

# Red Sea data Exploratory Analysis

Denis Dreano

April 26, 2015

```
library(ggplot2)
print(getwd())

## [1] "/Users/denis/Dropbox/repos/redseachl/scripts/analysis"

df <- read.csv('../data/merged/data_reduced.csv')
df$X <- as.Date(df$X)

df$wndpwr_1 <- df$uwnd1^2 + df$vwnd1^2
df$wndpwr_2 <- df$uwnd2^2 + df$vwnd2^2
df$wndpwr_3 <- df$uwnd3^2 + df$vwnd3^2
df$wndpwr_4 <- df$uwnd4^2 + df$vwnd4^2
```

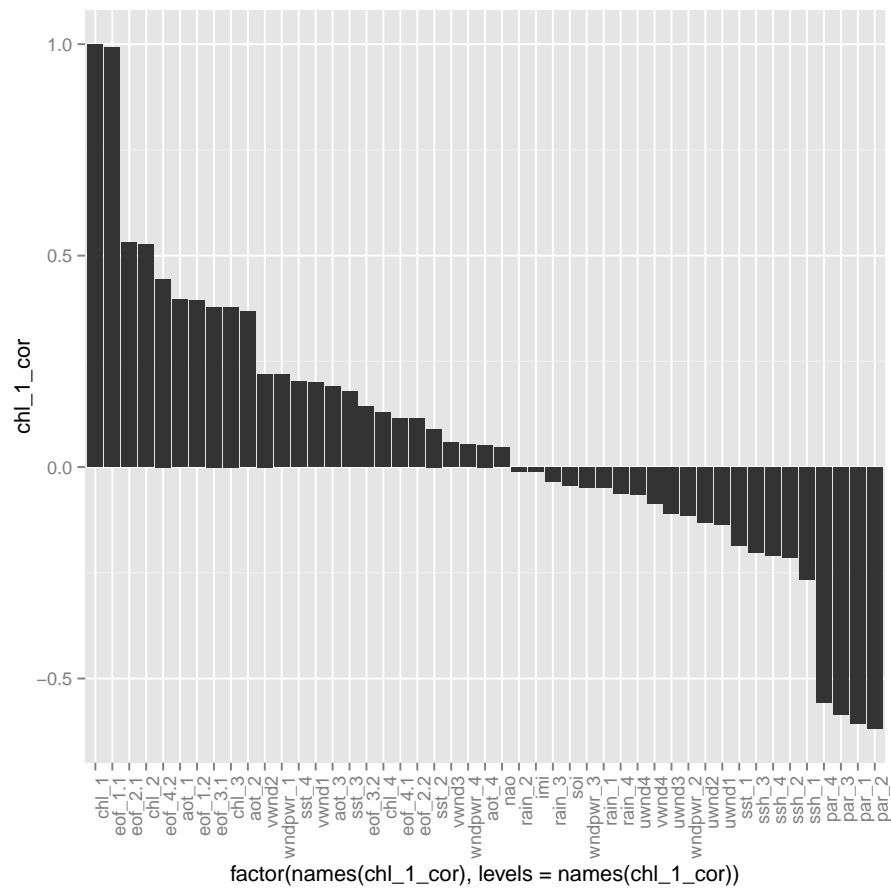
## 1 Region 1 (Southern Red Sea)

### 1.1 Correlation with other variables

The Correlation level between CHL in the south and other environmental variable. The most important are in the order: PAR, CHL (clusters 2 and 3), IMI, and wind power in cluster 1.

```
tot_cor <- cor(df[, -c(1,2,3)], use='complete.obs')
chl_1_cor <- tot_cor['chl_1',]
chl_1_cor <- sort(chl_1_cor, decreasing=TRUE)
names(chl_1_cor) <- factor(names(chl_1_cor), levels=names(chl_1_cor))
qplot(factor(names(chl_1_cor), levels=names(chl_1_cor)), chl_1_cor, geom='bar', stat='identity')

## Warning: Stacking not well defined when ymin != 0
```



## 1.2 Linear Regression

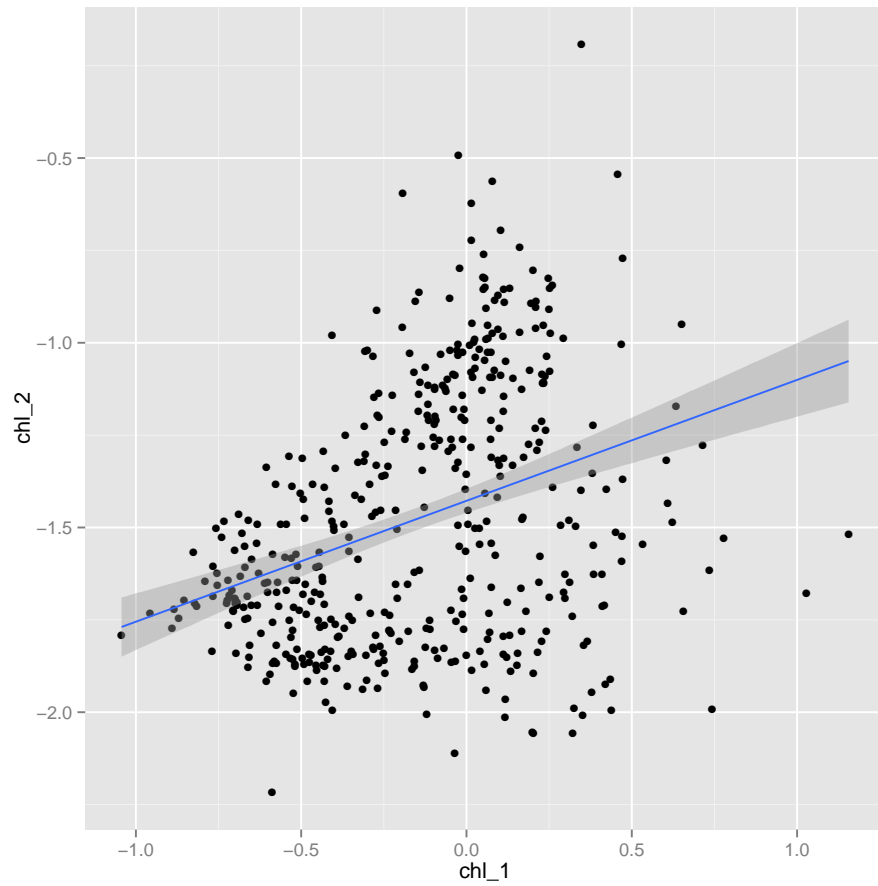
```
summary(lm(chl_1~sst_1+par_1+par_2+par_3+par_4+ssh_1+ssh_2+ssh_3+ssh_4+aot_1+chl_2+chl_3+aot_2,
data=df))
```

```
##
## Call:
## lm(formula = chl_1 ~ sst_1 + par_1 + par_2 + par_3 + par_4 +
##     ssh_1 + ssh_2 + ssh_3 + ssh_4 + aot_1 + chl_2 + chl_3 + aot_2,
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46880 -0.15882 -0.02295  0.12204  1.14062
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.724701    0.402686   4.283 2.40e-05 ***
## sst_1       -0.036525    0.015986  -2.285 0.022941 *
## par_1       -0.013776    0.007856  -1.754 0.080396 .
## par_2       -0.016272    0.010875  -1.496 0.135490
## par_3       -0.015442    0.013265  -1.164 0.245194
## par_4        0.032885    0.009272   3.547 0.000445 ***
## ssh_1       -2.507153    0.632757  -3.962 9.04e-05 ***
## ssh_2        1.546300    0.691516   2.236 0.025990 *
## ssh_3        0.934232    0.629504   1.484 0.138710
## ssh_4       -1.346539    0.524133  -2.569 0.010620 *
## aot_1        0.237522    0.554706   0.428 0.668779
## chl_2        0.460944    0.091576   5.033 7.80e-07 ***
## chl_3       -0.131964    0.088685  -1.488 0.137673
## aot_2        1.088669    0.520563   2.091 0.037236 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2364 on 342 degrees of freedom
## (108 observations deleted due to missingness)
## Multiple R-squared:  0.5333, Adjusted R-squared:  0.5156
## F-statistic: 30.07 on 13 and 342 DF,  p-value: < 2.2e-16
```

### 1.3 Chlorophyll and chlorophyll

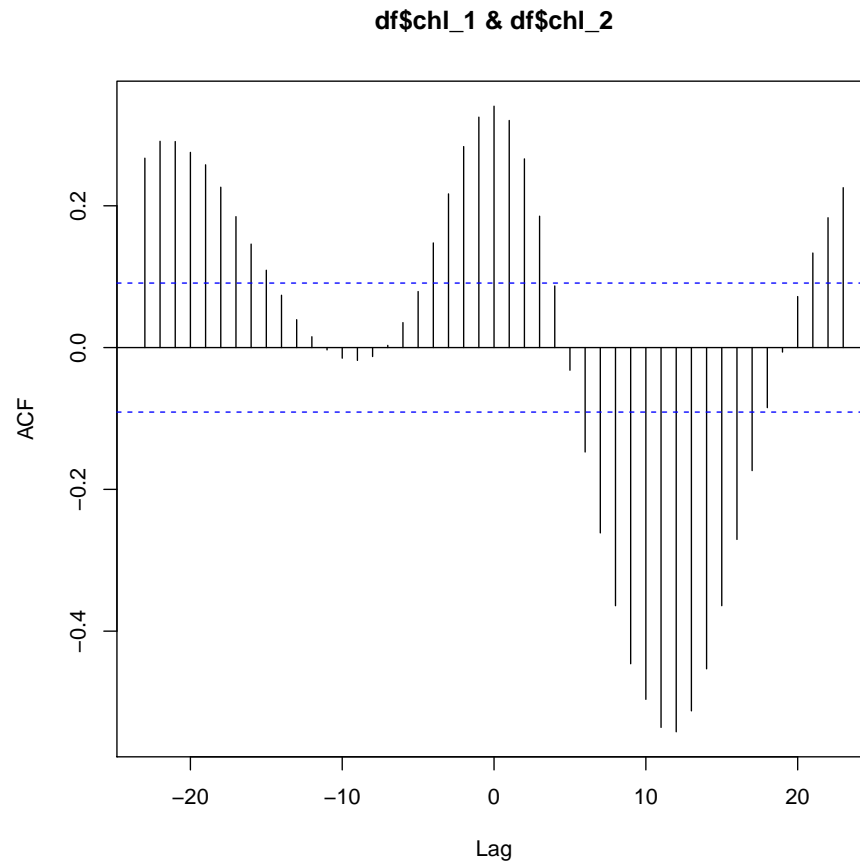
```
qplot(chl_1, chl_2, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



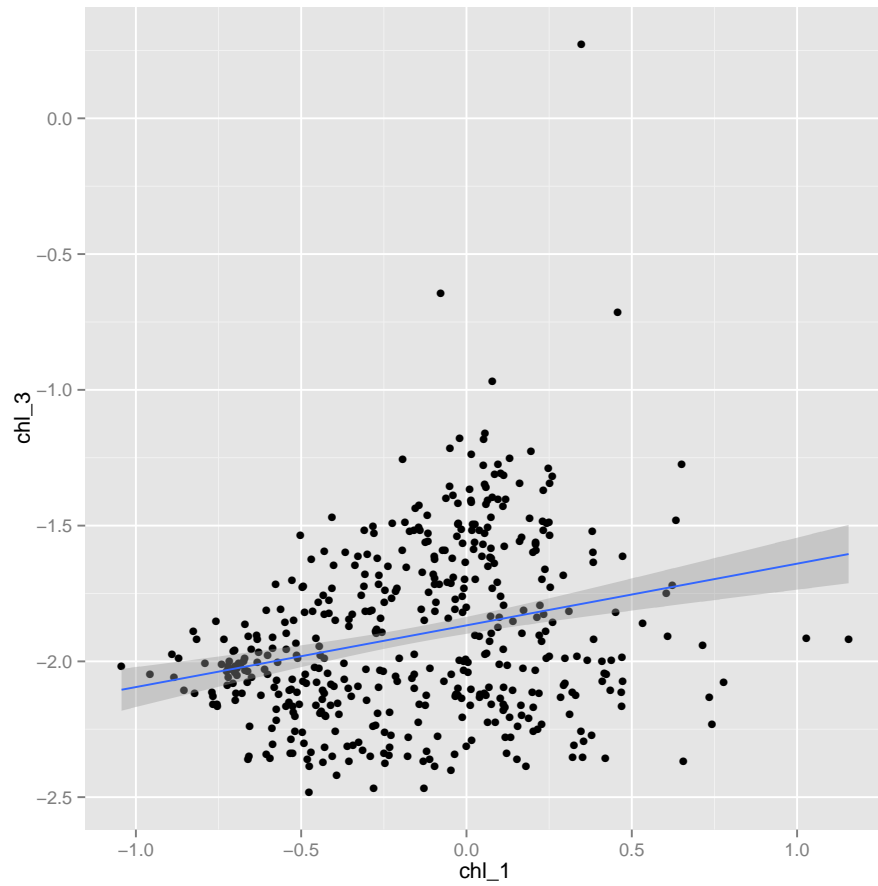
```
## Correlation coefficient
cor(df$chl_1, df$chl_2, use='complete.obs')

## [1] 0.3404985

ccf(df$chl_1, df$chl_2)
```



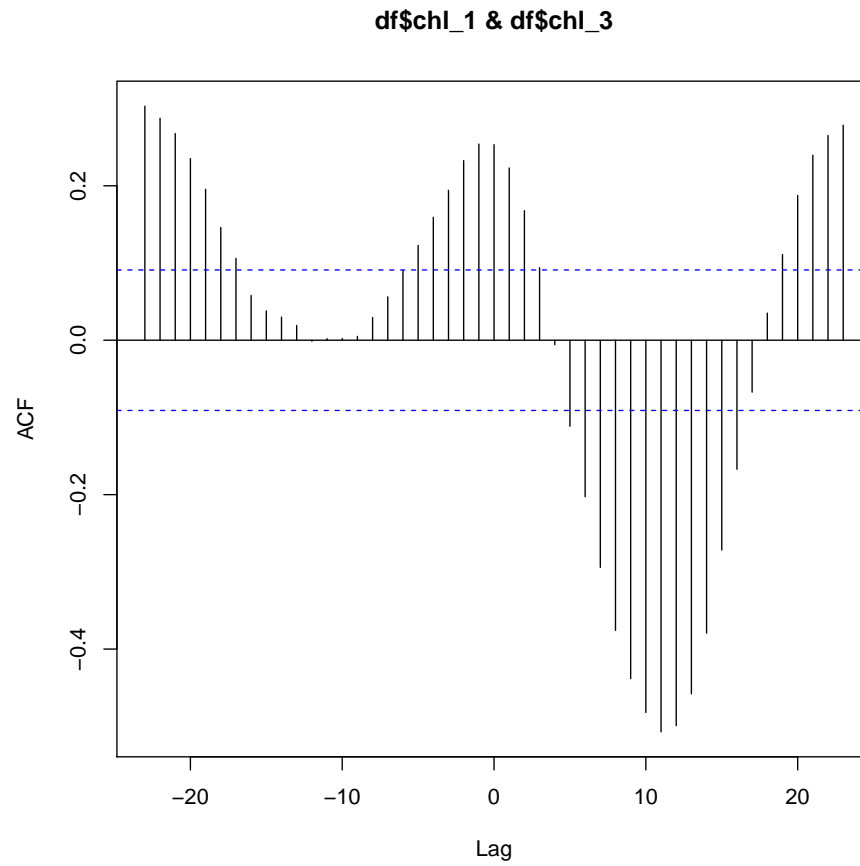
```
qplot(chl_1, chl_3, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



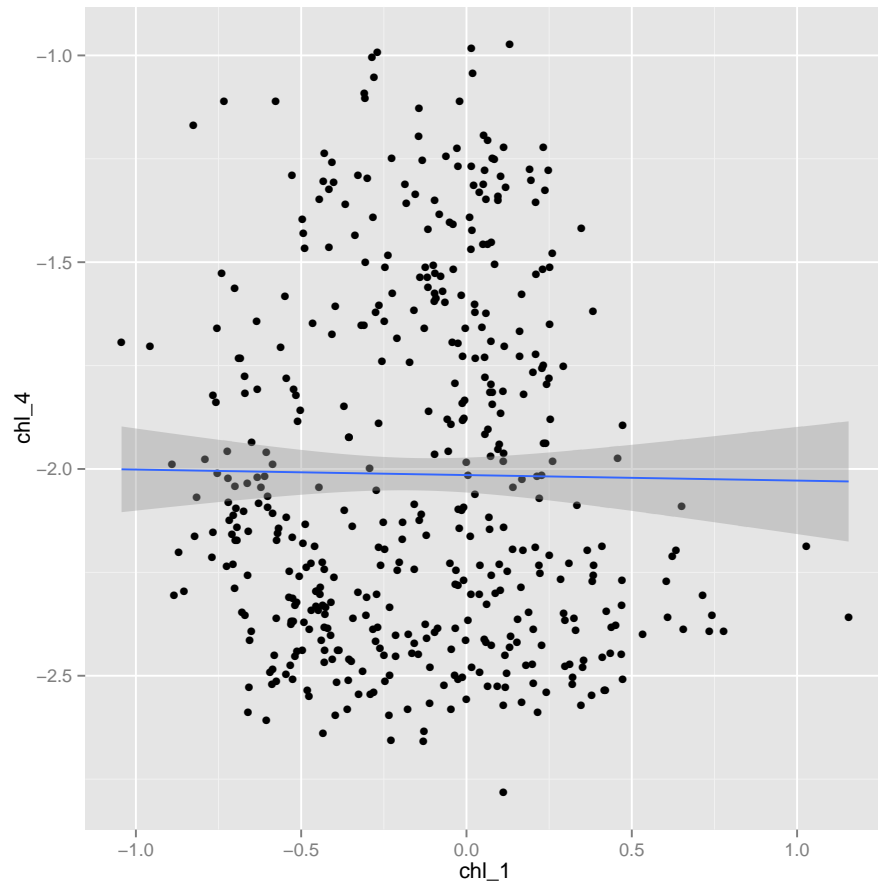
```
## Correlation coefficient
cor(df$chl_1, df$chl_3, use='complete.obs')

## [1] 0.2533124

ccf(df$chl_1, df$chl_3)
```

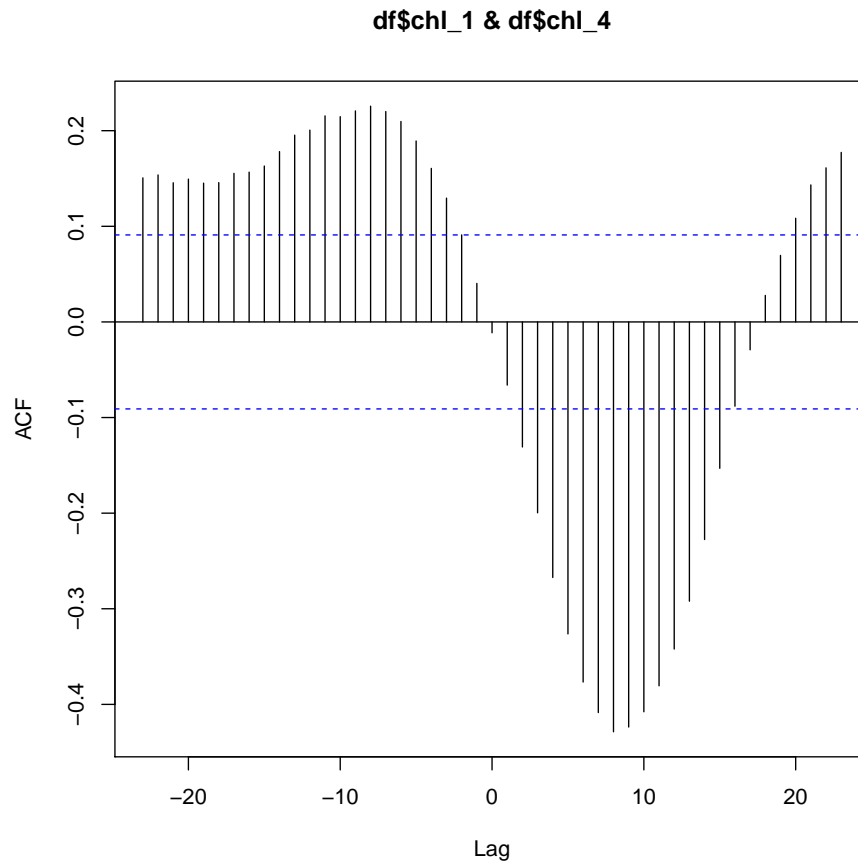


```
qplot(chl_1, chl_4, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



