Modelos Computacion

Miguel Diaz

2022-04-25

Carga de datos

```
PABLO <- read_excel("C:/Users/usuario/Desktop/RCOMPUTACION/XPABLO.xlsx")
PABLO
```

```
## # A tibble: 403 x 19
##       id  Long   Lat     z    MO    Ca    Mg      K    Na  CICE    CE    Fe
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
##  1     1 -72.6  8.08   120  2.09  7.83 1.56  0.175  0.291  9.85 0.130 133.
##  2     2 -72.6  8.08   119  1.65  3.95 0.771 0.496  0.136  5.36 0.126  29.7
##  3     3 -72.6  8.08   111  1.65  5.88 1.23  0.273  0.135  7.52 0.287 237.
##  4     4 -72.6  8.08   114  2.48  5.62 1.13  0.217  0.163  7.13 0.415 331.
##  5     5 -72.6  8.09   115  3.01 11.4  2.36  0.501  0.292 14.6  0.269 281.
##  6     6 -72.6  8.09   109  1.93  7.49 1.56  0.244  0.115  9.41 0.410 258.
##  7     7 -72.6  8.09   116  2.86 10.9  2.40  0.195  0.282 13.8  0.141 167.
##  8     8 -72.6  8.10   109  2.20 12.1  2.73  0.0438 0.420 15.3  0.163  54.5
##  9     9 -72.6  8.10   109  2.64 15.7  5.54  0.265  0.454 22.9  0.173  96.4
## 10    10 -72.6  8.10   115  2.06  7.96 1.78  0.133  0.308 10.2  0.245 446.
## # ... with 393 more rows, and 7 more variables: Cu <dbl>, Zn <dbl>, cos <dbl>,
## #   mod1 <dbl>, mod2 <dbl>, mod3 <dbl>, mod4 <chr>
```

Modelo Lineal K en relacion al Na (1)

```
model_1 <- lm(K ~ Na, data = PABLO)
summary(model_1)
```

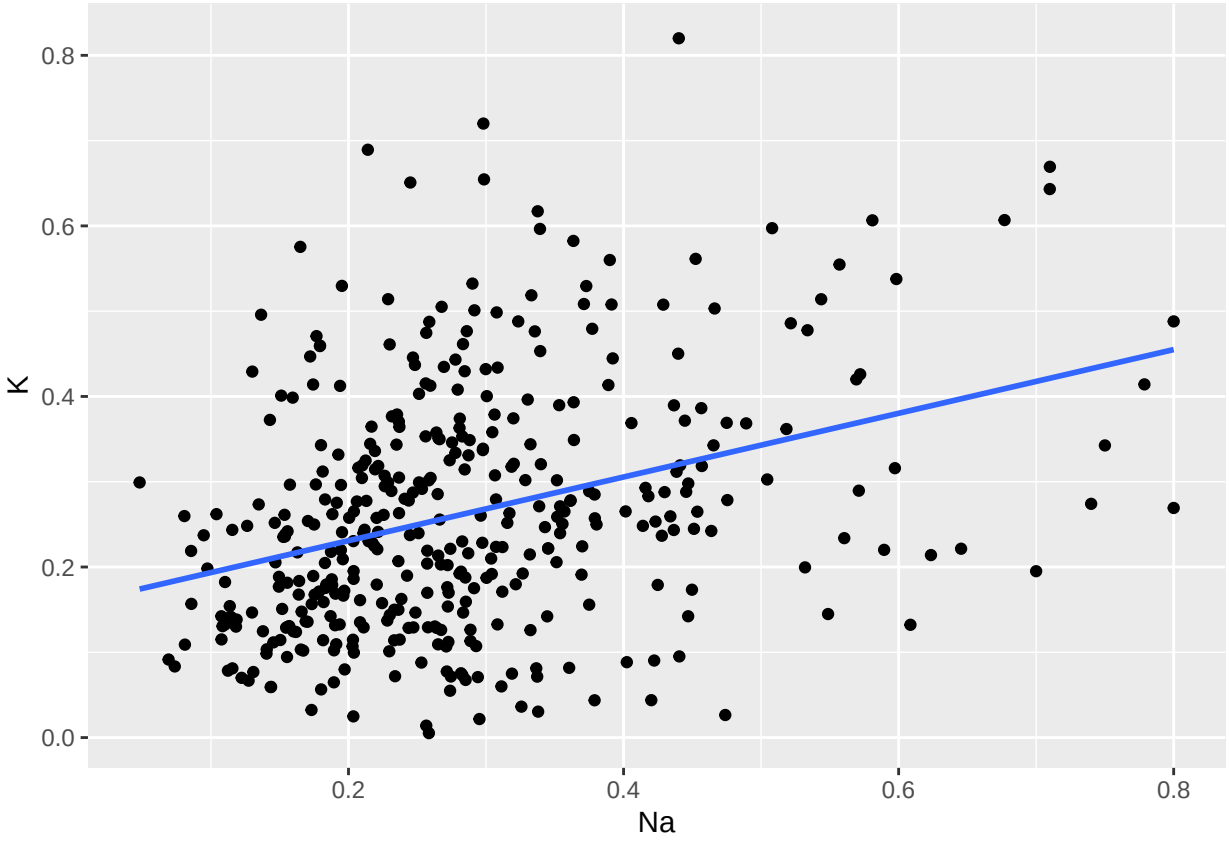
```
##
## Call:
## lm(formula = K ~ Na, data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30673 -0.09094 -0.02348  0.07565  0.49944
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.15615    0.01598   9.769  < 2e-16 ***
## Na           0.37352    0.05095   7.332 1.27e-12 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1357 on 401 degrees of freedom
## Multiple R-squared:  0.1182, Adjusted R-squared:  0.116
## F-statistic: 53.75 on 1 and 401 DF,  p-value: 1.267e-12
```

