Project - Milestone 1

k Michelle Patrick Krueger & Likhitha Mandapati

2023-03-10

Q1 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**.

Q2 Make an exploratory study by creating a data-summarizing visualization. We recommend that you: Create a four-row image to plot. Plot the trip distributions of the behavioral and physiological factors in the first row. Plot the trip distributions of the environmental variables in the second row. Plot the state psychometrics ranges in the third row. Plot the characteristic psychometric distributions in the fourth row. Make a table with these distributions' descriptive data as well. Describe your findings and ideas.

Q3 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.

Objective

This driving study's primary goal was to look at the behavioural and physiological trends among drivers over the course of a normal week. Considering the lack of physical activity involved in driving, any cardiovascular activation that is seen during driving is most likely caused by sympathetic arousal or potentially parasympathetic reduction. The research included a variety of trait, state, vehicular, observational, and environmental factors that have previously been associated with sympathetic reactions to account for this. These factors were chosen to account for a sizable portion of the autonomic activation in the drivers, which would account for a sizable portion of the observed cardiovascular activation. In order to account for cardiovascular activation brought on by a decrease in parasympathetic activity, the research also included indicators of circadian rhythms, such as morning and afternoon periods. Through complementary models centered on the participants' non-driving lives, the research also sought to find aspects of sympathetic activation and parasympathetic reduction that are not specific to driving.

Dataset Introduction

The dataset is a compilation of physiological, behavioral, and psychometric data collected over the course of seven days from 21 individuals. The dataset contains statistics on the age, gender, and trait anxiety score of each participant. Using an Apple Watch, physiological information like heart rate is also captured during the participants' everyday commutes. Behavioral data, such as vehicle speed and absolute throttle position, is captured using an OBD2 device, while environmental data, such as traffic conditions and weather, is recorded at each observation instant. The dataset also contains psychometric information, such as the NASA-TLX and State Anxiety scores of the participants, which evaluate perceived workload and anxiety levels at the beginning of each day, respectively. The Big-Five personality characteristics are also noted, which include agreeableness, conscientiousness, extraversion, neuroticism, and openness.