```
## Correlation coefficient  
cor(df$chl_1, df$chl_4, use='complete.obs')  
  
## [1] -0.01136072  
  
ccf(df$chl_1, df$chl_4)
```

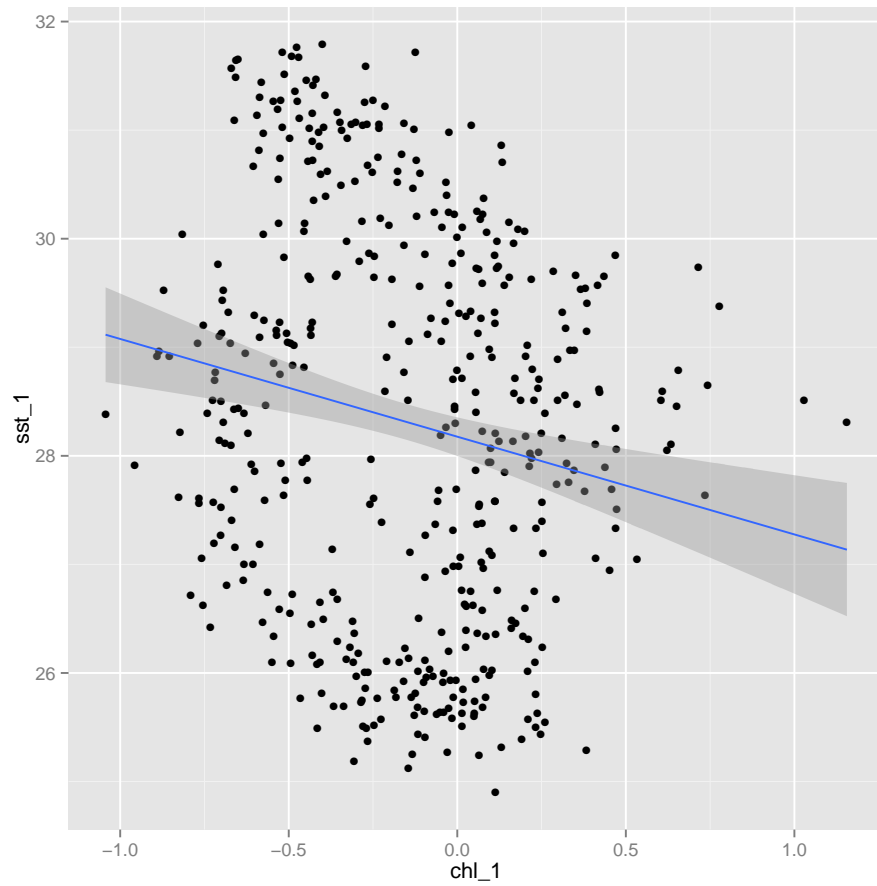




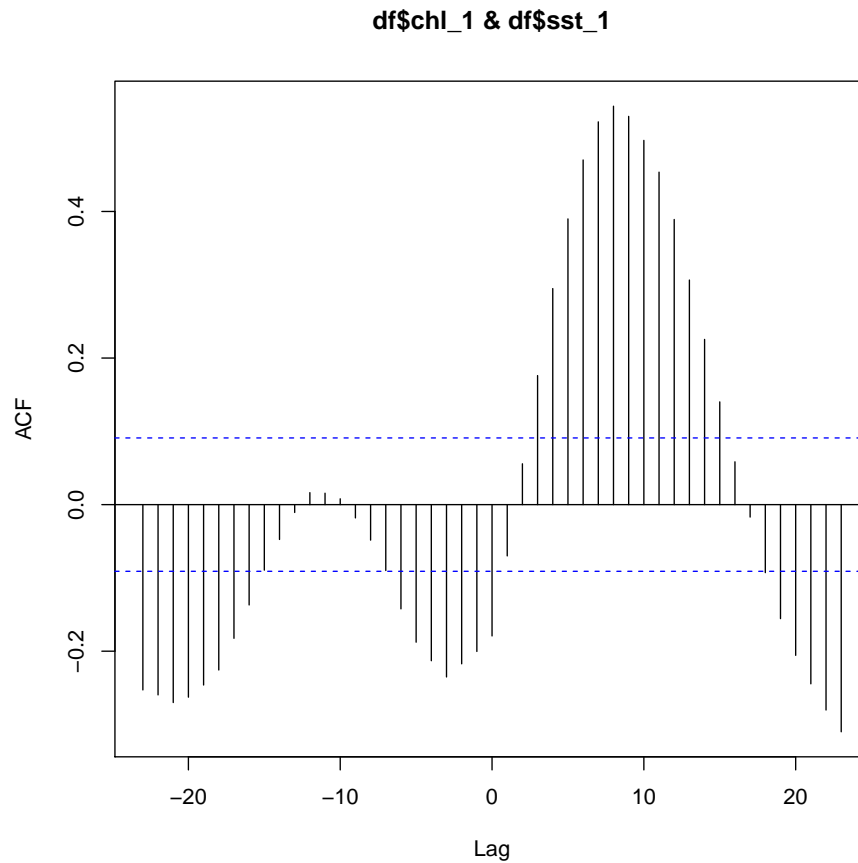
## 1.4 Chlorophyll and temperatures

```
qplot(chl_1, sst_1, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)

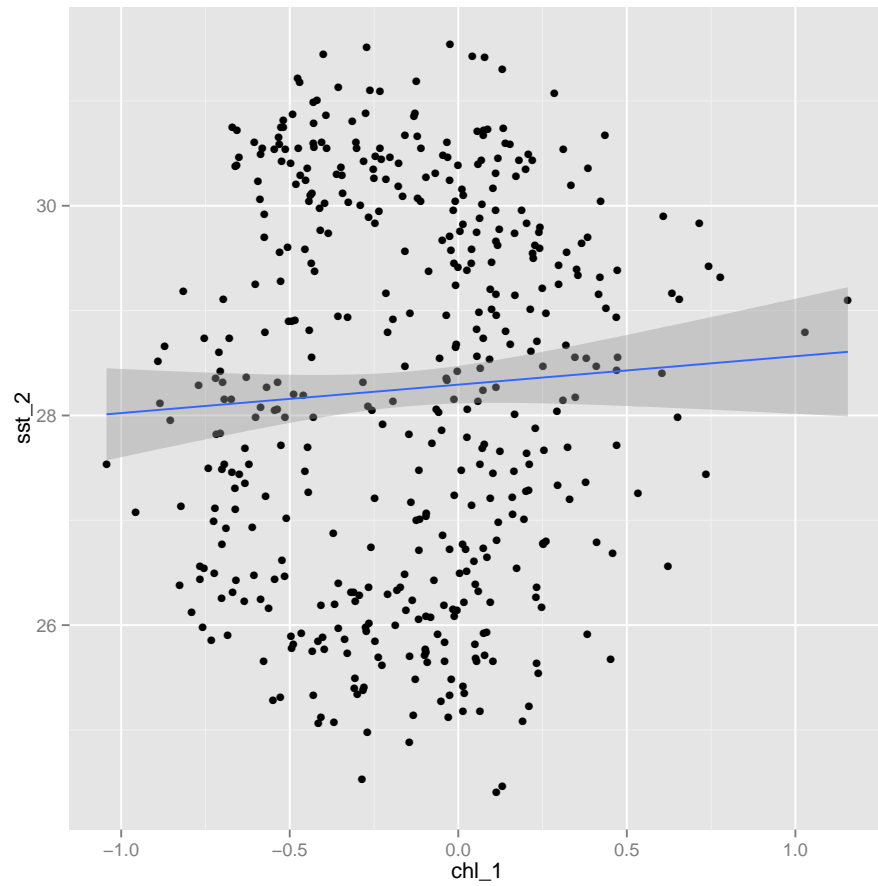
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## Correlation coefficient  
cor(df$chl_1, df$sst_1, use='complete.obs')  
  
## [1] -0.1792175  
  
ccf(df$chl_1, df$sst_1, na.action=na.pass)
```



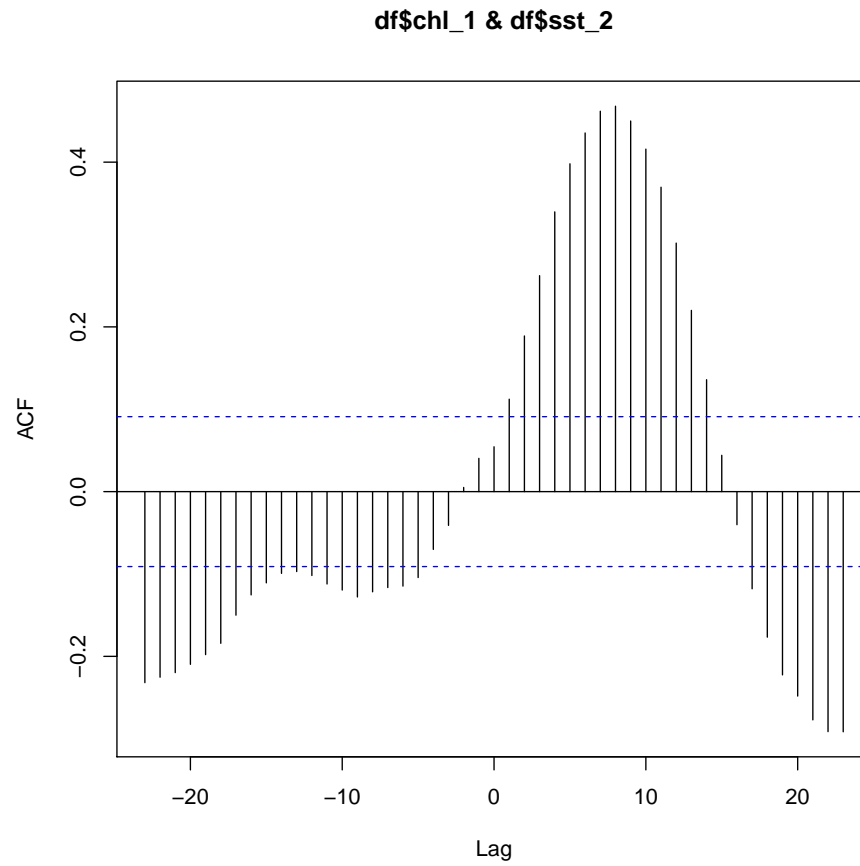
```
qplot(chl_1, sst_2, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
## Warning: Removed 1 rows containing missing values (stat.smooth).
## Warning: Removed 1 rows containing missing values (geom.point).
```



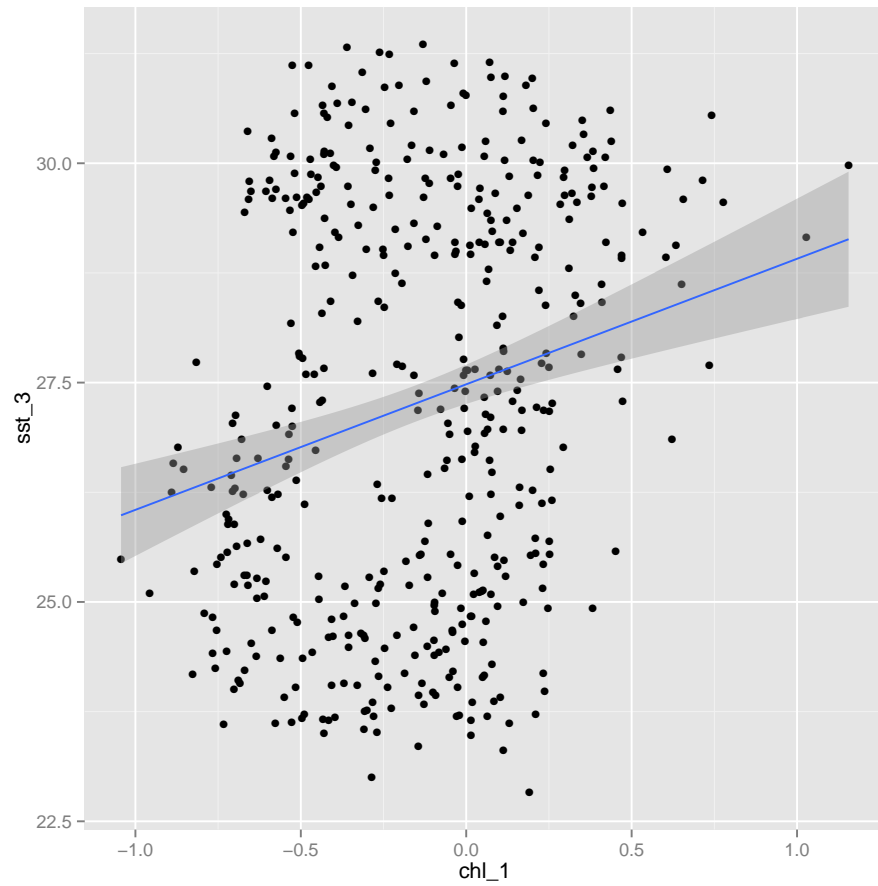
```
## Correlation coefficient
cor(df$chl_1, df$sst_2, use='complete.obs')

## [1] 0.05444485

ccf(df$chl_1, df$sst_2, na.action=na.pass)
```



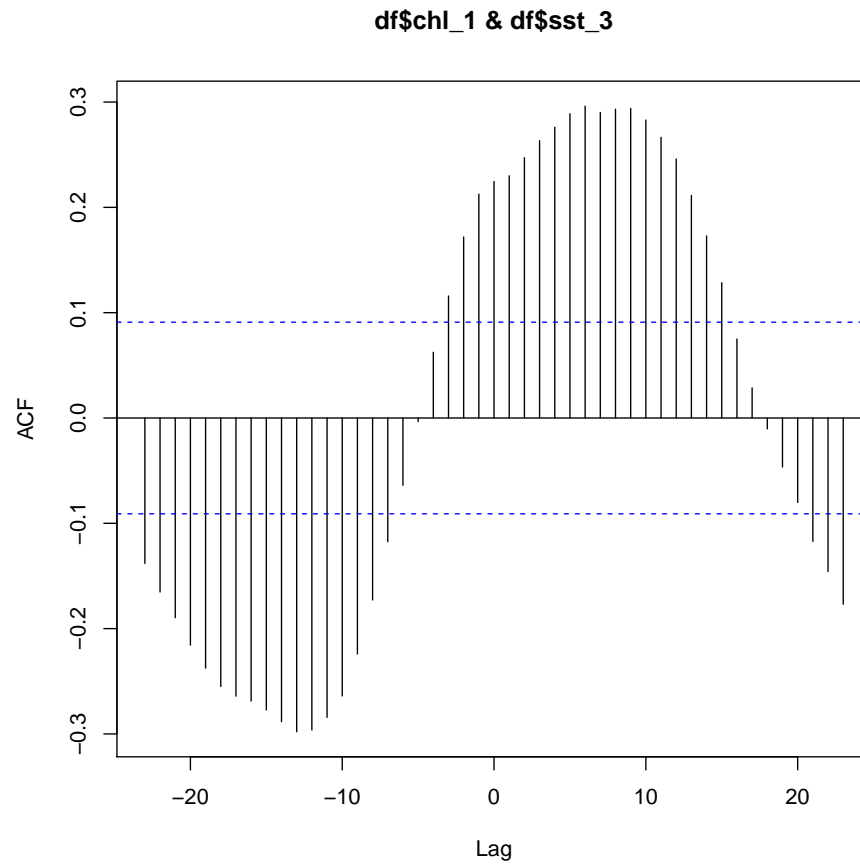
```
qplot(chl_1, sst_3, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
## Warning: Removed 1 rows containing missing values (stat.smooth).
## Warning: Removed 1 rows containing missing values (geom.point).
```



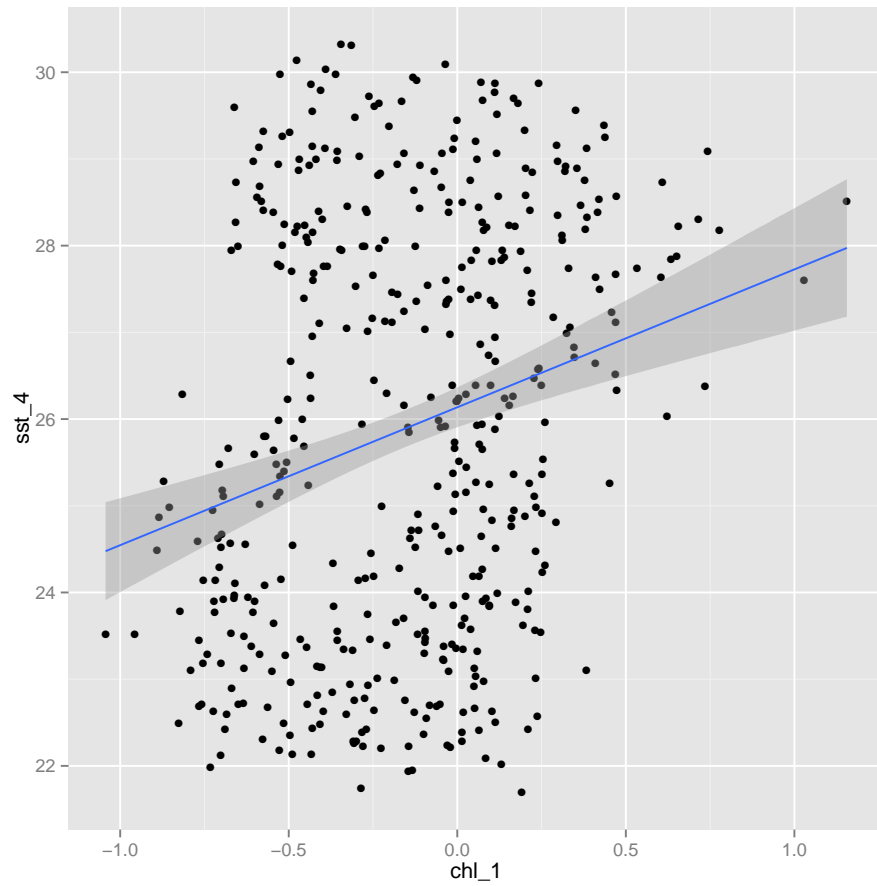
```
## Correlation coefficient
cor(df$chl_1, df$sst_3, use='complete.obs')

## [1] 0.2245816

ccf(df$chl_1, df$sst_3, na.action=na.pass)
```



```
qplot(chl_1, sst_4, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (geom_point).
```

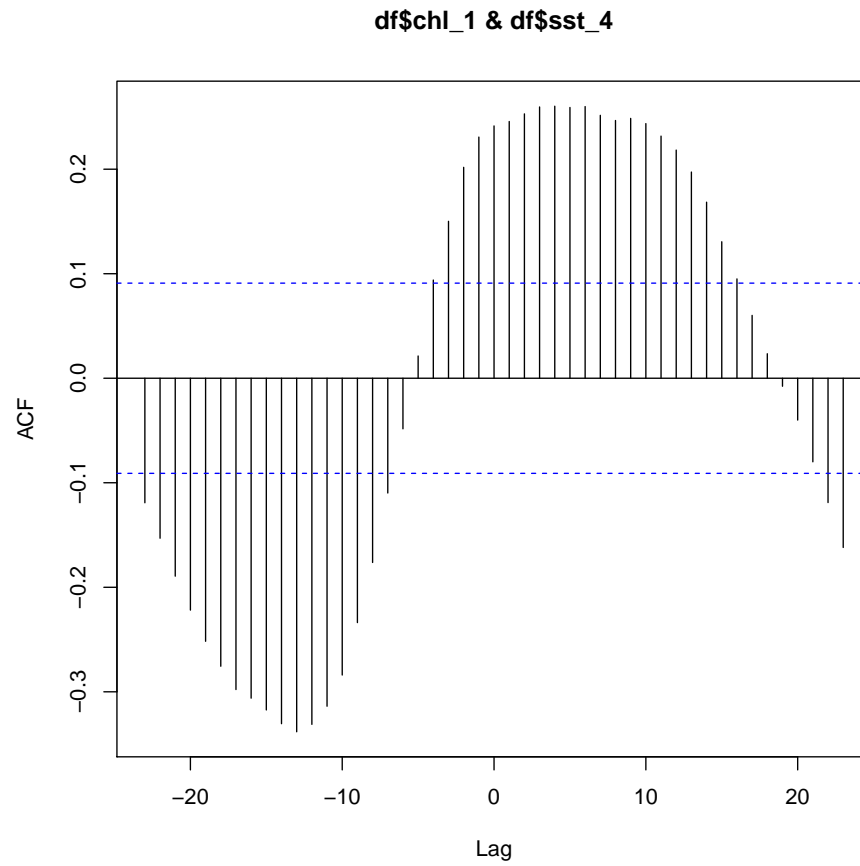


