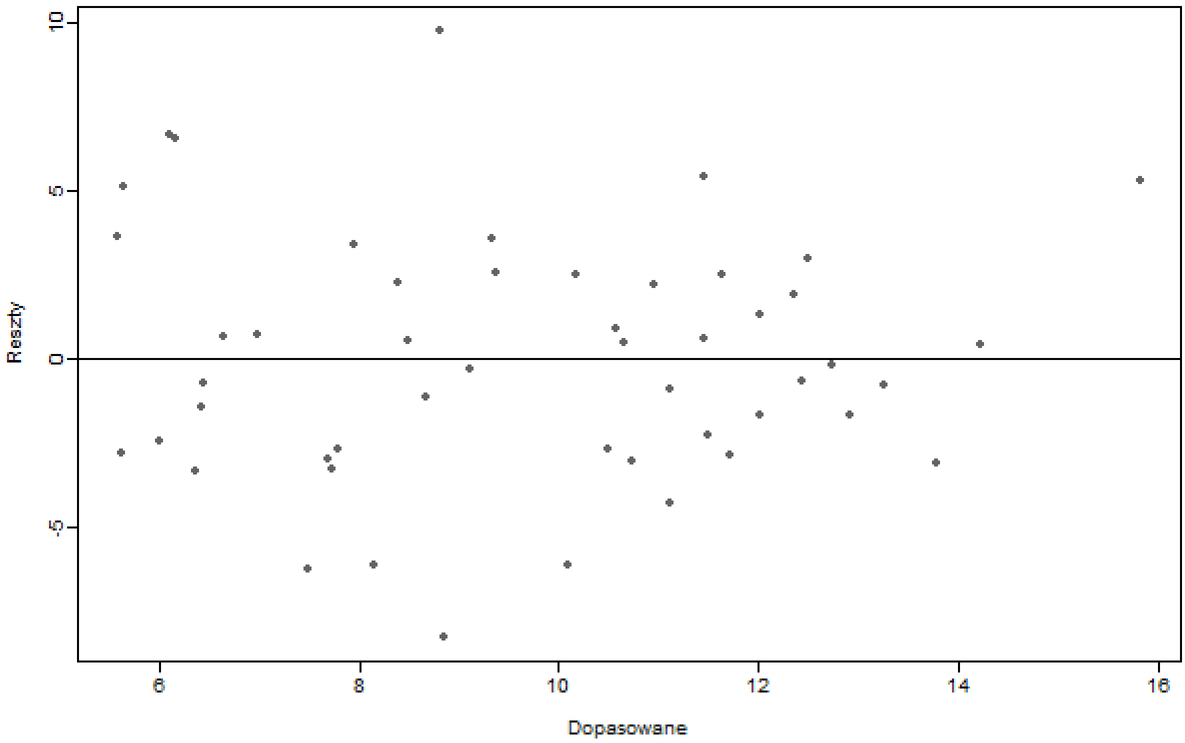
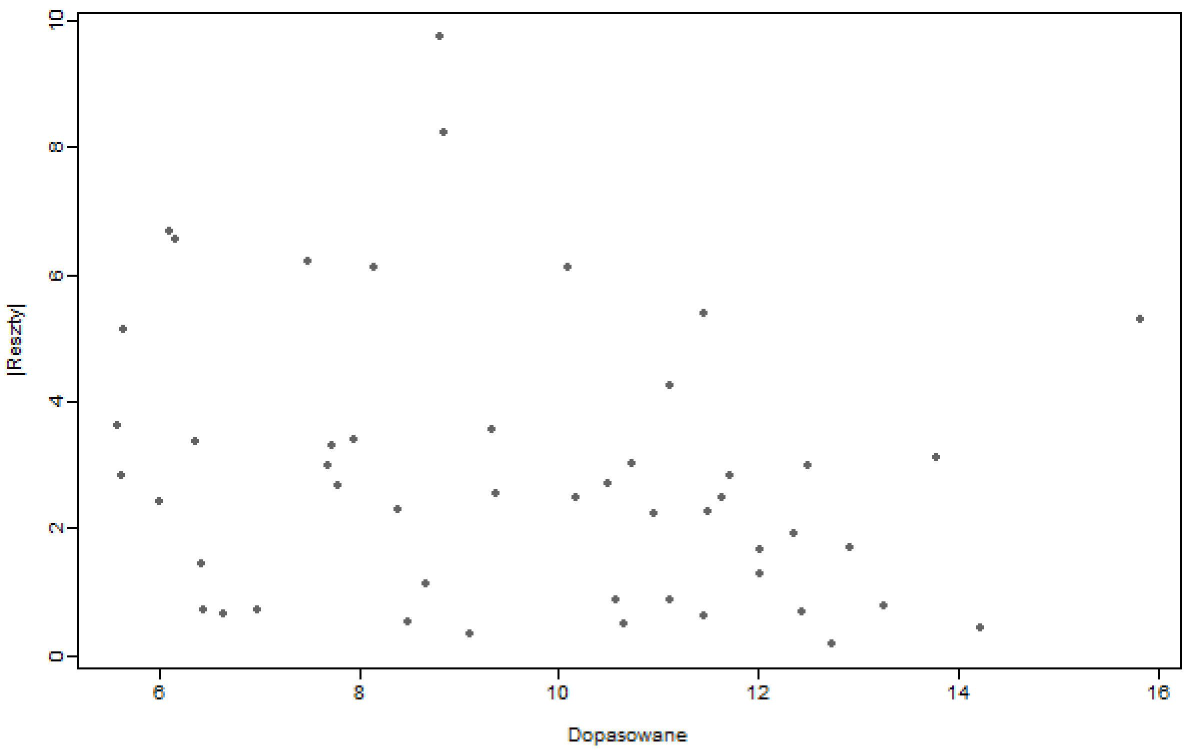
Reszty



