

### Problematic Words

Combining all strings into one column

Creating a regex pattern

Createing new column if problematic words exist & keep only problematic rows

Eliminate false positives

Problematic transactions

### Problematic MCC Codes

Tokenize the text

Remove stop words

Stemming and Lemmatizing

Sentiment total

Tokenize the text

Remove stop words

Stemming and Lemmatizing

Sentiment total

# P-Card Internal Control Analysis and Apple's MD&A Sentiment Analysis



## Authors

Stephen Bernardo, Ikhida Longe, Shelbie Nehra, Shannon Wigger, Morgan Zuidema



## Install and load packages

```
# uncomment install packages if necessary
# textdata package is required to download the Loughran-McDonald lexicon
# install.packages(c('tidyverse', 'textdata', 'tidytext'))
```

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.2      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr   1.5.0
## ✓ ggplot2    3.4.4      ✓ tibble    3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr     1.3.0
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(textdata)
library(tidytext)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##   set_names
##
## The following object is masked from 'package:tidyr':
##
##   extract
```



## Part 1 - P-Card Internal Control Test



### 1a: Creating a table of 10-20 most problematic transactions

These rows are a list of transactions, not a list of MCCs.

```
df_group <- read_rds("class_data_pcard_data_forNLP.rds")

# Clean data
df1_group <- df_group %>% select(-Agency.Number, -Agency.Description, -Posting.Date, -Month, -Weekday, -Day.Month)
```

## Problematic Words

# Combining all strings into one column

```
df1_group %<>%
  mutate(
    all_text = paste(Item.Description, Merchant, MCC, sep = '__')
  )
```

## Creating a regex pattern

```
pattern_group <- "\\bFEE\\b|DECORATIONS|DEPOSIT|PREPAYMENT|DONATION|\\bAWARD\\b|FUEL|\\bTAX\\b|AMAZON PRIME|SAM'S CLUB|MOVING|REGISTRATION|GASOLINE|FLOWERS|CANDY|UNIFORM|INSURANCE|BILLIARD"
```

## Createing new column if problematic words exist & keep only problematic rows

```
# Method A (just use one method): New column with count
df1a_group <- df1_group %>%
  mutate(problem = str_count(all_text, pattern_group))

# Method B (just use one method): New column with TRUE or FALSE
# df1b_group <- df1_group %>%
#   mutate(problem = str_detect(all_text, pattern_group))

# Filter
df_problems_group <- df1a_group %>% filter(problem > 0)
```

## Eliminate false positives

## Examine the problems to look for false positives

```
df_problems_group %>%
  select(Cardholder.Name, Item.Description, Amount, Merchant, MCC, problem) %>%
  arrange(MCC, desc(Amount))
```

Cardholder.Name <chr>	Item.Description <chr>	Am... <dbl>	Merchant <chr>	MCC <chr>
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900796680634	AIRLINES AIR CAF
REUTER, D	AIR TRAVEL	10.00	AGENT FEE 8900794193230	AIRLINES AIR CAF
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900794865396	AIRLINES AIR CAF
REUTER, D	AIR TRAVEL	10.00	AGENT FEE 8900793896988	AIRLINES AIR CAF

Cardholder.Name <chr>	Item.Description <chr>	Am... <dbl>	Merchant <chr>	MCC <chr>
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900794822068	AIRLINES AIR CAF
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900795060593	AIRLINES AIR CAF
REUTER, D	AIR TRAVEL	10.00	AGENT FEE 8900794193239	AIRLINES AIR CAF
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900796247382	AIRLINES AIR CAF
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900795350648	AIRLINES AIR CAF
MATTHEWS, W	AIR TRAVEL	10.00	AGENT FEE 8900794437599	AIRLINES AIR CAF

1-10 of 2,227 rows | 1-5 of 6 columns

Previous 1 2 3 4 5 6 ... 223 Next

## Eliminate rows that are not a problem and then check again

```
#safe_words <- "ISOPROPYL|SPRY GUN|CAULK GUN|BLOWGUN|HAND SANITIZER|HAND SANITIZI|A
LCOHOL WIPES|HEAT GUN|DISINFECTANT|MINERAL SPIRITS|DISINFECTA|ETHYL|GLUE GUN|STAPLE
GUN|RUBBING ALCOHOL|LCOHOL WET WIPES|ALCOHOL REAGENT|ALCOHOL PREP|DENATURED ALCOHOL
|BLOW GUN|VINYL ALCOHOL|ALCOHOL RESISTANT|ALCOHOL WIPE|ALCOHOL-RE|GUN BALLING|SPRAY
GUN|BEER PING|TAPE DISP|MONEY BANK BAG|RUBBERSTAMPS|AMYL ALCOHOL|RUBBER STAMP|ALCOH
OL WIP|GLUE GUN"
```

```
safe_words <- "AGENT FEE|ZOOM.US|PUBLISH|ADVERTISING|BASECAMP|MOVING PIXEL|MASK|TEL
EMEDICINE"
```

## Problematic transactions

```
# Keep only rows without safe words in them
df_problems_group %<>%
  filter(str_detect(all_text, safe_words) == FALSE)

# Look at the potential violations that are high in amount
df_problems_group %>%
  select(-Transaction.Date, -all_text) %>%
  filter(Amount > 1500) %>%
  arrange(desc(Amount))
```

Cardholder.Name <chr>	Item.Description <chr>	Amount <dbl>
HOLMES, K	GENERAL PURCHASE	21161.80
HOLMES, K	GENERAL PURCHASE	16018.47
KISSMAN CH, A	GENERAL PURCHASE	4622.82
BAKER, L	ANNUAL FEE 2020-2021	4313.00

Cardholder.Name	Item.Description	Amount
<chr>	<chr>	<dbl>
HERNDON, L	ANNUAL FEE 2020-2021	4313.00
LINSENMEYER, J	ALUMAGOAL MANCHESTER MATCH NMB TRANSPORT WHEELS (S	4158.00
FIELD, M	ANIMAL PRINTED TERRY KNIT EA SMOCKING ELASTIC CROP	4143.70
THARP, M	GENERAL PURCHASE	3174.70
GREER, B	GENERAL PURCHASE	3142.00
GERLT, T	GENERAL PURCHASE	3017.00
1-10 of 19 rows   1-3 of 6 columns		Previous 1 2 Next

# Problematic MCC Codes

## Explore unique MCCs

