Proyecto 2 - Analisis Exploratorio

Manuel Valenzuela 15072, Davis Álvarez 15842, Jorge Súchite 15293

23 de agosto de 2018

Descripción del Tema

Para este segundo proyecto propone observar la población de los paises del mundo a lo largo de los años y así mismo su área superficial, con estos datos obtener una densidad poblacional y ver si esta esta relacionada de alguna manera con la felicidad de las personas. Con estos datos se tratará de hacer 2 pronósticos. El primero de la población en base a su propio comportamiento; y el segundo, de la felicidad del mundo en base a su propio comportamiento y así mismo a la densidad poblacional a lo largo del tiempo.

Para hacer esto se uilizarán 3 datasets, el primero es World Happiness Report, el cual es un dataset que contiene una puntuación de felicidad anotada según producción económica, apoyo social, etc. El segundo es Countries Population, el cual contiene la población de cada país a lo largo de los años,y por último, el tercero sería Countries of the World, del cual se obtendría el área superficial de cada país.

Problema Científico

Lo que se buscará realizar con este proyecto es observar la densidad poblacional y la felicidad de las personas de cada país a lo largo del tiempo, ver si tienen algún tipo de relación entre sí y tratar de pronosticar que tan felices serán las personas en los años próximos. Así mismo, observando el comportamiento de la población a lo largo de los años se buscará pronosticar la población de los paises en los próximos años.

Objetivos

Generales

- Encontrar la Densidad Poblacional de los paises del mundo a lo largo de los años.
- Encontrar si hay una relación entre la densidad poblacional y la felicidad de las personas.

Especificos

- Describir el dataset de Countries Population, encontrar un patrón del comportamiento de la población y pronosticarla en los próximos años.
- Describir el dataset de World Happiness Report, encontrar las variables que más inciden en la felicidad de las personas y pronosticarla en los próximos años.
- Comparar los pronosticos y observar si efectivamente hay una relación entre la densidad poblacional y la felicidad de las personas.

Descripción de los Datos

A continuación se describen los datasets y así mismo cada una de las variables que estos tienen. Luego de eso es explicará el proceso de limpieza que se llevó a cabo para tomar en cuenta solo las variables que serán de utilidad para cumplir los objetivos planteados.

World Happiness Report

El World Happiness Report es una encuesta histórica sobre el estado de la felicidad global. El primer informe se publicó en 2012, el segundo en 2013, el tercero en 2015 y el cuarto en la Actualización de 2016. The World Happiness 2017, que clasifica a 155 países por su nivel de felicidad, fue lanzado en las Naciones Unidas en un evento que celebra el Día Internacional de la Felicidad el 20 de marzo.

Los rankings y puntajes de felicidad utilizan data de una encuesta de Gallup World. Las puntuaciones son basadas en las respuestas a las preguntas de evaluación de vida contenidas en esta encuesta. Estas preguntas se conocen como la Escalera de Cantril, esta les pide a los encuestados que piensen en una escalera con la mejor vida posible para ellos siendo un 10 y la peor vida posible siendo un 0. A continuación se describen las variables contenidas en este dataset.

No.	Variable	Tipo	Descripción
1	Country	Factor	Nombre del país
2	Region	Factor	Region a la que pertenece
3	Happiness.Rank	int	Rango del país basado en el Happiness Score
4	Happiness.Score	num	Puntuación obtenida de Encuesta
5	Standard.Error	num	La desviación Estándar del Happiness Score
6	Economy..GDP.per.Capita	num	El grado en que el Producto Interno Bruto del país contribuye al cálculo del puntaje de felicidad.
7	Family	num	El grado en que la familia contribuye al cálculo del puntaje de felicidad.
8	Health..Life.Expectancy	num	el grado en que la esperanza de vida contribuye al cálculo del puntaje de felicidad
9	Freedom	num	La medida en que la libertad contribuye al cálculo de la puntuación de la felicidad
10	Trust..Government.Corruption.	num	El grado en que la percepción de la corrupción contribuye al puntaje de felicidad
11	Generosity	num	El grado en que generosidad contribuye al cálculo del puntaje de felicidad
12	Dystopia.Residual	num	La medida en que la distopía residual contribuyó al cálculo de la puntuación de la felicidad
13	Year	num	Año en el que fue hecha la encuesta
14	Lower.Confidence.Interval	num	Intervalo de confianza más bajo del puntaje de felicidad
15	Upper.Confidence.Interval	num	Intervalo de confianza superior del puntaje de felicidad
16	Whisker.high	num	Bigote Superior
17	Whisker.low	num	Bigote Inferior

De las variables descritas en la tabla anterior se tenían 3 datasets, uno por cada año. El proceso de limpieza que se hizo con estos datos fue cuadrar las variables en todos los datasets, es decir, que se hizo que todos tuviesen las mismas variables con el mismo nombre. Luego se unieron los 3 datasets dejando al año como una variable más.

```
#Juntamos Datasets de Happines en una nueva Variable
h2015["year"] <- 2015
h2016["year"] <- 2016
h2017["year"] <- 2017

h2015[c("Lower.Confidence.Interval","Upper.Confidence.Interval","Whisker.high","Whisker.low")] <- NA
h2016[c("Standard.Error","Whisker.high","Whisker.low")] <- NA
h2017[c("Standard.Error","Lower.Confidence.Interval","Upper.Confidence.Interval","Region")] <- NA

hWorld <- rbind(h2015,h2016,h2017)
```

Countries Population

Este Dataset contiene la población de 217 paises a lo largo de 56 años, desde 1960 hasta el 2016. A continuación se detallan las variables de este dataset.

No.	Variable	Tipo	Descripción
1	i..Country	Factor	ID del país
2	Country.Code	Factor	Código del país
3	Indicator.Name	Factor	Nombre del indicador
4	Indicator.Code	Factor	Código del indicador
5	year	chr	Año de medida
6	population	num	Población

El proceso de limpieza que se hizo con este dataset fue poner los años como valor en una fila ya que estos venían como columnas. Obteniendo 6 variables con 12,369 observaciones.

```
#Ordenamos pWorld
colnames(pWorld)[5:61] <- c("1960","1961","1962","1963","1964","1965","1966","1967","1968","1969","1970
pWorld <- gather(pWorld, year, population, 5:61)
pWorld <- subset( pWorld, select = -X)
```

Countries of the World

Este dataset contiene información sobre población, región, tamaño del área, mortalidad infantil y más. Todos estos conjuntos de datos están formados por datos del gobierno de EE. UU. Las variables contenidas se detallan a continuación.

No.	Variable	Tipo	Descripción
1	Country	Factor	Nombre del país
2	Region	Factor	Región en la que se encuentra el país
3	Population	int	Población Actual
4	Area..sq..mi..	int	Área superficial en mi^2
5	Pop..Density..per.sq..mi..	num	Densidad Poblacional en $personas/mi^2$
6	Coastline..coast.area.ratio.	num	línea costera
7	Net.migration	num	Migración neta
8	Infant.mortality..per.1000.births.	num	Mortalidad infantil (por 1000 nacimientos)
9	GDP....per.capita.	int	Producto Interno Bruto
10	Literacy....	num	Porcentaje de Alfabetismo
11	Phones..per.1000.	num	Teléfonos por cada 1000 personas
12	Arable....	num	Porcentaje de tierra cultivable
13	Crops....	num	Porcentaje de cultivos
14	Other....	num	Otros
15	Climate	num	Clima
16	Birthrate	num	Tasa de Nacimiento
17	Deathrate	num	Tasa de Mortalidad
18	Agriculture	num	Agricultura
19	Industry	num	Industria
20	Service	num	Servicio

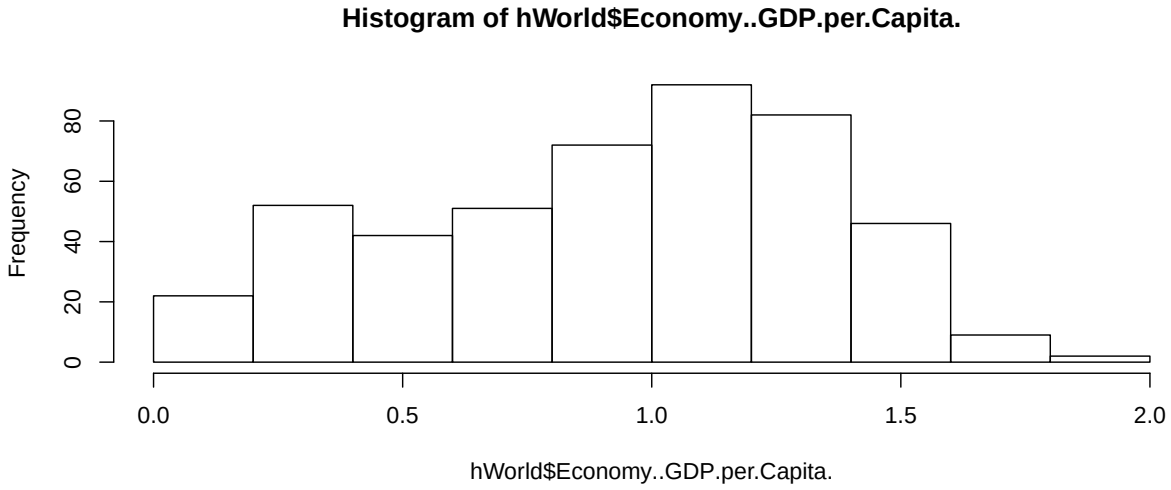
Para este dataset fue necesario convertir algunas columnas de tipo Factos a Numeric, por lo que se reemplazó la "," por ".". Obteniendo un dataset con 227 observaciones y 20 variables.

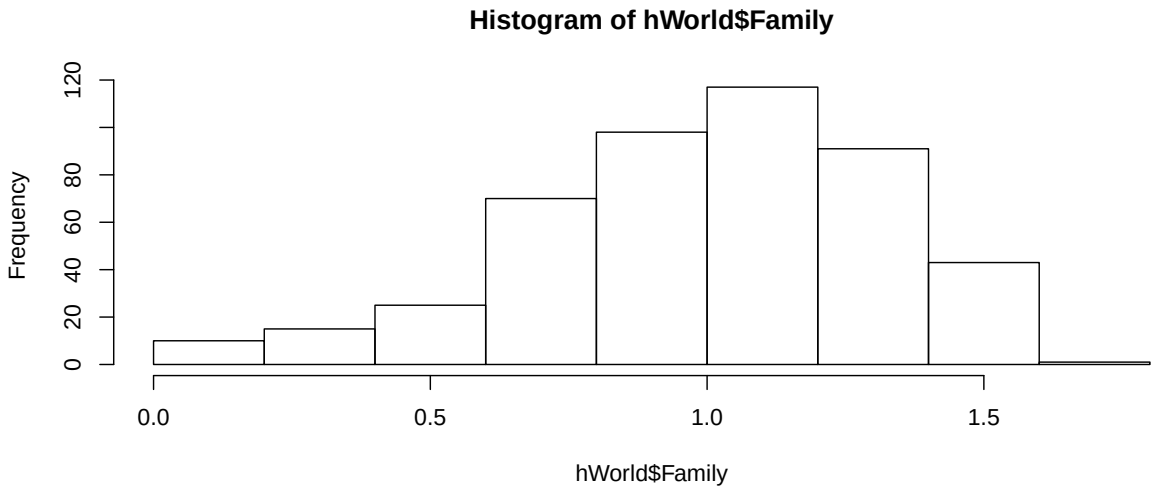
```r
#Convertimos Factors a num/int reemplazando "," por "."
cWorld$Coastline..coast.area.ratio.<- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Coastline..coast
cWorld$Pop..Density..per.sq..mi.. <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Pop..Density..per
cWorld$Net.migration <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Net.migration)))
cWorld$Infant.mortality..per.1000.births. <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Infant.mo
cWorld$Literacy.... <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Literacy....)))
cWorld$Phones..per.1000. <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Phones..per.1000.)))
cWorld$Crops.... <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Crops....)))
cWorld$Other.... <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Other....)))
cWorld$Climate <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Climate)))
cWorld$Arable.... <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Arable....)))
cWorld$Birthrate <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Birthrate)))
cWorld$Deathrate <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Deathrate)))
cWorld$Agriculture <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Agriculture)))
cWorld$Industry <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Industry)))
cWorld$Service <- as.numeric(gsub(",", ".", gsub("\\.", "", cWorld$Service)))

View(cWorld)
View(hWorld)
```

Exploración de Datos

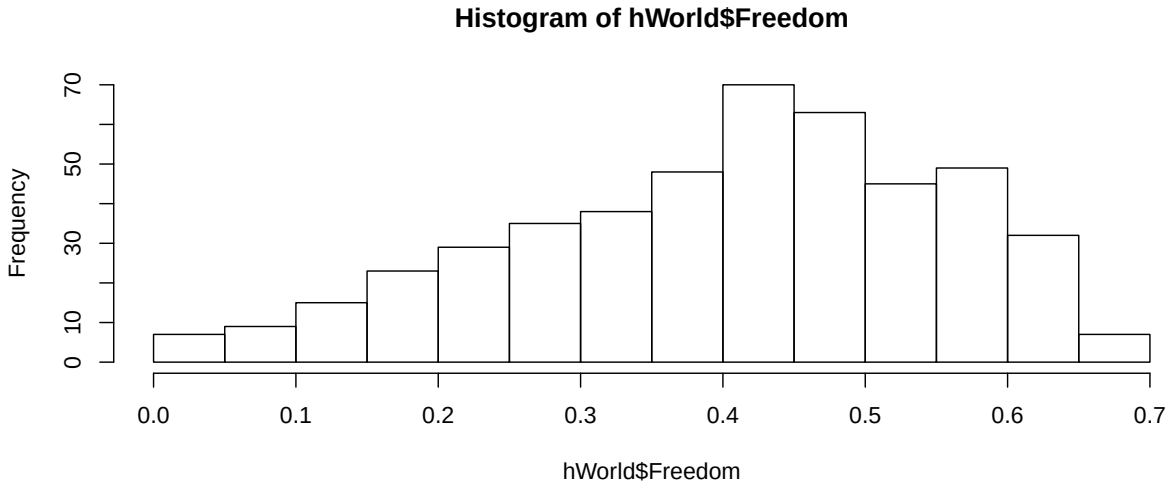
World Happiness Report

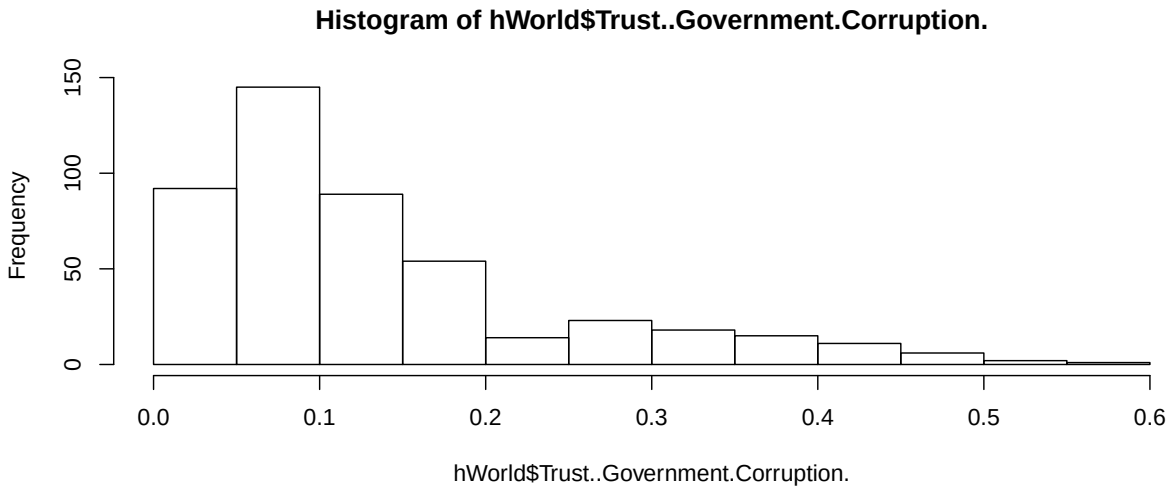
A continuación se muestran los histogramas de cada variable de *World Happiness Report*

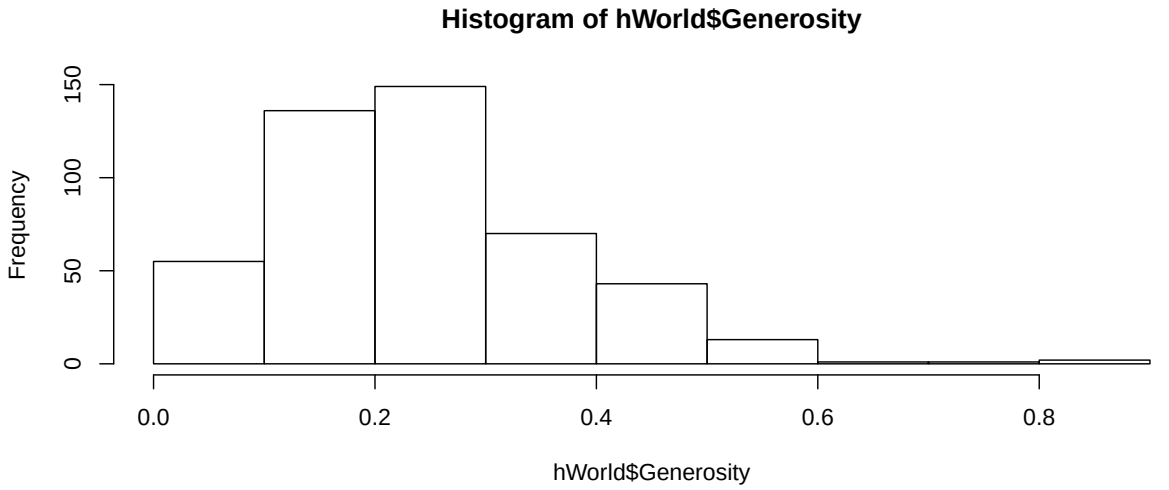

Histogram of hWorld$Economy..GDP.per.Capita.
Frequency
0
20
40
60
80
0.0
0.5
1.0
1.5
2.0
hWorld$Economy..GDP.per.Capita.


Histogram of hWorld$Family
Frequency
0
20
40
60
80
120
0.0
0.5
1.0
1.5
hWorld$Family


Histogram of hWorld$Health..Life.Expectancy.
Frequency
0
20
40
60
80
0.0
0.2
0.4
0.6
0.8
1.0
hWorld$Health..Life.Expectancy.







###Interpretación de histogramas

No.1 Punteo de Felicidad

Vemos que, en su mayoría existen más países en el rango de 5 y 6 en este rango.

No.2 Producto Interno Bruto

Se puede observar que, muy pocos países tienen un índice per capita entre 1.5 y 2 y que la mayoría de paises tienen un índice de 0.7 y 1.

No.3 Familia

Vemos que la mayoría de los datos, su índice de familia está entre 0.6 y 1.4

No.4 Esperanza de vida

La mayoría de los datos que se tienen de países nos muestran que la mayoría de las personas tienen una esperanza de vida entre el rango de 0.55 y un máximo de 0.9

No.5 Libertad

Vemos que, en su mayoría de todos lod datos, estos se encuentran con muy poca libertad puesto que la mayoría son menores a un rango de 0.4

No. 6 Trust Goverment Corruption

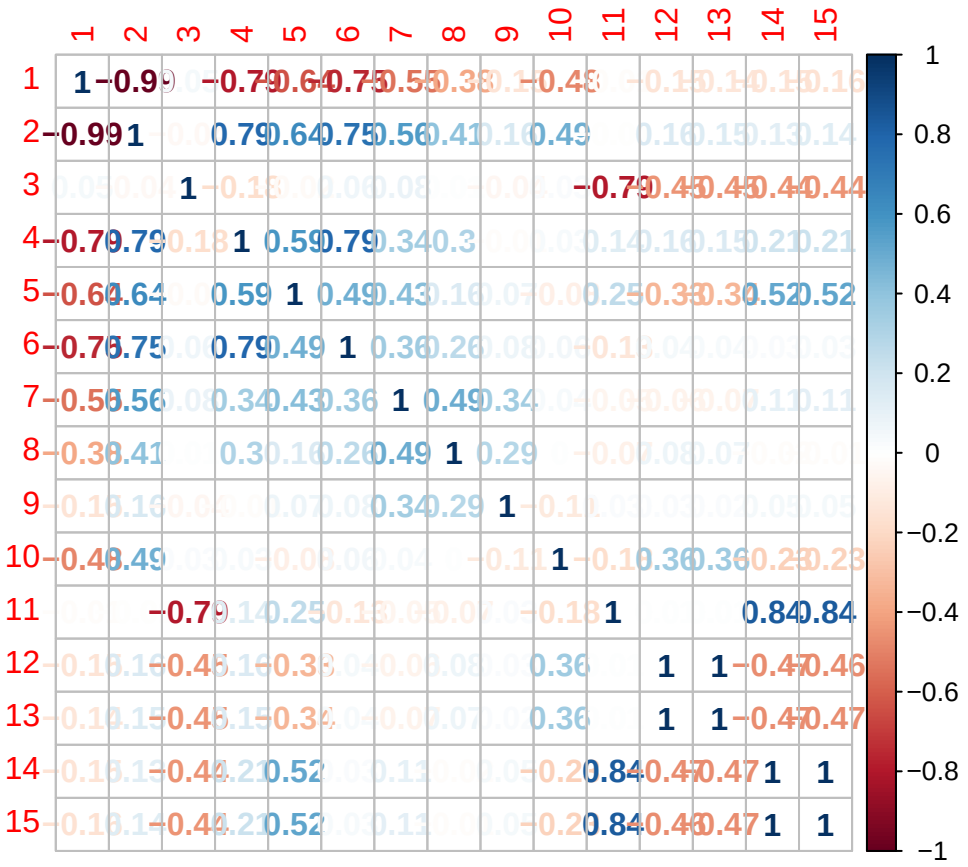
Vemos en la gráfica que muy pocos datos se tienen de países que son poco corruptos y que la mayoría tiene un índice de corrupción muy alto.

No. 7 Histograma de Generosidad

Se puede observar en el histograma que la mayoría de los datos que se tienen son datos de los países que no son tan generosos que digamos.

Correlación entre variables

Para poder observar si existe algún tipo de relación entre las variables numéricas del dataset, se realizó un gráfico de correlación en donde cada variable se puso contra cada variable del dataset. En el resultado podemos observar la correlación existente entre las variables:



En el diagrama de correlación se observó que existen coeficientes de correlaciones de varios valores, sin embargo, para este experimento los valores arriba de un 0.4 fueron considerados como significativos. Por lo que las relaciones que más se destacan son las siguientes:

- Happiness.Rank - Happiness.Score: 0.99
- Happiness.Rank - Economy..GDP.per.Capita: 0.79
- Happiness.Rank - Family: 0.64
- Happiness.Rank - Health..Life.Expectancy : 0.75
- Happiness.Rank - Freedom: 0.55
- Happiness.Rank - Dystopia.Residual: 0.55
- Happiness.Score - Economy..GDP.per.Capita: 0.79
- Happiness.Score - Family: 0.64
- Happiness.Score - Health..Life.Expectancy: 0.75
- Happiness.Score - Freedom: 0.56
- Happiness.Score - Trust..Government.Corruption: 0.41
- Happiness.Score - Dystopia.Residual: 0.49
- Economy..GDP.per.Capita - Family: 0.59
- Economy..GDP.per.Capita - Health..Life.Expectancy: 0.79

- Family - Health..Life.Expectancy: 0.49
- Family - Freedom: 0.43
- Freedom - Trust..Government.Corruption.: 0.49

Dentro de las correlaciones más significas, podemos resaltar que la existente entre *Happiness.Rank-Happiness.Score* es de 0.99, esto se debe a que el puesto del rank depende directamente de la puntuación obtenida en el test. Así mismo podemos observar que la familia, la economía y la esperanza de vida tienen gran influencia en el rank de felicidad de los países.

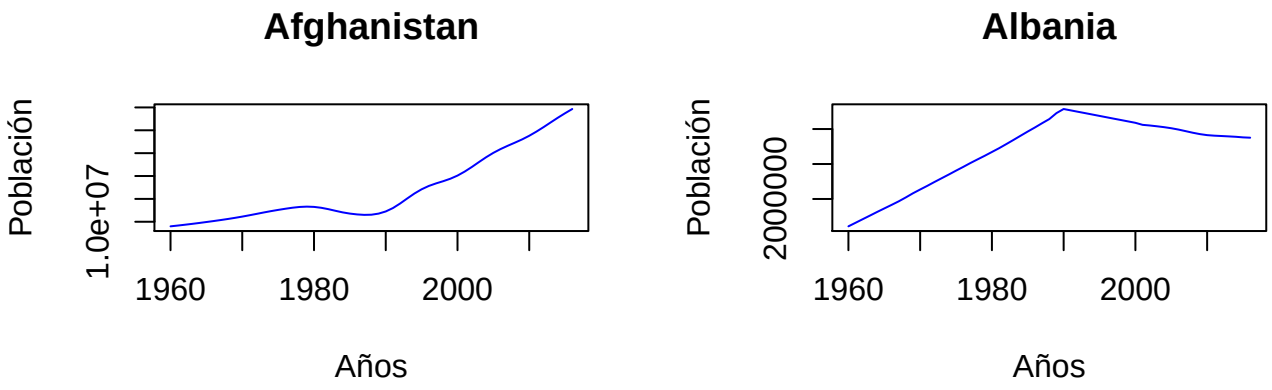
También es importante aclarar que las correlaciones existentes entre una variable y las variables *Lower.Confidence.Interval*, *Upper.Confidence.Interval*, *Whisker.high* y *Whisker.low* fueron omitidas debido a que estas variables no se encuentran registradas en más de un dataset.

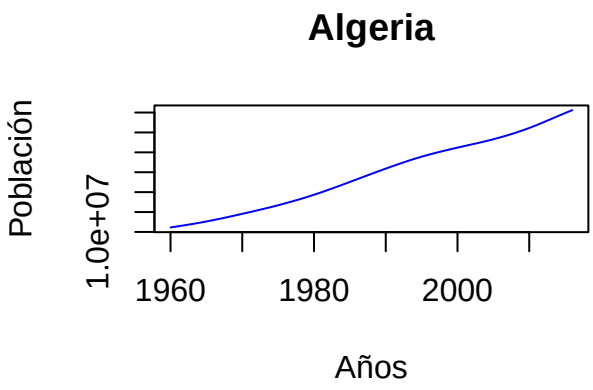
Countries Population

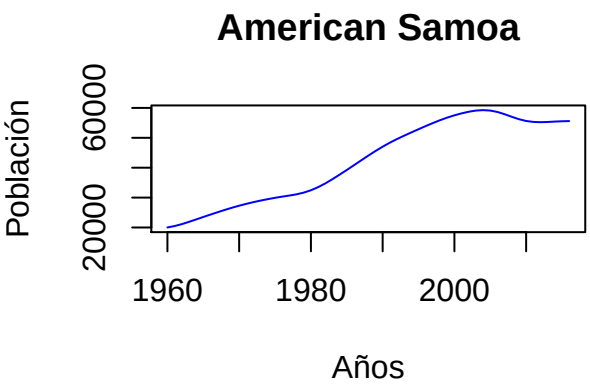
Los hallazgos encontrados en el dataset de Countries Population son simples, debido a que se encontró que la variable population, que describe la población de cada país, depende únicamente de la variable year. Por ende, crear un modelo que describa o pronostique la población de cada país consiste en tratar de encontrar un patrón y un modelo entre la variable population y la variable year. Debido a que solo se necesita de la variable year, tomada en diferentes unidades de tiempo, para describir cómo se comporta la población en cada país.

