Revised_scatterplots

Jackson Hoeke

5/18/2021

The goal of this revision is to reassess the regressions made between

Year, Diversity, Richness, and Latitude. There were several issues with the

regressions, including:

1) MPAs were sampled in different locations in different years with irregular

intervals.

2) Statistics were not recorded for the regressions.

3) The plots were usually too cluttered to read easily

Regressions will be performed on groups in the same region, and which the

intervals between sampling are (roughly) equal

The data is from summed transects, which have been combined into a

grand sum sheet to calculate diversity.

Groups will be:

Channel Islands: Carrington Point, Gull Island, Harris Point, South Point

Channel Islands with Anacapa: Same as above, but including Anacapa Island

Northern CA: Bodega Bay, Farallon Islands

Central CA: Pt. Buchon, Pt. Lobos, Pt. Sur

Other MPAs will not be used in the regression if they do not fit well, or

have less than 2 years of data collection, since they are more likely than

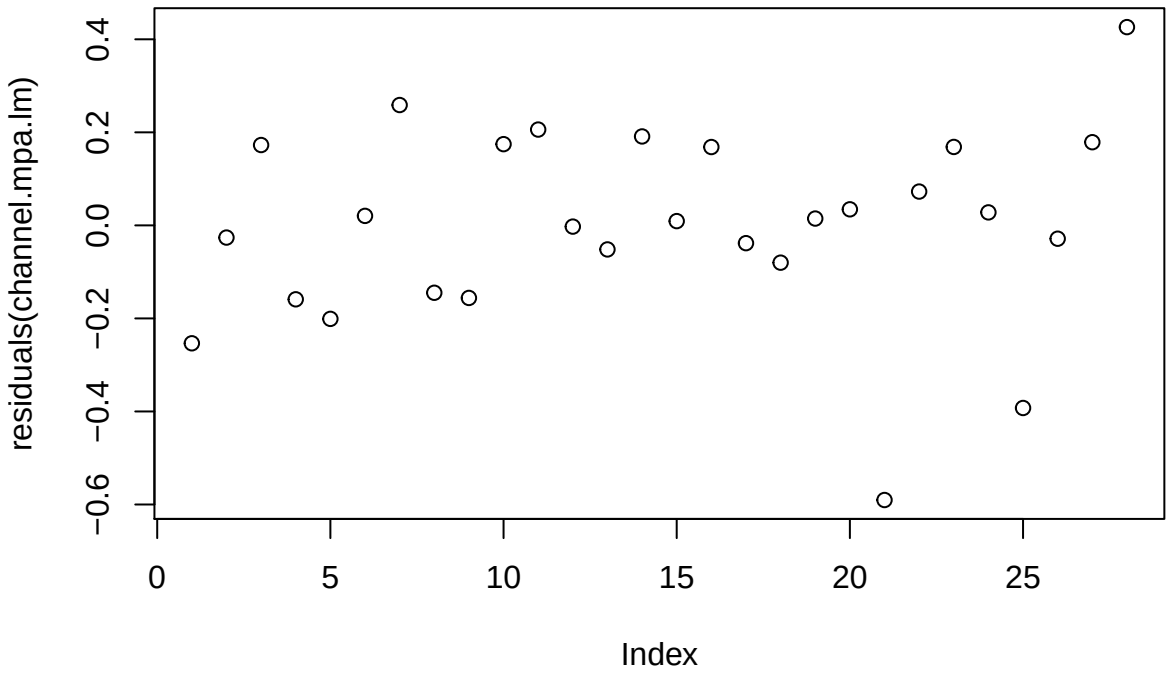
not to skew the data.

To start, average diversity over time for MPA and reference sites will be run

```r
# Import data

data <- data.frame(read.csv("Avg.transects.csv"))

# Add diversity metric

data <- mutate(data, diversity = diversity(data[10:166]))

# Add diversity average

grp.mean <- data %>%
  group_by_all() %>%
  mutate(newnames=paste0(Year, MPA_Group, Type))

split.mean <- split(grp.mean, grp.mean$newnames)
for (I in 1:length(split.mean)) {assign(unique(split.mean[[I]]$newnames),
                                        split.mean[[I]])}

mean.div <- lapply(split.mean, function(x){

  x <- mutate(x, avg.div=mean(x$diversity))
  x <- mutate(x, div.sd=sd(x$diversity))
  x <- mutate(x, div.SE=div.sd/length(x$diversity))

})

div.avgs <- bind_rows(mean.div)

# slim down data

p1 <- div.avgs[2:9]
p2 <- div.avgs[167:171]
mod.avg <- cbind(p1,p2)
mod.avg <- mod.avg[!duplicated(mod.avg[12]),]
mod.avg <- as.data.frame(mod.avg)

# Run linear regression, starting with the Channel Islands

Channel <- subset(mod.avg, MPA_Group == "Carrington Point" |
                    MPA_Group == "Gull Island" | MPA_Group == "Harris Point" |
                    MPA_Group == "South Point")

# Subset by designation

Channel.mpa <- subset(Channel, Designation == "MPA")
Channel.ref <- subset(Channel, Designation == "Reference")

# Run MPA first

set.seed(25)
channel.mpa.lm <- lm(formula = avg.div ~ Year, data = Channel.mpa)
summary(channel.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = Channel.mpa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59034 -0.09633  0.01194  0.16964  0.42619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.02024   22.39911   1.072    0.293
## Year        -0.01121    0.01115  -1.006    0.324
##
## Residual standard error: 0.2125 on 26 degrees of freedom
## Multiple R-squared:  0.03744,    Adjusted R-squared:  0.0004198
## F-statistic: 1.011 on 1 and 26 DF,  p-value: 0.3239
```

```r
# (Year: t = -1.006, p = 0.324)
# (df = 1,26, F = 1.011, p = 0.3239)
# (y~24.02-0.01121(Year))
# (R = 0.03744)

# Check assumptions

plot(residuals(channel.mpa.lm))
```



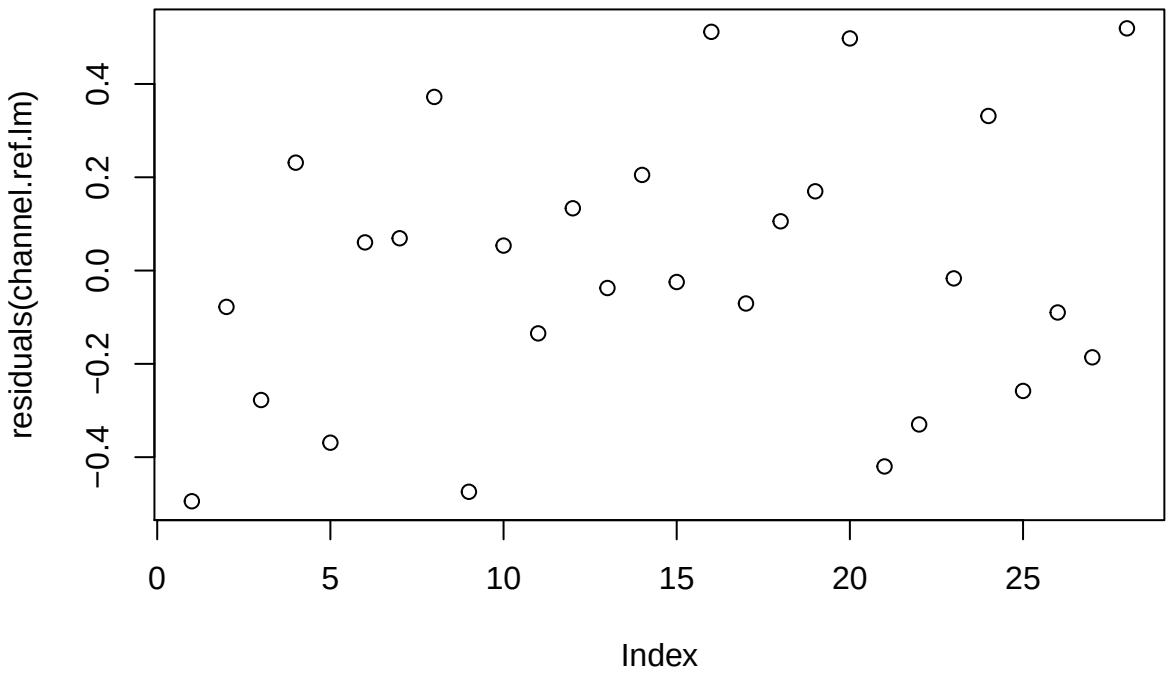
```r
shapiro.test(residuals(channel.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.mpa.lm)
## W = 0.95352, p-value = 0.2426
```

```r
# Residuals normal

# Run Ref

set.seed(25)
channel.ref.lm <- lm(formula = avg.div ~ Year, data = Channel.ref)
summary(channel.ref.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = Channel.ref)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49447 -0.20402 -0.02067  0.17868  0.51922
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.049391  31.349093  -0.225    0.824
## Year         0.004247   0.015603   0.272    0.788
##
## Residual standard error: 0.2975 on 26 degrees of freedom
## Multiple R-squared:  0.002841,   Adjusted R-squared:  -0.03551
## F-statistic: 0.07409 on 1 and 26 DF,  p-value: 0.7876
```

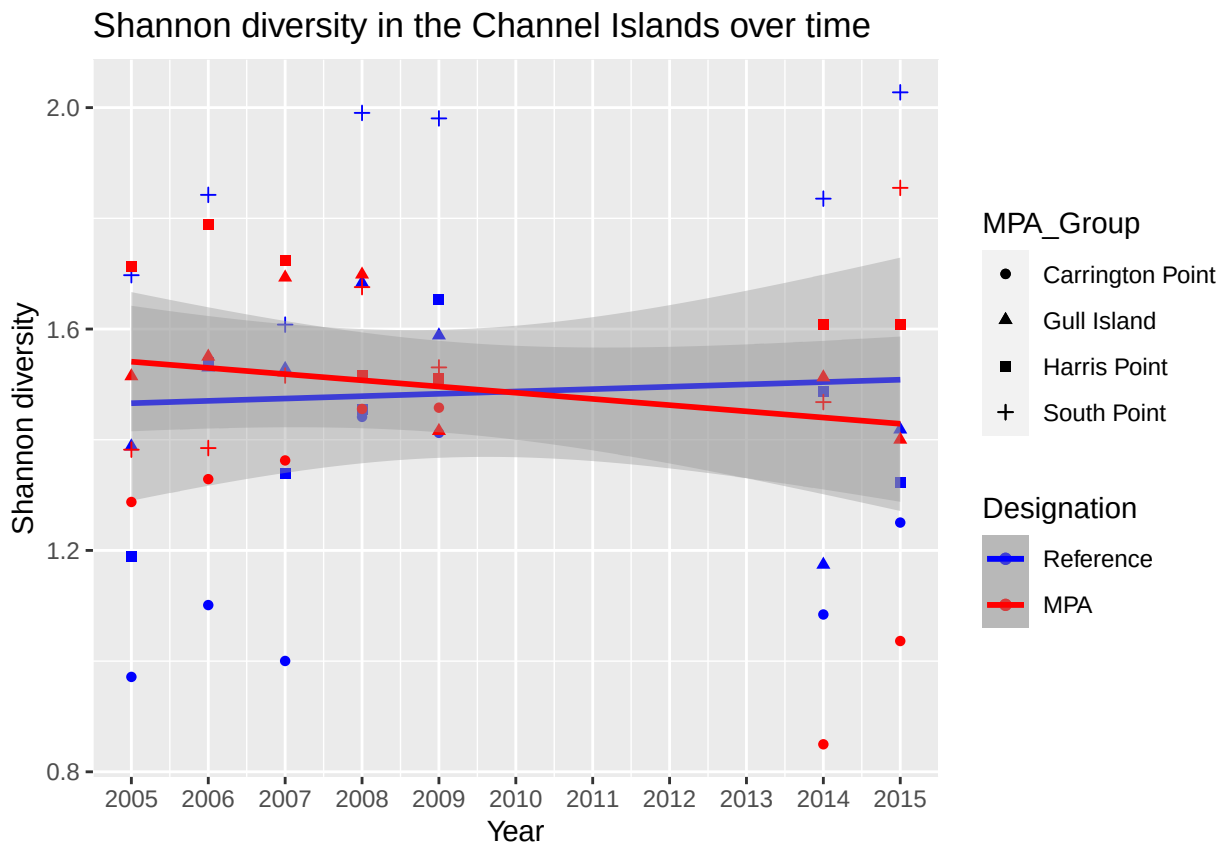
```r
# (Year: t = 0.272, p = 0.788)
# (df = 1,26, F = 0.07409, p = 0.7876)
# (R = 0.002841)

# Check assumptions

plot(residuals(channel.ref.lm))
```




```
shapiro.test(residuals(channel.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.ref.lm)
## W = 0.9691, p-value = 0.5566
```

```
# Residuals normal

# Plot MPA and Ref regressions for channel islands

ggplot() +
  geom_point(data=Channel.ref,aes(x=Year,y=avg.div,colour=Designation,
                                  shape=MPA_Group)) +
  stat_smooth(data=Channel.ref,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  geom_point(data=Channel.mpa,aes(x=Year,y=avg.div,colour=Designation,
                                  shape=MPA_Group)) +
  stat_smooth(data=Channel.mpa,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#0000ff","#ff0000"),
                     breaks = c("Reference","MPA")) +
  ggtitle("Shannon diversity in the Channel Islands over time") +
  xlab("Year") +
  ylab("Shannon diversity") +
  scale_x_continuous(breaks=seq(2005,2015,1))
```



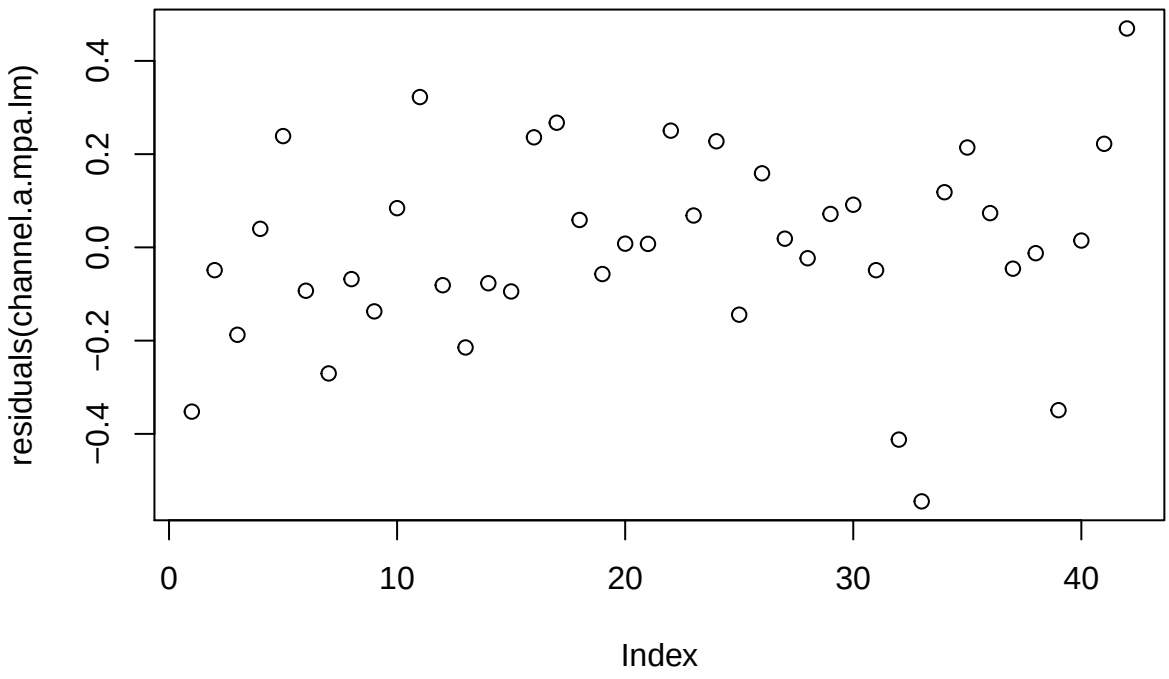
Run the channel islands with Anacapa Island

```
# Create Channel Islands (+ Ancacapa) data

Channel.a <- subset(mod.avg, MPA_Group == "Anacapa Island" |
                      MPA_Group == "Carrington Point" |
                      MPA_Group == "Gull Island" | MPA_Group == "Harris Point" |
                      MPA_Group == "South Point")
Channel.a <- subset(Channel.a, Year == "2005" | Year == "2006" |
                      Year == "2007" | Year == "2008" | Year == "2009" |
                      Year == "2014" | Year == "2015")

# Subset by designation

Channel.a.mpa <- subset(Channel.a, Designation == "MPA")

# Run model

set.seed(25)
channel.a.mpa.lm <- lm(formula = avg.div ~ Year, data = Channel.a.mpa)
summary(channel.a.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = Channel.a.mpa)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54474 -0.09000  0.00773  0.11160  0.46953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.411662  17.979294    1.08    0.287
## Year        -0.008946   0.008949   -1.00    0.323
##
## Residual standard error: 0.2089 on 40 degrees of freedom
## Multiple R-squared:  0.02438,    Adjusted R-squared:  -1.496e-05
## F-statistic: 0.9994 on 1 and 40 DF,  p-value: 0.3235
```

```
# (Year: t = -1.00, p = 0.323)
# (df = 1,40, F = 0.9994, p = 0.3235)
# (y~19.41-0.008946(Year))
# (R = 0.02438)

# Check assumptions

plot(residuals(channel.a.mpa.lm))
```

