## Multiple Linear Regression and Special Issues Assignment

# Lea Shipley

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

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

library(GGally)

## Warning: package 'GGally' was built under R version 3.6.2

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

##   
## Attaching package: 'GGally'

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

library(MASS)

##   
## Attaching package: 'MASS'

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

library(car)

## Warning: package 'car' was built under R version 3.6.2

## 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)

##   
## Attaching package: 'gridExtra'

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

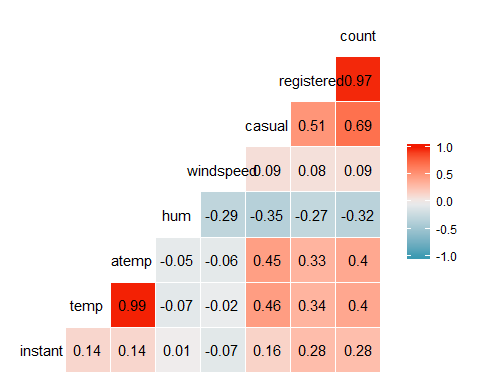
bike = read.csv("hour.csv")  
  
bike = bike %>% mutate(season = as\_factor(as.character(season))) %>%  
mutate(season = fct\_recode(season,  
"Spring" = "1",  
"Summer" = "2",  
"Fall" = "3",  
"Winter" = "4"))   
  
bike = bike %>% mutate(yr = as\_factor(as.numeric(yr))) %>%  
 mutate(mnth = as\_factor(as.numeric(mnth))) %>%  
 mutate(hr = as\_factor(as.numeric(hr))) %>%  
 mutate(holiday = as\_factor(as.numeric(holiday))) %>%  
 mutate(holiday = fct\_recode(holiday, "NotHoliday" = "0", "Holiday" = "1")) %>%  
 mutate(workingday = as.factor(as.numeric(workingday))) %>%   
 mutate(workingday = fct\_recode(workingday, "NotWorkingDay" = "0", "WorkingDay" = "1")) %>%  
 mutate(weathersit = as.factor(as.numeric(weathersit))) %>%  
 mutate(weathersit = fct\_recode(weathersit, "NoPrecip" = "1", "Misty" = "2", "LightPrecip" = "3" , "HeavyPrecip" = "4")) %>%  
 mutate(weekday = as\_factor(as.numeric(weekday))) %>%  
 mutate(weekday = fct\_recode(weekday, "Monday" = "1", "Tuesday" = "2", "Wednesday" = "3", "Thursday" = "4", "Friday" = "5", "Saturday" = "6", "Sunday" = "0"))

If we leave “yr,” “mnth,” and “hr” as numeric, R will treat them as quantitative. Converting them to factors will convert them into qualitative so that we can use them as categorical variables in predicting the count of bikes rented.

**Task 2**

ggcorr(bike, label = "TRUE", label\_round = 2)

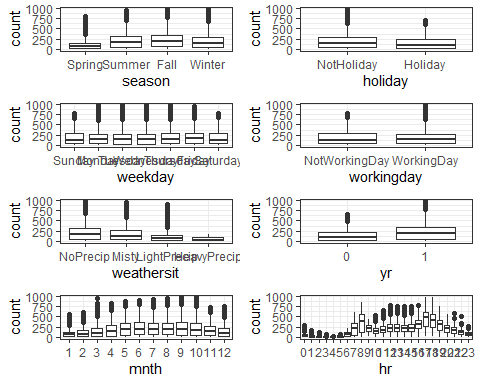
## Warning in ggcorr(bike, label = "TRUE", label\_round = 2): data in column(s)  
## 'dteday', 'season', 'yr', 'mnth', 'hr', 'holiday', 'weekday', 'workingday',  
## 'weathersit' are not numeric and were ignored



The quantitative variables “atemp” and “temp” appear to best correlated with “count.”

