

# Consumer Characteristics vs Number of Sales of Clothes

Jin Heo, Minseo Yoon, Seunghyun Ryu,  
ChaeJun Lim, Juheon Kim, Hyukjoon Choi

# Contents

## 1 Question of Interest

- Introduction of dataset

## 3 Visualization & Analysis

- Hypothesis 1
- Hypothesis 2
- Hypothesis 3

## 2 Preprocessing

- General relationship of variables in the dataset

## 4 Conclusions

- Result
- Implication

1

# Question of Interest

Introduction of dataset

# QUESTION OF INTEREST



**Spanish** Fast-Fashion Brand  
Founded in 1974 by Amancio  
Ortega

## WHY FASHION?

- Common interest
- Related to our daily life
- Member interested in fashion career

## WHY ZARA?

- Zara: well known brand in Korea
- Often go to Zara in Triple Street

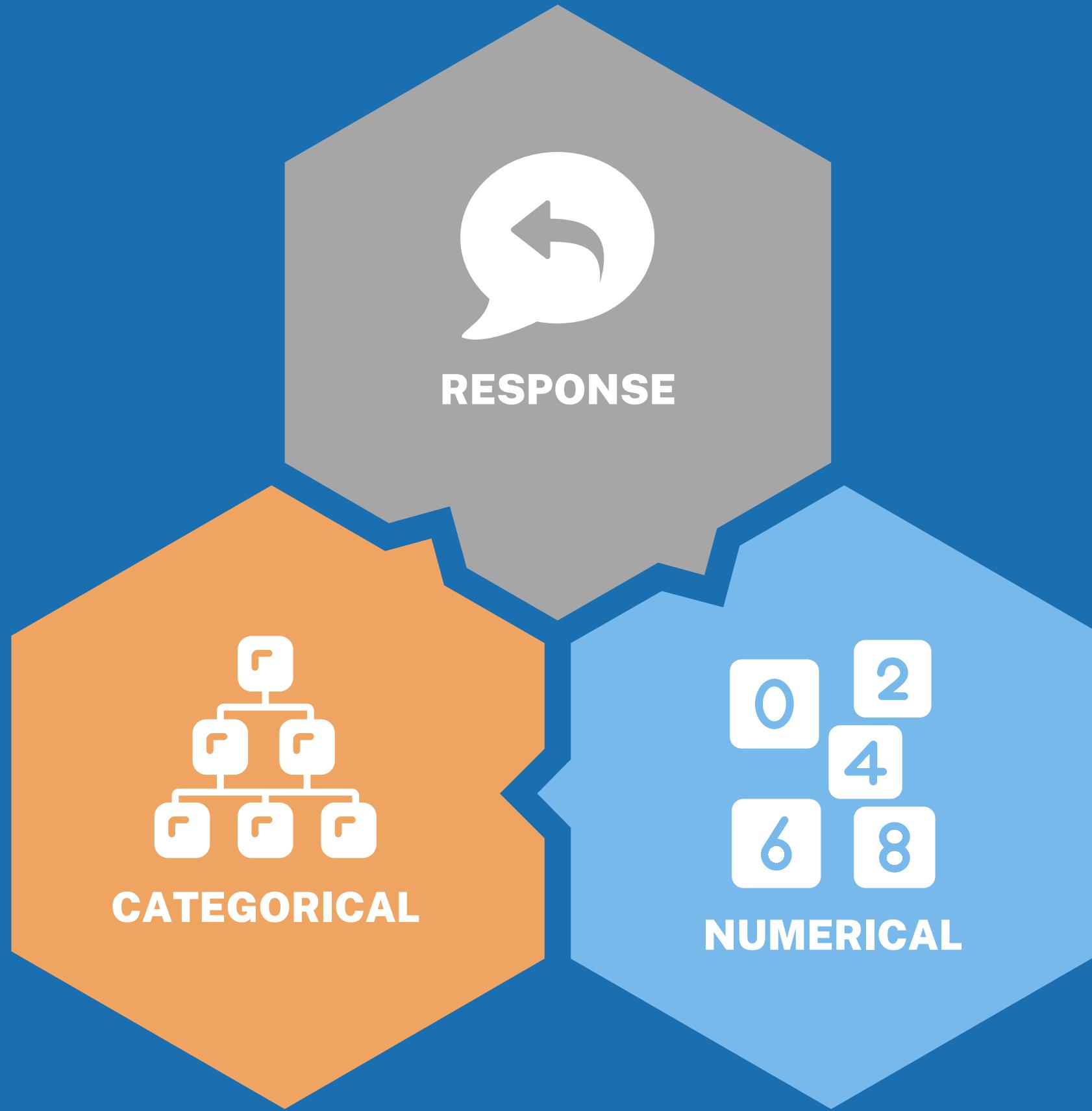
**DATA:**  
**FASHION\_2018\_2022**  
**(ZARA)**

**SOURCE : KAGGLE**  
by Fashion World DA

# Research Question

What is the **relationship** between **consumer characteristics** and the **number of sales** of clothes?

# TYPES OF VARIABLES



1

**RESPONSE**  
`sales_count`

3

**NUMERICAL**  
`price, reviews_count, wish_list_count`

10

**CATEGORICAL**  
`gender, category, pattern, color, age_group, season, material, average_rating, discount, out_of_stock_times`

**EXCEL**

|    | A          | B              | C     | D               | E             | F            | G         | H              | I        | J                  |
|----|------------|----------------|-------|-----------------|---------------|--------------|-----------|----------------|----------|--------------------|
| 1  | product_id | product_name   | brand | last_stock_date | month_of_sale | year_of_sale | age_group | average_rating | discount | out_of_stock_times |
| 2  | 1001       | Biker Jacket   | ZARA  | 1/28/2018       | 1             | 2018         | 25-35     | 4.9            | 20%      | 3                  |
| 3  | 1002       | Business Shirt | ZARA  | 1/28/2018       | 1             | 2018         | 18-24     | 3.5            | 0%       | 6                  |
| 4  | 1003       | Wool Jacket    | ZARA  | 1/7/2018        | 1             | 2018         | 18-24     | 4.3            | 5%       | 4                  |
| 5  | 1004       | Summer Dress   | ZARA  | 1/1/2018        | 1             | 2018         | 25-35     | 4.6            | 10%      | 3                  |
| 6  | 1005       | Casual Jeans   | ZARA  | 1/28/2018       | 1             | 2018         | 35-45     | 3.6            | 0%       | 4                  |
| 7  | 1006       | Business Shirt | ZARA  | 1/16/2018       | 1             | 2018         | 35-45     | 4.7            | 15%      | 6                  |
| 8  | 1007       | Casual Jeans   | ZARA  | 1/2/2018        | 1             | 2018         | 25-35     | 4              | 10%      | 3                  |
| 9  | 1008       | Heels          | ZARA  | 1/3/2018        | 1             | 2018         | 25-35     | 3.2            | 0%       | 3                  |
| 10 | 1009       | Beach Shirt    | ZARA  | 1/14/2018       | 1             | 2018         | 35-45     | 3.5            | 10%      | 6                  |
| 11 | 1010       | Pattern Skirt  | ZARA  | 1/21/2018       | 1             | 2018         | 35-45     | 4.1            | 0%       | 6                  |
| 12 | 1011       | Business Shirt | ZARA  | 1/1/2018        | 1             | 2018         | 25-35     | 3.7            | 5%       | 2                  |
| 13 | 1012       | Blouse         | ZARA  | 2/12/2018       | 2             | 2018         | 18-24     | 4.6            | 0%       | 5                  |
| 14 | 1013       | Chino Shorts   | ZARA  | 2/19/2018       | 2             | 2018         | 25-35     | 4.4            | 20%      | 4                  |
| 15 | 1014       | Pattern Skirt  | ZARA  | 2/21/2018       | 2             | 2018         | 18-24     | 4              | 10%      | 5                  |
| 16 | 1015       | Blouse         | ZARA  | 2/8/2018        | 2             | 2018         | 35-45     | 3.7            | 10%      | 3                  |
| 17 | 1016       | Blouse         | ZARA  | 2/10/2018       | 2             | 2018         | 35-45     | 4.6            | 10%      | 6                  |
| 18 | 1017       | Pattern Skirt  | ZARA  | 2/24/2018       | 2             | 2018         | 18-24     | 3              | 5%       | 6                  |
| 19 | 1018       | Summer Dress   | ZARA  | 2/18/2018       | 2             | 2018         | 18-24     | 3.3            | 5%       | 6                  |
| 20 | 1019       | Casual Jeans   | ZARA  | 2/6/2018        | 2             | 2018         | 18-24     | 4.8            | 15%      | 4                  |
| 21 | 1020       | Pattern Skirt  | ZARA  | 2/25/2018       | 2             | 2018         | 35-45     | 3.8            | 0%       | 2                  |
| 22 | 1021       | Blouse         | ZARA  | 2/22/2018       | 2             | 2018         | 35-45     | 3.6            | 10%      | 1                  |
| 23 | 1022       | Blouse         | ZARA  | 2/9/2018        | 2             | 2018         | 18-24     | 3.9            | 15%      | 6                  |
| 24 | 1023       | Heels          | ZARA  | 3/11/2018       | 3             | 2018         | 25-35     | 4.2            | 15%      | 2                  |

# EXCLUDED VARIABLES



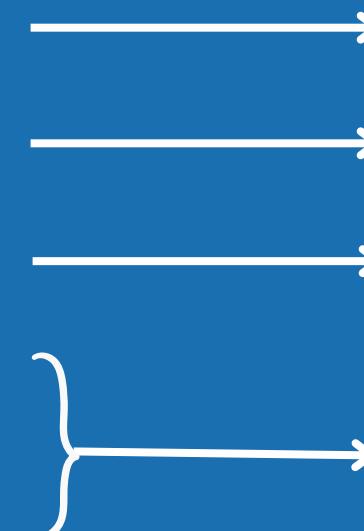
**NUMERICAL**  
last\_stock date



**Not Relevant to Data**



**CATEGORICAL**  
brand,  
product\_id,  
product\_name,  
month\_of\_sales,  
years\_of\_sales,



All **Same Identification** of Product  
Sub-version of “**category**”  
**Not Relevant to Data**



