Lab 02

CS3172-1, Spring 2023, Effat University

Hanin Alzaher

Packages

```
library(tidyverse)
library(scales)
```

Data

```
cas <- read_rds("data/canada_survey.rds")
```

Tasks

Task 1

```
cas <- filter(cas, energy_expense>0 , household_income>0)
```

```
cas <- mutate(cas, marital_status = factor(marital_status))
```

Task 2

```
cas <- mutate(cas, heat_equip= case_when(heat_equip==1~"steam",
                                         heat_equip==2~"forced air",
                                         heat_equip==3~"stove",
                                         heat_equip==4~"electric heating"))
```

```
cas <- mutate(cas, heat_fuel= case_when(heat_fuel==1~"oil",
                                        heat_fuel==2~"gas",
                                        heat_fuel==3~"electricity",
                                        heat_fuel==4~"other"))
```

Task 3

```
cas %>%
  group_by(heat_fuel, heat_equip)%>%
  summarise( avg = mean(energy_expense),
             median= median(energy_expense),
             standerd_deviation= sd(energy_expense))
```

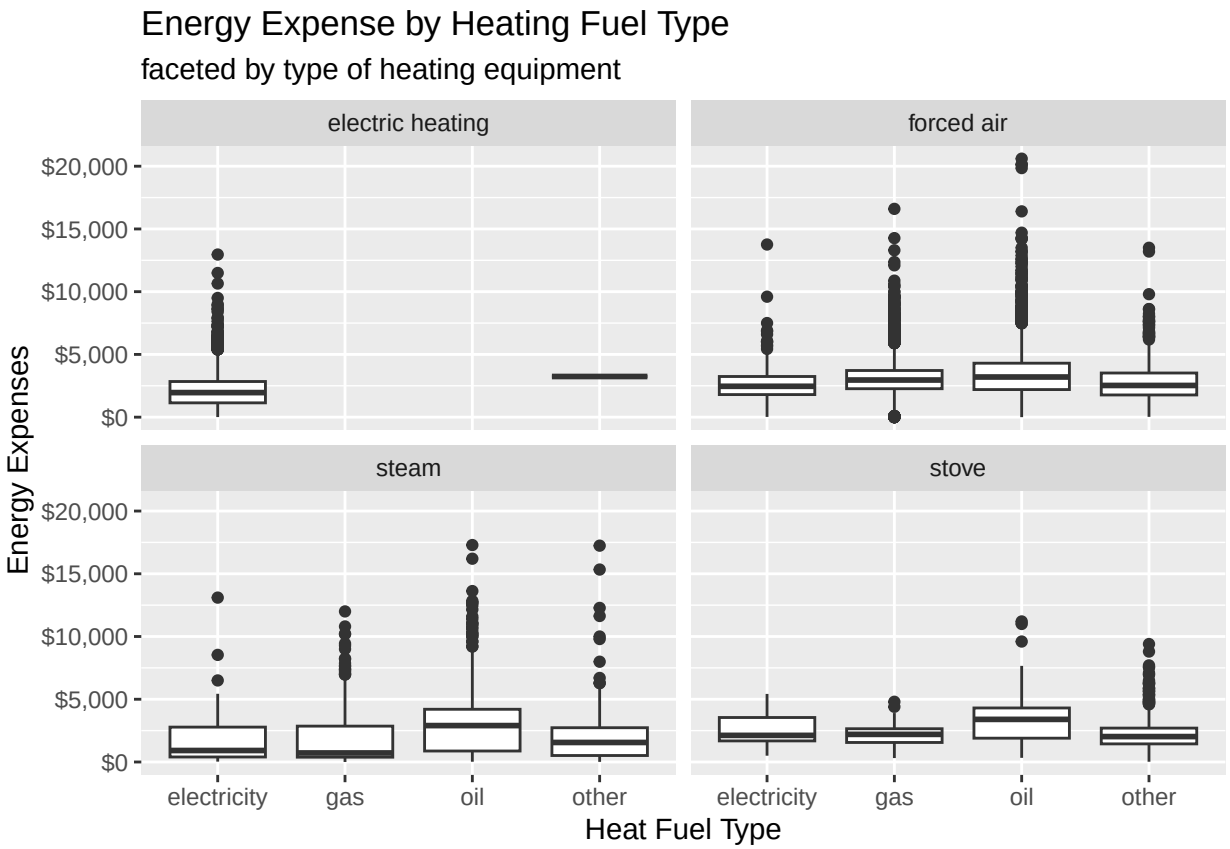
```
## `summarise()` has grouped output by 'heat_fuel'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 14 x 5
## # Groups:   heat_fuel [4]
##    heat_fuel   heat_equip         avg median standerd_deviation
##    <chr>       <chr>            <dbl>  <dbl>              <dbl>
##  1 electricity electric heating 2084.  1956               1270.
##  2 electricity forced air       2590.  2462.              1293.
##  3 electricity steam            1708.   915               1692.
##  4 electricity stove            2443.  2120               1229.
##  5 gas         forced air       3047.  2960               1395.
##  6 gas         steam            1698.   720               1820.
##  7 gas         stove            2178.  2202               1024.
##  8 oil         forced air       3499.  3200               2156.
##  9 oil         steam            2887.  2900               2142.
## 10 oil         stove            3396.  3395               2074.
## 11 other       electric heating 3240   3240                 NA
## 12 other       forced air       2861.  2526               1655.
## 13 other       steam            2047.  1555               2279.
## 14 other       stove            2210.  2025               1140.
```