*

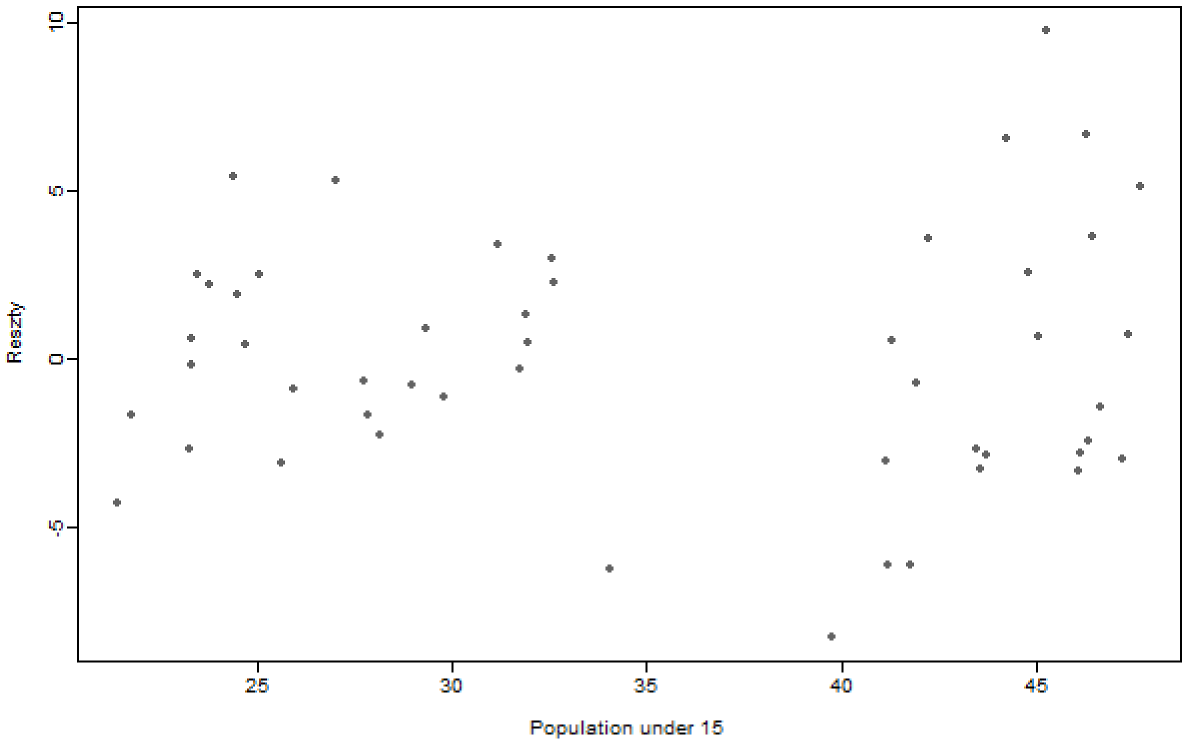
Dla oceny stałości wariancji:



> summary(lm(abs(g$res) ~ g$fit))

```
Call:
lm(formula = abs(g$res) ~ g$fit)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.8395 -1.6078 -0.3493  0.6625  6.7036 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.8398     1.1865   4.079  0.00017 ***
g$fit        -0.2035     0.1185  -1.717  0.09250 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 2.163 on 48 degrees of freedom
Multiple R-squared:  0.05784,   Adjusted R-squared:  0.03821 
F-statistic: 2.947 on 1 and 48 DF,  p-value: 0.0925
```

Reszty jako funkcje zmiennych objaśniających



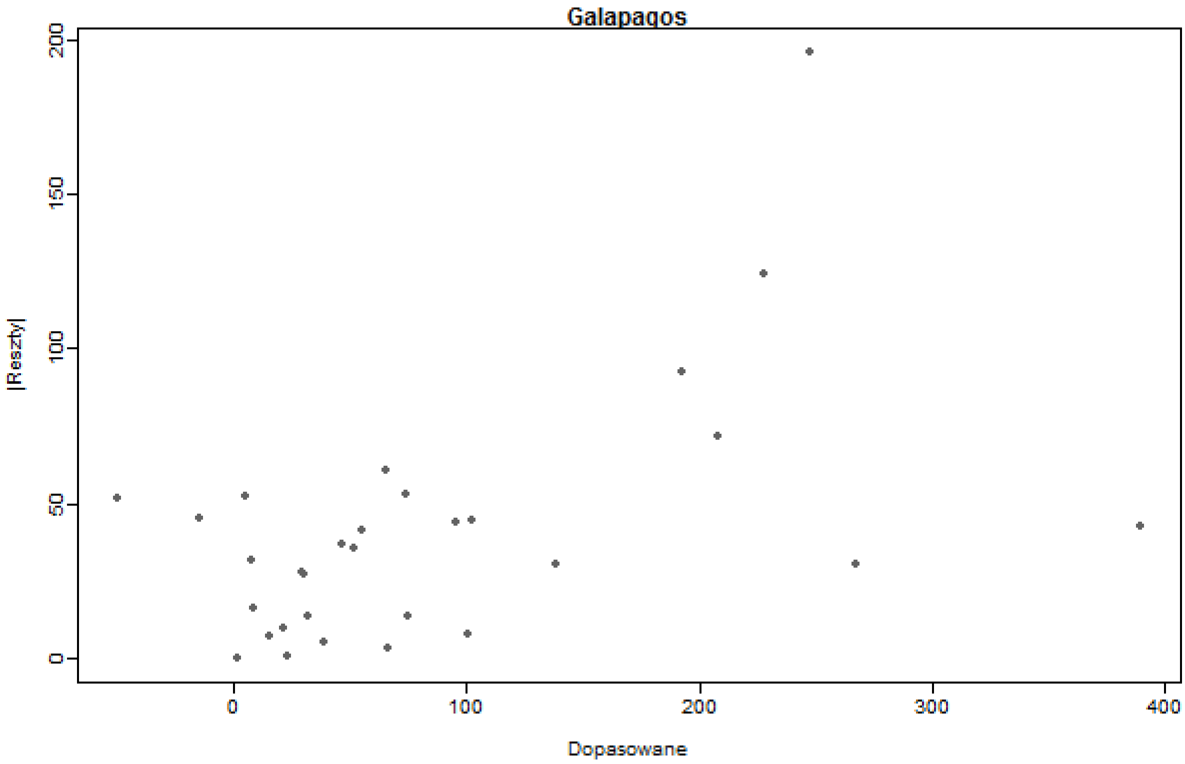
*

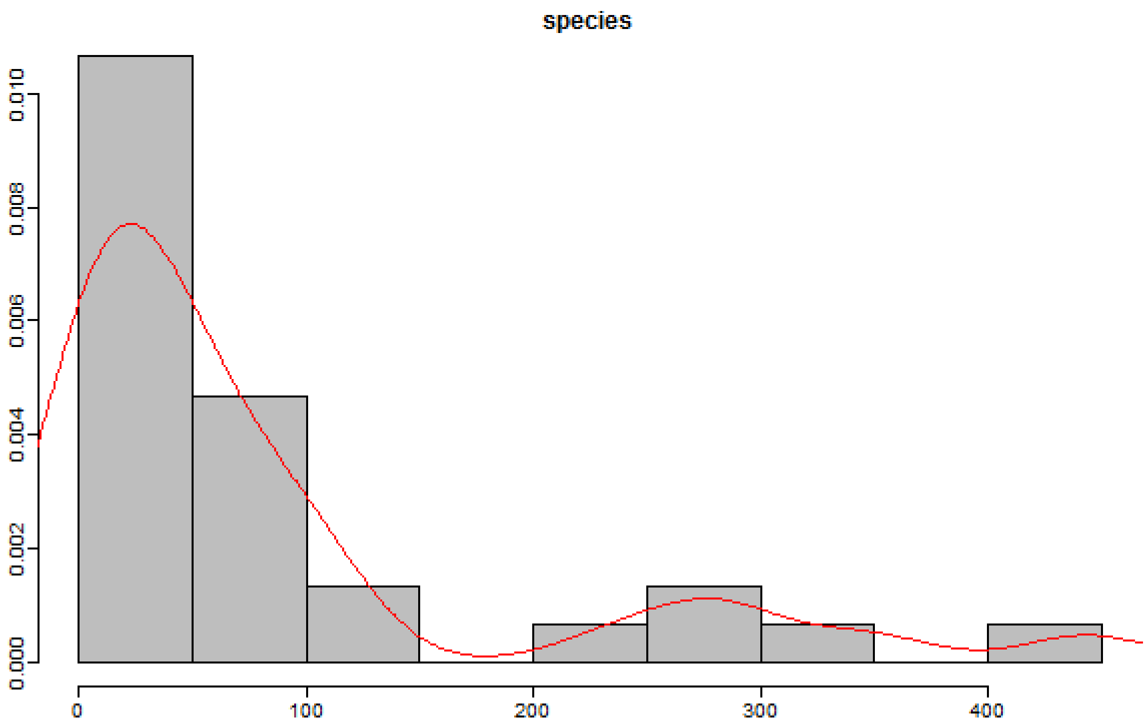
Tu widać dwie grupy!

```
> var(g$res[savings$pop15 > 35])/var(g$res[savings$pop15 <35])
[1] 2.785067

> table(savings$pop15 > 35)

FALSE  TRUE 
   27    23 
> 1-pf(2.7851,22,26)
[1] 0.006787451
```

Przekształcenie stabilizujące wariancję

```
> splm <- lm(species~., gala.eaa)
> summary(splm)

Call:
lm(formula = species ~ ., data = gala.eaa)

Residuals:
    Min       1Q   Median       3Q      Max
-124.064  -34.283   -8.733   27.972  195.973

