

ANALYSIS OF OIL PIPELINE ACCIDENTS FROM 2010-2017

CIS5270 PROJECT2

NAYANA BRAHMANAHALLI VENKATARAMANAPPA

CIN: 305837628

PRATHIBHA GUBBI PRAKASH

CIN: 305198899

1. **ABSTRACT:**

The oil pipeline has been used for transportation of oil as a relatively safe tool for many years. However, pipelines have caused a high number of accidents so far. Failures of oil pipelines can result from various causes, including technical reasons or external impacts, as well as human error [2]. These causes vary depending on the type of pipeline networks, local regulations ecological circumstances. A wide range of regulatory regimes has addressed safety requirements for the prevention of accidents based on potential risks. Nevertheless, over one billion gallons of oil and petroleum are still spilled every year [3]. For instance, just the Enbridge oil pipeline spillage in July 2010, near Marshall, Michigan in the US which was caused by a leak within a 30-inch pipeline, lead to 840,000 gallons of oil spilled into the Kalamazoo River and adjacent areas4.Therefore, although improvement of safety regulations has reduced the amount of oil spilled per transportation, pipeline accidents still occur everywhere. For avoiding an expansion of damage after oil pipeline accidents, it is held that more attention should be paid to response actions after spillage (ex-post regulations) for combating oil spillage.

1. **DATA SET URL’S:**

**URL:** <https://www.kaggle.com/usdot/pipeline-accidents>

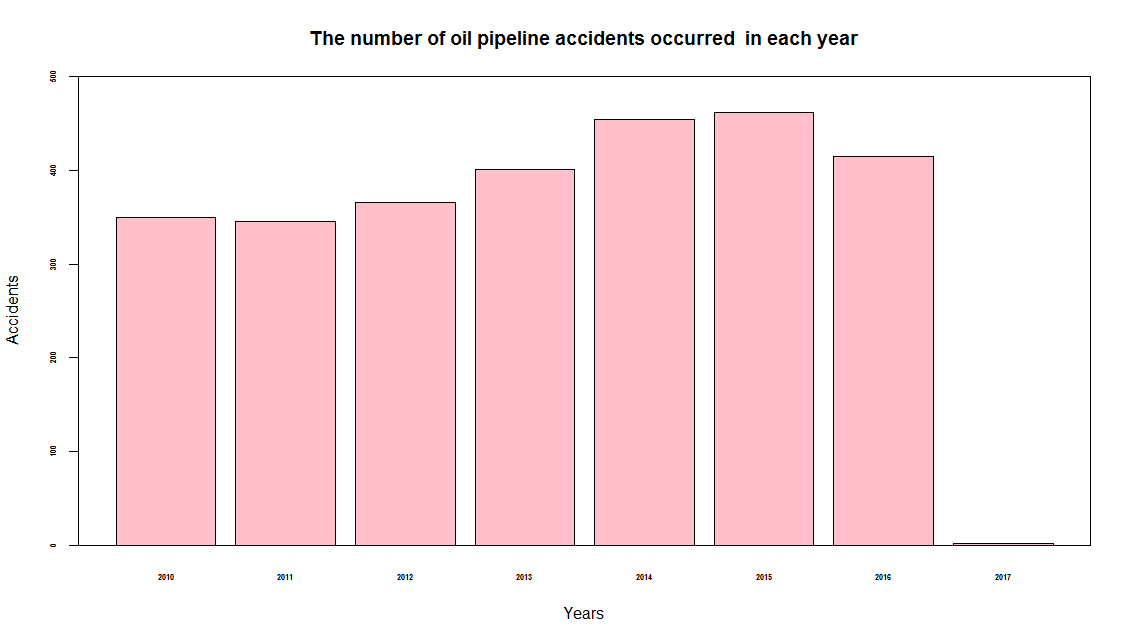
This database includes a record for each oil pipeline leak or spill reported to the Pipeline and Hazardous Materials Safety Administration since 2010. These records include the incident date and time, operator, and pipeline, cause of incident, type of hazardous liquid and quantity lost, injuries and fatalities, and associated costs. This data includes 2351 rows and 47 columns.The oil pipeline accident reports were collected and published by the DOT's Pipeline and Hazardous Materials Safety Administration. There are millions of kilometers of oil pipelines worldwide. It has been that there were 36,275 km of onshore oil pipelines in 2011. There were approximately 170,000 miles of hazardous liquid (mostly crude oil) pipelines in the US in 2012 [1]. This data includes both terrestrial and submarine pipelines.

