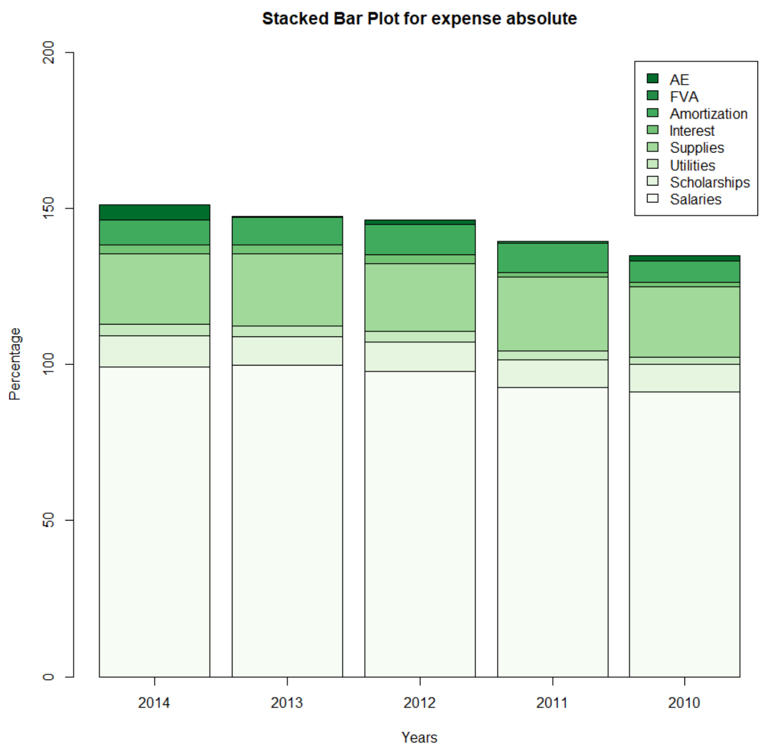
Question 3 (25 points)

For this question, we will use real-world data which came from a report to the Board of Governors in 2014. There are csv files on the learning system you can use that show revenue and expense as absolute and relative values. Using R:

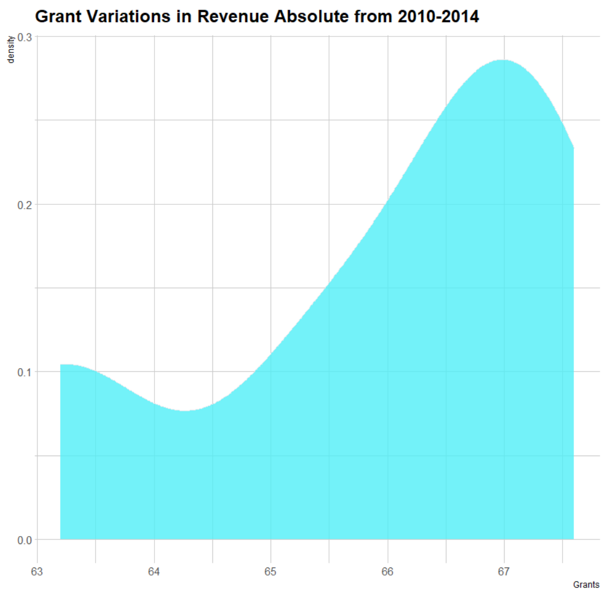
a) Show two different graphs that make the data look good

1 – Stacked Bar Plot



For the representation of Expense Absolute values for years 2010-2014, I used a stacked bar plot, with proper use of color palette, x and y axes labels, and legends. Here each bar displays the Salaries, Scholarships, Utilities, Supplies, Interest, Amortization, FVA and AE for each year. A stacked bar plot is a good graph if used properly and for a limited number of values. As we can see, the graph helps the viewer apprehend what area expense formulates most of the budget. For instance, we can clearly see that the white portion, which represents Salaries (as the legend says), takes the most area. Thus, most of the budget goes into Salaries expense. The trend is noticeable too, like the amount of budget going into Salaries increased by almost 10% from year 2012 to 2014.

