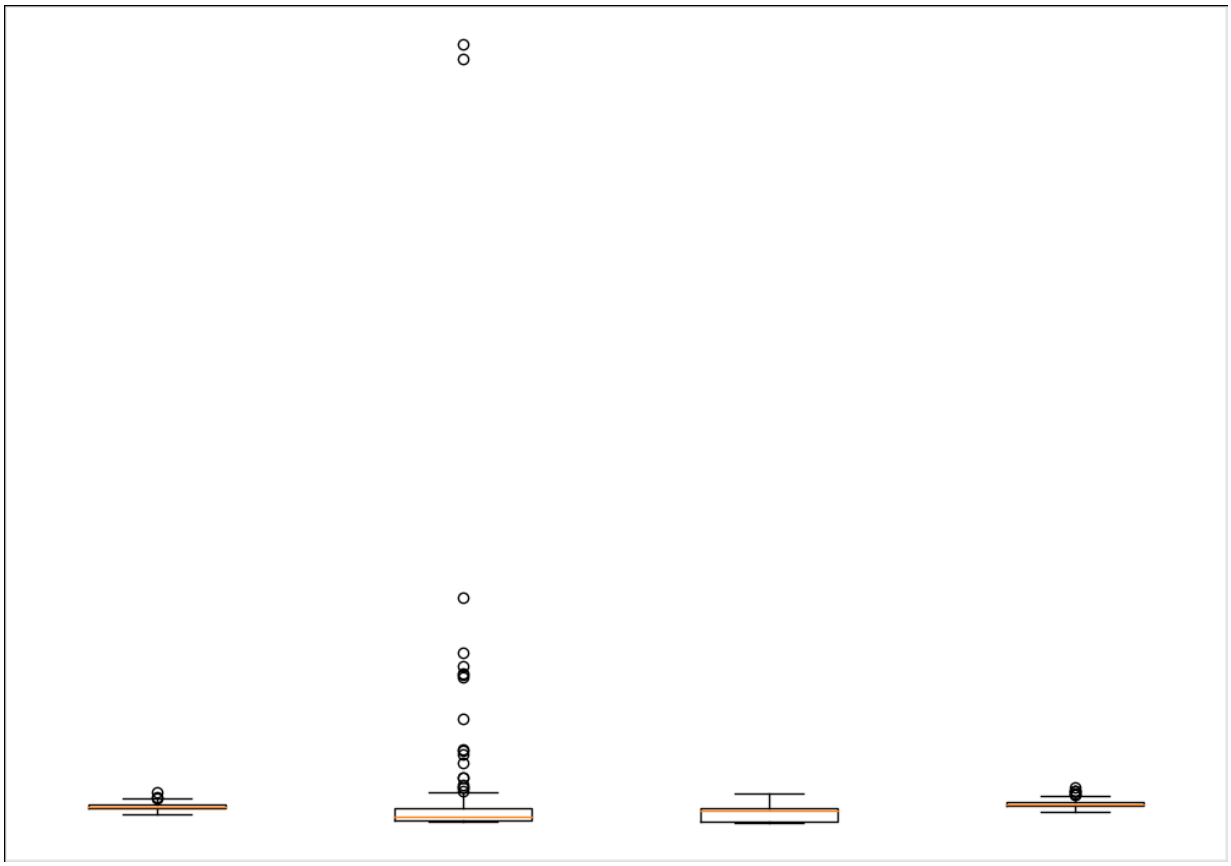
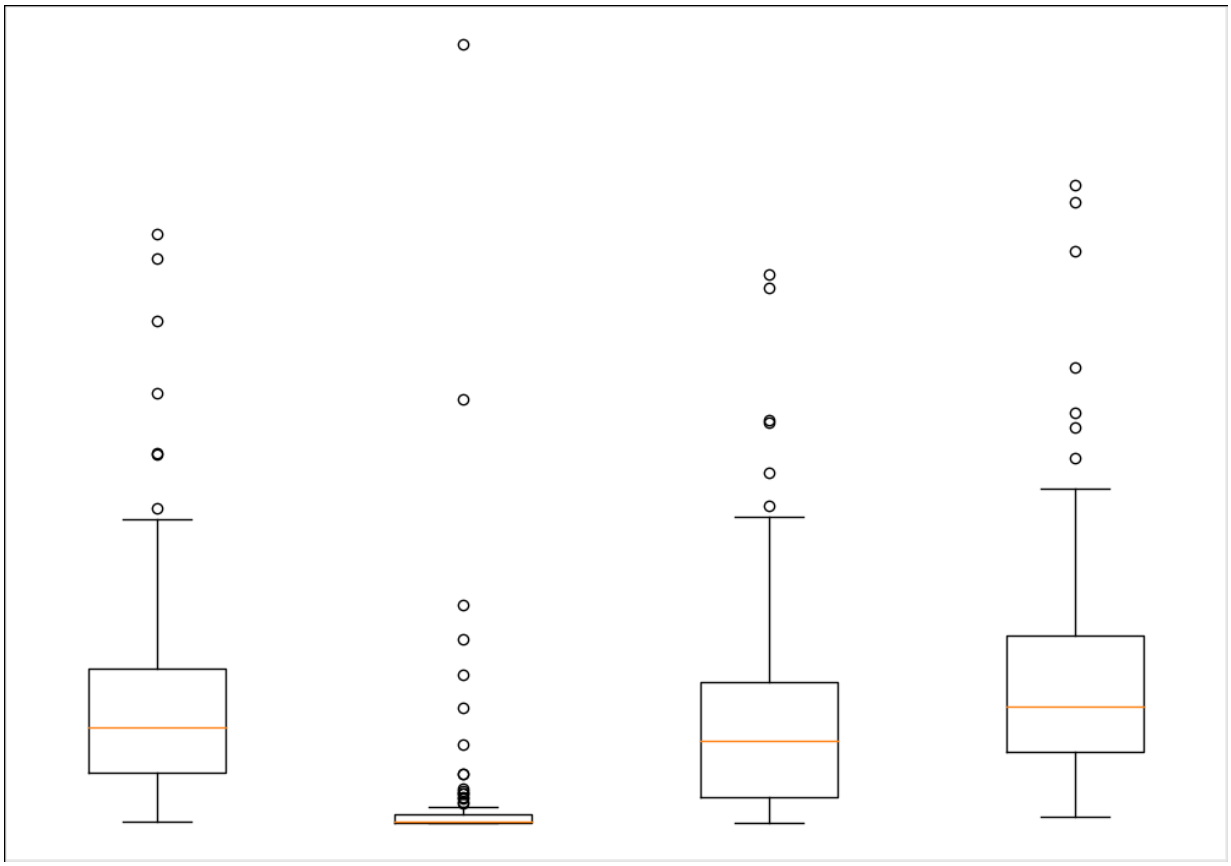
Project 2 Report - Robert Peterson

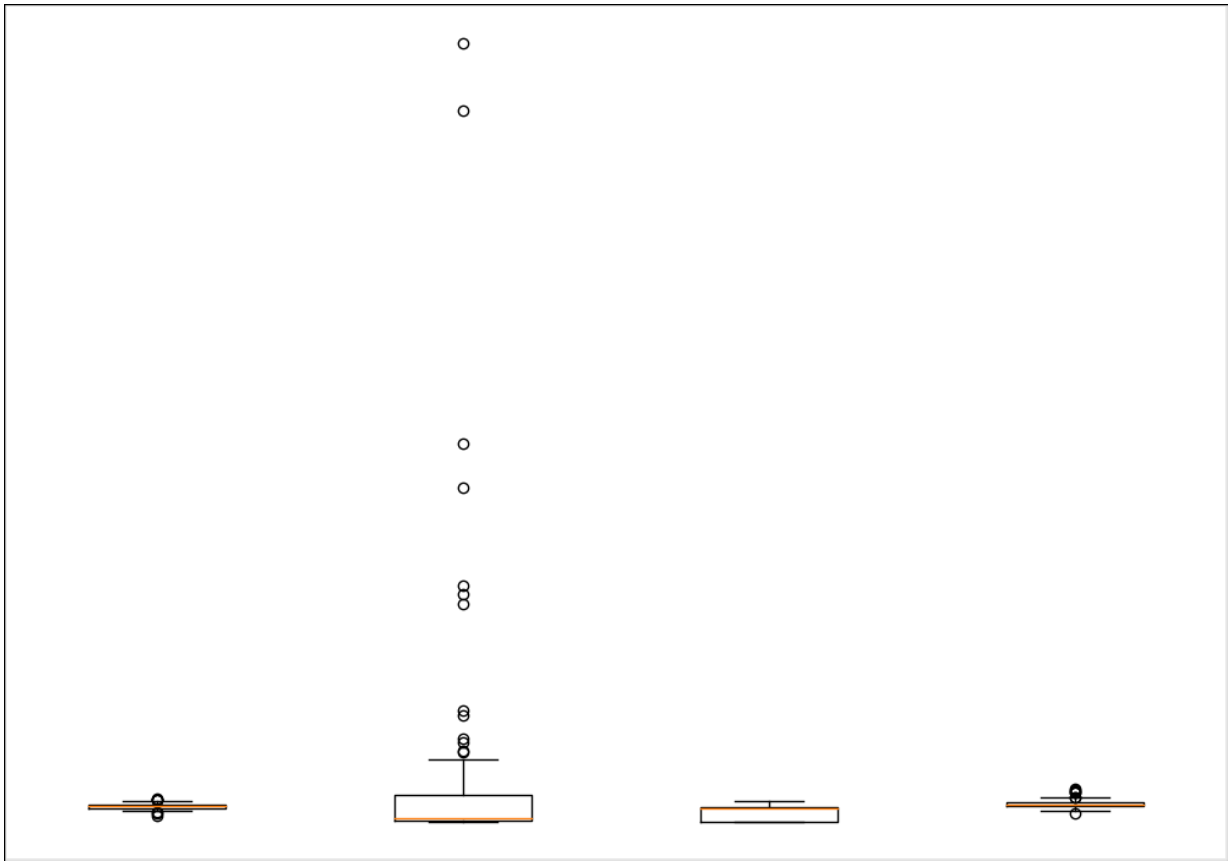
1. I chose to use Random Forest as my classifier after doing testing using SVM and TREE and realizing that RF was the highest accuracy, precision, and recall was with RF all data types. These metrics produced values of 79% for each category.
