Untitled

```r
library(dplyr)
library(ggplot2)
library(forcats)
library(janitor)
library(haven)
library(kableExtra)
```

```r
data(loan50, package = "openintro")
```

Pregunta 1

¿Cuántas observaciones tiene `loan50`? ¿Cuántas variables son cuantitativas? ¿Cuántas variables son categóricas nominales y ordinales?

```r
glimpse(loan50)
```

```
## Rows: 50
## Columns: 18
## $ state                  <fct> NJ, CA, SC, CA, OH, IN, NY, MO, FL, FL, MD,...
## $ emp_length             <dbl> 3, 10, NA, 0, 4, 6, 2, 10, 6, 3, 8, 10, 10,...
## $ term                   <dbl> 60, 36, 36, 36, 60, 36, 36, 36, 60, 60, 36,...
## $ homeownership          <fct> rent, rent, mortgage, rent, mortgage, mortg...
## $ annual_income          <dbl> 59000, 60000, 75000, 75000, 254000, 67000, ...
## $ verified_income        <fct> Not Verified, Not Verified, Verified, Not V...
## $ debt_to_income         <dbl> 0.55752542, 1.30568333, 1.05628000, 0.57434...
## $ total_credit_limit     <int> 95131, 51929, 301373, 59890, 422619, 349825...
## $ total_credit_utilized  <int> 32894, 78341, 79221, 43076, 60490, 72162, 2...
## $ num_cc_carrying_balance <int> 8, 2, 14, 10, 2, 4, 1, 3, 10, 4, 3, 4, 3, 2...
## $ loan_purpose           <fct> debt_consolidation, credit_card, debt_conso...
## $ loan_amount            <int> 22000, 6000, 25000, 6000, 25000, 6400, 3000...
## $ grade                  <fct> B, B, E, B, B, B, D, A, A, C, D, A, A, A, A...
## $ interest_rate          <dbl> 10.90, 9.92, 26.30, 9.92, 9.43, 9.92, 17.09...
## $ public_record_bankrupt <int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0...
## $ loan_status            <fct> Current, Current, Current, Current, Current...
## $ has_second_income      <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F...
## $ total_income           <dbl> 59000, 60000, 75000, 75000, 254000, 67000, ...
```

Pregunta 2

Construya una tabla de frecuencias, frecuencia relativa y frecuencia porcentual para la variable `homeownership`

```r
loan50 %>%
  group_by(homeownership) %>%
  tally(sort = TRUE) %>%
  mutate(frec_relativa = n/sum(n), frec_porcentual = frec_relativa*100) %>%
  kbl() %>%
```

```
kable_styling(bootstrap_options = "striped", font_size = 12)
```

homeownership	n	frec_relativa	frec_porcentual
mortgage	26	0.52	52
rent	21	0.42	42
own	3	0.06	6

Pregunta 3

Construya una tabla cruzada de frecuencias que muestre la relación entre las variables `loan_status` y `homeownership`.

```
loan50 %>% tabyl(loan_status, homeownership) %>% adorn_totals(c("row", "col")) %>%
  adorn_title("combined") %>% kbl() %>% kable_styling(bootstrap_options = "striped", font_size = 12)
```

loan_status/homeownership	rent	mortgage	own	Total
	0	0	0	0
Charged Off	0	0	0	0
Current	20	21	3	44
Fully Paid	1	5	0	6
In Grace Period	0	0	0	0
Late (16-30 days)	0	0	0	0
Late (31-120 days)	0	0	0	0
Total	21	26	3	50

