Final project

Group4

2024-06-03

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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 errors
```

```
library(infer)
library(dplyr)
library(ggplot2)
```

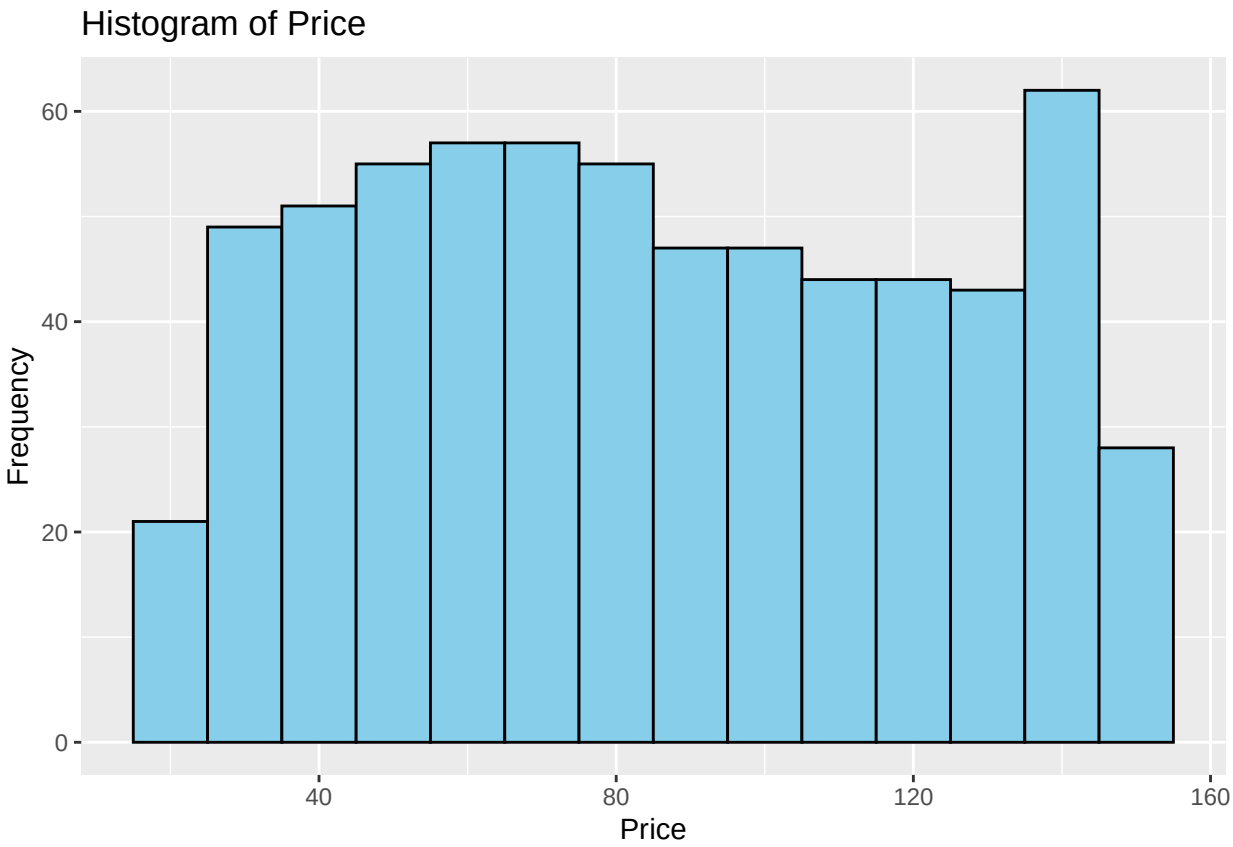
```
fashion <- readxl::read_excel("fashion_data_2018_2022.xls")
```

Numerical Variables

price - ChaeJun Lim

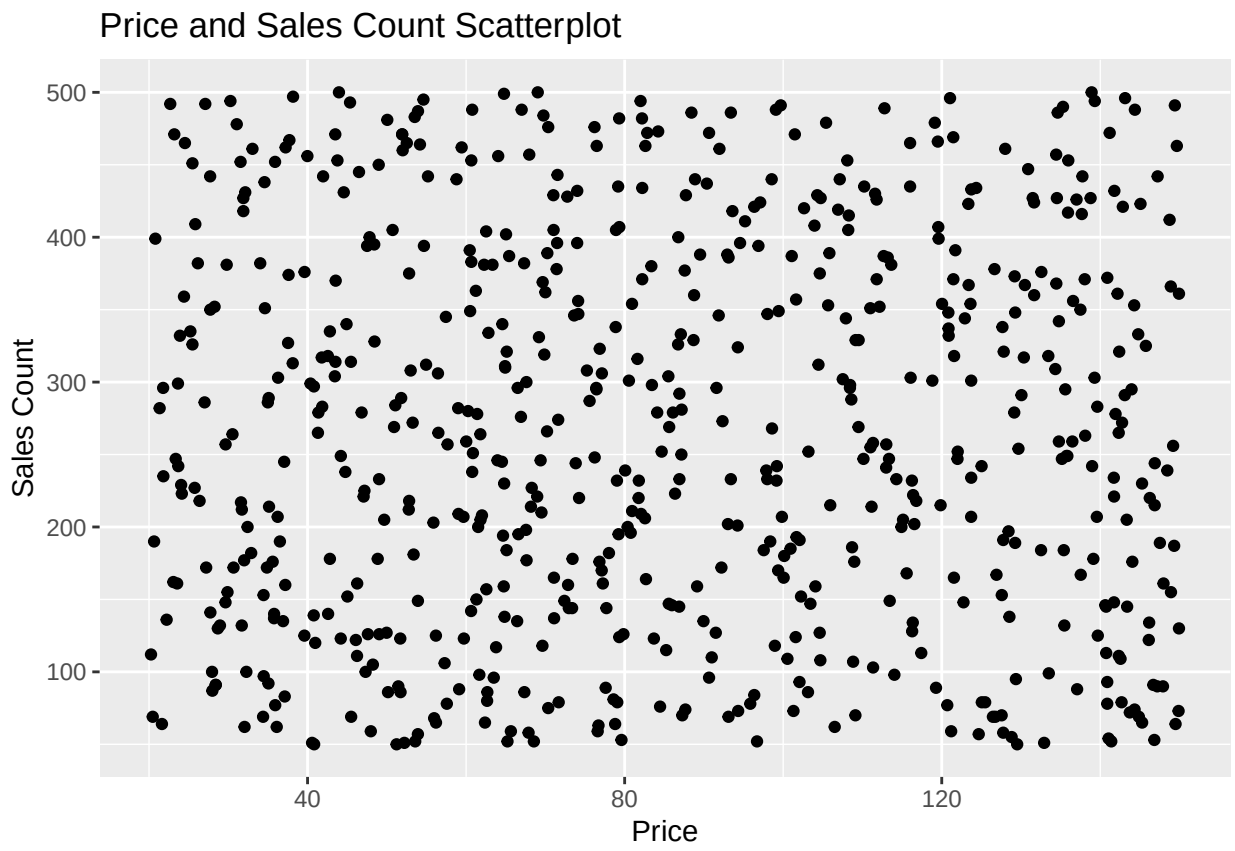
Histogram

```
fashion %>%
  ggplot()+
  geom_histogram(mapping = aes(x=price), binwidth=10, color="black",
                 fill="skyblue")+
  labs(
  title = "Histogram of Price",
  x = "Price",
  y = "Frequency"
  )
```



Scatterplot

```
fashion %>%
  ggplot()+
  geom_point(mapping = aes(x=price, y=sales_count))+
  labs(
  title = "Price and Sales Count Scatterplot",
  x = "Price",
  y = "Sales Count"
  )
```



reviews_count - ChaeJun Lim

Histogram

```
fashion %>%
  ggplot()+
  geom_histogram(mapping = aes(x=reviews_count), binwidth=20,
                 color = "black", fill = "skyblue")+
  labs(
  title = "Histogram of Reviews Count",
  x = "Reviews Count",
  y = "Frequency"
  )
```

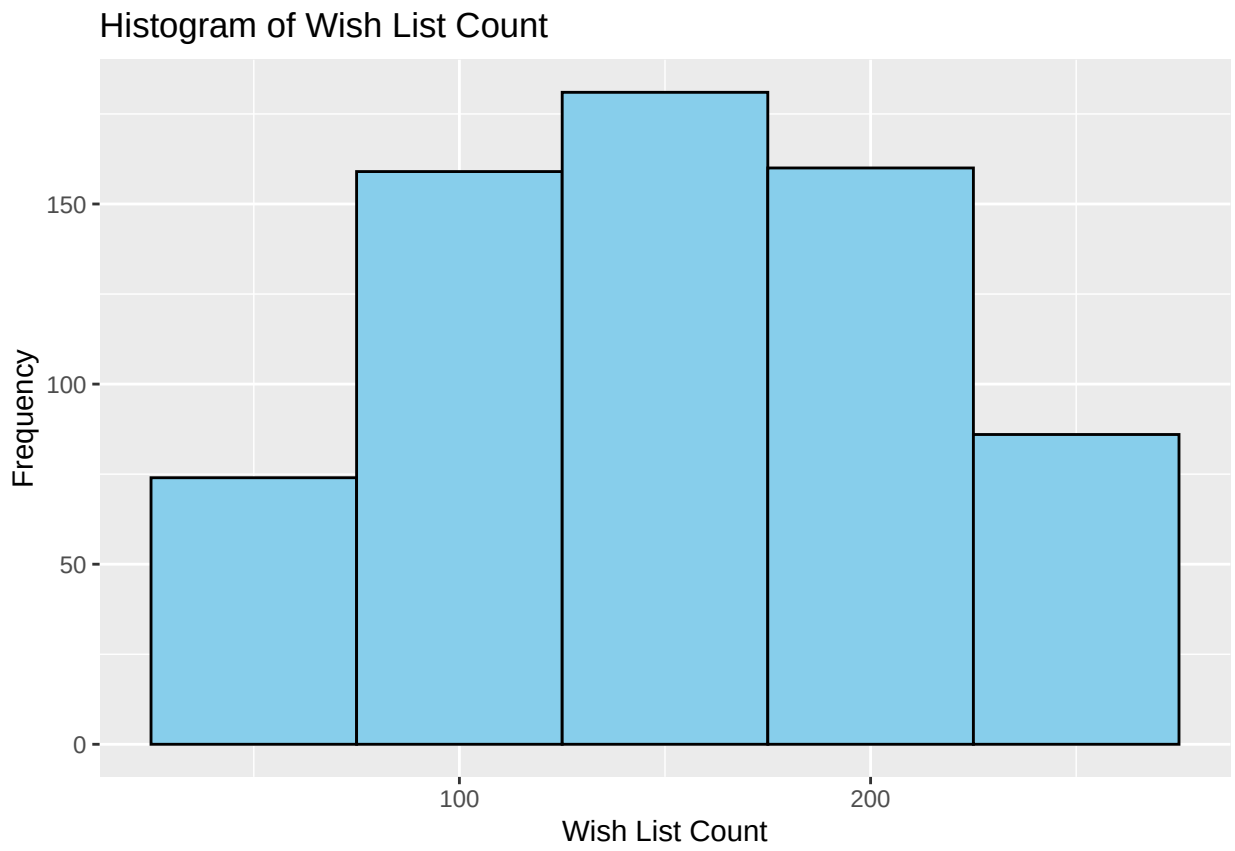
Scatterplot

```
fashion %>%
  ggplot()+
  geom_point(mapping = aes(x=reviews_count, y=sales_count))+
  labs(
  title = "Reviews Count and Sales Count Scatterplot",
  x = "Reviews Count",
  y = "Sales Count"
  )
```

wish_list_count - Jin Heo

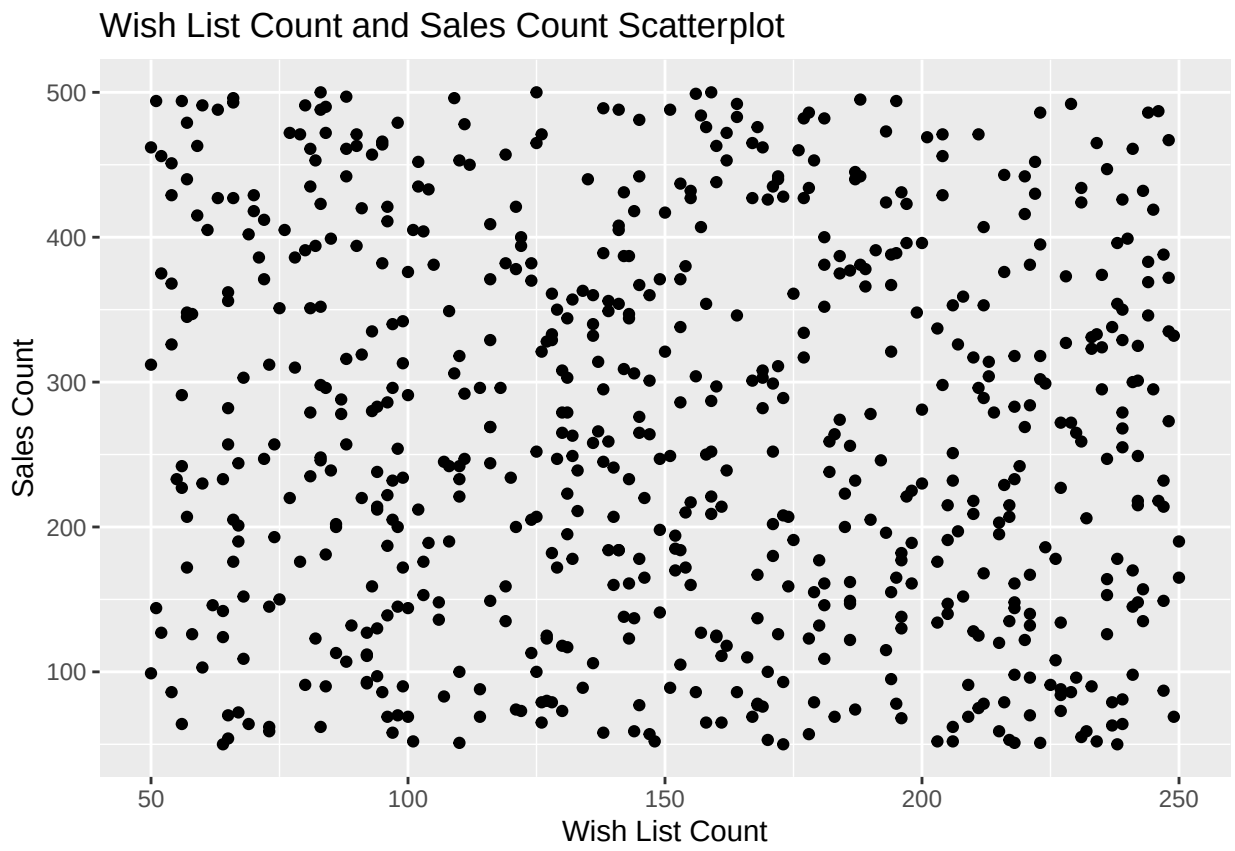
Histogram

```
fashion %>%
  ggplot()+
  geom_histogram(mapping = aes(x=wish_list_count), binwidth=50,
                 color="black", fill="skyblue")+
  labs(
  title = "Histogram of Wish List Count",
  x = "Wish List Count",
  y = "Frequency"
  )
```



Scatterplot

```
fashion %>%
  ggplot()+
  geom_point(mapping = aes(x=wish_list_count, y=sales_count))+
  labs(
  title = "Wish List Count and Sales Count Scatterplot",
  x = "Wish List Count",
  y = "Sales Count"
  )
```



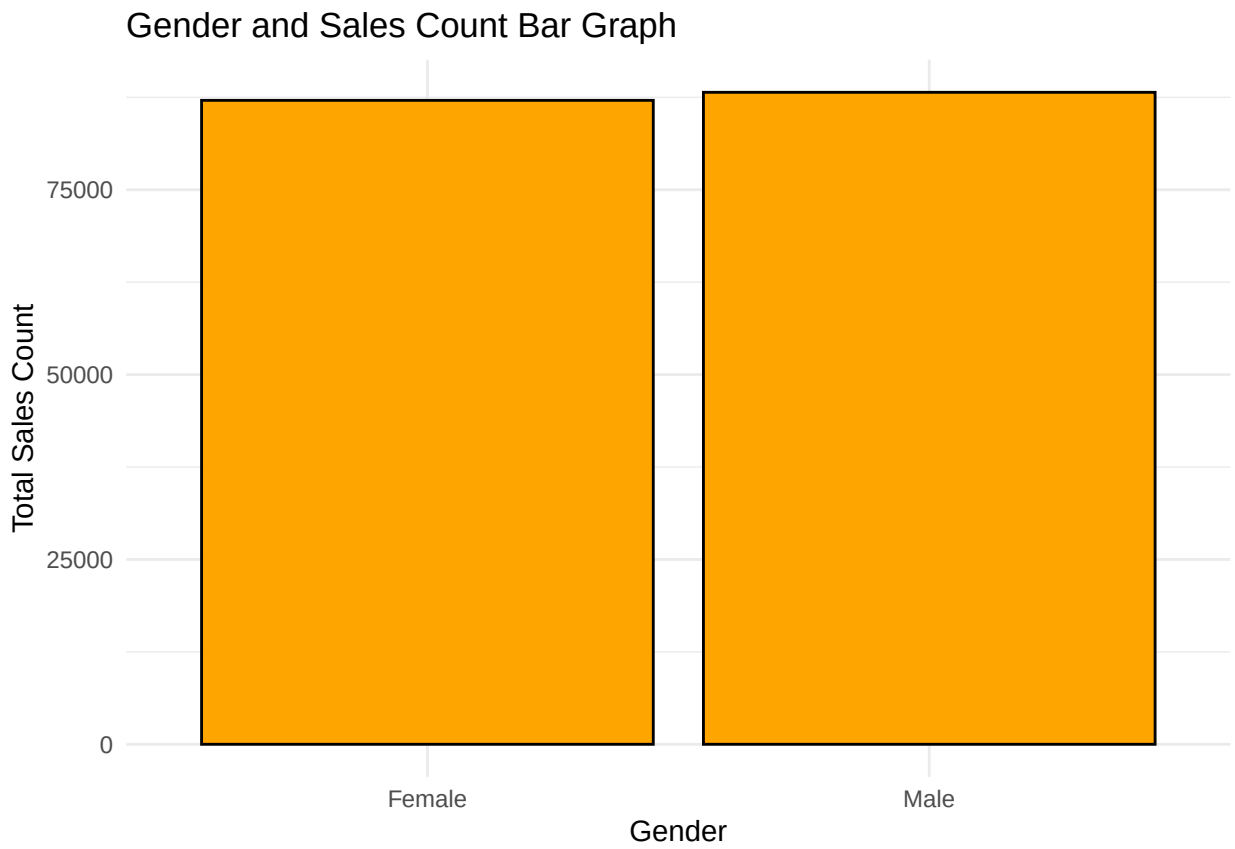
Categorical Variables Visualization

gender - Jin Heo

Bar graph

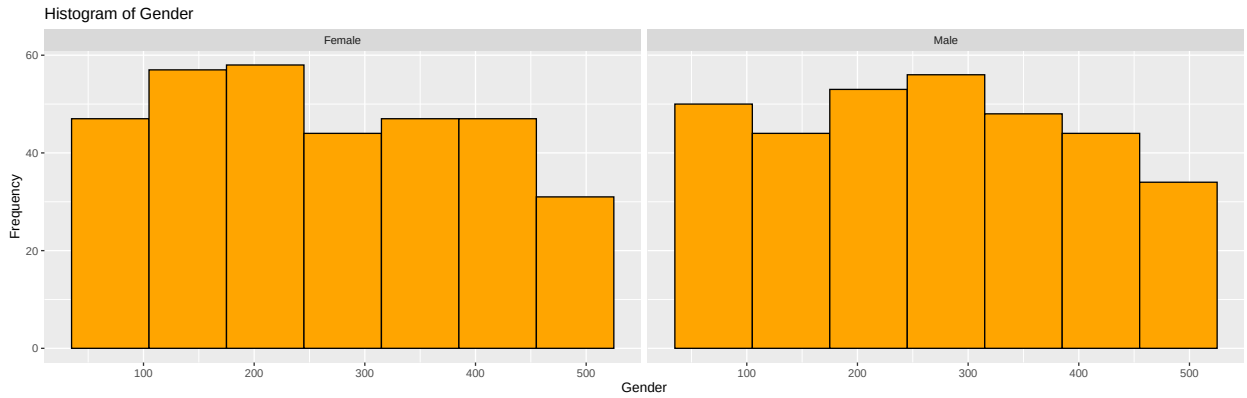
```
gender_sales <- fashion %>%
  group_by(gender) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(gender_sales) +
  geom_col(mapping = aes(x = gender, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Gender and Sales Count Bar Graph",
    x = "Gender",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



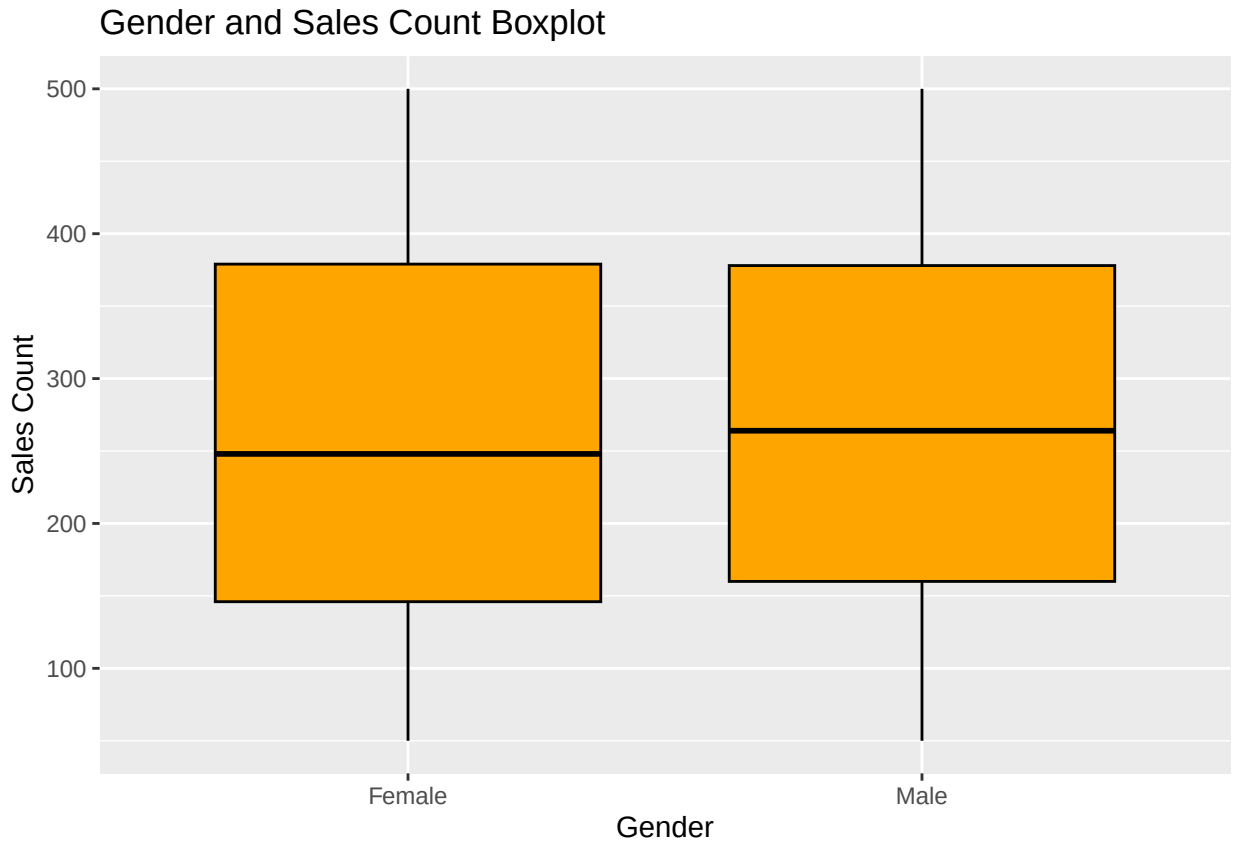
Histogram

```
fashion %>%
  ggplot()+
  geom_histogram(mapping = aes(x= sales_count), binwidth=70, color = "black",
                 fill = "orange") + facet_wrap(~ gender) +
  labs(
  title = "Histogram of Gender",
  x = "Gender",
  y = "Frequency"
  )
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(gender, sales_count, FUN=median)),
               color = "black", fill = "orange")+
  labs(
  title = "Gender and Sales Count Boxplot",
  x = "Gender",
  y = "Sales Count"
  )
```



