R Resit Task

library(tidyverse)

library(ggplot2)

library(dplyr)

df <- read.csv("./DATA/loan\_data.csv")

average\_loan\_amnt <- mean(df$loan\_amnt, na.rm = TRUE)

average\_loan\_amnt

unique\_purpose\_count <- length(unique(df$purpose))

unique\_purpose\_count

df2 <- df %>% filter(home\_ownership == 'RENT' | purpose != 'educational')

ggplot(df, aes(x = annual\_inc)) +

geom\_histogram(binwidth = 5000, fill = 'skyblue', color = 'black' +

labs(title = "Distribution of Annual Income", x = "Annual Income", y = "Frequency") +

theme\_minimal()

df3 = df %>%

group\_by(home\_ownership) %>%

summarise(std\_dev\_int\_rate = sd(int\_rate, na.rm = TRUE)) %>%

arrange(std\_dev\_int\_rate)

df3

df$issue\_date = as.Date(df$issue\_date)

defaults\_over\_time = df %>%

mutate(issue\_month = floor\_date(issue\_date,"month")) %>%

group\_by(issue\_month) %>%

summarise(default\_count = sum(default = 1, na.rm = TRUE)) %>%

arrange(issue\_month)

ggplot(defaults\_over\_time, aes(x = issue\_month, y = default\_count))+

geom\_line(colour = "blue")+

labs(title = "Evolution of Defaults Over Time",x = "Issue Month", y = "Number of Defaults")+

theme\_minimal()