SUBMITTED FOR THE DEGREE OF B.SC IN COMPUTING SCIENCE FOR THE YEAR

2018/2019

**dtools**

Random Patient Clinical Data Simulator

Student Name: Karol Groszewski

Registration Number: 201507145

“Except where explicitly stated all the work in this report, including appendices, is my own and was carried out during my final year. It has not been submitted for assessment in any other context.”

Signature: Karol Groszewski

Date: 25/07/2019

“I agree to this material being made available in whole or in part to benefit the education of future students.”

# Abstract

Access to real patient data is important for teaching and research purposes and research. However, due to legislation this data is protected and can only be used by people who have permission to access it. The solution to this problem was to create an application which simulates that data based on realistic statistics.

The resulting application turned out be a success as the simulated statistics are realistic. The application also allows users to view the data that was simulated. The user can access statistics based on various health conditions that affect the Scottish population. By choosing statistics on a specific health condition the user can view a chart and a table of records of the simulated data.

# Acknowledgements

I would like to acknowledge Marilyn Lennon, as a supervisor of this project, for her indispensable direction and support throughout the project.

# Table of Contents

[Abstract 2](#_Toc15007517)

[Acknowledgements 3](#_Toc15007518)

[Table of Contents 4](#_Toc15007519)

[Chapter 1 7](#_Toc15007520)

[Introduction 7](#_Toc15007521)

[1.1 Objectives 7](#_Toc15007522)

[1.2 Report Structure 7](#_Toc15007523)

[1.3 Marking Scheme 8](#_Toc15007524)

[Chapter 2 9](#_Toc15007525)

[Background Research/Related Work 9](#_Toc15007526)

[2.1 Simulating Patient Data 9](#_Toc15007527)

[2.2 Data Sources 9](#_Toc15007528)

[2.2.1 National Records of Scotland 9](#_Toc15007529)

[2.2.2 Scottish Government 10](#_Toc15007530)

[2.2.3 ISD Scotland 10](#_Toc15007531)

[2.2.4 Diabetes UK 10](#_Toc15007532)

[2.2.5 ScotPHO 10](#_Toc15007533)

[2.3 Related Work 10](#_Toc15007534)

[2.3.1 generatedata.com 11](#_Toc15007535)

[2.3.2 databasetestdata.com 12](#_Toc15007536)

[Chapter 3 13](#_Toc15007537)

[Problem Description and Specification 13](#_Toc15007538)

[3.1 Problem Description 13](#_Toc15007539)

[3.2 Where can this software be applied? 13](#_Toc15007540)

[3.3 Functional Requirements 13](#_Toc15007541)

[3.4 Non-Functional Requirements 14](#_Toc15007542)

[3.5 Approach taken to Solving the Problem 14](#_Toc15007543)

[Chapter 4 15](#_Toc15007544)

[Data Analysis and Creation 15](#_Toc15007545)

[4.1 Community Health Index (CHI) Number 15](#_Toc15007546)

[4.2 Statistics gathering 15](#_Toc15007547)

[4.2.1 Age Distribution 15](#_Toc15007548)

[4.2.2 Gender Distribution 16](#_Toc15007549)

[4.2.3 Body Mass Index (BMI) Distribution 16](#_Toc15007550)

[4.2.4 Circulatory and Respiratory 18](#_Toc15007551)

[4.2.5 Diseases of the Skin & Subcutaneous Tissue 19](#_Toc15007552)

[4.2.6 General Abnormal 20](#_Toc15007553)

[4.2.7 Digestive/Abdominal 21](#_Toc15007554)

[4.2.8 Diabetes Type 1 22](#_Toc15007555)

[4.2.9 Diabetes Type 2 23](#_Toc15007556)

[4.2.10 Hypertension 24](#_Toc15007557)

[Chapter 5 25](#_Toc15007558)

[Development Process of the Application 25](#_Toc15007559)

[5.1 Choice of Methodology Used – Waterfall Model 25](#_Toc15007560)

[5.2 Application Class Design 25](#_Toc15007561)

[5.3 Data set Design 26](#_Toc15007562)

[5.4 Graphical User Interface Design 28](#_Toc15007563)

[Chapter 6 30](#_Toc15007564)

[Detailed Design and Implementation 30](#_Toc15007565)

[6.1 Choice of Programming Languages Used 30](#_Toc15007566)

[6.1.1 HTML5 30](#_Toc15007567)

[6.1.2 CSS 30](#_Toc15007568)

[6.1.3 JavaScript 30](#_Toc15007569)

[6.2 Choice of Third-Party Tools 30](#_Toc15007570)

[6.2.1 Strathclyde DEVWEB 30](#_Toc15007571)

[6.2.2 FileZilla 31](#_Toc15007572)

[6.2.3 ag-Grid 31](#_Toc15007573)

[6.2.4 Chart.js 32](#_Toc15007574)

[6.2.5 Bootstrap 32](#_Toc15007575)

[6.2.6 jQuery 32](#_Toc15007576)

[6.2.7 Google Chrome 32](#_Toc15007577)

[6.3 Choice of Development Environment 33](#_Toc15007578)

[6.3.1 Operating System: Windows 10 33](#_Toc15007579)

[6.3.2 Mobile Operating System: Android 9 33](#_Toc15007580)

[6.3.3 IDE: Visual Studio Code 33](#_Toc15007581)

[6.3.4 Version Control: Git and GitHub 33](#_Toc15007582)

[6.4 Challenges of Design and Implementation 33](#_Toc15007583)

[6.4.1 Data Analysis and Clean-Up 33](#_Toc15007584)

[6.4.2 Storage of Large Data in Local Storage 34](#_Toc15007585)

[6.4.3 Manipulation of Large Volumes of Data 34](#_Toc15007586)

[Chapter 7 35](#_Toc15007587)

[Verification and Validation 35](#_Toc15007588)

[7.1 Verification 35](#_Toc15007589)

[7.2 Validation 35](#_Toc15007590)

[Chapter 8 37](#_Toc15007591)

[Results and Evaluation 37](#_Toc15007592)

[8.1 User Evaluation 37](#_Toc15007593)

[8.2 Evaluation of Planning 41](#_Toc15007594)

[8.3 Evaluation of End Product Functionality 42](#_Toc15007595)

[Chapter 9 46](#_Toc15007596)

[Summary and Conclusions 46](#_Toc15007597)

[9.1 Summary 46](#_Toc15007598)

[9.2 Issues and Limitations 46](#_Toc15007599)

[9.3 Future Work 46](#_Toc15007600)

[9.4 Final Conclusion 46](#_Toc15007601)

[Appendix A 47](#_Toc15007602)

[References 47](#_Toc15007603)

[Appendix B 48](#_Toc15007604)

[Detailed Test Strategy and Test Cases 48](#_Toc15007605)

[Appendix C 53](#_Toc15007606)

[Installation Guide 53](#_Toc15007607)

[User Guide 53](#_Toc15007608)

[Appendix D 56](#_Toc15007609)

[Participant Consent Form 56](#_Toc15007610)

[Appendix E 60](#_Toc15007611)

[User Survey 60](#_Toc15007612)

Chapter 1

## Introduction

Simulation of data can be a huge benefit for teaching and research purposes. In order to create data that can be used for that purpose, real statics will have to be used with generated patients in order to create this data as real data cannot be used due to laws protecting people’s privacy.

The simulated data will allow people to simulate patients based on realistic statistics. These statistics will allow researches to explore, study and utilize data which is based on variety of people affected by various health conditions. These simulated statistics will be displayed to the user using charts and a view of records.

### 1.1 Objectives

The main objective of this project is to develop a simulator that can create “random patient data” which then can be used for training, research and teaching purposes. The user will have access to information about various common health conditions based on Scottish statistics from the following health organizations:

* National Records of Scotland [3]
* Scottish Government [4]
* ISD Scotland [5]
* Diabetes UK [6]
* ScotPHO [7]

The next objective will be for the application to allow users to filter through that information in order to analyse the data about these conditions. While filtering through these statistics, appropriate graphs are going to be generated to aid ease of reading the data.

Additionally, an objective of the project is that the application shall be web-based so that it can be accessed remotely and on any type of device. Therefore, the user should be able to access the application on various electronic devices that have an internet access.

Furthermore, another objective of the project will be to provide sources that have been gathered to produce this application are going to be available to the user. If they would like to find out more information about these conditions or data used to generate these statistics.

### 1.2 Report Structure

The report is broken down into 9 different chapters where each chapter focuses on different development stage of the application. The first chapters focus on background research and the problem to be solved by the application. Followed by chapters focused on detailed analysis of design of the application, how implementation process has been executed, what issues have been encountered during the development of the application. The what were the results of testing stage and evaluation of the application. There report will come to an end with a summary and last comments before the final conclusion.

### 1.3 Marking Scheme

The marking scheme for this project is going to be “Experimentation-based with Significant Software Development”. This marking scheme was chosen because a large part of the project will be dedicated to researching, health condition statistics, their simulation, and development of application for their simulation.

Chapter 2

## Background Research/Related Work

The aim of this chapter is to describe in detail what research has been done in order to prepare for development of the Random Patient Clinical Data Simulator – “RPCDS”. Throughout this chapter we will discuss the benefits of simulating patient data in relation to teaching and research. Moreover, we will describe from where the statistics for this project were gathered.

### 2.1 Simulating Patient Data

Simulating patient data is the creation of personal data with details such as their first name, surname, gender, age and health conditions. Simulating patient data relates to my project as accessing real patient data is difficult due to legislation. Simulation-based learning in a health education environment has some advantages over other teaching methods used in that field depending on context, topic and method.

By creating this application, students who are studying to be a health care professional or medical professional can practice analysing patient data and statistics which can result in additional gains in knowledge, critical thinking ability, satisfaction or confidence based on studies [1]. Therefore, by developing an application that simulates patient data based on real data sources and statistics it will benefit the students in the medical industry.

Ideally researchers would want real patient data of a large number of people; however, this may not be possible due to legislations. Researchers may want data to identify trends in illnesses in various age groups and genders. Therefore, the application will be able to produce simulated data analytics based off the Scottish population for use by these researchers.

### 2.2 Data Sources

For “RPCDS” to be useful it needs to have some sort of data which comes from valid statistical sources. In this part of the chapter various data sources are going to be talked about and how they provide valid data for the project.

#### 2.2.1 National Records of Scotland

National Records of Scotland is a non-ministerial department of the Scottish Government which is responsible for civil registration, the census in Scotland, demography and statistics, family history, the national archives and historical records.

Since National Records of Scotland is a government organisation, they provide data which is used by various researchers and this would qualify as a high-quality data source for the application. The statistics that they produce play a vital role in underpinning decisions from national to local level and are a building block in the development of economic and social statistics.

#### 2.2.2 Scottish Government

The Scottish Government website provides a range of official statistics about Scotland from a variety of data producers, for information and re-use. They provide public access to data behind our official statistics in linked open data format.

#### 2.2.3 ISD Scotland

Scotland has some of the best health service data in the world. This is because Scotland has information which combines high quality data, consistency, national coverage and the ability to link data to allow patient-based analysis and follow up. The Information Services Division which is a part of NHS National Services Scotland provides health information, health intelligence, statistical services and advice that supports quality improvement in health and care and facilitates robust planning and decision making.

#### 2.2.4 Diabetes UK

Diabetes UK is a charity which is registered Scotland, England and Wales. The charities aim is to tackle the diabetes crisis. They are there to prevent Type 2 diabetes, campaign for and support everyone by diabetes, and fund research that will cure the condition.

Diabetes UK is UK’s leading charitable funder of diabetes research. They help others by pioneering research into all forms of diabetes and diabetes-related complications. The charity releases annual statistics on how and how many people are affected by the condition in the UK.

#### 2.2.5 ScotPHO

ScotPHO’s aim is to provide a clear picture of the health of the Scottish population and the factors that affect it. They contribute to improved collection and use of routine data on health, risk factors, behaviours and wider health determinants. ScotPHO takes a lead in determining Scotland’s future public health information needs, develop innovations in public health information and provide a focus for new routine public health information development where gaps exist.

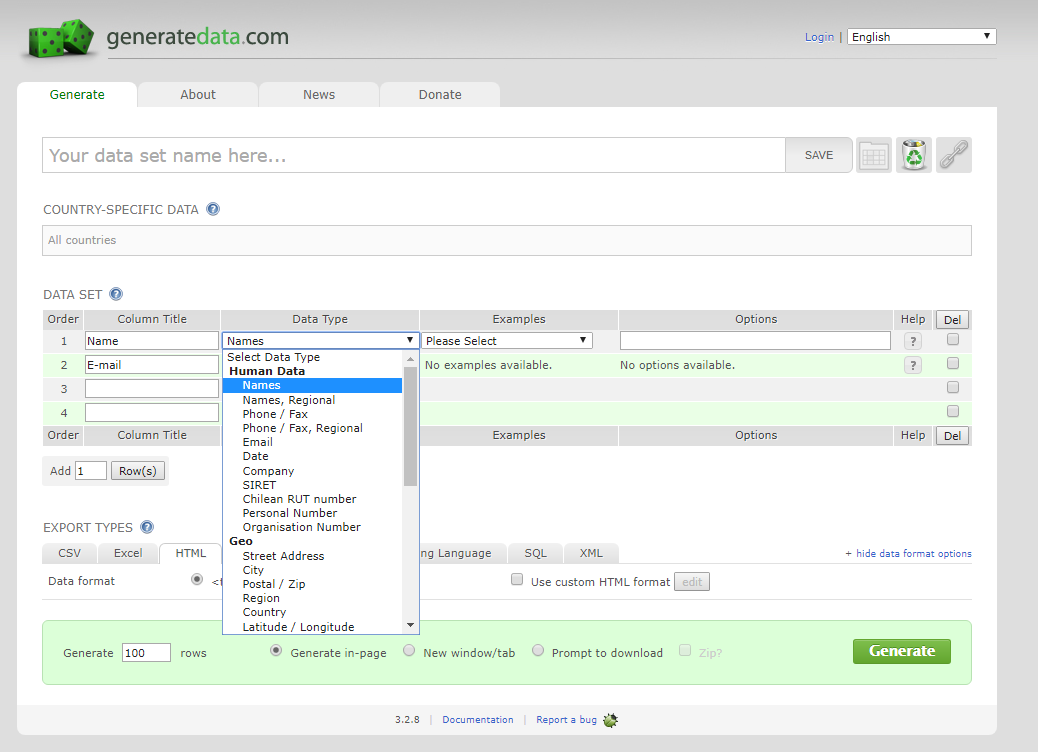
### 2.3 Related Work

In this part of the chapter different existing applications are going to be analysed in order to see how other people generate data for the users of their services. From there advantages and disadvantages of each application are going to be listed. The criteria that the applications were judge on are:

* Ease of use of the application
* Options available to the user
* In what formats can the data be outputted

#### 2.3.1 generatedata.com

generatedata.com is a free open-source tool which creates custom formatted samples and test data.



