Project2

Minh Tran

2024-10-03

```
#check for missing values
loans%>%
  count(across(everything(),~sum(is.na(.))))
```

```
##   employ_length annual_income debt_to_income loan_purpose loan_amount
## 1           410             0              0            0           0
##   interest_rate    n
## 1             0 5894
```

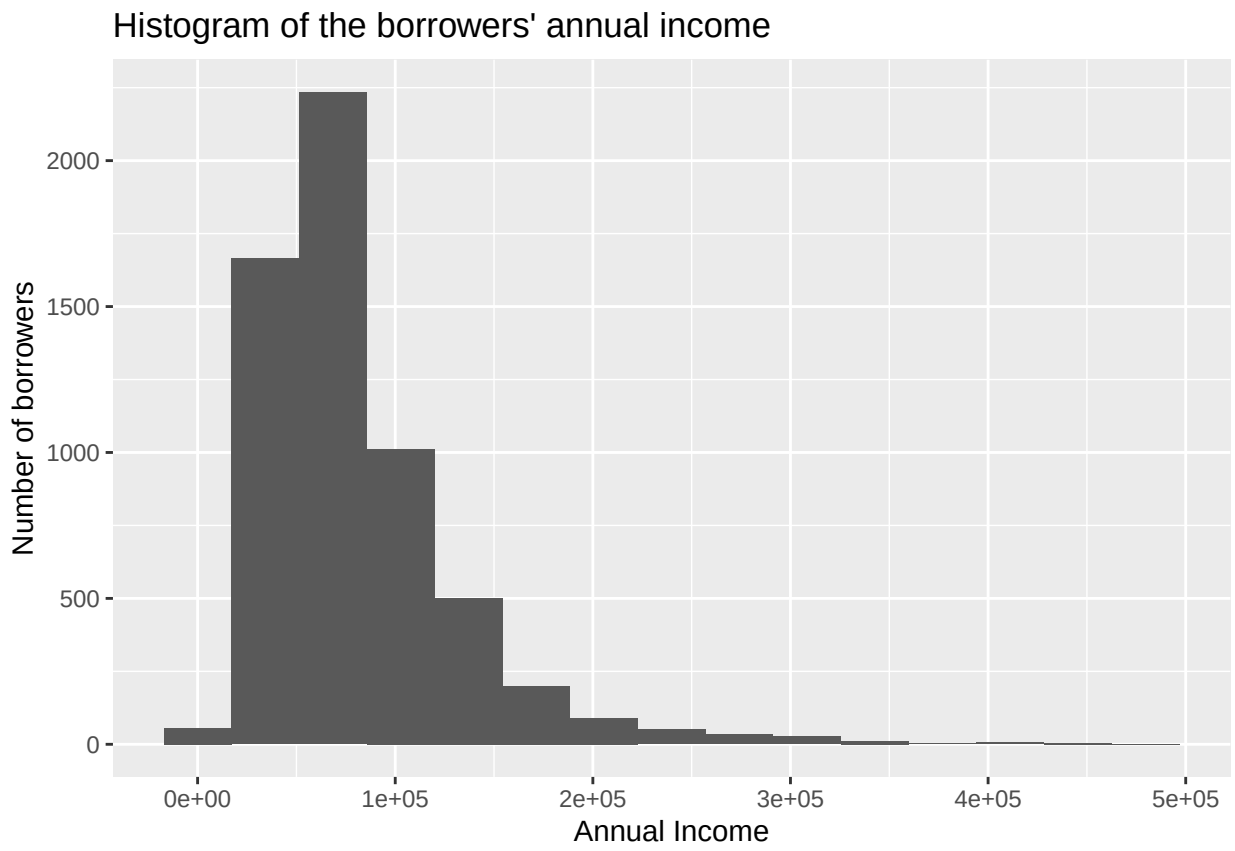
```
#frequency table for employment length
loans%>%
  tabyl(employ_length)%>%
  adorn_pct_formatting(digit=2)
```

```
##      employ_length    n percent valid_percent
##    2 or more years 4676  79.33%        85.27%
##  Less than 2 years  808  13.71%        14.73%
##               <NA>  410   6.96%            -
```

```
#summary statistics and histogram for annual income
favstats(loans$annual_income)
```

```
##   min    Q1 median    Q3    max     mean       sd    n missing
##  5235 49350  69000 98932 485000 80690.54 50524.35 5894       0
```

```
loans%>%
  ggplot(aes(x=annual_income))+
  geom_histogram(bins=15)+
  ggtitle("Histogram of the borrowers' annual income")+
  xlab("Annual Income")+
  ylab("Number of borrowers")
```

