

COVID-19_SincronizadoWeb

March 20, 2020

Coronavirus COVID-19 .

(Año 2020)

Jesús Octavio García Flores

(j.octavio.ibs@gmail.com) En este pequeño ejercicio de análisis se plantean las siguientes preguntas:

```
# ¿Cuántos casos confirmados de personas infectadas hay en el mundo?  
# ¿Qué país tiene el mayor número de casos?  
# ¿Cuántas personas han fallecido por causa del coronavirus?  
# ¿Qué porcentaje de la población infectada ha muerto?  
# ¿Cómo ha sido el crecimiento de casos en un país determinado?
```

```
[2]: import pandas as pd  
import matplotlib.pyplot as plt  
%matplotlib inline
```

Full data set

```
[3]: source_full_data = "https://covid.ourworldindata.org/data/ecdc/full_data.csv"  
data1 = pd.read_csv(source_full_data)  
full_data = pd.DataFrame(data1)  
full_data.head()
```

```
[3]:
```

	date	location	new_cases	new_deaths	total_cases	total_deaths
0	2019-12-31	Afghanistan	0	0	0	0
1	2020-01-01	Afghanistan	0	0	0	0
2	2020-01-02	Afghanistan	0	0	0	0
3	2020-01-03	Afghanistan	0	0	0	0
4	2020-01-04	Afghanistan	0	0	0	0

```
[4]: full_data[['new_cases', 'new_deaths', 'total_cases', 'total_deaths']].max()
```

```
[4]: new_cases      18345
      new_deaths     970
      total_cases    213254
      total_deaths    8843
      dtype: int64
```

```
[84]: deaths = full_data['total_deaths'].max()
      print("\033[1m \n EL TOTAL DE MUERTES HASTA EL MOMENTO ES DE " + str(deaths) + "\n
      ↪" PERSONAS \033[0m \n")
```

EL TOTAL DE MUERTES HASTA EL MOMENTO ES DE 8843 PERSONAS

Total cases set

```
[7]: source_total_cases = "https://covid.ourworldindata.org/data/ecdc/total_cases.
      ↪csv"
      data2 = pd.read_csv(source_total_cases)
      total_cases = pd.DataFrame(data2)
      total_cases.head()
```

```
[7]:
```

	date	World	Afghanistan	Albania	Algeria	Andorra	\
0	2019-12-31	27	0.0	NaN	0.0	NaN	
1	2020-01-01	27	0.0	NaN	0.0	NaN	
2	2020-01-02	27	0.0	NaN	0.0	NaN	
3	2020-01-03	44	0.0	NaN	0.0	NaN	
4	2020-01-04	44	0.0	NaN	0.0	NaN	

	Antigua and Barbuda	Argentina	Armenia	Australia	...	Ukraine	\
0	NaN	NaN	0.0	0	...	NaN	
1	NaN	NaN	0.0	0	...	NaN	
2	NaN	NaN	0.0	0	...	NaN	
3	NaN	NaN	0.0	0	...	NaN	
4	NaN	NaN	0.0	0	...	NaN	

	United Arab Emirates	United Kingdom	United States	Uruguay	Uzbekistan	\
0	0.0	0	0	NaN	NaN	
1	0.0	0	0	NaN	NaN	
2	0.0	0	0	NaN	NaN	
3	0.0	0	0	NaN	NaN	
4	0.0	0	0	NaN	NaN	

	Vatican	Venezuela	Vietnam	Zambia
0	NaN	NaN	0.0	NaN
1	NaN	NaN	0.0	NaN

2	NaN	NaN	0.0	NaN
3	NaN	NaN	0.0	NaN
4	NaN	NaN	0.0	NaN

[5 rows x 157 columns]

```
[12]: #Número de casos
max_world = total_cases['World'].max()
```

```
[170]: print("\033[1m \n HASTA AHORA, HA HABIDO UN TOTAL DE " + str(max_world) + "\n
↪CASOS EN TODO EL MUNDO \033[0m \n")
print("\033[1m \n (El " + str(round((deaths/max_world)*100, 2)) + "% HA\n
↪FALLECIDO) \033[0m \n")
```

HASTA AHORA, HA HABIDO UN TOTAL DE 213254 CASOS EN TODO EL MUNDO

(El 4.15% HA FALLECIDO)

```
[88]: total_cases.describe()
```

```
[88]:
```

	World	Afghanistan	Albania	Algeria	Andorra	\
count	80.000000	70.000000	11.000000	75.000000	6.000000	
mean	51040.925000	1.557143	30.000000	5.466667	14.833333	
std	54392.980669	4.915623	20.721969	14.925340	19.549084	
min	27.000000	0.000000	2.000000	0.000000	1.000000	
25%	234.000000	0.000000	10.500000	0.000000	2.750000	
50%	36242.500000	0.000000	33.000000	0.000000	9.500000	
75%	83824.500000	0.000000	46.500000	0.000000	14.000000	
max	213254.000000	22.000000	59.000000	73.000000	53.000000	

	Antigua and Barbuda	Argentina	Armenia	Australia	Austria	\
count	1.0	13.000000	71.000000	80.000000	80.000000	
mean	1.0	35.230769	4.507042	44.82500	92.000000	
std	NaN	31.188015	17.765837	100.13368	287.545704	
min	1.0	1.000000	0.000000	0.000000	0.000000	
25%	1.0	9.000000	0.000000	0.000000	0.000000	
50%	1.0	31.000000	0.000000	14.00000	0.000000	
75%	1.0	56.000000	0.000000	23.50000	5.500000	
max	1.0	97.000000	115.000000	565.00000	1646.000000	

	...	Ukraine	United Arab Emirates	United Kingdom	United States	\
count	...	6.000000	74.000000	80.000000	80.000000	
mean	...	7.500000	15.256757	154.025000	461.975000	

std	...	7.259477	27.702461	448.728893	1463.199131
min	...	1.000000	0.000000	0.000000	0.000000
25%	...	3.000000	0.000000	0.000000	0.000000
50%	...	4.000000	5.000000	3.500000	12.000000
75%	...	11.750000	13.000000	16.500000	61.500000
max	...	19.000000	113.000000	2630.000000	9415.000000

	Uruguay	Uzbekistan	Vatican	Venezuela	Vietnam	Zambia
count	5.000000	4.000000	5.0	5.000000	76.000000	1.0
mean	34.400000	13.250000	1.0	24.800000	13.539474	2.0
std	30.680613	7.804913	0.0	11.366618	16.934730	NaN
min	6.000000	6.000000	1.0	10.000000	0.000000	2.0
25%	8.000000	7.500000	1.0	15.000000	0.000000	2.0
50%	29.000000	12.000000	1.0	33.000000	11.000000	2.0
75%	50.000000	17.750000	1.0	33.000000	16.000000	2.0
max	79.000000	23.000000	1.0	33.000000	76.000000	2.0