```
## Correlation coefficient
cor(df$chl_1, df$sst_4, use='complete.obs')

## [1] 0.2414238

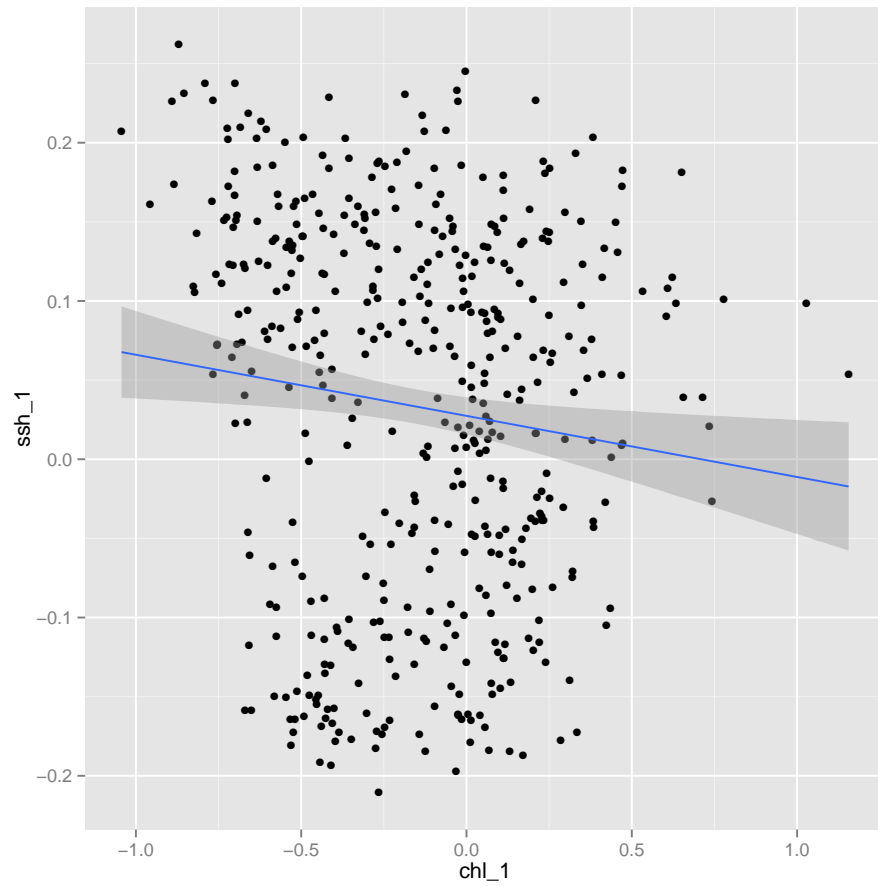
ccf(df$chl_1, df$sst_4, na.action=na.pass)
```





## 1.5 Chlorophyll and SLA

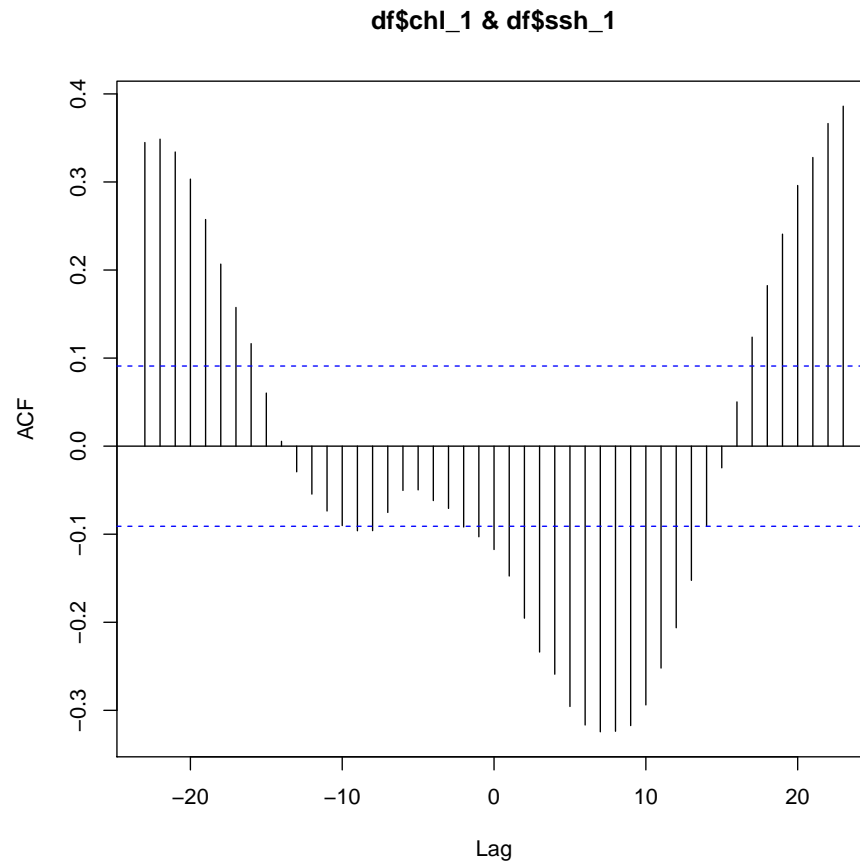
```
qplot(chl_1, ssh_1, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



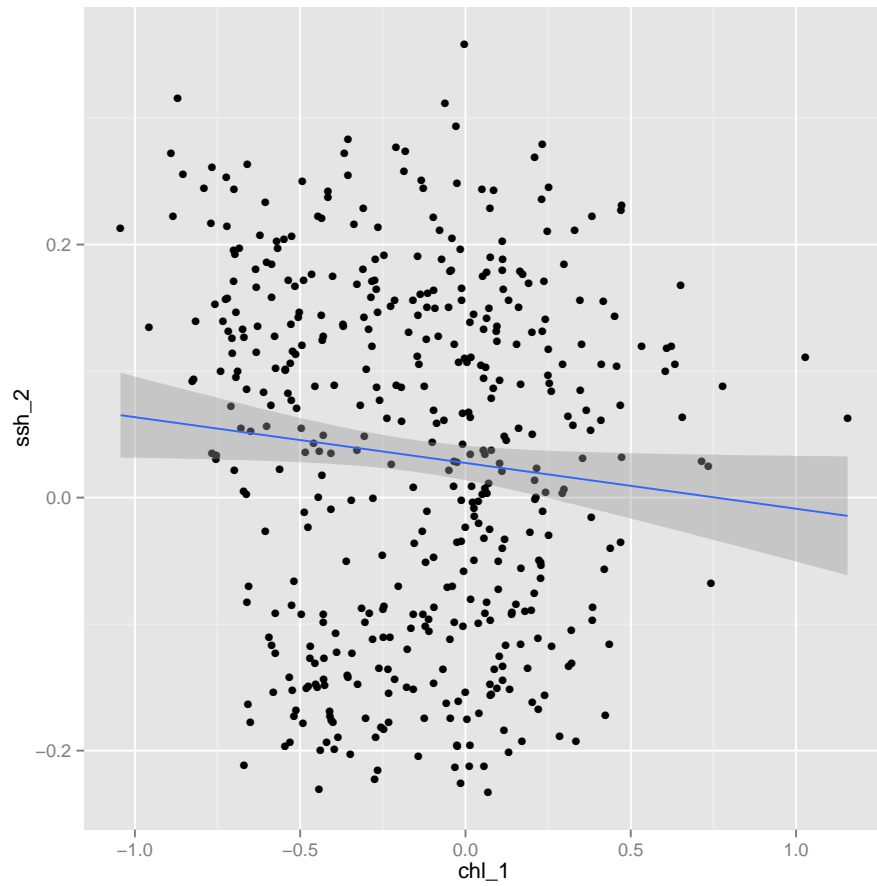
```
## Correlation coefficient
cor(df$chl_1, df$ssh_1, use='complete.obs')

## [1] -0.1173844

ccf(df$chl_1, df$ssh_1, na.action=na.pass)
```



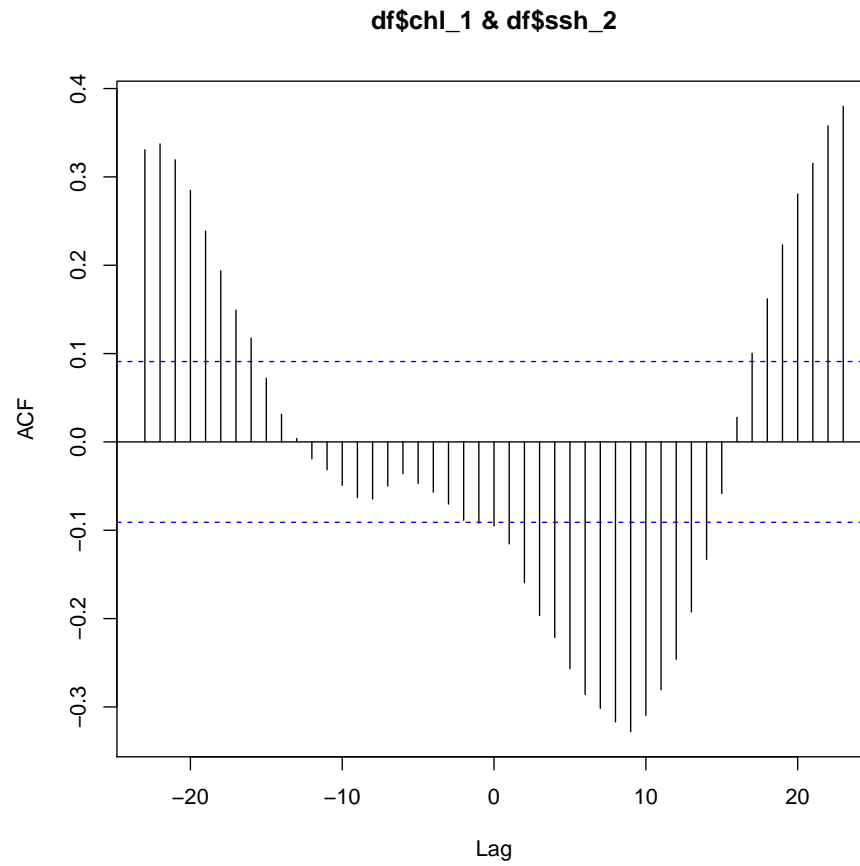
```
qplot(chl_1, ssh_2, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



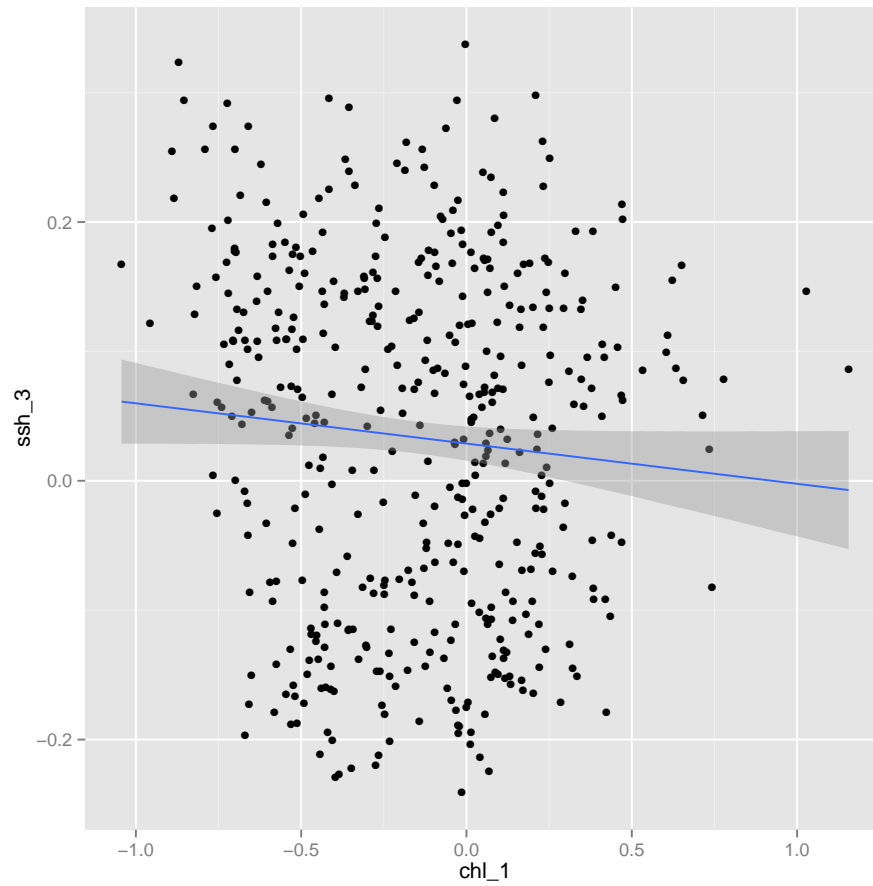
```
## Correlation coefficient
cor(df$chl_1, df$ssh_2, use='complete.obs')

## [1] -0.09501634

ccf(df$chl_1, df$ssh_2, na.action=na.pass)
```



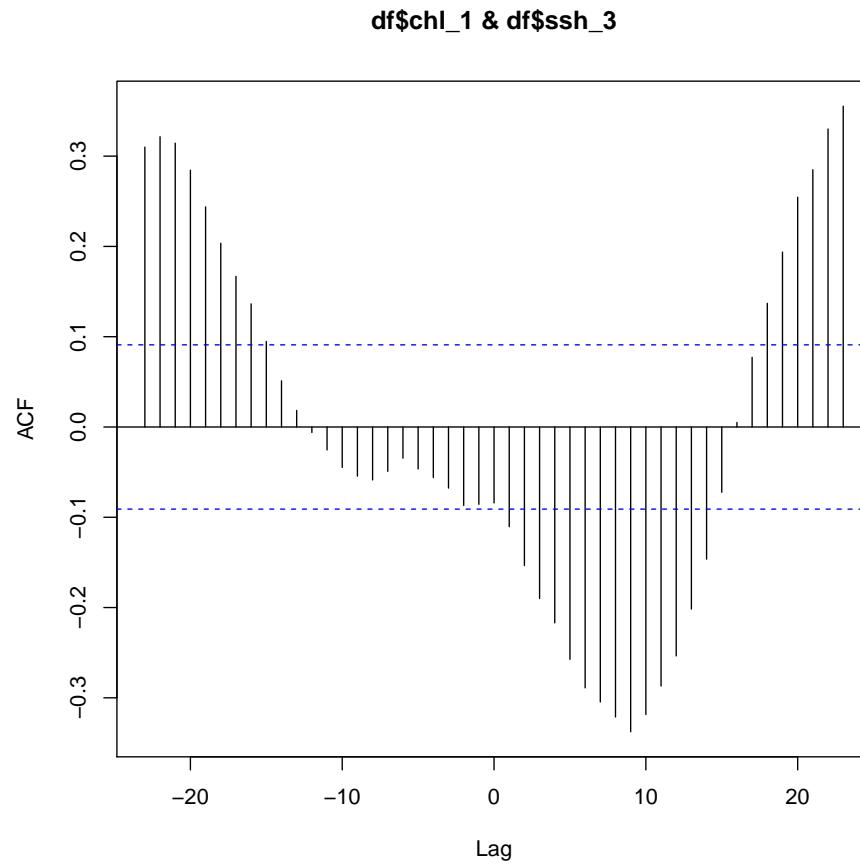
```
qplot(chl_1, ssh_3, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



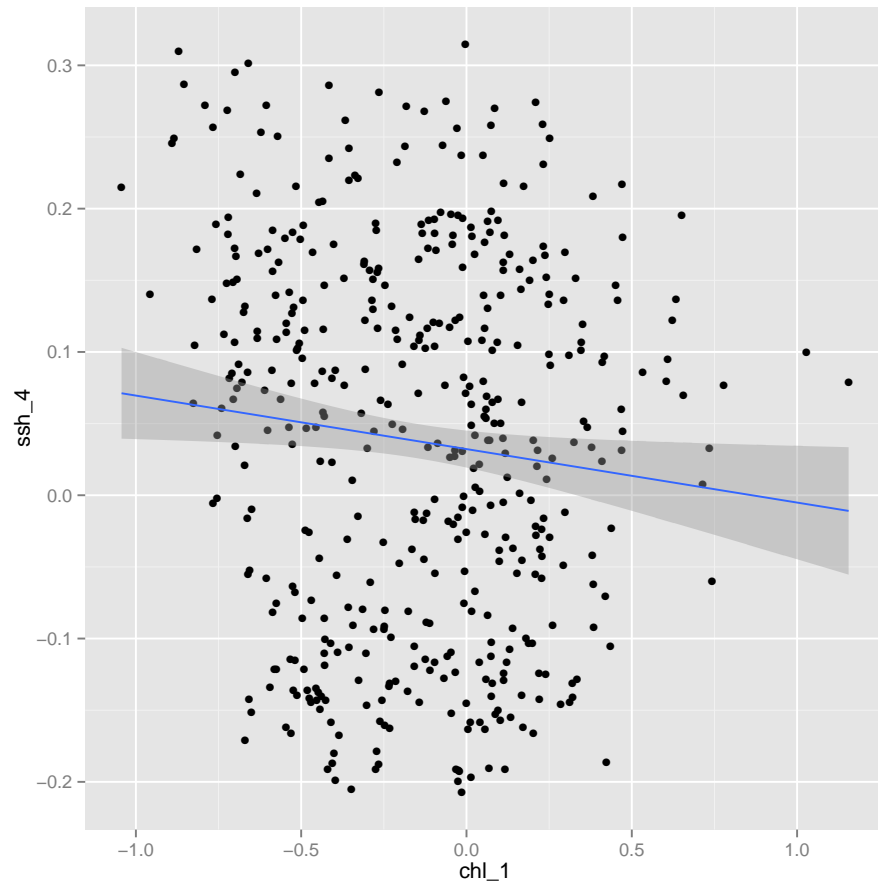
```
## Correlation coefficient
cor(df$chl_1, df$ssh_3, use='complete.obs')

## [1] -0.08408265

ccf(df$chl_1, df$ssh_3, na.action=na.pass)
```

