

Juntando todos los datos de todos los años, se obtiene informacion relevante sobre la base de datos

Estructura:

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## 'data.frame':   11966 obs. of  15 variables:
##  $ entidad: int  1 1 1 1 1 1 1 1 1 1 1 ...
##  $ edad   : int  52 75 67 65 66 39 47 42 67 45 ...
##  $ sexo   : int  2 1 1 2 1 1 2 1 1 1 1 ...
##  $ nivel  : int  3 3 2 2 6 3 3 2 2 2 2 ...
##  $ p34_1  : int  2 2 2 2 3 1 1 3 3 3 3 ...
##  $ p34_2  : int  2 1 1 1 3 1 1 3 3 1 1 ...
##  $ p34_3  : int  2 2 2 1 3 1 1 3 3 1 1 ...
##  $ p34_4  : int  2 1 1 1 1 1 1 1 1 3 1 ...
##  $ p2     : int  2 2 2 2 2 1 1 1 2 1 1 ...
##  $ p4     : int  0 0 0 0 0 0 1 0 0 0 0 ...
##  $ p5     : int  0 0 0 0 0 0 4 0 0 0 0 ...
##  $ p10    : int  0 0 0 0 0 2 0 0 0 0 0 ...
##  $ p11    : int  0 0 0 0 0 4 0 0 0 0 0 ...
##  $ p16    : int  0 0 1 0 0 1 0 3 0 2 2 ...
##  $ p17    : int  0 0 4 0 0 4 0 3 0 3 3 ...
```

Dimensión del data frame

Para encontrar la cantidad de columnas y renglones del data frame se usan las funciones de ncol y nrow

```
## [1] 15
## [1] 11966
```

Datos faltantes

Variables seleccionadas:

```
##   entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2 p4 p5 p10 p11 p16 p17
## 1      1   52    2     3     2     2     2     2  2  0  0  0  0  0  0
## 2      1   75    1     3     2     1     2     1  2  0  0  0  0  0  0
## 3      1   67    1     2     2     1     2     1  2  0  0  0  0  1  4
## 4      1   65    2     2     2     1     1     1  2  0  0  0  0  0  0
## 5      1   66    1     6     3     3     3     1  2  0  0  0  0  0  0
## 6      1   39    1     3     1     1     1     1  1  0  0  2  4  1  4
```

Analizando el sumario, no se notan discrepancias en los valores

```
##      entidad      edad      sexo      nivel
## Min.   : 1.00   Min.   :18.00   Min.   :1.000   Min.   : 0.000
```

```
## 1st Qu.: 9.00 1st Qu.:32.00 1st Qu.:1.000 1st Qu.: 3.000
## Median :15.00 Median :44.00 Median :2.000 Median : 4.000
## Mean :15.63 Mean :45.66 Mean :1.563 Mean : 4.581
## 3rd Qu.:20.00 3rd Qu.:58.00 3rd Qu.:2.000 3rd Qu.: 7.000
## Max. :32.00 Max. :97.00 Max. :2.000 Max. :99.000
## p34_1 p34_2 p34_3 p34_4 p2
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.00
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.00
## Median :2.000 Median :1.000 Median :2.000 Median :1.000 Median :1.00
## Mean :1.684 Mean :1.475 Mean :1.636 Mean :1.417 Mean :1.38
## 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.00
## Max. :3.000 Max. :3.000 Max. :3.000 Max. :3.000 Max. :2.00
## p4 p5 p10 p11
## Min. : 0.000 Min. :0.000 Min. : 0.0000 Min. :0.0000
## 1st Qu.: 0.000 1st Qu.:0.000 1st Qu.: 0.0000 1st Qu.:0.0000
## Median : 0.000 Median :0.000 Median : 0.0000 Median :0.0000
## Mean : 1.435 Mean :1.367 Mean : 0.9434 Mean :0.9041
## 3rd Qu.: 2.000 3rd Qu.:3.000 3rd Qu.: 1.0000 3rd Qu.:1.0000
## Max. :99.000 Max. :6.000 Max. :99.0000 Max. :6.0000
## p16 p17
## Min. : 0.0000 Min. :0.0000
## 1st Qu.: 0.0000 1st Qu.:0.0000
## Median : 0.0000 Median :0.0000
## Mean : 0.8276 Mean :0.8097
## 3rd Qu.: 0.0000 3rd Qu.:0.0000
## Max. :80.0000 Max. :6.0000
```

No hay valores vacios

