
bike = bike %>% mutate(workingday = as\_factor(as.character(workingday))) %>%  
mutate(workingday = fct\_recode(workingday,  
"NotWorkingDay" = "0",  
"WorkingDay" = "1"))  
  
  
bike = bike %>% mutate(weathersit = as\_factor(as.character(weathersit))) %>%  
mutate(weathersit = fct\_recode(weathersit,  
"NoPrecip" = "1",  
"Misty" = "2",  
"LightPrecip" = "3",  
"HeavyPrecip" = "4"))  
  
  
bike = bike %>% mutate(weekday = as\_factor(as.character(weekday))) %>%  
mutate(weekday = fct\_recode(weekday,  
"Monday" = "1",  
"Tuesday" = "2",  
"Wednesday" = "3",  
"Thursday" = "4",  
"Friday" = "5",  
"Saturday" = "6",  
"Sunday"= "0"))

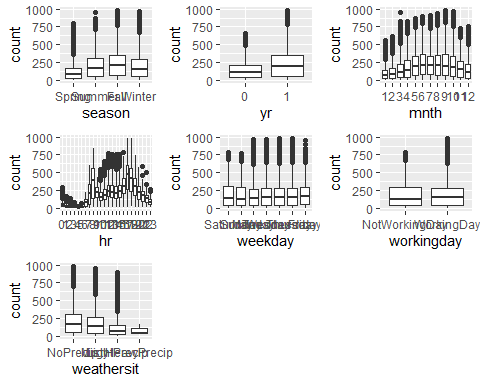
Task 1: When making a multi-linear regression model, if the data being used is categorical in nature, but is represented in the data as either binary or numerical, a conversion to a factor is required so that the model can interpret the fields and apply them to the model correctly.

bike\_quant = bike %>% dplyr::select(holiday, temp, atemp, hum, windspeed, count)  
  
ggpairs(bike\_quant, label=TRUE)

## Warning in warn\_if\_args\_exist(list(...)): Extra arguments: "label" are being  
## ignored. If these are meant to be aesthetics, submit them using the 'mapping'  
## variable within ggpairs with ggplot2::aes or ggplot2::aes\_string.

 Task 2: The temp and atemp variables are most correlated to the count variable. Around .40

season\_plt <- ggplot(bike, aes(x=season, y=count)) +   
 geom\_boxplot()  
  
yr\_plt <- ggplot(bike, aes(x=yr, y=count)) +   
 geom\_boxplot()   
  
mnth\_plt <- ggplot(bike, aes(x=mnth, y=count)) +   
 geom\_boxplot()  
  
hr\_plt <- ggplot(bike, aes(x=hr, y=count)) +   
 geom\_boxplot()  
  
weekdy\_plt <- ggplot(bike, aes(x=weekday, y=count)) +   
 geom\_boxplot()  
  
workingdy\_plt <- ggplot(bike, aes(x=workingday, y=count)) +   
 geom\_boxplot()  
  
weather\_plt <- ggplot(bike, aes(x=weathersit, y=count)) +   
 geom\_boxplot()  
  
  
grid.arrange(season\_plt, yr\_plt, mnth\_plt, hr\_plt, weekdy\_plt, workingdy\_plt, weather\_plt, nrow= 3 )



allmod = lm(count ~., bike)  
summary(allmod)

