Transformación de Datos

Gener Aviles-R

March 14, 2017

Considering the figure, please answer the following questions:

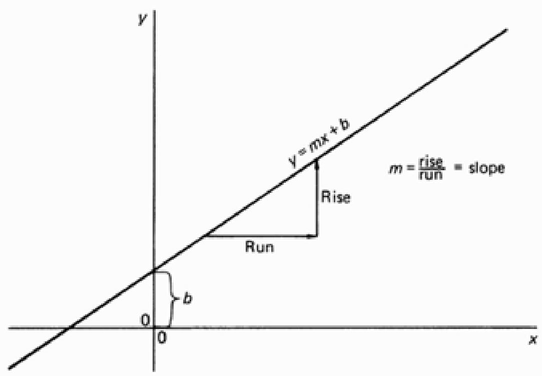


Figure 3-3 Graph of the equation $y = mx + b$.

1. Which of the following formulas define y as a linear function of x (m and b are constants)?

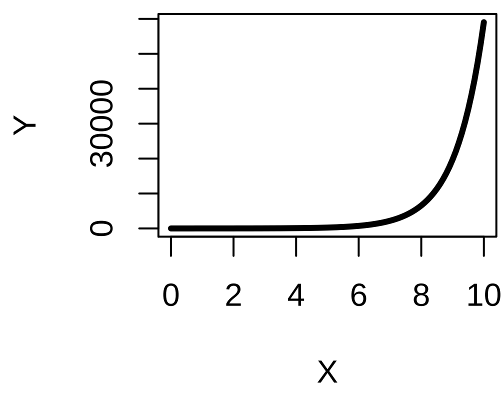
(a) $y = 3^x + 2$ (b) $y = \frac{1}{2}x - 5$ (c) $y = \dfrac{m}{x} + b$

(d) $y = m + x + b$ (e) $y = 3x^3 + 2$ (f) $y = \pi 10^x$

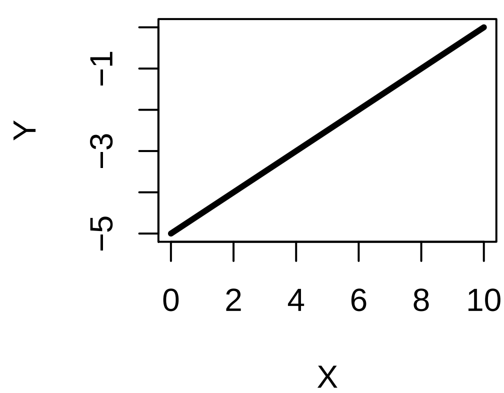
(g) $y = (x^2 - 2x)/(\sqrt{x})^2$

Graficación de las funciones

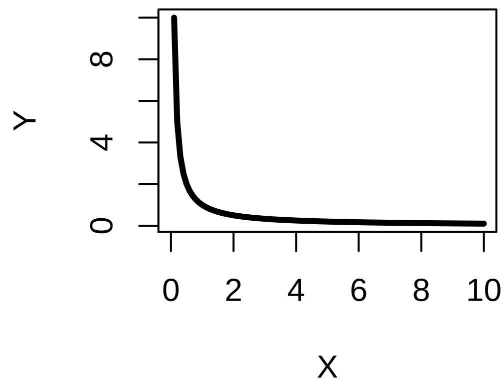
(A): y = 3^x + 2



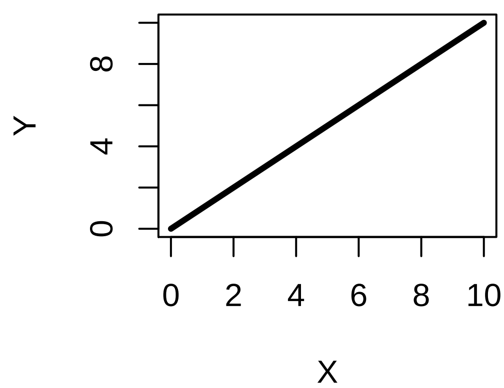
(B): y = 1/2X − 5

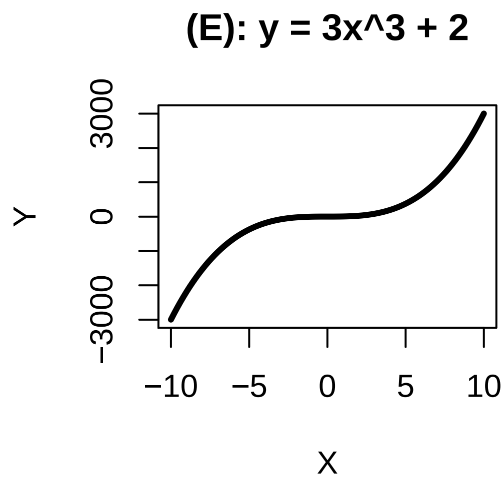


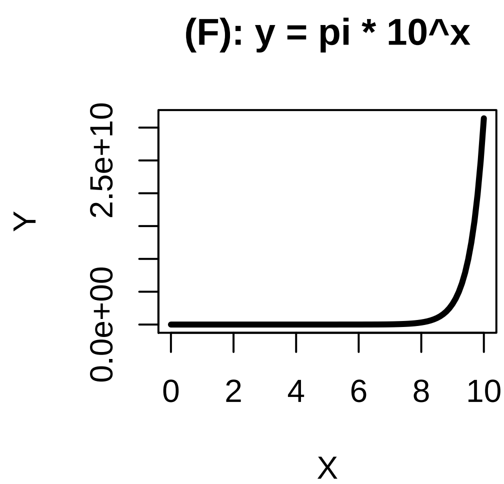
(C): y = m/x − b

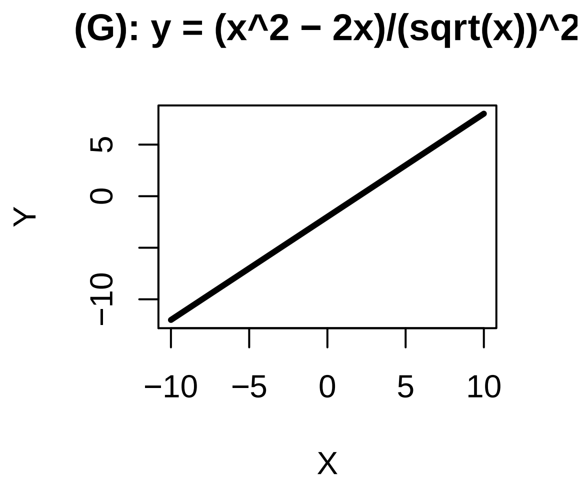


(D): y = m + x +b




(E): y = 3x^3 + 2
Y
3000
0
−3000
−10
−5
0
5
10
X


(F): y = pi * 10^x
Y
2.5e+10
0.0e+00
0
2
4
6
8
10
X



Respuesta

Basados en las graficaciones podemos ver que las ecuaciones lineales son las de los incisos:

- B
- D
- G

2. A straight line passes through the points $(3, 5)$ and $(4, 7)$. use these points to find a rise and corresponding run, the slope of the line, and the y intercept b.

Para resolver éste ejercicio se utilizará la aproximación de mínimos cuadrados, sobre la que se basa la regresión lineal. La función `lm()` en R hace éste cálculo a partir de un arreglo de datos que se le entregan, en éste caso los datos de las dos coordenadas del ejercicio.

```
###PENDIENTE AGREGAR LA TEORIA Y MATEMATICA DE LOS MINIMOS CUADRADOS

x_lm <- c(3,5)
y_lm <- c(4,7)

fit <- lm(x_lm ~ y_lm)
```

Mostrando los atributos del objeto `fit`, que es donde se guardó el modelo, vemos lo siguiente:

```
fit
```

```
##
## Call:
## lm(formula = x_lm ~ y_lm)
##
## Coefficients:
## (Intercept)         y_lm
##      0.3333       0.6667
```

Por lo tanto:

- y *intercept* $b = 0.3333$

- *Slope of the line* = 0.6667
- *Rise and corresponding run* = ??????????

For convenience in distinguishing the dependent variable (usually called y) in the following problems, we shall use the capital letter B, C, D, E and F instead of y.

