

Experiment 3

Question 1

```
[8]: library(tidyverse)
```

```
[9]: setwd("/home/asus/content/Notes/Semester 4/FDN Lab/Experiments/Experiment_3")
```

```
[10]: df <- data.frame(
  ID = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
  Name = c("Alice", "Bob", NA, "David", "Emma", "Frank", NA, "Hannah", "Ian", "Jack"),
  Age = c(25, NA, 30, 29, NA, 35, 40, NA, 50, 27),
  Salary = c(50000, 60000, 55000, NA, 70000, 75000, 80000, 65000, NA, 72000),
  Score = c(80, 90, NA, 85, 88, 92, NA, 77, 95, Inf)
)
```

```
[11]: #####
# Exploring Inbuilt Functions
#####
```

```
[12]: is.na(df)
```

	ID	Name	Age	Salary	Score
	FALSE	FALSE	FALSE	FALSE	FALSE
	FALSE	FALSE	TRUE	FALSE	FALSE
	FALSE	TRUE	FALSE	FALSE	TRUE
	FALSE	FALSE	FALSE	TRUE	FALSE
A matrix: 10 × 5 of type lgl	FALSE	FALSE	TRUE	FALSE	FALSE
	FALSE	FALSE	FALSE	FALSE	FALSE
	FALSE	TRUE	FALSE	FALSE	TRUE
	FALSE	FALSE	TRUE	FALSE	FALSE
	FALSE	FALSE	FALSE	TRUE	FALSE
	FALSE	FALSE	FALSE	FALSE	FALSE

```
[13]: complete.cases(df)
```

1. TRUE 2. FALSE 3. FALSE 4. FALSE 5. FALSE 6. TRUE 7. FALSE 8. FALSE
9. FALSE 10. TRUE

```
[14]: df[complete.cases(df), ]
```

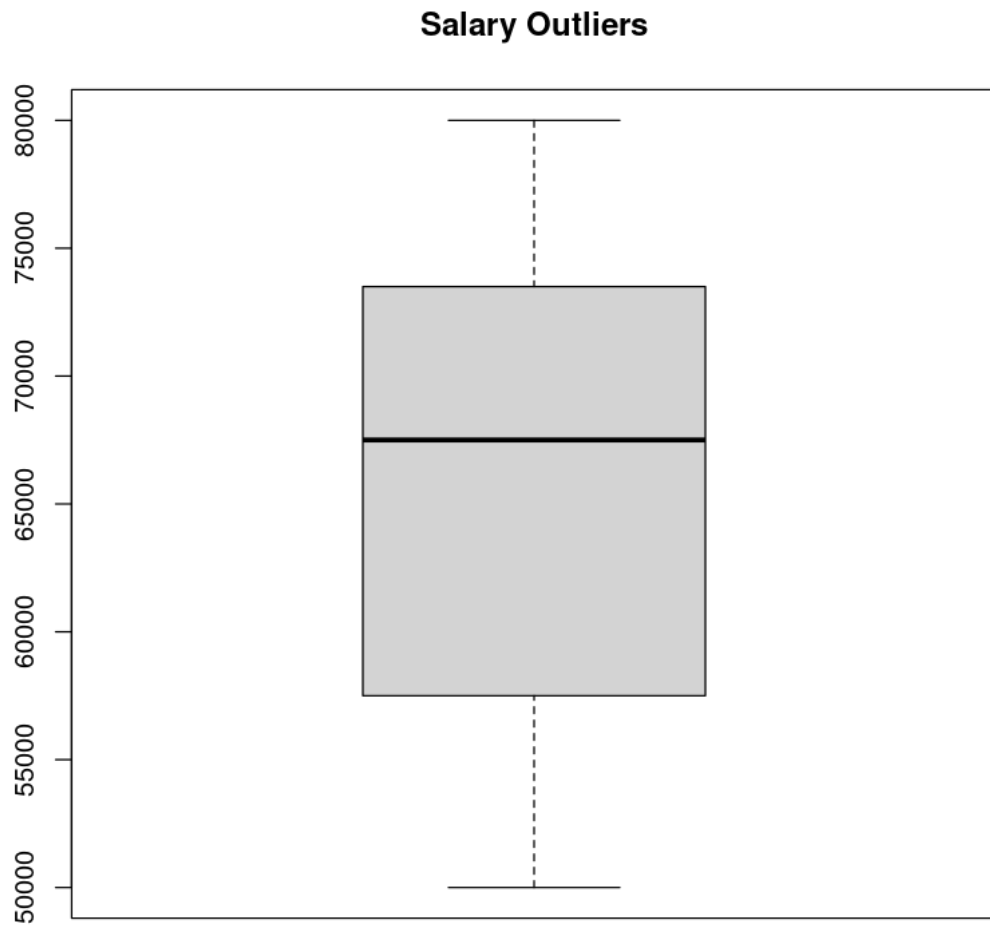
		ID <dbl>	Name <chr>	Age <dbl>	Salary <dbl>	Score <dbl>
A data.frame: 3 × 5	1	1	Alice	25	50000	80
	6	6	Frank	35	75000	92
	10	10	Jack	27	72000	Inf

```
[15]: summary(df)
```

