Final_INF 6480

Leslie Garcia

2025-12-13

Load packages

```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.1     ✔ stringr   1.5.2
## ✔ ggplot2   4.0.0     ✔ tibble    3.3.0
## ✔ lubridate 1.9.4     ✔ tidyr     1.3.1
## ✔ purrr     1.1.0
## ── Conflicts ──────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```r
library(lubridate)
library(broom)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```r
library(dplyr)
```

Read data file

```
border <- read.csv("Border_Crossing_Entry_Data_072025.csv")
```

Filter to California ports data only

```
border_ca <- border %>%
  filter(State == "California")

glimpse(border_ca)
```

```
## Rows: 22,015
## Columns: 10
## $ Port.Name <chr> "Calexico East", "Tecate", "Otay Mesa", "Calexico East", "Ca…
## $ State     <chr> "California", "California", "California", "California", "Cal…
## $ Port.Code <int> 2507, 2505, 2506, 2507, 2507, 2506, 2506, 2505, 2506, 2582, …
## $ Border    <chr> "US-Mexico Border", "US-Mexico Border", "US-Mexico Border", …
## $ Date      <chr> "Jan 2024", "Jan 2024", "Jan 2024", "Jan 2024", "Jan 2024", …
## $ Measure   <chr> "Personal Vehicle Passengers", "Truck Containers Loaded", "T…
## $ Value     <int> 435768, 5306, 101560, 18199, 247739, 79734, 283, 1815, 2, 20…
## $ Latitude  <dbl> 32.673, 32.576, 32.550, 32.673, 32.673, 32.550, 32.550, 32.5…
## $ Longitude <dbl> -115.388, -116.626, -116.939, -115.388, -115.388, -116.939, …
## $ Point     <chr> "POINT (-115.388018 32.673389)", "POINT (-116.626216 32.5764…
```

```
nrow(border_ca) #should be > 0
```

```
## [1] 22015
```

```
table(border_ca[['State']])
```

```
##
## California
##      22015
```

Clean data and create holiday indicators

```r
border_ca_clean <- border_ca %>%
mutate (
  Date = my(Date),
  Year = year(Date),
  Month = month(Date),
  Month_Name = month(Date, label = TRUE, abbr = TRUE),

# ORIGINAL holiday definition (July, Nov, Dec)
HolidayMonth = if_else(Month %in% c(7, 11, 12),
                        "Holiday", "Non-Holiday"),
HolidayMonth = factor(HolidayMonth),

# NEW: Spring/Easter months (March & April)
Spring_Easter_Month = Month %in% c(3, 4),

# NEW: Memorial Day month (May)
Memorial_Month = Month == 5,

# OPTIONAL: broader "major travel" months
# (Spring break/Easter, Memorial Day, July 4, Thanksgiving, Christmas/New Year)
Holiday_MajorTravel = if_else(Month %in% c(3, 4, 5, 7, 11, 12),
                               "Holiday", "Non-Holiday"),
Holiday_MajorTravel = factor(Holiday_MajorTravel),

Border = factor(Border),
Measure = factor(Measure),
`Port.Name` = factor(`Port.Name`),
Value_log = log1p(Value),
Lat_scaled = scale(Latitude)[,1],
Lon_scaled = scale(Longitude)[,1]

) %>%
filter(!is.na(Date), !is.na(Value))

summary(border_ca_clean)
```

```
##              Port.Name           State             Port.Code
##  Andrade           :3284   Length:22015      Min.   :2502
##  Calexico          :3285   Class :character   1st Qu.:2503
##  Calexico East     :4198   Mode  :character   Median :2505
##  Cross Border Xpress:  93                     Mean   :2505
##  Otay Mesa         :4042                      3rd Qu.:2506
##  San Ysidro        :3492                      Max.   :2582
##  Tecate            :3621
##               Border            Date
##  US-Mexico Border:22015   Min.   :1996-01-01
##                           1st Qu.:2002-05-01
##                           Median :2008-09-01
##                           Mean   :2009-01-25
##                           3rd Qu.:2015-02-01
##                           Max.   :2025-06-01
##
##                          Measure          Value             Latitude
##  Pedestrians               : 2210   Min.   :       0   Min.   :32.54
##  Personal Vehicle Passengers: 2123  1st Qu.:       0   1st Qu.:32.55
##  Personal Vehicles         : 2123   Median :    1063   Median :32.58
##  Truck Containers Empty    : 1802   Mean   :  140650   Mean   :32.62
##  Trucks                    : 1802   3rd Qu.:   63606   3rd Qu.:32.67
##  Truck Containers Loaded   : 1799   Max.   :3755585    Max.   :32.72
##  (Other)                   :10156
##    Longitude           Point              Year          Month
##  Min.   :-117.0   Length:22015       Min.   :1996   Min.   : 1.000
##  1st Qu.:-116.9   Class :character   1st Qu.:2002   1st Qu.: 3.000
##  Median :-116.6   Mode  :character   Median :2008   Median : 6.000
##  Mean   :-116.0                      Mean   :2009   Mean   : 6.444
##  3rd Qu.:-115.4                      3rd Qu.:2015   3rd Qu.: 9.000
##  Max.   :-114.7                      Max.   :2025   Max.   :12.000
##
##    Month_Name         HolidayMonth    Spring_Easter_Month Memorial_Month
##  Feb    : 1872   Holiday    : 5421   Mode :logical        Mode :logical
##  Mar    : 1868   Non-Holiday:16594   FALSE:18281          FALSE:20148
##  May    : 1867                       TRUE :3734           TRUE :1867
##  Apr    : 1866
##  Jun    : 1865
##  Jan    : 1862
##  (Other):10815
##   Holiday_MajorTravel   Value_log          Lat_scaled         Lon_scaled
##  Holiday    :11022   Min.   : 0.000   Min.   :-1.1397   Min.   :-1.1251
##  Non-Holiday:10993   1st Qu.: 0.000   1st Qu.:-1.0361   1st Qu.:-1.0222
##                      Median : 6.970   Median :-0.6514   Median :-0.6644
##                      Mean   : 6.463   Mean   : 0.0000   Mean   : 0.0000
##                      3rd Qu.:11.060   3rd Qu.: 0.7838   3rd Qu.: 0.7511
##                      Max.   :15.139   Max.   : 1.4496   Max.   : 1.5057
##
```