```
## entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2 p4
## 0 0 0 0 0 0 0 0 0 0
## p5 p10 p11 p16 p17
## 0 0 0 0 0
```

Chequeo de rango de los datos cuantitativos

Edad:

```
## [1] entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2
## [10] p4 p5 p10 p11 p16 p17
## <0 rows> (or 0-length row.names)

## [1] entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2
## [10] p4 p5 p10 p11 p16 p17
## <0 rows> (or 0-length row.names)
```

Cantidad de libros leidos:

```
## [1] entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2
## [10] p4 p5 p10 p11 p16 p17
## <0 rows> (or 0-length row.names)

## entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2 p4 p5 p10 p11 p16 p17
## 6634 12 23 1 7 2 2 1 1 1 99 3 0 0 0 0
## 9105 16 42 2 9 2 1 1 1 1 80 1 2 4 0 0
```

Revistas:

```
## [1] entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2
```

```
## [10] p4      p5      p10     p11     p16     p17
## <0 rows> (or 0-length row.names)

##      entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2 p4 p5 p10 p11 p16 p17
## 97         3  41   2    9     1     1     2     1  1  5  2  50  2  3  3
## 225        6  49   1    7     1     1     1     1  1  7  3  75  2  4  3
## 298        8  42   2    9     1     1     2     1  1  3  1  90  1  1  4
## 857       14  31   2    4     1     1     2     1  1  1  4  30  3  0  0
## 894       14  22   1    7     2     1     1     1  1  4  2  60  3  5  3
## 1032      15  52   1    7     1     2     2     1  1  8  4  30  3  0  0
## 1101      16  43   1    4     2     1     1     1  2  0  0  99  4  2  4
## 1256      19  70   1    6     2     2     2     2  1  3  3  30  4  1  3
## 2171       5  31   1    7     2     1     1     1  1 12  1  90  1  7  4
## 4127       4  48   1    3     2     1     2     1  1  0  0  30  3  0  0
## 4341       9  30   1    7     1     1     2     1  1  4  1  60  1  3  3
## 4588      11  22   1    7     1     1     1     1  1  0  0  30  1  0  0
## 5047      15  53   2    3     1     1     1     1  1 10  4  30  4  5  3
## 6175       6  68   2    4     1     1     1     1  1  1  4  60  4  0  0
## 6179       6  34   1    6     1     1     1     1  1  3  1  60  1  0  0
## 6223       7  24   1    7     2     1     1     1  1  8  2  37  2  0  0
## 6659      13  57   2    3     3     3     3     3  1  2  5  40  5  0  0
## 6742      14  72   1    7     2     1     2     1  1  4  3  50  2  0  0
## 7234      19  50   2    6     2     2     2     2  1 12  1  30  1  0  0
## 8357       9  34   2    7     1     1     2     1  1  1  1  60  1  0  0
## 8767      14  42   1    7     2     1     2     1  1 30  4  70  4  0  0
```

Periodicos:

```
## [1] entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2
## [10] p4      p5      p10     p11     p16     p17
## <0 rows> (or 0-length row.names)

##      entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2 p4 p5 p10 p11 p16 p17
## 3731      26  72   1    7     2     1     1     1  1  5  3  15  3  80  3
```

Conclusion: no hay discrepancias en la base de datos que podrian afectar el analisis

Prepararemos los datos para tener un mejor analisis, esto requiere que se combinen las columnas de la cantidad de libros, revistas y periodicos leidos en una sola columna mas general:

