

Student: Seif Kungulio
Date: 02/16/2025
Subject: Project 5
Class: DSCI 502
Section: 01W
Instructor: Sean Yang
File Name: Project5_Kungulio_Seif.docx

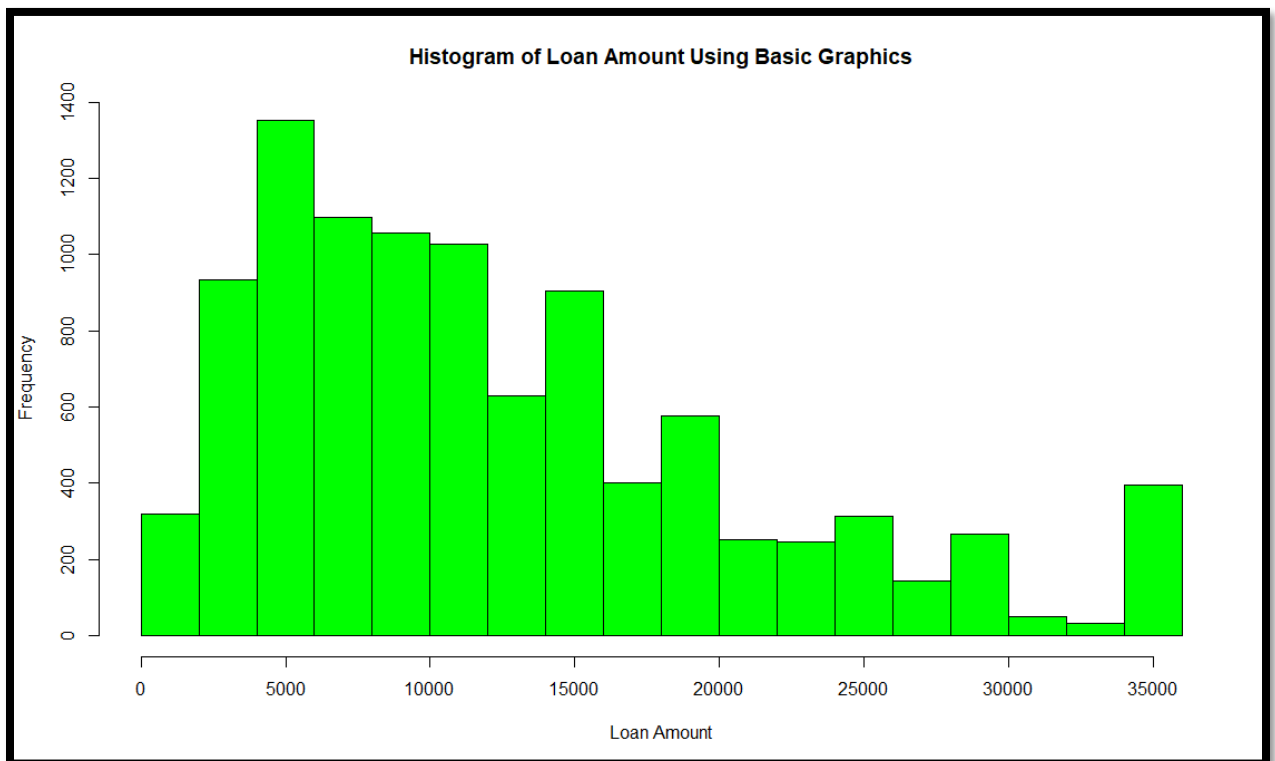
1. Read the dataset in loan.csv into R. Call the loaded data, loan. Make sure that you have the directory set to the correct location for the data.

