Sales Forecasting - Initial Results

Eunchong Jung

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Exploratory Data Analysis

Stores Data

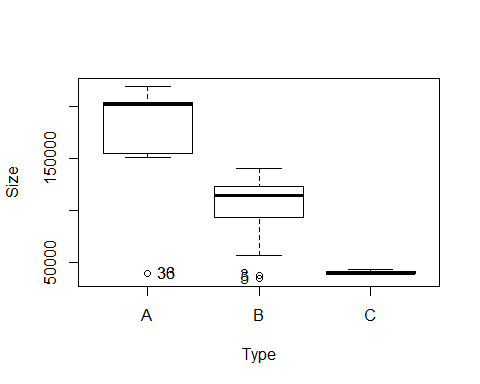
stores\_raw <- read.csv("stores.csv")  
stores\_raw$Store <- as.factor(stores\_raw$Store)  
str(stores\_raw)

## 'data.frame': 45 obs. of 3 variables:  
## $ Store: Factor w/ 45 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Type : Factor w/ 3 levels "A","B","C": 1 1 2 1 2 1 2 1 2 2 ...  
## $ Size : int 151315 202307 37392 205863 34875 202505 70713 155078 125833 126512 ...

summary(stores\_raw)

## Store Type Size   
## 1 : 1 A:22 Min. : 34875   
## 2 : 1 B:17 1st Qu.: 70713   
## 3 : 1 C: 6 Median :126512   
## 4 : 1 Mean :130288   
## 5 : 1 3rd Qu.:202307   
## 6 : 1 Max. :219622   
## (Other):39

# outliers  
Boxplot(Size~Type, data = stores\_raw, id.method = "y")



## [1] "33" "36" "3" "5"

stores\_raw[c(33, 36, 3, 5),]

## Store Type Size  
## 33 33 A 39690  
## 36 36 A 39910  
## 3 3 B 37392  
## 5 5 B 34875

stores\_clean <- stores\_raw  
aggregate(stores\_clean$Size, list(stores\_clean$Type), mean)

## Group.1 x  
## 1 A 177247.73  
## 2 B 101190.71  
## 3 C 40541.67

# replace outliers to Type "C" since values are similar to the mean of "C" stores  
stores\_clean[c(33, 36, 3, 5),2] <- "C"  
  
  
# check NAs  
sum(is.na(stores\_clean))

## [1] 0

summary(stores\_clean)

## Store Type Size   
## 1 : 1 A:20 Min. : 34875   
## 2 : 1 B:15 1st Qu.: 70713   
## 3 : 1 C:10 Median :126512   
## 4 : 1 Mean :130288   
## 5 : 1 3rd Qu.:202307   
## 6 : 1 Max. :219622   
## (Other):39

Training Data

train\_raw <- read.csv("train.csv")  
str(train\_raw)

## 'data.frame': 421570 obs. of 5 variables:  
## $ Store : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Dept : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Date : Factor w/ 143 levels "2010-02-05","2010-02-12",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Weekly\_Sales: num 24925 46039 41596 19404 21828 ...  
## $ IsHoliday : logi FALSE TRUE FALSE FALSE FALSE FALSE ...

summary(train\_raw)

## Store Dept Date Weekly\_Sales   
## Min. : 1.0 Min. : 1.00 2011-12-23: 3027 Min. : -4989   
## 1st Qu.:11.0 1st Qu.:18.00 2011-11-25: 3021 1st Qu.: 2080   
## Median :22.0 Median :37.00 2011-12-16: 3013 Median : 7612   
## Mean :22.2 Mean :44.26 2011-12-09: 3010 Mean : 15981   
## 3rd Qu.:33.0 3rd Qu.:74.00 2012-02-17: 3007 3rd Qu.: 20206   
## Max. :45.0 Max. :99.00 2011-12-30: 3003 Max. :693099   
## (Other) :403489   
## IsHoliday   
## Mode :logical   
## FALSE:391909   
## TRUE :29661   
##   
##   
##   
##

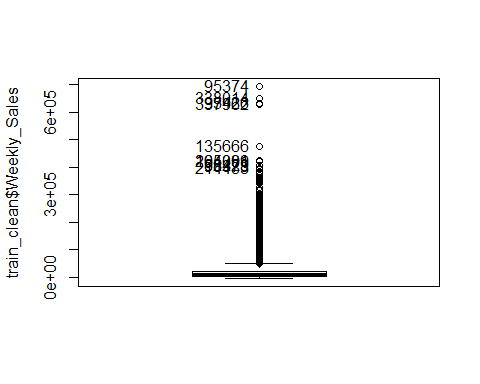
train\_clean <- train\_raw  
train\_clean$Store <- as.factor(train\_clean$Store)  
train\_clean$Dept <- as.factor(train\_clean$Dept)  
  
# check missing depts in each store  
stores\_unique <- levels(train\_clean$Store)  
depts\_unique <- levels(train\_clean$Dept)  
stores\_list <- as.list(stores\_unique)  
check\_depts <- function(x) {  
 identical(levels(train\_clean[train\_clean$Store == x, 2]), depts\_unique)  
}  
summary(sapply(stores\_list, check\_depts))

## Mode TRUE   
## logical 45

# check missing dates  
train\_clean$Date <- as.Date(train\_clean$Date)  
dates\_unique <- unique(train\_clean$Date)  
dates\_generated <- seq(min(train\_clean$Date), max(train\_clean$Date), by = "week")  
identical(dates\_unique, dates\_generated)

