

Analisi Recensioni Prodotto Amazon

Nintendo Switch

Emanuele Carta

05/07/2021

Sommario

1	<i>Introduzione</i>	3
1.2	<i>Descrizione del Prodotto</i>	3
2	<i>Descrizione dei Dati</i>	4
3	<i>Pulizia dei Dati</i>	4
4	<i>Analisi.....</i>	5
4.1	<i>Analisi Colonna RATING</i>	6
4.2	<i>Analisi Colonna DATE</i>	6
4.3	<i>Frequenza delle Parole</i>	7
4.4	<i>Topic Modelling</i>	9
4.5	<i>Sentiment Analysis</i>	11
4.6	<i>Analisi n-grammi</i>	20
5	<i>Conclusioni</i>	23
	<i>Fonti.....</i>	24

1 Introduzione

In questo breve elaborato andremo ad analizzare le recensioni di un prodotto in vendita su Amazon UK, sito sul quale è possibile trovare una vasta gamma di contenuti testuali, in lingua inglese, su cui poter testare le nostre conoscenze nel ambito del text mining.

L'oggetto del nostro studio sarà la Nintendo Switch, un prodotto di larghissima diffusione con oltre 13mila recensioni sul predetto sito Amazon. Si deve aggiungere doverosamente che l'analisi che è stata condotta ha potuto contare su un numero di gran lunga inferiore di recensioni, pari a 500.

Inizialmente andremo a descrivere in modo generale il prodotto, per avere un quadro generale su di esso. Seguirà un'analisi di carattere generale sulle classificazioni delle recensioni, per poi scendere nel dettaglio, evidenziando i principali punti di forza e di debolezza del prodotto.

Pur evidenziando la necessità di ulteriori approfondimenti, si è evidenziato un aumento del numero delle recensioni in un particolare segmento temporale, pari a un anno circa.

1.2 Descrizione del Prodotto

Nintendo Switch è una console di gioco ibrida ideata da Nintendo e rilasciata al pubblico il 3 Marzo 2017.

Da un punto di vista generale è un prodotto unico nel suo genere, poiché può essere utilizzato sia come una piattaforma di gioco fissa, ma anche come un dispositivo portatile. Questa sua caratteristica la rende un prodotto molto più flessibile rispetto alle sue dirette concorrenti, Ps4 e Xbox.

Quasi tre milioni di unità sono state spedite nel suo primo mese, superando la proiezione iniziale di Nintendo di due milioni ed entro un anno dal rilascio ha raggiunto oltre 14 milioni di unità vendute in tutto il mondo.

All'inizio del 2018, è diventata la console domestica e portatile più venduta sia in Giappone che negli Stati Uniti. A marzo 2021 ha venduto più di 84 milioni di unità in tutto il mondo. Le vendite della Nintendo Switch sono state fortemente legate alle vendite dei titoli proprietari di Nintendo, come: "The Legend of Zelda: Breath of the Wild", "Mario Kart 8 Deluxe", "Super Mario Odyssey", "Super Smash Bros. Ultimate", "Pokémon Sword and Shield e Animal Crossing: New Horizons", ognuno di essi ha venduto oltre venti milioni di unità ciascuno.

Nintendo è Switch composta da un'unità console, un dock e due controller Joy-Con.

La console consiste in uno schermo LCD che misura 6,2 pollici (160 mm). Lo schermo supporta il rilevamento capacitivo multi-touch a dieci punti e include la tecnologia tattile di Immersion Corporation . Lo schermo LCD supporta risoluzioni fino a 720p (1280×720 px).

La console, con o senza Joy-Con collegati, può essere posizionata nella docking station che permette di caricare la batteria e di collegarla a un televisore tramite una connessione HDMI.

Quando è collegata, l'unità può supportare risoluzioni fino a 1080p e una frequenza fotogrammi massima di 60 fotogrammi al secondo , anche se la risoluzione massima varia a seconda del gioco.

I controller , vengono chiamati Joy-Con. Questi ultimi, si collegano alla console Switch tramite guide laterali utilizzando un meccanismo di blocco con un piccolo pulsante di rilascio sul lato posteriore per consentire il distacco. Quando sono staccati, possono essere usati in coppia da un singolo giocatore o usati come controller separati da due giocatori individuali. In alternativa possono essere attaccati a un'impugnatura che emula un gamepad.

2 Descrizione dei Dati

L'estrazione di questi dati risale al 22/05/2021.

Andiamo adesso a descrivere come si presenta il nostro dataset di recensioni e cosa di preciso è contenuto al suo interno.

Il nome del nostro dataset è **nintendo_r**, e al suo interno sono contenute **5** colonne con **500** righe ciascuna.

```
> nintendo_r
# A tibble: 500 x 5
  RATING TITLE VERIFIED TEXT DATE
  <chr> <chr> <chr> <chr> <chr>
1 3.0 out ... A much needed Criti... NA "Im not sure how popula... Reviewed in th...
2 5.0 out ... Fantastic Console T... TOP 500 REVIEW... "Your browser does not ... Reviewed in th...
3 5.0 out ... Newest model/versio... NA "[Update Oct 2020]Appro... Reviewed in th...
4 1.0 out ... I wan Refund please... Verified Purch... "Your browser does not ... Reviewed in th...
5 1.0 out ... Not NEW Verified Purch... "Sold as NEW but damage... Reviewed in th...
6 5.0 out ... Fun in a box! Verified Purch... "Where to begin?This co... Reviewed in th...
7 5.0 out ... New Switch/Switch V... Verified Purch... "Great console, its the... Reviewed in th...
8 4.0 out ... Battery improved, t... TOP 500 REVIEW... "The Switch is an excel... Reviewed in th...
9 5.0 out ... Fantastic, would ha... Verified Purch... "Your browser does not ... Reviewed in th...
10 5.0 out ... Finally handheld ga... Verified Purch... "I have had many of the... Reviewed in th...
# ... with 490 more rows
```

Partiamo a descrivere dalla prima colonna a sinistra.

Questa colonna ha il nome **RATING** e in essa sono contenute le stelle date in base alla recensione. La struttura del testo è *"n.0 out of 5 stars"*, dove n è il numero di stelle, dato che varia ad ogni recensione.

Proseguendo troviamo la colonna **TITLE**: in questa colonna sono contenuti i titoli in base alla recensione, la forma del testo è variabile, ma in media troviamo sempre testi contenenti un numero basso di parole.

Dopo la colonna dei titoli troviamo **VERIFIED**: in questa colonna viene detto se la recensione è verificata, cioè se l'acquisto da parte del cliente è stato effettuato o no. La struttura del testo è quasi sempre la stessa, infatti nella maggior parte delle righe troviamo la dicitura: *"Verified Purchase"*; in altre troviamo la stessa, ma preceduta da *"TOP 500 REVIEWER|"* o *"TOP 1000 REVIEWER|"*.

Troviamo poi la colonna **TEXT** dove sono contenuti i vari testi delle recensioni. Come detto anche per i titoli, la lunghezza è variabile, ma la differenza tra i due è che in TEXT si possono trovare testi con svariate decine di parole.

L'ultima colonna che troviamo è **DATE**: in questa colonna sono presenti 2 informazioni: la prima è il luogo dove è stata effettuata la recensione e la seconda è la data. Troviamo una struttura fissa: *"Reviewed in the United Kingdom on Day Month Year"*, dove Day (scritto in valore numerico), Month (scritto a lettere), Year (scritto in valore numerico) variano a seconda della recensione.

3 Pulizia dei Dati

Dopo aver visualizzato il dataset a nostra disposizione, notiamo che nelle colonne **RATING** e **DATE** abbiamo una struttura e un testo extra che non vogliamo includere nella nostra analisi. Ad esempio, nella colonna **RATING**, è presente sempre la dicitura: *"n.0 out of 5 stars"*. Per far sì che al nostro occhio l'informazione sia più diretta, decidiamo di togliere il segmento in mezzo alla stringa, facendo in modo che a ogni riga della colonna vi sia scritto: *"n stars"*.

Nella colonna **DATE** la situazione è diversa, la pulizia che andremo a fare infatti ci servirà per eliminare elementi che disturbano l'analisi che faremo in seguito. Andremo quindi a rimuovere la località da cui è stata scritta la recensione, poiché notiamo che è sempre la medesima e rimuoviamo anche il giorno e il mese.

Questo tipo di pulizia generale sulle colonne può essere eseguito all'interno del pacchetto **textclean** utilizzando **mgsb()** per sostituire a una determinata stringa un'altra stringa a nostro piacimento.

```

delR <- c('.0 out of 5')
delD <- c('Reviewed in the United Kingdom on')
delDM <- mesi anno in inglese

n$RATING <- mgsub(pattern = delR, replacement = ' ', n$RATING)
n$DATE <- mgsub(pattern = delD, replacement = ' ', n$DATE)
n$DATE <- mgsub(pattern = delDM, replacement = '_', n$DATE)
n$DATE <- gsub('^.?*?_','_',n$DATE)
n$DATE <- mgsub(pattern = '_', replacement = '', n$DATE)

```

Ora che abbiamo cambiato la struttura e eliminato il testo extra di alcune colonne del nostro dataset, andiamo ad assegnare a ogni recensione un ID che ci servirà in seguito durante l'analisi. Fatto ciò, creiamo dei dataset specifici, prendendo solo le colonne che ci interessa analizzare.

```

#ASSEGNO AD OGNI RECENSIONE UN ID DA 1-500
n$id = (1:500)

#CREO UN DATASET FATTO SOLO DALLE RECENSIONI E DEI TITOLI
text <- n$TEXT
text <- as.data.frame(text)
title <- n$TITLE
title <- as.data.frame(title)

#CREO UN DATASET FATTO DAL RATING, DAL TESTO, ANNO, ED IL LORO ID
r <- n$RATING
t <- n$TEXT
d <- n$DATE
id <- n$id
rt <- data.frame(id,r,d,t)
colnames(rt) <- c('id','RATING','DATE','TEXT')

#CREO UN DATASET FATTO DAL RATING, DAL TITOLO, ANNO ED IL LORO ID
tt <- n$TITLE
rtt <- data.frame(id,r,d,tt)
colnames(rtt) <- c('id','RATING','DATE','TITLE')

```

Ancora i nostri dati però non sono pronti all'analisi.