##   
## Call:  
## lm(formula = count ~ ., data = bike)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.119e-11 -7.000e-14 1.000e-14 8.000e-14 3.855e-10   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.016e-09 7.434e-10 5.402e+00 6.67e-08 \*\*\*  
## instant 1.116e-14 2.080e-15 5.368e+00 8.07e-08 \*\*\*  
## dteday -2.679e-13 4.964e-14 -5.396e+00 6.88e-08 \*\*\*  
## seasonSummer 1.098e-12 1.459e-13 7.524e+00 5.56e-14 \*\*\*  
## seasonFall -1.434e-14 1.725e-13 -8.300e-02 0.933749   
## seasonWinter -4.265e-13 1.474e-13 -2.892e+00 0.003828 \*\*   
## yr1 1.085e-12 9.690e-13 1.120e+00 0.262888   
## mnth2 -1.282e-13 1.461e-13 -8.780e-01 0.380074   
## mnth3 5.138e-13 2.095e-13 2.452e+00 0.014211 \*   
## mnth4 2.560e-13 3.077e-13 8.320e-01 0.405524   
## mnth5 2.708e-13 3.770e-13 7.180e-01 0.472582   
## mnth6 1.003e-12 4.476e-13 2.242e+00 0.024984 \*   
## mnth7 1.217e-12 5.305e-13 2.295e+00 0.021761 \*   
## mnth8 1.264e-12 6.016e-13 2.101e+00 0.035693 \*   
## mnth9 1.310e-12 6.708e-13 1.953e+00 0.050832 .   
## mnth10 1.698e-12 7.485e-13 2.268e+00 0.023329 \*   
## mnth11 1.555e-12 8.279e-13 1.879e+00 0.060322 .   
## mnth12 1.458e-12 8.980e-13 1.623e+00 0.104541   
## hr1 -8.754e-13 1.602e-13 -5.465e+00 4.69e-08 \*\*\*  
## hr2 -9.176e-13 1.609e-13 -5.704e+00 1.19e-08 \*\*\*  
## hr3 -1.390e-12 1.622e-13 -8.569e+00 < 2e-16 \*\*\*  
## hr4 -1.859e-13 1.625e-13 -1.144e+00 0.252643   
## hr5 -1.296e-12 1.614e-13 -8.031e+00 1.03e-15 \*\*\*  
## hr6 -2.076e-12 1.614e-13 -1.287e+01 < 2e-16 \*\*\*  
## hr7 -1.643e-12 1.675e-13 -9.809e+00 < 2e-16 \*\*\*  
## hr8 -1.731e-12 1.811e-13 -9.557e+00 < 2e-16 \*\*\*  
## hr9 -1.441e-12 1.657e-13 -8.696e+00 < 2e-16 \*\*\*  
## hr10 -6.354e-13 1.641e-13 -3.872e+00 0.000108 \*\*\*  
## hr11 -1.009e-12 1.667e-13 -6.052e+00 1.46e-09 \*\*\*  
## hr12 -9.885e-13 1.700e-13 -5.814e+00 6.21e-09 \*\*\*  
## hr13 -1.094e-12 1.714e-13 -6.384e+00 1.77e-10 \*\*\*  
## hr14 -1.426e-12 1.723e-13 -8.275e+00 < 2e-16 \*\*\*  
## hr15 -1.055e-12 1.731e-13 -6.096e+00 1.11e-09 \*\*\*  
## hr16 -1.520e-12 1.757e-13 -8.650e+00 < 2e-16 \*\*\*  
## hr17 -1.202e-12 1.891e-13 -6.356e+00 2.13e-10 \*\*\*  
## hr18 -2.071e-12 1.860e-13 -1.113e+01 < 2e-16 \*\*\*  
## hr19 -2.269e-12 1.750e-13 -1.296e+01 < 2e-16 \*\*\*  
## hr20 -1.578e-12 1.698e-13 -9.291e+00 < 2e-16 \*\*\*  
## hr21 -1.252e-12 1.677e-13 -7.469e+00 8.49e-14 \*\*\*  
## hr22 -6.444e-13 1.669e-13 -3.860e+00 0.000114 \*\*\*  
## hr23 -9.672e-13 1.669e-13 -5.796e+00 6.91e-09 \*\*\*  
## holiday -1.120e-14 1.484e-13 -7.500e-02 0.939854   
## weekdaySunday -3.783e-13 8.632e-14 -4.382e+00 1.18e-05 \*\*\*  
## weekdayMonday -5.527e-13 9.507e-14 -5.814e+00 6.22e-09 \*\*\*  
## weekdayTuesday -5.927e-13 9.429e-14 -6.286e+00 3.33e-10 \*\*\*  
## weekdayWednesday -7.078e-13 9.433e-14 -7.504e+00 6.48e-14 \*\*\*  
## weekdayThursday -7.316e-13 9.421e-14 -7.766e+00 8.58e-15 \*\*\*  
## weekdayFriday -8.390e-13 9.158e-14 -9.161e+00 < 2e-16 \*\*\*  
## workingdayWorkingDay NA NA NA NA   
## weathersitMisty -4.984e-14 5.760e-14 -8.650e-01 0.386917   
## weathersitLightPrecip -7.161e-13 9.811e-14 -7.299e+00 3.02e-13 \*\*\*  
## weathersitHeavyPrecip -6.944e-13 1.764e-12 -3.940e-01 0.693931   
## temp -1.827e-12 8.872e-13 -2.059e+00 0.039498 \*   
## atemp 5.007e-13 9.186e-13 5.450e-01 0.585709   
## hum -1.027e-13 1.683e-13 -6.100e-01 0.541660   
## windspeed -1.281e-13 2.116e-13 -6.050e-01 0.545068   
## casual 1.000e+00 7.920e-16 1.263e+15 < 2e-16 \*\*\*  
## registered 1.000e+00 2.931e-16 3.412e+15 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.048e-12 on 17322 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.099e+30 on 56 and 17322 DF, p-value: < 2.2e-16

