Homework 3

Janavi Kolpekwar

2025-09-25

Loading Libraries & Reading in Data

```
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    3.2.1
## 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
```

```
library(colorfindr)
library(dplyr)
library(ggplot2)
df <- read.csv("homework3_data.csv")
head(df)
```

```
##      sales design items nps
## 1 32.55146      0     2   4
## 2 35.38214      0     4   5
## 3 30.87418      0     3   4
## 4 35.54265      0     3   6
## 5 32.07379      0     2   6
## 6 31.55580      0     1   4
```

Summary Statistics of the Company Data

```
summary(df$sales)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.83   30.86   33.76   33.66   35.57   44.30
```

```
table(df$design)
```

```
##
##   0   1
## 101  99
```

Distinct Color Selection (e-commerce Company of Choice: Etsy)

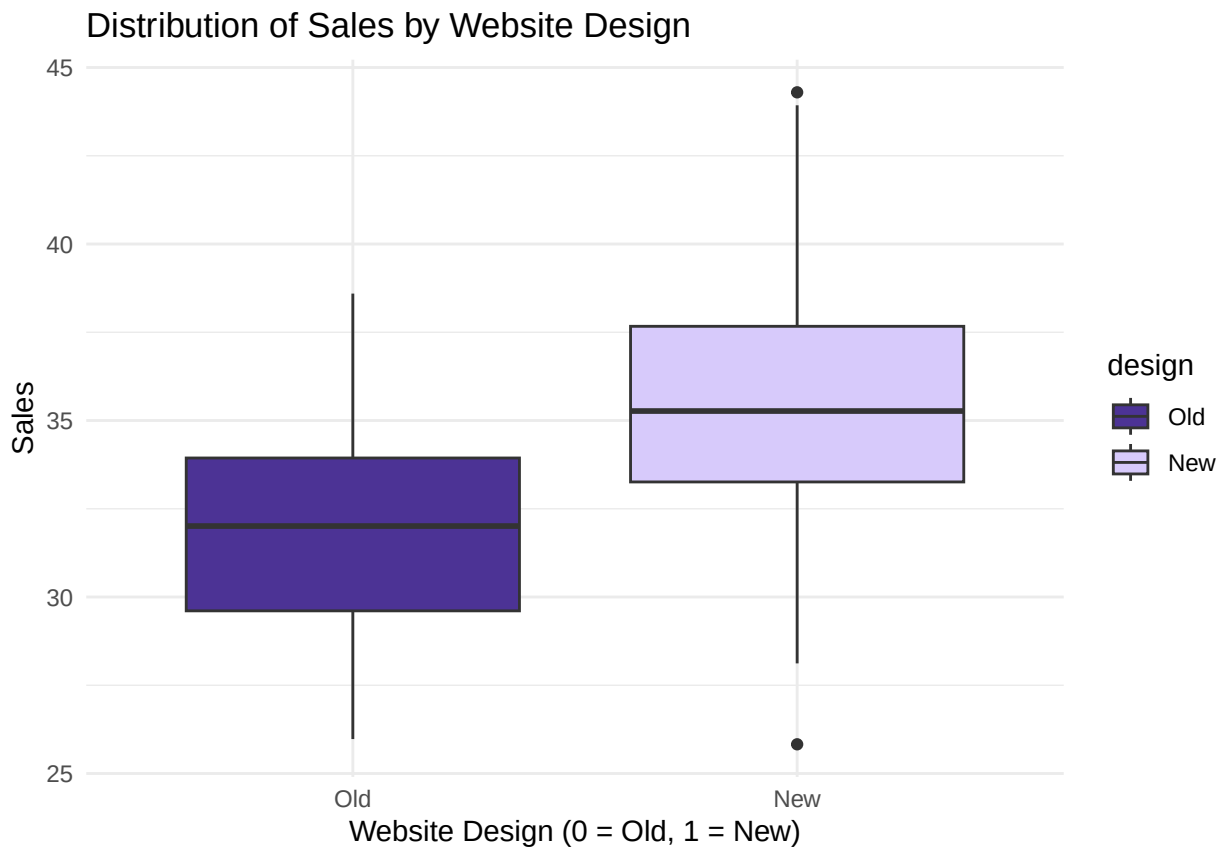
```
colors_data <- get_colors("~/Desktop/etsyimage.png")
print(colors_data)
```

```
## # A tibble: 105,697 x 3
##    col_hex col_freq col_share
##    <chr>      <int>     <dbl>
##  1 #FFFFFF   723905   0.430
##  2 #4C3395   339843   0.202
##  3 #D7CAFB    80684   0.0479
##  4 #FFFEFF    30103   0.0179
##  5 #4C3394    10028   0.00596
##  6 #FEFEFE     9648   0.00573
##  7 #FFFFFD     6157   0.00366
##  8 #FEFDFF     5628   0.00334
##  9 #4D3395     5062   0.00301
## 10 #FEFBFF     4751   0.00282
## # i 105,687 more rows
```

```
#Etsy company website color selection
custom_colors <- c("#4C3395", "#D7CAFB")
```

