Gulf Oil Spill 2010

Abstract:

The deepwater Horizon oil rig exploded in the Gulf of Mexico leading to an oil spill starting on 22nd April 2010 that lasted for 85 days. The goal of our paper was to find a model that fitted the data we have from the oil spill. We broke our data into three parts based on how it lined up with real events at the time of the spill and tested out various different single step, two step and three step models to find the model that fitted the data the best. We found that a three step piecewise model with the first and last part of the model using a linear and the middle part using a quadratic fit was the best fit for the data because it had the lowest total residual standard error out of all the different models we tried. Finally, we used our model to predict how the harmful financial and environmental effects of the oil spill could have been lessened.

Introduction:

On 20th April 2010, the Deepwater Horizon oil rig exploded in the Gulf of Mexico. This led to an oil leakage beginning on 22nd April 2010 which lasted for 85 days. The spill was eventually contained on 15th July 2010 by putting on a containment cap. The oil spill was estimated to have cost a loss of \$61.6 billion and around one million seabirds died from exposure due to the spill. We came up with a piecewise model broken down into three segments using differential equations to try to model the growth and size of the oil spill over time. We also used our model to suggest a different approach that could have been used to speed up the cleaning up process which would potentially have reduced the severe environmental and financial losses incurred due to the oil spill.

Dataset and Assumptions:

Data:

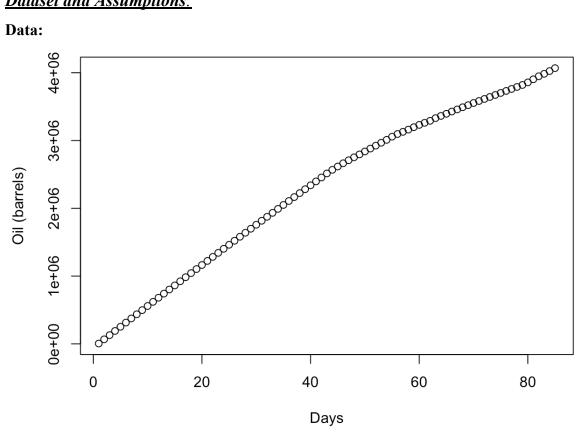


Fig 1. Graph showing oil spill in barrels of oil over a time period of 85 days.

Assumptions:

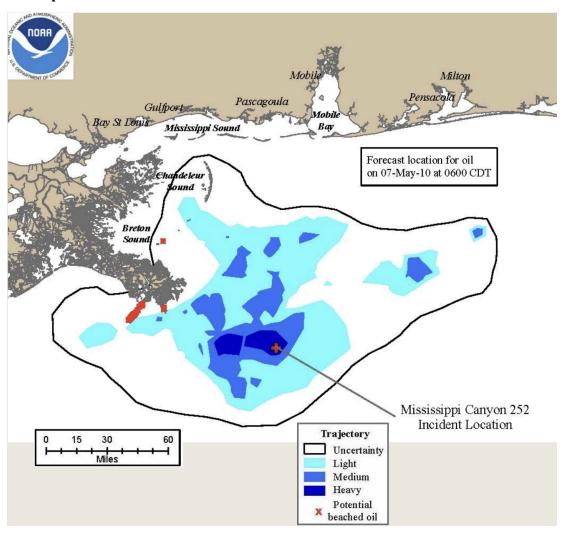


Fig 2. Map showing forecasted area of the oil spill using satellite imaging

For simplicity of our model, we assumed that the oil spill only occurred in the forecasted area as shown in the figure and did not spill to surrounding areas. This allowed us to assume that our dataset, provided to us by techniques of satellite forecasting, was accurate and gave us complete information about the oil spill. Additionally, we ignored the data from the first day of the oil spill because we assumed that this data was incomplete. Our research showed that the oil spill was first recorded around 8pm on 22nd April 2022 which would explain why the data from that day is about 1/6th of the oil spilled on the rest of the days. We also assumed that the oil initially being removed was negligible and did not affect the rate at which the oil spilled. Finally, we partitioned the dataset at three different points which we assumed represented significant

changes in either the flow of oil or the speed at which the oil was being removed. We assumed this led to the change in the trend of the data seen at these points.