**Task 3**

p1 = ggplot(bike, aes(x=season,y=count)) + geom\_boxplot() + theme\_bw()  
p2 = ggplot(bike, aes(x=holiday,y=count)) + geom\_boxplot() + theme\_bw()  
p3 = ggplot(bike, aes(x=weekday,y=count)) + geom\_boxplot() + theme\_bw()  
p4 = ggplot(bike, aes(x=workingday,y=count)) + geom\_boxplot() + theme\_bw()  
p5 = ggplot(bike, aes(x=weathersit,y=count)) + geom\_boxplot() + theme\_bw()  
p6 = ggplot(bike, aes(x=yr,y=count)) + geom\_boxplot() + theme\_bw()  
p7 = ggplot(bike, aes(x=mnth,y=count)) + geom\_boxplot() + theme\_bw()  
p8 = ggplot(bike, aes(x=hr,y=count)) + geom\_boxplot() + theme\_bw()  
  
grid.arrange(p1,p2,p3,p4,p5,p6,p7,p8, ncol = 2)



All of the categorical variables appear to affect count.

* Season- Summer and Fall with less cold temperatures have higher rentals. People are less likely to bike if weather is cold/chilly.
* Holiday- The higher rentals of “NotHoliday” could be attributed to people traveling out of town for the holiday, and less people commuting to work.
* Weekday-The higher count on weekdays is likely from people renting bikes to commute to/from work.
* Workingday - Like “weekday,” the higher count on working days is likely from people renting bikes to commute to/from work.
* Weathersit - People more likely to use different transportation if there is heavy rain.
* yr- Perhaps more popular in 2012 than in 2011 due to awareness of bike service increasing among residents of D.C.
* mnth- Rentals are higher in months that are not terribly cold in D.C.
* hr- Rentals are low late at night and early morning, and are higher at times when people are liking using them to commute to/from work, and also midday for lunch and/or errands.

**Task 4**

bike2 <- bike %>% dplyr::select(-c(instant, dteday, registered, casual))  
  
allmod = lm(count ~., bike2)   
summary(allmod)