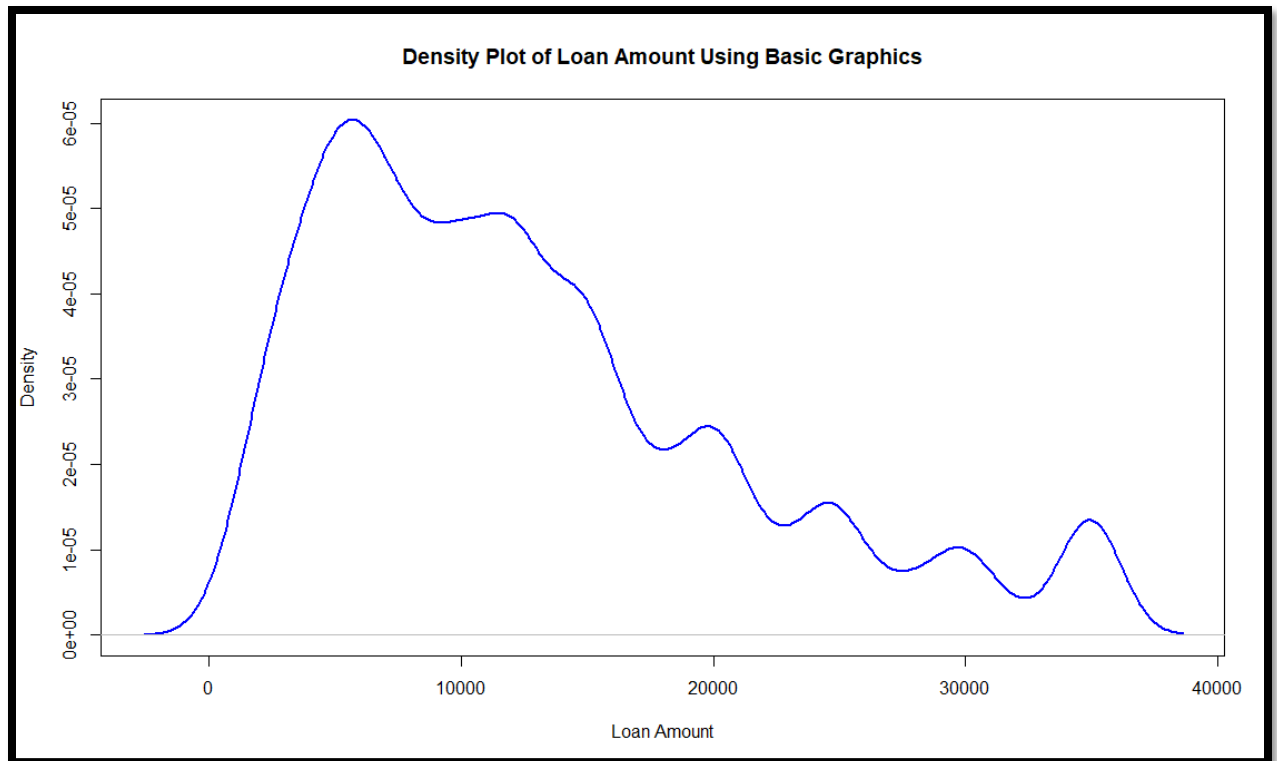
```
>
> ## 1. Read the dataset in loan.csv into R. Call the loaded data, loan.
> ### Make sure that you have the directory set to the correct location
> ### for the data.
>
> # Set the working directory to the correct location for the dataset.
> setwd("C:/PROJECTS/Maryville/DSCI 502/Week5")
>
> # Import necessary libraries
> library(ggplot2)
>
> # Load the data from loan.csv
> loan <- read.csv("loan.csv", stringsAsFactors = TRUE)
>
> # Display the dimensions (rows and columns) of the dataframe
> dim(loan) # Shows the number of rows and columns in the dataset.
[1] 10000 11
>
```

```
>
> ## 1. Read the dataset in loan.csv into R. Call the loaded data, loan.
> ### Make sure that you have the directory set to the correct location
> ### for the data.
>
> # Set the working directory to the correct location for the dataset.
> setwd("C:/PROJECTS/Maryville/DSCI 502/week5")
>
> # Import necessary libraries
> library(ggplot2)
>
> # Load the data from loan.csv
> loan <- read.csv("loan.csv", stringsAsFactors = TRUE)
>
> # Display the dimensions (rows and columns) of the dataframe
> dim(loan) # Shows the number of rows and columns in the dataset.
[1] 10000 11
> |
```

2. Please plot the histogram and density of the loan_amnt using basic graphics.

```
>  
> ## 2. Please plot the histogram and density of the loan_amnt using basic  
> ### graphics.  
>  
> # Histogram Plot using Basic Graphics  
> hist(loan$loan_amnt,  
+   main="Histogram of Loan Amount Using Basic Graphics",  
+   xlab="Loan Amount", col="green", border="black")  
>  
> # Density Plot using Basic Graphics  
> plot(density(loan$loan_amnt),  
+   main="Density Plot of Loan Amount Using Basic Graphics",  
+   xlab="Loan Amount", col="blue", lwd=2)  
>
```





3. Please plot the histogram and density of the loan_amnt and add the vertical line denoting the mean using ggplot2.

```
>
```

```
> ## 3. Please plot the histogram and density of the loan_amnt and add the
```

```
> #### vertical line denoting the mean using ggplot2.
```

```
>
```

```
> # Histogram Plot using ggplot2
```

```
> ggplot(loan, aes(x=loan_amnt)) +
```

```
+ geom_histogram(binwidth=2000, fill="green", color="black") +
```

```
+ geom_vline(aes(xintercept=mean(loan_amnt, na.rm=TRUE)),
```

```
+ color="red", linetype="dashed", size=1) +
```

```
+ ggtitle("Histogram of Loan Amount with Mean Using GGPlot2") +
```

```
+ xlab("Loan Amount") +
```

```
+ ylab("Count") +
```

```
+ theme_test()
```

```
>
```

```
> # Density Plot using ggplot2
```

```
> ggplot(loan, aes(x=loan_amnt)) +
```

