Lab 4

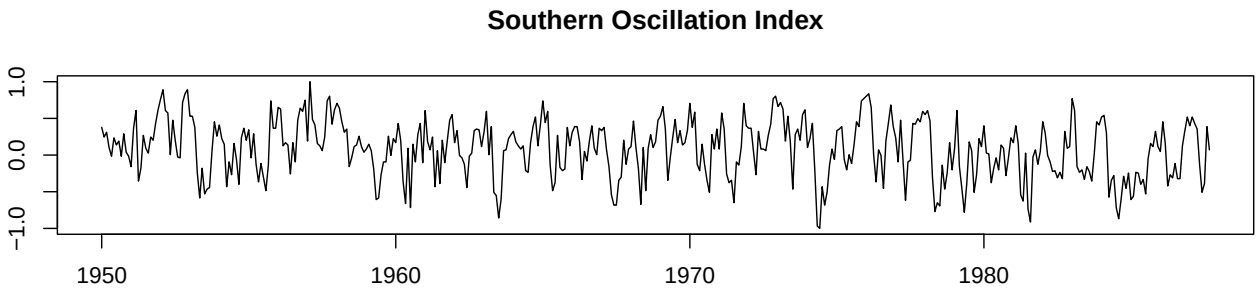
Arnab Hazra

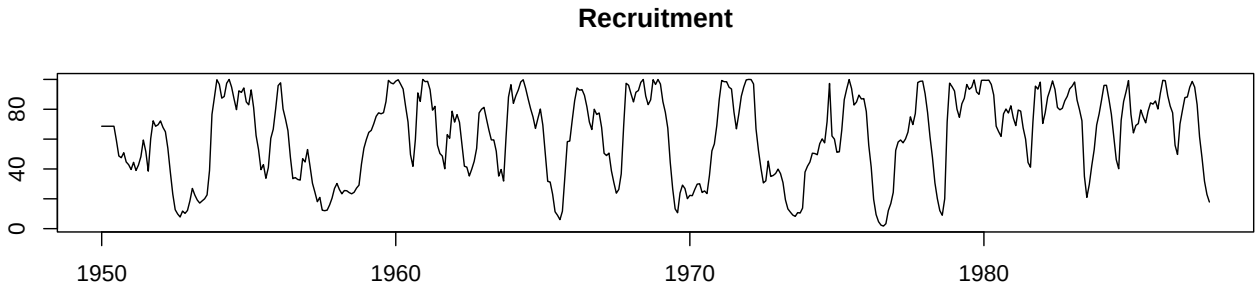
2024-09-03

We will spend five minutes for each question.

