

# Text Analysis on Animal Crossing Reviews

Nourhan Ghanima

2023-11-29

## Animal Crossing

Animal Crossing: New Horizons is a 2020 life simulation video game developed and published by Nintendo for the Nintendo Switch. It is the fifth main series title in the Animal Crossing series. New Horizons was released in all regions on March 20, 2020. It's been one of the most popular games globally since 2020.

New Horizons sees the player assuming the role of a customizable character who moves to a deserted island after purchasing a package from Tom Nook, a tanuki character who has appeared in every entry in the Animal Crossing series. Taking place in real-time, the player can explore the island in a nonlinear fashion, gathering and crafting items, catching insects and fish, and developing the island into a community of anthropomorphic animals.

## Basic Setup

I start as usual by loading the required libraries.

```
library(tidyverse)
library(tidytext)
library(widyr)
library(igraph)
library(ggraph)
```

```
user_reviews <- readr::read_tsv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/d
```

The reviews data has 4 columns, a user name, the grade each user assigned, the text of the review, and finally the date of the review. The grades range from 0 to 10, with 10 being the highest.

```
head(user_reviews)
```

```
## # A tibble: 6 x 4
##   grade user_name    text                                     date
##   <dbl> <chr>         <chr>                                     <date>
## 1     4 mds27272      My gf started playing before me. No option to c~ 2020-03-20
## 2     5 lolo2178    While the game itself is great, really relaxing~ 2020-03-20
## 3     0 Roachant     My wife and I were looking forward to playing t~ 2020-03-20
## 4     0 Houndf      We need equal values and opportunities for all ~ 2020-03-20
## 5     0 ProfessorFox BEWARE! If you have multiple people in your ho~ 2020-03-20
## 6     0 tb726       The limitation of one island per Switch (not pe~ 2020-03-20
```

The next step is to separate the reviews into words and getting rid of the stop words. Stop words are common words that have little or no significance in a text analysis; they are common words that by themselves don't carry much information. They're removed to improve efficiency and accuracy. This code chunk creates a list of words for each user that left a review.

```
review_words <- user_reviews |>
  unnest_tokens(output = word, input = text) |>
```

```
anti_join(stop_words, by = "word") |>
filter(str_detect(word, "[:alpha:]")) |>
distinct()
```

I then count the number of times a word is used by different users, filtering by words used by 100 or more users.

```
users_who_mention_word <- review_words |>
  dplyr::count(word, name = "users_n") |>
  filter(users_n >= 100)
```

I then look at the correlations of words across users.

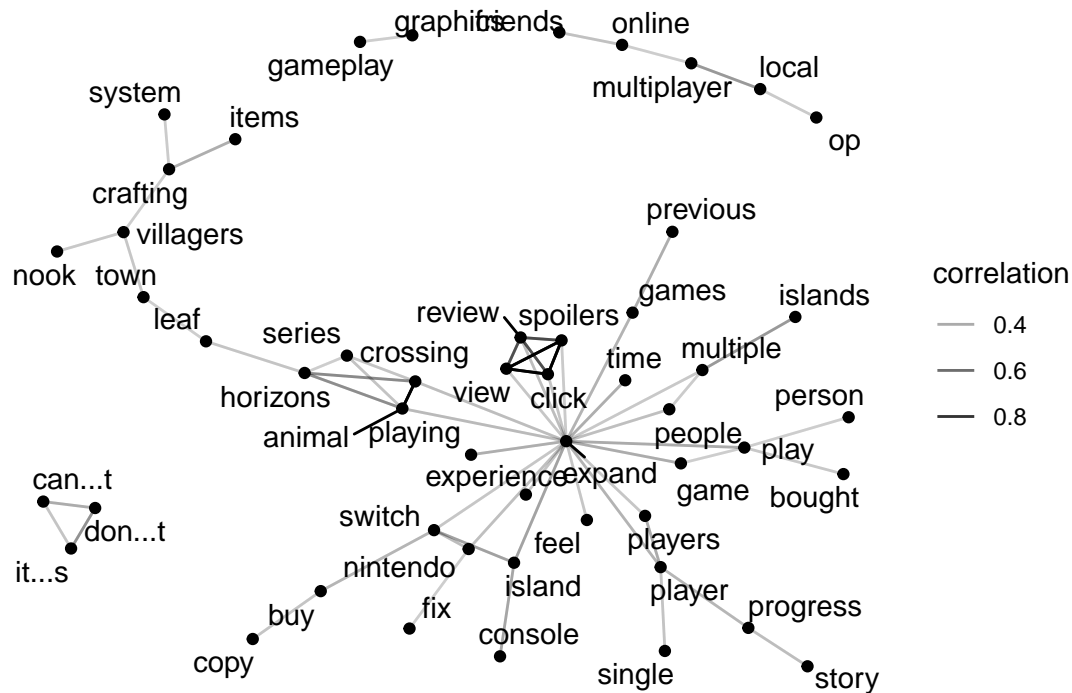
```
word_correlations <- review_words |>
  semi_join(users_who_mention_word, by = "word") |>
  pairwise_cor(item = word, feature = user_name) |>
  filter(correlation >= 0.2)
```

```
word_correlations
```

```
## # A tibble: 126 x 3
##   item1    item2 correlation
##   <chr>   <chr>         <dbl>
## 1 expand  playing      0.265
## 2 console island       0.282
## 3 expand  island       0.286
## 4 switch  island       0.292
## 5 expand  player       0.253
## 6 players player       0.272
## 7 progress player     0.247
## 8 single  player       0.227
## 9 island  console     0.282
## 10 switch buy         0.251
## # i 116 more rows
```

