Introducción al análisis de datos con R. Ejemplo: coronavirus

Probabilidad y Estadística (c)

12 de Abril, 2020

Cargamos los datos a R y los exploramos

- Cargamos los datos a R

```
datos<-read.table("datos_coronavirus.txt", header = TRUE, check.names = FALSE)
```

- Para ver los datos completos escribir en la consola

```
datos
```

- A veces nos alcanza con ver algunas filas, las primeras, por ejemplo:

```
head(datos)
```

- Si, en cambio, queremos ver las últimas, escribimos

```
tail(datos)
```

Para ver los nombres de las columnas

```
colnames(datos)
```

```
##  [1] "1/22/20" "1/23/20" "1/24/20" "1/25/20" "1/26/20" "1/27/20" "1/28/20"
##  [8] "1/29/20" "1/30/20" "1/31/20" "2/1/20"  "2/2/20"  "2/3/20"  "2/4/20"
## [15] "2/5/20"  "2/6/20"  "2/7/20"  "2/8/20"  "2/9/20"  "2/10/20" "2/11/20"
## [22] "2/12/20" "2/13/20" "2/14/20" "2/15/20" "2/16/20" "2/17/20" "2/18/20"
## [29] "2/19/20" "2/20/20" "2/21/20" "2/22/20" "2/23/20" "2/24/20" "2/25/20"
## [36] "2/26/20" "2/27/20" "2/28/20" "2/29/20" "3/1/20"  "3/2/20"  "3/3/20"
## [43] "3/4/20"  "3/5/20"  "3/6/20"  "3/7/20"  "3/8/20"  "3/9/20"  "3/10/20"
## [50] "3/11/20" "3/12/20" "3/13/20" "3/14/20" "3/15/20" "3/16/20" "3/17/20"
## [57] "3/18/20" "3/19/20" "3/20/20" "3/21/20" "3/22/20" "3/23/20" "3/24/20"
## [64] "3/25/20" "3/26/20" "3/27/20" "3/28/20" "3/29/20" "3/30/20" "3/31/20"
## [71] "4/1/20"  "4/2/20"  "4/3/20"  "4/4/20"  "4/5/20"  "4/6/20"  "4/7/20"
## [78] "4/8/20"  "4/9/20"  "4/10/20" "4/11/20" "4/12/20"
```

