CS109a Milestone #3: Predicting Seizures

Fernando Becerra, Jessica Cao, Yi Ding, Andrew Mayo

November 5, 2016

1 Introduction

For our CS109a project, we are studying seizures. The goal is build a model that can determine whether an individual will suffer from seizures given the individual's background, medical history, demographic, etc. In order to build a model that can accomplish this task as accurately as possible for any individual, we will rely on a number of different data sets. We use the Thall and Vail (1990) study, which tracks 59 epileptics over the course of 16 weeks (8 before a drug trial, 8 after). We also make use of the 1958 National Child Development Study (a.k.a. the 1958 Birth Cohort Study), which tracks every child born in the UK during a single week in 1958. Lastly, we are considering making use of EEG data taken for control and epileptic groups. Note that we may add additional data sets as the project progresses. Below we present a number of figures to explore the data that we will be working with to build our model.

2 Data Exploration

2.1 Thall and Vail (1990)

Figures 1-4 are the results of data exploration for the Thall and Vail (1990) data. From Fig. 1, we can see that there does not appear to be a strong variation over time in the number of seizures an individuals suffers from during the administration of prograbide (the drug being studied). In Fig. 2, we can see there is a wide distribution of ages for the group, ranging from 18-42 years of age. In Fig. 3, we can see how many seizures are suffered over 8 weeks prior to drug administration compared to an 8 week period during drug administration. Fig. 4 is a nice boxplot comparison of the control and test groups before and after drug administration (the groups should have the same distribution in the left plot, while the distribution should be different in the right plot if the drug is effective).

It should be noted that none of these three plots distinguish the control group from the test group. The effectiveness of the drug is not the goal here; we are merely hoping to develop a picture of the test subjects and their typical seizure patterns.

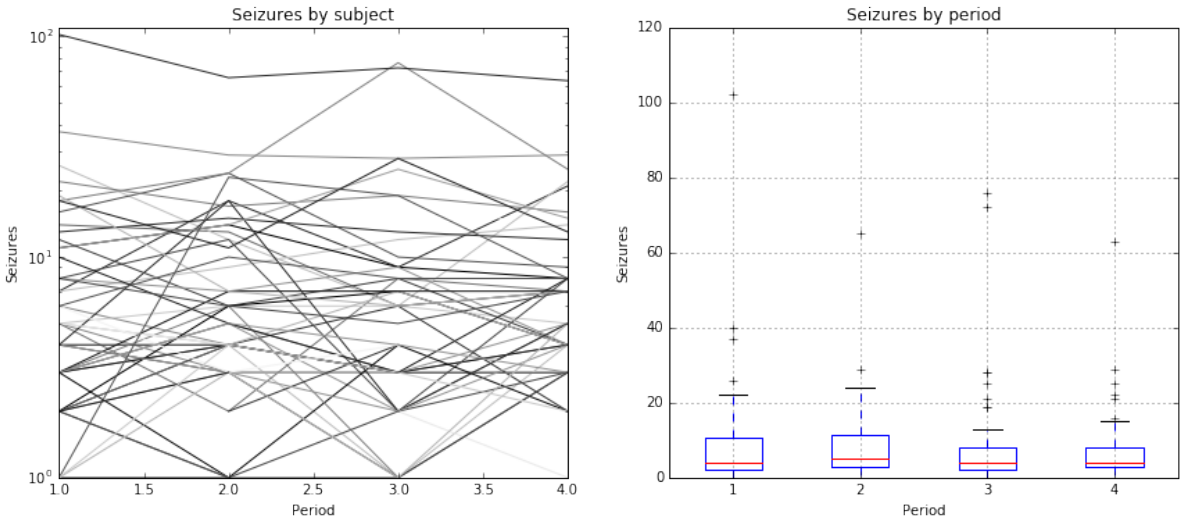

Figure 1: Left: the number of seizures for each patient is plotted for each of 4 2-week periods. Right: a boxcar plot of the number of seizures for the whole group for each two week period during drug administration.

