# STAT40840 - Assignment 1

Gargi - 23200711

*I have read and understood the Honesty Code and have neither received nor given assistance in any way with the work contained in this submission.*

# Question 1: How many variables are there? In what format is the month variable stored?

**The CONTENTS Procedure**

| | | | |
|---|---|---|---|
| **Data Set Name** | S40840.WEATHER | **Observations** | 3652 |
| **Member Type** | DATA | **Variables** | 27 |
| **Engine** | V9 | **Indexes** | 0 |
| **Created** | 23/06/2024 20:44:12 | **Observation Length** | 208 |
| **Last Modified** | 23/06/2024 20:44:12 | **Deleted Observations** | 0 |
| **Protection** | | **Compressed** | NO |
| **Data Set Type** | | **Sorted** | NO |
| **Label** | | | |
| **Data Representation** | SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64 | | |
| **Encoding** | utf-8 Unicode (UTF-8) | | |

| **Engine/Host Dependent Information** | |
|---|---|
| **Data Set Page Size** | 131072 |
| **Number of Data Set Pages** | 6 |
| **First Data Page** | 1 |
| **Max Obs per Page** | 629 |
| **Obs in First Data Page** | 604 |
| **Number of Data Set Repairs** | 0 |
| **Filename** | /home/u63919326/sasuser.v94/s40840/weather.sas7bdat |
| **Release Created** | 9.0401M7 |
| **Host Created** | Linux |
| **Inode Number** | 6961754177 |
| **Access Permission** | rw-r--r-- |
| **Owner Name** | u63919326 |
| **File Size** | 896KB |
| **File Size (bytes)** | 917504 |

## Question 1: How many variables are there? In what format is the month variable stored?

**The CONTENTS Procedure**

| Variables in Creation Order | | | | | |
|---|---|---|---|---|---|
| **#** | **Variable** | **Type** | **Len** | **Format** | **Informat** |
| **1** | date | Num | 8 | BEST12. | BEST32. |
| **2** | month | Char | 5 | $5. | $5. |
| **3** | year | Num | 8 | BEST12. | BEST32. |
| **4** | ind | Num | 8 | BEST12. | BEST32. |
| **5** | maxtp | Num | 8 | BEST12. | BEST32. |
| **6** | ind.1 | Num | 8 | BEST12. | BEST32. |
| **7** | mintp | Num | 8 | BEST12. | BEST32. |
| **8** | igmin | Num | 8 | BEST12. | BEST32. |
| **9** | gmin | Num | 8 | BEST12. | BEST32. |
| **10** | ind.2 | Num | 8 | BEST12. | BEST32. |
| **11** | rain | Num | 8 | BEST12. | BEST32. |
| **12** | cbl | Num | 8 | BEST12. | BEST32. |
| **13** | wdsp | Num | 8 | BEST12. | BEST32. |
| **14** | ind.3 | Num | 8 | BEST12. | BEST32. |
| **15** | hm | Num | 8 | BEST12. | BEST32. |
| **16** | ind.4 | Num | 8 | BEST12. | BEST32. |
| **17** | ddhm | Num | 8 | BEST12. | BEST32. |
| **18** | ind.5 | Num | 8 | BEST12. | BEST32. |
| **19** | hg | Num | 8 | BEST12. | BEST32. |
| **20** | sun | Char | 2 | $2. | $2. |
| **21** | glorad | Char | 2 | $2. | $2. |
| **22** | soil | Num | 8 | BEST12. | BEST32. |
| **23** | pe | Num | 8 | BEST12. | BEST32. |
| **24** | evap | Num | 8 | BEST12. | BEST32. |
| **25** | smd_wd | Num | 8 | BEST12. | BEST32. |
| **26** | smd_md | Num | 8 | BEST12. | BEST32. |
| **27** | smd_pd | Num | 8 | BEST12. | BEST32. |