```
##      entidad edad sexo nivel p34_1 p34_2 p34_3 p34_4 p2 p4 p5 p10 p11 p16 p17 TL
## 1         1  52   2    3     2     2     2     2  2  0  0  0  0  0  0  0
## 2         1  75   1    3     2     1     2     1  2  0  0  0  0  0  0  0
## 3         1  67   1    2     2     1     2     1  2  0  0  0  0  1  4  1
## 4         1  65   2    2     2     1     1     1  2  0  0  0  0  0  0  0
## 5         1  66   1    6     3     3     3     1  2  0  0  0  0  0  0  0
## 6         1  39   1    3     1     1     1     1  1  0  0  2  4  1  4  3
```

Descripción de variables

Exploracion de datos

Calculo de medias estadisticas

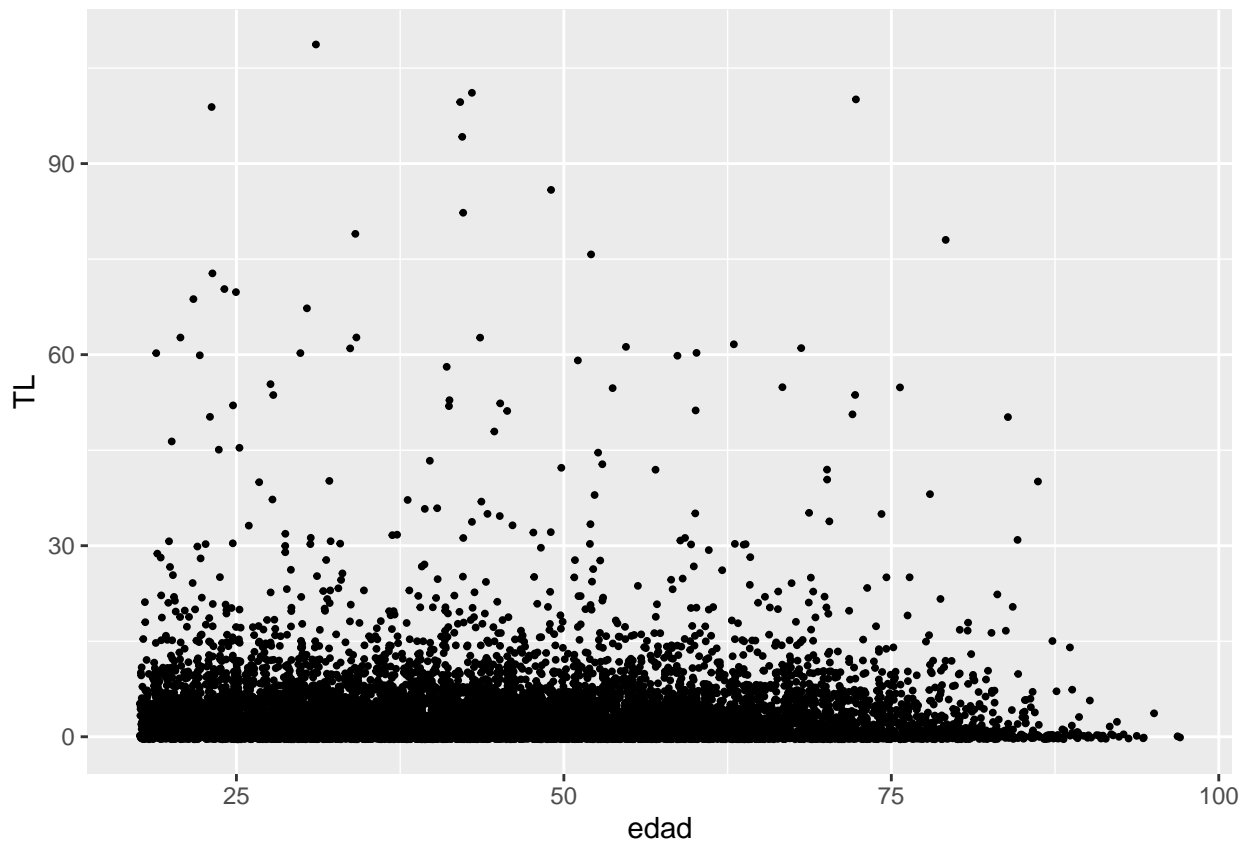
Variables cuantitativas

Variables cualitativas

Herramientas de visualizacion

Variables cuantitativas

Grafica de dispersion: Textos leidos vs edad



Gran cantidad de valores atipicos, mayor densidad cerca del 0, se nota una baja o nula coorrelacion debido a que es constante cerca del cero

Grafica de barras: Textos leidos vs cantidad de gente

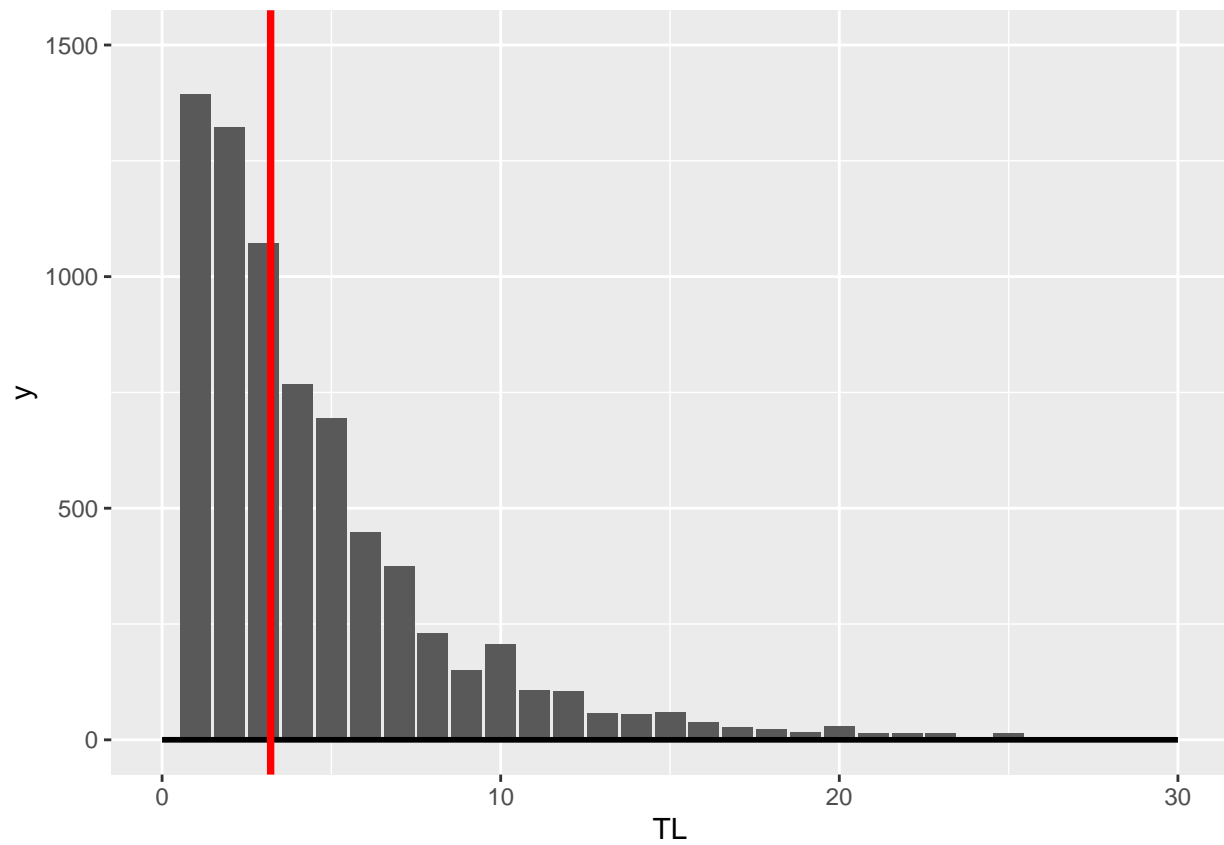
```
## Warning: `geom_vline()`: Ignoring `mapping` because `xintercept` was provided.
```