1. Plot the SOI and Recruitment data. The variables are given in the astsa package.

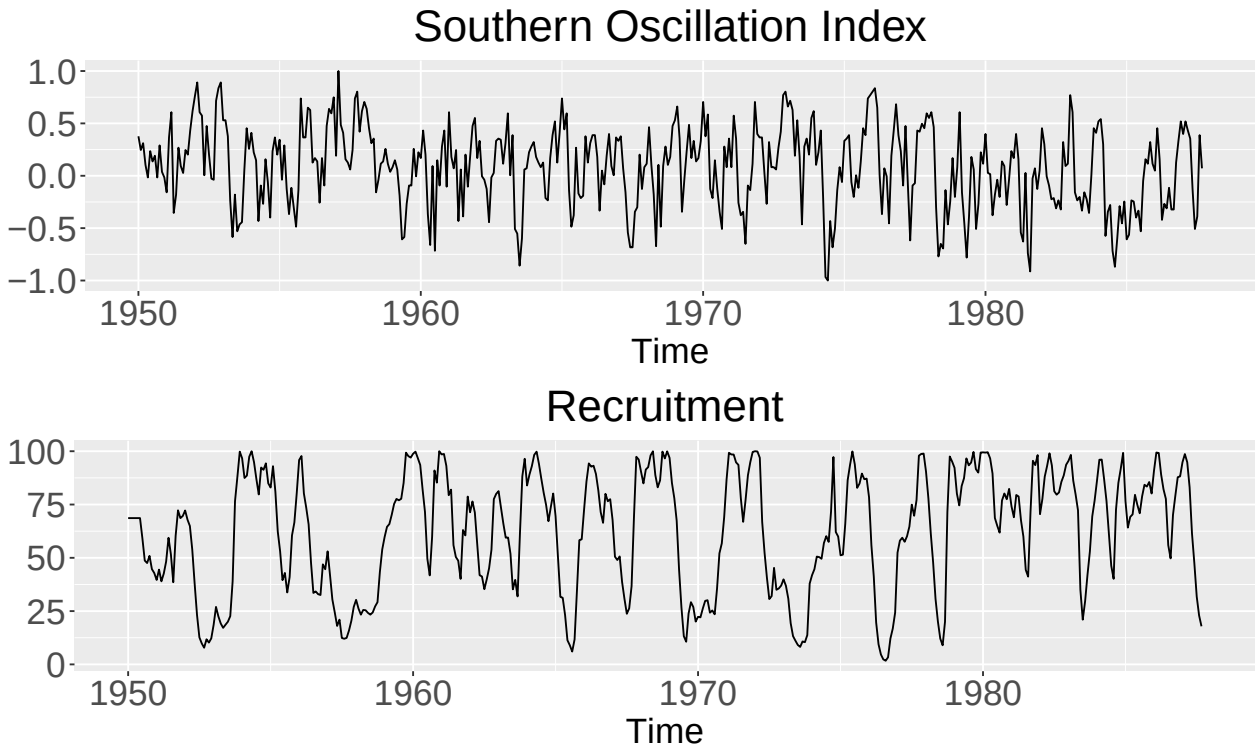
```
rm(list = ls())

library(astsa)

par(mfrow = c(2, 1))
plot(soi, ylab = "", xlab = "", main = "Southern Oscillation Index")
plot(rec, ylab = "", xlab = "", main = "Recruitment")

library(ggplot2)
```






```
library(gridExtra)

p1 <- ggplot() +
  geom_line(aes(x = as.vector(time(soi)), y = as.vector(soi))) +
```

```
  ggtitle("Southern Oscillation Index") + xlab("Time") + ylab(NULL) +
  theme(axis.text = element_text(size = 20),
        axis.title = element_text(size = 20),
        plot.title = element_text(hjust = 0.5, size = 25))

p2 <- ggplot() +
  geom_line(aes(x = as.vector(time(rec)), y = as.vector(rec))) +
  ggtitle("Recruitment") + xlab("Time") + ylab(NULL) +
  theme(axis.text = element_text(size = 20),
        axis.title = element_text(size = 20),
        plot.title = element_text(hjust = 0.5, size = 25))

grid.arrange(p1, p2, nrow = 2, ncol = 1)
```



2. Check if there is any long-term trend present in the data (for both series).

```
Z <- as.vector(time(soi))
soi.vec <- as.vector(soi)
rec.vec <- as.vector(rec)

# SOI trend

summary(lm(soi.vec ~ Z))
```

```
##
## Call:
## lm(formula = soi.vec ~ Z)
##
```

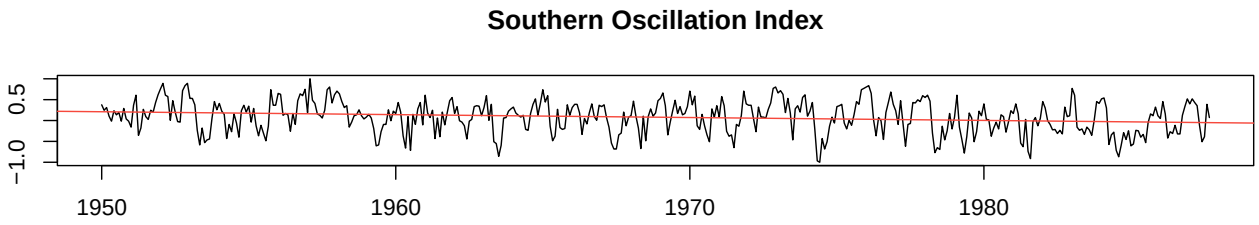
```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04140 -0.24183  0.01935  0.27727  0.83866
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.70367    3.18873   4.298 2.12e-05 ***
## Z           -0.00692    0.00162  -4.272 2.36e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3756 on 451 degrees of freedom
## Multiple R-squared:  0.0389, Adjusted R-squared:  0.03677
## F-statistic: 18.25 on 1 and 451 DF,  p-value: 2.359e-05
```

