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.

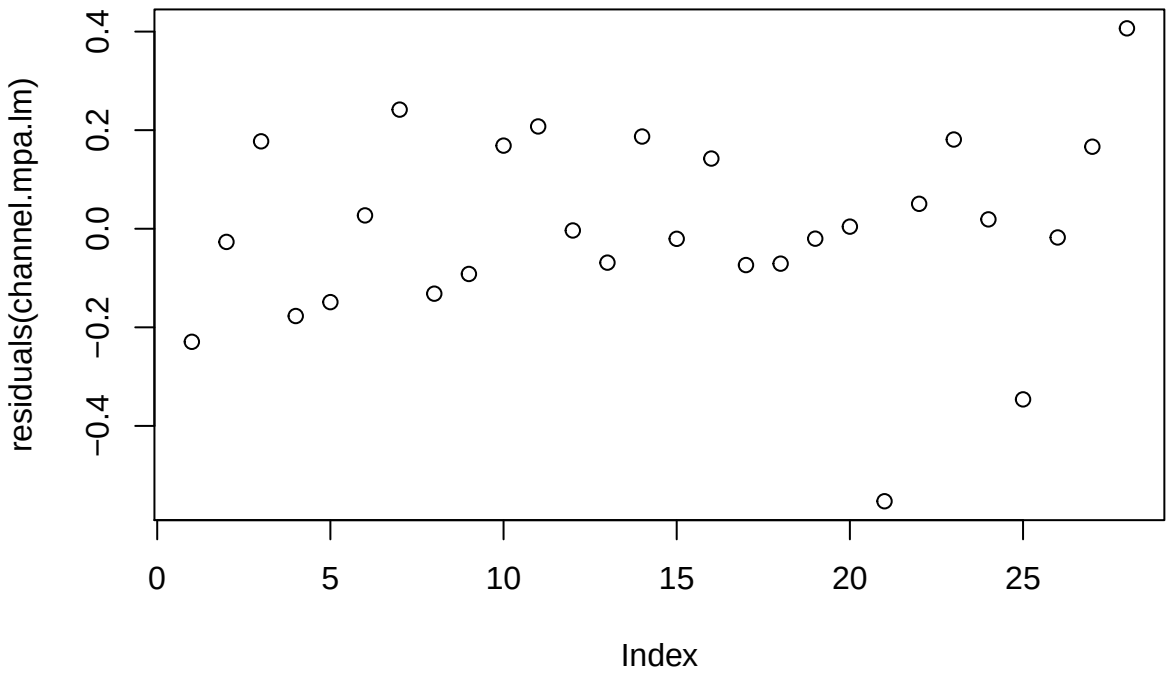
To start, average diversity over time for MPA and reference sites will be run

```
# Import data

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

# Add diversity metric

data <- mutate(data, diversity = diversity(data[9:165]))

# 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[1:8]
p2 <- div.avgs[166:170]
mod.avg <- cbind(p1,p2)
mod.avg <- mod.avg[!duplicated(mod.avg[11]),]
mod.avg <- as.data.frame(mod.avg)
```

```r
# 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

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
## -0.5528 -0.0782 -0.0106  0.1670  0.4066
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.22048   20.87972   1.304    0.204
## Year        -0.01273    0.01039  -1.225    0.231
##
## Residual standard error: 0.1981 on 26 degrees of freedom
## Multiple R-squared:  0.0546, Adjusted R-squared:  0.01824
## F-statistic: 1.502 on 1 and 26 DF,  p-value: 0.2314
```

```r
# (Year: t = -1.225, p = 0.231)
# (df = 1,26, F = 1.502, p = 0.2314)
# (y~27.22048-01273(Year))
# (R = 0.0546)

# Check assumptions

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




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

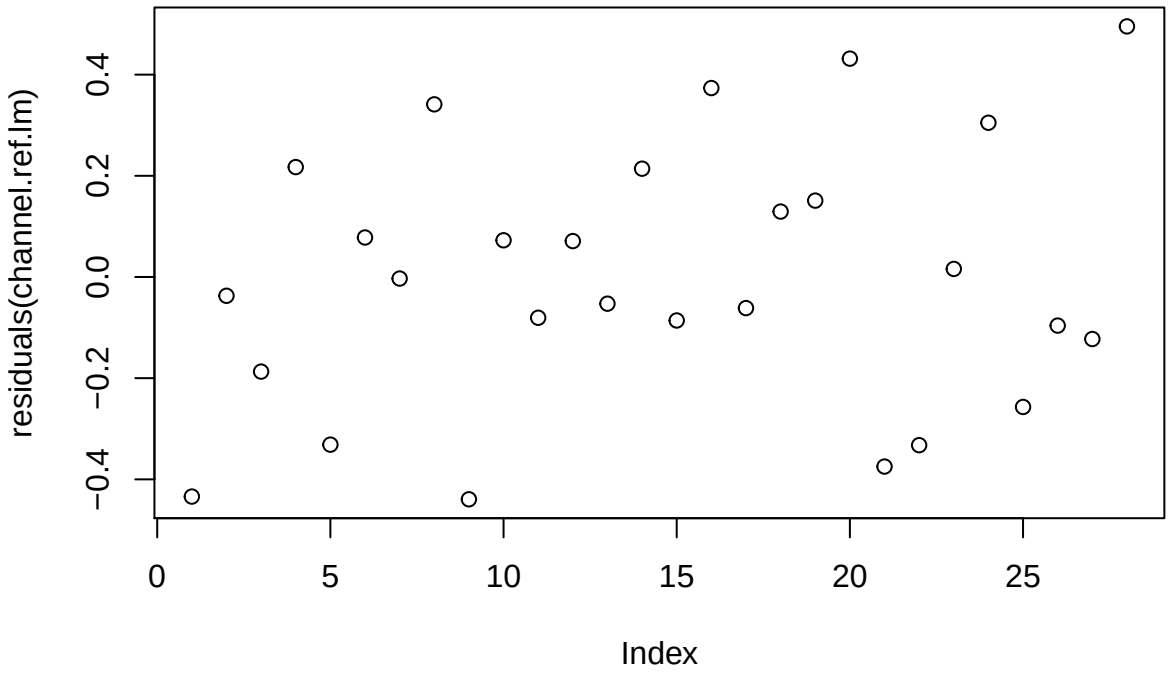
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.mpa.lm)
## W = 0.95641, p-value = 0.2857
```

```
# Residuals normal

# Run Ref

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
## -0.43944 -0.13876 -0.02013  0.16675  0.49534
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.268823  27.877540  -0.225    0.824
## Year         0.003932   0.013875   0.283    0.779
```

```
##
## Residual standard error: 0.2645 on 26 degrees of freedom
## Multiple R-squared:  0.003079,   Adjusted R-squared:  -0.03526
## F-statistic: 0.0803 on 1 and 26 DF,  p-value: 0.7791
```

```
# (Year: t = 0.283, p = 0.779)
# (df = 1,26, F = 0.0803, p = 0.7791)
# (y~-6.268823+0.003932(Year))
# (R = 0.003079)

# Check assumptions

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

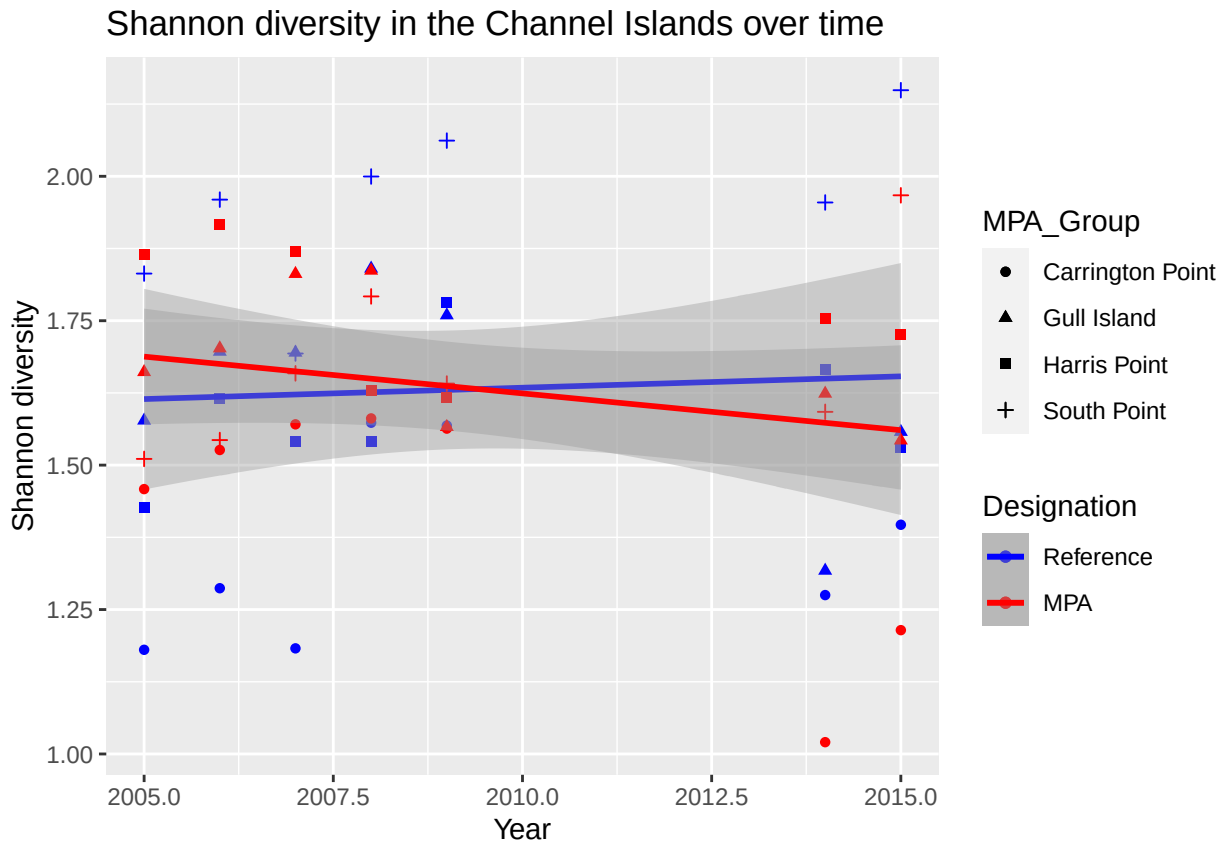



```
shapiro.test(residuals(channel.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.ref.lm)
## W = 0.97329, p-value = 0.6708
```

```
# 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") +
  xlim(2005,2015)
```



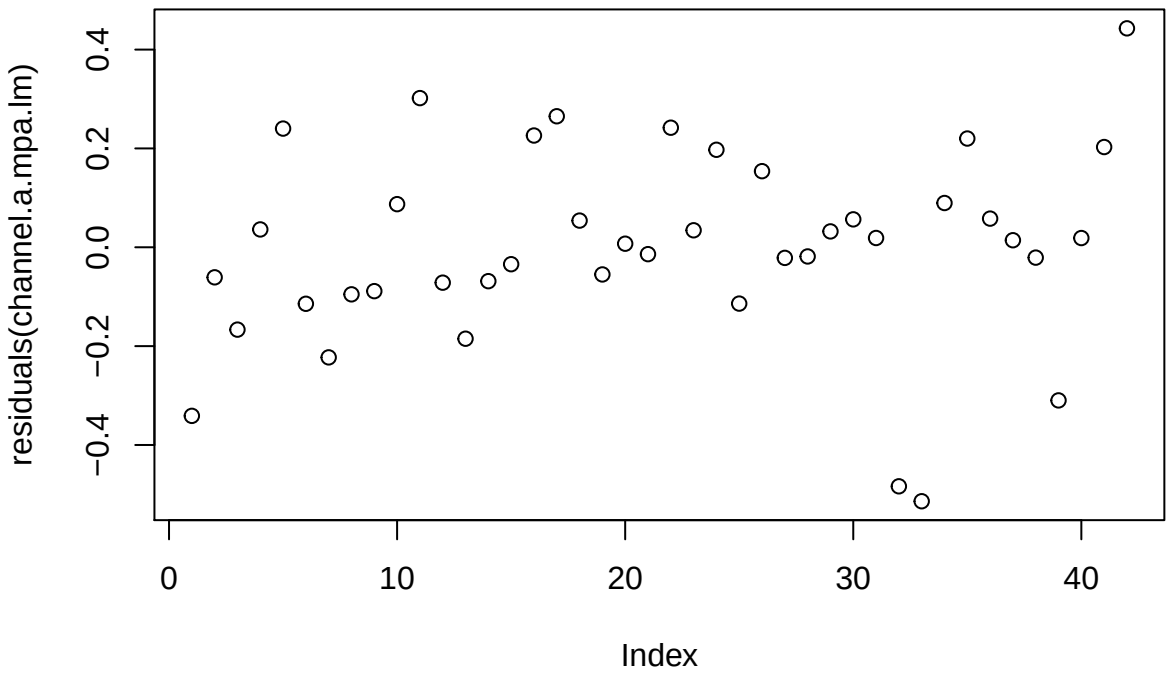
Run the channel islands with Anacapa Island

