Question 2 Output

Adam Whiteside

2025-01-10

```
#Question 2:
#reload data # replace with necessary file path
load("C:/Users/ADAMI/Downloads/upmc code/CodeChallenge2024.RData")

# Load necessary libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

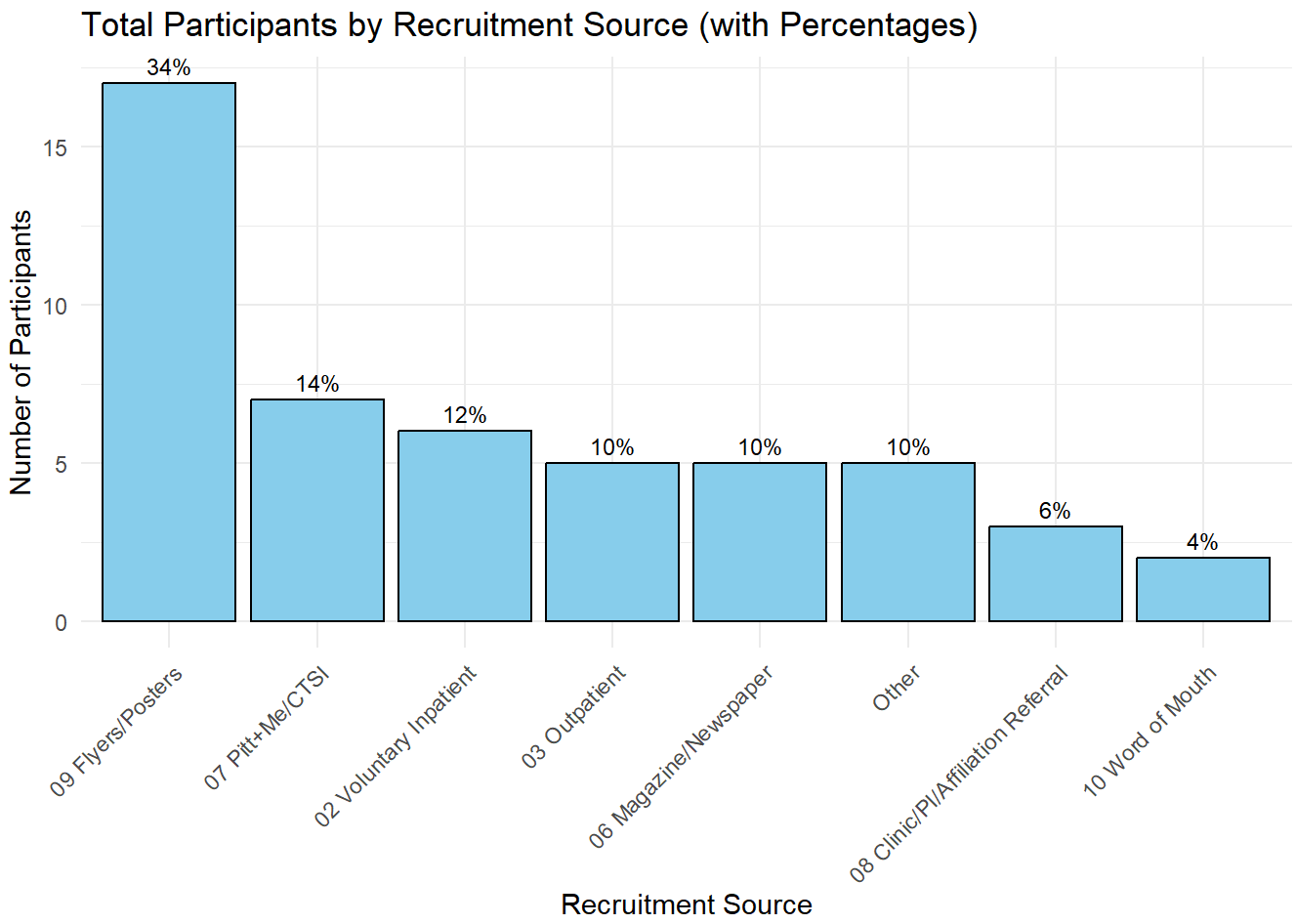
```
library(scales)  # For percentage formatting
library(patchwork)  # For combining plots

