HP-Exploratory analysis

preethi

11/23/2019

Loaded the required packages

```
library(googledrive)
```

```
## Warning: package 'googledrive' was built under R version 3.6.1
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.6.1
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.1
```

```
library(scales)
```

```
## Warning: package 'scales' was built under R version 3.6.1
```

```
library(reshape2)
```

Download file and read it

```
temp <- tempfile(fileext = ".zip")
dl <- drive_download(
  as_id("https://drive.google.com/open?id=1dyKTsCDegBCDDBDVJRJAf_JIUxd8YcET"), path = temp, overwrite =
out <- unzip(temp, exdir = tempdir())
expenditure <- read.csv(out, header = TRUE, stringsAsFactors = FALSE)

budget <- read.csv("hoa_wise_prep_data.csv", header = TRUE, stringsAsFactors = FALSE)
```

```
## change date format
expenditure$TRANSDATE <- as.Date(expenditure$TRANSDATE, format = "%Y-%m-%d")
budget$date <- as.Date(budget$date, format = "%Y-%m-%d")
```

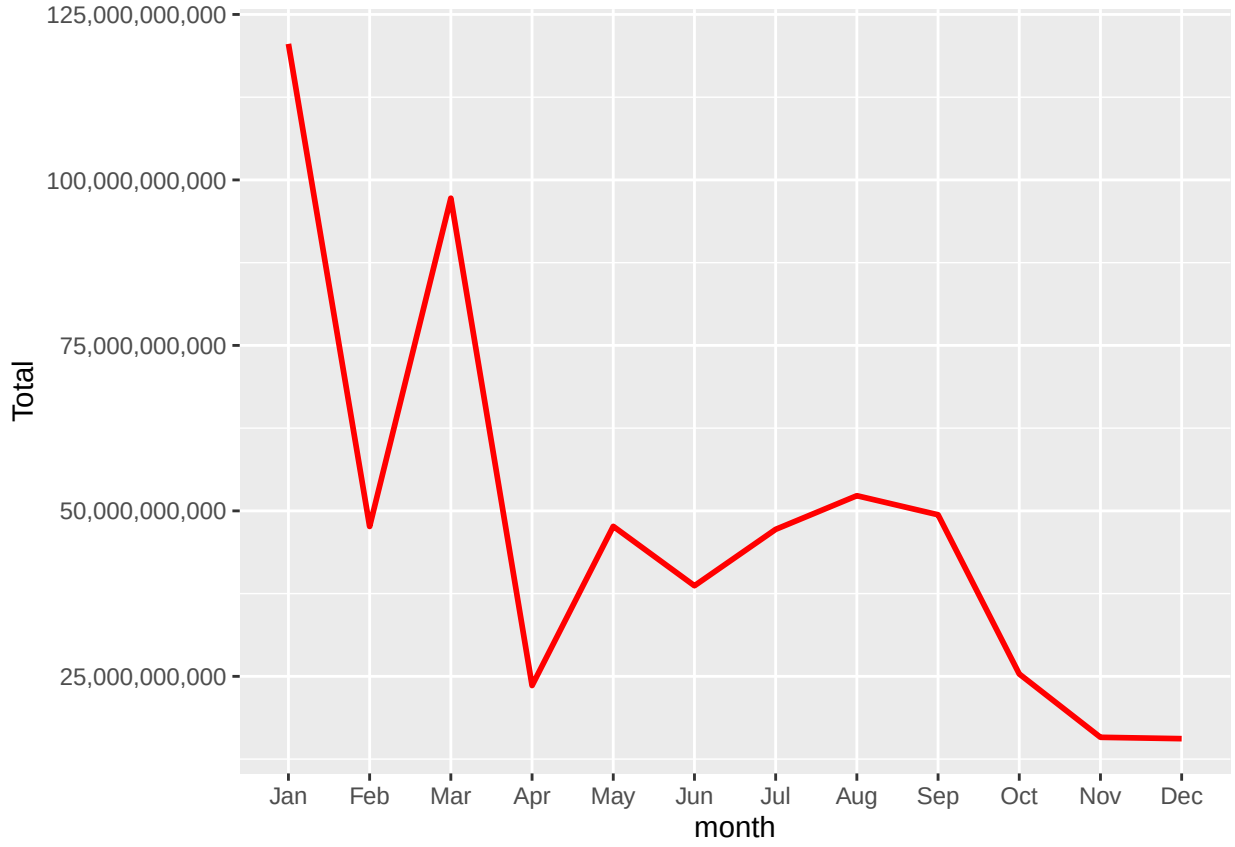
Add a month column to sort by month

```
expenditure <- expenditure %>% mutate(month = month(TRANSDATE , label = TRUE))
##expenditure$month <- as.Date(paste0("2018-", expenditure$month,"-01"),"%Y-%m-%d")
budget <- budget %>% mutate(month = month(date, label = TRUE))
```