$$Y_K = 0.156 + 0.374 X_{Na}$$

```
ggplot(PABLO, aes(y = K, x = Na)) +
  geom_point()+
  geom_smooth(method='lm', se = F)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Filtrado Na > 0.5

```
PABLO2 <- PABLO |>
  filter(Na <= 0.5)
PABLO2
```

```
## # A tibble: 374 x 19
##       id  Long   Lat     z    MO    Ca    Mg      K    Na  CICE    CE    Fe
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
##  1     1 -72.6  8.08   120  2.09  7.83 1.56  0.175  0.291  9.85 0.130 133.
##  2     2 -72.6  8.08   119  1.65  3.95 0.771 0.496  0.136  5.36 0.126  29.7
##  3     3 -72.6  8.08   111  1.65  5.88 1.23  0.273  0.135  7.52 0.287 237.
##  4     4 -72.6  8.08   114  2.48  5.62 1.13  0.217  0.163  7.13 0.415 331.
##  5     5 -72.6  8.09   115  3.01 11.4  2.36  0.501  0.292 14.6  0.269 281.
##  6     6 -72.6  8.09   109  1.93  7.49 1.56  0.244  0.115  9.41 0.410 258.
##  7     7 -72.6  8.09   116  2.86 10.9  2.40  0.195  0.282 13.8  0.141 167.
##  8     8 -72.6  8.10   109  2.20 12.1  2.73  0.0438 0.420 15.3  0.163  54.5
##  9     9 -72.6  8.10   109  2.64 15.7  5.54  0.265  0.454 22.9  0.173  96.4
## 10    10 -72.6  8.10   115  2.06  7.96 1.78  0.133  0.308 10.2  0.245 446.
## # ... with 364 more rows, and 7 more variables: Cu <dbl>, Zn <dbl>, cos <dbl>,
## #   mod1 <dbl>, mod2 <dbl>, mod3 <dbl>, mod4 <chr>
```

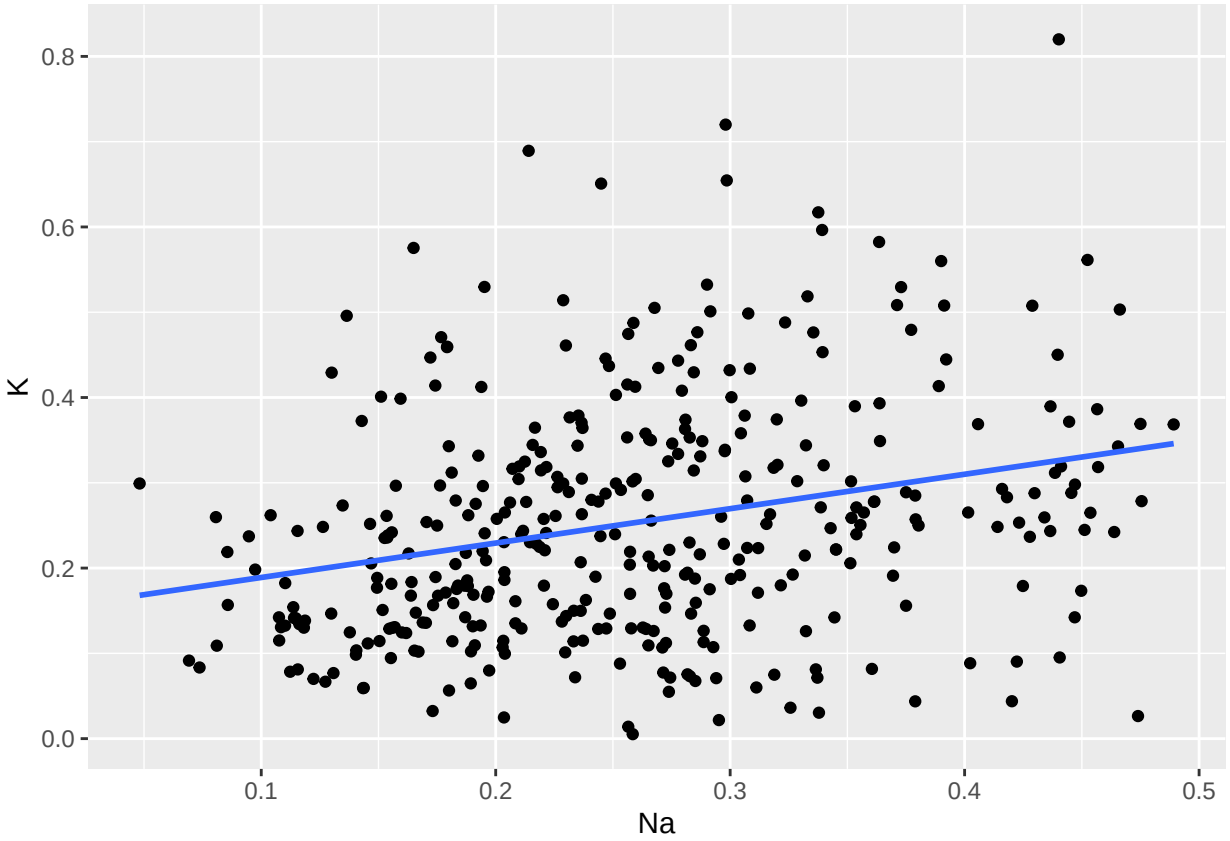
```
model_2 <- lm(K ~ Na, data = PABLO2)
summary(model_2)
```