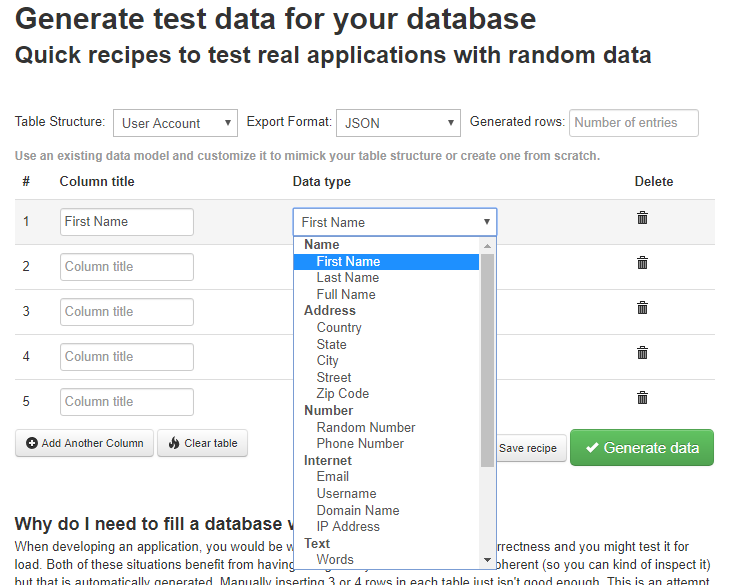
This data tool allows users to add various column names and select a data type which tell the website what data to generate for that column. It also has filtering. For example, if you want just to generate a female population you can add that statement to the options tab. After that the user can chose to what file the generated data should be exported to and how much of data should be generated by specifying the number of rows.

The advantage of using this website is that the user has various options in order to generate specific data that they would like. Another advantage is that the user can export the generated data to various formats for their uses. The data can be country specific. If you want to generate people who live in another country and you want to have their addresses, you can add that to the generation by changing that option.

The disadvantages of this website are that the user interface feels complicated to use. For example, if you want to generate just people with random names then you also need a provide an example to the website cause otherwise it won’t know how you would like to generate that data. Another disadvantage is that when you get an error the website won’t tell you what is causing it. When you click on the help button it gives the user a lot of information which is not related to the problem that the user is having.

#### 2.3.2 databasetestdata.com

databasetestdata.com is a free test data generator for databases. The aim of this test data is to test real applications with random test data.



This data tool allows users to generate specific data based on the “Data Types” chosen by the user. The website has some already pre-written styles such as “User Account” or “Shop Product”. The user also has the option to export the generate the data to JSON, CSV and XML. You also have the option to generate various number of entries. After clicking “Generate data” button the website takes the user to another page which shows the generated raw string in a textbox.

The advantages of this website are that the user has a wide range of options that they can generate. Another advantage is that the User Interface is simple to use as all the actions to generate data are simple to execute.

The disadvantage of this website is that the website does not export any of the data to actual files that can be downloaded by the user. Therefore, the user would have to use another service in order to convert the generated data into a file and store it on their system.

Chapter 3

## Problem Description and Specification

The aim of this chapter is to describe in detail what the problem is and how it is going to be approached. This will include a complete list of functional and non-functional requirements to be met by the software and how these requirements were identified and approached.

### 3.1 Problem Description

The challenge in this project is that the medical researchers cannot access real patient data for analysis. Another problem in Random Patient Clinical Data Simulator project will consist of creating a population which will have an age, gender, BMI[[1]](#footnote-1) and conditions based on statistics gathered for the project. Statistics come in various formats. Therefore, these statistics will have to be cleaned up in order to be used by the application. Statistical data isn’t always 100% accurate therefore, an error rate will have to implemented to make the application realistic and vary from simulation to simulation. Each person will also require a valid postcode which is in Scotland. Each person will have to have a unique CHI Number[[2]](#footnote-2) where the format will have to be the same for each person. The simulated data will have to be displayed using appropriate graphs based on the condition chosen by the user.

### 3.2 Where can this software be applied?

This software is going to be designed for researchers and health professionals who want to improve their analytical skills.

### 3.3 Functional Requirements

The functional requirements were decided on based on the problem description and related work discussed in the previous chapter.

The software will have to (listed from highest priority to lowest):

* Simulate data based on statistics gathered.
* Display the records stored in the data set.
* Assign postcodes to records in the data set that are valid postcodes in Scotland.
* Create a CHI Number for each record in the data set with appropriate format.
* Create a data set which is different each time the user decides to simulate the data.
* Implement error rate.
* Display graphical statistics based on each condition to the user.
* Let the data set to be retrieved to the user that has been simulated previously.

### 3.4 Non-Functional Requirements

The non-functional requirements were decided on based on the problem description and related work discussed in the previous chapter.

The software will have to (listed from highest priority to lowest):

* Load a data set automatically that has been previously simulated by the user on a specific device.
* Have an easy to use user interface so that they user can access simulated data quickly.
* Contain a user interface that support various screen sizes.

### 3.5 Approach taken to Solving the Problem

The approach taken to solve this problem was to create a website with generated user data and display statistics and records on the website using HTML5, CSS and JavaScript. A decision was made to use those programming languages for the website since I am comfortable in programming in those languages. The other reason why decision was made to use those languages since it will allow me to provide the application to a wide range of users since it will be available on the internet.

Before web application development, sources on age distribution had to be researched. This is done in order to know what percentage of people are of a specific age based upon their gender. Otherwise without that data it would be impossible to assign realistic age groups to the population with specific ages. If the data on age groups wouldn’t be available, then the application would lose its purpose as the age groups would be all over the places and after applying conditions statistics it would result in unrealistic or in data that would never be possible to achieve.

After gathering statistics on age distribution, statistics on various health conditions that affect the population of Scotland gad to be gathered. The only acceptable data would be that provides an age distribution and gender in order to keep the generated data usable.

From there the data was cleaned up since the statistics would only provide a number of people that are affected by the condition but in order for the data to be useable by the application, the statistics had to converted into percentages. For example, for one of the conditions the contact rate was per 1000 people was 400 males and that would be converted to percentage and stored in an Excel File created by myself.

After that stage implementation of the application could take place. The first step was to design the actual system and the way it will work. This has been done by just drawing diagrams and ideas. The next step was to decide on how to store the simulated data so that the user is going to be able to access it multiple times without the need to re-simulate the data. A decision was made to store all the simulated data in local storage. From there it was just the case of programming the application until completion.

Chapter 4

## Data Analysis and Creation

The aim of this chapter is to describe in detail step by step how data was gathered, cleaned, analysed and created for the application.

### 4.1 Community Health Index (CHI) Number

The CHI number is a population register used in Scotland for healthcare purposes. The CHI Number uniquely identifies a person on the index. This number is 10 digit long. The first 6 digits are persons Date of Birth (DDMMYY) followed by 2 random digits. The 9th digit is representing the persons sex which assigns an even number for females and an odd number for males. The final digit is a check digit [2].

### 4.2 Statistics gathering

Statistics for this project were gathered from sources listen in Chapter 2 to ensure the data would be valid. The statistics were gathered in a format which showed how many people were affected by the condition in different age groups based on gender. Therefore, each of the statistics had to be cleaned up and converted into percentages so that the statistics data could be used by the application. From that data graphs were created which are shown in each of the sections.

#### 4.2.1 Age Distribution

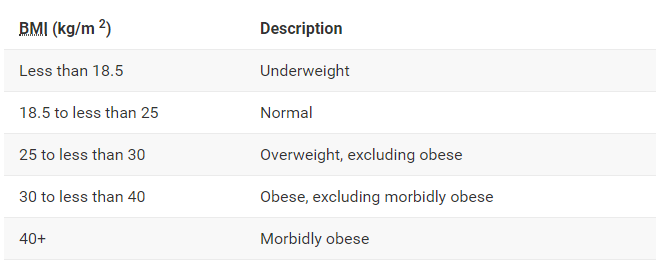
For the software to be useable it will have to have realistic statistics about how the age is distributed in Scotland. This is because conditions to individuals are going to be assigned by their age and gender. Below is a graph included which shows the percentage of age is distributed in Scotland based on Mid-2017 population estimates Scotland [3].

#### 4.2.2 Gender Distribution

To get accurate results in this project gender will have to be distributed accordingly. This is because if all age groups are split into 50% males and 50% females then wrong data will be outputted. In the graph below it is shown the percentage of males and females at different ages [3].

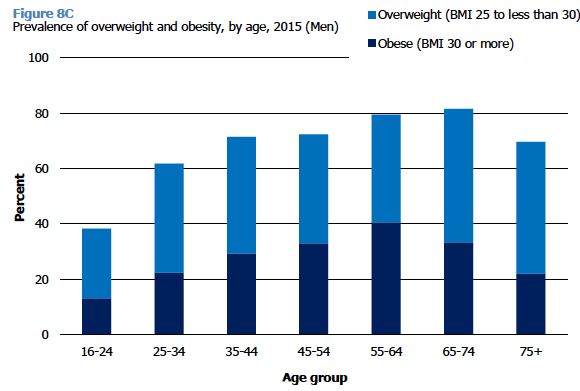
#### 4.2.3 Body Mass Index (BMI) Distribution

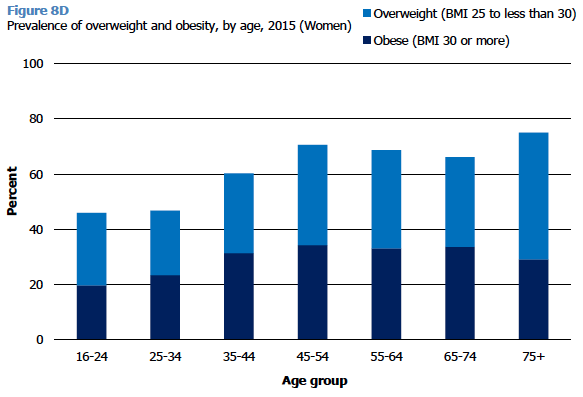
BMI is a widely accepted measure that allows for differences in weight based on persons height. This measure is important since people with higher BMI tend to be more affected by conditions such as Diabetes and Hypertension. Below is a BMI table for adults who are age 16 or higher.



Since people with an age lower than 16 are classified differently people who are younger than that will not have a BMI measure in the system. Below are charts which show peoples BMI based on age and gender. For people with a BMI that have lower than 25 will have a BMI which is either normal BMI or underweight BMI will have a random BMI generated between 15 and 25 [4].

For BMI data clean up didn’t have to take place as the data is already formatted into percentages of people, split into various age groups and gender.





#### 4.2.4 Circulatory and Respiratory

Circulatory and Respiratory [5] condition is one of the most common condition that affects people in Scotland which is the reason this statistic is going to be included in the application. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics. The following statistics includes signs and symptoms (S&S) that do not necessarily have a confirmed diagnosis, classified according to body system.

#### 4.2.5 Diseases of the Skin & Subcutaneous Tissue

Diseases of the Skin & Subcutaneous Tissue [5] condition is one of the most common condition that affects people in Scotland which is the reason this statistic is going to be included in the application. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics.

#### 4.2.6 General Abnormal

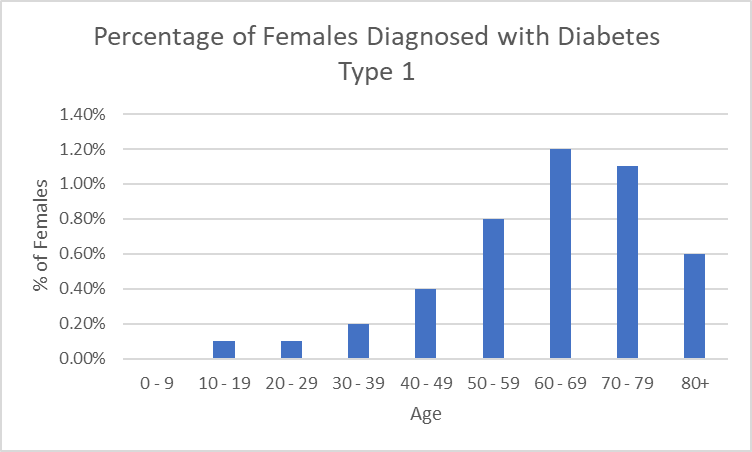
General Abnormal [5] condition is one of the most common condition that affects people in Scotland which is the reason this statistic is going to be included in the application. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics. The following statistics includes signs and symptoms (S&S) that do not necessarily have a confirmed diagnosis, classified according to body system.

#### 4.2.7 Digestive/Abdominal

Digestive/Abdominal [5] condition is one of the most common condition that affects people in Scotland which is the reason this statistic is going to be included in the application. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics. The following statistics includes signs and symptoms (S&S) that do not necessarily have a confirmed diagnosis, classified according to body system.

#### 4.2.8 Diabetes Type 1

Diabetes is a common condition in today’s society therefore, those statistics are going to be included in the application. Diabetes Type 1 [6] is less common than Diabetes Type 2. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on statistics from Diabetes UK.



#### 4.2.9 Diabetes Type 2

Diabetes Type 2 is more common than Diabetes Type 1 [6] because an obese person has a higher chance of being affected by this condition. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand.

#### 4.2.10 Hypertension

Hypertension [7] condition is one of the most common condition that affects mainly affects people who are older than 30 years old in Scotland which is the reason this statistic is going to be included in the application. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on ScotPHO statistics.

Chapter 5

## Development Process of the Application

The aim of this chapter is to describe how the application was designed. This will include what software development process was chosen, how was the structure of the data set was created for the application and how the graphical user interface was designed.

### Choice of Methodology Used – Waterfall Model

For this project it was decided that the application is going to be developed by following Waterfall Model. This methodology works by having a sequential approach to the software development cycle. The process consists of the following stages Requirements; Analysis; Design; Implementation; Testing.

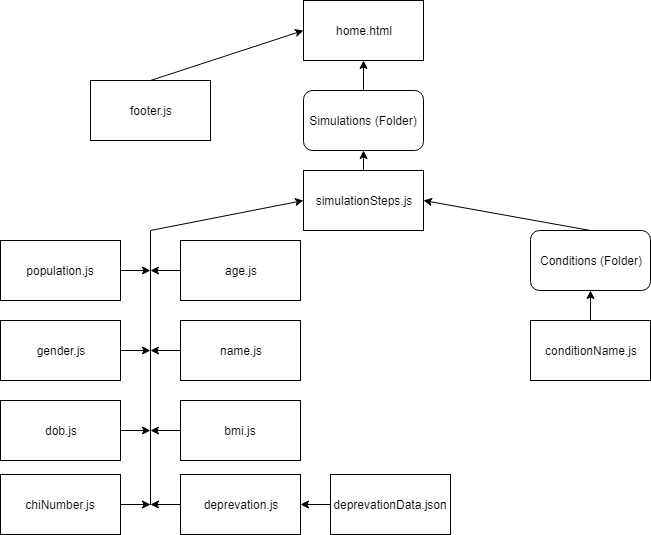
Information relating to these stages can be found at the following chapters:

* Requirements – Chapter 3
* Analysis – Chapter 2
* Design – Chapter 5
* Implementation – Chapter 6
* Testing – Chapter 7

The waterfall model is good for my project since it gives me a structure to follow in the development of this application. Furthermore, it allows me to incorporate change into the project by re-iterating stages in the waterfall model. Another reason why this model is good for this project is because in the problem description it was suggested to gather all the statistics (Requirements) before actual implementation as without these statistics you cannot implement realistic data.

### 5.2 Application Class Design

In the diagram below the class design for the simulation of data is shown. Home.html has all the scripts included shown by the arrows. Footer.js is responsible for showing text saying “CC – By Karol Groszewski ‘date’” and the current date. SimulationSteps.js will call functions from all the different classes connected to it. The population size is declared in this class too. Each class will be responsible for simulation specific data. For example, gender.js is going to be responsible for applying all the gender statistics to the population. Decision was made to choose the following design because it makes it flexible. To add new health condition statistics a new class will just have to be created with the same format. The only changes that must be made are going to be age group separation and the percentage of males and females affected by the condition. Deprevation.js loads in a JSON file called deprevationData.json. That JSON file will store all the postcodes and SIMID16 Ranks for each of the postcodes in Scotland. Age.js will include statistics on ages but also creates an error rate. The error rate is going to be created by taking away few records from each of the age groups and then the that simulation shouldn’t reach the 2000 records mark. In order to reach that 2000 records simulated mark a random function will be created where it will randomize an age between minimum and maximum and then add that person to the simulated data until there are 2000 records.



