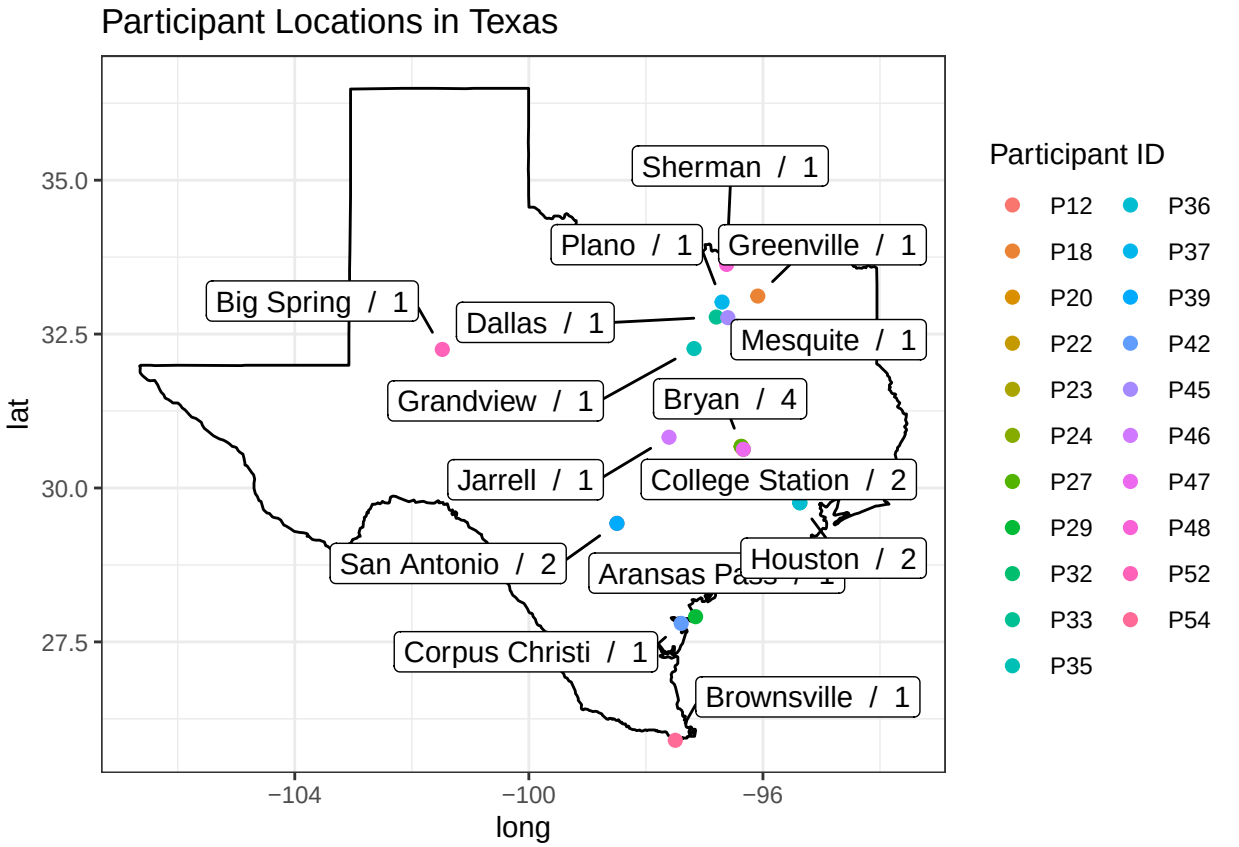
# PROJECT MILESTONE-1

Sai Sri Charishma vaddi- 2157099, Sivani Pamulapati - 2183785

2023-03-10

## Question1

This is a state-wide naturalistic driving study. To show how participants were distributed within the state of Texas, plot a map visualization. The latiitude and longitude for the city of each participant are provided in the file Affective_Driving_Dataset_Pariticpant_Cities.csv.




**Work flow and Observations**

By reading the particpiant data file in a data frame called participant_data, we grouped the data by city, latitude, and longitude, and count the number of unique participant IDs in each group. The map also higlights the count of number of drivers belong to the same city.

The result is stored in a data frame called driver_counts. The latitude and longitude for the city of each participant are taken on y-axis and x-axis respectively.

The map_data function is used to load map data for the state of Texas into a data frame called tx_map.

Overall, the plot shows the locations of participants in the state of Texas and provides a visual representation of the distribution of participants across different cities in the state.

The highest number of participants visited were in Bryan City with 4 trips.