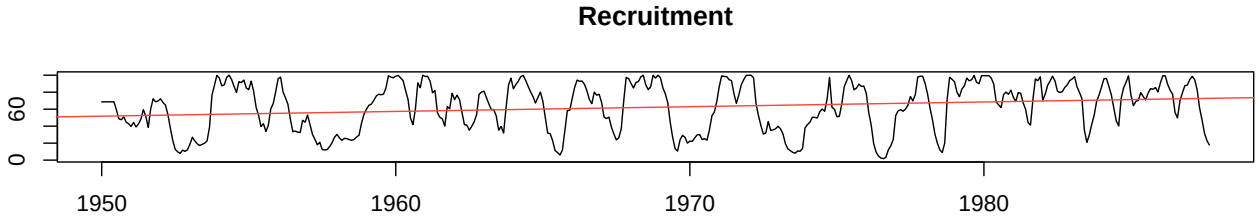
```
# Recruitment trend

summary(lm(rec.vec ~ Z))
```

```
##
## Call:
## lm(formula = rec.vec ~ Z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.864 -22.696   5.365  23.036  45.855
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1035.5774   232.0641  -4.462 1.02e-05 ***
## Z               0.5576     0.1179   4.731 3.00e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.34 on 451 degrees of freedom
## Multiple R-squared:  0.04728,    Adjusted R-squared:  0.04517
## F-statistic: 22.38 on 1 and 451 DF,  p-value: 2.996e-06
```

3. Overlay the fitted lines with both the series.

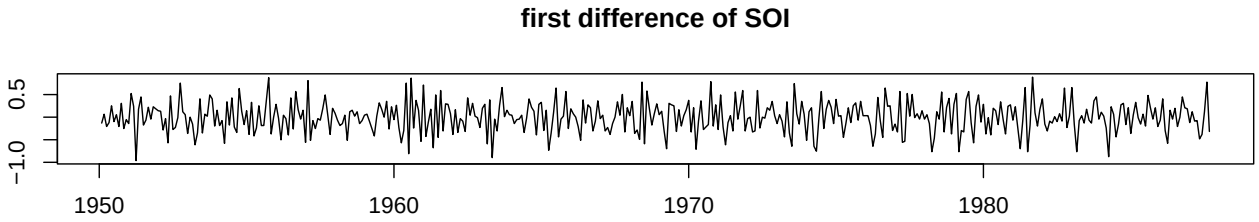
```
betahat.soi <- as.vector(lm(soi.vec ~ Z)$coefficients)
betahat.rec <- as.vector(lm(rec.vec ~ Z)$coefficients)

par(mfrow = c(2, 1))
plot(soi, ylab = "", xlab = "", main = "Southern Oscillation Index")
abline(a = betahat.soi[1], b = betahat.soi[2], col = 2)

plot(rec, ylab = "", xlab = "", main = "Recruitment")
abline(a = betahat.rec[1], b = betahat.rec[2], col = 2)
```

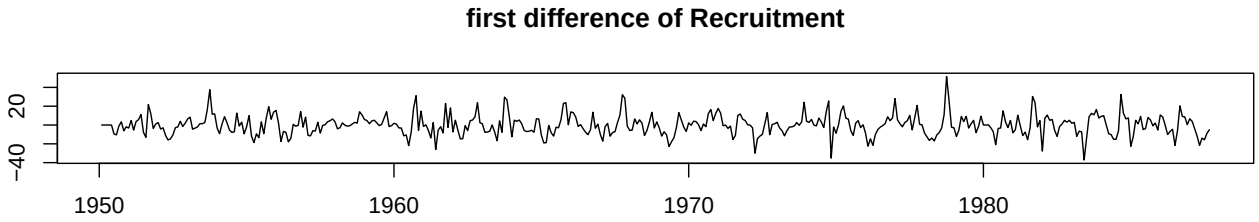




4. Plot the first differences of the two time series.

```
par(mfrow = c(2, 1))
plot(diff(soi), ylab = "", xlab = "", main = "first difference of SOI")
plot(diff(rec), ylab = "", xlab = "", main = "first difference of Recruitment")
```