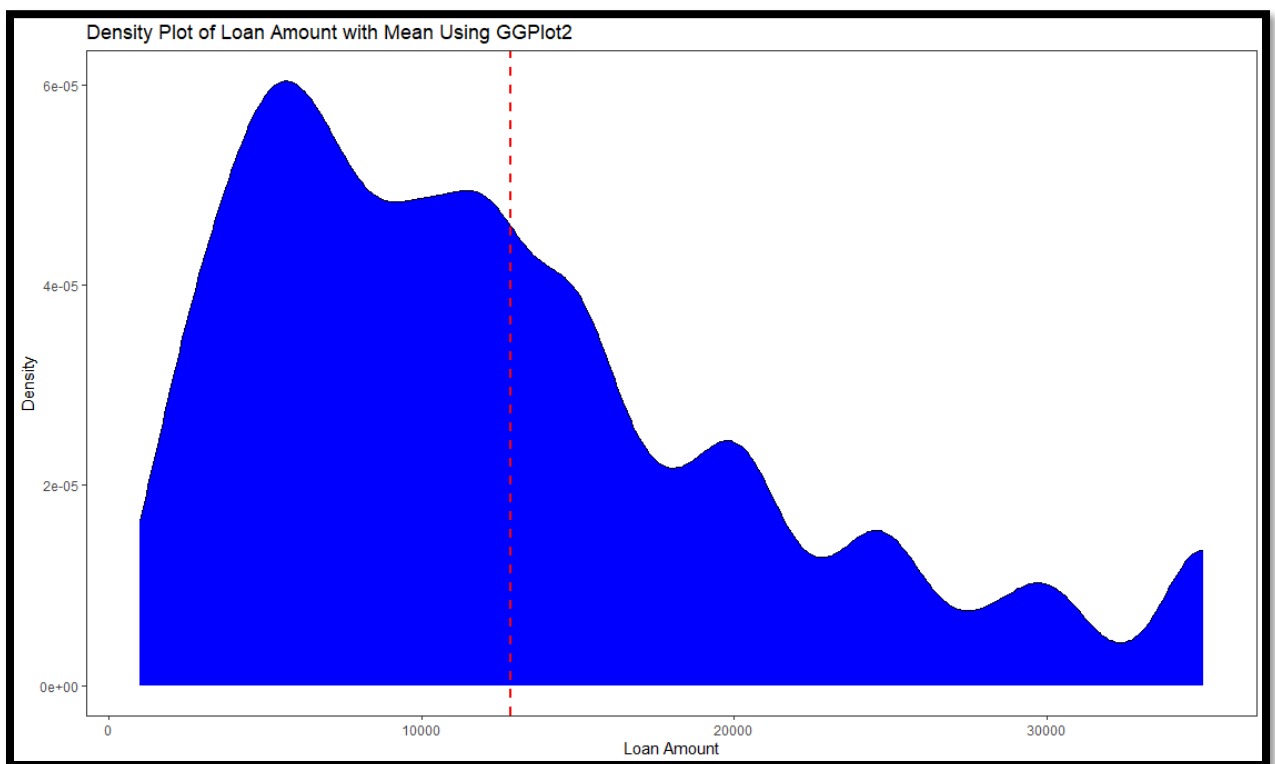
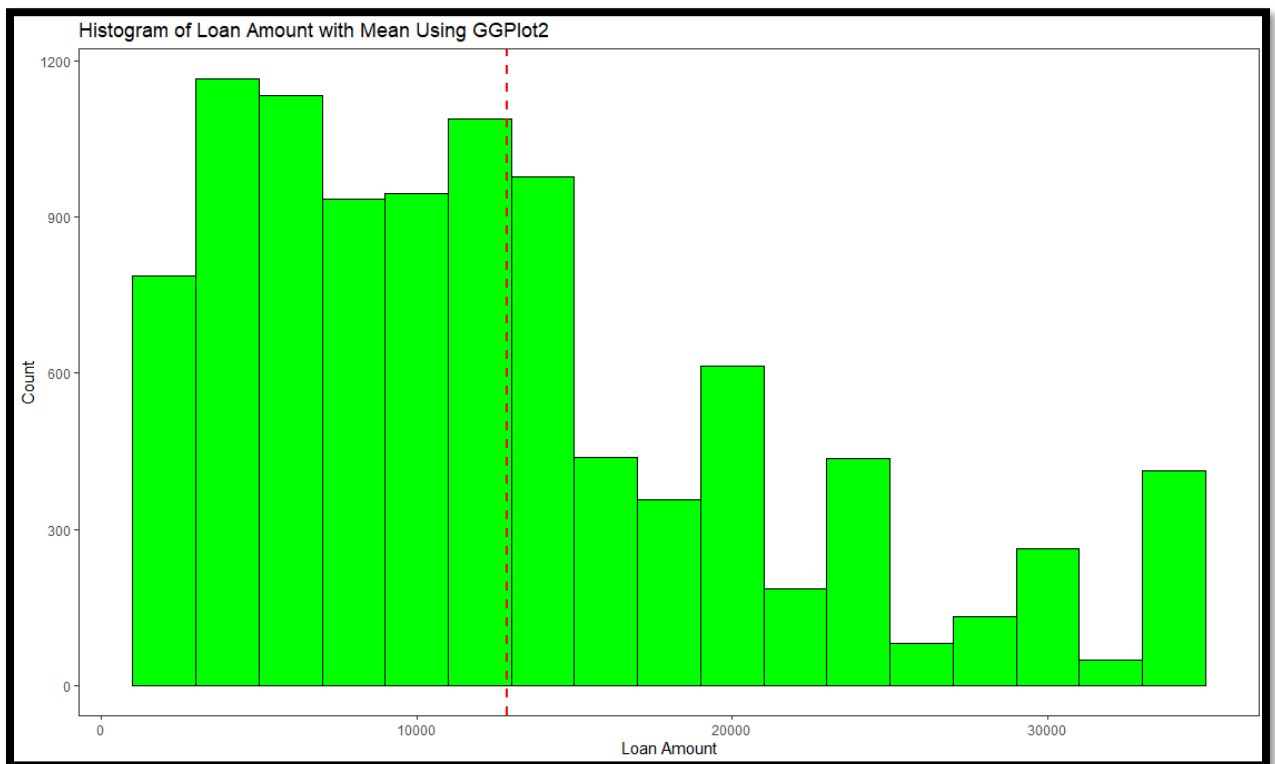
```
+ geom_density(fill="blue") +
```

```
+ geom_vline(aes(xintercept=mean(loan_amnt, na.rm=TRUE)),
```

```
+ color="red", linetype="dashed", size=1) +
```

