SDS 375 Data Product Development: Homework 3

Lauren Maurer

2025-10-01

```r
dat <- read.csv('homework3_data.csv')

library(colorfindr)
library(ggplot2)
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
## Creating color palette

web <- get_colors("web1.png")
web
```

```
## # A tibble: 504,132 x 3
##    col_hex col_freq col_share
##    <chr>      <int>     <dbl>
##  1 #FFFFFF   336136   0.178
##  2 #152D62   215561   0.114
##  3 #EE1298    22404   0.0118
##  4 #FCFCFC    17805   0.00942
##  5 #FEFEFE     9874   0.00522
##  6 #FBFBFB     4541   0.00240
##  7 #FDFDFD     3975   0.00210
##  8 #585858     3844   0.00203
##  9 #D2D3D4     3810   0.00201
## 10 #F9F9F9     3503   0.00185
## # i 504,122 more rows
```
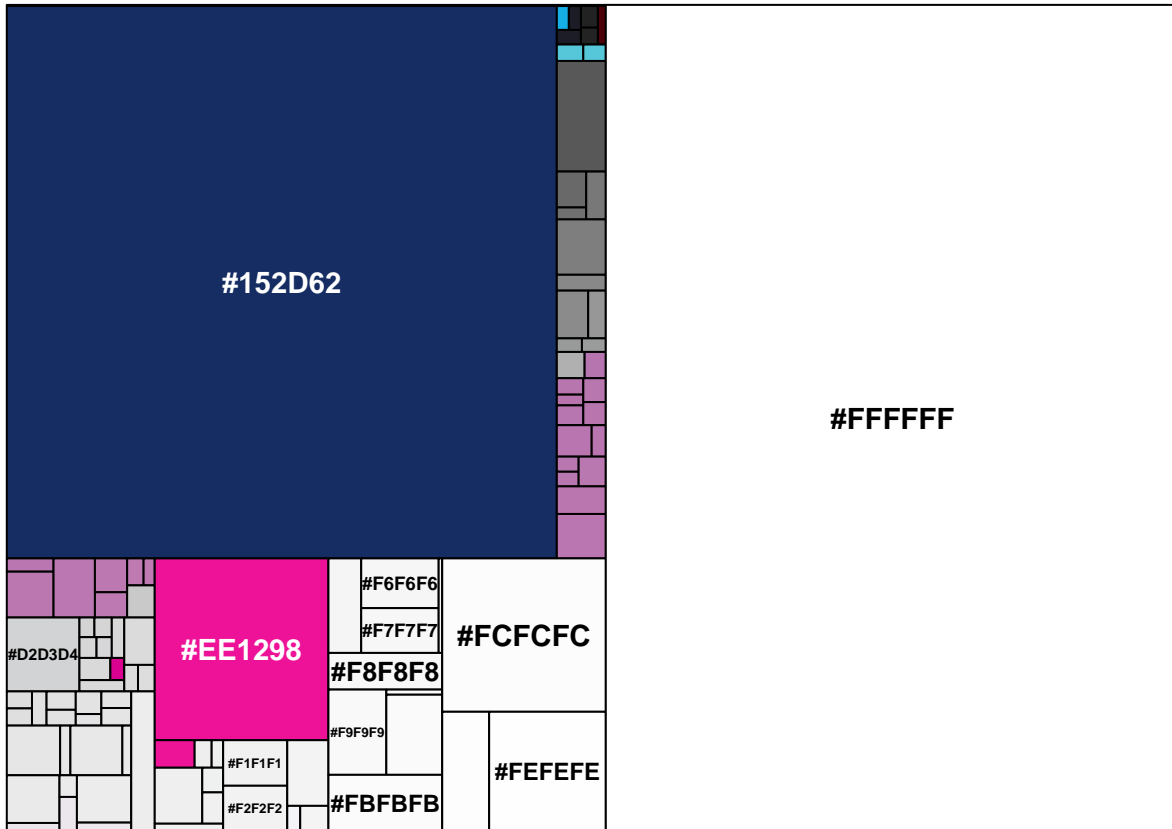
```r
plot_colors(web[1:100, ])
```

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

```
cols
```

```
##  [1] "#FFFFFF" "#152D62" "#EE1298" "#585858" "#D2D3D4" "#7E7E7E" "#BB77B1"
##  [8] "#8B8B8B" "#C9C9C9" "#56C6D9"
```

```
my_colors <- c(
  "Old Design" = '#BA78AF',
  "New Design" ='#6fc6d9'
)
```

Recommendation

The company should commit to the redesign for the website.