In each of these problems a precise relation exists between two variables. Find the relation. Trnasform the x variable in each problem to get the points on a straight line.

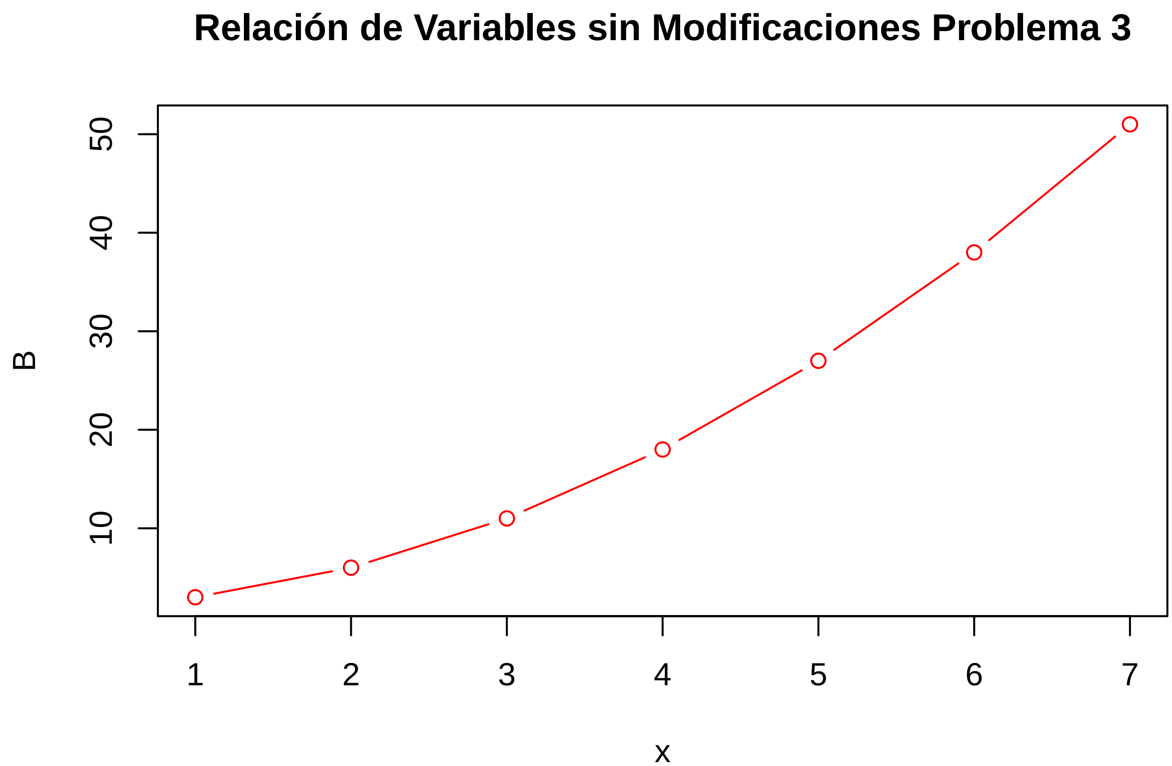
Problema 3:

B	3	6	11	18	27	38	51
x	1	2	3	4	5	6	7

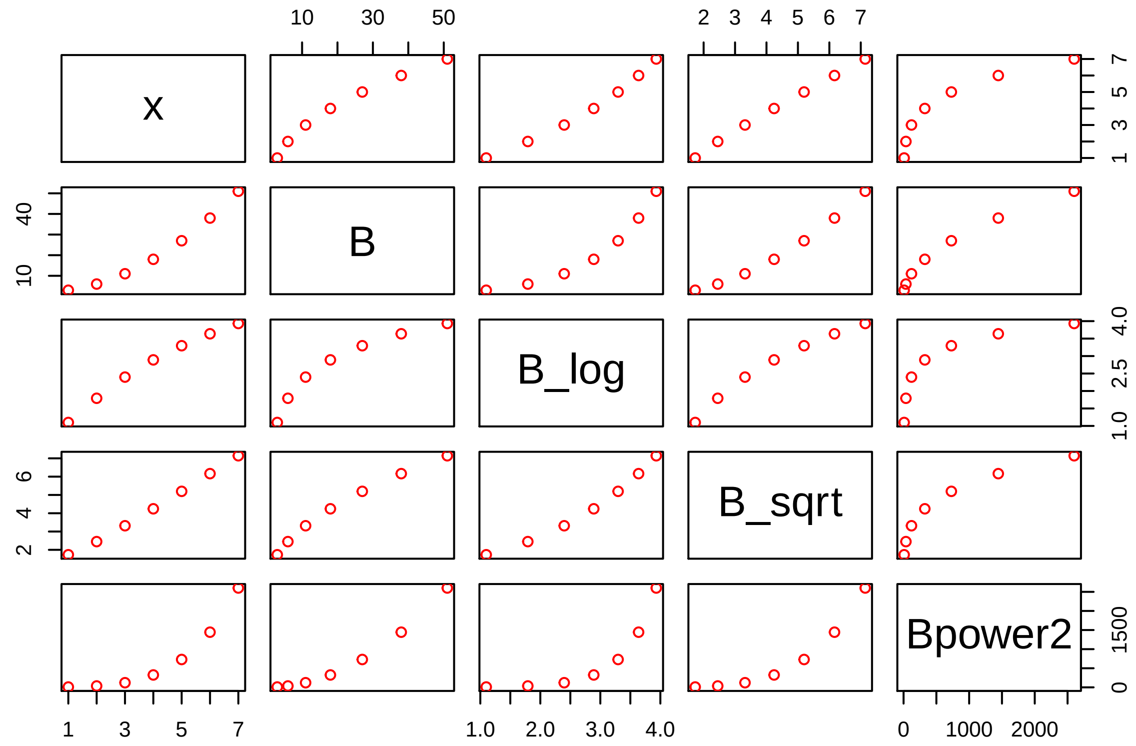
Relación inicial de variables:

Creando los objetos necesarios en R y visualización inicial:

```
B <- c(3,6,11,18,27,38,51)
x <- c(1,2,3,4,5,6,7)

plot(x, B,
     main = "Relación de Variables sin Modificaciones Problema 3",
     type = "b",
     col = "red")
```



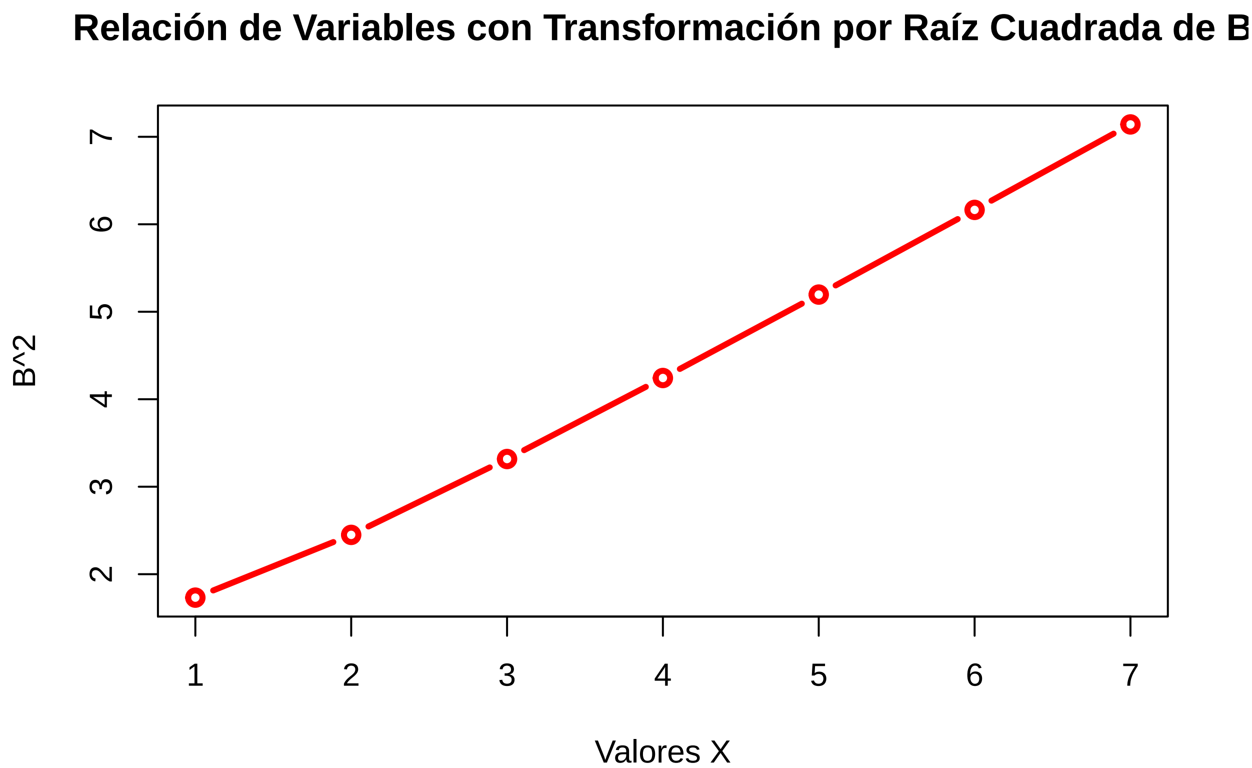
P3 Exploración de Transformaciones de Datos más Comunes



