Rehan Daya, Troy Walton, Yiwei Zhou, Emmanuella Eguche

Reducing Commercial Aviation Fatalities

Booz Allen Hamilton (BAH) is attempting to reduce aircraft fatalities caused by a loss of "airplane state awareness". This condition consists of numerous status effects such as drowsiness, distraction, being startled, etc. The train dataset is a compilation of three experiments applied to 18 pilots where electrocardiogram, electroencephalogram, and electrodermal recordings were collected to predict the probability of each event during these experiments. The three experiments are as follows:

1. **Channelized Attention (CA)** is being focused on one task, in this case, the subjects play an engaging puzzle-based video game.
2. **Diverted Attention (DA)** is having one's attention diverted by actions or thought processes, in this case, solving a math problem before returning to monitoring the task.
3. **Startle/Surprise (SS)** is caused by movie clip jump scares.

During experiment CA, one can experience event A or C. During experiment DA, one can experience event A or D and in experiment SS, event A or B. It helps to associate "event" with reaction to the experiment. These three experiments are split to allow for a controlled environment. If they were combined, these reactions would be read as noise. (https://www.kaggle.com/c/reducing-commercial-aviation-fatalities/data).

This dataset poses a personal satisfaction for us since it steps outside our comfort zone. We lack experience when dealing with biological data which we have found is noisy. There are an infinite number of things to account for when attempting to isolate a reaction during an experiment. A pilot may have stubbed his toe in the morning or gotten into a car crash on his way to the trial, either of which would throw off his results. Then we need to consider the effects of memories and emotional trauma that may have occurred in one's childhood but were blocked out. This can cause depression which results in a low EEG measure. It combines intuition and research when attempting to explain anomalies. However, there are certain variables that they could have easily added, such as age, gender, weight, and height, among other physical characteristics. Other than BAH, we expect neuroscientists and budding data analysts could benefit from the breakdown of this vast dataset. This data could be used to create a helmet for pilots that allows a model to predict signs of distractions, which may save lives one day.

While looking through the dataset, we attempted to learn where these EEG nodes were placed which can be seen in figure 1. We attempted to use this to predict which nodes should be lighting up during certain reactions since different parts of the brain take on different tasks. This would also show anomalies if the incorrect nodes were most prominent. There are four types of waves: Alpha, Beta, Theta, and Delta. It seems that your highest EEG occurs while you are awake. Low brain waves are related to being relaxed or "baseline", while high brain waves can be caused by paroxysmal surprise.

Data Preprocessing

We used KNN, logistic regression, and decision trees. The state of pilots was converted into binary: non-distracted (baseline, 0) and distracted (other three, 1).

Data Exploration

In figure 2, event A (baseline) was active most of the time. In figure 3, we can also see that each experiment was run for roughly the same amount of time, so there is no bias there. Figure 4 shows the amount of time the experiment specific event is occurring versus baseline. Keep in mind we assume each experiment's action occurred for the same amount of time though, it is against intuition, considering experiment CA event A occurred for .343% of the time and experiment DA event A occurred for 85.809% of the time. Figure 5 shows the total time in each seat for each experiment is similar. In figure 6, crew 1 had the least amount of time in these experiments. Additionally, considering the crew numbers are not 1-9, rather a max number of 13 is shown here, we assume BAH provided us a subset of data. BAH also took certain batches of the experiment and combined them which disrupted the continuity of the time column. Furthering our analysis from the prior graph, in figure 7, we can see that crew 1 had less time in experiment SS. The violin plot (figure 8) shows the distribution and relative amount of time each event occurred for each experiment. The plot on the left is for experiment CA, the middle is for DA, and right is for SS.

- In experiment CA, baseline event A occurs for 0.3431% of the time and event C occurs for 99.6568% of the time.
- In experiment DA, baseline event A occurs for 85.8098% of the time and event D occurs for 14.1901%
- In experiment SS, baseline event A occurs for 91.5779% of the time and event B occurs for 8.4221% of the time.

After breaking out the data into a large table we noticed when looking at the maximum mean for each column, that crew 1 had the highest EEG ratings. Since one of the indices is seat and we assume each member in the seat is swapped randomly, we are not sure if one or both pilots in this crew are displaying absence seizures. Absence seizures cause the person to be in daydreaming like trance for a couple of seconds but are unable to remember what happened. We originally thought it was someone with ADHD but that does not impact EEG ratings as much. Crew 1 was exhibiting higher mean

EEG ratings at baseline than other crew's EEG when getting jump scared. This could be an instance of "correlation doesn't equal causation".

When looking at the standard deviation table, we found the highest values pertained to the SS experiment. Jump scares result in an adrenaline rush causing increased blood flow to the brain, increased heart rate and causes the body to produce more sugar. All of this assists in increasing the focus and reaction speed of the individual, which coincides with a spike in EEG ratings.

From the 2-D scatter plot in figure 9, pilots' respiration rate and Galvanic Skin Response were used to show pilots tend to be distracted in two situations:

- Respiration sensor detects 600 microvolts or less.
- Respiration sensor detects higher than 800 microvolts and GSR sensor greater than 1500 microvolts.

