

HW3

Aishwarya Sreenivasan

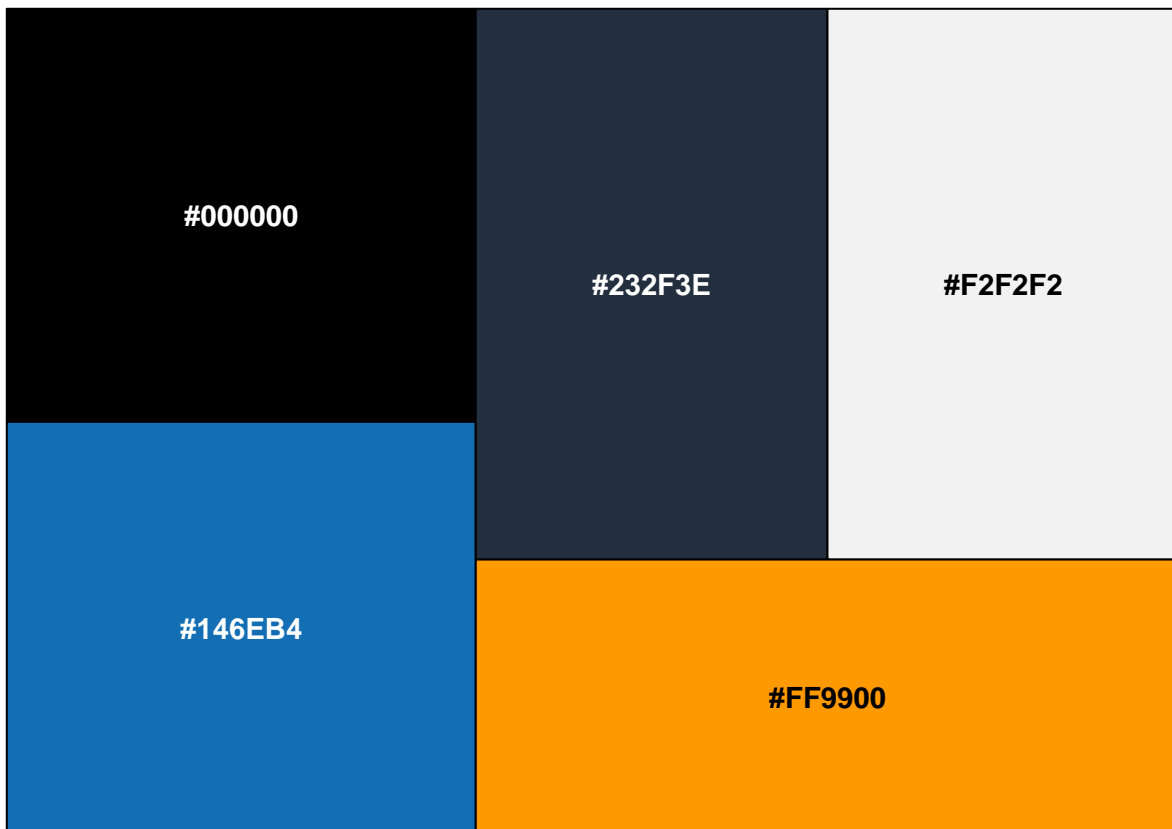
09-28-2025

```
# Loading libraries and dataset
```

```
library(colorfindr)  
library(tidyverse)  
library(ggplot2)  
dataset <- read_csv("homework3_data.csv")
```

```
# The color palette was sourced from Amazon, an eCommerce company that sells web-based products.
```

```
dat <- get_colors("C:\\Users\\Laddoo\\Documents\\SDS375Homework3\\amazon_col.png")  
plot_colors(dat[1:5, ])
```



Claim: The company should redesign their website, as doing so would lead to an average increase in sales of at least \$1.80 per customer.

Supporting Premise #1: The regression model indicates that the difference in sales with the redesign is greater than \$1.80.

```
# Regression model to find the relationship between design and sales.
```

```
lm(dataset$sales ~ dataset$design)
```

```
##
```

```
## Call:
```

```
## lm(formula = dataset$sales ~ dataset$design)
```