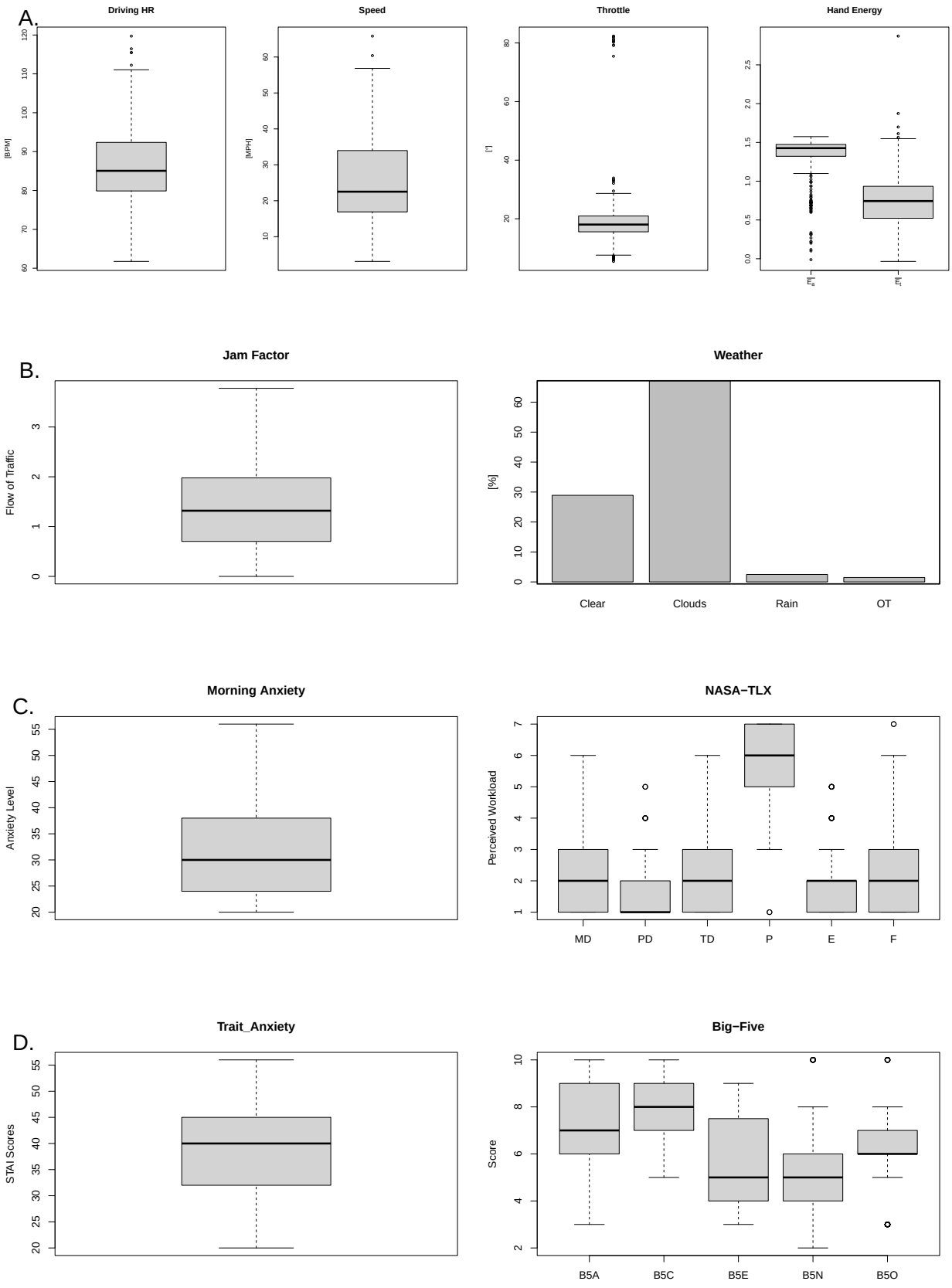
Below is the count of participants spread across different cities in texas state Bryan: 4 participants College Station: 2 participants San Antonio: 1 participant Houston: 1 participant Dallas: 1 participant Plano: 1 participant Big Spring: 1 participant Sherman: 1 participant Grandview: 1 participant Jarrell: 1 participant Corpus Christi: 1 participant Brownsville: 1 participant Aransas Pass: 1 participant Greenville: 1 participant

## Question 2.

Perform exploratory analysis by constructing summary visualization of the data. We suggest the following: Plot a figure with four rows.