```
## Warning: Removed 88 rows containing non-finite outside the scale range
```

```
## (`stat_count()`).
```

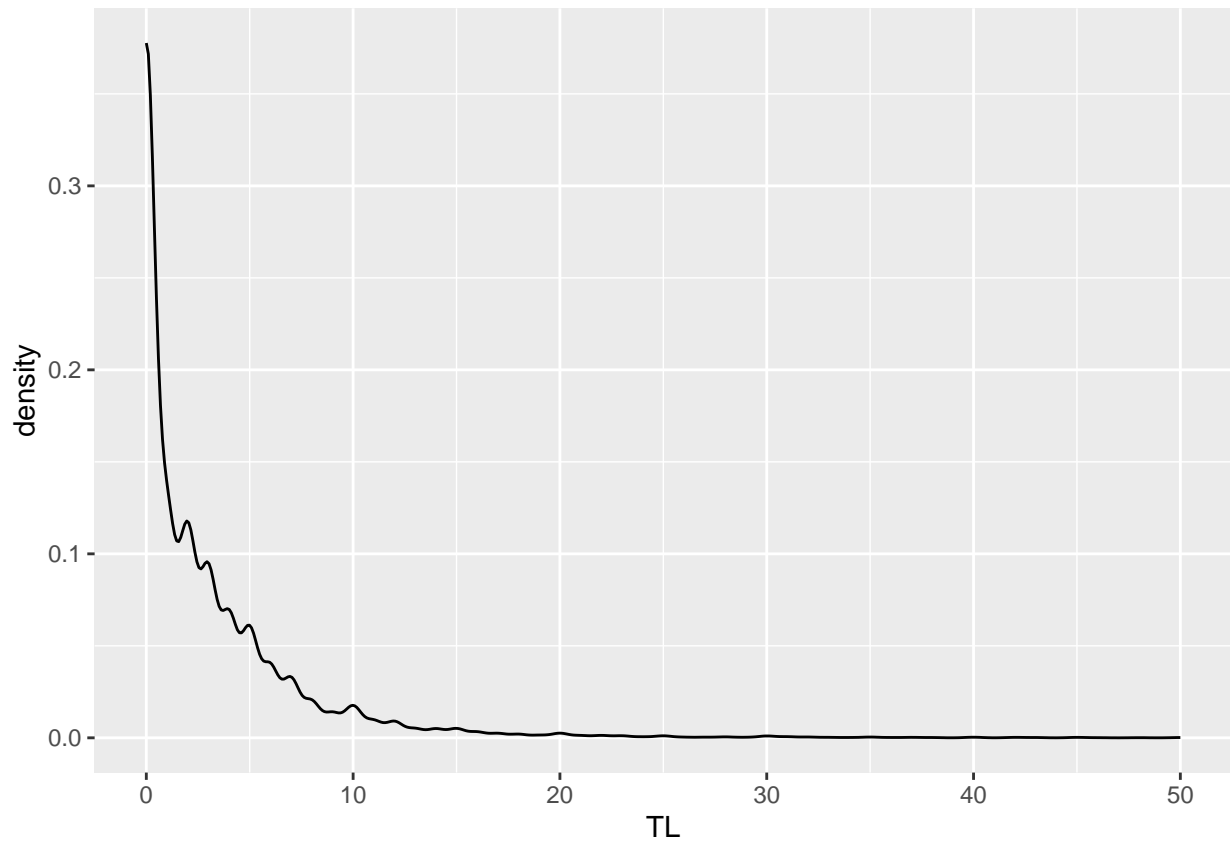
```
## Warning: Removed 2 rows containing missing values or values outside the scale range
```