emptymod = lm(count ~1, bike)  
summary(emptymod)

##   
## Call:  
## lm(formula = count ~ 1, data = bike)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -188.46 -149.46 -47.46 91.54 787.54   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 189.463 1.376 137.7 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 181.4 on 17378 degrees of freedom

#bike = bike %>% dplyr::select(-c(instant, dteday, registered, casual))  
  
forwardmod = stepAIC(emptymod, direction='forward', scope=list(upper=allmod,lower=emptymod),  
 trace = TRUE)

## Start: AIC=180764.7  
## count ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + registered 1 540358754 31402837 130336  
## + hr 23 286734681 285026910 168713  
## + casual 1 275828803 295932788 169321  
## + temp 1 93677759 478083832 177657  
## + atemp 1 91907421 479854170 177721  
## + hum 1 59618351 512143240 178853  
## + instant 1 44308491 527453100 179365  
## + dteday 1 44109491 527652100 179371  
## + mnth 11 42909976 528851615 179431  
## + season 3 37729358 534032233 179584  
## + yr 1 35876722 535884870 179641  
## + weathersit 3 12285030 559476561 180393  
## + windspeed 1 4970060 566791531 180615  
## + holiday 1 546889 571214702 180750  
## + workingday 1 524387 571237204 180751  
## + weekday 6 687929 571073662 180756  
## <none> 571761591 180765  
##   
## Step: AIC=130335.9  
## count ~ registered  
##   
## Df Sum of Sq RSS AIC  
## + casual 1 31402837 0 -922679  
## + hr 23 7374563 24028274 125730  
## + workingday 1 5857729 25545108 126750  
## + weekday 6 5711365 25691472 126859  
## + temp 1 3995235 27407602 127973  
## + atemp 1 3874353 27528484 128050  
## + mnth 11 2172469 29230368 129112  
## + hum 1 1980723 29422114 129206  
## + season 3 1642368 29760469 129408  
## + weathersit 3 361356 31041481 130141  
## + holiday 1 130653 31272184 130265  
## + windspeed 1 100387 31302450 130282  
## + instant 1 10893 31391944 130332  
## + dteday 1 10647 31392190 130332  
## + yr 1 9178 31393659 130333  
## <none> 31402837 130336  
##   
## Step: AIC=-922678.8  
## count ~ registered + casual

