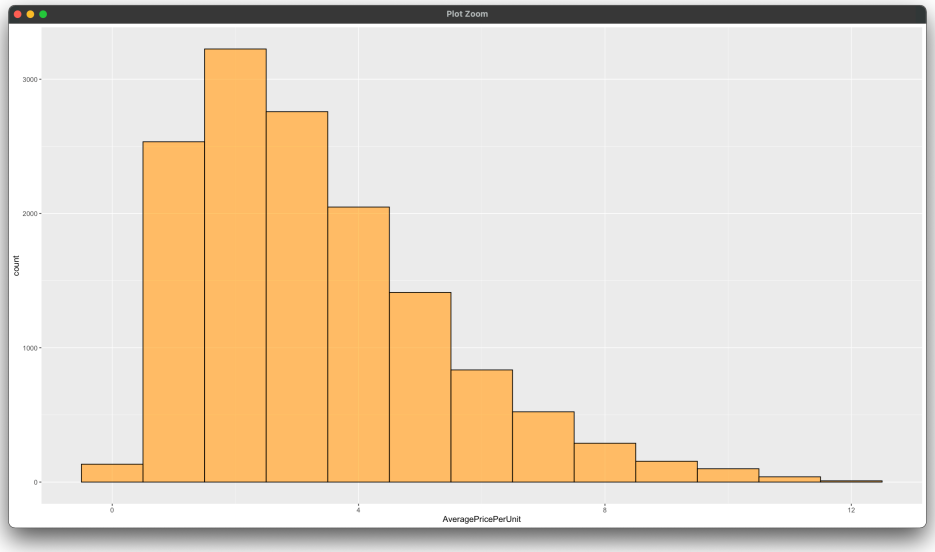
1. All code from your R script.

```
# YUEH-TING WU
# MIS 545 Section 02
# Lab03WuY.R
# This is R programming. Importing a CSV file and dplyr summarize() function to
# display statistic values. And also make these statistics visualization.

# Install the tidyverse package
# installed.packages("tidyverse")

# Load the package. This needs to be done every time when you want use the
# package.
library(tidyverse)

# Set a working directory to my Lab03 folder
setwd("~/MIS 545/Lab03")
print(getwd())

# Read GroceryTransactions.csv into a tibble called groceryTransactions1
groceryTransactions1 <- read_csv(file = "GroceryTransactions.csv",
                                 col_types = "ifffffiffffffin",
                                 col_names = TRUE)

# Display groceryTransaction1 in the console
print(groceryTransactions1)

# Display the first 20 rows of groceryTransaction1 in the console
head(groceryTransactions1, n = 20)

# Display the structure of groceryTransactions1 in the console
str(groceryTransactions1)

# Display the summary of groceryTransactions1 in the console
summary(groceryTransactions1)

# Use the dplyr summarize() function to display the following on the console
# Mean of revenue
print(summarize(.data = groceryTransactions1, mean(Revenue)))
# Median of units sold
print(summarize(.data = groceryTransactions1, median(UnitsSold)))
# Standard deviation of revenue
print(summarize(.data = groceryTransactions1, sd(Revenue)))
# Inter-quartile range of units sold
print(summarize(.data = groceryTransactions1, IQR(UnitsSold)))
# Minimum of revenue
print(summarize(.data = groceryTransactions1, min(Revenue)))
# Maximum of children
print(summarize(.data = groceryTransactions1, max(Children)))

# Create a new tibble called groceryTransaction2 that contains only columns of
# PurchaseDate, Homeowner, Children, AnnualIncome, UnitsSold, and Revenue
groceryTransactions2 <- select(.data = groceryTransactions1,
                               PurchaseDate,
                               Homeowner,
```

