bike <- read\_csv("~/Ban502/Module 2/bike\_cleaned.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## instant = col\_double(),  
## dteday = col\_character(),  
## season = col\_character(),  
## mnth = col\_character(),  
## hr = col\_double(),  
## holiday = col\_character(),  
## weekday = col\_character(),  
## workingday = col\_character(),  
## weathersit = col\_character(),  
## temp = col\_double(),  
## atemp = col\_double(),  
## hum = col\_double(),  
## windspeed = col\_double(),  
## casual = col\_double(),  
## registered = col\_double(),  
## count = col\_double()  
## )

bike = bike %>% mutate(dteday = mdy(dteday))  
  
bike = bike %>% mutate\_if(sapply(bike, is.character), as.factor)  
  
bike = bike %>% mutate(hr = as\_factor(hr))

## Task 1 Response:

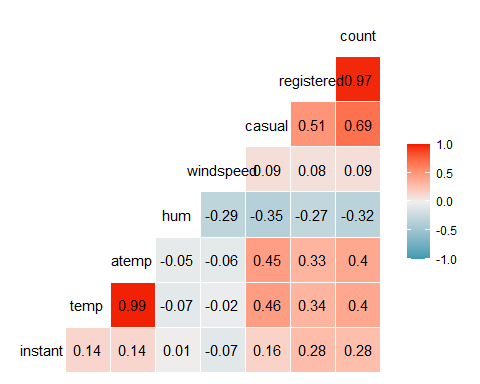
Hour is converted to a factor because it is better fitted as categorical than numeric due to its cyclical nature.

summary(bike)

## instant dteday season mnth   
## Min. : 1 Min. :2011-01-01 Fall :4232 Jul :1488   
## 1st Qu.: 4346 1st Qu.:2011-07-04 Spring:4409 May :1488   
## Median : 8690 Median :2012-01-02 Summer:4496 Dec :1483   
## Mean : 8690 Mean :2012-01-02 Winter:4242 Aug :1475   
## 3rd Qu.:13034 3rd Qu.:2012-07-02 Mar :1473   
## Max. :17379 Max. :2012-12-31 Oct :1451   
## (Other):8521   
## hr holiday weekday workingday   
## 16 : 730 Holiday : 500 Friday :2487 NotWorkingDay: 5514   
## 17 : 730 NotHoliday:16879 Monday :2479 WorkingDay :11865   
## 13 : 729 Saturday :2512   
## 14 : 729 Sunday :2502   
## 15 : 729 Thursday :2471   
## 12 : 728 Tuesday :2453   
## (Other):13004 Wednesday:2475   
## weathersit temp atemp hum   
## HeavyPrecip: 3 Min. :0.020 Min. :0.0000 Min. :0.0000   
## LightPrecip: 1419 1st Qu.:0.340 1st Qu.:0.3333 1st Qu.:0.4800   
## Misty : 4544 Median :0.500 Median :0.4848 Median :0.6300   
## NoPrecip :11413 Mean :0.497 Mean :0.4758 Mean :0.6272   
## 3rd Qu.:0.660 3rd Qu.:0.6212 3rd Qu.:0.7800   
## Max. :1.000 Max. :1.0000 Max. :1.0000   
##   
## windspeed casual registered count   
## Min. :0.0000 Min. : 0.00 Min. : 0.0 Min. : 1.0   
## 1st Qu.:0.1045 1st Qu.: 4.00 1st Qu.: 34.0 1st Qu.: 40.0   
## Median :0.1940 Median : 17.00 Median :115.0 Median :142.0   
## Mean :0.1901 Mean : 35.68 Mean :153.8 Mean :189.5   
## 3rd Qu.:0.2537 3rd Qu.: 48.00 3rd Qu.:220.0 3rd Qu.:281.0   
## Max. :0.8507 Max. :367.00 Max. :886.0 Max. :977.0   
##

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

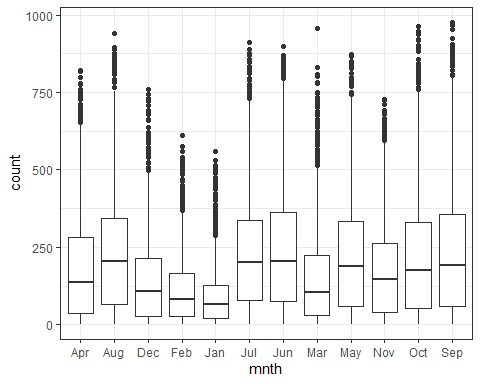
## Warning in ggcorr(bike, label = "TRUE", label\_round = 2): data in column(s)  
## 'dteday', 'season', 'mnth', 'hr', 'holiday', 'weekday', 'workingday',  
## 'weathersit' are not numeric and were ignored



