# Questions of Interest

We are interested in if the State Highway 1 (SH1) has any demographic impacts for communities located in the Whangārei District. Specifically, we want to know if incomes and population counts are affected by the presents of SH1.

## Total Income

### Read in and inspect the data:

```
income_tot.df = read.csv("Data/income_tot.csv", header = TRUE, stringsAsFactors = TRUE, na.strings = "NU
income_tot.df$Year = as.factor(income_tot.df$Year)    # converting int to Factor
income_tot.df = income_tot.df[4:21, 1:4]
income_tot.df$Area = factor(income_tot.df$Area, level = c("Raumanga East", "Raumanga West", "Raumanga",
str(income_tot.df)
```

```
## 'data.frame':    18 obs. of  4 variables:
##  $ Area  : Factor w/ 8 levels "Raumanga East",..: 1 2 3 3 4 4 5 5 5 6 ...
##  $ Year  : Factor w/ 3 levels "2001","2013",..: 1 1 2 3 2 3 1 2 3 1 ...
##  $ Income: int  1179 1962 2031 2388 1164 1413 48 33 93 651 ...
##  $ SH1   : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
```

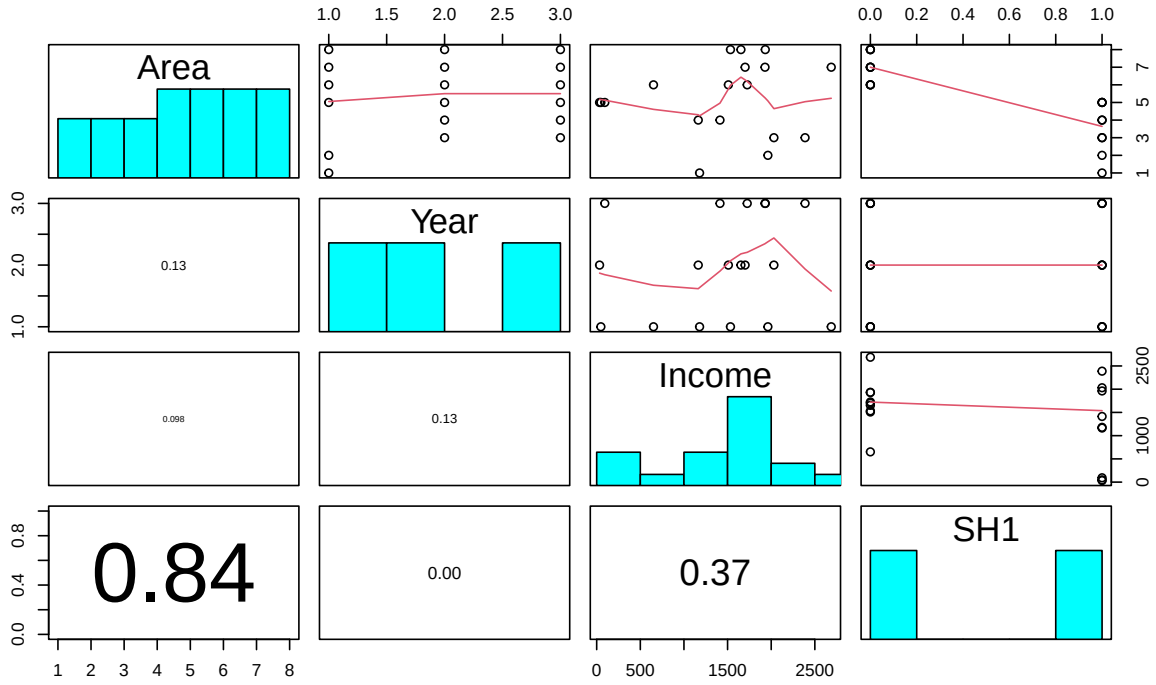
```
summary(income_tot.df)
```

```
##                Area     Year       Income        SH1
##  Port-Limeburners:3   2001:6   Min.   :  33   Mode :logical
##  Riverside       :3   2013:6   1st Qu.:1168   FALSE:9
##  Sherwood Rise   :3   2018:6   Median :1593   TRUE :9
##  Onerahi         :3            Mean   :1424
##  Raumanga        :2            3rd Qu.:1931
##  Tarewa          :2            Max.   :2688
##  (Other)         :2
```

```
levels(income_tot.df$Area)
```

