Media_landscape_visualization

Yvonne JIN

10/11/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
setwd("~/Documents/Media_project")

library(readr)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ dplyr   1.0.7
## ✓ tibble  3.1.5     ✓ stringr 1.4.0
## ✓ tidyr   1.1.4     ✓ forcats 0.5.1
## ✓ purrr   0.3.4
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(RColorBrewer)
df <- read_csv("video_database_1010.csv")
```

```
## Rows: 104 Columns: 13
```

```
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (11): 账号名称, 视频名称, 类别分类, 风格分类（playful patriotism）, 上游视频发布主体,
上游视频数量, 上游视频...
## dbl  (2): 视频观看量(W), 弹幕数（W）
```

```
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#class(df) a tibble
colnames(df)
```

```
##  [1] "账号名称"                          "视频名称"
##  [3] "视频观看量(W)"                     "弹幕数（W）"
##  [5] "类别分类"                          "风格分类（playful patriotism）"
##  [7] "上游视频发布主体"                  "上游视频数量"
##  [9] "上游视频浏览量"                    "下游视频发布主体"
## [11] "下游视频数量"                      "下游视频浏览量"
## [13] "所属四类类别"
```

```
summary(df)
```

```
##    账号名称            视频名称          视频观看量(W)      弹幕数（W）
##  Length:104         Length:104         Min.   : 439.8   Min.   :   0.000
##  Class :character   Class :character   1st Qu.: 527.9   1st Qu.:   0.912
##  Mode  :character   Mode  :character   Median : 631.3   Median :   2.400
##                                        Mean   : 708.1   Mean   :  93.864
##                                        3rd Qu.: 787.7   3rd Qu.:   5.200
##                                        Max.   :2645.6   Max.   :5600.000
##                                        NA's   :4        NA's   :5
##    类别分类          风格分类（playful patriotism）  上游视频发布主体
##  Length:104         Length:104                     Length:104
##  Class :character   Class :character               Class :character
##  Mode  :character   Mode  :character               Mode  :character
##
##
##
##
##  上游视频数量        上游视频浏览量      下游视频发布主体    下游视频数量
##  Length:104         Length:104         Length:104         Length:104
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  下游视频浏览量      所属四类类别
##  Length:104         Length:104
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

Plots

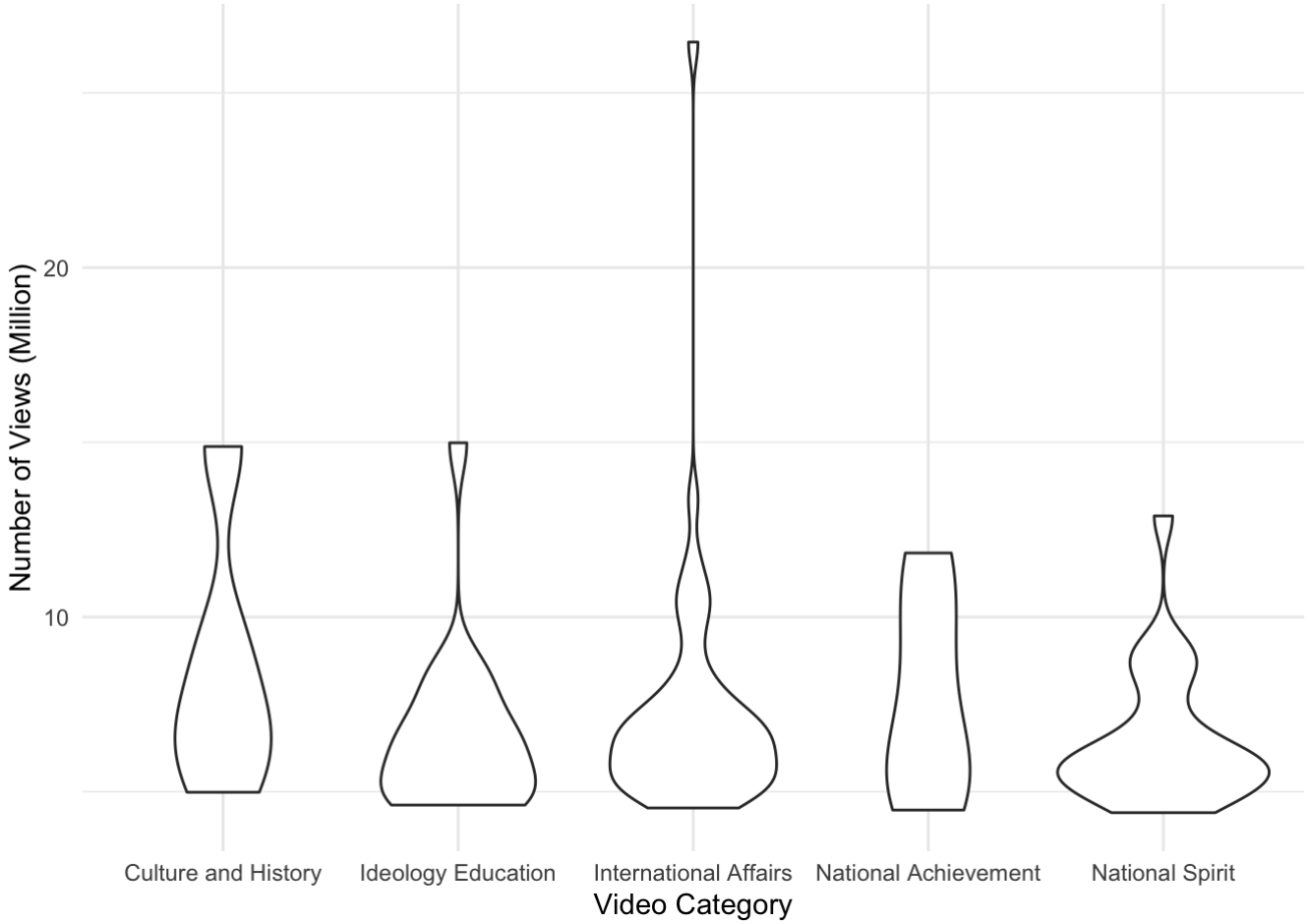
Violin plots showing number of views and number of videos in each category

Plot for views per category

```
##category
table(df_1$category)
```

```
##
##   Culture and History    Ideology Education International Affairs
##                     5                    18                    46
##  National Achievement       National Spirit
##                     8                    22
```

