

Análisis de la evolución de la incidencia de la COVID-19 en España

Juan Matorras Díaz-Caneja

16/11/2020

Introducción

Este es un ejercicio básico de análisis de los datos de la incidencia de la COVID-19 en España a lo largo de 2020. Habiendo la cantidad de informes y herramientas para el análisis de los datos sobre la incidencia de la COVID-19 que ya existen, este documento no pretende aportar nada singularmente nuevo y su razón de ser no es otra que poner en práctica y profundizar por mi parte en el aprendizaje de las técnicas de análisis de datos y el lenguaje R que he iniciado a finales de este año 2020.

Los datos de partida son los publicados por el Gobierno de España en la web **datos.gob.es** a través del siguiente enlace: <https://datos.gob.es/es/catalogo/e05070101-evolucion-de-enfermedad-por-el-coronavirus-covid-19>.

El grueso del informe se centra sobre los totales en España agregando los datos disponibles por Comunidades Autónomas, aunque también se muestran información de las CCAA de Madrid y Cantabria. La razón de la selección de estas dos comunidades y no otras es simple y llanamente que nací y crecí en la última y mantengo vínculos familiares y de amistad allí, mientras que en la primera he pasado básicamente la mitad de mi vida, sigo viviendo en ella y previsiblemente así seguirá siendo en los próximos años.

El archivo de datos no ha sido sometido a ningún tipo de modificación o alteración previa y su manipulación en este análisis es el mínimo imprescindible para permitir el tratamiento de los datos y obtención de resultados.

```
DatosCompleto <- read.csv(file.path("data", "datos_ccaas.csv"))
## Descartamos el método de detección (PCR, antígenos,...)
Datos <- DatosCompleto[, 1:3]
Datos$fecha <- as.Date(Datos$fecha, format = "%Y-%m-%d")
```

La fecha y hora de descarga de los datos que han sido utilizados para las tablas y gráficos incluidos en este informe ha sido (aaaa-mm-dd hh:mm:ss): 2020-11-16 13:57:10

El análisis se ha llevado a cabo utilizando el software libre para análisis estadístico **R**, versión 4.0.2. (1)

Se ha hecho uso también de los paquetes complementarios:

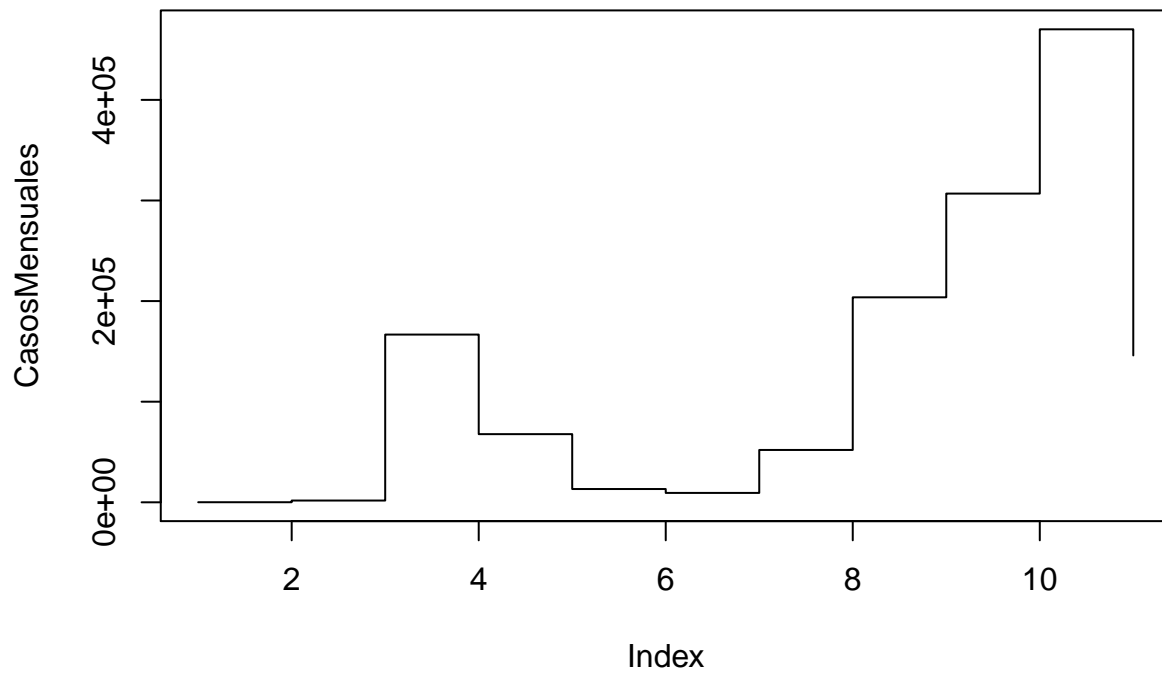
- **lubridate** para facilitar el manejo de fechas. (2)
- **knitr** para mejorar la apariencia de tablas. (3)

