PSY 5210 Exam 2

Prateek Kumar
11/11/18

PSY 5210 Exam 2

Submitted by: Prateek Kumar

Date: 11 November 2018

Table of Contents

Problem 1 2
A2_1_MOPP 3
B1_2_BDU 4
C2_2_MOPP 5
D2_2_MOPP 6
Problem 2 8
A2_1_MOPP 8
B1_2_BDU 9
C2_2_MOPP 10
D2_2_MOPP 11
Problem 3 12
A2_1_MOPP 12
B1_2_BDU 15
Problem 4 18

We are consulting with a medical group that is recording a number of physiological sensors on their patients using a mobile device. We have the sensor data for 4 patients.

Problem 1

Here we are dealing with 5 sensor data for each patient for each time interval, the sensor data we are considering here are: ECG.HR, EDR.BR, Belt.BR, CoreTemp, and Temp. The meaning of the following medical abbreviations is:

- ECG.HR: Heart rate measure of the patient from the ECG signal.
- EDR.BR: Breathing rate measured from the ECG signal
- Belt.BR: Breathing rate measured from the belt
- CoreTemp: Core Temperature of the patient's body
- Temp: Real Temperature of the patient's body

In our dataset we observed no missing values but there were some data points for CoreTemp which were out of their reasonable range, we observed the values as 0 but normal core body temperature of a healthy human being is stated to be 98.6°F or 37.0°C and if the core body temperature is 0 that means the patient froze to death, if by any case this happens then we won't get any data further but the 0 values are for certain time reading and we have readings after that as well so we can conclude that there might be some defect in the machine which altered or data.

So inorder to approximately visualize our data we converted the 0 values to NA values, thus this will result in blanks in our plot but we can atleast get an idea as how the values are changing.

Average measure of a normal person:

- ECG.HR average 60-100 per minute
- CoreTemp average 36.3-37.3 C
- Temp average 36.5-37.5 C

Now, looking at the 4 patient's data:

A2_1_MOPP

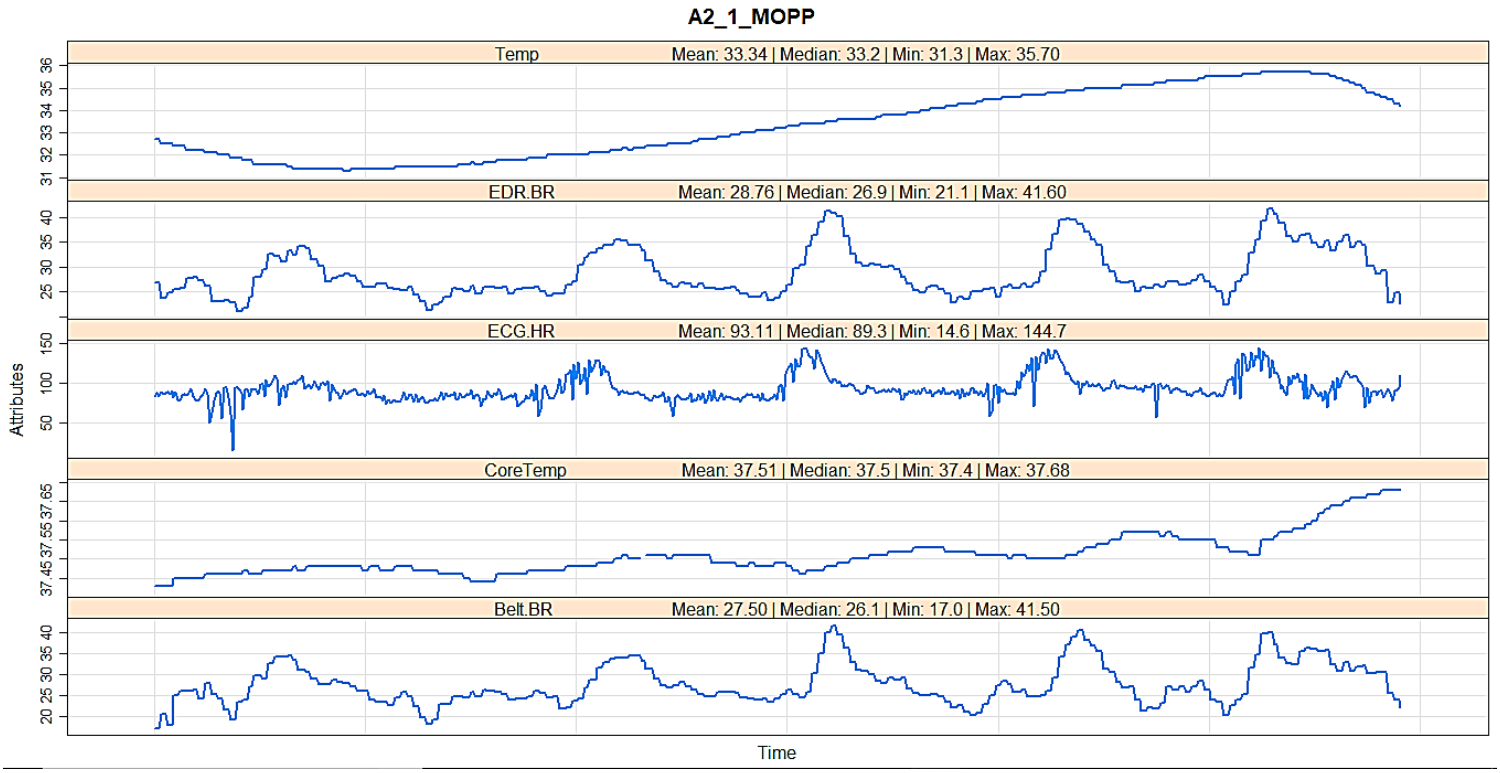

Figure 1: The plot for patient A2_1_MOPP

The above plot is of patient A2_1_MOPP. The data values are from time 12:59:28 to 13:48:43. Now talking about the 5 attributes:

- **ECG.HR**: The heartrate of the patient is normal at the later time interval, but in the beginning it dropped to a very low value which is 14.6 and also the heartrate maximum value here is 144.7. The average heartrate is 93.11 which is normal. So we can say that the heartrate of the patient is normal except for the starting reading which might be due to dropping heartrate or some defect in the machine.
- **EDR.BR & Belt.BR**: Breathing rate measured from the ECG signal and from belt have almost similar values except for some decimal difference.
- **CoreTemp**: The core temperature of the patient's body is low initially which is increasing by a small amount except at the end where it increased with a high rate. The minimum value is 37.4 and maximum value is 37.68 but the average is 37.51, here we see that the values are higher than the normal human body.
- **Temp**: The Real Temperature of the patient's body seems sinusoidal, it increases and then starts to decrease at the end. The minimum value is 31.3 and maximum value is 35.7 but the average is 33.34.

The data seems fine here and except for some values won't compromise our product.

B1_2_BDU

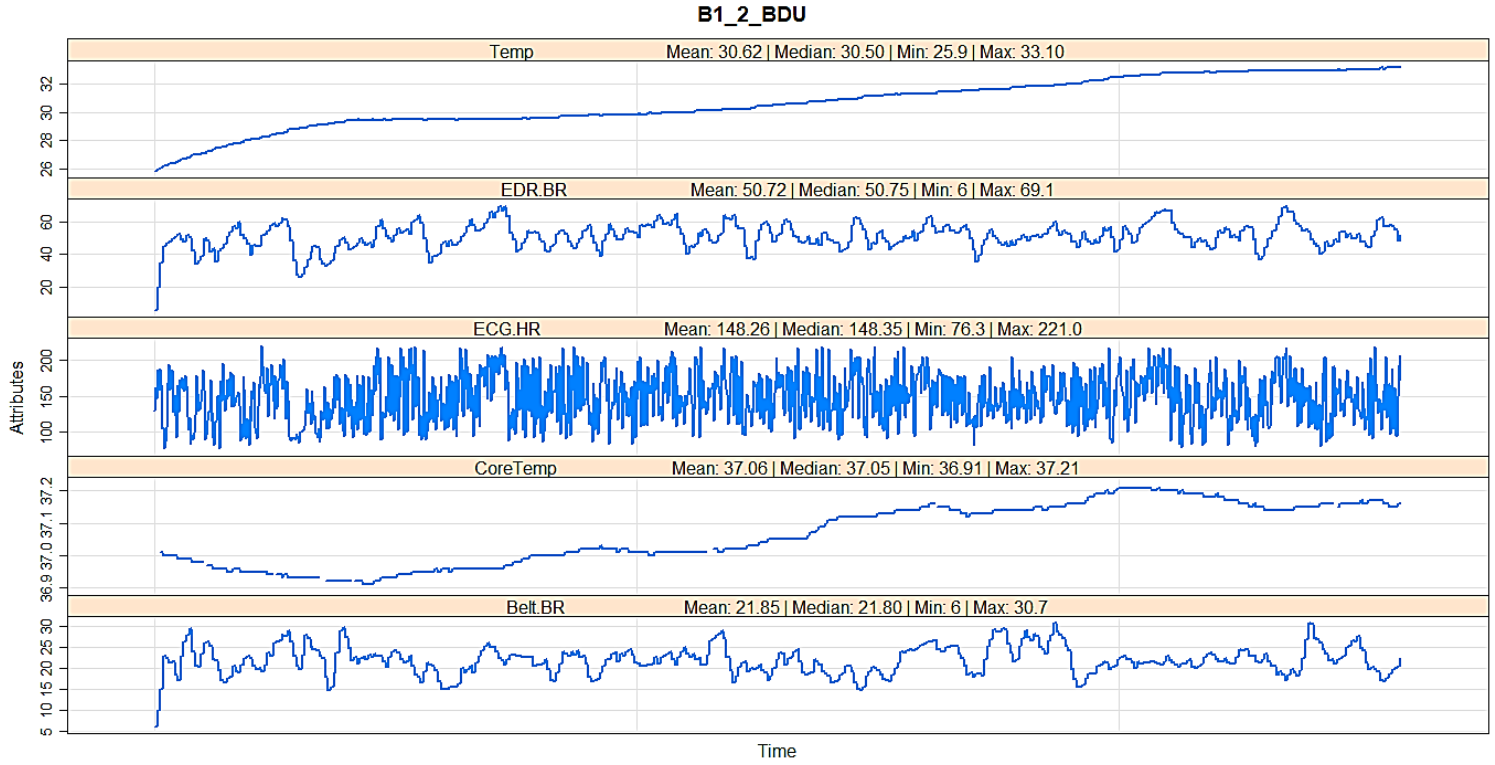

Figure 2: The plot for patient B1_2_BDU

The above plot is of patient B1_2_BDU. The data values are recorded in seconds interval but the time resets in the middle and starts over again. Now talking about the 5 attributes:

- **ECG.HR**: The heartrate of the patient appears very noisy here. The minimum value is 76.3 and maximum value is 221, the average heartrate is 148.35. We see from the above plot that the heartrate is oscillating constantly from low to high value within very small time frame. Also in general the heartrate of the patient is very high than the normal.