## [1] TRUE

# outliers; not removed due to vast difference of sales between stores  
Boxplot(train\_clean$Weekly\_Sales)



## [1] 95374 338014 95426 337962 135666 195089 264391 88429 95378 214433

# check NAs  
sum(is.na(train\_clean))

## [1] 0

# convert dates to day of week & week number  
# unused (all values are Friday): train\_clean$day\_of\_week <- wday(train\_clean$Date, label = TRUE)  
train\_clean$week\_number <- as.factor(week(train\_clean$Date))  
  
summary(train\_clean)

## Store Dept Date Weekly\_Sales   
## 13 : 10474 1 : 6435 Min. :2010-02-05 Min. : -4989   
## 10 : 10315 2 : 6435 1st Qu.:2010-10-08 1st Qu.: 2080   
## 4 : 10272 3 : 6435 Median :2011-06-17 Median : 7612   
## 1 : 10244 4 : 6435 Mean :2011-06-18 Mean : 15981   
## 2 : 10238 7 : 6435 3rd Qu.:2012-02-24 3rd Qu.: 20206   
## 24 : 10228 8 : 6435 Max. :2012-10-26 Max. :693099   
## (Other):359799 (Other):382960   
## IsHoliday week\_number   
## Mode :logical 7 : 8908   
## FALSE:391909 9 : 8899   
## TRUE :29661 6 : 8894   
## 10 : 8882   
## 15 : 8881   
## 16 : 8874   
## (Other):368232

str(train\_clean)

## 'data.frame': 421570 obs. of 6 variables:  
## $ Store : Factor w/ 45 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Dept : Factor w/ 81 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Date : Date, format: "2010-02-05" "2010-02-12" ...  
## $ Weekly\_Sales: num 24925 46039 41596 19404 21828 ...  
## $ IsHoliday : logi FALSE TRUE FALSE FALSE FALSE FALSE ...  
## $ week\_number : Factor w/ 53 levels "1","2","3","4",..: 6 7 8 9 10 11 12 13 14 15 ...

Features Data

features\_raw <- read.csv("features.csv")  
str(features\_raw)

## 'data.frame': 8190 obs. of 12 variables:  
## $ Store : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Date : Factor w/ 182 levels "2010-02-05","2010-02-12",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Temperature : num 42.3 38.5 39.9 46.6 46.5 ...  
## $ Fuel\_Price : num 2.57 2.55 2.51 2.56 2.62 ...  
## $ MarkDown1 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ MarkDown2 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ MarkDown3 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ MarkDown4 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ MarkDown5 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ CPI : num 211 211 211 211 211 ...  
## $ Unemployment: num 8.11 8.11 8.11 8.11 8.11 ...  
## $ IsHoliday : logi FALSE TRUE FALSE FALSE FALSE FALSE ...

features\_clean <- features\_raw  
features\_clean$Store <- as.factor(features\_clean$Store)  
features\_clean$Date <- as.Date(features\_clean$Date)  
  
  
# check NAs; CPI, Unemployment, IsHoliday have 585 NAs corresponding to 13 extra weeks \* 45 stores = 585, not found in training data; disregard  
sum(is.na(features\_clean))

## [1] 24040

# remove MarkDowns; too many NAs  
# explore MarkDowns later  
features\_clean <- features\_clean[, c(1:4, 10:12)]  
str(features\_clean)

## 'data.frame': 8190 obs. of 7 variables:  
## $ Store : Factor w/ 45 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Date : Date, format: "2010-02-05" "2010-02-12" ...  
## $ Temperature : num 42.3 38.5 39.9 46.6 46.5 ...  
## $ Fuel\_Price : num 2.57 2.55 2.51 2.56 2.62 ...  
## $ CPI : num 211 211 211 211 211 ...  
## $ Unemployment: num 8.11 8.11 8.11 8.11 8.11 ...  
## $ IsHoliday : logi FALSE TRUE FALSE FALSE FALSE FALSE ...

summary(features\_clean)

## Store Date Temperature Fuel\_Price   
## 1 : 182 Min. :2010-02-05 Min. : -7.29 Min. :2.472   
## 2 : 182 1st Qu.:2010-12-17 1st Qu.: 45.90 1st Qu.:3.041   
## 3 : 182 Median :2011-10-31 Median : 60.71 Median :3.513   
## 4 : 182 Mean :2011-10-31 Mean : 59.36 Mean :3.406   
## 5 : 182 3rd Qu.:2012-09-14 3rd Qu.: 73.88 3rd Qu.:3.743   
## 6 : 182 Max. :2013-07-26 Max. :101.95 Max. :4.468   
## (Other):7098   
## CPI Unemployment IsHoliday   
## Min. :126.1 Min. : 3.684 Mode :logical   
## 1st Qu.:132.4 1st Qu.: 6.634 FALSE:7605   
## Median :182.8 Median : 7.806 TRUE :585   
## Mean :172.5 Mean : 7.827   
## 3rd Qu.:213.9 3rd Qu.: 8.567   
## Max. :229.0 Max. :14.313   
## NA's :585 NA's :585

Combine datasets

df <- join(train\_clean, stores\_clean, by = "Store", type = "left")  
df <- join(df, features\_clean, by = c("Store", "Date"), type = "left")  
df <- df[-3]  
identical(df$IsHoliday, df$IsHoliday.1)