Pregunta 4

Presente una tabla con estadística descriptiva para la variable `interest_rate`.

```
loan50 %>% summarise(
    n = n(),                         # Número de observaciones
    mean = mean(interest_rate),      # Media aritmética
    sd = sd(interest_rate),          # Desviación estándar
    min = min(interest_rate),        # Mínimo
    p25 = quantile(interest_rate, 0.25), # Percentil 25 o primer cuartil
    p50 = median(interest_rate),     # Percentil 50 o segundo cuartil
    p75 = quantile(interest_rate, 0.75), # Percentil 75 o tercer cuartil
    max = max(interest_rate)) %>%
  kbl(digits = 2) %>% kable_styling(bootstrap_options = "striped", font_size = 13)
```

n	mean	sd	min	p25	p50	p75	max
50	11.57	5.05	5.31	7.96	9.93	13.71	26.3

Pregunta 3

Construya una matriz de correlaciones entre las variables

# Pregunta

```
loan50 %>% pull(loan_purpose) %>% levels()
```

```
##  [1] ""                  "car"               "credit_card"
##  [4] "debt_consolidation" "home_improvement"  "house"
##  [7] "major_purchase"    "medical"           "moving"
## [10] "other"             "renewable_energy"  "small_business"
## [13] "vacation"          "wedding"
```