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Df Sum of Sq RSS AIC  
## + workingday 1 2.1051e-23 1.5221e-19 -922679  
## + dteday 1 1.8546e-23 1.5221e-19 -922679  
## + instant 1 1.8258e-23 1.5221e-19 -922679  
## <none> 1.5223e-19 -922679  
## + windspeed 1 1.7239e-23 1.5221e-19 -922679  
## + temp 1 9.6350e-24 1.5222e-19 -922678  
## + atemp 1 5.4700e-24 1.5222e-19 -922677  
## + yr 1 4.7820e-24 1.5222e-19 -922677  
## + hum 1 4.1230e-24 1.5222e-19 -922677  
## + holiday 1 3.1100e-25 1.5222e-19 -922677  
## + season 3 2.0362e-23 1.5221e-19 -922675  
## + weathersit 3 5.5880e-24 1.5222e-19 -922673  
## + weekday 6 5.2639e-23 1.5217e-19 -922673  
## + mnth 11 8.4039e-23 1.5214e-19 -922666  
## + hr 23 1.9846e-22 1.5203e-19 -922656  
##   
## Step: AIC=-922679.2  
## count ~ registered + casual + workingday

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Df Sum of Sq RSS AIC  
## + dteday 1 2.0158e-23 1.5218e-19 -922680  
## + instant 1 1.9860e-23 1.5219e-19 -922680  
## <none> 1.5221e-19 -922679  
## + windspeed 1 1.7187e-23 1.5219e-19 -922679  
## + yr 1 5.4900e-24 1.5220e-19 -922678  
## + temp 1 5.0560e-24 1.5220e-19 -922678  
## + hum 1 3.0740e-24 1.5220e-19 -922678  
## + holiday 1 3.0140e-24 1.5220e-19 -922678  
## + atemp 1 2.2020e-24 1.5220e-19 -922677  
## + season 3 1.7452e-23 1.5219e-19 -922675  
## + weathersit 3 5.2990e-24 1.5220e-19 -922674  
## + weekday 6 3.1638e-23 1.5217e-19 -922671  
## + mnth 11 7.8979e-23 1.5213e-19 -922666  
## + hr 23 2.0033e-22 1.5200e-19 -922656  
##   
## Step: AIC=-922679.5  
## count ~ registered + casual + workingday + dteday

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Df Sum of Sq RSS AIC  
## + instant 1 1.1036e-22 1.5207e-19 -922690  
## + windspeed 1 2.1487e-23 1.5216e-19 -922680  
## <none> 1.5218e-19 -922680  
## + yr 1 8.3670e-24 1.5218e-19 -922678  
## + hum 1 4.9070e-24 1.5218e-19 -922678  
## + temp 1 4.0880e-24 1.5218e-19 -922678  
## + holiday 1 2.7240e-24 1.5218e-19 -922678  
## + atemp 1 1.5510e-24 1.5218e-19 -922678  
## + season 3 1.1185e-23 1.5217e-19 -922675  
## + weathersit 3 4.7300e-24 1.5218e-19 -922674  
## + weekday 6 3.1040e-23 1.5215e-19 -922671  
## + mnth 11 6.7539e-23 1.5212e-19 -922665  
## + hr 23 2.0964e-22 1.5198e-19 -922657  
##   
## Step: AIC=-922690.2  
## count ~ registered + casual + workingday + dteday + instant

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Df Sum of Sq RSS AIC  
## + windspeed 1 2.3512e-23 1.5205e-19 -922691  
## <none> 1.5207e-19 -922690  
## + hum 1 1.3827e-23 1.5206e-19 -922690  
## + holiday 1 2.6390e-24 1.5207e-19 -922688  
## + temp 1 2.0700e-24 1.5207e-19 -922688  
## + atemp 1 2.9600e-25 1.5207e-19 -922688  
## + yr 1 1.5900e-25 1.5207e-19 -922688  
## + weathersit 3 3.9560e-24 1.5207e-19 -922685  
## + season 3 3.1000e-24 1.5207e-19 -922685  
## + weekday 6 3.1485e-23 1.5204e-19 -922682  
## + hr 23 2.8055e-22 1.5179e-19 -922676  
## + mnth 11 2.3387e-23 1.5205e-19 -922671  
##   
## Step: AIC=-922690.8  
## count ~ registered + casual + workingday + dteday + instant +   
## windspeed

