Homework 3 | SDS 375

Nadine Angeles

2025-09-30

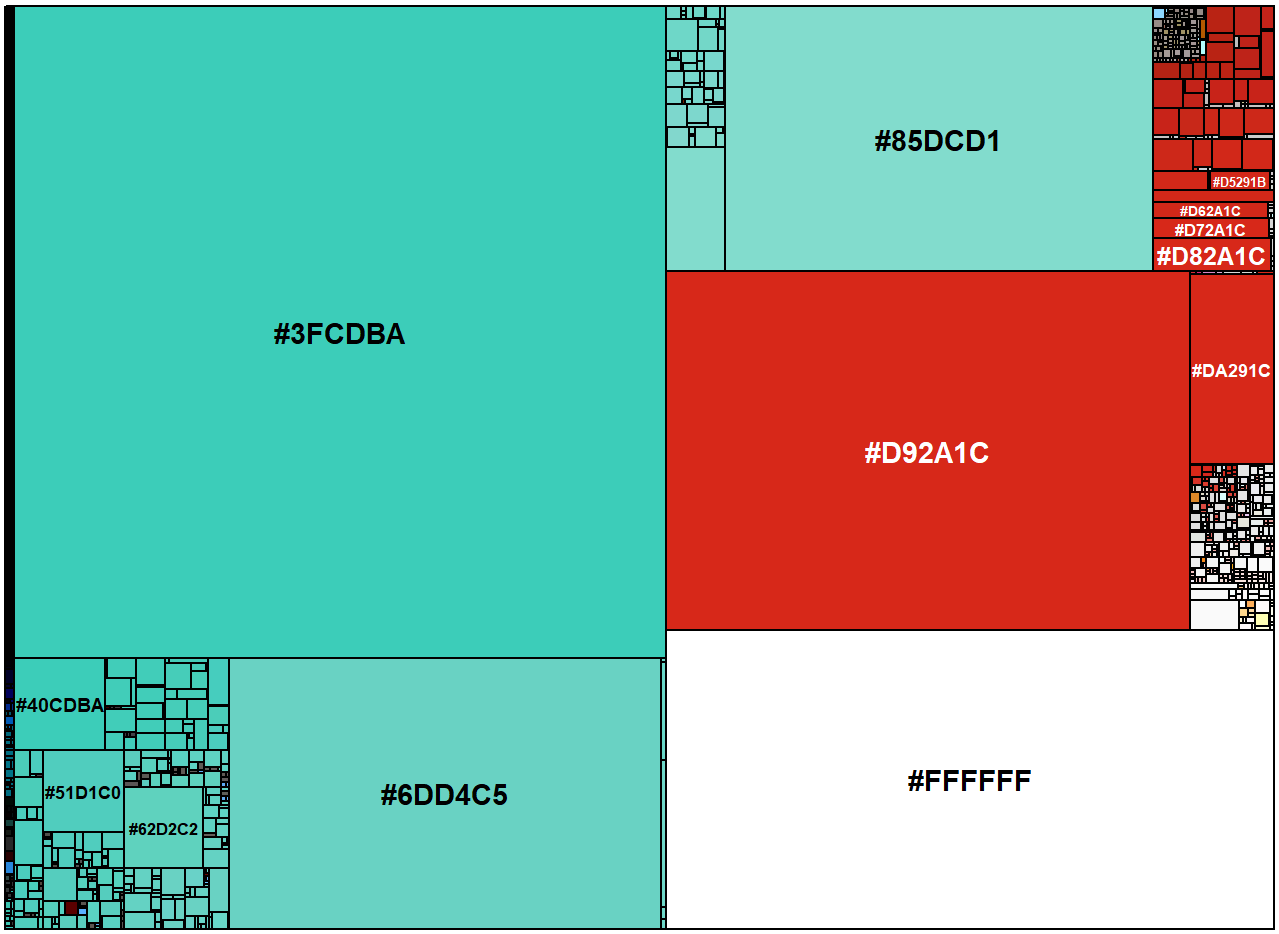
The Company should commit to redesigning the web site based on the criteria for success that they laid out.