P3 Respuesta óptima

Para éste problema, al reducción por **raíz cuadrada** produce el mejor resultado. Entonces desarrollamos una gráfica solo de ésta transformación.

```
plot(x, B_sqrt,
     type = "b",
     lwd = 3,
     col = "red",
     main = "Relación de Variables con Transformación por Raíz Cuadrada de B",
     xlab = "Valores X",
     ylab = "B^2")
```

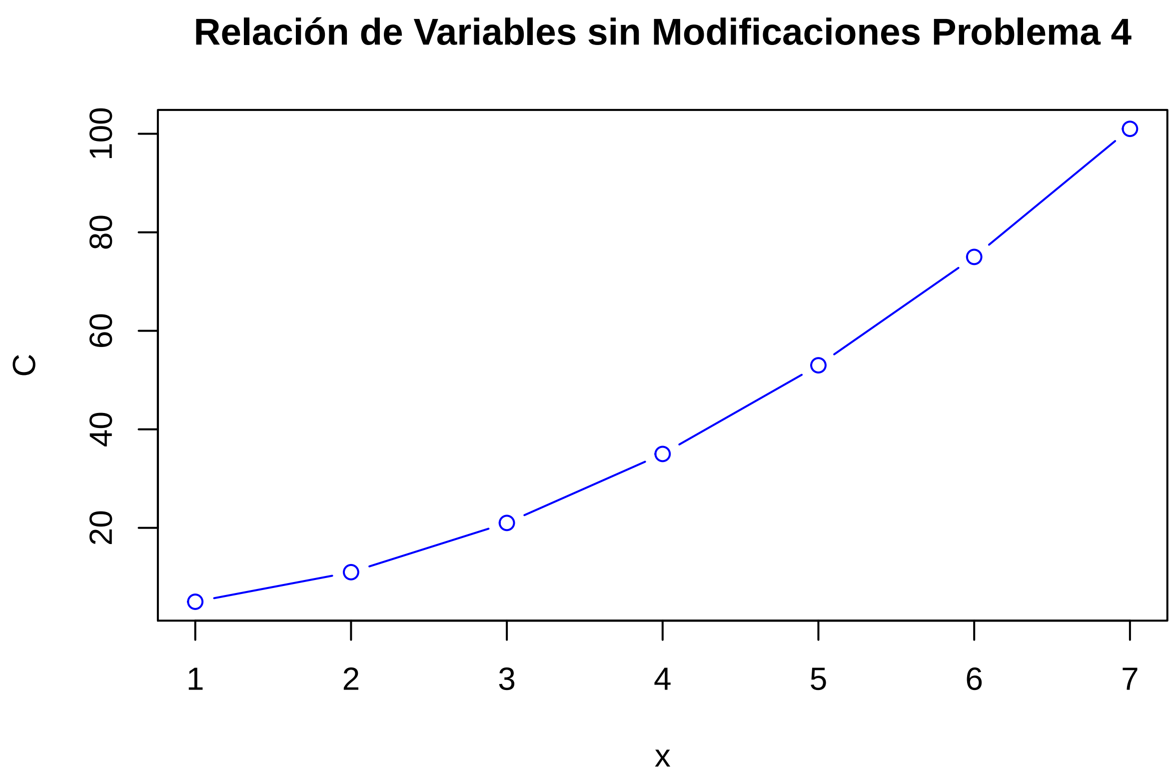


Problema 4

C	5	11	21	35	53	75	101
x	1	2	3	4	5	6	7

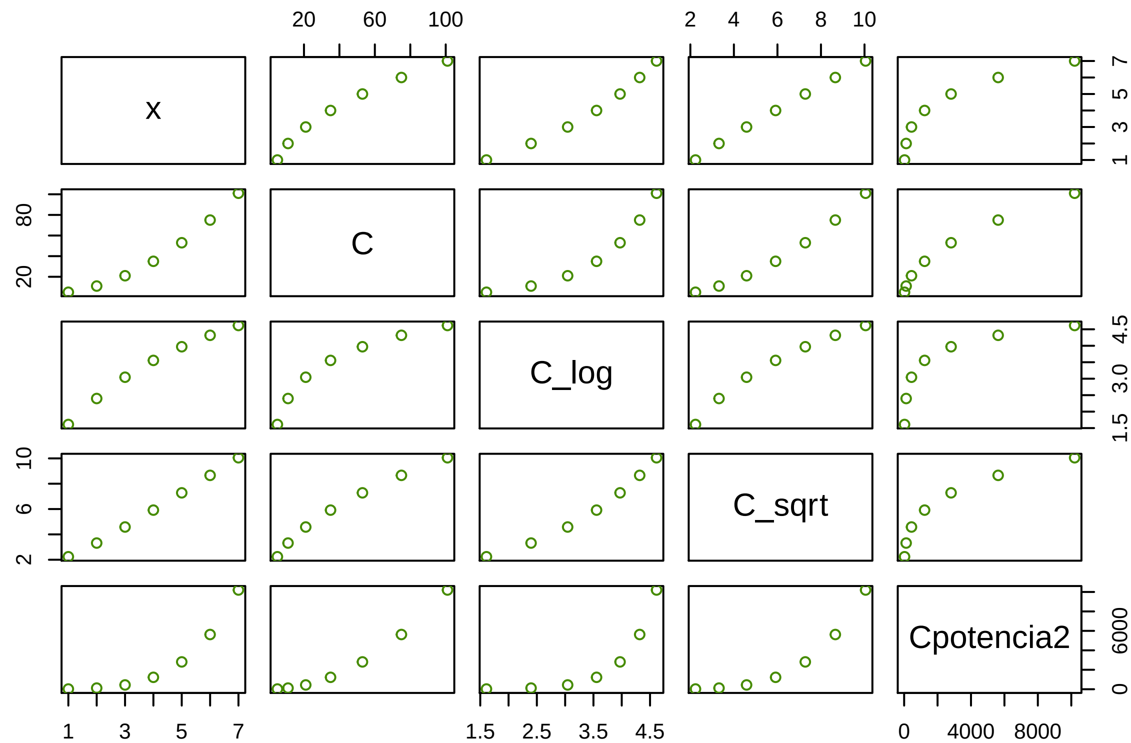
P4 Creando los objetos necesarios en R y visualización inicial:

```
C <- c(5,11,21,35,53,75,101)
x <- c(1,2,3,4,5,6,7)

cuatro <- data.frame(C,x)

plot(x, C,
     main = "Relación de Variables sin Modificaciones Problema 4",
     type = "b",
     col = "blue")
```