Supporting Evidence

```
# Supporting Plots

# Make design a factor for better labels and colors
dat$design_label <- factor(dat$design, levels = c(0, 1),
                           labels = c("Old Design", "New Design"))
```

```
# Histogram
ggplot(dat, aes(x = sales,)) +
  geom_histogram(bins=15, fill = "#152D62") +
  facet_wrap(~design_label, ncol = 1) +
  labs(title = "Sales by Design",
       x = "Design", y = "Sales")
```
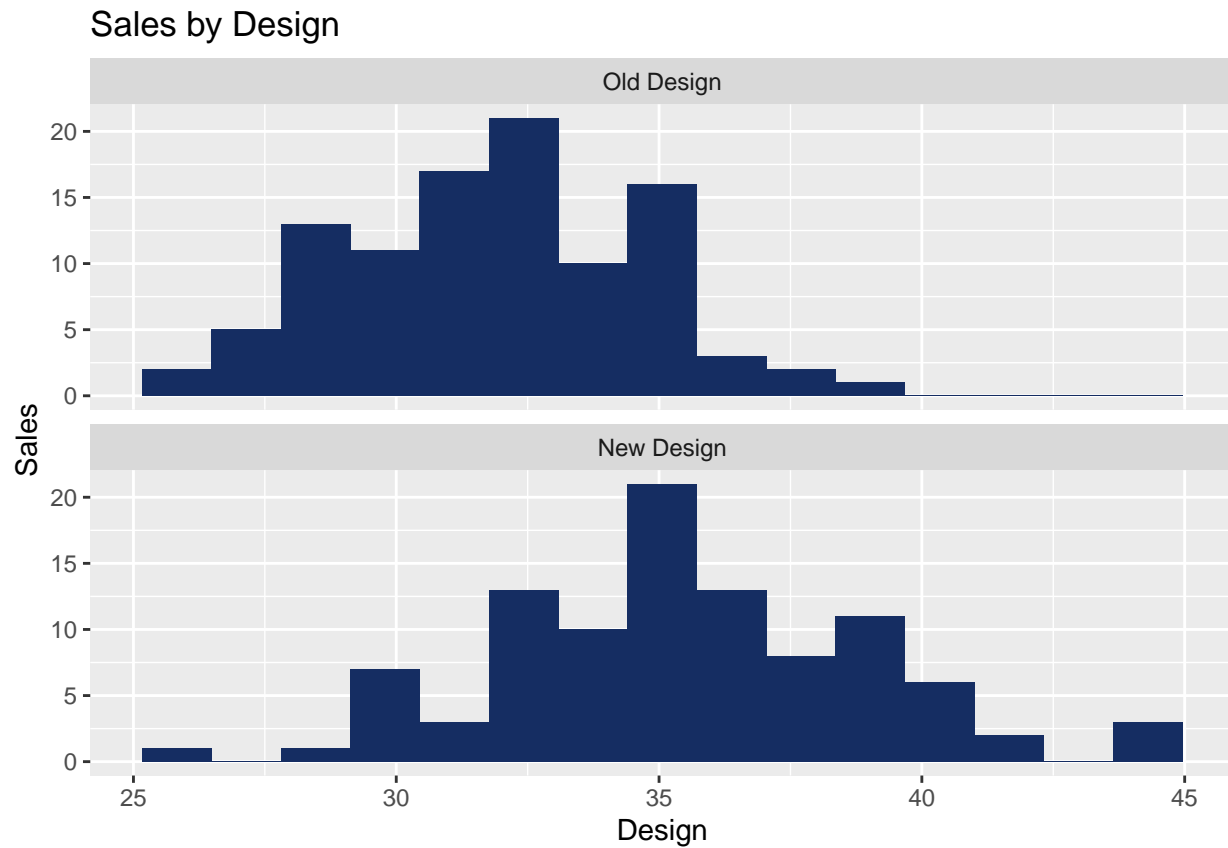

Sales by Design

```
# Box plot
ggplot(dat, aes(x = factor(design), y = sales, fill = factor(design_label)))+
  geom_boxplot() +
  labs(title = "Boxplot of Sales by Design",
       x = "Design", y = "Sales",
       fill = "Design Label") +
  scale_fill_manual(values = my_colors)
```

