bike <- read\_csv("~/Ban502/Module 3/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 %>% mutate(dteday = mdy(dteday))  
  
bike = bike %>% mutate\_if(sapply(bike, is.character), as\_factor)  
  
bike = bike %>% mutate(hr = as\_factor(hr))

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

## Task 2 Response: Training has 13,036 rows while Testing has 4,343.

bike\_recipe = recipe(count ~ season + mnth + hr + holiday + weekday + temp + weathersit, 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)

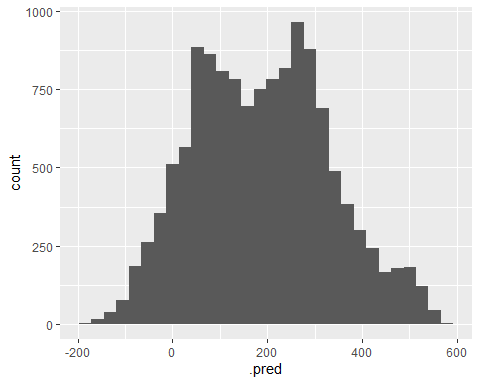
summary(lm\_fit$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -424.63 -62.16 -9.71 51.92 499.33   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -85.278 5.808 -14.684 < 2e-16 \*\*\*  
## seasonSpring 35.545 5.306 6.699 2.17e-11 \*\*\*  
## seasonSummer 26.998 6.289 4.293 1.77e-05 \*\*\*  
## seasonFall 65.129 5.330 12.219 < 2e-16 \*\*\*  
## mnthFeb 1.323 4.287 0.309 0.75768   
## mnthMar 5.078 4.818 1.054 0.29187   
## mnthApr -6.014 7.152 -0.841 0.40041   
## mnthMay -5.832 7.647 -0.763 0.44566   
## mnthJun -18.097 7.850 -2.305 0.02116 \*   
## mnthJul -41.455 8.813 -4.704 2.58e-06 \*\*\*  
## mnthAug -21.251 8.572 -2.479 0.01318 \*   
## mnthSep 4.316 7.622 0.566 0.57119   
## mnthOct -3.922 7.079 -0.554 0.57954   
## mnthNov -18.304 6.823 -2.683 0.00731 \*\*   
## mnthDec -15.180 5.411 -2.805 0.00503 \*\*   
## hr1 -17.912 5.849 -3.062 0.00220 \*\*   
## hr2 -26.901 5.868 -4.584 4.59e-06 \*\*\*  
## hr3 -37.809 5.909 -6.399 1.61e-10 \*\*\*  
## hr4 -41.087 5.912 -6.950 3.78e-12 \*\*\*  
## hr5 -24.888 5.873 -4.238 2.27e-05 \*\*\*  
## hr6 33.488 5.858 5.717 1.10e-08 \*\*\*  
## hr7 169.440 5.850 28.963 < 2e-16 \*\*\*  
## hr8 310.710 5.845 53.160 < 2e-16 \*\*\*  
## hr9 164.653 5.845 28.170 < 2e-16 \*\*\*  
## hr10 111.648 5.853 19.075 < 2e-16 \*\*\*  
## hr11 139.110 5.870 23.697 < 2e-16 \*\*\*  
## hr12 180.131 5.889 30.588 < 2e-16 \*\*\*  
## hr13 176.032 5.907 29.801 < 2e-16 \*\*\*  
## hr14 160.344 5.924 27.067 < 2e-16 \*\*\*  
## hr15 169.807 5.931 28.632 < 2e-16 \*\*\*  
## hr16 231.354 5.925 39.050 < 2e-16 \*\*\*  
## hr17 384.495 5.907 65.086 < 2e-16 \*\*\*  
## hr18 351.933 5.892 59.735 < 2e-16 \*\*\*  
## hr19 241.539 5.870 41.147 < 2e-16 \*\*\*  
## hr20 161.120 5.858 27.506 < 2e-16 \*\*\*  
## hr21 110.339 5.848 18.868 < 2e-16 \*\*\*  
## hr22 72.378 5.843 12.387 < 2e-16 \*\*\*  
## hr23 33.232 5.841 5.689 1.30e-08 \*\*\*  
## holidayHoliday -26.140 5.335 -4.899 9.71e-07 \*\*\*  
## weekdaySunday -15.873 3.148 -5.043 4.64e-07 \*\*\*  
## weekdayMonday -7.779 3.248 -2.395 0.01663 \*   
## weekdayTuesday -6.528 3.172 -2.058 0.03960 \*   
## weekdayWednesday -3.805 3.166 -1.202 0.22940   
## weekdayThursday -2.393 3.165 -0.756 0.44960   
## weekdayFriday 1.631 3.154 0.517 0.60515   
## temp 287.864 10.218 28.173 < 2e-16 \*\*\*  
## weathersitMisty -19.377 1.981 -9.782 < 2e-16 \*\*\*  
## weathersitLightPrecip -90.772 3.168 -28.650 < 2e-16 \*\*\*  
## weathersitHeavyPrecip -78.721 64.407 -1.222 0.22163   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 111.3 on 17330 degrees of freedom  
## Multiple R-squared: 0.6242, Adjusted R-squared: 0.6232   
## F-statistic: 599.8 on 48 and 17330 DF, p-value: < 2.2e-16

## Task 3 Response: We see that many factors included in the model are significant other than month and weekday. The adjusted R-squared value is 0.6232. This is an acceptable value for a model, but does not indicate much more insight to quality at this stage.

predict\_train = predict(lm\_fit,train)  
ggplot(predict\_train, aes(x = .pred )) +  
 geom\_histogram()

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



## Task 4 Response: The distribution of the prediction model is bimodal. According to <https://www.statisticshowto.com/what-is-a-bimodal-distribution/>, this indicates that there are two different groups represented in the data set. It also appears to be nearly semetric in shape, centering around 150 - 175 on the x axis.

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

## # A tibble: 3 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 112.   
## 2 rsq standard 0.625  
## 3 mae standard 81.4

## Task 5 Response: When the model is applied to the test data, the R-Squared value actually increased from 0.6232 to 0.624. Performance on the test set is similar to that on the training set. This suggests that our model is not overfitting.