From R:

* 1. All code from your R script (Code should be presented single-spaced in a fixed-width font. Adjust the font size so that no lines of code extend to the next line in the document):

# Sanchit Agarwala and Varun Harish Kapuria

# MIS 545 Section 1

# Lab05Group2AgarwalaKapuria.R

# This code is meant to provide an in-depth analysis on the zooSpending,scv

# dataset provided in MIS545 to conduct some data preprocessing tasks,

# and generating a multiple linear regression model.

# Pre-execution --------------------------------------------------------

# By utilizing installed.packages() we can establish if the package is

# installed and if the package isn't in installed can be downloaded using

# install.packages().

# installed.packages()

# install.packages("tidyverse")

# install.packages("dummies", repos = NULL, type="source")

# install.packages("corrplot")

# install.packages("olsrr")

# We can load/run the tidyverse/corrplot/olsrr library using the library()

# command and check if the package is installed can be checked under the

# packages tab on the right of the screen.

library(tidyverse)

library(dummies)

library(corrplot)

library(olsrr)

# By running the setwd() we can set the library as per request and can be

# then retrieved using the getwd() command.

setwd(paste0("/Users/sanchitagarwala/Documents/FallSemester/MIS545/",

"Deliverables/Lab05", sep=""))

# print(getwd())

# Understanding our data set ------------------------

# Read ZooVisitSpending.csv into a tibble called zooSpending (use the data

# types suggested at the bottom of these assignment instructions).

zooSpending <- read\_csv(file = "ZooVisitSpending.csv",

col\_types = "niil",

col\_names = TRUE)

# The print() command displays the tireTread1 in the console.

print(zooSpending)

# The str() command displays the structure of tireTread1 in the console

str(zooSpending)

# The summary() command displays the summary of tireTread1 in the console

summary(zooSpending)

# Recreating the displayAllHistograms() function

displayAllHistograms <- function(tibbleDataset) {

tibbleDataset %>%

keep(is.numeric) %>%

gather() %>%

ggplot() + geom\_histogram(mapping = aes(x=value,fill=key),

color = "black") +

facet\_wrap (~key, scales = "free") +

theme\_minimal()

}

# Calling the displayAllHistograms() function, passing in zooSpending as an

# argument

displayAllHistograms(zooSpending)

# Display a correlation matrix of zooSpending rounded to two decimal places

round(cor(zooSpending), 2)

# Displaying a correlation plot using the "number" method and limit output to

# the bottom left

corrplot(cor(zooSpending),

method = "number",

type = "lower")

# Generating the linear regression model and saving it in an object called

# zooSpendingModel

zooSpendingModel <- lm(data = zooSpending,

formula = VisitSpending ~ PartySize +

MilesFromZoo + Member)

# Alternative Method of the above

# zooSpendingModel <- lm(data = zooSpending,

# formula = VisitSpending ~ .)

# Displaying the beta coefficients for the model on the console

print(zooSpendingModel)

