

Supplementary Information 1: Lipid residues in pottery from the Indus Civilisation in northwest India

Suryanarayan et al. 2020

#Download data files, change source location #(e.g. 'C:/Users/pc/Dropbox/Dropbox/R') #to the name of the folder where you have saved the data files

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

Load packages

```
library(readr)
library(ade4)
library(ggplot2)
library(knitr)
library(tidyr)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##   smiths
```

```
library(RColorBrewer)
library(ggpubr)
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:readr':
##
##   col_factor
```

```
library(ggrepel)
library(Cairo)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggpubr':
##
##   get_legend
```

```
#Load data
```

```
data <- read_csv("PhD/Data/data.csv")
```

```
##
## -- Column specification -----
## cols(
##   .default = col_character(),
##   Rim_diam_cm = col_double(),
##   Sample_mass_mg = col_double(),
##   Area_GC_total = col_double(),
##   AreaC12 = col_double(),
##   AreaC14 = col_double(),
##   AreaC15 = col_double(),
##   AreaC161 = col_double(),
##   AreaC16 = col_double(),
##   AreaC17 = col_double(),
##   AreaC181 = col_double(),
##   AreaC18 = col_double(),
##   AreaC19 = col_double(),
##   AreaC201 = col_double(),
##   AreaC20 = col_double(),
##   AreaC21 = col_double(),
##   AreaC221 = col_double(),
##   AreaC22 = col_double(),
##   AreaC23 = col_double(),
##   AreaC24 = col_double(),
##   AreaC25 = col_double()
##   # ... with 24 more columns
## )
## i Use 'spec()' for the full column specifications.
```

```
save(data, file = "data.RData")
```

```
#convert data to factors
```

```
data$Site_name<- factor(data$Site_name)
data$Vessel_type <- factor(data$Vessel_type)
data$Rural_urban <-factor(data$Rural_urban)
data$Lipidconc_ug_g <- as.numeric(data$Lipidconc_ug_g)
```

```
#Prepare data #subset to NW India settlements
```

```
indus <- droplevels(subset(data, (Site_name=="RGR") | (Site_name=="MSDI") |  
                             (Site_name=="LHRI") | (Site_name=="MSDVII") |  
                             (Site_name=="ALM") |  
                             (Site_name=="FRN") | (Site_name=="KNK"))) )  
isotope_indus <- droplevels(subset(indus, !is.na(indus$delta13C_C16)))  
  
#Set order for site names  
isotope_indus$Site_name <- factor(isotope_indus$Site_name,  
                                levels = c("ALM", "MSDVII",  
                                           "MSDI", "LHRI",  
                                           "KNK", "FRN", "RGR"))
```

```
#convert variables to factors
```

```
indus$Vessel_type <- factor(indus$Vessel_type)  
indus$Site_name <- factor(indus$Site_name)  
isotope_indus$Vessel_type <- factor(isotope_indus$Vessel_type)  
isotope_indus$Site_name <- factor(isotope_indus$Site_name)  
isotope_indus$Chronology_details <- factor(isotope_indus$Chronology_details)  
isotope_indus$Vessel_form <- factor(isotope_indus$Vessel_form)
```

```
#Kruskal Wallis tests
```

```
#log of lipid concentrations  
indus$lipid_log <- log10(indus$Lipidconc_ug_g)  
  
#Kruskal-Wallis test for lipid concentrations across sites  
kruskal.test(lipid_log~Site_name,data= indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: lipid_log by Site_name  
## Kruskal-Wallis chi-squared = 11.799, df = 6, p-value = 0.06662
```

```
#kruskal wallis test for lipid concentration across vessel form  
kruskal.test(lipid_log~Vessel_type, data = indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: lipid_log by Vessel_type  
## Kruskal-Wallis chi-squared = 11.239, df = 10, p-value = 0.3392
```

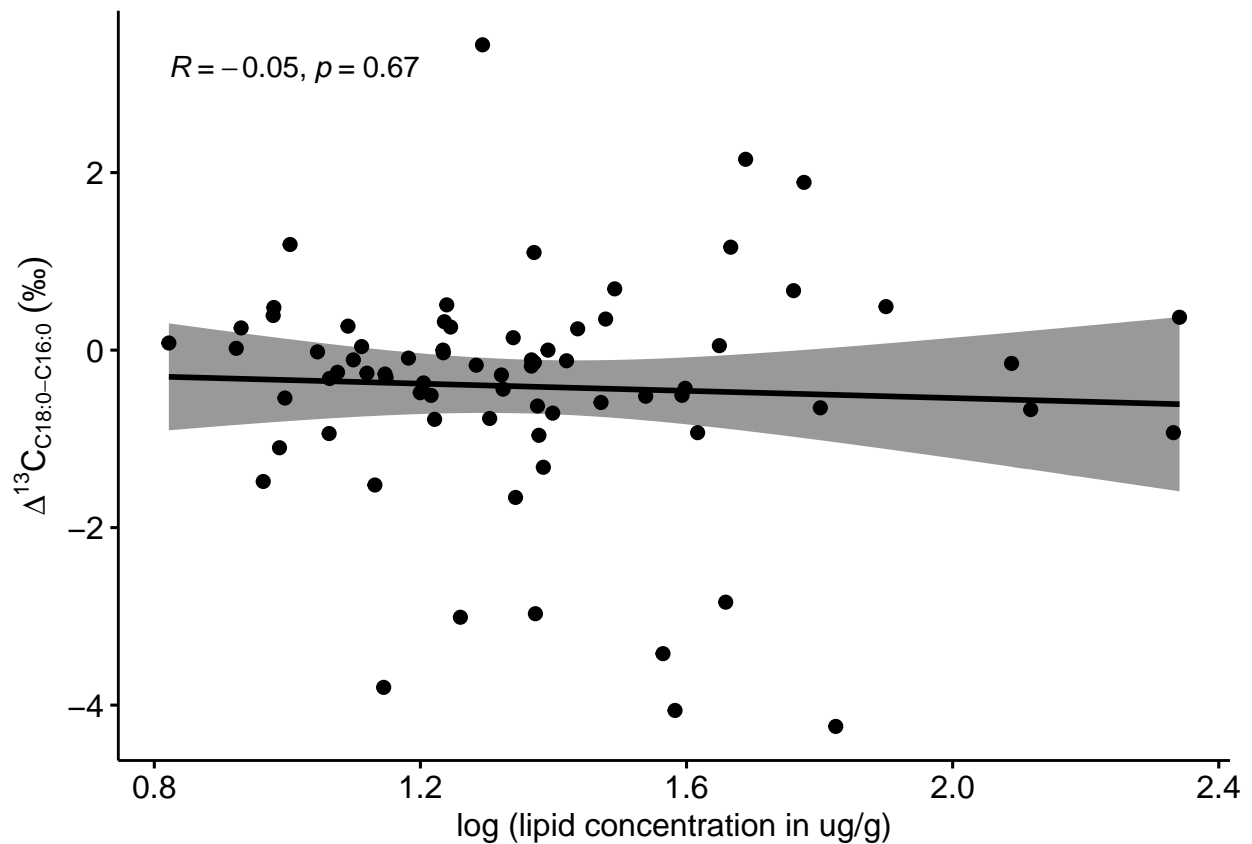
```
#Figure 1: Relationship between lipid yield and big delta
```

```
#convert lipid concentration and big delta to numeric values  
isotope_indus$Lipidconc_ug_g <- as.numeric(isotope_indus$Lipidconc_ug_g)  
isotope_indus$bigdelta <- as.numeric(isotope_indus$bigdelta)  
  
#convert lipid concentration to log values  
isotope_indus$lipidlog <- log10(isotope_indus$Lipidconc_ug_g)
```