# Calculate Summary Table
summary_table <- recruitment_data %>%
  group_by(RecruitSource, Gender, Group) %>%
  summarise(
    Total_Participants = n(),
    Percentage = n() / nrow(recruitment_data) * 100
  ) %>%
  arrange(desc(Total_Participants))
```

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

```
summary_table
```

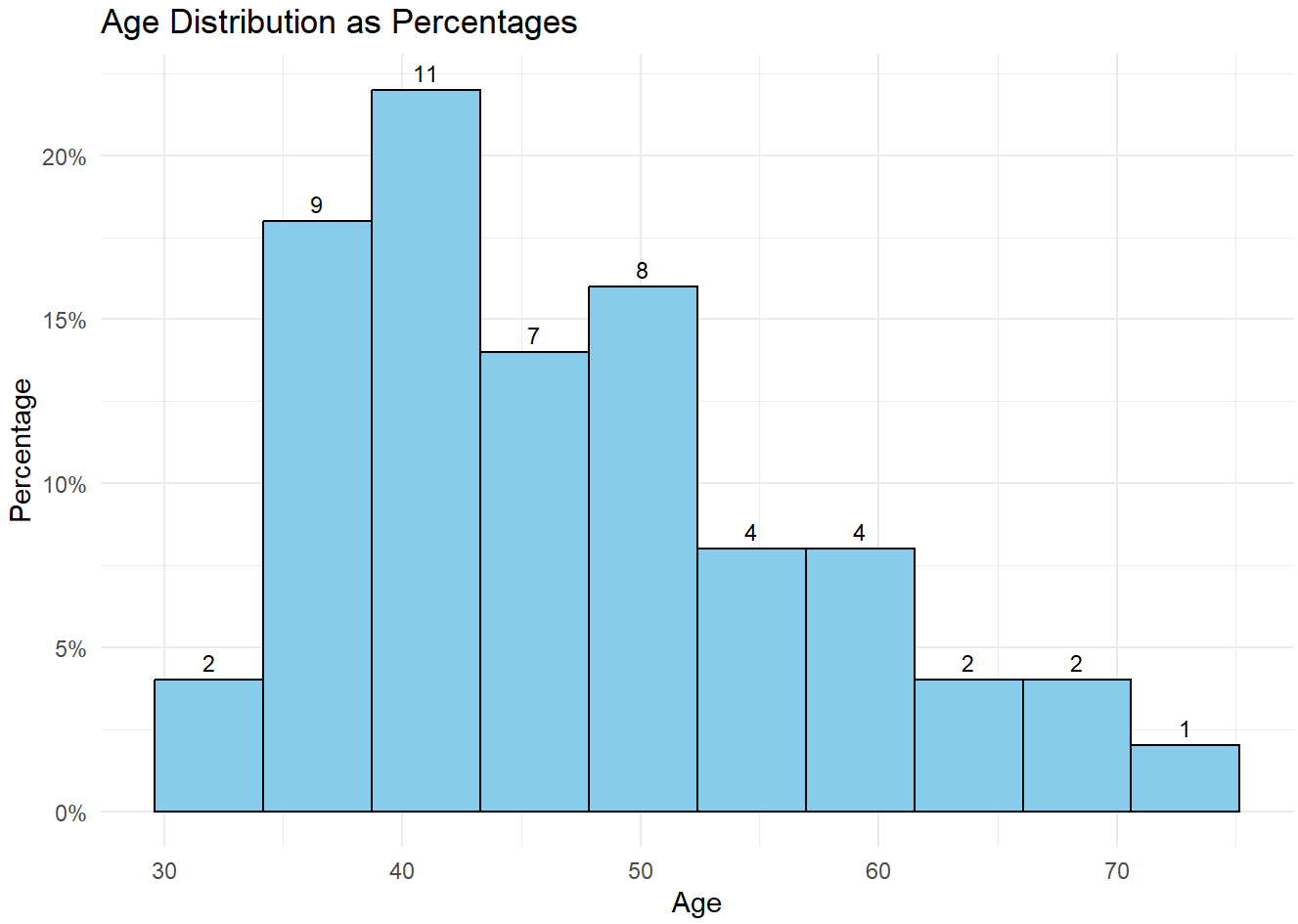
```
## # A tibble: 25 × 5
## # Groups:   RecruitSource, Gender [13]
##    RecruitSource          Gender Group Total_Participants Percentage
##    <chr>                  <chr>  <chr>              <int>      <dbl>
##  1 09 Flyers/Posters      F      3ATT                  10         20
##  2 02 Voluntary Inpatient F      3ATT                   4          8
##  3 03 Outpatient          F      3ATT                   4          8
##  4 09 Flyers/Posters      F      NON                    4          8
##  5 07 Pitt+Me/CTSI        F      NON                    3          6
##  6 06 Magazine/Newspaper  F      3ATT                   2          4
##  7 06 Magazine/Newspaper  M      NON                    2          4
##  8 09 Flyers/Posters      M      3ATT                   2          4
##  9 10 Word of Mouth       F      3ATT                   2          4
## 10 Other                  F      3ATT                   2          4
## # ℹ 15 more rows
```

