QBS181 HW2

Mark Taylor

Setup

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lubridate)
```

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

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(odbc)
library(sqldf)
```

```
## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite
```

```r
library(stringr)

# Connect to database
con <- DBI::dbConnect(odbc::odbc(),
                      Driver   = "/usr/local/lib/libmsodbcsql.17.dylib",
                      Server   = "qbs181-db.dartmouth.edu",
                      Database = "qbs181",
                      UID      = "mtaylor",
                      PWD      = "mtaylor@qbs181",
                      Port     = 40062)
```

1

```
# Left join Phonecall with PhoneCall_Encounter to get Encountercodes
# (aliases as EnrollmentGroup so we need not rename)
Phonecall_Encounter <- dbGetQuery(con, "SELECT CustomerId,
                                          EncounterCode as EnrollmentGroup
                                       FROM Phonecall_Encounter")

# Assign EnrollmentGroup based on the current EnrollmentGroup value
# (corresponding to the code)
Phonecall_Encounter$EnrollmentGroup <-
  ifelse(Phonecall_Encounter$EnrollmentGroup == 125060000, "Clinical Alert",
  ifelse(Phonecall_Encounter$EnrollmentGroup == 125060001, "Health Coaching",
  ifelse(Phonecall_Encounter$EnrollmentGroup == 125060002, "Technical Question",
  ifelse(Phonecall_Encounter$EnrollmentGroup == 125060003, "Administrative",
  ifelse(Phonecall_Encounter$EnrollmentGroup == 125060004, "Other",
  ifelse(Phonecall_Encounter$EnrollmentGroup == 125060004, "Lack of engagement",
                                                          "NULL")
  )))))

sample_n(Phonecall_Encounter, 10)
```

```
##                              CustomerId    EnrollmentGroup
## 1  6DF9BF37-E61D-E611-8128-C4346BB59854     Administrative
## 2  1FFD86B2-2D0B-E611-8120-C4346BAD2660     Administrative
## 3  87FB86B2-2D0B-E611-8120-C4346BAD2660     Administrative
## 4  968AE3CB-5111-E611-811B-C4346BAC02E8     Administrative
## 5  53EA1762-05E0-E511-8122-C4346BB59854     Clinical Alert
## 6  E540A719-F3C6-E611-80F5-5065F38A4B01     Administrative
## 7  99774C35-2945-E611-80E6-C4346BDC9111 Technical Question
## 8  C022FCAF-2616-E611-8128-C4346BB59854     Administrative
## 9  CC97A7B0-7148-E611-80E7-5065F38B3241               NULL
## 10 CA8EA7B0-7148-E611-80E7-5065F38B3241               NULL
```

2

```
# Return # of records for each EnrollmentGroup
sqldf("SELECT EnrollmentGroup, COUNT(*) recordNumber
FROM Phonecall_Encounter
GROUP BY EnrollmentGroup")
```

```
##      EnrollmentGroup recordNumber
## 1     Administrative         4480
## 2     Clinical Alert          453
## 3    Health Coaching          409
## 4               NULL         1824
## 5              Other          189
## 6 Technical Question         1059
```

3

```r
# import CallDuration
CallDuration <- dbGetQuery(con, "SELECT * FROM CallDuration")

# Merge with PhoneCall_Encounter
EncounterXDuration <- sqldf("SELECT PCE.*, CD.CallType, CD.CallDuration, CD.CallOutcome
          FROM CallDuration as CD
          LEFT JOIN Phonecall_Encounter as PCE
          ON CD.tri_CustomerIDEntityReference = PCE.CustomerId")

sample_n(EncounterXDuration, 10)
```

```
##                              CustomerId    EnrollmentGroup CallType
## 1  668EA7B0-7148-E611-80E7-5065F38B3241     Administrative        1
## 2  5BFA86B2-2D0B-E611-8120-C4346BAD2660     Administrative        1
## 3  78B7F1E3-BE43-E611-80E6-5065F38BA151               NULL        1
## 4  5693A7B0-7148-E611-80E7-5065F38B3241    Health Coaching        1
## 5  609CA7B0-7148-E611-80E7-5065F38B3241     Administrative        1
## 6  9DF9BF37-E61D-E611-8128-C4346BB59854     Administrative        1
## 7  5C5DD306-D547-E611-80E6-5065F38BA151     Administrative        1
## 8  6896A7B0-7148-E611-80E7-5065F38B3241 Technical Question        1
## 9  C4FDFEE3-D4E4-E511-8123-C4346BB59854     Administrative        1
## 10 B3FB86B2-2D0B-E611-8120-C4346BAD2660 Technical Question        1
##    CallDuration CallOutcome
## 1           102           2
## 2            93           2
## 3           107           2
## 4           869           1
## 5            91           2
## 6            38           3
## 7            76           2
## 8            53           3
## 9            93           2
## 10           50           1
```