```
# 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

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
## -0.51380 -0.08445  0.01075  0.08900  0.44296
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.863719  17.159406   1.274    0.210
## Year        -0.010094   0.008541  -1.182    0.244
##
## Residual standard error: 0.1994 on 40 degrees of freedom
## Multiple R-squared:  0.03374,    Adjusted R-squared:  0.009587
## F-statistic: 1.397 on 1 and 40 DF,  p-value: 0.2442
```

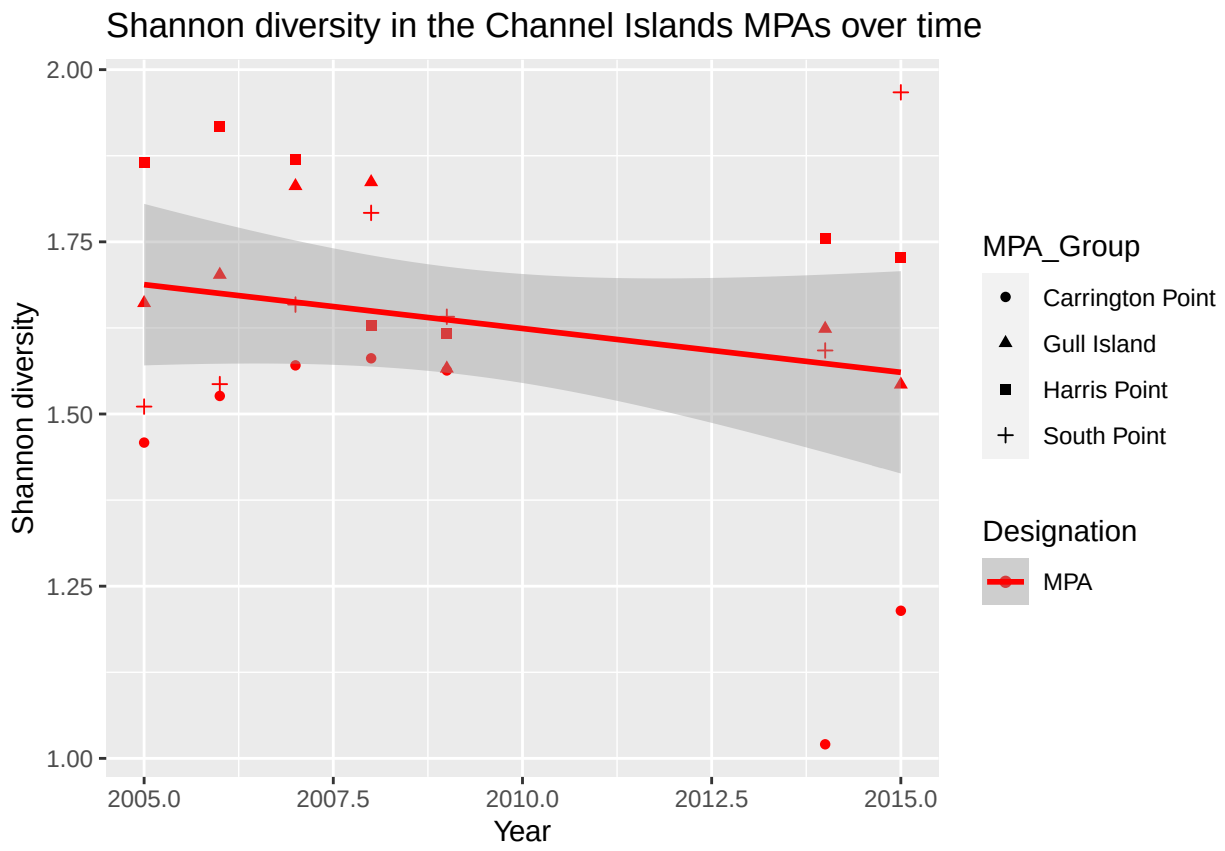
```
# (Year: t = -1.182, p = 0.244)
# (df = 1,40, F = 1.397, p = 0.2442)
# (y~21.863719-0.010094(Year))
# (R = 0.03374)

# 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 = 0.96342, p-value = 0.1955
```

```
# Residuals normal

# Plot

ggplot() +
  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("#ff0000"),
                     breaks = c("MPA")) +
  ggtitle("Shannon diversity in the Channel Islands MPAs over time") +
  xlab("Year") +
  ylab("Shannon diversity") +
  xlim(2005,2015)
```



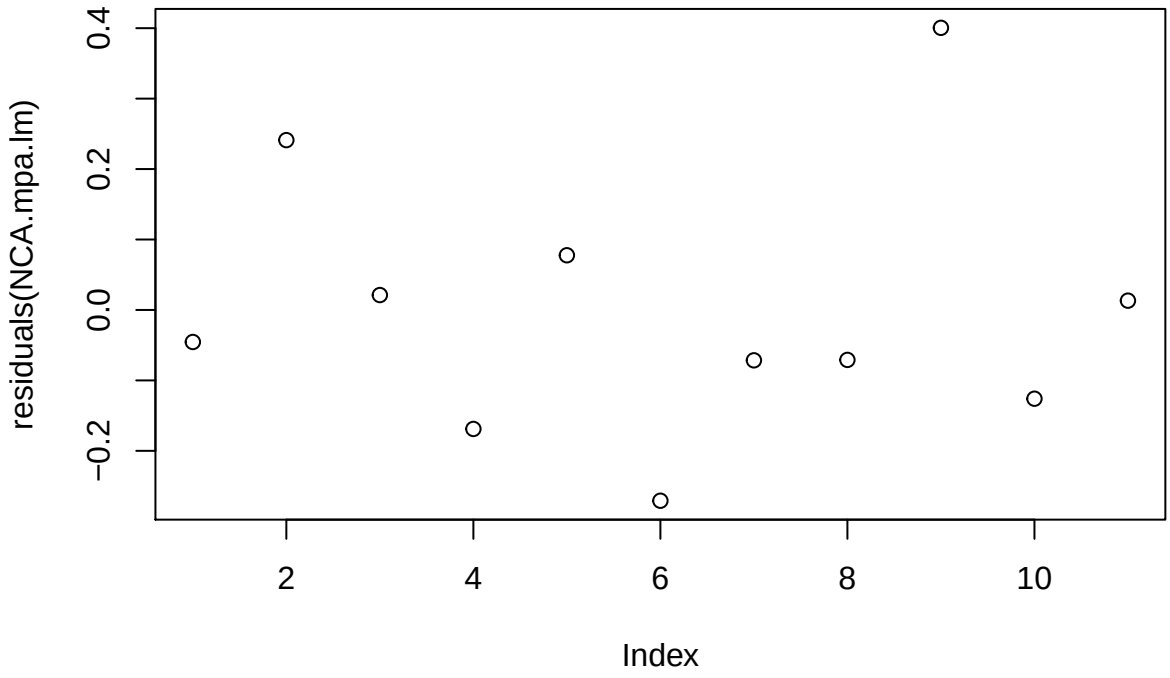
Run North CA

```
# 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

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
## -0.27080 -0.09874 -0.04547  0.04940  0.40048
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.860638  38.074789   0.154    0.881
## Year        -0.002087   0.018892  -0.110    0.914
##
## Residual standard error: 0.1986 on 9 degrees of freedom
## Multiple R-squared:  0.001354,   Adjusted R-squared:  -0.1096
## F-statistic: 0.0122 on 1 and 9 DF,  p-value: 0.9145
```

```
# (Year: t = -0.110, p = 0.914)
# (df = 1,9, F = 0.0122, p = 0.9145)
# (y~5.860638-0.002087(Year))
# (R = 0.001354)

# *** Significant ***

# Test assumptions

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



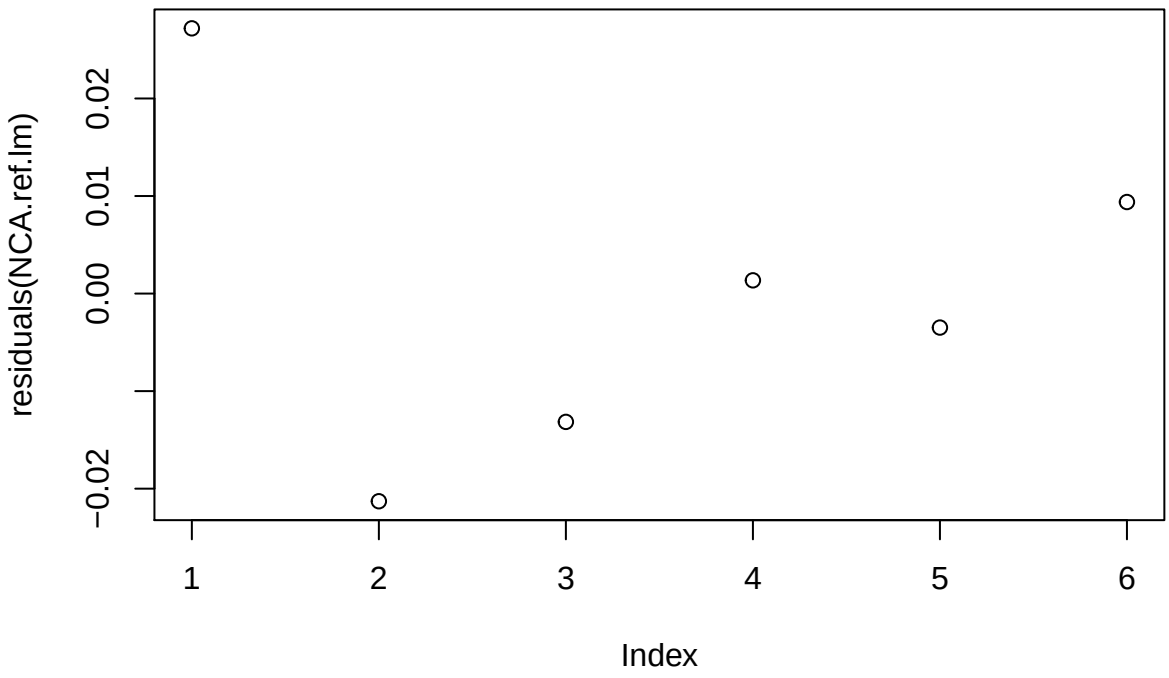

```
shapiro.test(residuals(NCA.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.mpa.lm)
## W = 0.93761, p-value = 0.4928
```

```r
# Residuals normal

# Run Ref

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
##  0.027190 -0.021290 -0.013155  0.001356 -0.003487  0.009387
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 28.038268   4.824751   5.811  0.00436 **
## Year        -0.013072   0.002394  -5.459  0.00547 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01916 on 4 degrees of freedom
## Multiple R-squared:  0.8817, Adjusted R-squared:  0.8521
## F-statistic: 29.81 on 1 and 4 DF,  p-value: 0.005472
```

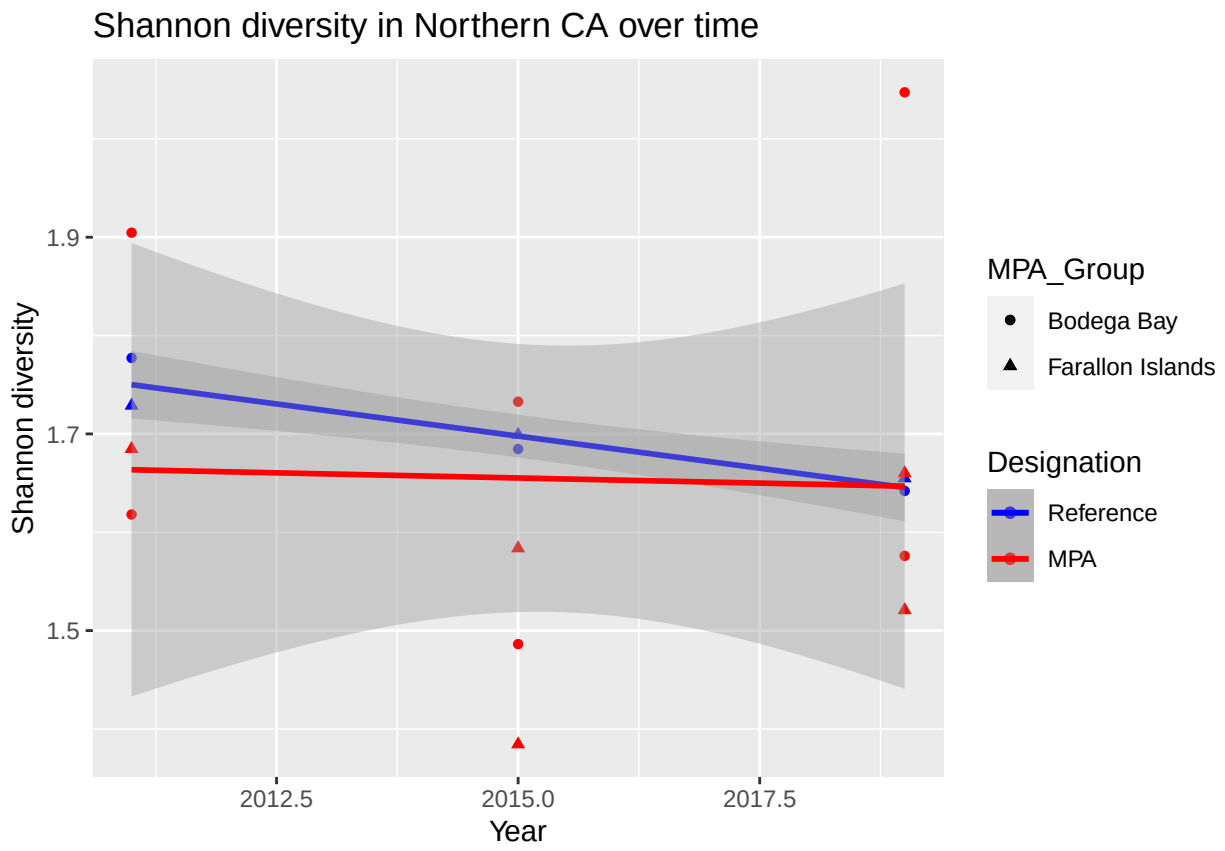
```r
# (Year: t = -5.459, p = 0.00547)
# (df = 1,4, F = 29.81, p = 0.005472)
# (y~28.038268-0.013072(Year))
# (R = 0.8817)

# Test assumptions

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




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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.ref.lm)
## W = 0.98003, p-value = 0.9517
```

```
# 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") +
  xlim(2011,2019)
```



