

# Building a Data Analysis and Visualization Web App for Mitronite INC.

Christopher Eeles

February 21, 2019

## 1 Personnel:

### 1.1 Roles

As a group, we decided that it is important to ensure each member can independently create a functional full stack web application using the R Shiny framework. Thus, we have agreed to complete a brief prototype application independently, at which point we will reconvene to compare our results, and incorporate the best ideas into the working version of our software.

Quality control and project management responsibilities will be shared between members. We will be enabling development collaboration and version control via BitBucket while implementing an AGILE development-cycle using the online SCRUM tool Trello for project management and task assignment. Ultimately we will develop a professional quality web application which maximizes utility to our project sponsor, Mitronite Inc, while also providing a meaningful and industry relevant learning experience for all team members.

Given the scope of this project, we have also decided to take on specialized software development roles to ensure that we can bring the appropriate depth of knowledge to the table for each of the application's layers. While we will all play a role in development of the database, back end, front end, statistical analysis and visualization, the sub-specialties of each group member are as follows:

#### Minoru Nakano

Interactive UI/UX/visualization Developer:

Minoru has the strongest background in professional web development. He has extensive experience with front-end UI technologies such as HTML, CSS, and Javascript. This makes him the most suitable team member to take on the specialization of 'Interactive

UI/UX/visualization Developer'. The responsibilities of this role are the construction of the front-end UI for the app using technologies such as HTML and built-in R Shiny functions. Other responsibilities include working with the data from the back-end, ensuring that it is formatted properly to work with the visualizations generated by the application framework.

#### Peter Vlasveld

Mechanical Back-end Developer:

Peter has a strong background with back-end programming languages such as Java, C# and SQL. This makes him the most suitable team member for the specialization of 'Mechanical Back-end Developer'. Responsibilities for this role include ensuring incoming data from .csv file is collected and placed into the relational database properly by the application. Other responsibilities include ensuring that data is properly collected from the relational database, and sent to the front-end for visualization. This role also emphasizes maintenance of data within the MySQL database; ensuring data integrity and security. Maintenance of the database will however be a common responsibility among all team members.

#### Christopher Eeles

Biological Expert/Statistical Analyst:

Chris has a strong background in the field of Biology as well as extensive statistical knowledge with technologies such as R. This makes him the most suitable team member for the specialization of 'Biological Expert/Statistical Analyst'. Responsibilities for this role include biological interpretation of the data using prior biological background, and statistical analysis. Other responsibilities include back-end statistical analysis of the relational data using technologies such

as R. This role will also focus on communication of biological interpretations to fellow team members, the PI, and our collaborators (Mitronite Inc.).

## 1.2 Sponsor

Our principle investigator (PI) is Dr. Frank Merante. Dr. Merante is a Lecturer at Seneca College as well as a Biotechnology Industrial and Academic Scientist who has worked for many biotechnology companies. He now teaches at Seneca College and works on collaborative research projects with outside clients. Our team is based out of Seneca, so Frank is easily accessible to us when we need him, however we do plan to schedule formal meetings with Frank and Eaton (The CEO of Mitronite, our collaboration partner on the project) at least once a month to discuss the progress of the project as well as provide them with an updated prototype at the end of a 3-4 week AGILE sprint.

## 2 Definition of Project:

The aim of this project is to provide a web application to the client (Mitronite Inc.) that allows them to input raw ICP data in the form of a .csv file into the application, which will then rapidly store, and visualize the data using a suitable web framework. The main goal is to produce informative visualizations that give the client insight into the data that would not have been able to be had without such a technology in place. Another goal is to store the data in such a way that it has proper integrity and security, as well as being easily accessible for analysis. The end deliverable should be easy to use, allow rapid access to the data, and on top of informative visualizations, allow the client to download data-sets back out as either .csv or .xlsx formats.

### 2.1 Background

In this project we will be assisting Mitronite Inc. and our PI, Dr. Frank Merante, by developing and implementing a full stack application for data cleaning, storage, annotation, analysis and visualization through an interactive, web browser based UI. The research team is interested in studying changes in the mineral content of sweat in a time-series over the course of exercise. This data will be used to infer the rates of mineral depletion in muscle tissue and surrounding extra-cellular fluid, in hopes of developing a product able to reduce or eliminate these losses and thereby increase exercise performance and recovery rate.

Given the low concentrations of these trace minerals in sweat, it was necessary to employ high accuracy ICP-SE technology—originally developed for archaeological analyses—to gather data for this study. The ICP-SE instrument allows for high-throughput mineral quantification, generating large volumes of data which need to be curated and interpreted to draw useful conclusions about the physiological impacts of mineral depletion and the efficacy of potential interventions.

Currently this data is being output as .csv files, which are manually curated, stored and visualized in Excel. With the constant influx of new data, such processes take significant time and resources away from more productive research and development activities and therefore constitute a considerable cost to the company.

Through our application, we will be able to remove this data processing bottleneck; thus, enabling research and development limited only by the rate of sample collection. Moreover, by embedding data visualization tools coupled to statistical metrics within our software we will facilitate rapid data interpretation; Thereby empowering researchers and decision makers with the information they need to efficiently direct research while expediting the the development and release of Mitronite’s product to market.