```
##
## Call:
## lm(formula = K ~ Na, data = PABLO2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31339 -0.08750 -0.02389  0.07271  0.49380
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.14861    0.02004   7.415 8.28e-13 ***
## Na           0.40346    0.07284   5.539 5.76e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1333 on 372 degrees of freedom
## Multiple R-squared:  0.0762, Adjusted R-squared:  0.07371
## F-statistic: 30.68 on 1 and 372 DF,  p-value: 5.755e-08
```

$$Y_K = 0.149 + 0.403 X_{Na}$$

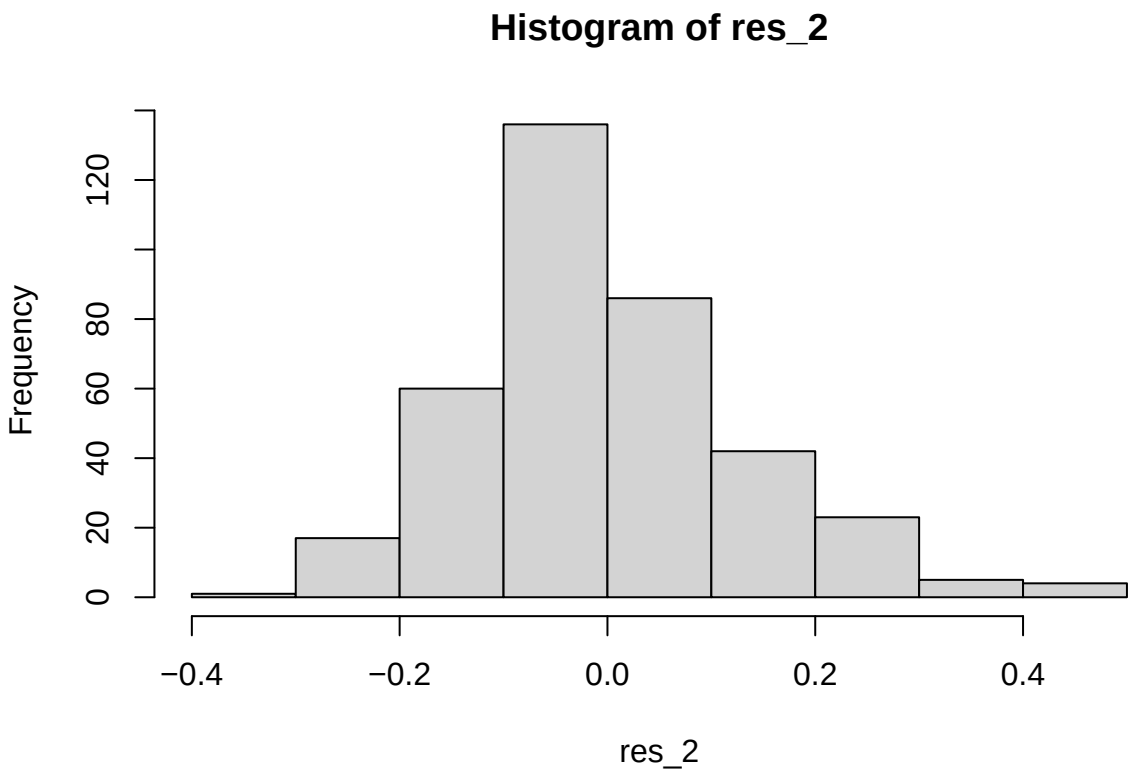
```
ggplot(PABLO2, aes(y = K, x = Na)) +
  geom_point()+
  geom_smooth(method='lm', se = F)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Residuales con y sin valor absoluto