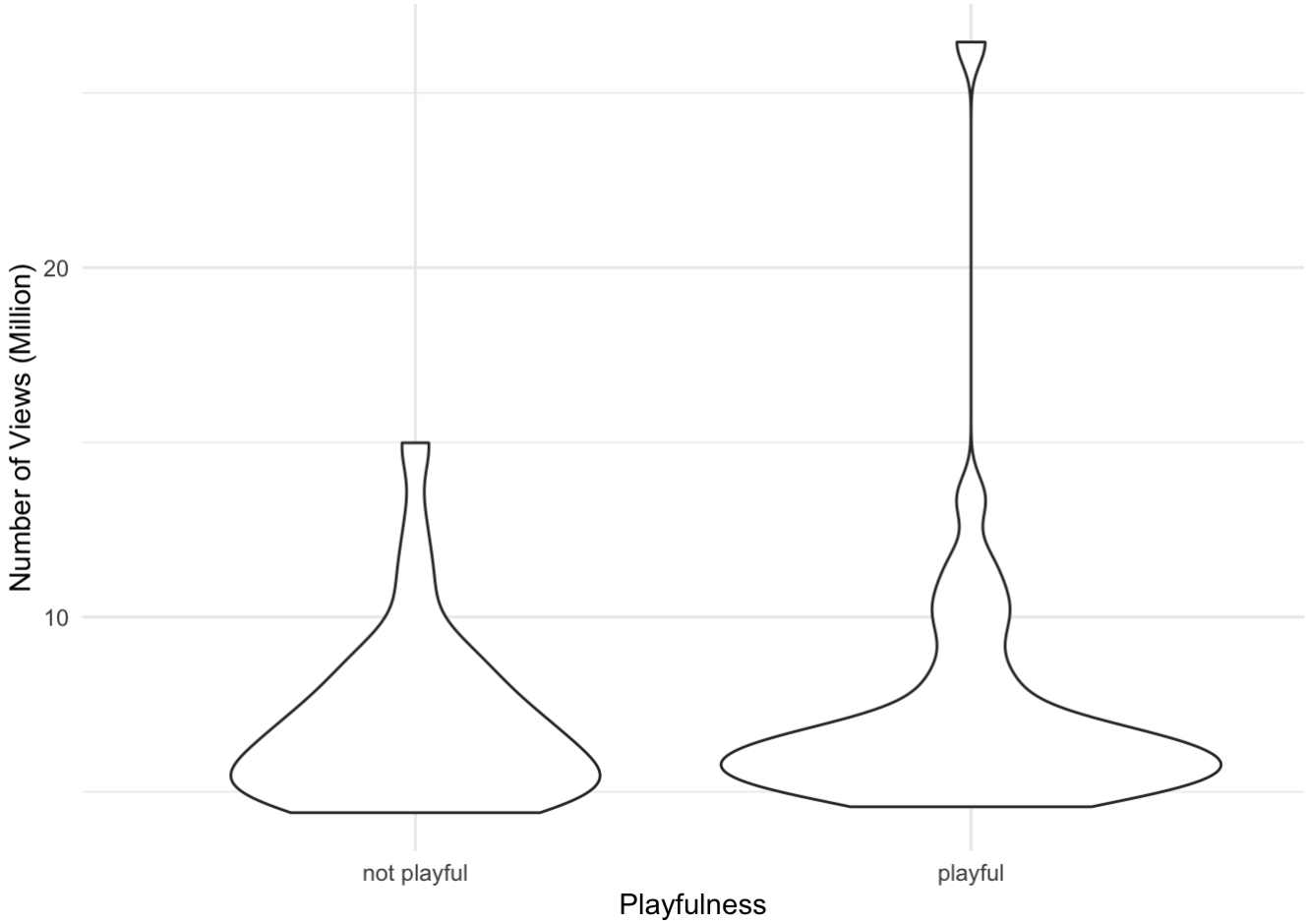
```
#translate category into english!

df_1 %>% ggplot(aes(category,view)) +
  geom_violin() +
  theme_minimal() +
  labs(x="Video Category", y="Number of Views (Million)")
```




```
##playfulness
table(df_1$playfulness)
```

```
##
## not playful     playful
##          62          37
```

```
df_1 %>% na.omit %>% ggplot(aes(playfulness,view)) +
  geom_violin() +
  theme_minimal() +
  labs(x="Playfulness", y="Number of Views (Million)")
```

