Health Index reconstruction

A.J.Nicholson

Introduction

The purpose of this analysis is to create an approximation of the health Index calculation in code. With the view of creating a unifying function that can be applied to any dataset with the required elements. The reconstruction of the health index analysis process will be based upon the description given in chapter 3 of The backbone of history: health and nutrition in the western hemisphere. (Steckel & Rose, 2002).

The data used in this document is the output from the end of the "CorrectData.rmd".

```
cols<-read.table("data/cols.txt",colClasses="character");cols<-as.character(cols[2,])
HIdata<-read.table("data/HIdata_c.txt",header=TRUE,colClasses=cols)
```

Sampling

I'm going to develop the code using a single population and then apply it back to the whole dataset.

```
kable(t(summary(HIdata$Site)[1:10]))# top 10 most populated
```

41D	W60	stt	ESB	CHB	301	WVO	HPK	WR0	214
1157	796	604	426	361	335	307	296	278	267

I need a population with a large number of observations to provide the best trouble shooting. So I will choose site 41D as my sample population. This sample comes from a cemetery in Dallas dating to 62 BP. It should have 1157 individuals 613 of whom are over 5 and the ancestry should be African American (3).

```
PopA<-HIdata[HIdata$Site=="41D",]
kable(data.frame(no=length(PopA$ID),noPlus5=(length(PopA$ID[PopA$Age>5]))))
```

no	noPlus5
1157	717

```
kable(t(summary(PopA$Ancestry)))
```

1	2	3	5	6
0	0	1157	0	0

The ancestry and total number match what is reported Steckel & Rose but there are more individuals over 5 than reported. This will not stop me using this population but it is interesting to note the difference.

Approach

The health Index is created in a multi-stage process. That I shall break into 3 sections:

1. Converting the recorded variables for each individual into scores. "for each individual, every observable attribute of the health index is scored on a scale of 0 to 100 percent"

2. Create Age specific scores for each attribute. This is an average of individuals scores, broken into age categories. For some attributes individuals contribute score for all age categories during their life; i.e. if they die at 30 and had a score of 50, they would contribute this same score for every age category up to and including the one containing 30. After this they contribute a score of zero. Other attributes only effect 10 years prior to death, before this period the individual would contribute a score of 100.

3. Combining Age specific scores to produce a single index score for each attribute that is averaged to get the health index.

To aid in this process I'm going to create empty data frames that will be filled during each section.

```
D1<-data.frame(Individual=PopA$ID,Age=NA,B2D=NA,Ten=NA,Stature=NA,Hypoplasia=NA,
               Anemia=NA,DentalHealth=NA,Infections=NA,DJD=NA,Trauma=NA)

kable(D1[1:10,])# first 10 lines
```

Individual	Age	B2D	Ten	Stature	Hypoplasia	Anemia	DentalHealth	Infections	DJD	Trauma
0001	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0002	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0003	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0004	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0005	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0006	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0007	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0008	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0009	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0010	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

```
D2<-data.frame(Trait=c("Stature","Hypoplasia","Anemia","DentalHealth",
                       "Infections","DJD","Trauma"),
               contributors=NA,Age0_4=NA,Age5_14=NA,Age15_24=NA,Age25_34=NA,
               Age35_44=NA,Age45=NA)

kable(D2)
```

Trait	contributors	Age0_4	Age5_14	Age15_24	Age25_34	Age35_44	Age45
Stature	NA	NA	NA	NA	NA	NA	NA
Hypoplasia	NA	NA	NA	NA	NA	NA	NA
Anemia	NA	NA	NA	NA	NA	NA	NA
DentalHealth	NA	NA	NA	NA	NA	NA	NA
Infections	NA	NA	NA	NA	NA	NA	NA
DJD	NA	NA	NA	NA	NA	NA	NA
Trauma	NA	NA	NA	NA	NA	NA	NA

```
HI<-data.frame(Pop=PopA$Site[1],HI=NA,QALY=NA,lifeEx=NA,pyl=NA,
               Stature=NA,Hypoplasia=NA,Anemia=NA,Dental=NA,Infection=NA,DJD=NA,Trauma=NA)

kable(HI)
```

Pop	HI	QALY	lifeEx	pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
41D	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Section 1 Individual scores

There are 7 potential attribute scores for each Individual:

1. Stature

2. Hypoplasia

3. Anemia

4. Dental Health

5. Infection

6. DJD

7. Trauma

Age

Due to Known issues in the accuracy of ages calculated from skeletal remains the reported ages are corrected according to the method outlined in Lovejoy et al (1985). This can be applied to the data set using the function below

```r
LovejoyAdj<-function(age,info=FALSE){# age= numeric, info=logical
  #info function provides details of the method used and reference
  #default value= FALSE
  if (info==TRUE){
    output<-"Age at death estimations from skeletal remains have been show to be inaccurate.
    Specifically there is a tendency to underestimate, particularly in older individuals.
    This function adjusts reported age using the method outlined in lovejoy et al. (1985).
    The function takes a single numeric input. To apply to dataframe the following steps are required:
    Dataframe$Adjusted<-NA
    for(i in 1:length(Dataframe$Age)){Dataframe$Adjusted[i]<-LovejoyAdj(Dataframe$Age[i])}"
    output<-strwrap(output)
  }else{
    if(is.na(age)){
      output<-NA
    }else{
      if(age>=18 & age<30){
        output<-age-1.4
      }else{
        if(age>=30 & age<40) {
          output<-age+0.8
        }else{
          if(age>=40 & age<50){
            output<-age+7.4
          }else{
            if(age>=50 & age<60){
              output<-age+6.8
            }else{
              if(age>=60){
                output<-age+13.6
              }else{output<-age}}}}}}
  output}
```

```r
for(i in 1:length(D1$Age)){
  D1$Age[D1$Individual==PopA$ID[i]]<-LovejoyAdj(PopA$Age[i])
}
```

From these ages I need to calculate what period constitutes birth to death(B2D) and what is 10 year prior to death(Ten). These are the two time periods that the attributes are are considered to act over. The age groups used by Steckle and Rose are "0-4","5-14","15-24","25-34","35-44"and "45+".

B2D will record the last period (i.e. the last time that individual will contribute to the heath index)

```
for(i in 1:length(D1$Age)){
  if(round(D1$Age[i])%in% 0:4){D1$B2D[i]<-1}
  if(round(D1$Age[i])%in% 5:14){D1$B2D[i]<-2}
  if(round(D1$Age[i])%in% 15:24){D1$B2D[i]<-3}
  if(round(D1$Age[i])%in% 25:34){D1$B2D[i]<-4}
  if(round(D1$Age[i])%in% 35:44){D1$B2D[i]<-5}
  if(round(D1$Age[i])>= 45){D1$B2D[i]<-6}
}
```

A Ten year period before death will either cover a one or two categories. So this is going to be a two digit number with lowest category coming first and 0 representing no category.

```
for(i in 1:length(D1$Age)){
  if((round(D1$Age[i])<=4)& (round(D1$Age[i]-10)<=4)){D1$Ten[i]<-01}
  if((round(D1$Age[i])%in% 5:14)& (round(D1$Age[i]-10)<=4)){D1$Ten[i]<-12}
  if((round(D1$Age[i])%in% 15:24)& (round(D1$Age[i]-10)%in% 5:14)){D1$Ten[i]<-23}
  if((round(D1$Age[i])%in% 25:34)& (round(D1$Age[i]-10)%in% 15:24)){D1$Ten[i]<-34}
  if((round(D1$Age[i])%in% 35:44)& (round(D1$Age[i]-10)%in% 25:34)){D1$Ten[i]<-45}
  if((round(D1$Age[i])>= 45)& (round(D1$Age[i]-10)%in% 35:44)){D1$Ten[i]<-56}
  if((round(D1$Age[i])>=45)& (round(D1$Age[i]-10)>=45)){D1$Ten[i]<-60}
}
```

```
kable(D1[1:10,])
```

Individual	Age	B2D	Ten	Stature	Hypoplasia	Anemia	DentalHealth	Infections	DJD	Trauma
0001	0.3	1	1	NA	NA	NA	NA	NA	NA	NA
0002	0.2	1	1	NA	NA	NA	NA	NA	NA	NA
0003	52.2	6	56	NA	NA	NA	NA	NA	NA	NA
0004	25.7	4	34	NA	NA	NA	NA	NA	NA	NA
0005	3.7	1	1	NA	NA	NA	NA	NA	NA	NA
0006	35.6	5	45	NA	NA	NA	NA	NA	NA	NA
0007	36.2	5	45	NA	NA	NA	NA	NA	NA	NA
0008	6.6	2	12	NA	NA	NA	NA	NA	NA	NA
0009	0.5	1	1	NA	NA	NA	NA	NA	NA	NA
0010	0.0	1	1	NA	NA	NA	NA	NA	NA	NA

Stature

In the original: 100 = attained or exceeded modern standard (Maresh, 1955), 0 if below 3 standard deviations and any where in-between is graded. I used Maresh(1970) values as reported in Schuer & Black, the standard deviation is estimated from given 10th and 90th percentiles (using z=(x-mean)/SD)

```
Maresh<-read.table("data/M70.txt",header=TRUE)
kable(Maresh)
```

age	ML	FL	AL	M3SD	F3SD	A3SD
0.125	86.0	87.2	86.60	70.43838	74.91451	72.67645
0.250	100.7	100.8	100.75	85.13838	88.86552	87.00195
0.500	112.2	111.1	111.65	96.17036	97.76147	96.96591

age	ML	FL	AL	M3SD	F3SD	A3SD
1.000	136.6	134.6	135.60	120.68736	121.02746	120.85741
1.500	155.4	153.9	154.65	135.86022	132.37114	134.11568
2.000	172.4	170.8	171.60	152.74321	148.68612	150.71466
2.500	187.2	185.2	186.20	165.55413	159.92699	162.74056
3.000	200.3	198.4	199.35	176.08003	170.43588	173.25796
3.500	212.1	211.1	211.60	181.91279	180.09376	181.00328
4.000	224.1	223.2	223.65	196.72090	189.61966	193.17028
4.500	235.7	235.5	235.60	206.09782	198.29251	202.19517
5.000	247.5	247.0	247.25	215.44072	214.00468	214.72270
6.000	269.7	268.9	269.30	232.60952	227.12933	229.86943
7.000	291.1	288.8	289.95	252.02044	247.84836	249.93440
8.000	312.1	309.8	310.95	268.22324	261.24306	264.73315
9.000	330.4	328.7	329.55	287.81030	274.05881	280.93455
10.000	349.3	347.9	348.60	301.32808	290.09969	295.71389
11.000	367.0	367.0	367.00	319.14509	292.81903	305.98206
12.000	386.1	387.6	386.85	332.62886	320.90733	326.76810
18.000	511.7	462.9	487.30	438.45507	377.95460	408.20484

The Maresh data provides separate male and female values, as sub-adult in this sample are not sexed and average value has also been included.

