Final submission homework 6 of Yue Ma

Yue Ma

10/25/2018

1. Words in Ulysses

a) Words with z

```
##   [1] "razor"                  "razorblade"
##   [3] "gazing"                 "gazed"
##   [5] "gaze"                   "seized"
##   [7] "zulus"                  "teazle"
##   [9] "brazen"                 "glazing"
##  [11] "puzzled"                "breezy"
##  [13] "crazy"                  "zut"
##  [15] "zarathustra"            "zeal"
##  [17] "nuzzling"               "frozen"
##  [19] "laissez"                "mürzsteg"
##  [21] "frauenzimmer"           "crozier"
##  [23] "wheezy"                 "razorshells"
##  [25] "maze"                   "savez"
##  [27] "zebra"                  "zmellz"
##  [29] "bloodz"                 "odz"
##  [31] "iridzman"               "fitzgerald"
##  [33] "blazing"                "milkoozing"
##  [35] "gizzards"               "dlugacz's"
##  [37] "puzzle"                 "oozed"
##  [39] "lazily"                 "sizeable"
##  [41] "sizzling"               "blazes"
##  [43] "trapeze"                "biz"
##  [45] "denzille"               "dlugacz"
##  [47] "prize"                  "bazaar"
##  [49] "gauze"                  "eczema"
##  [51] "lazy"                   "azotes"
##  [53] "wheeze"                 "hazard"
##  [55] "mazzoth"                "mozart's"
##  [57] "buzz"                   "benzoin"
##  [59] "embezzling"             "waltzing"
##  [61] "influenza"              "capsized"
##  [63] "ooze"                   "muzzle"
##  [65] "lazarus"                "puzzling"
##  [67] "oozing"                 "bronzefoil"
##  [69] "gazette"                "phiz"
##  [71] "quizzing"               "zephyrs"
##  [73] "zigzagging"             "breeze"
##  [75] "mazurka"                "entrez"
##  [77] "fitzharris"             "chapelizod"
##  [79] "fitzgibbon"             "unglazed"
##  [81] "fitzgibbon's"           "lozenge"
##  [83] "zion"                   "glazed"
```

```
##  [85] "dozen"                       "blizzard"
##  [87] "fitzmaurice"                 "lizzie"
##  [89] "graze"                       "freeze"
##  [91] "squeeze"                     "goerz"
##  [93] "cherchez"                    "gorgonzola"
##  [95] "zinfandel's"                 "fizz"
##  [97] "buzzed"                      "zinfandel"
##  [99] "zest"                        "sizing"
## [101] "doze"                        "size"
## [103] "assizes"                     "zealous"
## [105] "sizar's"                     "elizabethan"
## [107] "bronzelidded"                "aztec"
## [109] "panza"                       "eliza"
## [111] "azured"                      "lizards"
## [113] "elizabeth"                   "lopez"
## [115] "sneeze"                      "mezzo"
## [117] "horizon"                     "voulez"
## [119] "quizzer"                     "sacrifizio"
## [121] "grazie"                      "fitzgeralds"
## [123] "sneezed"                     "grizzled"
## [125] "citizens"                    "bronze"
## [127] "lozenges"                    "prizering"
## [129] "schoolprizes"                "lowsized"
## [131] "fitzsimons"                  "sonnez"
## [133] "liszt's"                     "bronzelydia"
## [135] "goldbronze"                  "bronzegold"
## [137] "bronze's"                    "buzzing"
## [139] "shebronze"                   "azure"
## [141] "blazure's"                   "bronzedouce"
## [143] "whizzed"                     "intermezzo"
## [145] "razzle"                      "spinoza"
## [147] "bronzelid"                   "glazily"
## [149] "herzog"                      "citizen"
## [151] "denizens"                    "breezes"
## [153] "fitzsimon"                   "goodsized"
## [155] "velasquez"                   "squeezed"
## [157] "skeezing"                    "sneezing"
## [159] "bézique"                     "dazzling"
## [161] "speranza's"                  "schwanzenbad"
## [163] "kisászony"                   "banzai"
## [165] "zivio"                       "rienzi"
## [167] "muzzling"                    "grazier"
## [169] "liz"                         "frenzied"
## [171] "zouave"                      "zaretsky"
## [173] "conspuez"                    "hazeleyes"
## [175] "hazelnuts"                   "zulu"
## [177] "gadzooks"                    "gonzaga"
## [179] "százharminczbrojúgulyás"     "rakóczsy's"
## [181] "visszontlátásra"             "chintz"
## [183] "lazenby's"                   "blaze"
## [185] "zrads"                       "furze"
## [187] "zoo"                         "snooze"
## [189] "mazer"                       "seize"
## [191] "coz"                         "cozening"
```

