Lillyvette Pecina

Question 1:

```
install.packages("readxl")
library(readxl)


#1. Load the data from SP500.xlsx into R:
SP500 <- read_excel(file.choose())
```

Question 2:

```
#2. How many rows and columns are in the dataset?
num_rows <- nrow(SP500)
num_columns <- ncol(SP500)
cat("Number of rows:", num_rows, "\n")
cat("Number of columns:", num_columns, "\n")
```

```
> #2. How many rows and columns are in the dataset?
> num_rows <- nrow(SP500)
> num_columns <- ncol(SP500)
> cat("Number of rows:", num_rows, "\n")
Number of rows: 1777
> cat("Number of columns:", num_columns, "\n")
Number of columns: 11
>
```

Question 3:

```
#3. Select the columns SP500, CPI, and Rate:
SP500_subset <- SP500[, c("SP500", "CPI", "Rate")]
```

Question 4:

```
#4. Select the 10th, 100th, 500th, and 1500th rows:
selected_rows <- SP500[c(10, 100, 500, 1500), ]
print(selected_rows)
```

```
> #4. Select the 10th, 100th, 500th, and 1500th rows:
> selected_rows <- SP500[c(10, 100, 500, 1500), ]
> print(selected_rows)
# A tibble: 4 × 11
  Date  SP500 Dividend Earnings   CPI Fraction  Rate RealPrice RealDividend RealEarnings `P/Eratio`
  <dbl>  <dbl>    <dbl>    <dbl> <dbl>    <dbl> <dbl>     <dbl>        <dbl>        <dbl>      <dbl>
1 1871.   4.59    0.26     0.4    12.4    1872.  5.35      93.1         5.27         8.11       11.5
2 1879.   3.77    0.187    0.333   8.18   1879.  4.17     116.          5.72        10.2        11.3
3 1912.   9.81    0.477    0.663   9.70   1913.  4.27     254.         12.3         17.1        14.8
4 1995.  615.     13.8     34.0   154.     1996.  5.71    1004.         22.5         55.5        18.1
```

Question 5:

```r
#5. Select all observations where SP500 is greater than 2000 or CPI is less than 100:
filtered_data_1 <- subset(SP500, SP500 > 2000 | CPI < 100)
print(filtered_data_1)
```

```
> #5. Select all observations where SP500 is greater than 2000 or CPI is less than 100:
> filtered_data_1 <- subset(SP500, SP500 > 2000 | CPI < 100)
> print(filtered_data_1)
# A tibble: 1,399 x 11
    Date SP500 Dividend Earnings   CPI Fraction  Rate RealPrice RealDividend RealEarnings `P/Eratio`
   <db l> <db l>    <db l>   <db l> <db l>    <db l> <db l>     <db l>        <db l>       <db l>
 1 1871.   4.44     0.26      0.4  12.5     1871.  5.32      89.4         5.23         8.05       11.1
 2 1871.   4.5      0.26      0.4  12.8     1871.  5.32      87.9         5.08         7.81       11.2
 3 1871.   4.61     0.26      0.4  13.0     1871.  5.33      88.7         5.00         7.70       11.5
 4 1871.   4.74     0.26      0.4  12.6     1871.  5.33      94.7         5.19         7.99       11.9
 5 1871.   4.86     0.26      0.4  12.3     1871.  5.33      99.3         5.31         8.17       12.2
 6 1871.   4.82     0.26      0.4  12.1     1871.  5.34     100.          5.40         8.30       12.0
 7 1871.   4.73     0.26      0.4  12.1     1872.  5.34      98.2         5.40         8.30       11.8
 8 1871.   4.79     0.26      0.4  11.9     1872.  5.34     101.          5.48         8.44       12.0
 9 1871.   4.84     0.26      0.4  12.2     1872.  5.35      99.7         5.35         8.24       12.1
10 1871.   4.59     0.26      0.4  12.4     1872.  5.35      93.1         5.27         8.11       11.5
# … with 1,389 more rows
```

Question 6:

```r
#6. Select data where Earnings is greater than 50 and Rate is less than 3 with only SP500 and Dividend columns:
filtered_data_2 <- subset(SP500, Earnings > 50 & Rate < 3, select = c("SP500", "Dividend"))
print(filtered_data_2)
```

```
#6. Select data where Earnings is greater than 50 and Rate is less than 3 with only SP500 and Dividend columns:
filtered_data_2 <- subset(SP500, Earnings > 50 & Rate < 3, select = c("SP500", "Dividend"))
print(filtered_data_2)
A tibble: 89 x 2
   SP500 Dividend
   <db l>    <db l>
 1 1087.      22.2
 2 1122.      22.4
 3 1172.      22.5
 4 1199.      22.6
 5 1185.      24.9
 6 1174.      25.2
 7 1207.      25.6
 8 1226.      26.0
 9 1243.      26.4
10 1301.      26.7
… with 79 more rows
```