```
n_distinct(df_group$MCC)
```

```
## [1] 394
```

```
unique(df_group$MCC)
```

```

## [1] "COURIER SERVICES-AIR OR GROUND,FREIGHT F"
## [2] "GROCERY STORES SUPERMARKETS"
## [3] "INDUSTRIAL SUPPLIES NOT ESLEWHERE CLASSI"
## [4] "BUSINESS SERVICES NOT ELSEWHERE CLASSIFI"
## [5] "AIRLINES AIR CARRIERS"
## [6] "MOTOR VEHICLE SUPPLIES AND NEW PARTS"
## [7] "COMBINATION CATALOG AND RETAIL MERCHANT"
## [8] "SCHOOLS AND EDUCATIONAL SERVICES NOT ELS"
## [9] "UNITED AIRLINES"
## [10] "GROCERY STORES, SUPERMARKETS"
## [11] "LUMBER AND BUILDING MATERIALS STORES"
## [12] "DRUG STORES PHARMACIES"
## [13] "LANDSCAPE AND HORTICULTURAL SERVICES"
## [14] "CABLE SATELLITE & OTHER PAY TV/RADIO SE"
## [15] "MISCELLANEOUS AND RETAIL STORES"
## [16] "LAB/MEDICAL/DENTAL/OPHTHALMIC HOSPITAL E"
## [17] "SERVICE STATIONS"
## [18] "BOOK STORES"
## [19] "MEN'S,WOMENS'AND CHILDREN'S UNIFORMS AND"
## [20] "COMMERCIAL EQUIPMENT, NOT ELSEWHERE CLAS"
## [21] "TELECOMMUNICATION SERV.INCLUD. LOCAL/L.D"
## [22] "CONTINUITY/SUBSCRIPTION MERCHANTS"
## [23] "AUTOMOTIVE PARTS, ACCESSORIES STORES"
## [24] "PLUMBING & HEATING EQUIPMENT AND SUPPLIE"
## [25] "GLASS PAINT WALLPAPER STORES"
## [26] "MEMBERSHIP ORGANIZATIONS NOT ELSEWHERE C"
## [27] "FAIRFIELD INN"
## [28] "MANAGEMENT CONSULTING AND PUBLIC RELATI"
## [29] "HOME SUPPLY WAREHOUSE STORES"
## [30] "SPORTING GOODS STORES"
## [31] "MEDICAL AND DENTAL LABORATORIES"
## [32] "PROFESSIONAL SERVICES NOT ELSEWHERE CLAS"
## [33] "DRUGS,DRUG PROPRIETARIES AND DRUGGIST'S"
## [34] "AMERICAN AIRLINES"
## [35] "HARDWARE EQUIPMENT AND SUPPLIES"
## [36] "HARDWARE STORES"
## [37] "DISCOUNT STORES"
## [38] "BOOKS PERIODICALS AND NEWSPAPERS"
## [39] "PUBLIC WAREHOUSING-FARM PROD REFRIDGERAT"
## [40] "STATIONARY OFFICE AND SCHOOL SUPPLY STO"
## [41] "HOLIDAY INNS"
## [42] "MISCELLANEOUS PUBLISHING & PRINTING"
## [43] "JEWELRY STORES-WATCHES, CLOCKES, AND SIL"
## [44] "QUICK-COPY AND REPRODUCTION SERVICES"
## [45] "DURABLE GOODS NOT ELSEWHERE CLASSIFIED"
## [46] "HILTON GARDEN INN"
## [47] "GOVERNMENT SERVICES NOT ELSEWHERE CLASSI"
## [48] "AUTO AND TRUCK DEALERS-(NEW&USED)-SALES"
## [49] "FAST FOOD RESTAURANTS"
## [50] "AUTOMOTIVE PARTS ACCESSORIES STORES"
## [51] "MOTOR FREIGHT CARRIERS, TRUCK-LOCAL/LONG"
## [52] "STATIONERY,OFFICE SUPPLIES,PRINTING AND"

```

## [53] "DRUGS DRUG PROPRIETARIES AND DRUGGIST S"  
## [54] "COURIER SERVICES-AIR OR GROUND FREIGHT F"  
## [55] "SOUTHWEST AIRLINES"  
## [56] "ELECTRONIC SALES"  
## [57] "COMPUTER SOFTWARE STORES"  
## [58] "ELECTRICAL PARTS AND EQUIPMENT"  
## [59] "COMPUTERS COMPUTER PERIPHERAL EQUIPMENT"  
## [60] "EATING PLACES, RESTAURANTS"  
## [61] "NURSERIES, LAWN AND GARDEN SUPPLY STORES"  
## [62] "UTILITIES-ELEC/GAS/HEAT OIL/SANITARY/WTR"  
## [63] "SPRINGHILL SUITES"  
## [64] "VARIETY STORES"  
## [65] "CATALOG MERCHANTS"  
## [66] "MISC FOOD STORES-SPECIALITY,CONVENIENCE,"  
## [67] "MISCELLANEOUS AUTOMOTIVE DEALERS"  
## [68] "SCANDINAVIAN AIRLINE SYSTEM (SAS)"  
## [69] "AUTOMOTIVE TIRE STORES"  
## [70] "OFFICE PHOTOGRAPHIC PHOTOCOPY AND MIC"  
## [71] "COSMETIC STORES"  
## [72] "NON-DURABLE GOODS NOT ELSEWHERE CLASSIFI"  
## [73] "CHARITABLE AND SOCIAL SERVICE ORGANIZATI"  
## [74] "PET SHOPS-PET FOOD AND SUPPLY STORES"  
## [75] "HAMPTON INNS HOTELS"  
## [76] "CIVIC, SOCIAL AND FRATERNAL ASSOCIATIONS"  
## [77] "COMMERCIAL PHOTOGRAPHY ART AND GRAPHIC"  
## [78] "CABLE, SATELLITE & OTHER PAY TV/RADIO SE"  
## [79] "ALL OTHER DIRECT MARKETERS"  
## [80] "PUBLIC WAREHOUSING-FARM PROD,REFRIDGERAT"  
## [81] "COMPUTER AND DATA PROCESSING SERVICES"  
## [82] "BAKERIES"  
## [83] "MISCELLANEOUS GENERAL MERCHANDISE STORES"  
## [84] "COMPUTER NETWORK/INFORMATION SERVICES"  
## [85] "EATING PLACES RESTAURANTS"  
## [86] "EMBASSY SUITES"  
## [87] "ORTHOPEDIC GOODS-ARTIFICIAL LIMB STORES"  
## [88] "HOUSEHOLD APPLIANCE STORES"  
## [89] "COMPUTERS,COMPUTER PERIPHERAL EQUIPMENT,"  
## [90] "OFFICE, PHOTOGRAPHIC, PHOTOCOPY, AND MIC"  
## [91] "COLLEGES UNIVERSITIES PROFESSIONAL SCH"  
## [92] "DIRECT MARKETING INSURANCE SERVICES"  
## [93] "TOURIST ATTRACTIONS AND EXHIBITS"  
## [94] "HOME2 SUITES"  
## [95] "NURSERIES LAWN AND GARDEN SUPPLY STORES"  
## [96] "ADVERTISING SERVICES"  
## [97] "RADIO, TELEVISION AND STEREO REPAIR SHOP"  
## [98] "SEWING, NEEDLEWORK, FABRIC AND PIECE GOO"  
## [99] "TESTING LABORATORIES (NON-MEDICAL)"  
## [100] "TRANSPORTATION SERVICES NOT ELSEWHERE CL"  
## [101] "DELTA"  
## [102] "AGRICULTURAL COOPERATIVES"  
## [103] "TAXICABS/LIMOUSINES"  
## [104] "POSTAGE STAMPS"

## [105] "DURABLE GOODS,NOT ELSEWHERE CLASSIFIED"  
## [106] "MISC FOOD STORES-SPECIALITY CONVENIENCE"  
## [107] "DEPARTMENT STORES"  
## [108] "HOBBY,TOY, AND GAME SHOPS"  
## [109] "STATIONERY OFFICE SUPPLIES PRINTING AND"  
## [110] "CIVIC SOCIAL AND FRATERNAL ASSOCIATIONS"  
## [111] "DIGITAL GOODS - MEDIA,BOOKS,MOVIES,MUSIC"  
## [112] "MEDICAL SERVICES & HEALTH PRACTITIONERS"  
## [113] "PIECE GOOD NOTIONS AND OTHER DRY GOODS"  
## [114] "MARRIOTT"  
## [115] "MEN'S AND WOMEN'S CLOTHING STORES"  
## [116] "ARTIST SUPPLY STORES CRAFT SHOPS"  
## [117] "BUSINESS AND SECRETARIAL SCHOOLS"  
## [118] "HEATING PLUMBING AIR CONDITIONING CONT"  
## [119] "TELECOMMUNICATION EQUIPMENT INCLUDING TE"  
## [120] "CAR WASHES"  
## [121] "TRAVEL AGENCIES"  
## [122] "CHEMICALS AND ALLIED PRODUCTS NOT ESLEWH"  
## [123] "MISCELLANEOUS PERSONAL SERVICES"  
## [124] "LA QUINTA INN AND SUITES"  
## [125] "LODGING HOTELS MOTELS RESORTS"  
## [126] "CLEANING AND MAINTENANCE JANITORIAL SER"  
## [127] "AIRLINES, AIR CARRIERS"  
## [128] "FUEL DEALERS-FUEL OIL, WOOD, COAL, LIQUE"  
## [129] "ELECTRICAL CONTRACTORS"  
## [130] "BEST WESTERN HOTELS"  
## [131] "STATIONARY, OFFICE AND SCHOOL SUPPLY STO"  
## [132] "HEATING, PLUMBING, AIR CONDITIONING CONT"  
## [133] "CONSTRUCTION MATERIALS"  
## [134] "MISCELLANEOUS HOUSE FURNISHING SPECIALIT"  
## [135] "ALITALIA"  
## [136] "SPORTING AND RECREATIONAL CAMPS"  
## [137] "COMMERCIAL EQUIPMENT NOT ELSEWHERE CLAS"  
## [138] "MISCELLENEOUS APPAREL AND ACCESSORY STOR"  
## [139] "METAL SERVICE CENTERS AND OFFICES"  
## [140] "DISINFECTING AND EXTERMINATING SERVICES"  
## [141] "RECORD SHOPS"  
## [142] "SPECIAL TRADE CONTRACTORS"  
## [143] "HYATT HOTELS"  
## [144] "ELECTRICAL AND SMALL APPLIANCE REPAIR SH"  
## [145] "COLLEGES, UNIVERSITIES, PROFESSIONAL SCH"  
## [146] "DIGITAL GOODS - APPS (EXCLUDES GAMES)"  
## [147] "CAMERA AND PHOTOGRAPHIC SUPPLY STORES"  
## [148] "GLASS, PAINT, WALLPAPER STORES"  
## [149] "PHOTOGRAPHIC STUDIOS"  
## [150] "HOBBY TOY AND GAME SHOPS"  
## [151] "TOLLS ROAD AND BRIDGE FEES"  
## [152] "BANDS ORCHESTRAS ENTERTAINERS"  
## [153] "MEN S WOMENS AND CHILDREN S UNIFORMS AND"  
## [154] "COMMERCIAL FURNITURE"  
## [155] "DISNEY RESORTS"  
## [156] "VETERINARY SERVICES"



## [157] "MUSIC STORES-MUSICAL INSTRUMENTS, PIANOS"  
## [158] "MICROTEL INN AND SUITES"  
## [159] "RESIDENCE INN"  
## [160] "FAMILY CLOTHING STORES"  
## [161] "TOLLS, ROAD AND BRIDGE FEES"  
## [162] "ENTERPRISE RENT-A-CAR"  
## [163] "MOTION PICTURE THEATERS"  
## [164] "CHILD CARE SERVICES"  
## [165] "AUTOMOTIVE REPAIR SHOPS (NON-DEALER)"  
## [166] "MANAGEMENT, CONSULTING AND PUBLIC RELATI"  
## [167] "MEN S AND WOMEN S CLOTHING STORES"  
## [168] "SHERATON"  
## [169] "WESTIN"  
## [170] "FLORISTS"  
## [171] "AUTOMOBILE PARKING LOTS AND GARAGES"  
## [172] "MISCELLANEOUS REPAIR SHOPS AND RELATED S"  
## [173] "MOTORCYCLE DEALERS"  
## [174] "IBERIA"  
## [175] "TRAILER PARKS AND CAMPGROUNDS"  
## [176] "BOOKS,PERIODICALS AND NEWSPAPERS"  
## [177] "BROADMOOR HOTEL"  
## [178] "TOWING SERVICES"  
## [179] "QUALITY INNS"  
## [180] "WHOLESALE CLUBS"  
## [181] "STENOGRAPHIC SERVICES"  
## [182] "RENAISSANCE HOTELS"  
## [183] "SPECIALITY CLEANING POLISHING & SANITATI"  
## [184] "CATERERS"  
## [185] "LODGING, HOTELS, MOTELS, RESORTS"  
## [186] "FURNITURE, HOME FURNISHINGS AND EQUIPMEN"  
## [187] "GIFT, CARD, NOVELTY, AND SOUVENIR SHOPS"  
## [188] "CROWNE PLAZA HOTELS"  
## [189] "BRITISH AIRWAYS - BRITISH A"  
## [190] "LARGE DIGITAL GOODS MERCHANT"  
## [191] "EQUIP RENTALS & LEASING SERVICES TOOL R"  
## [192] "DETECTIVE AGENCIES & PROTECTIVE AGENCY A"  
## [193] "GENERAL CONTRACTORS-RESIDENTIAL"  
## [194] "FLOOR COVERING, RUG AND CARPET STORES"  
## [195] "VIDEO RENTAL STORES"  
## [196] "OPTOMETRISTS, OPTHAMOLOGISTS"  
## [197] "AUTO AND TRUCK DEALERS-(NEW&USED)-SALES,"  
## [198] "INSURANCE-SALES & UNDERWRITING"  
## [199] "COURTYARD INNS"  
## [200] "HEARING AID SALES SERVICE SUPPLY STORE"  
## [201] "COMM. SPORTS PROF. SPORTS CLUBS ATHLETI"  
## [202] "RECREATIONAL AND UTILITY TRAILER, CAMPER"  
## [203] "COMMERCIAL PHOTOGRAPHY, ART, AND GRAPHIC"  
## [204] "FLOOR COVERING RUG AND CARPET STORES"  
## [205] "CLEANING AND MAINTENANCE, JANITORIAL SER"  
## [206] "GIFT CARD NOVELTY AND SOUVENIR SHOPS"  
## [207] "DETECTIVE AGENCIES & PROTECTIVE AGENCY,A"  
## [208] "VOCATIONAL AND TRADE SCHOOLS"

## [209] "PRECIOUS STONES AND METALS WATCHES & JE"  
## [210] "USED MERCHANDISE STORES, SECOND HAND STO"  
## [211] "RELIGIOUS GOODS STORES"  
## [212] "SPORTS APPAREL RIDING APPAREL STORES"  
## [213] "ELEMENTARY AND SECONDARY SCHOOLS"  
## [214] "DOCTORS PHYSICIANS"  
## [215] "FREEZER AND LOCKER MEAT PROVISIONERS"  
## [216] "SPECIALITY CLEANING,POLISHING & SANITATI"  
## [217] "INFORMATION RETRIEVAL SERVICES"  
## [218] "NORWEGIAN AIR SHUTTLE - NORWEGIAN"  
## [219] "THRIFTY RENT-A-CAR"  
## [220] "MEMBERSHIP CLUBS (SPORTS RECREATION ATHL"  
## [221] "PIECE GOOD,NOTIONS AND OTHER DRY GOODS"  
## [222] "LAUNDRY CLEANING AND GARMENT SERVICES"  
## [223] "CLARION HOTEL"  
## [224] "BANDS, ORCHESTRAS, ENTERTAINERS"  
## [225] "EMPLOYMENT AGENCIES, TEMPORARY HELP SUPP"  
## [226] "ACCOUNTING, AUDITING AND BOOKKEEPING SER"  
## [227] "FLORISTS SUPPLIES,NURSERY STOCK & FLOWER"  
## [228] "DIGITAL GOODS - GAMES"  
## [229] "DOUBLETREE"  
## [230] "EQUIP RENTALS & LEASING SERVICES, TOOL R"  
## [231] "AUTOMATED FUEL DISPENSERS"  
## [232] "MUSIC STORES-MUSICAL INSTRUMENTS PIANOS"  
## [233] "MEMBERSHIP CLUBS,(SPORTS,RECREATION,ATHL"  
## [234] "MOTOR FREIGHT CARRIERS TRUCK-LOCAL/LONG"  
## [235] "OMNI HOTELS"  
## [236] "PRECIOUS STONES AND METALS, WATCHES & JE"  
## [237] "BUDGET RENT-A-CAR"  
## [238] "AUTOMOTIVE BODY REPAIR SHOPS"  
## [239] "FLORISTS SUPPLIES NURSERY STOCK & FLOWER"  
## [240] "FURNITURE HOME FURNISHINGS AND EQUIPMEN"  
## [241] "WINDOW COVERING, UPHOLSTERY AND DRAPERY"  
## [242] "DRUG STORES, PHARMACIES"  
## [243] "THEATRICAL PRODUCERS (EXCEPT MOTION PICT"  
## [244] "COMFORT INNS"  
## [245] "NEWS DEALERS AND NEWSSTANDS"  
## [246] "TRUCK AND UTILITY TRAILER RENTAL"  
## [247] "PAINTS,VARNISHES AND SUPPLIES"  
## [248] "SEWING NEEDLEWORK FABRIC AND PIECE GOO"  
## [249] "HERTZ CORPORATION"  
## [250] "ANTIQUE SHOPS"  
## [251] "HILTON"  
## [252] "PACKAGE STORES BEER LIQUOR"  
## [253] "COMPUTER MAINTENANCE,REPAIR & SERVICES N"  
## [254] "REAL ESTATE AGENTS AND MANAGERS-RENTALS"  
## [255] "CIGAR STORES AND STANDS"  
## [256] "DRY CLEANERS"  
## [257] "DRURY INN"  
## [258] "ALASKA AIRLINES INC."  
## [259] "HYATT PLACE"  
## [260] "ARTIST SUPPLY STORES, CRAFT SHOPS"

## [261] "WYNDHAM"  
## [262] "WINDOW COVERING UPHOLSTERY AND DRAPERY"  
## [263] "RECREATIONAL AND UTILITY TRAILER CAMPER"  
## [264] "ENGINEERING, ARCHITECTURAL AND SURVEYING"  
## [265] "CANDY, NUT CONFECTIONERY STORES"  
## [266] "BOWLING ALLEYS"  
## [267] "AMUSEMENT RECREATION SERVICES (SWIMMING"  
## [268] "MANDALAY BAY RESORT"  
## [269] "AUTOMOBILE AND TRUCK DEALERS- (USED ONLY"  
## [270] "AVIS RENT-A-CAR"  
## [271] "COMPUTER MAINTENANCE REPAIR & SERVICES N"  
## [272] "CANDLEWOOD SUITES"  
## [273] "BICYCLE SHOPS-SALES AND SERVICE"  
## [274] "DAYS INNS"  
## [275] "JEWELRY STORES-WATCHES CLOCKS AND SIL"  
## [276] "PHOTOFINISHING LABORATORIES PHOTO DEVEL"  
## [277] "BUS LINES CHARTERS TOUR BUSES"  
## [278] "MARINAS, MARINE SERVICE/SUPPLIES"  
## [279] "MOTION PICTURE & VIDEO TAPE PRODUCTION &"  
## [280] "SWIMMING POOLS-SALES AND SUPPLIES"  
## [281] "SUPER 8 MOTELS"  
## [282] "BUS LINES, CHARTERS, TOUR BUSES"  
## [283] "RED ROOF INNS"  
## [284] "WOMEN'S ACCESSORY AND SPECIALITY SHOPS"  
## [285] "TAILORS SEAMSTRESS MENDING ALTERATION"  
## [286] "ALOFT"  
## [287] "MEN'S AND BOYS' CLOTHING AND ACCESSORIES"  
## [288] "QATAR AIRWAYS"  
## [289] "LAUNDRY, CLEANING, AND GARMENT SERVICES"  
## [290] "CONTRACTORS CONCRETE WORK"  
## [291] "SLEEP INN"  
## [292] "RELIGIOUS ORGANIZATIONS"  
## [293] "LEGAL SERVICES ATTORNEYS"  
## [294] "MEN S AND BOYS CLOTHING AND ACCESSORIES"  
## [295] "FUEL DEALERS-FUEL OIL WOOD COAL LIQUE"  
## [296] "AIR CANADA - AIR CAN"  
## [297] "DANCE HALLS, STUDIOS, AND SCHOOLS"  
## [298] "LEGAL SERVICES, ATTORNEYS"  
## [299] "WOMEN S READY TO WEAR STORES"  
## [300] "SPORTS APPAREL, RIDING APPAREL STORES"  
## [301] "ENGINEERING ARCHITECTURAL AND SURVEYING"  
## [302] "VIDEO AMUSEMENT GAME SUPPLIES"  
## [303] "THE RITZ-CARLTON"  
## [304] "NATIONAL CAR RENTAL"  
## [305] "CONTRACTORS, CONCRETE WORK"  
## [306] "HEARING AID SALES, SERVICE, SUPPLY STORE"  
## [307] "EMBASSY HOTELS"  
## [308] "BILLIARD AND POOL ESTABLISHMENTS"  
## [309] "AUTOMOTIVE PAINT SHOPS"  
## [310] "LOT (POLISH AIRLINES)"  
## [311] "ROOFING & SIDING, SHEET METAL WORK"  
## [312] "INBOUND TELESERVICES MERCHANTS"

## [313] "OPTICIANS"  
## [314] "EMPLOYMENT AGENCIES TEMPORARY HELP SUPP"  
## [315] "FUNERAL SERVICE AND CREMATORIES"  
## [316] "ECONOLODGES"  
## [317] "CLOTHING RENTAL-COSTUMES/UNIFORMS"  
## [318] "ACCOUNTING AUDITING AND BOOKKEEPING SER"  
## [319] "VIRGIN ATLANTIC"  
## [320] "ROOFING & SIDING SHEET METAL WORK"  
## [321] "LOCAL SUBURBAN COMMUTER PASSENGER TRANSP"  
## [322] "RADIO TELEVISION AND STEREO REPAIR SHOP"  
## [323] "GOLF COURSES-PUBLIC"  
## [324] "USED MERCHANDISE STORES SECOND HAND STO"  
## [325] "ALAMO RENT-A-CAR"  
## [326] "ART DEALERS AND GALLERIES"  
## [327] "DIGITAL GOODS - MEDIA BOOKS MOVIES MUSIC"  
## [328] "CARPET AND UPHOLSTERY CLEANING"  
## [329] "WELDING"  
## [330] "AIR CONDITIONING AND REFRIGERATION REPAI"  
## [331] "PASSENGER RAILWAYS"  
## [332] "HOTEL IBIS"  
## [333] "DOCTORS, PHYSICIANS"  
## [334] "SHOE STORES"  
## [335] "COMM. SPORTS, PROF. SPORTS CLUBS,ATHLETI"  
## [336] "BOAT DEALERS"  
## [337] "COMPANIA PANAMENA DE AVIACION (COPA)"  
## [338] "AMERICAS BEST VALUE INN"  
## [339] "DRINKING PLACES (ALCOHOLIC BEV.)-BARS TA"  
## [340] "TAILORS, SEAMSTRESS, MENDING, ALTERATION"  
## [341] "HOSPITALS"  
## [342] "HILTON INTERNATIONAL"  
## [343] "AUTOGRAPH"  
## [344] "WOMEN S ACCESSORY AND SPECIALITY SHOPS"  
## [345] "AMUSEMENT,RECREATION SERVICES (SWIMMING"  
## [346] "FRONTIER AIRLINES - FRONTIER"  
## [347] "TIRE RETREADING AND REPAIR SHOPS"  
## [348] "MOTEL 6"  
## [349] "MASSAGE PARLORS"  
## [350] "MARRIOT/TOWNPLACE SUITES"  
## [351] "MARINAS MARINE SERVICE/SUPPLIES"  
## [352] "EVA AIRLINES"  
## [353] "ELEMENT"  
## [354] "FIREPLACE FIREPLACE SCREENS AND ACCESS"  
## [355] "PETROLEUM AND PETROLEUM PRODUCTS"  
## [356] "PAINTS VARNISHES AND SUPPLIES"  
## [357] "CANDY NUT CONFECTIONERY STORES"  
## [358] "MIRAGE HOTEL AND CASINO"  
## [359] "WESTJETAIR"  
## [360] "BEAU RIVAGE HOTEL AND CASINO"  
## [361] "GAYLORD OPRYLAND"  
## [362] "AMERICINN"  
## [363] "WOMEN'S READY TO WEAR STORES"  
## [364] "HOTEL INDIGO"