- **EDR.BR & Belt.BR**: Breathing rate measured from the ECG signal and from belt are also very different in this case. The reading initially started as same but soon changed and both the readings have a drastic difference.

- **CoreTemp**: The core temperature of the patient's body seems very slightly sinusoidal but on compromise a straight line. It decreases in the beginning and then increases. The minimum value is 36.91 and the maximum value is 37.21 and the average is 37.06. The values here are same as that of a normal person.

- **Temp**: The Real Temperature of the patient's body is increasing constantly. Within the recorded time frame it rose from 25.9 to 33.10

The data seems inappropriate here and except for some values it will compromise our product.

C2_2_MOPP

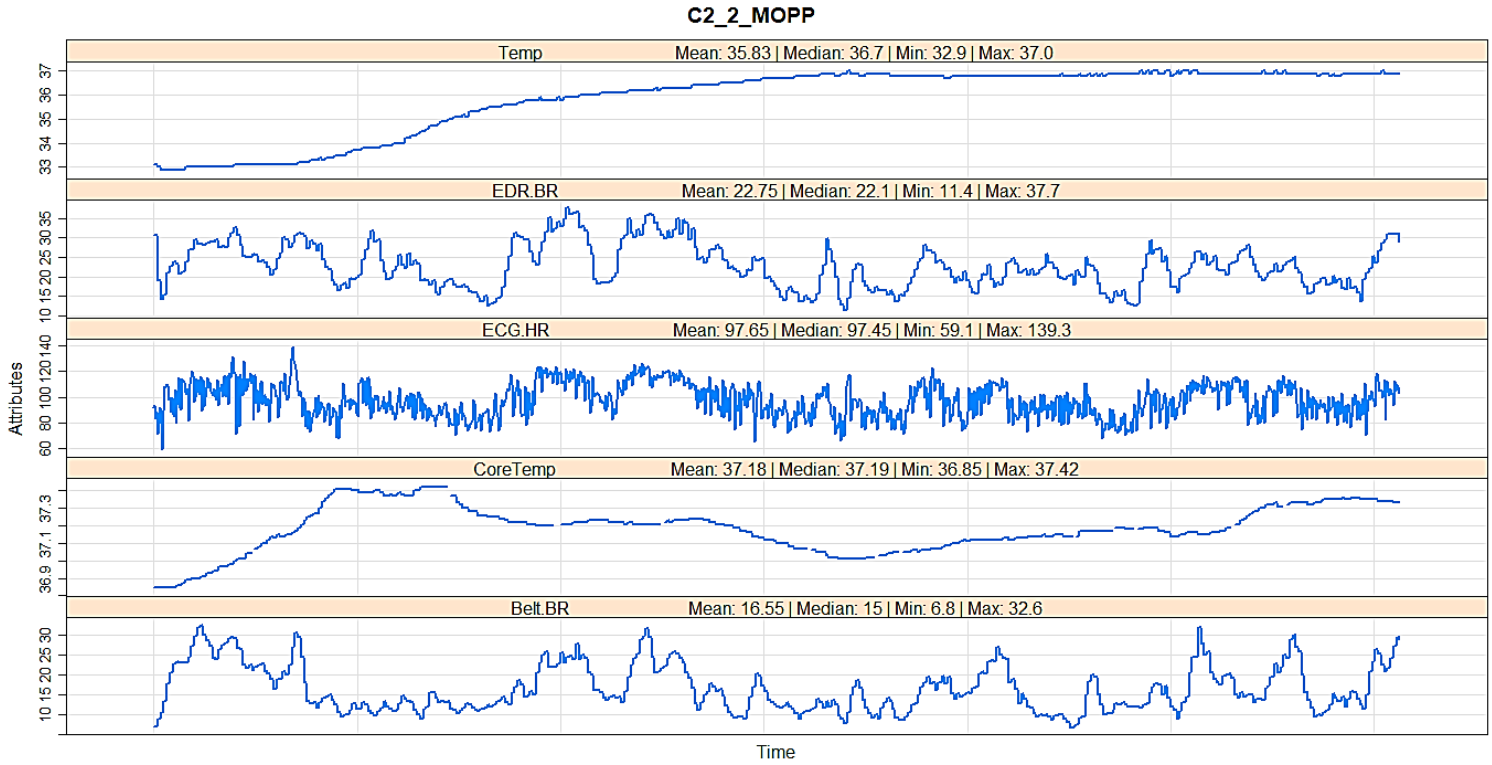

Figure 3: The plot for patient C2_2_MOPP

The above plot is of patient C2_2_MOPP. The data values are recorded in every 5 second interval from 11:01:19 to 12:43:24. Now talking about the 5 attributes:

- **ECG.HR**: The heartrate of the patient appears noisy here as well. The minimum value is 59.1 and maximum value is 139.3, the average heartrate is 97.65. We see from the above plot that the heartrate is oscillating constantly from low to high value within very small time frame. But in general the heartrate of the patient is normal.

- **EDR.BR & Belt.BR**: Breathing rate measured from the ECG signal and from belt are different in this case. The reading of the attributes is different from the start only.

- **CoreTemp**: The core temperature of the patient's body increases and then decreases but on compromise is a straight line because there is just a change of 1°. The minimum value is 36.85 and the maximum value is 37.42 and the average is 37.18. The values here are same as that of a normal person.

- **Temp**: The Real Temperature of the patient's body is increasing constantly. Within the recorded time frame it rose from 32.9 to 37

The data seems inappropriate here and except for some values it will compromise our product.