##   
## Call:  
## lm(formula = count ~ ., data = bike2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -393.87 -60.66 -7.96 51.31 439.18   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -83.630 6.633 -12.608 < 2e-16 \*\*\*  
## seasonSummer 38.178 4.856 7.862 4.00e-15 \*\*\*  
## seasonFall 32.055 5.749 5.575 2.51e-08 \*\*\*  
## seasonWinter 67.994 4.882 13.928 < 2e-16 \*\*\*  
## yr1 85.431 1.563 54.658 < 2e-16 \*\*\*  
## mnth2 3.426 3.920 0.874 0.382185   
## mnth3 14.299 4.407 3.244 0.001179 \*\*   
## mnth4 6.230 6.548 0.951 0.341438   
## mnth5 20.657 7.007 2.948 0.003201 \*\*   
## mnth6 6.238 7.205 0.866 0.386617   
## mnth7 -13.269 8.082 -1.642 0.100645   
## mnth8 7.897 7.879 1.002 0.316222   
## mnth9 32.269 7.001 4.609 4.07e-06 \*\*\*  
## mnth10 15.843 6.483 2.444 0.014549 \*   
## mnth11 -9.840 6.238 -1.577 0.114744   
## mnth12 -6.256 4.954 -1.263 0.206718   
## hr1 -17.294 5.345 -3.236 0.001216 \*\*   
## hr2 -26.369 5.364 -4.916 8.91e-07 \*\*\*  
## hr3 -37.112 5.403 -6.869 6.67e-12 \*\*\*  
## hr4 -40.263 5.408 -7.445 1.01e-13 \*\*\*  
## hr5 -23.501 5.373 -4.374 1.23e-05 \*\*\*  
## hr6 35.393 5.359 6.605 4.10e-11 \*\*\*  
## hr7 170.418 5.348 31.864 < 2e-16 \*\*\*  
## hr8 310.801 5.342 58.183 < 2e-16 \*\*\*  
## hr9 163.101 5.347 30.501 < 2e-16 \*\*\*  
## hr10 108.444 5.370 20.196 < 2e-16 \*\*\*  
## hr11 133.843 5.409 24.742 < 2e-16 \*\*\*  
## hr12 173.142 5.456 31.735 < 2e-16 \*\*\*  
## hr13 168.102 5.494 30.600 < 2e-16 \*\*\*  
## hr14 152.249 5.525 27.558 < 2e-16 \*\*\*  
## hr15 161.707 5.535 29.213 < 2e-16 \*\*\*  
## hr16 223.834 5.524 40.522 < 2e-16 \*\*\*  
## hr17 377.535 5.491 68.750 < 2e-16 \*\*\*  
## hr18 345.587 5.455 63.350 < 2e-16 \*\*\*  
## hr19 236.919 5.404 43.841 < 2e-16 \*\*\*  
## hr20 157.293 5.375 29.266 < 2e-16 \*\*\*  
## hr21 107.840 5.353 20.147 < 2e-16 \*\*\*  
## hr22 70.907 5.343 13.272 < 2e-16 \*\*\*  
## hr23 32.112 5.338 6.015 1.83e-09 \*\*\*  
## holidayHoliday -26.228 4.881 -5.374 7.81e-08 \*\*\*  
## weekdayMonday 9.275 2.973 3.120 0.001812 \*\*   
## weekdayTuesday 10.849 2.904 3.736 0.000187 \*\*\*  
## weekdayWednesday 13.625 2.900 4.698 2.64e-06 \*\*\*  
## weekdayThursday 13.149 2.901 4.532 5.87e-06 \*\*\*  
## weekdayFriday 17.445 2.892 6.032 1.65e-09 \*\*\*  
## weekdaySaturday 16.089 2.878 5.591 2.30e-08 \*\*\*  
## workingdayWorkingDay NA NA NA NA   
## weathersitMisty -10.409 1.920 -5.421 6.00e-08 \*\*\*  
## weathersitLightPrecip -65.189 3.236 -20.145 < 2e-16 \*\*\*  
## weathersitHeavyPrecip -62.580 58.893 -1.063 0.287970   
## temp 116.384 29.513 3.943 8.06e-05 \*\*\*  
## atemp 127.975 30.624 4.179 2.94e-05 \*\*\*  
## hum -82.802 5.554 -14.909 < 2e-16 \*\*\*  
## windspeed -29.167 7.052 -4.136 3.55e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 101.7 on 17326 degrees of freedom  
## Multiple R-squared: 0.6864, Adjusted R-squared: 0.6854   
## F-statistic: 729.1 on 52 and 17326 DF, p-value: < 2.2e-16

emptymod = lm(count ~1, bike2)   
summary(emptymod)

##   
## Call:  
## lm(formula = count ~ 1, data = bike2)  
##   
## 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

#forward  
forwardmod = stepAIC(emptymod, scope=list(upper=allmod,lower=emptymod),  
 trace = TRUE)

