

data processing for charts.R

jfinn — Oct 10, 2014, 10:29 AM

```
library(reshape2)
library(plyr)
library(XLConnect)
```

```
XLConnect 0.2-7 by Mirai Solutions GmbH
http://www.mirai-solutions.com ,
http://miraisolutions.wordpress.com
```

```
options(java.parameters = "-Xmx1024m" )

##### import data #####

setwd('P:/19251/Monitoring Results/Analysis Data')
wb1<-loadWorkbook('cleaned dataset.xlsx', create=TRUE)

soil<-readWorksheet(wb1, sheet="final.dataset")
colnames(soil)<-tolower(colnames(soil))
soil$medium<- "soil"

wipe<-readWorksheet(wb1, sheet="wipe.dataset")
colnames(wipe)<-tolower(colnames(wipe))
wipe$medium<- "wipe"
# need to convert EPA wipe results to ug/Wipe (mg/Wipe for metals, ng/Wipe for PAH)
wipe$result<-ifelse(wipe$collection.by=="EPA - Wipes",
                    ifelse(wipe$units=="ng/Wipe", wipe$result/1000, wipe$result*1000), wipe$result)

all.results<-rbind(soil, wipe)
# rename sample locations
all.results$location<-ifelse(all.results$location=="KCBX North - Coal" |
all.results$location=="KCBX South - Coal",
                            "Coal", ifelse(all.results$location=="KCBX North - Petcoke" |
all.results$location=="KCBX South - Petcoke",
                                            "Petcoke", all.results$location))
all.results$location<-ifelse(all.results$location=="Abutting",
                            ifelse(all.results$medium=="wipe",
                                    ifelse(all.results$collection.by=="EPA - Wipes", "Abutting (EPA
wipes)", "Abutting (WB wipes)"),
                                all.results$location), all.results$location)
all.results$location<-ifelse(all.results$location=="Control",
                            ifelse(all.results$medium=="wipe", "Control (WB wipes)", "Control
(soil)"),
                                all.results$location)
all.results$location<-ifelse(all.results$location=="Illinois Metro", "TACO
(soil)", all.results$location)
all.results$location<-ifelse(all.results$location=="USGS", "USGS (soil)", all.results$location)

##### PAH Profiles #####

pah<-subset(all.results, all.results$group=="PNA" & all.results$qa=="S")
# name correction
pah$parameter<-ifelse(pah$parameter=="Benzo (a) anthracene", "Benz(a)anthracene", pah$parameter)
# select only pah we've included in the profiles
parameter<-as.data.frame(c('2-Methylnaphthalene', 'Naphthalene',
                            'Acenaphthylene', 'Acenaphthene', 'Fluorene',
                            'Anthracene', 'Phenanthrene', 'Fluoranthene', 'Pyrene',
                            'Benz(a)anthracene',
                            'Chrysene', 'Benzo(a)pyrene', 'Benzo(b)fluoranthene',
                            'Benzo(k)fluoranthene', 'Benzo(g,h,i)perylene', 'Indeno(1,2,3-cd)pyrene',
                            'Dibenz(a,h)anthracene'))
param.rank<-as.data.frame(parameter)
```

```

col names( param. rank) <- c( "parameter")
pah<- merge( pah, param. rank, by=c( "parameter"))
pah<- pah[ order( pah$collection. by) , ]

# remove Test America petcoke samples
pah<- subset( pah, pah$collection. by!="Test America")

# samples with no PAHs detected
# 110 total samples with no PAHs detected: 1 SE Chicago soil, 1 Coal, 26 SE Chicago wipes,
# 82 S Deering/East Side wipes
# nd 1 = detect, nd 0 = non-detect
pah$nd<- ifelse( pah$location=="TACO (soil)", 1, pah$nd)
pah.nd<- ddply( pah, . ( location, medium, collection. by, full. id ), summarise,
               total=sum( !is. na( nd) , na. rm=T) ,
               nd=sum( nd, na. rm=T) )
pah.nd<- subset( pah.nd, pah.nd$nd!=0, select=c( "full. id", "medium"))
pah.detect<- merge( pah, pah.nd, by=c( "full. id", "medium"))

pah.summary<- ddply( pah.detect, . ( location, medium, parameter) , summarise,
                    n=length( result) ,
                    dt=sum( nd) ,
                    median=median( result, na. rm=TRUE) )
pah.totals<- ddply( pah.summary, . ( location, medium) , summarise,
                  total=sum( median, na. rm=TRUE) )
pah.plot<- merge( pah.summary, pah.totals, by=c( "location", "medium"))
pah.plot$percent<- pah.plot$median/pah.plot$total

##### V : Ni Ratio #####
# format v:ni ratio data
inorganic<- subset( all.results, all.results$group=="Inorganic" & all.results$qa=="S")
v.n.ratio<- subset( inorganic, inorganic$parameter=="Vanadium" | inorganic$parameter=="Nickel",

select=c( "full. id", "result", "nd", "parameter", "location", "medium", "collection. by"))
# remove Test America petcoke sample (not correlated) and WB abutting/control wipe samples (ND)
v.n.ratio<- subset( v.n.ratio, v.n.ratio$collection. by!="Test America" &
                  v.n.ratio$location!="Abutting (WB wipes)" & v.n.ratio$location!="Control (WB
wipes)")

nd.summary<- ddply( v.n.ratio, . ( location, collection. by, medium, parameter) , summarise,
                  total=sum( !is. na( nd) , na. rm=T) ,
                  nd=sum( nd, na. rm=T) )
nd.summary<- reshape( nd.summary, timevar = "parameter",
                     idvar = c( "location", "medium", "collection. by"), direction = 'wide')

# select samples with both V, Ni detect only, drop nd column (detect indicator)
v.n.ratio$nd<- ifelse( v.n.ratio$location=="TACO (soil)", 1, v.n.ratio$nd)
v.n.ratio$full. id<- ifelse( v.n.ratio$location=="TACO (soil)", "TACO (soil)", v.n.ratio$full. id)
v.n.ratio<- subset( v.n.ratio, select=- c( nd) )
v.n.ratio<- reshape( v.n.ratio, timevar = "parameter", idvar =
c( "location", "full. id", "medium", "collection. by"), direction = 'wide')
v.n.ratio$ratio<- v.n.ratio$result. Vanadium/v.n.ratio$result. Nickel
col names( v.n.ratio) <- c( "full. id", "location", "medium", "collection. by", "nickel", "vanadium", "ratio")

v.n.ratio<- subset( v.n.ratio, !is. na( v.n.ratio$ratio) )

v.n.barchart<- ddply( v.n.ratio, . ( location, collection. by, medium) , summarise,
                    median=median( ratio, na. rm=T) ,
                    mean=mean( ratio, na. rm=T) ,
                    max=max( ratio, na. rm=T) ,
                    min=min( ratio, na. rm=T) )

```