D2_2_MOPP

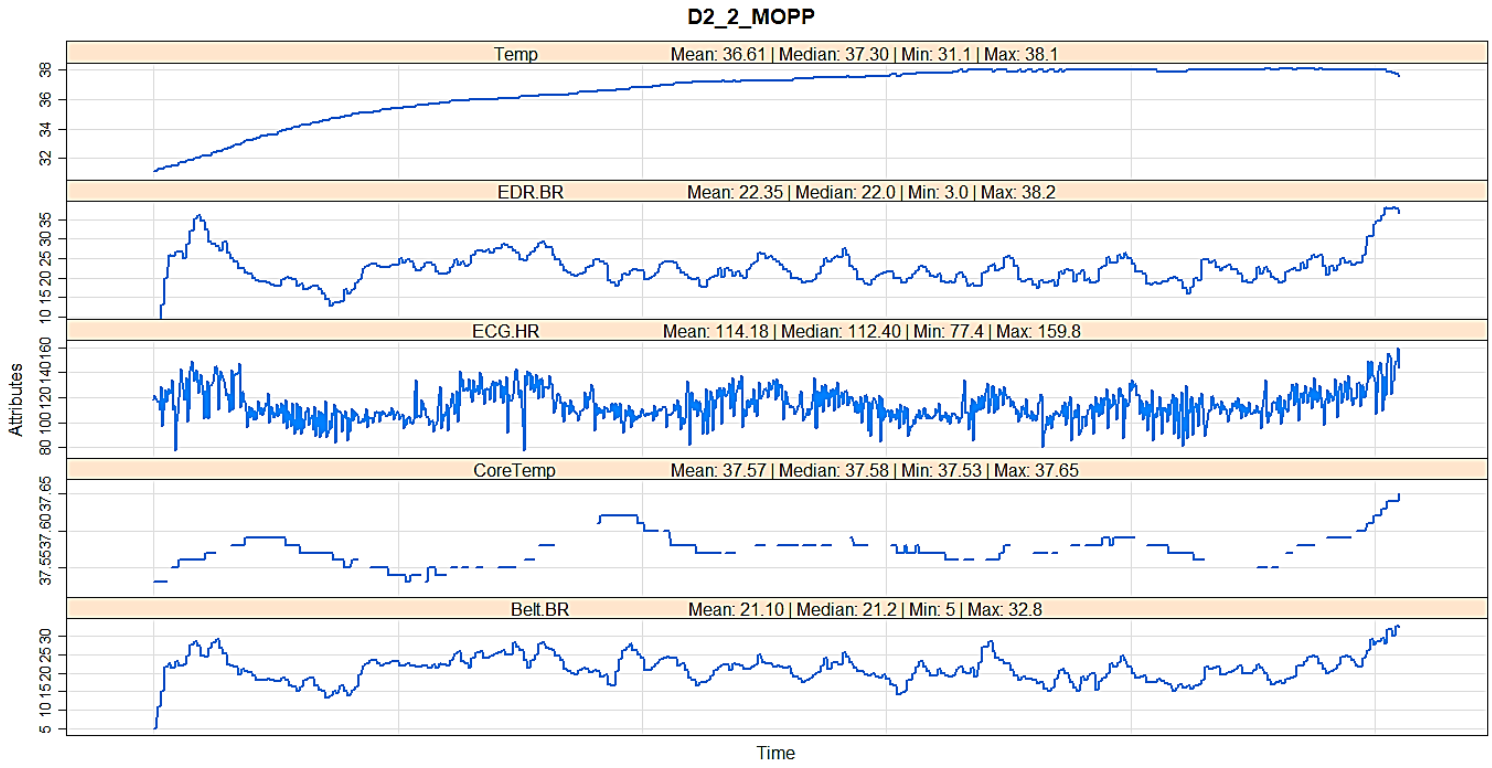

Figure 4: The plot for patient D2_2_MOPP

The above plot is of patient D2_2_MOPP. The data values are recorded in every 5 second interval from 16:22:52 to 17:47:52. Now talking about the 5 attributes:

- **ECG.HR**: The heartrate of the patient appears noisy here as well. The minimum value is 77.4 and maximum value is 159.8, the average heartrate is 114.18. We see from the above plot that the heartrate is oscillating constantly from low to high value within very small time frame. Also in general the heartrate of the patient is above normal.
- **EDR.BR & Belt.BR**: Breathing rate measured from the ECG signal and from belt are a bit similar in this case. The reading of the attributes has just a slight difference.
- **CoreTemp**: The core temperature of the patient's body increases and then decreases and continues as so but altogether is a straight line because there is just a change of 0.12°. The minimum value is 37.53 and the maximum value is 37.65 and the average is 37.57. The values here are same as that of a normal person.
- **Temp**: The Real Temperature of the patient's body increases and then appears constant. Within the recorded time frame it rose from 31.1 to 38.1

The data seems inappropriate here and except for some values it will compromise our product.

Additional Attributes:

- Overlaying each data on the same plot made the plot messy and a person cannot easily depict each of the attributes, so I plotted each attribute in different panel, trying to depict the real ECG monitor.

