Retina Image Analysis

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

Plan

- The Aim of the project
- Background
- Preliminary Results
- What needs to be done ?

The Aim of the Project

Aim

“We wanted to know if we can link tortuosity of the blood vessels in the eyes with cardiovascular diseases by using programming and bioinformatics.”

Background of the project

The importance of this subject

World Health Organization | Health Topics | Countries | Newsroom | Emergencies

Cardiovascular diseases (CVDs)

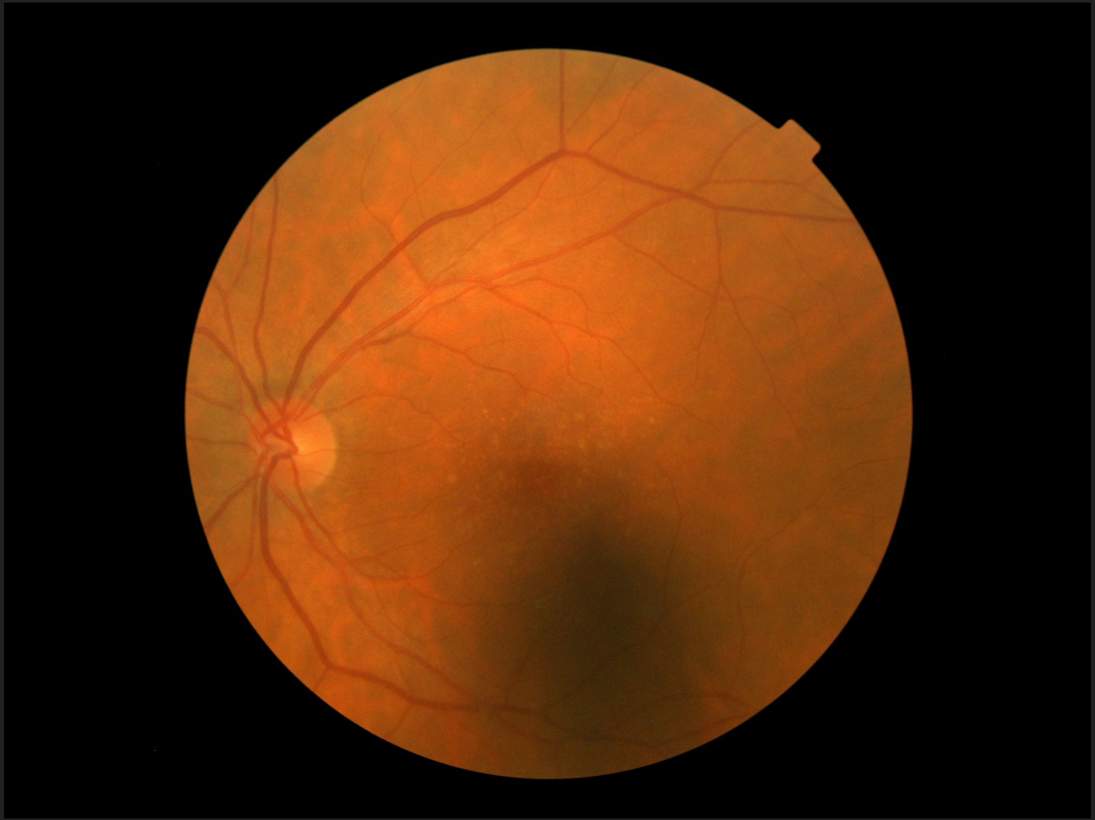
17 May 2017

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.

https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Why eye fundus ?

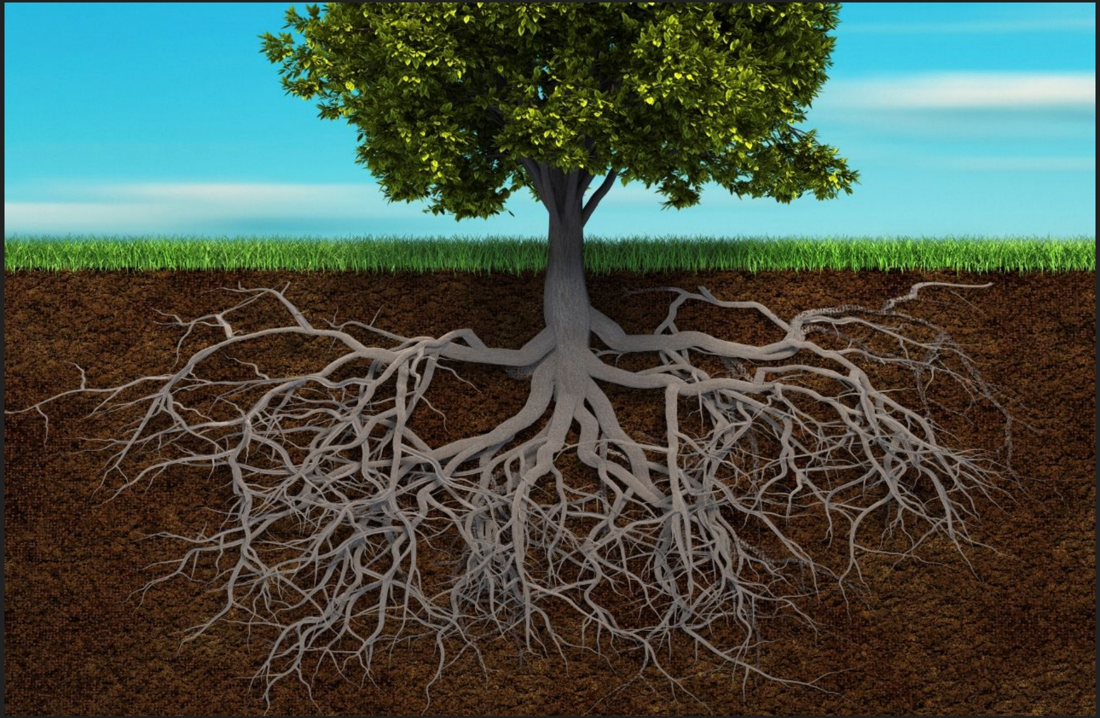


Our first eye fundus from the UK BioBank

What’s ARIA ?

ARIA

v1.0



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

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

What’s tortuosity and how was it calculated ?

We have two type of tortuosity :

- Distance Factor :

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

-

What software/programming languages did we used ?