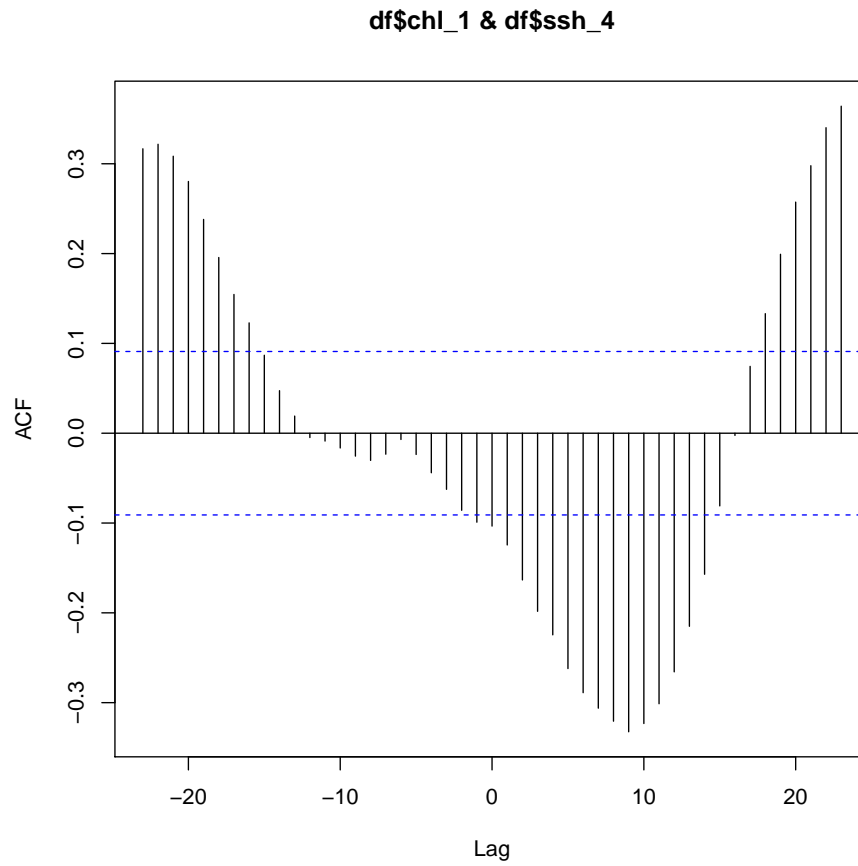


```
qplot(chl_1, ssh_4, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



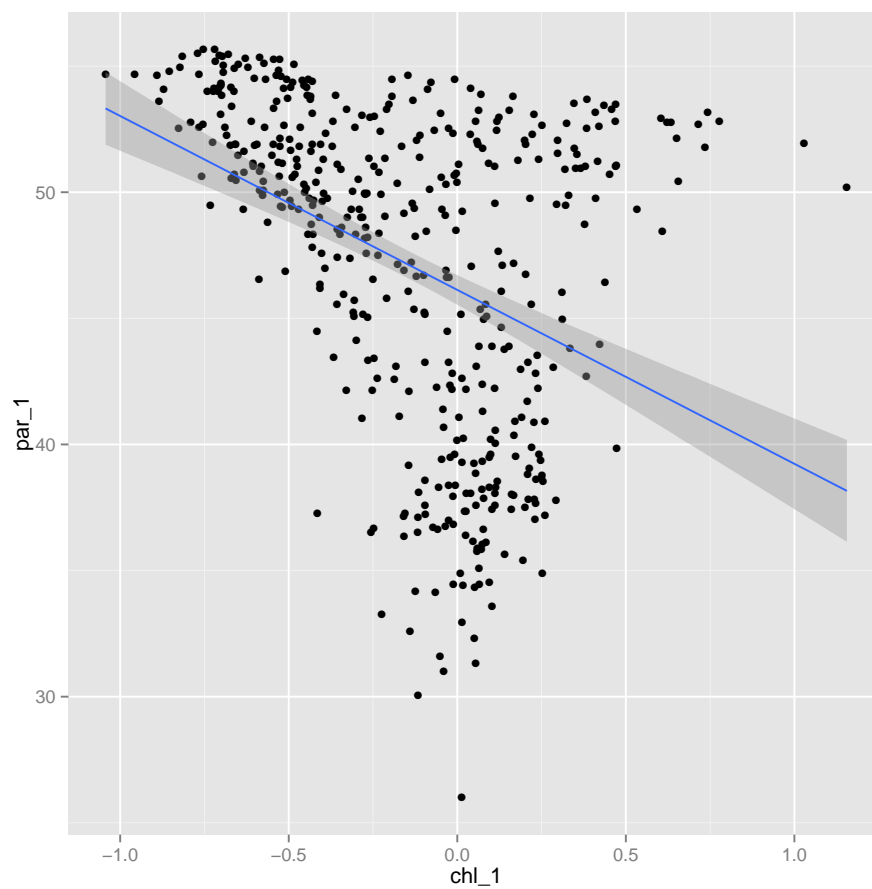
```
## Correlation coefficient  
cor(df$chl_1, df$ssh_4, use='complete.obs')  
  
## [1] -0.1034069  
  
ccf(df$chl_1, df$ssh_4, na.action=na.pass)
```





## 1.6 Chlorophyll and PAR

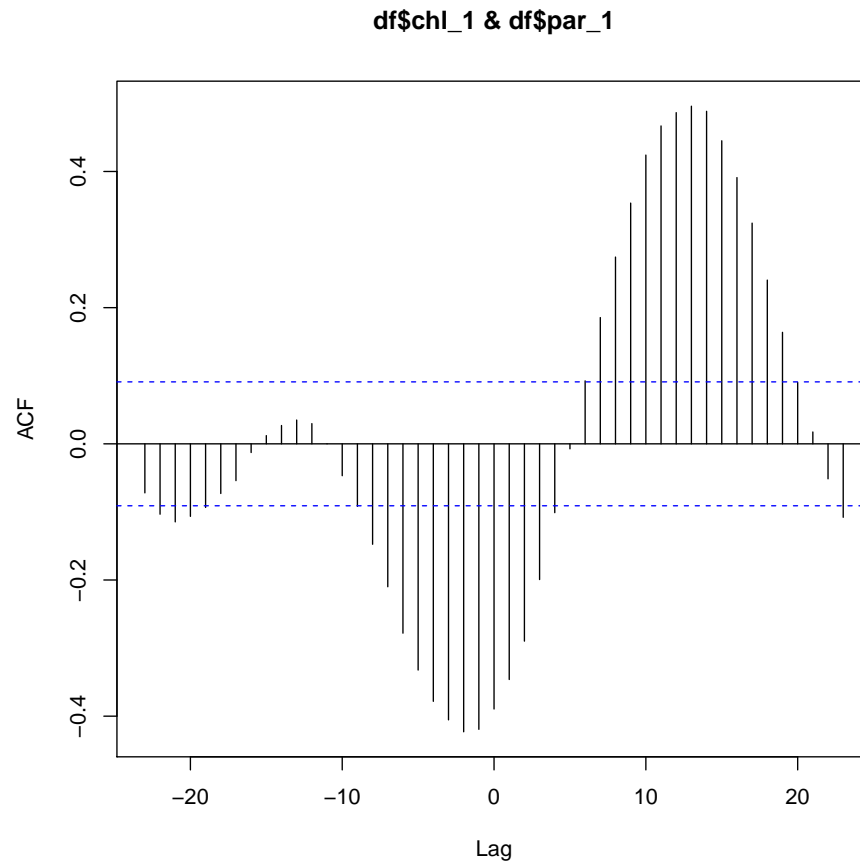
```
qplot(chl_1, par_1, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



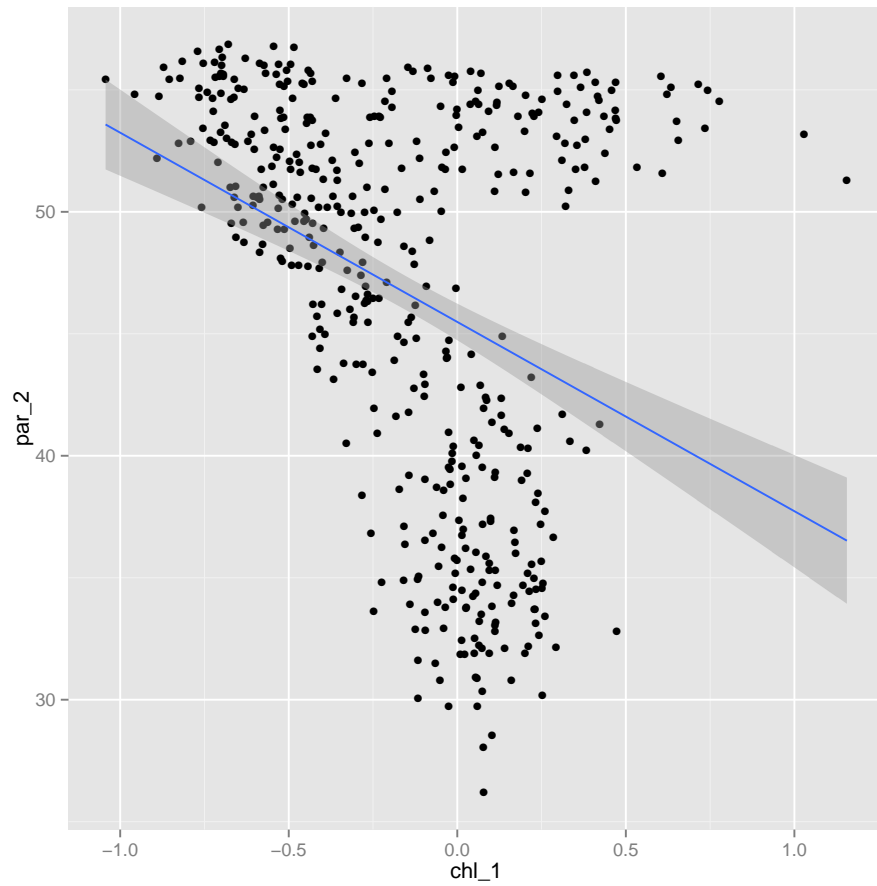
```
## Correlation coefficient
cor(df$chl_1, df$par_1, use='complete.obs')

## [1] -0.3894663

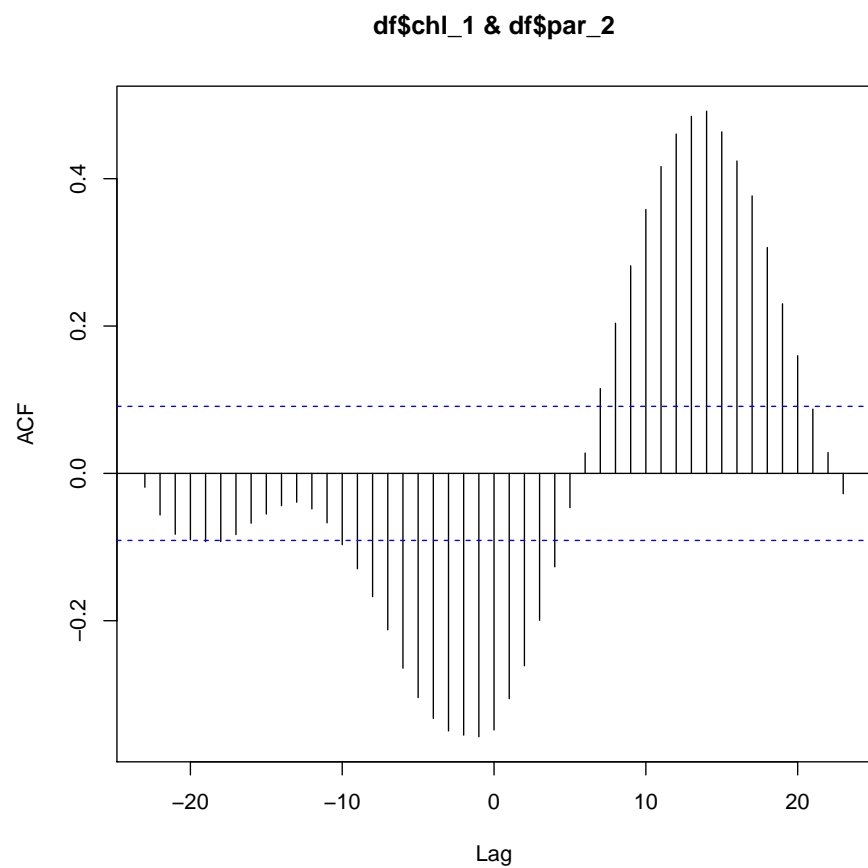
ccf(df$chl_1, df$par_1, na.action=na.pass)
```



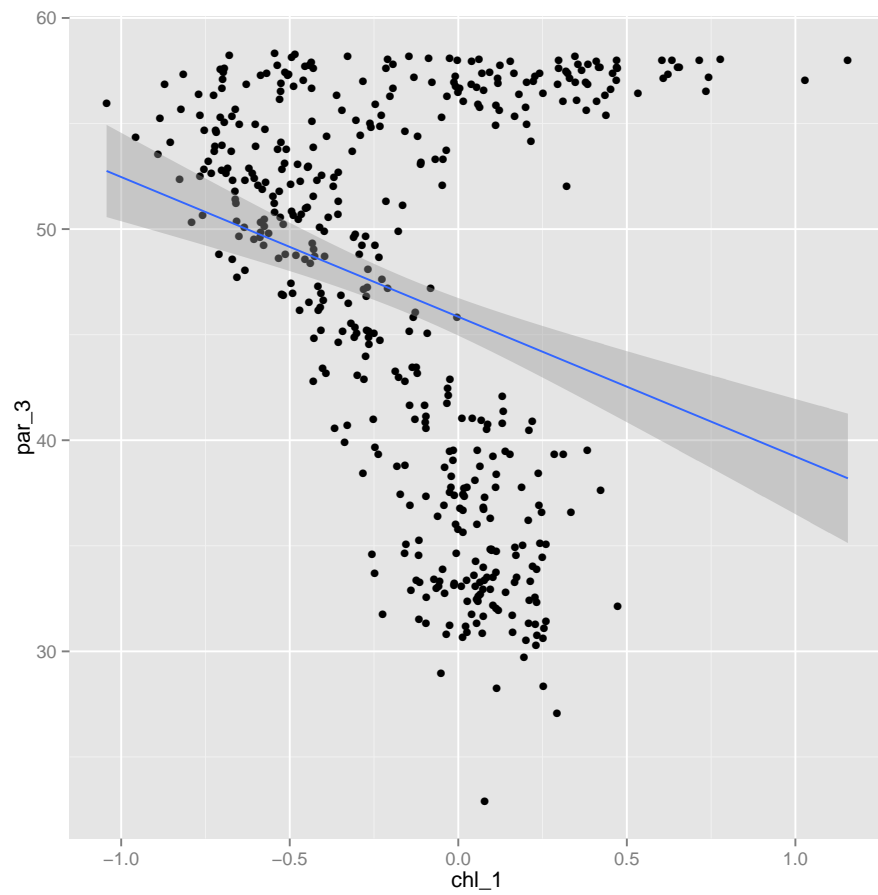
```
qplot(chl_1, par_2, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



```
## Correlation coefficient  
cor(df$chl_1, df$par_2, use='complete.obs')  
  
## [1] -0.3482226  
  
ccf(df$chl_1, df$par_2, na.action=na.pass)
```



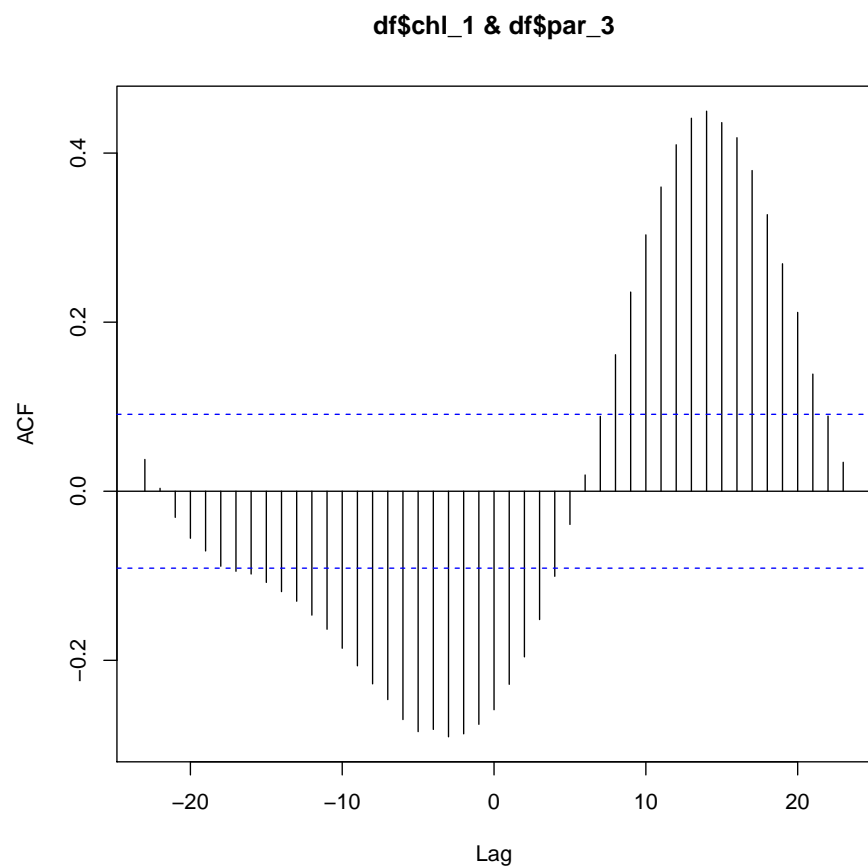
```
qplot(chl_1, par_3, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



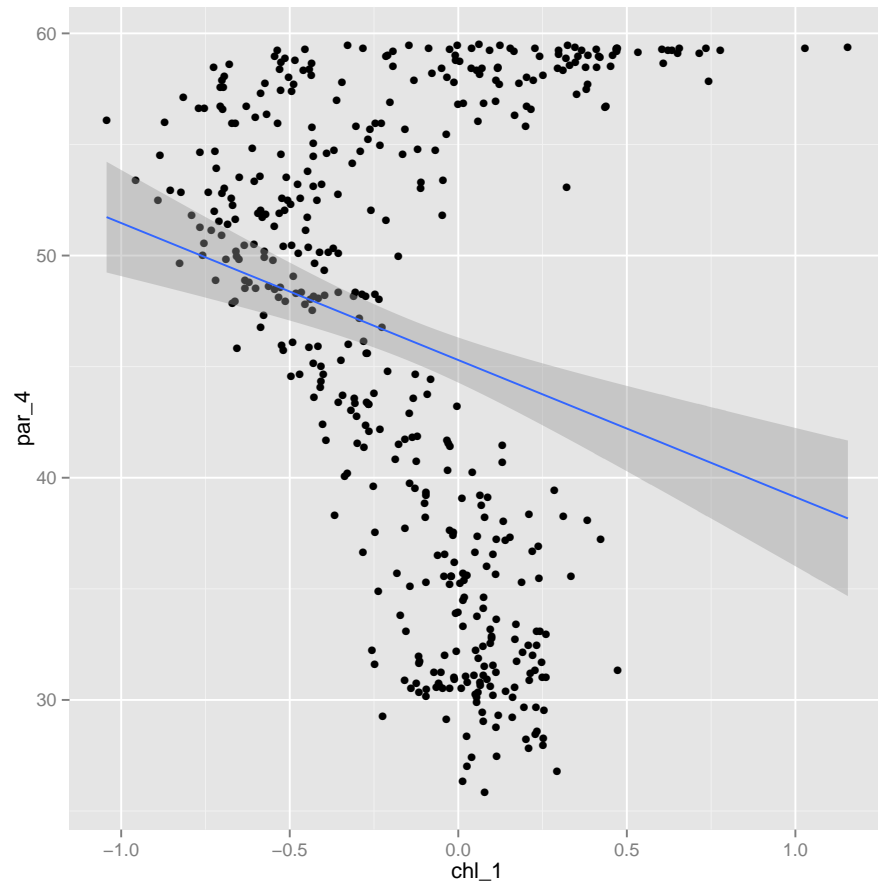
```
## Correlation coefficient
cor(df$chl_1, df$par_3, use='complete.obs')

## [1] -0.2584293

ccf(df$chl_1, df$par_3, na.action=na.pass)
```

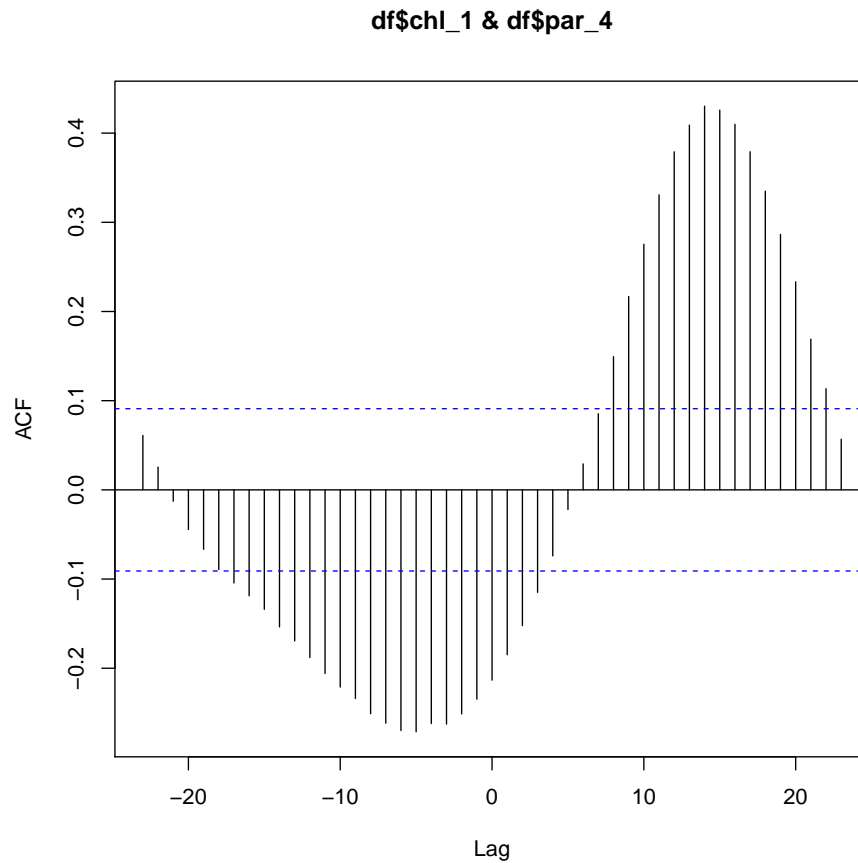


```
qplot(chl_1, par_4, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