In the first row, plot the trip distributions of the physiological and behavioral variables. Physio - HR_Raw Behavioral - Speed, ATP, Accel_Energy, Rot_Energy In the second row, plot the trip distributions of the environmental data. environmental- JF, Weather In the third row, plot the distributions of the state psychometrics. State PSychometrics - State_Anxiety,NASA-TLX In the fourth row, plot the distributions of the trait psychometrics. trait psychometrics - Trait_Anxiety, Big-Five

Construct also a table with the descriptive statistics of these distributions. State your observations and thoughts.


A.
Driving HR
[BPM]
Speed
[MPH]
Throttle
[°]
Hand Energy
[°]
B.
Jam Factor
Flow of Traffic
Weather
[%]
Clear
Clouds
Rain
OT
C.
Morning Anxiety
Anxiety Level
NASA−TLX
Perceived Workload
MD
PD
TD
P
E
F
D.
Trait_Anxiety
STAI Scores
Big−Five
Score
B5A
B5C
B5E
B5N
B5O

| | Trip1 | Trip2 | Trip3 | Trip4 | Trip5 | Trip6 | Trip7 |
|---|---|---|---|---|---|---|---|
| HR_Raw.mean | 84.767 | 85.388 | 91.000 | 88.235 | 90.149 | 84.352 | 81.958 |
| HR_Raw.sd | 11.403 | 11.390 | 11.547 | 11.051 | 8.267 | 2.674 | 3.664 |
| HR_Raw.med | 83 | 85 | 89 | 88 | 88 | 84 | 82 |
| Speed.mean | 34.651 | 29.279 | 27.105 | 22.336 | 25.514 | 17.659 | 16.189 |
| Speed.sd | 26.256 | 24.618 | 23.359 | 20.435 | 20.314 | 11.431 | 14.060 |
| Speed.med | 34.333 | 28.000 | 25.667 | 19.333 | 25.000 | 19.667 | 15.000 |
| ATP.mean | 21.978 | 22.622 | 21.338 | 16.306 | 12.899 | 19.323 | 19.130 |
| ATP.sd | 15.482 | 14.983 | 19.069 | 6.974 | 8.244 | 3.602 | 3.714 |
| ATP.med | 18.40 | 18.25 | 15.70 | 15.80 | 12.50 | 18.85 | 17.70 |
| Accel.mean | 1.232 | 1.203 | 1.346 | 1.128 | 1.425 | 1.415 | 1.430 |
| Accel.sd | 0.465 | 0.542 | 0.349 | 0.490 | 0.129 | 0.102 | 0.089 |
| Accel.med | 1.379 | 1.395 | 1.432 | 1.300 | 1.428 | 1.419 | 1.435 |
| Rot.mean | 0.633 | 0.676 | 0.793 | 0.639 | 0.770 | 0.824 | 0.433 |
| Rot.sd | 1.140 | 1.240 | 1.298 | 1.354 | 1.169 | 1.200 | 0.801 |
| Rot.med | 0.302 | 0.309 | 0.352 | 0.241 | 0.319 | 0.314 | 0.218 |

**Q2.Row1 - Observations** We can see that the heart rate's standard deviation (HR Raw.sd) is higher for the first three excursions (Trip1-Trip3) than it is for the next four (Trip4-Trip7), indicating that the heart rate was more variable earlier in the journeys. A declining tendency is also visible in the mean speed (Speed.mean) from the first to the last excursion. According to the ATP.mean, the total power drops from the first to the fourth trip before increasing over the final three visits. Similar to this, there is some variability among visits but no discernible pattern in Accel E.mean . We can see from the HR Raw variable that for Trips 1, 2, and 3, the standard deviation (sd) is relatively large, indicating that there was variability in heart rate during those travels. In a similar manner, when it comes to Speed, Trip1's mean is highest, indicating a higher driving speed, but Trip2's sd is greatest, indicating more variance in driving speed during that trip. With Trip1 having the greatest mean and Trip5 having the lowest, the ATP variable exhibits a declining trend in mean as the trips go on. The Accel E and Rot E variables have medians that are near to the means and relatively low standard deviations, which show that the data points are more evenly distributed and less likely to be outliers. Overall, the information provided by these descriptive statistics on the distribution and variability of variables across all the trips.

