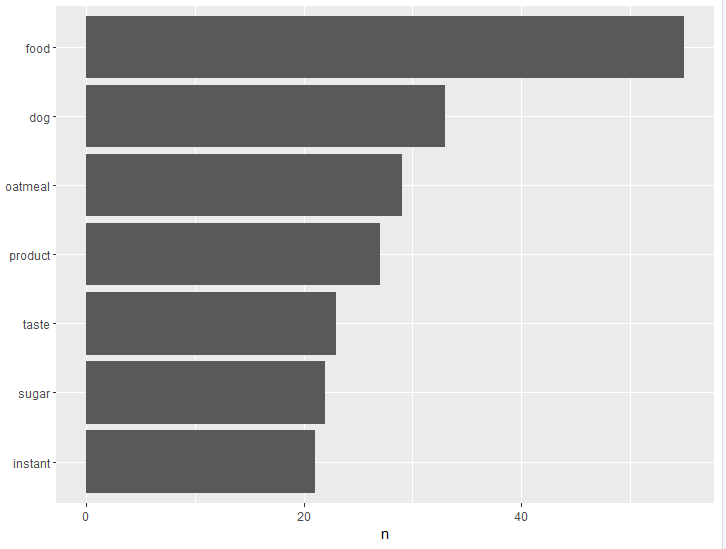
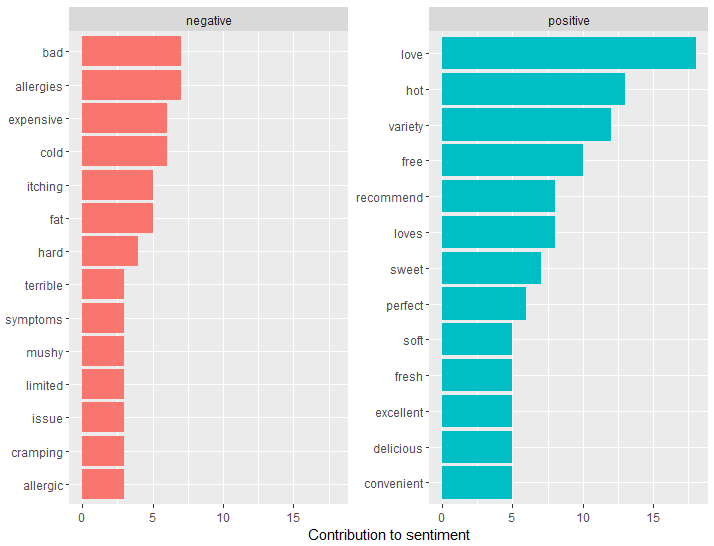
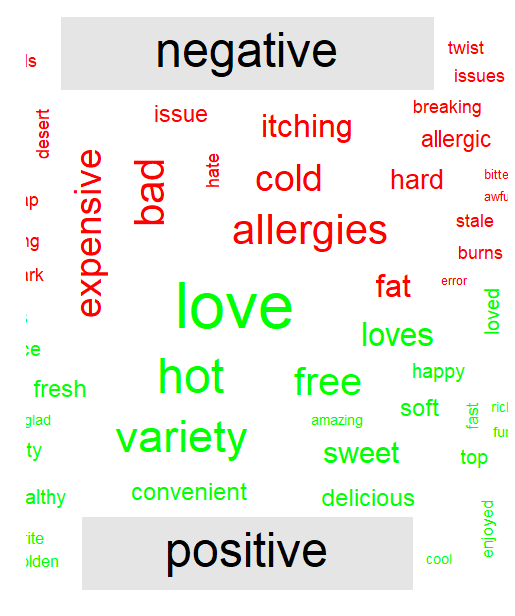
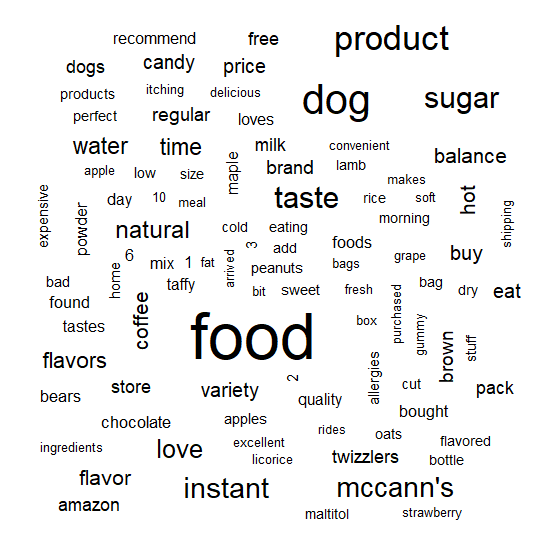
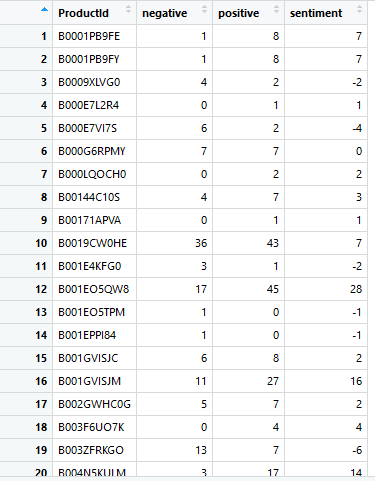
**W4Lab – Bruno Simione Beltrame**