https://fr.wikipedia.org/wiki/MATLAB

https://www.python.org/

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

What is the original dataset ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	822.23	822.5	822.75	822.99	823.2	823.4	823.59	823.76	823.93	824.1	824.26	824.42	824.58	824.74	824.91	825.08	825.26	825.46	825.66	825.89	826.13	826.39	826.65	826.93
2	737.15	737.61	738.08	738.55	739.02	739.5	739.99	740.47	740.97	741.46	741.96	742.46	742.97	743.48	743.99	744.5	745.01	745.53	746.05	746.57	747.09	747.6	748.12	748.62
3	619.6	620.48	621.33	622.15	622.96	623.75	624.53	625.3	626.06	626.82	627.58	628.34	629.11	629.89	630.68	631.48	632.3	633.12	633.96	634.8	635.64	636.48	637.33	638.17
4	973.74	972.85	971.96	971.08	970.19	969.31	968.43	967.55	966.66	965.78	964.89	964.01	963.12	962.24	961.36	960.48	959.6	958.73	957.86	956.99	956.13	955.27	954.4	953.54
5	540.71	541.62	542.53	543.42	544.31	545.18	546.04	546.9	547.74	548.58	549.4													
6	553.01	554.02	555.02	556	556.98	557.95	558.91	559.86	560.81	561.75	562.69	563.63	564.57	565.51	566.45	567.4	568.35	569.31	570.28	571.25	572.23	573.22	574.21	575.21
7	550.43	549.69	548.94	548.16	547.38	546.57	545.76	544.93	544.09	543.24	542.38	541.51	540.64	539.76	538.87	537.98	537.08	536.18	535.28	534.38	533.48	532.57	531.67	530.77
8	709.63	710.98	712.07	712.92	713.57	714.03	714.33	714.48	714.52	714.46	714.33	714.16	713.96	713.76	713.58	713.45	713.39	713.42	713.57	713.86	714.31	714.92	715.68	716.55
9	669.99	670.89	671.78	672.66	673.54	674.41	675.27	676.13	676.99	677.85	678.72	679.58	680.44	681.31	682.19	683.07	683.96	684.86	685.77	686.69	687.62	688.55	689.5	690.44
10	1083.4	1082.7	1081.9	1081.2	1080.4	1079.5	1078.7	1077.9	1077	1076.1	1075.2	1074.4	1073.4	1072.5	1071.6	1070.7	1069.8	1068.9	1068	1067.1	1066.2	1065.3	1064.4	1063.5
11	771.94	772.57	773.23	773.92	774.64	775.39	776.17	776.97	777.79	778.62	779.47	780.34	781.22	782.1	782.99	783.88	784.77	785.67	786.55	787.44	788.31	789.19	790.06	790.93
12	900.3	899.76	899.24	898.75	898.28	897.83	897.41	897.01																
13	895.05	894.19	893.32	892.43	891.51	890.58	889.64	888.69	887.72	886.75	885.76	884.78	883.79	882.8	881.81	880.82	879.84	878.87	877.9	876.94	876	875.06	874.13	873.21
14	358.56	359.52	360.48	361.45	362.42	363.39	364.37	365.34	366.32	367.29	368.27	369.24	370.22	371.19	372.16	373.13	374.1	375.07	376.04	377.01	377.98	378.96	379.93	380.91
15	819.06	818.17	817.3	816.43	815.58	814.73	813.9	813.08	812.27	811.47	810.68	809.9	809.14	808.38	807.64	806.9	806.18	805.47	804.77	804.07	803.4	802.73		
16	852.38	852.8	853.14	853.41	853.62	853.78	853.9	854	854.07	854.14	854.2	854.28	854.39	854.53	854.71	854.94	855.24	855.61	856.07	856.63	857.27	857.99	858.78	859.62
17	911.05	912	912.93	913.85	914.76	915.67	916.56	917.45	918.34	919.22	920.09	920.96	921.83	922.69	923.56	924.42	925.28	926.15	927.01	927.88	928.75	929.62	930.49	931.35
18	805.67	806.33	807.02	807.75	808.51	809.31	810.13	810.97	811.83	812.72	813.61	814.52	815.43	816.35	817.27	818.18	819.09	819.99	820.88	821.75	822.61	823.47	824.32	825.17
19	768.02	767.14	766.26	765.37	764.49	763.6	762.71	761.81	760.92	760.02	759.11	758.21	757.3	756.39	755.47	754.55	753.63	752.7	751.76	750.83	749.89	748.94	747.99	747.04
20	854.11	854.97	855.82	856.68	857.55	858.41	859.29	860.16	861.05	861.94	862.83	863.73	864.64	865.55	866.47	867.4	868.33	869.27	870.22	871.17	872.14	873.1	874.07	875.04
21	776.94	776.14	775.33	774.51	773.68	772.84	772	771.14	770.28	769.42	768.55	767.68	766.8	765.93	765.05	764.17	763.29	762.42	761.55	760.68	759.81	758.94	758.07	757.2
22	788.29	788.1	787.86	787.58	787.26	786.92	786.54	786.13	785.7	785.25	784.79	784.31	783.81	783.32	782.81	782.31	781.81	781.32	780.83	780.36	779.89	779.44	779	778.56
23	1275	1274.2	1273.3	1272.4	1271.5	1270.6	1269.7	1268.8	1267.9	1267	1266	1265.1	1264.2	1263.3	1262.3	1261.4	1260.5	1259.6	1258.7	1257.8	1256.9	1256	1255.2	1254.3
24	1006.1	1007	1008	1008.9	1009.8	1010.8	1011.7	1012.6	1013.6	1014.5	1015.4	1016.3	1017.2	1018.1	1019	1019.9	1020.8	1021.7	1022.6	1023.5	1024.3	1025.2	1026.1	1027
25	907.22	906.21	905.21	904.23	903.26	902.3	901.36	900.43	899.52	898.62	897.74	896.87	896.02	895.17	894.35	893.54								
26	1004.3	1003.9	1003.5	1003.2	1002.9	1002.6	1002.3	1002	1001.8	1001.6	1001.4	1001.3	1001.1	1001	1000.9									
27	1249.8	1250.2	1250.7	1251.1	1251.5	1252	1252.4	1252.8	1253.2	1253.6	1254	1254.3	1254.7	1255.1	1255.4	1255.7	1256	1256.3	1256.6	1256.8	1257	1257.3	1257.5	1257.7
28	791.12	791.8	792.47	793.12	793.78	794.42	795.07	795.71	796.36	797.02	797.68	798.36	799.04	799.75	800.47	801.22	801.99	802.78	803.6	804.45	805.31	806.19	807.08	807.99
29	890.97	890.05	889.12	888.2	887.28	886.36	885.45	884.53	883.61	882.7	881.78	880.87	879.95	879.03	878.11	877.19	876.26	875.34	874.4	873.47	872.53	871.59	870.64	869.7

Preliminary Results

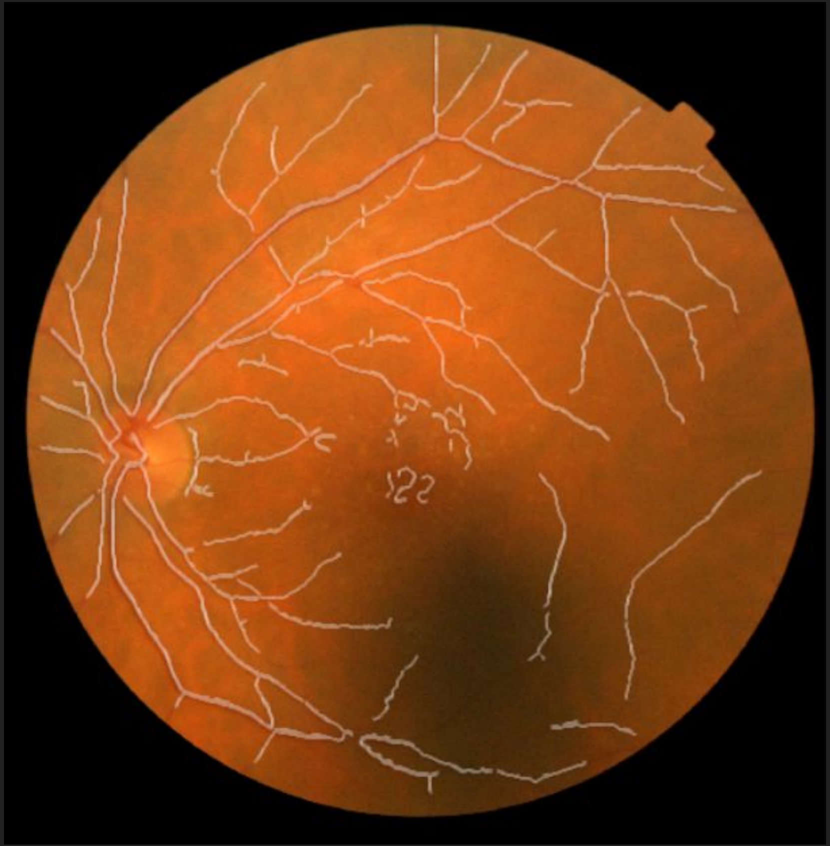
Plot of the blood vessels taking into account their diameter

We decided to use python with the following main packages :

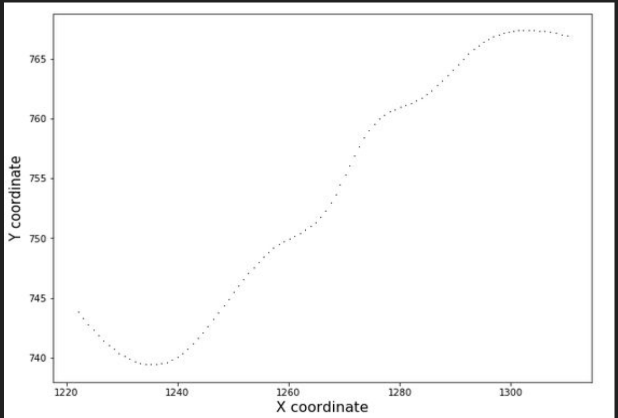
numpy

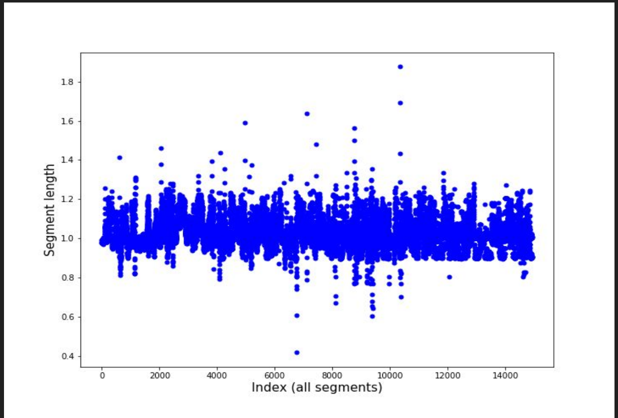
matplotlib

However, we did not manage to perfectly fit the diameter of points to the vessels on the image



