Lab 02

CS3172-1, Spring 2023, Effat University

Aicha Sidiya

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) %>%
  summarize(mean_expense = mean(energy_expense),
            median_expense = median(energy_expense),
            stdv_expense = 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      mean_expense median_expense stdv_expense
##    <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.
```

- **What combination of fuel type and equipment has the highest average energy expense?**

Oil and Forced air have the highest mean which is 3498.850.

- **Which combination has the most variability with regards to energy expense?**

Other and Steam have the highest variability with a standard deviation 2278.911.

- **Which type of heating equipment doesn't take all possible fuel types?**

The electric heating equipment use only electricity as fuel.

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")
```


Energy Expense by Heating Fuel Type
faceted by type of heating equipment
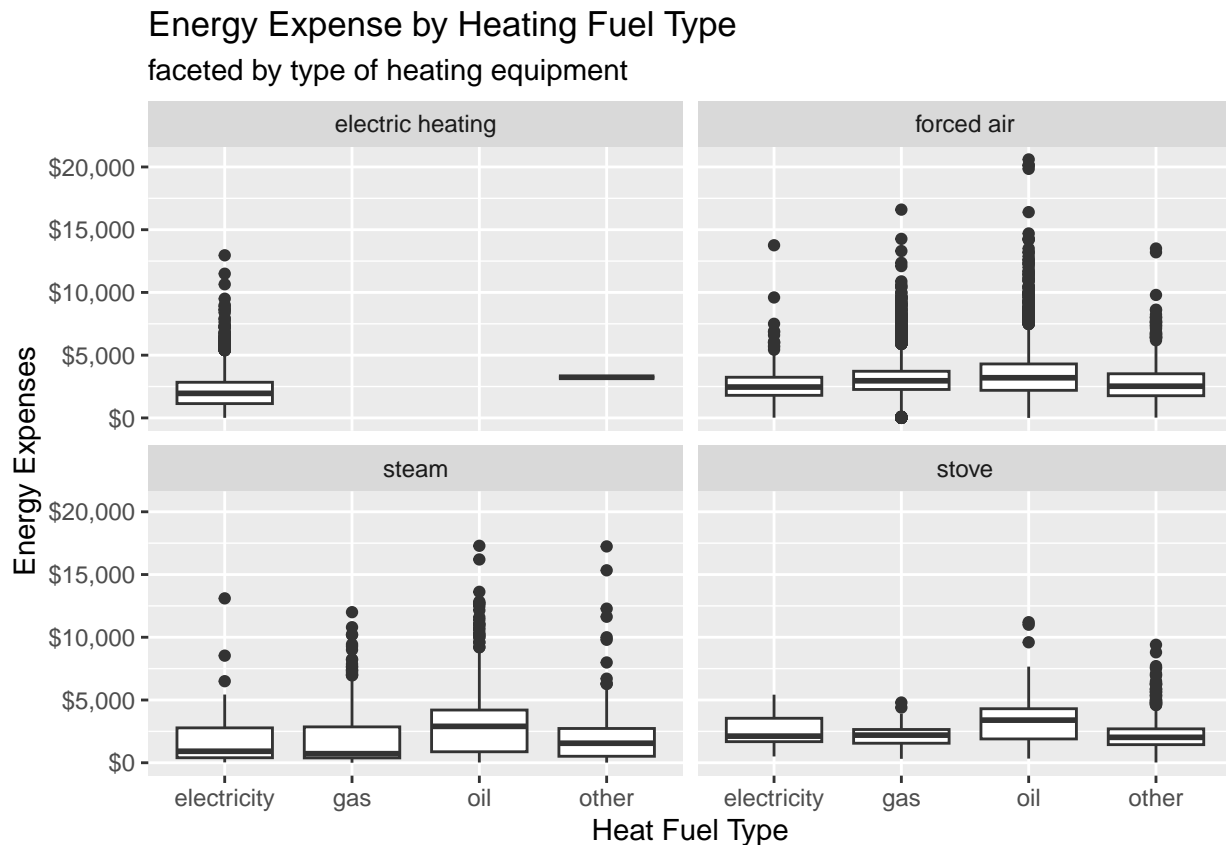
Task 5

```
cas <- mutate(cas, prop_energy_expense =
                (energy_expense / household_income))
cas2 <- cas %>%
  mutate(prop_energy_expense =
           (energy_expense / household_income)) %>%
  arrange(desc(prop_energy_expense)) %>%
  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_expense <dbl> 3.780000e+01, 1.492537e-05
```

```
#slice(1,n()) getting first and last column
```

The respondent with the lowest proportion of energy expense per household income which is 0.0000149 is a 85 year old divorced or widowed female that has a household income of 67000 $ and no degree, she lives in an apartment in Ontario that was built in 1971-1980, with 6 rooms, 2 beds and 1 bath she uses forced air heating equipment.