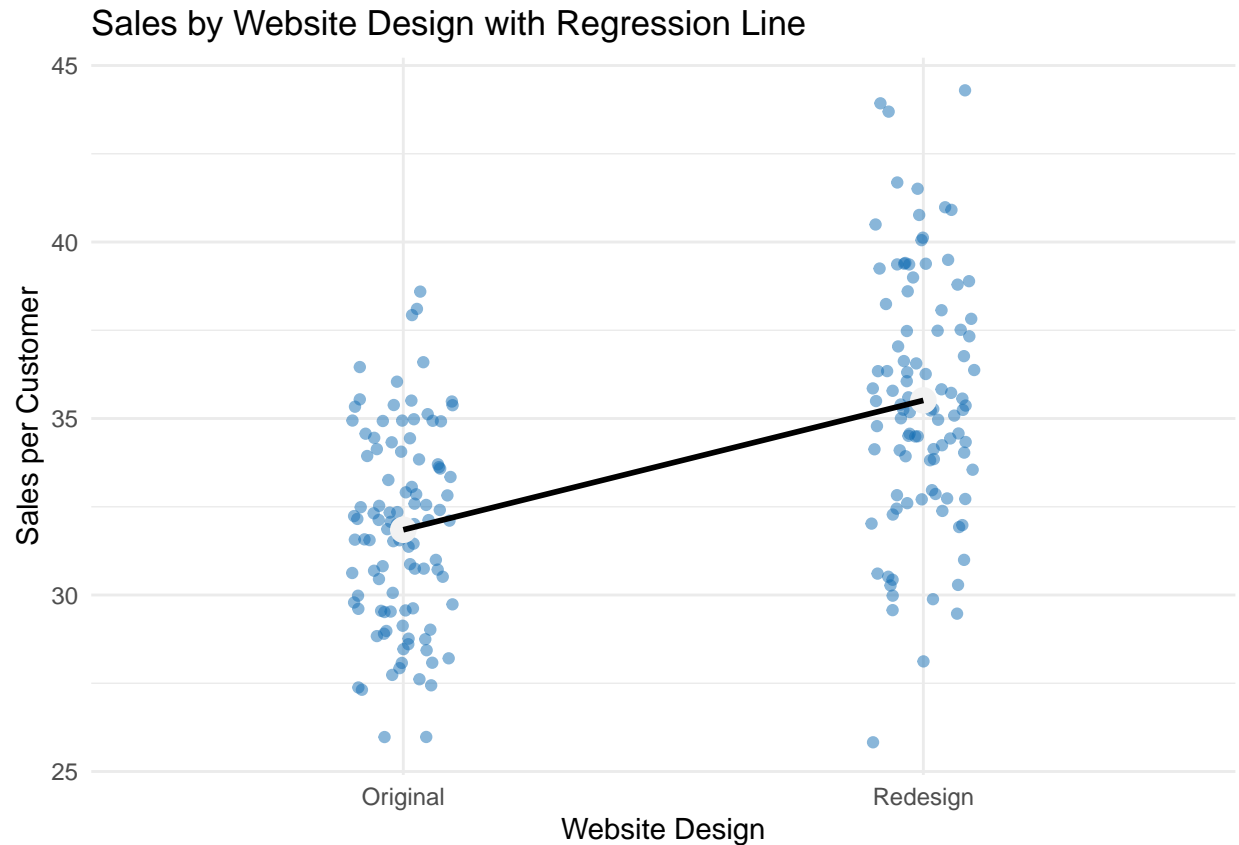
```
##
```

```
## Coefficients:
```

```
##      (Intercept)  dataset$design
```

```
##          31.848           3.665
```

```
ggplot(dataset, aes(x = factor(design), y = sales)) +  
  geom_jitter(width = 0.1, alpha = 0.5, color = "#146EB4") +  
  stat_summary(fun = mean, geom = "point", size = 4, color = "#F2F2F2") +  
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "#000000", size = 1) +  
  scale_x_discrete(labels = c("Original", "Redesign")) +  
  labs(  
    title = "Sales by Website Design with Regression Line",  
    x = "Website Design",  
    y = "Sales per Customer"  
  ) +  
  theme_minimal()
```



The estimated coefficient for the design variable is 3.665. The interpretation of the coefficient is that the difference between sales done with the original design, and the new design is \$3.67. This supports my claim that the redesign should be done, as the redesign leads to an average increase in sales greater than \$1.80 per customer.

Supporting Premise #2: The average sales between those with and without the redesign is greater than \$1.80, and is statistically significant.

```
# Calculates the difference in average sales value between those with the  
# original design and those with the redesign
```

```
original <- subset(dataset, design==0)  
redesign <- subset(dataset, design==1)  
mean_og <- mean(original$sales, na.rm = TRUE)  
mean_red <- mean(redesign$sales, na.rm = TRUE)  
print(mean_red-mean_og)
```

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

```
print(mean_red-mean_og > 1.80)
```

```
## [1] TRUE
```