2

# Preprocessing

General relationship of variables in dataset

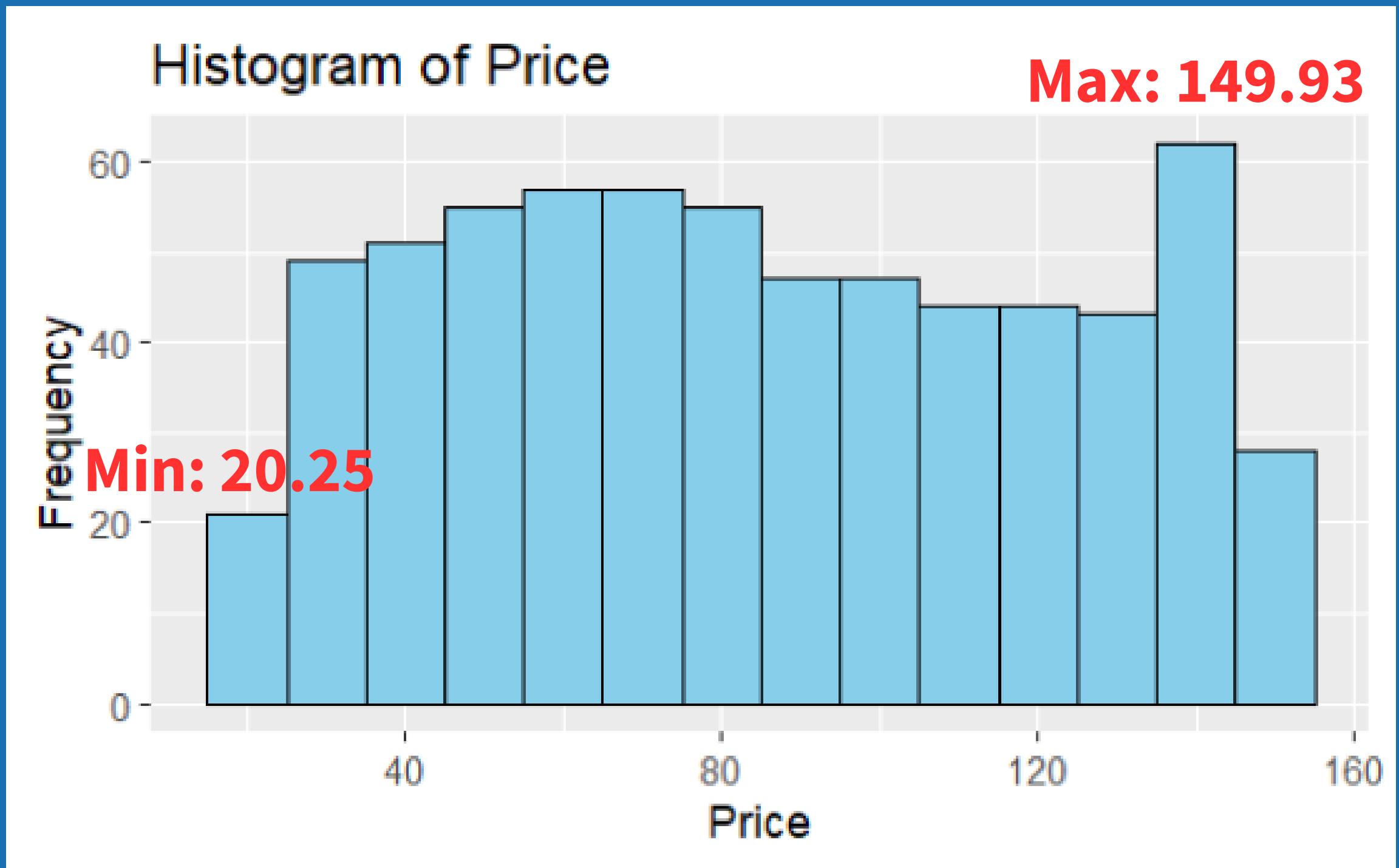
PRICE



Fast Fashion Brand

## HISTOGRAM

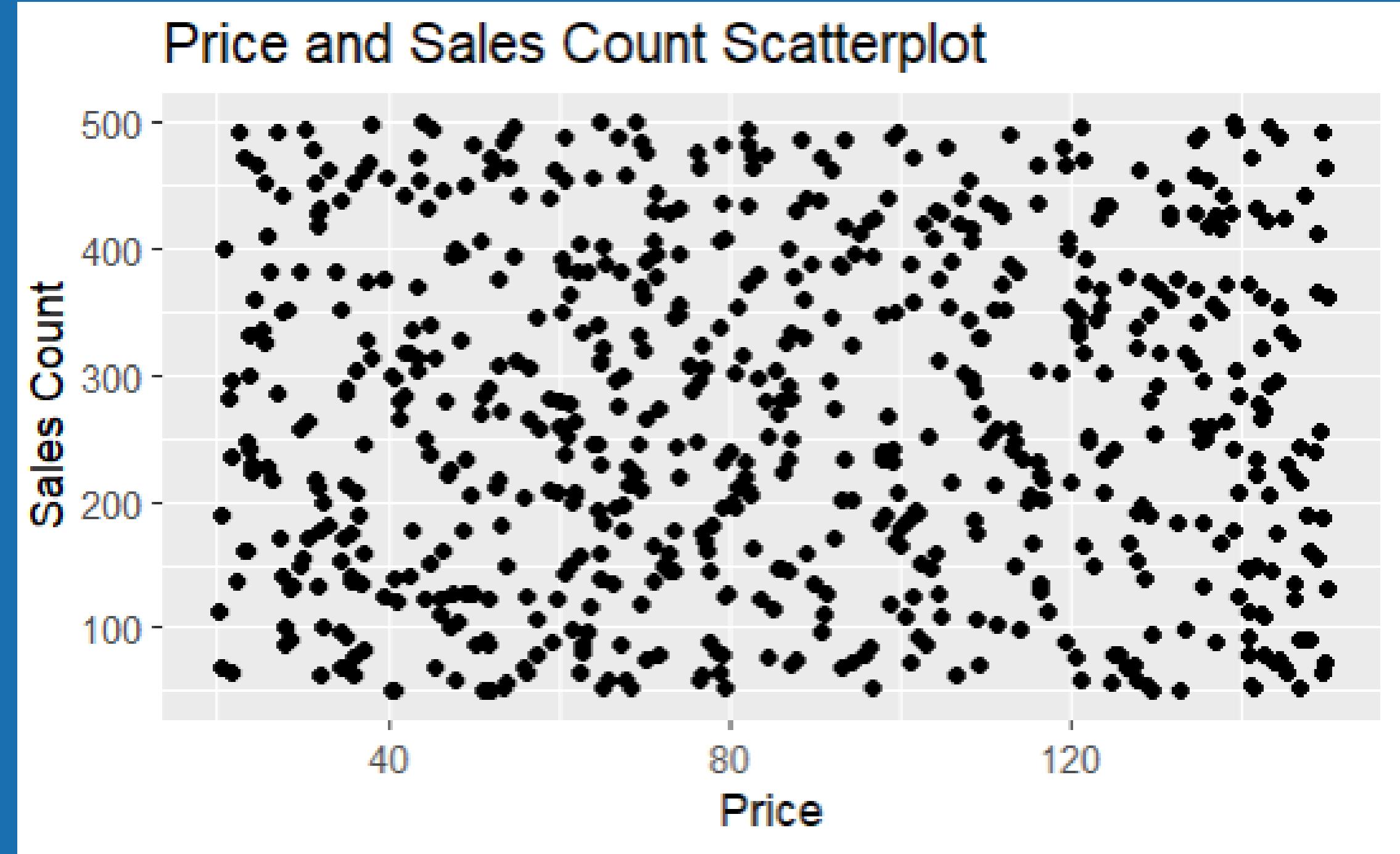
- Bimodal
- Not skewed



# PRICE

## SCATTERPLOT

- No correlation found
- Points are scattered without any discernible pattern



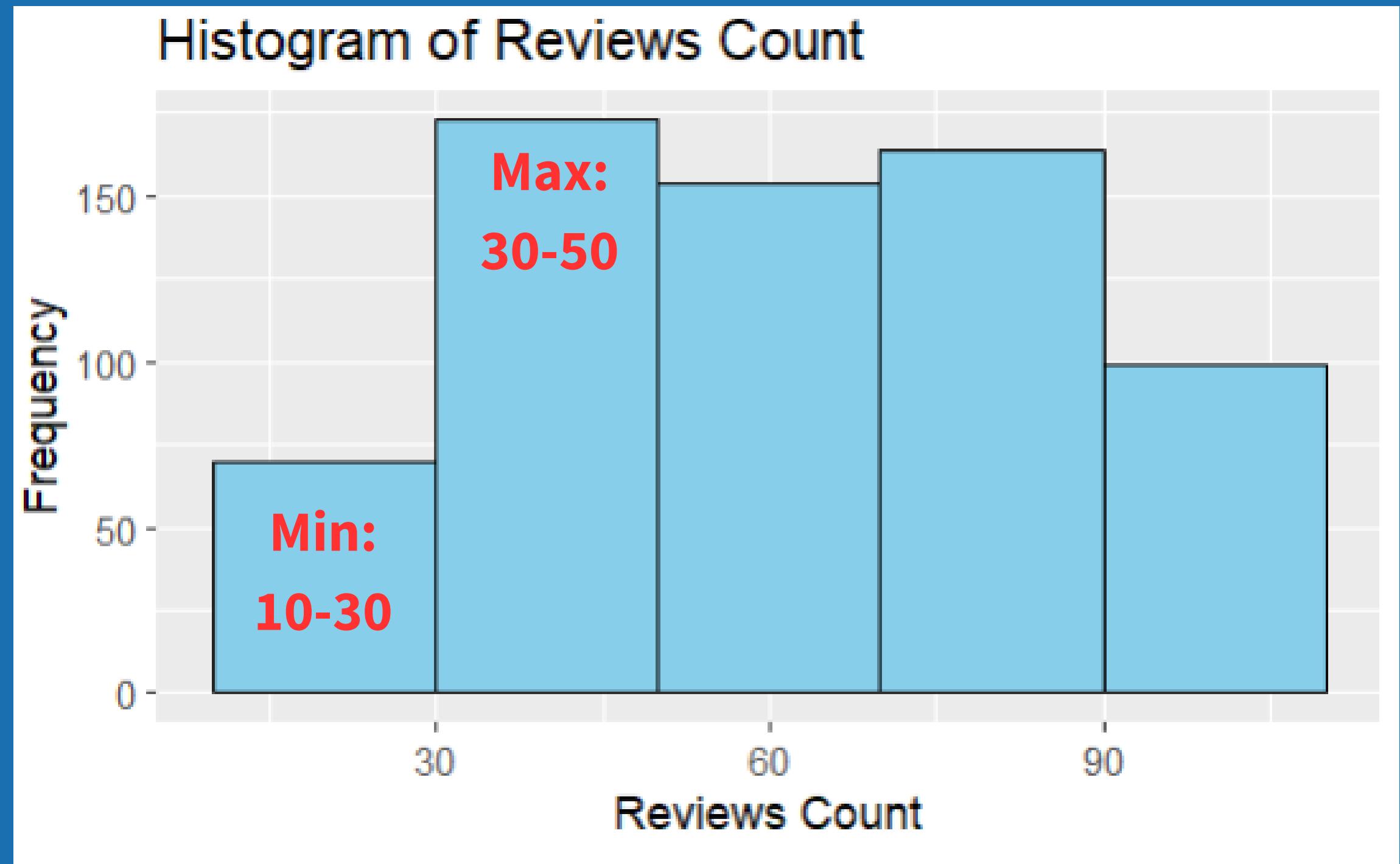
# REVIEWS\_COUNT



## High Product Turnover

### HISTOGRAM

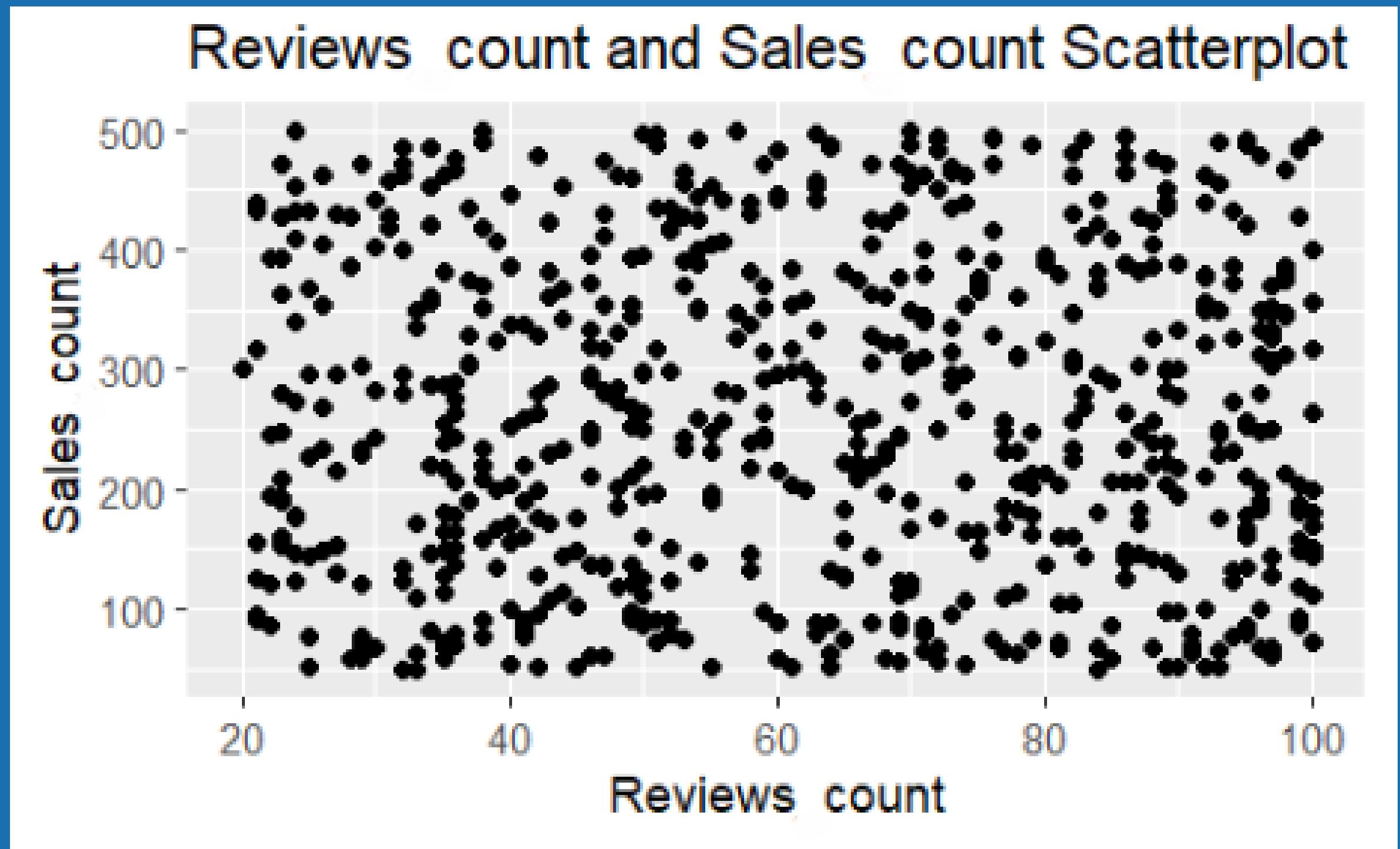
- Non-symmetric  
**bimodal** distribution



# REVIEWS\_COUNT

## SCATTERPLOT

- No correlation found



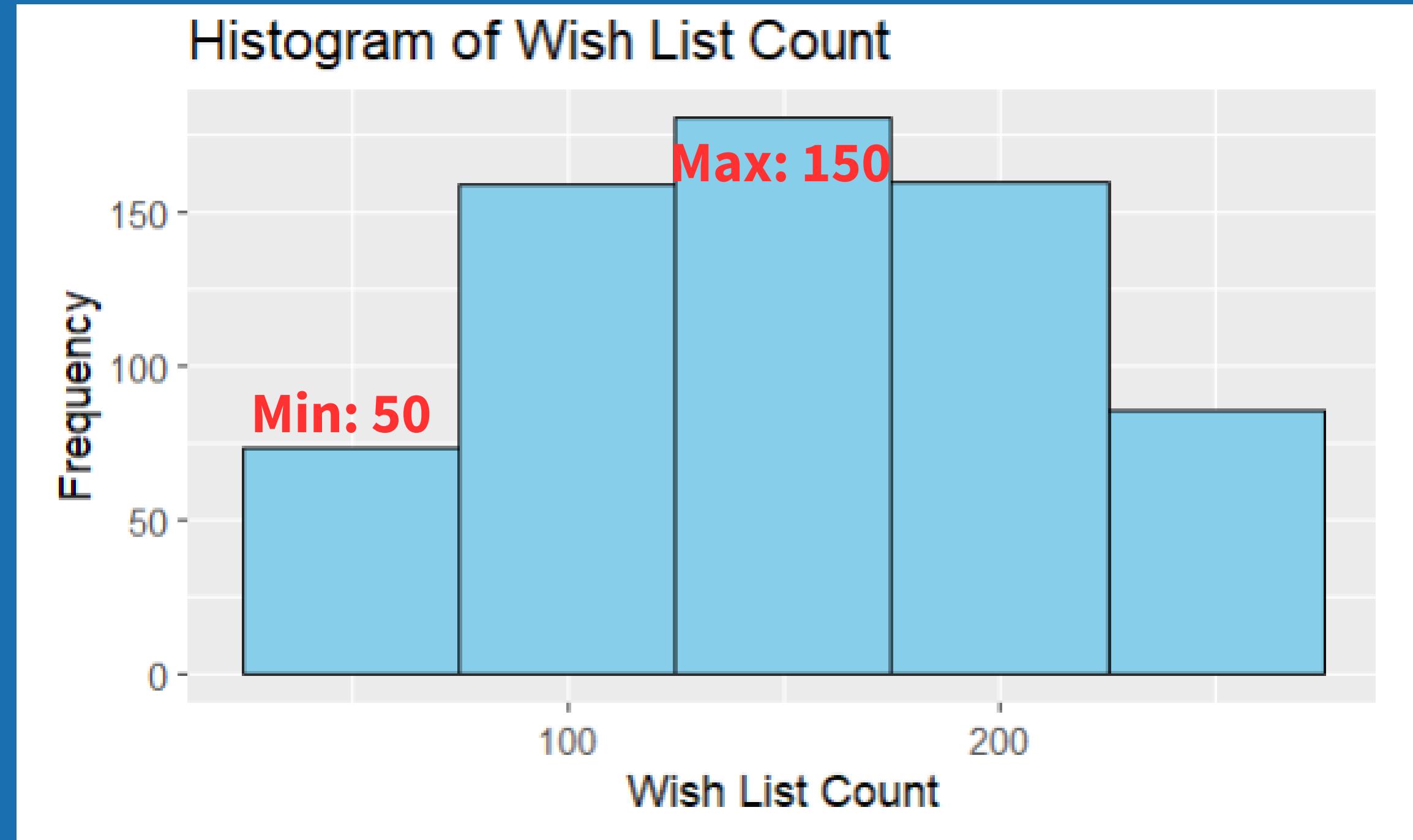
# WISH\_LIST\_COUNT



## Moderate Popularity

### HISTOGRAM

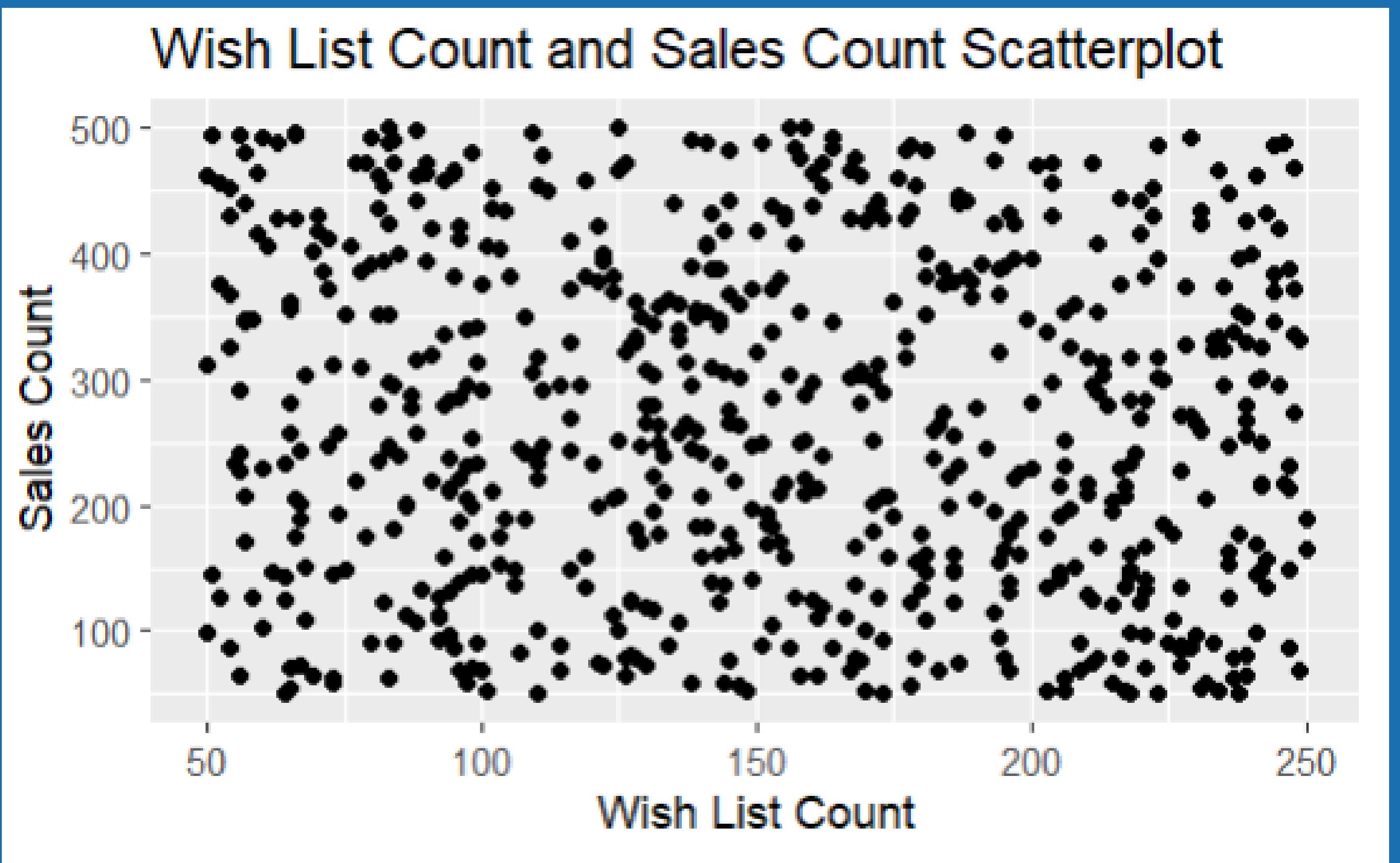
- Symmetric Unimodal



# WISH\_LIST\_COUNT

## SCATTERPLOT

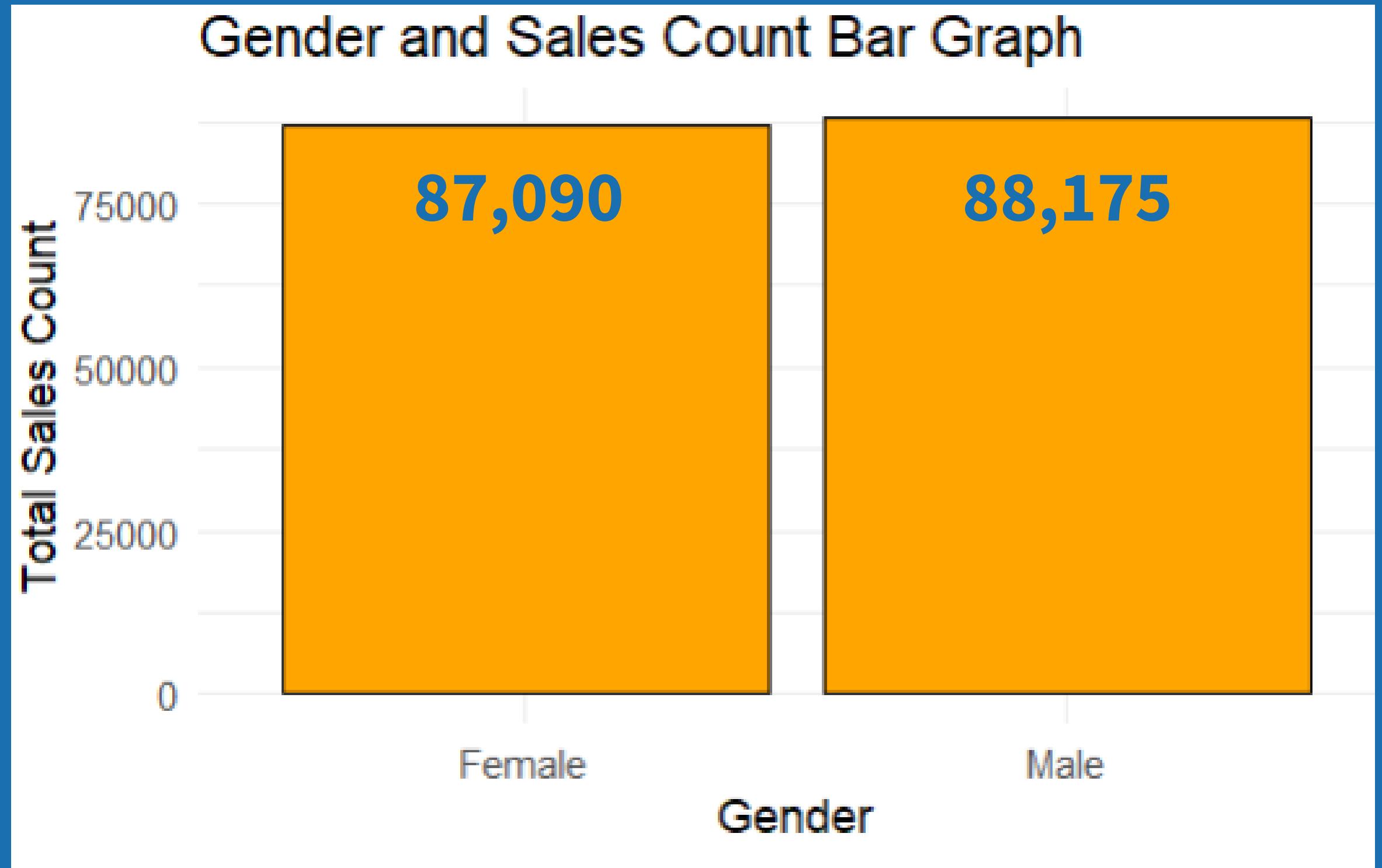
- No correlation found



# GENDER

## BAR GRAPH

- Male has slightly **higher** total sales count.



