

# Exam 01 solutions

STA 199, Spring 2021, Duke University

## Packages

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

## Data

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

## Tasks

### Task 1

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

```
cas <- cas %>%
  mutate(marital_status = factor(marital_status))
```

### Task 2

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

```
cas <- cas %>%
  mutate(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_equip, heat_fuel) %>%
```

```

summarise(mean_energy_exp = mean(energy_expense),
           median_energy_exp = median(energy_expense),
           sd_energy_exp = sd(energy_expense))

```

## `summarise()` regrouping output by 'heat\_equip' (override with `.groups` argument)

## # A tibble: 14 x 5

## # Groups: heat\_equip [4]

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

- Forced air w/ oil has highest mean energy expense
- Steam w/ other has the highest sd energy expense
- Electric heating only takes two fuels

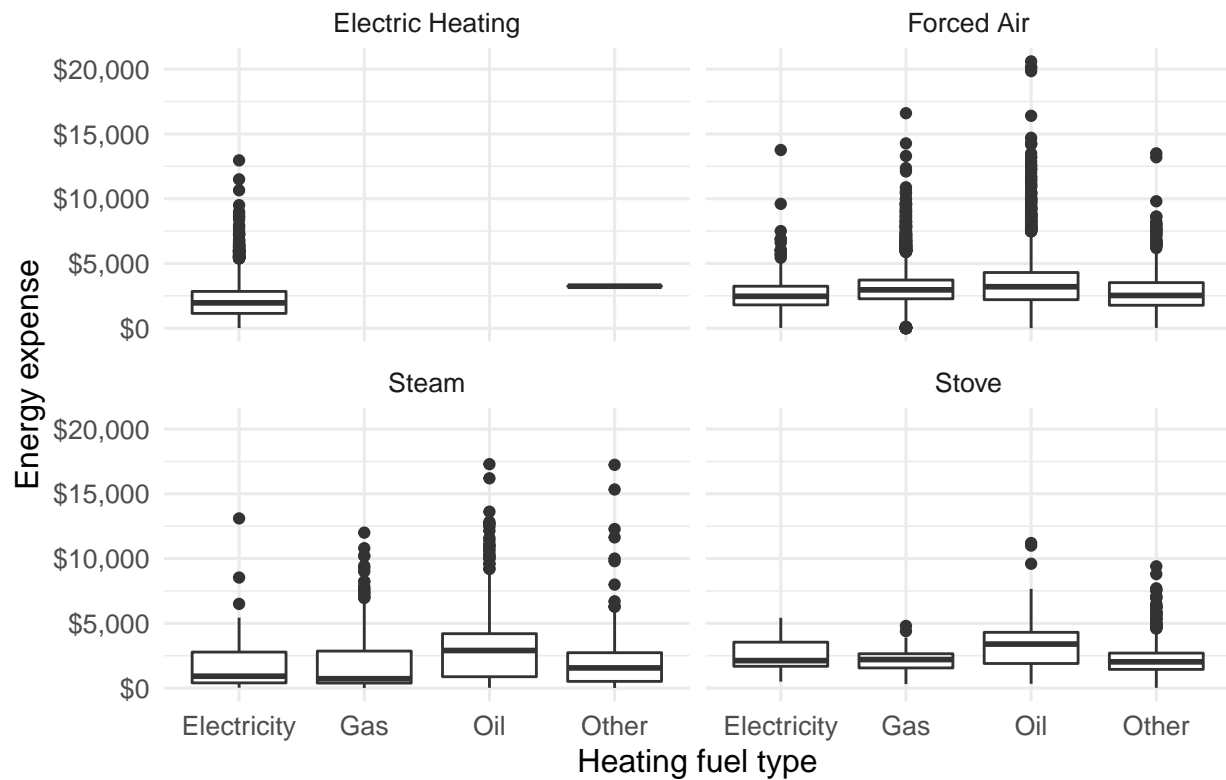
## Task 4

```

cas %>%
  mutate(heat_equip = str_to_title(heat_equip),
         heat_fuel = str_to_title(heat_fuel)) %>%
  ggplot(aes(x = heat_fuel, y = energy_expense)) +
  geom_boxplot() +
  scale_y_continuous(labels = label_dollar()) +
  facet_wrap(~ heat_equip, nrow = 2) +
  labs(x = "Heating fuel type", y = "Energy expense",
       title = "Facets by heating equipment type") +
  theme_minimal(base_size = 12)

