Data211FinalProject

Wendy Mendoza Gutierrez

2024-12-13

```r
library(dplyr)
```

```
##
## 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
```

```r
claim1 <- data.frame(
  Date = c("11/14", "11/15", "11/16", "11/17",
           "11/18", "11/19", "11/20", "11/21",
           "11/22", "11/23", "11/24", "11/25",
           "11/26", "11/27"),
  Worked = c("No", "Yes", "Yes", "Yes", "No", "No", "No", "No",
             "Yes", "Yes", "Yes", "No", "No", "Yes"),
  mLWater = c(3667.12, 2513.75, 3164.37, 2365.88, 2868.63, 2868.63,
              2632.04, 2957.35, 2513.75, 2720.76, 2336.31, 2957.35,
              3016.50, 2661.62)
)

# Calculate the mean of mLWater
mean_mLWater <- claim1 %>%
  group_by(Worked) %>%
  summarize(AverageWater = mean(mLWater), na.rm=TRUE)
mean_mLWater
```

```
## # A tibble: 2 x 3
##   Worked AverageWater na.rm
##   <chr>         <dbl> <lgl>
## 1 No            2995. TRUE
## 2 Yes           2611. TRUE
```

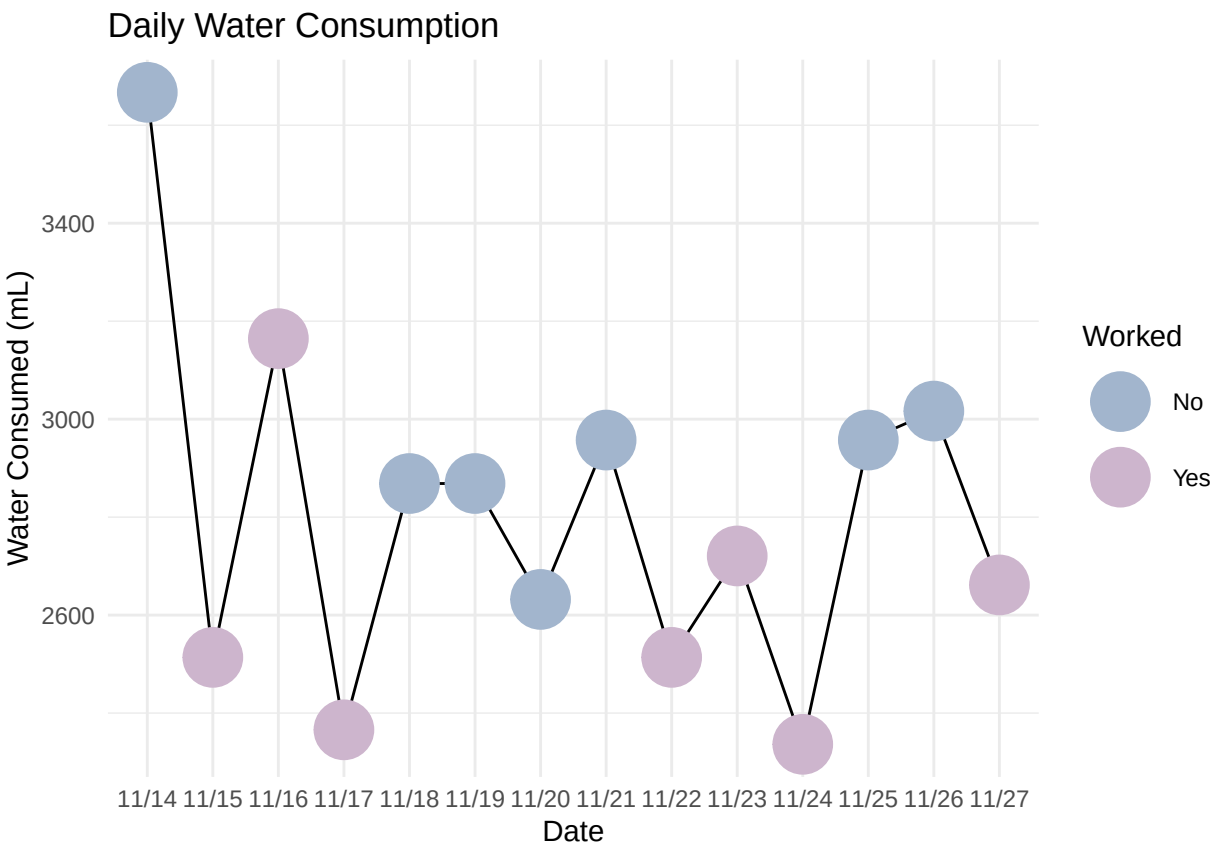
```r
t_test_result <- t.test(mLWater ~ Worked, data = claim1)
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  mLWater by Worked
## t = 2.3821, df = 11.797, p-value = 0.03496
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##   32.13778 736.77079
## sample estimates:
##  mean in group No mean in group Yes
##          2995.374          2610.920
```

