STOR 455 Group Project (Due 5pm on November 24th)

Old Geezers

Hunter Buresh (), Alex Georgiev (730483404), Zach Richardson ()

The Prediction (Required)

Our prediction of the cumulative domestic box office of "The Marvels" by December 8, 2023 is ...

Summary of Justification (Required)

Data (Required)

We obtained the dataset from: . ### Preview of dataset

```
data <- read.csv("marveldata.csv", quote="'")
summary(data)
```

```
##        id        title              totalgross        totaltheaters
##  Min.   : 1   Length:33          Min.   : 48483234   Min.   :3508
##  1st Qu.: 9   Class :character   1st Qu.:214504909   1st Qu.:4080
##  Median :17   Mode  :character   Median :333176600   Median :4275
##  Mean   :17                      Mean   :356173708   Mean   :4204
##  3rd Qu.:25                      3rd Qu.:411331607   3rd Qu.:4349
##  Max.   :33                      Max.   :858373000   Max.   :4662
##    opengross          opentheaters      date           distributer
##  Min.   : 46110859   Min.   :3505   Length:33          Length:33
##  1st Qu.: 80366312   1st Qu.:4030   Class :character   Class :character
##  Median :117027503   Median :4253   Mode  :character   Mode  :character
##  Mean   :132665838   Mean   :4196
##  3rd Qu.:179139142   3rd Qu.:4349
##  Max.   :357115007   Max.   :4662
##     season            production          marketing            ratings
##  Length:33          Min.   :130000000   Min.   : 65000000   Min.   :46.00
##  Class :character   1st Qu.:165000000   1st Qu.: 82500000   1st Qu.:76.00
##  Mode  :character   Median :200000000   Median :100000000   Median :83.00
##                     Mean   :204081818   Mean   :102040909   Mean   :80.88
##                     3rd Qu.:236200000   3rd Qu.:118100000   3rd Qu.:91.00
##                     Max.   :400000000   Max.   :200000000   Max.   :96.00
```

```
head(data, 5)
```

```
##   id                           title totalgross totaltheaters opengross
## 1 33                     The Marvels   48483234          4030  46110859
## 2 14    Guardians of the Galaxy Vol. 3  358995815          4450 118414021
## 3 25 Ant-Man and the Wasp: Quantumania  214504909          4345 106109650
## 4  7    Black Panther: Wakanda Forever  453829060          4396 181339761
## 5 15            Thor: Love and Thunder  343256830          4375 144165107
```

```
##   opentheaters       date                          distributer season production
## 1         4030 2023-11-10 Walt Disney Studios Motion Pictures   Fall  270000000
## 2         4450 2023-05-05 Walt Disney Studios Motion Pictures Spring  250000000
## 3         4345 2023-02-17 Walt Disney Studios Motion Pictures Winter  200000000
## 4         4396 2022-11-11 Walt Disney Studios Motion Pictures   Fall  250000000
## 5         4375 2022-07-08 Walt Disney Studios Motion Pictures Summer  250000000
##   marketing ratings
## 1 135000000      62
## 2 125000000      82
## 3 100000000      46
## 4 125000000      83
## 5 125000000      63
```

```
# We don't need the id, distributor, title, or date columns
modelData <- subset(data, select = -c(id, distributer, title, date))
summary(modelData)
```

```
##    totalgross        totaltheaters    opengross          opentheaters
##  Min.   : 48483234   Min.   :3508   Min.   : 46110859   Min.   :3505
##  1st Qu.:214504909   1st Qu.:4080   1st Qu.: 80366312   1st Qu.:4030
##  Median :333176600   Median :4275   Median :117027503   Median :4253
##  Mean   :356173708   Mean   :4204   Mean   :132665838   Mean   :4196
##  3rd Qu.:411331607   3rd Qu.:4349   3rd Qu.:179139142   3rd Qu.:4349
##  Max.   :858373000   Max.   :4662   Max.   :357115007   Max.   :4662
##     season            production          marketing            ratings
##  Length:33          Min.   :130000000   Min.   : 65000000   Min.   :46.00
##  Class :character   1st Qu.:165000000   1st Qu.: 82500000   1st Qu.:76.00
##  Mode  :character   Median :200000000   Median :100000000   Median :83.00
##                     Mean   :204081818   Mean   :102040909   Mean   :80.88
##                     3rd Qu.:236200000   3rd Qu.:118100000   3rd Qu.:91.00
##                     Max.   :400000000   Max.   :200000000   Max.   :96.00
```