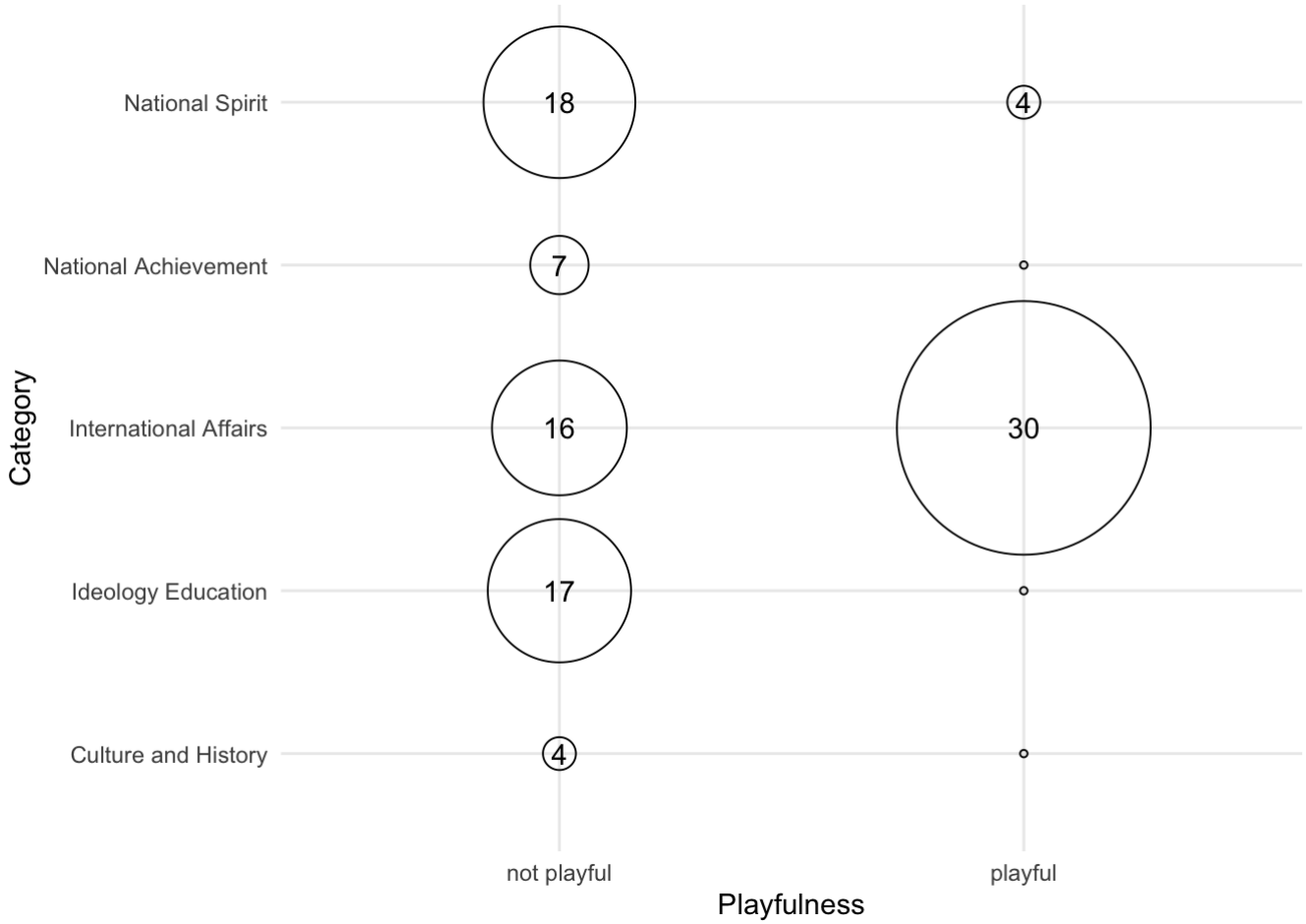


Playfulness X Category

```
#playful not playful cleaned-up


table(df_1$category, df_1$playfulness)
```

```
##
##                        not playful playful
##   Culture and History             4       1
##   Ideology Education             17       1
##   International Affairs          16      30
##   National Achievement            7       1
##   National Spirit                18       4
```

```
df_1 %>% na.omit() %>%
  ggplot(aes(x=playfulness,y=category)) +
  geom_count(shape = 1) +
  scale_shape(3,solid = FALSE) +
  scale_radius(range=c(1,45)) +
  theme_minimal() +
  annotate(geom = "text", x = "playful",y = "International Affairs", label = 30) +
  annotate(geom = "text", x = "not playful",y = "International Affairs", label = 16)
 +
  annotate(geom = "text", x = "playful",y = "National Spirit", label = 4) +
  annotate(geom = "text", x = "not playful",y = "National Spirit", label = 18) +
  annotate(geom = "text", x = "not playful",y = "Ideology Education", label = 17) +
  annotate(geom = "text", x = "not playful",y = "National Achievement", label = 7) +
  annotate(geom = "text", x = "not playful",y = "Culture and History", label = 4) +
  theme(legend.position = "none") +
  labs(x="Playfulness", y="Category")
```



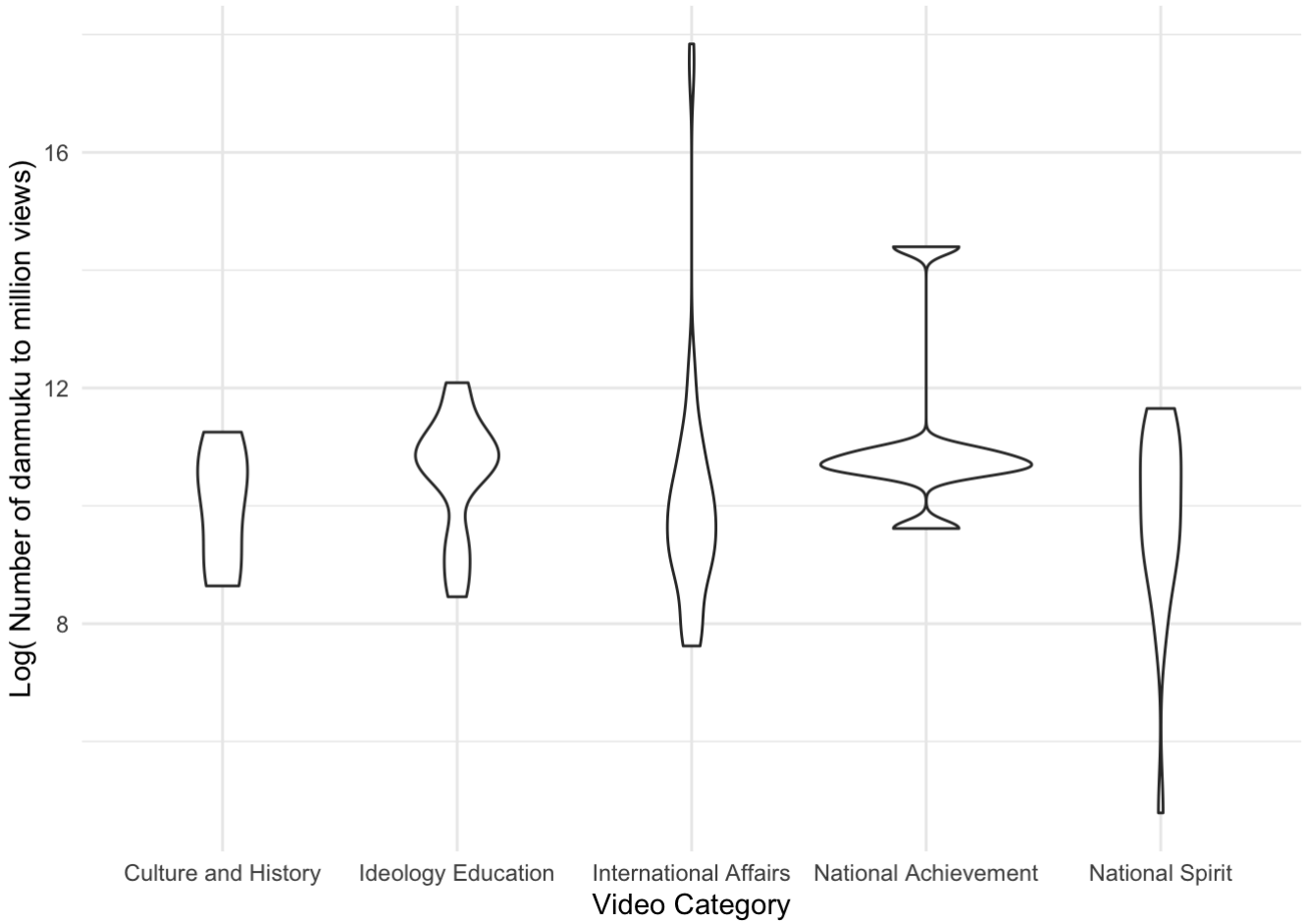

```
#add color
```

Heatmap (not in use)

Danmuku:Views - indicator for viewer participation

```
#added danmaku variable
df_1 %>% ggplot(aes(category,log(danmuku_n))) +
  geom_violin() +
  theme_minimal() +
  labs(x="Video Category", y="Log( Number of danmuku to million views)")
```

```
## Warning: Removed 4 rows containing non-finite values (stat_ydensity).
```