Claim 1

```
library(ggplot2)

claim1_graph <- claim1 %>%
  ggplot(aes(x = Date, y = mLWater, group = 1)) +
  geom_line() +
  geom_point(aes(color=Worked), size=10) +
  labs(title = "Daily Water Consumption",
       y = "Water Consumed (mL)",
       x = "Date") +
  scale_color_manual(values=c("No"="lightsteelblue3", "Yes"="thistle3")) +
  theme_minimal()
claim1_graph
```



Changed to left tail test

```
claim1b <- data.frame(
  mLWater_Worked = c(2513.75, 3164.37, 2365.88, 2513.75, 2720.76, 2336.31, 2661.62),
  mLWater_Not_Worked = c(3667.12, 2868.63, 2868.63, 2632.04, 2957.35, 2898.21, 3016.60)
)

average_worked <- mean(claim1$mLWater_Worked, na.rm = TRUE)
```

```
## Warning in mean.default(claim1$mLWater_Worked, na.rm = TRUE): argument is not
## numeric or logical: returning NA
```

```
average_not_worked <- mean(claim1$mLWater_Not_Worked, na.rm = TRUE)
```

```
## Warning in mean.default(claim1$mLWater_Not_Worked, na.rm = TRUE): argument is
## not numeric or logical: returning NA
```

```
plot1 <- data.frame(
  Worked = c("Yes", "No"),
  AverageWater = c(average_worked, average_not_worked)
)
plot1
```

```
##   Worked AverageWater
## 1    Yes           NA
## 2     No           NA
```

