Lab 02

CS3172-1, Spring 2023, Effat University

Razan Almahdi

# 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))

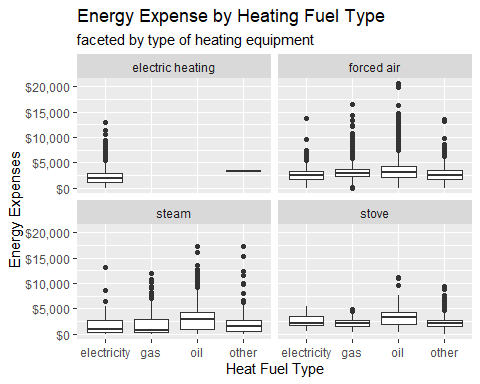
## `summarise()` has grouped output by 'heat\_fuel'. You can override using the  
## `.groups` argument.

## # A tibble: 14 × 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")



## 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

## 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 × 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\_prop = energy\_expense / household\_income,  
 vulnerable = if\_else(energy\_prop > 0.05, "vulnerable", "not")) %>%  
 group\_by(education, age) %>%  
 summarize(prop\_vulnerable = mean(vulnerable == "vulnerable")) %>%  
 ungroup() %>%  
 ggplot(aes( x= education, y = age, fill = prop\_vulnerable)) +  
 geom\_raster() +  
 scale\_fill\_gradient(low = "white", high = "red") +  
 labs( x = "Education", y= "Age", fill = "Proportion\nenergy\nvulnerable") +  
 theme\_bw()

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