Check for months labeled as holiday

```
border_ca_clean %>%
  count(Year, Month, Month_Name, HolidayMonth, Holiday_MajorTravel,
        Spring_Easter_Month, Memorial_Month) %>%
  arrange(Year, Month) %>%
  head(50)
```

```
##    Year Month Month_Name HolidayMonth Holiday_MajorTravel Spring_Easter_Month
## 1  1996     1        Jan  Non-Holiday         Non-Holiday               FALSE
## 2  1996     2        Feb  Non-Holiday         Non-Holiday               FALSE
## 3  1996     3        Mar  Non-Holiday             Holiday                TRUE
## 4  1996     4        Apr  Non-Holiday             Holiday                TRUE
## 5  1996     5        May  Non-Holiday             Holiday               FALSE
## 6  1996     6        Jun  Non-Holiday         Non-Holiday               FALSE
## 7  1996     7        Jul      Holiday             Holiday               FALSE
## 8  1996     8        Aug  Non-Holiday         Non-Holiday               FALSE
## 9  1996     9        Sep  Non-Holiday         Non-Holiday               FALSE
## 10 1996    10        Oct  Non-Holiday         Non-Holiday               FALSE
## 11 1996    11        Nov      Holiday             Holiday               FALSE
## 12 1996    12        Dec      Holiday             Holiday               FALSE
## 13 1997     1        Jan  Non-Holiday         Non-Holiday               FALSE
## 14 1997     2        Feb  Non-Holiday         Non-Holiday               FALSE
## 15 1997     3        Mar  Non-Holiday             Holiday                TRUE
## 16 1997     4        Apr  Non-Holiday             Holiday                TRUE
## 17 1997     5        May  Non-Holiday             Holiday               FALSE
## 18 1997     6        Jun  Non-Holiday         Non-Holiday               FALSE
## 19 1997     7        Jul      Holiday             Holiday               FALSE
## 20 1997     8        Aug  Non-Holiday         Non-Holiday               FALSE
## 21 1997     9        Sep  Non-Holiday         Non-Holiday               FALSE
## 22 1997    10        Oct  Non-Holiday         Non-Holiday               FALSE
## 23 1997    11        Nov      Holiday             Holiday               FALSE
## 24 1997    12        Dec      Holiday             Holiday               FALSE
## 25 1998     1        Jan  Non-Holiday         Non-Holiday               FALSE
## 26 1998     2        Feb  Non-Holiday         Non-Holiday               FALSE
## 27 1998     3        Mar  Non-Holiday             Holiday                TRUE
## 28 1998     4        Apr  Non-Holiday             Holiday                TRUE
## 29 1998     5        May  Non-Holiday             Holiday               FALSE
## 30 1998     6        Jun  Non-Holiday         Non-Holiday               FALSE
## 31 1998     7        Jul      Holiday             Holiday               FALSE
## 32 1998     8        Aug  Non-Holiday         Non-Holiday               FALSE
## 33 1998     9        Sep  Non-Holiday         Non-Holiday               FALSE
## 34 1998    10        Oct  Non-Holiday         Non-Holiday               FALSE
## 35 1998    11        Nov      Holiday             Holiday               FALSE
## 36 1998    12        Dec      Holiday             Holiday               FALSE
## 37 1999     1        Jan  Non-Holiday         Non-Holiday               FALSE
## 38 1999     2        Feb  Non-Holiday         Non-Holiday               FALSE
## 39 1999     3        Mar  Non-Holiday             Holiday                TRUE
## 40 1999     4        Apr  Non-Holiday             Holiday                TRUE
## 41 1999     5        May  Non-Holiday             Holiday               FALSE
## 42 1999     6        Jun  Non-Holiday         Non-Holiday               FALSE
## 43 1999     7        Jul      Holiday             Holiday               FALSE
## 44 1999     8        Aug  Non-Holiday         Non-Holiday               FALSE
## 45 1999     9        Sep  Non-Holiday         Non-Holiday               FALSE
## 46 1999    10        Oct  Non-Holiday         Non-Holiday               FALSE
## 47 1999    11        Nov      Holiday             Holiday               FALSE
## 48 1999    12        Dec      Holiday             Holiday               FALSE
## 49 2000     1        Jan  Non-Holiday         Non-Holiday               FALSE
## 50 2000     2        Feb  Non-Holiday         Non-Holiday               FALSE
##    Memorial_Month  n
## 1          FALSE 72
## 2          FALSE 72
```

```
## 3              FALSE 72
## 4              FALSE 72
## 5               TRUE 72
## 6              FALSE 72
## 7              FALSE 72
## 8              FALSE 72
## 9              FALSE 72
## 10             FALSE 72
## 11             FALSE 72
## 12             FALSE 72
## 13             FALSE 72
## 14             FALSE 72
## 15             FALSE 72
## 16             FALSE 72
## 17              TRUE 72
## 18             FALSE 72
## 19             FALSE 72
## 20             FALSE 72
## 21             FALSE 72
## 22             FALSE 72
## 23             FALSE 72
## 24             FALSE 72
## 25             FALSE 72
## 26             FALSE 72
## 27             FALSE 72
## 28             FALSE 72
## 29              TRUE 72
## 30             FALSE 72
## 31             FALSE 72
## 32             FALSE 72
## 33             FALSE 72
## 34             FALSE 72
## 35             FALSE 72
## 36             FALSE 72
## 37             FALSE 72
## 38             FALSE 72
## 39             FALSE 72
## 40             FALSE 72
## 41              TRUE 72
## 42             FALSE 72
## 43             FALSE 72
## 44             FALSE 72
## 45             FALSE 72
## 46             FALSE 72
## 47             FALSE 72
## 48             FALSE 72
## 49             FALSE 72
## 50             FALSE 72
```

