

Retina Image Analysis

Students : Alexandre Jann, Maylis Touya, Paola Zanchi
Teaching Assistant : Michael Beyeler

Emails : alexandre.jann@unil.ch, maylis.touya@unil.ch, paola.zanchi@unil.ch.

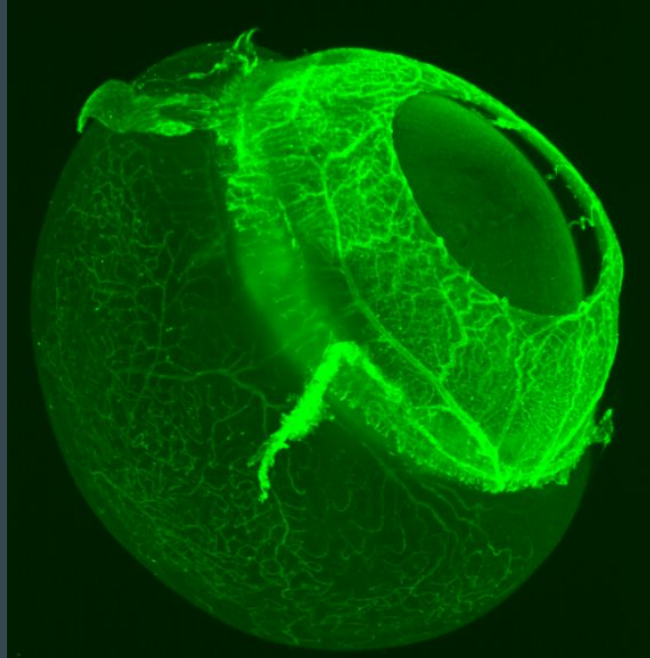


Image of mouse eyeball taken with light-sheet
fluorescent microscopy, with the blood vessels
shown in green.

Prahst et al: eLife paper 2020

Plan

1. Goals of the project
2. Introduction
3. Material and Methods
4. Results
5. Discussion
6. Challenges
7. Feedback
8. Conclusion
9. Questions ?

Goals of the project

- *Linking the tortuosity of blood vessels in the eyes with cardiovascular diseases.*
- *Finding and verifying our results by using machine learning algorithm.*

Introduction

Cardiovascular diseases (CVDs)

17 May 2017

Key facts

- CVDs are the number 1 cause of death globally: more people die annually from CVDs than from any other cause.
- An estimated 17.9 million people died from CVDs in 2016, representing 31% of all global deaths. Of these deaths, 85% are due to heart attack and stroke.
- Over three quarters of CVD deaths take place in low- and middle-income countries.
- Out of the 17 million premature deaths (under the age of 70) due to noncommunicable diseases in 2015, 82% are in low- and middle-income countries, and 37% are caused by CVDs.
- Most cardiovascular diseases can be prevented by addressing behavioural risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies.
- People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management using counselling and medicines, as appropriate.

Hypothesis

“Can we predict if someone had a cardiovascular disease by looking at its eye fundus blood vessels’ tortuosity and diameter plus other features ?”