Model:

Description:

Our model was broken down into three different parts on the basis of the trends we saw in our dataset. The first part of the dataset consisted of the oil spill from Day 1- Day 43. We used a linear model to fit that part of the data. The second part of the data was broken down into days 44-79 and we used a quadratic fit for the data. This part of the model stated that oil was added at a constant rate but removed at a rate that was a function of time squared. The equation used was

 $\frac{dy}{dy} = A - Dt^2$. This was solved by integrating the function with respect to time to give the solution $y(t) = At - \frac{Dt^2}{2} + C$. The last part of the model looked at the oil spilled from Day 80- Day 85 and we again used a linear model to fit the data. The linear model used in the first and last part of the model was solved in R which estimated the coefficients for the model using

Ordinary Least Squares Regression
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

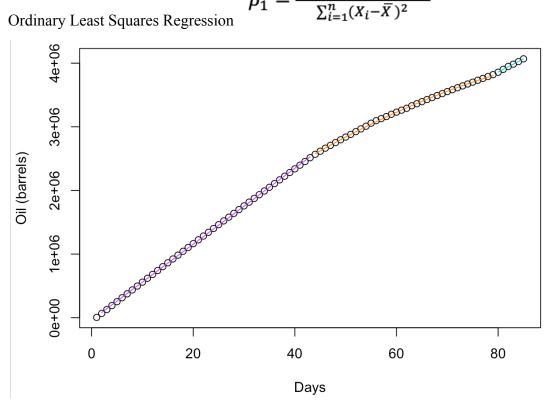


Fig 3. Graph showing our three step piecewise model fitted to the dataset

Analysis:

To quantify what was the best fit for the different segments of our dataset, we focused on the residual standard errors (RSE) the different fits gave us. The fit with the lowest standard error was considered to be the best fit for that part of the data and was selected. As can be seen in Fig 4, the first part of the model from Day 1- Day 43 has a linear fit and our line goes through all of the data points indicating that this is a good fit for our data. Furthermore, this fit resulted in a residual standard error of 7461 on 40 degrees of freedom (as seen in Fig 5) which was the lowest for this section of the data. This led us to conclude that it was the best fit for our data from Day 1- Day 43. The linear fit also gave a slope of 59644.22 which meant that for every additional day approximately 60000 barrels of oil were added to the spill.

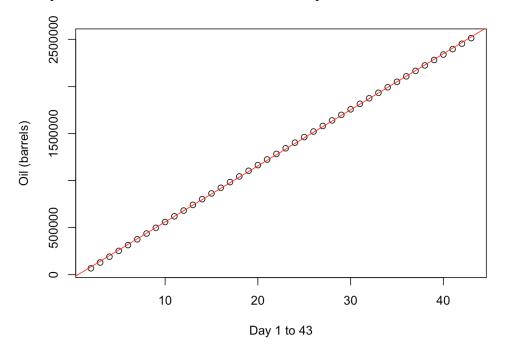


Fig 4. Graph showing linear fit of data for day 1-43 for our selected model

```
Residuals:
  Min
           1Q Median
                         30
                               Max
       -5361
                              8200
-15276
                2009
                       6530
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -37013.55
                         2427.43
                                  -15.25
                                           <2e-16 ***
                                           <2e-16 ***
day43
             59644.92
                           94.98 627.97
                 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 7461 on 40 degrees of freedom
Multiple R-squared: 0.9999,
                                Adjusted R-squared: 0.9999
F-statistic: 3.943e+05 on 1 and 40 DF, p-value: < 2.2e-16
```

Fig 5. Figure showing model summary and residual standard error for linear fit of data from Day 1-43

For the second segment of the data, from Day 44- day 79, we used a quadratic fit for the data and as seen in Fig 6, the fit included almost all of the data points in this segment. The residual standard error for this part of the data set was 7099 on 33 degrees of freedom as shown below in Fig 7. This was the lowest residual standard error recorded for any of the models used for this part of the data set so we concluded that this fit was the best fit for this segment.