1. **DATA CLEANING:**

|  |  |  |
| --- | --- | --- |
| **Scope/Problem** | **Dirty Data** | **Cleaned Data/ Remarks** |
| **Columns with empty values**  **(removed unwanted columns)** |  |  |
| **Missing Values** | **C:\Users\USER\Desktop\5270\R-project\na.PNG** | **C:\Users\USER\Desktop\5270\R-project\na removed.PNG** |
| **Splitting Date and Time Columns** | **C:\Users\USER\Desktop\5270\R-project\date.PNG** | **C:\Users\USER\Desktop\5270\R-project\splited.PNG** |

1. **ANALYSIS AND VISUALIZATION:**

**1.Total number of oil pipeline accidents that took place in each year from 2010 to 2017 based on different liquid type and pipeline type**



**Graph Used:** [Bar graph]

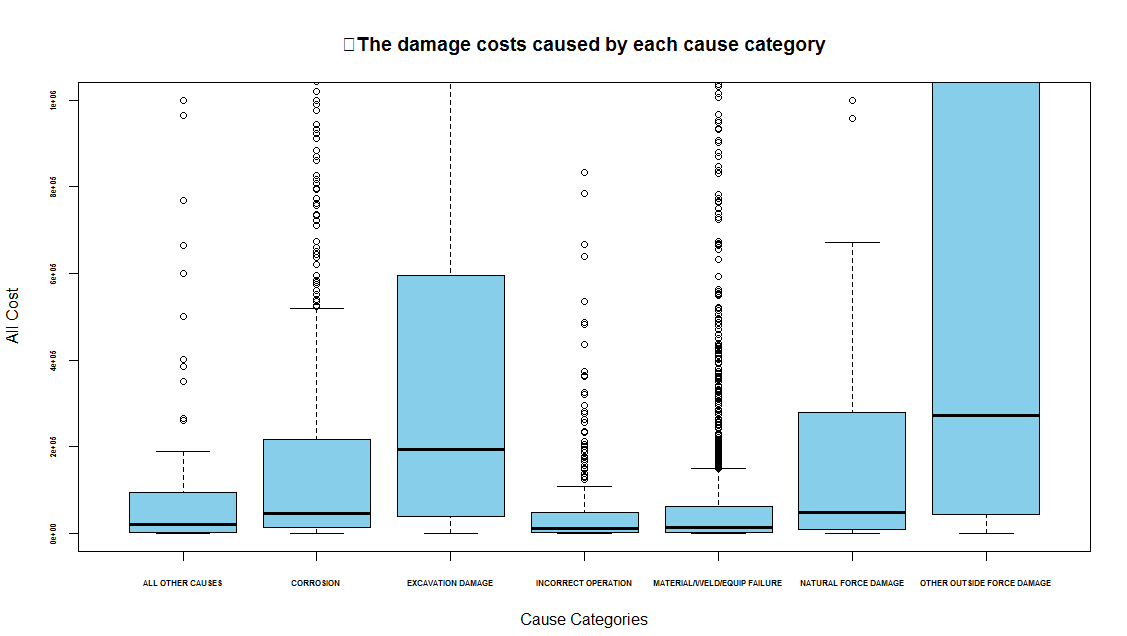
**Parameters Used:** [Xlab, Yalb, Cex.axis, Cex.lab, Col, Ylim, main]

**Axis Specification:**

|  |  |
| --- | --- |
| x-axis | Years from 2010 to 2017 |
| y-axis | Total number of oil pipeline accidents |

Bar graph represents increase in accidents from 2010-2015 after which there is a small decrease form 2015-2016. In the year 2015, the number of oil pipeline accidents that occurred was maximum when compared to all other years. As seen in the above graph the year 2017 has least number of oil pipeline accidents in its first quarter. It can be inferred from the above graph that some precautionary measures were taken to successfully reduce the number of oil pipeline accidents from the year 2015.

**2.Oil pipeline accidents cause categories that were responsible for the all type of damage costs incurred during accidents.**



**Graph Used:** [Box and whisker graph]

**Parameters Used:** [ylim, cex.lab, cex.axis, xlab,ylab, font.axis, col, main]

**Axis Specification:**

|  |  |
| --- | --- |
| x-axis | Different cause categories responsible for oil pipeline accidents |
| y-axis | Different type of costs incurred due to the accidents |