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

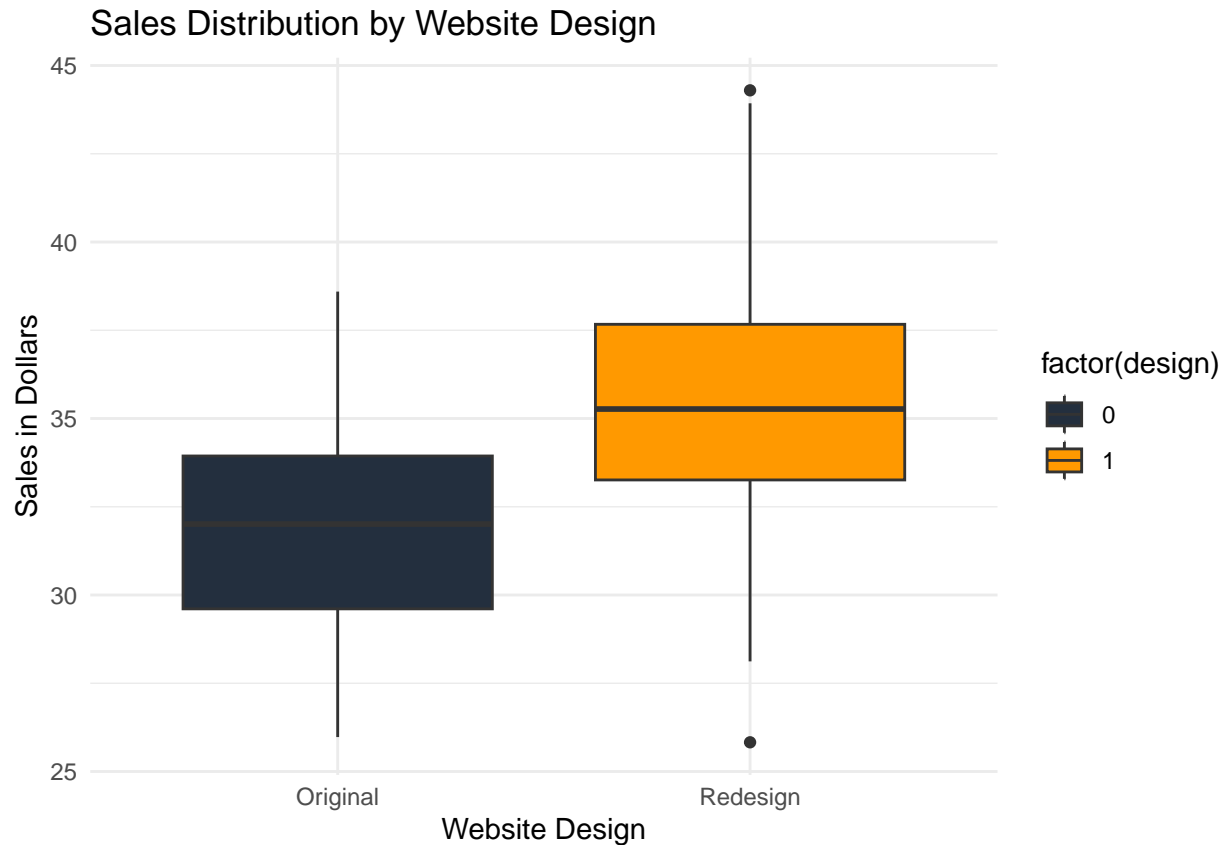
```
##  
## Welch Two Sample t-test  
##  
## data: sales by design  
## t = -8.1554, df = 186.01, p-value = 5.042e-14  
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
## 95 percent confidence interval:  
## -4.551445 -2.778364  
## sample estimates: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

This supports the claim, as the average difference between sales of the redesign and original design is \$3.67, which is greater than \$1.80, and that the difference is significant since the p-value is less than 0.05.

Supporting Premise #3: The median sales between those with and without the redesign is greater than \$1.80.