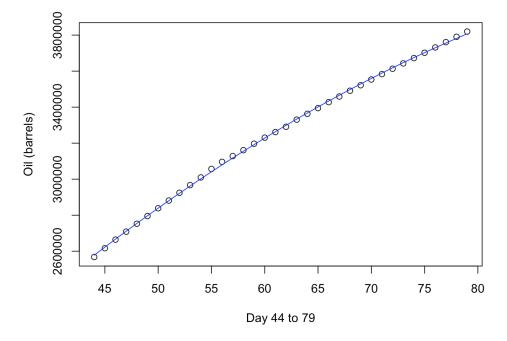


Fig 6. Graph showing quadratic fit of data for day 44-79 for our selected model

```
Parameters:
    Estimate Std. Error t value Pr(>|t|)
                         46.843
                                  <2e-16 ***
mya 70909.95
      581.89
                  24.54
                         23.708
                                  <2e-16
myc 21087.40
               45648.44
                          0.462
                                   0.647
Signif. codes:
                       0.001 "** 0.01 "* 0.05 ". 0.1 " 1
Residual standard error: 7099 on 33 degrees of freedom
Number of iterations to convergence: 1
Achieved convergence tolerance: 1.163e-07
```

Fig 7. Figure showing model summary and residual standard error for quadratic fit of data from Day 44-79

For the final part of the data (Day 80- Day 85), we used a linear fit as well. The fitted line passed through almost all of the data points as seen in Fig 8. The linear model also led to a residual standard error of 2943 on 4 degrees of freedom shown below in Fig 9. This was the lowest residual standard error for any of the different fits used for this part of the dataset which allowed us to deduce that this was the best for this part of the data. The linear fit also gave a slope of 41435.1 which meant that for every additional day approximately 41000 barrels of oil were added to the spill. This rate was lesser than the amount of oil being spilled in the first part of the model which might be due to the increased efforts in removing the oil being spilled.

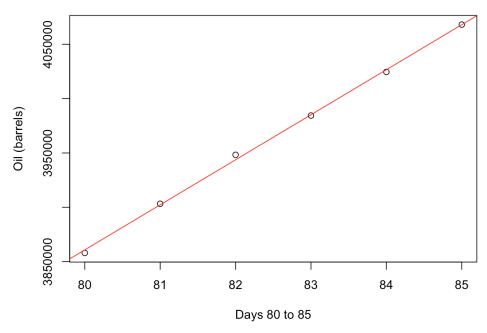


Fig 8. Graph showing linear fit of data for day 80-85 for our selected model

```
Residuals:
          989.96 4575.88 -760.21 -2037.30
                                               57.62
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
                                 9.408 0.000712 ***
(Intercept) 546071.1
                       58045.0
day80t85
            41435.1
                         703.4 58.905 4.97e-07 ***
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
Residual standard error: 2943 on 4 degrees of freedom
Multiple R-squared: 0.9988,
                               Adjusted R-squared:
F-statistic: 3470 on 1 and 4 DF. p-value: 4.974e-07
```

Fig 9. Figure showing model summary and residual standard error for linear fit of data from Day 80-85

Thus, the final form of our best fit model would be:

(a) Day 1-43 :
$$y(t) = 59645(t) - 37014$$

(b) Day 44-79: $y(t) = 70910(t) - \frac{582}{2}(t^2) + 21087$
(c) Day 80-85: $y(t) = 41435(t) + 546071$
Total Residual standard error (TRSE) = RSE(a) + RSE(b) + RSE(c)
TRSE = 7461 + 7099 + 2943
TRSE = 17503

Thus, we concluded our model to be the best fit for the data because it gave us the lowest TRSE out of all of the different models we used for our given dataset.