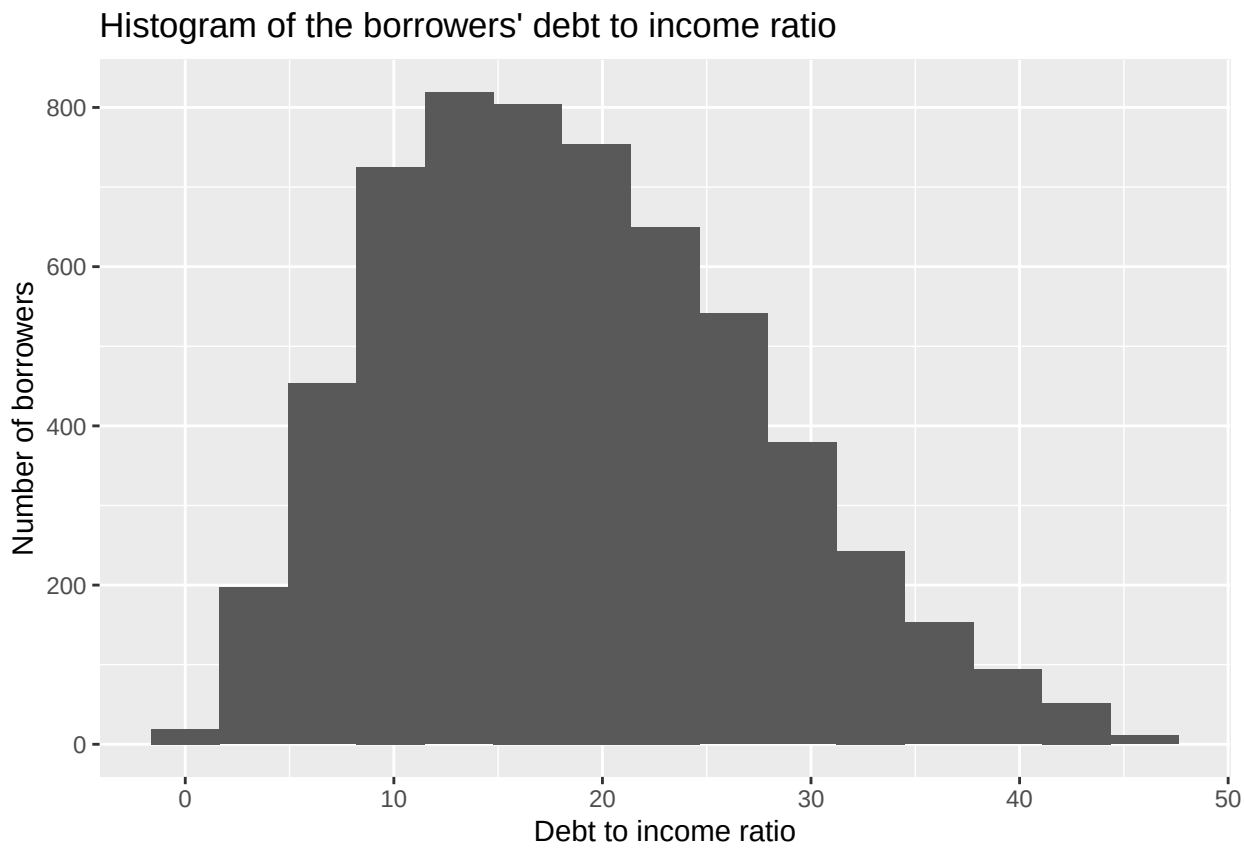



```
#IQR = Q3 - Q1 = 98932 - 49350 = 49,582
```

```
#checking summary statistics for debt to income ratio
favstats(loans$debt_to_income)
```

```
##  min    Q1 median      Q3 max     mean       sd    n missing
##    0 11.83  17.79 24.6375  46 18.61585 8.827795 5894       0
```

```
loans%>%
  ggplot(aes(x=debt_to_income))+
  geom_histogram(bins=15)+
  ggtitle("Histogram of the borrowers' debt to income ratio")+
  xlab("Debt to income ratio")+
  ylab("Number of borrowers")
```




```
#IQR = 24.6375 - 11.83 = 12.8075
```