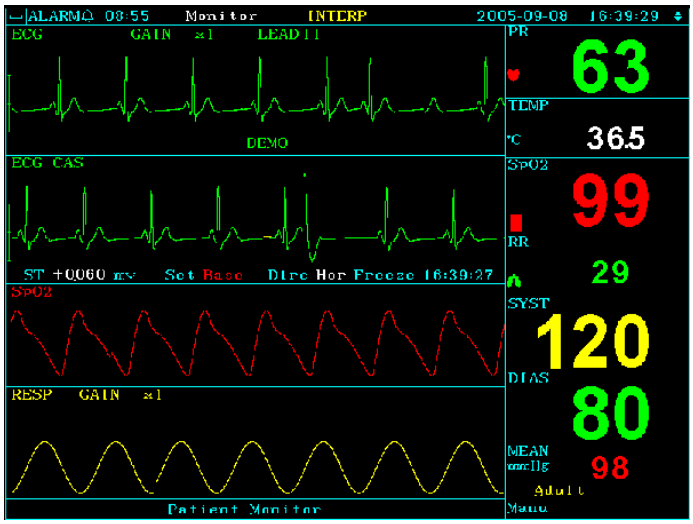

Figure 5: Real ECG monitor

- Zooming in the scales for each attribute, because when we plot all the attributes on the same scale then for the attributes (here Temp) which has a small range seems to be a straight line and we miss the minute details about the data.
- Setting up names of attributes in the panels(Legend) to identify which plot is about which attribute
- Putting grid lines for proper view.
- Instead of point I am displaying the line to depict the ECG monitor visual
- Increasing the thickness of the lines for proper view.
- Displaying the title to identify which plot corresponds to which patient
- Displaying min/max value for each data to get the minimum and maximum readings of the attribute
- Displaying average value for each data to get the average readings of that attribute
- Displaying median value for each data to get the median readings of that attribute

Problem 2

Here we have to create a loess regression that estimates the midline of the ECG.HR values based on time.

Here for each of the 4 data I used enough polynomial predictors (span values) to capture the overall patterns of the data over time inorder to get major peaks and valleys and came up with the best estimator midline.

I checked up with the R-squared values and stopped when I got the predictor which gave the maximum R-squared value.

Below are the loess regression plots of ECG.HR vs Time for each of the 4 patient data:

A2_1_MOPP

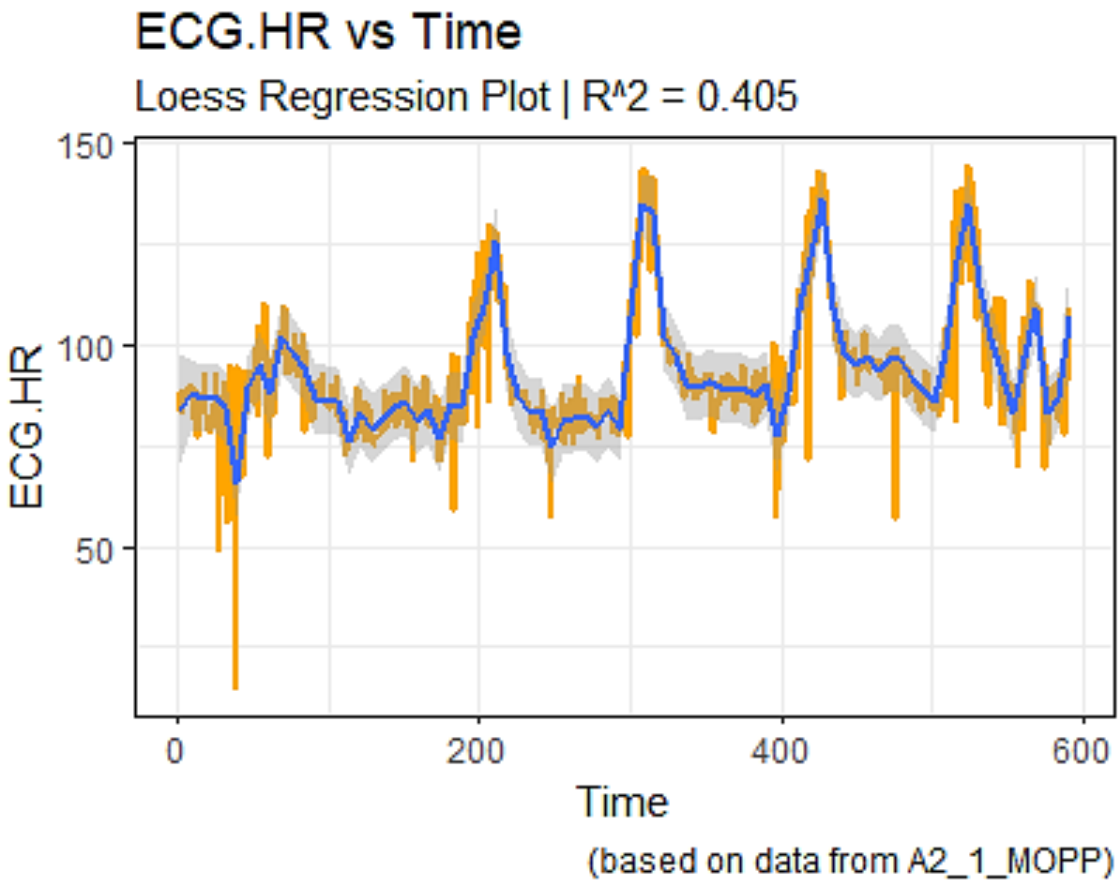

Figure 6: Loess Regression plot for A2_1_MOPP between ECG.HR and Time

Here is used the span value of 0.02 and got the above loess line(blue) with $R^2 = 0.405$