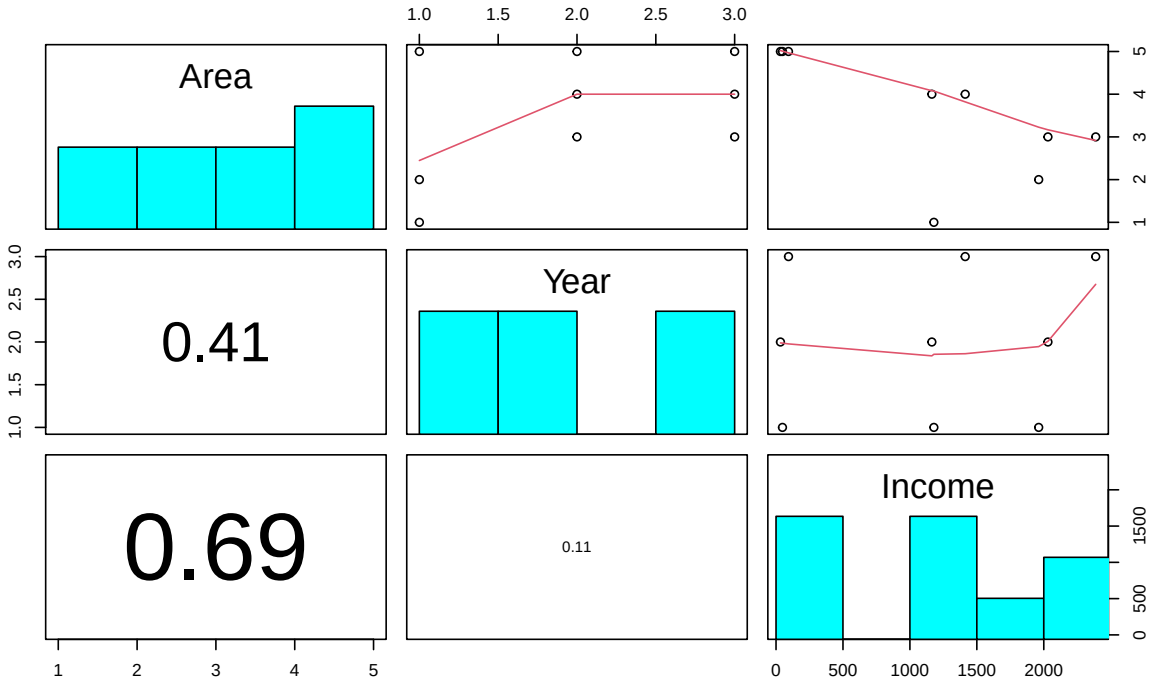
```
## [1] "Raumanga East"    "Raumanga West"    "Raumanga"         "Tarewa"
## [5] "Port-Limeburners" "Riverside"        "Sherwood Rise"    "Onerahi"
```

```
#incomes.tbl = with(income_tot.df, table(Intersect_SH1, Area))
#incomes.tbl
pairs20x(income_tot.df)
```

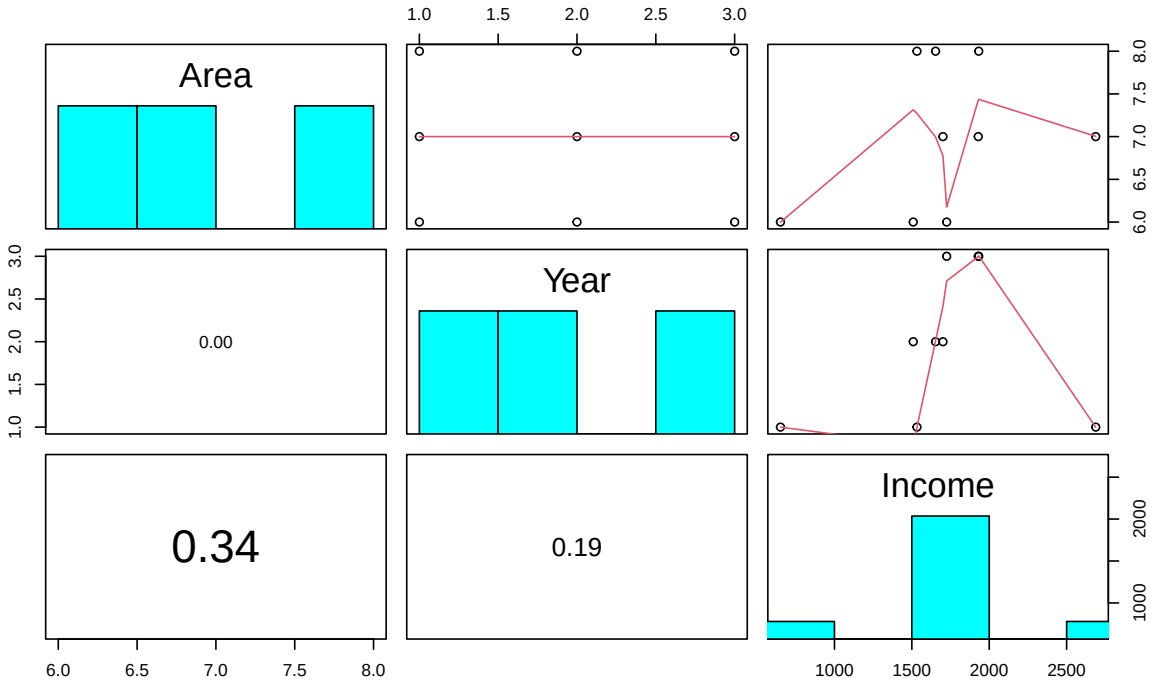



**Comment on the pairs plot**

```
income_tot_sh1.df = income_tot.df[which(income_tot.df$SH1 == TRUE), names(income_tot.df) %in% c("Area",
pairs20x(income_tot_sh1.df)
```




```
income_tot_nsh1.df = income_tot.df[which(income_tot.df$SH1 == FALSE), names(income_tot.df) %in% c("Area
pairs20x(income_tot_nsh1.df)
```