When we introduce the subsequent 3-D scatter plot in figure 10, including seat as the third dimension, the previous conclusions are still valid. But given the seat difference, we see that seat = 1 (right) tends to distract the pilots. This corresponds to the results in the analysis.

<u>Models/Results</u>

Due to machine constraints, we were unable to produce robust models. Given the datasets totaled over 8gb and python utilizes only 1 core from the CPU, we were severely bottlenecked. Additionally, we ran into problems when attempting to predict probabilities to match Kaggle's desired output (we successfully did it with the Bagging Random Forest model). We worked around this by utilizing the train dataset as the whole, then split a new train and test portion from that. This allowed us to create a single column for predictions, which showed the event occurring.

Logistic regression

The accuracy of prediction without looking at the features is 62.62%. In the training set, data points abundance is ensured by 340171 distracting events and 393831 baseline events. In the testing set, the number of baseline events is 168800 with approximately 54% of the whole testing dataset. Including all the factors into the model, we have the following observations from the logistic regression:

- EEG_fp2 and respiration have the most negative weights. A higher rating of respiration and EEG_fp2 means the pilot is less likely to be distracted

- seat = 1 has high positive weight: pilots are more likely to be distracted when seat = 1 (right seat). This illustrates why the captain is always seated on the left.
- EEG_c4, EEG_p3, EEG_f8 have the most positive weightings among all the EEG readings at around 0.2%-0.25%. They are the most predictive factors of pilots' distractedness. We find that:
    - c4 is related to the inhibition of brain alpha waves (lack of concentration).
    - p3 node detects reactivity to external sensory stimulation.
    - f8 is correlated with frequent judgement control.
- Time and galvanic skin response also have relatively high weight. The more time into experiment and the more sweat from emotional fluctuation, the more distracted the pilots are.

KNN

16GB of RAM was not enough to create models where KNN>1, so we had to settle with a single model. Additionally, we settled with K-fold=3 to prevent our calculation times from exceeding 45 minutes. This resulted in an accuracy of 98.7% out-of-sample.

Tree Models

All the predictors, except the categorical variable differentiating each crew, were used in these models. Initially, we created tree models with a maximum of 2 to 3 branches. These limited models, however, had training and testing accuracies of around 61% to 64%. A bagging method was also attempted with models limited to a depth of 2-3 branches, but it did not show any noticeable improvements. At the risk of creating an over-fitted model, we developed a tree model with no depth limits. This model ended up with a training accuracy of 100%, which was not unexpected. However, this model also gave a testing accuracy of 99.96%, which was surprisingly good. This could be because the predictors included in developing this model were sufficient enough to represent the underlying relationships and principles that define, or at least are correlated with, a pilot’s state of mind.

From all the tree models, we found that certain predictors were the best to split the data as they led to the greatest drops in entropy, which represent the purest splits of the data. These predictors include:

- Electrocardiogram Signal (ECR)
- Galvanic Skin Response (GSR)

With ECR being the first predictor to split the data. To give a sense of the size and scope of this model, we included a low-resolution image of the model as figure 15.