Graphical Evidence 1:

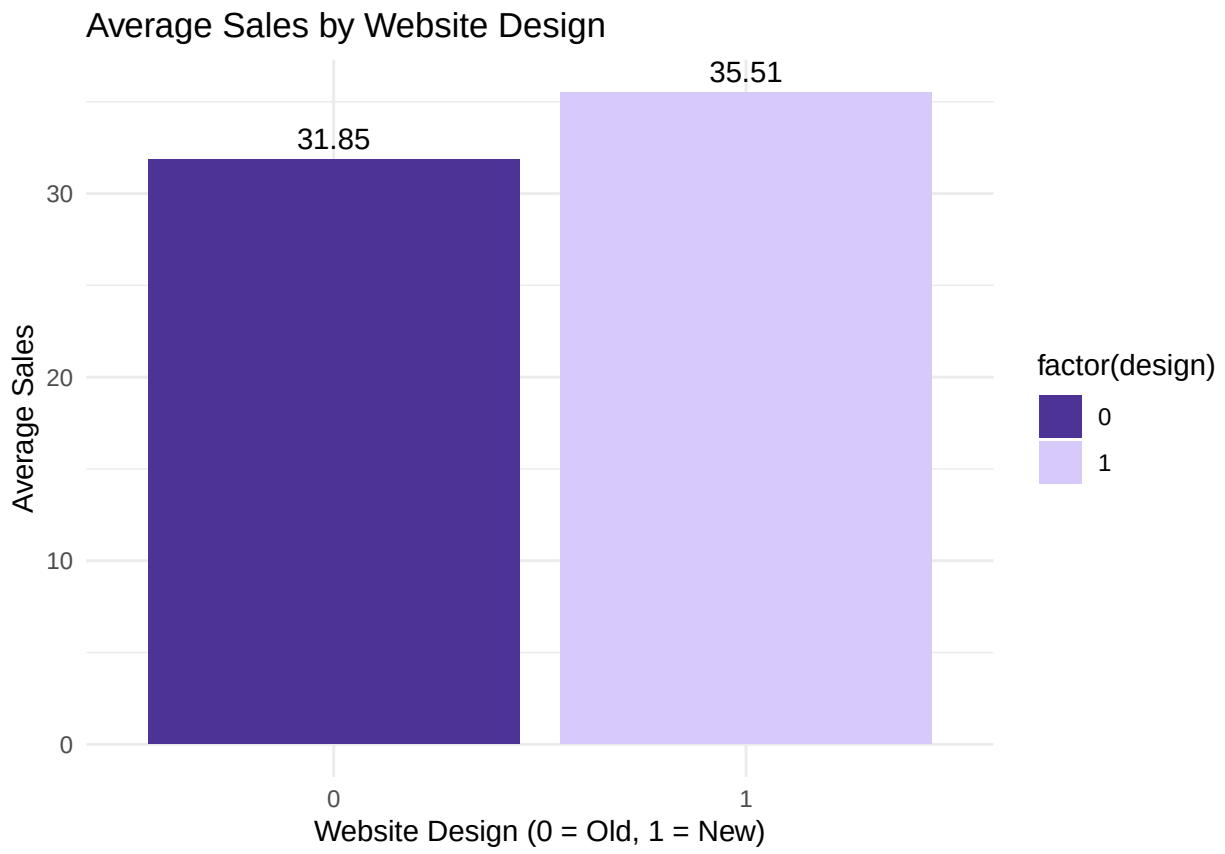
```
# Boxplot of sales by design
df %>%
  mutate(design= factor(design, labels = c("Old", "New"))) %>%
  ggplot(aes(x = design, y = sales, fill = design)) +
  geom_boxplot() +
  scale_fill_manual(values = custom_colors) +
  labs(x = "Website Design (0 = Old, 1 = New)", y = "Sales",
       title = "Distribution of Sales by Website Design") +
  theme_minimal()
```