Comparing with other models:

We came up with our model as the best fit after comparing it to other models and finding that they had higher TRSE than our chosen model. One of the models we compared our model to was another three step model. We chose the same linear model for Days 1-43 and Days 80-85 as we had for our selected model. However, we chose a cubed model for days 44-79. The equation for this model was $y(t) = At - \frac{Dt^3}{3} + C$. We found a residual standard error of 8361 (shown below in Fig 10) for this middle part of the model (this was more than that of our selected model

for the middle part of the dataset) which led to a TRSE of 18765 which was more than the TRSE of 17503 for our chosen model.

```
Parameters:
    Estimate Std. Error t value Pr(>|t|)
mya 5.309e+04   9.129e+02   58.16   < 2e-16 ***
myd 4.671e+00   2.347e-01   19.90   < 2e-16 ***
myc 3.782e+05   3.674e+04   10.29   7.82e-12 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8361 on 33 degrees of freedom

Number of iterations to convergence: 1
Achieved convergence tolerance: 1.68e-07
```

Fig 10. Figure showing model summary and residual standard error for cubed fit of data from Day 44-79

We also tried to fit a two step model to the data and partitioned the dataset into two parts from Day 1-43 and Day 44-85. For the first part of this model, we chose the same linear model as our final selected model but for the second part of this model we chose a quadratic fit of the form $y(t) = At - \frac{Dt^2}{2} + C$. This quadratic fit did not visually seem to be the best fit for the data from Day 44-85 either as we show below in Fig 11. The model seemed to especially not fit the data after Day 80. Our results for this residual standard error were higher as expected (shown below in Fig 12) and we got a residual standard error of 18360. This led to a TRSE of 25821 which was much higher than the TRSE of our selected model which was 17503.

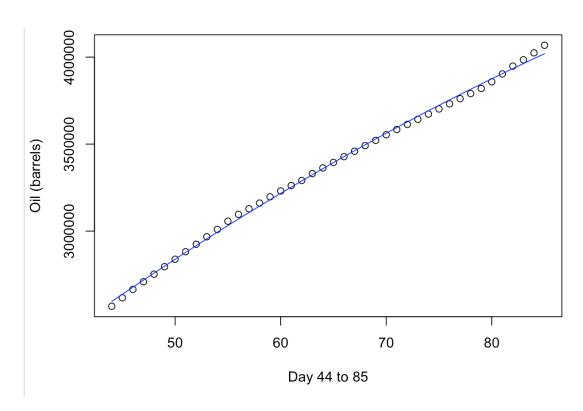


Fig 11. Graph showing data from Day 44-85 with a quadratic fit

```
Parameters:
    Estimate Std. Error t value Pr(>|t|)
    55867.92
                2793.49 19.999 < 2e-16 ***
       327.68
              43.16
                          7.592 3.34e-09 ***
myd
myc 454453.26
               87952.00
                          5.167 7.38e-06 ***
Signif. codes:
               0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
Residual standard error: 18360 on 39 degrees of freedom
Number of iterations to convergence: 1
Achieved convergence tolerance: 3.213e-07
```

Fig 12. Figure showing model summary and residual standard error for quadratic fit of data from Day 44-85

Moreover, we also tried another two step model with the same linear model as our final selected model from Days 1-43 but a cubed fit for the data from Day 44-85. As expected, the cubed equation of $y(t) = At - \frac{Dt^3}{3} + C$ did not turn out to be a better fit for our data than our already selected final model. We got a residual standard error of 19480 for the cubed fit in our model which we have shown below in Fig 13. Combined with the residual standard error of

the linear model, the TRSE added up to 26941 which was again much higher than our TRSE for our selected final model of 17503.

```
Parameters:
    Estimate Std. Error t value Pr(>|t|)
mya 4.504e+04 1.526e+03 29.514 < 2e-16 ***
myd 2.426e+00 3.544e-01 6.845 3.50e-08 ***
myc 6.872e+05 6.416e+04 10.711 3.53e-13 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19480 on 39 degrees of freedom

Number of iterations to convergence: 1
Achieved convergence tolerance: 4.064e-08
```

Fig 13. Figure showing model summary and residual standard error for cubed fit of data from Day 44-85