-Check the counts

```
table(border_ca_clean[["HolidayMonth"]])
```

```
## 
##      Holiday Non-Holiday
##       5421       16594
```

```
table(border_ca_clean[["Holiday_MajorTravel"]])
```

```
## 
##      Holiday Non-Holiday
##      11022       10993
```
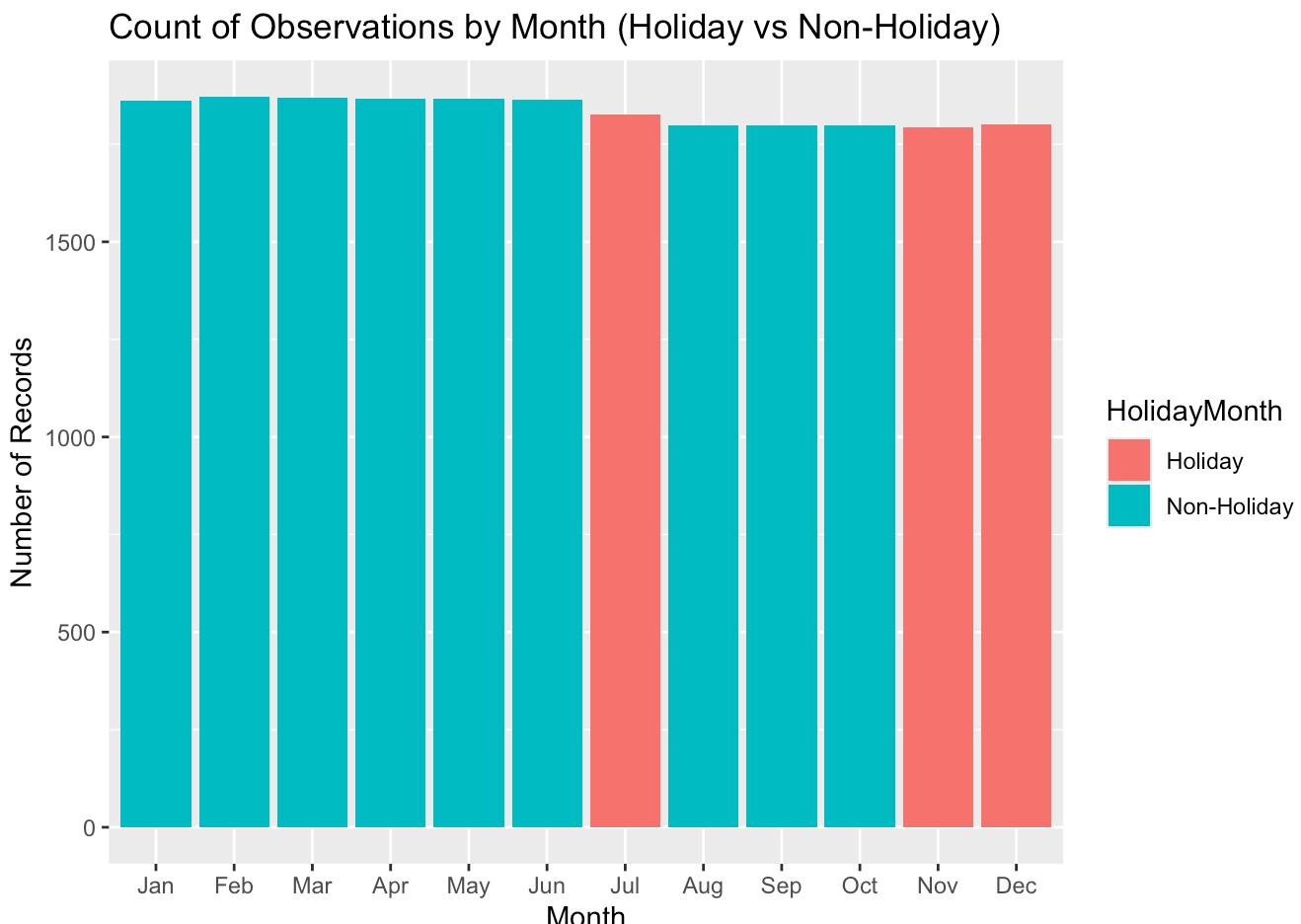
-quick barplot to check labeling

```
table(border_ca_clean[["Holiday_MajorTravel"]])
```

```
## 
##      Holiday Non-Holiday
##      11022       10993
```

#quick viz to check holiday labeling
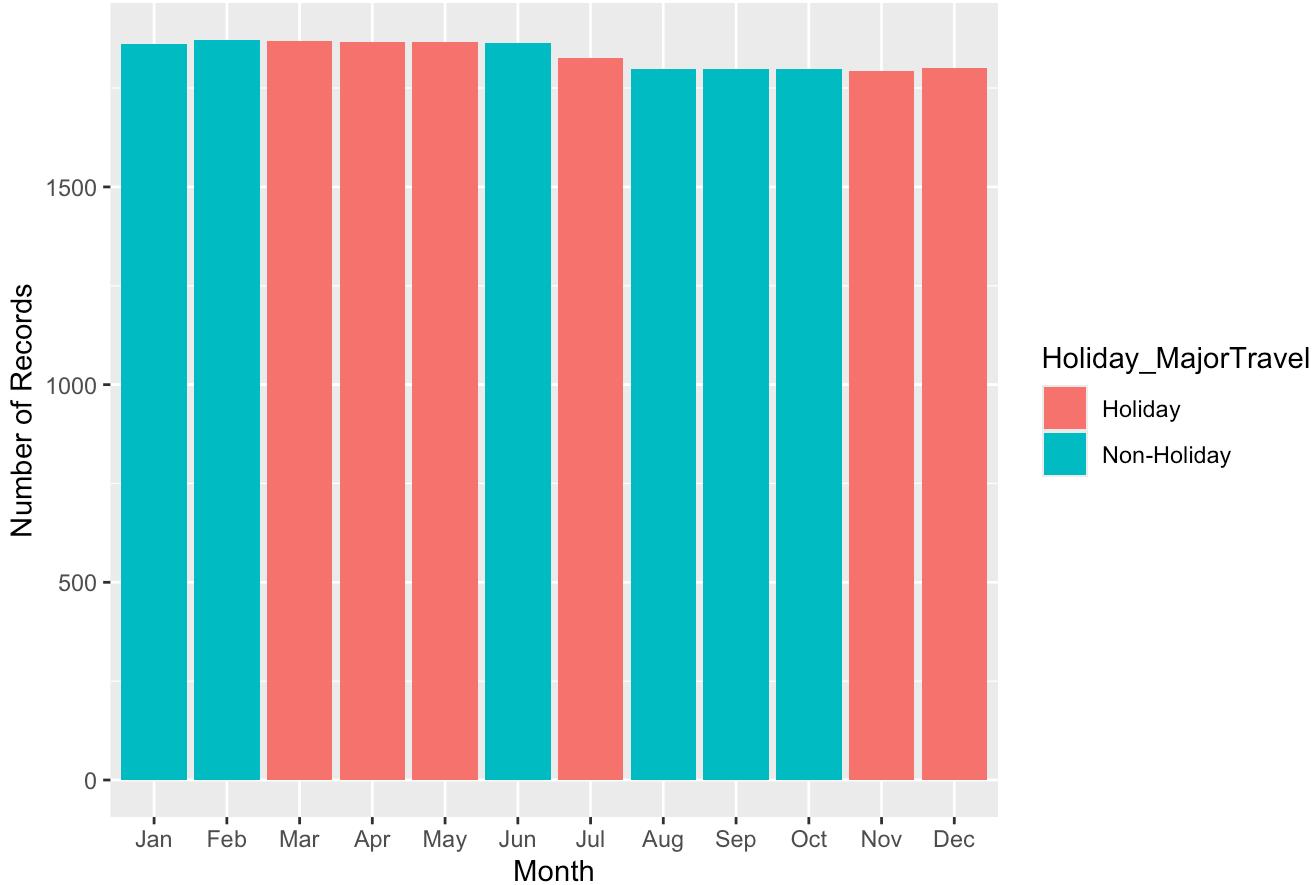
```
ggplot(border_ca_clean, aes(x = Month_Name, fill = HolidayMonth)) +
  geom_bar() +
  labs(title = "Count of Observations by Month (Holiday vs Non-Holiday)",
       x = "Month", y = "Number of Records")
```

