Lab 02

CS3172-1, Spring 2023, Effat University

Duaa Suroor

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

```
cas1<- cas %>% group_by(heat_equip,heat_fuel) %>% summarise(
  mean_ener_exp = mean(energy_expense),
  median_ener_exp = median(energy_expense),
  sd_ener_exp = sd(energy_expense))
```

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

- Provide the answer to the theoretical questions here

Task 4

```
ggplot(cas, mapping = aes(x=heat_fuel, y=energy_expense))+
  geom_boxplot()+
  scale_y_continuous(labels = scales::dollar_format())+
  facet_wrap(~ heat_equip, nrow = 2) +
  labs(title = "energy expense vs heating type",
       x="heat fuel type",
       y="enery expense")
```
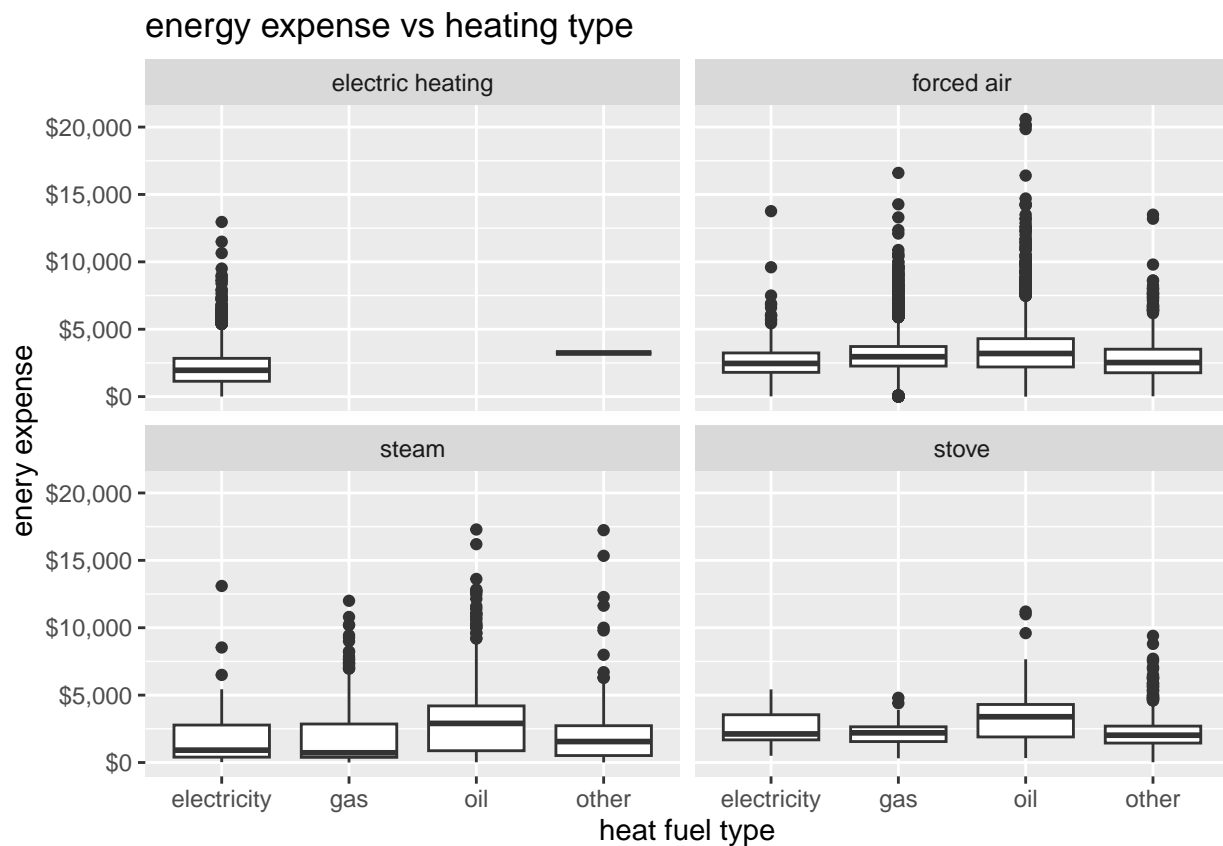
Task 5

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

Task 6

```
cas %>%
  group_by(year, province) %>%
  summarize(median_energy_expense_per_room = median(energy_expense/rooms))%>%
  group_by(year) %>%
  filter(median_energy_expense_per_room == min(median_energy_expense_per_room)) %>%
  select(year, province, median_energy_expense_per_room) %>%
  arrange(year)
```

```
## # 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_prop=energy_expense/household_income,vulnerable = if_else(energy_prop>0.05,"vulnerable"
  group_by(education, age)%>%
  summarise(prop_vulnerable = mean(vulnerable == "vulnerable"))%>%
  ungroup()%>%
  ggplot(aes(x = education, y = age, fill = prop_vulnerable))+geom_raster()+scale_fill_gradient(low="wh
```