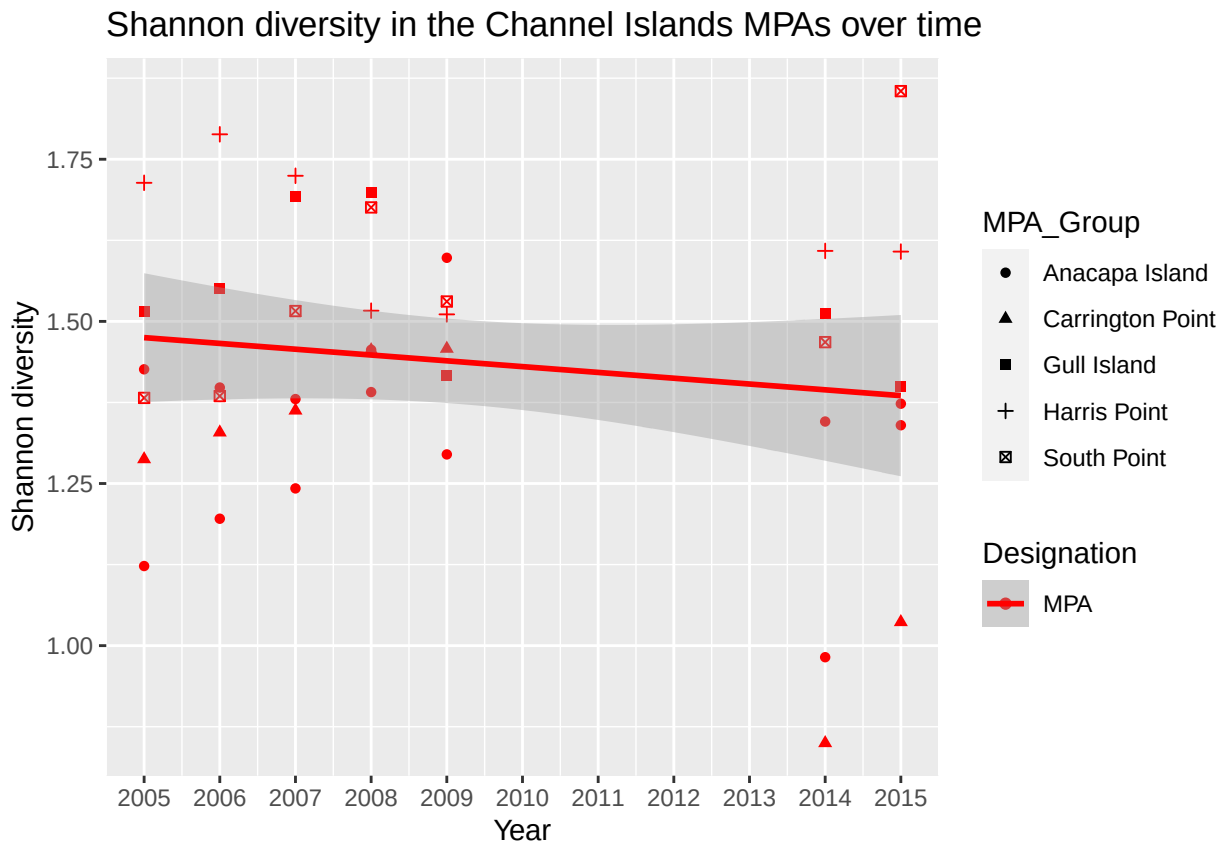



```
shapiro.test(residuals(channel.a.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  residuals(channel.a.mpa.lm)
## W = 0.97893, p-value = 0.6203
```

```
# Residuals normal

# Plot

ggplot() +
  geom_point(data=Channel.a.mpa,aes(x=Year,y=avg.div,colour=Designation,
                                    shape=MPA_Group)) +
  stat_smooth(data=Channel.a.mpa,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#ff0000"),
                     breaks = c("MPA")) +
  ggtitle("Shannon diversity in the Channel Islands MPAs over time") +
  xlab("Year") +
  ylab("Shannon diversity") +
  scale_x_continuous(breaks=seq(2005,2015,1))
```



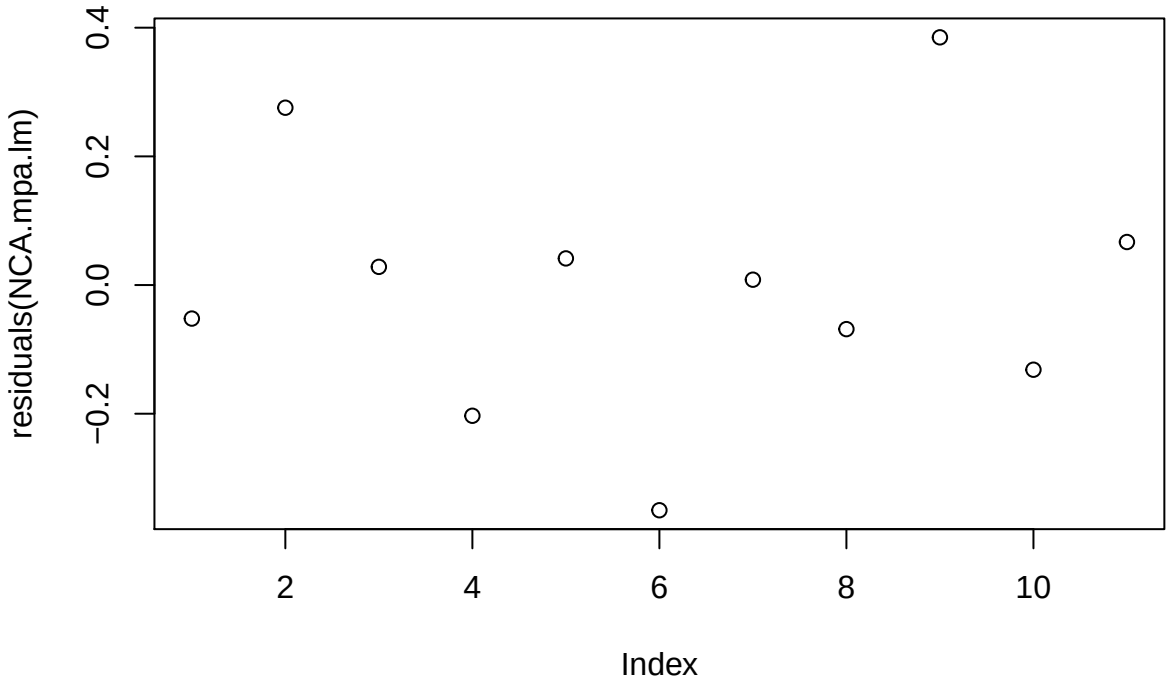
Run North CA

```
# Create northern CA data

NCA <- subset(mod.avg, MPA_Group == "Bodega Bay" |
```

```
                MPA_Group == "Farallon Islands")

# Subset by designation

NCA.mpa <- subset(NCA, Designation == "MPA")
NCA.ref <- subset(NCA, Designation == "Reference")

# Run MPA first

set.seed(25)
NCA.mpa.lm <- lm(formula = avg.div ~ Year, data = NCA.mpa)
summary(NCA.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = NCA.mpa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35003 -0.10006  0.00831  0.05414  0.38500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.06522   41.39742  -0.485    0.639
## Year          0.01072    0.02054   0.522    0.614
##
## Residual standard error: 0.216 on 9 degrees of freedom
## Multiple R-squared:  0.02936,    Adjusted R-squared:  -0.07849
## F-statistic: 0.2722 on 1 and 9 DF,  p-value: 0.6144
```

```
# (Year: t = 0.522, p = 0.614)
# (df = 1,9, F = 0.2722, p = 0.6144)
# (y~-20.06522+0.01072(Year))
# (R = 0.02936)

# Test assumptions

plot(residuals(NCA.mpa.lm))
```




```
shapiro.test(residuals(NCA.mpa.lm))
```

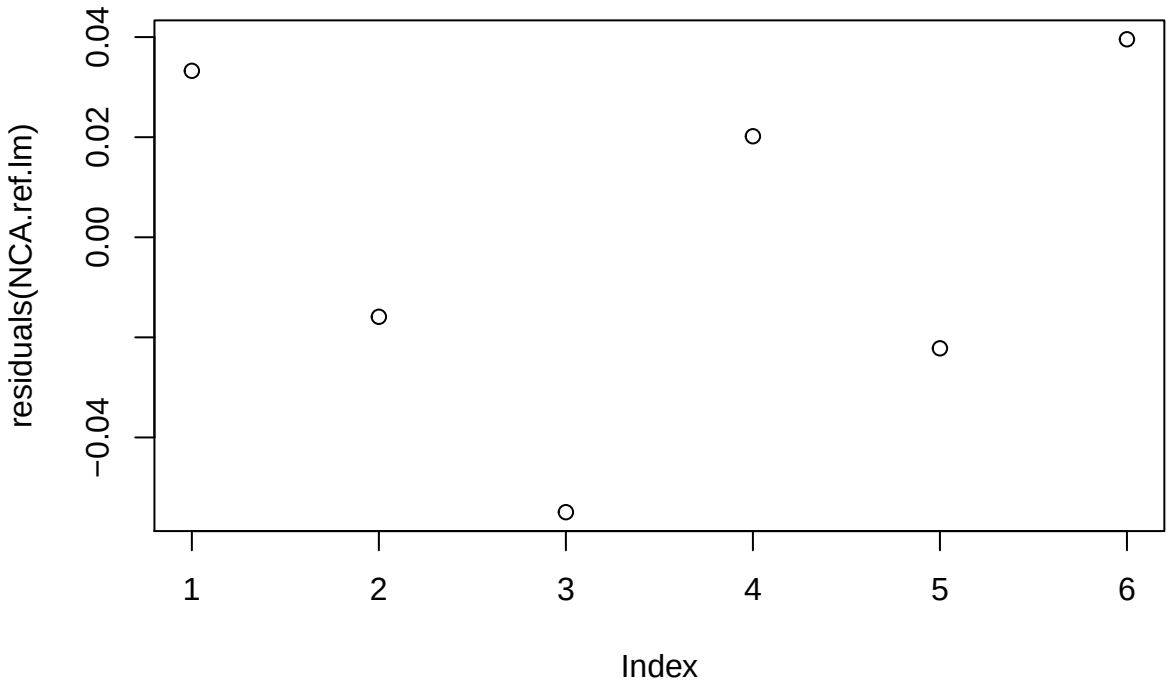
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.mpa.lm)
## W = 0.96533, p-value = 0.8359
```

```
# Residuals normal

# Run Ref

set.seed(25)
NCA.ref.lm <- lm(formula = avg.div ~ Year, data = NCA.ref)
summary(NCA.ref.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = NCA.ref)
##
## Residuals:
##       58       61      101      106      142      147
##  0.03326 -0.01589 -0.05493  0.02019 -0.02219  0.03956
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.12499   10.41683  -0.204    0.848
```

```
## Year         0.00184    0.00517   0.356    0.740
##
## Residual standard error: 0.04136 on 4 degrees of freedom
## Multiple R-squared:  0.03069,    Adjusted R-squared:  -0.2116
## F-statistic: 0.1266 on 1 and 4 DF,  p-value: 0.7399
```

```
# (Year: t = 0.356, p = 0.740)
# (df = 1,4, F = 0.1266, p = 0.7399)
# (y~-2.12499+0.00184(Year))
# (R = 0.03069)

# Test assumptions

plot(residuals(NCA.ref.lm))
```

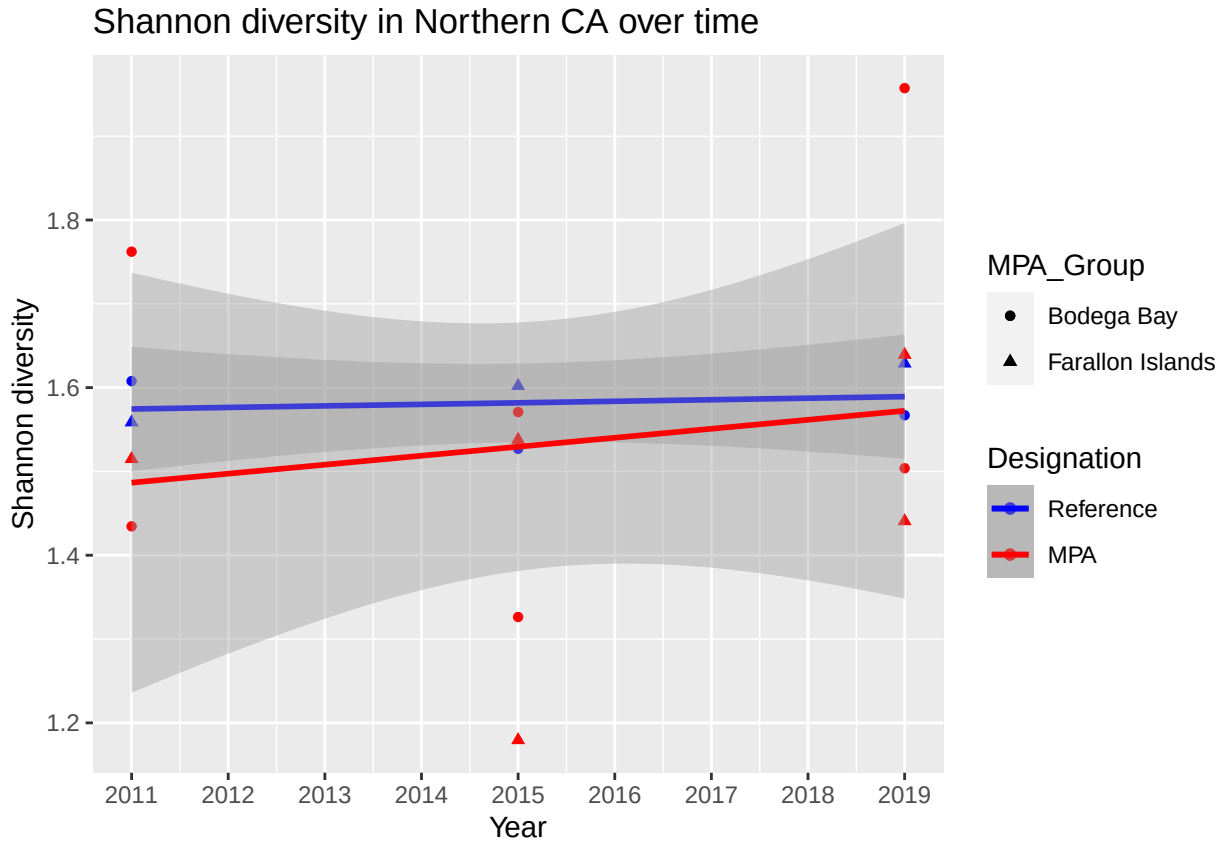



```
shapiro.test(residuals(NCA.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.ref.lm)
## W = 0.92398, p-value = 0.5345
```