[8 rows x 156 columns]

Total deaths

```
[123]: source_total_deaths = "https://covid.ourworldindata.org/data/ecdc/total_deaths.
      ↪CSV"
data3 = pd.read_csv(source_total_deaths)
total_deaths = pd.DataFrame(data3)
#total_deaths.head()
```

```
[125]: # Muertos en China a causa del coronavirus.
deaths_china = total_deaths['China'].max()
```

```
[131]: # Buscamos el país con mayor número de casos. Los datos son acumulados por día,
      ↪por eso buscamos el
      # máximo de cada país (porque es el acumulado total).
      # Indexamos con "loc" por el renglón "max" y hacemos la búsqueda a partir de la
      ↪columna "Afghanistan"

      # data_frame.idxmax(axis=1) devuelve el índice del elemento máximo por columna.

number_max_case = total_cases.describe().loc['max', 'Afghanistan:'].max()
name_max_case = total_cases.describe().loc['max', 'Afghanistan:'].idxmax(axis=1)

print("\033[1m \n EL PAÍS CON MAYOR NÚMERO DE CASOS HASTA EL MOMENTO ES ",
      ↪name_max_case, " CON ", number_max_case, " CASOS, \033[0m \n")

# Casos de muerte comparados con su población
porcentaje_deaths = (deaths_china/number_max_case)*100
```

```

print("\033[1m CON UN NÚMERO DE MUERTOS de", deaths_china , "PERSONAS (EL",
↳round(porcentaje_deaths,2), "% DE SU POBLACIÓN DE INFECTADOS). \033[0m \n")

# Casos totales comparados con los de China
porcentaje = (number_max_case / max_world)*100      # Porcentaje de infectados en
↳China
print("\033[1m ABARCANDO EL ", round(porcentaje,2), "% DEL TOTAL MUNDIAL.
↳\033[0m \n")

```

EL PAÍS CON MAYOR NÚMERO DE CASOS HASTA EL MOMENTO ES China CON 81238.0 CASOS,

CON UN NÚMERO DE MUERTOS de 3250 PERSONAS (EL 4.0 % DE SU POBLACIÓN DE INFECTADOS).

ABARCANDO EL 38.09 % DEL TOTAL MUNDIAL.

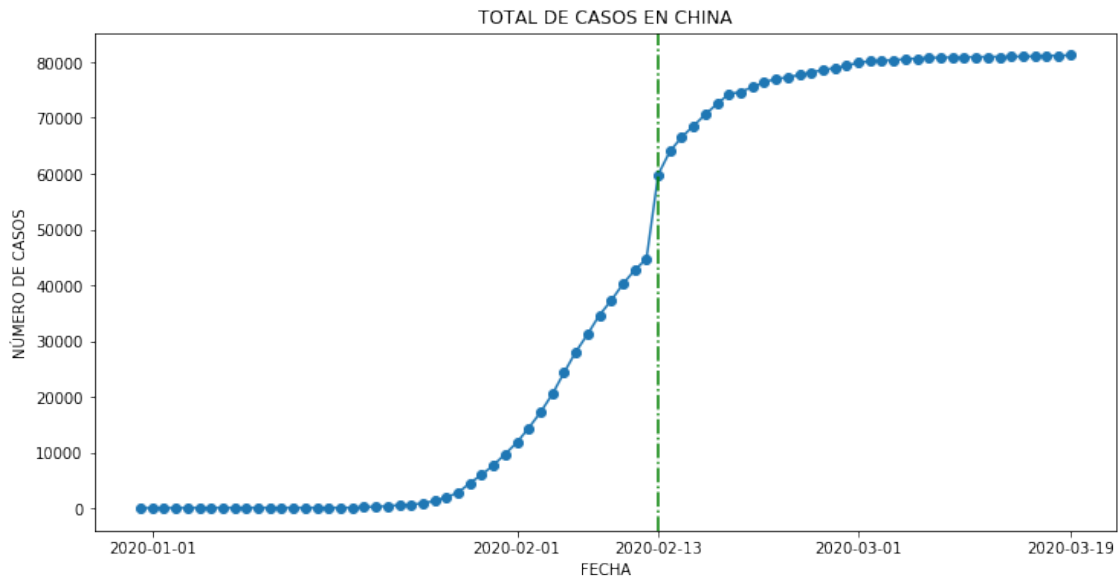
0.0.1 Enseguida se expone una gráfica del crecimiento que ha tenido la epidemia en este país.

```

[91]: plt.figure(figsize=(12,6))
plt.plot(total_cases['date'], total_cases['China'], marker='o')
plt.axvline(x='2020-02-13', color='g', ls='-.') #Grafica línea donde está el
↳máximo incremento

plt.xticks(['2020-01-01', '2020-02-01', '2020-02-13', '2020-03-01',
↳total_cases['date'][len(total_cases)-1]])
plt.xlabel('FECHA')
plt.ylabel('NÚMERO DE CASOS')
plt.title('TOTAL DE CASOS EN CHINA');

```



0.1 China ha disminuído las infecciones por contagio en los últimos días. Sus estrategias de contingencia le están dando muy buenos resultados.

Veamos un gráfico que representa la curva de los casos en China. Para esto, calculamos el número de infectados que aumentan por día en el país. Es decir, como los datos que tenemos son acumulados, hacemos la diferencia por cada día que pasa. Por ejemplo, si el primer día habían 10 casos y al segundo 15 casos, entonces hubo un aumento de 5 personas.

```
[161]: # Vamos a calcular los incrementos por día
china = total_cases[['date', 'China']]
china[0:5]
```

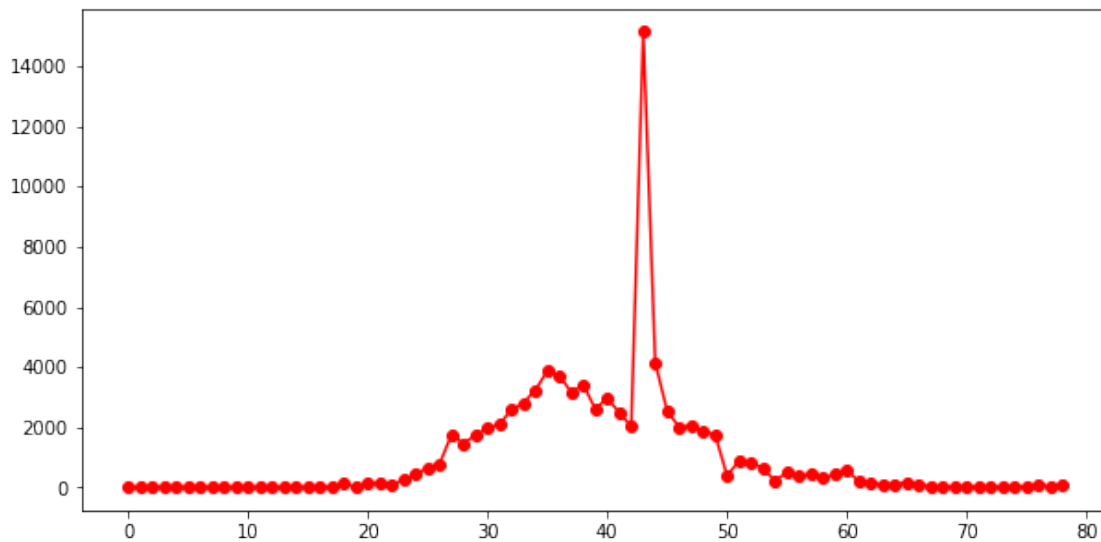
```
[161]:      date  China
0  2019-12-31    27
1  2020-01-01    27
2  2020-01-02    27
3  2020-01-03    44
4  2020-01-04    44
```

```
[93]: incremento = []

for n in range(1, len(china['China'])):
    diferencia = china['China'][n] - china['China'][n-1]
    incremento.append(diferencia)

#len(china['China']), len(incremento)
```

```
plt.figure(figsize=(10,5))
plt.plot(incremento, color='red', marker='o');
```



```
[94]: print("\033[1m \n EL INCREMENTO MÁXIMO APARENTE FUE DE", max(incremento), "\n"
        ↪ "INFECCIONES. \033[0m \n")
```