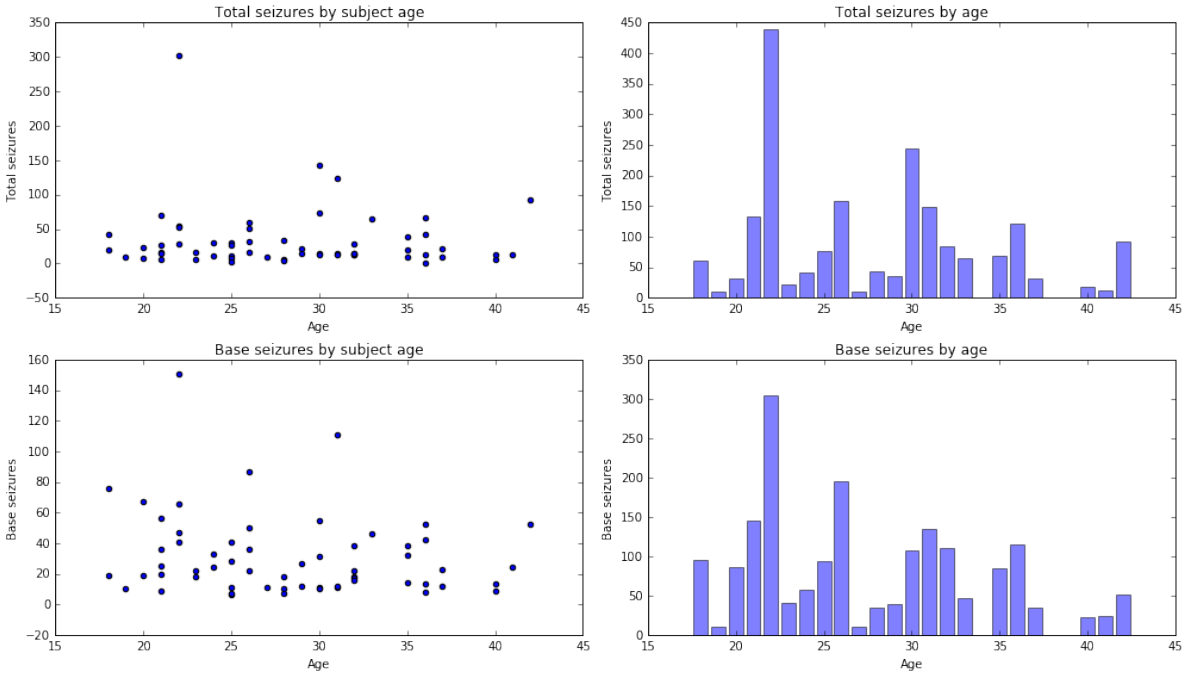

Figure 2: Top: The number of seizures in the final 8 week period of the Thall and Vail (1990) study as a function of age (left is a scatter plot where each point is an individual, whereas right is a histogram for better visualization of overall trends). Bottom: The number of seizures in the first 8 week period of the study as a function of age (prior to the drug trial). Left and right are again a scatter plot and histogram, respectively.