```
## (`geom_bar()`).
```



Densidad de cantidad de libros leídos

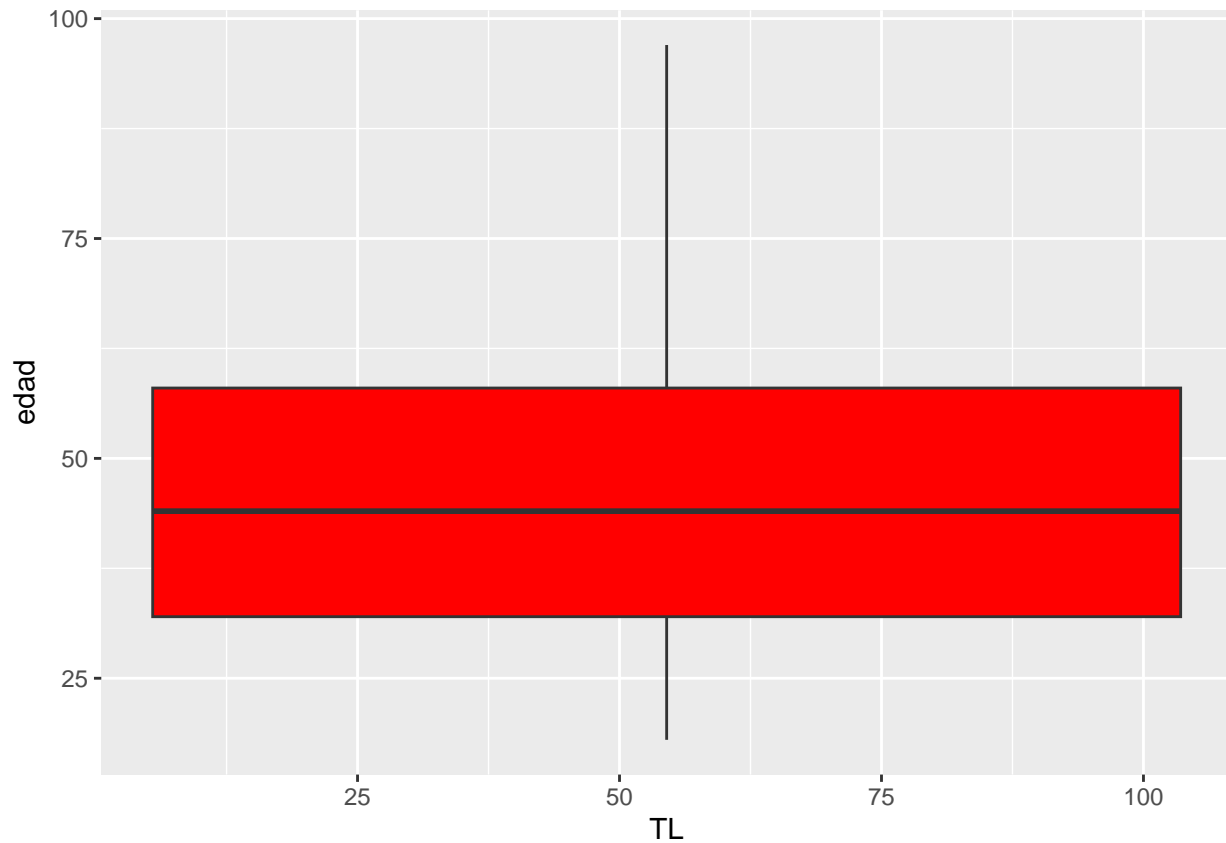
```
## Warning: Removed 43 rows containing non-finite outside the scale range  
## (`stat_density()`).
```



Distribucion asimetrica con sesgo a la izquierda

Boxplot:

```
## Warning: Continuous x aesthetic  
## i did you forget `aes(group = ...)`?
```



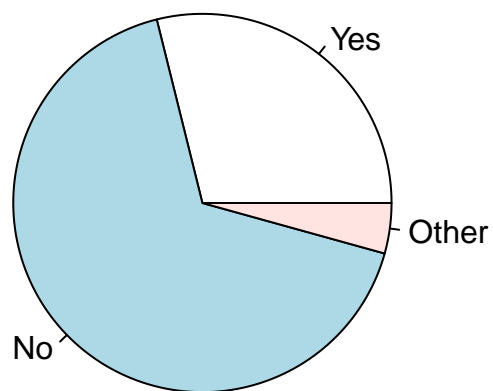
Análisis de correlacion

Como ya se vio en las graficas anteriores, la probabilidad que haya una correlacion entre la edad y la cantidad de textos leídos de una persona es muy baja. Tambien se puede calcular con la funcion `cor()`, con esta se observa que la correlacion es virtualmente nula.