Furthermore, we tried different single step models to fit our data. A quadratic fit was used to fit our data and did not seem to be the best fit for the data as seen below in Fig 14. Our single quadratic fit model returned a residual standard error of 33640 which was much higher than the TRSE of 17503 for our selected final model. Additionally, a single cubed model was also tried to fit the data and we saw that it was not the best fit for our data for a lot of different data points as we can see in Fig 15. The single cubed model gave a residual standard error of 30350 which was again higher than the TRSE of our selected model. Finally, we tried a single linear model to fit our dataset. As we have shown below in Fig 16, the single linear model clearly did not fit the data well at all and returned a residual standard error of 156800 which was much higher than the TRSE of our selected model.

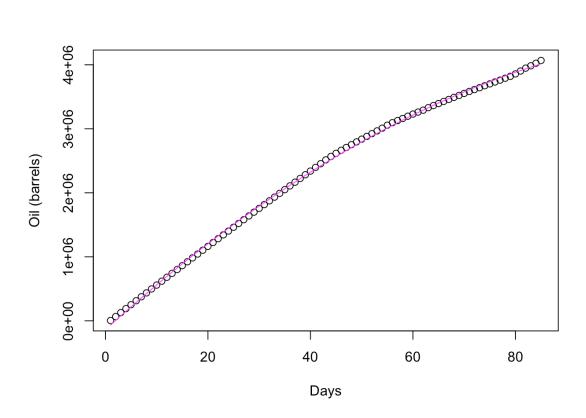


Fig 14. Graph showing a quadratic fit applied to the whole dataset

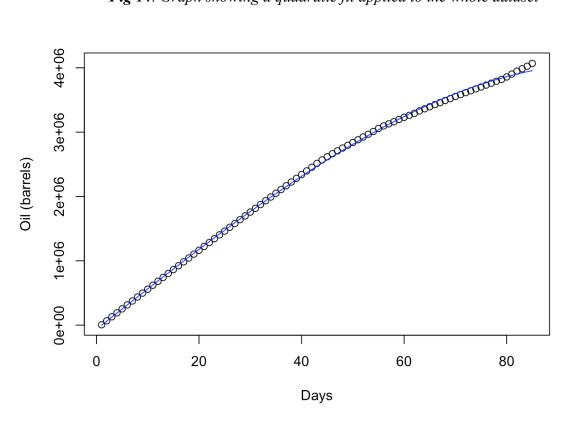


Fig 15. Graph showing a cubed fit applied to the whole dataset

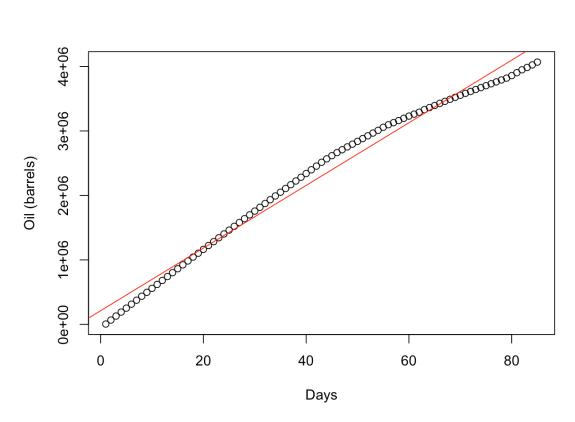


Fig 16. Graph showing a linear fit applied to the whole dataset

Discussion:

The comparison of our three step model with other models and its lowest total residual standard error showed that it is indeed the best fit to our data. Our model was further strengthened by how closely it tied to real world events that happened at the time of the oil spill(Reuters, 2010). For the first 43 days of the spill, the cleanup efforts did not do much to remove the oil that had spilled. This was reflected in the linear part of our model from Day 1-43 where oil was being added at a constant rate everyday (approximately 60,000 barrels per day) and the amount of oil was constantly increasing. From day 44-79 we used a quadratic model to fit the data. This is because on April 15th 2010 (Day 45) a temporary cap was put on to contain the well which led the rate of oil added to decrease (Reuters, 2010). On day 80 the temporary cap was removed and a permanent cap was put in place on 15th July 2010 (day 85) to permanently prevent the oil from spilling. Within these 6 days, the oil started to spill again at a constant rate and the amount of oil started to increase linearly. This was something that was reflected in the linear model we used from Day 80-Day 85.

