



UNIVERSIDAD  
NEBRIJA

MÁSTER EN  
DATA  
SCIENCE Y  
ANÁLISIS DE  
DATOS

# ESTUDIO DE LA INFLUENCIA ENTRE LA TEMPERATURA MEDIA DE UN LUGAR Y LA INCIDENCIA DEL CORONAVIRUS SARS-COV2

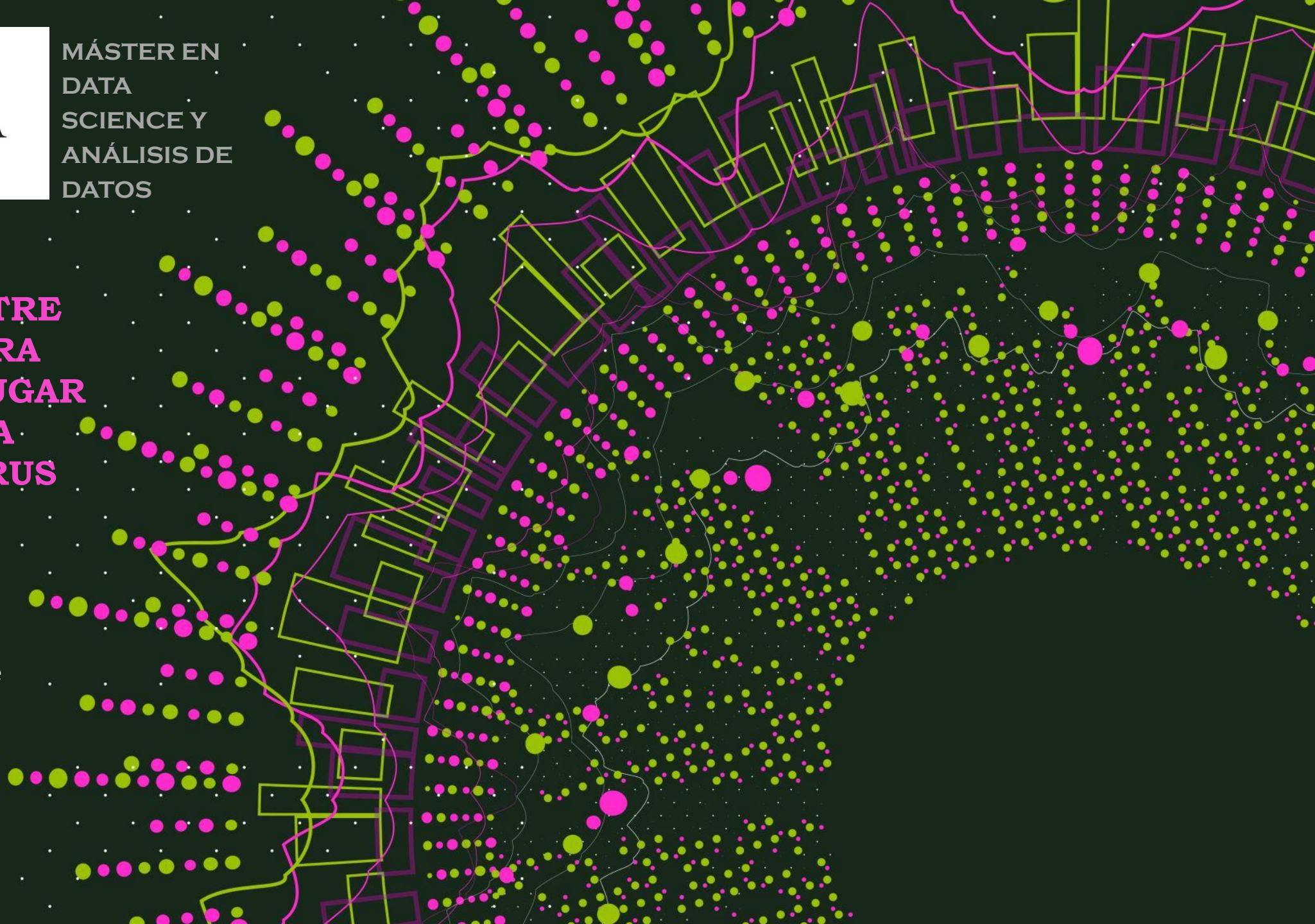
AUTOR: Beatriz Cubillas García

DIRECTOR: Rafael Marín Sastre

FECHA: Septiembre 2021

ENTIDAD COLABORADORA

Instituto Europeo de Estudios  
Empresariales



# RESUMEN DEL CONTENIDO

- PROYECTO FIN DE MASTER DE DATA SCIENCE
- ESTUDIO DE INVESTIGACIÓN
- PERSPECTIVA DE CORRELACION: relación de influencia entre dos o mas variables.
- Estudio de seis provincias españolas

## ► ANTECEDENTES

- Pandemia mundial en el último año y medio
- Datos y estadísticas diarias: n° de contagios diarios, incidencia a 14 días, personas hospitalizadas, ...
- “Ola” : episodios de contagio masivo sin control en la población.
- Variaciones de la incidencia en las distintas épocas del año.

## ► OBJETIVOS

- Análisis de correlación entre dos variables:
  - ✓ N° de casos positivos nuevos diarios
  - ✓ Temperatura media diaria
- Concluir si es posible tomar medidas de control de la pandemia en función de la temperatura media.



