Question 2 Analysis

```
knitr::opts_chunk$set(
  fig.width = 8,
  fig.asp = 0.618,
  fig.align = "center",
  fig.retina = 3,
  dpi = 300,
  out.width = "60%"
)

ggplot2::theme_set(ggplot2::theme_minimal(base_size = 16))
```

```
superbowl<- read.csv("data/superbowl.csv")
```

```
superbowl <- superbowl %>%
  mutate(ratio = like_count / (like_count + dislike_count))
```

```
outliers <- boxplot.stats(superbowl$view_count)$out


superbowl_no_outliers <- superbowl %>%
  filter(!(view_count %in% outliers))


quantile(superbowl_no_outliers$view_count,na.rm = TRUE)
```

```
##       0%      25%      50%      75%     100%
##     10.0   3852.5  29219.0  88066.0 403641.0
```

```
boxplot.stats(superbowl$view_count)$out
```

```
##  [1]  26727063    865781   1990447    669906   3464175    576696   1683994
##  [8]   2319854   6428474   1939823  28785122    503550   1274288   4921309
## [15]   1452877    598260   7952240    555734    582575   1404745    640393
## [22]    729583   1060001   1046640    955616  22849816    746836    491630
## [29]   7658201 176373378   3624622   1214968
```

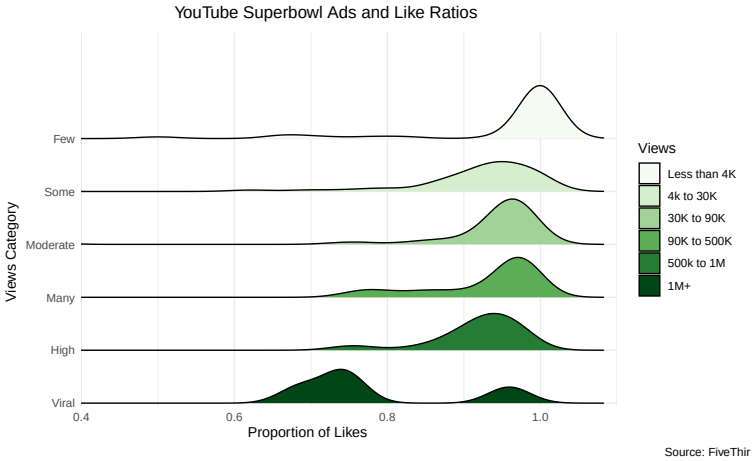
```
superbowl <-superbowl %>%
  mutate(view_category = case_when(
    view_count < 4000 ~ "Few",
    view_count >=4000 & view_count < 30000 ~ "Some",
    view_count >= 30000 &view_count < 90000 ~ "Moderate",
    view_count >= 90000 &view_count < 500000 ~ "Many",
    view_count >= 500000 &view_count < 10000000 ~ "High",
    TRUE ~ "Viral"
  ))

superbowl<-superbowl%>%
  mutate(interactions= like_count+dislike_count+comment_count)
```

```
superbowl %>%
  mutate(Views = fct_relevel(
    view_category, "Viral", "High", "Many", "Moderate","Some","Few"
  )) %>%
  ggplot(aes(x = ratio, y = Views, fill = Views)) +
    geom_density_ridges(scale = 1, show.legend = TRUE) +
    scale_fill_discrete_sequential(palette = "Greens", order = c(6:1),
                                   labels = c("1M+", "500k to 1M",
                                              "90K to 500K", "30K to 90K",
                                              "4k to 30K", "Less than 4K")) +
    scale_x_continuous(expand = c(0, 0.1)) +
    scale_y_discrete(expand = expand_scale(mult = c(0.01, .4))) +
    coord_cartesian(xlim = c(0.5, 1.0)) +
    labs(
      title = "YouTube Superbowl Ads and Like Ratios",
      x = "Proportion of Likes",
      y = "Views Category",
      caption = "Source: FiveThirtyEight") +
    guides(fill = guide_legend(reverse = TRUE)) +
    theme_minimal() +
    theme(plot.title = element_text(hjust = .4, vjust = 2),
          plot.caption = element_text(hjust = 1.41),
          axis.title.x = element_text(hjust = .4),
          axis.title.y = element_text(hjust = .38, vjust = 2))
```

```
## Warning: `expand_scale()` is deprecated; use `expansion()` instead.
```

```
## Picking joint bandwidth of 0.0275
```

