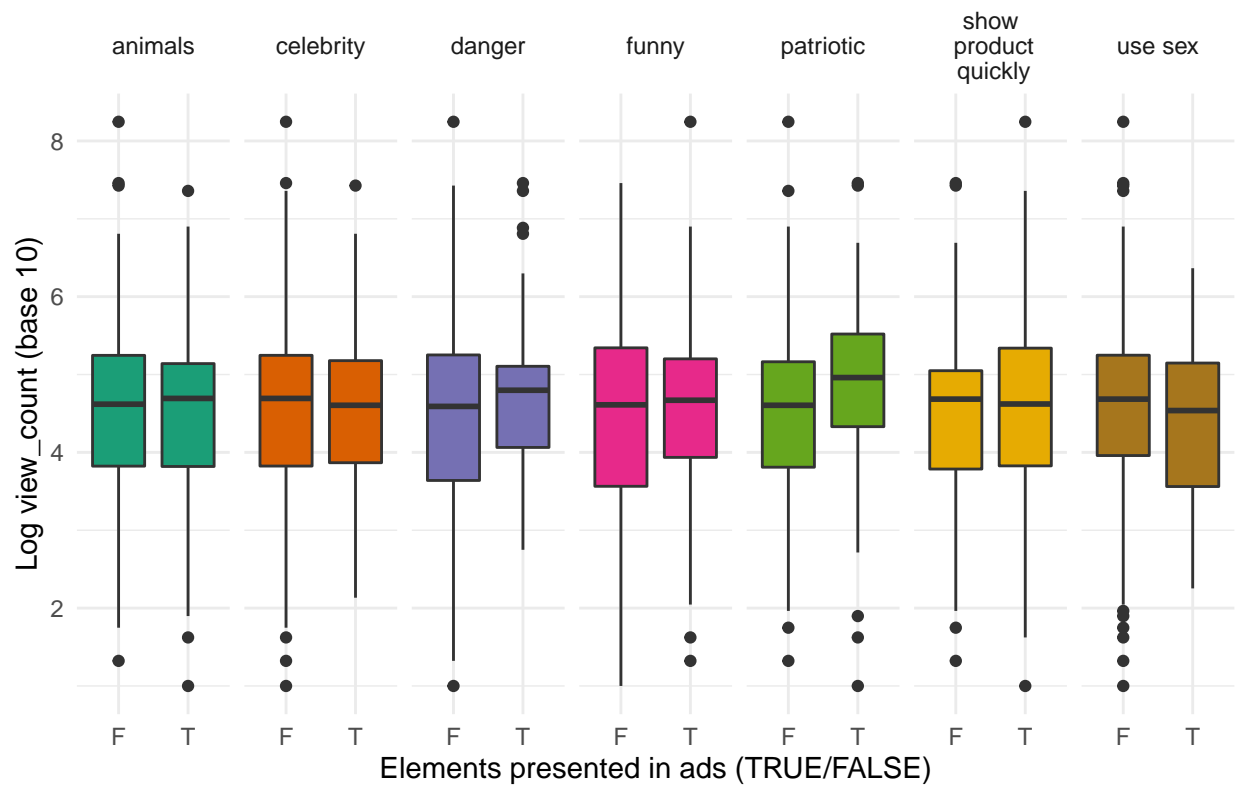# Superbowl Commercials

Yuzhou Pan

2021-03-05

Plot

Box plot of log view_count vs. each category of elements

Conclusion & What to improve

- The patriotic element contributes the most positively to the view counts.
- Check how to handle extreme values in the data set, conceptually. Maybe using logarithm is not wise here.

Codes

```
youtube <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/20
```

```r
library(tidyverse)
youtube %>%
  drop_na(view_count, like_count, dislike_count) -> youtube_new

youtube_new %>%
  select(c("brand", "view_count", "funny", "show_product_quickly", "patriotic", "celebrity", "danger",

rename(youtube_selected,
       "show\n product\n quickly" = "show_product_quickly",
       "use sex" = "use_sex") -> youtube_selected

youtube_selected %>%
  pivot_longer(
    !c(brand, view_count),
    names_to = "the_factor",
    values_to = "bool_value") %>%
  mutate(bool_value = if_else(bool_value == TRUE, "T", "F"),
         log_view_count = log(view_count, base = 10)) -> youtube_selected_shaped

youtube_selected_shaped %>%
  ggplot(aes(x = bool_value, y = log_view_count, fill = the_factor) ) +
  geom_boxplot() +
  facet_wrap(~the_factor, nrow = 1) +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(title = "Box plot of log view_count vs. each category of elements",
       x = "Elements presented in ads (TRUE/FALSE)",
       y = "Log view_count (base 10)") +
  scale_fill_brewer(palette = "Dark2")
```