Abundance box plots

Kelsey Jesser

10/5/2022

MST marker abundance in Ecuadorian household samples

Load libraries

```
library(ggplot2)
library(magrittr)
library(reshape2)
library(tidyr)
library(dplyr)
library(tibble)
library(janitor)
library(ggpubr)
library(ggthemes)
```

Clear environment

```
#rm(list = ls())
```

Import data

```
data<-read.csv("Phase II 09022022.csv")
```
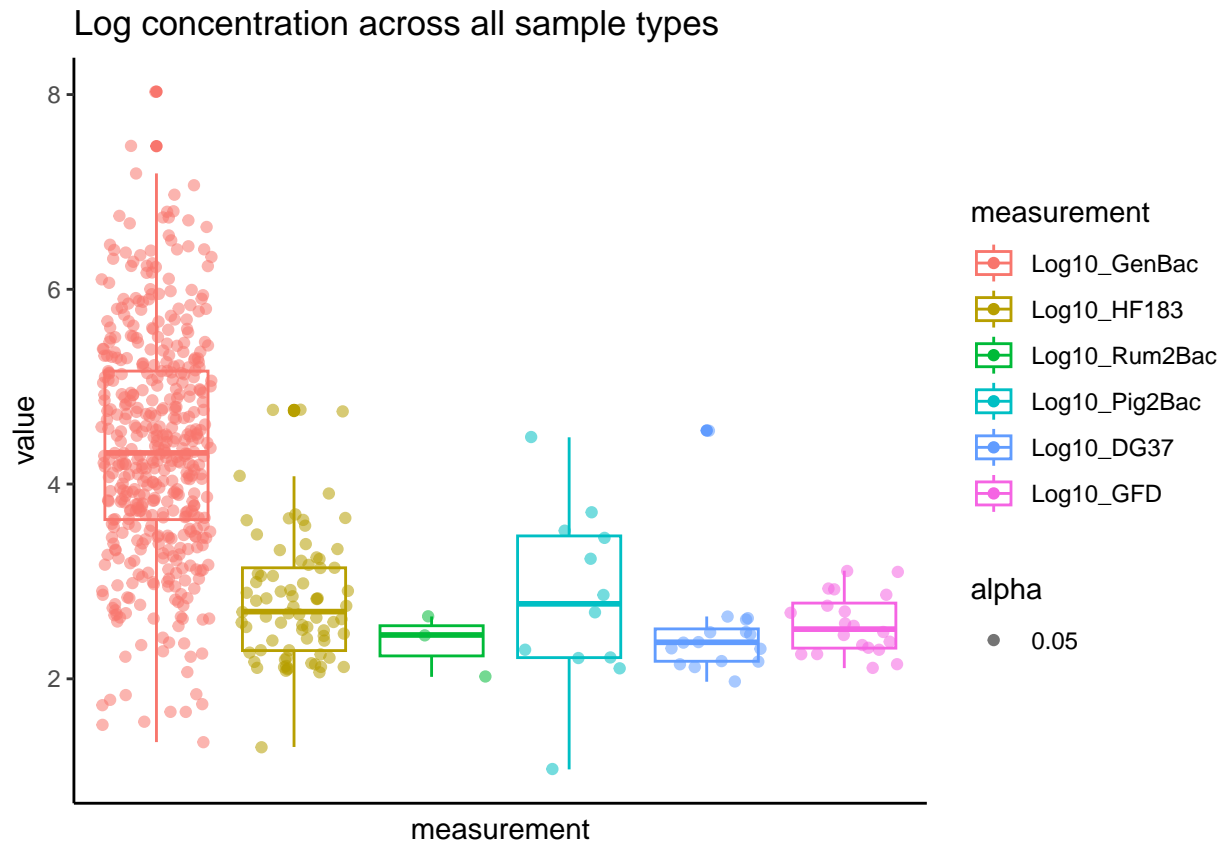
Concentration box plots

```
data_melt<-data %>%
  subset(select=-c(GenBac, HF183, Rum2Bac, Pig2Bac, DG37, GFD)) %>%
  melt(id.vars=c("Sample_Code", "Sample_Type", "Site", "Household"), variable.name=("measurement"), valu
data_melt$value<-as.numeric(data_melt$value)