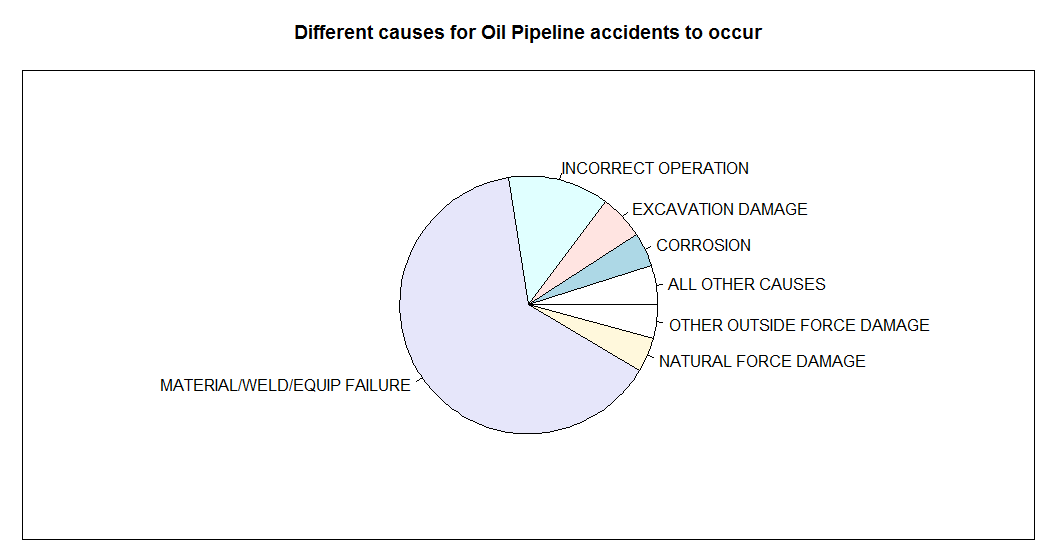
The above Box plot represents the damage costs caused by each cause category

during oil pipeline accidents. It helps them to investigate the cause of the accident and to see what should be done to avoid this in the future. The boxplot is a convenient way of graphically

depicting groups of the total cost through their quartiles. The lines extending vertically from

the boxes (whiskers) indicates variability outside the upper and lower quartiles. The bottom and top of the box are the first and third quartiles, and the band inside the box is the second quartile (the median).

**3.What is the main cause for most of the oil pipeline accidents to occur in the span of the years 2010 to 2017?**



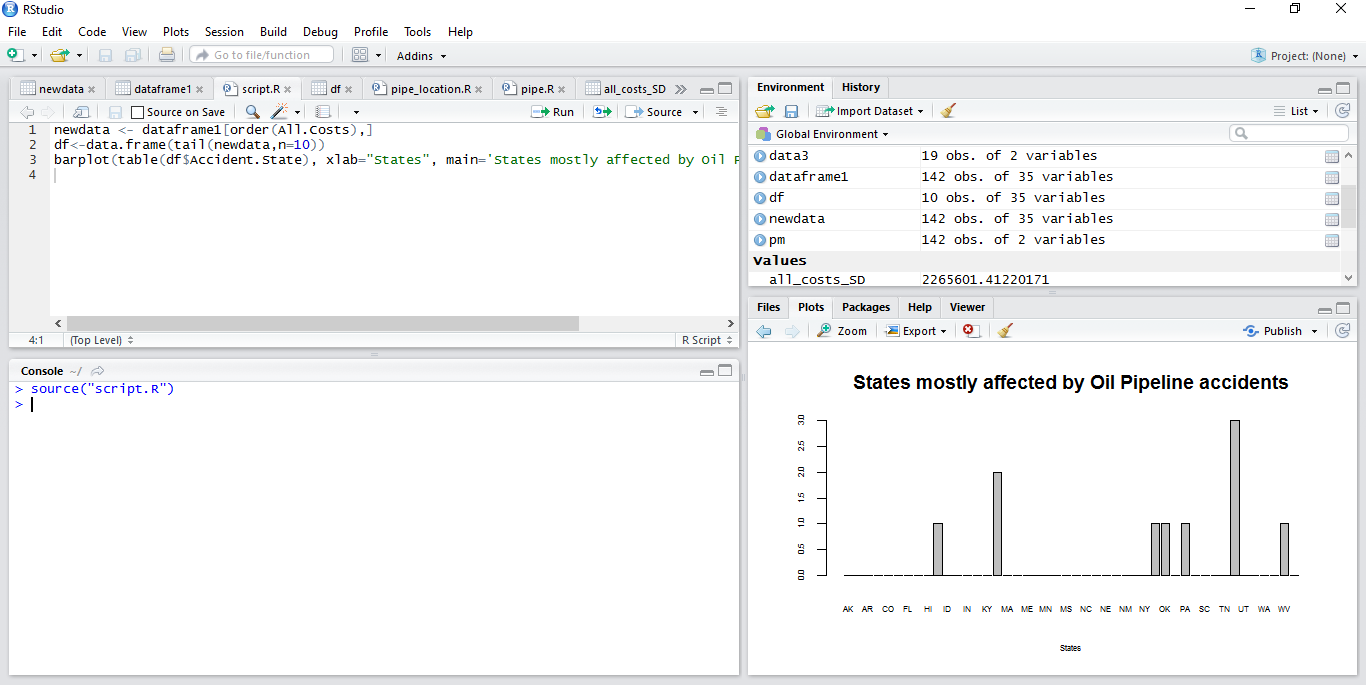
**Graph Used:** [pie Chart]

