



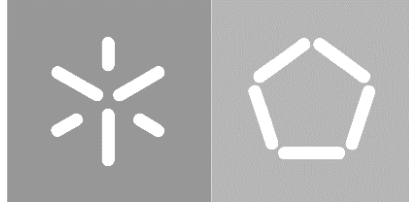
**Universidade do Minho**

Escola de Engenharia

Samuel Gustavo Correia Nogueira Alves

**Dynamic application to identify percentiles  
related to placenta parameters according to  
gestational age**

December 2021



**Universidade do Minho**

Escola de Engenharia

Samuel Gustavo Correia Nogueira Alves

**Dynamic application to identify percentiles  
related to placenta parameters according to  
gestational age**

Master dissertation in Bioinformatics

Dissertation supervised by

**Professor Ana Cristina da Silva Braga**

**Professor Rosete Maria Amorim Novais Nogueira Cardoso**

December 2021

## **DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS**

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

### **Licença concedida aos utilizadores deste trabalho**



**Atribuição-NãoComercial-SemDerivações**  
**CC BY-NC-ND**

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

## **ACKNOWLEDGEMENTS**

Firstly, I would like to thank my main tutor Prof. Dr. Ana Cristina Braga for the unconditional help and support during the thesis. I am grateful for all the dedication shown since the beginning, always present to help when I needed.

Also I would like to thank my co-tutor Prof. Dr. Rosete Nogueira for valuable help and inputs to help me.

To my parents and sister for their support during the different projects throughout my life.

And lastly but definitely not least my girlfriend which during all my ups and downs always had a word of love and support when I needed it the most giving me motivation to keep working to achieve my goals.

## **STATEMENT OF INTEGRITY**

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

## RESUMO

Durante anos, o crescimento do feto e o seu desenvolvimento eram as medidas que os médicos usavam para estimar o desenvolvimento e crescimento do mesmo. No entanto em estudos recentes tem-se vindo a provar que o desenvolvimento e crescimento da placenta é de grande importância para estudar a saúde e crescimento do feto.

Para a correta análise do desenvolvimento da placenta, curvas de crescimento precisam de ser criadas para comparação com diferentes percentis. Métodos de regressão linear foram usados para a criação de curvas de crescimento para a população Portuguesa (Nogueira et al., 2019).

Neste trabalho, usando os mesmos dados usados por Nogueira et al. foram criadas curvas de crescimento usando um método estatístico conhecido por regressão de quantis que permite criar curvas usando um método mais robusto que o anteriormente usado.

Foi também objetivo deste trabalho a criação de uma aplicação que permita ao utilizador colocar os valores de crescimento da placenta e poder comparar com as curvas de crescimento. Para isto, duas aplicações foram criadas. A primeira denominada APP1 tenta simular as ligações a bases de dados existentes em hospitais, tentando ao máximo recriar as condições presentes num hospital. A segunda aplicação, APP2, é mais simples e só precisa de documentos .csv com os valores da placenta, e permite a sua utilização sem necessidade de criação de algum tipo de ligação (<https://samuelalves.shinyapps.io/APP2/>).

Ambos os objetivos foram conseguidos com sucesso, permitindo um avanço na área que cada vez mais se mostra de grande utilidade.

## PALAVRAS-CHAVE

Placenta, Curvas de Crescimento, Desenvolvimento, Shiny, Regressão de Quantis

## **ABSTRACT**

For years, fetal growth and development were the measures that the doctors used to estimate fetal development and growth. However, recent studies have shown that the development and growth of the placenta is of great importance for studying the health and growth of the fetus.

For correct analysis of placental development, growth curves need to be created for comparison with different percentiles. Linear regression methods were used to create growth curves for the Portuguese population (Nogueira et al., 2019).

In this work, using the same data used by Nogueira et al., growth curves were created using a statistical method known as quantile regression, which allows the creation of curves using a more robust method than the one previously used (Nogueira et al., 2019).

It was also the objective of this work to create an application that allows the user to enter placental growth values and compare them with the growth curves. For this, two applications were created. The first one called APP1 tries to simulate the connections to existing databases in hospitals, trying as much as possible to recreate the conditions present in a hospital. The second application, APP2, is simpler and only needs .csv documents with the placenta values and allows its use without the need to create any kind of connection (<https://samuelalves.shinyapps.io/APP2/>).

Both goals were successfully achieved, allowing for a breakthrough in the area that is increasingly proving to be of great use.

## **KEYWORDS**

Placenta, Growth Curves, Development, Shiny, Quantile Regression

## **ABBREVIATIONS**

COVID-19	Coronavirus Disease 19
GA	Gestational Age
GPIO	General Purpose Input/Output
IoT	Internet of Things
IP address	Internet Protocol address
LED	Light-emitting diode
OWASP	Open Web Application Security Project
RAM	Random-access memory
RNA	Ribonucleic Acid
SQL	Structured Query Language
SSH	Secure Shell
UI	User Interface
USB	Universal Serial Bus



**CONTENTS**

Acknowledgements..... iv

Resumo..... vi

Abstract..... vii

Contents ..... ix

List of Figures..... xi

List of Tables..... xii

1. Introduction ..... 13

    1.1 Motivation and context..... 13

    1.2 Objectives ..... 14

    1.3 Organization..... 14

2. State of the art..... 16

    2.1 Models and statistical methods..... 16

        2.1.1 Previous placenta studies ..... 16

        2.1.2 Placental growth curves ..... 16

        2.1.3 Regression methods ..... 17

    2.2 Technological considerations ..... 18

        2.2.1 Raspberry Pi..... 18

        2.2.2 MariaDB..... 19

        2.2.3 SQL injection ..... 19

        2.2.4 Shiny applications..... 21

3. Regression methods ..... 21

    3.1 Linear regression ..... 22

        3.1.1 Least squares..... 23

        3.1.2 Quantile regression..... 25

4. Case of study..... 29

    4.1 Dataset used..... 29

    4.2 Growth curves ..... 34

5. Application development .....	37
5.1 System architecture .....	37
5.1.1 General overview .....	37
5.1.2 APP1.....	37
5.1.3 Packages used in the APP1 .....	37
5.1.4 APP1.....	38
5.1.5 Database.....	40
5.1.6 APP2 code .....	41
5.2 Manual to use APP1 and APP2.....	42
5.2.1 APP1 Instructions.....	42
5.2.2 APP2 Instructions.....	44
6. Conclusion and Future Work .....	48
References .....	49
Attachments .....	51
APP1 Code.....	51
APP2 Code.....	55

**LIST OF FIGURES**

Figure 1 - Schematics of APP1’s data interaction with the server. .... 19

Figure 2 – Representation of residuals for the OLS method. .... 25

Figure 3- Representation of examples of differences on distribution such as mean A) or quantile B). .. 27

Figure 4 – Distribution of women according Gestational Age. .... 31

Figure 5 – Distribution of women according to GA and fetal gender. .... 31

Figure 6 – Distribution of Maternal Age. .... 32

Figure 7 – Association between diameter 1 and GA. .... 33

Figure 8 – Association between diameter 2 and GA. .... 33

Figure 9 - Growth curve for diameter1 vs gestational age. .... 34

Figure 10 - Growth curve for diameter2 vs gestational age. .... 35

Figure 11 - Growth curve for thickness vs gestational age. .... 35

Figure 12 - Growth curve for placental weight vs gestational age. .... 36

Figure 13 – Representation of the application layout. .... 38

Figure 14 – APP1’s workflow’s fluxogram..... 39

Figure 15 – Example of database created. .... 40

Figure 16 – APP1’s main window with client information. .... 42

Figure 17- APP1’s main window with client information editing. .... 43

Figure 18 – Changes to APP1’s data..... 44

Figure 19 – APP2’s input file. .... 44

Figure 20 – APP2’s main window. .... 45

Figure 21 – APP2’s main window with client information. .... 45

Figure 22 - APP1’s main window with client information editing. .... 46

Figure 23 – APP2’s client information before editing. .... 47

Figure 24 - APP2’s client information after editing. .... 47

**LIST OF TABLES**

Table 1 – Description of variables available on dataset. .... 30

Table 2 – Descriptive statistics for the quantitative measures. .... 32

## **1. INTRODUCTION**

In this chapter the dissertations' motivation and context will be presented mainly the importance of the placenta as an organ capable of regulating the release of different hormones and protect from xenobiotics, and as an indicator of different possible diseases throughout the newborn's life. It will also be presented different challenges of conducting this type of study.

It will also state the objectives and the organization of this thesis.

### **1.1 Motivation and context**

Placenta is the biggest and one of the first organs to develop during a pregnancy and considered by some experts one of the most important organs due to its importance to a good development of the baby during the pregnancy which in mammals what determines the growth of the fetus the most is the supply of nutrients by the placenta to the fetus.

Its growth and development are highly regulated and coordinated such as the exchange of excretion products and nutrients between the mother and the baby has a high efficiency.

As such, the placenta can react to changes in the circulatory and nutritional environment adjusting its nutrients' transfer capabilities by changing the relative proportions of different kinds of nutrients supplied to the fetus.

The placenta has also a protector effect against certain xenobiotics and diseases and infects that could come from the mother. It is also the source of many hormones released to both the mother's and fetus' blood flow to allow a correct pregnancy.

Studies have shown the relationship between the several measures of the placenta with pregnancy problem, newborn mortality and the appearance of diseases that appear later in life (Asgharnia et al., 2008; Barker et al., 1990; Hindmarsh et al., 2001). As such, changes in the development of the placenta have negative effects in the fetus and its ability to develop in the intrauterine environment.

There has been an increase in the interest by researchers in better understand this organ but despite that, there is still a gap in the knowledge about it and its biological significance due to the previous lack of interest to study it compared to the fetus.

Moreover, there is an added difficulty to get the measurements of the placenta due to the inexistence of a standard on the methods to get the measurements like having or not the membranes and the umbilical cord or even blood since even after removing it there could still be clots that can change its weight.

Besides that, there is still a huge lack of data on the measurements of the placenta (and its implications on the fetus health or even when as an adult) adding the fact that different races have different placenta measurements, and the available data is already old.

If in one hand the percentile curves of the weight of the baby are common, percentile curves of the different measurements of the placenta are rare and can be outdated adding the fact that there is not much data to allow the creation of these studies.

Despite that, a work previously done by Doctor Rosete Nogueira (Nogueira et al., 2019), allowed the creation of percentile curves of the placenta of the Portuguese population.

Thus, the following research question was considered: How can a dynamic application help professionals to determine the stage of development of the fetus based on the parameters of the placenta?

## **1.2 Objectives**

Continuing the work previously done, the main goal of this dissertation is the creation of a dynamic application that allow the user to determine using the gestational age and evaluated parameters the classification of the baby and to systematize the collection of the parameters and its analysis through the use of the created percentile curves.

To create and implement the application it is necessary to evaluate the percentiles curves based on the available data. The methodologies used to do this are based on the statistical methods like regression techniques.

It is also a goal to create on the app previously referred a way to allow the professionals to create an entry on the database for each patient and its placenta measure values. It must allow the user to create a new entry and to retrieve each value to then plot it and compare to the percentile curves.

## **1.3 Organization**

This document will be organized in 6 chapters. The first one starts by explaining the motivation behind this work and its objectives. The second chapter will talk about the state of the art on the subject of the thesis. The third chapter will be to explain the regression methods. On the fourth chapter the case of study and dataset used are presented. On this chapter the newly created growth curves are also presented. The fifth chapter will be about the applications development, both the APP1 and APP2, and

how they were developed. It also has instructions on how to use both applications. The sixth and final chapter has the conclusion and possible improvements for the future.

## **2. STATE OF THE ART**

In this chapter it will be discussed the state of the art about placenta growth curves and its importance, after that it will be discussed the regression methods and how they are used in the creation of growth charts.

This chapter explores, based on the literature, a set of definitions relevant to the study in question, namely the concept associated with placental growth curves (2.1) and regression models (2.2). The concepts associated with the development of dynamic applications in Shiny are also approached (2.3).

### **2.1 Models and statistical methods**

#### 2.1.1 Previous placenta studies

Despite the different problems associated with the study of the placenta as previously stated on chapter 1, to work alongside the increased interest in the study of this organ, there are some studies already being done. Some studies such as the work done by Nogueira et al., aim to provide new data and insights from it of the placentas' different measurements.

Other studies have been done to allow the researchers to use the available technologies to better study the placenta while still taking in considerations the main objective and interest behind studying this organ (Salafia et al., 2005).

To corroborate studies like of Salafia et al., other studies have been done to show how studies of the placenta can be done using ultrasound techniques to accompany the development of the placenta throughout the pregnancy (Isakov et al., 2018).

#### 2.1.2 Placental growth curves

According to Nogueira et al., 2019, placenta has increasingly gained interest among the researchers. These authors pointed out that one of the reasons to the increased interest is the fact that although more studies need to be done, the percentile curves are useful on the evaluation of fetal follow-up and diseases on children. Their study consisted in evaluating several placenta's metrics such as placenta weight, placental diameters, placental thickness and the placental weight ratio and birth/placental weight ratio, to create percentile curves of the Portuguese population (Nogueira et al., 2019).

There are studies that show that the placenta can be a predictor of perinatal mortality and morbidity (O'Brien et al., 2020) such as pre-eclampsia or fetal growth restriction (Turco & Moffett, 2019), and several diseases in adult life such as cardiovascular problems in patients with low birthweight and



relatively big placentas which is indicated with an increased placental ratio which is the ratio between placental weight to birthweight (Lao & Wong, 1999).