## [1] TRUE

df$IsHoliday.1 <- NULL

Multiple Regression

# Split the dataset to 80% training and 20% test sets  
rn\_train <- sample(nrow(df), floor(nrow(df)\*0.8))  
train <- df[rn\_train,]  
test <- df[-rn\_train,]  
  
model <- lm(Weekly\_Sales ~ ., data = train)  
summary(model)

##   
## Call:  
## lm(formula = Weekly\_Sales ~ ., data = train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -65443 -5798 -474 5061 627401   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 34363.487 2687.755 12.785 < 2e-16 \*\*\*  
## Store2 5144.641 206.946 24.860 < 2e-16 \*\*\*  
## Store3 -16546.275 217.863 -75.948 < 2e-16 \*\*\*  
## Store4 4828.970 1128.622 4.279 1.88e-05 \*\*\*  
## Store5 -18597.327 223.335 -83.271 < 2e-16 \*\*\*  
## Store6 -332.810 211.785 -1.571 0.116080   
## Store7 -13230.068 357.687 -36.988 < 2e-16 \*\*\*  
## Store8 -9580.790 222.310 -43.097 < 2e-16 \*\*\*  
## Store9 -14887.660 226.064 -65.856 < 2e-16 \*\*\*  
## Store10 3551.192 1104.773 3.214 0.001307 \*\*   
## Store11 -2958.079 212.353 -13.930 < 2e-16 \*\*\*  
## Store12 -6155.978 1051.598 -5.854 4.81e-09 \*\*\*  
## Store13 4254.401 1110.869 3.830 0.000128 \*\*\*  
## Store14 7078.203 418.906 16.897 < 2e-16 \*\*\*  
## Store15 -14159.335 1032.902 -13.708 < 2e-16 \*\*\*  
## Store16 -15130.020 382.058 -39.601 < 2e-16 \*\*\*  
## Store17 -10809.001 1121.889 -9.635 < 2e-16 \*\*\*  
## Store18 -6952.194 1010.534 -6.880 6.01e-12 \*\*\*  
## Store19 -2604.240 1032.616 -2.522 0.011670 \*   
## Store20 8276.686 238.641 34.683 < 2e-16 \*\*\*  
## Store21 -11059.070 210.285 -52.591 < 2e-16 \*\*\*  
## Store22 -8113.940 976.132 -8.312 < 2e-16 \*\*\*  
## Store23 -4897.425 1091.747 -4.486 7.26e-06 \*\*\*  
## Store24 -3637.287 1024.583 -3.550 0.000385 \*\*\*  
## Store25 -11627.754 244.528 -47.552 < 2e-16 \*\*\*  
## Store26 -8701.184 1029.193 -8.454 < 2e-16 \*\*\*  
## Store27 2039.831 984.793 2.071 0.038329 \*   
## Store28 -2089.391 1051.193 -1.988 0.046853 \*   
## Store29 -14520.428 996.876 -14.566 < 2e-16 \*\*\*  
## Store30 -17847.660 228.517 -78.102 < 2e-16 \*\*\*  
## Store31 -2249.636 207.132 -10.861 < 2e-16 \*\*\*  
## Store32 -4922.900 339.209 -14.513 < 2e-16 \*\*\*  
## Store33 -23039.652 1110.301 -20.751 < 2e-16 \*\*\*  
## Store34 -9285.970 1057.172 -8.784 < 2e-16 \*\*\*  
## Store35 -9330.757 962.740 -9.692 < 2e-16 \*\*\*  
## Store36 -19306.825 239.640 -80.566 < 2e-16 \*\*\*  
## Store37 -16181.012 228.666 -70.763 < 2e-16 \*\*\*  
## Store38 -17589.891 1055.594 -16.664 < 2e-16 \*\*\*  
## Store39 -1296.772 209.089 -6.202 5.58e-10 \*\*\*  
## Store40 -10961.526 1092.125 -10.037 < 2e-16 \*\*\*  
## Store41 -4298.861 366.118 -11.742 < 2e-16 \*\*\*  
## Store42 -16253.032 1110.009 -14.642 < 2e-16 \*\*\*  
## Store43 -12661.850 262.227 -48.286 < 2e-16 \*\*\*  
## Store44 -22594.171 1121.203 -20.152 < 2e-16 \*\*\*  
## Store45 -10484.504 420.074 -24.959 < 2e-16 \*\*\*  
## Dept2 24443.716 260.939 93.676 < 2e-16 \*\*\*  
## Dept3 -7494.242 261.092 -28.703 < 2e-16 \*\*\*  
## Dept4 6601.263 260.811 25.311 < 2e-16 \*\*\*  
## Dept5 1925.574 261.571 7.362 1.82e-13 \*\*\*  
## Dept6 -15385.957 266.751 -57.679 < 2e-16 \*\*\*  
## Dept7 4821.193 260.161 18.532 < 2e-16 \*\*\*  
## Dept8 10992.569 260.876 42.137 < 2e-16 \*\*\*  
## Dept9 694.997 261.493 2.658 0.007866 \*\*   
## Dept10 -1025.221 261.282 -3.924 8.72e-05 \*\*\*  
## Dept11 -4784.327 261.450 -18.299 < 2e-16 \*\*\*  
## Dept12 -15227.261 262.179 -58.080 < 2e-16 \*\*\*  
## Dept13 11339.931 261.193 43.416 < 2e-16 \*\*\*  