### Fit and check model

```
#library(dplyr)
#incomes.grouped.df = incomes.df %>% group_by("SH1", "Area", "Year") %>% summarize(TRUE=)    # https://

income_tot.fit = glm(Income ~ SH1 * Area, family = poisson, data = income_tot.df)
#plot(income_tot.fit, which = 1)    # identically distributed
#plot(income_tot.fit, which = 2)    # normality assumption
summary(income_tot.fit)
```

```
##
## Call:
## glm(formula = Income ~ SH1 * Area, family = poisson, data = income_tot.df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -19.813   -3.781   -0.645    4.103   12.156
##
## Coefficients: (8 not defined because of singularities)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    7.44191    0.01398 532.395   <2e-16 ***
## SH1TRUE                       -0.36948    0.03230 -11.438   <2e-16 ***
## AreaRaumanga West              0.50930    0.03685  13.821   <2e-16 ***
## AreaRaumanga                   0.62810    0.03278  19.162   <2e-16 ***
## AreaTarewa                     0.08881    0.03516   2.526   0.0115 *
## AreaPort-Limeburners          -3.01198    0.08121 -37.088   <2e-16 ***
## AreaRiverside                 -0.27564    0.02128 -12.954   <2e-16 ***
## AreaSherwood Rise              0.21064    0.01881  11.201   <2e-16 ***
## AreaOnerahi                         NA         NA      NA       NA
## SH1TRUE:AreaRaumanga West           NA         NA      NA       NA
## SH1TRUE:AreaRaumanga                NA         NA      NA       NA
## SH1TRUE:AreaTarewa                  NA         NA      NA       NA
## SH1TRUE:AreaPort-Limeburners        NA         NA      NA       NA
## SH1TRUE:AreaRiverside               NA         NA      NA       NA
## SH1TRUE:AreaSherwood Rise           NA         NA      NA       NA
## SH1TRUE:AreaOnerahi                 NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 10057.77  on 17  degrees of freedom
## Residual deviance:   935.72  on 10  degrees of freedom
## AIC: 1107.5
##
## Number of Fisher Scoring iterations: 4
```

```
1 - pchisq(109.96, 4)    # residual deviance
```

```
## [1] 0
```

```
#library(lme4)    # https://uoftcoders.github.io/rcourse/lec08-linear-mixed-effects-models.html
library(lmerTest)    # https://rcompanion.org/handbook/G_03.html
```

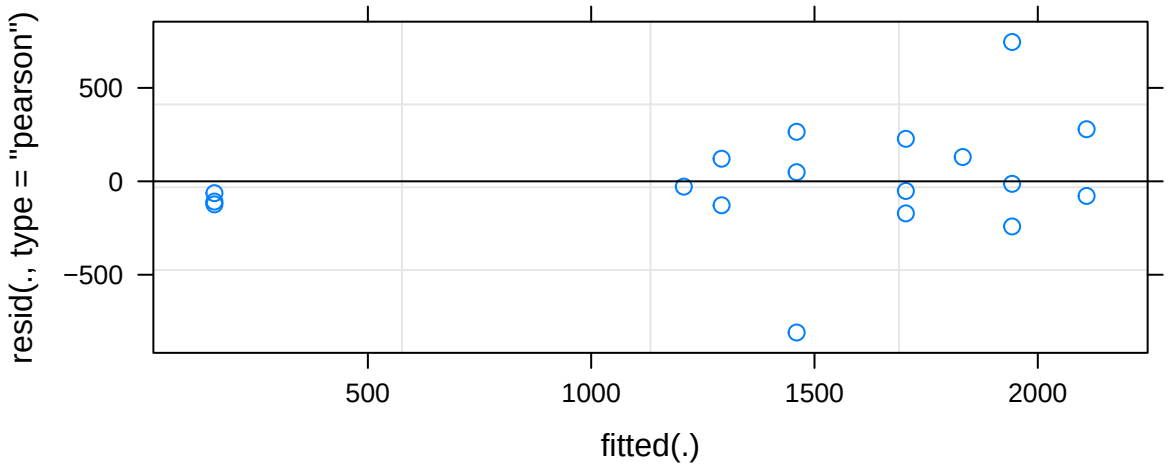
```
## Loading required package: lme4
```

```
## Loading required package: Matrix
##
## Attaching package: 'lmerTest'
## The following object is masked from 'package:lme4':
##
##     lmer
## The following object is masked from 'package:stats':
##
##     step
```

