R Notebook

Load required libraries

```
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
```

```
library(readxl)
library(stringr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v readr   1.4.0
## v tibble  3.0.4     v purrr   0.3.4
## v tidyr   1.1.2     v forcats 0.5.0
```

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

setup list of purchase card files

```
cc_folder = "../../../data/data.birmingham.gov.uk"
cc_files = c(
  "2019-10:purchase-card-publish-spend-october-2019.xls",
  "2019-11:publish-spend-november-2019-mod.xls",
  "2019-12:publish-spend-december-2019.xls",
  "2020-01:publish-spend-january-2020.xls",
  "2020-02:publish-spend-february-2020.xls",
  "2020-03:publish-spend-march-2020.xls",
  "2020-04:publish-spend-april-2020.xls",
  "2020-05:publish-spend-may-2020.xls",
  "2020-06:publish-spend-june-2020.xls",
  "2020-07:publish-spend-july-2020.xls",
  "2020-08:publish-spend-august-2020.xls",
  "2020-09:publish-spend-purchase-card-itemised-transactions-september-20-all-directorates.xls"
  )
```

read purchase card files into a single dataset

```
## function to make sure a column exists (adds it if it does exist)
fncols <- function(data, cname) {
  add <-cname[!cname%in%names(data)]
```

```
  if(length(add)!=0) data[add] <- NA
  data
}

# read a purchase card file, normalize columns (as different month files vary)
read_cc_file <- function(folder, file) {
  path <- paste(folder, word(file, 2, sep=":"), sep="/")
  print(path)
  card <- read_excel(path)
  card <- fncols(card, "MERCHANT NAME")
  card.norm <- card %>% select(`TRANS DATE`, `BILLING GROSS AMT`, `MERCHANT NAME`, `CARD NUMBER`, `TRANS
  card.norm$Directorate = toupper(card.norm$Directorate)
  card.norm$month = word(file, 1, sep=":")
  print(word(file, 1, sep=":"))

  card.norm
}

# loop on all files and aggregate into a single dataset called cc_all
all <- read_cc_file(cc_folder, cc_files[[1]])
```

```
## [1] "../../../data/data.birmingham.gov.uk/purchase-card-publish-spend-october-2019.xls"
## [1] "2019-10"
```

```
for ( file in cc_files[-1] ) {
  more <- read_cc_file(cc_folder, file)
  all <- rbind(all, more)
}
```

```
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-november-2019-mod.xls"
## [1] "2019-11"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-december-2019.xls"
## [1] "2019-12"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-january-2020.xls"
## [1] "2020-01"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-february-2020.xls"
## [1] "2020-02"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-march-2020.xls"
## [1] "2020-03"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-april-2020.xls"
## [1] "2020-04"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-may-2020.xls"
## [1] "2020-05"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-june-2020.xls"
## [1] "2020-06"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-july-2020.xls"
## [1] "2020-07"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-august-2020.xls"
## [1] "2020-08"
## [1] "../../../data/data.birmingham.gov.uk/publish-spend-purchase-card-itemised-transactions-septembe
## [1] "2020-09"
```

```
rm(more)
```

build aggregation subsets

```r
# all expenses with amount 3000
all3000 <- all %>% filter(`BILLING GROSS AMT`==3000)

# expenses by directorate
all_by_dir <- all %>%
                count(month, Directorate, wt=`BILLING GROSS AMT`) %>%
                rename(total=n)
all_by_dir <- all_by_dir[order(-all_by_dir$total),]

# NEIGHBOURHOODS directorate by month
all.n = all %>% filter(Directorate=="NEIGHBOURHOODS")
all.n.c <- all.n %>%
            count(month, Directorate, wt=`BILLING GROSS AMT`) %>%
            rename(total=n)

# expenses with amount 3000 by month
all.3000.c = all3000 %>%
              count(month, Directorate, wt=`BILLING GROSS AMT`) %>%
              rename(total=n)

# summary dataset build: all expenses (counts) by month
all.c <- all %>%
              count(month, wt=`BILLING GROSS AMT`) %>%
              rename(total=n)


# add neighboorhoods and c3000 columns to summary dataset
all.c$neighboorhoods = all.n.c$total
all.c$c3000 = all.3000.c$total

# calc homeless expenses and add as column to summary
all.n.hless = all.n %>% filter(str_detect(`TRANS CAC DESC 2`, "less|Less|LESS"))
all.n.hless.c = all.n.hless %>%
                count(month, wt=`BILLING GROSS AMT`) %>%
                rename(total=n)
all.c$hless <- all.n.hless.c$total

all.c
```

```
## # A tibble: 12 x 5
##    month      total neighboorhoods  c3000    hless
##    <chr>      <dbl>          <dbl>  <dbl>    <dbl>
##  1 2019-10 1565013        1160354. 228000 1089820.
##  2 2019-11 1477212.       1100618. 231000 1019944.
##  3 2019-12  868353.         527778.  75000  466781.
##  4 2020-01 1134460.         753754. 234000  678942.
##  5 2020-02  930214.         567480. 138000  460275.
##  6 2020-03  604724.         304718. 147000  246495.
##  7 2020-04  319038.          80624. 114000   42195.
##  8 2020-05  519127.         222860. 144000  162882.
##  9 2020-06  540431.         153568.  72000  113468.
## 10 2020-07  601437.         250375. 162000  186046.
## 11 2020-08  383225.         171272.  75000  101536.
## 12 2020-09  612692.         234208. 129000  159134.
```

sumamrise

```
all.c %>% summarize_if(is.numeric, sum, na.rm=TRUE)
```

```
## # A tibble: 1 x 4
##      total neighboorhoods   c3000    hless
##      <dbl>          <dbl>   <dbl>    <dbl>
## 1 9555925.       5527610. 1749000 4727517.
```

Write files for outside analysis

```
write.csv(all, "output/bcc-cc-all.csv")
write.csv(all.n, "output/bcc-cc-all-neighboorhoods.csv")
```