Some NLP

Yuchen Hu

11/26/2018

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------- tidyverse 1.2.1 --
```

```
## <U+221A> ggplot2 3.0.0     <U+221A> purrr   0.2.4
## <U+221A> tibble  1.4.2     <U+221A> dplyr   0.7.5
## <U+221A> tidyr   0.8.1     <U+221A> stringr 1.3.1
## <U+221A> readr   1.1.1     <U+221A> forcats 0.3.0
```

```
## -- Conflicts -------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

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

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:purrr':
##
##     compose, simplify
```

```
## The following object is masked from 'package:tidyr':
##
##     crossing
```

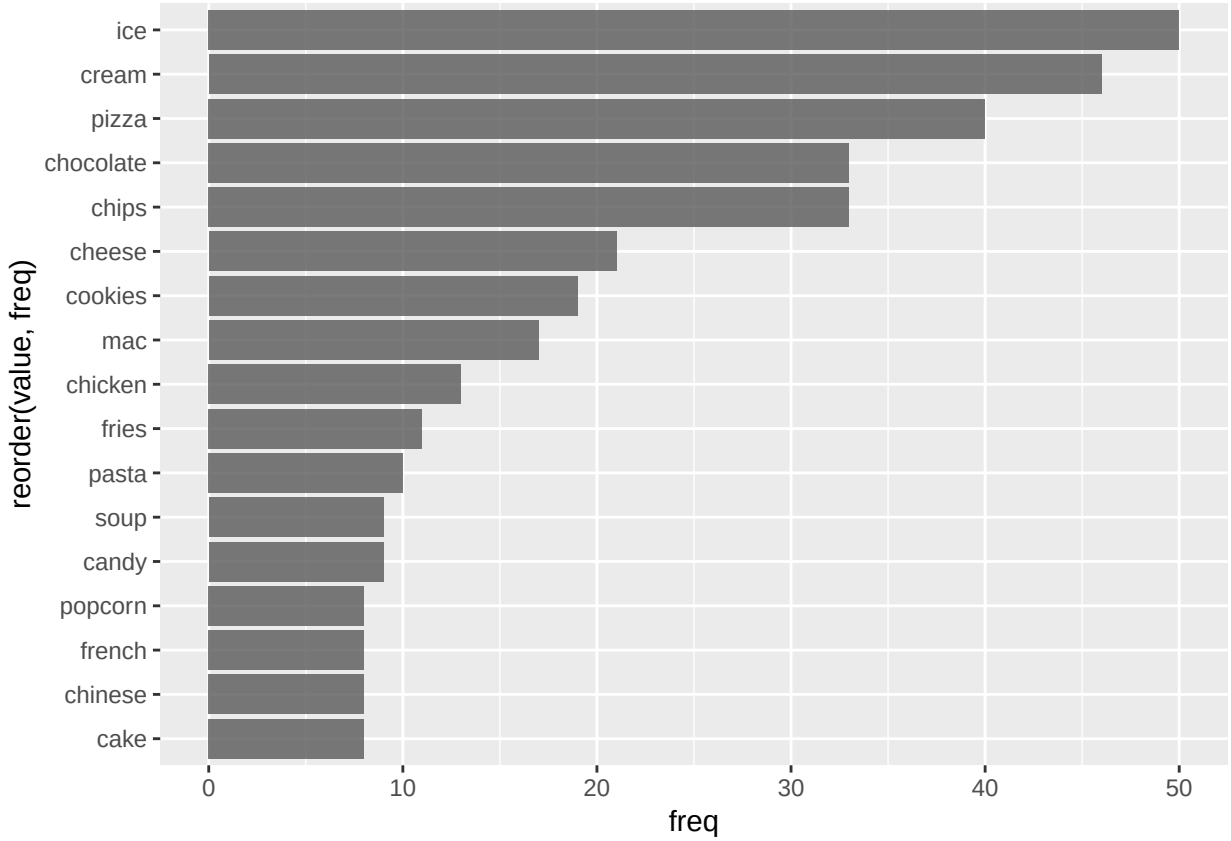
```
## The following object is masked from 'package:tibble':
##
##     as_data_frame
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