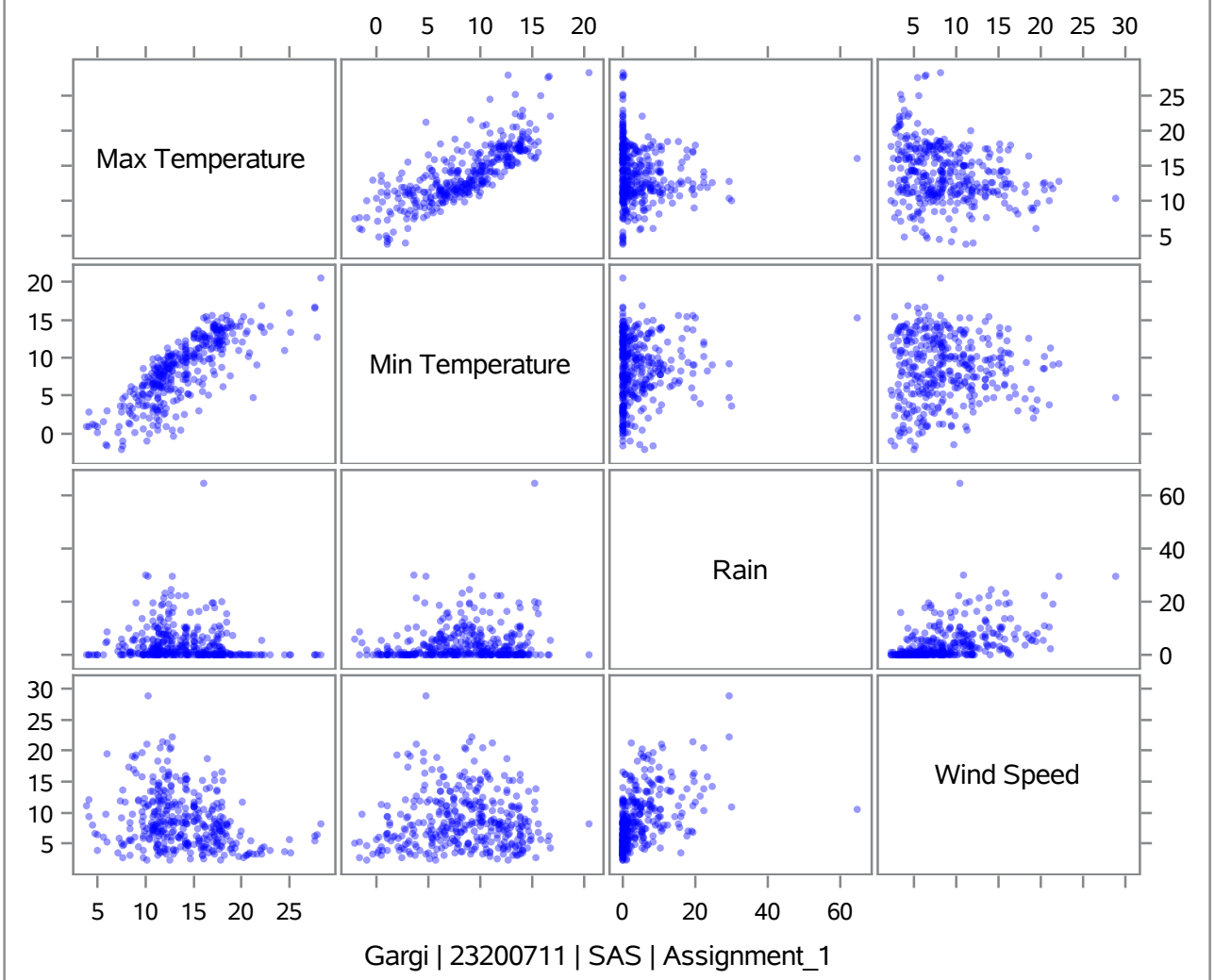
Answer: As we can see, there are 27 variables and the month variable is stored in the character format

## Question 2: Print first 15 rows. What is the value of rain in the 10th row?

| Obs | date | month | year | maxtp | mintp | rain |
|---|---|---|---|---|---|---|
| **1** | 1 | jan | 2012 | 10.9 | 4.1 | 5.7 |
| **2** | 2 | jan | 2012 | 12.1 | 3.3 | 19.1 |
| **3** | 3 | jan | 2012 | 12.8 | 4.8 | 5 |
| **4** | 4 | jan | 2012 | 10.9 | 7.6 | 3.6 |
| **5** | 5 | jan | 2012 | 11.2 | 6.9 | 2.5 |
| **6** | 6 | jan | 2012 | 11.5 | 8.2 | 3.4 |
| **7** | 7 | jan | 2012 | 10.9 | 9 | 0.2 |
| **8** | 8 | jan | 2012 | 12.2 | 10.2 | 0.2 |
| **9** | 9 | jan | 2012 | 11.3 | 8.4 | 3.8 |
| **10** | 10 | jan | 2012 | 11.4 | 9.8 | 6.1 |
| **11** | 11 | jan | 2012 | 11.2 | 10 | 3.9 |
| **12** | 12 | jan | 2012 | 10.9 | 9.6 | 2.6 |
| **13** | 13 | jan | 2012 | 10.2 | 7.2 | 0.6 |
| **14** | 14 | jan | 2012 | 10.5 | 8.2 | 0.3 |
| **15** | 15 | jan | 2012 | 9.7 | 6.8 | 14.3 |

Answer: The value of rain in the 10th row is 6.1

## Question 3: Comment on the relationships between the variables: maximum temperature, minimum temperature, rain, and wind speed




Overall, the most significant relationship observed is between maximum temperature (Maxtp) and minimum temperature (Mintp), which shows a positive correlation. The other pairs of variables do not exhibit clear correlations based on the scatter plots.