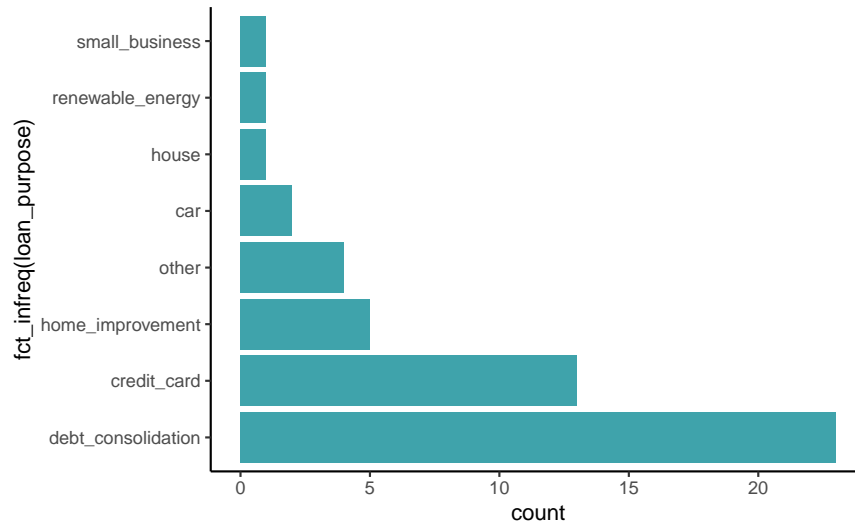
```
loan50 %>% count(loan_purpose, sort = TRUE)
```

```
## # A tibble: 8 x 2
##   loan_purpose          n
##   <fct>             <int>
## 1 debt_consolidation   23
## 2 credit_card          13
## 3 home_improvement      5
## 4 other                 4
## 5 car                   2
## 6 house                 1
## 7 renewable_energy      1
## 8 small_business        1
```
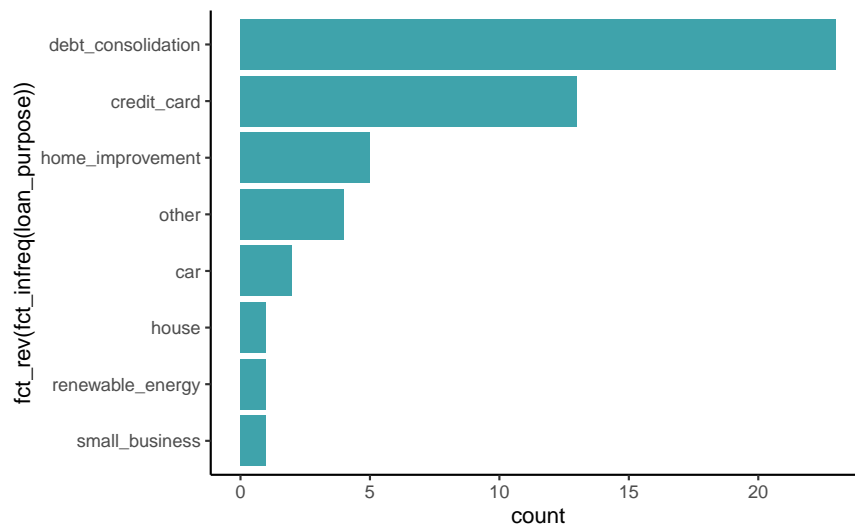
```
loan50 %>% ggplot(aes(x = loan_purpose)) +
  geom_bar(fill = "#3fa3ab") +
  theme_classic()
```
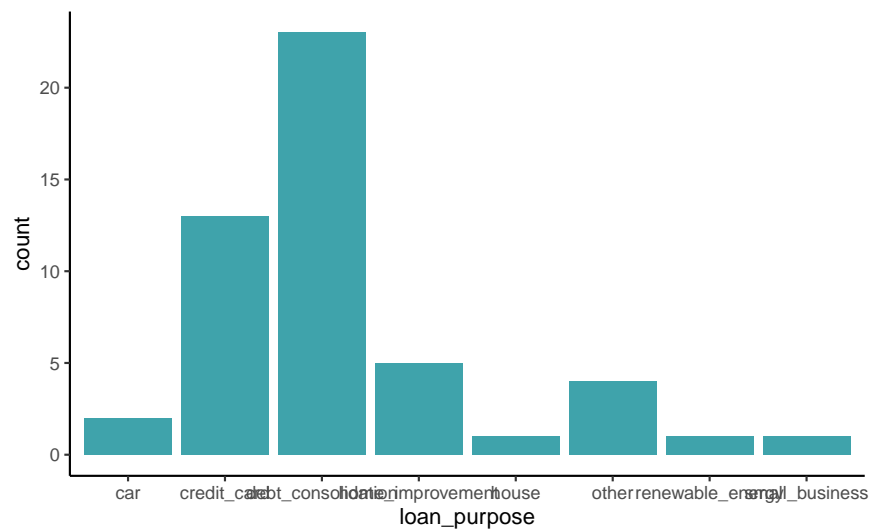


```
loan50 %>% ggplot(aes(x = fct_infreq(loan_purpose))) +
  geom_bar(fill = "#3fa3ab") +
  theme_classic() +
  coord_flip()
```

```
loan50 %>% ggplot(aes(x = fct_rev(fct_infreq(loan_purpose)))) +
  geom_bar(fill = "#3fa3ab") +
  theme_classic() +
  coord_flip()
```



```
loan50 %>%
  mutate(loan_purpose = fct_lump(loan_purpose, n = 5, other_level = "Extra")) %>%
  count(loan_purpose, sort = TRUE)
```

```