Observation of the distance between each point of a vessel

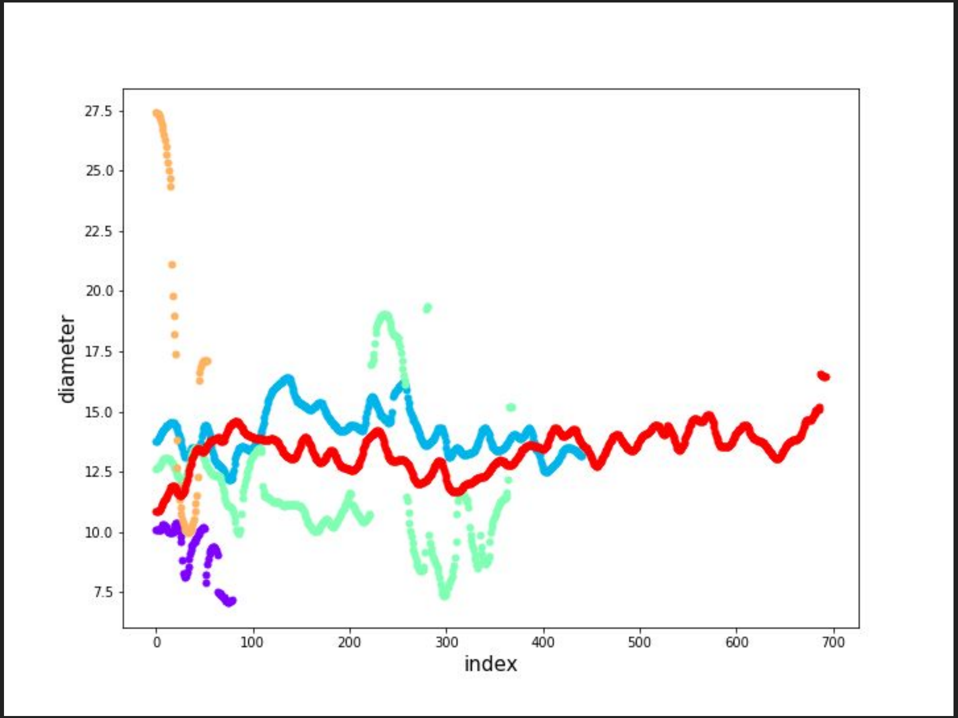




The vessels to be seems to be continuous, without big gaps between points.

Therefore, we decided that euclidean distance would be used to measure distances.

Observation of the diameter of vessels



The visualisation of the diameter showed some big gaps along some segments (green points)

Whereas, some segments seemed to have continuous diameter values

We wondered whether these gaps were visible on the Image

Observation of gaps in the diameter of vessels

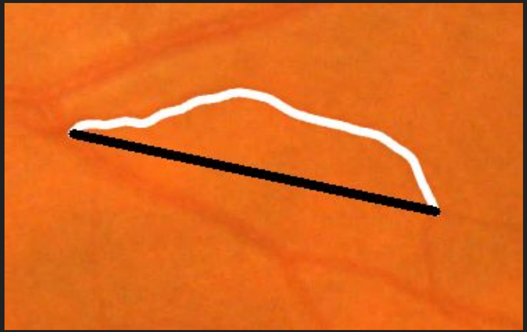
We mathematically extracted these gaps and plotted them on the image.

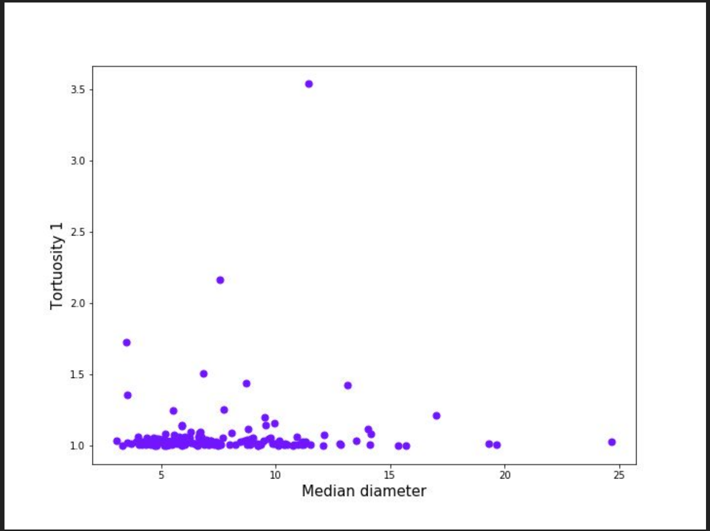
We didn't see any visual correlation between these gaps and a sudden change in diameter of a subpart of a vessel.

Therefore, we decided to ignore these values and to use the median of diameters to represent this feature.

Tortuosity measurement implementation : 1st method (Distance Factor)

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

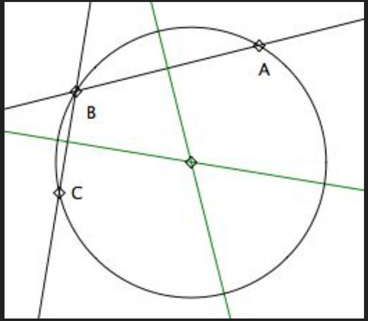




We can notice some very tortuous vessels, but most of the other vessels have a tortuosity close to 1.

It seems that there is no correlation between this tortuosity and the diameter of a vessel.

Tortuosity measurement implementation : 2^{nd} method

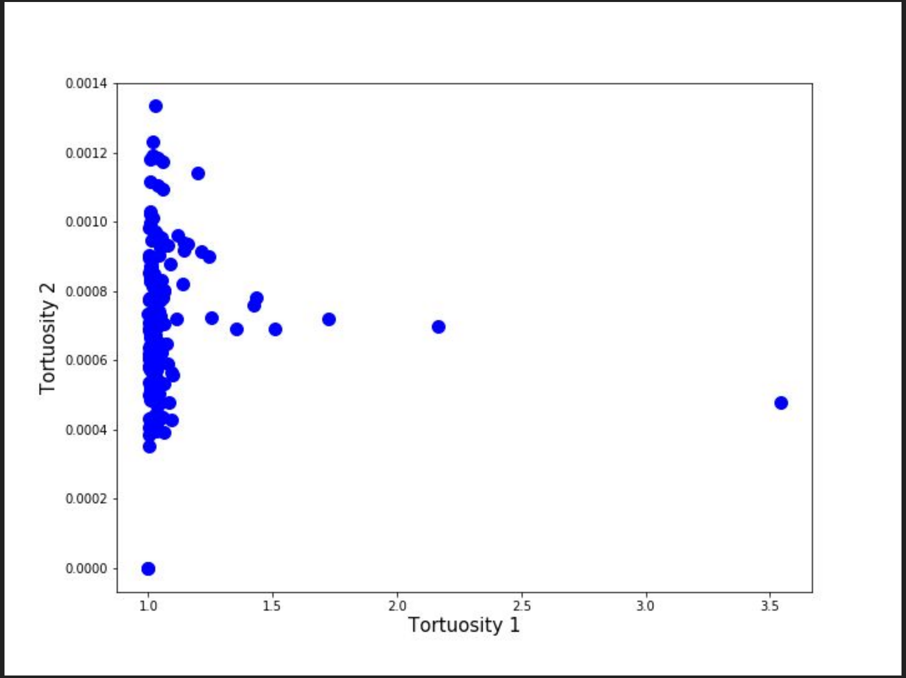


Algorithm:

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

Compute and sum 1/radius

Divide by the length of the vessel

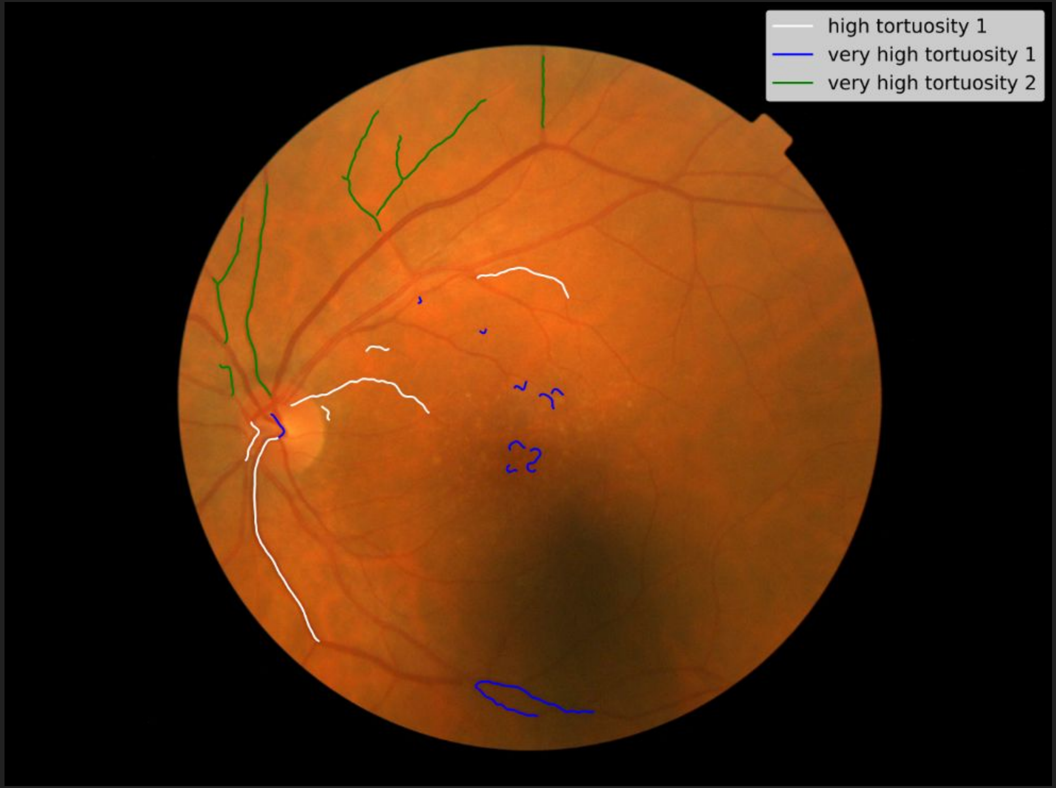


Visualisation of highly tortuous vessels

Most tortuous vessel with the 1st method seem to highlight artifacts produced by ARIA software

Highly tortuous vessels (1st method) seem to be reflect tortuous vessels

Most tortuous vessels obtained with the 2nd method don't seem to be very tortuous, maybe because of many small curves



Calculation of Distance Factors for 50 eye fundus

Algorithm :

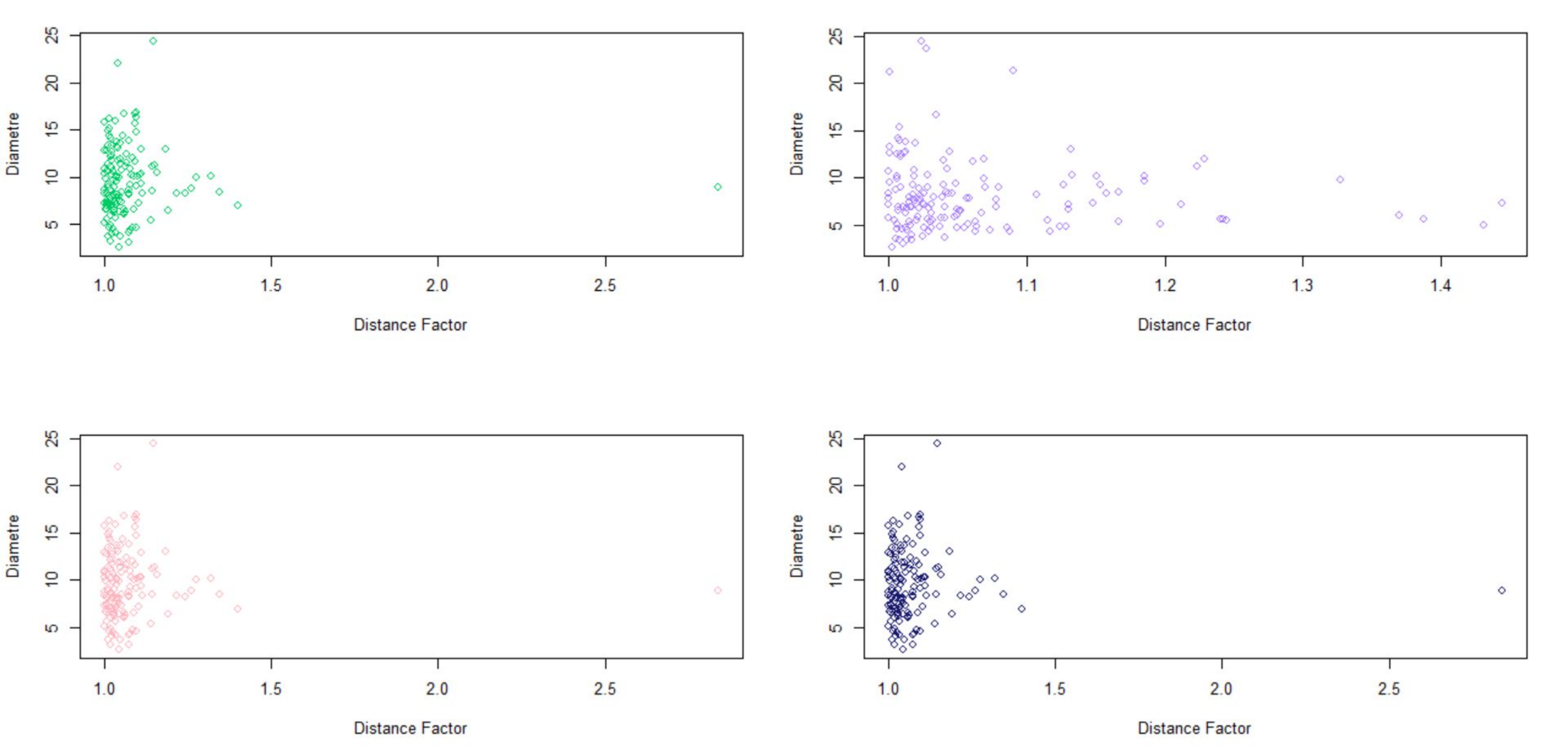
- Bash command : `ls > preprocess.txt`
- Preprocessing : identify pair of eyes or single eyes *(this step enables easy parallelisation)*
- Calculated the Distance Factor and the median diameter for each vessel
- Extract the median difference of tortuosity and diameter between pair of eyes
- Write the output in a well formatted file

