Week 7 Challenge

Mong Si Ok
Mong Si Ok

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

II. Code to edit and execute using the Code-along-6.Rmd file

```
# Installing the package and dependent
# dependencies
install.packages("ggplot2")

# or, alternatively, install the dependencies
# ggplot2 depends on, imports, and suggests
dependencies = TRUE, dep=T

# Load the ggplot2 package
library(ggplot2)

# Alternatively, load the tidyverse package
library(tidyverse)

# Check the version of ggplot2 package
packageVersion("ggplot2")

# Load the iris dataset
data(iris)
head(iris, n=10L)
# Alternatively
# iris <- read.csv(file, header=T)
# Check the structure of the iris dataset
str(iris)
# OR type(iris) to calculate the iris column data-frame for iris data,
# and then type(iris$Sepal.Length) to calculate the Sepal.Length column data-frame.

# Alternatively
# Load the iris dataset from the ggplot2 package
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
# Total loan count
# Total loan amount
ggplot(loans,
       aes(total_amount,
           loan_amount)) +
  geom_point() +
  facet_grid(home_ownership ~ purpose_cleaned)
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
# Total loan count
# Total loan amount
# Home ownership
ggplot(loans,
       aes(total_amount,
           loan_amount,
           color = home_ownership)) +
  geom_point() +
  facet_grid(home_ownership ~ purpose_cleaned)
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
# Total loan count
# Total loan amount
# Home ownership
# Loan purpose
ggplot(loans,
       aes(total_amount,
           loan_amount,
           color = home_ownership,
           shape = purpose_cleaned)) +
  geom_point()
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
# Load ggplot2 and dplyr packages
library(ggplot2)
library(dplyr)

# Loading required package: grDevices

# Loading required package: dplyr

# Attaching package: 'dplyr'

# The following objects are masked from 'package:stats':
#
#     filter, lag

# The following objects are masked from 'package:base':
#
#     intersect, setdiff, setequal, union

# Load the loans data
loans <- read_csv("loans_full_schema.csv")
# Rows: 10,000
# Columns: 40
# Chapter 2: Exploring categorical variables
# Table of counts for home_ownership
table(loans$home_ownership)
#
# mortgage     own    rent
#     4654     871    4484
# Chapter 2: Exploring categorical variables
# Table of counts for home_ownership, sorted
sort(table(loans$home_ownership))
#
#     own    rent
#     871    4484
# mortgage
#     4654
# Chapter 2: Exploring categorical variables
# Table of counts for home_ownership, sorted, stored
home_ownership_counts <- sort(table(loans$home_ownership))
# Chapter 2: Exploring categorical variables
# Bar chart of home_ownership
barplot(home_ownership_counts)
```

```
# Bar chart of home_ownership, with horizontal bars
barplot(home_ownership_counts, horiz = TRUE)
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels
barplot(home_ownership_counts, horiz = TRUE, las = 1)
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans")
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title, and a color
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans", col = "lightblue")
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title, and a color, and a legend
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans", col = "lightblue",
        legend.text = TRUE)
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title, and a color, and a legend, and a box around the legend
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans", col = "lightblue",
        legend.text = TRUE, bty = "o")
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title, and a color, and a legend, and a box around the legend, and a smaller legend
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans", col = "lightblue",
        legend.text = TRUE, bty = "o", cex.name = 0.7)
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title, and a color, and a legend, and a box around the legend, and a smaller legend, and a smaller legend
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans", col = "lightblue",
        legend.text = TRUE, bty = "o", cex.name = 0.7, cex.legend = 0.7)
```

```
# Bar chart of home_ownership, with horizontal bars, and rotated labels, and a title, and a color, and a legend, and a box around the legend, and a smaller legend, and a smaller legend, and a smaller legend
barplot(home_ownership_counts, horiz = TRUE, las = 1, main = "Amounts of Lending Club Loans", col = "lightblue",
        legend.text = TRUE, bty = "o", cex.name = 0.7, cex.legend = 0.7, cex.main = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership
pie(home_ownership_counts)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored
home_ownership_pie <- pie(home_ownership_counts)
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,")
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie, and a smaller pie, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7, cex = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7, cex = 0.7, cex.pie = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7, cex = 0.7, cex.pie = 0.7, cex.main = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7, cex = 0.7, cex.pie = 0.7, cex.main = 0.7, cex.pie = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7, cex = 0.7, cex.pie = 0.7, cex.main = 0.7, cex.pie = 0.7,
                          cex.pie = 0.7)
```

```
# Chapter 2: Exploring categorical variables
# Pie chart of home_ownership, stored, with a title, and a smaller title, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie, and a smaller pie
home_ownership_pie <- pie(home_ownership_counts, main = "Interest rate (%) at 3 NBCs,",
                          cex.main = 0.7, radius = 0.7, cex = 0.7, cex.pie = 0.7, cex.main = 0.7, cex.pie = 0.7,
                          cex.pie = 0.7, cex.pie = 0.7)
```



```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
# Hex plot
# Enter code here
ggplot(loans, aes(x = debt_to_income, y = interest_rate)) +
  geom_hex()
```

```
## Warning: Removed 24 rows containing non-finite values (`stat_binhex()`).
```

```
# Hex plot
# Enter code here
ggplot(loans %>% filter(debt_to_income < 100),
       aes(x = debt_to_income, y = interest_rate)) +
geom_hex()
```

```
# Bar plot
# Enter code here
ggplot(loans, aes(x = homeownership)) +
  geom_bar()
```

```
# Segmented Bar plot
# Enter code here
ggplot(loans, aes(x = homeownership,
                  fill = grade)) +
  geom_bar()
```

```
# Segmented Bar plot with equal heights for better comparison
# Enter code here
ggplot(loans, aes(x = homeownership, fill = grade)) +
  geom_bar(position = "fill")
```

```
# Customising bar plots
# Enter code here
ggplot(loans, aes(y = homeownership, fill = grade)) + geom_bar(position = "fill") +
  labs(x = "Proportion", y = "Homeownership", fill = "Grade", title = "Grades of Lending Club Loans")
```

```
# Violin plots
# Enter code here
ggplot(loans, aes(x = homeownership, y = loan_amount)) +
  geom_violin()
```

```
# Ridge plots
# Enter code here
#install.packages('ggridges')
library(ggridges)
library(ggplot2)
ggplot(loans, aes(x = loan_amount, y = grade, fill = grade, colour = grade)) +
  geom_density_ridges(alpha = 0.5)
```

```
## Picking joint bandwidth of 2360
```