### 5.3 Data set Design

The initial idea was to use MongoDB to store all the simulated data. However, I wanted to make the application user friendly. Therefore, I have decided to store all the simulated data in the user’s device local storage. This will allow users to access simulated data just by simulating it once and from there they can access it at any time they would like. They will only have to re-simulate the data if they want a different data set or the local storage has been cleared and there is nothing stored there.

Before any storing I would create an object called person which would follow the following default layout:

|  |  |
| --- | --- |
| **Person Key** | **Person Defualt Key Value** |
| firstName | “” |
| secondName | “” |
| gender | “M” |
| age | age – this is a parameter |
| DoB (Date of Birth) | 000000 |
| postcode | “” |
| SIMD16\_Rank | “” |
| CHINumber | 0 |
| BMI | “N/A” |
| CaR (Condition) | “N” |
| DotSaST (Condition) | “N” |
| GA (Condition) | “N” |
| DA (Condition) | “N” |
| Di1 (Condition) | “N” |
| Di2 (Condition) | “N” |
| HT (Condition) | “N” |

As you can see for the persons key “age” I am passing in a parameter which sets an objects age. This is because this data is being create one by one until a certain number of people is reached with a specific age and then moves one to the next age. This makes data creation simpler as after that step I don’t need to do any changes to the persons age.

For persons key “gender” the default value is “M” standing for male. I have decided to set this as a default value as from there I will just have to apply female gender statistics. This is because let’s say there are 60% Females at an age of 5 and 40% males. I only must change 60% of people with an age 5 to “F” which sets their gender to female. This just makes the simulation cleaner as you only need to work with one gender statistic instead of two at the same time.

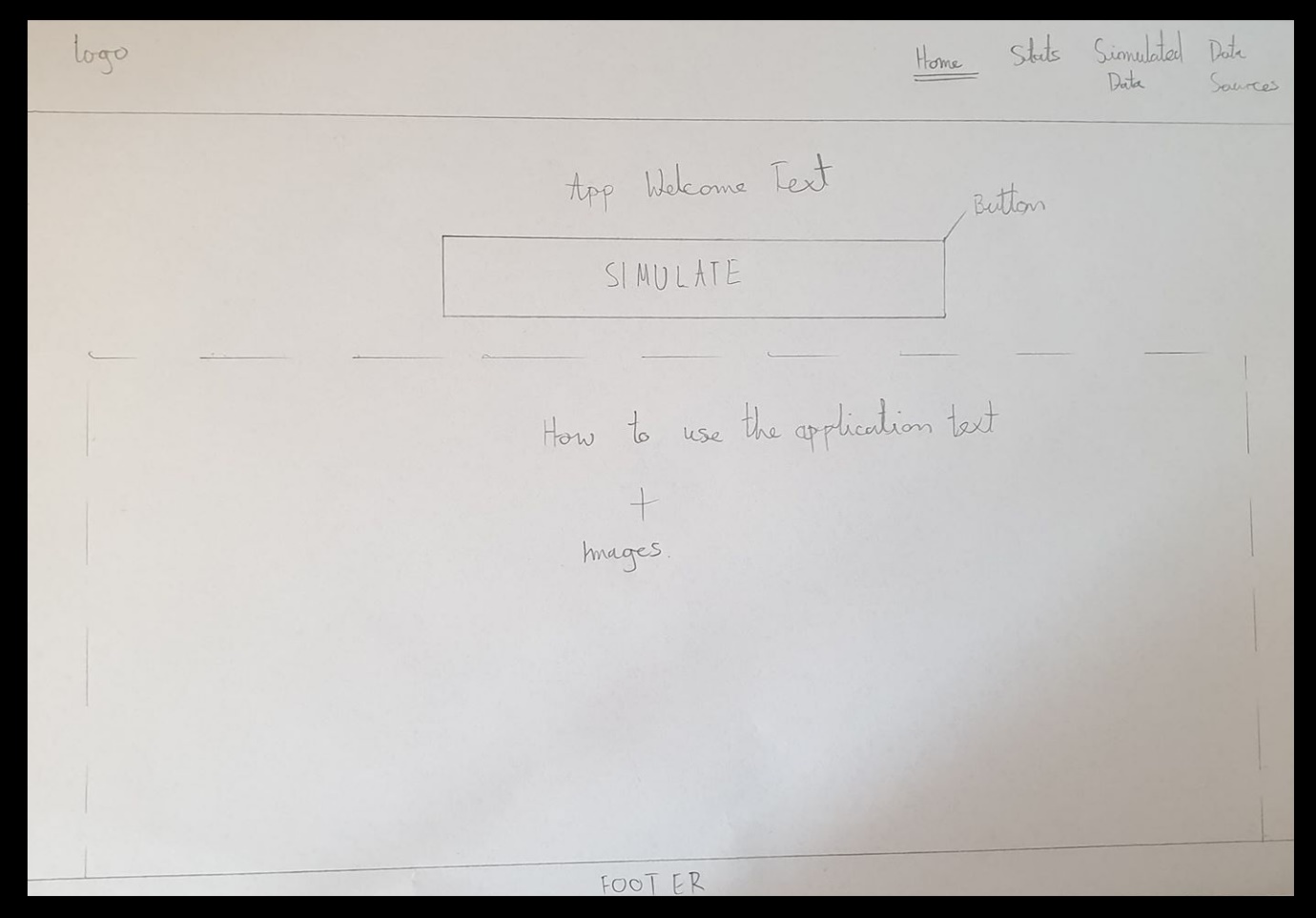
The rest of the keys default values are just initializations for the persons keys.

From there all the other statistics would be applied to the person objects.

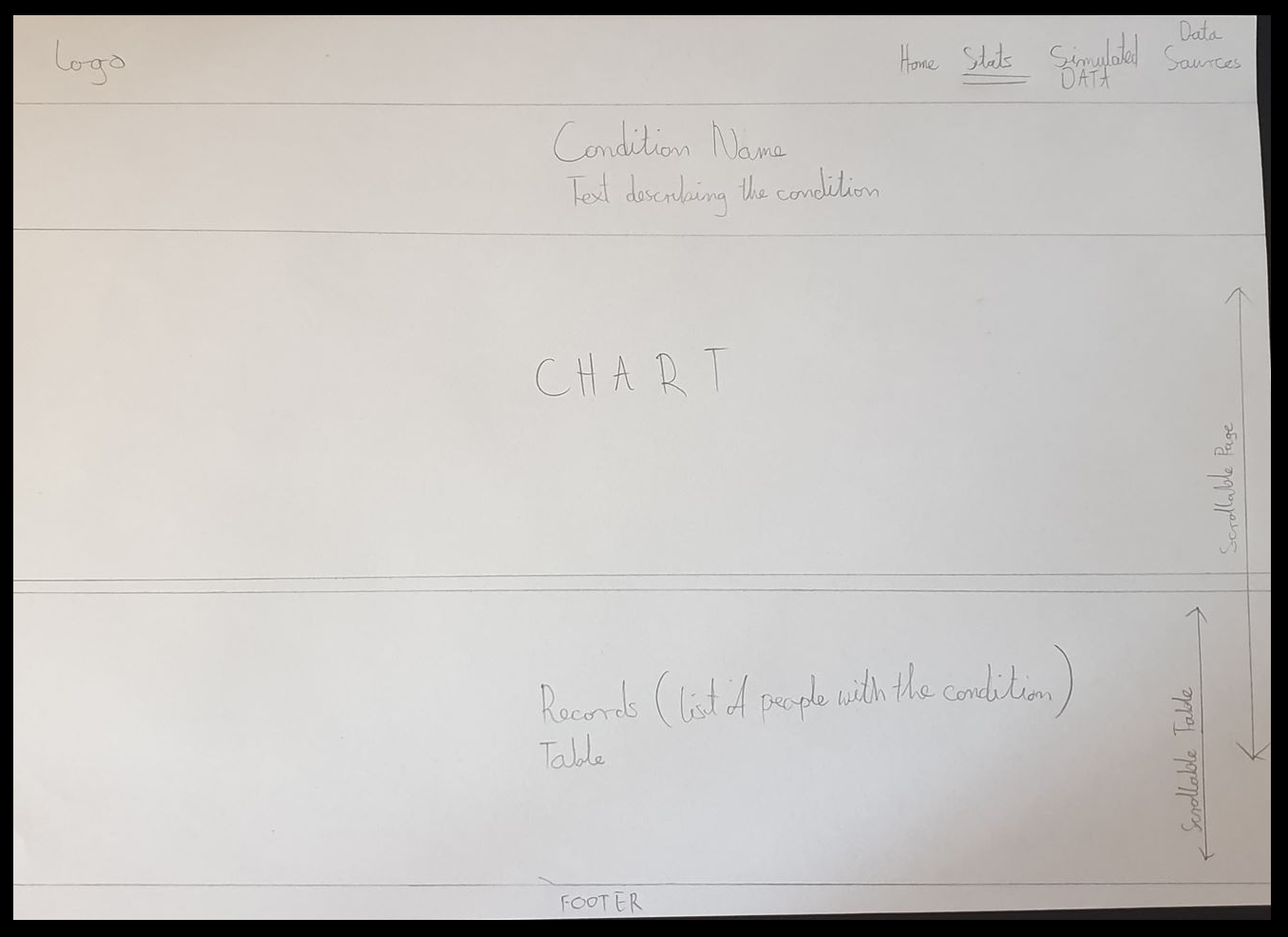
Finally, after all simulation has been completed. The data set would be converted to a JSON string and later stored in local storage so that data can be accessed by the user at any time.

### 5.4 Graphical User Interface Design

Before coding a decision was made to design the web pages before implementation. This is because less time would have to be spent thinking on how to make a web page look. Below is a template that was drawn for the home page of the application. I have made the “SIMULATE” button large in size so that the user will have a straight idea that they have to click on it in order to start the application.



In the image below a statistics page for a health condition shown. The chart and the records table look large on the web page as I want them to stand out, so it makes readability easier of the data. The user will be able to scroll down the page until the end of the table is shown. However, the table is going to be made scrollable on its own so that the page itself isn’t forever scrollable.



Chapter 6

## Detailed Design and Implementation

The aim of this chapter is to describe in detail what programming languages, tools, environment has been chosen in order to develop the application. In the final topic I will discuss the challenges I had to face in the design and implementation of this application.

### 6.1 Choice of Programming Languages Used

For this application development the following programming languages have been used.

#### 6.1.1 HTML5

HTML5 is a web development language that defines the properties and behaviours of web page content by implementing a mark-up-based pattern to it. HTML5 was used for the project in order to display web pages to the user since that’s currently the industry standard for display web pages to users in a web browser.

#### 6.1.2 CSS

CSS is a style sheet language used for describing the presentation of a document written in mark-up language such as HTML. Most of the styling such as the navigation bar on the top of the page has been created in CSS.

#### 6.1.3 JavaScript

JavaScript is a high-level, interpreted programming language. I have used JavaScript as my main back-end language as I am familiar with it and it’s a flexible language. JavaScript is a flexible programming language as over the years people made more and more libraries for different problems available to developers.

### 6.2 Choice of Third-Party Tools

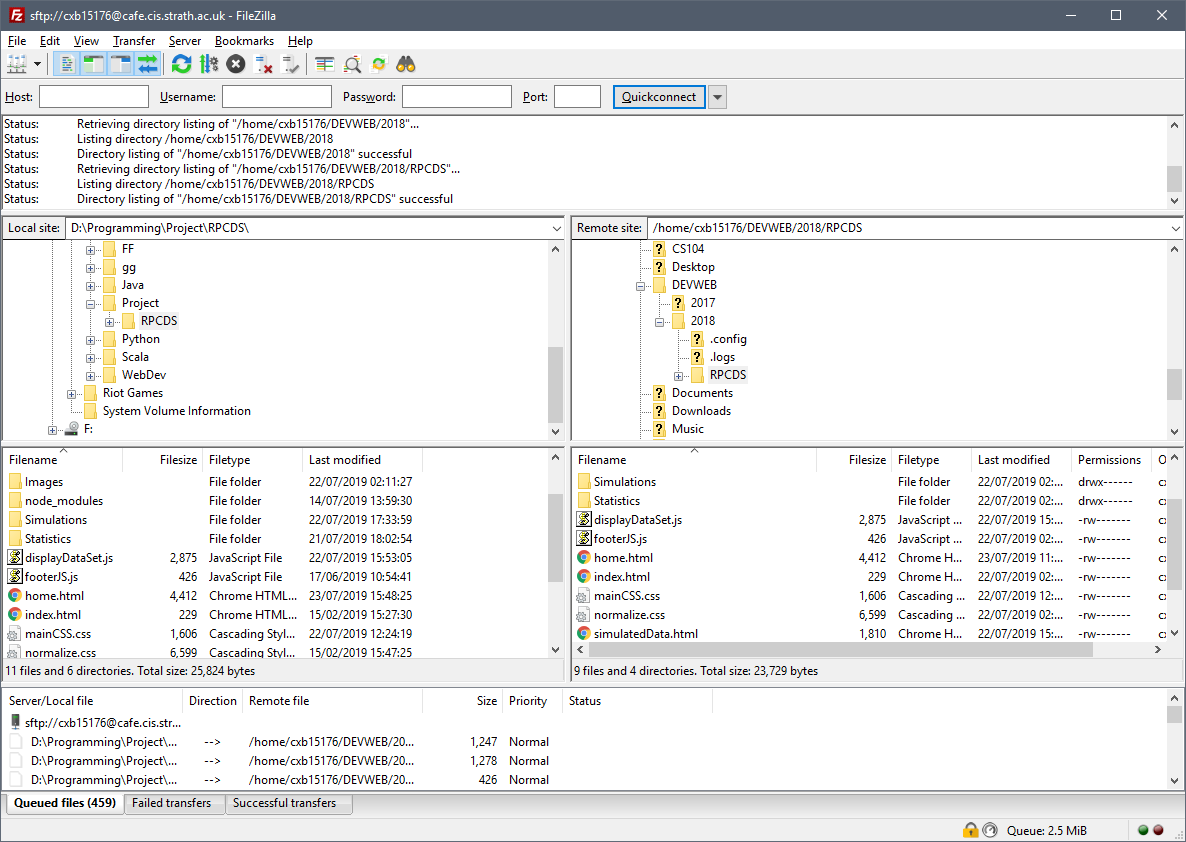
For this application I have used the following third-party tools.

#### 6.2.1 Strathclyde DEVWEB

I have used Strathclyde DEVWEB which is a university web server that allows students to host student work on it. I have decided to host my application on this web server since its free to use for students. The other reason for hosting it on this service is that people can access the application just by accessing it through a web link. This results in users being able to access the application on various devices.

#### 6.2.2 FileZilla

FileZilla is a free software which is a cross-platform FTP application. Since I was developing my project from home and wanted to host it on Strathclyde’s DEVWEB I had to FileZilla. FileZilla allowed me to access university’s web server from home, so I didn’t have to travel to university to put my project on my university computer account.