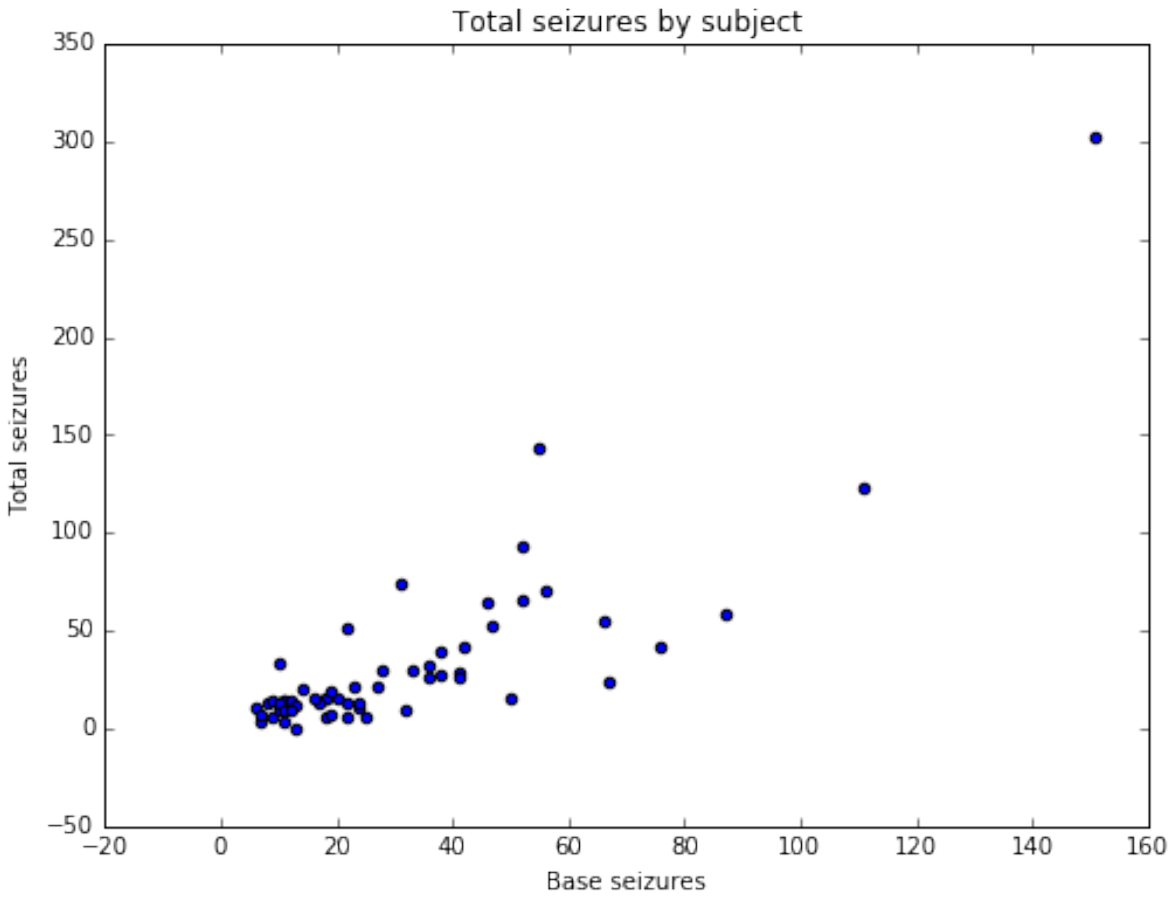

Figure 3: This figure shows the number of seizures individuals suffer in the 8 weeks during administration of the drug as a function of number of seizures prior to drug administration (first 8 weeks).

