Midterm

Cody Evans

Importing data

```
options(tinytex.verbose = TRUE)

suppressMessages(library("tidyverse"))
suppressMessages(register <- read_csv("register.csv"))
suppressMessages(sales <- read_csv("sales.csv"))
#N <- nrow(register)
```

Original column names for the two data sets Register and Sales

Register

```
colnames(register)
```

```
##  [1] "purchase"
##  [2] "item"
##  [3] "charge"
##  [4] "price"
##  [5] "is the customer a student/faculty/staff (0) or unaffiliated (1)"
##  [6] "customer id"
##  [7] "receipt"
##  [8] "contact preference"
##  [9] "newsletter"
## [10] "sales"
## [11] "preferred customer discount"
```

Sales

```
colnames(sales)
```

```
##  [1] "category of inventory goods" "1-2018"
##  [3] "2-2018"                      "3-2018"
##  [5] "4-2018"                      "5-2018"
##  [7] "6-2018"                      "7-2018"
##  [9] "8-2018"                      "9-2018"
## [11] "10-2018"                     "11-2018"
## [13] "12-2018"                     "1-2019"
## [15] "2-2019"                      "3-2019"
## [17] "4-2019"                      "5-2019"
## [19] "6-2019"                      "7-2019"
## [21] "8-2019"                      "9-2019"
## [23] "10-2019"
```

Changing column names for better understanding

```
colnames(register)[colnames(register) == "is the customer a student/faculty/staff (0) or unaffiliated (
colnames(register)[colnames(register) == "customer id"] <- "cid"
colnames(register)[colnames(register) == "preferred customer discount"] <- "discount"
colnames(register)[colnames(register) == "contact preference"] <- "contact"

cat("Improved column names for the register data\n")
```

```
## Improved column names for the register data
```

```
colnames(register)
```

```
##  [1] "purchase"     "item"         "charge"       "price"
##  [5] "unaffiliated" "cid"          "receipt"      "contact"
##  [9] "newsletter"   "sales"        "discount"
```

```
colnames(sales)[colnames(sales) == "category of inventory goods"] <- "goods"

cat("Improved column names for the sales data\n")
```

```
## Improved column names for the sales data
```

```
colnames(sales)
```

```
##  [1] "goods"   "1-2018"  "2-2018"  "3-2018"  "4-2018"  "5-2018"  "6-2018"
##  [8] "7-2018"  "8-2018"  "9-2018"  "10-2018" "11-2018" "12-2018" "1-2019"
## [15] "2-2019"  "3-2019"  "4-2019"  "5-2019"  "6-2019"  "7-2019"  "8-2019"
## [22] "9-2019"  "10-2019"
```

Making some Tibbles

"{r} people <- tibble(generator = surv$generator, name = surv$name, sex = surv$sex, major = surv$major, grade_level = surv$grade_level, year_born = surv$year_born, instrument = surv$instrument, time = surv$time_sub)

song_list <- tibble(song_artist = surv$song_artist, song = surv$song, link = surv$link, name = surv$name)

tempPref <- pref[2:45]

temp <- gather(tempPref, key = "song_name", value = "rating", -name)

ratings <- tibble(persName = temp$name, songName = temp$song_name, rating = temp$rating)

people$time <- as.POSIXlt(parse_datetime(people$time, format = "%D %H:%M"))

people$major <- as.factor(people$major)

levels(people$major) "'