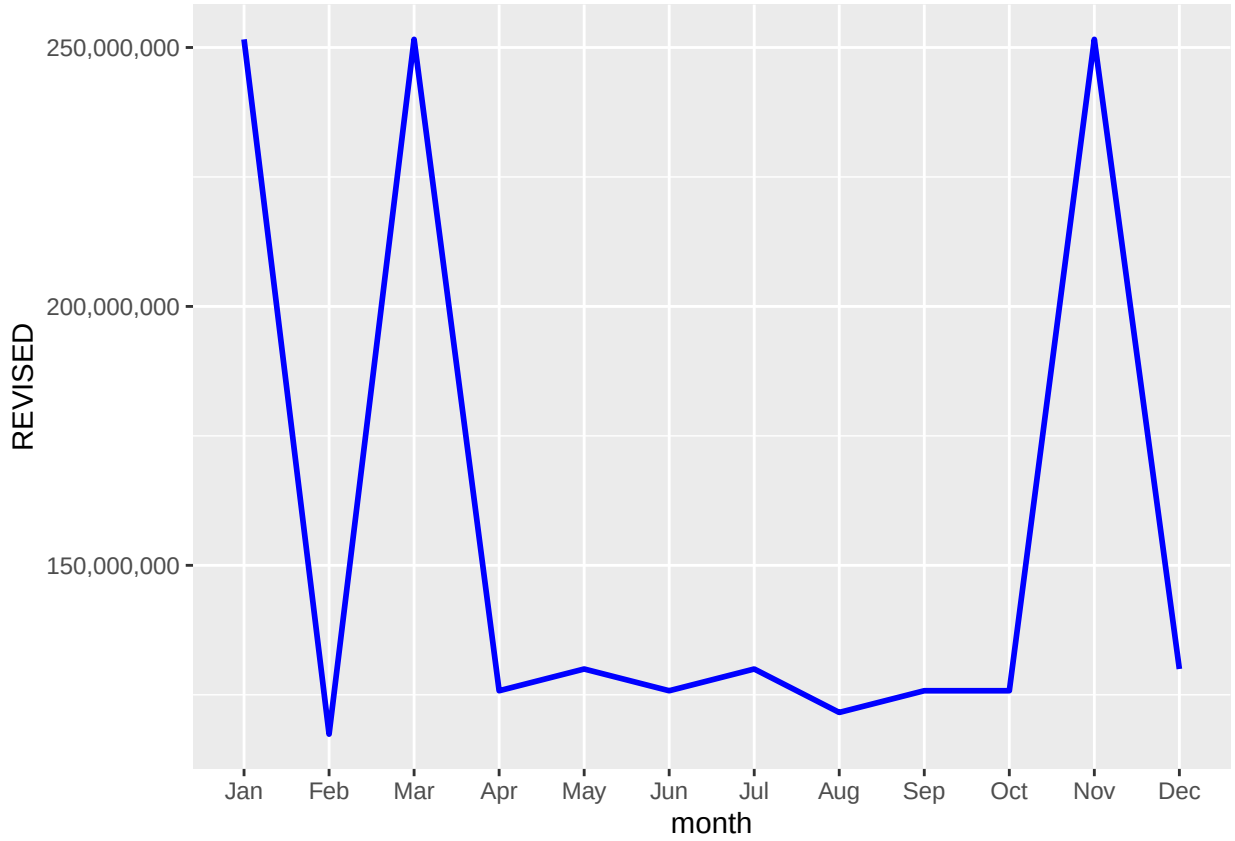
create new tables with the different groups of interest

```
## expenditure data grouped by month and netpayment
Sumpayment <- expenditure %>% group_by(month)%>%summarise(Total = sum(NETPAYMENT, na.rm = TRUE))

##budget data grouped by month and revised and sanctioned amounts
BudgetSum <- budget %>% group_by(month) %>% summarise_at(c("REVISED","SANCTION"), sum, na.rm = TRUE)


BudgetSum.long <- melt(BudgetSum, id = "month", measure = c("REVISED","SANCTION"))
```