I want to check the suitability of this as a reference guide. Firstly for Sub Adult remains. For this I will explore using the average values or the range between the minimum possible values for a given age as reference values.

```
stature<-HIdata[,c("Age","Sex","FLEN","FDIAP")]


plot(Maresh$age,Maresh$AL,type="n",xlab="Age",ylab="femur diaphyseal length(mm)",
     main="subadult comparison")
points(stature$Age,stature$FDIAP,col="black",pch=20)
lines(Maresh$age,Maresh$AL,col="green",lwd=2)
lines(Maresh$age,Maresh$A3SD,col="red",lwd=2)

MinL<-rep(NA,length(Maresh$AL));Min3SD<-rep(NA,length(Maresh$AL))
for(i in 1:length(Maresh$age)){
  MinL[i]<-min(Maresh$ML[i],Maresh$FL[i])
  Min3SD[i]<-min(Maresh$M3SD[i],Maresh$F3SD[i])
  }
lines(Maresh$age,MinL,col="dark green",lwd=2)
lines(Maresh$age,Min3SD,col="dark red",lwd=2)
legend("topleft",c("100(average)","0(average)","100(min)","0(min)"),
       fill=c("green","red","dark green","dark red"),bg="white",title="Score")
```

subadult comparison
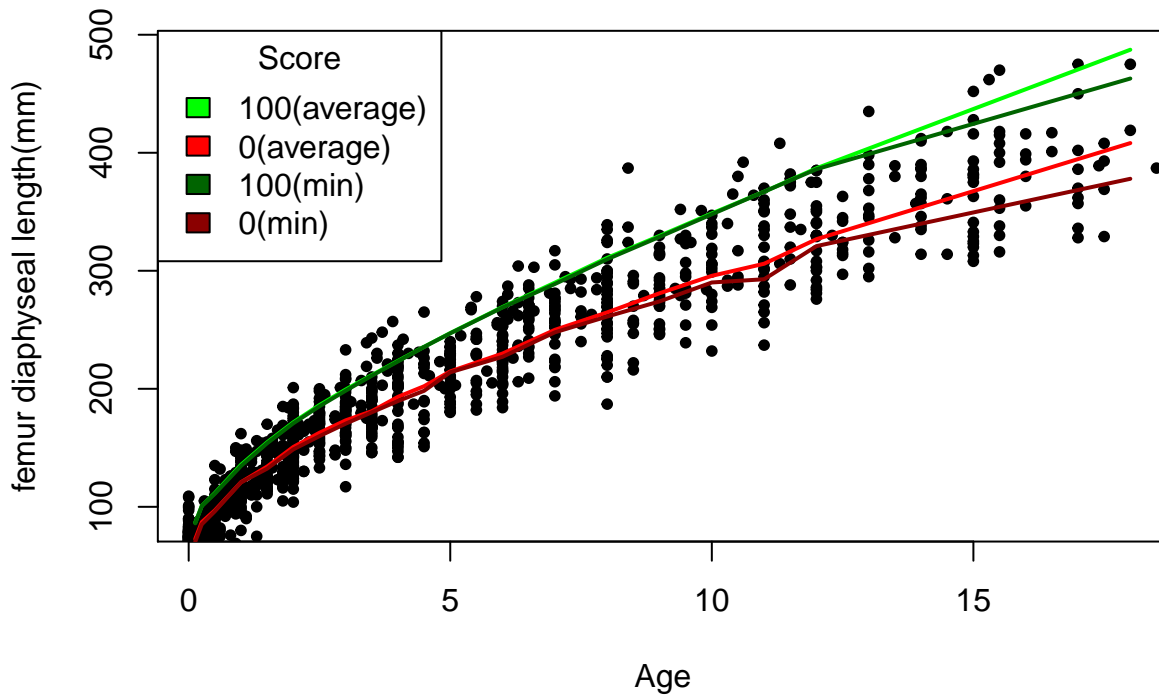
Anything above the green line would score 100 and anything below the red line would score 0. The difference between the two possible ranges is negligible for the younger individuals but increase as age does.

For adult remains the final row of the Maresh data frame is used.

```
kable(Maresh[20,])
```

|	age	ML	FL	AL	M3SD	F3SD	A3SD
20	18	511.7	462.9	487.3	438.4551	377.9546	408.2048

For these individual the different male and female values can be used for those of known sex. For remains of unknown sex I will again compare the effect of using the minimum and average values.

```
par(mfrow=c(1,3))
hist(stature$FLEN[stature$Sex==1|stature$Sex==2 ],main="Female",xlab="Femur length(mm)")
abline(v=Maresh$FL[20],col="green");abline(v=Maresh$F3SD[20],col="red")
hist(stature$FLEN[stature$Sex==3|stature$Sex==4 ],main="male",xlab="Femur length(mm)")
abline(v=Maresh$ML[20],col="green");abline(v=Maresh$M3SD[20],col="red")
hist(stature$FLEN[stature$Sex==6|is.na(stature$Sex) ],main="Unknown",xlab="Femur length(mm)")
abline(v=Maresh$AL[20],col="green");abline(v=Maresh$A3SD[20],col="red")
abline(v=Maresh$FL[20],col="dark green");abline(v=Maresh$F3SD[20],col="dark red")
```
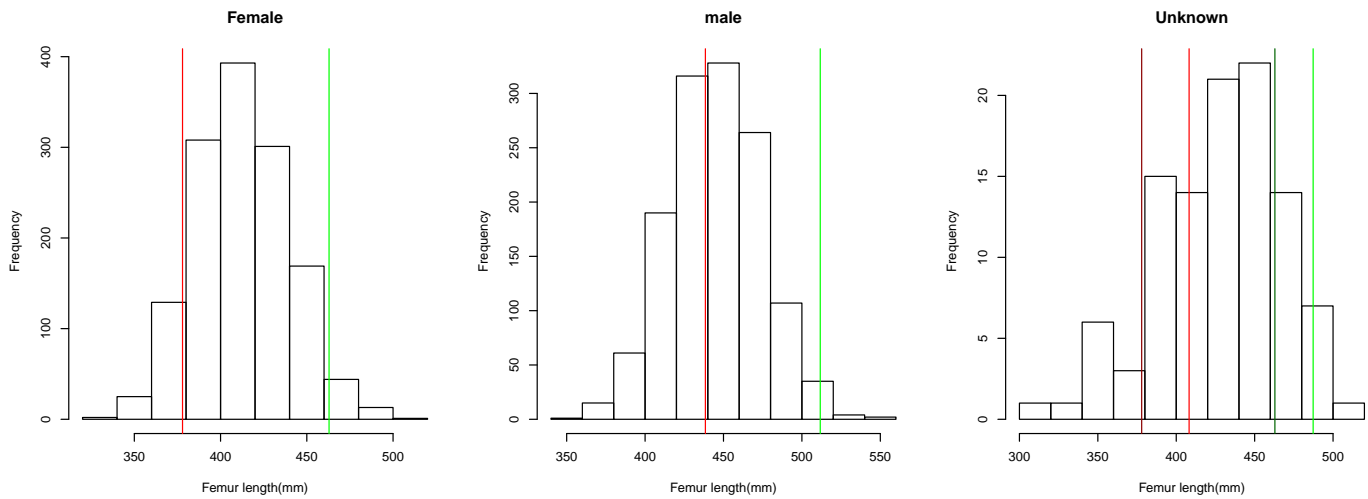
Ideally the center of the histograms should fall within the two lines. This is the case for the female remains and the unknown remains when using female value but not for the male remains; where there a large number of individuals that fall under the red (0 score) line. So I shall explore alternative ranges on this data.

```
hist(stature$FLEN[stature$Sex==3|stature$Sex==4 ],main="male",xlab="Femur length(mm)")
abline(v=Maresh$ML[20],col="green",lty=2)
abline(v=Maresh$M3SD[20],col="green",lty=2)
abline(v=Maresh$AL[20],col="red",lty=2)
abline(v=Maresh$A3SD[20],col="red",lty=2)
abline(v=Maresh$FL[20],col="blue",lty=2)
abline(v=Maresh$F3SD[20],col="blue",lty=2)
legend("topright",c("male","average","female"),
        fill=c("green","red","blue"))
```
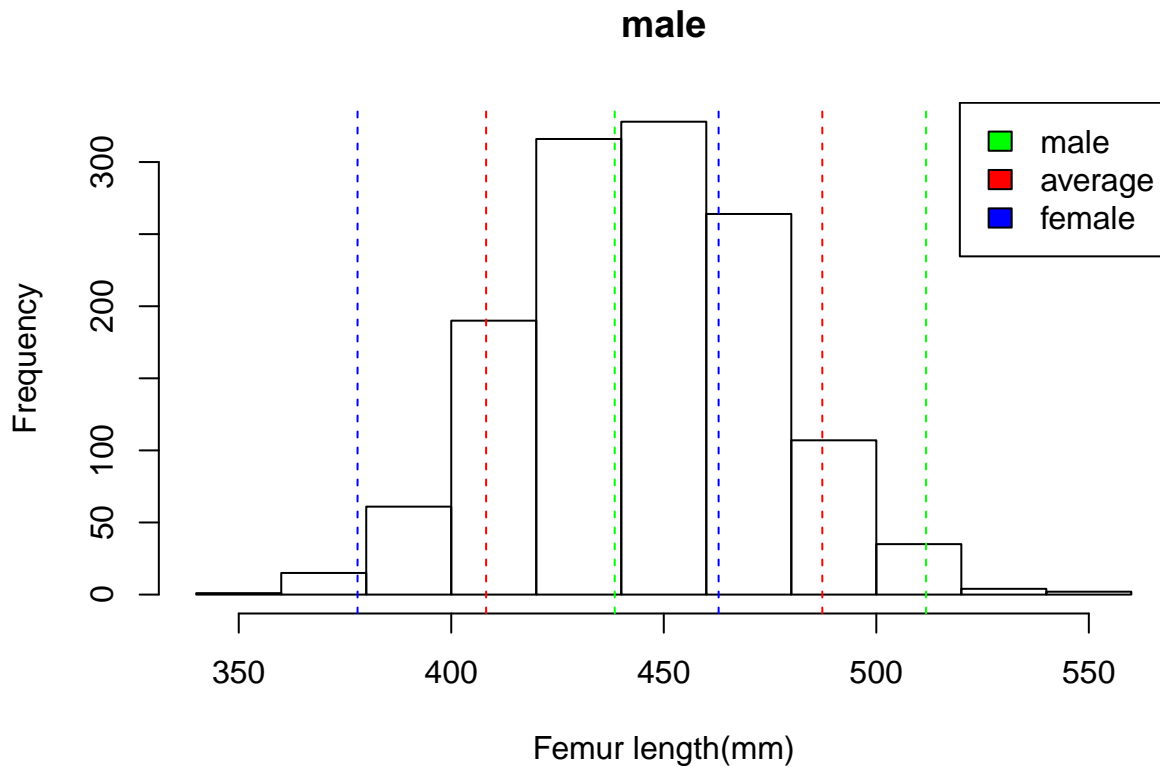
The male range although most technically appropriate would lead to a disproportionately large number of individuals score as 0. The female range similar would lead to more individuals than perhaps can be justified as being score as 100.

The range that seems to best fit the data is the average range. However there isn't a clear objective reason for using an average range. Particularly if minimum values are used for unknown data. I propose a compromise, I will continue to use the male value for the upper 100 score limit but lower the 0 score limit to the average value.