4

```r
# Replace numbers with call types
EncounterXDuration$CallType <-
  ifelse(EncounterXDuration$CallType == 1, "Inbound", "Outbound")

# Report count by call type
sqldf("SELECT CallType, COUNT(*) recordNumber
FROM EncounterXDuration
GROUP BY CallType")
```

```
##   CallType recordNumber
## 1  Inbound         9875
## 2 Outbound          833
```

```r
# Replace numbers with call outcomes
EncounterXDuration$CallOutcome <-
  ifelse(EncounterXDuration$CallOutcome == 1, "No response",
  ifelse(EncounterXDuration$CallOutcome == 2, "Left voice mail", "successful"))

# Report count by call outcome
sqldf("SELECT CallOutcome, COUNT(*) recordNumber
FROM EncounterXDuration
GROUP BY CallOutcome")
```

```
##       CallOutcome recordNumber
## 1 Left voice mail         4741
## 2     No response         5203
## 3      successful          764
```

```r
# Report sum of call duration by EnrollmentGroup
sqldf("SELECT EnrollmentGroup, SUM(CallDuration)
FROM EncounterXDuration
GROUP BY EnrollmentGroup")
```

```
##      EnrollmentGroup SUM(CallDuration)
## 1               <NA>               549
## 2     Administrative            708638
## 3     Clinical Alert            268958
## 4    Health Coaching            267033
## 5               NULL            380474
## 6              Other            200487
## 7 Technical Question            442126
```

5

```r
# merge Conditions and Demographics as we import
DemoXCond <- dbGetQuery(con, "SELECT D.*, C.tri_name
FROM Conditions as C
LEFT JOIN Demographics as D
ON C.tri_patientid = D.contactid")

# import Text
Text <- dbGetQuery(con, "SELECT * FROM Text")

# merge Conditions and Demographics (already merged) with Text
DemoXCondXText <- sqldf("SELECT DC.*, T.SenderName, T.TextSentDate
      FROM Text as T
      LEFT JOIN DemoXCond as DC
      ON T.tri_contactId = DC.contactid")

# add aggregated Week column
DemoXCondXText <- DemoXCondXText %>% group_by(Week=floor_date(TextSentDate, "7 days"))

# count of texts by sender type for each week
```

```
WeeklyTextsBySender <- DemoXCondXText %>% count(Week, SenderName) %>% ungroup()

sample_n(WeeklyTextsBySender, 10)
```

```
## # A tibble: 10 x 3
##    Week                SenderName     n
##    <dttm>              <chr>      <int>
##  1 2016-09-22 00:00:00 System       739
##  2 2016-09-08 00:00:00 Clinician    507
##  3 2016-12-29 00:00:00 Customer      55
##  4 2016-03-15 00:00:00 System        78
##  5 2016-07-15 00:00:00 Customer     486
##  6 2016-07-08 00:00:00 Clinician    311
##  7 2016-08-01 00:00:00 System      1081
##  8 2016-09-29 00:00:00 Clinician     67
##  9 2016-10-15 00:00:00 Customer     165
## 10 2016-05-22 00:00:00 System      1309
```

6

```
# count of texts by condition for each week
WeeklyTextsByCondition <- DemoXCondXText %>% count(Week, tri_name) %>% ungroup()

sample_n(WeeklyTextsByCondition, 10)
```

```
## # A tibble: 10 x 3
##    Week                tri_name                     n
##    <dttm>              <chr>                    <int>
##  1 2016-04-15 00:00:00 Activity Monitoring        633
##  2 2016-03-29 00:00:00 Activity Monitoring        346
##  3 2017-01-01 00:00:00 Diabetes                    56
##  4 2016-05-15 00:00:00 Congestive Heart Failure    34
##  5 2016-10-29 00:00:00 Diabetes                     6
##  6 2016-04-08 00:00:00 COPD                        16
##  7 2016-09-01 00:00:00 COPD                       102
##  8 2017-01-22 00:00:00 Diabetes                   173
##  9 2016-06-01 00:00:00 Diabetes                    89
## 10 2016-07-01 00:00:00 Hypertension               140
```