## # A tibble: 6 x 2
##   loan_purpose        n
##   <fct>           <int>
## 1 debt_consolidation  23
## 2 credit_card         13
## 3 home_improvement     5
## 4 other                4
## 5 Extra                3
## 6 car                  2
```

```
loan50 %>% ggplot(aes(x = homeownership, y = loan_amount, colour = homeownership)) +
  geom_jitter()
```
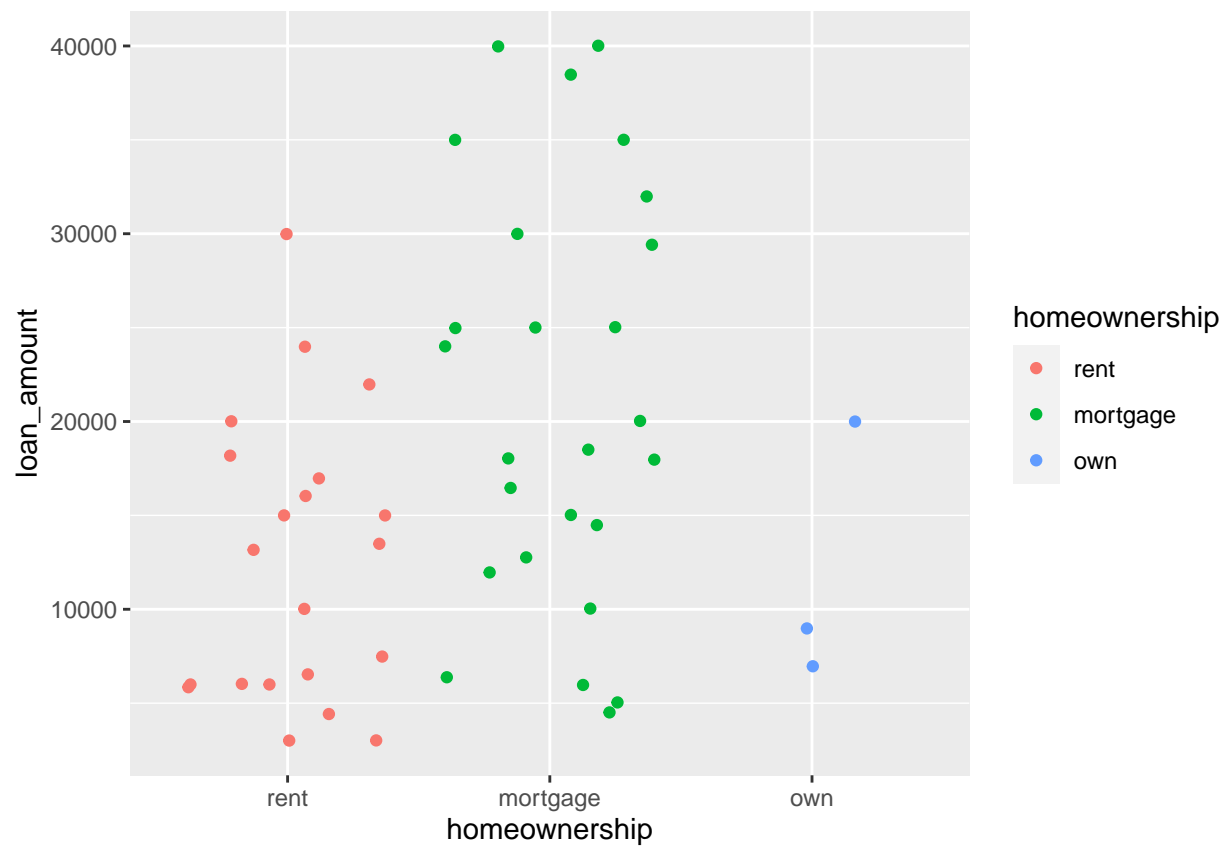


```
loan50 %>%
  mutate(homeownership = fct_reorder(homeownership, loan_amount)) %>%
  ggplot(aes(x = homeownership, y = loan_amount, colour = homeownership)) +
  geom_jitter()
```
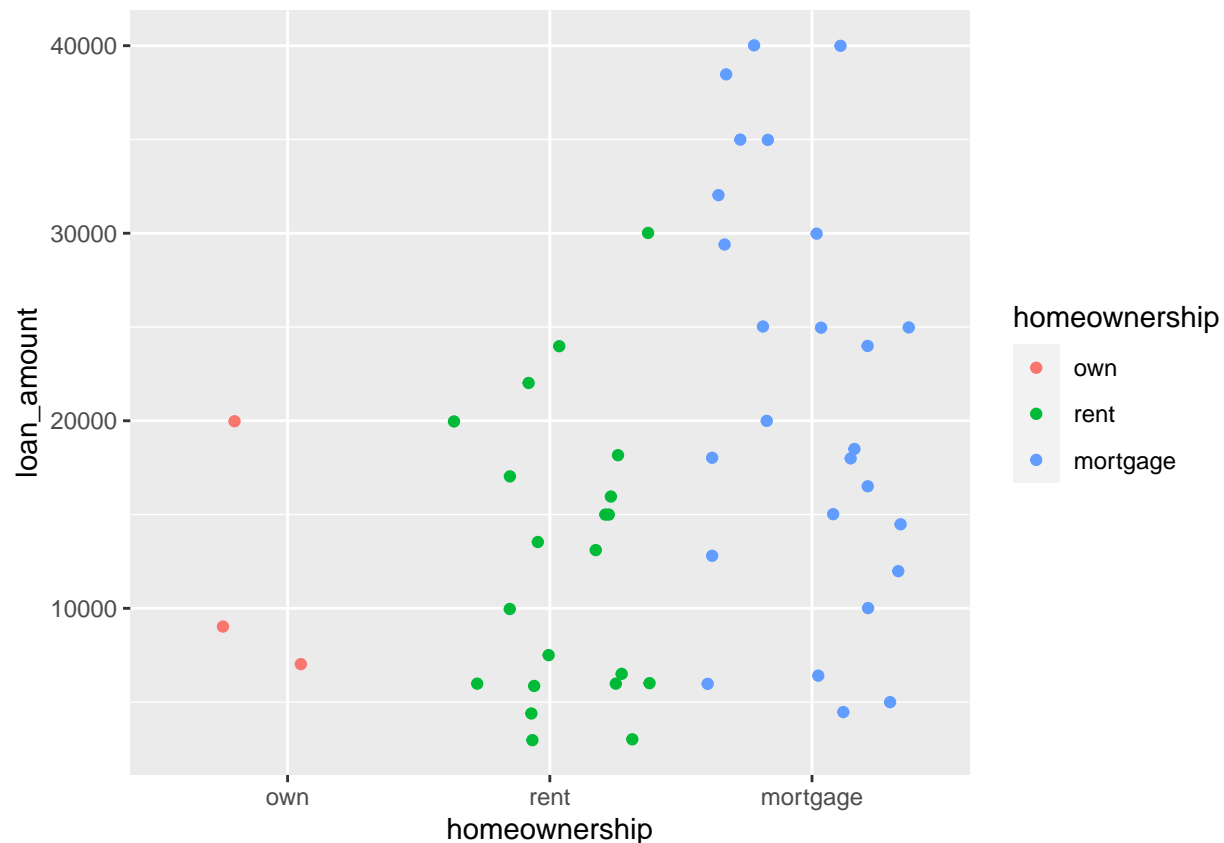
```r
mediana_ingreso <- loan50 %>% pull(total_income) %>% median()

loan50 %>% mutate(sup_median = if_else(total_income >= mediana_ingreso, "Mayor a mediana", "Menor a med
                  sup_median = factor(sup_median)) %>%
  tabyl(sup_median)
```

```
##        sup_median  n percent
##  Mayor a mediana 25     0.5
##  Menor a mediana 25     0.5
```

```r
ene <- read_dta("sample5_NDE2021.dta")
```

```r
ene %>% tabyl(activ)
```

```
## activ    n    percent valid_percent
##     1 1506 0.38794436    0.46959775
##     2  174 0.04482226    0.05425631
##     3 1527 0.39335394    0.47614593
##    NA  675 0.17387944           NA
```