Analysis (Required)

We noticed that the opening gross for The Marvels was only $46 million. Looking at the distrbution of the opening gross of other Marvel movies shows that The Marvels has the lowest opening gross out of all the movies in the dataset. ### Distribution of opening gross

```
openGrossMil <- data$opengross/1000000
summary(openGrossMil)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   46.11   80.37  117.03  132.67  179.14  357.12
```

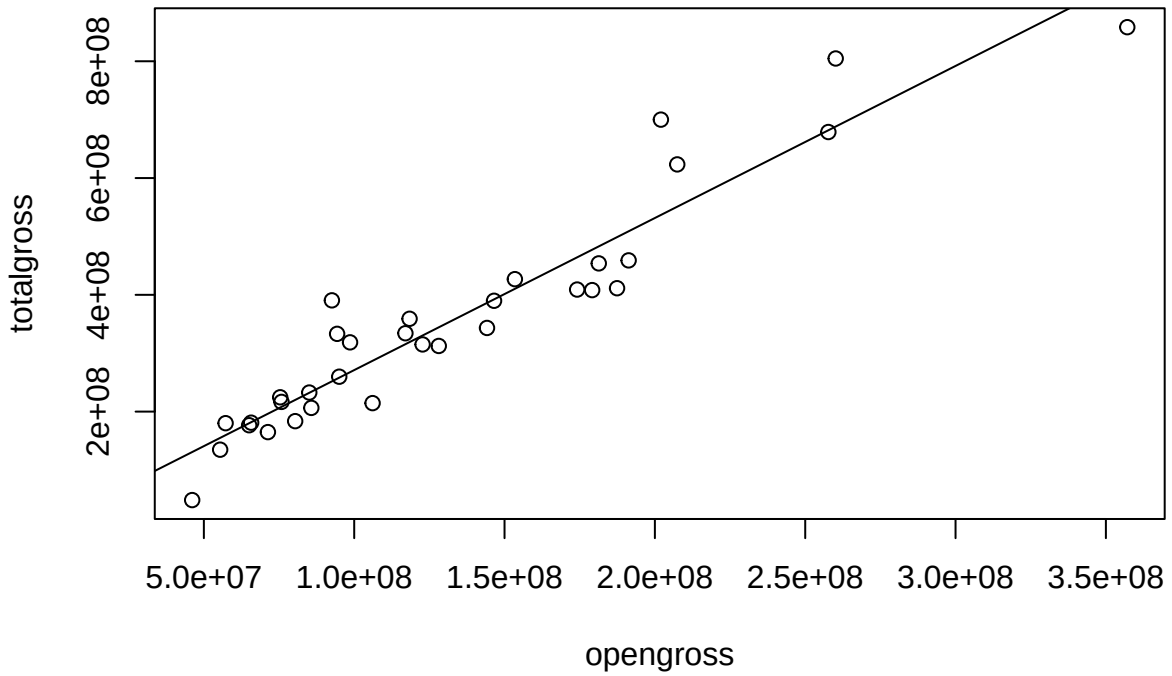
Full model

```
# Fit a model with all data
fullModel = lm(totalgross~opengross, data=modelData)
summary(fullModel)
```

```
##
## Call:
## lm(formula = totalgross ~ opengross, data = modelData)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -87453553 -36306375  -3008707  18757942 163292669
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.064e+07  2.343e+07   0.454    0.653
## opengross   2.604e+00  1.565e-01  16.641   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62350000 on 31 degrees of freedom
## Multiple R-squared:  0.8993, Adjusted R-squared:  0.8961
## F-statistic: 276.9 on 1 and 31 DF,  p-value: < 2.2e-16
```

```
# Plot regression line
plot(totalgross~opengross, data=modelData)
abline(fullModel)
```