```
# Boxplot of original design and redesign to compares sales
```

```
ggplot(dataset, aes(x = factor(design), y = sales, fill = factor(design))) +  
  geom_boxplot() +  
  scale_x_discrete(labels = c("Original", "Redesign")) +  
  labs(title = "Sales Distribution by Website Design",  
        x = "Website Design",  
        y = "Sales in Dollars") +  
  scale_fill_manual(values = c("#232F3E", "#FF9900")) +  
  theme_minimal()
```



This boxplot supports the claim that the company should redesign the website, as the median difference in sales is more than \$1.80, with the redesign have a higher range of sales values than the original website design.

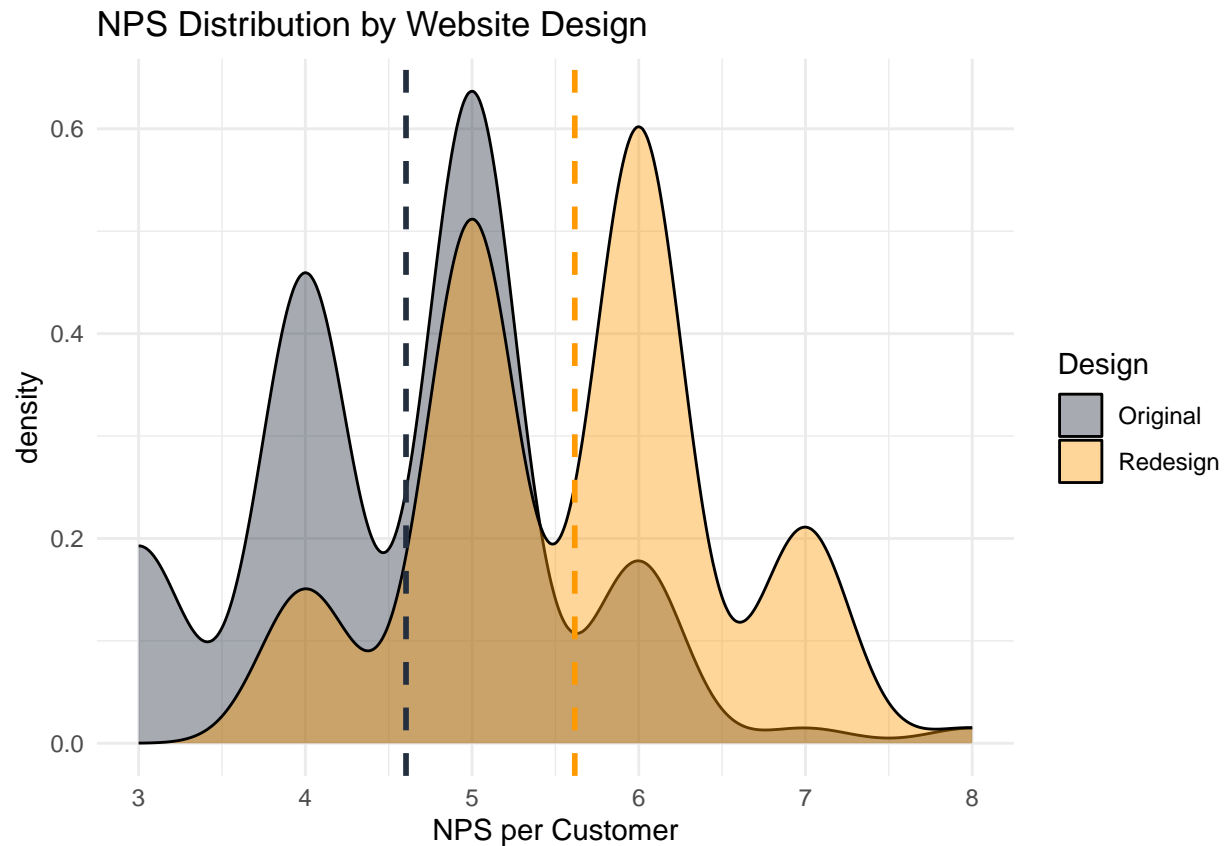
Supporting Premise #4: The average NPS per customer is greater with the website redesign than with the original design.

```
# Density plot showing distribution of NPS score per customer based on design.

mean_ogn <- mean(original$nps, na.rm = TRUE)
mean_redn <- mean(redesign$nps, na.rm = TRUE)

ggplot(dataset, aes(x = nps, fill = factor(design))) +
  geom_density(alpha = 0.4) +
  geom_vline(xintercept = mean_ogn, color = "#232F3E", linetype = "dashed", size = 1) +
  geom_vline(xintercept = mean_redn, color = "#FF9900", linetype = "dashed", size = 1) +
  scale_fill_manual(
    values = c("#232F3E", "#FF9900"),
    labels = c("Original", "Redesign")
  ) +
  labs(
    title = "NPS Distribution by Website Design",
    x = "NPS per Customer",
    fill = "Design"
  )
```

```
) +  
theme_minimal()
```



This graph supports the claim, as it shows that the average NPS per customer is higher by about one point with the redesign compared to the original design, making it more likely that the website with the new design will be more popular.

Recommendation:

The company should commit to redesigning the website. The evidence shows that the average difference in sales with the website redesign is \$3.67, which exceeds the threshold of a \$1.80 difference in sales set by the finance people (Premise #1). A t-test shows that the increase in sales with the redesign is statistically significant (Premise #2). Additionally, there is a greater median in sales and higher range of sales with the website redesign (Premise #3). Furthermore, the distribution of NPS per customer has a higher range with the website redesign, with the average score having an approximately 1 point increase, indicating that the website might become more popular with the website redesign (Premise #4). Thus, the company should redesign their website.

Alternative Statement:

The company should not redesign the website. It is possible that the difference in sales does not reflect true performance gains or is skewed by outliers or short-term factors. Additionally, if the customers surveyed after the redesign happen to be more loyal, high-spending, or tech-savvy, their responses may not generalize to the broader customer base. Furthermore, if the data was collected during a promotional period, holiday

season, or shortly after a product launch, the observed sales and NPS improvements may be confounded by external factors rather than the redesign itself. These data quality issues and external factors could lead to the company nonetheless choosing to stick with the original design.