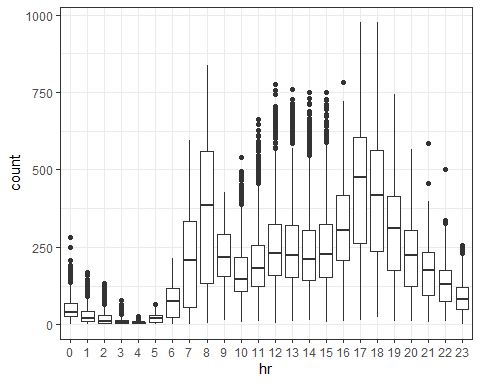
## Task 2 response:

Setting aside the Registered and Casual variables, atemp and temp appear to be the most highly correlated at 0.4.

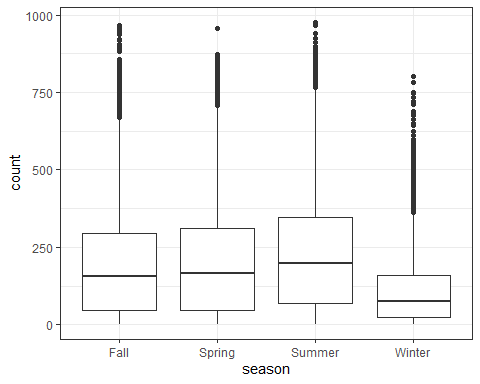
ggplot(bike,aes(x=mnth,y=count)) + geom\_boxplot() + theme\_bw()



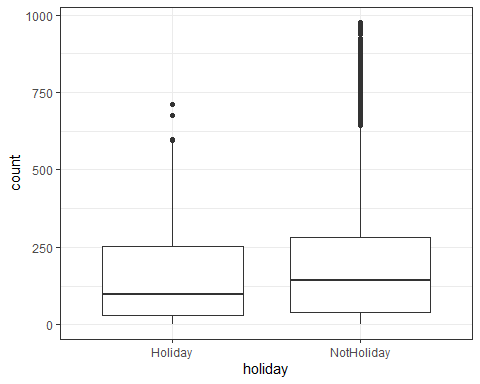
ggplot(bike,aes(x=hr,y=count)) + geom\_boxplot() + theme\_bw()



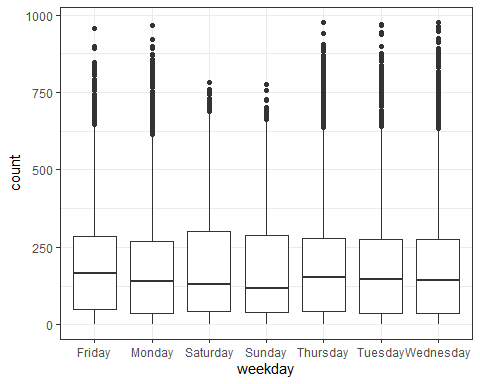
ggplot(bike,aes(x=season,y=count)) + geom\_boxplot() + theme\_bw()



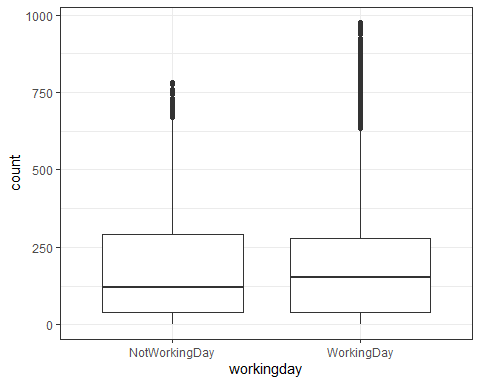
ggplot(bike,aes(x=holiday,y=count)) + geom\_boxplot() + theme\_bw()