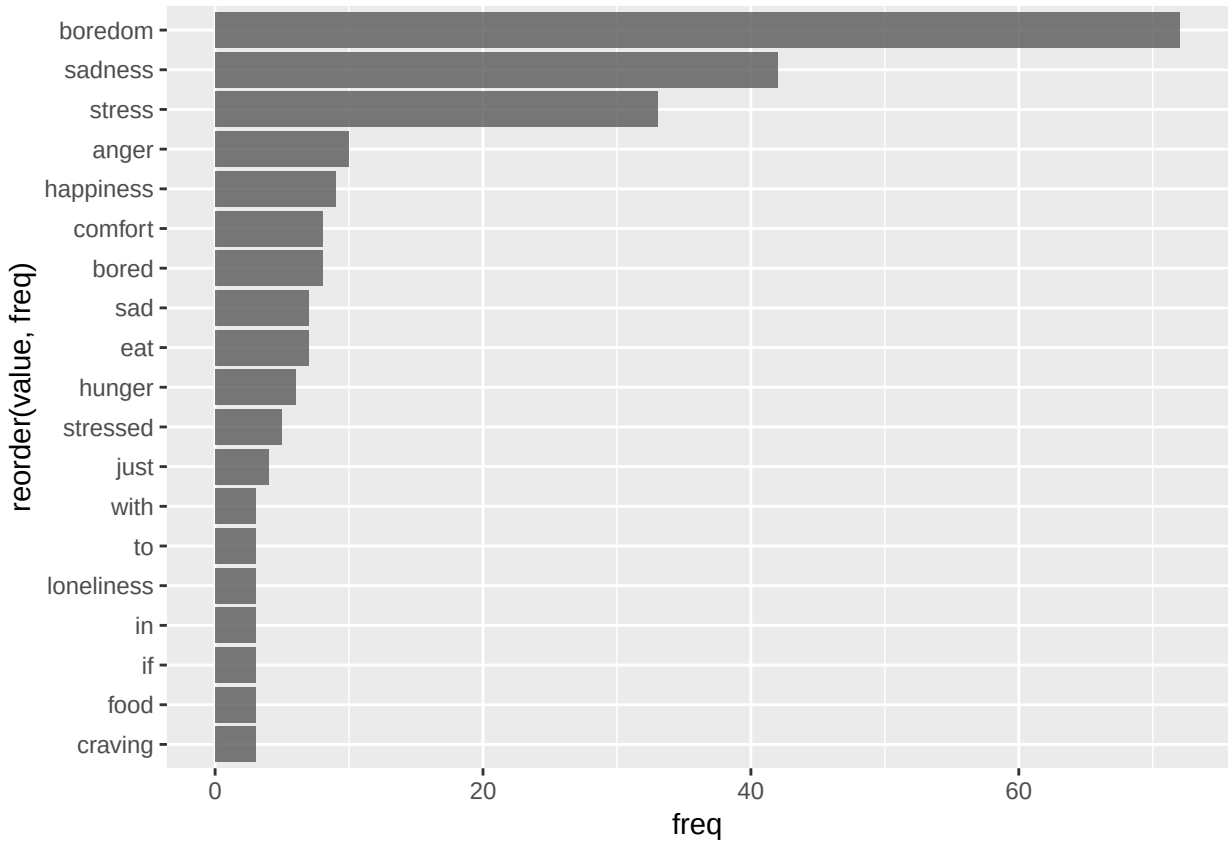
```
## The following object is masked from 'package:base':
##
##     union
```

```
food <- read.csv("food_coded.csv",stringsAsFactors = FALSE)

comfort_food_reason <- food$comfort_food_reasons %>% str_split(" ")

comfort_food <- food$comfort_food %>%
  str_to_lower() %>%
```