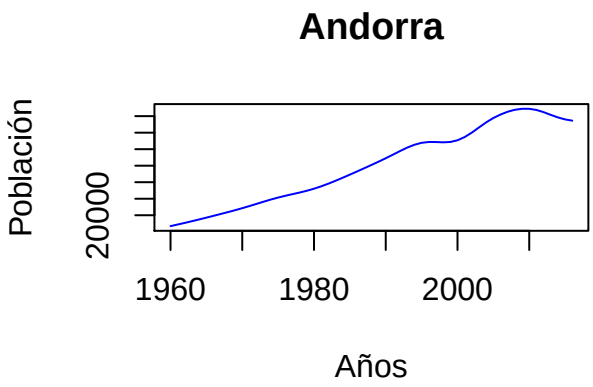
Gráficas de Población

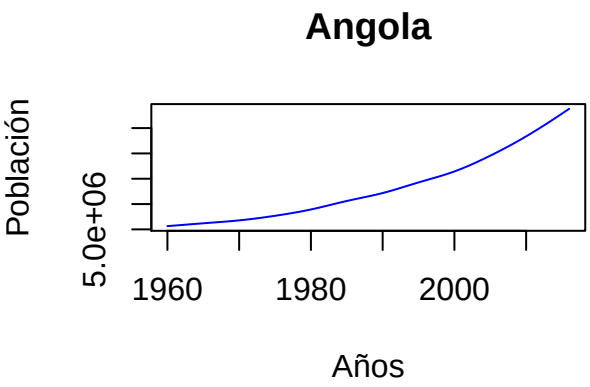
A continuación se muestran las gráficas de población de algunos países:

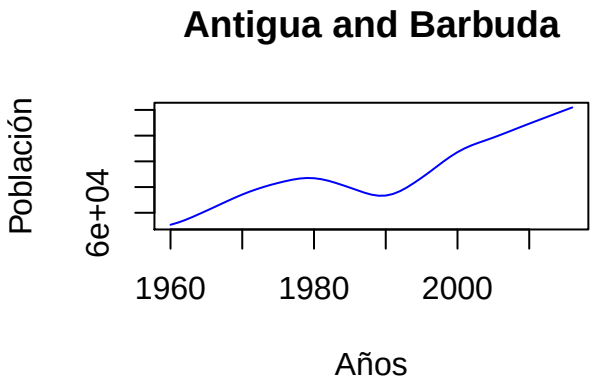


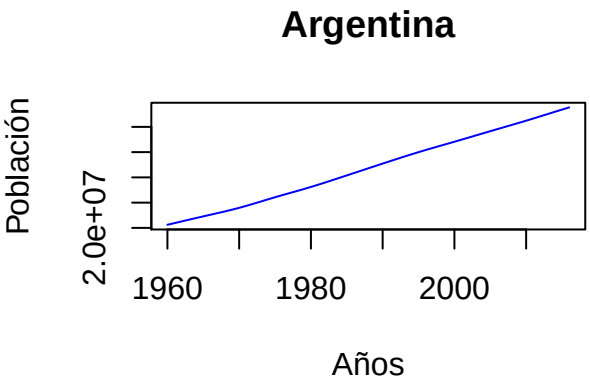

Algeria
Población
1.0e+07
1960
1980
2000
Años

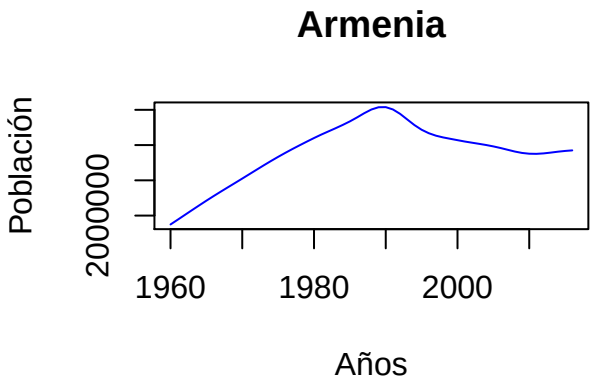

American Samoa
Población
20000
60000
1960
1980
2000
Años

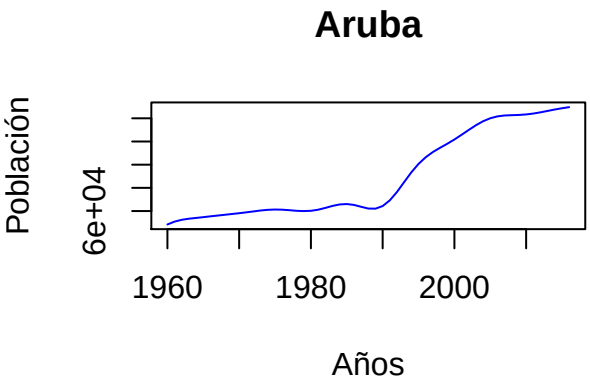

Andorra
Población
20000
1960
1980
2000
Años

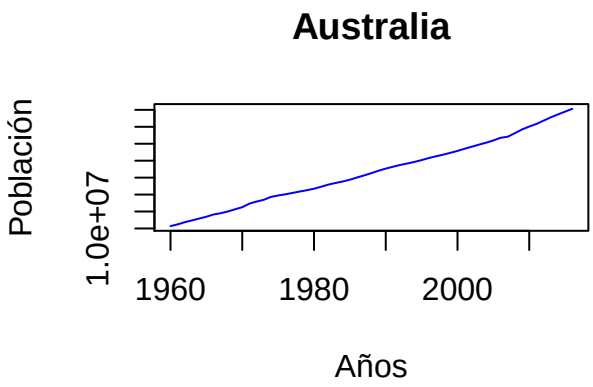

Angola
Población
5.0e+06
1960
1980
2000
Años

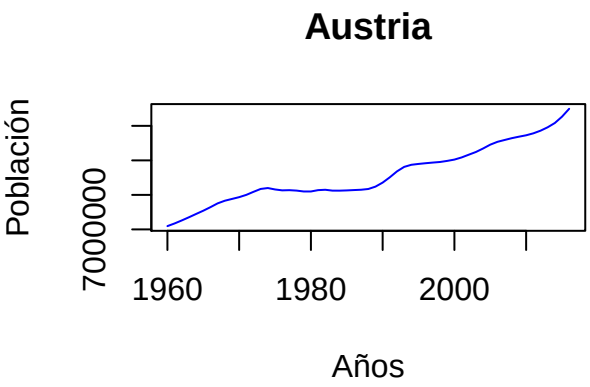

Antigua and Barbuda
Población
6e+04
1960
1980
2000
Años

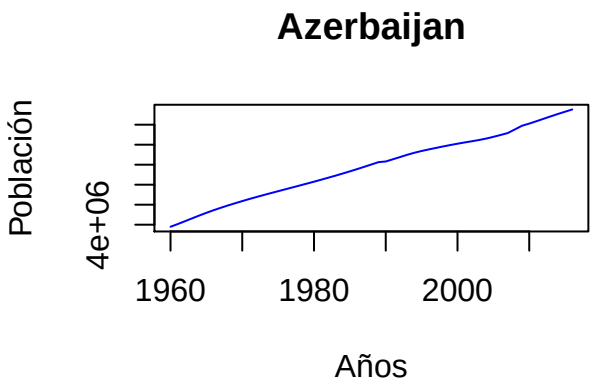

Argentina
Población
2.0e+07
1960
1980
2000
Años

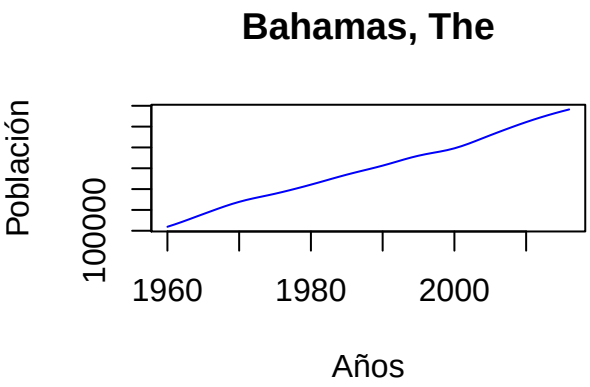

Armenia
Población
2000000
1960
1980
2000
Años

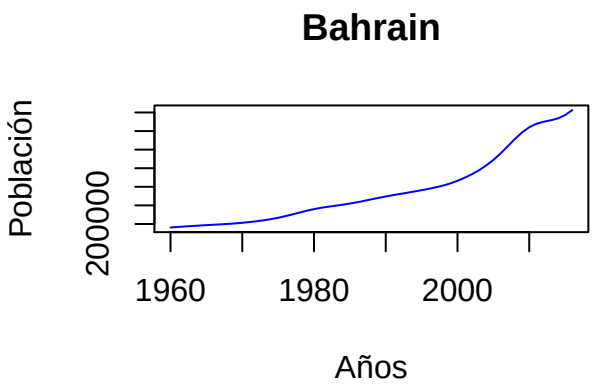

Aruba
Población
6e+04
1960
1980
2000
Años

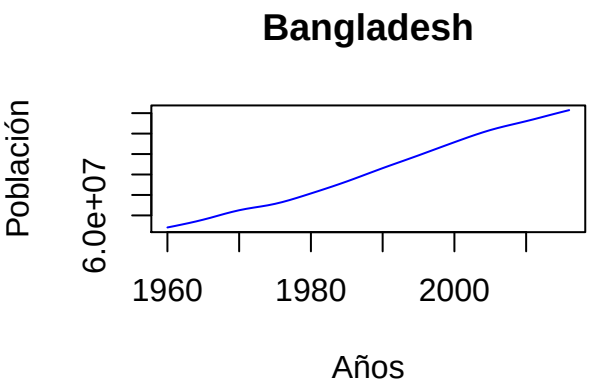

Australia
Población
1.0e+07
1960
1980
2000
Años

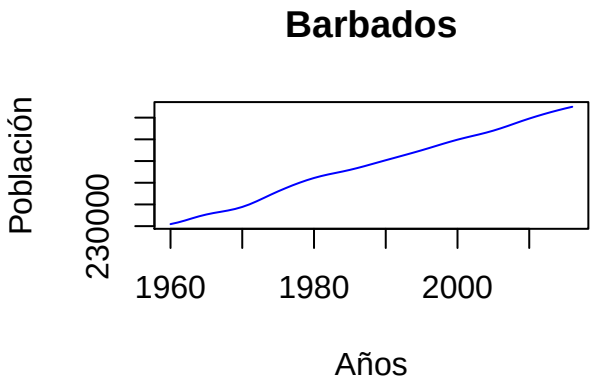

Austria
Población
7000000
1960
1980
2000
Años

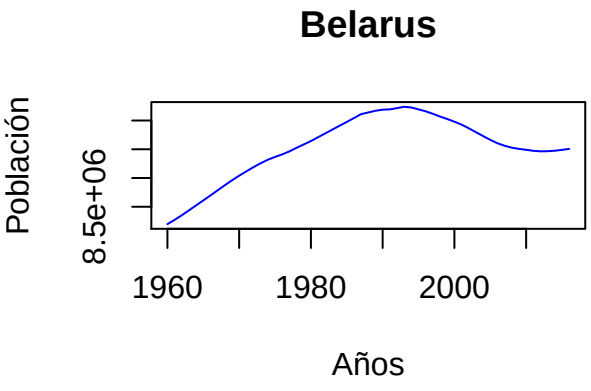

Azerbaijan
Población
4e+06
1960
1980
2000
Años

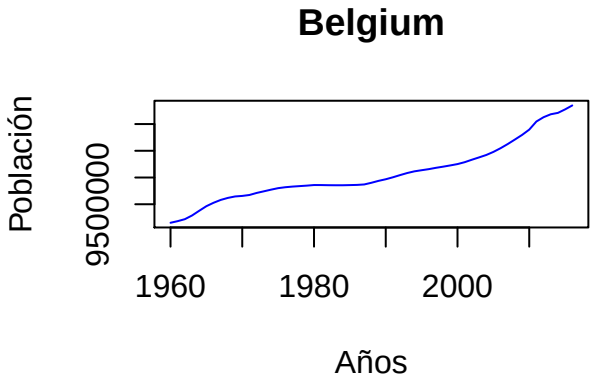

Bahamas, The
Población
100000
1960
1980
2000
Años

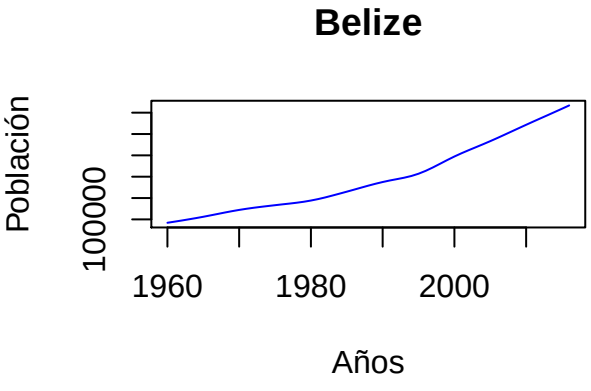

Bahrain
Población
200000
1960
1980
2000
Años

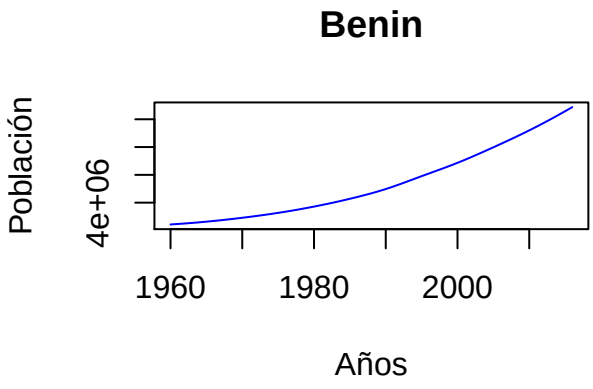

Bangladesh
Población
6.0e+07
1960
1980
2000
Años

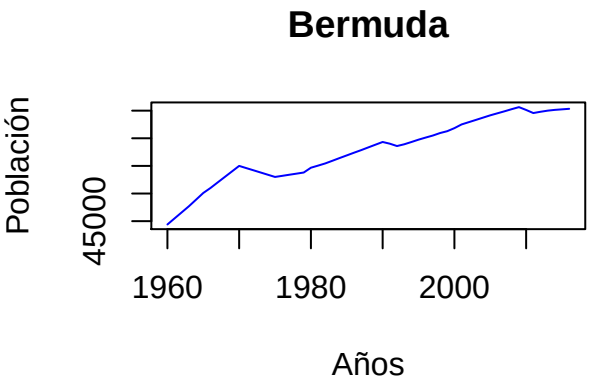

Barbados
Población
230000
1960
1980
2000
Años

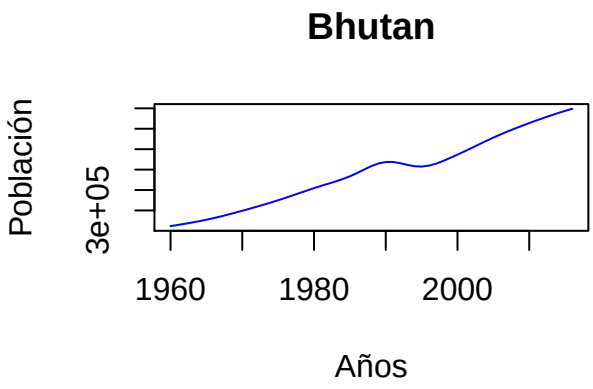

Belarus
Población
8.5e+06
1960
1980
2000
Años

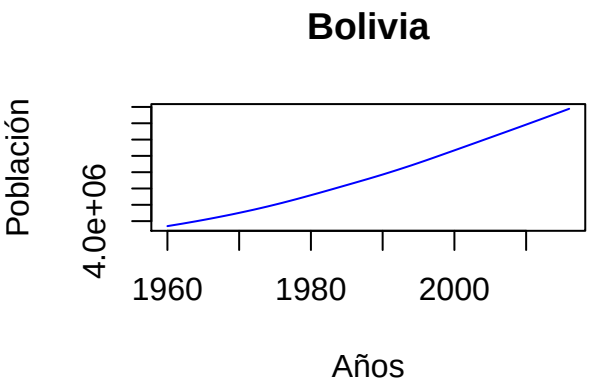

Belgium
Población
9500000
1960
1980
2000
Años

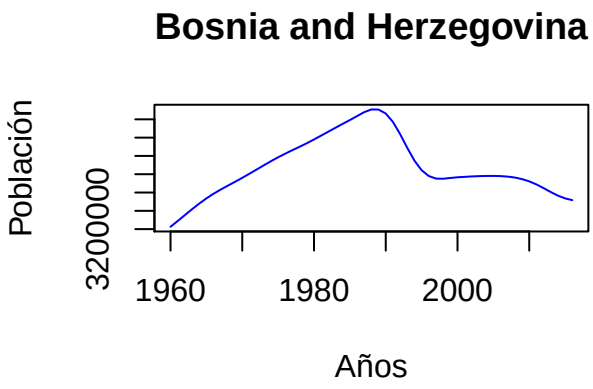

Belize
Población
100000
1960
1980
2000
Años

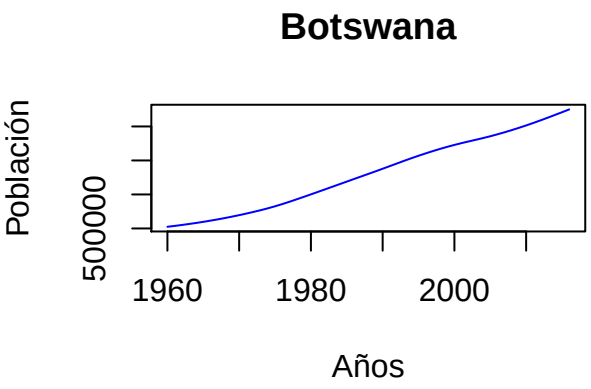

Benin
Población
4e+06
1960
1980
2000
Años

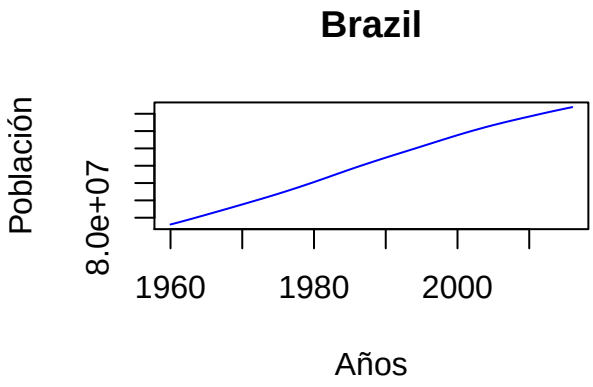

Bermuda
Población
45000
1960
1980
2000
Años

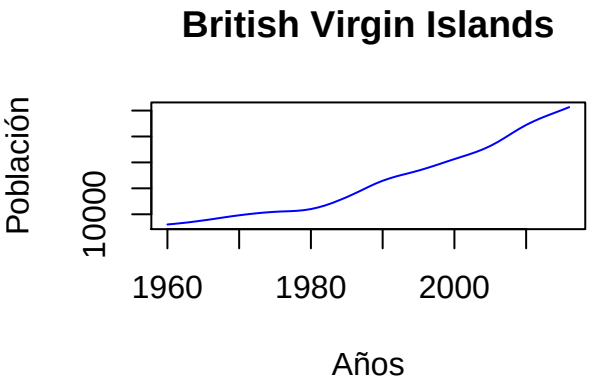

Bhutan
Población
3e+05
1960
1980
2000
Años

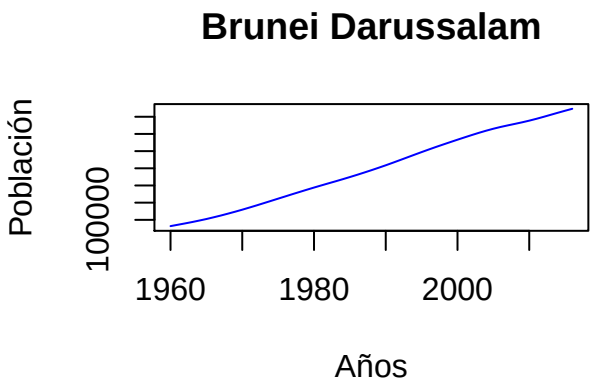

Bolivia
Población
4.0e+06
1960
1980
2000
Años

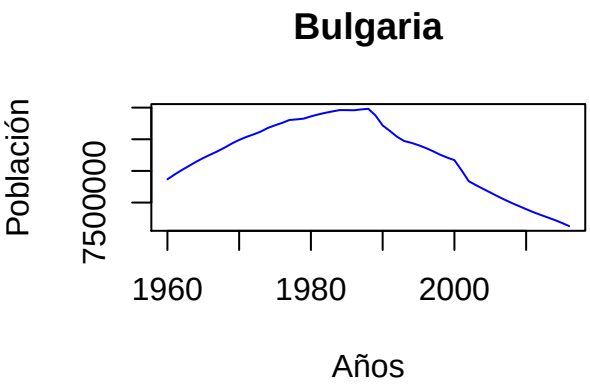

Bosnia and Herzegovina
Población
3200000
1960
1980
2000
Años

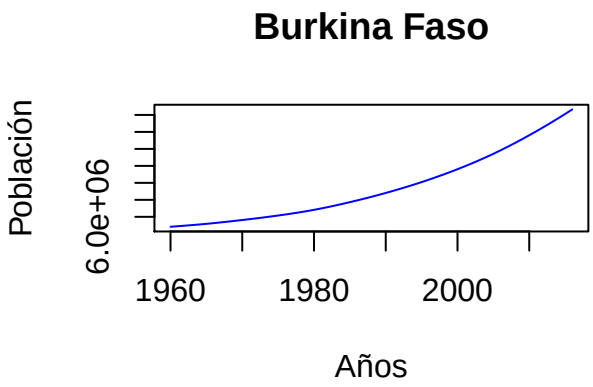

Botswana
Población
500000
1960
1980
2000
Años

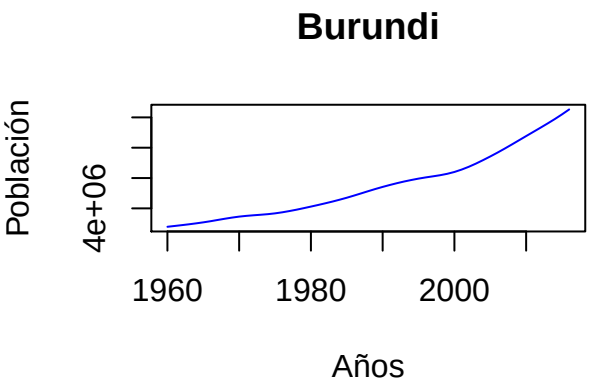

Brazil
Población
8.0e+07
1960
1980
2000
Años

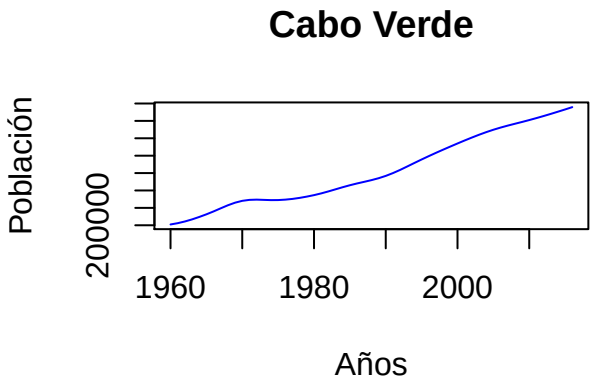

British Virgin Islands
Población
10000
1960
1980
2000
Años

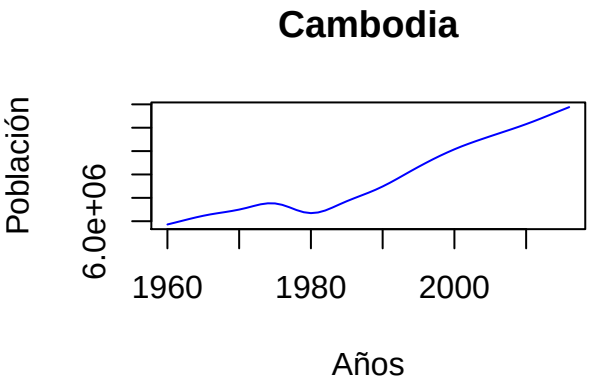

Brunei Darussalam
Población
100000
1960
1980
2000
Años

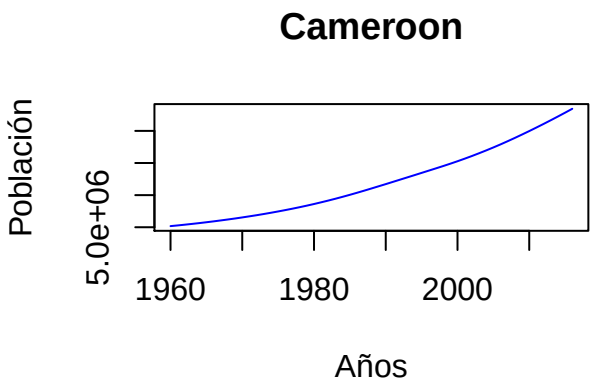

Bulgaria
Población
7500000
1960
1980
2000
Años

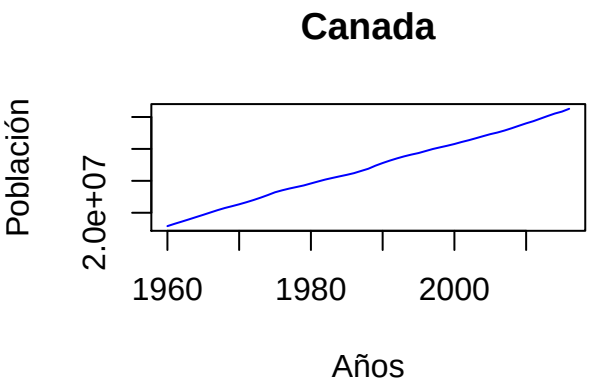

Burkina Faso
Población
6.0e+06
1960
1980
2000
Años

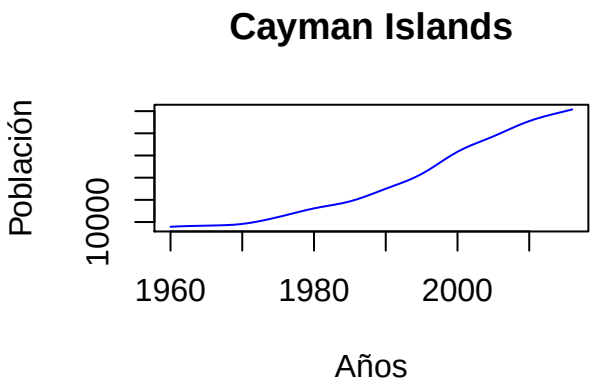

Burundi
Población
4e+06
1960
1980
2000
Años

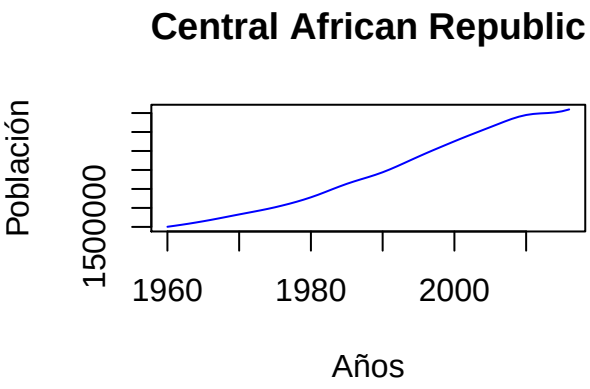

Cabo Verde
Población
200000
1960
1980
2000
Años

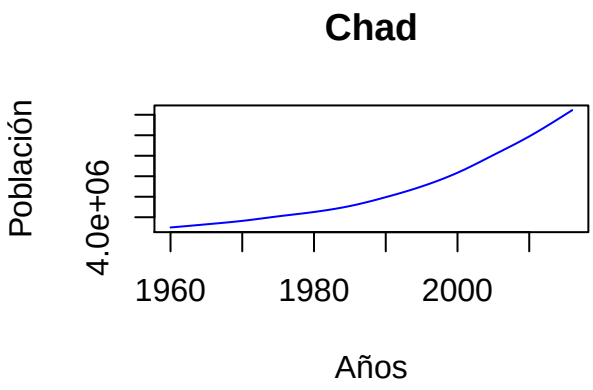

Cambodia
Población
6.0e+06
1960
1980
2000
Años

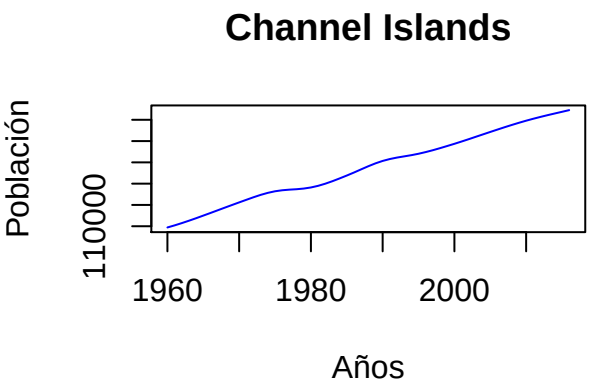

Cameroon
Población
5.0e+06
1960
1980
2000
Años

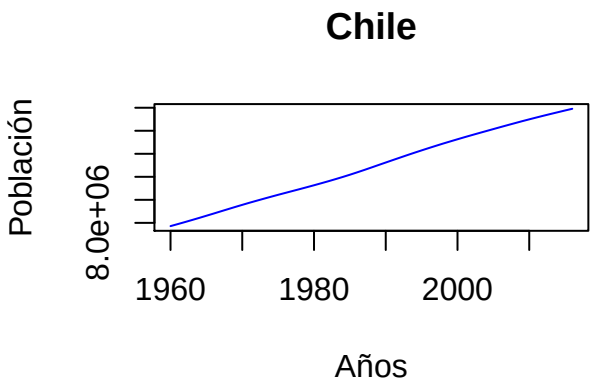

Canada
Población
2.0e+07
1960
1980
2000
Años

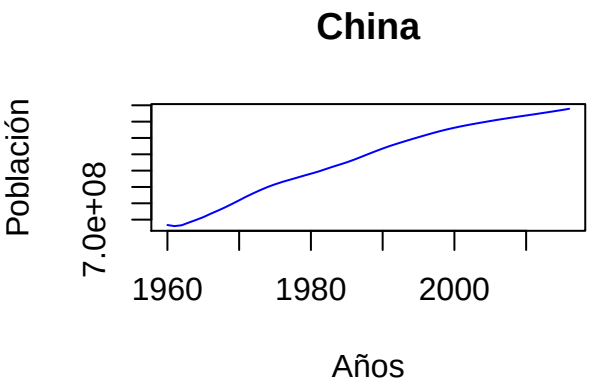

Cayman Islands
Población
10000
1960
1980
2000
Años

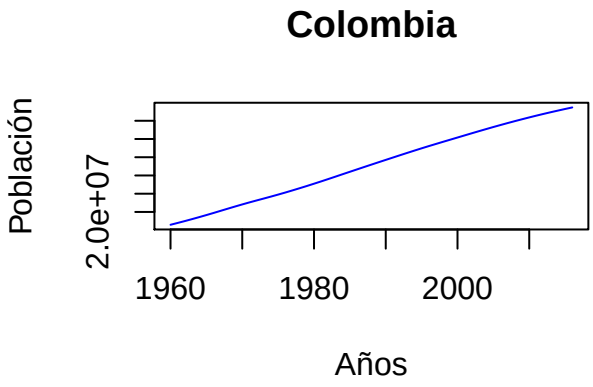

Central African Republic
Población
1500000
1960
1980
2000
Años

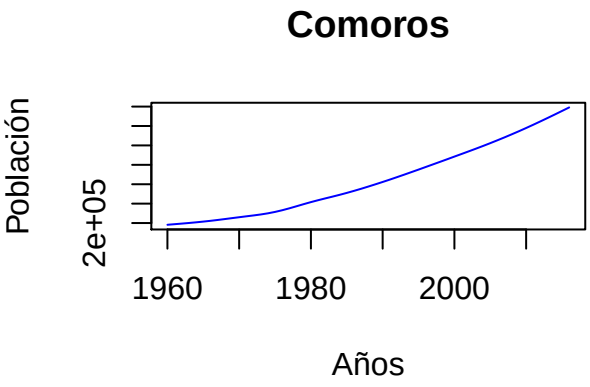

Chad
Población
4.0e+06
1960
1980
2000
Años

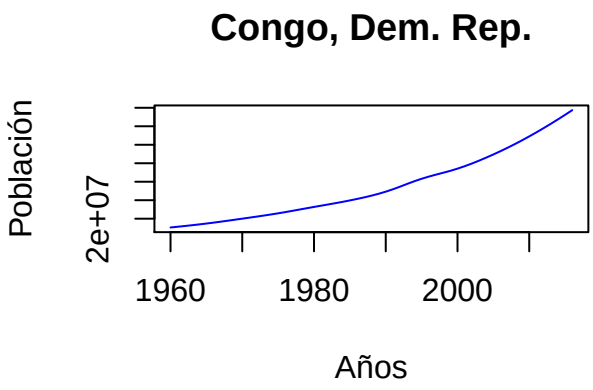

Channel Islands
Población
110000
1960
1980
2000
Años

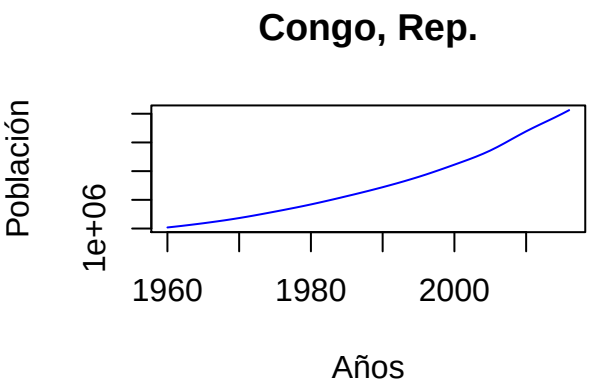

Chile
Población
8.0e+06
1960
1980
2000
Años

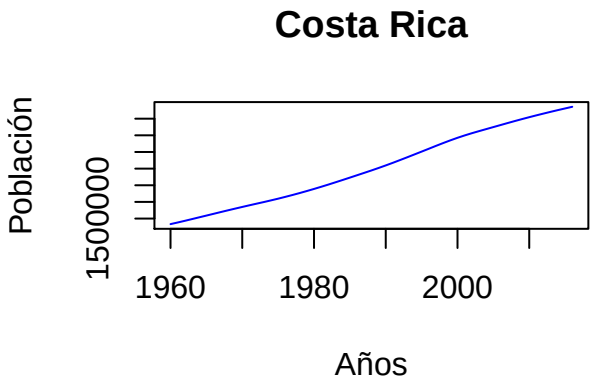

China
Población
7.0e+08
1960
1980
2000
Años

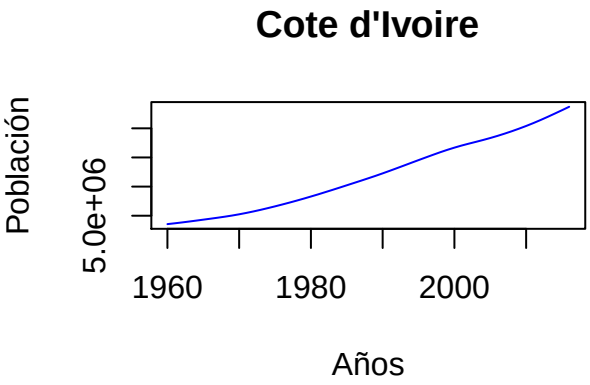

Colombia
Población
2.0e+07
1960
1980
2000
Años

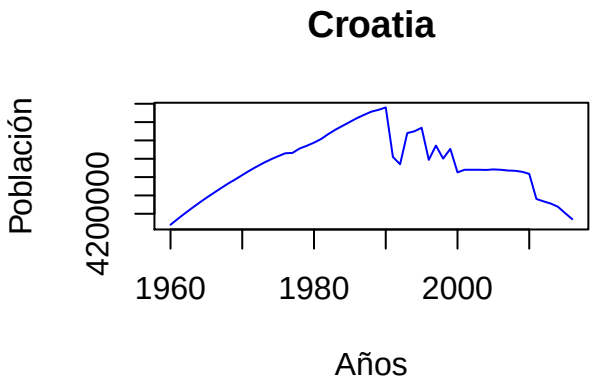

Comoros
Población
2e+05
1960
1980
2000
Años

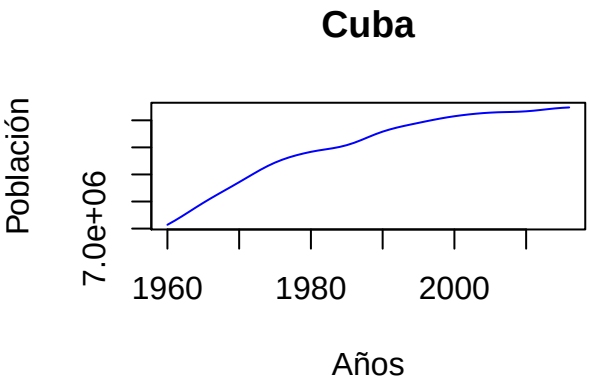

Congo, Dem. Rep.
Población
2e+07
1960
1980
2000
Años

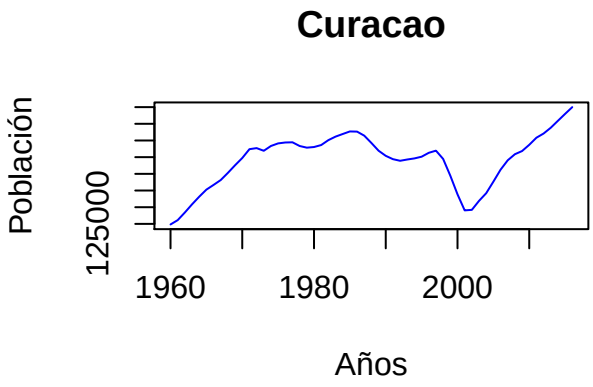

Congo, Rep.
Población
1e+06
1960
1980
2000
Años

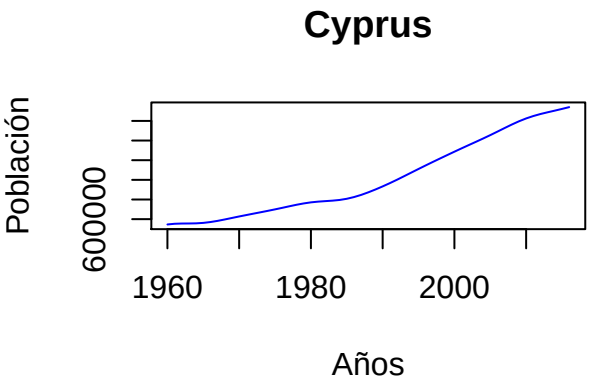

Costa Rica
Población
1500000
1960
1980
2000
Años

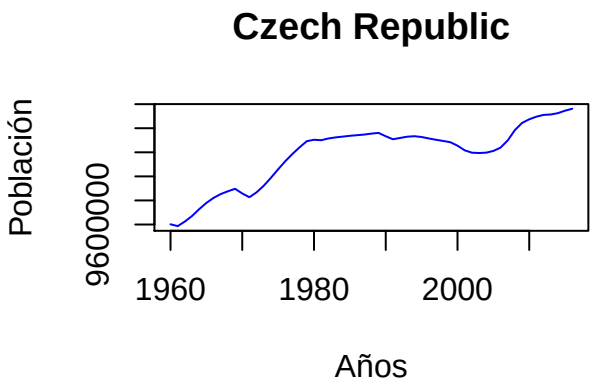

Cote d'Ivoire
Población
5.0e+06
1960
1980
2000
Años

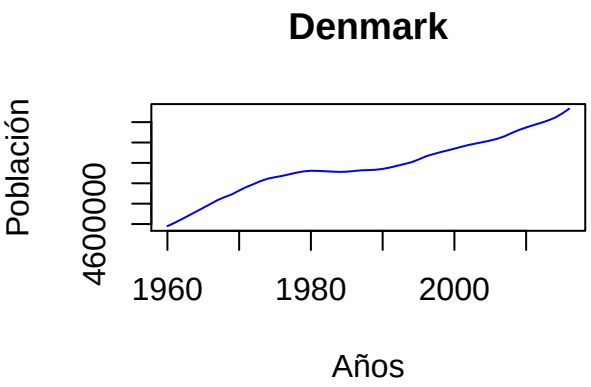

Croatia
Población
4200000
1960
1980
2000
Años

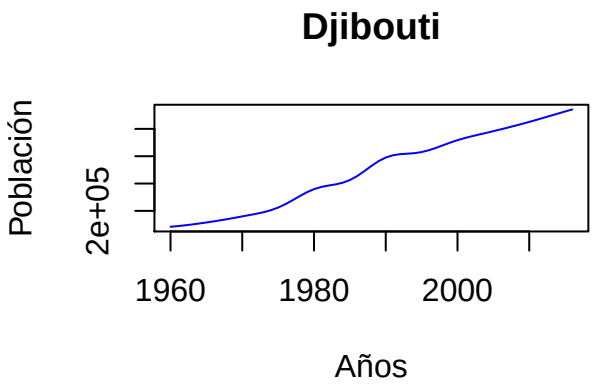

Cuba
Población
7.0e+06
1960
1980
2000
Años