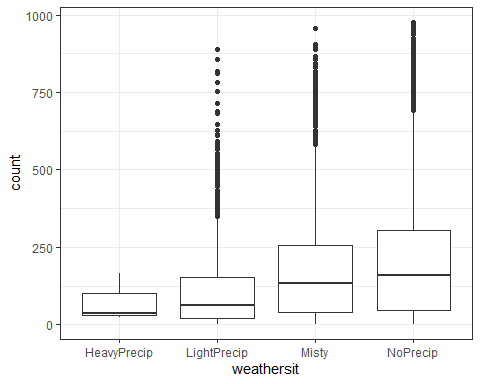
ggplot(bike,aes(x=weekday,y=count)) + geom\_boxplot() + theme\_bw()



ggplot(bike,aes(x=workingday,y=count)) + geom\_boxplot() + theme\_bw()



ggplot(bike,aes(x=weathersit,y=count)) + geom\_boxplot() + theme\_bw()



## Task 3 Response:

Month, Hour, Season, and weathersit all appear to affect the count of rental bikes while holiday, weekday, and workingday appear to have less of an impact. This is most likely associated with the common sense aspect that the conditions outside (influenced by month, season, and weathersit) influence how everyone is willing to travel or what activity they will seek out while people spend their holidays, weekdays, and workingdays in different manners. People are also more likely to ride during the day (hr) than at night.

bike\_recipe = recipe(count ~ hr, bike)  
  
lm\_model =   
 linear\_reg() %>%   
 set\_engine("lm")   
  
lm\_wflow =   
 workflow() %>%   
 add\_model(lm\_model) %>%   
 add\_recipe(bike\_recipe)  
  
lm\_fit = fit(lm\_wflow, bike)

summary(lm\_fit$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -446.45 -60.99 -6.01 50.10 551.49   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 53.898 4.756 11.332 < 2e-16 \*\*\*  
## hr1 -20.522 6.731 -3.049 0.002300 \*\*   
## hr2 -31.028 6.752 -4.595 4.35e-06 \*\*\*  
## hr3 -42.171 6.796 -6.205 5.58e-10 \*\*\*  
## hr4 -47.545 6.796 -6.996 2.73e-12 \*\*\*  
## hr5 -34.008 6.747 -5.040 4.70e-07 \*\*\*  
## hr6 22.146 6.729 3.291 0.000999 \*\*\*  
## hr7 158.167 6.724 23.523 < 2e-16 \*\*\*  
## hr8 305.113 6.724 45.377 < 2e-16 \*\*\*  
## hr9 165.411 6.724 24.600 < 2e-16 \*\*\*  
## hr10 119.770 6.724 17.812 < 2e-16 \*\*\*  
## hr11 154.245 6.724 22.939 < 2e-16 \*\*\*  
## hr12 199.418 6.722 29.668 < 2e-16 \*\*\*  
## hr13 199.763 6.719 29.729 < 2e-16 \*\*\*  
## hr14 187.051 6.719 27.838 < 2e-16 \*\*\*  
## hr15 197.335 6.719 29.368 < 2e-16 \*\*\*  
## hr16 258.085 6.717 38.422 < 2e-16 \*\*\*  
## hr17 407.554 6.717 60.674 < 2e-16 \*\*\*  
## hr18 371.613 6.722 55.286 < 2e-16 \*\*\*  
## hr19 257.625 6.722 38.327 < 2e-16 \*\*\*  
## hr20 172.132 6.722 25.608 < 2e-16 \*\*\*  
## hr21 118.416 6.722 17.617 < 2e-16 \*\*\*  
## hr22 77.437 6.722 11.520 < 2e-16 \*\*\*  
## hr23 33.933 6.722 5.048 4.50e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 128.2 on 17355 degrees of freedom  
## Multiple R-squared: 0.5015, Adjusted R-squared: 0.5008   
## F-statistic: 759.1 on 23 and 17355 DF, p-value: < 2.2e-16

## Task 4 Response:

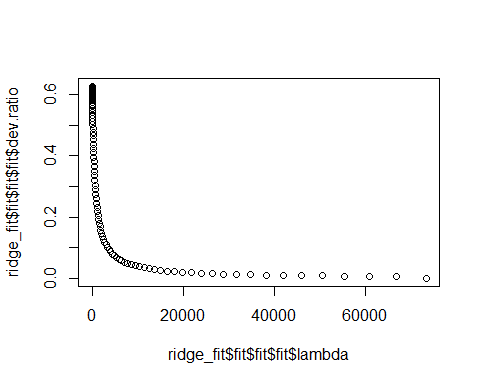
This model shows that every level of the hour variable is significant and the adjusted R-Squared value is .5008. This is an acceptable value for a predictor, but does not indicate much insight at this stage.