Recode for interaction modes

```
# recode upstream & downstream -> four categories?

colnames(df)
```

```
##  [1] "账号名称"                          "视频名称"
##  [3] "视频观看量(W)"                     "弹幕数（W）"
##  [5] "类别分类"                          "风格分类（playful patriotism）"
##  [7] "上游视频发布主体"                  "上游视频数量"
##  [9] "上游视频浏览量"                    "下游视频发布主体"
## [11] "下游视频数量"                      "下游视频浏览量"
## [13] "所属四类类别"
```

```
df_2 <- select(df, uploader = "账号名称", view = "视频观看量(W)", category = "类别分类",danmuku_n = "弹幕数（W）", playfulness = "风格分类（playful patriotism）", upstream_uploader = "上游视频发布主体", upstream_n = "上游视频数量", upstream_view = "上游视频浏览量", downstream_uploader = "下游视频发布主体", downstream_n = "下游视频数量", downstream_view = "下游视频浏览量")

summary(df_2)
```

```
##    uploader              view          category           danmuku_n
##  Length:104         Min.   : 439.8   Length:104         Min.   :   0.000
##  Class :character   1st Qu.: 527.9   Class :character   1st Qu.:   0.912
##  Mode  :character   Median : 631.3   Mode  :character   Median :   2.400
##                     Mean   : 708.1                      Mean   :  93.864
##                     3rd Qu.: 787.7                      3rd Qu.:   5.200
##                     Max.   :2645.6                      Max.   :5600.000
##                     NA's   :4                           NA's   :5
##  playfulness        upstream_uploader   upstream_n        upstream_view
##  Length:104         Length:104         Length:104         Length:104
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  downstream_uploader downstream_n       downstream_view
##  Length:104          Length:104         Length:104
##  Class :character    Class :character   Class :character
##  Mode  :character    Mode  :character   Mode  :character
##
##
##
##
```

```
# same as df_1
df_2 <- df_2 %>% mutate(
    category =  recode(category," (1) 国家、民族精神体现类"= "National Spirit",
                       " (2) 国家成就类" = "National Achievement",
                       " (3) 政治外交类" = "International Affairs",
                       " (4) 国家民族历史与文化类" = "Culture and History",
                       " (5) 思想教育、价值观引导类" = "Ideology Education"),
  view = view*0.01, # unit: million
  danmuku_n = danmuku_n*10000, # unit: individual
  d_to_view = danmuku_n/view
)

#disable scientific notation
df_2$d_to_view <- format(df_2$d_to_view, scientific = FALSE)

# check notation consistency
table(df_2$upstream_uploader)
```

```
##
##                     N/A            Non official            Non-official
##                      68                       6                       8
##                Official Official & Non-official   Official&Non-official
##                      13                       3                       1
```

```
table(df_2$downstream_uploader)
```

```
##
##                     N/A            Non-official Non-official & official
##                      37                      19                       1
##                Official Official & Non-official
##                      15                      27
```

```
df_2 <- df_2 %>% mutate(
  upstream_uploader = recode(upstream_uploader,"Non official" = "Non-official", "Offi
cial&Non-official" = "Official & Non-official"), # fix notation inconsistency
  downstream_uploader = recode(downstream_uploader, "Non-official & official" = "Offi
cial & Non-official"))

# check notation consistency
table(df_2$upstream_uploader)
```

```
##
##                     N/A            Non-official                Official
##                      68                      14                      13
## Official & Non-official
##                       4
```

```
table(df_2$downstream_uploader)
```

```
##
##                     N/A            Non-official                Official
##                      37                      19                      15
## Official & Non-official
##                      28
```

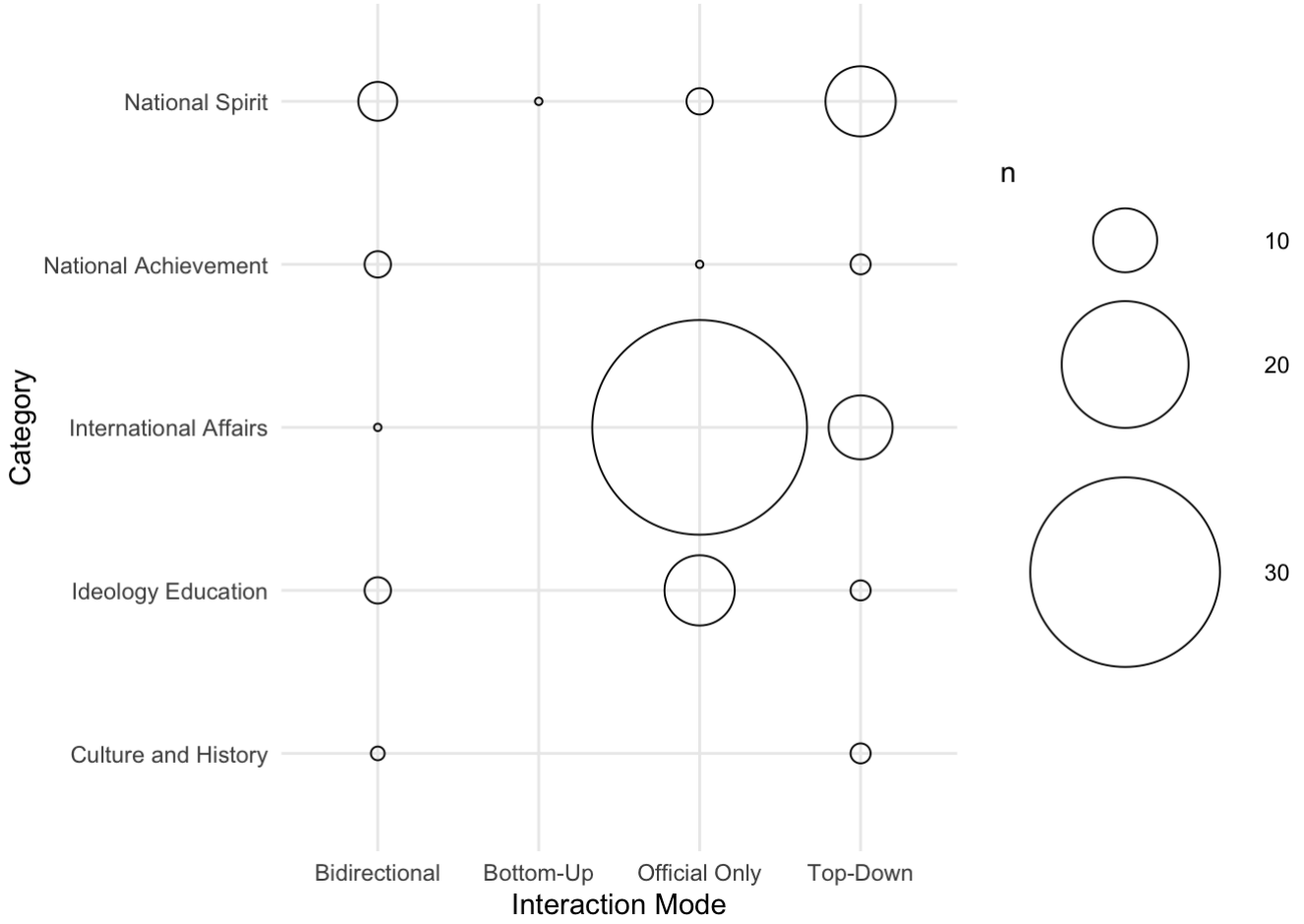
```
df_2 <- df_2 %>% mutate(
  interaction_mode = case_when(
    (upstream_uploader == "N/A" |upstream_uploader == "Official" ) & (downstream_uplo
ader == "N/A" | downstream_uploader == "Official") ~ "Official Only",
    (upstream_uploader == "N/A" |upstream_uploader == "Official" ) & (downstream_uplo
ader == "Non-official" | downstream_uploader == "Official & Non-official") ~ "Top-Dow
n",
    (upstream_uploader == "Non-official" | upstream_uploader == "Official & Non-offic
ial") & (downstream_uploader == "N/A" | downstream_uploader == "Official") ~ "Bottom-
Up",
    (upstream_uploader == "Non-official" | upstream_uploader == "Official & Non-offic
ial") & (downstream_uploader  == "Non-official" | downstream_uploader == "Official &
 Non-official" ) ~ "Bidirectional"
  ))

# Official Only: re-upload and interaction between official accounts
# Top-down: Official account upload first, non-official account imitate
# Bottom-up: Non-official account upload first, official account re-upload
# Bidirectional: Non-official account upload first, official account re-upload, non-o
fficial account then imitate
```