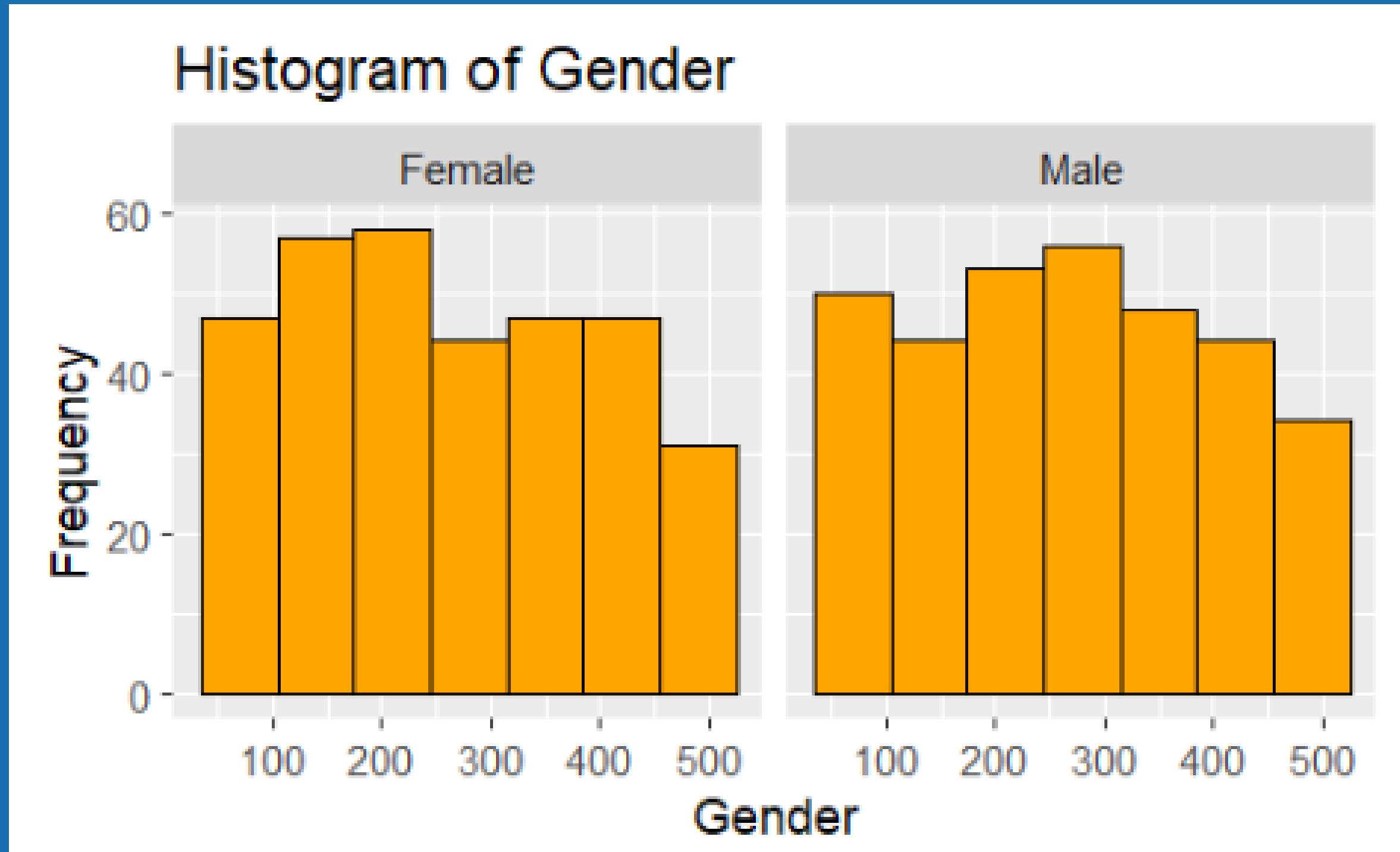
# GENDER



## Balanced Marketing & Product Offering

### HISTOGRAM

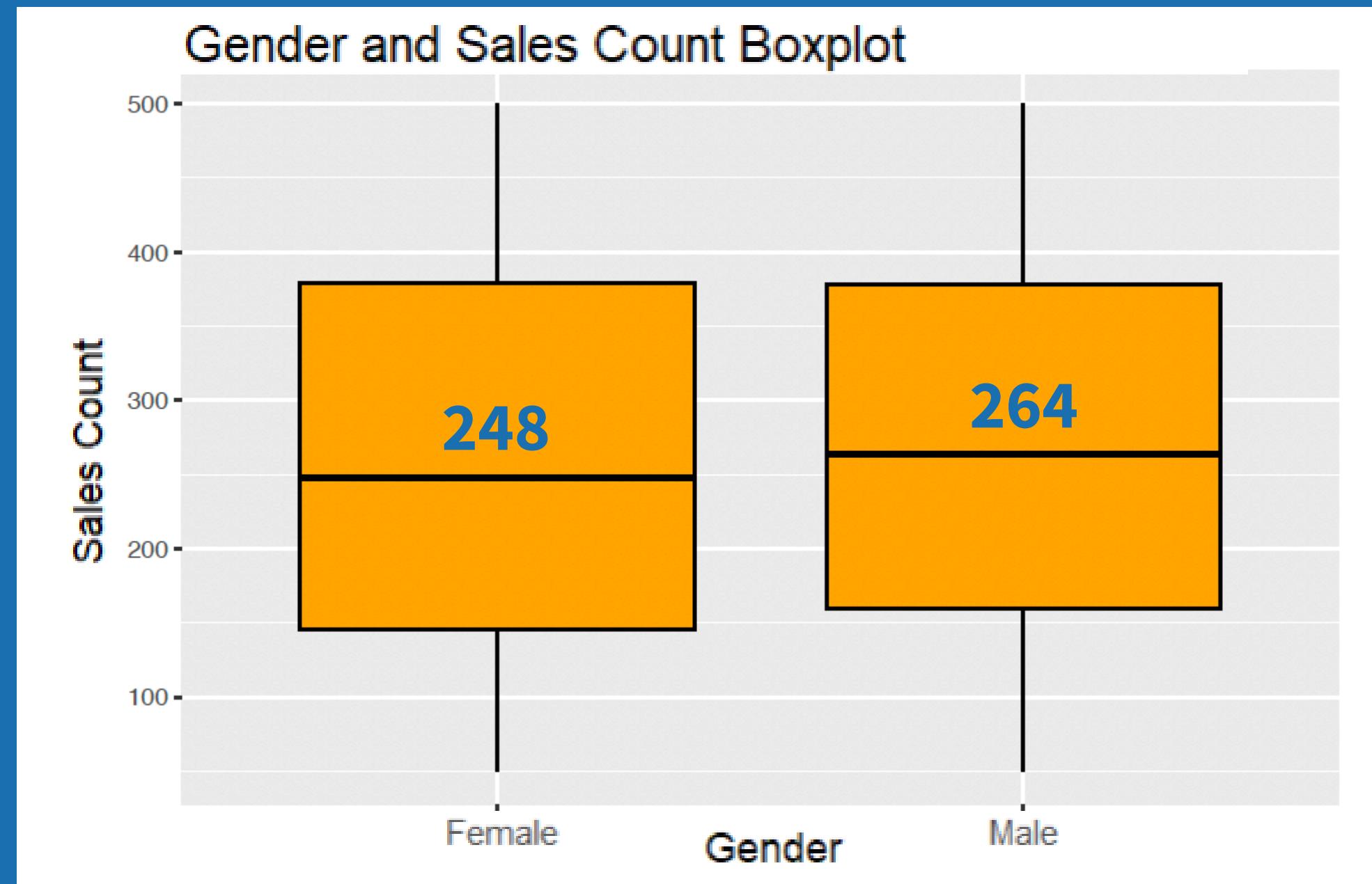
- **Highest Frequent:**
  - 180~230(Female) / 250~320(Male)
- **Least Frequent:**
  - 450~500(Female/Male)
- **Female:**
  - **Symmetric Bimodality**
- **Male:**
  - **Unimodality**



# GENDER

## BOX PLOT

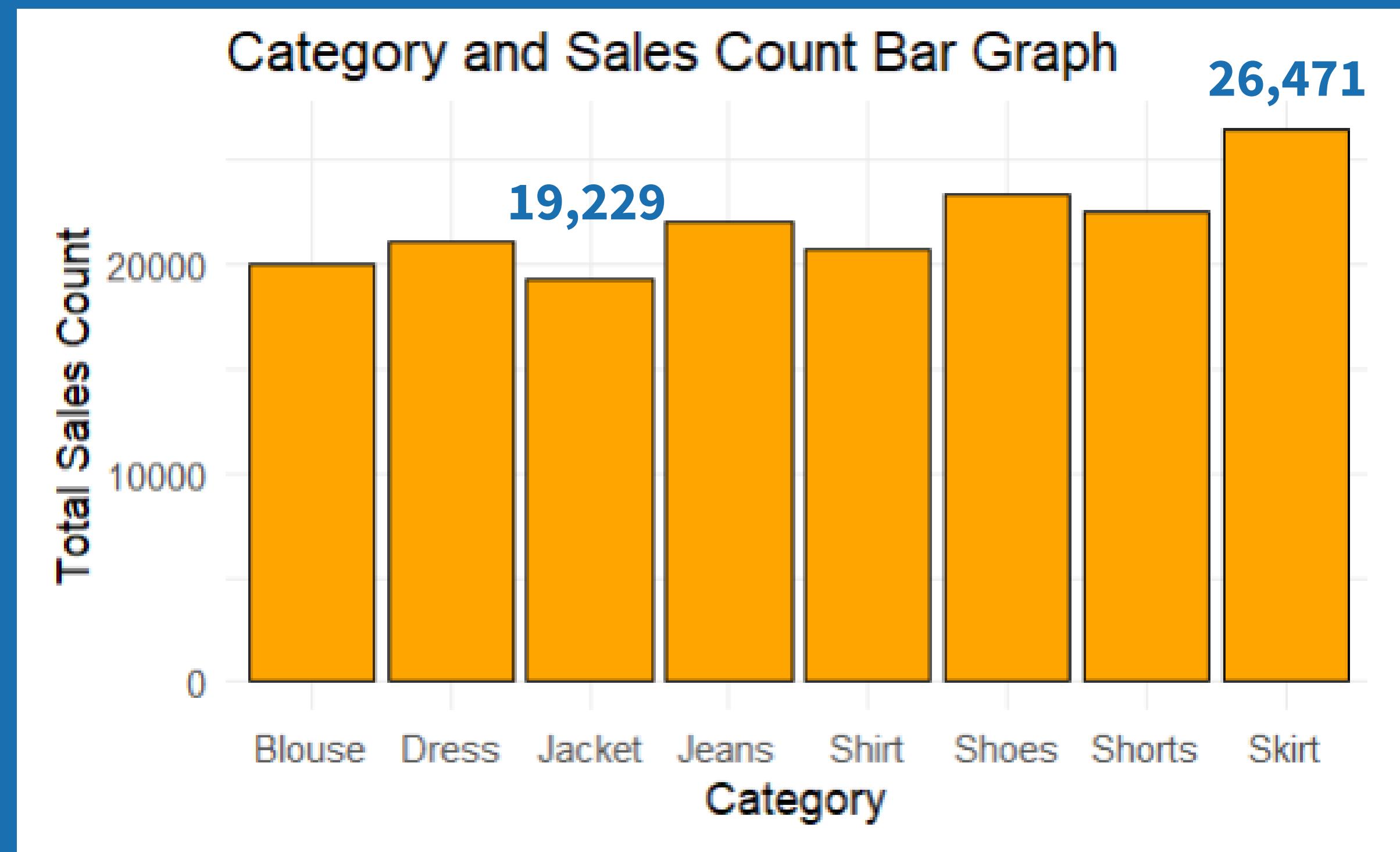
- Male: Higher median & Narrow IQR
  - Consistent purchasing with less variation in products
- Female: Lower median & Wider IQR
  - Larger variety of products  
→ some high sale, others do not



# CATEGORY

## BAR GRAPH

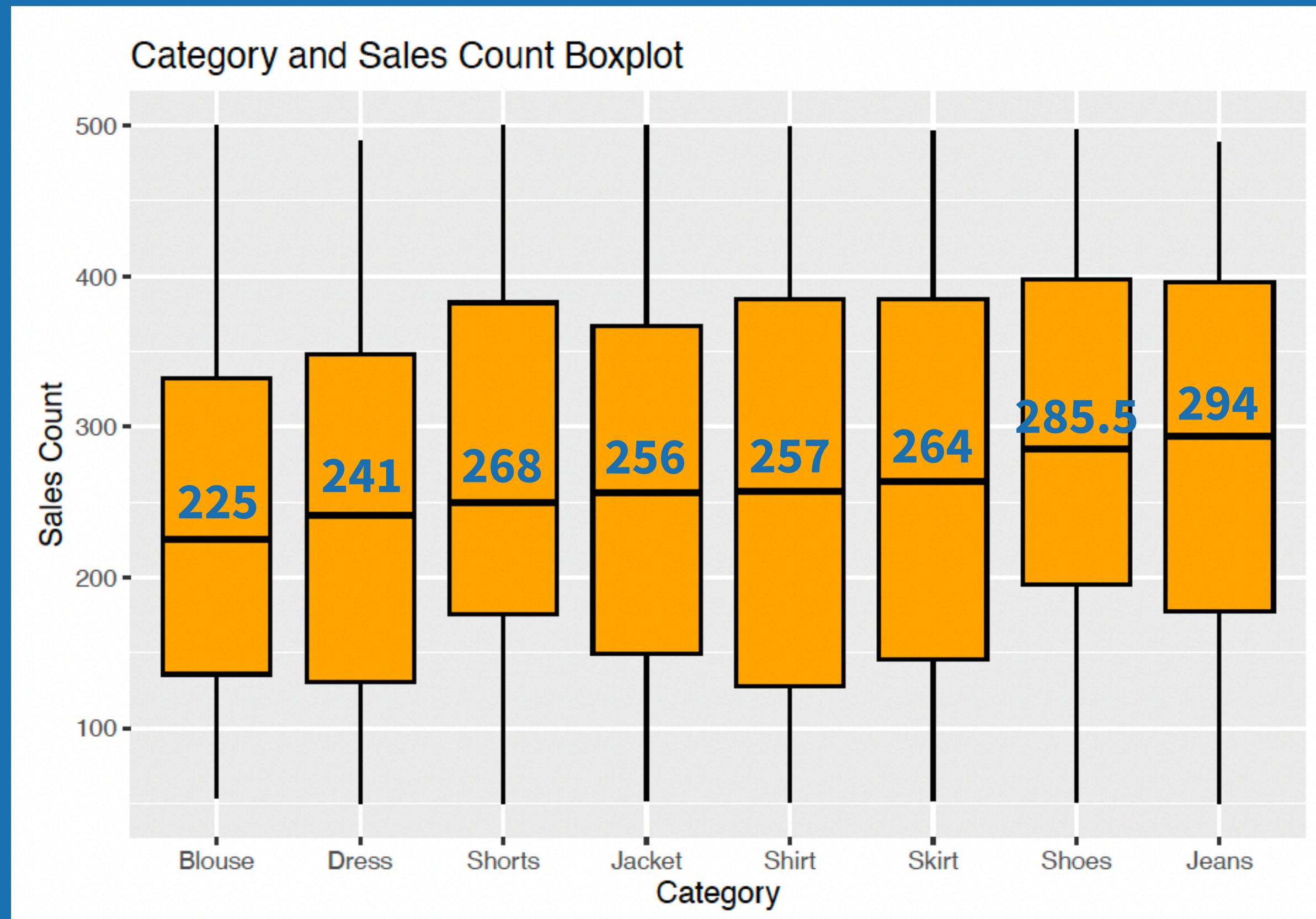
- Total sales counts across different categories between 19,000 & 27,000.
- Highest: Skirt
- Lowest: Jacket



# CATEGORY

## BOX PLOT

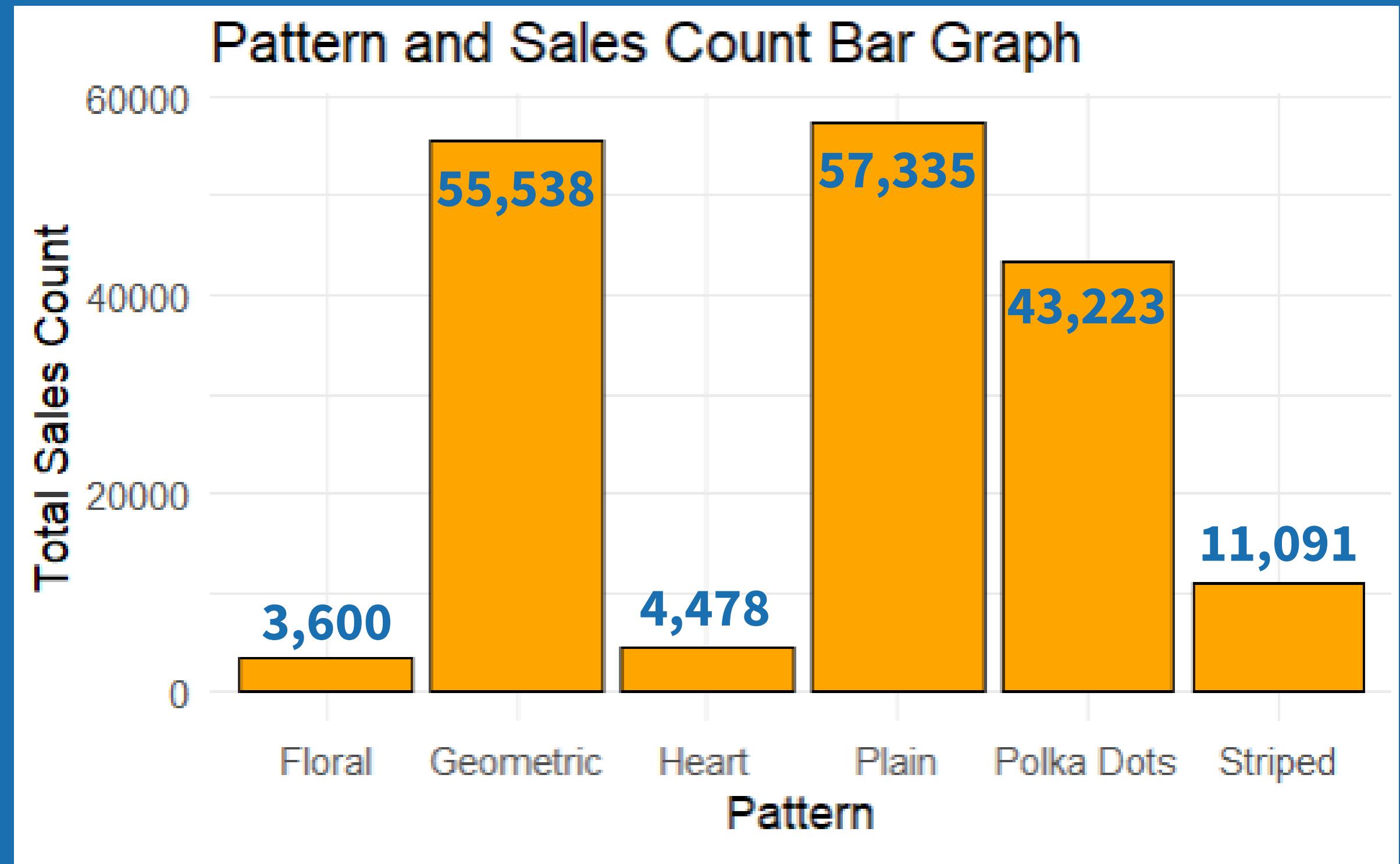
- No outliers
- Highest Median Value:
  - Jeans
- Lowest Median Value:
  - Blouse
- Medians: 200-300



# PATTERN

## BAR GRAPH

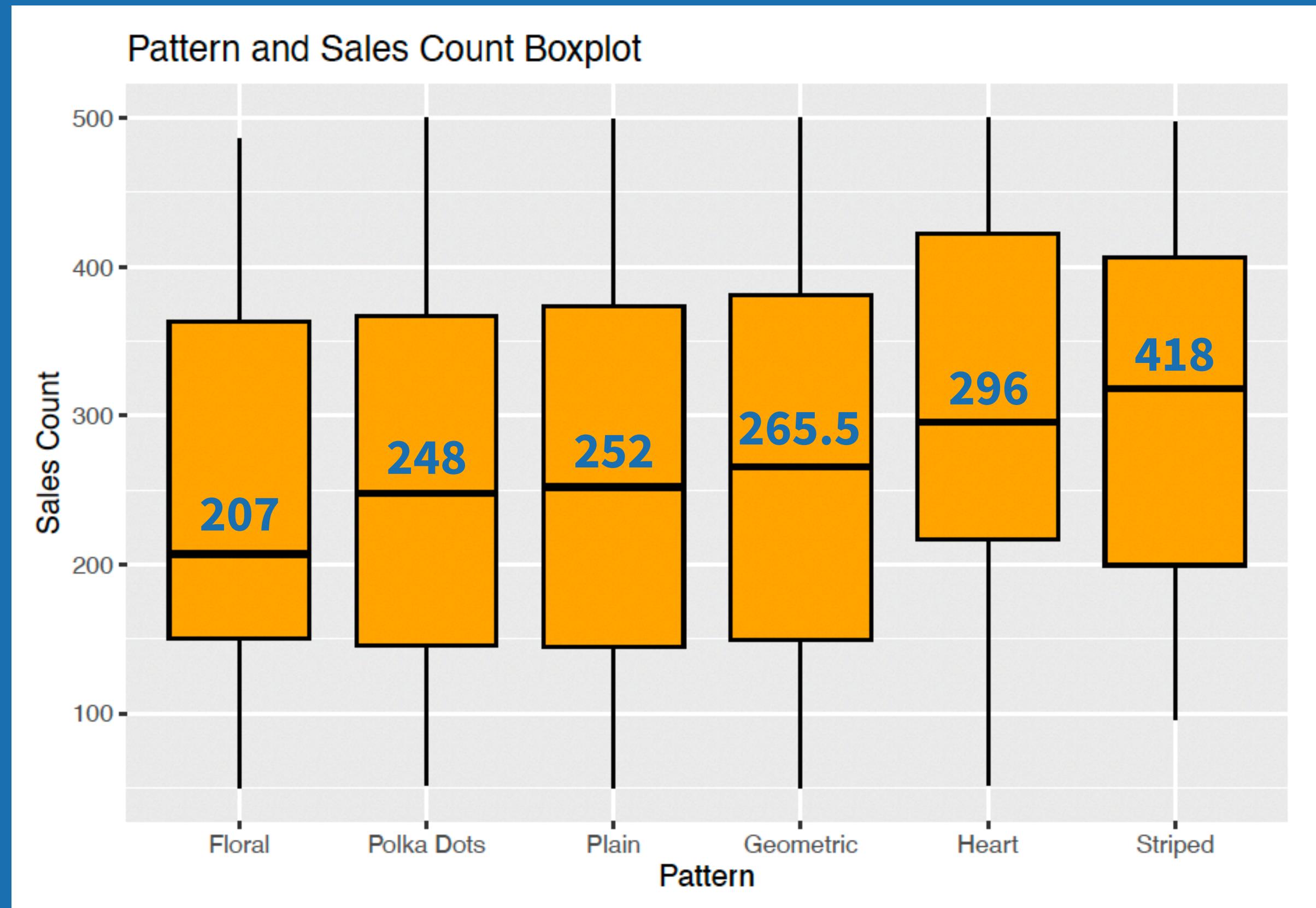
- Certain **patterns** have **higher** sales volume than other
- **Highest:** Plain
- **Lowest:** Floral