```
# Residuals normal

# Plot MPA and Ref regressions for channel islands

ggplot() +
```

```
  geom_point(data=NCA.ref,aes(x=Year,y=avg.div,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=NCA.ref,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  geom_point(data=NCA.mpa,aes(x=Year,y=avg.div,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=NCA.mpa,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#0000ff","#ff0000"),
                     breaks = c("Reference","MPA")) +
  ggtitle("Shannon diversity in Northern CA over time") +
  xlab("Year") +
  ylab("Shannon diversity") +
  scale_x_continuous(breaks=seq(2011,2019,1))
```



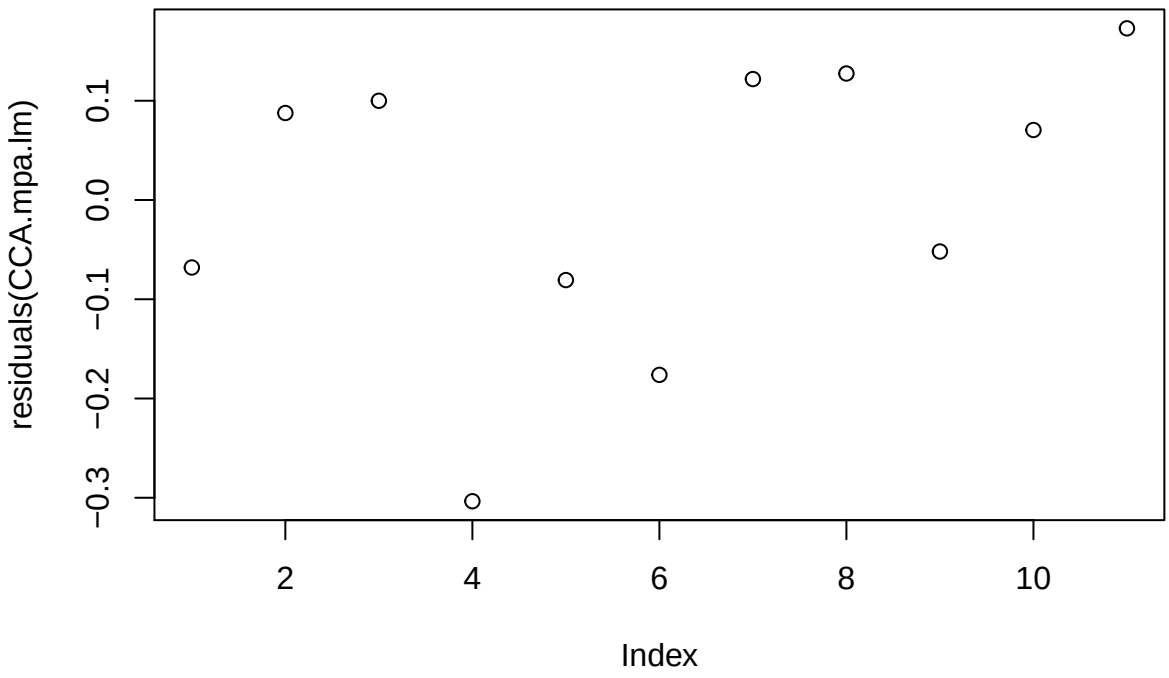
Run Central CA

```
# Create central CA data

CCA <- subset(mod.avg, MPA_Group == "Point Buchon" |
                MPA_Group == "Point Lobos" | MPA_Group == "Point Sur")
CCA <- subset(CCA, Year == "2008" | Year == "2016" | Year == "2019")

# Subset by designation

CCA.mpa <- subset(CCA, Designation == "MPA")
CCA.ref <- subset(CCA, Designation == "Reference")
```

```r
# Run MPA first

set.seed(25)
CCA.mpa.lm <- lm(formula = avg.div ~ Year, data = CCA.mpa)
summary(CCA.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = CCA.mpa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30350 -0.07434  0.07053  0.11090  0.17292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -29.35270   21.51890  -1.364    0.206
## Year          0.01544    0.01068   1.446    0.182
##
## Residual standard error: 0.1566 on 9 degrees of freedom
## Multiple R-squared:  0.1884, Adjusted R-squared:  0.09825
## F-statistic:  2.09 on 1 and 9 DF,  p-value: 0.1822
```

```r
# (Year: t = 1.446, p = 0.182)
# (df = 1,9, F = 2.09, p = 0.1822)
# (y~-29.35270+0.01544(Year))
# (R = 0.1884)

# Check assumptions

plot(residuals(CCA.mpa.lm))
```




```
shapiro.test(residuals(CCA.mpa.lm))
```

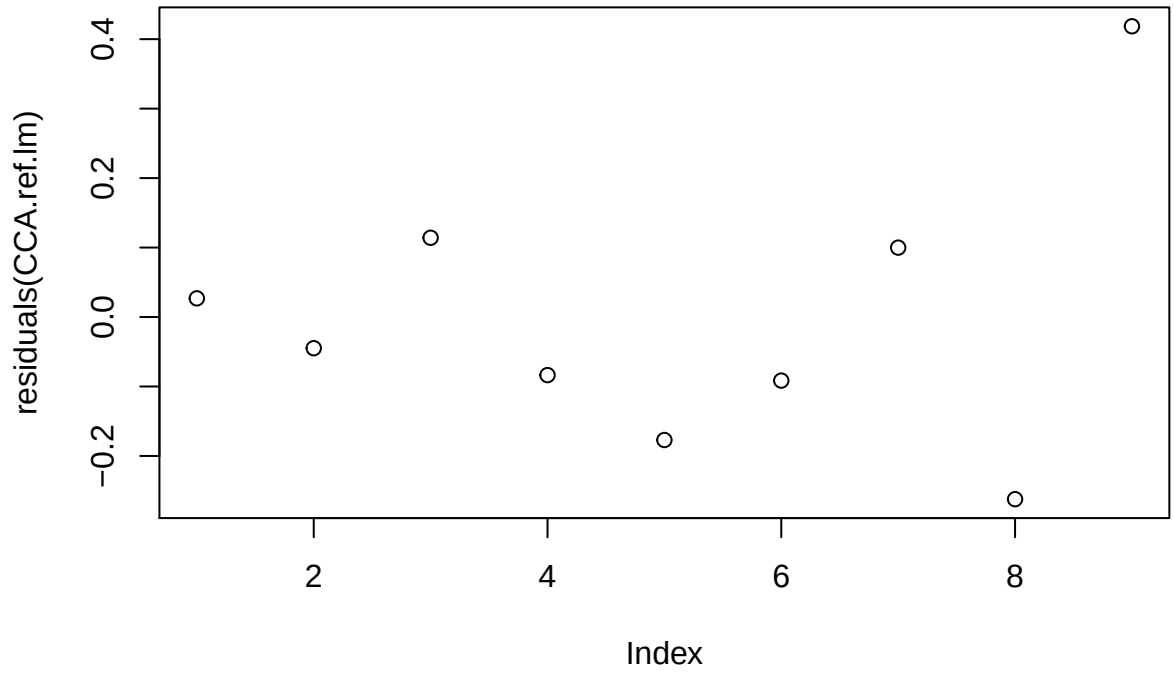
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.mpa.lm)
## W = 0.90582, p-value = 0.2174
```

```
# Residuals normal

# Run Ref

set.seed(25)
CCA.ref.lm <- lm(formula = avg.div ~ Year, data = CCA.ref)
summary(CCA.ref.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Year, data = CCA.ref)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26223 -0.09154 -0.04487  0.09987  0.41853
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -36.60770   30.74587  -1.191    0.273
```

```
## Year          0.01906    0.01526   1.249    0.252
##
## Residual standard error: 0.2126 on 7 degrees of freedom
## Multiple R-squared:  0.1822, Adjusted R-squared:  0.06541
## F-statistic:  1.56 on 1 and 7 DF,  p-value: 0.2518
```

```
# (Year: t = 1.249, p = 0.252)
# (df = 1,7, F = 1.56, p = 0.2518)
# (y~-36.60770+0.01906(Year))
# (R = 0.1822)

# Check assumptions

plot(residuals(CCA.ref.lm))
```

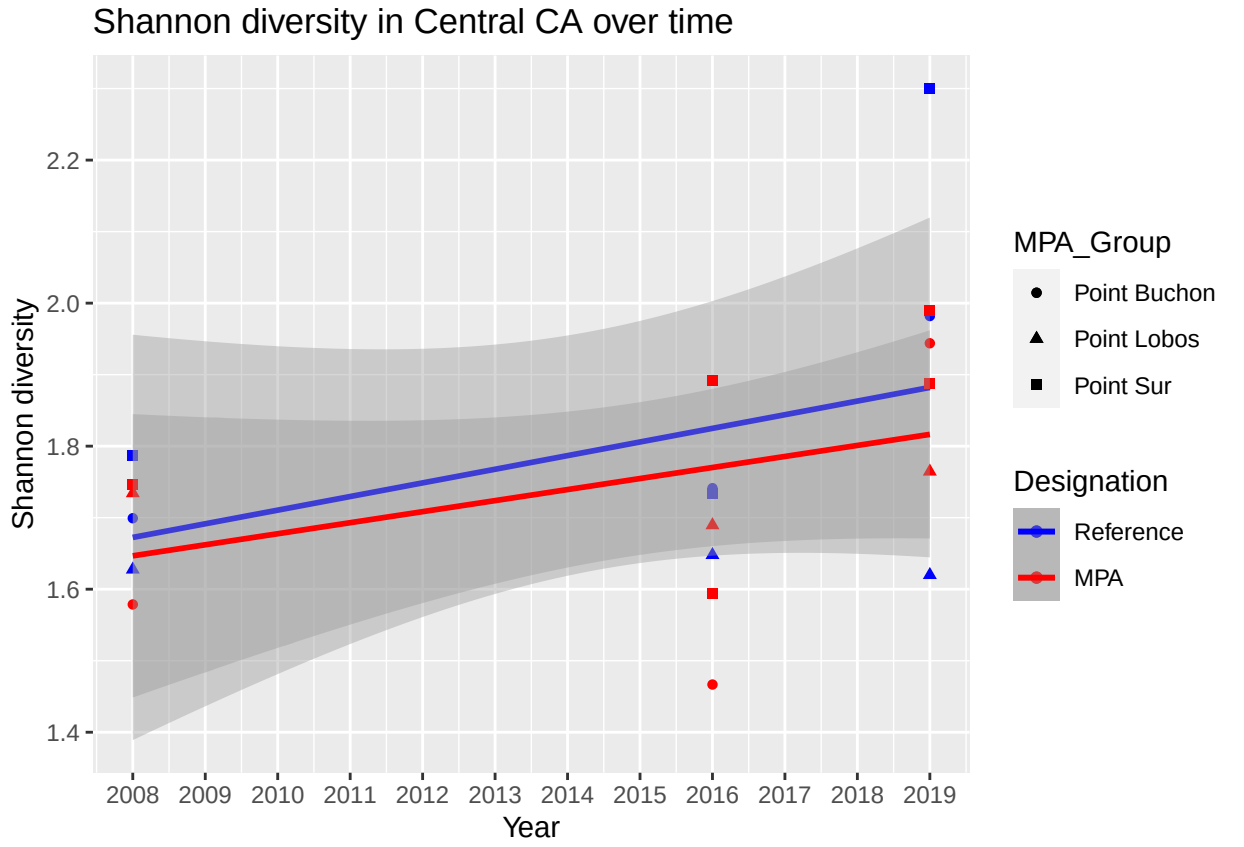



```
shapiro.test(residuals(CCA.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.ref.lm)
## W = 0.93398, p-value = 0.5202
```

```
# Residuals normal

# Plot MPA and Ref regressions for channel islands

ggplot() +
```

```
  geom_point(data=CCA.ref,aes(x=Year,y=avg.div,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=CCA.ref,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  geom_point(data=CCA.mpa,aes(x=Year,y=avg.div,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=CCA.mpa,aes(x=Year,y=avg.div,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#0000ff","#ff0000"),
                     breaks = c("Reference","MPA")) +
  ggtitle("Shannon diversity in Central CA over time") +
  xlab("Year") +
  ylab("Shannon diversity") +
  scale_x_continuous(breaks=seq(2007,2019,1))
```



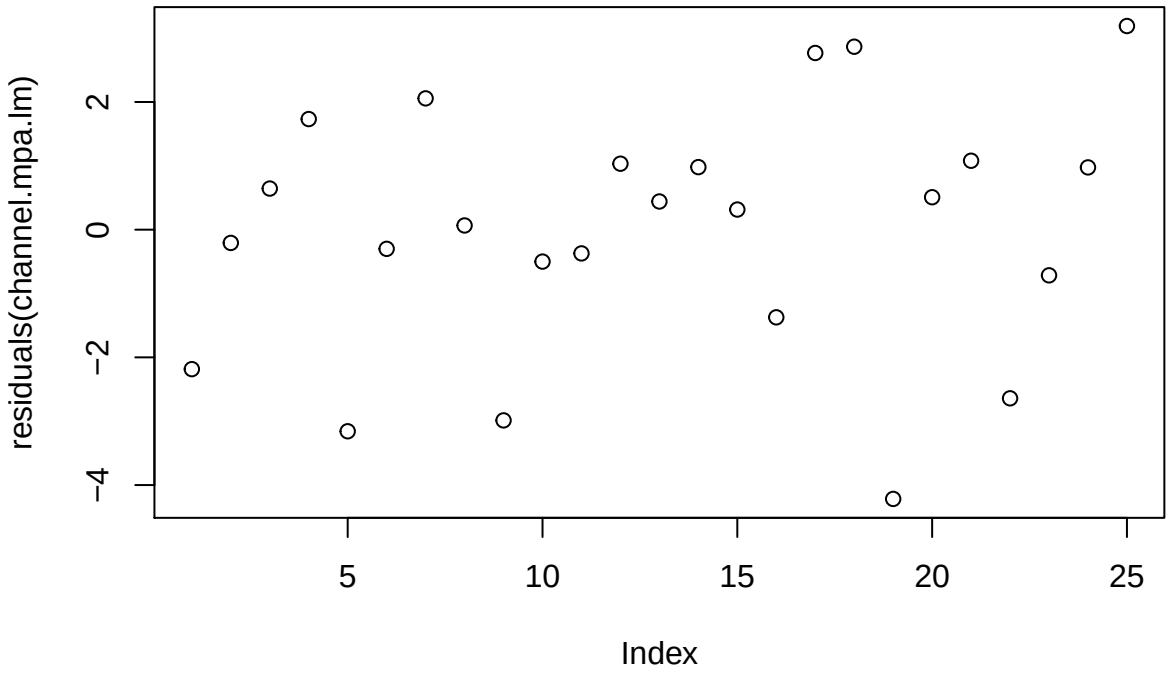
Run again, but using species richness instead of diversity

```
# The following 4 data blocks repeat the above code, but using richness

# Add richness metric

data <- mutate(data, richness = rowSums(data[10:166]!=0))

# Add richness average

grp.mean <- data %>%
  group_by_all() %>%
```

```
  mutate(newnames=paste0(Year, MPA_Group, Type))

split.mean <- split(grp.mean, grp.mean$newnames)
for (I in 1:length(split.mean)) {assign(unique(split.mean[[I]]$newnames),
                                        split.mean[[I]])}

mean.rich <- lapply(split.mean, function(x){

  x <- mutate(x, avg.rich=mean(x$richness))
  x <- mutate(x, rich.sd=sd(x$richness))
  x <- mutate(x, rich.SE=rich.sd/length(x$richness))

})

rich.avgs <- bind_rows(mean.rich)
p1 <- rich.avgs[2:8]
p2 <- rich.avgs[167:172]
mod.avg <- cbind(p1,p2)
mod.avg <- mod.avg[!duplicated(mod.avg[11]),]
mod.avg <- as.data.frame(mod.avg)

# Run linear regression, starting with the Channel Islands

Channel <- subset(mod.avg, MPA_Group == "Carrington Point" |
                    MPA_Group == "Gull Island" | MPA_Group == "Harris Point" |
                    MPA_Group == "South Point")

# Subset by designation

Channel.mpa <- subset(Channel, Designation == "MPA")
Channel.ref <- subset(Channel, Designation == "Reference")

# Run MPA first

set.seed(25)
channel.mpa.lm <- lm(formula = avg.rich ~ Year, data = Channel.mpa)
summary(channel.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.rich ~ Year, data = Channel.mpa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2171 -0.7151  0.3153  1.0333  3.1896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -52.09036  216.18383  -0.241    0.812
## Year          0.03285    0.10761   0.305    0.763
##
## Residual standard error: 1.975 on 23 degrees of freedom
## Multiple R-squared:  0.004035,   Adjusted R-squared:  -0.03927
## F-statistic: 0.09318 on 1 and 23 DF,  p-value: 0.7629
```

```
# (Year: t = 0.305, p = 0.763)
# (df = 1,23, F = 0.09318, p = 0.7629)
# (y~-52.09036+0.03285(Year))
# (R = 0.004035)

# Check assumptions

plot(residuals(channel.mpa.lm))
```




```
shapiro.test(residuals(channel.mpa.lm))
```