## Dept14 -4430.532 261.708 -16.929 < 2e-16 \*\*\*  
## Dept16 -5074.589 260.074 -19.512 < 2e-16 \*\*\*  
## Dept17 -9112.998 260.484 -34.985 < 2e-16 \*\*\*  
## Dept18 -12834.105 279.554 -45.909 < 2e-16 \*\*\*  
## Dept19 -20443.632 296.261 -69.005 < 2e-16 \*\*\*  
## Dept20 -14410.721 266.301 -54.114 < 2e-16 \*\*\*  
## Dept21 -14255.565 260.713 -54.679 < 2e-16 \*\*\*  
## Dept22 -10252.460 271.307 -37.789 < 2e-16 \*\*\*  
## Dept23 3477.728 269.329 12.913 < 2e-16 \*\*\*  
## Dept24 -14842.075 273.654 -54.237 < 2e-16 \*\*\*  
## Dept25 -10465.541 261.867 -39.965 < 2e-16 \*\*\*  
## Dept26 -12864.891 268.378 -47.936 < 2e-16 \*\*\*  
## Dept27 -19284.977 269.938 -71.442 < 2e-16 \*\*\*  
## Dept28 -19289.261 264.539 -72.917 < 2e-16 \*\*\*  
## Dept29 -15868.685 275.614 -57.576 < 2e-16 \*\*\*  
## Dept30 -17435.129 274.960 -63.410 < 2e-16 \*\*\*  
## Dept31 -17723.405 264.514 -67.004 < 2e-16 \*\*\*  
## Dept32 -13340.456 265.705 -50.208 < 2e-16 \*\*\*  
## Dept33 -14873.916 273.927 -54.299 < 2e-16 \*\*\*  
## Dept34 -5865.211 273.870 -21.416 < 2e-16 \*\*\*  
## Dept35 -18598.042 274.513 -67.749 < 2e-16 \*\*\*  
## Dept36 -19452.013 275.693 -70.557 < 2e-16 \*\*\*  
## Dept37 -22464.540 345.682 -64.986 < 2e-16 \*\*\*  
## Dept38 42039.376 261.155 160.975 < 2e-16 \*\*\*  
## Dept39 -28376.135 3827.257 -7.414 1.23e-13 \*\*\*  
## Dept40 25622.035 261.168 98.106 < 2e-16 \*\*\*  
## Dept41 -19280.310 273.575 -70.476 < 2e-16 \*\*\*  
## Dept42 -14191.019 261.079 -54.355 < 2e-16 \*\*\*  
## Dept43 -16076.471 4418.101 -3.639 0.000274 \*\*\*  
## Dept44 -16485.644 271.560 -60.707 < 2e-16 \*\*\*  
## Dept45 -22362.235 382.277 -58.498 < 2e-16 \*\*\*  
## Dept46 605.316 260.509 2.324 0.020148 \*   
## Dept47 -22843.611 606.330 -37.675 < 2e-16 \*\*\*  
## Dept48 -23921.690 400.009 -59.803 < 2e-16 \*\*\*  
## Dept49 -14450.819 288.287 -50.127 < 2e-16 \*\*\*  
## Dept50 -23536.218 417.277 -56.404 < 2e-16 \*\*\*  
## Dept51 -21351.105 439.944 -48.531 < 2e-16 \*\*\*  
## Dept52 -17761.519 263.047 -67.522 < 2e-16 \*\*\*  
## Dept54 -21897.556 282.670 -77.467 < 2e-16 \*\*\*  
## Dept55 -9913.339 270.518 -36.646 < 2e-16 \*\*\*  
## Dept56 -16520.610 267.267 -61.813 < 2e-16 \*\*\*  
## Dept58 -19289.903 290.393 -66.427 < 2e-16 \*\*\*  
## Dept59 -19135.751 263.898 -72.512 < 2e-16 \*\*\*  
## Dept60 -18909.975 267.536 -70.682 < 2e-16 \*\*\*  
## Dept65 27267.809 1246.374 21.878 < 2e-16 \*\*\*  
## Dept67 -11915.440 261.050 -45.644 < 2e-16 \*\*\*  
## Dept71 -15804.904 275.059 -57.460 < 2e-16 \*\*\*  
## Dept72 30668.472 264.883 115.781 < 2e-16 \*\*\*  
## Dept74 -5483.760 260.584 -21.044 < 2e-16 \*\*\*  
## Dept77 -24272.205 1195.471 -20.303 < 2e-16 \*\*\*  
## Dept78 -23021.802 993.691 -23.168 < 2e-16 \*\*\*  
## Dept79 2692.148 260.610 10.330 < 2e-16 \*\*\*  
## Dept80 -7765.458 266.795 -29.106 < 2e-16 \*\*\*  
## Dept81 -3867.585 260.535 -14.845 < 2e-16 \*\*\*  
## Dept82 -3895.578 260.862 -14.933 < 2e-16 \*\*\*  
## Dept83 -16645.785 266.792 -62.392 < 2e-16 \*\*\*  
## Dept85 -17776.316 265.135 -67.046 < 2e-16 \*\*\*  
## Dept87 -6016.268 261.838 -22.977 < 2e-16 \*\*\*  
## Dept90 26009.181 261.066 99.627 < 2e-16 \*\*\*  
## Dept91 14202.824 261.117 54.393 < 2e-16 \*\*\*  
## Dept92 56002.631 259.716 215.630 < 2e-16 \*\*\*  
## Dept93 7071.888 266.893 26.497 < 2e-16 \*\*\*  
## Dept94 13589.277 269.386 50.445 < 2e-16 \*\*\*  
## Dept95 50486.249 261.283 193.224 < 2e-16 \*\*\*  
## Dept96 -3825.393 281.895 -13.570 < 2e-16 \*\*\*  
## Dept97 -5263.239 261.981 -20.090 < 2e-16 \*\*\*  