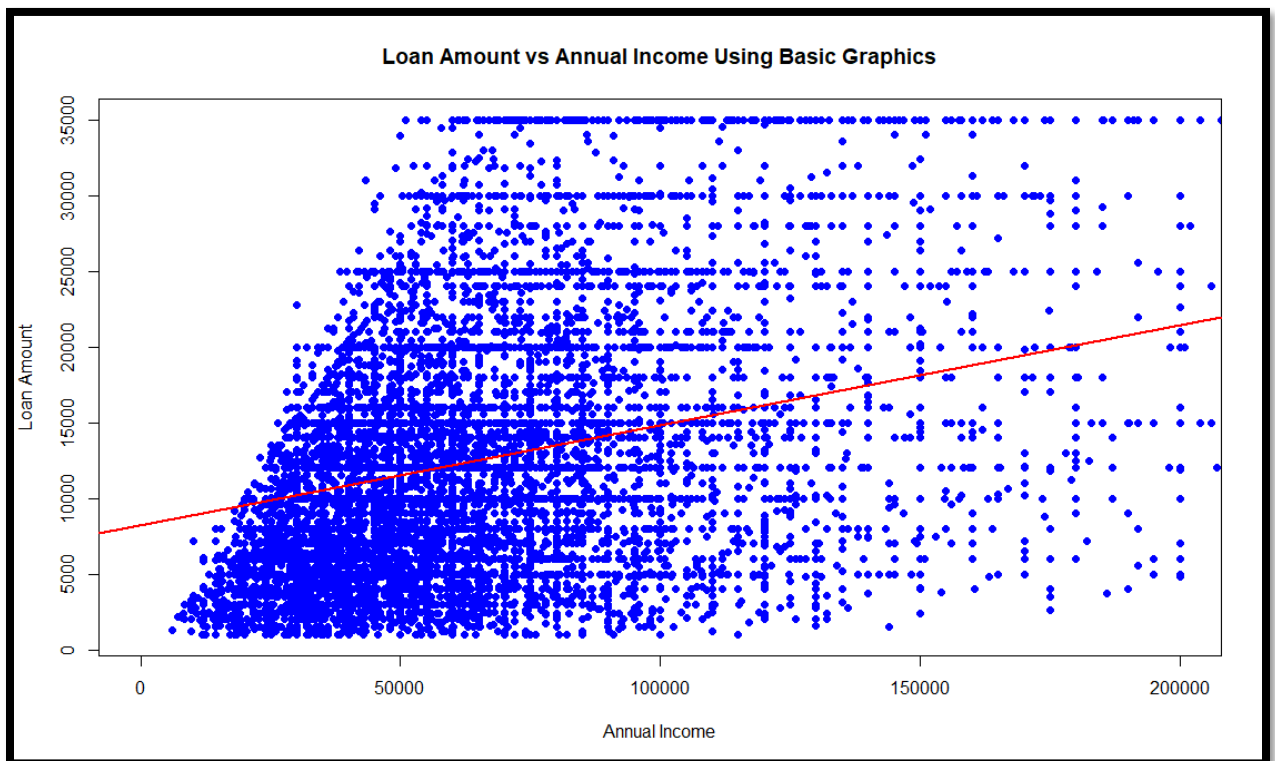
```
+ ggtitle("Density Plot of Loan Amount with Mean Using GGPlot2") +
```

```
+ xlab("Loan Amount") + ylab("Density") +  
+ theme_test()  
>
```



4. Please scatter plot of loan_amnt (y-axis) against annual_inc (x-axis) and add the trend line using basic graphics.

```
>
> ## 4. Please scatter plot of loan_amnt (y-axis) against annual_inc (x-axis)
> ### and add the trend line using basic graphics.
>
> # Scatter Plot using Basic Graphics
> plot(loan$annual_inc, loan$loan_amnt,
+   main="Loan Amount vs Annual Income Using Basic Graphics",
+   xlab="Annual Income",
+   ylab="Loan Amount",
+   pch=19, col="blue",
+   xlim = c(0, 200000))
> abline(lm(loan_amnt ~ annual_inc, data=loan), col="red", lwd=2)
>
```