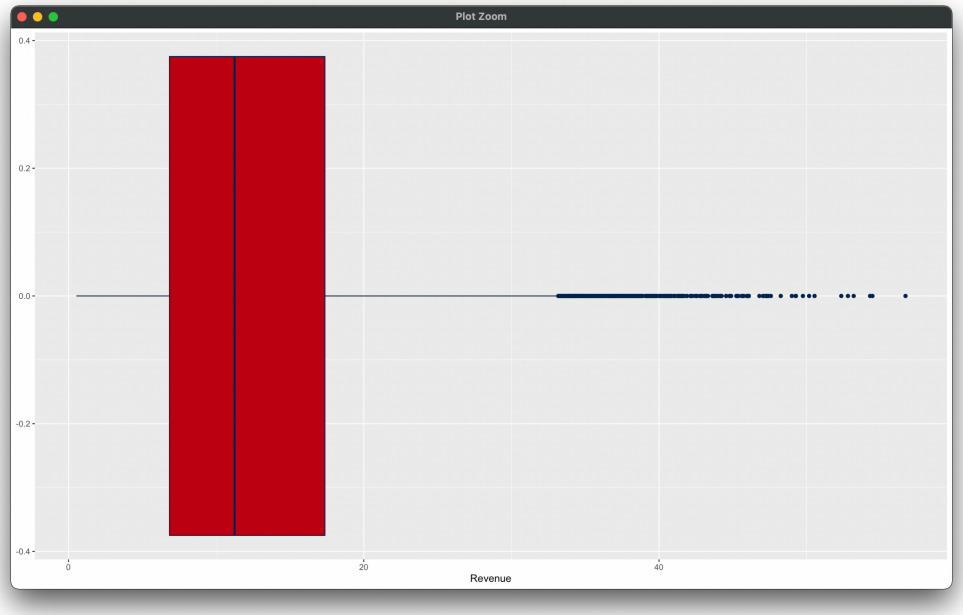
```
                    Children,
                    AnnualIncome,
                    UnitsSold,
                    Revenue)

# Display all of the features in groceryTransactions2 for transactions made by
# non-homeowner with at least 4 children.
# Use filter() to get the result
print(filter(.data = groceryTransactions2,
          Homeowner == "N" &
            Children >= "4"))

# Display all of the records and features in groceryTransaction2 that were
# either made by customers in the $150K + annual income category OR had more
# than 6 units sold.
# Use "pipe" %>% to filter the result
print(groceryTransactions2 %>%
      select(PurchaseDate,
            Homeowner,
            Children,
            AnnualIncome,
            UnitsSold,
            Revenue) %>%
      filter(AnnualIncome == "$150K +" |
              UnitsSold > "6"))

# Display the average transaction revenue grouped by annual income level.
# Sort the results by average transaction revenue from largest to smallest
print(groceryTransactions1 %>%
      group_by(AnnualIncome) %>%
      summarize(averageTransactionRevenue = mean(Revenue)) %>%
      arrange(averageTransactionRevenue),
    n = Inf)

# Create a new tibble called grocerytransaction3 that contains all of the
# features in groceryTransaction2 along with a new feature
# AveragePricePerUnit
groceryTransactions3 <- groceryTransactions2 %>%
# Use the mutate(), AveragePricePerUnit calculated by dividing Revenue by
# UnitsSold
mutate(AveragePricePerUnit = Revenue / UnitsSold)

# Display the groceryTransaction3 in the console
print(groceryTransactions3)

# Use ggplot() to create a histogram of AveragePricePerUnit
histogramAveragePricePerUnit = ggplot(data = groceryTransactions3,
                              aes(x = AveragePricePerUnit))
histogramAveragePricePerUnit + geom_histogram(binwidth = 1,
                                    color = "black",
                                    fill = "orange",
                                    alpha = 0.6)

# Use ggplot() to create a boxplot of revenue
```

```
BoxplotRevenue = ggplot(data = groceryTransactions3,
                        aes(x = Revenue))
BoxplotRevenue + geom_boxplot(color = "#0C234B",
                              fill = "#AB0520")
```

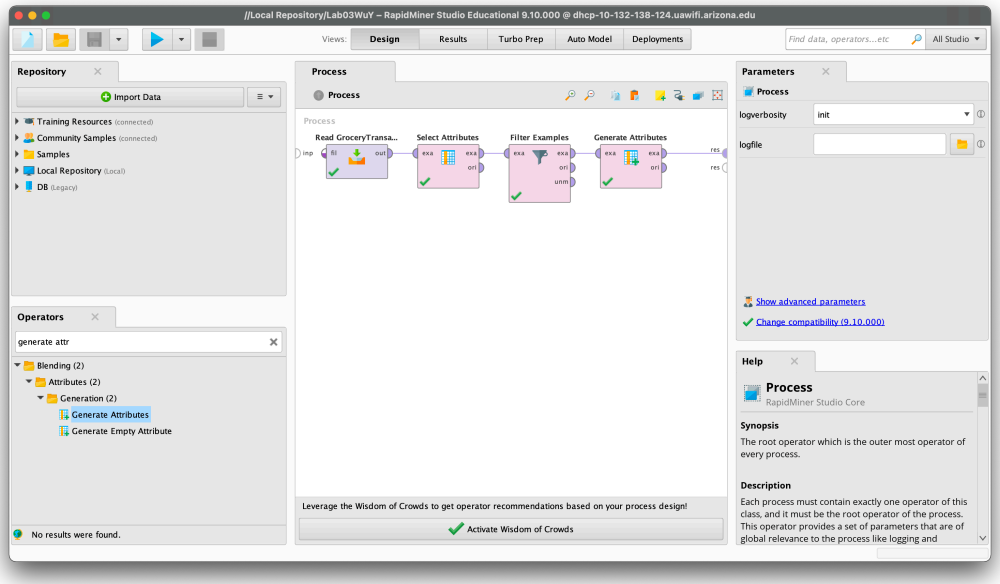
2. Histogram copy/pasted from R




3. Boxplot copy/pasted from R



4. A screenshot of your RapidMiner process




5. A screenshot of your RapidMiner results