P4 Exploración de Transformaciones de Datos más Comunes

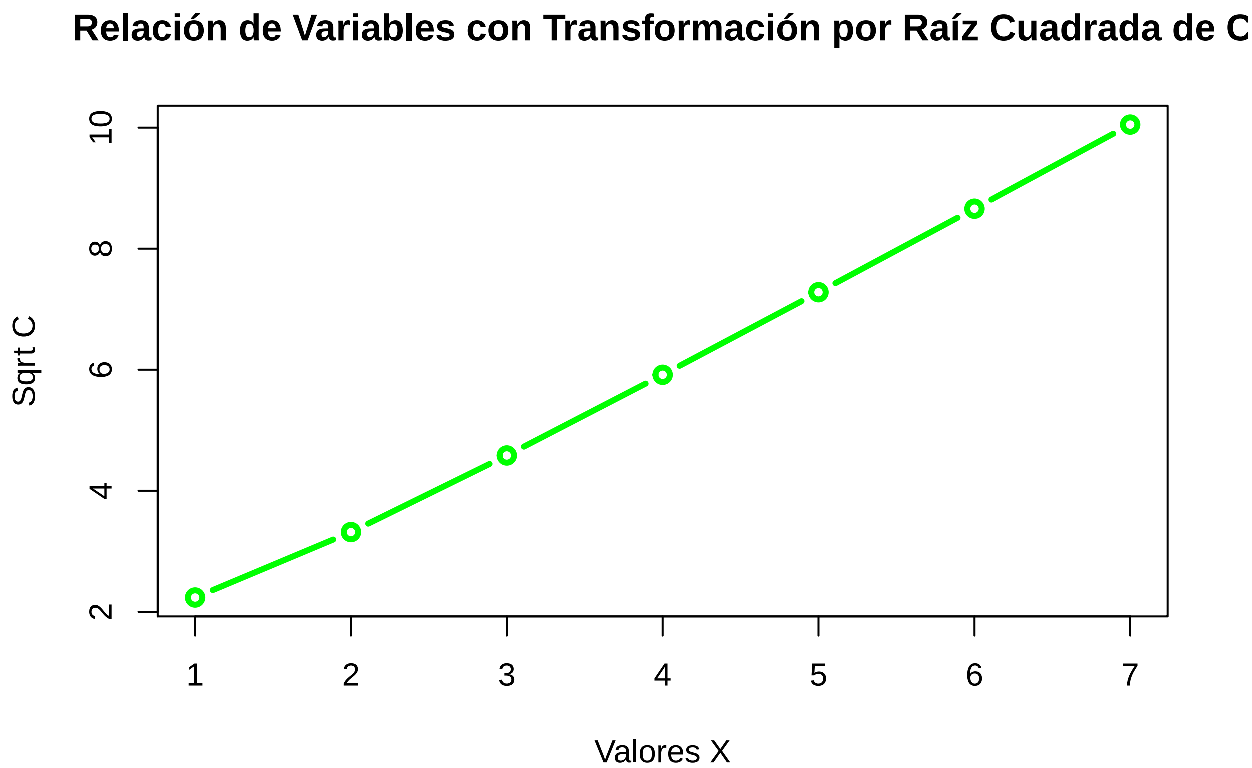
```
##Raíz Cuadrada
C_sqrt <- sqrt(C)

#plot(x, B_sqrt, main = "Transformación con Raíz Cuadrada")
#lines(x,B_sqrt, type = "l")

##Potencia al cuadrado
Cpotencia2 <- C^2
#plot(x, Bpower2, main = "Transformación con Poder a la 2")
#lines(x,Bpower2, type = "l")

##Logaritmo
C_log <- log(C)
#plot(x, Bpower2, main = "Transformación por Logaritmos Naturales")
#lines(x,Bpower2, type = "l")

##Añadiendo todos los valores a una sola tabla
cuatro <- data.frame(x, C, C_log, C_sqrt, Cpotencia2)

##Una gráfica inicial
plot(cuatro,
     col = "chartreuse4")
```



P4 Respuesta óptima

Para éste problema, al reducción por **raíz cuadrada** produce el mejor resultado. Entonces desarrollamos una gráfica solo de ésta transformación.

```
plot(x, C_sqrt,
     type = "b",
     lwd = 3,
     col = "green",
     main = "Relación de Variables con Transformación por Raíz Cuadrada de C",
     xlab = "Valores X",
     ylab = "Sqrt C")
```