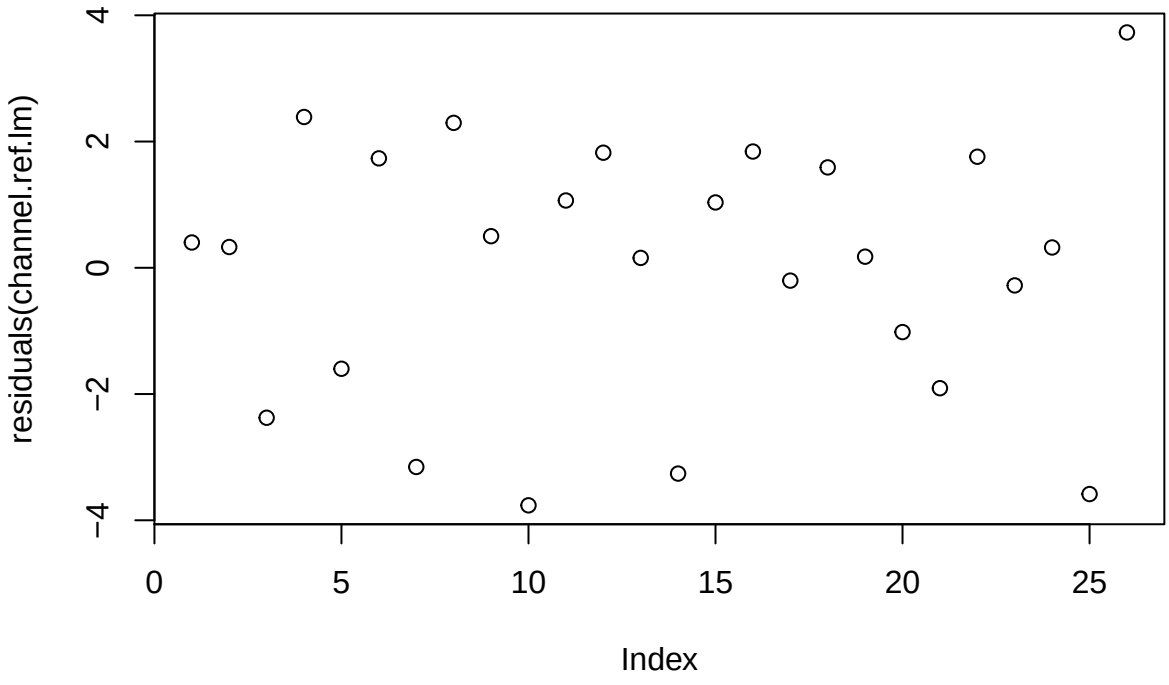
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.mpa.lm)
## W = 0.96335, p-value = 0.4851
```

```
# Residuals normal

# Run Ref

set.seed(25)
channel.ref.lm <- lm(formula = avg.rich ~ Year, data = Channel.ref)
summary(channel.ref.lm)
```

```
##
## Call:
```

```
## lm(formula = avg.rich ~ Year, data = Channel.ref)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7632 -1.4545  0.3253  1.6978  3.7283
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -232.0613   220.5525  -1.052    0.303
## Year           0.1218     0.1098   1.110    0.278
##
## Residual standard error: 2.079 on 24 degrees of freedom
## Multiple R-squared:  0.0488, Adjusted R-squared:  0.009162
## F-statistic: 1.231 on 1 and 24 DF,  p-value: 0.2782
```

```
# (Year: t = 1.110, p = 0.278)
# (df = 1,24, F = 1.009, p = 0.2782)
# (y~-232.0613+0.1218(Year))
# (R = 0.0488)

plot(residuals(channel.ref.lm))
```

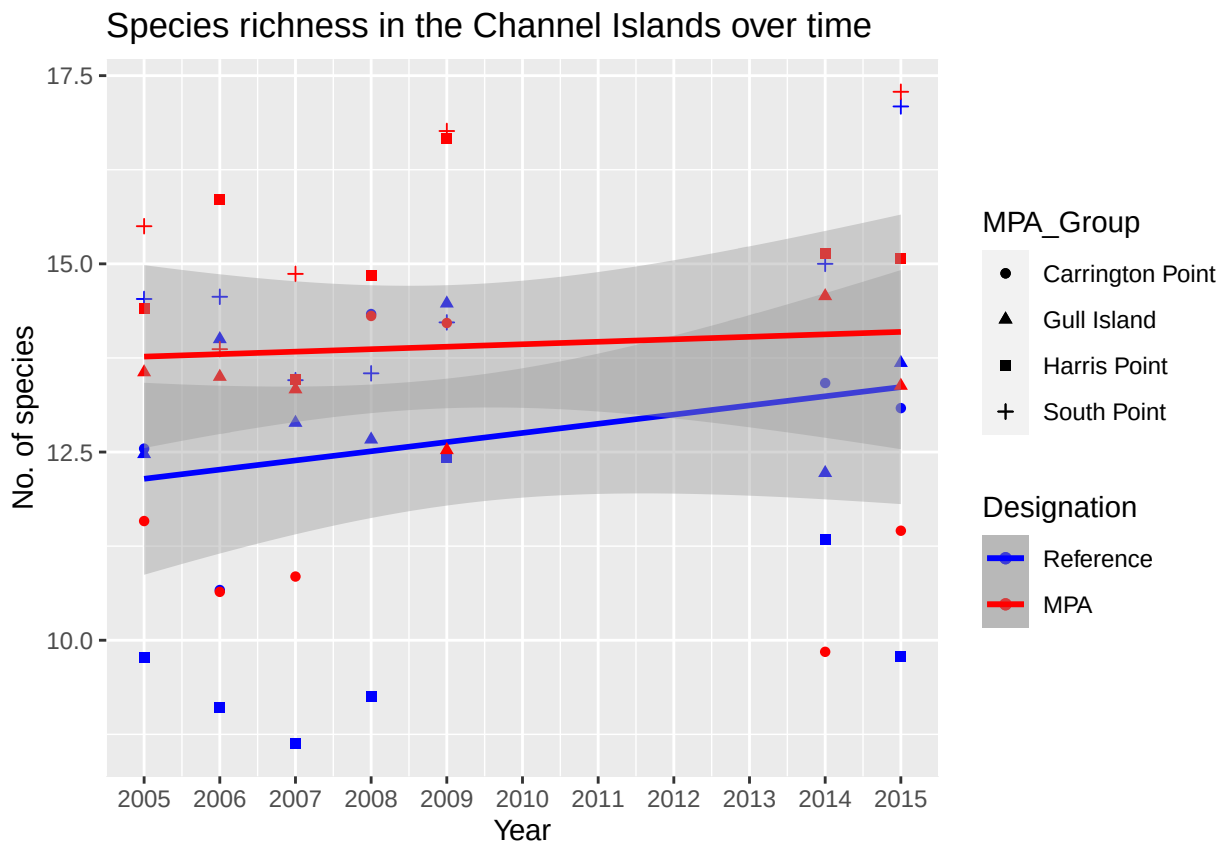



```
shapiro.test(residuals(channel.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  residuals(channel.ref.lm)
## W = 0.94498, p-value = 0.1766
```

```
# Residuals normal

# Plot MPA and Ref regressions for channel islands

ggplot() +
  geom_point(data=Channel.ref,aes(x=Year,y=avg.rich,colour=Designation,
                                  shape=MPA_Group)) +
  stat_smooth(data=Channel.ref,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  geom_point(data=Channel.mpa,aes(x=Year,y=avg.rich,colour=Designation,
                                  shape=MPA_Group)) +
  stat_smooth(data=Channel.mpa,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#0000ff","#ff0000"),
                     breaks = c("Reference","MPA")) +
  ggtitle("Species richness in the Channel Islands over time") +
  xlab("Year") +
  ylab("No. of species") +
  scale_x_continuous(breaks=seq(2005,2015,1))
```



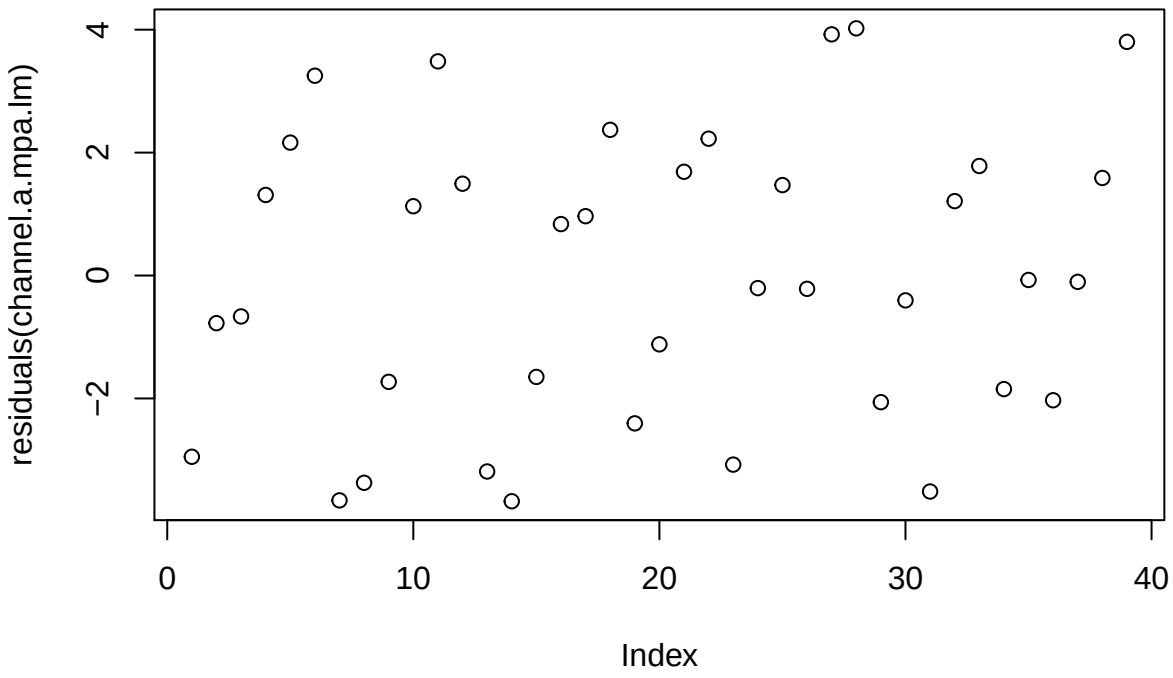
Channel Islands, with Anacapa

```
# Create channel islands (+ Anacapa) data
Channel.a <- subset(mod.avg, MPA_Group == "Anacapa Island" |
                      MPA_Group == "Carrington Point" |
```

```
                      MPA_Group == "Gull Island" | MPA_Group == "Harris Point" |
                      MPA_Group == "South Point")
Channel.a <- subset(Channel.a, Year == "2005" | Year == "2006" |
                      Year == "2007" | Year == "2008" | Year == "2009" |
                      Year == "2014" | Year == "2015")

# Subset by designation

Channel.a.mpa <- subset(Channel.a, Designation == "MPA")

# Run model

set.seed(25)
channel.a.mpa.lm <- lm(formula = avg.rich ~ Year, data = Channel.a.mpa)
summary(channel.a.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.rich ~ Year, data = Channel.a.mpa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6726 -1.9386 -0.1031  1.6377  4.0216
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -235.3671   209.0338  -1.126    0.267
## Year           0.1235     0.1040   1.187    0.243
##
## Residual standard error: 2.37 on 37 degrees of freedom
## Multiple R-squared:  0.03668,    Adjusted R-squared:  0.01065
## F-statistic: 1.409 on 1 and 37 DF,  p-value: 0.2428
```

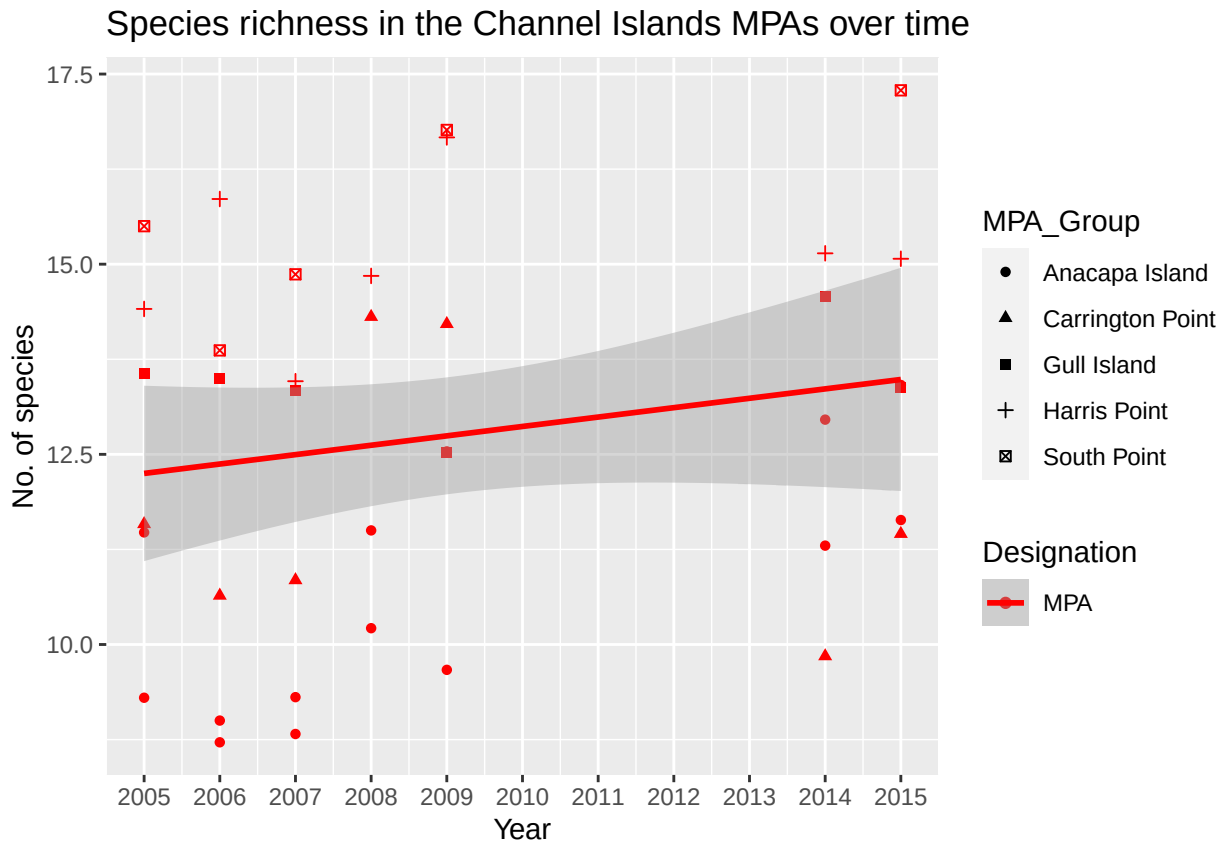
```
# (Year: t = 1.187, p = 0.243)
# (df = 1,37, F = 1.409, p = 0.2428)
# (y~-235.3671+0.1235(Year))
# (R = 0.03668)

plot(residuals(channel.a.mpa.lm))
```




```
shapiro.test(residuals(channel.a.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.a.mpa.lm)
## W = 0.95148, p-value = 0.0922
```