```
## [365] "AIRPORTS, FLYING FIELDS, AIRPORT TERMINA"
## [366] "JETBLUE AIRLINES"
## [367] "LUXOR HOTEL AND CASINO"
## [368] "ARIA"
## [369] "RADISSON"
## [370] "AVIANCA"
## [371] "AMUSEMENT PARKS CIRCUSES CARNIVALS FO"
## [372] "ROSEN HOTELS & RESORTS"
## [373] "LUGGAGE AND LEATHER GOODS STORES"
## [374] "STAYBRIDGE SUITES"
## [375] "MOTOR HOME AND RECREATIONAL VEHICLE RENT"
## [376] "CORRESPONDENCE SCHOOLS"
## [377] "OPTOMETRISTS OPTHAMOLOGISTS"
## [378] "RAMADA INNS"
## [379] "PHOTOFINISHING LABORATORIES, PHOTO DEVEL"
## [380] "BUYING/SHOPPING SERVICES CLUBS"
## [381] "TYPESETTING PLATE MAKING & RELATED SERVI"
## [382] "HOMEWOOD SUITES"
## [383] "FINNAIR"
## [384] "DANCE HALLS STUDIOS AND SCHOOLS"
## [385] "COMMERCIAL FOOTWEAR"
## [386] "COUNSELING SERVICE-DEBT MARRIAGE PERSO"
## [387] "BEAUTY SHOPS AND BARBER SHOPS"
## [388] "FINES"
## [389] "AIR NEW ZEALAND LTD. INTERNATIONAL"
## [390] "INTERCONTINENTAL HOTELS"
## [391] "LAUNDRIES SERVICES-FAMILY AND COMMERCIAL"
## [392] "SN BRUSSELS AIRLINES - SNBRU AIR"
## [393] "HEALTH AND BEAUTY SPAS"
## [394] "SHOE REPAIR SHOPS, SHOE SHINE PARLORS &"
```

## Combine some codes

Combine codes by trimming spaces and eliminating punctuation.

```
# Trim white spaces
df2_group <- df_group %>% mutate(MCC2 = str_trim(MCC, side='both'))

# Replace punctuation with ''
df2_group <- df2_group %>% mutate(MCC2 = str_replace_all(MCC2, '[:punct:]', ''))

n_distinct(df2_group$MCC2)
```

```
## [1] 341
```

```
sort(unique(df2_group$MCC2))
```

```
## [1] "ACCOUNTING AUDITING AND BOOKKEEPING SER"
## [2] "ADVERTISING SERVICES"
## [3] "AGRICULTURAL COOPERATIVES"
## [4] "AIR CANADA AIR CAN"
## [5] "AIR CONDITIONING AND REFRIGERATION REPAI"
## [6] "AIR NEW ZEALAND LTD INTERNATIONAL"
## [7] "AIRLINES AIR CARRIERS"
## [8] "AIRPORTS FLYING FIELDS AIRPORT TERMINA"
## [9] "ALAMO RENTACAR"
## [10] "ALASKA AIRLINES INC"
## [11] "ALITALIA"
## [12] "ALL OTHER DIRECT MARKETERS"
## [13] "ALOFT"
## [14] "AMERICAN AIRLINES"
## [15] "AMERICAS BEST VALUE INN"
## [16] "AMERICINN"
## [17] "AMUSEMENT PARKS CIRCUSES CARNIVALS FO"
## [18] "AMUSEMENT RECREATION SERVICES SWIMMING"
## [19] "AMUSEMENTRECREATION SERVICES SWIMMING"
## [20] "ANTIQUE SHOPS"
## [21] "ARIA"
## [22] "ART DEALERS AND GALLERIES"
## [23] "ARTIST SUPPLY STORES CRAFT SHOPS"
## [24] "AUTO AND TRUCK DEALERSNEWUSEDSESALES"
## [25] "AUTOGRAPH"
## [26] "AUTOMATED FUEL DISPENSERS"
## [27] "AUTOMOBILE AND TRUCK DEALERS USED ONLY"
## [28] "AUTOMOBILE PARKING LOTS AND GARAGES"
## [29] "AUTOMOTIVE BODY REPAIR SHOPS"
## [30] "AUTOMOTIVE PAINT SHOPS"
## [31] "AUTOMOTIVE PARTS ACCESSORIES STORES"
## [32] "AUTOMOTIVE REPAIR SHOPS NONDEALER"
## [33] "AUTOMOTIVE TIRE STORES"
## [34] "AVIANCA"
## [35] "AVIS RENTACAR"
## [36] "BAKERIES"
## [37] "BANDS ORCHESTRAS ENTERTAINERS"
## [38] "BEAU RIVAGE HOTEL AND CASINO"
## [39] "BEAUTY SHOPS AND BARBER SHOPS"
## [40] "BEST WESTERN HOTELS"
## [41] "BICYCLE SHOPSSALES AND SERVICE"
## [42] "BILLIARD AND POOL ESTABLISHMENTS"
## [43] "BOAT DEALERS"
## [44] "BOOK STORES"
## [45] "BOOKS PERIODICALS AND NEWSPAPERS"
## [46] "BOOKSPERIODICALS AND NEWSPAPERS"
## [47] "BOWLING ALLEYS"
## [48] "BRITISH AIRWAYS BRITISH A"
## [49] "BROADMOOR HOTEL"
## [50] "BUDGET RENTACAR"
## [51] "BUS LINES CHARTERS TOUR BUSES"
## [52] "BUSINESS AND SECRETARIAL SCHOOLS"
```