Dobbiamo convertire la colonna **TEXT** in modo che per ogni riga ci sia un token (parola). Per far ciò, dobbiamo effettuare la tokenizzazione, cioè il processo di suddivisione del testo in token.

Proseguiamo quindi anche alla rimozione delle stop-words.

Per effettuare la tokenizzazione usiamo la funzione `unnest_tokens()` di *tidytext*; invece per eliminare le stop-words la funzione `str_detect()` di *stringr*.

```

#TOKENIZZAZIONE E PULIZIA DELLE RECENSIONI
rev_words <- rt %>%
  unnest_tokens(word, TEXT) %>%
  filter(str_detect(word, "[a-z']$"),
         !word %in% stop_words$word)

#TOKENIZZAZIONE E PULIZIA DEI TITOLI DELLE RECENSIONI
rev_words <- rtt %>%
  unnest_tokens(word, TITLE) %>%
  filter(str_detect(word, "[a-z']$"),
         !word %in% stop_words$word)

```

4 Analisi

La prima analisi che faremo, sarà uno studio generico del dataset. In particolare, ci concentreremo sulla colonna **RATING** e sulla colonna **DATE**, per iniziare a capire cosa ci aspetterà durante lo studio più approfondito.

4.1 *Analisi Colonna RATING*

Per analizzare questa colonna ci è bastato trovare la frequenza del numero delle stelle in base al numero delle recensioni .

Rendendo i dati che abbiamo trovato sotto forma di barplot salta subito all'occhio che su le **500** recensioni che avevamo nel nostro dataset, la maggior parte sono **5** stelle. Ovviamente questo dato non ci sorprende poiché la Nintendo Switch, come detto nel paragrafo introduttivo del prodotto, è una delle console di maggior successo nel mercato attuale.

Questa prima analisi che abbiamo effettuato ci fa intuire che andando a studiare più approfonditamente le recensioni troveremo, forse, più parole positive che negative nei confronti del prodotto.

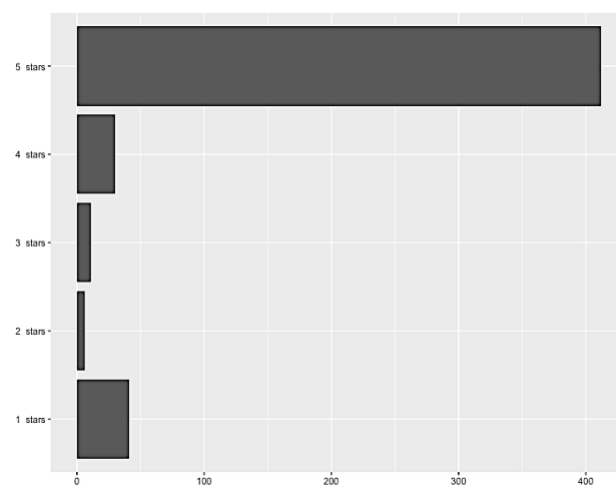


Figura 1: Grafico delle Stelle

4.2 *Analisi Colonna DATE*

Anche per questa colonna, l'analisi è stata eseguita facendo la frequenza degli anni in base al numero delle recensioni. Dal grafico si capisce subito che la maggior parte delle recensioni è stata effettuata durante il **2020**. Guardando la differenza sostanziale che vi è tra il numero di recensioni effettuate nel **2020** e nel **2019**, potremmo ipotizzare che durante il **2020** vi siano stati più acquisti e quindi anche più recensioni in confronto dell'anno precedente.

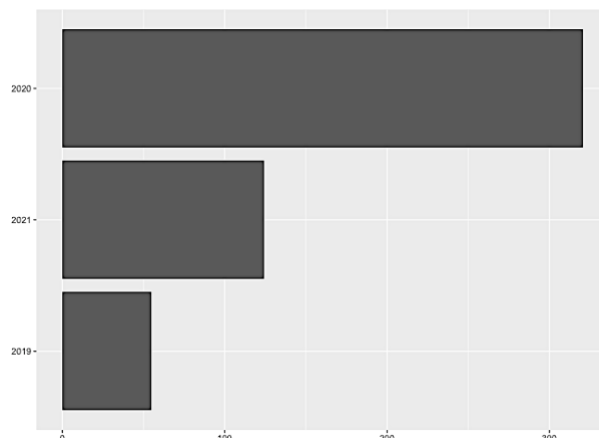


Figura 2: Grafico delle Recensioni in Base agli Anni

Per capire se la nostra ipotesi potesse essere reale abbiamo cercato nel WEB informazioni riguardo un particolare incremento delle vendite durante il **2020** del nostro prodotto.

BBC Sign in Home News Sport Reel Worklife Travel Future Culture More Search

NEWS

Home Coronavirus Video World UK Business Tech Science Stories Entertainment & Arts Health More


Business Market Data New Economy New Tech Economy Companies Entrepreneurship Technology of Business Global Education Economy Global Car Industry Business of Sport

Covid-19: Nintendo profits triple as games boom continues

By Tom Espiner
Business reporter, BBC News

5 November 2020 | Comments

Coronavirus pandemic



Top Stories

Ethiopia rebels seize capital of devastated region
Supporters celebrate in Mekelle as officials flee the city and the government calls for a ceasefire.
2 hours ago

What could have caused Miami building collapse?
44 minutes ago

South Africa ex-President Zuma sentenced to jail
2 hours ago

Features

S&P Global Market Intelligence Who We Serve Solutions News & Insights Events Product Login Request Demo

6 Aug, 2020

IN THIS LIST

- Nintendo profits up over 400% as COVID-19 drives game sales
- VIDEO State of the Market – Mining Q1-2021
- BLOG State of the Market: Mining Q1-2021
- BLOG Earnings & Results Roundtable: How banks have outperformed during the COVID-19 pandemic
- BLOG Q1: U.S. Solar and Wind Power by the Numbers

Nintendo profits up over 400% as COVID-19 drives game sales

Author: Anser Haider

Nintendo Co. Ltd.'s profits surged more than 400% in its fiscal first quarter ended June 30, as demand for the company's Switch console and Animal Crossing video game skyrocketed amid the COVID-19 pandemic.

The Japanese video game company shipped more than 5.7 million Switch consoles in the quarter, up 166.6% year over year. The shipments included about 3 million of the original Switch device and 2.6 million units of the portable-only Switch Lite model. As of June 30, lifetime-to-date sales of the Switch consoles reached 61.44 million units worldwide.

Come si evince da gli articoli, sopra riportati, la nostra ipotesi era esatta: troviamo che le vendite della Nintendo Switch durante l'anno **2020** hanno avuto un boom aiutato anche dalla pandemia.

4.3 *Frequenza delle Parole*

Ripartendo del nostro dataset di parole tokenizzate e pulite, andiamo a cercare le parole che hanno una frequenza maggiore nelle recensioni.

```
#FREQUENZA PAROLE NELLE RECENSIONE
freq <- rev_words %>%
  count(word, sort = TRUE)
```

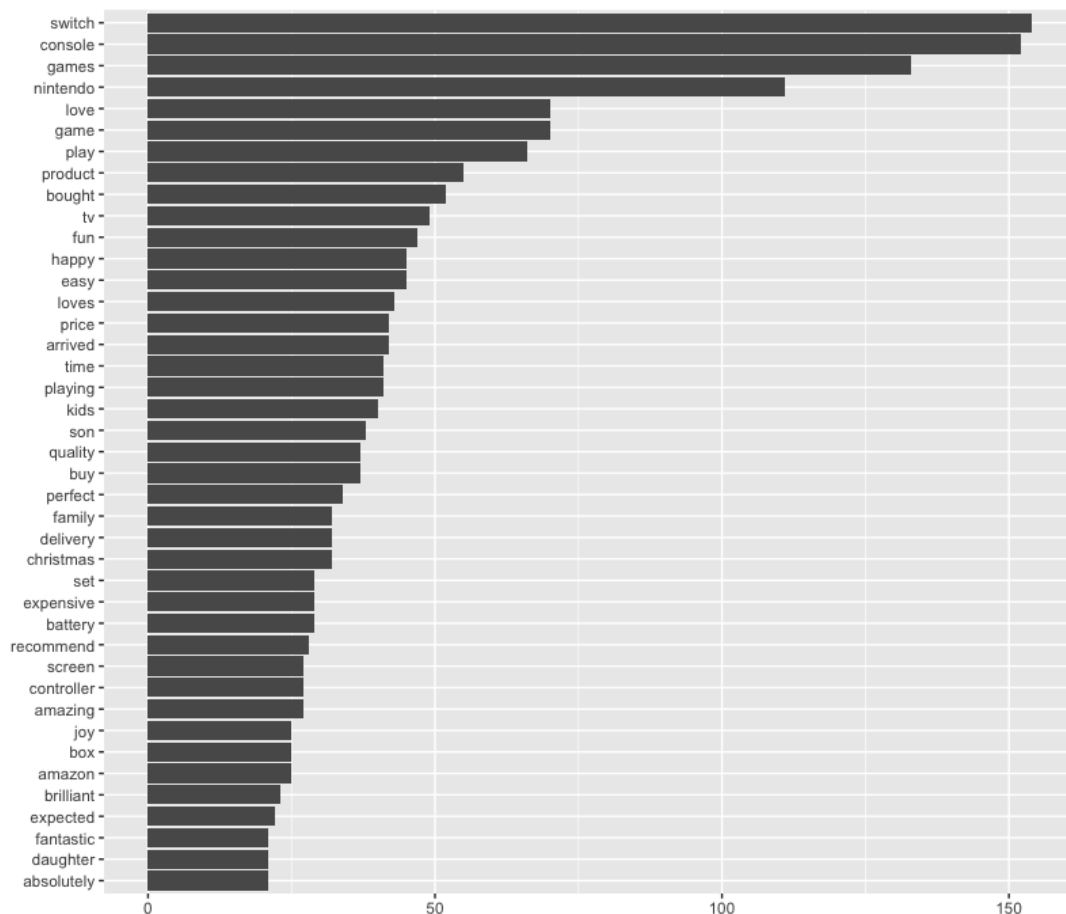


Figura 3: Grafico della Frequenza delle Parole nei Testi

Rileviamo subito che tra i primi posti delle parole più usate ci sono: “switch”, “console”, “games” e “nintendo”. Ovviamente c’era da aspettarselo poiché andando ad analizzare le recensioni di uno specifico prodotto, sarà sempre lui il soggetto sia delle recensioni buone che di quelle cattive.