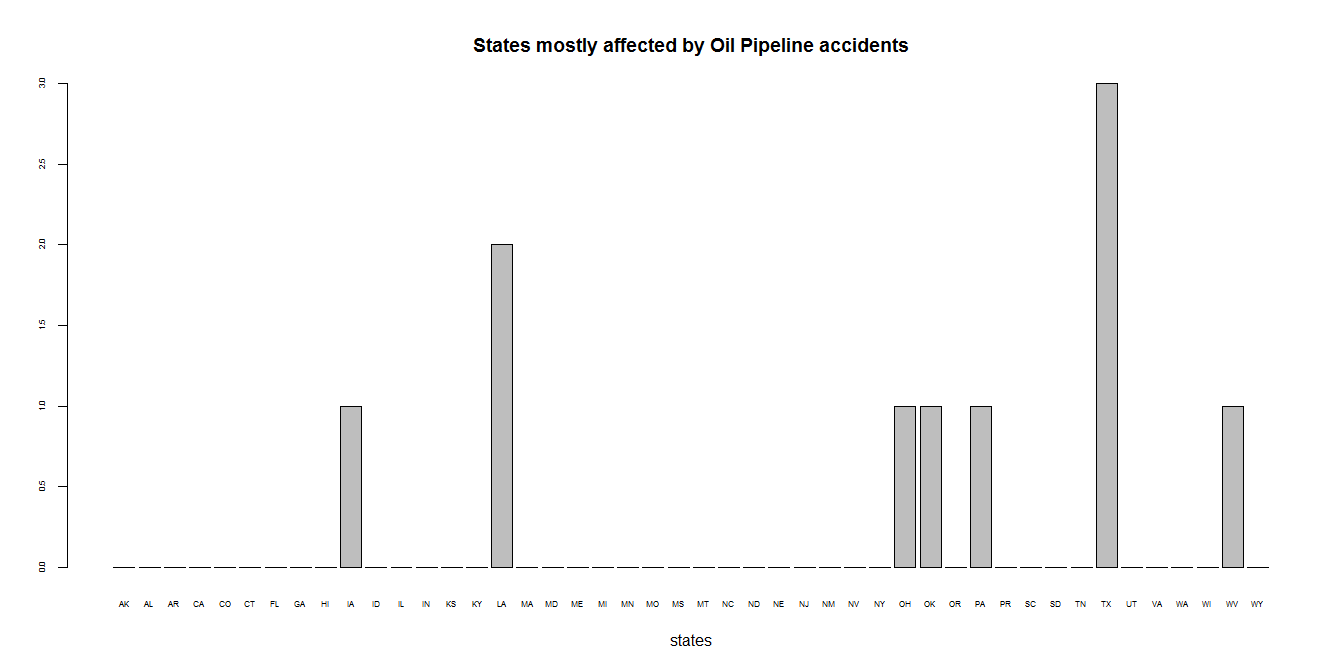
**Parameters Used:** [table, main]

The above pie chart helps in identifying the different type of causes for oil pipeline accidents to occur. It also represents the percentage by which each category was responsible for the accidents to take place. From the above graph, more number of accidents are due to material/equipment failure followed by incorrect operation. More than 50% of the accidents were caused due to the low quality of material used in building the oil pipelines for transportation of oil across different places.

**4.States that are bearing the top 10 highest damage cost due to the oil pipeline accidents over 7 years (2010 to 2017).**