As a result of this suitability check I shall modify the reference database as such:

```
Maresh2<-Maresh[,c("age","FL","F3SD","ML","M3SD")]
Maresh2$M3SD<-Maresh$A3SD# lower male 0 score to average

#replace the average values with minimum values to be used for unkown sex
for(i in 1:length(Maresh$age)){
  Maresh2$UL[i]<-min(Maresh$ML[i],Maresh$FL[i])
  Maresh2$U3SD[i]<-min(Maresh$M3SD[i],Maresh$F3SD[i])
  }

kable(Maresh2)
```

age	FL	F3SD	ML	M3SD	UL	U3SD
0.125	87.2	74.91451	86.0	72.67645	86.0	70.43838
0.250	100.8	88.86552	100.7	87.00195	100.7	85.13838
0.500	111.1	97.76147	112.2	96.96591	111.1	96.17036
1.000	134.6	121.02746	136.6	120.85741	134.6	120.68736
1.500	153.9	132.37114	155.4	134.11568	153.9	132.37114
2.000	170.8	148.68612	172.4	150.71466	170.8	148.68612
2.500	185.2	159.92699	187.2	162.74056	185.2	159.92699
3.000	198.4	170.43588	200.3	173.25796	198.4	170.43588
3.500	211.1	180.09376	212.1	181.00328	211.1	180.09376
4.000	223.2	189.61966	224.1	193.17028	223.2	189.61966
4.500	235.5	198.29251	235.7	202.19517	235.5	198.29251
5.000	247.0	214.00468	247.5	214.72270	247.0	214.00468
6.000	268.9	227.12933	269.7	229.86943	268.9	227.12933
7.000	288.8	247.84836	291.1	249.93440	288.8	247.84836
8.000	309.8	261.24306	312.1	264.73315	309.8	261.24306
9.000	328.7	274.05881	330.4	280.93455	328.7	274.05881
10.000	347.9	290.09969	349.3	295.71389	347.9	290.09969
11.000	367.0	292.81903	367.0	305.98206	367.0	292.81903
12.000	387.6	320.90733	386.1	326.76810	386.1	320.90733
18.000	462.9	377.95460	511.7	408.20484	462.9	377.95460

```
write.table(Maresh2,"data/Maresh.txt",row.names=FALSE)
Maresh<-Maresh2
```

Now I shall begin to use this reference data to turn the femur lengths recorded into stature scores. Starting by creating a function to define the logical test at the heart of the assessment.

```
StatureLogic<-function(x,A,B){# x = value, A=modern standard, B= 10th percentile
if(x>=A){output<-100}# if reaches or exceeds modern standard
else{
if(x<=B){output<-0}# if less than or equal to 1oth percentile
else{output<-100-((A-x)/(A-B)*100)}# overwise graded score between 0 and 100
}
output
}
```

This test then need to be applied using the appropriate reference data. The following function applies the test to an

individual

```r
Stature<-function(femur,Age,Sex,Maresh){# reference data, femur length age and sex
  if(anyNA(c(Age,femur))|round(Age) %in% 12:17){# age excluded in S&R 11.75-17.75
    s<-NA
  }else{
    if(round(Age)>=18){# adult values
      if(Sex==6|is.na(Sex)){s<-StatureLogic(femur,Maresh$UL[Maresh$age==18],
                                            Maresh$U3SD[Maresh$age==18])}#unknown sex
      if(Sex==1|Sex==2){s<-StatureLogic(femur,Maresh$FL[Maresh$age==18],
                                        Maresh$F3SD[Maresh$age==18])}#female
        if(Sex==3|Sex==4){s<-StatureLogic(femur,Maresh$ML[Maresh$age==18],
                                          Maresh$M3SD[Maresh$age==18])}#male
    }else{s<-StatureLogic(femur,Maresh$UL[which.min(abs(Maresh$age-Age))],
                          Maresh$U3SD[which.min(abs(Maresh$age-Age))])}#juvenilles
  }
  round(s,digits=2)
}
```

and the code below applies this function to the population

```r
for (i in 1:length(D1$Individual)){
if(PopA$Age[i]>=18){
  D1$Stature[D1$Individual==PopA$ID[i]]<-Stature(PopA$FLEN[i],
                                                  PopA$Age[i],PopA$Sex[i],Maresh)
}else{
  D1$Stature[D1$Individual==PopA$ID[i]]<-Stature(PopA$FDIAP[i],PopA$Age[i],
                                                  PopA$Sex[i],Maresh)}
}
summary(D1$Stature)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   33.88   59.71   54.95   77.16  100.00    1006
```

```r
hist(D1$Stature)
```

Histogram of D1$Stature
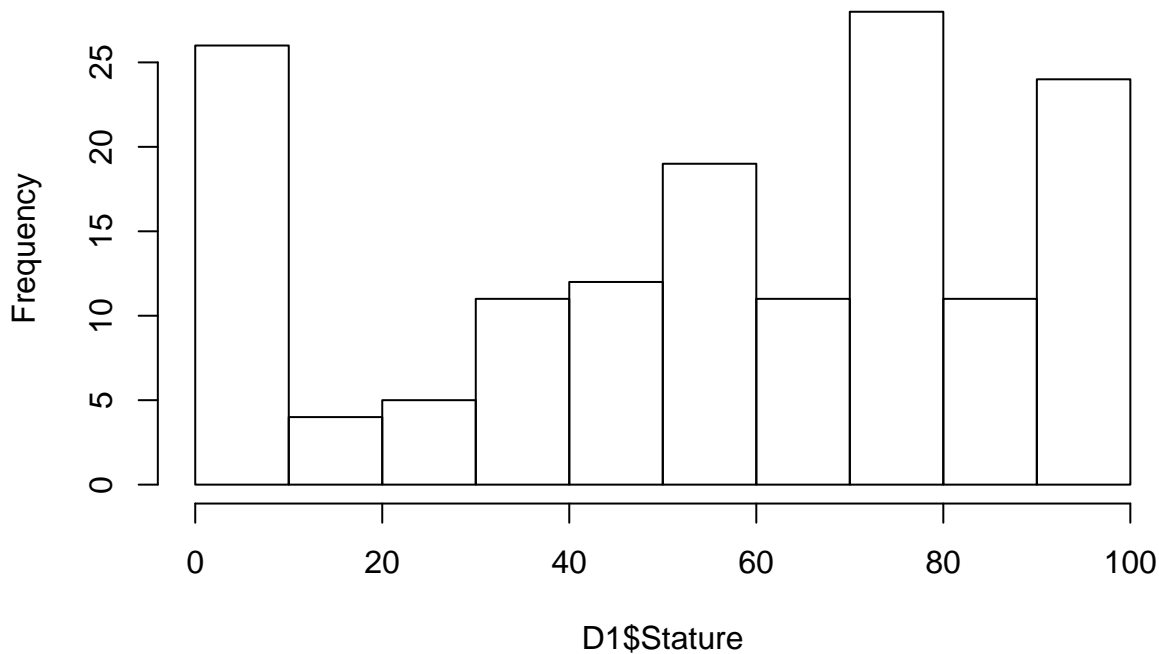
Hypoplasia

This is the first of series of categorical variables that can be dealt with in similar manners. For each category 1 equates to a score of 100 and the highest category a score of 0. The other categories are assigned scores that follow regular intervals determined by the number of categories. I.e. if there are 3 categories then the 2nd category scores 50 and if there are 4 categories it would score 67.

This basic idea is defined in the function below.

```
Cat2Percent<-function(x,c){# value and number of categories
ref<-data.frame(cat=c(0,1,2,3,4,5),two=c(NA,100,0,NA,NA,NA),three=c(NA,100,50,0,NA,NA),
              four=c(NA,100,67,33,0,NA),five=c(NA,100,75,50,25,0))
ref[ref$cat==x,c]
}
```

There are four fields in the original dataset that correspond to linear enamel hypoplasia, each of which has 3 categories. The lowest score from these 4 categories becomes the hypoplasia score.

```
for(i in 1:length(PopA[,1])){
x<-c(Cat2Percent(PopA$LDI[i],3),Cat2Percent(PopA$LDC[i],3),
     Cat2Percent(PopA$LPI[i],3),Cat2Percent(PopA$LPC[i],3))
if(sum(is.na(x))==length(x)){D1$Hypoplasia[D1$Individual==PopA$ID[i]]<-NA}# if all NA's
else{D1$Hypoplasia[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}
}

summary(D1$Hypoplasia)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   50.00  100.00   64.93  100.00  100.00     340
```

```
hist(D1$Hypoplasia)
```

Histogram of D1$Hypoplasia
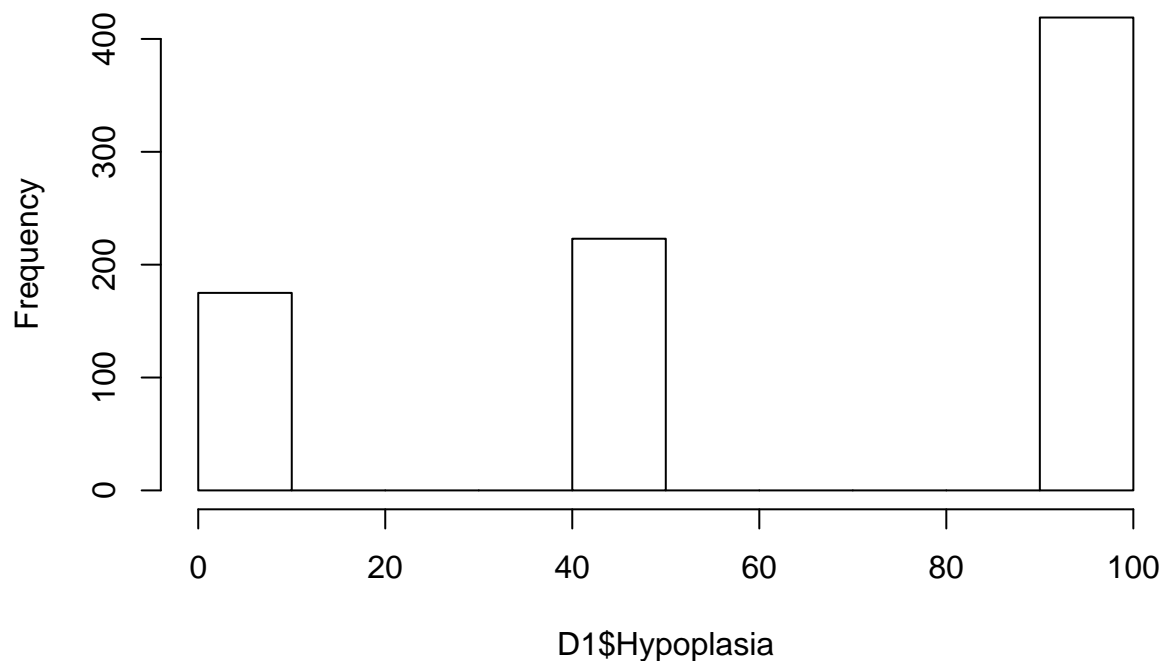
Anemia

Like hypoplasia this is a categorical variable in which the lowest score counts.

```
for(i in 1:length(PopA[,1])){
x<-c(Cat2Percent(PopA$CROB[i],3),Cat2Percent(PopA$PORHY[i],2))
if(sum(is.na(x))==length(x)){D1$Anemia[D1$Individual==PopA$ID[i]]<-NA}
else{D1$Anemia[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}
}

summary(D1$Anemia)
```
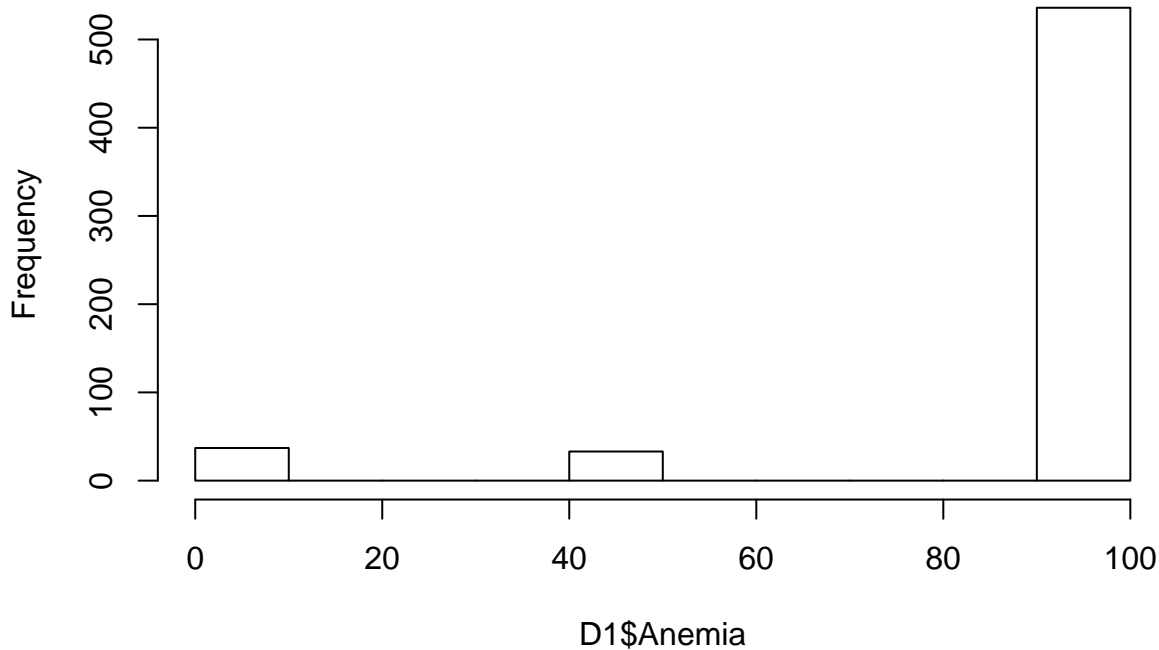
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00  100.00  100.00   91.17  100.00  100.00     551
```

```
hist(D1$Anemia)
```

Histogram of D1$Anemia

Dental health

This has two weighted components completeness(75%) and abscesses (25%).

Completeness is defined as 1-((Antemortem loss + cavities)/(antemortem loss+teeth present)).For the abscess score 100= no abscesses, 50= one abscess and 0= more than one abscess.

There is an additional consideration for this measure; if the number of teeth recorded (SUMTET + SUMPRE) is less than 8 then you cannot score the attribute.