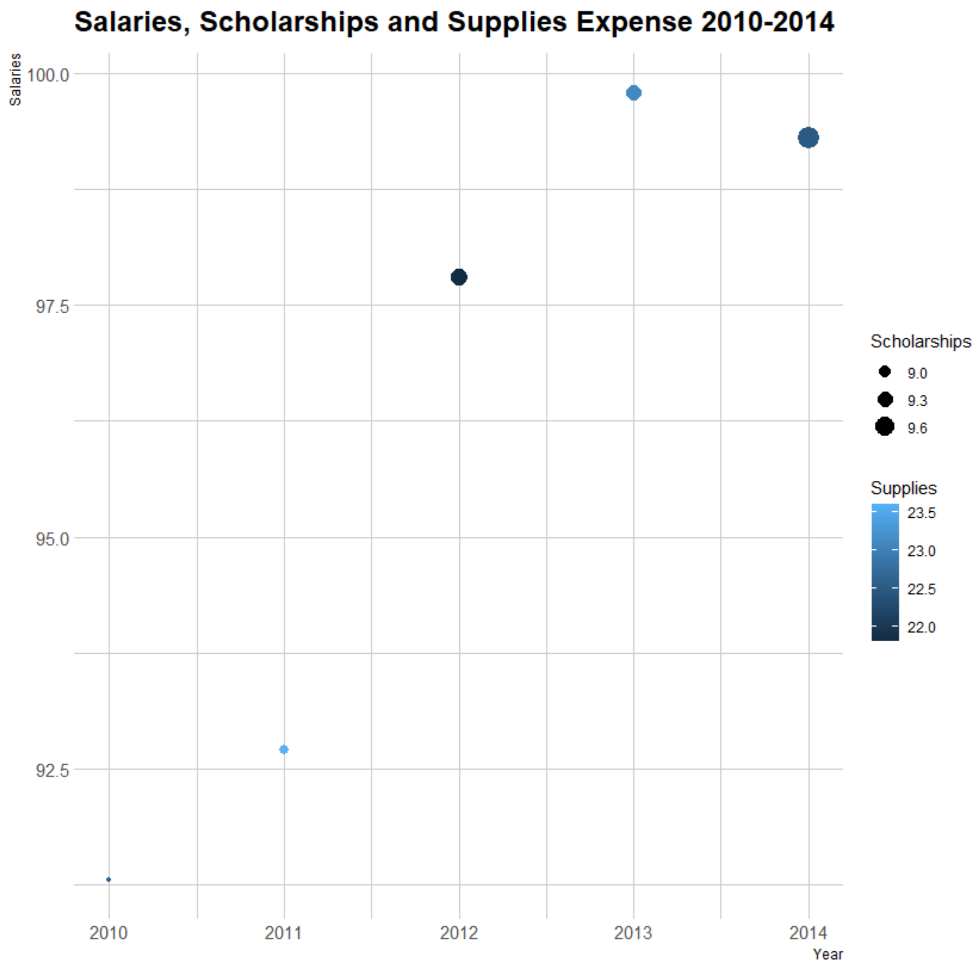
2 – *Basic Density Plot*



Another graph that would represent the data in a good way would be a Basic Density Plot. A density plot is a good way to analyse the trend of a particular attribute through a period of time. For example, considering Grants in the Revenue Absolute data, one might want to know how their Revenue value is being influenced by Grants. According to the graph, Over the course of four years (2010-2014), the Grants have been concentrated into values over 66 – thus helping in perceiving the revenue by the viewer. Since, I did a density plot of one variable, there is no legend. I have added labels for x and y axes.