Appendix

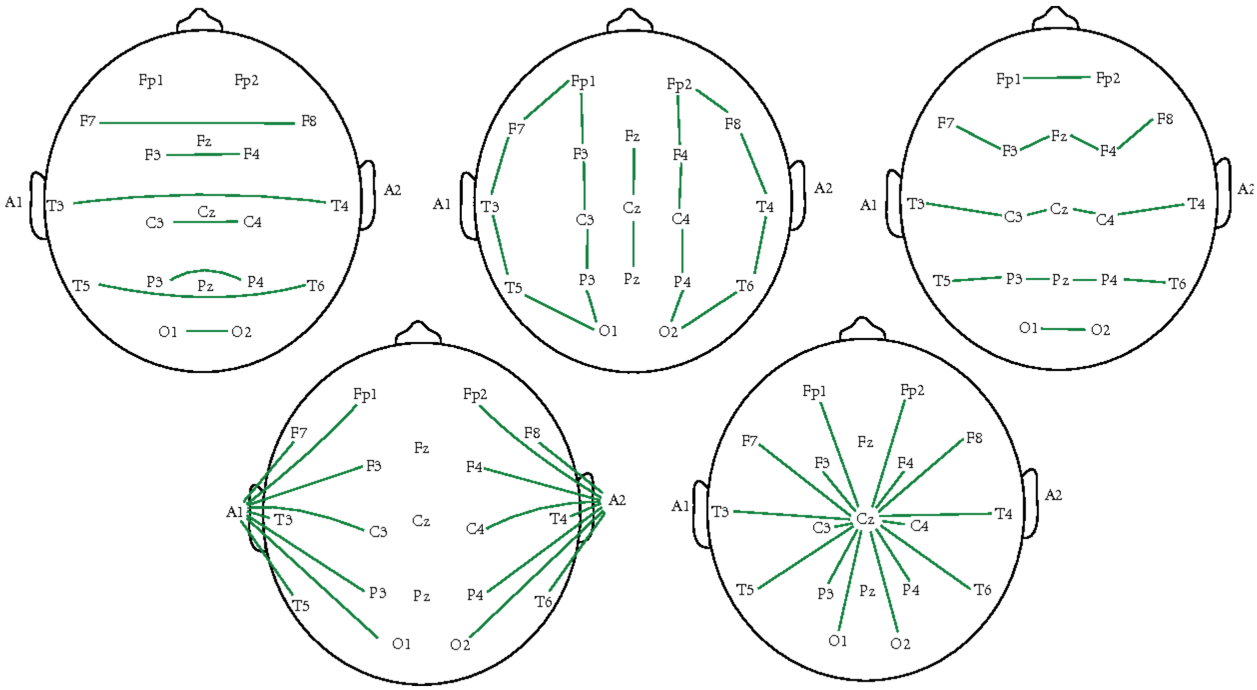

Fig 1. Location of EEG nodes

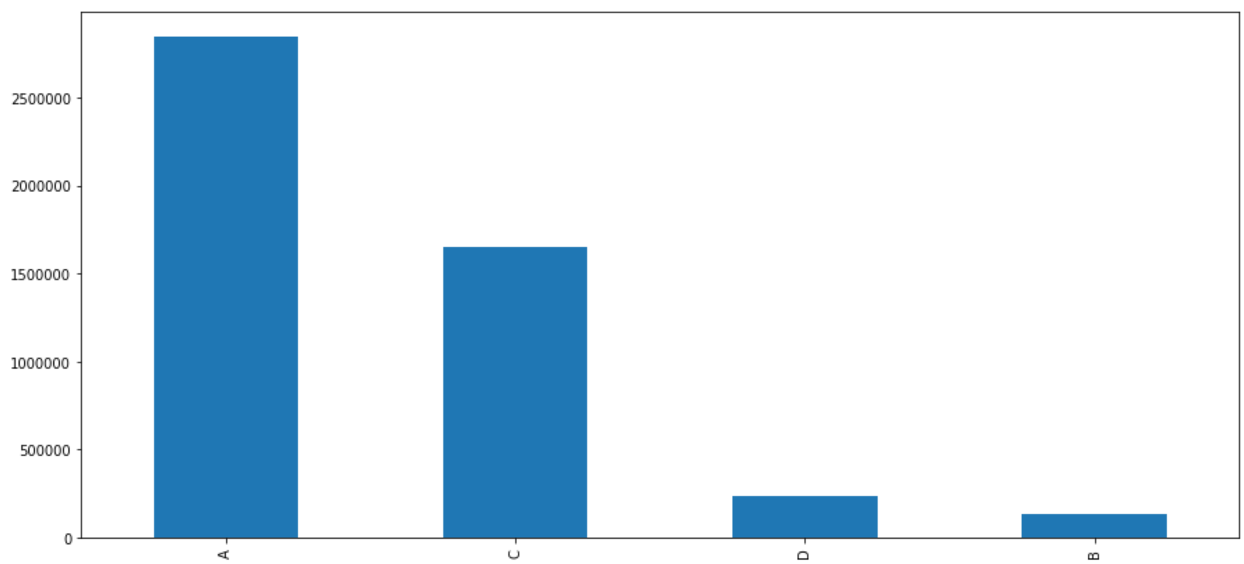

Fig 2. Number of instances of each event, without separating by experiment