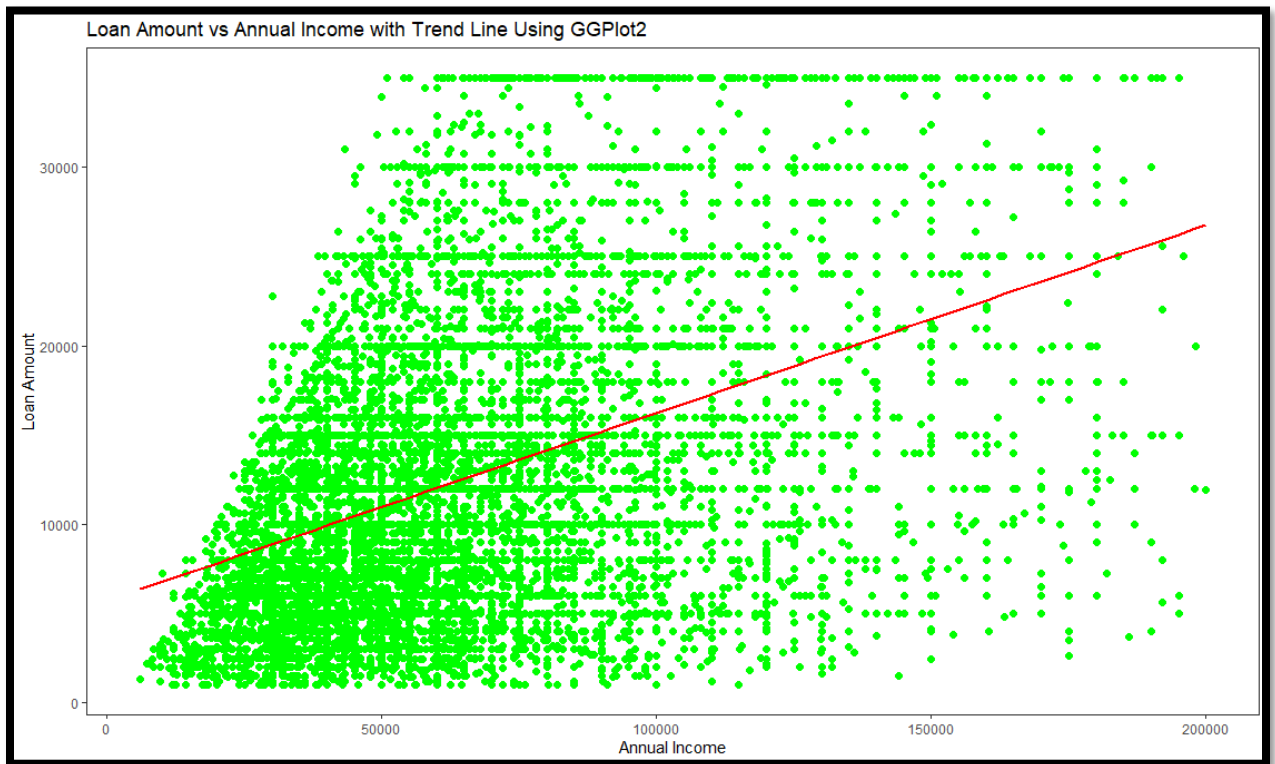
5. Please scatter plot of loan_amnt (y-axis) against annual_inc (x-axis) and add the trend line using ggplot2.

```
>
> ## 5. Please scatter plot of loan_amnt (y-axis) against annual_inc (x-axis)
> ### and add the trend line using ggplot2.
>
> # Scatter Plot using ggplot2
```

```

> ggplot(data = loan[which(loan$annual_inc < 200000),],
+   aes(x=annual_inc, y=loan_amnt)) +
+   geom_point(color="green", size = 2) +
+   geom_smooth(method="lm", color="red", se=FALSE, lwd = 1) +
+   ggtitle("Loan Amount vs Annual Income with Trend Line Using GGPlot2") +
+   xlab("Annual Income") + ylab("Loan Amount") + theme_test()
`geom_smooth()` using formula = 'y ~ x'
>