B1_2_BDU

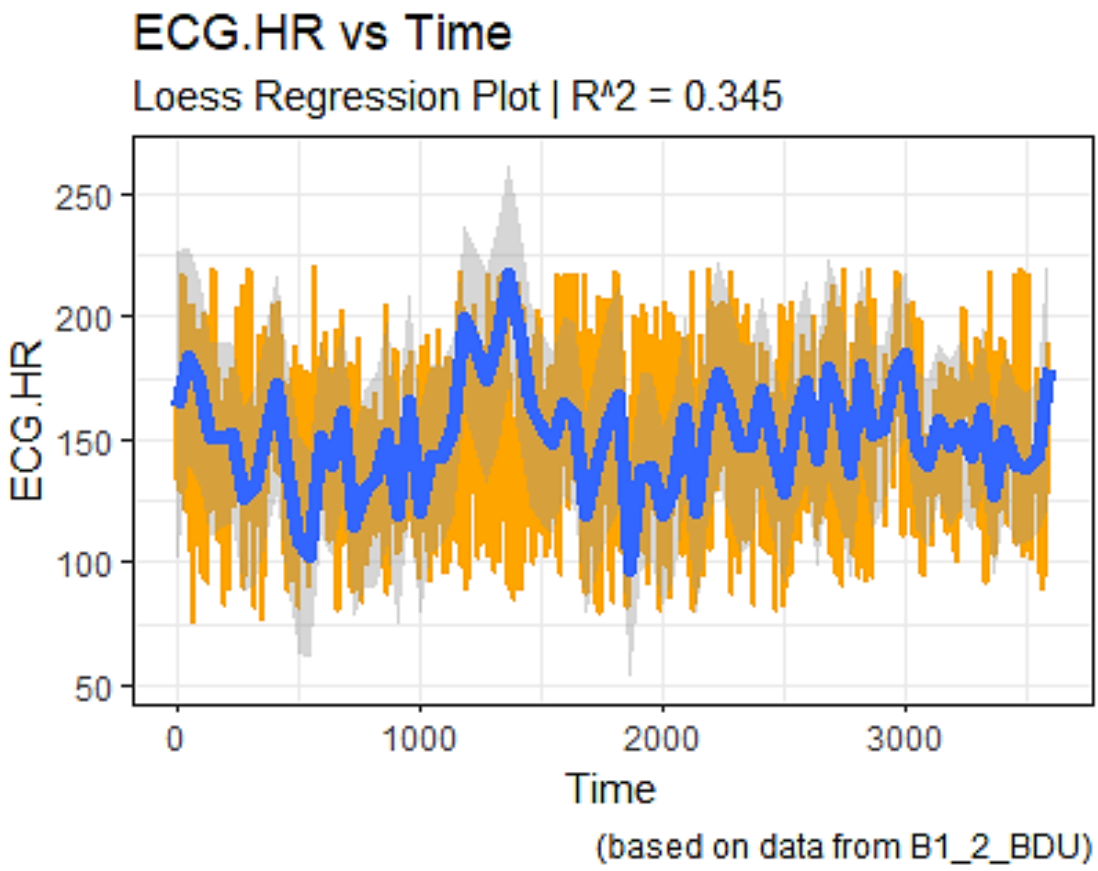

Figure 7: Loess Regression plot for B1_2_BDU between ECG.HR and Time

Here is used the span value of 0.008 and got the above loess line(blue) with $R^2 = 0.345$

C2_2_MOPP

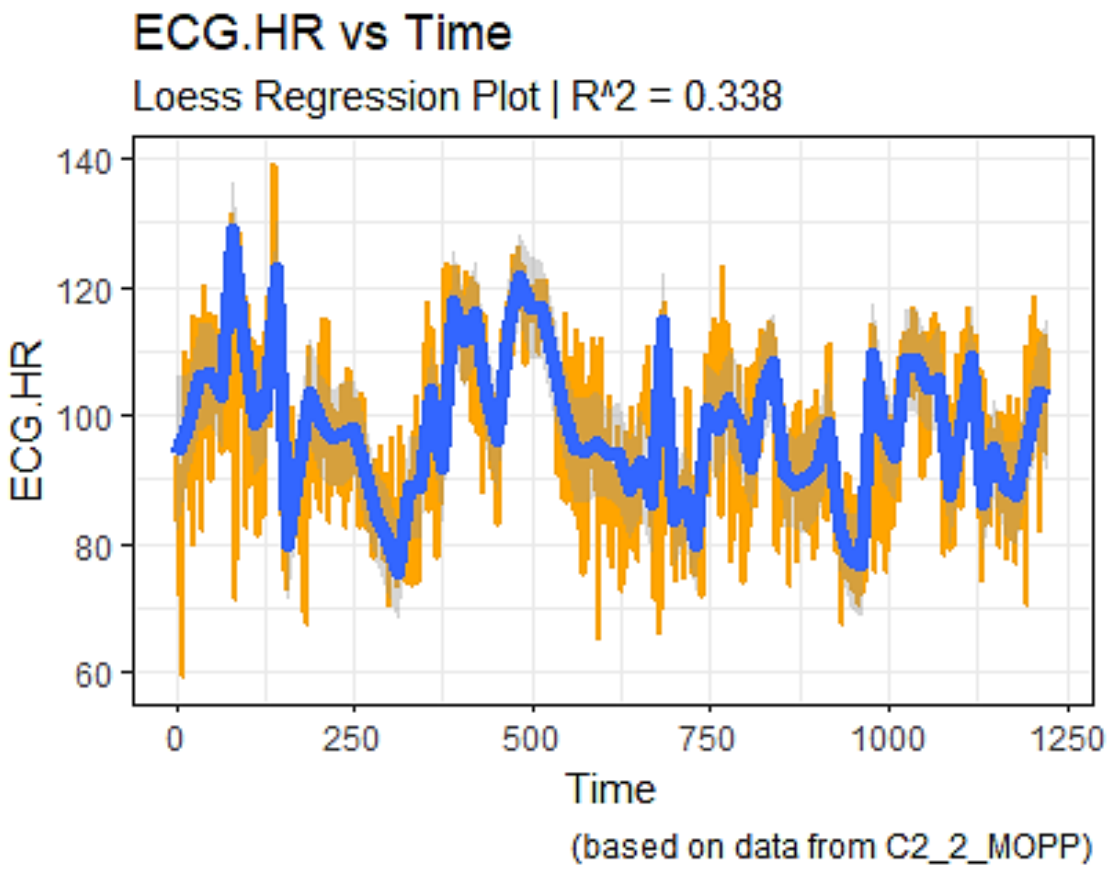

Figure 8: Loess Regression plot for C2_2_MOPP between ECG.HR and Time

Here is used the span value of 0.008 and got the above loess line(blue) with $R^2 = 0.338$