Growth charts are important by providing a comparison to a reference allowing the professionals to visually see the current values of the fetus/baby and its growth trajectory (Fenton, 2003).

The updating and creation of the percentile curves of the placenta and comparing them between regions or countries (a study in Ireland showed that the tenth centile of a previously created percentile curves created with data from North Americans for the placental weight at 39 weeks was higher than the growth chart of the Irish population's placenta (O'Brien et al., 2020)) are important to evaluate pregnancy problems and increase the mother education and healthcare (Nogueira et al., 2019).

Given the scarcity of studies to assess the parameters associated with the placenta and its development with gestational age, the bibliographic review will be addressed in terms of the evolution charts of development parameters of the fetus and/or newborn.

### 2.1.3 Regression methods

Most of the studies associated with growth charts use regression models.

Regression is a statistical method used to investigate the relationships between variables. Quantile regression is a method used to estimate functional relations between variables for all portions of a probability distribution (Cade & Noon, 2003).

Through the years there have been innumerable ways to create growth charts. One of the first methods fit smoothing curves on sample quantiles of the segmented age groups. The problem with these methods arise with the fact that they are not robust (robustness is considered the resilience of statistical procedures to deviations from the assumptions of hypothetical models (Koenker & Bassett, 1978)) to outliers and a large sample is needed to estimate the percentiles with good precision (Chen, 2005).

In the 19th century Quetelet created anthropometric methods to construct reference growth charts. According to the models created, growth curves usually are constructed based on the assumptions that the different measurements are normally distributed. If in one hand, adult heights from relatively homogeneous populations are close to normal, on the other hand, children's heights are usually non-normally distributed. Other measurements such as weight are potentially more problematic. As such several methods have been proposed to solved this problem and normalize the measurements (Wei et al., 2006).

Quantile regression is a nonparametric method which is becoming more frequently used thanks to the increase in computing power to solve the previously stated problem. Its main advantage is that it is

independent of any distribution or transformation to normality due to the fact that it estimates the distributions directly determining the quantiles and as such it is a more direct method of representation. It is also more robust against outliers. Due to the fact that standard deviation and mean are used in normal distributions, both of these won't be suited to use with quantile regression and as such, quantile estimates instead of Zscore the user obtains a quantile (Kiserud et al., 2018).

## **2.2 Technological considerations**

### 2.2.1 Raspberry Pi

The computer used to create and host the database has been the Raspberry Pi Model B.

Raspberry pi consists of computers with a single board (single-board computer) that has several versions, each with different specifications produced by the Raspberry Pi Foundation <https://www.raspberrypi.org/about/>. To operate the Raspberry Pi, it is used the same components as a normal computer such as a keyboard and mouse for the input and a monitor to be displayed the output. It also contains the usual pieces found on normal computers to operate such as graphics card, processor, program memory (RAM). The hard drive for the Raspberry Pi is an SD card and the whole computer is powered by USB (Universal Serial Bus) (Raspberry Pi as Internet of Things hardware: Performances and Constraints). Due to several specifications and features such as GPIO (General Purpose Input/Output) pins (that allows the user to control different components such as LEDs (Light-emitting diode) and relays), its own operating system (raspbian) that is based off Linux (although there are other operating systems available that are not Linux based) and offers 'out of the box' several open source resources which allow the user to create several projects such as IoT (Internet of Things) devices like a computer which automates a water treatment plant (Raspberry Pi as IoT hardware: Performances and Constraints; Raspberry Pi for Automation of Water Treatment Plant).

Compared with other products similar to a Raspberry Pi, other products have a better performance, although they are more expensive, and as such Raspberry Pi is good choice due to the low price and the ability to connect to other computers in a given network just like an usual computer would.

From the variety of options on the operating system to use, Raspbian was the chosen one due to its optimization for the Raspberry Pi's hardware.

To interact with the Raspberry Pi without needing to setup the input (keyboard and mouse) and output (screen) peripherals, on a computer running the operating system windows 10 it was used the application

PuTTY which enables the user to connect and control the Raspberry Pi with a SSH (secure shell) connection.

### 2.2.2 MariaDB

After MySQL was bought by Oracle, one of MySQL's developers created a new database management system which he called MariaDB.

Several companies already changed their database management system to MariaDB such as Wikipedia (<https://diff.wikimedia.org/2013/04/22/wikipedia-adopts-mariadb/>), WordPress.com or Google (<https://mariadb.org/about/>). Originally created as an enhanced drop-in replacement for MySQL, it now has new features not available on MySQL (<https://mariadb.com/kb/en/mariadb-vs-mysql-compatibility/>). This database management system is available in the Raspbian repository and as such it was the database management system chosen to be used on this work. The schematics of the interaction of the APP1. The schematics are shown on Figure 1.

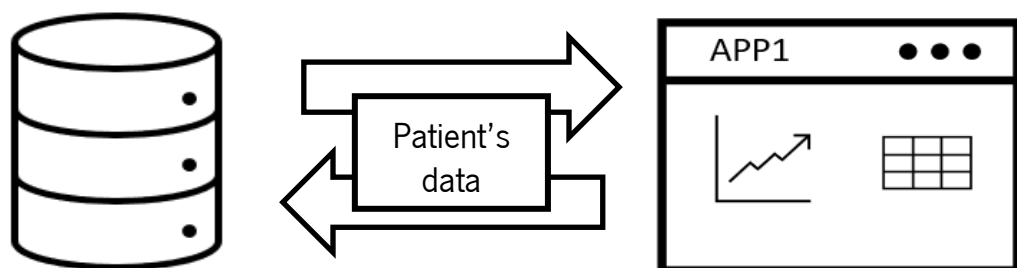


Figure 1 - Schematics of APP1's data interaction with the server.

### 2.2.3 SQL injection

As said by Balasundaram & Ramaraj (Balasundaram & Ramaraj, 2012), no database is safe from attackers.

Open Web Application Security Project (OWASP), one of the main sources for guidelines on web application security for many companies related to web security, as of 2021 claims that the 3<sup>rd</sup> highest incident rate security flaws involve code injection (<https://owasp.org/Top10/>).

Tiwari (Tiwari & Tiwari, 2015), state that, SQL, Structured Query Language is the most used language to interact with relational databases both to define and structure the database and its manipulation. SQL injection consists on using SQL statement as an input to the application with malicious intentions.

These statements are created based on the pre-defined logical expressions from a pre-defined query and the injection alters it to a true or false statement. The error messages returned by a database can also

assist the attacker and that can be useful when the attacker does not know how the database is structured and its queries (Anley, 2002; Boyd & Keromytis, 2004). Defines the query as a collection of statements that usually return a single “result set”.

The following explanation of an example of a SQL injection is from (Anley, 2002).

It starts by showing a normal SQL statement:

```
select id, forename, surname from authors
```

With this statement it will be retrieved the *id*, *forename* and *surname* columns from the *authors* table. This statement could be restricted to a given *author*:

```
select id, forename, surname, from authors where forename = 'john' and  
surname = 'smith'
```

It is important to note that the string literals *john* and *smith* are delimited with single quotes. Assuming that both the *forename* and *surname* fields are being gathered from an user input, it can be injected an SQL query by changing the input values such as:

Forename: jo'hn

Surname: smith

Then the query is created as:

```
Select id, forename, surname from authors where forename = 'jo'hn' and  
surname = 'smith'
```

When running this query an error might be returned stating that it has an incorrect syntax near *hn*.

This happens because with the query submitted, the *single quote* character *breaks out* of the single-quoted data. The database then tried to execute *hn* and failed.

This would be an example where nothing wrong happens if the attacker wanted to delete the *author* table, it could change the query to:

Forename: jo'; drop table authors–

Surname:

Related to this work, the database produced with the APP1 could be a target for SQL injections, To try to mitigate some of the attacks, the input is coerced to an integer and the first element is the chosen to complete the query from the user's input.

#### 2.2.4 Shiny applications

Shiny is a package to use in R that allows the user to create web applications relatively easily with R code. Authors have already stated its usefulness and that it may revolutionize the way models simulations are shared for example in the pharmacometrics' field making it accessible to bigger group of researchers or clinicians, since the ability of sharing and communicate the information and results is quite challenging, and although creating these web applications requires from the researchers' high competency, these skills are becoming more common in these fields. It is also shown by the authors different ways that these applications can be shared such as ShinyApps.io, Shiny Server and Shiny Server Pro which allows the creators to share the web application created to a wider audience since these solutions are not local but instead are hosted by a web server that allows user from around the world to access it (Wojciechowski et al., 2015).

In the bioinformatics there are a lot of shiny web applications already created. For example, the ShinyCircos that generates a circos plot which is considered one of the best ways to present genomic data, Cipr that is a web application to annotate cell clusters from single cell rna sequences (Ekiz et al., 2020; Y. Yu et al., 2018). BEST is another example that is a web application used to retrieve cross-related enzyme functional parameters and information from BRENDA (Hidalgo et al., 2020). Another one is the robvis which is a web application used to visualize risk-of-bias assessments (McGuinness & Higgins, 2021).

Due to the pandemic of COVID-19 (Coronavirus Disease 19) several web applications were created that are used to model and show in different ways the data gathered from COVID-19 (Mercatelli et al., 2020; Salehi et al., 2021).

### **3. REGRESSION METHODS**

As most of the statistical methodology associated with growth curves involves regression methods, this chapter will be dedicated to the main concepts associated with this methodology.

### 3.1 Linear regression

Regression is a statistical method used to investigate the relationships between variables and can give answers related to a response variable. The classical regression (commonly known as the regression function) is focused on the expectation of variable  $Y$  conditional on the values of a set of variables  $X$ ,  $E(Y|X)$  (Weisberg, 2005). This function can have different levels of complexity although it restricts a specific location of the  $Y$  conditional distribution exclusively.

According with Su et al., linear regression delivers the simplest model form to model the regression function as a linear combination of predictors. This regression is used as the basis for many modern modelling tools. Linear regression is also used due to its modelling when the sample size is small or the signal is relatively weak, providing the user with positive results (Su et al., 2012).

Considering the data  $D = \{(y_i, x_i) : i = 1, \dots, n\}$ , where  $y_i$  is the  $i$ th response, measured on a continuous scale,  $x_i = (x_{i1}, \dots, x_{ip})^t \in \mathbb{R}^p$  is the associated predictor vector; and  $n(\gg p)$  is the sample size, the linear model is specified as

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \text{ with } \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \quad (1)$$

for  $i = 1, \dots, n$ . In matrix form,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ with } \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}), \quad (2)$$

Where  $\mathbf{y} = [y_i]_{n \times 1}$  is the  $n$ -dimensional response vector;  $\mathbf{X} = (x_{ij})_{n \times (p+1)}$  with  $x_{i0} = 1$  is often called the design matrix; and  $\boldsymbol{\varepsilon} = [\varepsilon_i]_{n \times 1}$ . There are four major statistical assumptions involved in the specification of model (1) or (2):

1. Linearity where  $\boldsymbol{\mu} \equiv [E(y_i|x_i)]_{n \times 1} = \mathbf{X}\boldsymbol{\beta}$ ;
2. Independence  $\varepsilon_i$ 's are independent of each other;
3. Homoscedasticity  $\varepsilon_i$ 's have equal variance  $\sigma^2$ ;
4. Normality  $\varepsilon_i$ 's are normally distributed.

To allow model interpretation, it can be introduced a generic notation  $\mu_x$  to denote the conditional mean response as

$$\mu_x = E(Y|X_1, \dots, X_p) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (3)$$

Where  $X_j$  is an  $n \times 1$  vector for  $j = 1, \dots, p$ . As  $\partial\mu_x/\partial X_j = \beta_j$  the regression parameters can be interpreted in terms of change rate:  $\beta_j$  corresponds to the amount of change in the conditional mean response  $\mu_x$  with one unit increase in  $X_j$ , since all other predictors are fixed (Su et al., 2012).

According to Weisberg (Weisberg, 2005), the variance equation,

$$\text{Var}(Y|X = x) = \sigma^2 \quad (4)$$

Is assumed to be constant and a positive unknown value. Since the variance  $\sigma^2 > 0$  the value observed on the  $i$ th response  $y_i$  will usually not equal its expected value  $E(Y|X = x_i)$ . Considering that difference, it was created a quantity called a *statistical error*, or  $e_i$ , for case  $i$  defined implicitly by  $y_i = E(Y|X = x_i) + e_i$  or explicitly by  $e_i = y_i - E(Y|X = x)$ , i.e, the difference between the observed and the expected value of the response variable.

This error depends on unknown parameters in the mean function, thus, are not observable quantities but random variables that correspond to the *vertical distance between the point  $y_i$  and the mean function  $E(Y|X = x)$* .

Weisberg makes two assumptions concerning the errors that he considers important:

- $E(e_i|x_i) = 0$ , thus if drawn a scatterplot of the  $e_i$  versus the  $x_i$  it would be a null scatterplot with no patterns;
- Errors are all independent, so, the value of the error in one case gives no information about the error's value of another case.

Another important topic to consider is the model estimation which has several methods of estimating both  $\beta$  and  $\sigma^2$  for linear models (examples include least squares, Bayesian approach, ridge regression, maximum likelihood, robust estimation, etc.). The one it will be characterized is the least squares.