EL INCREMENTO MÁXIMO APARENTE FUE DE 15141 infecciones

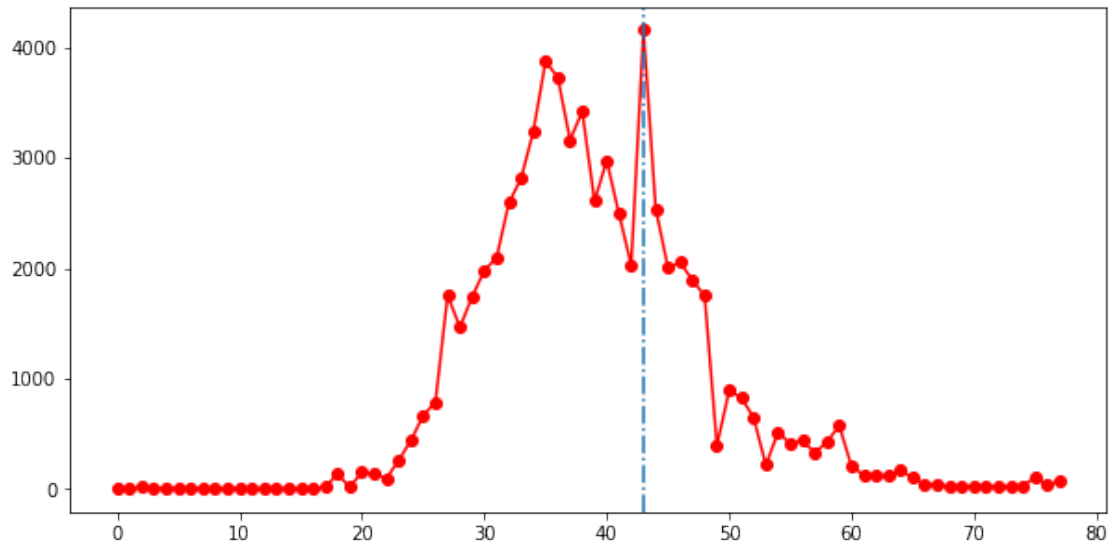
Este cambio intenso no es fidedigno, puesto que hubo un cambio en la forma de contar los casos en China por cuestiones de laboratorio. Si eliminamos los valores atípicos obtenemos lo siguiente:

```
[95]: def outlier_China():

    #Eliminamos el outlier (sin sustituirlo)
    incremento.pop(incremento.index(max(incremento)))

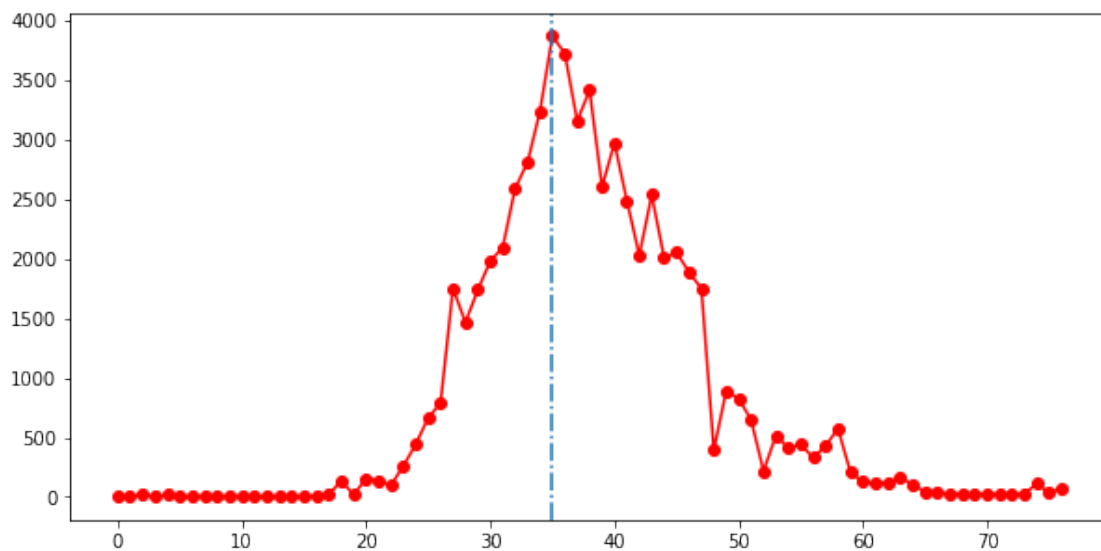
    #Calculamos nuevamente la corrdenada x del nuevo valor máximo y graficamos.
    x_max = incremento.index(max(incremento))
    plt.figure(figsize=(10,5))
    plt.plot(incremento, color='red', marker='o')
    plt.axvline(x=x_max, ymin=0, ls='-.')
    plt.show

    outlier_China()
```



```
[96]: # Parece que hay dos valores que se salen del rango abitual.
      # Así que volvemos a ajustar la gráfica quitando ese valor.
```

```
outlier_China()
```



```
[97]: print("\033[1m \nEL INCREMENTO MÁXIMO FUE DE", max(incremento), "INFECCIONES_
      ↪\033[0m \n")
```

EL INCREMENTO MÁXIMO FUE DE 3872 INFECCIONES


```
[98]: x_max
      # El maximo incremento está entre el índice 42 y 43:
      china['date'][43]
```

```
[98]: '2020-02-12'
```