```
# Residuals normal
# Plot

ggplot() +
  geom_point(data=Channel.a.mpa,aes(x=Year,y=avg.rich,colour=Designation,
                                    shape=MPA_Group)) +
  stat_smooth(data=Channel.a.mpa,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#ff0000"),
                     breaks = c("MPA")) +
  ggtitle("Species richness in the Channel Islands MPAs over time") +
  xlab("Year") +
  ylab("No. of species") +
  scale_x_continuous(breaks=seq(2005,2015,1))
```



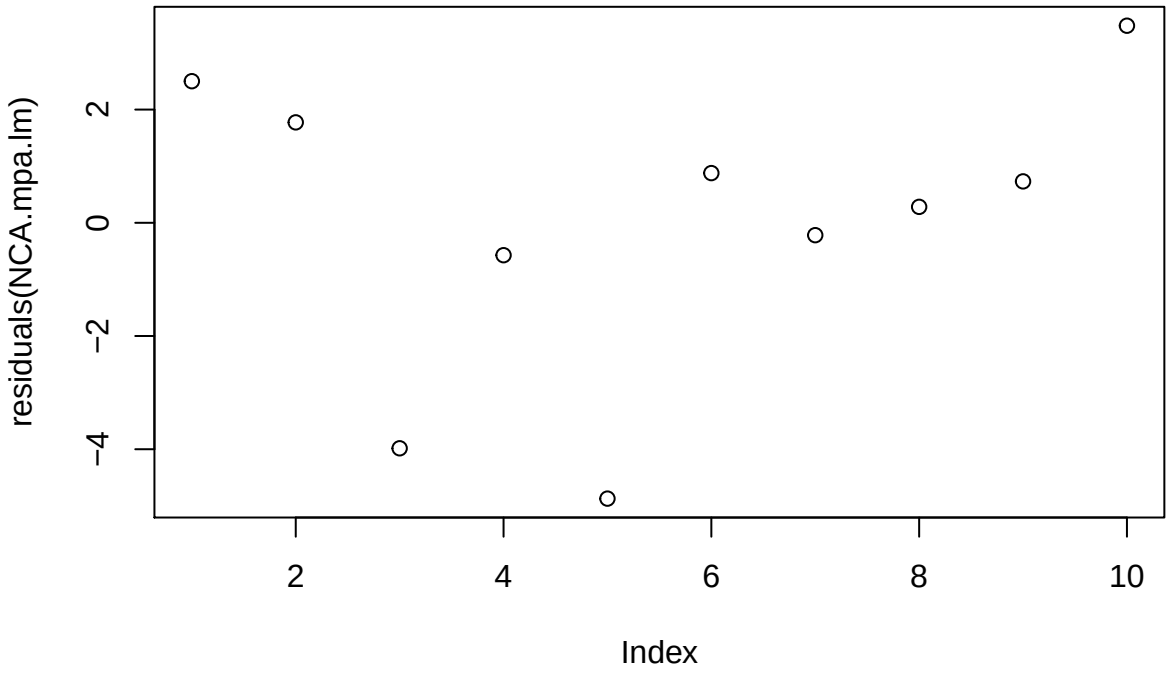
Run North CA

```
# Create northern CA data

NCA <- subset(mod.avg, MPA_Group == "Bodega Bay" |
                MPA_Group == "Farallon Islands")

# Subset by designation

NCA.mpa <- subset(NCA, Designation == "MPA")
NCA.ref <- subset(NCA, Designation == "Reference")

# Run MPA first

set.seed(25)
NCA.mpa.lm <- lm(formula = avg.rich ~ Year, data = NCA.mpa)
summary(NCA.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.rich ~ Year, data = NCA.mpa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8723 -0.4839  0.5063  1.5499  3.4813
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2828.4706   597.0911  -4.737  0.00147 **
## Year            1.4116     0.2962   4.766  0.00142 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.804 on 8 degrees of freedom
## Multiple R-squared:  0.7395, Adjusted R-squared:  0.7069
## F-statistic: 22.71 on 1 and 8 DF,  p-value: 0.001417
```

```r
# (Year: t = 5.486, p = 0.00142)
# (df = 1,8, F = 22.71, p = 0.001417)
# (y~-2828.4706+1.4116(Year))
# (R = 0.7395)

# *** Significant ***

# Check assumptions

plot(residuals(NCA.mpa.lm))
```




```r
shapiro.test(residuals(NCA.mpa.lm))
```

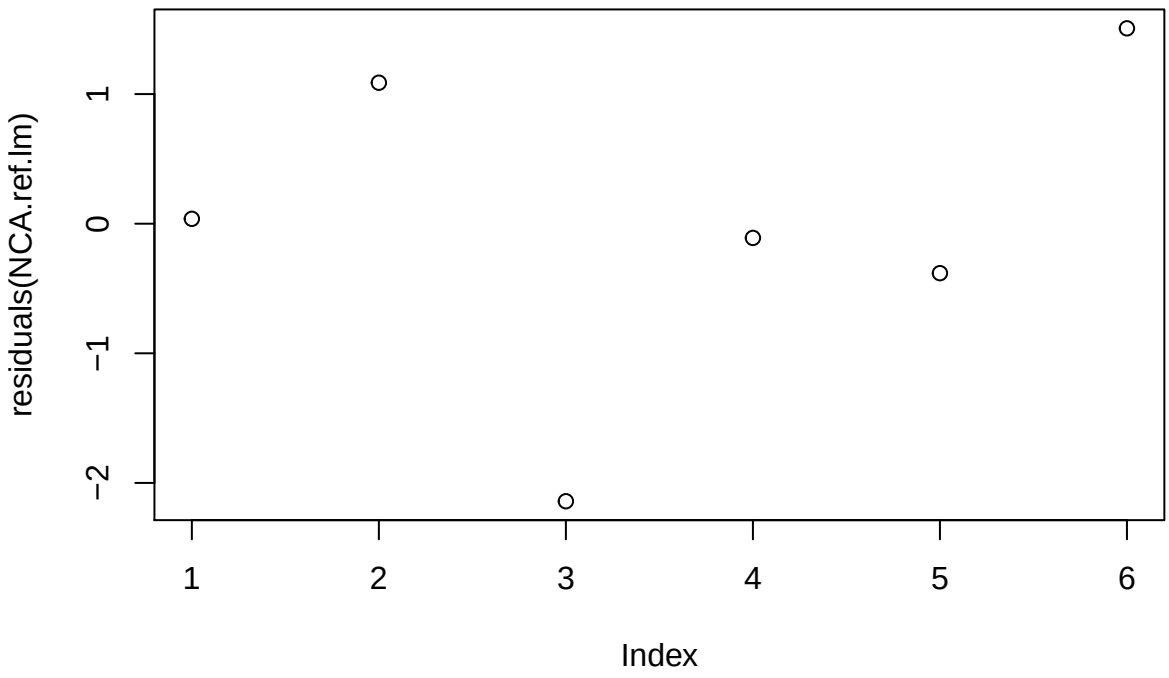
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.mpa.lm)
```

```
## W = 0.91396, p-value = 0.3093
```

```
# Residuals normal

# Run Ref

set.seed(25)
NCA.ref.lm <- lm(formula = avg.rich ~ Year, data = NCA.ref)
summary(NCA.ref.lm)
```

```
##
## Call:
## lm(formula = avg.rich ~ Year, data = NCA.ref)
##
## Residuals:
##       53       55       87       92      123      127
##  0.03772  1.08772 -2.14128 -0.10961 -0.38172  1.50717
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2669.473    360.607  -7.403  0.00178 **
## Year            1.333      0.179   7.448  0.00174 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.432 on 4 degrees of freedom
## Multiple R-squared:  0.9328, Adjusted R-squared:  0.9159
## F-statistic: 55.48 on 1 and 4 DF,  p-value: 0.001735
```

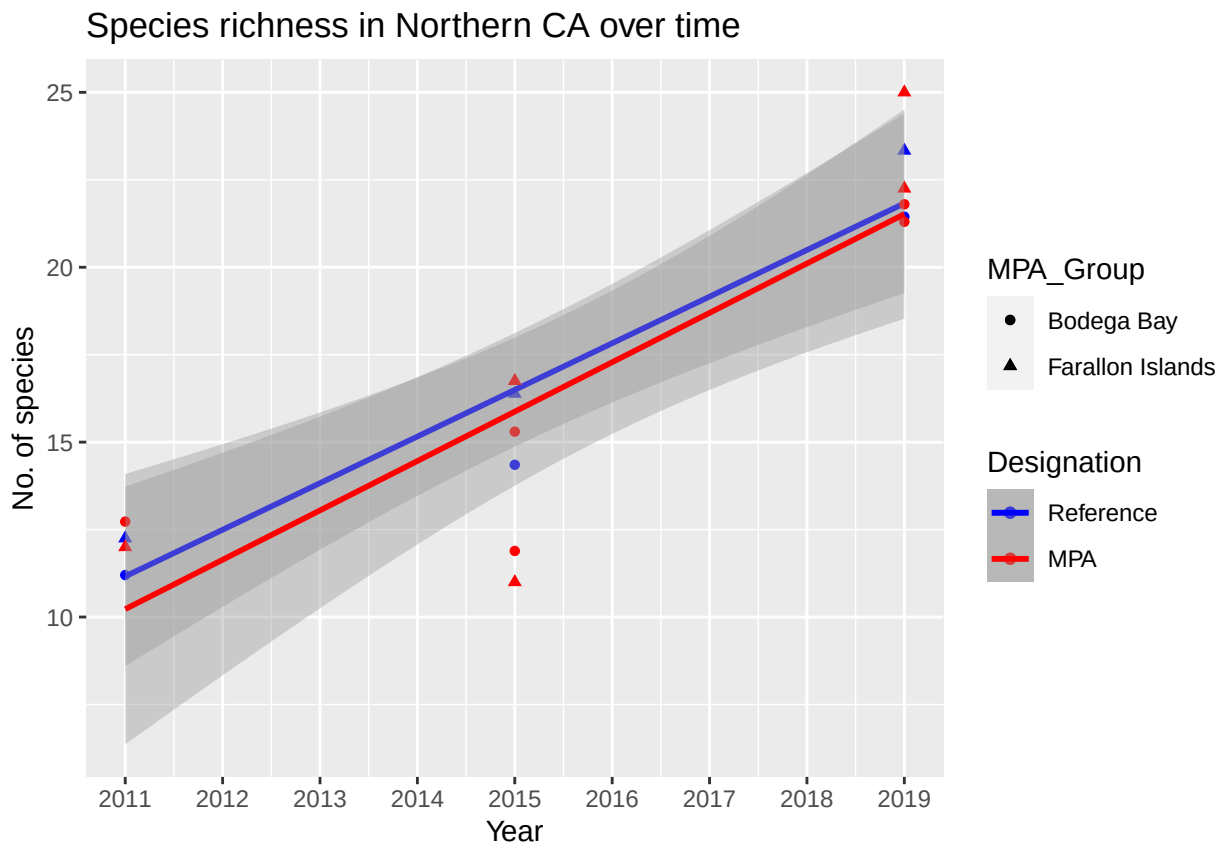
```
# (Year: t = 7.448, p = 0.00174)
# (df = 1,4, F = 55.48, p = 0.001735)
# (y~-2669.473+1.333(Year))
# (R = 0.9328)

# Check assumptions

plot(residuals(NCA.ref.lm))
```




```
shapiro.test(residuals(NCA.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.ref.lm)
## W = 0.93675, p-value = 0.6331
```

```
# Residuals normal

# Plot MPA and Ref regressions

ggplot() +
  geom_point(data=NCA.ref,aes(x=Year,y=avg.rich,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=NCA.ref,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  geom_point(data=NCA.mpa,aes(x=Year,y=avg.rich,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=NCA.mpa,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#0000ff","#ff0000"),
                     breaks = c("Reference","MPA")) +
  ggtitle("Species richness in Northern CA over time") +
  xlab("Year") +
  ylab("No. of species") +
  scale_x_continuous(breaks=seq(2011,2019,1))
```



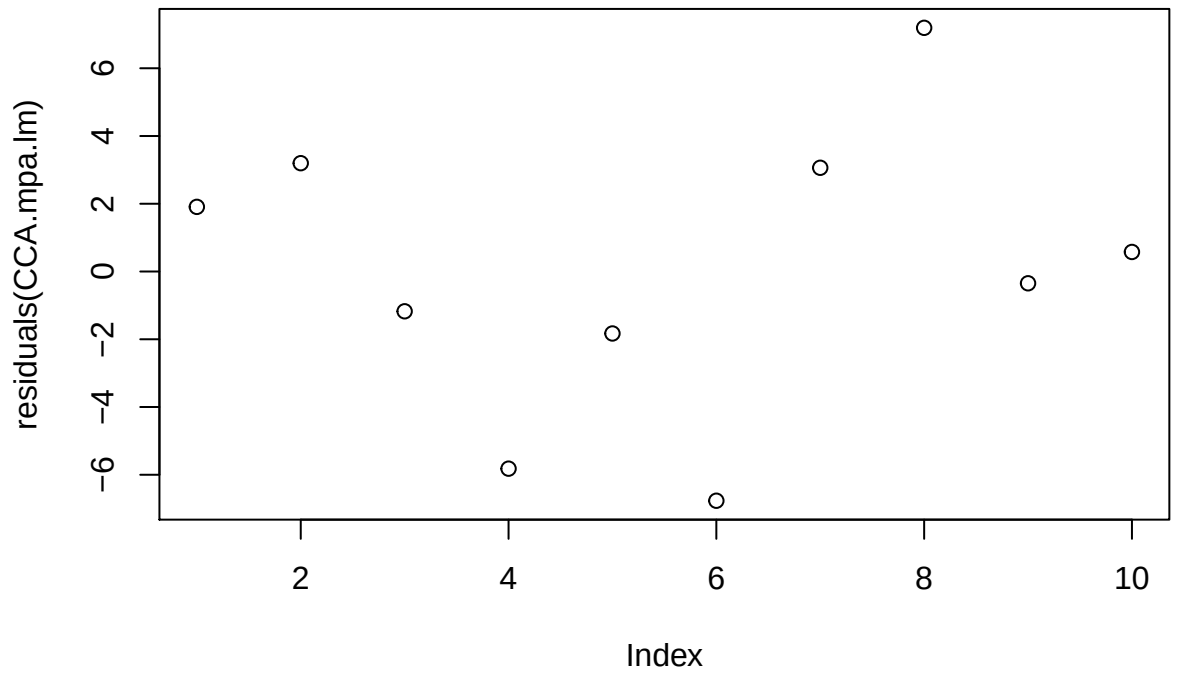
Run Central CA

```
CCA <- subset(mod.avg, MPA_Group == "Point Buchon" |
                MPA_Group == "Point Lobos" | MPA_Group == "Point Sur")
CCA <- subset(CCA, Year == "2008" | Year == "2016" | Year == "2019")

# Subset by designation

CCA.mpa <- subset(CCA, Designation == "MPA")
CCA.ref <- subset(CCA, Designation == "Reference")

# Run MPA first

set.seed(25)
CCA.mpa.lm <- lm(formula = avg.rich ~ Year, data = CCA.mpa)
summary(CCA.mpa.lm)
```

```
##
## Call:
## lm(formula = avg.rich ~ Year, data = CCA.mpa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7668 -1.6656  0.1143  2.7741  7.1935
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1212.3093   616.2127  -1.967   0.0847 .
## Year            0.6105     0.3058   1.996   0.0810 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.47 on 8 degrees of freedom
## Multiple R-squared:  0.3324, Adjusted R-squared:  0.249
## F-statistic: 3.984 on 1 and 8 DF,  p-value: 0.08101
```

```
# (Year: t = 1.996, p = 0.0810)
# (df = 1,8, F = 3.984, p = 0.08101)
# (y~-1212.3093+0.6105(Year))
# (R = 0.3324)

# Test assumptions

plot(residuals(CCA.mpa.lm))
```