bike\_recipe2 = recipe(count ~., bike) %>%  
 step\_rm(instant,dteday,registered,casual) %>%  
 step\_dummy(all\_nominal()) %>%  
 step\_center(all\_predictors()) %>%  
 step\_scale(all\_predictors())   
   
ridge\_model =  
 linear\_reg(mixture = 0) %>%  
 set\_engine("glmnet")  
  
ridge\_wflow =  
 workflow() %>%  
 add\_model(ridge\_model) %>%  
 add\_recipe(bike\_recipe2)  
  
ridge\_fit = fit(ridge\_wflow, bike)

ridge\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit")

##   
## Call: glmnet::glmnet(x = maybe\_matrix(x), y = y, family = "gaussian", alpha = ~0)   
##   
## Df %Dev Lambda  
## 1 52 0.00 73420  
## 2 52 0.61 66900  
## 3 52 0.67 60950  
## 4 52 0.74 55540  
## 5 52 0.81 50600  
## 6 52 0.88 46110  
## 7 52 0.97 42010  
## 8 52 1.06 38280  
## 9 52 1.16 34880  
## 10 52 1.27 31780  
## 11 52 1.39 28960  
## 12 52 1.53 26390  
## 13 52 1.67 24040  
## 14 52 1.83 21910  
## 15 52 2.00 19960  
## 16 52 2.19 18190  
## 17 52 2.40 16570  
## 18 52 2.62 15100  
## 19 52 2.86 13760  
## 20 52 3.13 12540  
## 21 52 3.41 11420  
## 22 52 3.72 10410  
## 23 52 4.06 9482  
## 24 52 4.43 8640  
## 25 52 4.83 7872  
## 26 52 5.26 7173  
## 27 52 5.72 6536  
## 28 52 6.22 5955  
## 29 52 6.76 5426  
## 30 52 7.34 4944  
## 31 52 7.96 4505  
## 32 52 8.62 4105  
## 33 52 9.33 3740  
## 34 52 10.09 3408  
## 35 52 10.90 3105  
## 36 52 11.76 2829  
## 37 52 12.67 2578  
## 38 52 13.63 2349  
## 39 52 14.65 2140  
## 40 52 15.72 1950  
## 41 52 16.83 1777  
## 42 52 18.01 1619  
## 43 52 19.23 1475  
## 44 52 20.49 1344  
## 45 52 21.81 1225  
## 46 52 23.16 1116  
## 47 52 24.56 1017  
## 48 52 25.98 926  
## 49 52 27.44 844  
## 50 52 28.93 769  
## 51 52 30.43 701  
## 52 52 31.95 639  
## 53 52 33.48 582  
## 54 52 35.01 530  
## 55 52 36.53 483  
## 56 52 38.04 440  
## 57 52 39.54 401  
## 58 52 41.01 365  
## 59 52 42.44 333  
## 60 52 43.84 303  
## 61 52 45.20 276  
## 62 52 46.51 252  
## 63 52 47.77 230  
## 64 52 48.96 209  
## 65 52 50.10 190  
## 66 52 51.18 174  
## 67 52 52.19 158  
## 68 52 53.14 144  
## 69 52 54.02 131  
## 70 52 54.83 120  
## 71 52 55.59 109  
## 72 52 56.28 99  
## 73 52 56.91 91  
## 74 52 57.49 82  
## 75 52 58.01 75  
## 76 52 58.48 68  
## 77 52 58.91 62  
## 78 52 59.30 57  
## 79 52 59.64 52  
## 80 52 59.96 47  
## 81 52 60.24 43  
## 82 52 60.49 39  
## 83 52 60.72 36  
## 84 52 60.93 33  
## 85 52 61.11 30  
## 86 52 61.28 27  
## 87 52 61.44 25  
## 88 52 61.58 22  
## 89 52 61.71 20  
## 90 52 61.83 19  
## 91 52 61.95 17  
## 92 52 62.05 15  
## 93 52 62.14 14  
## 94 52 62.23 13  
## 95 52 62.32 12  
## 96 52 62.40 11  
## 97 52 62.47 10  
## 98 52 62.54 9  
## 99 52 62.60 8  
## 100 52 62.66 7

plot(ridge\_fit$fit$fit$fit$lambda,ridge\_fit$fit$fit$fit$dev.ratio)



ridge\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit") %>%   
 coef(s = 36)

