

module2.R

phanthihonghuong

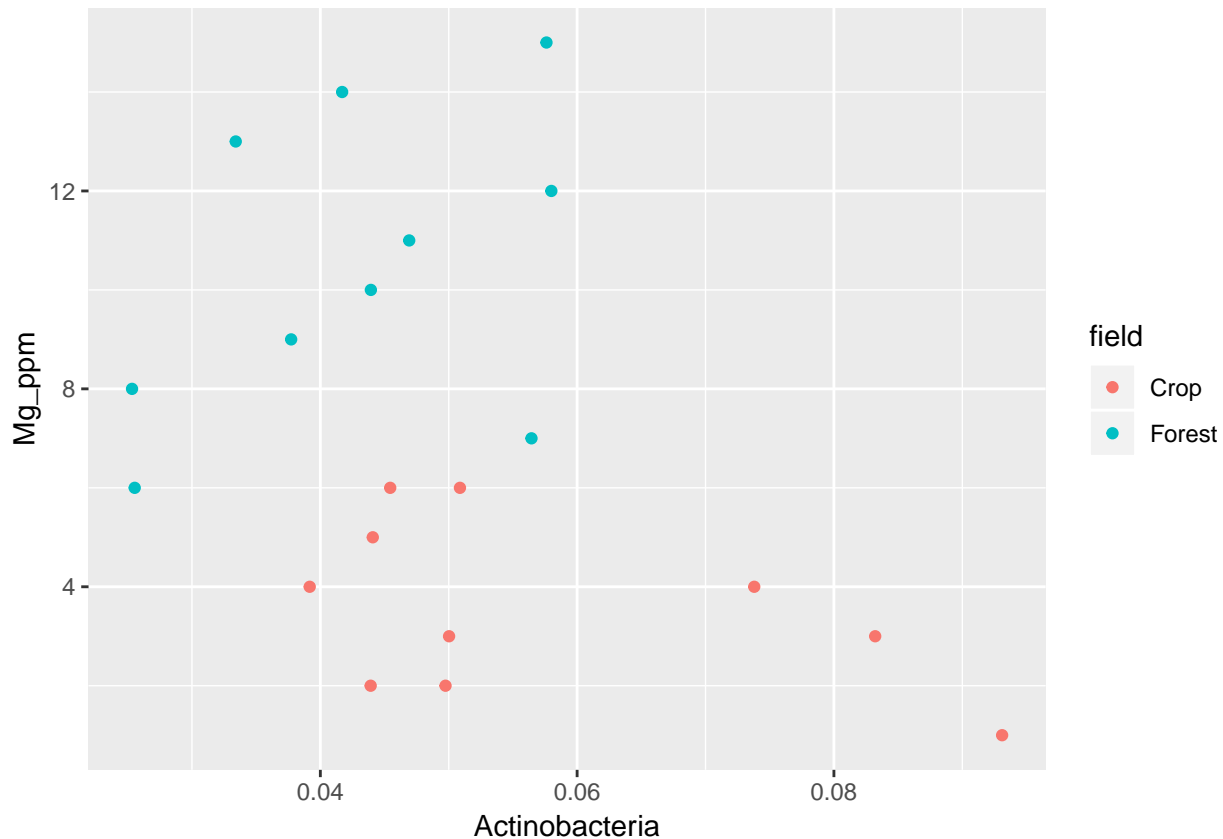
Sun Mar 24 21:03:03 2019

```
library(ggplot2)
chemdata <- as.matrix(read.table('SoilChemData2014_reduced.txt', header=TRUE))

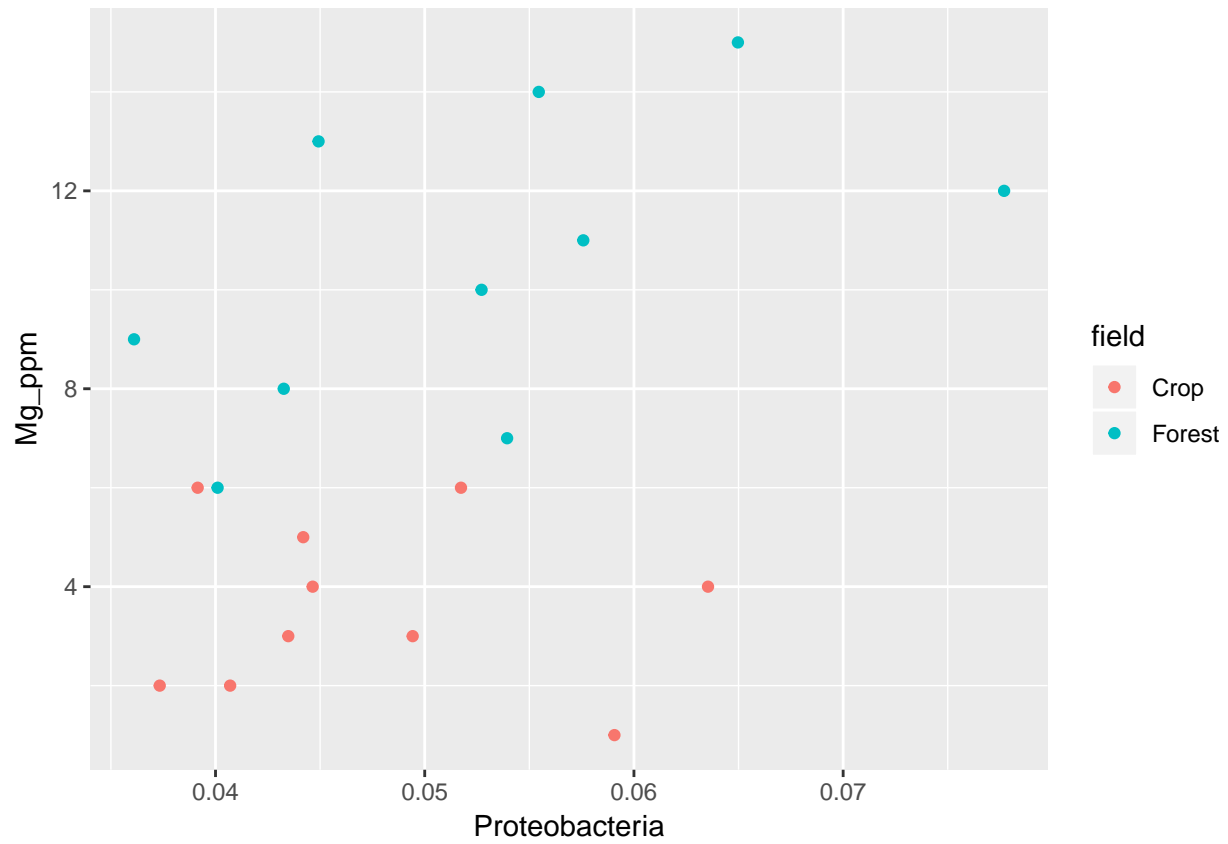
phyla <- read.table('fake.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.tax.summary.txt',
#phyla
phyla_taxlevel2 <- phyla[which(phyla$taxlevel == '2'),]

a <- phyla_taxlevel2[which(phyla_taxlevel2$taxon == 'Actinobacteria' | phyla_taxlevel2$taxon == 'Proteobacteria'),]
rownames(a) <- c('Actinobacteria', 'Proteobacteria')
a <- as.data.frame(t(a))
a[['Sample']] <- c('C1', 'C10', 'C2', 'C3', 'C4', 'C5', 'C6', 'C7', 'C8', 'C9', 'F1', 'F10', 'F2', 'F3')
b <- merge(a, chemdata, by='Sample')
field <- c(rep('Crop', 10), rep('Forest', 10))
b <- cbind(b, field)
b[,2] <- b[,2]/sum(b[,2])
b[,3] <- b[,3]/sum(b[,3])
b[,7] <- as.numeric(b[,7])

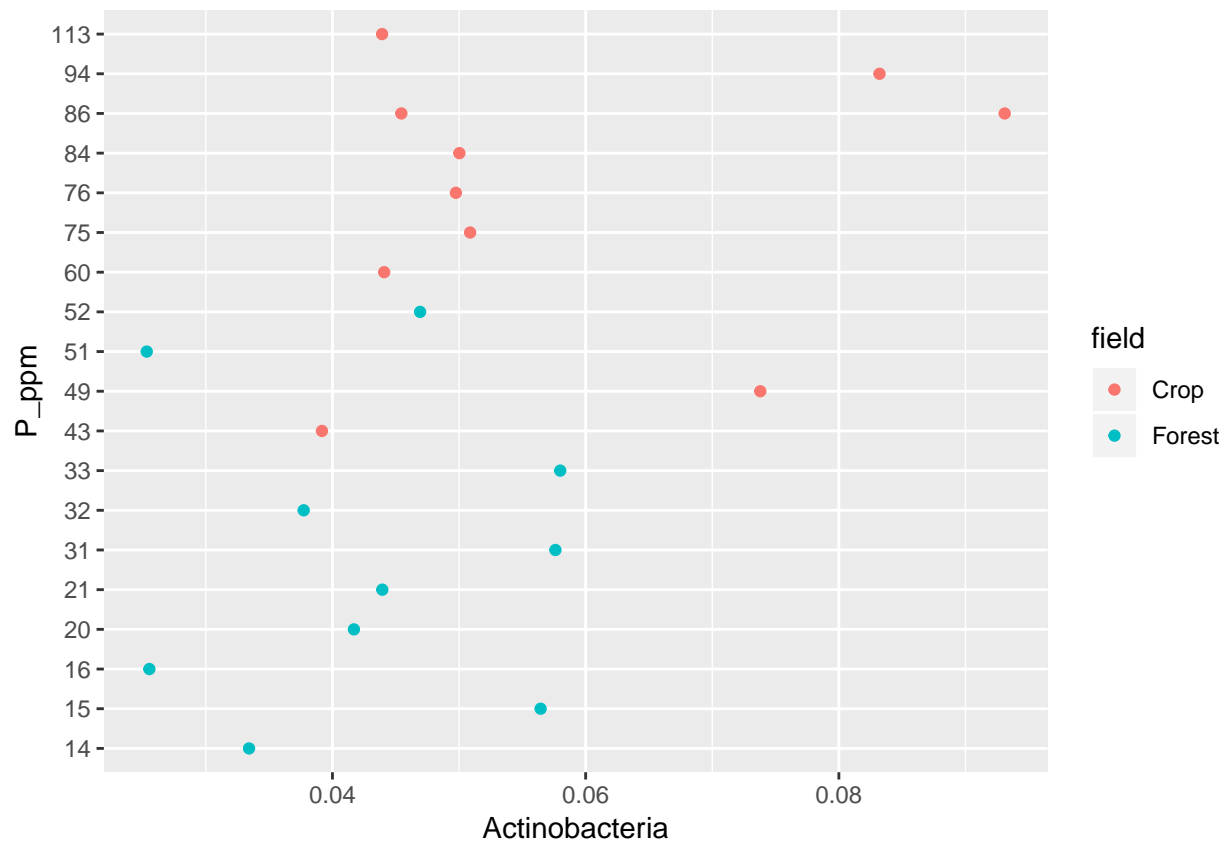
ggplot(b, aes(Actinobacteria, Mg_ppm, color=field)) + geom_point()
```



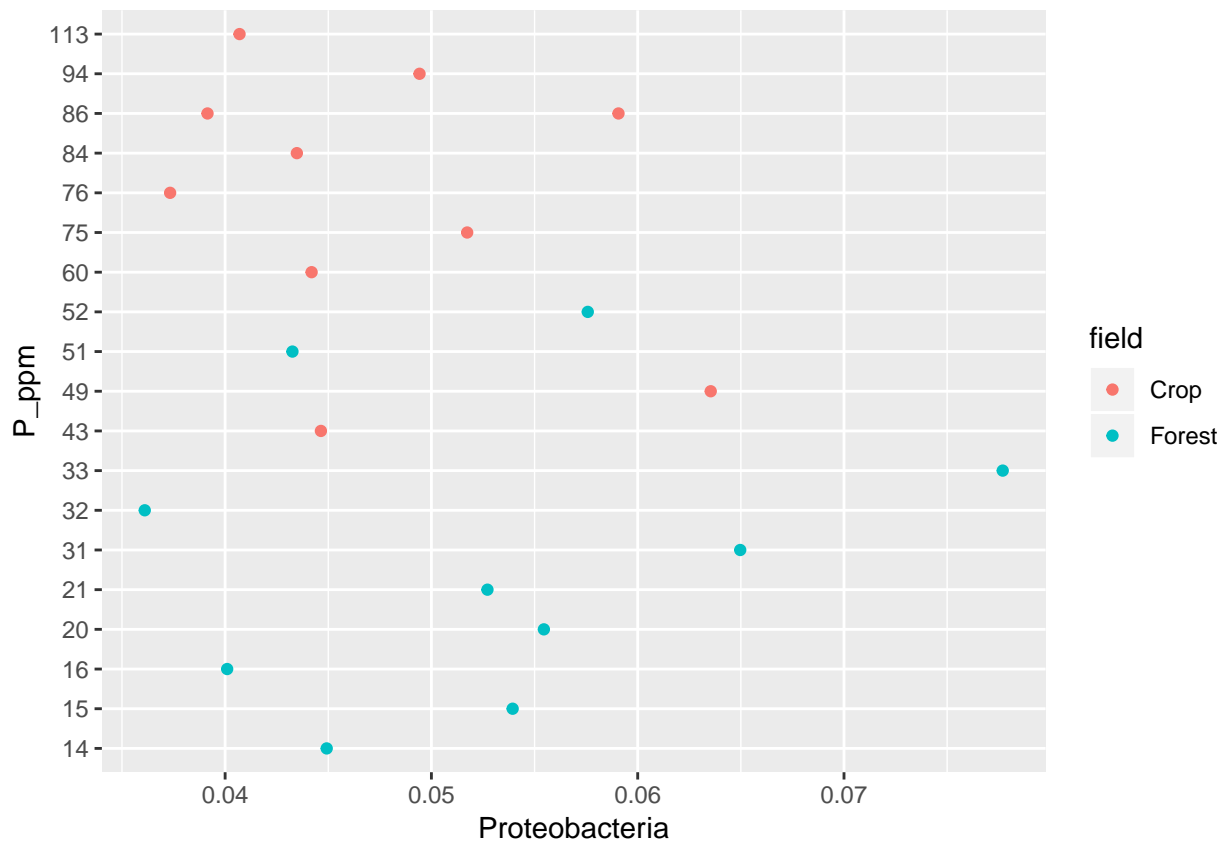
```
ggplot(b, aes(Proteobacteria, Mg_ppm,color=field)) + geom_point()
```



```
ggplot(b, aes(Actinobacteria, P_ppm,color=field)) + geom_point()
```



```
ggplot(b, aes(Proteobacteria, P_ppm,color=field)) + geom_point()