Category X Interaction Mode

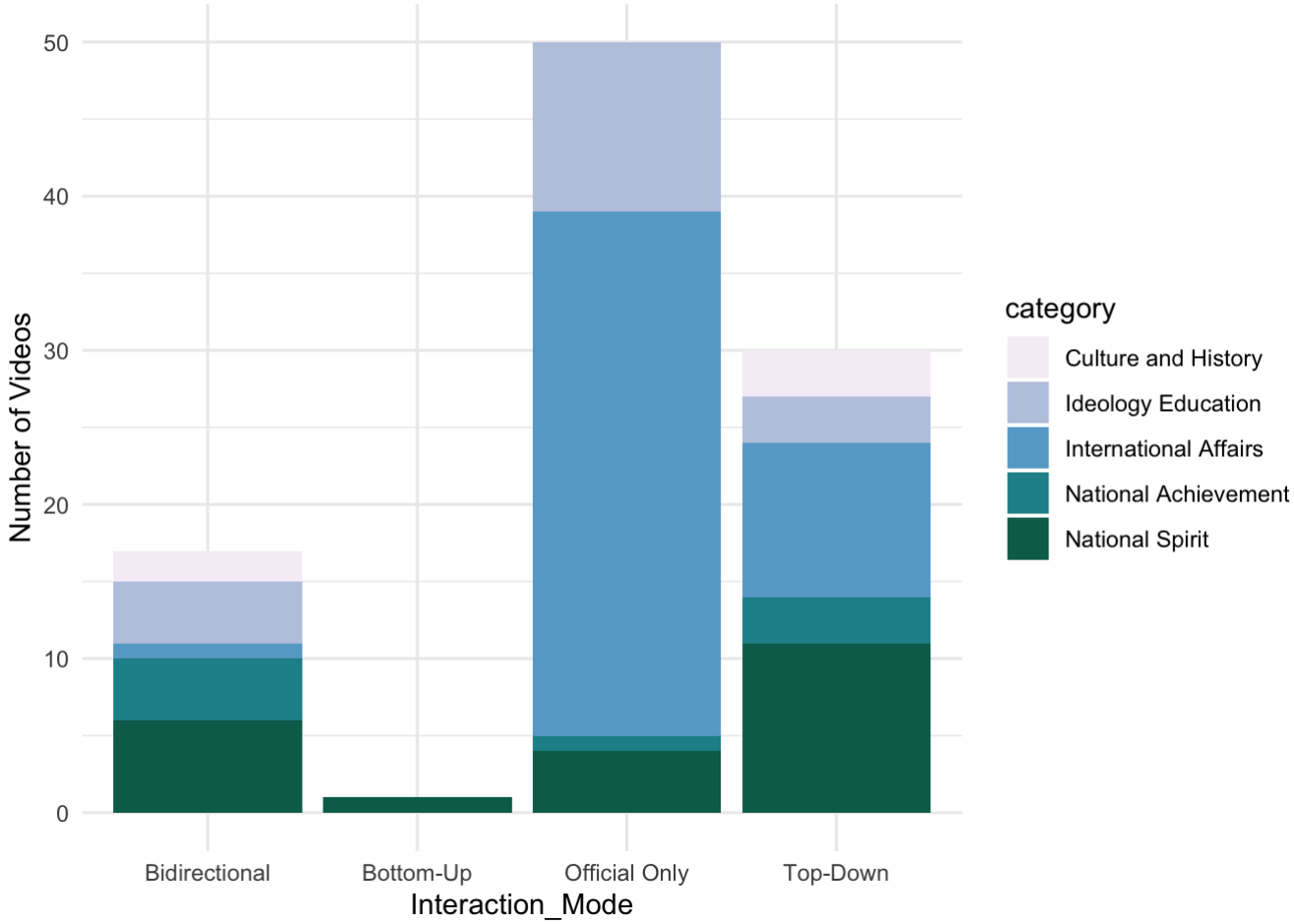
```
table(df_2$category,df_2$interaction_mode)
```

```
##
##                         Bidirectional Bottom-Up Official Only Top-Down
##   Culture and History               2         0             0        3
##   Ideology Education                4         0            11        3
##   International Affairs             1         0            35       10
##   National Achievement              4         0             1        3
##   National Spirit                   6         1             4       11
```