## Start: AIC=180764.7  
## count ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + hr 23 286734681 285026910 168713  
## + temp 1 93677759 478083832 177657  
## + atemp 1 91907421 479854170 177721  
## + hum 1 59618351 512143240 178853  
## + 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=168712.5  
## count ~ hr  
##   
## Df Sum of Sq RSS AIC  
## + atemp 1 50518941 234507969 165324  
## + temp 1 50101685 234925225 165355  
## + mnth 11 44822160 240204750 165761  
## + season 3 39619754 245407156 166117  
## + yr 1 36875130 248151780 166307  
## + weathersit 3 13766672 271260238 167858  
## + hum 1 4924310 280102600 168412  
## + windspeed 1 1476211 283550699 168624  
## + holiday 1 561784 284465126 168680  
## + weekday 6 719530 284307380 168681  
## + workingday 1 485366 284541544 168685  
## <none> 285026910 168713  
## - hr 23 286734681 571761591 180765  
##   
## Step: AIC=165324  
## count ~ hr + atemp  
##   
## Df Sum of Sq RSS AIC  
## + yr 1 33463769 201044200 162650  
## + weathersit 3 9227265 225280704 164632  
## + hum 1 7008684 227499285 164799  
## + season 3 6580442 227927527 164835  
## + mnth 11 5854560 228653409 164907  
## + weekday 6 607638 233900331 165291  
## + holiday 1 274006 234233963 165306  
## + temp 1 152153 234355816 165315  
## + windspeed 1 120557 234387412 165317  
## + workingday 1 90170 234417799 165319  
## <none> 234507969 165324  
## - atemp 1 50518941 285026910 168713  
## - hr 23 245346201 479854170 177721  
##   
## Step: AIC=162650.2  
## count ~ hr + atemp + yr  
##   
## Df Sum of Sq RSS AIC  
## + weathersit 3 8408358 192635842 161914  
## + season 3 7190305 193853896 162023  
## + mnth 11 6486062 194558138 162102  
## + hum 1 4341837 196702363 162273  
## + weekday 6 641648 200402552 162607  
## + holiday 1 324763 200719438 162624  
## + windspeed 1 109311 200934889 162643  
## + workingday 1 106404 200937797 162643  
## + temp 1 91735 200952465 162644  
## <none> 201044200 162650  
## - yr 1 33463769 234507969 165324  
## - atemp 1 47107580 248151780 166307  
## - hr 23 247247710 448291911 176541  
##   
## Step: AIC=161913.7  
## count ~ hr + atemp + yr + weathersit  
##   
## Df Sum of Sq RSS AIC  
## + season 3 7771024 184864818 161204  
## + mnth 11 7464989 185170852 161249  
## + hum 1 805099 191830743 161843  
## + weekday 6 686172 191949670 161864  
## + holiday 1 413536 192222305 161878  
## + workingday 1 212428 192423414 161897  
## + temp 1 134482 192501360 161904  
## + windspeed 1 44407 192591435 161912  
## <none> 192635842 161914  
## - weathersit 3 8408358 201044200 162650  
## - yr 1 32644862 225280704 164632  
## - atemp 1 42889218 235525060 165405  
## - hr 23 249611395 442247237 176311  
##   
## Step: AIC=161204.1  
## count ~ hr + atemp + yr + weathersit + season  
##   
## Df Sum of Sq RSS AIC  
## + mnth 11 2051323 182813495 161032  
## + hum 1 1810161 183054657 161035  
## + weekday 6 704303 184160515 161150  
## + holiday 1 392702 184472116 161169  
## + temp 1 352584 184512234 161173  
## + workingday 1 214973 184649845 161186  
## <none> 184864818 161204  
## + windspeed 1 158 184864660 161206  
## - season 3 7771024 192635842 161914  
## - weathersit 3 8989078 193853896 162023  
## - atemp 1 12095973 196960791 162304  
## - yr 1 33305296 218170114 164081  
## - hr 23 244894689 429759507 175819  
##   
## Step: AIC=161032.2  
## count ~ hr + atemp + yr + weathersit + season + mnth  
##   
## Df Sum of Sq RSS AIC  
## + hum 1 2356411 180457084 160809  
## + weekday 6 692672 182120823 160978  
## + holiday 1 312321 182501174 161004  
## + temp 1 233052 