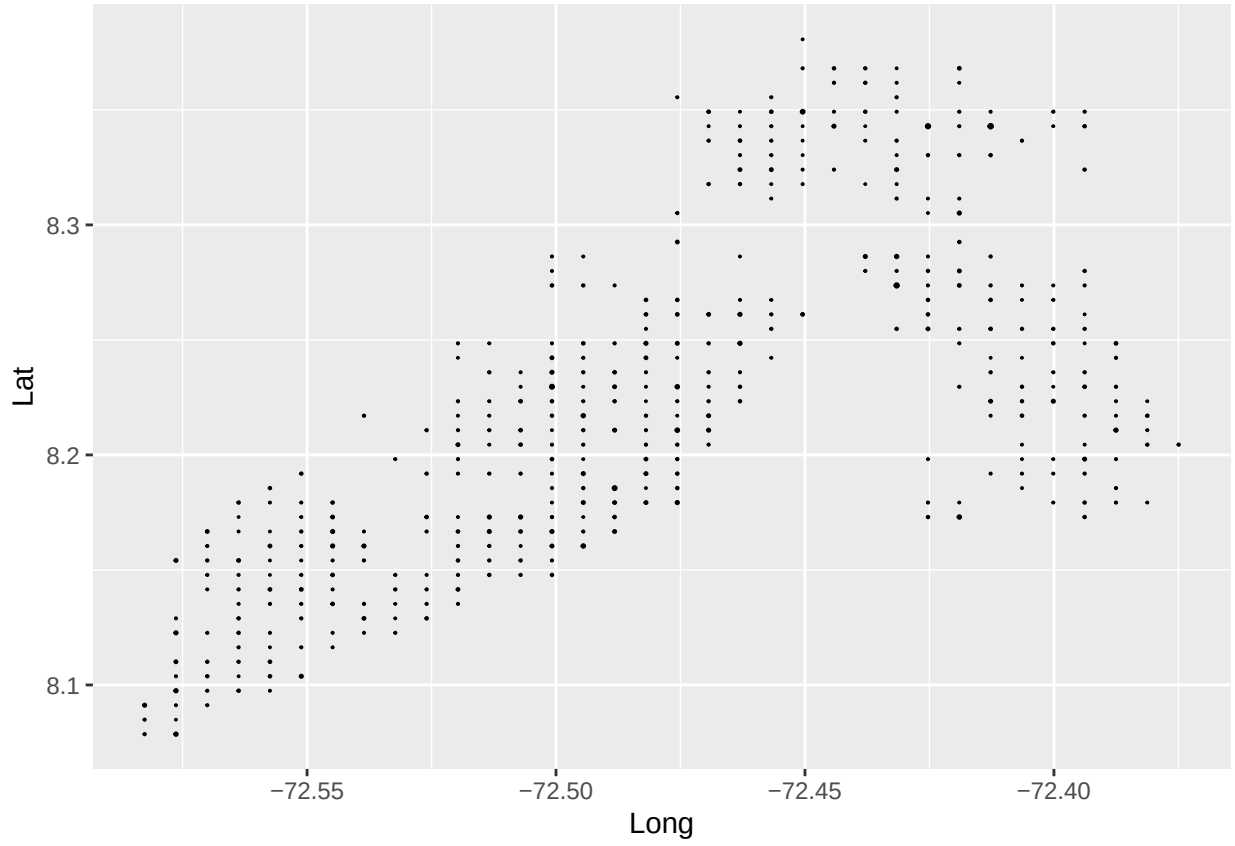
```
res_2 <- model_2$residuals
hist(res_2)
```