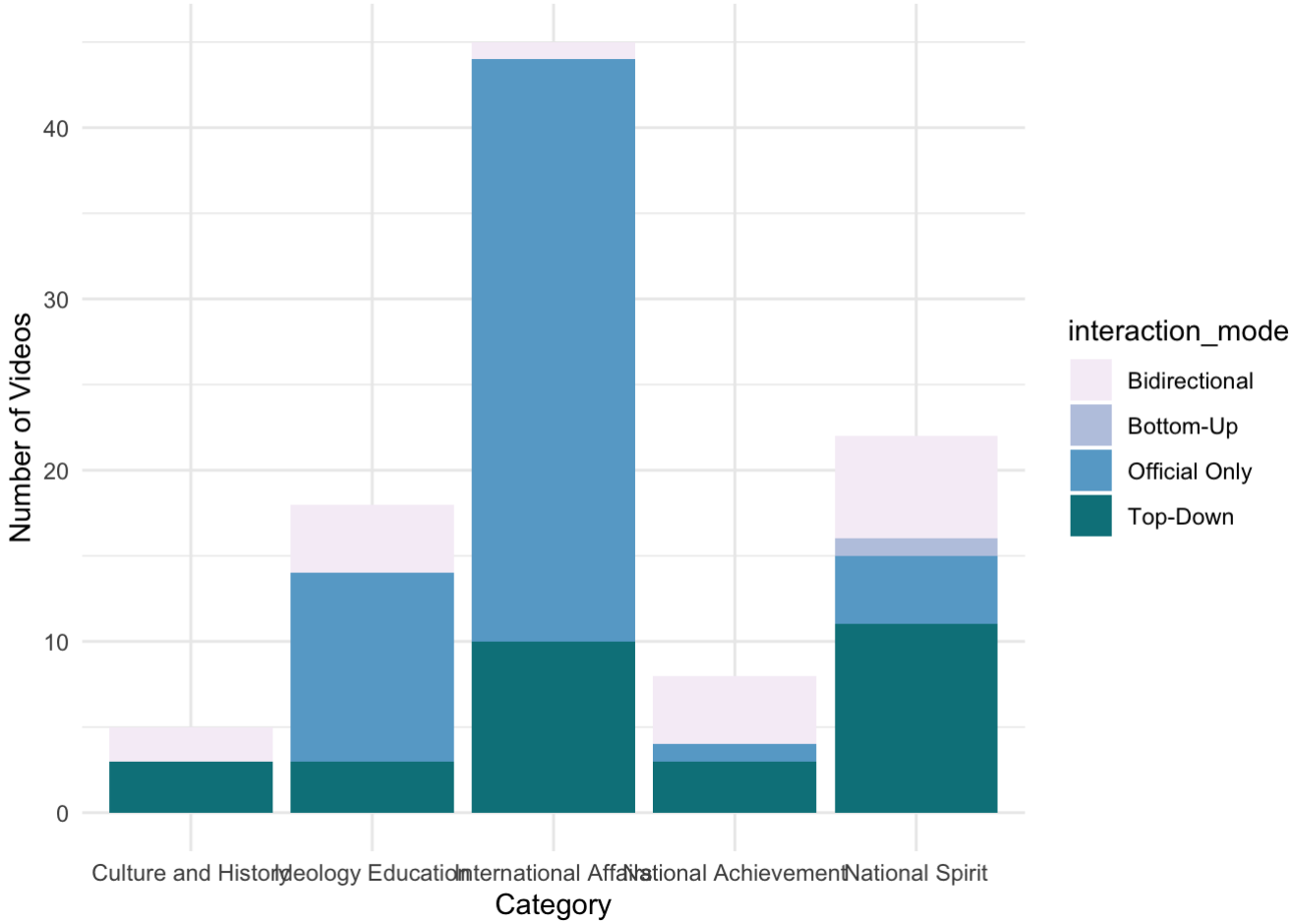
```
# count
df_2 %>% na.omit() %>%
  ggplot(aes(x=interaction_mode,y=category)) +
  geom_count(shape = 1) +
  scale_shape(3,solid = FALSE) +
  scale_radius(range=c(1,38)) +
  theme_minimal() +
  labs(x="Interaction Mode", y="Category")
```



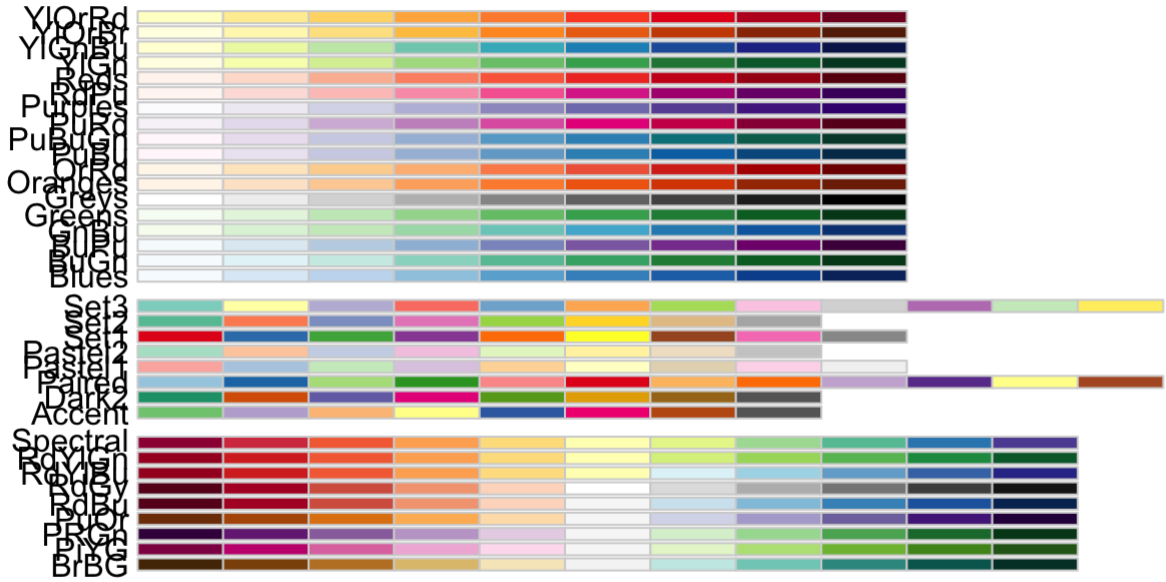

```
# bar chart
# 6 fill 4
df_2 %>% na.omit %>%
  ggplot(aes(x=interaction_mode,fill=category)) +
  geom_bar() +
  scale_fill_brewer(palette = "PuBuGn") +
  theme_minimal() +
  labs(x="Interaction_Mode", y="Number of Videos")
```




```
# 4 fill 6
df_2 %>% na.omit %>%
  ggplot(aes(x=category,fill=interaction_mode)) +
  geom_bar() +
  scale_fill_brewer(palette = "PuBuGn") +
  theme_minimal() +
  labs(x="Category", y="Number of Videos")
```