```
# Plot 1: Total participants by Recruitment Source with percentages
source_plot <- recruitment_data %>%
  count(RecruitSource) %>%
  ggplot(aes(x = reorder(RecruitSource, -n), y = n, label = scales::percent(n / sum(n)))) +
  geom_bar(stat = "identity", fill = "skyblue", color = "black") +
  geom_text(vjust = -0.5, size = 3) +
  labs(
    title = "Total Participants by Recruitment Source (with Percentages)",
    x = "Recruitment Source",
    y = "Number of Participants"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(source_plot)
```

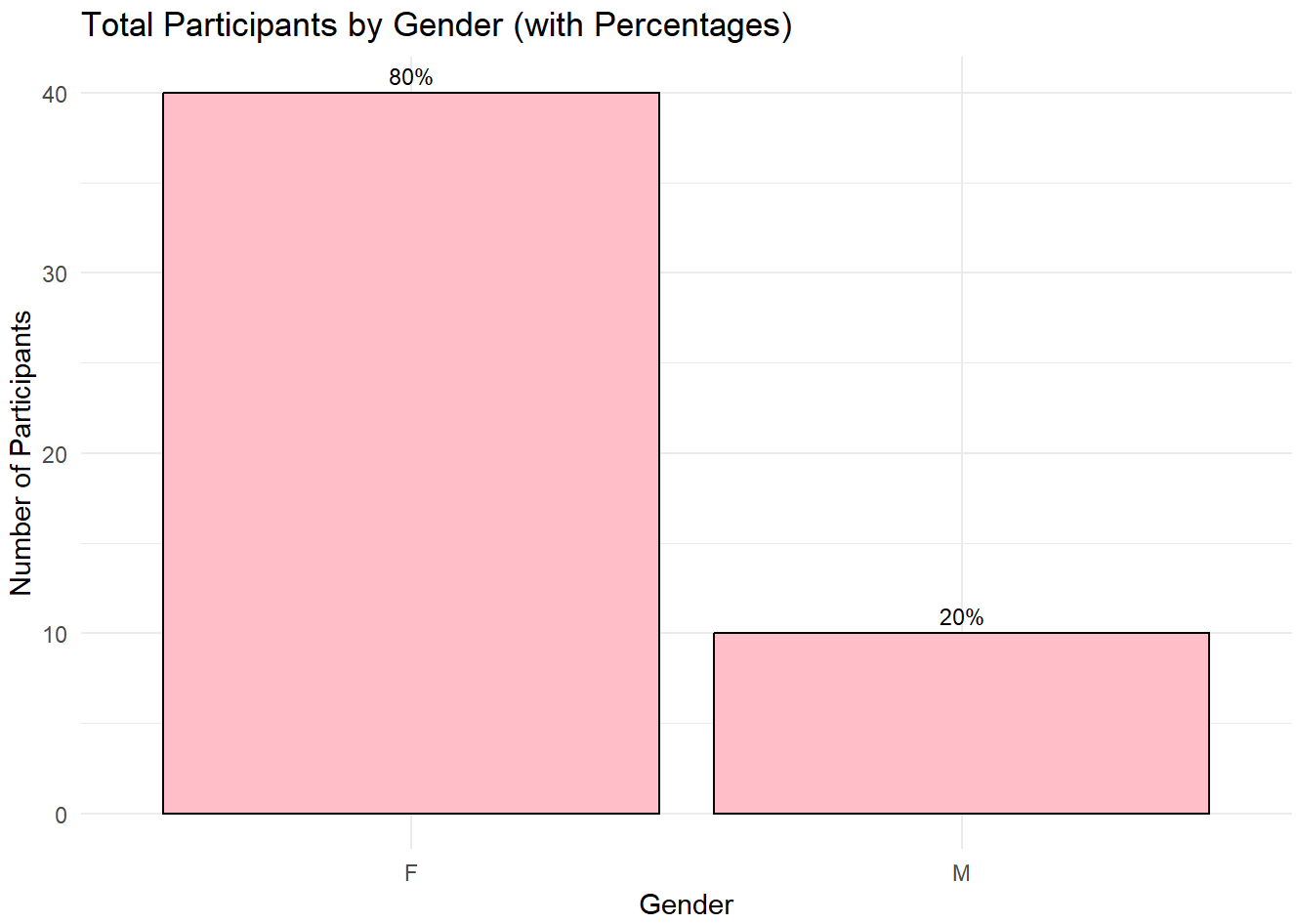