```
DentalHealth<-function(SUMTET,SUMPRE,SUMCAV,SUMABS){
x<-c(SUMTET,SUMPRE,SUMCAV,SUMABS)
if(sum(is.na(x))>0|SUMTET+SUMPRE<8){DH<-NA}
else{
completeness<-1-((SUMPRE+SUMCAV)/(SUMPRE+SUMTET))
if(SUMABS==0){Abscess<-1}
if(SUMABS==1){Abscess<-0.5}
if(SUMABS>1){Abscess<-0}
DH<-(completeness*75)+(Abscess*25)
}
round(DH,digits=2)
}
```

```
for(i in 1:length(PopA[,1])){
D1$DentalHealth[D1$Individual==PopA$ID[i]]<-DentalHealth(PopA$SUMTET[i],PopA$SUMPRE[i],
                                              PopA$SUMCAV[i],PopA$SUMABS[i])
}
```

```
summary(D1$DentalHealth)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
```

```
##    15.52   71.15   87.07   80.46   95.31  100.00      649
```

```
hist(D1$DentalHealth)
```

Histogram of D1$DentalHealth
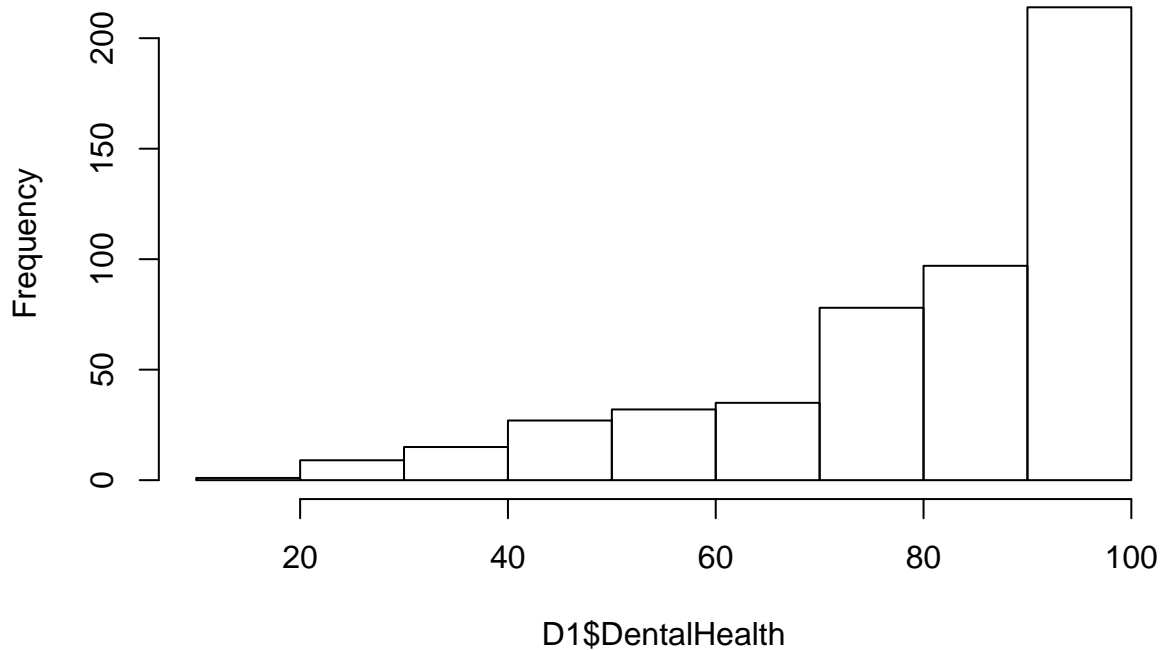
Infection

This is another attribute based on categorical variables.

```
for(i in 1:length(PopA[,1])){
x<-c(Cat2Percent(PopA$SKELINF[i],3),Cat2Percent(PopA$TIBINF[i],4))
if(sum(is.na(x))==length(x)){D1$Infections[D1$Individual==PopA$ID[i]]<-NA}
else{D1$Infections[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}
}

summary(D1$Infections)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00  100.00   65.78  100.00  100.00
```

```
hist(D1$Infections)
```

Histogram of D1$Infections
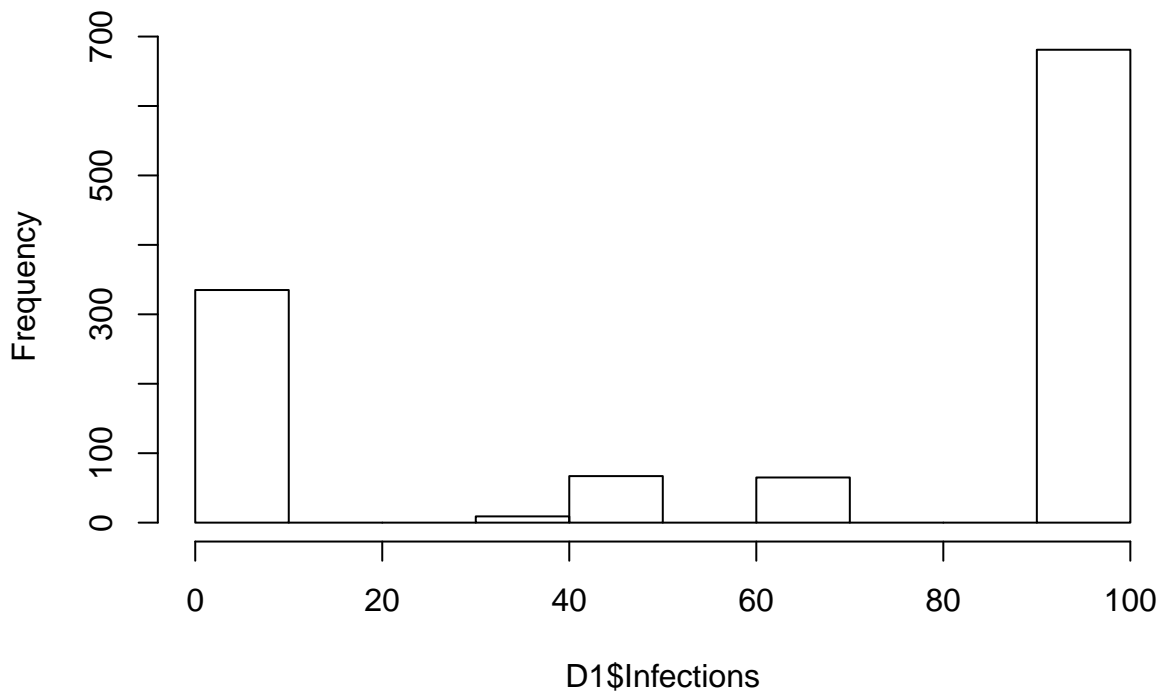
DJD

The Degenerative joint disease attribute score is the lowest score of from 8 variables:DJSH,DJHK,DJCER,DJTHO,DJLUM,DJTMJ,DJW
and DJHAN.

```
for(i in 1:length(PopA[,1])){
x<-c(Cat2Percent(PopA$DJSH[i],5),Cat2Percent(PopA$DJHK[i],5),Cat2Percent(PopA$DJCER[i],4),
     Cat2Percent(PopA$DJTHO[i],4),Cat2Percent(PopA$DJLUM[i],4),Cat2Percent(PopA$DJTMJ[i],2),
     Cat2Percent(PopA$DJWR[i],2),Cat2Percent(PopA$DJHAN[i],2))
if(sum(is.na(x))==length(x)){D1$DJD[D1$Individual==PopA$ID[i]]<-NA}
else{D1$DJD[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}
}

summary(D1$DJD)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00    0.00   67.00   53.48  100.00  100.00     757
```

```
hist(D1$DJD)
```

Histogram of D1$DJD
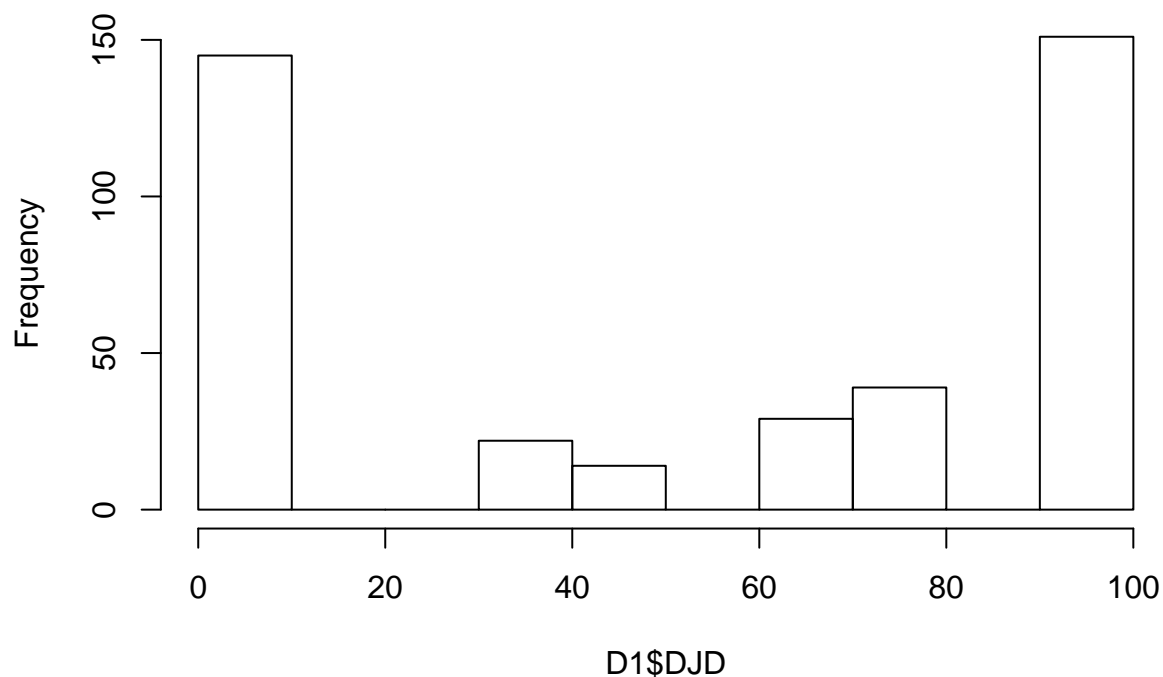
Trauma

This is also recorded in categorical variables but it is scored differently. An individual receives 100 if trauma is absent and 0 if there is trauma on any part.

