



VIT[®]

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

Lab Exercise

Code/Course	:	CSE3020 – Data Visualization	Date	:	24/08/2021
			Slot	:	L9+L10
Lab Experiment	:	Creating the various visual analytics using “plot”, “ggplot2 and ggvis” packages in R			

Pre-requisite: We will assume you are moderately familiar with basic concepts in R, including variables and functions, and with RStudio, the integrated development environment for programming in R.

In data analysis more than anything, a picture really is worth a thousand words. When you start analyzing data in R, your first step shouldn't be to run a complex statistical test: first, you should visualize your data in a graph. This lets you understand the basic nature of the data, so that you know what tests you can perform, and where you should focus your analysis. Introduce you to ggplot2, a powerful R package that produces data visualizations easily and intuitively. We will assume you are moderately familiar with basic concepts in R, including variables and functions, and with RStudio, the integrated development environment for programming in R. So, ggplot2 is a third party package: that means its code that doesn't come built into the language. This means you have to install it.

You can do that with one line of R code here in your interactive terminal, which is:

```
install.packages("ggplot2")
```

 and hit return. Or you can go to the Tools->Install Packages menu, where here you type "ggplot2" and hit install.

Each time you reopen R, you need to load the library using the library function before you use it.

So here that's:

```
library("ggplot2")
```

Now we're ready to use it. ggplot2 comes with some data available to use as a demonstration:

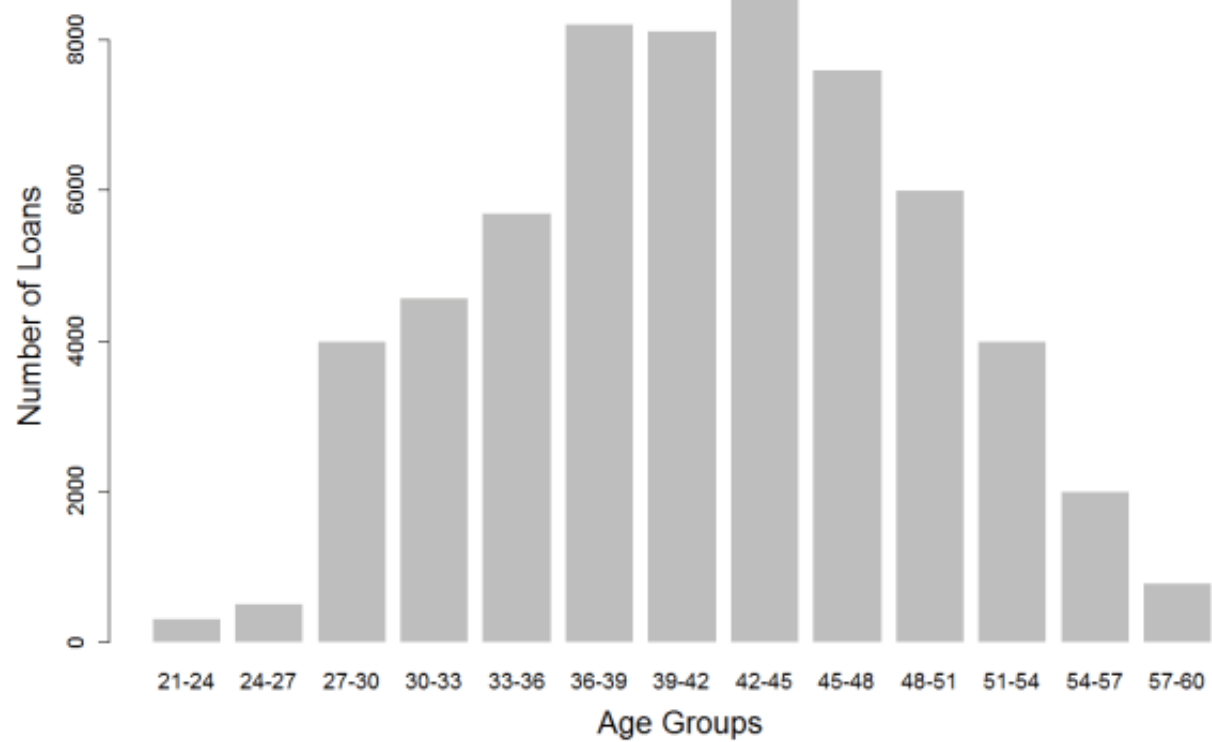
particularly, the "diamonds" dataset, containing information about several attributes of 54000 diamonds. We can access it using the data function:

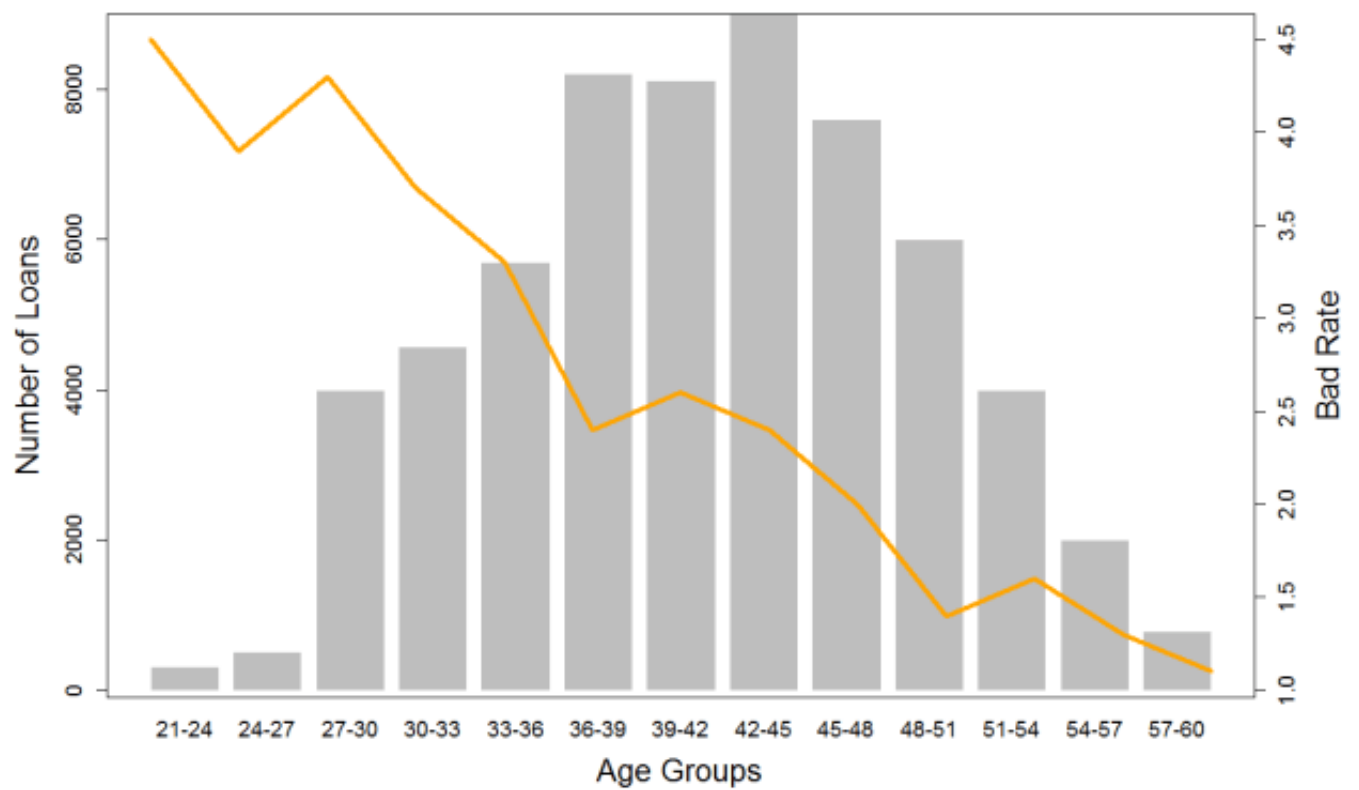
