

# project\_2.Rmd

Ken Wood

8/10/2024

In Part Two of the course project, you will create data visualizations of categorical and quantitative variables using the Credit data set. You'll use these visualizations to perform exploratory data analysis that will help you understand into which income brackets students and non-students fall. Your analyses will also help you understand the relationship between the credit limit and credit balance of students vs. non-students.

First, run the following code chunk to see the Credit data set:

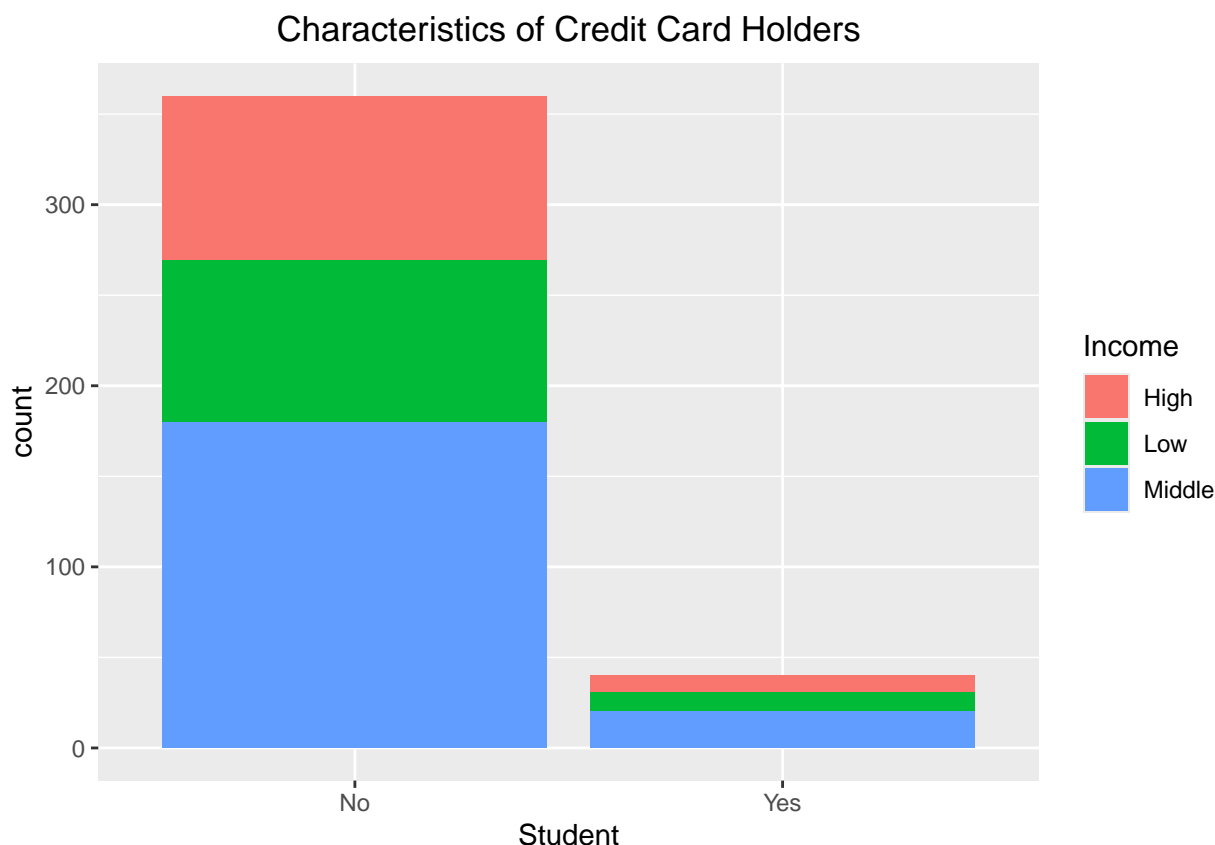
```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

##   ID Income Limit Rating Cards Age Education Student Married Balance
## 1  1   Low  3606   283    2  34         11      No    Yes    333
## 2  2  High  6645   483    3  82         15     Yes    Yes    903
## 3  3  High  7075   514    4  71         11     No    No    580
## 4  4  High  9504   681    3  36         11     No    No    964
## 5  5 Middle 4897   357    2  68         16     No    Yes    331
## 6  6  High  8047   569    4  77         10     No    No   1151
```

## Question 1

Create a stacked barplot titled “Characteristics of Credit Card Holders” that shows the number of credit card holders who are students and the number who are non-students, broken down by income level. In other words, your plot will have two bars — one for students and one for non-students — and each bar will consist of three different-colored layers corresponding to the three income levels (low, middle, and high). The y-axis label should be “Number of Card Holders.”

```
ggplot(Credit, aes(x = Student, fill = Income)) +
  geom_bar() +
  ggtitle("Characteristics of Credit Card Holders") +
  theme(plot.title = element_text(hjust=0.5))
```



### Question 2

Based on the plot you made in Question 1, in which category do most of the credit card holders fall (students or non-students)? For this category, do most of the credit card holders have a low, middle, or high income?