Model with low opening gross

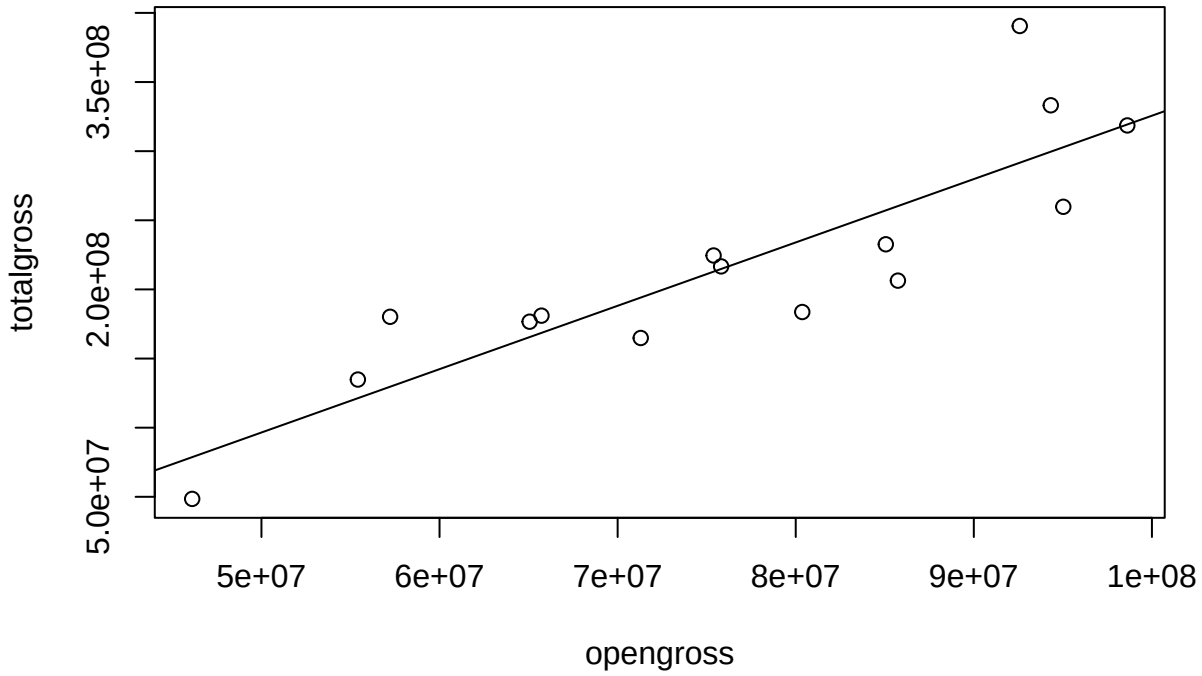
```
# Get data with opening gross < 100 mil
lowOpeners = subset(modelData, opengross<=100000000)

# Fit model with low openers
lowOpenModel = lm(totalgross~opengross, data=lowOpeners)
summary(lowOpenModel)
```

```
##
## Call:
## lm(formula = totalgross ~ opengross, data = lowOpeners)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -53920824 -29717738   1852240  12979095  98896996
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.326e+08  5.563e+07  -2.384   0.0331 *
## opengross    4.583e+00  7.150e-01   6.409 2.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42880000 on 13 degrees of freedom
## Multiple R-squared:  0.7596, Adjusted R-squared:  0.7411
## F-statistic: 41.08 on 1 and 13 DF,  p-value: 2.31e-05
```

```
# Plot regression line
plot(totalgross~opengross, data=lowOpeners)
abline(lowOpenModel)
```