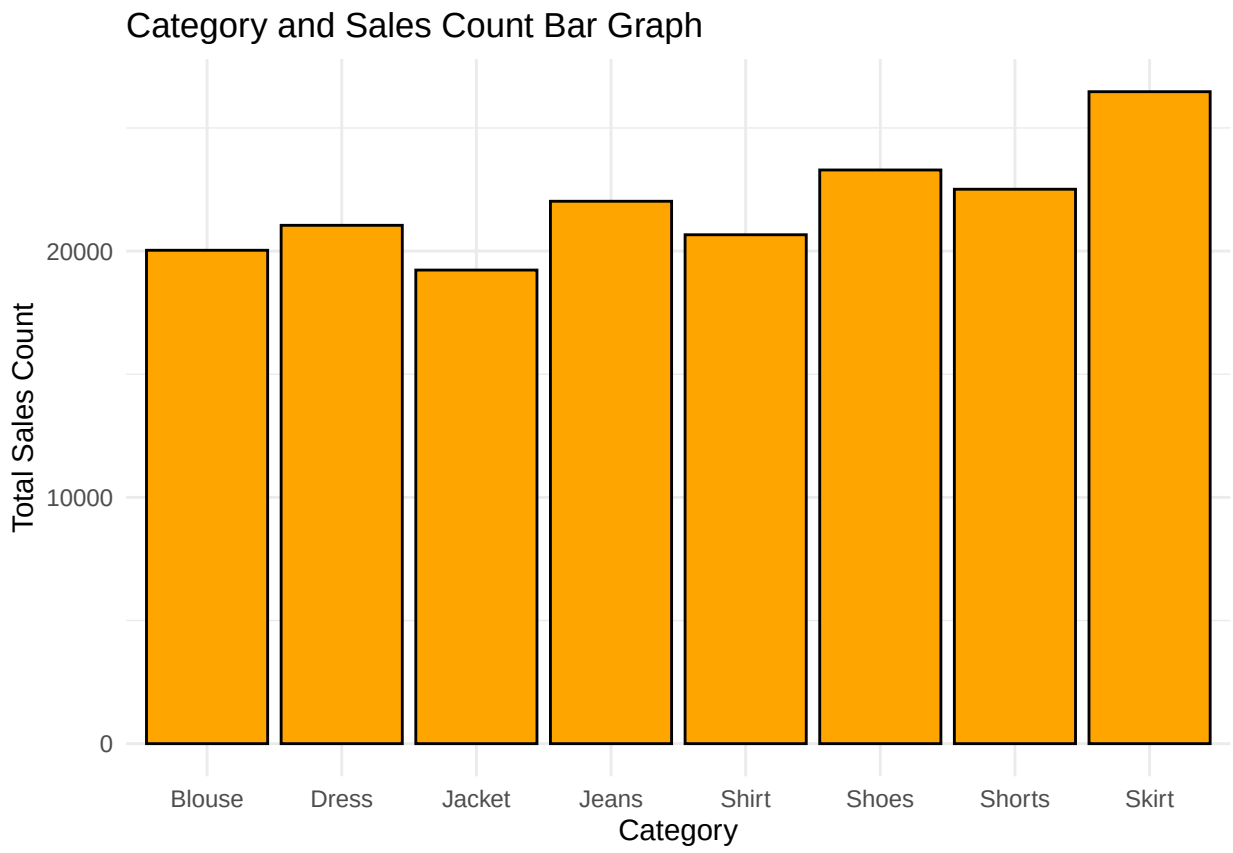
category - Minseo Yoon

Bar graph

```
category_sales <- fashion %>%
  group_by(category) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

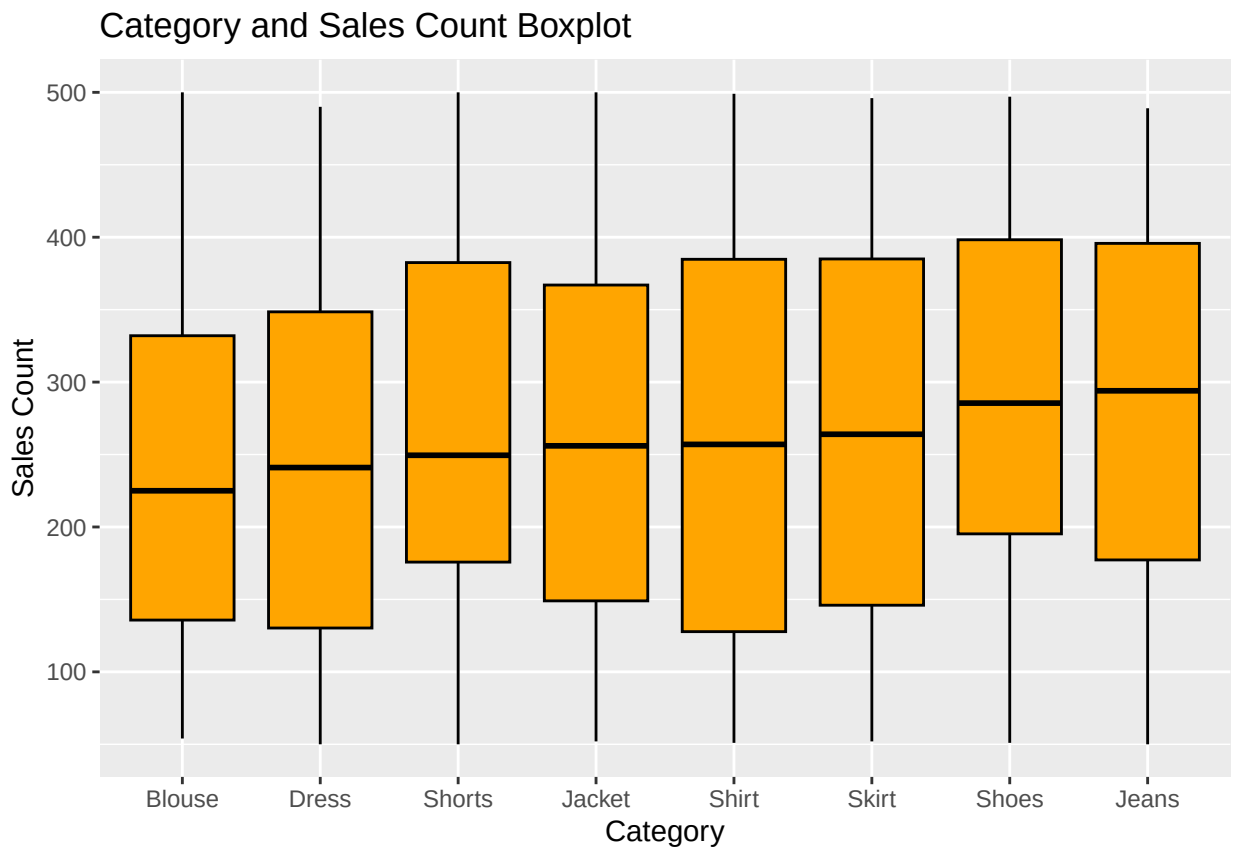
```
ggplot(category_sales) +
  geom_col(mapping = aes(x = category, y = total_sales_count),
```

```
          color = "black", fill = "orange") +
  labs(
    title = "Category and Sales Count Bar Graph",
    x = "Category",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(category, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Category and Sales Count Boxplot",
  x = "Category",
  y = "Sales Count"
  )
```

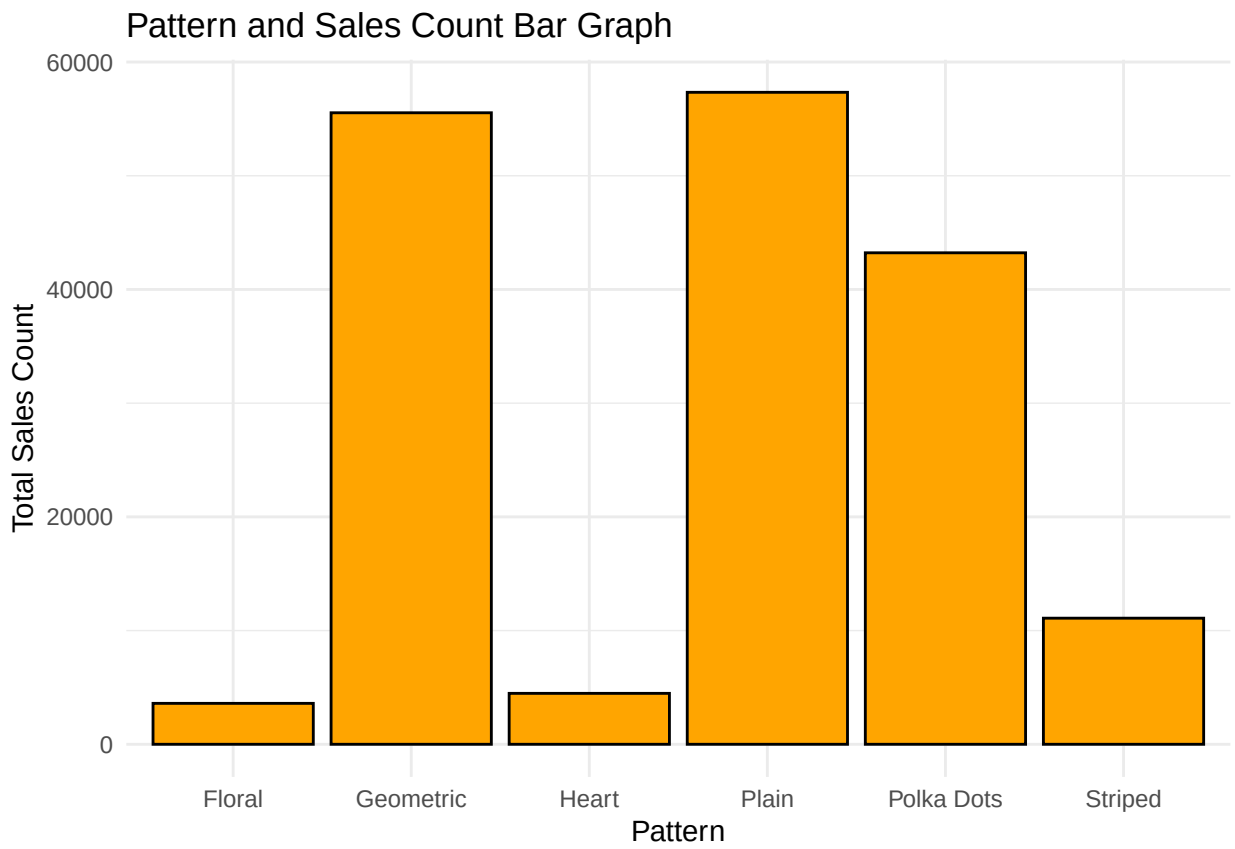


pattern - Minseo Yoon

Bar graph

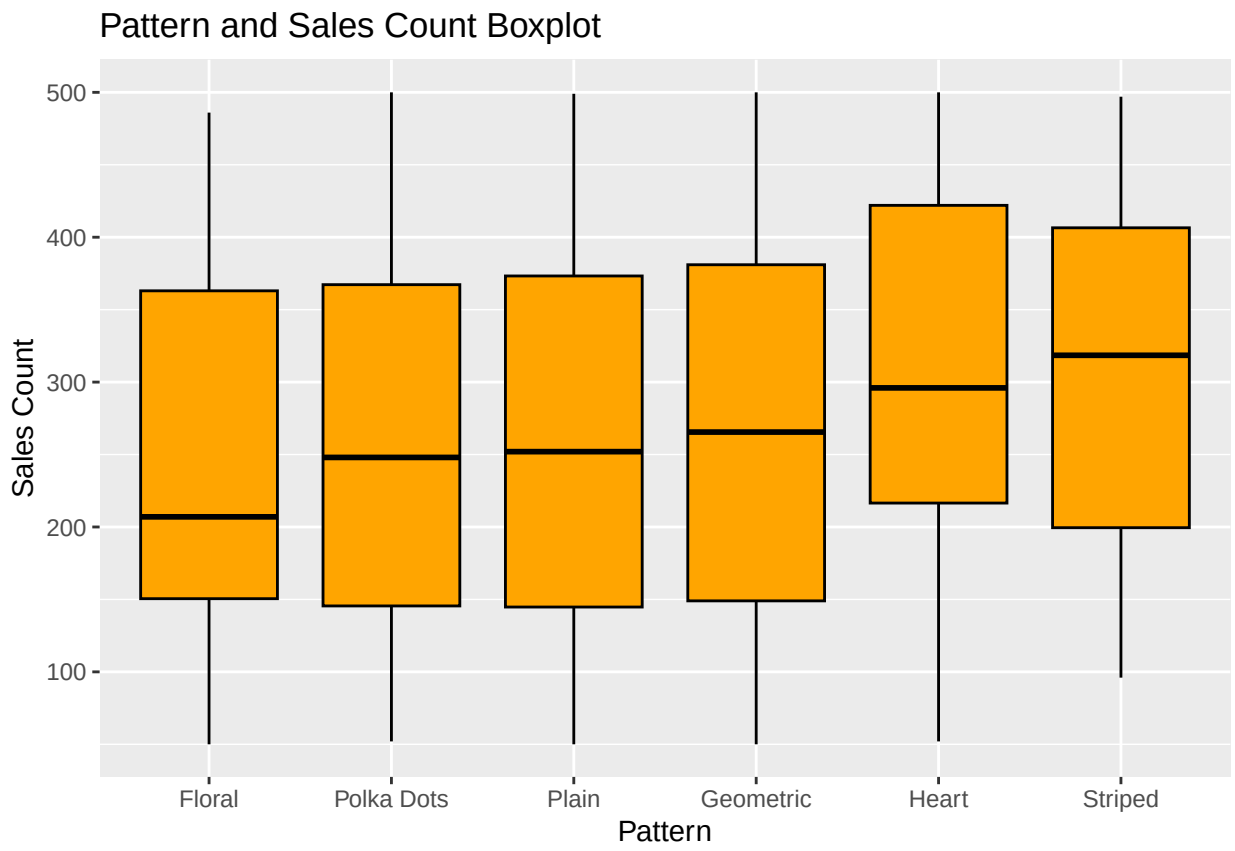
```
pattern_sales <- fashion %>%
  group_by(pattern) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(pattern_sales) +
  geom_col(mapping = aes(x = pattern, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Pattern and Sales Count Bar Graph",
    x = "Pattern",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(pattern, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Pattern and Sales Count Boxplot",
  x = "Pattern",
  y = "Sales Count"
  )
```

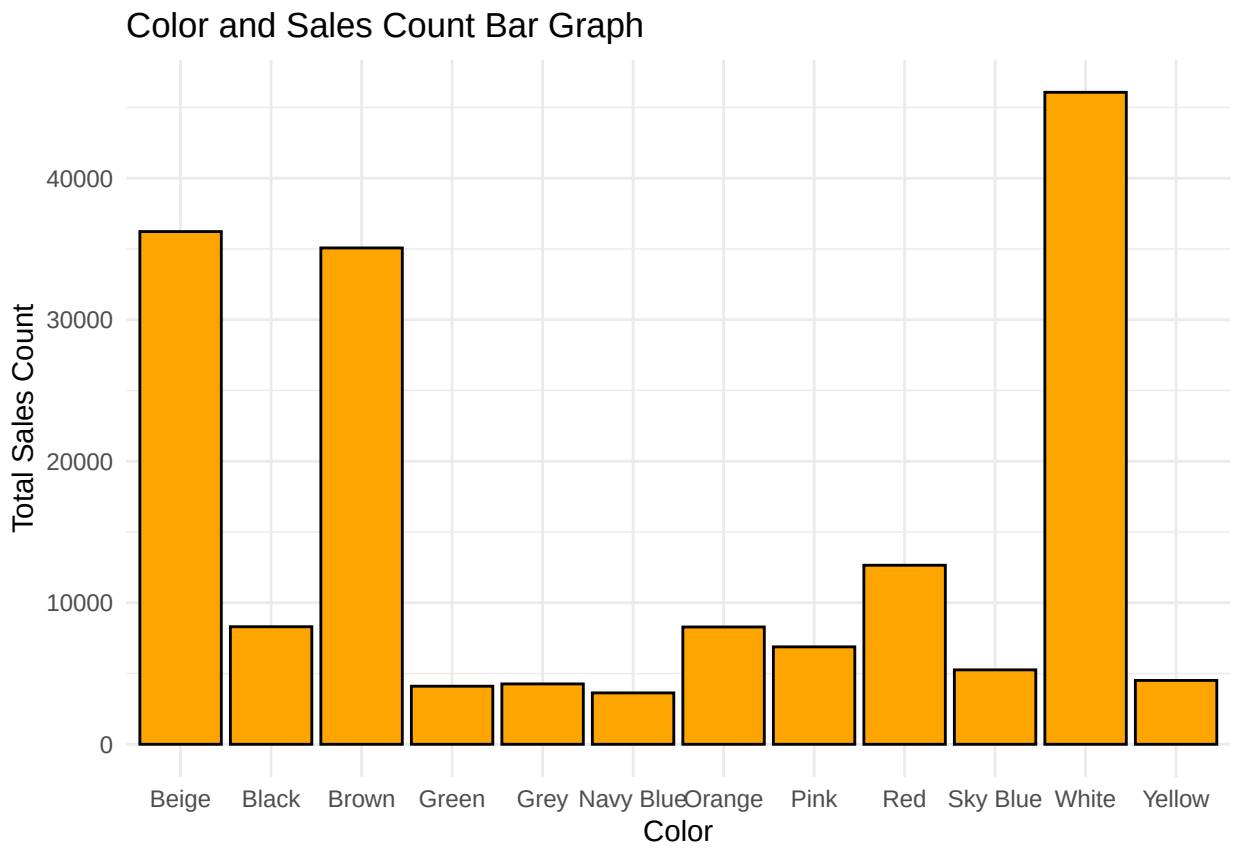


color - Seunghyun Ryu

Bar graph

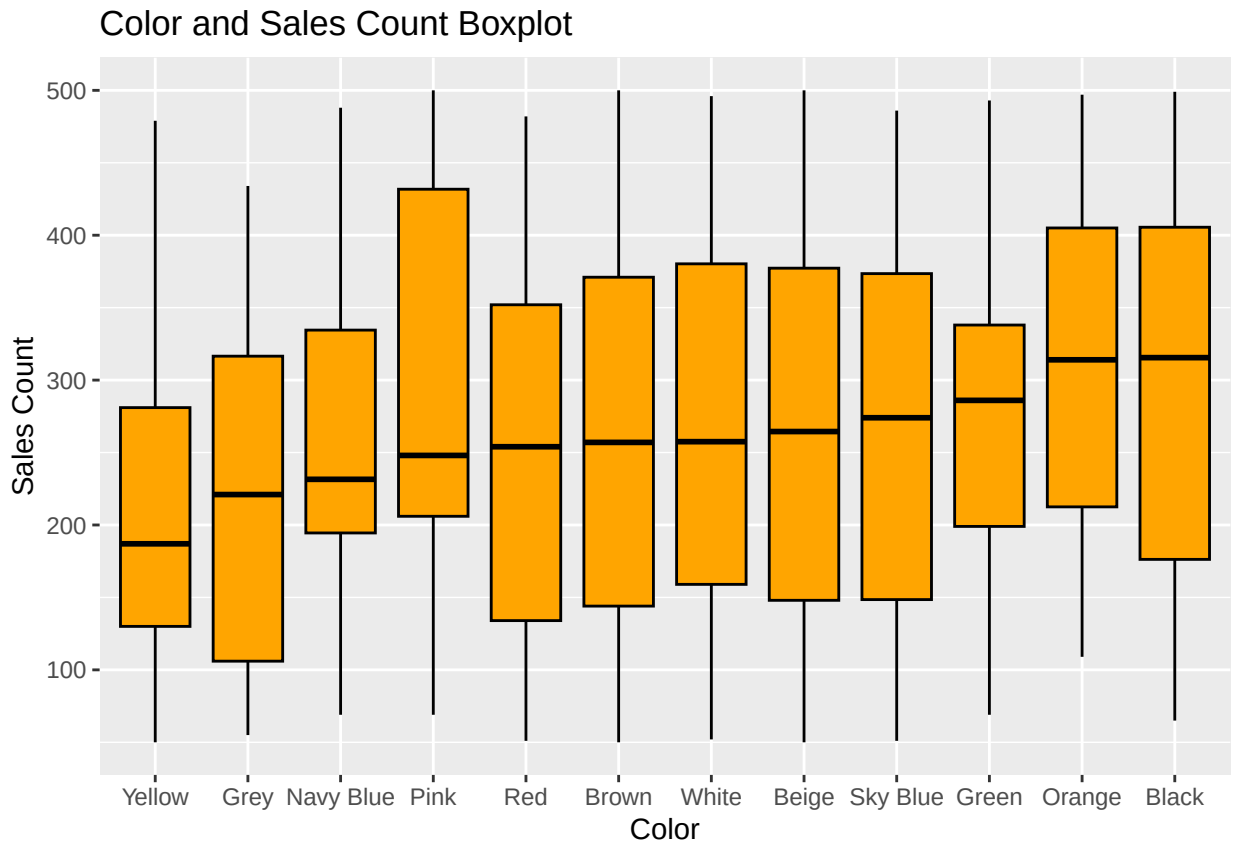
```
color_sales <- fashion %>%
  group_by(color) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(color_sales) +
  geom_col(mapping = aes(x = color, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Color and Sales Count Bar Graph",
    x = "Color",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(color, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Color and Sales Count Boxplot",
  x = "Color",
  y = "Sales Count"
  )
```

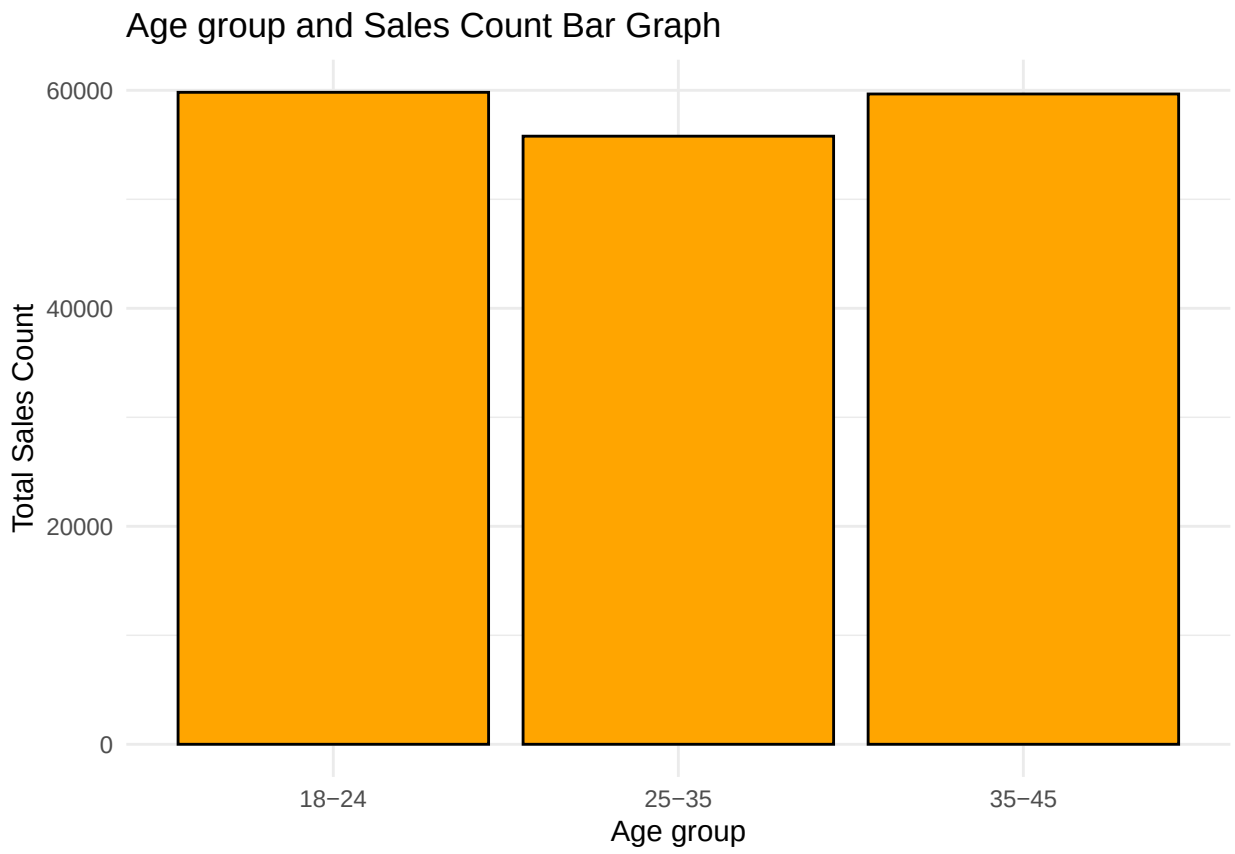


age group - Seunghyun Ryu

Bar graph

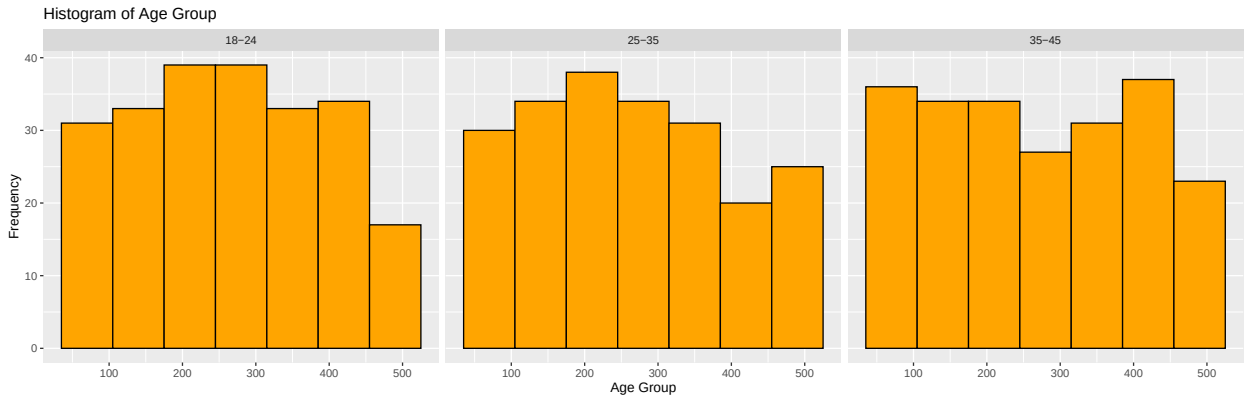
```
age_group_sales <- fashion %>%
  group_by(age_group) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(age_group_sales) +
  geom_col(mapping = aes(x = age_group, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Age group and Sales Count Bar Graph",
    x = "Age group",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