Continuando a scorrere tra le parole più utilizzate, iniziamo a notare che un gran numero sono aggettivi per lo più positivi come: “amazing”, “fun”, “brilliant”, “fantastic”. Come detto in precedenza, nell’analisi preliminare, essendo per lo più recensioni con 5 stelle, è normale che tra le parole con una frequenza alta siano presenti parole positive riguardo alla console.

Infine possiamo notare un altro sottogruppo di parole che appartiene allo stesso campo semantico, cioè: “son”, “kid”, “family”, “daughter”. Essendo Nintendo Switch una console adatta a un pubblico molto vasto, grazie alla varietà dei titoli presenti nello store aventi PEGI¹ dai 3 ai 18 anni, rende così questo prodotto usufruibile da ogni singolo componente della famiglia.

[1] PEGI: fornisce una classificazione dei videogiochi in base all’età in 38 paesi europei. La classificazione conferma che il gioco è adeguato agli utenti di una determinata età.

Continuando ad analizzare la frequenza delle parole siamo andati ad isolare quelle in base all’anno e alle stelle.

```
#FREQUENZA DELLE PAROLE DELLE RECENSIONI IN BASE ALLE STELLE
words_by_stars <- rev_words %>%
  count(RATING, word, sort = TRUE) %>%
  ungroup()
```

Dalla singola analisi di **words_by_star** non vi sono risultati degni di nota; d’altro canto nell’analisi della frequenza delle parole in base all’anno, emerge un dato molto importante.

	DATE	word	n
295	2020	lockdown	4
320	2021	lockdown	4

Andando ad esaminare **words_by_year** alle righe **295** e **320** troveremo la parola lockdown. Questa parola non ha niente a che vedere con il linguaggio videoludico e nessun tipo di caratteristica tecnica. Il ritrovamento di questa parola avvalorava ancora di più la tesi descritta in precedenza nell'analisi preliminare. Infine andiamo ad analizzare la frequenza delle parole nei titoli delle recensioni.

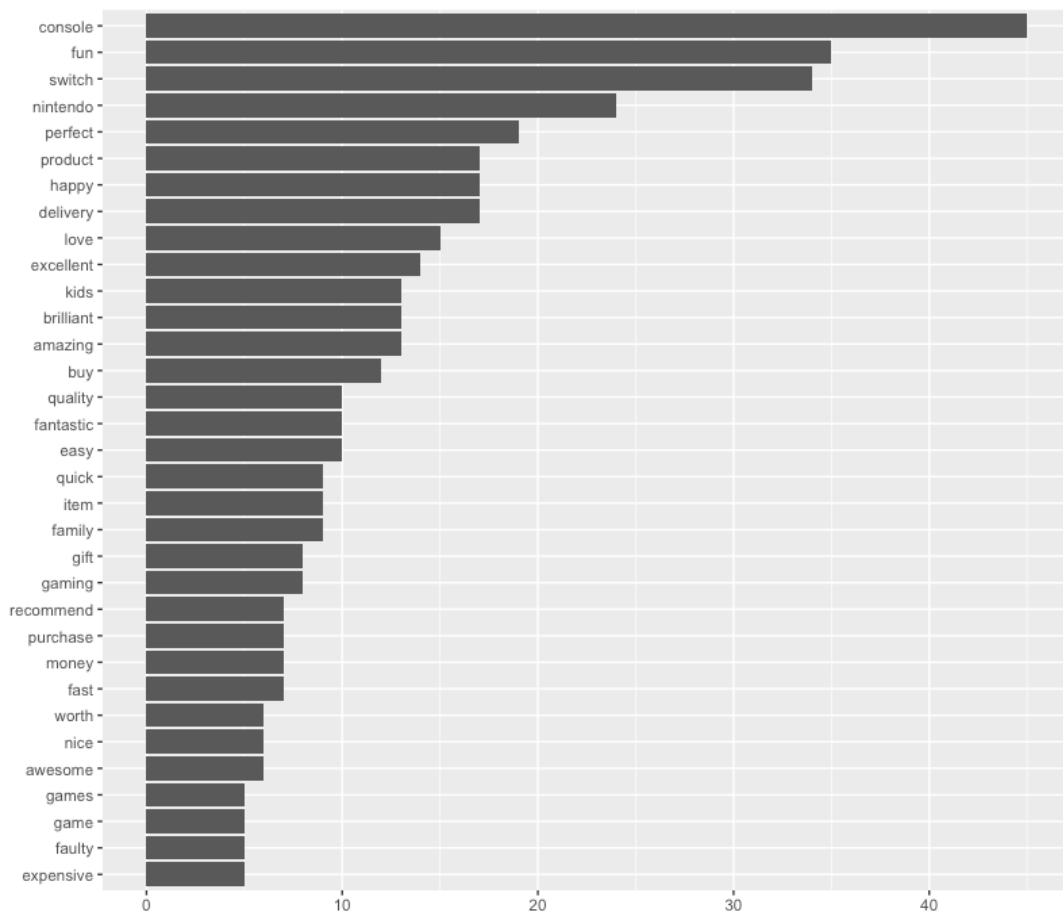


Figura 4: Grafico della Frequenza delle Parole nei Titoli delle Recensioni

4.4 Topic Modelling

La modellazione degli argomenti è un metodo per la classificazione dei topic presenti in un documento. Prima di tutto trasformiamo il nostro dataset tokenizzato e pulito in una Document-Term Matrix tramite la funzione `cast_dtm()`, appartenente alla libreria `tidytext`.

Dopodiché utilizziamo la funzione `LDA()` della libreria `topicmodels` con **k = 6**, per creare un modello LDA con 6 argomenti.

Questa funzione restituisce un oggetto contenente tutti i dettagli dell'adattamento del modello, per esempio, le parole che sono associate agli argomenti.

```
#CONVERTIAMO I NOSTRI DATI IN UNA DTM
rev_dtm <- rev_words %>%
  unite(document, RATING, id) %>%
  count(document, word) %>%
  cast_dtm(document, word, n)

#LDA
rev_lda <- LDA(rev_dtm, k = 6, control = list(seed = 2410))
#FREQUENZA DELLE PAROLE DELLE RECENSIONI IN BASE ALL'ANNO
words_by_year <- rev_words %>%
  count(DATE, word, sort = TRUE) %>%
  ungroup()
```

Per trasformare una Document-Term Matrix in una tabella di dati ordinata usiamo la funzione `tidy()`. Il pacchetto *tidytext* fornisce questo metodo per estrarre le probabilità *per-topic-per-word*, chiamata β . Ciò che abbiamo fatto è stato trasformare il modello in un formato *one-topic-per-term-per-row*. Per ogni combinazione, il modello calcola la probabilità che quel termine venga generato da quell'argomento. In seguito usiamo la funzione `slice_max()`, contenuta nella libreria *dplyr*, per trovare i **10** termini più comuni all'interno di ciascun argomento.

```
#ORDINIAMO DTM
rev_topics <- tidy(rev_lda, matrix = "beta")

#TOP 10 TERMINI PIÙ COMUNI
rev_top_terms <- rev_topics %>%
  group_by(topic) %>%
  slice_max(beta, n = 10) %>%
  ungroup() %>%
  arrange(topic, -beta)
```

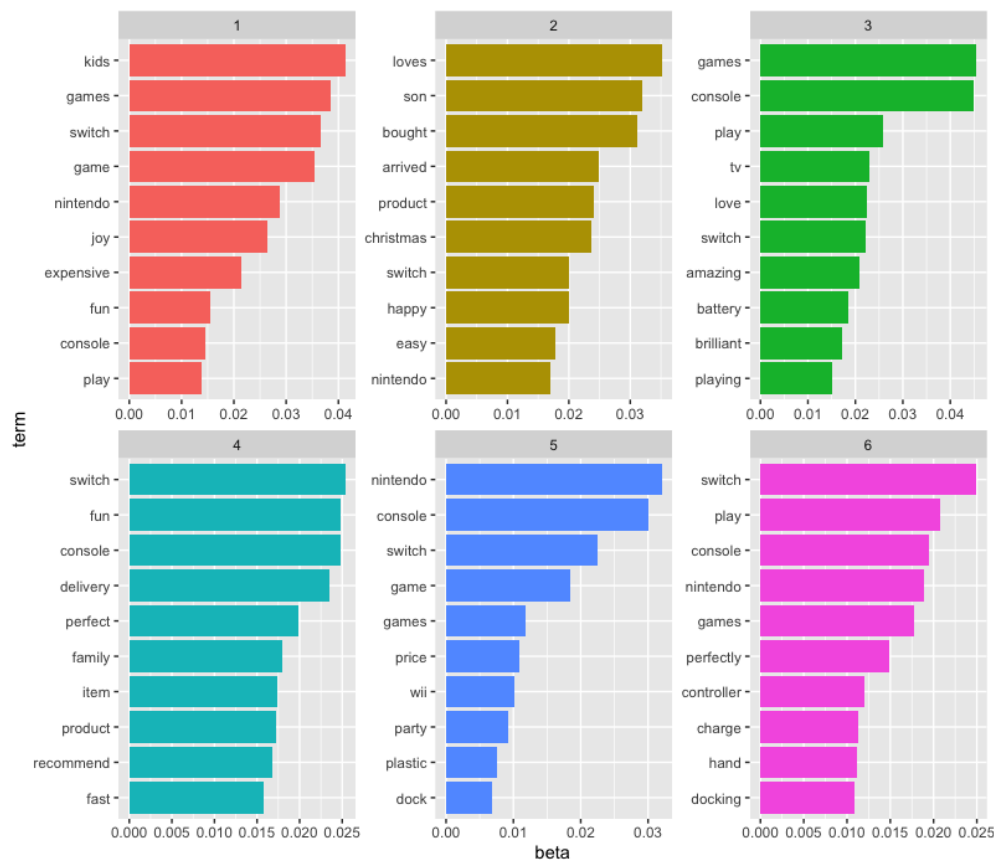


Figura 5: Topic Modelling

Questa visualizzazione permette di comprendere i vari argomenti delle recensioni. Per esempio nel 2° argomento le parole più comuni sono: “loves”, “son”, “bought”, “arrived”, “christmas”; questo suggerisce che probabilmente vari clienti abbiano acquistato la Nintendo Switch per il proprio figlio per Natale e gli sia arrivata in tempo per l’occasione.