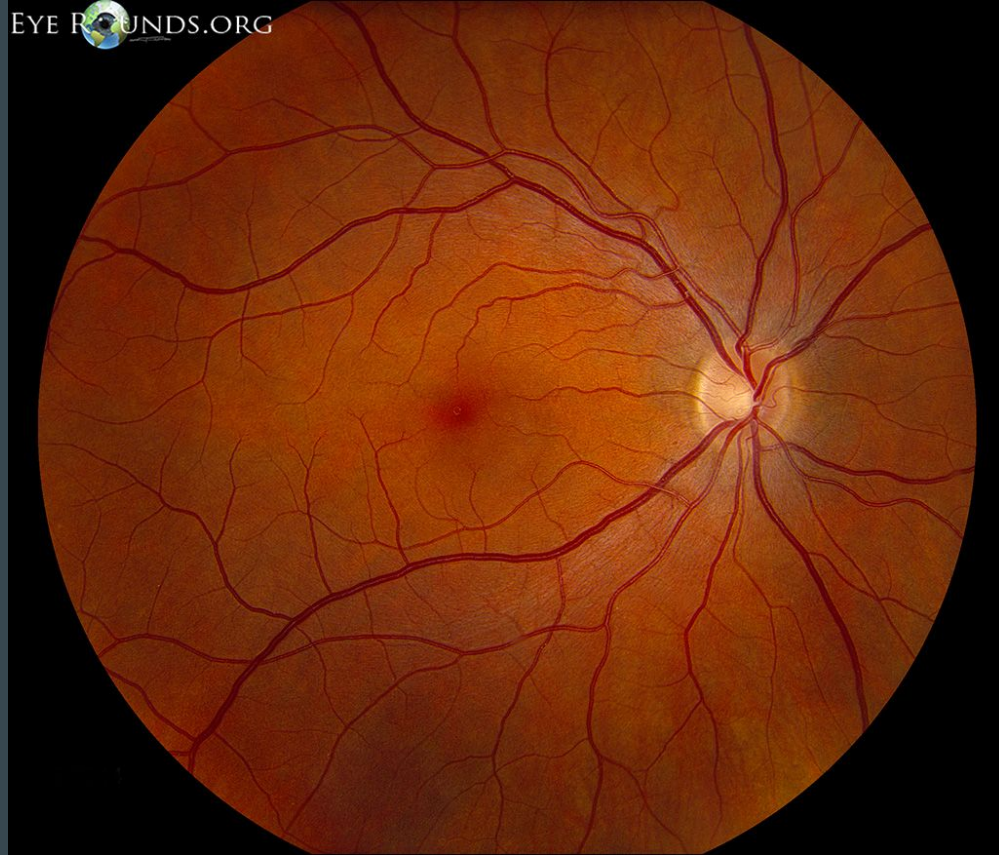
Material and Methods

Material : Eye Fundus Snapshot



https://fyidoctors-assets.s3.ca-central-1.amazonaws.com/assets/media/dilated-eye-exams-fyi-no-logo_VisiqueBlogThumbnail.jpg

EYE FUNDUS.ORG



<https://webeye.opth.uiowa.edu/eyeforum/atlas/LARGE/Normal-fundus-LRG.jpg>

Material and Methods

Material : UK BioBank



https://images.squarespace-cdn.com/content/v1/57b3657b6a496370ca7dfb39/1528916706931-GMQ9MXM8RAIO35FIIIT7I/ke17ZwdGBToddI8pDm48kJyIDfcmoFwXcVU90f0-meIUqsxRUqqbr1mOJYKfIPR7LoDQ9mXPOjoJoqy81S2I8N_N4V1vUb5AoIIbLZhVYxCRW4BPu10St3TBAUQYVKcJ4Sq8bkbnF_lwCSUC3pZDwFbu8TocOAAEacIx-zfpkYS68d3vaqp5M25InYIQyD/UKB_banner.jpg

Material and Methods

Material : Software and Group's Server



<https://upload.wikimedia.org/wikipedia/commons/thumb/c/c3/Python-logo-notext.svg/768px-Python-logo-notext.svg.png>



<http://clisp.fr/wp-content/uploads/2020/03/RStudio-Ball.png>



https://images.idgesg.net/images/article/2020/05/server_racks_close-up_perspective_shot_by_monsitj_gettyimages-918951042_cw_cio_2400x1600-100841601-large.jpg