Boxplot of Sales by Design
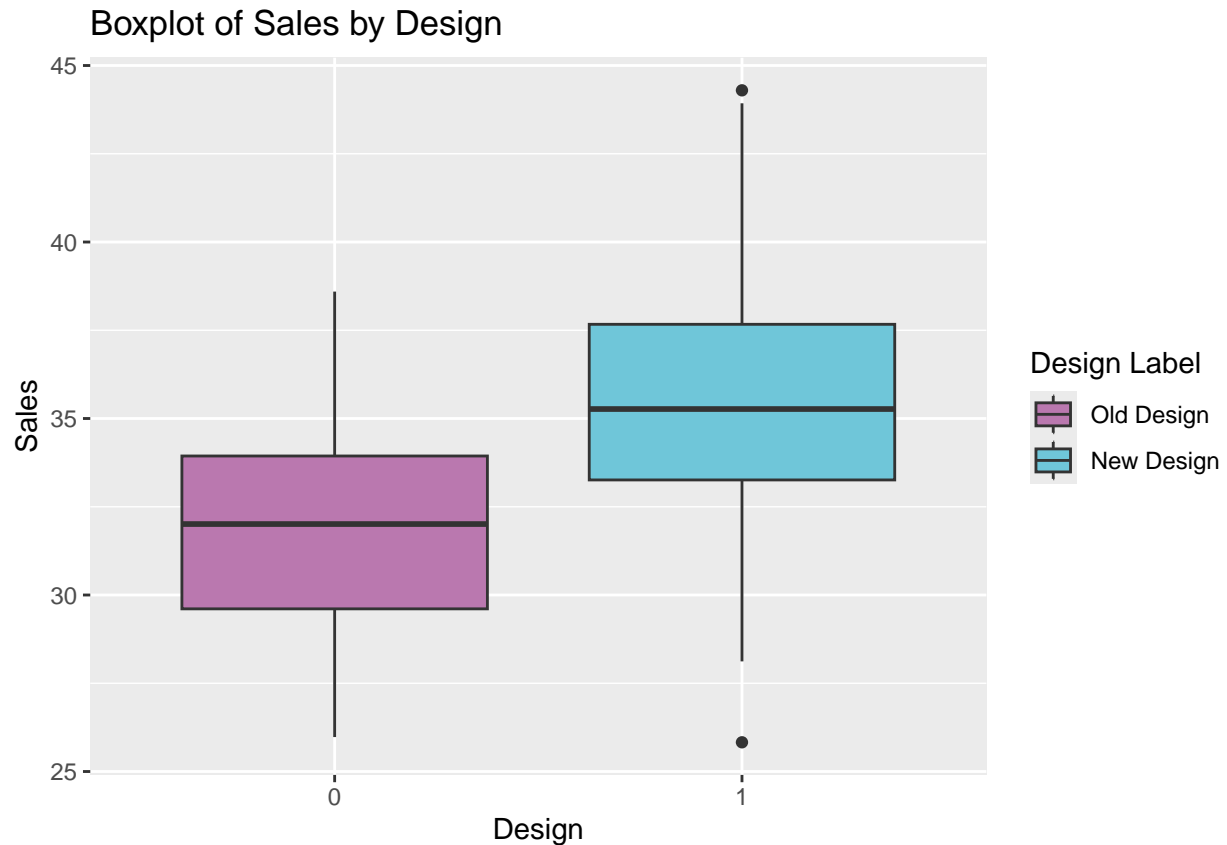
In the faceted histogram the graph of the new design is shifted to the right showing an average increase in sales compared to the old design. The box plot of the new design is shifted up indicating an increase in median sales. The median sales for the old design is around 32 and the median sales for the new design is around 35. Both of these plots support that the website should be redesigned.

```
# Linear model
model <- lm(sales ~ design, data = dat)
summary(model)
```

```
##
## Call:
## lm(formula = sales ~ design, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6855 -2.2511 -0.0673  2.1205  8.7838
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.8482     0.3154 100.969  < 2e-16 ***
## design        3.6649     0.4483   8.175 3.49e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.17 on 198 degrees of freedom
## Multiple R-squared:  0.2523, Adjusted R-squared:  0.2486
```

```
## F-statistic: 66.83 on 1 and 198 DF,  p-value: 3.489e-14
```

```r
# Calculate difference in means
mean_new <- mean(dat$sales[dat$design == 0])
mean_old <- mean(dat$sales[dat$design == 1])
mean_diff <- mean_new - mean_old


old <- dat$sales[dat$design == 0]
new <- dat$sales[dat$design == 1]

# One-sided t-test
ttest <- t.test(new, old, alternative = "greater", mu = 1.80)
```

Conducting a t-test for the difference in means between the old design and the new design shows that there is statistical evidence that there is an average increase of at least \$1.80. The p-value is 2.53×10^{-5}, which is very small and indicates statistical significance. If the redesign is done there is an estimated \$3.66 increase for each customer.

Alternative Statement

Recommendation: Do the redesign because it will increase sales.

Alternative: The redesign would not increase sales even though the data suggests it would.