In-class Assignment 7

Andrew Shao

2024-10-16

7.1.4 Q1

```
library(r02pro)
```

```
## Warning: package 'r02pro' was built under R version 4.3.3
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
my_ahp <- ahp %>% filter(yr_built < 2000,
                         yr_sold >= 2009,
                         bedroom  %in% c(2, 3))

my_ahp
```

```
## # A tibble: 407 x 56
##    dt_sold    yr_sold mo_sold yr_built yr_remodel bldg_class bldg_type
##    <date>       <dbl>   <dbl>    <dbl>      <dbl>      <dbl> <chr>
##  1 2010-03-25    2010       3     1976       2005         60 1Fam
##  2 2010-01-15    2010       1     1953       2007         20 1Fam
##  3 2010-03-22    2010       3     1900       1993         50 1Fam
##  4 2010-06-06    2010       6     1966       2002         20 1Fam
##  5 2010-05-08    2010       5     1959       1959         20 1Fam
##  6 2009-06-28    2009       6     1963       1963         20 1Fam
##  7 2010-03-02    2010       3     1992       2007         20 1Fam
##  8 2009-04-24    2009       4     1961       1961         20 1Fam
##  9 2010-04-05    2010       4     1965       2009         20 1Fam
## 10 2009-04-26    2009       4     1958       1958         20 1Fam
```

```
## # i 397 more rows
## # i 49 more variables: house_style <chr>, zoning <chr>, neighborhd <chr>,
## #   oa_cond <dbl>, oa_qual <dbl>, func <chr>, liv_area <dbl>, `1fl_area` <dbl>,
## #   `2fl_area` <dbl>, tot_rms <dbl>, bedroom <dbl>, bathroom <dbl>, kit <dbl>,
## #   kit_qual <chr>, central_air <chr>, elect <chr>, bsmt_area <dbl>,
## #   bsmt_cond <chr>, bsmt_exp <chr>, bsmt_fin_qual <chr>, bsmt_ht <chr>,
## #   ext_cond <chr>, ext_cover <chr>, ext_qual <chr>, fdn <chr>, ...
```

7.1.4 Q2

```
my_gm <- gm %>% filter(continent %in% c('Asia', 'Africa'),
                       HDI_category %in% c('medium', 'high'),
                       year == 2006)

my_gm
```

```
## # A tibble: 40 x 33
##    country     year smoking_female smoking_male lungcancer_newcases_female
##    <chr>      <dbl>          <dbl>        <dbl>                       <dbl>
##  1 Armenia     2006             NA           NA                        10.6
##  2 Azerbaijan  2006             NA           NA                         7.93
##  3 Bahrain     2006             NA           NA                        16.3
##  4 Botswana    2006             NA           NA                        10
##  5 China       2006             NA           NA                        23.1
##  6 Cape Verde  2006             NA           NA                         7.7
##  7 Algeria     2006             NA           NA                         4.02
##  8 Egypt       2006             NA           NA                         4.95
##  9 Gabon       2006             NA           NA                         9.21
## 10 Georgia     2006             NA           NA                         7.79
## # i 30 more rows
## # i 28 more variables: lungcancer_newcases_male <dbl>, owid_edu_idx <dbl>,
## #   food_supply <dbl>, average_daily_income <dbl>, sanitation <dbl>,
## #   child_mortality <dbl>, income_per_person <dbl>, HDI <dbl>,
## #   alcohol_male <dbl>, alcohol_female <dbl>, livercancer_newcases_male <dbl>,
## #   livercancer_newcases_female <dbl>, mortality_male <dbl>,
## #   mortality_female <dbl>, cholesterol_fat_in_blood_male <dbl>, ...
```

7.2.5 Q1

```
ahp %>%
  select(date_sold = dt_sold,
         house_type = house_style,
         liv_area,
         sale_price)
```

```
## # A tibble: 2,048 x 4
##    date_sold  house_type liv_area sale_price
##    <date>     <chr>         <dbl>      <dbl>
##  1 2010-03-25 2Story         1479       130.
```

```
##  2 2009-04-10 2Story         2122        NA
##  3 2010-01-15 1Story         1057       109
##  4 2010-04-19 2Story         1444       174
##  5 2010-03-22 1.5Fin         1445       138.
##  6 2010-06-06 1Story         1888       190
##  7 2006-06-14 SFoyer         1072       140
##  8 2010-05-08 1Story         1188       142
##  9 2007-06-14 1Story          924       112.
## 10 2007-09-01 2Story         2080       135
## # i 2,038 more rows
```

