

Group Id: A-199

Name of Student Presenting: Danial Zaf.

Visualization and Analysis –

Tutorial Presentation for Feedback

University of
Hertfordshire **UH**

Date: 26 November 2025

Part 1: VISUALISATION

We are using the dataset **DS134 - london_weather.csv** to answer our Research Question “Is there a difference in the mean of daily precipitation among the seasons in London from 1979 to 2020?”

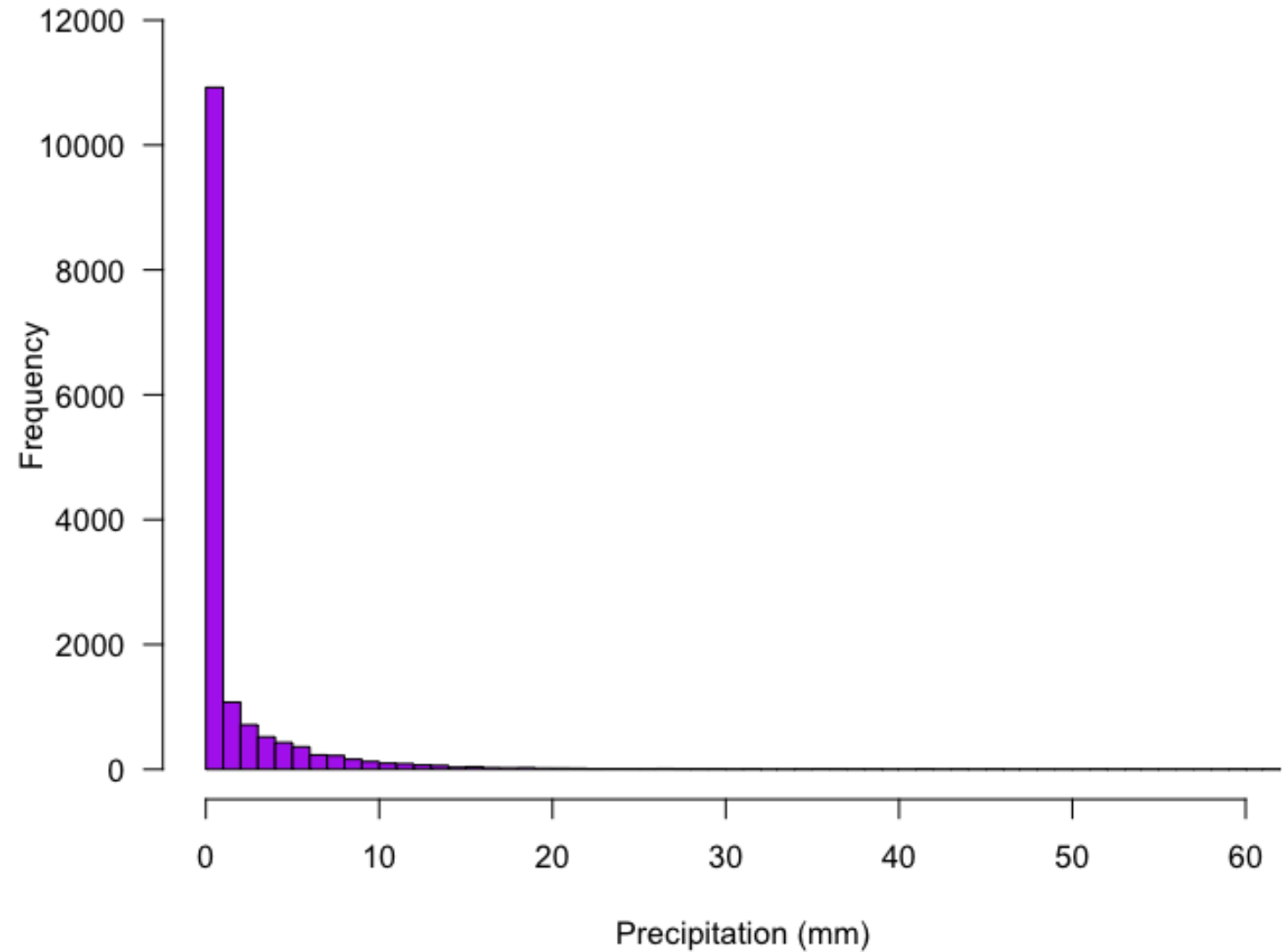
	date <dbl>	cloud_cover <dbl>	sunshine <dbl>	global_radiation <dbl>	max_temp <dbl>	mean_temp <dbl>	min_temp <dbl>	precipitation <dbl>	pressure <dbl>	snow_depth <dbl>
1	19790101	2	7	52	2.3	-4.1	-7.5	0.4	101900	9
2	19790102	6	1.7	27	1.6	-2.6	-7.5	0	102530	8
3	19790103	5	0	13	1.3	-2.8	-7.2	0	102050	4
4	19790104	8	0	13	-0.3	-2.6	-6.5	0	100840	2
5	19790105	6	2	29	5.6	-0.8	-1.4	0	102250	1