**Using Script:**





**Graph Used:** [Bar graph]

**Parameters Used:** [order, tail, table, xlab, main]

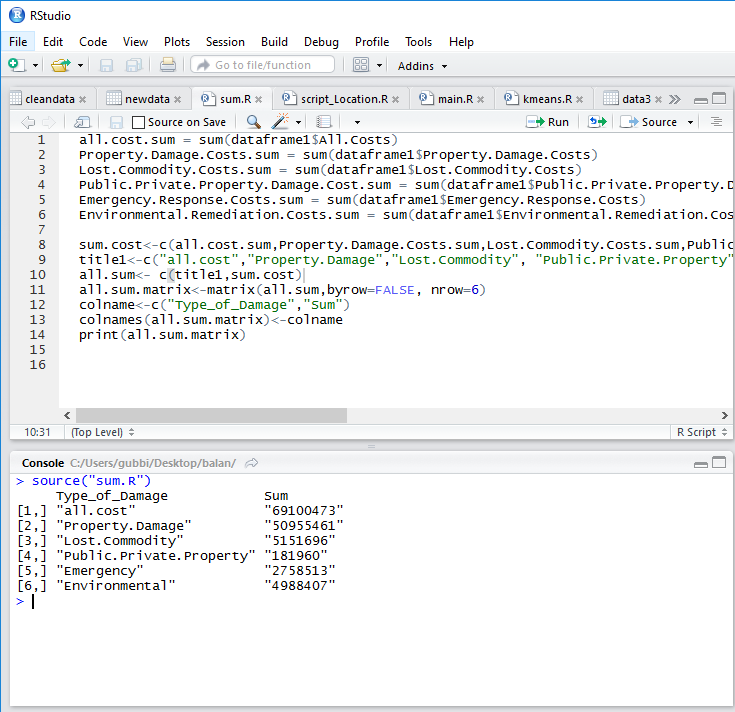
**Axis Specification:**

|  |  |
| --- | --- |
| x-axis | Different states affected by oil pipeline accidents |
| y-axis | All cost of damages |

The above graph helps in analyzing the states that bear the most amount of damage costs caused by the oil pipeline accidents. In order to retrieve this data, we first ordered the entire data frame in increasing order of the damage cost and took the last 10 rows having the highest damage costs. We then represented the states that have incurred these highest damage costs. According to the above graph the state Texas had the maximum damage cost when compared to all other states followed by Louisiana.

**Script:**

**To display a matrix repressing the sum of different type of damage costs over all pipe line accidents that took place from the year 2010 to 2017.**

