

assignment5.2_Chattapadhyay_Kausik.R

kausik

2022-09-30

```
# Assignment: ASSIGNMENT 5.2
# Name: Chattapadhyay, Kausik
# Date: 2022-09-29

## Load the ggplot2 package
library(ggplot2)
library(plyr)
library(dplyr)
library(readxl)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory
setwd("/Users/kausik/desktop/MS Data Science/DSC 520/dsc520-stats-r-assignments")

## Load the Housing Dataset
housingDF <- read_excel("data/week-7-housing.xlsx", sheet="Sheet2")
# Renaming the field names
colnames(housingDF)[2] <- "Sale_Price"
colnames(housingDF)[1] <- "Sale_Date"
str(housingDF)

## tibble [12,865 x 24] (S3: tbl_df/tbl/data.frame)
##  $ Sale_Date      : POSIXct[1:12865], format: "2006-01-03" "2006-01-03" ...
##  $ Sale_Price      : num [1:12865] 698000 649990 572500 420000 369900 ...
##  $ sale_reason      : num [1:12865] 1 1 1 1 1 1 1 1 1 1 ...
##  $ sale_instrument  : num [1:12865] 3 3 3 3 3 15 3 3 3 3 ...
##  $ sale_warning     : chr [1:12865] NA NA NA NA ...
##  $ sitetype         : chr [1:12865] "R1" "R1" "R1" "R1" ...
##  $ addr_full        : chr [1:12865] "17021 NE 113TH CT" "11927 178TH PL NE" "13315 174TH AVE I
##  $ zip5             : num [1:12865] 98052 98052 98052 98052 98052 ...
##  $ ctyname          : chr [1:12865] "REDMOND" "REDMOND" NA "REDMOND" ...
##  $ postalctyn       : chr [1:12865] "REDMOND" "REDMOND" "REDMOND" "REDMOND" ...
##  $ lon              : num [1:12865] -122 -122 -122 -122 -122 ...
##  $ lat              : num [1:12865] 47.7 47.7 47.7 47.6 47.7 ...
##  $ building_grade   : num [1:12865] 9 9 8 8 7 7 10 10 9 8 ...
##  $ square_feet_total_living: num [1:12865] 2810 2880 2770 1620 1440 4160 3960 3720 4160 2760 ...
##  $ bedrooms         : num [1:12865] 4 4 4 3 3 4 5 4 4 4 ...
##  $ bath_full_count  : num [1:12865] 2 2 1 1 1 2 3 2 2 1 ...
##  $ bath_half_count  : num [1:12865] 1 0 1 0 0 1 0 1 1 0 ...
##  $ bath_3qtr_count  : num [1:12865] 0 1 1 1 1 1 1 0 1 1 ...
##  $ year_built       : num [1:12865] 2003 2006 1987 1968 1980 ...
##  $ year_renovated    : num [1:12865] 0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ current_zoning      : chr [1:12865] "R4" "R4" "R6" "R4" ...
## $ sq_ft_lot           : num [1:12865] 6635 5570 8444 9600 7526 ...
## $ prop_type           : chr [1:12865] "R" "R" "R" "R" ...
## $ present_use         : num [1:12865] 2 2 2 2 2 2 2 2 2 2 ...
```

a. Using the dplyr package, use the 6 different operations to analyze/transform the data - GroupBy, Summarize, Mutate, Filter, Select, and Arrange - Remember this isn't just modifying data, you are learning about your data also - so play around and start to understand your dataset in more detail.

```
housingDF %>%
  # Select few important fields from housing dataset
  select(Sale_Date, Sale_Price, zip5, lon, lat, square_feet_total_living:year_built) %>%
  # Calculate new field bath_total
  mutate(bath_total = bath_full_count + bath_half_count + bath_3qtr_count) %>%
  # Remove unnecessary bath related fields.
  select(-bath_full_count, -bath_half_count, -bath_3qtr_count) %>%
  # Filter the records from Jan 2015 having sale price 100K and living area > 1500 SQFT
  filter(Sale_Date >= "2015-01-01", Sale_Price > 100000,
         square_feet_total_living > 1500) %>%
  # Grouping the data by year_built and zip codes
  group_by(year_built, zip5) %>%
  # Summarize the data by calculating median sales price and average square feet
  summarize(median_sales_price = median(Sale_Price),
            avg_sqft = mean(square_feet_total_living)) %>%
  # Sort the data by median sales price in descending order
  arrange(desc(median_sales_price))
```