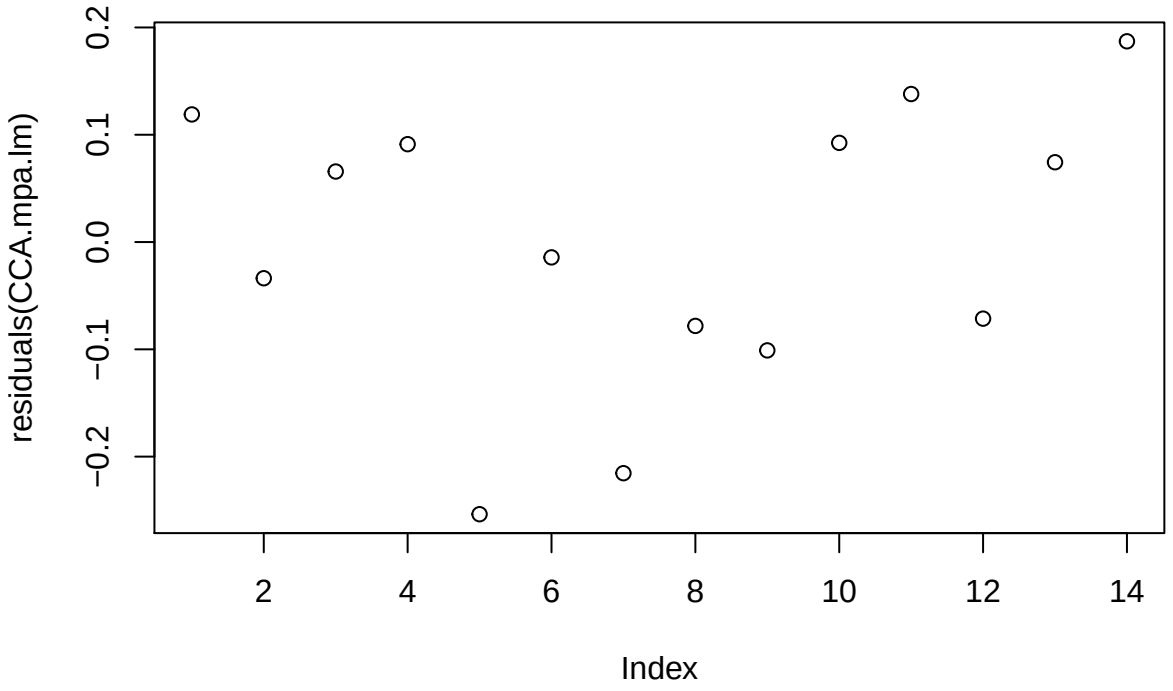
Run Central CA

```
# Create central CA data

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

# Subset by designation

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

# Run MPA first

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
## -0.25367 -0.07648  0.02571  0.09214  0.18708
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -19.439771  16.577573  -1.173    0.264
## Year          0.010569   0.008231   1.284    0.223
##
## Residual standard error: 0.138 on 12 degrees of freedom
## Multiple R-squared:  0.1208, Adjusted R-squared:  0.04752
## F-statistic: 1.649 on 1 and 12 DF,  p-value: 0.2234
```

```
# (Year: t = 1.284, p = 0.223)
# (df = 1,12, F = 1.649, p = 0.2234)
# (y~-19.439771+0.010569(Year))
# (R = 0.1208)

# Check assumptions

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




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

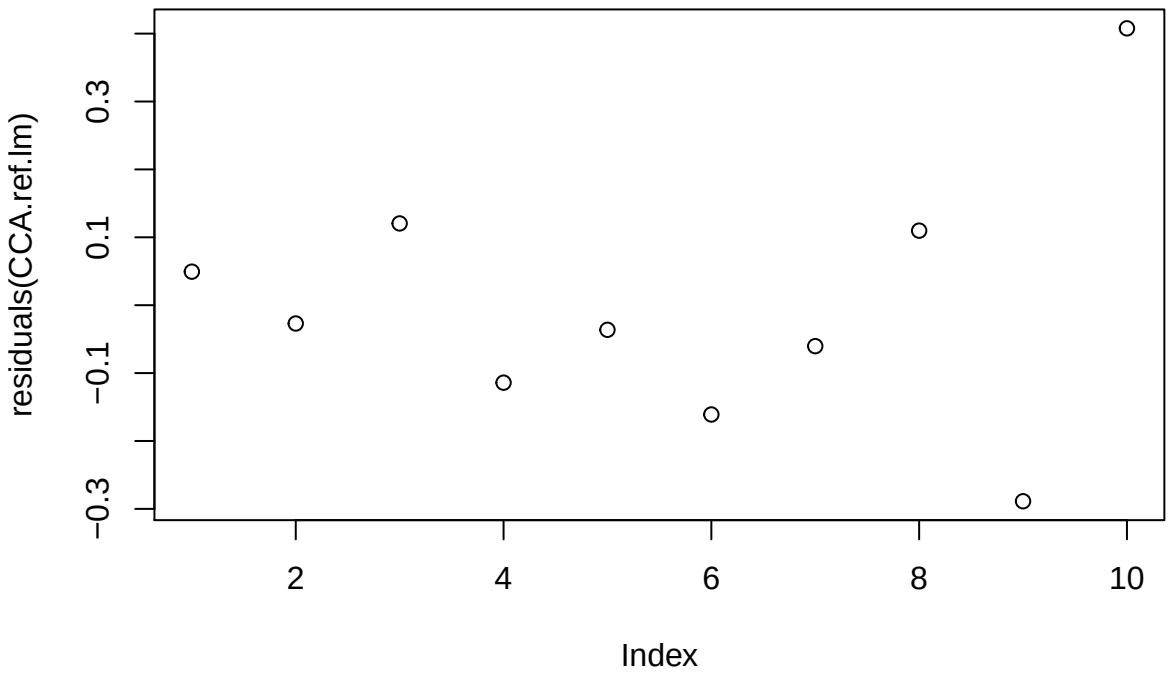
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.mpa.lm)
## W = 0.94057, p-value = 0.4257
```

```
# Residuals normal

# Run Ref
```

```r
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
## -0.28877 -0.10067 -0.03159  0.09470  0.40778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.68876   28.70878  -0.895    0.397
## Year          0.01369    0.01425   0.961    0.365
##
## Residual standard error: 0.201 on 8 degrees of freedom
## Multiple R-squared:  0.1034, Adjusted R-squared:  -0.008637
## F-statistic: 0.9229 on 1 and 8 DF,  p-value: 0.3648
```

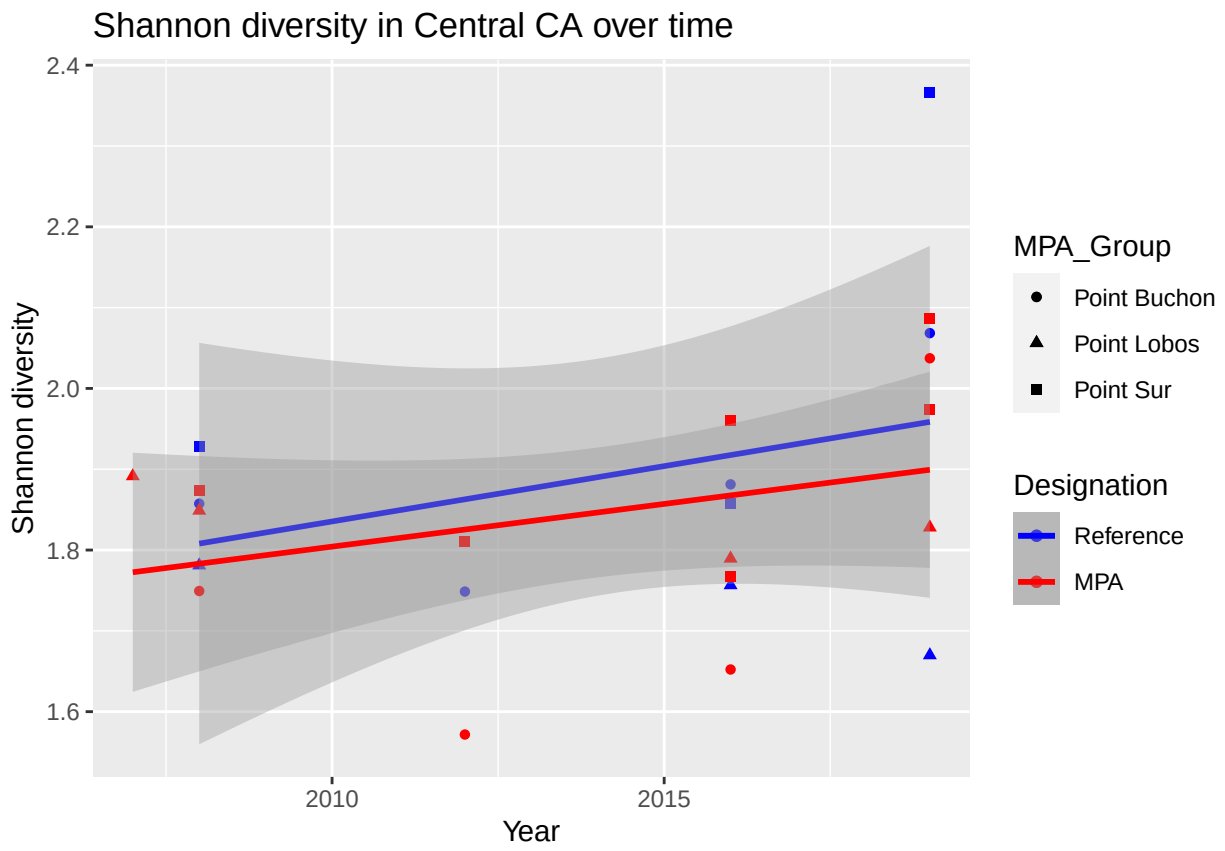
```r
# (Year: t = 0.961, p = 0.365)
# (df = 1,8, F = 0.9229, p = 0.3648)
# (y~-25.68876+0.01369(Year))
# (R = 0.1034)

# Check assumptions

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




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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.ref.lm)
## W = 0.94949, p-value = 0.6625
```

```
# 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") +
  xlim(2007,2019)
```



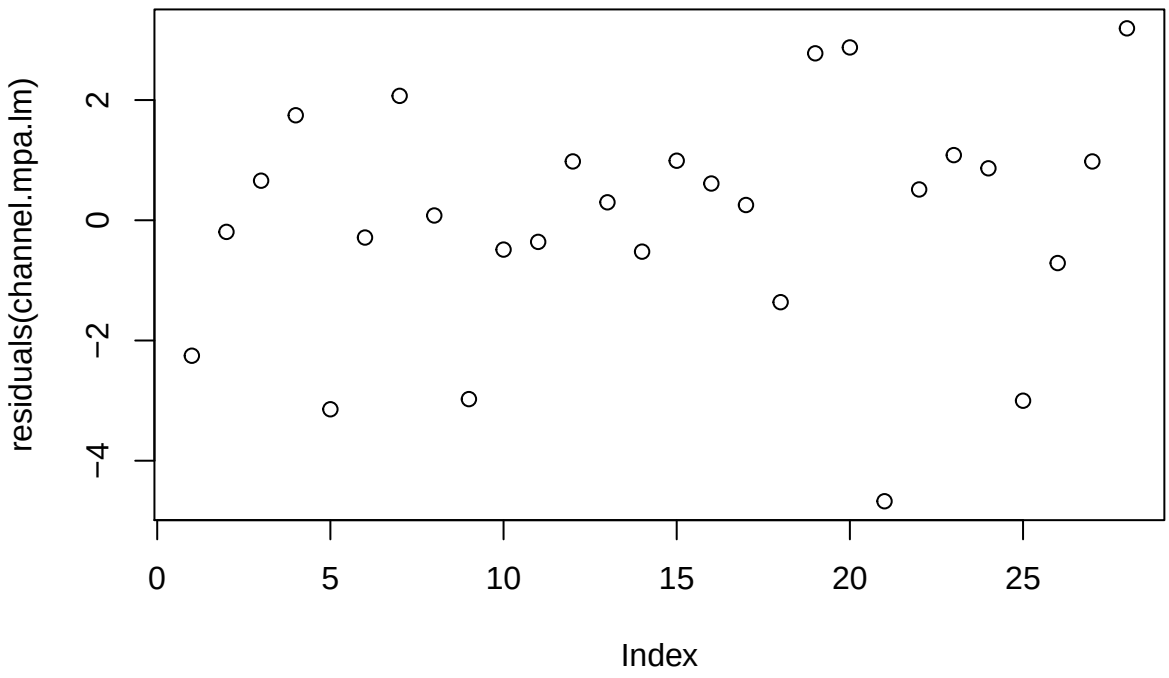
Run again, but using species richness instead of diversity

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

# Add richness metric

data <- mutate(data, richness = rowSums(data[9:165]!=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)
```

```r
p1 <- rich.avgs[1:8]
p2 <- rich.avgs[167:171]
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

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
## -4.6743 -0.5693  0.2762  0.9818  3.1929
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -53.31948  203.94446  -0.261    0.796
## Year          0.03395    0.10151   0.334    0.741
##
## Residual standard error: 1.935 on 26 degrees of freedom
## Multiple R-squared:  0.004284,   Adjusted R-squared:  -0.03401
## F-statistic: 0.1119 on 1 and 26 DF,  p-value: 0.7407
```

```r
# (Year: t = 0.334, p = 0.741)
# (df = 1,23, F = 0.1119, p = 0.7407)
# (y~-53.31948+0.03395(Year))
# (R = 0.004284)