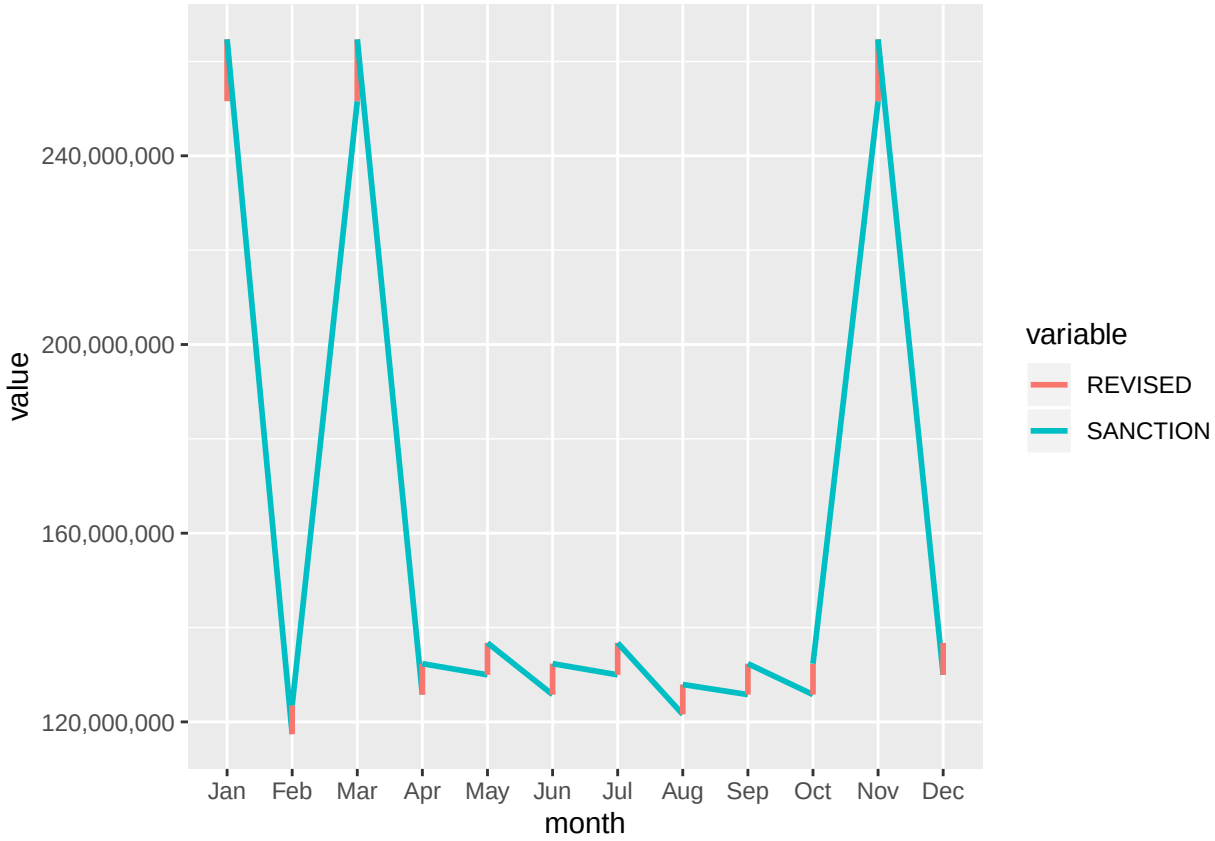
```
## budget data grouped by medical budget for  revised and sanctioned
medical_budget <- medical_budget <- budget %>% group_by(month) %>% filter(major == 2210) %>% summarise_a

medical_budget.long <- melt(medical_budget, id = "month", measure = c("REVISED","SANCTION"))

##medical budget
medical_budget_type <-subset(budget, major == 2210)

## plot of total expenditure monthly for state

plot1 <- ggplot (data = Sumpayment, aes(x = month, y=Total, group = 1)) + geom_line(color = "red", size=

print (plot1)
```



