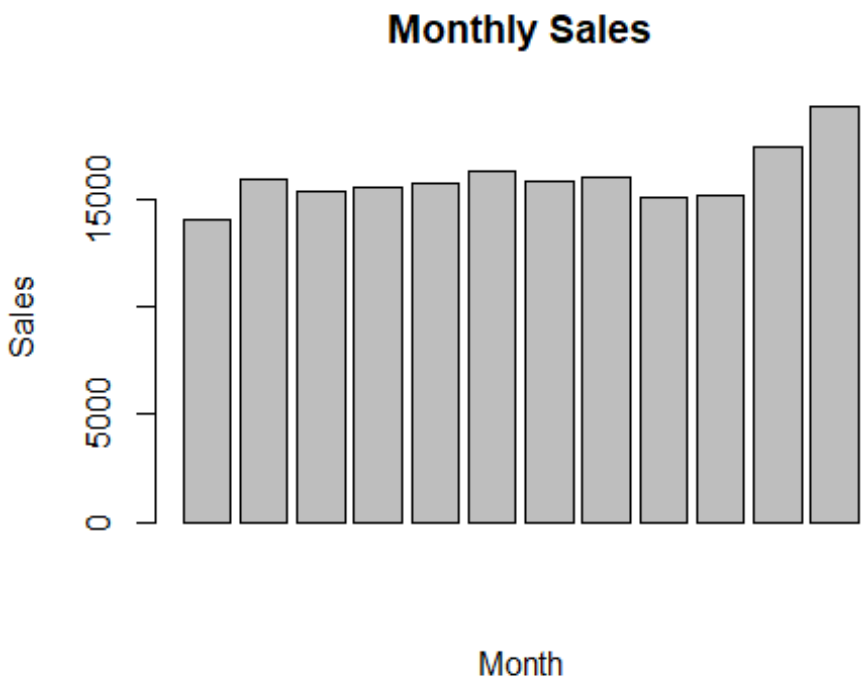
MVA project

Lokesh Arora, Ankita Shinde

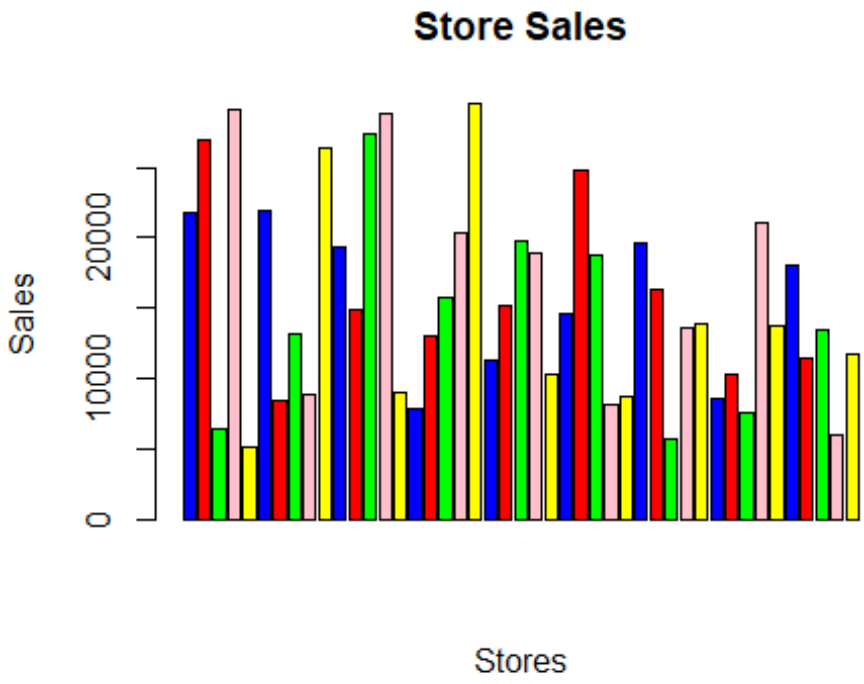
10/1/2020

https://github.com/lokesh2334/walmart

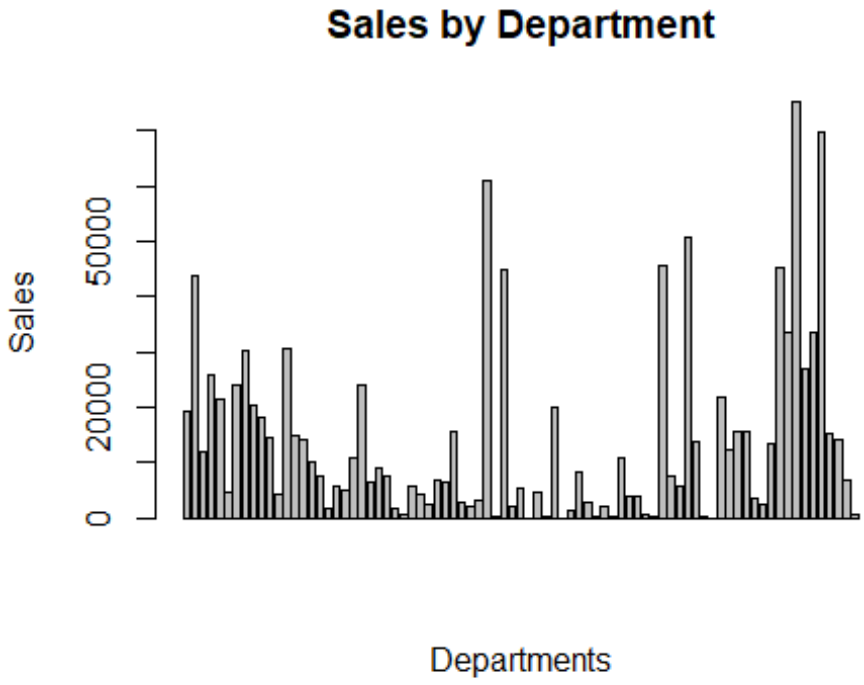
```
features = read.csv("D:/Ankita/SEMESTER 2/MVA/features.csv",header = T)
train = read.csv("D:/Ankita/SEMESTER 2/MVA/train.csv",header = T)
test = read.csv("D:/Ankita/SEMESTER 2/MVA/test.csv",header = T)
stores = read.csv("D:/Ankita/SEMESTER 2/MVA/stores.csv",header = T)

dataset = merge(train,stores, by="Store")

#dataset = merge(dataset,features, by="Store")

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

dataset$year <- year(ymd(dataset$Date))
dataset$month <- month(ymd(dataset$Date))
dataset$day <- day(ymd(dataset$Date))

library("plyr")
month_wise =ddply(dataset, .(month), summarize,  Sales=mean(Weekly_Sales))
barplot(month_wise$Sales, main="Monthly Sales", xlab="Month", ylab="Sales")
```




```
store_wise =ddply(dataset, .(Store), summarize,  Sales=mean(Weekly_Sales))

barplot(store_wise$Sales, main="Store Sales", xlab="Stores", ylab="Sales", co
l = c('blue', 'red', 'green', 'pink', 'yellow'))
```




```
dept_wise =ddply(dataset, .(Dept), summarize,  Sales=mean(Weekly_Sales))
barplot(dept_wise$Sales, main="Sales by Department", xlab="Departments", ylab
="Sales")
```



```r
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
dept_sales =ddply(dataset, c("year","Dept"), summarize,  Sales=sum(Weekly_Sal
es))
top_depts = top_n(dept_sales, 30, Sales)
```