

Untitled

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(forcats)
library(janitor)

##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

library(haven)

data(loan50, package = "openintro")

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, ...
```

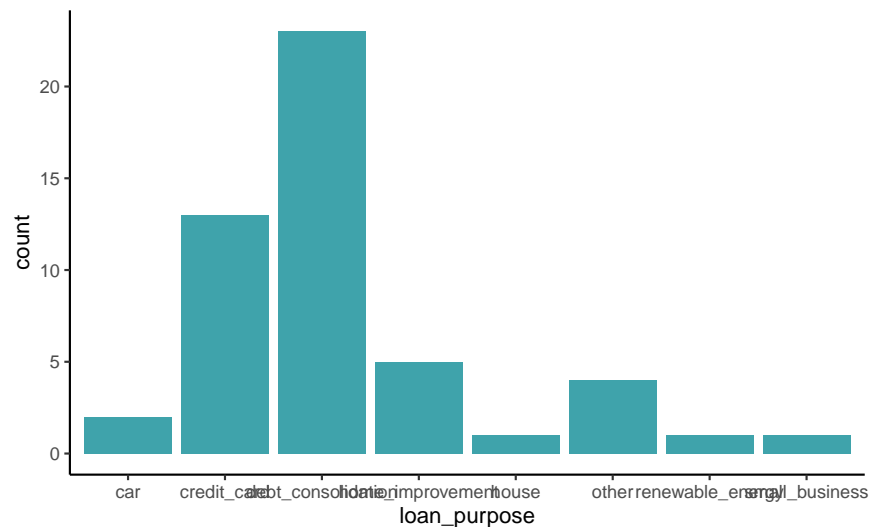
```
loan50 %>% pull(loan_purpose) %>% levels()
```

```
## [1] "" "car" "credit_card"
## [4] "debt_consolidation" "home_improvement" "house"
## [7] "major_purchase" "medical" "moving"
## [10] "other" "renewable_energy" "small_business"
## [13] "vacation" "wedding"
```

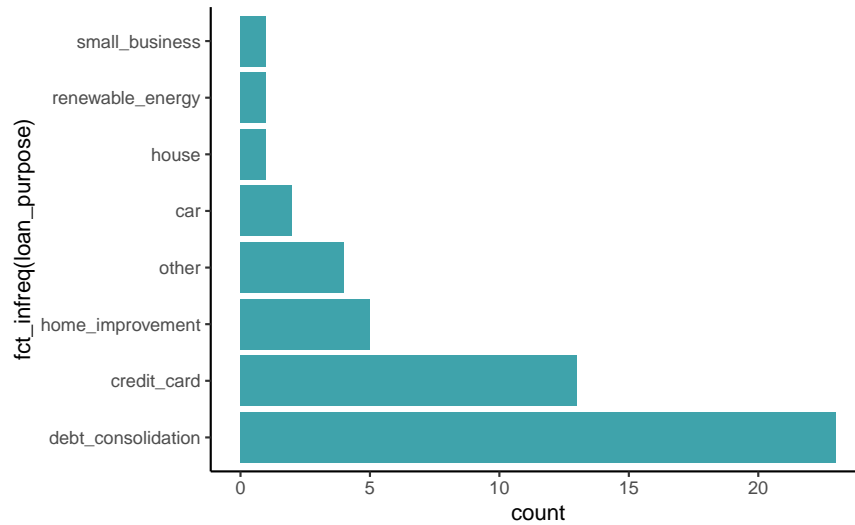
```
loan50 %>% count(loan_purpose, sort = TRUE)
```