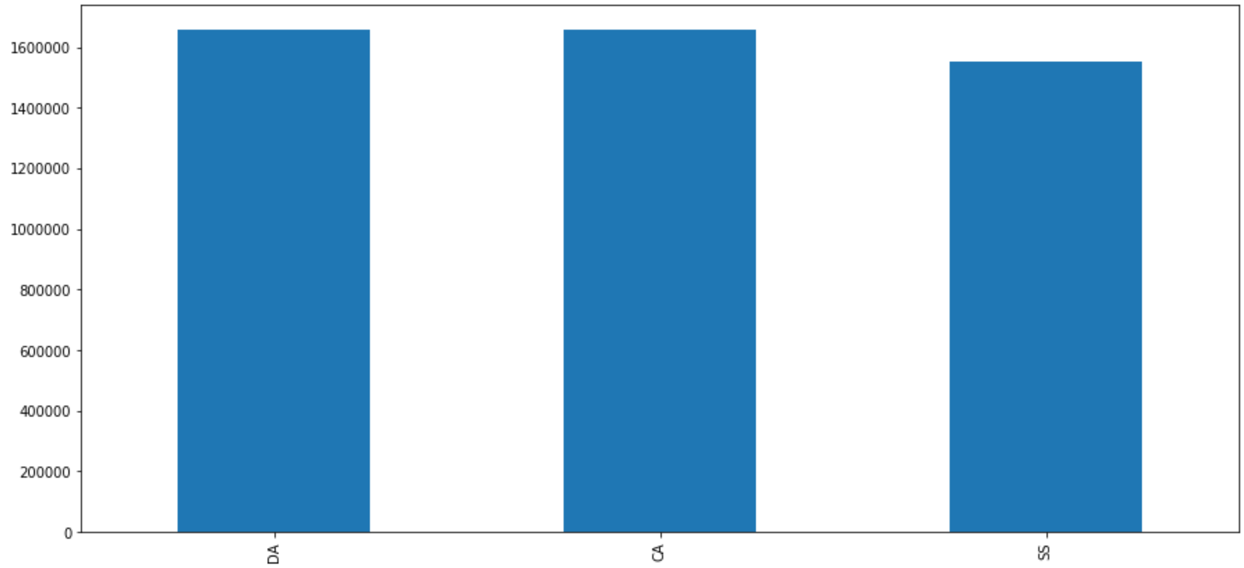

Fig 3. Number of instances each experiment ran for

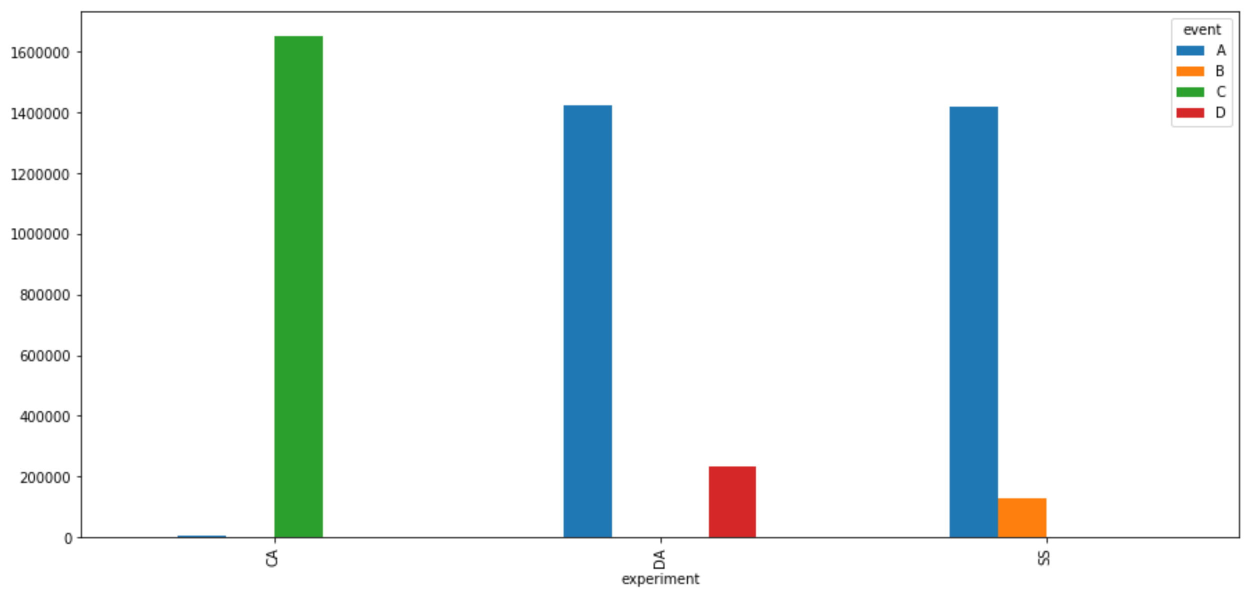

Fig 4. Number of instances of events, broken down by experiment.