# DELIMITACIÓN DEL TEMA DE ESTUDIO

## CONTEXTO GEOGRÁFICO

- Se eligen seis provincias que tienen una población similar:

- León
- Huelva
- Cantabria
- Lleida
- Ciudad Real
- Castellón

Tienen diferentes condiciones climáticas y de situación en la geografía española

## CONTEXTO TEMPORAL

- Duración del estudio: 1 año.
  - Desde el 1 de julio de 2020 (fin del confinamiento estricto)
  - Hasta el 30 de junio de 2021.

# DATOS DE LAS PROVINCIAS SELECCIONADAS

PROVINCIA	POBLACIÓN PROVINCIAL	ALTITUD MEDIA	COMUNIDAD AUTÓNOMA	SITUACIÓN
LEÓN	455.066	912	Castilla y León	Centro-Norte
HUELVA	530.386	19	Andalucía	Sur- Océano Atlántico
CANTABRIA	583.904	52	Cantabria	Norte – Mar Cantábrico
LLEIDA	436.911	185	Cataluña	Noreste
CIUDAD REAL	492.065	628	Castilla – La Mancha	Centro
CASTELLÓN	576.941	43	Comunidad Valenciana	Este- Mediterráneo

# DESCRIPCIÓN DEL MÉTODO UTILIZADO

## PROCESO ETL

### EXTRACCIÓN

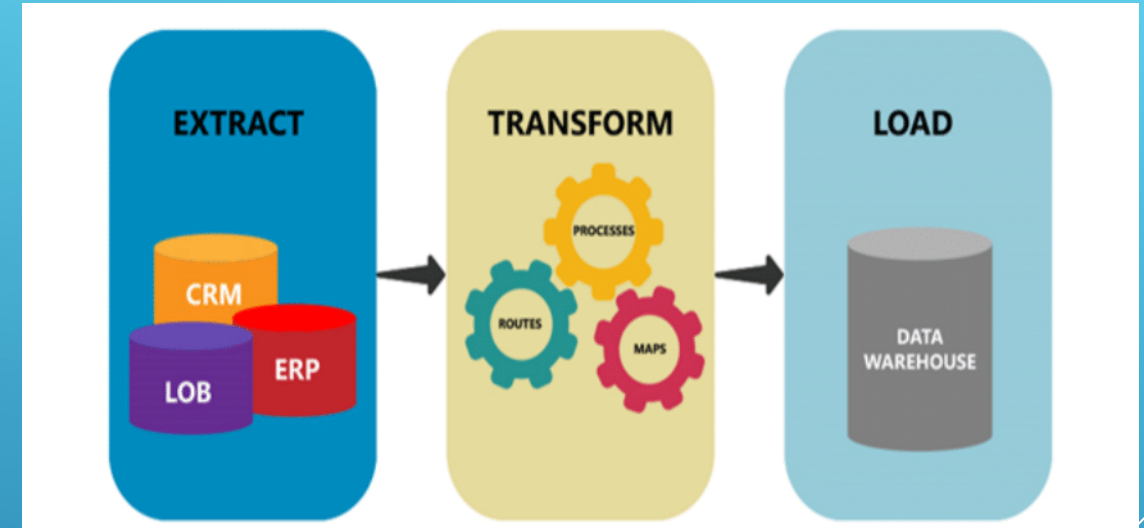
- Datos climatológicos del AEMET
- Datos de incidencia de Covid de las distintas comunidades autónomas.