Our research showed that there were multiple alternative measures used to try to cover the well before the containment cap was put on. On May 7th 2010 (15 days after the start of the

spill), BP tried to cover the well by shooting it with mud and cement which did not work. Two days later (17 days after the start of the spill) BP tried to shoot the well with golf and rubber balls to plug the spill (Weiner, 2010). All of these techniques proved to be futile as the oil continued to spill at a constant rate as shown by our model and eventually a temporary containment cap was put on 6th June 2010 which caused the rate of the oil spill to decrease.

We used our model to show how much money lost could have been prevented had the 28 days spent on futile methods (May 7th- June 6th) not been wasted and the temporary cap would have been put on earlier. The linear part of our model from Day 1-43 had a slope of 59644.92 which meant that 59644 barrels of oil were being spilled everyday during the first 43 days. The cost of cleaning up oil was estimated to be around \$75-\$80 per barrel (Cohen, 2010). This means (59644 barrels / day * 28 days) 1670032 less barrels of oil would have been spilled and (1670032 barrels * \$75/barrel) \$125,252,400 could have been saved if the cap had been put on earlier.

Furthermore, research has shown that 5000-6500 seagulls, 165,000 sea turtles, 26,000 marine animals and 82,000 birds were killed by the spill (O'Mara, 2015). We also used our model to see how these numbers could have reduced to approximately 1700 less seagulls, 54,000 less sea turtles, 8500 less marine animals and 27000 less birds killed (28/85 * number of animals killed) had the containment cap been put on earlier.

To end our paper,

A minute of silence for the birds.

References

- 1. Cohen, M. A. (2010). A Taxonomy of Oil Spill Costs. *Resources for the Future*. https://doi.org/https://media.rff.org/documents/RFF-BCK-Cohen-DHCosts.pdf
- 2. O'Mara, C. (2015). *Deepwater horizon's impact on Wildlife*. National Wildlife Federation. Retrieved September 20, 2022, from https://www.nwf.org/oilspill
- 3. Reuters. (2010, July 30). *Timeline of the Gulf of Mexico oil spill*. CNBC. Retrieved September 19, 2022, from https://www.cnbc.com/id/37448876
- 4. Weiner, M. (2010, May 31). *How golf balls took on the oil spill and lost*. CNN. Retrieved September 20, 2022, from

 $http://www.cnn.com/2010/SPORT/golf/05/31/golf.oil.spill/index.html\#: \sim: text = BP\%20 press\%20 \\ officer\%20 Sheila\%20 Williams, keep\%20 the\%20 junk\%20 in\%20 place$