#Figure 1: Correlation between D13C values and lipid yields

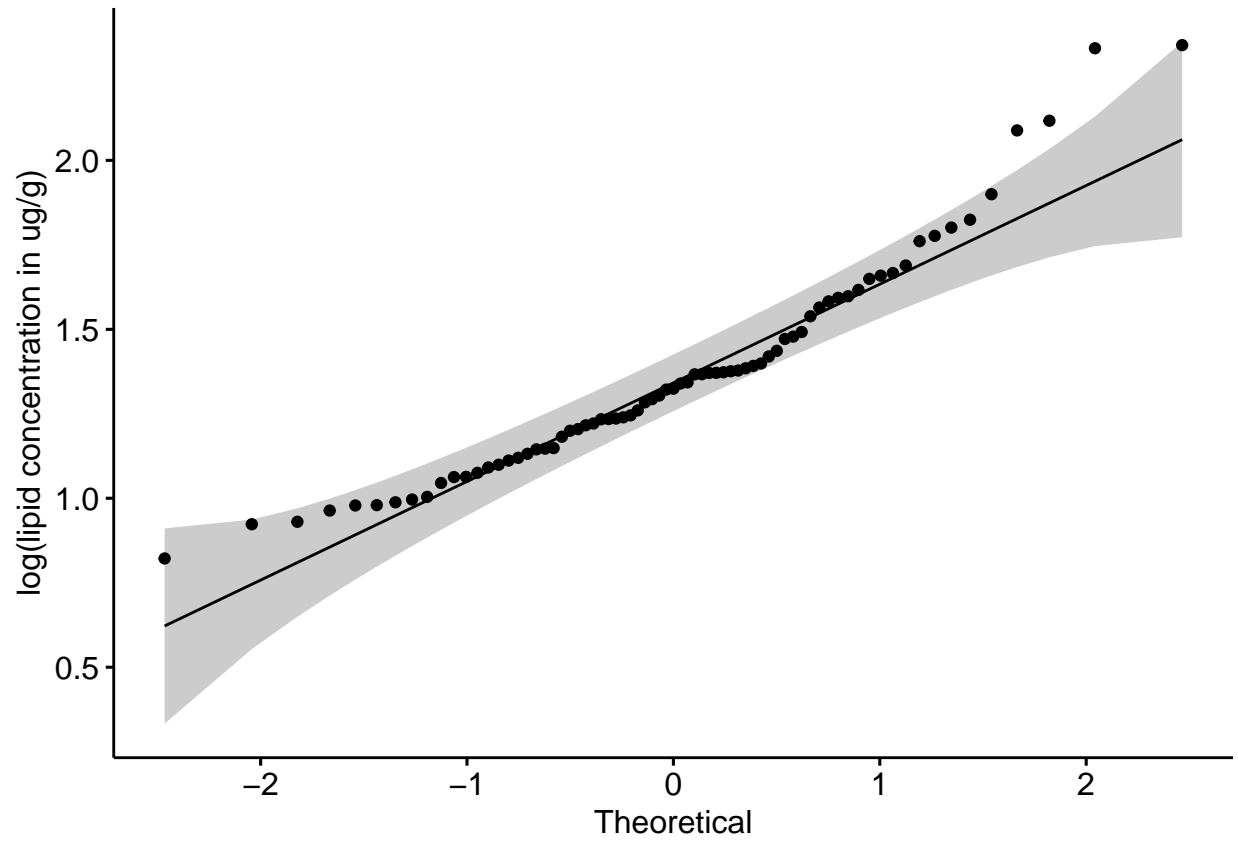
```
fig1 <- ggscatter(isotope_indus, x = "lipidlog",  
                 y = "bigdelta",  
                 add = "reg.line", conf.int = TRUE,  
                 cor.coef = TRUE, cor.method = "pearson",  
                 xlab = "log (lipid concentration in ug/g)",  
                 ylab = expression(Delta13* C [C18:0-C16:0]* " "("\u2030"))  
fig1
```

'geom_smooth()' using formula 'y ~ x'