```

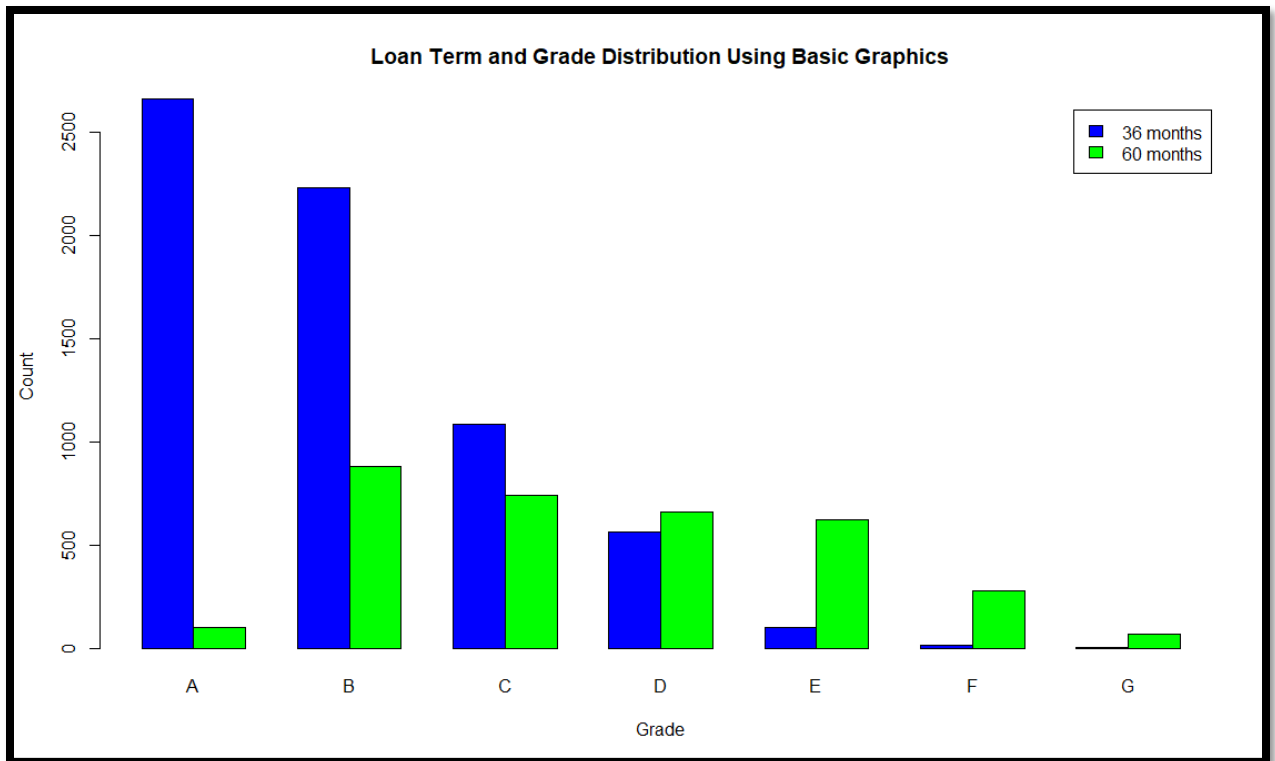


6. Please plot the barplot of term and grade on the same barplot using basic graphics

```

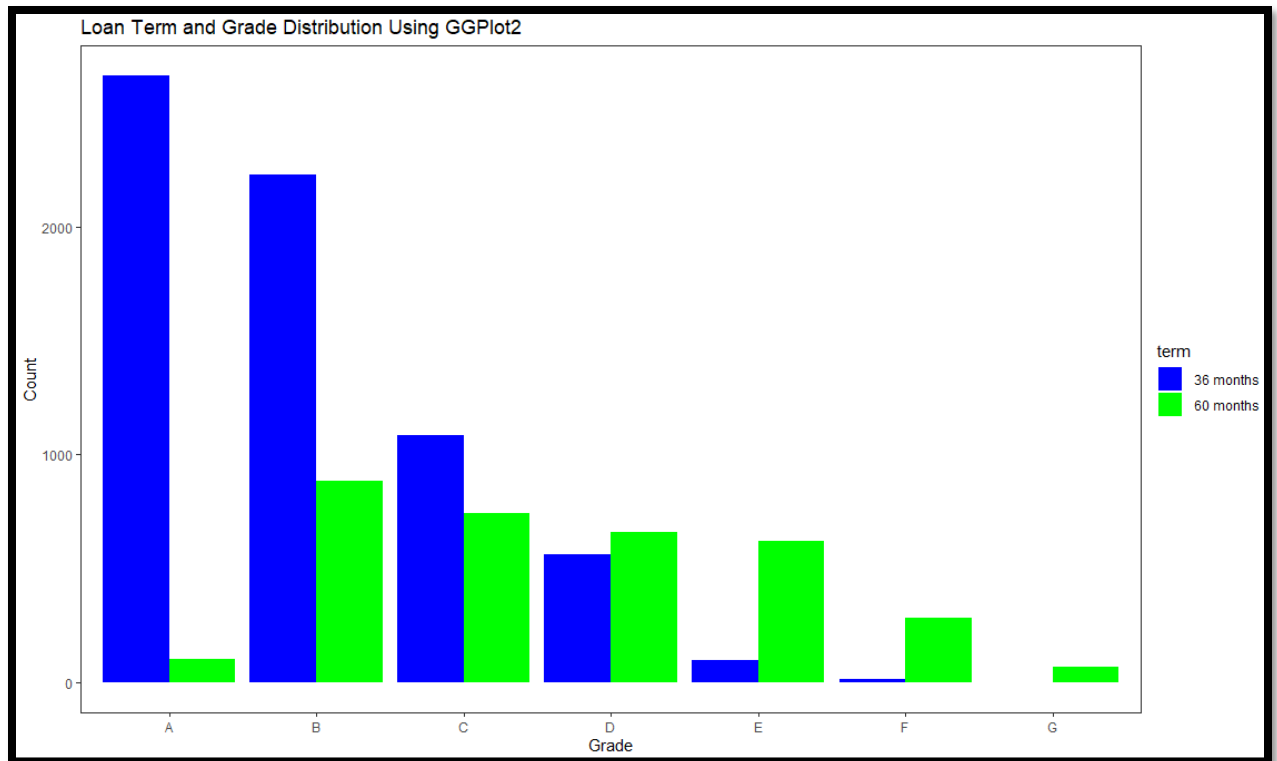
>
> ## 6. Please plot the barplot of term and grade on the same barplot using
> ### basic graphics
>
> # Barplot of Term and Grade using Basic Graphics
> barplot(table(loan$term, loan$grade), beside=TRUE,
+   legend=rownames(table(loan$term, loan$grade)),
+   col=c("blue", "green"),
+   main="Loan Term and Grade Distribution Using Basic Graphics",
+   xlab="Grade", ylab="Count")
>