Incidencia mensual y número total de casos detectados desde el inicio de 2020

La evolución de número de casos notificados por meses se reflejan en la siguiente tabla y el gráfico que la acompaña:

```
CasosMensuales <- tapply(Datos$num_casos, month(Datos$fecha), sum)
TotalCasosOrigen <- sum(Datos$num_casos)
PorcentajePoblacion <- paste(format(TotalCasosOrigen / 47.4E+06 *100, digits = 4), "%")
TCasosMensuales <- cbind(head(Meses, length(CasosMensuales)),
  format(CasosMensuales, big.mark = ".", decimal.mark = ","))
colnames(TCasosMensuales) <- c("Mes", "Nº Casos Mensuales")
```

Mes	Nº Casos Mensuales
E	81
F	1.742
M	166.711
A	67.732
M	13.188
J	9.310
J	52.032
A	203.738
S	306.863
O	470.128
N	145.975



El número total de casos acumulados desde el 1 de enero de 2020 hasta la fecha indicada en el punto anterior según los datos oficiales disponibles en ese momento ascienden a un total de 1.437.500.

Considerando una población en España de 47,4 millones de personas según los datos publicados por el INE (Instituto Nacional de Estadística) correspondiente al inicio de año, el porcentaje de contagio de la población es 3.033 %.

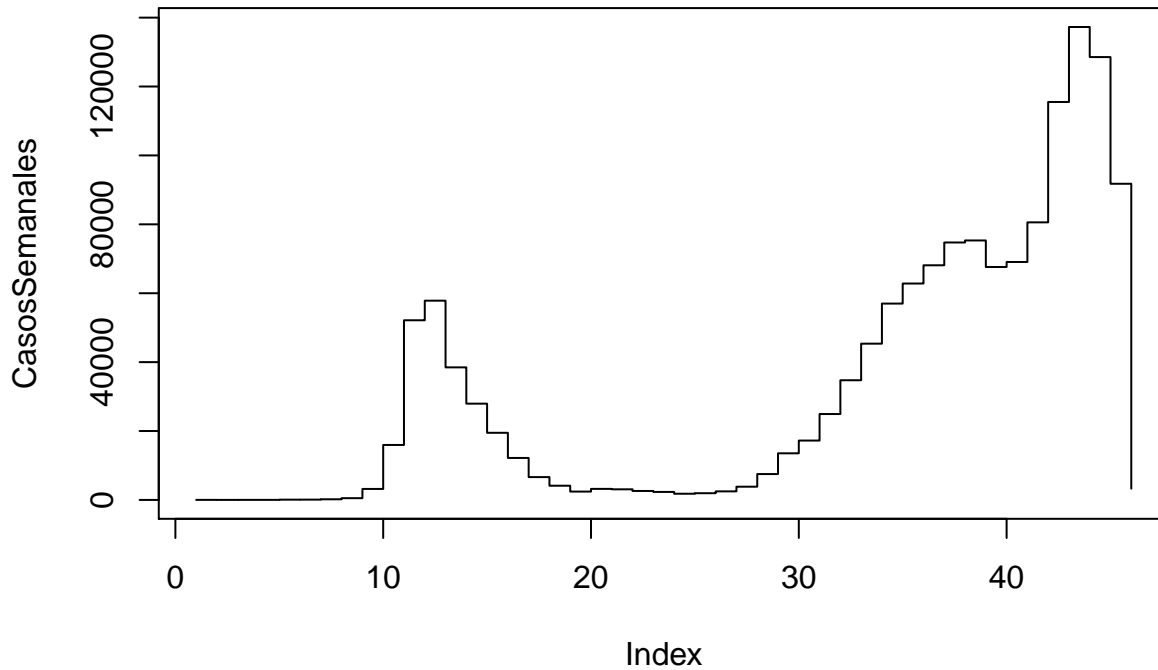
Llama singularmente la atención la baja tasa de contagio entre la población en contraste con el colapso hospitalario sufrido tanto en el los primeros meses del año como el que se parece estar sufriendo en estas últimas semanas. Antes de llegar a una conclusión precipitada sobre una insuficiencia estructural de medios de atención, hay que tener en cuenta dos factores fundamentales de la COVID-19 que no han sido analizados en este estudio que probablemente la distinguan de otras enfermedades y patologías:

- Aunque el porcentaje de personas infectadas sea bajo, el porcentaje de infectados que requieran atención especial tipo UCI puede ser, y probablemente así sea, muy elevado.
- Otro factor a tener en cuenta es el tiempo por el que se van a tener que prolongar esos cuidados especiales, al menos en comparación con las patologías más comunes que normalmente se atienden en las UCIs.