## A Word Network Plot

I then create a word network plot, including only the words that are correlated and excluding any words that appear by themselves. Darker lines show a higher word correlation.



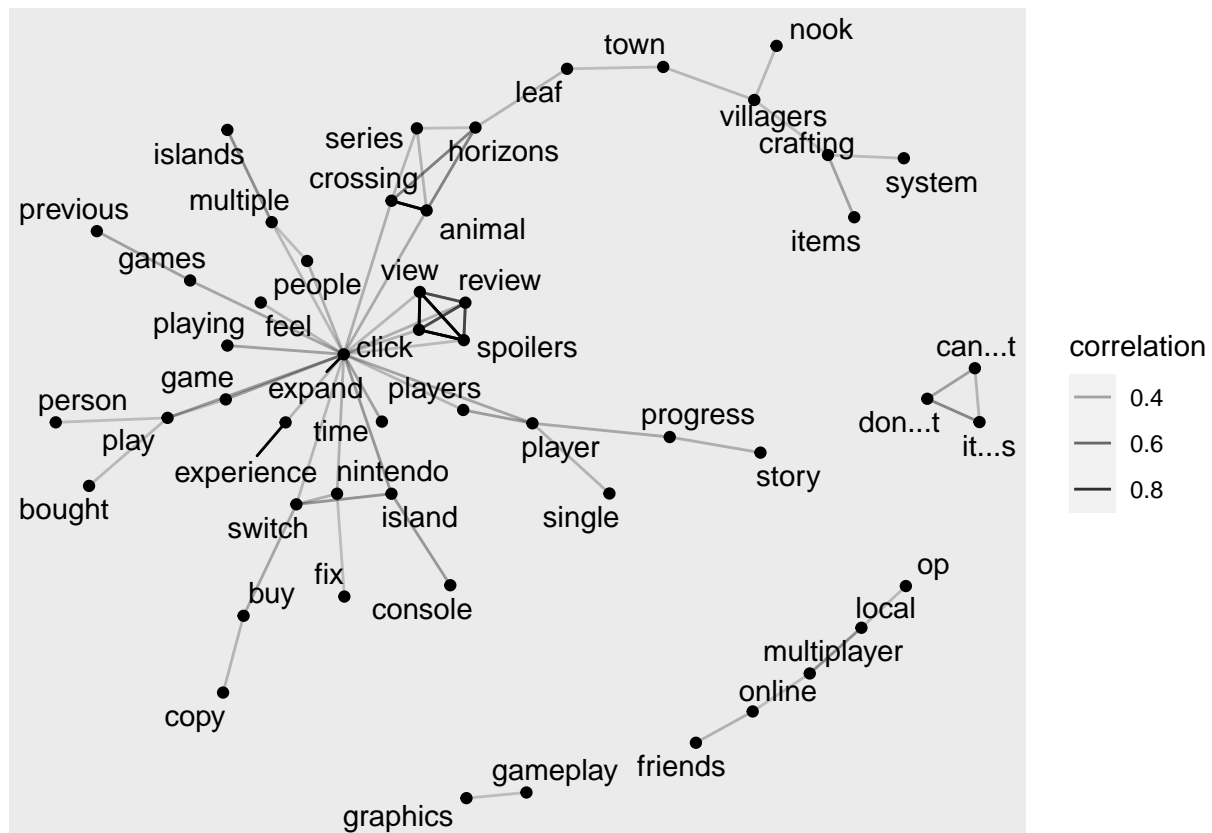
## Creating a function

```
generate_word_graph <- function (review_words,
                                minimum_users_n = 100,
                                minimum_correlation = 0.2){
  users_who_mention_word <- review_words |>
    dplyr::count(word, name = "users_n") |>
    filter(users_n >= minimum_users_n)

  word_correlations <- review_words |>
    semi_join(users_who_mention_word, by = "word") |>
    pairwise_cor(item = word, feature = user_name) |>
    filter(correlation >= minimum_correlation)

  graph_from_data_frame(d = word_correlations,
                        vertices = users_who_mention_word |>
                          semi_join(word_correlations, by = c("word" = "item1")) |>
                          ggraph(layout = "fr") +
                          geom_edge_link(aes(alpha = correlation)) +
                          geom_node_point() +
                          geom_node_text(aes(label = name), repel = TRUE)
  }

review_words |>
  generate_word_graph(
    minimum_users_n = 100,
    minimum_correlation = 0.2
  )
```