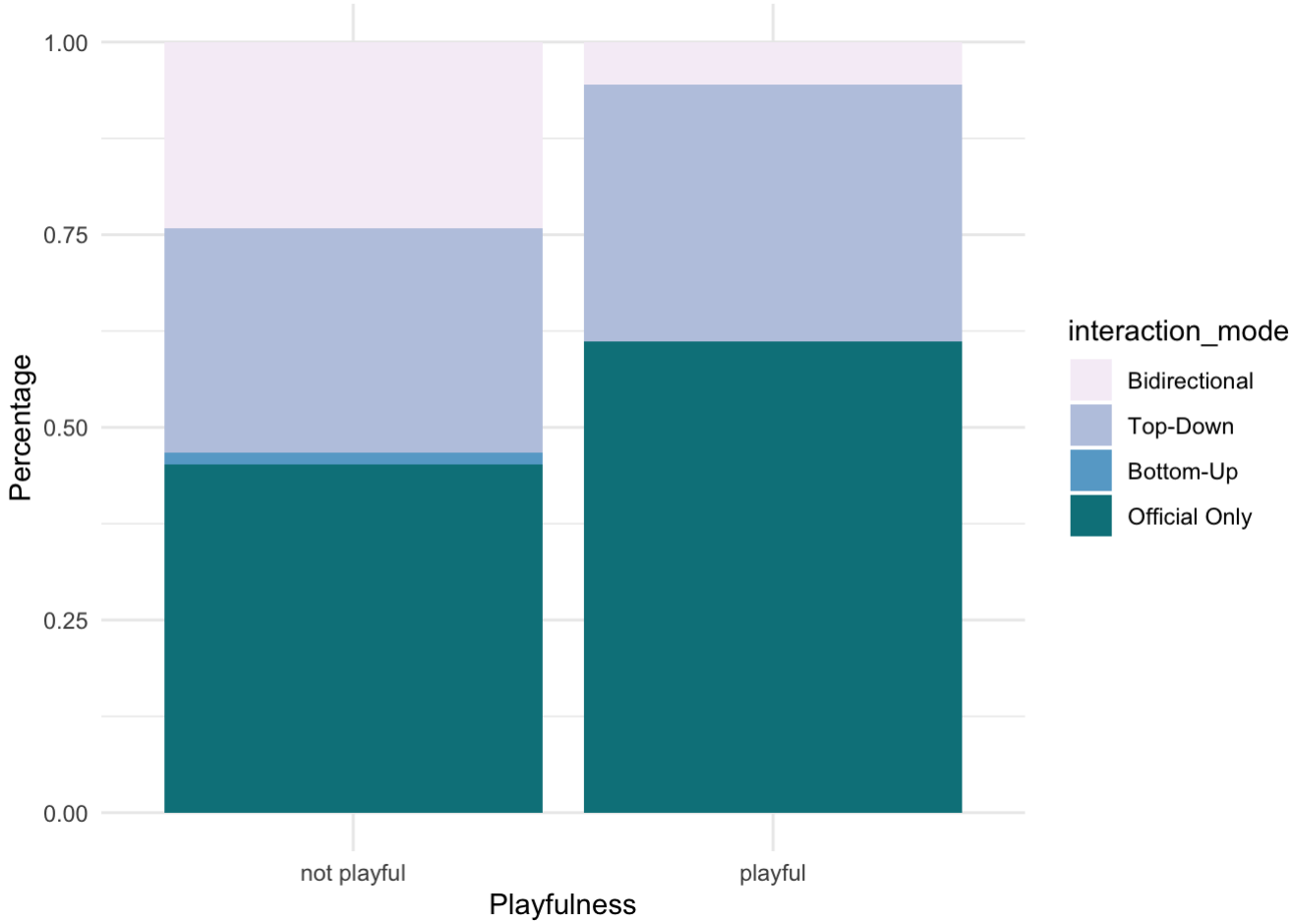



```
#heatmap
#df_2 %>% na.omit %>% ggplot(aes(x=category,y=interaction_mode,fill=view)) +
#  geom_tile()


#df_1 %>% ggplot(aes(x=category,y=playfulness, fill=view)) +
#  geom_tile() +
#  scale_fill_distiller(palette = "RdPu")

display.brewer.all()
```



playfulness X interactive mode

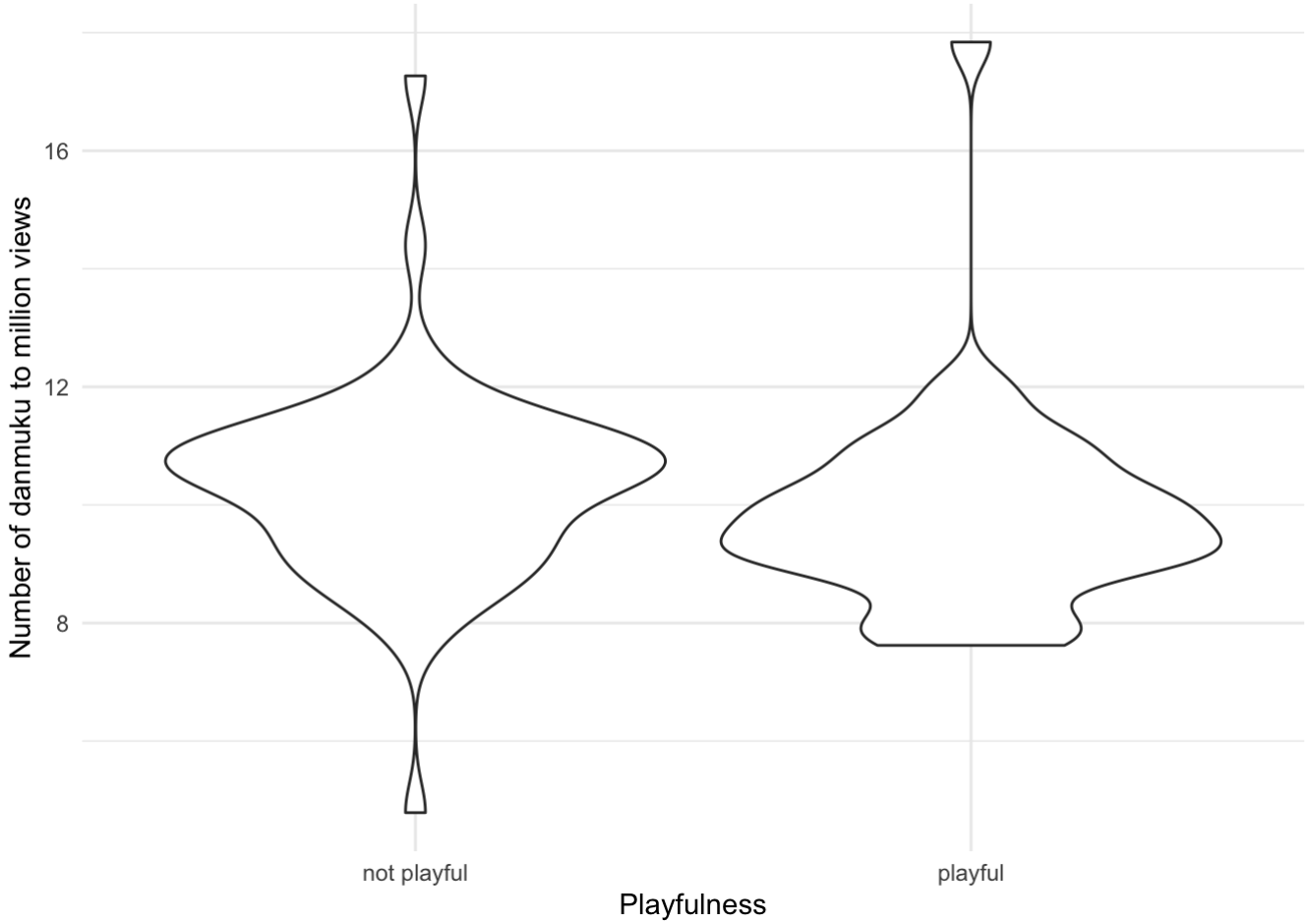
```
df_2 %>% na.omit %>%
  mutate(interaction_mode = factor(interaction_mode, levels=c("Bidirectional", "Top-D
own", "Bottom-Up","Official Only"))) %>%
  ggplot(aes(x=playfulness,fill=interaction_mode)) +
  geom_bar(position="fill") +
  scale_fill_brewer(palette = "PuBuGn") +
  theme_minimal() +
  labs(x="Playfulness", y="Percentage")
```