Check major travel holiday labels

```
ggplot(border_ca_clean, aes(x = Month_Name, fill = Holiday_MajorTravel)) +
geom_bar() +
labs(title = "Observations by Month (Major Travel Holiday Definition)",
x = "Month", y = "Number of Records")
```

Observations by Month (Major Travel Holiday Definition)

Monthly totals

```
monthly_ca <- border_ca_clean %>%
group_by(Date, Month, HolidayMonth) %>%
summarise(total_value = sum(Value), .groups = "drop")

max_total <- max(monthly_ca[["total_value"]], na.rm = TRUE)
```

DESCRIPTIVE STATISTICS

```
#overall summary
desc_ca <- border_ca_clean %>%
  summarise(
    mean_value = mean(Value),
    median_value = median(Value),
    sd_value = sd(Value),
    var_value = var(Value)
  )
desc_ca
```

```
##   mean_value median_value sd_value     var_value
## 1   140649.9         1063 369900.6 136826430875
```

```
#Holiday/non-holiday
desc_ca_holiday <- border_ca_clean %>%
  group_by(HolidayMonth) %>%
  summarise(
    mean_value = mean(Value),
    median_value = median(Value),
    sd_value = sd(Value),
    n = n()
  )
desc_ca_holiday
```

```
## # A tibble: 2 × 5
##   HolidayMonth mean_value median_value sd_value     n
##   <fct>             <dbl>        <dbl>    <dbl> <int>
## 1 Holiday          143023.          944   379348.  5421
## 2 Non-Holiday      139875.         1113   366770. 16594
```

```
#Frequency of transportation modes
table(border_ca_clean[["Measure"]])
```

```
## 
##             Bus Passengers                        Buses
##                       1742                         1757
##                Pedestrians Personal Vehicle Passengers
##                       2210                         2123
##          Personal Vehicles         Rail Containers Empty
##                       2123                         1695
##     