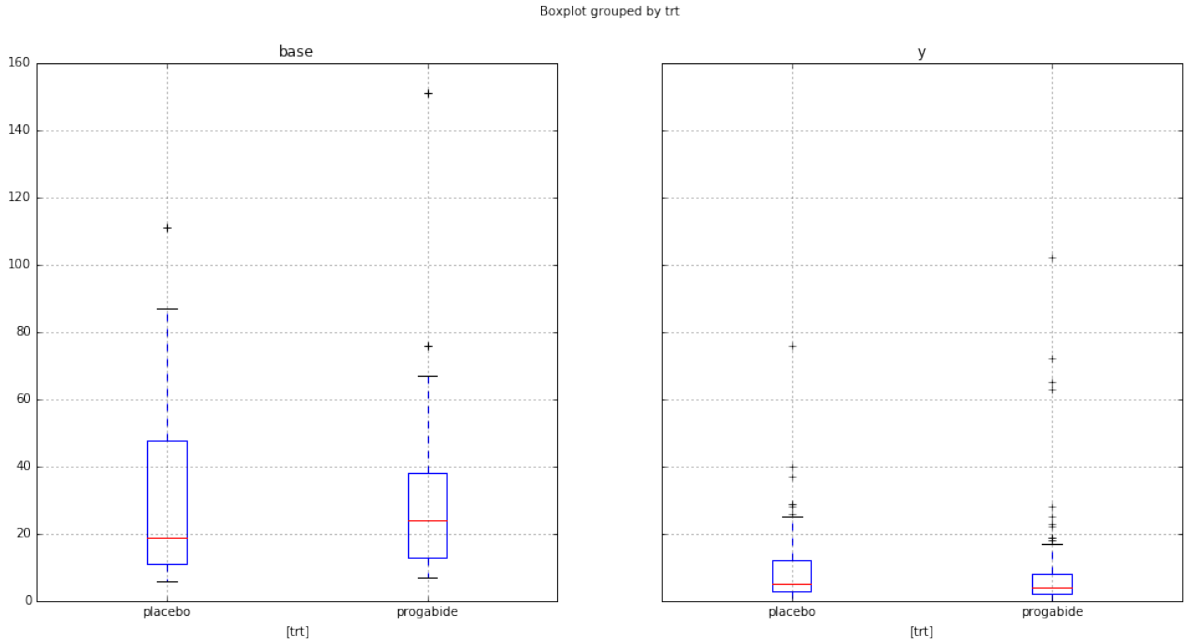

Figure 4: Both the left and right panel are boxplots of the number of seizures suffered for each individual depending on whether the individual was in the control or test group. Left: a comparison of groups with the baseline number of seizures (first 8 weeks); because this is prior to drug administration, the two groups are expected to be identically distributed. Right: a comparison of the two groups during drug administration (last 8 weeks). If the drug is effective, the two groups should not have the same distribution: the distribution of the progabide group should tend to smaller number of seizures if the drug helps prevent/reduce seizures.

2.2 1958 Birth Cohort Study

Figs. 5-6 come from data collected through the 1958 Birth Cohort Study. The data set is quite massive, so we are currently only looking at a very small subset of the data. We may include more data later if it seems relevant.

Fig. 5 examines the relationship between seizures and alcohol intake frequency. Interestingly, individuals who do not drink (though may have in the past) appear to be at a greater risk for seizures. It is unlikely that drinking prevents seizures; rather, individuals prone to seizures may choose not to drink because of an increased risk of seizures, potentially life-threatening seizures, or a risk or mixing alcohol with medication to treat seizures. Of course, these are just hypotheses; we will be reading studies related to this data and analyzing it in more detail in the future.

In Fig. 6, a relationship between risk of seizures and headache severity is

examined. Risk of seizure increases significantly with the severity of headache.

Figs. 5-6 both point to very interesting features we may want to investigate. Is there predictive power for seizures in alcohol intake frequency and headache severity? Now that we have this data at our disposal, we can answer those questions and build our model based on the answers.