```
## Correlation coefficient  
cor(df$chl_1, df$par_4, use='complete.obs')  
  
## [1] -0.213377  
  
ccf(df$chl_1, df$par_4, na.action=na.pass)
```

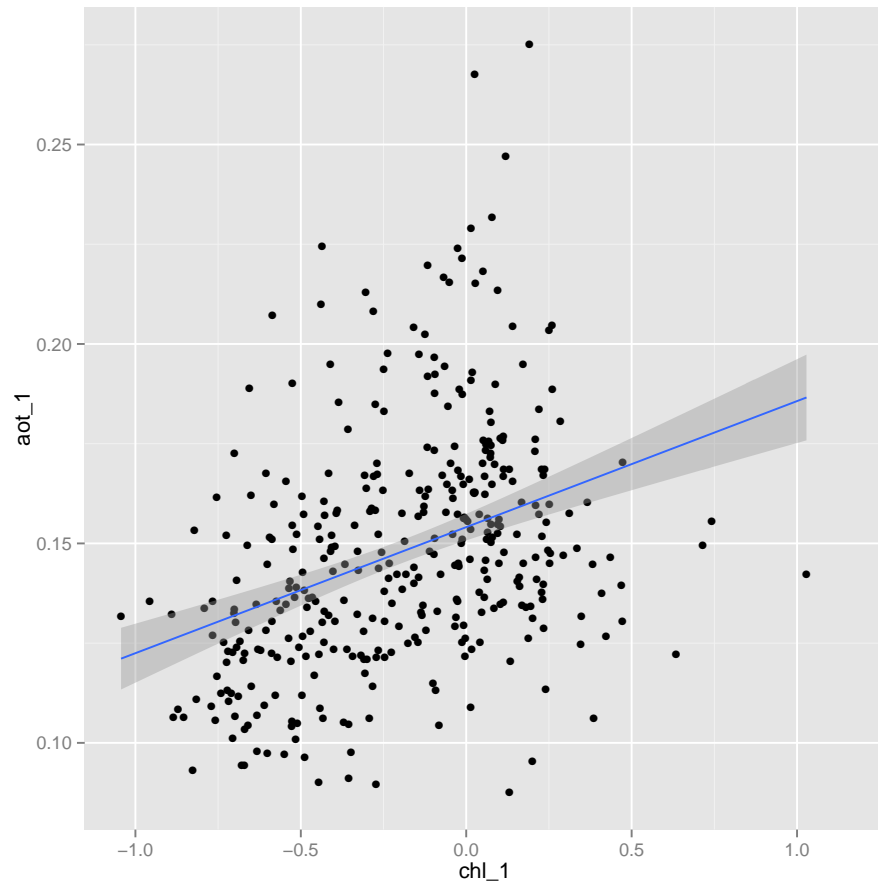




## 1.7 Chlorophyll and AOT

```
qplot(chl_1, aot_1, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)

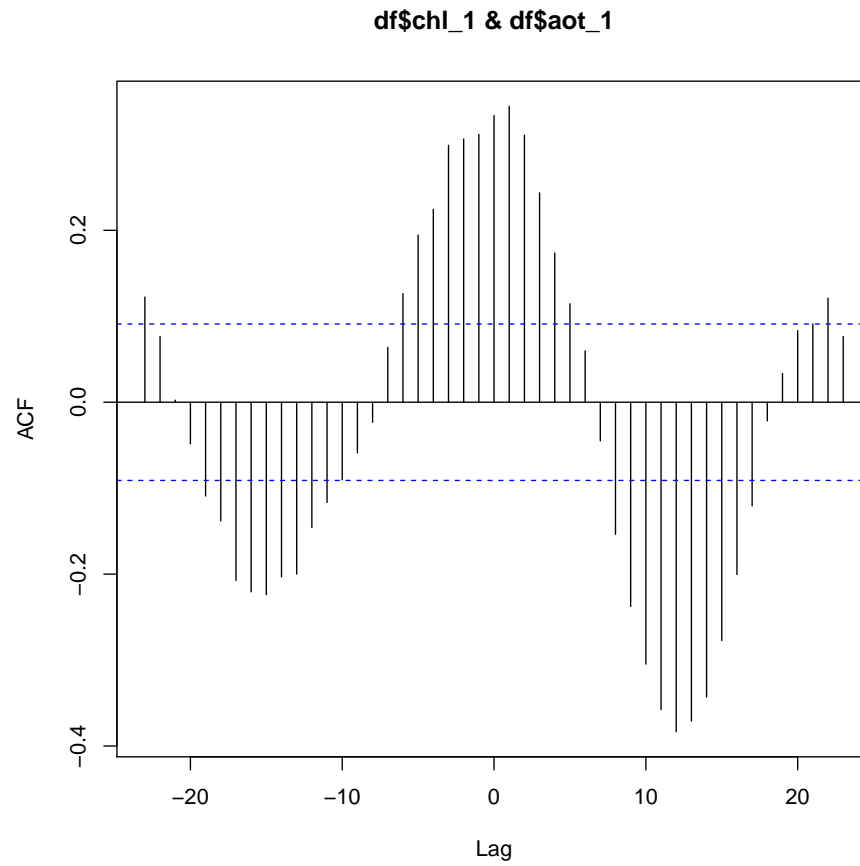
## Warning: Removed 90 rows containing missing values (stat_smooth).
## Warning: Removed 90 rows containing missing values (geom_point).
```



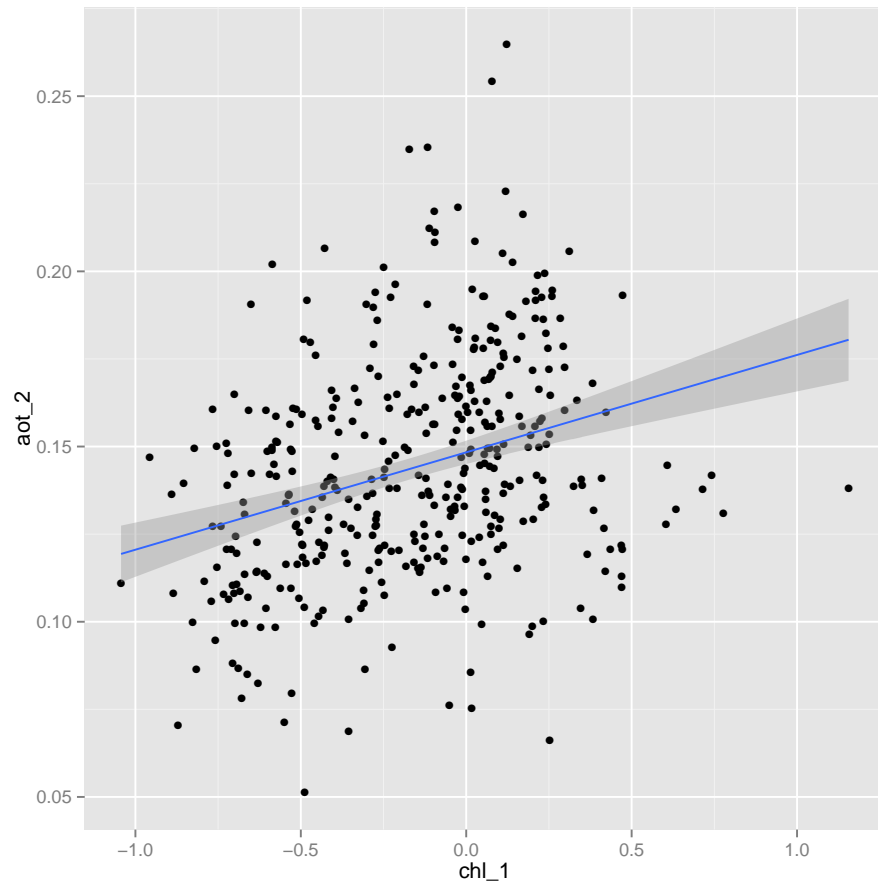
```
## Correlation coefficient
cor(df$chl_1, df$aot_1, use='complete.obs')

## [1] 0.3559261

ccf(df$chl_1, df$aot_1, na.action=na.pass)
```



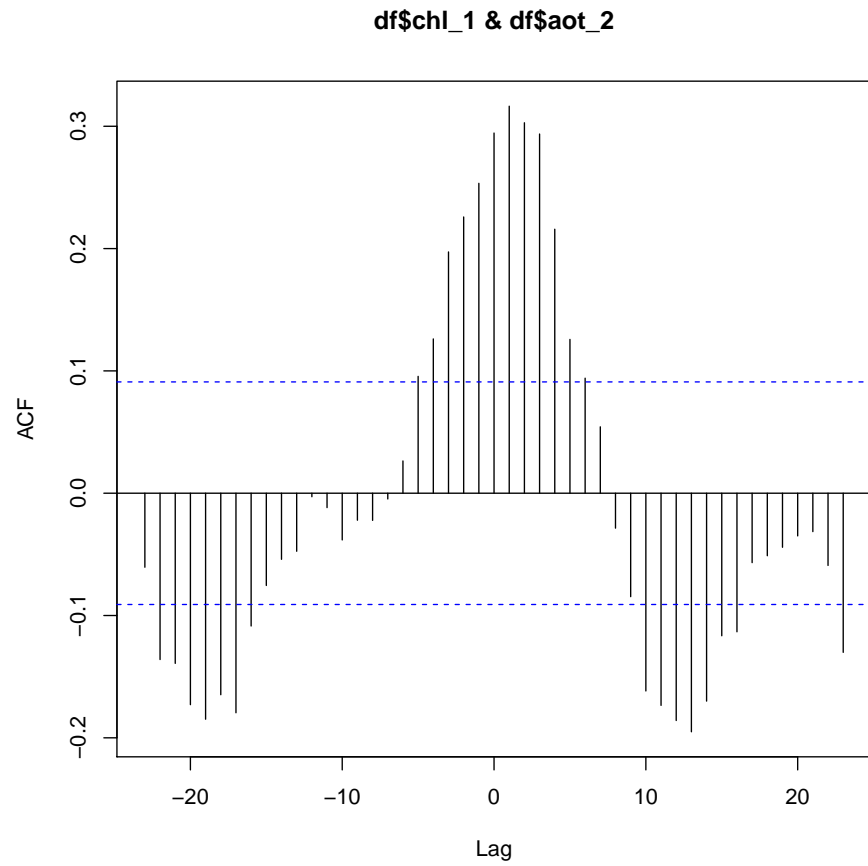
```
qplot(chl_1, aot_2, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
## Warning: Removed 62 rows containing missing values (stat.smooth).
## Warning: Removed 62 rows containing missing values (geom.point).
```



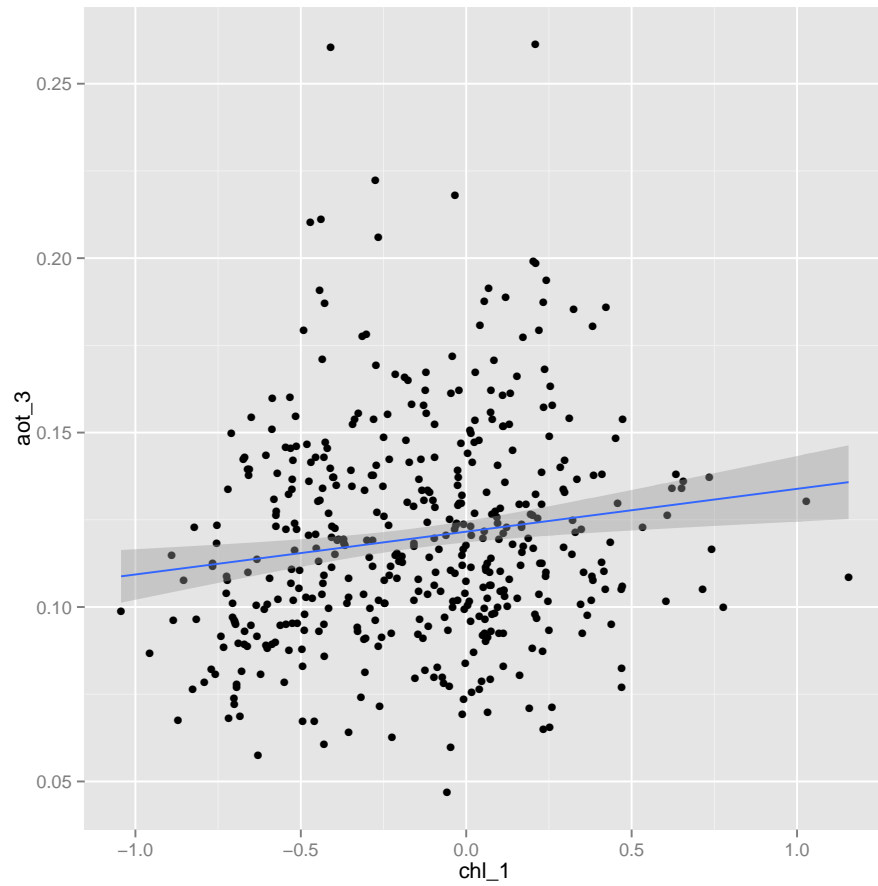
```
## Correlation coefficient
cor(df$chl_1, df$aot_2, use='complete.obs')

## [1] 0.3044926

ccf(df$chl_1, df$aot_2, na.action=na.pass)
```



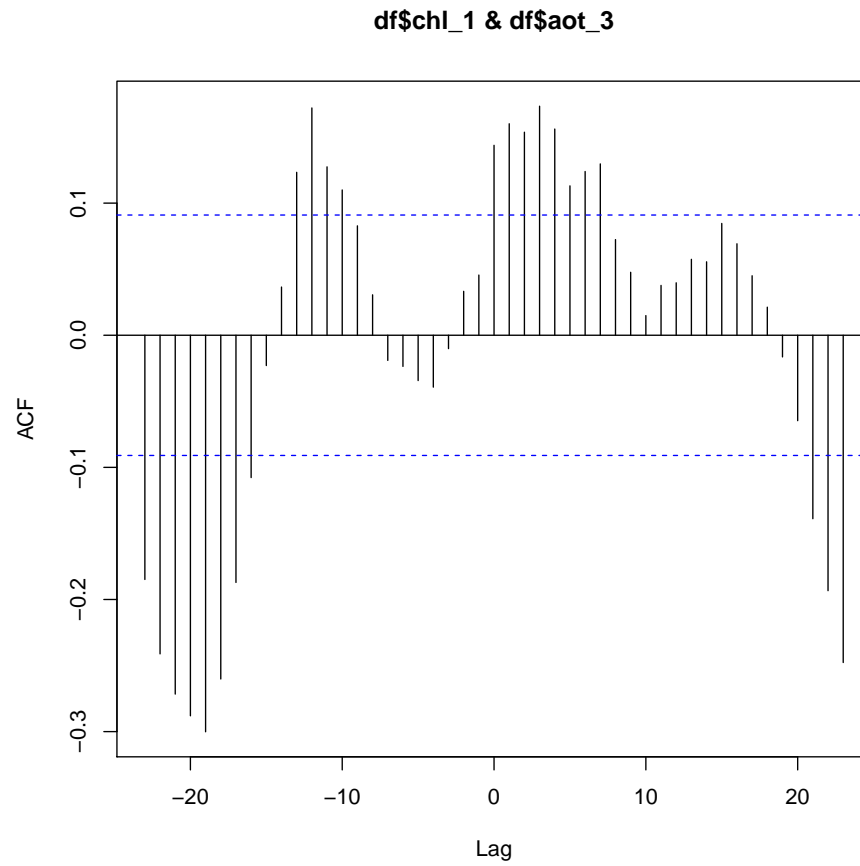
```
qplot(chl_1, aot_3, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
## Warning: Removed 9 rows containing missing values (stat_smooth).
## Warning: Removed 9 rows containing missing values (geom_point).
```



```
## Correlation coefficient
cor(df$chl_1, df$aot_3, use='complete.obs')

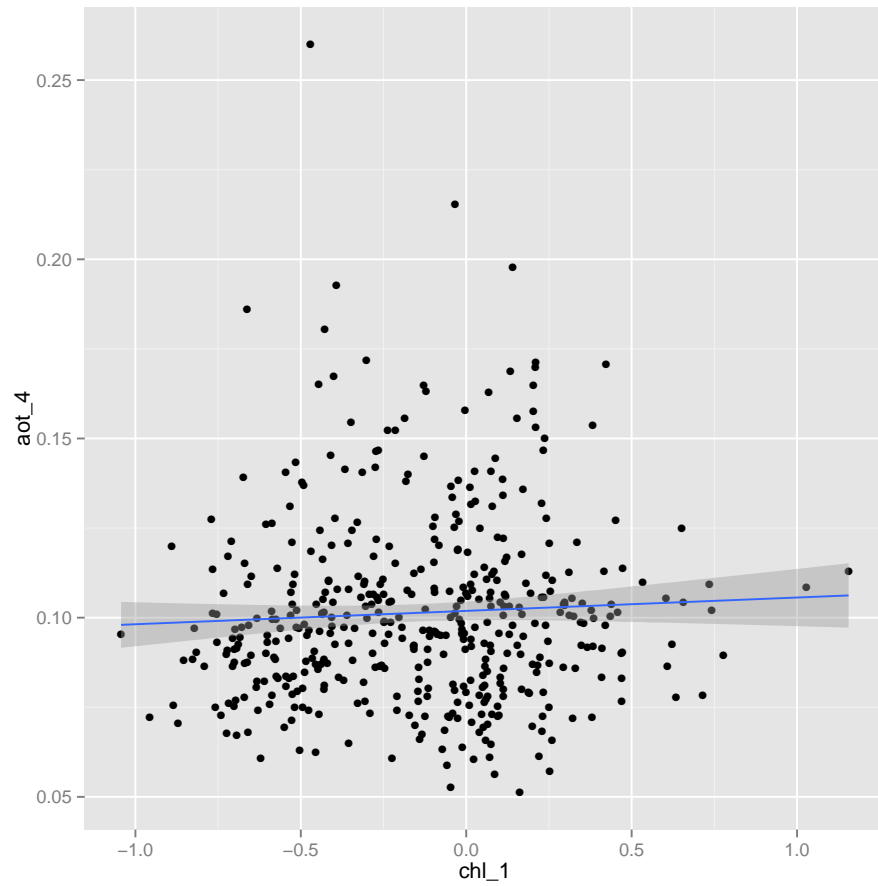
## [1] 0.1435728

ccf(df$chl_1, df$aot_3, na.action=na.pass)
```



```
qplot(chl_1, aot_4, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)

## Warning: Removed 5 rows containing missing values (stat_smooth).
## Warning: Removed 5 rows containing missing values (geom_point).
```

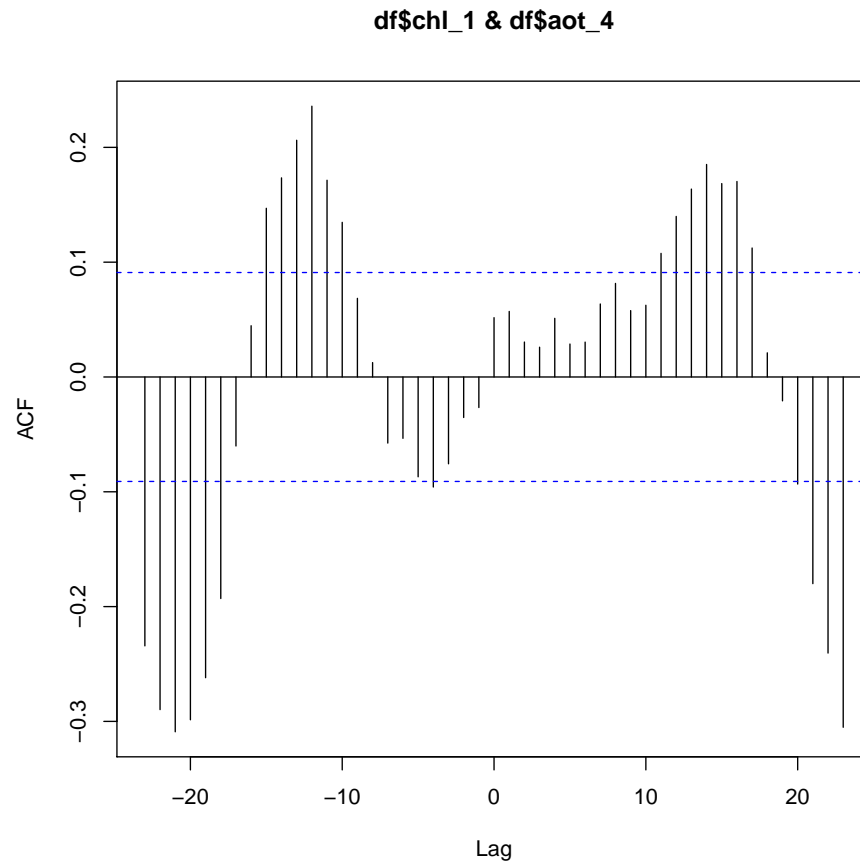