a<-ggplot(data=data_melt, aes(x=measurement, y=value, fill=measurement))+
  geom_boxplot()+
  theme_classic()+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())
  a+ggtitle("Concentration across all sample types")
```
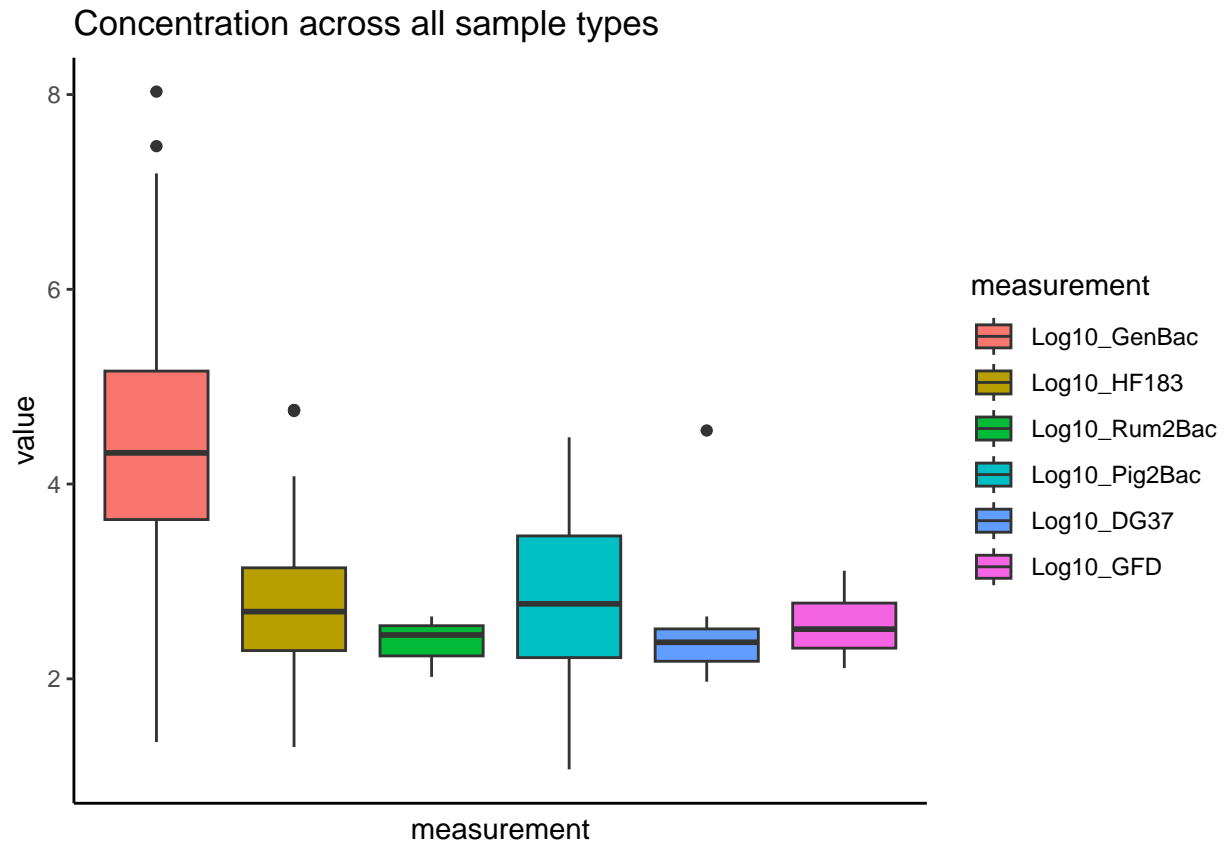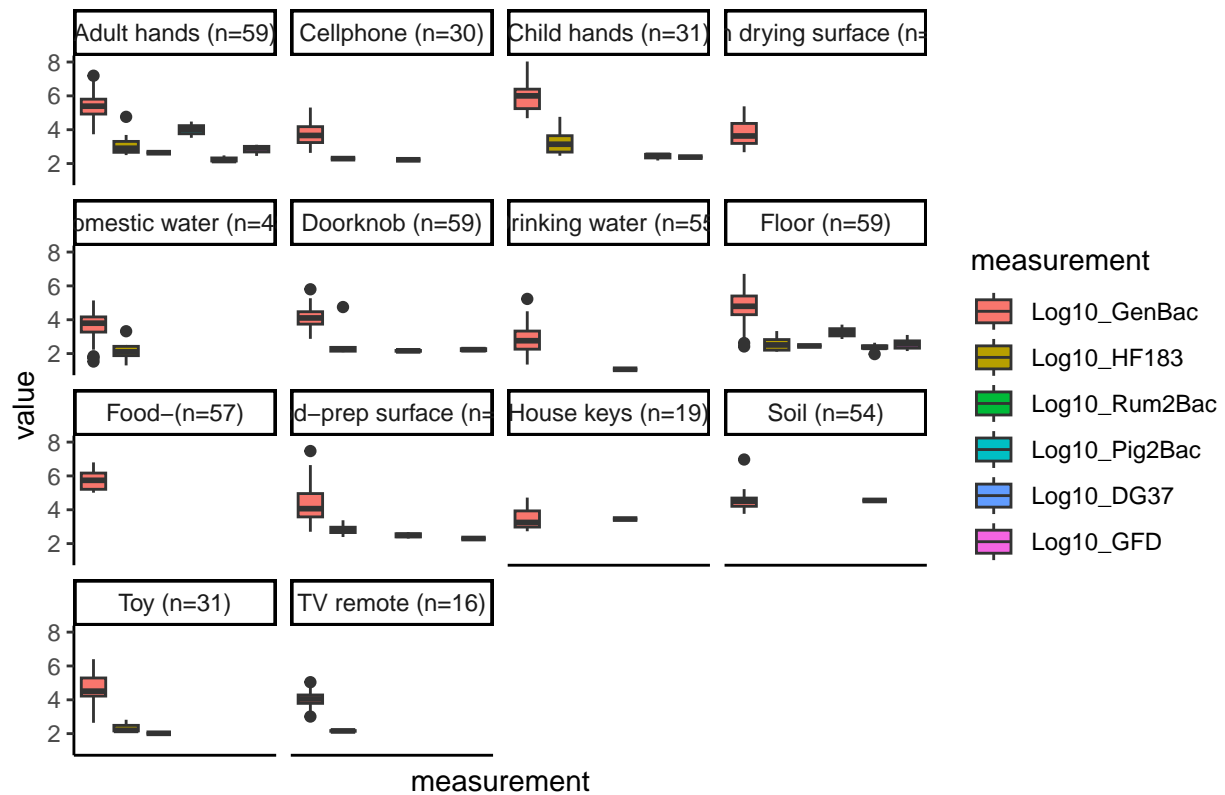
Concentration across all sample types


```
b<-ggplot(data=data_melt, aes(x=measurement, y=value, color=measurement))+
  geom_boxplot()+
  geom_jitter(aes(color=measurement, height = 0, width = .4, alpha=0.05)) +
  theme_classic()+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())
b+ggtitle("Log concentration across all sample types")
```