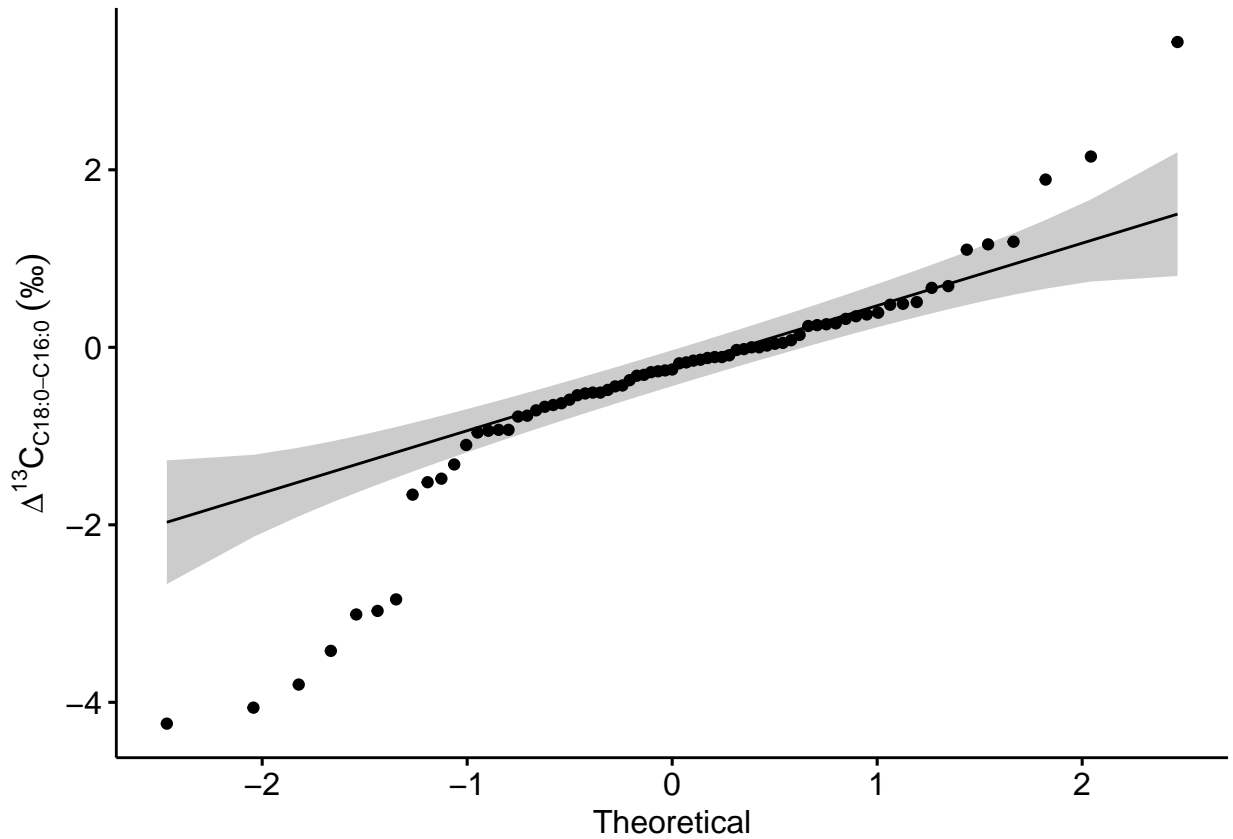


#Figure 1 continued

```
ggqqplot(isotope_indus$lipidlog, ylab = "log(lipid concentration in ug/g)")
```



```
# big delta  
ggqqplot(isotope_indus$bigdelta,  
          ylab = expression(Delta13* C [C18:0-C16:0]* " ("\"u2030\""))
```



```
res <- cor.test(isotope_indus$lipidlog, isotope_indus$bigdelta,
               method = "pearson")
res
```

```
##
## Pearson's product-moment correlation
##
## data: isotope_indus$lipidlog and isotope_indus$bigdelta
## t = -0.42221, df = 71, p-value = 0.6742
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.2769228 0.1821203
## sample estimates:
##          cor
## -0.05004383
```

#Prepare mixing curves for Figure 2

```
#Values for fatty acids in sources
plantC4_I_16 <- (-18)
plantC4_I_18 <- (-18)

rumC4_I_16 <- (-18)
rumC4_I_18 <- (-20)
```

```

dairyC4_I_16 <- (-18)
dairyC4_I_18 <- (-23)

plantC3_I_16 <- (-30)
plantC3_I_18 <- (-30)

rumC3_I_16 <- (-30)
rumC3_I_18 <- (-32)

dairyC3_I_16 <- (-30)
dairyC3_I_18 <- (-35)

nonrumC3_I_16 <- (-30)
nonrumC3_I_18 <- (-30)

#Ranges for FA concentrations in ruminant
plantC4_C_16 <- (0.8)
plantC4_C_18 <- (0.2)

rumC4_C_16 <- (0.6)
rumC4_C_18 <- (0.4)

dairyC4_C_16 <- (0.6)
dairyC4_C_18 <- (0.4)

rumC3_C_16 <- (0.6)
rumC3_C_18 <- (0.4)

plantC3_C_16 <- (0.8)
plantC3_C_18 <- (0.2)

dairyC3_C_16 <- (0.6)
dairyC3_C_18 <- (0.4)

nonrumC3_C_16 <- (0.5)
nonrumC3_C_18 <- (0.5)

#C3rum vs C4plant

A_I_16 <- rumC3_I_16
A_C_16 <- rumC3_C_16
A_I_18 <- rumC3_I_18
A_C_18 <- rumC3_C_18

B_I_16 <- plantC4_I_16
B_C_16 <- plantC4_C_16
B_I_18 <- plantC4_I_18
B_C_18 <- plantC4_C_18

#specify the A contribution
contribution_A <- seq(0, 1, length.out = 11)
#the B contribution is A -1

