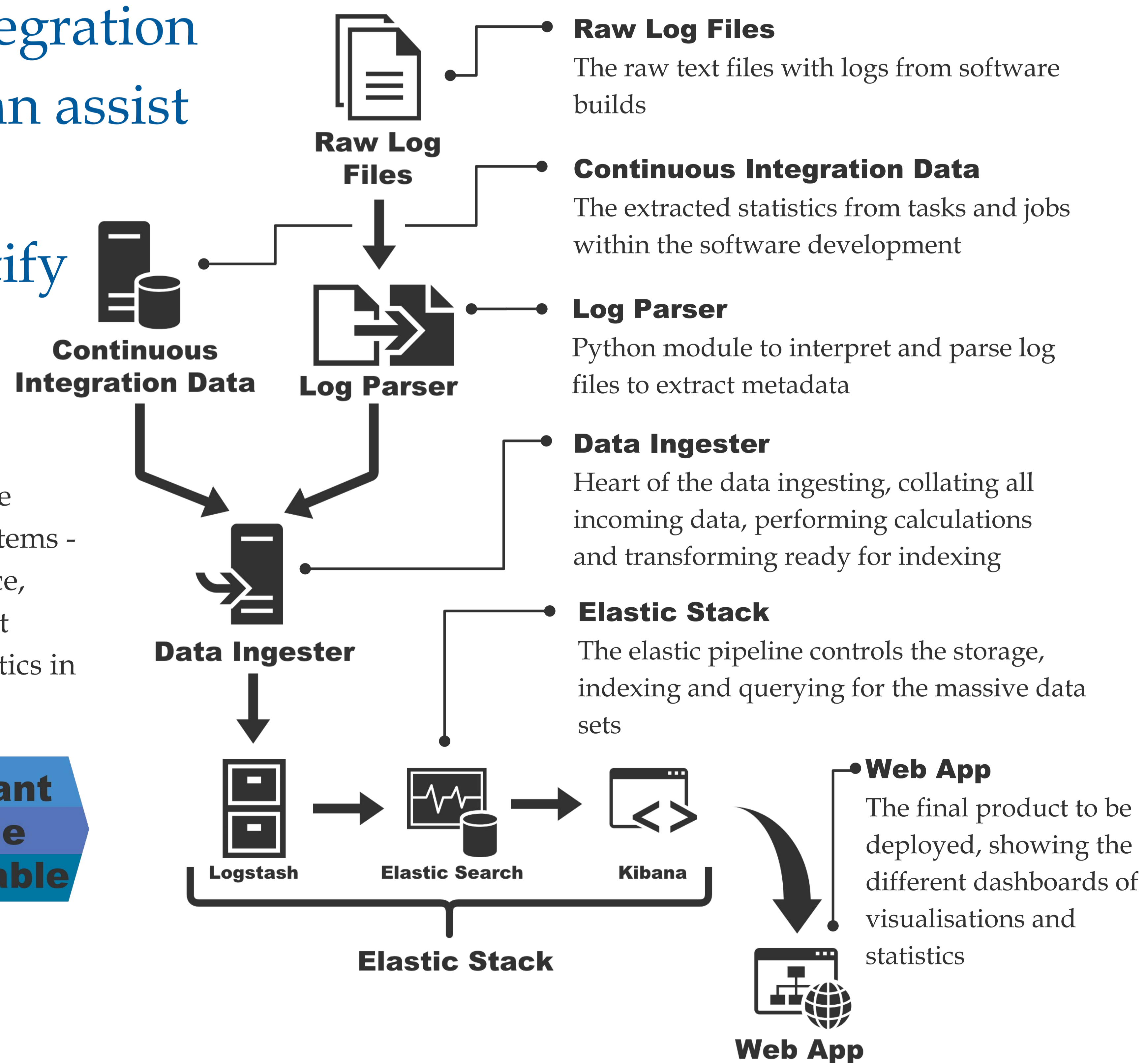# Viewing Big Data in Software Builds

Markus Andersons, Jake Lane, Samuel Lewis {a1687605, a1686679, a1668737}@student.adelaide.edu.au

Supervised by Christoph Treude, Meredith Lane, Jase Rieger

Through collaboration with an industry client, Maptek, a web app was created which enables the trends and patterns in big data from continuous integration systems to be identified. This can assist in the optimisation of software development processes to identify potential process bottlenecks.