### TRANSFORMACIÓN

- Adecuación de los datos para su tratamiento

### CARGA (Load)

- Carga de los datos en RStudio para obtener análisis estadísticos y gráficos



# PROCEDIMIENTO DE TRABAJO

## ► PROCESO ETL

- ❑ DATOS CLIMATOLÓGICOS
- ❑ DATOS DE INCIDENCIA “COVID”
- ❑ TRANSFORMACIÓN DATOS
- ❑ CARGA DE DATOS EN RSTUDIO
- ❑ GENERAR DATAFRAME

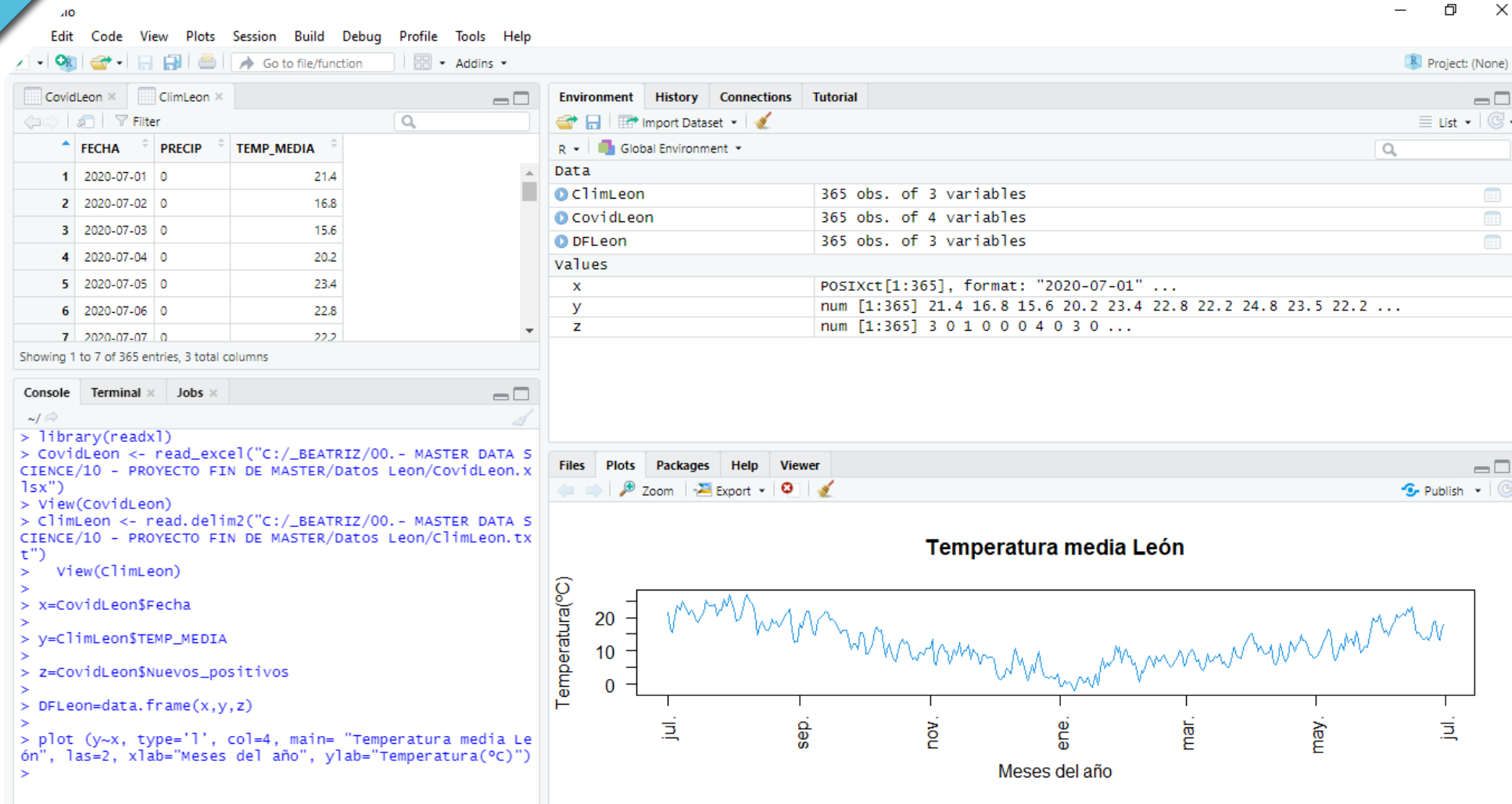
## ► EN R

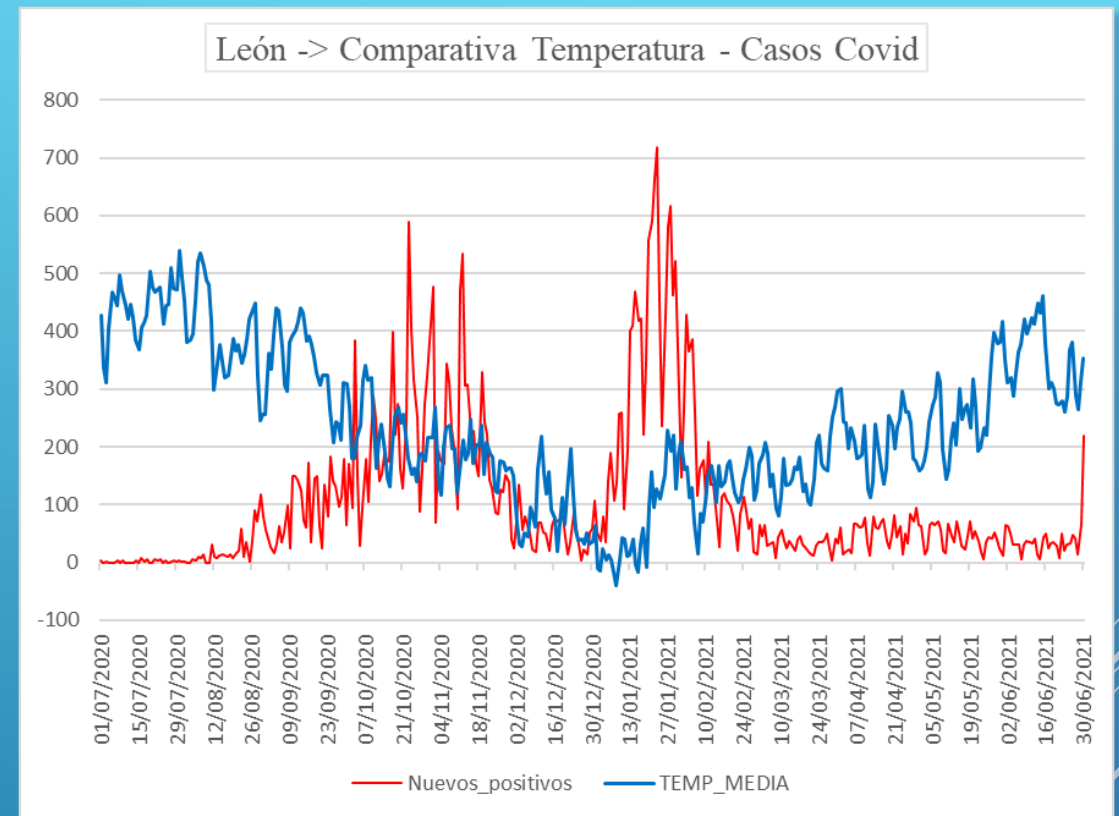
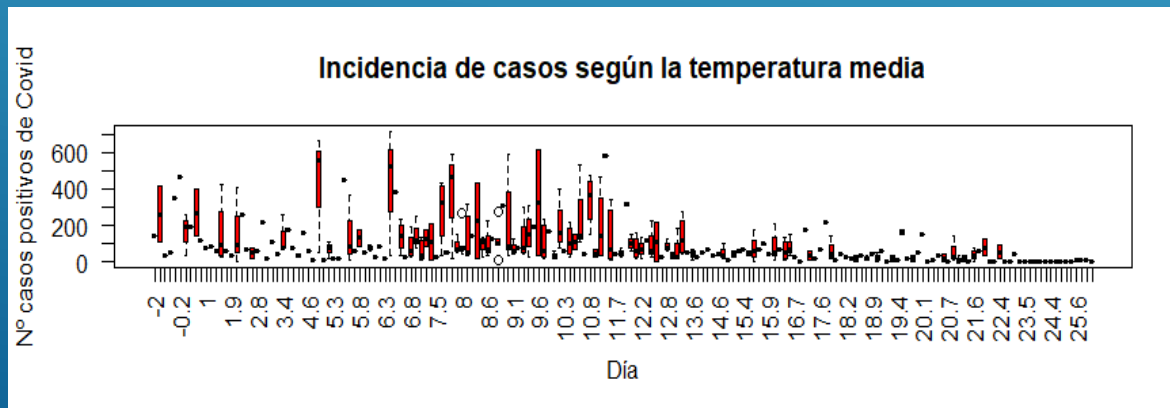