Basic dataset for statistical work

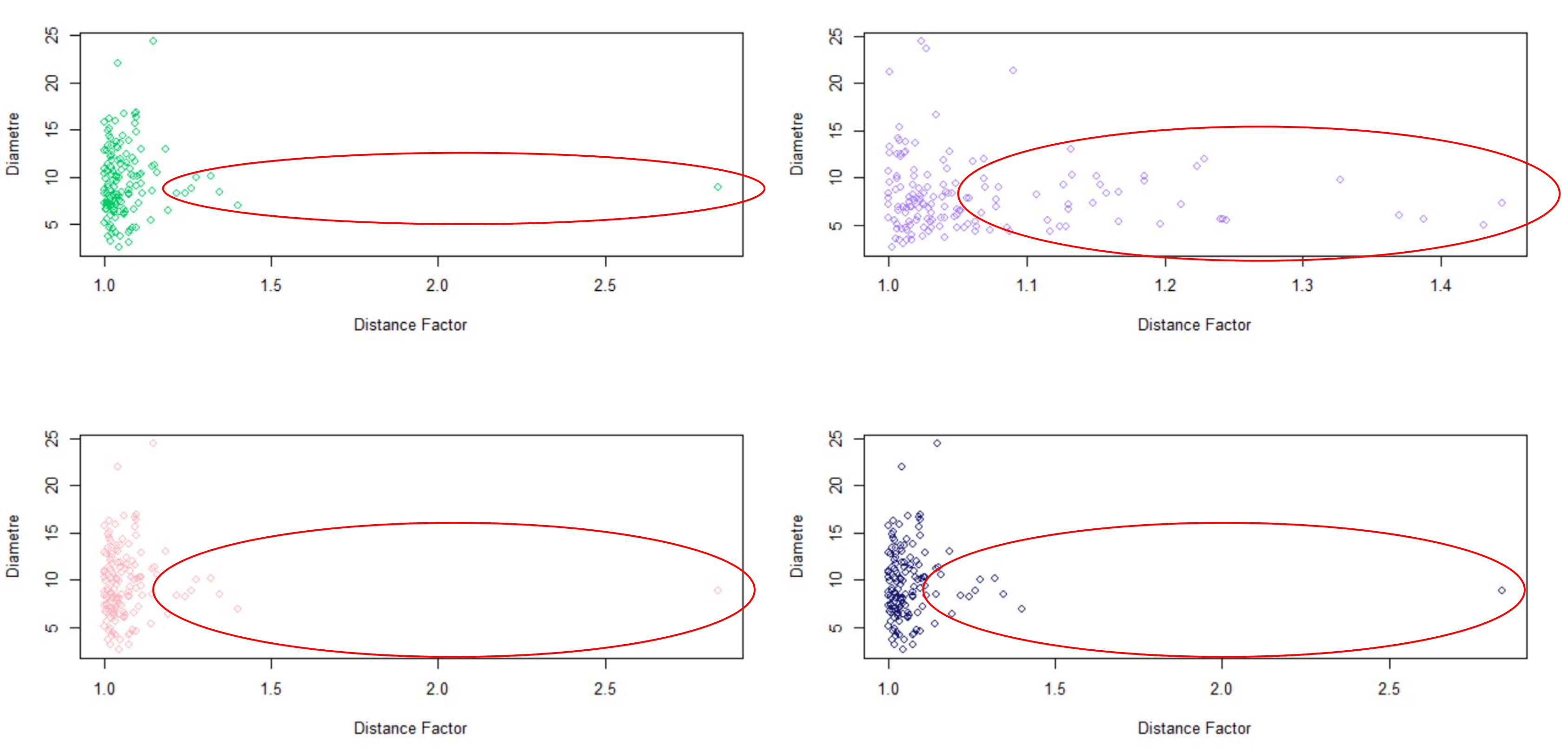
	ID	Side															
1	>1019547	r	1.006299	1.021242	1.142341	1.261676	1.009143	1.079102	1.029170	1.047745	2.838191	1.010656	1.029861	1.041200	1.039319	1.031920	1.059805
2	>1019547	r	6.911900	13.356500	8.500700	8.834800	10.620000	10.878000	8.956800	11.344000	8.881900	11.328000	6.865300	11.842000	22.049000	15.929500	16.746000
3	>1018862	l	1.185534	1.078195	1.004660	1.147832	1.153889	1.090755	1.061292	1.056791	1.026316	1.011266	1.001292	1.017183	1.049949	1.240660	1.035205
4	>1018862	l	9.584000	6.946800	5.464900	7.247500	9.213000	21.300000	11.659000	7.861650	7.192500	2.975300	21.173000	6.951700	6.628400	5.576800	16.671500
5	>1018862	r	1.026515	1.010779	1.052191	1.195384	1.056794	1.059250	1.002805	1.000234	1.050733	1.015383	1.027681	1.028471	1.017299	1.077176	1.108564
6	>1018862	r	4.858200	5.414200	5.366800	4.690150	4.485500	5.015950	3.717700	5.328050	8.820900	9.773250	6.285500	5.381000	6.961800	4.850600	10.886000
7	>1020524	l	1.021254	1.020257	1.029458	1.093540	1.020476	1.151714	1.017795	1.000000	1.180976	1.067312	1.008310	1.099414	1.007082	1.340435	1.025845
8	>1020524	l	7.664400	9.290400	10.766000	8.529950	11.359500	7.892000	3.719200	2.007050	10.240500	8.145700	5.558900	14.813500	11.491000	3.214500	4.873500
9	>1020524	r	1.178633	1.038090	1.020981	1.000529	1.087153	1.034402	1.130432	1.011414	1.017042	1.028285	1.192193	1.004737	1.055921	1.008089	1.008345
10	>1020524	r	9.826600	5.761900	6.234400	11.157500	6.138550	5.669700	4.982800	9.609800	10.820000	5.163950	6.475900	3.875300	3.782250	6.544600	4.311700
11	>1018843	l	1.023258	1.053818	1.013497	1.017013	1.011421	1.008219	1.001477	1.006291	1.000848	1.032527	1.160356	1.030133	1.023504	1.023684	1.005872
12	>1018843	l	6.664300	7.180500	10.031500	10.101000	4.123350	4.771300	5.222200	10.508000	9.671000	4.825300	20.496000	5.618400	21.928000	8.429750	6.120350
13	>1018843	r	1.014311	1.019705	1.017277	1.064496	1.317888	1.026727	1.063519	1.000733	1.012324	1.010156	1.021020	1.025619	1.007111	1.030620	1.016689

Showing 1 to 13 of 94 entries, 222 total columns

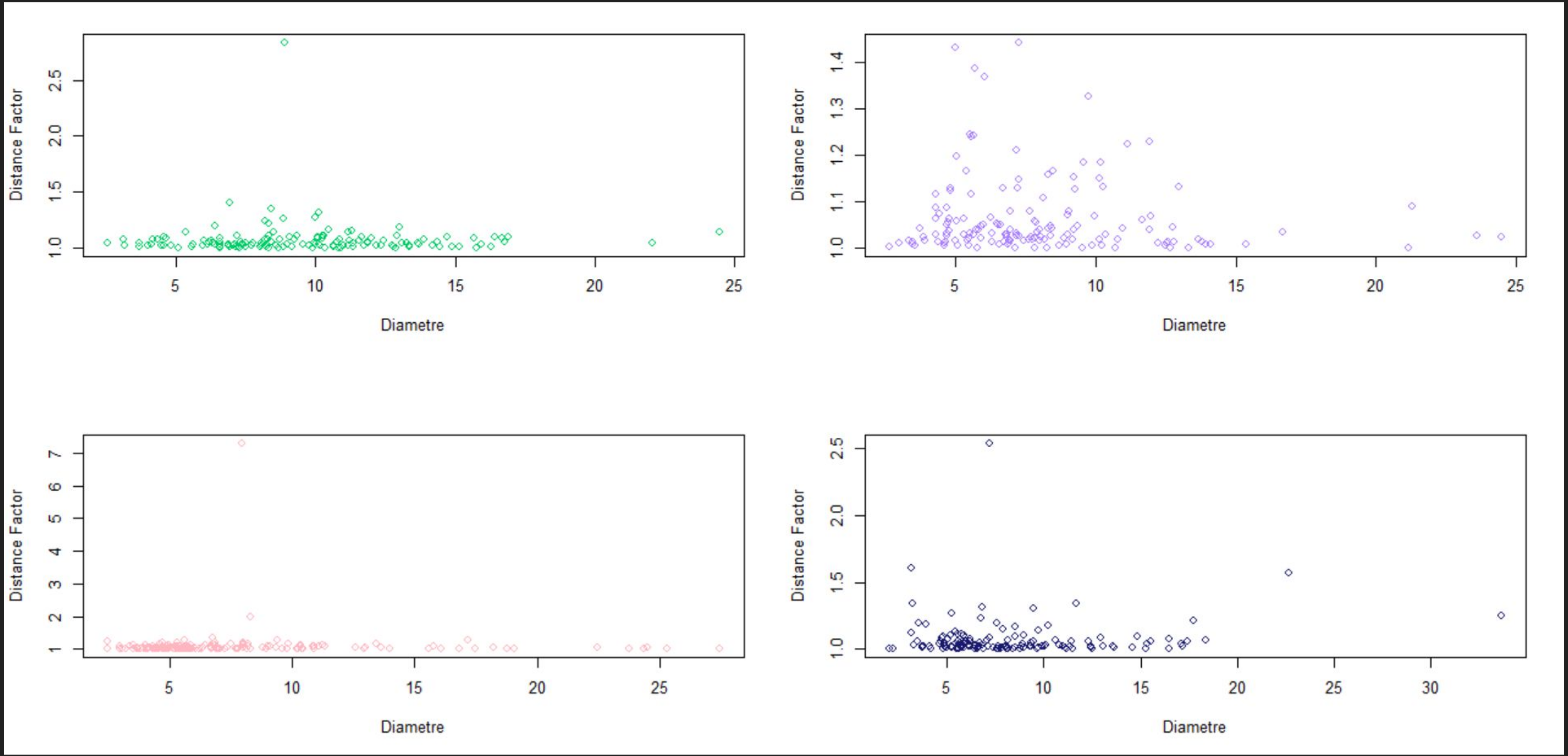
Using R for statistical analysis : plot 1