```
## # 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
```

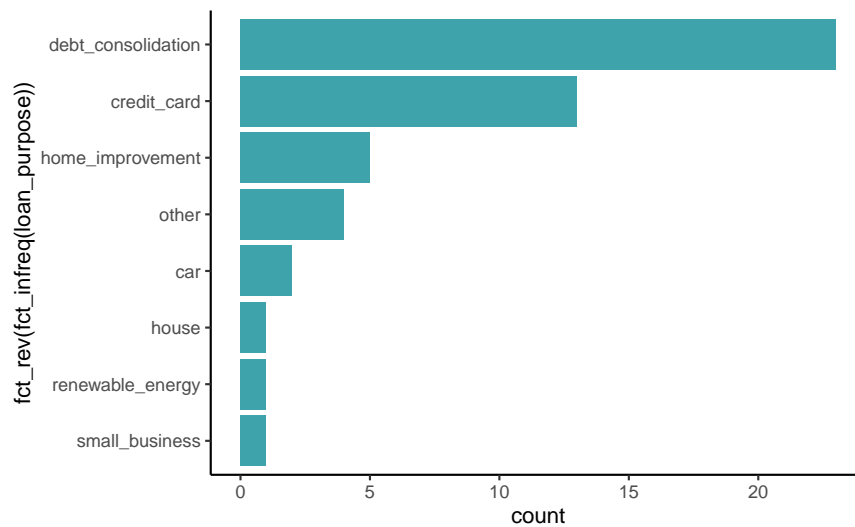
```
loan50 %>% ggplot(aes(x = loan_purpose)) +
  geom_bar(fill = "#3fa3ab") +
  theme_classic()
```



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



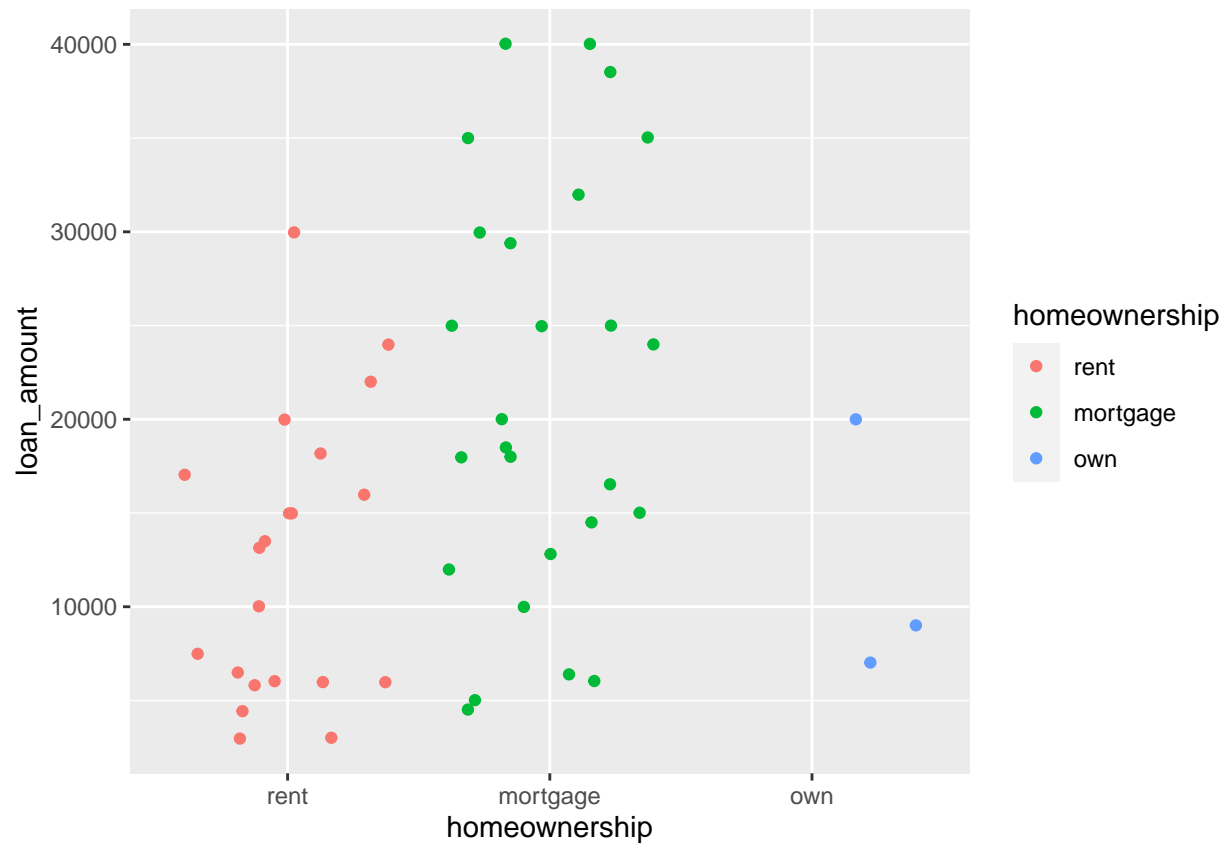
```
loan50 %>% ggplot(aes(x = fct_rev(fct_infreq(loan_purpose)))) +
  geom_bar(fill = "#3fa3ab") +
  theme_classic() +
  coord_flip()
```



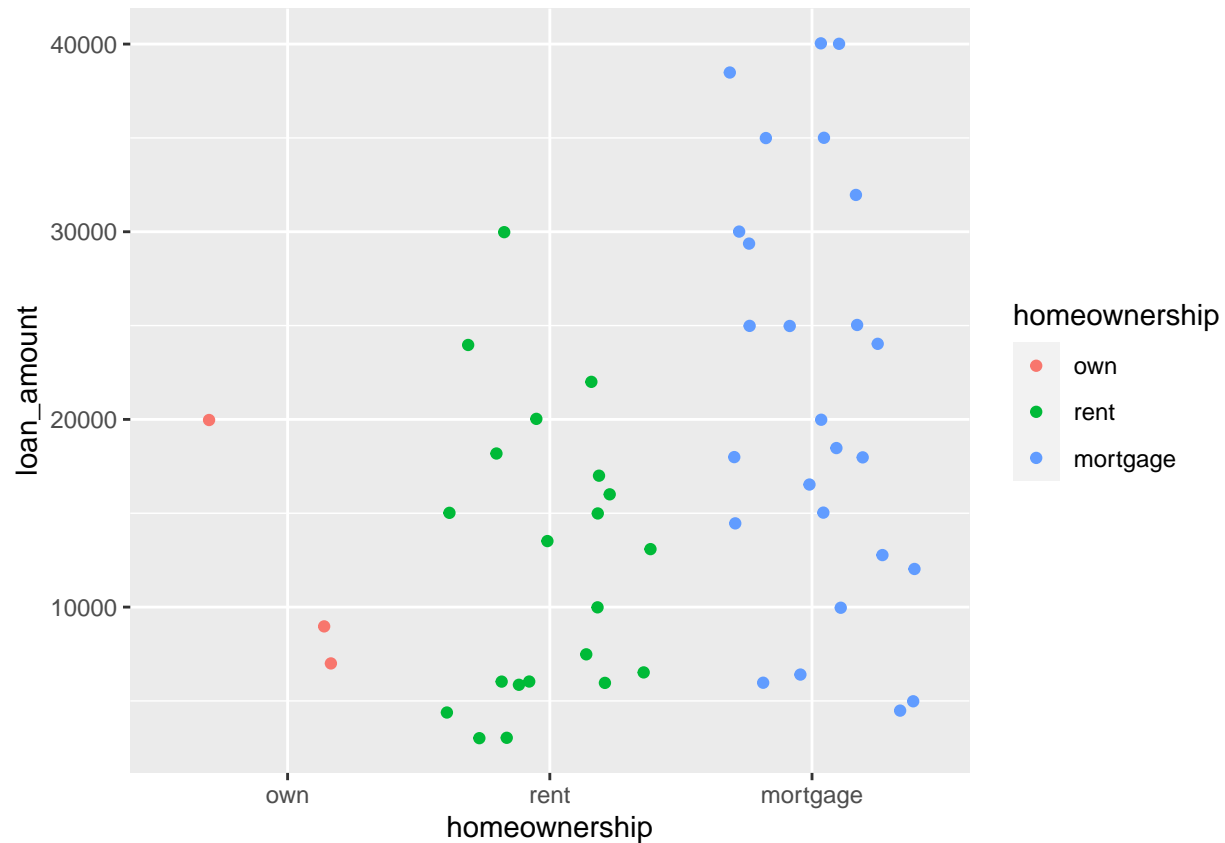
```
loan50 %>%
  mutate(loan_purpose = fct_lump(loan_purpose, n = 5, other_level = "Extra")) %>%
  count(loan_purpose, sort = TRUE)
```

