

HW3

Nancy Yeung (ny3257)

```
library(ggplot2)
library(RColorBrewer)
library(dplyr)
library(kableExtra)
library(knitr)
library(tidyverse)
library(mosaic)
library(boot)
library(colorfindr)

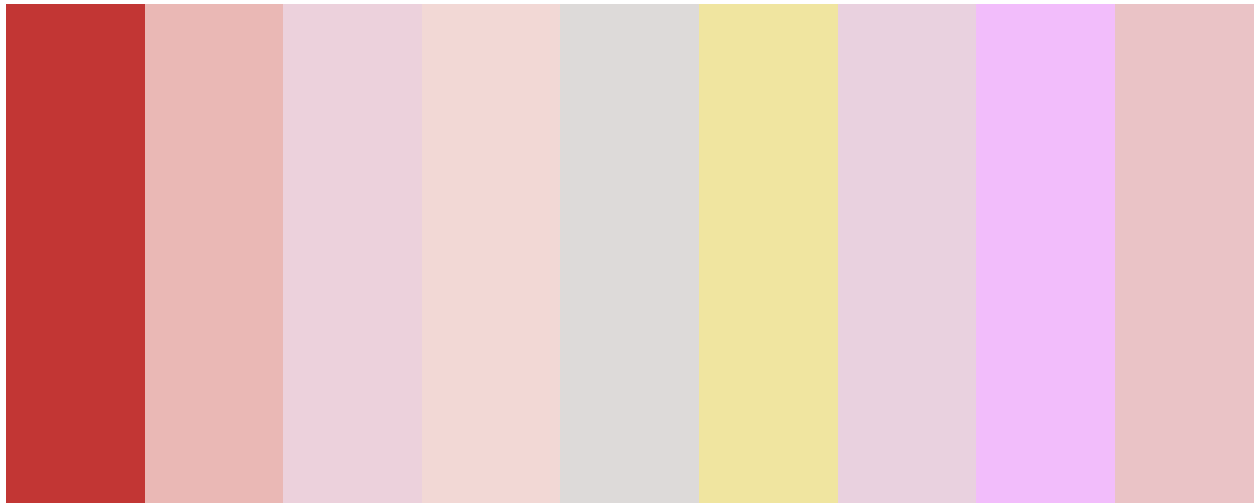
data <- read.csv("homework3_data.csv")
```

Question 1:

Yes the company should commit to redesigning their web site.

Question 2 and 3:

```
dat <- get_colors("tesolife.png")
cols <- make_palette(dat[1:100, ])
```



```
cols
```

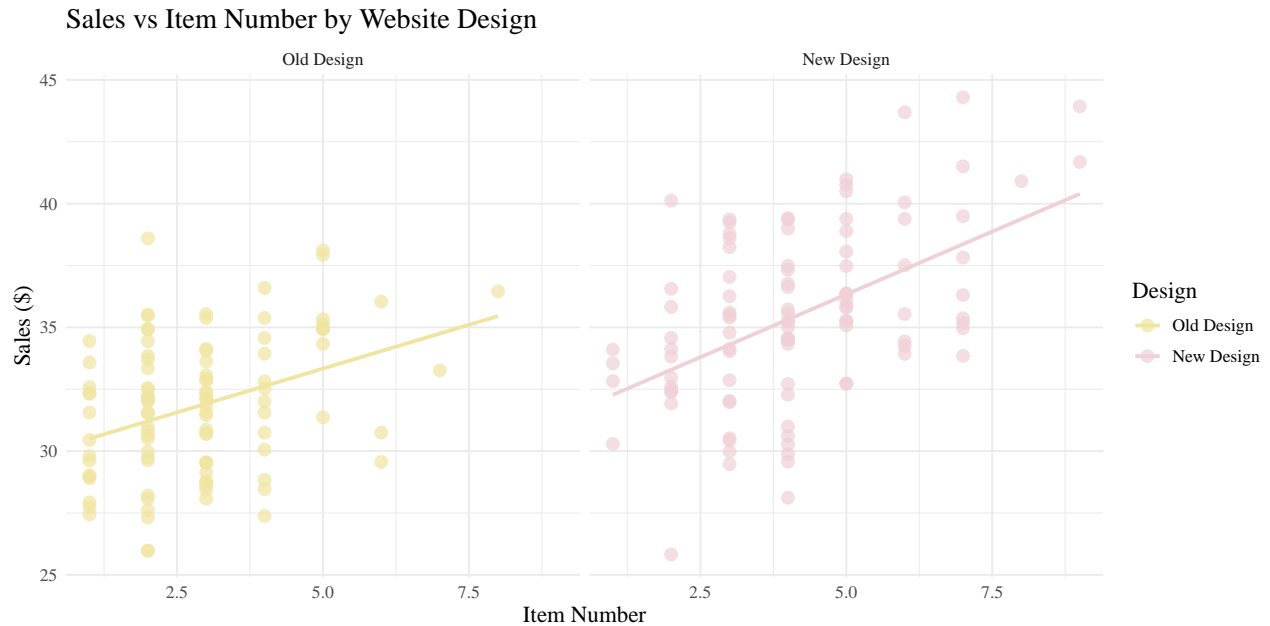
```
## [1] "#FFFFFF" "#C23634" "#EAB8B5" "#ECD1DC" "#F2D8D5" "#DDDAD9" "#F0E5A0"
## [8] "#E9D1DF" "#F2BDFB" "#EAC3C6"
```

```
# First supporting evidence
ggplot(data, aes(x = items, y = sales, color = factor(design))) +
  geom_point(alpha = 0.7, size = 3) +
  geom_smooth(method = "lm", se = FALSE) + # Line of best fit
labs(
```

```

x = "Item Number",
y = "Sales ($)",
color = "Design",
title = "Sales vs Item Number by Website Design"
) +
scale_color_manual(values = c("0" = "#F0E5A0", "1" = "#EFD0D6"),
                    labels = c("Old Design", "New Design")) +
facet_wrap(~ design, labeller = labeller(design = c("0" = "Old Design", "1" = "New Design")) +
theme_minimal() +
theme(text = element_text(size = 13, family = "serif"))