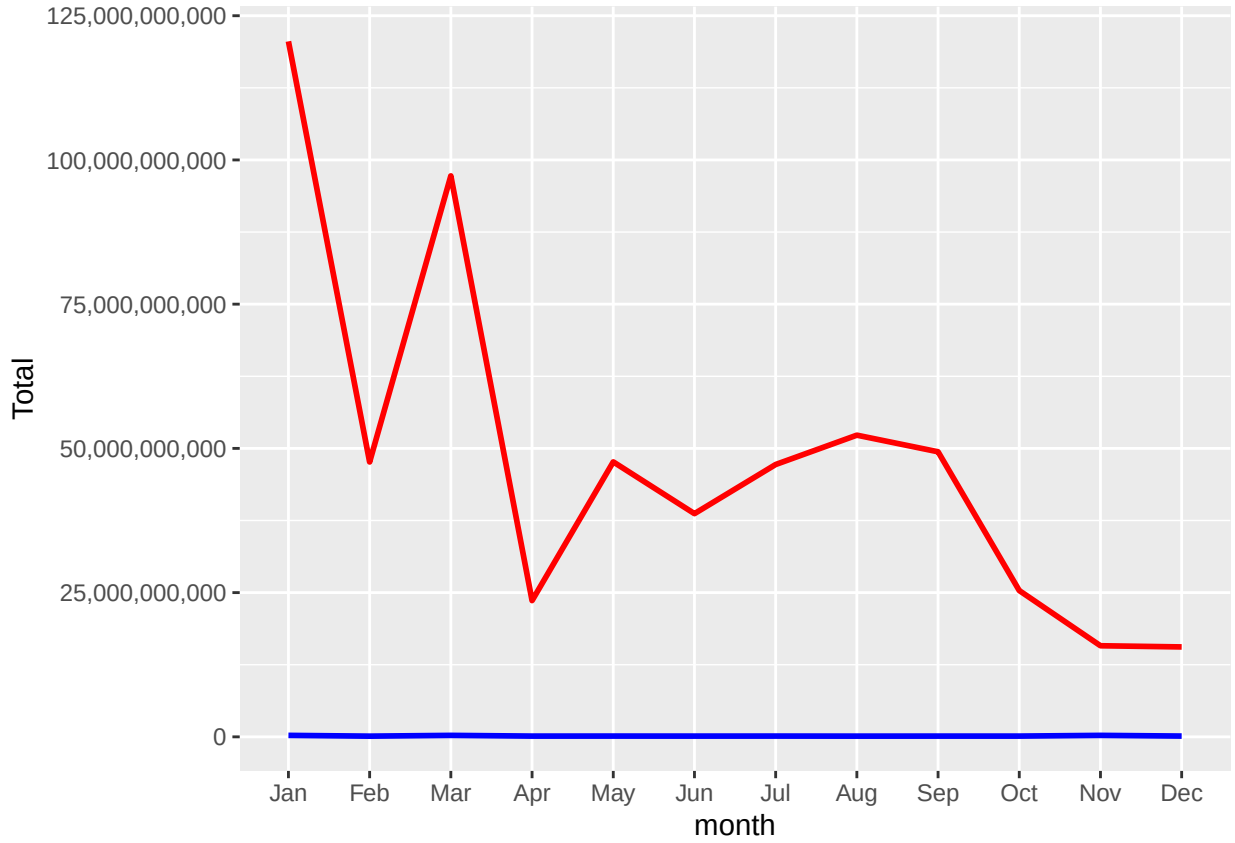

```
##plot of monthly budget data

plot2a <- ggplot (data = BudgetSum, aes(x = month, y=REVISED, group = 1)) + geom_line(color = "blue", si

print (plot2a)
```