In the above image the application is displayed with project files on my computer (left side of the application) and project files hosted on Strathclyde’s DEVWEB (right side of the application).

#### 6.2.3 ag-Grid

ag-Grid is an open source JavaScript framework which allows developers to display data grids without the need of doing all the data formatting on your own. Before finding out about this framework I have developed my own data display grid however, when I have tried to add features such as data sorting, I have encountered a problem. The problem was that since I was trying to sort 2000 rows of data it would either crash the website or it would take few minutes to do it since there is so much data. Therefore, I had to look for other methods of display my data. After doing research of this problem, I have found ag-Grid and decided to use it for my application. ag-Grid has various built in customization such as table sorting, data filtering, data export and many other features.

#### 6.2.4 Chart.js

Chart.js is an open source JavaScript framework which allows developers to display simple and clean HTML5 based charts. This framework is simple to use as it uses HTML5’s canvas to display the chart. Chart.js provides a wide range of features such as various graphs available and animation. I have decided to use this framework because it supports charts on mobile devices.

#### 6.2.5 Bootstrap

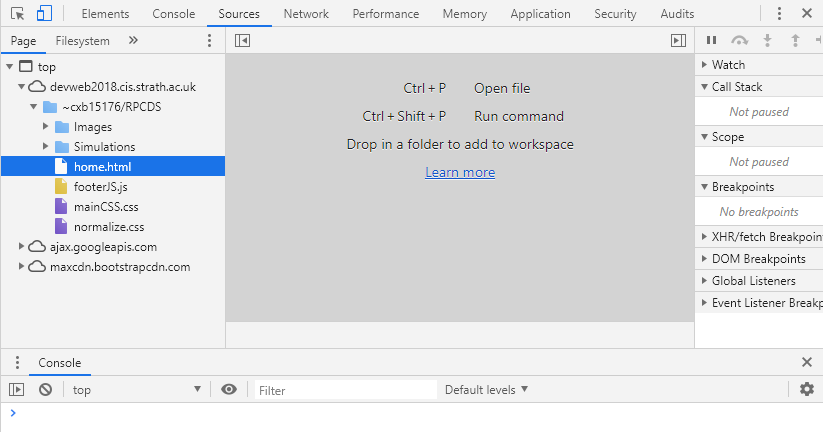
Bootstrap is an open source CSS framework which provides developers with high quality design templates available to use. It contains CSS and JavaScript design templates for typographs, forms, buttons, navigation and other interface components. I have decided to use this framework as the design templates look professional and save me time developing the application as I had to spend less time on styling.

#### 6.2.6 jQuery

jQuery is a lightweight “write less, do more” JavaScript library available to developers. I have used jQuery in my project in order to access the Bootstrap framework so I could use their design templates. Without jQuery I would have to write more complicated code just to access that framework.

#### 6.2.7 Google Chrome

Google Chrome is a popular cross platform web browser. This web browser provides developers with an “Inspect” tool shown in the image below. This tool is useful to developers as they can see what is happening in their website by seeing what is loaded by it. There is also a built-in console that messages can we outputted from the website which makes it a good tool for quick testing. Since my application is using local storage for data storage, I can check how much data is stored by clicking on the “Application” tab.



### 6.3 Choice of Development Environment

For this application I have used the following environments.

#### 6.3.1 Operating System: Windows 10

Windows 10 is a popular computer operating system. I have decided to use this operating system instead of Linux because although I have experience using Linux, I prefer using Windows operating systems as I am more used to their operations.

#### 6.3.2 Mobile Operating System: Android 9

The mobile that I am using uses Android 9 as its mobile operating system. I have used my mobile phone for the development of this application in order to test its mobile application display. On the phone I have used in the built in “Internet” browser and Google Chrome in order to test the mobile functionality of the application.

#### 6.3.3 IDE: Visual Studio Code

As my choice of Integrated Development Environment was Visual Studio Code. This is a source-code editor developed by Microsoft. I have decided to use this as my editor as it has embedded Git control and GitHub, syntax highlighting, intelligent code completion, snippers and code refactoring.

#### 6.3.4 Version Control: Git and GitHub

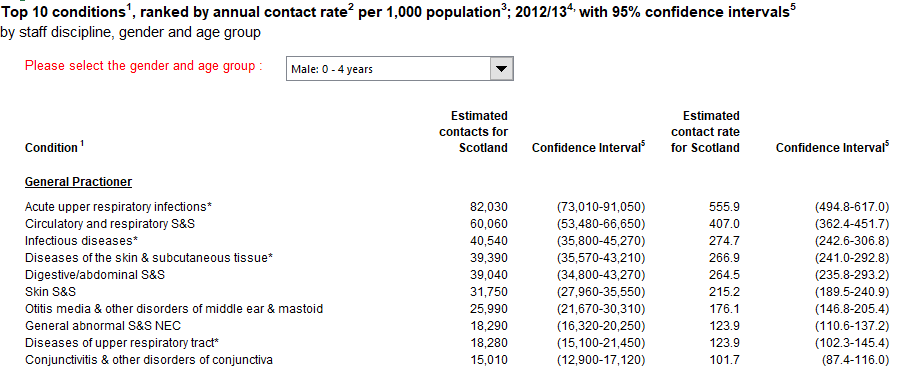
Version control is a system that records changes to a file or set of files over time so that at later stages the specific version can be recalled. Git is one of those systems. GitHub is a service that provides for hosting for software development version control using Git. I have decided to use those two systems as that’s the industry standard currently.

### 6.4 Challenges of Design and Implementation

During the development of the application I have encountered the following challenges.

#### 6.4.1 Data Analysis and Clean-Up

Data analysis and clean-up was a huge challenge in the development of this application. This is because there isn’t a set standard for statistics therefore, organizations release them in different formats.



Above is an image of statistics that I have used for some of the conditions. I had to take those numbers that have been provided to me and convert these into percentages, so I knew the percentage of people that were affected by certain conditions based on age and gender. This was process was time consuming as I had to make sure that my calculations were correct in order for those statistics to be useful.

#### 6.4.2 Storage of Large Data in Local Storage

At the start of the implementation of the project I would only simulate 200 people to test the application. However, by the end of the implementation stage I wanted to increase the number of people to be simulated to 4000. After setting that, when I’ve clicked the simulate button the website would crash as there would be too much data being simulated for a website. Therefore, I had to reduce that number to 2000 people.

#### 6.4.3 Manipulation of Large Volumes of Data

Manipulation of Large Volumes of Data was a huge challenge too. Since I had 2000 simulated people, I wanted to have a feature where user could do simple data manipulation like sorting and filtering. However, since I was using just a standard HTML table to display that data, sorting and filtering would take way too much time. Therefore, I had to use something else rather than just a standard HTML table. I have decided to use ag-Grid which is a JavaScript library that creates a data grid for the developer’s data. After putting my simulated data into that data grid, it turned out well as data sorting and filtering was instant.

Chapter 7

## Verification and Validation

The aim of this chapter is to describe how the project was verified and validated. For more detailed testing strategy and test cases please look at “Appendix B”.

### 7.1 Verification

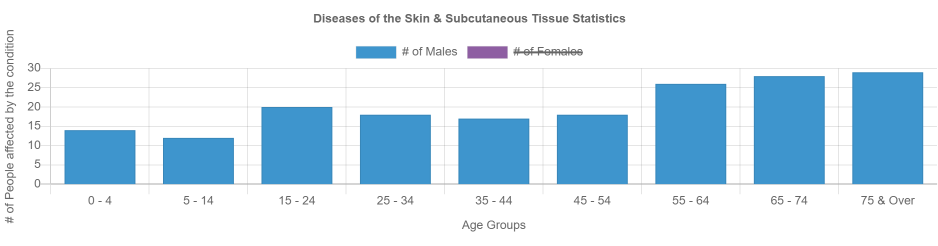
A decision was made to test the application using black box testing. Black box testing is where tests are carried out without knowing any technical information about the application. Test cases were carried out by deploying the most recent application files to the Strathclyde’s DEVWEB server. Test scenarios were created for each feature of the application. Black box testing is good since there are no biases towards the way the system was developed.

### 7.2 Validation

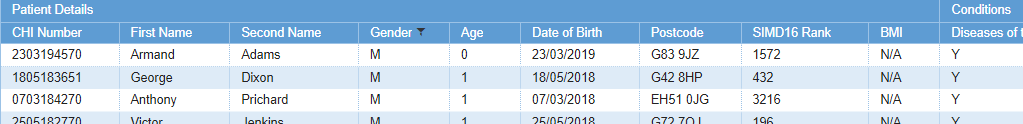
The application meets the specifications that I have set out at the beginning of the project. The application was validated by checking if each of the functional and non-functional requirements were implemented into the application. There was a lot of debugging in order to make sure that the simulated was appropriate.

In the chart below is the statistics on male “Diseases of the Skin & Subcutaneous Tissue” that was created before implementation but after data clean up stage.

The next chart was created by the application on the same statistic.



Comparing the two charts above it is shown that the bar structure is similar. However, not the same but this is due to the error rate that was implemented. Overall, the simulated graph by the application is valid as the statistics look realistic. Below an image is show which shows simulated people based on the simulated graph above.



This table shows that the application solves the problem of simulating random patient clinical data as patients are randomly simulated and statistics on these simulated patients can be gathered.

Chapter 8

## Results and Evaluation

The aim of this chapter is to describe how the project was evaluated and what the results were. This will include user evaluation and feedback, evaluation of the original project plan and the evaluation of the final product.

### 8.1 User Evaluation

The aim of the user evaluation was to ensure that the application solves the problem. Therefore, after the application was completed to a satisfactory degree where only a series of slight styling changes still to be made, the application was prepared for testing of a series of short scenarios. The user evaluation consisted of an idea where I would send out e-mails to university students (e.g. 4th year friends) and my supervisor. From there they to get a wider range of responses they could pass on the invitation to other people they know. My target users were mainly people who have a basic knowledge of data analytics which is basically looking at a graph and if it provides enough information to the user and from where that data has come from.

However, before I could send my invitations out, I had to submit an “Ethics Application Form” to the Strathclyde Ethics Committee. I had to do this to prove that I am not using volunteer’s identity or survey responses for any personal reasons other than the project user evaluation.

During the creation of the survey I have decided that I will not take volunteers names as all I wanted from the them would be the answers to the survey questions on the application and simulated data. This decision was taken to protect volunteer’s anonymity. However, before taking part in the survey, in the invitation e-mail there would be a “Participation Document” which gives potential volunteers information on what is the application survey is about and how I will handle their information. In the e-mail I have also stated that they should provide a signature at the bottom of that participation document just to make sure that they have read the contents of the document and so that they will know in detail how that information provided will be handled by me. After signing the document, I have asked them to send that document back to my university e-mail.

Please see “Appendix D” for the “Participant Consent Form”.

The main aim in the creation of the application survey was to create a survey which took 5 – 10 minutes to complete and to pull out as much information from the volunteer as possible. Therefore, I have decided to use Strathclyde’s Qualtrics Website.

Please see “Appendix E” to see the user survey that was sent to the volunteers.

Qualtrics is a website designed for user friendly and web-based surveys. It only took me few minutes in order to get familiar with the website and tools available to me which helped me to design questions for the users. In the table below I have shown what questions I have asked the volunteers. For those questions the volunteers had either a slider or a multiple-choice statement to choose their answers to the questions. In the survey the volunteers would be provided with detailed instructions for each of the questions to get them familiar with the website user interface. At the time of writing this chapter I have received 11 recorded responses from the survey.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Question** | **Average Rating** | **Minimum Rating** | **Maximum Rating** |
| 1 | The look of the “Home” page | 3.64 | 2 | 5 |
| 2 | Was the simulation timing reasonable? | 3.64 | 2 | 5 |
| 3 | Was “Simulated Data” page helpful? | 4 | 1 | 5 |
| 4 | Was “Simulated Data” page well displayed? | 4.45 | 4 | 5 |
| 5 | Was it easy to access data on health conditions? | 4 | 4 | 4 |
| 6 | Was the chart on the health condition well displayed? | 4.36 | 3 | 5 |
| 7 | Was the data table on the health condition well displayed? | 4.36 | 3 | 5 |
| 8 | Could the health condition chart be saved as an image? | 4 | 4 | 4 |
| 9 | When clicked on a data set of the chart did the data set change? | 4 | 4 | 4 |

For the survey questions listed above the questions that I have received the worst responses where questions 1, 2 and 3. For question 1 I can admit that the home page could look a bit better since it looks cluttered. However, I have weak skills for creating styles for web pages. For question 2 some volunteers felt that the simulation timing was taking too long. This is because the application access and loads a large file with 300 thousand rows of data in order to provide data on Scottish postcodes and SIMD16 Ranks. For question 3 some volunteers gave a low rating as data simulation on Scottish Health statistics is a niche area of interest. This is because the people who would be interested in this type of data would be people who are interested in data analytics or statistics on the Scottish Population.

Overall, for the above questions the feedback from the survey volunteers was positive.

Other than the questions above I have also asked volunteers the following textual feedback questions:

1. Is the “Data Sources” page helpful or more details should be added to it?
2. What is the most important improvement you feel this application needs?
3. How does the simulated data produced by the application help you?

For question 1 I have overall received positive responses. Below is one of the responses:

**“I think this page is useful as it provides sources for all the information used within this application.**

This volunteer has stated that the “Data Sources” page was a useful feature to the application as they can see what sources were used in the application. Another response was:

**“The data sources page provides all the needed information except for the now dead link relating diabetes.”**

This volunteer has that all the important information is provided on the page. However, the diabetes source link doesn’t work anymore as Diabetes UK have taken down that file and its no longer available. At the requirements stage and implementation stage the data was still available but that was back in March 2019. If I had more time, I would look for another source for that health condition but due to the amount of time it takes to analyse the data I wouldn’t be able to complete the project on time.

For question 2 volunteers overall have suggest good improvements to the website. One of them was:

**“I think it would be useful to allow users to export the data for use within an application like Microsoft Excel. It would make the transfer of information between the site and local much, much faster!”**

This volunteer has suggested to add the export feature to the website. Since I still had a bit of time left before submission, I have decided to add that feature in. I have decided to add that feature in as it makes the application more fit for purpose as the main aim of it is to simulate data and provide it to the user for their uses. Another response for this question was:

**“The ability to add new statistics easily to the applications.”**

This is another great suggestion that could be added to the application. The application could have this functionality as I have made it easy to add new statistics. However, for this project I have decided to stick with the main specifications. This feature would be a great addition to the application for future work.

For question 3 volunteers have provided mixed responses. One of the responses was:

**“It allows for the generation of a large number of patients all with differing conditions. This random generation of data could aid the teaching of statistical analysis for students.”**

Whereas another volunteer stated:

**“I don't find it helpful. Sorry.”**

This question has received mixed responses because not everyone is going to be interested in data on health conditions that affect the population of Scotland. However, in this survey there were still some volunteers that are interested in that area.

### 8.2 Evaluation of Planning

For this project this was the original project plan with timescales:

**24th October – 21st November (4 Weeks)**

The most important part of this project is to gather all requirements before starting to develop the application as otherwise there is going to be time wasted on changing values created before which might not even be anywhere close to the real-world statistics.

* At this stage of the project the time is going to be used to gather all the requirements to generate usable data.
* After gathering all the requirements design of the application can take place.
* After having a design of the application early stages of application development can take place.
* Research what is the best way to generate statistical data and visualise it.

**21st November – 1st of December (1 Week)**

* Background review chapter draft submitted to supervisor.

