McKibben Ex. 5.2 R Markdown

Makayla McKibben

2024-07-10

```
knitr::opts_chunk$set(echo = TRUE)
# Makayla McKibben
# DSC520
# Week 5
# Assignment 5.2

#install.packages("purrr")
library(purrr)
```

```
## Warning: package 'purrr' was built under R version 4.4.1
```

```
#install.packages("tidyverse")
library(stringr)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.1
```

```
##
## 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
```

```
# Importing the data set
housing <- read.csv(file = 'Housing.csv', header = TRUE,
                    sep =",", stringsAsFactors = FALSE)

# Checking what's in the data set
head(str(housing))
```

```
## 'data.frame':    12865 obs. of  24 variables:
##  $ Sale.Date              : chr  "1/3/2006" "1/3/2006" "1/3/2006" "1/3/2006" ...
##  $ Sale.Price             : int  698000 649990 572500 420000 369900 184667 1050000 875000 660000 65
##  $ sale_reason            : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
##  $ sale_instrument         : int  3 3 3 3 3 15 3 3 3 3 ...
##  $ sale_warning            : chr  "" "" "" "" ...
##  $ sitetype                : chr  "R1" "R1" "R1" "R1" ...
##  $ addr_full               : chr  "17021 NE 113TH CT" "11927 178TH PL NE" "13315 174TH AVE NE" "3303
##  $ zip5                    : int  98052 98052 98052 98052 98052 98053 98053 98053 98053 98052 ...
##  $ ctyname                 : chr  "REDMOND" "REDMOND" "" "REDMOND" ...
##  $ postalctyn              : chr  "REDMOND" "REDMOND" "REDMOND" "REDMOND" ...
##  $ lon                     : num  -122 -122 -122 -122 -122 ...
##  $ lat                     : num  47.7 47.7 47.7 47.6 47.7 ...
##  $ building_grade          : int  9 9 8 8 7 7 10 10 9 8 ...
##  $ square_feet_total_living: int  2810 2880 2770 1620 1440 4160 3960 3720 4160 2760 ...
##  $ bedrooms                : int  4 4 4 3 3 4 5 4 4 4 ...
##  $ bath_full_count         : int  2 2 1 1 1 2 3 2 2 1 ...
##  $ bath_half_count         : int  1 0 1 0 0 1 0 1 1 0 ...
##  $ bath_3qtr_count         : int  0 1 1 1 1 1 1 0 1 1 ...
##  $ year_built              : int  2003 2006 1987 1968 1980 2005 1993 1988 1978 1976 ...
##  $ year_renovated          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ current_zoning          : chr  "R4" "R4" "R6" "R4" ...