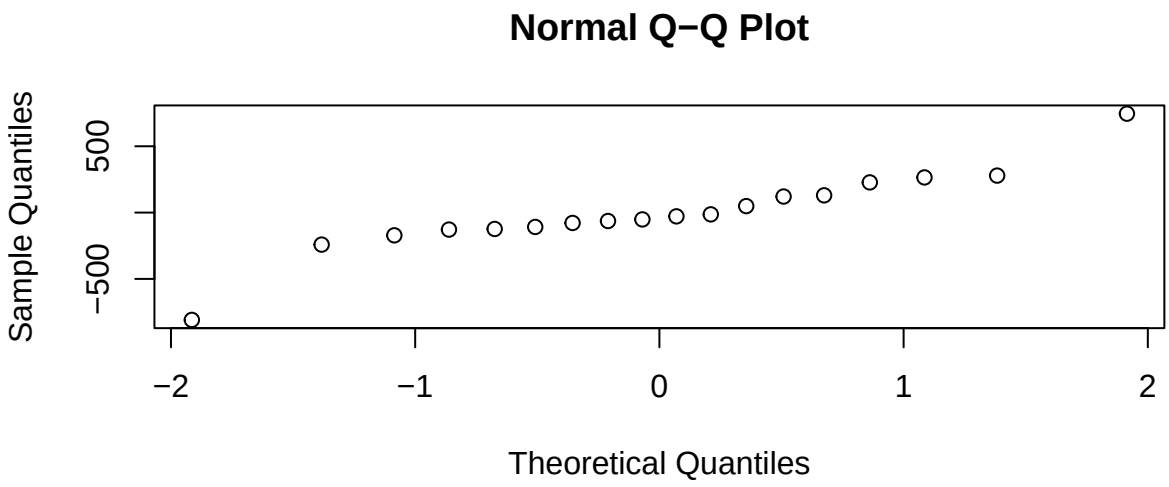
```
income_tot.lmer = lmer(Income ~ SH1 + (1 + SH1 | Area) + (1 + SH1 | Year), data = income_tot.df, REML =
```

```
## boundary (singular) fit: see ?isSingular
```

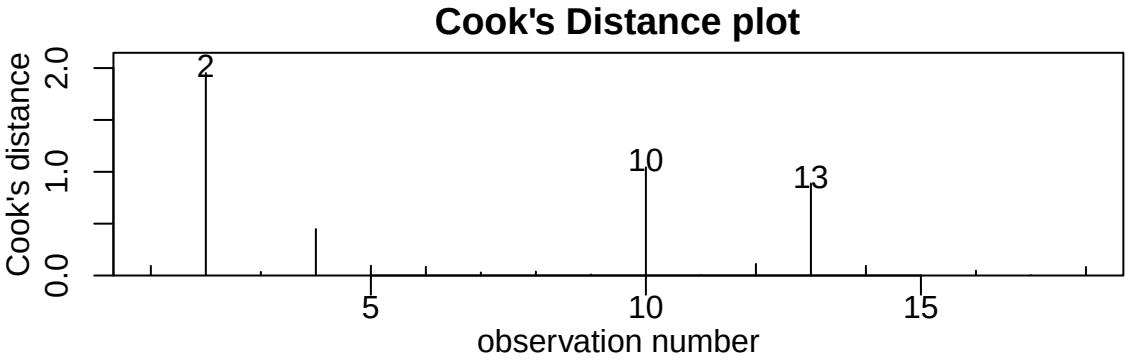
```
plot(income_tot.lmer)    # https://stats.stackexchange.com/questions/376273/assumptions-for-lmer-models
```




```
qqnorm(resid(income_tot.lmer))    # https://stats.stackexchange.com/questions/77891/checking-assumption
```




```
cooks20x(income_tot.lmer)
```




```
summary(income_tot.lmer)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: Income ~ SH1 + (1 + SH1 | Area) + (1 + SH1 | Year)
##    Data: income_tot.df
##
##      AIC      BIC   logLik deviance df.resid
##    294.3    302.3   -138.2    276.3        9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2163 -0.3279 -0.1095  0.3502  2.0424
##
## Random effects:
```

```
##  Groups   Name        Variance  Std.Dev.  Corr
##  Area     (Intercept) 6.521e+04 2.554e+02
##           SH1TRUE     4.079e+05 6.387e+02 0.16
##  Year     (Intercept) 0.000e+00 0.000e+00
##           SH1TRUE     1.469e-06 1.212e-03  NaN
##  Residual             1.333e+05 3.650e+02
## Number of obs: 18, groups:  Area, 8; Year, 3
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 1702.333    191.160    3.000   8.905  0.00299 **
## SH1TRUE     -382.947    398.529    7.814  -0.961  0.36539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## SH1TRUE -0.480
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
#isSingular(incomes.lmer, tol = 1e-4)
#VarCorr(incomes.lmer)    # https://rstudio-pubs-static.s3.amazonaws.com/63556_e35cc7e2dfb54a5bb551f3fa
```