Rail Containers Loaded            Train Passengers
##                       1594                         1675
##                     Trains      Truck Containers Empty
##                       1693                         1802
##    Truck Containers Loaded                      Trucks
##                       1799                         1802
```

```
prop.table(table(border_ca_clean[["Measure"]]))
```

```
##
##          Bus Passengers                              Buses
##             0.07912787                          0.07980922
##             Pedestrians  Personal Vehicle Passengers
##             0.10038610                          0.09643425
##        Personal Vehicles          Rail Containers Empty
##             0.09643425                          0.07699296
##    Rail Containers Loaded                Train Passengers
##             0.07240518                          0.07608449
##                  Trains          Truck Containers Empty
##             0.07690211                          0.08185328
##   Truck Containers Loaded                          Trucks
##             0.08171701                          0.08185328
```
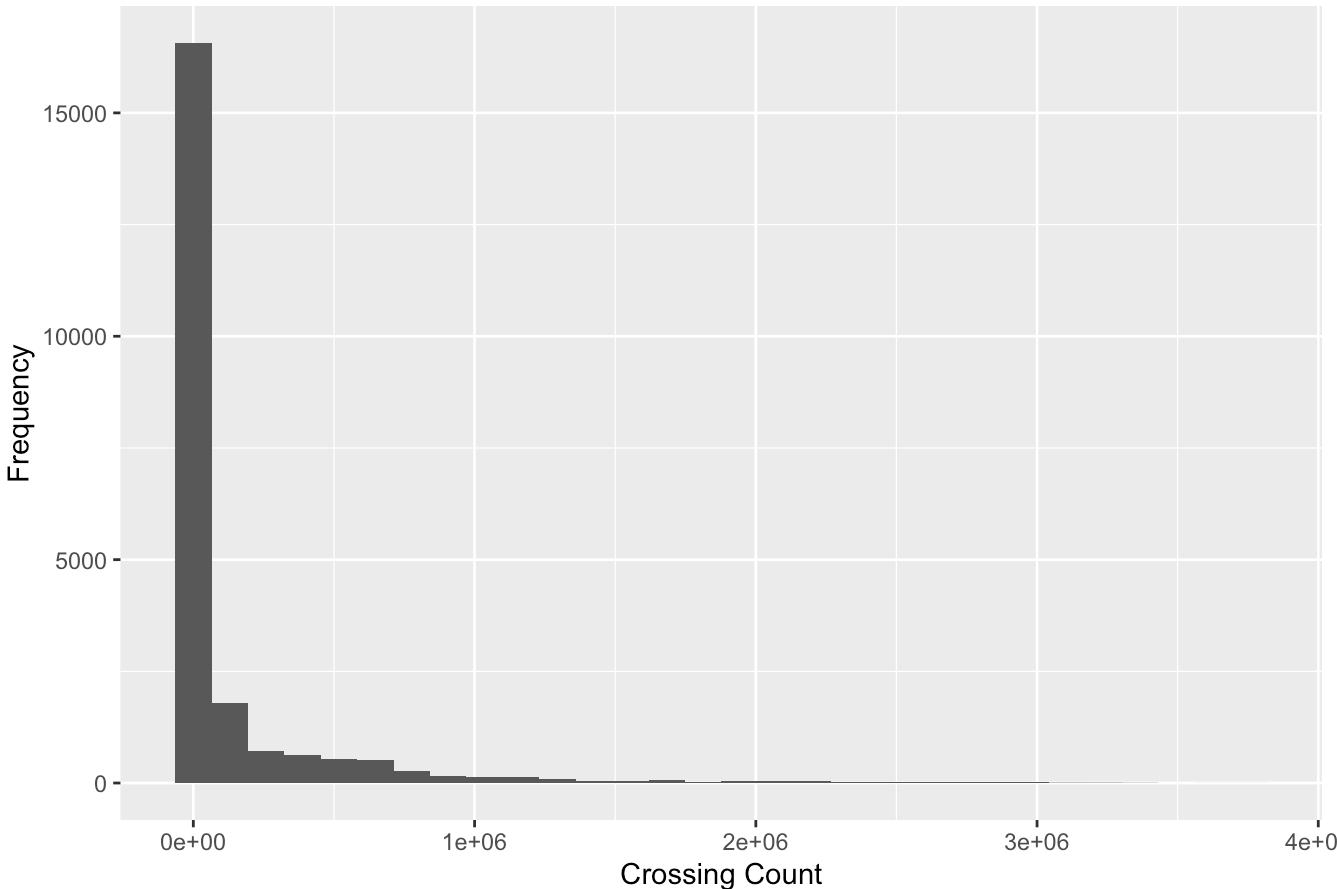
VISUALIZATION - # Histogram of crossing counts

```
ggplot(border_ca_clean, aes(Value)) +
  geom_histogram(bins = 30) +
  labs(title = "Distribution of California Border Crossing Counts",
       x = "Crossing Count", y = "Frequency")
```
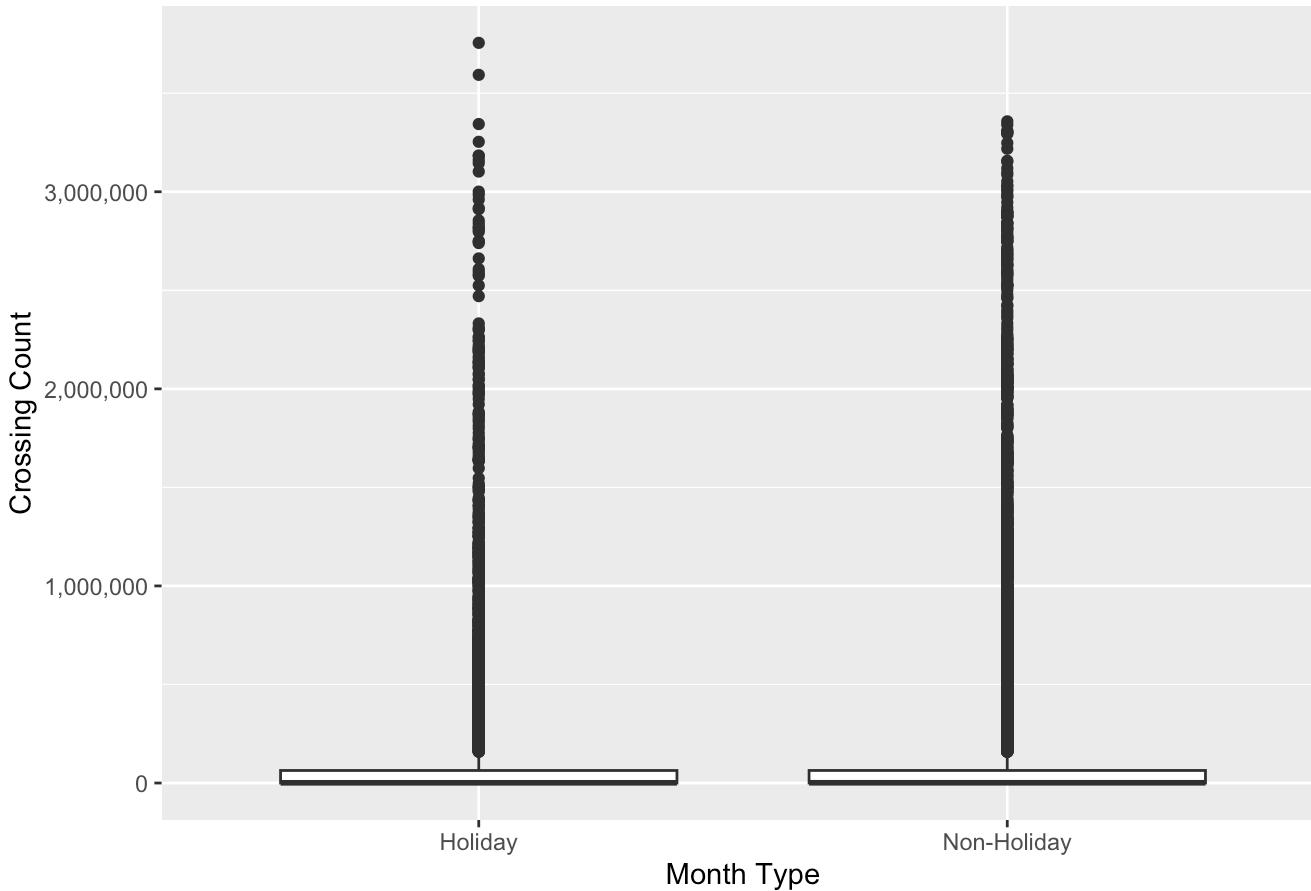

Distribution of California Border Crossing Counts

- Boxplot: Holiday vs Non-Holiday

```
ggplot(border_ca_clean, aes(HolidayMonth, Value)) +
  geom_boxplot() +
  scale_y_continuous(labels = comma) +
  labs(title = "California Border Crossings: Holiday vs Non-Holiday",
       x = "Month Type", y = "Crossing Count")
```