```r
str(ene$activ)
```

```
##  dbl+lbl [1:3882] NA,  2,  3,  1,  1,  1,  3,  3, NA, NA,  1,  1, NA, NA,  ...
##  @ label      : chr "Condición de actividad"
##  @ format.stata: chr "%29.0g"
##  @ labels     : Named num [1:3] 1 2 3
##   ..- attr(*, "names")= chr [1:3] "Ocupados/as" "Desocupados/as" "Fuera de la fuerza de trabajo"
```

```
ene1 <- ene %>% filter(!is.na(activ)) %>% mutate(activ = as_factor(activ))
```

```
ene1 %>% tabyl(activ)
```

```
##                         activ    n    percent
##                   Ocupados/as 1506 0.46959775
##                 Desocupados/as  174 0.05425631
##  Fuera de la fuerza de trabajo 1527 0.47614593
```

```
ene <- ene %>% zap_label()
str(ene$activ)
```

```
##  dbl+lbl [1:3882] NA,  2,  3,  1,  1,  1,  3,  3, NA, NA,  1,  1, NA, NA,  ...
##  @ format.stata: chr "%29.0g"
##  @ labels      : Named num [1:3] 1 2 3
##   ..- attr(*, "names")= chr [1:3] "Ocupados/as" "Desocupados/as" "Fuera de la fuerza de trabajo"
```

```
ene <- ene %>% zap_labels()
str(ene$activ)
```

```
##  num [1:3882] NA 2 3 1 1 1 3 3 NA NA ...
##  - attr(*, "format.stata")= chr "%29.0g"
```