| | Trip1 | Trip2 | Trip3 | Trip4 | Trip5 | Trip6 | Trip7 |
|---|---|---|---|---|---|---|---|
| JF.mean | 1.134 | 1.354 | 1.469 | 1.431 | 1.872 | 1.637 | 1.452 |
| JF.sd | 1.495 | 1.387 | 1.349 | 1.376 | 1.441 | 0.383 | 0.562 |
| JF.med | 0.717 | 1.021 | 1.278 | 1.224 | 1.686 | 1.457 | 1.324 |
| weather.mean | 1.721 | 1.806 | 1.766 | 1.895 | 1.567 | 2.000 | 1.000 |
| weather.sd | 0.660 | 0.451 | 0.495 | 0.426 | 0.496 | 0.000 | 0.000 |
| weather.med | 2 | 2 | 2 | 2 | 2 | 2 | 1 |

**Q2.Row2 - Observations**

JF: When comparing Trips 1 through 3, JF.mean rises, while when comparing Trips 3 through 7, it falls.Except from a drop in Trip6, JF.sd is mostly steady between trips.JF.med typically rises from Trip 1 to Trip 5, then slightly falls in Trips 6 and 7.

Weather: Throughout visits, Weather.mean varies a little bit but is generally stable. With the exception of Trips 6 and 7, where it declines, weather.sd is largely stable between journeys. Only a few instances of 1 may be found across weather.med's data.

Overall, it appears that the variables JF and Weather differ a little from trip to trip but are generally constant, although the descriptive statistics for some variables, such as Weather.sd and Weather.med, indicate more differences between trips. More research into these factors and any effects they may have overall analysis should be done.

| | Trip1 | Trip2 | Trip3 | Trip4 | Trip5 | Trip6 | Trip7 |
|---|---|---|---|---|---|---|---|
| SA.mean | 30.349 | 30.871 | 32.588 | 33.806 | 32.484 | 24.000 | 24.000 |
| SA.sd | 9.017 | 9.660 | 9.387 | 11.054 | 12.742 | 0.000 | 0.000 |
| SA.med | 29 | 30 | 33 | 31 | 24 | 24 | 24 |
| MD.mean | 2.099 | 2.060 | 2.042 | 2.273 | 1.622 | 1.000 | 1.000 |
| MD.sd | 1.215 | 1.115 | 1.154 | 1.224 | 0.734 | 0.000 | 0.000 |
| MD.med | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| PD.mean | 1.923 | 1.765 | 1.661 | 1.929 | 1.470 | 1.000 | 1.000 |
| PD.sd | 1.153 | 0.770 | 0.746 | 0.869 | 0.499 | 0.000 | 0.000 |
| PD.med | 2 | 2 | 1 | 2 | 1 | 1 | 1 |
| TD.mean | 2.114 | 1.812 | 2.169 | 2.391 | 2.080 | 1.000 | 1.000 |
| TD.sd | 1.366 | 1.095 | 1.427 | 1.104 | 1.318 | 0.000 | 0.000 |
| TD.med | 2 | 1 | 2 | 3 | 2 | 1 | 1 |
| P.mean | 5.631 | 5.570 | 5.554 | 5.361 | 5.935 | 4.000 | 4.000 |
| P.sd | 1.129 | 1.311 | 1.211 | 1.350 | 1.193 | 0.000 | 0.000 |
| P.med | 6 | 6 | 6 | 6 | 6 | 4 | 4 |
| E.mean | 1.914 | 2.160 | 1.836 | 2.197 | 1.152 | 1.000 | 1.000 |
| E.sd | 0.983 | 1.161 | 0.956 | 0.934 | 0.359 | 0.000 | 0.000 |
| E.med | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| F.mean | 1.826 | 2.006 | 2.630 | 2.698 | 1.154 | 1.000 | 1.000 |
| F.sd | 1.198 | 1.172 | 1.817 | 1.495 | 0.361 | 0.000 | 0.000 |
| F.med | 1 | 2 | 2 | 2 | 1 | 1 | 1 |

**Q2.Row3 - Observations**

There appears to be fluctuation in the mean and standard deviation for the trip-related variables (Trip1, Trip2,. . . , Trip7), indicating that the experiences on each trip may differ from one another.

The mean and median values for State A appear to remain fairly constant across the many journeys, but the standard deviation fluctuates widely. The mean and standard deviation for MD, PD, TD, E, and F varied between the several trips, indicating that the experiences connected to these variables may vary depending on the trip.