```



```
summary(lm(b[1:10,2] ~ b[1:10,7]))
```

```
##
## Call:
## lm(formula = b[1:10, 2] ~ b[1:10, 7])
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.021527	-0.014361	-0.002944	0.015288	0.022835

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.075585	0.013646	5.539	0.000548 ***
b[1:10, 7]	-0.005069	0.003455	-1.467	0.180544

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01775 on 8 degrees of freedom
## Multiple R-squared:  0.212, Adjusted R-squared:  0.1135
## F-statistic: 2.152 on 1 and 8 DF, p-value: 0.1805
```

```
summary(lm(b[1:10,3] ~ b[1:10,7]))
```

```
##
## Call:
## lm(formula = b[1:10, 3] ~ b[1:10, 7])
##
## Residuals:
```

```
##           Min           1Q       Median           3Q           Max
## -0.010654 -0.006414 -0.002536  0.004512  0.016380
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0488141  0.0069959   6.978 0.000115 ***
## b[1:10, 7]  -0.0004134  0.0017713  -0.233 0.821316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009101 on 8 degrees of freedom
## Multiple R-squared:  0.006763, Adjusted R-squared:  -0.1174
## F-statistic: 0.05447 on 1 and 8 DF, p-value: 0.8213
```

```
summary(lm(b[,2] ~ b[,7]))
```

```