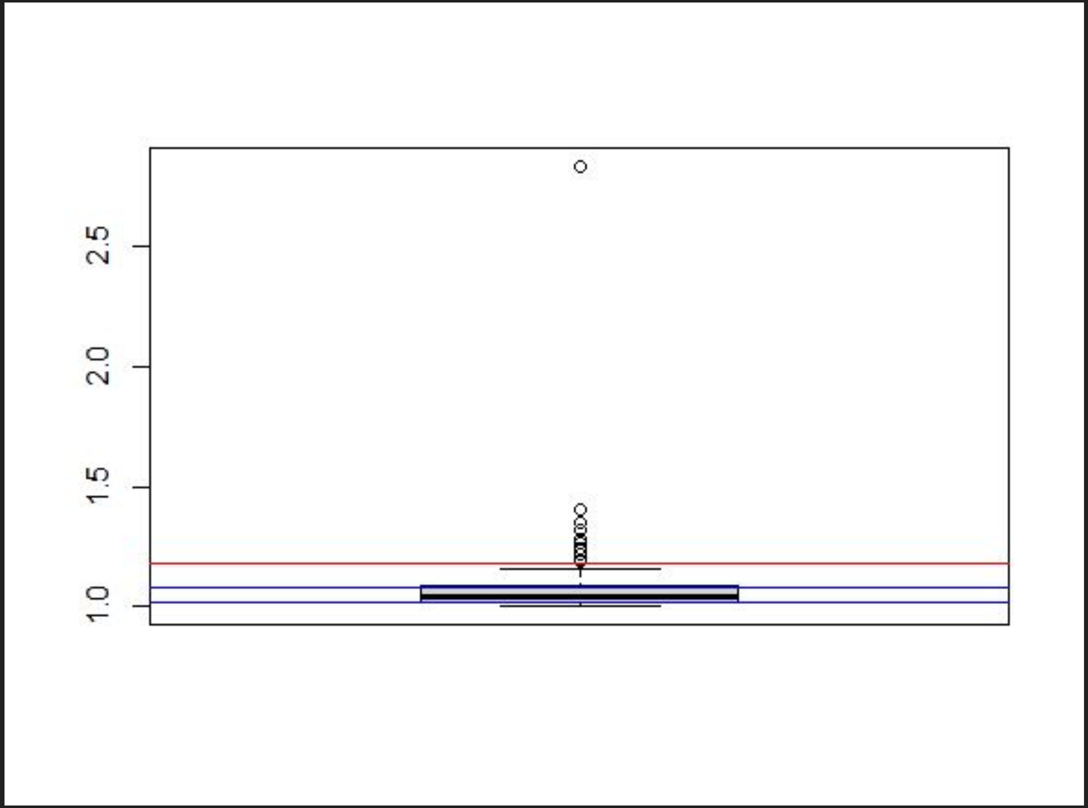
Using R for statistical analysis : plot 1