# PATTERN

## BOX PLOT

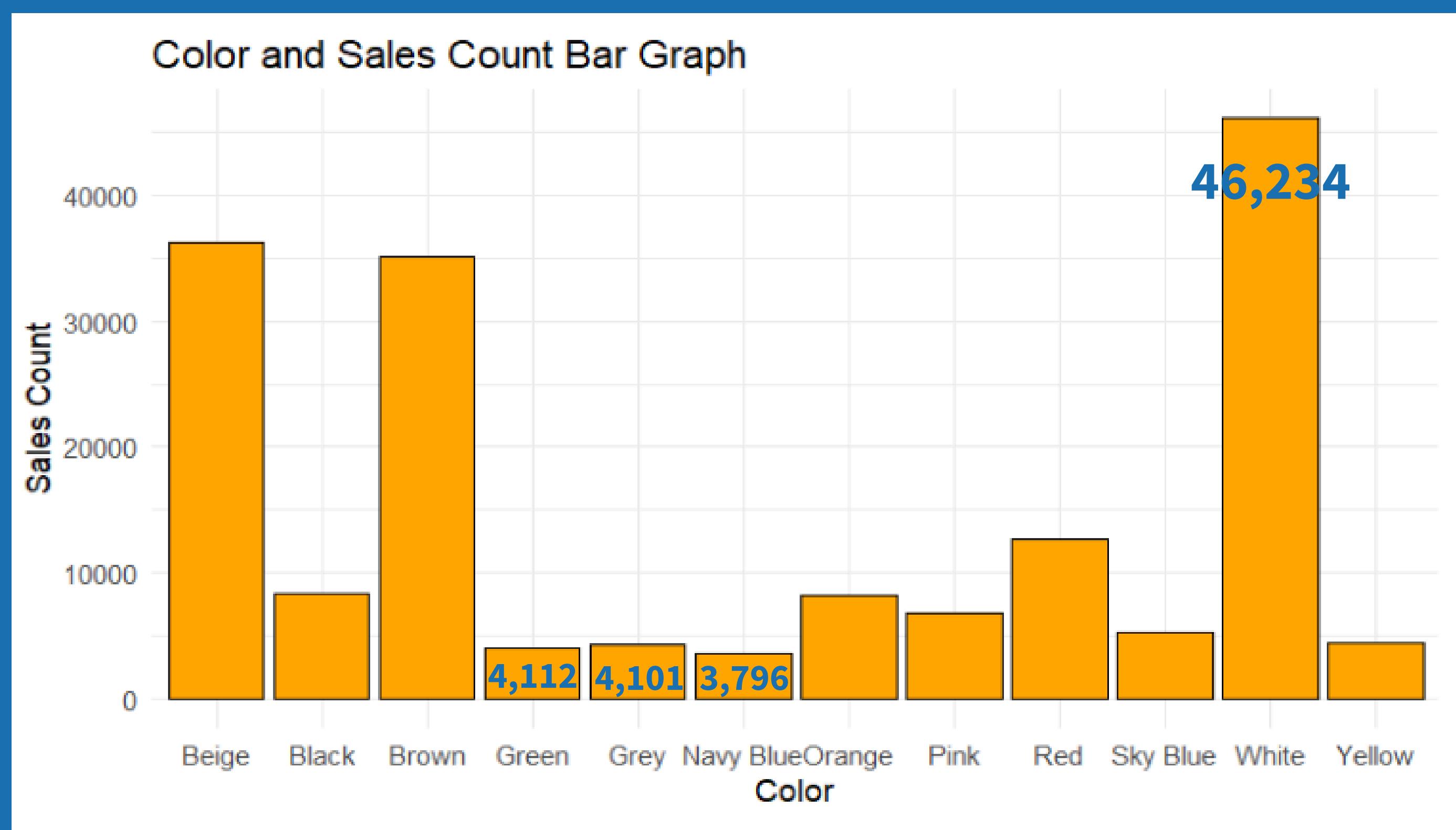
- Highest Median Value:  
Striped
- Lowest Median Value:  
Floral



# COLOR

## BAR GRAPH

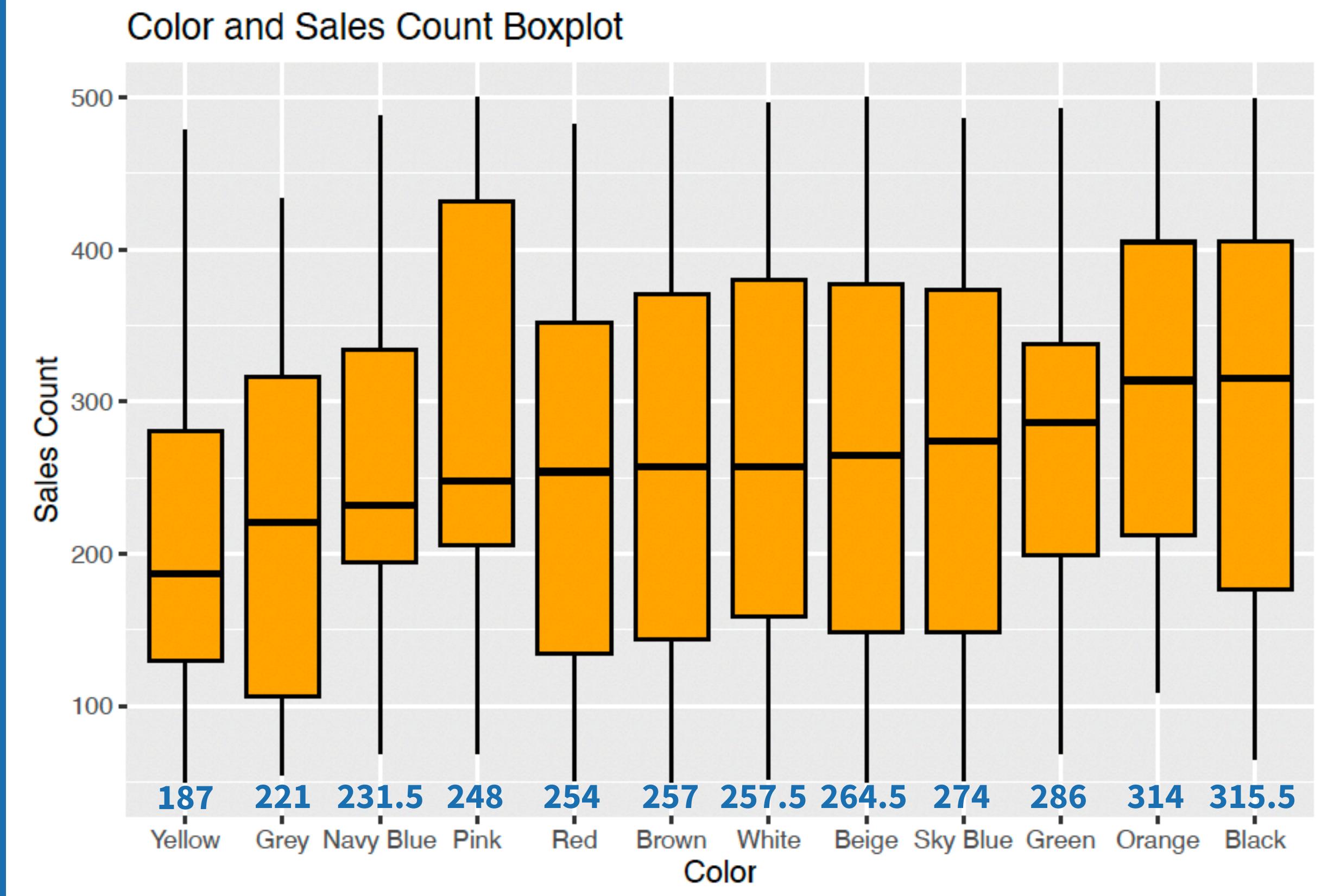
- **Highest:**
  - White
- **High:**
  - Beige, Brown
- **Low:**
  - other values



# COLOR

## BOXPLOT

- **Highest Value:** Black
- **Lowest Value:** Yellow
- **No outliers**



→ Basic colors are more preferred than seasonal colors

# AGE GROUP

## Younger adults:

- more fashion-conscious & experiment trends

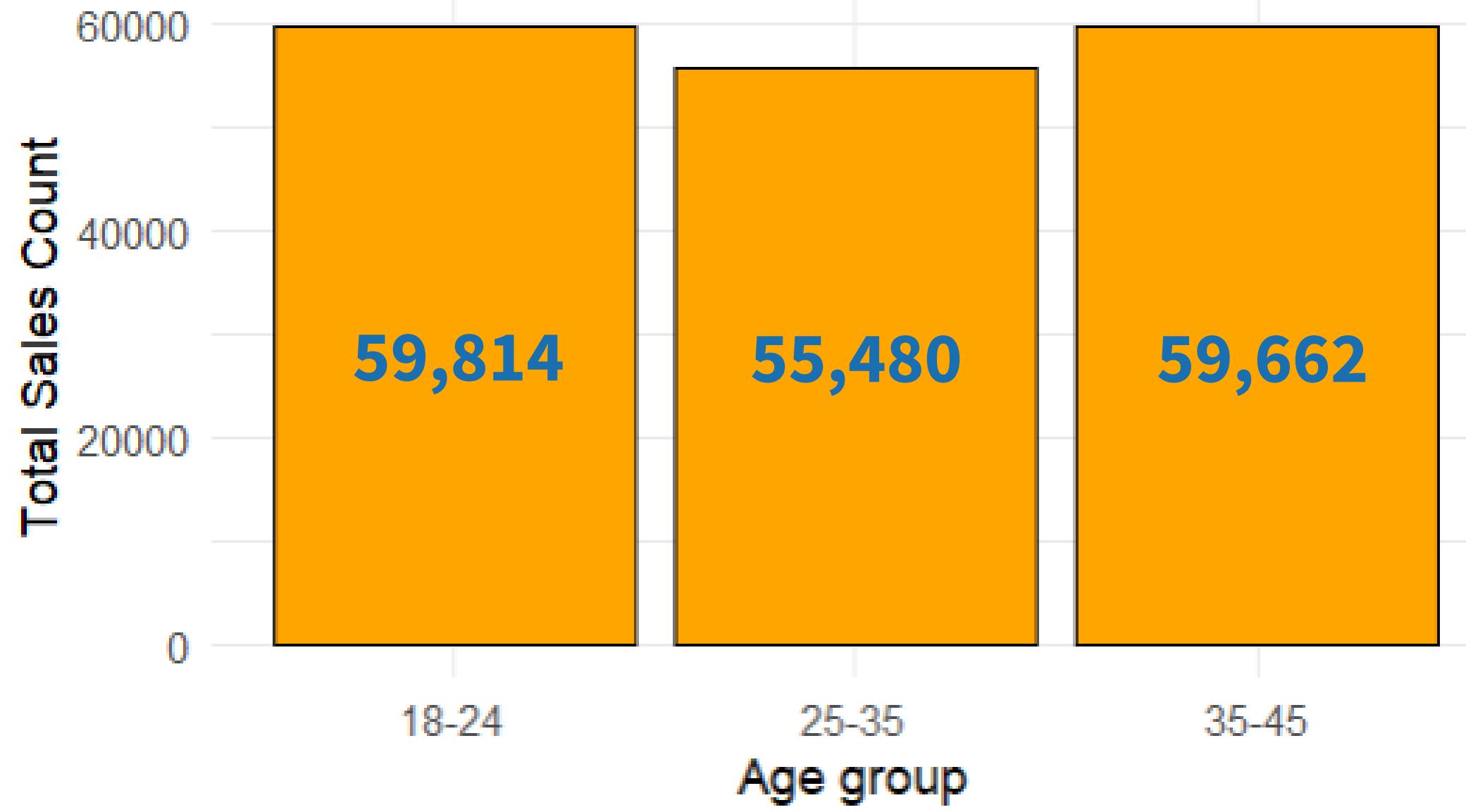
## 25-35:

- transitional phase of starting families or investing in careers

## Older adults:

- More purchasing power & seek quality & style

Age group and Sales Count Bar Graph



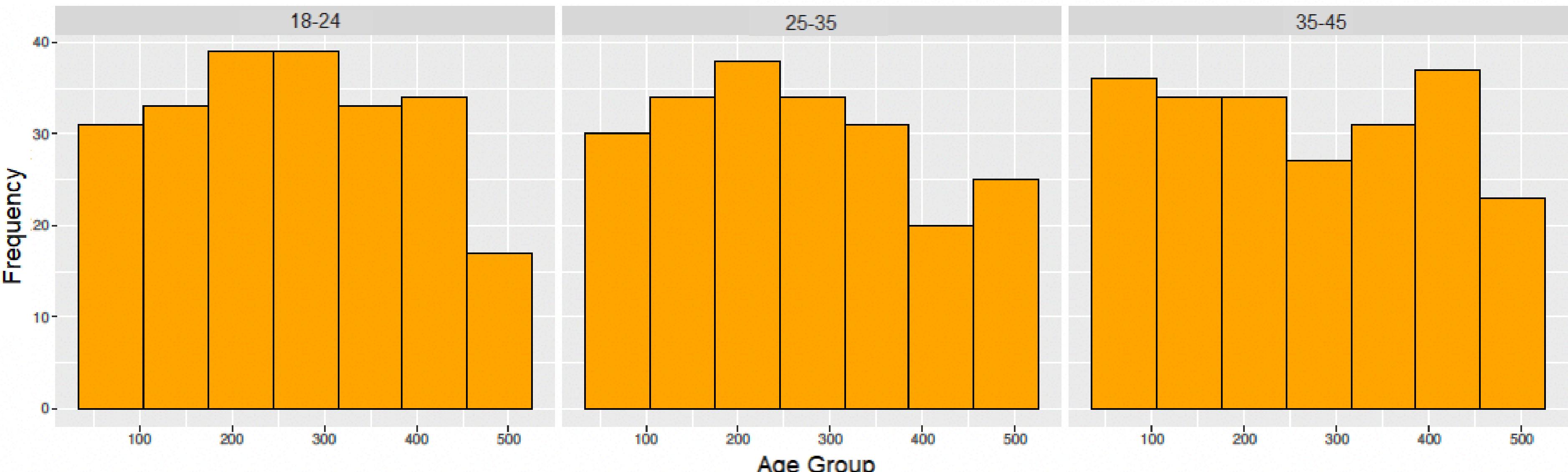
## BAR GRAPH

- Each age group shows **similar** sales\_count
- **25-35** has the **lowest** sales\_count compared to others

# HISTOGRAM

- **18-24:** varied purchasing depending on trend & budget
- **25-35:** steady & moderate purchasing, possibly due to financial responsibility
- **35-45:** High purchasing power

Histogram of Age Group



Multimodal

Slightly right-skewed unimodal

Multimodal

# AGE GROUP

## BOXPLOT

**18-24:** **Highest** median, Moderate IQR

- Purchase frequently, consistent level of purchasing across different products

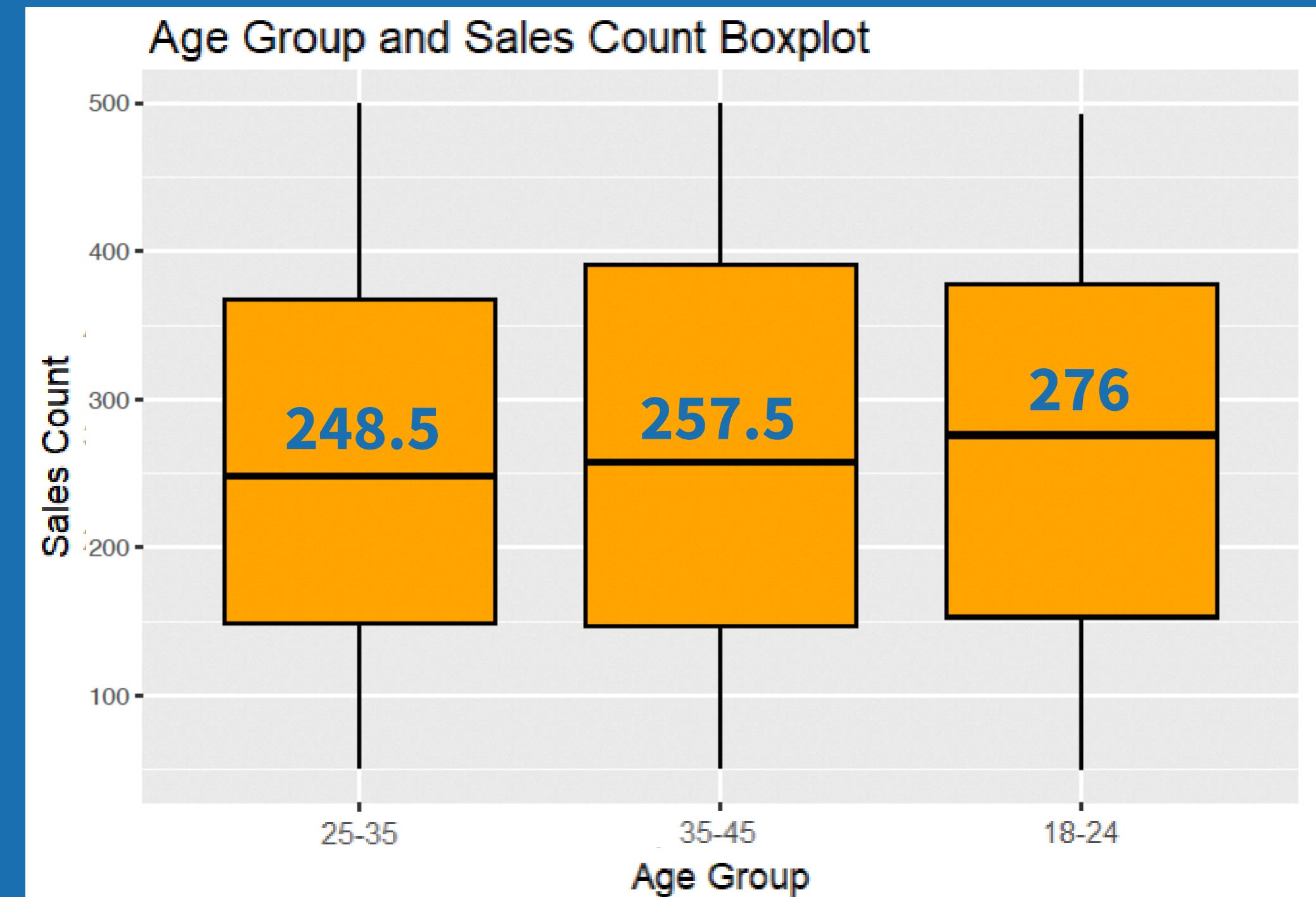
**25-35:** **Lowest** median, **Smallest** IQR