```
## [1] -0.05596988
```

Variables categoricas

Cantidad de personas que fueron a una libreria en su infancia



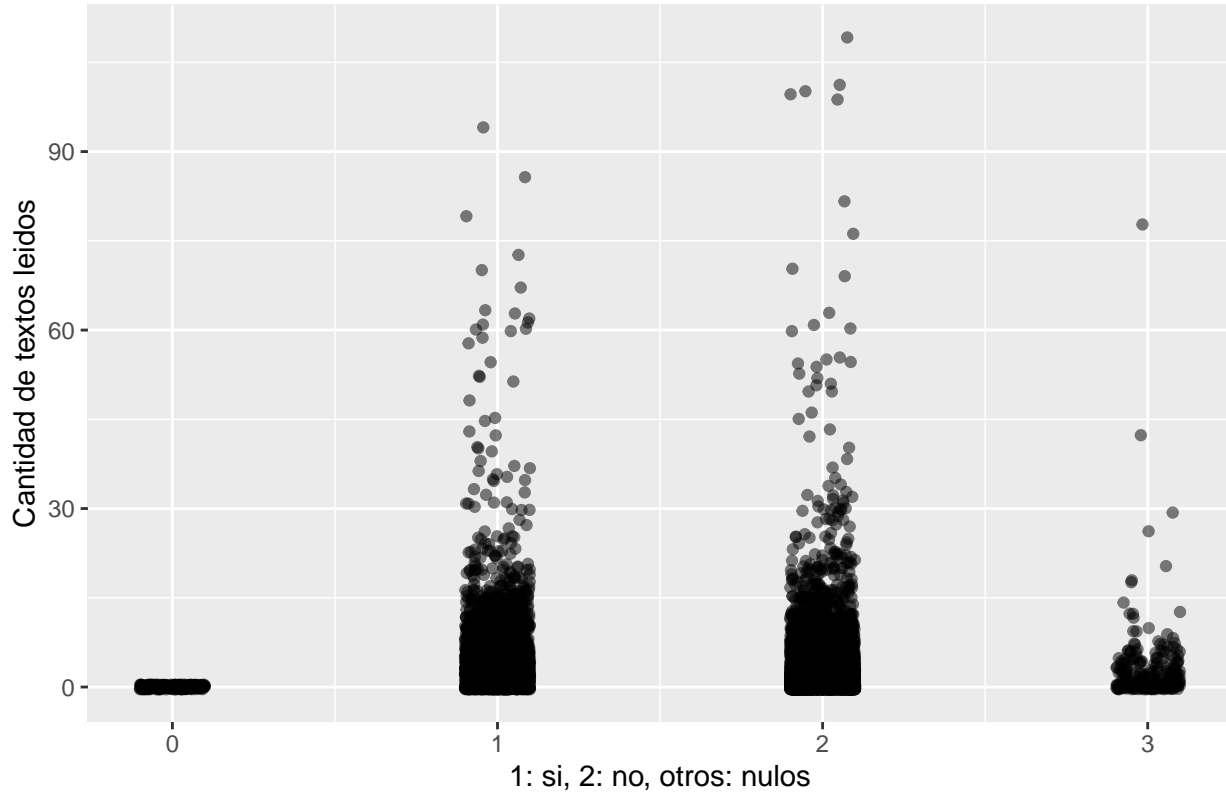
Nuestro objetivo es analizar como se relacionan las tendencias de lectura con variables categoricas como la manera en la que desarrollaron su infancia y educacion.

A continuacion se realizan graficos de dispersion relacionando las variables categoricas con la cantidad de textos leidos