```r
Fact2Num<- function(x){
  a<-as.character(x)
  a<-as.numeric(a)
  a
}

PopA[,39:45]<-apply(PopA[,39:45],2,FUN=Fact2Num)
for(i in 1:length(PopA[,1])){
  if(PopA$TRARM[i]>1|PopA$TRLEG[i]>1|PopA$TRNAS[i]>1|PopA$TRFAC[i]>1|PopA$TRSKUL[i]>1|
       PopA$TRHAN[i]>1|PopA$TRWEAP[i]>1){# if any trauma

    D1$Trauma[D1$Individual==PopA$ID[i]]<-0

    }else{
      if(sum(c(PopA$TRARM[i],PopA$TRLEG[i],PopA$TRNAS[i],PopA$TRFAC[i],PopA$TRSKUL[i],
             PopA$TRHAN[i],PopA$TRWEAP[i]),na.rm=TRUE)==0){#if all unrecorded

        D1$Trauma[D1$Individual==PopA$ID[i]]<-NA

        }else{D1$Trauma[D1$Individual==PopA$ID[i]]<-100}# otherwise trauma =100
      }}

summary(D1$Trauma)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##    0.00  100.00  100.00   94.04  100.00  100.00
```

```
hist(D1$Trauma)
```

Histogram of D1$Trauma
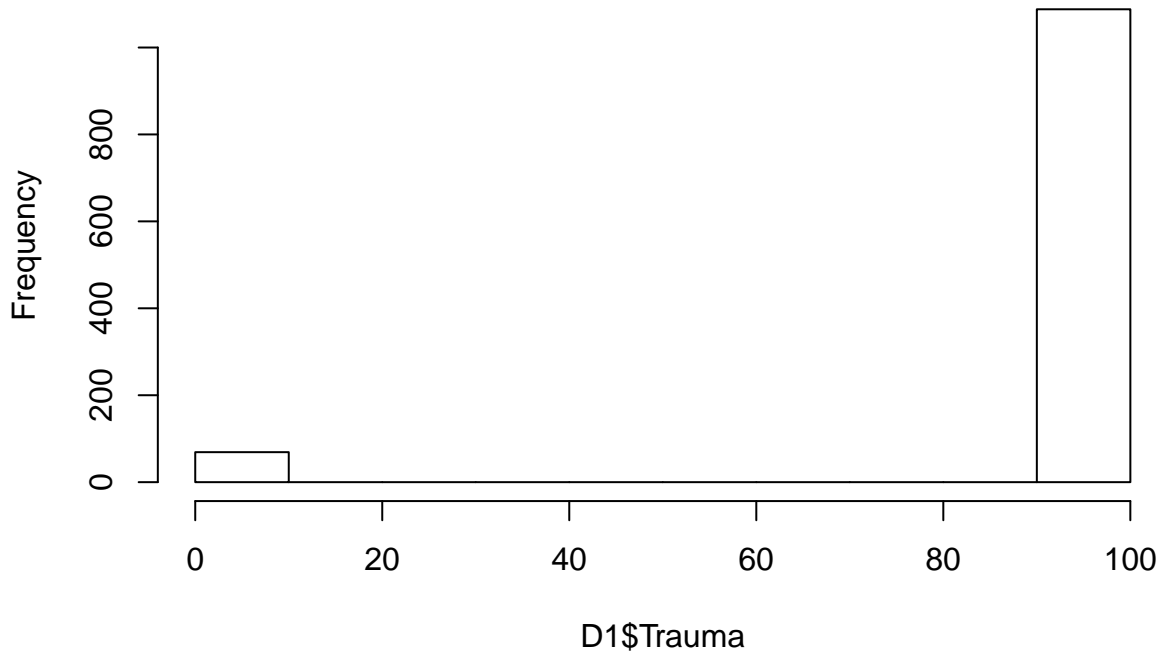
Section 2: Age Specific score.

For each attribute I will calculate the total score at each age and the number of contributors to that score.

For this section the attributes can be divided into 2 groups:

1. Stature, anemia and hypoplasia. These are considered to take effect from birth to death. This means any contributor (any individual without an NA for that attribute) will add their score to the every age category which they are alive during.

2.Everything else effects a ten year period prior to death. The contributors score is added to the total for the age categories that represent this period. Prior to this period(from birth to 10 years before death) any contributor will add a score of 100 to the total. As with the previous group an individual stops contributing on their death.

DJD and dental health which fall in the second group have to additional caveat that they are not considered to effect an individual below 18 years old.

Birth to death

```
B2DTotal<-function(D1,D2,D2row,D1column){#D2row & D1 Column= number
D2$contributors[D2row]<-length(D1[!(is.na(D1[,D1column])),D1column])
if(D2$contributors[D2row]==0){D2<-D2}# if no contributors don't change
else{
```

16

```r
D2$Age0_4[D2row]<-sum(D1[,D1column],na.rm=TRUE)# everyone who can contributes to the first age period
D2$Age5_14[D2row]<-sum(D1[D1$B2D %in% 2:6,D1column],na.rm=TRUE)#everyone not dead
D2$Age15_24[D2row]<-sum(D1[D1$B2D %in% 3:6,D1column],na.rm=TRUE)
D2$Age25_34[D2row]<-sum(D1[D1$B2D %in% 4:6,D1column],na.rm=TRUE)
D2$Age35_44[D2row]<-sum(D1[D1$B2D %in% 5:6,D1column],na.rm=TRUE)
D2$Age45[D2row]<-sum(D1[D1$B2D==6,D1column],na.rm=TRUE)
}
D2
}
```

```r
D2<-B2DTotal(D1,D2,D2row=1,D1column=5)# stature
D2<-B2DTotal(D1,D2,D2row=2,D1column=6)# Hypoplasia
D2<-B2DTotal(D1,D2,D2row=3,D1column=7)# anemia
```

```r
kable(D2[1:3,])
```

Trait	contributors	Age0_4	Age5_14	Age15_24	Age25_34	Age35_44	Age45
Stature	151	8298.14	7162.98	6567.01	5685.33	3305.8	1583.97
Hypoplasia	817	53050.00	34200.00	31650.00	27150.00	19150.0	9800.00
Anemia	606	55250.00	46900.00	43350.00	37250.00	25950.0	13200.00

Ten years prior to death (part 1)

Trauma and infection have an effect over a 10 year period prior to death, with no restrictions.

```r
TenTotal<-function(D1,D2,D2row,D1column){
D2$contributors[D2row]<-length(D1[!(is.na(D1[,D1column])),D1column])
D1<- D1[!(is.na(D1[,D1column])),]
if(D2$contributors[D2row]==0){D2<-D2# if no contributors don't change
}else{
  # pattern: a-b+c
  #a=100* number of contributors
  #b=100* number of dead and those who's ten years include that period
  #c=sum of scores of those whos ten years include that period
D2$Age0_4[D2row]<-(100*D2$contributors[D2row])-
  (100*length(D1[D1$Ten %in% c(01,12),D1column]))+
  sum(D1[D1$Ten %in% c(01,12),D1column])

D2$Age5_14[D2row]<-(100*D2$contributors[D2row])-
  (100*length(D1[D1$Ten %in% c(01,12,23),D1column]))+
  sum(D1[D1$Ten %in% c(12,23),D1column])

D2$Age15_24[D2row]<-(100*D2$contributors[D2row])-
  (100*length(D1[D1$Ten %in% c(01,12,23,34),D1column]))+
  sum(D1[D1$Ten %in% c(23,34),D1column])

D2$Age25_34[D2row]<-(100*D2$contributors[D2row])-
  (100*length(D1[D1$Ten %in% c(01,12,23,34,45),D1column]))+
  sum(D1[D1$Ten %in% c(34,45),D1column])

D2$Age35_44[D2row]<-(100*D2$contributors[D2row])-
  (100*length(D1[D1$Ten %in% c(01,12,23,34,45,56),D1column]))+
  sum(D1[D1$Ten %in% c(45,56),D1column])

D2$Age45[D2row]<-sum(D1[D1$Ten %in% c(56,60),D1column])
```

```
}
D2
}
```

```
D2<-TenTotal(D1,D2,D2row=5,D1column=9)# infection
D2<-TenTotal(D1,D2,D2row=7,D1column=11)# Trauma
kable(D2[c(5,7),])
```

	Trait	contributors	Age0_4	Age5_14	Age15_24	Age25_34	Age35_44	Age45
5	Infections	1157	101286	65901	56039	44928	29175	11804
7	Trauma	1157	115600	71800	64400	53900	37100	17400

Ten years prior to death (part 2)

DJD and Dental health also effect a period 10 years prior to death but are not considered to effect anyone under 18. There for the code above need to be adjusted slightly so that the first two age category (0 to 4 and 5 to 14) calculations do not include the contributor section of the equation.

```
TenTotal2<-function(D1,D2,D2row,D1column){
  D2$contributors[D2row]<-length(D1[!(is.na(D1[,D1column])),D1column])
  D1<- D1[!(is.na(D1[,D1column])),]
  if(D2$contributors[D2row]==0){D2<-D2
  }else{
    D2$Age0_4[D2row]<-(100*D2$contributors[D2row])

    D2$Age5_14[D2row]<-(100*D2$contributors[D2row])-
      (100*length(D1[D1$Ten %in% c(01),D1column]))

    D2$Age15_24[D2row]<-(100*D2$contributors[D2row])-
      (100*length(D1[D1$Ten %in% c(01,12,23,34),D1column]))+
      sum(D1[D1$Ten %in% c(23,34),D1column])

    D2$Age25_34[D2row]<-(100*D2$contributors[D2row])-
      (100*length(D1[D1$Ten %in% c(01,12,23,34,45),D1column]))+
      sum(D1[D1$Ten %in% c(34,45),D1column])

    D2$Age35_44[D2row]<-(100*D2$contributors[D2row])-
      (100*length(D1[D1$Ten %in% c(01,12,23,34,45,56),D1column]))+
      sum(D1[D1$Ten %in% c(45,56),D1column])

    D2$Age45[D2row]<-sum(D1[D1$Ten %in% c(56,60),D1column])
  }
  D2
}
```

```
D2<-TenTotal2(D1,D2,D2row=4,D1column=8)# DentalHealth
```

```
D2<-TenTotal2(D1,D2,D2row=6,D1column=10)# DJD
```

```
kable(D2[c(4,6),])
```

	Trait	contributors	Age0_4	Age5_14	Age15_24	Age25_34	Age35_44	Age45
4	DentalHealth	508	50800	50800	46956.26	37387.13	23687.89	10671.05
6	DJD	400	40000	40000	36651.00	26836.00	13011.00	3958.00

Average age score

Currently the D2 data frame contains total attribute scores for each age category. To turn these into average scores they need to be divided by the number of contributors for that attribute.