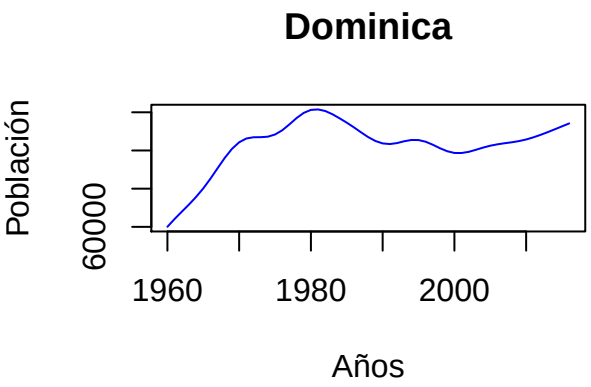

Curacao
Población
125000
1960
1980
2000
Años

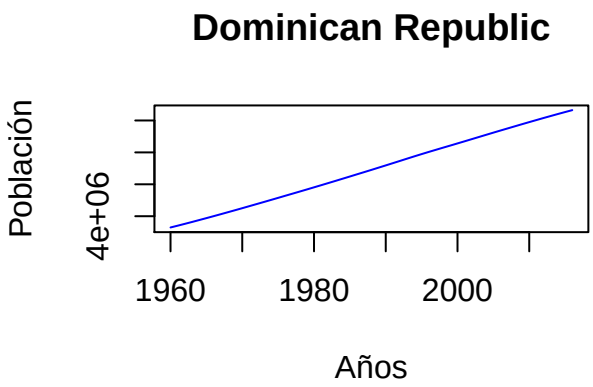

Cyprus
Población
600000
1960
1980
2000
Años

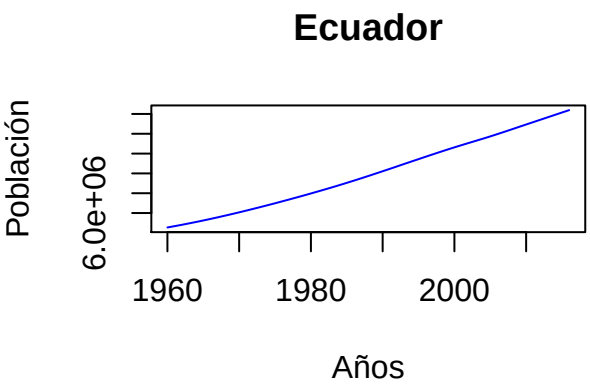

Czech Republic
Población
9600000
1960
1980
2000
Años

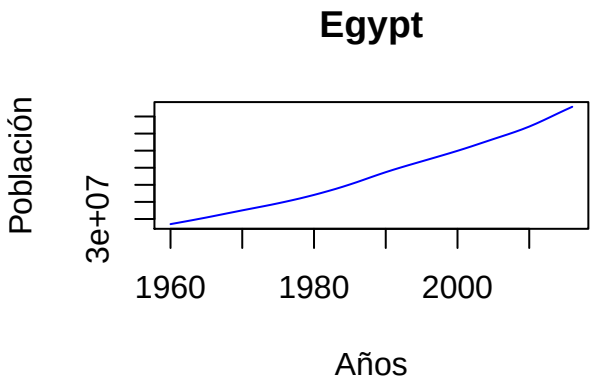

Denmark
Población
4600000
1960
1980
2000
Años

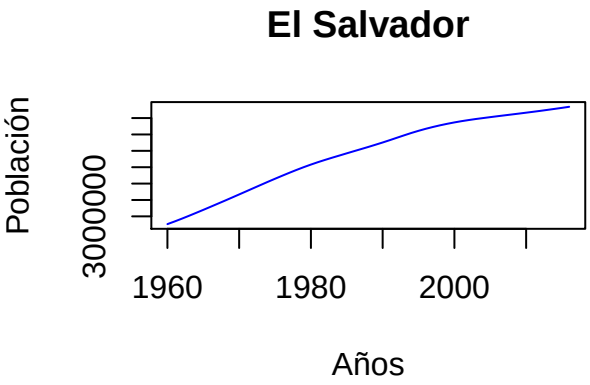

Djibouti
Población
2e+05
1960
1980
2000
Años

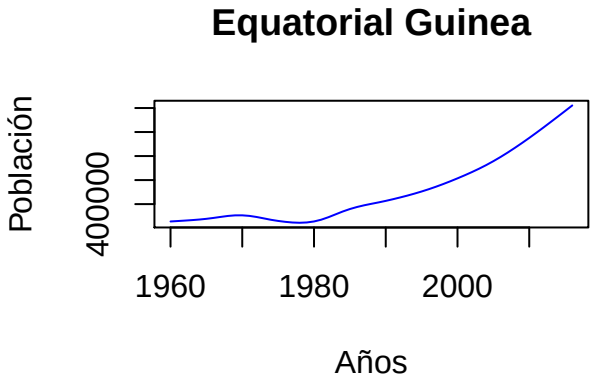

Dominica
Población
60000
1960
1980
2000
Años

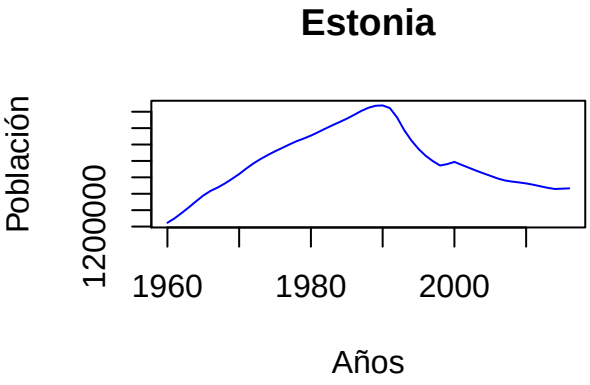

Dominican Republic
Población
4e+06
1960
1980
2000
Años

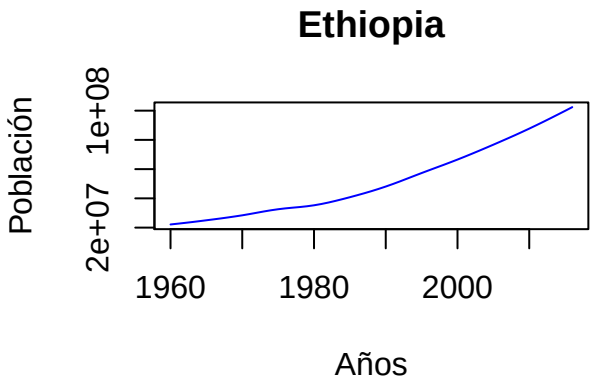

Ecuador
Población
6.0e+06
1960
1980
2000
Años

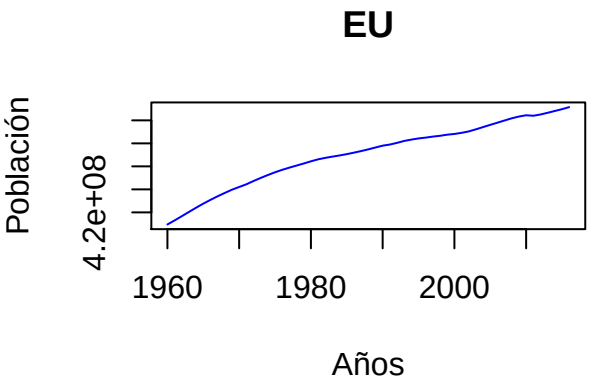

Egypt
Población
3e+07
1960
1980
2000
Años

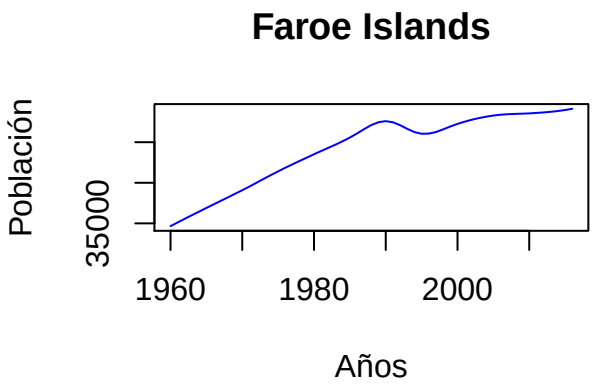

El Salvador
Población
3000000
1960
1980
2000
Años

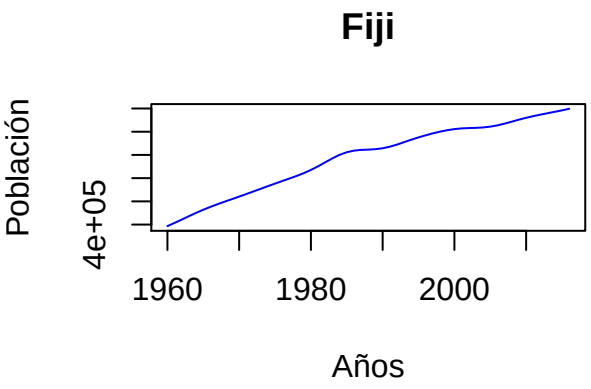

Equatorial Guinea
Población
400000
1960
1980
2000
Años

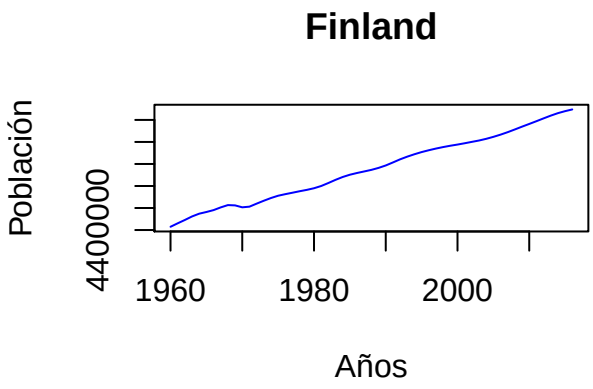

Estonia
Población
1200000
1960
1980
2000
Años

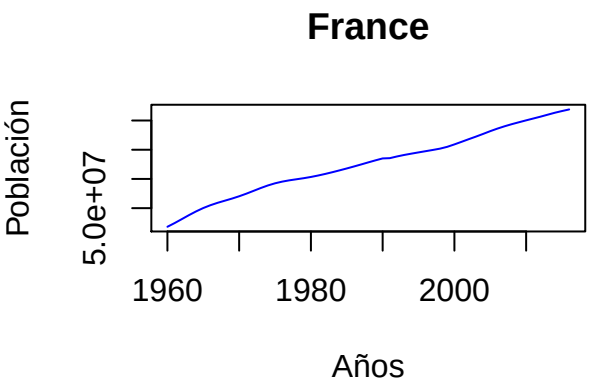

Ethiopia
Población
2e+07
1e+08
1960
1980
2000
Años

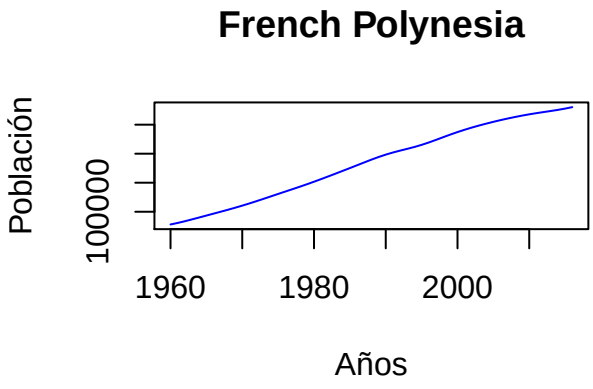

EU
Población
4.2e+08
1960
1980
2000
Años

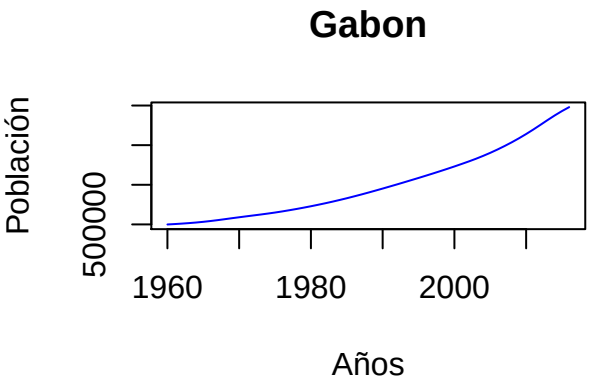

Faroe Islands
Población
35000
1960
1980
2000
Años

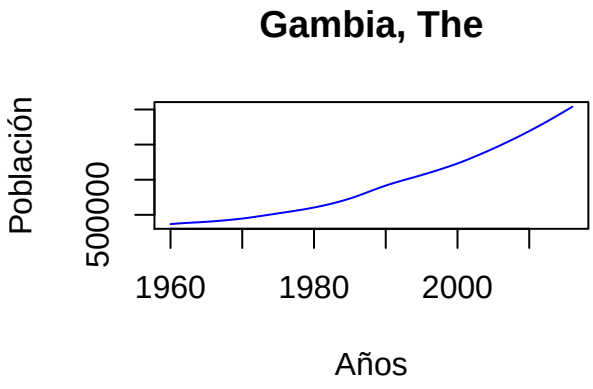

Fiji
Población
4e+05
1960
1980
2000
Años

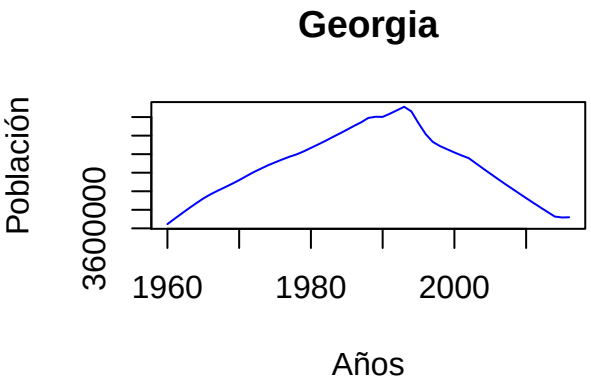

Finland
Población
4400000
1960
1980
2000
Años

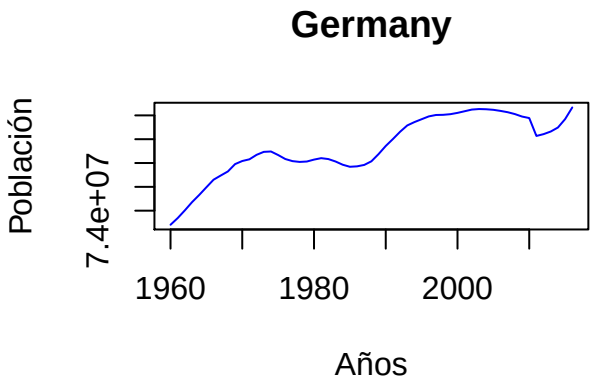

France
Población
5.0e+07
1960
1980
2000
Años

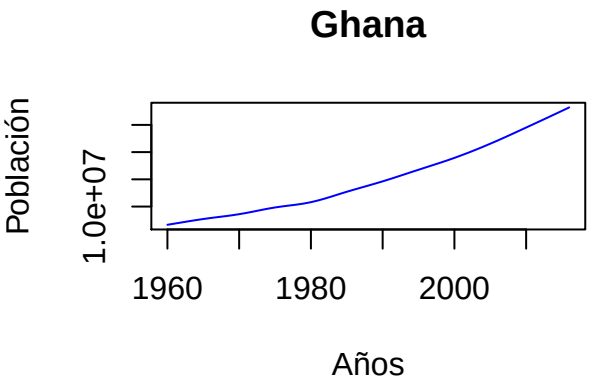

French Polynesia
Población
100000
1960
1980
2000
Años

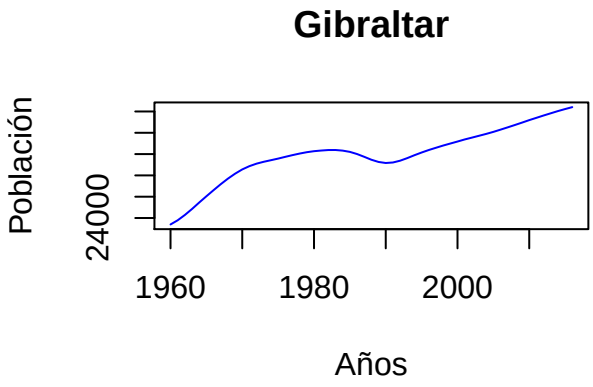

Gabon
Población
500000
1960
1980
2000
Años

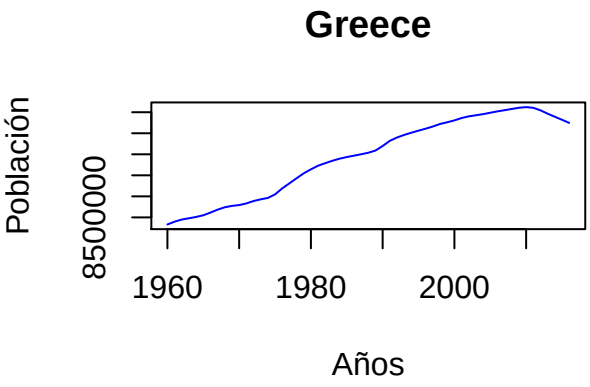

Gambia, The
Población
500000
1960
1980
2000
Años

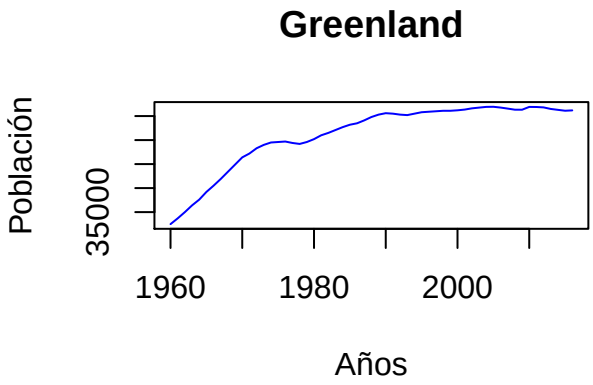

Georgia
Población
3600000
1960
1980
2000
Años

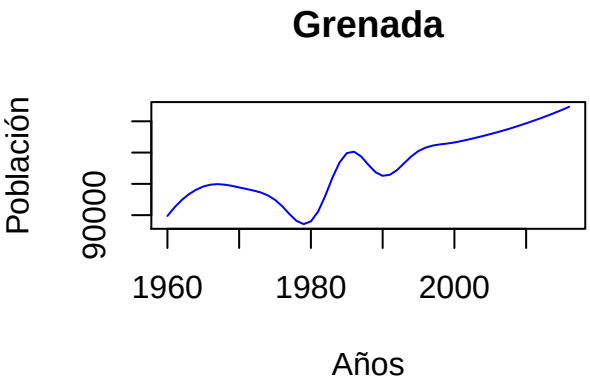

Germany
Población
7.4e+07
1960
1980
2000
Años

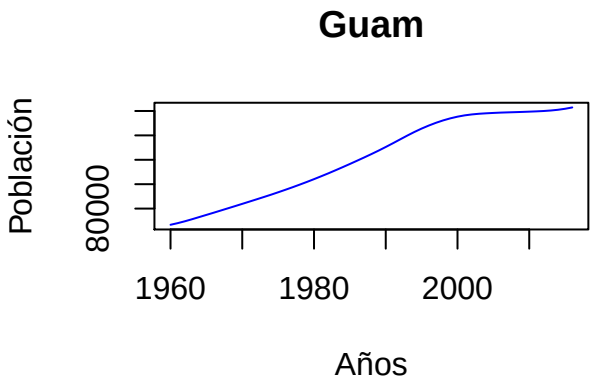

Ghana
Población
1.0e+07
1960
1980
2000
Años

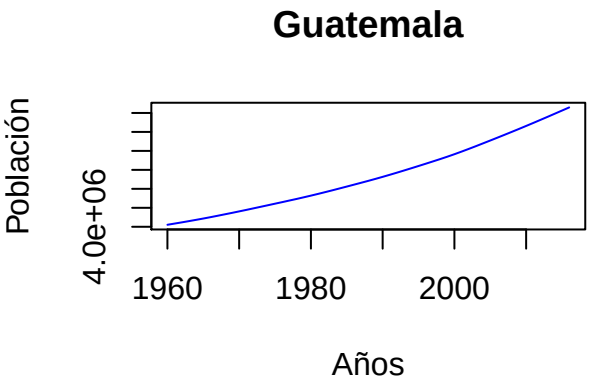

Gibraltar
Población
24000
1960
1980
2000
Años

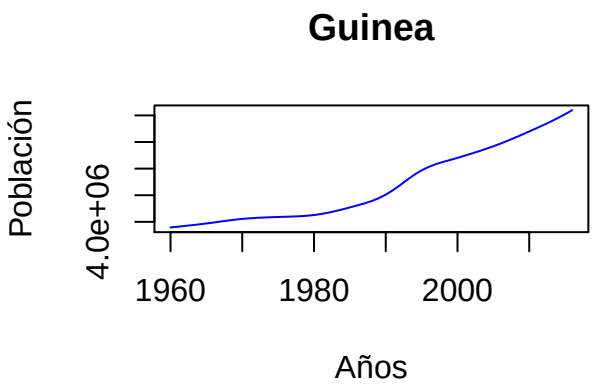

Greece
Población
8500000
1960
1980
2000
Años

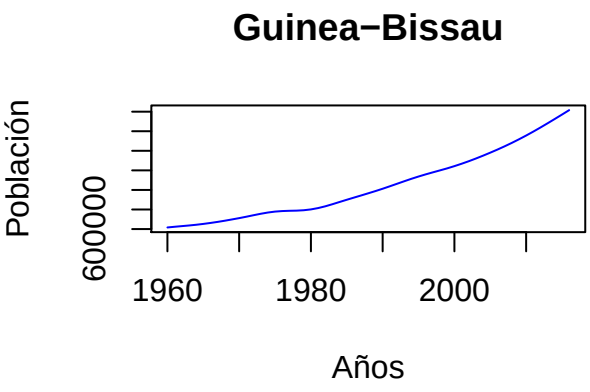

Greenland
Población
35000
1960
1980
2000
Años

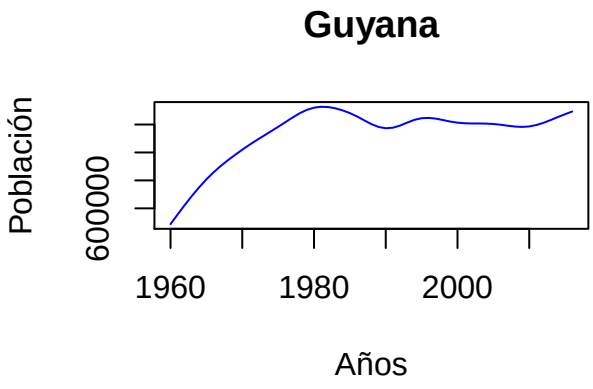

Grenada
Población
90000
1960
1980
2000
Años

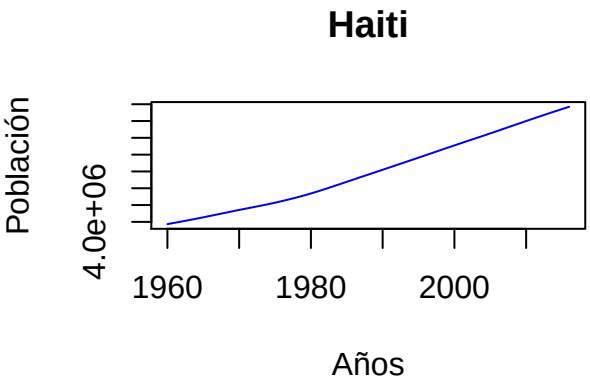

Guam
Población
80000
1960
1980
2000
Años

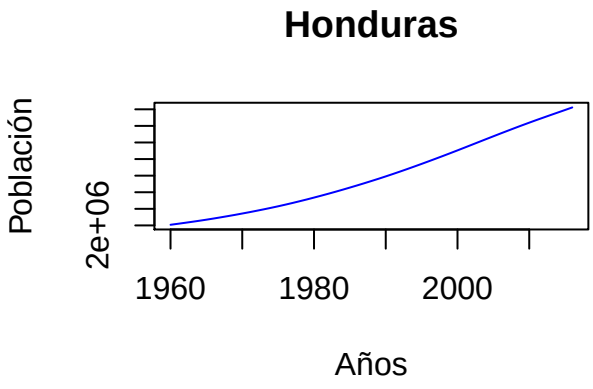

Guatemala
Población
4.0e+06
1960
1980
2000
Años

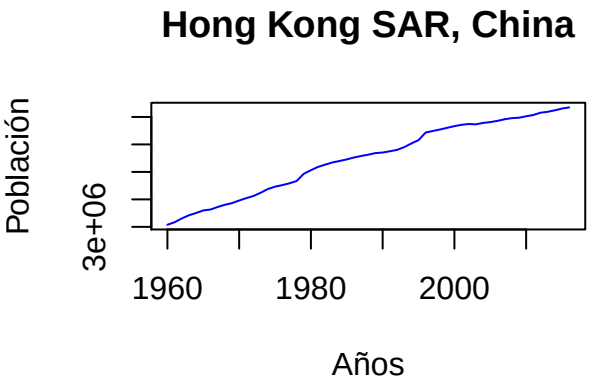

Guinea
Población
4.0e+06
1960
1980
2000
Años

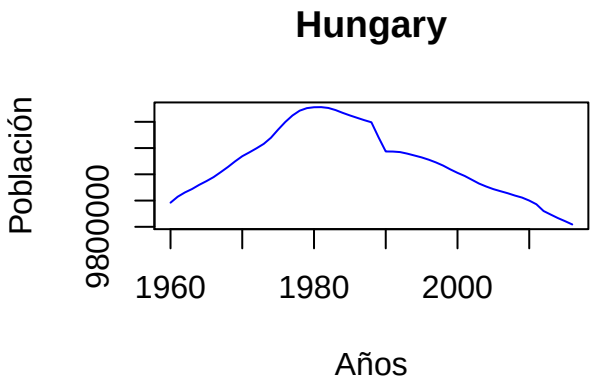

Guinea−Bissau
Población
600000
1960
1980
2000
Años

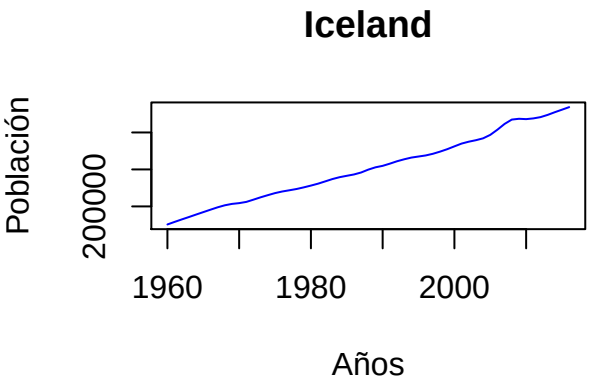

Guyana
Población
600000
1960
1980
2000
Años

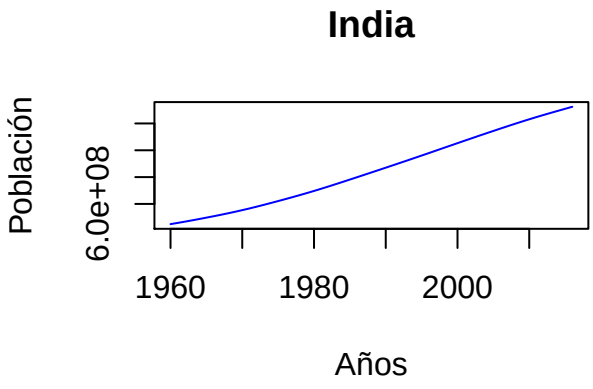

Haiti
Población
4.0e+06
1960
1980
2000
Años

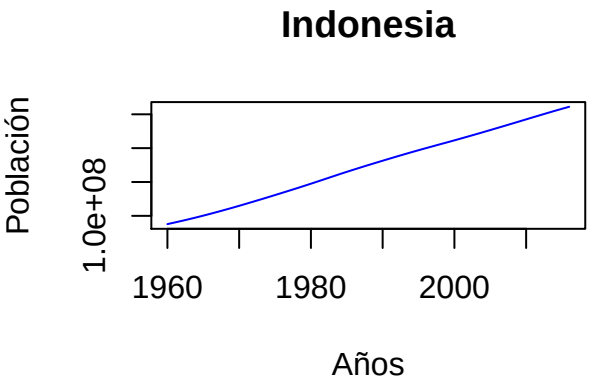

Honduras
Población
2e+06
1960
1980
2000
Años

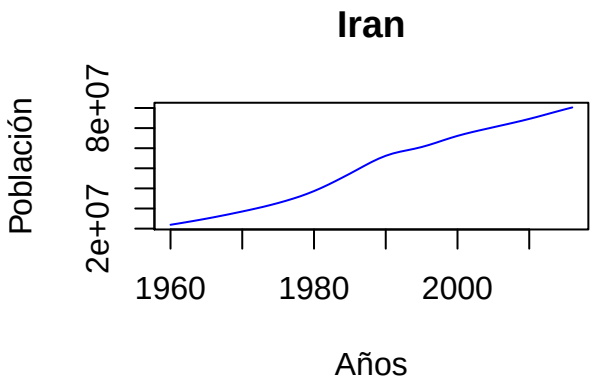

Hong Kong SAR, China
Población
3e+06
1960
1980
2000
Años

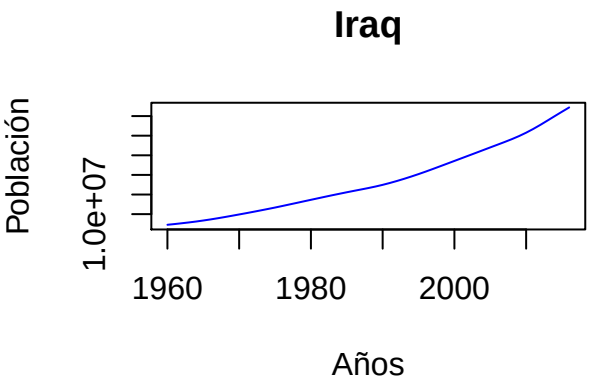

Hungary
Población
9800000
1960
1980
2000
Años

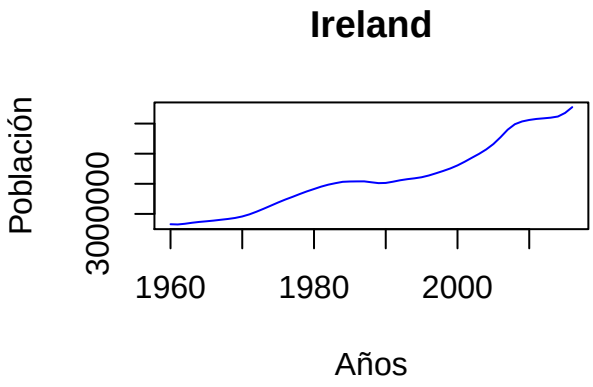

Iceland
Población
200000
1960
1980
2000
Años

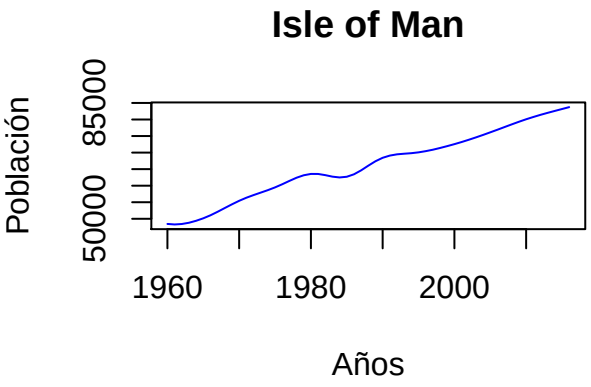

India
Población
6.0e+08
1960
1980
2000
Años

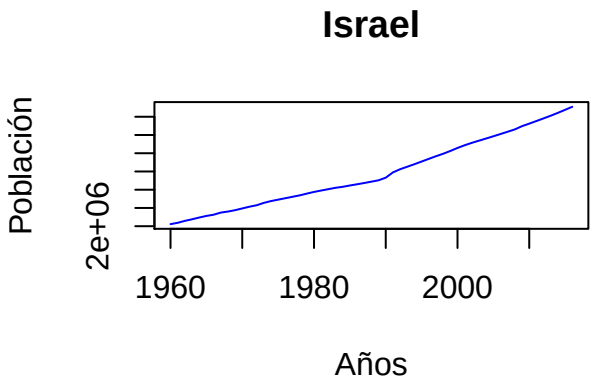

Indonesia
Población
1.0e+08
1960
1980
2000
Años

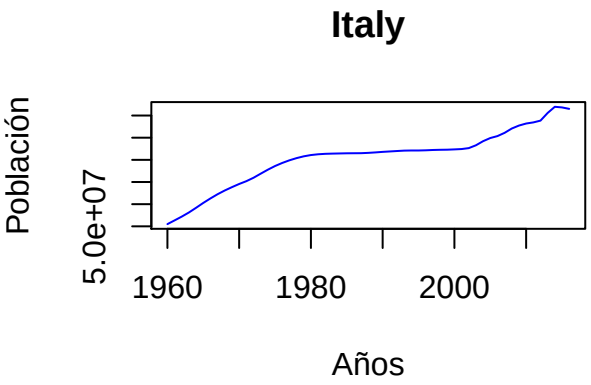

Iran
Población
2e+07
8e+07
1960
1980
2000
Años

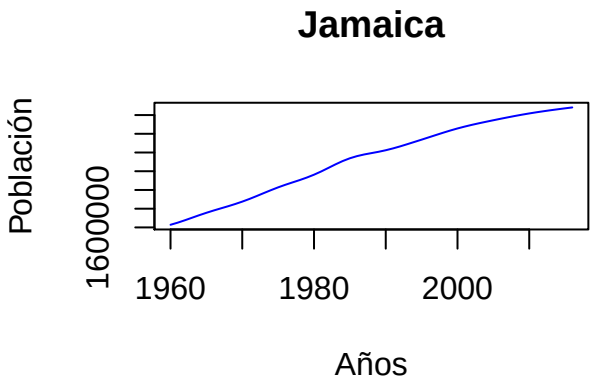

Iraq
Población
1.0e+07
1960
1980
2000
Años

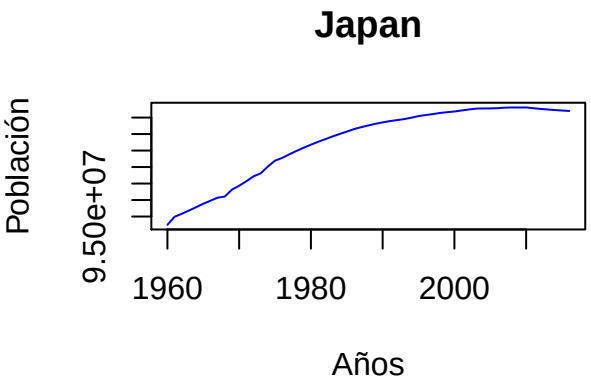

Ireland
Población
3000000
1960
1980
2000
Años

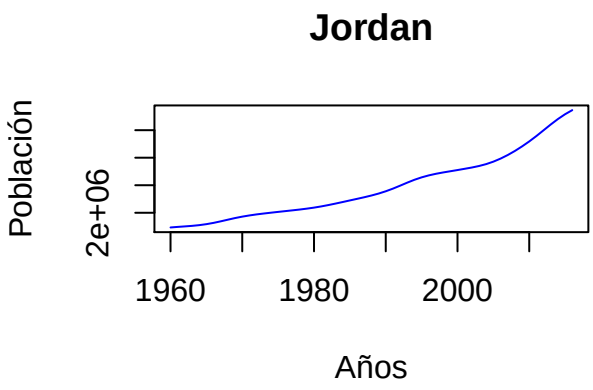

Isle of Man
Población
50000
85000
1960
1980
2000
Años

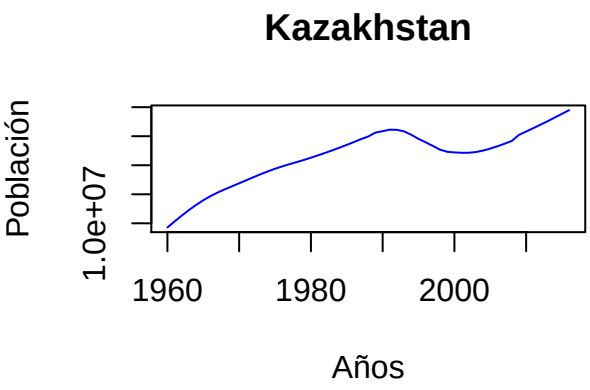

Israel
Población
2e+06
1960
1980
2000
Años

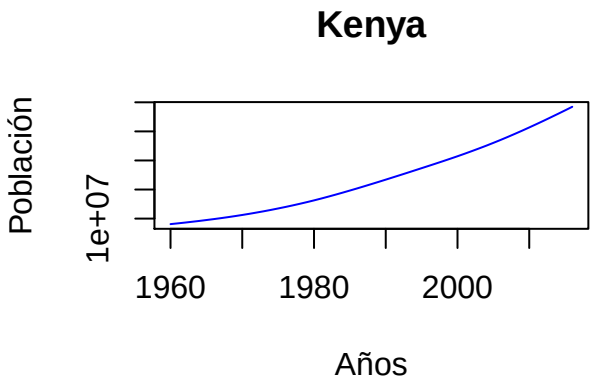

Italy
Población
5.0e+07
1960
1980
2000
Años

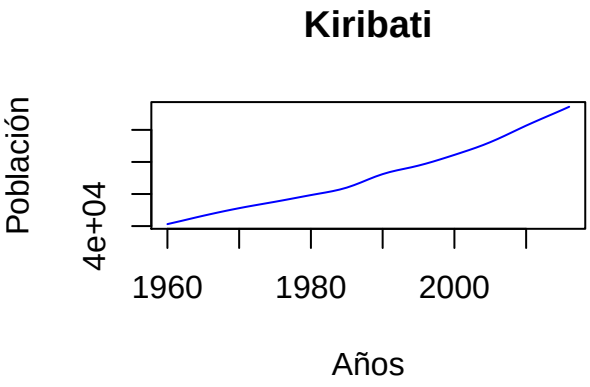

Jamaica
Población
1600000
1960
1980
2000
Años

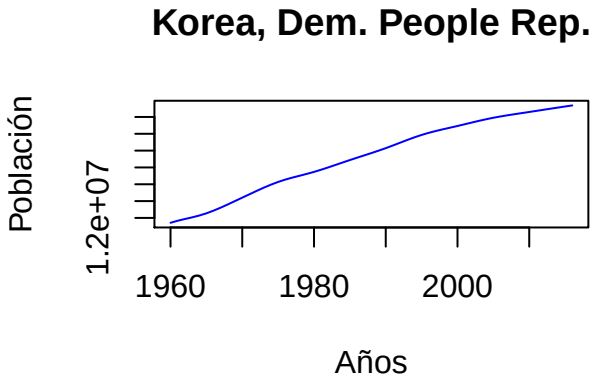

Japan
Población
9.50e+07
1960
1980
2000
Años

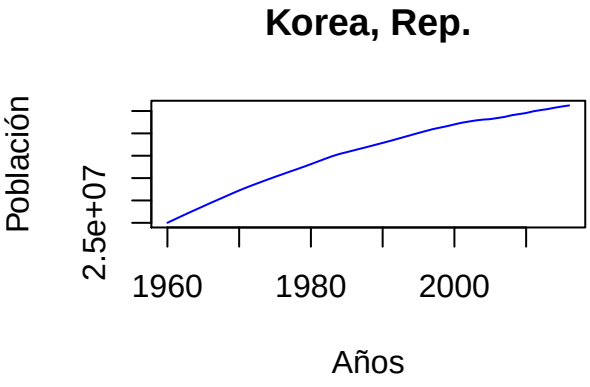

Jordan
Población
2e+06
1960
1980
2000
Años

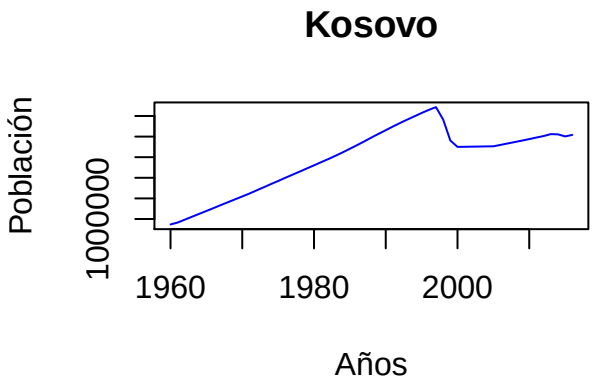