# Check assumptions

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




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

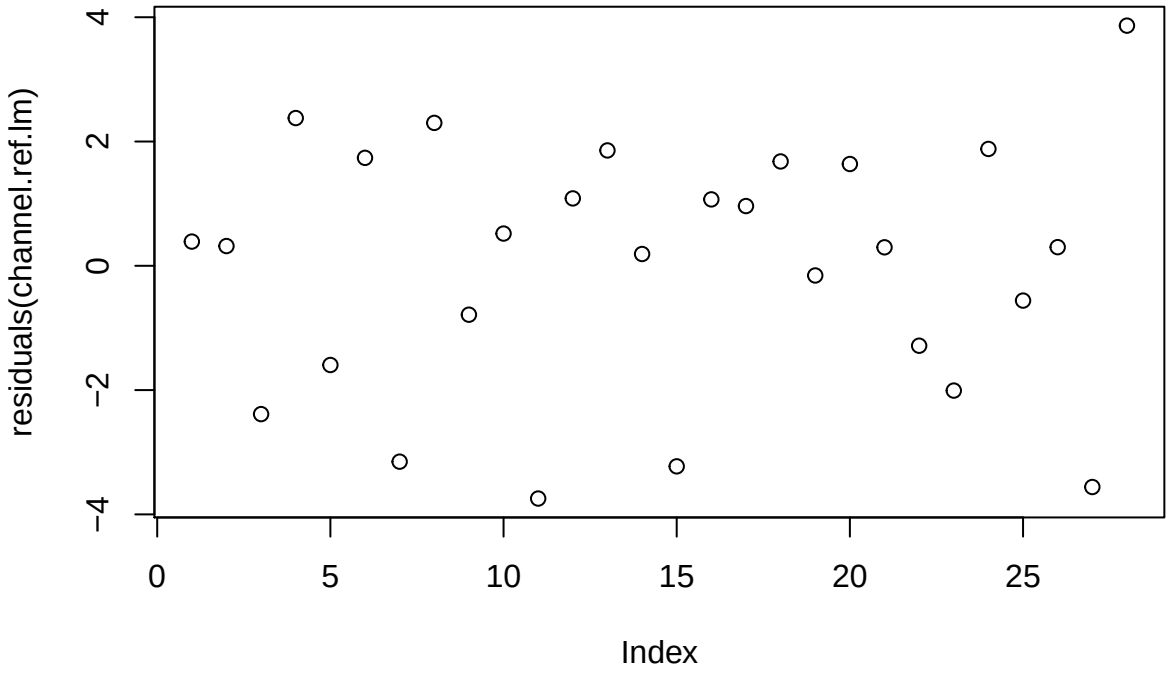
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.mpa.lm)
## W = 0.95194, p-value = 0.2216
```

```
# Residuals normal

# Run Ref

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
## -3.7449 -1.3635  0.3089  1.6485  3.8646
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -201.4798   214.0869  -0.941    0.355
## Year           0.1071     0.1066   1.005    0.324
```

```
##
## Residual standard error: 2.031 on 26 degrees of freedom
## Multiple R-squared:  0.03737,    Adjusted R-squared:  0.0003441
## F-statistic: 1.009 on 1 and 26 DF,  p-value: 0.3243
```

```
# (Year: t = 1.005, p = 0.324)
# (df = 1,24, F = 1.009, p = 0.3243)
# (y~-201.4798+0.1071(Year))
# (R = 0.03737)

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

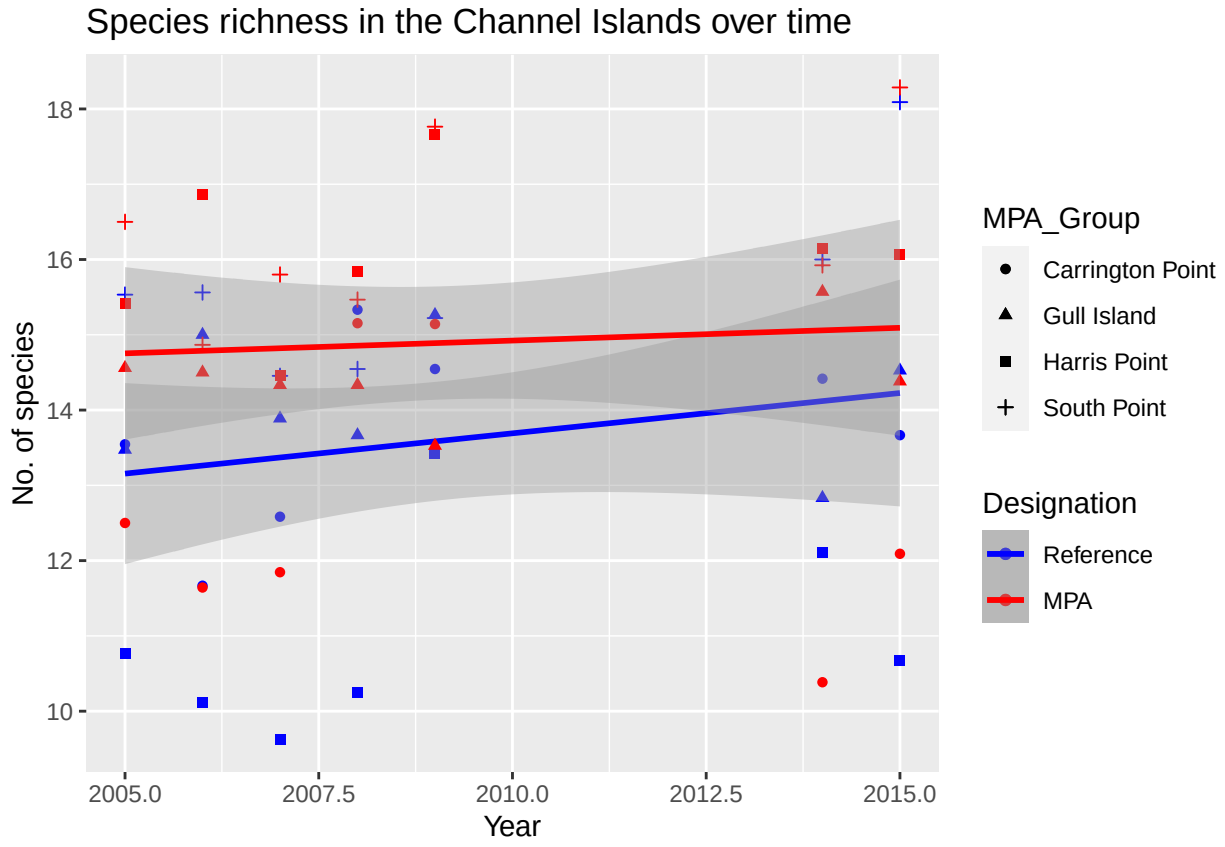



```
shapiro.test(residuals(channel.ref.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.ref.lm)
## W = 0.95382, p-value = 0.2467
```

```
# 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") +
  xlim(2005,2015)
```



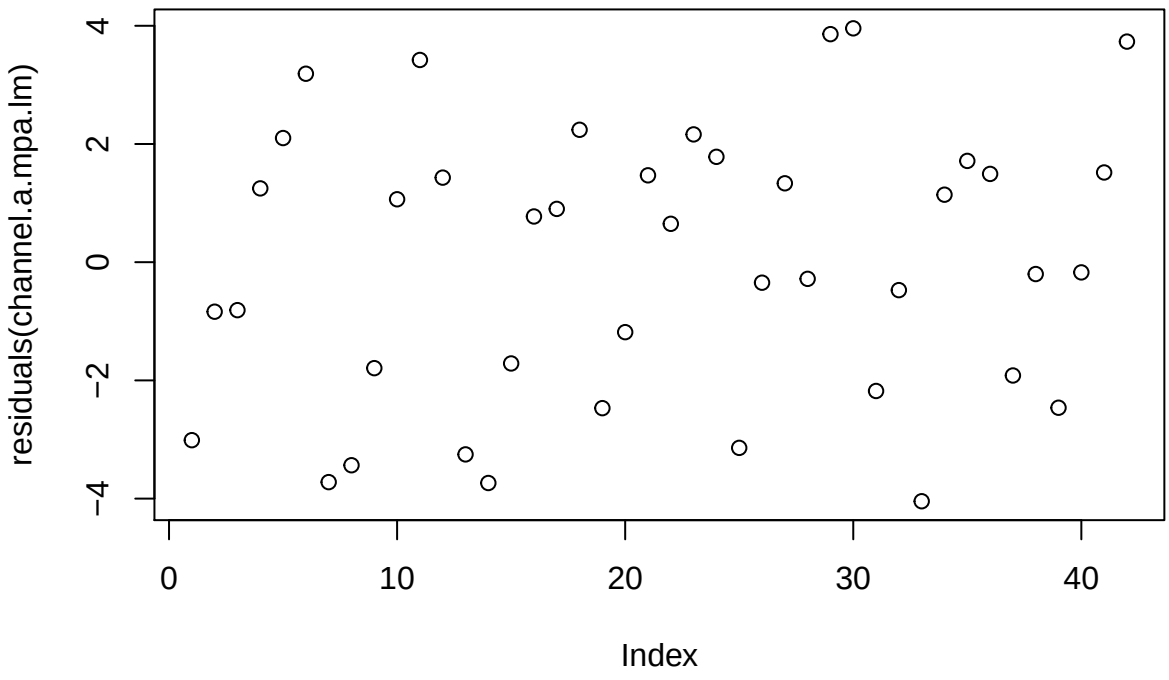
Channel Islands, with Anacapa

```
# 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")
```

```r
# Run model

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
## -4.0446 -1.8860  0.2384  1.5119  3.9565
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -235.7342   201.0714  -1.172    0.248
## Year           0.1242     0.1001   1.241    0.222
##
## Residual standard error: 2.337 on 40 degrees of freedom
## Multiple R-squared:  0.03708,    Adjusted R-squared:  0.01301
## F-statistic:  1.54 on 1 and 40 DF,  p-value: 0.2218
```

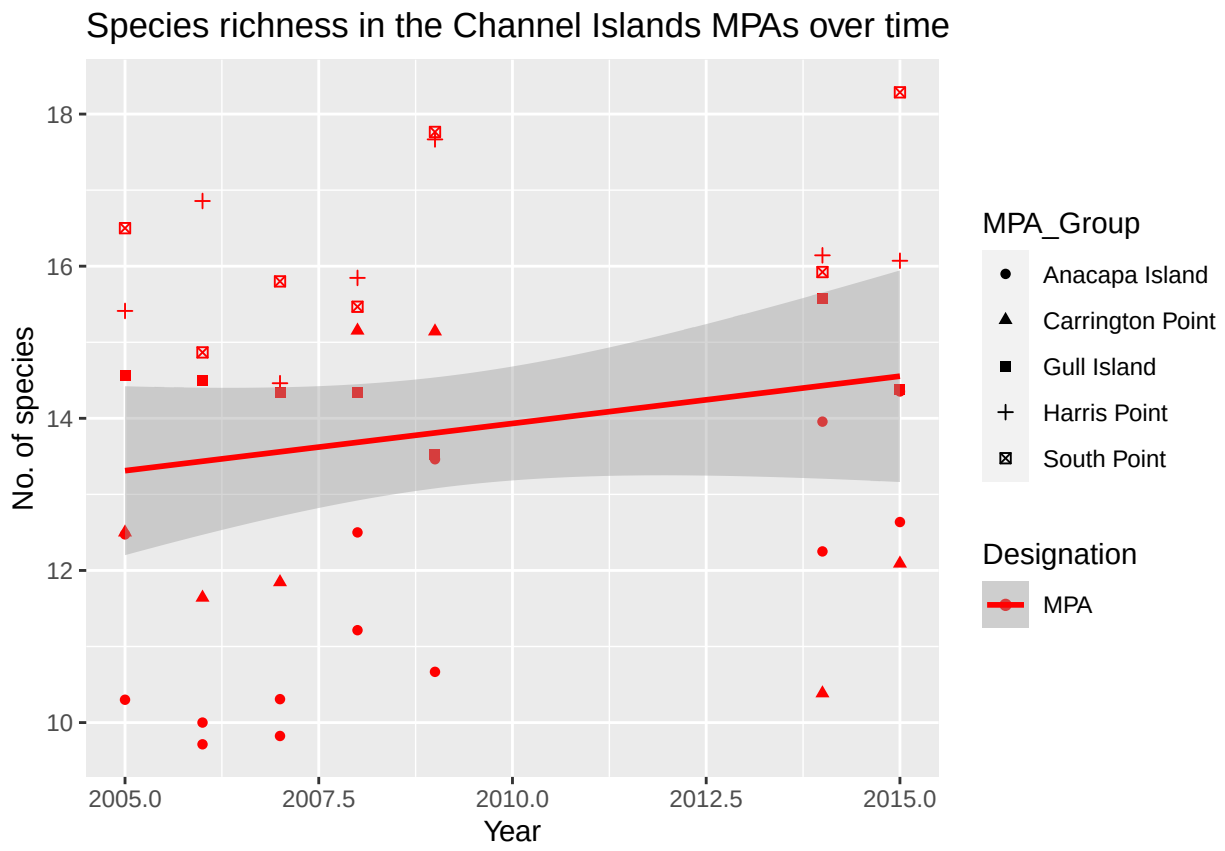
```r
# (Year: t = 1.241, p = 0.222)
# (df = 1,40, F = 1.54, p = 0.2218)
# (y~-235.7342-0.1242(Year))
# (R = 0.03708)

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




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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(channel.a.mpa.lm)
## W = 0.95548, p-value = 0.1016
```

```
# 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") +
  xlim(2005,2015)
```