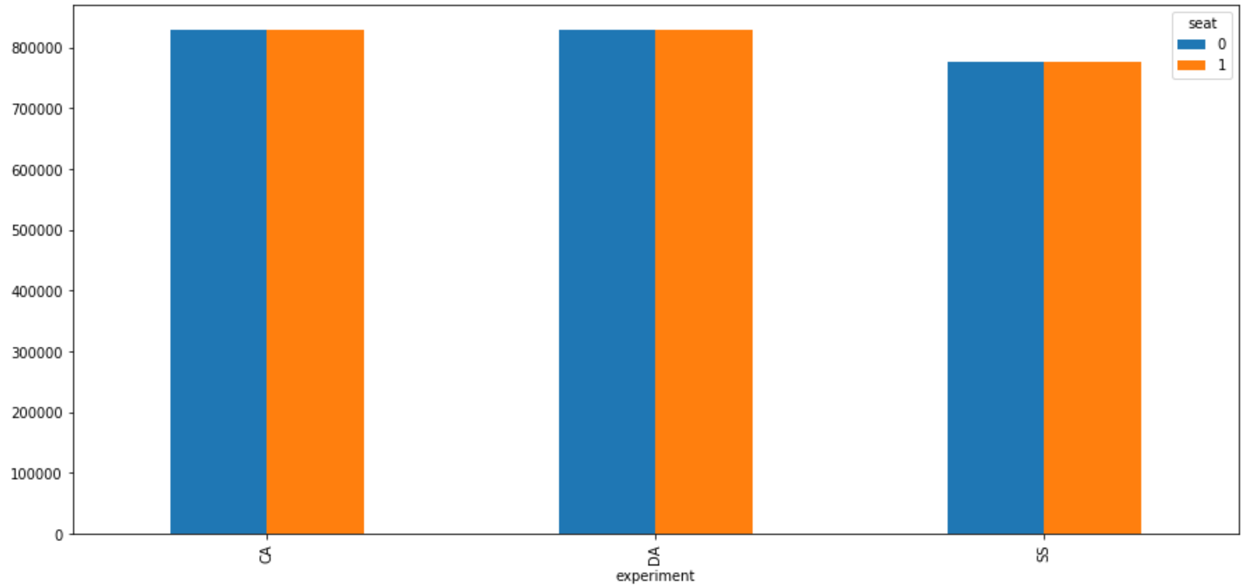

Fig 5. Number of instances of seat variable broken down by experiment

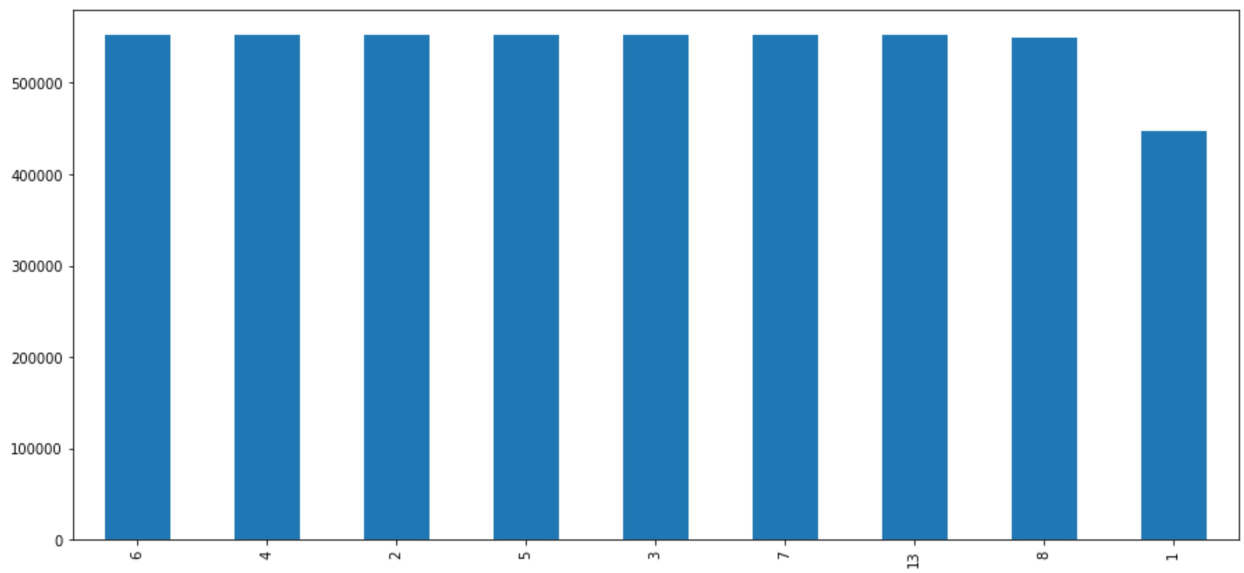

Fig 6. Number of instances of crew variable, not broken down by experiment

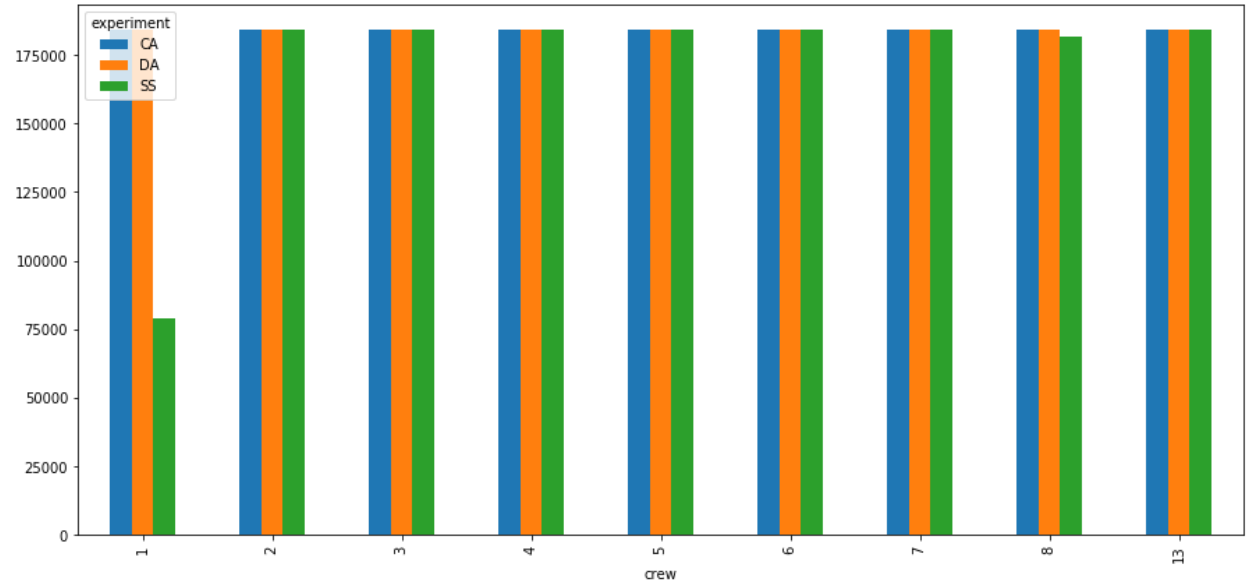