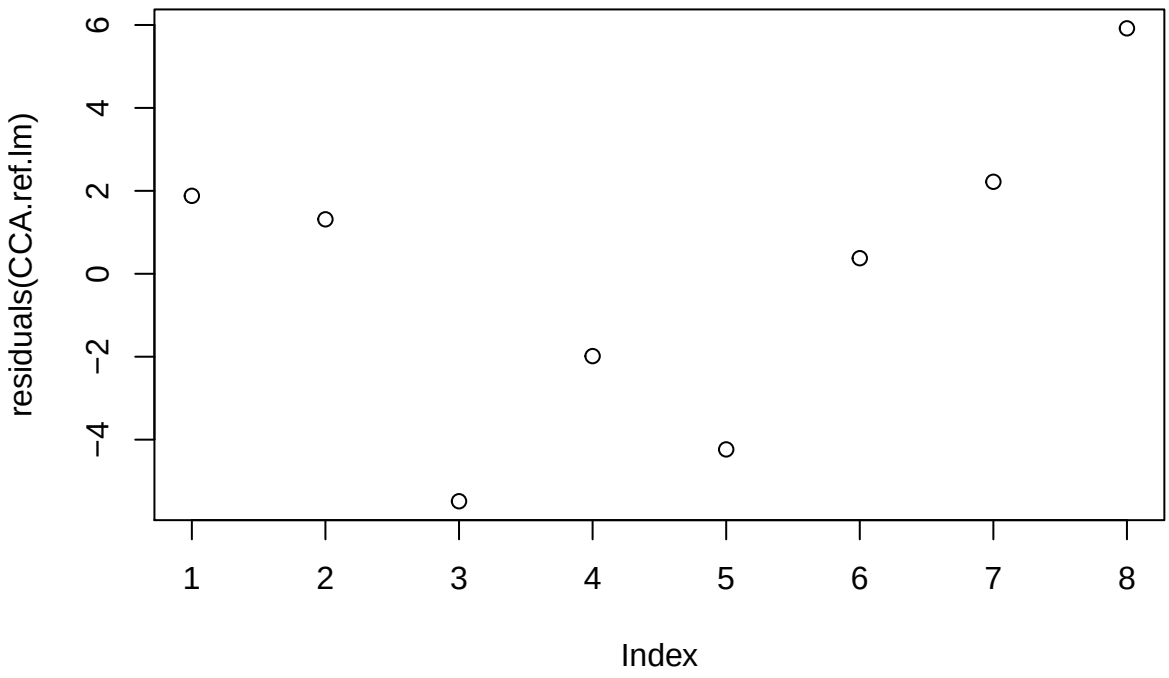

```
shapiro.test(residuals(CCA.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.mpa.lm)
## W = 0.9682, p-value = 0.8737
```

```r
# Residuals normal

# Run Ref

set.seed(25)
CCA.ref.lm <- lm(formula = avg.rich ~ Year, data = CCA.ref)
summary(CCA.ref.lm)
```

```
##
## Call:
## lm(formula = avg.rich ~ Year, data = CCA.ref)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4852 -2.5477  0.8443  1.9643  5.9187
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1574.3024   665.1320  -2.367   0.0558 .
## Year            0.7904     0.3301   2.395   0.0537 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.027 on 6 degrees of freedom
## Multiple R-squared:  0.4887, Adjusted R-squared:  0.4034
## F-statistic: 5.734 on 1 and 6 DF,  p-value: 0.05369
```

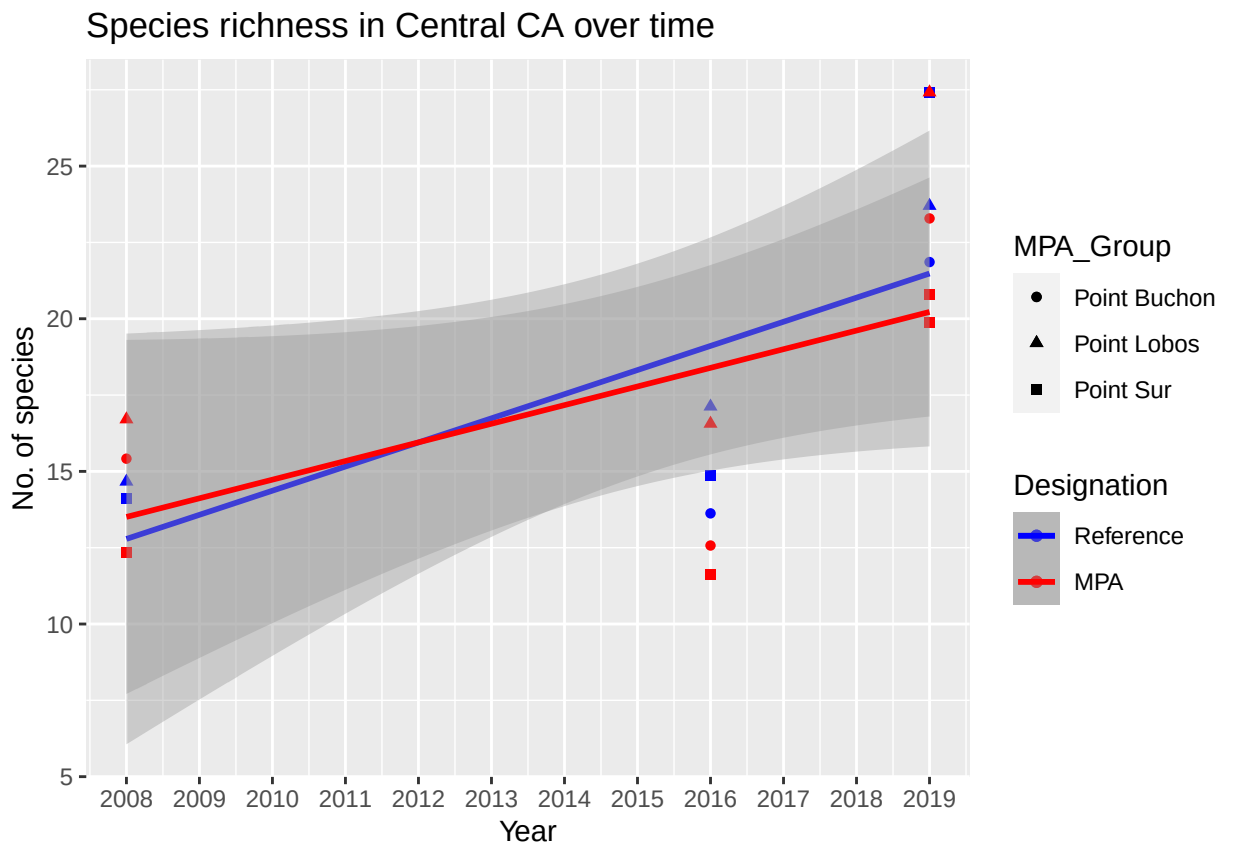
```r
# (Year: t = 2.395, p = 0.0537)
# (df = 1,6, F = 5.734, p = 0.05369)
# (y~-1574.3024+0.7904(Year))
# (R = 0.4887)

# Test assumptions

plot(residuals(CCA.ref.lm))
```




```
shapiro.test(residuals(CCA.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.ref.lm)
## W = 0.9611, p-value = 0.8205
```

```
# Residuals normal

# Plot MPA and Ref regressions

ggplot() +
  geom_point(data=CCA.ref,aes(x=Year,y=avg.rich,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=CCA.ref,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  geom_point(data=CCA.mpa,aes(x=Year,y=avg.rich,colour=Designation,
                              shape=MPA_Group)) +
  stat_smooth(data=CCA.mpa,aes(x=Year,y=avg.rich,colour=Designation),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#0000ff","#ff0000"),
                     breaks = c("Reference","MPA")) +
  ggtitle("Species richness in Central CA over time") +
  xlab("Year") +
  ylab("No. of species") +
  scale_x_continuous(breaks=seq(2008,2019,1))
```



Run ANCOVA and linear regression with latitude data to determine relationship # between latitude and diversity

```
# Data with latitude

data <- data.frame(read.csv("Avg.transects.lat.csv"))

# Add diversity metric

data <- mutate(data, diversity = diversity(data[10:166]))

# Add diversity average

grp.mean <- data %>%
  group_by_all() %>%
  mutate(newnames=paste0(Year, MPA_Group, Type))

split.mean <- split(grp.mean, grp.mean$newnames)
for (I in 1:length(split.mean)) {assign(unique(split.mean[[I]]$newnames),
                                        split.mean[[I]])}

mean.div <- lapply(split.mean, function(x){

  x <- mutate(x, avg.div=mean(x$diversity))
  x <- mutate(x, div.sd=sd(x$diversity))
  x <- mutate(x, div.SE=div.sd/length(x$diversity))
```

```
})

div.avgs <- bind_rows(mean.div)

# Remove excess data

p1 <- div.avgs[2:9]
p2 <- div.avgs[167:171]
mod.avg <- cbind(p1,p2)
mod.avg <- mod.avg[!duplicated(mod.avg[10]),]
mod.avg <- as.data.frame(mod.avg)
mod.avg$Year <- as.factor(mod.avg$Year)
mod.avg <- subset(mod.avg, Designation == "MPA" | Designation == "Reference")

# ANCOVA
# Two-way ANOVA with no interaction

set.seed(25)
MPA.aov = lm(avg.div~Latitude+Year,data=mod.avg)
Anova(MPA.aov,type="III")
```

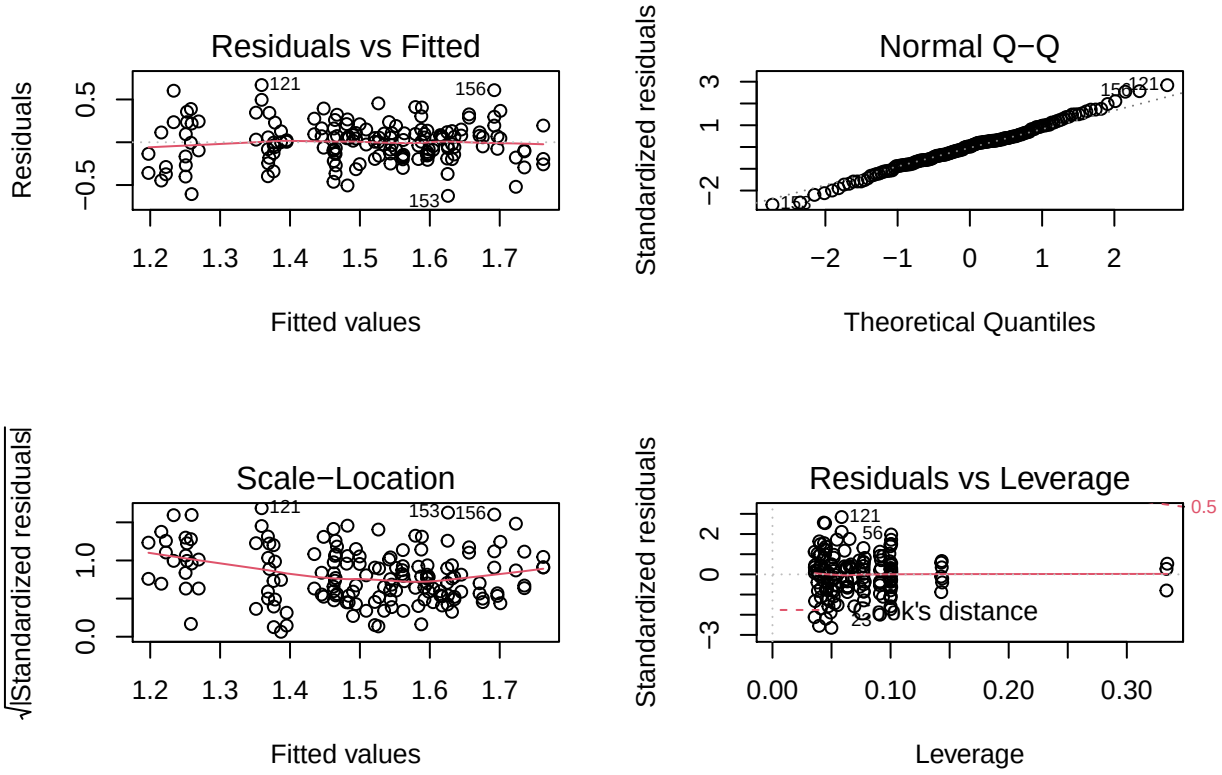
```
## Anova Table (Type III tests)
##
## Response: avg.div
##             Sum Sq  Df F value    Pr(>F)
## (Intercept) 0.0155   1  0.2630  0.608832
## Latitude    0.6320   1 10.7564  0.001298 **
## Year        2.4612  10  4.1886 4.063e-05 ***
## Residuals   8.6378 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(MPA.aov)
```

```
##
## Call:
## lm(formula = avg.div ~ Latitude + Year, data = mod.avg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62620 -0.14366  0.00553  0.12279  0.66821
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.188729   0.368006   0.513  0.60883
## Latitude     0.034685   0.010576   3.280  0.00130 **
## Year2006     0.096763   0.108407   0.893  0.37354
## Year2007     0.093874   0.105937   0.886  0.37700
## Year2008     0.227950   0.097987   2.326  0.02137 *
## Year2009     0.175045   0.108407   1.615  0.10852
## Year2011     0.118376   0.126876   0.933  0.35234
## Year2012     0.120215   0.160298   0.750  0.45449
## Year2014    -0.117802   0.090971  -1.295  0.19737
## Year2015     0.008051   0.094598   0.085  0.93229
```

```
## Year2016     0.144514   0.103920   1.391  0.16644
## Year2019     0.248936   0.094367   2.638  0.00924 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2424 on 147 degrees of freedom
## Multiple R-squared:  0.2667, Adjusted R-squared:  0.2118
## F-statistic: 4.861 on 11 and 147 DF,  p-value: 2.194e-06
```

```
# Latitude and Year are both significant
# Test normality of residuals

par(mfrow=c(2,2))
plot(MPA.aov)
```




```
shapiro.test(MPA.aov$residuals)
```