```
## # 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()
```



```
mediana_ingreso <- loan50 %>% pull(total_income) %>% median()
```

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

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

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

```
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
```

```
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"

sexo <- labelled(
  c(2, 1, 1, 2, 2),
  labels = c(Mujer = 1, Hombre = 2),
  label = "Sexo de la persona"
)
edad <- c(21, 32, 28, 45, 24)

tb <- tibble::tibble(sexo, edad)
tb

## # A tibble: 5 x 2
##       sexo  edad
##   <dbl+lbl> <dbl>
## 1 2 [Hombre]    21
## 2 1 [Mujer]    32
## 3 1 [Mujer]    28
## 4 2 [Hombre]    45
## 5 2 [Hombre]    24

str(tb)

## tibble [5 x 2] (S3: tbl_df/tbl/data.frame)
## $ sexo: dbl+lbl [1:5] 2, 1, 1, 2, 2
##   ..@ labels: Named num [1:2] 1 2
##   ..- attr(*, "names")= chr [1:2] "Mujer" "Hombre"
##   ..@ label : chr "Sexo de la persona"
## $ edad: num [1:5] 21 32 28 45 24

tb %>% tabyl(sexo)

##  sexo n percent
##    1 2     0.4
##    2 3     0.6

tb <- tb %>% mutate(sexo = as_factor(sexo))
tb

```

```
## # A tibble: 5 x 2
##   sexo    edad
##   <fct>  <dbl>
## 1 Hombre    21
## 2 Mujer     32
## 3 Mujer     28
## 4 Hombre    45
## 5 Hombre    24
```

```
str(tb)
```

```
## tibble [5 x 2] (S3: tbl_df/tbl/data.frame)
##  $ sexo: Factor w/ 2 levels "Mujer","Hombre": 2 1 1 2 2
##    ..- attr(*, "label")= chr "Sexo de la persona"
##  $ edad: num [1:5] 21 32 28 45 24
```

```
tb %>% tabyl(sexo)
```

```
##   sexo n percent
##   Mujer 2    0.4
##   Hombre 3    0.6
```