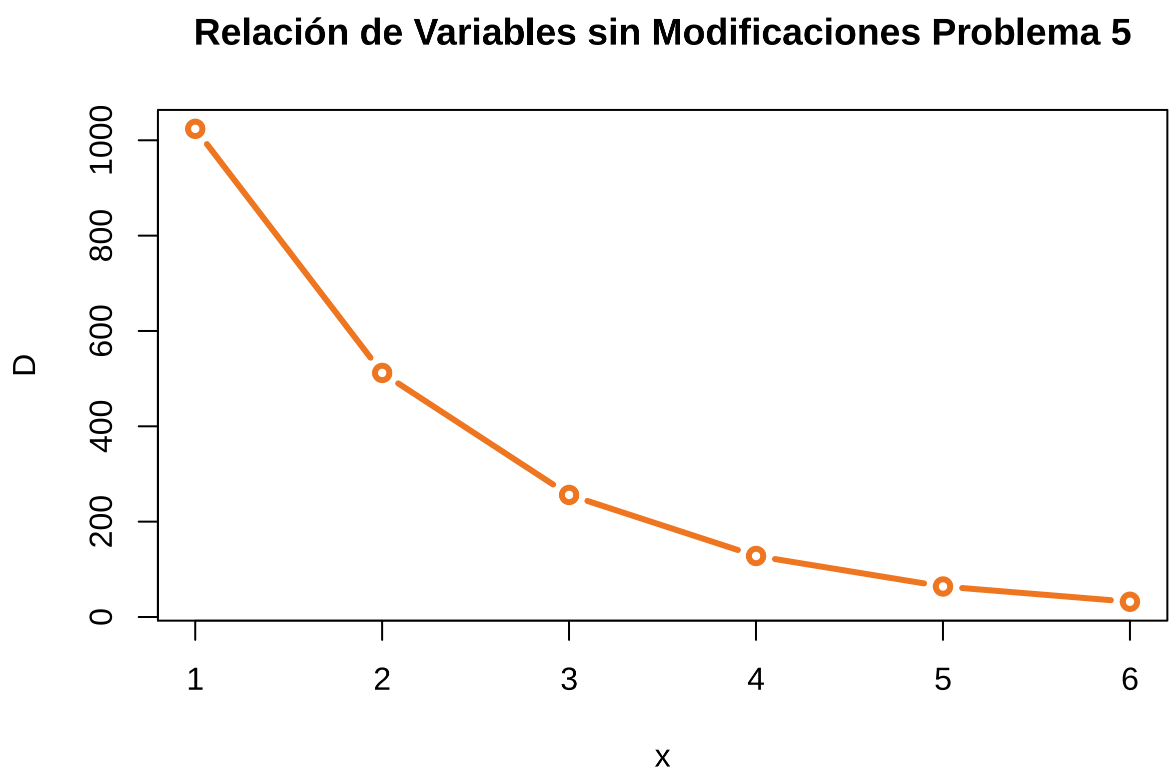


Problema 5

D	1024	512	256	128	64	32
x	1	2	3	4	5	6

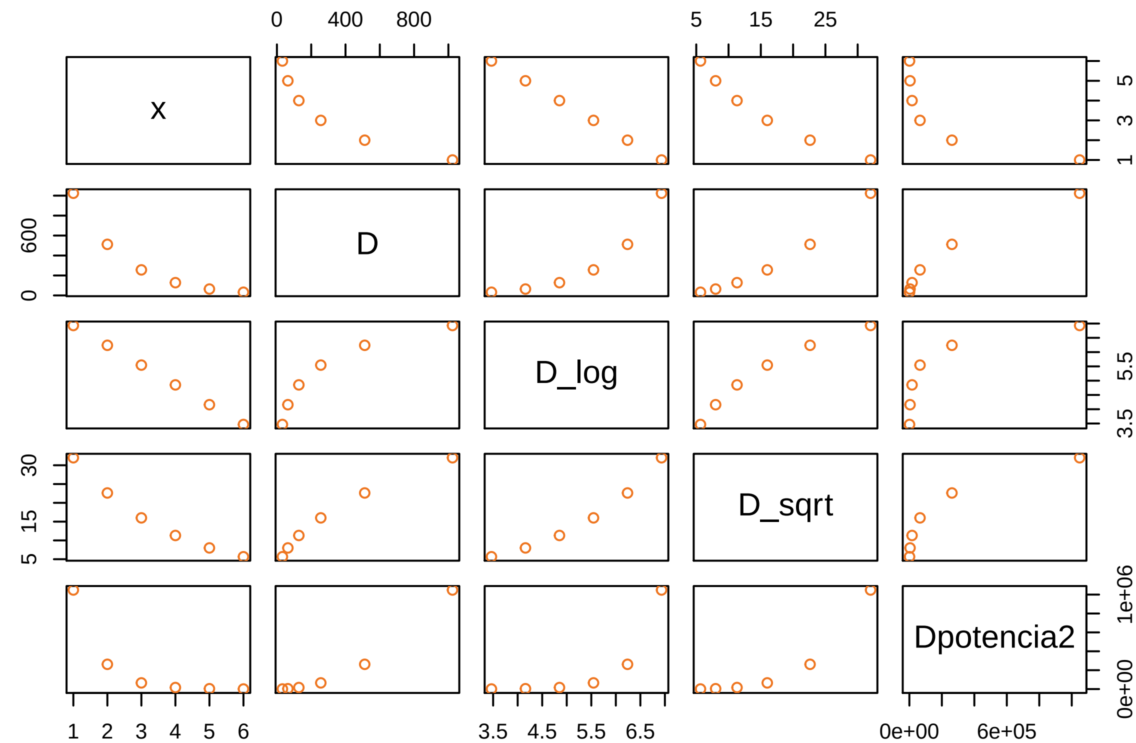
P5 Creando los objetos necesarios en R y visualización inicial:

```
D <- c(1024, 512, 256, 128, 64, 32)
x <- c(1,2,3,4,5,6)

cinco <- data.frame(D,x)

plot(x, D,
     type = "b",
     lwd = 3,
     col = "chocolate2",
     main = "Relación de Variables sin Modificaciones Problema 5")
```



P5 Exploración de Transformaciones de Datos más Comunes

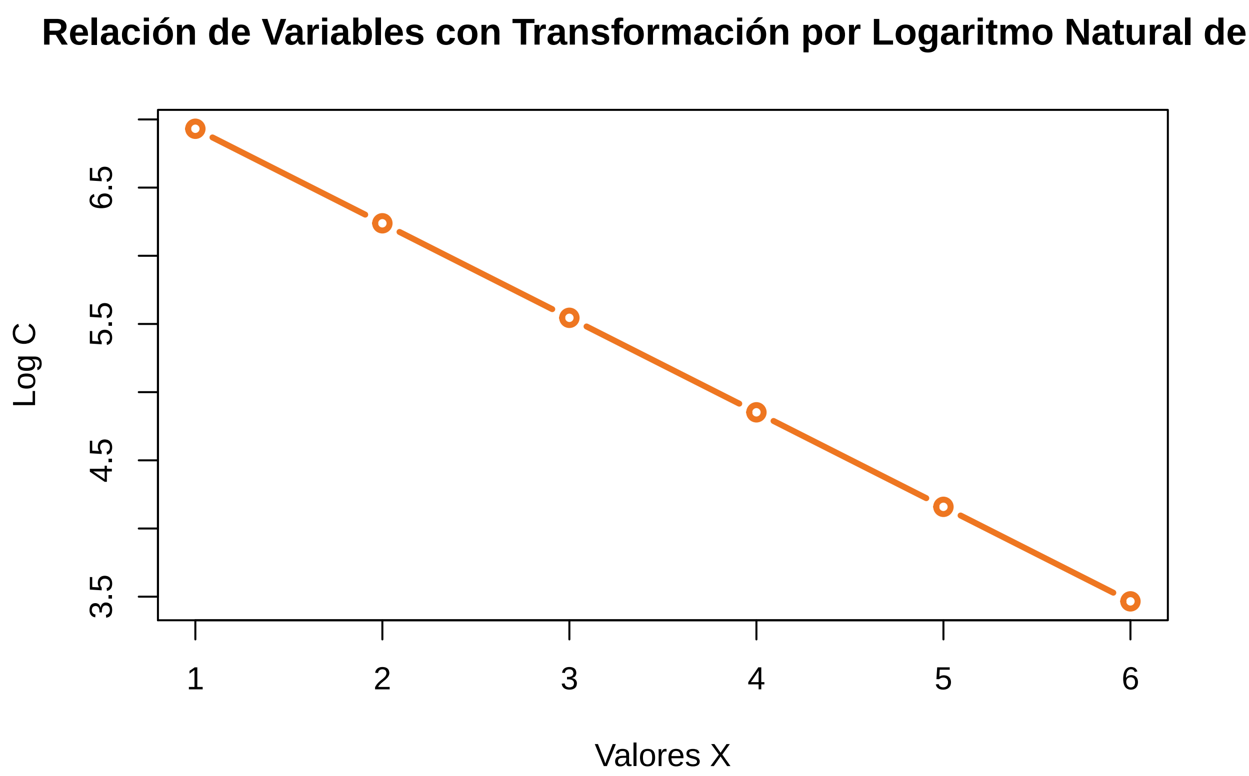
```
##Raíz Cuadrada
D_sqrt <- sqrt(D)

#plot(x, B_sqrt, main = "Transformación con Raíz Cuadrada")
#lines(x,B_sqrt, type = "l")

##Potencia al cuadrado
Dpotencia2 <- D^2
#plot(x, Bpower2, main = "Transformación con Poder a la 2")
#lines(x,Bpower2, type = "l")

##Logaritmo
D_log <- log(D)
#plot(x, Bpower2, main = "Transformación por Logaritmos Naturales")
#lines(x,Bpower2, type = "l")

##Añadiendo todos los valores a una sola tabla
cinco <- data.frame(x, D, D_log, D_sqrt, Dpotencia2)

##Una gráfica inicial
plot(cinco,
     col = "chocolate2")
```



P5 Respuesta óptima

Para éste problema, al reducción por **logaritmo natural** produce el mejor resultado. Entonces desarrollamos una gráfica solo de ésta transformación.

```
plot(x, D_log,
     type = "b",
     lwd = 3,
     col = "chocolate2",
     main = "Relación de Variables con Transformación por Logaritmo Natural de D",
     xlab = "Valores X",
     ylab = "Log C")
```

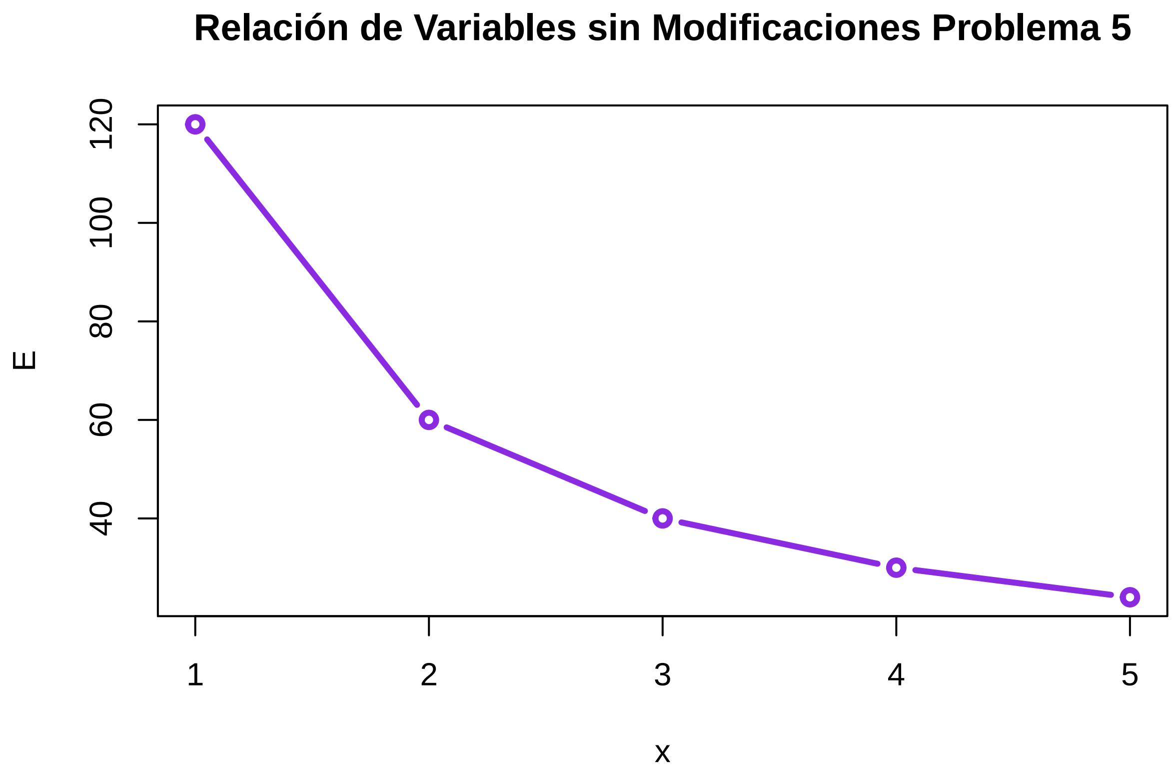


Problema 6

E	120	60	40	30	24
x	1	2	3	4	5

P6 Creando los objetos necesarios en R y visualización inicial:

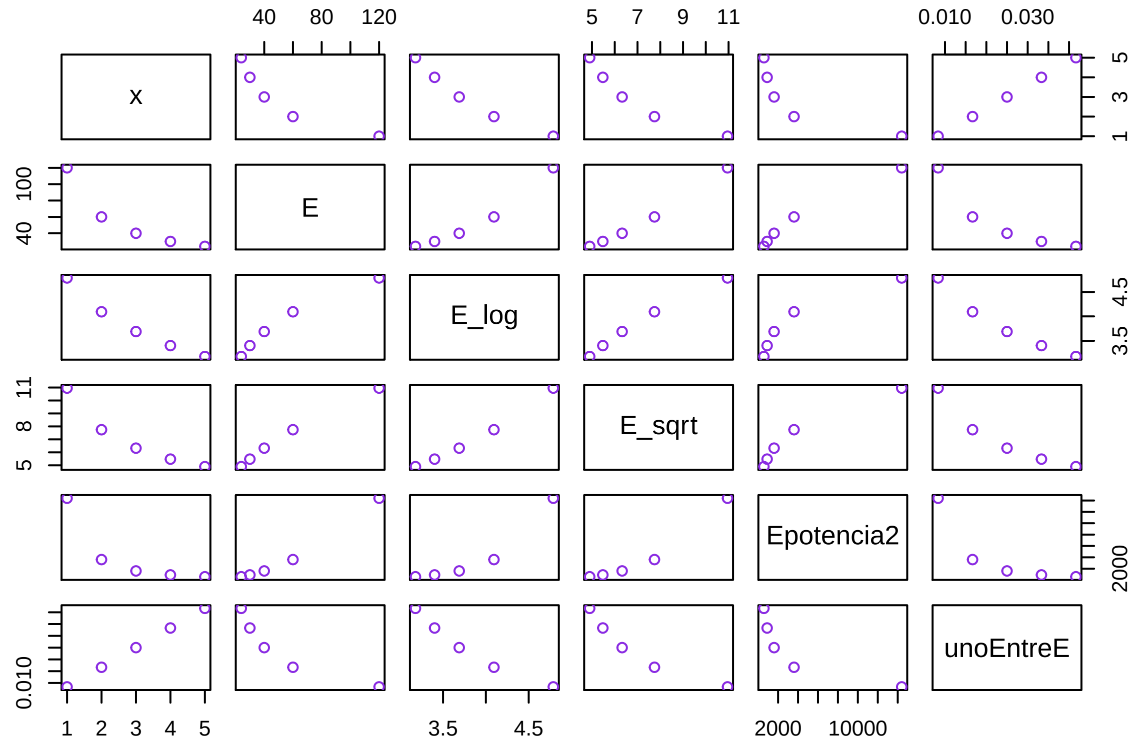
```
E <- c(120,60,40,30,24)
x <- c(1,2,3,4,5)

cuatro <- data.frame(E,x)

plot(x, E,
     main = "Relación de Variables sin Modificaciones Problema 5",
     lwd = 3,
     type = "b",
     col = "blueviolet")
```



P6 Exploración de Transformaciones de Datos más Comunes

```
##Raíz Cuadrada
E_sqrt <- sqrt(E)

#plot(x, B_sqrt, main = "Transformación con Raíz Cuadrada")
#lines(x,B_sqrt, type = "l")

##Potencia al cuadrado
Epotencia2 <- E^2
#plot(x, Bpower2, main = "Transformación con Poder a la 2")
#lines(x,Bpower2, type = "l")

##Logaritmo
E_log <- log(E)
#plot(x, Bpower2, main = "Transformación por Logaritmos Naturales")
#lines(x,Bpower2, type = "l")

##1/x

unoEntreE <- 1/E

##Añadiendo todos los valores a una sola tabla
seis <- data.frame(x, E, E_log, E_sqrt, Epotencia2, unoEntreE)
```

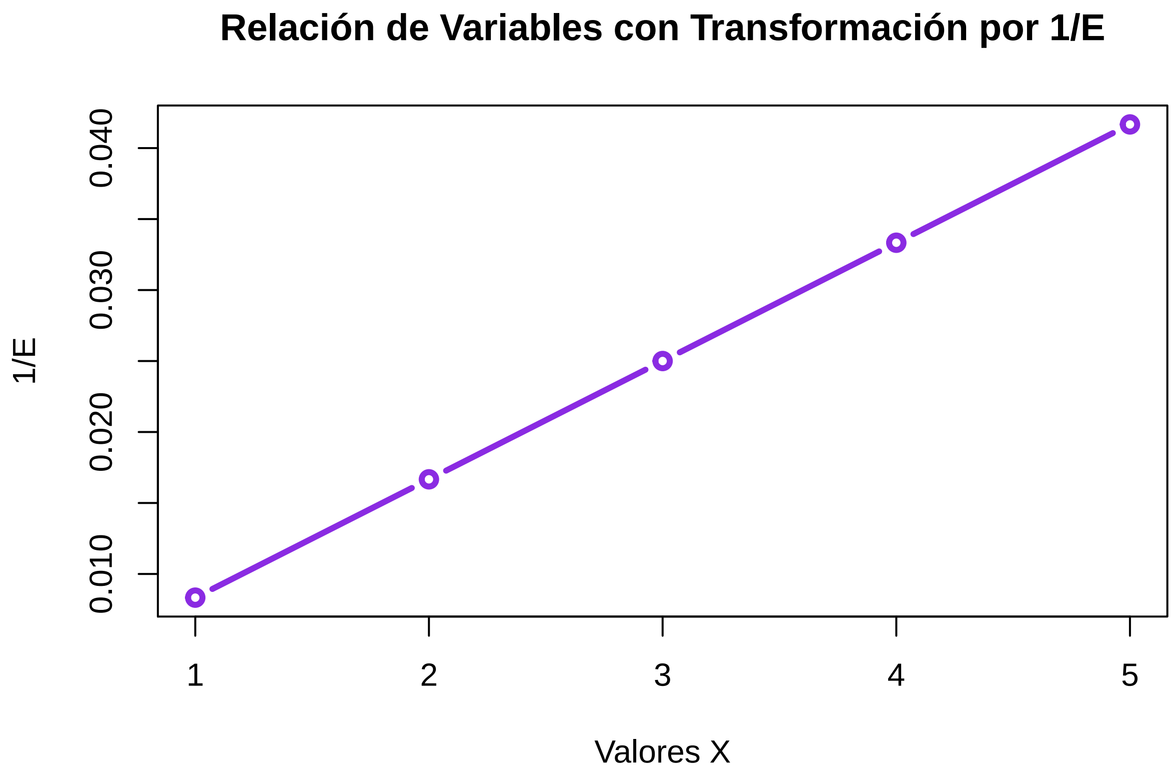
```r
##Una gráfica inicial
plot(seis,
     col = "blueviolet")
```



P6 Respuesta óptima

Para éste problema, al reducción por $1/E$ produce el mejor resultado. Entonces desarrollamos una gráfica solo de ésta transformación.

```r
plot(x, unoEntreE,
     type = "b",
     lwd = 3,
     col = "blueviolet",
     main = "Relación de Variables con Transformación por 1/E",
     xlab = "Valores X",
     ylab = "1/E")
```



Problema 7

Encuentra la pendiente cuando F es graficada contra x^2, después da la fórmula para F y x^2.

