```
Section 1: Importing and Summarizing Data

Q1

# Load tidyverse

library(tidyverse)



# Read in the data
taxis <- read_csv("/usr/local/share/datasets/taxis.csv")



# Explore the data

head(taxis)

Q2

# Take a glimpse at your data

glimpse(taxis)



# Check out the structure of your data

str(taxis)



# Summarize your data

summary(taxis)

Q3

# Create total_amount_col

total_amount_col <- taxis[,18]

total_amount_col


# Extract the 125th row
```

```
row_125 <- taxis[125,]

row_125


#Find passenger_count_1031

passenger_count_1031 <- taxis[1031,4]

passenger_count_1031

Q4

# Find the mean and standard deviation of trip_distance

mean(taxis$trip_distance)

sd(taxis$trip_distance)



# Find the 25th, 50th, and 75th percentiles of trip_distance

quantile(taxis$trip_distance, c(.25,.50,.75))

Q5

# Find the numeric columns

numeric_cols <- sapply(taxis, is.numeric)

numeric_cols



# Create taxis_numeric

taxis_numeric <-
taxis,c("VendorID","passenger_count","trip_distance","pickup_longitude","pickup_lati
tude","RateCodeID","dropoff_longitude","dropoff_latitude","fare_amount","extra","mta
_tax","tip_amount","tolls_amount","total_amount")]

taxis_numeric


# Find means of all numeric columns

all_means <- sapply(taxis_numeric,mean)
```

```
all_means

Q6

# Define columns

columns <- c("trip_distance", "total_amount", "passenger_count")



# Create summary function

taxis_summary <- function(col, data) {

c(
    mean = mean(data[[col]]),

      sd = sd(data[[col]]),

      quantile(data[[col]], c(.25, .5, .75))
    )
}



# Use sapply to summarize columns


sapply(columns, taxis_summary, taxis)

Q7

# Find the correlation between trip_distance and total_amount

cor(taxis$trip_distance,taxis$total_amount)


# Find the correlation of a different pair of columns

cor(taxis$extra,taxis$total_amount)
```