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Df Sum of Sq RSS AIC  
## <none> 1.5205e-19 -922691  
## + hum 1 6.3300e-24 1.5204e-19 -922690  
## + temp 1 3.2790e-24 1.5205e-19 -922689  
## + holiday 1 2.5320e-24 1.5205e-19 -922689  
## + atemp 1 1.2880e-24 1.5205e-19 -922689  
## + yr 1 8.6700e-25 1.5205e-19 -922689  
## + season 3 5.2190e-24 1.5205e-19 -922685  
## + weathersit 3 3.6250e-24 1.5205e-19 -922685  
## + weekday 6 3.2150e-23 1.5202e-19 -922683  
## + hr 23 2.7339e-22 1.5178e-19 -922676  
## + mnth 11 2.6318e-23 1.5202e-19 -922672

summary(forwardmod)

##   
## Call:  
## lm(formula = count ~ registered + casual + workingday + dteday +   
## instant + windspeed, data = bike)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.440e-11 -6.000e-14 -1.000e-14 4.000e-14 3.861e-10   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.890e-09 5.258e-10 3.595e+00 0.000325 \*\*\*  
## registered 1.000e+00 1.911e-16 5.234e+15 < 2e-16 \*\*\*  
## casual 1.000e+00 5.869e-16 1.704e+15 < 2e-16 \*\*\*  
## workingdayWorkingDay -8.353e-14 5.400e-14 -1.547e+00 0.121918   
## dteday -1.260e-13 3.511e-14 -3.589e+00 0.000333 \*\*\*  
## instant 5.269e-15 1.470e-15 3.583e+00 0.000340 \*\*\*  
## windspeed -3.039e-13 1.854e-13 -1.639e+00 0.101231   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.958e-12 on 17372 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.089e+31 on 6 and 17372 DF, p-value: < 2.2e-16

vif(forwardmod)

## registered casual workingday dteday instant windspeed   
## 1.660553e+00 1.662572e+00 1.254262e+00 1.079987e+05 1.080380e+05 1.021416e+00

Task 4: These are the variables used in the forward model: hr + atemp + yr + weathersit + season + mnth + hum + weekday + holiday + windspeed + temp

This model has an adjusted R-Squared of ~.69 which is an indicator of a good model. Based on the VIF metric, there are instances of multi-collinearity between atemp and temp. Common sense also has me thinking there is a high degree of multi-collinearity between the season and month variables as well as the working day and day variables as these two sets are naturally tightly coupled.

backmod = stepAIC(allmod, direction = "backward", trace = TRUE)

## Start: AIC=-921608.8  
## count ~ instant + dteday + season + yr + mnth + hr + holiday +   
## weekday + workingday + weathersit + temp + atemp + hum +   
## windspeed + casual + registered

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Warning in stepAIC(allmod, direction = "backward", trace = TRUE): 0 df terms are  
## changing AIC

## Df Sum of Sq RSS AIC  
## - instant 1 0 0 -921730  
## - windspeed 1 0 0 -921704  
## - atemp 1 0 0 -921672  
## <none> 0 -921609  
## - temp 1 0 0 -921548  
## - season 3 0 0 -921472  
## - dteday 1 0 0 -921458  
## - mnth 11 0 0 -921282  
## - hum 1 0 0 -921250  
## - weekday 5 0 0 -920949  
## - yr 1 0 0 -920866  
## - weathersit 3 0 0 -920795  
## - hr 23 0 0 -917981  
## - casual 1 14805609 14805609 117377  
## - registered 1 108129815 108129815 151932  
##   
## Step: AIC=-921730.3  
## count ~ dteday + season + yr + mnth + hr + holiday + weekday +   
## workingday + weathersit + temp + atemp + hum + windspeed +   
## casual + registered