Para ver los nombres de las filas

```
rownames(datos)
```

```
##   [1] "Afghanistan"                "Albania"
##   [3] "Algeria"                    "Andorra"
##   [5] "Angola"                     "Antigua and Barbuda"
##   [7] "Argentina"                  "Armenia"
##   [9] "Australia"                  "Austria"
##  [11] "Azerbaijan"                 "Bahamas"
##  [13] "Bahrain"                    "Bangladesh"
##  [15] "Barbados"                   "Belarus"
##  [17] "Belgium"                    "Benin"
##  [19] "Bhutan"                     "Bolivia"
##  [21] "Bosnia and Herzegovina"     "Brazil"
##  [23] "Brunei"                     "Bulgaria"
##  [25] "Burkina Faso"               "Cabo Verde"
##  [27] "Cambodia"                   "Cameroon"
##  [29] "Canada"                     "Central African Republic"
##  [31] "Chad"                       "Chile"
##  [33] "China"                      "Colombia"
##  [35] "Congo (Brazzaville)"        "Congo (Kinshasa)"
##  [37] "Costa Rica"                 "Cote d'Ivoire"
##  [39] "Croatia"                    "Diamond Princess"
##  [41] "Cuba"                       "Cyprus"
##  [43] "Czechia"                    "Denmark"
##  [45] "Djibouti"                   "Dominican Republic"
##  [47] "Ecuador"                    "Egypt"
##  [49] "El Salvador"                "Equatorial Guinea"
##  [51] "Eritrea"                    "Estonia"
##  [53] "Eswatini"                   "Ethiopia"
##  [55] "Fiji"                       "Finland"
##  [57] "France"                     "Gabon"
##  [59] "Gambia"                     "Georgia"
##  [61] "Germany"                    "Ghana"
##  [63] "Greece"                     "Guatemala"
##  [65] "Guinea"                     "Guyana"
##  [67] "Haiti"                      "Holy See"
##  [69] "Honduras"                   "Hungary"
##  [71] "Iceland"                    "India"
##  [73] "Indonesia"                  "Iran"
##  [75] "Iraq"                       "Ireland"
##  [77] "Israel"                     "Italy"
##  [79] "Jamaica"                    "Japan"
##  [81] "Jordan"                     "Kazakhstan"
##  [83] "Kenya"                      "Korea, South"
##  [85] "Kuwait"                     "Kyrgyzstan"
##  [87] "Latvia"                     "Lebanon"
##  [89] "Liberia"                    "Liechtenstein"
##  [91] "Lithuania"                  "Luxembourg"
##  [93] "Madagascar"                 "Malaysia"
##  [95] "Maldives"                   "Malta"
##  [97] "Mauritania"                 "Mauritius"
##  [99] "Mexico"                     "Moldova"
## [101] "Monaco"                     "Mongolia"
## [103] "Montenegro"                 "Morocco"
## [105] "Namibia"                    "Nepal"
## [107] "Netherlands"                "New Zealand"
```

```
## [109] "Nicaragua"                    "Niger"
## [111] "Nigeria"                      "North Macedonia"
## [113] "Norway"                       "Oman"
## [115] "Pakistan"                     "Panama"
## [117] "Papua New Guinea"             "Paraguay"
## [119] "Peru"                         "Philippines"
## [121] "Poland"                       "Portugal"
## [123] "Qatar"                        "Romania"
## [125] "Russia"                       "Rwanda"
## [127] "Saint Lucia"                  "Saint Vincent and the Grenadines"
## [129] "San Marino"                   "Saudi Arabia"
## [131] "Senegal"                      "Serbia"
## [133] "Seychelles"                   "Singapore"
## [135] "Slovakia"                     "Slovenia"
## [137] "Somalia"                      "South Africa"
## [139] "Spain"                        "Sri Lanka"
## [141] "Sudan"                        "Suriname"
## [143] "Sweden"                       "Switzerland"
## [145] "Taiwan*"                      "Tanzania"
## [147] "Thailand"                     "Togo"
## [149] "Trinidad and Tobago"          "Tunisia"
## [151] "Turkey"                       "Uganda"
## [153] "Ukraine"                      "United Arab Emirates"
## [155] "United Kingdom"               "Uruguay"
## [157] "US"                           "Uzbekistan"
## [159] "Venezuela"                    "Vietnam"
## [161] "Zambia"                       "Zimbabwe"
## [163] "Dominica"                     "Grenada"
## [165] "Mozambique"                   "Syria"
## [167] "Timor-Leste"                  "Belize"
## [169] "Laos"                         "Libya"
## [171] "West Bank and Gaza"           "Guinea-Bissau"
## [173] "Mali"                         "Saint Kitts and Nevis"
## [175] "Kosovo"                       "Burma"
## [177] "MS Zaandam"                   "Botswana"
## [179] "Burundi"                      "Sierra Leone"
## [181] "Malawi"                       "South Sudan"
## [183] "Western Sahara"               "Sao Tome and Principe"
## [185] "Yemen"
```

Cómo extraer algunas filas

- Extraemos las filas 15 y 50:

```
c(15,50)
```

```
## [1] 15 50
```

```
datos[c(15,50),]
```

- Extraemos las primeras 3 filas:

```
1:3
```

```
## [1] 1 2 3
```

```
datos[1:3,]
```

Cómo extraer algunas columnas

- Usando su número

```
datos[,4]
```

- Usando su nombre

```
datos$"4/12/20"
```

¿Cuántos países tienen más de 5000 casos al día de hoy?

```
ncol(datos)
```

```
## [1] 82
```

```
datos[ , ncol(datos)]
```

```
##   [1]    607    446   1914    638     19     21   2142   1013   6315  13945
##  [11]   1098     46   1136    621     71   2578  29647     35      5    300
##  [21]   1009  22192    136    675    497      8    122    820  24298      8
##  [31]     18   7213  83134   2776     60    234    595    574   1600    712
##  [41]    669    633   5991   6369    214   2967   7466   2065    125     21
##  [51]     34   1309     14     71     16   2974 133670     49      9    257
##  [61] 127854    566   2114    155    250     45     33      8    393   1410
##  [71]   1701   9205   4241  71686   1352   9655  11145 156363     69   6748
##  [81]    389    951    197  10512   1234    377    651    630     50     79
##  [91]   1053   3281    106   4683     20    378      7    324   4219   1662
## [101]     93     16    272   1661     16     12  25746   1330      9    529
## [111]    323    828   6525    599   5230   3234      2    134   7519   4648
## [121]   6674  16585   2979   6300  15770    126     15     12    356   4462
## [131]    280   3630     11   2532    742   1205     25   2173 166831    210
## [141]     19     10  10483  25415    388     32   2551     76    113    707
## [151]  56956     54   2777   4123  85206    480 555313    865    181    262
## [161]     43     14     16     14     21     25      2     14     19     25
## [171]    290     38    105     12    283     41      9     13      5     10
## [181]     13      4      6      4      1
```