The respondent with the highest proportion of energy expense per household income 37.8 is a 55 year old divorced female that has a household income of 100 $ wih a bachelor of university diplome , she lives in a single detached in Saskatchewan that was built in 1971-1980, with 7 rooms, 3 beds and 1 bath she uses forced air heating equipment.
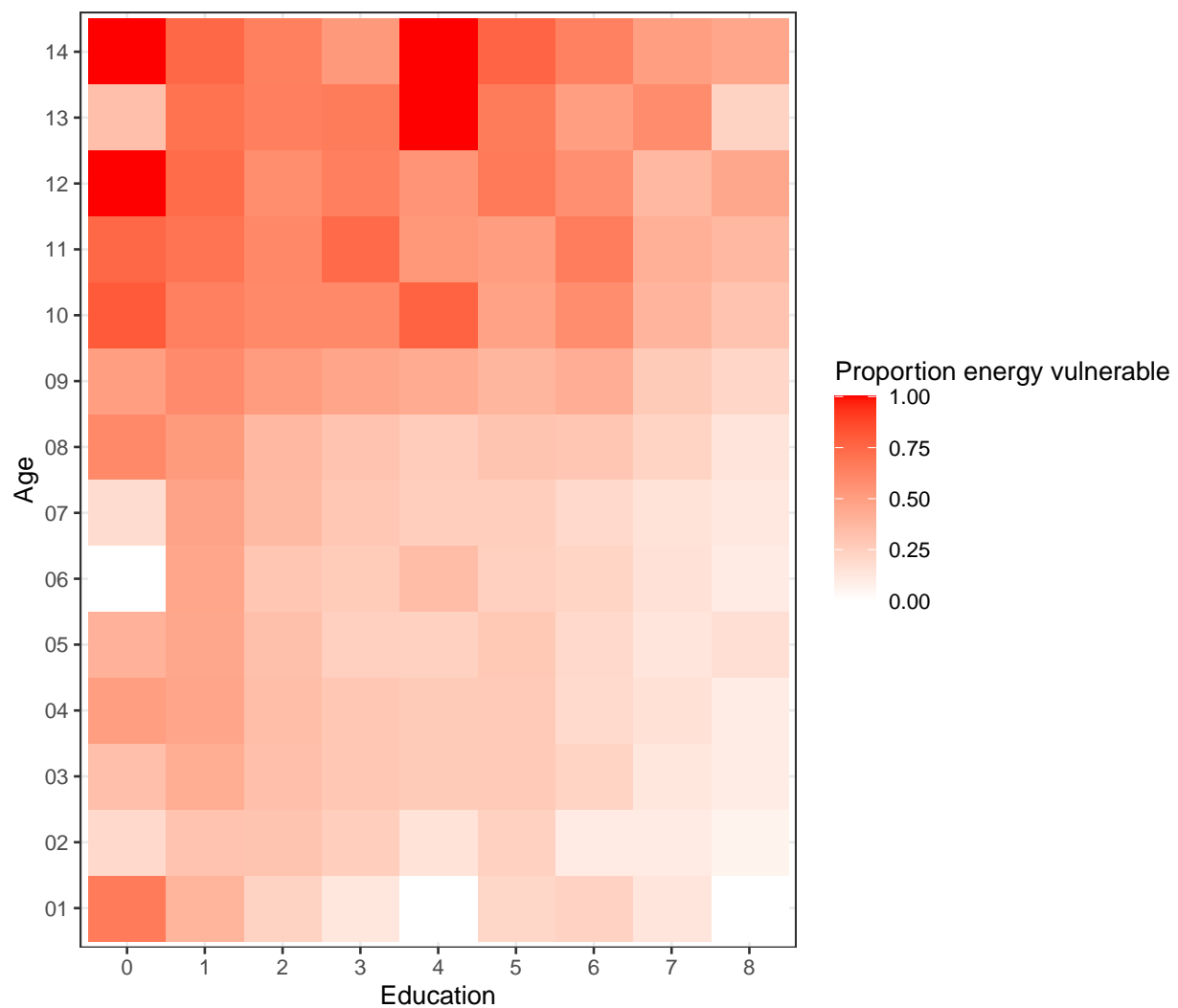
Task 6

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

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

Task 7

```
cas %>% mutate(energy_vulnerable =
                ifelse(prop_energy_expense > 0.05, "Yes", "No")) %>%
  group_by(education, age) %>%
  summarise(prop_vulnerable =
                mean(energy_vulnerable == "Yes")) %>%
  ungroup() %>%
ggplot(aes(x = education,
            y = age,
            fill = prop_vulnerable)) +
  geom_raster() +
  scale_fill_gradient( low = "white",
                        high = "red") +
  theme_bw() +
  labs(x = "Education",
        y = "Age",
        fill = "Proportion energy vulnerable")
```

Task 8