5501-01 programming assignment

AUTHOR
Steve Simon and Leroy Wheeler

PUBLISHED
August 18, 2024

This program reads data on housing prices in Albuquerque, New Mexico in 1993. Find more information in the data dictionary.

Load the tidyverse library

For most of your programs, you should load the tidyverse library. The messages and warnings are suppressed.

```
        library(tidyverse)
```

Read the data and view a brief summary

Use the read_csv function to read the data. The glimpse function will produce a brief summary.

```
        alb <- read_csv(
          file="../data/albuquerque-housing.csv",
          col_names=TRUE,
          col_types="nnnnccc",
          na=".")
        glimpse(alb)
```

```
Rows: 117
Columns: 7
$ price        <dbl> 205000, 208000, 215000, 215000, 199900, 190000, 180000, 1…
$ sqft         <dbl> 2650, 2600, 2664, 2921, 2580, 2580, 2774, 1920, 2150, 171…
$ age          <dbl> 13, NA, 6, 3, 4, 4, 2, 1, NA, 1, 4, 8, 15, 14, 18, NA, 16…
$ features     <dbl> 7, 4, 5, 6, 4, 4, 4, 5, 4, 3, 5, 6, 3, 5, 8, 3, 4, 3, 4, …
$ northeast    <chr> "yes", "yes", "yes", "yes", "yes", "yes", "yes", "yes", "…
```

```
$ custom_build <chr> "yes", "yes", "yes", "yes", "yes", "no", "no", "yes", "no…
$ corner_lot   <chr> "no", "no", "no", "no", "no", "no", "no", "no", "no", "no…
```

Calculate overall means

The summarize_if function produces means, but only for numeric data. You wouldn't want to compute means for data with values "yes" and "no".

```
        alb |>
          summarise_if(is.numeric, mean, na.rm = TRUE)
```

```
# A tibble: 1 × 4
    price  sqft   age features
    <dbl> <dbl> <dbl>    <dbl>
1 106274. 1654.  15.0     3.53
```

Summarize price

The average price of a home, 106 thousand dollars, is quite low because the data comes from 1993.

Summarize sqft

The average size of a house is about 1,600 square feet, which is small by today's standards.

Summarize age

The average age of a home was 15 years, which is a relatively new house.

Summarize features

The average number of features is 3.5, which is low.