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.71893   16.90706  -0.338  0.73789
area        -0.02031    0.02181  -0.931  0.36034
elevation    0.31498    0.05211   6.044  2.2e-06 ***
adjacent    -0.07528    0.01698  -4.434  0.00015 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 61.01 on 26 degrees of freedom
Multiple R-squared:  0.746,     Adjusted R-squared:  0.7167
F-statistic: 25.46 on 3 and 26 DF,  p-value: 6.683e-08
```





Przekształcenie stabilizujące wariancję
Można pokazać, ze wariancja $h(\bar{X}) \approx \sigma^2(h'(\mu))^2$. Przekształcenie h stabilizujące wariancję $h(\bar{X})$ musi, dla pewnej stałej c, spełniać równanie

$$\sigma^2(h'(\mu))^2 = c^2$$

Gdy h jest rosnąca to musi spełniać równanie

$$h'(\mu) = \frac{c}{\sigma}$$

Np dla rodziny rozkładów Poissona z parametrem λ, $\mu(\lambda) = \lambda$, $\sigma(\lambda) = \sqrt{\lambda}$. Wstawiając do równania, otrzymamy

$$h'(\lambda) = \frac{c}{\sqrt{\lambda}},$$

$$h(\lambda) = 2c\sqrt{\lambda}$$

a więc przekształcenie pierwiatkowe stabilizuje wariancję.
Gdy $log(\sigma) = a\,log(\mu) + b$, czyli gdy $\sigma = e^b\mu^a$ to