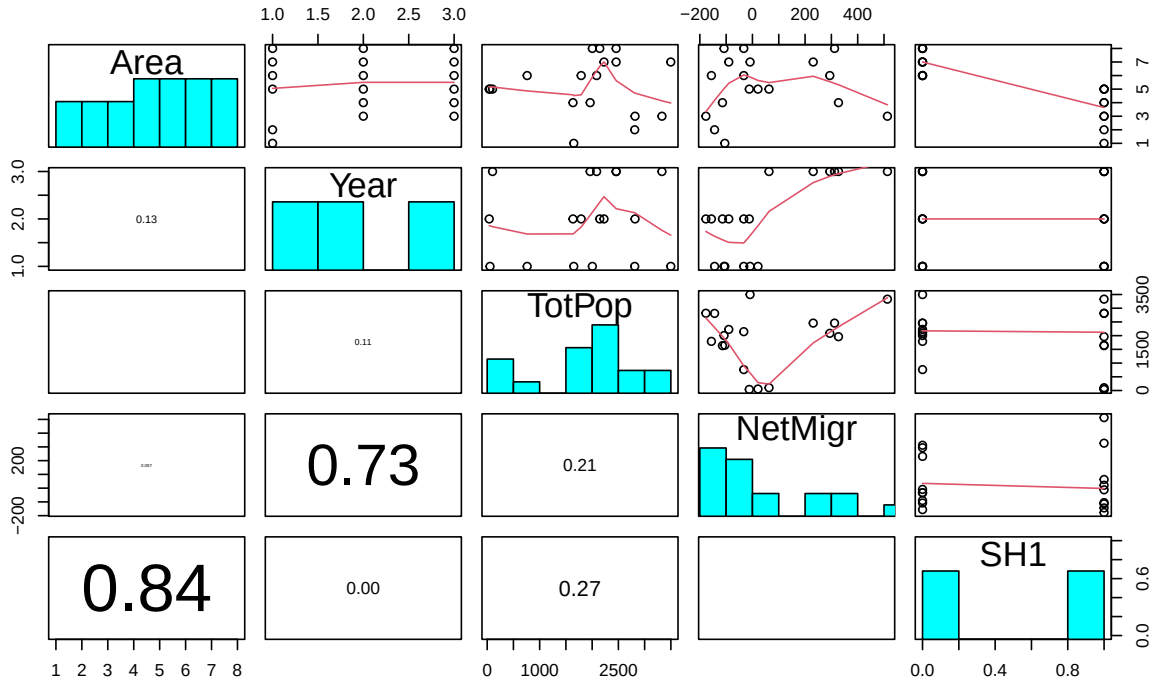
## Populations

**Read in and inspect the data:**

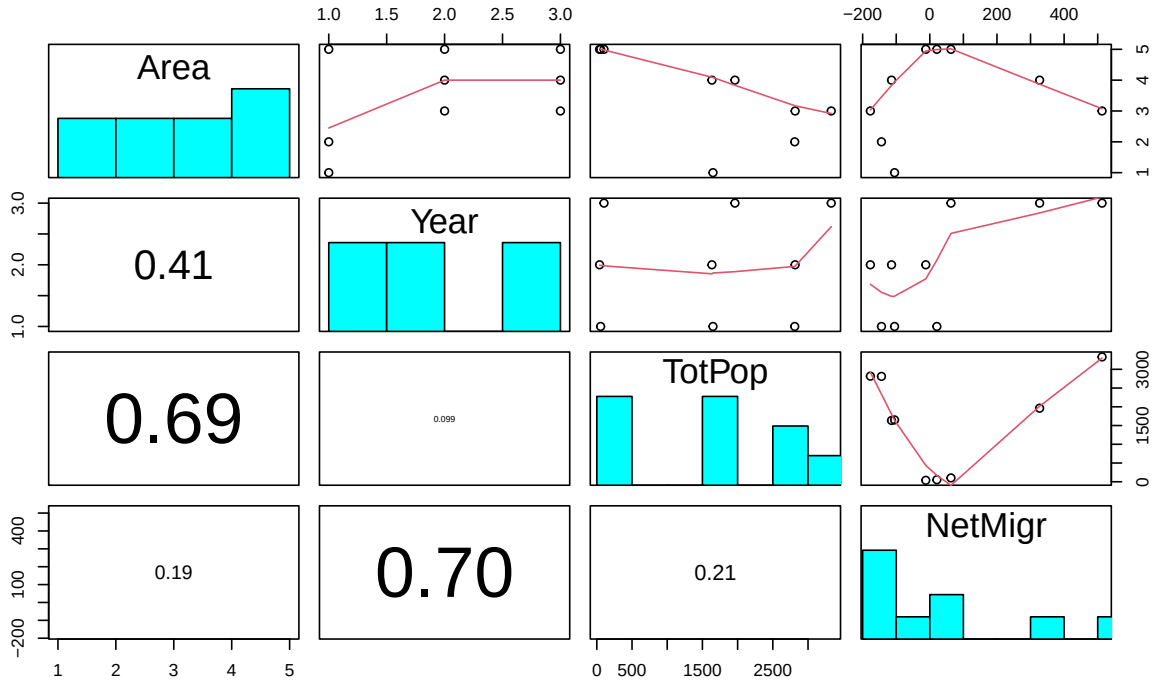
```
pop.df = read.csv("Data/pop_migr.csv", header = TRUE, stringsAsFactors = TRUE, na.strings = "NULL")
pop.df$Year = as.factor(pop.df$Year)    # converting int to Factor
pop.df = pop.df[4:21,]
pop.df$Area = factor(pop.df$Area, level = c("Raumanga East", "Raumanga West", "Raumanga", "Tarewa", "Po
str(pop.df)
```

```
## 'data.frame':    18 obs. of  5 variables:
##  $ Area   : Factor w/ 8 levels "Raumanga East",..: 1 2 3 3 4 4 5 5 5 6 ...
##  $ Year   : Factor w/ 3 levels "2001","2013",..: 1 1 2 3 2 3 1 2 3 1 ...
##  $ TotPop : int  1650 2811 2817 3330 1635 1962 54 39 102 762 ...
##  $ NetMigr: int  -105 -144 -177 513 -114 327 21 -12 63 -33 ...
##  $ SH1    : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
```