Using R for statistical analysis : plot 2



Using R for statistical analysis : boxplot for the first eye fundus

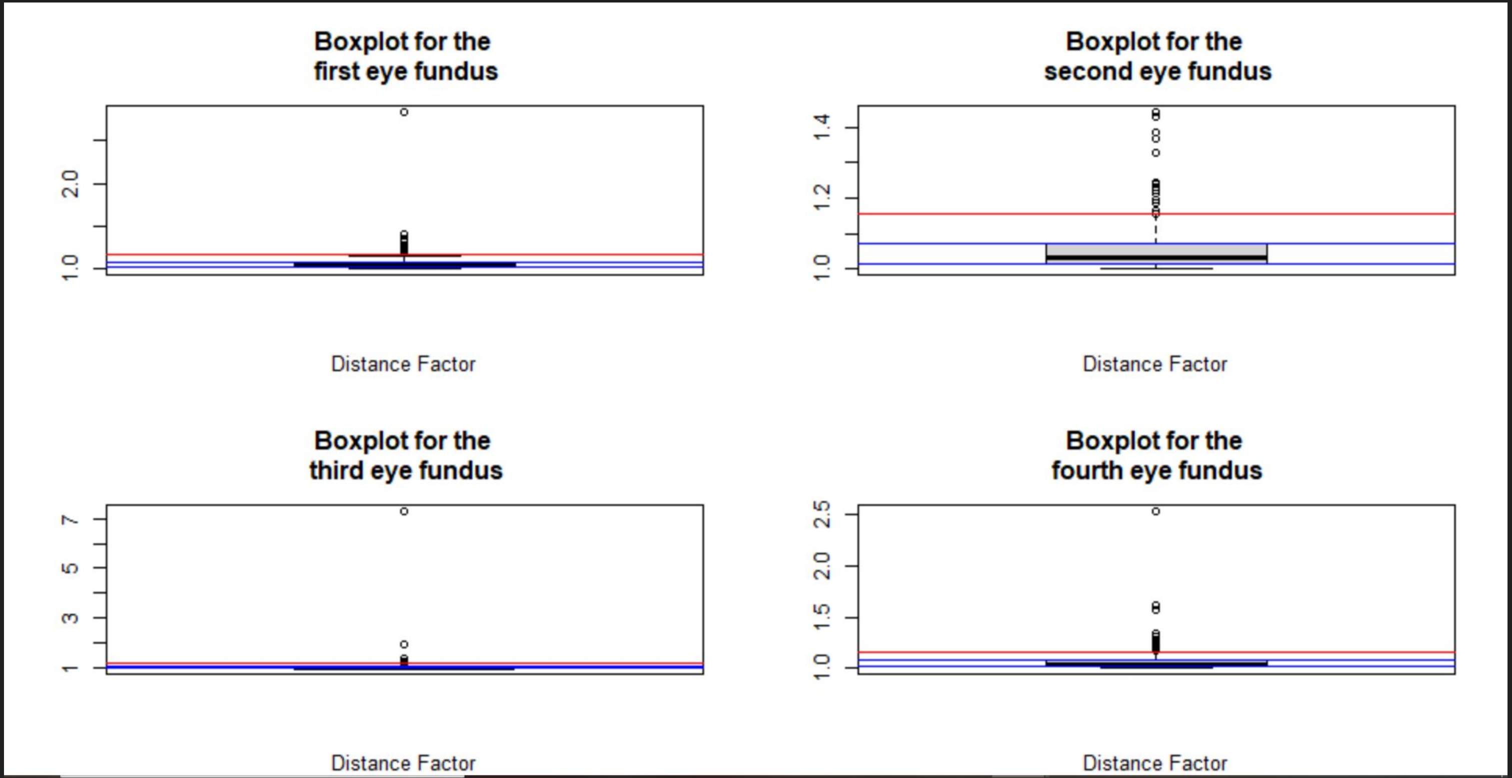


$$\textit{Removing the outliers} = Q3 + 1{,}5 \cdot (Q3 - Q1)$$

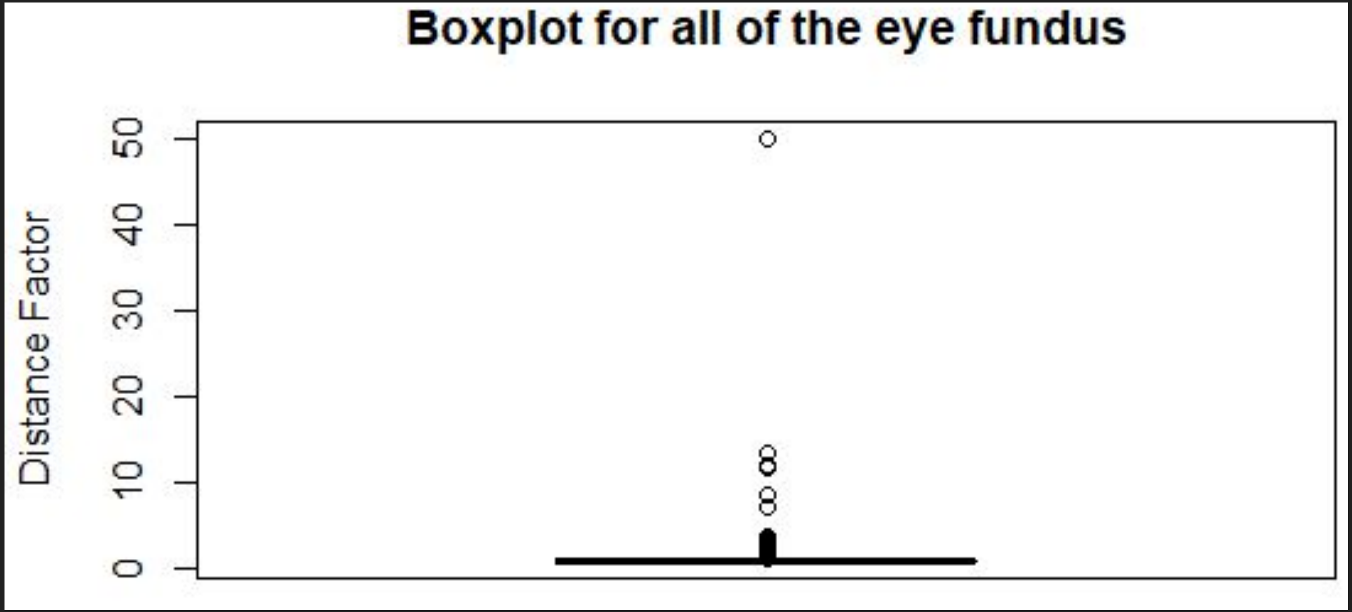
The data is very crushed and the outliers cause a complete unreadability of our results.

```
> summary(DFwork$V1, na.omit=TRUE)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   1.019   1.039   1.074   1.083   2.838 
```

Using R for statistical analysis : boxplot for the 4th first eye fundus

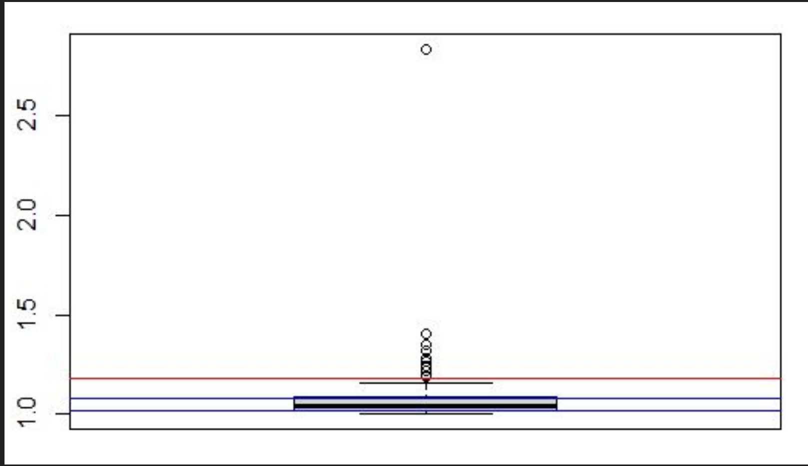


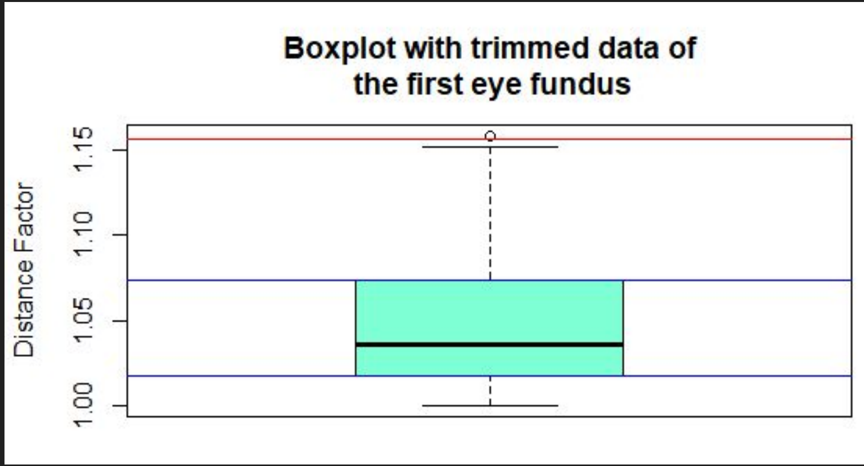
Boxplot for all of the values (DF)




```
> summary(vectot, na.omit=TRUE)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.012   1.026   1.069   1.060  50.020
```

Removal of outliers of the first fundus

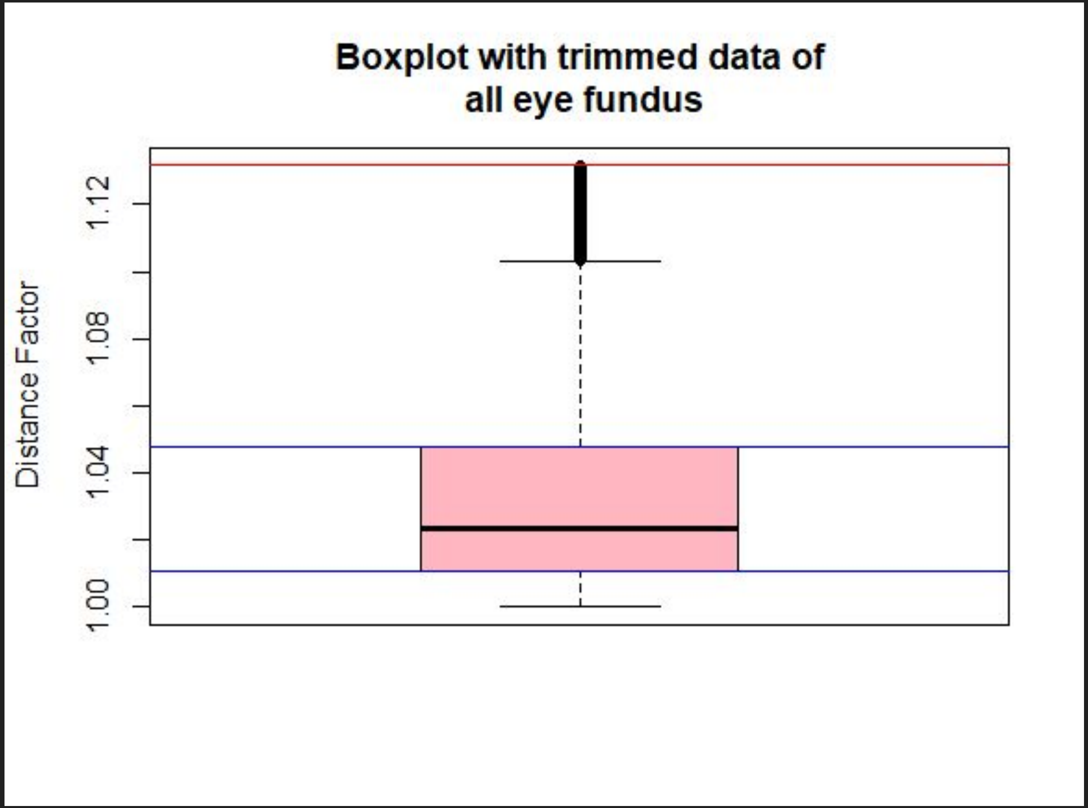




$$\textit{Removing the outliers} = Q3 + 1{,}5 \cdot (Q3 - Q1)$$

```
> summary(vf)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   1.018   1.036   1.046   1.073   1.158 
```