### 3.1.1 Least squares

The least squares is considered one of the most popular methods for estimating  $\beta$ . This method minimizes the distance from the observed response to the predicted values

$$Q(\beta) = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \quad (5)$$

And when differentiated relatively for  $\beta$  :

$$\frac{\partial Q(\beta)}{\partial \beta} = -2(\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \beta) \quad (6)$$

And when setting this equation to zero, it is gotten the normal equation  $\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \mathbf{X} \beta$

Assuming that  $\mathbf{X}$  is a complete column of rank  $p$ , the Gram matrix  $\mathbf{X}^T \mathbf{X}$  is positive definite. Thus, the least squares estimator  $\hat{\beta}$  as a unique solution to the normal equation given by

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (7)$$

And the vector of fitted values,  $\hat{\mathbf{y}}$  is

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\beta} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{H} \mathbf{y} \quad (8)$$

Where  $\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  is called sometimes the hat matrix or the projection matrix.

The Figure 2 shows a representation of the residuals. The residuals are the basis for the criterion function to allow the obtaining of the estimators that from a geometrical point of view, are the vertical distances between the fitted line and the actual y-values. On the Figure 2, each data point is represented by a circle and the line represents the candidate ordinary least square line. The vertical lines that are drawn between the line and points are the residuals. The points that are below the line are negative results and the above ones are positive residuals.



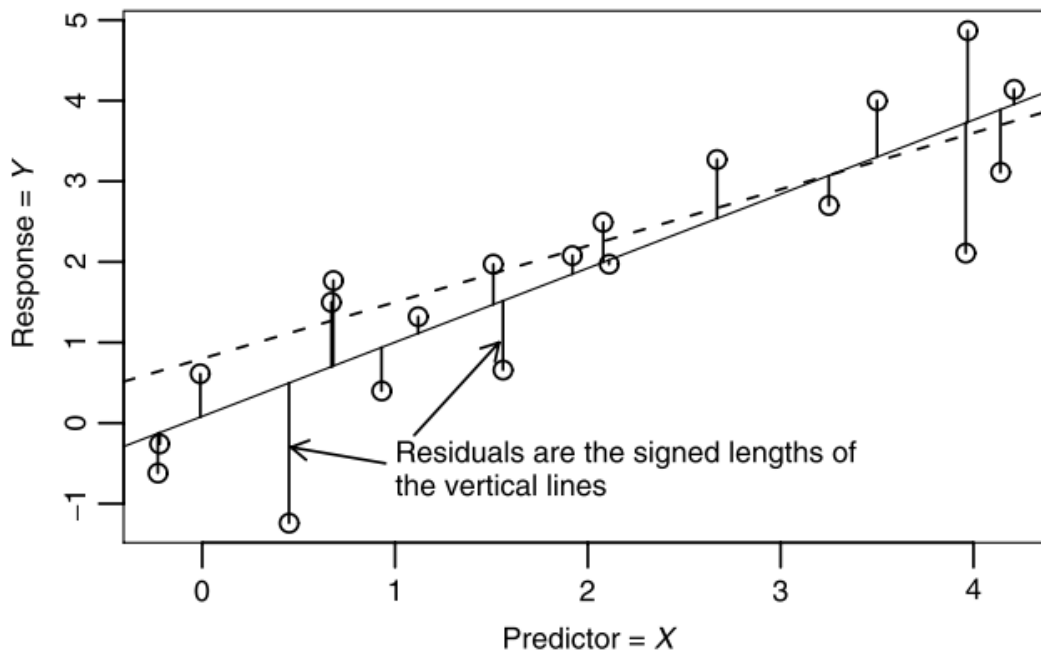


Figure 2 – Representation of residuals for the OLS method.

(Source: Weisberg, 2005)

### 3.1.2 Quantile regression

Quantile regression, introduced by Koenker and Bassat is a method used to estimate functional relations between variables for all portions of a probability distribution and allows the studying of the conditional distribution of  $Y$  on  $X$  at different location, offering a global view on the interrelations of  $X$  and  $Y$  (Cade & Noon, 2003).

As stated by Buhai, considering the ordinary quantile which has a real value random variable  $Y$  has the following distribution

$$F(y) = \Pr(Y \leq y) \quad (9)$$

Thus for any  $\tau \in (0,1)$ , the  $\tau$ -th quantile of  $Y$  is defined as

$$Q(\tau) = \inf\{y: F(y) \geq \tau\} \quad (10)$$

The median is  $Q(1/2)$ , the first quartile  $Q(1/4)$  and the first decile  $Q(1/10)$ . The quantile function allows the user to create a complete characterization of  $Y$ , like the distribution of function  $F$ . The quantiles can be written as solutions of this optimization:  $\tau \in (0,1)$

$$\rho_\tau(u) = u(\tau - I(i < 0)) \quad (11)$$

Where  $I(\cdot)$  is the usual indicator function. Thus, the solution to the minimization is

$$\hat{\alpha}(\tau) = \arg \min_{\xi \in \mathbb{R}} E[\rho_\tau(Y - \xi)] \quad (12)$$

The sample analogue of  $Q(\tau)$  is based on a random sample  $\{y_1, \dots, y_n\}$  of  $Y$ . the  $\tau$ -th quantile can be identified as any solution to:

$$\hat{\alpha}_\tau = \arg \min_{\xi \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(Y - \xi) \quad (13)$$

Let  $\mathbf{x}_i$ ,  $i = \overline{1 \dots n}$ , a  $K \times 1$  vector of regressors. It can be written the equivalent to the first expression as

$$F_{u_\tau}(\tau - \mathbf{x}_i^T \boldsymbol{\beta}_\tau | \mathbf{x}_i) = \Pr(y_i \leq \tau | \mathbf{x}_i) \quad (14)$$

Which is a different form of:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_\tau + u_{\tau_i} \quad (15)$$

considering the distribution of the error term,  $u_{\tau_i}$  is left unspecified, the only constraint being the quantile restriction  $Q_\tau(u_{\tau_i} | \mathbf{x}_i) = 0$ .

Using as analogy the estimation of conditional mean functions as in

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^K} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \quad (16)$$

The linear conditional quantile function

$$Q_Y(\tau|\mathbf{X} = \mathbf{x}) = \mathbf{x}_i^T \boldsymbol{\beta}_\tau \quad (17)$$

Can be estimated by solving the equivalent of expression 16.

$$\hat{\boldsymbol{\beta}}_\tau = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^K} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \quad (18)$$

According with Yu et al., to discuss about model estimation we need to take in mind the principles of the least squares estimation.

Considering the regression model that is the simplest, the  $\boldsymbol{\beta}$  is estimated through the quadratic loss function  $r(u) = u^2$  and given a data set with observations  $\{x_i, y_i\}_{i=1}^n$ , its estimation is done by minimizing

$$\sum_{i=1}^n r(y_i - x_i^T \boldsymbol{\beta}) = \sum_{i=1}^n r(y_i - x_i^T \boldsymbol{\beta})^2 \quad (19)$$

It is of great importance to take in consideration that this work is being done on a non-normal distributed sample as previously stated since it is also common on infants' studies to show a non-normal distribution contrary to what usually happens on adults. The distribution can also be different about the mean or a specific quantile as shown on Figure 3 (Beyerlein, 2014).

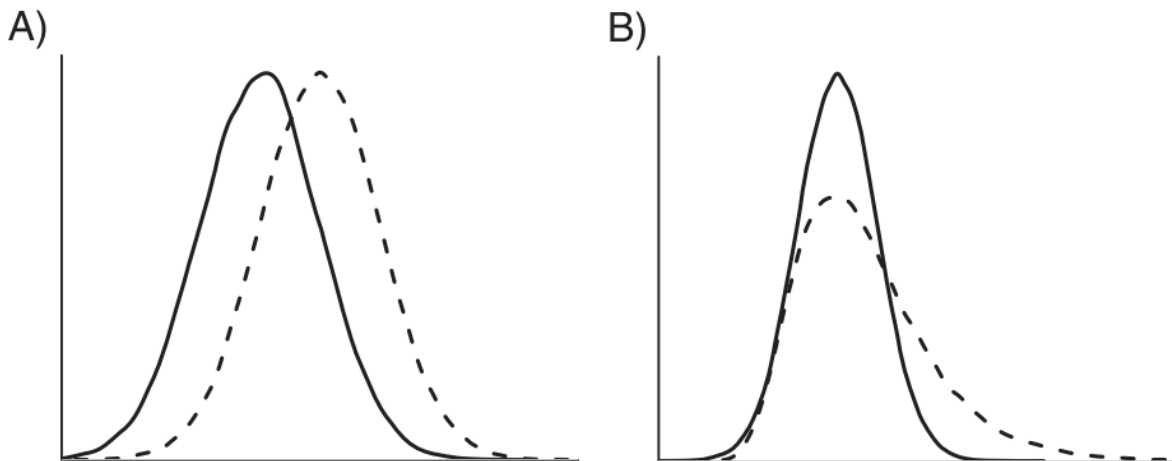


Figure 3- Representation of examples of differences on distribution such as mean A) or quantile B).

(Source: Beyerlein, 2014)

Taking this premise in account, it is important to correctly choose a regression method that can fill the gaps from other methods and bring a better solution to the researchers.

As such, modelling using only the mean that is what happens on linear regression can lead to missing important aspects of the association between the outcome and its predictors (Beyerlein, 2014).

An example of this would be as presented by Beyerlein, where a large meta-data study showed that breastfeeding was associated with a significant reduction of a child's risk of overweight later in life although there was no difference in the mean body mass index (BMI) values between breastfed and formula-fed children. Although this seems contradictory, the results fit well together analysing together with quantile regression since the data set showed that breastfeeding was associated with both a decrease in upper BMI percentiles and increase in the lower BMI percentiles at the age of 5-6 years as such there was no difference in mean BMI.

Thus, the quantile regression allows for examination of changes in distribution and provides more specific information than linear regression, allowing to better understand how pertinent variables affect the entire distribution of an outcome variable rather than just the mean of the distribution (Hajovsky et al., 2020).

## 4. CASE OF STUDY

In this chapter it will be presented the case of study. It uses a dataset created Doctor Rosete Nogueira by morphologically analyse placentas from Portuguese women collected at different pregnancy stages that were later studied at a specialized laboratory. The data collected contains information about the mother, the fetus and the placenta. This study was conducted with the goal to create the placenta's growth curves of the Portuguese population.

### 4.1 Dataset used

The dataset used was collected by Doctor Rosete Nogueira by analysing placentas of 2248 singleton pregnancies, collected from Portuguese hospitals and analysed at CGC Genetics, from 12 to 41 gestational ages (these gestational ages were chosen since before the 12 weeks the placenta's growth curves follow a different distribution and 41 weeks due to the end of the pregnancies), only once when morphologically analysed by Doctor Rosete Nogueira, of which only 1951 had the complete information needed for the study. The name and information of the patient was changed to a process number, not allowing its direct identification. It has the maternal age being the lowest 15.1 years and the highest 48.2 years with an average of 35.6 years. It has several information about the fetus such as gender, weight, eventual pathologies. It has also information about the umbilical cord such as diameter, number of vessels, length, and insertion type. Finally, the dataset has information about the placenta such as diameter (both the smallest and biggest), thickness, shape, volume, and weight.

The principal variables as well its scale of measurement are listed on Table 1.

The Figure 4 shows that there is a reduced amount of data on the interval of 12 to 15 and 25 to 32 gestational ages compared to other intervals. It is also observed that the frequency not counting with these intervals is relatively similar and with a frequency with a value usually higher than 65.

The Figure 5 shows that until the gestational age of 24 weeks there is still occurrences of ambiguous genders. On the interval of 16 to 24 weeks of gestational age there is approximately the same frequency between male and female genders.

Table 1 – Description of variables available on dataset.

<b>Variable name</b>	<b>Designation</b>	<b>Scale of measurement</b>	<b>Values</b>
<b>MaternalAge</b>	Maternal Age	Proportional	years
<b>GA</b>	Gestational age	Proportional	From 12 to 41
<b>Fetalgender</b>	Fetus gender	Nominal	Levels: M – Male, F – Female, A – Ambiguous; 9 – Missing information
<b>Fetalweightg</b>	Birthweight	Proportional	Grams, g
<b>Placentalshape</b>	Placental Shape	Nominal	Levels: Normal, Bilobate, Circunvalate, Membranous
<b>Placentalweight</b>	Placental Weight	Proportional	Centimetre, cm
<b>Diameter1</b>	Diameter 1	Proportional	Centimetre, cm
<b>Diameter2</b>	Diameter 2	Proportional	Centimetre, cm
<b>Placentalthickness</b>	Placental Thickness	Proportional	Centimetre, cm
<b>Placental_vol</b>	Placental Volume	Proportional	Cubic Centimetre, cm <sup>3</sup>

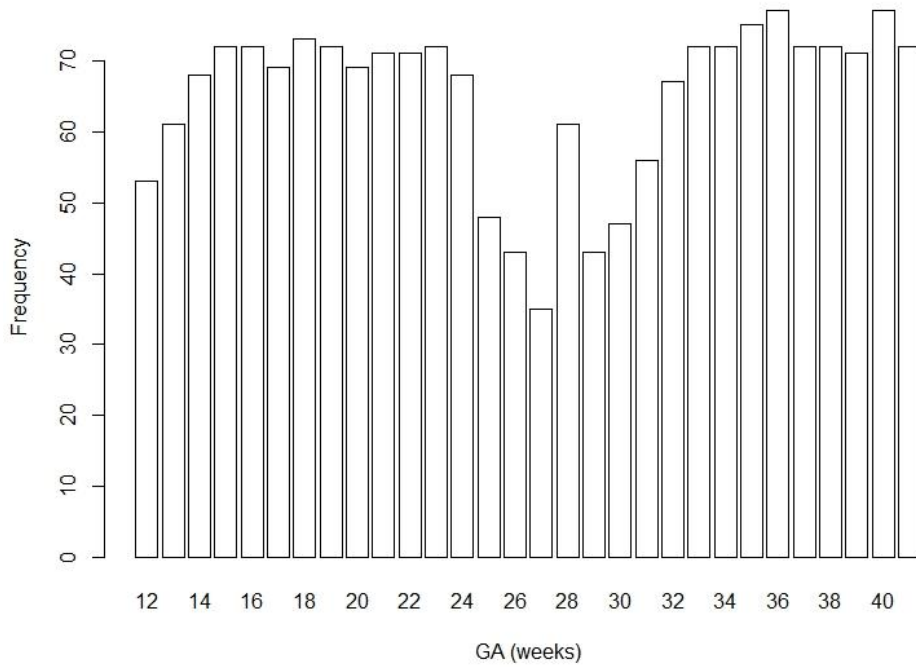


Figure 4 – Distribution of women according Gestational Age.

