clean_rows_name <- function(char){ partial_clean_names <- function(.data, unique = FALSE) { n <- if (is.data.frame(.data)) colnames(.data) else .data

```
n <- gsub("%+", "_pct_", n)
n <- gsub("\\$+", "_dollars_", n)
n <- gsub("\\++", "_plus_", n)
n <- gsub("-+", "_minus_", n)
n <- gsub("\\*+", "_star_", n)
n <- gsub("#+", "_cnt_", n)
n <- gsub("&+", "_and_", n)
n <- gsub("@+", "_at_", n)

n <- gsub("[^a-zA-Z0-9_]+", "_", n)
n <- gsub("([A-Z][a-z])", "_\\1", n)
n <- tolower(trimws(n))

n <- gsub("(^_+|_+$)", "", n)

n <- gsub("_+", "_", n)

if (unique) n <- make.unique(n, sep = "_")

if (is.data.frame(.data)) {
  colnames(.data) <- n
  .data
} else {
  n
}
```

}

char %>% partial_clean_names() %>% str_replace_all('_',' ') %>% toTitleCase() %>% str_replace_all(' i ',' I ') %>% StrCap() %>% return() }

Package

pacman::p_load(here, data.table, magrittr, tidyverse, janitor, ggplot2, ggthemes, tools, DescTools)

Read

spotify_data <- fread(here('data.csv'))

Clean

spoti_clean <- spotify_data %>% filter(Region == 'global') %>% clean_names() %>% mutate(date = as.Date(date, '%Y-%m-%d'), year = year(date), month = month(date), track_name = clean_rows_name(track_name), artist = clean_rows_name(artist), track_name = ifelse(str_detect(track_name, 'Despacito') == T, 'Despacito', track_name), general_name = paste0(artist, ':', track_name)) %>% filter(track_name != '' & artist != '') # Rows without a name/artist

We select the top 1 position songs

tops <- spoti_clean %>% filter(position == 1) %>% select(general_name) %>% unique()

tops <- tops %>% bind_cols(colors = rainbow(nrow(tops)))

Sum all the objects with the same features

general_data <- spoti_clean %>%
group_by(general_name, artist, track_name, date, region, year, month) %>% summarise(streams = sum(streams)) %>% ungroup()

Select the top

top_tracks <- general_data %>% filter(general_name %in% (tops %>% select(general_name) %>% pull())) %>% left_join(tops)

top_positions <- general_data %>% group_by(date) %>% mutate(max = ifelse(max(streams) == streams, 1, 0)) %>% ungroup() %>% filter(max ==1) %>% left_join(tops)

Plot

top_tracks %>% select(general_name, colors) %>% unique()

top_tracks %>% ggplot(aes(x = date, y = streams, group = colors, color = general_name)) + geom_line(alpha = 0.3, show.legend = FALSE, size = 0.6) + geom_line(data = top_positions, aes(x = date, y = streams), size = 0.9) + labs(title = "Top songs in Spotify", subtitle = '2017 - 2018') + theme(axis.title = element_blank(),

```
axis.text.x = element_text(size = 10, color = "gray20", angle = 0),
axis.text.y = element_blank(),

axis.ticks.x = element_blank(),
axis.ticks.y = element_blank(),

panel.background = element_blank(),
panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),

legend.title = element_text(size=10, color = "gray30", face="bold"),
legend.text = element_text(size=8, color = "gray30", face="bold"),
legend.justification=c(1,3),
#legend.position=c(0.05, 0.95),
legend.background = element_blank(),
legend.key = element_blank()
```

) + scale_color_manual("Name:", values = top_tracks$colors, labels = top_tracks$track_name)