danmuku & playfulness

```
df_1 %>% ggplot(aes(playfulness,log(danmuku_n))) +
  geom_violin() +
  theme_minimal() +
  labs(x="Playfulness", y="Number of danmuku to million views")
```

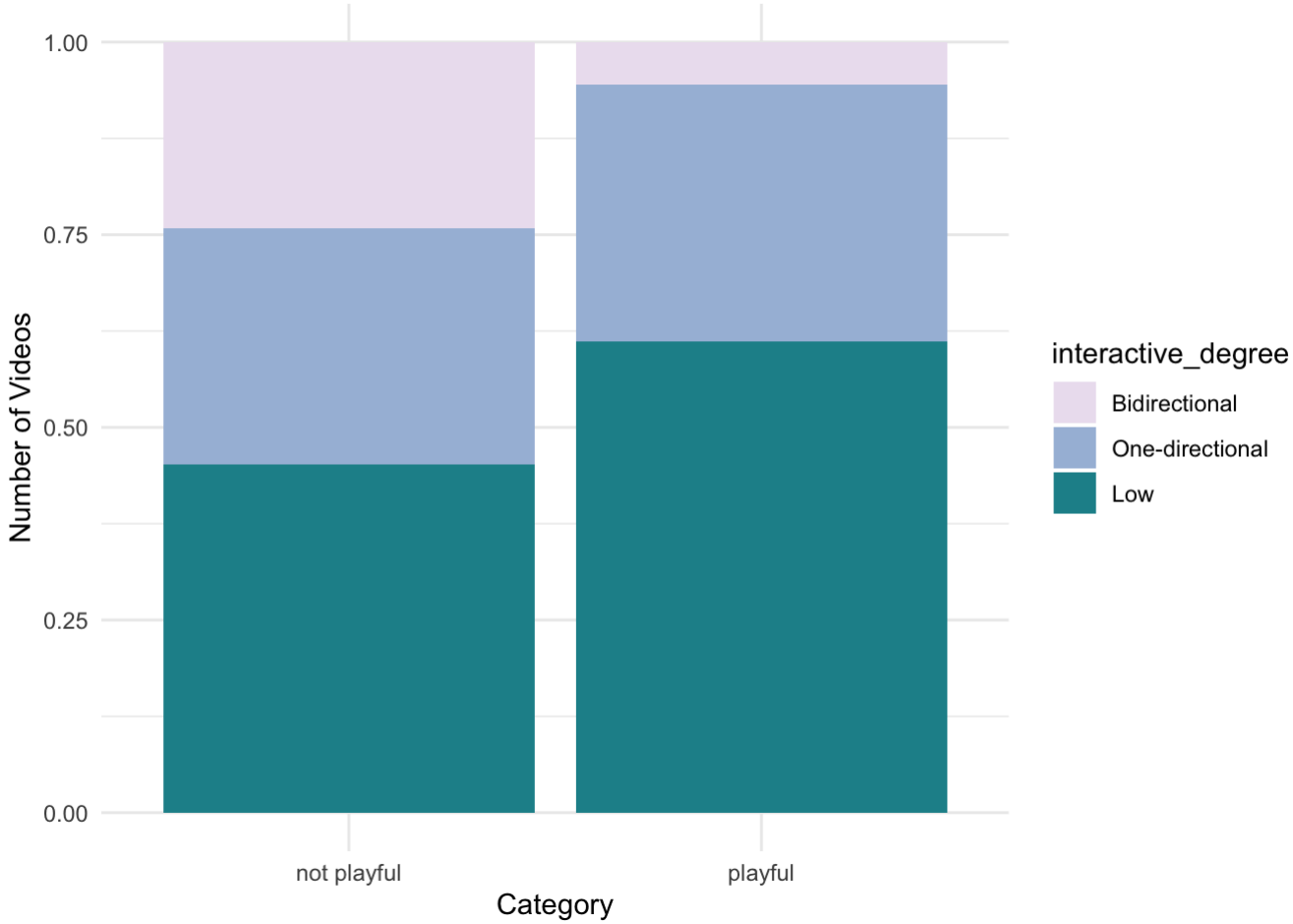
```
## Warning: Removed 4 rows containing non-finite values (stat_ydensity).
```



statistical analysis

playfulness -> more interactive?

```
df_2 <- df_2 %>% mutate(
  interactive_degree = case_when(
    interaction_mode == "Official Only" ~ "Low",
    interaction_mode == "Top-Down" | interaction_mode == "Bottom-Up" ~ "One-direction
al",
    interaction_mode == "Bidirectional" ~ "Bidirectional"
  )
)

df_2 %>% na.omit %>%
  mutate(interactive_degree = factor(interactive_degree, levels=c("Bidirectional", "O
ne-directional", "Low"))) %>%
  ggplot(aes(x=playfulness,fill=interactive_degree)) +
  geom_bar(position="fill") +
  scale_fill_brewer(palette = "PuBuGn") +
  theme_minimal() +
  labs(x="Category", y="Number of Videos")
```



playfulness -> more danmuku?

```
#In Jamovi
```

output

```
write.csv(df_2, "database_coded.csv",row.names = F, fileEncoding = "UTF-8")
```