D2_2_MOPP

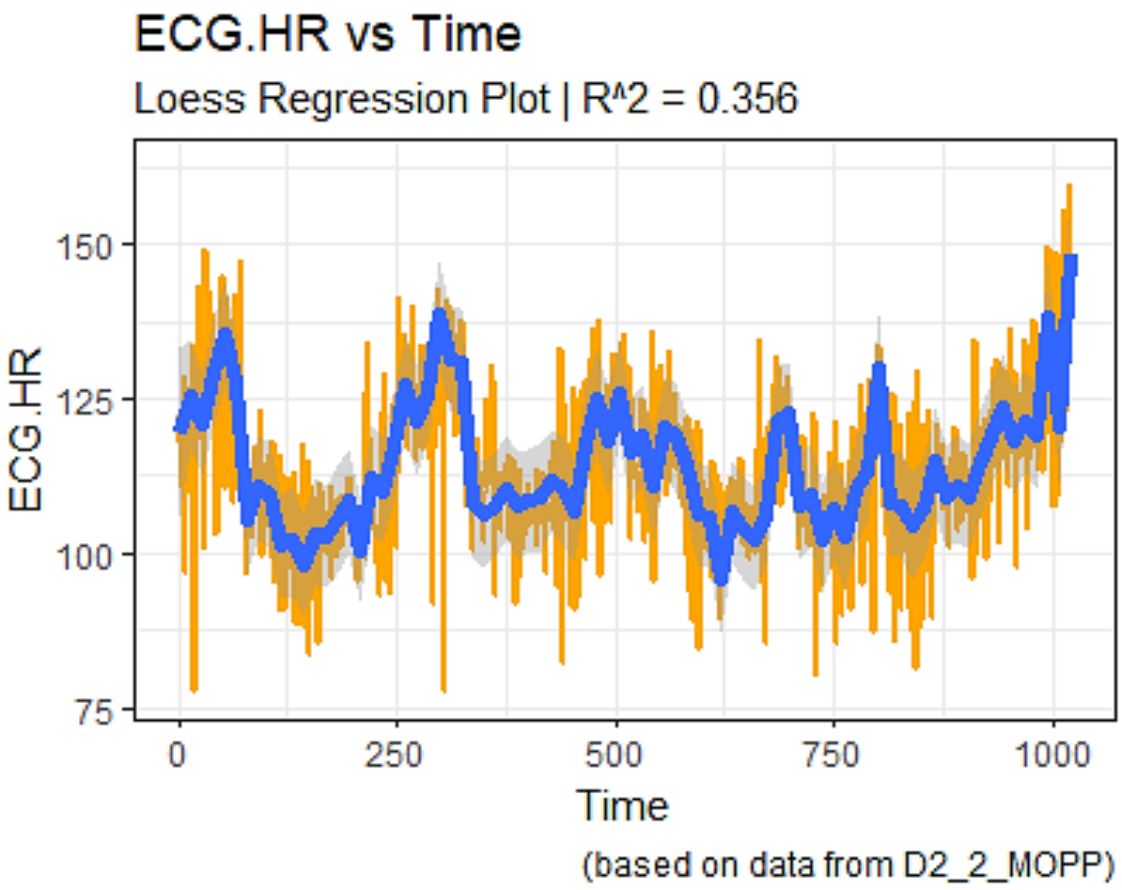

Figure 9: Loess Regression plot for D2_2_MOPP between ECG.HR and Time

Here is used the span value of 0.011 and got the above loess line(blue) with $R^2 = 0.356$

We use the argument "span" in order to tell R how much smoothing we want. The larger the span, the more points that are included in the weighted estimation, and the smoother the plot will look. The limitation though, is making the span too small might over fit the data, but we want it to give us some idea of the pattern that we see.

So if we fit every point on its own, we may as well just look at a scatterplot. On the other hand, if we smooth too much, we may as well just estimate a linear regression. The idea is to find a balance between the two using the smoothing parameter.

Here in the above plots we see that the plots are very noisy, it fluctuates from high to low constantly within a small time frame, which leads the loess curve to move up and down. While we eyeball we see that for the span value the loess curve best fits the plot but still we get a low R-squared value for each plot because of the noisy nature of the data, which seems to be a limitation in this case.

Problem 3

Breathing rate was measured twice, once from a chest belt sensor, and once inferred from the ECG sensors (EDR.BR, and Belt.BR).

Separately for the A2 and B1 data files, we took the two BR variables, and combined them into a single data vector (one dimensional). Next we created a time predictor and a categorical predictor specifying which type of Breathing measure and made 2 data frames each from A2 and B1 patients data.

We then converted time and rate to numeric and left the categorical predictor(type) as factor itself.

Next we created polynomial regression model that attempts to predict breathing rate over time, using time (and polynomials of time) and type as predictors. We changed the degree value for each polynomial regression inorder to best fit the model.