**1st December – 15th of December (2 Weeks)**

* Problem specification / Software specification Chapter submitted to supervisor.
* Prepare project progress presentation.

**15th December – 26th January (6 Weeks)**

* Start programming the project.
* Implement random patient data details without any conditions.
* After having random patient data created conditions will be added and randomized based on statistics.

**26th January – 15th February (3 Weeks)**

* Implement data visualization of patient’s data using appropriate graphs.

**15th February – 22nd February (1 Week)**

* Software testing.

**22nd February – 1st March (1 Week)**

* Software development chapter submitted to supervisor.

**1st March – 15th March (2 Weeks)**

* Analytic evaluation chapter submitted to supervisor.

**15th March – 22nd March (1 Week)**

* Full report draft submitted to supervisor for comments.

**22nd March – 25th March (3 Days)**

* Submit the project and the project report.

Due to personal circumstances I have gotten an extension to the project deadline which changed the dates above. However, the times stated to complete each of the tasks didn’t change too much. The

Since for this project I have chosen the Waterfall Model methodology I had to gather my requirements for the application before doing anything else. This is because without those statistics I wouldn’t be able to do the project as that’s the main requirement of it. I have stated that for this stage it would take 4 weeks to complete that task and it did take me that much time to gather all the requirements. It took me this much time to complete this task because I was searching for statistics that were high quality and detailed as I was mainly looking for statistics that gave me an age rage and the number of people affected by the condition based on gender.

The next important tasks in the plan were programming and data visualization. The implementation of data simulation did take me 6 weeks to do it since I was testing the application each time I have uploaded it to the deployment server (Strathclyde DEVWEB) to make sure that the statistics were simulated correctly. For the visualisation of data in the format of a chart and a table with all the records I have stated that it will take me 3 weeks to do it. However, the actual time taken to complete this task was only a week since I have used a third-part libraries to display that data to the user.

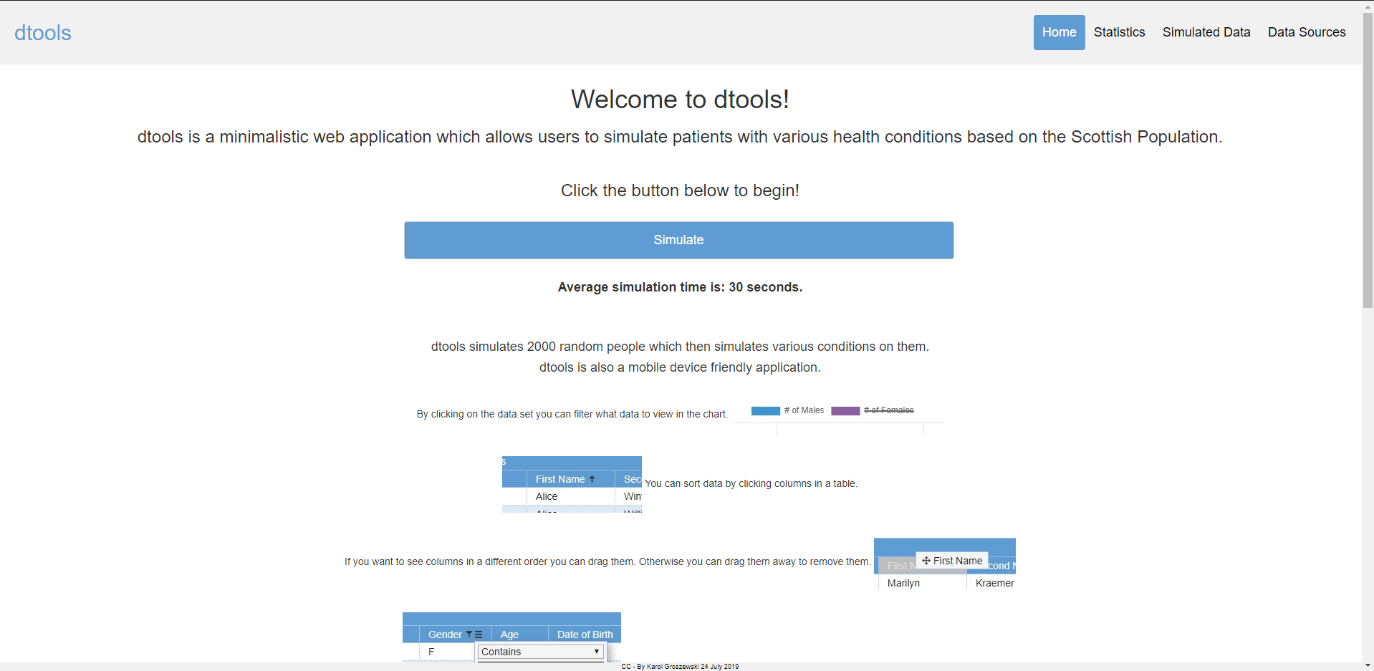
The report writing tasks were hard to judge due to my university exams and personal circumstances but to write the report I had to half each of the time scales to complete it on time.

Overall, I have stuck with the initial plan to complete this project and I think I have judged the time to complete each of the tasks correctly as I gave myself enough time to complete each of the tasks.

### 8.3 Evaluation of End Product Functionality

The resulting web application created by this project was successfully created due to the high-quality statistics that were available to me to use to develop this application. Without having these statistics, the application wouldn’t be able to provide users with detailed information on health conditions affecting the Scottish population which would result in the loss of users interested in the application.

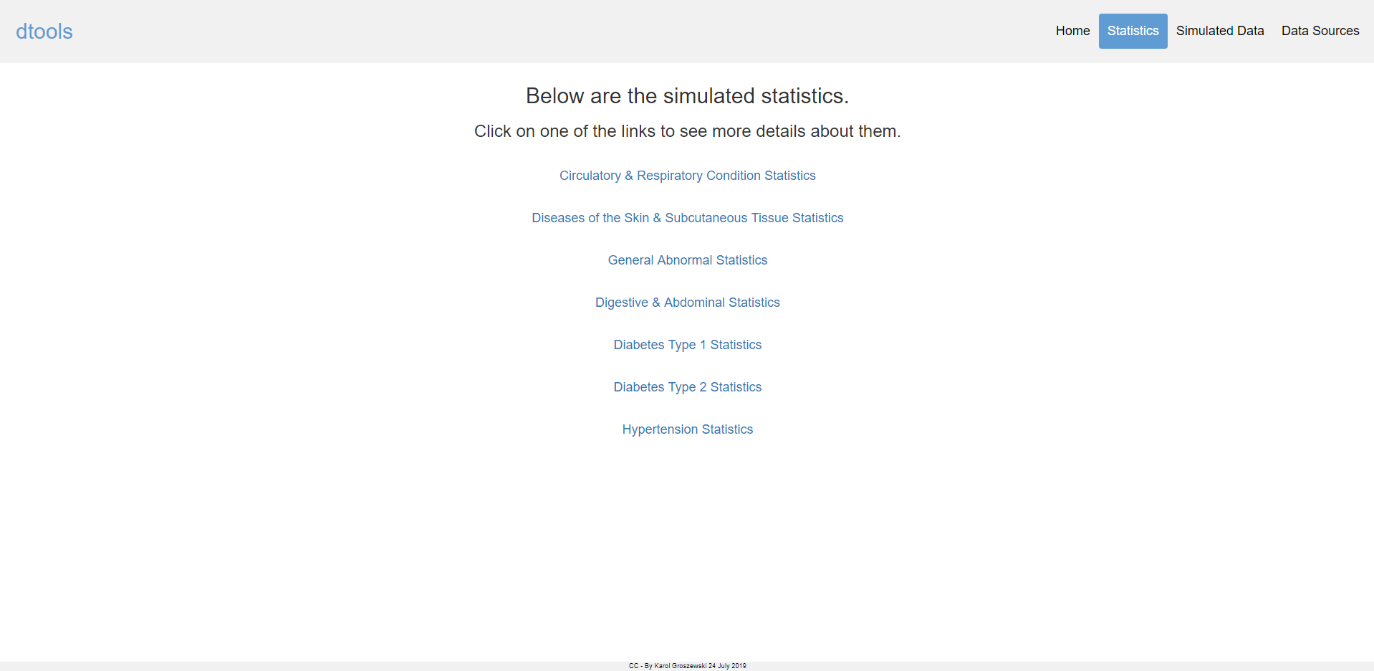
I have decided to produce a web application that supports both local computers and mobile devices. This application meets this requirement as the user can simulate data on various devices available to them.



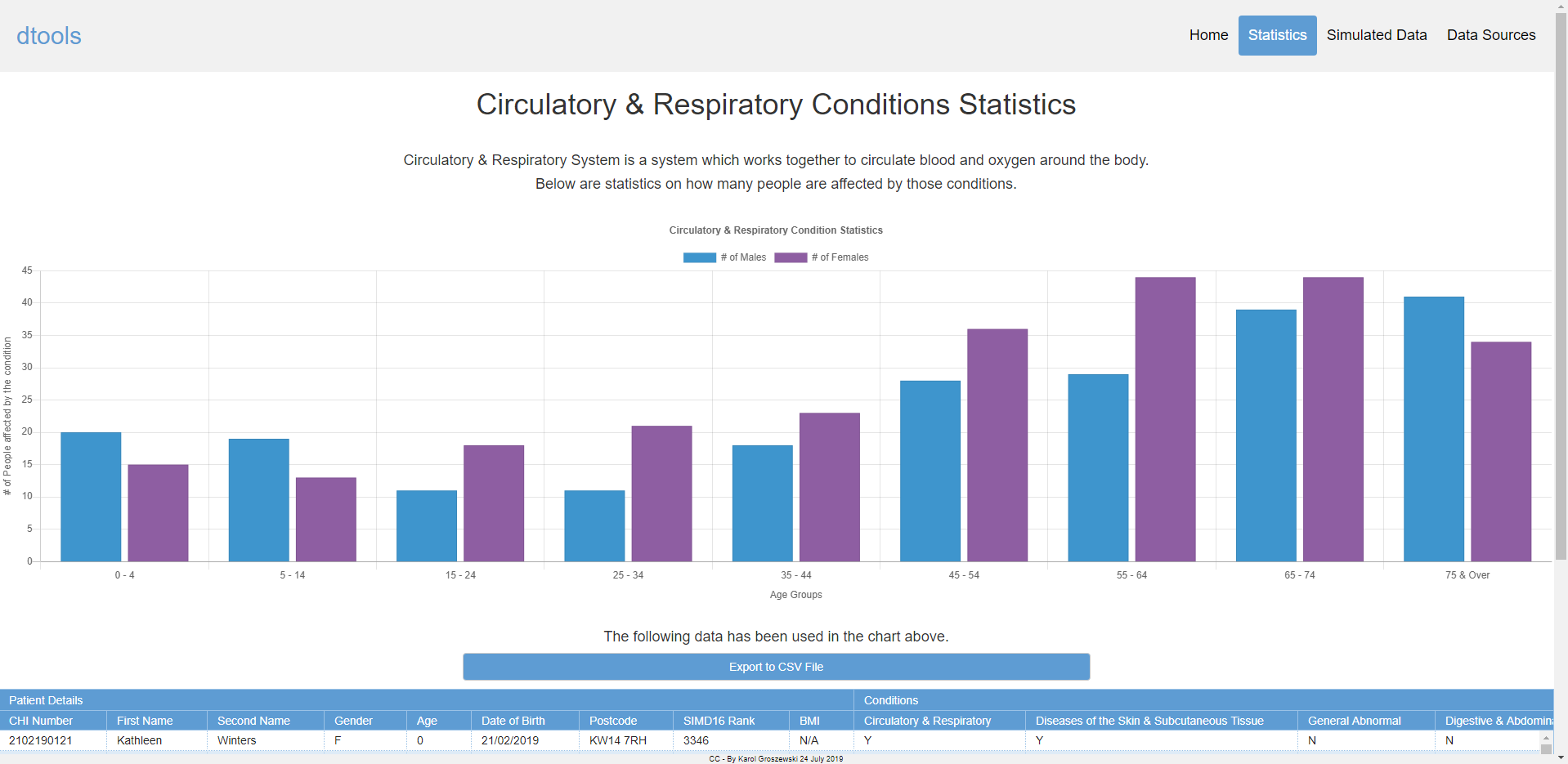
The image above is the screenshot of the applications “Home” page. I like minimalistic designed applications since they are easy to navigate for new users therefore, I have decided to take the same approach here to keep the design on the home page as simple as possible. To help the user out with what the application to them I have added pictures to the page and descriptions on what the user is able to do in this web application to improve user friendliness.

I have made the “Simulate” button large as I wanted the users to know straight away that they need to click this to begin. Since the application is loading in and simulating a lot of data, I have included a average simulation time to make sure that the user needs to wait for a short while in order for the application to simulate data for them. When the user clicks on the “Simulate” button the average simulation time text changes to a different text telling the user that after the data has simulated, they will be redirected to the “Simulated Data” page.

I have added this redirection since the purpose of this application is to simulate data and after simulation they would most likely click on the “Simulated Data” page anyways.



In the image above “Statistics” page is shown. This is a nice and simple page as the user has large headings about what is on this page and what to do next. I have made links large to so its easy for the user to click on them. Although I like designing minimalistic web pages this page looks too empty. However, since I wanted the application to support mobile devices too, I had to keep this page the way it is shown above so it doesn’t look too cluttered on mobile device displays.

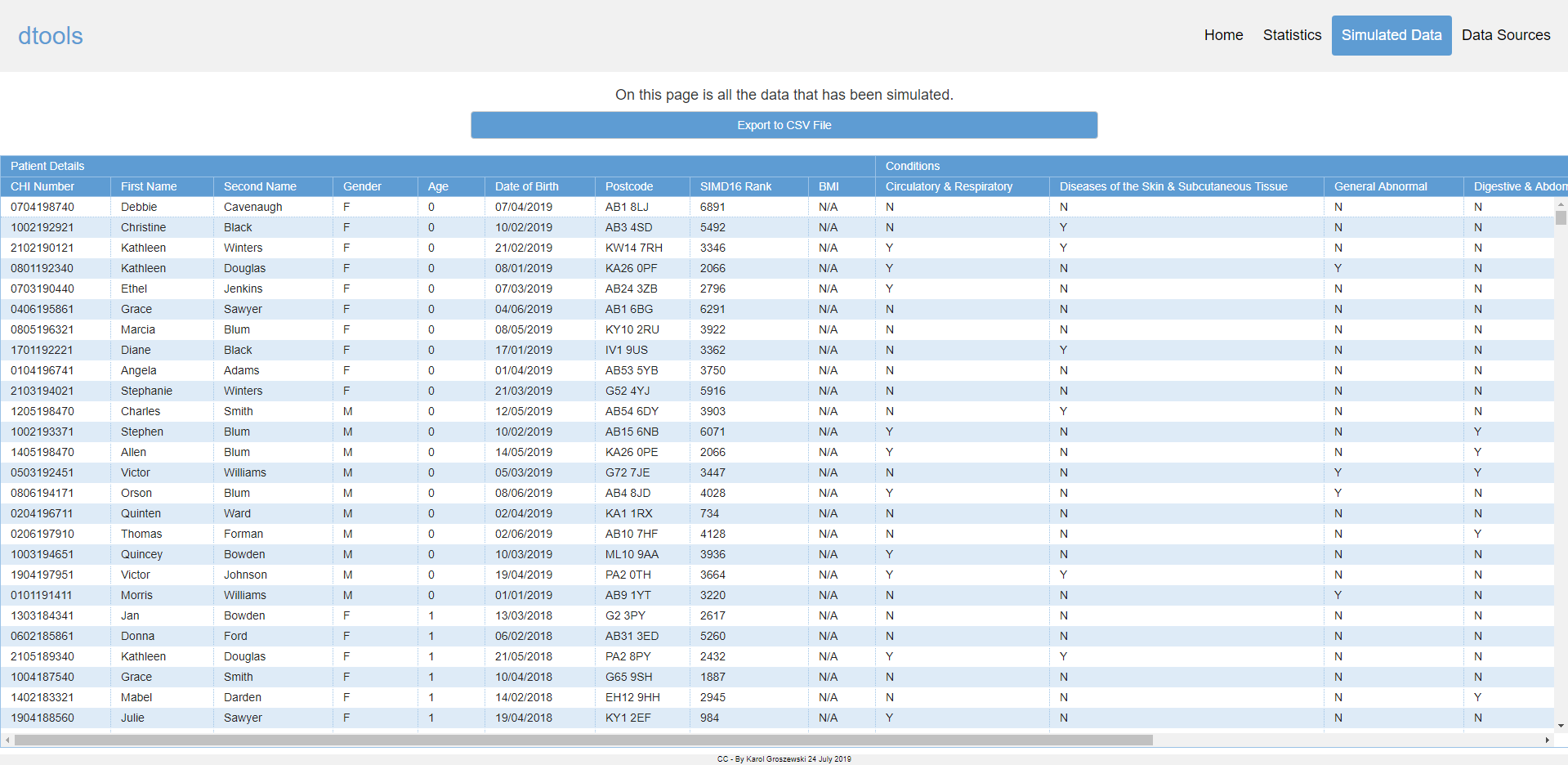


In the image above a “Circulatory & Respiratory Conditions Statistics” page is shown which can be accessed from the “Statistics” page. This page has a big heading showing the user what statistics are shown to the user on this page. Below the main heading there is short description on what the condition is just to improve the feel of the site.