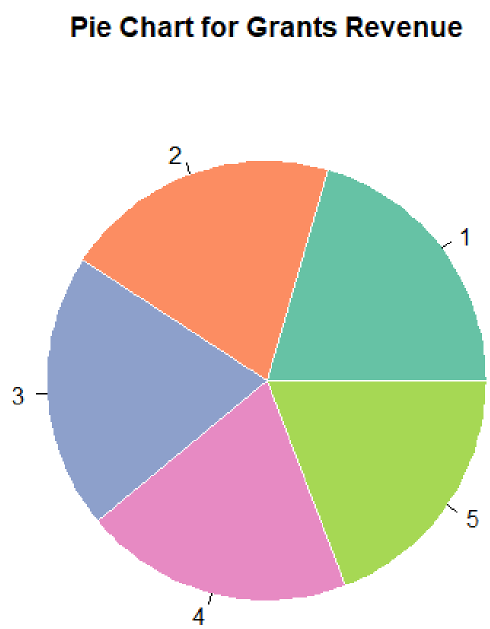
b) Show two different graphs that make the data look bad

1 – Scatter Plot



This is a scatter plot for data Expense Absolute. Scatter Plot is a good way to display multiple variables via multiple attributes like color, shape, size etc. However, improper use of pre-attentive features will result in a bad graph that make it difficult to extract values from it. For example, in this graph, the Salaries of the expense absolute data have been represented on the y-axis. The x-axis has the Year. Scholarships are displayed using size and the color displays Supplies. Now consider year 2010, the size of the dot is the smallest, hence depicting that the scholarship were the least in 2010, also being at the very bottom means, that the Salaries expense was least in that year. However, such a small size makes it hard to extract the exact color of the circle – is it less than or more than the Supplies value in the year 2012 and 2013 – the information extracted can be wrong.