'summarise()' has grouped output by 'year_built'. You can override using the '.groups' argument.

```
## # A tibble: 152 x 4
## # Groups:   year_built [81]
##   year_built zip5 median_sales_price avg_sqft
##   <dbl> <dbl>          <dbl>    <dbl>
## 1    1948 98053          2025000    2610
## 2    1938 98052          1700000    2770
## 3    1955 98052          1562500    2550
## 4    1956 98053          1500000    4280
## 5    1949 98052          1300000    3840
## 6    1956 98074          1300000    3470
## 7    2015 98053          1244995    4412.
## 8    1988 98074          1150000    3450
## 9    1999 98074          1060850    4205
## 10   2014 98053          1014990    4058.
## # ... with 142 more rows
```

b. Using the purrr package - perform 2 functions on your dataset. You could use zip_n, keep, discard, compact, etc.

```
library(purrr)
```

```
housingDF %>%
  select(Sale_Date, Sale_Price, zip5, square_feet_total_living:year_built) %>%
```

```
split(.$zip5) %>%
map(summary)
```

```
## $'98052'
##   Sale_Date           Sale_Price      zip5
##   Min.   :2006-01-03 00:00:00   Min.   : 2031   Min.   :98052
##   1st Qu.:2008-07-04 12:00:00   1st Qu.: 463000 1st Qu.:98052
##   Median :2011-12-23 00:00:00   Median : 599950 Median :98052
##   Mean   :2011-08-27 13:20:00   Mean   : 649375 Mean   :98052
##   3rd Qu.:2014-07-25 00:00:00   3rd Qu.: 740000 3rd Qu.:98052
##   Max.   :2016-12-16 00:00:00   Max.   :4400000 Max.   :98052
##   square_feet_total_living bedrooms bath_full_count bath_half_count
##   Min.   : 310           Min.   : 0.00   Min.   :0.000   Min.   :0.0000
##   1st Qu.: 1860           1st Qu.: 3.00   1st Qu.:1.000   1st Qu.:0.0000
##   Median : 2430           Median : 4.00   Median :2.000   Median :1.0000
##   Mean   : 2499           Mean   : 3.68   Mean   :1.738   Mean   :0.6116
##   3rd Qu.: 3070           3rd Qu.: 4.00   3rd Qu.:2.000   3rd Qu.:1.0000
##   Max.   :13210          Max.   :11.00   Max.   :6.000   Max.   :6.0000
##   bath_3qtr_count year_built
##   Min.   :0.0000   Min.   :1900
##   1st Qu.:0.0000   1st Qu.:1976
##   Median :0.0000   Median :1986
##   Mean   :0.5776   Mean   :1988
##   3rd Qu.:1.0000   3rd Qu.:2006
##   Max.   :7.0000   Max.   :2016
##
## $'98053'
##   Sale_Date           Sale_Price      zip5
##   Min.   :2006-01-03 00:00:00   Min.   : 698   Min.   :98053
##   1st Qu.:2008-07-12 12:00:00   1st Qu.: 455000 1st Qu.:98053
##   Median :2011-09-07 00:00:00   Median : 584000 Median :98053
##   Mean   :2011-06-15 19:46:28   Mean   : 672624 Mean   :98053
##   3rd Qu.:2014-02-25 00:00:00   3rd Qu.: 752500 3rd Qu.:98053
##   Max.   :2016-12-15 00:00:00   Max.   :3850000 Max.   :98053
##   square_feet_total_living bedrooms bath_full_count bath_half_count
##   Min.   : 240           Min.   :0.00   Min.   : 0.000   Min.   :0.0000
##   1st Qu.: 1720           1st Qu.:2.00   1st Qu.: 2.000   1st Qu.:0.0000
##   Median : 2390           Median :3.00   Median : 2.000   Median :1.0000
##   Mean   : 2580           Mean   :3.19   Mean   : 1.875   Mean   :0.6134
##   3rd Qu.: 3200           3rd Qu.:4.00   3rd Qu.: 2.000   3rd Qu.:1.0000
##   Max.   :13540          Max.   :8.00   Max.   :23.000   Max.   :8.0000
##   bath_3qtr_count year_built
##   Min.   :0.0000   Min.   :1900
##   1st Qu.:0.0000   1st Qu.:1994
##   Median :0.0000   Median :2004
##   Mean   :0.3812   Mean   :1999
##   3rd Qu.:1.0000   3rd Qu.:2008
##   Max.   :8.0000   Max.   :2016
##
## $'98059'
##   Sale_Date           Sale_Price      zip5 square_feet_total_living
##   Min.   :2013-05-23   Min.   :645000   Min.   :98059   Min.   :4360
##   1st Qu.:2013-05-23   1st Qu.:645000   1st Qu.:98059   1st Qu.:4360
```

