Lab 2 Solution

Joash Mochogi

```
library(openintro)
library(tidyverse)
```

Problem 1

There is a data frame called loan50 that is contained in the openintro package. load the data frame into your work space.

```
data("loan50")
```

Problem 2

Give a brief description of this data frame. How many variables does the data frame have? How many cases/observations does it have?

```
?loan50
# student may or may not run this command
```

The dataset represents a sample of loans made through a lending club platform that allows individuals to lend money to other individuals. The dataset has 50 observations and 18 variables.

Problem 3

Do people that have been employed for long tend to get lower interest rates? To answer this question, create a scatter plot for the variables emp_length and interest_rate.

```
g1 <- ggplot(data = loan50, aes(x = emp_length, y = interest_rate)) +
  geom_point()
print(g1)
```

The scatter plot shows that there is no clear relationship between the length of employment and the interest rate. People who have been employed for long do not necessarily get lower interest rates.

Problem 4

Recreate the following plot using the loan50 data frame.

```
ggplot(data = loan50, aes(x = emp_length, y = interest_rate, colour = homeownership)) +
  geom_point()
```

Problem 5

Create a bar plot (with differently colored bars) to visualize the distribution of the loans by homeownership. What insights can you draw from this bar plot?

```
ggplot(data = loan50, aes(x = homeownership, fill = homeownership)) +
  geom_bar()
```

The bar plot shows that most loans are held by people with a mortgage or who rent their home. People who own their homes have the least number of loans.

Problem 6

Recreate the following plot using the loan50 data frame.

```
ggplot(data = loan50, aes(x = homeownership, fill = loan_purpose)) +
  geom_bar(position = "dodge")
```

```
print(g1)
```