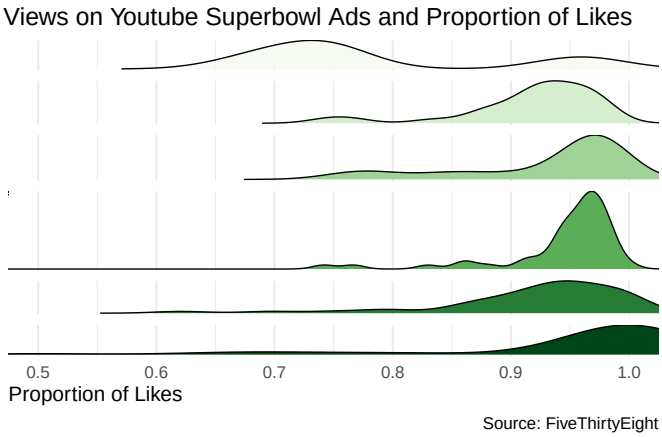
```
## Warning: Removed 31 rows containing non-finite values (stat_density_ridges).
```




```
superbowl %>%
  mutate(view_category = fct_relevel(
    view_category, "Viral", "High", "Many", "Moderate","Some","Few"
  )) %>%
  ggplot(aes(x = ratio, y = view_category, fill = view_category)) +
    geom_density_ridges(show.legend = FALSE) +
    geom_text(aes(label = view_category), x = 0.45, y = Inf) +
    scale_fill_discrete_sequential(palette = "Greens") +
    facet_grid(view_category ~., scales = "free_y", space = "free_y", switch = "x") +
```

```
    coord_cartesian(xlim = c(0.5, 1.0), ) +
    labs(
      title = "Views on Youtube Superbowl Ads and Proportion of Likes",
      x = "Proportion of Likes",
      y = NULL,
      caption = "Source: FiveThirtyEight") +
    theme(
      plot.margin = unit(c(1, 1, 1, 4), "lines"),
      plot.title.position = "plot",
      plot.caption.position = "plot",
      strip.background = element_blank(),
      strip.text = element_blank(),
      axis.title.x = element_text(hjust = 0),
      axis.text.y = element_blank(),
      axis.ticks.y = element_blank(),
      panel.grid.major.y = element_blank(),
      panel.grid.minor.y = element_blank()
  )
```

```
## Picking joint bandwidth of 0.0382
## Picking joint bandwidth of 0.0182
## Picking joint bandwidth of 0.025
## Picking joint bandwidth of 0.00935
## Picking joint bandwidth of 0.0227
## Picking joint bandwidth of 0.0517
## Warning: Removed 31 rows containing non-finite values (stat_density_ridges).
```



note: geom_grid interaction = likes + dislikes+ comments

```
superbowl%>%
  group_by(view_category)%>%
  select(interactions)
```

```
## Adding missing grouping variables: `view_category`
## # A tibble: 247 x 2
## # Groups:   view_category [6]
```

```
##    view_category interactions
##    <chr>                <int>
##  1 Many                    NA
##  2 Moderate               513
##  3 Many                   153
##  4 Few                      2
##  5 Some                    25
##  6 Some                   139
##  7 Many                  2081
##  8 Some                    90
##  9 Moderate               379
## 10 Few                      7
## # ... with 237 more rows
```

```
superbowl%>%
 group_by(view_category)%>%
  filter(!is.na(interactions))%>%
  summarise(vti= sum(view_count)/sum(interactions))
```

```
## # A tibble: 6 x 2
##   view_category   vti
##   <chr>         <dbl>
## 1 Few            149.
## 2 High           169.
## 3 Many           232.
## 4 Moderate       222.
## 5 Some           199.
## 6 Viral          330.
```

```
superbowl%>%
  mutate(view_category = paste(view_category, "Views"))%>%
   mutate(view_category = fct_relevel(
    view_category, "Few Views","Some Views","Moderate Views","Many Views", "High Views",  "Viral Views"
  ))%>%
ggplot(aes(x=view_count, y=interactions))+
    geom_area(aes(fill = view_category, alpha=0.5), show.legend = NULL)+
  geom_line()+
  facet_wrap(vars(view_category),
             scales="free",
             nrow=3,
             strip.position="top")+
  tag_facets(tag_pool = c("149","199","222", "232","169", "330"),
             position = list(x=.1, y=.76))+
    scale_y_continuous(label = label_number_si()) +
     scale_x_continuous(label = label_number_si())+
  labs(x= "View Count", y="Number of Interactions", title="Assessing Superbowl Ads Interactions",
       subtitle="Interactions is defined by sum of \nlikes, dislikes and comments on a video",
       caption="Number on Each Panel represents \nthe View To Interations Ratio of that view category")+
  theme(panel.grid.minor = element_blank(),
        strip.background = element_rect(colour = "black"),
        strip.placement = "inside",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        panel.spacing = unit(1.4, "lines"),
        tagger.panel.tag.text = element_text(color = "blue", size = 6),
```

```
      plot.caption = element_text(color = "blue", size = 8, hjust =0.5),
      plot.caption.position = "panel")+
  scale_fill_viridis(discrete = T)
```

```
## Warning: Removed 28 rows containing missing values (position_stack).
```

```
## Warning: Removed 16 row(s) containing missing values (geom_path).
```

```
## Warning: `show.legend` must be a logical vector.
```