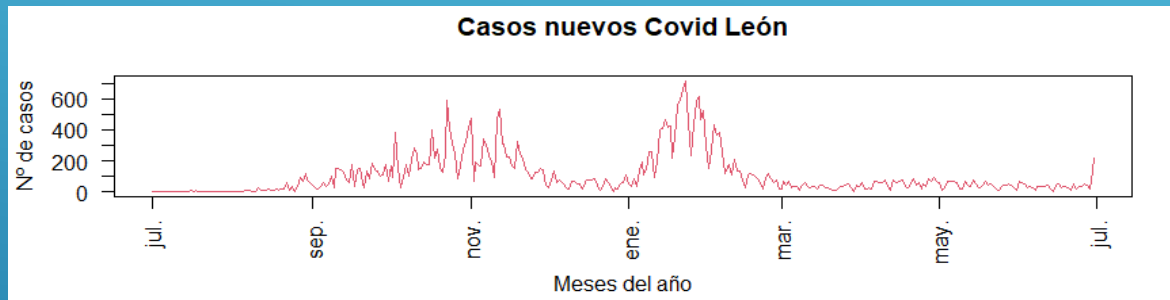
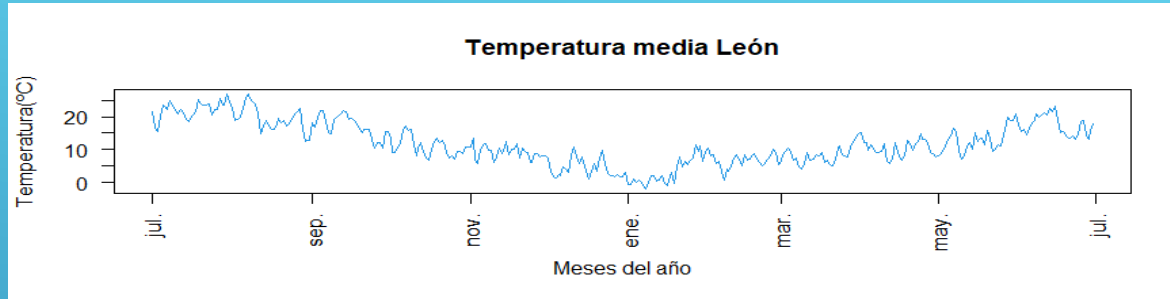
- GRÁFICA TEMPERATURA
- GRÁFICA INCIDENCIA DE CASOS “COVID”
- COMPARATIVA TEMPERATURA – INCIDENCIA CASOS “COVID”
- GRAFICO DEL NÚMERO DE CASOS POR DÍA EN FUNCIÓN DE LA TEMPERATURA MEDIA