```

```

contribution_B<- 1-contribution_A

#calculate mixed 16:0 value
final_16 <- (contribution_A*A_I_16*A_C_16 +
             contribution_B*B_I_16*B_C_16)/(contribution_A*A_C_16 + contribution_B*B_C_16)

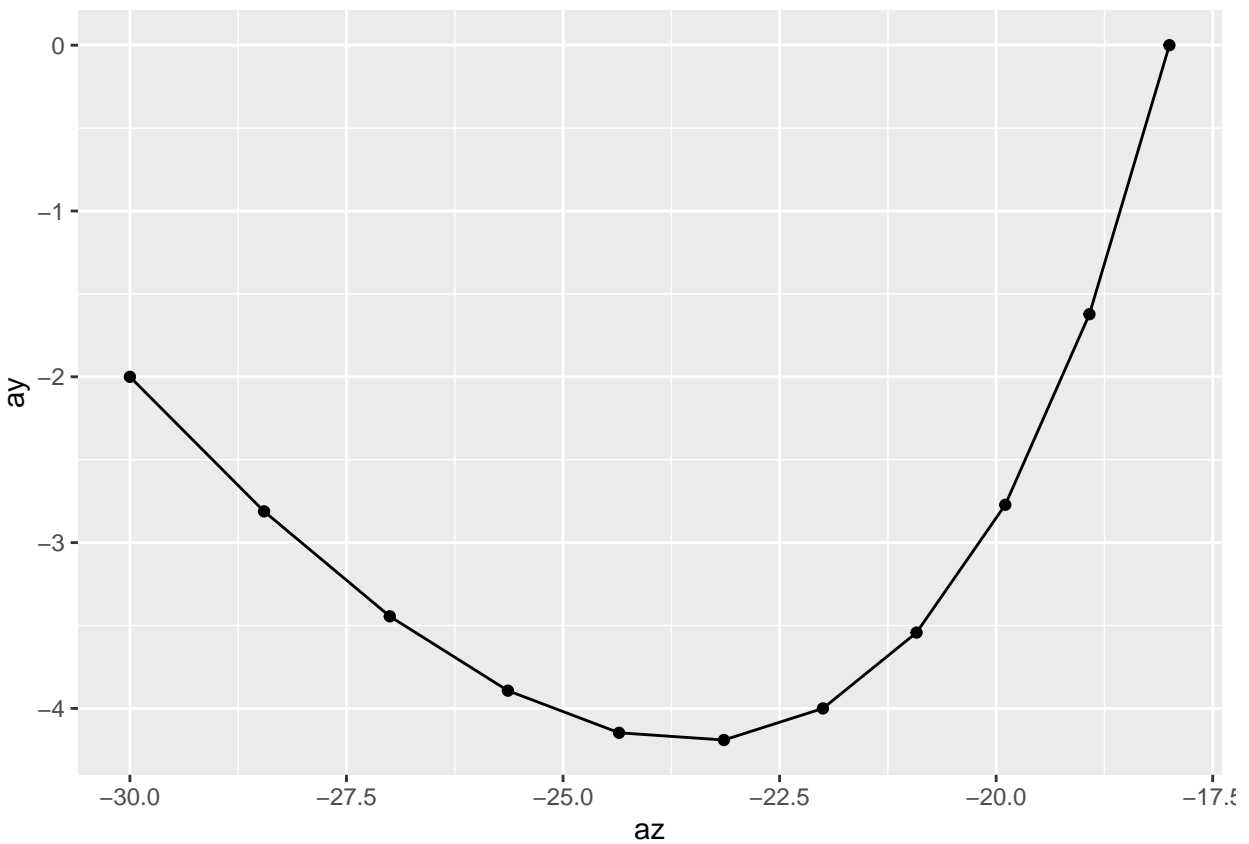
#calculate mixed 18:0 value
final_18 <- (contribution_A*A_I_18*A_C_18 +
             contribution_B*B_I_18*B_C_18)/(contribution_A*A_C_18 + contribution_B*B_C_18)

#calculate big delta
big_delta <- final_18 - final_16

ax <- contribution_A
ay <- big_delta
az <- final_16
aa <- final_18

#Plot mixing curve #C3rum vs C4dairy
ggplot() + geom_line(aes(x=az, y=ay)) +
  geom_point(aes(x=az, y=ay))

```



#Prepare mixing curves for Figure 2


```

#C3nonrum vs C4plant
A_I_16 <- nonrumC3_I_16
A_C_16 <- nonrumC3_C_16
A_I_18 <- nonrumC3_I_18
A_C_18 <- nonrumC3_C_18

B_I_16 <- plantC4_I_16
B_C_16 <- plantC4_C_16
B_I_18 <- plantC4_I_18
B_C_18 <- plantC4_C_18

#specify the A contribution
contribution_A <- seq(0, 1, length.out = 11)
#the B contribution is A -1
contribution_B<- 1-contribution_A

#calculate mixed 16:0 value
final_16 <- (contribution_A*A_I_16*A_C_16 +
             contribution_B*B_I_16*B_C_16)/(contribution_A*A_C_16 + contribution_B*B_C_16)

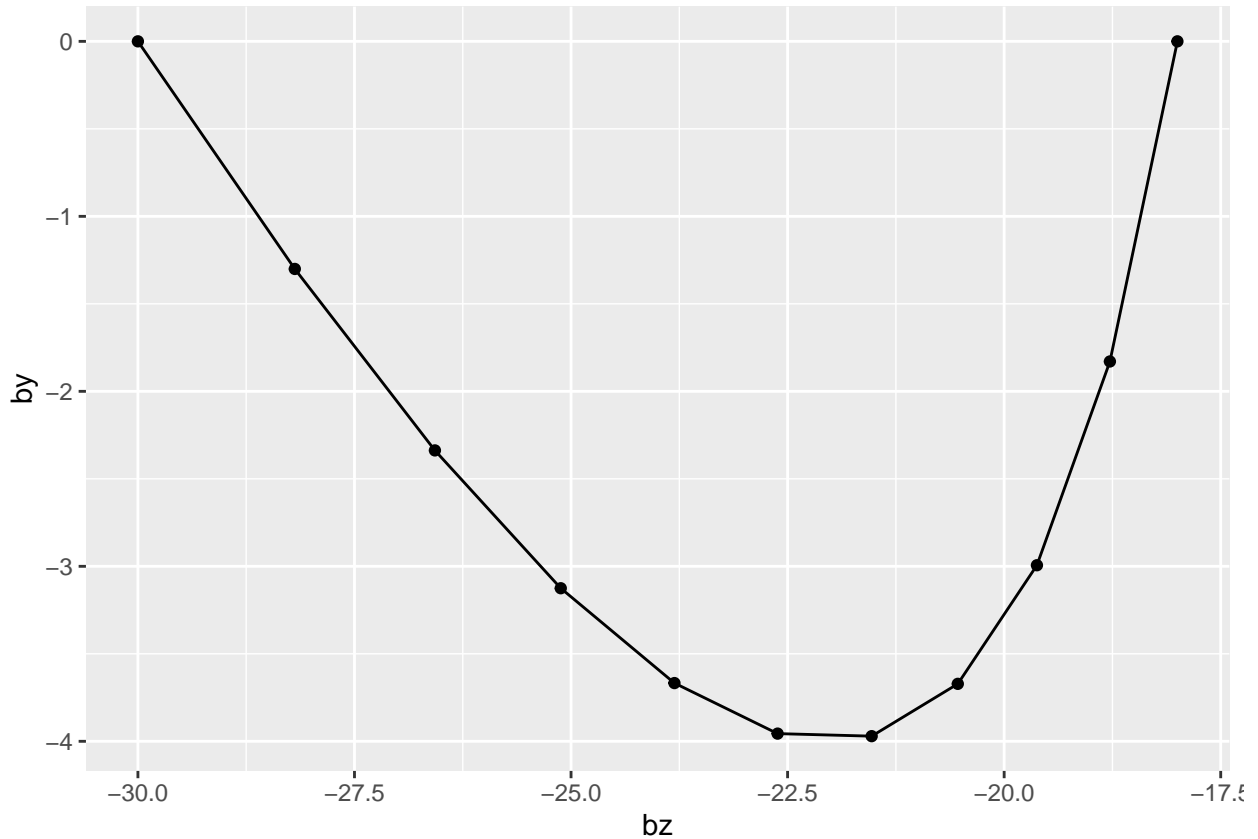
#calculate mixed 18:0 value
final_18 <- (contribution_A*A_I_18*A_C_18 +
             contribution_B*B_I_18*B_C_18)/(contribution_A*A_C_18 + contribution_B*B_C_18)

#calculate big delta
big_delta <- final_18 - final_16

bx <- contribution_A
by <- big_delta
bz <- final_16
ba <- final_18

#Plot mixing curve #C3non-rum vs C4plant
ggplot() + geom_line(aes(x=bz, y=by)) +
  geom_point(aes(x=bz, y=by))