## Question 4: Rename and Print the first 10 rows

| Obs | date | month | year | max_temp | min_temp | wind_speed |
|---|---|---|---|---|---|---|
| **1** | 1 | jan | 2012 | 10.9 | 4.1 | 17.2 |
| **2** | 2 | jan | 2012 | 12.1 | 3.3 | 18.6 |
| **3** | 3 | jan | 2012 | 12.8 | 4.8 | 25.7 |
| **4** | 4 | jan | 2012 | 10.9 | 7.6 | 22.1 |
| **5** | 5 | jan | 2012 | 11.2 | 6.9 | 15.8 |
| **6** | 6 | jan | 2012 | 11.5 | 8.2 | 10.6 |
| **7** | 7 | jan | 2012 | 10.9 | 9 | 8.9 |
| **8** | 8 | jan | 2012 | 12.2 | 10.2 | 9.1 |
| **9** | 9 | jan | 2012 | 11.3 | 8.4 | 6.9 |
| **10** | 10 | jan | 2012 | 11.4 | 9.8 | 12.8 |

Success!

## Question 5: Append new variable: wind_speed1 = winds peed divided by 2 and round the record to 1 decimal

| Obs | year | wind_speed | wind_speed1 |
|---|---|---|---|
| 1 | 2012 | 17.2 | 8.6 |
| 2 | 2012 | 18.6 | 9.3 |
| 3 | 2012 | 25.7 | 12.9 |
| 4 | 2012 | 22.1 | 11.1 |
| 5 | 2012 | 15.8 | 7.9 |
| 6 | 2012 | 10.6 | 5.3 |
| 7 | 2012 | 8.9 | 4.5 |
| 8 | 2012 | 9.1 | 4.6 |
| 9 | 2012 | 6.9 | 3.5 |
| 10 | 2012 | 12.8 | 6.4 |

Done!

## Question 6: Print a table containing only those observations where the maximum temperature was greater than 25 C. On how many of these days was rain recorded?

| Obs | date | month | year | max_temp | min_temp | rain |
|---|---|---|---|---|---|---|
| **145** | 25 | may | 2012 | 25.9 | 7.3 | 0.1 |
| **222** | 10 | aug | 2012 | 25.5 | 13.1 | 0 |
| **555** | 9 | jul | 2013 | 27.6 | 14.4 | 0 |
| **565** | 19 | jul | 2013 | 26 | 13.4 | 0 |
| **566** | 20 | jul | 2013 | 27.5 | 16.2 | 0 |
| **1661** | 19 | jul | 2016 | 25.2 | 13.6 | 0 |
| **1997** | 20 | jun | 2017 | 26 | 15.2 | 0.7 |
| **2368** | 26 | jun | 2018 | 26.6 | 12.8 | 0 |
| **2369** | 27 | jun | 2018 | 28.2 | 13.2 | 0 |
| **2370** | 28 | jun | 2018 | 28.4 | 10.9 | 0 |
| **2371** | 29 | jun | 2018 | 25.4 | 13.5 | 0 |
| **2734** | 27 | jun | 2019 | 28 | 17.7 | 0 |
| **3485** | 17 | jul | 2021 | 25.1 | 13.4 | 0 |
| **3489** | 21 | jul | 2021 | 27.6 | 16.5 | 0 |
| **3490** | 22 | jul | 2021 | 28.3 | 20.5 | 0 |
| **3491** | 23 | jul | 2021 | 27.7 | 16.6 | 0 |
| **3537** | 7 | sep | 2021 | 27.9 | 12.7 | 0.2 |

## Question 6: Print a table containing only those observations where the maximum temperature was greater than 25 C. On how many of these days was rain recorded?

| Obs | date | month | year | max_temp | min_temp | rain |
|---|---|---|---|---|---|---|
| **145** | 25 | may | 2012 | 25.9 | 7.3 | 0.1 |
| **1997** | 20 | jun | 2017 | 26 | 15.2 | 0.7 |
| **3537** | 7 | sep | 2021 | 27.9 | 12.7 | 0.2 |