- More cautious spending, very consistent purchasing behavior within narrow range

**35-45:** Moderate median, **Widest** IQR

- Moderate purchasing behavior, high level of variability

**no outliers**



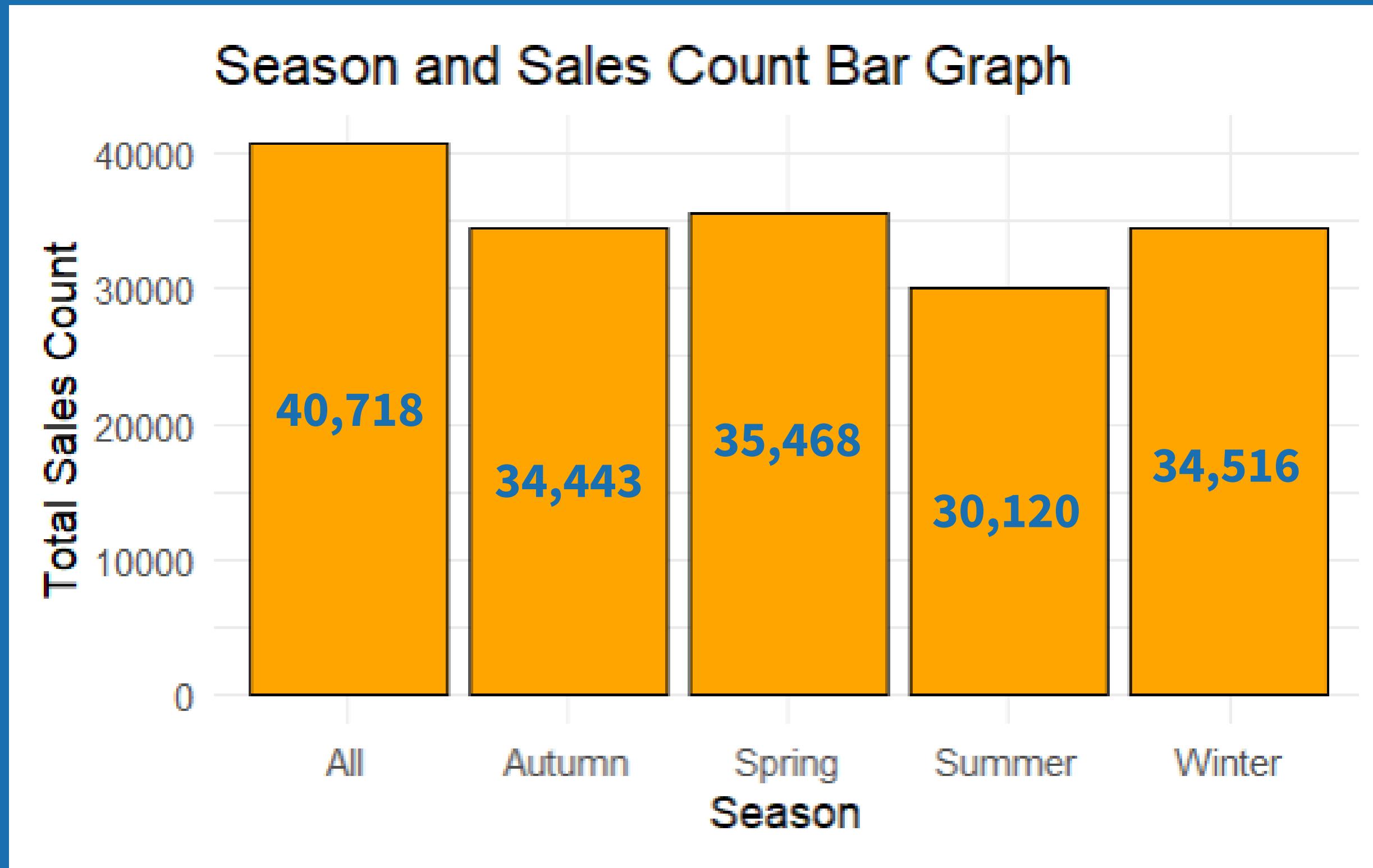
# SEASON



Basics or essentials for all seasons are more preferred

## BAR GRAPH

- Sales count between 30,000 & 41,000
- **Highest: All**
- **Lowest : Summer**



# SEASON



Fewer products achieving very **high sales** except for **All season**

## HISTOGRAM

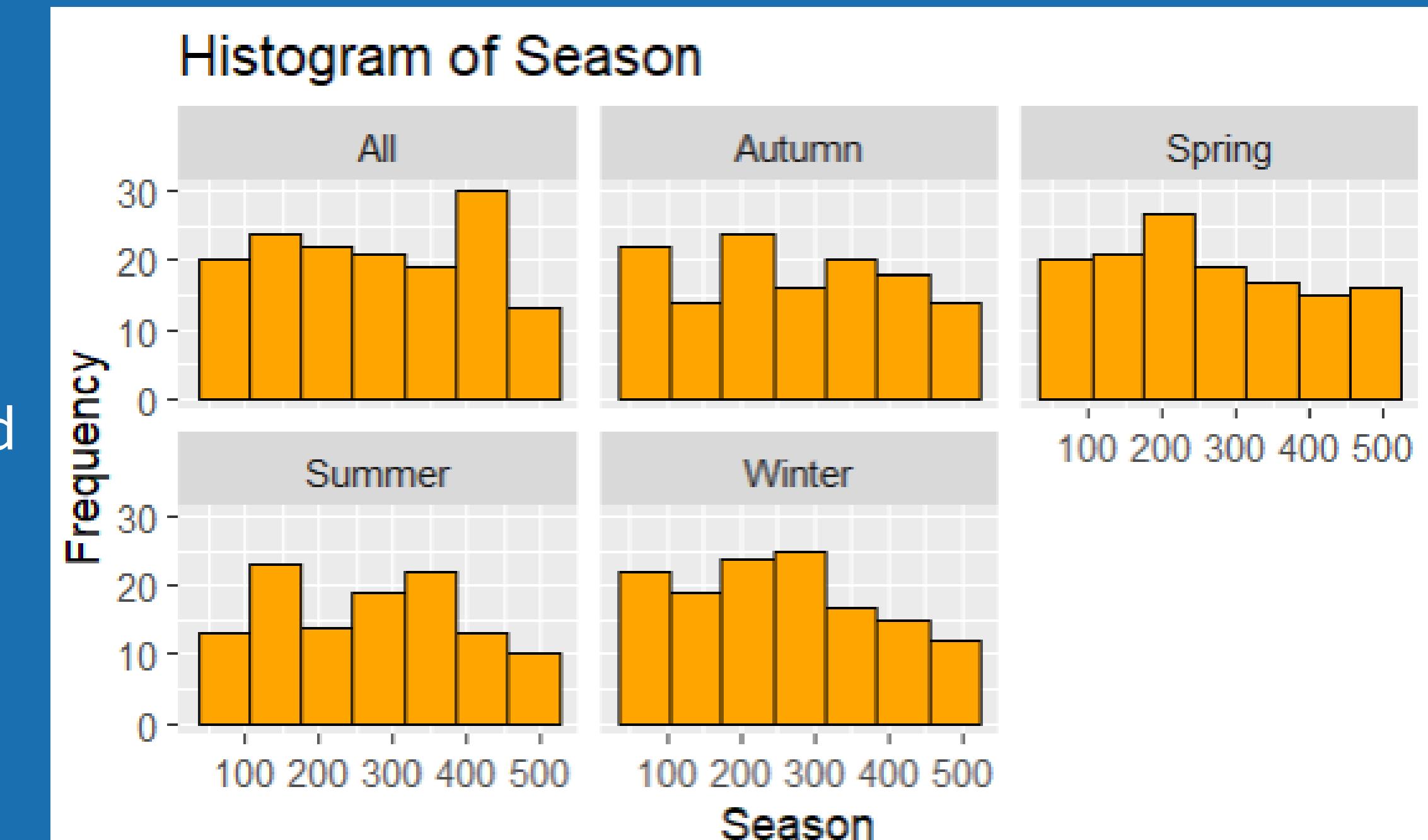
**All:** Bimodal

**Autumn:** Multi-modal

**Spring:** Unimodal, right-skewed

**Summer:** Symmetric Bimodal

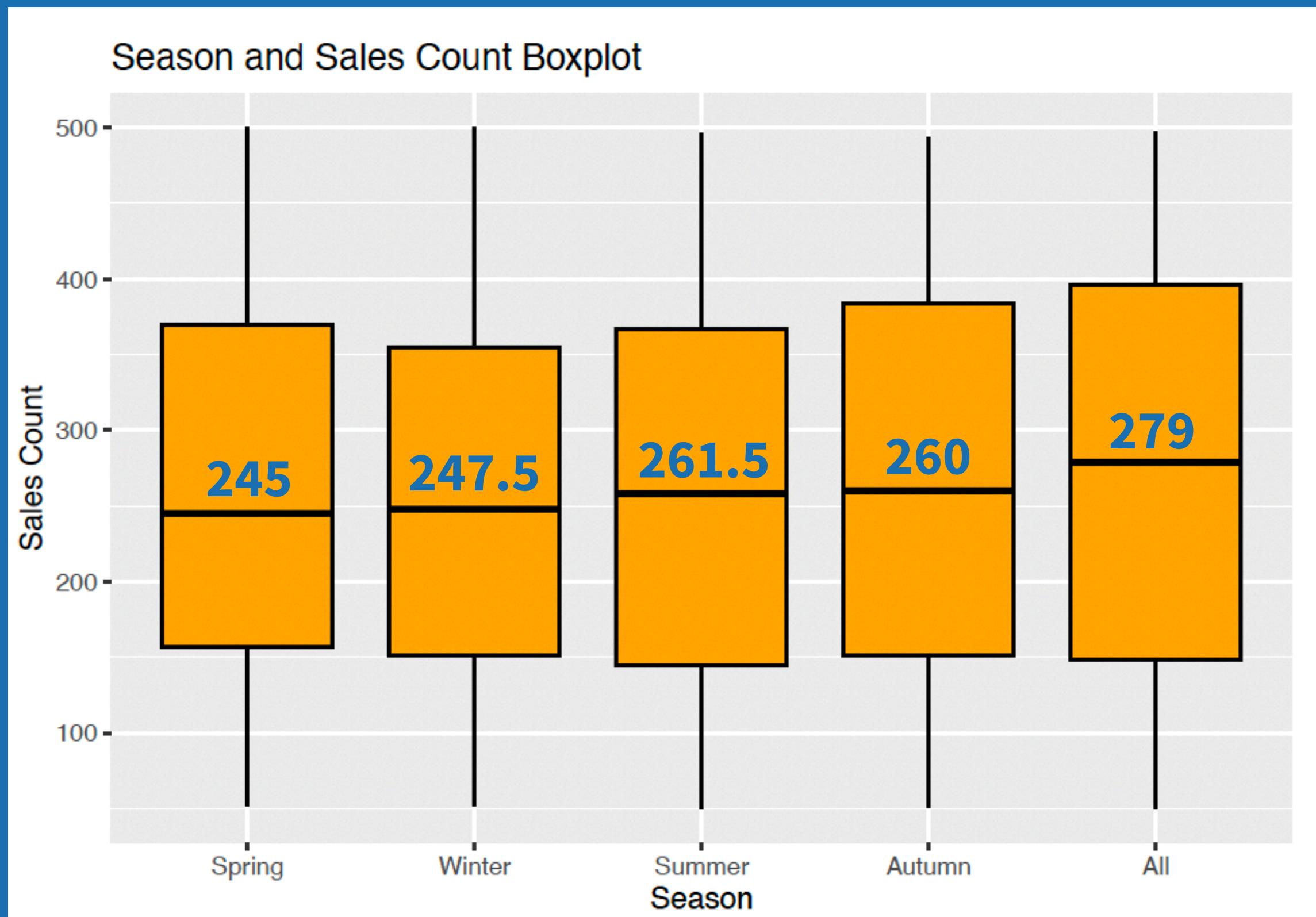
**Winter:** Bimodal, right-skewed



# SEASON

## BOXPLOT

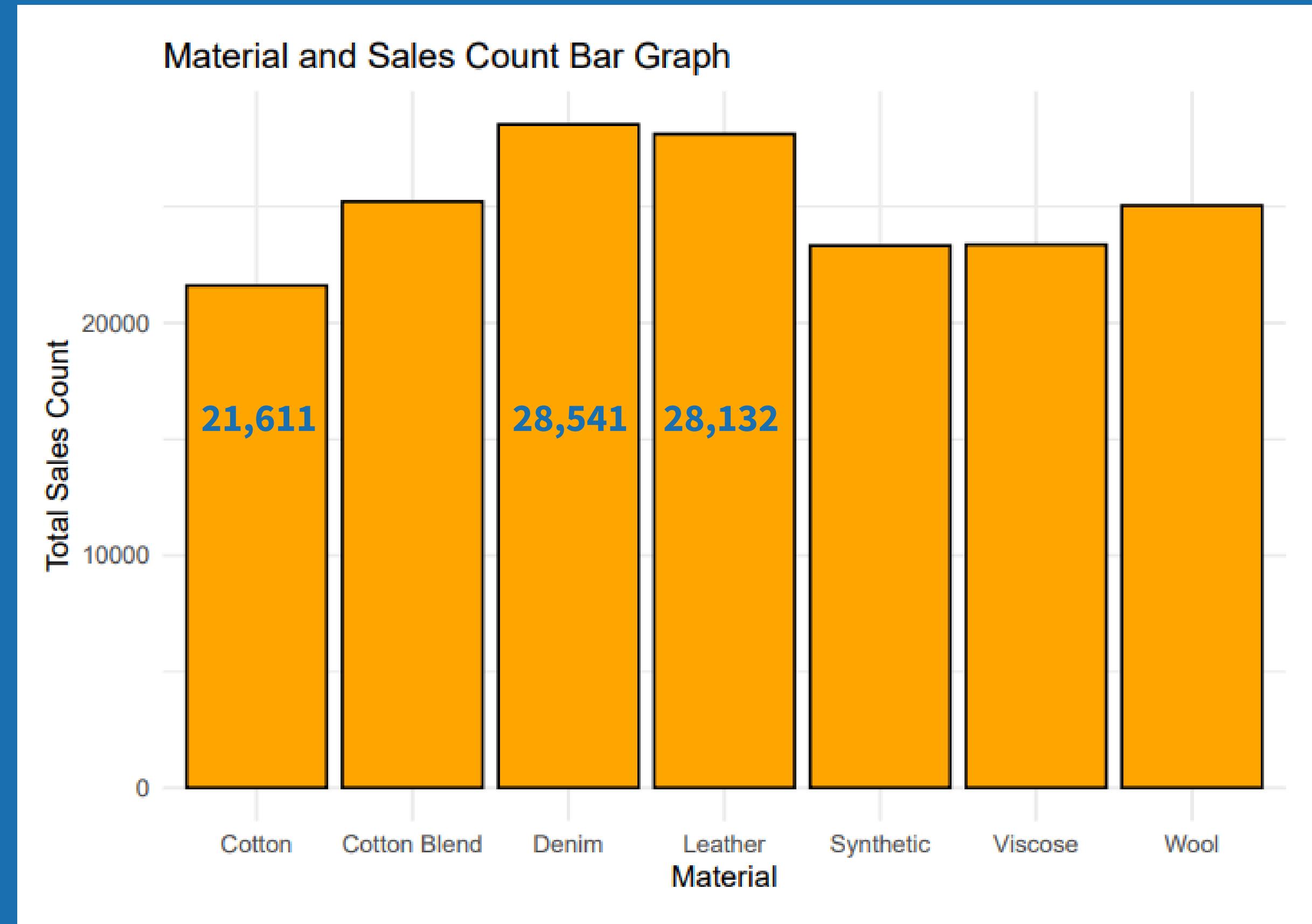
- No **outliers**
- Small difference in **median**
- **Highest** median: **All**
- **Lowest** median: **Spring**



# MATERIAL

## BAR GRAPH

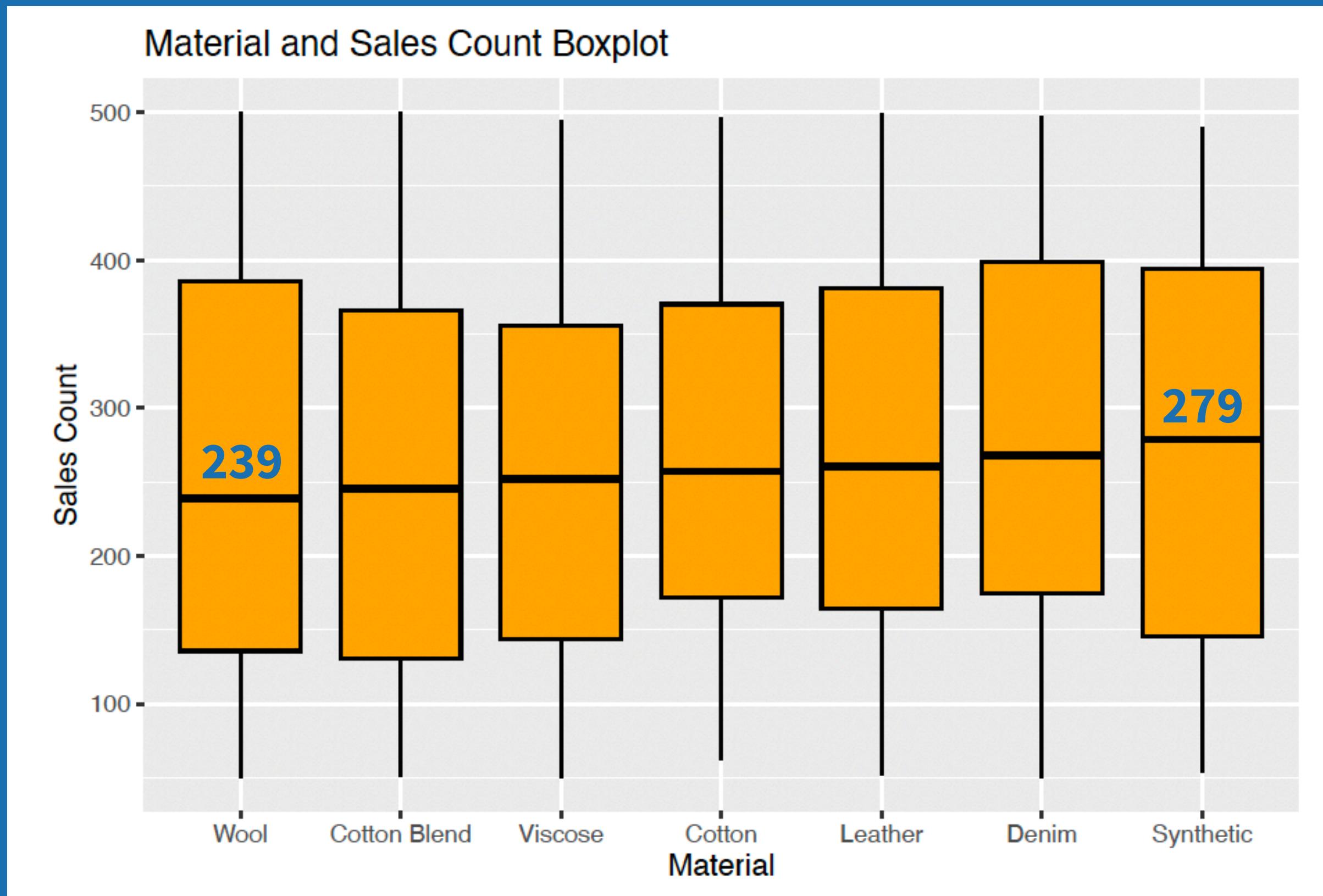
- All material's sales\_count is between 21,000 and 29,000.
- **Highest:** Denim
- **Lowest:** Cotton