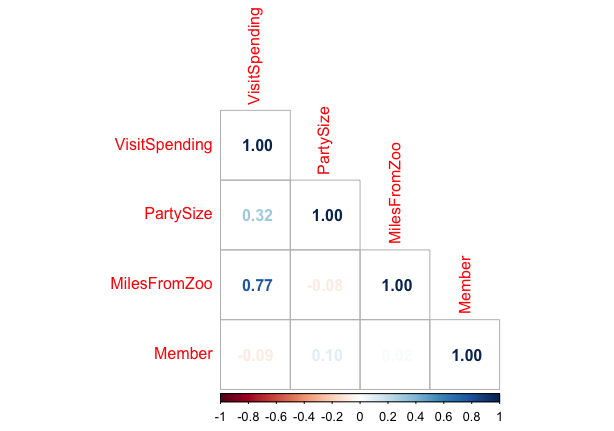
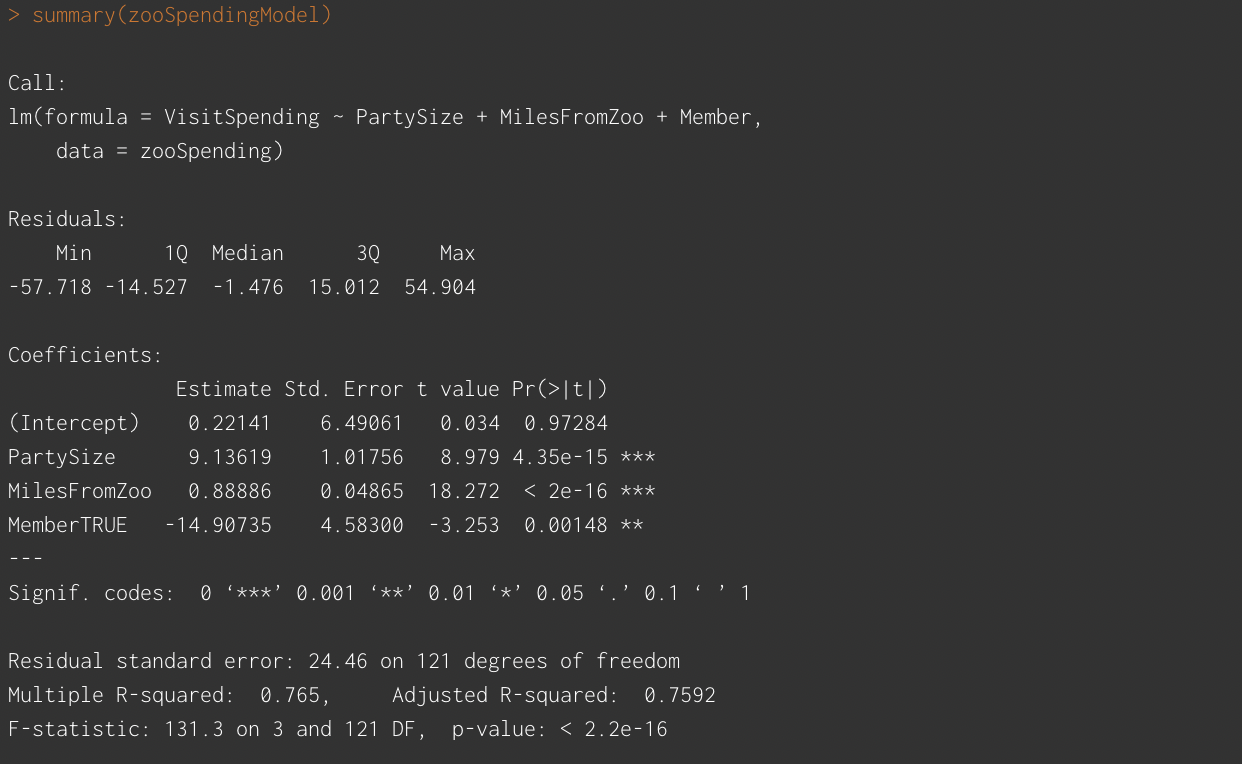
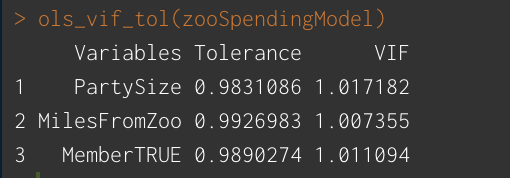
# Displaying the linear regression model results using the summary() function

summary(zooSpendingModel)