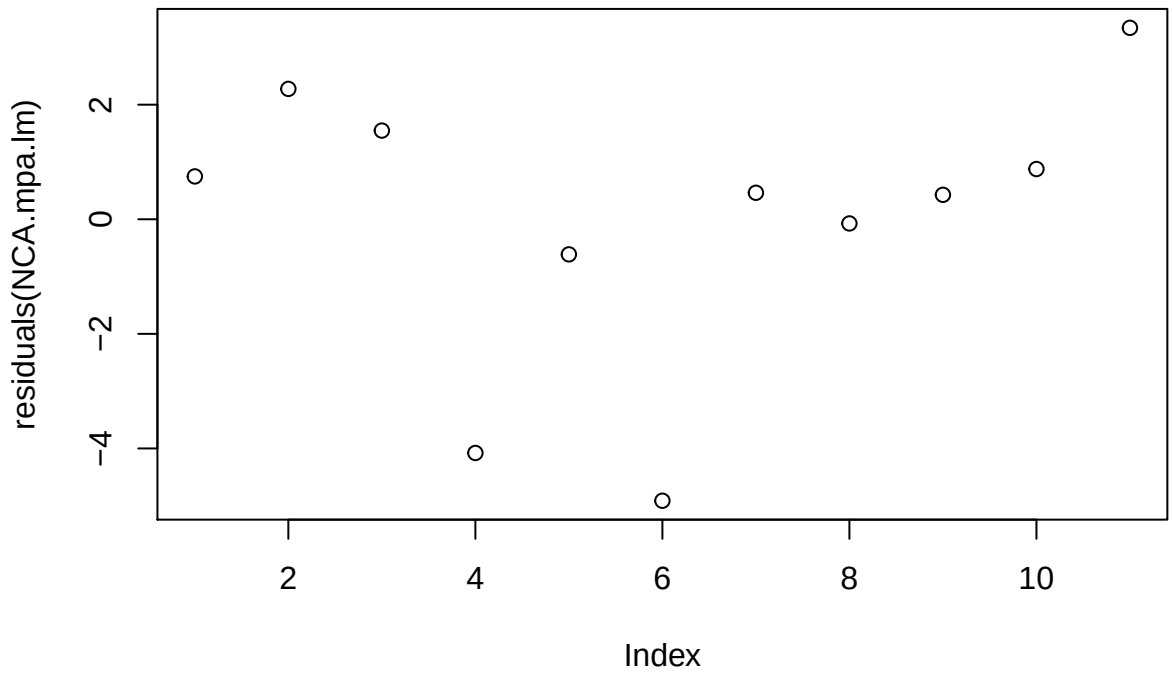
Run North CA

```
# 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

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
## -4.9127 -0.3430  0.4623  1.2123  3.3410
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -2733.8755   501.5398  -5.451 0.000405 ***
## Year            1.3652     0.2489   5.486 0.000387 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.617 on 9 degrees of freedom
## Multiple R-squared:  0.7698, Adjusted R-squared:  0.7442
## F-statistic: 30.09 on 1 and 9 DF,  p-value: 0.0003873
```

```r
# (Year: t = 5.486, p = 0.000387)
# (df = 1,9, F = 30.09, p = 0.0003873)
# (y~-2733.8755+1.3652(Year))
# (R = 0.7698)

# *** Significant ***

# Check assumptions

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



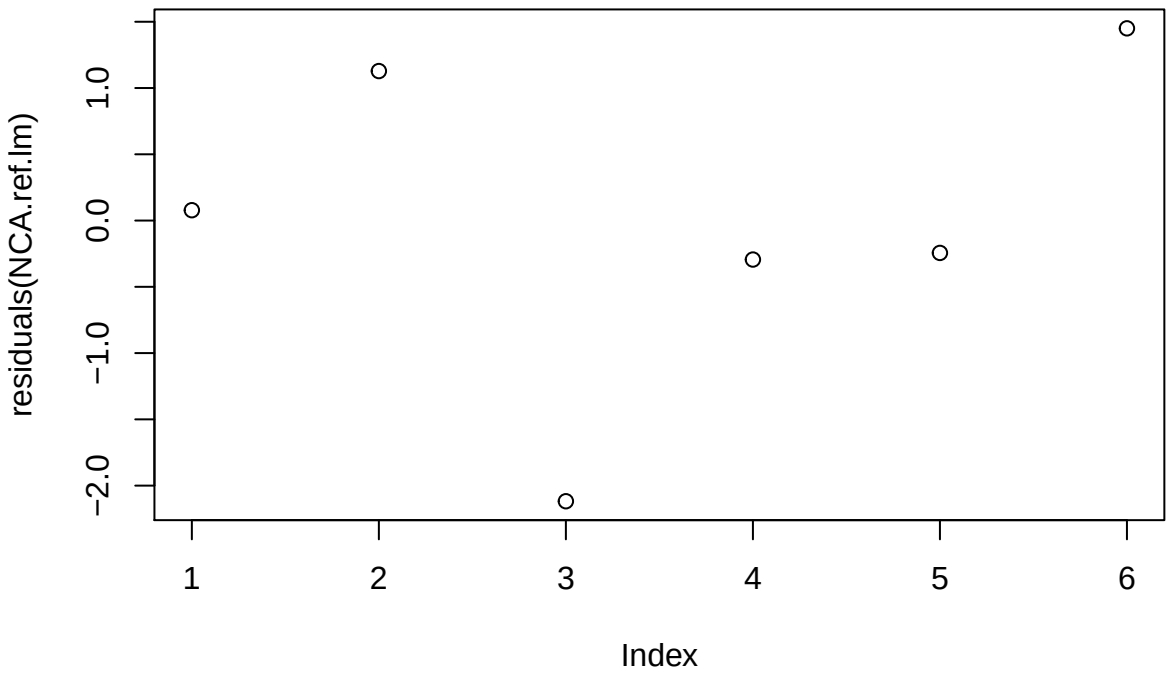

```r
shapiro.test(residuals(NCA.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.mpa.lm)
## W = 0.88291, p-value = 0.1133
```

```r
# Residuals normal

# Run Ref

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:
##      58      61     101     106     142     147
##  0.0780  1.1280 -2.1178 -0.2942 -0.2442  1.4502
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2588.2127   356.4880  -7.260  0.00191 **
## Year            1.2931     0.1769   7.309  0.00186 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.415 on 4 degrees of freedom
## Multiple R-squared:  0.9303, Adjusted R-squared:  0.9129
## F-statistic: 53.42 on 1 and 4 DF,  p-value: 0.001864
```

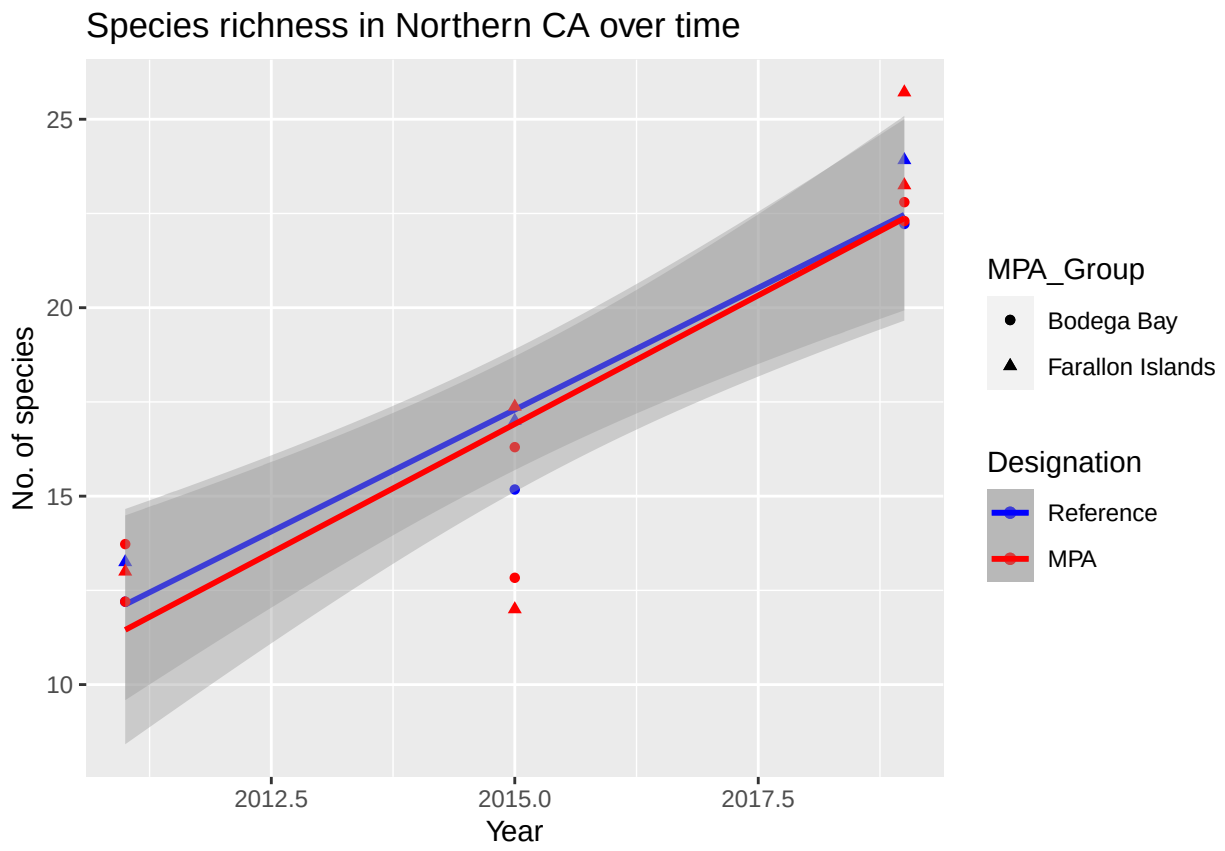
```r
# (Year: t = 7.309, p = 0.001864)
# (df = 1,4, F = 53.42, p = 0.001864)
# (y~-2588.2127+1.2931(Year))
# (R = 0.9303)

# *** Significant ****

# Check assumptions

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




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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(NCA.ref.lm)
## W = 0.92454, p-value = 0.5387
```

```
# 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") +
  xlim(2011,2019)
```



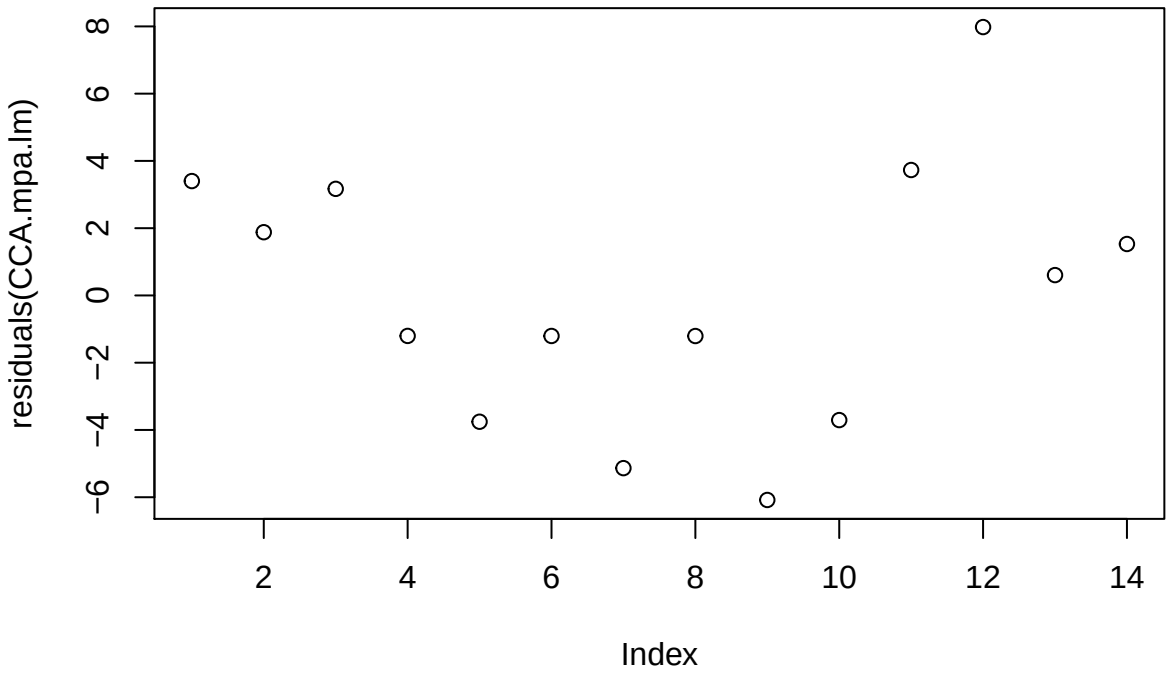
Run Central CA

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

# Subset by designation

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

# Run MPA first

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
## -6.0816 -3.0816 -0.2994  2.8467  7.9797
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -1032.0771   491.7473  -2.099   0.0577 .
## Year            0.5212     0.2442   2.135   0.0541 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.093 on 12 degrees of freedom
## Multiple R-squared:  0.2752, Adjusted R-squared:  0.2148
## F-statistic: 4.557 on 1 and 12 DF,  p-value: 0.05411
```

```
# (Year: t = 2.135, p = 0.05411)
# (df = 1,12, F = 4.557, p = 0.05411)
# (y~-1032.0771+0.5212(Year))
# (R = 0.2752)

# Test assumptions

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



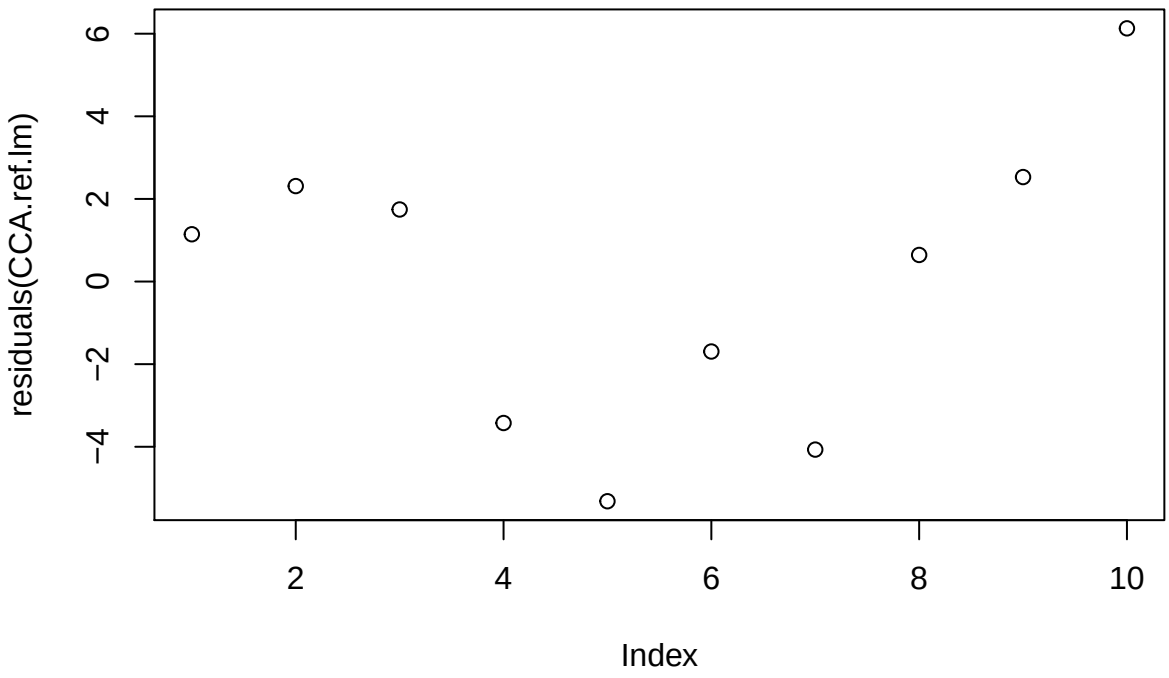