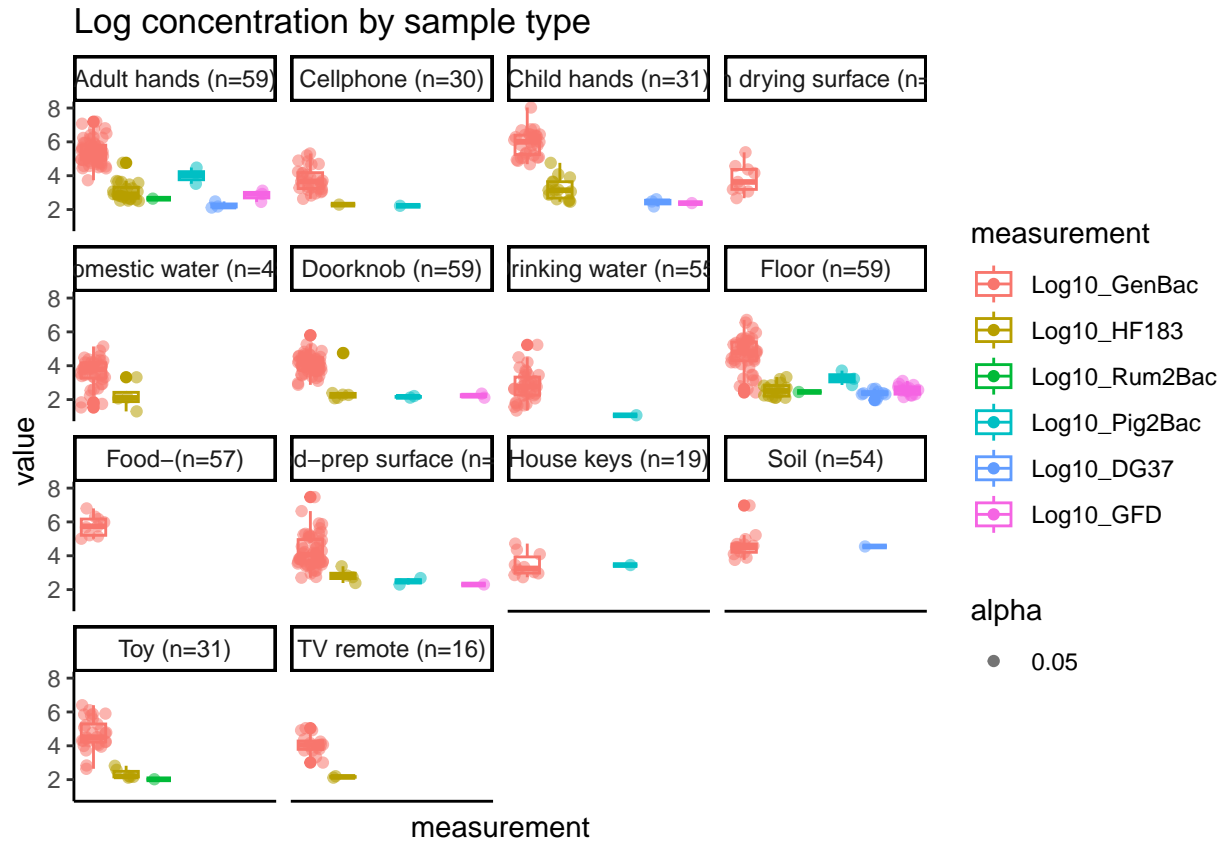# Log concentration across all sample types


```
c<-ggplot(data=data_melt, aes(x=measurement, y=value, fill=measurement))+
  geom_boxplot()+
  facet_wrap( ~Sample_Type)+
  theme_classic()+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())
c+ggtitle("Log concentration by sample type")
```
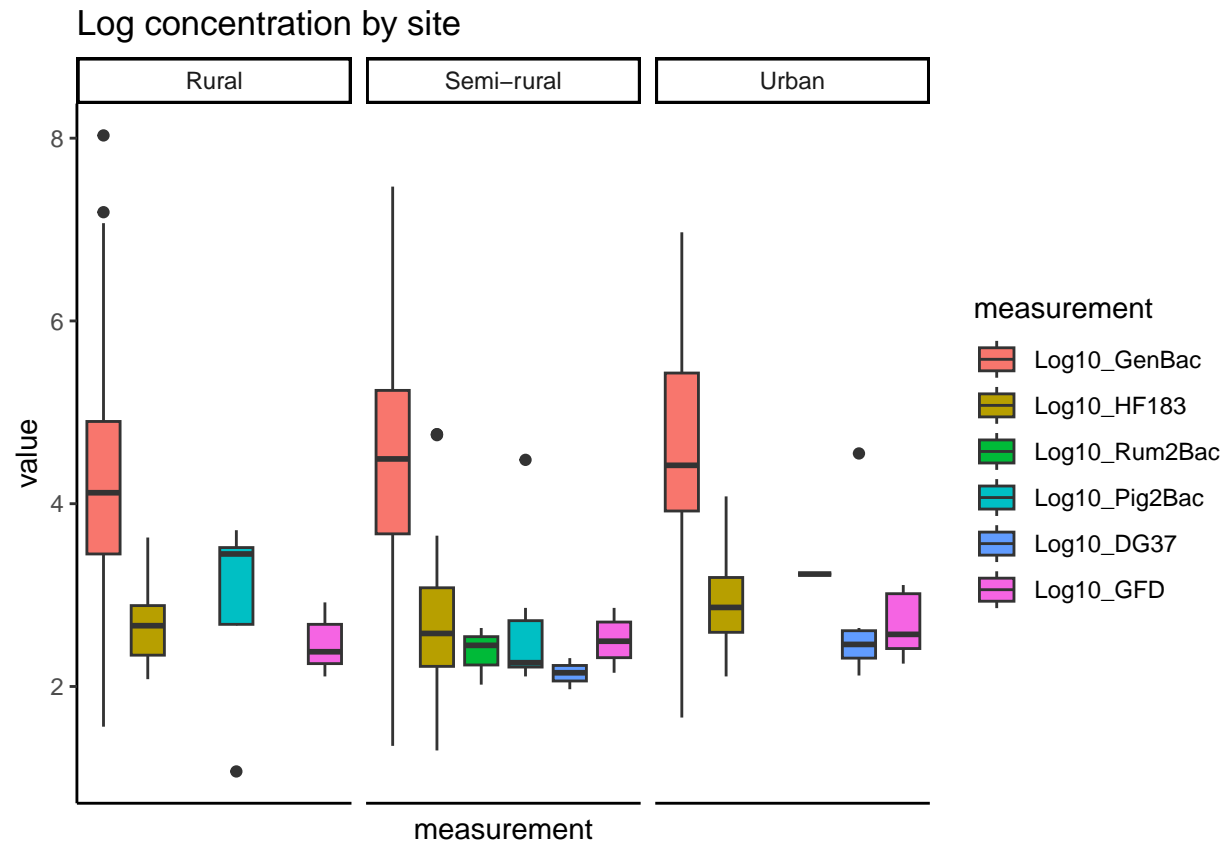
Log concentration by sample type


```
d<-ggplot(data=data_melt, aes(x=measurement, y=value, color=measurement))+
  geom_boxplot()+
  geom_jitter(aes(color=measurement, height = 0, width = .4, alpha=0.05)) +
  facet_wrap( ~Sample_Type)+
  theme_classic()+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())
d+ggtitle("Log concentration by sample type")
```