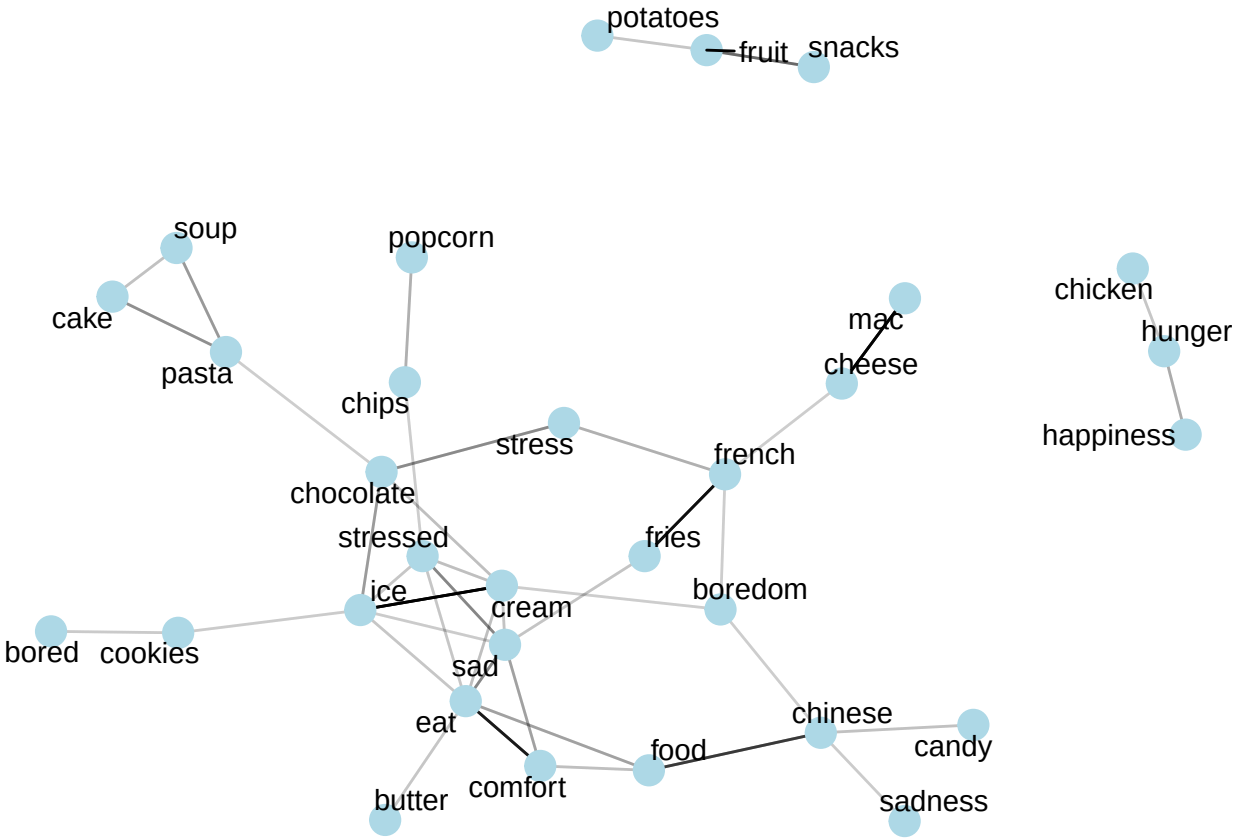
```
  str_split(",|\\|/| ") %>%
  map_df(enframe, .id = 'student') %>%
  unnest %>%
  filter(value != '' & value != 'and') %>%
  select(student, value)
comfort_food %>%
  group_by(value) %>%
  summarize(freq = n()) %>%
  top_n(15, wt = freq) %>%
  ggplot(aes(reorder(value, freq), freq)) +
  geom_bar(stat = "identity", alpha = .8, show.legend = FALSE) +
  coord_flip()
```




```
comfort_food_reason <- food$comfort_food_reasons %>%
  str_to_lower() %>%
  str_split(",|\\|/| ") %>%
  map_df(enframe, .id = 'student') %>%
  unnest %>%
  filter(!(value%in%c('','and','i','when','am','or','they','i\'m','my','usually','are','a'))) %>%
  select(student, value)
comfort_food_reason %>%
  group_by(value) %>%
  summarize(freq = n()) %>%
  top_n(15, wt = freq) %>%
  ggplot(aes(reorder(value, freq), freq)) +
  geom_bar(stat = "identity", alpha = .8, show.legend = FALSE) +
  coord_flip()
```




```
comfort <- rbind(comfort_food,comfort_food_reason)
comfort_pairs <- comfort %>%
  pairwise_count(value, student, sort = TRUE)
comfort_cor <- comfort %>%
  group_by(value) %>%
  filter(n() >= 5) %>%
  pairwise_cor(value, student) %>%
  filter(!is.na(correlation), correlation > .15)

set.seed(123)
comfort_cor %>%
  graph_from_data_frame() %>%
  ggraph(layout = "fr") +
  geom_edge_link(aes(edge_alpha = correlation), show.legend = FALSE) +
  geom_node_point(color = "lightblue", size = 5) +
  geom_node_text(aes(label = name), repel = TRUE) +
  theme_void()
```


potatoes
fruit
snacks
soup
popcorn
cake
pasta
chips
mac
cheese
chicken
hunger
happiness
stress
french
chocolate
stressed
fries
ice
cream
boredom
bored
cookies
sad
eat
chinese
candy
food
comfort
butter
sadness