```

## [53] "BUSINESS SERVICES NOT ELSEWHERE CLASSIFI"
## [54] "BUYINGSHOPPING SERVICES CLUBS"
## [55] "CABLE SATELLITE  OTHER PAY TVRADIO SE"
## [56] "CAMERA AND PHOTOGRAPHIC SUPPLY STORES"
## [57] "CANDLEWOOD SUITES"
## [58] "CANDY NUT CONFECTIONERY STORES"
## [59] "CAR WASHES"
## [60] "CARPET AND UPHOLSTERY CLEANING"
## [61] "CATALOG MERCHANTS"
## [62] "CATERERS"
## [63] "CHARITABLE AND SOCIAL SERVICE ORGANIZATI"
## [64] "CHEMICALS AND ALLIED PRODUCTS NOT ESLEWH"
## [65] "CHILD CARE SERVICES"
## [66] "CIGAR STORES AND STANDS"
## [67] "CIVIC SOCIAL AND FRATERNAL ASSOCIATIONS"
## [68] "CLARION HOTEL"
## [69] "CLEANING AND MAINTENANCE JANITORIAL SER"
## [70] "CLOTHING RENTALCOSTUMESUNIFORMS"
## [71] "COLLEGES UNIVERSITIES PROFESSIONAL SCH"
## [72] "COMBINATION CATALOG AND RETAIL MERCHANT"
## [73] "COMFORT INNS"
## [74] "COMM SPORTS PROF SPORTS CLUBS ATHLETI"
## [75] "COMM SPORTS PROF SPORTS CLUBSATHLETI"
## [76] "COMMERCIAL EQUIPMENT NOT ELSEWHERE CLAS"
## [77] "COMMERCIAL FOOTWEAR"
## [78] "COMMERCIAL FURNITURE"
## [79] "COMMERCIAL PHOTOGRAPHY ART AND GRAPHIC"
## [80] "COMPANIA PANAMENA DE AVIACION COPA"
## [81] "COMPUTER AND DATA PROCESSING SERVICES"
## [82] "COMPUTER MAINTENANCE REPAIR  SERVICES N"
## [83] "COMPUTER MAINTENANCEREPAIR  SERVICES N"
## [84] "COMPUTER NETWORKINFORMATION SERVICES"
## [85] "COMPUTER SOFTWARE STORES"
## [86] "COMPUTERS COMPUTER PERIPHERAL EQUIPMENT"
## [87] "COMPUTERSCOMPUTER PERIPHERAL EQUIPMENT"
## [88] "CONSTRUCTION MATERIALS"
## [89] "CONTINUITYSUBSCRIPTION MERCHANTS"
## [90] "CONTRACTORS CONCRETE WORK"
## [91] "CORRESPONDENCE SCHOOLS"
## [92] "COSMETIC STORES"
## [93] "COUNSELING SERVICEDEBT MARRIAGE PERSO"
## [94] "COURIER SERVICESAIR OR GROUND FREIGHT F"
## [95] "COURIER SERVICESAIR OR GROUNDFREIGHT F"
## [96] "COURTYARD INNS"
## [97] "CROWNE PLAZA HOTELS"
## [98] "DANCE HALLS STUDIOS AND SCHOOLS"
## [99] "DAYS INNS"
## [100] "DELTA"
## [101] "DEPARTMENT STORES"
## [102] "DETECTIVE AGENCIES  PROTECTIVE AGENCY A"
## [103] "DETECTIVE AGENCIES  PROTECTIVE AGENCYA"
## [104] "DIGITAL GOODS  APPS EXCLUDES GAMES"

```

## [105] "DIGITAL GOODS GAMES"  
## [106] "DIGITAL GOODS MEDIA BOOKS MOVIES MUSIC"  
## [107] "DIGITAL GOODS MEDIABOOKSMOVIESMUSIC"  
## [108] "DIRECT MARKETING INSURANCE SERVICES"  
## [109] "DISCOUNT STORES"  
## [110] "DISINFECTING AND EXTERMINATING SERVICES"  
## [111] "DISNEY RESORTS"  
## [112] "DOCTORS PHYSICIANS"  
## [113] "DOUBLETREE"  
## [114] "DRINKING PLACES ALCOHOLIC BEVBARS TA"  
## [115] "DRUG STORES PHARMACIES"  
## [116] "DRUGS DRUG PROPRIETARIES AND DRUGGIST S"  
## [117] "DRUGSDRUG PROPRIETARIES AND DRUGGISTS"  
## [118] "DRURY INN"  
## [119] "DRY CLEANERS"  
## [120] "DURABLE GOODS NOT ELSEWHERE CLASSIFIED"  
## [121] "DURABLE GOODSNOT ELSEWHERE CLASSIFIED"  
## [122] "EATING PLACES RESTAURANTS"  
## [123] "ECONOLODGES"  
## [124] "ELECTRICAL AND SMALL APPLIANCE REPAIR SH"  
## [125] "ELECTRICAL CONTRACTORS"  
## [126] "ELECTRICAL PARTS AND EQUIPMENT"  
## [127] "ELECTRONIC SALES"  
## [128] "ELEMENT"  
## [129] "ELEMENTARY AND SECONDARY SCHOOLS"  
## [130] "EMBASSY HOTELS"  
## [131] "EMBASSY SUITES"  
## [132] "EMPLOYMENT AGENCIES TEMPORARY HELP SUPP"  
## [133] "ENGINEERING ARCHITECTURAL AND SURVEYING"  
## [134] "ENTERPRISE RENTACAR"  
## [135] "EQUIP RENTALS LEASING SERVICES TOOL R"  
## [136] "EVA AIRLINES"  
## [137] "FAIRFIELD INN"  
## [138] "FAMILY CLOTHING STORES"  
## [139] "FAST FOOD RESTAURANTS"  
## [140] "FINES"  
## [141] "FINNAIR"  
## [142] "FIREPLACE FIREPLACE SCREENS AND ACCESS"  
## [143] "FLOOR COVERING RUG AND CARPET STORES"  
## [144] "FLORISTS"  
## [145] "FLORISTS SUPPLIES NURSERY STOCK FLOWER"  
## [146] "FLORISTS SUPPLIESNURSERY STOCK FLOWER"  
## [147] "FREEZER AND LOCKER MEAT PROVISIONERS"  
## [148] "FRONTIER AIRLINES FRONTIER"  
## [149] "FUEL DEALERSFUEL OIL WOOD COAL LIQUE"  
## [150] "FUNERAL SERVICE AND CREMATORIES"  
## [151] "FURNITURE HOME FURNISHINGS AND EQUIPMEN"  
## [152] "GAYLORD OPRYLAND"  
## [153] "GENERAL CONTRACTORSRESIDENTIAL"  
## [154] "GIFT CARD NOVELTY AND SOUVENIR SHOPS"  
## [155] "GLASS PAINT WALLPAPER STORES"  
## [156] "GOLF COURSEPUBLIC"



## [157] "GOVERNMENT SERVICES NOT ELSEWHERE CLASSI"  
## [158] "GROCERY STORES SUPERMARKETS"  
## [159] "HAMPTON INNS HOTELS"  
## [160] "HARDWARE EQUIPMENT AND SUPPLIES"  
## [161] "HARDWARE STORES"  
## [162] "HEALTH AND BEAUTY SPAS"  
## [163] "HEARING AID SALES SERVICE SUPPLY STORE"  
## [164] "HEATING PLUMBING AIR CONDITIONING CONT"  
## [165] "HERTZ CORPORATION"  
## [166] "HILTON"  
## [167] "HILTON GARDEN INN"  
## [168] "HILTON INTERNATIONAL"  
## [169] "HOBBY TOY AND GAME SHOPS"  
## [170] "HOBBYTOY AND GAME SHOPS"  
## [171] "HOLIDAY INNS"  
## [172] "HOME SUPPLY WAREHOUSE STORES"  
## [173] "HOME2 SUITES"  
## [174] "HOMWOOD SUITES"  
## [175] "HOSPITALS"  
## [176] "HOTEL IBIS"  
## [177] "HOTEL INDIGO"  
## [178] "HOUSEHOLD APPLIANCE STORES"  
## [179] "HYATT HOTELS"  
## [180] "HYATT PLACE"  
## [181] "IBERIA"  
## [182] "INBOUND TELESERVICES MERCHANTS"  
## [183] "INDUSTRIAL SUPPLIES NOT ESLEWHERE CLASSI"  
## [184] "INFORMATION RETRIEVAL SERVICES"  
## [185] "INSURANCESALES UNDERWRITING"  
## [186] "INTERCONTINENTAL HOTELS"  
## [187] "JETBLUE AIRLINES"  
## [188] "JEWELRY STORESWATCHES CLOCKES AND SIL"  
## [189] "LA QUINTA INN AND SUITES"  
## [190] "LABMEDICALDENTALOPHTHALMIC HOSPITAL E"  
## [191] "LANDSCAPE AND HORTICULTURAL SERVICES"  
## [192] "LARGE DIGITAL GOODS MERCHANT"  
## [193] "LAUNDRIES SERVICESFAMILY AND COMMERCIAL"  
## [194] "LAUNDRY CLEANING AND GARMENT SERVICES"  
## [195] "LEGAL SERVICES ATTORNEYS"  
## [196] "LOCAL SUBURBAN COMMUTER PASSENGER TRANSP"  
## [197] "LODGING HOTELS MOTELS RESORTS"  
## [198] "LOT POLISH AIRLINES"  
## [199] "LUGGAGE AND LEATHER GOODS STORES"  
## [200] "LUMBER AND BUILDING MATERIALS STORES"  
## [201] "LUXOR HOTEL AND CASINO"  
## [202] "MANAGEMENT CONSULTING AND PUBLIC RELATI"  
## [203] "MANDALAY BAY RESORT"  
## [204] "MARINAS MARINE SERVICESUPPLIES"  
## [205] "MARRIOTT"  
## [206] "MARRIOTTOWNPLACE SUITES"  
## [207] "MASSAGE PARLORS"  
## [208] "MEDICAL AND DENTAL LABORATORIES"

## [209] "MEDICAL SERVICES HEALTH PRACTITIONERS"  
## [210] "MEMBERSHIP CLUBS SPORTS RECREATION ATHL"  
## [211] "MEMBERSHIP CLUBSSPORTSRECREATIONATHL"  
## [212] "MEMBERSHIP ORGANIZATIONS NOT ELSEWHERE C"  
## [213] "MEN S AND BOYS CLOTHING AND ACCESSORIES"  
## [214] "MEN S AND WOMEN S CLOTHING STORES"  
## [215] "MEN S WOMENS AND CHILDREN S UNIFORMS AND"  
## [216] "MENS AND BOYS CLOTHING AND ACCESSORIES"  
## [217] "MENS AND WOMENS CLOTHING STORES"  
## [218] "MENSWOMENSAND CHILDRENS UNIFORMS AND"  
## [219] "METAL SERVICE CENTERS AND OFFICES"  
## [220] "MICROTEL INN AND SUITES"  
## [221] "MIRAGE HOTEL AND CASINO"  
## [222] "MISC FOOD STORESSPECIALITY CONVENIENCE"  
## [223] "MISC FOOD STORESSPECIALITYCONVENIENCE"  
## [224] "MISCELLANEOUS AND RETAIL STORES"  
## [225] "MISCELLANEOUS AUTOMOTIVE DEALERS"  
## [226] "MISCELLANEOUS GENERAL MERCHANDISE STORES"  
## [227] "MISCELLANEOUS HOUSE FURNISHING SPECIALIT"  
## [228] "MISCELLANEOUS PERSONAL SERVICES"  
## [229] "MISCELLANEOUS PUBLISHING PRINTING"  
## [230] "MISCELLANEOUS REPAIR SHOPS AND RELATED S"  
## [231] "MISCELLENEOUS APPAREL AND ACCESSORY STOR"  
## [232] "MOTEL 6"  
## [233] "MOTION PICTURE VIDEO TAPE PRODUCTION "  
## [234] "MOTION PICTURE THEATERS"  
## [235] "MOTOR FREIGHT CARRIERS TRUCKLOCALLONG"  
## [236] "MOTOR HOME AND RECREATIONAL VEHICLE RENT"  
## [237] "MOTOR VEHICLE SUPPLIES AND NEW PARTS"  
## [238] "MOTORCYCLE DEALERS"  
## [239] "MUSIC STORESMUSICAL INSTRUMENTS PIANOS"  
## [240] "NATIONAL CAR RENTAL"  
## [241] "NEWS DEALERS AND NEWSSTANDS"  
## [242] "NONDURABLE GOODS NOT ELSEWHERE CLASSIFI"  
## [243] "NORWEGIAN AIR SHUTTLE NORWEGIAN"  
## [244] "NURSERIES LAWN AND GARDEN SUPPLY STORES"  
## [245] "OFFICE PHOTOGRAPHIC PHOTOCOPY AND MIC"  
## [246] "OMNI HOTELS"  
## [247] "OPTICIANS"  
## [248] "OPTOMETRISTS OPTHAMOLOGISTS"  
## [249] "ORTHOPEDIC GOODSARTIFICIAL LIMB STORES"  
## [250] "PACKAGE STORES BEER LIQUOR"  
## [251] "PAINTS VARNISHES AND SUPPLIES"  
## [252] "PAINTSVARNISHES AND SUPPLIES"  
## [253] "PASSENGER RAILWAYS"  
## [254] "PET SHOPSPET FOOD AND SUPPLY STORES"  
## [255] "PETROLEUM AND PETROLEUM PRODUCTS"  
## [256] "PHOTOFINISHING LABORATORIES PHOTO DEVEL"  
## [257] "PHOTOGRAPHIC STUDIOS"  
## [258] "PIECE GOOD NOTIONS AND OTHER DRY GOODS"  
## [259] "PIECE GOODNOTIONS AND OTHER DRY GOODS"  
## [260] "PLUMBING HEATING EQUIPMENT AND SUPPLIE"

## [261] "POSTAGE STAMPS"  
## [262] "PRECIOUS STONES AND METALS WATCHES JE"  
## [263] "PROFESSIONAL SERVICES NOT ELSEWHERE CLAS"  
## [264] "PUBLIC WAREHOUSINGFARM PROD REFRIDGERAT"  
## [265] "PUBLIC WAREHOUSINGFARM PRODREFRIDGERAT"  
## [266] "QATAR AIRWAYS"  
## [267] "QUALITY INNS"  
## [268] "QUICKCOPY AND REPRODUCTION SERVICES"  
## [269] "RADIO TELEVISION AND STEREO REPAIR SHOP"  
## [270] "RADISSON"  
## [271] "RAMADA INNS"  
## [272] "REAL ESTATE AGENTS AND MANAGERSRENTALS"  
## [273] "RECORD SHOPS"  
## [274] "RECREATIONAL AND UTILITY TRAILER CAMPER"  
## [275] "RED ROOF INNS"  
## [276] "RELIGIOUS GOODS STORES"  
## [277] "RELIGIOUS ORGANIZATIONS"  
## [278] "RENAISSANCE HOTELS"  
## [279] "RESIDENCE INN"  
## [280] "ROOFING SIDING SHEET METAL WORK"  
## [281] "ROSEN HOTELS RESORTS"  
## [282] "SCANDINAVIAN AIRLINE SYSTEM SAS"  
## [283] "SCHOOLS AND EDUCATIONAL SERVICES NOT ELS"  
## [284] "SERVICE STATIONS"  
## [285] "SEWING NEEDLEWORK FABRIC AND PIECE GOO"  
## [286] "SHERATON"  
## [287] "SHOE REPAIR SHOPS SHOE SHINE PARLORS "  
## [288] "SHOE STORES"  
## [289] "SLEEP INN"  
## [290] "SN BRUSSELS AIRLINES SNBRU AIR"  
## [291] "SOUTHWEST AIRLINES"  
## [292] "SPECIAL TRADE CONTRACTORS"  
## [293] "SPECIALITY CLEANING POLISHING SANITATI"  
## [294] "SPECIALITY CLEANINGPOLISHING SANITATI"  
## [295] "SPORTING AND RECREATIONAL CAMPS"  
## [296] "SPORTING GOODS STORES"  
## [297] "SPORTS APPAREL RIDING APPAREL STORES"  
## [298] "SPRINGHILL SUITES"  
## [299] "STATIONARY OFFICE AND SCHOOL SUPPLY STO"  
## [300] "STATIONERY OFFICE SUPPLIES PRINTING AND"  
## [301] "STATIONERYOFFICE SUPPLIESPRINTING AND"  
## [302] "STAYBRIDGE SUITES"  
## [303] "STENOGRAPHIC SERVICES"  
## [304] "SUPER 8 MOTELS"  
## [305] "SWIMMING POOLSSALES AND SUPPLIES"  
## [306] "TAILORS SEAMSTRESS MENDING ALTERATION"  
## [307] "TAXICABSLIMOUSINES"  
## [308] "TELECOMMUNICATION EQUIPMENT INCLUDING TE"  
## [309] "TELECOMMUNICATION SERVINCLUDE LOCALLD"  
## [310] "TESTING LABORATORIES NONMEDICAL"  
## [311] "THE RITZCARLTON"  
## [312] "THEATRICAL PRODUCERS EXCEPT MOTION PICT"