California Border Crossings: Holiday vs Non-Holiday
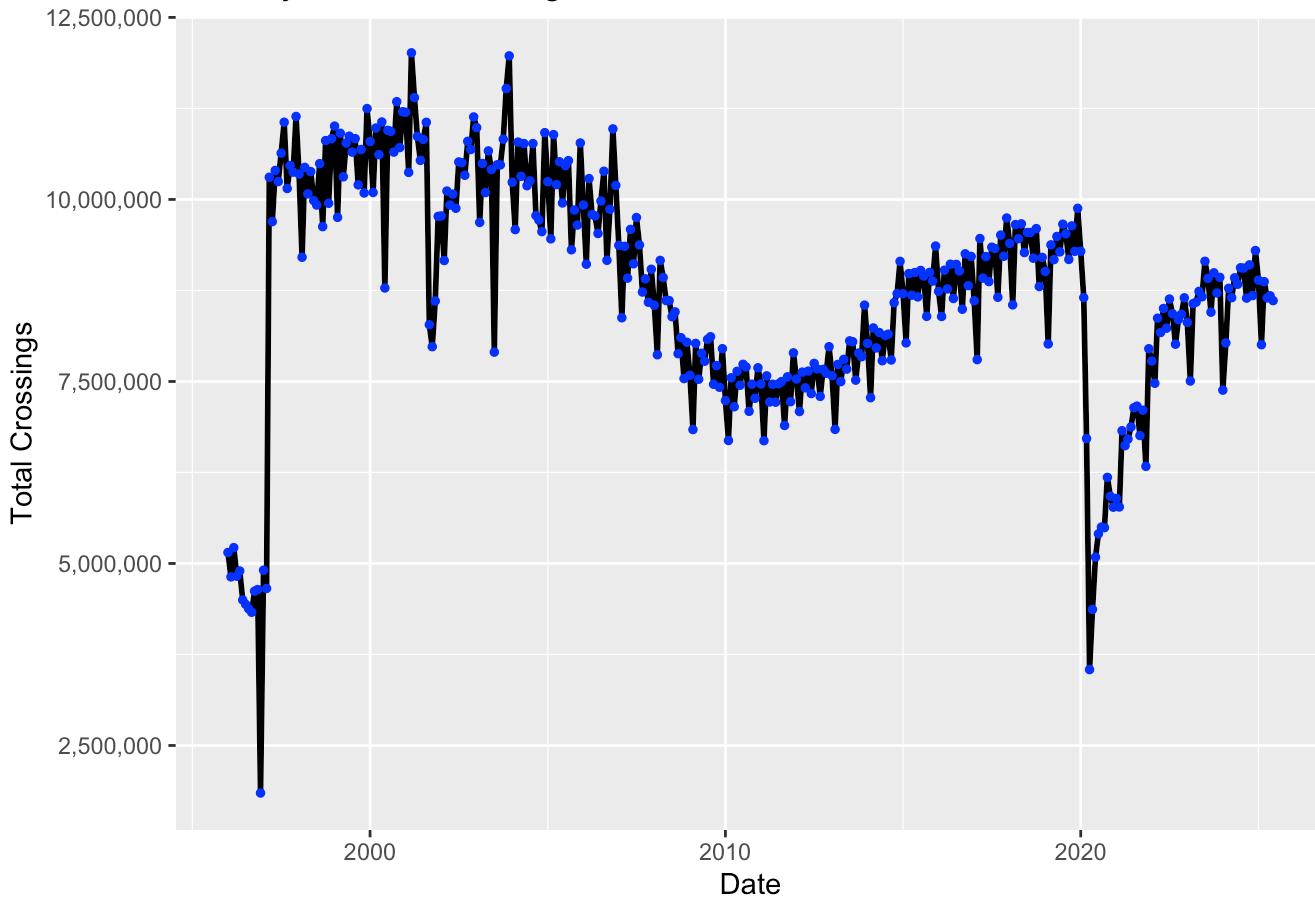

- Time series trend

```
monthly_ca <- border_ca_clean %>%
  group_by(Date) %>%
  summarise(total_value = sum(Value))

ggplot(monthly_ca, aes(Date, total_value)) +
  geom_line(size = 1) +
  geom_point(color = "blue", size = 1) +
  scale_y_continuous(labels = comma) +
  labs(title = "Monthly Border Crossings in California Over Time",
       x = "Date", y = "Total Crossings")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Monthly Border Crossings in California Over Time
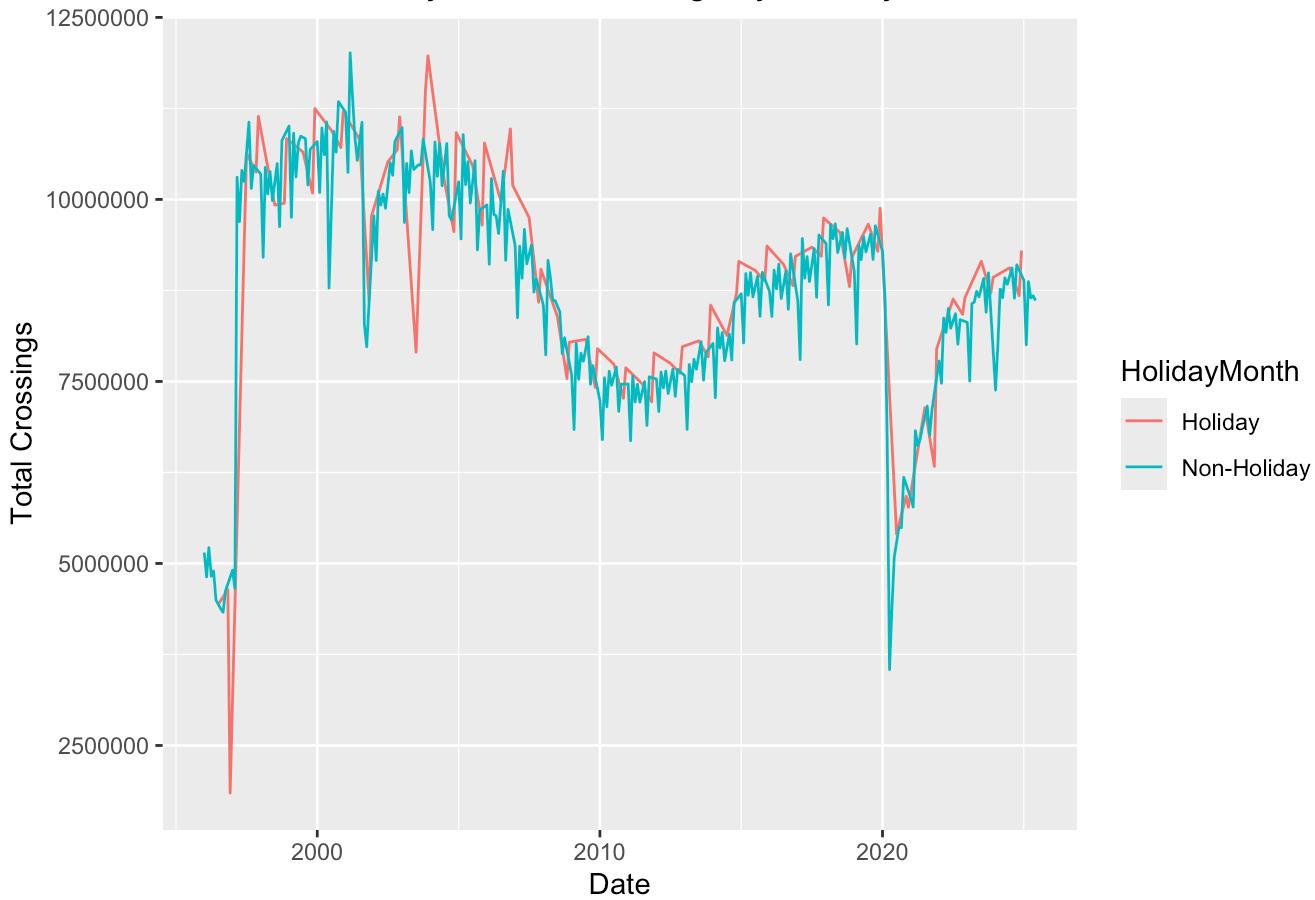
- Trend by holiday