Fig 7. Number of instances of crew variable, broken down by experiment

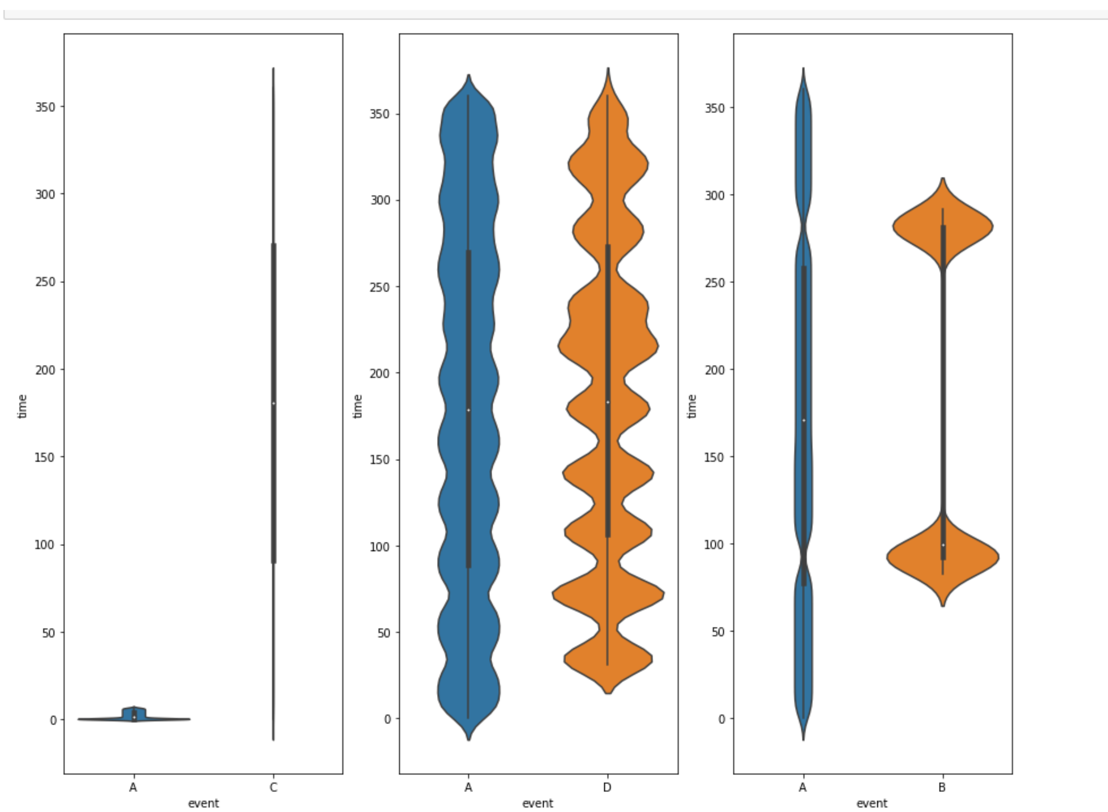

Fig 8. Event grouped by experiment with values utilizing time. This is slightly inaccurate due to BAH disrupting time column continuum