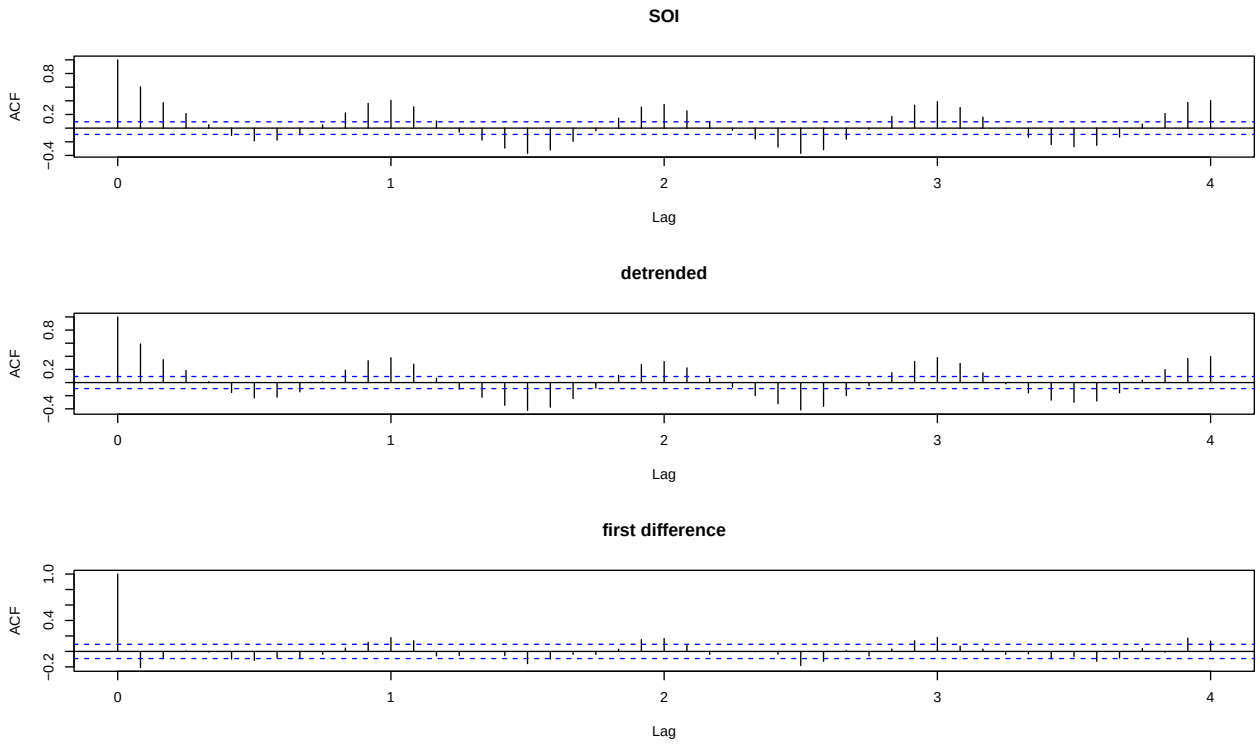




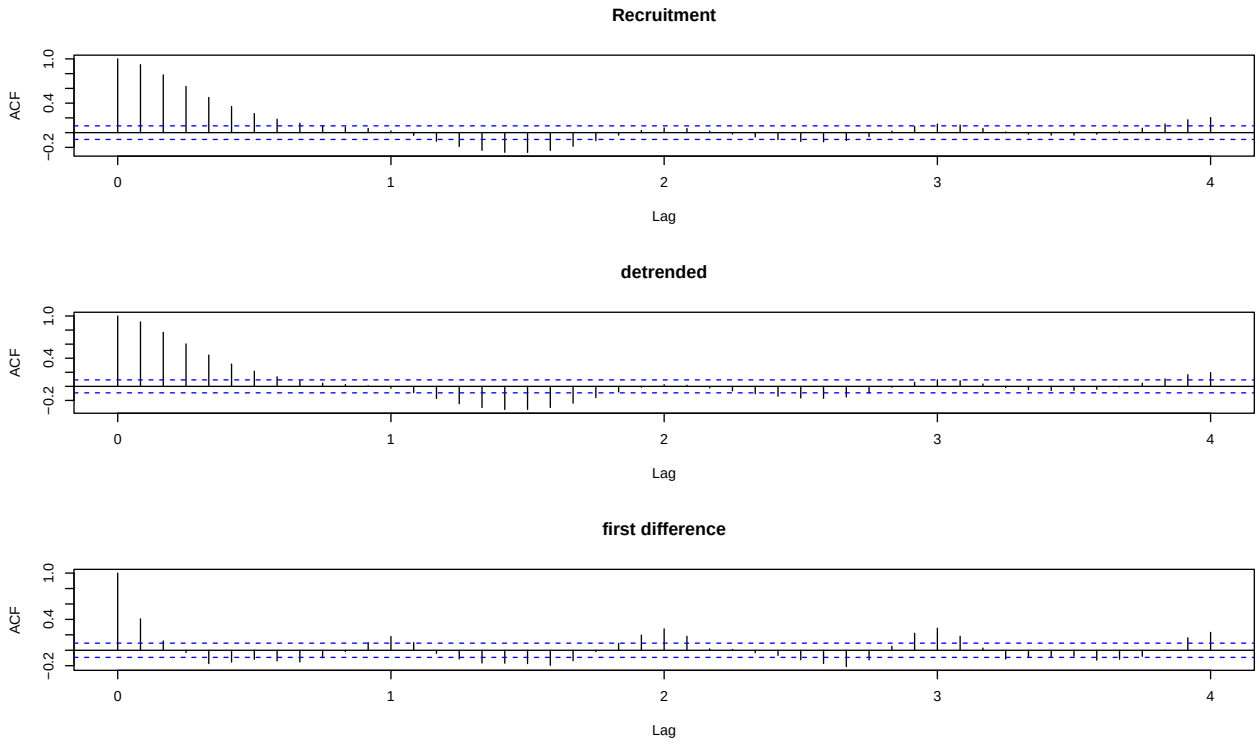
5. Plot the ACF for original SOI series, detrended series, and first differences. Repeat the same for Recruitment series.

```
lmm.soi <- lm(soi.vec ~ Z)
lmm.rec <- lm(rec.vec ~ Z)

resids.soi <- resid(lmm.soi)
resids.rec <- resid(lmm.rec)

resids.soi.ts <- ts(resids.soi, frequency = 12, start = c(1950, 0))
```

```
resids.rec.ts <- ts(resids.rec, frequency = 12, start = c(1950, 0))

par(mfrow = c(3, 1))
acf(soi, 48, main = "SOI")
acf(resids.soi.ts, 48, main = "detrended")
acf(diff(soi), 48, main = "first difference")
```




```
par(mfrow = c(3, 1))
acf(rec, 48, main = "Recruitment")
acf(resids.rec.ts, 48, main = "detrended")
acf(diff(rec), 48, main = "first difference")
```



6. Plot the CCF between SOI and Recruitment. Use the `ccf` function.

```
ccf(soi, rec, 48, main="SOI vs Recruitment", ylab="CCF")
```

