Comparative Analysis of Coffee Habits Between Remote and In-Person Workers

Table of contents

Executive Summary

Introduction

Methodology

The survey categorizes respondents into three groups: those who work from home, those who work in an office, and those with a hybrid work arrangement. Using a questionnaire format, the survey focuses on participants' coffee consumption habits.

```r
library(ggplot2)
library(readr)
library(tidyverse)
library(patchwork)
library(knitr)
```

```r
coffee_survey <- read_csv(here::here("data/coffee_survey.csv"))

coffee_survey_selected <- coffee_survey %>%
  select("submission_id", "wfh", "cups", "total_spend", "most_paid") %>%
```

```
  mutate(wfh = recode(wfh, "I do a mix of both" = "both",
                      "I primarily work from home" = "remote",
                      "I primarily work in person" = "in_person"))
```

```
work_mode_summary <- coffee_survey_selected %>%
  filter(!is.na(wfh)) %>%
  count(wfh) %>%
  mutate(work_mode = wfh,
         proportion = scales::percent(n/sum(n))) %>%
  select(work_mode, n, proportion)

knitr::kable(work_mode_summary)
```

Table 1: Proportion of Participants by Work Mode

work_mode	n	proportion
both	883	25.1%
in_person	1049	29.8%
remote	1592	45.2%

Table 1 is used to show the number of people in each group, and calculate the percentage of them within the total sample. So that we can compare the proportion of different groups.

```
cup_sum <- coffee_survey_selected %>%
  filter(!is.na(cups), !is.na(wfh)) %>%
  count(cups, wfh) %>%
  mutate(cups = factor(
    cups,
    levels = c("Less than 1", "1", "2", "3", "4", "More than 4"))
    )

p1 <- cup_sum %>%
  ggplot(aes(x = cups, y = n, fill = wfh)) +
  geom_col(position = "dodge", width = 0.6) +
  labs(
    title = "Cups of coffee per day",
    x = "Cups",
    y = "Number of people",
    fill = "Work Mode") +
  theme_minimal()
```

```r
total_spend_sum <- coffee_survey_selected %>%
  filter(!is.na(total_spend), !is.na(wfh)) %>%
  count(total_spend, wfh) %>%
  mutate(total_spend = factor(
    total_spend, levels = c(
      "<$20", "$20-$40", "$40-$60", "$60-$80", "$80-$100", ">$100"))
    )

p2 <- total_spend_sum %>%
  ggplot(aes(x = total_spend, y = n, fill = wfh)) +
  geom_col(position = "dodge", width = 0.6) +
  labs(
    title = "Money spent on coffee per month",
    x = "Spending Range", y = "Number of People", fill = "Work Mode"
    ) +
  theme_minimal()

most_paid_sum <- coffee_survey_selected %>%
  filter(!is.na(most_paid), !is.na(wfh)) %>%
  count(most_paid, wfh) %>%
  mutate(most_paid = factor(most_paid, levels = c(
    "Less than $2", "$2-$4", "$4-$6", "$6-$8", "$8-$10",
    "$10-$15", "$15-$20", "More than $20"))
    )

p3 <- most_paid_sum %>%
  ggplot(aes(x = most_paid, y = n, fill = wfh)) +
  geom_col(position = "dodge", width = 0.6) +
  labs(title = "Maximum Price Paid for a Cup of Coffee",
       x = "Spending Range", y = "Number of People", fill = "Work Mode"
       ) +
  theme_minimal()
```

```r
p1 <- p1 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
p2 <- p2 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
p3 <- p3 + theme(axis.text.x = element_text(angle = 45, hjust = 1))

combined <- (p1 | p2) / (p3 | patchwork::plot_spacer())

ggsave("combined_plot.png", combined, width = 14, height = 10)
```
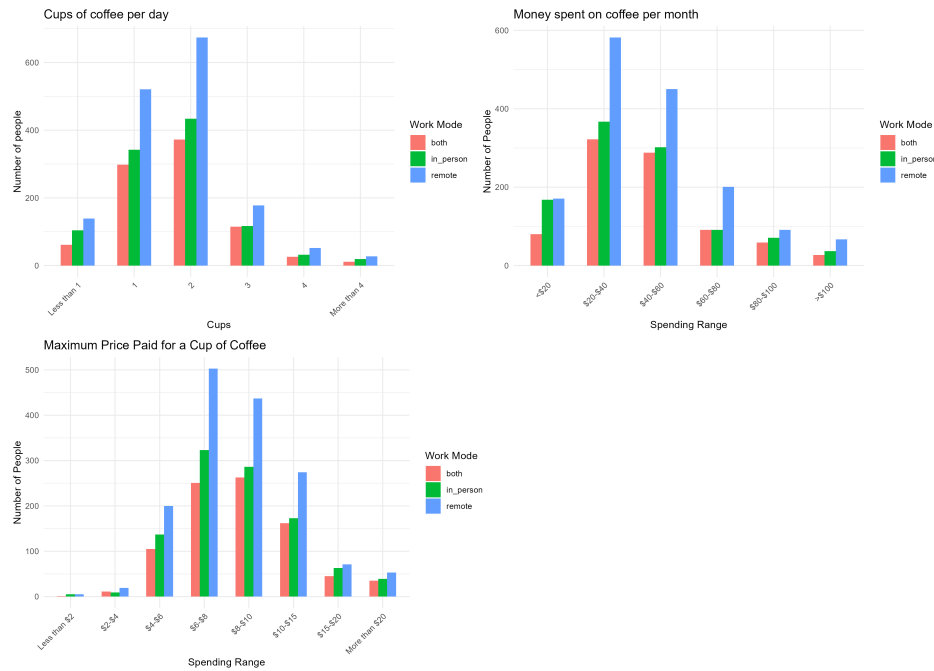
```
knitr::include_graphics("combined_plot.png")
```


Figure 1: Research results

According to Figure 1 , we can make comparisons across the three groups, showing their daily coffee consumption, monthly coffee spending, and the maximum price paid for a cup of coffee. This allows us to observe respondents' preferences and behavior regarding coffee consumption. Building on these insights, we can further apply cluster analysis to categorize consumers into different types, which helps us draw more conclusions.

Results

Conclusion