```
## [193] "pizzle"                  "bonzes"
## [195] "poyntz"                  "rendezvoused"
## [197] "grazing"                 "sturzgeburt"
## [199] "drizzling"               "zodiacal"
## [201] "huzzah"                  "mediumsized"
## [203] "spermatozoa"             "spallanzani"
## [205] "piazzetta"               "zermatt"
## [207] "kreutzer"                "guzzling"
## [209] "uz"                      "amazingly"
## [211] "rheumatiz"               "nachez"
## [213] "largesize"               "jenatzy"
## [215] "woozy"                   "winefizzling"
## [217] "ginsizzling"             "booseguzzling"
## [219] "sneezes"                 "seizes"
## [221] "frieze"                  "gazelle"
## [223] "squeezes"                "gazes"
## [225] "fritz"                   "fez"
## [227] "nozzle"                  "dazedly"
## [229] "haze"                    "amazon"
## [231] "chimpanzee"              "zoe"
## [233] "gazelles"                "vizier"
## [235] "howitzers"               "razed"
## [237] "graziers"                "mitzvah"
## [239] "askenazim"               "bronzed"
## [241] "zouave's"                "smerdoz"
## [243] "schwarz"                 "aranjuez"
## [245] "szombathely"             "azazel"
## [247] "mizraim"                 "gizzard"
## [249] "jetez"                   "zenith"
## [251] "lizardlettered"          "zodiac"
## [253] "dazed"                   "zoe's"
## [255] "graize"                  "zigzag"
## [257] "quinze"                  "furzebush"
## [259] "drizzle"                 "horizontal"
## [261] "shulomowitz"             "abramovitz"
## [263] "chazen"                  "godblazeqrukbrukarchkrasht"
## [265] "fizzing"                 "lifesize"
## [267] "trapezes"                "beelzebub"
## [269] "buzzard"                 "wizard"
## [271] "waltz"                   "dansez"
## [273] "changez"                 "donnez"
## [275] "remerciez"               "waltzes"
## [277] "florryzoe"               "zigzags"
## [279] "tweezers"                "blazer"
## [281] "patrizio"                "franz"
## [283] "razors"                  "hashbaz"
## [285] "fitzedward"              "rendezvous"
## [287] "brazier"                 "dizzily"
## [289] "bamboozled"              "choza"
## [291] "lazarillo"               "bazan"
## [293] "ozone"                   "capsize"
## [295] "squeezing"               "gazers"
## [297] "aztecs"                  "amazing"
## [299] "fitz"                    "tidysized"
```

```
## [301] "halfcrazy"                 "viz"
## [303] "sized"                     "bandez"
## [305] "largesized"                "conversaziones"
## [307] "smallsized"                "zones"
## [309] "ahorizontal"               "sizes"
## [311] "prizes"                    "bezique"
## [313] "colza"                     "fitzpatrick"
## [315] "zoological"                "horizontally"
## [317] "oz"                        "mendoza"
## [319] "l'zamatejch"               "piazzi"
## [321] "frenzy"                    "freezing"
## [323] "zero"                      "mizrach"
## [325] "hozier's"                  "mezzanine"
## [327] "mizpah"                    "szesfehervar"
## [329] "herz"                      "plaza"
## [331] "outsize"                   "embezzlement"
## [333] "doz"                       "brazenfaced"
## [335] "zingari"                   "gomez"
## [337] "embarazada"                "switzers"
## [339] "delapaz"                   "zurich"
```

```
## number_as
##   1   2   3
## 293  46   1
```