Nel 3° argomento, le parole più comuni sono: “games”, “console”, “play”, “tv”, “love”. La prima cosa che viene in mente, leggendo queste parole e conoscendo il prodotto, è che la grande versatilità che la console ha, potendo essere giocata sia su televisore che anche come dispositivo portatile sia stato un argomento di discussione.

Mediante tutti gli argomenti che abbiamo individuato sono indirizzati verso l’approvazione del prodotto, descrivendolo in modo positivo. La cosa non ci stupisce poiché, come abbiamo visto, sia nell’analisi preliminare sia nello studio delle frequenze, le parole positive verso la Nintendo Switch e le recensioni ottime non mancano.

4.5 Sentiment Analysis

Possiamo utilizzare le tecniche di analisi del sentiment per esaminare la frequenza con cui appaiono parole positive e negative nelle recensioni e nei loro titoli.

Il primo studio che andremo a fare sarà quello di analizzare le parole più o meno positive in base alle stelle. Per questa analisi utilizzeremo il lessico dei sentimenti **AFINN**, che fornisce valori di positività numerica per ogni parola e lo visualizzeremo con un grafico a barre.

Per eseguire l'analisi del sentiment utilizzeremo la funzione `inner_join()` della libreria *dplyr*.

```
#ANALISI DELLE PAROLE PIU' O MENO NEGATIVE IN BASE ALLE STELLE
#RECENSIONI
rating_sentiments <- words_by_stars %>%
  inner_join(get_sentiments("afinn"), by = "word") %>%
  group_by(RATING) %>%
  summarize(value = sum(value * n) / sum(n))

#TITOLI
Per l'analisi dei titoli è stata usata la medesima riga di codice usata per le
recensioni, cambiando ovviamente i dati.
```

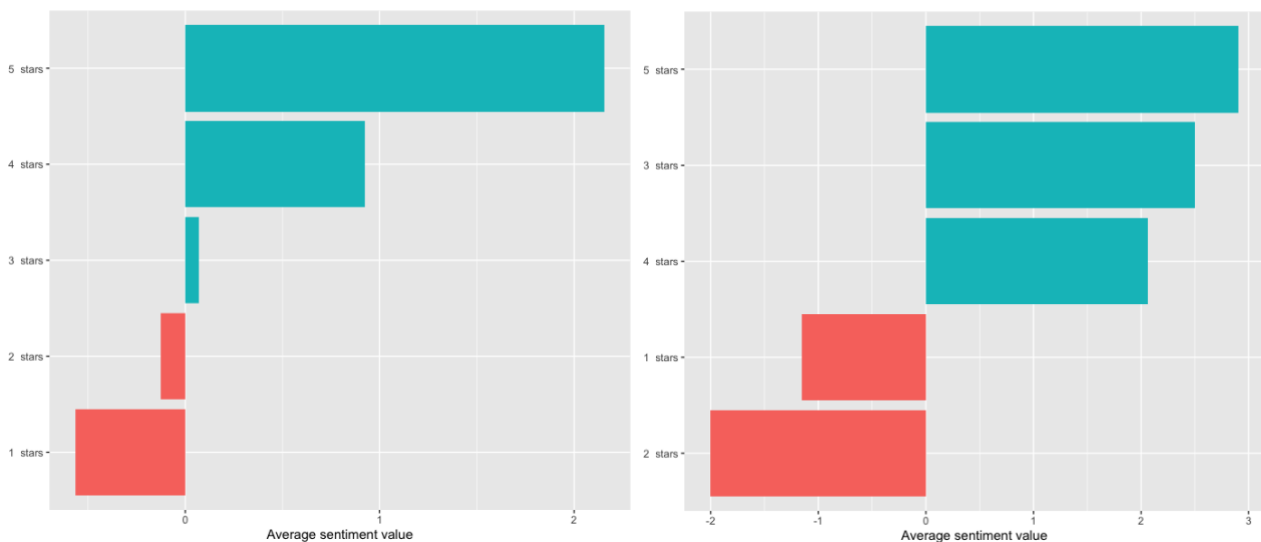


Figura 6: Grafico sulle Stelle (RECENSIONI) (Sinistra), Grafico sulle Stelle (TITOLI) (Destra)

Nel primo grafico a sinistra dove viene espresso il sentiment dei testi delle recensioni in base alle stelle, notiamo che la situazione sembra normale. La positività delle parole espresse nelle recensioni tende ad essere più alta sulle recensioni a 5 stelle, che gradualmente scende fino a diventare negativa nelle recensioni a 1 e 2 stelle.

Troviamo una situazione simile nel grafico a destra, dove viene descritto il sentiment dei titoli delle recensioni in base alle stelle. Ciò che cambia è il valore del sentiment che si alza e anche la posizione delle stelle, che se prima la trovavamo in ordine decrescente, adesso la troviamo mista. Queste differenze non cambiano però il risultato dell'analisi ottenuta sui testi delle recensioni, infatti i titoli delle recensioni con 5, 4 e 3 stelle rimangono positivi.

Per capire per quale motivo alcuni gruppi di recensioni siano risultati più positivi o negativi di altri, andiamo a esaminare i contributi positivi e negativi totali di ogni parola in modo generale, per poi incentrare l'analisi sulle stelle.

```
#ANALISI DEL SENTIMENT PAROLA PER PAROLA
#RECENSIONI
contributions <- rev_words %>%
  inner_join(get_sentiments("afinn"), by = "word") %>%
  group_by(word) %>%
  summarize(occurrences = n(),
    contribution = sum(value))
```

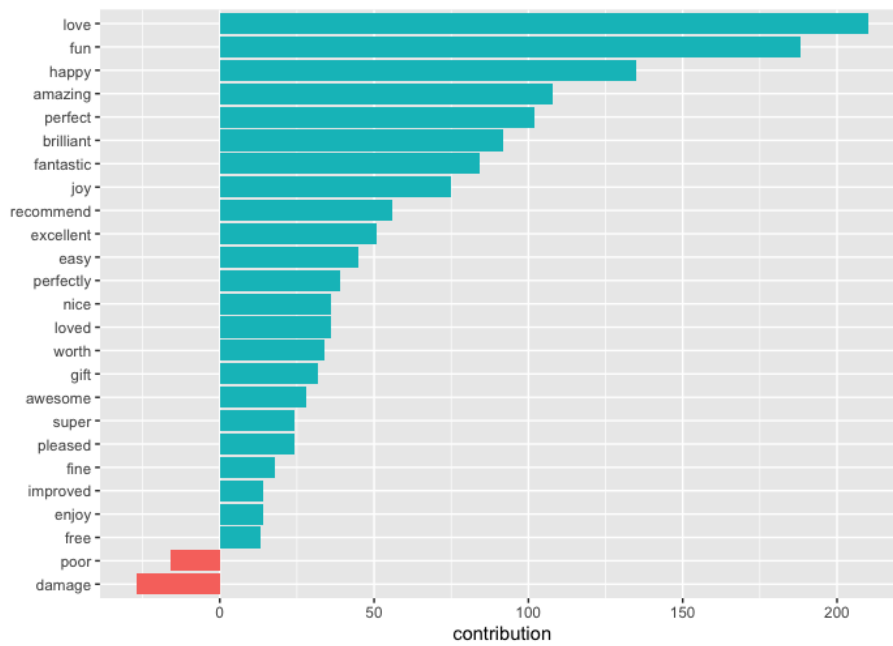


Figura 7: Grafico Parole che Contribuiscono al Sentiment(RECENSIONI)

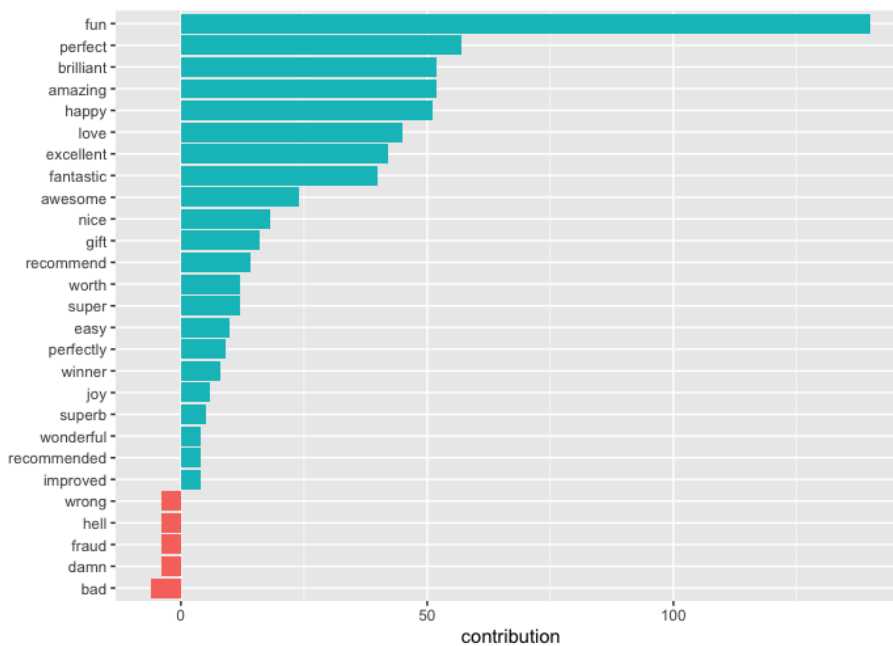


Figura 8: Grafico Parole che Contribuiscono al Sentiment(TITOLI)

Come possiamo vedere, la maggior parte delle parole usate sono principalmente positive, sia nel grafico sopra, che rappresenta le parole nelle recensioni, sia nel grafico sotto, che rappresenta le parole nei titoli delle recensioni.

In alcune occasioni potremo incontrare degli errori, infatti se andiamo a vedere le parole utilizzate molte di essere potrebbero essere usate anche in senso negativo. Per esempio *“happy”*, possiamo usare questa parola in modo positivo come: *“I was happy to receive this product”* (sono felice di aver ricevuto questo prodotto), ma possiamo usarla anche in modo negativo come: *“I am not happy about this product”* (non sono felice di questo prodotto).