##
## Call:
## lm(formula = b[, 2] ~ b[, 7])
##
## Residuals:
##           Min           1Q       Median           3Q           Max
## -0.025958 -0.009004 -0.003803  0.008545  0.034462
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0600652  0.0072838   8.246 1.59e-07 ***
## b[, 7]       -0.0014277  0.0008895  -1.605   0.126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01657 on 18 degrees of freedom
## Multiple R-squared:  0.1252, Adjusted R-squared:  0.07659
## F-statistic: 2.576 on 1 and 18 DF, p-value: 0.1259
```

```
summary(lm(b[11:20,2] ~ b[11:20,7]))
```

```
##
## Call:
## lm(formula = b[11:20, 2] ~ b[11:20, 7])
##
## Residuals:
##           Min           1Q       Median           3Q           Max
## -0.0136379 -0.0086943 -0.0000743  0.0061348  0.0199251
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.024234   0.014171   1.710   0.126
## b[11:20, 7]  0.001755   0.001302   1.348   0.215
##
## Residual standard error: 0.01182 on 8 degrees of freedom
## Multiple R-squared:  0.1851, Adjusted R-squared:  0.08329
## F-statistic: 1.818 on 1 and 8 DF, p-value: 0.2145
```

```
summary(lm(b[11:20,3] ~ b[11:20,7]))
```

```