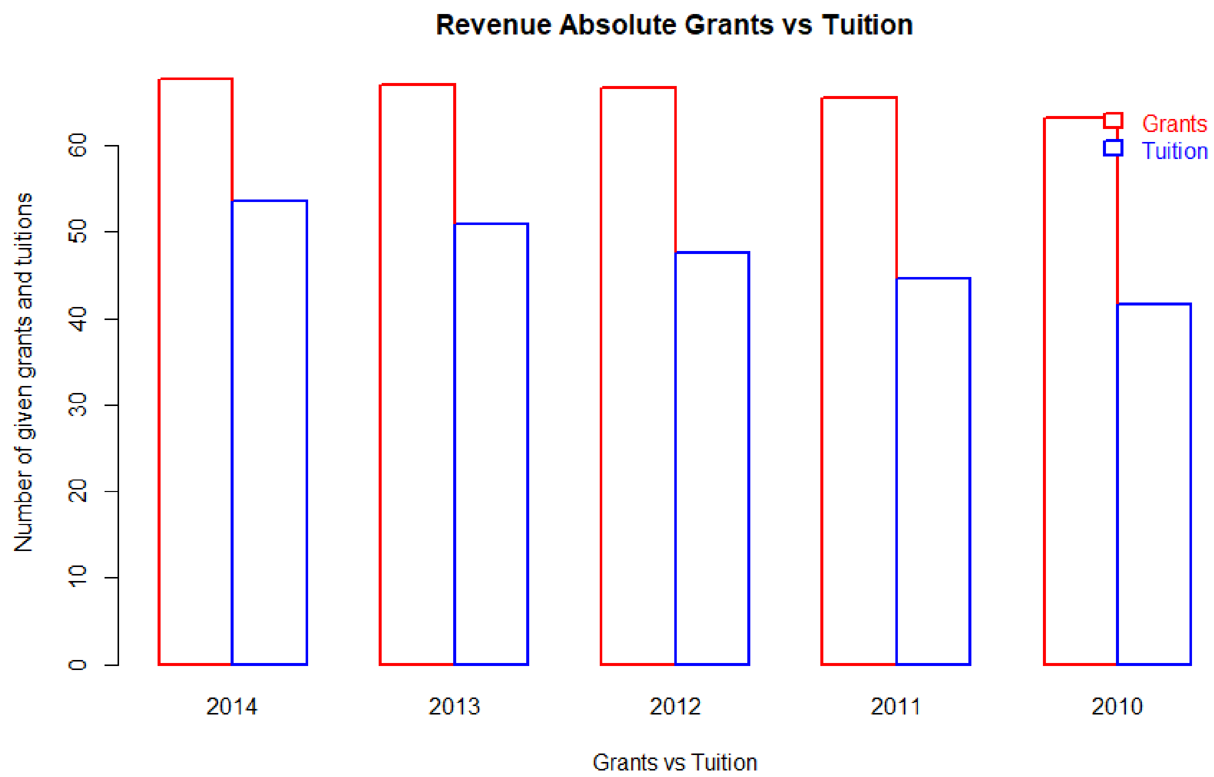
2- Pie Chart



To represent graph in a bad way, I chose a Pie Chart for Revenue Absolute Grants data. A Pie chart is never a good option to represent comparisons – especially when the data values are very close to each other. For example, considering Grants in Revenue, the values over the years 2010-2014 vary between 63 to 68. Now, in a pie chart, this low difference can not be displayed properly. Hence, when we try to extract information from the pie chart, we can not exactly extract any information since all the triangles look the same sizes. Also, there is no mention of which triangle represents what. The numbers 1-5 actually point towards the years 2010-2014, but the user has no way to know that.

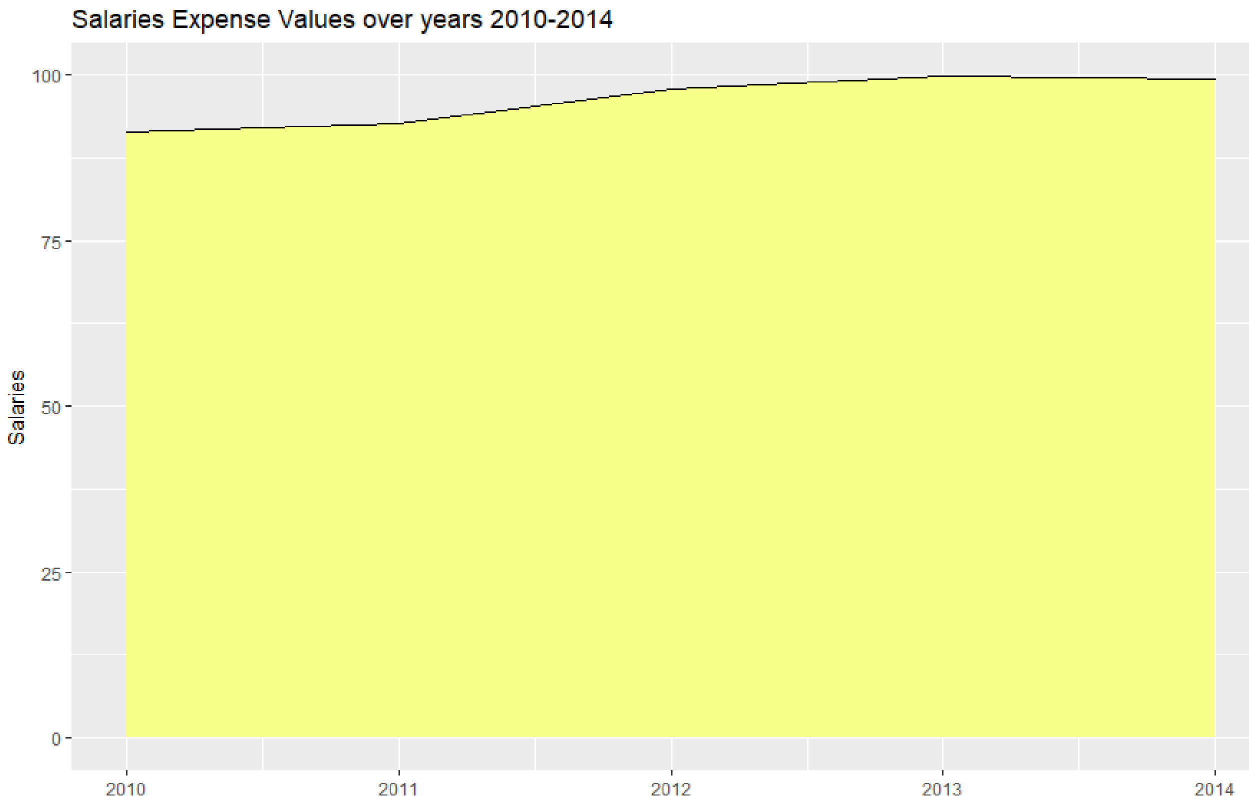
c) Show two different graphs that show things are pretty much the same over those 5 years