## 53 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 189.46308763  
## temp 24.17009340  
## atemp 23.90423004  
## hum -24.36786720  
## windspeed -1.41387864  
## season\_Spring -1.58525595  
## season\_Summer -5.25298675  
## season\_Winter -14.20110243  
## mnth\_Aug -0.22523984  
## mnth\_Dec 0.72076504  
## mnth\_Feb -2.75554295  
## mnth\_Jan -3.12773901  
## mnth\_Jul -5.89600528  
## mnth\_Jun -1.60093013  
## mnth\_Mar -0.08066607  
## mnth\_May 2.83154981  
## mnth\_Nov 2.24406354  
## mnth\_Oct 8.01872626  
## mnth\_Sep 7.76778585  
## hr\_X1 -19.50885577  
## hr\_X2 -20.70925543  
## hr\_X3 -22.07657200  
## hr\_X4 -22.40814850  
## hr\_X5 -20.00280082  
## hr\_X6 -10.33865017  
## hr\_X7 12.15276688  
## hr\_X8 35.49757114  
## hr\_X9 10.23454114  
## hr\_X10 0.62509662  
## hr\_X11 4.50444543  
## hr\_X12 10.75071505  
## hr\_X13 9.69359595  
## hr\_X14 6.86057932  
## hr\_X15 8.40434934  
## hr\_X16 18.92098053  
## hr\_X17 44.99413662  
## hr\_X18 39.80576842  
## hr\_X19 21.92202326  
## hr\_X20 8.81626418  
## hr\_X21 0.81187810  
## hr\_X22 -5.19346023  
## hr\_X23 -11.56834223  
## holiday\_NotHoliday 3.27872387  
## weekday\_Monday -1.58977766  
## weekday\_Saturday 1.47923921  
## weekday\_Sunday -2.65253470  
## weekday\_Thursday -0.60387127  
## weekday\_Tuesday -1.00076839  
## weekday\_Wednesday -0.21223082  
## workingday\_WorkingDay 2.06578074  
## weathersit\_LightPrecip -10.89451797  
## weathersit\_Misty 2.08171919  
## weathersit\_NoPrecip 4.35356364

## Task 5 Response:

By graphing the ridge model, we see that the optimal R-Sqaured value is going to hover around .6, which means that the count and hr model was a pretty good model after all. We also see that the intercept is 189.46 and temp, atemp, hum, hr, season, and weathersit(precipitation) all have the largest coefficients (slope).

bike\_recipe3 = recipe(count ~., bike) %>%   
 step\_rm(instant,dteday,registered,casual) %>%  
 step\_dummy(all\_nominal()) %>%   
 step\_center(all\_predictors()) %>%   
 step\_scale(all\_predictors())  
   
lasso\_model =   
 linear\_reg(mixture = 1) %>%   
 set\_engine("glmnet")   
  
lasso\_wflow =   
 workflow() %>%   
 add\_model(lasso\_model) %>%   
 add\_recipe(bike\_recipe3)  
  
lasso\_fit = fit(lasso\_wflow, bike)

lasso\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit")