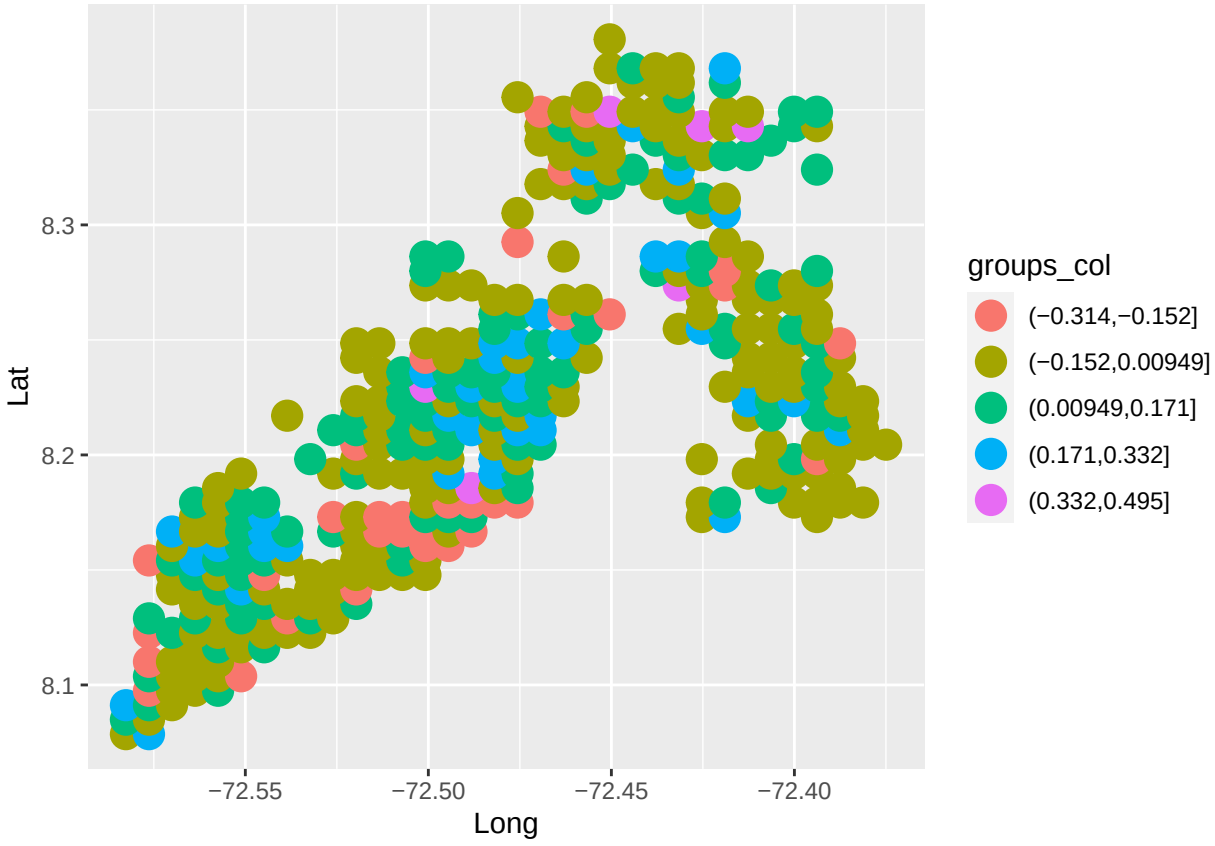

```
res_2abs <- abs(res_2)
hist(res_2abs)
```