Question-1

## Participant Locations in Texas
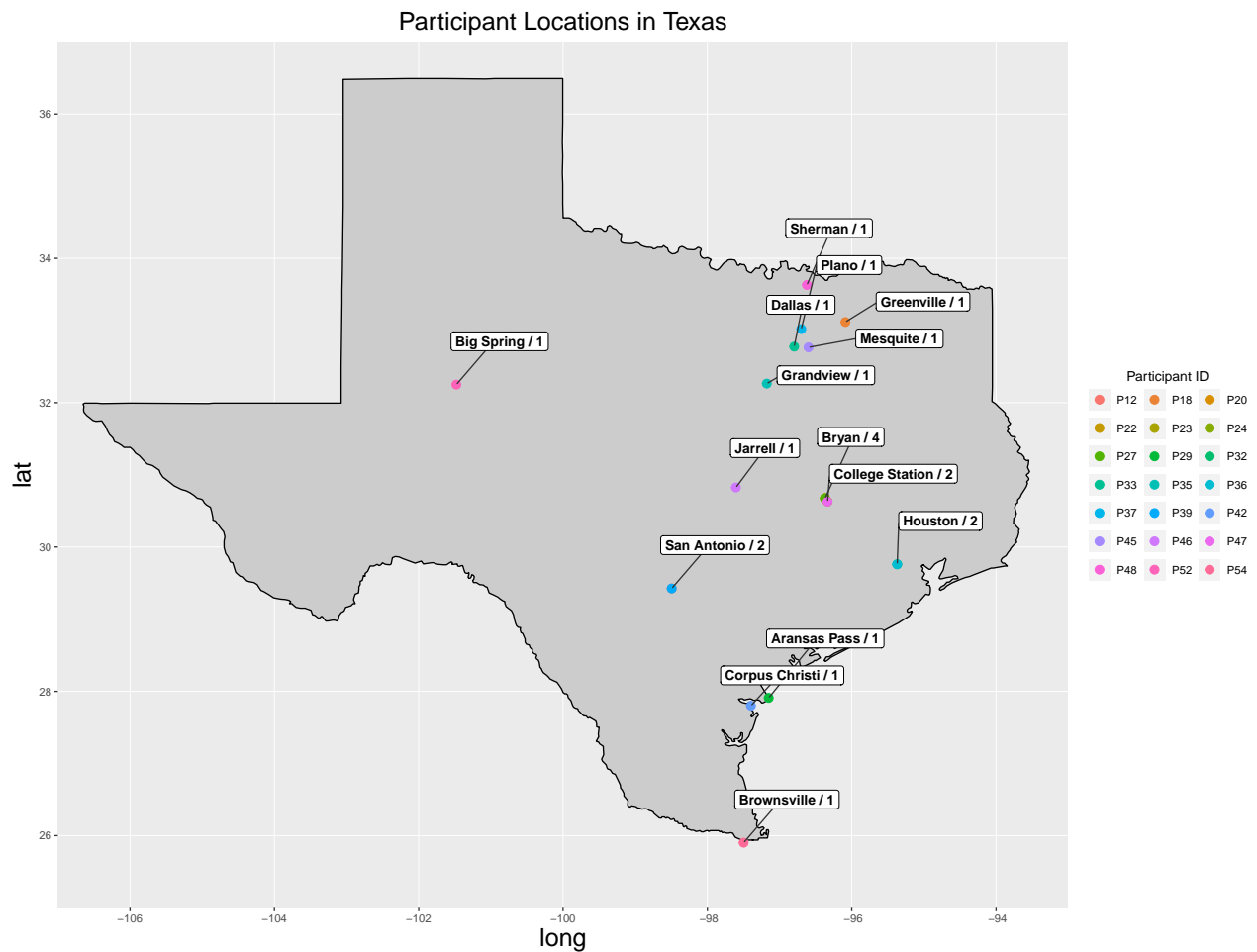
Observations

TThe map visualization displays the distribution of the 21 participants in a Texas research on statewide naturalistic driving. A number of the participant towns are shown on the map, and the size of the label next to each city's name indicates how many participants are from that city. With 4 and 2 participants each, Bryan and College Station are home to the bulk of the participants. San Antonio, Houston, Aransas Pass, Corpus Christi, and Brownsville each have two participants, while Sherman, Greenville, Plano, Dallas, Mesquite, Grandview, Big Spring, and Jarrell each have one member.

The geographical distribution of the individuals across Texas is also shown on the map. There are no participants in the state's western or northern areas; instead, they are all concentrated in the state's eastern, central, and southern regions. Each participant's city's longitude and latitude are taken into account, and the map is scaled to display the complete state of Texas.

Question-2

- Plot of all the physiological, behavioral, enviornmental, and psychometric variables.
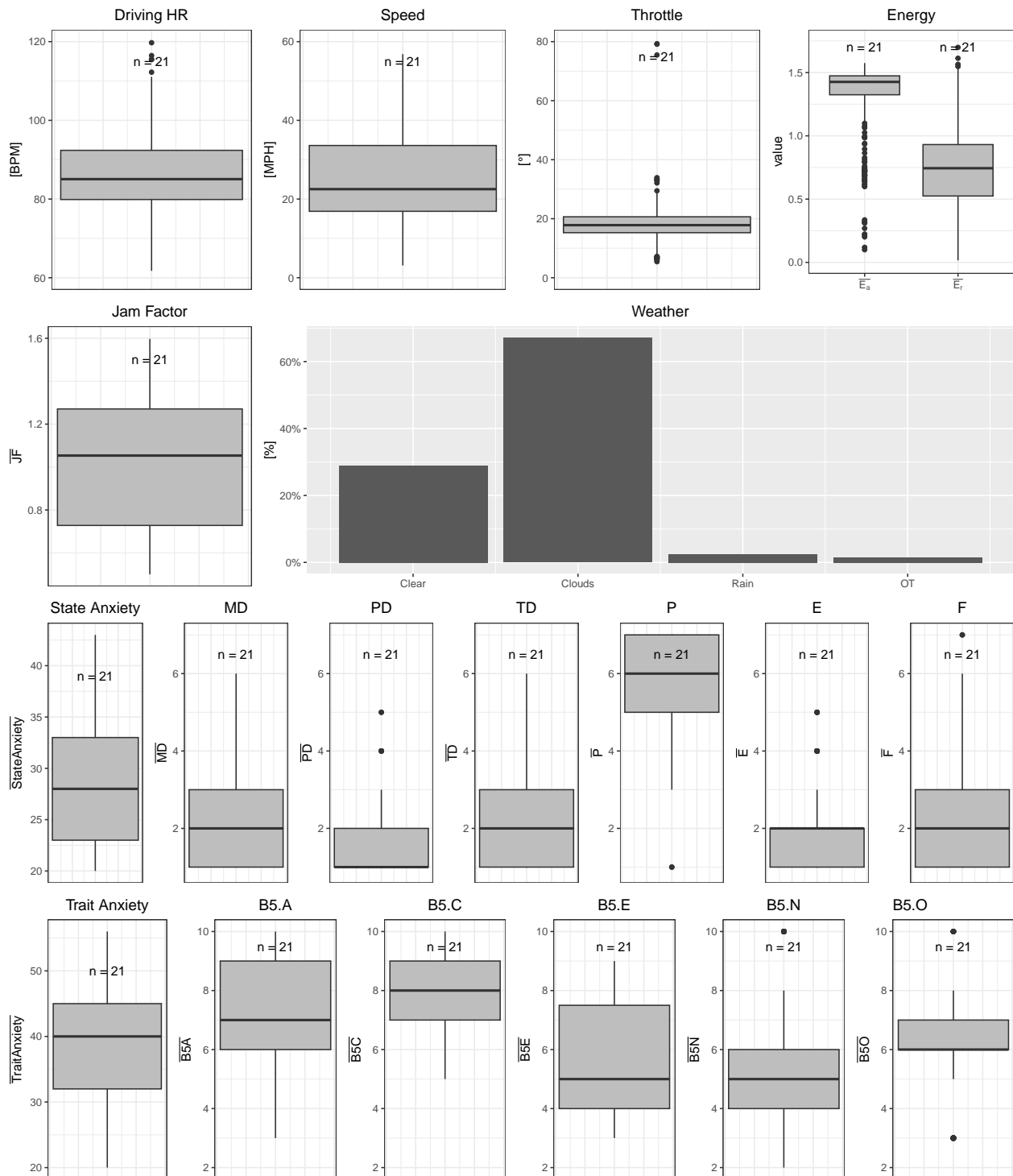
trip distributions of physiological and behavior variables