Incidencia semanal

Evolución de número de casos notificados por semanas:

```
CasosSemanales <- tapply(Datos$num_casos, week(Datos$fecha), sum)
plot(CasosSemanales, type = "s")
```



Incidencia diaria

Curva epidémica de los casos notificados por días:

```
CasosDiarios <- tapply(Datos$num_casos, yday(Datos$fecha), sum)
plot(CasosDiarios, type = "s")
```

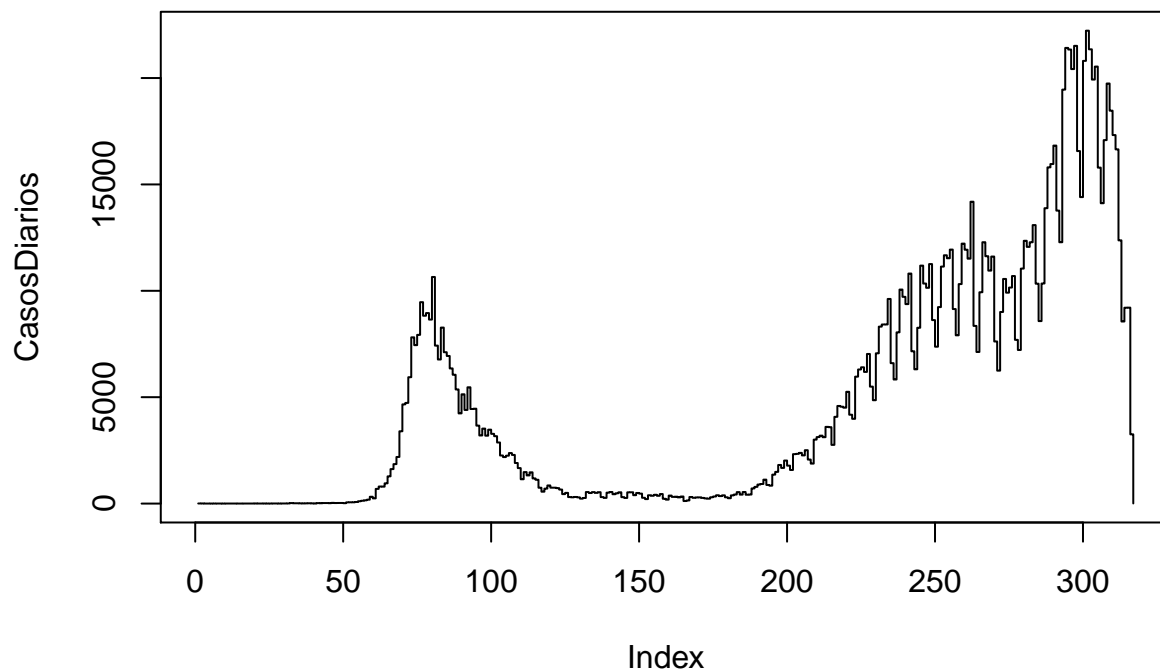
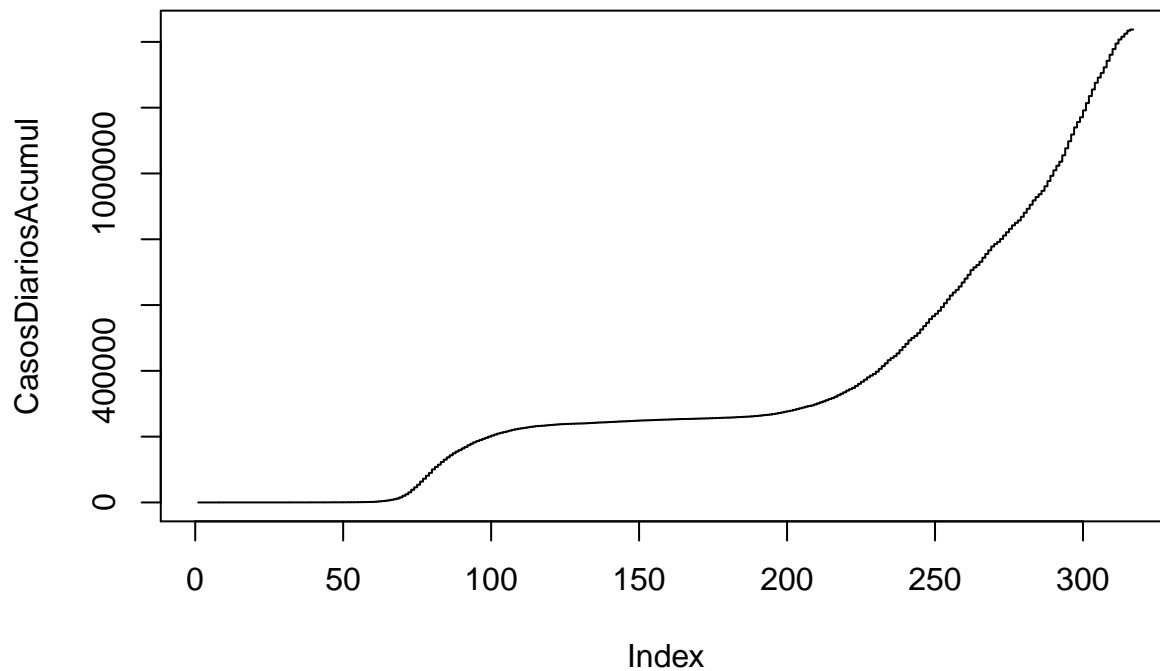