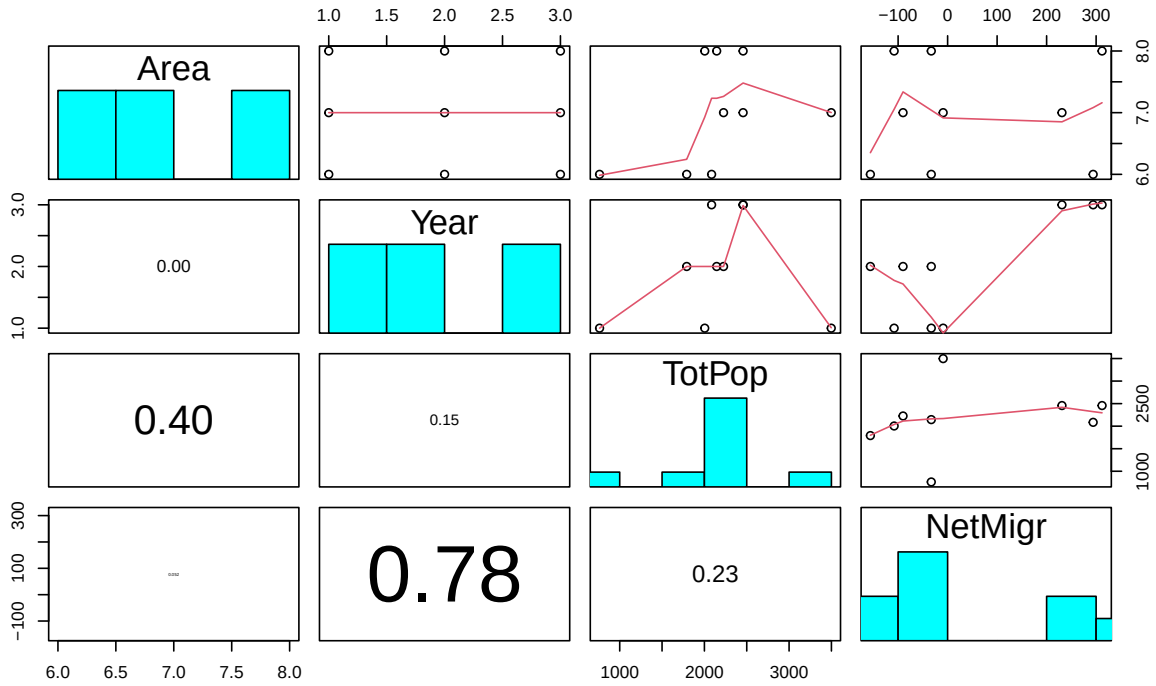
```
pairs20x(pop.df)
```




```
pop_sh1.df = pop.df[which(pop.df$SH1 == TRUE), names(pop.df) %in% c("Area", "Year", "TotPop", "NetMigr")]
pairs20x(pop_sh1.df)
```



```
pop_nsh1.df = pop.df[which(pop.df$SH1 == FALSE), names(pop.df) %in% c("Area", "Year", "TotPop", "NetMigr
pairs20x(pop_nsh1.df)
```




**Comment on the pairs plot**

**Fit and check model**

```
pop.fit = glm(TotPop ~ SH1, family = poisson, data = pop.df)
#plot(income_tot.fit, which = 1)    # identically distributed
#plot(income_tot.fit, which = 2)    # normality assumption
summary(income_tot.fit)
```

```
##
## Call:
## glm(formula = Income ~ SH1 * Area, family = poisson, data = income_tot.df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -19.813   -3.781   -0.645    4.103   12.156
##
## Coefficients: (8 not defined because of singularities)
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   7.44191    0.01398 532.395   <2e-16 ***
## SH1TRUE                      -0.36948    0.03230 -11.438   <2e-16 ***
## AreaRaumanga West             0.50930    0.03685  13.821   <2e-16 ***
## AreaRaumanga                  0.62810    0.03278  19.162   <2e-16 ***
## AreaTarewa                    0.08881    0.03516   2.526   0.0115 *
## AreaPort-Limeburners         -3.01198    0.08121 -37.088   <2e-16 ***
## AreaRiverside                -0.27564    0.02128 -12.954   <2e-16 ***
```

```
## AreaSherwood Rise              0.21064    0.01881  11.201   <2e-16 ***
## AreaOnerahi                         NA         NA      NA       NA
## SH1TRUE:AreaRaumanga West           NA         NA      NA       NA
## SH1TRUE:AreaRaumanga                NA         NA      NA       NA
## SH1TRUE:AreaTarewa                  NA         NA      NA       NA
## SH1TRUE:AreaPort-Limeburners        NA         NA      NA       NA
## SH1TRUE:AreaRiverside               NA         NA      NA       NA
## SH1TRUE:AreaSherwood Rise           NA         NA      NA       NA
## SH1TRUE:AreaOnerahi                 NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 10057.77  on 17  degrees of freedom
## Residual deviance:   935.72  on 10  degrees of freedom
## AIC: 1107.5
##
## Number of Fisher Scoring iterations: 4
```