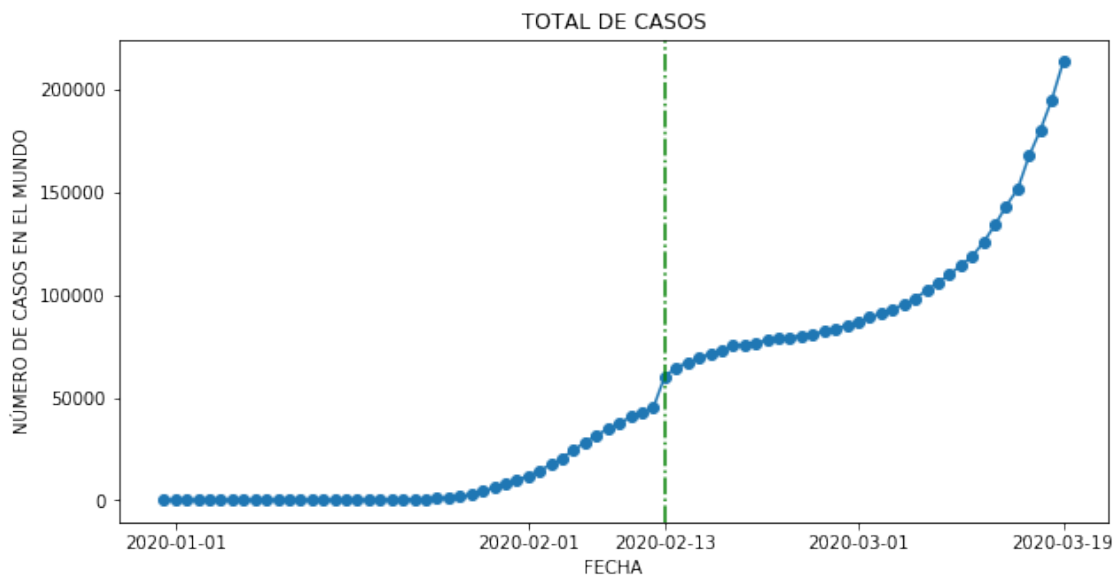
El primer incremento (17 nuevos infectados) se dio del 2 al 3 de enero 2020. El mayor incremento se dio del 11 al 12 de febrero.

0.2 Crecimiento mundial de la epidemia

```
[104]: cases_world = total_cases[['date', 'World']]
```

```
[153]: plt.figure(figsize=(10,5))
      plt.plot(cases_world['date'], cases_world['World'], marker='o')
      plt.axvline(x='2020-02-13', color='g', ls='-.')

      plt.xticks(['2020-01-01', '2020-02-01', '2020-02-13', '2020-03-01',
      ↪ cases_world['date'][len(total_cases)-1]])
      plt.xlabel('FECHA')
      plt.ylabel('NÚMERO DE CASOS EN EL MUNDO')
      plt.title('TOTAL DE CASOS');
```



El crecimiento exponencial de infecciones asegura la categoría de PANDEMIA.

1 Por país

Hacemos una función que haga los cálculos respectivo por país, así como lo hicimos en el caso c

```
[154]: # Casos totales comparados con los del país elegido

def estadisticos(str_country):

    data_frame = total_cases.copy()

    max_world = data_frame['World'].max() # Total de casos mundial
    number_max_cases = data_frame.describe().loc['max',str_country] # Casos
    ↪ totales en el país elegido
    deaths_country = total_deaths[str_country].max()

    porcentaje = (number_max_cases / max_world)*100 # Porcentaje de infectados
    ↪ a nivel mundial

    porcentaje_deaths = (deaths_country/number_max_case)*100 # Casos de
    ↪ muerte comparados con la población local

    #Grafica el número de casos diarios (acumulado)
    plt.figure(figsize=(10,5))
    plt.plot(data_frame['date'], data_frame[str_country], marker='o')
    plt.axvline(x='2020-02-13', color='g', ls='-.')

    plt.xticks(['2020-01-01', '2020-02-01', '2020-02-13', '2020-03-01',
    ↪ data_frame['date'][len(total_cases)-1]])
    plt.xlabel('Fecha')
    plt.ylabel('Número de casos')
    plt.title(f'Total de casos en {str_country}')
    plt.show

    # Cambia los NaN por ceros.
    country = data_frame[['date',str_country]] #Subconjunto con los datos
    ↪ del país que necesitamos.
    country = country.fillna(value = 0)

    # Calcula los incrementos diarios
    incremento = []
    for n in range(1,len(country[str_country])):
        diferencia = country[str_country][n] - country[str_country][n-1]
        incremento.append(diferencia)

    plt.figure(figsize=(10,5))
```

```

#Grafica una línea vertical donde se encuentra el máximo de la gráfica
x_max = incremento.index(max(incremento))
plt.axvline(x=x_max, ymin=0)

#Grafica la curva de incrementos
plt.plot(incremento, color='red', marker='o')
plt.xlabel('Días transcurridos')
plt.ylabel('Incremento')
plt.title(f'Incremento diario en {str_country}')
plt.show

print("\nEl número total de casos en", str_country, "es de",
↪number_max_cases, "infectados.")

print("\nEl incremento máximo fue de ", max(incremento), "infecciones")

print("\nCon un número de muertos de", deaths_country , "personas (el",
↪round(porcentaje_deaths,2), "% de su población infectada).\n")

print(str_country, "abarca el", round(porcentaje,2), "% del total de casos",
↪confirmados a nivel mundial.")