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - season 3 0 0 -922431  
## - hr 23 0 0 -922309  
## - yr 1 0 0 -922147  
## - temp 1 0 0 -922094  
## - windspeed 1 0 0 -922089  
## - mnth 11 0 0 -922028  
## - hum 1 0 0 -921999  
## - atemp 1 0 0 -921889  
## - weathersit 3 0 0 -921776  
## <none> 0 -921730  
## - weekday 5 0 0 -921547  
## - dteday 1 0 0 -917612  
## - casual 1 14832540 14832540 117406  
## - registered 1 108187774 108187774 151939  
##   
## Step: AIC=-922431.4  
## count ~ dteday + yr + mnth + hr + holiday + weekday + workingday +   
## weathersit + temp + atemp + hum + windspeed + casual + registered

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## <none> 0 -922431  
## - yr 1 0 0 -922300  
## - weathersit 3 0 0 -922285  
## - weekday 5 0 0 -922115  
## - temp 1 0 0 -922058  
## - hum 1 0 0 -921987  
## - atemp 1 0 0 -921679  
## - windspeed 1 0 0 -921225  
## - mnth 11 0 0 -921205  
## - dteday 1 0 0 -920676  
## - hr 23 0 0 -919887  
## - casual 1 14883576 14883576 117460  
## - registered 1 109963085 109963085 152216

summary(backmod)