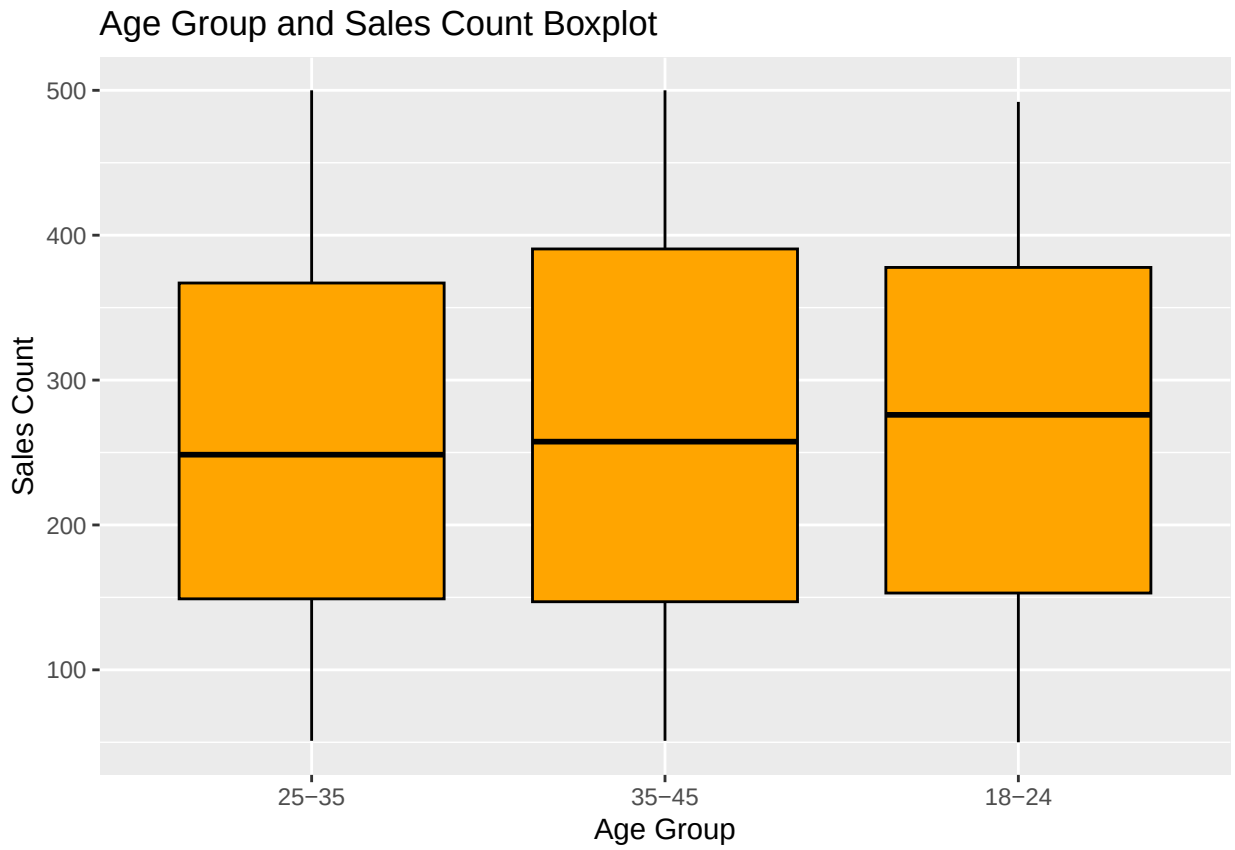
Histogram

```
fashion %>%
  ggplot()+
  geom_histogram(mapping = aes(x= sales_count), binwidth=70, color = "black",
                 fill = "orange") + facet_wrap(~ age_group) +
  labs(
  title = "Histogram of Age Group",
  x = "Age Group",
  y = "Frequency"
  )
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x= reorder(age_group, sales_count,
                                             FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Age Group and Sales Count Boxplot",
  x = "Age Group",
  y = "Sales Count"
  )
```

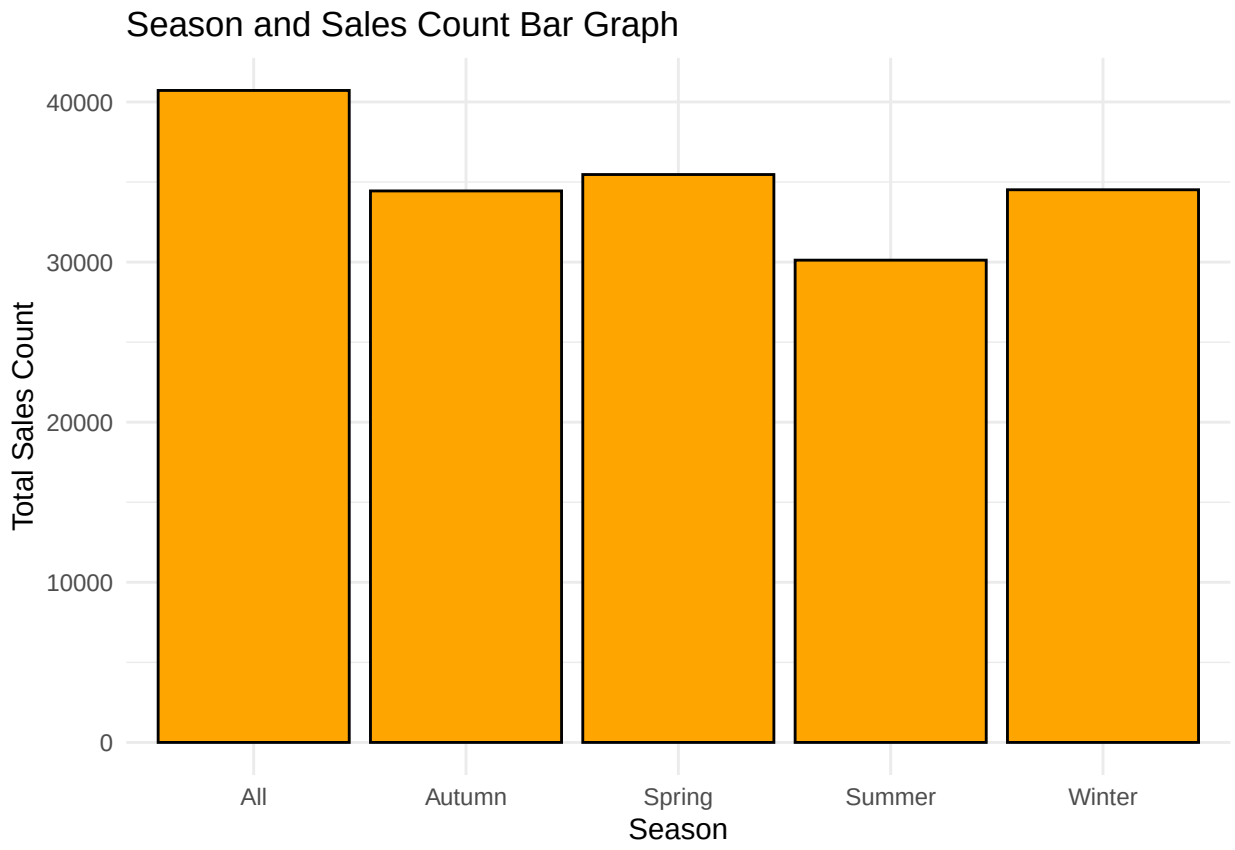


season - Juheon Kim

Bar graph

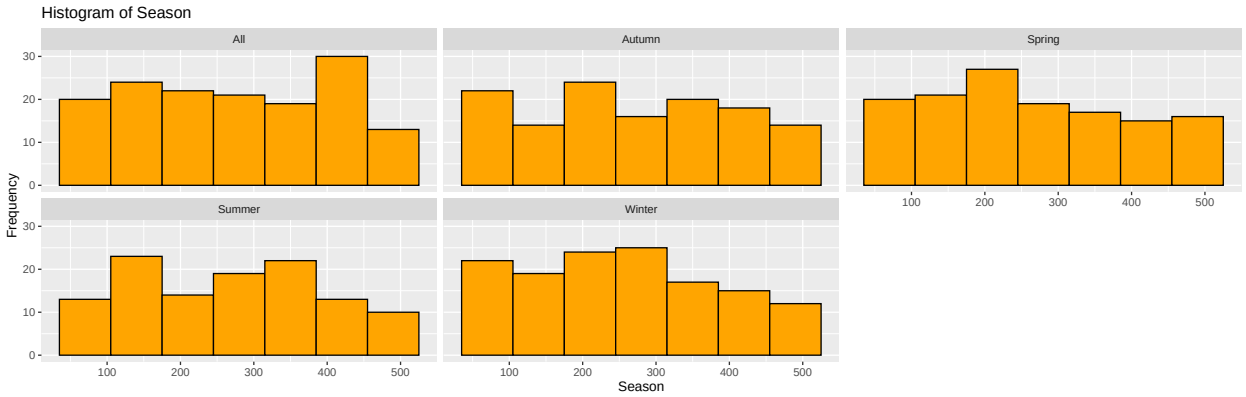
```
season_sales <- fashion %>%
  group_by(season) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```r
ggplot(season_sales) +
  geom_col(mapping = aes(x = season, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Season and Sales Count Bar Graph",
    x = "Season",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



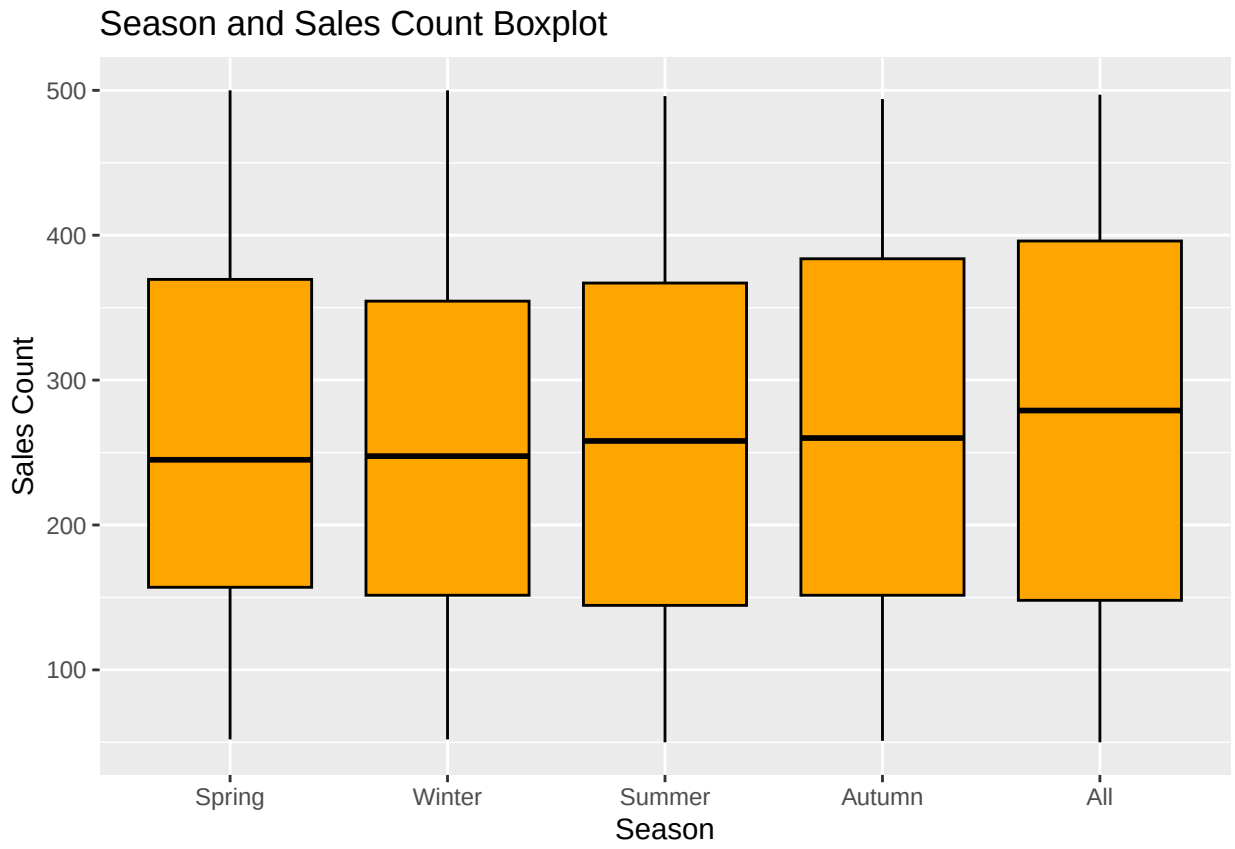
Histogram