```
D3<-D2
D3[,3:8]<-D2[,3:8]/D2[,2]

kable(D3)
```

Trait	contributors	Age0_4	Age5_14	Age15_24	Age25_34	Age35_44	Age45
Stature	151	54.95457	47.43695	43.49013	37.65119	21.89272	10.48987
Hypoplasia	817	64.93268	41.86047	38.73929	33.23133	23.43941	11.99510
Anemia	606	91.17162	77.39274	71.53465	61.46865	42.82178	21.78218
DentalHealth	508	100.00000	100.00000	92.43358	73.59671	46.62970	21.00600
Infections	1157	87.54192	56.95851	48.43474	38.83146	25.21608	10.20225
DJD	400	100.00000	100.00000	91.62750	67.09000	32.52750	9.89500
Trauma	1157	99.91357	62.05704	55.66119	46.58600	32.06569	15.03889

```
plot(1:7,rep(100,7),ylim=c(0,100),ylab="Score",xlab="Age cateogory",xaxt="n",type="n",main="Attribute score b
axis(1,at=1:6,labels=c("0-4","5-14","15-24","25-34","35-44","45+"))
colours<-c("indianred","cadetblue","seagreen","tan","orchid","orange","navy")
lines(1:6,D3[1,3:8],col=colours[1]);lines(1:6,D3[2,3:8],col=colours[2])
lines(1:6,D3[3,3:8],col=colours[3]);lines(1:6,D3[4,3:8],col=colours[4])
lines(1:6,D3[5,3:8],col=colours[5]);lines(1:6,D3[6,3:8],col=colours[6])
lines(1:6,D3[7,3:8],col=colours[7])
legend("topright",legend=D3$Trait,fill=colours,bg="white")
```
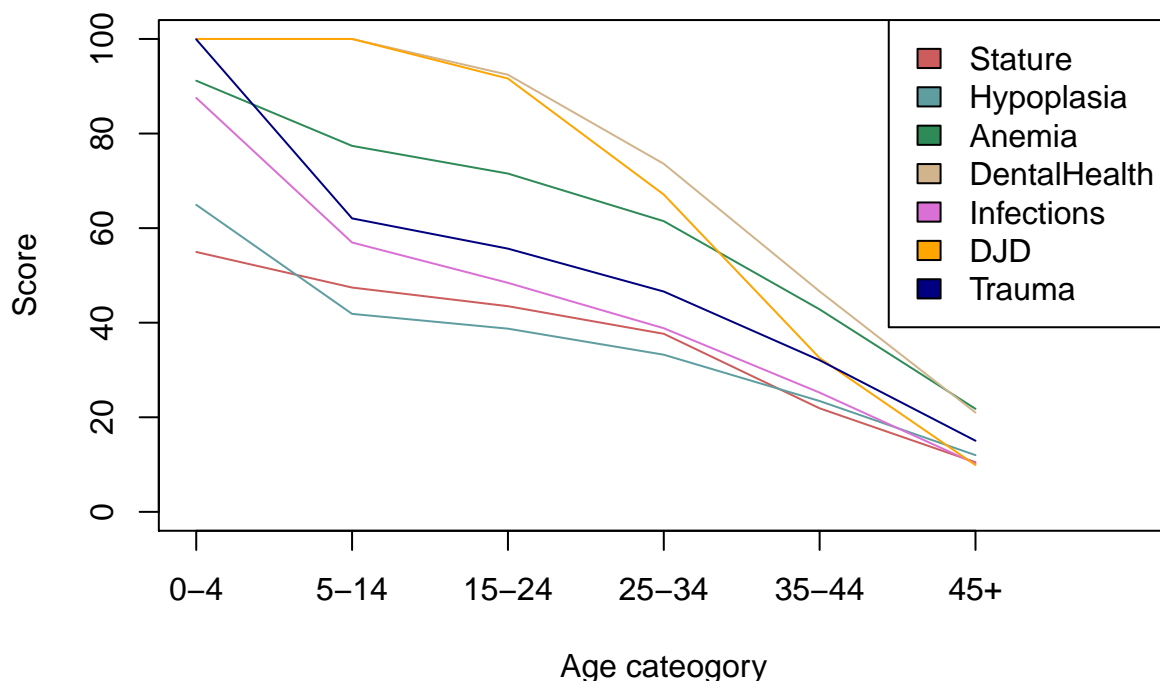
Section 3 creating the index.

Average attribute score

The first step in calculating is creating a single average score for each attribute. This is done by multiplying each age specific rate by the person-years lived in each age category in the reference population summing then dividing by life expectancy at birth.

Although a life expectancy is given for the reference population I can find no person-years lived data in S&R so I am going to estimate both of these quantities from the population.

```
PYL<-function(D1){# uses info in D1 database to calculate the person years live in each age category
  pyl<-data.frame(age=c("0_4","5_14","15_24","25_34","35_44","45+"),pyl=NA)
  #pyl= (number of years in period x number of people who survived till the end)
  #+(the years lived during that period by those who died in that period)
  pyl$pyl[1]<-(4*length(D1$Age[D1$B2D!=1]))+sum(D1$Age[D1$B2D==1])# 0-4
  pyl$pyl[2]<-(10*length(D1$Age[D1$B2D %in% 3:6]))+sum((D1$Age[D1$B2D==2]-5))#(-5= years before)
  pyl$pyl[3]<-(10*length(D1$Age[D1$B2D %in% 4:6]))+sum((D1$Age[D1$B2D==3]-15))
  pyl$pyl[4]<-(10*length(D1$Age[D1$B2D %in% 5:6]))+sum((D1$Age[D1$B2D==4]-25))
  pyl$pyl[5]<-(10*length(D1$Age[D1$B2D %in% 6]))+sum((D1$Age[D1$B2D==5]-35))
  pyl$pyl[6]<-sum((D1$Age[D1$B2D==6]-44))
  # n.b. no survivors in last period everyone who isn't already dead must die during this period
  pyl$Averagepyl<-pyl$pyl/length(D1$Age)# number of years lived in the category by the average person
  pyl

}
```

```
pyl<-PYL(D1)
kable(pyl)
```

age	pyl	Averagepyl
0_4	3236.3	2.797148
5_14	6888.3	5.953587
15_24	6132.4	5.300259
25_34	4865.5	4.205272
35_44	2642.6	2.284010
45+	1904.0	1.645635

from this an approximation of life expectancy can be calculated by summing the averagepyl column

```
HI$lifeEx<-sum(pyl$Averagepyl)
HI$pyl<-sum(pyl$pyl)
```

For this population that gives a life expectancy value of 22.1859118. This and the averagepyl values can be used to calculate the average attribute scores

```
for(i in 1:7){
  D3$AverageScore[i]<-sum(D3[i,3:8]*pyl$Averagepyl)/HI$lifeEx
}
```

```
HI[,6:12]<-D3$AverageScore
kable(HI[,6:12])
```

Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
40.2167	38.2764	67.02814	81.83394	48.60615	78.13208	55.7943

```
plot(1:7,rep(100,7),ylim=c(0,100),ylab="Score",xlab="Age cateogory",xaxt="n",type="n",main="Attribute scores
axis(1,at=1:6,labels=c("0-4","5-14","15-24","25-34","35-44","45+"))

abline(h=D3$AverageScore[1],col=colours[1],lty=2);abline(h=D3$AverageScore[2],col=colours[2],lty=2)
abline(h=D3$AverageScore[3],col=colours[3],lty=2);abline(h=D3$AverageScore[4],col=colours[4],lty=2)
abline(h=D3$AverageScore[5],col=colours[5],lty=2);abline(h=D3$AverageScore[6],col=colours[6],lty=2)
abline(h=D3$AverageScore[7],col=colours[7],lty=2)

lines(1:6,D3[1,3:8],col=colours[1],lwd=2);lines(1:6,D3[2,3:8],col=colours[2],lwd=2)
lines(1:6,D3[3,3:8],col=colours[3],lwd=2);lines(1:6,D3[4,3:8],col=colours[4],lwd=2)
lines(1:6,D3[5,3:8],col=colours[5],lwd=2);lines(1:6,D3[6,3:8],col=colours[6],lwd=2)
lines(1:6,D3[7,3:8],col=colours[7],lwd=2)

legend("topright",legend=D3$Trait,fill=colours,bg="white")
```


Attribute scores (by age and average)
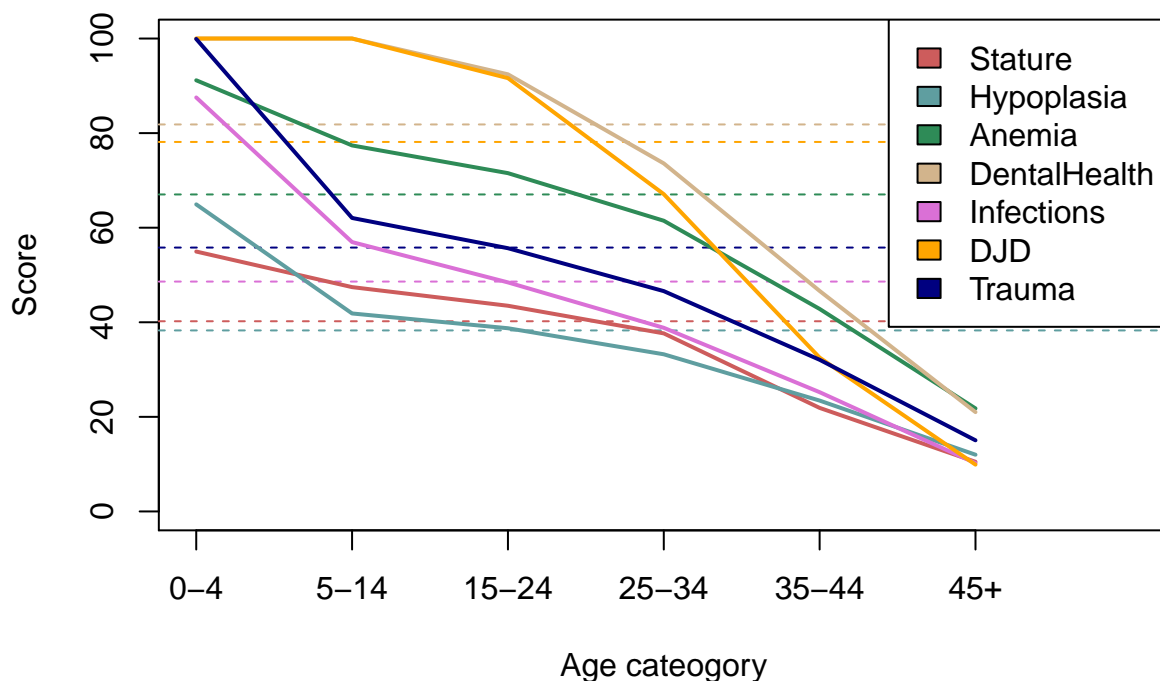
Health Index

The health index is a simple average of these average attribute scores.

```
HI$HI<-mean(D3$AverageScore)
```

This gives a value of 58.5553867.

This should be times by the life expectancy in the reference population(26.38) to get the comparable quality adjusted life years.