**Part A**







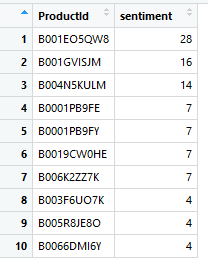


**Challenge Questions**

**1 – Top 10**

> combinedReview\_top10 <- combinedReview\_sentiment[order(combinedReview\_sentiment$sentiment, decreasing = TRUE), ]

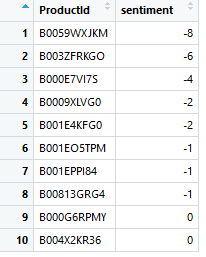
> View(combinedReview\_top10[c(1:10), c(1,4)])



**2 – Worst 10**

> combinedReview\_worst10 <- combinedReview\_sentiment[order(combinedReview\_sentiment$sentiment), ]

> View(combinedReview\_worst10[c(1:10), c(1,4)])



**3 – Big Review**

> reviewsAll <- reviews[, c(2,10)]

> View(reviewsAll)

> combinedReviewAll <- reviewsAll %>% group\_by(ProductId) %>% summarise(Text = paste(Text, collapse = " "))

> combinedReviewAll$ProductId <- as.factor(combinedReviewAll$ProductId)

> combinedReviewAll2 <- combinedReviewAll %>% unnest\_tokens(word, Text)

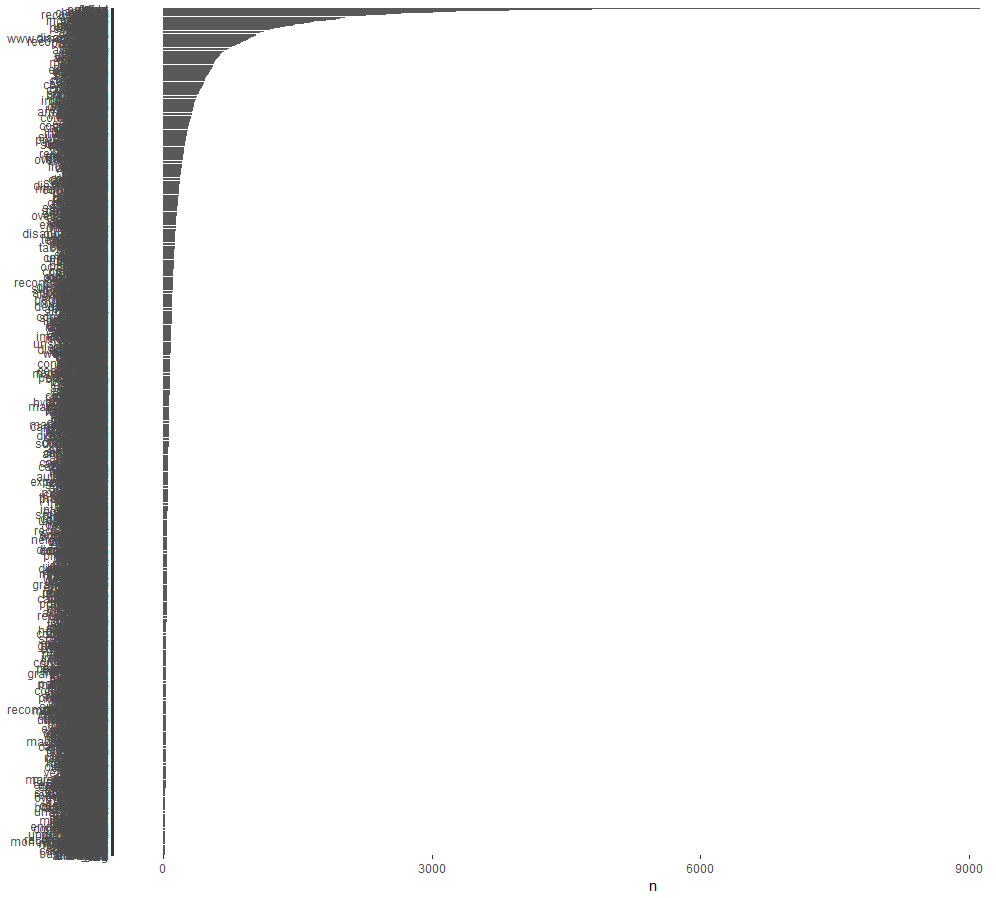
> View(combinedReviewAll2)