Model with only Fall releases

```
# Only take movies that had fall releases
fallData = subset(modelData, season=="Fall")
summary(fallData)
```

```
##    totalgross        totaltheaters    opengross          opentheaters
##  Min.   : 48483234   Min.   :3841   Min.   : 46110859   Min.   :3841
##  1st Qu.:185616187   1st Qu.:3956   1st Qu.: 73342954   1st Qu.:3956
##  Median :224543292   Median :4080   Median : 85058311   Median :4080
##  Mean   :235112596   Mean   :4088   Mean   : 95382524   Mean   :4088
##  3rd Qu.:273850104   3rd Qu.:4195   3rd Qu.:104241415   3rd Qu.:4195
##  Max.   :453829060   Max.   :4396   Max.   :181339761   Max.   :4396
##     season            production          marketing           ratings
##  Length:7           Min.   :150000000   Min.   : 75000000   Min.   :47.00
##  Class :character   1st Qu.:157500000   1st Qu.: 78750000   1st Qu.:64.50
```

```
##  Mode  :character   Median :180000000   Median : 90000000   Median :83.00
##                     Mean   :200171429   Mean   :100085714   Mean   :76.14
##                     3rd Qu.:243100000   3rd Qu.:121550000   3rd Qu.:90.50
##                     Max.   :270000000   Max.   :135000000   Max.   :93.00
```

```
head(modelData)
```

```
##   totalgross totaltheaters opengross opentheaters season production marketing
## 1   48483234          4030  46110859         4030   Fall  270000000 135000000
## 2  358995815          4450 118414021         4450 Spring  250000000 125000000
## 3  214504909          4345 106109650         4345 Winter  200000000 100000000
## 4  453829060          4396 181339761         4396   Fall  250000000 125000000
## 5  343256830          4375 144165107         4375 Summer  250000000 125000000
## 6  411331607          4534 187420998         4534 Spring  200000000 100000000
##   ratings
## 1      62
## 2      82
## 3      46
## 4      83
## 5      63
## 6      73
```

```
# Fit model with fall releases
fallModel = lm(totalgross~opengross, data=fallData)
summary(fallModel)
```

```
##
## Call:
## lm(formula = totalgross ~ opengross, data = fallData)
##
## Residuals:
##         1         4         8         9        17        20        26
## -50747029 -18338034  -3819451  44570077   4484968  26001635  -2152166
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.794e+07  3.209e+07   -0.87 0.423893
## opengross    2.758e+00  3.092e-01    8.92 0.000295 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33510000 on 5 degrees of freedom
## Multiple R-squared:  0.9409, Adjusted R-squared:  0.929
## F-statistic: 79.56 on 1 and 5 DF,  p-value: 0.000295
```

```
# Plot regression line
plot(totalgross~opengross, data=fallData)
abline(fallModel)
```