```
HI$QALY<-26.38*(HI$HI/100)
HI<-round(HI[,sapply(HI, is.numeric)],2)
kable(HI)
```

	HI	QALY	lifeEx	pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
	58.56	15.45	22.19	25669.1	40.22	38.28	67.03	81.83	48.61	78.13	55.79

Appendix: Application to multiple populations

The code above can be distilled into the following function for calculating the health index on a single population:

```r
HealthIndex<-function(Pop){
  # pop= dataframe with Info for a single population

#source functions/ data
source('HIfunctions.R')# file containing all functions used e.g Fact2Num,Stature ect.
Maresh<-read.table("data/Maresh.txt",header=TRUE)

#adjust input dataframe
PopA<-Pop[!(is.na(Pop$Age)),]#rm observations without age
PopA[,39:45]<-apply(PopA[,39:45],2,FUN=Fact2Num)# change trauma field to factors

#set up empty data sets
D1<-data.frame(Individual=PopA$ID,Age=NA,B2D=NA,Ten=NA,Stature=NA,Hypoplasia=NA,
            Anemia=NA,DentalHealth=NA,Infections=NA,DJD=NA,Trauma=NA)

D2<-data.frame(Trait=c("Stature","Hypoplasia","Anemia","DentalHealth",
                    "Infections","DJD","Trauma"),
            contributors=NA,Age0_4=NA,Age5_14=NA,Age15_24=NA,Age25_34=NA,
            Age35_44=NA,Age45=NA)

HI<-data.frame(Pop=PopA$Site[1],num=length(PopA$Age),HI=NA,QALY=NA,lifeEx=NA,Pyl=NA,
            Stature=NA,Hypoplasia=NA,Anemia=NA,Dental=NA,Infection=NA,DJD=NA,Trauma=NA)
#added num to HI dataframe, number of individuals recorded (with an age)

#filling D1
for(i in 1:length(D1$Individual)){
  #Age
  D1$Age[D1$Individual==PopA$ID[i]]<-round(LovejoyAdj(PopA$Age[i]))# move rounding age to earlier in code

  #B2D
  if(D1$Age[i]%in% 0:4){D1$B2D[i]<-1}
  if(D1$Age[i]%in% 5:14){D1$B2D[i]<-2}
  if(D1$Age[i]%in% 15:24){D1$B2D[i]<-3}
  if(D1$Age[i]%in% 25:34){D1$B2D[i]<-4}
  if(D1$Age[i]%in% 35:44){D1$B2D[i]<-5}
  if(D1$Age[i]>= 45){D1$B2D[i]<-6}

  #Ten
  if((D1$Age[i]<=4)& (D1$Age[i]-10<=4)){D1$Ten[i]<-01}
  if((D1$Age[i]%in% 5:14)& ((D1$Age[i]-10)<=4)){D1$Ten[i]<-12}
  if((D1$Age[i]%in% 15:24)& ((D1$Age[i]-10)%in% 5:14)){D1$Ten[i]<-23}
  if((D1$Age[i]%in% 25:34)& ((D1$Age[i]-10)%in% 15:24)){D1$Ten[i]<-34}
  if((D1$Age[i]%in% 35:44)& ((D1$Age[i]-10)%in% 25:34)){D1$Ten[i]<-45}
  if((D1$Age[i]>= 45)& ((D1$Age[i]-10)%in% 35:44)){D1$Ten[i]<-56}
  if((D1$Age[i]>=45)& ((D1$Age[i]-10)>=45)){D1$Ten[i]<-60}

  #Stature
  if(PopA$Age[i]>=18){
    D1$Stature[D1$Individual==PopA$ID[i]]<-Stature(PopA$FLEN[i],
```

```r
                                                   PopA$Age[i],PopA$Sex[i],Maresh)
  }else{
    D1$Stature[D1$Individual==PopA$ID[i]]<-Stature(PopA$FDIAP[i],PopA$Age[i],
                                                   PopA$Sex[i],Maresh)}


  #hypoplasia
  x<-c(Cat2Percent(PopA$LDI[i],3),Cat2Percent(PopA$LDC[i],3),
       Cat2Percent(PopA$LPI[i],3),Cat2Percent(PopA$LPC[i],3))
  if(sum(is.na(x))==length(x)){D1$Hypoplasia[D1$Individual==PopA$ID[i]]<-NA
  }else{D1$Hypoplasia[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}


  #Anemia
  x<-c(Cat2Percent(PopA$CROB[i],3),Cat2Percent(PopA$PORHY[i],2))
  if(sum(is.na(x))==length(x)){D1$Anemia[D1$Individual==PopA$ID[i]]<-NA
  }else{D1$Anemia[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}


  #DentalHealth
  D1$DentalHealth[D1$Individual==PopA$ID[i]]<-DentalHealth(PopA$SUMTET[i],PopA$SUMPRE[i],
                                                          PopA$SUMCAV[i],PopA$SUMABS[i])


  #Infections
  x<-c(Cat2Percent(PopA$SKELINF[i],3),Cat2Percent(PopA$TIBINF[i],4))
  if(sum(is.na(x))==length(x)){D1$Infections[D1$Individual==PopA$ID[i]]<-NA
  }else{D1$Infections[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}


  #DJD
  x<-c(Cat2Percent(PopA$DJSH[i],5),Cat2Percent(PopA$DJHK[i],5),Cat2Percent(PopA$DJCER[i],4),
       Cat2Percent(PopA$DJTHO[i],4),Cat2Percent(PopA$DJLUM[i],4),Cat2Percent(PopA$DJTMJ[i],2),
       Cat2Percent(PopA$DJWR[i],2),Cat2Percent(PopA$DJHAN[i],2))
  if(sum(is.na(x))==length(x)){D1$DJD[D1$Individual==PopA$ID[i]]<-NA
  }else{D1$DJD[D1$Individual==PopA$ID[i]]<-min(x,na.rm=TRUE)}


  #Trauma
  if(PopA$TRARM[i]>1|PopA$TRLEG[i]>1|PopA$TRNAS[i]>1|PopA$TRFAC[i]>1|PopA$TRSKUL[i]>1|
     PopA$TRHAN[i]>1|PopA$TRWEAP[i]>1){
  D1$Trauma[D1$Individual==PopA$ID[i]]<-0
  }else{
    if(sum(c(PopA$TRARM[i],PopA$TRLEG[i],PopA$TRNAS[i],PopA$TRFAC[i],PopA$TRSKUL[i],
             PopA$TRHAN[i],PopA$TRWEAP[i]),na.rm=TRUE)==0){
      D1$Trauma[D1$Individual==PopA$ID[i]]<-NA
    }else{D1$Trauma[D1$Individual==PopA$ID[i]]<-100}}
}


#filling in D2
D2<-B2DTotal(D1,D2,D2row=1,D1column=5)# stature
D2<-B2DTotal(D1,D2,D2row=2,D1column=6)# Hypoplasia
D2<-B2DTotal(D1,D2,D2row=3,D1column=7)# anemia
D2<-TenTotal(D1,D2,D2row=5,D1column=9)# infection
D2<-TenTotal(D1,D2,D2row=7,D1column=11)# Trauma
D2<-TenTotal2(D1,D2,D2row=4,D1column=8)# DentalHealth
D2<-TenTotal2(D1,D2,D2row=6,D1column=10)# DJD
D2[,3:8]<-D2[,3:8]/D2[,2]

#Filling HI
pyl<-PYL(D1)
HI$lifeEx<-sum(pyl$Averagepyl)
HI$Pyl<-sum(pyl$pyl)
for(i in 1:7){
```

```
  if(D2$contributors[i]==0){D2$AverageScore[i]<-NA}
  else{D2$AverageScore[i]<-sum(D2[i,3:8]*pyl$Averagepyl,na.rm=TRUE)/HI$lifeEx}
}
HI[7:13]<-D2$AverageScore
HI$HI<-mean(D2$AverageScore,na.rm=TRUE)
HI$QALY<-26.38*(HI$HI/100)
HI<-round(HI[,sapply(HI, is.numeric)],2)

#return HI dataset
HI
}
```

To apply this is the whole Health Index dataset the following function is used:

```
MultiHI<-function(x){
  x<-x[!(is.na(x$Age)),]#rm no age
  x<-split(x,x$Site)#split by site, this will create a list of dataframes
  x<-Filter(function(x)length(x[,1])>0,x)# rm empty dataframes
  x<-lapply(x,FUN=HealthIndex)#apply function to the list
  x<-do.call("rbind",x)#joins output into single dataframe
  x<-x[order(-x$HI),]# orders by Health Index
  x
}
```

This has been appied to the corrected health index data and the following results were produced:

```
Results<-read.table("data/results.txt",header=TRUE)
```

The top populations with 100% result from small numbers of individuals, incomplete data and an absence of evidence of "abnormality".

```
kable(Results[1:10,])
```

Site	num	HI	QALY	lifeEx	Pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
103	1	100	26.38	15	15	NA	NA	100	100	100	NA	100
108	1	100	26.38	16	16	NA	NA	NA	100	100	NA	100
W0D	8	100	26.38	37	296	NA	NA	NA	NA	100	NA	100
W13	1	100	26.38	37	37	NA	NA	NA	100	100	NA	100
W20	3	100	26.38	37	111	NA	NA	NA	100	100	NA	100
W26	1	100	26.38	37	37	NA	NA	NA	100	100	NA	100
W28	1	100	26.38	48	48	NA	NA	NA	NA	100	NA	100
W31	1	100	26.38	37	37	NA	NA	NA	NA	100	NA	100
W42	2	100	26.38	37	74	NA	NA	NA	100	100	NA	100
W4a	2	100	26.38	37	74	NA	NA	NA	100	100	NA	100

Because of this issue I have remove all populations with 10 or less contributors from the following table. The first table is the result of running HealthIndex() on the whole Health Index dataset; i.e. it the result you would get if all the data were from a single population.

```
Results<-Results[Results$num>10,]
Results2<-read.table("data/results2.txt",header=TRUE)
kable(Results2)
```

	num	HI	QALY	lifeEx	Pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
	11829	62.49	16.49	28.42	336125	28.6	53.96	59.93	78.79	69.83	73.19	73.15