Valor absoluto para los residuales del modelo

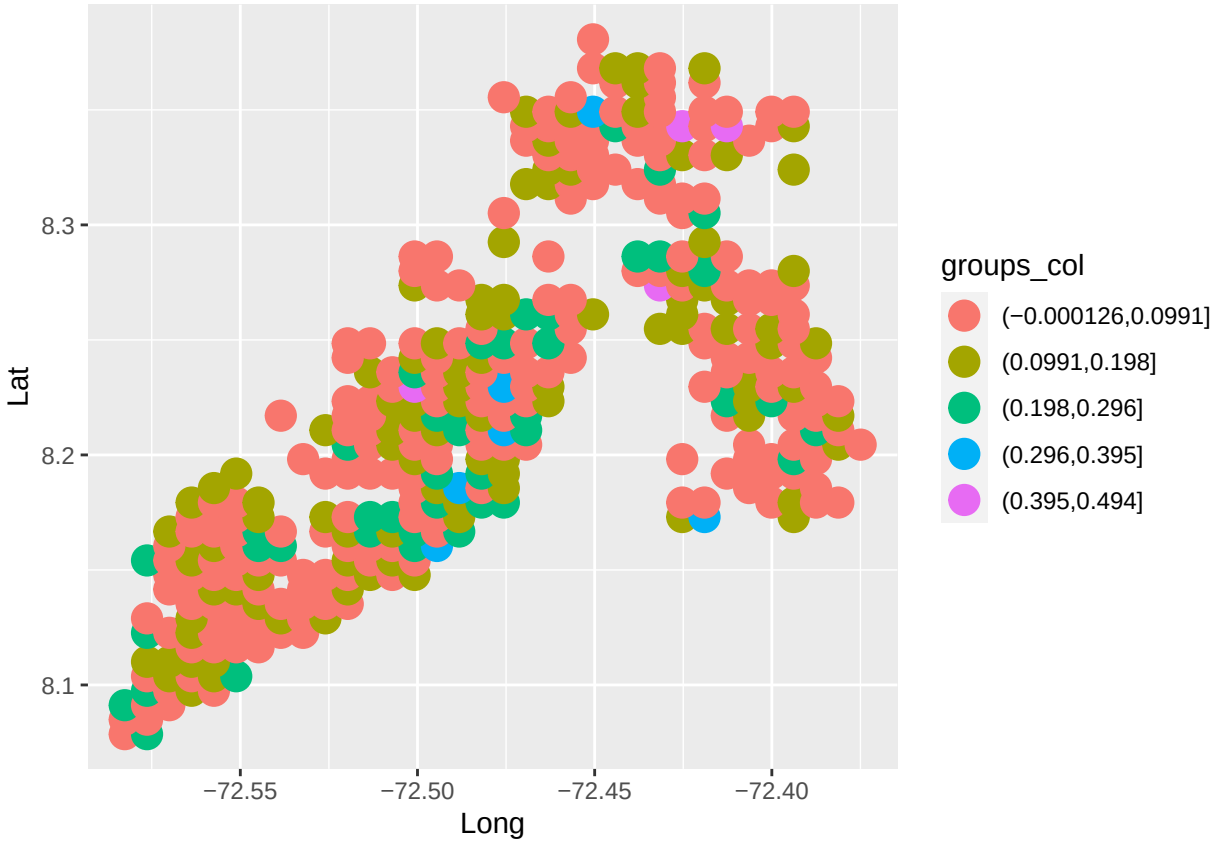
```
ggplot(PABLO2, aes(Long, Lat))+
  geom_point(size = abs(res_2))
```




```
groups_col <- cut(res_2, breaks = 5)
ggplot(PABLO2, aes(Long, Lat, color = groups_col))+
  geom_point(size = 5)
```




```
groups_col <- cut(res_2abs, breaks = 5)
ggplot(PABLO2, aes(Long, Lat, color = groups_col))+
  geom_point(size = 5)
```



Indice de Moran para residuales del modelo lineal

```
dist_matr <- as.matrix(dist(cbind(x = PABLO2$Long, y = PABLO2$Lat)))
dist_matr_inv<- 1/dist_matr
diag(dist_matr_inv) <- 0
Moran.I(res_2, dist_matr_inv)
```

```
## $observed
## [1] 0.03228386
##
## $expected
## [1] -0.002680965
##
## $sd
## [1] 0.004625763
##
## $p.value
## [1] 4.063416e-14
```

```
shapiro.test(res_2)
```

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

```
##
## data:  res_2
## W = 0.97014, p-value = 6.028e-07
```

Modelo de regresión multiple (2)

```
model_3 <- lm(K ~ Na + Mg, data = PABLO)
summary(model_3)
```

```
##
## Call:
## lm(formula = K ~ Na + Mg, data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28354 -0.07296 -0.01698  0.06061  0.54313
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.087702   0.014155   6.196 1.44e-09 ***
## Na          0.068130   0.047773   1.426    0.155
## Mg          0.077376   0.005686  13.608  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1123 on 400 degrees of freedom
## Multiple R-squared:  0.3972, Adjusted R-squared:  0.3942
## F-statistic: 131.8 on 2 and 400 DF,  p-value: < 2.2e-16
```

$$Y_K = 0.088 + 0.068X_{Na} + 0.077Z_{Mg}$$

```
res_3 <- model_3$residuals
```

Indice de Moran para el modelo de regresion multiple

```
dist_matr2 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv2<- 1/dist_matr2
diag(dist_matr_inv2) <- 0
Moran.I(res_3, dist_matr_inv2)
```

```
## $observed
## [1] 0.01983024
##
## $expected
## [1] -0.002487562
##
## $sd
```

```
## [1] 0.004246057
##
## $p.value
## [1] 1.471244e-07
```

```
shapiro.test(res_3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_3
## W = 0.94411, p-value = 3.458e-11
```

