EDA_MIDTERMS_KHAFAJI

Exploratory Data Analysis of customer purchasing behavior in an e-commerce platform.

The data set that we will use to analyze customer purchasing behaviour in an e-commerce platform has the following variables, or columns:

- Customer_ID: Unique identifier for each customer
- Gender: Male or Female
- Age: Customer's age in years
- Browsing_Time: Average time spent on the website per visit (in minutes)
- Purchase_Amount: Total amount spent in a single transaction (in USD)
- Number_of_Items: Number of items purchased per transaction
- Discount_Applied: Discount percentage applied to the transaction
- Total_Transactions: Total number of transactions by the customer
- Category: Product category (e.g., Electronics, Clothing, Home & Kitchen, etc.)
- Satisfaction_Score: Customer satisfaction score (1-5 scale)

Let's first load up the data as a tibble, saving it as ecom_pbh.

```
ecom_pbh <- read_csv('EDA_Ecommerce_Assessment.csv', show_col_types = FALSE)

print("The structure of the data set looks like this:")
```

```
## [1] "The structure of the data set looks like this:"
```

```
print(spec(ecom_pbh))
```

```
## cols(
##   Customer_ID = col_double(),
##   Gender = col_character(),
##   Age = col_double(),
##   Browsing_Time = col_double(),
##   Purchase_Amount = col_double(),
##   Number_of_Items = col_double(),
##   Discount_Applied = col_double(),
##   Total_Transactions = col_double(),
##   Category = col_character(),
##   Satisfaction_Score = col_double()
## )
```

```
cat("\n\nMissing Values in Each Column:\n")
```

```
##
##
```

```
## Missing Values in Each Column:
```

```r
print(colSums(is.na(ecom_pbh)))
```

```
##        Customer_ID             Gender                Age      Browsing_Time
##                  0                  0                  0                  0
##    Purchase_Amount    Number_of_Items   Discount_Applied Total_Transactions
##                  0                  0                  0                  0
##           Category Satisfaction_Score
##                  0                  0
```

```r
cat("\n\nSummary of each column: \n")
```

```
##
##
## Summary of each column:
```

```r
print(summary(ecom_pbh))
```

```
##   Customer_ID        Gender               Age        Browsing_Time
##  Min.   :   1.0   Length:3000        Min.   :18.00   Min.   :  1.00
##  1st Qu.: 750.8   Class :character   1st Qu.:31.00   1st Qu.: 29.98
##  Median :1500.5   Mode  :character   Median :44.00   Median : 59.16
##  Mean   :1500.5                      Mean   :43.61   Mean   : 59.87
##  3rd Qu.:2250.2                      3rd Qu.:57.00   3rd Qu.: 89.33
##  Max.   :3000.0                      Max.   :69.00   Max.   :119.95
##  Purchase_Amount  Number_of_Items Discount_Applied Total_Transactions
##  Min.   :  5.03   Min.   :1.00    Min.   : 0.00    Min.   : 1.00
##  1st Qu.:128.69   1st Qu.:3.00    1st Qu.:12.00    1st Qu.:12.00
##  Median :245.09   Median :5.00    Median :24.00    Median :24.00
##  Mean   :247.96   Mean   :4.99    Mean   :24.34    Mean   :24.68
##  3rd Qu.:367.20   3rd Qu.:7.00    3rd Qu.:37.00    3rd Qu.:37.00
##  Max.   :499.61   Max.   :9.00    Max.   :49.00    Max.   :49.00
##    Category         Satisfaction_Score
##  Length:3000        Min.   :1.000
##  Class :character   1st Qu.:2.000
##  Mode  :character   Median :3.000
##                     Mean   :3.066
##                     3rd Qu.:4.000
##                     Max.   :5.000
```

```r
cat("\n\nwhile the first few columns of the data set looks like this:")
```

```
##
##
## while the first few columns of the data set looks like this:
```

```r
print(head(ecom_pbh))
```

```
## # A tibble: 6 x 10
##   Customer_ID Gender   Age Browsing_Time Purchase_Amount Number_of_Items
##         <dbl> <chr>  <dbl>         <dbl>           <dbl>           <dbl>
## 1           1 Male      65          46.6           232.                6
## 2           2 Female    19          98.8           473.                8
## 3           3 Male      23          79.5           338.                1
## 4           4 Male      45          95.8            37.1               7
## 5           5 Male      46          33.4           236.                3
## 6           6 Female    43          83.4           124.                9
```

```
## # i 4 more variables: Discount_Applied <dbl>, Total_Transactions <dbl>,
## #   Category <chr>, Satisfaction_Score <dbl>
```

We can see that the data set has 3,000 data points, with 2 categorical variables and 8 numerical variables. There are no missing values in the data set, and the set up of the data set is tidy.

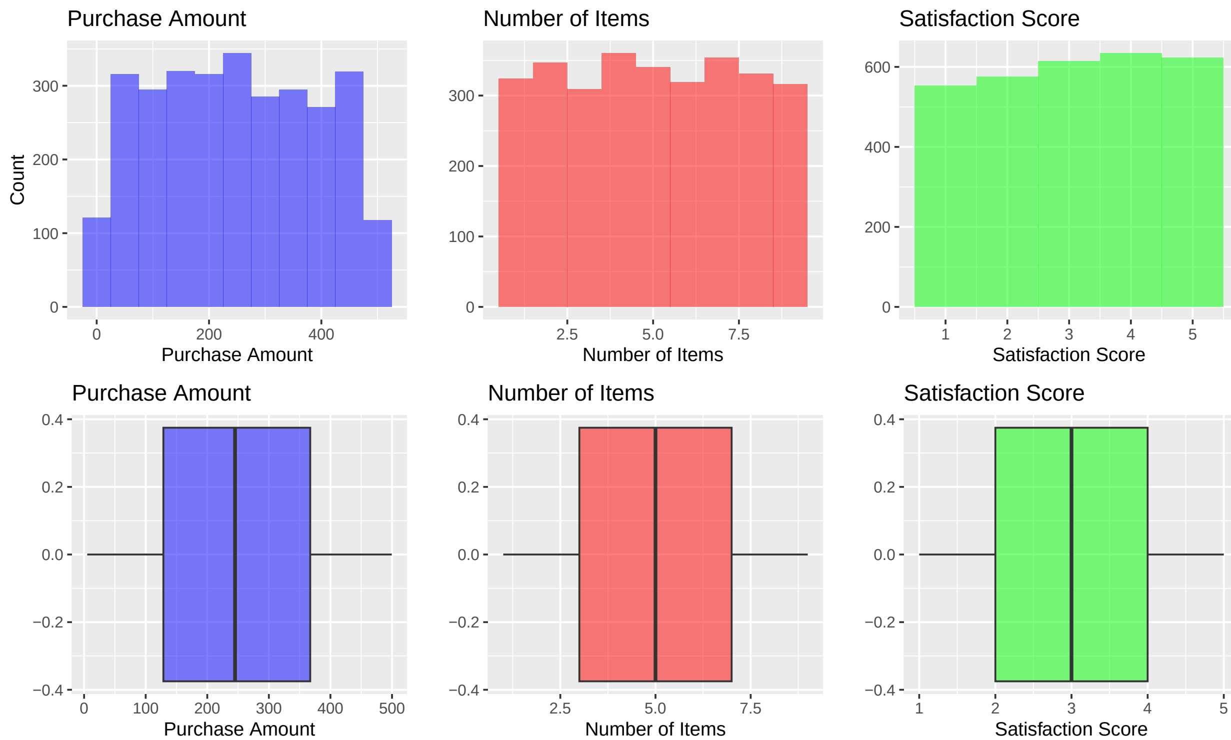
Univariate Data Analysis

Create histograms and boxplots to visualize the distribution of Purchase_Amount, Number_of_Items, and Satisfaction_Score.