Graphical Evidence 2:

```r
# Bar chart of average sales and NPS
summary_stats <- df %>%
  group_by(design) %>%
  summarise(
    avg_sales = mean(sales),
    avg_nps   = mean(nps),
    .groups = "drop"
  )

ggplot(summary_stats, aes(x = factor(design), y = avg_sales, fill = factor(design))) +
  geom_col(position = "dodge") +
  scale_fill_manual(values = custom_colors) +
  geom_text(aes(label = round(avg_sales, 2)), vjust = -0.5) +
  labs(x = "Website Design (0 = Old, 1 = New)", y = "Average Sales",
       title = "Average Sales by Website Design") +
  theme_minimal()
```



Estimated Sales Change from the group means:

```
sales_total_stats <- df %>%
  group_by(design) %>%
  summarize(mean_sales = mean(sales), groups = "drop")
sales_total_stats
```

```
## # A tibble: 2 x 3
##   design mean_sales groups
##    <int>      <dbl> <chr>
## 1      0       31.8 drop
## 2      1       35.5 drop
```

Average Increase in Sales:

```
avg_sales_increase <- sales_total_stats$mean_sales[2] - sales_total_stats$mean_sales[1]
avg_sales_increase
```

```
## [1] 3.664904
```

t-test Summary Stat:

```
t.test(sales ~ design, data=df)
```

```
##
##  Welch Two Sample t-test
##
## data:  sales by design
## t = -8.1554, df = 186.01, p-value = 5.042e-14
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -4.551445 -2.778364
## sample estimates:
## mean in group 0 mean in group 1
##        31.84819        35.51309
```

Final Company Report with follow-up considerations:

Primary Recommendation

- Based on the statistical analysis report, the company should commit to redesigning the website. The analysis strongly supports adopting the New website design, as both the boxplot and bar chart show consistently higher sales (35.5 vs. 31.8) with an average increase of about 3.66 units. The Welch t-test (p 5×10^{1}) provides overwhelming statistical evidence that this difference is not due to chance, confirming the redesign has a significant and practically meaningful impact on sales.

Alternative Statement

- Although the results strongly favor the New design, it's worth considering other possible explanations for the sales increase. Seasonal shifts, marketing promotions, or broader market conditions may have played a role, and the initial novelty of the redesign could temporarily boost engagement without lasting effects. The very low p-value points to a real difference between designs, but further testing and tracking additional measures like NPS would give a clearer picture of the redesign's long-term impact.

Graphical Evidence 1

- Based on the boxplot, the distribution of sales is clearly higher under the New website design, with both the median and overall range shifted upward compared to the Old design. This suggests that the redesign not only raised average performance but also improved consistency in sales outcomes.

Graphical Evidence 2

- Based on the histogram, the difference in mean sales is striking as the New design generated an average of about 35.5 units compared to 31.8 under the Old design, a meaningful increase of roughly 3.7 units. This provides strong visual evidence that the redesign has a positive effect on performance.

Statistical Standpoint

- From a statistical standpoint,the Welch two-sample t-test confirms that this difference is highly significant. The p-value of $\sim 5 \times 10^{1}$ and a confidence interval excluding zero reinforce that the observed improvement is not due to random variation, but reflects a consistent and reliable gain.

p-val Considerarion

- This very low p-value indicates overwhelming evidence in favor of the New design, making it highly unlikely that the difference arose by chance. The findings therefore support moving forward with the redesign as a data-driven decision.

Overal Analysis

- While the evidence support redesign, it's important to consider the alternative explanations for the observed increase in sales: Factors such as seasonal demand, short-term promotions, or the novelty of a new interface could also play a role in boosting sales. To ensure long-term success, the company should continue monitoring both sales and other key metrics such as Net Promoter Score (NPS) to validate that the redesign sustains its impact over time.