##  $ sq_ft_lot               : int  6635 5570 8444 9600 7526 7280 97574 30649 42688 94889 ...
##  $ prop_type               : chr  "R" "R" "R" "R" ...
##  $ present_use             : int  2 2 2 2 2 2 2 2 2 2 ...
```

```
## NULL
```

```
# Searching for houses in my price range
f_house <- filter(housing, between(Sale.Price,
                                   400000, 500000), .by = Sale.Price) %>%
  arrange( , desc(Sale.Price), .by_group = TRUE, .locale = "en")

# Checking the results
head(f_house)
```

```
##   Sale.Date Sale.Price sale_reason sale_instrument sale_warning sitetype
## 1 2/21/2006     500000           1               3                    R1
## 2 3/14/2006     500000           1               3           51       R1
## 3  5/5/2006     500000           1               3                    R1
## 4 5/11/2006     500000           1               3                    R1
## 5 6/25/2006     500000           1               3                    R1
## 6 6/27/2006     500000           1               2           15       R1
##            addr_full  zip5 ctyname postalctyn       lon      lat building_grade
## 1   18612 NE 25TH ST 98052 REDMOND    REDMOND -122.0903 47.63186              8
## 2  13822 181ST LN NE 98052           REDMOND -122.0990 47.72501              9
## 3   9903 170TH PL NE 98052 REDMOND    REDMOND -122.1138 47.68884              8
## 4 13621 179TH AVE NE 98052           REDMOND -122.1016 47.72238              8
## 5  8030 199TH AVE NE 98053           REDMOND -122.0750 47.67598              7
## 6  16615 NE 108TH PL 98052 REDMOND    REDMOND -122.1180 47.69673              7
##   square_feet_total_living bedrooms bath_full_count bath_half_count
## 1                     2340        4               1               0
## 2                     2720        3               2               1
## 3                     2560        4               1               0
## 4                     1940        4               2               1
## 5                     2900        5               2               0
## 6                     1850        3               1               0
```

```
##   bath_3qtr_count year_built year_renovated current_zoning sq_ft_lot prop_type
## 1               1       1975              0             R3     11745         R
## 2               0       1984              0            RA5     42848         R
## 3               1       1979              0             R5      7425         R
## 4               0       1983              0             R6      8743         R
## 5               1       1952              0            RA5     24750         R
## 6               1       1977              0             R4      9600         R
##   present_use
## 1           2
## 2           2
## 3           2
## 4           2
## 5           2
## 6           2
```

```
# Splitting the data
head(select(housing, c(Sale.Price, square_feet_total_living, sq_ft_lot)))
```

```
##   Sale.Price square_feet_total_living sq_ft_lot
## 1     698000                     2810      6635
## 2     649990                     2880      5570
## 3     572500                     2770      8444
## 4     420000                     1620      9600
## 5     369900                     1440      7526
## 6     184667                     4160      7280
```

```
# Organizing the data
head(arrange(housing, desc(square_feet_total_living),
        .by_group = TRUE, .locale = "en"))
```

```
##    Sale.Date Sale.Price sale_reason sale_instrument sale_warning sitetype
## 1 11/19/2012    2300000           4              18        13 31       R1
## 2  5/21/2013    1300000           1              22                    R1
## 3   5/5/2014    2280000           1               3                    R1
## 4 11/30/2012    3000000           4              18        15 31       R1
## 5 12/11/2012    2491149           4              18     13 15 31       R1
## 6  2/26/2008    3995000           1               3                    R1
##                      addr_full  zip5 ctyname postalctyn       lon      lat
## 1              26408 NE 70TH ST 98053           REDMOND -121.9857 47.66751
## 2              26408 NE 70TH ST 98053           REDMOND -121.9857 47.66751
## 3              26408 NE 70TH ST 98053           REDMOND -121.9857 47.66751
## 4 2222 W LAKE SAMMAMISH PKWY NE 98052 REDMOND    REDMOND -122.0874 47.62981
## 5 2222 W LAKE SAMMAMISH PKWY NE 98052 REDMOND    REDMOND -122.0874 47.62981
## 6            13707 160TH AVE NE 98052           REDMOND -122.1272 47.72393
##   building_grade square_feet_total_living bedrooms bath_full_count
## 1             12                    13540        7               1
## 2             12                    13540        7               1
## 3             12                    13540        7               1
## 4             12                    13210        4               2
## 5             12                    13210        4               2
## 6             13                    11810        7               4
##   bath_half_count bath_3qtr_count year_built year_renovated current_zoning
## 1               2               8       1999              0           RA10
```

```
