```r
# CREATE DATA FOR JOYPLOT

## Load Data ####
genre_df <- read_csv('../data/cleaned_skatedata.csv')

## Create Dataframe of genre percentages by year
genre_pct_by_year <- genre_df %>%
  select(genre, year) %>%
  filter(year > 1988) %>%
  group_by(year) %>%
  mutate(year_cnt = n()) %>%
  ungroup() %>%
  group_by(genre, year, year_cnt) %>%
  summarise(gen_cnt = n()) %>%
  ungroup() %>%
  group_by(genre, year) %>%
  summarise(genre_pct = gen_cnt / year_cnt) %>%
  ungroup() %>%
  arrange(year) %>%
  rename(activity = genre,
         time = year,
         p = genre_pct)

## Create smooth dataset of areas for joyplot ####
joyplot_data <- genre_pct_by_year %>%
  group_by(activity) %>%
  filter(!is.na(activity)) %>%
  arrange(activity, time) %>%
  mutate(p_peak = p/max(p), # Normalize as percentage of peak popularity
         p_smooth = (lag(p_peak) + p_peak + lead(p_peak)) / 3, # Moving average
         p_smooth = coalesce(p_smooth, p_peak)) %>% # When there's no lag or lead, we get NA. Use the pointwise data
  ungroup()  %>%
  mutate(activity = reorder(activity, p, FUN=which.max))

## Save data to tsv ####
write_tsv(joyplot_data, "data.tsv")

```