182580443 161012  
## + workingday 1 203953 182609542 161015  
## <none> 182813495 161032  
## + windspeed 1 68 182813428 161034  
## - mnth 11 2051323 184864818 161204  
## - season 3 2357357 185170852 161249  
## - atemp 1 6141071 188954566 161604  
## - weathersit 3 9426109 192239605 161900  
## - yr 1 33450244 216263739 163950  
## - hr 23 239078958 421892454 175520  
##   
## Step: AIC=160808.7  
## count ~ hr + atemp + yr + weathersit + season + mnth + hum  
##   
## Df Sum of Sq RSS AIC  
## + weekday 6 581105 179875980 160765  
## + holiday 1 322997 180134087 160780  
## + workingday 1 194139 180262945 160792  
## + windspeed 1 114287 180342797 160800  
## + temp 1 100025 180357059 160801  
## <none> 180457084 160809  
## - hum 1 2356411 182813495 161032  
## - mnth 11 2597573 183054657 161035  
## - season 3 2570704 183027788 161049  
## - weathersit 3 4315093 184772178 161213  
## - atemp 1 6511986 186969071 161423  
## - yr 1 31244888 211701972 163582  
## - hr 23 199208665 379665749 173689  
##   
## Step: AIC=160764.7  
## count ~ hr + atemp + yr + weathersit + season + mnth + hum +   
## weekday  
##   
## Df Sum of Sq RSS AIC  
## + workingday 1 274717 179601263 160740  
## + holiday 1 274717 179601263 160740  
## + windspeed 1 112085 179763895 160756  
## + temp 1 77171 179798809 160759  
## <none> 179875980 160765  
## - weekday 6 581105 180457084 160809  
## - hum 1 2244844 182120823 160978  
## - mnth 11 2576962 182452941 160990  
## - season 3 2600222 182476201 161008  
## - weathersit 3 4422816 184298796 161181  
## - atemp 1 6413325 186289305 161372  
## - yr 1 31316082 211192062 163552  
## - hr 23 199421146 379297125 173684  
##   
## Step: AIC=160740.1  
## count ~ hr + atemp + yr + weathersit + season + mnth + hum +   
## weekday + workingday  
##   
## Df Sum of Sq RSS AIC  
## + windspeed 1 111562 179489701 160731  
## + temp 1 95460 179505803 160733  
## <none> 179601263 160740  
## - workingday 1 274717 179875980 160765  
## - weekday 6 661682 180262945 160792  
## - hum 1 2262012 181863275 160956  
## - mnth 11 2516670 182117933 160960  
## - season 3 2452566 182053829 160970  
## - weathersit 3 4455530 184056793 161160  
## - atemp 1 6487847 186089110 161355  
## - yr 1 31329410 210930673 163533  
## - hr 23 199272899 378874162 173667  
##   
## Step: AIC=160731.3  
## count ~ hr + atemp + yr + weathersit + season + mnth + hum +   
## weekday + workingday + windspeed  
##   
## Df Sum of Sq RSS AIC  
## + temp 1 160954 179328746 160718  
## <none> 179489701 160731  
## - windspeed 1 111562 179601263 160740  
## - workingday 1 274194 179763895 160756  
## - weekday 6 661738 180151438 160783  
## - mnth 11 2539680 182029381 160954  
## - season 3 2387332 181877033 160955  
## - hum 1 2373546 181863247 160958  
## - weathersit 3 4166478 183656179 161124  
## - atemp 1 6397019 185886720 161338  
## - yr 1 31213710 210703411 163516  
## - hr 23 198496857 377986557 173628  
##   
## Step: AIC=160717.7  
## count ~ hr + atemp + yr + weathersit + season + mnth + hum +   
## weekday + workingday + windspeed + temp  
##   
## Df Sum of Sq RSS AIC  
## <none> 179328746 160718  
## - temp 1 160954 179489701 160731  
## - windspeed 1 177057 179505803 160733  
## - atemp 1 180751 179509498 160733  
## - workingday 1 298893 179627639 160745  
## - weekday 6 664366 179993112 160770  
## - mnth 11 2426171 181754917 160929  
## - hum 1 2300667 181629413 160937  
## - season 3 2398467 181727213 160943  
## - weathersit 3 4208731 183537478 161115  
## - yr 1 30920851 210249597 163480  
## - hr 23 196741474 376070220 173542