Upon examination of the plot created in Question 1, we see that most credit card holders are non-students. Most of the non-student credit card holders have a ‘middle’ income.

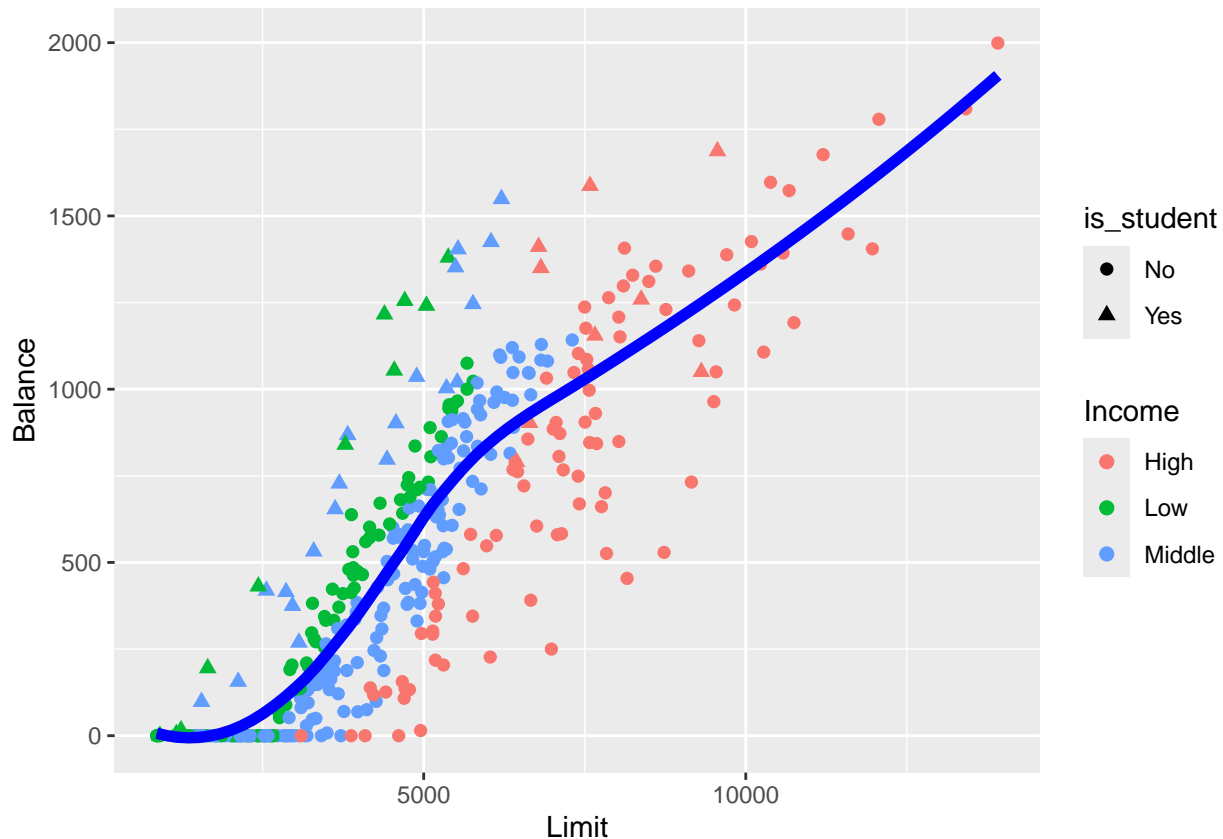
### Question 3

Create a scatterplot titled “Credit Card Balance vs. Limit” that displays the relationship between credit limit (on the x-axis) and credit card balance (on the y-axis). Set the color of the points to reflect the card holder’s income and the shape of the points to reflect whether that person is a student. To help you visualize the relationship between credit limit and card balance, add a single blue trendline to your plot.

```
is_student <- as.factor(Credit$Student)
ggplot(data = Credit, aes(x = Limit, y = Balance)) +
  geom_point(aes(color = Income, shape=is_student), size = 2) +
  geom_smooth(se = FALSE, size = 2, color = "blue")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```



#### Question 4

Based on the scatterplot you made in Question 3, write one or two sentences describing the income level and student status (yes or no) of individuals who have a credit limit above \$10,000. Do the same for individuals who have a credit limit below \$5,000.

We can see from the plot generated in Question 3 that there are no students who have a credit limit  $> \$10,000$ . The individuals who do have a credit limit  $> \$10,000$  are not student and have a high income. Conversely, the individuals who have a credit limit  $< \$5,000$  comprise both students and non-students, most of whom have a low to middle income level.

#### Question 5

We might expect that as a person's credit limit increases, their credit card balance is likely to increase as well. After all, as your limit increases, you're able to charge more and more to your credit card(s). Does the curve you drew in Question 3 indicate that credit card balance tends to increase as credit limit increases? Provide a brief explanation.

Yes the curve has a non-zero, positive slope which indicates that, as a person's credit limit increases, their credit card balance typically increases.