```
## [313] "THRIFTY RENTACAR"
## [314] "TIRE RETREADING AND REPAIR SHOPS"
## [315] "TOLLS ROAD AND BRIDGE FEES"
## [316] "TOURIST ATTRACTIONS AND EXHIBITS"
## [317] "TOWING SERVICES"
## [318] "TRAILER PARKS AND CAMPGROUNDS"
## [319] "TRANSPORTATION SERVICES NOT ELSEWHERE CL"
## [320] "TRAVEL AGENCIES"
## [321] "TRUCK AND UTILITY TRAILER RENTAL"
## [322] "TYPESETTING PLATE MAKING RELATED SERVI"
## [323] "UNITED AIRLINES"
## [324] "USED MERCHANDISE STORES SECOND HAND STO"
## [325] "UTILITIESELEGASHEAT OILSANITARYWTR"
## [326] "VARIETY STORES"
## [327] "VETERINARY SERVICES"
## [328] "VIDEO AMUSEMENT GAME SUPPLIES"
## [329] "VIDEO RENTAL STORES"
## [330] "VIRGIN ATLANTIC"
## [331] "VOCATIONAL AND TRADE SCHOOLS"
## [332] "WELDING"
## [333] "WESTIN"
## [334] "WESTJETAIR"
## [335] "WHOLESALE CLUBS"
## [336] "WINDOW COVERING UPHOLSTERY AND DRAPERY"
## [337] "WOMEN S ACCESSORY AND SPECIALITY SHOPS"
## [338] "WOMEN S READY TO WEAR STORES"
## [339] "WOMENS ACCESSORY AND SPECIALITY SHOPS"
## [340] "WOMENS READY TO WEAR STORES"
## [341] "WYNDHAM"
```

## List of problematic MCCs

```
mcc_list_group = c('ART DEALERS AND GALLERIES',
  'AUTOMOTIVE PAINT SHOPS',
  'AUTOMOTIVE BODY REPAIR SHOPS',
  'BILLIARD AND POOL ESTABLISHMENTS',
  'BOWLING ALLEYS',
  'CHILD CARE SERVICES',
  'COSMETIC STORES',
  'DOCTORS PHYSICIANS',
  'DISNEY RESORTS',
  'FUNERAL SERVICE AND CREMATORIES',
  'HOSPITALS',
  'JEWELRY STORESWATCHES CLOCKES AND SIL',
  'MANDALAY BAY RESORT',
  'MIRAGE HOTEL AND CASINO',
  'PET SHOPSPET FOOD AND SUPPLY STORES',
  'SWIMMING POOLSSALES AND SUPPLIES',
  'VIDEO AMUSEMENT GAME SUPPLIES',
  'WHOLESALE CLUBS',
  'TRAILER PARKS AND CAMPGROUNDS',
  'OPTOMETRISTS OPTHAMOLOGISTS',
  'VIDEO RENTAL STORES',
  'VETERINARY SERVICES')
```

## Problematic transactions

```
# Keep only problematic ones
mcc_problems_group <- df2_group %>% filter(MCC2 %in% mcc_list_group)

# view
mcc_problems_group %>%
  select(Cardholder.Name, Item.Description, Amount, Merchant, MCC2) %>%
  arrange(MCC2, Merchant, desc(Amount))
```

Cardholder.Name <chr>	Item.Description <chr>	Amount <dbl>
LONGAN, S	GENERAL PURCHASE	275.00
COOPER, S	FRAMED 20 X 24 GICLEE PRIN NMB	300.00
TECKEMEYER, J	GENERAL PURCHASE	75.00
TECKEMEYER, J	GENERAL PURCHASE	35.00
PESTRAK, S	GENERAL PURCHASE	3000.00
PHILLIPS, K	GENERAL PURCHASE	1347.35
HUNT, K	GENERAL PURCHASE	3020.00

Cardholder.Name	Item.Description	Amount
<chr>	<chr>	<dbl>
HUNT, K	GENERAL PURCHASE	1000.00
HOFFMAN, C	GENERAL PURCHASE	900.00
WASHBURN, M	GENERAL PURCHASE	117.47

1-10 of 954 rows | 1-3 of 5 columns

Previous 1 2 3 4 5 6 ... 96 Next

## Especially problematic transactions

# *Especially problematic*

```
mcc_problems_group %>% select(Cardholder.Name, Item.Description, Amount, Merchant,
Transaction.Date, MCC2) %>%
  filter(Amount > 2100) %>%
  arrange(desc(Amount))
```

Cardholder.Name	Item.Description	Amount	Merchant	Tran
<chr>	<chr>	<dbl>	<chr>	
PERRY, S	ROOM CHARGES	12395.38	DISNEY RESORTS-RESE	
DAVIS, M	GENERAL PURCHASE	4857.29	TRI-C CLUB SUPPLY INC	
MASSEY, J	GENERAL PURCHASE	4383.45	CIMARRON VALLEY COOP	
ROBISON, C	GENERAL PURCHASE	4290.00	CATTLE STATS LLC	
LEISTER, J	GENERAL PURCHASE	3894.90	BIOZYME INCORPORATED	
LEISTER, J	GENERAL PURCHASE	3865.00	BIOZYME INCORPORATED	
LEISTER, J	GENERAL PURCHASE	3310.00	BIOZYME INCORPORATED	
DAVIS, G	GENERAL PURCHASE	3100.00	IN PREFERRED PDR	
PHILLIPS, K	GENERAL PURCHASE	3091.76	G & M BODY SHOP WEST	
HUNT, K	GENERAL PURCHASE	3020.00	BOBSSPEEDOMETERAUTOMO	

1-10 of 20 rows | 1-5 of 6 columns

Previous 1 2 Next

## 👉 1b: How the top 10-20 transactions were identified

We compiled 2 separate tables of top problematic transactions: 1 using the regex technique and 1 using MCC filters. For the regex technique, we first identified potentially problematic words (such as Insurance, Uniform, Flowers, Fee, and Fuel) based on our interpretation of the OSU P-card guidelines. We then eliminated likely false positives that included safe words including "Agent Fee", "Mask", "Publish" and "Advertising" as we concluded these could be categorized as reasonable and allowed business transactions. There were numerous 5-10 dollars agent fees that seemed acceptable especially if OSU works with a dedicated travel agency for employee travel. The other safe words seemed to be potentially OK as they were likely business-related (publishing or advertising expenses) or due to the Covid-19 pandemic (mask-related or telemedicine expenses). The final step was to filter for amounts > 1500 dollars. For the MCC filters, we independently identified potentially problematic MCC codes by scanning the complete list of codes (sorted A-Z). We determined the MCC codes (Bowling Alleys, Cosmetic Stores, etc.) given they are unlikely to be serving a business purpose. Other MCC codes were included in the list group Wholesale Clubs especially since the P-Card guidelines specifically prohibit purchases related to Amazon Prime and Sam's Club memberships. The final step was to filter for amounts > 2100 dollars.

## 👉 1c: Questions to ask management to determine if these violate PCard controls

We created a list of questions we would ask management to determine if these transactions are indeed problematic and violations of the P-card controls. The questions would include:

1. When are entertainment and recreation purchases allowed? Is there are maximum dollar amount allowed? (for example, for employee relations or team events). For example, we identified high value purchase at a Billiards establishment.
2. Do theme park tickets or resorts, for example Disney, qualify as valid purchases?
3. When are swimming pool-related purchases allowed?
4. Regarding the uniform purchases: Do you have a list of approved taxability forms for non-taxable clothing?
5. For the pet food and supply stores & veterinary services: Are there any animals at the university that need to be cared for (vet clinic, vet program, etc) that would qualify as valid purchases?
6. When are art dealer and gallery purchases allowed? For example, is this allowed for university buildings? Is this covered in the "Decorations" section of the P-card prohibited purchases list?
7. When are automotive body repair shop purchases allowed?
8. When are aircraft fuel purchases allowed?
9. Is membership to APA.org a permitted institutional purchase?



## Part 2 - Apple's MD&A Sentiment Analysis

### 👉 2a: MD&A text sentiment

```
mda <- read_file("MDA.txt")
nchar(mda)
```

```
## [1] 2693
```

# Tokenize the text

```
# Change to a tibble (tidy dataframe)
tokens_mda <- tibble(mda)