summary(forwardmod)

##   
## Call:  
## lm(formula = count ~ hr + atemp + yr + weathersit + season +   
## mnth + hum + weekday + workingday + windspeed + temp, data = bike2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -393.87 -60.66 -7.96 51.31 439.18   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -83.630 6.633 -12.608 < 2e-16 \*\*\*  
## hr1 -17.294 5.345 -3.236 0.00122 \*\*   
## hr2 -26.369 5.364 -4.916 8.91e-07 \*\*\*  
## hr3 -37.112 5.403 -6.869 6.67e-12 \*\*\*  
## hr4 -40.263 5.408 -7.445 1.01e-13 \*\*\*  
## hr5 -23.501 5.373 -4.374 1.23e-05 \*\*\*  
## hr6 35.393 5.359 6.605 4.10e-11 \*\*\*  
## hr7 170.418 5.348 31.864 < 2e-16 \*\*\*  
## hr8 310.801 5.342 58.183 < 2e-16 \*\*\*  
## hr9 163.101 5.347 30.501 < 2e-16 \*\*\*  
## hr10 108.444 5.370 20.196 < 2e-16 \*\*\*  
## hr11 133.843 5.409 24.742 < 2e-16 \*\*\*  
## hr12 173.142 5.456 31.735 < 2e-16 \*\*\*  
## hr13 168.102 5.494 30.600 < 2e-16 \*\*\*  
## hr14 152.249 5.525 27.558 < 2e-16 \*\*\*  
## hr15 161.707 5.535 29.213 < 2e-16 \*\*\*  
## hr16 223.834 5.524 40.522 < 2e-16 \*\*\*  
## hr17 377.535 5.491 68.750 < 2e-16 \*\*\*  
## hr18 345.587 5.455 63.350 < 2e-16 \*\*\*  
## hr19 236.919 5.404 43.841 < 2e-16 \*\*\*  
## hr20 157.293 5.375 29.266 < 2e-16 \*\*\*  
## hr21 107.840 5.353 20.147 < 2e-16 \*\*\*  
## hr22 70.907 5.343 13.272 < 2e-16 \*\*\*  
## hr23 32.112 5.338 6.015 1.83e-09 \*\*\*  
## atemp 127.975 30.624 4.179 2.94e-05 \*\*\*  
## yr1 85.431 1.563 54.658 < 2e-16 \*\*\*  
## weathersitMisty -10.409 1.920 -5.421 6.00e-08 \*\*\*  
## weathersitLightPrecip -65.189 3.236 -20.145 < 2e-16 \*\*\*  
## weathersitHeavyPrecip -62.580 58.893 -1.063 0.28797   
## seasonSummer 38.178 4.856 7.862 4.00e-15 \*\*\*  
## seasonFall 32.055 5.749 5.575 2.51e-08 \*\*\*  
## seasonWinter 67.994 4.882 13.928 < 2e-16 \*\*\*  
## mnth2 3.426 3.920 0.874 0.38219   
## mnth3 14.299 4.407 3.244 0.00118 \*\*   
## mnth4 6.230 6.548 0.951 0.34144   
## mnth5 20.657 7.007 2.948 0.00320 \*\*   
## mnth6 6.238 7.205 0.866 0.38662   
## mnth7 -13.269 8.082 -1.642 0.10065   
## mnth8 7.897 7.879 1.002 0.31622   
## mnth9 32.269 7.001 4.609 4.07e-06 \*\*\*  
## mnth10 15.843 6.483 2.444 0.01455 \*   
## mnth11 -9.840 6.238 -1.577 0.11474   
## mnth12 -6.256 4.954 -1.263 0.20672   
## hum -82.802 5.554 -14.909 < 2e-16 \*\*\*  
## weekdayMonday -16.953 5.081 -3.337 0.00085 \*\*\*  
## weekdayTuesday -15.379 5.641 -2.726 0.00641 \*\*   
## weekdayWednesday -12.603 5.639 -2.235 0.02544 \*   
## weekdayThursday -13.079 5.600 -2.336 0.01952 \*   
## weekdayFriday -8.783 5.599 -1.569 0.11676   
## weekdaySaturday 16.089 2.878 5.591 2.30e-08 \*\*\*  
## workingdayWorkingDay 26.228 4.881 5.374 7.81e-08 \*\*\*  
## windspeed -29.167 7.052 -4.136 3.55e-05 \*\*\*  
## temp 116.384 29.513 3.943 8.06e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 101.7 on 17326 degrees of freedom  
## Multiple R-squared: 0.6864, Adjusted R-squared: 0.6854   
## F-statistic: 729.1 on 52 and 17326 DF, p-value: < 2.2e-16

vif(forwardmod)