## Dept98 -13043.480 267.678 -48.728 < 2e-16 \*\*\*  
## Dept99 -25353.584 539.900 -46.960 < 2e-16 \*\*\*  
## IsHolidayTRUE -44.588 133.399 -0.334 0.738193   
## week\_number2 -926.910 273.415 -3.390 0.000699 \*\*\*  
## week\_number3 -957.536 273.173 -3.505 0.000456 \*\*\*  
## week\_number4 -1605.413 273.367 -5.873 4.29e-09 \*\*\*  
## week\_number5 778.083 272.466 2.856 0.004294 \*\*   
## week\_number6 1862.791 264.596 7.040 1.92e-12 \*\*\*  
## week\_number7 1668.909 253.312 6.588 4.45e-11 \*\*\*  
## week\_number8 501.426 249.785 2.007 0.044705 \*   
## week\_number9 630.057 250.387 2.516 0.011859 \*   
## week\_number10 660.608 251.734 2.624 0.008685 \*\*   
## week\_number11 401.636 257.771 1.558 0.119208   
## week\_number12 -128.736 260.075 -0.495 0.620605   
## week\_number13 -182.171 257.990 -0.706 0.480116   
## week\_number14 1778.008 263.664 6.743 1.55e-11 \*\*\*  
## week\_number15 352.343 267.501 1.317 0.187784   
## week\_number16 345.057 270.269 1.277 0.201703   
## week\_number17 -504.467 272.823 -1.849 0.064450 .   
## week\_number18 240.686 272.307 0.884 0.376764   
## week\_number19 408.515 281.374 1.452 0.146542   
## week\_number20 -39.957 278.760 -0.143 0.886022   
## week\_number21 -6.731 291.047 -0.023 0.981550   
## week\_number22 748.421 299.111 2.502 0.012345 \*   
## week\_number23 986.431 301.550 3.271 0.001071 \*\*   
## week\_number24 598.464 304.407 1.966 0.049299 \*   
## week\_number25 239.197 307.995 0.777 0.437379   
## week\_number26 32.465 317.905 0.102 0.918661   
## week\_number27 837.429 317.219 2.640 0.008293 \*\*   
## week\_number28 6.238 318.258 0.020 0.984363   
## week\_number29 -452.362 321.519 -1.407 0.159442   
## week\_number30 -1049.690 322.283 -3.257 0.001126 \*\*   
## week\_number31 -63.579 320.956 -0.198 0.842974   
## week\_number32 120.687 322.061 0.375 0.707859   
## week\_number33 49.434 317.648 0.156 0.876327   
## week\_number34 64.336 313.466 0.205 0.837384   
## week\_number35 -276.781 313.788 -0.882 0.377742   
## week\_number36 276.653 322.737 0.857 0.391332   
## week\_number37 -658.275 299.113 -2.201 0.027754 \*   
## week\_number38 -985.695 290.024 -3.399 0.000677 \*\*\*  
## week\_number39 -1309.365 294.215 -4.450 8.57e-06 \*\*\*  
## week\_number40 -181.391 282.656 -0.642 0.521044   
## week\_number41 -230.267 271.558 -0.848 0.396467   
## week\_number42 -557.138 271.168 -2.055 0.039920 \*   
## week\_number43 -335.472 267.563 -1.254 0.209914   
## week\_number44 399.799 283.450 1.410 0.158401   
## week\_number45 860.305 279.251 3.081 0.002065 \*\*   
## week\_number46 697.084 280.013 2.489 0.012794 \*   
## week\_number47 3615.687 286.250 12.631 < 2e-16 \*\*\*  
## week\_number48 4397.181 284.006 15.483 < 2e-16 \*\*\*  
## week\_number49 2896.590 271.756 10.659 < 2e-16 \*\*\*  
## week\_number50 4599.506 271.414 16.946 < 2e-16 \*\*\*  
## week\_number51 8418.520 272.174 30.931 < 2e-16 \*\*\*  
## week\_number52 6734.002 279.847 24.063 < 2e-16 \*\*\*  
## week\_number53 -1116.144 361.842 -3.085 0.002038 \*\*   
## TypeB NA NA NA NA   
## TypeC NA NA NA NA   
## Size NA NA NA NA   
## Temperature 25.205 4.807 5.244 1.58e-07 \*\*\*  
## Fuel\_Price -290.104 89.893 -3.227 0.001250 \*\*   
## CPI -21.883 12.331 -1.775 0.075957 .   
## Unemployment -549.048 50.140 -10.950 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13240 on 337074 degrees of freedom  
## Multiple R-squared: 0.6619, Adjusted R-squared: 0.6617   
## F-statistic: 3645 on 181 and 337074 DF, p-value: < 2.2e-16