	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

Observations

Heart Rate (HR):

Trip1's mean heart rate is the lowest (84.76 bpm), and Trip3's mean heart rate is the greatest (91.00 bpm). The HR's standard deviation (SD), which ranges from 11.40 to 11.55 bpm, is reasonably constant between trips. The median heart rate, which ranges from 82 to 89 bpm, is also constant between journeys. Conclusion: There are only slight variations in the heart rates of the participants during the various trips. **Speed:**

The mean speed of rip1 is the fastest (34.65 mph), and the mean speed of trip7 is the slowest (16.19 mph). Each trip's SD for speed is different, with Trip1's SD (26.26 mph) being the greatest and Trip6's SD being the lowest (11.43 mph). With a range of 15 to 34 mph, the median speed is comparatively constant across trips. Conclusion: The speed of the participants changes considerably between the trips, with some trips involving fast driving and others featuring slow driving. **ATP (Absolute Throttle Position):**

The mean ATP of Trip1 is the greatest (21.98%), while the mean ATP of Trip4 is the lowest (16.31%). Trip1 had the highest SD (15.48%) and Trip4 had the lowest SD (6.97%) for ATP, which fluctuates between trips. The median ATP, which ranges from 12.50% to 18.40%, is fairly constant between visits. Conclusion: The use of the accelerator pedal by the participants differs considerably across trips, with some trips having more frequent and intense acceleration than others.

Acceleration:

The mean acceleration for Trip1 is the greatest (1.23 m/s2), while the mean acceleration for Trip4 is the lowest (1.13 m/s2). Trip1 had the highest SD (0.47 m/s2) and Trip5 had the lowest SD (0.13 m/s2). The SD of acceleration changes between trips. The median acceleration, which ranges from 1.30 to 1.43 m/s2, is fairly constant between journeys. Inference: The measured acceleration, which varies slightly between trips, and the participants' use of the accelerator pedal are consistent.

Rotation:

The mean rotation for Trip6 is the greatest (0.82 deg/s), while the mean rotation for Trip7 is the lowest (0.43 deg/s). Trip1 had the highest SD (1.14 deg/s), while Trip7 had the lowest SD (0.80 deg/s). The SD of rotation changes between trips. The median rotation, which ranges from 0.22 deg/s to 0.35 deg/s, is fairly constant across journeys. Conclusion: The participants' rotational use of the steering wheel changes slightly between trips.

trip distributions of environmental variables

	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

Observations

Jam Factor

The average JF values for each trip range from 1.134 to 1.872, which is a low figure. Each trip's JF standard deviation is fairly constant, with values varying from 1.349 to 1.495. The median JF values for each journey, which range from 0.717 to 1.686, are also quite low. Conclusion: Participants usually drove steadily and avoided making sudden, erratic movements while driving, according to the low values of JF. According to the consistent standard deviation of JF across journeys, participants' driving behavior was largely consistent under various driving circumstances. The low median values of JF imply that participants' driving behavior rarely changed dramatically or abruptly during each journey. **Weather**

Each journey's weather has a mean value between 1.000 and 2.000, with trip 4 having the highest mean value.