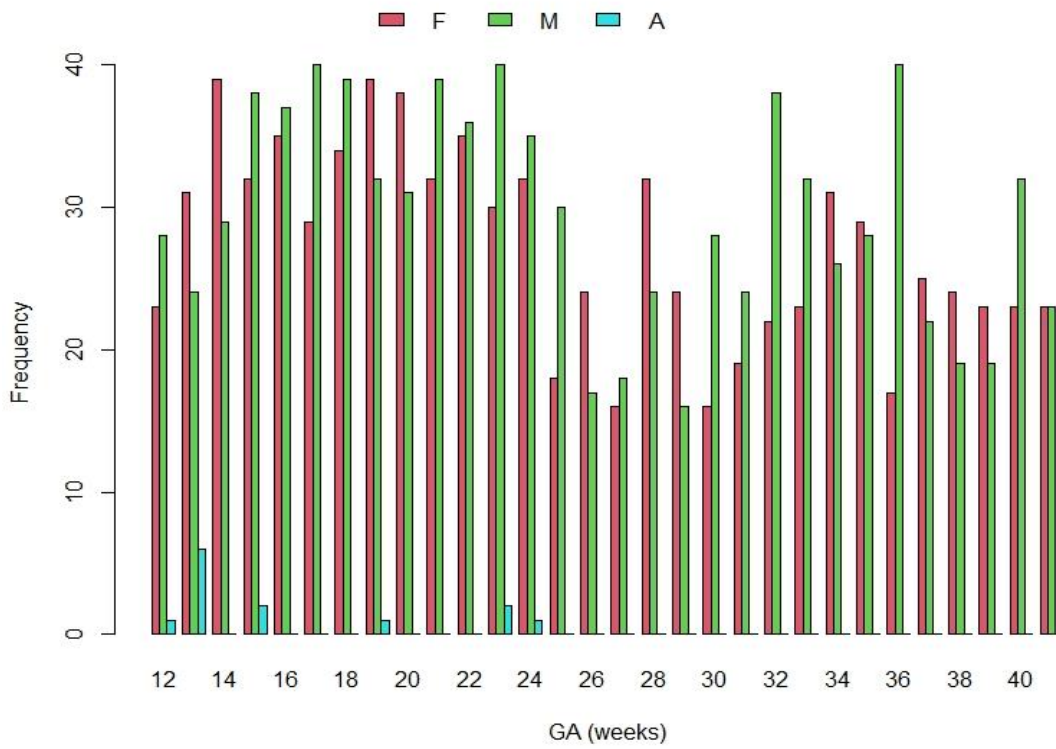


Figure 5 – Distribution of women according to GA and fetal gender.

According with the plots from Figure 4 and Figure 5, it is noted that there is enough information on each gestational age value for the construction of the percentile curves for the babies of both masculine and feminine gender.

On the Table 2 the descriptive statistics for the quantitative metrics, gestational age (GA), Fetalweight, Placentalweight and both Diameters 1 and 2 are shown, in specifics the minimum and maximum value, median and mean and the 1<sup>st</sup> and 3<sup>rd</sup> Quantile.

Table 2 – Descriptive statistics for the quantitative measures.

	<b>GA</b>	<b>Fetalweight</b>	<b>Placentalweight</b>	<b>Diameter1</b>	<b>Diameter2</b>
<b>Min.</b>	12.00	5.4	6.0	1.70	1.5
<b>1st Q</b>	19.00	235.5	101.0	10.00	8.0
<b>Median</b>	26.00	766.0	195.0	13.00	11.0
<b>Mean</b>	26.71	1248.7	233.2	14.48	12.8
<b>3rd Q</b>	35.00	2220.0	345.5	17.00	14.6
<b>Max.</b>	41.00	4880.0	995.0	32.00	30.00

Figure 6 illustrates the age distribution of women in histogram form as well as the density function. Through this image it can be seen that this is practically symmetric.

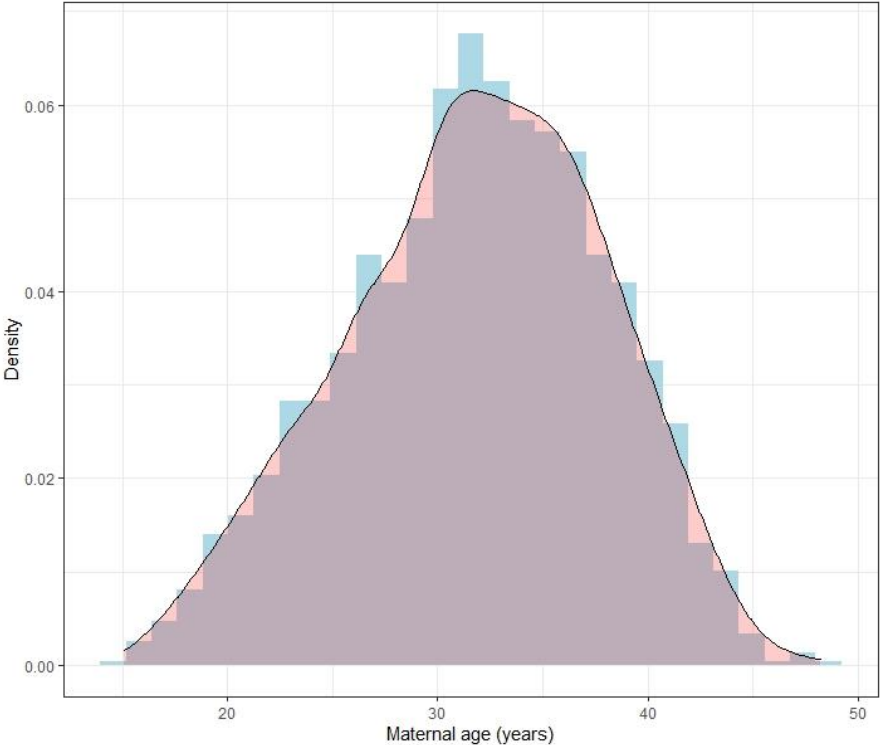


Figure 6 – Distribution of Maternal Age.



The graphs in the Figure 7 and Figure 8 illustrates the associations of diameters 1 and 2 with gestational age according fetus gender. As can be seen from the graphs, both the diameter 1 and 2 increase its size as the gestational age progresses for all the sexes (Female, Male and Ambiguous).

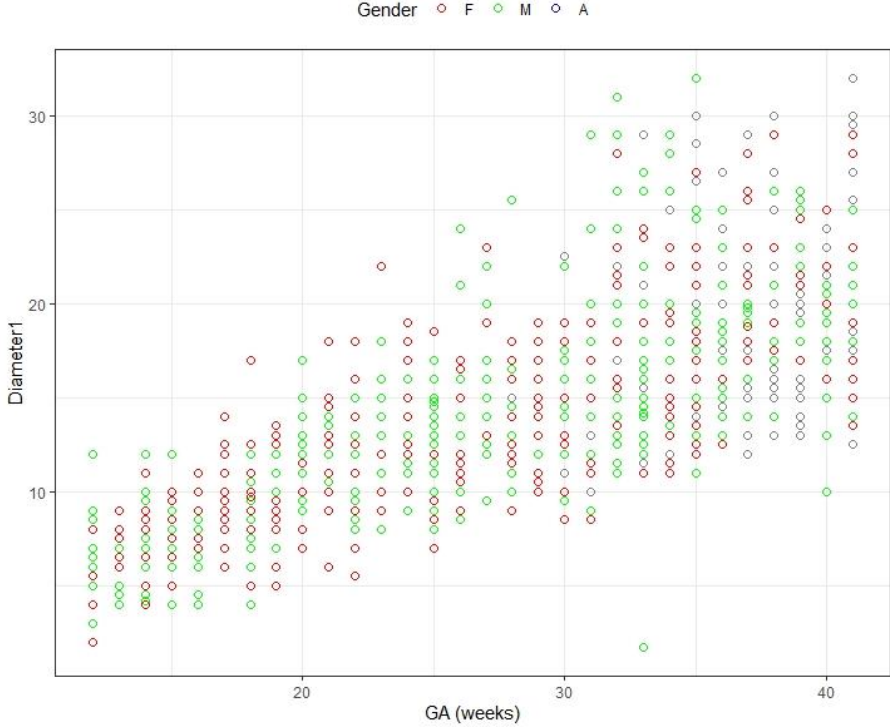


Figure 7 – Association between diameter 1 and GA.

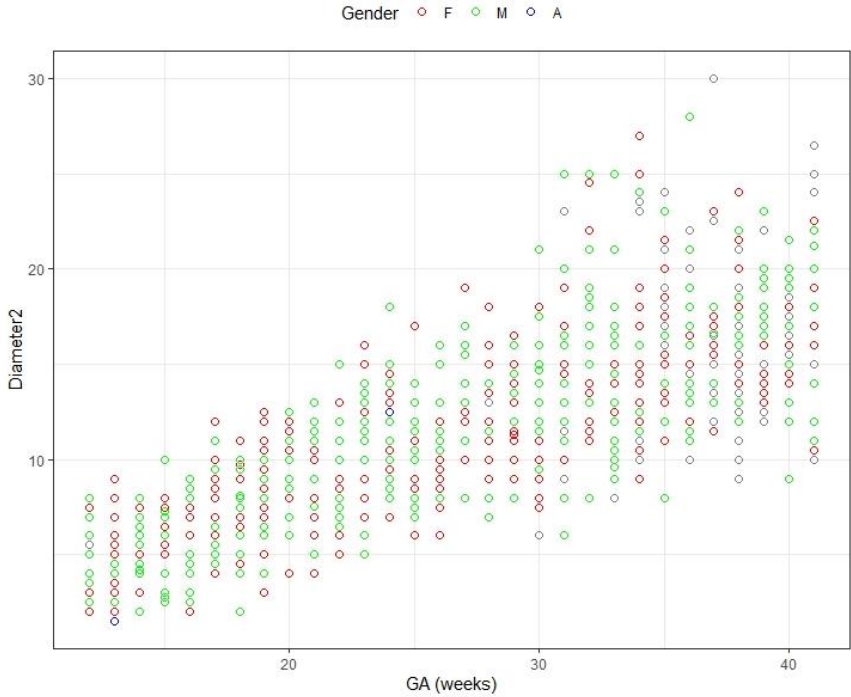


Figure 8 – Association between diameter 2 and GA.

## 4.2 Growth curves

Using the previously described dataset, and using the packages `quantregGrowth` (which uses the package `quantreg`), and `ggplot2` the growth chart was made.

To do that, the dataset was loaded into R using the `read.csv()` function, then using `quantregGrowth`'s package `gcrq` function it was created an object that contains the coefficients and was later plotted using the `plot.gcrq` function. This was made for every chart made, only changing the `gcrq` function to the corresponding column.

The plots created are in the following images (Figure 9 to Figure 12). These plots do not differentiate between the genders since that would split the data and thus creating a less robust plot, and due to the fact that there were no observable differences between the genders that would justify the separation.

The 97<sup>th</sup> percentile on both the Figure 9 and Figure 10 show a decline on the interval of 35 to 40 weeks of gestational age. Although still not sure about the real cause, there is 2 hypotheses. From a statistical point of view, there could be the need for more data to allow to create a better curve. From a biological point of view, this could show alterations on the placenta on the late stages of the pregnancy.