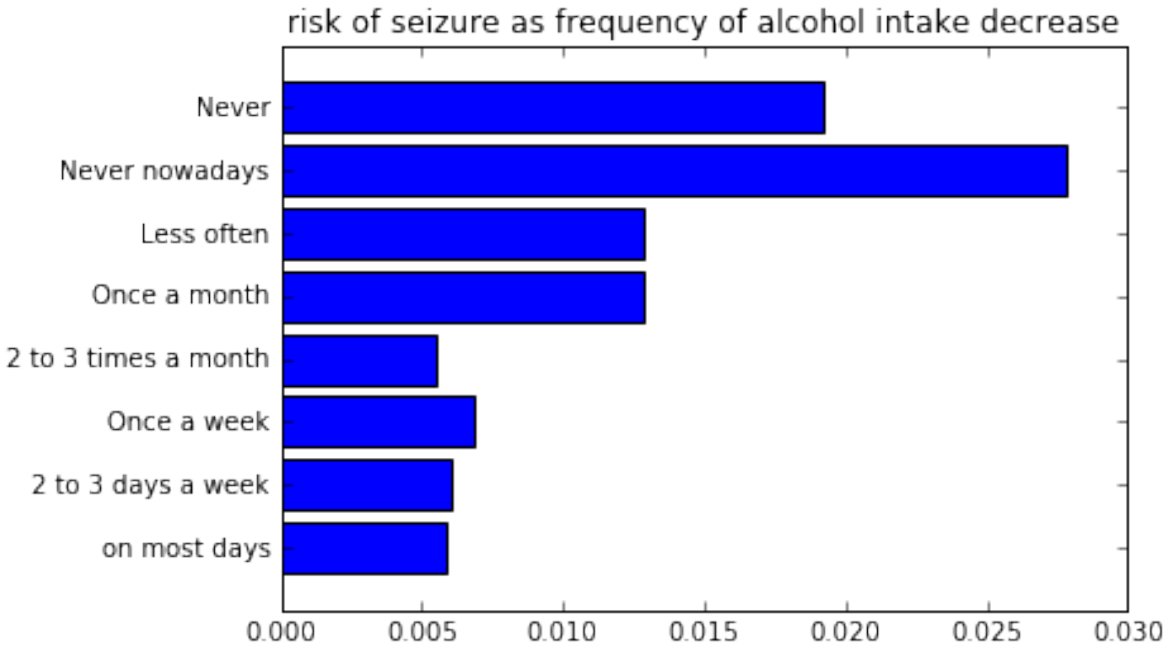

Figure 5: A bar plot of the risk of seizure for a given frequency of alchohol intake. Higher intake of alcohol appears to be directly correlated with a greater risk of seizure.

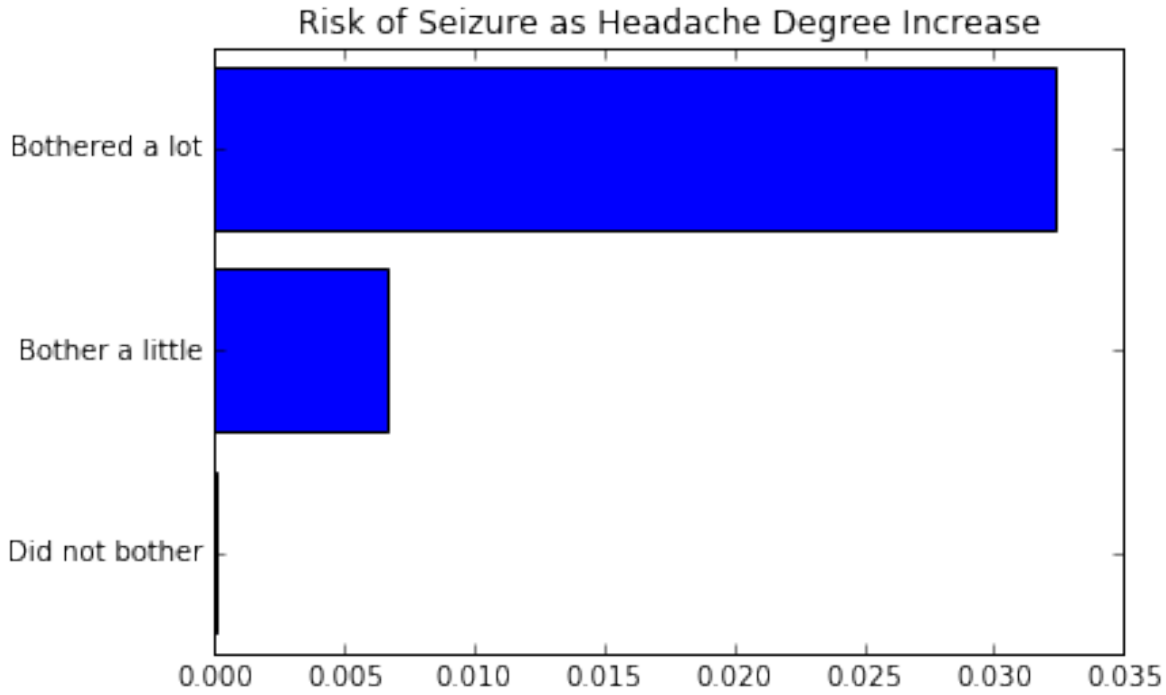

Figure 6: A bar plot of the risk of seizure for a given level of headache severity. There is an obvious direct correlation between the severity of headache and the risk of seizure.