Kazakhstan
Población
1.0e+07
1960
1980
2000
Años

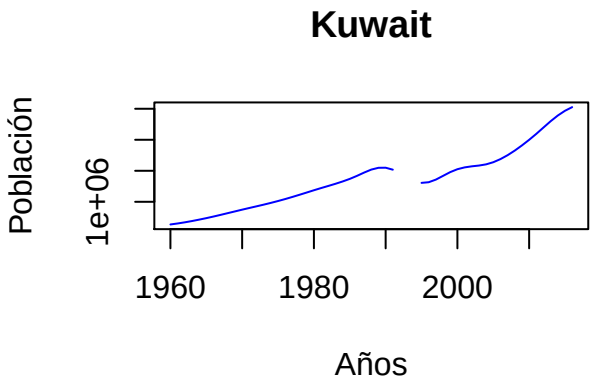

Kenya
Población
1e+07
1960
1980
2000
Años

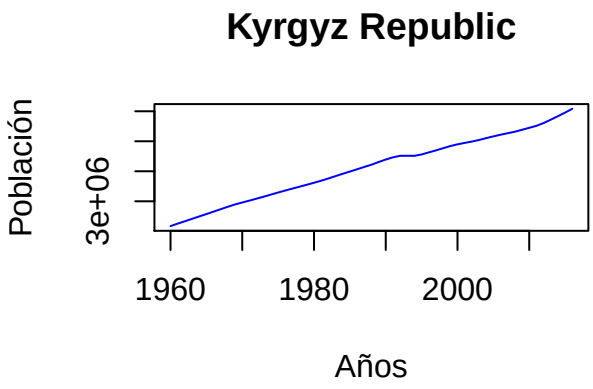

Kiribati
Población
4e+04
1960
1980
2000
Años

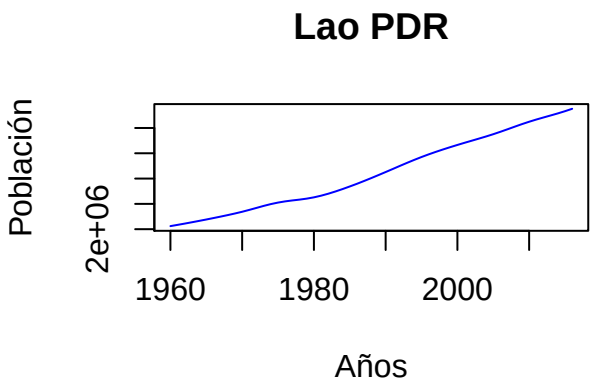

Korea, Dem. People Rep.
Población
1.2e+07
1960
1980
2000
Años

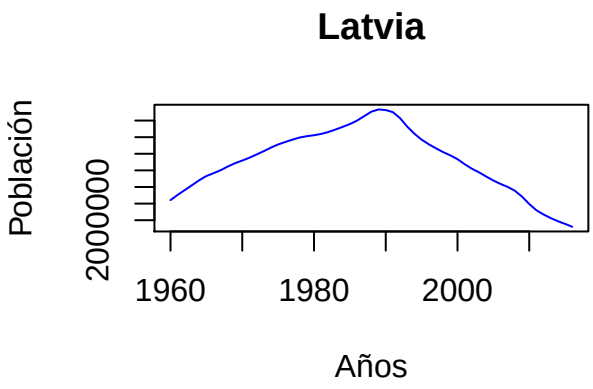

Korea, Rep.
Población
2.5e+07
1960
1980
2000
Años

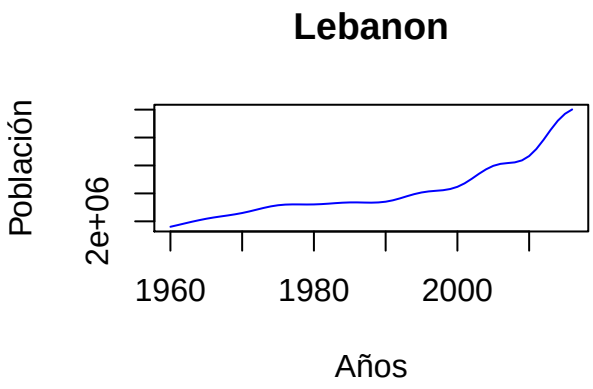

Kosovo
Población
1000000
1960
1980
2000
Años

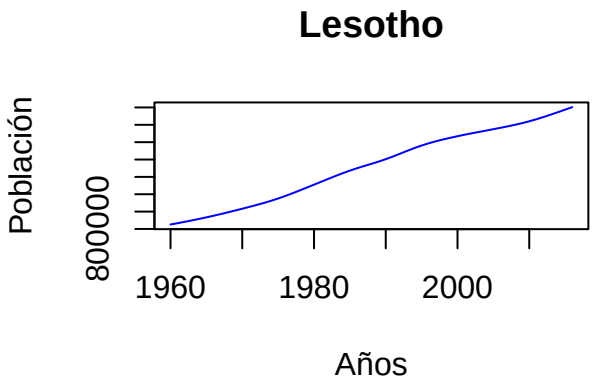

Kuwait
Población
1e+06
1960
1980
2000
Años

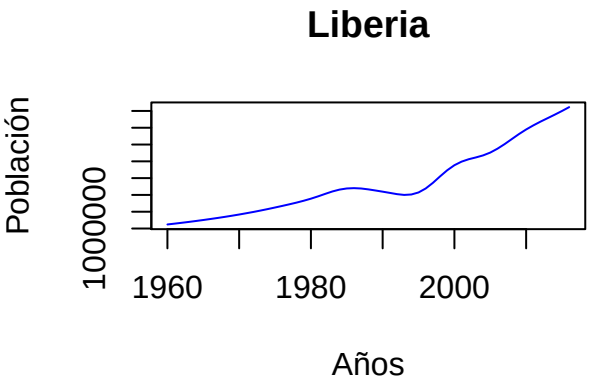

Kyrgyz Republic
Población
3e+06
1960
1980
2000
Años

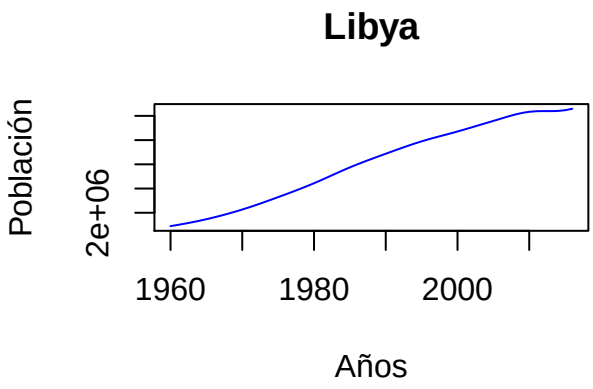

Lao PDR
Población
2e+06
1960
1980
2000
Años

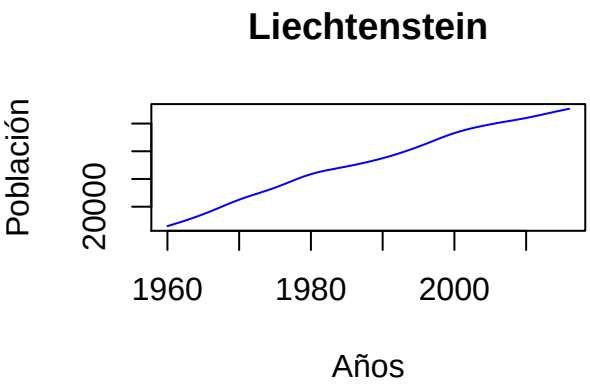

Latvia
Población
2000000
1960
1980
2000
Años

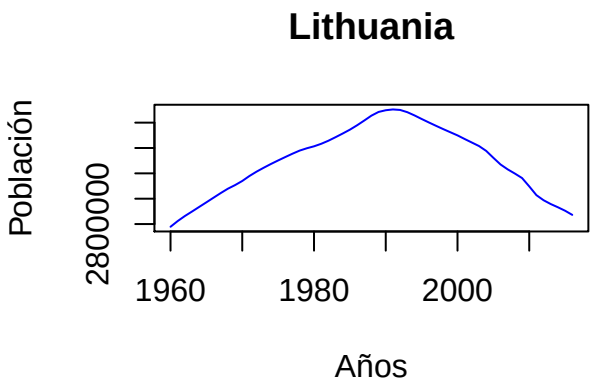

Lebanon
Población
2e+06
1960
1980
2000
Años

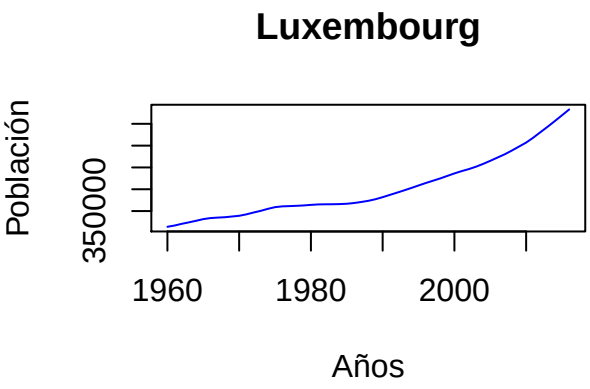

Lesotho
Población
800000
1960
1980
2000
Años

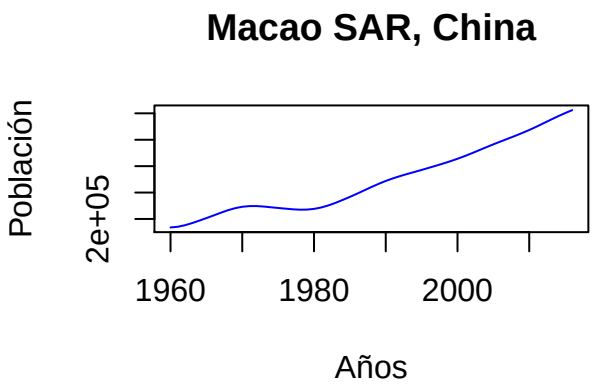

Liberia
Población
1000000
1960
1980
2000
Años

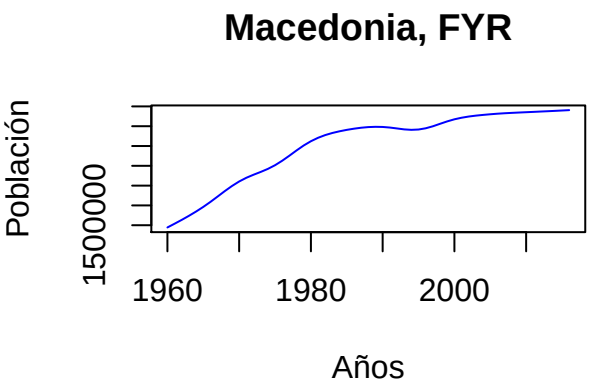

Libya
Población
2e+06
1960
1980
2000
Años

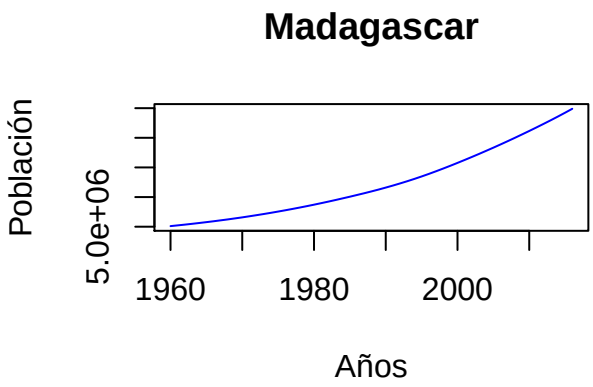

Liechtenstein
Población
20000
1960
1980
2000
Años

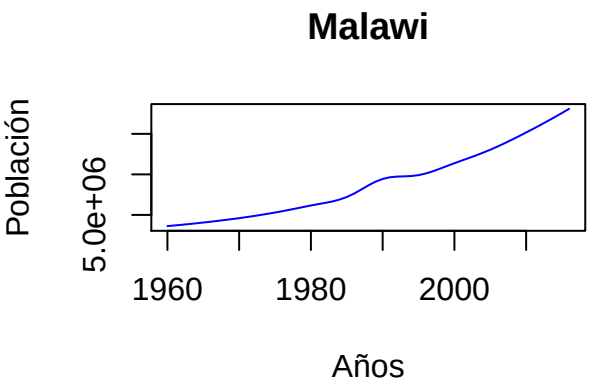

Lithuania
Población
2800000
1960
1980
2000
Años

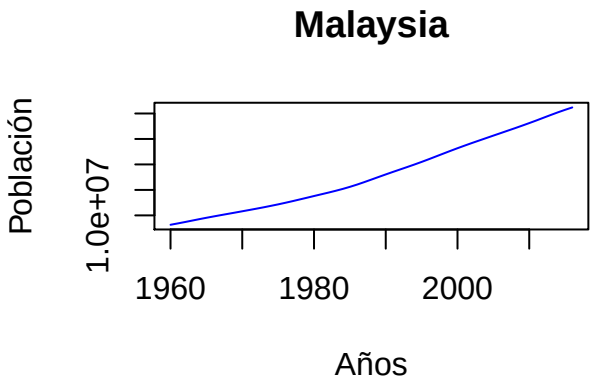

Luxembourg
Población
350000
1960
1980
2000
Años

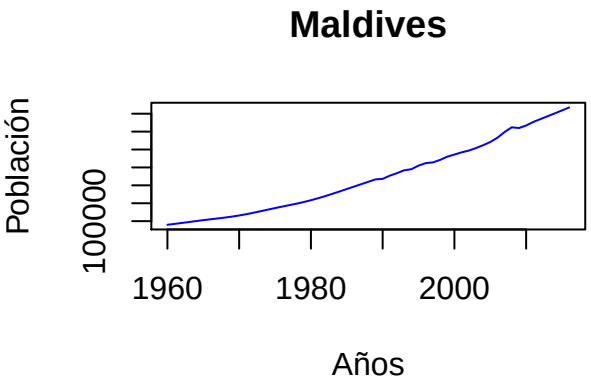

Macao SAR, China
Población
2e+05
1960
1980
2000
Años

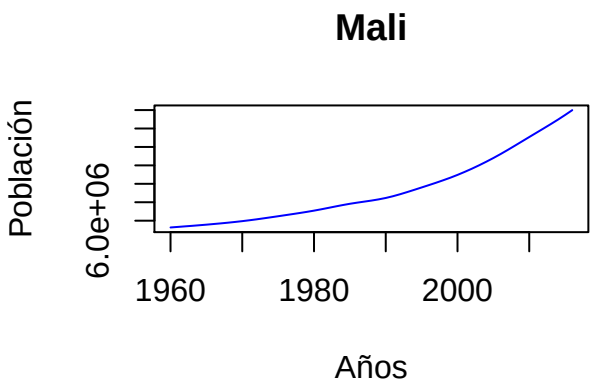

Macedonia, FYR
Población
1500000
1960
1980
2000
Años

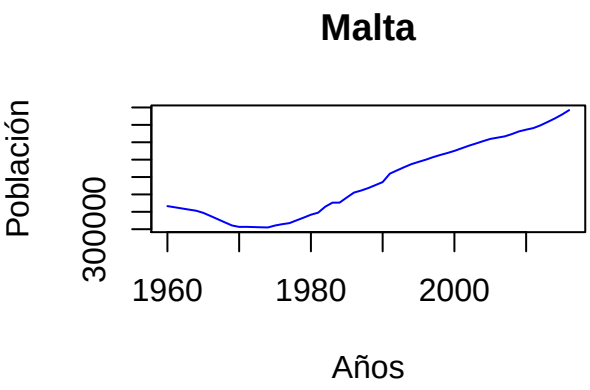

Madagascar
Población
5.0e+06
1960
1980
2000
Años

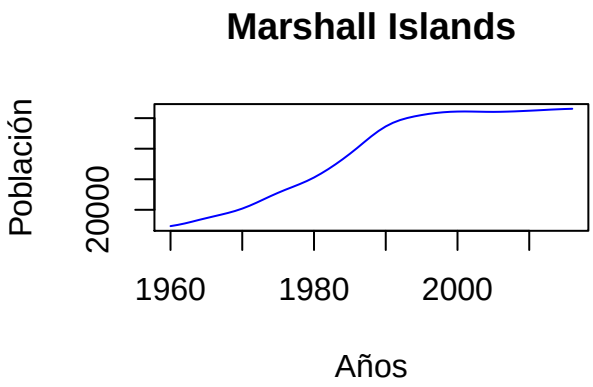

Malawi
Población
5.0e+06
1960
1980
2000
Años

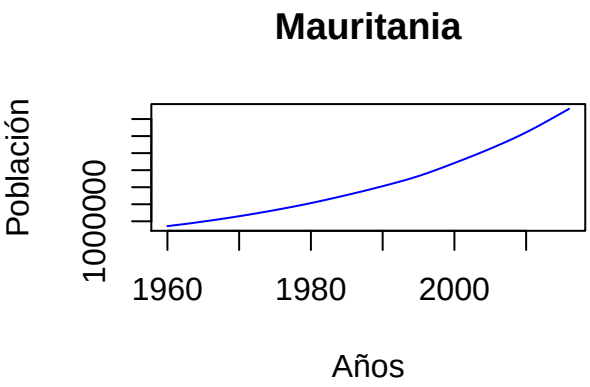

Malaysia
Población
1.0e+07
1960
1980
2000
Años

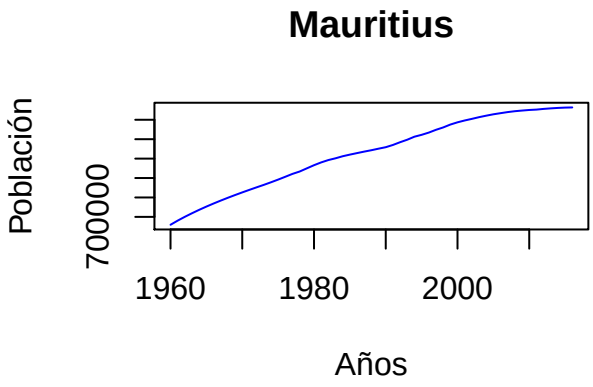

Maldives
Población
100000
1960
1980
2000
Años

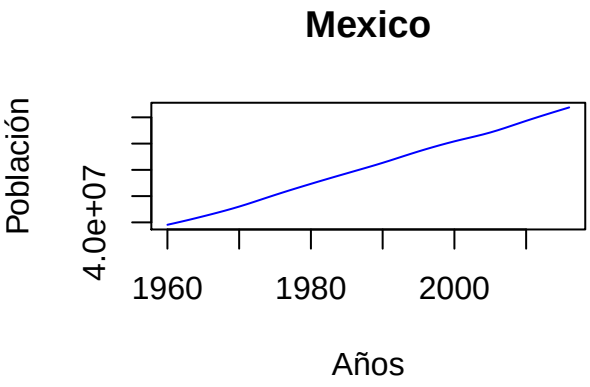

Mali
Población
6.0e+06
1960
1980
2000
Años

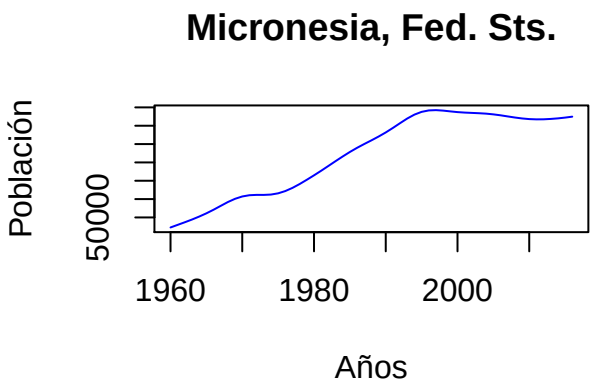

Malta
Población
300000
1960
1980
2000
Años

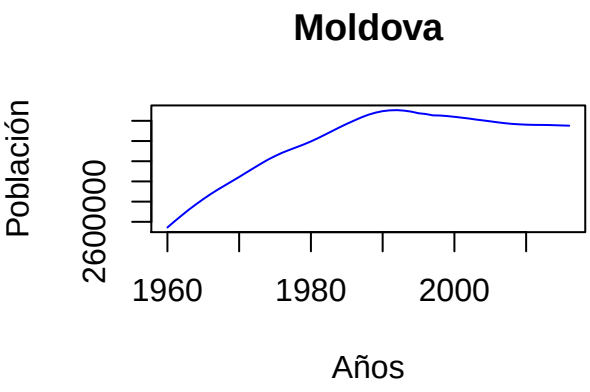

Marshall Islands
Población
20000
1960
1980
2000
Años

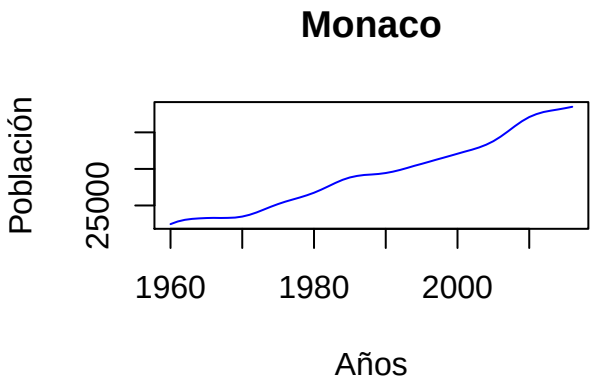

Mauritania
Población
1000000
1960
1980
2000
Años

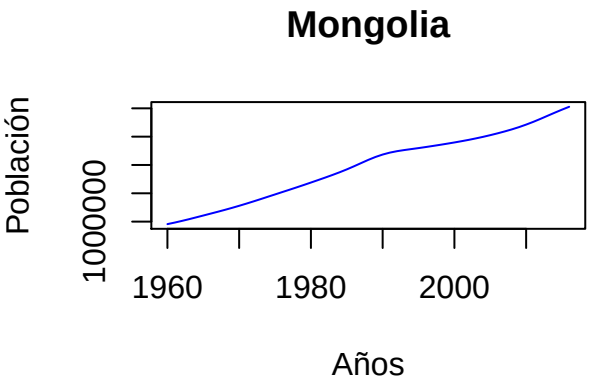

Mauritius
Población
700000
1960
1980
2000
Años

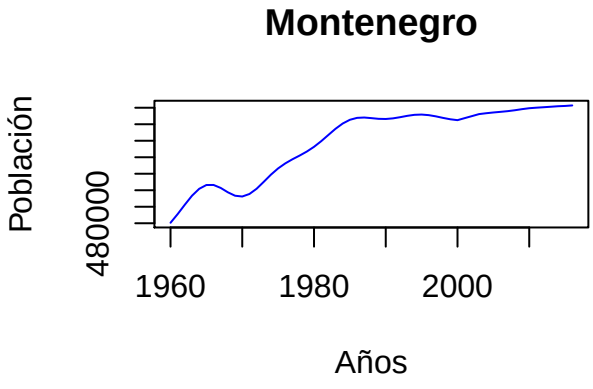

Mexico
Población
4.0e+07
1960
1980
2000
Años

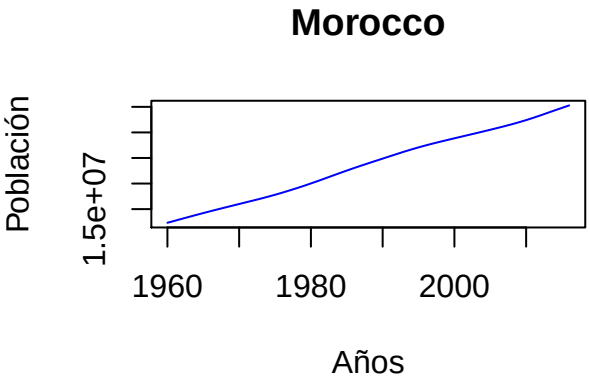

Micronesia, Fed. Sts.
Población
50000
1960
1980
2000
Años

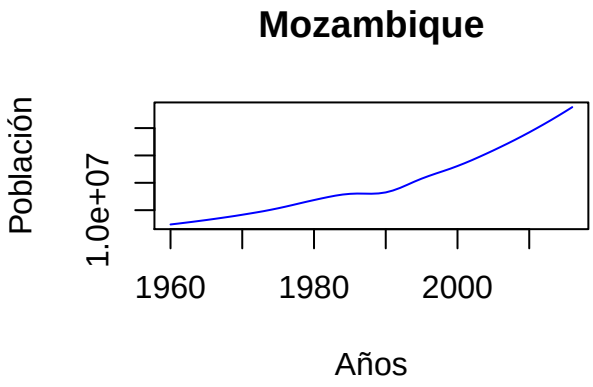

Moldova
Población
2600000
1960
1980
2000
Años

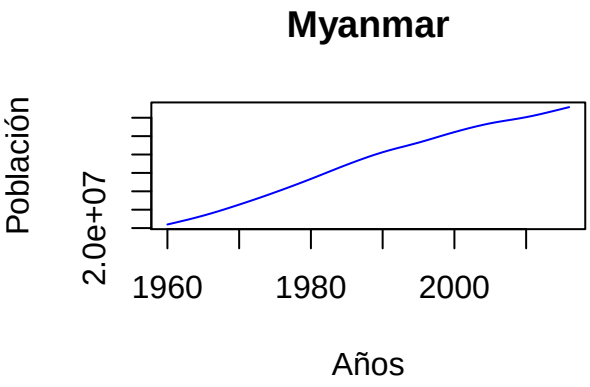

Monaco
Población
25000
1960
1980
2000
Años

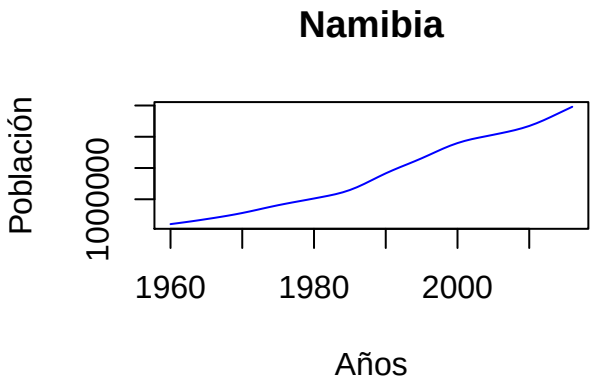

Mongolia
Población
1000000
1960
1980
2000
Años

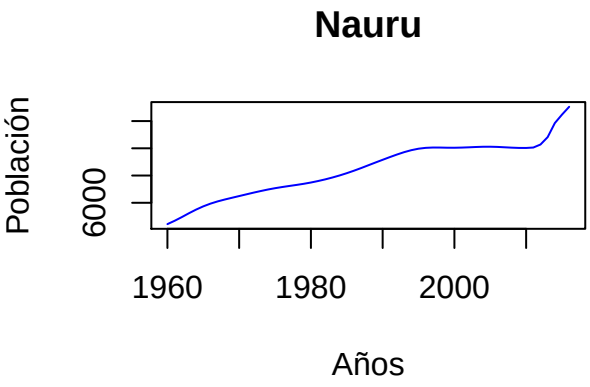

Montenegro
Población
480000
1960
1980
2000
Años

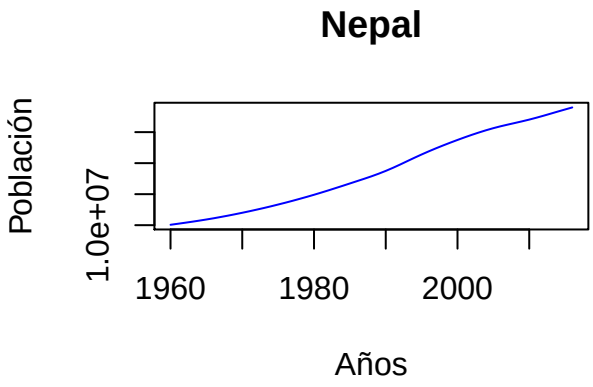

Morocco
Población
1.5e+07
1960
1980
2000
Años

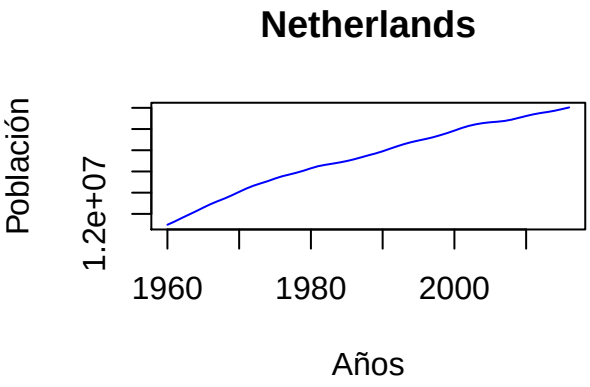

Mozambique
Población
1.0e+07
1960
1980
2000
Años

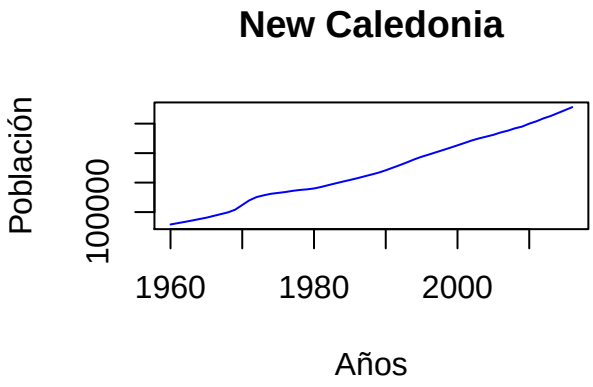

Myanmar
Población
2.0e+07
1960
1980
2000
Años

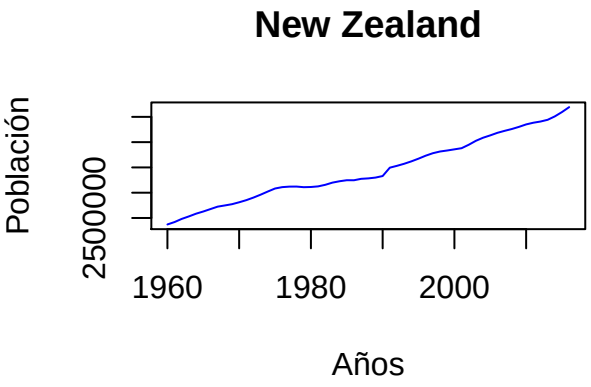

Namibia
Población
1000000
1960
1980
2000
Años

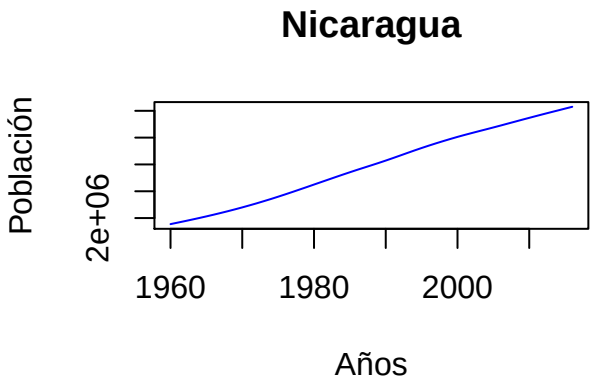

Nauru
Población
6000
1960
1980
2000
Años

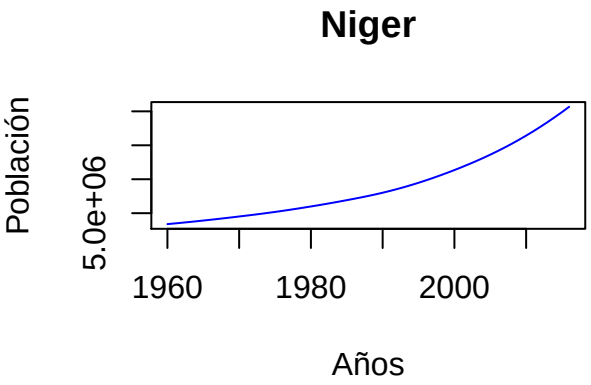

Nepal
Población
1.0e+07
1960
1980
2000
Años

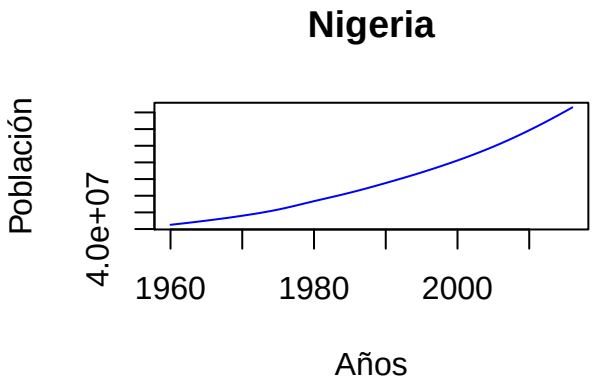

Netherlands
Población
1.2e+07
1960
1980
2000
Años

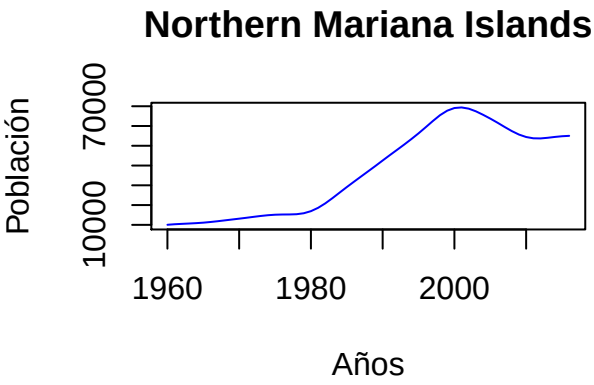

New Caledonia
Población
100000
1960
1980
2000
Años

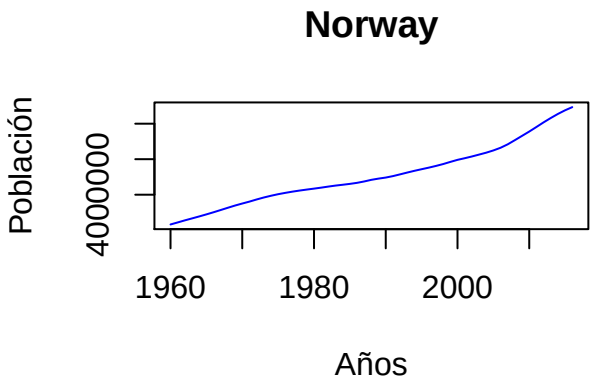

New Zealand
Población
2500000
1960
1980
2000
Años

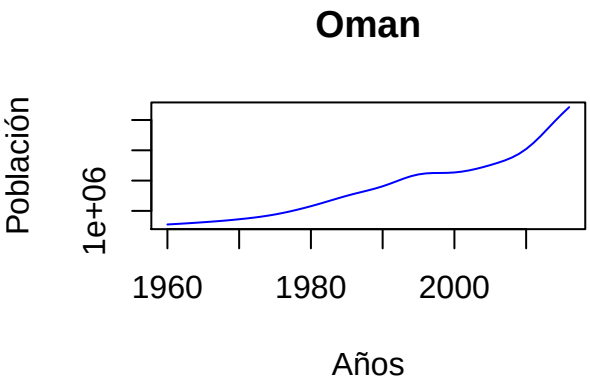

Nicaragua
Población
2e+06
1960
1980
2000
Años

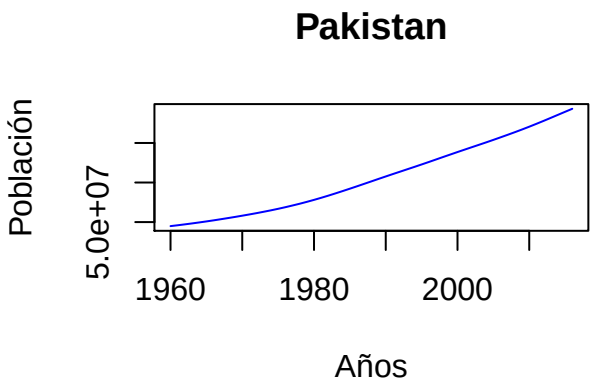

Niger
Población
5.0e+06
1960
1980
2000
Años

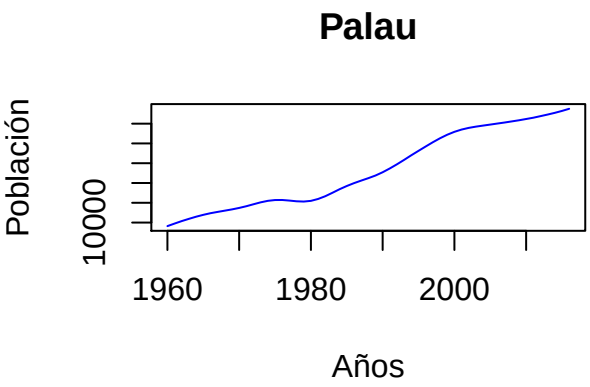

Nigeria
Población
4.0e+07
1960
1980
2000
Años

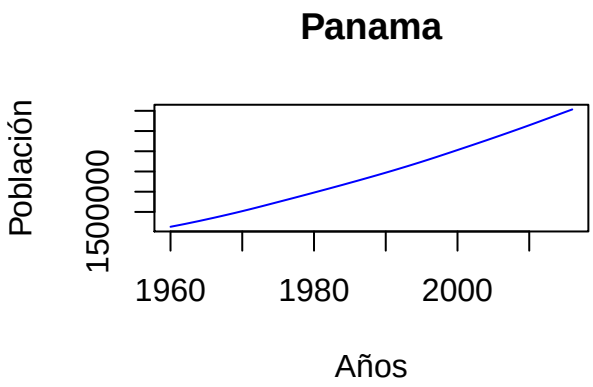

Northern Mariana Islands
Población
10000
70000
1960
1980
2000
Años

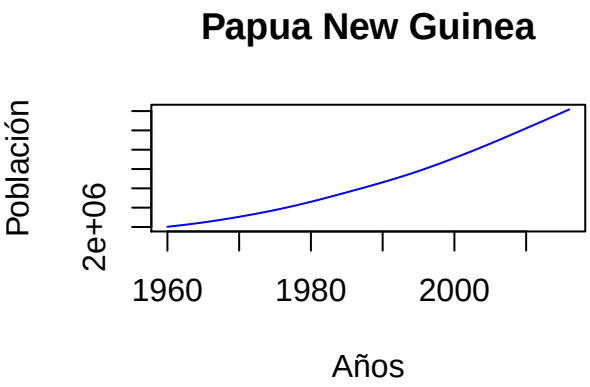

Norway
Población
4000000
1960
1980
2000
Años

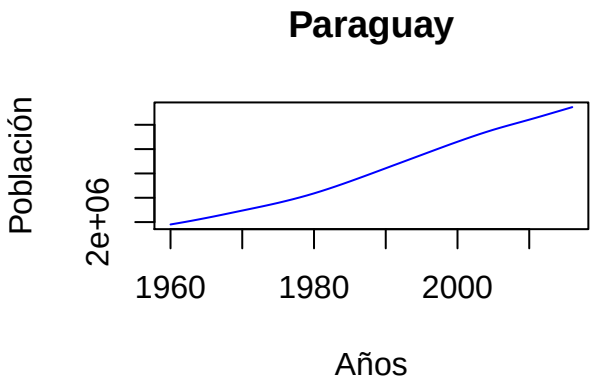

Oman
Población
1e+06
1960
1980
2000
Años

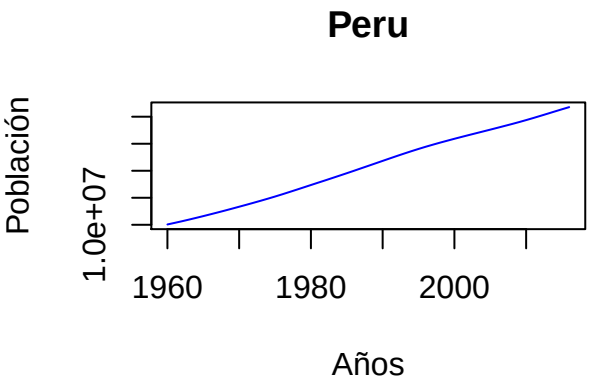