In terms of consistency across journeys, the weather standard deviation ranges from 0.426 to 0.660.

For the first five trips, the weather's median number is consistently 2.000; however, for trips six and seven, it falls to 0.000.

Conclusion: The participants may have faced more difficult driving circumstances during trip 4 due to the high mean value of the weather, such as heavy rain or strong winds. The constant standard deviation of the weather indicates that the driving circumstances were generally constant throughout various journeys. While the median value of 2.000 for the first five trips indicates that participants encountered some degree of unfavorable weather conditions during these trips, the drop in the median value of weather to 0.000 for trips 6 and 7 suggests that participants encountered clear weather conditions during these trips.

trip distributions of state psychometrics

	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

Observations

State Anxiety

Trips 3 and 4 have the greatest SA mean values (32.588 and 33.806, respectively), while Trips 6 and 7 have the lowest values (24.000 for both).

Trip5 has the greatest SA standard deviation values (12.742), followed by Trip4 (11.054), and Trip6 and Trip7 have the lowest values (0.000 for both).

Situational awareness (SA) mean values for trips 3 and 4 are greater than those for the other trips, suggesting that the participants were more conscious of their surroundings and the road conditions during these trips. The participants may have been less involved or attentive during trips 6 and 7, which have the lowest SA mean values. There may have been more variation in the participants' situational awareness levels on Trip 5, as evidenced by the trip with the greatest SA standard deviation.

MD(Mental Demand)

Trip4 has the greatest MD mean value (2.273), while Trip6 and Trip7 have the lowest values (1.000 for both).

Trip1 has the highest MD standard deviation number (1.215), while Trip6 and Trip7 have the lowest values (0.000 for both).

Conclusion: Trip 4 has the highest mean mental demand (MD) value, indicating that participants may have felt more mental stress and cognitive load during this journey. The participants may have encountered lower amounts of mental demand on trips 6 and 7, as those trips have the lowest MD mean values.

PD(Physical Demand)

Trip1 has the greatest PD mean value (1.923), and Trip6 and Trip7 have the lowest values (1.000 for both). Trip1 has the greatest PD standard deviation (1.153), and Trip6 and Trip7 have the lowest values (0.000 for both). Conclusion: Trip 1 has the highest mean physical demand (PD) value, indicating that the participants may have performed more strenuous driving activities on this journey. The participants may have performed less physically demanding driving duties on trips 6 and 7, as these have the lowest PD mean values.

TD(Temporal Demand)

The fact that Trip4's TD mean is higher than those of the other trips suggests that drivers were under more time constraint on this trip. Additionally, Trip3's TD mean is comparatively greater. The fact that Trip1's TD standard deviation is higher suggests that participants' levels of temporal demand varied more on this journey. Conclusion: Trips 3 and 4 have greater temporal demand (TD) mean values than the other trips, which may indicate that the participants were under more time pressure and were required to meet higher temporal demands. The participants may have encountered less temporal pressure on trips 6 and 7, which have the lowest TD mean values.

P(Performance)

The fact that Trip1 and Trip5 have higher P means that the drivers performed better on these journeys. P standard deviation is higher for Trips 2 and 3, which suggests that participant performance varied more on these excursions. Conclusion: Trip 5 has the highest performance (P) mean value, indicating that participants may have driven more skillfully on this journey than on the others. The participants may have displayed lower levels of driving ability on trips 6 and 7, which have the lowest P mean values.

E(Effort)

Trips 2 and 4 have higher E means, suggesting that the drivers worked harder on these trips. For Trips 1, 2, and 3, the E standard deviation is comparatively higher, suggesting that participant effort varied more on these trips. Conclusion: The physical demand (E) dimension has the greatest mean values for Trips 2 and 4, suggesting that the participants may have performed more physically taxing driving duties on these trips. The E mean values for trips 5 and 6 are the lowest, indicating that the participants may have performed less physically demanding driving duties on these trips.