```



7. Please plot the barplot of term and grade on the same barplot using ggplot2

```
>
> ## 7. Please plot the barplot of term and grade on the same barplot
> ### using ggplot2
>
> # Barplot of Term and Grade using ggplot2
> ggplot(loan, aes(x=grade, fill=term)) +
+   geom_bar(position="dodge") +
+   ggtitle("Loan Term and Grade Distribution Using GGPlot2") +
+   xlab("Grade") + ylab("Count") +
+   scale_fill_manual(values=c("blue", "green")) +
+   theme_test()
>
```



8. Please boxplot loan_amnt (y-axis) against term (x-axis) and save the graph in a file, loanterm.jpg, using basic graphics.

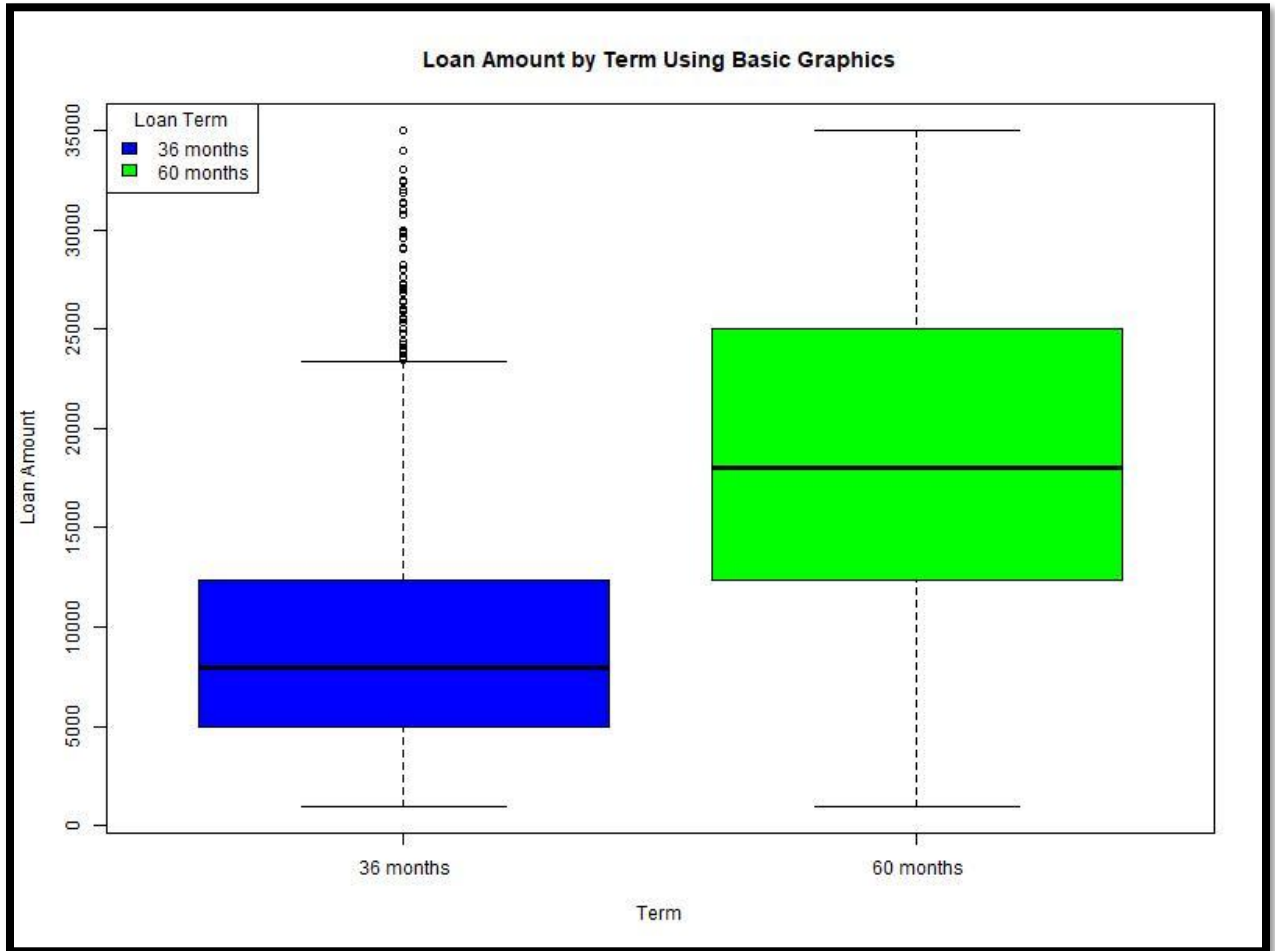
```
>
> ## 8. Please boxplot loan_amnt (y-axis) against term (x-axis)
> ### and save the graph in a file, loanterm.jpg, using basic graphics.
>
> # Open a graphics device to save plots as a JPEG image
> jpeg("loanterm.jpg", width=800, height=600)
>
> # Manually set fill colors for each level of 'term'
> box_colors <- c("blue", "green")
>
> # Boxplot of Loan Amount by Term using Basic Graphics
> boxplot(loan$loan_amnt ~ loan$term,
+   main="Loan Amount by Term Using Basic Graphics",
+   xlab="Term",
+   ylab="Loan Amount",
+   col=box_colors[as.factor(loan$term)], # Apply colors based on 'term'
+   border="black") # Set border color
>
> # Add the legend
> legend("topleft", legend=levels(as.factor(loan$term)),
```



```

+ fill=box_colors, title="Loan Term", border="black")
> dev.off() # Close the active graphical device
RStudioGD
2
>