> combinedReviewAll3 <- combinedReviewAll2 %>% anti\_join(stop\_words, by = "word")

> combinedReviewAll3 <- filter(combinedReviewAll3, word!="br")

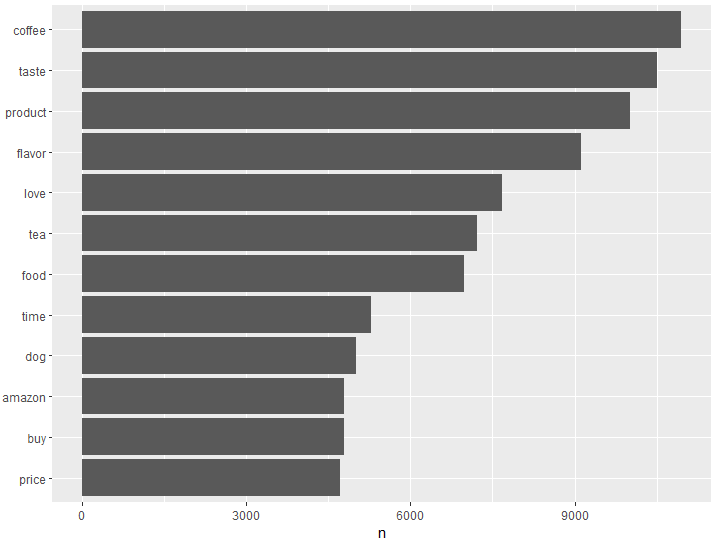
> combinedReviewAll3 %>% count(word, sort = TRUE) %>% filter(n>20) %>% mutate(word = reorder(word,n)) %>% ggplot(aes(word, n)) + geom\_col() + xlab(NULL) + coord\_flip()

N >20 🡪 Impossible to read



> combinedReviewAll3 %>% count(word, sort = TRUE) %>% filter(n>4000) %>% mutate(word = reorder(word,n)) %>% ggplot(aes(word, n)) + geom\_col() + xlab(NULL) + coord\_flip()

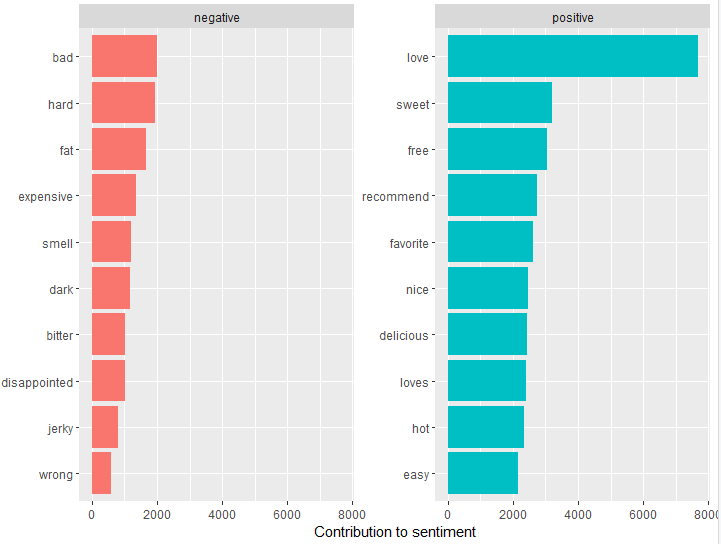
N > 4000



> bing\_words\_counts\_all <- combinedReviewAll3 %>% inner\_join(get\_sentiments("bing")) %>% count(word, sentiment, sort = TRUE) %>% ungroup()