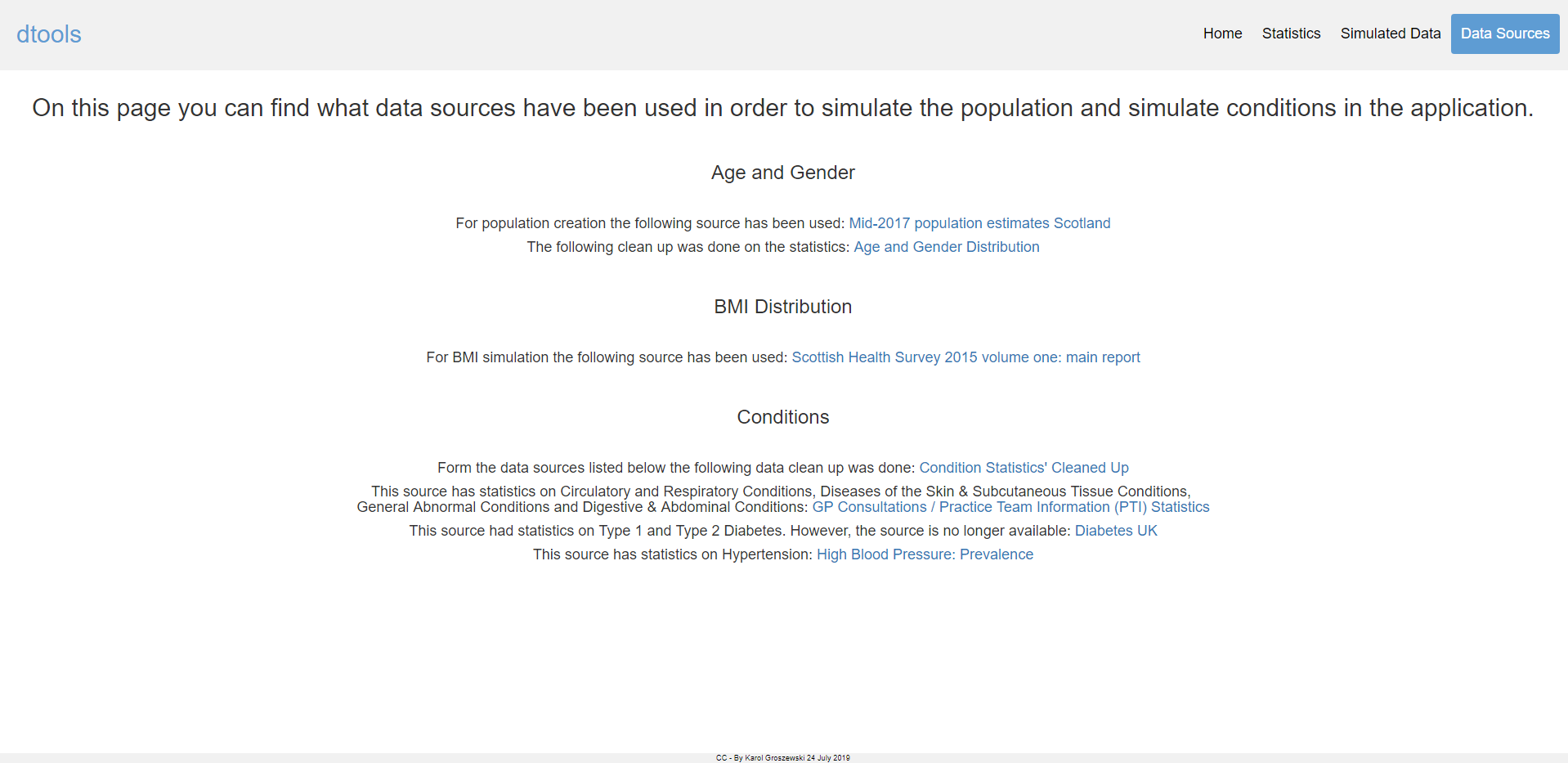
The bar chart shown to the user is large. I have decided to make the chart this size to make sure that the simulated data is readable. After clicking on the data set (e.g. # of Males) the data shown to the user on the chart will change. This adds more functionality to the application and use. As for example some users might just be interested in seeing only data on the female population of Scotland. Another additional functionality of this chart is that while hovering over a bar of data with the mouse, the user will be able to see how many people are affected by the condition in the age group.

From the user evaluation survey, one of the volunteers suggested that the simulated data should be exported to a local file so then the user can use it for other purposes. I have decided to add that feature to the site since I still had enough time to implement it before the submission of the project.

When the page is scrolled down the table of records is shown to the user with all the simulated people who are affected by the condition. The user has various options to edit the table. By clicking on the column, the records are going to be sorted by the user’s choice. When a user hovers over their mouse over the column a filter button will appear, which will filter out the table to show what data the user wants. If user isn’t interested in peoples age for example, they can click and drag that column off the table to remove it from the view.



In the image above “Simulated Data” page is shown which displays all the data that has been simulated by the application. Again, the user has the same options as in the health condition specific page just without having a chart displayed.



In the image above “Data Sources” page is shown which displays to the user what sources have been used to simulate the data. This page is useful since a user might be interested in the data or would like to validate statistics available to the user. Therefore, I have decided to add these references that I have used to this page.

Overall, I am happy with the way the final product turned out to be as it is simple to use for users and it simulates realistic data based on the statistics gathered from the requirements stage of the waterfall model methodology.

Chapter 9

## Summary and Conclusions

The aim of this chapter is state my summary of the project and its conclusions.

### 9.1 Summary

To summarize this report, the application that was developed is a success. This is because the application succeeds as it provides the functionality necessary to create simulated patient data that would allow third parties to see and explore the data to their use. The final project is a great example how data should be simulated for medical and research purposes as the data can be analysed and exported by users for their needs.

### 9.2 Issues and Limitations

The issue with the application is that the number of people to be simulated is limited to 2000 records as otherwise if the number of people is higher the website would crash. I believe that the website is crashing because this data that is simulated is trying to be stored in local storage but there isn’t enough storage for it.

The limitation within the application is that one of the links to the data source is dead as the provider took down the statistics file. The limitation could be solved by finding another source for that statistic. However, due to time constraints it would take too long to find a new source and manually clean up the data to be used in the application.

### 9.3 Future Work

For future work there could be a few new features that would improve the application. One of these features could be where a user can simulate a higher number of people that it currently is (2000 records) and let the user decide how many records to simulate.

Another feature that could be added to the application is where the users are able to enter their own statistics. Right now, the statistics that are entered are all given to the user, but they cannot enter their own to customize it.

### 9.4 Final Conclusion

To conclude the project succeeds in achieving the requirements set at the start of this project. This project circumvents difficulties and administrative tasks presented when attempting to acquire the necessary patient data for a study. The resulting project has created an application that allows researchers and medical personnel to simulate patient data that can be used for their research needs.

Appendix A

## References

1. Cant, R. P. and Cooper, S. J. (2010), Simulation‐based learning in nurse education: systematic review. Journal of Advanced Nursing, 66: 3-15. doi:[10.1111/j.1365-2648.2009.05240.x](https://doi.org/10.1111/j.1365-2648.2009.05240.x) , [Accessed: 12th March 2019]
2. Information Services Division Scotland, Community Health Index (CHI) Number, [Online], Available from: <https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/Patient-Identification-and-Demographic-Information/Community-Health-Index-Number/> , [Accessed: 12th March 2019]
3. National Records of Scotland, Mid-2017 population estimates Scotland, [Online], Available from: <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2017> , [Accessed: 12th March 2019]
4. Scottish Government, Scottish Health Survey 2015 volume one: main report – 8: Obesity, [Online], Available from: <https://www.gov.scot/publications/scottish-health-survey-2015-volume-1-main-report/pages/71/> , [Accessed: 12th March 2019]
5. Information Services Division Scotland, GP Consultations / Practice Team Information (PTI) Statistics, [Online], Available from: <https://www.isdscotland.org/Health-Topics/General-Practice/GP-Consultations/> , [Accessed: 12th March 2019]
6. Diabetes UK, Version 3. Revised March 2014, [Online], Available from: <https://www.diabetes.org.uk/resources-s3/2017-11/diabetes-key-stats-guidelines-april2014.pdf> , [Accessed: 12th March 2019]
7. ScotPHO, High Blood Pressure: prevalence, [Online], Available from: <https://www.scotpho.org.uk/clinical-risk-factors/high-blood-pressure/data/prevalence/> , [Accessed: 12th March 2019]

Appendix B

## Detailed Test Strategy and Test Cases

To test the application black box testing was used which is shown below.

|  |  |
| --- | --- |
| Test | Simulate data on the home page. |
| Description | Test to ensure that the website simulates the full data. |
| Preconditions | The user must be on the “Home” page. |
| Steps | 1. Click on the “Simulate” button. 2. Wait to be redirected to the “Simulated Data” page. |
| Expected Result | The user gets redirected to the “Simulated Data” page and all the simulated data is displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Simulated Data” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Simulated Data” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test if people with the following condition are displayed only on the “Circulatory & Respiratory Conditions Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “Circulatory & Respiratory Conditions Statistics” page. |
| Steps | 1. Open the “Circulatory & Respiratory Conditions Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Circulatory & Respiratory Conditions Statistics” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Circulatory & Respiratory Conditions Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |
| Test | Test if people with the following condition are displayed only on the “Diseases of the Skin & Subcutaneous Tissue Conditions Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “Diseases of the Skin & Subcutaneous Tissue Conditions Statistics” page. |
| Steps | 1. Open the “Diseases of the Skin & Subcutaneous Tissue Conditions Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Diseases of the Skin & Subcutaneous Tissue Conditions Statistics” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Diseases of the Skin & Subcutaneous Tissue Conditions Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test if people with the following condition are displayed only on the “General Abnormal Conditions Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “General Abnormal Conditions Statistics” page. |
| Steps | 1. Open the “General Abnormal Conditions Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “General Abnormal Conditions Statistics” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “General Abnormal Conditions Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test if people with the following condition are displayed only on the “Digestive & Abdominal Conditions Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “Digestive & Abdominal Conditions Statistics” page. |
| Steps | 1. Open the “Digestive & Abdominal Conditions Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Digestive & Abdominal Conditions Statistics”. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Digestive & Abdominal Conditions Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test if people with the following condition are displayed only on the “Diabetes Type 1 Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “Diabetes Type 1 Statistics” page. |
| Steps | 1. Open the “Diabetes Type 1 Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Diabetes Type 1 Statistics” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Diabetes Type 1 Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test if people with the following condition are displayed only on the “Diabetes Type 2 Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “Diabetes Type 2 Statistics” page. |
| Steps | 1. Open the “Diabetes Type 2 Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Diabetes Type 2 Statistics” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Diabetes Type 2 Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test if people with the following condition are displayed only on the “Hypertension Statistics” page. |
| Description | Test to ensure that all the people with the type of condition are displayed. |
| Preconditions | The user must be on the “Hypertension Statistics” page. |
| Steps | 1. Open the “Hypertension Statistics” page. |
| Expected Result | People that are affected by the following condition should only be displayed. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Test | Test export to CSV on the “Hypertension Statistics” page. |
| Description | Test to ensure that the data displayed gets exported to a CSV file. |
| Preconditions | The user must be on the “Hypertension Statistics” page. |
| Steps | 1. Click on the “Export to CSV File” button 2. The file should be downloaded. |
| Expected Result | The exported CSV file should be exported to the users “Downloads” folder or will be promoted asking where the file should be saved. The exported file will contain the statistics shown on the page. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Description | Test sorting of the columns of the data table by clicking the column. |
| Preconditions | The user must be on any page that displays a data table. |
| Steps | 1. Click on the any of the data columns. |
| Expected Result | The data should be sorted depending on which column the user clicks. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Description | Test filtering of data. |
| Preconditions | The user must be on any page that displays a data table. |
| Steps | 1. Hover over a column with the mouse. 2. Click an options button that appears after hovering over the column. 3. Enter a filter. |
| Expected Result | The data should be filtered depending on any input the user gives. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Description | Test chart data tooltip. |
| Preconditions | The user must be on any page that displays a chart. |
| Steps | 1. Hover a bar on the chart. |
| Expected Result | A tooltip should pop up stating the number of people in that age range. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Description | Test chart data set filtering. |
| Preconditions | The user must be on any page that displays a chart. |
| Steps | 1. Click on one of the data sets on the chart. |
| Expected Result | The data displayed by the chart should change based on the data set clicked. |
| Actual Results | + Result as expected. |

|  |  |
| --- | --- |
| Description | Test saving a chart as an image. |
| Preconditions | The user must be on any page that displays a chart. |
| Steps | 1. Right click a chart. 2. Click “Save image as…”. 3. Save the image. |
| Expected Result | The image should be saved to a folder that the user desires. |
| Actual Results | + Result as expected. |

Appendix C

## Installation Guide

The application is hosted on the following link: <https://devweb2018.cis.strath.ac.uk/~cxb15176/RPCDS/home.html>