¿Cuántos países tienen más de 5000 casos al día de hoy?

```
sum(datos[, ncol(datos)]>5000)
```

```
## [1] 33
```

¿Qué países tienen más de 150000 casos?

```
  which(datos[,ncol(datos)]>150000)
```

```
## [1]  78 139 157
```

```
rownames(datos)[c(78,139,157)]
```

```
## [1] "Italy" "Spain" "US"
```

o lo hacemos en dos pasos

```
cuales <- which( datos[ncol(datos)]>150000)
cuales
```

```
## [1]  78 139 157
```

```
rownames(datos)[cuales]
```

```
## [1] "Italy" "Spain" "US"
```

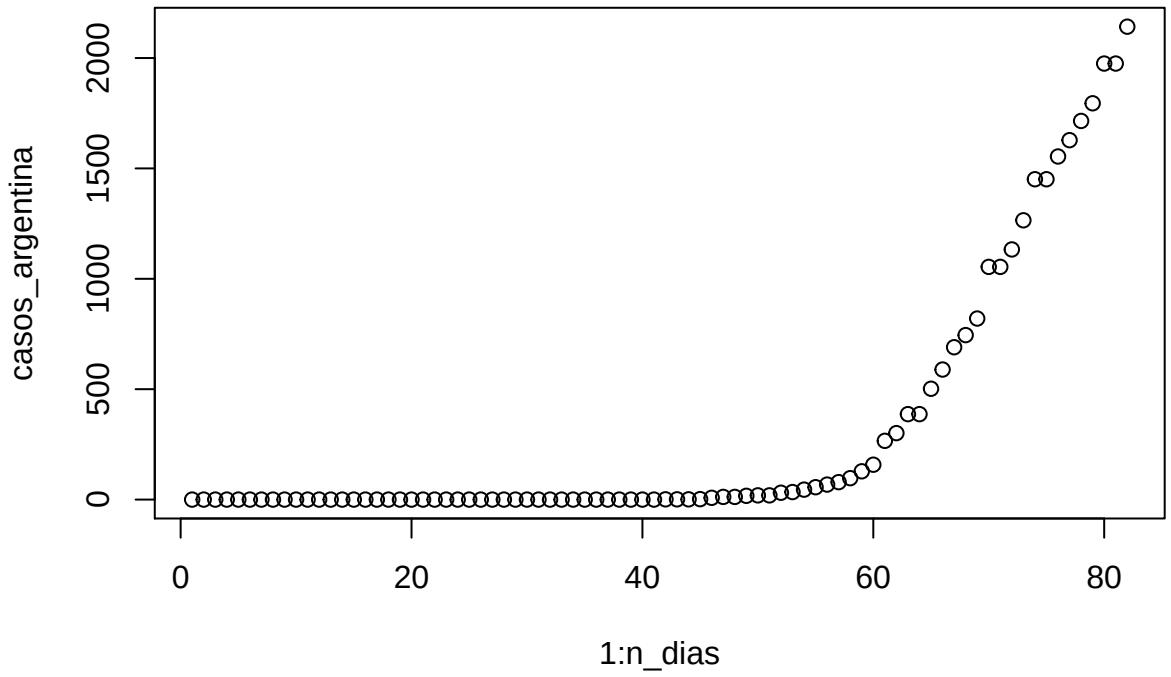
¿Cuántos casos confirmados hay hoy en el mundo?

```
sum(datos[,ncol(datos)])
```

```
## [1] 1846679
```

Graficamos la cantidad de casos confirmados en Argentina por día

```
i_arg <- which(rownames(datos)=="Argentina")
n_dias <- ncol(datos)
casos_argentina <- datos[i_arg,]
plot(1:n_dias, casos_argentina)
```



Graficamos la cantidad de casos confirmados en Argentina por día hasta el 29 de marzo

```
i_arg <- which(rownames(datos)=="Argentina")
colnames(datos)
```