```r
fashion %>%
  ggplot()+
  geom_histogram(mapping = aes(x= sales_count), binwidth=70, color = "black",
                 fill = "orange") + facet_wrap(~ season) +
  labs(
  title = "Histogram of Season",
  x = "Season",
  y = "Frequency"
  )
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(season, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Season and Sales Count Boxplot",
  x = "Season",
  y = "Sales Count"
  )
```

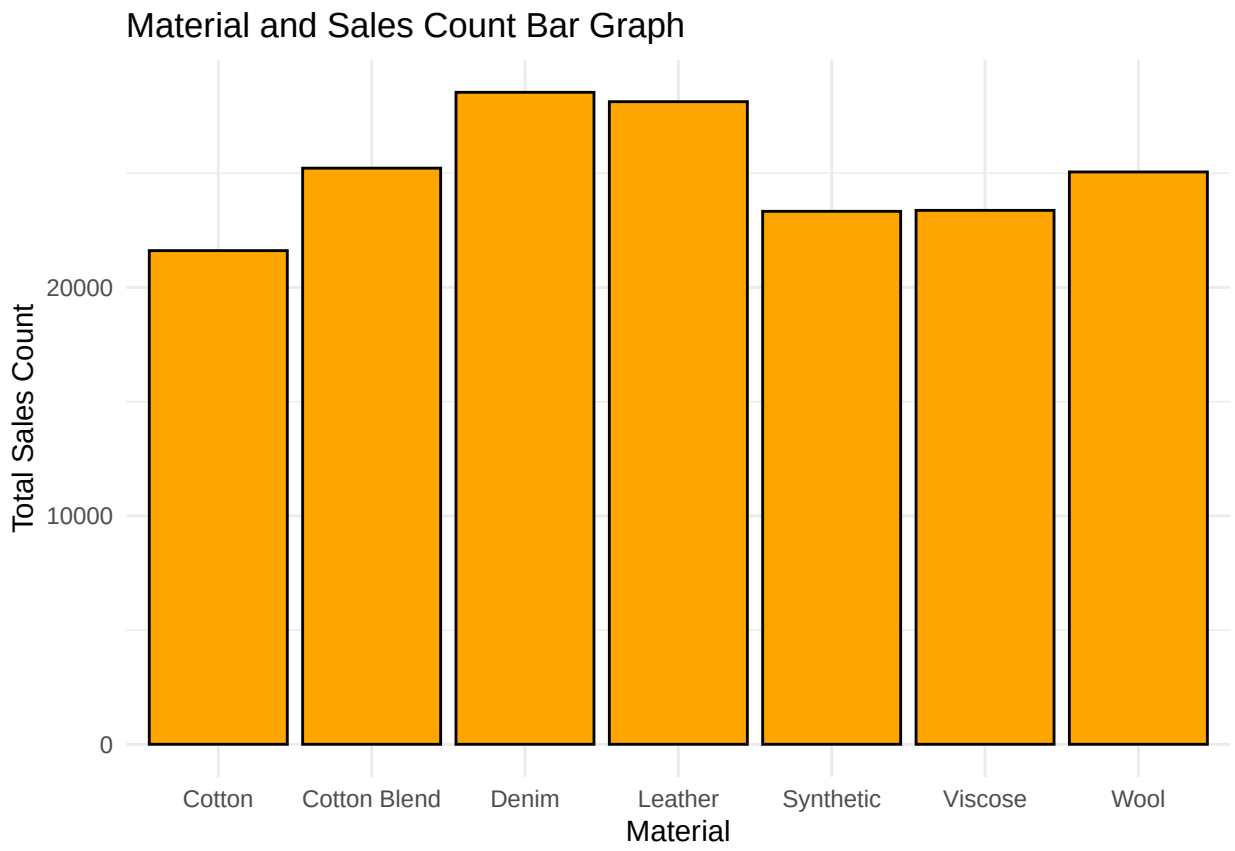


material - Juheon Kim

Bar graph

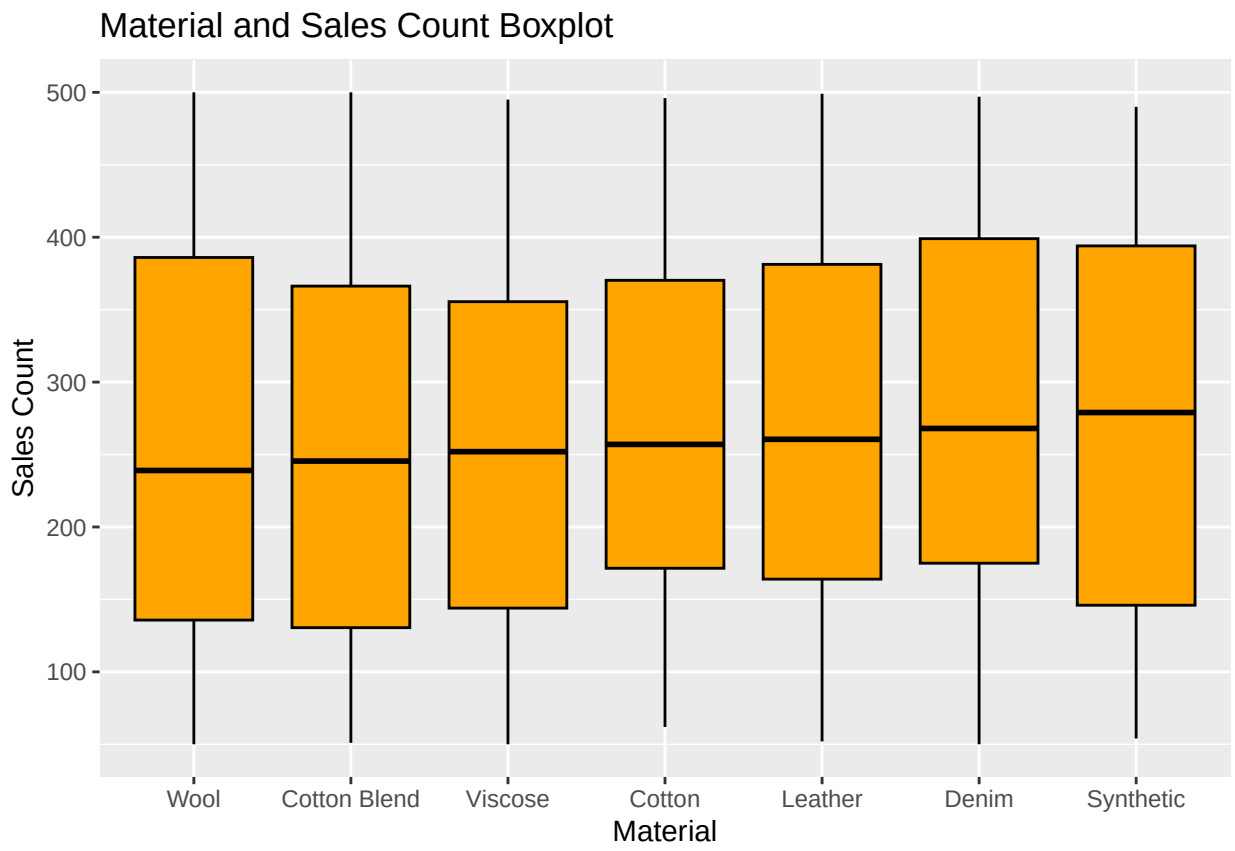
```
material_sales <- fashion %>%
  group_by(material) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(material_sales) +
  geom_col(mapping = aes(x = material, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Material and Sales Count Bar Graph",
    x = "Material",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(material, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Material and Sales Count Boxplot",
  x = "Material",
  y = "Sales Count"
  )
```

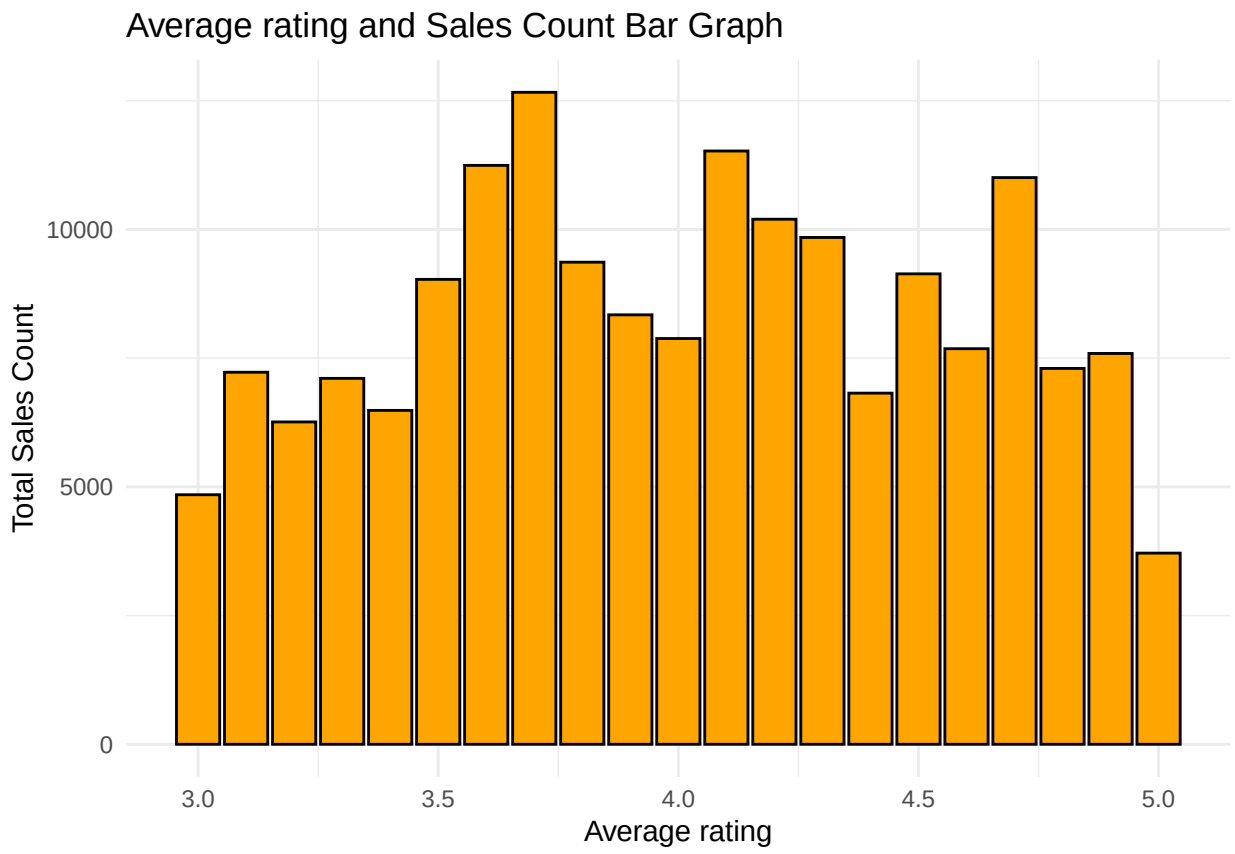


average rating - Hyukjoon Choi

Bar graph

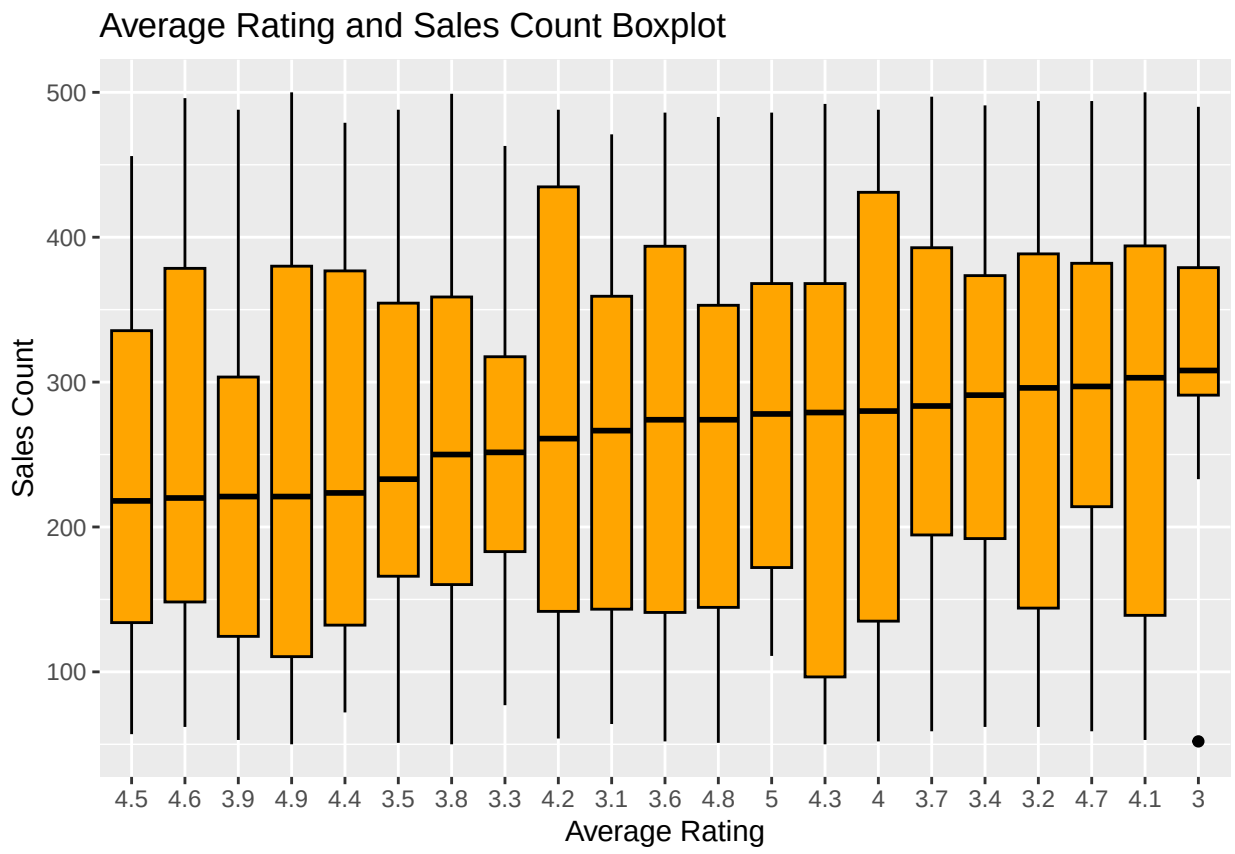
```
average_rating_sales <- fashion %>%
  group_by(average_rating) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(average_rating_sales) +
  geom_col(mapping = aes(x = average_rating, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Average rating and Sales Count Bar Graph",
    x = "Average rating",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(average_rating, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Average Rating and Sales Count Boxplot",
  x = "Average Rating",
  y = "Sales Count"
  )
```

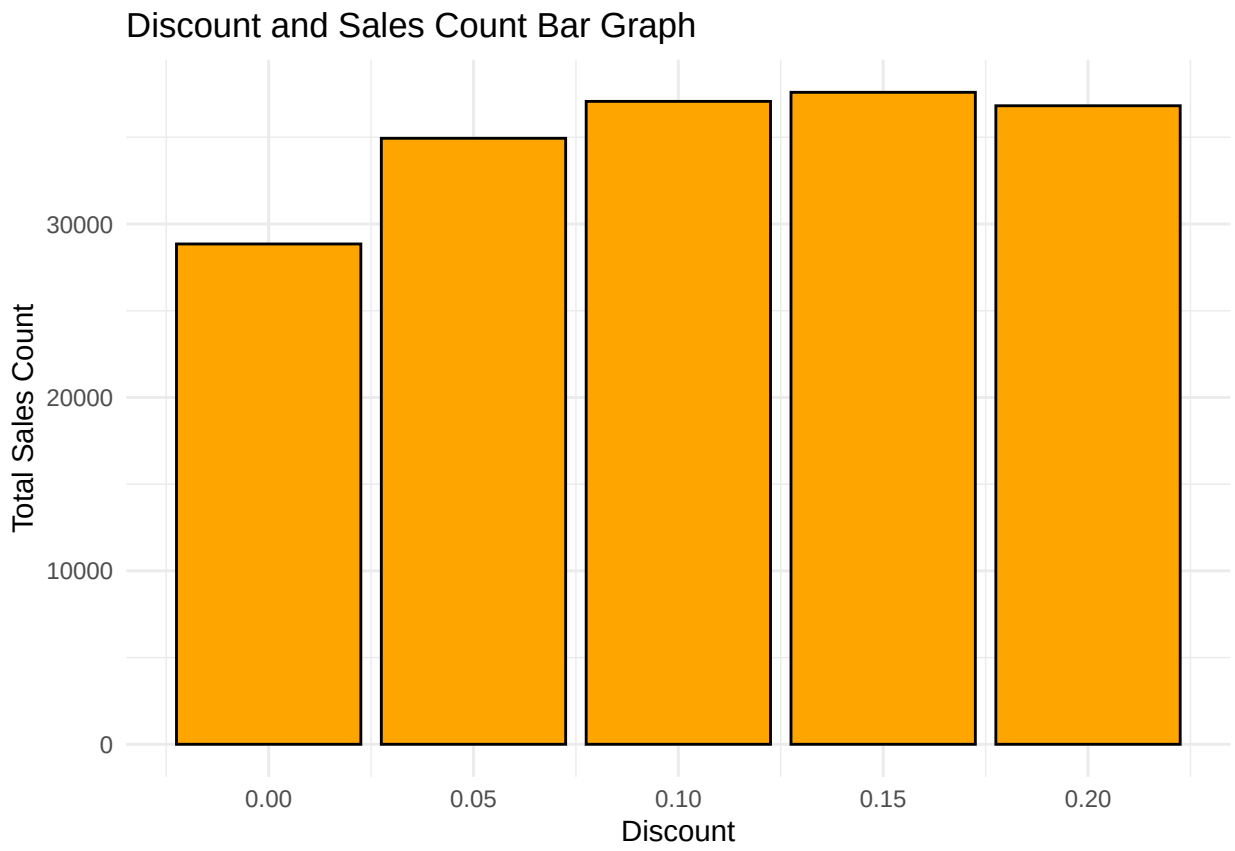


discount - Hyukjoon Choi

Bar graph

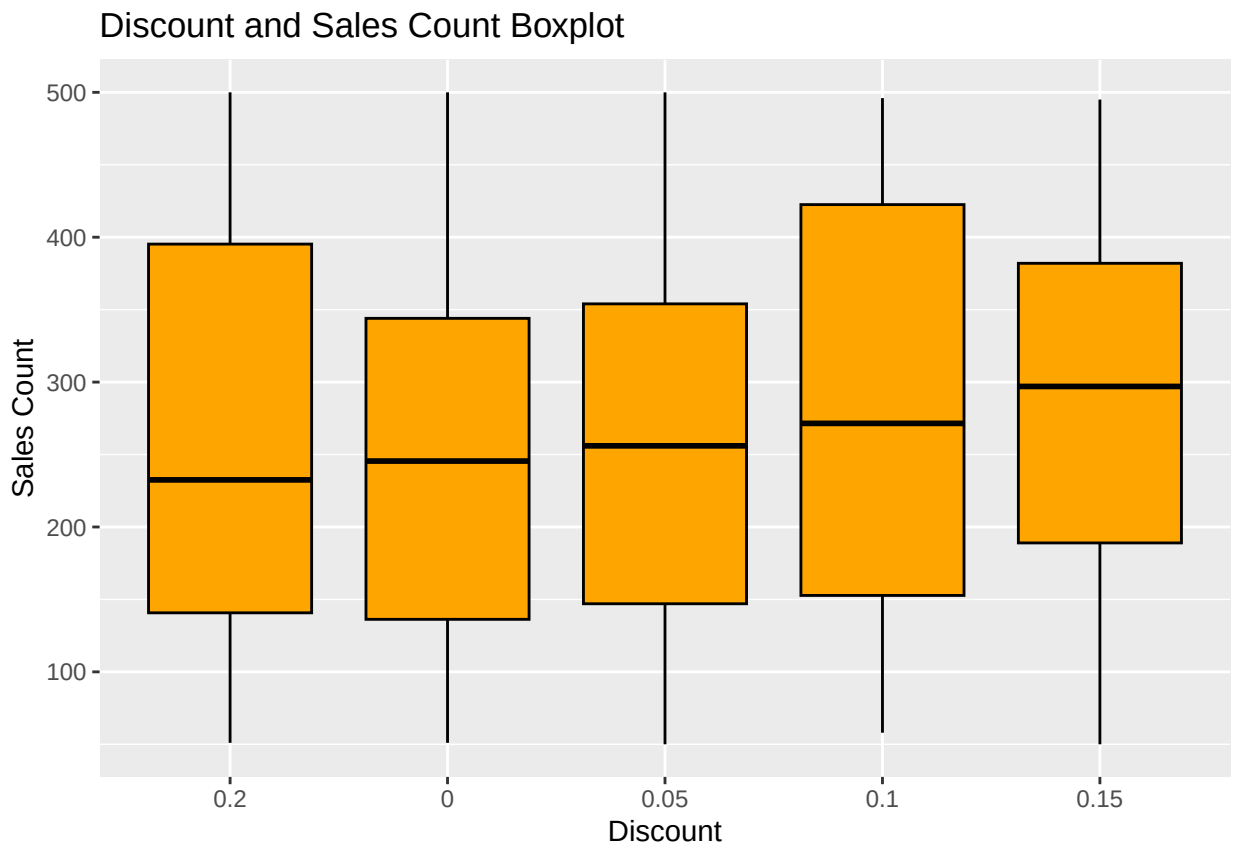
```
discount_sales <- fashion %>%
  group_by(discount) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(discount_sales) +
  geom_col(mapping = aes(x = discount, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Discount and Sales Count Bar Graph",
    x = "Discount",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(discount, sales_count,
                                            FUN=median)), color = "black",
                                            fill = "orange")+
  labs(
  title = "Discount and Sales Count Boxplot",
  x = "Discount",
  y = "Sales Count"
  )
```

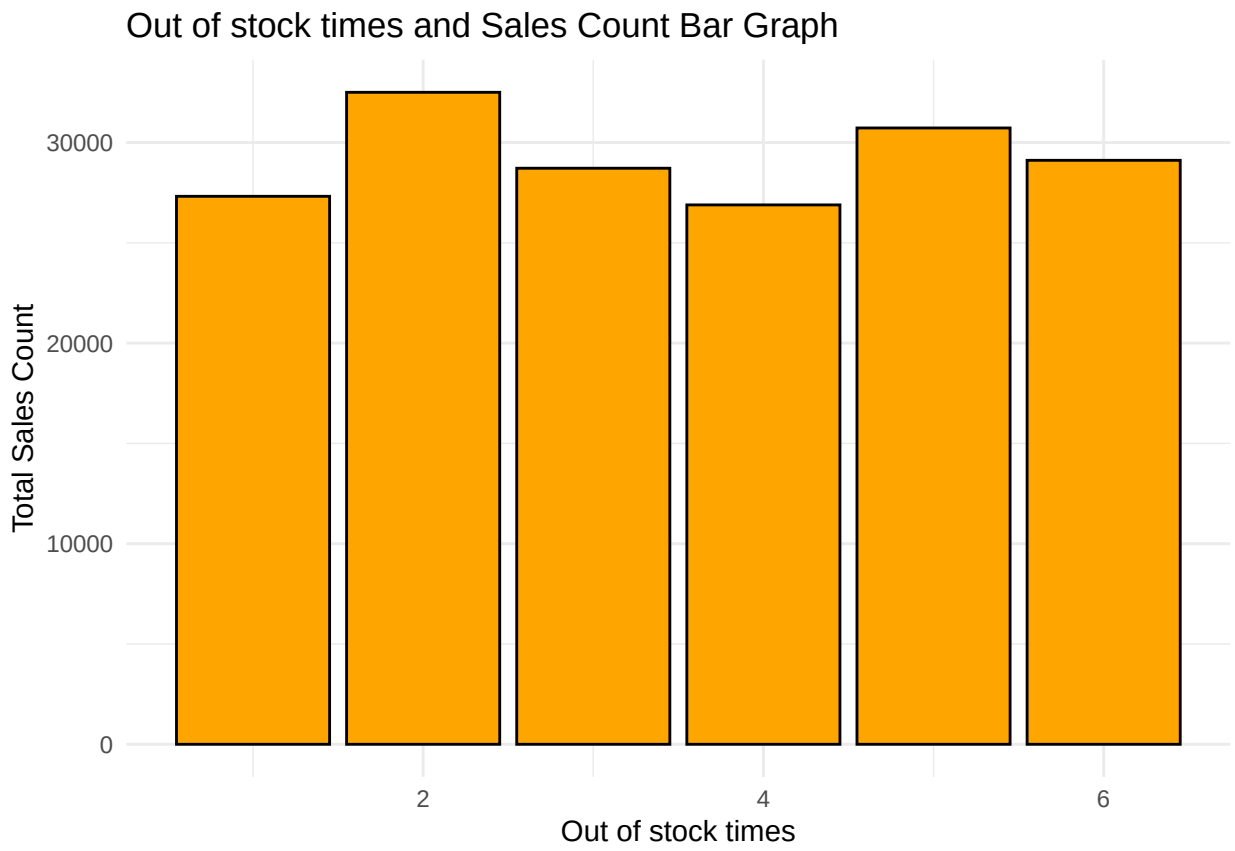


out_of_stock_times - ChaeJun Lim

Bar Graph

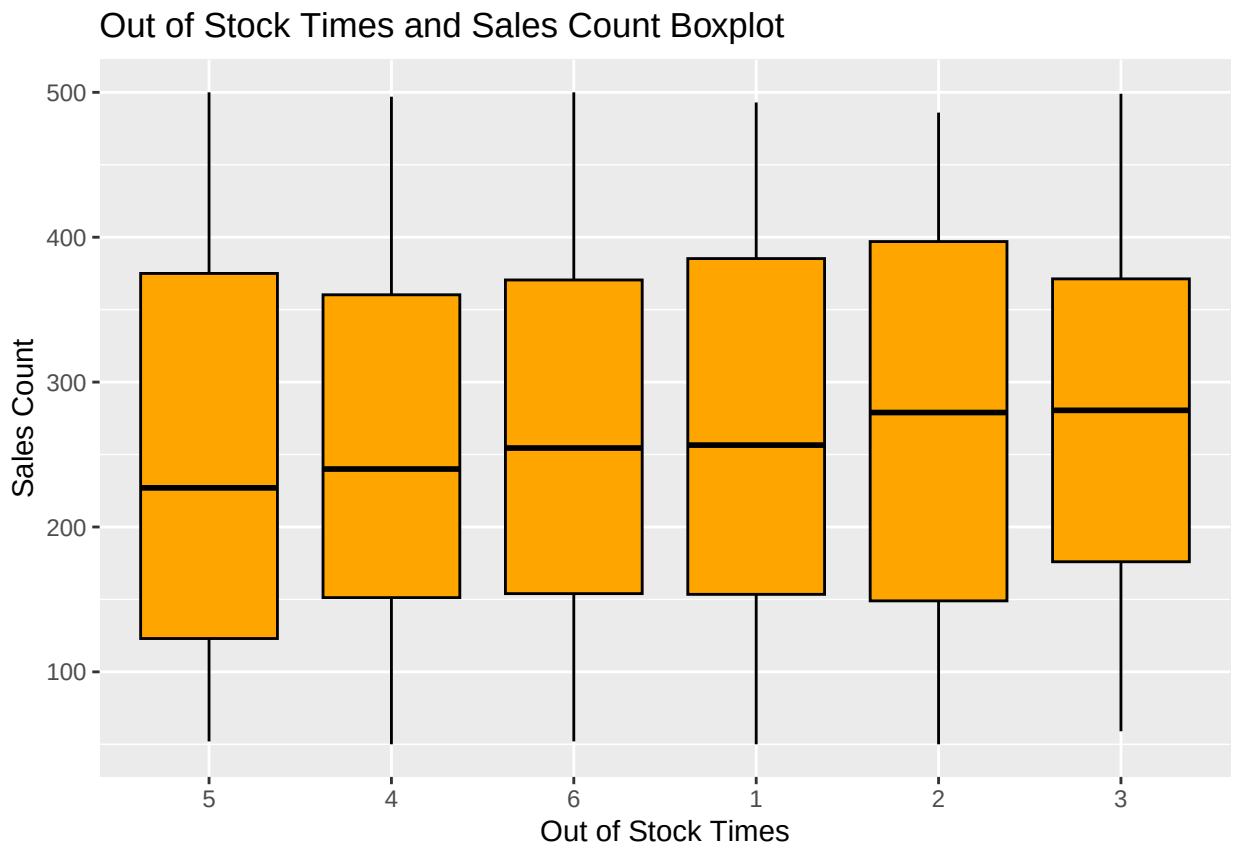
```
out_of_stock_times_sales <- fashion %>%
  group_by(out_of_stock_times) %>%
  summarize(total_sales_count = sum(sales_count, na.rm = TRUE))
```

```
ggplot(out_of_stock_times_sales) +
  geom_col(mapping = aes(x = out_of_stock_times, y = total_sales_count),
           color = "black", fill = "orange") +
  labs(
    title = "Out of stock times and Sales Count Bar Graph",
    x = "Out of stock times",
    y = "Total Sales Count"
  ) +
  theme_minimal()
```



Boxplot

```
fashion %>%
  ggplot()+
  geom_boxplot(aes(y=sales_count, x=reorder(out_of_stock_times,
                                            sales_count, FUN=median)),
               color = "black", fill = "orange")+
  labs(
  title = "Out of Stock Times and Sales Count Boxplot",
  x = "Out of Stock Times",
  y = "Sales Count"
  )
```



Hypothesis 1

Gender & Sales Count - ChaeJun Lim, Hyukjoon Choi

Hypothesis: There is at least one relationship between gender and sales count.

Alternative: There is at least relationship between gender and sales count.

Male & Female

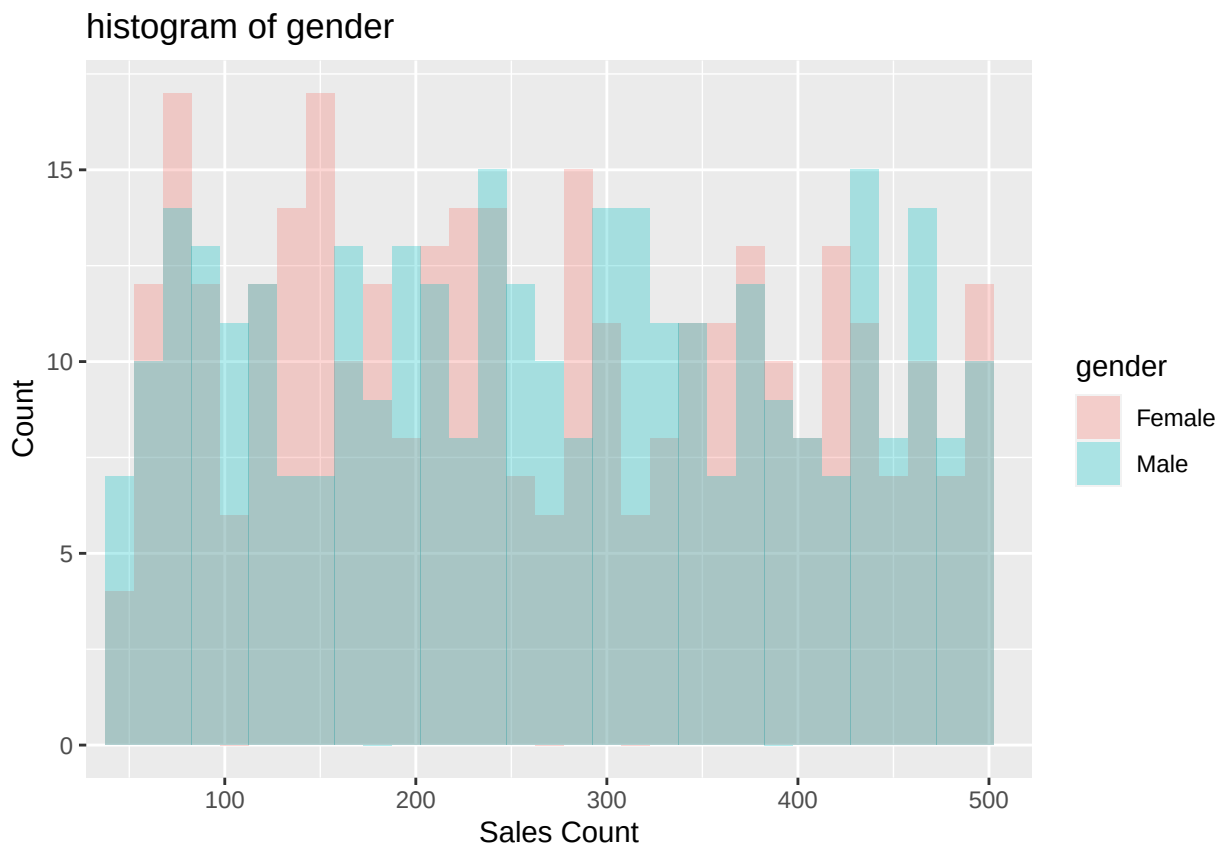
```
fashion %>%
  select(sales_count) %>%
  arrange(desc(sales_count)) %>%
  head(5)
```

```
## # A tibble: 5 x 1
##   sales_count
##         <dbl>
## 1         500
## 2         500
## 3         500
## 4         499
## 5         497
```

```
fashion %>%
  select(sales_count) %>%
  arrange(sales_count) %>%
  head(5)
```

```
## # A tibble: 5 x 1
##   sales_count
##         <dbl>
## 1          50
## 2          50
## 3          50
## 4          51
## 5          51
```

```
fashion %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = gender),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of gender",
  x = "Sales Count",
  y = "Count"
  )
```