Now, let's first analyze three of our main metrics: the purchase amount for each transaction, the number of items for each transaction, and the satisfaction score of the buyers.

Let's start first with the purchase amount

```r
pur_amount_dist <- ecom_pbh %>% ggplot(aes(x=Purchase_Amount)) + geom_histogram(alpha=0.5, fill="blue",
  ylab("Count") + xlab("Purchase Amount") + ggtitle("Purchase Amount")

pur_amount_box <- ecom_pbh %>% ggplot(aes(x=Purchase_Amount)) + geom_boxplot(alpha=0.5, fill="blue") +
  ylab("") + xlab("Purchase Amount") + ggtitle("Purchase Amount")


num_items_dist <- ecom_pbh %>% ggplot(aes(x=Number_of_Items)) + geom_histogram(alpha=0.5, fill="red", bi
  ylab("") + xlab("Number of Items") + ggtitle("Number of Items")

num_items_box <- ecom_pbh %>% ggplot(aes(x=Number_of_Items)) + geom_boxplot(alpha=0.5, fill="red") +
  ylab("") + xlab("Number of Items") + ggtitle("Number of Items")


satisfaction_score_dist <- ecom_pbh %>% ggplot(aes(x=Satisfaction_Score)) + geom_histogram(alpha=0.5, fi

satisfaction_score_box <- ecom_pbh %>% ggplot(aes(x=Satisfaction_Score)) + geom_boxplot(alpha=0.5, fill=

plot_grid(pur_amount_dist, num_items_dist, satisfaction_score_dist, pur_amount_box, num_items_box, satis
```



Most orders have a purchase amount between 100 USD and 300 USD, with the median being at around 250 USD. The number of items usually bought in one order is between 3 and 7 items, with the median being 7 items bought in one order. For the satisfaction score, most users of the e-commerce app are neutral about their satisfactions, with a median score of 3, although the mode is 4, as can be seen in the histogram.

Compute measures of central tendency (mean, median, mode) and spread (variance, standard deviation, IQR) for Purchase_Amount.

```
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

ecom_pbh_summ <- ecom_pbh %>% summarise(
  Max = max(Purchase_Amount),
  Min = min(Purchase_Amount),
  Mean = mean(Purchase_Amount),
  Median = median(Purchase_Amount),
  Mode = Mode(Purchase_Amount),
  Variance = var(Purchase_Amount),
  Standard_deviation = sd(Purchase_Amount),
  Interquartile_range = IQR(Purchase_Amount)
  )

ecom_pbh_summ %>% pivot_longer(colnames(ecom_pbh_summ), names_to = "Purchase Amount Statistics", values_
```

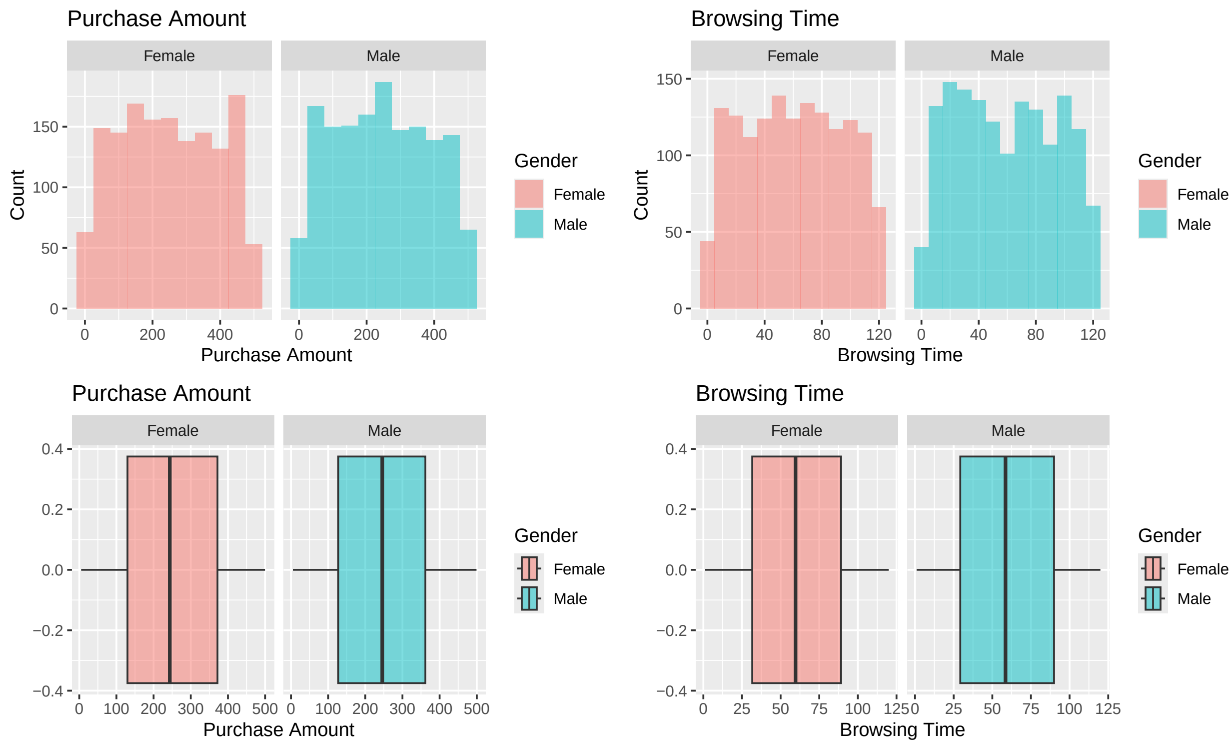
```
## # A tibble: 8 x 2
##   `Purchase Amount Statistics`    Value
##   <chr>                           <dbl>
## 1 Max                            500.
```

```
## 2 Min                           5.03
## 3 Mean                        248.
## 4 Median                      245.
## 5 Mode                         29.3
## 6 Variance                  19846.
## 7 Standard_deviation          141.
## 8 Interquartile_range         239.
```

The customers of the e-commerce website has a mean purchase amount of 247.96 USD, a median of 245.09 USD, and a mode of 29.33 USD. The data set has a variance of 19,845.99, while having a standard deviation of 140.876. It's interquartile range agrees with the central tendency measures, with an interquartile range of 238.50.