# MATERIAL

## BOXPLOT

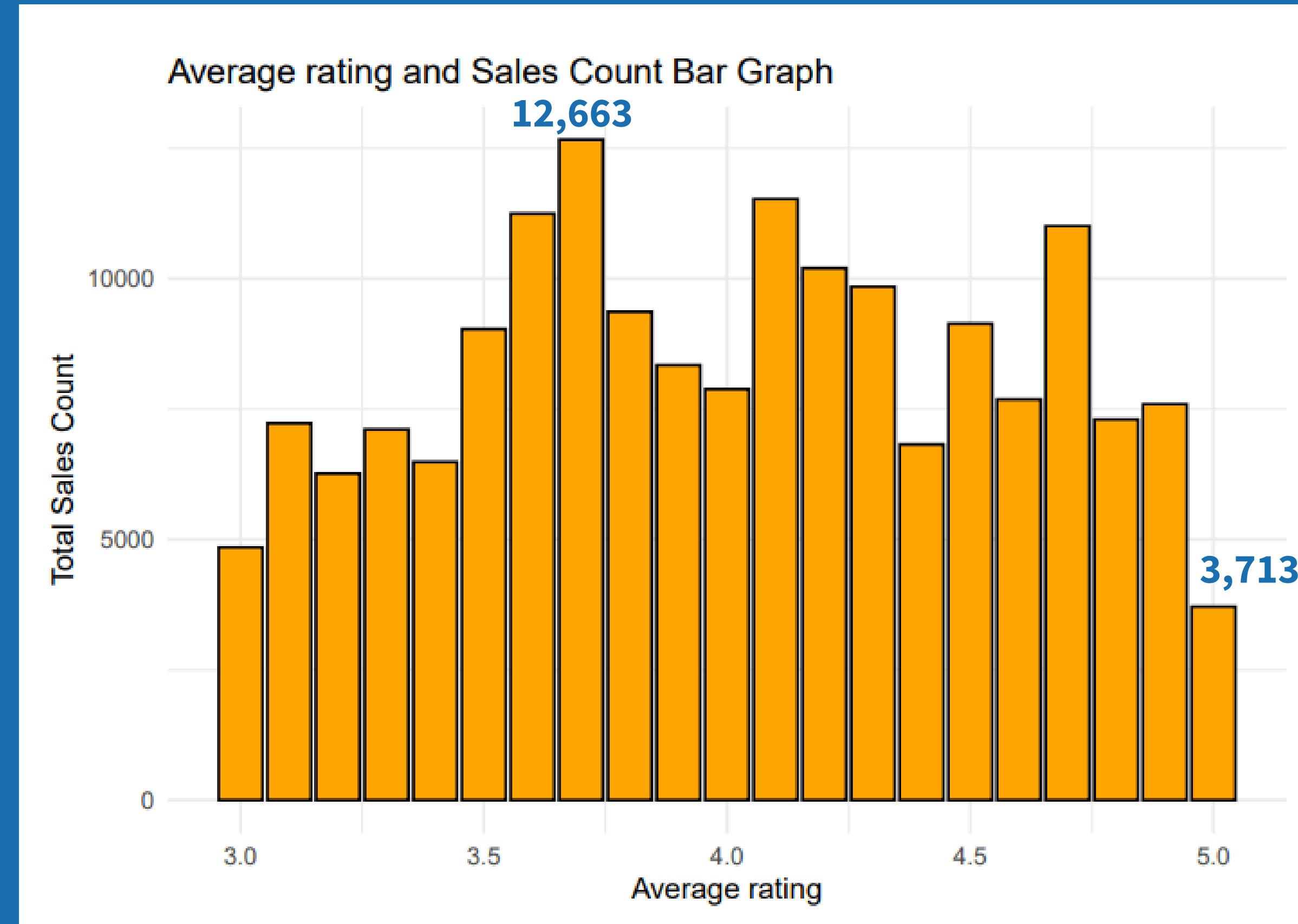
- no **outliers**
- All have a **median** value between 200~300
- **Highest median value:** Synthetic
- **Lowest median value:** Wool



# AVERAGE\_RATING

## BAR GRAPH

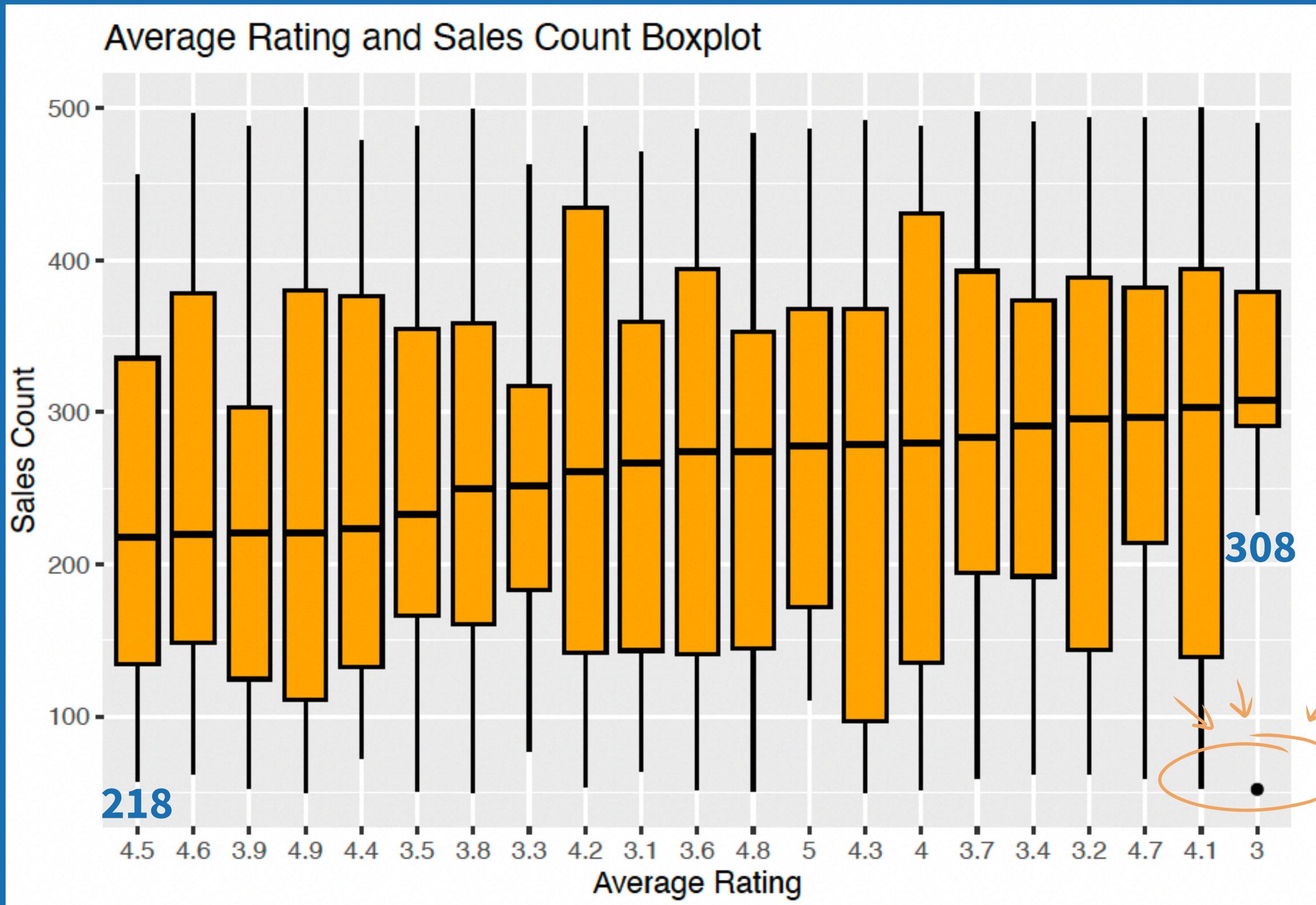
- Highest at 3.7
- Lowest at 5.0



# AVERAGE\_RATING

## BOXPLOT

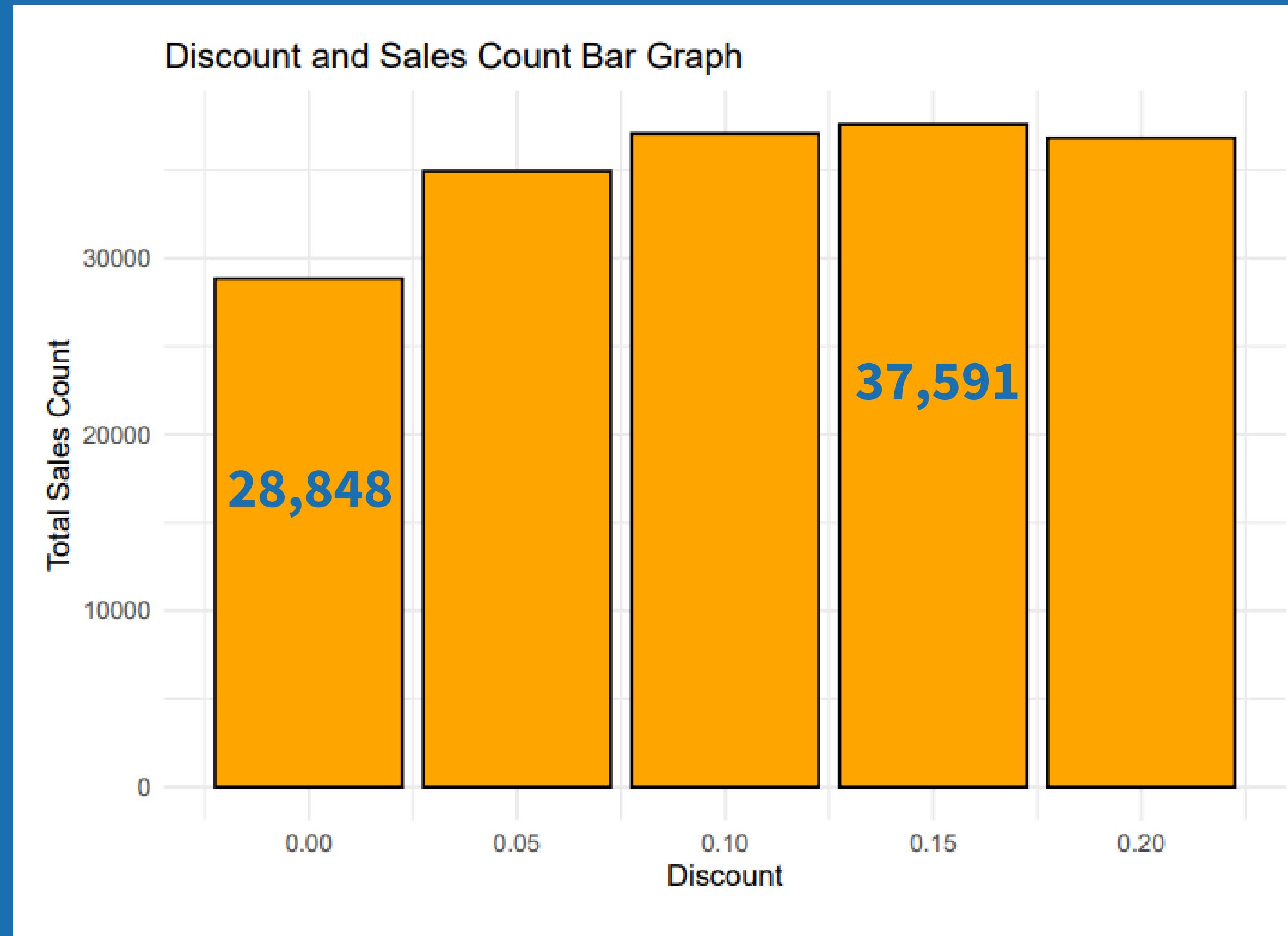
- Reveals a **significant outlier**
- All have a **median** value between 200~300
- **Highest** median: 3
- **Lowest** median: 4.5



# DISCOUNT

## BAR GRAPH

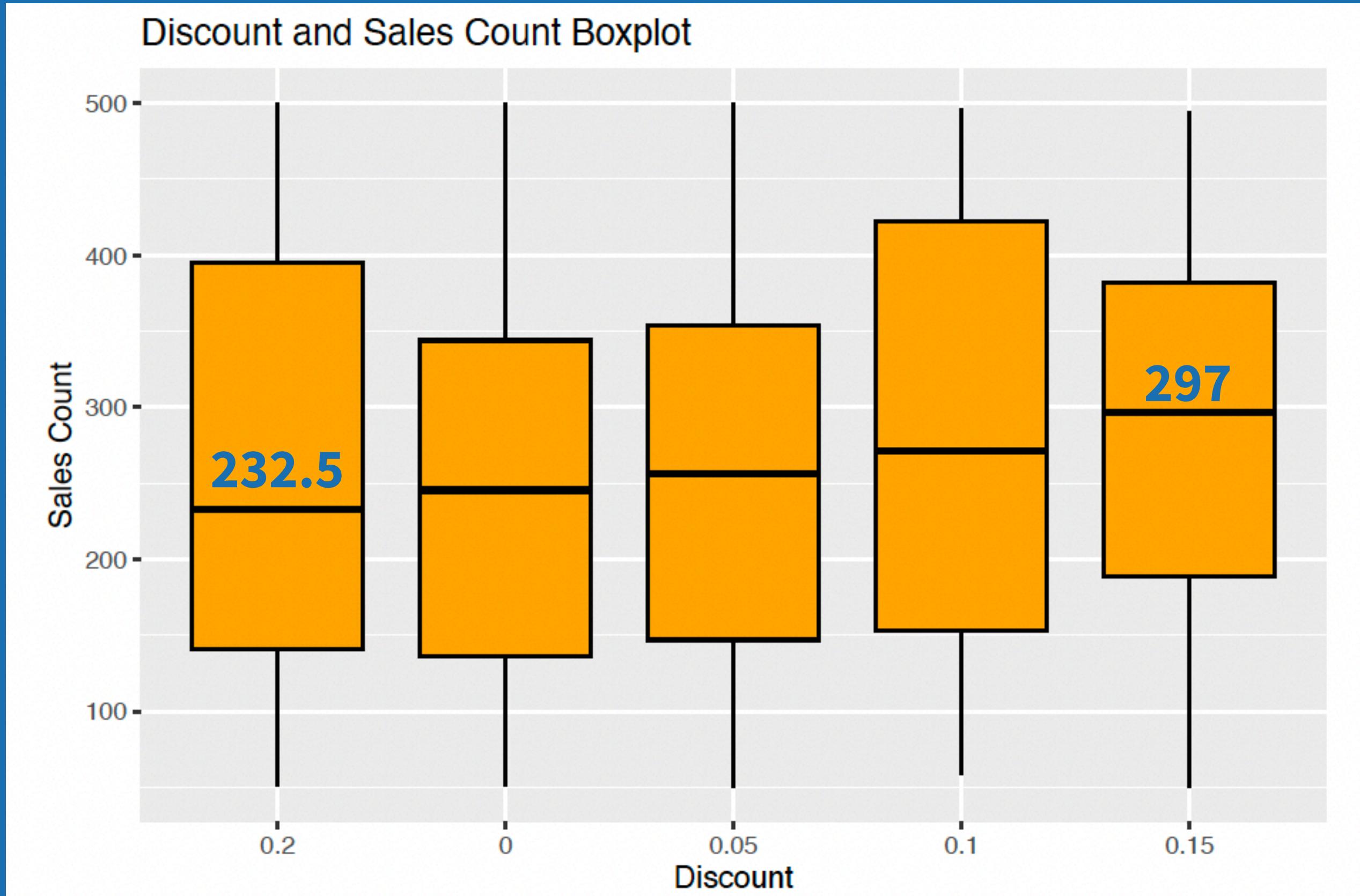
- From **5%** to **20%** discount, the sales count remains relatively **stable**
- **Highest discount:**
  - **15%** discount



# DISCOUNT

## BOXPLOT

- no **outliers**
- **Median: 100 to 500**
- **Highest Median: 0.15**
- **Lowest Median: 0.2**