Log concentration by sample type


```
e<-ggplot(data=data_melt, aes(x=measurement, y=value, fill=measurement))+
  geom_boxplot()+
  facet_wrap( ~Site)+
  theme_classic()+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())
e+ggtitle("Log concentration by site")
```

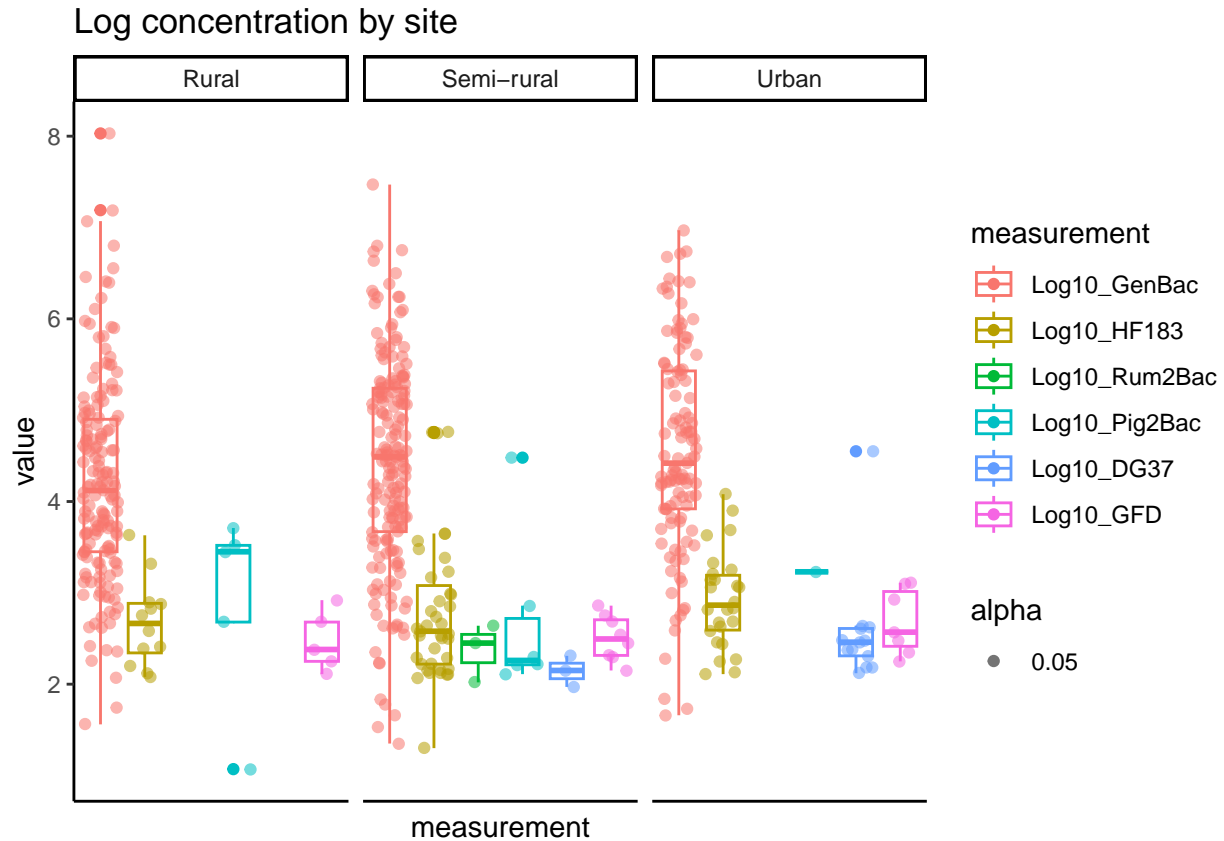## Log concentration by site


```
f<-ggplot(data=data_melt, aes(x=measurement, y=value, color=measurement))+
  geom_boxplot()+
  geom_jitter(aes(color=measurement, height = 0, width = .4, alpha=0.05)) +
  facet_wrap( ~Site)+
  theme_classic()+
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank())
f+ggtitle("Log concentration by site")
```