F(Frustration)

Trips 3 and 4 have higher F means than the other trips, which suggests that the drivers were more irritated on these travels. Additionally, F mean for Trips 6 and 7 is significantly lower, suggesting that the drivers experienced less annoyance on these trips. For Trips 3 and 4, the F standard deviation is comparatively higher, suggesting that participant frustration varied more on these trips. Conclusion: The functional task (F) dimension has the greatest mean values for Trips 3 and 4, suggesting that the participants may have engaged in more difficult and complex driving tasks during these trips. The participants may have engaged in simpler driving activities on Trip 1 based on the trip's lowest F mean value.

trip distributions of trait psychometrics

	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

Observations

B5A

For performance metric B5A, which gauges the driver's control over speed, trips 4 and 5 have the lowest and greatest mean values, respectively. This implies that the participants may have driven more carefully and slowly during Trip 4 and more quickly during Trip 5. For performance metric B5A, Trips 5 and 6 have the greatest and lowest standard deviation values, respectively, suggesting that the driver's speed control may have been more variable on Trip 5 than on Trip 6.

B5C

The mean values for performance metric B5C, which gauges the driver's lateral position control, are lowest and greatest for trips 4 and 5, respectively. This indicates that the participants' lateral positions may have been more consistent during Trip4 and more variable during Trip5. For performance measure B5C, Trips 5 and 6 have the greatest and lowest standard deviation values, respectively, suggesting that the driver's lateral position control may have been more variable during Trip 5 than it was during Trip 6.

B5E

Trips 1, 2, and 3 have greater mean values than the other trips for the performance metric B5E, which gauges the driver's ability to control longitudinal acceleration. This implies that the participants may have been accelerating and decelerating more forcefully during these journeys. The performance metric B5E for Trip 5 has the highest standard deviation value, suggesting that the driver's control of longitudinal acceleration may have been less consistent during this journey.

B5N

Trips 4 and 5 have the lowest and highest mean values for performance metric B5N, which gauges how frequently the driver interacts with the secondary controls of the car, respectively (e.g., radio, HVAC, etc.). This indicates that compared to Trip4, the participants on Trip5 might have been more easily sidetracked.

B5O

Trips 1, 2, 3, and 7 have greater mean values than the other trips for the performance metric B5O, which gauges the driver's steering wheel reversals. This indicates that the participants may have reversed the steering wheel more frequently during these journeys, possibly as a result of navigating intersections with more turns or driving on curvier roads.