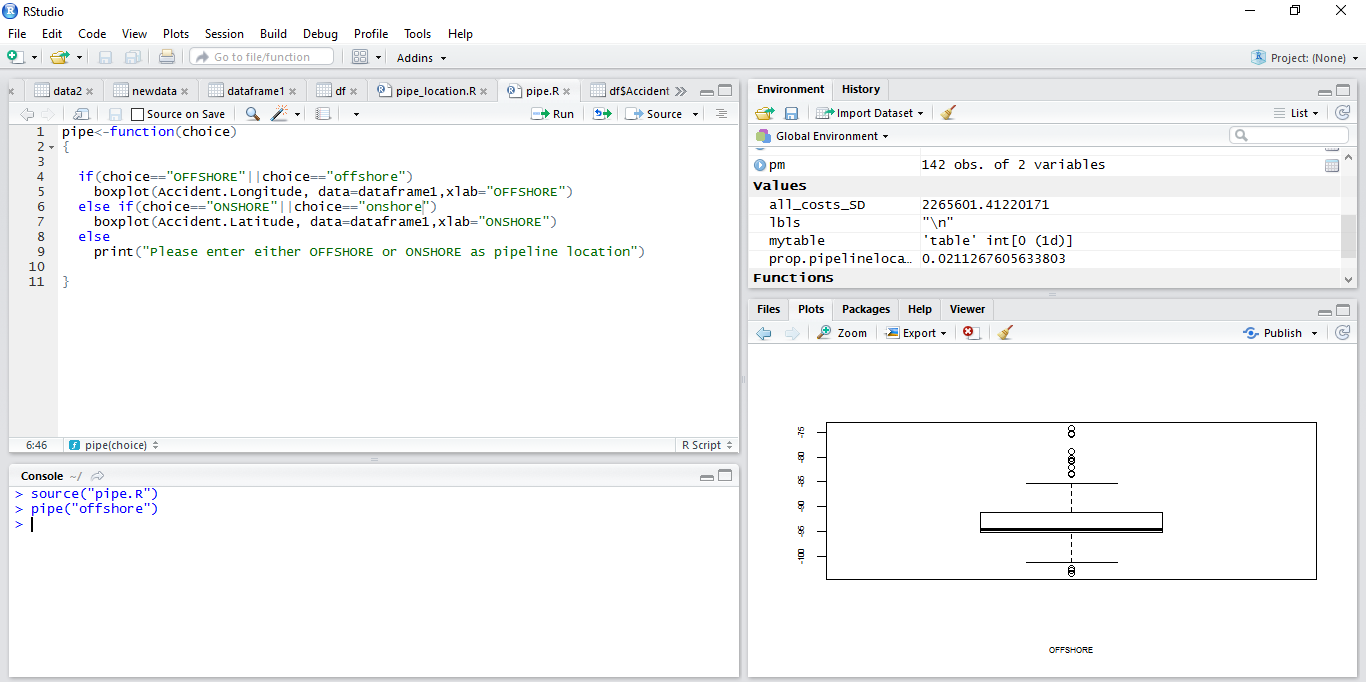
****

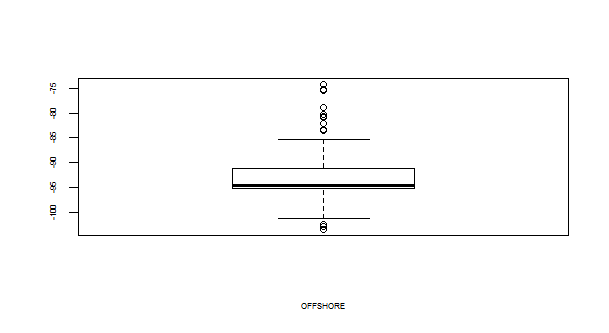
In our dataset, we have different type of damage cost that have occurred due to oil pipeline accidents like property damage cost, lost commodity damage cost, public/private property damage cost, emergency response damage cost, environmental damage cost. The above script helps in displaying the matrix representing the different type of damages and the sum of cost incurred by each damage type. Using this matrix displayed by the script, we can infer property damage was maximum when compared to all other type of damages.

**Function:**

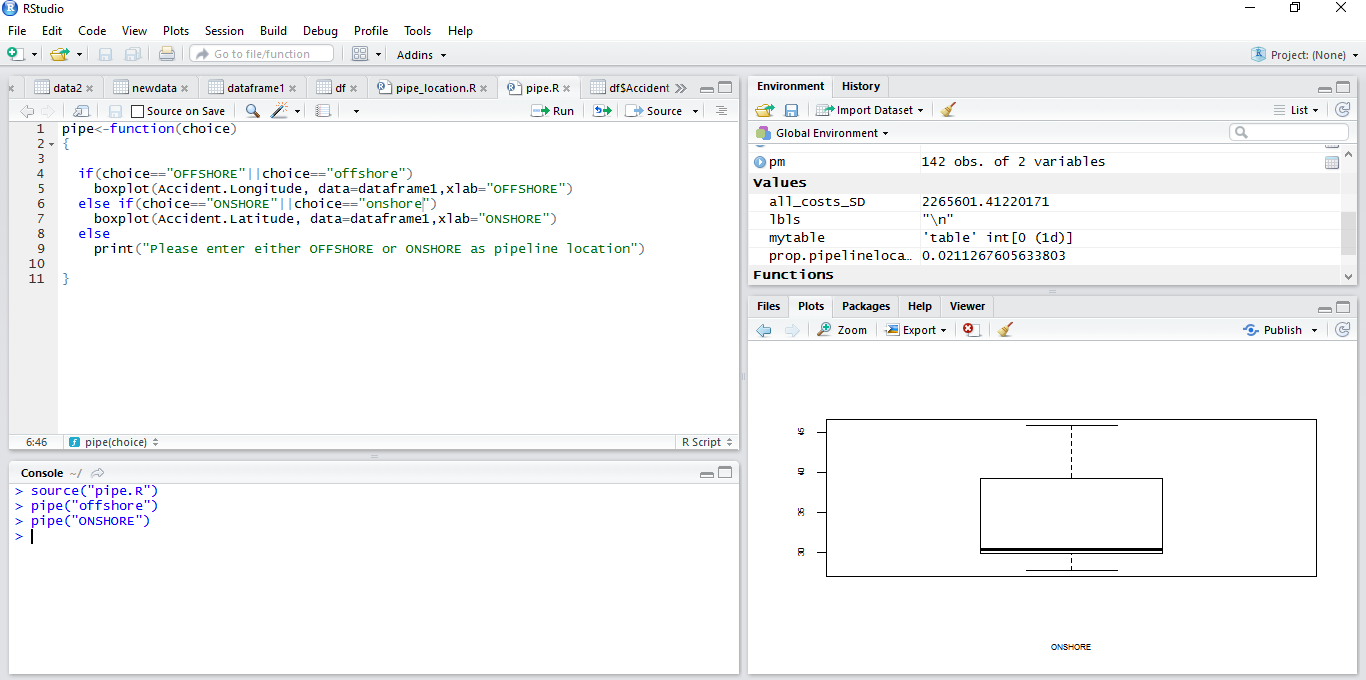
**Function to display a graph for the oil pipeline accidents that have occurred ONSHORE and OFFSHORE based on the input choice of pipeline location type.**

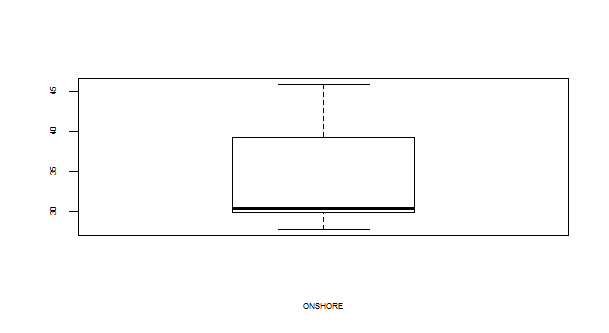
* Input Choice: Offshore
* Output: Box and whisker plot repressing relationship between the offshore pipeline location and number of accidents occurred.