Pakistan
Población
5.0e+07
1960
1980
2000
Años

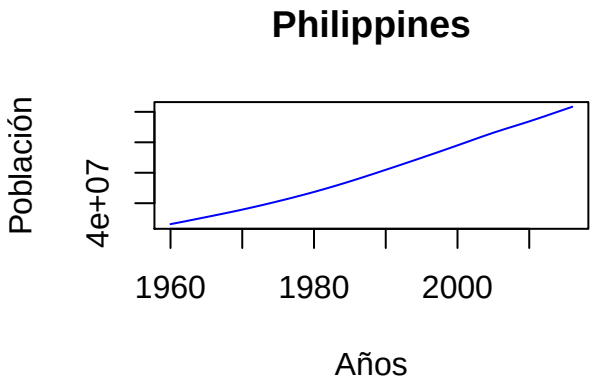

Palau
Población
10000
1960
1980
2000
Años

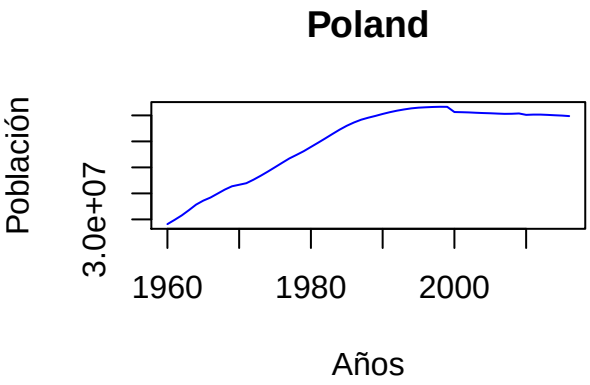

Panama
Población
1500000
1960
1980
2000
Años

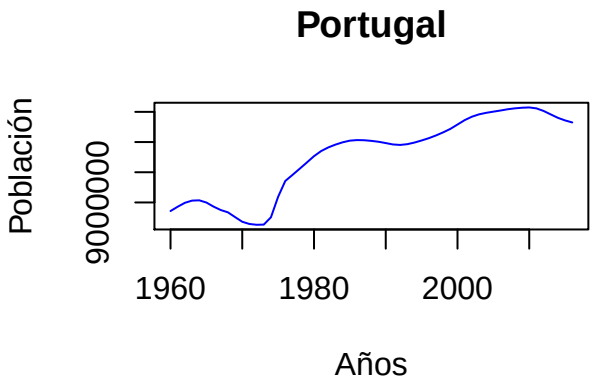

Papua New Guinea
Población
2e+06
1960
1980
2000
Años

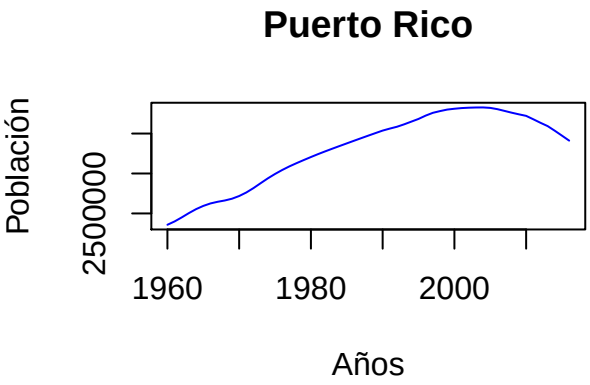

Paraguay
Población
2e+06
1960
1980
2000
Años

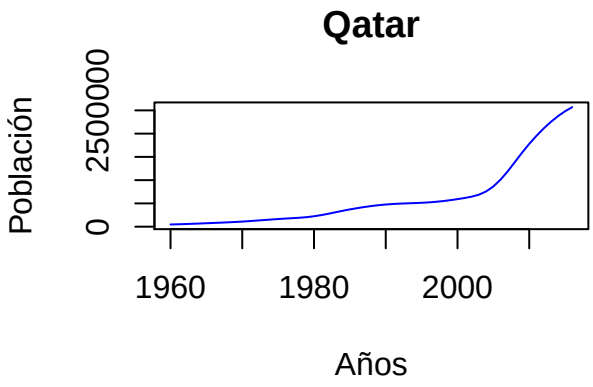

Peru
Población
1.0e+07
1960
1980
2000
Años

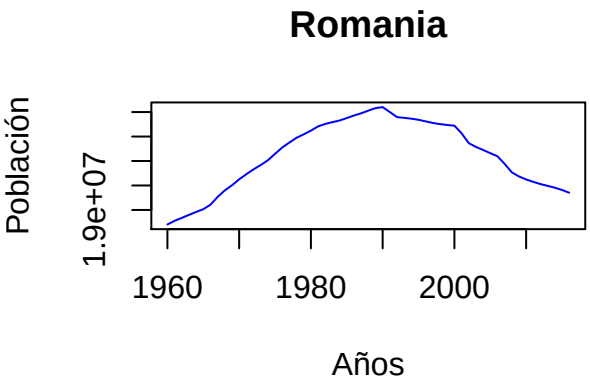

Philippines
Población
4e+07
1960
1980
2000
Años

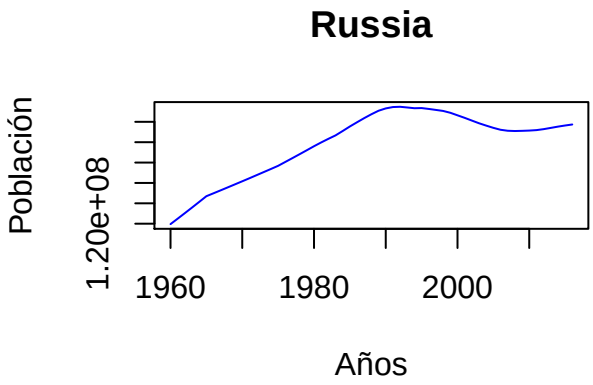

Poland
Población
3.0e+07
1960
1980
2000
Años

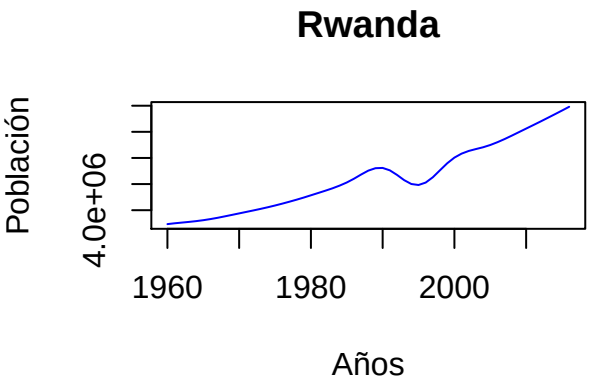

Portugal
Población
9000000
1960
1980
2000
Años

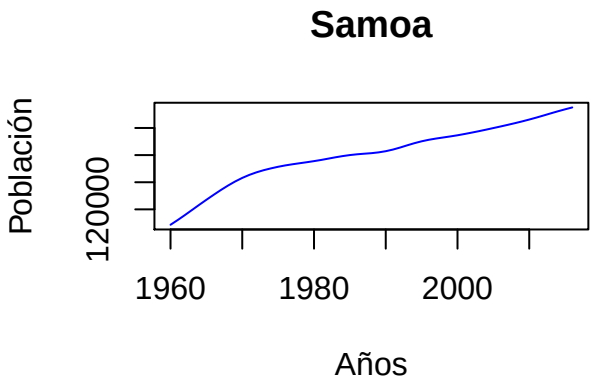

Puerto Rico
Población
2500000
1960
1980
2000
Años

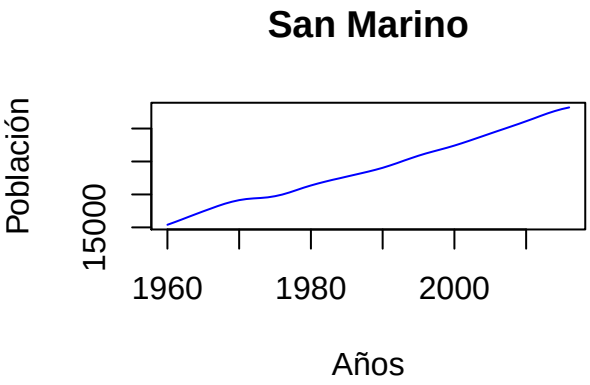

Qatar
Población
0
2500000
1960
1980
2000
Años

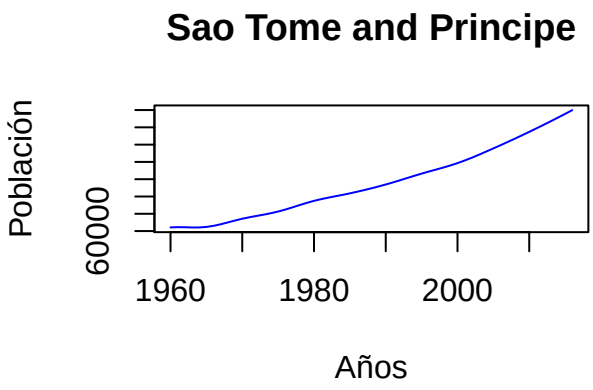

Romania
Población
1.9e+07
1960
1980
2000
Años

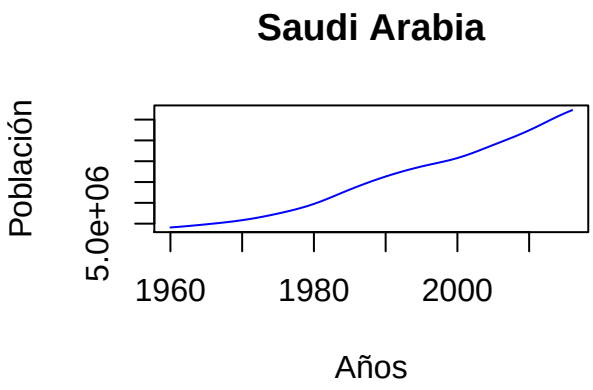

Russia
Población
1.20e+08
1960
1980
2000
Años

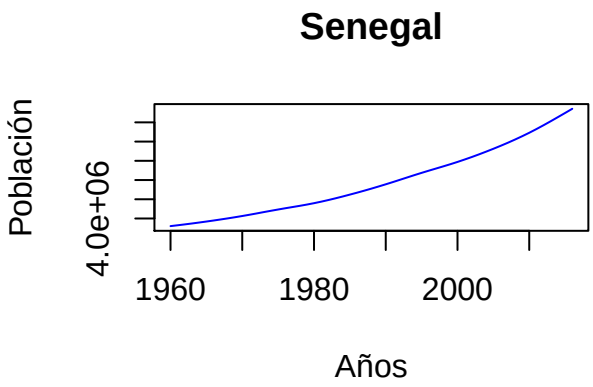

Rwanda
Población
4.0e+06
1960
1980
2000
Años

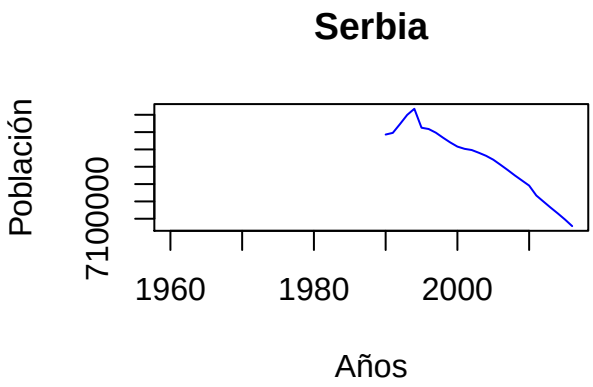

Samoa
Población
120000
1960
1980
2000
Años

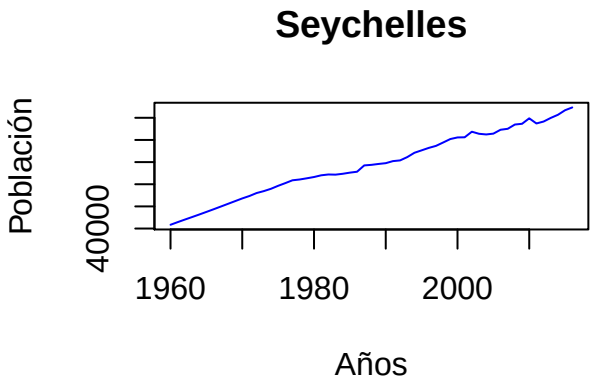

San Marino
Población
15000
1960
1980
2000
Años

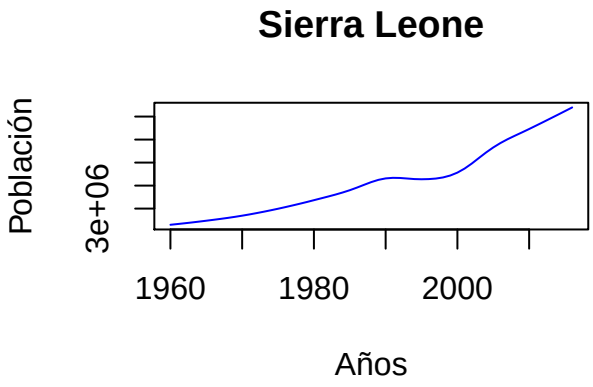

Sao Tome and Principe
Población
60000
1960
1980
2000
Años

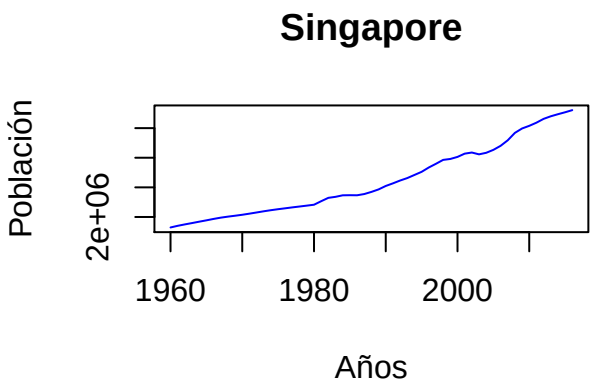

Saudi Arabia
Población
5.0e+06
1960
1980
2000
Años

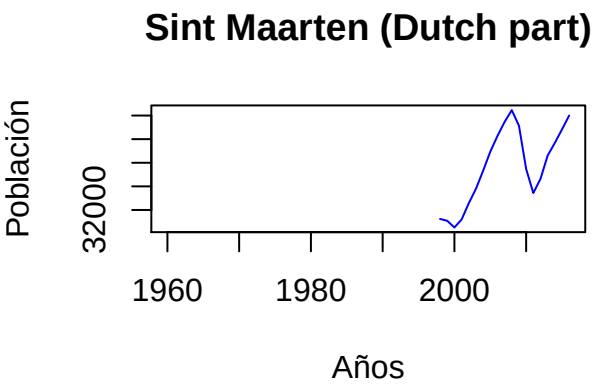

Senegal
Población
4.0e+06
1960
1980
2000
Años

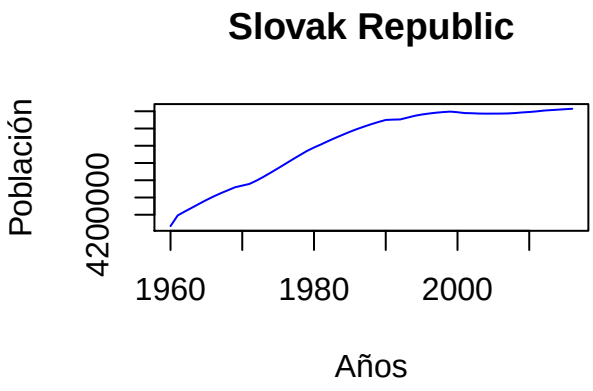

Serbia
Población
7100000
1960
1980
2000
Años

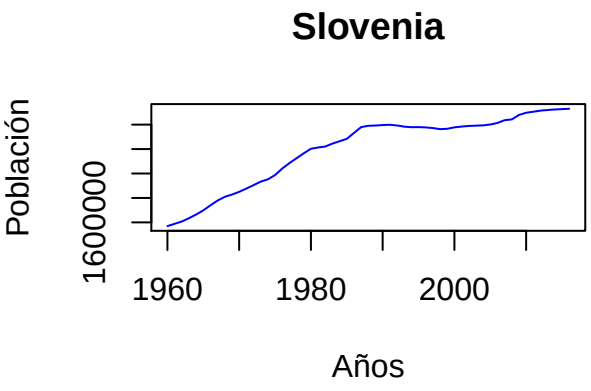

Seychelles
Población
40000
1960
1980
2000
Años

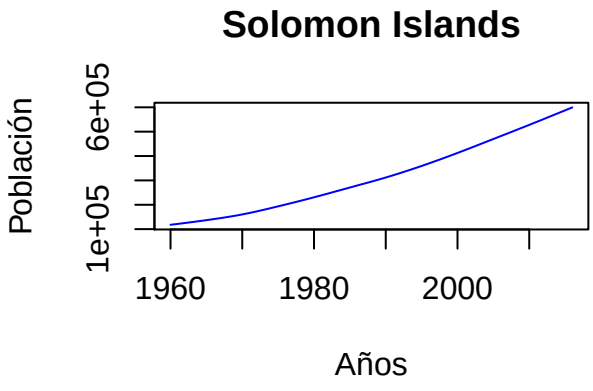

Sierra Leone
Población
3e+06
1960
1980
2000
Años

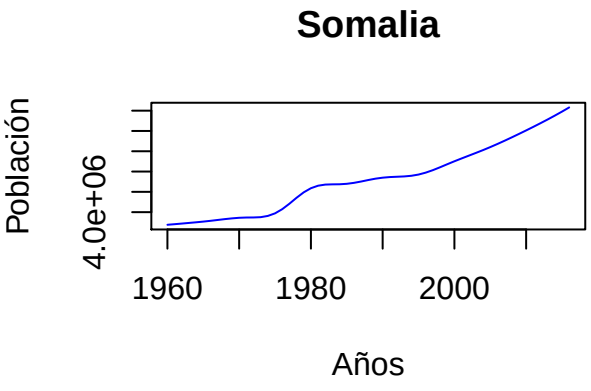

Singapore
Población
2e+06
1960
1980
2000
Años

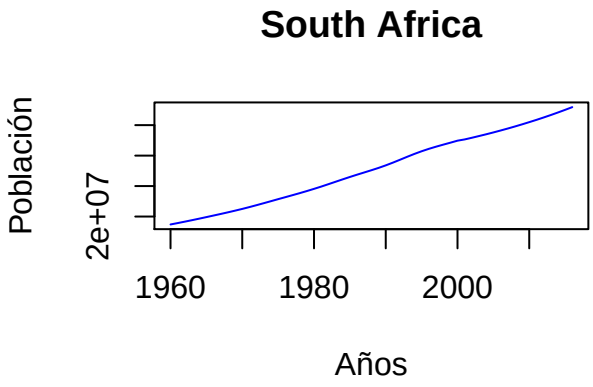

Sint Maarten (Dutch part)
Población
32000
1960
1980
2000
Años

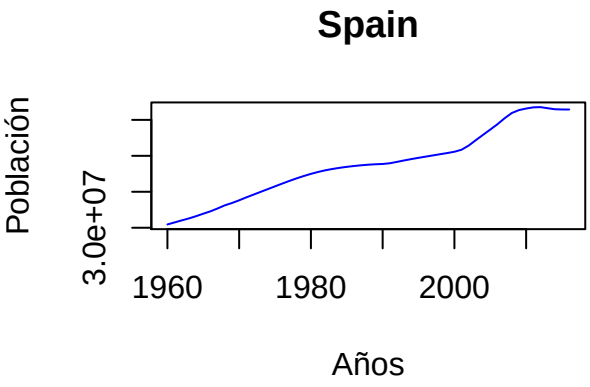

Slovak Republic
Población
4200000
1960
1980
2000
Años

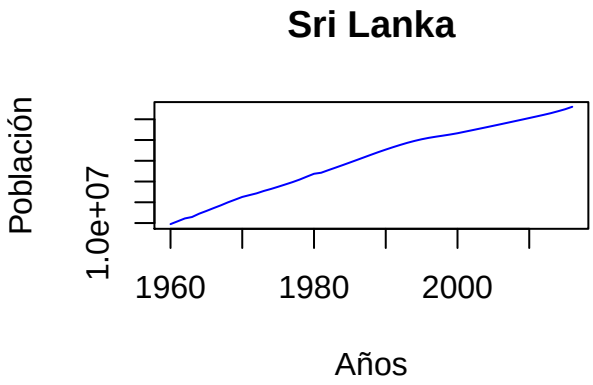

Slovenia
Población
1600000
1960
1980
2000
Años

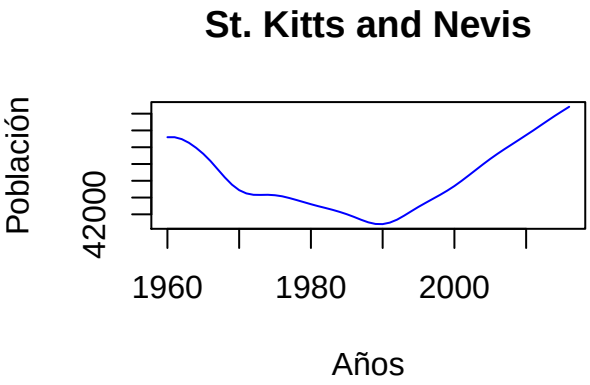

Solomon Islands
Población
1e+05
6e+05
1960
1980
2000
Años

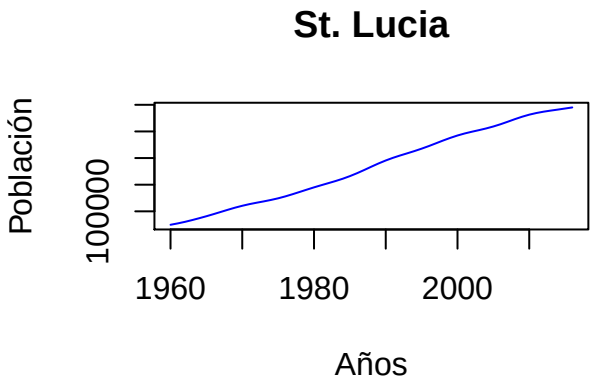

Somalia
Población
4.0e+06
1960
1980
2000
Años

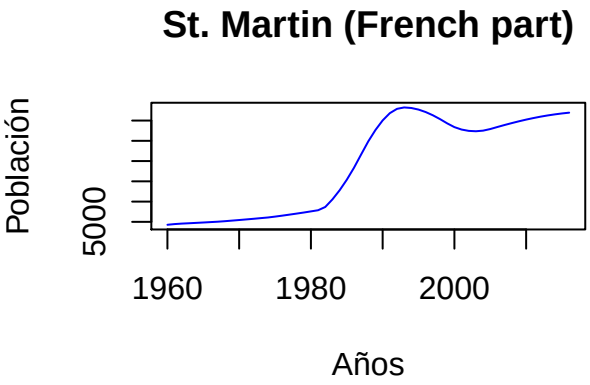

South Africa
Población
2e+07
1960
1980
2000
Años

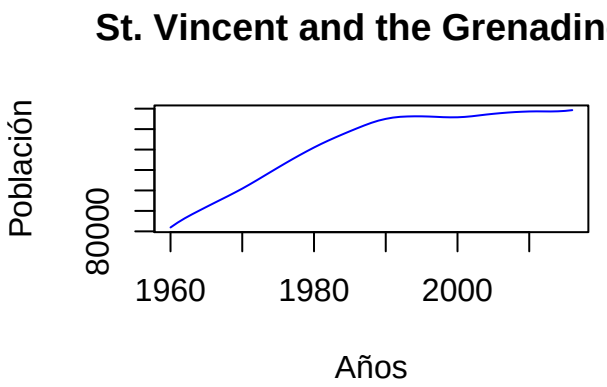

Spain
Población
3.0e+07
1960
1980
2000
Años

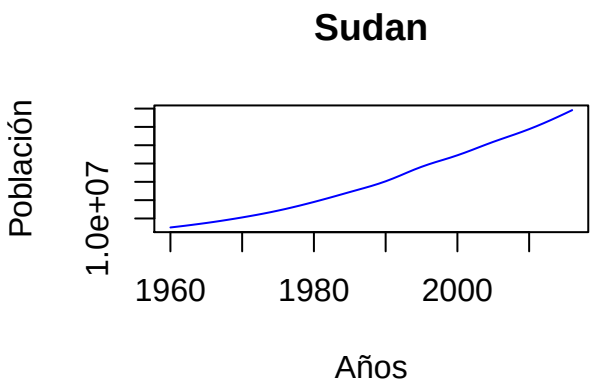

Sri Lanka
Población
1.0e+07
1960
1980
2000
Años

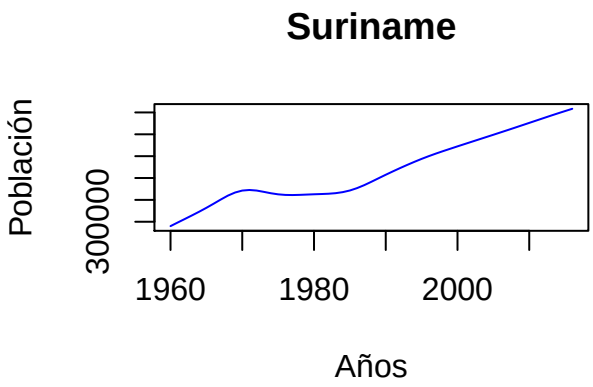

St. Kitts and Nevis
Población
42000
1960
1980
2000
Años

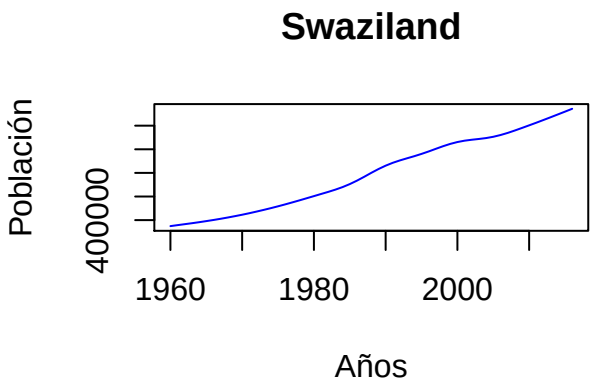

St. Lucia
Población
100000
1960
1980
2000
Años

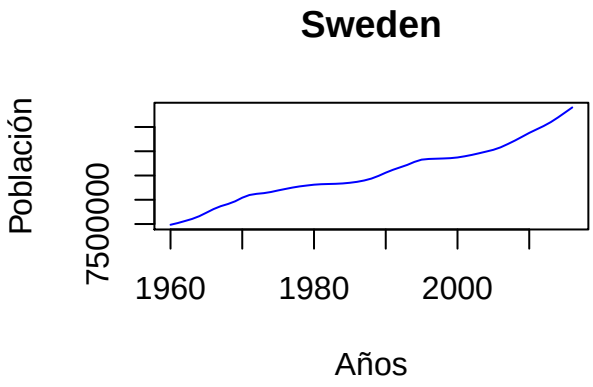

St. Martin (French part)
Población
5000
1960
1980
2000
Años

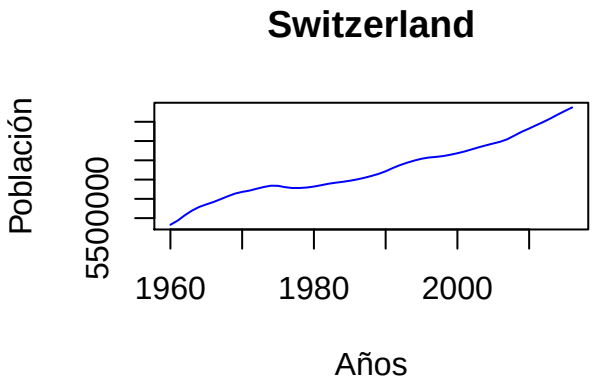

St. Vincent and the Grenadines
Población
80000
1960
1980
2000
Años

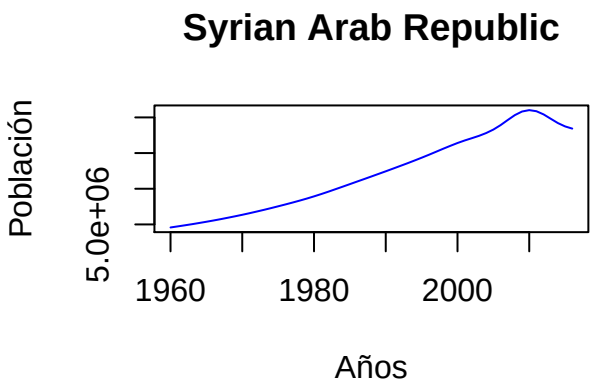

Sudan
Población
1.0e+07
1960
1980
2000
Años

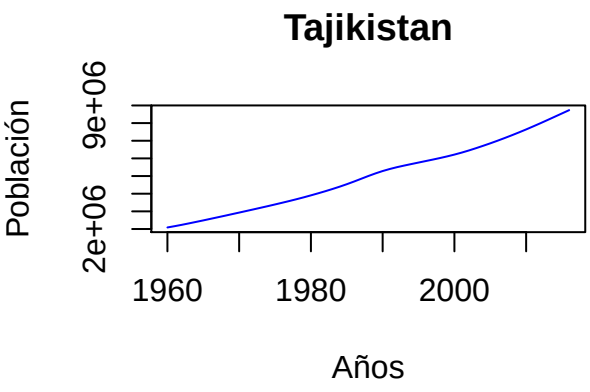

Suriname
Población
300000
1960
1980
2000
Años

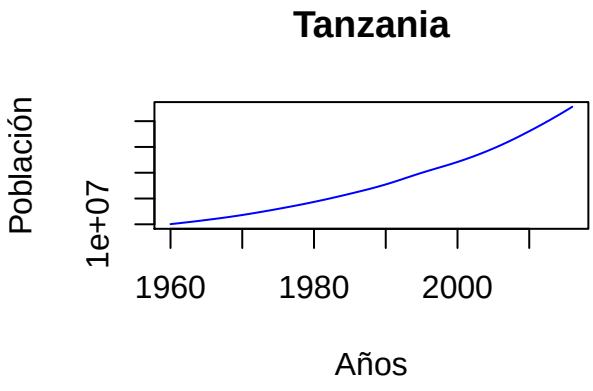

Swaziland
Población
400000
1960
1980
2000
Años

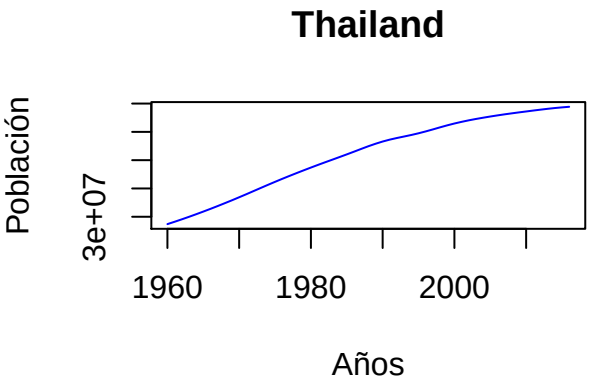

Sweden
Población
7500000
1960
1980
2000
Años

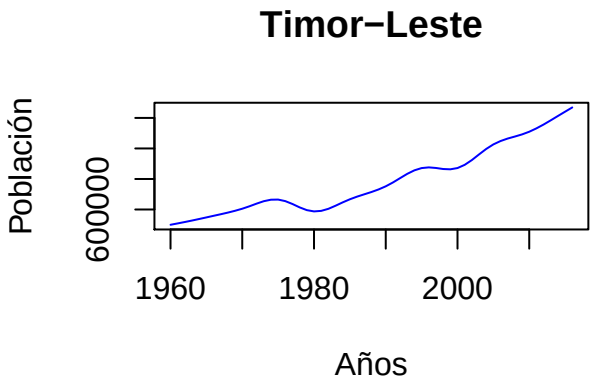

Switzerland
Población
5500000
1960
1980
2000
Años

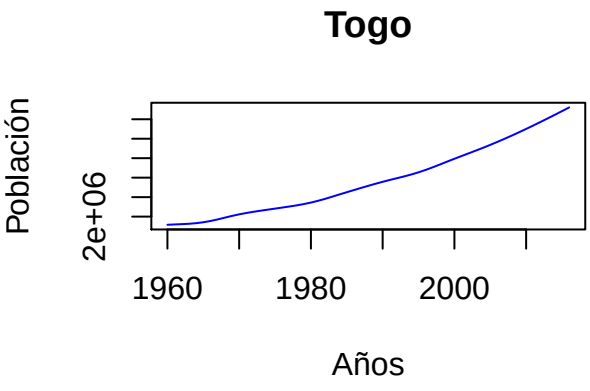

Syrian Arab Republic
Población
5.0e+06
1960
1980
2000
Años

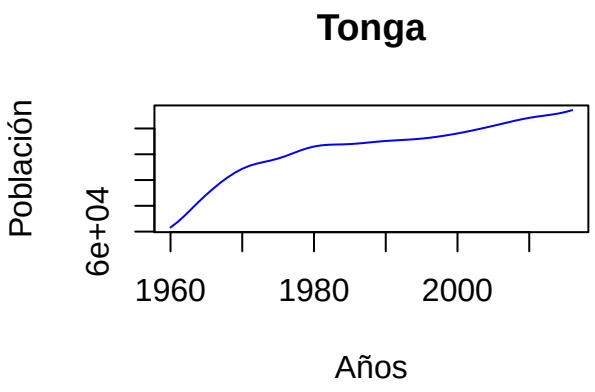

Tajikistan
Población
2e+06
9e+06
1960
1980
2000
Años

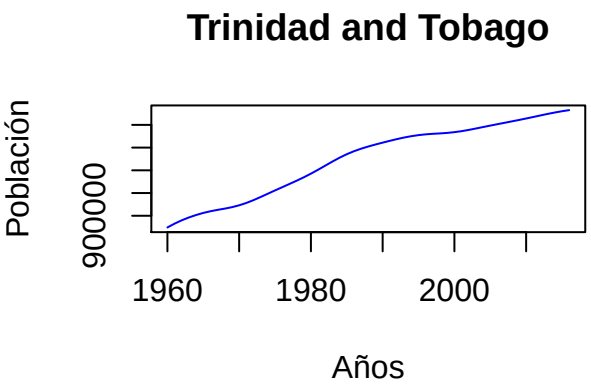

Tanzania
Población
1e+07
1960
1980
2000
Años

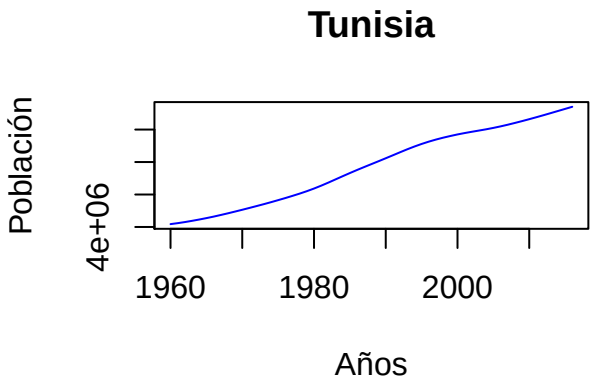

Thailand
Población
3e+07
1960
1980
2000
Años

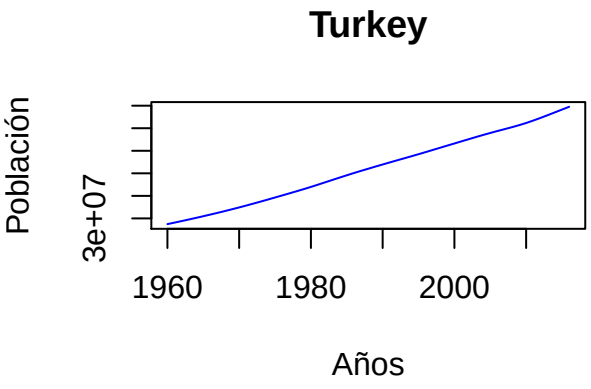

Timor−Leste
Población
600000
1960
1980
2000
Años

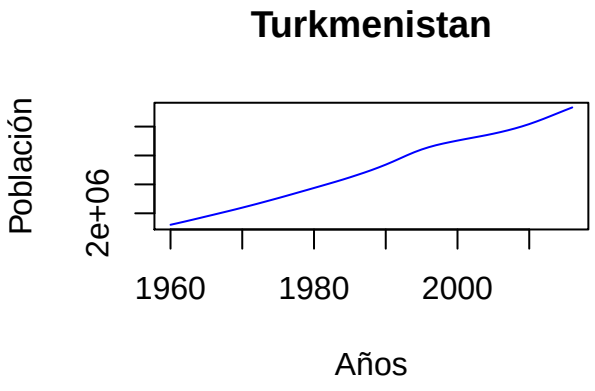

Togo
Población
2e+06
1960
1980
2000
Años

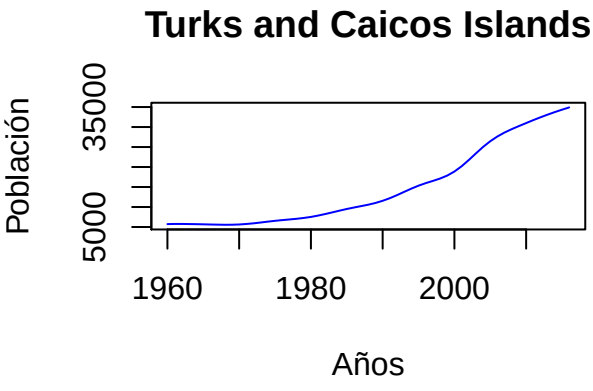

Tonga
Población
6e+04
1960
1980
2000
Años

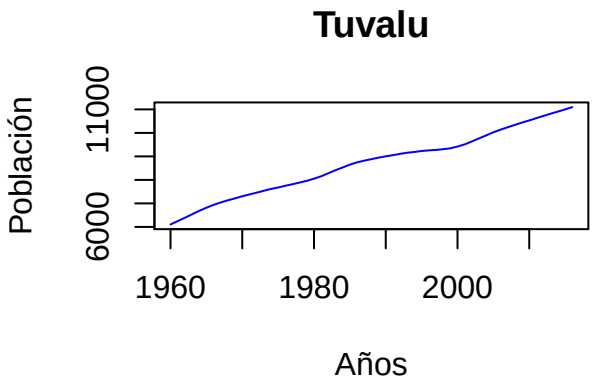

Trinidad and Tobago
Población
900000
1960
1980
2000
Años

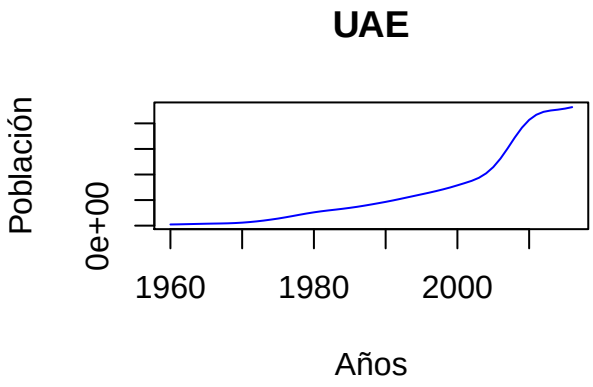