```
kable(Results)
```

	Site	num	HI	QALY	lifeEx	Pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
16	W81	17	99.70	26.30	37.00	629	NA	NA	NA	99.10	100.00	NA	100.00
19	W38	47	99.54	26.26	37.00	1739	NA	NA	NA	98.62	100.00	NA	100.00
20	WR6	11	99.30	26.19	37.00	407	NA	NA	NA	97.89	100.00	NA	100.00
30	WC6	15	98.60	26.01	37.00	555	NA	NA	NA	95.81	100.00	NA	100.00
32	WA7	12	98.49	25.98	37.00	444	NA	NA	NA	95.47	100.00	NA	100.00
34	W4A	24	98.34	25.94	37.00	888	NA	NA	NA	95.02	100.00	NA	100.00
50	W52	73	94.14	24.83	36.92	2695	NA	86.66	98.83	95.06	93.73	NA	96.43
52	Wr3	62	93.80	24.74	37.23	2308	NA	86.48	98.41	91.63	96.69	NA	95.76
53	WO0	17	92.98	24.53	34.41	585	NA	NA	NA	NA	92.98	NA	92.98
56	W71	13	92.16	24.31	38.62	502	NA	NA	NA	90.73	92.87	NA	92.87
62	W77	63	88.02	23.22	39.56	2492	NA	NA	NA	NA	88.34	NA	87.70
64	LNC	22	87.17	22.99	42.64	938	NA	70.02	92.69	84.63	91.76	92.71	91.18
66	WOO	116	85.67	22.60	33.77	3917	NA	NA	NA	NA	89.08	NA	82.26
67	sfc	29	85.52	22.56	44.45	1289	NA	92.00	67.45	89.84	90.33	81.86	91.63
68	WM2	12	84.84	22.38	38.33	460	NA	NA	NA	NA	93.70	NA	75.98
69	105	13	83.29	21.97	25.92	337	42.04	NA	93.47	91.20	91.33	90.37	91.33
70	sfo	35	83.16	21.94	46.94	1643	NA	66.67	90.65	77.22	90.30	NA	90.99
74	W46	23	82.00	21.63	31.91	734	NA	NA	NA	82.55	82.32	NA	81.13
75	WMO	27	80.35	21.20	30.59	826	NA	NA	NA	NA	81.38	NA	79.32
76	snt	14	80.19	21.15	43.43	608	NA	73.05	NA	NA	93.24	61.22	93.24
82	cot	89	77.02	20.32	34.16	3040	19.32	81.69	96.51	87.30	83.06	88.77	82.48
84	flo	57	76.44	20.16	19.61	1118	6.03	89.73	87.69	91.09	87.30	91.30	81.93
85	WO7	230	76.22	20.11	32.37	7446	47.27	78.45	80.22	88.74	79.31	80.25	79.29
86	W60	787	75.55	19.93	36.59	28797	19.85	77.23	78.99	86.95	94.81	NA	95.47
88	LNP	126	75.29	19.86	36.11	4550	NA	60.36	62.91	79.57	83.77	83.93	81.20
89	WC8	158	74.57	19.67	34.96	5524	21.96	70.73	77.10	82.28	90.14	90.85	88.95
92	ftl	19	74.17	19.57	23.42	445	49.88	88.05	88.79	80.78	87.03	89.01	35.65
93	3C9	16	73.77	19.46	24.62	394	49.94	31.73	94.75	84.40	78.58	87.70	89.28
94	WC3	134	73.77	19.46	32.09	4300	31.19	85.00	71.20	81.34	88.56	NA	85.35
95	SNK	25	73.63	19.42	24.68	617	52.28	88.10	61.93	83.96	82.97	86.48	59.70
96	W3c	82	73.47	19.38	31.46	2580	34.18	72.80	75.58	73.82	85.34	86.70	85.88
98	lat	47	73.05	19.27	34.98	1644	12.83	78.84	86.30	88.03	81.89	82.08	81.39
99	WLE	87	72.93	19.24	35.53	3091	60.98	52.54	81.58	78.64	85.10	71.77	79.90
100	W43	197	72.76	19.19	39.58	7797	38.87	79.59	NA	79.13	81.86	76.07	81.02
101	WM5	25	72.65	19.17	37.00	925	8.25	82.35	NA	NA	100.00	NA	100.00
102	WVO	297	72.61	19.15	34.23	10167	2.42	77.54	81.57	85.48	86.92	85.50	88.82
103	BU4	20	72.54	19.14	24.65	493	36.68	NA	66.21	85.93	78.67	89.06	78.67
104	WR0	276	72.24	19.06	34.53	9530	19.55	73.86	75.24	80.98	85.35	84.05	86.65
105	SF1	19	72.17	19.04	47.05	894	37.43	76.27	84.85	70.58	81.64	80.35	74.04
106	WI2	15	71.30	18.81	38.73	581	13.61	NA	73.55	83.40	91.15	75.25	90.85
107	205	23	71.03	18.74	31.22	718	37.36	NA	85.74	79.82	78.96	61.56	82.78
108	WB6	12	70.79	18.68	23.92	287	81.51	77.24	70.80	78.34	52.90	78.61	56.16
109	GPS	30	70.56	18.61	23.07	692	59.66	50.00	83.21	86.03	87.76	88.21	39.07
111	osg	75	70.28	18.54	31.75	2381	11.61	72.60	85.83	89.76	72.79	87.18	72.20
112	CHY	42	69.81	18.42	33.69	1415	64.98	NA	75.91	76.05	75.46	71.22	55.23
114	lib	16	68.53	18.08	37.25	596	9.50	79.56	78.80	71.22	82.11	75.84	82.70
115	AM2	17	68.39	18.04	33.53	570	43.15	NA	71.91	66.08	76.96	80.70	71.52
116	220	30	68.10	17.96	30.73	922	41.67	NA	80.33	79.88	66.57	69.64	70.50
117	CO1	25	67.96	17.93	37.60	940	17.56	73.07	78.15	78.04	72.08	72.07	84.74
118	301	316	67.91	17.92	21.40	6761	34.42	77.85	72.25	80.72	74.33	NA	NA
119	214	170	67.34	17.76	19.95	3391	30.78	NA	76.68	81.65	65.19	78.29	71.42

	Site	num	HI	QALY	lifeEx	Pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
120	SAB	25	67.29	17.75	21.80	545	26.63	64.99	60.46	77.10	76.60	90.55	74.66
121	211	13	67.10	17.70	29.15	379	37.15	NA	82.56	73.22	69.70	71.38	68.58
122	109	31	67.00	17.67	26.84	832	22.40	NA	85.72	77.47	72.16	74.24	69.98
123	221	24	66.85	17.63	25.46	611	49.78	NA	59.29	78.78	68.04	71.94	73.25
125	WS4	108	66.47	17.54	29.42	3177	28.43	68.49	42.79	78.26	87.26	80.67	79.41
126	ST6	24	66.04	17.42	25.75	618	27.09	NA	76.80	80.44	69.23	73.93	68.78
127	CRW	73	65.74	17.34	37.49	2737	62.92	NA	63.13	78.05	61.06	63.93	65.38
128	HPK	275	65.61	17.31	34.84	9582	42.71	69.81	68.71	73.18	58.51	70.72	75.62
129	AYA	56	65.44	17.26	36.48	2043	17.19	75.92	59.88	77.11	81.63	NA	80.91
130	sal	51	65.31	17.23	31.63	1613	14.18	71.02	51.35	76.45	78.30	87.86	78.00
131	WIB	101	65.01	17.15	32.17	3249	9.53	75.43	59.07	81.03	83.75	63.93	82.36
132	MR6	31	64.82	17.10	35.16	1090	3.31	NA	84.43	76.35	NA	79.55	80.45
133	CUB	18	64.44	17.00	27.22	490	11.42	80.60	46.63	78.95	74.29	77.98	81.21
134	TL2	163	64.37	16.98	34.60	5639	28.24	57.02	73.33	77.11	69.93	69.92	75.06
135	rea	51	63.85	16.84	33.14	1690	5.85	66.71	73.45	80.94	71.40	79.09	69.51
136	116	34	63.62	16.78	24.00	816	47.62	NA	66.08	83.10	58.52	65.81	60.62
137	110	22	63.46	16.74	26.82	590	12.69	NA	73.58	85.03	67.40	73.79	68.25
139	XCA	35	62.74	16.55	39.17	1371	35.16	58.76	42.34	82.93	71.65	69.82	78.54
140	WC2	16	62.66	16.53	31.25	500	22.04	51.57	NA	NA	88.51	NA	88.51
141	303	115	62.63	16.52	32.57	3746	32.89	NA	62.48	72.63	60.03	79.26	68.52
142	206	22	62.45	16.47	24.27	534	41.73	NA	61.39	77.04	66.05	NA	66.05
144	TL3	11	62.31	16.44	35.91	395	43.10	47.83	61.81	74.69	58.69	74.10	75.95
146	218	15	61.57	16.24	25.53	383	14.84	NA	71.57	78.51	61.71	76.86	65.95
147	DUF	70	61.52	16.23	20.17	1412	49.03	55.45	57.27	78.62	61.78	65.74	62.73
148	SA1	53	61.23	16.15	30.91	1638	6.34	65.08	58.74	74.01	69.84	82.47	72.11
149	CUI	116	60.77	16.03	36.59	4245	19.28	58.21	49.69	79.40	70.54	70.72	77.53
150	115	26	60.17	15.87	20.12	523	29.34	NA	53.85	81.54	60.13	73.20	62.97
152	stt	523	59.44	15.68	23.06	12060	36.03	62.62	50.58	76.41	45.02	63.72	81.67
153	MR1	122	59.32	15.65	25.50	3111	6.11	NA	62.91	83.46	60.63	74.12	68.69
154	TL4	155	59.22	15.62	35.60	5518	22.99	57.10	52.96	73.63	64.10	69.05	74.67
156	FAB	85	58.77	15.50	28.18	2395	49.87	46.58	60.13	70.72	55.22	72.60	56.28
157	BOO	38	58.46	15.42	21.29	809	36.74	57.28	61.19	83.91	54.76	56.95	58.39
159	41D	1157	58.22	15.36	22.58	26130	40.10	38.13	66.86	81.62	48.00	77.85	54.96
160	ESB	426	57.94	15.28	22.15	9436	NA	45.40	48.15	NA	55.85	82.72	57.58
161	KIT	26	57.72	15.23	21.31	554	45.82	56.11	62.60	72.64	57.70	53.98	55.17
162	cry	14	57.24	15.10	30.00	420	52.51	18.91	46.96	62.70	82.63	65.02	71.92
164	SA2	21	57.13	15.07	22.57	474	2.12	59.92	49.86	74.55	66.40	84.62	62.45
167	dol	43	55.88	14.74	25.70	1105	27.33	38.51	39.20	80.44	74.20	50.77	80.74
168	AZ1	118	55.10	14.53	27.08	3195	9.09	NA	64.16	70.75	62.91	60.66	63.00
169	BFT	67	54.93	14.49	34.48	2310	3.79	NA	63.17	75.21	63.94	63.08	60.41
170	SGB	99	54.65	14.42	25.17	2492	9.69	NA	39.60	94.54	55.17	69.69	59.19
171	BUF	101	54.62	14.41	20.71	2092	35.03	51.62	51.91	72.94	56.64	57.21	57.02
172	sfb	17	54.46	14.37	44.24	752	30.87	NA	4.26	NA	91.36	NA	91.36
173	PEA	96	54.13	14.28	20.31	1950	35.35	42.90	60.06	70.94	54.00	60.34	55.35
174	dk2	39	54.02	14.25	14.77	576	23.05	NA	46.69	91.56	49.43	62.10	51.30
175	3C7	35	53.91	14.22	33.74	1181	41.59	26.16	48.95	68.34	59.22	62.22	70.90
176	WW7	149	53.66	14.16	22.24	3314	19.71	NA	57.64	77.72	55.39	58.56	52.95
178	Y1B	29	52.89	13.95	23.17	672	13.88	NA	52.93	NA	61.58	71.79	64.24
179	cop	45	52.82	13.93	47.33	2130	20.44	36.79	22.41	74.36	64.99	74.60	76.14
180	KX1	59	52.26	13.79	18.59	1097	15.30	NA	52.00	85.69	51.92	55.80	52.87
181	MON	121	51.76	13.65	21.95	2656	27.35	50.44	51.71	70.38	51.64	56.40	54.40
182	COY	77	50.81	13.40	29.53	2274	7.93	41.77	61.60	68.35	62.94	52.49	60.58
183	CO2	200	50.26	13.26	23.88	4777	11.74	52.32	53.48	76.10	48.81	53.16	56.22
184	teo	50	49.84	13.15	29.56	1478	17.34	14.34	49.56	81.02	53.78	75.17	57.69
186	STI	55	49.54	13.07	26.35	1449	14.47	42.43	46.51	68.22	58.04	52.49	64.64
187	CHB	361	48.85	12.89	21.32	7697	12.68	32.83	37.93	81.75	49.65	72.61	54.49
188	SPR	74	47.66	12.57	21.51	1592	7.06	54.25	53.25	68.15	48.78	49.93	52.15

	Site	num	HI	QALY	lifeEx	Pyl	Stature	Hypoplasia	Anemia	Dental	Infection	DJD	Trauma
191	BU1	72	46.89	12.37	20.88	1503	22.66	NA	43.22	75.67	44.95	49.31	45.51
192	la8	259	46.77	12.34	22.71	5882	10.28	32.41	21.94	76.84	57.14	70.57	58.20
193	wt1	63	46.63	12.30	25.67	1617	19.72	NA	45.43	75.08	45.18	49.58	44.81
194	co9	239	46.40	12.24	27.77	6637	20.55	16.90	38.28	66.16	44.34	73.84	64.71
195	haw	186	46.08	12.16	22.15	4119	12.99	20.12	28.43	73.95	56.18	70.35	60.54
196	QUI	46	45.99	12.13	25.70	1182	12.03	28.15	59.32	58.14	53.73	51.80	58.76
197	SUN	138	45.11	11.90	17.22	2377	26.00	39.97	40.93	75.18	42.40	45.41	45.90
198	ST5	22	43.41	11.45	22.09	486	19.19	NA	47.36	57.55	42.51	51.78	42.04
199	Y2B	48	41.31	10.90	19.90	955	9.40	NA	33.22	NA	43.85	70.34	49.72
200	3La	80	40.81	10.76	21.14	1691	42.63	7.58	35.60	72.95	37.67	45.72	43.50
201	k10	67	40.24	10.61	17.48	1171	19.71	NA	35.76	69.88	37.86	41.83	36.38
203	DW2	72	39.05	10.30	11.53	830	19.77	NA	41.74	66.33	32.88	39.05	34.53
204	J73	56	38.78	10.23	17.96	1006	12.56	30.58	24.20	77.13	36.73	46.14	44.11
205	gua	25	38.57	10.17	11.04	276	4.69	36.69	6.52	81.85	30.93	78.99	30.32
206	J74	49	38.39	10.13	22.59	1107	6.49	25.50	25.86	71.00	39.96	54.04	45.88