Primer Modelo con ajustes para datos georreferenciados (3)

```
model_4 <- lm(K ~ Na + Long + Lat , data = PABLO)
summary(model_4)
```

```
##
## Call:
## lm(formula = K ~ Na + Long + Lat, data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29097 -0.09088 -0.02648  0.07669  0.47447
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.44715   12.28362  -0.362   0.7175
## Na           0.36562    0.05088   7.186 3.31e-12 ***
## Long        -0.03653    0.15977  -0.229   0.8193
## Lat          0.23812    0.12566   1.895   0.0588 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1351 on 399 degrees of freedom
## Multiple R-squared:   0.13,  Adjusted R-squared:  0.1234
## F-statistic: 19.87 on 3 and 399 DF,  p-value: 5.045e-12
```

```
res_4 <- model_4$residuals
dist_matr3 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv3<- 1/dist_matr3
diag(dist_matr_inv3) <- 0
Moran.I(res_4, dist_matr_inv3)
```

```
## $observed
## [1] 0.03529946
##
## $expected
## [1] -0.002487562
```

```
##
## $sd
## [1] 0.00426002
##
## $p.value
## [1] 0
```

```
shapiro.test(res_4)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_4
## W = 0.97071, p-value = 3.108e-07
```

Segundo Modelo con ajustes para datos georreferenciados (4)

```
model_5 <- lm(K ~ Na + I(Long**2) + I(Lat**2), data = PABLO)
summary(model_5)
```

```
##
## Call:
## lm(formula = K ~ Na + I(Long^2) + I(Lat^2), data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29110 -0.09076 -0.02668  0.07651  0.47447
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.0455279  6.1320789  -0.334    0.739
## Na           0.3657398  0.0508830   7.188 3.26e-12 ***
## I(Long^2)    0.0002352  0.0011006   0.214    0.831
## I(Lat^2)     0.0143172  0.0076200   1.879    0.061 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1351 on 399 degrees of freedom
## Multiple R-squared:  0.1299, Adjusted R-squared:  0.1233
## F-statistic: 19.85 on 3 and 399 DF,  p-value: 5.194e-12
```

```
res_5 <- model_5$residuals
dist_matr4 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv4<- 1/dist_matr4
diag(dist_matr_inv4) <- 0
Moran.I(res_5, dist_matr_inv4)
```

```
## $observed
## [1] 0.03535691
##
```

```
## $expected
## [1] -0.002487562
##
## $sd
## [1] 0.004260021
##
## $p.value
## [1] 0
```

```
shapiro.test(res_5)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_5
## W = 0.9707, p-value = 3.099e-07
```

Tercer Modelo con ajustes para datos georreferenciados (5)

```
model_6 <- lm(K ~ Na + Long + Lat + I(Long**2) + I(Lat**2), data = PABLO)
summary(model_6)
```

```
##
## Call:
## lm(formula = K ~ Na + Long + Lat + I(Long^2) + I(Lat^2), data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27044 -0.08839 -0.01944  0.07835  0.49936
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.483e+04  1.559e+04   0.951  0.34214
## Na           3.434e-01  5.119e-02   6.708 6.82e-11 ***
## Long         4.176e+02  4.317e+02   0.967  0.33402
## Lat          7.081e+01  2.490e+01   2.844  0.00468 **
## I(Long^2)    2.883e+00  2.979e+00   0.968  0.33376
## I(Lat^2)    -4.274e+00  1.507e+00  -2.836  0.00480 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.134 on 397 degrees of freedom
## Multiple R-squared:  0.1478, Adjusted R-squared:  0.1371
## F-statistic: 13.77 on 5 and 397 DF,  p-value: 2.05e-12
```

```
res_6 <- model_6$residuals
dist_matr5 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv5<- 1/dist_matr5
diag(dist_matr_inv5) <- 0
Moran.I(res_6, dist_matr_inv5)
```

```
## $observed
## [1] 0.02585439
##
## $expected
## [1] -0.002487562
##
## $sd
## [1] 0.004258878
##
## $p.value
## [1] 2.836975e-11
```

```
shapiro.test(res_6)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_6
## W = 0.97042, p-value = 2.754e-07
```