# Testing for multicollinearity using the ols\_vif\_tol() function

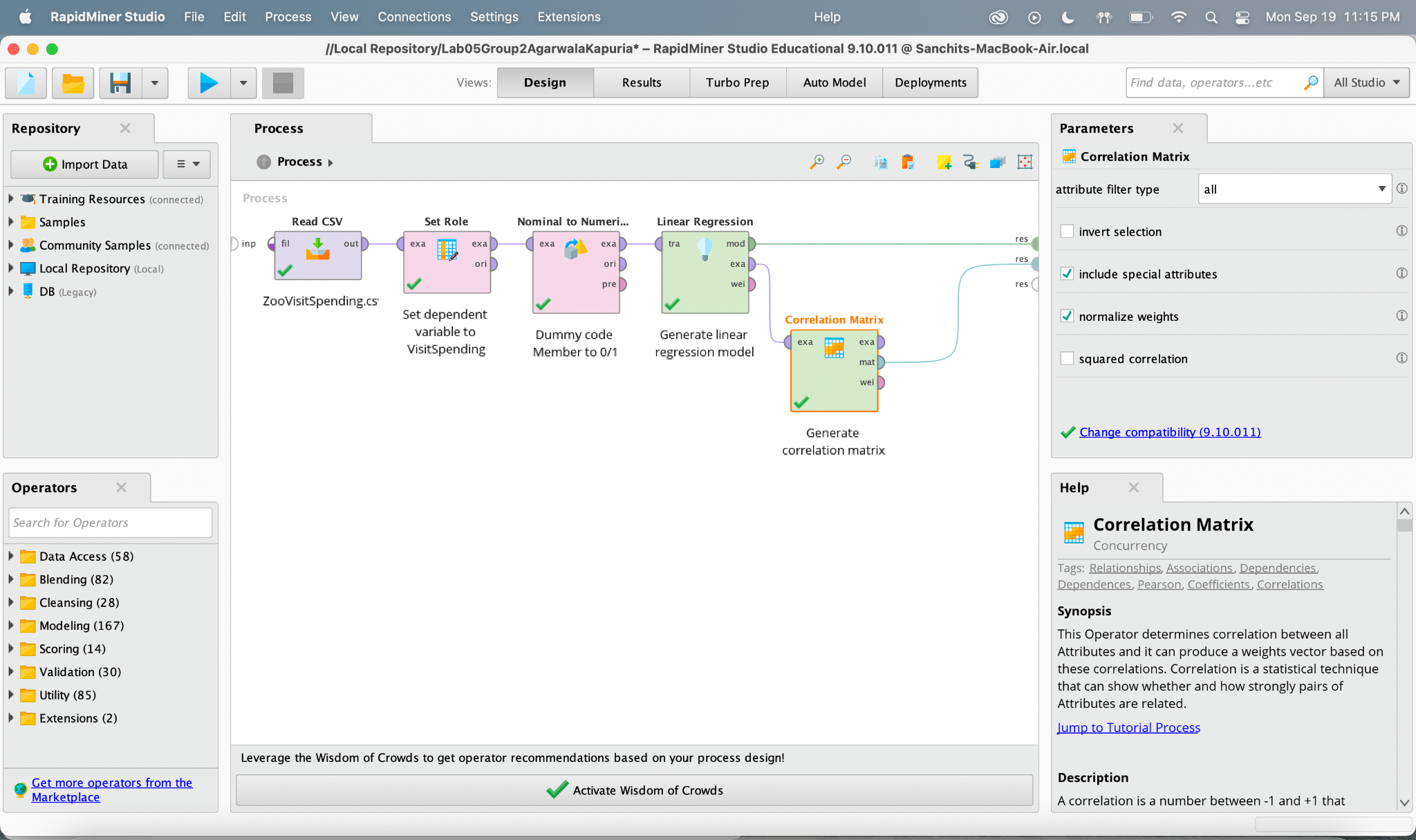
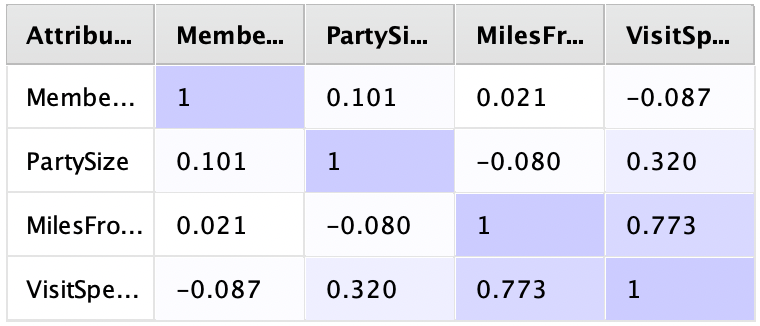
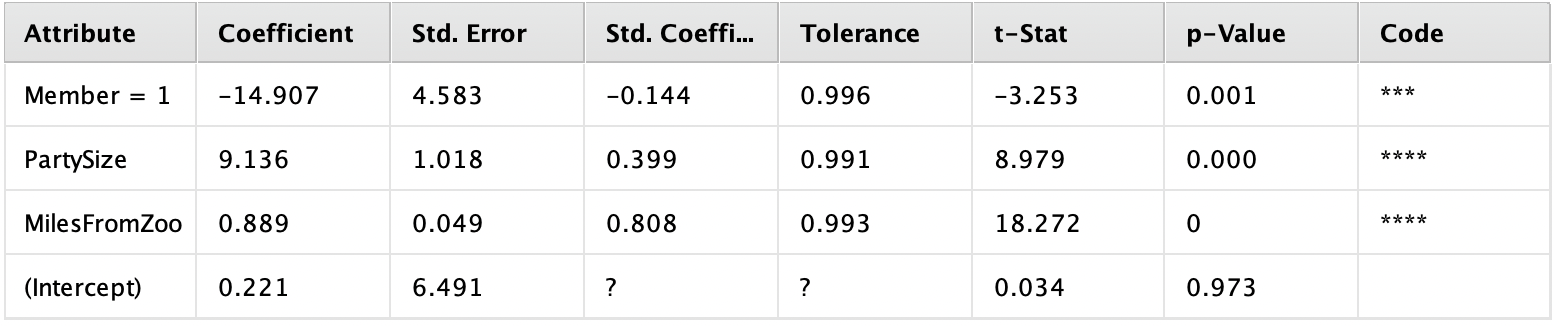
ols\_vif\_tol(zooSpendingModel)