2.3 EEG data

Figs. 7 and 8 are of EEG data. They compare the brain activity of someone suffering from a seizure against someone who is not. There is a significant spike in brain activity during seizure. It will be interesting to investigate whether there are signs in the EEG data before or after a seizure of evidence that a seizure will/did occur. It would be even more interesting to investigate whether there are any differences in long-term brain activity between individuals who suffer seizures and those who do not (there probably isn't, but it's worth a look!).

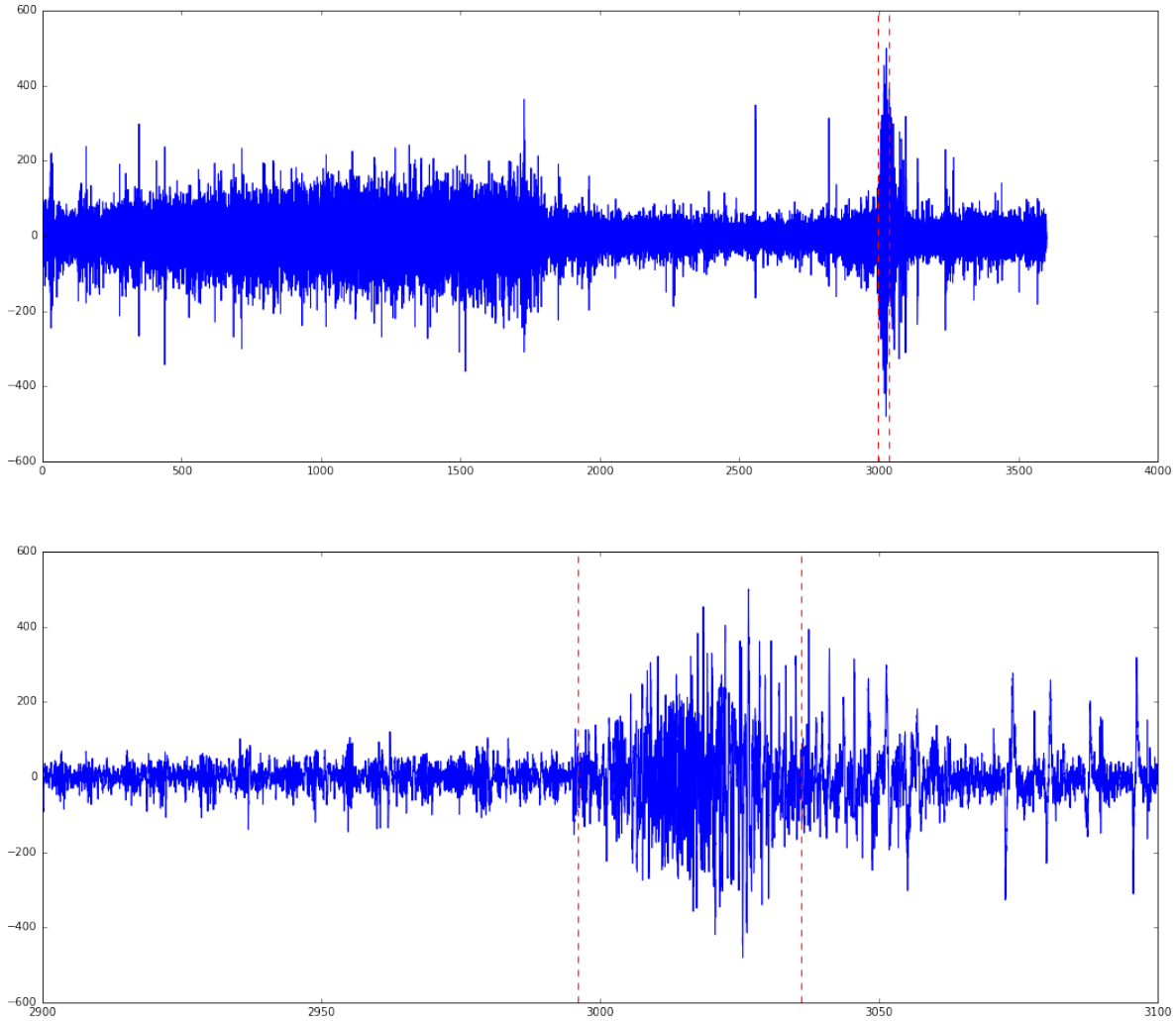

Figure 7: A graph of EEG data for a patient suffering a seizure. There appears to be a spike in brain activity during the time of the seizure (between the dotted red lines). The bottom panel is a zoomed in section of the top panel.

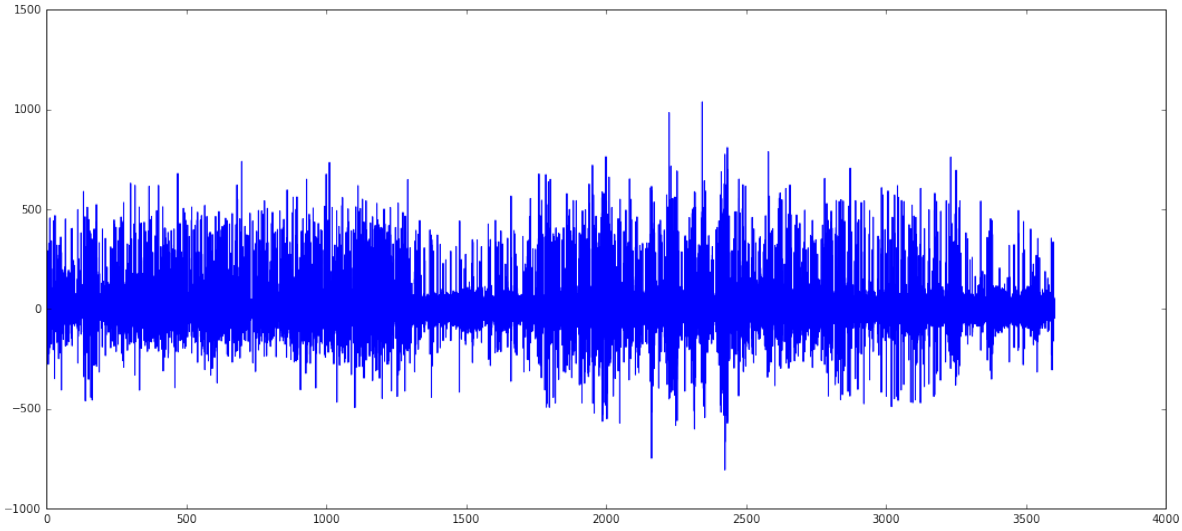

Figure 8: A graph of EEG data for a patient with no seizure. There are no spikes in brain activity as large as the one seen during a seizure in Fig. 7.

3 Conclusion

We hope to use the data sets explored here to develop a model to predict seizures. The data from the Thall and Vail (1990) study may be useful in giving us a (hopefully) unbiased sample of those suffering from seizures from which we can look for commonalities. The results of the drug trial itself will likely not be as useful, but we will still into the results of the trial to be certain. The 1958 Birth Cohort Study will be very useful to us as we can compare samples of thousands of individuals and look for correlations between seizures and dozens of other factors, including medical history, socioeconomic status, age, gender, etc. Lastly, the EEG data gives us an additional way to look at seizures and could be a very intriguing additional source of information. These data sets and any others we choose to include will be very helpful to us as we develop our predictive model for seizures.