

Homework3

```
# Load libraries
library(tidyverse)
library(colorfindr)
library(ggplot2)
```

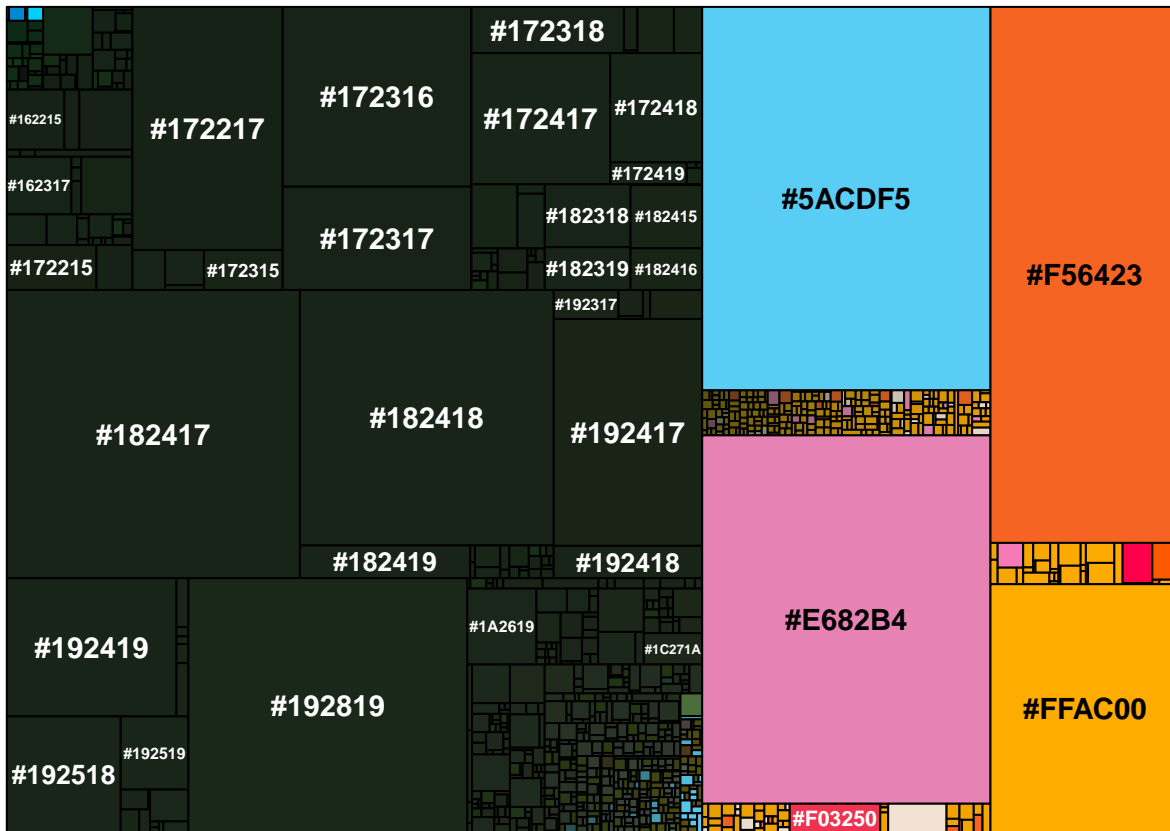
```
# Read in the data
df = read.csv("homework3_data.csv")
```

Color Palette Creation

```
dat <- get_colors("acl_website.png")
dat
```

```
## # A tibble: 35,461 x 3
##   col_hex col_freq col_share
##   <chr>    <int>    <dbl>
## 1 #5ACDF5  223893    0.108
## 2 #E682B4  215086    0.103
## 3 #F56423  204867    0.0986
## 4 #182417  171053    0.0823
## 5 #192819  144032    0.0693
## 6 #182418  131336    0.0632
## 7 #FFAC00   95119    0.0458
## 8 #172217   73998    0.0356
## 9 #172316   68740    0.0331
## 10 #192417   68227    0.0328
## # i 35,451 more rows
```

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



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



```
cols
```

```
## [1] "#5ACDF5" "#F56423" "#182417" "#192819" "#FFAC00" "#172217" "#172316"
## [8] "#192417" "#192419" "#142917"
```

