```
## Project:  STA 215, Fall 2023, Final Project
# Located:   Kline TCNJ Google Drive
# File Name: STA215-Lopez.R
# Date:      2023_11_20
# Who:       Jillian Lopez

## Load packages
# NOTE: Run base.R if these commands return an error!
library(readr)
library(dplyr)
library(tidytext)
library(tidyverse)
library(ggplot2)
library(haven)
library(forcats)
library(psych)

# Load data
data <- read_delim("raw_data.csv")

##############################################################################
############### STEP 1: Table 1 (Descriptive Table)          ###################
##############################################################################

# EXAMINE Time (Quantitative Variable 1)
table(data$Time)
mean(data$Time)
sd(data$Time)
summary(data$Time)

# EXAMINE Streams (Quantitative Variable 2)
table(data$Streams)
mean(data$Streams)
sd(data$Streams)

# EXAMINE Swifts_Emotions (Qualitative Variable 1)
table(data$Swifts_Emotions)

# EXAMINE Theme (Qualitative Variable 2)
table(data$Theme)

##############################################################################
###################  STEP 2: Table  2 (Contingency Table)    ###################
##############################################################################

table(data$Swifts_Emotions, data$Theme)

##############################################################################
###################   STEP 3: Chi squared test               ###################
##############################################################################

chisq.test(data$Swifts_Emotions, data$Theme)

##############################################################################
###################  STEP 4: ANOVA                           ###################
##############################################################################

# Perform ANOVA
anova <- aov(data$Streams ~ data$Theme)
# Summarize ANOVA results
summary(anova)

##############################################################################
###################  STEP 5: Correlation                     ###################
```

```
################################################################################

cor(data$Streams, data$Time)

################################################################################
####################  STEP 6: Linear Regression               ####################
################################################################################

## examine the scatter plot ##
linear_plot <- plot(data$Time, data$Streams)
print(linear_plot)

## calculate linear regression ##
linear_relationship <- lm(data$Streams ~ data$Time)
summary(linear_relationship)

################################################################################
####################  STEP 7: Figure 1                        ####################
################################################################################

## slope ##
abline(linear_relationship, col = "red")

## mean of X on the x-axis ##
abline(v = 232.5, col = "blue")

## mean of Y on the y-axis ##
abline(h = 264782281, col = "green")

################################################################################
####################  STEP 8: Examine residuals               ####################
################################################################################

## plot the residuals ##
plot(data$Time, residuals(linear_relationship))

## add a horizontal line at zero to indicate the baseline ##
abline(h = 0, col = "red")
```