# Importing the dataset

library(readr)  
data <- read\_csv("data/german\_data\_clean.csv")

## Rows: 1000 Columns: 21  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (13): checking\_account, credit\_history, purpose, savings, present\_employ...  
## dbl (8): duration\_months, credit\_amount, installment\_rate, present\_residenc...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

Show the entire data frame

data

## # A tibble: 1,000 × 21  
## checking\_account duration\_months credit\_history purpose credit\_amount savings  
## <chr> <dbl> <chr> <chr> <dbl> <chr>   
## 1 ... < 0 euro 6 critical acco… radio/… 1169 unknow…  
## 2 0 <= ... < 102… 48 existing cred… radio/… 5951 ... < …  
## 3 No checking acc… 12 critical acco… educat… 2096 ... < …  
## 4 ... < 0 euro 42 existing cred… furnit… 7882 ... < …  
## 5 ... < 0 euro 24 delay in payi… car (n… 4870 ... < …  
## 6 No checking acc… 36 existing cred… educat… 9055 unknow…  
## 7 No checking acc… 24 existing cred… furnit… 2835 255 eu…  
## 8 0 <= ... < 102… 36 existing cred… car (u… 6948 ... < …  
## 9 No checking acc… 12 existing cred… radio/… 3059 .. >= …  
## 10 0 <= ... < 102… 30 critical acco… car (n… 5234 ... < …  
## # ℹ 990 more rows  
## # ℹ 15 more variables: present\_employment\_since <chr>, installment\_rate <dbl>,  
## # personal\_status\_sex <chr>, other\_deptors <chr>, present\_residence <dbl>,  
## # property <chr>, age\_years <dbl>, other\_installment\_plans <chr>,  
## # housing <chr>, existing\_credits <dbl>, job <chr>,  
## # people\_liable\_maintenance <dbl>, telephone <chr>, foreign\_worker <chr>,  
## # response <dbl>

Selecting a column from a DataFrame

mean(data$age\_years)

## [1] 35.546

Overview of the DataFrame

str(data)

## spc\_tbl\_ [1,000 × 21] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ checking\_account : chr [1:1000] "... < 0 euro" "0 <= ... < 102 euro" "No checking account" "... < 0 euro" ...  
## $ duration\_months : num [1:1000] 6 48 12 42 24 36 24 36 12 30 ...  
## $ credit\_history : chr [1:1000] "critical account/other credits existing (not at this bank)" "existing credits paid back duly till now" "critical account/other credits existing (not at this bank)" "existing credits paid back duly till now" ...  
## $ purpose : chr [1:1000] "radio/television" "radio/television" "education" "furniture/equipment" ...  
## $ credit\_amount : num [1:1000] 1169 5951 2096 7882 4870 ...  
## $ savings : chr [1:1000] "unknown/ no savings account" "... < 51 euro" "... < 51 euro" "... < 51 euro" ...  
## $ present\_employment\_since : chr [1:1000] ".. >= 7 years" "1 <= ... < 4 years" "4 <= ... < 7 years" "4 <= ... < 7 years" ...  
## $ installment\_rate : num [1:1000] 4 2 2 2 3 2 3 2 2 4 ...  
## $ personal\_status\_sex : chr [1:1000] "male : single" "female : divorced/separated/married" "male : single" "male : single" ...  
## $ other\_deptors : chr [1:1000] "none" "none" "none" "guarantor" ...  
## $ present\_residence : num [1:1000] 4 2 3 4 4 4 4 2 4 2 ...  
## $ property : chr [1:1000] "real estate" "real estate" "real estate" "(if not real estate) building society savings agreement/ life insurance" ...  
## $ age\_years : num [1:1000] 67 22 49 45 53 35 53 35 61 28 ...  
## $ other\_installment\_plans : chr [1:1000] "none" "none" "none" "none" ...  
## $ housing : chr [1:1000] "own" "own" "own" "for free" ...  
## $ existing\_credits : num [1:1000] 2 1 1 1 2 1 1 1 1 2 ...  
## $ job : chr [1:1000] "skilled employee / official" "skilled employee / official" "unskilled - resident" "skilled employee / official" ...  
## $ people\_liable\_maintenance: num [1:1000] 1 1 2 2 2 2 1 1 1 1 ...  
## $ telephone : chr [1:1000] "yes, registered under the customers name" "none" "none" "none" ...  
## $ foreign\_worker : chr [1:1000] "yes" "yes" "yes" "yes" ...  
## $ response : num [1:1000] 0 1 0 0 1 0 0 0 0 1 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. checking\_account = col\_character(),  
## .. duration\_months = col\_double(),  
## .. credit\_history = col\_character(),  
## .. purpose = col\_character(),  
## .. credit\_amount = col\_double(),  
## .. savings = col\_character(),  
## .. present\_employment\_since = col\_character(),  
## .. installment\_rate = col\_double(),  
## .. personal\_status\_sex = col\_character(),  
## .. other\_deptors = col\_character(),  
## .. present\_residence = col\_double(),  
## .. property = col\_character(),  
## .. age\_years = col\_double(),  
## .. other\_installment\_plans = col\_character(),  
## .. housing = col\_character(),  
## .. existing\_credits = col\_double(),  
## .. job = col\_character(),  
## .. people\_liable\_maintenance = col\_double(),  
## .. telephone = col\_character(),  
## .. foreign\_worker = col\_character(),  
## .. response = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

Statistical overview

summary(data)

## checking\_account duration\_months credit\_history purpose   
## Length:1000 Min. : 4.0 Length:1000 Length:1000   
## Class :character 1st Qu.:12.0 Class :character Class :character   
## Mode :character Median :18.0 Mode :character Mode :character   
## Mean :20.9   
## 3rd Qu.:24.0   
## Max. :72.0   
## credit\_amount savings present\_employment\_since installment\_rate  
## Min. : 250 Length:1000 Length:1000 Min. :1.000   
## 1st Qu.: 1366 Class :character Class :character 1st Qu.:2.000   
## Median : 2320 Mode :character Mode :character Median :3.000   
## Mean : 3271 Mean :2.973   
## 3rd Qu.: 3972 3rd Qu.:4.000   
## Max. :18424 Max. :4.000   
## personal\_status\_sex other\_deptors present\_residence property   
## Length:1000 Length:1000 Min. :1.000 Length:1000   
## Class :character Class :character 1st Qu.:2.000 Class :character   
## Mode :character Mode :character Median :3.000 Mode :character   
## Mean :2.845   
## 3rd Qu.:4.000   
## Max. :4.000   
## age\_years other\_installment\_plans housing existing\_credits  
## Min. :19.00 Length:1000 Length:1000 Min. :1.000   
## 1st Qu.:27.00 Class :character Class :character 1st Qu.:1.000   
## Median :33.00 Mode :character Mode :character Median :1.000   
## Mean :35.55 Mean :1.407   
## 3rd Qu.:42.00 3rd Qu.:2.000   
## Max. :75.00 Max. :4.000   
## job people\_liable\_maintenance telephone   
## Length:1000 Min. :1.000 Length:1000   
## Class :character 1st Qu.:1.000 Class :character   
## Mode :character Median :1.000 Mode :character   
## Mean :1.155   
## 3rd Qu.:1.000   
## Max. :2.000   
## foreign\_worker response   
## Length:1000 Min. :0.0   
## Class :character 1st Qu.:0.0   
## Mode :character Median :0.0   
## Mean :0.3   
## 3rd Qu.:1.0   
## Max. :1.0

summary(data$credit\_amount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 250 1366 2320 3271 3972 18424