The database has 10 columns and 15,341 Rows. We will use precipitation(mm) and date columns.

Independent Variable: date ==> Season (Nominal) | Dependent Variable: precipitation (Interval)

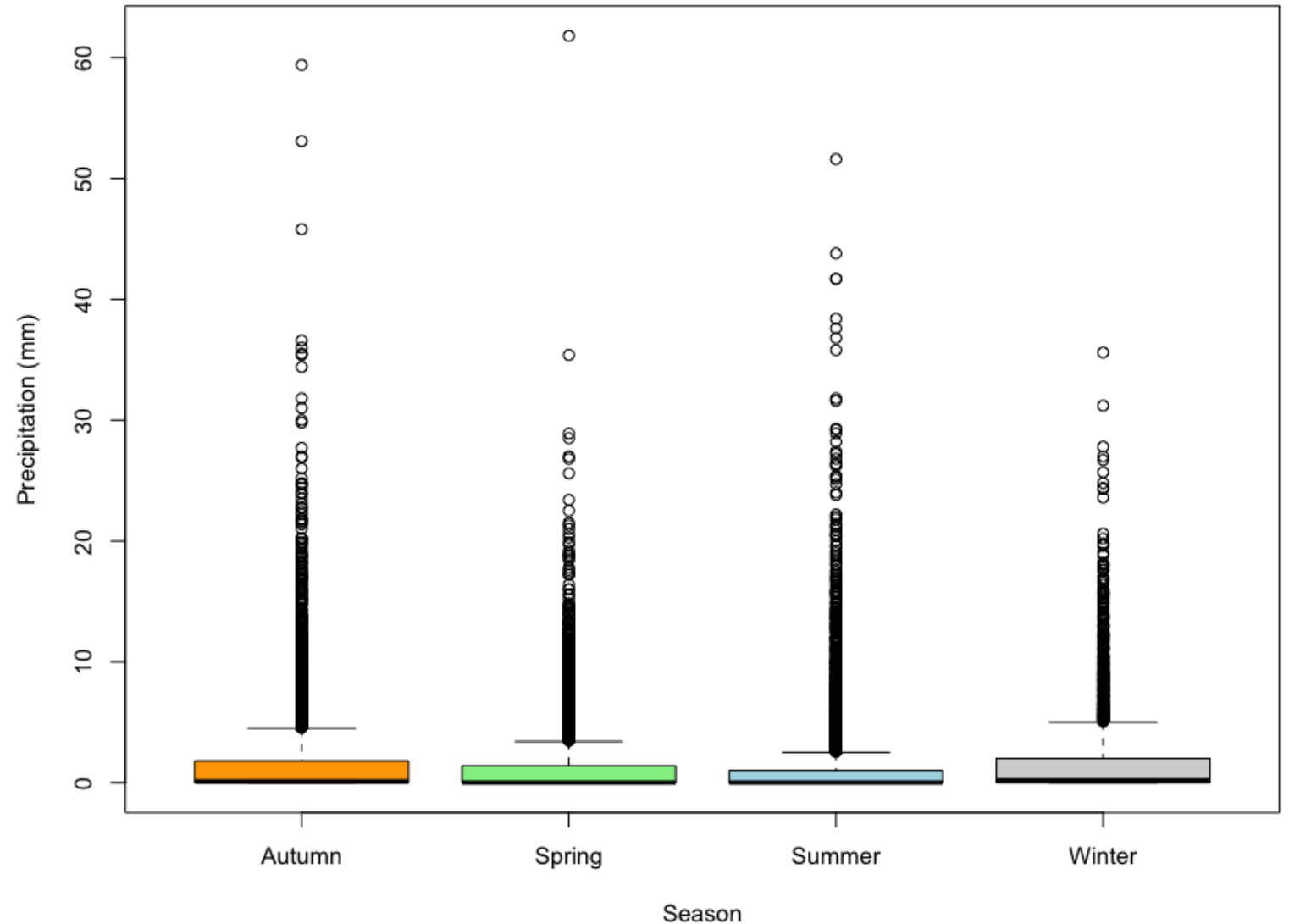
Histogram of Daily Precipitation

Variable `precipitation`(mm) is **not normally** distributed.



\$Autumn						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.000	0.000	0.100	1.901	1.800	59.400	
\$Spring						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.000	0.000	0.000	1.462	1.400	61.800	
\$Summer						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.000	0.000	0.000	1.573	1.000	51.600	
\$Winter						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.000	0.000	0.200	1.742	2.000	35.600	

Daily Precipitation in London by Season from 1979 to 2020



Part 2: Analysis (building on your Visualizations)

~~Null hypothesis (H_0): There is **no difference** in the mean of daily precipitation among the seasons in London.~~

Alternative hypothesis (H_1): There is **a difference** in the mean of daily precipitation among the seasons in London.

```
> pairwise.wilcox.test(df$precipitation, df$season, p.adjust.method = "holm")
```

```
Pairwise comparisons using Wilcoxon rank sum test with continuity correction
```

```
data: df$precipitation and df$season
```

```
      Autumn Spring Summer
Spring 8.0e-07 -         -
Summer 2.8e-16 0.0021  -
Winter 0.0021 2.8e-16 < 2e-16
```

```
P value adjustment method: holm
```

This means that precipitation patterns in London from 1979 to 2020 are specific to each of the four seasons.

Winter	1			
Spring	8.0e-07 Very significant difference	1		