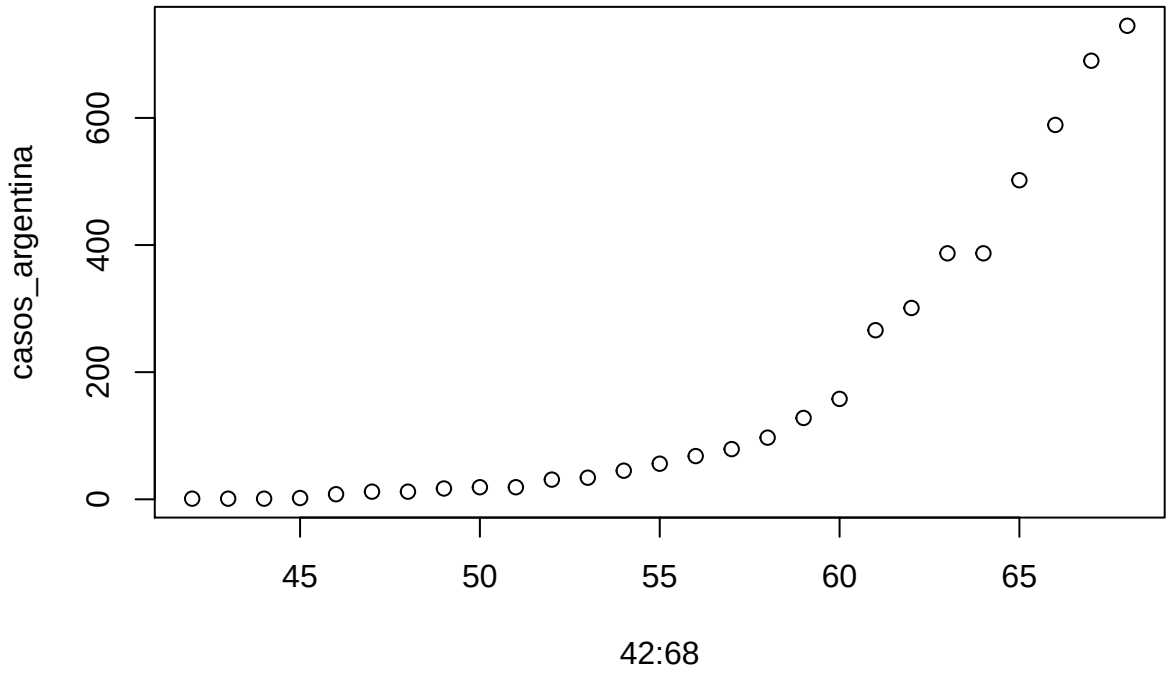
```
##  [1] "1/22/20" "1/23/20" "1/24/20" "1/25/20" "1/26/20" "1/27/20" "1/28/20"
##  [8] "1/29/20" "1/30/20" "1/31/20" "2/1/20"  "2/2/20"  "2/3/20"  "2/4/20"
## [15] "2/5/20"  "2/6/20"  "2/7/20"  "2/8/20"  "2/9/20"  "2/10/20" "2/11/20"
## [22] "2/12/20" "2/13/20" "2/14/20" "2/15/20" "2/16/20" "2/17/20" "2/18/20"
## [29] "2/19/20" "2/20/20" "2/21/20" "2/22/20" "2/23/20" "2/24/20" "2/25/20"
## [36] "2/26/20" "2/27/20" "2/28/20" "2/29/20" "3/1/20"  "3/2/20"  "3/3/20"
## [43] "3/4/20"  "3/5/20"  "3/6/20"  "3/7/20"  "3/8/20"  "3/9/20"  "3/10/20"
## [50] "3/11/20" "3/12/20" "3/13/20" "3/14/20" "3/15/20" "3/16/20" "3/17/20"
## [57] "3/18/20" "3/19/20" "3/20/20" "3/21/20" "3/22/20" "3/23/20" "3/24/20"
## [64] "3/25/20" "3/26/20" "3/27/20" "3/28/20" "3/29/20" "3/30/20" "3/31/20"
## [71] "4/1/20"  "4/2/20"  "4/3/20"  "4/4/20"  "4/5/20"  "4/6/20"  "4/7/20"
## [78] "4/8/20"  "4/9/20"  "4/10/20" "4/11/20" "4/12/20"
```

```
casos_argentina <- datos[i_arg, 42:68]
```

Graficamos la cantidad de casos confirmados en Argentina por día hasta el 29 de marzo

```
plot(42:68, casos_argentina)
```