```r
fashion %>%
  group_by(gender) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   gender  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Female  263.    248  132.    50   500   233
## 2 Male    268.    264  132.    50   500   218
```

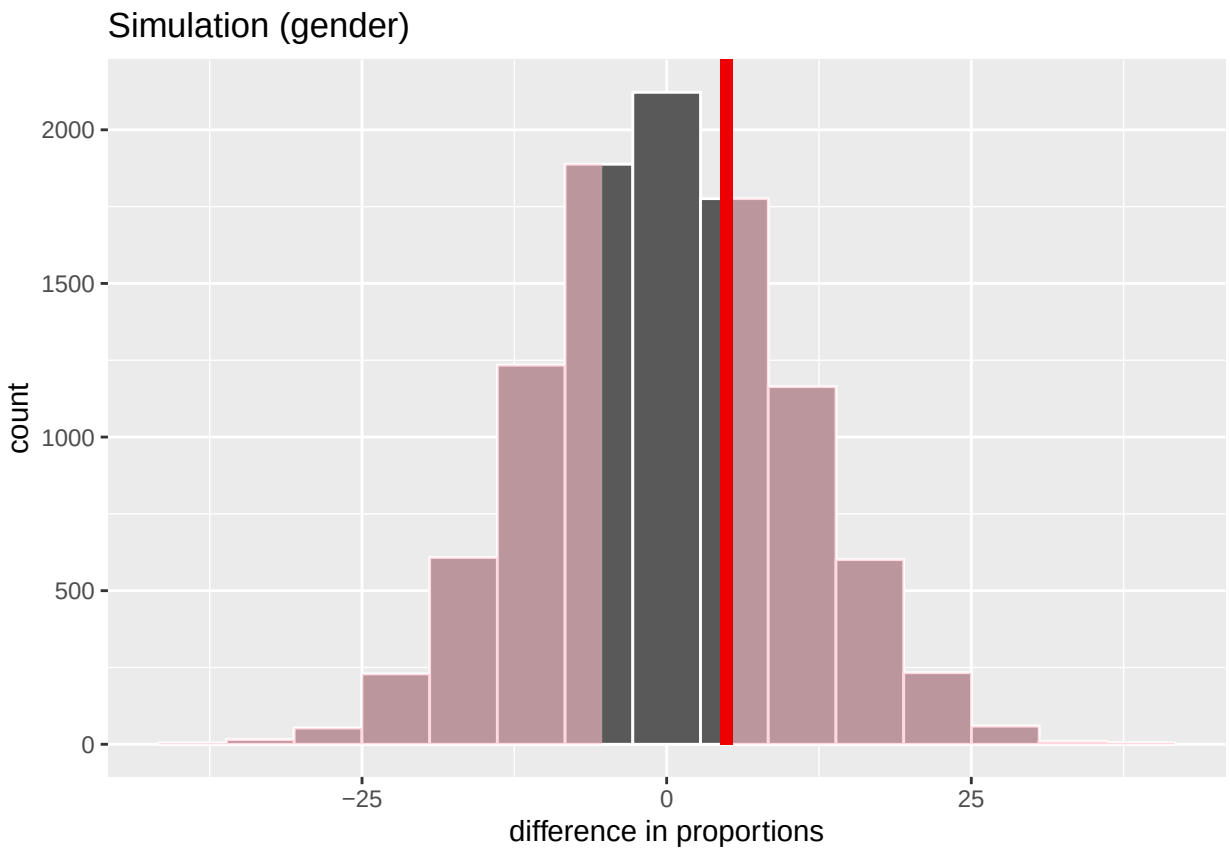
```r
fashion_gender_null <- fashion %>%
  specify(formula = sales_count ~ gender) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Male", "Female"))
```

```r
fashion_gender_obs_stat <- fashion %>%
  specify(sales_count ~ gender) %>%
  calculate(stat = "diff in means", order = c("Male", "Female"))
```

```r
fashion_gender_null %>%
  get_p_value(obs_stat = fashion_gender_obs_stat, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.625
```

```r
visualize(fashion_gender_null) +
  shade_p_value(obs_stat = fashion_gender_obs_stat, direction = "two_sided") +
  labs(
  title = "Simulation (gender)",
  x = "difference in proportions",
  y = "count"
  )
```



Hypothesis 2

Age & Sales Count - Seunghyun Ryu, Miseo Yoon

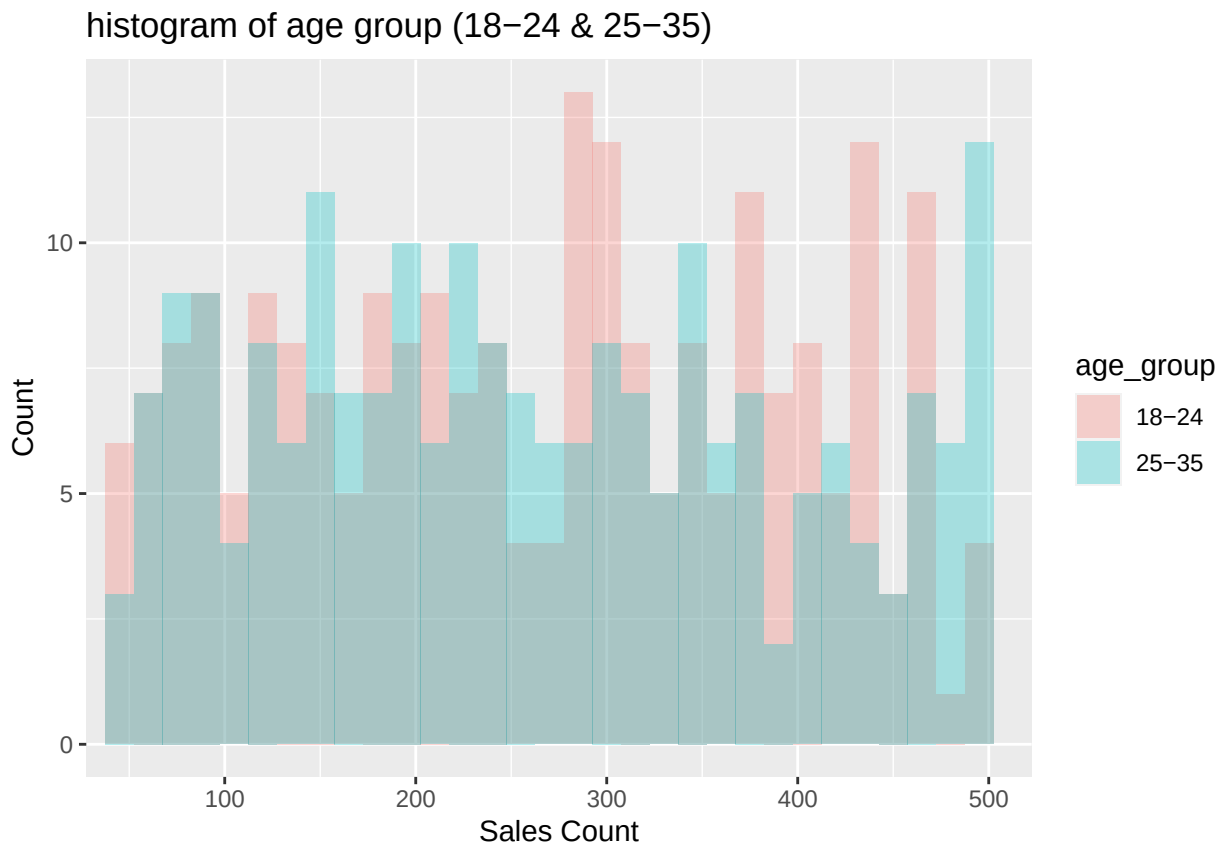
Hypothesis: There is a relationship between age and sales count.

Alternative: There is no relationship between age and sales count.

18-24 & 25-35

```
fashion_age <- fashion %>%
  filter(age_group == "18-24" | age_group == "25-35")
```

```
fashion_age %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = age_group),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of age group (18-24 & 25-35)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_age %>%
  group_by(age_group) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   age_group  mean median    sd   min   max   IQR
##   <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 18-24      265.   276   128.    50   492  225.
## 2 25-35      263.   248.  133.    51   500  218
```

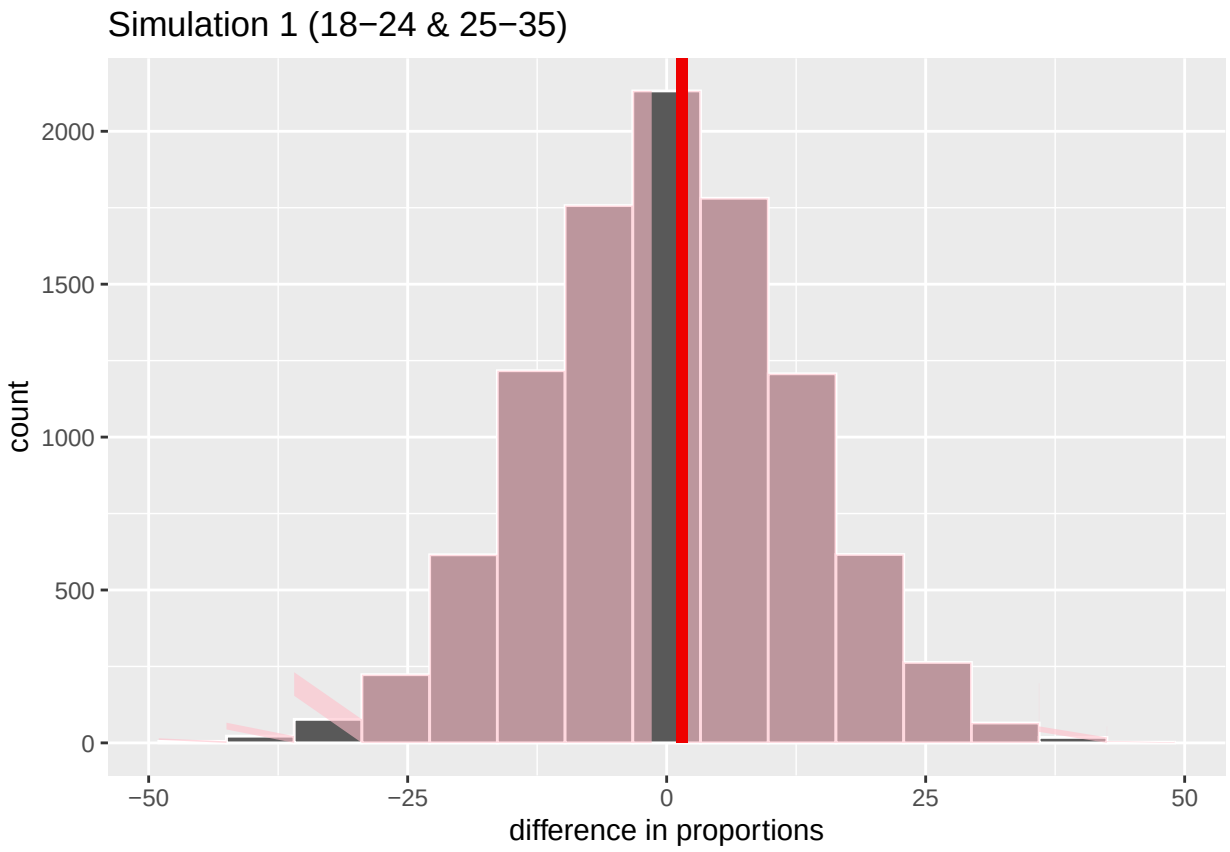
```
fashion_age_null <- fashion_age %>%
  specify(formula = sales_count ~ age_group) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("18-24", "25-35"))
```

```
fashion_age_obs_stat <- fashion_age %>%
  specify(sales_count ~ age_group) %>%
  calculate(stat = "diff in means", order = c("18-24", "25-35"))
```

```
fashion_age_null %>%
  get_p_value(obs_stat = fashion_age_obs_stat, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.899
```

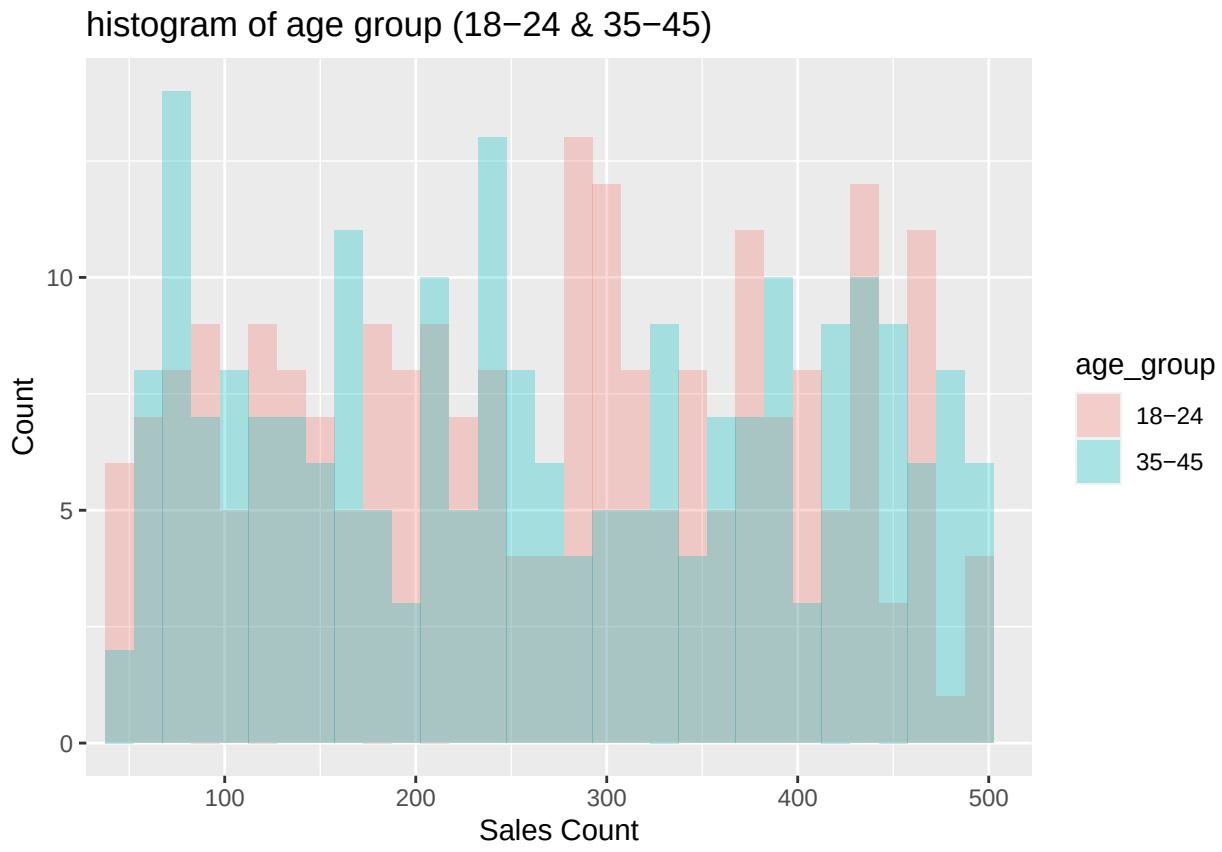
```
visualize(fashion_age_null) +
  shade_p_value(obs_stat = fashion_age_obs_stat, direction = "two_sided") +
  labs(
  title = "Simulation 1 (18-24 & 25-35)",
  x = "difference in proportions",
  y = "count"
  )
```



18-24 & 35-45

```r
fashion_age_2 <- fashion %>%
  filter(age_group == "18-24" | age_group == "35-45")
```

```r
fashion_age_2 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = age_group),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of age group (18-24 & 35-45)",
  x = "Sales Count",
  y = "Count"
  )
```




```r
fashion_age_2 %>%
  group_by(age_group) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   age_group  mean median    sd   min   max   IQR
##   <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 18-24      265.   276   128.    50   492  225.
## 2 35-45      269.   258.  137.    51   500  244.
```

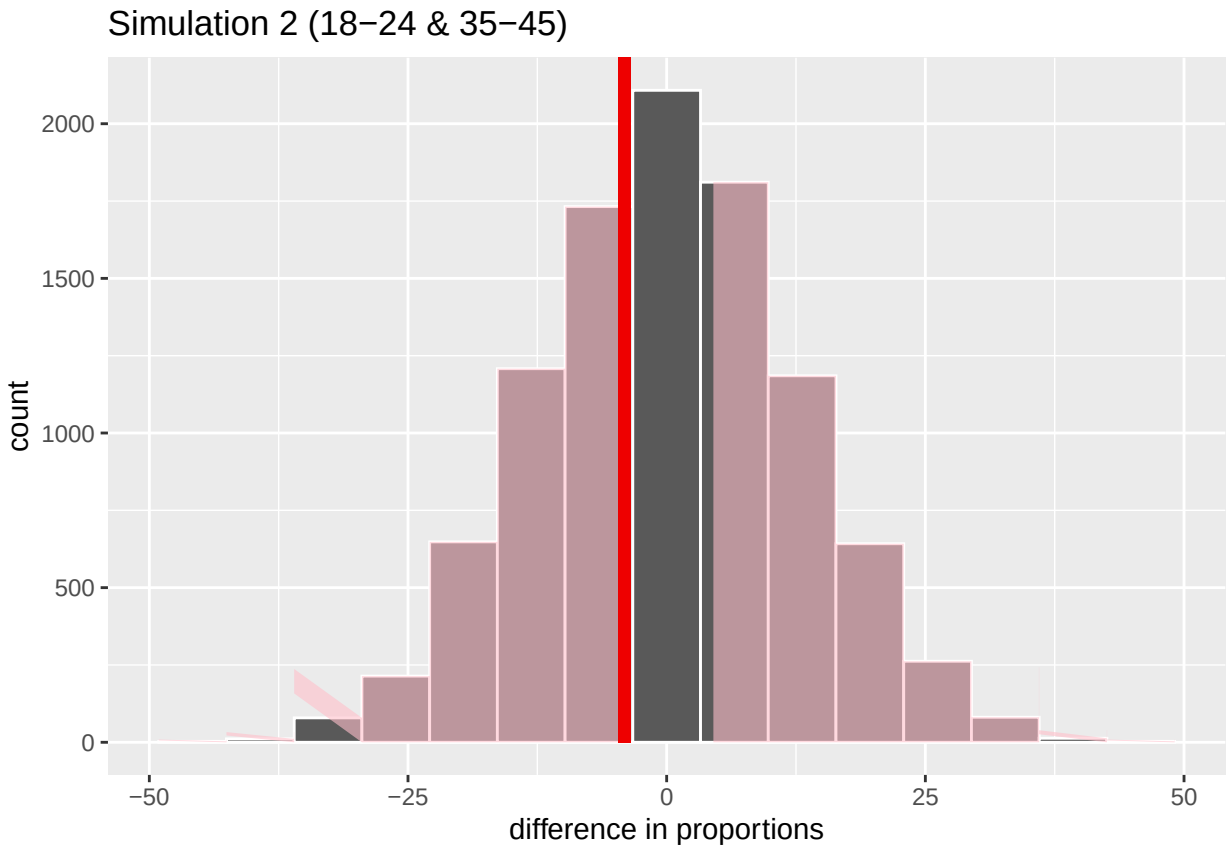
```
fashion_age_null_2 <- fashion_age_2 %>%
  specify(formula = sales_count ~ age_group) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("18-24", "35-45"))
```

```
fashion_age_obs_stat_2 <- fashion_age_2 %>%
  specify(sales_count ~ age_group) %>%
  calculate(stat = "diff in means", order = c("18-24", "35-45"))
```

```
fashion_age_null_2 %>%
  get_p_value(obs_stat = fashion_age_obs_stat_2, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.729
```

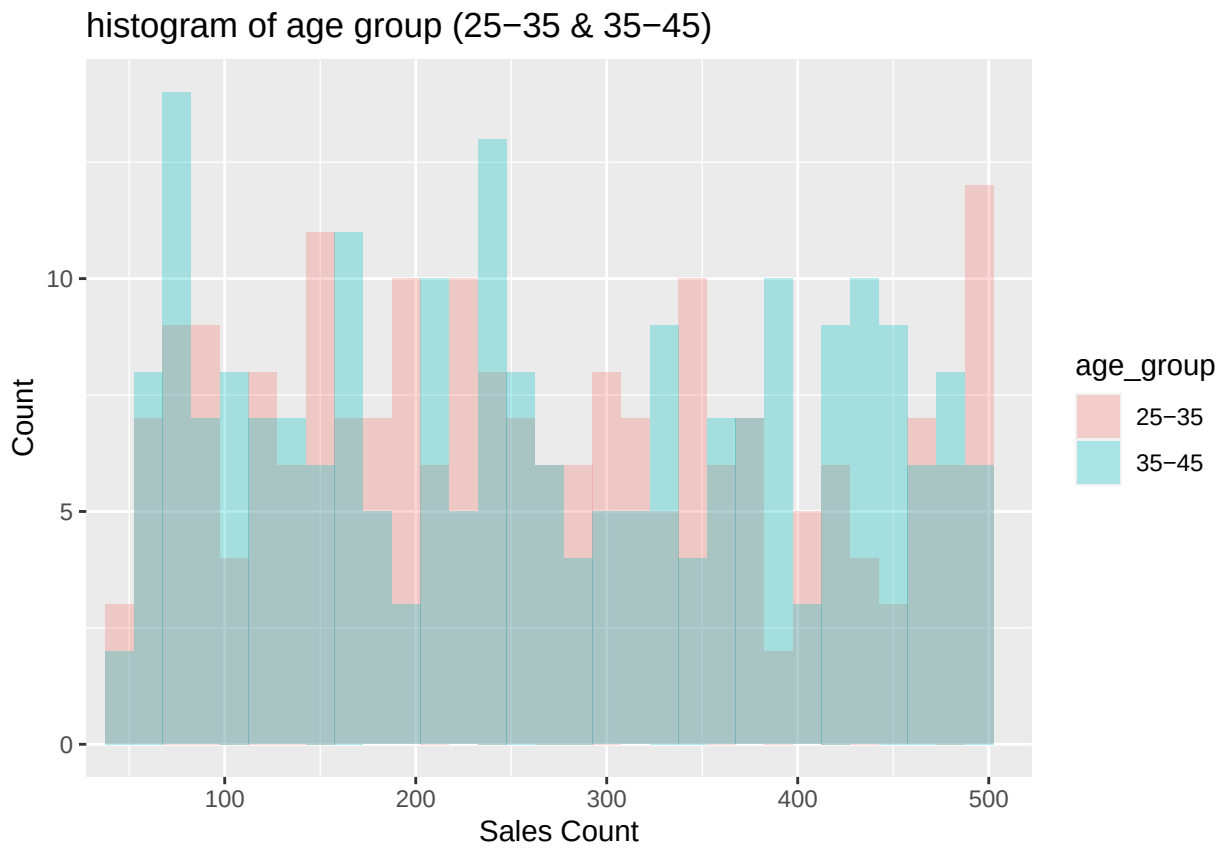
```
visualize(fashion_age_null_2) +
  shade_p_value(obs_stat = fashion_age_obs_stat_2, direction = "two_sided") +
  labs(
  title = "Simulation 2 (18-24 & 35-45)",
  x = "difference in proportions",
  y = "count"
  )
```



25-35 & 35-45

```
fashion_age_3 <- fashion %>%
  filter(age_group == "25-35" | age_group == "35-45")
```

```
fashion_age_3 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = age_group),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of age group (25-35 & 35-45)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_age_3 %>%
  group_by(age_group) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   age_group  mean median    sd   min   max   IQR
##   <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 25-35      263.   248.  133.    51   500  218
## 2 35-45      269.   258.  137.    51   500  244.
```

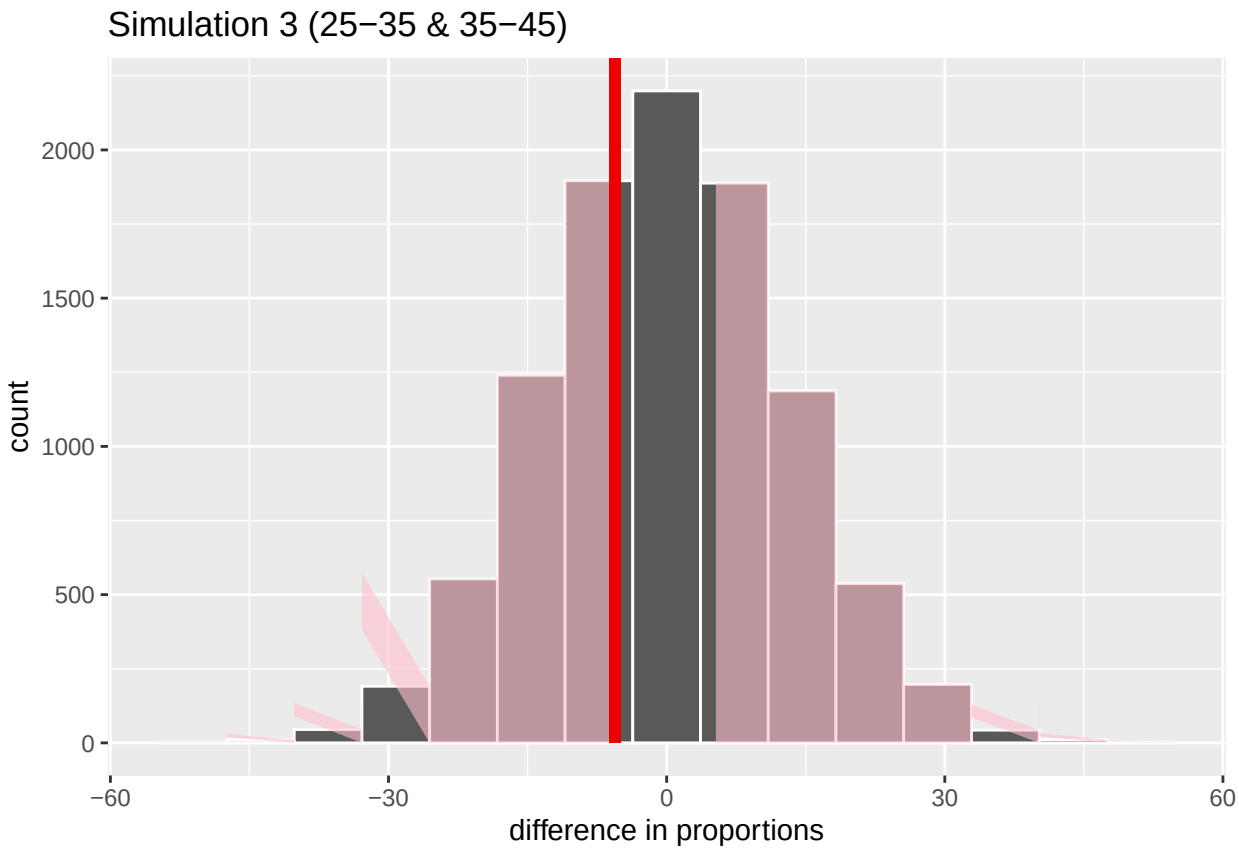
```
fashion_age_null_3 <- fashion_age_3 %>%
  specify(formula = sales_count ~ age_group) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("25-35", "35-45"))
```