```
shapiro.test(residuals(CCA.mpa.lm))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.mpa.lm)
## W = 0.96807, p-value = 0.8498
```

```r
# Residuals normal

# Run Ref

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
## -5.3184 -2.9916  0.8944  2.1698  6.1298
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1577.5313   537.9606  -2.932   0.0189 *
## Year            0.7923     0.2671   2.966   0.0180 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.767 on 8 degrees of freedom
## Multiple R-squared:  0.5238, Adjusted R-squared:  0.4642
## F-statistic: 8.799 on 1 and 8 DF,  p-value: 0.01797
```

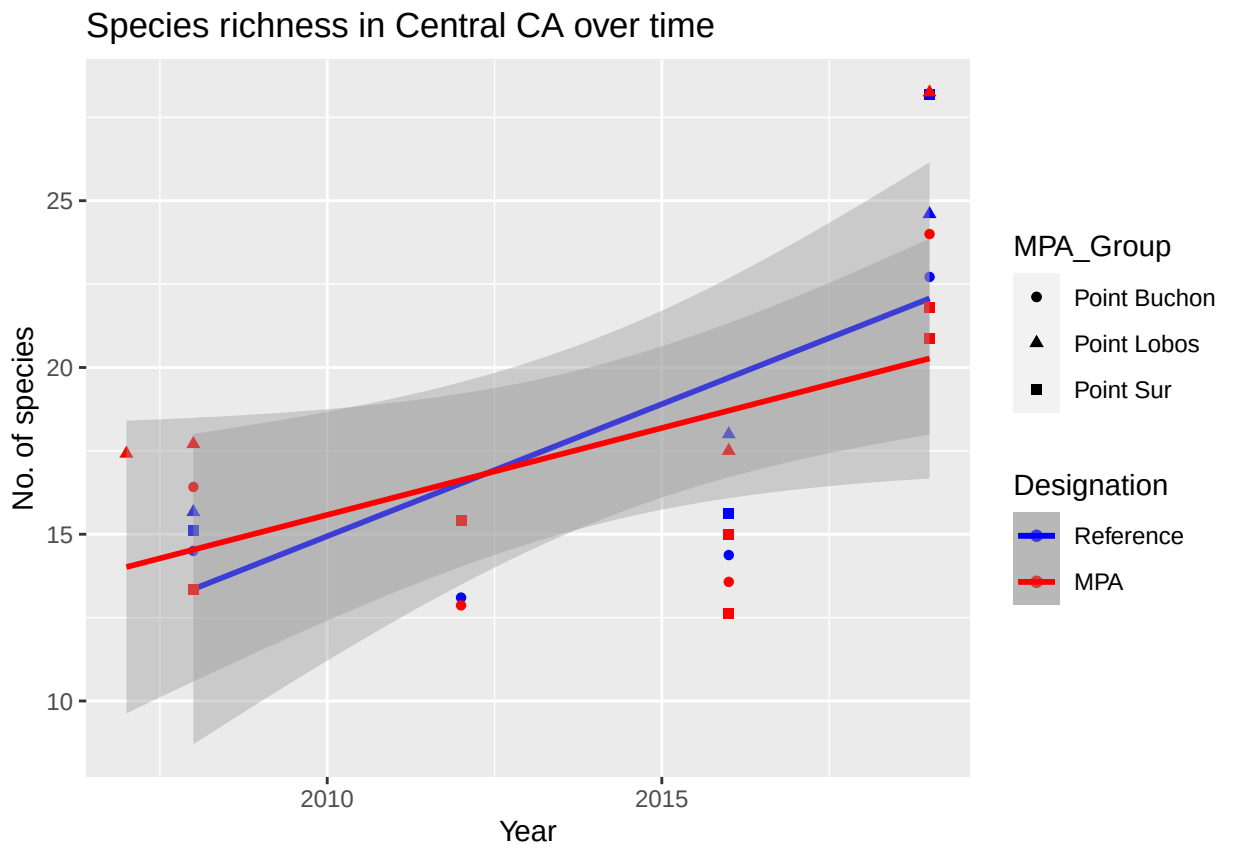
```r
# (Year: t = 2.966, p = 0.0180)
# (df = 1,8, F = 8.799, p = 0.01797)
# (y~-1577.5313+0.7923(Year))
# (R = 0.5238)

# *** Significant ***

# Test assumptions

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




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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CCA.ref.lm)
## W = 0.9547, p-value = 0.7241
```

```
# 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") +
  xlim(2007,2019)
```



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[9:166]))

# 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[1:9]
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)
mod.avg$Year <- as.factor(mod.avg$Year)
mod.avg <- subset(mod.avg, Designation == "MPA" | Designation == "Reference")

# ANCOVA
# Two-way ANOVA with no interaction

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

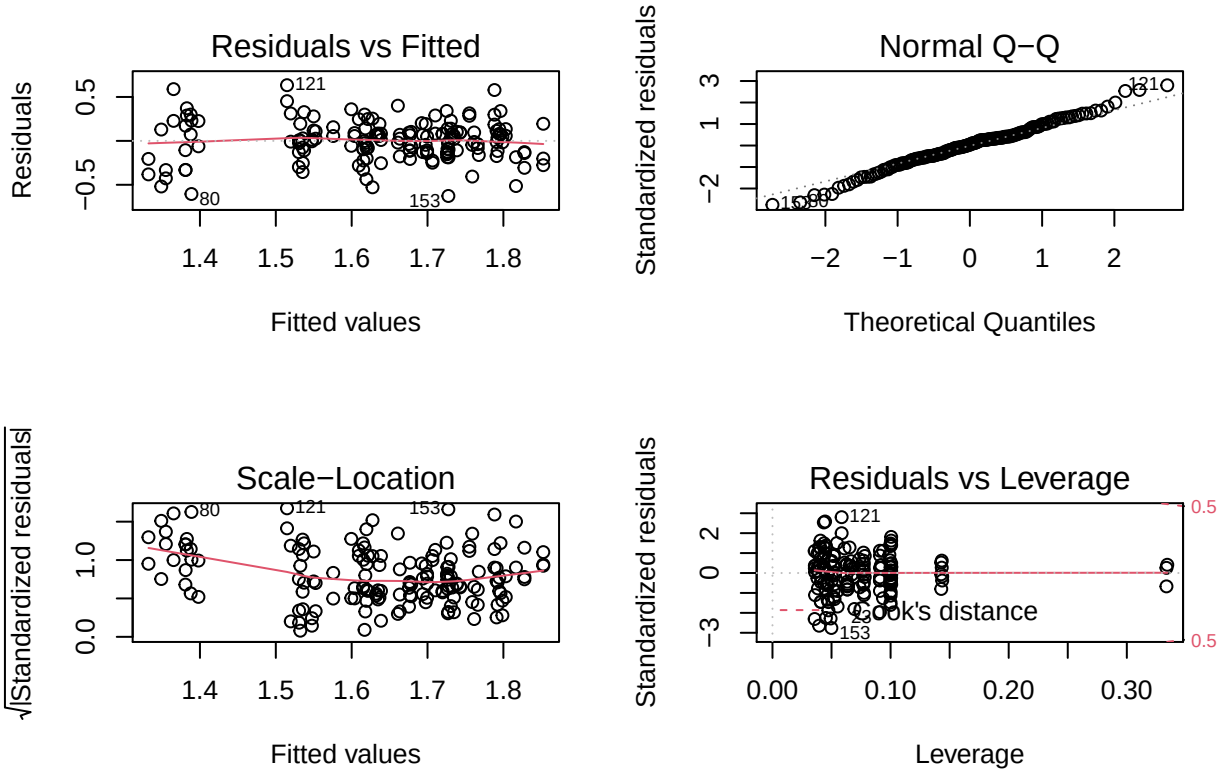
```
## Anova Table (Type III tests)
##
## Response: avg.div
##             Sum Sq  Df F value   Pr(>F)
## (Intercept) 0.0889   1  1.6322 0.203419
## Latitude    0.5325   1  9.7822 0.002125 **
## Year        2.2293  10  4.0950 5.48e-05 ***
## Residuals   8.0026 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 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
## -0.62728 -0.12779  0.01005  0.12337  0.63423
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.452534   0.354217   1.278  0.20342
## Latitude     0.031838   0.010179   3.128  0.00212 **
## Year2006     0.079773   0.104345   0.765  0.44579
## Year2007     0.084119   0.101967   0.825  0.41073
## Year2008     0.190491   0.094315   2.020  0.04523 *
## Year2009     0.141219   0.104345   1.353  0.17801
## Year2011     0.122660   0.122122   1.004  0.31683
## Year2012     0.128441   0.154292   0.832  0.40650
## Year2014    -0.154507   0.087563  -1.765  0.07972 .
## Year2015    -0.004936   0.091054  -0.054  0.95684
## Year2016     0.126609   0.100026   1.266  0.20760
```

```
## Year2019     0.184185   0.090832   2.028  0.04439 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2333 on 147 degrees of freedom
## Multiple R-squared:  0.2587, Adjusted R-squared:  0.2032
## F-statistic: 4.664 on 11 and 147 DF,  p-value: 4.282e-06
```

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

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




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

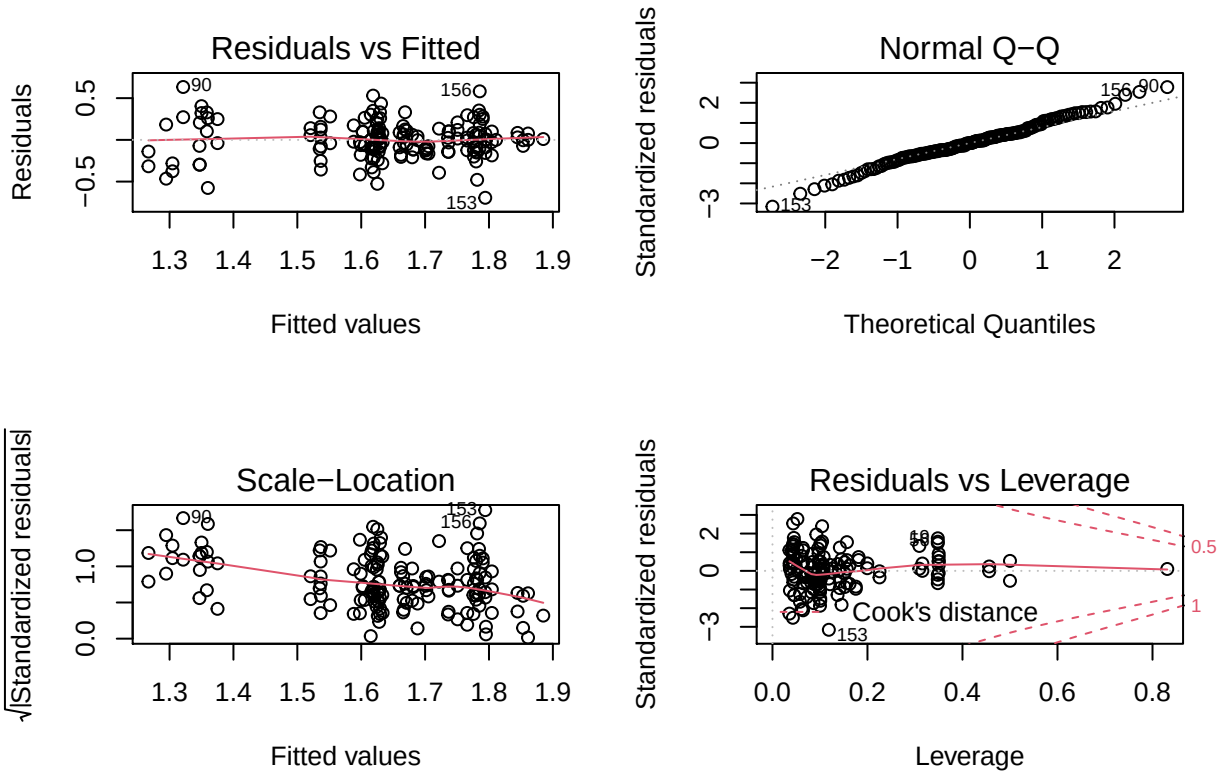
```
##
##  Shapiro-Wilk normality test
##
## data:  MPA.aov$residuals
## W = 0.98976, p-value = 0.304
```