### 2.2 Goal

Our goal is to automate and streamline the process of data input, curation, storage, annotation and analysis while developing visualization tools which allow rapid interpretation of biological data. In doing so we will provide Mitronite with the information necessary to efficiently allocate resources in the development of an effective solution to mineral depletion in athletes and active lifestyle enthusiasts alike.

Ideally our product will remove the technical and administrative burdens of data analysis, freeing up Mitronite’s research team to draw meaningful conclusions from the mineral concentration data. This information can then be used to formulate and implement experimental interventions to find correlation between mineral depletion and the physiology of muscle cells. Such studies will contribute to their goal of providing evidence based nutraceutical solutions to health and exercise problems in both the consumer and professional health and wellness markets.

### 2.3 Scope

The scope of this project is to produce a functional full-stack web-based application that performs the following tasks:

- Storing ICP data from a CSV file into a relational database.
- Retrieving the ICP data from the database based on a user input.
- Performing statistical analyses on the retrieved data set.
- Visualizing the resulting data on a web browser.

The end-point of this project is delivery of the fully functional web application that accomplishes the tasks outlined above.

## 2.4 Target audience

The target audiences of this project are the PI (Dr. Frank Merante), and the project sponsor (Mitronite Inc.), as the finished web application will be used by these two parties. The web application will assist our audiences in the areas of data storage, manipulation, analysis and visualization.

## 2.5 Skills

The essential skills required for this project are as follows:

### 2.5.1 Proficiency in database design

Efficient data storage in a relational database is a key component to this project. All of the team members have been trained in relational database design and database technologies such as SQL, and database management systems such as Oracle and MySQL to accomplish this task.

### 2.5.2 Proficiency in statistics

While our Biological Expert/Statistical Analyst has extensive training in statistical analysis and R programming, the team will be learning essential statistical methodologies independently, as well as in BIF705 Statistics course to achieve sufficient proficiency in statistics, in order to complete this project.

### 2.5.3 Proficiency in R Shiny

R Shiny is our candidate web application framework to be used for this project. All of the team members have been trained in the fundamental concepts of programming which applies to any programming languages or web development. Since the team members have experience in either working with R programming language to perform statistical analyses, or working with full stack web application frameworks, we will

train independently to become proficient in the framework before the execution phase of this project in order to proceed with the subsequent development effort.

## 3 Approach:

### 3.1 Plan

Our project is comprised of four major phases: Initiation, Planning, Execution and Closure. The action plan for each of these phases is outlined as follows:

#### 3.1.1 Initiation

During the initiation phase, we will gather and identify key requirements for the project. We will study a sample data set provided by our PI, in order to derive an appropriate database scheme, and to implement meaningful statistical analytic functionalities in the web application. We will also explore available database and web application framework, and choose the most appropriate technology stack for this project. Our current candidate framework is R Shiny. Upon completion of the initiation phase, we will have derived a prototype database design, and chosen a web application framework to work with.

#### 3.1.2 Planning

During the planning phase, we will organize the key requirements into two subcategories: "baseline" and "additional". The baseline requirements are the items that must be completed in order to successfully deliver the project, while the additional requirements will be individually implemented so long as time and resources permit. We will also define use cases by examining the requirements, in order to outline the interactions between the users and the system. We will derive system sequence diagram(s) to visualize how the system executes and responds to the user inputs.

#### 3.1.3 Execution

During the execution phase, we will develop the web application by implementing the use cases identified in the planning phase. Agile development methodology will be used for this purpose, which takes an iterative approach to the development. We will begin the process by developing a functional prototype, provide it to the client to receive feedback. The client will be given access to the prototype application which will be hosted in our testing server. We will then start the next iteration of the development to

implement additional use cases, or make modifications to the existing implementations based on the feedback. Our plan is to perform three to four iterations of two- to three-week development cycles to complete the execution phase.

#### **3.1.4 Closure**

During the closure phase, the application will be installed and set up in the client's system environment. We will perform functional and installation testing to ensure that the application functions as expected. Any issues found after the testing will be resolved before the end of the closure phase. We will also finalize an installation and user documentations for the client.

### **3.2 Resources**

The following resources will be required to complete the project, and will be provided by either the client or the team:

#### **3.2.1 Resources Provided by the clients**

- ICP data output to be used for testing and development.
- Information on any other data that needs to be inserted into the database.
- A production server environment at the client's location to install the web application.

#### **3.2.2 Resources Provided by the team**

- A testing server environment to host the prototype application, and to perform testing for each iteration.
- Appropriate development environment and hardware to develop the application.

### **3.3 Sub-tasks**

The project tasks can be categorized into four major groups as follows:

#### **3.3.1 Management**

Management tasks include ensuring the project time line, defining the features to be implemented for each iteration of development, and setting up team meetings. These tasks will be shared equally among the teammates.

#### **3.3.2 Correspondence**

Correspondence tasks include maintaining frequent contact with the stake holders, namely the PI and the client, setting up regular meetings to update on the project progress, and clarifying any uncertain requirements. These tasks will be shared equally among the teammates.