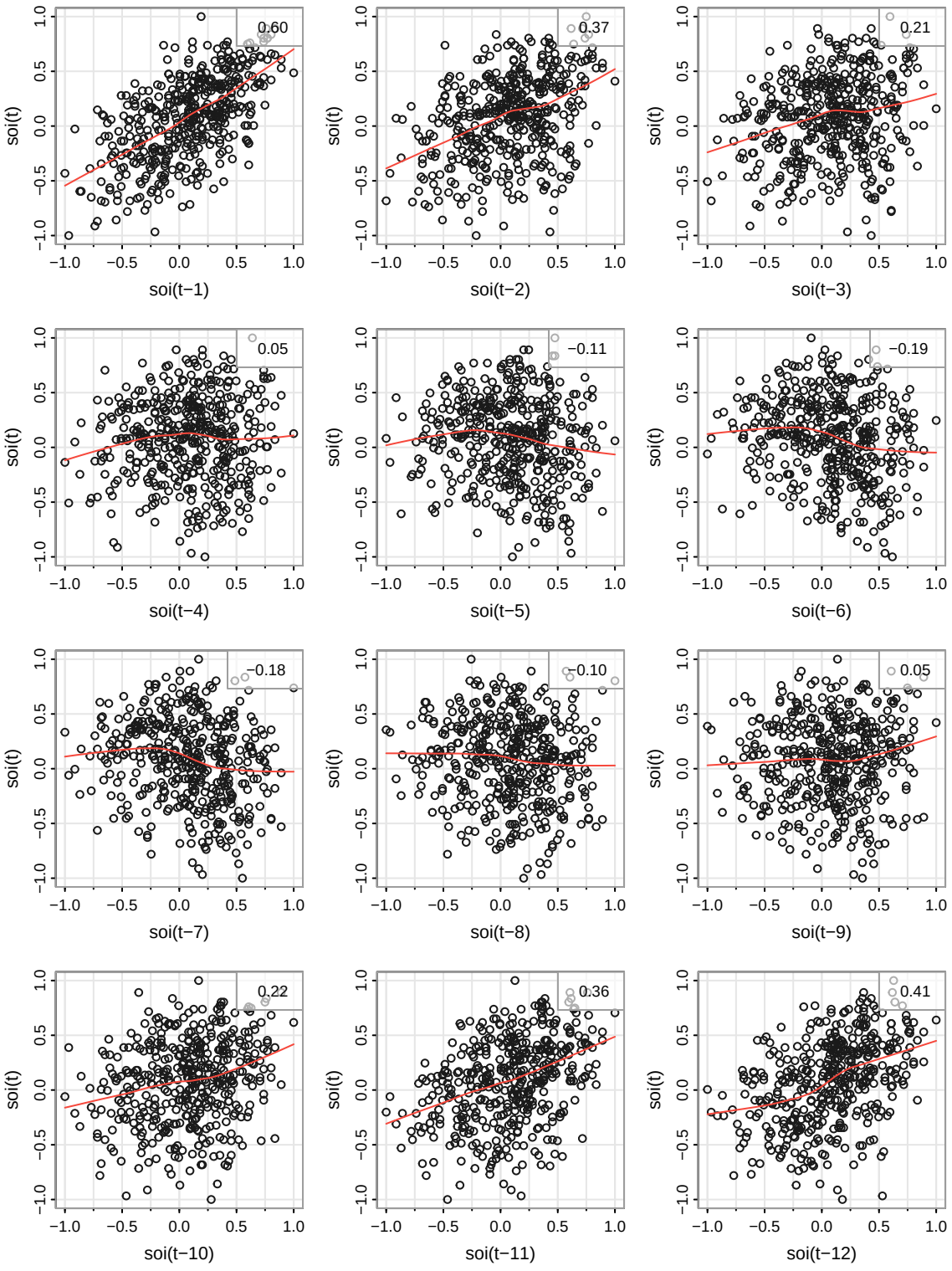
7. Perform regression of Recruitment R_t based on SOI at lag 6, i.e., S_{t-6}. Use the function `ts.intersect` to create the data frame.

```
fish <- ts.intersect(rec, soiL6 = lag(soi, -6), dframe = TRUE)
summary(fit1 <- lm(rec ~ soiL6, data = fish, na.action = NULL))
```

```
##
## Call:
## lm(formula = rec ~ soiL6, data = fish, na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.187 -18.234   0.354  16.580  55.790
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   65.790      1.088   60.47   <2e-16 ***
## soiL6        -44.283      2.781  -15.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.5 on 445 degrees of freedom
## Multiple R-squared:  0.3629, Adjusted R-squared:  0.3615
## F-statistic: 253.5 on 1 and 445 DF,  p-value: < 2.2e-16
```