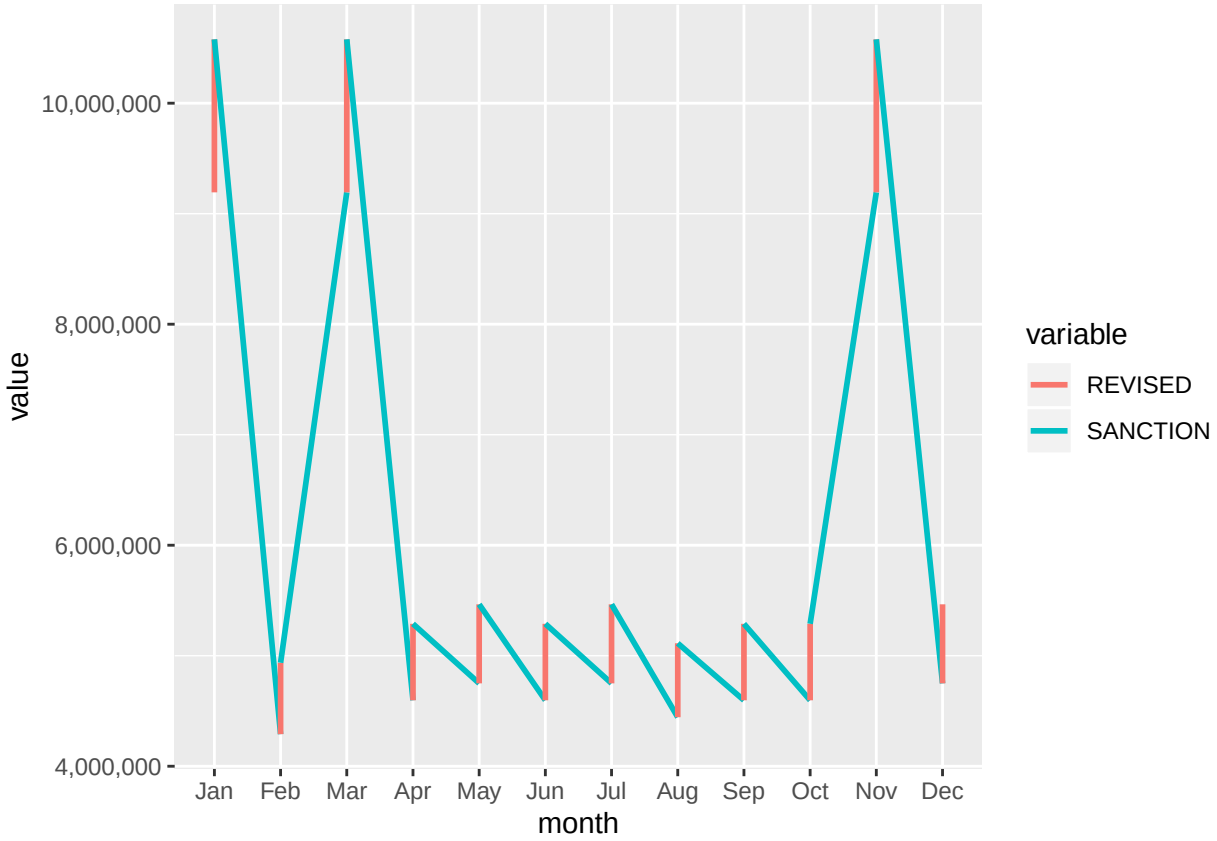

```
plot2 <- ggplot(data = BudgetSum.long, aes(x= month,y=value,color = variable, group = 1))+ geom_line(si

print(plot2)
```




```
##plot of budget and expenditure data
plot3 <- ggplot() + geom_line(data = Sumpayment, aes(x = month, y=Total, group = 1), color = "red", size

print(plot3)
```




```
##plot of monthly medical budget estimates (revised and sanctioned)
plot4a <- ggplot(data = medical_budget.long, aes(x= month,y=value,color = variable, group = 1))+ geom_li

print(plot4a)
```



Districtwise Spending

There are supposed to be 12 districts in Himachal. There is a lot more data here.

data that is not working yet

Districtspend <- expenditure %>% group_by(District)%>% summarise(Total = sum(NETPAYMENT, na.rm = TRUE))

District_spending_month <- expenditure %>% group_by(month,District,SOE_description) %>% summarise(Total = sum(NETPAYMENT, na.rm = TRUE))

##total expenditure monthly per district

district_plot <- ggplot(data = District_spending_month, aes(x=month,y=Total, group =1))+ geom_line(color = "darkorchid4") + facet_wrap(~ District, ncol = 7) + labs(title = "Total Expenditure by district", subtitle = "Data plotted by month", y = "total expenditure", x = "month") + theme_bw(base_size = 15) + scale_y_continuous(labels = scales::comma)

print(district_plot)

##pie chart with district wise spending

barplot <- ggplot(subset(District_spending_month, District == "AMB"), aes(x="",y = Total, fill = SOE_description))+geom_bar(stat="identity", width = 1)

pie <- barplot + coord_polar("y", start=0)

pie

##not clean