```r
fashion_age_obs_stat_3 <- fashion_age_3 %>%
  specify(sales_count ~ age_group) %>%
  calculate(stat = "diff in means", order = c("25-35", "35-45"))
```

```r
fashion_age_null_3 %>%
  get_p_value(obs_stat = fashion_age_obs_stat_3, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.674
```

```r
visualize(fashion_age_null_3) +
  shade_p_value(obs_stat = fashion_age_obs_stat_3, direction = "two_sided") +
  labs(
  title = "Simulation 3 (25-35 & 35-45)",
  x = "difference in proportions",
  y = "count"
  )
```



Hypothesis 3

Season & Sales Count - Jin Heo, Juheon Kim

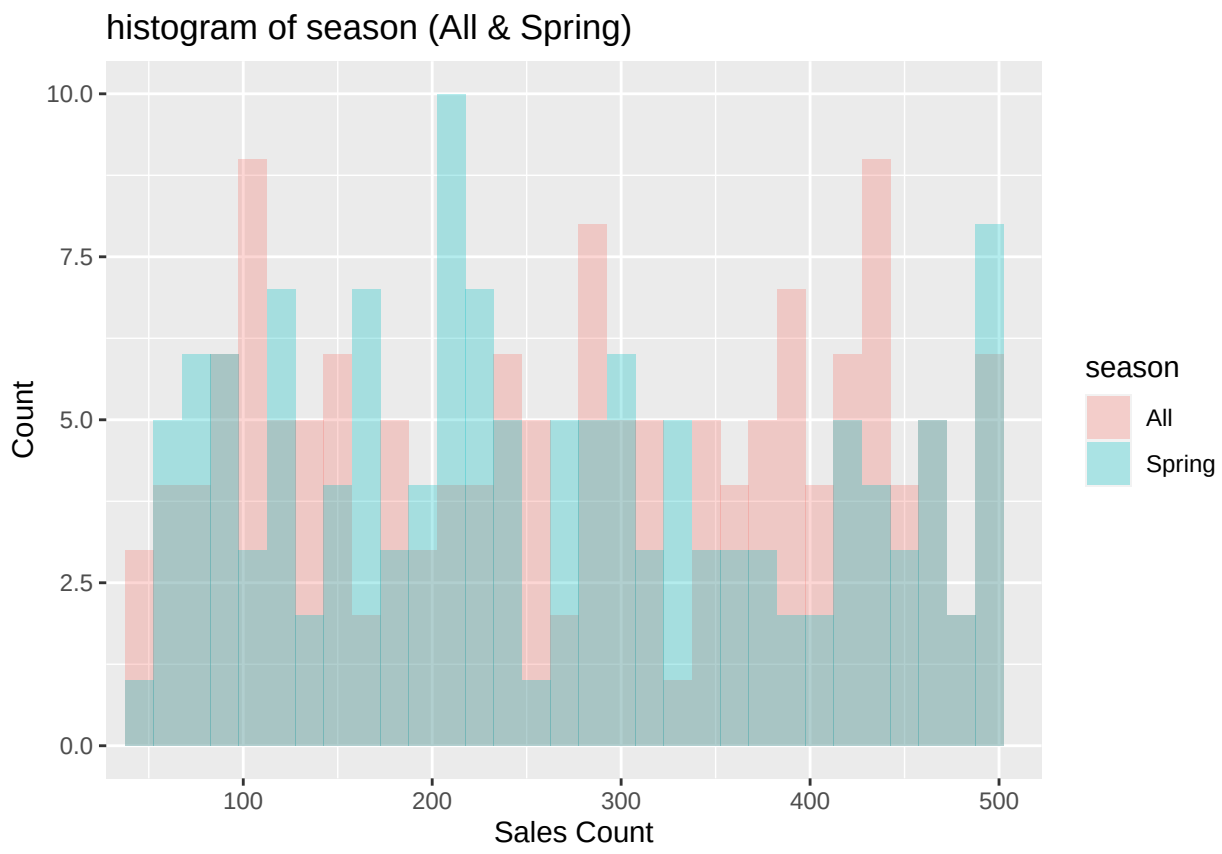
Hypothesis: There is at least one relationship between age and sales count.

Alternative: There is at least relationship between age and sales count.

All & Spring

```
fashion_season <-fashion %>%
  filter(season == "All" | season == "Spring")
```

```
fashion_season %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (All & Spring)",
  x = "Sales Count",
  y = "Count"
  )
```



```r
fashion_season %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 All     273.    279  136.    50   497  248
## 2 Spring  263.    245  133.    52   500  212.
```

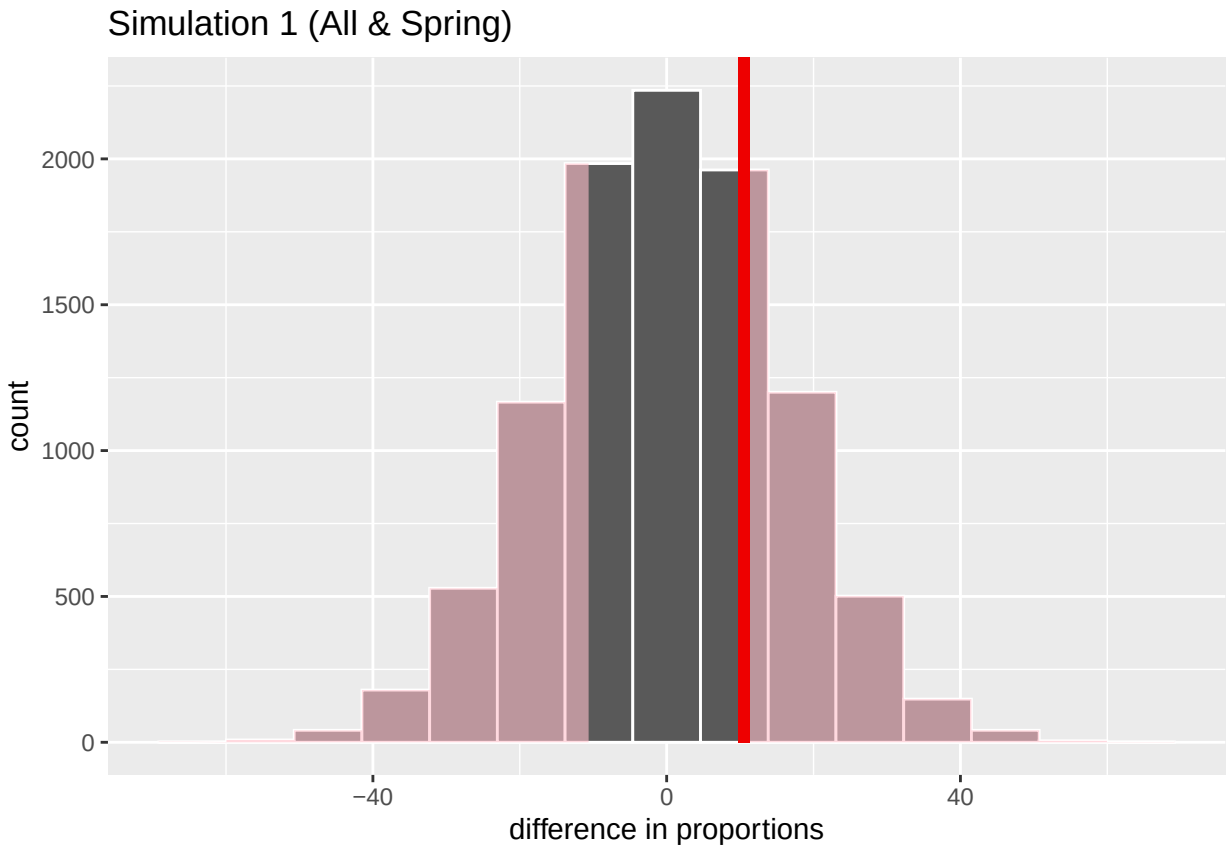
```r
fashion_season_null <- fashion_season %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("All", "Spring"))
```

```r
fashion_season_obs_stat <- fashion_season %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("All", "Spring"))
```

```r
fashion_season_null %>%
  get_p_value(obs_stat = fashion_season_obs_stat, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.508
```

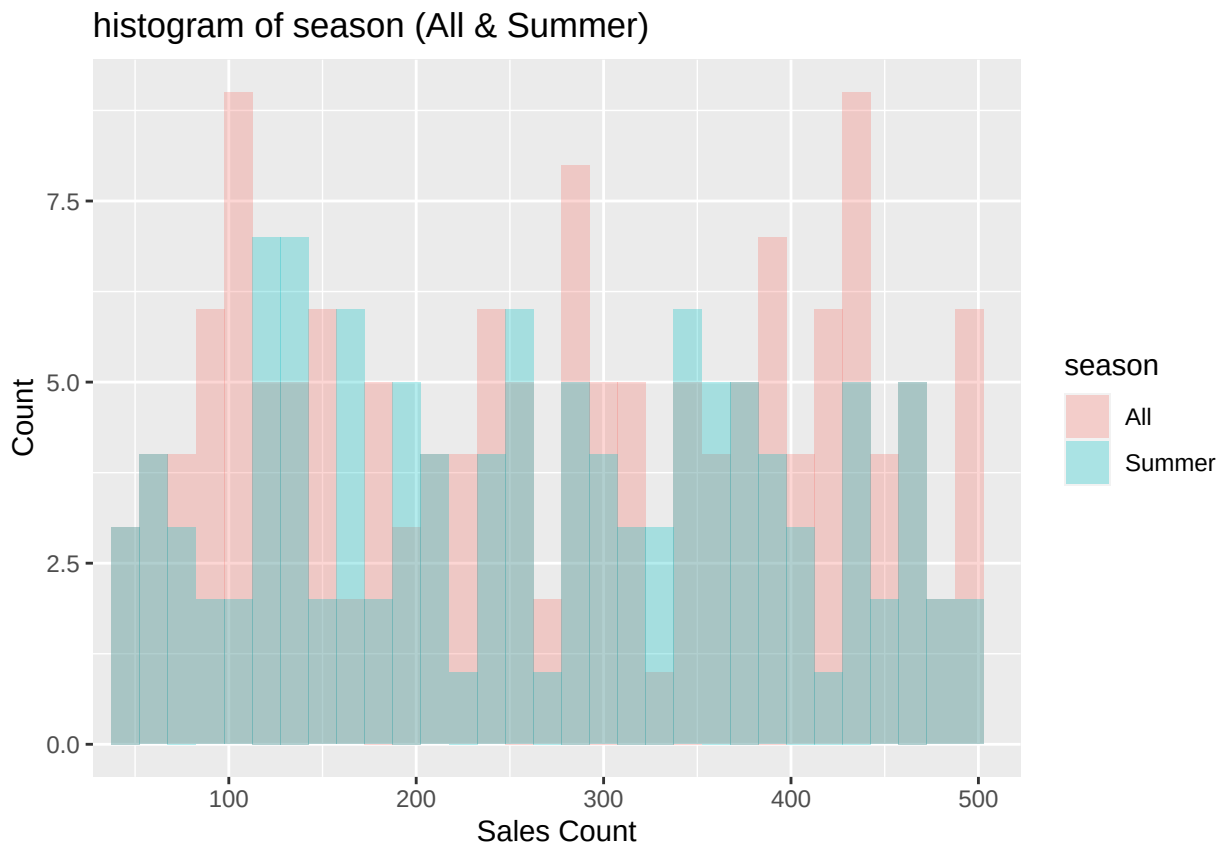
```r
visualize(fashion_season_null) +
  shade_p_value(obs_stat = fashion_season_obs_stat, direction = "two_sided") +
  labs(
  title = "Simulation 1 (All & Spring)",
  x = "difference in proportions",
  y = "count"
  )
```



All & Summer

```
fashion_season_2 <-fashion %>%
  filter(season == "All" | season == "Summer")
```

```
fashion_season_2 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (All & Summer)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_2 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 All     273.    279  136.    50   497  248
## 2 Summer  264.    258  128.    50   496  222.
```

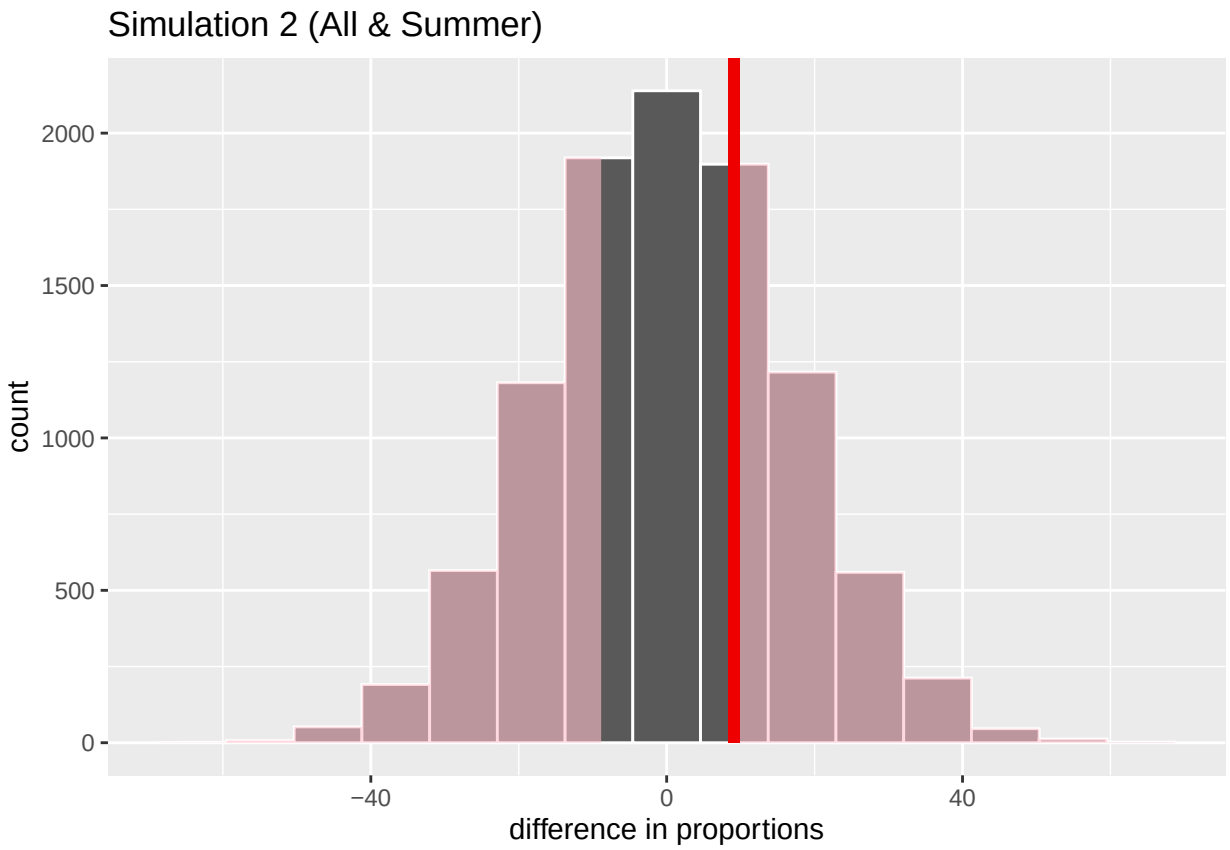
```
fashion_season_null_2 <- fashion_season_2 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("All", "Summer"))
```

```r
fashion_season_obs_stat_2 <- fashion_season_2 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("All", "Summer"))
```

```r
fashion_season_null_2 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_2, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.587
```

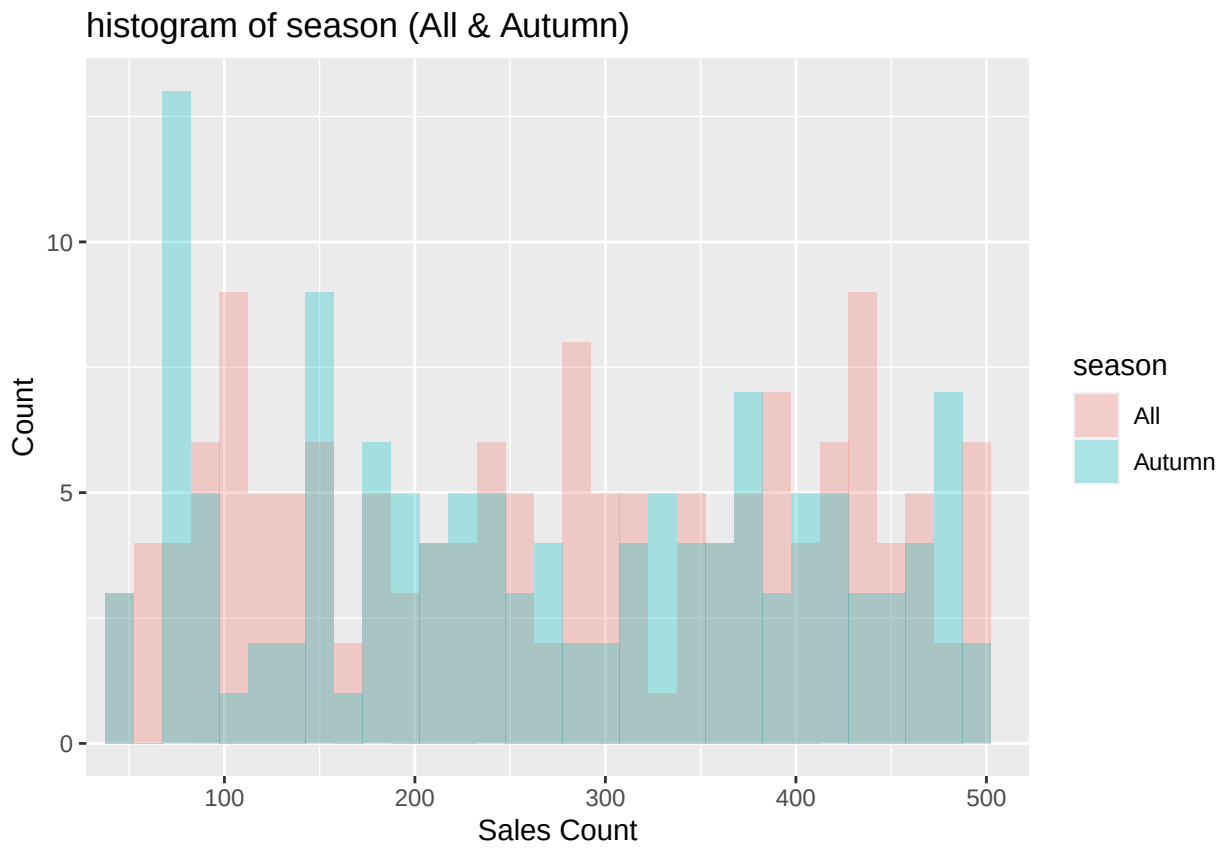
```r
visualize(fashion_season_null_2) +
  shade_p_value(obs_stat = fashion_season_obs_stat_2, direction = "two_sided") +
  labs(
  title = "Simulation 2 (All & Summer)",
  x = "difference in proportions",
  y = "count"
  )
```



All & Autumn

```
fashion_season_3 <-fashion %>%
  filter(season == "All" | season == "Autumn")
```

```
fashion_season_3 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (All & Autumn)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_3 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 All     273.    279  136.    50   497  248
## 2 Autumn  269.    260  136.    51   494  232.
```

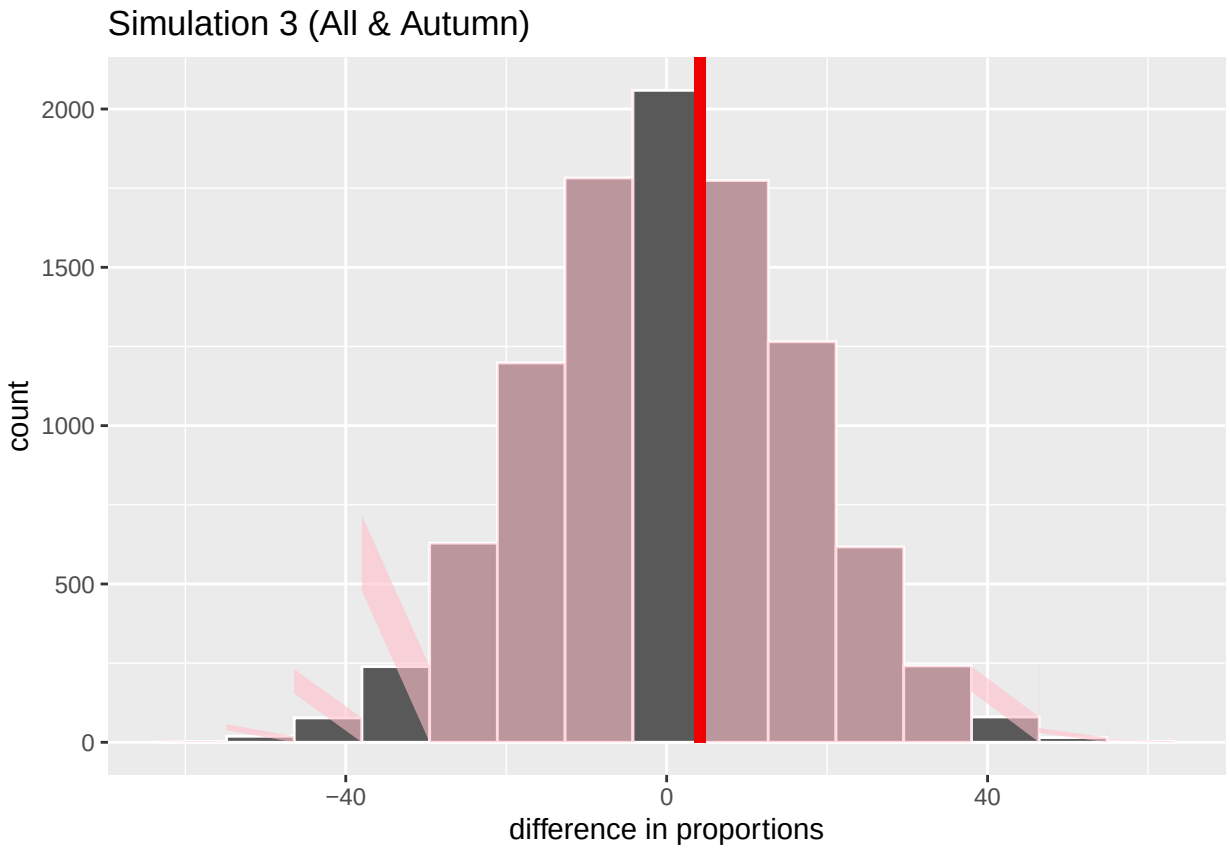
```
fashion_season_null_3 <- fashion_season_3 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("All", "Autumn"))
```

```
fashion_season_obs_stat_3 <- fashion_season_3 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("All", "Autumn"))
```