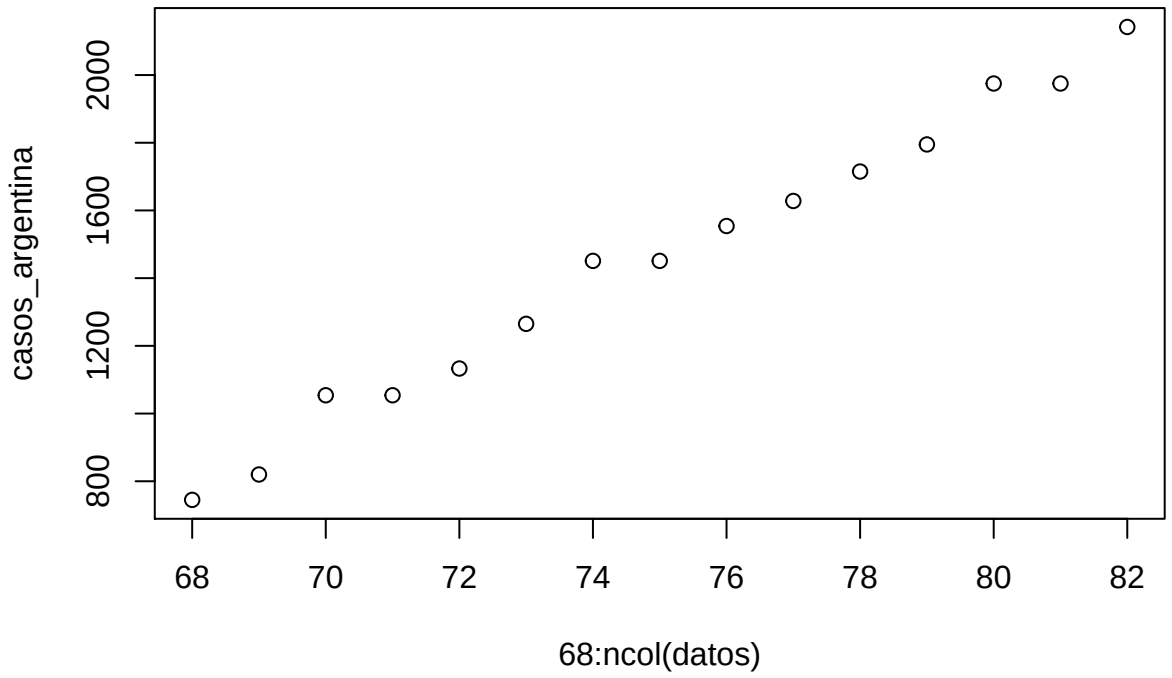
Graficamos la cantidad de casos confirmados en Argentina por día hasta el 29 de marzo



Graficamos la cantidad de casos confirmados en Argentina por día desde el 29 de marzo

```
i_arg <- which(rownames(datos)=="Argentina")
casos_argentina <- datos[i_arg, 68:ncol(datos)]
plot(68:ncol(datos), casos_argentina)
```

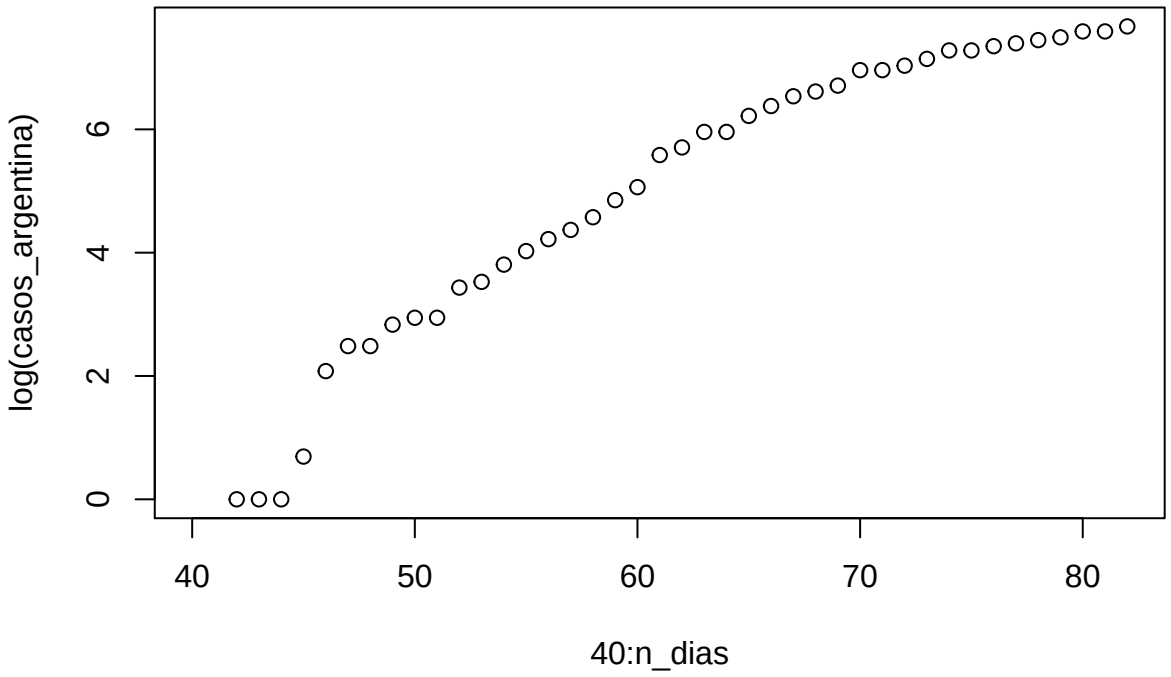
Graficamos la cantidad de casos confirmados en Argentina por día desde el 29 de marzo