The results above show the all unique words that contain at least one z. The table shows how many z's the words contain: 293 words contain one z, 46 words contain two z's, 1 word contain 3 z's.

```
## [1] "zmellz"                  "zigzagging"
## [3] "százharminczbrojúgulyás" "azazel"
## [5] "zigzag"                  "zigzags"
```

From the results above, we can see that word - "százharminczbrojúgulyás" has the farest distance between consecutive z's. There are 7 characters bewteen consectutive z's.

b) Vowels

```
## v1
##     0     1
## 29094   876
```

From the results, we can see that there 876 unique words start and end with a vowel.

```
## [1] "air"  "out"  "oak"  "ouns" "aunt" "our"
```

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

There are 341 unique words that start with two or more vowels. For instance, "air" "out" "oak" "ouns" "aunt" "our", etc.

```
##
##   3   4   5   6   7   8   9  11  13  16  21
##  53 118 119  47  12   5   3   3   3   1   1
```

```
## [1] "frseeeeeeeeeeeeeeeeeeeefrong"
```

From the results, we can see that word - "frseeeeeeeeeeeeeeeeeeefrong" has the most consecutive vowels with 20 consecutive vowels.

c) English spelling

```
## [1] "Number of words without rules: 201"
```

```
## [1] "Number of words with rules: 799"
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
##                                       content proportion
## 1    Number of words with rules(total words):      2.67%
## 2 Number of words without rules(total words):     0.671%
```

```
##                           content proportion
## 1    Number of words with rules:      79.9%
## 2 Number of words without rules:      20.1%
```

From the table, we can see that there are 2.67% of the words when the rule holds and 0.671% of the words when the rule does not hold. If we just look at the words with ei or ie, there are 79.9% of the words when the rule holds and 20.1% of the words when the rule does not hold.

2. MTA Delays

a) Clean the dataset

```
## # A tibble: 6 x 2
##   text                                                  created_at
##   <chr>                                                 <dttm>
## 1 Service Update: Northbound C trains continue to run~ 2018-04-12 01:09:29
## 2 "Service update: Some northbound E trains are runni~ 2018-02-13 18:01:50
## 3 "4 and 6 train service has resumed. \nhttps://t.co/~ 2018-07-09 06:12:11
## 4 7 and 7 express trains are running with delays in b~ 2018-01-11 12:52:10
## 5 "F train service has resumed. \nhttps://t.co/i80bMp~ 2018-08-22 23:03:09
## 6 Southbound 3 trains are running with delays because~ 2018-04-11 14:19:49
```

The table above is the head of the clean table of MTA twitter concent.

b) Time of Delay

```
##
## Attaching package: 'lubridate'
```

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

```
## ------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:lubridate':
##
##     here
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following object is masked from 'package:purrr':
##
##     compact
```

```
##
##             night mornings mid-day afternoon evening
##   Sunday      137       65     128        44      90
##   Monday      134       54     207        94     147
##   Tuesday     146       37     166       107     152
##   Wednesday   135       23     161        83     167
##   Thursday    139       32     200        70     128
##   Friday      143       39     176        83     139
##   Saturday    113       55     122        62      80
```

From the table above, we can see that there are most delays during mid-day on Monday. On a day, there are most delays during mid-days and least delays on morning on average. Meanwhile, there are more delays on weekdays than weekends on average.

c) Type of Delay

```
## delay_reason
##                   because of signal problems at
##                                             981
## because of a train with mechanical problems at
##                                             491
##                   because of switch problems at
##                                             194
##                 because of a sick passenger at
##                                             178
##                    because of NYPD activity at
##                                             176
```

From the table above, we can see that the top 5 reasons of delays are: 1. signal problems; 2. mechanical problems; 3. switch problems; 4. sick passengers; 5. NYPD activities.

d) Provide a summary of which train lines are affected by weekday vs. weekend.

```
##     weekday weekend
```

```
## 1       195      69
## 2       492     113
## 3       322      70
## 4       376     103
## 5       379      89
## 6       200      54
## 7       272      76
## A       320     126
## B       207       4
## C       228      53
## D       278     100
## E       325      86
## F       400     107
## G        75      22
## J       136      42
## L       148      62
## M       248      19
## N       300      89
## Q       223      68
## R       285      80
## S        10       9
## SIR      23       5
## W       188      13
## Z         7       0
```

From the table above, we can see that line 2 is affected most on weekdays and line A is affected most at weekends. Line Z is affected least both on weekdays and weekends.