Risolveremo questo problema nelle analisi successive.

Studiamo ora quali parole hanno contribuito maggiormente all'interno di ciascun gruppo di stelle , in modo da poter vedere quali gruppi potrebbero essere stimati in modo errato. Andiamo quindi a calcolare il contributo di ogni parola al sentiment in ogni gruppo di stelle.

```
#PAROLE CON MAGGIOR EFFETTO SUL SENTIMENT IN OGNI SINGOLO RATING
#RECENSIONI
top_sentiment_words <- words_by_stars %>%
  inner_join(get_sentiments("afinn"), by = "word") %>%
  mutate(contribution = value * n / sum(n))
```

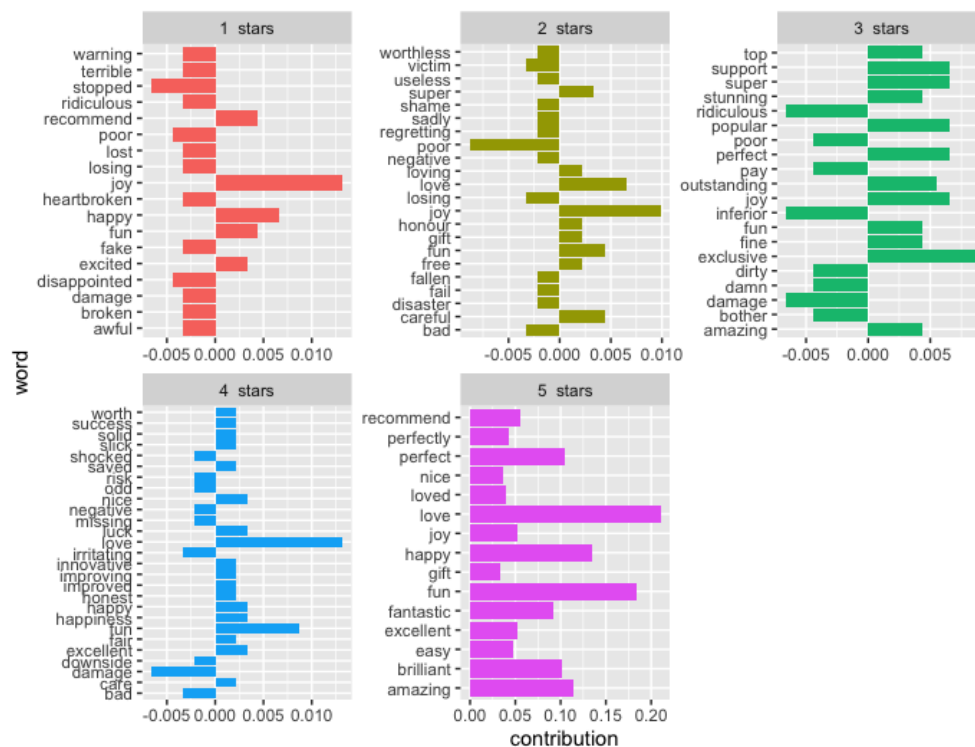


Figura 9: Grafico Parole che Contribuiscono al Sentiment nei Vari Gruppi di Stelle(RECENSIONI)

Come possiamo vedere lo studio che abbiamo effettuato delle recensioni nei diversi gruppi di stelle è risultato corretto ed equilibrato. Non vi sono stati contaminazioni di parole troppo positive che hanno modificato il risultato finale di ogni gruppo di stelle.

Andando adesso ad analizzare lo stesso tipo di studio che abbiamo fatto sulle recensioni, ma sul titolo.

Prima avevamo notato che, nel sentiment diviso per gruppi di stelle, nei titoli le posizioni delle stelle cambiavano.

Avendo ora questo grafico sotto mano, possiamo capire il perché.

In effetti, i gruppi che cambiavano posizione erano proprio i gruppi delle 2 e 3 stelle.

Come notiamo, l'analisi su questi ultimi non è molto accurata, poiché le parole positive e negative sono di gran lunga minori a quelle degli altri gruppi, arrivando ad essere solo positive o solo negative.

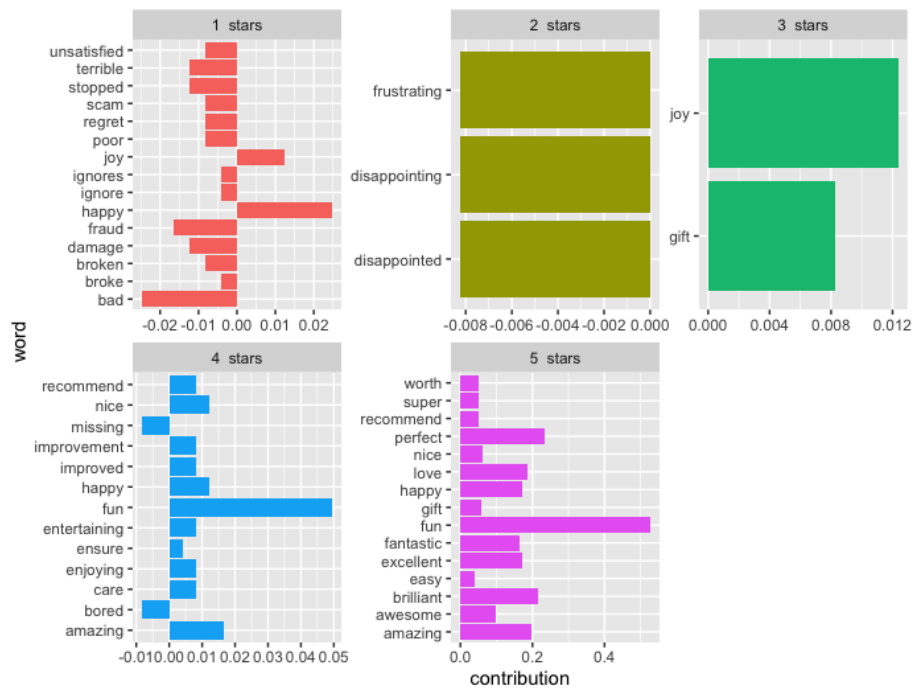


Figura 10: Grafico Parole che Contribuiscono al Sentiment nei Vari Gruppi di Stelle(RECENSIONI)

Passiamo ora ad analizzare ogni singola recensione e ogni singolo titolo. Per effettuare questa analisi, abbiamo usato due tipi di studio della sentiment: il primo è quello che abbiamo usato fino ad adesso della libreria *tidytext*, il secondo che andremo ad usare è la funzione `sentiment_by()` della libreria *sentimentr*. Con l'utilizzo di `sentiment_by()` siamo andati a studiare il sentiment delle singole recensioni e dei singoli titoli con il numero di parole contenute nel testo indefinito.

```
#SENTIMENT GENERALE SU RECENSIONI DI OGNI DIMENSIONE
#RECENSIONI
text_dtm <- text %>%
  as.matrix()
sentiment_text <- text_dtm %>%
  sentiment_by()
```

```
> sentiment_text %>%
+   arrange(desc(ave_sentiment))
  element_id word_count      sd ave_sentiment
1:         283         5      NA      1.7441330
2:         306         7      NA      1.7197384
3:         259         6      NA      1.5717559
4:         260         5      NA      1.5652476
5:         480         3      NA      1.5011107
...
496:        344         5 0.006186218 -0.4286384
497:        393         1      NA -0.5000000
498:        402         1      NA -0.5000000
499:        125        20      NA -0.5170907
500:        425        12 0.058946609 -0.5416815
```

```
> sentiment_title %>%
+   arrange(desc(ave_sentiment))
  element_id word_count      sd ave_sentiment
1:         260         5      NA      1.565248
2:         444         5      NA      1.498166
3:          62         5      NA      1.448972
4:        136         2      NA      1.414214
5:          65         8      NA      1.396536
...
496:        263         1      NA     -0.750000
497:        314         1      NA     -0.750000
498:         40         1      NA     -1.000000
499:        134         1      NA     -1.000000
500:        425         4      NA     -1.125000
```

Da questi dati dunque possiamo individuare le recensioni e di titoli siano stati i più positivi e più negativi. Andiamo a vederli un po' più nel dettaglio.
La recensione più positiva è la **n.283**, leggiamola:

	RATING	TITLE	TEXT
283	5 Stars	Clever bit of kit	Very good and very clever

Il testo di questa recensione sembra ottimo, in **5** parole **2** sono aggettivi molto positivi.
Estraiamo anche le altre due recensioni presenti sul podio delle più positive:

	RATING	TITLE	TEXT
306	5 Stars	great console	very good console very pleased with it
259	5 Stars	Excellent, highly recommend	Excellent product, great quality, swift delivery

Anche la seconda e la terza classificata rispecchiano le nostre aspettative. Tutte e tre le nostre recensioni sono sia positive nel testo che positivi nelle stelle date, rimanendo coerenti nella loro generalità.

Ora passiamo ai migliori titoli positivi:

	RATING	TITLE	TEXT
260	5 Stars	Brilliant quality and good fun	Brilliant quality and good fun
444	5 Stars	Good quality and fast delivery	Very good package and well delivery. Good quality product, I like it!
62	5 Stars	very well packaged - brand new.	I got this Nintendo Switch from amazon as the seller, so directly from amazon as a new product. I had my doubts [...]

Come possiamo vedere, tutti e tre i titoli sono positivi, i primi due riguardano il prodotto, invece il 3° si concentra più sulla spedizione. Anche qui si rimane coerenti nella positività calcolata.

Cambiamo sponda, andando a vedere i podi delle recensioni e dei titoli negativi.
Partiamo dalle recensioni:

	RATING	TITLE	TEXT
425	1 Stars	Arrived but not working	Arrived and not working. Seller refusing to refund for no hood reason
125	5 Stars	Quality system for the family	A great system for the whole family but not happy that Amazon increased the price from the RRP of 279.99 to 299.99.
402	3 Stars	Daylight robbery	Overpriced...