```
## Median :2013-05-23 Median :645000 Median :98059 Median :4360
## Mean :2013-05-23 Mean :645000 Mean :98059 Mean :4360
## 3rd Qu.:2013-05-23 3rd Qu.:645000 3rd Qu.:98059 3rd Qu.:4360
## Max. :2013-05-23 Max. :645000 Max. :98059 Max. :4360
## bedrooms bath_full_count bath_half_count bath_3qtr_count year_built
## Min. :4 Min. :2 Min. :0 Min. :1 Min. :2003
## 1st Qu.:4 1st Qu.:2 1st Qu.:0 1st Qu.:1 1st Qu.:2003
## Median :4 Median :2 Median :0 Median :1 Median :2003
## Mean :4 Mean :2 Mean :0 Mean :1 Mean :2003
## 3rd Qu.:4 3rd Qu.:2 3rd Qu.:0 3rd Qu.:1 3rd Qu.:2003
## Max. :4 Max. :2 Max. :0 Max. :1 Max. :2003
##
## $'98074'
## Sale_Date Sale_Price zip5
## Min. :2006-04-14 00:00:00 Min. : 434000 Min. :98074
## 1st Qu.:2007-08-27 00:00:00 1st Qu.: 650000 1st Qu.:98074
## Median :2012-04-24 00:00:00 Median : 820000 Median :98074
## Mean :2011-10-03 09:51:46 Mean : 951544 Mean :98074
## 3rd Qu.:2015-02-06 00:00:00 3rd Qu.:1082500 3rd Qu.:98074
## Max. :2016-10-07 00:00:00 Max. :2160200 Max. :98074
## square_feet_total_living bedrooms bath_full_count bath_half_count
## Min. :1010 Min. :2 Min. :0.000 Min. :0.0000
## 1st Qu.:2550 1st Qu.:4 1st Qu.:2.000 1st Qu.:0.0000
## Median :3810 Median :4 Median :2.000 Median :1.0000
## Mean :3682 Mean :4 Mean :2.411 Mean :0.7945
## 3rd Qu.:4510 3rd Qu.:4 3rd Qu.:3.000 3rd Qu.:1.0000
## Max. :6280 Max. :6 Max. :4.000 Max. :2.0000
## bath_3qtr_count year_built
## Min. :0.0000 Min. :1955
## 1st Qu.:0.0000 1st Qu.:1998
## Median :0.0000 Median :1999
## Mean :0.2055 Mean :1996
## 3rd Qu.:0.0000 3rd Qu.:2002
## Max. :2.0000 Max. :2006
```

```
## keep() and discard() allow you to filter a vector based on a predicate function.
housingDF %>%
  select(Sale_Date, Sale_Price, zip5, square_feet_total_living:year_built) %>%
  keep(function(x) mean(x) > 200000)
```

```
## # A tibble: 12,865 x 2
## Sale_Date Sale_Price
## <dtm> <dbl>
## 1 2006-01-03 00:00:00 698000
## 2 2006-01-03 00:00:00 649990
## 3 2006-01-03 00:00:00 572500
## 4 2006-01-03 00:00:00 420000
## 5 2006-01-03 00:00:00 369900
## 6 2006-01-03 00:00:00 184667
## 7 2006-01-04 00:00:00 1050000
## 8 2006-01-04 00:00:00 875000
## 9 2006-01-04 00:00:00 660000
## 10 2006-01-04 00:00:00 650000
## # ... with 12,855 more rows
```

```
housingDF %>%
  select(Sale_Date, Sale_Price, zip5, square_feet_total_living:year_built) %>%
  discard(function(x) mean(x) > 200000)
```

```
## # A tibble: 12,865 x 7
##   zip5 square_feet_total_living bedrooms bath_full_count bath_half_~1 bath_~2 year_~3
##   <dbl>           <dbl>         <dbl>         <dbl>         <dbl>   <dbl>   <dbl>
## 1 98052             2810           4             2             1       0     2003
## 2 98052             2880           4             2             0       1     2006
## 3 98052             2770           4             1             1       1     1987
## 4 98052             1620           3             1             0       1     1968
## 5 98052             1440           3             1             0       1     1980
## 6 98053             4160           4             2             1       1     2005
## 7 98053             3960           5             3             0       1     1993
## 8 98053             3720           4             2             1       0     1988
## 9 98053             4160           4             2             1       1     1978
## 10 98052            2760           4             1             0       1     1976
## # ... with 12,855 more rows, and abbreviated variable names 1: bath_half_count,
## # 2: bath_3qtr_count, 3: year_built
```