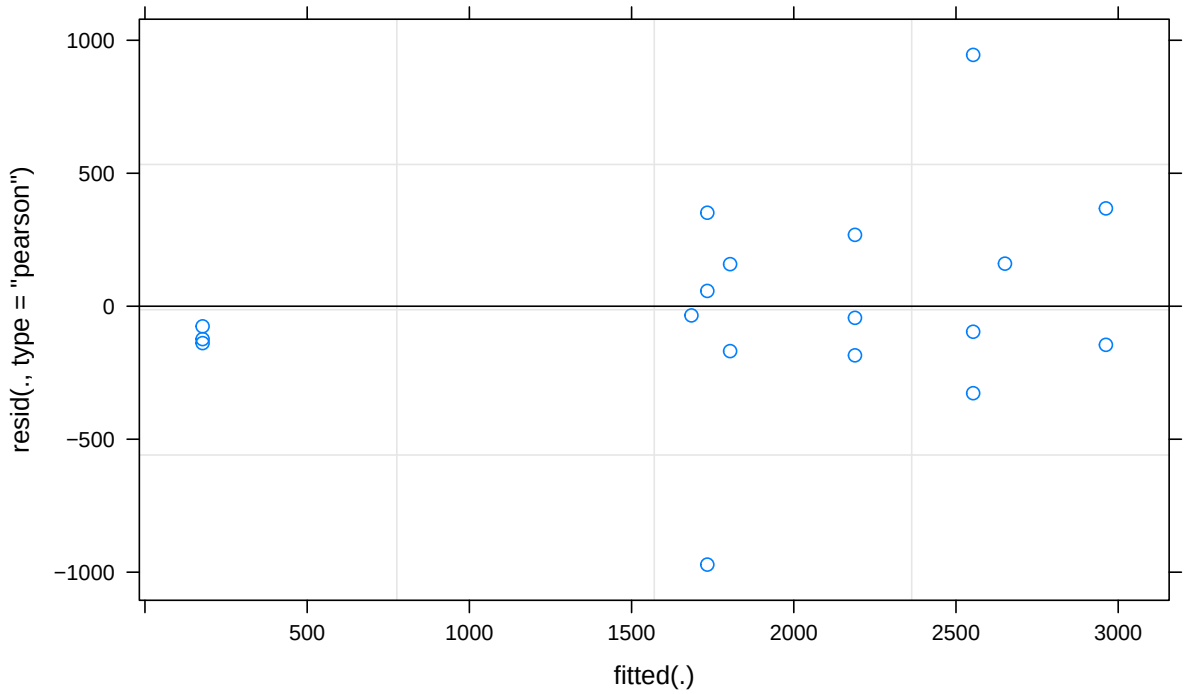
```
1 - pchisq(109.96, 4)     # residual deviance, https://stats.stackexchange.com/questions/108995/interpre
```

```
## [1] 0
```