##   
## Call: glmnet::glmnet(x = maybe\_matrix(x), y = y, family = "gaussian", alpha = ~1)   
##   
## Df %Dev Lambda  
## 1 0 0.00 73.420  
## 2 1 2.78 66.900  
## 3 1 5.09 60.950  
## 4 3 7.60 55.540  
## 5 3 11.69 50.600  
## 6 4 15.44 46.110  
## 7 4 19.18 42.010  
## 8 6 22.56 38.280  
## 9 6 26.23 34.880  
## 10 6 29.28 31.780  
## 11 8 32.06 28.960  
## 12 11 34.97 26.390  
## 13 12 38.11 24.040  
## 14 12 40.86 21.910  
## 15 14 43.28 19.960  
## 16 14 45.50 18.190  
## 17 15 47.37 16.570  
## 18 15 49.03 15.100  
## 19 16 50.55 13.760  
## 20 16 51.81 12.540  
## 21 18 52.98 11.420  
## 22 19 54.01 10.410  
## 23 21 54.90 9.482  
## 24 24 55.78 8.640  
## 25 25 56.58 7.872  
## 26 26 57.29 7.173  
## 27 27 57.91 6.536  
## 28 27 58.47 5.955  
## 29 28 58.95 5.426  
## 30 28 59.38 4.944  
## 31 29 59.74 4.505  
## 32 31 60.09 4.105  
## 33 32 60.41 3.740  
## 34 32 60.69 3.408  
## 35 32 60.92 3.105  
## 36 33 61.11 2.829  
## 37 36 61.30 2.578  
## 38 37 61.60 2.349  
## 39 36 61.82 2.140  
## 40 36 61.98 1.950  
## 41 38 62.13 1.777  
## 42 39 62.25 1.619  
## 43 40 62.36 1.475  
## 44 41 62.46 1.344  
## 45 42 62.58 1.225  
## 46 42 62.69 1.116  
## 47 42 62.77 1.017  
## 48 41 62.84 0.926  
## 49 42 62.89 0.844  
## 50 42 62.92 0.769  
## 51 42 62.96 0.701  
## 52 42 62.98 0.639  
## 53 42 63.01 0.582  
## 54 42 63.04 0.530  
## 55 42 63.05 0.483  
## 56 43 63.07 0.440  
## 57 44 63.09 0.401  
## 58 45 63.11 0.365  
## 59 45 63.13 0.333  
## 60 45 63.14 0.303  
## 61 46 63.15 0.276  
## 62 49 63.16 0.252  
## 63 49 63.17 0.230  
## 64 49 63.18 0.209  
## 65 49 63.19 0.190  
## 66 49 63.19 0.174  
## 67 49 63.20 0.158  
## 68 49 63.20 0.144  
## 69 49 63.21 0.131  
## 70 48 63.21 0.120  
## 71 48 63.21 0.109  
## 72 48 63.21 0.099  
## 73 48 63.22 0.091  
## 74 49 63.22 0.082  
## 75 49 63.22 0.075  
## 76 49 63.22 0.068  
## 77 49 63.22 0.062  
## 78 49 63.22 0.057  
## 79 50 63.22 0.052  
## 80 50 63.22 0.047  
## 81 50 63.22 0.043

lasso\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit") %>%   
 coef(s = 0.209)

## 53 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 189.4630876  
## temp 34.4913037  
## atemp 19.5538246  
## hum -22.5725057  
## windspeed -4.8780834  
## season\_Spring -8.7291602  
## season\_Summer -14.6648672  
## season\_Winter -24.0996980  
## mnth\_Aug -0.6600791  
## mnth\_Dec .   
## mnth\_Feb 0.4340047  
## mnth\_Jan 0.4620123  
## mnth\_Jul -7.6399443  
## mnth\_Jun -2.6797887  
## mnth\_Mar 2.3567539  
## mnth\_May 2.3836670  
## mnth\_Nov -0.3946005  
## mnth\_Oct 4.8791023  
## mnth\_Sep 7.6190457  
## hr\_X1 -6.0781708  
## hr\_X2 -7.6077530  
## hr\_X3 -9.4575705  
## hr\_X4 -9.8013379  
## hr\_X5 -6.7082150  
## hr\_X6 4.5520354  
## hr\_X7 31.4038806  
## hr\_X8 59.1879182  
## hr\_X9 29.1345538  
## hr\_X10 17.6789279  
## hr\_X11 22.2547005  
## hr\_X12 29.7220165  
## hr\_X13 28.4625866  
## hr\_X14 25.1165495  
## hr\_X15 26.9612759  
## hr\_X16 39.5096359  
## hr\_X17 70.5697093  
## hr\_X18 64.3574488  
## hr\_X19 43.0130206  
## hr\_X20 27.4010703  
## hr\_X21 17.8011183  
## hr\_X22 10.6408589  
## hr\_X23 3.0553800  
## holiday\_NotHoliday 4.3114634  
## weekday\_Monday -1.7470310  
## weekday\_Saturday 0.1974065  
## weekday\_Sunday -4.5657282  
## weekday\_Thursday -0.6577382  
## weekday\_Tuesday -1.2366061  
## weekday\_Wednesday -0.2130957  
## workingday\_WorkingDay .   
## weathersit\_LightPrecip -14.6389163  
## weathersit\_Misty .   
## weathersit\_NoPrecip 2.6340613

## Task 6 Response: What are the implications of the model results from the ridge and lasso methods?

These models show that temp, atemp, hum, hr, season, and weathersit(precipitation) all have the largest coefficients (slope). In the Lasso model, Working day was kicked out as a variable entirely and month, weekday,a nd holidays all have lower coefficent values. These results would imply that the count of bike rentals is most closely correlated with aspects of weather and time of day. If I was going to take the model further, I would examine the combination of these variables to find which combination maximized the R-squared value while considering multicollinearity, as many aspects of weather can be related.