```
## Warning: Use of `df_c$p34_1` is discouraged.  
## i Use `p34_1` instead.
```

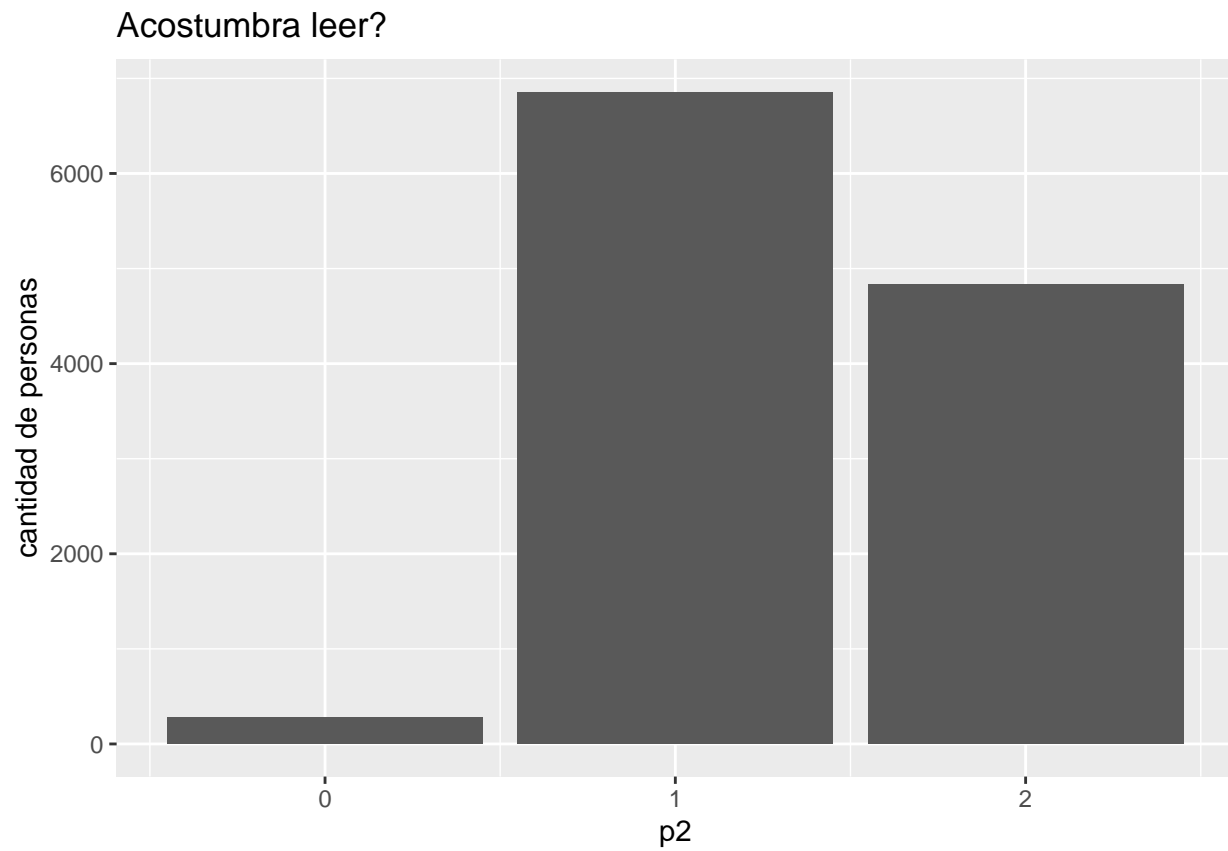
```
## Warning: Use of `df_c$TL` is discouraged.  
## i Use `TL` instead.
```

lo llevaban a bibliotecas o librerias?



Porcentaje de personas que acostumbran leer

```
## Ignoring unknown labels:  
## * xv : "1: si, 2: no, 0: nulo"
```

Vinculacion de datos categoricos

Medidas de variables cuantitativas

Graficos cuantitativos

Graficos comparativos