```
monthly_ca_holiday <- border_ca_clean %>%
  group_by(Date, HolidayMonth) %>%
  summarise(total_value = sum(Value), .groups="drop")

ggplot(monthly_ca_holiday,
       aes(Date, total_value, color = HolidayMonth)) +
  geom_line() +
  labs(title = "California Monthly Border Crossings by Holiday Status",
       x = "Date", y = "Total Crossings")
```

California Monthly Border Crossings by Holiday Status

- Correlation heat map for numeric variables
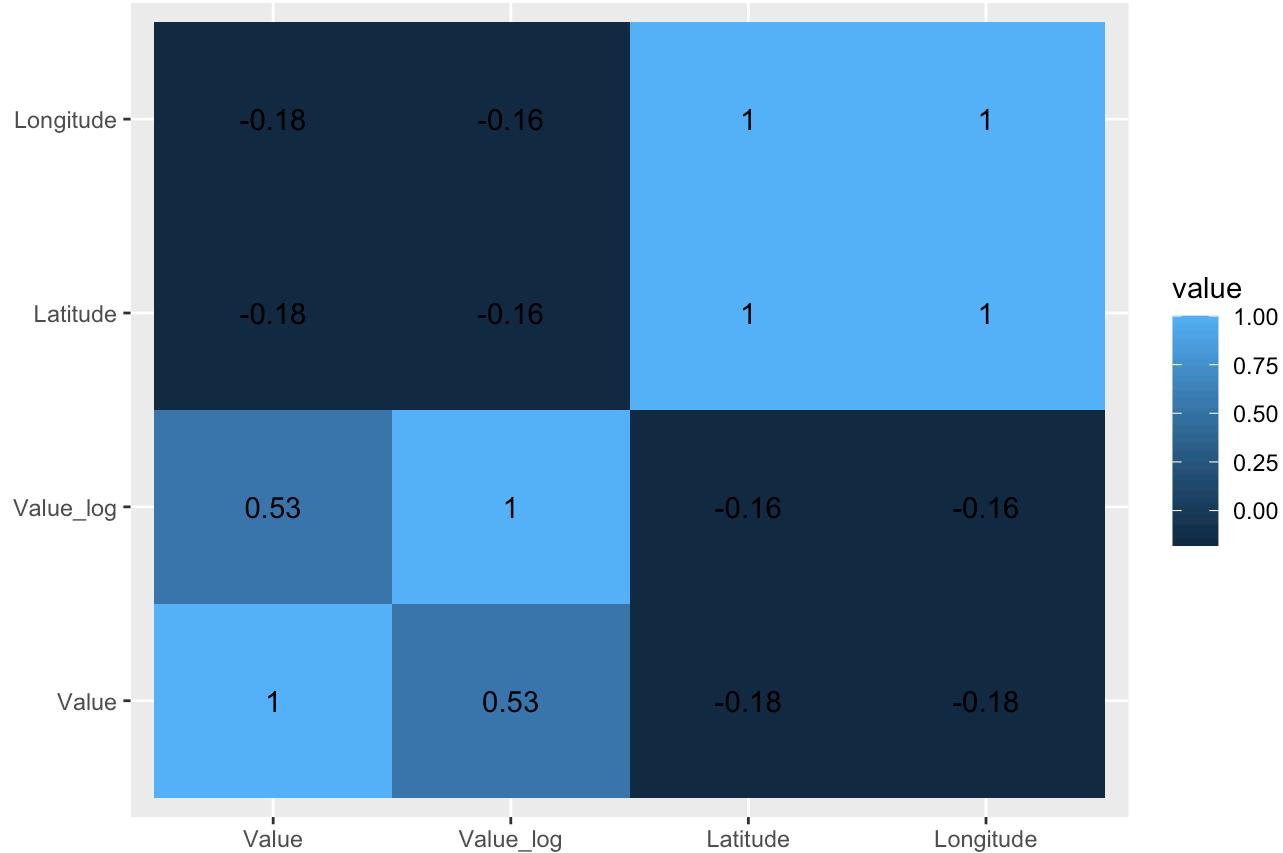
```
num_vars_ca <- border_ca_clean %>%
  select(Value, Value_log, Latitude, Longitude)

cor_mat_ca <- cor(num_vars_ca, use = "complete.obs")
cor_melt_ca <- melt(cor_mat_ca)

ggplot(cor_melt_ca, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  geom_text(aes(label = round(value, 2))) +
  labs(title = "Correlation Heatmap (California Data)",
       x = "", y = "")
```

Correlation Heatmap (California Data)

HYPOTHESIS TESTING - T-test: Are crossings higher during holiday months?

