

---

# EE / CPRE / SE 491 - Weekly Report 05

11/5/2019 – 11/19/2019

**Group number:** sdmay20-41  
**Project title:** Machine Learning for Understanding Aging  
**Client &/Advisor:** Dr. Julie Dickerson  
**Team Members/Role:**

- Ian Simon / Chief Engineer
- Jacob Laing / Chief Engineer
- Nathan Carter / Test Engineer
- Samantha Williams / Meeting Scribe
- Scott Rose / Meeting Facilitator
- Aria Sheets / Report Manager

---

## Weekly Summary:

- This week, we spent time together in order to properly insert the TSV data into our database. This required a lot of effort from all of us, which is why we came together. We searched online for tools that could help us, and we found a website that properly converts TSV files into MySQL statements that we could execute. We also used SQL statements in order to remove users who were not present in all three of the data files we used. This is the first kind of filtering that we will be using for our data, and it allows us to only use data that is relevant.

## Past Week Accomplishments:

- Ian Simon: This week I worked on getting certified through IRB. This organization regulates and monitors practices used in collecting data in research studies. This is relevant to our project since we are dealing with data collected from human studies.
- Jacob Laing: This week I worked on continuing to set up the vm so that we can use MySQL workbench to store our data. Along with that I worked on the CTI training that was provided to us from Dr. Dickerson.
- Nathan Carter: This week I continued work on sample data sets using plotly. I also worked on parsing tsv files and putting the data into our database. I also began and completed the IRB training on the CITI website.
- Samantha Williams: This week I aided in the parsing of tsv files into our database. We then used a join to find the users that were included in all three files. I also continued working on figuring out how to parse DDI files, but didn't have much success. I also began the CTI training course that Dr. Dickerson requested we do.
- Scott Rose: My main task was creating a script that parsed the data we were given. Based on the data, the script would create SQL tables to store the data in a server. To get this script to run on the team's virtual machine, I had to do some configuration on the vm. This configuration included installing the java jdk and jre on the machine.

- Aria Sheets: I worked on trying to find a resource that we could utilize to easily convert TSV files into MySQL statements that we could execute to put the data on our database. I also did some The tool was perfect for the job with the exceptions of some hiccups while trying to convert. After multiple attempts of converting, it finally worked. We assume the issue was due to the large amount of columns in our file. I also finished the CITI Program training for the Social/Behavioral Research Course. This course helped me understand how I should be careful when using data about humans and their bodies.

**Pending Issues:**

- Ian Simon: I will be completing the IBM training this upcoming week, and I will have to help with reading the data we need into our database.
- Jacob Laing: Still having issues connecting to the MySQL database from workbench.
- Nathan Carter: There is an issue with figuring out how exactly to get our data into the database. We are mostly having issue with formatting and missing data.
- Samantha Williams: Still need to figure out if there is a way to parse DDI files so that we know how to read the values that are present in the tables. Most of the entries have different scales that they are based off of, so it is important for us to know these values without having to manually comb through the files. I also need to complete the CITI training course that I began earlier this week.
- Scott Rose: The SQL tables were too big to be used. As a result, we will be parsing the data we have and eliminating anything that isn't necessary. This should cut down our data size to something that is able to be put into an SQL database.
- Aria Sheets: I'm still working on trying to understand how our data will be formatted and then input into the database. While we did insert the data, we know that not all of it will be needed, which is why I'm still trying to figure out how it will be formatted, and what data we want to keep and discard.

**Individual Contributions:**

<b><u>Name</u></b>	<b><u>Individual Contributions</u></b>	<b><u>Hours this week</u></b>	<b><u>Hours Cumulative</u></b>
Ian Simon	Worked on IRB training and group assignments	2	27
Jacob Laing	This week I worked on continuing to set up the vm so that we can use MySQL workbench to store our data. Along with that I worked on the CTI training that was provided to us from Dr. Dickerson.	4	28
Nathan Carter	Completed IRB training and worked on reading data into the database.	4	31

Samantha Williams	I aided in parsing the data, but did not have a super hands-on role this week on the group work. To compensate, I did more of the heavy work on the youtube video assignment.	2	27
Scott Rose	I created a script that parses our tsv data and turns it into and MySQL tables. I also did some configuring on the team vm so that it would have the Java 8 JDK and JRE.	6	31
Aria Sheets	I helped in finding a tool that would allow us to parse the TSV data into MySQL statements. I also finished the CITI course that helped me understand what we need to consider when using data about humans.	5	35

**Comments and Extended Discussion:**

- We spent some extra time refining our design document this week. We also ran into some issues dealing with column limits in MySQL, but we're working on a solution currently.

**Plans for the Upcoming Week:**

- Ian Simon: This upcoming week I plan to help with the data reading into SQL and working with the TSV files.
- Jacob Laing: I will continue working on finishing the setup of the VM so that we can start easily storing data on our database.
- Nathan Carter: This week I will continue working on reading the data into the database. I will also continue to work with plotly. Work on plotly was put on hold while we figured out our database and data. I will also begin to work on filtering the data in our database based on our client wants.
- Samantha Williams: This week we are going to try to see which IDs are present in all three files and create a new database table with these IDs and columns we are interested in. This may involve creating our own tsv files with the data we want. We will also look into different statistical variables that Dr. Dickerson mentioned.
- Scott Rose: I will be making a script that eliminates all unnecessary columns in our data. The formatted data will be put into a new tsv file.
- Aria Sheets: This week I will work on processing the data for columns that we care about. I will also help in taking this formatted data and inserting it into our MySQL database.