Material and Methods

Methods : ARIA



<https://www.me.iastate.edu/bglab/files/2014/10/AriaLogo.png>



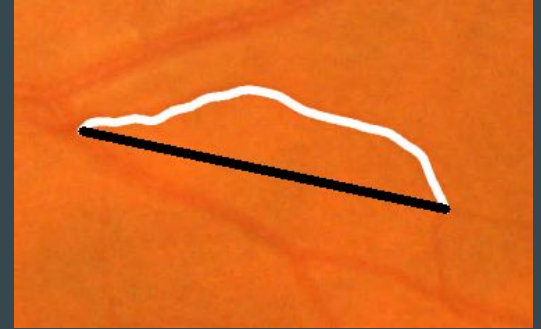
<https://summitlife.org/good-roots-summit-life-today/>

Material and Methods

Methods : Tortuosity Measurement

Distance Factor (1st tortuosity measure)

$$DF = \frac{\text{Total length of the blood vessel}}{\text{Length between start and end point of the blood vessel}}$$



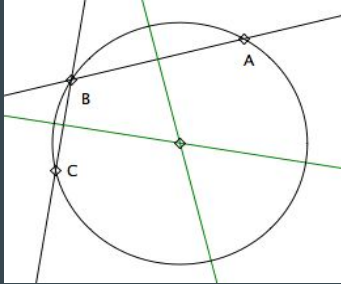
In a more mathematical way :

$$DF = \frac{\sum_{i=0}^{l-1} \text{distance}(\text{point}_i, \text{point}_{i+1})}{\text{distance}(\text{point}_0, \text{point}_l)}$$

Material and Methods

Methods : Tortuosity Measurement

Curvature tortuosity (2nd tortuosity measure)

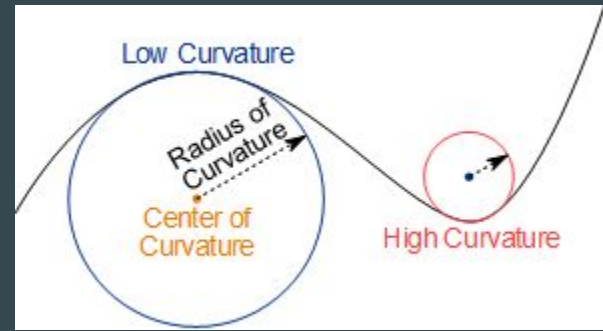


Algorithm:

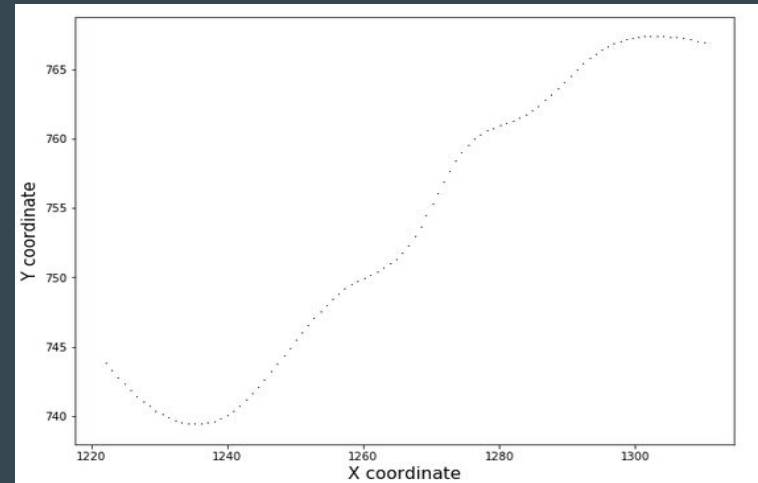
Find the center of the circle for each set of 3 points along a vessel