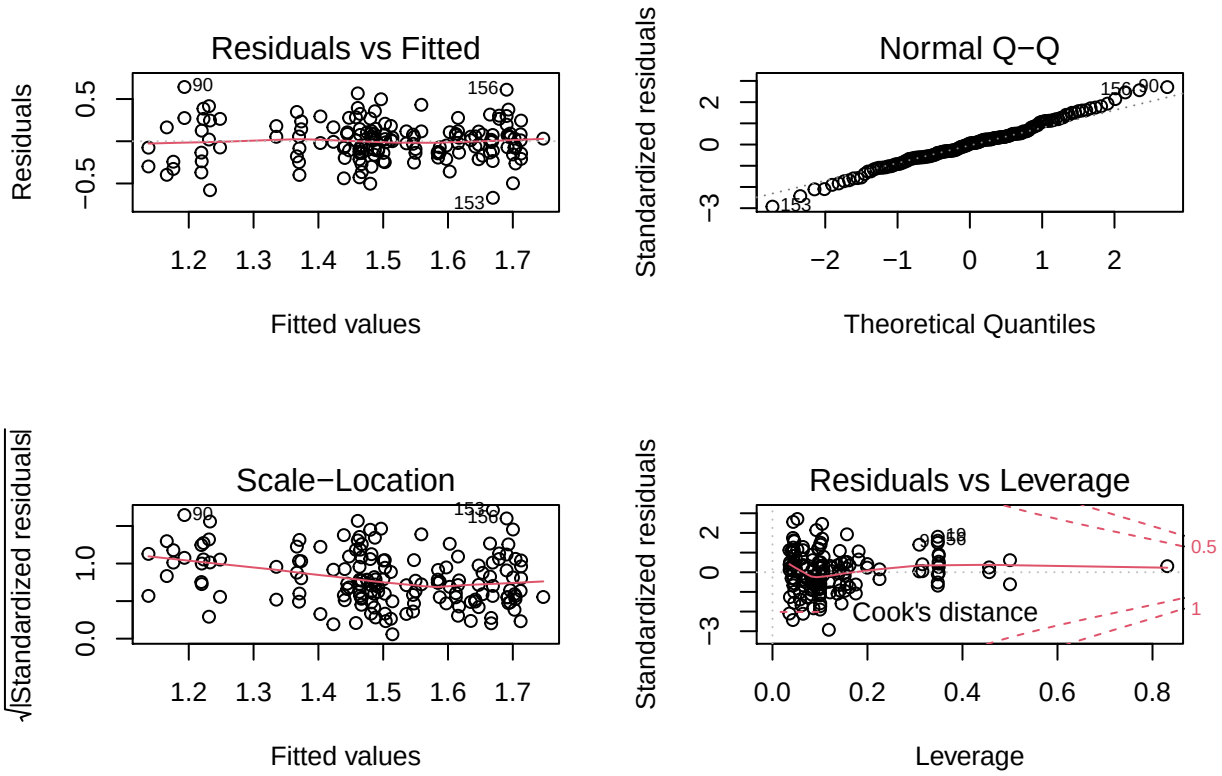
```
##
##  Shapiro-Wilk normality test
##
## data:  MPA.aov$residuals
## W = 0.99359, p-value = 0.7097
```

```
# Data is normal
# Proceed to interaction

MPA.aov2 = aov(avg.div~Latitude*Year,data=mod.avg)
summary(MPA.aov2)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Latitude        1  0.680  0.6805  11.466 0.000925 ***
## Year           10  2.461  0.2461   4.147 5.11e-05 ***
## Latitude:Year  10  0.507  0.0507   0.854 0.577606
## Residuals     137  8.131  0.0593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Interaction is NOT significant
# Assess the assumptions

plot(MPA.aov2)
```

```
## Warning: not plotting observations with leverage one:
##   67
```




```
shapiro.test(MPA.aov2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  MPA.aov2$residuals
## W = 0.99165, p-value = 0.4813
```

```
leveneTest(mod.avg$Latitude, mod.avg$Year)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
```

```
## group  10  5.0726 2.403e-06 ***
##       148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Variance is roughly equal and residuals are roughly normal considering the
# data.

# Run post-hoc on model with no interactions

set.seed(25)
Tukey=glht(MPA.aov, linfct = mcp(Year ="Tukey"))
summary(Tukey)
```

```
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = avg.div ~ Latitude + Year, data = mod.avg)
##
## Linear Hypotheses:
##                    Estimate Std. Error t value Pr(>|t|)
## 2006 - 2005 == 0  0.096763   0.108407   0.893   0.9980
## 2007 - 2005 == 0  0.093874   0.105937   0.886   0.9981
## 2008 - 2005 == 0  0.227950   0.097987   2.326   0.4010
## 2009 - 2005 == 0  0.175045   0.108407   1.615   0.8622
## 2011 - 2005 == 0  0.118376   0.126876   0.933   0.9971
## 2012 - 2005 == 0  0.120215   0.160298   0.750   0.9995
## 2014 - 2005 == 0 -0.117802   0.090971  -1.295   0.9647
## 2015 - 2005 == 0  0.008051   0.094598   0.085   1.0000
## 2016 - 2005 == 0  0.144514   0.103920   1.391   0.9434
## 2019 - 2005 == 0  0.248936   0.094367   2.638   0.2240
## 2007 - 2006 == 0 -0.002889   0.105937  -0.027   1.0000
## 2008 - 2006 == 0  0.131187   0.097987   1.339   0.9556
## 2009 - 2006 == 0  0.078282   0.108407   0.722   0.9997
## 2011 - 2006 == 0  0.021613   0.126876   0.170   1.0000
```

```
## 2012 - 2006 == 0  0.023452   0.160298   0.146   1.0000
## 2014 - 2006 == 0 -0.214565   0.090971  -2.359   0.3804
## 2015 - 2006 == 0 -0.088712   0.094598  -0.938   0.9969
## 2016 - 2006 == 0  0.047751   0.103920   0.459   1.0000
## 2019 - 2006 == 0  0.152173   0.094367   1.613   0.8635
## 2008 - 2007 == 0  0.134076   0.095080   1.410   0.9380
## 2009 - 2007 == 0  0.081171   0.105937   0.766   0.9995
## 2011 - 2007 == 0  0.024503   0.124023   0.198   1.0000
## 2012 - 2007 == 0  0.026341   0.158429   0.166   1.0000
## 2014 - 2007 == 0 -0.211675   0.087583  -2.417   0.3435
## 2015 - 2007 == 0 -0.085823   0.091013  -0.943   0.9968
## 2016 - 2007 == 0  0.050640   0.100908   0.502   1.0000
## 2019 - 2007 == 0  0.155062   0.090999   1.704   0.8175
## 2009 - 2008 == 0 -0.052905   0.097987  -0.540   1.0000
## 2011 - 2008 == 0 -0.109574   0.115436  -0.949   0.9966
## 2012 - 2008 == 0 -0.107735   0.152719  -0.705   0.9997
## 2014 - 2008 == 0 -0.345752   0.076634  -4.512    <0.01 ***
## 2015 - 2008 == 0 -0.219899   0.079652  -2.761   0.1707
## 2016 - 2008 == 0 -0.083436   0.091415  -0.913   0.9975
## 2019 - 2008 == 0  0.020986   0.080237   0.262   1.0000
## 2011 - 2009 == 0 -0.056669   0.126876  -0.447   1.0000
## 2012 - 2009 == 0 -0.054830   0.160298  -0.342   1.0000
## 2014 - 2009 == 0 -0.292847   0.090971  -3.219   0.0529 .
## 2015 - 2009 == 0 -0.166994   0.094598  -1.765   0.7826
## 2016 - 2009 == 0 -0.030531   0.103920  -0.294   1.0000
## 2019 - 2009 == 0  0.073891   0.094367   0.783   0.9993
## 2012 - 2011 == 0  0.001839   0.169519   0.011   1.0000
## 2014 - 2011 == 0 -0.236178   0.105534  -2.238   0.4617
## 2015 - 2011 == 0 -0.110325   0.103082  -1.070   0.9912
## 2016 - 2011 == 0  0.026137   0.115878   0.226   1.0000
## 2019 - 2011 == 0  0.130560   0.106717   1.223   0.9762
## 2014 - 2012 == 0 -0.238017   0.147274  -1.616   0.8615
## 2015 - 2012 == 0 -0.112164   0.148121  -0.757   0.9995
## 2016 - 2012 == 0  0.024299   0.155339   0.156   1.0000
## 2019 - 2012 == 0  0.128721   0.148943   0.864   0.9984
## 2015 - 2014 == 0  0.125853   0.066251   1.900   0.6977
## 2016 - 2014 == 0  0.262315   0.081401   3.223   0.0518 .
## 2019 - 2014 == 0  0.366738   0.068359   5.365    <0.01 ***
## 2016 - 2015 == 0  0.136463   0.082112   1.662   0.8394
## 2019 - 2015 == 0  0.240885   0.068865   3.498   0.0229 *
## 2019 - 2016 == 0  0.104422   0.084129   1.241   0.9736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
# The only years that're different are 2014 from 2008 & 2019 and
# 2015 from 2019
# 2014 had very low diversity, probably due to the heat wave (p < 0.05).

# There was a significant difference found between (average) diversity
# for different latitudes (df = 1,147, F = 10.7564, p = 0.001298) and year
# (df = 10,147, F = 4.1886, p = 4.063e-05).

# Linear regression between diversity and latitude
```

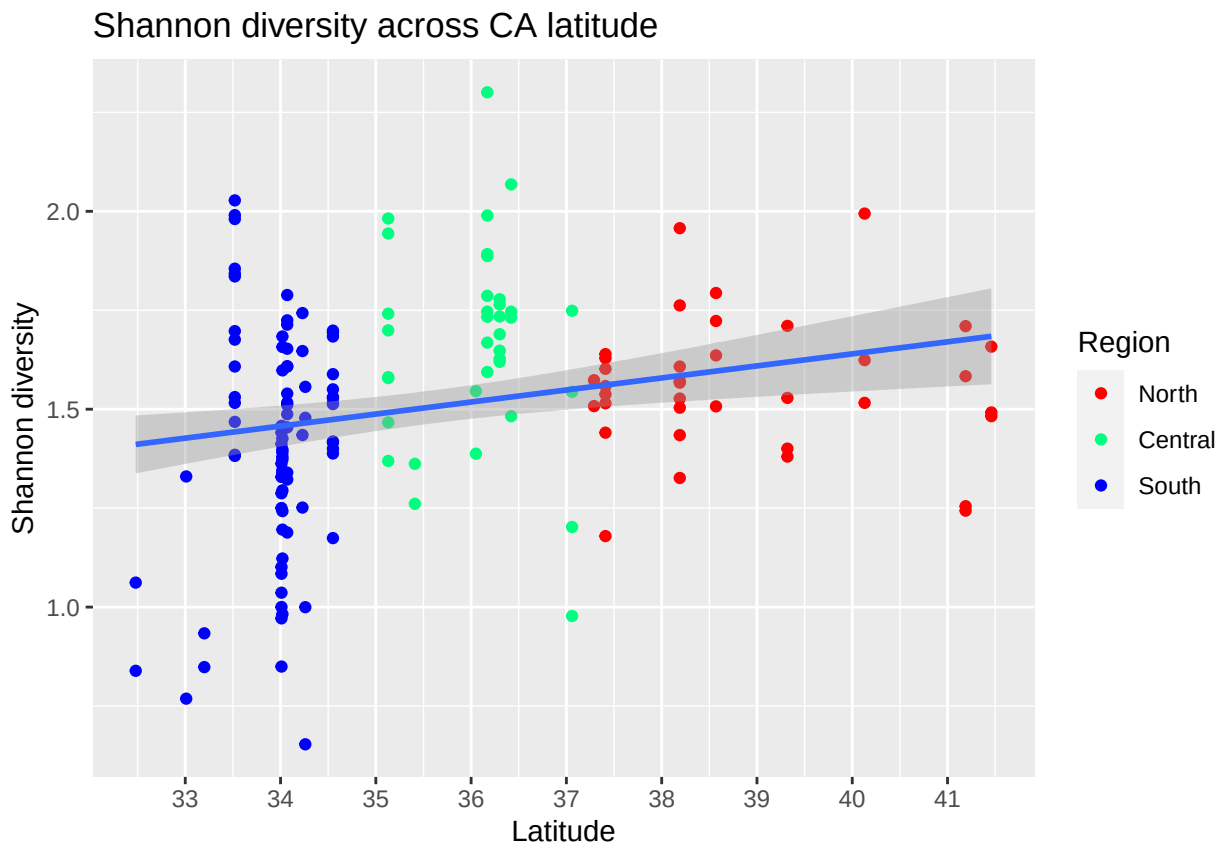
```r
set.seed(25)
lat.lm <- lm(formula = avg.div ~ Latitude, data = mod.avg)
summary(lat.lm)
```

```
##
## Call:
## lm(formula = avg.div ~ Latitude, data = mod.avg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81189 -0.12509 -0.00181  0.18006  0.77732
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.423105   0.349507   1.211  0.22788
## Latitude    0.030417   0.009804   3.103  0.00228 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2659 on 157 degrees of freedom
## Multiple R-squared:  0.05777,    Adjusted R-squared:  0.05177
## F-statistic: 9.626 on 1 and 157 DF,  p-value: 0.002276
```

```r
# (Year: t = 3.103, p = 0.00228)
# (df = 1,157, F = 9.626, p = 0.002276)
# (y~0.423105+0.030417(Latitude))
# (R = 0.05777)

plot(residuals(lat.lm))
ggqqplot(lat.lm$residuals)
shapiro.test(residuals(lat.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lat.lm)
## W = 0.98628, p-value = 0.1197
```

```r
# Residuals are normal

# Plot MPA and Ref regressions for channel islands

ggplot() +
  geom_point(data=mod.avg, aes(x=Latitude,y=avg.div,colour=Region)) +
  stat_smooth(data=mod.avg,aes(x=Latitude,y=avg.div),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#ff0000","#00ff7f","#0000ff"),
                     breaks = c("North","Central","South")) +
  ggtitle("Shannon diversity across CA latitude") +
  xlab("Latitude") +
  ylab("Shannon diversity") +
  scale_x_continuous(breaks=seq(32,42,1))
```



Repeat above, but using richness instead of diversity

```
# Add richness metric

data <- mutate(data, richness = rowSums(data[10:166]!=0))

# Add richness average

grp.mean <- data %>%
  group_by_all() %>%
  mutate(newnames=paste0(Year, MPA_Group, Type))

split.mean <- split(grp.mean, grp.mean$newnames)
for (I in 1:length(split.mean)) {assign(unique(split.mean[[I]]$newnames), split.mean[[I]])}

mean.rich <- lapply(split.mean, function(x){

  x <- mutate(x, avg.rich=mean(x$richness))
  x <- mutate(x, rich.sd=sd(x$richness))
  x <- mutate(x, rich.SE=rich.sd/length(x$richness))

})

rich.avgs <- bind_rows(mean.rich)

# Remove excess data
```

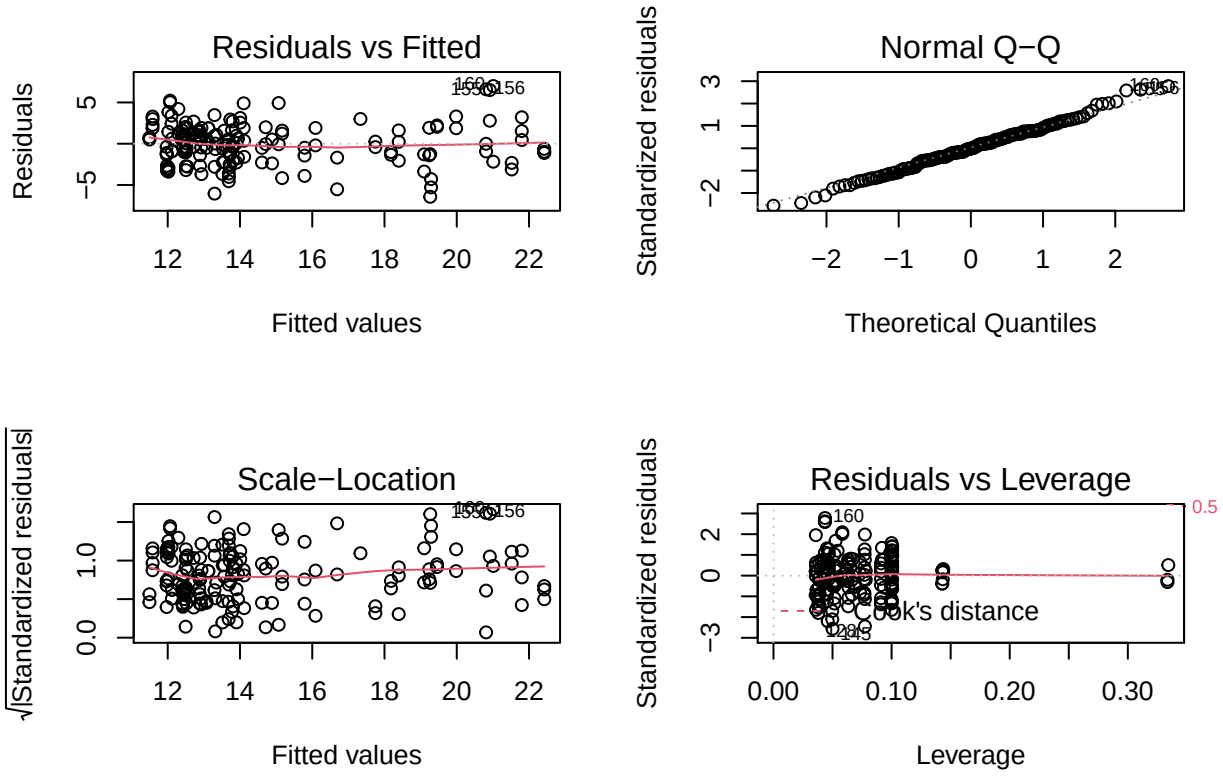
```r
p1 <- rich.avgs[2:9]
p2 <- rich.avgs[167:171]
mod.avg <- cbind(p1,p2)
mod.avg <- mod.avg[!duplicated(mod.avg[11]),]
mod.avg$Year <- as.factor(mod.avg$Year)

# ANCOVA
# Two-way ANOVA with no interaction

set.seed(25)
MPA.aov = lm(avg.rich~Latitude+Year,data=mod.avg)
Anova(MPA.aov,type="III")
```

```
## Anova Table (Type III tests)
##
## Response: avg.rich
##              Sum Sq  Df F value    Pr(>F)
## (Intercept)   93.86   1  14.191 0.0002378 ***
## Latitude     338.25   1  51.142 3.669e-11 ***
## Year        1003.29  10  15.169 < 2.2e-16 ***
## Residuals    978.88 148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(MPA.aov)
```

```
##
## Call:
## lm(formula = avg.rich ~ Latitude + Year, data = mod.avg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4319 -1.4127 -0.0285  1.6126  6.9889
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -14.6173     3.8802  -3.767 0.000238 ***
## Latitude      0.7972     0.1115   7.151 3.67e-11 ***
## Year2006     -0.5229     1.1501  -0.455 0.650017
## Year2007     -0.5256     1.1239  -0.468 0.640740
## Year2008      0.3904     1.0395   0.376 0.707813
## Year2009      1.1897     1.1501   1.034 0.302654
## Year2011     -3.7118     1.3451  -2.760 0.006520 **
## Year2012     -0.8707     1.7006  -0.512 0.609405
## Year2014      1.0187     0.9588   1.062 0.289764
## Year2015     -0.0402     1.0029  -0.040 0.968078
## Year2016     -0.3149     1.1023  -0.286 0.775542
## Year2019      6.5939     1.0008   6.588 7.30e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.572 on 148 degrees of freedom
## Multiple R-squared:  0.6022, Adjusted R-squared:  0.5726
## F-statistic: 20.37 on 11 and 148 DF,  p-value: < 2.2e-16
```

```
# Latitude and Year are both significant
# Test normality of residuals

par(mfrow=c(2,2))
plot(MPA.aov)
```




```
shapiro.test(MPA.aov$residuals)
```