Compare the distribution of Browsing_Time and Purchase_Amount across different Gender groups using density plots.

```r
pur_amount_dist_gender <- ecom_pbh %>% ggplot(aes(x=Purchase_Amount, fill=Gender)) + geom_histogram(alpl
  ylab("Count") + xlab("Purchase Amount") + ggtitle("Purchase Amount") + facet_wrap(~Gender)

pur_amount_box_gender <- ecom_pbh %>% ggplot(aes(x=Purchase_Amount, fill=Gender)) + geom_boxplot(alpha=(
  ylab("") + xlab("Purchase Amount") + ggtitle("Purchase Amount") + facet_wrap(~Gender)


browse_dist_gender <- ecom_pbh %>% ggplot(aes(x=Browsing_Time, fill=Gender)) + geom_histogram(alpha=0.5
  ylab("Count") + xlab("Browsing Time") + ggtitle("Browsing Time") + facet_wrap(~Gender)

browse_box_gender <- ecom_pbh %>% ggplot(aes(x=Browsing_Time, fill=Gender)) + geom_boxplot(alpha=0.5 ) +
  ylab("") + xlab("Browsing Time") + ggtitle("Browsing Time") + facet_wrap(~Gender)


plot_grid(pur_amount_dist_gender, browse_dist_gender, pur_amount_box_gender, browse_box_gender, label_s
```

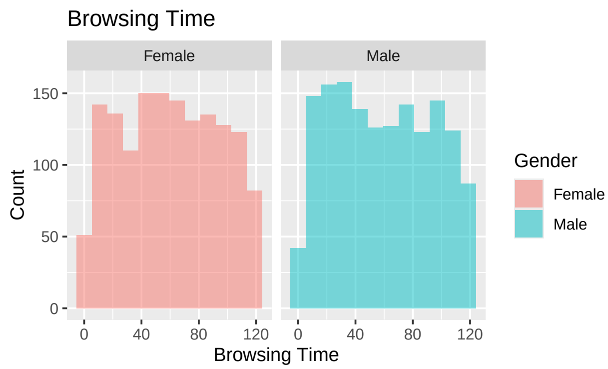


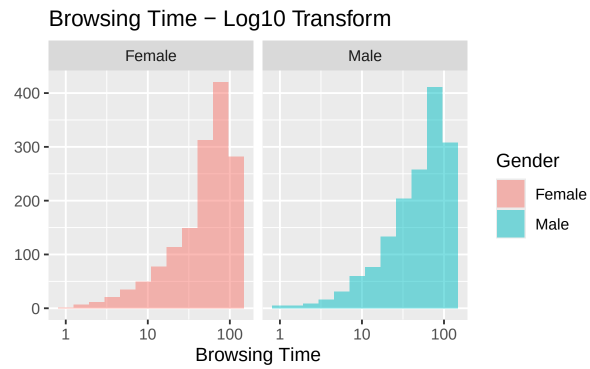
For purchasing amount, there seems to be no significant difference in the median purchase amount for men and women, although there are more women with purchases above 400 USD.

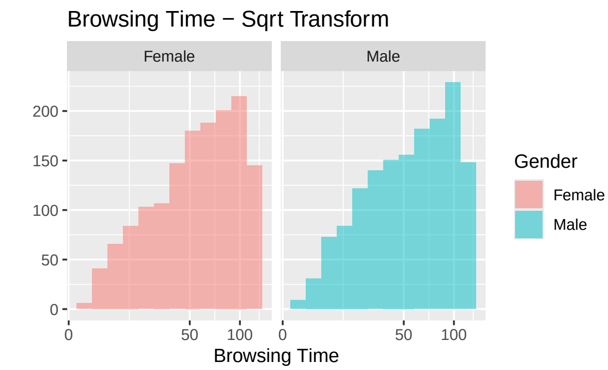
Regarding browsing time. On median, women and men both spend around 60 minutes browsing the e-commerce application. However, the distribution for males is very slightly skewed to the left (although somewhat equalized); the female distribution is more uniform.

Apply a logarithmic or square root transformation on Browsing_Time and evaluate changes in skewness.

```
browse_dist_gender_noscale <- ecom_pbh %>% ggplot(aes(x=Browsing_Time, fill=Gender)) + geom_histogram(al

browse_dist_gender_log10 <- ecom_pbh %>% ggplot(aes(x=Browsing_Time, fill=Gender)) + geom_histogram(alph

browse_dist_gender_sqrt <- ecom_pbh %>% ggplot(aes(x=Browsing_Time, fill=Gender)) + geom_histogram(alpha

plot_grid(browse_dist_gender_noscale, browse_dist_gender_log10, browse_dist_gender_sqrt)
```







Both transformations skewed the distribution female and male browsing times to the right.

Fit a simple linear regression model predicting Purchase_Amount based on Browsing_Time. Interpret the results.

```
pur_vs_browse_model <- lm(Purchase_Amount~Browsing_Time, data = ecom_pbh)

summary(pur_vs_browse_model)
```

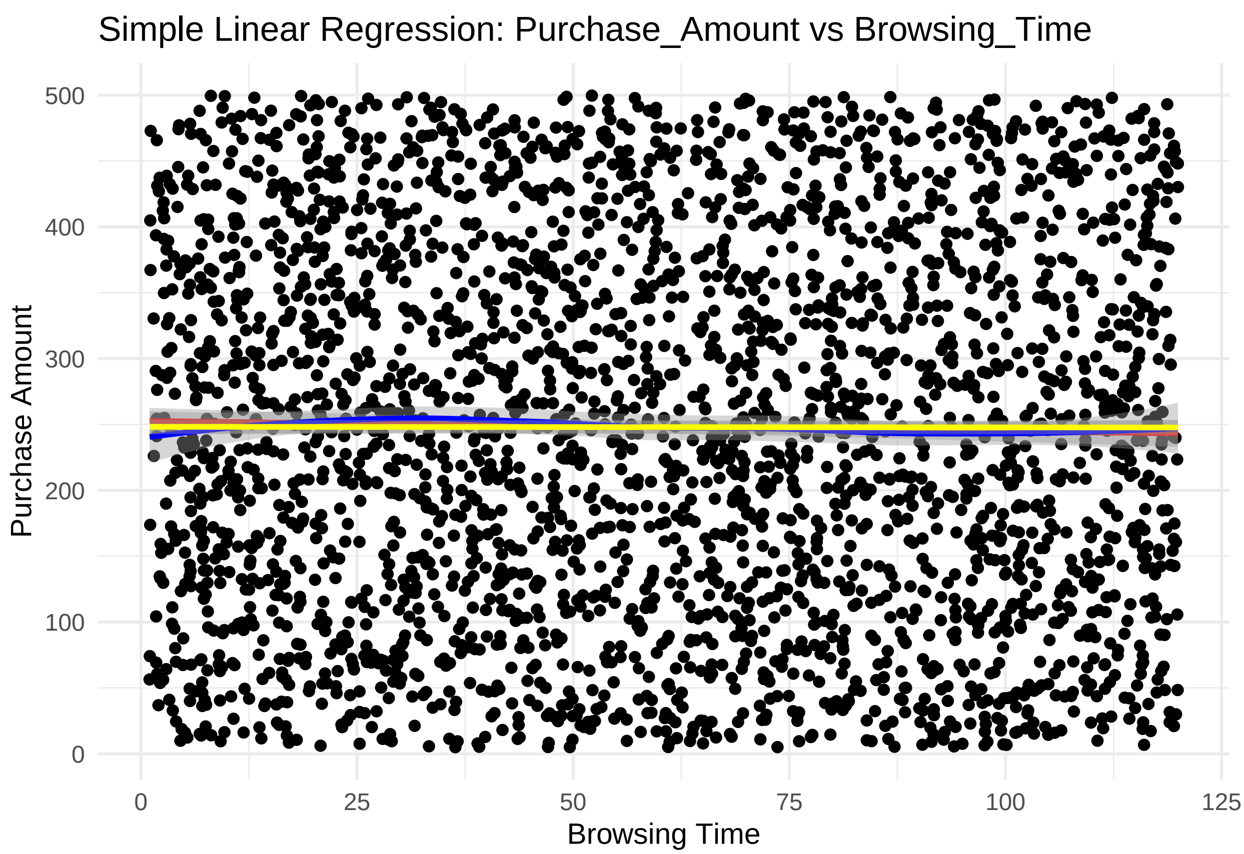
```
##
## Call:
## lm(formula = Purchase_Amount ~ Browsing_Time, data = ecom_pbh)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -244.867 -120.473   -2.946  118.246  254.069
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   252.65596    5.17524  48.820   <2e-16 ***
## Browsing_Time  -0.07839    0.07501  -1.045    0.296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 140.9 on 2998 degrees of freedom
## Multiple R-squared:  0.0003642,  Adjusted R-squared:  3.075e-05
## F-statistic: 1.092 on 1 and 2998 DF,  p-value: 0.2961
```

From the summary of the linear regression model, we can see that the intercept has a very significant effect, given by the p-value. However, we can see that the p-value of Browsing time has no significant effect on the purchase amount, leading us to conclude that the browsing time of the customers of the e-commerce site has no significant effect, or significant association, on the purchase amount.