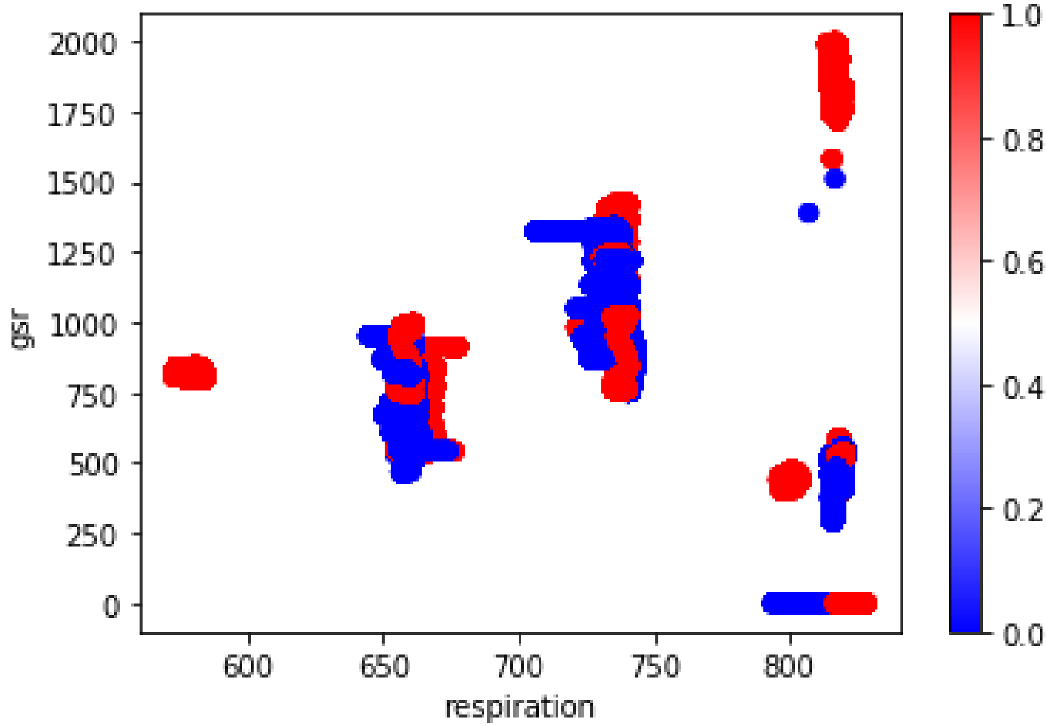

Fig 9. Relationship between GSR and respiration

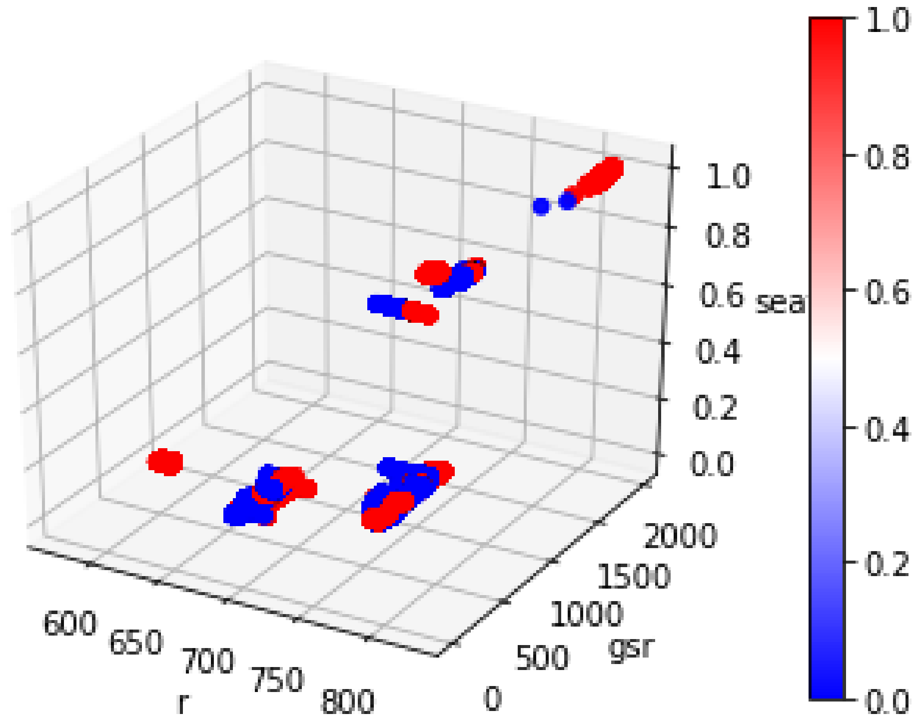

Fig 10. Relationship between R, GSR, and seat

	weight
seat	1.921266
eeg_f8	0.002541
eeg_p3	0.002463
eeg_c4	0.002124
time	0.002060
gsr	0.001774
eeg_t4	0.001755
eeg_c3	0.001397
eeg_fz	0.000845
eeg_f7	0.000822
eeg_fp1	0.000617
eeg_t5	0.000329
eeg_poz	0.000111
ecg	-0.000039
eeg_cz	-0.000069
eeg_t3	-0.000293
eeg_pz	-0.000551
eeg_f4	-0.000671
eeg_f3	-0.000868
eeg_o2	-0.000992
eeg_t6	-0.001248
eeg_o1	-0.001390
eeg_p4	-0.001414
r	-0.003680
eeg_fp2	-0.004827

Fig 11. Data frame with model factors and weights

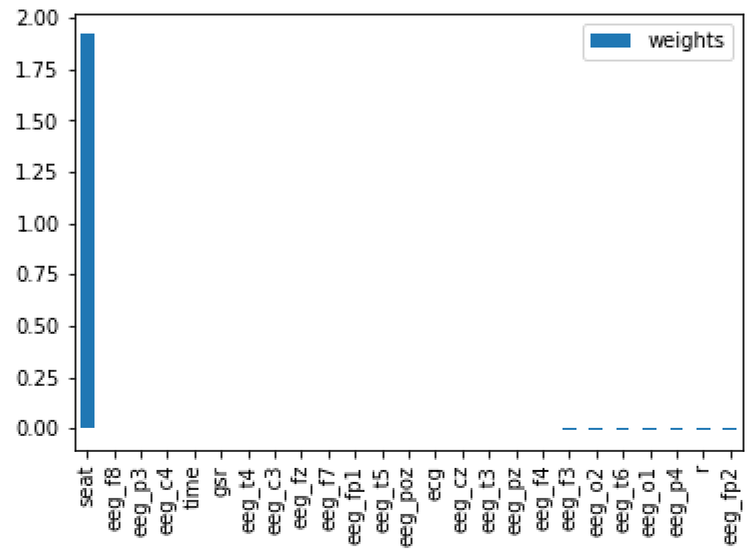

Fig 12. Logistic Regression: Factor weight bar plot including "seat"