A differenza delle recensioni positive qui la situazione cambia, infatti notiamo subito che le stelle date a queste recensioni non rispecchiano le nostre previsioni, tranne la recensione **n.425** che le rispecchia in pieno. Nella recensione **n.125** le stelle date sono ben **5**. La negatività della recensione è data dalla “cattiva” gestione dei prezzi da parte di Amazon e non quindi da problematiche avute con il nostro prodotto.

La recensione **n.402** rimane molto ambigua, probabilmente il cliente, anche qui, non ritiene che il prezzo sia giusto.

Passiamo ai titoli più negativi:

	RATING	TITLE	TEXT
425	1 Stars	Arrived but not working	[...]
134	2 Stars	Disappointed	The Nintendo Switch is a great product in terms if gaming. But my one has arrived with a problem. When I play with [...]
40	2 Stars	Disappointing	The concept of the Switch is great, but unfortunately ours suffered with a faulty red [...]

Guardando i risultati ottenuti notiamo che il titolo più negativo è proprio quello della recensione più negativa, cioè la numero **n.425**. Scorrendo i titoli delle recensioni **n.134** e **n.40** troviamo che si assomigliano molto, ma la cosa più curiosa è che leggendo la recensione per intero notiamo che tutti e due i clienti sono soddisfatti della console, ma si lamentano entrambi di aver avuto problemi con i controller.

Terminata l'analisi sulle singole recensioni e sui loro titoli , aventi un numero di parole indefinito, passiamo ad evidenziare quali recensioni con un maggior quantitativo di parole sono state classificate più positive e più negative.

Questa procedura è stata condotta solo sulle recensioni, poiché sono quelle che teoricamente hanno più parole a confronto ai titoli.

Per fare ciò torniamo ad usare la funzione contenuta in *tidytext*, usando sempre lo stesso lessico per l'analisi.

```
#SENTIMENT SU RECENSIONI CHE UN NUMERO DI PAROLE ELEVATO
sentiment_big_text <- rev_words %>%
  inner_join(get_sentiments("afinn"), by = "word") %>%
  group_by(id) %>%
  summarize(sentiment = mean(value),
            words = n()) %>%
  ungroup() %>%
  filter(words >= 10)
```

```
> sentiment_big_text %>%
+   arrange(desc(sentiment))
# A tibble: 6 x 3
   id sentiment words
<int>      <dbl> <int>
1     9      1.68    37
2    17      1.47    19
3     3      0.75    16
4     1      0.211   90
5   499     -0.583    12
6    50     -0.6     10
```


Sono poche le recensioni aventi un numero di parole positive molto alto.

	9
RATING	5 Stars
TITLE	Fantastic, would have been hard to survive lockdown without this
TEXT	<p>Your browser does not support HTML5 video. This was a birthday present to the 5 year old, that came at the right time for us, as we are spending lockdown in a rather small flat, with three kids. Before I describe the specs of the Switch, I'd like to explain why this was selected instead of the Switch Light or the more expensive PS or Xbox. Our older kids are 5 and 6, which is almost the age when they can "do it themselves", but not quite there yet. The operation functions of Nintendo Switch are easily learnt, the kids can set it up themselves and the variety of games Nintendo offers is often in the right mind for children this age - i.e. not so sophisticated or complicated that they would need an adult's help. The Switch Light is a good option of course, especially for a kid with no siblings, but without the ability to hook the console to the tv, it's going to be harder to turn this into a multi participation venue, i.e. only one kids at a time. With the Nintendo Switch, we often see the kids opt for TV game mode, and we can all participate in active games such as "Just Dance" or "Tokyo 2020 Olympics", which the game rewards them through actual movement. Why not PS or Xbox? Well, they're more expensive and as described above sophisticated. So, if you have a very technical child of 10+ maybe they are the better option. For us, the fact that they can hook up the controller themselves and the games are plain and simple was the selling point. What's in the box? The Switch console, a tv hookup hub, a charging cable, 2 Joy Cons (note - 2 Joy Cons makes one whole controller in essence, and some games would require them as such, i.e. you need both per user), and one Joy Con Grip (to combine them to one controller). Game modes: The solo mode, as seen on the second picture/the video, provides the kids hours of fun on their own. - The TV hookup (photo 3) 1) You can play solo with a big screen. Easy to setup - you just need an empty socket near the tv and an HDMI connection. You then can use the two Joy-Con's with the the controller hub that comes with the switch. 2) Another option (for games which allow it) is to use each Joy-Con as a solo controller. The Joy-Cons are gyroscopic and feel the hand movement, so you can play "Just Dance" or "Mario Kart" by moving your hands around. they even provide feedback in terms of vibrations, and remind of the old Wii consoles. - Detaching the JoyCons and using the console as a small screen, this is the worst option but it works too, there's a small leg that allows the console to stand upright (but it tends to fall off, a bit of a design flaw that is negligible since this game mode is not the best anyways). To sum up all the Pro's: hours of fun, interactive games which will keep the kids fit not just in 'zombie' mode, dual game mode as mentioned and adequate for kids up from the age of 5 to setup on their own. To add to that: there are dozens of game around, with genres that fit just about anyone. Con's: Not all is perfect with the Nintendo Switch and here are some faults and the way to overcome them (if applicable): - The games are expensive. A new game is around £40, which is not optimal when you want the kids to have a variety. Also, if you bought a game that they didn't like/can't handle - that's money down the toilet. Luckily, the Nintendo Shop has previous and free periods that allow you to test games (but we don't really bother with these). - The Switch comes with two Joy Cons. If you want to play more than two people you will have to get another set, and another grip. while this is relatively cheap, you will then want a charging dock, so you won't have to rotate between them, and it doesn't really end... - You would probably want a screen protector, a travel case, and a few more games... so again when it comes to costs, add at least £100-150 to the original price (and prob more). - The worst thing is the need to subscribe to the Nintendo Shop to be able to play online multi games in several games. this is downright infuriating, as you need to pay for something you already paid for. However, we found that we enjoy the games without the online option as it is. Good games to start with: - Animal Crossings New Horizons - especially for calmer kids who likes building and strategy. - Mario Kart - this is brilliant, the superlatives this game has received are very just, it provides several game modes and hours of fun for all of the family (I enjoy playing with the kids myself, and we get quite competitive)- Just Dance - a good workout in form of dancing with the Joy Cons which will provide hours of fun, but each participant needs a JoyCon. Hope this helped someone to decide whether to get one or not. and If not - feel free to ask questions or comment, or vote this 'helpful' if indeed it was. Thanks and Game on!</p>

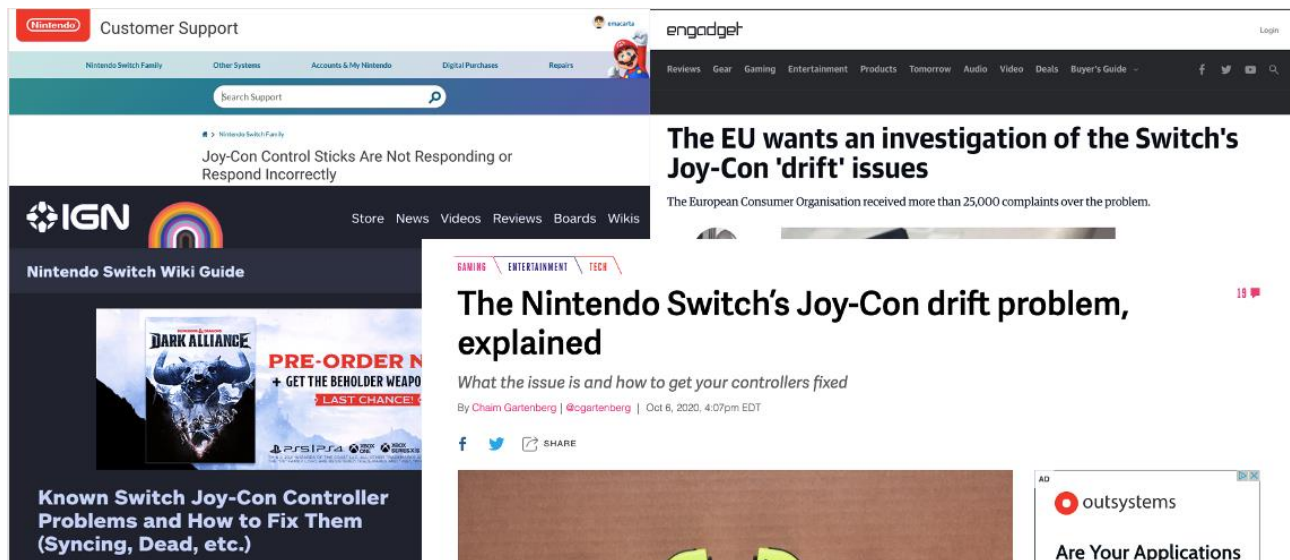
Questa è la recensione più positiva, con un numero di parole dello stesso tenore molto alto. Il cliente ha descritto analiticamente la Nintendo Switch, andando a confrontarla anche con le altre console in circolazione. Come vediamo dal titolo l'acquisto è stato fatto durante il un lockdown, e soprattutto è stato un regalo per i suoi figli piccoli.

Tutte le informazioni che leggiamo in questo testo le abbiamo spesso trovate durante l'analisi.

Andiamo ad individuare ora la recensione con il maggior numero di parole negative.

	50
RATING	2 Stars
TITLE	Controller crashed in less than a month
TEXT	Word ont he street is Joy Con controllers are LOUSY and fail... and that was exactly what we experienced too... Less than a month into ownership with very careful use, the left controller just stopped communicating with the main console. USEless waste of dollar. RETurned ... Amazons service was / is great, they refunded straight away... bu Switch to somethig reliable. POOR POOR POOR ...

Anche qua, come nell'analisi generale delle recensioni, troviamo che la lamentela riguarda proprio il funzionamento del controller. Decidiamo quindi di cercare più informazioni al riguardo tramite il WEB.



La ricerca ha evidenziato svariati articoli riguardanti questo problema, che affligge moltissime unità, alcuni appaiono addirittura sul sito della ufficiale Customer Support della Nintendo. Capiamo quindi che non si tratta di casi isolati ma di una vero e proprio difetto del prodotto.

L'ultima analisi che andremo a fare riguardo al sentiment, sarà quella di individuare le parole positive e negative più comuni.