```
# Data is approximately normal for the number of data points
# 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  0.563  0.5635  10.289 0.00167 **
## Year           10  2.229  0.2229   4.071 6.5e-05 ***
## Latitude:Year  10  0.500  0.0500   0.913 0.52315
## Residuals     137  7.503  0.0548
## ---
## 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 = 0.98899, p-value = 0.249
```

```
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
```

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

# Run post-hoc on model with no interactions

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

##
##   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  0.079773   0.104345   0.765   0.9995
## 2007 - 2005 == 0  0.084119   0.101967   0.825   0.9990
## 2008 - 2005 == 0  0.190491   0.094315   2.020   0.6149
## 2009 - 2005 == 0  0.141219   0.104345   1.353   0.9525
## 2011 - 2005 == 0  0.122660   0.122122   1.004   0.9947
## 2012 - 2005 == 0  0.128441   0.154292   0.832   0.9989
## 2014 - 2005 == 0 -0.154507   0.087563  -1.765   0.7834
## 2015 - 2005 == 0 -0.004936   0.091054  -0.054   1.0000
## 2016 - 2005 == 0  0.126609   0.100026   1.266   0.9697
## 2019 - 2005 == 0  0.184185   0.090832   2.028   0.6098
## 2007 - 2006 == 0  0.004346   0.101967   0.043   1.0000
## 2008 - 2006 == 0  0.110719   0.094315   1.174   0.9823
## 2009 - 2006 == 0  0.061447   0.104345   0.589   0.9999
## 2011 - 2006 == 0  0.042888   0.122122   0.351   1.0000
## 2012 - 2006 == 0  0.048668   0.154292   0.315   1.0000
## 2014 - 2006 == 0 -0.234279   0.087563  -2.676   0.2069
## 2015 - 2006 == 0 -0.084709   0.091054  -0.930   0.9971
## 2016 - 2006 == 0  0.046837   0.100026   0.468   1.0000
## 2019 - 2006 == 0  0.104412   0.090832   1.150   0.9849
## 2008 - 2007 == 0  0.106372   0.091518   1.162   0.9836
## 2009 - 2007 == 0  0.057100   0.101967   0.560   1.0000
```

```
## 2011 - 2007 == 0  0.038541   0.119376   0.323   1.0000
## 2012 - 2007 == 0  0.044322   0.152492   0.291   1.0000
## 2014 - 2007 == 0 -0.238626   0.084301  -2.831   0.1454
## 2015 - 2007 == 0 -0.089055   0.087603  -1.017   0.9941
## 2016 - 2007 == 0  0.042491   0.097127   0.437   1.0000
## 2019 - 2007 == 0  0.100066   0.087590   1.142   0.9855
## 2009 - 2008 == 0 -0.049272   0.094315  -0.522   1.0000
## 2011 - 2008 == 0 -0.067831   0.111110  -0.610   0.9999
## 2012 - 2008 == 0 -0.062050   0.146997  -0.422   1.0000
## 2014 - 2008 == 0 -0.344998   0.073763  -4.677    <0.01 ***
## 2015 - 2008 == 0 -0.195427   0.076667  -2.549   0.2678
## 2016 - 2008 == 0 -0.063882   0.087990  -0.726   0.9997
## 2019 - 2008 == 0 -0.006307   0.077230  -0.082   1.0000
## 2011 - 2009 == 0 -0.018559   0.122122  -0.152   1.0000
## 2012 - 2009 == 0 -0.012778   0.154292  -0.083   1.0000
## 2014 - 2009 == 0 -0.295726   0.087563  -3.377   0.0332 *
## 2015 - 2009 == 0 -0.146155   0.091054  -1.605   0.8668
## 2016 - 2009 == 0 -0.014610   0.100026  -0.146   1.0000
## 2019 - 2009 == 0  0.042965   0.090832   0.473   1.0000
## 2012 - 2011 == 0  0.005781   0.163167   0.035   1.0000
## 2014 - 2011 == 0 -0.277167   0.101580  -2.729   0.1842
## 2015 - 2011 == 0 -0.127596   0.099219  -1.286   0.9663
## 2016 - 2011 == 0  0.003949   0.111537   0.035   1.0000
## 2019 - 2011 == 0  0.061525   0.102719   0.599   0.9999
## 2014 - 2012 == 0 -0.282948   0.141756  -1.996   0.6320
## 2015 - 2012 == 0 -0.133377   0.142571  -0.936   0.9970
## 2016 - 2012 == 0 -0.001832   0.149518  -0.012   1.0000
## 2019 - 2012 == 0  0.055744   0.143362   0.389   1.0000
## 2015 - 2014 == 0  0.149571   0.063769   2.346   0.3887
## 2016 - 2014 == 0  0.281116   0.078351   3.588   0.0177 *
## 2019 - 2014 == 0  0.338691   0.065798   5.147    <0.01 ***
## 2016 - 2015 == 0  0.131545   0.079036   1.664   0.8384
## 2019 - 2015 == 0  0.189121   0.066284   2.853   0.1385
## 2019 - 2016 == 0  0.057575   0.080977   0.711   0.9997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
# The only year that's different is 2014 from 2008, 2009, 2016 & 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 = 9.7822, p = 0.002125) and year
# (df = 10,147, F = 4.0950, p = 5.48e-05).

# *** Significant ***

# Linear regression between diversity and latitude

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

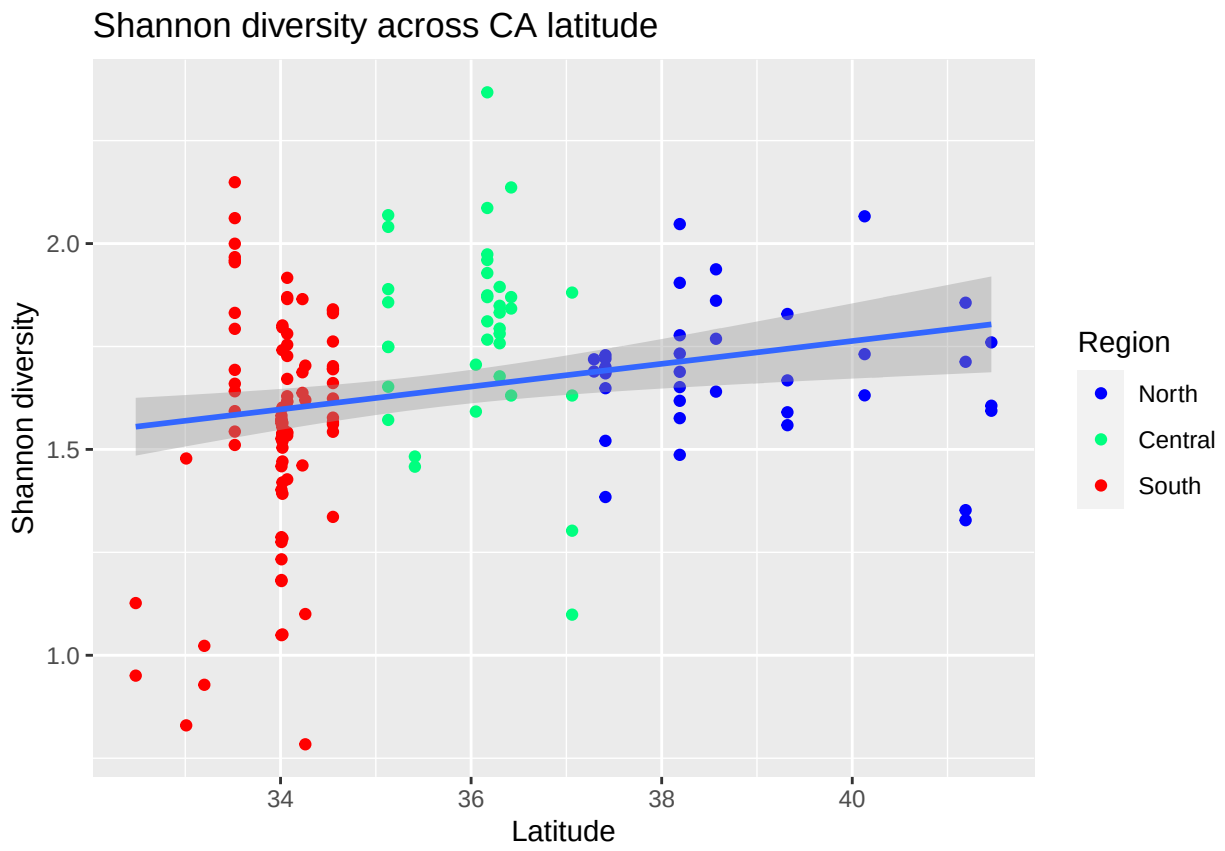
```
##
## Call:
## lm(formula = avg.div ~ Latitude, data = mod.avg)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82066 -0.09273  0.00906  0.16326  0.71001
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.656167   0.335577   1.955  0.05232 .
## Latitude    0.027678   0.009413   2.940  0.00377 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2553 on 157 degrees of freedom
## Multiple R-squared:  0.05219,    Adjusted R-squared:  0.04616
## F-statistic: 8.646 on 1 and 157 DF,  p-value: 0.003774
```

```
# (Year: t = 2.940, p = 0.00377)
# (df = 1,157, F = 8.646, p = 0.003774)
# (y~0.656167+0.027678(Latitude))
# (R = 0.05219)

# *** Significant ***

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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lat.lm)
## W = 0.97011, p-value = 0.001597
```

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



Repeat above, but using richness instead of diversity

```
# Add richness metric

data <- mutate(data, richness = rowSums(data[9:165]!=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
```

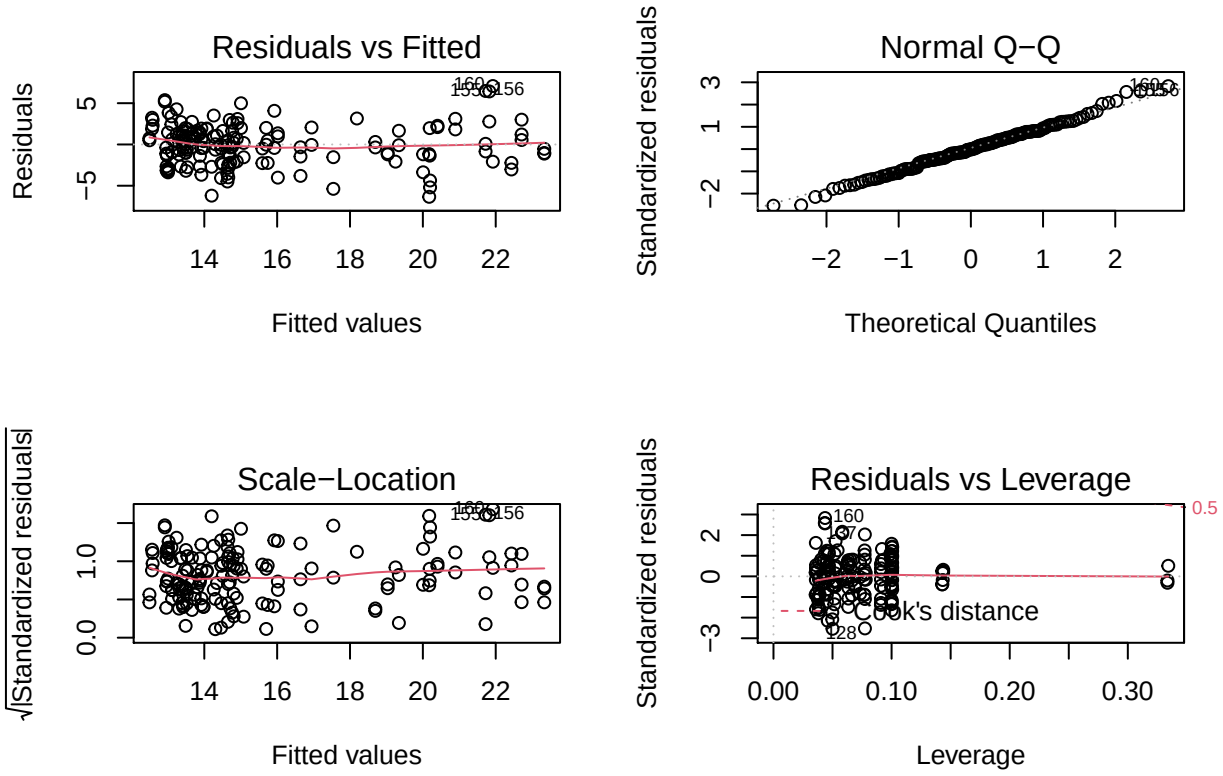
```r
p1 <- rich.avgs[1:9]
p2 <- rich.avgs[167:171]
mod.avg <- cbind(p1,p2)
mod.avg <- mod.avg[!duplicated(mod.avg[12]),]
mod.avg$Year <- as.factor(mod.avg$Year)