```



#Prepare mixing curves for Figure 2

```

#C3dairy vs C4plant
A_I_16 <- dairyC3_I_16
A_C_16 <- dairyC3_C_16
A_I_18 <- dairyC3_I_18
A_C_18 <- dairyC3_C_18

B_I_16 <- plantC4_I_16
B_C_16 <- plantC4_C_16
B_I_18 <- plantC4_I_18
B_C_18 <- plantC4_C_18

#specify the A contribution
contribution_A <- seq(0, 1, length.out = 11)
#the B contribution is A -1
contribution_B <- 1 - contribution_A

#calculate mixed 16:0 value
final_16 <- (contribution_A * A_I_16 * A_C_16 +
             contribution_B * B_I_16 * B_C_16) / (contribution_A * A_C_16 + contribution_B * B_C_16)

#calculate mixed 18:0 value
final_18 <- (contribution_A * A_I_18 * A_C_18 +
             contribution_B * B_I_18 * B_C_18) / (contribution_A * A_C_18 + contribution_B * B_C_18)

```

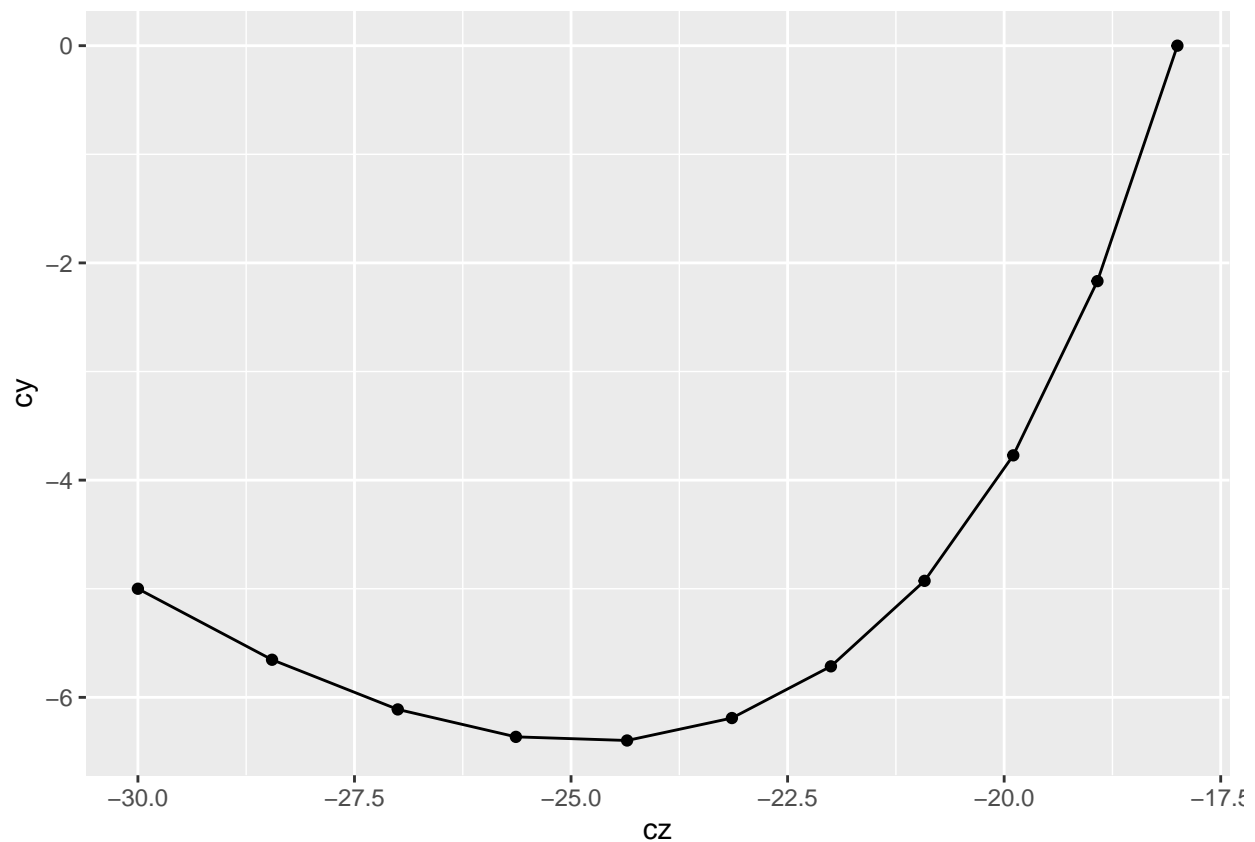
```

#calculate big delta
big_delta <- final_18 - final_16

cx <- contribution_A
cy <- big_delta
cz <- final_16
ca <- final_18

#Plot mixing curve #C3non-rum vs C4plant
ggplot() + geom_line(aes(x=cz, y=cy)) +
  geom_point(aes(x=cz, y=cy))