Compute and sum $1/\text{radius}$

Divide by the length of the vessel



From Michael Beyeler's presentation



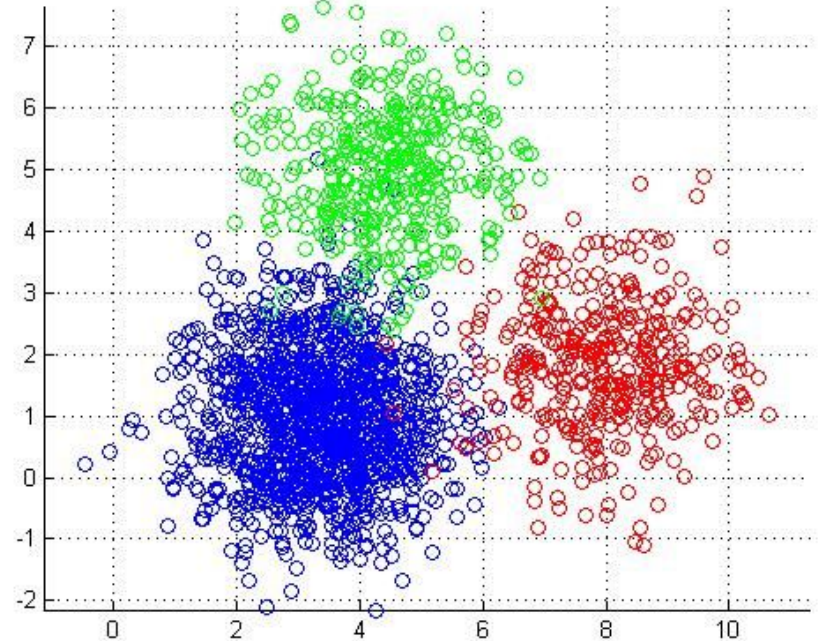
Material and Methods

Methods : K-Means

Machine learning Algorithm:

- A way to cluster data and make groups
- It's based on nearest mean values
- In k-means, the k stands for the number of clusters

→ Is, in the end, a table of truth, and therefore linked to an accuracy measurement



<https://it.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/24616/versions/14/screenshot.jpg>

Material and Methods

Methods : Accuracy measurement

Is used to know if a denotes **the closeness** of computations or estimates to the exact or **true values**

Formula :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

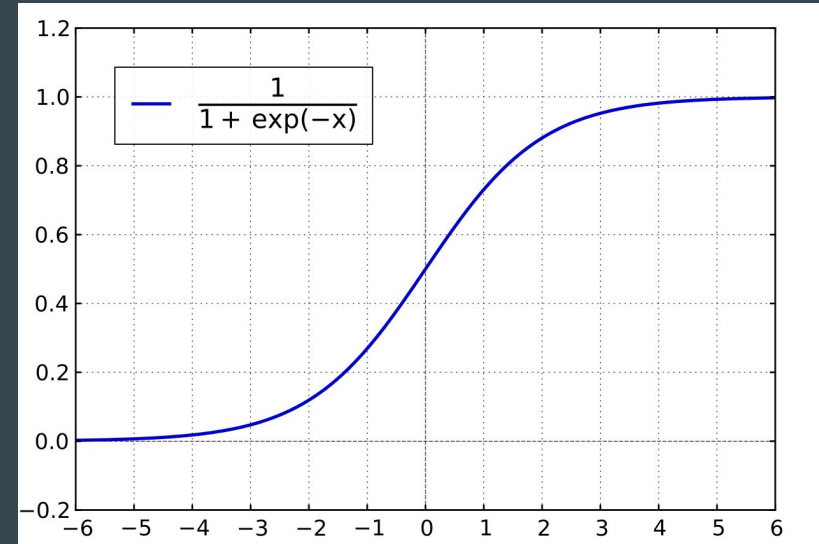
https://miro.medium.com/max/512/1*-BkpqhN-5fPicMifDQ0SwA.png

Material and Methods

Methods : Logistic Regression

- Statistical model used to predict the probability of binary events (such as sick vs healthy).
- It is more suited to this type of prediction than linear regression