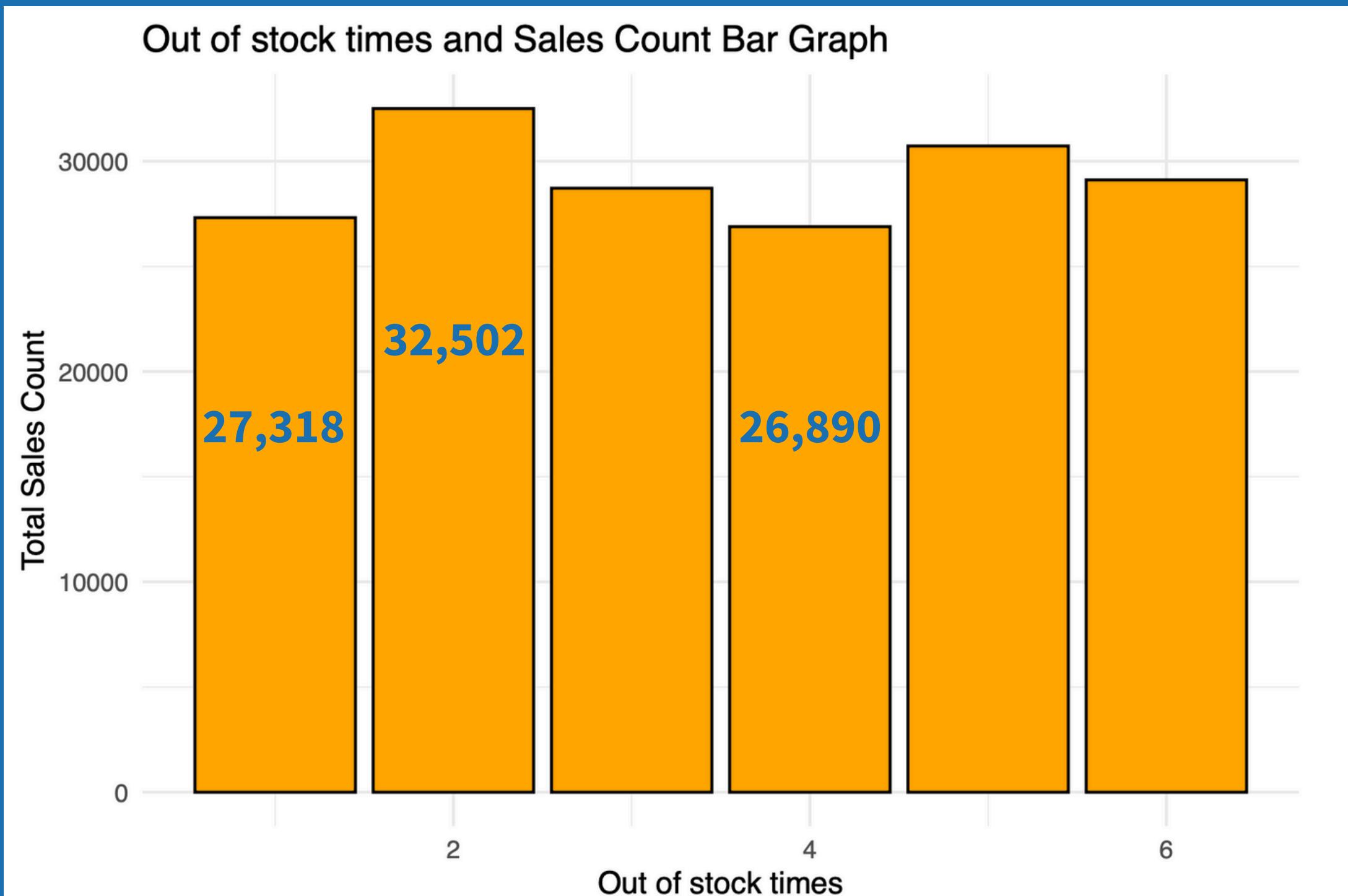
# OUT\_OF\_STOCK\_TIMES



How many times that the product was out of stock

## BAR GRAPH

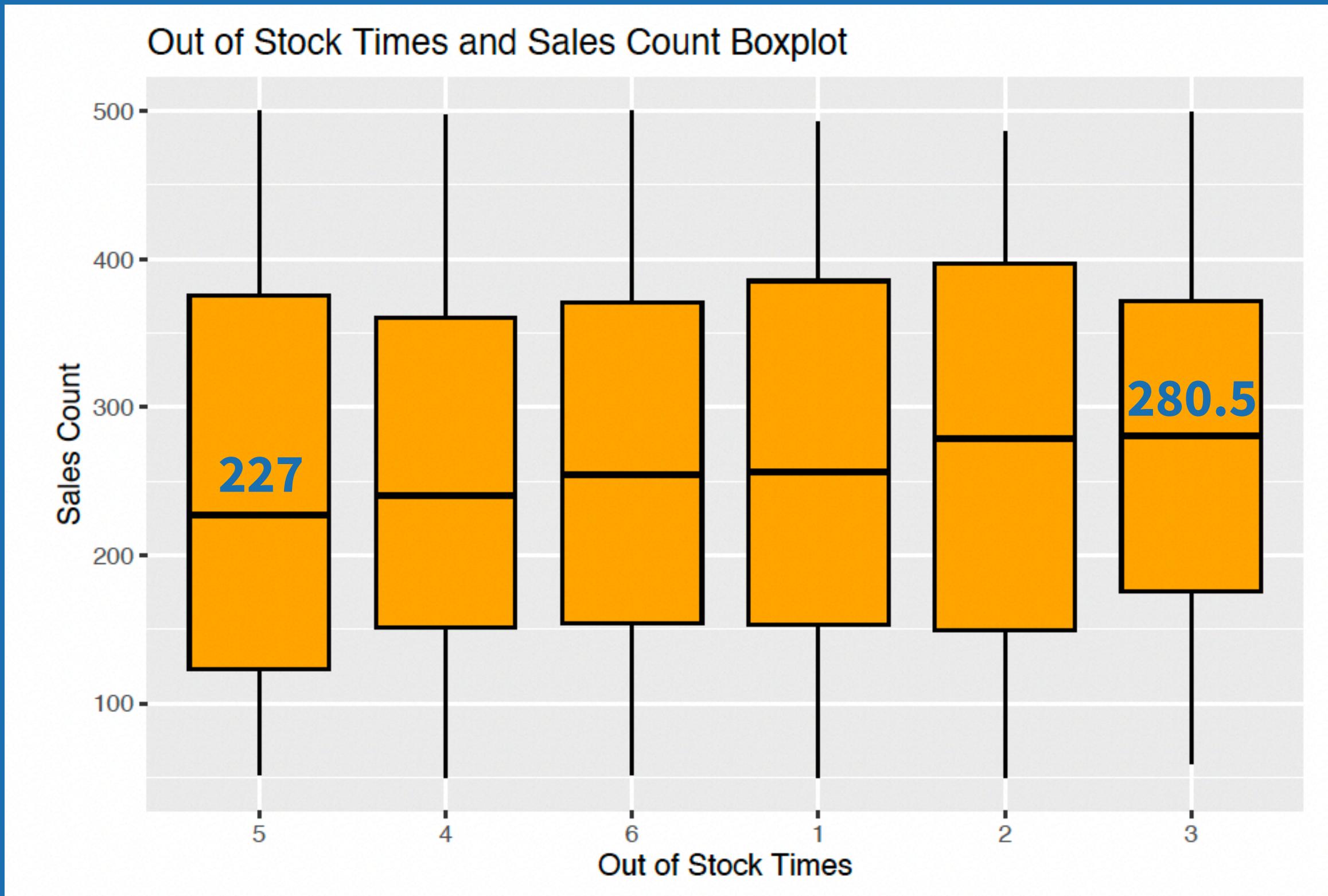
- **Highest** value: 2
- **Low** value: 4 < 1



# OUT\_OF\_STOCK\_TIMES

## BOXPLOT

- Median sales counts for each out\_of\_stock\_times are **similar**
- **Highest median:** 3
- **Lowest median:** 5



# ③ Visualization & Analysis

# HYPOTHESIS 1: GENDER VS. SALES\_COUNT

## 1. Hypothesis: Null vs. Alternative

**Null:** There is **no** relationship between gender and sales count.

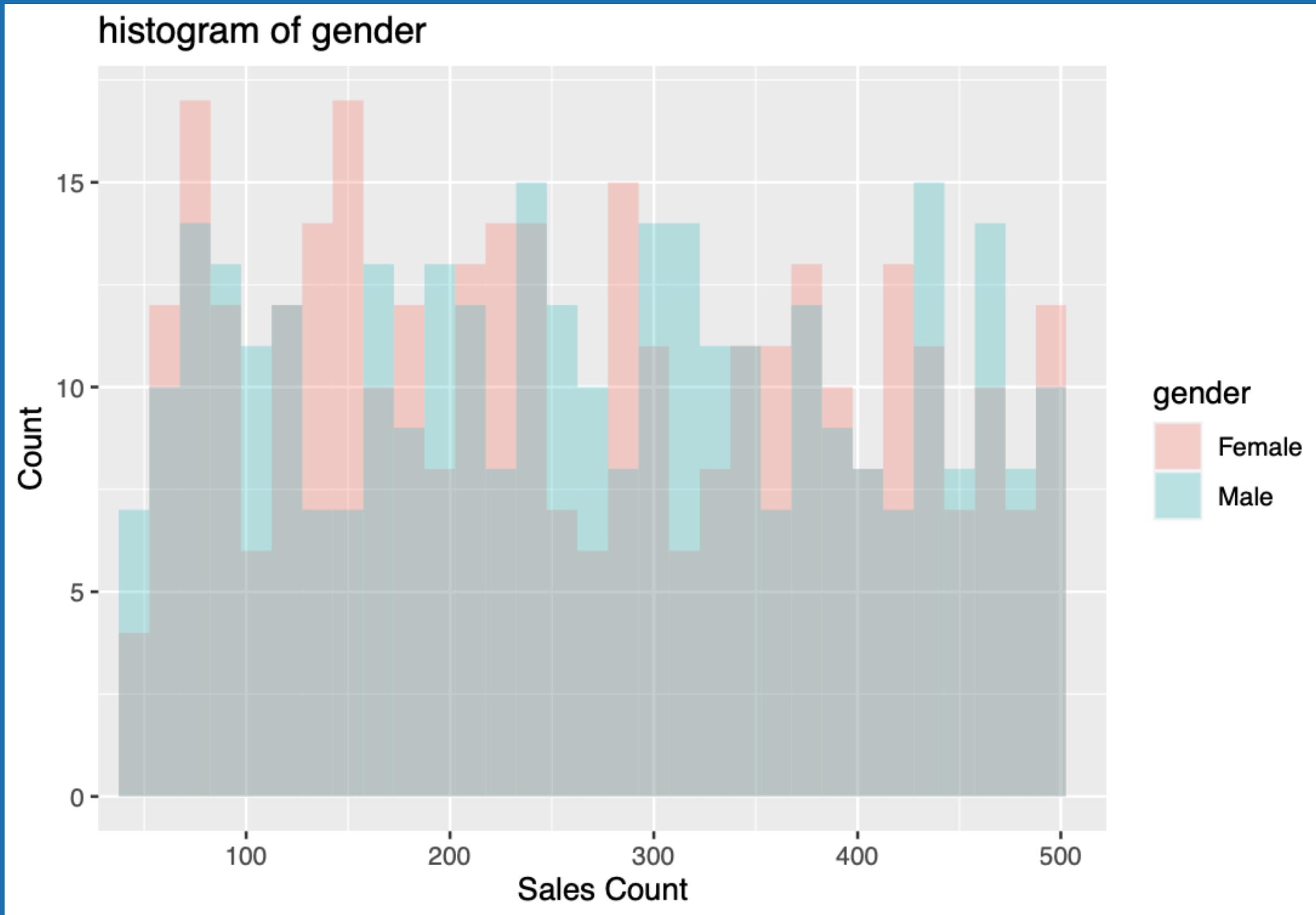
**Alternative:** There is relationship between gender and sales count.

## 2. Histogram of Gender

### Male & Female

### HISTOGRAM

- **Higest value:**  
Male(200~300, 400~500),  
Female(50~200)
- **Lowest value:**  
Male(0~100),  
Female(0~100)

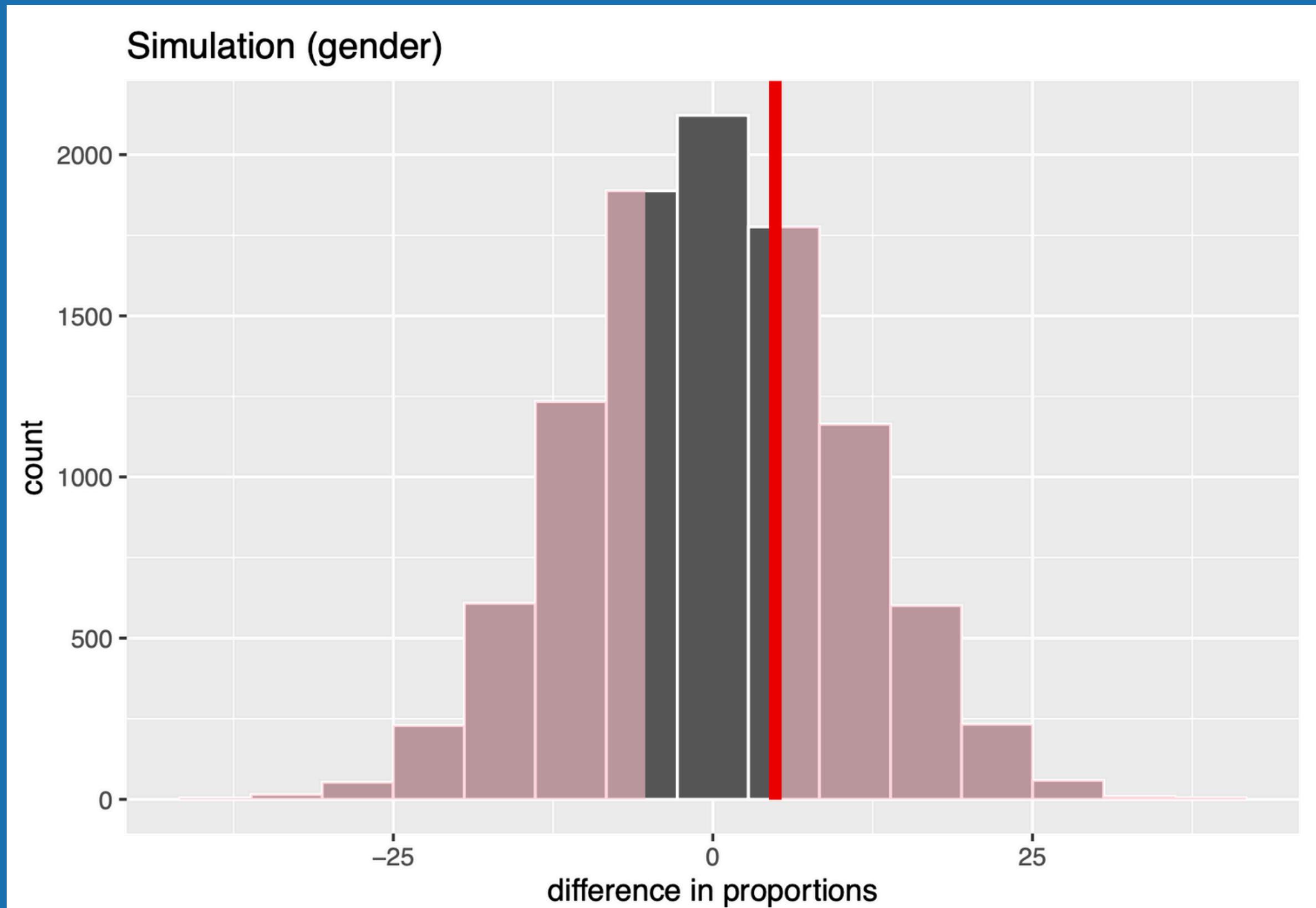


### 3. calculate “p-value ” & Simulation

Male & Female

p-value: 0.625

→ fail to reject H1



# HYPOTHESIS 2: AGE VS. SALES\_COUNT

**Null Hypothesis:** There is no relationship between age and sales count.

**Alternative:** There is relationship between age and sales count.

## 1. Filter data

**18-24 & 25-35**

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

```
f_ag_2 <- fashion %>%  
  filter(age_group == "18-24" | age_group == "35-45")
```

**18-24 & 35-45**

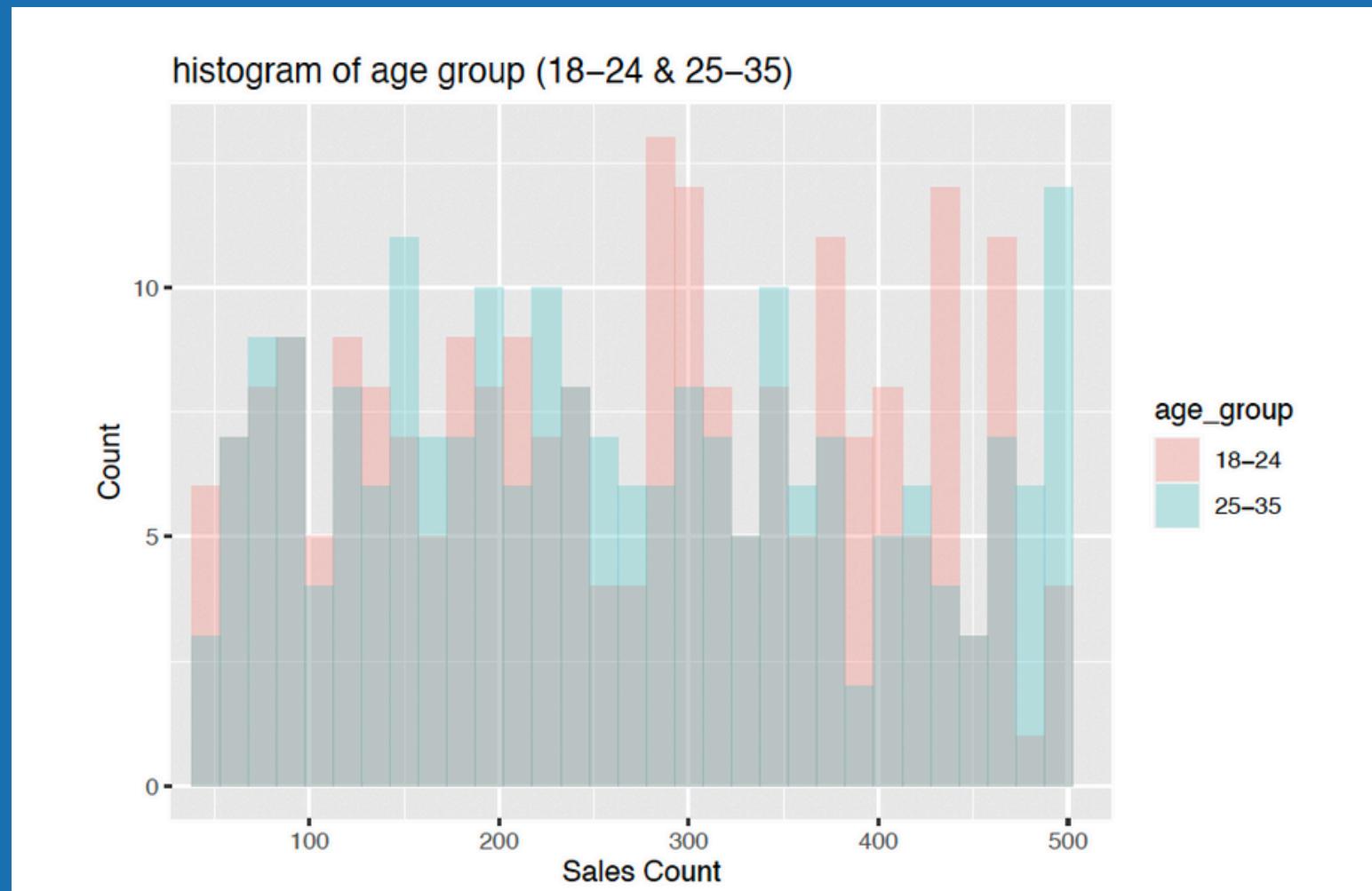
**25-35 & 35-45**

```
f_ag_3 <- fashion %>%  
  filter(age_group == "25-35" | age_group == "35-45")
```

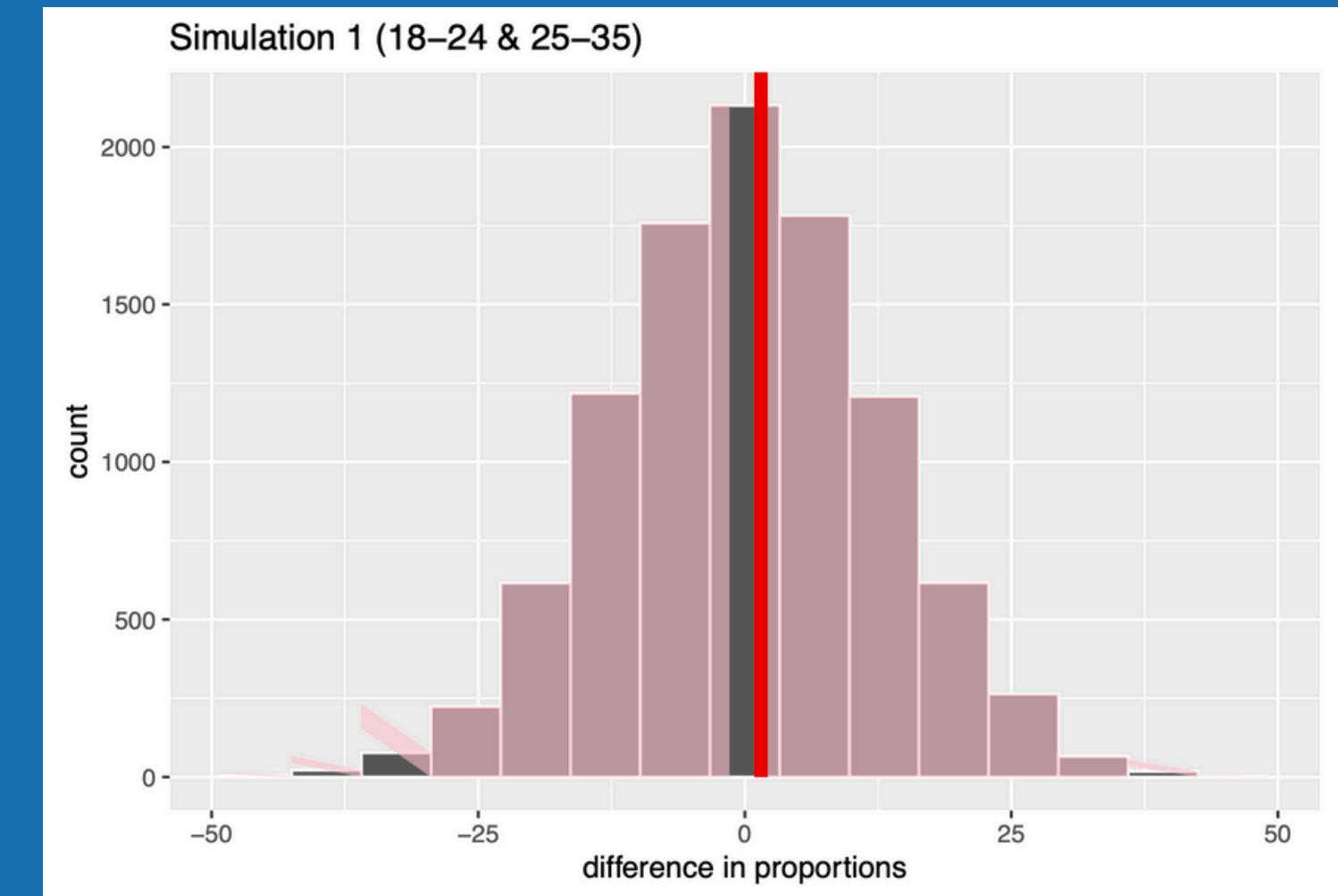
HYPOTHESIS 2: AGE VS. SALES\_COUNT

## 2. Histogram, “stat”, “p-value” of 18-24 & 25-35

### 18-24 & 25-35



### 18-24 & 25-35



**Highest Value:**

Male(480~500)/Female(280~300)