F	0.283	0.785	1.54	2.54
x	0.3	0.5	0.7	0.9

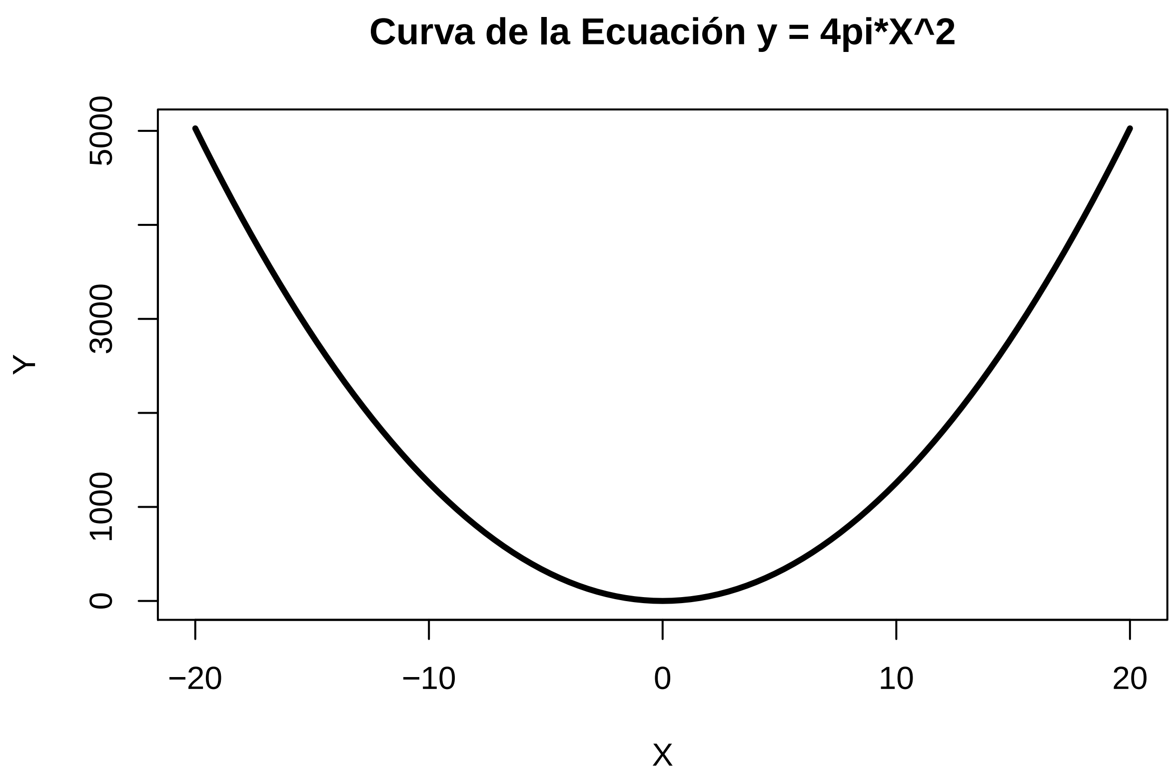
NO JALA P7 Creando los objetos necesarios en R y visualización inicial: NO JALA

```
#F <- c(0.283,0.785,1.54,2.54)
#x_1 <- c(0.3,0.5,0.7,0.9)
#xsqr <- x_1^2

#cuatro <- data.frame(E,x)

#plot(x, E,
 #    main = "Relación de Variables sin Modificaciones Problema 5",
  #   lwd = 3,
   #  type = "b",
    # col = "blueviolet")
```

Problema 8

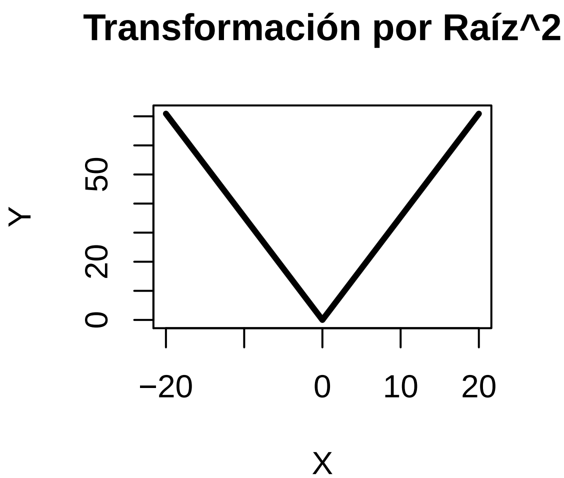
$$y = 4\pi x^2$$

```
eq = function(x){x^2*(pi*4)}
test8 <- curve(eq,
      from=-20, to=20,
      lwd = 3,
      main = "Curva de la Ecuación y = 4pi*X^2",
      xlab="X",
      ylab="Y")
```



Aplicando una transformación de datos por raíz cuadrada obtenemos la siguiente gráfica:

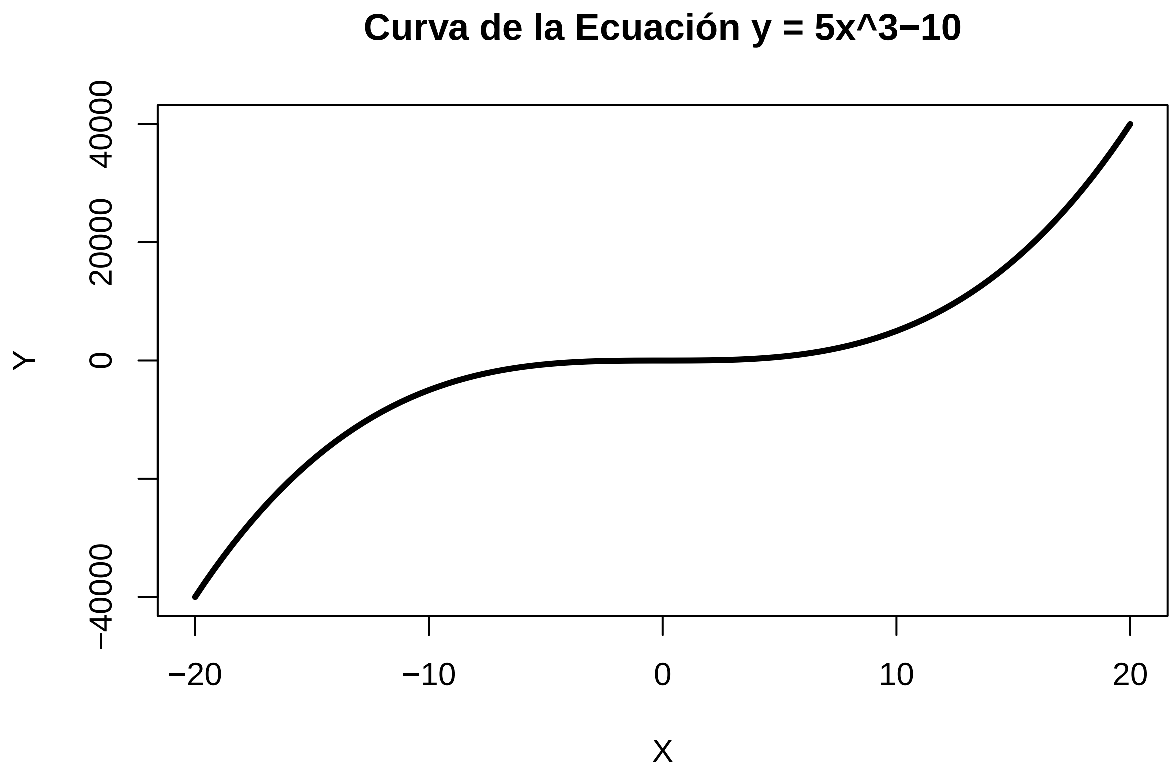
```
plot(test8$x, sqrt(test8$y),
     type = "l",
     lwd = 3,
     main = "Transformación por Raíz^2",
     xlab = "X",
     ylab = "Y")
```



Problema 9

$$y = 5x^3 - 10$$

```
eq = function(x){((x^3)*5)-10}
test9 <- curve(eq,
      from=-20, to=20,
      lwd = 3,
      main = "Curva de la Ecuación y = 5x^3-10",
      xlab="X",
      ylab="Y")
```