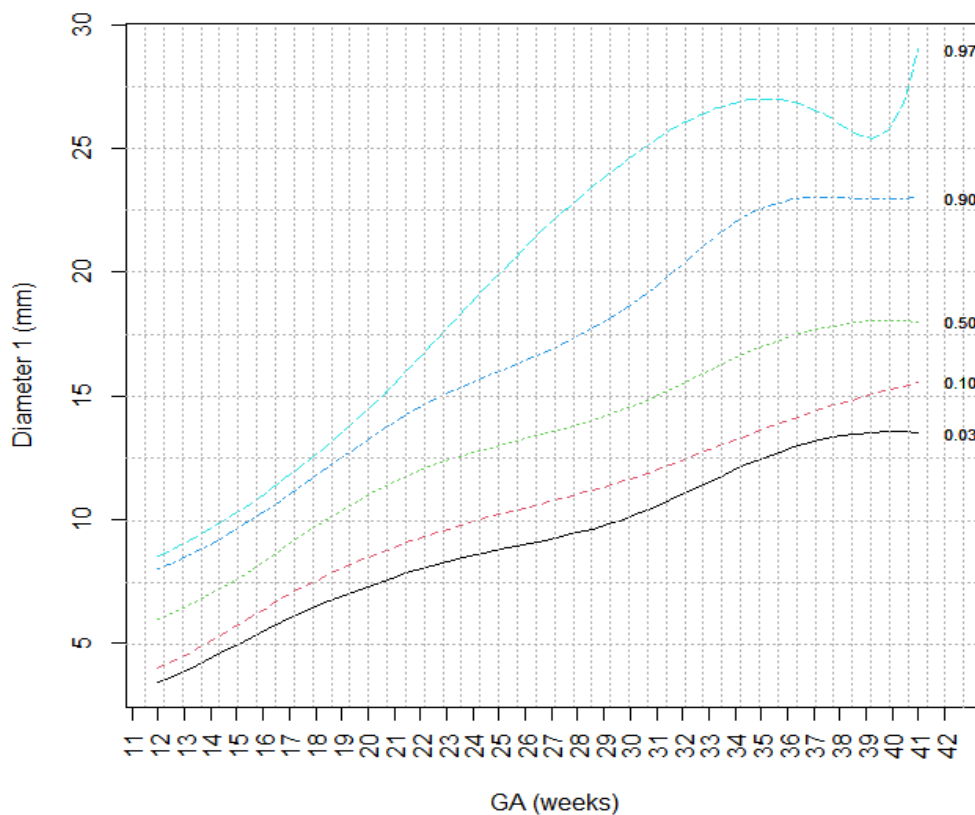


Figure 9 - Growth curve for diameter1 vs gestational age.

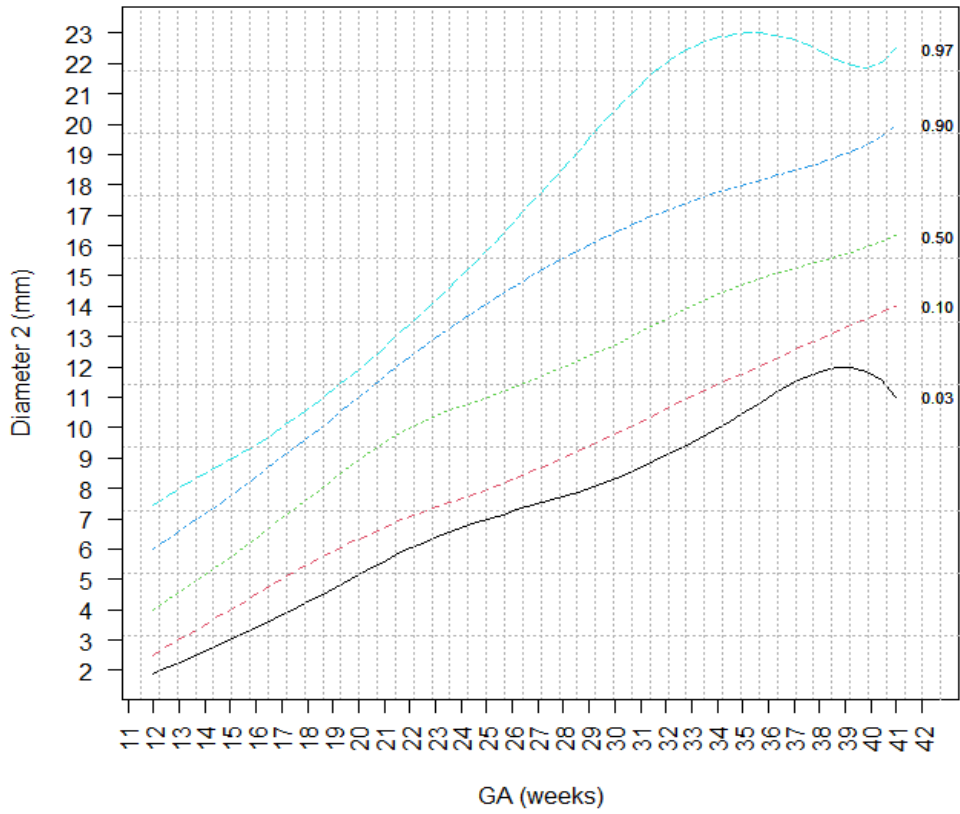


Figure 10 - Growth curve for diameter2 vs gestational age.

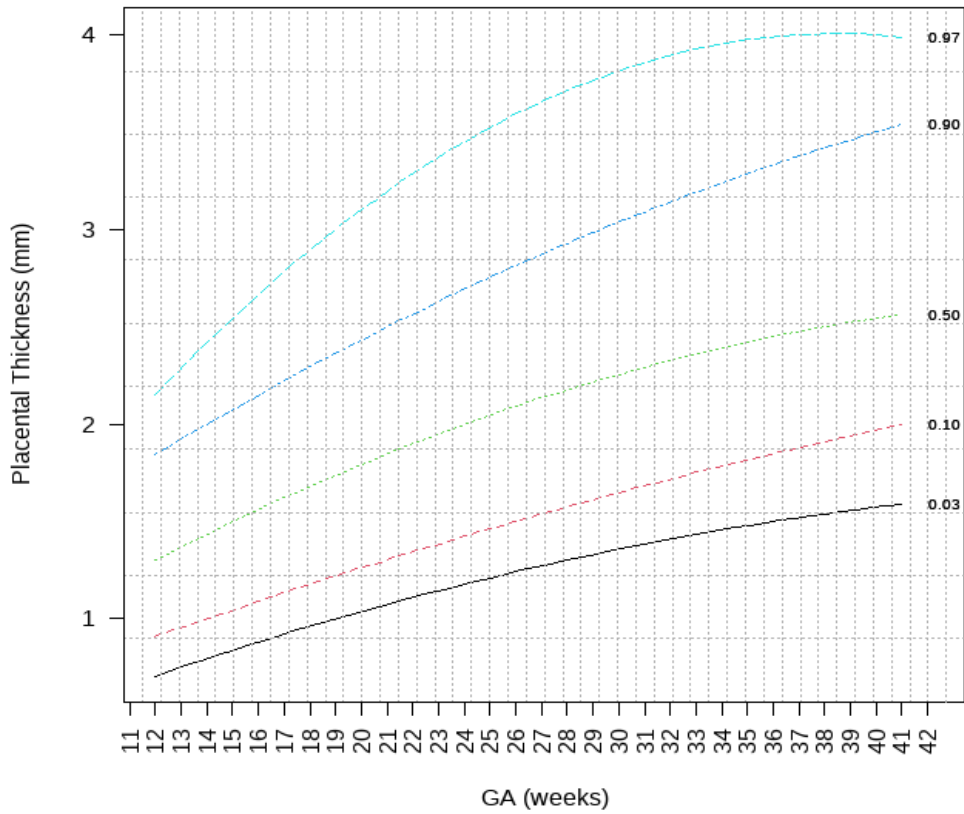


Figure 11 - Growth curve for thickness vs gestational age.

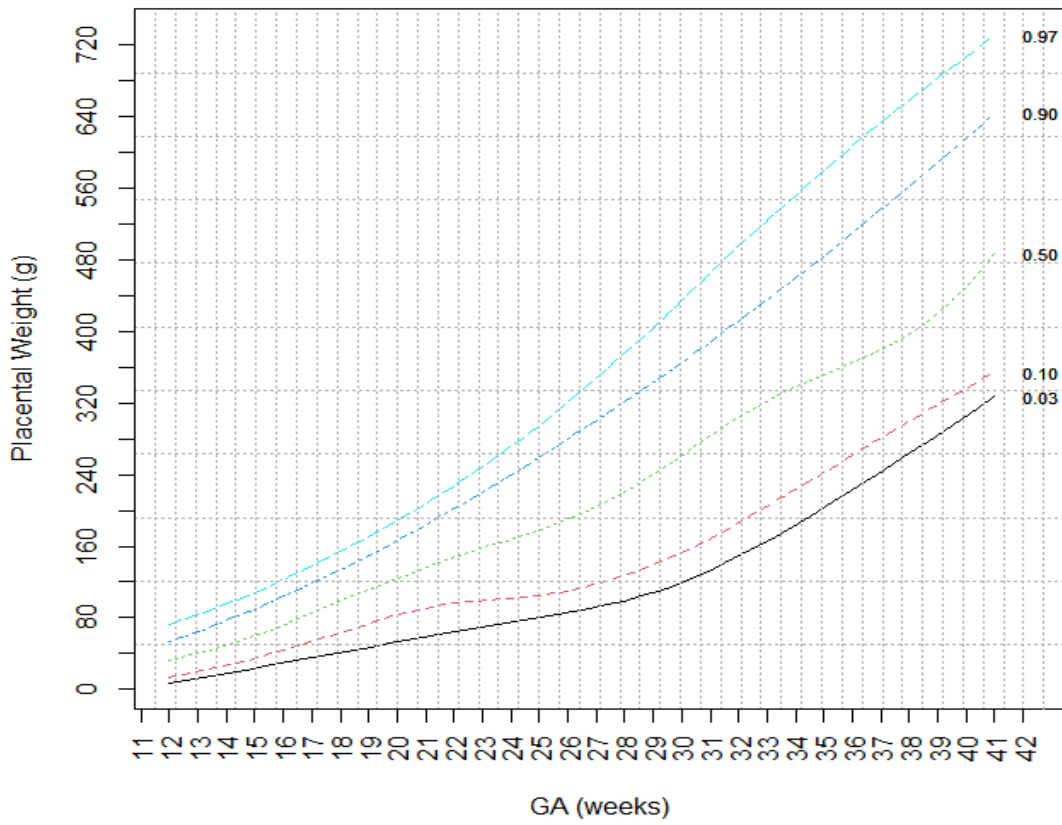


Figure 12 - Growth curve for placental weight vs gestational age.

## 5. APPLICATION DEVELOPMENT

### 5.1 System architecture

#### 5.1.1 General overview

Since the objective of the application created is to allow the doctors, and other professionals in the obstetrician field to be able to create a chart for each client's placenta on a previously created growth chart, great efforts have been made to simulate as much as possible the systems found in hospital and obstetric clinics to create an application that could be used in a real-world situation.

As such, two applications have been made: one that connects to a database to get the patient's information like it would in a hospital (which will from now on called APP1), and the second application that uses the information of the patient on a .csv file (which will from now on called APP2). Both allow the user to see the corresponding values of the placenta on a previously created placenta's growth chart. Both applications' code is on the attachments.

#### 5.1.2 APP1

The app1 was as previously stated, created to as much as possible mimic what would be the setup in an hospital: Connect do a database with the patient's information which will be retrieved and used individually by the patient's doctor to analyse and compare to the previously created placenta's growth chart.

As such, to simulate the hospital's database a Raspberry Pi was used to host a MariaDB database that can be accessed by the application.

The application was created with Rstudio's package *Shiny* and shinydashboard.

#### 5.1.3 Packages used in the APP1

The core of the application is made with Rstudio's package Shiny. This package as previously stated, allows the creation of webapps using R programming language. Another package used was shinydashboard which allows the user to create a shiny webapp, but the layout is already predefined thus allowing the user to create a good looking webapp.

The application layout is defined by three parts: header, sidebar and body as shown in Figure 13.

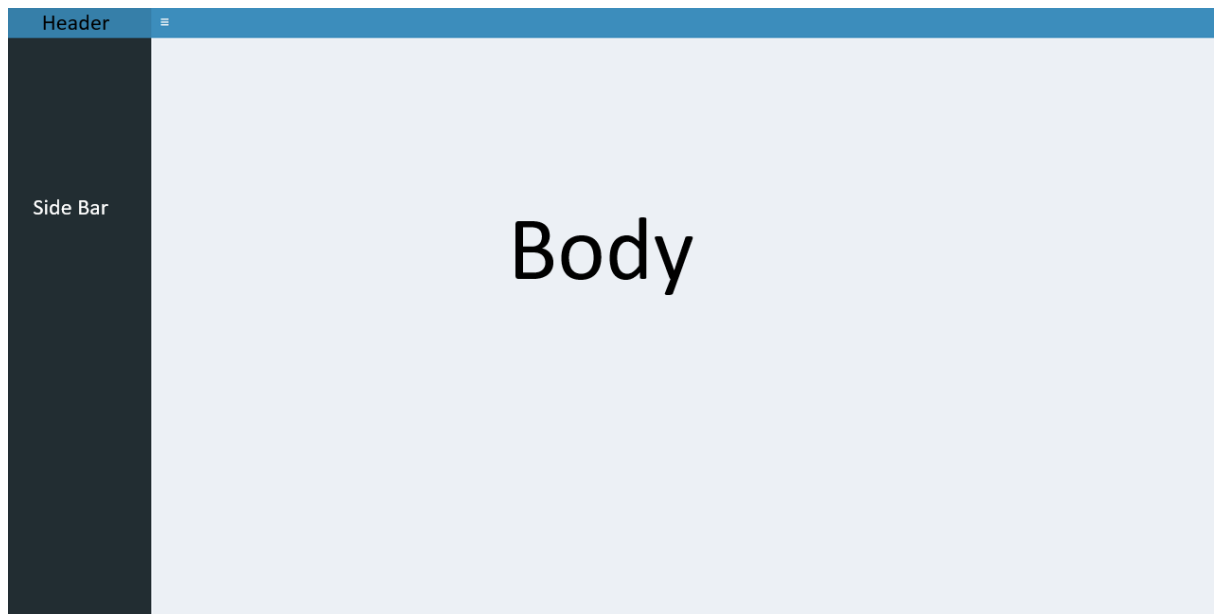


Figure 13 – Representation of the application layout.