Summer	2.8e-16 Very significant difference	0.0021 significant difference	1	
Autumn	0.0021 significant difference	2.8e-16 Very significant difference	< 2e-16 Very significant difference	1
	Winter	Spring	Summer	Autumn

```

1  # R codes used for analysis and visualization
2
3  library(readr)
4  df <- read_csv("london_weather.csv")
5
6  # create year and month column
7  df$year <- as.numeric(substr(as.character(df$date), 1, 4))
8  df$month <- as.numeric(substr(as.character(df$date), 5, 6))
9
10 df$season <- "Autumn"
11 df$season[df$month %in% c(12, 1, 2)] <- "Winter"
12 df$season[df$month %in% c(3, 4, 5)] <- "Spring"
13 df$season[df$month %in% c(6, 7, 8)] <- "Summer"
14
15 df$season_year <- paste(df$season, df$year)
16
17 # Clear NA from the data :
18
19 sum(is.na(df$date))
20 sum(is.na(df$precipitation)) #We have 6 NA
21
22 df <- df[!is.na(df$date) & !is.na(df$precipitation), ]
23 sum(is.na(df$precipitation))
24
25 View(df)
26 |

```

```

66 hist(df$precipitation,
67       main = "Histogram of Daily Precipitation",
68       xlab = "Precipitation (mm)",
69       ylab = "Frequency",
70       col = "darkorchid2",
71       breaks = 50,
72       ylim = c(0, 12000),
73       las = 1,
74 )
75
76 boxplot(precipitation ~ season, data = df,
77         main = "Daily Precipitation in London by Season from 1979 to 2020",
78         xlab = "Season",
79         ylab = "Precipitation (mm)",
80         col = c("orange", "lightgreen", "lightblue", "lightgray"),
81         border = "black")
82
83 tapply(df$precipitation, df$season, summary)
84
85 # Pairwise Wilcoxon Test
86
87 pairwise.wilcox.test(df$precipitation, df$season, p.adjust.method = "holm")

```