Homework3

Selah Norton

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(colorfindr)
dataset <- read_csv('homework3_data.csv')
```

```
## Rows: 200 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (4): sales, design, items, nps
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

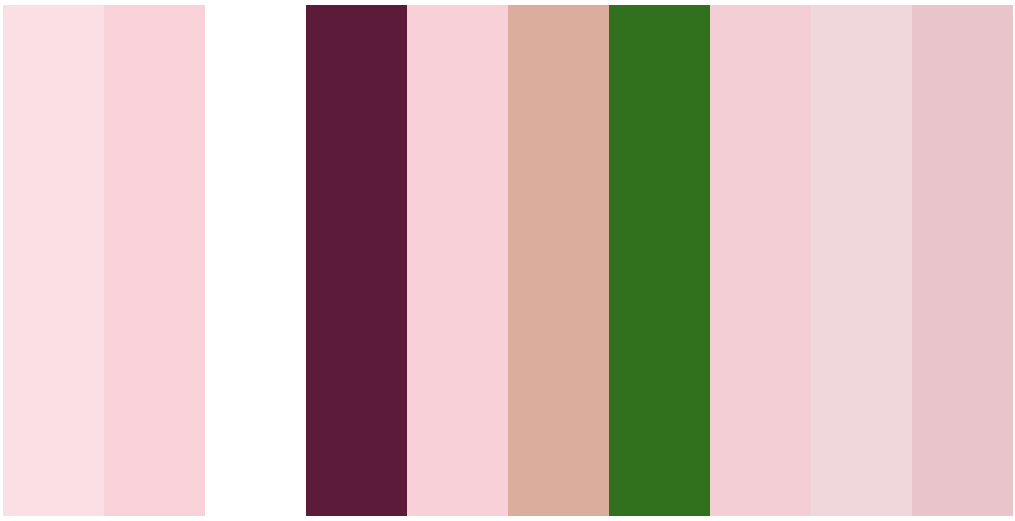
Recommendation/Primary Statement

Based on the old data when a small website redesign occurred and sales were collected, the company should commit to the full redesign as this old data supports that this will lead to an increase in sales of at least $1.80.

Supporting Evidence

The following code supports this recommendation. The box plot graph of sales comparing the old design vs new design, show a higher median in the new design, by about 2.5 dollars (1.8) and both of these box plots are symmetric. The histograms of sales in the old design and new design, show a higher mode for the new design and both of the histograms have a bell shape. When looking at the difference of average sales of the old vs new design, using the mean function, the new design has an higher average sales by $3.66 (1.8). Fitting a linear regression model with sales as the outcome and the design as a predictor, estimate a $3.66 increase (1.8) in sales with the new design on average.

