Modelling loan interest rates

In this application exercise we will be studying loan interest rates. The dataset is one you've come across before in your reading – the dataset about loans from the peer-to-peer lender, Lending Club, from the **openintro** package. We will use **tidyverse** and **tidymodels** for data exploration and modeling, respectively.

```
library(tidyverse)
library(tidymodels)
library(openintro)
```

Before we use the dataset, we'll make a few transformations to it.

- **Your turn:** Review the code below with your neighbor and write a summary of the data transformation pipeline.

Add response here.

```
loans <- loans_full_schema |>
  mutate(
    credit_util = total_credit_utilized / total_credit_limit,
    bankruptcy = as.factor(if_else(public_record_bankrupt == 0, 0, 1)),
    verified_income = droplevels(verified_income),
    homeownership = str_to_title(homeownership),
    homeownership = fct_relevel(homeownership, "Rent", "Mortgage", "Own")
  ) |>
  rename(credit_checks = inquiries_last_12m) |>
  select(
    interest_rate, loan_amount, verified_income,
    debt_to_income, credit_util, bankruptcy, term,
    credit_checks, issue_month, homeownership
  )
```

Here is a glimpse at the data:

```
glimpse(loans)
```

```
Rows: 10,000
Columns: 10
$ interest_rate   <dbl> 14.07, 12.61, 17.09, 6.72, 14.07, 6.72, 13.59, 11.99, ~
$ loan_amount     <int> 28000, 5000, 2000, 21600, 23000, 5000, 24000, 20000, 2~
$ verified_income <fct> Verified, Not Verified, Source Verified, Not Verified,~
$ debt_to_income  <dbl> 18.01, 5.04, 21.15, 10.16, 57.96, 6.46, 23.66, 16.19, ~
$ credit_util     <dbl> 0.54759517, 0.15003472, 0.66134832, 0.19673228, 0.7549~
$ bankruptcy      <fct> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, ~
$ term            <dbl> 60, 36, 36, 36, 36, 36, 60, 60, 36, 36, 60, 60, 36, 60~
$ credit_checks   <int> 6, 1, 4, 0, 7, 6, 1, 1, 3, 0, 4, 4, 8, 6, 0, 0, 4, 6, ~
$ issue_month     <fct> Mar-2018, Feb-2018, Feb-2018, Jan-2018, Mar-2018, Jan-~
$ homeownership   <fct> Mortgage, Rent, Rent, Rent, Rent, Own, Mortgage, Mortg~
```

Get to know the data

- **Your turn:** What is a typical interest rate in this dataset? What are some attributes of a typical loan and a typical borrower. Give yourself no more than 5 minutes for this exploration and share 1-2 findings.

```
# add code to explore loans here
```

```
# add code to explore borrowers here
```

Interest rate vs. credit utilization ratio

The regression model for interest rate vs. credit utilization is as follows.

```
rate_util_fit <- linear_reg() |>
  fit(interest_rate ~ credit_util, data = loans)

tidy(rate_util_fit)
```
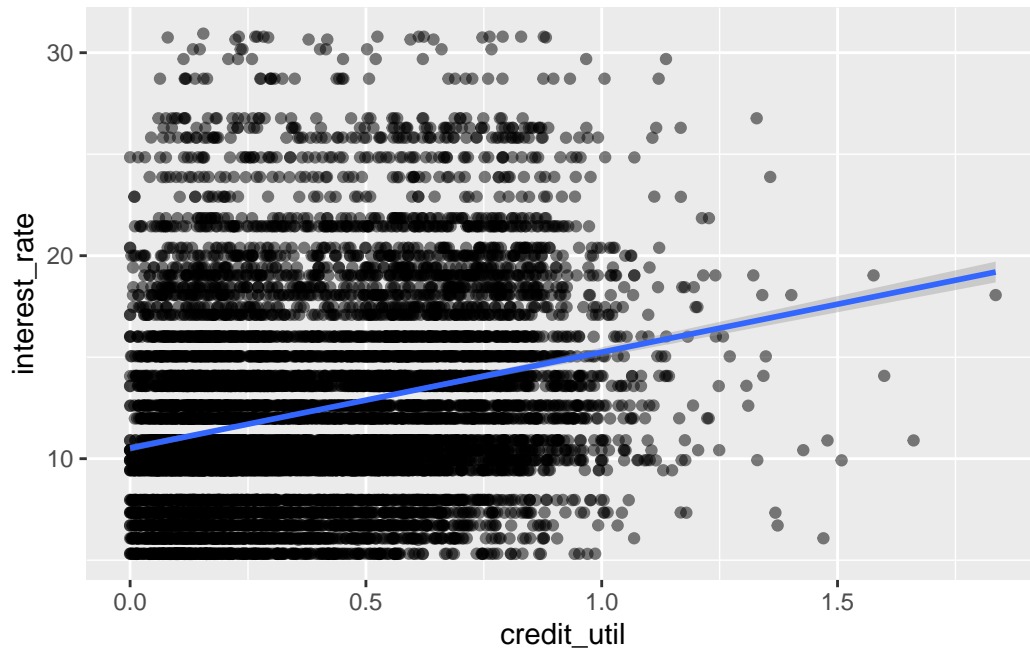
```
# A tibble: 2 x 5
  term        estimate std.error statistic   p.value
  <chr>          <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)    10.5     0.0871     121.  0
2 credit_util     4.73    0.180       26.3 1.18e-147
```

And here is the model visualized:

```
ggplot(loans, aes(x = credit_util, y = interest_rate)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm")
```


- **Your turn:** Interpret the intercept and the slope.

Add response here.

Interest rate vs. homeownership

Next we predict interest rates from homeownership, which is a categorical predictor with three levels:

```
levels(loans$homeownership)
```

```
[1] "Rent"     "Mortgage" "Own"
```

- **Demo:** Fit the linear regression model to predict interest rate from homeownership and display a tidy summary of the model. Write the estimated model output below.

```
# add code here.
```

- **Your turn:** Interpret each coefficient in context of the problem.

Add response here.

Interest rate vs. credit utilization and homeownership

Main effects model

- **Demo:** Fit a model to predict interest rate from credit utilization and homeownership, **without** an interaction effect between the two predictors. Display the summary output and write out the estimated regression equation.

```
# add code here
```

Add response here.

- **Demo:** Write the estimated regression equation for loan applications from each of the homeownership groups separately.

  - Rent: *Add response here.*
  - Mortgage: Add response here.
  - Own: Add response here.

- **Question:** How does the model predict the interest rate to vary as credit utilization varies for loan applicants with different homeownership status. Are the rates the same or different?

Add response here.

Interaction effects model

- **Demo:** Fit a model to predict interest rate from credit utilization and homeownership, **with** an interaction effect between the two predictors. Display the summary output and write out the estimated regression equation.

```
# add code here
```

Add response here.

- **Demo:** Write the estimated regression equation for loan applications from each of the homeownership groups separately.

    - Rent: *Add response here.*
    - Mortgage: Add response here.
    - Own: Add response here.

- **Question:** How does the model predict the interest rate to vary as credit utilization varies for loan applicants with different homeownership status. Are the rates the same or different?

Add response here.

Choosing a model

Rule of thumb: **Occam's Razor** - Don't overcomplicate the situation! We prefer the *simplest* best model.

```
# add code here
```

- **Review:** What is R-squared? What is adjusted R-squared?

Add response here.

- **Question:** Based on the adjusted R^2s of these two models, which one do we prefer?

Add response here.

Another model to consider

- **Your turn:** Let's add one more model to the variable – issue month. Should we add this variable to the interaction effects model from earlier?

```
# add code here
```

Add response here.