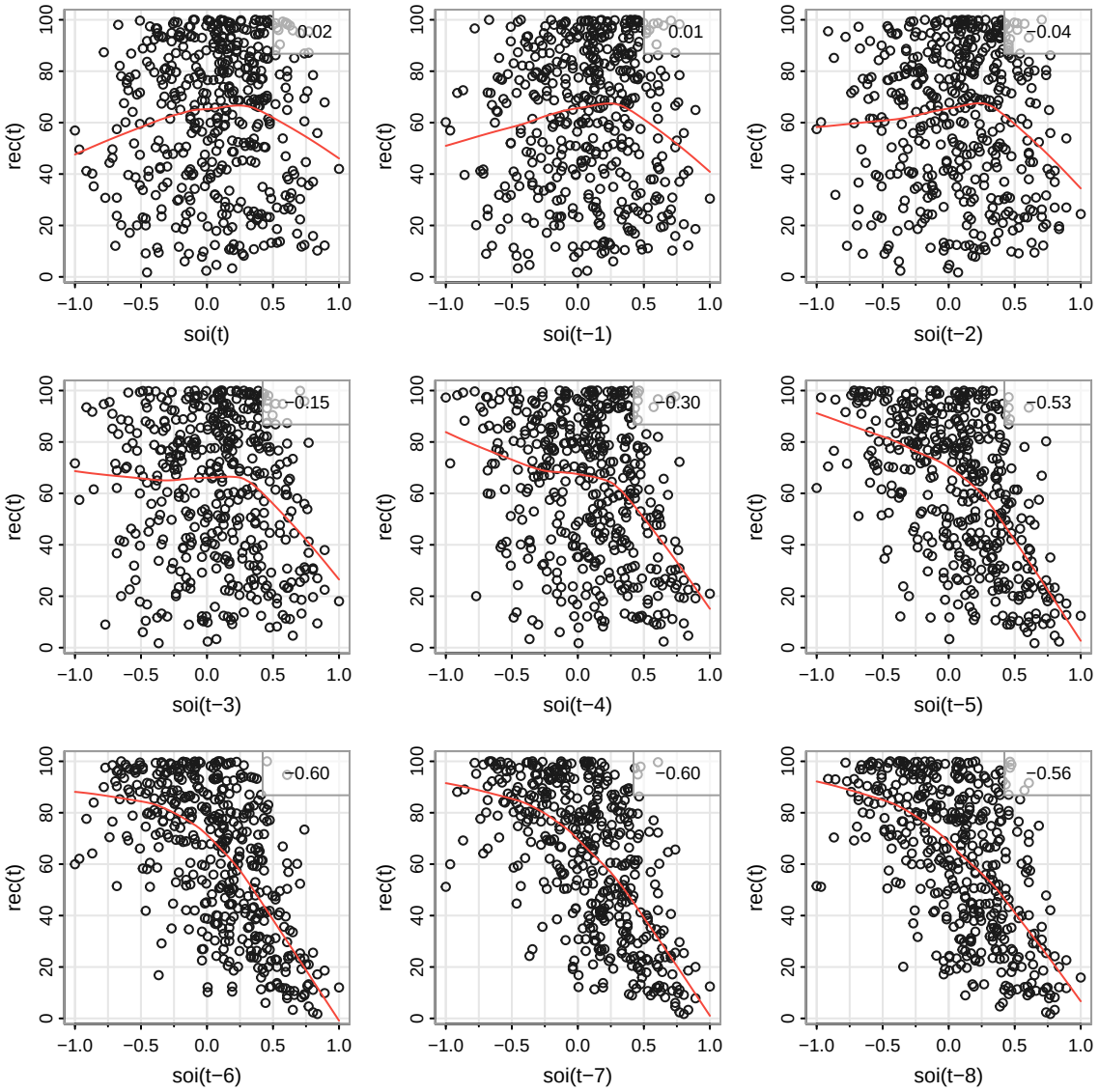
8. Plot the scatter-plot matrix of the lagged SOI. Use the function `lag1.plot`.

```
lag1.plot(soi, 12, cex.axis = 2)
```



9. Plot the scatter-plot matrix of SOI and lagged Recruitment. Use the function `lag2.plot`.

```
lag2.plot(soi, rec, 8)
```



10. Go back to Question 7. Perform the regression based on the dummy variable $D_t = I(S_t < 0)$. Thus, create $D_{t-6} = I(S_{t-6} < 0)$.

```
dummy <- ifelse(soi < 0, 0, 1)
fish <- ts.intersect(rec, soiL6 = lag(soi, -6), dL6 = lag(dummy, -6), dframe = TRUE)
summary(fit <- lm(rec ~ soiL6*dL6, data=fish, na.action = NULL))
```

```
##
## Call:
## lm(formula = rec ~ soiL6 * dL6, data = fish, na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.291 -15.821   2.224  15.791  61.788
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   74.479      2.865  25.998  < 2e-16 ***
## soiL6        -15.358      7.401  -2.075   0.0386 *
## dL6           -1.139      3.711  -0.307   0.7590
## soiL6:dL6    -51.244      9.523  -5.381  1.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.84 on 443 degrees of freedom
## Multiple R-squared:  0.4024, Adjusted R-squared:  0.3984
## F-statistic: 99.43 on 3 and 443 DF,  p-value: < 2.2e-16
```

```
plot(fish$soiL6, fish$rec)
points(fish$soiL6, fitted(fit), pch = '+', col = 2)
```