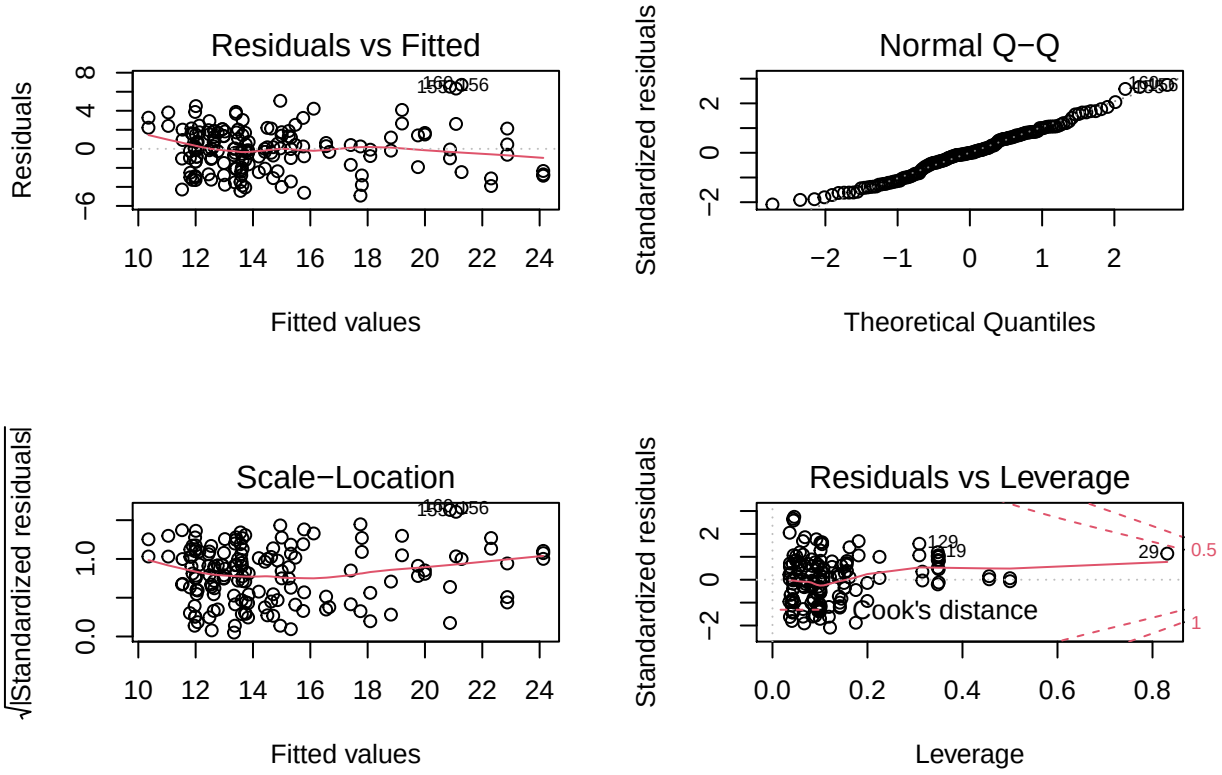
```
##
##  Shapiro-Wilk normality test
##
## data:  MPA.aov$residuals
## W = 0.99397, p-value = 0.7516
```

```
# Data is approximately normal for the number of data points
# Proceed to interaction

MPA.aov2 = aov(avg.rich~Latitude*Year,data=mod.avg)
summary(MPA.aov2)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Latitude        1  478.4   478.4  76.148 7.63e-15 ***
## Year           10 1003.3   100.3  15.970  < 2e-16 ***
## Latitude:Year  10  111.9    11.2   1.781   0.0695 .
## Residuals     138  867.0     6.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Interaction is NOT significant
# Assess the assumptions

plot(MPA.aov2)
```

```
## Warning: not plotting observations with leverage one:
##   67
```




```
shapiro.test(MPA.aov2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  MPA.aov2$residuals
## W = 0.98486, p-value = 0.07874
```

```
leveneTest(mod.avg$Latitude, mod.avg$Year)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
## group  10  4.9085 3.993e-06 ***
##       149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Variance is roughly equal and residuals are roughly normal considering the
# data.
```

```r
# Run post-hoc on model with no interactions

set.seed(25)
Tukey=glht(MPA.aov, linfct = mcp(Year ="Tukey"))
summary(Tukey)
```

```
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = avg.rich ~ Latitude + Year, data = mod.avg)
##
## Linear Hypotheses:
##                   Estimate Std. Error t value Pr(>|t|)
## 2006 - 2005 == 0 -0.52292    1.15013  -0.455  1.00000
## 2007 - 2005 == 0 -0.52558    1.12393  -0.468  0.99999
## 2008 - 2005 == 0  0.39037    1.03954   0.376  1.00000
## 2009 - 2005 == 0  1.18965    1.15013   1.034  0.99322
## 2011 - 2005 == 0 -3.71183    1.34509  -2.760  0.17186
## 2012 - 2005 == 0 -0.87072    1.70057  -0.512  0.99999
## 2014 - 2005 == 0  1.01867    0.95880   1.062  0.99168
## 2015 - 2005 == 0 -0.04020    1.00293  -0.040  1.00000
## 2016 - 2005 == 0 -0.31487    1.10227  -0.286  1.00000
## 2019 - 2005 == 0  6.59392    1.00084   6.588  < 0.001 ***
## 2007 - 2006 == 0 -0.00265    1.12393  -0.002  1.00000
## 2008 - 2006 == 0  0.91329    1.03954   0.879  0.99821
## 2009 - 2006 == 0  1.71258    1.15013   1.489  0.91319
## 2011 - 2006 == 0 -3.18890    1.34509  -2.371  0.37242
## 2012 - 2006 == 0 -0.34779    1.70057  -0.205  1.00000
## 2014 - 2006 == 0  1.54160    0.95880   1.608  0.86509
## 2015 - 2006 == 0  0.48272    1.00293   0.481  0.99999
## 2016 - 2006 == 0  0.20806    1.10227   0.189  1.00000
## 2019 - 2006 == 0  7.11685    1.00084   7.111  < 0.001 ***
## 2008 - 2007 == 0  0.91594    1.00873   0.908  0.99764
## 2009 - 2007 == 0  1.71523    1.12393   1.526  0.89971
## 2011 - 2007 == 0 -3.18625    1.31491  -2.423  0.33931
## 2012 - 2007 == 0 -0.34514    1.68076  -0.205  1.00000
## 2014 - 2007 == 0  1.54425    0.92291   1.673  0.83327
## 2015 - 2007 == 0  0.48537    0.96496   0.503  0.99999
## 2016 - 2007 == 0  0.21071    1.07036   0.197  1.00000
## 2019 - 2007 == 0  7.11950    0.96516   7.376  < 0.001 ***
## 2009 - 2008 == 0  0.79929    1.03954   0.769  0.99943
## 2011 - 2008 == 0 -4.10220    1.22396  -3.352  0.03617 *
```

```
## 2012 - 2008 == 0 -1.26108    1.62023  -0.778  0.99937
## 2014 - 2008 == 0  0.62831    0.80665   0.779  0.99937
## 2015 - 2008 == 0 -0.43057    0.84457  -0.510  0.99999
## 2016 - 2008 == 0 -0.70523    0.96974  -0.727  0.99965
## 2019 - 2008 == 0  6.20356    0.85108   7.289  < 0.001 ***
## 2011 - 2009 == 0 -4.90148    1.34509  -3.644  0.01415 *
## 2012 - 2009 == 0 -2.06037    1.70057  -1.212  0.97773
## 2014 - 2009 == 0 -0.17098    0.95880  -0.178  1.00000
## 2015 - 2009 == 0 -1.22986    1.00293  -1.226  0.97572
## 2016 - 2009 == 0 -1.50452    1.10227  -1.365  0.94968
## 2019 - 2009 == 0  5.40427    1.00084   5.400  < 0.001 ***
## 2012 - 2011 == 0  2.84111    1.79819   1.580  0.87774
## 2014 - 2011 == 0  4.73050    1.11810   4.231  0.00184 **
## 2015 - 2011 == 0  3.67162    1.09355   3.358  0.03497 *
## 2016 - 2011 == 0  3.39696    1.22910   2.764  0.16984
## 2019 - 2011 == 0 10.30575    1.13193   9.105  < 0.001 ***
## 2014 - 2012 == 0  1.88939    1.55977   1.211  0.97779
## 2015 - 2012 == 0  0.83051    1.57136   0.529  0.99998
## 2016 - 2012 == 0  0.55585    1.64804   0.337  1.00000
## 2019 - 2012 == 0  7.46464    1.58018   4.724  < 0.001 ***
## 2015 - 2014 == 0 -1.05888    0.69909  -1.515  0.90395
## 2016 - 2014 == 0 -1.33354    0.85928  -1.552  0.88960
## 2019 - 2014 == 0  5.57525    0.72033   7.740  < 0.001 ***
## 2016 - 2015 == 0 -0.27466    0.87106  -0.315  1.00000
## 2019 - 2015 == 0  6.63413    0.73053   9.081  < 0.001 ***
## 2019 - 2016 == 0  6.90879    0.89256   7.740  < 0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```r
# 2019 is different from all other years (p < 0.05).
# 2011 is different from 2008,2009,2014,2015 (p < 0.05)

# There was a significant difference found between (average) richness
# for different latitudes (df = 1,148, F = 51.142, p = 3.669e-11) and year
# (df = 10,148, F = 15.169, p = 2.2e-16).

# Linear regression between richness and latitude

set.seed(25)
lat.lm <- lm(formula = avg.rich ~ Latitude, data = mod.avg)
summary(lat.lm)
```

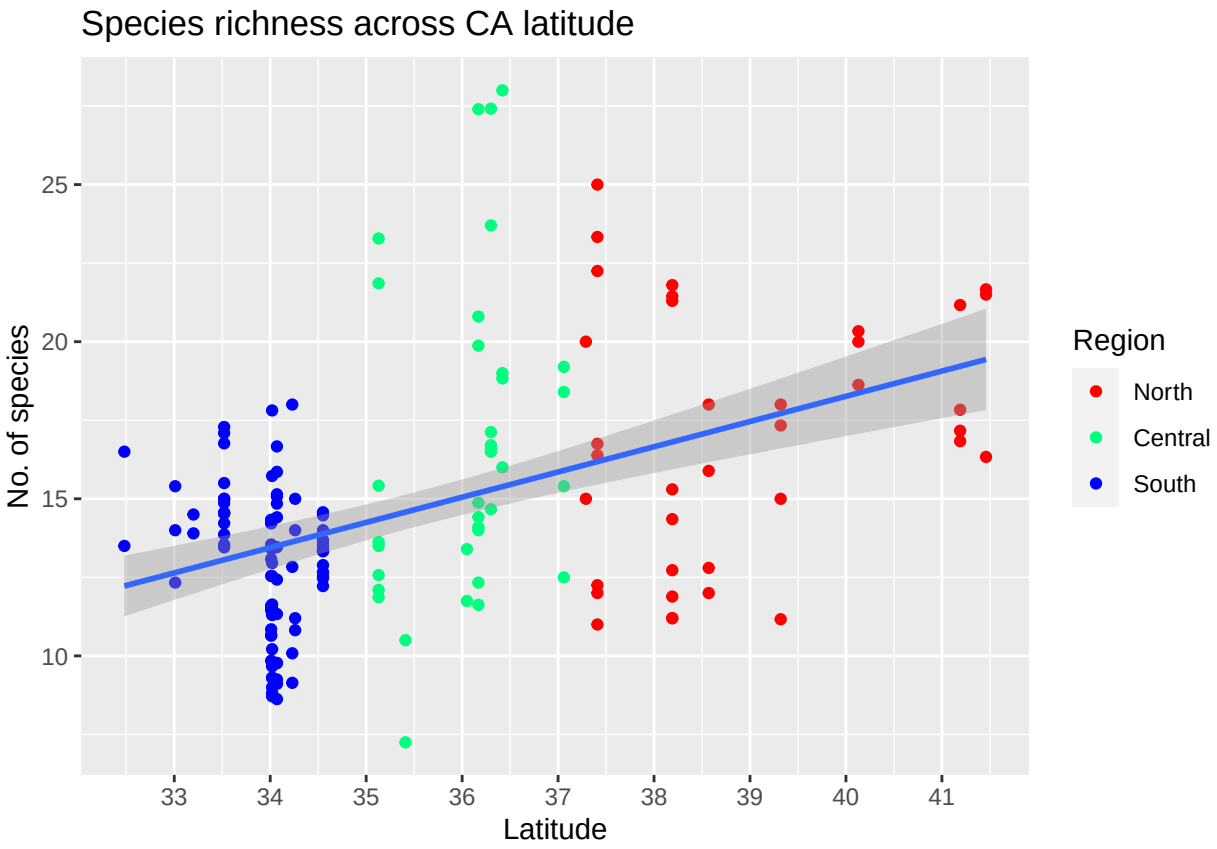
```
##
## Call:
## lm(formula = avg.rich ~ Latitude, data = mod.avg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3269 -2.4017 -0.3479  1.6355 12.6122
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.8545     4.6332  -2.990  0.00323 **
## Latitude      0.8029     0.1300   6.175 5.36e-09 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.542 on 158 degrees of freedom
## Multiple R-squared:  0.1944, Adjusted R-squared:  0.1893
## F-statistic: 38.13 on 1 and 158 DF,  p-value: 5.361e-09
```

```
# (Year: t = 6.175, p = 5.36e-09)
# (df = 1,158, F = 38.13, p = 5.361e-09)
# (y~-13.8545+0.8029(Latitude))
# (R = 0.1944)

plot(residuals(lat.lm))
ggqqplot(lat.lm$residuals)
shapiro.test(residuals(lat.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lat.lm)
## W = 0.94707, p-value = 1.016e-05
```

```
# Residuals are about as close to normal as you can ask for this data set

# Plot MPA and Ref regressions for channel islands

ggplot() +
  geom_point(data=mod.avg, aes(x=Latitude,y=avg.rich,colour=Region)) +
  stat_smooth(data=mod.avg,aes(x=Latitude,y=avg.rich),
              method = lm, formula = y~x) +
  scale_color_manual(values= c("#ff0000","#00ff7f","#0000ff"),
                     breaks = c("North","Central","South")) +
  ggtitle("Species richness across CA latitude") +
  xlab("Latitude") +
  ylab("No. of species") +
  scale_x_continuous(breaks=seq(32,42,1))
```


Species richness across CA latitude
No. of species
25
20
15
10
33
34
35
36
37
38
39
40
41
Latitude
Region
North
Central
South