##
## Call:
## lm(formula = b[11:20, 3] ~ b[11:20, 7])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0133920 -0.0047836 -0.0006242  0.0033680  0.0216351
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.028977   0.013209   2.194  0.0596 .
## b[11:20, 7]  0.002257   0.001213   1.860  0.0999 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01102 on 8 degrees of freedom
## Multiple R-squared:  0.3019, Adjusted R-squared:  0.2146
## F-statistic: 3.459 on 1 and 8 DF,  p-value: 0.09994
```

```
summary(lm(b[,3] ~ b[,7]))
```

```
##
## Call:
## lm(formula = b[, 3] ~ b[, 7])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.015918 -0.007503 -0.001974  0.003715  0.022561
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0426891   0.0044313   9.634 1.58e-08 ***
## b[, 7]       0.0010370   0.0005412   1.916  0.0714 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01008 on 18 degrees of freedom
## Multiple R-squared:  0.1694, Adjusted R-squared:  0.1233
## F-statistic: 3.672 on 1 and 18 DF,  p-value: 0.07135
```

```
summary <- read.table('fake.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti_mcc.0.0')
#summary
summary2 <- cbind(summary, field)
#summary2

library(data.table)
phylotypes <- fread('fake.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti_mcc.0.0')
#phylotypes
```