# MAPA DE LOS TIPOS DE CLIMA EN ESPAÑA









## TEMPERATURA Y CLIMA

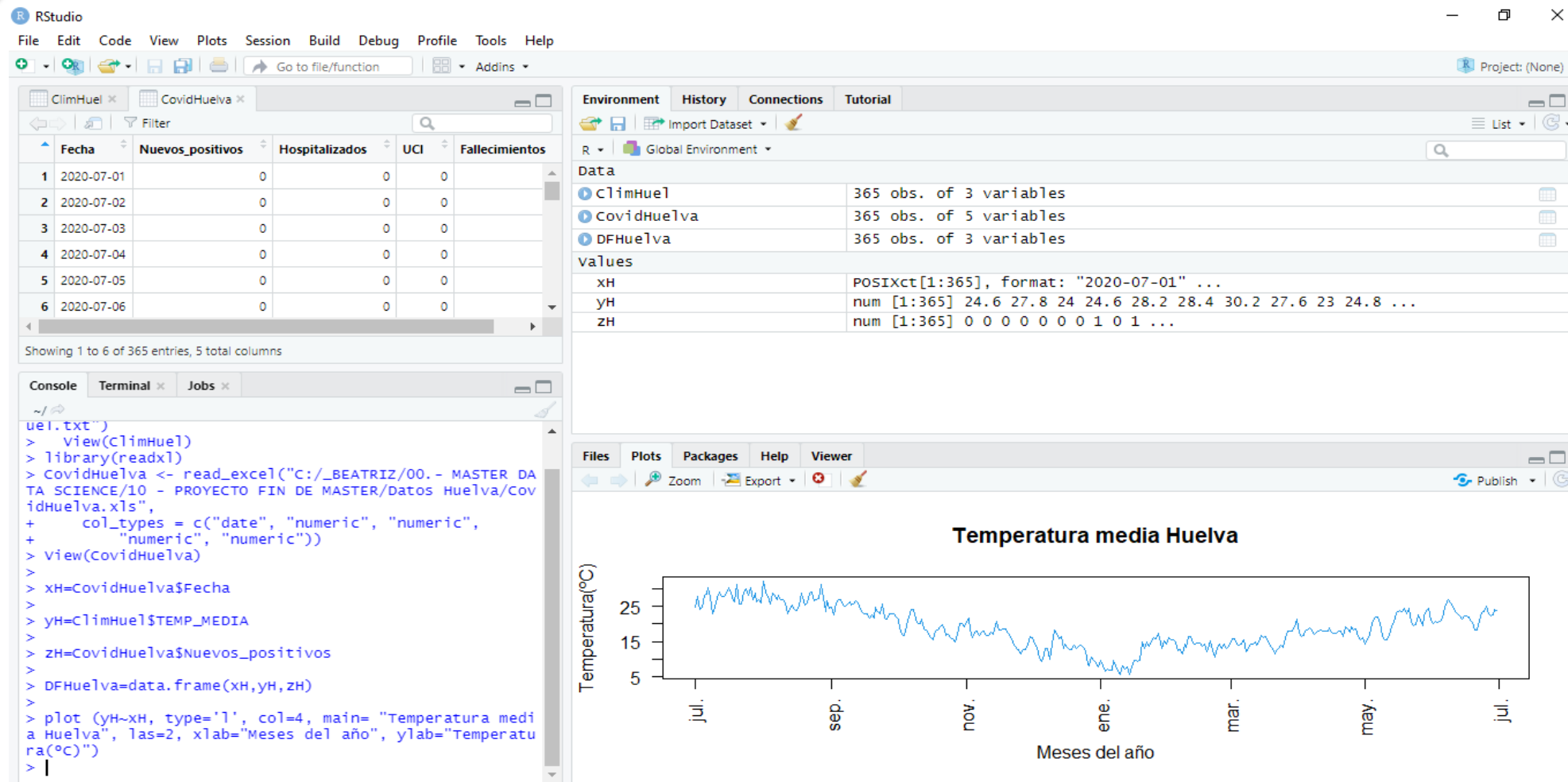
- ▶ Clima Mediterráneo Interior
- ▶ Veranos frescos (22° mes mas cálido)
- ▶ Inviernos considerablemente fríos (entre 6°C y -3°C en el mes mas frío)