compact() is a helpful wrapper that throws away empty elements of a list.

```
housingDF %>%
  select(Sale_Date, Sale_Price, zip5, square_feet_total_living:year_built) %>%
  compact()
```

```
## # A tibble: 12,865 x 9
##   Sale_Date      Sale_Price zip5 squar~1 bedro~2 bath_~3 bath_~4 bath_~5 year_~6
##   <dtm>           <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 2006-01-03 00:00:00  698000 98052   2810     4       2       1       0     2003
## 2 2006-01-03 00:00:00  649990 98052   2880     4       2       0       1     2006
## 3 2006-01-03 00:00:00  572500 98052   2770     4       1       1       1     1987
## 4 2006-01-03 00:00:00  420000 98052   1620     3       1       0       1     1968
## 5 2006-01-03 00:00:00  369900 98052   1440     3       1       0       1     1980
## 6 2006-01-03 00:00:00  184667 98053   4160     4       2       1       1     2005
## 7 2006-01-04 00:00:00  1050000 98053   3960     5       3       0       1     1993
## 8 2006-01-04 00:00:00   875000 98053   3720     4       2       1       0     1988
## 9 2006-01-04 00:00:00   660000 98053   4160     4       2       1       1     1978
## 10 2006-01-04 00:00:00   650000 98052   2760     4       1       0       1     1976
## # ... with 12,855 more rows, and abbreviated variable names
## # 1: square_feet_total_living, 2: bedrooms, 3: bath_full_count, 4: bath_half_count,
## # 5: bath_3qtr_count, 6: year_built
```

c. Use the cbind and rbind function on your dataset

Splitting the data frame in two data frames.

```
housingDF1 <- housingDF %>%
  select(Sale_Date, Sale_Price, zip5)

housingDF2 <- housingDF %>%
  select(square_feet_total_living:year_built)
```

cbind() function to join two sets of columns together into a single dataframe.

```
housingDF3 <- cbind(housingDF1, housingDF2)
head(housingDF3)
```

```
##   Sale_Date Sale_Price zip5 square_feet_total_living bedrooms bath_full_count
## 1 2006-01-03   698000 98052             2810           4             2
## 2 2006-01-03   649990 98052             2880           4             2
## 3 2006-01-03   572500 98052             2770           4             1
## 4 2006-01-03   420000 98052             1620           3             1
## 5 2006-01-03   369900 98052             1440           3             1
## 6 2006-01-03   184667 98053             4160           4             2
##   bath_half_count bath_3qtr_count year_built
## 1                1                0       2003
## 2                0                1       2006
## 3                1                1       1987
## 4                0                1       1968
## 5                0                1       1980
## 6                1                1       2005
```