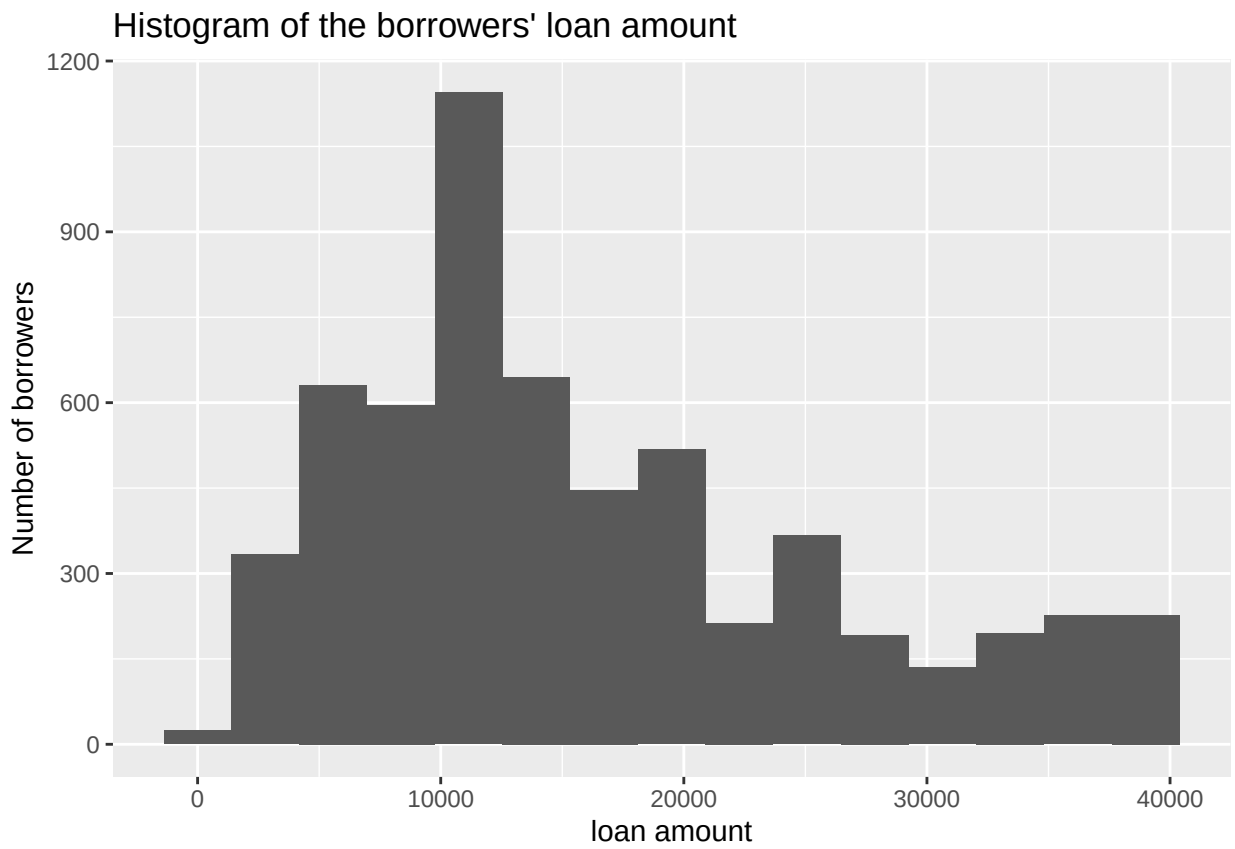
```
#frequency table for purpose of loan
loans%>%
  tabyl(loan_purpose)%>%
  adorn_pct_formatting(digit=2)
```

```
##        loan_purpose    n percent
##         credit_card 1705  28.93%
##  debt_consolidation 3719  63.10%
##    home_improvement  470   7.97%
```

```
#summary stats and histogram for loan amount
favstats(loans$loan_amount)
```

```
##   min   Q1 median    Q3   max  mean       sd    n missing
##  1000 9000  14000 22000 40000 16184 9886.623 5894       0
```

