Purrr and factors

Bryana Benson
Conner Bryan

Libraries

```
library("readxl")
library("purrr")
library("tibble")
library("tidyverse")
library("dplyr")
```

Combining multiple spreadsheets in a workbook

Download the spreadsheet https://www.census.gov/retail/mrts/www/mrtssales92-present.xls. Combine the Agjusted Range of each spreadsheet in the excel workbook to a single dataframe.

```
#Excel spreadsheet in Project Directory
path <- "mrtssales92-present.xls"

#Combine data in each worksheet in excel workbook using the specified range
data <- path %>%
  excel_sheets() %>%
  set_names() %>%
  map_df(~ read_excel(path = path, sheet = .x, range = "A72:N110"), .id = "sheet")
```

```
## readxl works best with a newer version of the tibble package.
## You currently have tibble v1.4.2.
## Falling back to column name repair from tibble <= v1.4.2.
## Message displays once per session.
```

```
#Convert to tibble
as_tibble(data)
```

```
## # A tibble: 1,026 x 15
##    sheet X__1  `ADJUSTED(2)`   X__2   X__3   X__4   X__5   X__6   X__7
##    <chr> <chr> <chr>          <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 2018  <NA>  Retail and f~ 492034 492530 496077 497776 503955 505168
##  2 2018  <NA>  Retail sales~ 390976 392572 394170 395627 401215 402499
##  3 2018  <NA>  Retail sales~ 450425 450842 454370 455899 460881 461838
##  4 2018  <NA>  Retail sales~ 349367 350884 352463 353750 358141 359169
##  5 2018  <NA>  Retail sales~ 434632 434859 437764 439586 444141 444367
##  6 2018  <NA>  Retail sales~ 333574 334901 335857 337437 341401 341698
##  7 2018  <NA>  GAFO(1)       108084 108483 108939 109266 110515 110162
##  8 2018  441   Motor vehicl~ 101058  99958 101907 102149 102740 102669
##  9 2018  4411~ Automobile a~  93580  92557  94411  94653  95173  95159
## 10 2018  4413  Automotive p~   7478   7401   7496   7496   7567   7510
## # ... with 1,016 more rows, and 6 more variables: X__8 <dbl>, X__9 <dbl>,
## #   X__10 <dbl>, X__11 <dbl>, X__12 <dbl>, X__13 <dbl>
```

Normalizing data

Select the columns that include Year, Code, Kind of Business, and Months. Assign names to columns.

```
#Select columns
data <- data[,c(3,1,4:15)]

#Add names to columns
names(data)[1:14] <- c("Business","Year", "Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep"

#Preview data
head(data)
```

```
## # A tibble: 6 x 14
##   Business Year     Jan    Feb    Mar    Apr    May    Jun    Jul    Aug
##   <chr>    <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Retail ~ 2018  492034 492530 496077 497776 503955 505168 508230 507872
## 2 Retail ~ 2018  390976 392572 394170 395627 401215 402499 405525 405935
## 3 Retail ~ 2018  450425 450842 454370 455899 460881 461838 464728 463666
## 4 Retail ~ 2018  349367 350884 352463 353750 358141 359169 362023 361729
## 5 Retail ~ 2018  434632 434859 437764 439586 444141 444367 446711 446408
## 6 Retail ~ 2018  333574 334901 335857 337437 341401 341698 344006 344471
## # ... with 4 more variables: Sep <dbl>, Oct <dbl>, Nov <dbl>, Dec <dbl>
```

Gather data and make 'Month' a factor.

```
#Gather data: Each month column becomes a single column called 'Month' and the values associated are re
data.g<- gather(data,'Month','Sale',3:14)
```