Per effettuare quest'analisi, useremo il lessico **bing** che permette la classificazione delle parole in modo binario in categorie positive e negative.

```
#PAROLE PIU' POSITIVE E PIU' NEGATIVE USATE NELLE RECENSIONI
#RECENSIONI
bing_word_counts <- rev_words %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()
```

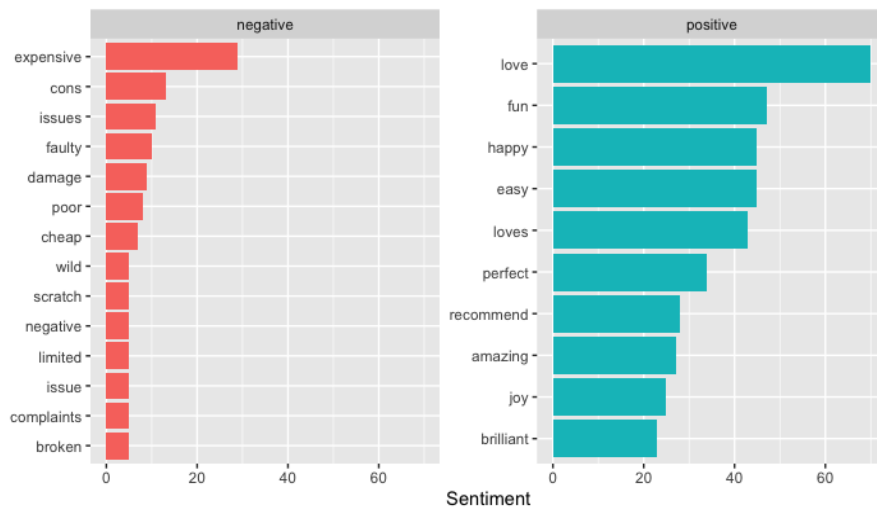


Figura 11: Grafico Parole Positive/Negative(RECENSIONI)

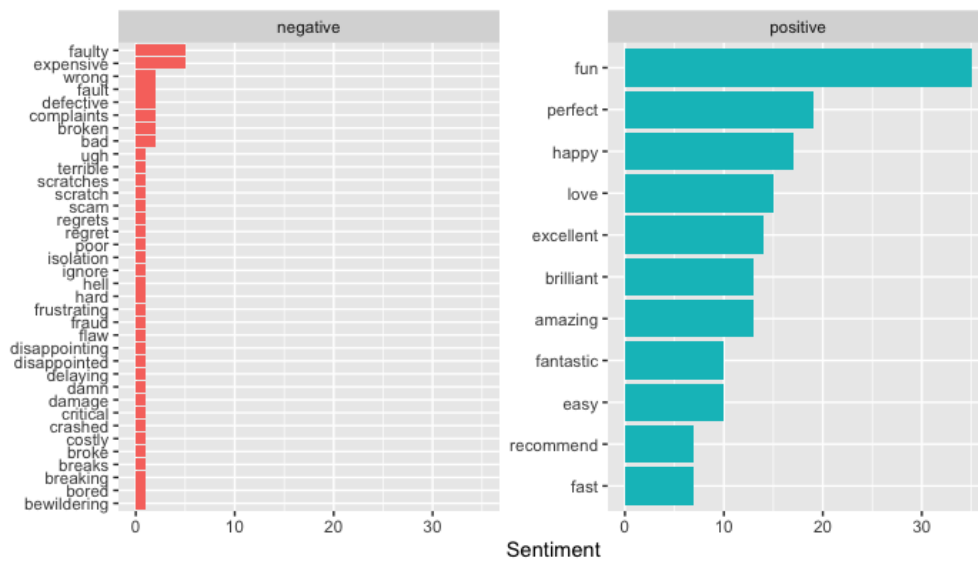


Figura 12: Grafico Parole Positive/Negative(TITOLI)



Figura 13: Wordcloud Parole Positive/Negative(RECENSIONI)

4.6 Analisi n-grammi

Finora abbiamo considerato le parole come singole unità. Tuttavia, per esaminare al meglio in nostro dataset andremo ad analizzare le relazioni tra le parole, osserviamo quali parole tendano a essere spesso vicine. Per eseguire questo studio, useremo i metodi offerti da *tidytext* per calcolare e visualizzare le relazioni tra le parole nel nostro dataset.

Per tokenizzare i testi delle nostre recensioni in bi-grammi andremo ad usare sempre la funzione `unnest_tokens()`, aggiungendo `token = "bigrams"`, per far sì che le parole vengano tokenizzate per coppie, piuttosto che per singole parole.

```
rev_bigrams <- text %>%  
  unnest_tokens(bigram, text, token = "ngrams", n = 2)
```

Andiamo ora ad esaminare i bi-grammi più comuni ed attraverso la funzione `count()`, della libreria *dplyr*. In seguito, usiamo la funzione `separate()`, della libreria *tidyr*.

Quest'ultima divide la colonna in più parti attraverso un separatore . Si creano pertanto due colonne, *"word1"* e *"word2"*. Giunti a questo punto possiamo rimuovere i casi in cui una delle due parole è una stop-word.

```
rev_bigram_counts <- rev_bigrams %>%  
  count(bigram, sort = TRUE) %>%  
  separate(bigram, c("word1", "word2"), sep = " ")  
  
bigrams_filtered <- rev_bigram_counts %>%  
  filter(!word1 %in% stop_words$word) %>%  
  filter(!word2 %in% stop_words$word)
```

Creati i nostri subset composti da bi-grammi, possiamo ora andare a svolgere le varie analisi.

Il primo caso che andremo a studiare, è l'uso di alcune parole che possono essere scambiate per positive o negative se non viste senza la parola che li precede.

Le parole che andremo a cercare sono:

```
negate_words <- c('not', 'no', "don't")  
good_words <- c('good', 'fine', 'nice', 'perfect')
```

Il codice per il grafici qui sotto troverà le righe dove sono contenute le parole che ci interessano e poi andrà a calcolare il sentiment, con l'uso il lessico **AFINN**, delle parole a cui sono correlate. Per eseguire queste analisi, immetteremo il subset `rev_bigram_counts` poiché ancora provvisto di stopwords.

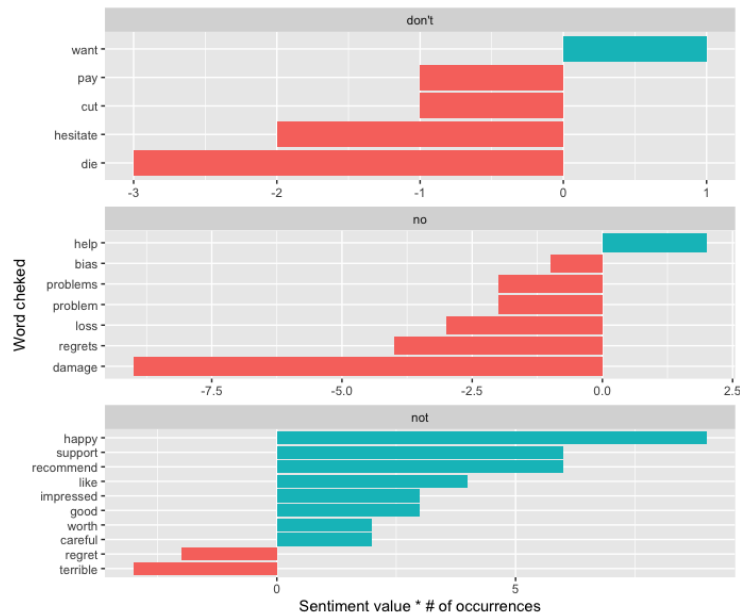


Figura 14: Grafico Parole di Negazione

Nel il primo grafico, cioè quello dei lemmi di negazione, vediamo chiaramente che la maggior parte degli stessi che a cui vengono attribuite le negazioni sono spesso a parole con sentimento negativo. L'unica che fa eccezione è "not", che è usata spesso vicino a parole positive.

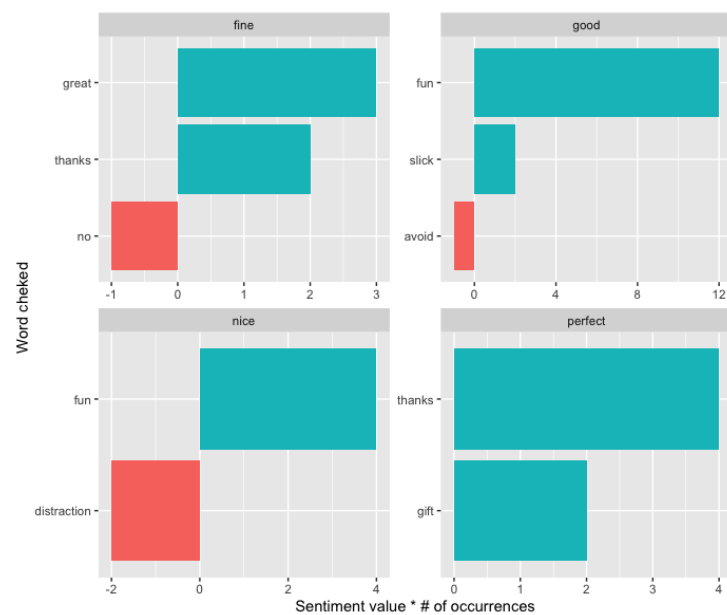


Figura 15: Grafico Parole Positive

Nelle parole positive, vengono associate per lo più altri lemmi con un sentiment positivo. Notiamo un particolare molto simpatico: andando a guardare il grafico delle parole associate a "nice" troviamo "distraction", potremmo classificarlo un pregio o un difetto?

La seconda analisi che realizzeremo, sarà incentrata sulla problematica che abbiamo riscontrato più volte durante l'analisi delle recensioni, cioè il cattivo funzionamento del controller della Nintendo Switch.

Le parole che cercheremo saranno quindi:

```
problems_words <-c('joystick','drift')
```

Il risultato è proprio quello che ci aspettavamo, infatti andando a cercare proprio il termine che descrive in pratica la problematica effettiva del controller, cioè “drift” 2, troviamo che vien associata alla parola “problem”.

[2] Drift: quando una delle due levette per controllare il movimento e la visuale in un videogioco iniziano muoversi in modo anomalo senza avergli dato alcun input fisico.