Tunisia
Población
4e+06
1960
1980
2000
Años

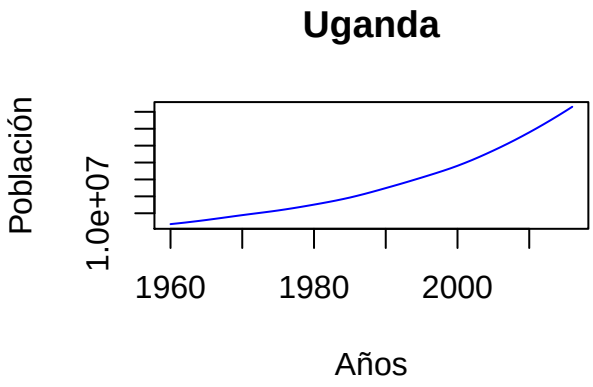

Turkey
Población
3e+07
1960
1980
2000
Años

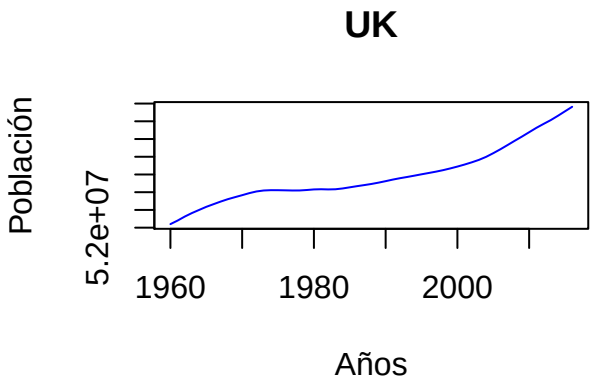

Turkmenistan
Población
2e+06
1960
1980
2000
Años

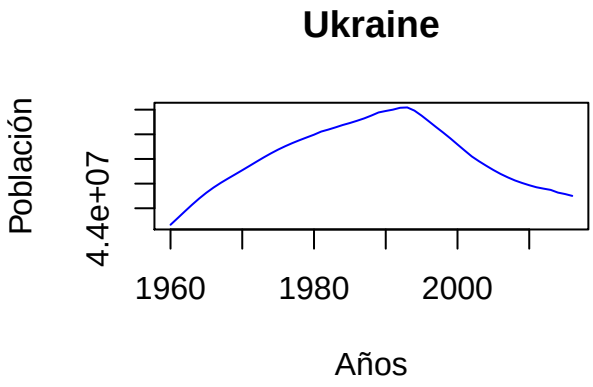

Turks and Caicos Islands
Población
5000
35000
1960
1980
2000
Años

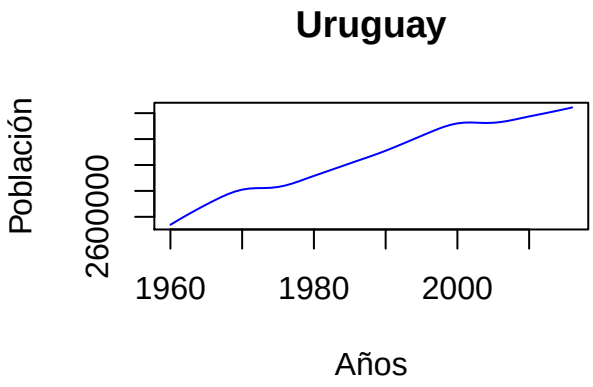

Tuvalu
Población
6000
11000
1960
1980
2000
Años

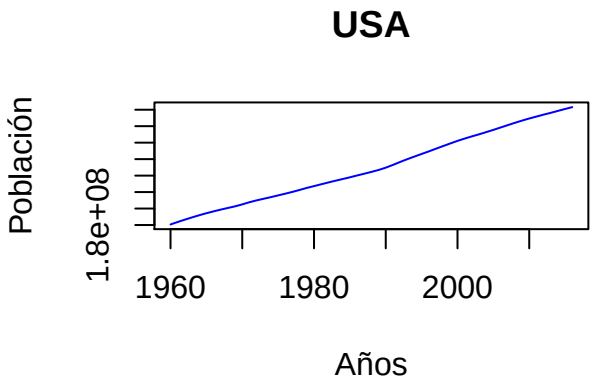

UAE
Población
0e+00
1960
1980
2000
Años

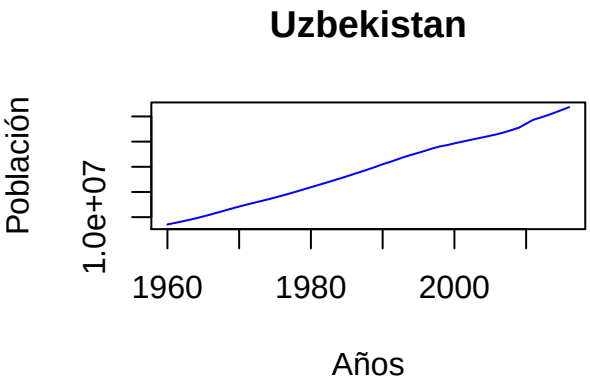

Uganda
Población
1.0e+07
1960
1980
2000
Años

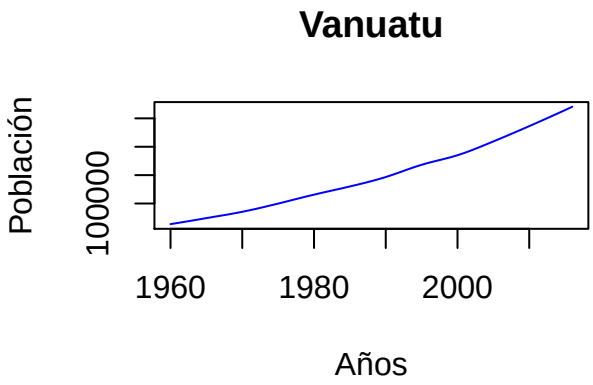

UK
Población
5.2e+07
1960
1980
2000
Años

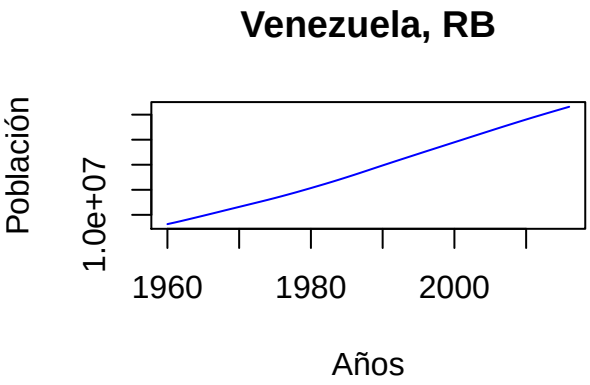

Ukraine
Población
4.4e+07
1960
1980
2000
Años

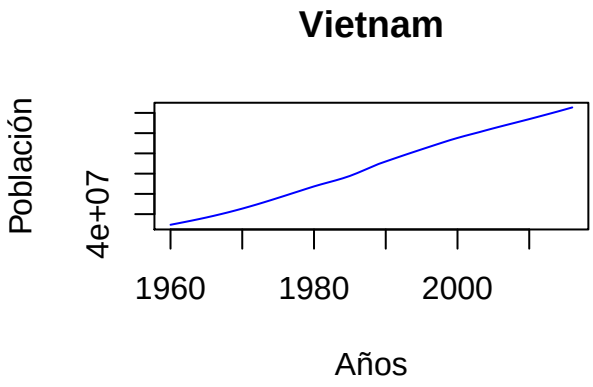

Uruguay
Población
2600000
1960
1980
2000
Años

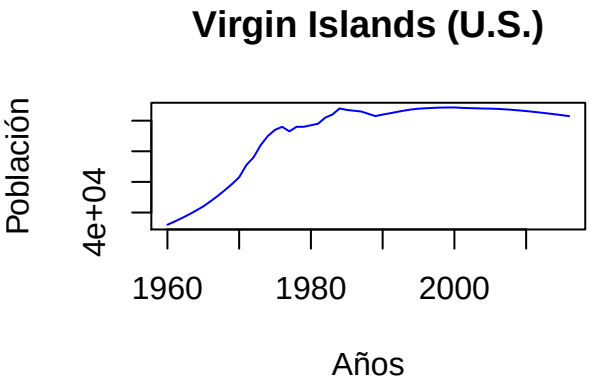

USA
Población
1.8e+08
1960
1980
2000
Años

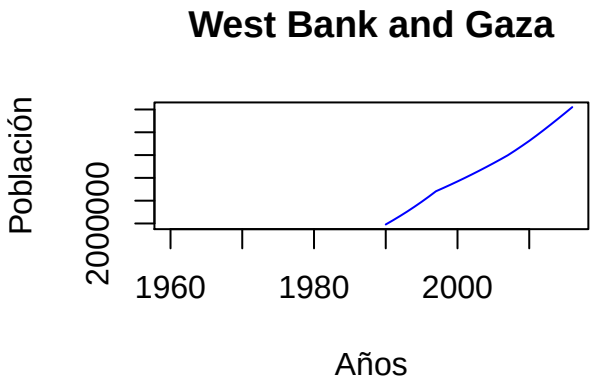

Uzbekistan
Población
1.0e+07
1960
1980
2000
Años

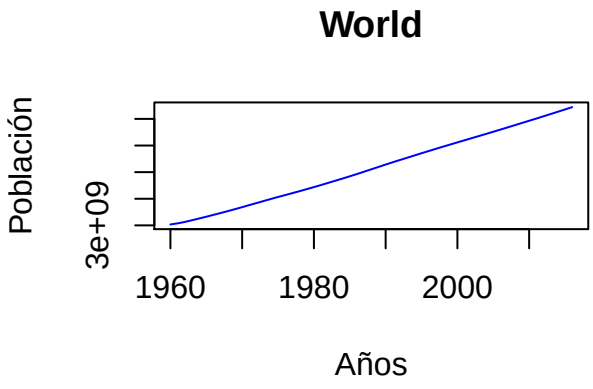

Vanuatu
Población
100000
1960
1980
2000
Años

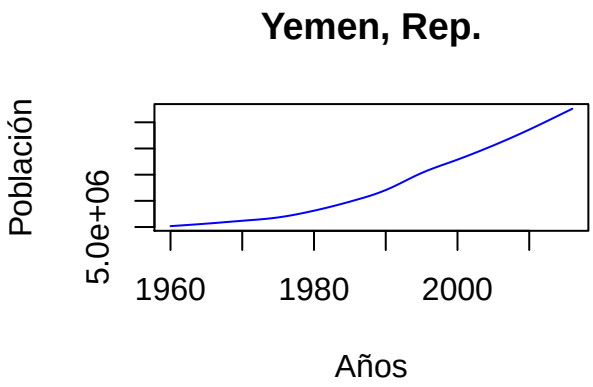

Venezuela, RB
Población
1.0e+07
1960
1980
2000
Años

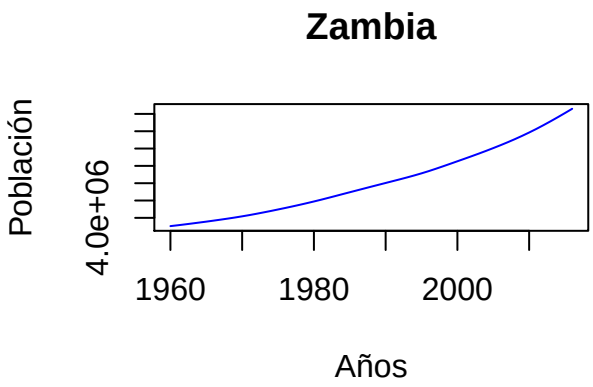

Vietnam
Población
4e+07
1960
1980
2000
Años

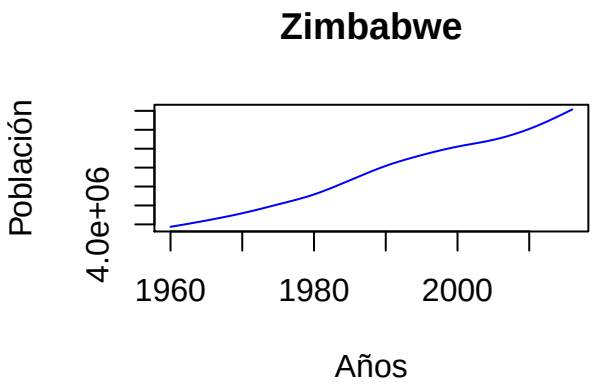

Virgin Islands (U.S.)
Población
4e+04
1960
1980
2000
Años


West Bank and Gaza
Población
2000000
1960
1980
2000
Años


World
Población
3e+09
1960
1980
2000
Años


Yemen, Rep.
Población
5.0e+06
1960
1980
2000
Años


Zambia
Población
4.0e+06
1960
1980
2000
Años


Zimbabwe
Población
4.0e+06
1960
1980
2000
Años

Frecuencias entre las Variables

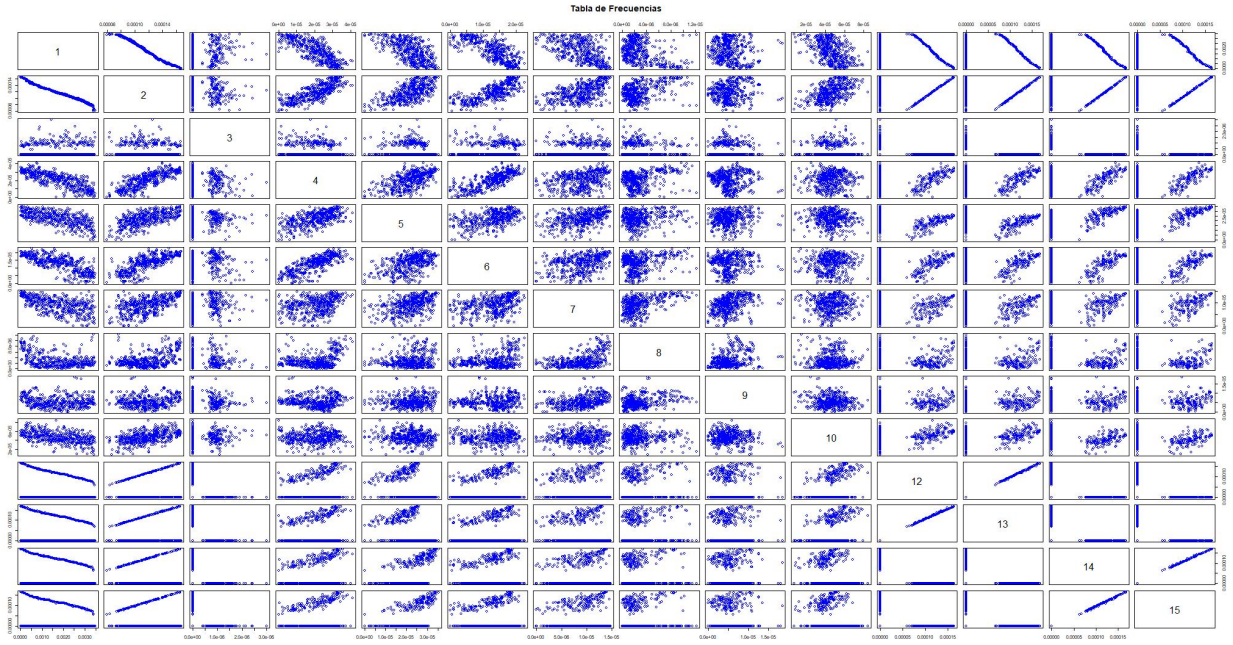


Tabla de Frecuencias

Se puede observar que en la tabla de frecuencias las variables que son correlacionadas y las que son inversamente proporcional. Por ejemplo:

- La variable 8 y 7 tienen una correlación de 0.4904 las cuales serían "libertad" "Confianza de un gobierno corrupto"
- Las otras variables, 4 y 5 tienen una correlación de 0.585449 las cuales serían "Producto interno Bruto" "Familia"
- Las otras variables, 7 y 2 tienen una correlación de 0.585449 las cuales serían "Libertad" "Factor de felicidad"
- Las otras variables, 5 y 6 tienen una correlación de 0.4863827 las cuales serían "Familia" "Esperanza de vida"
- Las otras variables, 4 y 6 tienen una correlación de 0.7854496 las cuales serían "Producto Interno Bruto" "Esperanza de vida"
- Las otras variables, 5 y 7 tienen una correlación de 0.42537669 las cuales serían "Familia" "Libertad"
- Las otras variables, 8 y 4 tienen una correlación de 0.42537669 las cuales serían "Confianza del gobierno corrupto" "Producto Interno Bruto"

Countries of the World

Como se mencionó anteriormente este dataset contiene distintos datos de los paises del mundo. A continuación se muestra un resumen de cada variable de este dataset:

Country	Region	Population	Area..sq..mi..
Afghanistan : 1	SUB-SAHARAN AFRICA :51	Min. :7.026e+03	Min. : 2
Albania : 1	LATIN AMER. & CARIB :45	1st Qu.:4.376e+05	1st Qu.: 4648
Algeria : 1	ASIA (EX. NEAR EAST) :28	Median :4.787e+06	Median : 86600
American Samoa : 1	WESTERN EUROPE :28	Mean :2.874e+07	Mean : 598227
Andorra : 1	OCEANIA :21	3rd Qu.:1.750e+07	3rd Qu.: 441811
Angola : 1	NEAR EAST :16	Max. :1.314e+09	Max. :17075200
(Other) :221	(Other) :38	NA	NA

Pop..Density..per.sq..mi..	Coastline..coast.area.ratio.	Net.migration	Infant.mortality..per.1000.births.
13,8 : 2	0,00 : 44	0 : 62	: 3
2,7 : 2	0,09 : 4	: 3	9,95 : 3
372,5 : 2	0,13 : 4	-0,07 : 2	12,62 : 2
49,6 : 2	0,10 : 3	-0,31 : 2	4,39 : 2
66,6 : 2	0,15 : 3	-0,39 : 2	10,03 : 1
69,8 : 2	0,21 : 3	-0,71 : 2	10,09 : 1
(Other):215	(Other):166	(Other):154	(Other):215

GDP....per.capita.	Literacy....	Phones..per.1000.	Arable....
Min. : 500	: 18	: 4	0 : 9
1st Qu.: 1900	99,0 : 13	2,3 : 2	20 : 3
Median : 5550	97,0 : 11	2,7 : 2	: 2
Mean : 9690	98,0 : 10	255,6 : 2	1,64 : 2
3rd Qu.:15700	100,0 : 7	26,8 : 2	1,67 : 2
Max. :55100	98,6 : 4	269,5 : 2	10 : 2
NA's :1	(Other):164	(Other):213	(Other):207

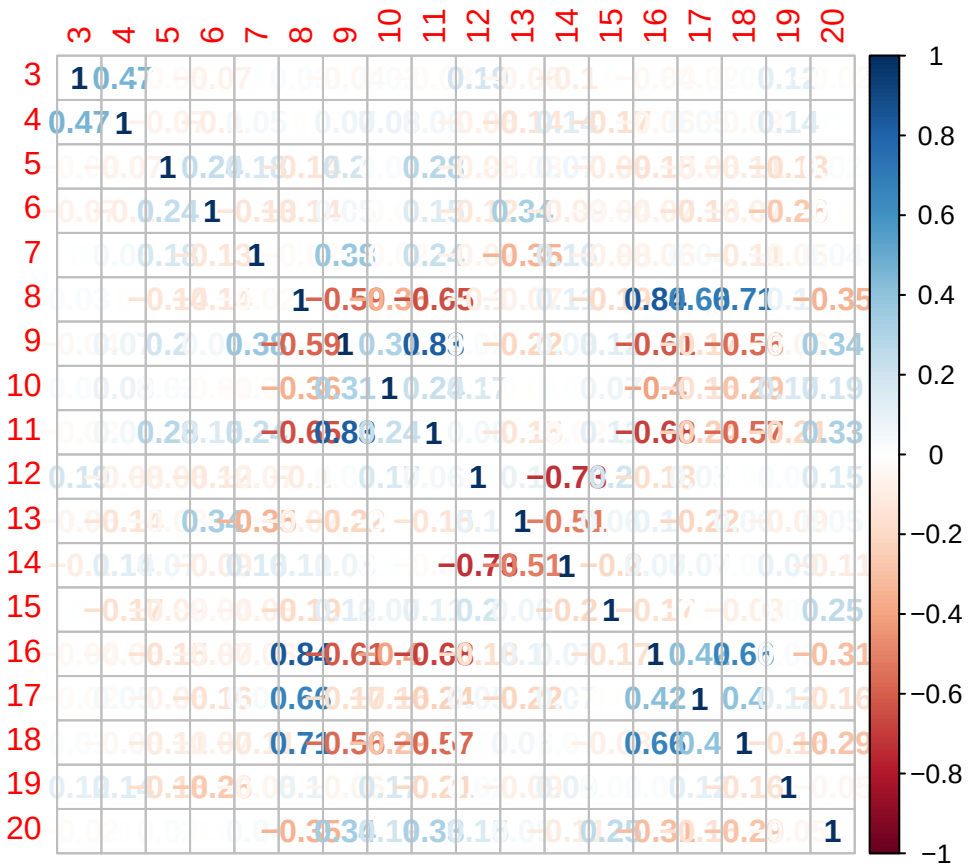
Crops....	Other....	Climate	Birthrate
0 : 28	100 : 8	: 22	: 3
0,03 : 5	75 : 3	1 : 29	12,56 : 2
0,01 : 4	: 2	1,5: 8	18,02 : 2
0,19 : 4	70 : 2	2 :111	18,79 : 2
0,4 : 4	70,44 : 2	2,5: 3	20,48 : 2
0,14 : 3	73,33 : 2	3 : 48	10 : 1
(Other):179	(Other):208	4 : 6	(Other):215

Deathrate	Agriculture	Industry	Service