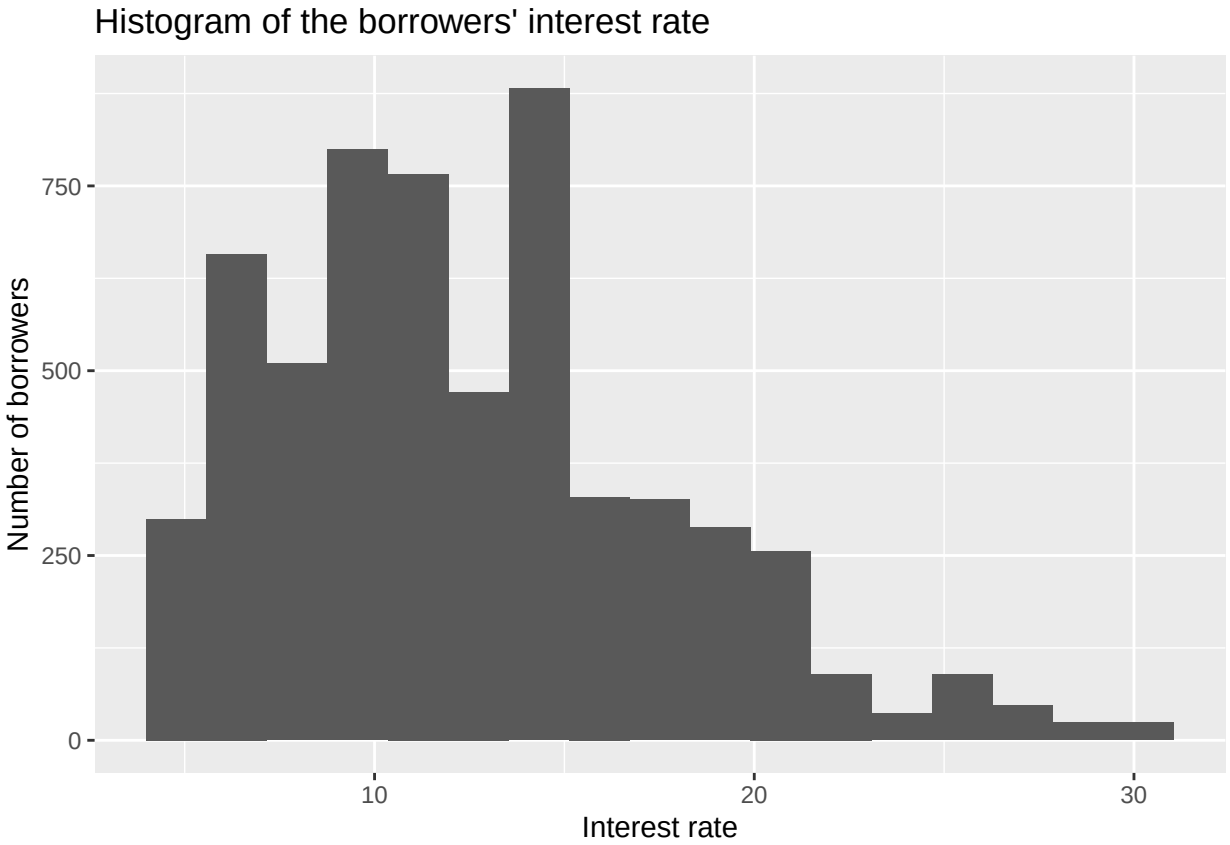
```
loans%>%
  ggplot(aes(x=loan_amount))+
  geom_histogram(bins=15)+
  ggtitle("Histogram of the borrowers' loan amount")+
  xlab("loan amount")+
  ylab("Number of borrowers")
```




```
favstats(loans$interest_rate)
```

```
##   min   Q1 median    Q3   max     mean       sd    n missing
##  5.31 9.43  10.91 16.01 30.79 12.45205 5.162209 5894       0
```

```
loans%>%
  ggplot(aes(x=interest_rate))+
  geom_histogram(bins=17)+
  ggtitle("Histogram of the borrowers' interest rate")+
  xlab("Interest rate")+
  ylab("Number of borrowers")
```


Histogram of the borrowers' interest rate
Number of borrowers
750
500
250
0
10
20
30
Interest rate