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

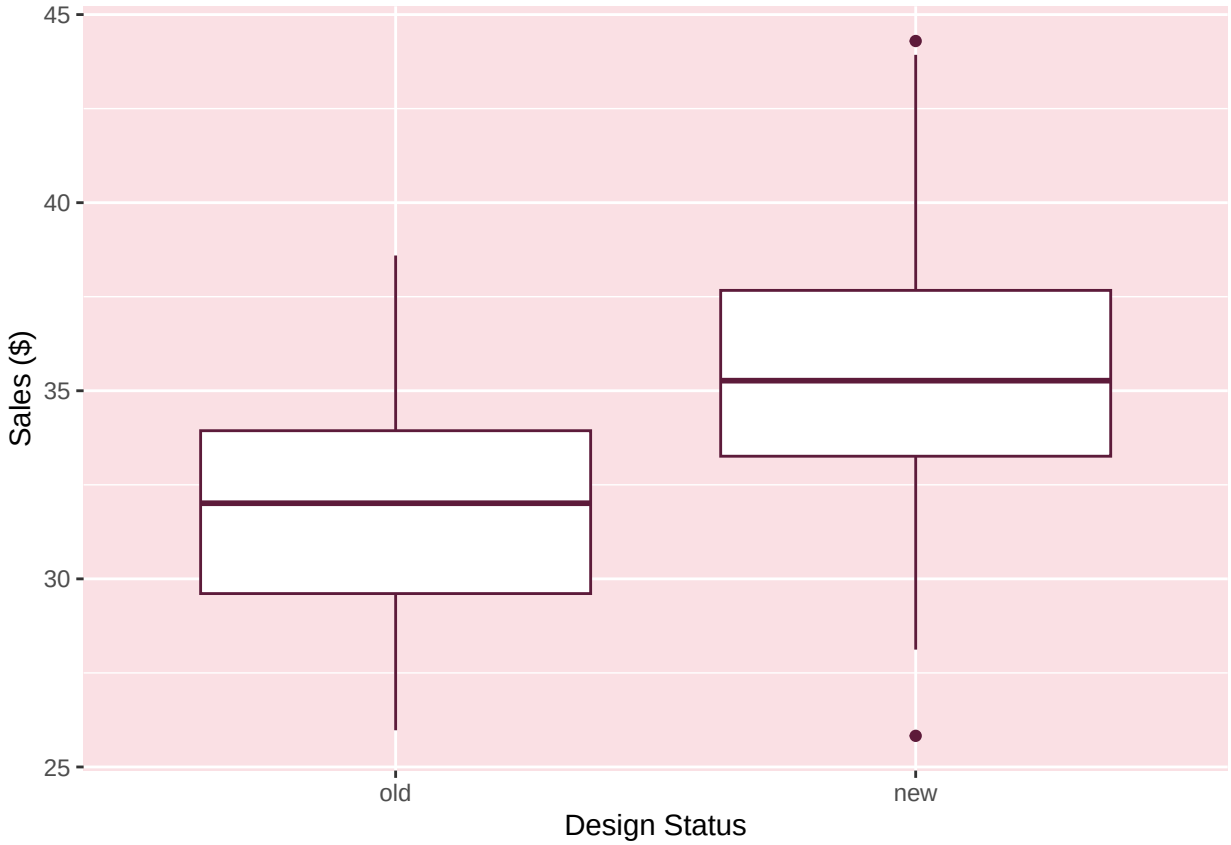


```
cols
```

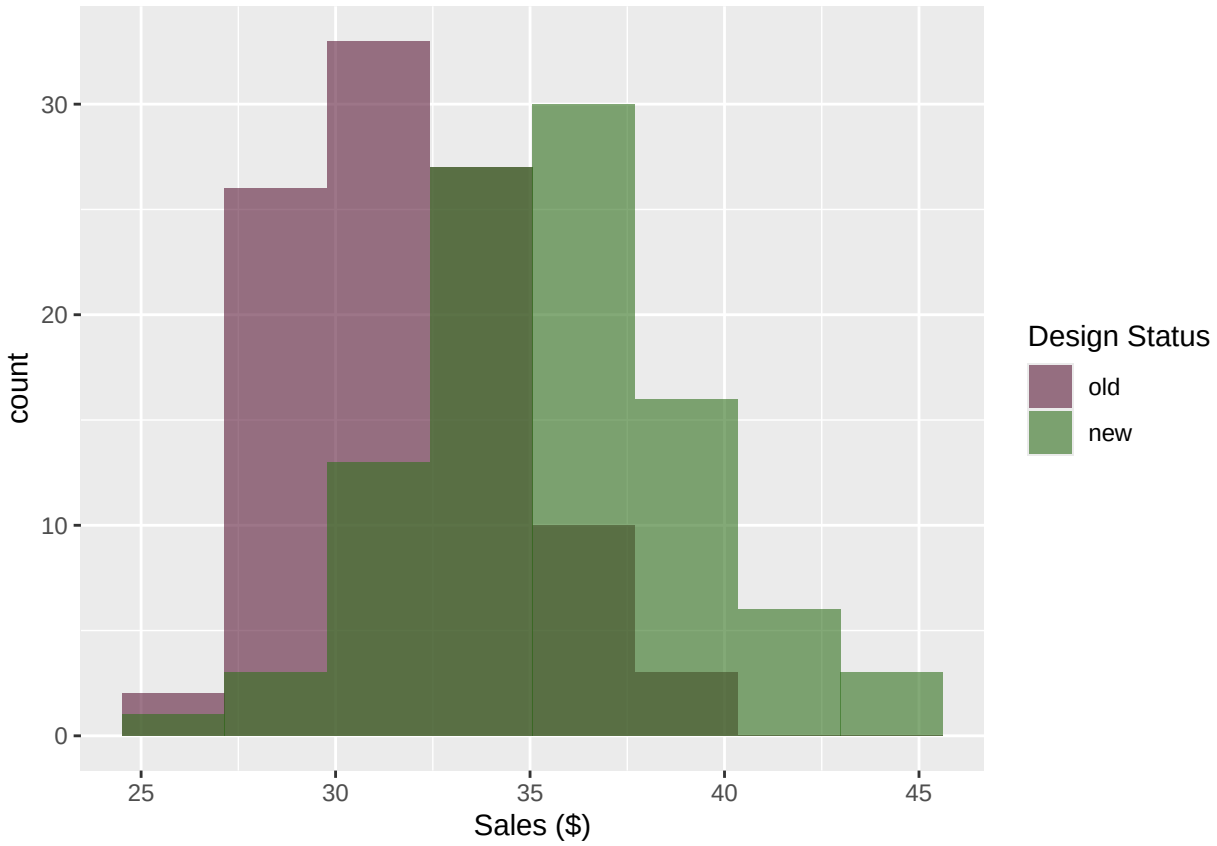
```
##  [1] "#FAE0E4" "#F8D2D8" "#FFFFFF" "#5D1B3A" "#F7D1D7" "#DAAD9D" "#31711D"
##  [8] "#F3CED4" "#F0D7DB" "#E9C5CB"
```

```
p1 <- dataset %>%
  # Make design into a factor
  ggplot(aes(x=as.factor(design), y=sales)) +
  geom_boxplot(color=c('#5D1B3A')) +
  # Label x and y axis
  scale_x_discrete(name='Design Status',
                   labels= c('old', 'new')) +
  scale_y_continuous(name='Sales ($)') +
  # Add background color
  theme(panel.background = element_rect(fill = '#FAE0E4'))

p1
```




```
p2 <- dataset %>%
  # Ensure x/sales is numeric
  ggplot(aes(x=as.numeric(sales),fill=factor(design))) +
  geom_histogram(bins=8, alpha=0.6, position='identity') +
  # Label x axis
  scale_x_continuous(name='Sales ($)') +
  # Add color to plot and edit legend
  scale_fill_manual(values=c('#5D1B3A','#31711D'),
                    name = 'Design Status',
                    labels = c('0'='old', '1'='new'))
p2
```




```
# Compute means
old_sales <- mean(dataset$sales[dataset$design == 0])
new_sales <- mean(dataset$sales[dataset$design == 1])
# Compare
diff_sales <- new_sales - old_sales
diff_sales
```

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

```
# Create a linear regression model
model <- lm(sales ~ design, data=dataset)
# Extract coefficients
coef(model)
```

```
## (Intercept)      design
##   31.848190    3.664904
```

Alternative Statement and Analysis

The redesign will decrease sales even though the data suggests that it would increase sales. This could be because the data used from the original redesign was a smaller scale, it does not necessarily imply that a major redesign would be received the same way. It could also be due to the fact that customers simply liked the first redesign and are opposed to this particular redesign because of personal preferences. Maybe this data from the original redesign was taken from a time where they could expect an increase in sales, such as November or December, when people do holiday gift shopping, making it seem like in the old data the redesign increased sales but it was actually due to this confounder. Along the same line, this data could have been taken when a limited time product was available, another potential confounding variable.