: 4	: 15	: 16	: 15
10,31 : 2	0,01 : 8	0,11 : 4	0,684 : 5
12,25 : 2	0,04 : 6	0,17 : 4	0,55 : 4
14,02 : 2	0,03 : 5	0,1 : 3	0,62 : 4
3,92 : 2	0,018 : 4	0,12 : 3	0,46 : 3
5,28 : 2	0,06 : 4	0,18 : 3	0,549 : 3
(Other):213	(Other):185	(Other):194	(Other):193

Correlación entre variables

Para poder observar si existe algún tipo de relación entre las variables numéricas del dataset , se realizó un gráfico de correlación en donde cada variable se puso contra cada variable del dataset. En el resultado

podemos observar la correlación existente entre las variables:



En el diagrama de correlación se observó que existen coeficientes de correlaciones de varios valores, sin embargo, para este experimento los valores arriba de un 0.4 fueron considerados como significativos. Por lo que las relaciones que más se destacan son las siguientes:

- Population - Area..sq..mi..: 0.47
- Infant.mortality..per.1000.births. - GDP....per.capita.: 0.59
- Infant.mortality..per.1000.births. - Phones..per.1000.: 0.65
- Infant.mortality..per.1000.births. - Birthrate: 0.84
- Infant.mortality..per.1000.births. - Deathrate: 0.66
- Infant.mortality..per.1000.births. - Agriculture: 0.71
- GDP....per.capita. - Phones..per.1000.: 0.83
- GDP....per.capita. - Birthrate: 0.61
- GDP....per.capita. - Agriculture: 0.56
- Literacy.... - Birthrate: 0.4
- Phones..per.1000. - Birthrate: 0.68
- Phones..per.1000. - Agriculture: 0.57
- Arable.... - Other....: 0.73

- Crops.... - Other....: 0.51
- Birthrate - Deathrate: 0.42
- Birthrate - Agriculture: 0.66
- Deathrate - Agriculture: 0.4

En los resultados obtenidos en el diagrama de correlación, se logró observar que la variable *Infant.mortality..per.1000.births* es la variable que tiene mayor relación con otras variables dentro del dataset, seguida de *GDP. . . .per.capita.*. También, la correlación de mayor magnitud es la existente entre las variables *Infant.mortality..per.1000.births.* y *Birthrate.*

Análisis de PCA

Dados los datasets que se tienen *World Happiness Report*, *World Population* y *Countries of the World*; el único dataset en el que valdría la pena analizar sus componentes principales es el de *World Happiness Report*; eso es porque es el único dataset en el que tiene sentido una reducción de dimensionalidad ya que tiene bastantes variables que podrían estar relacionadas con la felicidad de la gente. Por otro lado, los otros 2 Datasets no tienen una dimensión muy alta (en el caso de *World Population*) o no se están relacionando las variables (en el caso de *Countries of the world*).

Previo a realizar el análisis de PCA se midió la adecuación muestral utilizando los coeficientes de Kaiser-Meyer-Olkin, la prueba de esfericidad de Bartlett y el estadístico de Barlett. Los resultados fueron los siguientes:

KMO	Barlett	RMS
0.329113	21154	0.045415

Debido a que la adecuación muestral es inaceptable no se llevó a cabo el analisis de PCA ni un análisis factorial.

Conclusiones y Hallazgos

Para el dataset World Happiness Report, se puede comenzar mencionando que las variables *Area..sq..mi..*, *Pop..Density..per.sq..mi..* y *Crops. . . .* presentan tener multicolinealidad. Por lo que se recomiendo que sean removidas en el caso que se planee calcular un modelo lineal múltiple o que esta característica sea tomada en cuenta al realizar algun analisis con el dataset.

Continuamos mencionando la relación existente entre entre variables como *Happiness.Rank*, la cual describe la posición en la que se encuentra cada pais segun su felicidad, en donde el tema familiar, la expectancia de vida y la economía tiene impacto directo. Así com que el grado que impacto el producto interno bruto en la felicidad es bastante parecido al grado de impacto de la expectancia de vida.

Los hallazgos encontrados en el dataset de *Countries Population* son simples, debido a que se encontró que la variable *population*, que describe la población de cada país, depende únicamente de la variable *year*. Ya que se encontró en el diagrama de correlación que cada año presenta tener una relación estrecha con los demás años.

Por ende, crear un modelo que describa o pronostique la población de cada país consiste en realizar una Regresión Lineal Simple entre la variable *population* y la variable *year*. Debido a que solo se necesita de la variable year, tomada en diferentes unidades de tiempo, para describir cómo se comporta la población en cada país.