```

2 Algunos países como ejemplo:

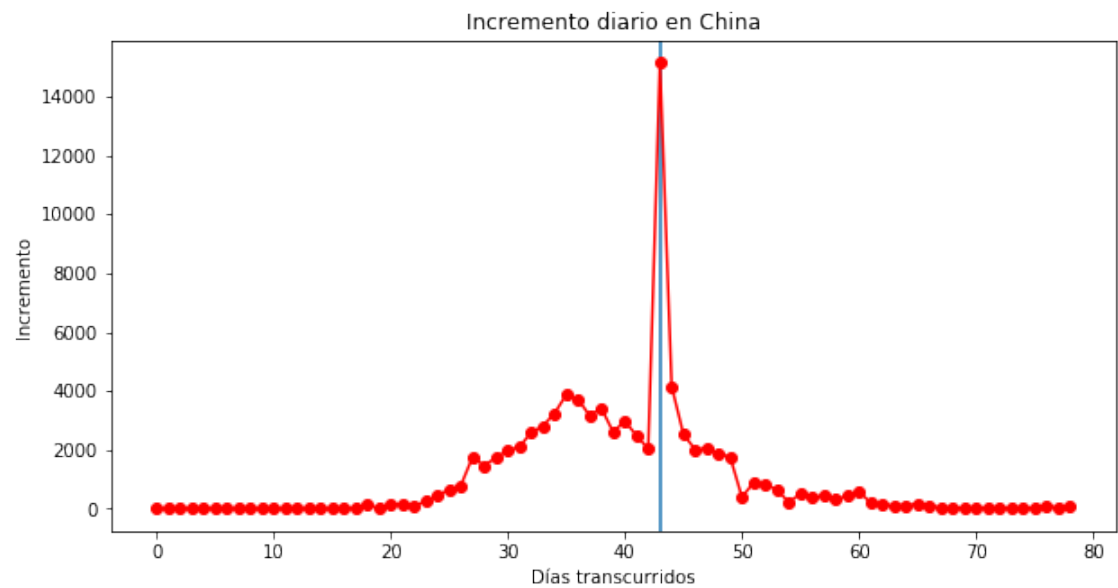
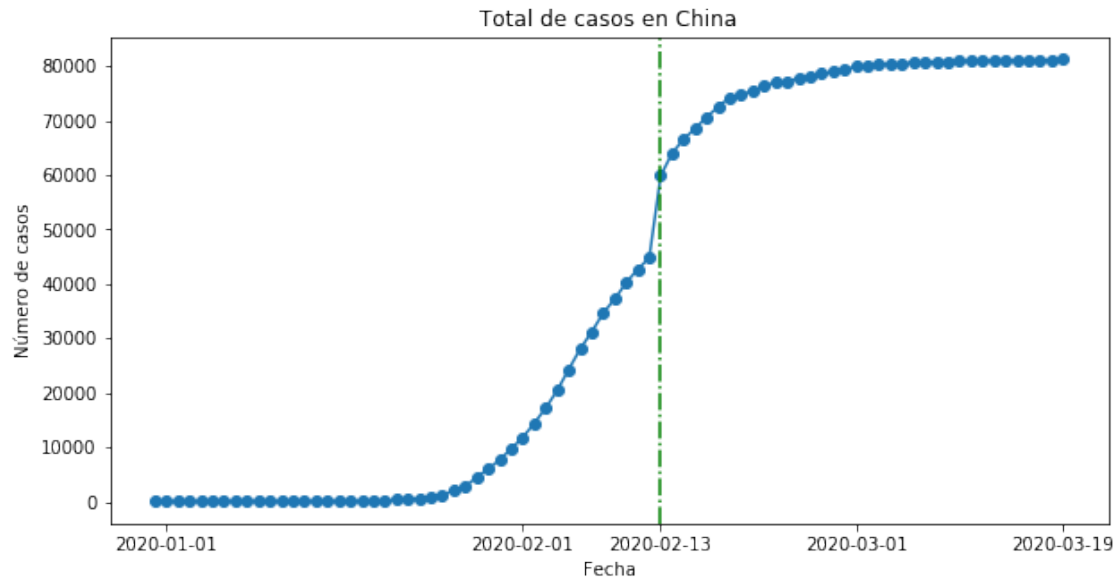
```
[168]: estadisticos('China')
```

El número total de casos en China es de 81238.0 infectados.

El incremento máximo fue de 15141 infecciones

Con un número de muertos de 3250 personas (el 4.0 % de su población infectada).

China abarca el 38.09 % del total de casos confirmados a nivel mundial.



2.1 México

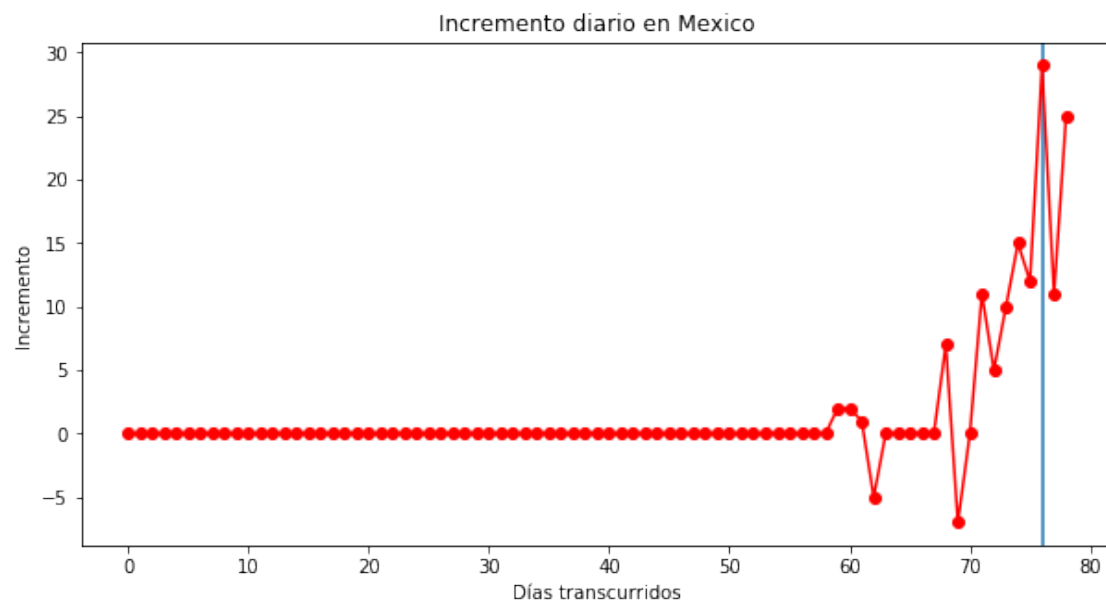
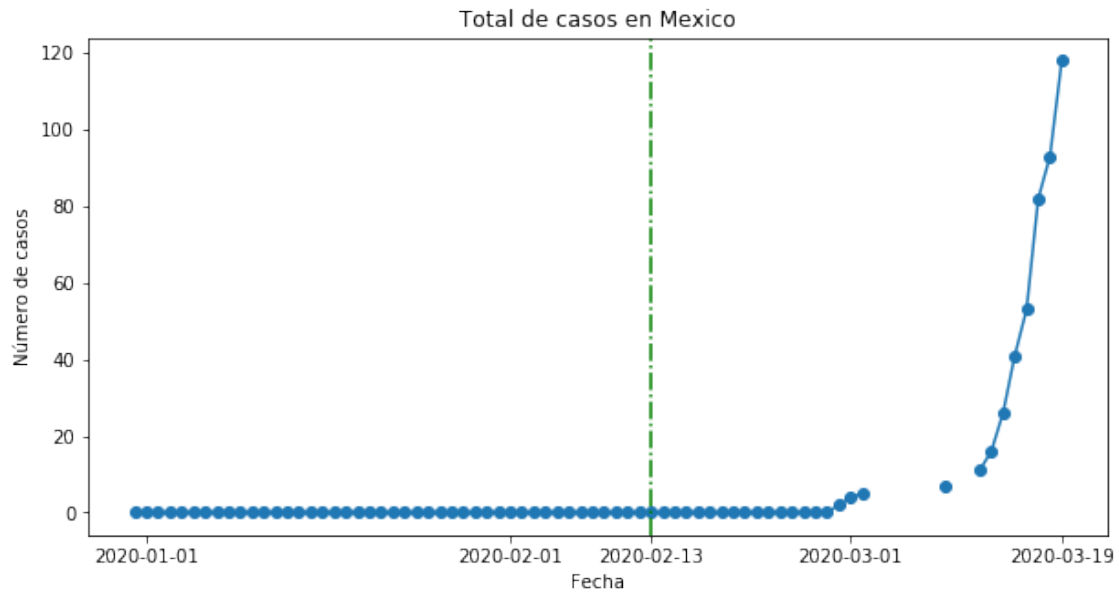
```
[155]: estadisticos('Mexico')
```

El número total de casos en Mexico es de 118.0 infectados.

