# Module 3 Assignment 1

## BAN 502

## UNCW

## Gotovich, James

library(lubridate)

##   
## Attaching package: 'lubridate'

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

library(tidyverse)

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

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.1.0 v dplyr 1.0.5  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x lubridate::as.difftime() masks base::as.difftime()  
## x lubridate::date() masks base::date()  
## x dplyr::filter() masks stats::filter()  
## x lubridate::intersect() masks base::intersect()  
## x dplyr::lag() masks stats::lag()  
## x lubridate::setdiff() masks base::setdiff()  
## x lubridate::union() masks base::union()

library(tidymodels)

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

## -- Attaching packages -------------------------------------- tidymodels 0.1.3 --

## v broom 0.7.6 v rsample 0.1.0   
## v dials 0.0.9 v tune 0.1.5   
## v infer 0.5.4 v workflows 0.2.2   
## v modeldata 0.1.0 v workflowsets 0.0.2   
## v parsnip 0.1.5 v yardstick 0.0.8   
## v recipes 0.1.16

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

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

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

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

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

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

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

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

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

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

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

## -- Conflicts ----------------------------------------- tidymodels\_conflicts() --  
## x scales::discard() masks purrr::discard()  
## x dplyr::filter() masks stats::filter()  
## x recipes::fixed() masks stringr::fixed()  
## x dplyr::lag() masks stats::lag()  
## x yardstick::spec() masks readr::spec()  
## x recipes::step() masks stats::step()  
## \* Use tidymodels\_prefer() to resolve common conflicts.

bike\_cleaned\_2 <- read\_csv("bike\_cleaned-2.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\_cleaned\_2%>%  
mutate(dteday=mdy(dteday))%>%  
mutate(season = as\_factor(season))%>%  
mutate(mnth = as\_factor(mnth))%>%  
mutate(hr = as\_factor(hr))%>%  
mutate(holiday=as\_factor(holiday))%>%  
mutate(weekday=as\_factor(weekday))%>%  
mutate(workingday=as\_factor(workingday))%>%  
mutate(weathersit=as\_factor(weathersit))

set.seed(1234)  
bike\_split = initial\_split(bike, prob = 0.70, strata = count)  
train = training(bike\_split)  
test = testing(bike\_split)

The test contains 4347 rows, and the test 13032.

bike\_recipe = recipe(count ~ season+mnth+hr+holiday+weekday+temp+weathersit, train)   
  
 lm\_model=  
 linear\_reg()%>%  
 set\_engine("lm")  
   
 lm\_wflow=  
 workflow()%>%  
 add\_model(lm\_model)%>%  
 add\_recipe(bike\_recipe)  
   
 lm\_fit=fit(lm\_wflow, train)