```
## Correlation coefficient
cor(df$chl_1, df$aot_4, use='complete.obs')

## [1] 0.05161445

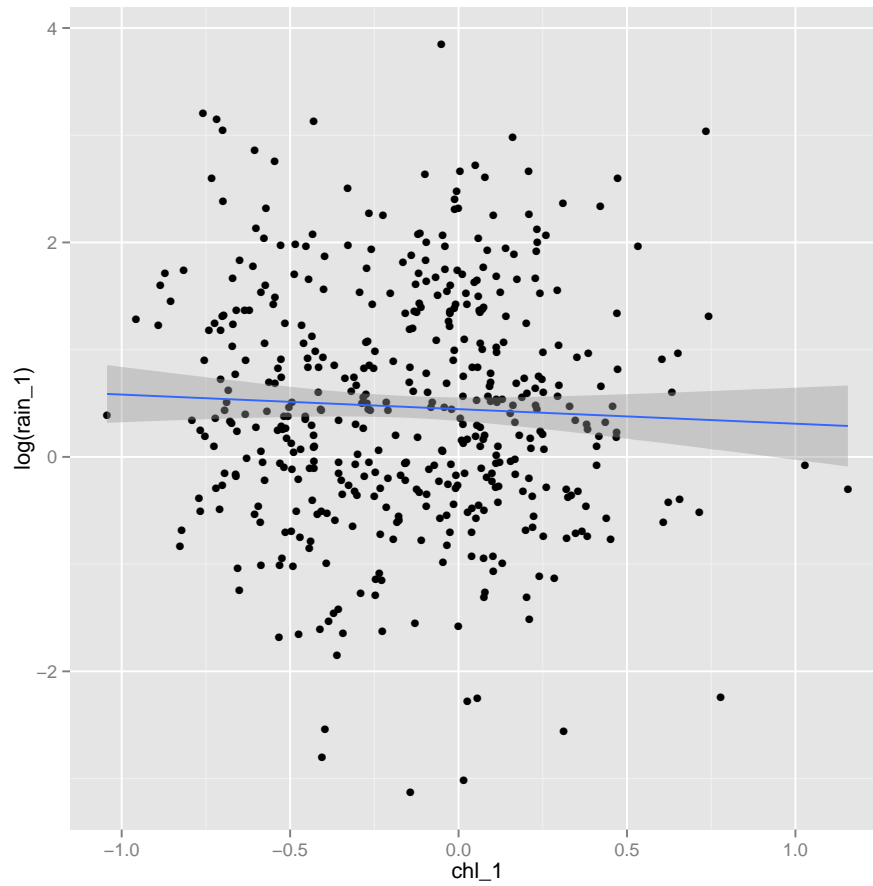
ccf(df$chl_1, df$aot_4, na.action=na.pass)
```





## 1.8 Chlorophyll and Rain

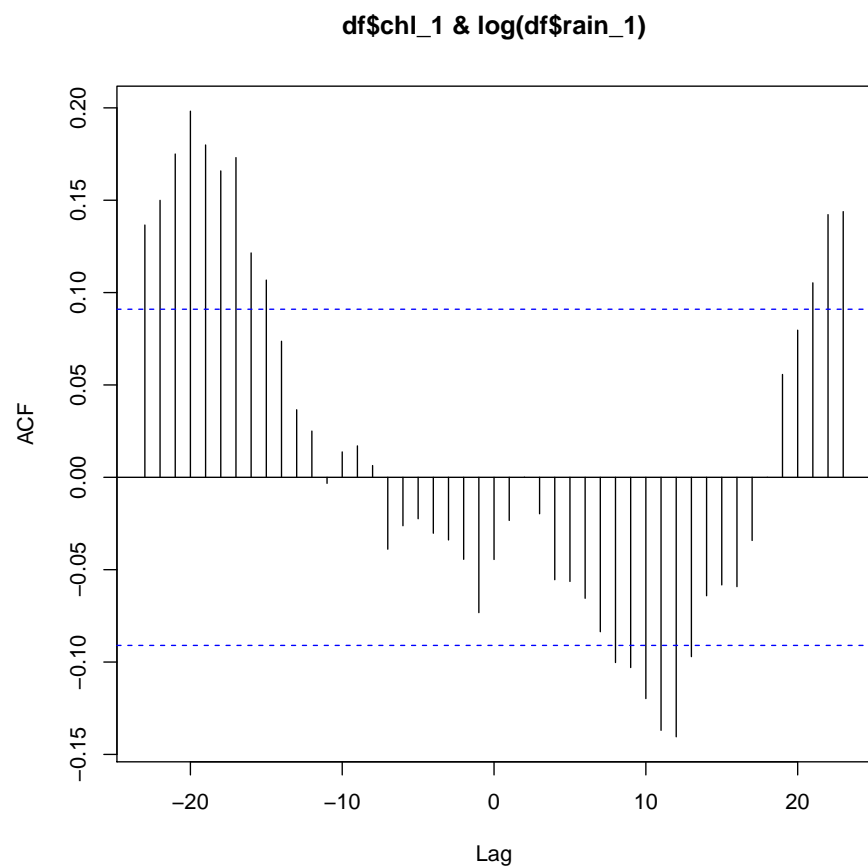
```
qplot(chl_1, log(rain_1), data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



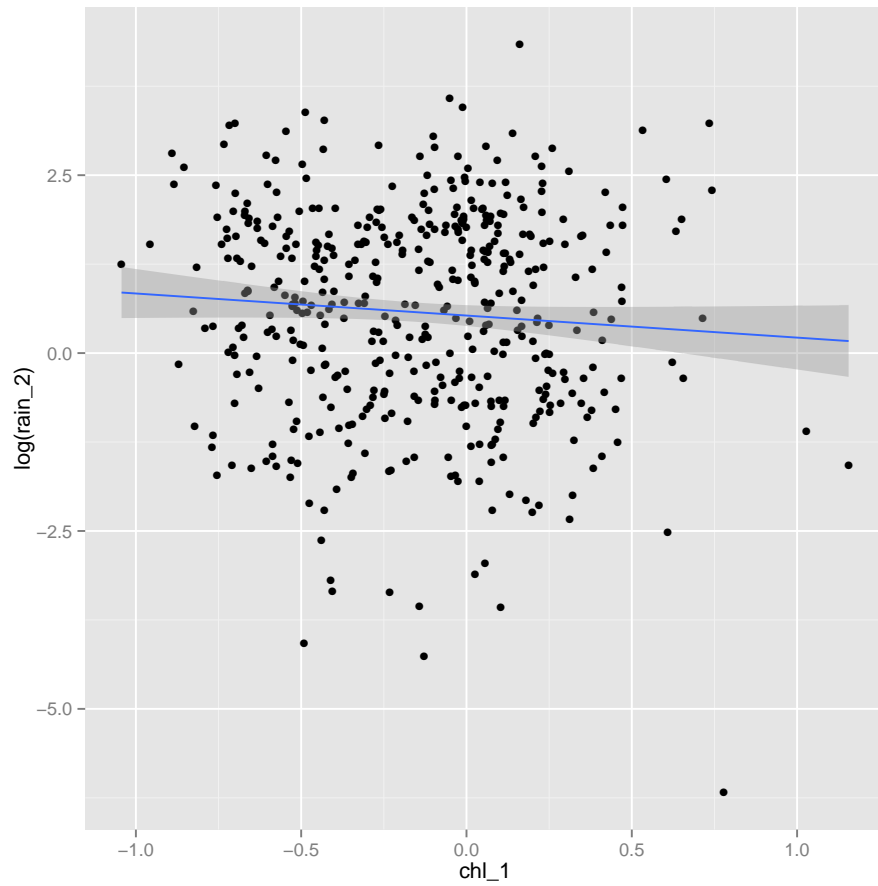
```
## Correlation coefficient
cor(df$chl_1, log(df$rain_1), use='complete.obs')

## [1] -0.04447456

ccf(df$chl_1, log(df$rain_1), na.action=na.pass)
```



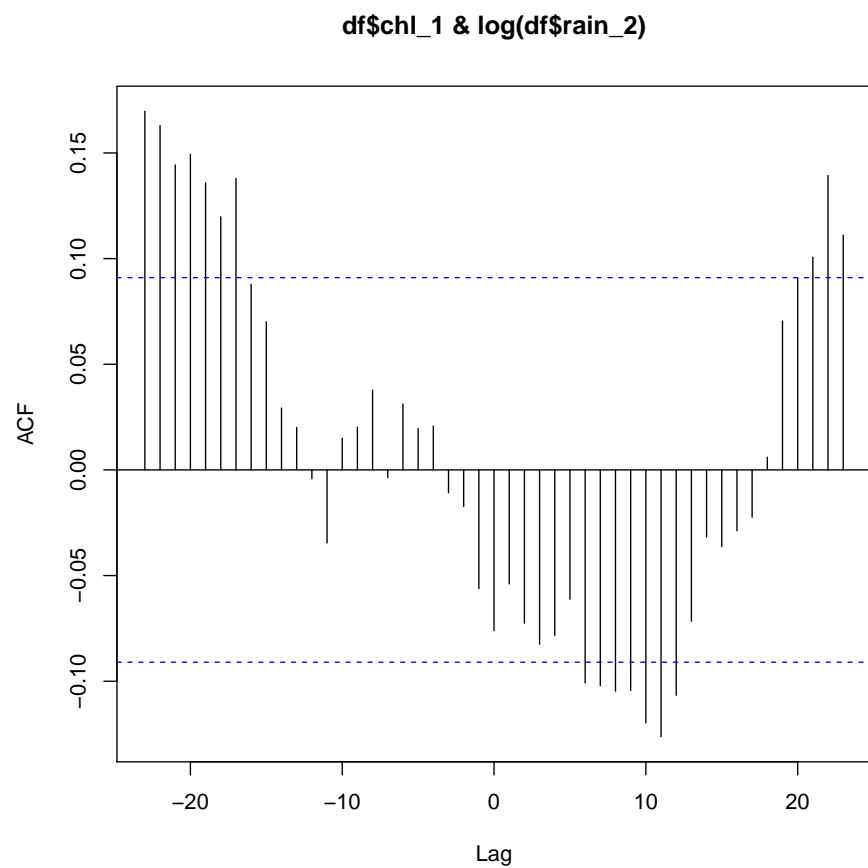
```
qplot(chl_1, log(rain_2), data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



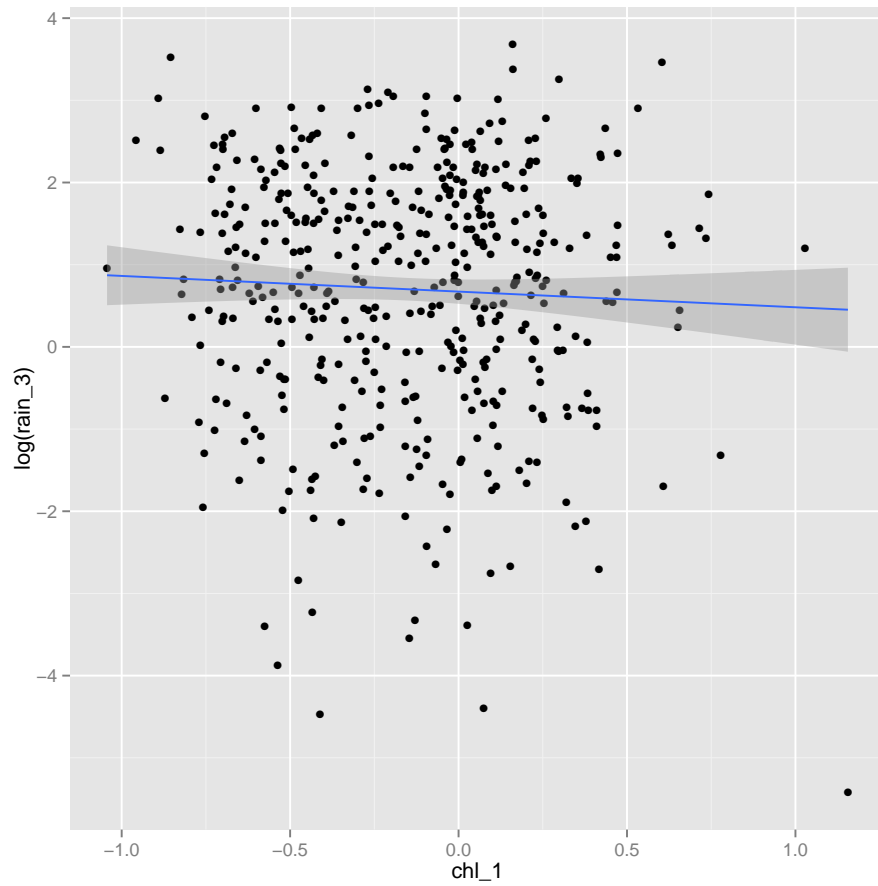
```
## Correlation coefficient
cor(df$chl_1, log(df$rain_2), use='complete.obs')

## [1] -0.07603666

ccf(df$chl_1, log(df$rain_2), na.action=na.pass)
```



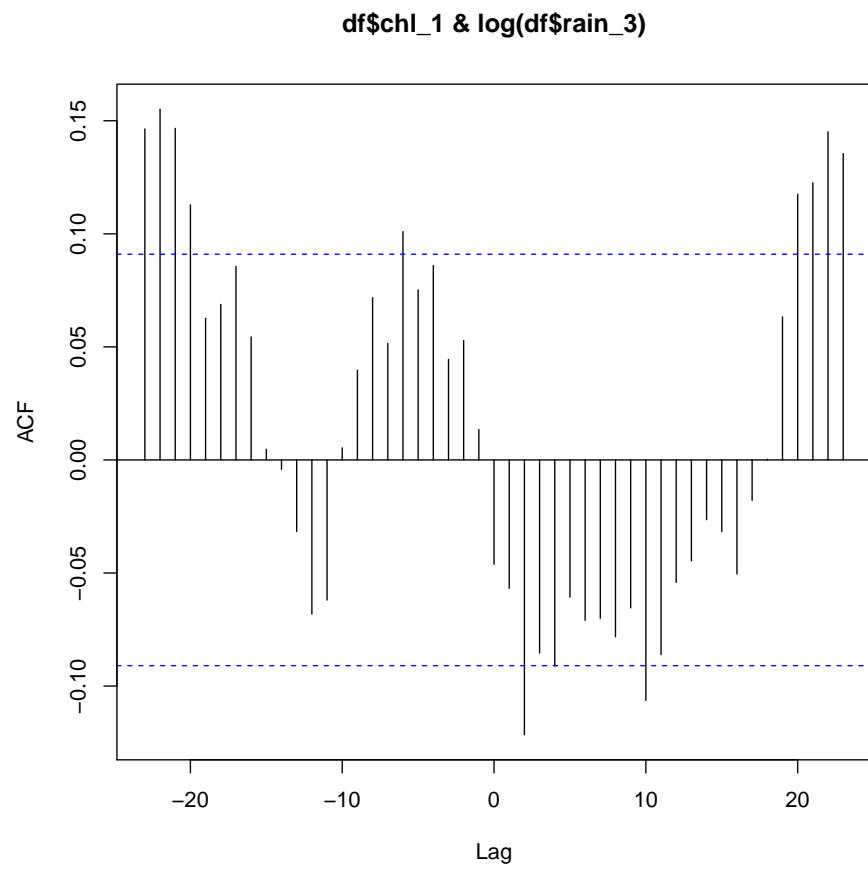
```
qplot(chl_1, log(rain_3), data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



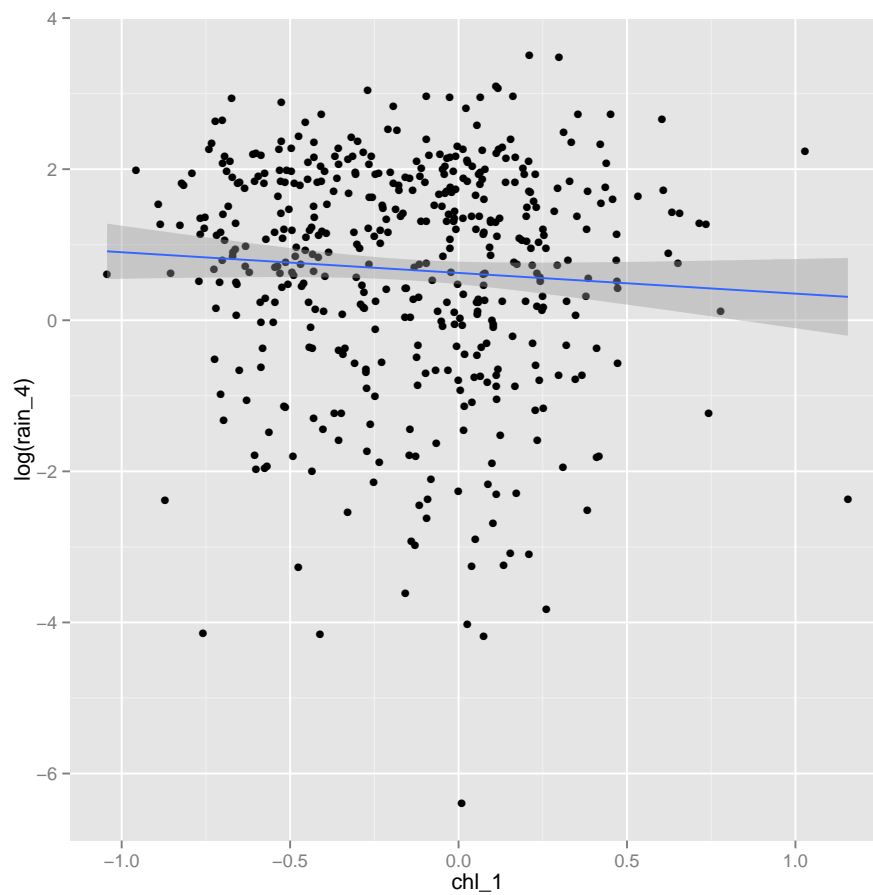
```
## Correlation coefficient
cor(df$chl_1, log(df$rain_3), use='complete.obs')

## [1] -0.046129

ccf(df$chl_1, log(df$rain_3), na.action=na.pass)
```