```
fashion_season_null_3 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_3, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.803
```

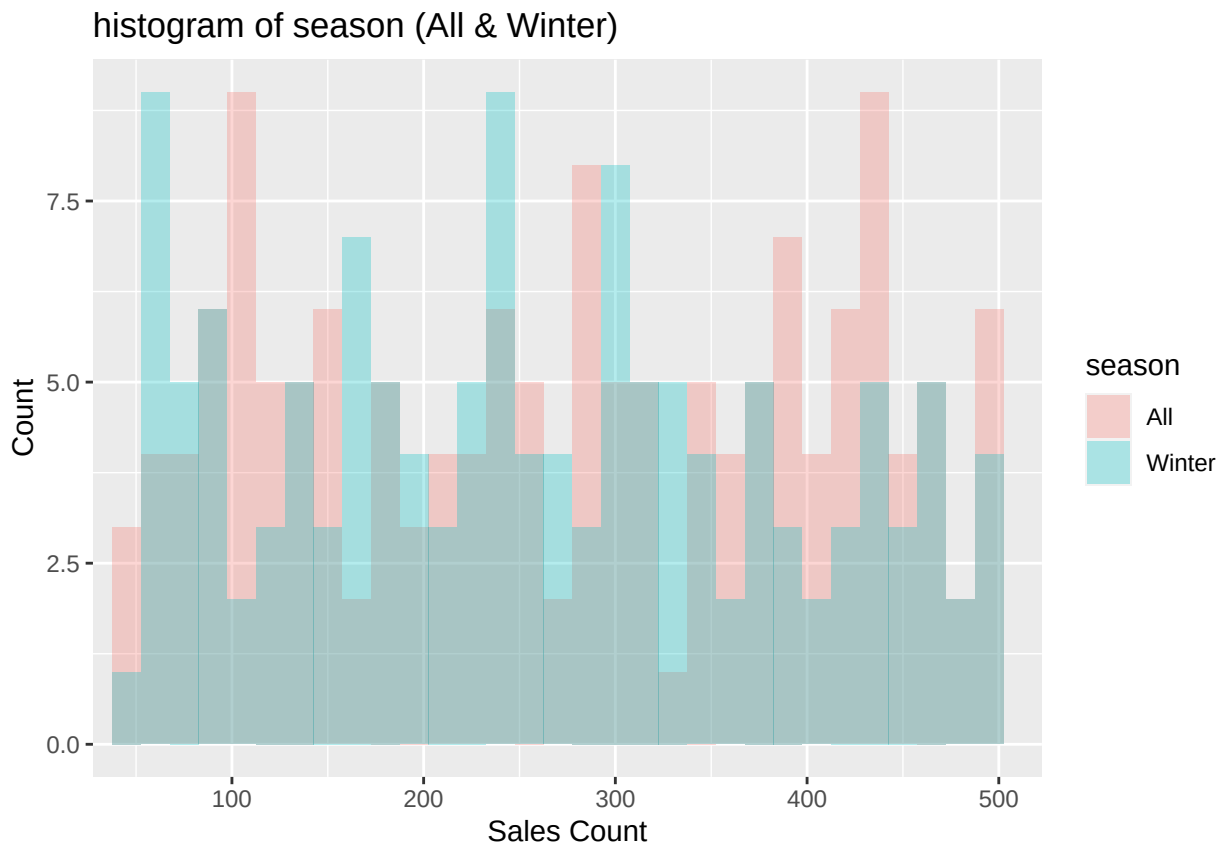
```
visualize(fashion_season_null_3) +
  shade_p_value(obs_stat = fashion_season_obs_stat_3, direction = "two_sided") +
  labs(
  title = "Simulation 3 (All & Autumn)",
  x = "difference in proportions",
  y = "count"
  )
```



All & Winter

```
fashion_season_4 <-fashion %>%
  filter(season == "All" | season == "Winter")
```

```
fashion_season_4 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (All & Winter)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_4 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 All     273.   279   136.    50   497   248
## 2 Winter  258.   248.  130.    52   500   203
```

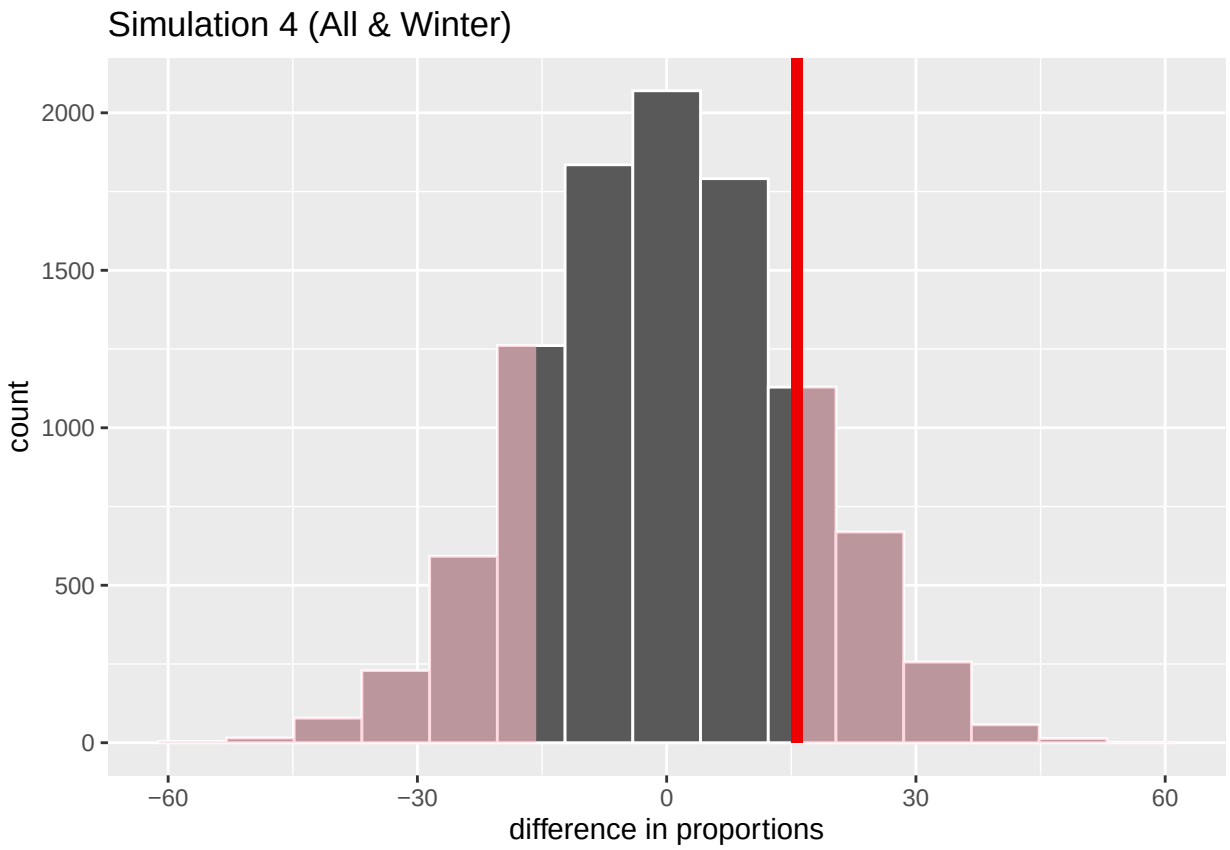
```
fashion_season_null_4 <- fashion_season_4 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("All", "Winter"))
```

```
fashion_season_obs_stat_4 <- fashion_season_4 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("All", "Winter"))
```

```
fashion_season_null_4 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_4, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.311
```

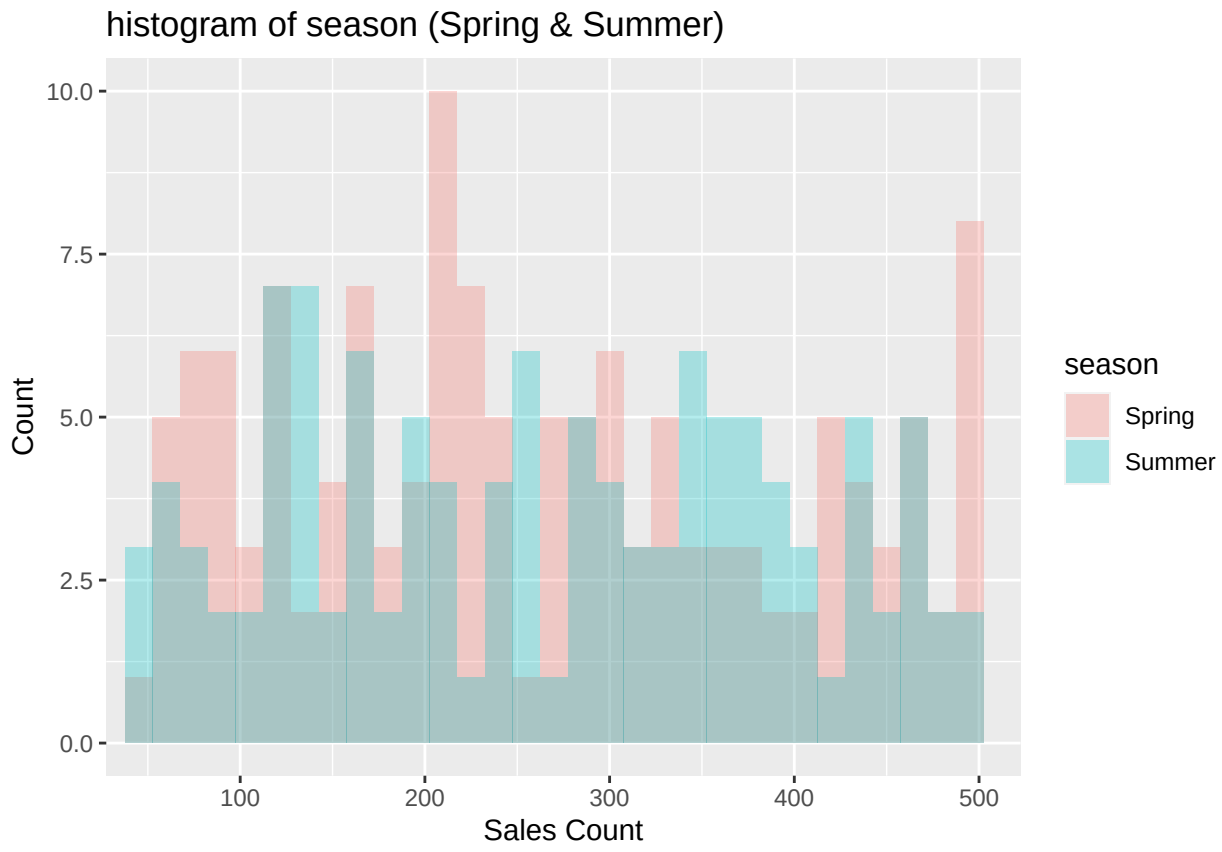
```
visualize(fashion_season_null_4) +
  shade_p_value(obs_stat = fashion_season_obs_stat_4, direction = "two_sided") +
  labs(
  title = "Simulation 4 (All & Winter)",
  x = "difference in proportions",
  y = "count"
  )
```



Spring & Summer

```r
fashion_season_5 <-fashion %>%
  filter(season == "Spring" | season == "Summer")
```

```r
fashion_season_5 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (Spring & Summer)",
  x = "Sales Count",
  y = "Count"
  )
```




```r
fashion_season_5 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Spring  263.    245  133.    52   500  212.
## 2 Summer  264.    258  128.    50   496  222.
```

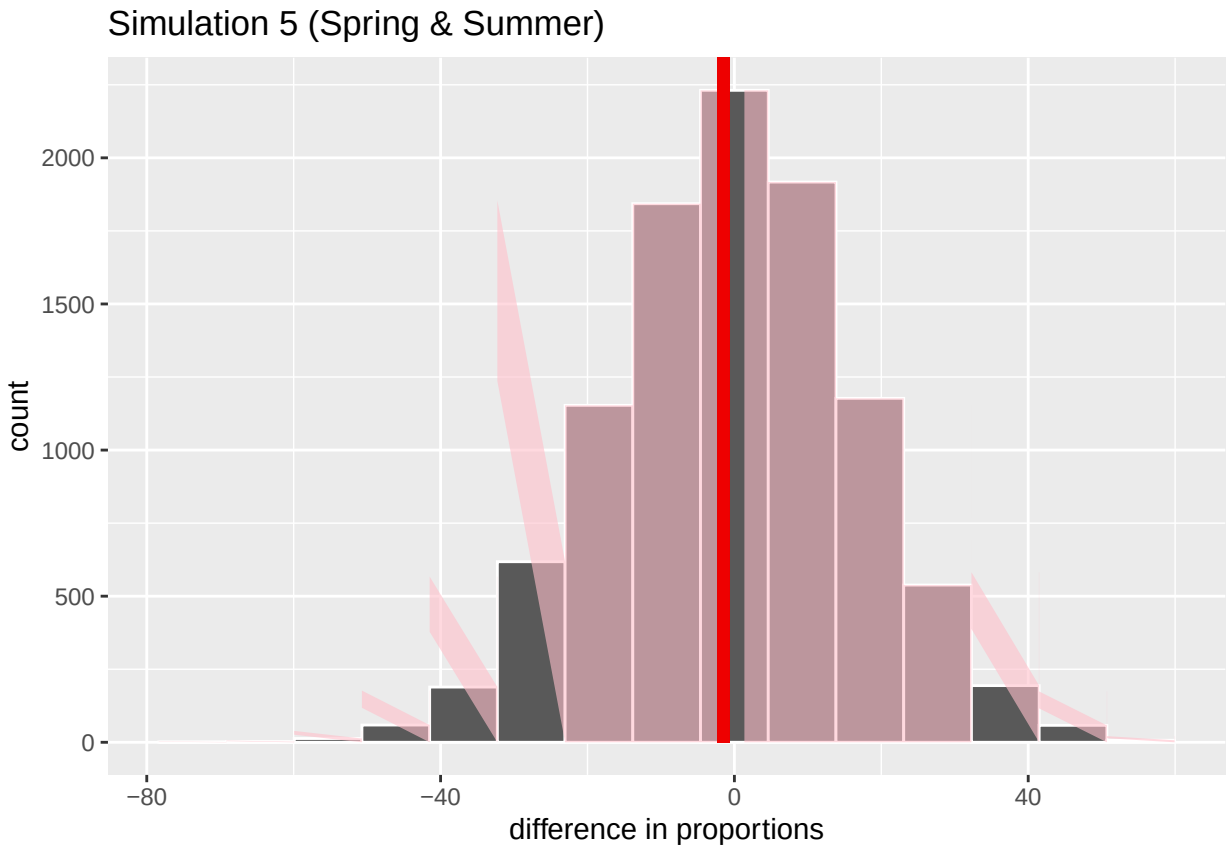
```
fashion_season_null_5 <- fashion_season_5 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Spring", "Summer"))
```

```
fashion_season_obs_stat_5 <- fashion_season_5 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("Spring", "Summer"))
```

```
fashion_season_null_5 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_5, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.932
```

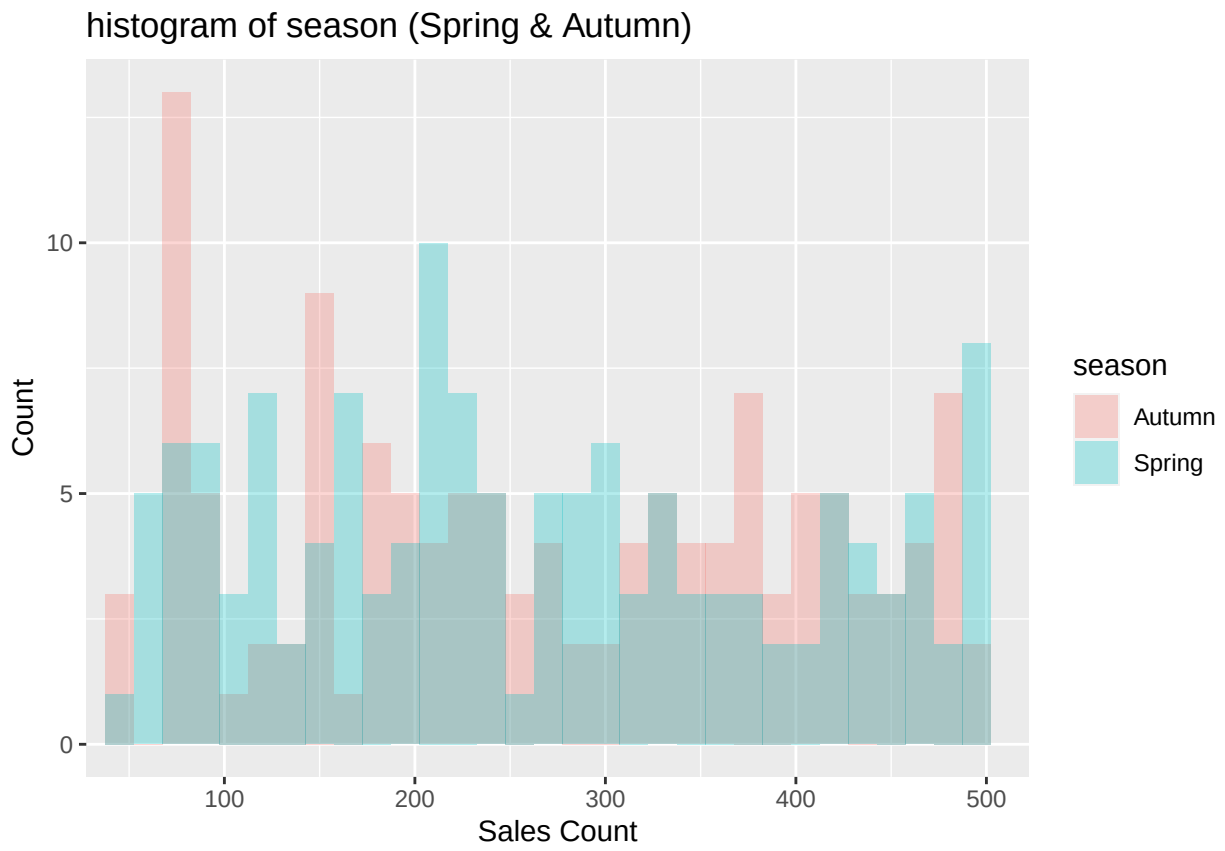
```
visualize(fashion_season_null_5) +
  shade_p_value(obs_stat = fashion_season_obs_stat_5, direction = "two_sided") +
  labs(
  title = "Simulation 5 (Spring & Summer)",
  x = "difference in proportions",
  y = "count"
  )
```



Spring & Autumn

```
fashion_season_6 <- fashion %>%
  filter(season == "Spring" | season == "Autumn")
```

```
fashion_season_6 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (Spring & Autumn)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_6 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Autumn  269.    260  136.    51   494  232.
## 2 Spring  263.    245  133.    52   500  212.
```

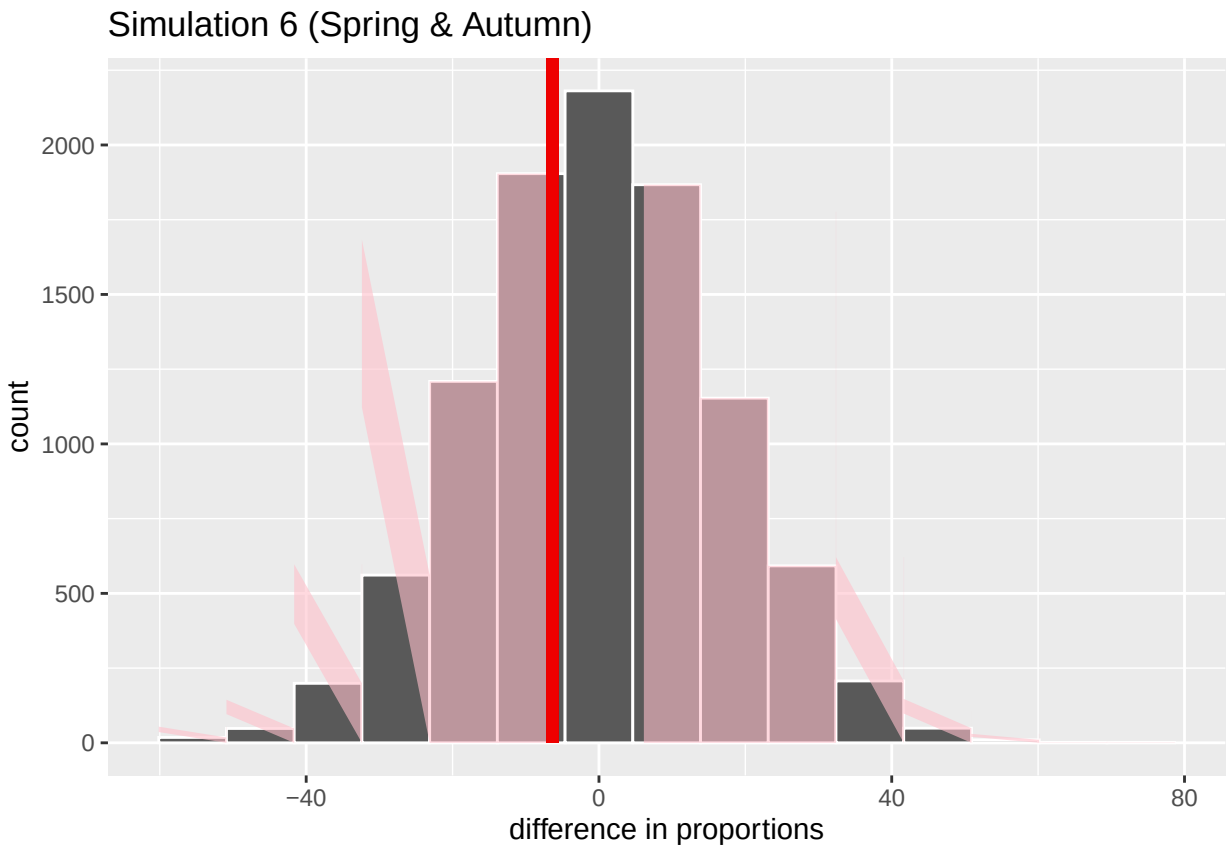
```
fashion_season_null_6 <- fashion_season_6 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Spring", "Autumn"))
```

```r
fashion_season_obs_stat_6 <- fashion_season_6 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("Spring", "Autumn"))
```

```r
fashion_season_null_6 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_6, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.709
```

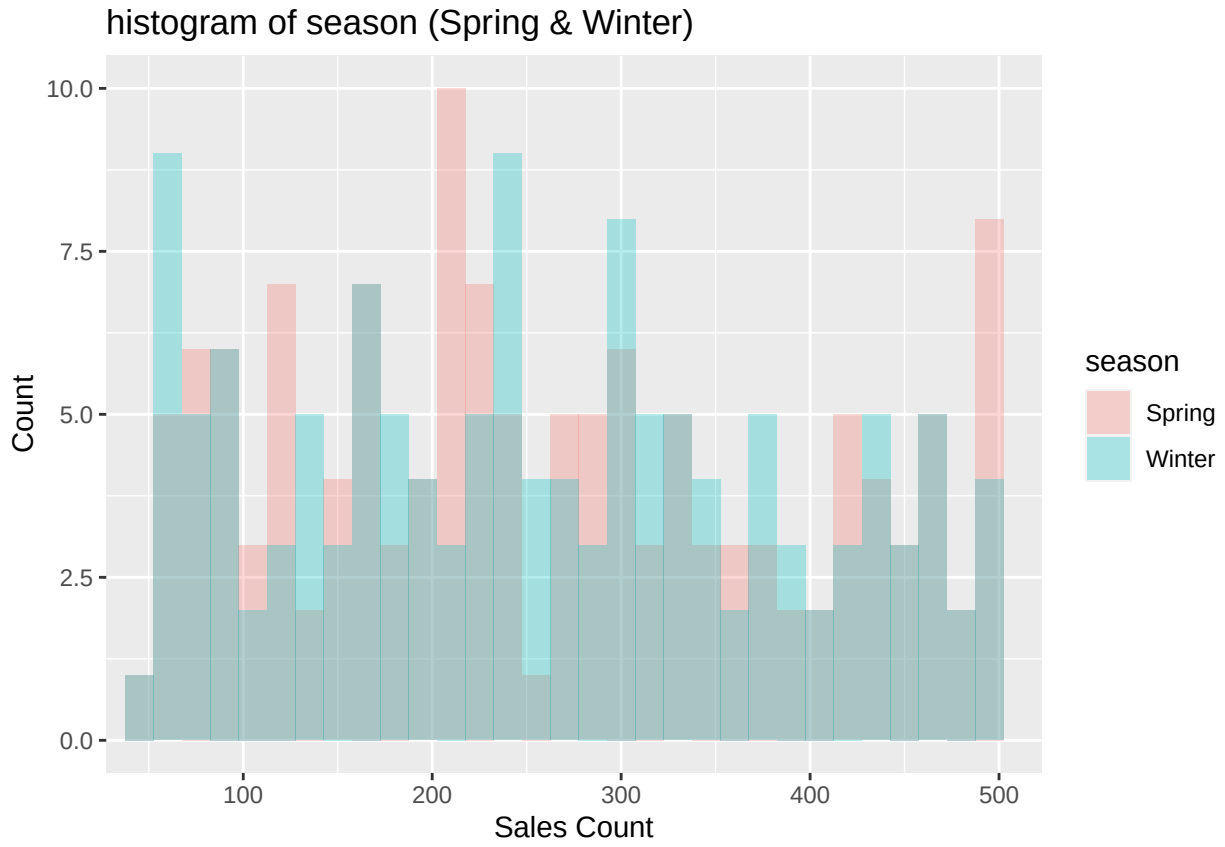
```r
visualize(fashion_season_null_6) +
  shade_p_value(obs_stat = fashion_season_obs_stat_6,
                direction = "two_sided") +
  labs(
  title = "Simulation 6 (Spring & Autumn)",
  x = "difference in proportions",
  y = "count"
  )
```



Spring & Winter