There appears to be variance in the mean and standard deviation for P (which is likely an acronym for "price") among the many excursions, indicating that the price of each trip may be different

| | Trip1 | Trip2 | Trip3 | Trip4 | Trip5 | Trip6 | Trip7 |
|---|---|---|---|---|---|---|---|
| B5A.mean | 7.112 | 7.587 | 7.630 | 6.554 | 8.203 | 6.000 | 6.000 |
| B5A.sd | 2.120 | 1.689 | 1.484 | 2.190 | 1.594 | 0.000 | 0.000 |
| B5A.med | 7 | 8 | 7 | 7 | 7 | 6 | 6 |
| B5C.mean | 7.472 | 7.546 | 8.106 | 7.961 | 8.746 | 5.000 | 5.000 |
| B5C.sd | 1.342 | 1.339 | 1.377 | 1.155 | 1.581 | 0.000 | 0.000 |
| B5C.med | 7 | 7 | 8 | 8 | 9 | 5 | 5 |
| B5E.mean | 5.693 | 5.841 | 5.661 | 5.316 | 5.006 | 4.000 | 4.000 |
| B5E.sd | 2.100 | 2.180 | 2.227 | 2.001 | 2.746 | 0.000 | 0.000 |
| B5E.med | 6 | 6 | 6 | 5 | 3 | 4 | 4 |
| B5N.mean | 4.559 | 4.706 | 5.148 | 6.012 | 4.527 | 5.000 | 5.000 |
| B5N.sd | 1.673 | 1.798 | 1.835 | 1.819 | 0.499 | 0.000 | 0.000 |
| B5N.med | 4 | 4 | 5 | 5 | 5 | 5 | 5 |
| B5O.mean | 6.965 | 6.495 | 5.867 | 5.571 | 6.413 | 7.000 | 7.000 |
| B5O.sd | 1.890 | 1.703 | 1.618 | 1.595 | 0.492 | 0.000 | 0.000 |
| B5O.med | 7 | 6 | 6 | 6 | 6 | 7 | 7 |

**Q2.Row4 - Observations**

Trait_A.mean seems to vary across trips, with the highest mean in Trip4 and the lowest in Trip5. The standard deviation for this variable also varies across trips, with the highest in Trip3 and the lowest in Trip5.

B5E.mean seems to be relatively consistent across trips, with a mean around 5.5-6 and a standard deviation around 2.

B5A.mean also seems to be relatively consistent across trips, with a mean around 6.5-7.5 and a standard deviation around 1.5-2.

B5C.mean shows some variation across trips, with the highest mean in Trip5 and the lowest in Trip7. The standard deviation is relatively consistent around 1.3-1.6.

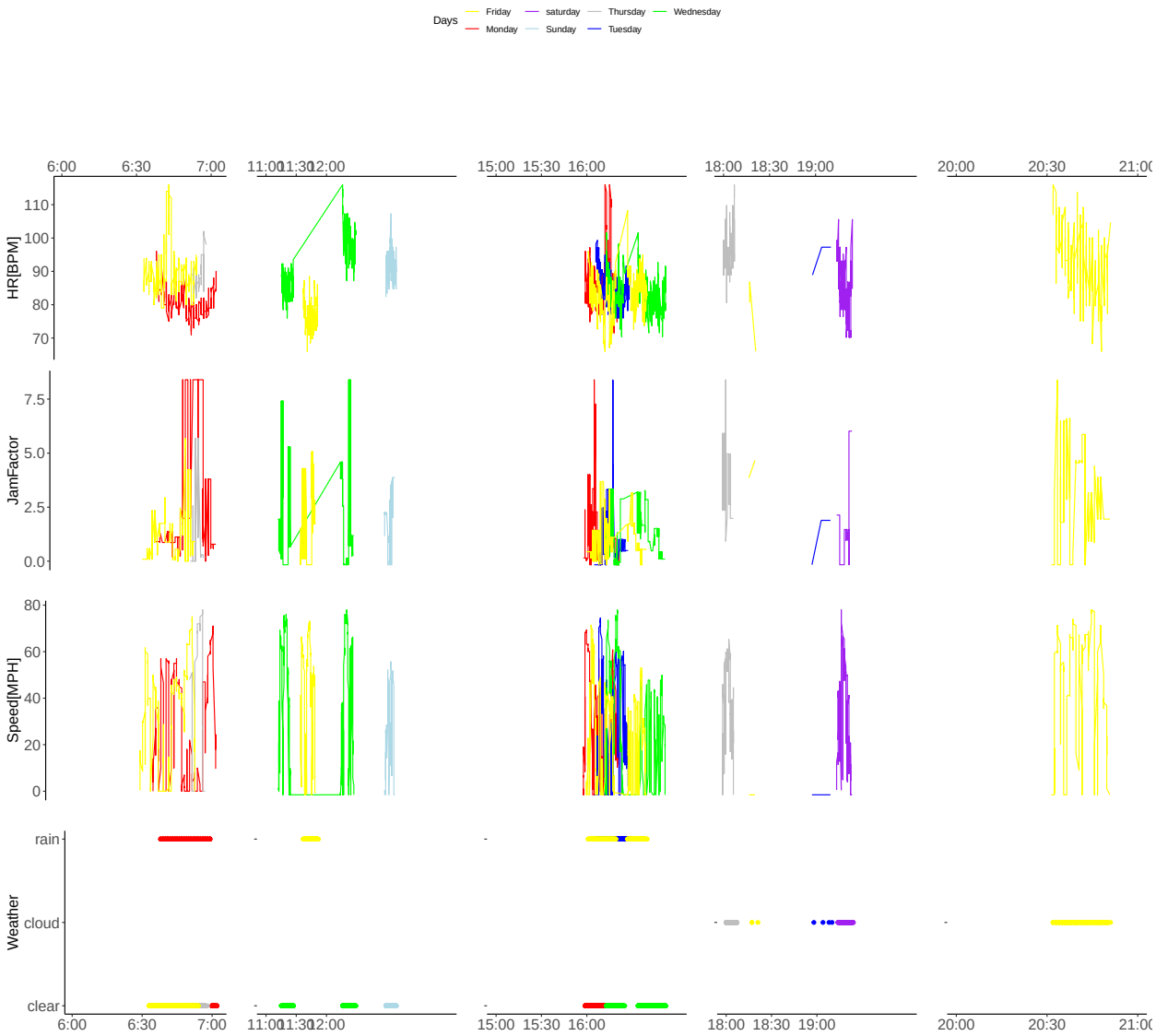
B5N.mean seems to be relatively consistent across trips, with a mean around 4.5-5 and a standard deviation around 1.7-1.8.