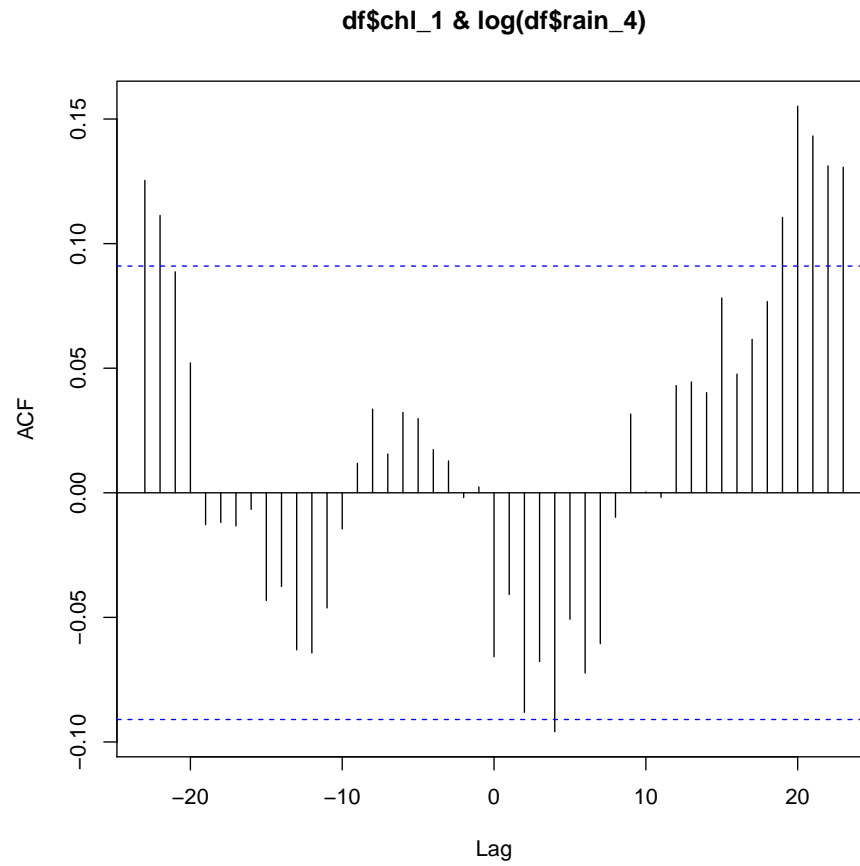


```
qplot(chl_1, log(rain_4), data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



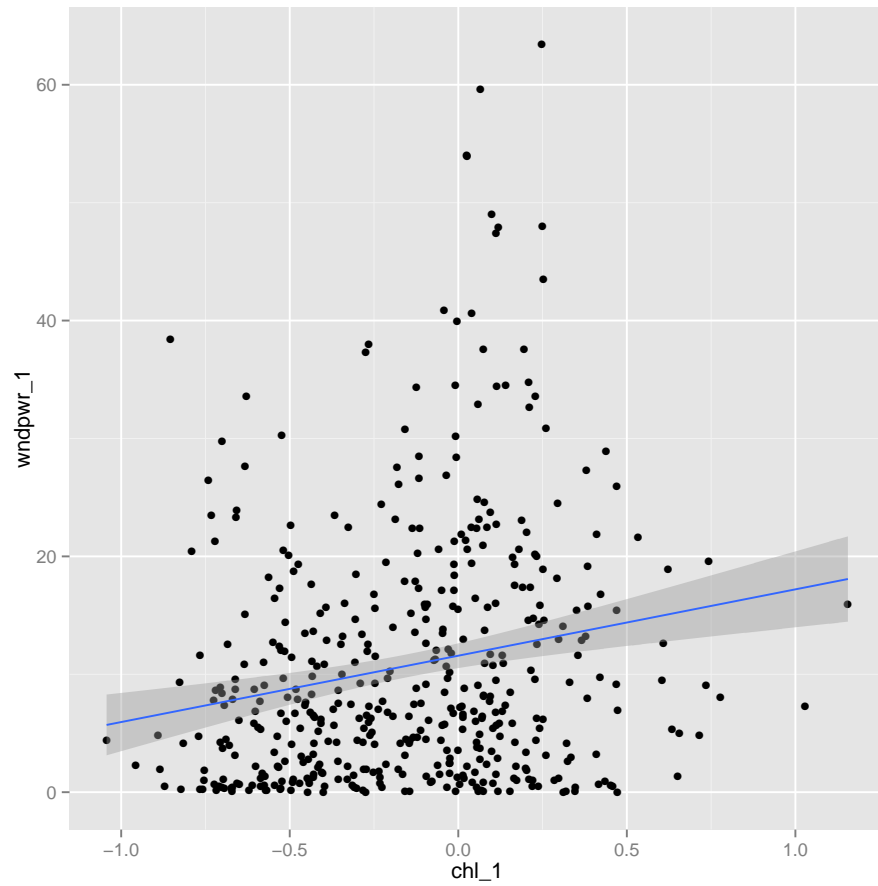
```
## Correlation coefficient  
cor(df$chl_1, log(df$rain_4), use='complete.obs')  
  
## [1] -0.0657788  
  
ccf(df$chl_1, log(df$rain_4), na.action=na.pass)
```





## 1.9 Chlorophyll and wind

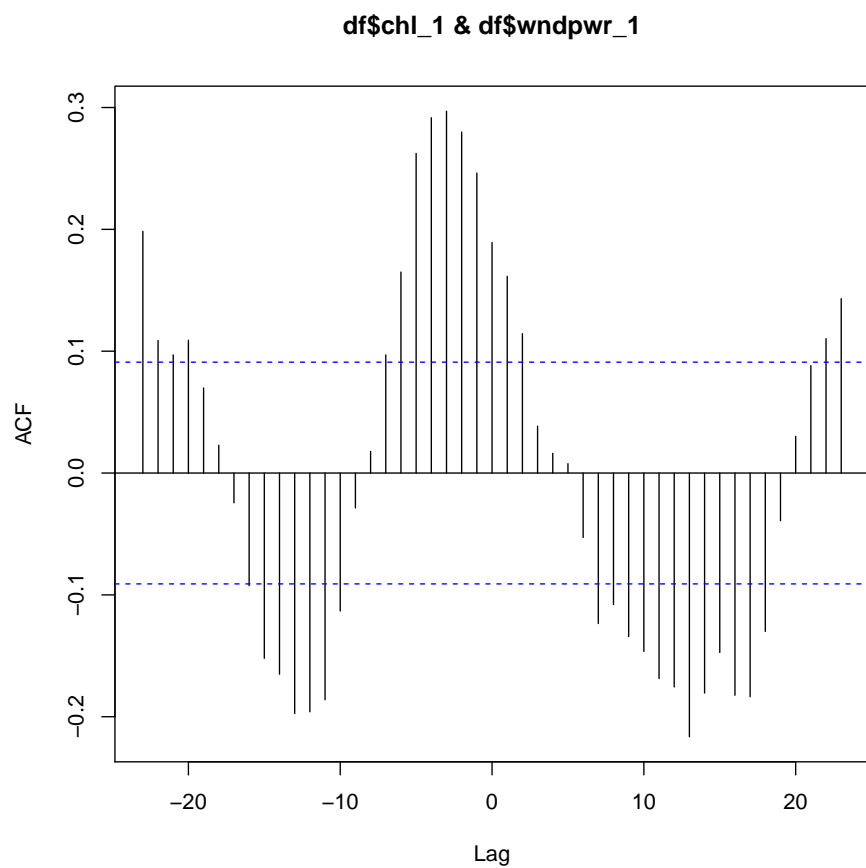
```
qplot(chl_1, wndpwr_1, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



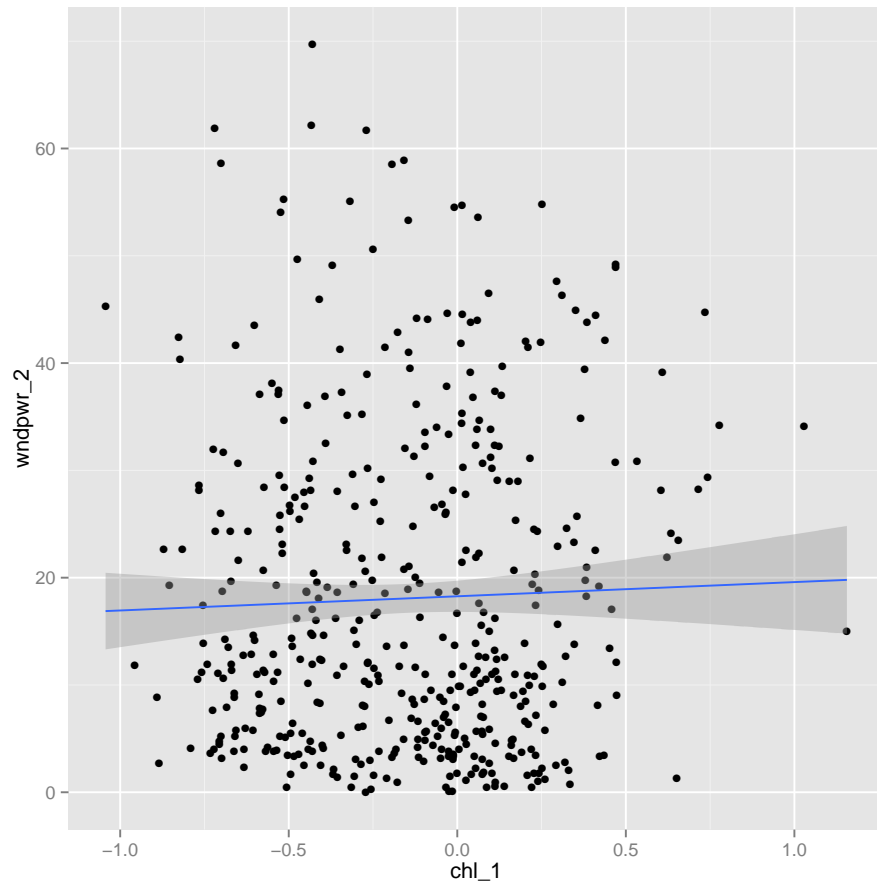
```
## Correlation coefficient
cor(df$chl_1, df$wndpwr_1, use='complete.obs')

## [1] 0.1892394

ccf(df$chl_1, df$wndpwr_1, na.action=na.pass)
```



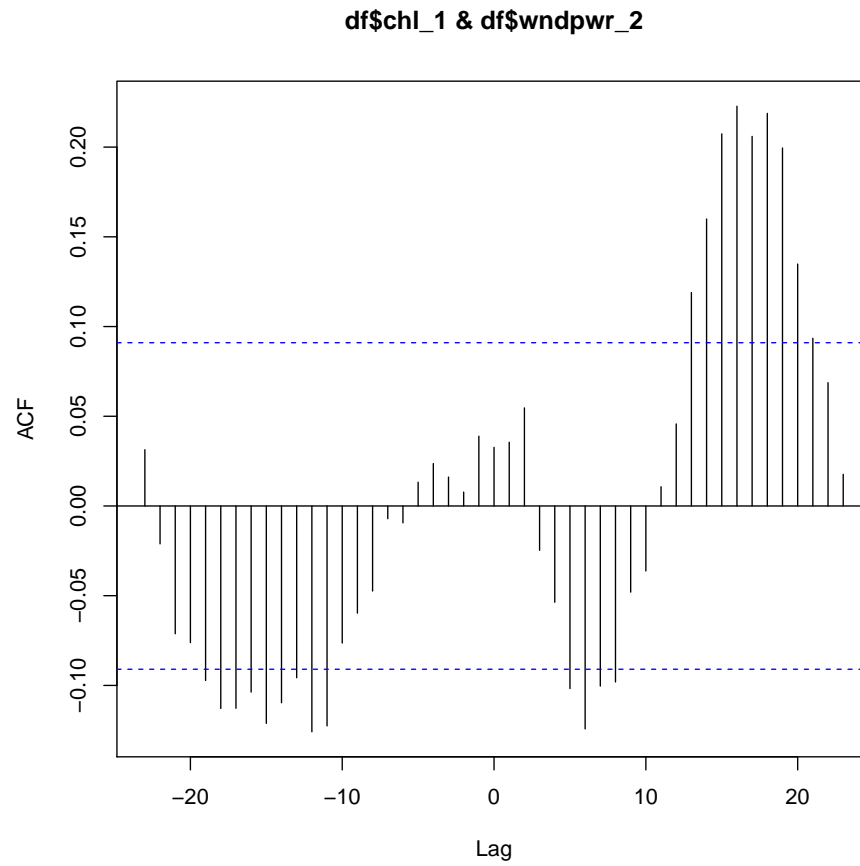
```
qplot(chl_1, wndpwr_2, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



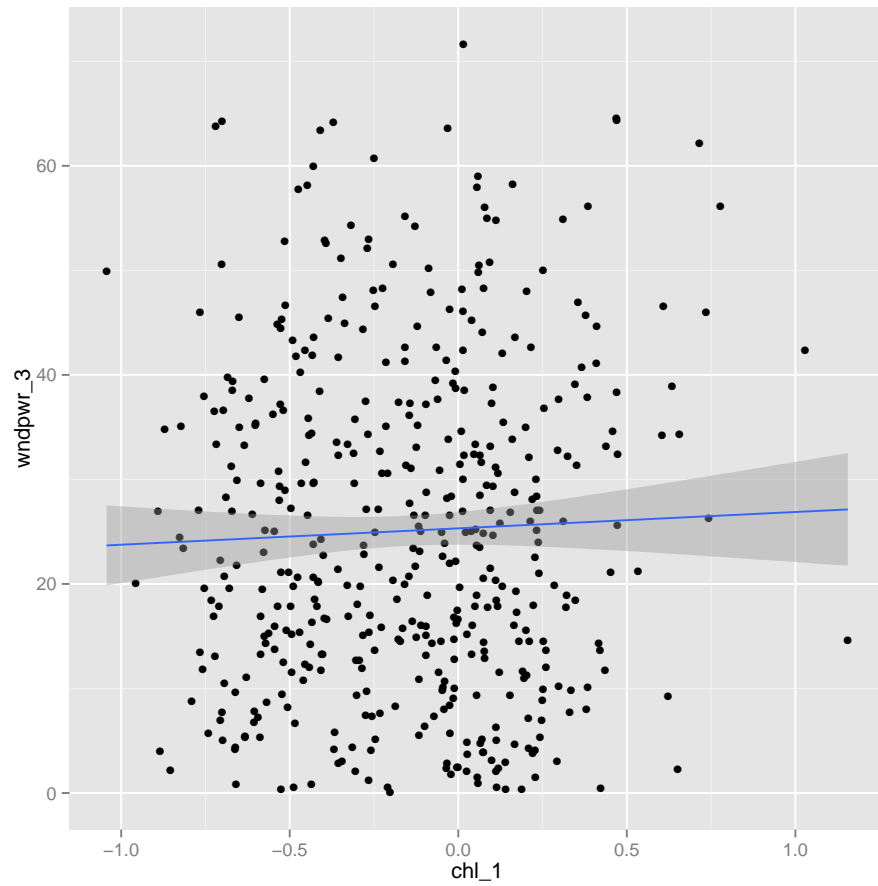
```
## Correlation coefficient
cor(df$chl_1, df$wndpwr_2, use='complete.obs')

## [1] 0.03268013

ccf(df$chl_1, df$wndpwr_2, na.action=na.pass)
```



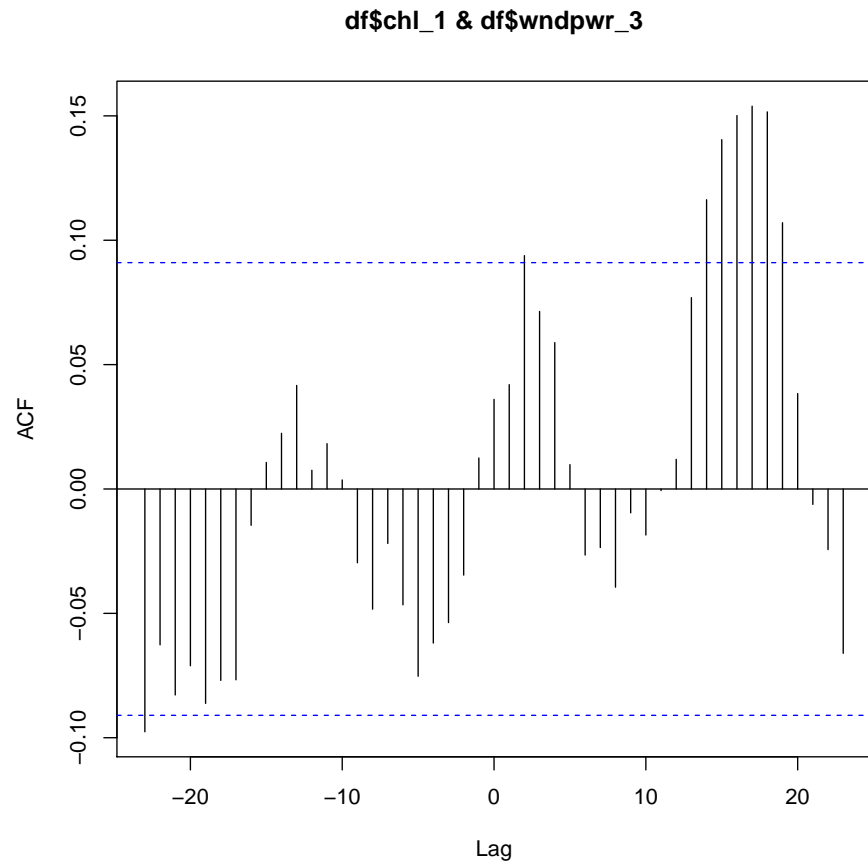
```
qplot(chl_1, wndpwr_3, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



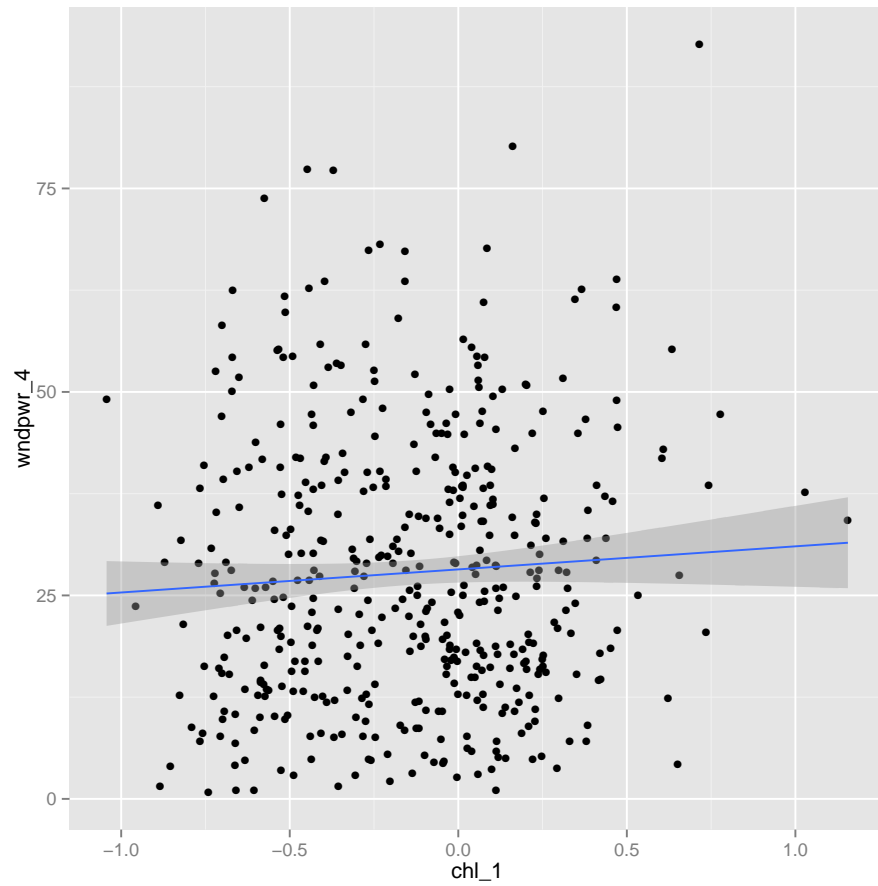
```
## Correlation coefficient
cor(df$chl_1, df$wndpwr_3, use='complete.obs')

## [1] 0.03603483

ccf(df$chl_1, df$wndpwr_3, na.action=na.pass)
```

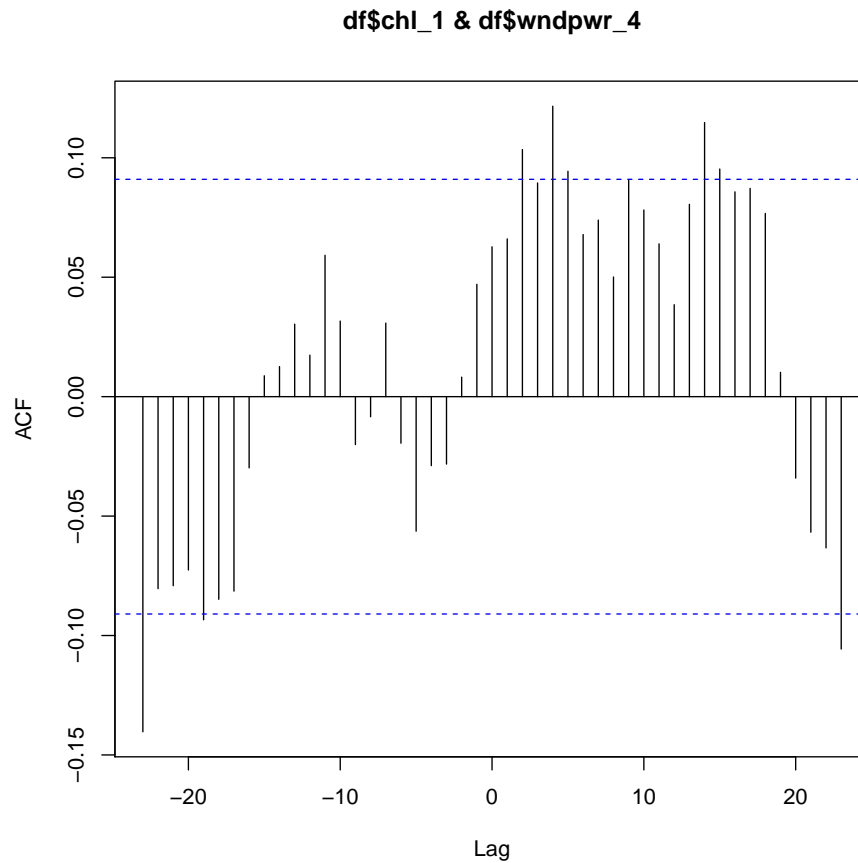


```
qplot(chl_1, wndpwr_4, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