El incremento máximo fue de 29.0 infecciones

Con un número de muertos de 0.0 personas (el 0.0 % de su población infectada).

Mexico abarca el 0.06 % del total de casos confirmados a nivel mundial.



```
[167]: total_cases['Mexico'].values
```

```
[167]: array([ 0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  0.,  0.,  2.,  4.,  5., nan, nan, nan,
        nan, nan, nan,  7., nan, nan, 11., 16., 26., 41., 53.,
        82., 93., 118.]
```

2.1.1 Como puede verse, faltan registros para México. Pueden ajustarse con ayuda de la primer gráfica haciendo un seguimiento de su forma, pero apesar de ello, no cabe duda que el crecimiento es exponencial. Y podríamos decir (a groso modo) que se esperarían entre 50 y 200 casos nuevos de infectados del 20 al 21 de marzo, ya que ha tenido un control escaso (y notemos que la información que ofrece la OMS ha tenido por lo menos un día de retraso).

2.2 Canadá

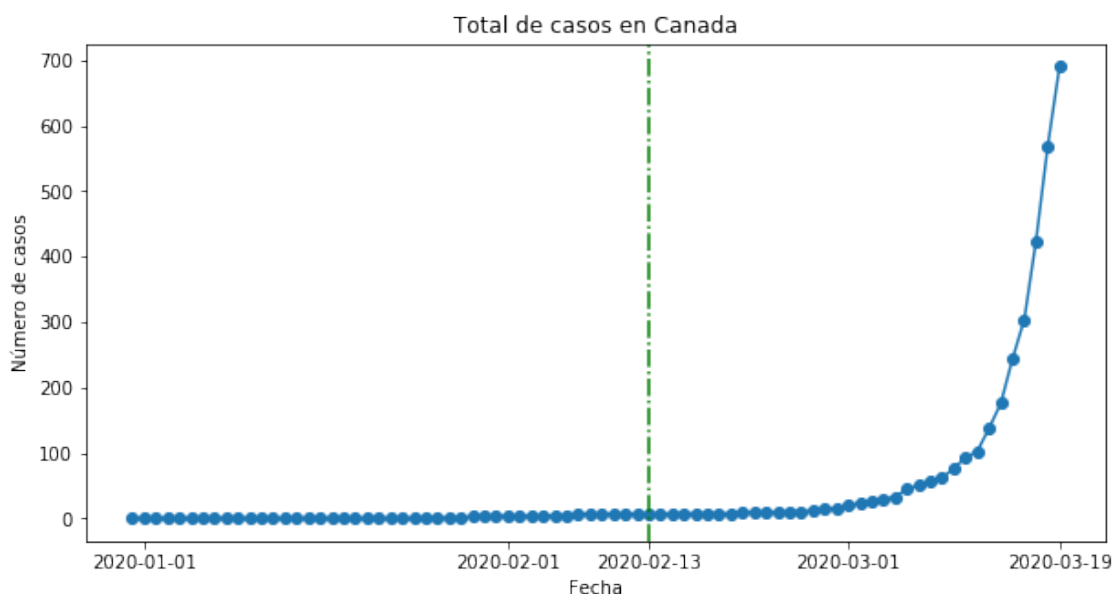
```
[156]: estadisticos('Canada')
```

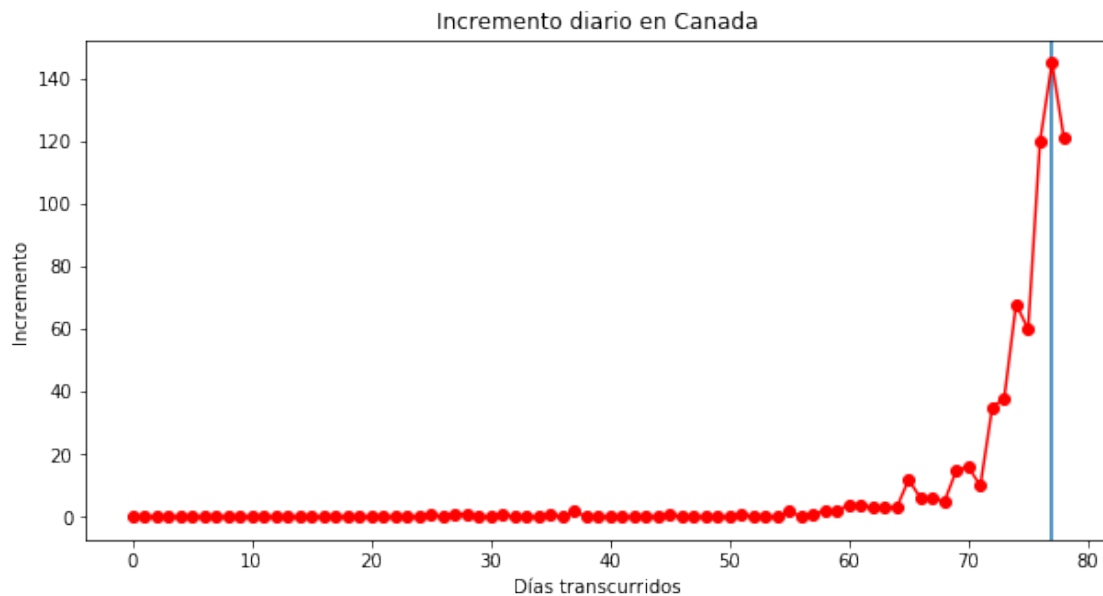
El número total de casos en Canada es de 690.0 infectados.

El incremento máximo fue de 145 infecciones

Con un número de muertos de 9 personas (el 0.01 % de su población infectada).

Canada abarca el 0.32 % del total de casos confirmados a nivel mundial.