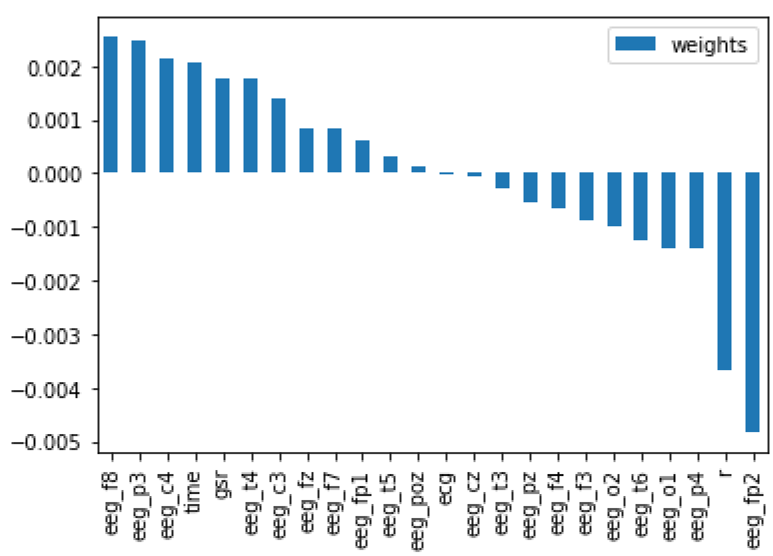

Fig 13. Logistic Regression: Factor weight bar plot excluding "seat"

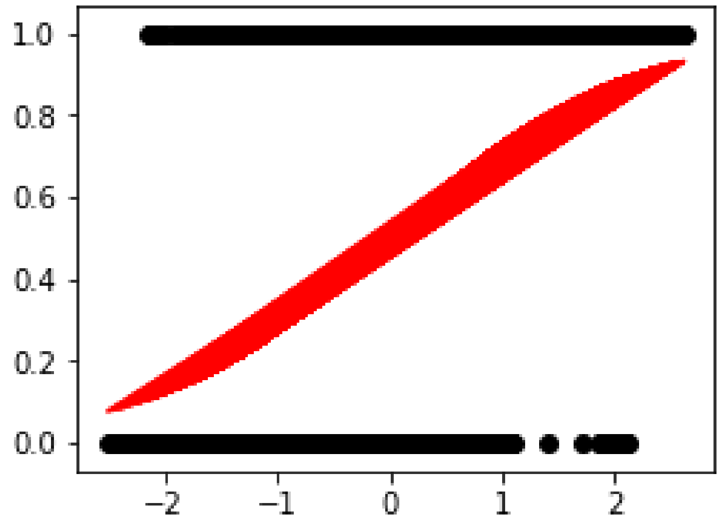

Fig 14. Logistic Regression: Sigmoid function

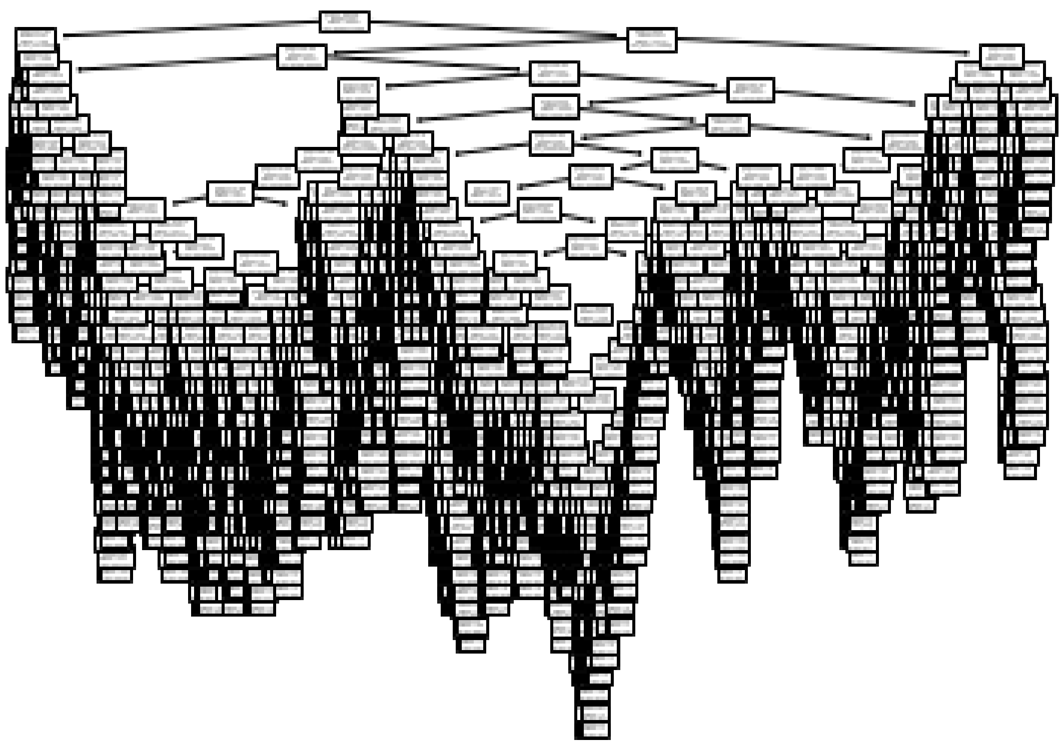

Fig 15. The graph above is the actual final tree model but is not legible due to necessary reductions to the file size to get it from Python to this document.