Gráfico de casos acumulados a origen por días:

```
CasosDiariosAcumul <- cumsum(CasosDiarios)
plot(CasosDiariosAcumul, type = "s")
```



Detalle del número de casos en los dos primeros meses de 2020

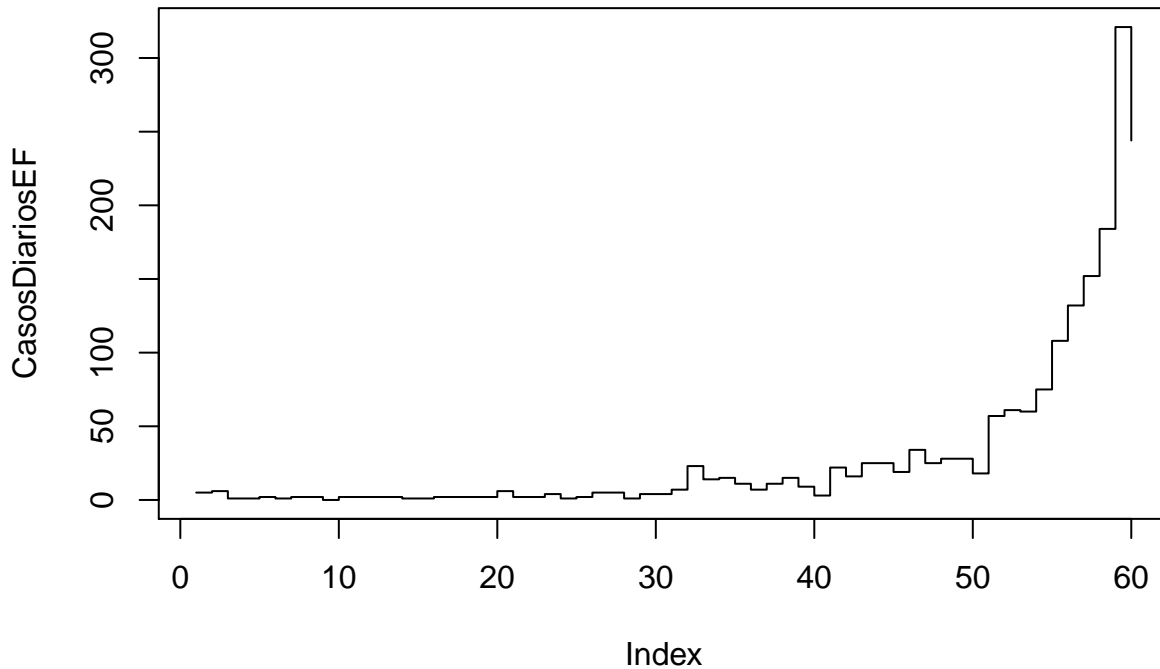
Evolución diaria del número de casos durante los dos primeros meses del año:

```
PeriodoEF <- seq.Date(
  from = as.Date("2020-01-01"),
  to = as.Date("2020-02-29"),
```

```

    by = "day"
  )
  DatosEF <- subset(Datos, fecha %in% PeriodoEF)
  CasosDiariosEF <- tapply(DatosEF$num_casos, yday(DatosEF$fecha), sum)
  TotalCasosEF <- sum(DatosEF$num_casos)
  plot(CasosDiariosEF, type = "s")