```



#Figure 2: Mixing curves

```

p2 <-ggplot() +
  geom_line(aes(x=az, y=ay), linetype = "dashed") +
  geom_line(aes(x=bz, y=by), linetype = "dashed") +
  geom_line(aes(x=cz, y=cy), linetype = "dashed") +
  geom_point (aes(x=az, y=ay)) +
  geom_point (aes(x=bz, y=by)) +
  geom_point (aes(x=cz, y=cy)) +
  labs(x=expression(delta^{13}*C[16:0]*"(\u2030)"),
       y=expression(Delta^{13}*C*" (\u2030)))+
  scale_y_continuous(position = "right", limits=c(-8,4))+

```

```

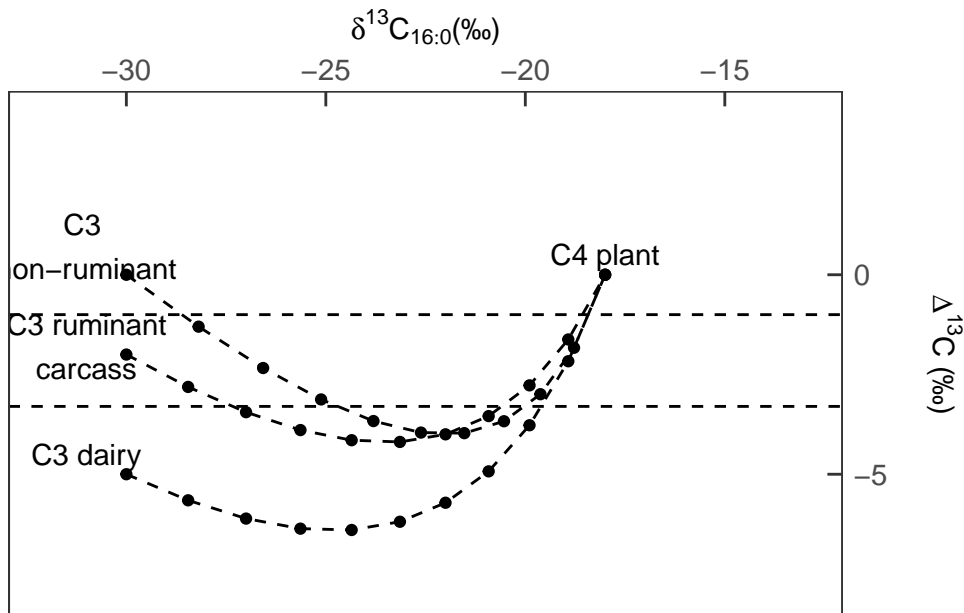
scale_x_continuous(position="top", limits=c(-32,-13))+
theme_bw() +
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()+
theme(axis.ticks.length=unit(-0.2,"cm"))+
theme(plot.title=element_text(hjust=0.5))+
theme(axis.text=element_text(size=11),
axis.title = element_text(size=11),
axis.text.x.top = element_text(margin = margin(b = 10)),
axis.text.y.right = element_text(margin = margin(l = 10)),
axis.title.x = element_text(margin = margin(t=10)),
axis.title.y.right = element_text(margin = margin(l=15)),
title = element_text(size=11),
legend.title = element_blank()+
geom_hline(yintercept = -1, linetype="dashed")+
geom_hline(yintercept = -3.3, linetype="dashed")+
theme(text=element_text()+
annotate("text", x=-31, y=-4.5, label="C3 dairy", size=4)+
annotate("text", x=-31, y=-1.8, label="C3 ruminant\ncarcass", size=4)+
annotate("text", x=-31, y=0.7, label="C3 \nnon-ruminant", size=4)+
annotate("text", x=-18, y=0.5, label="C4 plant", size=4)+
coord_fixed(ratio = 1)
p2 + ggtitle("Figure S2", subtitle = "Hypothetical mixing lines between different animal fats and a C4 plant
NB. This figure is illustrative; both the stable isotope and concentration values of fatty acids in each
theme (plot.subtitle = element_text(size =6)) + theme(plot.title = element_text(size = 8))

```

Figure S2

Hypothetical mixing lines between different animal fats and a C4 plant oil. 10% incremental contributions of C4 plant oil to the hypothetical mix are shown by points. The hypothetical $\delta^{13}\text{C}_{16:0}$ and $\delta^{13}\text{C}_{18:0}$ values for each source are: C3 non-ruminants (-30‰, -30‰), C3 ruminant carcass (-30‰, -32‰), C3 dairy (-30‰, -35‰), C4 plant (-18‰, -18‰). The relative amounts of C16:0 and C18:0 in each source are: C3 non-ruminants (1:1), C3 ruminant carcass (3:2), C3 dairy (3:2), C4 plant (4:1) and are approximated from data reported in the USDA database (<https://fdc.nal.usda.gov/>). The range limits typically used to discriminate dairy, ruminant carcass and non-ruminant products from their ^{13}C are shown (horizontal dashed line).

NB. This figure is illustrative; both the stable isotope and concentration values of fatty acids in each source vary according to local environmental variables, husbandry practices and/or growing conditions and ideally need to be defined according to the context.