2.3 Estados Unidos

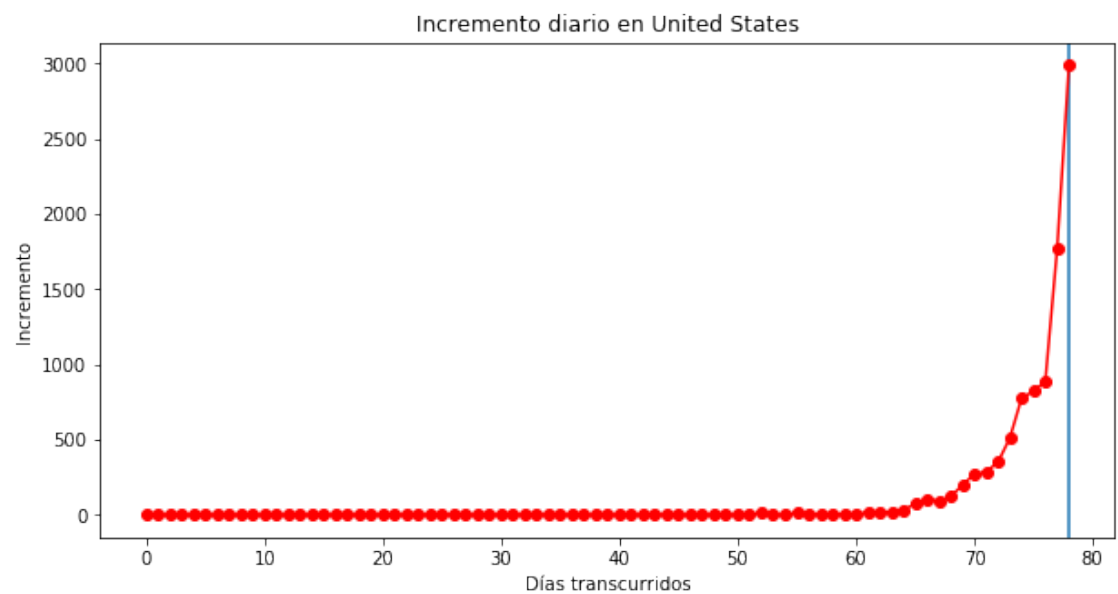
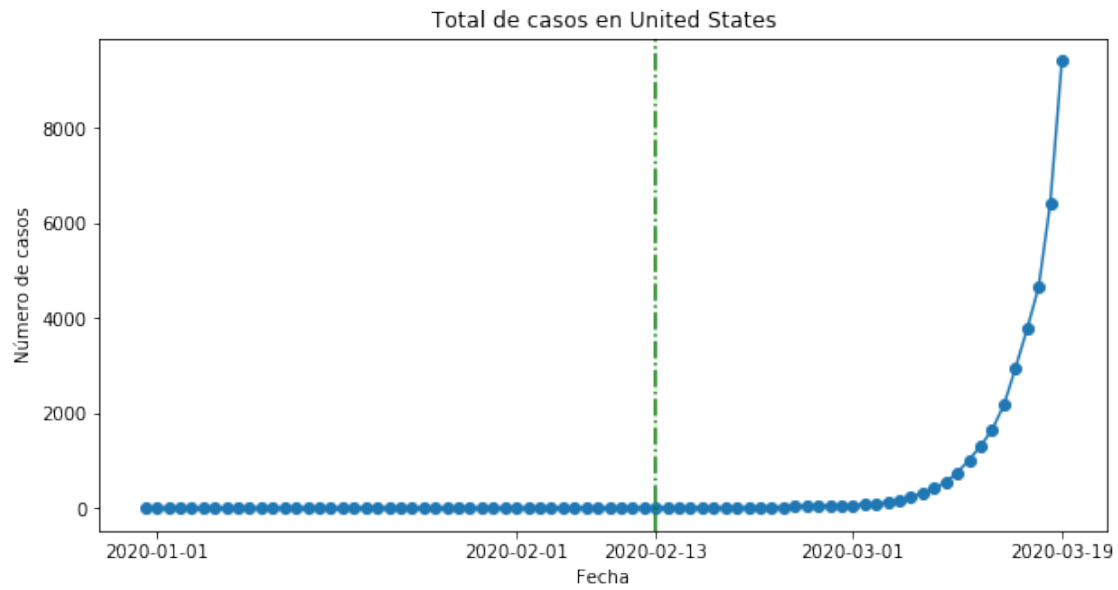
```
[160]: estadisticos('United States')
```

El número total de casos en United States es de 9415.0 infectados.

El incremento máximo fue de 2988 infecciones

Con un número de muertos de 150 personas (el 0.18 % de su población infectada).

United States abarca el 4.41 % del total de casos confirmados a nivel mundial.



2.4 Japón

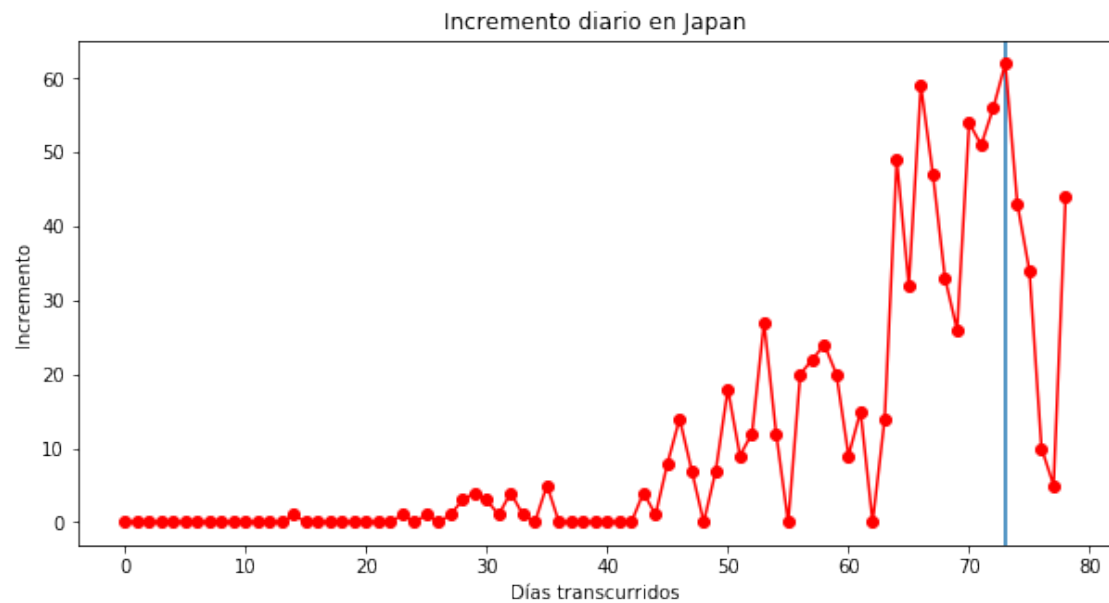
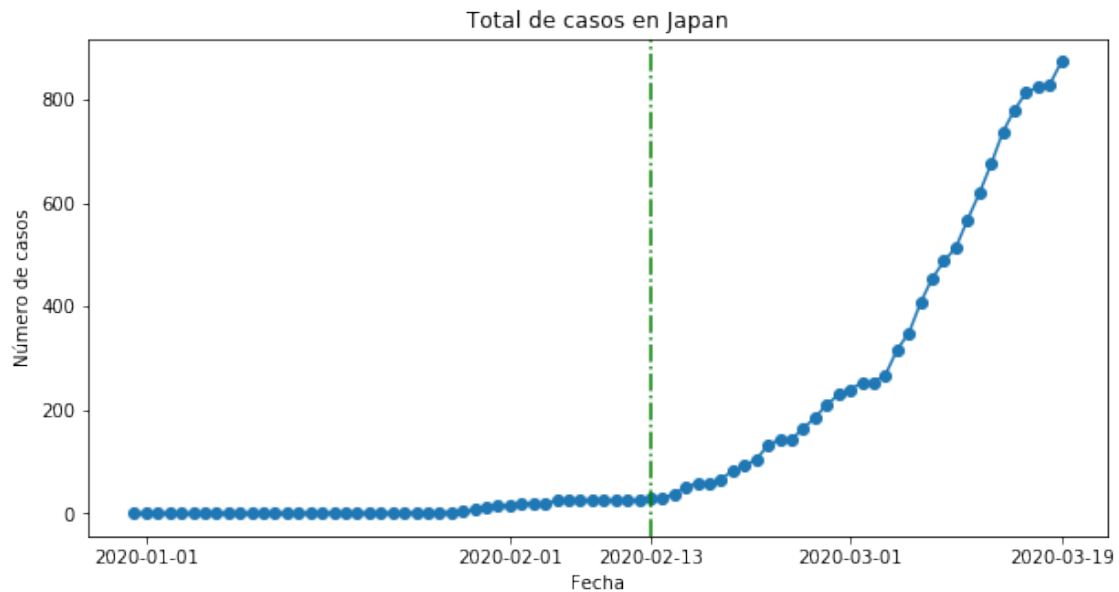
```
[157]: estadisticos('Japan')
```