* Input Choice: Onshore
* Output: Box and whisker plot repressing relationship between the onshore pipeline location and number of accidents occurred.

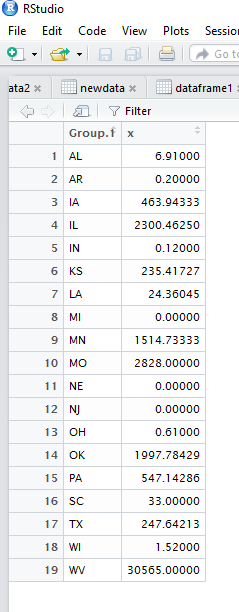




**Statistical Averages:**

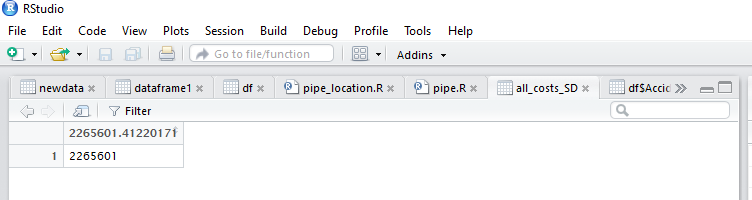
We have applied three types of statistical averages for our data shown as below

* Mean
* Standard Deviation
* Proportion
* **Mean:** Mean is used to find the average value of net loss of quantity of the oil barrels that occurred for each state due to oil pipeline accidents. Mean is calculated using the aggregate function on Net loss of barrels’ field and listed according to each state. From the output of the mean function we can predict the loss in the quantity of oil for each state.

****

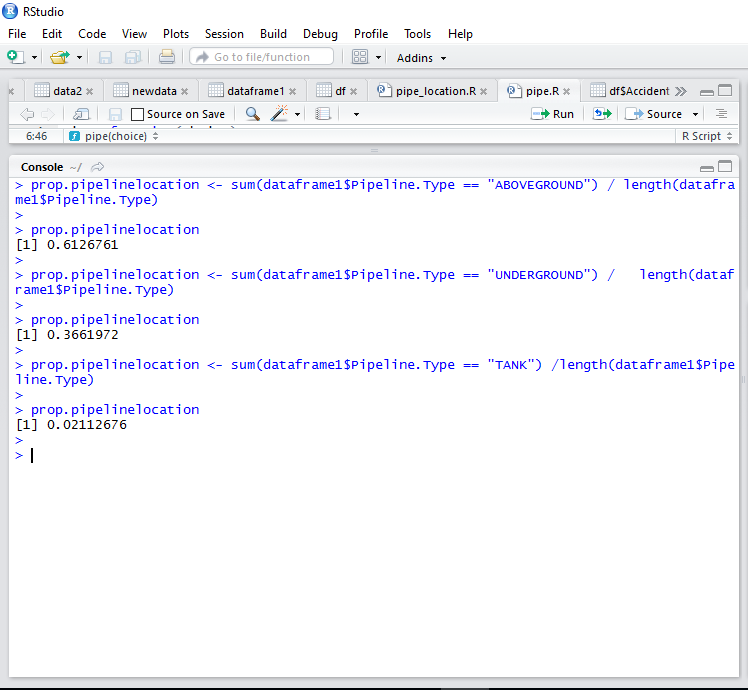
* **Standard Deviation:**

Standard Deviation function is used to calculate the extent of deviation of all damage costs from other damage costs caused by oil pipeline accidents. The standard deviation of all costs was found to be 2265601.



* **Proportion:**

Proportion function is used to find the proportion of the oil pipeline accidents that have occurred based on the different pipeline types like Underground pipeline, aboveground pipeline and tanks.



The above proportions helps us to analyze 0.6% percent of oil pipeline accidents have occurred Aboveground, 0.3% of oil pipeline accidents have occurred in the underground and 0.02% of accidents have occurred to explosion of tanks. From this information, we can predict which is the major cause for accidents to occur and take precautions based on it.