## 2              2              8       1999             0           RA10
## 3              2              8       1999             0           RA10
## 4              3              4       2008             0             R4
## 5              3              4       2008             0             R4
## 6              1              4       2000             0        RA2.5SO
##   sq_ft_lot prop_type present_use
## 1    307752         R           2
## 2    307752         R           2
## 3    307752         R           2
## 4     29728         R           2
## 5     29728         R           2
## 6    139392         R           2
```

```
head(arrange(housing, desc(sq_ft_lot), .by_group = TRUE, .locale = "en"))
```

```
##   Sale.Date Sale.Price sale_reason sale_instrument sale_warning sitetype
## 1 6/13/2013      14000           1              26                    R1
## 2  3/2/2010    4400000           1               3        35 45       R1
## 3 3/29/2016    2165000           1               3                    R1
## 4 3/29/2016    2165000           1               3                    R1
## 5  7/6/2010        698           1              26           24       R1
## 6  7/6/2010        698           1              26           24       R1
##                  addr_full  zip5 ctyname postalctyn       lon      lat
## 1         20210 NE 85TH ST 98053            REDMOND -122.0724 47.68034
## 2        12053 154TH PL NE 98052            REDMOND -122.1345 47.70950
## 3       11207 248TH AVE NE 98053            REDMOND -122.0090 47.69878
## 4       11207 248TH AVE NE 98053            REDMOND -122.0090 47.69878
## 5 19805 NE NOVELTY HILL RD 98053            REDMOND -122.0741 47.68643
## 6 19805 NE NOVELTY HILL RD 98053            REDMOND -122.0741 47.68643
##   building_grade square_feet_total_living bedrooms bath_full_count
## 1             12                     8750        5               2
## 2              6                     2410        3               1
## 3             11                     3690        4               2
## 4              9                     1230        1               1
## 5             12                     5830        4               4
## 6              5                     1040        3               1
##   bath_half_count bath_3qtr_count year_built year_renovated current_zoning
## 1               2               3       1996             0            RA5
## 2               0               1       1935             0          A10SO
## 3               1               2       1999             0          RA10P
## 4               0               0       1999             0          RA10P
## 5               0               1       1969             0           RA5P
## 6               0               0       1900             0           RA5P
##   sq_ft_lot prop_type present_use
## 1   1631322         R           2
## 2   1327090         R           2
## 3   1166246         R         300
## 4   1166246         R         300
## 5   1127205         R           2
## 6   1127205         R           2
```

```
# Grabbing some information by using the mean function
summarize(housing, avg = mean(Sale.Price))
```

```
##        avg
## 1 660737.7
```

```
summarize(housing, avg = mean(square_feet_total_living))
```

```
##        avg
## 1 2539.506
```

```
summarize(housing, avg = mean(sq_ft_lot))
```

```
##        avg
## 1 22228.57
```

```
# Grouping and checking the data
g_sale_price <- group_by(housing, Sale.Price,
                         square_feet_total_living, sq_ft_lot,
                         .add = TRUE, .drop = TRUE)
head(g_sale_price)
```

```
## # A tibble: 6 x 24
## # Groups:   Sale.Price, square_feet_total_living, sq_ft_lot [6]
##   Sale.Date Sale.Price sale_reason sale_instrument sale_warning sitetype
##   <chr>          <int>       <int>           <int> <chr>        <chr>
## 1 1/3/2006      698000           1               3 ""           R1
## 2 1/3/2006      649990           1               3 ""           R1
## 3 1/3/2006      572500           1               3 ""           R1
## 4 1/3/2006      420000           1               3 ""           R1
## 5 1/3/2006      369900           1               3 "15"         R1
## 6 1/3/2006      184667           1              15 "18 51"      R1
## # i 18 more variables: addr_full <chr>, zip5 <int>, ctyname <chr>,
## #   postalctyn <chr>, lon <dbl>, lat <dbl>, building_grade <int>,
## #   square_feet_total_living <int>, bedrooms <int>, bath_full_count <int>,
## #   bath_half_count <int>, bath_3qtr_count <int>, year_built <int>,
## #   year_renovated <int>, current_zoning <chr>, sq_ft_lot <int>,
## #   prop_type <chr>, present_use <int>
```

```
# Cutting the data down to fields of interest
g_sale_price <- mutate(g_sale_price, .keep = c("used"))

# Split the data to keep the prices that are higher than average
higher_than_avg <- keep(housing$Sale.Price, housing$Sale.Price > mean(housing$Sale.Price))
head(higher_than_avg)
```

```
## [1]  698000 1050000  875000  803000  765000  765000
```

```
# Split the data to remove prices higher than average
lower_than_avg <- discard(housing$Sale.Price, housing$Sale.Price > mean(housing$Sale.Price))
head(lower_than_avg)