Part A- use at least 2 graphical presentations that support your recommendation

Five Number Summary Table

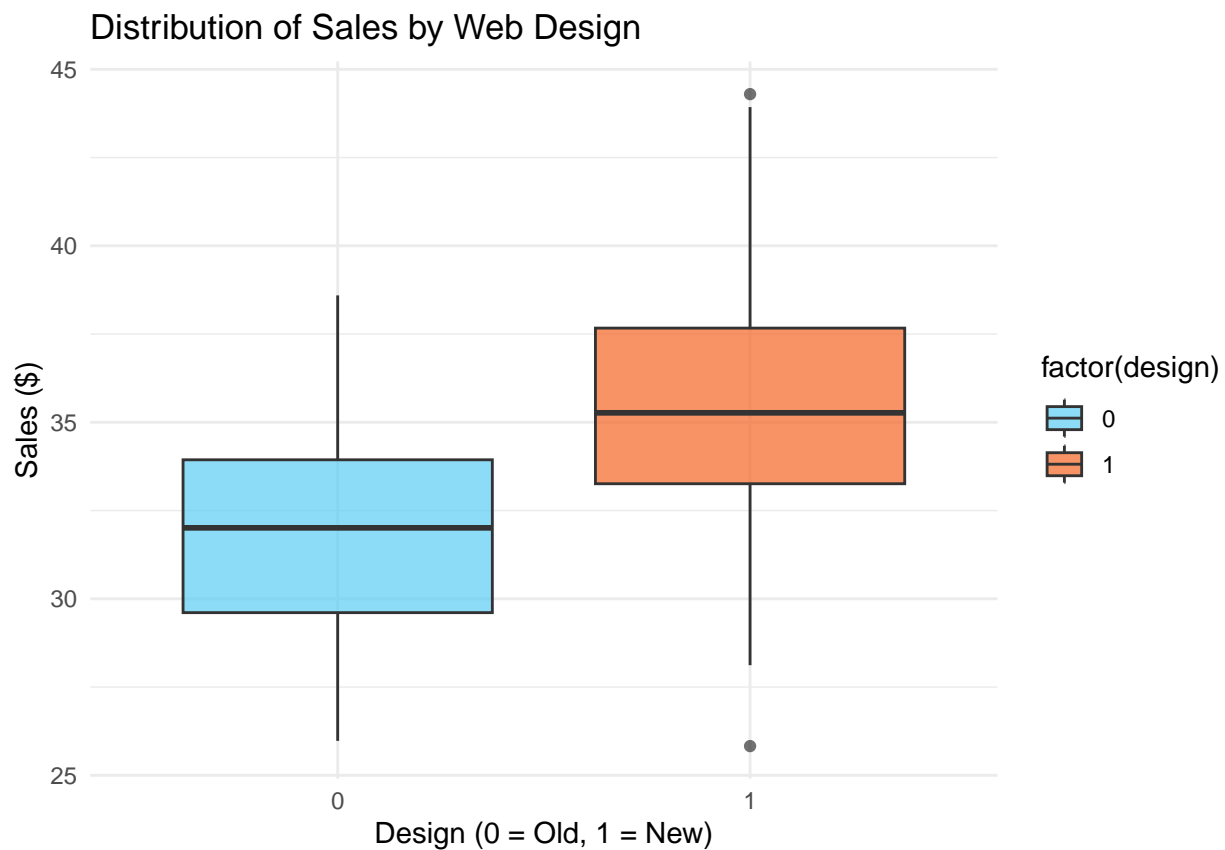
```
df %>%
  group_by(design) %>%
  summarise(
    Min = min(sales),
    Q1 = quantile(sales, 0.25),
    Median = median(sales),
    Q3 = quantile(sales, 0.75),
    Max = max(sales)
  )
```

```
## # A tibble: 2 x 6
##   design    Min    Q1 Median    Q3    Max
```

```
##      <int> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      0  26.0  29.6   32.0  33.9  38.6
## 2      1  25.8  33.3   35.3  37.7  44.3
```