#### 5.1.4 APP1

As seen in Figure 14, the workflow on how the APP1 works is shown. It is created a dataframe that will later be plotted and used in the programs (both APP1 and APP2) outside this code. Firstly, the code starts by reading the csv file using the command “read.csv” and the directory is specified. It was then created columns from the file it was opened and the first line removal since that is the intercept line created by the gcrq command. It was also created a column named “GA” and for each iteration on the for loop, creates a line with “GA” and the respective number. It is finally created the dataframe in this example “qr4d1”. The same procedure is then repeated to create the remaining dataframes.

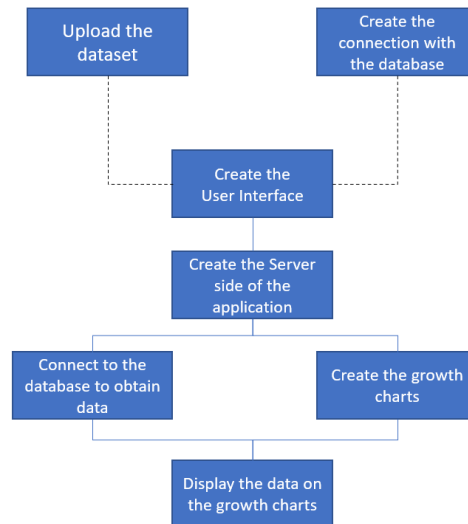


Figure 14 – APP1’s workflow’s fluxogram..

The connection object is created. To do this, it is used the `dbConnect` from the library “`RMariaDB`”. The argument `dbname` indicates the name of the database, the host has the IP address (Internet Protocol address) of the raspberry pi, `username` and `password` are the respective arguments to create the connection.

The next step is the creation of the user interface. In this case it starts by defining the sidebar that has a text input which by default has the word “ID” written that once the user starts to write it disappears. This text input is used to define what is the identification of the mother the user wants to see, that being a number with six digits. The side bar also has action button that when it is pressed it creates a new user. The body has a tab that contains all the plots and tables about the diameter of the placenta.

The server is created by a function with both the input and the output.

After that, the output for the `tbl1` is generated using the library “`DT`” and using its function “`renderDataTable`”. It is then needed to get the necessary data to be displayed. To do that, we get it using a function from the already mentioned library “`RMariaDB`”, “`dbGetQuery`” that sends a query to the data base and using the `id` input given by the user gets the wanted data. The argument `editable` is “`TRUE`” so the table can be edited by the user.

The `observe` function allows the creation of a dynamic function that is waiting for a change in its state and when it does, it triggers an event. In this case the function is waiting for a change in the values of the table. When it does, creates the variable “`info`” which contains information about the edit done by the user, and it has 3 information:

- The column that can be accessed by using “info\$column” which is usually incremented by 1 due to column index offset of 1;
- The row that can be accessed by using “info\$row”;
- Value edited by the user that can be accessed by using “info\$value”.

The value is transformed by using the “as.numeric” function to update the database.

To create the plot, firstly it is needed to get the data by sending a query to the database.

It is then used the library “ggplot” to plot all the growth curves lines and the points.

### 5.1.5 Database

As previously stated, the database was created using MariaDB which since it was created by a former MySQL developer, will have a lot of similarities. The examples in the Figure 15 are of 2 patients (111111 and 222222) and a column for the diameter 1 and diameter 2.

GA	1111111	1111112	2222221	2222222
1	2	2	0	0
2	3	0	0	0
3	4	0	0	0
4	6	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	3	0	0	0
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0
14	0	0	0	0
15	0	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	0	0	0	0
20	8	0	0	0
21	0	0	0	0
22	0	0	0	0

Figure 15 – Example of database created.



### 5.1.6 APP2 code

Like with APP1, the files used to make the growth curves are first uploaded .

The same is done to the values of the d2's growth curves.

This UI (User Interface) does not use the package “shinydashboard” to keep it simpler to the user. As such the webapp is created using only the “shiny” package built-in functions. This UI is composed by 4 tabs: 1 tab for the plot and other for the table for both diameters. It has also a sidebar composed of 2 forms to allow the user to upload the d1 and d2 .csv file containing their values. It also has a button that allows the user to save the changes made on the files.

The server starts by making the input work. The files can be uploaded in a reactive function to allow the data to be dynamically changeable during the execution of the app. This is done for both diameters.

Next 2 reactive objects are created to store the values changed by the user.

The next function is for when the button is pressed, change the .csv file with the corresponding changes. To this, the function from the package rhandsontable. This only happens if there is any change stored in either value1 or value2.

## 5.2 Manual to use APP1 and APP2

This is a small introduction on how to use the applications developed. The values showed on the different figures are not real and have no biological significance.

### 5.2.1 APP1 Instructions

This application as previously stated needs a Raspberry Pi properly set up with the databased and the connection created on the file that needs to be ran on R. This connection on a real-world situation like in an Hospital, could be made with a server, not necessarily a Raspberry Pi.

After the connection is set up, the user can open the application.

After entering the ID of the patient on the box highlighted by the red arrow on Figure 16 the table and plot for both the Diameter 1 and Diameter 2 are shown. The table is divided in pages composed of 10 entries by default and the gestational age can be searched by using the box highlighted by the orange arrow. The plots on the side of the table are the corresponding ones and show the values from the table.



Figure 16 – APP1's main window with client information.

The values on the table can be changed as shown on the entry highlighted by the grey rectangle on the Figure 17. This change will not update the values on the plot.

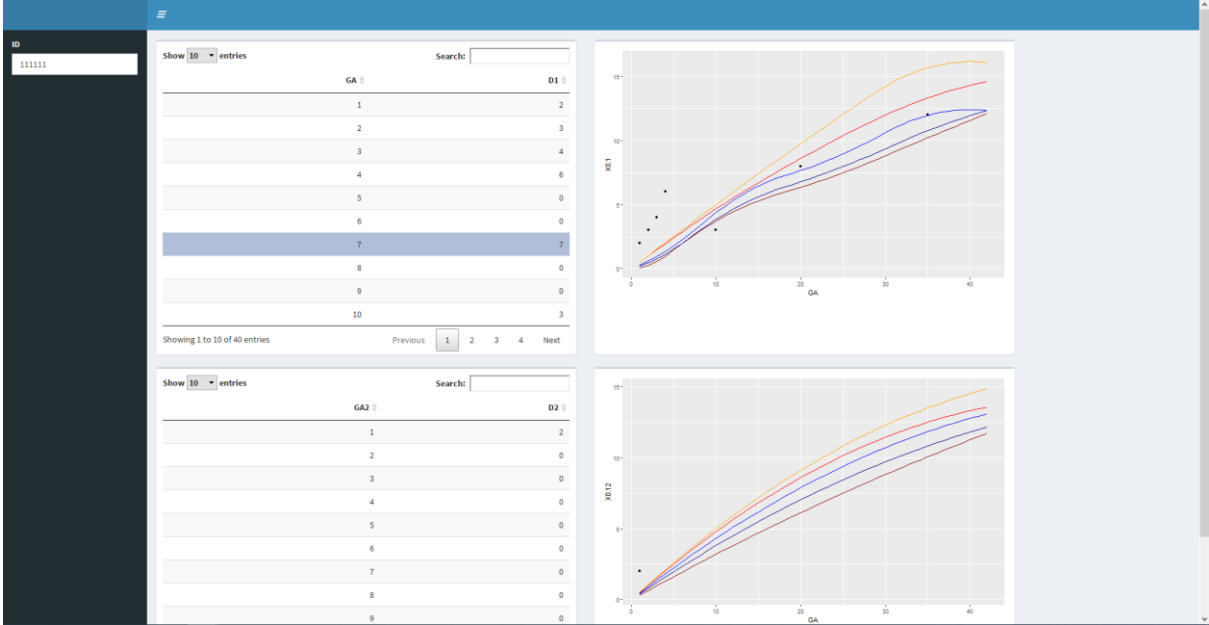


Figure 17- APP1's main window with client information editing.

To update the values on the plot the user needs to reopen the application and the values changed will appear as so on both the table and the plot as shown on Figure 18 highlighted by the red arrows. In this hypothetical scenario the patient would have values in the first gestational weeks higher than the 97<sup>th</sup> percentile and would then stabilize the values to stay on the 50<sup>th</sup> percentil.



The application when open appears as shown on Figure 20. To start the user must upload on the boxes highlighted by a red arrow the previously created csv files for both Diameter 1 and Diameter 2. As seen on Figure 20 this application is divided in tabs to allow a cleaner look to the user.

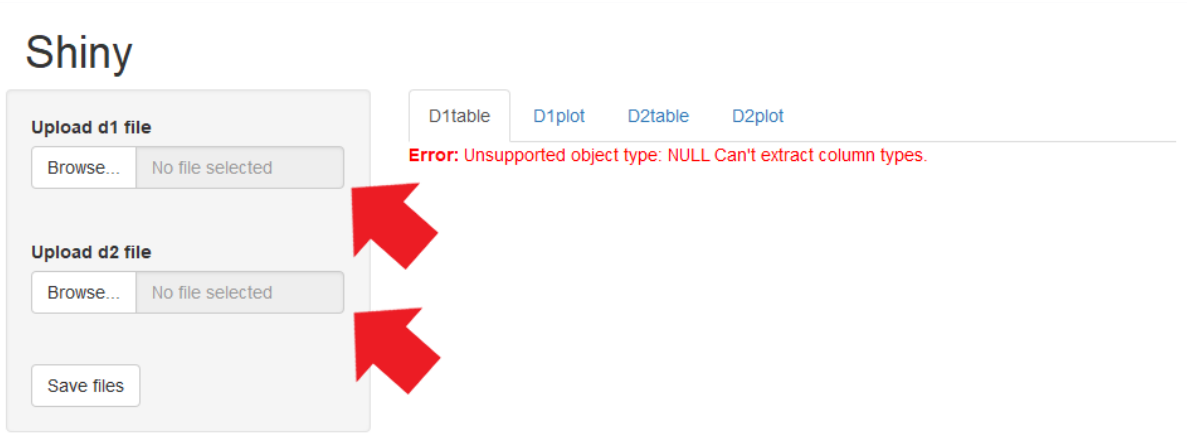


Figure 20 – APP2’s main window.

After uploading the files, the different tabs will show the corresponding table or plot as shown on Figure 21.

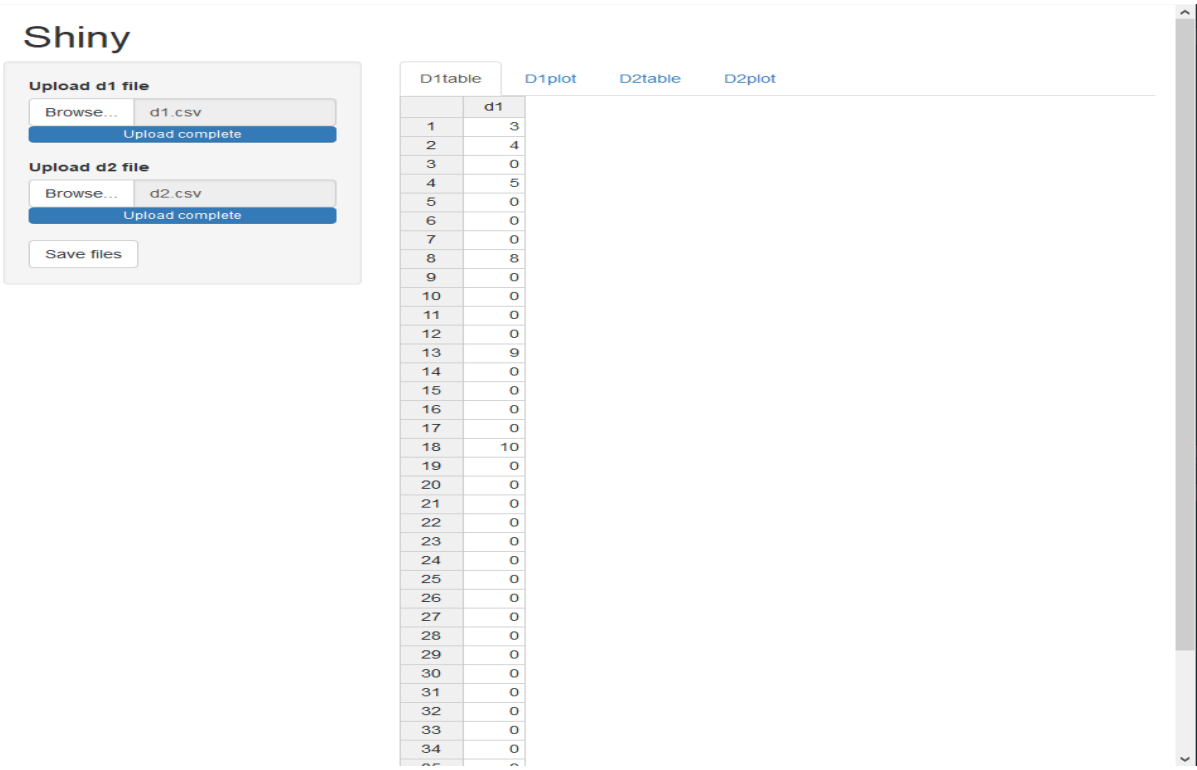


Figure 21 – APP2’s main window with client information.