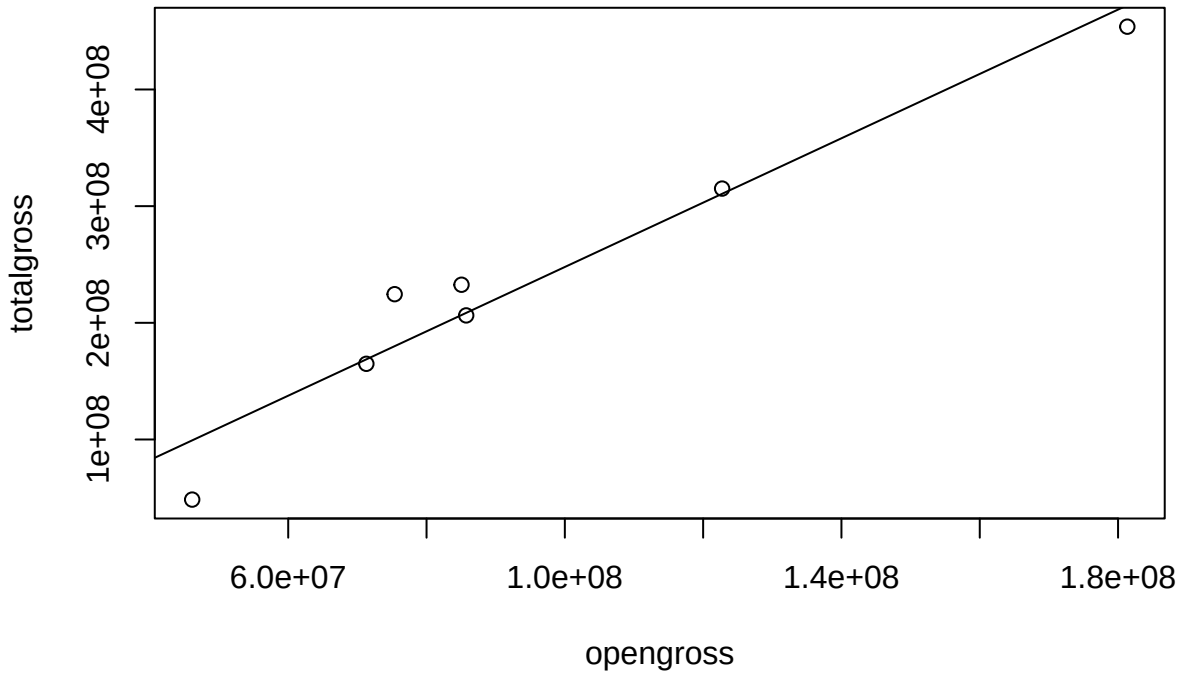


Table with prediction from all models

```
# Get The Marvels opening gross
theMarvelsOpenGross = data.frame(opengross=(subset(data, title=="The Marvels"))$opengross)

# Predictions for each model
fullPred = predict(fullModel, newdata = theMarvelsOpenGross)
lowOpenPred = predict(lowOpenModel, newdata = theMarvelsOpenGross)
fallPred = predict(fallModel, newdata = theMarvelsOpenGross)

# Create a data frame for the table
predictionTable <- data.frame(
  FullModel = fullPred,
  LowOpenModel = lowOpenPred,
  FallModel = fallPred
)

# Display the table
print(predictionTable)
```

```
##   FullModel LowOpenModel FallModel
## 1 130738678     78683385  99230263
```

Predict only with daily earnings

```
# Convert table from website into dataframe
dailyEarnings <- data.frame(
  Date = c("Nov 20", "Nov 19", "Nov 18", "Nov 17", "Nov 16", "Nov 15", "Nov 14", "Nov 13", "Nov 12", "N
  Day = c("Monday", "Sunday", "Saturday", "Friday", "Thursday", "Wednesday", "Tuesday", "Monday", "Sunda
  Rank = c(4, 3, 3, 4, 1, 1, 1, 1, 1, 1, 1),
  Revenue = c(1137196, 2910248, 4453682, 2756659, 1251387, 1789239, 3300946, 2372375, 9247703, 15260052
  Change_Daily = c("-60.9%", "-34.7%", "+61.6%", "+120.3%", "-30.1%", "-45.8%", "+39.1%", "-74.3%", "-3
```

```
  Change_LastWeek = c("-52.1%", "-68.5%", "-70.8%", "-87.2%", "-", "-", "-", "-", "-", "-", "-"),
  Theaters = c(4030, 4030, 4030, 4030, 4030, 4030, 4030, 4030, 4030, 4030, 4030),
  Avg = c(282, 722, 1105, 684, 310, 443, 819, 588, 2294, 3786, 5360),
  TotalRevenue = c(66082591, 64945395, 62035147, 57581465, 54824806, 53573419, 51784180, 48483234, 4611
  Week = c(11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1)
)

# Preview data
summary(dailyEarnings)
```

```
##      Date               Day                 Rank          Revenue
##  Length:11          Length:11          Min.   :1.000   Min.   : 1137196
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.: 2080807
##  Mode  :character   Mode  :character   Median :1.000   Median : 2910248
##                                        Mean   :1.909   Mean   : 6007508
##                                        3rd Qu.:3.000   3rd Qu.: 6850692
##                                        Max.   :4.000   Max.   :21603104
##  Change_Daily       Change_LastWeek       Theaters         Avg
##  Length:11          Length:11          Min.   :4030   Min.   : 282.0
##  Class :character   Class :character   1st Qu.:4030   1st Qu.: 515.5
##  Mode  :character   Mode  :character   Median :4030   Median : 722.0
##                                        Mean   :4030   Mean   :1490.3
##                                        3rd Qu.:4030   3rd Qu.:1699.5
##                                        Max.   :4030   Max.   :5360.0
##   TotalRevenue           Week
##  Min.   :21603104   Min.   : 1.0
##  1st Qu.:47297046   1st Qu.: 3.5
##  Median :53573419   Median : 6.0
##  Mean   :51262487   Mean   : 6.0
##  3rd Qu.:59808306   3rd Qu.: 8.5
##  Max.   :66082591   Max.   :11.0
```

```
head(dailyEarnings)
```

```
##     Date       Day Rank Revenue Change_Daily Change_LastWeek Theaters  Avg
## 1 Nov 20    Monday    4 1137196       -60.9%          -52.1%     4030  282
## 2 Nov 19    Sunday    3 2910248       -34.7%          -68.5%     4030  722
## 3 Nov 18  Saturday    3 4453682       +61.6%          -70.8%     4030 1105
## 4 Nov 17    Friday    4 2756659      +120.3%          -87.2%     4030  684
## 5 Nov 16  Thursday    1 1251387       -30.1%               -     4030  310
## 6 Nov 15 Wednesday    1 1789239       -45.8%               -     4030  443
##   TotalRevenue Week
## 1     66082591   11
## 2     64945395   10
## 3     62035147    9
## 4     57581465    8
## 5     54824806    7
## 6     53573419    6
```