B5O.mean shows some variation across trips, with the highest mean in Trip1 and the lowest in Trip4. The standard deviation is relatively consistent around 1.6-1.9.

Overall, these descriptive statistics provide a summary of the central tendency, variability, and distribution of each variable across the different trips.

## Question 3

The summary visualization gives a mean sense of the data at the trip level. This is an appropriate level to perform statistical analysis. However, it is useful as an analyst to have a sense of the underlying moment by moment data for quality control and other reasons. Accordingly, construct a signal figure for participant P27. The figure should show the participant's signals of heart rate, jam factor, vehicle speed, and weather throughout the day for all seven days of the observational period.




## Observations:

The signal figure for participant P27 is shown above. It displays the person's signals for heart rate, jam factor, vehicle speed, and weather throughout the day on each of the seven observational days.

Jam factor: The jam factor describes the flow of traffic at the time of observation. The highest Jam Factor ever recorded was 9.34248 on Day 1, or Monday. It suggests that Monday's traffic was moving very slowly. and the lowest jam factor, which was measured on day two, Tuesday, was 0.00, indicating that traffic flowed freely on that day. The plots show that there was more heavy traffic on weekdays than on weekends. Daily, the traffic volume varied more in the mornings between 7:00 and 8:00 and in the evenings between 4:00 and 5:00.

Weather: It is clear from the graph that there are three different sorts of weather. cloud, rain, and clarity. Weather was a little gloomy on Saturday and Friday, with rains reported on Monday, Wednesday, and Friday.

Day 3 - Wednesday - saw a largely clear sky. On Friday, there was a greater range in the weather; it was clear in the morning, it rained at 11:00 and 4:00, and it was cloudy at night.

Speed:

On day three, Wednesday afternoon, 82 was the fastest speed ever recorded. When comparing weekends, Saturday had larger speed fluctuations than Sunday. In comparison to prior days, the speed was more inconsistent on Friday. Because of the greater variations in traffic from 4:00 to 5:00 pm, the speed also varied more at that time.

Heart rate:

The highest HeartRate recorded value is 117 on Day 1 and the lowest value is 66 on Day 6. On Day 5, there was higher fluctuation in heart rate from morning 6:30 to 7:00 am, afternoon 4:00 to 5:00 pm, and evening 8:00 to 9:00 pm. More variations in heart rate may be seen between 4:00 and 5:00 pm, which corresponds to participant 27's heart rate on Days 1, 2, 3, and 5. Only on Days 3 and 5 did the heart rate fluctuate higher between 11:00 am and 12:00 pm.