Project Goals

Modern software development companies generate massive amounts of Big Data from their Continuous Integration systems - the systems that run tests over newly submitted code. Hence, much of the project is focussed around determining the best mechanism to display this large amount of visualised statistics in a timely fashion. The hierarchy of these goals is:

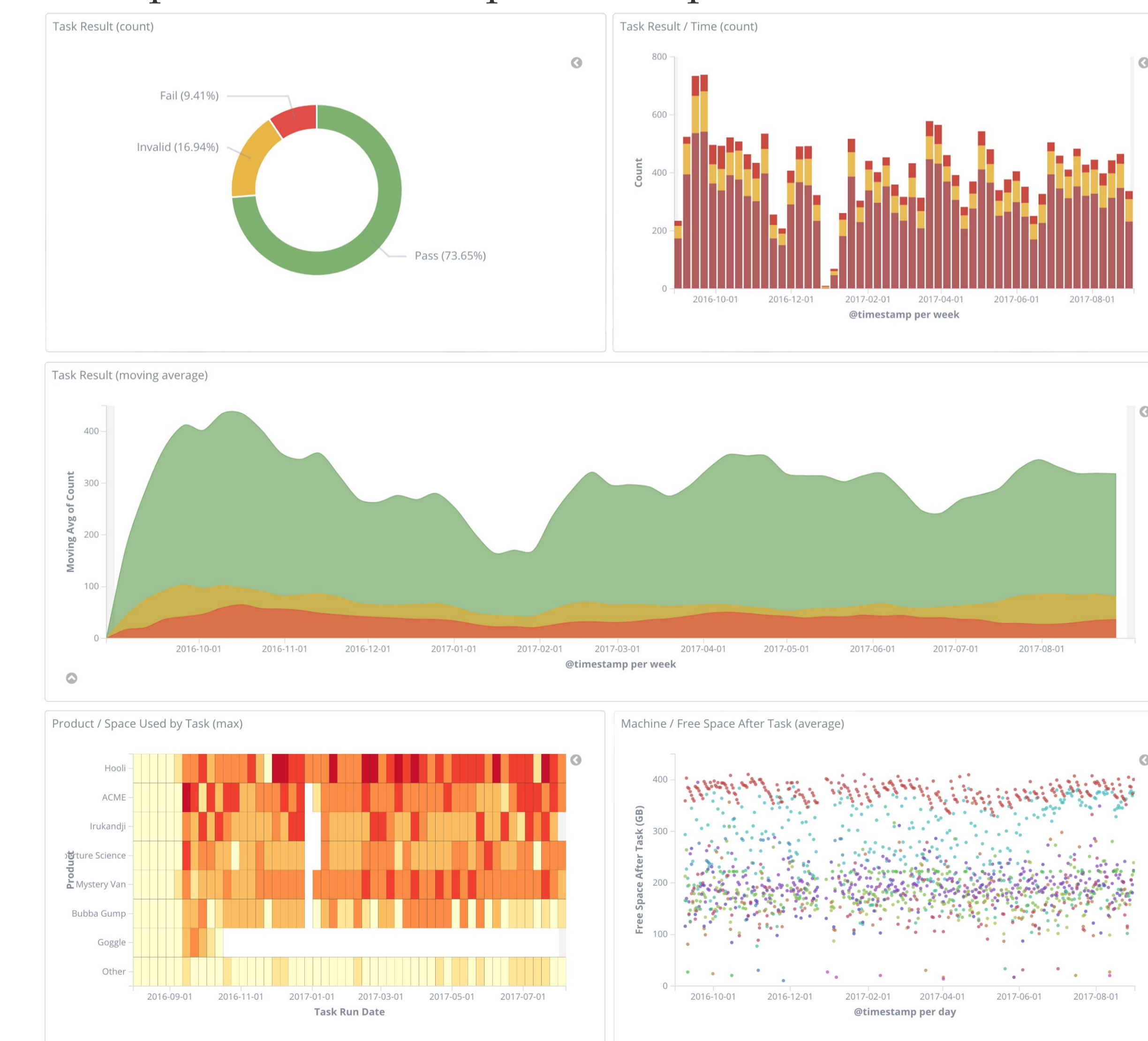**Trends from Big Data** → **Visualisation** → **Performant Portable Maintainable**

Method

To highlight bottlenecks in the development process, we utilised a combination of off-the-shelf and custom solutions as shown below. This data is shown via a large range of interactive graphs which enable users to manually identify these areas of interest.

Raw Log Files
The raw text files with logs from software builds

Continuous Integration Data
The extracted statistics from tasks and jobs within the software development

Log Parser
Python module to interpret and parse log files to extract metadata

Data Ingester
Heart of the data ingesting, collating all incoming data, performing calculations and transforming ready for indexing

Elastic Stack
The elastic pipeline controls the storage, indexing and querying for the massive data sets

Web App
The final product to be deployed, showing the different dashboards of visualisations and statistics

Results

The created web application has been deployed to our client and is currently in production use to gain valuable feedback from the users. By visualising the big data, it can easily be interpreted by human users and the trends and important statistics can be identified. While it is too early to gauge the complete effectiveness of the system, the initial results are proving promising, with the visualised data already generating a number of interesting insights which highlight some possible areas for process improvement.

Acknowledgements

We would like to thank Maptek for their time and assistance throughout this project.