7.2.5 Q2

```
ahp %>%
  select(starts_with('yr'),
         ends_with('qual'))
```

```
## # A tibble: 2,048 x 10
##    yr_sold yr_built yr_remodel oa_qual kit_qual bsmt_fin_qual ext_qual fp_qual
##      <dbl>    <dbl>      <dbl>   <dbl> <chr>    <chr>         <chr>    <chr>
##  1    2010     1976       2005       6 Good     GLQ           Good     <NA>
##  2    2009     1996       1997       7 Good     GLQ           Good     Average
##  3    2010     1953       2007       5 Good     GLQ           Average  <NA>
##  4    2010     2006       2007       5 Average  Unf           Average  <NA>
##  5    2010     1900       1993       6 Average  Unf           Average  <NA>
##  6    2010     1966       2002       6 Good     ALQ           Good     Good
##  7    2006     2005       2006       6 Average  GLQ           Average  <NA>
##  8    2010     1959       1959       5 Average  Rec           Average  <NA>
##  9    2007     1952       1952       5 Average  Rec           Average  <NA>
## 10    2007     1969       1969       5 Fair     Rec           Average  Average
## # i 2,038 more rows
## # i 2 more variables: gar_qual <chr>, heat_qual <chr>
```

7.3.1 Q1

```
ahp %>%
  filter(yr_built == 2008,
         house_style == '2Story') %>%
  arrange(yr_remodel, desc(sale_price))
```

```
## # A tibble: 8 x 56
##   dt_sold    yr_sold mo_sold yr_built yr_remodel bldg_class bldg_type
##   <date>       <dbl>   <dbl>    <dbl>      <dbl>      <dbl> <chr>
## 1 2009-06-19    2009       6     2008       2008         60 1Fam
## 2 2008-11-16    2008      11     2008       2008         60 1Fam
## 3 2009-07-11    2009       7     2008       2008         60 1Fam
## 4 2009-04-13    2009       4     2008       2008         60 1Fam
## 5 2009-06-17    2009       6     2008       2008         60 1Fam
## 6 2008-01-27    2008       1     2008       2008         60 1Fam
```

```
## 7 2009-01-07    2009       1     2008       2009         60 1Fam
## 8 2009-07-16    2009       7     2008       2009         60 1Fam
## # i 49 more variables: house_style <chr>, zoning <chr>, neighborhd <chr>,
## #   oa_cond <dbl>, oa_qual <dbl>, func <chr>, liv_area <dbl>, `1fl_area` <dbl>,
## #   `2fl_area` <dbl>, tot_rms <dbl>, bedroom <dbl>, bathroom <dbl>, kit <dbl>,
## #   kit_qual <chr>, central_air <chr>, elect <chr>, bsmt_area <dbl>,
## #   bsmt_cond <chr>, bsmt_exp <chr>, bsmt_fin_qual <chr>, bsmt_ht <chr>,
## #   ext_cond <chr>, ext_cover <chr>, ext_qual <chr>, fdn <chr>, fence <chr>,
## #   fp <dbl>, fp_qual <chr>, gar_area <dbl>, gar_car <dbl>, gar_cond <chr>, ...
```

7.3.1 Q2

```
ahp %>%
  filter(yr_sold == 2009,
         house_style == '1Story') %>%
  arrange(desc(sale_price))
```

```
## # A tibble: 236 x 56
##    dt_sold    yr_sold mo_sold yr_built yr_remodel bldg_class bldg_type
##    <date>       <dbl>   <dbl>    <dbl>      <dbl>      <dbl> <chr>
##  1 2009-04-13    2009       4     2008       2008         20 1Fam
##  2 2009-02-27    2009       2     2004       2005         20 1Fam
##  3 2009-07-24    2009       7     2008       2009         20 1Fam
##  4 2009-10-01    2009      10     2005       2006         20 1Fam
##  5 2009-07-21    2009       7     2003       2003         20 1Fam
##  6 2009-05-17    2009       5     2007       2007         20 1Fam
##  7 2009-07-04    2009       7     2008       2008         20 1Fam
##  8 2009-05-20    2009       5     2008       2008         20 1Fam
##  9 2009-06-18    2009       6     2003       2003         20 1Fam
## 10 2009-07-06    2009       7     2003       2004         20 1Fam
## # i 226 more rows
## # i 49 more variables: house_style <chr>, zoning <chr>, neighborhd <chr>,
## #   oa_cond <dbl>, oa_qual <dbl>, func <chr>, liv_area <dbl>, `1fl_area` <dbl>,
## #   `2fl_area` <dbl>, tot_rms <dbl>, bedroom <dbl>, bathroom <dbl>, kit <dbl>,
## #   kit_qual <chr>, central_air <chr>, elect <chr>, bsmt_area <dbl>,
## #   bsmt_cond <chr>, bsmt_exp <chr>, bsmt_fin_qual <chr>, bsmt_ht <chr>,
## #   ext_cond <chr>, ext_cover <chr>, ext_qual <chr>, fdn <chr>, ...
```