#Kruskal-Wallis tests testing effect of settlement type on compound-specific isotopic values

```
kruskal.test(delta13C_C16~Rural_urban, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: delta13C_C16 by Rural_urban  
## Kruskal-Wallis chi-squared = 5.945, df = 1, p-value = 0.01476
```

```
kruskal.test(delta13C_C18~Rural_urban, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: delta13C_C18 by Rural_urban  
## Kruskal-Wallis chi-squared = 5.2806, df = 1, p-value = 0.02156
```

```
kruskal.test(bigdelta~Rural_urban, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: bigdelta by Rural_urban  
## Kruskal-Wallis chi-squared = 0.21381, df = 1, p-value = 0.6438
```

```
#Kruskal-wallis tests testing effect of time period on compound-specific isotopic values
```

```
kruskal.test(delta13C_C16~Chronology_details, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: delta13C_C16 by Chronology_details  
## Kruskal-Wallis chi-squared = 4.2038, df = 2, p-value = 0.1222
```

```
kruskal.test(delta13C_C18~Chronology_details, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: delta13C_C18 by Chronology_details  
## Kruskal-Wallis chi-squared = 4.4388, df = 2, p-value = 0.1087
```

```
kruskal.test(bigdelta~Chronology_details, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: bigdelta by Chronology_details  
## Kruskal-Wallis chi-squared = 0.77642, df = 2, p-value = 0.6783
```

```
#Kruskal-Wallis tests testing effect of vessel form on compound-specific isotopic values
```

```
kruskal.test(delta13C_C16~Vessel_form, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: delta13C_C16 by Vessel_form  
## Kruskal-Wallis chi-squared = 10.416, df = 7, p-value = 0.1662
```

```
kruskal.test(delta13C_C18~Vessel_form, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: delta13C_C18 by Vessel_form  
## Kruskal-Wallis chi-squared = 10.138, df = 7, p-value = 0.1809
```

```
kruskal.test(bigdelta~Vessel_form, data = isotope_indus)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: bigdelta by Vessel_form  
## Kruskal-Wallis chi-squared = 6.4045, df = 7, p-value = 0.4934
```

#Figure 3: Comparison of GC-C-IRMS results from before and after 4.2 ka BP ##Prepare data

```
#change column names  
names(isotope_indus)[names(isotope_indus) == "delta13C_C16"] <- "C16"  
names(isotope_indus)[names(isotope_indus) == "delta13C_C18"] <- "C18"  
  
#reshape data for boxplots  
isotope_reshape<- gather(data = isotope_indus, key = d13C, value = d13C_value, C16:C18)  
droplevels(isotope_reshape)
```

```
## # A tibble: 146 x 90  
##   Site_name Sample Rural_urban Trench_Context Context_Notes Chronology  
##   <fct>      <chr> <fct>      <chr>          <chr>          <chr>  
## 1 ALM      ALM11~ Rural      SC-114        114            Late_Hara~  
## 2 ALM      ALM11~ Rural      SC-117        117            Late_Hara~  
## 3 ALM      ALM11~ Rural      SC-117        117            Late_Hara~  
## 4 ALM      ALM11~ Rural      SC-119        119            Late_Hara~  
## 5 ALM      ALM12~ Rural      SC-122        122            Late_Hara~  
## 6 ALM      ALM12~ Rural      SC-124        124            Late_Hara~  
## 7 ALM      ALM12~ Rural      SC-125        125            Late_Hara~  
## 8 ALM      ALM12~ Rural      SC-125        125            Late_Hara~  
## 9 ALM      ALM12~ Rural      SC-125        125            Late_Hara~  
## 10 ALM     ALM12~ Rural      SC-125        125            Late_Hara~  
## # ... with 136 more rows, and 84 more variables: Before_After_4.2_kya <chr>,  
## #   Chronology_details <fct>, Artefact_type <chr>, Rim_base_body <chr>,  
## #   Rim_diam_cm <dbl>, Vesseltype <chr>, Vessel_form <fct>, Vessel_type <fct>,  
## #   Vessel_category <chr>, Haryana_or_Classic <chr>, Saturated <chr>,
```

```
## # Unsaturated <chr>, Branched <chr>, Date_of_analysis <chr>,
## # Sample_mass_mg <dbl>, Area_GC_total <dbl>, AreaC12 <dbl>, AreaC14 <dbl>,
## # AreaC15 <dbl>, AreaC161 <dbl>, AreaC16 <dbl>, AreaC17 <dbl>,
## # AreaC181 <dbl>, AreaC18 <dbl>, AreaC19 <dbl>, AreaC201 <dbl>,
## # AreaC20 <dbl>, AreaC21 <dbl>, AreaC221 <dbl>, AreaC22 <dbl>, AreaC23 <dbl>,
## # AreaC24 <dbl>, AreaC25 <dbl>, AreaC26 <dbl>, AreaC28 <dbl>,
## # Areaphthalates <dbl>, FAMEarea <dbl>, Area_IS_C36 <dbl>, Area_IS_C34 <dbl>,
## # Mass_IS_C34_ug <dbl>, Mass_IS_C36_ug <dbl>, Lipidconc_ug_vial <dbl>,
## # Lipidconc_ug_g <dbl>, Interpretable <lgl>, C12concentration_ug_vial <chr>,
## # C14concentration_ug_vial <chr>, C15concentration_ug_vial <chr>,
## # C161concentration_ug_vial <chr>, C16concentration_ug_vial <dbl>,
## # C17concentration_ug_vial <chr>, C181concentration_ug_vial <chr>,
## # C18concentration_ug_vial <dbl>, C19concentration_ug_vial <chr>,
## # C201concentration <chr>, C20concentration_ug_vial <chr>,
## # C21concentration_ug_vial <chr>, C221concentration <chr>, 'C22concentration
## # (ug/vial)' <chr>, C23concentration_ug_vial <chr>,
## # C24concentration_ug_vial <chr>, C25concentration_ugvial <chr>,
## # C26concentration_ug_vial <chr>, C28concentration_ug_vial <chr>,
## # C34concentration_ug_vial <chr>, FAMEconcentration <dbl>,
## # Cholesterol_derivatives <chr>, 'longchain_Alkanes' <chr>, Alcohols <chr>,
## # Sulphur <chr>, qty_C16_inj_dil_100ul_ng <dbl>,
## # qty_C18_inj_dil_100ul_ng <dbl>, C16conc_ug_vial <dbl>,
## # C18conc_ug_vial <dbl>, dilution_ul <dbl>, qty_C16_inj_ng <chr>,
## # qty_C18_inj_ng <chr>, PS_ratio <dbl>, C12_C14 <dbl>, C161_C181 <chr>,
## # C15_C17_C18 <chr>, bigdelta <dbl>, lipidlog <dbl>, d13C <chr>,
## # d13C_value <dbl>
```

```
#reorder names of sites
isotope_reshape$Site_name <- factor(isotope_reshape$Site_name,
                                   levels = c("ALM", "MSDVII", "MSDI", "LHRI", "KNK",
                                              "FRN", "RGR"))
```

```
#subset for groups that has evidence pre and post 4.2 ka
climate <- droplevels(subset(isotope_reshape, (Site_name=="ALM") |
                             (Site_name=="MSDVII")))
```