Appendix

Code for Project:

```
#Loading Data
setwd("~/Desktop/Math 301/Oil Spill Project")
oilspill = read.csv("oil data.csv", header=T)
attach(oilspill)
plot(day,oil, xlab = "Days", ylab = "Oil (barrels)")
#Our Best Model (Three Parts)
#(Residual standard error: 2943 on 4 degrees of freedom), (Residual standard error: 7099 on 33
degrees of freedom),
# and (Residual standard error: 7461 on 40 degrees of freedom) (17,503 Total)
#Partition
#Linear model from 2 to 43, day 1 not full day (Residual standard error: 7461 on 40 degrees of
freedom)
oilspill43 = subset(oilspill, (day < 44 & day>1), select = c(day,oil)) #Data from day 2 until 43
day43 = oilspill43$day
oil43 = oilspill43$oil
plot(day43,oil43,main = "First Linear Model",xlab = "Day 1 to 43", ylab = "Oil (barrels)")
modelin = lm(oil43 \sim day43)
summary(modelin)
abline(modelin, col="red")
#Separating the Remaining Data
#Three way partition (2nd part)
oilspill44t79 = \text{subset}(\text{oilspill}, (\text{day} > 43 \& \text{day} < 80), \text{select} = c(\text{day}, \text{oil}))
day44t79 = oilspill44t79$day
oil44t79 = oilspill44t79$oil
plot(day44t79, oil44t79,xlab = "Days from 43 to 79", ylab = "Oil (barrels)")
#Three way partition (3rd part)
```

```
oilspill80t85 = \text{subset}(\text{oilspill}, (\text{day} > 79), \text{select} = c(\text{day}, \text{oil}))
day80t85 = oilspill80t85$day
oil80t85 = oilspill80t85$oil
plot(day80t85, oil80t85, xlab = "Day 80 to 85", ylab = "Oil (barrels)")
#Middle part 3ac of Best Model (Residual standard error: 7099 on 33 degrees of freedom)
modelcm= function(t,a,d,c) (a*t-(d*t^2)/2+c)
modelcmfit = nls(oil ~ modelcm(day,mya,myd,myc), data=oilspill44t79, start=c(mya=60000,
myd=10, myc=1)
summary(modelcmfit)
points(day44t79,predict(modelcmfit), col="red")
plot(day44t79,oil44t79,main = "Middle Quadratic Model", xlab = "Day 44 to 79", ylab = "Oil
(barrels)")
lines(day44t79,predict(modelcmfit),col = "blue")
#3rd part linear of Best Model (Residual standard error: 2943 on 4 degrees of freedom)
plot(day80t85,oil80t85, main = "End Linear Model",xlab = "Days 80 to 85", ylab = "Oil
(barrels)")
modelin3 = lm(oil80t85 \sim day80t85)
summary(modelin3)
abline(modelin3, col="red")
#All Models on one graph
plot(day,oil,main = "Entire Model", xlab = "Days", ylab = "Oil (barrels)")
lines(day43,predict(modelin), col="darkorchid1")
lines(day44t79,predict(modelcmfit),col = "darkorange1")
lines(day80t85,predict(modelin3),col = "darkturquoise")
#Inferior Models
#Middle part of other inferior threeway model
#The first and last linear sections remained the same (modelin and modelin3)
```

```
#(Residual standard error: 2943 on 4 degrees of freedom), (Residual standard error: 8361 on 33
degrees of freedom),
# and (Residual standard error: 7461 on 40 degrees of freedom) (18,765 Total)
#Middle part 3bc (Residual standard error: 8361 on 33 degrees of freedom)
modelbcmt= function(t,a,d,c) (a*t-(d*t^3)/3+c)
modelbcmtfit = nls(oil ~ modelbcmt(day,mya,myd,myc), data=oilspill44t79, start=c(mya=60000,
myd=5, myc=1)
summary(modelbcmtfit)
points(day44t79,predict(modelbcmtfit), col="red")
plot(day44t79, oil44t79, xlab = "Day 44 to 79", ylab = "Oil (barrels)")
lines(day44t79,predict(modelbcmtfit),col = "blue")
#Two Step Models
#Both of these use the first linear model (modelin)
#Second Half of Data partition
oilspill44t85 = \text{subset}(\text{oilspill}, \text{day} > 43, \text{select} = \text{c}(\text{day}, \text{oil}))
