Public Opinion on Geothermal Energy: A Comparative Analysis

Veronika Slavíková

2025-08-22

Project Overview

This project provides a comparative analysis of public opinion on a hypothetical geothermal energy initiative. The data used in this case study was randomly generated and does not represent real-world events or public sentiment. The purpose of this analysis is to demonstrate key data processing and visualization skills, using sample data collected from two fictional public events: an "Open Day" and a "Public Hearing".

Key Skills Demonstrated

- **Data Processing and Cleaning:** Efficiently handling and preparing raw data from multiple sources.
- **Data Merging:** Combining datasets for a comprehensive comparative analysis.
- **Data Visualization:** Creating clear and informative plots to present survey results.
- **R Programming:** Utilizing `tidyverse` libraries for data manipulation and visualization.

Methodology and Data Limitations

The data for this analysis was generated for a sample project and does not represent a real-world population. The survey was conducted at project-specific events, which likely attracted individuals already interested in the topic. As such, the results may not be representative of the general population and should be considered a supplement to a more extensive sociological study.

To ensure data quality, values with `0`, `NA`, or `-` were removed, and white spaces were trimmed from the response columns. For questions that allowed multiple answers, the data was processed to reflect the total number of responses rather than the number of participants.

Part 1: Comparative Analysis of Shared Questions

This section compares public opinion from the Open Day and Public Hearing events using four identical questions. The data from both sources was combined to create comparative bar plots.

Data Loading and Preparation

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(dplyr)
library(tidyr)
library(ggplot2)
library(stringr)

Open_day <- read_delim("Data_Open_day.csv", delim = ";")
```

```
## Rows: 123 Columns: 6
## -- Column specification ------------------------------------------------------
## Delimiter: ";"
## chr (6): otazka_1, otazka_2, otazka_3, otazka_4, otazka_6, otazka_7
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
Hearing <- read_delim("Data_public_hearing.csv", delim = ";")
```

```
## New names:
## Rows: 159 Columns: 6
## -- Column specification
## ---------------------------------------------------- Delimiter: ";" chr
## (5): otazka_1, otazka_2, otazka_3, otazka_4, otazka_5 lgl (1): ...6
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...6`
```

Data processing:

```r
# Function to create and save bar plots
create_bar_plot <- function(data, question_code, title, x_label, y_label, recode_labels, source_levels)

  # Process and filter data
  plot_df <- data %>%
    pivot_longer(
      cols = starts_with("otazka_"),
      names_to = "question_code",
      values_to = "response_code"
    ) %>%
    mutate(response_code = str_trim(response_code)) %>%
    filter(question_code == !!question_code, !(response_code %in% c("-", "NA")), !is.na(response_code))
```

```
    mutate(
      response_label = recode(response_code, !!!recode_labels),
      source_label = recode(source, "Open Day" = "Open day", "Public Hearing" = "Public hearing")
    ) %>%
    mutate(
      response_label = factor(response_label, levels = recode_labels),
      source_label = factor(source_label, levels = source_levels)
    )
  # Create and print the plot
  plot <- ggplot(plot_df, aes(x = response_label, fill = source_label)) +
    geom_bar(position = position_dodge(width = 0.9), width = 0.8) +
    labs(title = title, x = x_label, y = y_label, fill = "Data source") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))

  print(plot)
}
```

Merging data for questions 1–4:

```
# Combine and process data for questions 1-4
combined_df_1_4 <- bind_rows(
  Open_day %>% select(otazka_1, otazka_2, otazka_3, otazka_4) %>% mutate(source = "Open Day"),
  Hearing %>% select(otazka_1, otazka_2, otazka_3, otazka_4) %>% mutate(source = "Public Hearing")
)
```
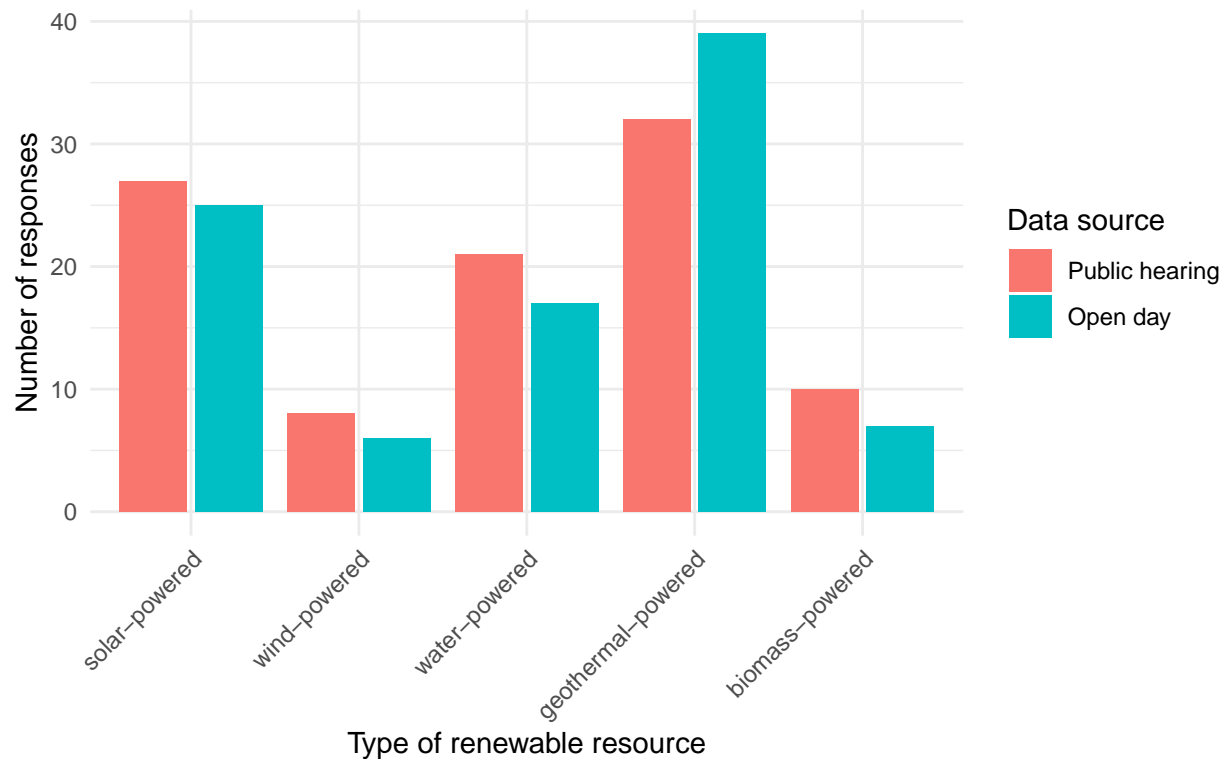
Generating a graph for question no. 1:

```
# Call the function for each bar plot
create_bar_plot(
  data = combined_df_1_4,
  question_code = "otazka_1",
  title = "Which of these renewable sources do you consider to be the most
  promising in our climate conditions?",
  x_label = "Type of renewable resource",
  y_label = "Number of responses",
  recode_labels = c("a" = "solar-powered", "b" = "wind-powered", "c" = "water-powered", "d" = "geotherma
  source_levels = c("Public hearing", "Open day")
)
```

Which of these renewable sources do you consider to be the most promising in our climate conditions?
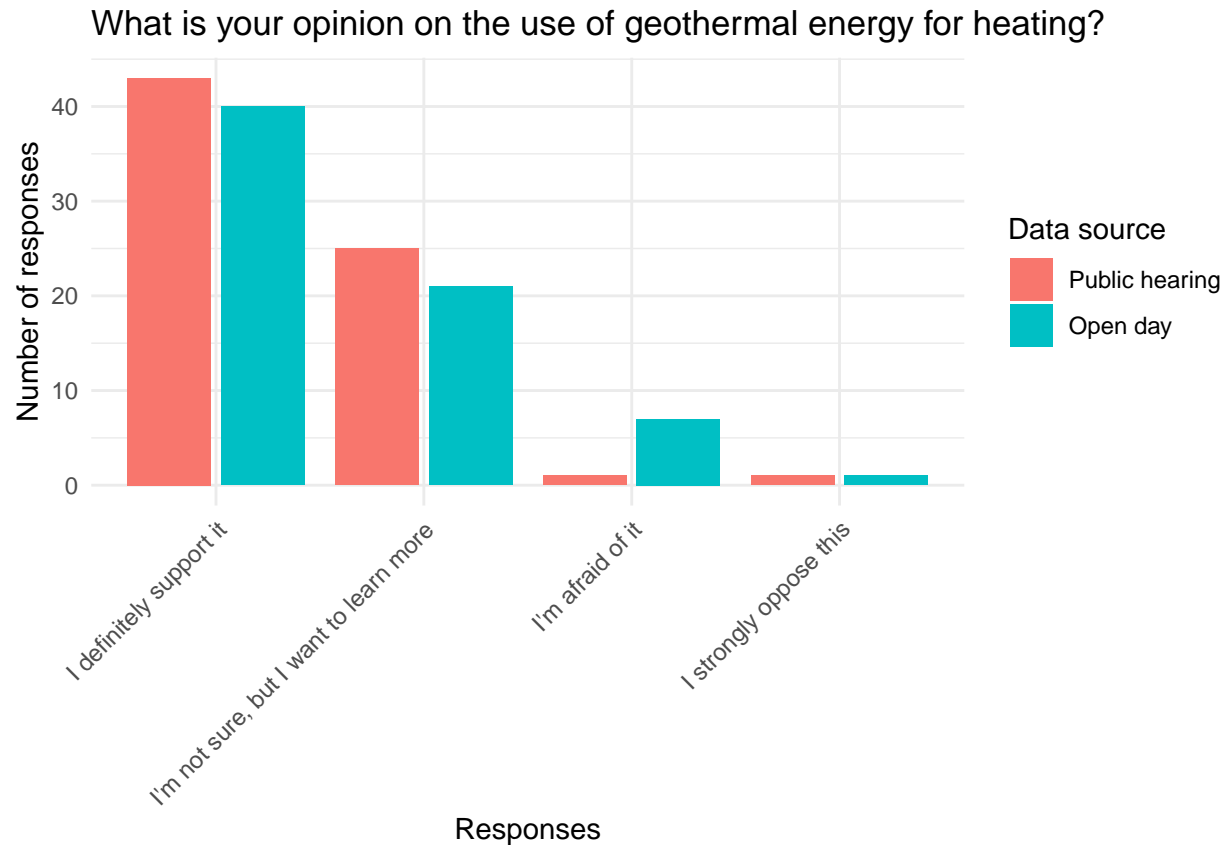

Generating a graph for question no. 2:

```
create_bar_plot(
  data = combined_df_1_4,
  question_code = "otazka_2",
  title = "Do you consider geothermal energy to be a promising source?",
  x_label = "Responses",
  y_label = "Number of responses",
  recode_labels = c("a" = "yes", "b" = "rather yes", "c" = "rather no", "d" = "no", "e" = "I don´t know
  source_levels = c("Public hearing", "Open day")
)
```

Do you consider geothermal energy to be a promising source?
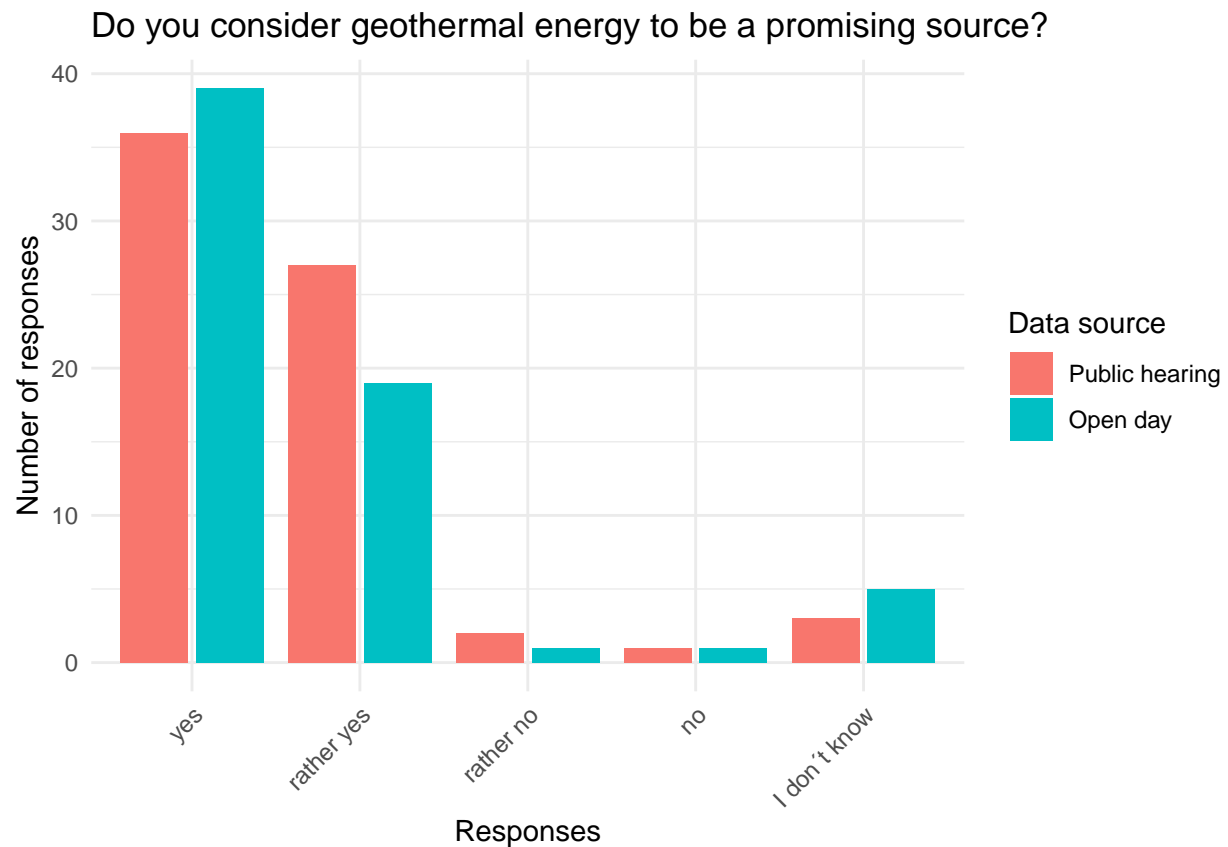
Generating a graph for question no. 3:

```
create_bar_plot(
  data = combined_df_1_4,
  question_code = "otazka_3",
  title = "What is your opinion on the use of geothermal energy for heating?",
  x_label = "Responses",
  y_label = "Number of responses",
  recode_labels = c("a" = "I definitely support it", "b" = "I'm not sure, but I want to learn more", "c"
  source_levels = c("Public hearing", "Open day")
)
```

What is your opinion on the use of geothermal energy for heating?

Generating a graph for question no. 4:

```
create_bar_plot(
  data = combined_df_1_4,
  question_code = "otazka_4",
  title = "Do you know what the approximate temperature of the Earth's crust is
  at a depth of 1 km?",
  x_label = "Responses",
  y_label = "Number of responses",
  recode_labels = c("a" = "3,2 °C", "b" = "32 °C", "c" = "320 °C"),
  source_levels = c("Public hearing", "Open day")
)
```

Do you know what the approximate temperature of the Earth's crust is at a depth of 1 km?
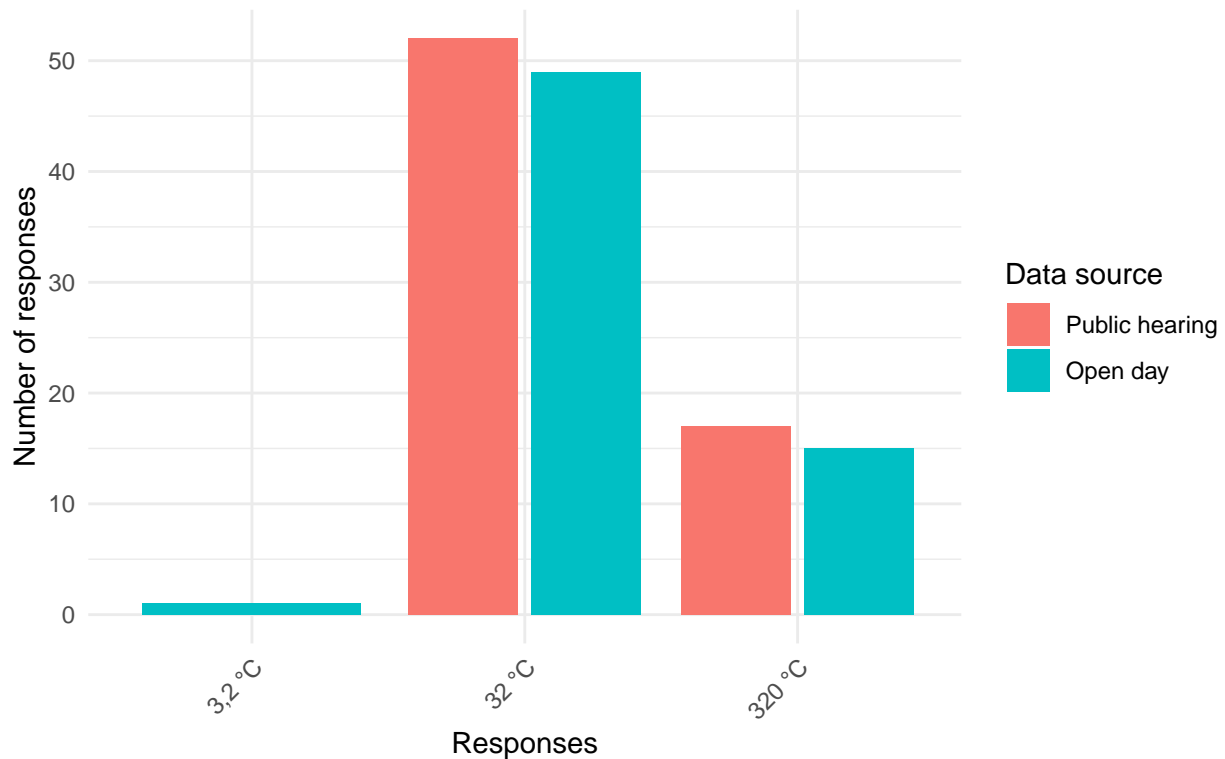
Preparation of data for creating pie charts for questions 5–7:

```r
# Function to create and save pie charts
create_pie_chart <- function(data, question_code, title, recode_labels, source_label) {

  # Process and filter data
  plot_df <- data %>%
    select(!!question_code) %>%
    mutate(source = source_label) %>%
    pivot_longer(
      cols = starts_with("otazka_"),
      names_to = "question_code",
      values_to = "response_code"
    ) %>%
    mutate(response_code = str_trim(response_code)) %>%
    filter(!(response_code %in% c("-", "NA")), !is.na(response_code)) %>%
    mutate(response_label = recode(response_code, !!!recode_labels)) %>%
    mutate(source_label = recode(source, "Open Day" = "", "Public Hearing" = ""))

  # Reorder levels
  response_levels <- plot_df %>% distinct(response_label) %>% pull(response_label) %>% setdiff("other")
  plot_df <- plot_df %>% mutate(response_label = factor(response_label, levels = response_levels))

  # Summarize data
```

```
  plot_summary <- plot_df %>%
    group_by(source_label, response_label) %>%
    summarise(count = n(), .groups = "drop") %>%
    group_by(source_label) %>%
    mutate(percent = count / sum(count), label = paste0(round(percent * 100), "%"))
  # Create and print the plot
  plot <- ggplot(plot_summary, aes(x = "", y = count, fill = response_label)) +
    geom_bar(stat = "identity", width = 1) +
    coord_polar("y") +
    facet_wrap(~ source_label) +
    geom_text(
      aes(label = label),
      position = position_stack(vjust = 0.5),
      color = "white",
      size = 3
    ) +
    labs(title = title, fill = "Responses") +
    theme_minimal() +
    theme(
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      axis.text = element_blank(),
      axis.ticks = element_blank(),
      panel.grid = element_blank()
    )
  print(plot)
}
```
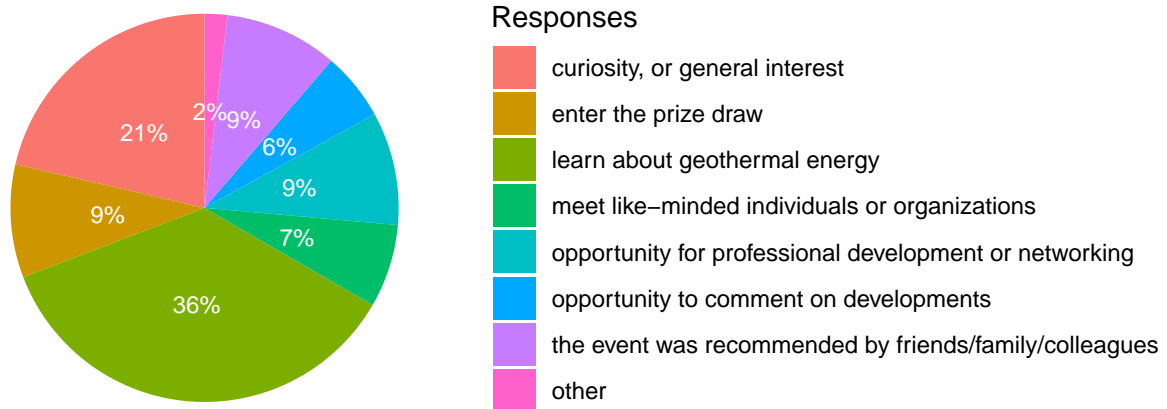
Generating a graph for question no. 5:

```
# Call the function for each pie chart
create_pie_chart(
  data = Hearing,
  question_code = "otazka_5",
  title = "What are the reasons for your participation at the public hearing?",
  recode_labels = c("a" = "learn about geothermal energy", "b" = "opportunity to comment on developments
  source_label = "Public Hearing"
)
```

What are the reasons for your participation at the public hearing?

Responses

- curiosity, or general interest
- enter the prize draw
- learn about geothermal energy
- meet like−minded individuals or organizations
- opportunity for professional development or networking
- opportunity to comment on developments
- the event was recommended by friends/family/colleagues
- other

Generating a graph for question no. 6:

```
create_pie_chart(
  data = Open_day,
  question_code = "otazka_6",
  title = "I found out about the open day from this source:",
  recode_labels = c("a" = "poster", "b" = "video", "c" = "personal invitation", "d" = "local newspaper"
  source_label = "Open Day"
)
```

I found out about the open day from this source:
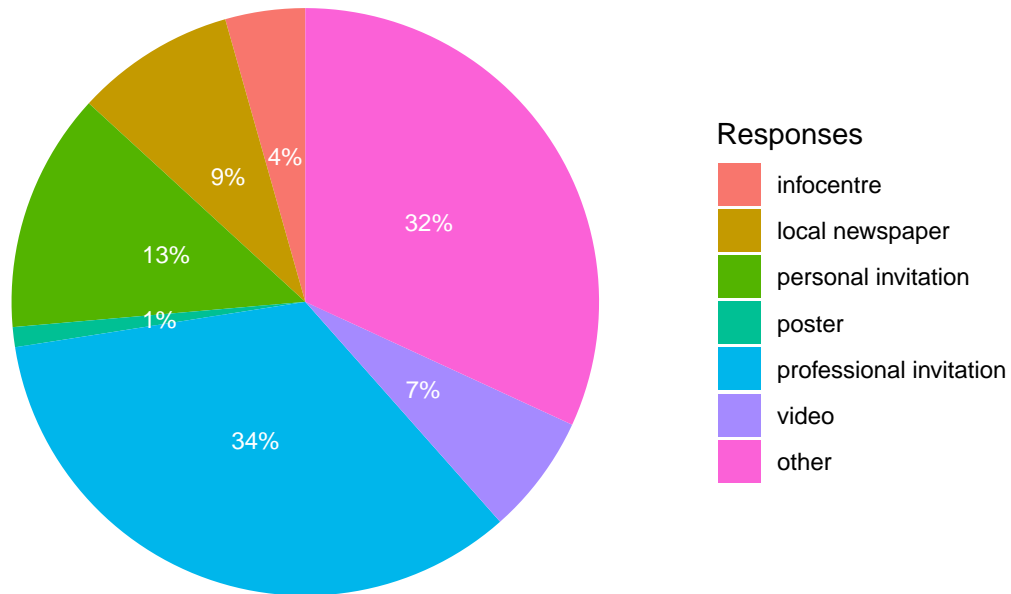
Generating a graph for question no. 7:

```
create_pie_chart(
  data = Open_day,
  question_code = "otazka_7",
  title = "My motivation to attend the open day:",
  recode_labels = c("a" = "I support the use of geothermal energy", "b" = "I am concerned about the use
  source_label = "Open Day"
)
```

My motivation to attend the open day:

Responses
- general interest
- I'm here out of curiosity
- I support the use of geothermal energy
- I want to see the geothermal laboratory and drill cores
- other

19% 5% 3% 41% 33%