Joining, by = "word"

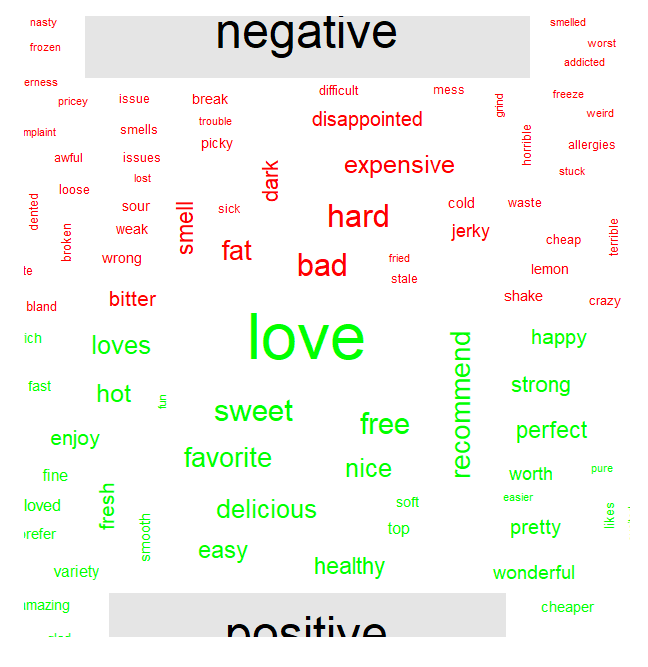
> bing\_words\_counts\_all %>% group\_by(sentiment) %>% top\_n(10) %>% ungroup() %>% mutate(word = reorder(word, n)) %>% ggplot(aes(word, n, fill = sentiment)) + geom\_col(show.legend = FALSE) + facet\_wrap(~sentiment, scales = "free\_y") + labs(y = "Contribution to sentiment", x = NULL) + coord\_flip()



> combinedReviewAll3 %>% count(word, sort = TRUE) %>% with(wordcloud(word, n, max.words = 100))



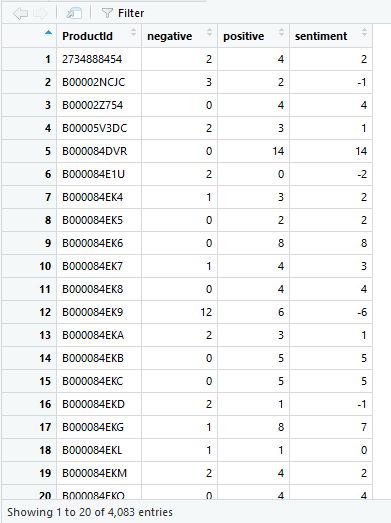
> combinedReviewAll3 %>% inner\_join(get\_sentiments("bing")) %>% count(word, sentiment, sort = TRUE) %>% acast(word ~ sentiment, value.var = "n", fill = 0) %>% comparison.cloud(colors = c("red", "green"), max.words = 100)



> combinedReviewAll\_sentiment <- combinedReviewAll3 %>% inner\_join(get\_sentiments("bing")) %>% count(ProductId, sentiment) %>% spread(sentiment, n, fill = 0) %>% mutate(sentiment = positive - negative)

Joining, by = "word"

> View(combinedReviewAll\_sentiment)



**Sorted:**

**Worst**

> combinedReviewAll\_sentiment\_sorted <- combinedReviewAll\_sentiment[order(combinedReviewAll\_sentiment$sentiment),]

> View(combinedReviewAll\_sentiment\_sorted[,c(1,4)])

**Top**

> combinedReviewAll\_top10 <- combinedReviewAll\_sentiment[order(combinedReviewAll\_sentiment$sentiment, decreasing = TRUE), ]

> View(combinedReviewAll\_top10[, c(1,4)])