# Tokenize
tokens_mda <- tokens_mda %>% unnest_tokens(output=word, input=mda, token='words', to_lower=TRUE)

# add order of the words
tokens_mda <- tokens_mda %>% mutate(order = row_number())

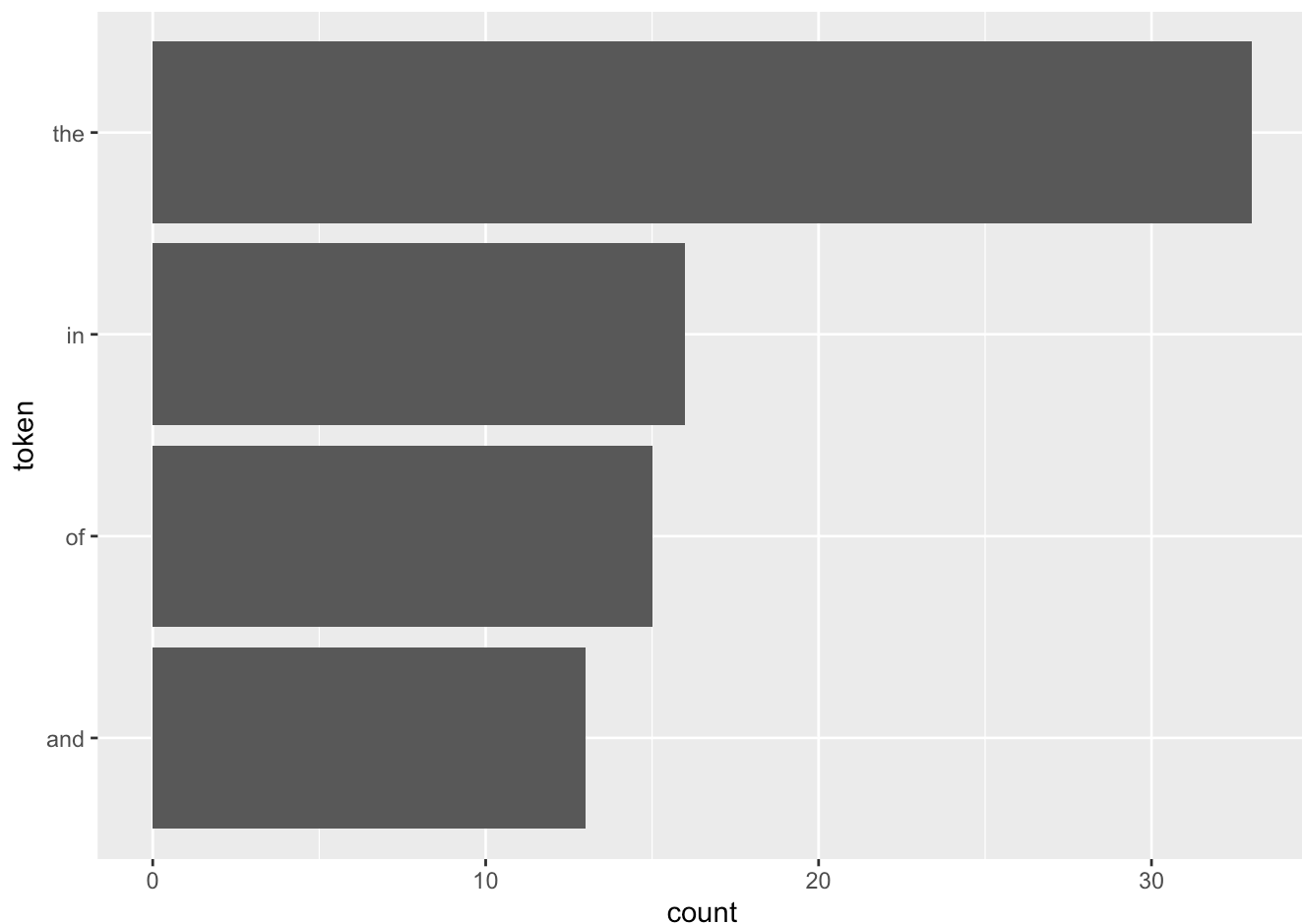
# Count tokens
tokens_mda %>% nrow()
```

```
## [1] 410
```

# Remove stop words

```
# Look at the most important frequent words
tokens_mda %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 10) %>%
  mutate(token = reorder(word, count)) %>%
  ggplot(aes(x=count, y=token)) +
  geom_col()
```





```
# Load custom stopwords
custom_stop_words <- read_csv("stop_words_list.csv", col_names = FALSE)
```

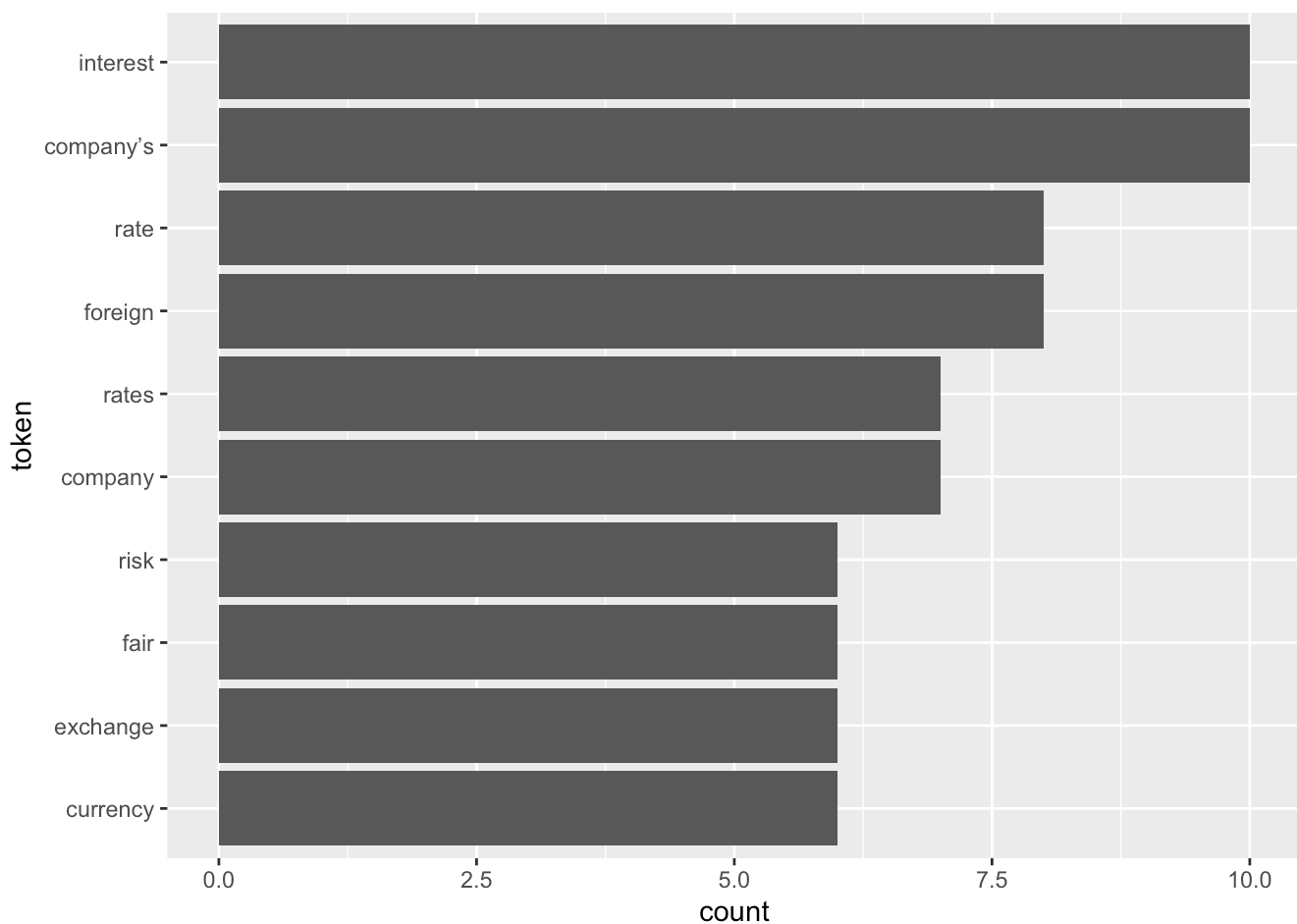
```
## Rows: 1087 Columns: 1
## — Column specification —————
## Delimiter: ","
## chr (1): X1
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Remove stop words
tokens_mda <- tokens_mda %>%
  anti_join(custom_stop_words, by = c('word'='X1'))

tokens_mda %>% nrow()
```

```
## [1] 250
```

```
tokens_mda %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 5) %>%
  mutate(token = reorder(word, count)) %>%
  ggplot(aes(x=count, y=token)) +
  geom_col()
```



## Stemming and Lemmatizing

```
# look at similar words
arrange(tokens_mda, word)[100:120, ]
```

word	order
<chr>	<int>
exposures	274
exposures	391
expressed	216
fair	34

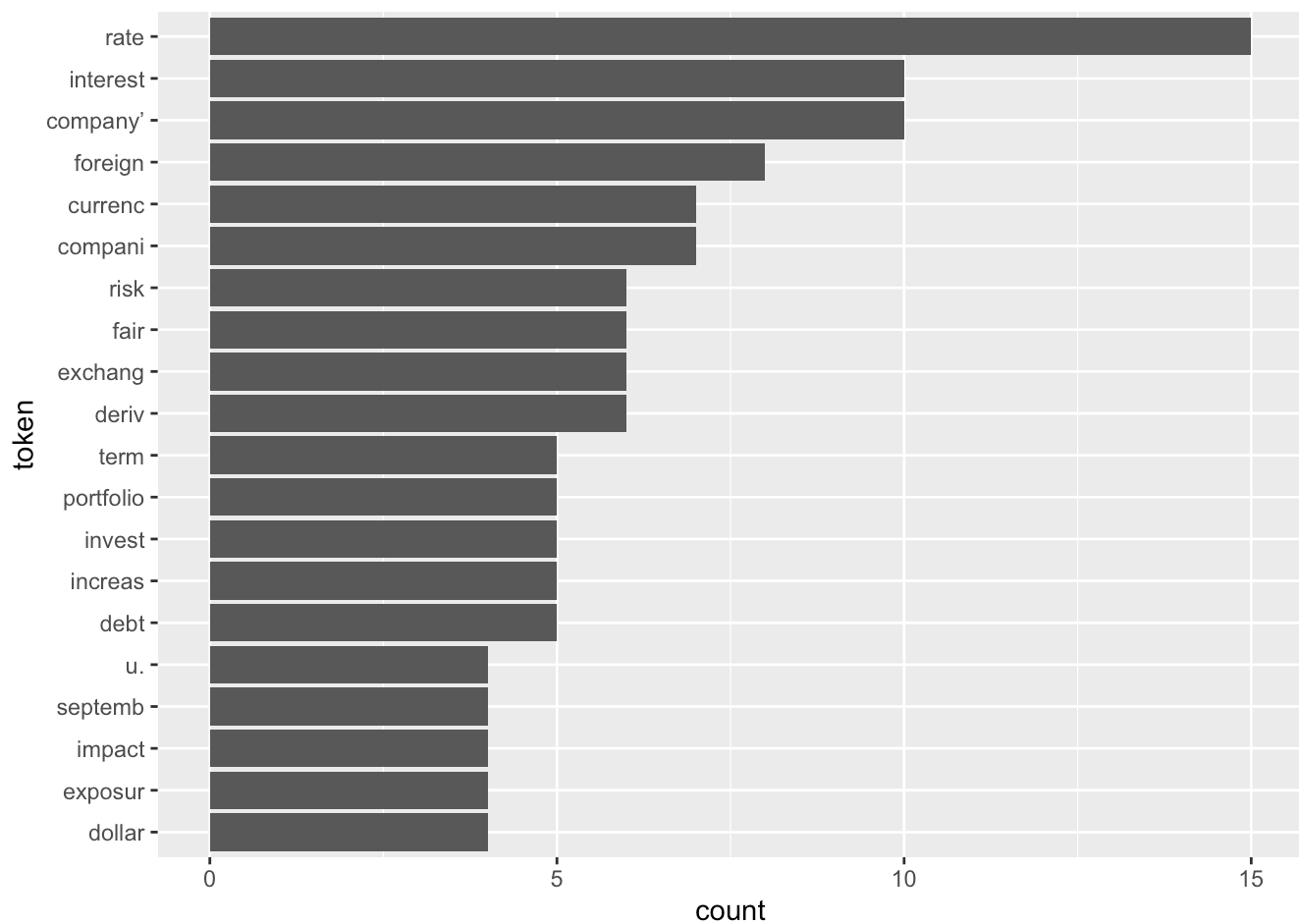
word <chr>	order <int>
fair	152
fair	228
fair	327
fair	366
fair	403
fluctuations	10
1-10 of 21 rows	Previous 1 2 3 Next

```
#Stem the tokens
stemmed_mda <- tokens_mda %>% mutate(stem = SnowballC::wordStem(word))

# look at similar words now
arrange(stemmed_mda, word)[100:120, ]
```

word <chr>	order <int>	stem <chr>
exposures	274	exposur
exposures	391	exposur
expressed	216	express
fair	34	fair
fair	152	fair
fair	228	fair
fair	327	fair
fair	366	fair
fair	403	fair
fluctuations	10	fluctuat
1-10 of 21 rows	Previous 1 2 3 Next	