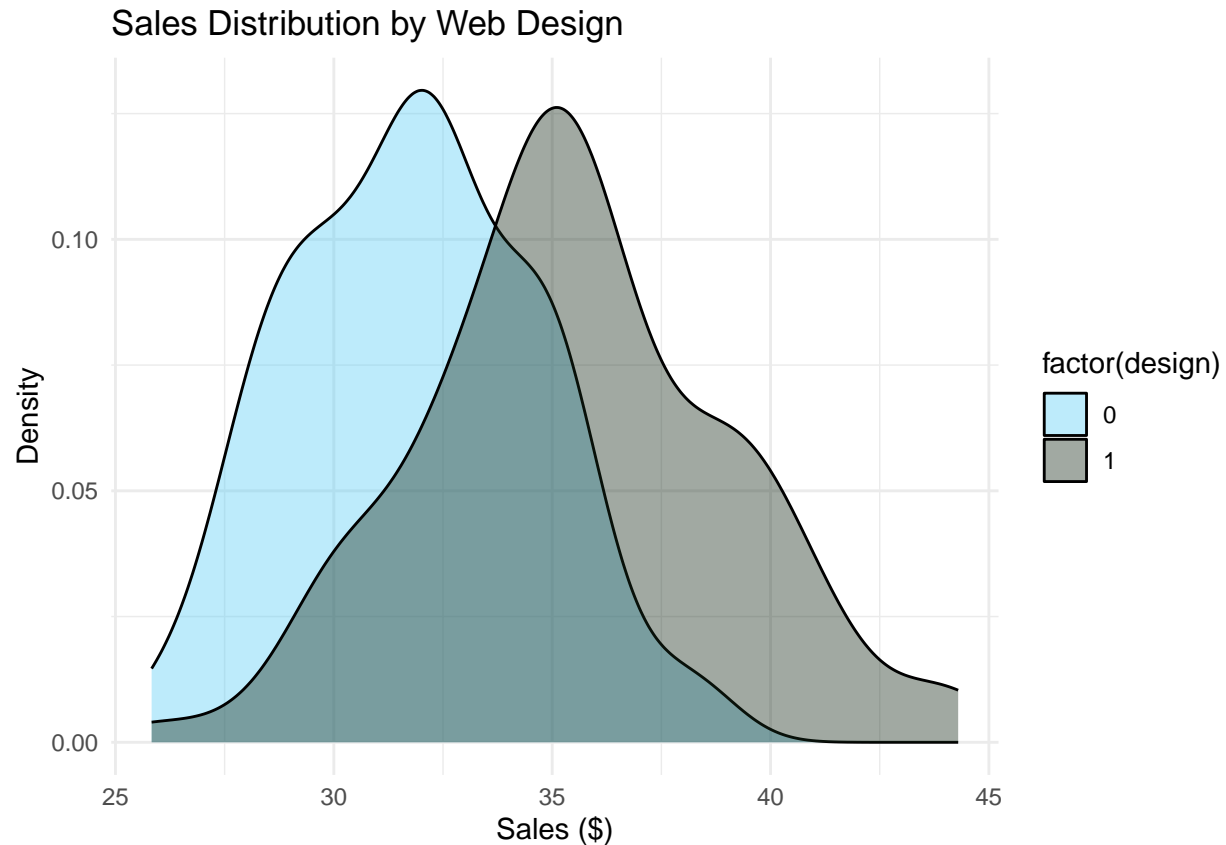
Graphical Representation 1

```
ggplot(df, aes(x = factor(design), y = sales, fill = factor(design))) +
  geom_boxplot(alpha = 0.7) +
  scale_fill_manual(values = cols[1:2]) +
  labs(x = "Design (0 = Old, 1 = New)", y = "Sales ($)",
       title = "Distribution of Sales by Web Design") +
  theme_minimal()
```



Graphical Representation 2

```
ggplot(df, aes(x = sales, fill = factor(design))) +
  geom_density(alpha = 0.4) +
  scale_fill_manual(values = c(cols[1], cols[10])) +
  labs(x = "Sales ($)", y = "Density",
       title = "Sales Distribution by Web Design") +
  theme_minimal()
```



Part B - Estimate how much sales will increase/decrease if the redesign is done

T-Test

```
t_test <- t.test(sales ~ design, data = df,
                 alternative = "greater", mu = 1.80, var.equal = TRUE)
t_test
```

```
##
## Two Sample t-test
##
## data: sales by design
## t = -12.19, df = 198, p-value = 1
## alternative hypothesis: true difference in means between group 0 and group 1 is greater than 1.8
## 95 percent confidence interval:
## -4.405799 Inf
## sample estimates:
## mean in group 0 mean in group 1
## 31.84819 35.51309
```

New design mean sales = \$35.51 Old design mean sales = \$31.85 Difference = +\$3.66 per customer

The t-test (unadjusted) did not find strong statistical evidence that the increase was $> \$1.80$, likely because of how the test was set up relative to the null as p-value came out as 1.

Linear Regression

```
model <- lm(sales ~ design + items, data = df)
summary(model)
```

```
##
## Call:
## lm(formula = sales ~ design + items, data = df)
##
## 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 ***
## design        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
```

Redesign effect = $+\$2.52$ per customer.

The regression analysis (which adjusts for cart size) shows that the new design is linked to an average increase of $\$2.52$ per customer, statistically significant ($p = 2.44e-08$) and above the $\$1.80$ requirement.

Part C- whether the redesign will lead to an average increase in sales of at least $\$1.80$ per customer

Recommendation for the Company

The two visuals show that the prior redesign resulted in an average increase in sales of more than $\$1.80$ per customer.

The regression results are more reliable because they account for customer behavior (items purchased). Based on this, the company has good evidence to proceed with the redesign, as it likely boosts sales above the target threshold of $\$1.80$.

Primary and Alternative Statements

Based on the historical data, the redesign is likely to increase average sales per customer by approximately $\$2.50$, which exceeds the company's required threshold of $\$1.80$. Therefore, the evidence supports moving forward with the full redesign, as it is expected to generate meaningful sales growth.

Alternative Statement

Even though the analysis suggests the redesign increases sales above the threshold, there is a possibility that the observed increase is due to confounding factors (e.g., differences in customer purchasing behavior, seasonality, promotions, or net promoter score), and the redesign may not truly increase sales.