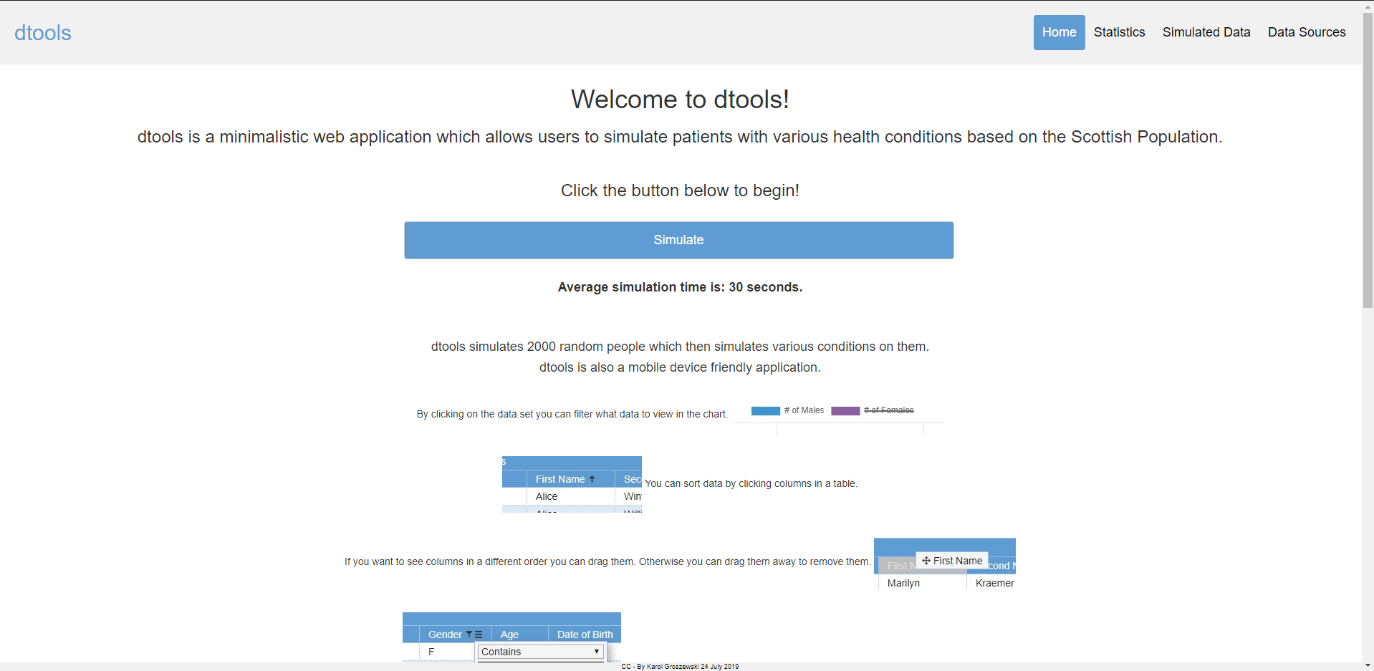
However, if you would like to host the website on your own web hosting server then:

1. Download RPCDS.zip file.
2. Extract the files.
3. Upload the extracted files to the web hosting server.
4. Done!

After those steps the application should be ready to use.

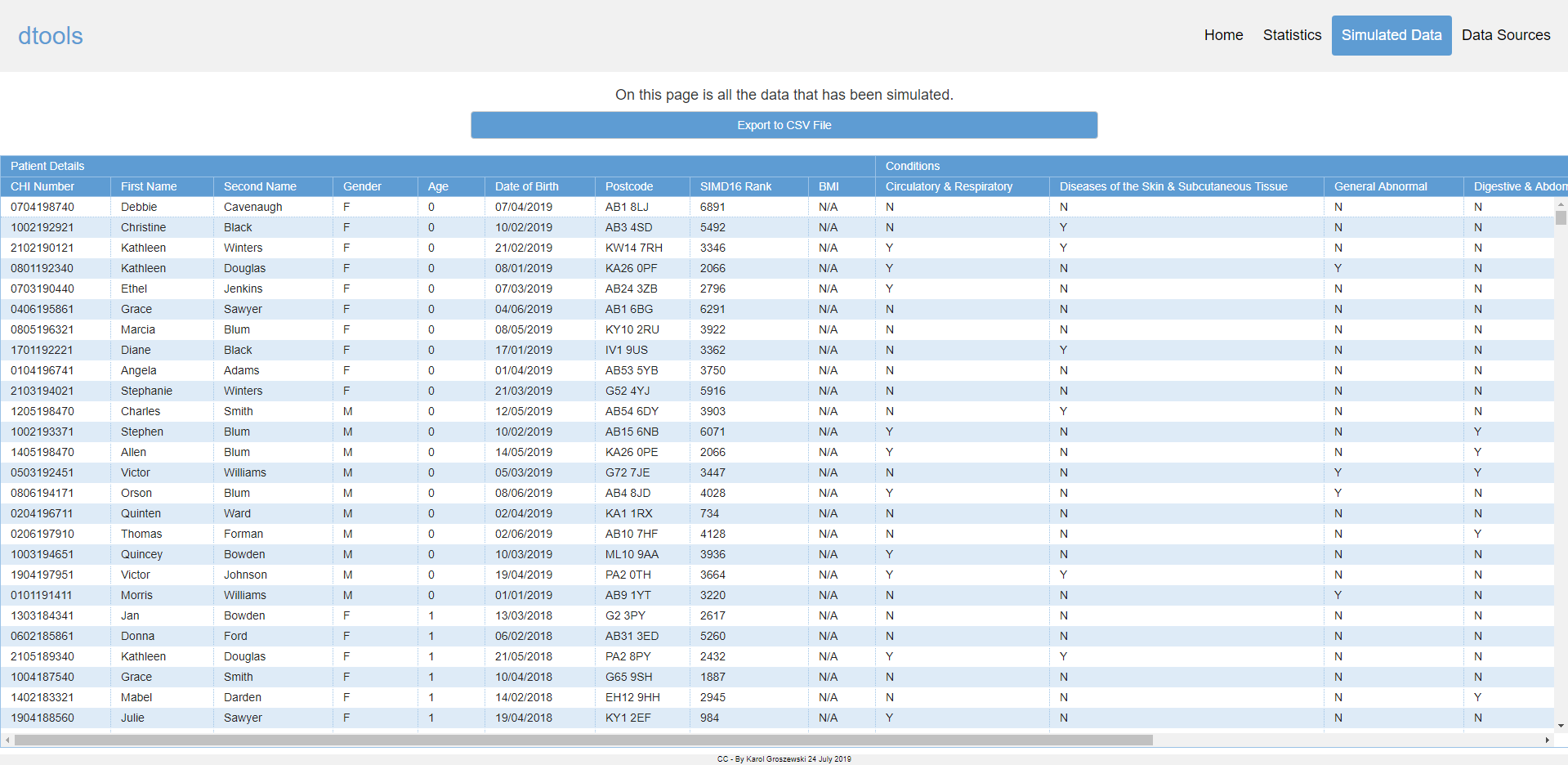
## User Guide

To start the application on the home page of it click on the “Simulate” button.



After clicking the button please wait to be re-directed to the Simulated Data page.

On the Simulated Data page shown below the user has various options.

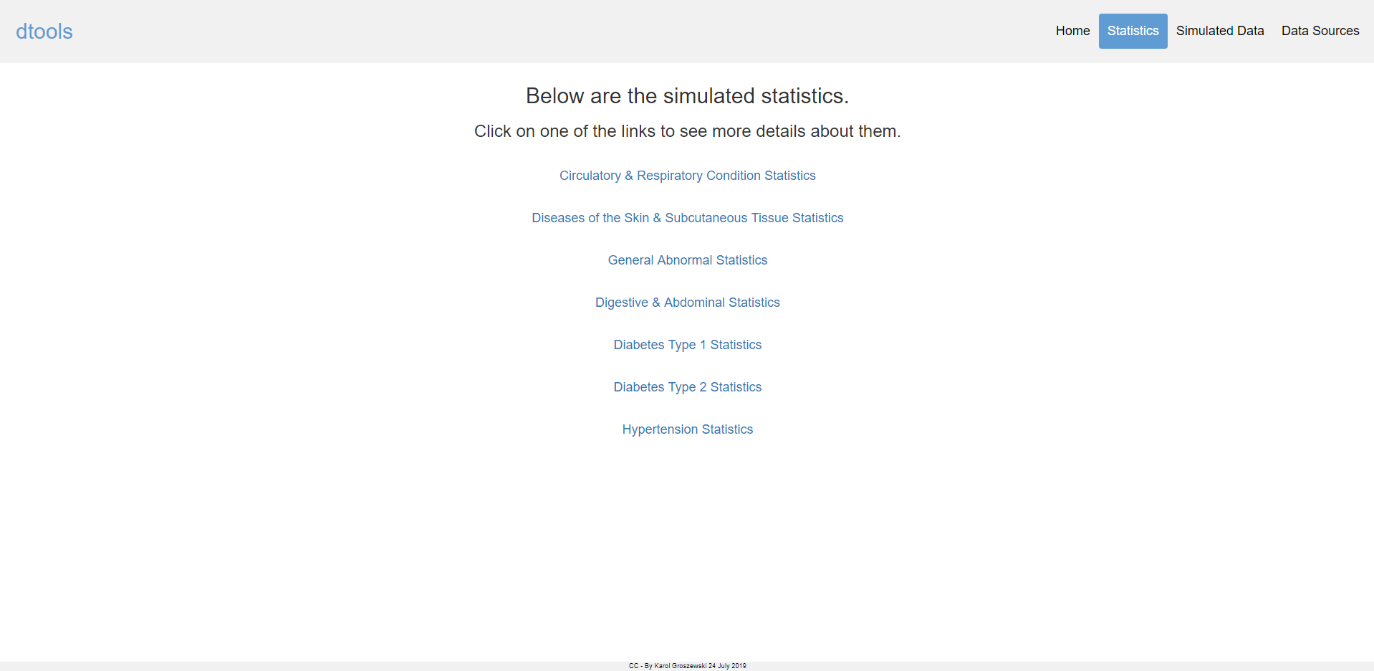


The user can download the full simulated data to their local machine by clicking on the “Export to CSV File” button.

The user can sort data by clicking on the columns. For example, clicking the “First Name” column.

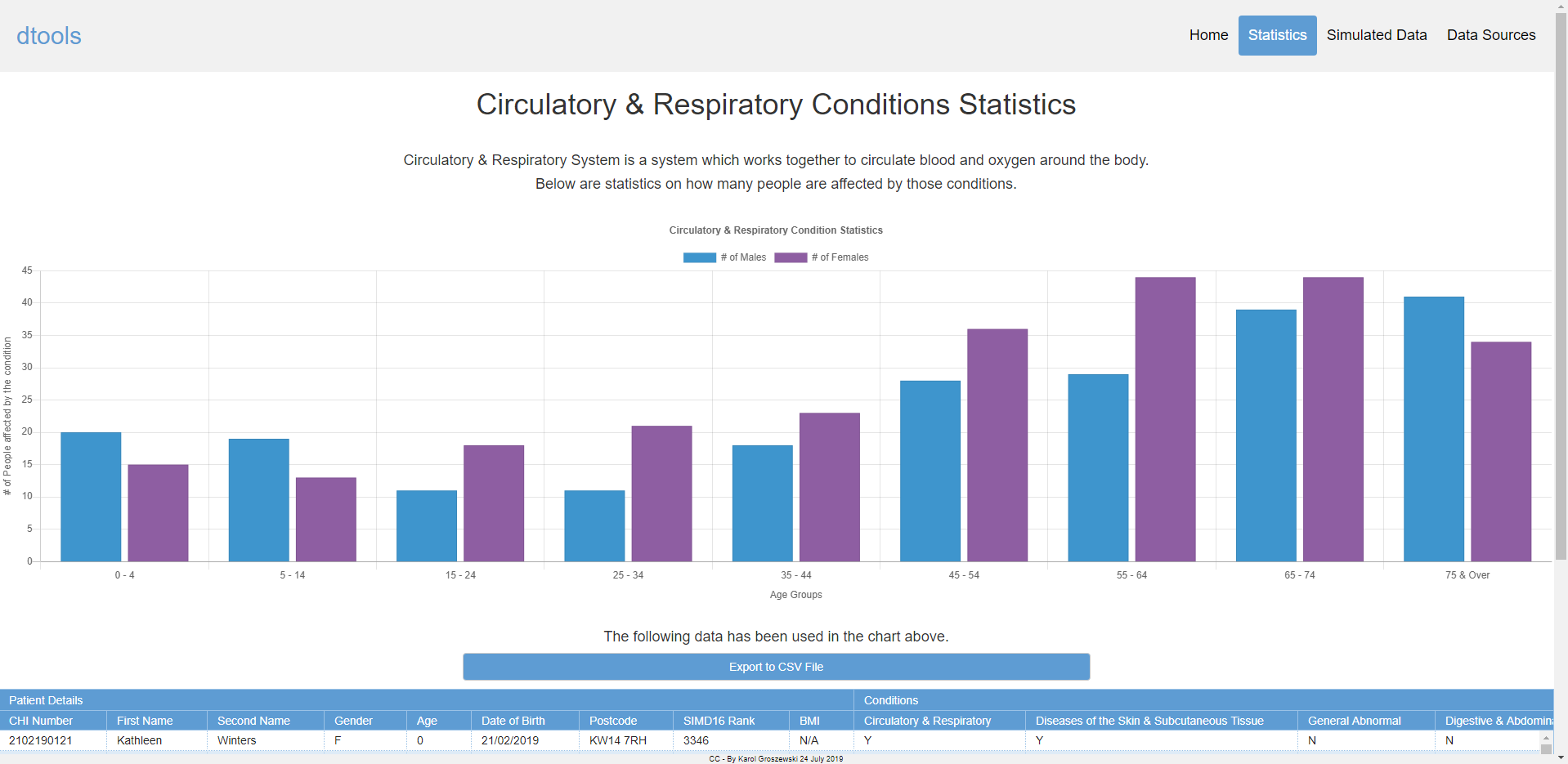
The user can filter the data in each of the columns by hovering over the column and additional button will appear and then clicking on it.

The user can see statistics on various health conditions by clicking on the “Statistics” button in the navigation bar.



On the “Statistics Page” the user is able to view various health conditions simulated.

By clicking on one of the hyperlinks, the user will be redirected to a page about statistics on that health condition.