```
stemmed_mda %>%
  group_by(stem) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 3) %>%
  mutate(token =reorder(stem, count)) %>%
  ggplot(aes(x=count, y=token)) +
  geom_col()
```



## Sentiment total

```
# load finance sentiment list and explore it
lm_dict <- tidytext::get_sentiments('loughran')

# view dictionary
lm_dict %>% group_by(sentiment) %>% summarize(count = n())
```

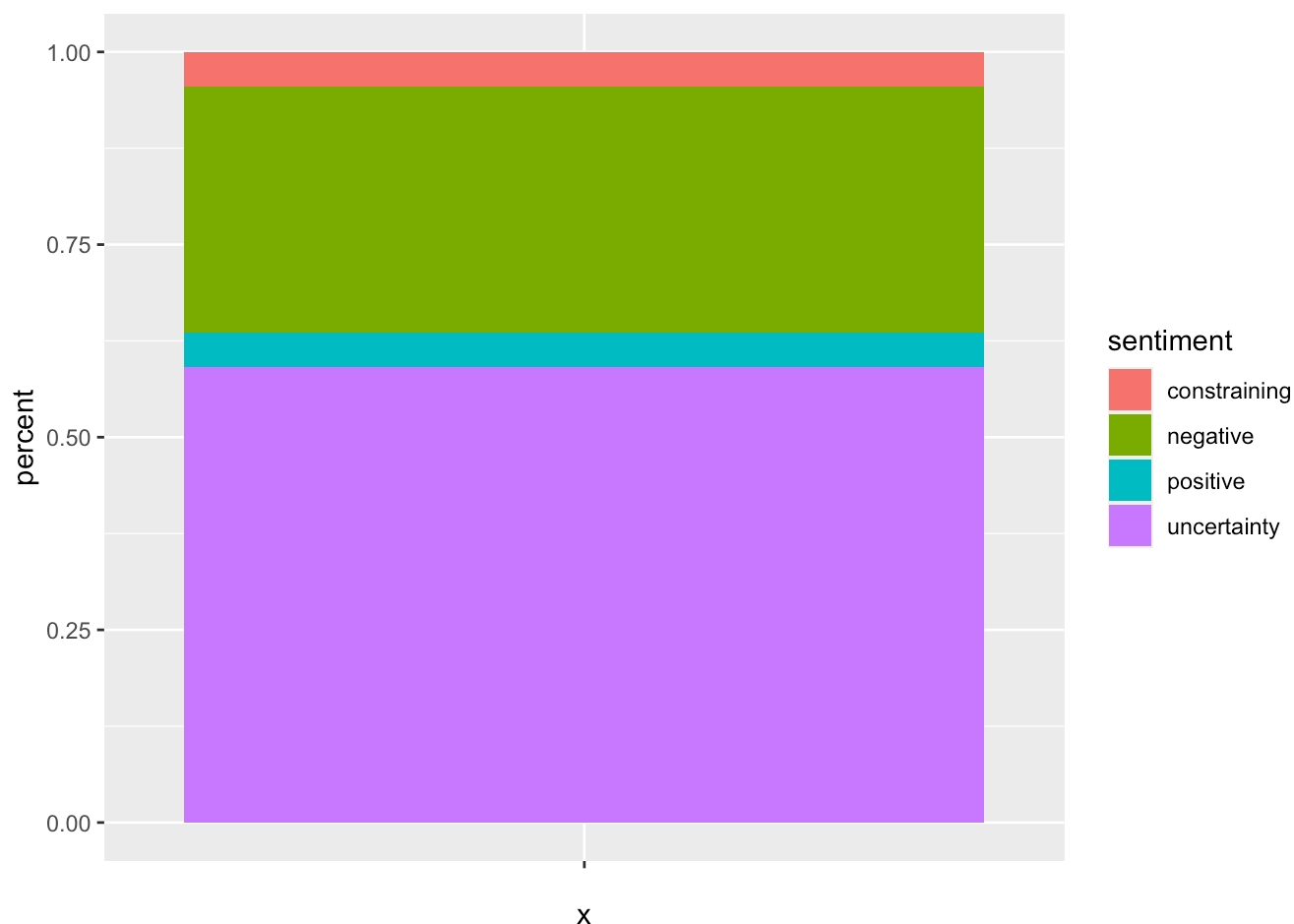
sentiment	count
<chr>	<int>
constraining	184
litigious	904
negative	2355
positive	354
superfluous	56
uncertainty	297
6 rows	

```
# Add sentiment
sentimented_mda <- stemmed_mda %>%
  inner_join(lm_dict, by = 'word')

# Explore totals
sentimented_mda %>%
  group_by(sentiment) %>%
  summarize(count = n(), percent = count/nrow(sentimented_mda))
```

<b>sentiment</b> <chr>	<b>count</b> <int>	<b>percent</b> <dbl>
constraining	1	0.04545455
negative	7	0.31818182
positive	1	0.04545455
uncertainty	13	0.59090909
4 rows		

```
sentimented_mda %>%
  group_by(sentiment) %>%
  summarize(count = n(), percent = count/nrow(sentimented_mda)) %>%
  ggplot(aes(x='', y=percent, fill=sentiment)) +
  geom_bar(width=1, stat='identity')
```



## 👉 2b: Press release sentiment

```
pr <- read_file("press_release.txt")
nchar(pr)
```

```
## [1] 2093
```

## Tokenize the text

```
# Change to a tibble (tidy dataframe)
tokens_pr <- tibble(pr)

# Tokenize
tokens_pr <- tokens_pr %>% unnest_tokens(output=word, input=pr, token='words', to_lower=TRUE)

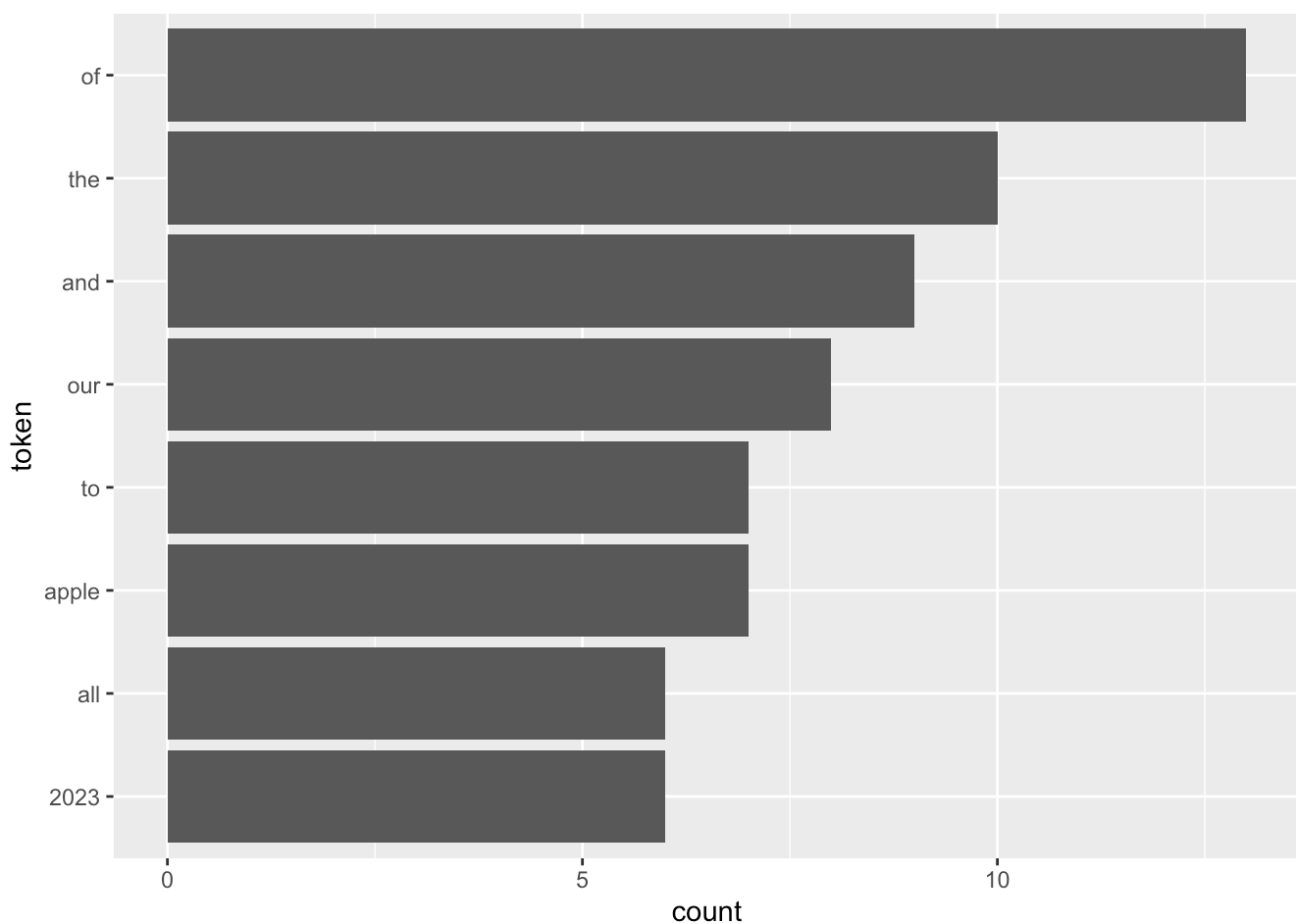
# add order of the words
tokens_pr <- tokens_pr %>% mutate(order = row_number())

# Count tokens
tokens_pr %>% nrow()
```

## [1] 326

# Remove stop words

```
# Look at the most important frequent words
tokens_pr %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 5) %>%
  mutate(token = reorder(word, count)) %>%
  ggplot(aes(x=count, y=token)) +
  geom_col()
```



```
# Load custom stopwords
custom_stop_words <- read_csv("stop_words_list.csv", col_names = FALSE)
```

```
## Rows: 1087 Columns: 1
## — Column specification —————
## Delimiter: ","
## chr (1): X1
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

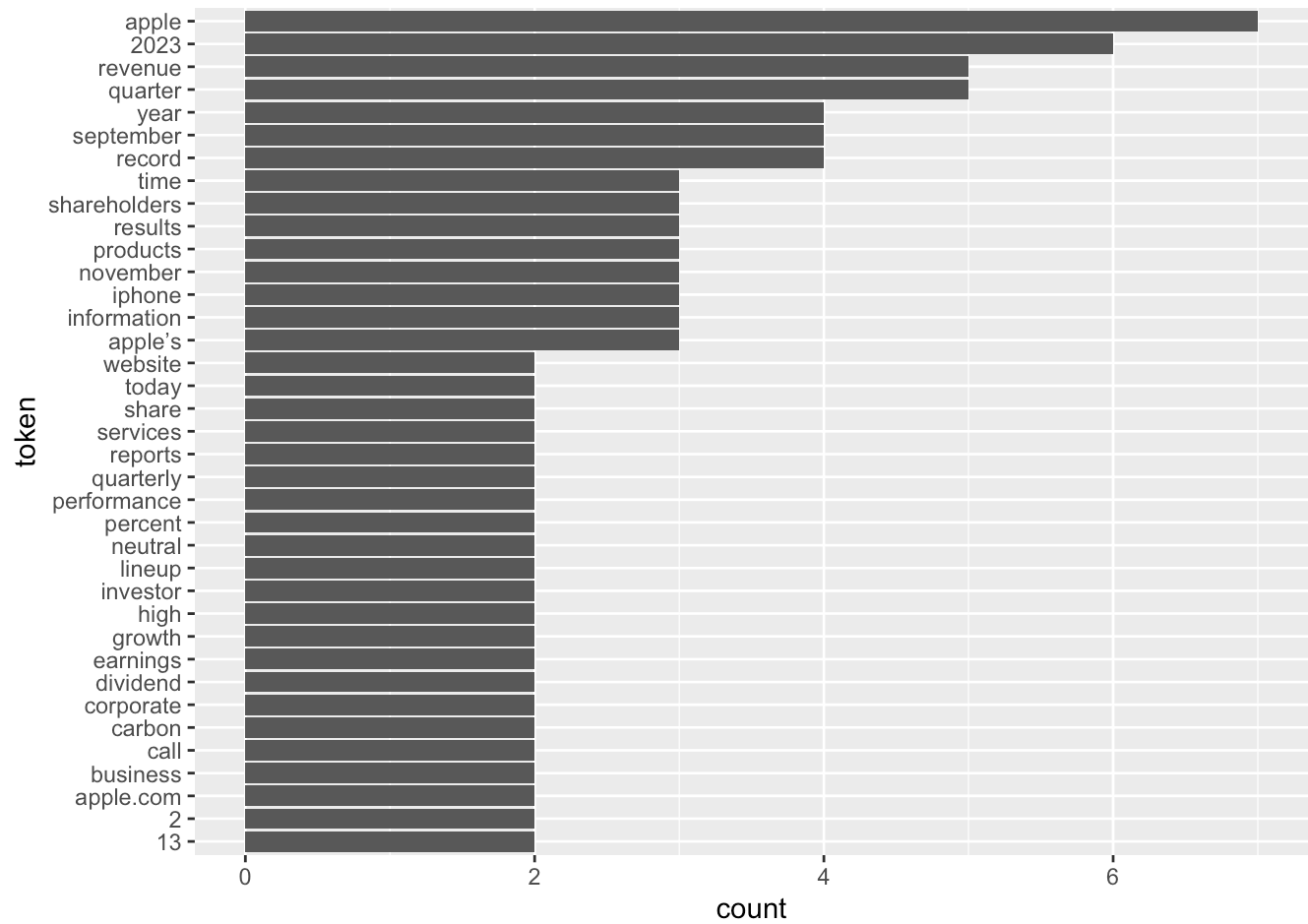
```
# Remove stop words
tokens_pr <- tokens_pr %>%
  anti_join(custom_stop_words, by = c('word'='X1'))

tokens_pr %>% nrow()
```

```
## [1] 196
```

```
tokens_pr %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 1) %>%
  mutate(token = reorder(word, count)) %>%
  ggplot(aes(x=count, y=token)) +
  geom_col()
```