# ANCOVA
# Two-way ANOVA with no interaction

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)  81.76   1  12.481 0.0005486 ***
## Latitude    338.66   1  51.703 2.968e-11 ***
## Year        992.36  10  15.150 < 2.2e-16 ***
## Residuals   969.43 148
## ---
## 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
## -6.3407 -1.4012 -0.0375  1.6530  7.0790
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.6420     3.8614  -3.533 0.000549 ***
## Latitude      0.7977     0.1109   7.190 2.97e-11 ***
## Year2006     -0.5146     1.1446  -0.450 0.653660
## Year2007     -0.5310     1.1185  -0.475 0.635680
## Year2008      0.3846     1.0345   0.372 0.710605
## Year2009      1.1621     1.1446   1.015 0.311612
## Year2011     -3.7054     1.3386  -2.768 0.006358 **
## Year2012     -0.8631     1.6923  -0.510 0.610819
## Year2014      0.9720     0.9542   1.019 0.310027
## Year2015     -0.1809     0.9981  -0.181 0.856418
## Year2016     -0.4026     1.0969  -0.367 0.714098
## Year2019      6.5109     0.9960   6.537 9.53e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.559 on 148 degrees of freedom
## Multiple R-squared:  0.6011, Adjusted R-squared:  0.5715
## F-statistic: 20.28 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)
```

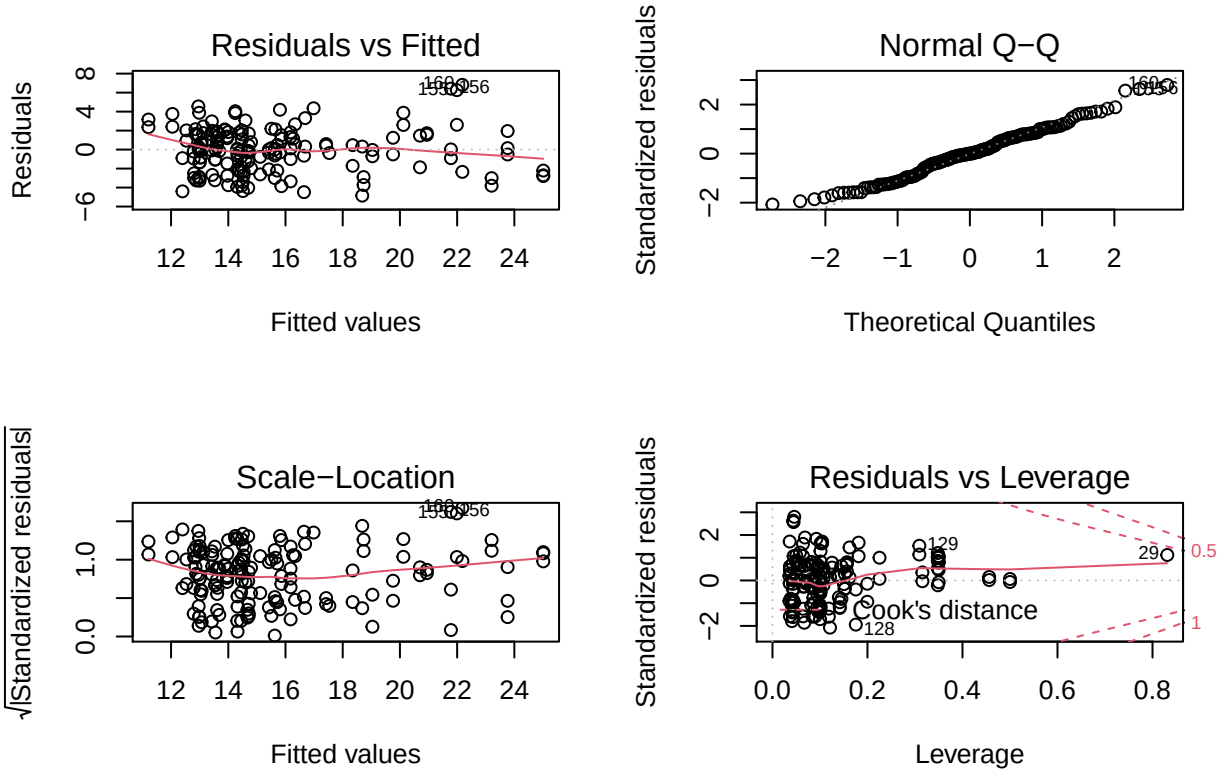
```
##
##  Shapiro-Wilk normality test
##
## data:  MPA.aov$residuals
## W = 0.99425, p-value = 0.7851
```

```
# 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  468.7   468.7  75.423 9.67e-15 ***
## Year           10  992.4    99.2  15.971  < 2e-16 ***
## Latitude:Year  10  111.9    11.2   1.802   0.0657 .
## Residuals     138  857.5     6.2
## ---
## 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 = 0.98478, p-value = 0.07697
```

```
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

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

##
##   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 -0.51459    1.14457  -0.450   1.0000
## 2007 - 2005 == 0 -0.53098    1.11849  -0.475   1.0000
## 2008 - 2005 == 0  0.38459    1.03451   0.372   1.0000
## 2009 - 2005 == 0  1.16210    1.14457   1.015   0.9941
## 2011 - 2005 == 0 -3.70544    1.33858  -2.768   0.1681
## 2012 - 2005 == 0 -0.86308    1.69234  -0.510   1.0000
## 2014 - 2005 == 0  0.97196    0.95416   1.019   0.9940
## 2015 - 2005 == 0 -0.18090    0.99808  -0.181   1.0000
## 2016 - 2005 == 0 -0.40264    1.09693  -0.367   1.0000
## 2019 - 2005 == 0  6.51094    0.99600   6.537    <0.01 ***
## 2007 - 2006 == 0 -0.01639    1.11849  -0.015   1.0000
## 2008 - 2006 == 0  0.89918    1.03451   0.869   0.9984
## 2009 - 2006 == 0  1.67669    1.14457   1.465   0.9213
## 2011 - 2006 == 0 -3.19085    1.33858  -2.384   0.3641
## 2012 - 2006 == 0 -0.34848    1.69234  -0.206   1.0000
## 2014 - 2006 == 0  1.48656    0.95416   1.558   0.8870
## 2015 - 2006 == 0  0.33369    0.99808   0.334   1.0000
## 2016 - 2006 == 0  0.11195    1.09693   0.102   1.0000
## 2019 - 2006 == 0  7.02554    0.99600   7.054    <0.01 ***
## 2008 - 2007 == 0  0.91557    1.00385   0.912   0.9975
## 2009 - 2007 == 0  1.69308    1.11849   1.514   0.9042
## 2011 - 2007 == 0 -3.17446    1.30854  -2.426   0.3380
## 2012 - 2007 == 0 -0.33210    1.67263  -0.199   1.0000
## 2014 - 2007 == 0  1.50294    0.91845   1.636   0.8522
## 2015 - 2007 == 0  0.35007    0.96029   0.365   1.0000
## 2016 - 2007 == 0  0.12834    1.06518   0.120   1.0000
## 2019 - 2007 == 0  7.04192    0.96049   7.332    <0.01 ***
## 2009 - 2008 == 0  0.77751    1.03451   0.752   0.9995
## 2011 - 2008 == 0 -4.09003    1.21804  -3.358   0.0349 *
## 2012 - 2008 == 0 -1.24766    1.61239  -0.774   0.9994
## 2014 - 2008 == 0  0.58738    0.80274   0.732   0.9996
## 2015 - 2008 == 0 -0.56549    0.84048  -0.673   0.9998
```

```
## 2016 - 2008 == 0 -0.78723    0.96505  -0.816   0.9990
## 2019 - 2008 == 0  6.12636    0.84697   7.233    <0.01 ***
## 2011 - 2009 == 0 -4.86754    1.33858  -3.636   0.0144 *
## 2012 - 2009 == 0 -2.02518    1.69234  -1.197   0.9797
## 2014 - 2009 == 0 -0.19014    0.95416  -0.199   1.0000
## 2015 - 2009 == 0 -1.34300    0.99808  -1.346   0.9541
## 2016 - 2009 == 0 -1.56474    1.09693  -1.426   0.9333
## 2019 - 2009 == 0  5.34884    0.99600   5.370    <0.01 ***
## 2012 - 2011 == 0  2.84236    1.78949   1.588   0.8738
## 2014 - 2011 == 0  4.67741    1.11269   4.204    <0.01 **
## 2015 - 2011 == 0  3.52454    1.08826   3.239   0.0489 *
## 2016 - 2011 == 0  3.30280    1.22315   2.700   0.1948
## 2019 - 2011 == 0 10.21639    1.12645   9.070    <0.01 ***
## 2014 - 2012 == 0  1.83504    1.55223   1.182   0.9814
## 2015 - 2012 == 0  0.68217    1.56376   0.436   1.0000
## 2016 - 2012 == 0  0.46044    1.64007   0.281   1.0000
## 2019 - 2012 == 0  7.37402    1.57254   4.689    <0.01 ***
## 2015 - 2014 == 0 -1.15287    0.69570  -1.657   0.8418
## 2016 - 2014 == 0 -1.37461    0.85512  -1.607   0.8656
## 2019 - 2014 == 0  5.53898    0.71684   7.727    <0.01 ***
## 2016 - 2015 == 0 -0.22174    0.86685  -0.256   1.0000
## 2019 - 2015 == 0  6.69185    0.72699   9.205    <0.01 ***
## 2019 - 2016 == 0  6.91359    0.88824   7.783    <0.01 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
# 2019 is different from 2005, 2006, 2007, 2008, 2009, 2012, 2014, 2015 & 2016.
# 2014 is different from 2011 (p < 0.05).

# There was a significant difference found between (average) richness
# for different latitudes (df = 1,114, F = 74.086, p = 4.73e-14) and year
# (df = 10,956.1, F = 14.402, p = 3.22e-16).

# *** Significant ***

# Linear regression between richness and latitude

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

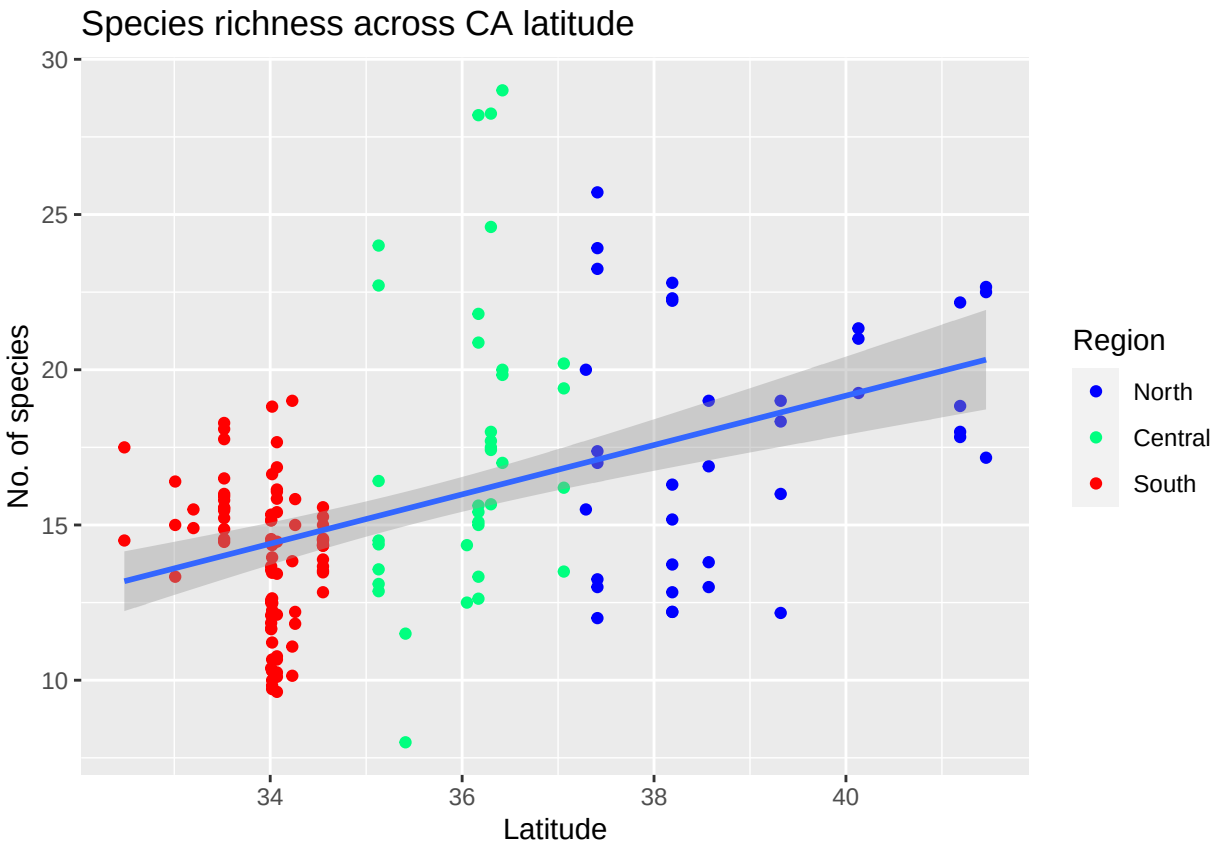
```
##
## Call:
## lm(formula = avg.rich ~ Latitude, data = mod.avg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5167 -2.3562 -0.3201  1.7014 12.6807
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12.6236     4.6093  -2.739  0.00688 **
## Latitude      0.7947     0.1294   6.144  6.3e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.524 on 158 degrees of freedom
## Multiple R-squared:  0.1928, Adjusted R-squared:  0.1877
## F-statistic: 37.74 on 1 and 158 DF,  p-value: 6.296e-09
```

```
# (Year: t = 6.144, p = 6.3e-09)
# (df = 1,158, F = 37.74, p = 6.296e-09)
# (y~-12.6237+0.7947(Latitude))
# (R = 0.1928)

# *** Significant ***

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

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lat.lm)
## W = 0.94872, p-value = 1.4e-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("#0000ff","#00ff7f","#ff0000"),
                     breaks = c("North","Central","South")) +
  ggtitle("Species richness across CA latitude") +
  xlab("Latitude") +
  ylab("No. of species")
```


Species richness across CA latitude
No. of species
30
25
20
15
10
34
36
38
40
Latitude
Region
North
Central
South