ID		Name	Age		Salary
Min.	: 1.00	Length:10	Min.	:25.00	Min. :50000
1st Qu.	: 3.25	Class :character	1st Qu.	:28.00	1st Qu.:58750
Median	: 5.50	Mode :character	Median	:30.00	Median :67500
Mean	: 5.50		Mean	:33.71	Mean :65875
3rd Qu.	: 7.75		3rd Qu.	:37.50	3rd Qu.:72750
Max.	:10.00		Max.	:50.00	Max. :80000
			NA's	:3	NA's :2

Score	
Min.	:77.00
1st Qu.	:83.75
Median	:89.00
Mean	: Inf
3rd Qu.	:92.75
Max.	: Inf
NA's	:2

```
[16]: # Boxplot to detect outliers
boxplot(df$Salary, main = "Salary Outliers", horizontal = FALSE)
```



```
[17]: # Identify outliers using IQR
Q1 <- quantile(df$Salary, 0.25, na.rm = TRUE)
Q3 <- quantile(df$Salary, 0.75, na.rm = TRUE)
IQR <- Q3 - Q1
lower_bound <- Q1 - 1.5 * IQR
upper_bound <- Q3 + 1.5 * IQR
outliers <- df$Salary[df$Salary < lower_bound | df$Salary > upper_bound]
print(outliers)
```

```
[1] NA NA
```

```
[18]: iqr_value <- IQR(df$Salary, na.rm=TRUE)
print(iqr_value)
```

```
[1] 14000
```

```
[19]: df_clean <- na.omit(df)
      print(df_clean)
```

	ID	Name	Age	Salary	Score
1	1	Alice	25	50000	80
6	6	Frank	35	75000	92
10	10	Jack	27	72000	Inf

```
[20]: df$Age[is.na(df$Age)] <- mean(df$Age, na.rm = TRUE)
      df$Salary[is.na(df$Salary)] <- mean(df$Salary, na.rm = TRUE)
      df$Score[is.na(df$Score)] <- mean(df$Score, na.rm = TRUE)
      print(df)
```

	ID	Name	Age	Salary	Score
1	1	Alice	25.00000	50000	80
2	2	Bob	33.71429	60000	90
3	3	<NA>	30.00000	55000	Inf
4	4	David	29.00000	65875	85
5	5	Emma	33.71429	70000	88
6	6	Frank	35.00000	75000	92
7	7	<NA>	40.00000	80000	Inf
8	8	Hannah	33.71429	65000	77
9	9	Ian	50.00000	65875	95
10	10	Jack	27.00000	72000	Inf