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

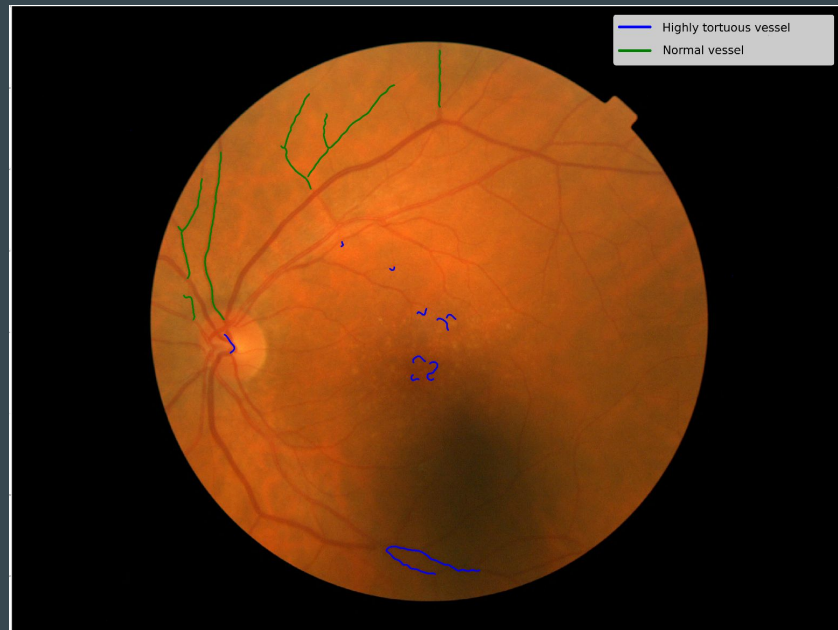


<https://d33wubrfki0l68.cloudfront.net/9297bafa66fc69e71ee94cc17972a8c60fe75768/b74cb/wp-content/uploads/2018/07/logistic-regression.png>

Results - ARIA's Artifacts discovery

We highlighted some artifacts thanks to our measure of tortuosity

Highly tortuous vessels appears to be artifacts generated by ARIA



Results - model construction

Compute the Distance factor and mean diameter for each vessel of each eye fundus



<https://upload.wikimedia.org/wikipedia/commons/thumb/c/c3/Python-logo-notext.svg/768px-Python-logo-notext.svg.png>

First filter of the data using bash to avoid having to load too much unnecessary data in R
(to save time)



<https://www.linuxjournal.com/content/understanding-bash-elements-programming>

Select one eye for each participant :

- compute the median tortuosity
- or remove outliers vessels and compute the mean tortuosity

Select features from the UK BioBank database and add eye fundus measurement to it

Fit the data into a linear or a logistic regression model



<https://stat.ethz.ch/R-manual/R-devel/doc/html/NEWS.html>

Results

- Eventually, we settled on a **logistic regression** model where we **removed outlier vessels** and compute the **mean** which is more suited for the kind of predictions that we wanted to make.

```
logistic regression : disease ~ vessel tortuosity, vessel diameter, diastolic pressure,  
                        systolic pressure, BMI, age, sex
```

Angina results:

angina.pred	FALSE	TRUE
FALSE	12145	322
TRUE	1	1

Accuracy : 0.974

Strokes results:

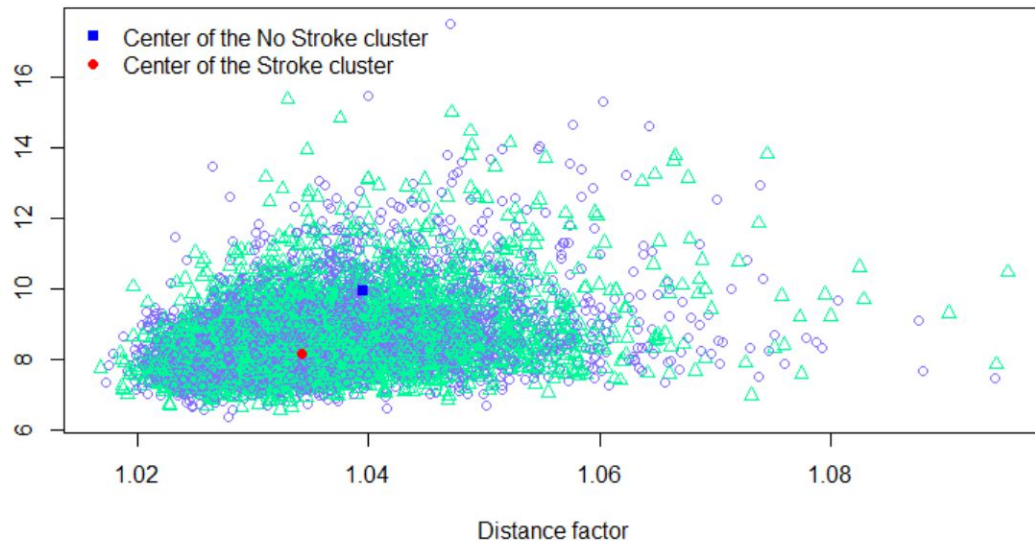
strokes.pred	FALSE	TRUE
FALSE	12282	185
TRUE	2	0

Accuracy : 0.985

Results - K-means clustering

true	cluster	
	No Stroke Cluster	Stroke Cluster
No Stroke	2081	7328
Stroke	51	122

Plot of the distance factor against the Diameter with
as a determining factor the presence of not of strokes



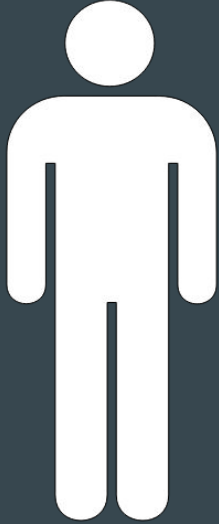
$$Accuracy = \frac{2081 + 122}{2081 + 122 + 7328 + 51}$$

$$Accuracy \cong 0,23$$

Results

- The first model we created was designed to predict the diastolic pressure with a linear model but it wasn't very smart because you can measure it precisely with a simple instrument.
- Attempt to construct the model in R took more than 12 hours on the Brynhild server to eventually give no output (not even error or warning). We guess it crashed but it was working on a smaller sample.
- Then we rewrote the script to make use of R function instead of coding every with for loops and ... it worked ! Scripts took from 7 to 12 minutes depending on the algorithm to give us linear models.

Discussion - A biased dataset



503 806 white people

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRSBCXqRv9Dlp2-Lqu3sXy9Ghib5b7AziL_4lnF-NSrCHuOOW_oJjEjsy6nyr_WzYOzq1U&usqp=CAU