* 1. The correlation plot

1. 
   1. The model summary
2. 
   1. The results from the test for multicollinearity
3. 

From RapidMiner:

* 1. A screenshot of your process (include the entire RapidMiner window including the title bar of the application)

1. 
   1. The correlation matrix
2. 
   1. A screenshot of the linear regression model results
3. 
4. Answer the following question in a sentence: Within the model, which variables are statistically significant?
5. Member, PartySize and, MilesFromZoo are statistically significant
6. Answer the following question in a sentence: How much of the variance in zoo spending can be explained by the variance in party size, miles from the zoo, and zoo membership?
7. 76.5% of variance in zoo spending can be explained by the variance in party size, miles from the zoo, and zoo membership
8. Answer the following question in a sentence: Within the model, how much more/less will zoo spending be with each additional guest in a party?
9. With each additional guest, the zoo spending will increase by $9.136
10. Answer the following question in a sentence: Within the model, how much more/less is zoo spending for members compared with non-members? Explain why this might be the case.
11. Members spend $14.90 less than a non-member, this might be because of some member-only discounts
12. Answer the following question in a sentence: Within the model, how much more/less will spending be for each additional mile traveled to visit the zoo? Explain why this might be the case.
13. For each mile extra traveled, the zoo spending will increase by $0.88. This might be because people who have traveled far have the urge to spend more as they will have the chance to visit the zoo less.
14. Answer the following question in a sentence: Does the model suffer from multicollinearity? If so, what could be done to rectify it? If not, why?
15. No, it does not suffer from multicollinearity, as the tolerance values are greater than 0.2 and VIF values are less than 5. To rectify it, we will remove the correlated variables.