Question-3
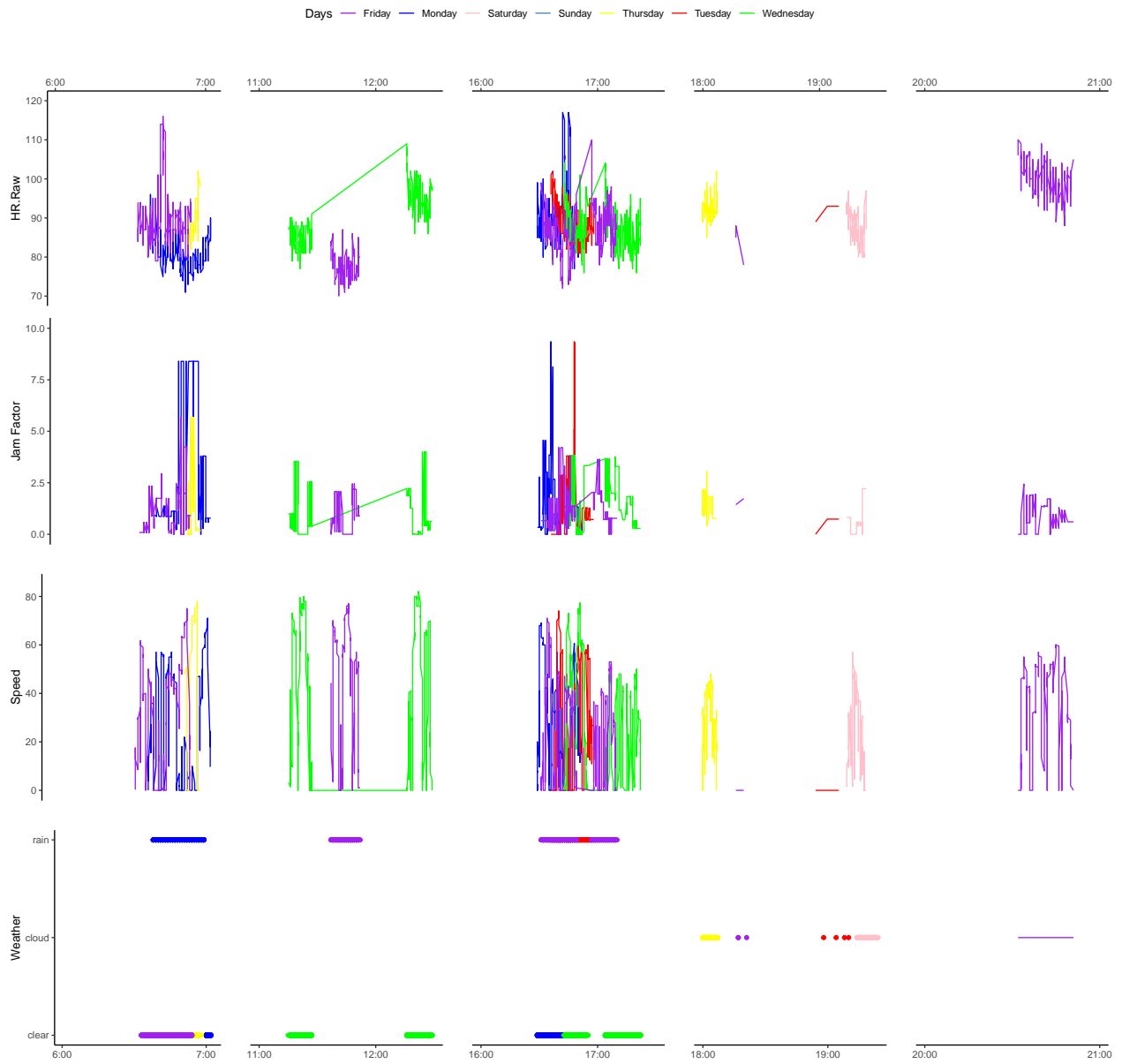
The plots show the relationship between time and weather across different days. Each point on the plot represents the weather condition at a specific time. The x-axis represents the time of the day while the y-axis represents the weather condition. There are different colors used for different days to distinguish them from one another. The plots show that the weather conditions are not constant throughout the day, and there are fluctuations in the weather pattern over time. The plots reveal that the weather conditions are more stable during the day and become more unpredictable in the evening and night. There is a consistent pattern of clear weather conditions in the morning and early afternoon, which tends to transition into cloudy and rainy conditions in the later part of the day. The plots provide insight into the range of weather conditions experienced during the observed period and how they change over time. The plots show that on some days, the weather conditions follow a particular pattern, while on other days, they are more random and unpredictable. Overall, the plots provide a comprehensive visualization of the weather patterns over time and help to identify trends and changes in the weather conditions.

Observations

The above plot is a signal figure for participant P27, that shows the participant-27's signals of heart rate, jam factor, vehicle speed, and weather throughout the day for all seven days of the observational period.

HeartRate

The maximum value of HeartRate registered is 117 registered on Day1 and minimum value is 66 on day 6. On Day 5, Heart rate was more fluctuating from morning 6:30 -7:00 am, 4:00 pm - 5:00 pm and 8:00pm to 9:00 pm. From 4:00pm - 5:00pm, we can see more fluctuations in Heart rate which means the Heart rate for Participant 27 on days Day1, day2, Day3, day5.From 11:00 am - 12:00 pm, Heart rate was more fluctuating only on Days 3 and 5.

Jam factor

Jam factor indicates traffic conditions in each observational moment. Highest Jam Factor recorded is 9.34248 and was recorded on Day1 i.e. Monday. which indicates that there wass very slow flow of traffic on Monday. and Lowest Jam factor was 0.00 recorded on day2 - tuesday, which means there was free flow of traffic on Tuesday.From the plots, it is evident that there was heavy traffic recorded on weekdays than weekends. On daily basis, the traffic level was fluctuating more morning 6-7 am and evening 4:00pm - 5:00 pm

Weather

From graph, it is evident that there are 3 types of weathers Cloud, rain, clear. Rains recorded on Monday, Wednesday and Friday weather was a bit cloudy on Saturday and Friday. The sky was mostly clear on Day3 - Wednesday.The weather was more fluctuating on Friday, as it was clear sky in the morning, rained at 11:00am and 4:00pm and then it was cloudy at night.

Speed

The highest speed recorded was 82 on day3 - Wednesday afternoon.Comparing weekends, the speed was more fluctuating on Saturday than Sunday.On Friday, the speed was more fluctuating comparating to other days. As there was more fluctuations in traffic from 4:00pm - 5:00 pm, the speed was also more fluctuating from 4:00 - 5 pm