2. The data type with the highest accuracy was all. However for a single data type Diastolic was the best, followed by Systolic. This type of Blood Pressure reading is indicative of severe pain. According to *Understanding the physiological effects of unrelieved pain*, pain produces a response that increases blood pressure. In *The Relationship Between Blood Pressure and Pain* there was an increase in base systolic BP for subjects with constant pain. All data types being used together makes sense as well as the Blood Pressure is probably the strongest indicator, and uses other data to support the RF decision or correct it if all other signs point to the correct classification.
3. All performed much better for each classifier. This is because the model can learn from the correlational information and weed out the complementary information that may hold less emphasis on decision making than the correlational information. It is great for decision making models as the model begins to function similarly to a human's decision making process as a person will use many forms of information, biased and unbiased, intrinsic and extrinsic, to determine a decision.
4. There was a lot of variability in EDA and Res. I think this is because the breathing rates and electrical responses of subjects are different among subjects based on bodily responses. Breathing is very different based on a subject's health and can be widely different and varied over the course of pain recognition. However in blood pressure the differences can be seen much easier as all metrics are about the same except for blood pressure variance which has many outliers.
5. The physiological signal with the most variability was the Diastolic Blood Pressure. 1 = No pain Dia, 2 = No pain EDA, 3 = No pain Sys, 4 = No pain Res, 5 = Pain Dia, 6 = Pain EDA, 7 = Pain Sys, 8 = Pain Res (I couldn't find out how to change the legend easily so sorry :/). Diastolic BP is the amount of pressure in your arteries between beats of your heart. The light blue and blue lines have the greatest gap between them, suggesting that the greatest difference is in the Diastolic BP of a subject to show the difference in pain vs no pain. All other groups do not show much difference from pain and no pain. Even in the computed metrics of a Random Forest on just Diastolic BP, it returned 69% accuracy, recall, and precision. This was highest among all other data types (closely followed by Systolic BP).



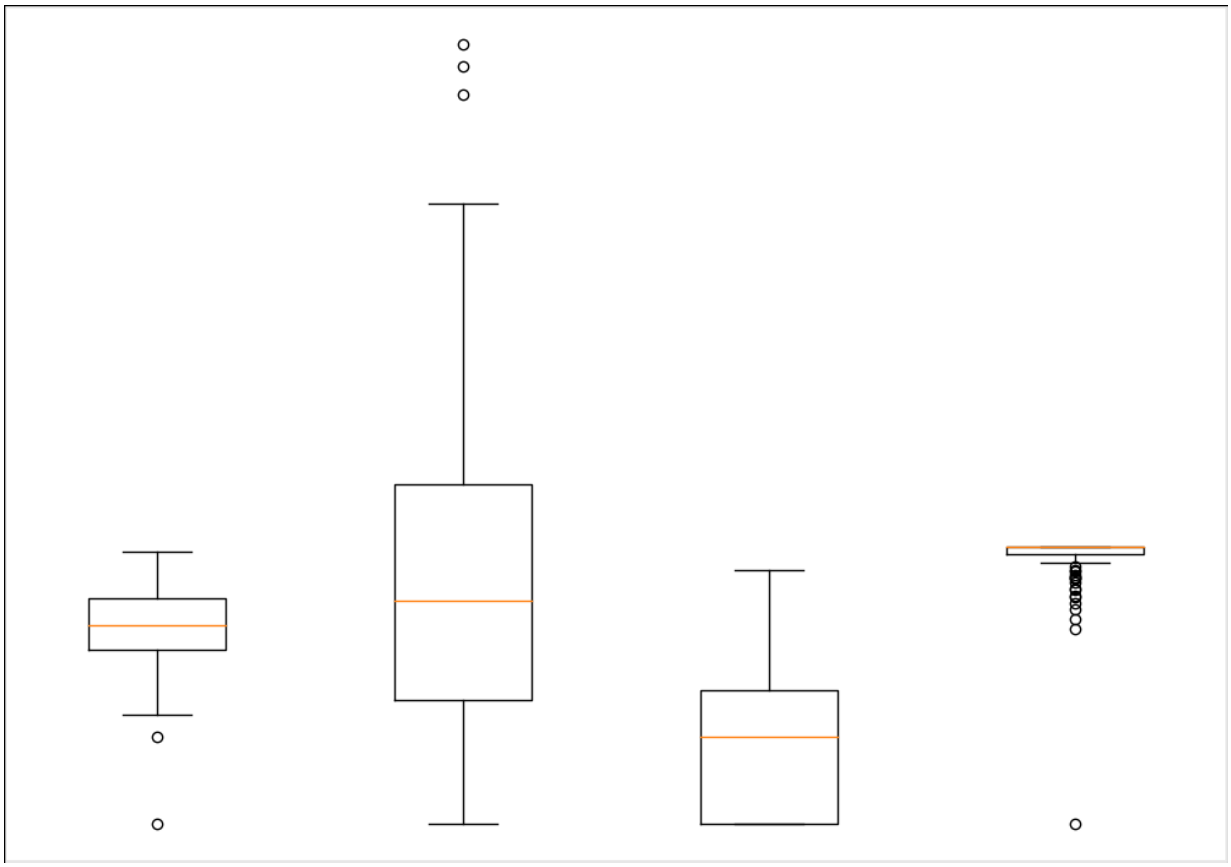
Diastolic Box Plot (Mean, Variance, Min, Max) - Same order of metrics for all box plots



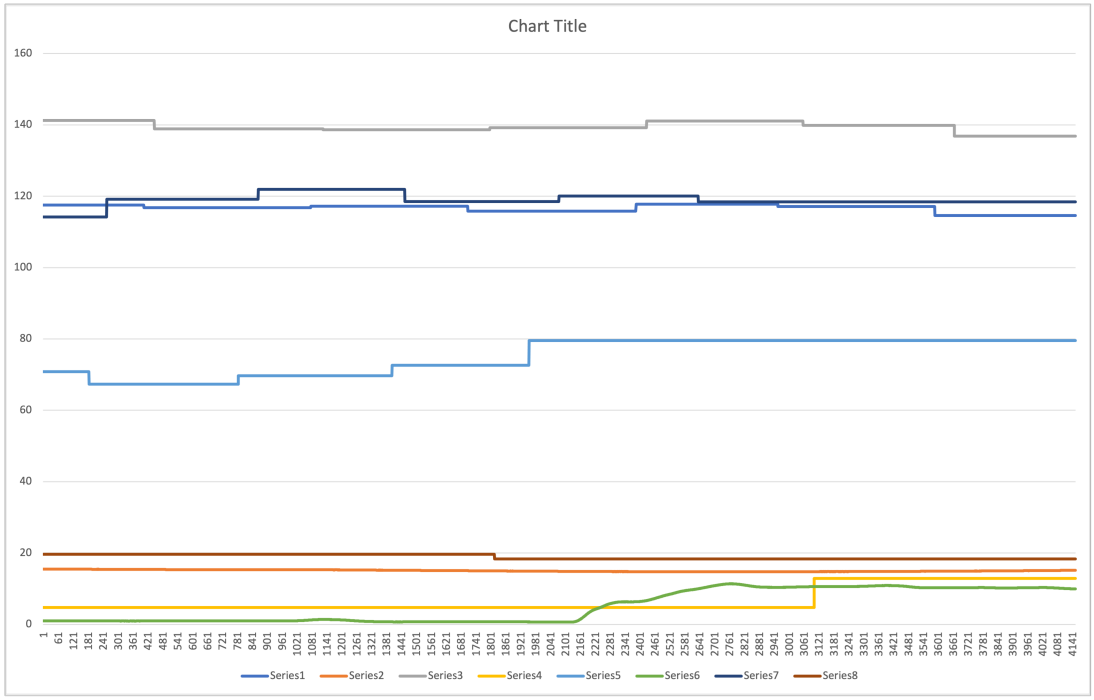
EDA Box Plot



Systolic Box Plot



Respiration Box Plot



Line Chart of Subject F027
No Pain (Dia, EDA, Sys, Res) Pain (Dia, EDA, Sys, Res)