```

## Facets by heating equipment type



## Task 5

```
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
```

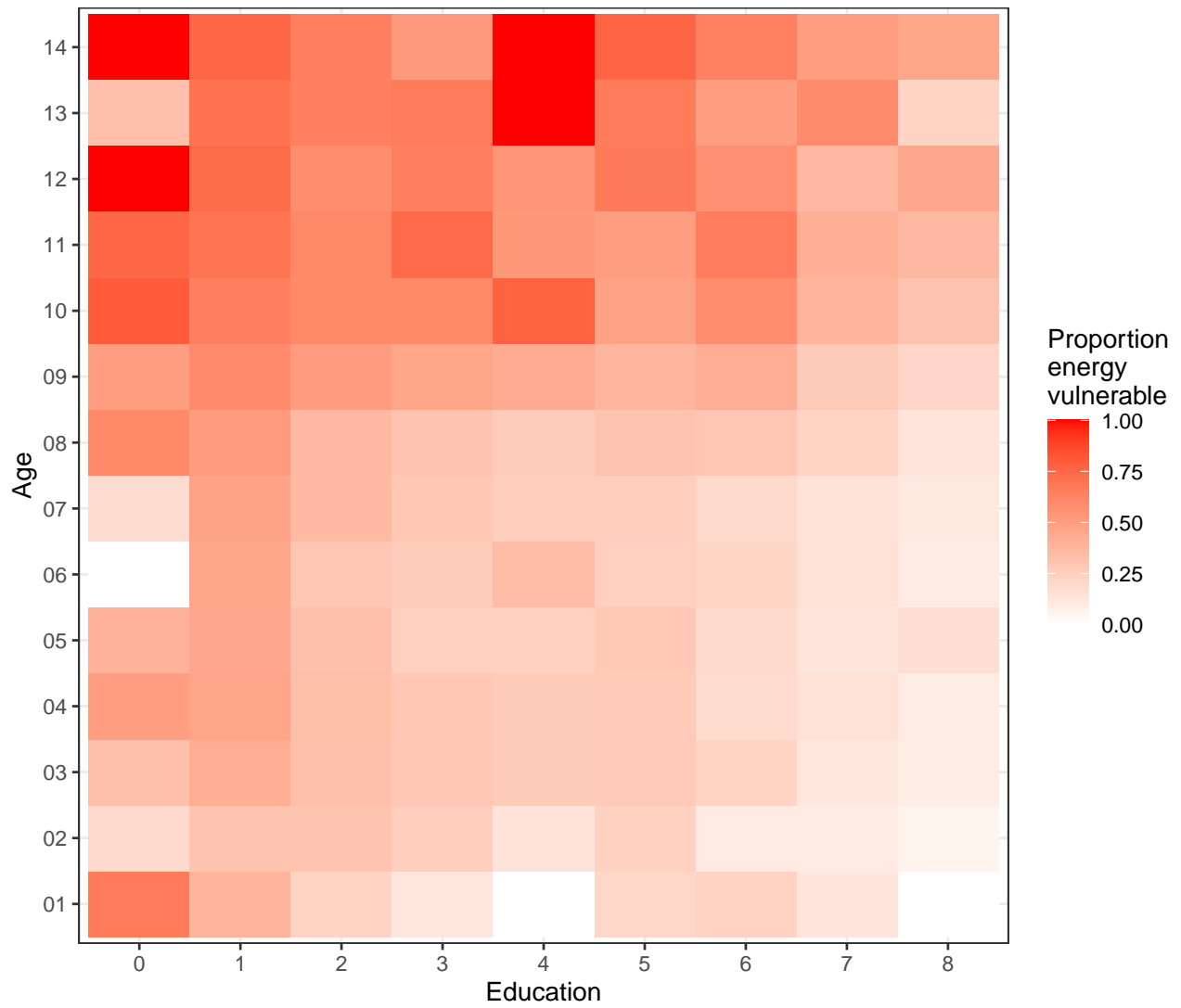
## Task 6

```
cas %>%
  mutate(eepr = energy_expense / rooms) %>%
  group_by(year, province) %>%
  summarise(median_energy_expense_per_room = median(eepr)) %>%
  arrange(median_energy_expense_per_room) %>%
  slice(1) %>%
  ungroup()
```

```
## # A tibble: 2 x 3
##   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", "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()
```



## Task 8

Answers will vary.