## GVIF Df GVIF^(1/(2\*Df))  
## hr 1.749392 23 1.012232  
## atemp 46.501555 1 6.819205  
## yr 1.025491 1 1.012665  
## weathersit 1.394544 3 1.056993  
## season 164.047934 3 2.339709  
## mnth 320.419121 11 1.299859  
## hum 1.927632 1 1.388392  
## weekday 8.853571 6 1.199296  
## workingday 8.664184 1 2.943499  
## windspeed 1.249705 1 1.117902  
## temp 54.224994 1 7.363762

The forward model includes the variables “hr,” “atemp,” “yr,” “weathersit,” “season,” “mnth,” “hum,” “weekday,” “workingday,” “windspeed,” and “temp.” The model has a good R-squared value, and is statistcally significant with a p-value less than 0.05. The model does match intuition/common sense, as the variables in the model are likely to influence a person’s decision to rent a bike. The variables “atemp,” “season,” “mnth,” “weekday,” “workingday,” and “temp” all have VIF values greater than 5, indicating multicollinearity.

**Task 5**

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

## Start: AIC=160717.7  
## count ~ season + yr + mnth + hr + holiday + weekday + workingday +   
## weathersit + temp + atemp + hum + windspeed  
##   
##   
## Step: AIC=160717.7  
## count ~ season + yr + mnth + hr + holiday + weekday + weathersit +   
## temp + atemp + hum + windspeed  
##   
## Df Sum of Sq RSS AIC  
## <none> 179328746 160718  
## - temp 1 160954 179489701 160731  
## - windspeed 1 177057 179505803 160733  
## - atemp 1 180751 179509498 160733  
## - holiday 1 298893 179627639 160745  
## - weekday 6 498795 179827541 160754  
## - mnth 11 2426171 181754917 160929  
## - hum 1 2300667 181629413 160937  
## - season 3 2398467 181727213 160943  
## - weathersit 3 4208731 183537478 161115  
## - yr 1 30920851 210249597 163480  
## - hr 23 196741474 376070220 173542

summary(backmod)