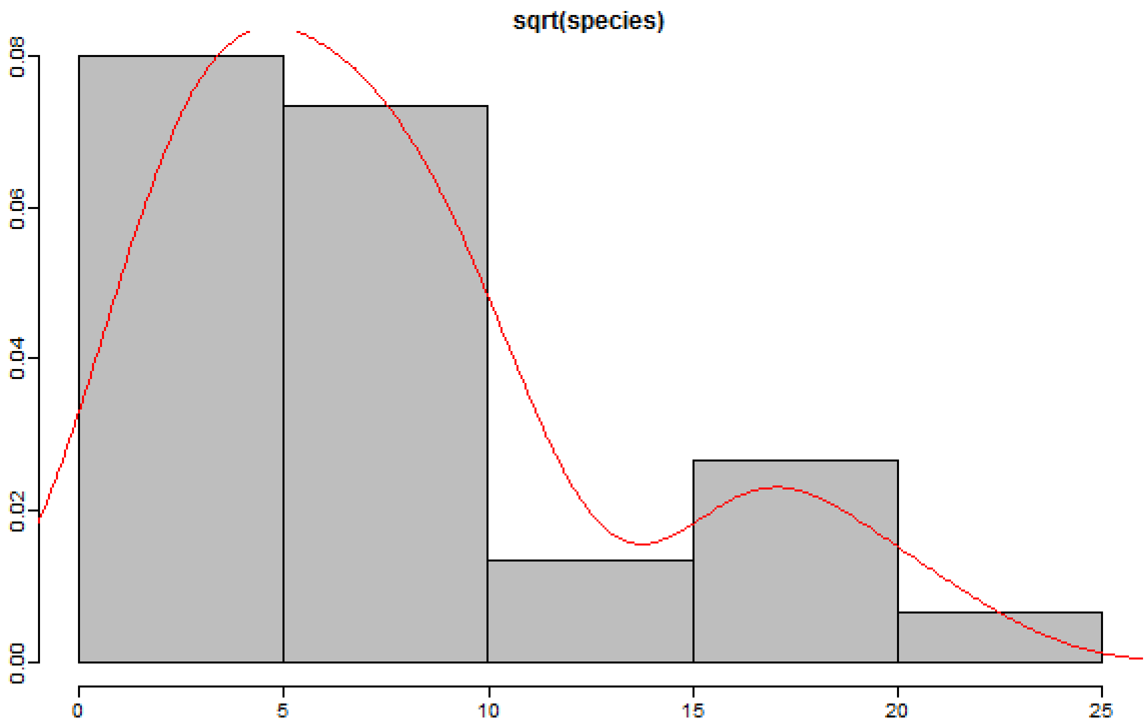
$$h'(\mu) = ce^{-b}\mu^{-a},$$

$$h(\mu) = \frac{ce^{-b}}{1-a}\mu^{1-a}$$

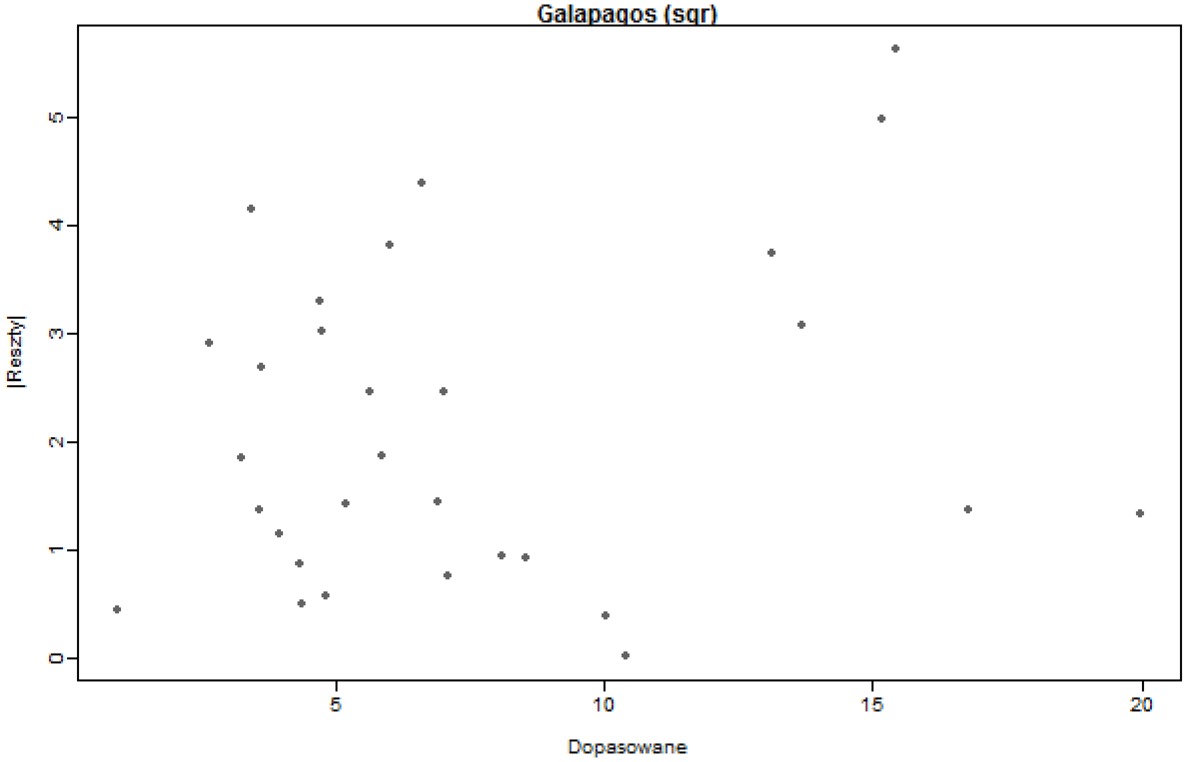
czyli przekształcenie stabilizujące wariancję jest potęgowe o potędze $1-a$.
Zgadza się to z teorią Tukeya, że gdy $log(\sigma) = a\,log(\mu) + b$, to przekształcenie stabilizujące wariancję jest potęgowe o potędze $1-a$
Wg rozdz. 1.5 Bickel, Doksum

```
> sp_sort <- sort(gala.eaa$species)

> (sp_mt <- matrix(sp_sort,nrow = 6,ncol = 5))
     [,1] [,2] [,3] [,4] [,5]
[1,]    2   10   25   58  108
[2,]    2   12   31   62  237
[3,]    2   16   40   70  280
[4,]    3   18   44   93  285
[5,]    5   21   51   97  347
[6,]    8   24   58  104  444


> x <- log(apply(X = sp_mt,MARGIN = 2,median))
> y <- log(apply(X = sp_mt,MARGIN = 2,IQR))
> (rbind(x,y))
       [,1]     [,2]     [,3]     [,4]     [,5]
x 0.9162907 2.833213 3.737670 4.400603 5.643679
y 0.9162907 1.981001 2.772589 3.465736 4.427836


> summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.06591    0.20275   0.325 0.766485
x            0.75487    0.05268  14.328 0.000737 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1869 on 3 degrees of freedom
Multiple R-squared:  0.9856,    Adjusted R-squared:  0.9808
F-statistic: 205.3 on 1 and 3 DF,  p-value: 0.0007368
```