```



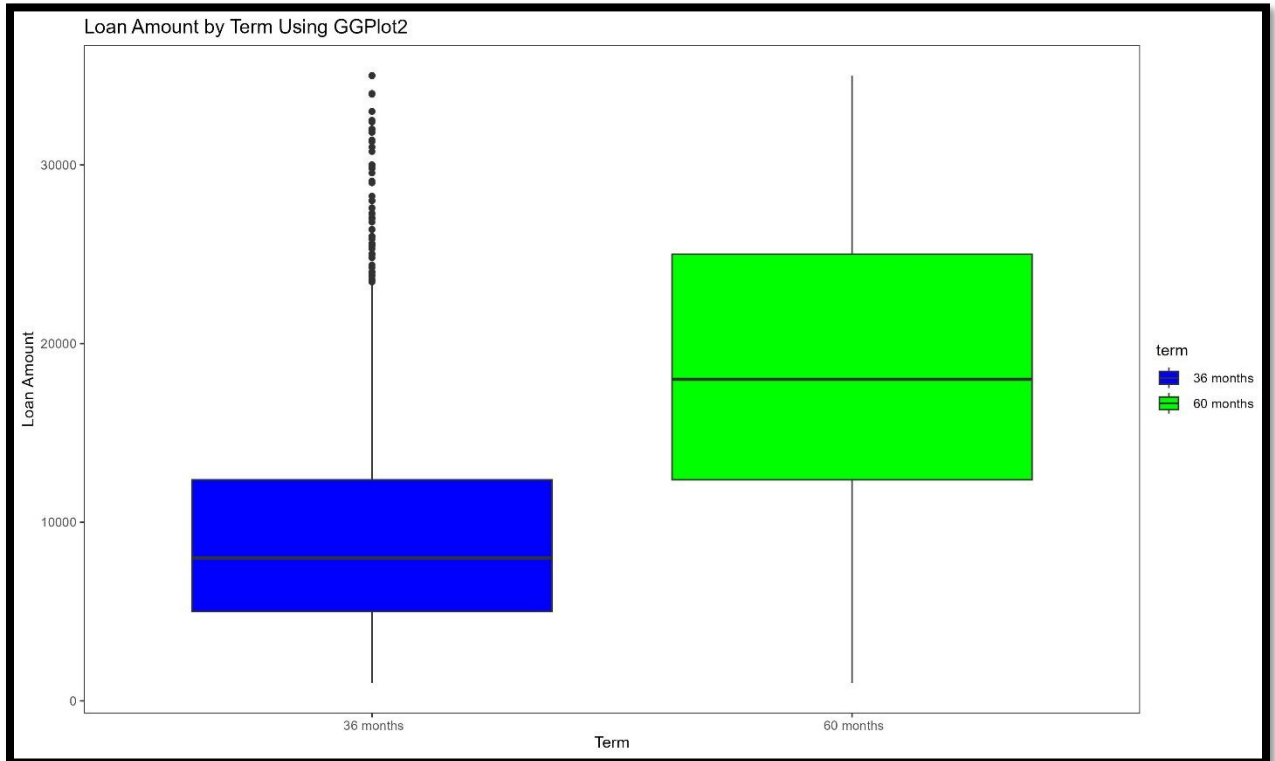
9. Please boxplot loan_amnt (y-axis) against term (x-axis) and save the graph in a file, loanterm.jpg, using ggplot2. Are there any differences in loan amount with respect to term?

```

>
> ## 9. Please boxplot loan_amnt (y-axis) against term (x-axis) and
> ### save the graph in a file, loanterm.jpg, using ggplot2.
> ### Are there any differences in loan amount with respect to term?
>
> # Boxplot of Loan Amount by Term using ggplot2
> ggplot(loan, aes(x=term, y=loan_amnt, fill = term)) +
+ geom_boxplot() +
+ scale_fill_manual(values=c("blue", "green")) +

```

```
+ ggtitle("Loan Amount by Term Using GGPlot2") +  
+ xlab("Term") + ylab("Loan Amount") + theme_test()  
>  
> # Save the plot generated by ggplot2 to a file  
> ggsave("loanterm.jpg")  
Saving 11.8 x 7.04 in image  
>
```



Yes, there are noticeable differences in loan amounts with respect to the term:

- 36-month term loans (blue boxplot) tend to have lower loan amounts. The median loan amount is lower compared to the 60-month term loans. There are also several outliers on the higher end, but the overall range is lower.
- 60-month term loans (green boxplot) generally have higher loan amounts. The median loan amount is significantly higher than the 36-month term loans. The interquartile range (middle 50% of the data) is also higher, indicating that longer-term loans are often associated with larger loan amounts.

This suggests that borrowers opting for a longer repayment term (60 months) tend to borrow larger amounts compared to those choosing a shorter term (36 months).