##   
## Call:  
## lm(formula = count ~ dteday + yr + mnth + hr + holiday + weekday +   
## workingday + weathersit + temp + atemp + hum + windspeed +   
## casual + registered, data = bike)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.903e-11 -7.000e-14 1.000e-14 8.000e-14 3.861e-10   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.506e-11 3.864e-11 1.166e+00 0.243551   
## dteday -2.679e-15 2.578e-15 -1.039e+00 0.298673   
## yr1 8.004e-13 9.449e-13 8.470e-01 0.396922   
## mnth2 1.049e-13 1.379e-13 7.610e-01 0.446659   
## mnth3 1.110e-13 1.933e-13 5.740e-01 0.565705   
## mnth4 3.927e-13 2.641e-13 1.487e+00 0.136965   
## mnth5 6.597e-13 3.413e-13 1.933e+00 0.053231 .   
## mnth6 7.807e-13 4.192e-13 1.862e+00 0.062569 .   
## mnth7 1.288e-12 4.965e-13 2.594e+00 0.009498 \*\*   
## mnth8 1.609e-12 5.687e-13 2.830e+00 0.004662 \*\*   
## mnth9 8.013e-13 6.413e-13 1.249e+00 0.211539   
## mnth10 9.397e-13 7.145e-13 1.315e+00 0.188510   
## mnth11 9.050e-13 7.919e-13 1.143e+00 0.253131   
## mnth12 8.559e-13 8.693e-13 9.850e-01 0.324862   
## hr1 -6.258e-13 1.564e-13 -4.000e+00 6.35e-05 \*\*\*  
## hr2 -4.262e-13 1.571e-13 -2.714e+00 0.006661 \*\*   
## hr3 -6.752e-13 1.583e-13 -4.266e+00 2.00e-05 \*\*\*  
## hr4 -9.898e-13 1.585e-13 -6.245e+00 4.32e-10 \*\*\*  
## hr5 2.120e-13 1.573e-13 1.348e+00 0.177626   
## hr6 -9.907e-13 1.572e-13 -6.304e+00 2.97e-10 \*\*\*  
## hr7 -1.341e-12 1.630e-13 -8.230e+00 < 2e-16 \*\*\*  
## hr8 -1.419e-12 1.761e-13 -8.063e+00 7.95e-16 \*\*\*  
## hr9 -1.136e-12 1.609e-13 -7.060e+00 1.73e-12 \*\*\*  
## hr10 -9.534e-13 1.593e-13 -5.987e+00 2.18e-09 \*\*\*  
## hr11 -1.185e-12 1.617e-13 -7.326e+00 2.47e-13 \*\*\*  
## hr12 -1.182e-12 1.647e-13 -7.175e+00 7.51e-13 \*\*\*  
## hr13 -1.270e-12 1.658e-13 -7.658e+00 1.98e-14 \*\*\*  
## hr14 -1.158e-12 1.664e-13 -6.957e+00 3.59e-12 \*\*\*  
## hr15 -1.052e-12 1.669e-13 -6.304e+00 2.98e-10 \*\*\*  
## hr16 -1.012e-12 1.691e-13 -5.984e+00 2.22e-09 \*\*\*  
## hr17 -2.569e-12 1.820e-13 -1.411e+01 < 2e-16 \*\*\*  
## hr18 -1.370e-12 1.785e-13 -7.675e+00 1.74e-14 \*\*\*  
## hr19 -1.024e-12 1.671e-13 -6.129e+00 9.05e-10 \*\*\*  
## hr20 -1.606e-12 1.613e-13 -9.954e+00 < 2e-16 \*\*\*  
## hr21 -1.209e-12 1.585e-13 -7.625e+00 2.56e-14 \*\*\*  
## hr22 -1.286e-12 1.571e-13 -8.186e+00 2.88e-16 \*\*\*  
## hr23 -6.345e-13 1.564e-13 -4.058e+00 4.98e-05 \*\*\*  
## holiday 1.270e-13 1.446e-13 8.780e-01 0.379846   
## weekdaySunday -3.752e-13 8.428e-14 -4.452e+00 8.56e-06 \*\*\*  
## weekdayMonday -2.760e-13 9.283e-14 -2.973e+00 0.002954 \*\*   
## weekdayTuesday -3.263e-13 9.209e-14 -3.544e+00 0.000396 \*\*\*  
## weekdayWednesday -2.807e-13 9.208e-14 -3.048e+00 0.002306 \*\*   
## weekdayThursday -3.569e-13 9.199e-14 -3.879e+00 0.000105 \*\*\*  
## weekdayFriday -4.765e-13 8.940e-14 -5.329e+00 9.97e-08 \*\*\*  
## workingdayWorkingDay NA NA NA NA   
## weathersitMisty 2.620e-14 5.625e-14 4.660e-01 0.641408   
## weathersitLightPrecip 2.822e-13 9.581e-14 2.945e+00 0.003230 \*\*   
## weathersitHeavyPrecip 1.888e-13 1.723e-12 1.100e-01 0.912768   
## temp -2.699e-12 8.648e-13 -3.121e+00 0.001803 \*\*   
## atemp 1.160e-12 8.970e-13 1.293e+00 0.195960   
## hum -4.279e-13 1.638e-13 -2.612e+00 0.009005 \*\*   
## windspeed -1.495e-13 2.064e-13 -7.240e-01 0.468827   
## casual 1.000e+00 7.716e-16 1.296e+15 < 2e-16 \*\*\*  
## registered 1.000e+00 2.839e-16 3.523e+15 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.977e-12 on 17326 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.241e+30 on 52 and 17326 DF, p-value: < 2.2e-16

Task 5: The backward stepwise model is the same as the forward stepwise model.

Task 6: After analyzing the allmod model, my assumption is that the weekday variable is the culprit that is kicking out/overriding the workingday variable. This is because weekday is most likely the driver to encode what a working day is or is not.

Task 7: Overall, the model does appropriately describe a relationship between the predictors and the response variable. With an adjusted R-squred of nearly .69 I think this model is pretty strong. However, as noted above, there are some tightly coupled variables that should be pruned and of course the model could be tweaked a bit further. But the first iteration is a good one.