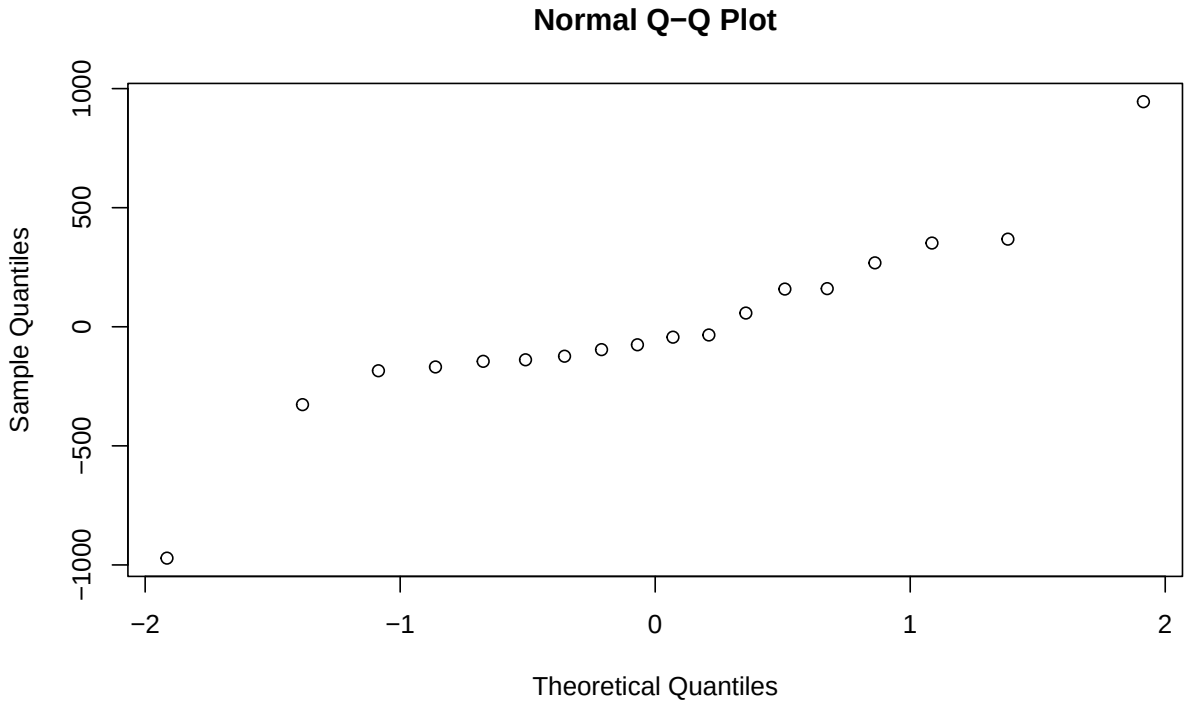
```
pop.lmer = lmer(TotPop ~ SH1 + (1 + SH1 | Area) + (1 + SH1 | Year), data = pop.df, REML = FALSE)
```

```
## boundary (singular) fit: see ?isSingular
```

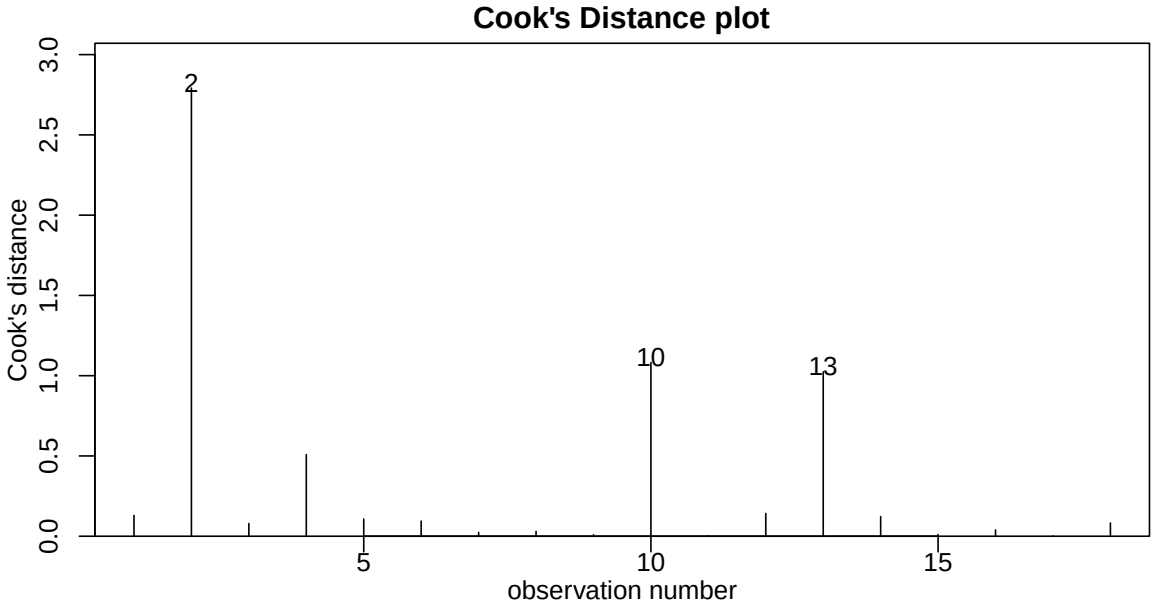
```
plot(pop.lmer)
```



```
qqnorm(resid(pop.lmer))
```




```
cooks20x(pop.lmer)
```



```
summary(pop.lmer)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: TotPop ~ SH1 + (1 + SH1 | Area) + (1 + SH1 | Year)
##    Data: pop.df
##
##      AIC      BIC   logLik deviance df.resid
##    304.7    312.7   -143.4    286.7        9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0983 -0.3100 -0.1288  0.3448  2.0413
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.  Corr
##  Area     (Intercept) 1.620e+05 4.025e+02
##           SH1TRUE     5.630e+05 7.503e+02 0.56
##  Year     (Intercept) 0.000e+00 0.000e+00
##           SH1TRUE     4.358e-05 6.601e-03  NaN
##  Residual             2.143e+05 4.630e+02
## Number of obs: 18, groups:  Area, 8; Year, 3
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 2158.333    278.936    3.000   7.738  0.00449 **
## SH1TRUE     -302.556    564.551    7.833  -0.536  0.60689
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## SH1TRUE -0.494
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```