These tables can be edited as shown on Figure 22 on the marked value of the table. After editing every value the user must click the button “Save files” to save the changes made.

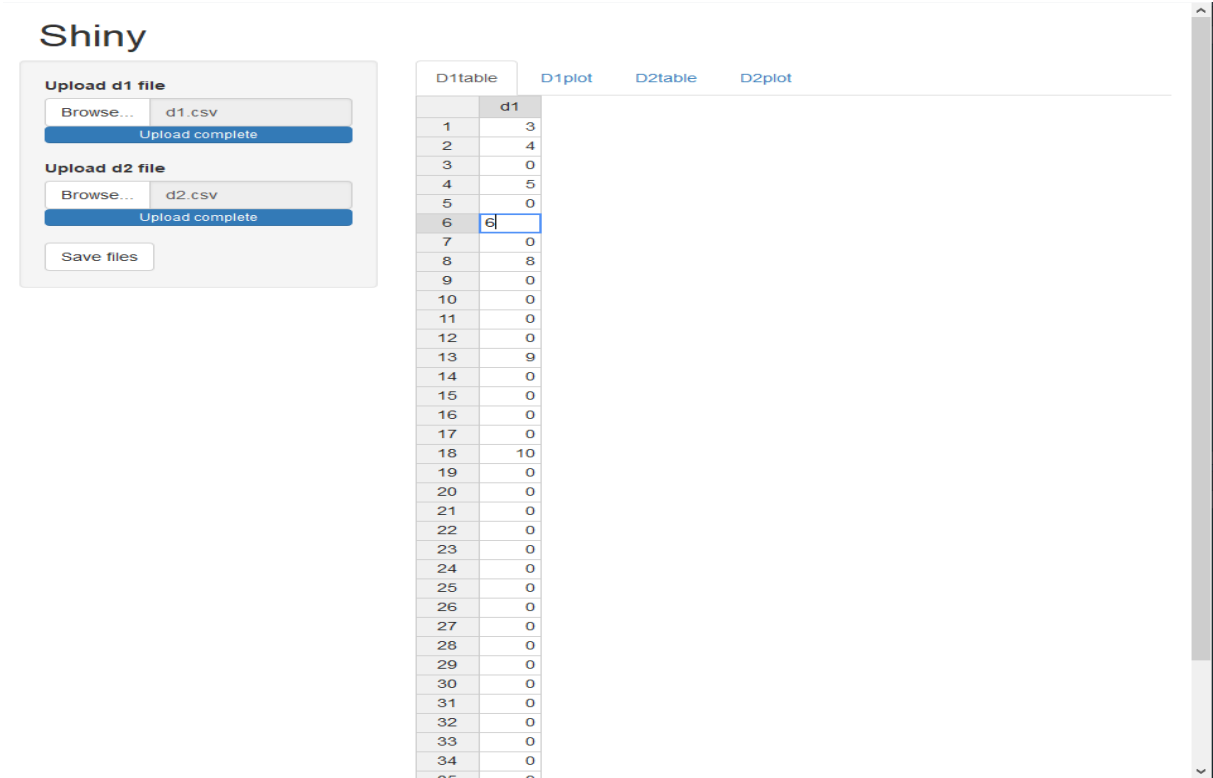


Figure 22 - APP1's main window with client information editing.

There are two important points to be made about the plot as shown on the Figure 23:

- All the values are shown on the plot, including the values that are considered as zero;
- After saving the changes the value is not automatically updated on the plot and as such the user must reopen the application to see the changes made on the plot as seen on the Figure 24.

# Shiny

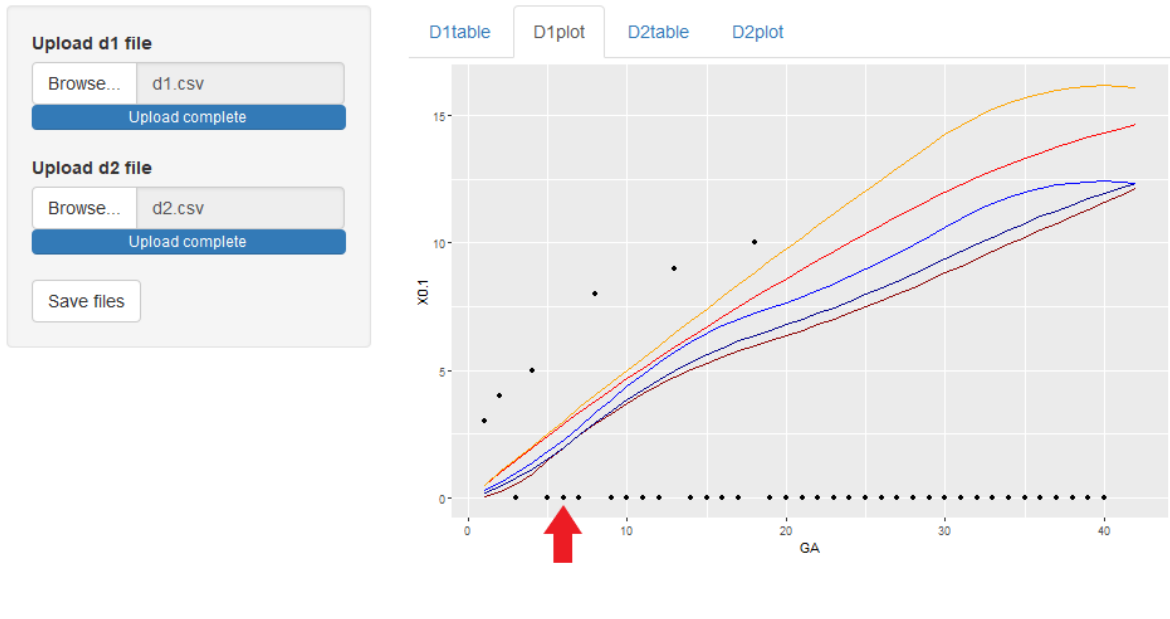


Figure 23 – APP2's client information before editing.

# Shiny

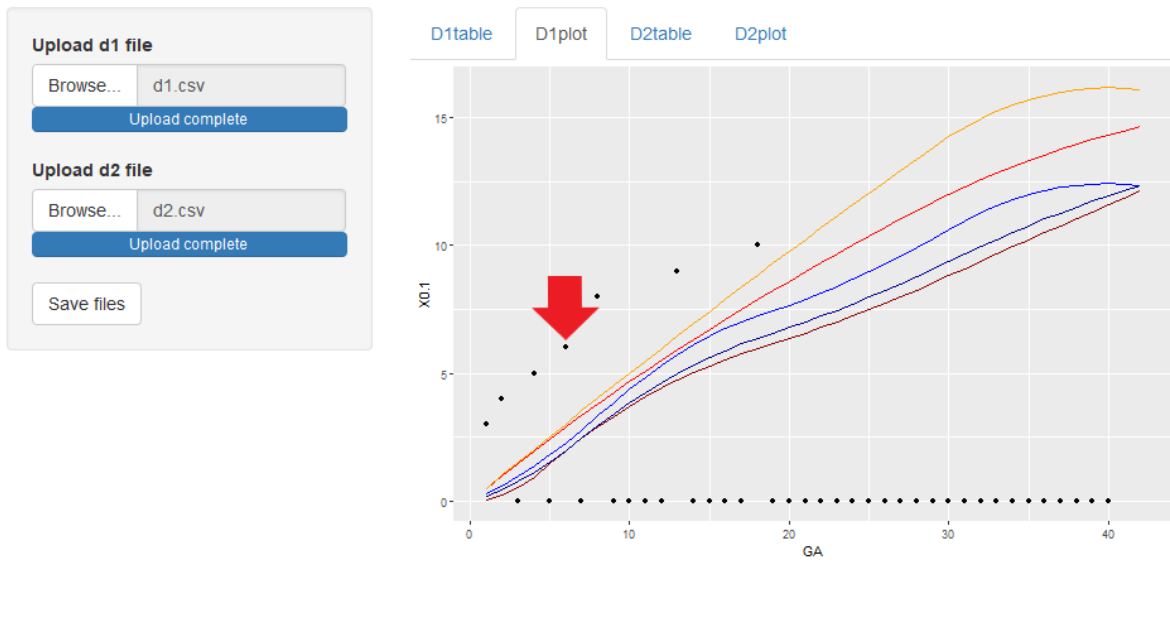


Figure 24 - APP2's client information after editing.

## 6. CONCLUSION AND FUTURE WORK

It is widely accepted that the measurements of the fetus are good indicators of the current and future health of it, but as more research is being done about placentas, new discoveries are being released to the scientific community stating that the placenta and its measurements are also good indicators of the fetus' health.

With this premise in mind, scientists are creating placenta's growth curves for different populations and the work about the Portuguese population was already published by Nogueira et al. (Nogueira et al., 2019). To give a new perspective and possibly improve the work done by Nogueira et al. and allow researchers and doctors to use the technology to their aid, the premises were set for this work: Create new growth curves using a different statistical method compared to the work already published and create a web application that allows the doctors to see the progression of the patient's placenta's measurements and compare them with the newly created curves.

Both objectives were successfully completed. The newly created placenta's growth curves were generated using a more robust regression method consolidating the Portuguese population's placenta growth curves allowing the researchers and doctors to start using these on their studies and patient's appointments to better evaluate the fetus current and future health.

The two applications created were successful at their objectives that is to allow the user to create and work with the measurements gotten by the doctor allowing both the doctor and the researcher to better study the progression of the measurements compared to the growth curves created. APP1 was able to recreate what would happen in a hospital by recreating a connection to a database. APP2 on the other hand was made to be easier to set up and work with when the user is not using a hospital database.

Regarding future works, besides the UI that could be eventually improved, allowing the user to change the values and see the changes immediately on the plots without needing to re-upload the file or update the application. It would also be interesting to set up the applications to a real-world situation and be used by many researchers and doctors and hear their opinions about the applications and start improving based on their insights. An important aspect to consider is the security and as such both applications should go through an analysis such as a static one to be able to detect security flaws (the main one of the APP1 would be SQL injection and as such improve the sanitization of the queries deeper than what is already done). Another interesting idea would be to use machine learning algorithms to predict different outcomes of the fetus although that, besides other constraints, would need to work with a big dataset and that might not yet be available with the size needed.



## REFERENCES

- Anley, C. (2002). Advanced SQL injection in SQL server applications. *NGSSoftware Insight Security Research*, 25.
- Asgharnia, M., Esmailpour, N., Poorghorban, M., & Atrkar-Roshan, Z. (2008). Placental weight and its association with maternal and neonatal characteristics. *Acta Medica Iranica*, 46(6), 467–472.
- Balasundaram, I., & Ramaraj, E. (2012). An efficient technique for detection and prevention of SQL injection attack using ASCII based string matching. *Procedia Engineering*, 30, 183–190.
- Barker, D. J. P., Bull, A. R., Osmond, C., & Simmonds, S. J. (1990). Fetal and placental size and risk of hypertension in adult life. *British Medical Journal*, 301(6746), 259–262.
- Beyerlein, A. (2014). Quantile regression - Opportunities and challenges from a user's perspective. *American Journal of Epidemiology*, 180(3), 330–331.
- Boyd, S. W., & Keromytis, A. D. (2004). SQLrand: Preventing SQL injection attacks. *Proceedings of the 2<sup>nd</sup> Applied Cryptography and Network Security (ACNS) Conference*, 292–302.
- Buhai, S. (2004). Quantile regression: Overview and selected applications. *Ad-Astra-The Young Romanian Scientists' Journal*, February, 1–20.
- Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8), 412–420.
- Chen, C. (2005). Growth charts of body mass index (BMI) with quantile regression. *Proceedings of the 2005 International Conference on Algorithmic Mathematics and Computer Science, AMCS'05*, 1, 114–120.
- Ekiz, H. A., Conley, C. J., Stephens, W. Z., & O'Connell, R. M. (2020). CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. *BMC Bioinformatics*, 21(1), 191.
- Fenton, T. R. (2003). A new growth chart for preterm babies: Babson and Benda's chart updated with recent data and a new format. *BMC Pediatrics*, 3, 1–10.
- Hajovsky, D. B., Villeneuve, E. F., Schneider, W. J., & Caemmerer, J. M. (2020). An Alternative Approach to Cognitive and Achievement Relations Research: An Introduction to Quantile Regression. *Journal of Pediatric Neuropsychology*, 6(2), 83–95.
- Hidalgo, J. S., Oróstica, K. Y., Sanchez-Daza, A., & Olivera-Nappa, Á. (2020). BEST: a Shiny/R web-based application to easily retrieve cross-related enzyme functional parameters and information from BRENDA. *Bioinformatics*, 1–2.
- Hindmarsh, P. C., Geary, M. P. P., Rodeck, C. H., Jackson, M. R., & Kingdom, J. C. P. (2001). Effect of Early Maternal Iron Stores on Placental Weight and Structure. *Obstetric and Gynecologic Survey*, 56(2), 66–67.
- Isakov, K. M. M., Emerson, J. W., Campbell, K. H., Galerneau, F., Anders, A. M., Lee, Y. K., Subramanyam, P., Roberts, A. E., & Kliman, H. J. (2018). Estimated Placental Volume and Gestational Age. *American Journal of Perinatology*, 35(8), 748–757.
- Kiserud, T., Benachi, A., Hecher, K., Perez, R. G., Carvalho, J., Piaggio, G., & Platt, L. D. (2018). The World Health Organization fetal growth charts: concept, findings, interpretation, and application. *American Journal of Obstetrics and Gynecology*, 218(2), S619–S629.
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33–50.
- Lao, T. T., & Wong, W. M. (1999). The neonatal implications of a high placental ratio in small-for-gestational age infants. *Placenta*, 20(8), 723–726.
- McGuinness, L. A., & Higgins, J. P. T. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, 12(1), 55–61.
- Mercatelli, D., Holding, A. N., & Giorgi, F. M. (2020). Web tools to fight pandemics: the COVID-19 experience. *Briefings in Bioinformatics*, 1–11.