day44t85 = oilspill44t85$day
oil44t85 = oilspill44t85$oil
plot(day44t85, oil44t85)
#Second half model with 3ac (Quadratic) (7461 on 40 degrees of freedom) and (18360 on 39
degrees of freedom)
modelct= function(t,a,d,c) (a*t-(d*t^2)/2+c)
modelctfit = nls(oil ~ modelct(day,mya,myd,myc), data=oilspill44t85, start=c(mya=60000,
myd=10, myc=1)
summary(modelctfit)
points(day44t85,predict(modelctfit), col="red")
plot(day44t85,oil44t85, xlab = "Day 44 to 85", ylab = "Oil (barrels)")
lines(day44t85,predict(modelctfit),col = "blue")
```

```
#Second half model with 3bc (Cubed) (7461 on 40 degrees of freedom) and (19480 on 39
degrees of freedom)
modelbct= function(t,a,d,c) (a*t-(d*t^3)/3+c)
modelbctfit = nls(oil ~ modelbct(day,mya,myd,myc), data=oilspill44t85, start=c(mya=60000,
myd=5, myc=1)
summary(modelbctfit)
points(day44t85,predict(modelbctfit), col="red")
plot(day44t85, oil44t85, xlab = "Day 44 to 85", ylab = "Oil (barrels)")
lines(day44t85,predict(modelbctfit),col = "blue")
#Single model for all of the data
#Model from 3ac (Quadratic) (Residual standard error: 33640 on 82 degrees of freedom)
modelc= function(t,a,d,c) (a*t-(d*t^2)/2+c)
modelcfit = nls(oil ~ modelc(day,mya,myd,myc), data=oilspill, start=c(mya=60000, myd=10,
myc=1)
summary(modelcfit)
points(day,predict(modelcfit), col="red")
plot(day,oil,xlab = "Days", ylab = "Oil (barrels)")
lines(day,predict(modelcfit),col = "blue")
#Model from 3bc (Cubed) (Residual standard error: 30350 on 82 degrees of freedom)
plot(day,oil)
modelbc= function(t,a,d,c) (a*t-(d*t^3)/3+c)
modelbcfit = nls(oil ~ modelbc(day,mya,myd,myc), data=oilspill, start=c(mya=60000, myd=5,
myc=1)
summary(modelbcfit)
points(day,predict(modelbcfit), col="red")
plot(day,oil, xlab = "Days", ylab = "Oil (barrels)")
lines(day,predict(modelbcfit),col = "blue")
#Linear Model (Residual standard error: 156800 on 83 degrees of freedom)
```

```
plot(day,oil,xlab = "Days", ylab = "Oil (barrels)")
modellinb = lm(oil\sim day)
summary(modellinb)
abline(modellinb, col="red")
#All Model Summaries
#Our Best Model
summary(modelin)
summary(modelcmfit)
summary(modelin3)
#Inferior 3 Step Model
summary(modelin)
summary(modelbcmtfit)
summary(modelin3)
#Two Step Model 1
summary(modelin)
summary(modelctfit)
#Two Step Model 2
summary(modelin)
summary(modelbctfit)
#One Step Model 1
summary(modelcfit)
#One Step Model 2
summary(modelbcfit)
#One Step Model 3
summary(modellinb)
#Failed Models
#Model from 2e
#plot(day,oil)
\#modele = function(t,a,b,c) (a*(t/b-1/b^2)+1+c/(exp(b*t)))
```

```
#modelefit= nls(oil~modele(day,mya,myb,myc), data=oilspill,
start=c(mya=60000,myb=2000,myc=-10000))
#Model from 2a
#plot(day,oil)
\#modela = function(t,a,b,c) (a*t-b*t+c)
#modelafit = nls(oil~ modela(day,mya,myb,myc),data=oilspill,
start=c(mya=2000000,myb=2000,myc=-10000))
#Model from 2b
#plot(day,oil)
\#modelb = function(t,a,b,c) (a/b+c/(exp(b*t)))
#modelbfit = nls(oil ~modelb(day,mya,myb,myc),data=oilspill, start=c(mya=120000, myb=2,
myc=10)
#Model 3bd
#plot(day,oil)
\#modelbd = function(t,a,k,c,d) ((a-d*t^2)/k+((2*d)/k^2)*t-(2*d)/k^3+c*exp(-k*t))
#modelbdfit = nls(oil ~ modelbd(day,mya,myk,myc,myd), data=oilspill, start=c(mya=50000,
myk=10, myc=1000, myd=2000))
#Model from 3ad
#plot(day,oil)
\#modeld = function(t,a,b,c,d) (a/b-d*(t/b-1/b^2)+(c/exp(b*t)))
#modeldfit = nls(oil ~ modeld(day,mya,myb,myc,myd), data=oilspill, start=c(mya=600000,
myb=12, myc=2000, myd=5000))
```