```
fashion_season_7 <- fashion %>%
  filter(season == "Spring" | season == "Winter")
```

```
fashion_season_7 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (Spring & Winter)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_7 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
```

```
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Spring  263.   245   133.    52   500  212.
## 2 Winter  258.   248.  130.    52   500  203
```

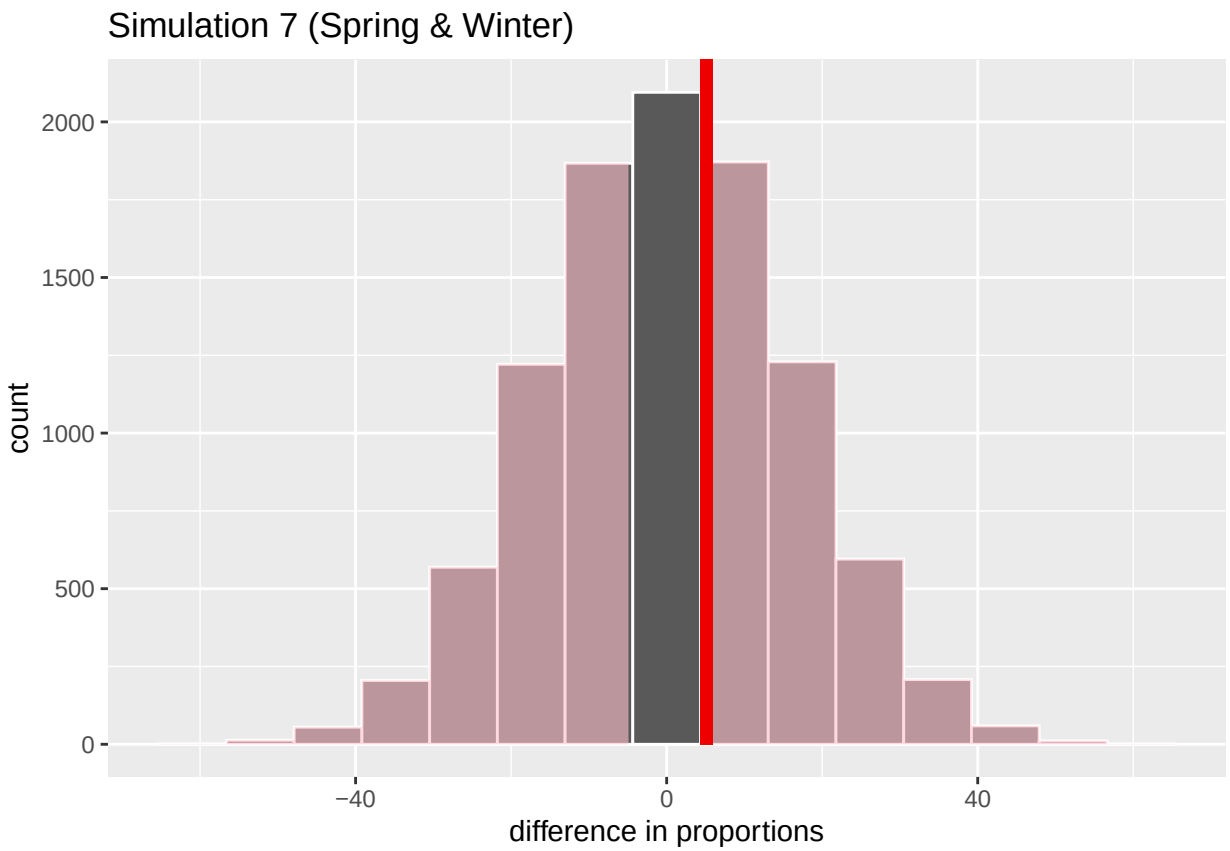
```
fashion_season_null_7 <- fashion_season_7 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Spring", "Winter"))
```

```
fashion_season_obs_stat_7 <- fashion_season_7 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("Spring", "Winter"))
```

```
fashion_season_null_7 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_7, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.757
```

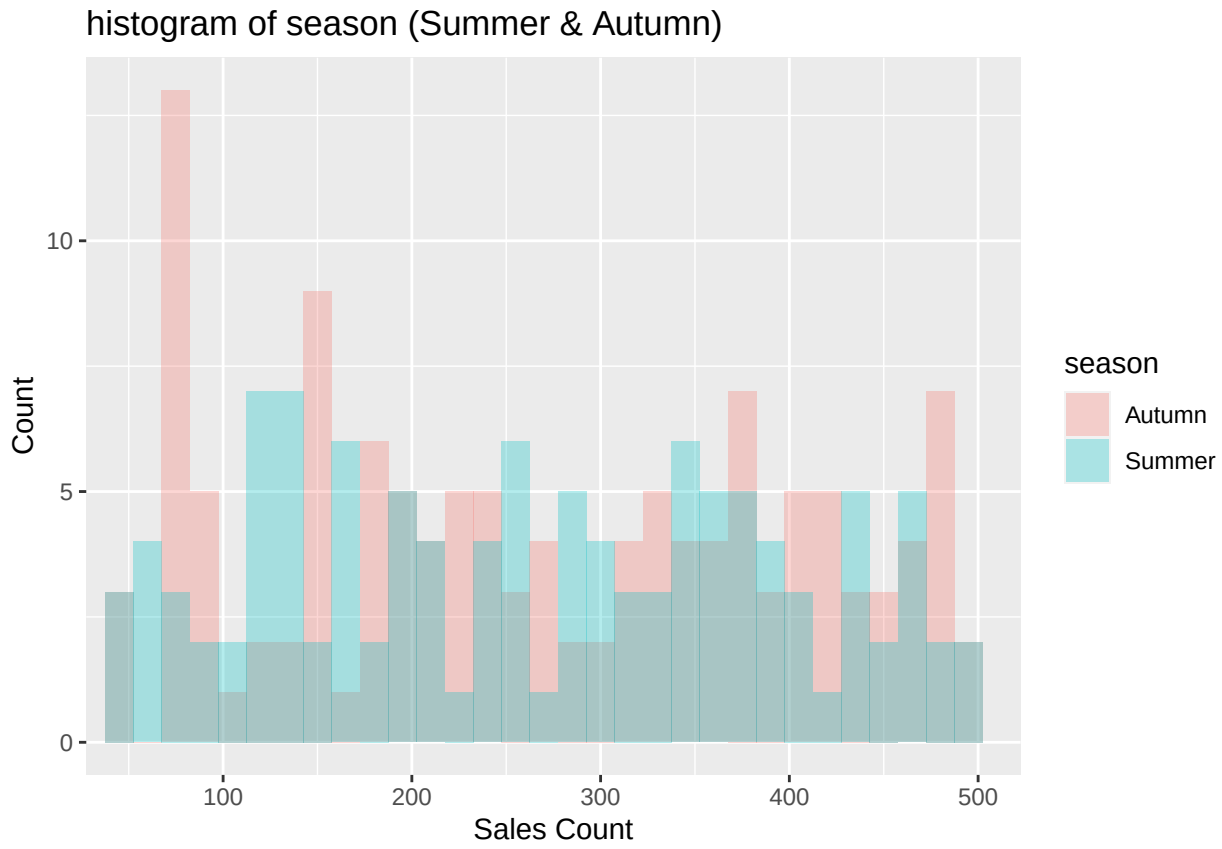
```
visualize(fashion_season_null_7) +
  shade_p_value(obs_stat = fashion_season_obs_stat_7, direction = "two_sided") +
  labs(
  title = "Simulation 7 (Spring & Winter)",
  x = "difference in proportions",
  y = "count"
  )
```



Summer & Autumn

```
fashion_season_8 <- fashion %>%
  filter(season == "Summer" | season == "Autumn")
```

```
fashion_season_8 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (Summer & Autumn)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_8 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Autumn  269.    260  136.    51   494  232.
## 2 Summer  264.    258  128.    50   496  222.
```

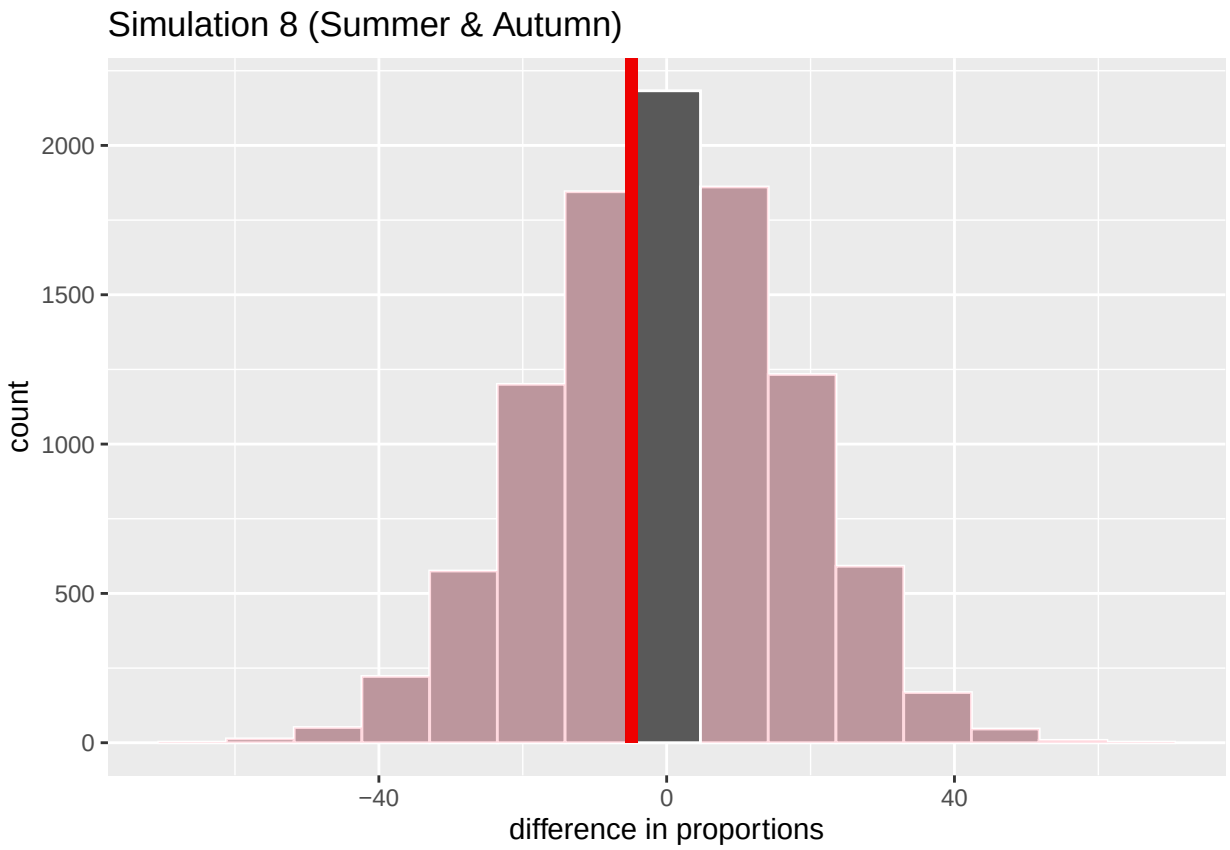
```
fashion_season_null_8 <- fashion_season_8 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Summer", "Autumn"))
```

```
fashion_season_obs_stat_8 <- fashion_season_8 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("Summer", "Autumn"))
```

```
fashion_season_null_8 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_8, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.775
```

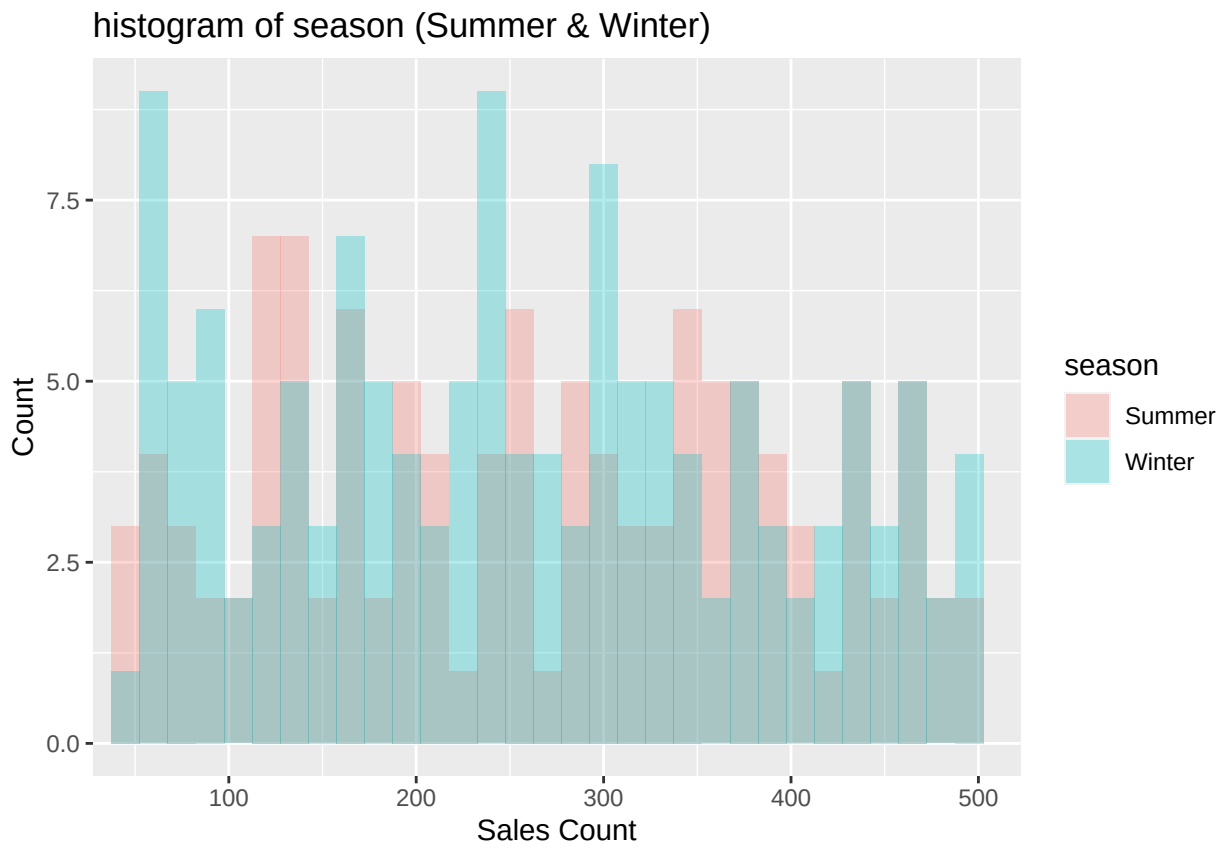
```
visualize(fashion_season_null_8) +
  shade_p_value(obs_stat = fashion_season_obs_stat_8, direction = "two_sided") +
  labs(
  title = "Simulation 8 (Summer & Autumn)",
  x = "difference in proportions",
  y = "count"
  )
```



Summer & Winter

```r
fashion_season_9 <- fashion %>%
  filter(season == "Summer" | season == "Winter")
```

```r
fashion_season_9 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (Summer & Winter)",
  x = "Sales Count",
  y = "Count"
  )
```




```r
fashion_season_9 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Summer  264.   258   128.    50   496  222.
## 2 Winter  258.   248.  130.    52   500  203
```

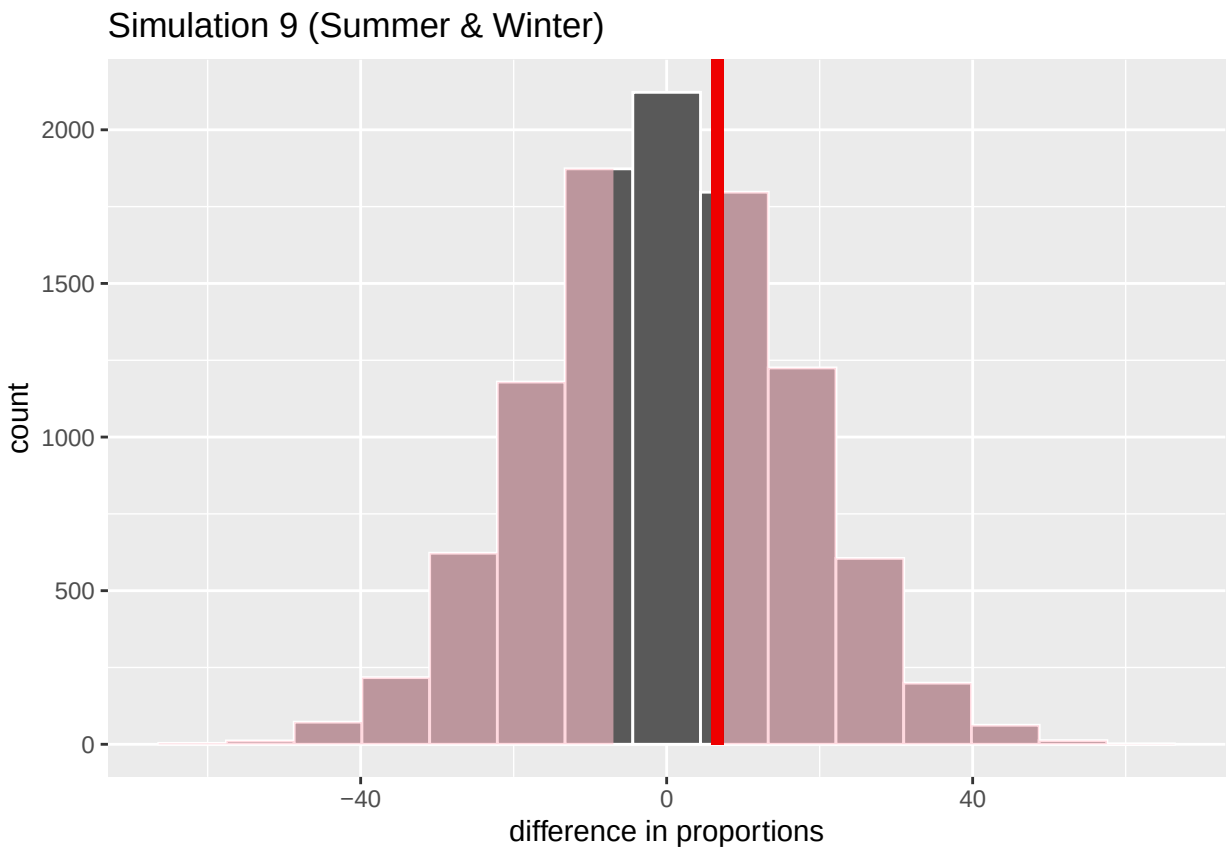
```
fashion_season_null_9 <- fashion_season_9 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Summer", "Winter"))
```

```
fashion_season_obs_stat_9 <- fashion_season_9 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("Summer", "Winter"))
```

```
fashion_season_null_9 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_9, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.684
```

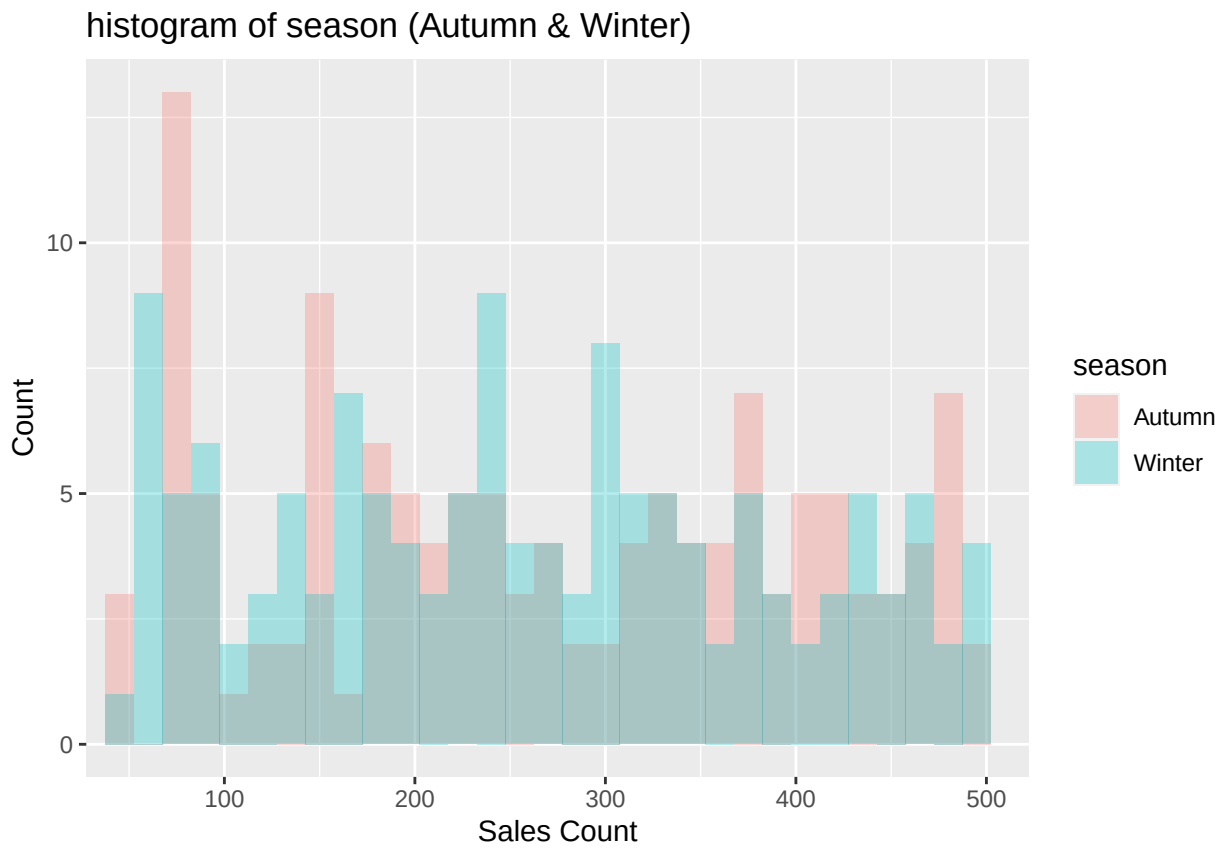
```
visualize(fashion_season_null_9) +
  shade_p_value(obs_stat = fashion_season_obs_stat_9, direction = "two_sided") +
  labs(
  title = "Simulation 9 (Summer & Winter)",
  x = "difference in proportions",
  y = "count"
  )
```



Autumn & Winter

```
fashion_season_10 <- fashion %>%
  filter(season == "Autumn" | season == "Winter")
```

```
fashion_season_10 %>%
  ggplot() +
  geom_histogram(mapping = aes(x = sales_count, fill = season),
  binwidth = 15, alpha = 0.3, position = "identity") +
  coord_cartesian(xlim = c(50,500)) +
  labs(
  title = "histogram of season (Autumn & Winter)",
  x = "Sales Count",
  y = "Count"
  )
```




```
fashion_season_10 %>%
  group_by(season) %>%
  summarize(
    mean = mean(sales_count),
    median = median(sales_count),
    sd = sd(sales_count),
    min = min(sales_count),
    max = max(sales_count),
    IQR = IQR(sales_count)
  )
```

```
## # A tibble: 2 x 7
##   season  mean median    sd   min   max   IQR
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Autumn  269.   260   136.    51   494  232.
## 2 Winter  258.   248.  130.    52   500  203
```

```
fashion_season_null_10 <- fashion_season_10 %>%
  specify(formula = sales_count ~ season) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("Autumn", "Winter"))
```

```
fashion_season_obs_stat_10 <- fashion_season_10 %>%
  specify(sales_count ~ season) %>%
  calculate(stat = "diff in means", order = c("Autumn", "Winter"))
```

```
fashion_season_null_10 %>%
  get_p_value(obs_stat = fashion_season_obs_stat_10, direction = "two_sided")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.481
```

```
visualize(fashion_season_null_10) +
  shade_p_value(obs_stat = fashion_season_obs_stat_10, direction = "two_sided") +
  labs(
  title = "Simulation 10 (Autumn & Winter)",
  x = "difference in proportions",
  y = "count"
  )
```