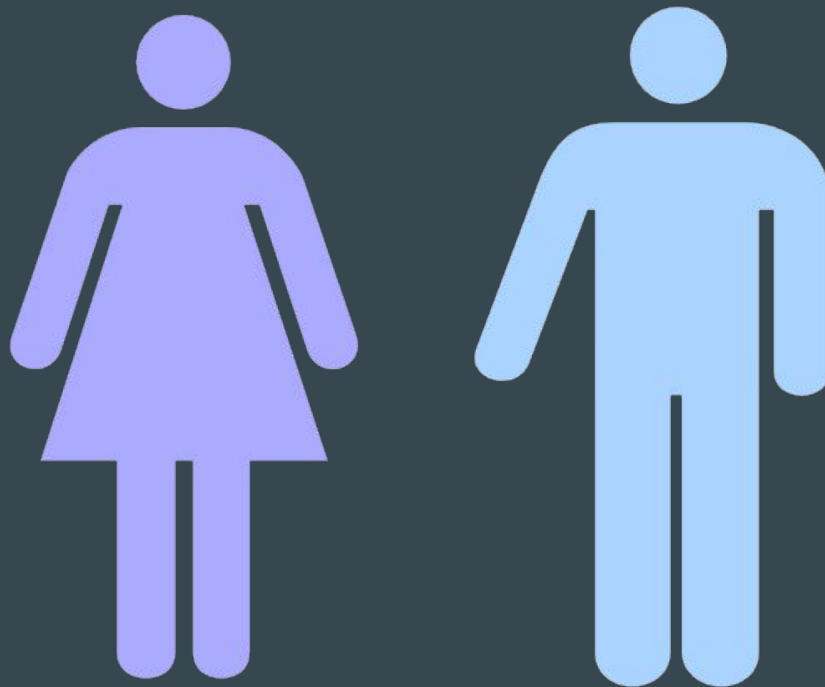


3 095 mixed people
10 000 Asian or Asian British people
8 221 Black or Black British people
1 653 Chinese people
etc ...

https://www.vippng.com/png/full/1-12562_dance-color-silhouette-diagram-play-png-image-with.png

Discussion - A biased dataset

These numbers have been taken from the “Genetic sex” data field of the UK Biobank

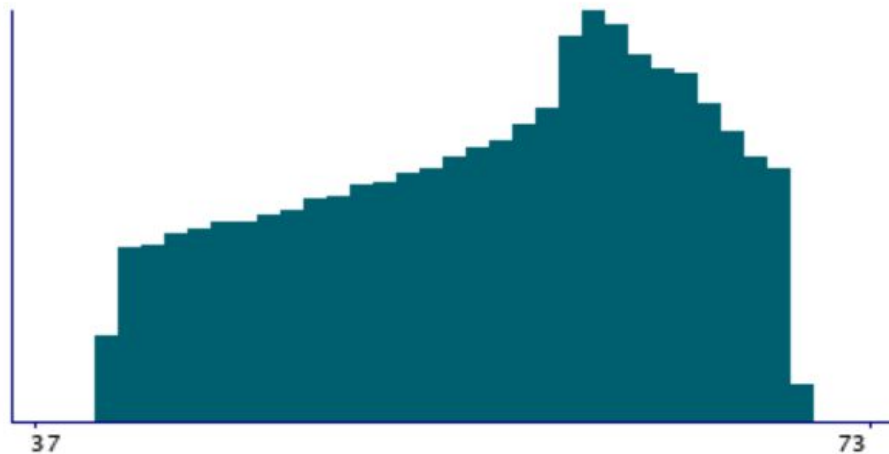


264 780 females for 223 457 males

Discussion - A biased dataset

502,478 items of data are available, covering 502,478 participants.
Units of measurement are years.

Maximum	73
Decile 9	67
Decile 8	64
Decile 7	62
Decile 6	60
Median	58
Decile 4	55
Decile 3	52
Decile 2	48
Decile 1	44
Minimum	37



- There are 37 distinct values.
- Mean = 56.5287
- Std.dev = 8.09522

Discussion - ARIA : a good or a bad choice ?