```
ttest_ca <- t.test(Value_log ~ HolidayMonth, data = border_ca_clean)
ttest_ca
```

```
##
##  Welch Two Sample t-test
##
## data:  Value_log by HolidayMonth
## t = -0.65946, df = 9186.3, p-value = 0.5096
## alternative hypothesis: true difference in means between group Holiday and group Non-Holiday i
s not equal to 0
## 95 percent confidence interval:
##  -0.2080624  0.1033098
## sample estimates:
##     mean in group Holiday mean in group Non-Holiday
##                  6.423566                  6.475942
```

- ANOVA: Differences by transportation mode

```
anova_ca <- aov(Value_log ~ Measure, data = border_ca_clean)
summary(anova_ca)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Measure       11 363408   33037    3605 <2e-16 ***
## Residuals   22003 201630       9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##if significant add
#TukeyHSD(anova_ca)
```

- Chi-square: Is HolidayMonth associated with Measure type?

```
ca_tab <- table(border_ca_clean[["Measure"]], border_ca_clean[["HolidayMonth"]])
chisq.test(ca_tab)
```

```
##
##  Pearson's Chi-squared test
##
## data:  ca_tab
## X-squared = 0.049906, df = 11, p-value = 1
```

- Correlation: Latitude vs crossing counts

```
cor(border_ca_clean[["Value"]], border_ca_clean[["Latitude"]])
```

```
## [1] -0.1802671
```

LINEAR REGRESION MODEL - Modeling dataset

```
model_ca <- border_ca_clean %>%
  select(Value_log, HolidayMonth, Measure,
         Year, Month, Lat_scaled, Lon_scaled) %>%
  na.omit()

nrow(model_ca)
```

```
## [1] 22015
```

- Train-test split

```
set.seed(123)
n <- nrow(model_ca)
train_idx <- sample(seq_len(n), size = 0.7*n)

train_ca <- model_ca[train_idx, ]
test_ca <- model_ca[-train_idx, ]
```

- Fit linear regression model

```
lm_ca <- lm(
  Value_log ~ HolidayMonth + Measure +
    Year + Month + Lat_scaled + Lon_scaled,
  data = train_ca
)

summary(lm_ca)
```

```
##
## Call:
## lm(formula = Value_log ~ HolidayMonth + Measure + Year + Month +
##       Lat_scaled + Lon_scaled, data = train_ca)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.8927  -1.5316   0.4565   1.8139   6.1402
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -99.067100   5.866140 -16.888  < 2e-16 ***
## HolidayMonthNon-Holiday           0.099541   0.065292   1.525    0.127
## MeasureBuses                     -1.996011   0.114643 -17.411  < 2e-16 ***
## MeasurePedestrians                4.303350   0.108690  39.593  < 2e-16 ***
## MeasurePersonal Vehicle Passengers 5.968575  0.109743  54.387  < 2e-16 ***
## MeasurePersonal Vehicles          5.339536   0.109413  48.802  < 2e-16 ***
## MeasureRail Containers Empty     -4.157098   0.115982 -35.843  < 2e-16 ***
## MeasureRail Containers Loaded    -5.663742   0.117785 -48.086  < 2e-16 ***
## MeasureTrain Passengers          -4.900437   0.115503 -42.427  < 2e-16 ***
## MeasureTrains                    -5.364269   0.114710 -46.764  < 2e-16 ***
## MeasureTruck Containers Empty    -1.181196   0.113169 -10.437  < 2e-16 ***
## MeasureTruck Containers Loaded   -0.788943   0.113701  -6.939 4.11e-12 ***
## MeasureTrucks                    -0.514230   0.113583  -4.527 6.02e-06 ***
## Year                              0.052667   0.002920  18.034  < 2e-16 ***
## Month                             0.007399   0.008192   0.903    0.366
## Lat_scaled                       14.665729   0.751391  19.518  < 2e-16 ***
## Lon_scaled                      -15.449658   0.751539 -20.557  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.838 on 15393 degrees of freedom
## Multiple R-squared:  0.6865, Adjusted R-squared:  0.6862
## F-statistic:  2107 on 16 and 15393 DF,  p-value: < 2.2e-16
```

- Predictions on test set

```
test_ca[["pred_log"]] <- predict(lm_ca, newdata = test_ca)
head(test_ca)
```

```
##    Value_log HolidayMonth                     Measure Year Month Lat_scaled
## 3  11.528415  Non-Holiday    Truck Containers Loaded 2024     1 -1.0360889
## 8   7.504392  Non-Holiday    Truck Containers Empty 2024      1 -0.6514024
## 12 10.821237  Non-Holiday            Personal Vehicles 2024   1  1.4495777
## 23 14.516173  Non-Holiday Personal Vehicle Passengers 2024    1 -1.1396584
## 24  8.489205  Non-Holiday                      Trucks 2024    1 -0.6514024
## 29 12.969336  Non-Holiday Personal Vehicle Passengers 2024    4  0.7837741
##     Lon_scaled  pred_log
## 3  -1.0222204  7.446872
## 8  -0.6643559  7.167444
## 12  1.5056977 10.973995
## 23 -1.1251207 14.275244
## 24 -0.6643559  7.834409
## 29  0.7510953 13.519084
```

- RMSE

```
rmse_ca <- sqrt(mean((test_ca[["pred_log"]] - test_ca[["Value_log"]])^2))
rmse_ca
```

```
## [1] 2.867307
```

- Test R^2

```
sst <- sum((test_ca[["Value_log"]] - mean(test_ca[["Value_log"]]))^2)
sse <- sum((test_ca[["Value_log"]] - test_ca[["pred_log"]])^2)
r2_test_ca <- 1 - sse/sst
r2_test_ca
```

```
## [1] 0.6799512
```

- Predicted vs Actual Plot

```
ggplot(test_ca, aes(Value_log, pred_log)) +
  geom_point(alpha = 0.4) +
  geom_abline(slope=1, intercept=0, linetype="dashed") +
  labs(title = "Predicted vs Actual (California Border Crossings)",
       x = "Actual log(1 + Value)", y = "Predicted log(1 + Value)")
```

Predicted vs Actual (California Border Crossings)