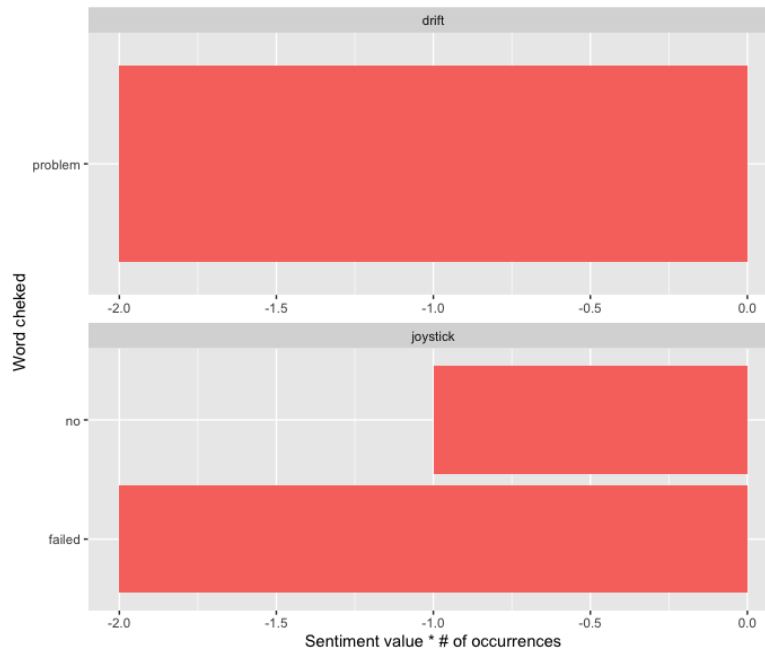


Figura 16: Grafico Ricerca Problema

L'ultimo studio che andremo a svolgere sarà quello di visualizzare tutte le relazioni tra le parole contemporaneamente.

Partiamo dal subset **bigrams_filtered**, che contiene i bi-grammi senza le stopwords.

Usiamo la libreria **igraph**, che contiene funzioni per la manipolazione e l'analisi delle reti. Applicando la funzione **graph_from_data_frame()**, quest'ultima crea un grafico da uno o due frame di dati contenenti l'elenco dei bordi e gli attributi dei bordi/vertici.

```
bigram_graph <- bigrams_filtered %>%
  filter(n > 2) %>%
  graph_from_data_frame()
```

igraph possiede funzioni di stampa integrate, ma non sono delle migliori. Useremo quindi il pacchetto **ggraph** perché implementa queste visualizzazioni in termini di grammatica della grafica.

Possiamo convertire un oggetto **igraph** in un **ggraph** con la funzione **ggraph()**, dopodiché aggiungiamo livelli: nodi, bordi e testo.

In seguito attraverso la funzione **edge_alpha()** rendiamo i collegamenti più chiari o più scuri in base a quanto il bi-gramma sia comune.

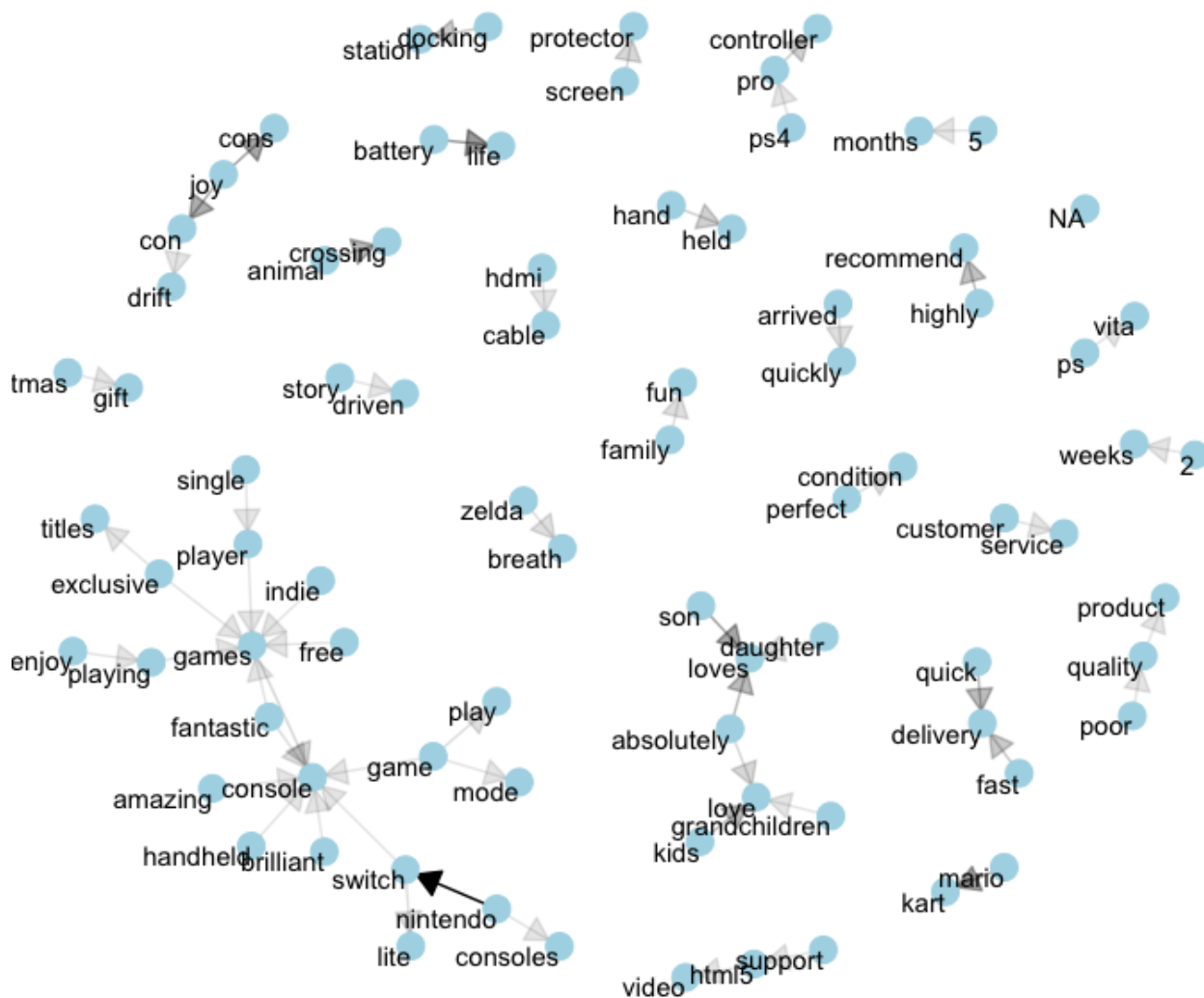
Usando **grid::arrow()**, aggiungiamo la direzionalità della freccia.

Infine aggiungiamo un tema utile per tracciare reti, **theme_void()**.

```
set.seed(1223)

a <- grid::arrow(type = "closed", length = unit(.15, "inches"))

ggraph(bigram_graph, layout = "fr") +
  geom_edge_link(aes(edge_alpha = n), show.legend = FALSE,
    arrow = a, end_cap = circle(.07, 'inches')) +
  geom_node_point(color = "lightblue", size = 5) +
  geom_node_text(aes(label = name), vjust = 1, hjust = 1) +
  theme_void()
```



5 Conclusioni

Da tutti gli studi che abbiamo eseguito siamo riusciti a capire molte informazioni riguardo al prodotto di cui abbiamo analizzato le recensioni.

Andando a fare un sunto generale di ciò che abbiamo appreso:

- si conferma che la Nintendo Switch è molto amata dai suoi acquirenti, dato il suo elevato rateo di stelle e una netta minoranza di recensioni negative, con circa il **90%** di recensioni con **5** stelle;
- le recensioni (e, probabilmente, anche le vendite) durante l'anno **2020** sembrano aumentate a causa della pandemia (oltre i **3/5** delle recensioni presenti nel nostro dataset sono dell'anno **2020**). L'evento ha costretto molte persone a rimanere a casa, le quali avrebbero acquistato la console per passare il tempo da soli o in compagnia durante i lockdown, imposti dalle autorità governative per l'emergenza pandemica;
- questo prodotto sembra essere utilizzato da tutto il nucleo familiare, nessuno escluso (poiché la frequenza nelle recensioni delle parole riguardanti il campo semantico della famiglia è molto elevata);
- le recensioni negative sembrano addensarsi sul malfunzionamento del controller, problema strutturale che pare non essere stato ancora risolto dalla Nintendo (poiché presenti ancora recensioni negative con questo argomento anche nell'anno in corso).

Fonti

Estrazione dei dati: Pagina Nintendo Switch Amazon.uk

- https://www.amazon.co.uk/Nintendo-Switch-Neon-Red-blue/dp/B07W4CK8KR/ref=cm_cr_ar_p_d_product_top?ie=UTF8

Articoli riguardante la maggiore vendita di Nintendo Switch durante la pandemia

- <https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/nintendo-profits-up-over-400-as-covid-19-drives-game-sales-59801601>
- <https://www.bbc.com/news/business-54813841>

Articoli riguardante il problema dei controller della Nintendo Switch

- <https://www.theverge.com/21504741/nintendo-switch-joy-con-drift-problem-explained>
- https://en-americas-support.nintendo.com/app/answers/detail/a_id/46903/~/joy-con-control-sticks-are-not-responding-or-respond-incorrectly
- [https://www.ign.com/wikis/nintendo-switch/Known_Switch_Joy_Con_Controller_Problems_and_How_to_Fix_Them_\(Syncing,_Dead,_etc.\)](https://www.ign.com/wikis/nintendo-switch/Known_Switch_Joy_Con_Controller_Problems_and_How_to_Fix_Them_(Syncing,_Dead,_etc.))
- https://www.engadget.com/nintendo-switch-joy-con-drift-european-union-151738223.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xLmNvbS8&guce_referrer_sig=AQA-AAJIX72IS9N2FwB0byrS66Fk-eXbPeQ1WD8UXiyeh97DohSux6RKSON5IOUJgNxFlyyi642xU4rZ7TAwElRnYo4SG1N4NMSWTq8hMGYiSW2tf46dE1SJiCwtSYvVVJMyRNOHa-Hbl-GnrkQ7yeY419ibZzVcgWabmXXI1fpx4mU9P