# Stemming and Lemmatizing

```
# look at similar words
arrange(tokens_pr, word)[100:120, ]
```

word	order
<chr>	<int>
lineup	95
lineup	108
live	243
long	196
loyalty	163
maestri	166
major	118
make	124
meeting	324
milestone	119

1-10 of 21 rows

Previous 1 2 3 Next

```
#Stem the tokens
stemmed_pr <- tokens_pr %>% mutate(stem = SnowballC::wordStem(word))

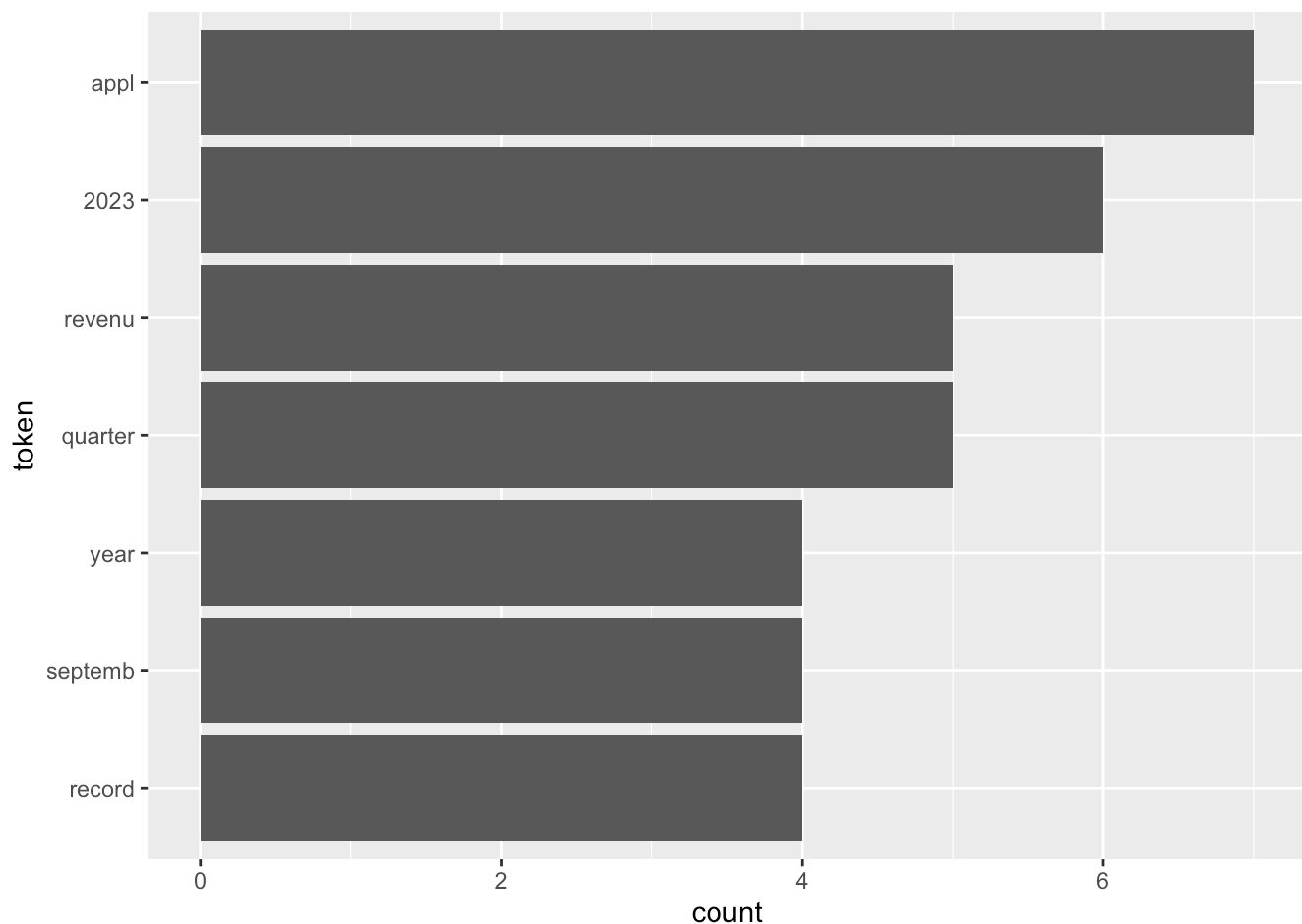
# look at similar words now
arrange(stemmed_pr, word)[100:120, ]
```

word <chr>	order <int>	stem <chr>
lineup	95	lineup
lineup	108	lineup
live	243	live
long	196	long
loyalty	163	loyalti
maestri	166	maestri
major	118	major
make	124	make
meeting	324	meet
milestone	119	mileston

1-10 of 21 rows

Previous 1 2 3 Next

```
stemmed_pr %>%
  group_by(stem) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 3) %>%
  mutate(token = reorder(stem, count)) %>%
  ggplot(aes(x=count, y=token)) +
  geom_col()
```



## Sentiment total

```
# load finance sentiment list and explore it
lm_dict <- tidytext::get_sentiments('loughran')

# view dictionary
lm_dict %>% group_by(sentiment) %>% summarize(count = n())
```

sentiment	count
<chr>	<int>
constraining	184
litigious	904
negative	2355
positive	354
superfluous	56
uncertainty	297
6 rows	

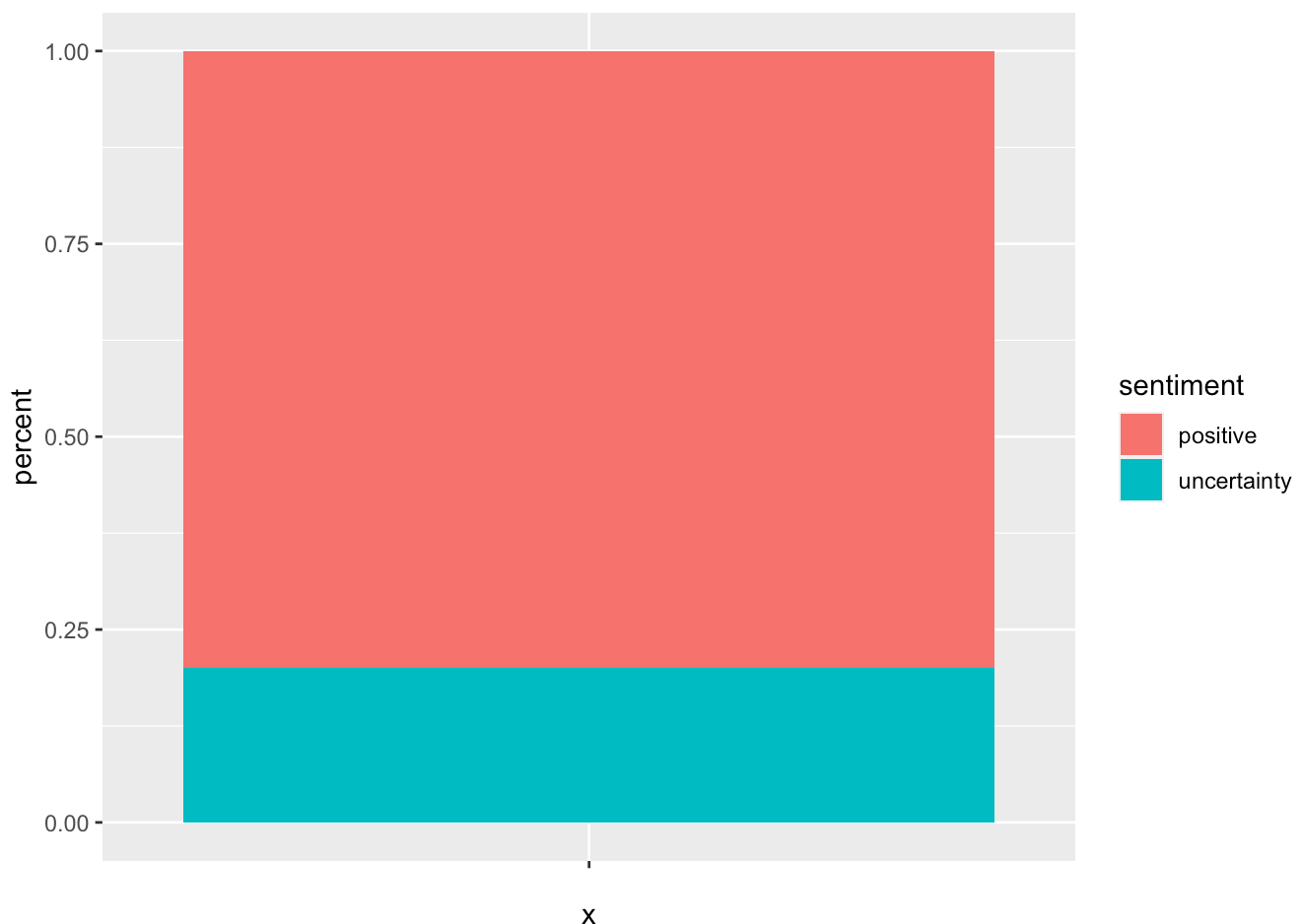
```
# Add sentiment
sentimented_pr <- stemmed_pr %>%
  inner_join(lm_dict, by = 'word')

# Explore totals
sentimented_pr %>%
  group_by(sentiment) %>%
  summarize(count = n(), percent = count/nrow(sentimented_pr))
```

<b>sentiment</b> <chr>	<b>count</b> <int>	<b>percent</b> <dbl>
positive	4	0.8
uncertainty	1	0.2

2 rows

```
sentimented_pr %>%
  group_by(sentiment) %>%
  summarize(count = n(), percent = count/nrow(sentimented_pr)) %>%
  ggplot(aes(x='', y=percent, fill=sentiment)) +
  geom_bar(width=1, stat='identity')
```



## 👉 2c: Explaining why the sentiments differ

The management, discussion and analysis (MD&A) section contains narratives on the analysis of a company's performance, which includes discussion of risks (including material weaknesses in the company's internal controls), compliance, and future plans (e.g., goals and new milestones). On the other hand, the press release reports earnings and contains forward looking statements regarding the company's operations and strategy.

There is more negative (~30%) and uncertainty (>50%) sentiment in the MD&A text transcripts compared to the press release which has mostly positive (~80%) sentiment. The difference is likely due to the fact the MD&A section includes additional disclosures of the company, which tackles risks and possible adverse effects of economic changes, which may result to a negative sentiment. In Apple's MD&A statement they discussed how they are exposed to fluctuations in the US interest rate and foreign exchange rate and how these can potentially negatively impact the company. They also discussed how strengthening of the US dollar will negatively affect net sales. Although they discuss the future plan to try to mitigate these risks, it is mostly going to have a negative sentiment. In contrast, the press release is optimistic and has an overall positive (~80%) and promotional sentiment. It highlights achievements like the iPhone revenue record and the all-time high in services revenue. The language used is optimistic and intended to instill confidence in the company's performance and future prospects. The text reads phrases such as "pleased to report", "strongest lineup", "unparalleled customer loyalty", and "long-term growth". This type of positive sentiment seems reasonable as with press releases, managers mainly use positive language. Additionally the tone is positive as the press release offers quotes from the CEO and CFO about increasing shareholder return. When there is more positive sentiment in a press release, research has shown that it reflects positively on the stock price for the company.