Use ggplot2 (or equivalent) to create scatter plots and regression lines.

```
ecom_pbh %>% ggplot(aes(x=Browsing_Time, y=Purchase_Amount)) +
  geom_point() +
  geom_smooth(method="lm", color="red") +
  geom_smooth(method="loess", color="blue") +
  geom_smooth(method="gam", color="yellow") +
  labs(title = "Simple Linear Regression: Purchase_Amount vs Browsing_Time",
       x = "Browsing Time",
       y = "Purchase Amount") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
```

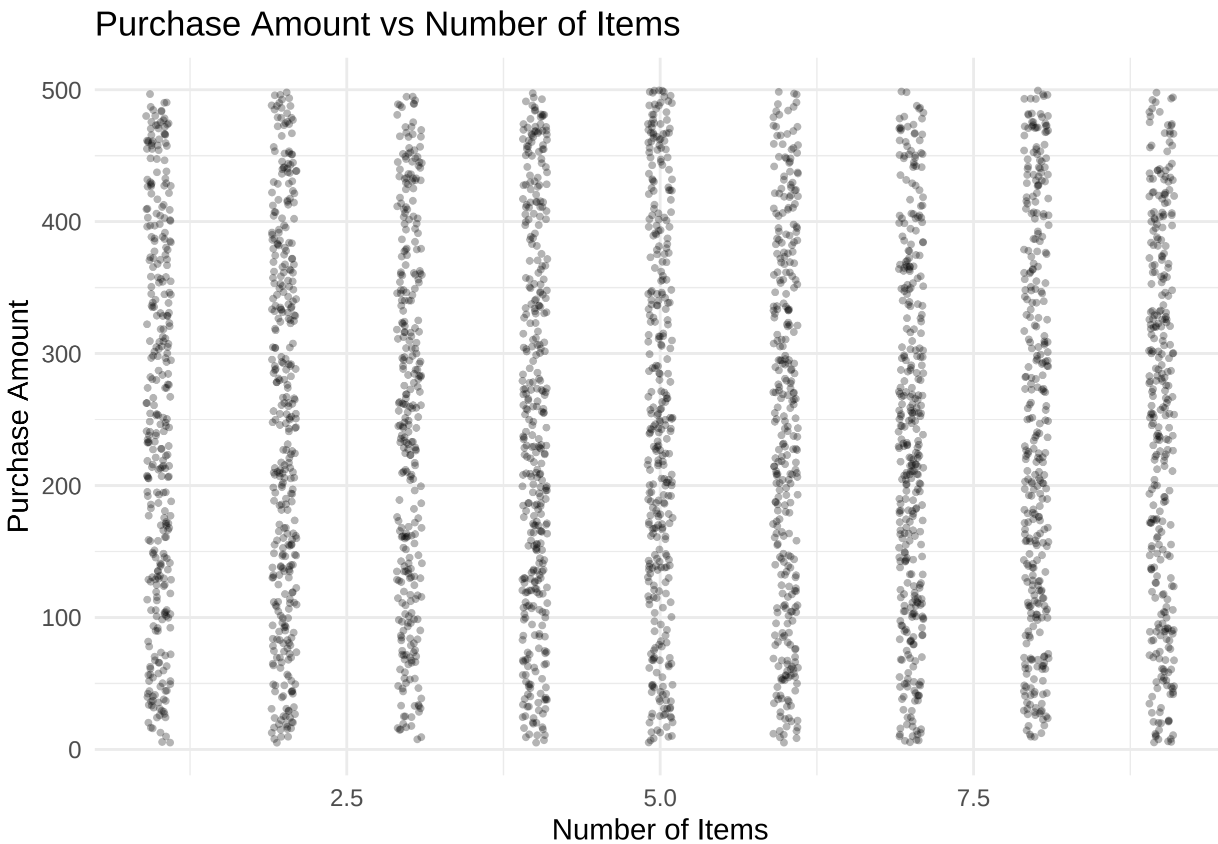


All regression lines are stuck dead center of the plot, agreeing with our regression model above that there is no meaningful relationship between browsing time and purchase amount.

Bivariate Data Analysis

Create scatter plots to explore the relationship between Purchase_Amount and Number_of_Items.

```
ecom_pbh %>% ggplot(aes(x=Number_of_Items, y=Purchase_Amount)) +
  #geom_hex(bins=30)+
  geom_jitter(alpha = 0.3, size = 0.7, width = 0.1, height = 0.1) +
  labs(title="Purchase Amount vs Number of Items", x="Number of Items", y="Purchase Amount") +
  theme_minimal()
```



Fit a polynomial regression model for Purchase_Amount and Browsing_Time and compare it with a simple linear model.

```
pur_vs_browse_model_poly <- lm(Purchase_Amount ~ poly(Browsing_Time, degree = 2), data = ecom_pbh)
summary(pur_vs_browse_model_poly)
```

```
##
## Call:
## lm(formula = Purchase_Amount ~ poly(Browsing_Time, degree = 2),
##     data = ecom_pbh)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -245.47 -120.41   -3.49  118.25  255.85
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
```