- Provide the answer to the theoretical questions here
- What combination of fuel type and equipment has the highest average energy expense? oil and forced air has the the highest average energy expense with 3498.85
- Which combination has the most variability with regards to energy expense? other and steam with standerd deviation 2278.9
- Which type of heating equipment doesn't take all possible fuel types? electricity heating.

Task 4

```
ggplot(data = cas, mapping = aes(x = heat_fuel, y= energy_expense)) +
  geom_boxplot() + facet_wrap(~heat_equip, nrow = 2)+
scale_y_continuous(labels = scales :: dollar_format())+
labs(title = "Energy Expense by Heating Fuel Type",
subtitle = "faceted by type of heating equipment",
x="Heat Fuel Type",
y="Energy Expenses")
```



Task 5

```
cas <- mutate(cas, prop_energy =
(energy_expense / household_income))

cas2 <- cas %>%
  arrange(desc(prop_energy))%>%
  slice(1,n()) %>% glimpse()
```

```
## Rows: 2
## Columns: 25
## $ year                <fct> 2009, 2009
## $ province            <fct> Saskatchewan, Ontario
## $ dwelling_type       <fct> "Single detached", "Apartment"
## $ year_built          <fct> 1971-1980, 1971-1980
## $ rooms               <dbl> 7, 6
## $ beds                <dbl> 3, 2
## $ baths               <dbl> 1, 1
## $ heat_equip          <chr> "forced air", "forced air"
## $ heat_age            <fct> 2, 5
## $ heat_fuel           <chr> "gas", "gas"
## $ water_fuel          <fct> 2, 4
## $ cook_fuel           <fct> 2, 2
```

```
## $ income              <dbl> 100, 67000
## $ marital_status      <fct> 3, 3
## $ age                 <fct> 08, 14
## $ sex                 <fct> 2, 2
## $ education           <fct> 6, 1
## $ household_income    <dbl> 100, 67000
## $ energy_expense      <dbl> 3780, 1
## $ water_expense       <dbl> 540, 1
## $ electricity_expense <dbl> 1716, 0
## $ nat_gas_expense     <dbl> 1524, 0
## $ other_fuel_expense  <dbl> 0, 0
## $ consumption         <dbl> 19908, 16423
## $ prop_energy         <dbl> 3.780000e+01, 1.492537e-05
```