1. **R CODE:**

**Code for Data Upload:**

> setwd("~/")

> data1<-read.csv("database.csv")

> View(data1)

**Data Cleaning:**

* **Code to remove unwanted Columns:**

> data3<-data1[,(-27:-41)]

> View(data3)

* **Code to remove N/A values from rows:**

>newdata<-na.omit(data3)

> View(newdata)

* **Code to split Date and Time Columns:**

> dataframe1$Date <- sapply(strsplit(as.character(dataframe1$Accident.Date.Time), " "), "[", 1)

> dataframe1$Time <- sapply(strsplit(as.character(dataframe1$Accident.Date.Time), " "), "[", 2)

**Code to Add Data to Dataframe:**

> dataframe1<-data.frame(newdata)

**Code for Visualizations:**

* **Pie Chart to represent different causes for Oil Pipeline accidents**

**And which is the main cause for most of the oil pipeline accidents to**

**Occur.**

> pie(table(dataframe1$Cause.Category),main="Different causes for Oil Pipeline accidents to occur")

> box()

* **Box plot to represent the damage costs caused by each cause category during oil pipeline accidents.**

> boxplot(data1$All.Costs~data1$Cause.Category,ylim=c(0,1000000),par(cex.lab=0.5),par(cex.axis=0.5), xlab="Cause Categories", ylab="All Cost", cex.lab=1, col="sky blue", par(font.axis=2), main=("The damage costs caused by each cause category")

* **Bar Plot to represent the number of oil pipeline accidents occurred in each year.**

> barplot(table(data1$Accident.Year), ylim = c(0,500), col="pink", xlab="Years", ylab = "Accidents", cex.lab=1, main="The number of oil pipeline accidents occurred in each year")

**Script:**

* **Bar Plot to represent the States having top 10 highest costs of damages caused by different cause categories**

> newdata <- dataframe1[order(All.Costs),]

> View(newdata)

> df<-data.frame(tail(newdata,n=10))

> View(df)

> barplot(table(df$Accident.State),xlab=’States’,main=”States mostly affected by Oil Pipeline accidents”)

* **To display a matrix repressing the sum of different type of damage costs over all pipe line accidents that took place from the year 2010 to 2017.**

all.cost.sum = sum(dataframe1$All.Costs)

Property.Damage.Costs.sum = sum(dataframe1$Property.Damage.Costs)

Lost.Commodity.Costs.sum = sum(dataframe1$Lost.Commodity.Costs)

Public.Private.Property.Damage.Cost.sum = sum(dataframe1$Public.Private.Property.Damage.Costs)

Emergency.Response.Costs.sum = sum(dataframe1$Emergency.Response.Costs)

Environmental.Remediation.Costs.sum = sum(dataframe1$Environmental.Remediation.Costs)

sum.cost<-c(all.cost.sum,Property.Damage.Costs.sum,Lost.Commodity.Costs.sum,Public.Private.Property.Damage.Cost.sum,Emergency.Response.Costs.sum,Environmental.Remediation.Costs.sum)

title1<-c("all.cost","Property.Damage","Lost.Commodity", "Public.Private.Property","Emer gency","Environmental")

all.sum<- c(title1,sum.cost)

all.sum.matrix<-matrix(all.sum,byrow=FALSE, nrow=6)

colname<-c("Type\_of\_Damage","Sum")

colnames(all.sum.matrix)<-colname

print(all.sum.matrix)

**Function:**

Function to display the oil pipeline accidents that have occurred ONSHORE and OFFSHOR. Function takes pipeline location as the input. The output of the function display the box graph of the oil pipeline accidents individually for OFFSHORE location and ONSHORE location**.**

pipe\_location<-function(choice)

{

if(choice=="OFFSHORE"||choice==”offshore”)

boxplot(Accident.Longitude , data=dataframe1,xlab=”OFFSHORE”)

else if(choice=="ONSHORE"||choice==”onshore”)

boxplot(Accident.Latitude , data=dataframe1,xlab=”ONSHORE”)

else

print("Please enter either offshore or onshore as your location choice")

}

**Statistical Averages: Mean, Standard deviation and Proportion**

**Mean:**

> data3<-aggregate(dataframe1$Net.Loss..Barrels., by=list(dataframe1$Accident.State), FUN=mean)

> View(data3)

**Standard Deviation:**

> all\_costs\_SD<-sd(All.Costs)

> all\_costs\_SD

**Proportion:**

>prop.pipelinelocation <- sum(dataframe1$Pipeline.Type == "ABOVEGROUND") /

length(dataframe1$Pipeline.Type)

> prop.pipelinelocation

>prop.pipelinelocation <- sum(dataframe1$Pipeline.Type == "UNDERGROUND") / length(dataframe1$Pipeline.Type)

> prop.pipelinelocation

> prop.pipelinelocation <- sum(dataframe1$Pipeline.Type == "TANK") /

length(dataframe1$Pipeline.Type)

> prop.pipelinelocation

1. **REFERENCES:**

* <http://primis.phmsa.dot.gov/comm/reports/safety/PSI.html>
* Cheremisinoff, N. et al (2010), Emergency Response Management of Offshore Oil Spills: Guidelines for Emergency Responders. The US: John Wiley & Sons.
* Department of environmental quality & Enbridge (2010), Administrative Consent Order & Partial Settlement Agreement, Michigan the US.