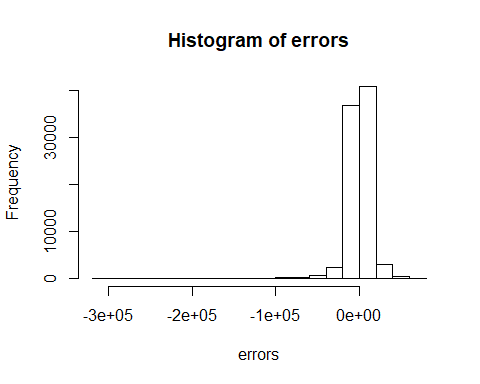
# CPI & Size not statistically significant; remove  
train$CPI <- NULL  
train$Size <- NULL  
  
model <- lm(Weekly\_Sales ~ ., data = train)  
summary(model)

##   
## Call:  
## lm(formula = Weekly\_Sales ~ ., data = train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -65373 -5799 -472 5060 627397   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 29720.632 616.021 48.246 < 2e-16 \*\*\*  
## Store2 5151.689 206.909 24.898 < 2e-16 \*\*\*  
## Store3 -16604.723 215.359 -77.102 < 2e-16 \*\*\*  
## Store4 6792.694 222.161 30.576 < 2e-16 \*\*\*  
## Store5 -18565.100 222.596 -83.403 < 2e-16 \*\*\*  
## Store6 -332.208 211.786 -1.569 0.116741   
## Store7 -12783.343 254.116 -50.305 < 2e-16 \*\*\*  
## Store8 -9606.326 221.844 -43.302 < 2e-16 \*\*\*  
## Store9 -14916.734 225.470 -66.158 < 2e-16 \*\*\*  
## Store10 5474.623 214.060 25.575 < 2e-16 \*\*\*  
## Store11 -3015.861 209.842 -14.372 < 2e-16 \*\*\*  
## Store12 -4391.492 342.463 -12.823 < 2e-16 \*\*\*  
## Store13 6187.046 219.194 28.226 < 2e-16 \*\*\*  
## Store14 7709.416 221.296 34.838 < 2e-16 \*\*\*  
## Store15 -12371.126 227.039 -54.489 < 2e-16 \*\*\*  
## Store16 -14610.218 245.308 -59.559 < 2e-16 \*\*\*  
## Store17 -8863.649 238.759 -37.124 < 2e-16 \*\*\*  
## Store18 -5206.211 230.706 -22.566 < 2e-16 \*\*\*  
## Store19 -815.745 225.020 -3.625 0.000289 \*\*\*  
## Store20 8452.773 217.036 38.946 < 2e-16 \*\*\*  
## Store21 -11051.619 210.244 -52.566 < 2e-16 \*\*\*  
## Store22 -6427.229 222.450 -28.893 < 2e-16 \*\*\*  
## Store23 -3015.908 260.478 -11.578 < 2e-16 \*\*\*  
## Store24 -1863.784 225.974 -8.248 < 2e-16 \*\*\*  
## Store25 -11453.121 223.854 -51.163 < 2e-16 \*\*\*  
## Store26 -6925.866 241.837 -28.639 < 2e-16 \*\*\*  
## Store27 3744.319 217.546 17.212 < 2e-16 \*\*\*  
## Store28 -325.285 341.841 -0.952 0.341317   
## Store29 -12806.664 247.390 -51.767 < 2e-16 \*\*\*  
## Store30 -17840.268 228.480 -78.082 < 2e-16 \*\*\*  
## Store31 -2241.859 207.086 -10.826 < 2e-16 \*\*\*  
## Store32 -4471.088 224.157 -19.946 < 2e-16 \*\*\*  
## Store33 -21118.256 246.073 -85.821 < 2e-16 \*\*\*  
## Store34 -7458.129 238.264 -31.302 < 2e-16 \*\*\*  
## Store35 -7670.085 226.264 -33.899 < 2e-16 \*\*\*  
## Store36 -19288.665 239.422 -80.563 < 2e-16 \*\*\*  
## Store37 -16162.072 228.418 -70.757 < 2e-16 \*\*\*  
## Store38 -15824.198 352.630 -44.875 < 2e-16 \*\*\*  
## Store39 -1276.562 208.780 -6.114 9.70e-10 \*\*\*  
## Store40 -9080.172 262.431 -34.600 < 2e-16 \*\*\*  
## Store41 -3793.877 230.373 -16.468 < 2e-16 \*\*\*  
## Store42 -14328.341 236.388 -60.614 < 2e-16 \*\*\*  
## Store43 -12559.810 255.847 -49.091 < 2e-16 \*\*\*  
## Store44 -20651.567 242.608 -85.123 < 2e-16 \*\*\*  
## Store45 -9853.338 223.542 -44.078 < 2e-16 \*\*\*  
## Dept2 24444.180 260.940 93.678 < 2e-16 \*\*\*  
## Dept3 -7494.173 261.093 -28.703 < 2e-16 \*\*\*  
## Dept4 6601.334 260.812 25.311 < 2e-16 \*\*\*  
## Dept5 1925.106 261.571 7.360 1.85e-13 \*\*\*  
## Dept6 -15385.578 266.752 -57.677 < 2e-16 \*\*\*  
## Dept7 4821.198 260.162 18.532 < 2e-16 \*\*\*  
## Dept8 10992.369 260.877 42.136 < 2e-16 \*\*\*  
## Dept9 694.809 261.494 2.657 0.007883 \*\*   
## Dept10 -1025.152 261.283 -3.924 8.73e-05 \*\*\*  
## Dept11 -4784.258 261.451 -18.299 < 2e-16 \*\*\*  
## Dept12 -15227.131 262.179 -58.079 < 2e-16 \*\*\*  
## Dept13 11339.980 261.194 43.416 < 2e-16 \*\*\*  
## Dept14 -4430.857 261.708 -16.931 < 2e-16 \*\*\*  
## Dept16 -5074.553 260.074 -19.512 < 2e-16 \*\*\*  
## Dept17 -9112.762 260.485 -34.984 < 2e-16 \*\*\*  
## Dept18 -12833.746 279.555 -45.908 < 2e-16 \*\*\*  
## Dept19 -20443.353 296.262 -69.004 < 2e-16 \*\*\*  
## Dept20 -14411.431 266.301 -54.117 < 2e-16 \*\*\*  
## Dept21 -14255.944 260.713 -54.681 < 2e-16 \*\*\*  
## Dept22 -10252.576 271.308 -37.789 < 2e-16 \*\*\*  
## Dept23 3478.460 269.330 12.915 < 2e-16 \*\*\*  
## Dept24 -14841.647 273.655 -54.235 < 2e-16 \*\*\*  
## Dept25 -10465.247 261.868 -39.964 < 2e-16 \*\*\*  
## Dept26 -12864.485 268.378 -47.934 < 2e-16 \*\*\*  
## Dept27 -19284.663 269.939 -71.441 < 2e-16 \*\*\*  
## Dept28 -19289.148 264.539 -72.916 < 2e-16 \*\*\*  
## Dept29 -15868.801 275.614 -57.576 < 2e-16 \*\*\*  
## Dept30 -17434.991 274.961 -63.409 < 2e-16 \*\*\*  
## Dept31 -17723.114 264.515 -67.002 < 2e-16 \*\*\*  
## Dept32 -13339.980 265.706 -50.206 < 2e-16 \*\*\*  
## Dept33 -14873.711 273.927 -54.298 < 2e-16 \*\*\*  
## Dept34 -5864.620 273.871 -21.414 < 2e-16 \*\*\*  
## Dept35 -18597.720 274.514 -67.748 < 2e-16 \*\*\*  
## Dept36 -19452.103 275.694 -70.557 < 2e-16 \*\*\*  
## Dept37 -22464.867 345.683 -64.987 < 2e-16 \*\*\*  
## Dept38 42039.348 261.156 160.974 < 2e-16 \*\*\*  
## Dept39 -28382.420 3827.268 -7.416 1.21e-13 \*\*\*  
## Dept40 25622.383 261.169 98.107 < 2e-16 \*\*\*  
## Dept41 -19280.147 273.575 -70.475 < 2e-16 \*\*\*  
## Dept42 -14190.790 261.079 -54.354 < 2e-16 \*\*\*  
## Dept43 -16068.815 4418.113 -3.637 0.000276 \*\*\*  
## Dept44 -16485.298 271.561 -60.706 < 2e-16 \*\*\*  
## Dept45 -22358.147 382.271 -58.488 < 2e-16 \*\*\*  
## Dept46 605.380 260.510 2.324 0.020135 \*   
## Dept47 -22845.447 606.331 -37.678 < 2e-16 \*\*\*  
## Dept48 -23925.348 400.005 -59.813 < 2e-16 \*\*\*  
## Dept49 -14454.267 288.281 -50.139 < 2e-16 \*\*\*  
## Dept50 -23535.466 417.278 -56.402 < 2e-16 \*\*\*  
## Dept51 -21343.140 439.922 -48.516 < 2e-16 \*\*\*  
## Dept52 -17761.333 263.048 -67.521 < 2e-16 \*\*\*  
## Dept54 -21895.752 282.669 -77.461 < 2e-16 \*\*\*  
## Dept55 -9913.353 270.519 -36.646 < 2e-16 \*\*\*  
## Dept56 -16520.615 267.268 -61.813 < 2e-16 \*\*\*  
## Dept58 -19289.417 290.393 -66.425 < 2e-16 \*\*\*  
## Dept59 -19134.436 263.898 -72.507 < 2e-16 \*\*\*  
## Dept60 -18910.669 267.536 -70.684 < 2e-16 \*\*\*  
## Dept65 27267.583 1246.378 21.877 < 2e-16 \*\*\*  
## Dept67 -11915.529 261.051 -45.644 < 2e-16 \*\*\*  
## Dept71 -15804.946 275.059 -57.460 < 2e-16 \*\*\*  
## Dept72 30668.623 264.883 115.782 < 2e-16 \*\*\*  
## Dept74 -5483.810 260.584 -21.044 < 2e-16 \*\*\*  
## Dept77 -24288.776 1195.439 -20.318 < 2e-16 \*\*\*  
## Dept78 -23029.438 993.685 -23.176 < 2e-16 \*\*\*  
## Dept79 2692.309 260.611 10.331 < 2e-16 \*\*\*  
## Dept80 -7765.982 266.795 -29.108 < 2e-16 \*\*\*  
## Dept81 -3867.612 260.536 -14.845 < 2e-16 \*\*\*  
## Dept82 -3895.769 260.863 -14.934 < 2e-16 \*\*\*  
## Dept83 -16646.515 266.792 -62.395 < 2e-16 \*\*\*  
## Dept85 -17776.159 265.136 -67.045 < 2e-16 \*\*\*  
## Dept87 -6016.395 261.839 -22.977 < 2e-16 \*\*\*  
## Dept90 26009.379 261.067 99.627 < 2e-16 \*\*\*  
## Dept91 14203.242 261.117 54.394 < 2e-16 \*\*\*  
## Dept92 56003.122 259.717 215.632 < 2e-16 \*\*\*  
## Dept93 7071.386 266.894 26.495 < 2e-16 \*\*\*  
## Dept94 13588.104 269.386 50.441 < 2e-16 \*\*\*  
## Dept95 50486.729 261.284 193.226 < 2e-16 \*\*\*  
## Dept96 -3826.120 281.896 -13.573 < 2e-16 \*\*\*  
## Dept97 -5263.510 261.982 -20.091 < 2e-16 \*\*\*  
## Dept98 -13044.848 267.678 -48.733 < 2e-16 \*\*\*  
## Dept99 -25369.678 539.826 -46.996 < 2e-16 \*\*\*  
## IsHolidayTRUE -53.650 133.302 -0.402 0.687340   
## week\_number2 -925.768 273.415 -3.386 0.000709 \*\*\*  
## week\_number3 -956.453 273.173 -3.501 0.000463 \*\*\*  
## week\_number4 -1605.907 273.368 -5.875 4.24e-09 \*\*\*  
## week\_number5 776.076 272.465 2.848 0.004395 \*\*   
## week\_number6 1869.029 264.573 7.064 1.62e-12 \*\*\*  