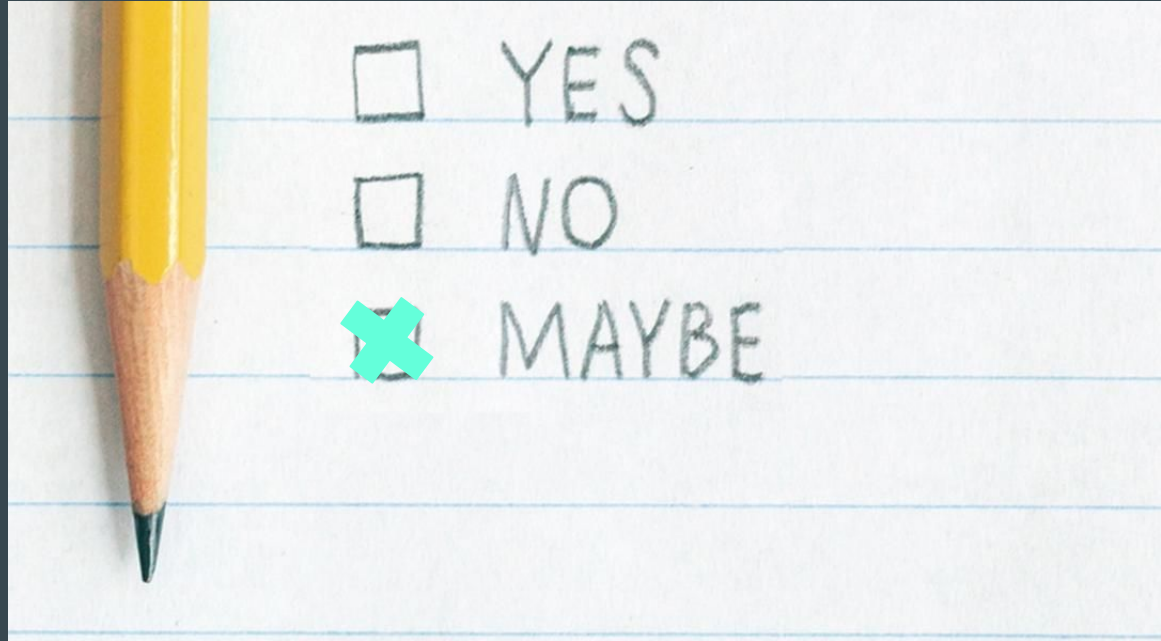
Pros:

- Made for measuring tortuosity
- Is an “easy to use” software
- Only needs images
- Gives a good general idea of the tortuosity of the blood vessels
- Gives a general idea of the diameter of the blood vessels

Cons:

- Does not “understand” when blood vessels cross one another
- Can’t “read” on the edge of the image
- Is bad at giving a precise diameter the precise path of the blood vessels
- Creates artifacts

Discussion - What did we found and what to do with it ?



Discussion - What could have been done better ?



https://www.sciencemag.org/sites/default/files/styles/article_main_large/public/images/c_careers_highlighting-lines-16x9.jpg?itok=A6YowzTM

Challenges

- The current global pandemic of COVID-19
- Working as a team of future scientist with everyone's strength and weaknesses
- Uncooperative computer and serious computer issues
- Issue to connect to the server
- 12 hours running code for nothing
- R programming on the server without RStudio IDE
- Filtering and modifying data using basic bash commands
- Using several different programming languages while keeping the analysis reproducible (creating scripts instead of using one time commands)



<https://greekmythology.wikia.org/wiki/File:ATLAS-ET-LES-HESPERIDES.jpg>

Feedback

- Freedom to explore the data on our own and develop ideas to extract the informations from eye fundus was stimulating
- Working on real data was useful because it showed us that sometimes everything does not go as planned
- Not focused enough on a single path at the end of the project so we tried a lot of different things that leads to disappointment
- Not having a UNIL provided Matlab license was a bit sad, but it's no-one's fault

Conclusion

- Even if we did not find anything accurate enough, eye fundi might be useful,
- The fact that eye fundus snapshots are a non invasive method is a huge advantage, showing that this method might have more potential in the future.
- Maybe if we had more precise data on cardiovascular diseases it would have been easier and it might still be useful as a proxy of the cardiovascular system

→ More in depth research can be done.

Take-Home Messages

- Programming on a server is different from programming in your own development environment,
- The final goal of every statistical analysis is the biological or clinical application and that is something that we shall not forget,
- Science being science, the results are not always good,
- A data set might be enormous but that does not mean it might be useful,
- Eye fundi might have a true potential in diagnosing diseases but we did not find it.

Questions ?



<https://www.avisto.com/wp-content/uploads/2018/11/questions-poseur-entretien.jpg>

Emails : alexandre.jann@unil.ch, maylis.touya@unil.ch, paola.zanchi@unil.ch.