```



Los casos reportados totales a lo largo de esos dos meses son 1.823, si bien es claro que se produce una acusada inflexión en la pendiente de evolución a partir del día 50.

```

Periodo1_50 <- seq.Date(
  from = as.Date("2020-01-01"),
  to = as.Date("2020-02-19"),
  by = "day")
Periodo51_60 <- seq.Date(
  from = as.Date("2020-02-20"),
  to = as.Date("2020-02-29"),
  by = "day")
Datos1_50 <- subset(Datos, fecha %in% Periodo1_50)
Datos51_60 <- subset(Datos, fecha %in% Periodo51_60)
TotalCasos1_50 <- sum(Datos1_50$num_casos)
TotalCasos51_60 <- sum(Datos51_60$num_casos)

```

Siendo así que el desglose de número agregado de casos identificados en dichos primeros 50 días y los siguientes 10 días queda de la siguiente manera:

- Periodo 1-50: 429
- Periodo 51-60: 1.394

```

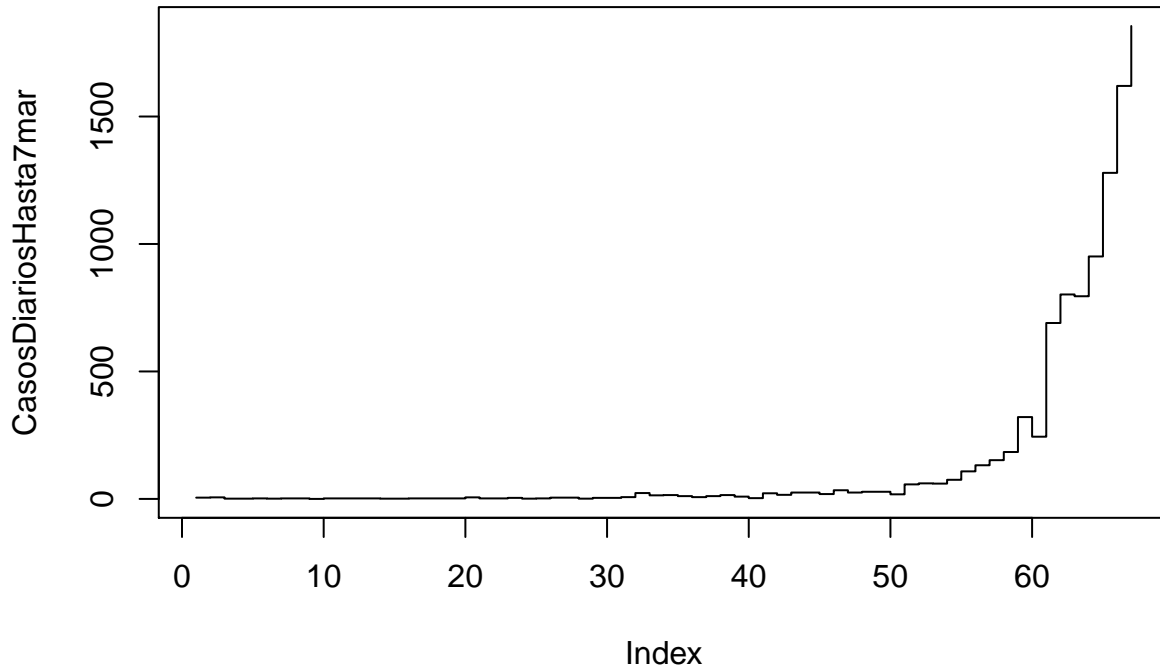
PeriodoSem1mar <- seq.Date(
  from = as.Date("2020-03-01"),
  to = as.Date("2020-03-07"),
  by = "day")
DatosSem1mar <- subset(Datos, fecha %in% PeriodoSem1mar)

```

```
CasosDiariosSem1mar <- tapply(DatosSem1mar$num_casos, yday(DatosSem1mar$fecha), sum)
TotalCasosSem1mar <- sum(DatosSem1mar$num_casos)
```

En los primeros siete días de marzo la progresión diaria de nuevos casos siguió disparándose, resultando un total de 7.992 casos a añadir al total anterior, siendo éstos 4.4 veces los registrados a lo largo de enero y febrero.

Gráfico del número de casos diarios desde inicio de año 2020 hasta el 7 de marzo:



.....

- (1) R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- (2) Garrett Golemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL: <http://www.jstatsoft.org/v40/i03/>
- (3) Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.