Log concentration by site
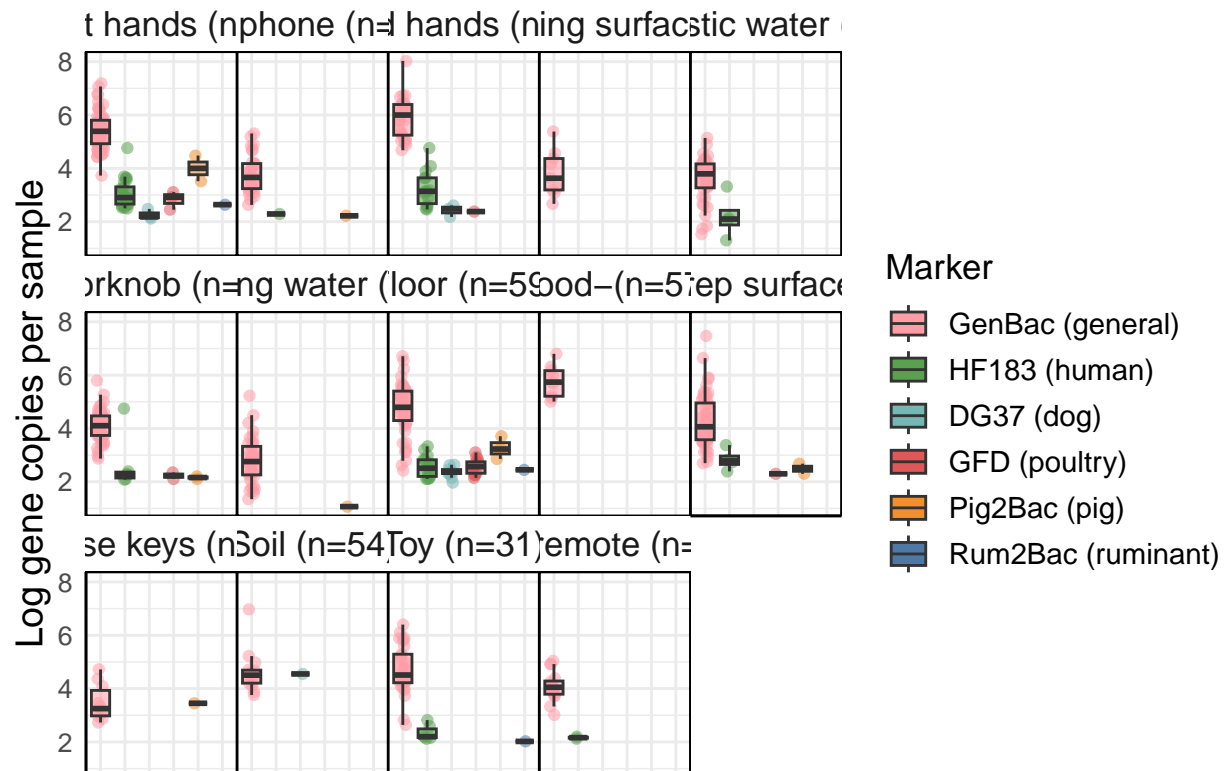
Nicely formatted multiplot

```r
data_melt$measurement <- factor(data_melt$measurement, levels = c("Log10_GenBac", "Log10_HF183", "Log10_

labels<-c("GenBac (general)", "HF183 (human)", "DG37 (dog)", "GFD (poultry)", "Pig2Bac (pig)", "Rum2Bac

cols <- c("Log10_GenBac"="#FF9DA7", "Log10_HF183"="#59A14F", "Log10_DG37"="#76B7B4", "Log10_GFD"="#E1575

abundance_boxplot<-ggplot(data=data_melt,
                          aes(x=measurement, y=value))+
  geom_jitter(aes(color=measurement,alpha=1), width=0.15, size=1.5) +
  geom_boxplot(outlier.shape=NA,  aes(fill=measurement, alpha=1))+
  theme_minimal()+
  scale_color_manual(values=cols, labels=labels)+
  scale_fill_manual(values=cols, labels=labels)+
  scale_x_discrete(labels=labels)+
  guides(alpha=FALSE, color=FALSE)+
  guides(fill=guide_legend(title="Marker"))+
  facet_wrap(~Sample_Type, nrow=3)+
  labs(y="Log gene copies per sample")+
  theme(axis.line=element_line(),
        axis.text.y=element_text(size=10),
        axis.title.y=element_text(size=14),
        legend.background=element_rect(color=NA),
        legend.text=element_text(size=11),
        legend.title=element_text(size=13),
        axis.text.x=element_blank(),
```

```
        axis.ticks.x=element_blank(),
        axis.title.x=element_blank(),
        strip.text.x=element_text(size=13),
        panel.spacing.x=unit(0, "line"),
        panel.spacing.y=unit(0, "line"),
        panel.border=element_rect(color="black", fill=NA))
abundance_boxplot+ggtitle("Marker abundance by sample type")
```