1 – Double Column Chart



To show that the data has remained pretty much same over the 5 years 2010-2014, I chose a double column chart. I used the Grants and Tuition data from the Revenue Absolute data. We can see that over the years, the Grants value has decreased by a very small amount – almost negligible if the graph is seen at one glance. For comparison I added the Tuition value. We can clearly see that the Tuition value raises from just over 40 to more than 50. This comparison also helps us see that the data has stayed pretty much same for Grants given from 2010 to 2014.

2 – Area Chart



I used an Area Chart to show that the Salaries Expense value has stayed pretty much the same over 5 years (2010-2014). The black line over the graph shows that the change has been very less and s it can be said that there was no sudden increase in the expense of Salaries.

R source code attached below in this same document/pdf.

R source code :

```
expabs <- read.csv("C://Users//punya//Documents//Priyam//2020 FALL//COIS-
3510H//assn2//ExpenseAbsolute.csv")
expper <- read.csv("C://Users//punya//Documents//Priyam//2020 FALL//COIS-
3510H//assn2//ExpensePercent.csv")
revabs <- read.csv("C://Users//punya//Documents//Priyam//2020 FALL//COIS-
3510H//assn2//RevenueAbsolute.csv")
revper <- read.csv("C://Users//punya//Documents//Priyam//2020 FALL//COIS-
3510H//assn2//RevenuePercent.csv")
head(expabs)
head(expper)
head(revabs)
head(revper)

library(ggplot2)
library(dplyr)
library(hrbrthemes)
library(fmsb)
library(RColorBrewer)
brewer.pal.info

#good graph 1 - bar plot
coul <- brewer.pal(9, "Greens")
bp <- barplot(t(expabs[ , -1]),legend=TRUE, ylim = c(0,200), col = coul,main="Stacked Bar Plot for expense
absolute",xlab = "Years", ylab = "Percentage")
axis(side = 1, at = bp, labels = expabs$Year)

#good graph 2 – Scatter Plot
#source : https://www.r-graph-gallery.com/21-distribution-plot-using-ggplot2.html
revabs %>%
  ggplot( aes(x=Grants)) +
  geom_density(fill="#50EFF7", color="#e9ecef", alpha=0.8) +
  ggtitle("Grant Variations in Revenue Absolute from 2010-2014") +
  theme_ipsum()

#bad graph 1 - scatter plot
#source : https://www.r-graph-gallery.com/274-map-a-variable-to-ggplot2-scatterplot.html
ggplot(expabs, aes(x=Year, y=Salaries, size=Scholarships, color=Supplies)) +
```

```
  ggtitle("Salaries, Scholarships and Supplies Expense 2010-2014") +
  geom_point() +
  theme_ipsum()

#bad graph 2 - Pie Chart
piedata <- revabs$Grants
pie(piedata,border="white", col=myPalette, main="Pie Chart for Grants Revenue")

#data same through years graph 1 -double column chart
cols <- c('red','blue');
ylim <- c(0,max(revabs[c('Grants','Tuition')])); par(lwd=2); barplot(
  t(revabs[c('Grants','Tuition')]), legend=TRUE,  beside=T,   ylim=ylim,   border=cols,
  col='white',   names.arg=expabs$Year,
  main = "Revenue Absolute Grants vs Tuition",   xlab='Grants vs Tuition',
  ylab='Number of given grants and tuitions',
  args.legend=list(text.col=cols,col=cols,border=cols,bty='n')
)

#data same through years graph 2 - Area Graph
#source : https://www.r-graph-gallery.com/164-area-chart-ggplot2.html
xValue <- 1:50
yValue <- cumsum(rnorm(50))
data <- data.frame(xValue,yValue)
ggplot(expabs, aes(x=Year, y=Salaries)) +
  geom_area(fill="#F7FE89", color="#000000") +
  ggtitle("Salaries Expense Values over years 2010-2014")
```