##   
## Call:  
## lm(formula = count ~ season + yr + mnth + hr + holiday + weekday +   
## weathersit + temp + atemp + hum + windspeed, data = bike2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -393.87 -60.66 -7.96 51.31 439.18   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -83.630 6.633 -12.608 < 2e-16 \*\*\*  
## seasonSummer 38.178 4.856 7.862 4.00e-15 \*\*\*  
## seasonFall 32.055 5.749 5.575 2.51e-08 \*\*\*  
## seasonWinter 67.994 4.882 13.928 < 2e-16 \*\*\*  
## yr1 85.431 1.563 54.658 < 2e-16 \*\*\*  
## mnth2 3.426 3.920 0.874 0.382185   
## mnth3 14.299 4.407 3.244 0.001179 \*\*   
## mnth4 6.230 6.548 0.951 0.341438   
## mnth5 20.657 7.007 2.948 0.003201 \*\*   
## mnth6 6.238 7.205 0.866 0.386617   
## mnth7 -13.269 8.082 -1.642 0.100645   
## mnth8 7.897 7.879 1.002 0.316222   
## mnth9 32.269 7.001 4.609 4.07e-06 \*\*\*  
## mnth10 15.843 6.483 2.444 0.014549 \*   
## mnth11 -9.840 6.238 -1.577 0.114744   
## mnth12 -6.256 4.954 -1.263 0.206718   
## hr1 -17.294 5.345 -3.236 0.001216 \*\*   
## hr2 -26.369 5.364 -4.916 8.91e-07 \*\*\*  
## hr3 -37.112 5.403 -6.869 6.67e-12 \*\*\*  
## hr4 -40.263 5.408 -7.445 1.01e-13 \*\*\*  
## hr5 -23.501 5.373 -4.374 1.23e-05 \*\*\*  
## hr6 35.393 5.359 6.605 4.10e-11 \*\*\*  
## hr7 170.418 5.348 31.864 < 2e-16 \*\*\*  
## hr8 310.801 5.342 58.183 < 2e-16 \*\*\*  
## hr9 163.101 5.347 30.501 < 2e-16 \*\*\*  
## hr10 108.444 5.370 20.196 < 2e-16 \*\*\*  
## hr11 133.843 5.409 24.742 < 2e-16 \*\*\*  
## hr12 173.142 5.456 31.735 < 2e-16 \*\*\*  
## hr13 168.102 5.494 30.600 < 2e-16 \*\*\*  
## hr14 152.249 5.525 27.558 < 2e-16 \*\*\*  
## hr15 161.707 5.535 29.213 < 2e-16 \*\*\*  
## hr16 223.834 5.524 40.522 < 2e-16 \*\*\*  
## hr17 377.535 5.491 68.750 < 2e-16 \*\*\*  
## hr18 345.587 5.455 63.350 < 2e-16 \*\*\*  
## hr19 236.919 5.404 43.841 < 2e-16 \*\*\*  
## hr20 157.293 5.375 29.266 < 2e-16 \*\*\*  
## hr21 107.840 5.353 20.147 < 2e-16 \*\*\*  
## hr22 70.907 5.343 13.272 < 2e-16 \*\*\*  
## hr23 32.112 5.338 6.015 1.83e-09 \*\*\*  
## holidayHoliday -26.228 4.881 -5.374 7.81e-08 \*\*\*  
## weekdayMonday 9.275 2.973 3.120 0.001812 \*\*   
## weekdayTuesday 10.849 2.904 3.736 0.000187 \*\*\*  
## weekdayWednesday 13.625 2.900 4.698 2.64e-06 \*\*\*  
## weekdayThursday 13.149 2.901 4.532 5.87e-06 \*\*\*  
## weekdayFriday 17.445 2.892 6.032 1.65e-09 \*\*\*  
## weekdaySaturday 16.089 2.878 5.591 2.30e-08 \*\*\*  
## weathersitMisty -10.409 1.920 -5.421 6.00e-08 \*\*\*  
## weathersitLightPrecip -65.189 3.236 -20.145 < 2e-16 \*\*\*  
## weathersitHeavyPrecip -62.580 58.893 -1.063 0.287970   
## temp 116.384 29.513 3.943 8.06e-05 \*\*\*  
## atemp 127.975 30.624 4.179 2.94e-05 \*\*\*  
## hum -82.802 5.554 -14.909 < 2e-16 \*\*\*  
## windspeed -29.167 7.052 -4.136 3.55e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 101.7 on 17326 degrees of freedom  
## Multiple R-squared: 0.6864, Adjusted R-squared: 0.6854   
## F-statistic: 729.1 on 52 and 17326 DF, p-value: < 2.2e-16

The backward model is also good with the same R-squared value and p-value as the forward model.

The two models use the same variables with the exception of one. The backward model uses the variable “holiday,” whereas the forward model uses the variable “workingday.”

“Workingday” is represented in the model via “holiday” and “weekday,” as typically weekdays and non-holiday days would be classified as a working day.

I would recommend this model for use, as the p-value is statistically significant and the R-squared value is good. A cautionary factor would be the presence of multicollinearity between the independent variables, however intuition/common sense would still suggest that the model is good when considering the association of independent variables with the dependent variable for this dataset.