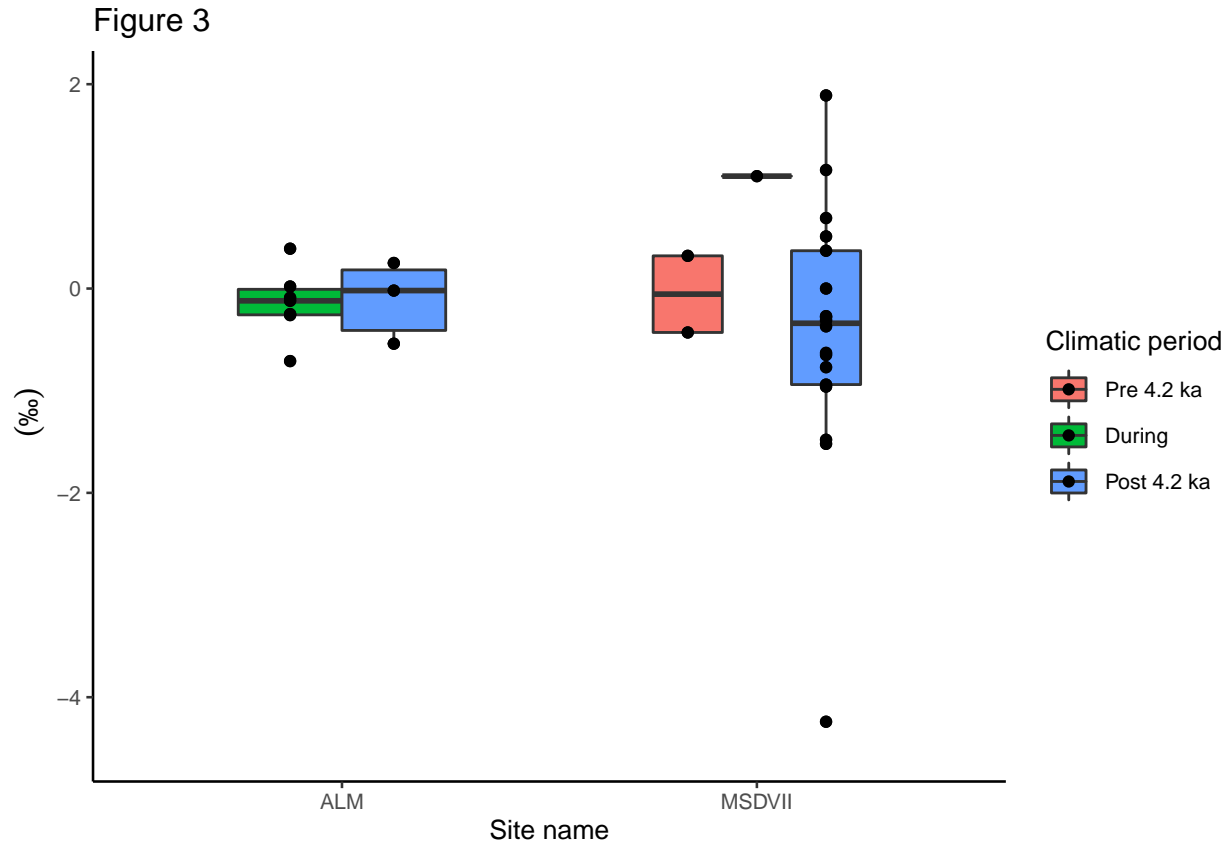
```
#set order of climatic period
climate$Before_After_4.2_kya <- factor(climate$Before_After_4.2_kya,
                                       levels=c("Pre 4.2 ka", "During", "Post 4.2 ka"))
```

#Figure 3

```
fig3 <- ggplot(climate, aes(x = Site_name, y = bigdelta, fill = Before_After_4.2_kya))+
  theme(plot.title = element_text(hjust=0, size=16)) +
  labs(x = "Site name",
       y = expression(" "("\u2030")), fill = "Climatic period" ) +
  ggtitle("Figure 3") +
  geom_boxplot(aes(fill = Before_After_4.2_kya),
              position = position_dodge(0.5), width = .5,
              outlier.shape = NA) +
  geom_point(aes(y = bigdelta, group = Before_After_4.2_kya),
            position = position_dodge(width=0.5)) +
  theme_classic(base_size = 10) +
```

```
scale_y_continuous(breaks=c(-4,-2, 0, 2)) +
expand_limits(y = c(-4.5, 2)) +
theme (axis.text.x = element_text())
```

fig3



#Figure 4: Lipid concentration across vessel forms

```
fig4 <- ggplot(data = indus,
               aes(x = Vessel_type, y = Lipidconc_ug_g)) +
  geom_boxplot(aes(x=Vessel_type, y=Lipidconc_ug_g),
              position = position_dodge(0.5), width = .2, outlier.shape = NA) +
  scale_y_log10(breaks = c(5, 10, 20, 30, 50, 100, 200),
              limits = c(2, 300)) +
  xlab("Vessel form") + ylab(expression("Lipid concentration (ug/g)")) +
  ggtitle("Figure 4") +
  stat_summary(fun=mean, geom="point", shape=18, size=3, color="red")+
  geom_jitter(aes(y = Lipidconc_ug_g), position = position_jitter(0.1)) +
  scale_x_discrete(labels = paste(levels(indus$Vessel_type),
                                "\n(N=", table(indus$Vessel_type), ")", sep="")) +
  theme_classic(base_size = 10) + theme(axis.text.x=element_text(angle=90,hjust=1,
                                                                    vjust=0.5))
```

fig4

Figure 4