## week\_number7 1673.236 253.301 6.606 3.96e-11 \*\*\*  
## week\_number8 503.010 249.784 2.014 0.044034 \*   
## week\_number9 639.234 250.335 2.554 0.010665 \*   
## week\_number10 676.448 251.577 2.689 0.007171 \*\*   
## week\_number11 422.757 257.497 1.642 0.100633   
## week\_number12 -105.881 259.757 -0.408 0.683556   
## week\_number13 -156.080 257.571 -0.606 0.544535   
## week\_number14 1811.031 263.007 6.886 5.75e-12 \*\*\*  
## week\_number15 390.781 266.623 1.466 0.142740   
## week\_number16 386.754 269.246 1.436 0.150880   
## week\_number17 -463.331 271.838 -1.704 0.088300 .   
## week\_number18 282.628 271.281 1.042 0.297491   
## week\_number19 451.973 280.307 1.612 0.106871   
## week\_number20 2.453 277.734 0.009 0.992954   
## week\_number21 32.512 290.207 0.112 0.910800   
## week\_number22 782.452 298.497 2.621 0.008760 \*\*   
## week\_number23 1015.776 301.098 3.374 0.000742 \*\*\*  
## week\_number24 623.822 304.072 2.052 0.040213 \*   
## week\_number25 259.733 307.778 0.844 0.398727   
## week\_number26 51.362 317.728 0.162 0.871578   
## week\_number27 854.385 317.076 2.695 0.007048 \*\*   
## week\_number28 24.431 318.094 0.077 0.938779   
## week\_number29 -432.645 321.328 -1.346 0.178165   
## week\_number30 -1027.071 322.032 -3.189 0.001426 \*\*   
## week\_number31 -42.106 320.729 -0.131 0.895552   
## week\_number32 142.135 321.836 0.442 0.658752   
## week\_number33 71.300 317.410 0.225 0.822267   
## week\_number34 84.770 313.255 0.271 0.786692   
## week\_number35 -257.470 313.600 -0.821 0.411639   
## week\_number36 302.868 322.400 0.939 0.347519   
## week\_number37 -638.692 298.910 -2.137 0.032620 \*   
## week\_number38 -971.807 289.919 -3.352 0.000802 \*\*\*  
## week\_number39 -1300.813 294.176 -4.422 9.79e-06 \*\*\*  
## week\_number40 -173.858 282.625 -0.615 0.538454   
## week\_number41 -225.369 271.545 -0.830 0.406566   
## week\_number42 -549.628 271.135 -2.027 0.042649 \*   
## week\_number43 -332.184 267.558 -1.242 0.214407   
## week\_number44 408.053 283.413 1.440 0.149929   
## week\_number45 863.750 279.245 3.093 0.001981 \*\*   
## week\_number46 700.416 280.008 2.501 0.012370 \*   
## week\_number47 3621.161 286.234 12.651 < 2e-16 \*\*\*  
## week\_number48 4400.288 284.001 15.494 < 2e-16 \*\*\*  
## week\_number49 2888.590 271.720 10.631 < 2e-16 \*\*\*  
## week\_number50 4593.695 271.395 16.926 < 2e-16 \*\*\*  
## week\_number51 8410.419 272.136 30.905 < 2e-16 \*\*\*  
## week\_number52 6728.074 279.828 24.044 < 2e-16 \*\*\*  
## week\_number53 -1080.607 361.289 -2.991 0.002781 \*\*   
## TypeB NA NA NA NA   
## TypeC NA NA NA NA   
## Temperature 24.824 4.802 5.170 2.35e-07 \*\*\*  
## Fuel\_Price -393.638 68.390 -5.756 8.63e-09 \*\*\*  
## Unemployment -515.065 46.339 -11.115 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13240 on 337075 degrees of freedom  
## Multiple R-squared: 0.6619, Adjusted R-squared: 0.6617   
## F-statistic: 3665 on 180 and 337075 DF, p-value: < 2.2e-16

prediction <- predict(model, interval = "prediction", newdata = test)

## Warning in predict.lm(model, interval = "prediction", newdata = test):  
## prediction from a rank-deficient fit may be misleading

# Errors  
errors <- prediction[, "fit"] - test$Weekly\_Sales  
hist(errors)



# Root Mean Squared Error & Percentage of cases with less than 25% error   
rmse <- sqrt(sum((prediction[,"fit"] - test$Weekly\_Sales)^2)/nrow(test))  
rel\_change <- 1 - ((test$Weekly\_Sales - abs(errors)) / test$Weekly\_Sales)  
pred25 <- table(rel\_change<0.25)["TRUE"] / nrow(test)  
paste("RMSE:", rmse)

## [1] "RMSE: 13129.726377009"

paste("PRED(25):", pred25)

## [1] "PRED(25): 0.258830087529948"