Marker abundance by sample type

Nicely formatted multiplot that includes only sample types with 3 or more host-associated detections

```
data_melt_sub<-data_melt%>%
  subset(Sample_Type=="Adult hands (n=59)" | Sample_Type=="Child hands (n=31)" | Sample_Type=="Doorknob

abundance_boxplot_sub<-ggplot(data=data_melt_sub,
                        aes(x=measurement, y=value))+
  geom_jitter(aes(color=measurement,alpha=1), width=0.15, size=1.5) +
  geom_boxplot(outlier.shape=NA,  aes(fill=measurement, alpha=1))+
  theme_minimal()+
  scale_color_manual(values=cols, labels=labels)+
  scale_fill_manual(values=cols, labels=labels)+
  scale_x_discrete(labels=labels)+
  guides(alpha="none", color="none")+
  guides(fill=guide_legend(title="MST Marker"))+
  facet_wrap(~Sample_Type, nrow=2)+
  labs(y="Log abundance per sample")+
  theme(axis.line=element_line(),
```

```
        axis.text.y=element_text(size=10),
        axis.title.y=element_text(size=14),
        legend.background=element_rect(color=NA),
        legend.text=element_text(size=11),
        legend.title=element_text(size=13),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.x=element_blank(),
        strip.text.x=element_text(size=13),
        panel.spacing.x=unit(0, "line"),
        panel.spacing.y=unit(0, "line"),
        panel.border=element_rect(color="black", fill=NA))
abundance_boxplot_sub+ggtitle("Marker abundance by sample type")
```


Marker abundance by sample type