Make Month a factor and specify levels.

```
#Check data types
str(data.g)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    12312 obs. of  4 variables:
##  $ Business: chr  "Retail and food services sales, total" "Retail sales and food services excl motor
##  $ Year    : chr  "2018" "2018" "2018" "2018" ...
##  $ Month   : chr  "Jan" "Jan" "Jan" "Jan" ...
##  $ Sale    : num  492034 390976 450425 349367 434632 ...
```

```
#Specify levels and Make Month a factor
levels <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")

data.g$Month <- factor(data.g$Month, levels)

#Check data types again
str(data.g)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    12312 obs. of  4 variables:
##  $ Business: chr  "Retail and food services sales, total" "Retail sales and food services excl motor
##  $ Year    : chr  "2018" "2018" "2018" "2018" ...
##  $ Month   : Factor w/ 12 levels "Jan","Feb","Mar",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Sale    : num  492034 390976 450425 349367 434632 ...
```

```
#Preview 20 rows of data
head(data.g, n = 20)
```

```
## # A tibble: 20 x 4
##    Business                                          Year  Month   Sale
##    <chr>                                             <chr> <fct>  <dbl>
##  1 Retail and food services sales, total             2018  Jan   492034
##  2 Retail sales and food services excl motor vehicle an~ 2018  Jan   390976
```

```
##  3 Retail sales and food services excl gasoline stations 2018  Jan   450425
##  4 Retail sales and food services excl motor vehicle an~ 2018  Jan   349367
##  5 Retail sales, total                                   2018  Jan   434632
##  6 Retail sales, total (excl. motor vehicle and parts d~ 2018  Jan   333574
##  7 GAFO(1)                                               2018  Jan   108084
##  8 Motor vehicle and parts dealers                       2018  Jan   101058
##  9 Automobile and other motor vehicle dealers            2018  Jan    93580
## 10 Automotive parts, acc., and tire stores               2018  Jan     7478
## 11 Furniture, home furn, electronics, and appliance sto~ 2018  Jan    18208
## 12 Furniture and home furnishings stores                 2018  Jan     9984
## 13 Electronics and appliance stores                      2018  Jan     8224
## 14 Building mat. and garden equip. and supplies dealers  2018  Jan    31975
## 15 Building mat. and supplies dealers                    2018  Jan    27658
## 16 Food and beverage stores                              2018  Jan    61145
## 17 Grocery stores                                        2018  Jan    54380
## 18 Beer, wine and liquor stores                          2018  Jan     4743
## 19 Health and personal care stores                       2018  Jan    28136
## 20 Pharmacies and drug stores                            2018  Jan    23345
```

Calculations

Compute average "Retail and food services sales, total" for each month.

```
data.g %>%
  filter(str_detect(Business, "Retail and food services")) %>%
  group_by(Month) %>%
  summarize("Average Sales" = mean(Sale))
```

```
## # A tibble: 12 x 2
##    Month `Average Sales`
##    <fct>           <dbl>
##  1 Jan           319512.
##  2 Feb           320487.
##  3 Mar           322163.
##  4 Apr           323272.
##  5 May           324319.
##  6 Jun           325738
##  7 Jul           326883.
##  8 Aug           328044.
##  9 Sep           328608.
## 10 Oct           330086.
## 11 Nov               NA
## 12 Dec               NA
```

Get rid of null values in the dataset that is causing

```
#omit NAs
data.g <- na.omit(data.g)
```

```
#Rerun the previous code
data.g %>%
  filter(str_detect(Business, "Retail and food services")) %>%
  group_by(Month) %>%
  summarize("Average Sales" = mean(Sale))
```

```
## # A tibble: 12 x 2
##    Month `Average Sales`
```

```
##    <fct>          <dbl>
##  1 Jan          319512.
##  2 Feb          320487.
##  3 Mar          322163.
##  4 Apr          323272.
##  5 May          324319.
##  6 Jun          325738 
##  7 Jul          326883.
##  8 Aug          328044.
##  9 Sep          328608.
## 10 Oct          330086.
## 11 Nov          323837.
## 12 Dec          324600.
```