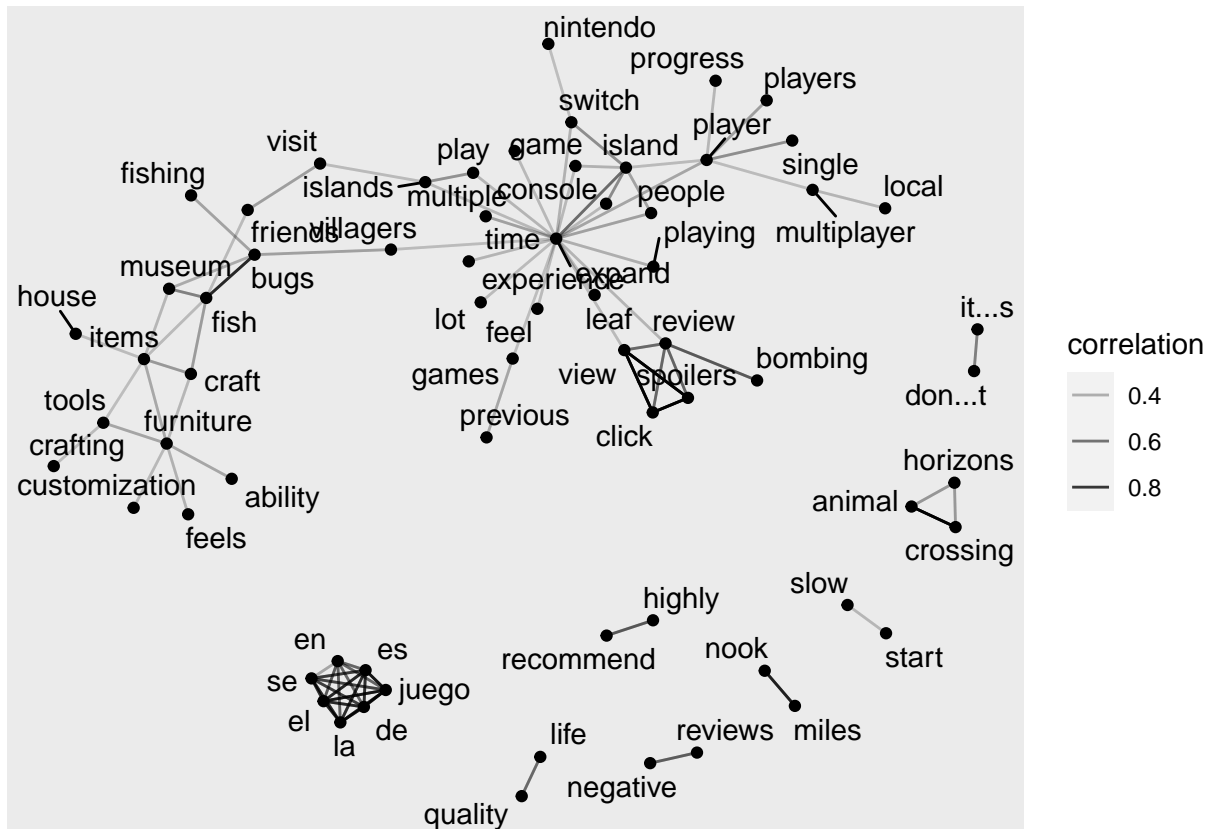
## Generating plots for positive and negative reviews

```
review_words_negative <- review_words |>
  filter(grade < 5)
```

```
review_words_positive <- review_words |>
  filter(grade >= 5)
```

```
review_words_negative |>
  generate_word_graph(
    minimum_users_n = 40,
    minimum_correlation = 0.2
  )
```





The positive reviews seem to mention review bombing often, a problem not uncommon in gaming and media more generally. Review bombing in gaming refers to a coordinated effort by a large number of individuals to leave negative reviews or ratings of a game on online platforms.

```
review_words |>
  distinct(user_name, .keep_all = TRUE) |>
  select(grade, user_name) |>
  filter (grade < 1) |>
  dplyr::count(grade)
```

```
## # A tibble: 1 x 2
##   grade     n
##   <dbl> <int>
## 1      0 1158
```

1158 users gave the game a grade of 0, which is often argued by the gaming community that it's a score that should solely be given if the game were actually broken. The review bombing seems to have been brought on by the developing team's decision to limit its users to one island per console, which is an obvious theme when looking at the graph for the negative reviews above.