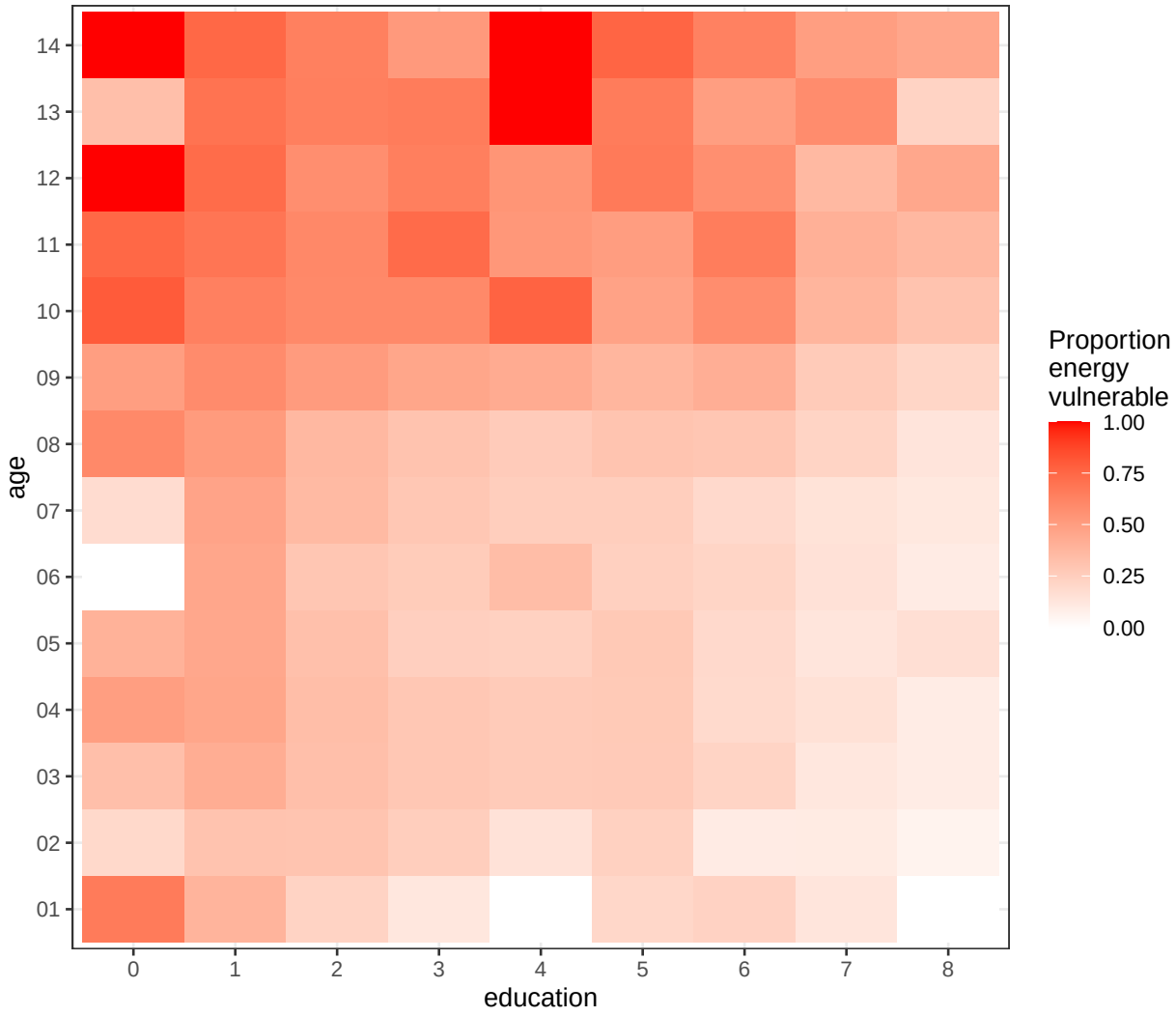
Task 6

```
cas%>%
  group_by(year,province )%>%
  summarise(median_energy_expense_per_room=median(energy_expense/rooms))%>%
  slice_min(median_energy_expense_per_room)
```

```
## # A tibble: 2 x 3
## # Groups:   year [2]
##   year  province median_energy_expense_per_room
##   <fct> <fct>                             <dbl>
## 1 2007  Quebec                             275
## 2 2009  Quebec                             269.
```

Task 7

```
cas %>% mutate( energy_vulnerable= if_else(prop_energy>0.05,"vulnerable","not"))%>%
  group_by(education,age)%>%
  summarise(prop_vul= mean(energy_vulnerable == "vulnerable"))%>%
  ungroup()%>%
  ggplot( aes(x=education,y=age, fill=prop_vul)) +
  geom_raster()  + scale_fill_gradient(low = "white",high = "red") +
  theme_bw() + labs(fill = "Proportion\nenergy\nvulnerable")
```



what we observe from the plot that the older people have vulnerable energy proportion.