

R Dataframe Categorical Variables with forcats

Fan Wang

2021-10-28

Contents

1	Factor Label and Combine	1
1.1	Factor, Label, Cross and Graph	1

1 Factor Label and Combine

Go to the [RMD](#), [R](#), [PDF](#), or [HTML](#) version of this file. Go back to [fan's REconTools Package](#), [R Code Examples Repository \(bookdown site\)](#), or [Intro Stats with R Repository \(bookdown site\)](#).

1.1 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)

```