Przekształcenie o potędze 1-0.75=0.25



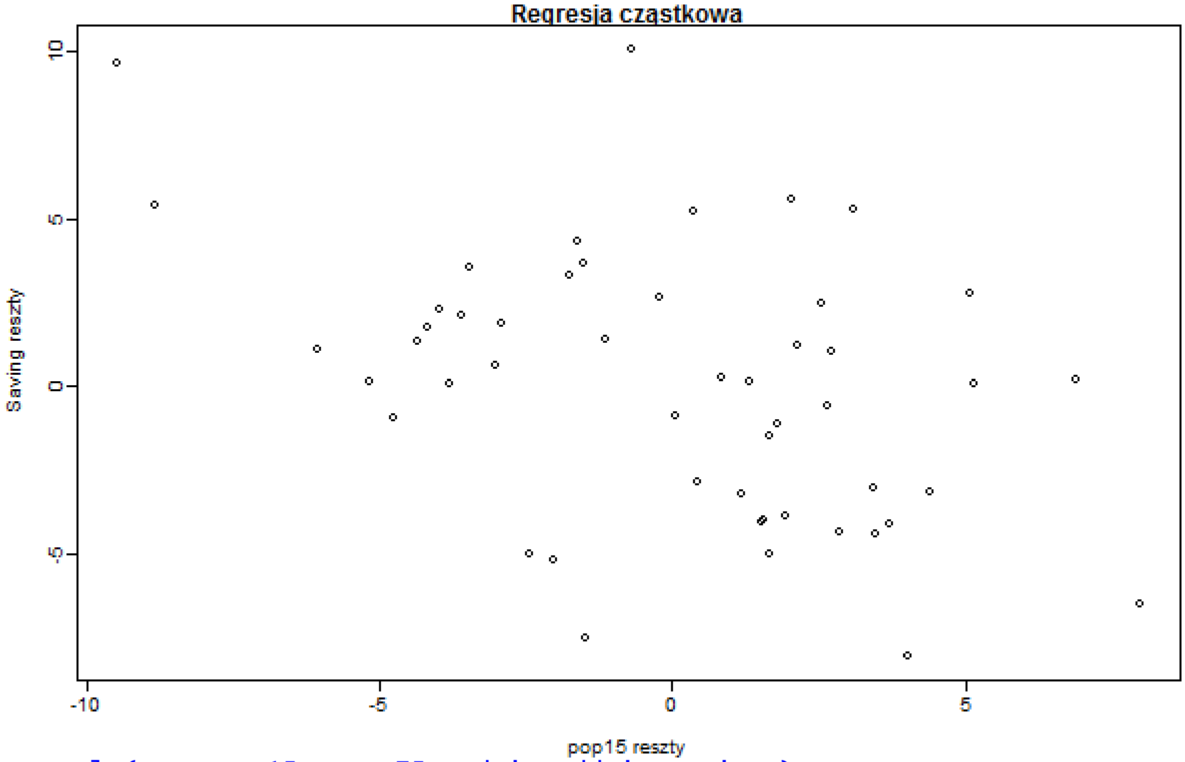
```
> splm.sqr <- lm(sqrt(species)~., gala.eaa)
> summary(splm.sqr)

Call:
lm(formula = sqrt(species) ~ ., data = gala.eaa)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.9752 -1.7377 -0.4927  1.4342  5.6267 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.8611345  0.7698707   3.716 0.000975 ***
area        -0.0019221  0.0009933  -1.935 0.063931 .  
elevation    0.0165770  0.0023729   6.986 2.04e-07 ***
adjacent    -0.0035043  0.0007731  -4.533 0.000115 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 2.778 on 26 degrees of freedom
Multiple R-squared:  0.7638,	Adjusted R-squared:  0.7366 
F-statistic: 28.03 on 3 and 26 DF,  p-value: 2.63e-08
```



Nieliniowość

Partial Regression or *Added Variable* plots can help isolate the effect of x_i on y.

1. Regress y on all x except x_i, get residuals $\hat{\delta}$. This represents y with the other X-effect taken out.
2. Regress x_i on all x except x_i, get residuals $\hat{\gamma}$ This represents x_i with the other X-effect taken out.
3. Plot $\hat{\delta}$ against $\hat{\gamma}$