Company Colors:

```
palette <- get_colors("P.Terry.png")
palette
```

```
## # A tibble: 112,911 × 3
##    col_hex col_freq col_share
##    <chr>      <int>     <dbl>
##  1 #3FCDBA   617997   0.332
##  2 #D92A1C   273501   0.147
##  3 #FFFFFF   263543   0.142
##  4 #6DD4C5   170043   0.0915
##  5 #85DCD1   165070   0.0888
##  6 #DA291C    23138   0.0124
##  7 #40CDBA    12079   0.00650
##  8 #84DCD1    10641   0.00572
##  9 #51D1C0     9825   0.00529
## 10 #62D2C2     9379   0.00505
## # ℹ 112,901 more rows
```

```
plot_colors(palette[1:1000, ])
```




```
cols <- make_palette(palette[1:100, ])
```

```
cols
```

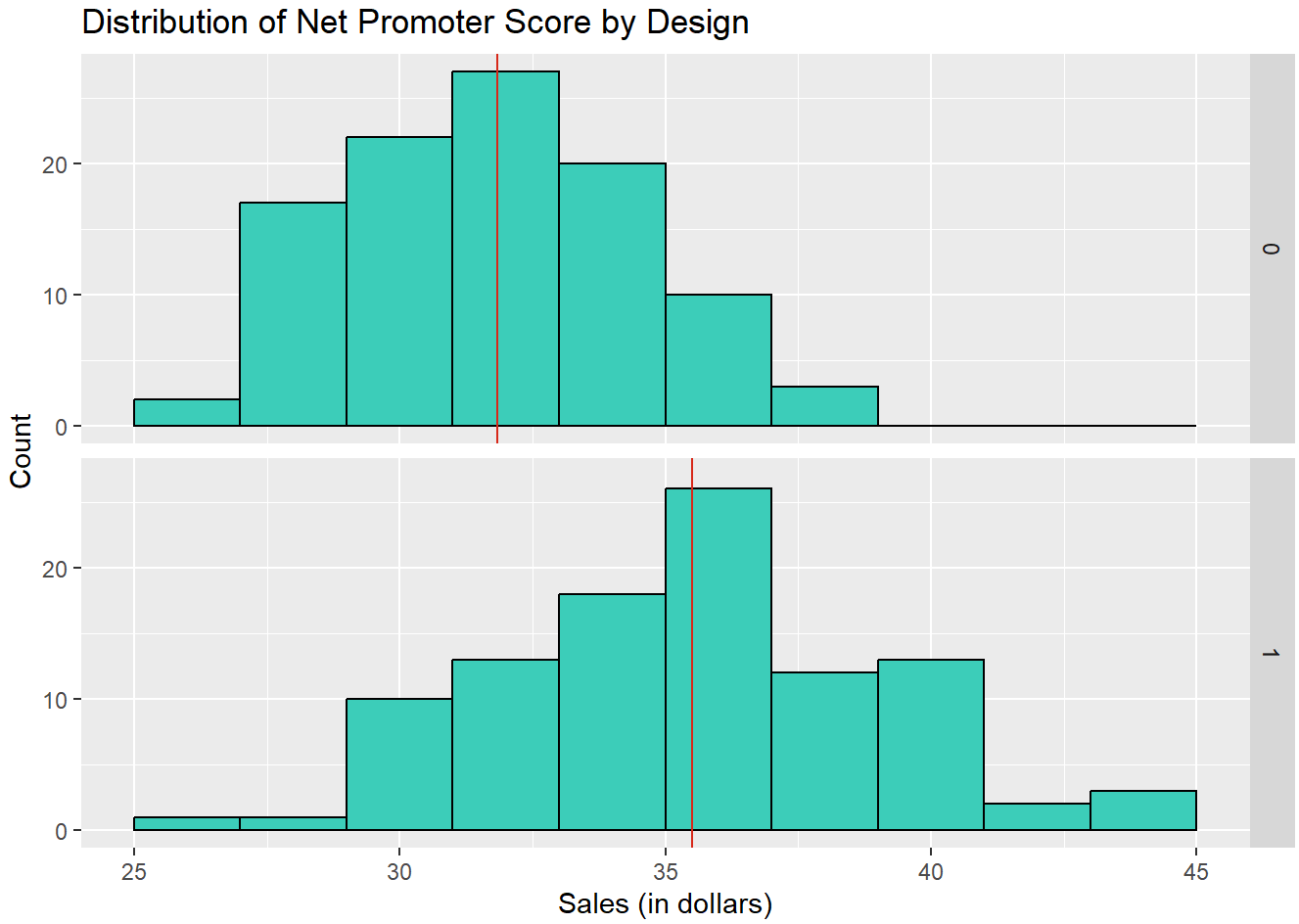
```
##  [1] "#3FCDBA" "#D92A1C" "#FFFFFF" "#6DD4C5" "#85DCD1" "#51D1C0" "#62D2C2"
##  [8] "#000000" "#C52619" "#69D3C4"
```