Cuarto Modelo con ajustes para datos georreferenciados (6)

```
model_7 <- lm(K ~ Na + I(Long**2) + I(Lat**2) + I(Na**2)+ Long + Lat , data = PABLO)
summary(model_7)
```

```
##
## Call:
## lm(formula = K ~ Na + I(Long^2) + I(Lat^2) + I(Na^2) + Long +
##     Lat, data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27515 -0.08739 -0.02026  0.07826  0.49453
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13223.1636 15698.8317   0.842  0.40013
## Na              0.5068     0.1907   2.658  0.00818 **
## I(Long^2)       2.5744     3.0000   0.858  0.39134
## I(Lat^2)       -4.0511     1.5281  -2.651  0.00834 **
## I(Na^2)        -0.2206     0.2480  -0.890  0.37417
## Long          372.8637   434.7656   0.858  0.39162
## Lat            67.1257    25.2447   2.659  0.00816 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1341 on 396 degrees of freedom
## Multiple R-squared:  0.1495, Adjusted R-squared:  0.1366
## F-statistic:  11.6 on 6 and 396 DF,  p-value: 5.589e-12
```

```
res_7 <- model_7$residuals
dist_matr6 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv6<- 1/dist_matr6
diag(dist_matr_inv6) <- 0
Moran.I(res_7, dist_matr_inv6)
```

```
## $observed
## [1] 0.02559224
##
## $expected
## [1] -0.002487562
##
## $sd
## [1] 0.004258926
##
## $p.value
## [1] 4.305445e-11
```

```
shapiro.test(res_7)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_7
## W = 0.97185, p-value = 5.047e-07
```

Modelos de regresión espacial

```
xy = as.matrix(PABLO[,c(2,3)])
contnb <- dnearneigh(coordinates(xy),0,380000,longlat = F)
dlist <- nbdists(contnb, xy)
dlist <- lapply(dlist, function(x) 1/x)
Wve <- nb2listw(contnb,glist=dlist,style = "W")
```

Modelo autoregresivo puro (7)

```
model_auto <- spautolm(K ~ 1,data = PABLO,listw=Wve)
summary(model_auto)
```

```
##
## Call: spautolm(formula = K ~ 1, data = PABLO, listw = Wve)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.261475 -0.106784 -0.022284  0.075295  0.551268
##
## Coefficients:
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.20872    0.10283  2.0297  0.04239
##
## Lambda: 0.93328 LR test value: 31.281 p-value: 2.2325e-08
## Numerical Hessian standard error of lambda: 0.065062
##
## Log likelihood: 224.5228
## ML residual variance (sigma squared): 0.018969, (sigma: 0.13773)
## Number of observations: 403
## Number of parameters estimated: 3
## AIC: -443.05
```

```
res_8 <- model_auto$fit$residuals
dist_matr7 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv7<- 1/dist_matr7
diag(dist_matr_inv7) <- 0
Moran.I(res_8, dist_matr_inv7)
```

```
## $observed
## [1] 0.02593183
##
## $expected
## [1] -0.002487562
##
## $sd
## [1] 0.004258952
##
## $p.value
## [1] 2.50866e-11
```

```
shapiro.test(res_8)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_8
## W = 0.9606, p-value = 6.342e-09
```

```
Moran.I(res_8, dist_matr_inv7)
```

```
## $observed
## [1] 0.02593183
##
## $expected
## [1] -0.002487562
##
## $sd
## [1] 0.004258952
##
## $p.value
## [1] 2.50866e-11
```

$$Y_K = \alpha_0 + \lambda W Y_K + uu = \rho W u + \epsilon$$

Si $\rho = 0$, $\mathbf{u} = \epsilon$

$$Y_K = \alpha_0 + \lambda W Y_K + \epsilon$$

Ultimo intento con regresion multiple (8)

```
model_9 <- lm(K ~ Na + Mg + CICE + Fe, data = PABLO)
summary(model_9)
```

```
##
## Call:
## lm(formula = K ~ Na + Mg + CICE + Fe, data = PABLO)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29687 -0.07014 -0.01673  0.05653  0.56863
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.289e-02  1.734e-02   1.320  0.18750
## Na          1.441e-02  4.696e-02   0.307  0.75915
## Mg          5.808e-02  6.580e-03   8.827  < 2e-16 ***
## CICE        8.952e-03  1.565e-03   5.720  2.1e-08 ***
## Fe          5.469e-05  2.029e-05   2.696  0.00732 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1076 on 398 degrees of freedom
## Multiple R-squared:  0.4498, Adjusted R-squared:  0.4443
## F-statistic: 81.35 on 4 and 398 DF,  p-value: < 2.2e-16
```

```
res_9 <- model_9$residuals
dist_matr8 <- as.matrix(dist(cbind(x = PABLO$Long, y = PABLO$Lat)))
dist_matr_inv8<- 1/dist_matr8
diag(dist_matr_inv8) <- 0
shapiro.test(res_9)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_9
## W = 0.93569, p-value = 3.529e-12
```

```
Moran.I(res_9, dist_matr_inv8)
```

```
## $observed
## [1] 0.01172778
##
## $expected
```

```
## [1] -0.002487562
##
## $sd
## [1] 0.004243019
##
## $p.value
## [1] 0.0008072705
```