```

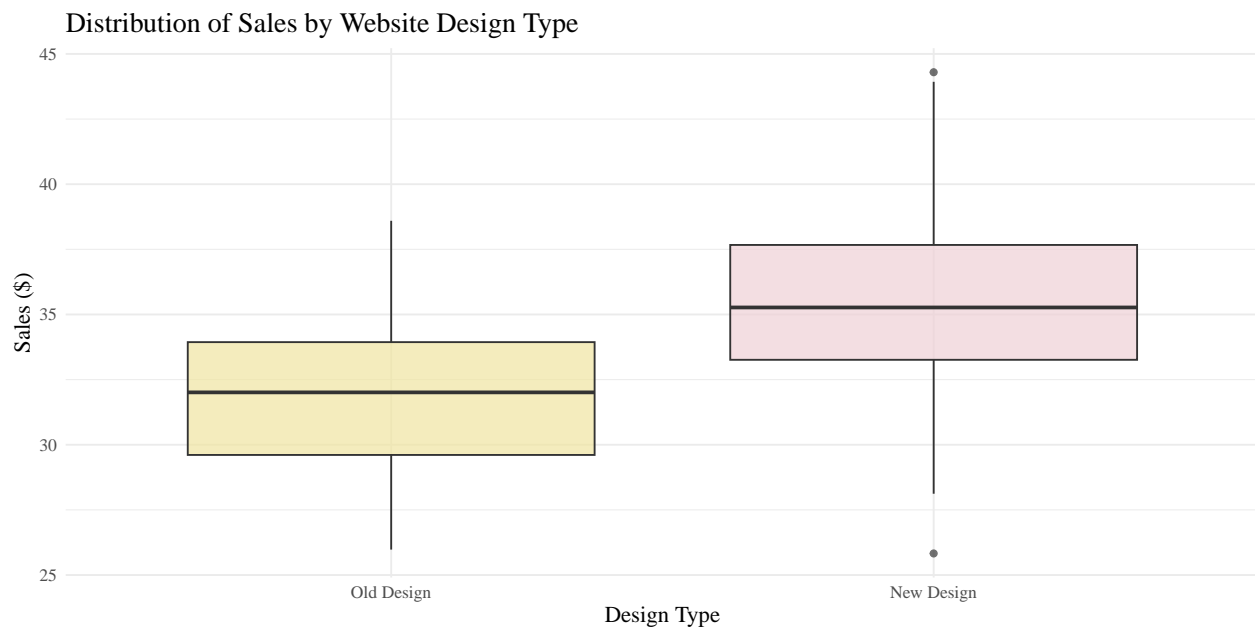


The fitted regression line is noticeably steeper for the new design, indicating that higher ratings translate into a greater increase in sales compared to the old design. Additionally, the points for the old design are clustered more heavily on the left of the x-axis, showing that it generally received less items sold.

```

# Second supporting evidence
ggplot(data, aes(x = factor(design), y = sales, fill = factor(design))) +
  geom_boxplot(alpha = 0.7) +
  scale_fill_manual(values = c("0" = "#F0E5A0", "1" = "#EFD0D6"),
                    labels = c("Old Design", "New Design")) +
  labs(
    x = "Design Type",
    y = "Sales ($)",
    title = "Distribution of Sales by Website Design Type"
  ) +
  scale_x_discrete(labels = c("0" = "Old Design", "1" = "New Design")) +
  theme_minimal() +
  theme(text = element_text(size = 13, family = "serif"),
        legend.position = "none")

```



The box plot above shows sales are higher with the new design.

```
# Third supporting evidence
meansales <- data %>%
  group_by(design) %>% summarize(MeanSales = mean(sales))
meansales
```

```
## # A tibble: 2 x 2
##   design MeanSales
##   <int>     <dbl>
## 1     0     31.8
## 2     1     35.5
```

The average amount of sales is ~\$3.66 higher for the new design (design 1).

```
# Fourth supporting evidence
set.seed(500)

boot_diff <- function(data, i) {
  d <- data[i, ]
  mean(d$sales[d$design == 1]) - mean(d$sales[d$design == 0])
}

b_res <- boot(data, boot_diff, R = 10000)

boot.ci(b_res, type = c("perc", "bca"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = b_res, type = c("perc", "bca"))
##
## Intervals :
```

Level	Percentile	BCa
95%	(2.801, 4.528)	(2.804, 4.535)

Calculations and Intervals on Original Scale

The 95% bootstrap confidence interval for the difference in mean sales between the new and old design ranges from approximately 2.80 to 4.50 per person (which is also the increase in profit) which is higher than the \$1.80 increase per customer required.

Question 4:

Alternative Statement: The company should not redesign its website because the projected increase in sales does not meet the necessary threshold.