Now while changing the degree we constantly monitored the R-squared value and the value increased as we increased the degree.

A2_1_MOPP

Below is the table for the A2 data with their degree and corresponding R-squared value

Degree	R-squared
1	0.124
2	0.127
3	0.129
4	0.164
5	0.203
6	0.203
7	0.204
8	0.215
9	0.221
10	0.221
20	0.586
30	0.606
50	0.609

70	0.615
110	0.654

Figure 10: Table of Degree of the polynomial regression and the corresponding R-squared value for A2_1_MOPP

We see in the above table that as the degree of the polynomial regression increases the R-squared value also increases. I tried comparing the above results with the AIC, BIC and ANOVA values and observed a similar result.

Below is the table for the AIC values.

Degree	AIC
1	3592.892
2	3590.008
3	3588.004
4	3540.478
5	3484.557
6	3486.269
7	3485.601
8	3470.394
9	3461.773
10	3463.435
20	2719.201
30	2663.670
50	2657.552
70	2643.376
110	2521.038

Figure 11: Table of Degree of the polynomial regression and the corresponding AIC value for A2_1_MOPP

We can see above that the AIC value is decreasing and our final model's AIC value is the lowest.

Now, comparing with the BIC values.

Degree	BIC
1	3608.122
2	3610.315
3	3613.387
4	3570.938
5	3520.094

6	3526.882
7	3531.291
8	3521.161
9	3517.616
10	3524.355
20	2810.581
30	2770.280
50	2779.392
70	2780.445
110	2673.338

Figure 12: Table of Degree of the polynomial regression and the corresponding BIC value for A2_1_MOPP

Looking at the above table of BIC values we see that we get suboptimal values in the starting as values decrease and then increase but as we proceed till the end we get the least value.

So we can conclude that the AIC and BIC results support our R-squared results.

But while comparing with ANOVA we get a very absurd result.

Degree	Sum of Sq
1	
2	100.8
3	82.4
4	995.9
5	1113.0
6	5.4
7	50.0
8	319.5
9	194.9
10	6.2
20	10207.4
30	578.2
50	110.4
70	181.3
110	1083.5

Figure 13: Table of Degree of the polynomial regression and the corresponding Sum of Sq value(ANOVA) for A2_1_MOPP

Here we see a high oscillating values for sum of squares with respect to the models. This might happen because ANOVA is not able to support this kind of the regression models.

B1_2_BDU

Below is the table for the B1 data with their degree and corresponding R-squared value

Degree	R-squared
1	0.8541
2	0.8547
3	0.8548
4	0.8547
5	0.8547
6	0.8573
7	0.8579
8	0.8579
9	0.8585
10	0.8588
20	0.8657
30	0.8681
50	0.8685

Figure 14: Table of Degree of the polynomial regression and the corresponding R-squared value for B1_2_BDU

We see in the above table that as the degree of the polynomial regression increases the R-squared value also increases. I tried comparing the above results with the AIC, BIC and ANOVA values and observed a similar result.

Below is the table for the AIC values.

Degree	AIC
1	9245.922
2	9238.585

3	9237.629
4	9239.385
5	9239.924
6	9193.687
7	9185.111
8	9186.723
9	9175.522
10	9171.586
20	9047.470
30	9003.785
50	8999.176
70	8970.397
90	8974.368

Figure 15: Table of Degree of the polynomial regression and the corresponding AIC value for B1_2_BDU

We can see above that the AIC value is decreasing and our final model's AIC value is the lowest.

Now, comparing with the BIC values.

Degree	BIC
1	9263.498
2	9262.019
3	9266.922
4	9274.537
5	9280.935
6	9240.556
7	9237.838
8	9245.310
9	9239.967
10	9241.890
20	9152.926
30	9126.816
50	9139.783

Figure 16: Table of Degree of the polynomial regression and the corresponding BIC value for B1_2_BDU

Looking at the above table of BIC values we see that we get the best model at degree 30 and it increases as we proceed to higher degree order.

So we can conclude that the AIC and BIC results support our R-squared results.

But while comparing with ANOVA we get a very absurd result.

Degree	Sum of Sq
1	
2	331.1
3	104.6
4	8.6
5	51.6
6	1688.7
7	366.1
8	13.4
9	454.7
10	203.7
20	4545.7
30	1600.6
50	337.8
70	1097.8
90	0.9

Figure 17: Table of Degree of the polynomial regression and the corresponding Sum of Sq value(ANOVA) for B1_2_BDU

Here we see a high oscillating values for sum of squares with respect to the models. This might happen because ANOVA is not able to support this kind of the regression models.

Problem 4

Here we are considering the 4th patient's data.

Here we have to create a regression model that predicts breathing rate as measured by the belt (Belt.BR) based on the other attributes.

So I created the model predicting Belt.BR values with respect to all other attributes. I obtained a R-squared value of **0.7525**. Now the model considering all the attributes is a very complex model so I attempted to find a simpler model.

I used BIC for this task and finally the simplest model just contained 6 sensor attributes (Temp, Vbat, BR.consist, ECG.HR, EDR.BR, PWI.Conf).

I then calculated the R-squared value of this model and the value was **0.7506.** We can see that there is not much of a difference between the 2 R-squared values so we can go with the new simpler model.

Finally, I tried to improve the model by applying polynomial regression. I applied polynomial regression on 3 attributes (Temp, ECG.HR, EDR.BR) with the degree values of 10, 2 and 2 respectively, this improved our model a bit and the new R-squared value turned out to be **0.7692.**