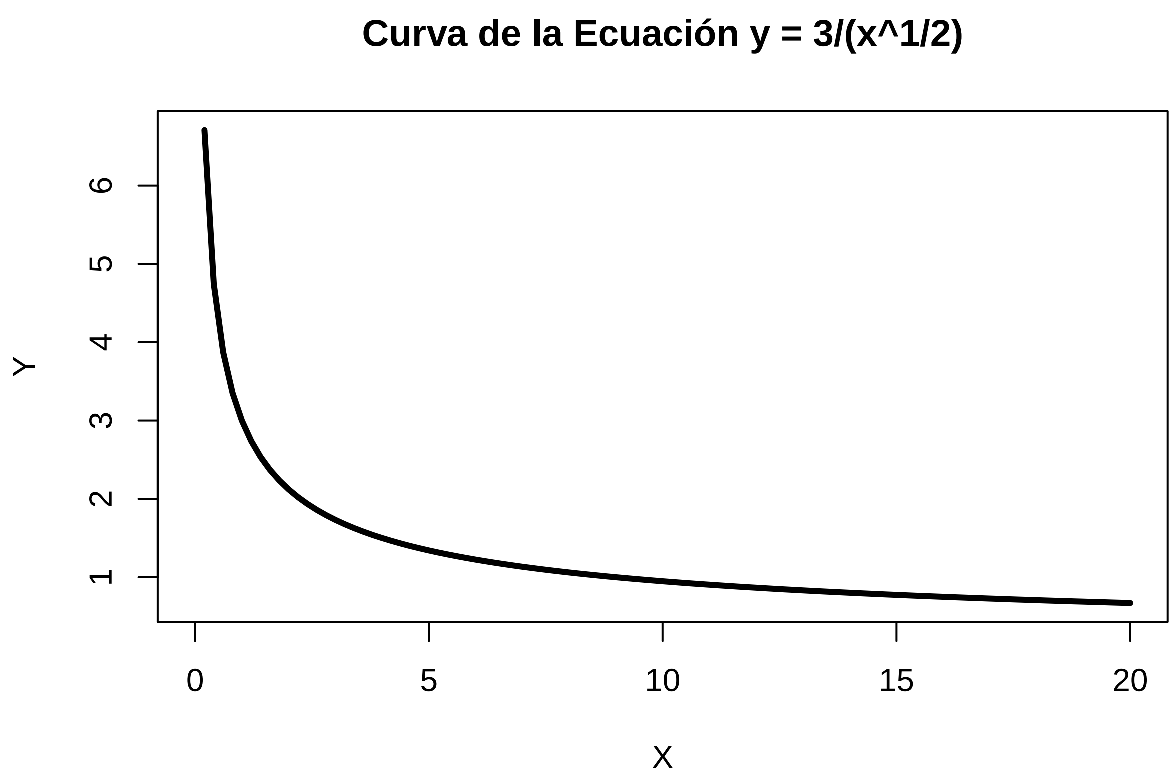
Aplicando una transformación de datos por raíz cúbica obtenemos la siguiente gráfica:

```r
plot(test9$x, (test9$y)^(1/3),
     type = "l",
     lwd = 3,
     main = "Transformación por Raíz^3",
     xlab = "X",
     ylab = "Y")
```

Problema 10

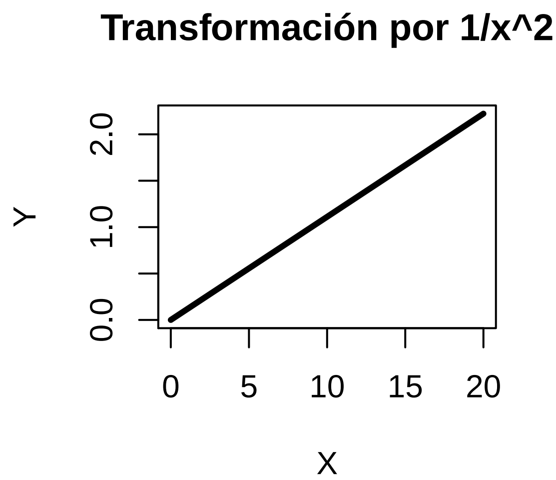
$$y = \frac{3}{x^{1/2}}$$

```r
eq = function(x){3/(x^(1/2))}
test10 <- curve(eq,
      from=0, to=20,
      lwd = 3,
      main = "Curva de la Ecuación y = 3/(x^1/2)",
      xlab="X",
      ylab="Y")
```



Aplicando una transformación de datos por raíz cúbica obtenemos la siguiente gráfica:

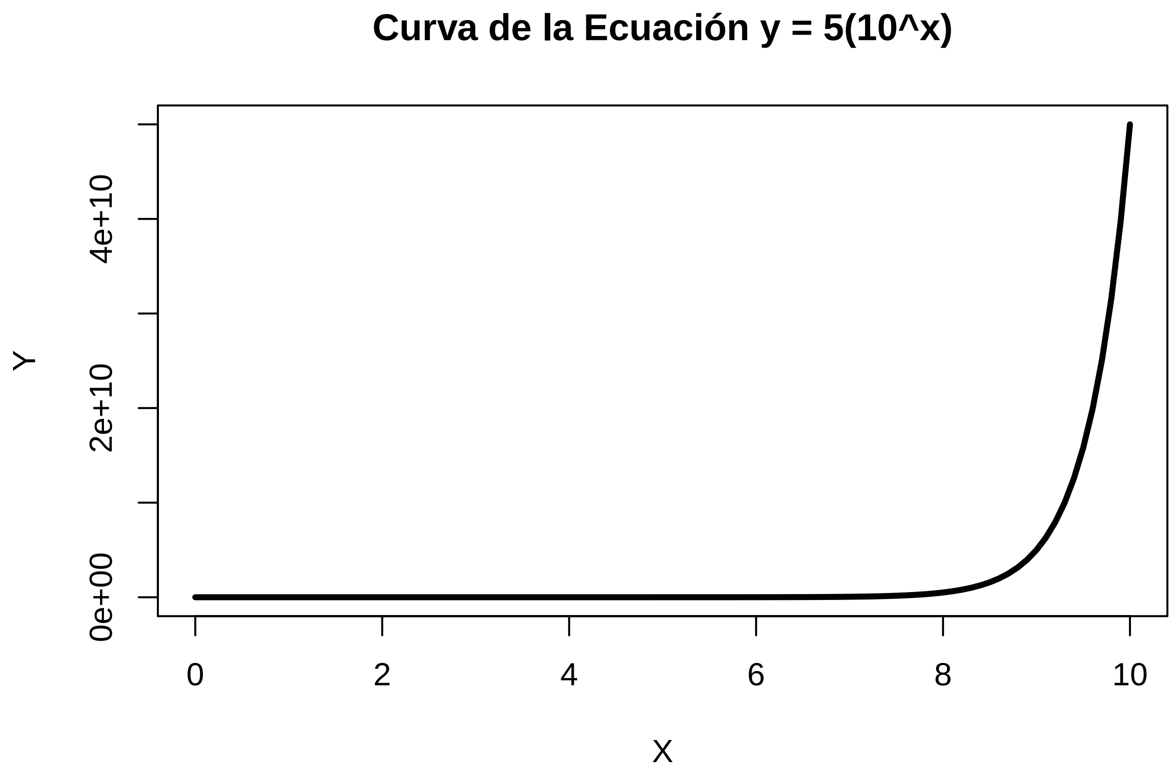
```r
plot(test10$x, 1/(test10$y)^2,
     type = "l",
     lwd = 3,
     main = "Transformación por 1/x^2",
     xlab = "X",
     ylab = "Y")
```



Problema 11

$$y = 5(10^x)$$

```
eq = function(x){10^x * 5}
test11 <- curve(eq,
      from=0, to=10,
      lwd = 3,
      main = "Curva de la Ecuación y = 5(10^x)",
      xlab="X",
      ylab="Y")
```




```r
plot(test11$x, log(test11$y),
     type = "l",
     lwd = 3,
     main = "Transformación por Log",
     xlab = "X",
     ylab = "Y")
```