```
# Plot 2: Age Distribution as Percentages with Counts as Labels
age_percentage_plot <- ggplot(recruitment_data, aes(x = Age)) +
  geom_histogram(aes(y = (..count..) / sum(..count..)),
                 bins = 10, fill = "skyblue", color = "black") +
  scale_y_continuous(labels = scales::percent_format()) +
  labs(
    title = "Age Distribution as Percentages",
    x = "Age",
    y = "Percentage"
  ) +
  theme_minimal() +
  geom_text(
    aes(y = (..count..) / sum(..count..), label = ..count..),
    stat = "bin",
    bins = 10,
    vjust = -0.5,
    size = 3
  )

print(age_percentage_plot)
```

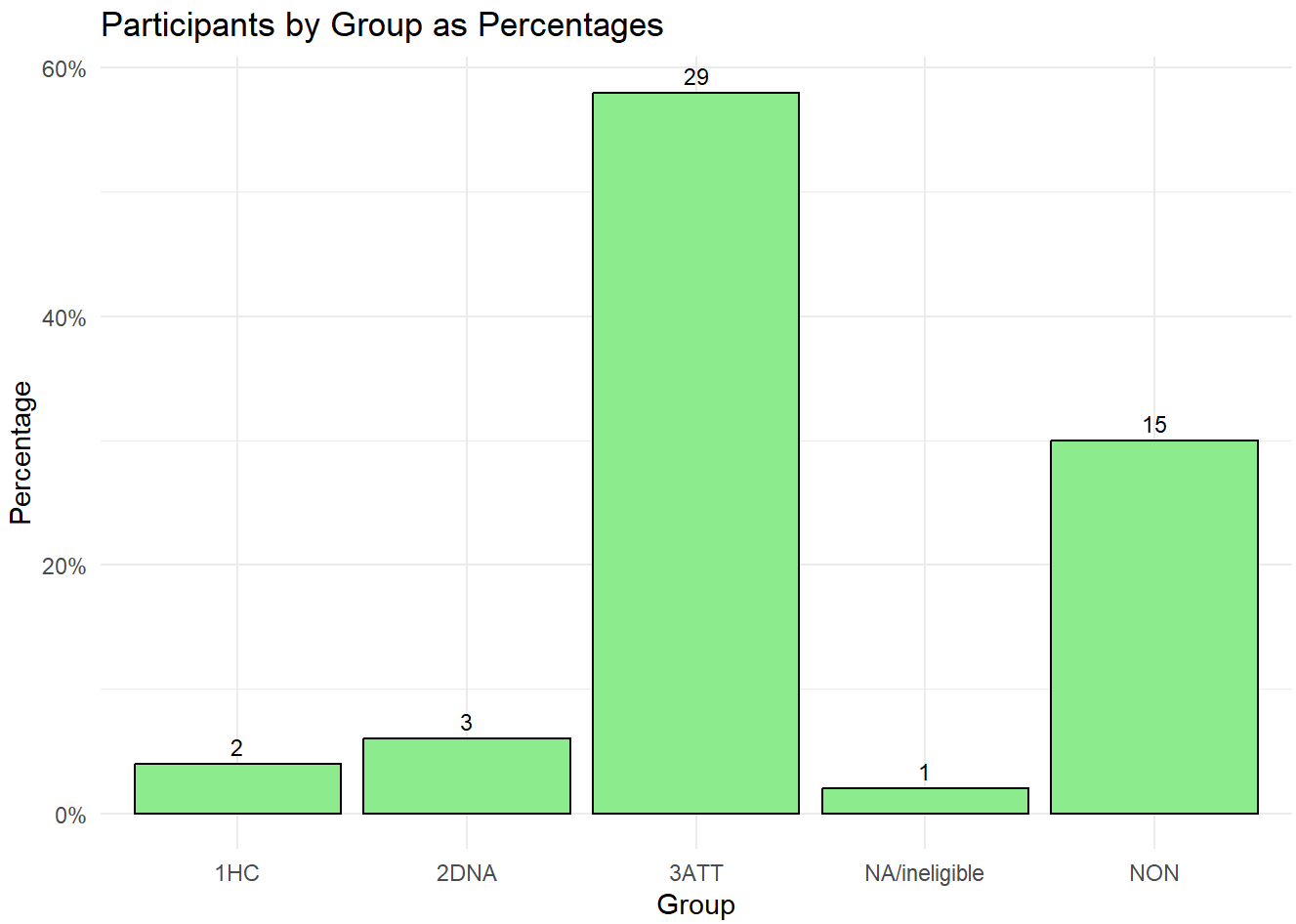
```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## ℹ Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
# Plot 3: Total participants by Gender (with percentages)
gender_plot <- recruitment_data %>%
  count(Gender) %>%
  ggplot(aes(x = Gender, y = n, label = scales::percent(n / sum(n)))) +
  geom_bar(stat = "identity", fill = "pink", color = "black") +
  geom_text(vjust = -0.5, size = 3) +
  labs(
    title = "Total Participants by Gender (with Percentages)",
    x = "Gender",
    y = "Number of Participants"
  ) +
  theme_minimal()

print(gender_plot)
```




```
# Plot 4: Participants by Group as Percentages with Counts as Labels
group_percentage_plot <- ggplot(recruitment_data, aes(x = Group)) +
  geom_bar(aes(y = (..count..) / sum(..count..)),
           fill = "lightgreen", color = "black") +
  scale_y_continuous(labels = scales::percent_format()) +
  labs(
    title = "Participants by Group as Percentages",
    x = "Group",
    y = "Percentage"
  ) +
  theme_minimal() +
  geom_text(
    aes(y = (..count..) / sum(..count..), label = ..count..),
    stat = "count",
    vjust = -0.5,
    size = 3
  )

print(group_percentage_plot)
```


Participants by Group as Percentages
60%
40%
20%
0%
Percentage
29
15
2
3
1
1HC
2DNA
3ATT
NA/ineligible
NON
Group