El número total de casos en Japan es de 873.0 infectados.

El incremento máximo fue de 62 infecciones

Con un número de muertos de 29 personas (el 0.04 % de su población infectada).

Japan abarca el 0.41 % del total de casos confirmados a nivel mundial.



2.5 España

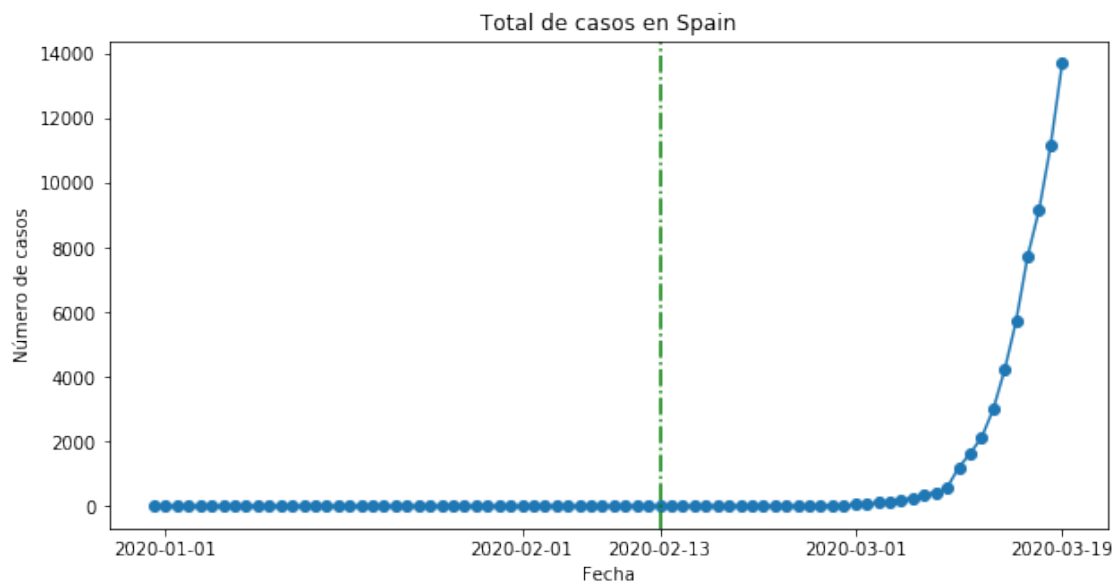
```
[158]: estadisticos('Spain')
```

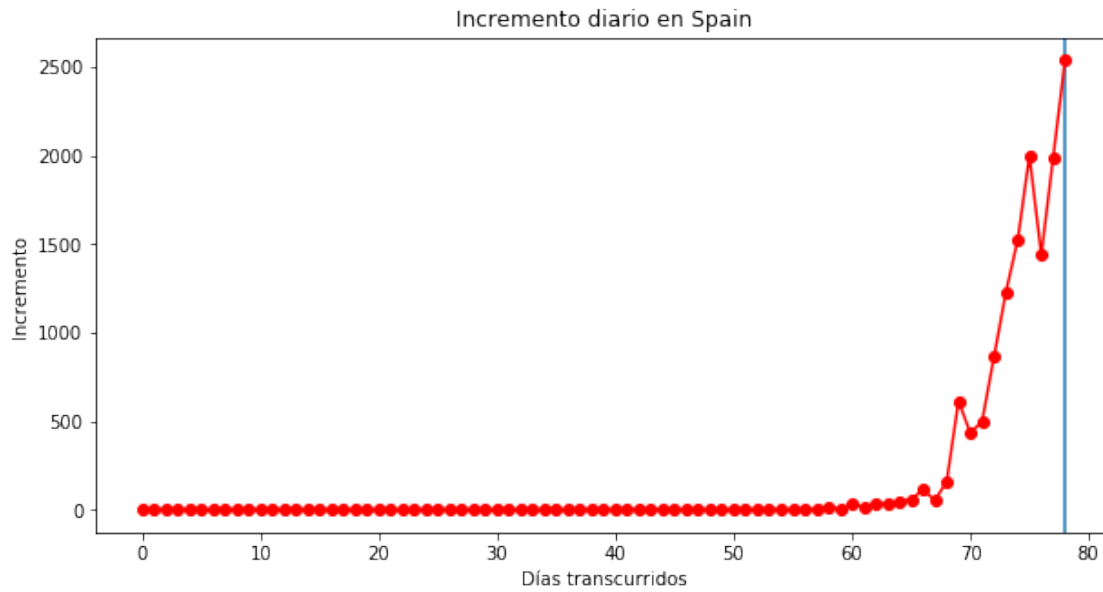
El número total de casos en Spain es de 13716.0 infectados.

El incremento máximo fue de 2538 infecciones

Con un número de muertos de 598 personas (el 0.74 % de su población infectada).

Spain abarca el 6.43 % del total de casos confirmados a nivel mundial.





2.6 Italia

```
[159]: estadisticos('Italy')
```

El número total de casos en Italy es de 35713.0 infectados.

El incremento máximo fue de 6230 infecciones

Con un número de muertos de 2978 personas (el 3.67 % de su población infectada).

Italy abarca el 16.75 % del total de casos confirmados a nivel mundial.