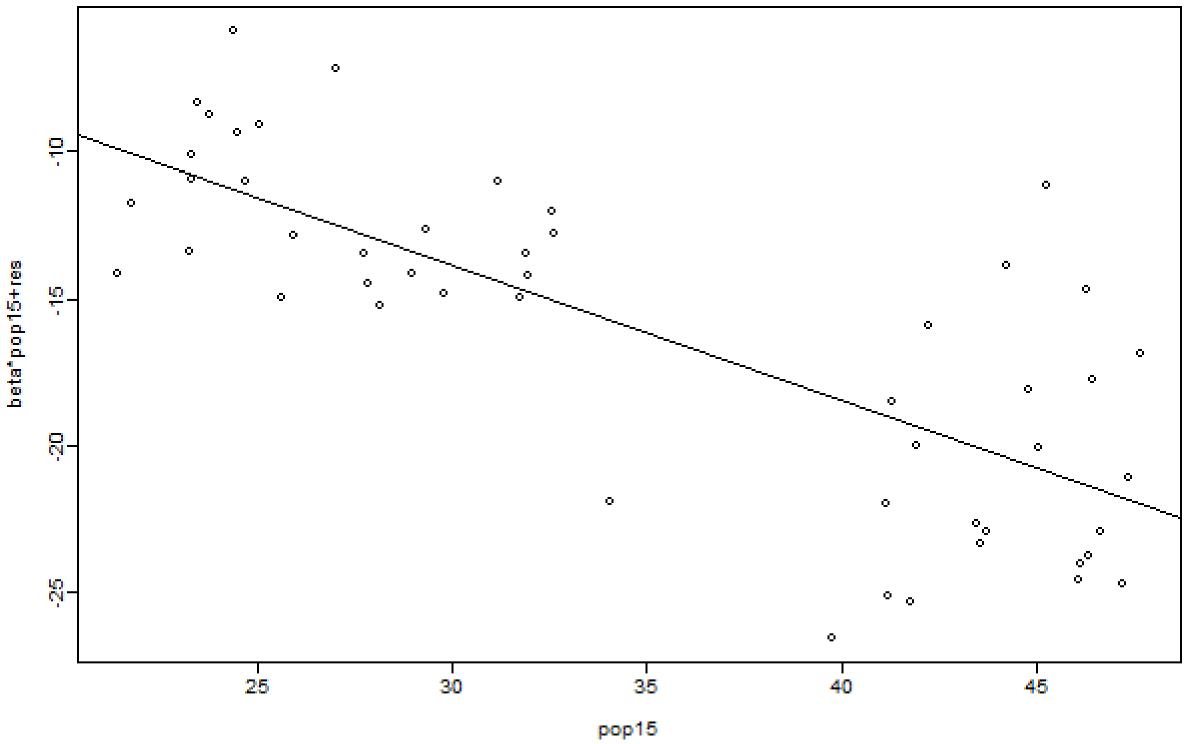



```
> g <- lm(sr ~ pop15 + pop75 + dpi + ddpi, savings)
> d <- lm(sr ~ pop75 + dpi + ddpi,savings)$res
> m <- lm(pop15 ~ pop75 + dpi + ddpi,savings)$res
> lm(d ~m)$coef
 (Intercept)             m
 9.907998e-17 -4.611931e-01
> g$coef
 (Intercept)         pop15         pop75           dpi          ddpi
28.5660865407 -0.4611931471 -1.6914976767 -0.0003369019  0.4096949279
```

Można bezpośrednio skorzystać z pakietu *faraway*:

```
> prplot(g,1)
```

$$y-\sum_{j\neq i} x_j\hat{\beta}_j = \hat{y}+\hat{\varepsilon}-\sum_{j\neq i} x_j\hat{\beta}_j = x_i\hat{\beta}_i+\hat{\varepsilon}$$



Jeszcze raz widać dwie grupy:

```
> g1 <- lm(sr ~ pop15+pop75+dpi+ddpi,savings,subset=(pop15 > 35))
> g2 <- lm(sr ~ pop15+pop75+dpi+ddpi,savings,subset=(pop15 < 35))
> summary(g1)

Call:
lm(formula = sr ~ pop15 + pop75 + dpi + ddpi, data = savings,
  subset = (pop15 > 35))

Residuals:
    Min      1Q  Median      3Q     Max
-5.5511 -3.5101  0.0443  2.6764  8.4983

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.4339689 21.1550278  -0.115    0.910
pop15        0.2738537  0.4391910   0.624    0.541
pop75       -3.5484769  3.0332806  -1.170    0.257
dpi          0.0004208  0.0050001   0.084    0.934
ddpi         0.3954742  0.2901012   1.363    0.190

Residual standard error: 4.454 on 18 degrees of freedom
Multiple R-squared:  0.1558,    Adjusted R-squared:  -0.03185
F-statistic: 0.8302 on 4 and 18 DF,  p-value: 0.5233

> summary(g2)

Call:
lm(formula = sr ~ pop15 + pop75 + dpi + ddpi, data = savings,
  subset = (pop15 < 35))

Residuals:
    Min      1Q  Median      3Q     Max
-5.5890 -1.5015  0.1165  1.8857  5.1466
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.9617950  8.0837502   2.964  0.00716 **
pop15       -0.3858976  0.1953686  -1.975  0.06092 .
pop75       -1.3277421  0.9260627  -1.434  0.16570
dpi         -0.0004588  0.0007237  -0.634  0.53264
ddpi         0.8843944  0.2953405   2.994  0.00668 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.772 on 22 degrees of freedom
Multiple R-squared:  0.5073,    Adjusted R-squared:  0.4177
F-statistic: 5.663 on 4 and 22 DF,  p-value: 0.002734
```