rbind() to combine dataframes by rows.

```
housingDFX <- housingDF %>%
  select(Sale_Date, Sale_Price, zip5, square_feet_total_living:year_built) %>%
  filter(Sale_Date > "2016-01-01")
tail(housingDFX)
```

```
## # A tibble: 6 x 9
##   Sale_Date      Sale_Price zip5 square~1 bedro~2 bath_~3 bath_~4 bath_~5 year_~6
##   <dtm>          <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 2016-12-15 00:00:00   824000 98052    1980     3       2       1       0    2013
## 2 2016-12-15 00:00:00   798930 98053    2920     3       2       1       0    2001
## 3 2016-12-15 00:00:00   750000 98052    2320     4       1       1       1    1980
## 4 2016-12-15 00:00:00   629000 98052    2000     4       1       0       1    1967
## 5 2016-12-16 00:00:00   835000 98052    2460     4       2       1       0    1990
## 6 2016-12-16 00:00:00   455500 98052    1150     3       1       0       0    1961
## # ... with abbreviated variable names 1: square_feet_total_living, 2: bedrooms,
## #   3: bath_full_count, 4: bath_half_count, 5: bath_3qtr_count, 6: year_built
```

```
housingDFY <- housingDF %>%
  select(Sale_Date, Sale_Price, zip5, square_feet_total_living:year_built) %>%
  filter(Sale_Date > "2015-01-01", Sale_Date < "2016-01-01")
tail(housingDFY)
```

```
## # A tibble: 6 x 9
##   Sale_Date      Sale_Price zip5 square~1 bedro~2 bath_~3 bath_~4 bath_~5 year_~6
##   <dtm>          <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 2015-12-30 00:00:00   612500 98052    1360     3       1       0       1    1978
## 2 2015-12-30 00:00:00   602000 98053    2560     4       2       1       0    1991
## 3 2015-12-30 00:00:00   535000 98052    1610     3       1       1       0    1961
## 4 2015-12-31 00:00:00   934900 98052    3570     4       2       1       1    2015
## 5 2015-12-31 00:00:00   775000 98052    2980     5       2       1       0    2007
## 6 2015-12-31 00:00:00   691000 98052    2310     4       2       1       0    1988
## # ... with abbreviated variable names 1: square_feet_total_living, 2: bedrooms,
## #   3: bath_full_count, 4: bath_half_count, 5: bath_3qtr_count, 6: year_built
```

```
housingDFXY <- rbind(housingDFX, housingDFY)
tail(housingDFXY)
```

```
## # A tibble: 6 x 9
##   Sale_Date      Sale_Price zip5 square~1 bedro~2 bath~3 bath~4 bath~5 year~6
##   <dtm>          <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 2015-12-30 00:00:00    612500 98052    1360     3       1       0       1    1978
## 2 2015-12-30 00:00:00    602000 98053    2560     4       2       1       0    1991
## 3 2015-12-30 00:00:00    535000 98052    1610     3       1       1       0    1961
## 4 2015-12-31 00:00:00    934900 98052    3570     4       2       1       1    2015
## 5 2015-12-31 00:00:00    775000 98052    2980     5       2       1       0    2007
## 6 2015-12-31 00:00:00    691000 98052    2310     4       2       1       0    1988
## # ... with abbreviated variable names 1: square_foot_total_living, 2: bedrooms,
## #   3: bath_full_count, 4: bath_half_count, 5: bath_3qtr_count, 6: year_built
```

d. Split a string, then concatenate the results back together

```
library(stringr)

addrList <- str_split(housingDF$addr_full, pattern=" ")
addrMatrix <- data.frame(Reduce(rbind, addrList))
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```


[illegible]

[illegible]

[illegible]

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

[illegible]

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```



```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

[illegible]

[illegible]

[illegible]

[illegible]


```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

[illegible]

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

[illegible]


```
## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)

## Warning in f(init, x[[i]]): number of columns of result is not a multiple of vector
## length (arg 2)
```

```
names(addrMatrix) <- c("House_Number", "Street_Addr_1", "Street_Addr_2", "Street_Addr_3")
housingDFA <- housingDF %>%
  select(Sale_Date, Sale_Price, zip5, addr_full)

housingDFA <- cbind(housingDFA, addrMatrix)

housingDFA$full_addr_2 <- paste(housingDFA$House_Number, housingDFA$Street_Addr_1,
                               housingDFA$Street_Addr_2, housingDFA$Street_Addr_3)

head(housingDFA)
```

```
##      Sale_Date Sale_Price zip5      addr_full House_Number Street_Addr_1
## init 2006-01-03   698000 98052 17021 NE 113TH CT      17021          NE
## X    2006-01-03   649990 98052 11927 178TH PL NE      11927        178TH
## X.1  2006-01-03   572500 98052 13315 174TH AVE NE      13315        174TH
## X.2  2006-01-03   420000 98052 3303 178TH AVE NE       3303        178TH
## X.3  2006-01-03   369900 98052 16126 NE 108TH CT      16126          NE
```

## X.4	2006-01-03	184667 98053	8101 229TH DR NE	8101	229TH
##	Street_Addr_2	Sreet_Addr_3	full_addr_2		
## init	113TH	CT	17021 NE 113TH		
## X	PL	NE	11927 178TH PL		
## X.1	AVE	NE	13315 174TH AVE		
## X.2	AVE	NE	3303 178TH AVE		
## X.3	108TH	CT	16126 NE 108TH		
## X.4	DR	NE	8101 229TH DR		