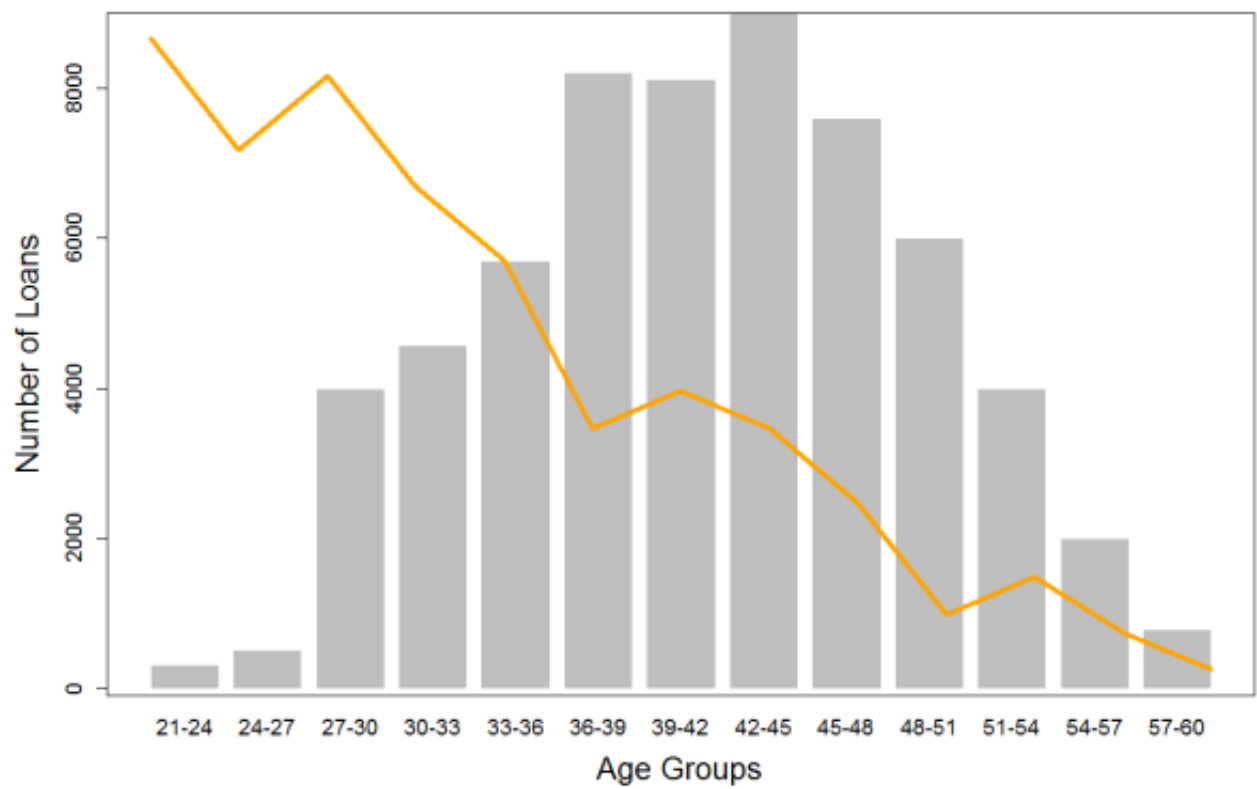
Practical Exercise:

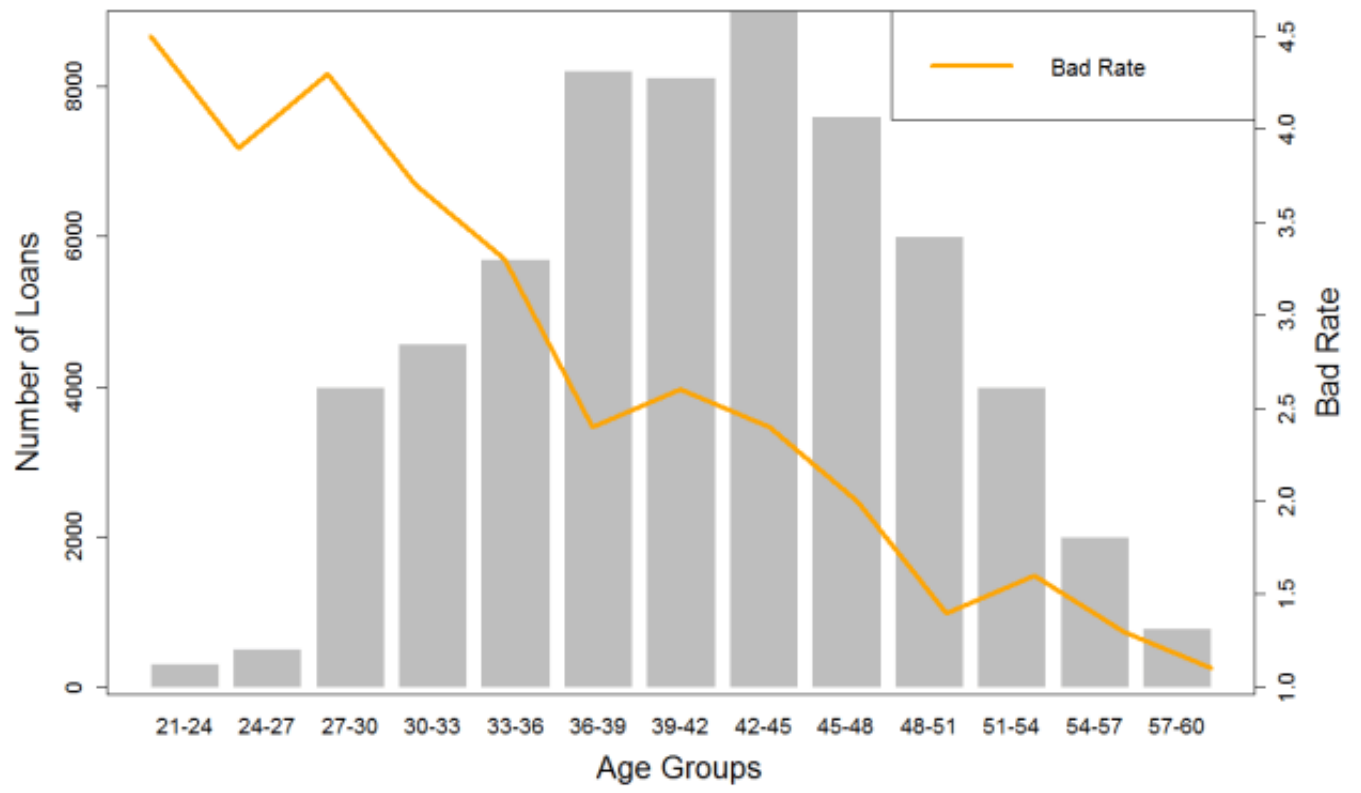
Consider the below sample Bank dataset

Age Group	Number of Loans	Bad Loans	Good Loans	Bad Rate
21-24	310	14	296	4.5
24-27	511	20	491	3.9
27-30	4000	172	3828	4.3
30-33	4568	169	4399	3.7
33-36	5698	188	5510	3.3
36-39	8209	197	8012	2.4
39-42	8117	211	7906	2.6
42-45	9000	216	8784	2.4
45-48	7600	152	7448	2
48-51	6000	84	5916	1.4
51-54	4000	64	3936	1.6
54-57	2000	26	1974	1.3
57-60	788	9	779	1.1

And Construct the following charts using “ggplot2 and ggvis” packages







Practical Exercise - II

Use the HR dataset to do the visual analytics with respect to the following points.

1. Plot the number of employees belong to male and female category
2. Plot the number of employees recruited through various recruitment sources
3. Plot the no. of employees under each marital categories
4. Plot the no. of employees under each marital categories with different colours
5. Plot the no. of employees under each marital categories in horizontal manner
6. Plot pie chart to represent employees belongs to various departments
7. Use histogram and density to plot employment status