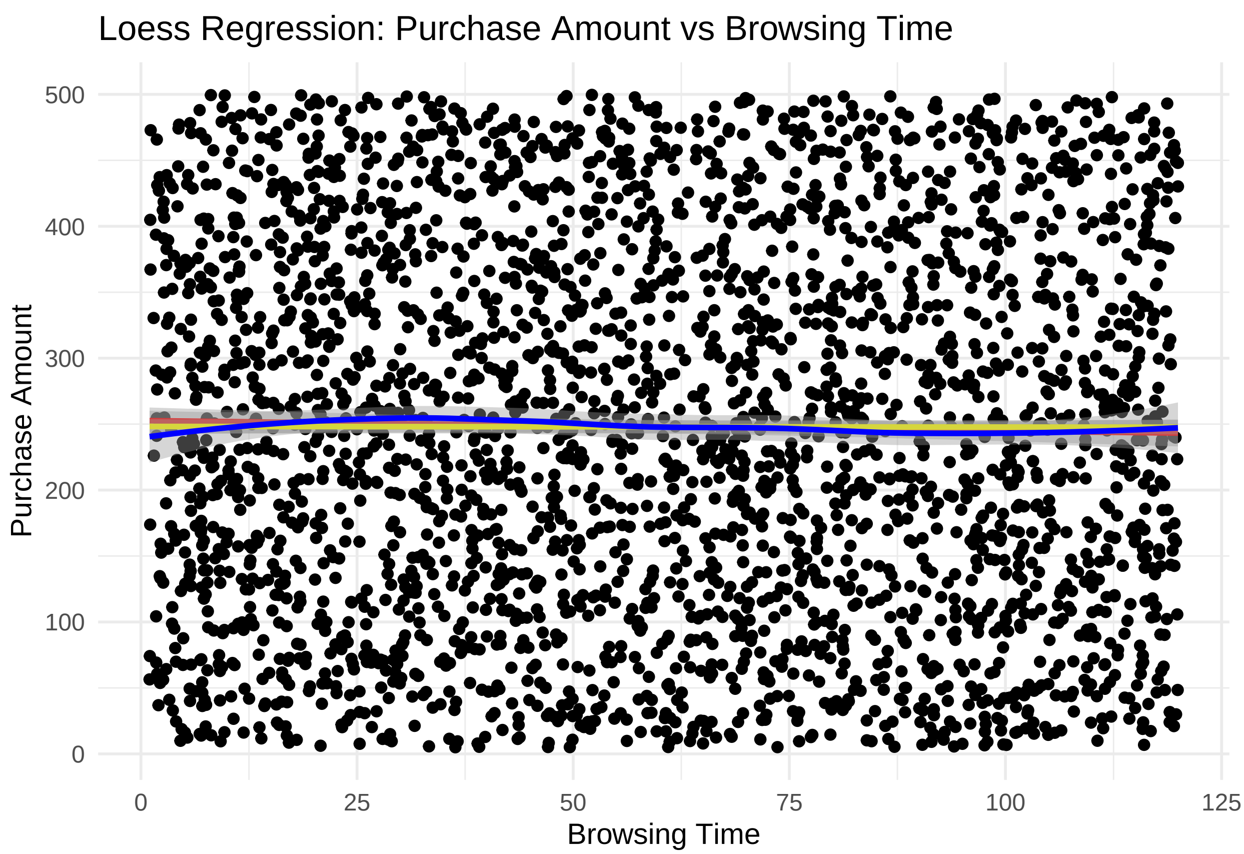
```
## (Intercept)                      247.963      2.572  96.397   <2e-16 ***
## poly(Browsing_Time, degree = 2)1 -147.227    140.892  -1.045    0.296
## poly(Browsing_Time, degree = 2)2  -68.129    140.892  -0.484    0.629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 140.9 on 2997 degrees of freedom
## Multiple R-squared:  0.0004422,  Adjusted R-squared:  -0.0002249
## F-statistic: 0.6629 on 2 and 2997 DF,  p-value: 0.5154
```

The model with the polynomial fit (F-stat = 0.6629, p-value = 0.5154) has a worse F-statistic than the earlier model made with the regular regression fit (F-stat = 1.092, p-value: 0.2961).

Apply LOESS (Locally Estimated Scatterplot Smoothing) to Purchase_Amount vs. Browsing_Time and visualize the results.

```
ecom_pbh %>% ggplot(aes(x=Browsing_Time, y=Purchase_Amount)) +
  geom_point() +
  geom_smooth(method="lm", color="red") +
  geom_smooth(method="gam", color="yellow") +
  geom_smooth(method="loess", color="blue") +
  labs(title = "Loess Regression: Purchase Amount vs Browsing Time",
       x = "Browsing Time",
       y = "Purchase Amount") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using formula = 'y ~ x'
```



As seen earlier, the LOESS application on Purchase Amount vs Browsing time scatter does not have much difference from other regression mapping.

Compare robust regression methods (Huber or Tukey regression) with ordinary least squares (OLS).

```r
cat("Robust Regression with Huber:")
```

```
## Robust Regression with Huber:
```

```r
# Fit Huber regression
model_huber <- rlm(Purchase_Amount ~ Browsing_Time, data = ecom_pbh, method = "M")
summary(model_huber)
```

```
##
## Call: rlm(formula = Purchase_Amount ~ Browsing_Time, data = ecom_pbh,
##     method = "M")
## Residuals:
##      Min       1Q   Median       3Q      Max
## -244.818 -120.331   -2.848  118.291  254.289
##
## Coefficients:
##                Value    Std. Error t value
## (Intercept)   252.6462   5.3363    47.3448
## Browsing_Time  -0.0803   0.0773    -1.0378
##
## Residual standard error: 176.9 on 2998 degrees of freedom
```

```r
cat("\n\nRobust Regression with Tukey")
```

```
##
##
## Robust Regression with Tukey
```

```r
# Fit Tukey regression
model_tukey <- rlm(Purchase_Amount ~ Browsing_Time, data = ecom_pbh, method = "MM")
summary(model_tukey)
```

```
##
## Call: rlm(formula = Purchase_Amount ~ Browsing_Time, data = ecom_pbh,
##     method = "MM")
## Residuals:
##      Min       1Q   Median       3Q      Max
## -244.818 -119.795   -2.611  118.545  255.129
##
## Coefficients:
##                Value    Std. Error t value
## (Intercept)   252.8370   5.6124    45.0499
## Browsing_Time  -0.0894   0.0813    -1.0995
##
## Residual standard error: 169.1 on 2998 degrees of freedom
```

```r
cat("\n\nRegression with OLS")
```

```
##
##
## Regression with OLS
```

```r
summary(pur_vs_browse_model)
```

```
##
## Call:
## lm(formula = Purchase_Amount ~ Browsing_Time, data = ecom_pbh)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -244.867 -120.473   -2.946  118.246  254.069
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   252.65596    5.17524  48.820   <2e-16 ***
## Browsing_Time  -0.07839    0.07501  -1.045    0.296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 140.9 on 2998 degrees of freedom
## Multiple R-squared:  0.0003642,  Adjusted R-squared:  3.075e-05
## F-statistic: 1.092 on 1 and 2998 DF,  p-value: 0.2961
```

