Case Study

Load in the important libraries.

```r
library(tidyverse)
library(dplyr)
library(tidyr)
library(ggplot2)
library(lubridate)
```

Problem 1 – Data handling, analysis and plotting

The first problem of the case study builds on the data in the files p01-02_portfolio.csv and p01-02_rates.csv.
One file contains membership information for a Group Life portfolio and one has information on the rates
which should be charged.

```r
# Load the CSV file
portfolio_data <- read_delim(
  "Case Study/data/p01-02_portfolio.csv",
  delim = ";",
  show_col_types = FALSE
)

# View the data
head(portfolio_data)
```

```r
# Load the CSV file
rates_data <- read_delim(
  "Case Study/data/p01-02_rates.csv",
  delim = ";",
  show_col_types = FALSE
)

# View the data
head(rates_data)
```

Question a.

Read the data from the two files into R's memory. The rates are applicable to each individual in the portfolio,
depending on that individual's age and gender. Combine the two datasets into a single table by looking up
the rate for each line of the portfolio.

```r
# Step 1: Convert the Date.of.Birth column to Date format
portfolio_data$Date.of.Birth <- dmy(portfolio_data$Date.of.Birth)  # dmy is used for "day-month-year" f
```

```
# Step 2: Calculate the time difference in years
portfolio_data$age <- ceiling(interval(portfolio_data$Date.of.Birth, today()) / years(1))

# Step 3: View the updated data with age column
head(portfolio_data)
```

```
# Inner join the two datasets
joined_data <- inner_join(
  portfolio_data,
  rates_data,
  by = c("age" = "Age", "Gender" = "Gender"),
  relationship = "many-to-many"
)

# Check if the row count of the joined data matches the original portfolio data
if (nrow(joined_data) == nrow(portfolio_data)) {
  message("Sanity Check Passed: The row count of joined_data matches portfolio_data.")
} else {
  message("Sanity Check Failed: The row count of joined_data does not match portfolio_data.")
  message("Rows in portfolio_data: ", nrow(portfolio_data))
  message("Rows in joined_data: ", nrow(joined_data))
}
```

```
## Sanity Check Failed: The row count of joined_data does not match portfolio_data.
```

```
## Rows in portfolio_data: 177922
```

```
## Rows in joined_data: 145616
```

```
# View the result
head(joined_data)
```

Question b.

Group the Industry field into common-sense based groupings and determine the mean, standard deviation and quintiles of DeathSI for each of your industry groups.

2