```

```
## [1] 649990 572500 420000 369900 184667 660000
```

```r
# rbind
highest_priced <- arrange(housing, desc(Sale.Price),
                          .by_group = TRUE, .locale = "en")
lowest_priced <- arrange(housing, desc(Sale.Price),
                         .by_group = TRUE, .locale = "en")
top_and_bottom <- rbind(highest_priced %>% slice(1:5),
                        lowest_priced %>% slice((nrow(lowest_priced)-4):nrow(lowest_priced)))
top_and_bottom
```

```
##     Sale.Date Sale.Price sale_reason sale_instrument sale_warning sitetype
## 1    3/2/2010    4400000           1               3        35 45       R1
## 2    3/2/2010    4400000           1               3        35 45       R1
## 3  11/17/2011    4380542           1              22        11 45       R1
## 4  11/17/2011    4380542           1              22        11 45       R1
## 5  11/17/2011    4380542           1              22        11 45       R1
## 6  12/22/2009        998           1              26           24       R1
## 7  12/29/2009        873           1              26           24       R1
## 8   1/28/2010        873           1              26        24 32       R1
## 9    7/6/2010        698           1              26           24       R1
## 10   7/6/2010        698           1              26           24       R1
##                   addr_full  zip5 ctyname postalctyn       lon      lat
## 1         12025 154TH PL NE 98052            REDMOND -122.1350 47.70801
## 2         12053 154TH PL NE 98052            REDMOND -122.1345 47.70950
## 3         17137 NE 120TH ST 98052 REDMOND    REDMOND -122.1113 47.70674
## 4         11818 171ST PL NE 98052 REDMOND    REDMOND -122.1119 47.70639
## 5        17011 NE 118TH WAY 98052 REDMOND    REDMOND -122.1134 47.70580
## 6         8226 196TH AVE NE 98053            REDMOND -122.0777 47.67746
## 7         8332 196TH AVE NE 98053            REDMOND -122.0782 47.67802
## 8         8340 196TH AVE NE 98053            REDMOND -122.0784 47.67845
## 9  19805 NE NOVELTY HILL RD 98053            REDMOND -122.0741 47.68643
## 10 19805 NE NOVELTY HILL RD 98053            REDMOND -122.0741 47.68643
##    building_grade square_feet_total_living bedrooms bath_full_count
## 1              11                     5790        3               2
## 2               6                     2410        3               1
## 3               8                     3290        4               2
## 4               8                     2450        4               2
## 5               8                     2750        4               2
## 6               7                     1850        3               1
## 7               7                     2160        2               1
## 8               7                     3430        3               1
## 9              12                     5830        4               4
## 10              5                     1040        3               1
##    bath_half_count bath_3qtr_count year_built year_renovated current_zoning
## 1                1               1       1999              0            A10
## 2                0               1       1935              0          A10SO
## 3                1               0       2012              0             R4
## 4                1               0       2010              0             R4
## 5                1               0       2012              0             R4
## 6                1               0       1960           1989            RA5
## 7                0               1       1968              0            RA5
## 8                1               0       1955              0            RA5
## 9                0               1       1969              0          RA5P
## 10               0               0       1900              0          RA5P
```

```
##    sq_ft_lot prop_type present_use
## 1     657816         R           2
## 2    1327090         R           2
## 3       6712         R           2
## 4       4749         R           2
## 5       5816         R           2
## 6     209589         R           2
## 7     102505         R           2
## 8     105660         R           2
## 9    1127205         R           2
## 10   1127205         R           2
```

```
# cbind
sale_price <- housing$Sale.Price
sq_footage_interior <- housing$square_feet_total_living
head(cbind(sale_price, sq_footage_interior))
```

```
##      sale_price sq_footage_interior
## [1,]     698000                2810
## [2,]     649990                2880
## [3,]     572500                2770
## [4,]     420000                1620
## [5,]     369900                1440
## [6,]     184667                4160
```

```
# String split and concatenate
string_5.2 <- "This1is1week15's1assignment"
string_5.2.1 <- str_split_i(string_5.2, "1", 1)
string_5.2.2 <- str_split_i(string_5.2, "1", 2)
string_5.2.3 <- str_split_i(string_5.2, "1", 3)
string_5.2.4 <- str_split_i(string_5.2, "1", 4)
string_5.2.5 <- str_split_i(string_5.2, "1", 5)
string_5.2 <- paste(string_5.2.1, string_5.2.2,
                    string_5.2.3, string_5.2.4,
                    string_5.2.5, collapse = " ")
string_5.2
```

```
## [1] "This is week 5's assignment"
```