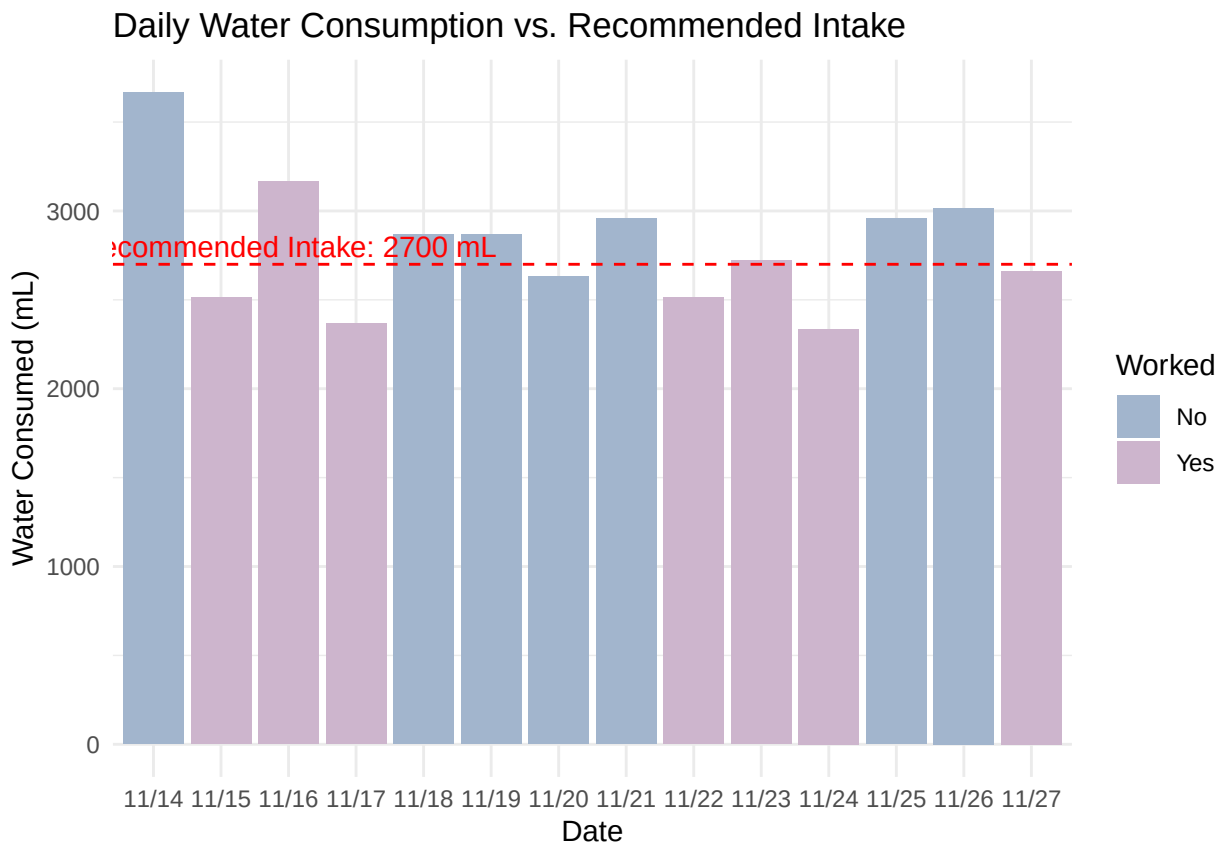
```r
t.test(claim1b$mLWater_Worked, claim1b$mLWater_Not_Worked, alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  claim1b$mLWater_Worked and claim1b$mLWater_Not_Worked
## t = -2.3219, df = 11.779, p-value = 0.0195
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -86.93474
## sample estimates:
## mean of x mean of y
##   2610.92   2986.94
```

```r
summary_data <- claim1 %>%
  group_by(Worked) %>%
  summarize(AverageWater = mean(mLWater, na.rm = TRUE))

ggplot(summary_data, aes(x = Worked, y = AverageWater, fill = Worked)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Water Consumption",
       y = "Average Water Consumed (mL)",
       x = "Work") +
  theme_minimal() +
  scale_fill_manual(values = c("Yes" = "thistle3", "No" = "lightsteelblue3")) +
  scale_y_continuous(limits = c(0, 3250), breaks = seq(0, 3500, by = 250))
```


```
rec_intake <- 2700
ggplot(claim1, aes(x = Date, y = mLWater, fill = Worked)) +
  geom_bar(stat = "identity") +
  geom_hline(yintercept = rec_intake, linetype = "dashed", color = "red") +
  labs(title = "Daily Water Consumption vs. Recommended Intake",
       y = "Water Consumed (mL)",
       x = "Date") +
  theme_minimal() +
  scale_fill_manual(values = c("Yes" = "thistle3", "No" = "lightsteelblue3")) +
  annotate("text", x = 3, y = rec_intake + 100,
           label = "Recommended Intake: 2700 mL", color = "red")
```




```
t.test(claim1$mLWater, mu=2700, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  claim1$mLWater
## t = 1.0962, df = 13, p-value = 0.2929
## alternative hypothesis: true mean is not equal to 2700
## 95 percent confidence interval:
##  2599.873 3006.422
## sample estimates:
## mean of x
##  2803.147
```