```
## Correlation coefficient  
cor(df$chl_1, df$wndpwr_4, use='complete.obs')  
  
## [1] 0.06274202  
  
ccf(df$chl_1, df$wndpwr_4, na.action=na.pass)
```

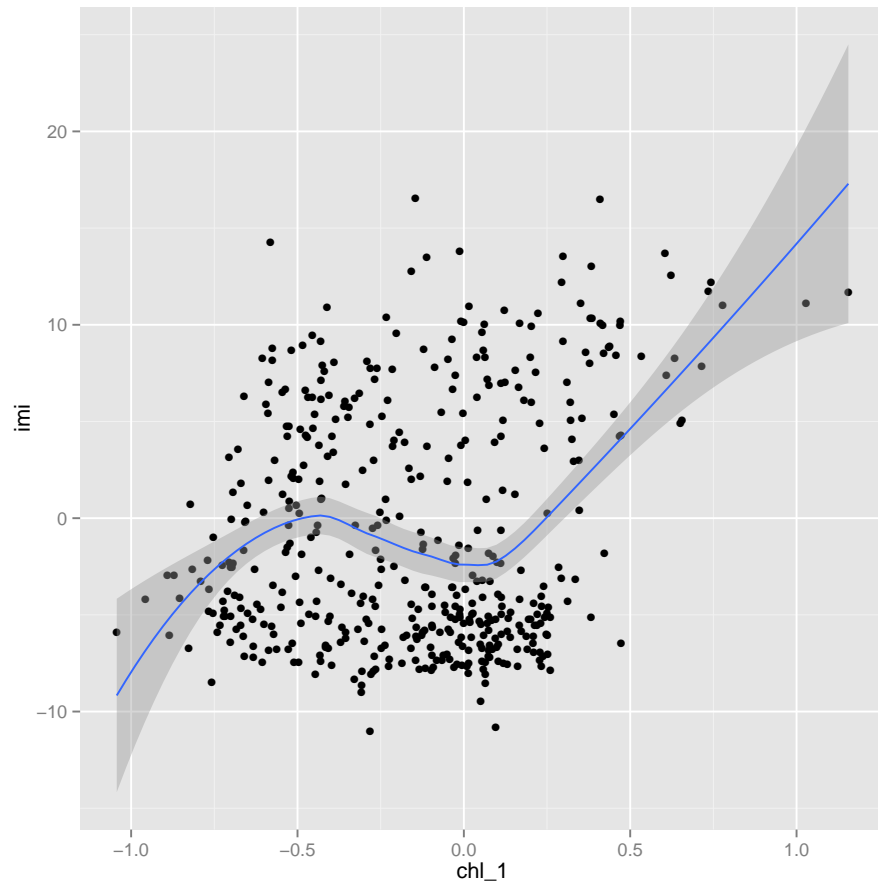




## 1.10 Chlorophyll and indices

```
qplot(chl_1, imi, data=df, geom=c('point', 'smooth'))#, method='lm', formula=y~x)

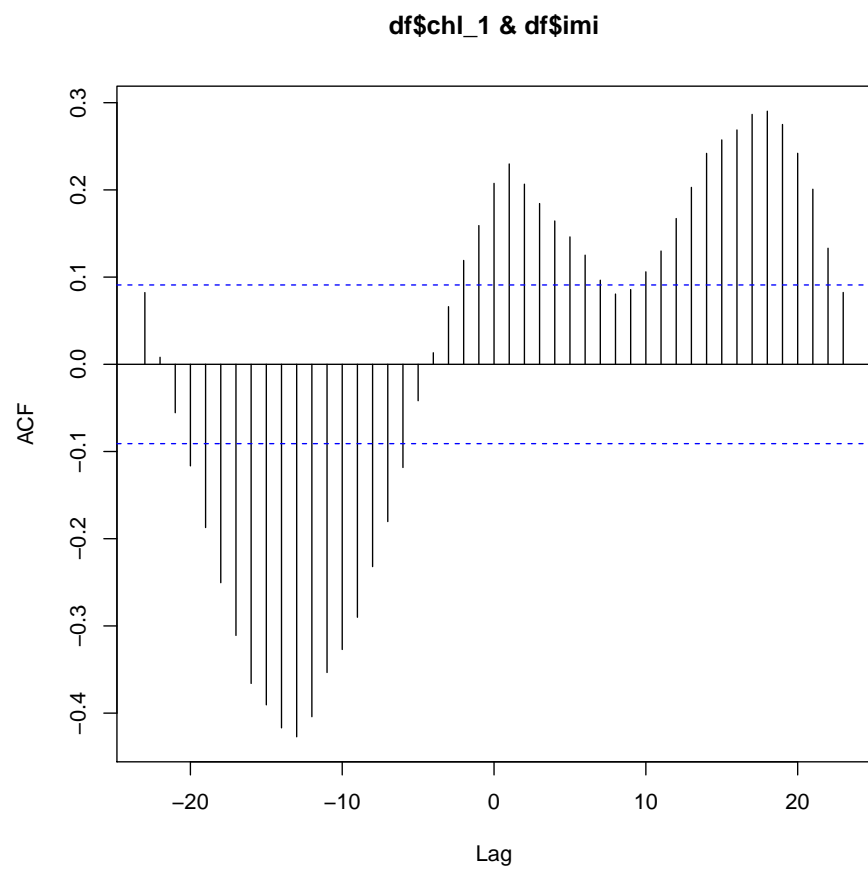
## geom_smooth: method="auto" and size of largest group is <1000,
so using loess. Use 'method = x' to change the smoothing method.
```



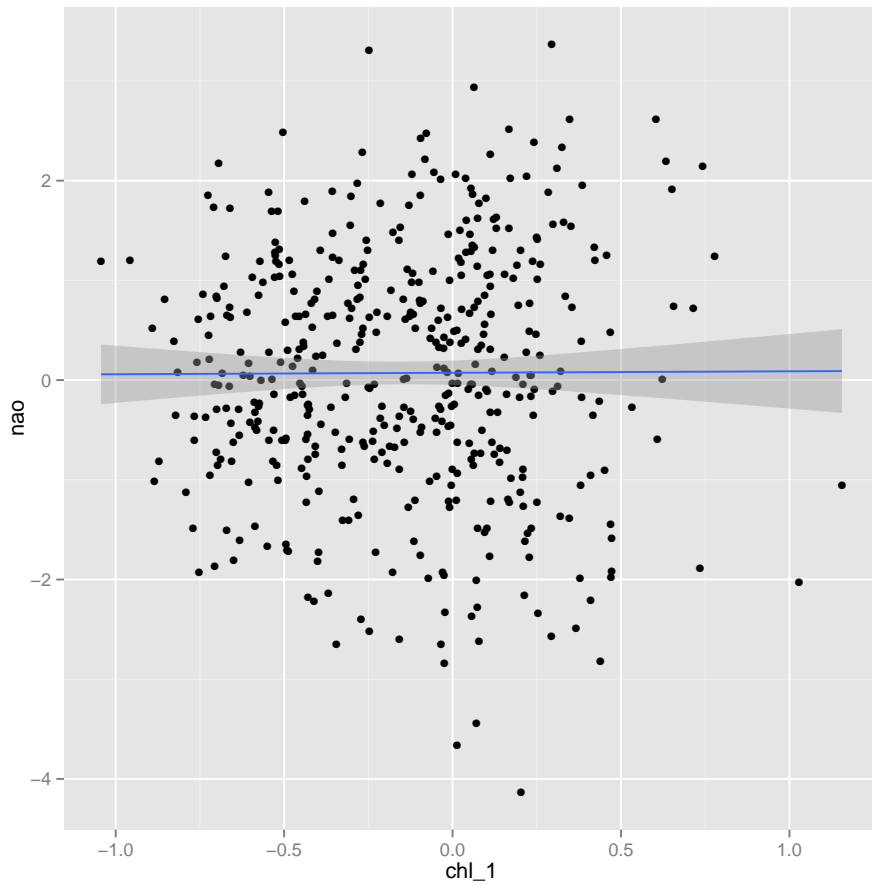
```
## Correlation coefficient
cor(df$chl_1, df$imi, use='complete.obs')

## [1] 0.2074861

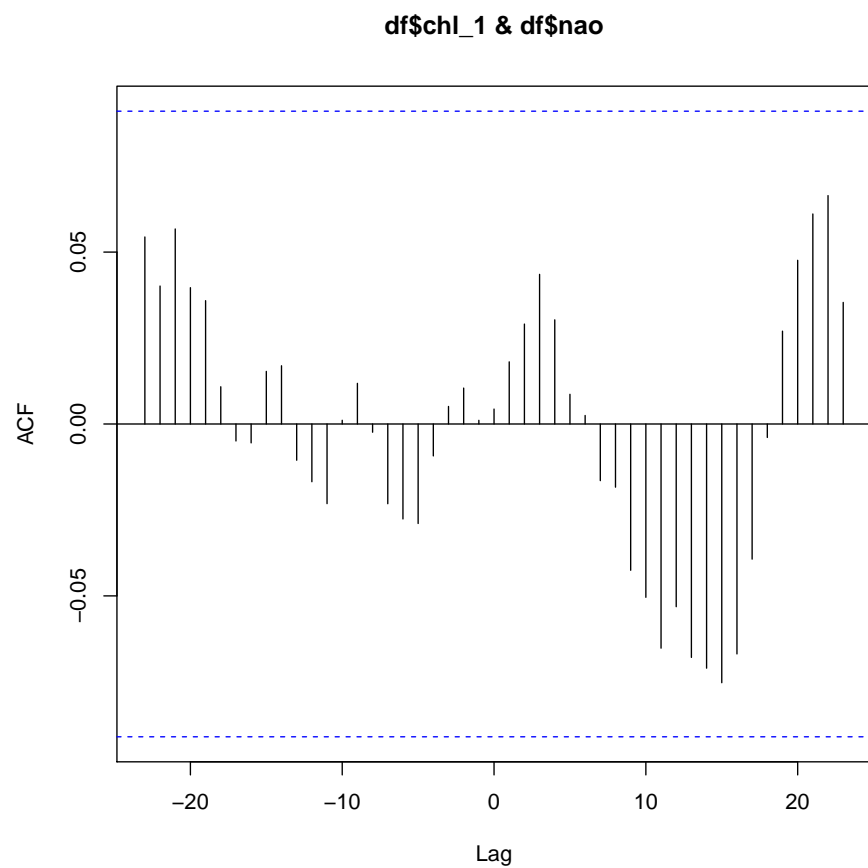
ccf(df$chl_1, df$imi, na.action=na.pass)
```



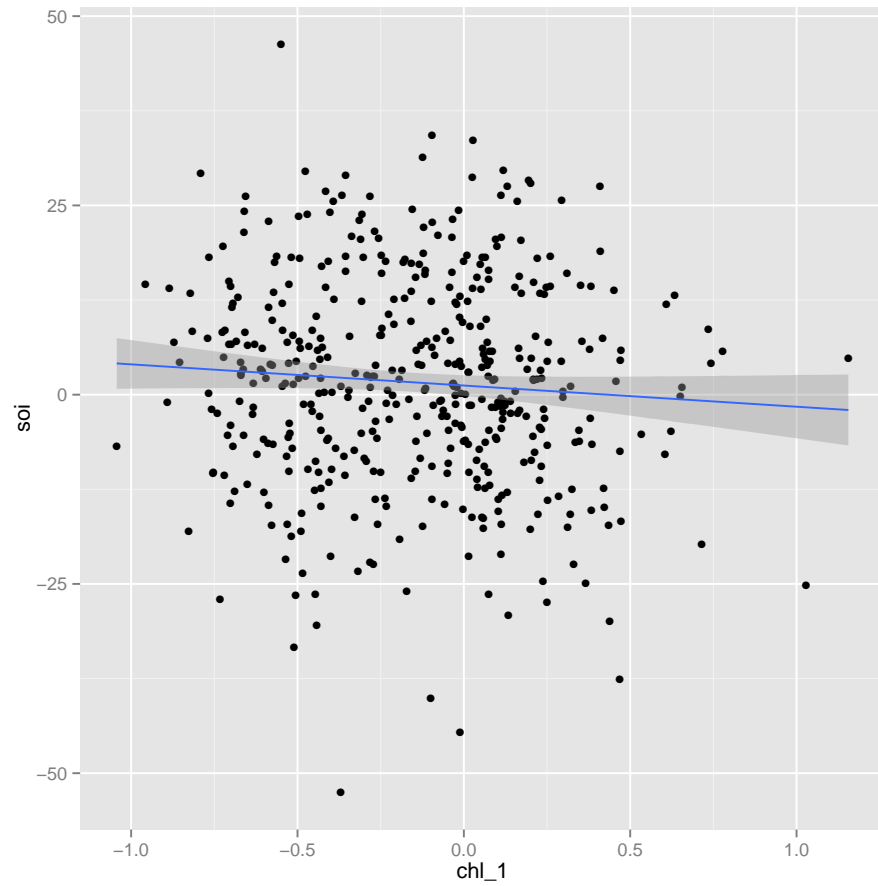
```
qplot(chl_1, nao, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



```
## Correlation coefficient  
cor(df$chl_1, df$nao, use='complete.obs')  
  
## [1] 0.004341686  
  
ccf(df$chl_1, df$nao, na.action=na.pass)
```



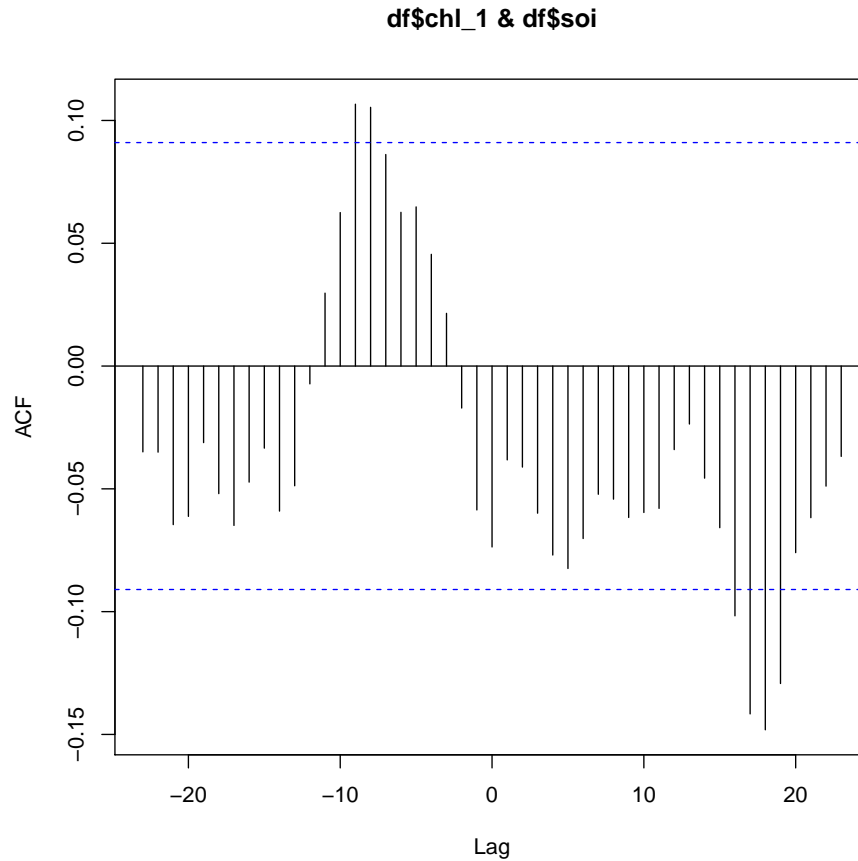
```
qplot(chl_1, soi, data=df, geom=c('point', 'smooth'), method='lm', formula=y~x)
```



```
## Correlation coefficient
cor(df$chl_1, df$soi, use='complete.obs')

## [1] -0.07367003

ccf(df$chl_1, df$soi, na.action=na.pass)
```

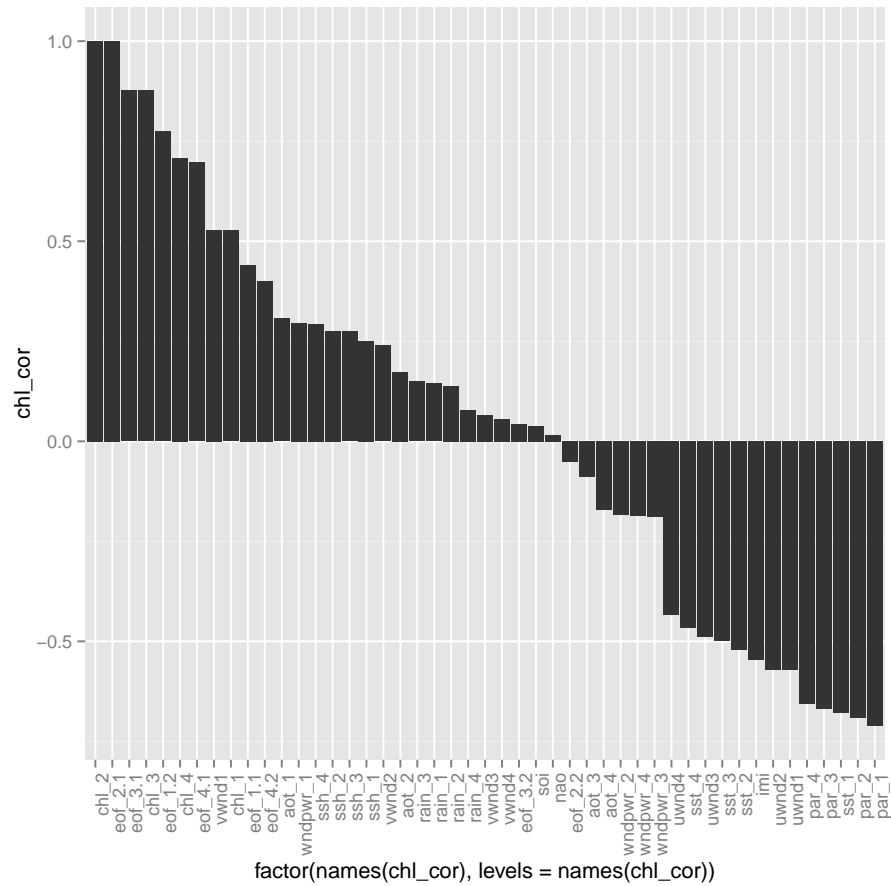


## 2 Region 2 (Southern Central Red Sea)

### 2.1 Correlation with other variables

The Correlation level between CHL in the southern central cluster and other environmental variable. The most important are in the order: CHL (clusters 3 and 4), PAR, wind, IMI, SSH, rain.

```
tot_cor <- cor(df[, -c(1, 2, 3)], use='complete')
chl_cor <- tot_cor['chl_2', ]
chl_cor <- sort(chl_cor, decreasing=TRUE)
names(chl_cor) <- factor(names(chl_cor), levels=names(chl_cor))
qplot(factor(names(chl_cor), levels=names(chl_cor)), chl_cor, geom='bar', stat='identity') +
  ## Warning: Stacking not well defined when ymin != 0
```



### 3 Region 3 (Northern Central Red Sea)

#### 3.1 Correlation with other variables

The Correlation level between CHL in the northern central cluster and other environmental variable. The most important are in the order: CHL (clusters 2 and 4), wind, PAR, IMI, SSH, rain.

```
tot_cor <- cor(df[, -c(1,2,3)], use='complete')
chl_cor <- tot_cor['chl_3',]
chl_cor <- sort(chl_cor, decreasing=TRUE)
names(chl_cor) <- factor(names(chl_cor), levels=names(chl_cor))
qplot(factor(names(chl_cor), levels=names(chl_cor)), chl_cor, geom='bar', stat='identity') +
  ## Warning: Stacking not well defined when ymin != 0
```





