

Interact and Cut Variables to Generate Categorical Variables

Fan Wang

2023-03-09

Contents

1	Generate Categorical Variables	1
1.1	Cut Continuous Variable to Categorical Variable	1
1.2	Factor, Label, Cross and Graph	2

1 Generate Categorical Variables

Go to the [RMD](#), [R](#), [PDF](#), or [HTML](#) version of this file. Go back to [fan's REconTools](#) research support package, [R4Econ](#) examples page, [PkgTestR](#) packaging guide, or [Stat4Econ](#) course page.

1.1 Cut Continuous Variable to Categorical Variable

We have a continuous variable, we cut it with explicitly specified cuts to generate a categorical variable, and label it. We will use `base::cut()`.

```
# break points to specific
fl_min_mpg <- min(mtcars$mpg)
fl_max_mpg <- max(mtcars$mpg)
ar_fl_cuts <- c(10, 20, 30, 40)
# generate labels
ar_st_cuts_lab <- c("10<=mpg<20", "20<=mpg<30", "30<=mpg<40")
# generate new variable
mtcars_cate <- mtcars %>%
  tibble::rownames_to_column(var = "cars") %>%
  mutate(mpg_grp = base::cut(mpg,
    breaks = ar_fl_cuts,
    labels = ar_st_cuts_lab,
    # if right is FALSE, interval is closed on the left
    right = FALSE
  ))
) %>% select(cars, mpg_grp, mpg) %>%
  arrange(mpg) %>% group_by(mpg_grp) %>%
  slice_head(n=3)
# Display
st_caption <- "Cuts a continuous var to a categorical var with labels"
kable(mtcars_cate,
  caption = st_caption
) %>% kable_styling_fc()
```

Cuts a continuous var to a categorical var with labels

cars	mpg_grp	mpg
Cadillac Fleetwood	10<=mpg<20	10.4
Lincoln Continental	10<=mpg<20	10.4
Camaro Z28	10<=mpg<20	13.3
Mazda RX4	20<=mpg<30	21.0
Mazda RX4 Wag	20<=mpg<30	21.0
Hornet 4 Drive	20<=mpg<30	21.4
Honda Civic	30<=mpg<40	30.4
Lotus Europa	30<=mpg<40	30.4
Fiat 128	30<=mpg<40	32.4

1.2 Factor, Label, Cross and Graph

Generate a Scatter plot with different colors representing different categories. There are multiple underlying factor/categorical variables, for example two binary variables. Generate scatter plot with colors for the combinations of these two binary variables.

We combine here the *vs* and *am* variables from the *mtcars* dataset. *vs* is engine shape, *am* is auto or manual shift. We will generate a scatter plot of *mpg* and *qsec* over four categories with different colors.

- *am*: Transmission (0 = automatic, 1 = manual)
- *vs*: Engine (0 = V-shaped, 1 = straight)
- *mpg*: miles per gallon
- *qsec*: 1/4 mile time

```
# First make sure these are factors
tb_mtcars <- as_tibble(mtcars) %>%
  mutate(vs = as_factor(vs), am = as_factor(am))

# Second Label the Factors
am_levels <- c(auto_shift = "0", manual_shift = "1")
vs_levels <- c(vshaped_engine = "0", straight_engine = "1")
tb_mtcars <- tb_mtcars %>%
  mutate(vs = fct_recode(vs, !!!vs_levels),
         am = fct_recode(am, !!!am_levels))

# Third Combine Factors
tb_mtcars_selected <- tb_mtcars %>%
  mutate(vs_am = fct_cross(vs, am, sep='_', keep_empty = FALSE)) %>%
  select(mpg, qsec, vs_am)

# relabel interaction variables
am_vs_levels <- c("vshape (engine) and auto (shift)" = "vshaped_engine_auto_shift",
                  "vshape (engine) and manual (shift)" = "vshaped_engine_manual_shift",
                  "straight (engine) and auto (shift)" = "straight_engine_auto_shift",
                  "straight (engine) and manual (shift)" = "straight_engine_manual_shift")
tb_mtcars_selected <- tb_mtcars_selected %>%
  mutate(vs_am = fct_recode(vs_am, !!!am_vs_levels))

# Show
print(tb_mtcars_selected[1:10,])
```

Now we generate scatter plot based on the combined factors

```

# Labeling
st_title <- paste0('Distribution of MPG and QSEC from mtcars')
st_subtitle <- paste0('https://fanwangecon.github.io/',
                      'R4Econ/amto/tibble/htmlpdf/fs_tib_factors.html')
st_caption <- paste0('mtcars dataset, ',
                     'https://fanwangecon.github.io/R4Econ/')
st_x_label <- 'MPG = Miles per Gallon'
st_y_label <- 'QSEC = time for 1/4 Miles'

# Graphing
plt_mtcars_scatter <-
  ggplot(tb_mtcars_selected,
        aes(x=mpg, y=qsec, colour=vs_am, shape=vs_am)) +
  geom_jitter(size=3, width = 0.15) +
  labs(title = st_title, subtitle = st_subtitle,
       x = st_x_label, y = st_y_label, caption = st_caption) +
  theme_bw()

# show
print(plt_mtcars_scatter)

```