Graficamos el logaritmo de la cantidad de casos confirmados en Argentina por día

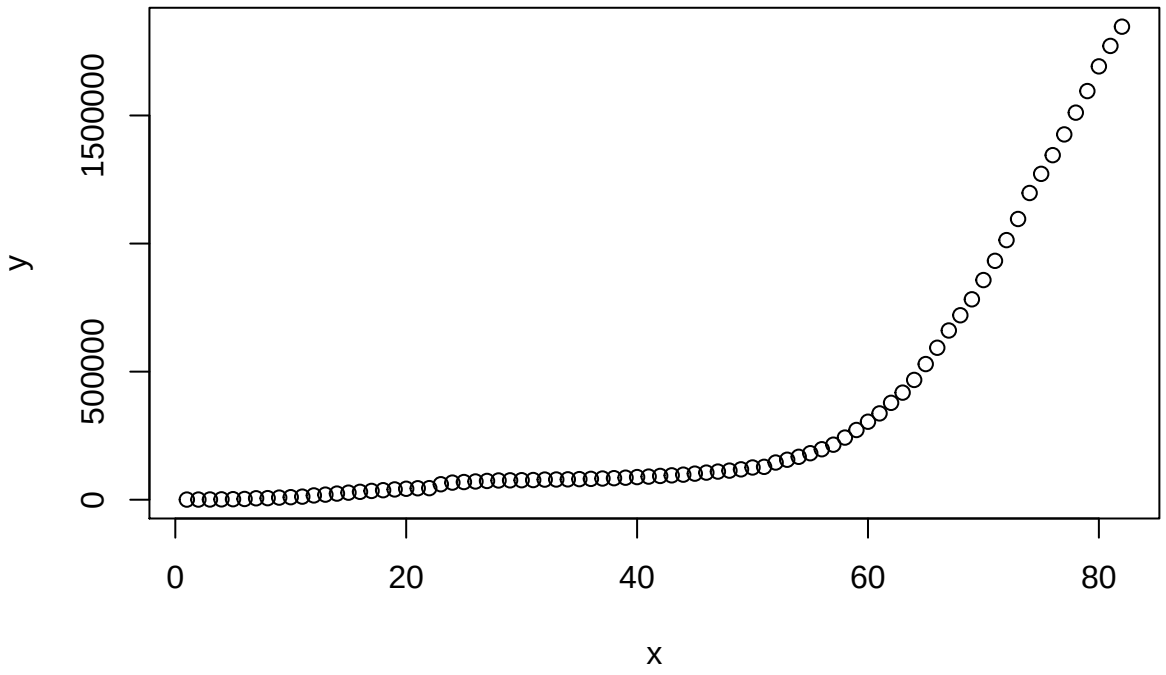
```
i_arg <- which(rownames(datos)=="Argentina")
n_dias <- ncol(datos)
casos_argentina <- datos[i_arg,40:n_dias]
plot(40:n_dias, log(casos_argentina))
```

Graficamos el logaritmo de la cantidad de casos confirmados en Argentina por día



Graficamos los casos totales en el mundo desde el 22 de enero

```
x<-1:ncol(datos)
y<-colSums(datos)
plot(x,y)
```



Graficamos el logaritmo de los casos totales en el mundo desde el 22 de enero

```
x<-1:ncol(datos)
y<-colSums(datos)
plot(x, log(y))
```