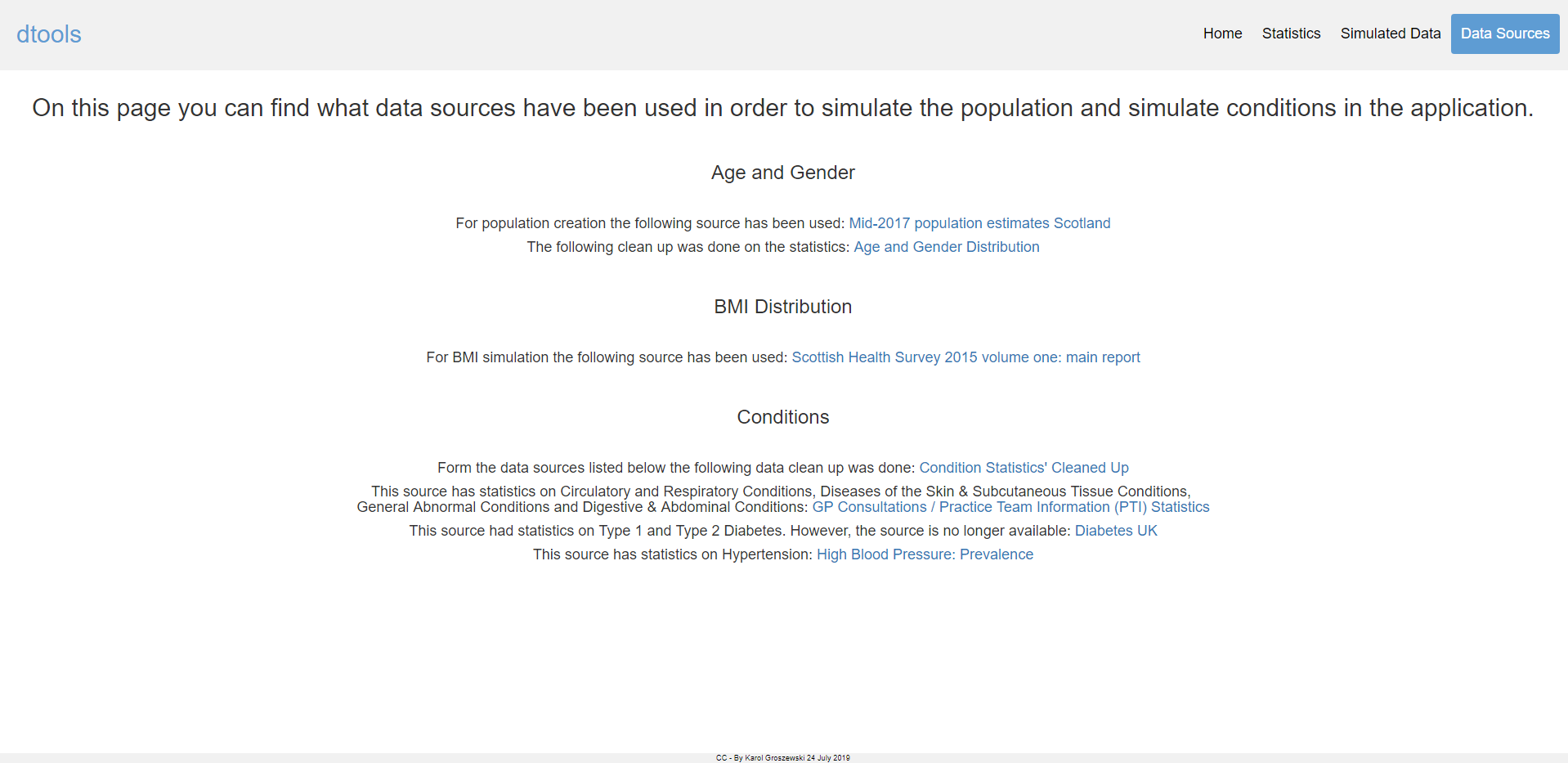


In the above image “Circulatory & Respiratory conditions Statistics” page is shown on that health condition.

The user can hover the mouse over the bars to see an accurate number of people in each category.

The user can click on a data set (e.g. # of Males) to remove a data set from being displayed in the chart.

The same rules apply for the “Export to CSV File” and the data table as in the “Simulated Data” page.



After clicking on the “Data Sources” page the user will be presented to the data sources that were used for simulation process in this application.

Appendix D

## Participant Consent Form

**Participant Information Sheet for Users**

**Name of department: Department of Computer and Information Services  
Title of the study: Random Patient Clinical Data Simulator**

**Introduction**

I am Karol Groszewski a 4th year student currently undertaking their dissertation project which focuses on developing a “random patient data” simulator which should be able to generate random but usable patient data.

What is a Random Patient Clinical Data Simulator?

Random Patient Clinical Data Simulator is a web application which allows users to simulate data. After data has been simulated the user will be able to retrieve the simulated data and analyse it.

The simulated data is going to be displayed using graphs for easier understanding.

**What is the purpose of this research?**The purpose of this research is to simulate data based on real world statistics for teaching and research purposes. Currently patient data is hard to access due to law protecting patient’s data privacy. However, statistics on patients across Scotland are available.

**Do you have to take part?**A participant’s decision to take part in the research is voluntary and refusing to participate or withdrawing participation will not affect any other aspects of the way a person is treated.

**What will you do in the project?**If you decide to participate in this research you will be provided with a document which has tasks for the user to complete on the web application. After completing these tasks, the user will be able to fill out a survey which ask questions on the application such as:

* How useful this application is to you?
* Is the application easy to navigate?
* Is the data readable?
* Are there any improvement suggestions?

**Why have you been invited to take part?**As a user of the application you have been invited to participate in this research in order to give feedback on the usability of the application. This feedback is important to help create and improve this application which can be used for teaching and research purposes.

**What are the potential risks to you in taking part?**There are no risks associated with this research, and you are not required to perform any preparatory activities. You are free to skip any questions you are not comfortable with and may stop at any time without having to give a reason.

**What information is being collected in the project?**

In the project there is no information collected on the participant. The survey will not ask participants for any private information such as participants name. The only piece of information collected from the participants are going to be their answers to the survey questions.

**Who will have access to the information?**

The information collected in this study will be anonymous and shared within a report and will be accessed by the lead researcher and supervisor.

**Where will the information be stored and how long will it be kept for?**

The information collected will be stored on the free surveys service. After the project has been graded the results from the survey are going to be deleted.

**What happens next?**If you participate in this study you will be asked to sign an online consent form, which confirms your rights discussed above and agreement to participate.

If you wish to withdraw data from the survey you may do so at any time without detriment. Contact the lead researcher.

If you do not wish to participate no further action is required, thank you for your attention and consideration.

**Lead researcher contact details:**

* Mr Karol Groszewski
* Student at the Computer Science department of the University of Strathclyde
* [karol.groszewski.2015@uni.strath.ac.uk](mailto:karol.groszewski.2015@uni.strath.ac.uk)

**Supervisor contact details:**

* Dr Marilyn Lennon
* Project Supervisor
* Computer and Information Sciences
* [marilyn.lennon@strath.ac.uk](mailto:marilyn.lennon@strath.ac.uk)

This research was granted ethical approval by the University of Strathclyde Ethics Committee. If you have any questions/concerns, during or after the research, or wish to contact an independent person to whom any questions may be directed or further information may be sought from, please contact:

Secretary to the Departmental Ethics Committee

Department of Computer and Information Sciences,

Livingstone Tower

Richmond Street

Glasgow

G1 1XH

email:ethics@cis.strath.ac.uk

**Consent Form for Users**

**Name of department: Computer Science  
Title of the study: Random Patient Clinical Data Simulator**

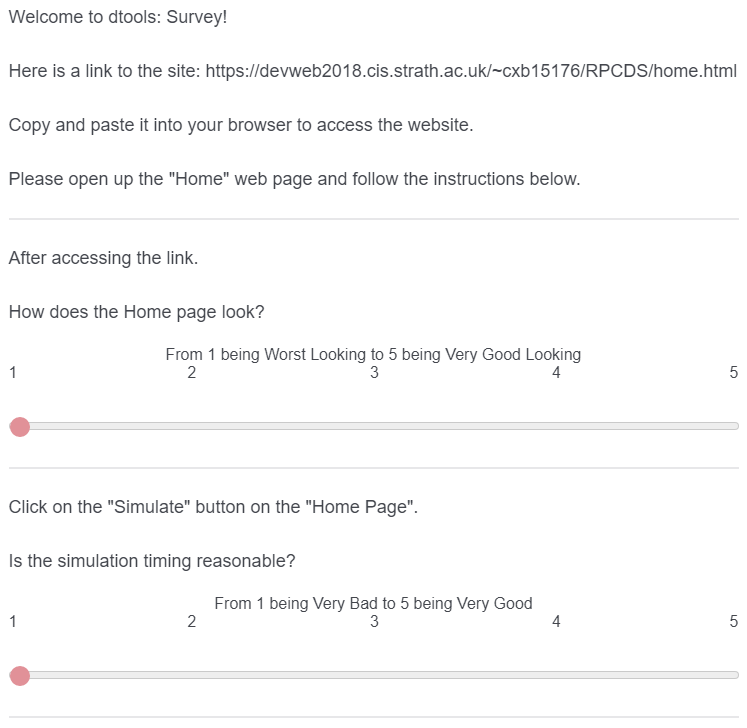
* I confirm that I have read and understood the Participant Information Sheet for the above project and the researcher has answered any queries to my satisfaction.
* I confirm that I have read and understood how my personal information will be used and what will happen to it (i.e. how it will be stored and for how long).
* I understand that my participation is voluntary and that I am free to withdraw from the project at any time, up to the point of completion, without having to give a reason and without any consequences.
* I understand that I can request the withdrawal from the study of some personal information and that whenever possible researchers will comply with my request.
* I understand that anonymised data (i.e. data that do not identify me personally) cannot be withdrawn once they have been included in the study.
* I understand that any information recorded in the research will remain confidential and no information that identifies me will be made publicly available.
* I consent to be a participant in the project.

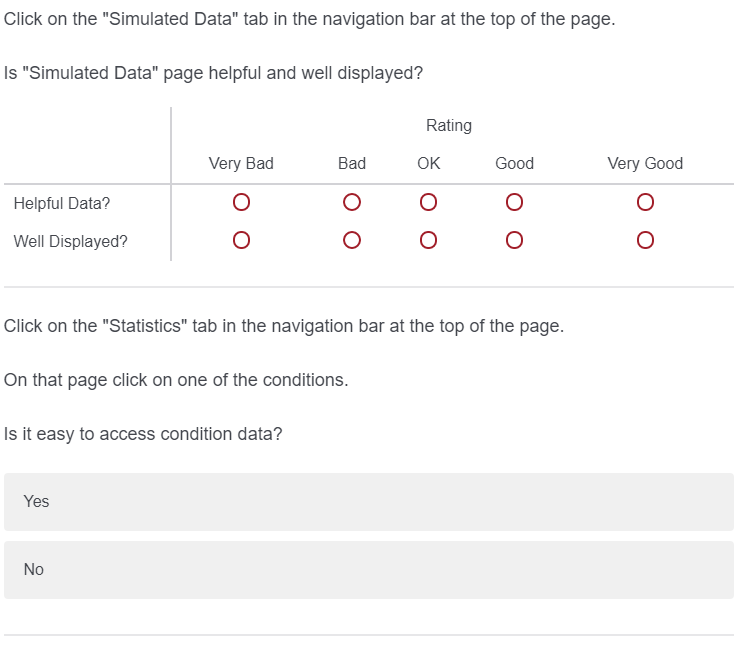
|  |  |
| --- | --- |
| Signature of Participant: …………………………. | Date: ……………………………………… |

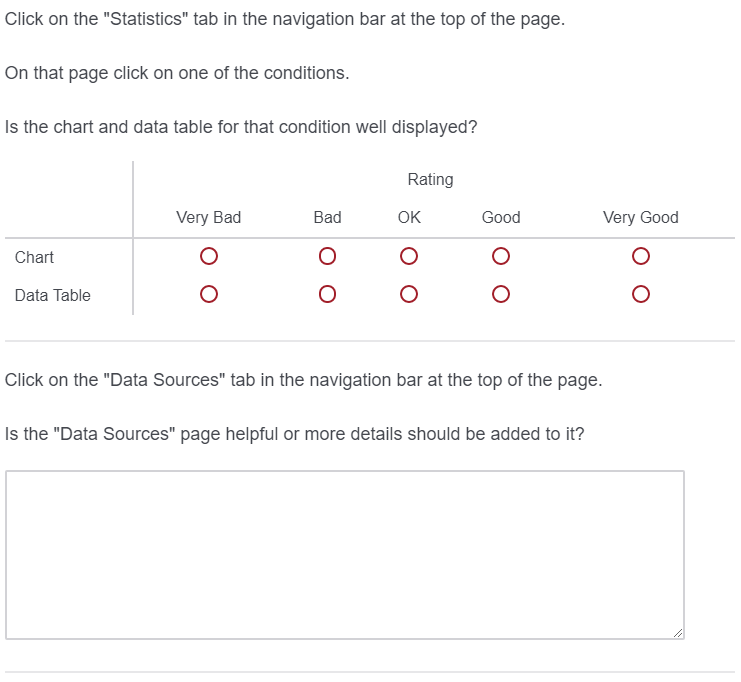
Appendix E

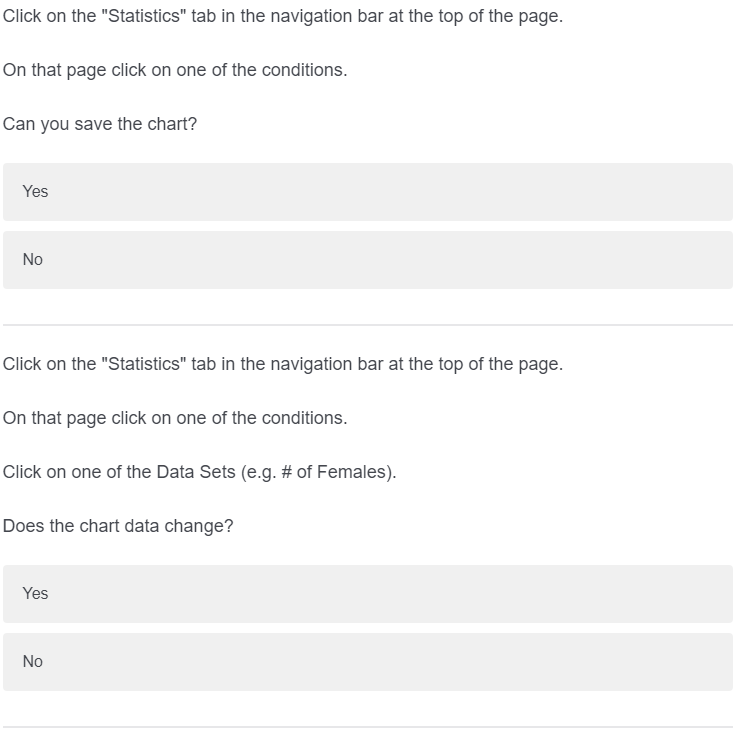
## User Survey

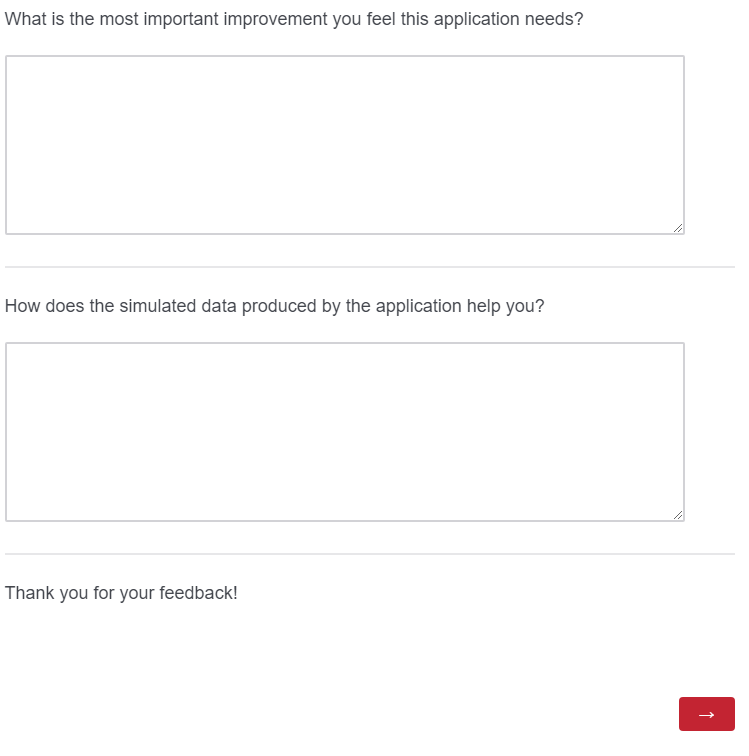
Link to the survey: <https://strathsci.qualtrics.com/jfe/form/SV_eVwIdavJBlhuk9n>











1. Body Mass Index [↑](#footnote-ref-1)
2. Community Health Index Number [↑](#footnote-ref-2)