Lab 10 - Studying Correlated Variables

Ariane Ducellier

10/24/2023

In this lab, we are going to visualize two highly correlated variables from the loan data set.

Libraries

Load the necessary libraries.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## 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
```

Data

Load the data.

```
df <- read.csv("../data/LoanStats.csv")
df <- subset(df, grade=="A")
```

Plot

Complete the cell with the code to make the plot. We need to use highly correlated variables, try: - total_rec_prncp and total_pymnt_int - funded_amnt, total_pymnt_inv