SLR based on earnings since Nov 12

```
# Data that only contains Nov 12 and later
sinceNov12 = subset(dailyEarnings, (Date != "Nov 10") & (Date != "Nov 11"))
```

```r
# Add column that represents days since Nov 12
sinceNov12$DaysAfterOpeningWknd <- c(8, 7, 6, 5, 4, 3, 2, 1, 0)

# Show data
summary(sinceNov12)
```

```
##      Date               Day                 Rank          Revenue
##  Length:9           Length:9           Min.   :1.000   Min.   :1137196
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.:1789239
##  Mode  :character   Mode  :character   Median :1.000   Median :2756659
##                                        Mean   :2.111   Mean   :3246604
##                                        3rd Qu.:3.000   3rd Qu.:3300946
##                                        Max.   :4.000   Max.   :9247703
##  Change_Daily       Change_LastWeek       Theaters          Avg
##  Length:9           Length:9           Min.   :4030   Min.   : 282.0
##  Class :character   Class :character   1st Qu.:4030   1st Qu.: 443.0
##  Mode  :character   Mode  :character   Median :4030   Median : 684.0
##                                        Mean   :4030   Mean   : 805.2
##                                        3rd Qu.:4030   3rd Qu.: 819.0
##                                        Max.   :4030   Max.   :2294.0
##   TotalRevenue           Week    DaysAfterOpeningWknd
##  Min.   :46110859   Min.   : 3   Min.   :0
##  1st Qu.:51784180   1st Qu.: 5   1st Qu.:2
##  Median :54824806   Median : 7   Median :4
##  Mean   :56157900   Mean   : 7   Mean   :4
##  3rd Qu.:62035147   3rd Qu.: 9   3rd Qu.:6
##  Max.   :66082591   Max.   :11   Max.   :8
```

```r
sinceNov12
```

```
##     Date       Day Rank Revenue Change_Daily Change_LastWeek Theaters  Avg
## 1 Nov 20    Monday    4 1137196       -60.9%          -52.1%     4030  282
## 2 Nov 19    Sunday    3 2910248       -34.7%          -68.5%     4030  722
## 3 Nov 18  Saturday    3 4453682       +61.6%          -70.8%     4030 1105
## 4 Nov 17    Friday    4 2756659      +120.3%          -87.2%     4030  684
## 5 Nov 16  Thursday    1 1251387       -30.1%               -     4030  310
## 6 Nov 15 Wednesday    1 1789239       -45.8%               -     4030  443
## 7 Nov 14   Tuesday    1 3300946       +39.1%               -     4030  819
## 8 Nov 13    Monday    1 2372375       -74.3%               -     4030  588
## 9 Nov 12    Sunday    1 9247703       -39.4%               -     4030 2294
##   TotalRevenue Week DaysAfterOpeningWknd
## 1     66082591   11                     8
## 2     64945395   10                     7
## 3     62035147    9                     6
## 4     57581465    8                     5
## 5     54824806    7                     4
## 6     53573419    6                     3
## 7     51784180    5                     2
## 8     48483234    4                     1
## 9     46110859    3                     0
```

```r
# Fit model passed on days since opening weekend
sinceNov12Model = lm(TotalRevenue~DaysAfterOpeningWknd, data=sinceNov12)

# Summarize model
```

```r
summary(sinceNov12Model)
```

```
##
## Call:
## lm(formula = TotalRevenue ~ DaysAfterOpeningWknd, data = sinceNov12)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1333094  -327535    14504   751134  1098326
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          45905674     547891   83.79 9.08e-12 ***
## DaysAfterOpeningWknd  2563056     115080   22.27 9.30e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 891400 on 7 degrees of freedom
## Multiple R-squared:  0.9861, Adjusted R-squared:  0.9841
## F-statistic:   496 on 1 and 7 DF,  p-value: 9.301e-08
```

```r
# Convert to millions for easier viewing
plotData = sinceNov12
plotData$TotalRevenue = plotData$TotalRevenue / 1000000

# Plot regression line
milModel = lm(TotalRevenue~DaysAfterOpeningWknd, data=plotData)
plot(TotalRevenue~DaysAfterOpeningWknd, data=plotData, yaxp=c(40,70,5), ylim=c(40, 70),
     main="Revenue Since Opening Weekend", xlab="Days Since Opening Weekend", ylab="Total Revenue")
abline(milModel)
```