# Data Analytics and Visualisation (COMP09014)

# MSc in Computing (Data Science)

# Final Project Proposal

# Andre McQuaid (S00216958)

## Introduction

Covid-19 has been with us for almost 2 years now and Johns Hopkins university has become an important and trusted data source for many analyses of Covid-19 related data. [CSSEGISandData/COVID-19: Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE (github.com)](https://github.com/CSSEGISandData/COVID-19) is a Github project where John Hopkins makes Covid-19 data publicly available. The goal of this project is to analyse, pre-process and consume this data with a view to developing a Shiny dashboard to visualise trends in Covid-19 cases and deaths (and potentially recoveries too) across the globe.

The R statistical programming language will be used extensively as part of this project and this will feed into the Shiny dashboard, as Shiny is also based on the R language.

The project will also look to leverage ML and AI techniques in order to add a predictive element to the project. For example, predict regional cases and deaths.

## Data

The following key data points are available and will be leveraged as part of the project:

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| Date | Data snapshot date - MM-DD-YYYY.csv in UTC |
| Province/State | Province, state or dependency name |
| Country/Region | Country, region or sovereignty name. |
| Last Update | MM/DD/YYYY HH:mm:ss (24 hour format, in UTC) |
| Confirmed | Counts include confirmed and probable (where reported). |
| Deaths | Counts include confirmed and probable (where reported). |
| Recovered | Recovered cases are estimates based on local media reports |
| Latitude | Location Data |
| Longitude | Location Data |

## Objectives

* Visualise the prevalence of Covid cases, deaths and recoveries (where available) across the various regions and countries of the world
* Allow rich filtering capabilities in the dashboard, for example:
  + Switch between cases and deaths
  + Region\country level filtering
  + Trending – filter on whether cases\deaths are increasing or decreasing
* Show the evolution of cases and deaths over time, given access to this time series data. For example, ability to show an animation of the progression of cases\deaths over time
* Provide predictive capabilities around regional cases and deaths, leveraging the existing data. For example, predict the regions\countries where an increase in cases and\or deaths could be anticipated

## References

[COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University](https://github.com/CSSEGISandData/COVID-19)

[The R Project for Statistical Computing](https://www.r-project.org/)

[Shiny (rstudio.com)](https://shiny.rstudio.com/)

O. Sarkar, M. F. Ahamed and P. Chowdhury, "[Forecasting & Severity Analysis of COVID-19 Using Machine Learning Approach with Advanced Data Visualization](https://ieeexplore.ieee.org/document/9392704)," 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392704.

A. Kunjir, D. Joshi, R. Chadha, T. Wadiwala and V. Trikha, "[A Comparative Study of Predictive Machine Learning Algorithms for COVID-19 Trends and Analysis,](https://ieeexplore.ieee.org/document/9282953)" 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 3407-3412, doi: 10.1109/SMC42975.2020.9282953.

Hamzah, F.B., Lau, C., Nazri, H., Ligot, D.V., Lee, G., Tan, C.L., Shaib, M.K.B.M., Zaidon, U.H.B., Abdullah, A.B. and Chung, M.H., 2020. [CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction](https://cdn.spotle.ai/projects/296083/10079/20_255695.pdf). *Bull World Health Organ*, *1*(32).

Roy, A.N., Jose, J., Sunil, A., Gautam, N., Nathalia, D. and Suresh, A., 2020. [Prediction and spread visualization of COVID-19 pandemic using machine learning](https://www.preprints.org/manuscript/202005.0147/v1).

Sengupta, S., Mugde, S. and Sharma, G., 2020. [Covid-19 pandemic data analysis and forecasting using machine learning algorithms](https://www.medrxiv.org/content/10.1101/2020.06.25.20140004v2.full.pdf). *medRxiv*.