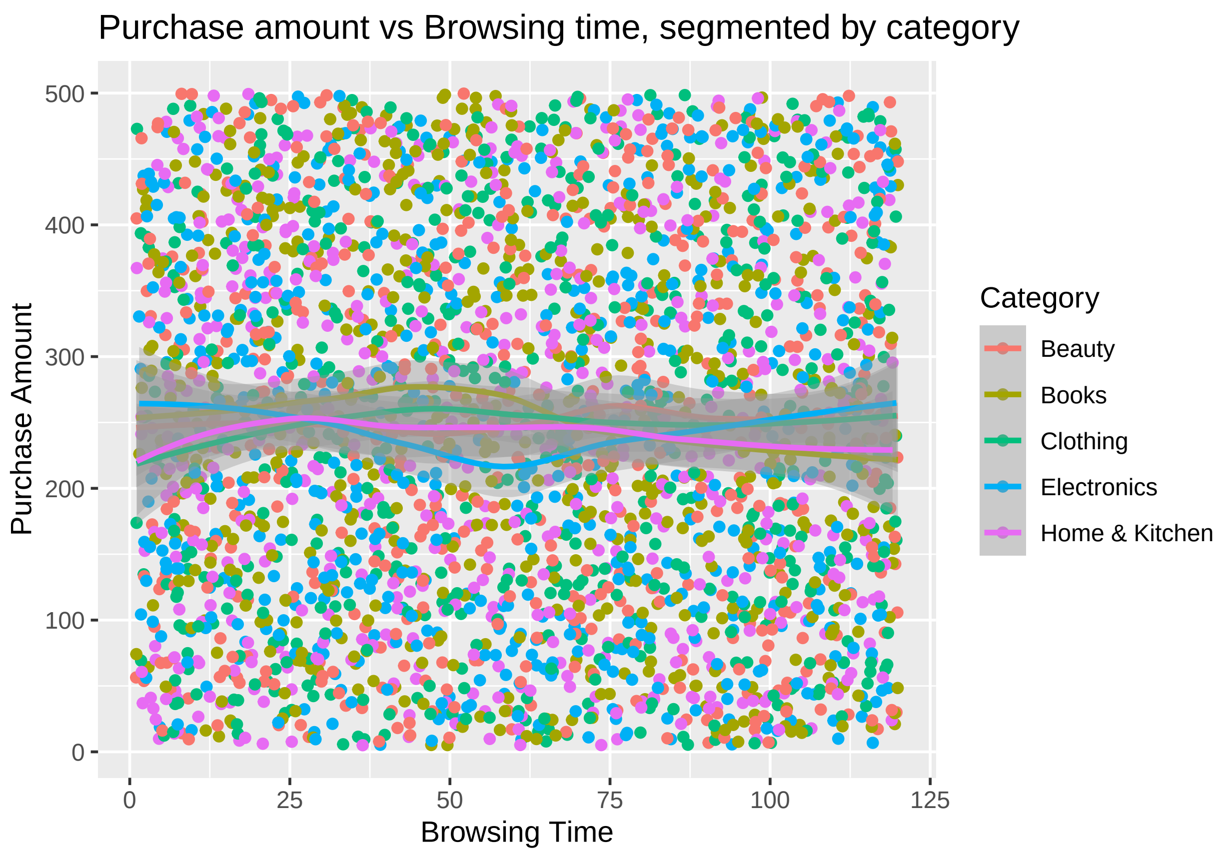
Both robust regression methods yielded similar t-values for browsing time with the OLS model, which indicates that all models does not find a reliable relationship between browsing time and purchase amount.

Trivariate/Hypervariate Data Analysis

Explore interaction effects between Browsing_Time and Category on Purchase_Amount using interaction plots.

```
ecom_pbh %>% ggplot(aes(x=Browsing_Time, y=Purchase_Amount, color=Category, group = Category)) +
  geom_point() +
  geom_smooth(method = "loess")+
  labs(title="Purchase amount vs Browsing time, segmented by category", x="Browsing Time", y="Purchase A
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

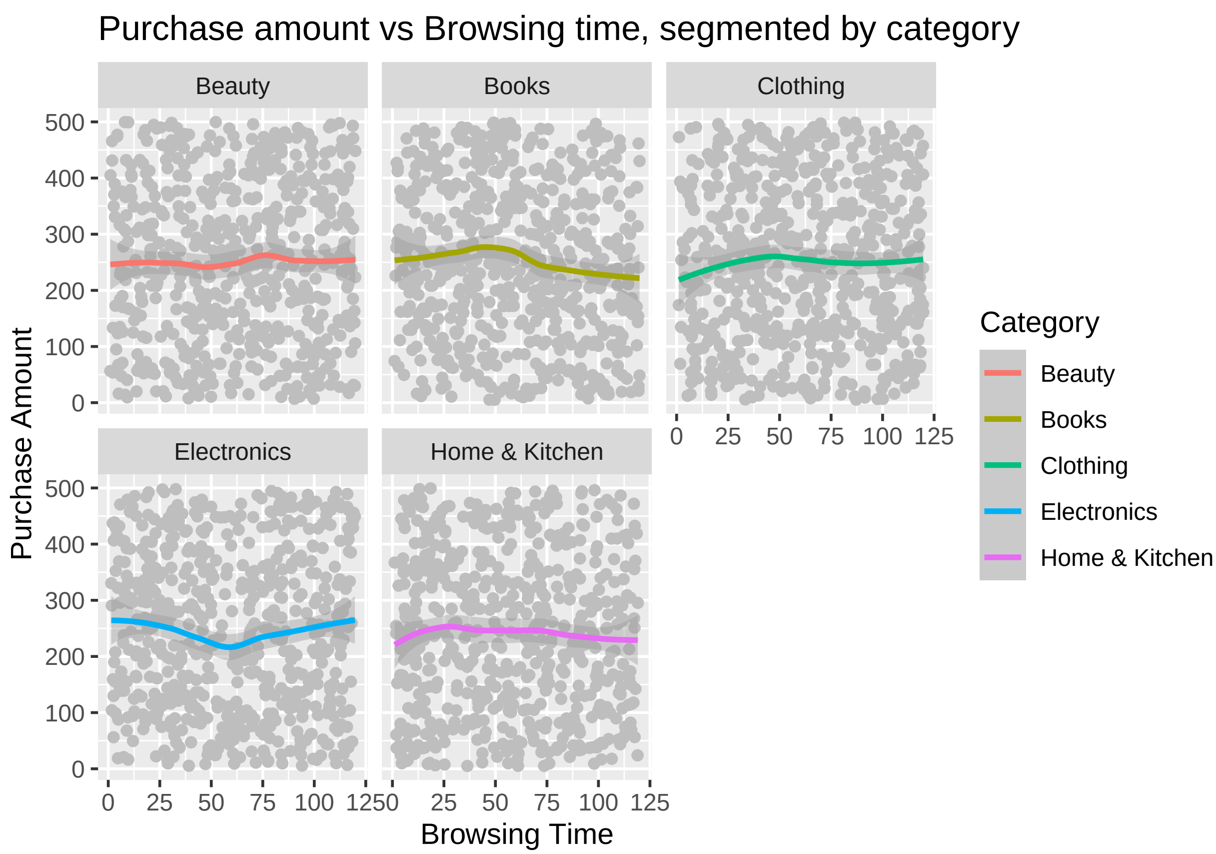


The LOESS fit of each category does not differ much with each other, as all the data points for all categories are widely distributed.

Create coplots of Purchase_Amount against Browsing_Time for different levels of Category.