Question 7:

```r
#7. Remove the entire column Rate:
SP500$Rate <- NULL
```

Question 8:

```r
#8. Add a RealPrice column:
CPI_2018_09 <- tail(SP500$CPI, n=1)
SP500$RealPrice <- (SP500$SP500 * SP500$CPI) / CPI_2018_09
```

Question 9:

```r
#9. Add a RealEarnings column:
SP500$RealEarnings <- (SP500$Earnings * SP500$CPI) / CPI_2018_09
```

Question 10:

```r
#10. Add a PERatio column:
SP500$PERatio <- SP500$RealPrice / SP500$RealEarnings
```

Question 11:

```r
#11. Find the average earnings using a for loop:
total_earnings <- 0
count <- 0

for (i in 1:nrow(SP500)) {
  if (!is.na(SP500$Earnings[i])) {
    total_earnings <- total_earnings + SP500$Earnings[i]
    count <- count + 1
  }
}

if (count > 0) {
  average_earnings <- total_earnings / count
  cat("Average earnings using for loop:", average_earnings, "\n")
} else {
  cat("No valid earnings data found.\n")
}
```

```
+ }
Average earnings using for loop: 13.35659
>
>
```

Question 12:

```r
#12. Find the average earnings using a while loop:
total_earnings <- 0
count <- 0
i <- 1

while (i <= nrow(SP500)) {
  if (!is.na(SP500$Earnings[i])) {
    total_earnings <- total_earnings + SP500$Earnings[i]
    count <- count + 1
  }
  i <- i + 1
}

if (count > 0) {
  average_earnings <- total_earnings / count
  cat("Average earnings using while loop:", average_earnings, "\n")
} else {
  cat("No valid earnings data found.\n")
}
```

```
+ }
Average earnings using while loop: 13.35659
>
>
```

Question 13:

```r
#13. Find the average earnings using a repeat loop:
total_earnings <- 0
count <- 0
i <- 1

repeat {
  if (!is.na(SP500$Earnings[i])) {
    total_earnings <- total_earnings + SP500$Earnings[i]
    count <- count + 1
  }

  i <- i + 1
  if (i > nrow(SP500)) break
}


if (count > 0) {
  average_earnings <- total_earnings / count
  cat("Average earnings using repeat loop:", average_earnings, "\n")
} else {
  cat("No valid earnings data found.\n")
}
```

```
+    cat("No valid earnings data found.\n")
+ }
Average earnings using repeat loop: 13.35659
>
>
```

Question 14:

```r
#14. Find the average earnings for dividends over 25.00 using the for loop:
total_earnings <- 0
count <- 0

for (i in 1:nrow(SP500)) {
  if (!is.na(SP500$Dividend[i]) && SP500$Dividend[i] > 25.00 && !is.na(SP500$Earnings[i])) {
    total_earnings <- total_earnings + SP500$Earnings[i]
    count <- count + 1
  }
}

if (count > 0) {
  average_earnings <- total_earnings / count
  cat("Average earnings for dividends over 25 using for loop:", average_earnings, "\n")
} else {
  cat("No valid earnings for dividends over 25 were found.\n")
}
```

```
+ }
Average earnings for dividends over 25 using for loop: 84.87601
>
>
```

Question 15:

```r
#15. Find the average earnings for dividends over 25.00 using the while loop:
total_earnings <- 0
count <- 0
i <- 1

while (i <= nrow(SP500)) {
  if (!is.na(SP500$Dividend[i]) && SP500$Dividend[i] > 25.00 && !is.na(SP500$Earnings[i])) {
    total_earnings <- total_earnings + SP500$Earnings[i]
    count <- count + 1
  }
  i <- i + 1
}

if (count > 0) {
  average_earnings <- total_earnings / count
  cat("Average earnings for dividends over 25 using while loop:", average_earnings, "\n")
} else {
  cat("No valid earnings for dividends over 25 were found.\n")
}
```

```
+ }
Average earnings for dividends over 25 using while loop: 84.87601
>
```

Question 17:

```r
#16. Find the average earnings for dividends over 25.00 using the repeat loop:
total_earnings <- 0
count <- 0
i <- 1

repeat {
  if (!is.na(SP500$Dividend[i]) && SP500$Dividend[i] > 25.00 && !is.na(SP500$Earnings[i])) {
    total_earnings <- total_earnings + SP500$Earnings[i]
    count <- count + 1
  }

  i <- i + 1
  if (i > nrow(SP500)) break

if (count > 0) {
  average_earnings <- total_earnings / count
  cat("Average earnings for dividends over 25 using repeat loop:", average_earnings, "\n")
} else {
  cat("No valid earnings for dividends over 25 were found.\n")
}
```

```
Average earnings for dividends over 25 using repeat loop: 84.87601
>
```