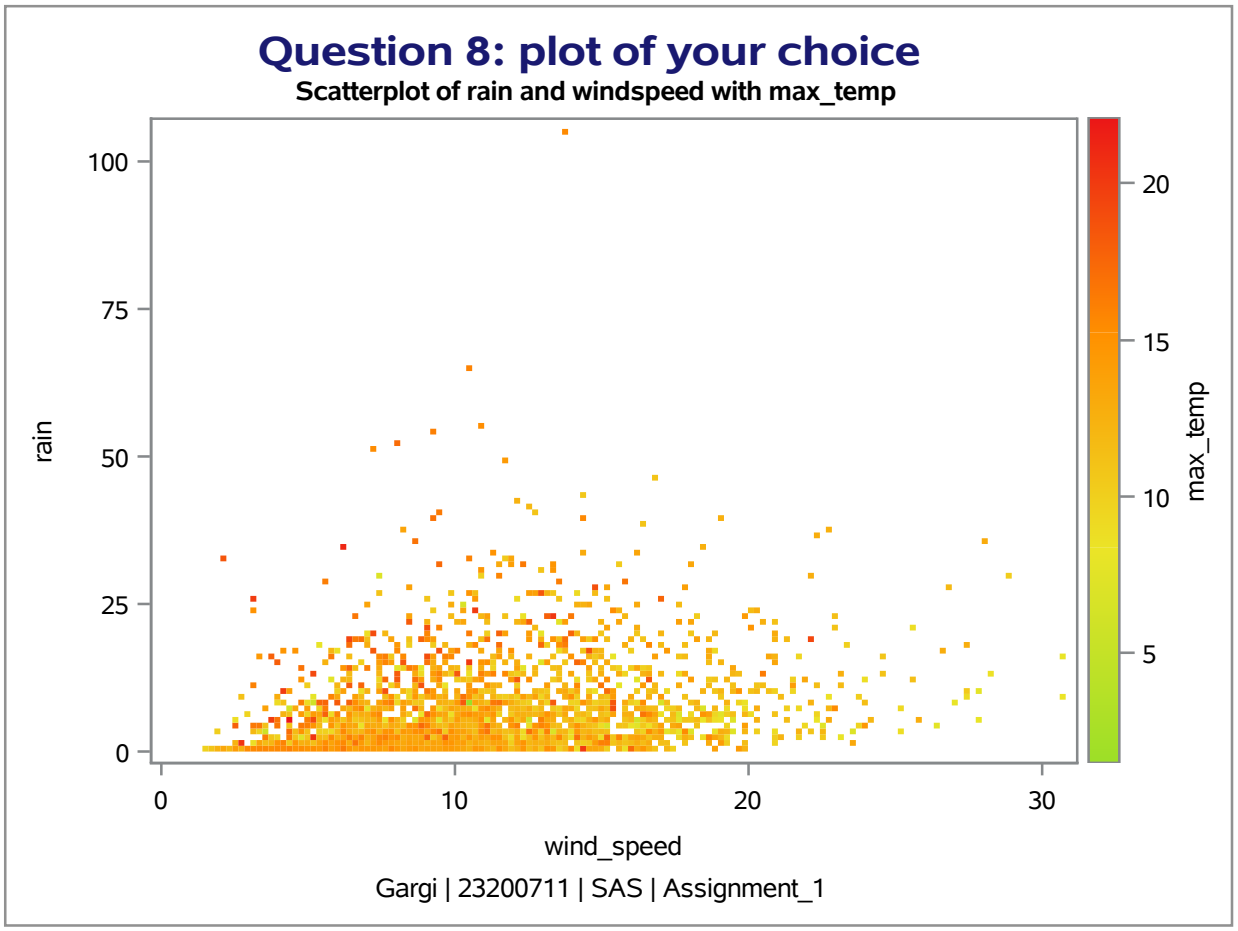
Answer: As we can see, there are 3 days where rain was recorded!

## Question 7: What was the mean rain recorded? What was the maximum of the minimum temperature?

**The MEANS Procedure**

| Variable | Mean | Std Dev | Maximum | Median |
|---|---|---|---|---|
| max_temp | 13.67 | 3.69 | 28.40 | 13.30 |
| min_temp | 8.38 | 3.94 | 20.50 | 8.50 |
| rain | 4.58 | 6.91 | 105.50 | 1.90 |

Answer: The mean rain recorded is 4.58 and the maximum of the minimum temperature is 20.50




The scatter plot indicates a low moderate positive correlation among wind speed, rain, and the heatmap/gradient colors of the points indicate the average maximum temperature. Regions experiencing higher wind speeds exhibit greater rainfall. The decrease (as the color gradient seems to cool) in average maximum temperature seems to show a more noticeable trend with increasing wind speed compared to increasing rainfall.