Visualizations:

```
# Distribution of price by design type

mean_sales <- company |>
  group_by(design) |>
  summarise(avg_sales = mean(sales, na.rm = TRUE)) |>
  arrange(avg_sales)

ggplot(company, aes(x = sales)) +
  geom_histogram(binwidth = 2, fill = "#3FCDBA", color = "#000000") +
  facet_grid(design ~ .) +
  labs(
    x = "Sales (in dollars)",
    y = "Count",
    title = "Distribution of Net Promoter Score by Design"
  ) +
  geom_vline(
    data = mean_sales,
    aes(xintercept = avg_sales),
    lwd = .5,
    linetype = 1,
    color = "#D92A1C")
```

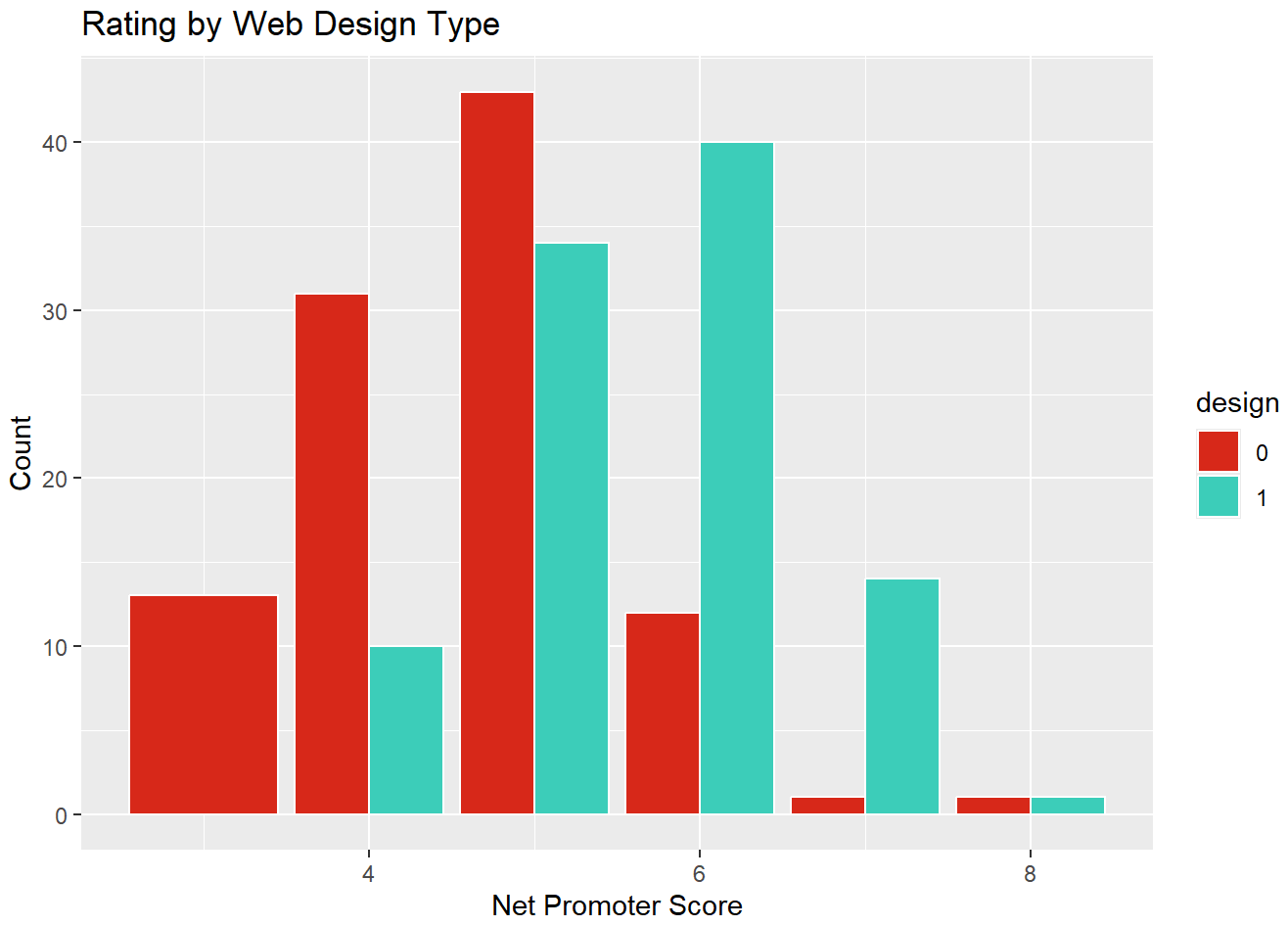


Our graph shows the distribution of sales per web design, with “0” representing the old web design and “1” representing the new design.

Design one has an average sale of about 35 dollars while Design zero has an average sale of about 31 dollars.

```
# Net Promoter Score by design type

ggplot(company, aes(x = nps, fill = design)) +
  geom_bar(position = "dodge", color = "#FFFFFF") +
  labs(
    x = "Net Promoter Score",
    y = "Count",
    title = "Rating by Web Design Type"
  ) +
  scale_fill_manual(
    values = c("1" = "#3FCDBA", "0" = "#D92A1C"))
```



We can see that design 0 tends to have a lower Net Promoter Score, while design 1 is shown to have higher ratings.