#### **3.3.3 Development**

Development tasks focus on implementing the features to the web application. These tasks will be divided among the teammates based on the three components of the application: back-end database communication, statistical analyses and data visualization. The back-end database communication will be mainly performed by the Mechanical Back-end Developer, the statistical analyses will be implemented by the Biological Expert/Statistical Analyst, and the data visualization will be implemented by the Interactive UI/UX/visualization Developer.

#### **3.3.4 Testing**

Testing tasks ensure the functional integrity of every version of the web application provided to the stake holders throughout the development process. Each team member responsible for the assigned component will be testing any feature that is implemented in it, in order to identify and resolve any issues with the application before releasing a version at the end of each iteration.

### **3.4 Time-line**

Our project will utilize an AGILE framework to develop software in multi week 'sprints' after which the software will be implemented to generate feedback from our sponsor and PI. The current goal is to conclude the planning stage of our project life-cycle by December 15<sup>th</sup> and begin our first sprint, where we will generate a working prototype, ending the first week of January. After this we can implement the system for our sponsor and PI, allowing them a week of use to generate feedback and inform further changes. The third week of January we will meet with our sponsor and PI to define any modifications or additional features that are deemed necessary.

With this information in hand, we will proceed on a two week 'sprint' before providing another week for the end-user to experience our product. This process will repeat until we have satisfied, and hopefully exceeded, the success criteria for the project. Under the AGILE system, essentially undergo all four steps of planning over the course of our sprint.

## 3.5 Roadblocks

### 3.5.1 Time-management

Given the demanding nature of the bioinformatics program, we will need to ensure that we can manage our time wisely to meet the timeline outlined above. We have attempted to accelerate progress towards measurable goals by implementing the AGILE framework. This development process prioritizes generating useful prototypes quickly, without getting bogged down in the details, then adjust the product according to feedback from its use in the field. As a team, we will define the progress required each week, assigning tasks through our Trello board which indicates whether they have been completed.

### 3.5.2 Lack of Information

In order to proceed with development we must ensure all relevant questions are answered during our pre- and post-sprint meetings with Mitronite. This will require thorough preparation to ensure that we do not forget to ask about an aspect of our product; otherwise progress will be impeded. To solve this problem we will need to ensure that the team has a pre-meeting huddle to define the information we need to move forward, write it down, and assign questions to each group member. In this way inter-group accountability can be used as a tool to gather complete and relevant information about the needs of our sponsor and PI during each meeting.

### 3.5.3 Hardware Available

We currently have no information about the hardware on which we will be implementing our product. It is necessary to ensure this is specified early allow for designing and optimizing our software according to the memory, processing power, and disk space which will be available. We may choose to go with cloud based access, but for this we would need to find out the budget available, and limit our application to operate within the specifications each plan provides. It is unclear at this moment where Mitronite plans to host the final deliverable. The solution to this is clear and early communication with the Mitronite team so that we know exactly what hardware environment we are developing the application for.

## 4 Deliverable:

The final deliverable for this project will be a full-stack web application encompassing database design, back-end code to run on the

server and an interactive UI which embeds useful visualization methods and statistical tests in a way that makes analysis seem effortless.

## 4.1 Final Product

Our final product is a browser based application which will be accessible to all members of the Mitronite team via a local network or the cloud. Logins will be assigned and managed to ensure the data is only available to authorized personnel. The database side of our app will ensure data integrity, security, and recoverability through implementation of a regular back-up schedule. Data will be uploadable via .csv files, with batch loading available; alternatively, it may be possible to pull data off of the ICP-SE analyzer to automate data-entry. Database design will allow expansion should new instrumentation or other data be required.

The back-end, written in R, will control the server's general operation as well as powering the statistical and visualization tools which will be displayed in the UI. This will be invisible to the user while allowing powerful and customized tools to be accessed with the click of a mouse. Because of the wealth of packages available for R and R Shiny, we will develop visualization tools including histograms, 2D and 3D water-fall plots, as well as enable UI customizable statistical tests such as t-tests and linear regression to determine if trends observed in the visualization are significant enough to draw conclusions from. The back-end will allow addition of new visualization tools and statistical tests based on the requirements of database growth, mainly via R-packages and modification of existing code.

The UI will feature a polished and stylish look, allowing point-and-click interfaces in a 'dashboard' style web-page. As R-Shiny automates much of the front end design, we are able to provide a professional quality UI with significantly less effort than other frameworks. The dashboard will feature dynamically loading tabs to navigate through each section of the application while check-boxes, drop-down menus and text-boxes will allow real-time updates to statistical tests and visualizations based on the users-selection.

## 4.2 Product Demo

As discussed above we will be utilizing the AGILE framework for development, so we will be interactively demoing and implementing our software for our PI and sponsor as we complete each sprint. The initial demo will occur in the first half of January and will be a guided tour through

the components and features of our product, followed by hands-on-training for the staff members who will be using it. Further demos will focus on new features or changes and will highlight how we have met the criteria laid out in each of our post-sprint meetings. Ideally, the software will be implemented on the system so we can open associated code on a projector, then navigate to the host to access the application.