```
ecom_pbh %>% ggplot(aes(x=Browsing_Time, y=Purchase_Amount, color=Category, group = Category)) +
  geom_point(color="grey") +
  geom_smooth(method = "loess")+
  labs(title="Purchase amount vs Browsing time, segmented by category", x="Browsing Time", y="Purchase A
  facet_wrap(~Category)
```

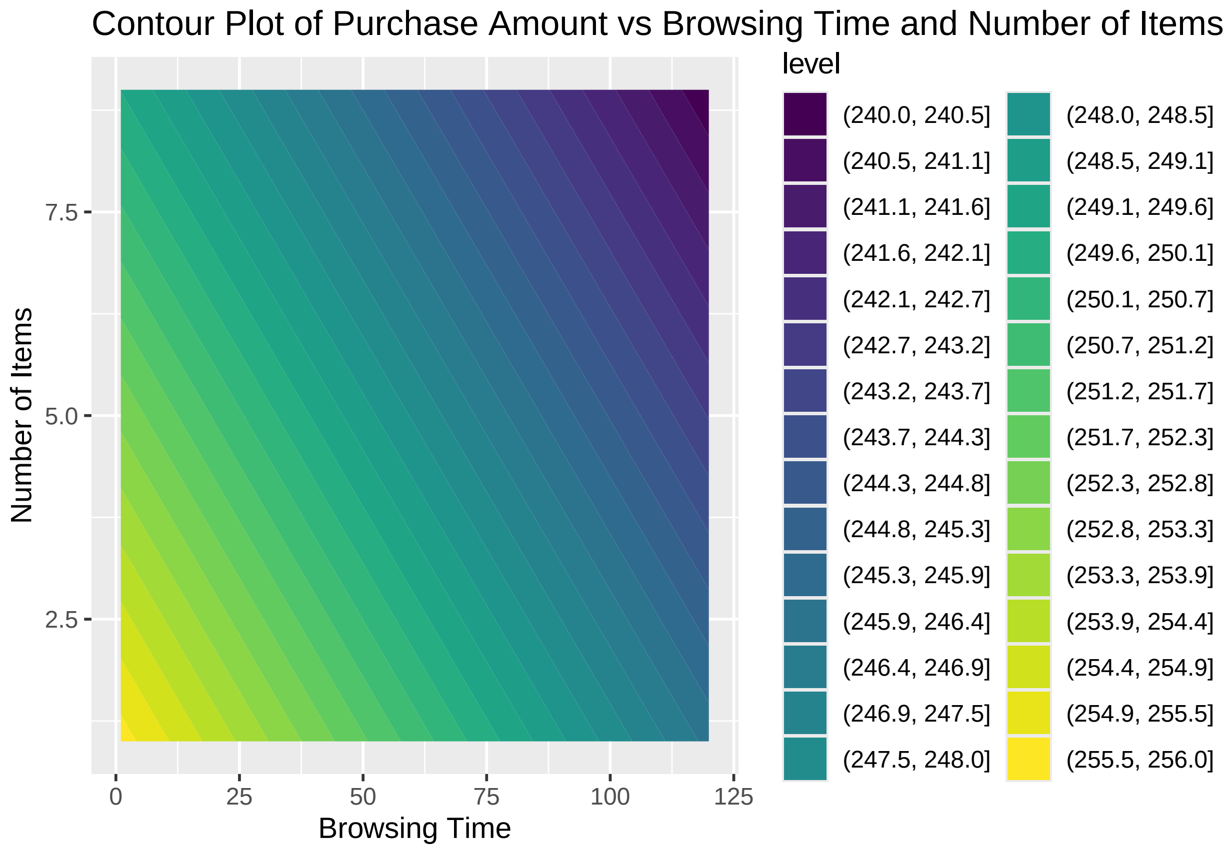
```
## `geom_smooth()` using formula = 'y ~ x'
```



As from above, the LOESS fit hugs the central horizontal line at around y=250, indicating that, for all categories, the browsing time does not significantly affect the purchase amount. Although, it can be noted that for beauty, clothing, and electronics, the trend does increase slightly.

Use level plots or contour plots to visualize relationships between Browsing_Time, Number_of_Items, and Purchase_Amount.

```
purchase_browse_numItems_model <- lm( Purchase_Amount ~ Browsing_Time + Number_of_Items, data=ecom_pbh)

grid <- expand.grid(
  Browsing_Time = seq(min(ecom_pbh$Browsing_Time), max(ecom_pbh$Browsing_Time), length.out = 100),
  Number_of_Items = seq(min(ecom_pbh$Number_of_Items), max(ecom_pbh$Number_of_Items), length.out = 100)
)

grid$Purchase_Amount <- predict(purchase_browse_numItems_model, newdata = grid)

grid %>% ggplot(aes(x=Browsing_Time, y=Number_of_Items, z=Purchase_Amount)) +
  geom_contour_filled(bins=30) +
  labs(title="Contour Plot of Purchase Amount vs Browsing Time and Number of Items", x="Browsing Time",
```



As we can see from the contour plot, there are diagonal ridges forming across the plot, appearing equally spaced. as both browsing time and number of items increase, so does the purchase amount of the customer/order.

Perform multiple regression with Purchase_Amount as the dependent variable and Browsing_Time, Number_of_Items, and Satisfaction_Score as predictors. Perform model selection and assess variable importance.

First, let's use the regsubsets function from the leaps package to pick our subsets for the model.

```
ecom_pbh_multReg_features <- ecom_pbh %>% select(Browsing_Time, Number_of_Items, Satisfaction_Score, Pu

ecom_pbh_best_subset <- regsubsets(Purchase_Amount ~ ., data = ecom_pbh_multReg_features)

summary(ecom_pbh_best_subset)$which
```

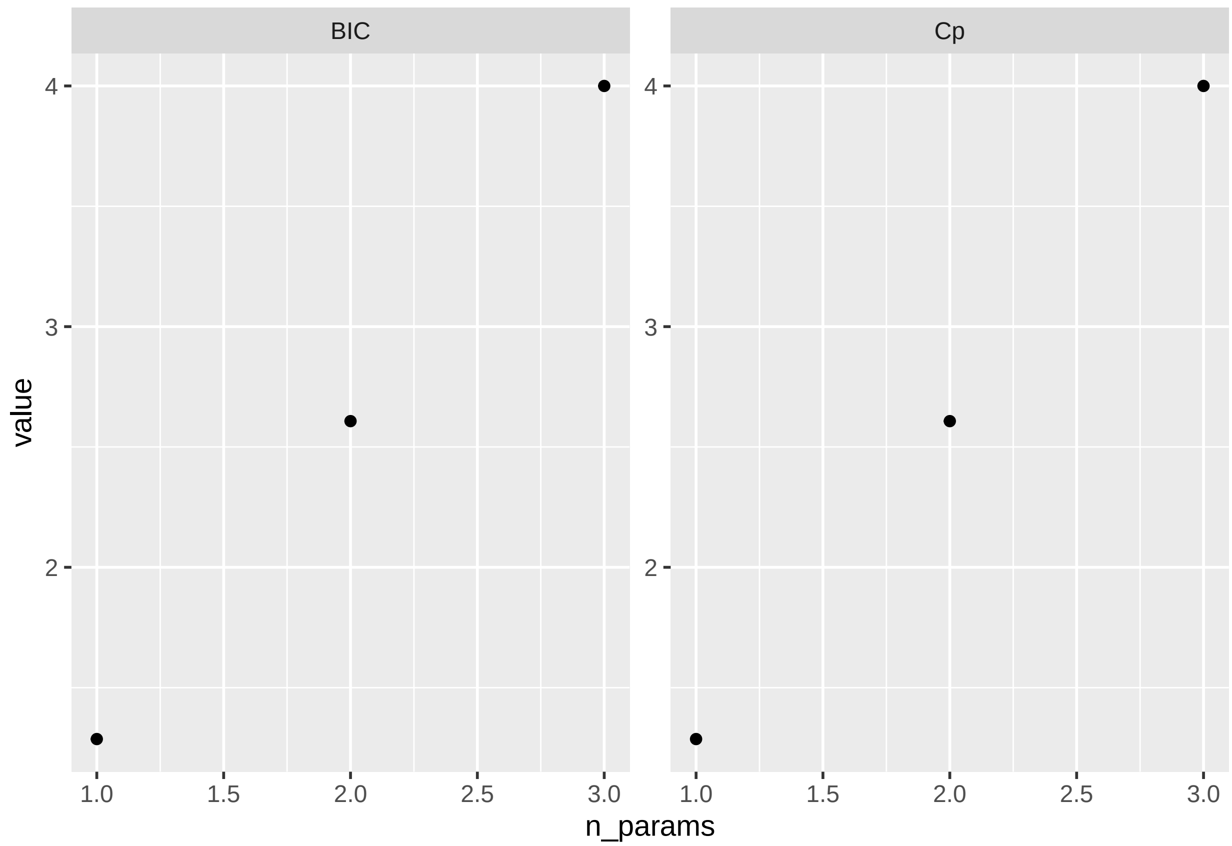
```
##   (Intercept) Browsing_Time Number_of_Items Satisfaction_Score
## 1        TRUE          TRUE           FALSE              FALSE
## 2        TRUE          TRUE           FALSE               TRUE
## 3        TRUE          TRUE            TRUE               TRUE
```

We find that, for one predictor, Browsing_Time is most effective. For two: Browsing Time and Satisfaction Score; for three, all predictor variables.

Let's now use Mallow's CP and Bayesian Information Criterion (BIC) to figure out how many predictors in the model works best:

```
cp_df = data.frame(value  = summary(ecom_pbh_best_subset)$cp,
                   n_params = seq_along(summary(ecom_pbh_best_subset)$cp),
                   type = "Cp")
```

```
bic_df = data.frame(value  = summary(ecom_pbh_best_subset)$cp,
                    n_params = seq_along(summary(ecom_pbh_best_subset)$bic),
                    type = "BIC")

model_selection_criterion_df = rbind(cp_df, bic_df)
model_selection_criterion_df %>% ggplot(aes(x = n_params, y = value)) +
    geom_point() + facet_wrap(~ type, scales = "free_y")
```

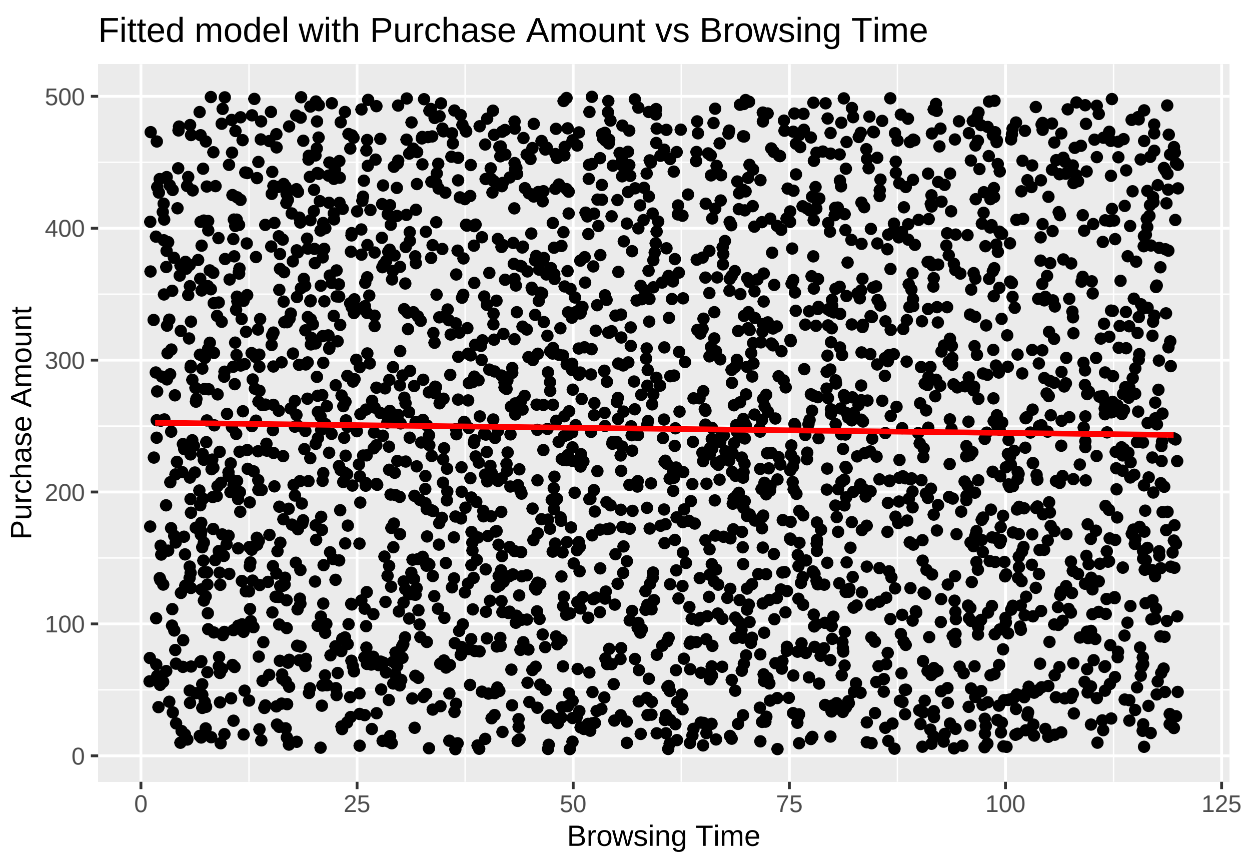


Our BIC and Mallow's Cp plot tells us that we should use only one predictor for our model. Given the previous results in finding the best parameter subset, we end up with the formula: "Purchase Amount ~ Browsing Time".

Finally, let's create our model:

```
ecom_pbh_final_model <- lm(Purchase_Amount ~ Browsing_Time, data = ecom_pbh_multReg_features)
final_fits <- augment(ecom_pbh_final_model)

ecom_pbh_multReg_features_wBrowsingTime_Bins <-  ecom_pbh_multReg_features %>%
  mutate(Browsing_Time_bins = cut_number(Browsing_Time, n=120)) %>%
  separate(Browsing_Time_bins, into = c(NA, "lo", "hi", NA), remove = FALSE, sep = "\\[|\\(|\\)|\\]|,")
  mutate(bin_mean = (as.numeric(lo) + as.numeric(hi)) / 2)

ecom_pbh_grid_final <- expand.grid(
  Browsing_Time = unique(ecom_pbh_multReg_features_wBrowsingTime_Bins$bin_mean)
  ) %>% tibble()
```

```r
ecom_pbh_fits_on_grid <- augment(ecom_pbh_final_model, newdata = ecom_pbh_grid_final)

# ecom_pbh_fits_on_grid <- merge(ecom_pbh_fits_on_grid,
#                                unique(ecom_pbh_multReg_features_wBrowsingTime_Bins[,c("Browsing_Time_b
#                                by.x = "Browsing_Time", by.y = "bin_mean")

ecom_pbh_multReg_features_wBrowsingTime_Bins %>% ggplot(aes(x=Browsing_Time, y=Purchase_Amount)) +
  geom_point() +
  geom_line(aes(y = .fitted), data = ecom_pbh_fits_on_grid, color="red", linewidth = 1)+
  labs(title = "Fitted model with Purchase Amount vs Browsing Time", x = "Browsing Time", y = "Purchase
```




```r
summary(ecom_pbh_final_model)
```

```
##
## Call:
## lm(formula = Purchase_Amount ~ Browsing_Time, data = ecom_pbh_multReg_features)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -244.867 -120.473   -2.946  118.246  254.069
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   252.65596    5.17524  48.820   <2e-16 ***
## Browsing_Time  -0.07839    0.07501  -1.045    0.296
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 140.9 on 2998 degrees of freedom
## Multiple R-squared:  0.0003642,  Adjusted R-squared:  3.075e-05
## F-statistic: 1.092 on 1 and 2998 DF,  p-value: 0.2961
```

Even if our analysis have found that Browsing time as our only predictor for Amount Purchased is the best model, it is still very far off from being an accurate predictive model of amount purchased. Our regression table agrees with this insight, showing no association with browsing time and amount purchased (F-statistic = 1.092, p value = 0.296)