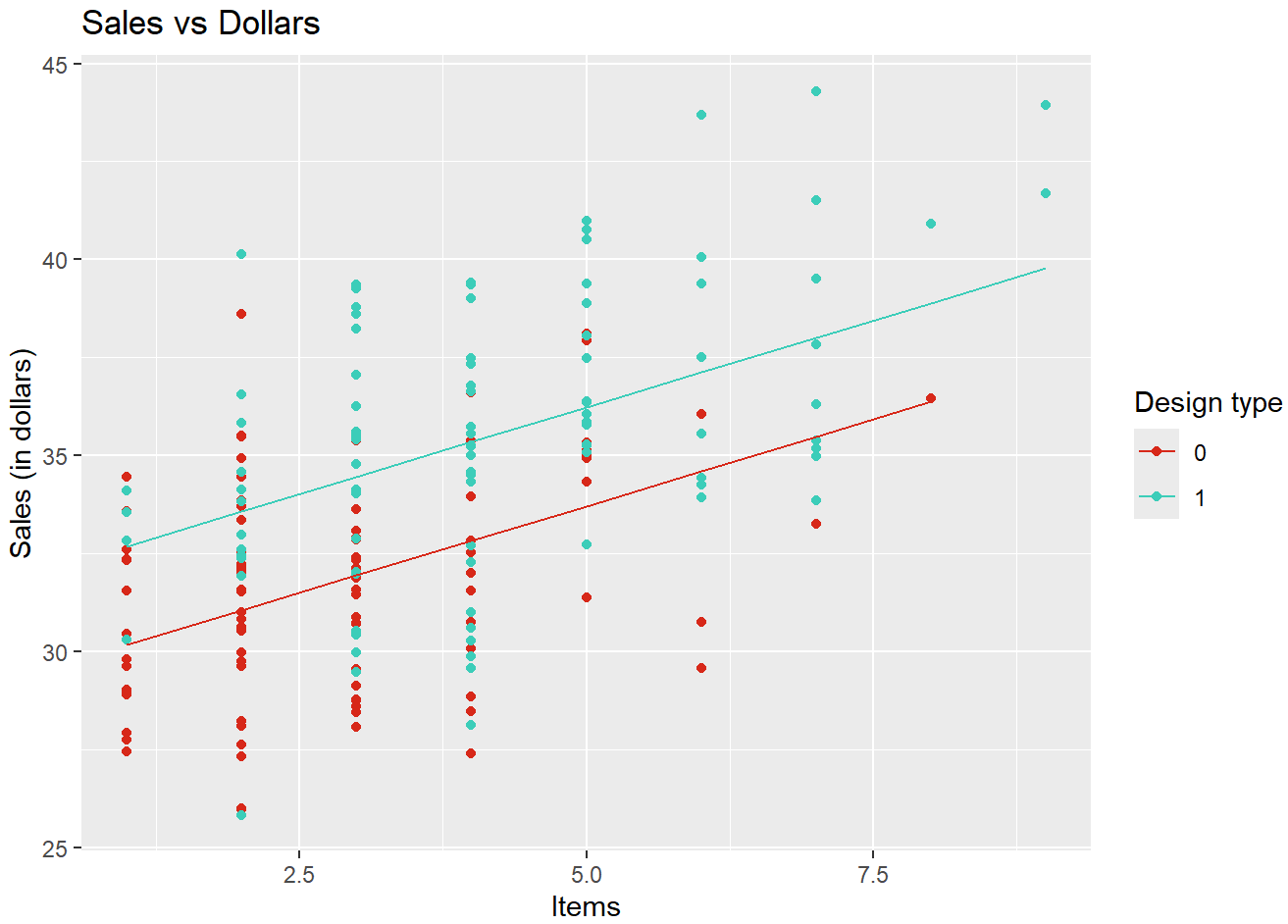
Conclusion:

```
# Regression for sales by design
sales_lm <- lm(sales ~ design + items, data = company)
summary(sales_lm)
```

```
##
## Call:
## lm(formula = sales ~ design + items, data = company)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.7398 -2.2291  0.0971  1.8135  7.5470
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  29.2730     0.4620  63.368  < 2e-16 ***
## design1       2.5190     0.4333   5.813 2.44e-08 ***
## items         0.8877     0.1260   7.047 3.00e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.84 on 197 degrees of freedom
## Multiple R-squared:  0.4029, Adjusted R-squared:  0.3968
## F-statistic: 66.46 on 2 and 197 DF,  p-value: < 2.2e-16
```

```
# Add predicted values (yhat) into our data frame
company$yhat <- sales_lm$fitted.values

# Plot
ggplot(company, aes(x = items, y = sales, col = design)) +
  geom_point() +
  labs(
    x = "Items",
    y = "Sales (in dollars)",
    title = "Sales vs Dollars",
    col = "Design type"
  ) +
  geom_line(aes(y = yhat, col = design)) +
  scale_color_manual(
    values = c("1" = "#3FCDBA", "0" = "#D92A1C"))
```



Holding the number of items constant, implementing design 1 is expected to increase sales by $2.52 per customer compared to design 0. Thus the redesign will lead to an average increase in sales of at least $1.80 per customer.

Recommendation: Commit to the redesign because it will increase sales Alternative statement: The redesign would not increase sales even though the data suggest it would.