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

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -425.40 -62.01 -9.19 51.82 498.22   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -84.6667 6.7301 -12.580 < 2e-16 \*\*\*  
## seasonSpring 30.3018 6.1412 4.934 8.15e-07 \*\*\*  
## seasonSummer 22.4951 7.2458 3.105 0.001910 \*\*   
## seasonFall 62.6073 6.1966 10.104 < 2e-16 \*\*\*  
## mnthFeb 0.7343 4.9253 0.149 0.881489   
## mnthMar 4.8147 5.5624 0.866 0.386735   
## mnthApr -1.5806 8.2916 -0.191 0.848819   
## mnthMay -3.7640 8.8780 -0.424 0.671596   
## mnthJun -12.6290 9.1016 -1.388 0.165297   
## mnthJul -35.9128 10.1886 -3.525 0.000425 \*\*\*  
## mnthAug -17.6135 9.9313 -1.774 0.076163 .   
## mnthSep 5.0866 8.8359 0.576 0.564843   
## mnthOct -3.0935 8.2266 -0.376 0.706900   
## mnthNov -18.4377 7.9346 -2.324 0.020156 \*   
## mnthDec -14.6731 6.2960 -2.331 0.019793 \*   
## hr1 -20.0010 6.7788 -2.951 0.003178 \*\*   
## hr2 -28.2104 6.7758 -4.163 3.16e-05 \*\*\*  
## hr3 -39.9847 6.8862 -5.806 6.53e-09 \*\*\*  
## hr4 -40.0952 6.8248 -5.875 4.33e-09 \*\*\*  
## hr5 -26.3996 6.7869 -3.890 0.000101 \*\*\*  
## hr6 31.6189 6.8577 4.611 4.05e-06 \*\*\*  
## hr7 164.7441 6.7880 24.270 < 2e-16 \*\*\*  
## hr8 306.2584 6.8106 44.968 < 2e-16 \*\*\*  
## hr9 162.9738 6.7725 24.064 < 2e-16 \*\*\*  
## hr10 109.6755 6.7060 16.355 < 2e-16 \*\*\*  
## hr11 138.2796 6.7761 20.407 < 2e-16 \*\*\*  
## hr12 179.9545 6.7808 26.539 < 2e-16 \*\*\*  
## hr13 177.7657 6.8321 26.019 < 2e-16 \*\*\*  
## hr14 155.5932 6.9615 22.350 < 2e-16 \*\*\*  
## hr15 167.9333 6.8691 24.447 < 2e-16 \*\*\*  
## hr16 229.8491 6.8955 33.333 < 2e-16 \*\*\*  
## hr17 385.3796 6.8544 56.224 < 2e-16 \*\*\*  
## hr18 348.4015 6.7443 51.659 < 2e-16 \*\*\*  
## hr19 241.7780 6.8337 35.380 < 2e-16 \*\*\*  
## hr20 162.3119 6.8192 23.802 < 2e-16 \*\*\*  
## hr21 108.5944 6.7785 16.020 < 2e-16 \*\*\*  
## hr22 71.4647 6.7990 10.511 < 2e-16 \*\*\*  
## hr23 33.0981 6.7487 4.904 9.49e-07 \*\*\*  
## holidayHoliday -21.7636 6.0393 -3.604 0.000315 \*\*\*  
## weekdaySunday -10.9450 3.6314 -3.014 0.002584 \*\*   
## weekdayMonday -5.1098 3.7595 -1.359 0.174112   
## weekdayTuesday -4.3280 3.6663 -1.180 0.237828   
## weekdayWednesday -1.2262 3.6497 -0.336 0.736896   
## weekdayThursday -1.1646 3.6618 -0.318 0.750459   
## weekdayFriday 2.2216 3.6488 0.609 0.542619   
## temp 285.8953 11.7631 24.305 < 2e-16 \*\*\*  
## weathersitMisty -17.7668 2.2849 -7.776 8.06e-15 \*\*\*  
## weathersitLightPrecip -91.5764 3.6605 -25.017 < 2e-16 \*\*\*  
## weathersitHeavyPrecip -78.3382 64.4412 -1.216 0.224138   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 111.3 on 12983 degrees of freedom  
## Multiple R-squared: 0.6237, Adjusted R-squared: 0.6223   
## F-statistic: 448.4 on 48 and 12983 DF, p-value: < 2.2e-16

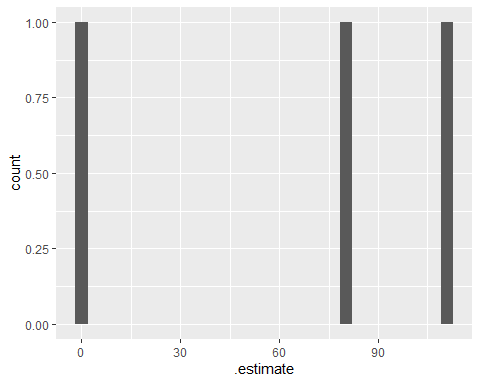
Because of all the factors involved its a bigger model, but significance codes are mostly high and the r value is .62.

predict\_train=  
 lm\_fit%>%predict(test)%>%bind\_cols(test)%>%metrics(truth=count, estimate=.pred)  
show(predict\_train)

## # A tibble: 3 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 112.   
## 2 rsq standard 0.625  
## 3 mae standard 81.0

ggplot(predict\_train, aes(x=.estimate))+  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



The r value of the esimate set matched the one that was assesed earlier at .62