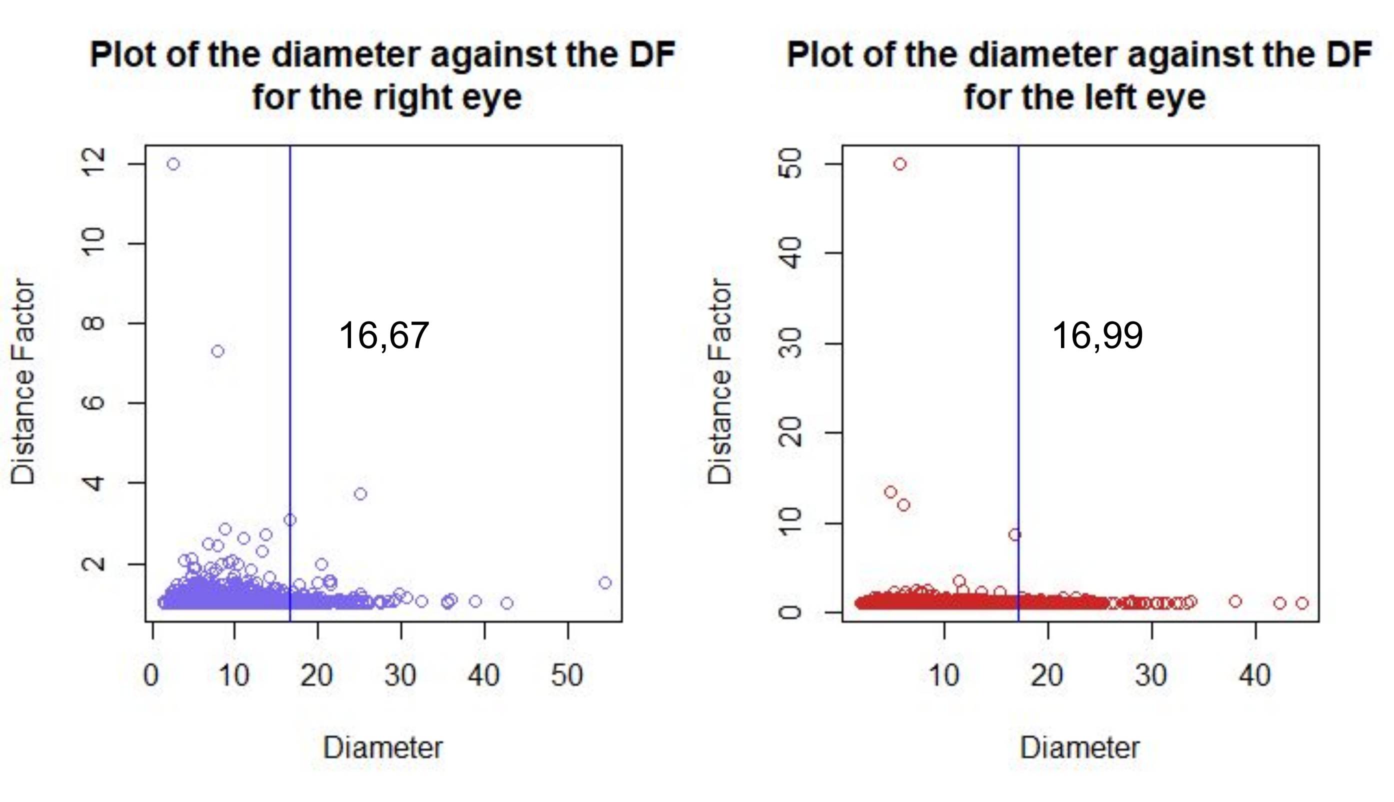
We removed the outliers for the 47 others




```
> lvt
[1] 1.131727
```

```
> length(vectot)
[1] 8239
> length (vf)
[1] 7493
```

Remark for the left and right eyes : are they different ?



Are they now significantly different ?

We used a Welch t-test with non paired data because the sample did not have the same size.

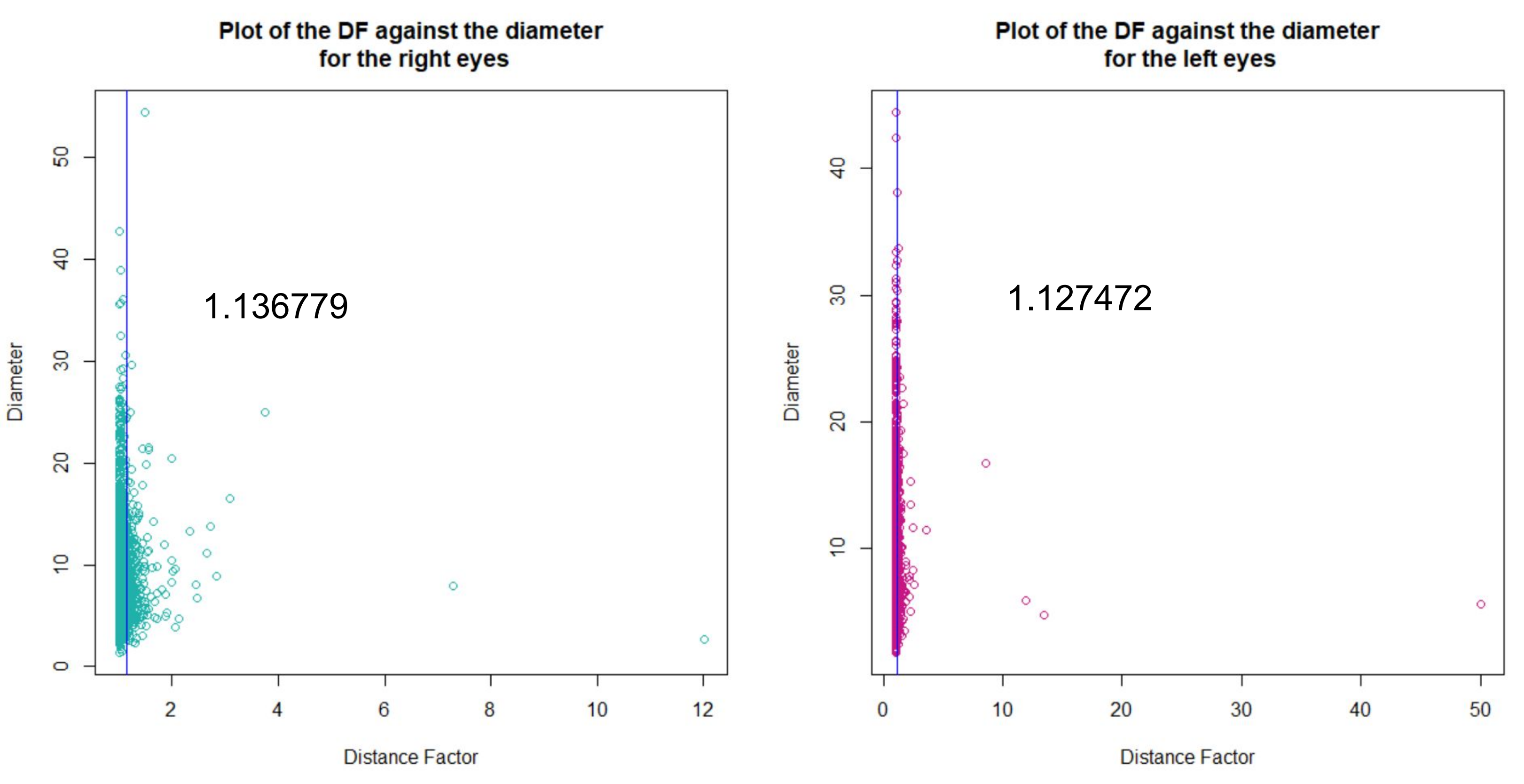
Hypothesis : *“There are no differences between the right and the left eye”*

```
        Welch Two Sample t-test

data:  DDIA and GDIA
t = -2.4505, df = 8120.6, p-value = 0.01429
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.40322320 -0.04481512
sample estimates:
mean of x mean of y
 8.177839  8.401858
```

→ They are significantly different in term of diameter

And for the Distance Factor ?



Are they now significantly different ?

We used a Welch t-test with non paired data because the sample did not have the same size.

Hypothesis : *“There are no differences between the right and the left eye”*

```
        Welch Two Sample t-test

data:  DDF and GDF
t = -0.97432, df = 4501.6, p-value = 0.33
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.04026799  0.01353110
sample estimates:
mean of x mean of y 
 1.062223  1.075591 
```

→ They are not significantly different in term of distance factor

Hypothesis

- If we take the ratio between arteriolar and venular widths as an index of generalized arteriolar narrowing, will this be linked to the presence of CVDs ?
- We know that we can link narrower retinal arterioles and wider retinal venules to long-term risk of mortality and ischemic stroke in both sexes, but how much is it predictive in a specific sex ?
- Is this biomarker reproductible for both man and women and is it sensitive ?
- On a totally different level, knowing that the pressure inside the arterioles, does that mean that they are less tortuous ?
- If we have a high tortuosity in a vein/ an artery, what does that mean ?
- Is the length or other features of the blood vessels linked to something specific ? Sex ? Wearing glasses ? Age ? CVDs ?

What needs to be done ?

Graphically

- Mapping each outlier segment in order to know if they are relevant and find a threshold above which we can exclude values
- Apply all updates to the Wiki

Mathematically

- Seeing if the right and the left eye are significantly different for a given feature
- Linking diseases with vessels tortuosity/diameter/length phenotypes
- If we don’t find significant results we could add another feature which is the classification of vessels into Veins or Artery

Maybe try to implement a simple machine learning method to our diseases/features…

Questions ?