- Nogueira, R., Cardoso, P. L., Azevedo, A., Gomes, M., Almeida, C., Varela, C., Braga, A. C., & Pinto, J. C. (2019). Placental Biometric Parameters: The Usefulness of Placental Weight Ratio and Birth/Placental Weight Ratio Percentile Curves for Singleton Gestations as a Function of Gestational age. *Journal of Clinical and Anatomic Pathology*, *4*, 1–15.
- O'Brien, O., Higgins, M. F., & Mooney, E. E. (2020). Placental weights from normal deliveries in Ireland. *Irish Journal of Medical Science*, *189*(2), 581–583.
- Salafia, C. M., Maas, E., Thorp, J. M., Eucker, B., Pezzullo, J. C., & Savitz, D. A. (2005). Measures of placental growth in relation to birth weight and gestational age. *American Journal of Epidemiology*, *162*(10), 991–998
- Salehi, M., Arashi, M., Bekker, A., Ferreira, J., Chen, D.-G., Esmaili, F., & Frances, M. (2021). A Synergetic R-Shiny Portal for Modeling and Tracking of COVID-19 Data. *Frontiers in Public Health*, *8*, 1–10.
- Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(3), 275–294.
- Tiwari, Y., & Tiwari, M. (2015). A Study of SQL of Injections Techniques and their Prevention Methods. *International Journal of Computer Applications*, *114*(17), 31–33.
- Turco, M. Y., & Moffett, A. (2019). Development of the human placenta. *Development (Cambridge)*, *146*(22), 1–14.
- Wei, Y., Pere, A., Koenker, R., & He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, *25*(8), 1369–1382.
- Weisberg, S. (2005). *Applied Linear Regression (Third Edition)* Hoboken, New Jersey: John Wiley & Sons, Inc.
- Wojciechowski, J., Hopkins, A. M., & Upton, R. N. (2015). Interactive pharmacometric applications using R and the Shiny package. *CPT: Pharmacometrics and Systems Pharmacology*, *4*(3), 146–159.
- Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society Series D: The Statistician*, *52*(3), 331–350.
- Yu, Y., Ouyang, Y., & Yao, W. (2018). ShinyCircos: An R/Shiny application for interactive creation of Circos plot. *Bioinformatics*, *34*(7), 1229–1231.

## ATTACHMENTS

### APP1 Code

```
library(shiny)
library(shinydashboard)
library(RMariaDB)
library(pool)
library(DBI)
library(dbplyr)
library(dplyr)
library(tidyverse)
library(ggplot2)
library(plotly)
library(DT)

qr3d1=read.csv(Path_to_file)
X0.1=qr3d1$X0.1
X0.1=X0.1[-1]
X0.25 = qr3d1$X0.25
X0.25=X0.25[-1]
X0.5=qr3d1$X0.5
X0.5= X0.5[-1]
X0.75=qr3d1$X0.75
X0.75=X0.75[-1]
X0.9=qr3d1$X0.9
X0.9=X0.9[-1]
GA = c()
for (i in 1:42){GA[[i]] = i}
qr4d1 = data.frame(GA, X0.1, X0.25, X0.5, X0.75, X0.9)
qr3d2=read.csv(Path_to_file)
X0.12=qr3d2$X0.1
X0.12=X0.12[-1]
X0.252 = qr3d2$X0.25
```

```

X0.252=X0.252[-1]
X0.52=qr3d2$X0.5
X0.52= X0.52[-1]
X0.752=qr3d2$X0.75
X0.752=X0.752[-1]
X0.92=qr3d2$X0.9
X0.92=X0.92[-1]
GA = c()
for (i in 1:42){GA[[i]] = i}
qr4d2 = data.frame(GA, X0.12, X0.252, X0.52, X0.752, X0.92)

```

```

conn <- dbConnect(
  RMariaDB::MariaDB(),
  dbname = "DataBaseName",
  host = "IP",
  username = "UserName",
  password = "Password")
on.exit(dbDisconnect(conn), add = TRUE)

```

```

ui <- dashboardPage(
  dashboardHeader(),
  dashboardSidebar(sidebarMenu(
    textInput("id", "ID")
  )),
  dashboardBody(
    tabItem(tabName = 'diameter',
      fluidRow(box(DT::dataTableOutput('tbl1'), width = 5, height = 500),
        box(plotOutput('plt1'), width = 5, height = 500),
        box(DT::dataTableOutput("tbl2"), width = 5, height = 500),
        box(plotOutput('plt2'), width = 5, height = 500))
    ))
)

```

```

server <- function(input, output) {

output$tbl1 <- DT::renderDataTable(datatable({
dbb <- dbGetQuery(conn, paste0(
"SELECT GA, `", as.integer(input$id)[1], "1` FROM placenta5"))
}, rownames= FALSE, colnames = c('GA', 'D1'), editable = TRUE))

observe(str(input$tbl1_cell_edit))

observeEvent(input$tbl1_cell_edit, {
info = input$tbl1_cell_edit
i = info$row
j = info$col = info$col + 1 # column index offset by 1
v = info$value
info$value1d <- as.numeric(info$value)

dbGetQuery(conn, paste0(
"UPDATE placenta5 SET`", as.integer(input$id)[1], "1`=", info$value1d, " WHERE GA
=", info$row))
}))

output$tbl2 <- DT::renderDataTable(datatable({

dbb <- dbGetQuery(conn, paste0(
"SELECT GA, `", input$id, "2` FROM placenta5"))
}, rownames= FALSE, colnames = c('GA2', 'D2'), editable = TRUE))

observe(str(input$tbl2_cell_edit))

observeEvent(input$tbl2_cell_edit, {
info = input$tbl2_cell_edit
i = info$row
j = info$col = info$col + 1 # column index offset by 1
v = info$value
info$value2d <- as.numeric(info$value)

```

```

dbGetQuery(conn, paste0(
"UPDATE placenta5 SET`", as.integer(input$Id)[1],"2`=",info$value2d," WHERE GA
=",info$row))
})

```

```

output$plt1 <- renderPlot({

```

```

dbq1 <- data.frame(dbGetQuery(conn, paste0(
"SELECT GA, `", as.integer(input$Id)[1],"1` FROM placenta5 WHERE `", input$Id,"1`
>0")))

```

```

p <- ggplot(qr4d1, aes(x=GA))+ geom_line(aes(y=X0.1), color="darkred")+
geom_line(aes(y=X0.25), color="darkblue")+ geom_line(aes(y=X0.5), color="blue")+
geom_line(aes(y=X0.75), color="red")+geom_line(aes(y=X0.9), color="orange")+
geom_point(dbq1, mapping =aes(x=GA, y=dbq1[[2]]))

```

```

print(p)

```

```

})

```

```

output$plt2 <- renderPlot({

```

```

dbq2 <- data.frame(dbGetQuery(conn, paste0(
"SELECT GA, `", as.integer(input$Id)[1],"2` FROM placenta5 WHERE `", input$Id,"2`
>0")))

```

```

p2 <- ggplot(qr4d2, aes(x=GA))+ geom_line(aes(y=X0.12), color="darkred")+
geom_line(aes(y=X0.252), color="darkblue")+ geom_line(aes(y=X0.52), color="blue")+
geom_line(aes(y=X0.752), color="red")+geom_line(aes(y=X0.92), color="orange")+
geom_point(dbq2, mapping =aes(x=GA, y=dbq2[[2]]))

```

```

print(p2)

```

```

})

```

```

}

```

```

shinyApp(ui, server)

```

## APP2 Code

```
library(shiny)
```

```
library(DT)
```

```
library(tidyverse)
```

```
library(ggplot2)
```

```
library(plotly)
```

```
library(rhandsontable)
```

```
qr3d1 = read.csv(Path_to_file)
```

```
X0.1 = qr3d1$X0.1
```

```
X0.1 = X0.1[-1]
```

```
X0.25 = qr3d1$X0.25
```

```
X0.25 = X0.25[-1]
```

```
X0.5 = qr3d1$X0.5
```

```
X0.5 = X0.5[-1]
```

```
X0.75 = qr3d1$X0.75
```

```
X0.75 = X0.75[-1]
```

```
X0.9 = qr3d1$X0.9
```

```
X0.9 = X0.9[-1]
```

```
GA = c()
```

```
for (i in 1:40) {
```

```
GA[[i]] = i
```

```
}
```

```
qr4d1 = data.frame(GA, X0.1, X0.25, X0.5, X0.75, X0.9)
```

```
qr3d2 = read.csv(Path_to_file)
```

```
X0.12 = qr3d2$X0.1
```

```
X0.12 = X0.12[-1]
```

```
X0.252 = qr3d2$X0.25
```

```
X0.252 = X0.252[-1]
```

```
X0.52 = qr3d2$X0.5
```

```
X0.52 = X0.52[-1]
```

```

X0.752 = qr3d2$X0.75
X0.752 = X0.752[-1]
X0.92 = qr3d2$X0.9
X0.92 = X0.92[-1]
GA = c()
for (i in 1:40) {
  GA[[i]] = i
}
qr4d2 = data.frame(GA, X0.12, X0.252, X0.52, X0.752, X0.92)

```

```

ui <- fluidPage(headerPanel(title = "APP2"),
  sidebarLayout(
    sidebarPanel (
      shiny::fileInput(inputId = "file_id1", label = "Upload d1 file"),
      shiny::fileInput(inputId = "file_id2", label = "Upload d2 file"),
      actionButton("do", "Save files")
    ),
    mainPanel(tabsetPanel(
      type = "tab",
      tabPanel("D1table", rHandsontableOutput('d1tbl')),
      tabPanel("D1plot",plotOutput('plt1')),
      tabPanel("D2table", rHandsontableOutput('d2tbl')),
      tabPanel("D2plot",plotOutput('plt2'))
    ))
  ))

```

```

server <- function(input, output, session) {
  d1 = reactive({
    file = input$file_id1
    if (is.null(file)) {
      return()
    }
  })
}

```



```

else {
read.csv(file = file$datapath)
}
})

d2 = reactive({
file = input$file_id2
if (is.null(file)) {
return()
}
else {
read.csv(file = file$datapath)
}
})

# d1 <- reactive({
# df1()
# })
# d2 <- reactive({
# df2()
# })
#
#
values1 <- reactiveValues()
values2 <- reactiveValues()

observeEvent(input$do, {
values1$data <- hot_to_r(input$d1tbl)
values2$data <- hot_to_r(input$d2tbl)
if (is.null(values1$data)) {
return()
}
else {

```

```

write.csv(
values1$data,
Path_to_file,
row.names = FALSE
)
}
if (is.null(values2$data)) {
return()
}
else {
write.csv(
values2$data,
Path_to_file,
row.names = FALSE
)
}
})

```

```

output$d1tbl = renderRHandsontable(rhandsontable(d1(), rownames = FALSE, editable =
TRUE))

```

```

output$d2tbl = renderRHandsontable(rhandsontable(d2(), rownames = FALSE, editable =
TRUE))

```

```

output$NewIris <- DT::renderDataTable({
values$data
})

```

```

output$plt1 <- renderPlot({
# a<- dbqr$NIF2
#print(qr)
p <-

```

```

ggplot(qr4d1, aes(x = GA)) + geom_line(aes(y = X0.1), color = "darkred") +
geom_line(aes(y = X0.25), color = "darkblue") + geom_line(aes(y =
X0.5), color = "blue") +
geom_line(aes(y = X0.75), color = "red") +
geom_line(aes(y = X0.9), color = "orange") +
geom_point(d1(), mapping = aes(x = GA, y =
d1()[[1]]))
#p<-plotly(p)
print(p)
#plot(dbqr)

observe(str(input$tbl1_cell_edit))

observeEvent(input$tbl1_cell_edit, {
info = input$tbl1_cell_edit
i = info$row
j = info$col = info$col + 1 # column index offset by 1
v = info$value
info$value1d <-
as.numeric(info$value)

# dbGetQuery(conn, paste0(
# "UPDATE placenta5 SET ", input$id, "1`=",info$value1d," WHERE GA =",info$row))

write.csv(x = d1(), file = "data_final_d2.csv")

})

})

output$plt2 <- renderPlot({
# a<- dbqr$NIF2
#print(qr)
p <-

```

```
ggplot(qr4d2, aes(x = GA)) + geom_line(aes(y = X0.1), color = "darkred") +
geom_line(aes(y = X0.25), color = "darkblue") + geom_line(aes(y =
X0.5), color = "blue") +
geom_line(aes(y = X0.75), color = "red") +
geom_line(aes(y = X0.9), color = "orange") +
geom_point(d2(), mapping = aes(x =
GA, y = d2()[[1]]))
#p<-plotly(p)
print(p)
#plot(dbqr)
})

}
shinyApp(ui, server)
```