**Lowest Value:**

Male(380~400) / Female(430~450)

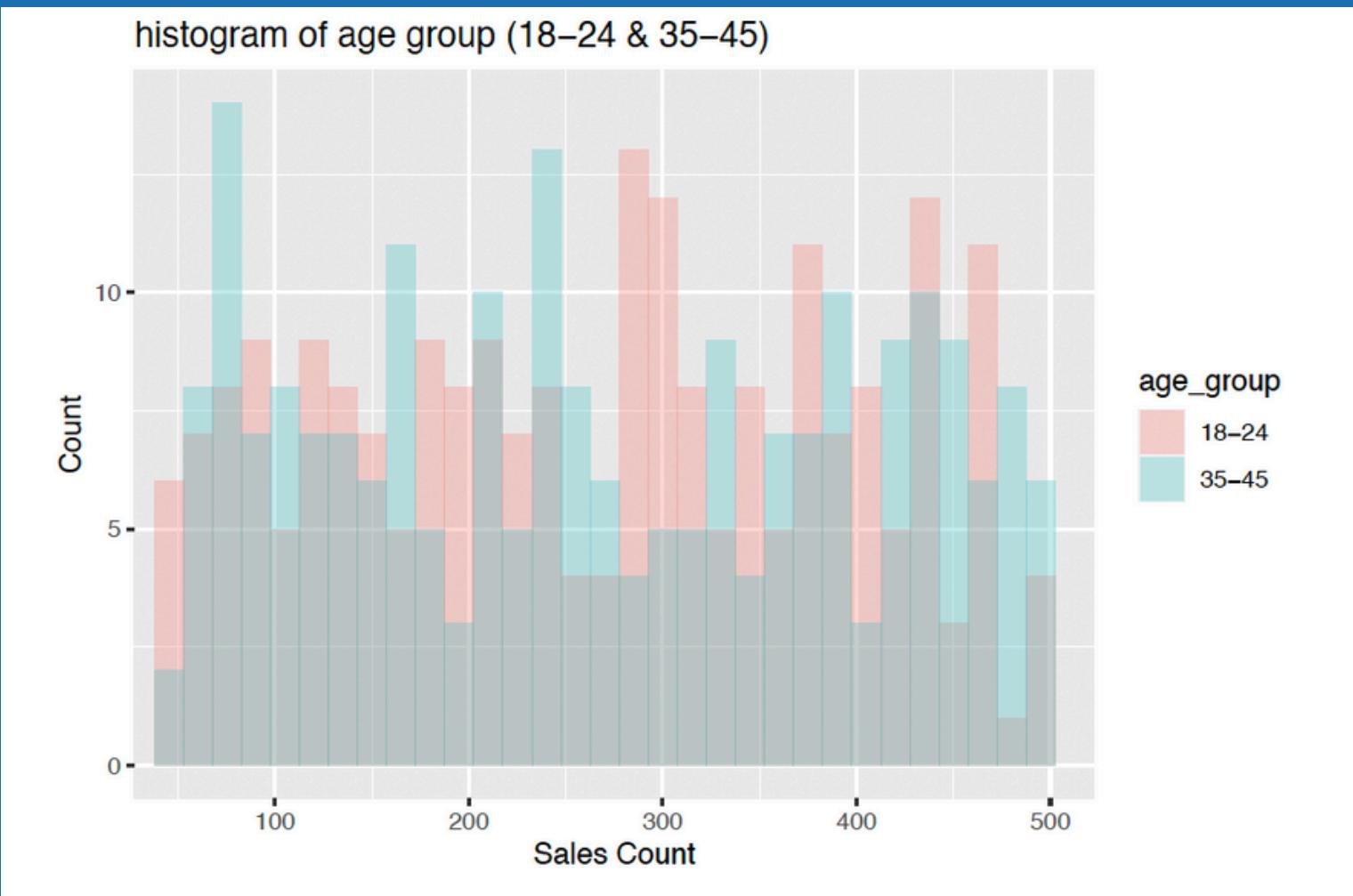
**p-value: 0.899**

→ fail to reject H<sub>2</sub>

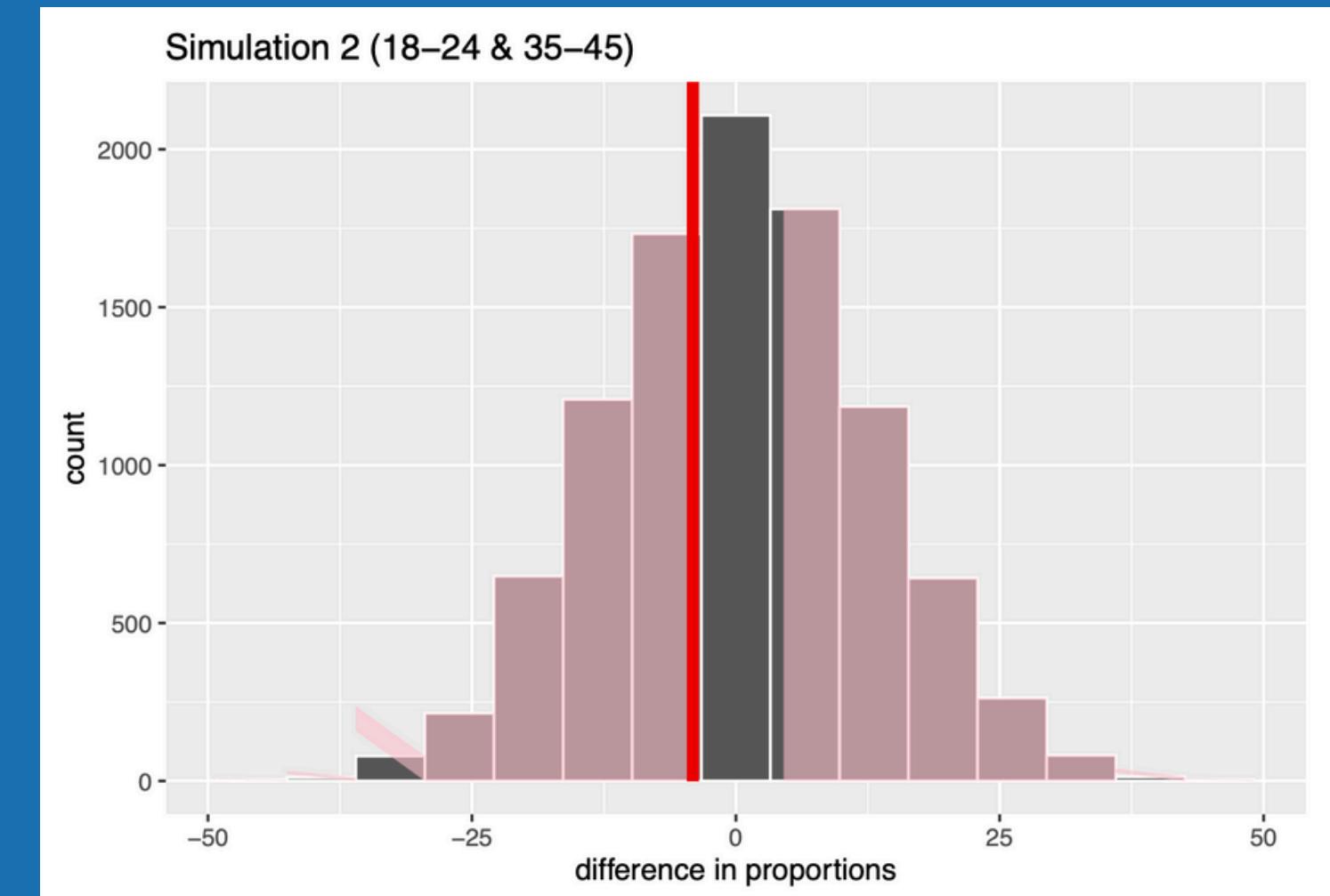
HYPOTHESIS 2: AGE VS. SALES\_COUNT

## 2. Histogram, “stat”, “p-value” of 18-24 & 35-45

### 18-24 & 35-45



### 18-24 & 35-45



**Highest Value:**

Male(30~70)/Female(280~300)

**Lowest Value:**

Male(400~430) / Female(450~480)

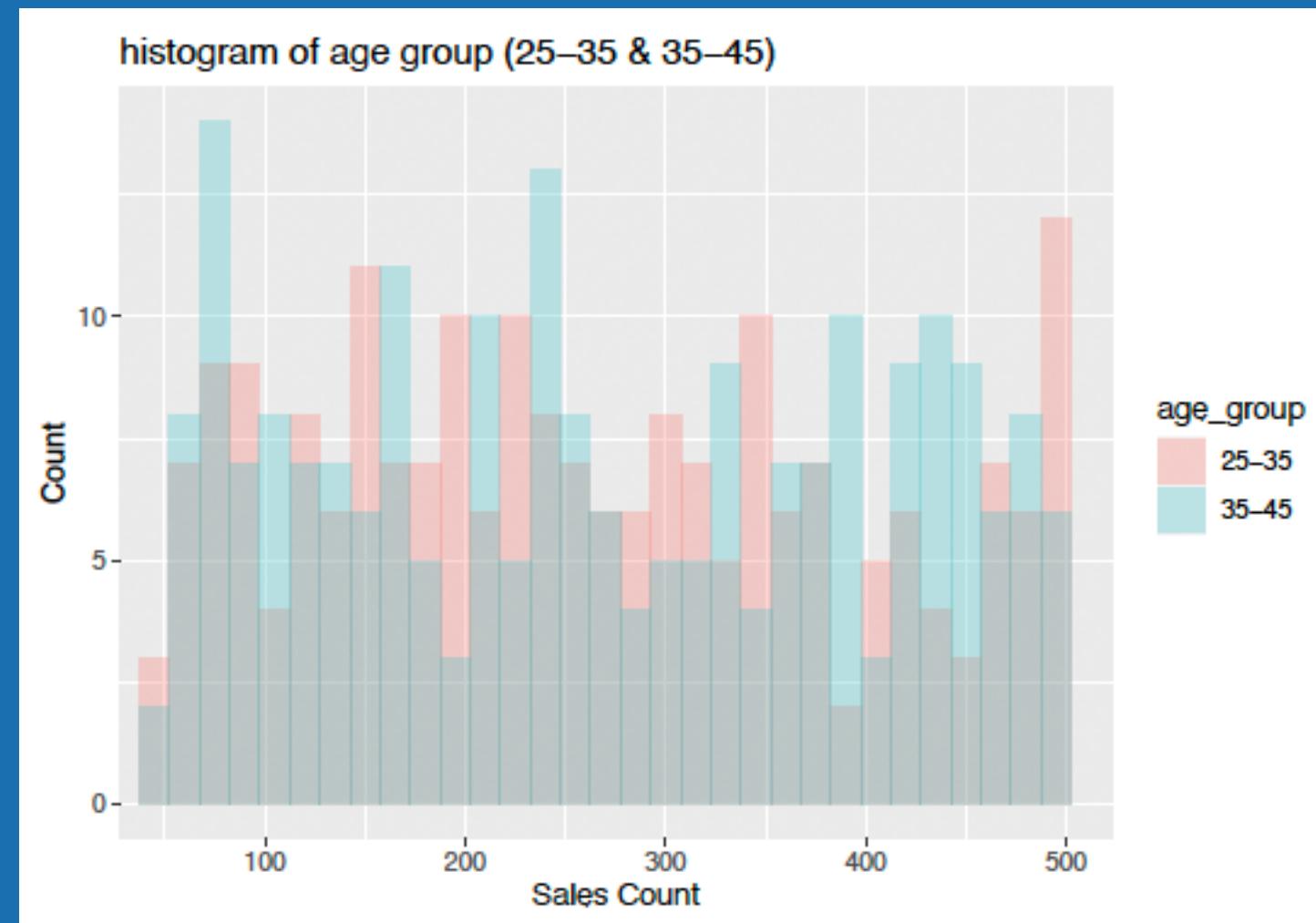
**p-value: 0.729**

→ fail to reject H<sub>2</sub>

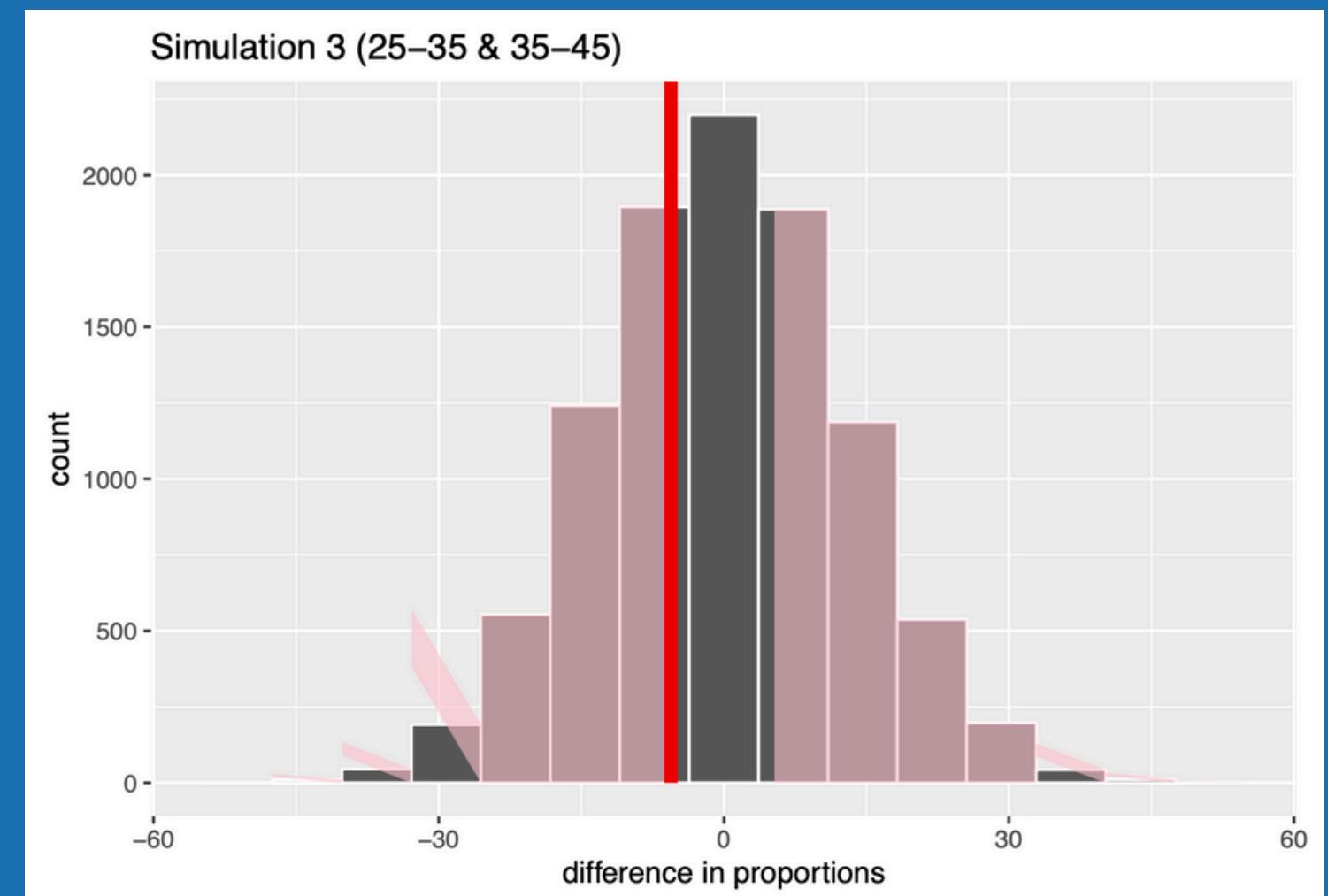
HYPOTHESIS 2: AGE VS. SALES\_COUNT

## 2. Histogram, “stat”, “p-value” of 25-35 & 35-45

### 25-35 & 35-45



### 25-35 & 35-45



**Highest Value:**

Male(30~70)/Female(480~500)

**Lowest Value:**

Male(0~30) / Female(380~400)

**p-value: 0.674**

**→ fail to reject H2**

# HYPOTHESIS 3: SEASON VS. SALES\_COUNT

## 1. Null & Alternative Hypothesis

**Null Hypothesis:** There is a relationship between season and sales\_count.

**Alternative:** There is no relationship between season and sales\_count.

HYPOTHESIS 3: SEASON VS. SALES\_COUNT

## 2. calculate “p-value ” & Simulation

| Season          | P - value | Result         |
|-----------------|-----------|----------------|
| All & Spring    | 0.508     | Fail to reject |
| All & Summer    | 0.587     | Fail to reject |
| All & Autumn    | 0.803     | Fail to reject |
| All & Winter    | 0.311     | Fail to reject |
| Spring & Summer | 0.932     | Fail to reject |

HYPOTHESIS 3: SEASON VS. SALES\_COUNT

## 2. calculate “p-value ” & Simulation

| Season          | P - value | Result         |
|-----------------|-----------|----------------|
| Spring & Autumn | 0.709     | Fail to reject |
| Spring & Winter | 0.757     | Fail to reject |
| Summer & Autumn | 0.775     | Fail to reject |
| Summer & Winter | 0.684     | Fail to reject |
| Autumn & Winter | 0.481     | Fail to reject |

4

# Conclusion

# 1.RESULT

## 1.GENDER

All P-value > 0.05

**Fail to reject**

**There is no relationship  
between  
Gender & Sales\_count**

## 2. AGE

All P-value > 0.05

**Fail to reject**

**There is no relationship  
between  
Age & Sales\_count**

## 3. SEASON

All P-value > 0.05

**Fail to reject**

**There is no relationship  
between  
Season & Sales\_count**

## 2. IMPLICATIONS

### 1. GENDER

- Purchasing behavior is similar across all genders
- Marketing & product offerings might not need to be heavily changed

### 2. AGE

- Products appeal equally across various age demographics
- Diverse product range without the need to target specific age groups

### 3. SEASON

- Relatively stable sales performance throughout the year
- ZARA is doing great for its sales strategies of balancing and producing clothes seasonally

## 2. IMPLICATIONS



- ZARA's successful utilization of meeting the desires of customers regardless of their age, gender, and season

### 1. Stability & Consistency:

- a. ZARA's stable sales across gender, age, season

### 2. Marketing Strategy:

- a. Adopting a broad marketing strategy rather than heavily rely on specific segments

### 3. Product Offering:

- a. Well-rounded product lineup that appeals to a wide audience

**THANKS FOR WATCHING**