## DATOS COVID

- ▶ Dos “olas” de contagio al final de Octubre de 2020 y en febrero de 2021.

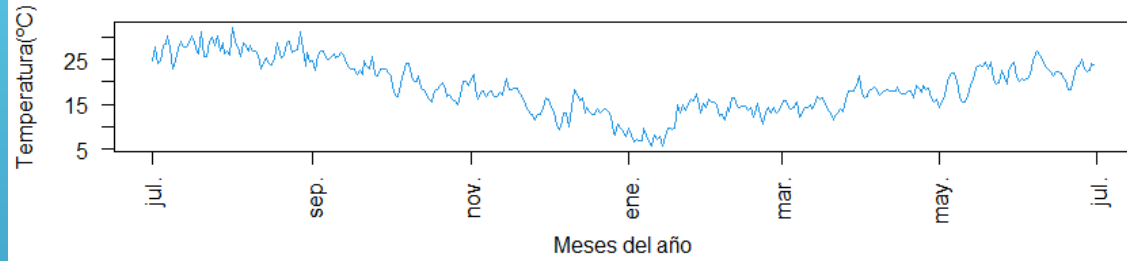
## Observaciones

- ▶ De julio a septiembre de 2020 y entre febrero y junio de 2021, se observa que “a mayor temperatura media, menor incidencia de nuevos casos”.
- ▶ Los días con temperatura media por debajo de los 12°C tienen una mayor incidencia de casos.

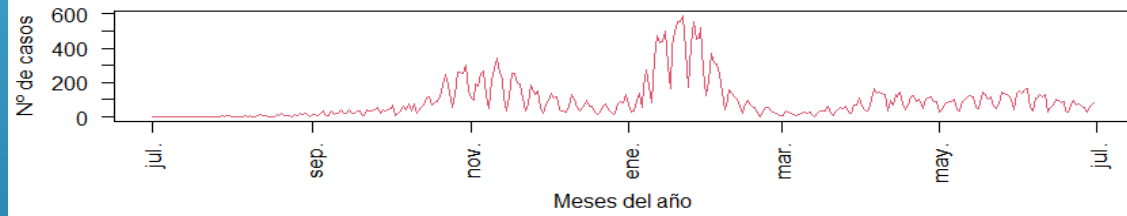




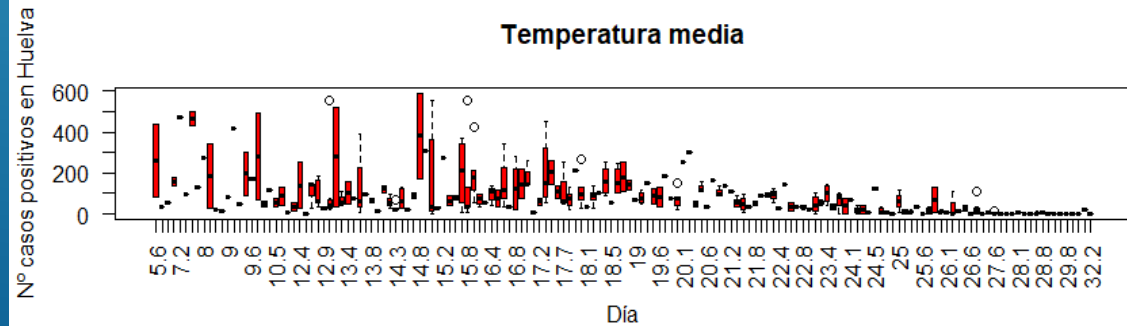
## Temperatura media Huelva



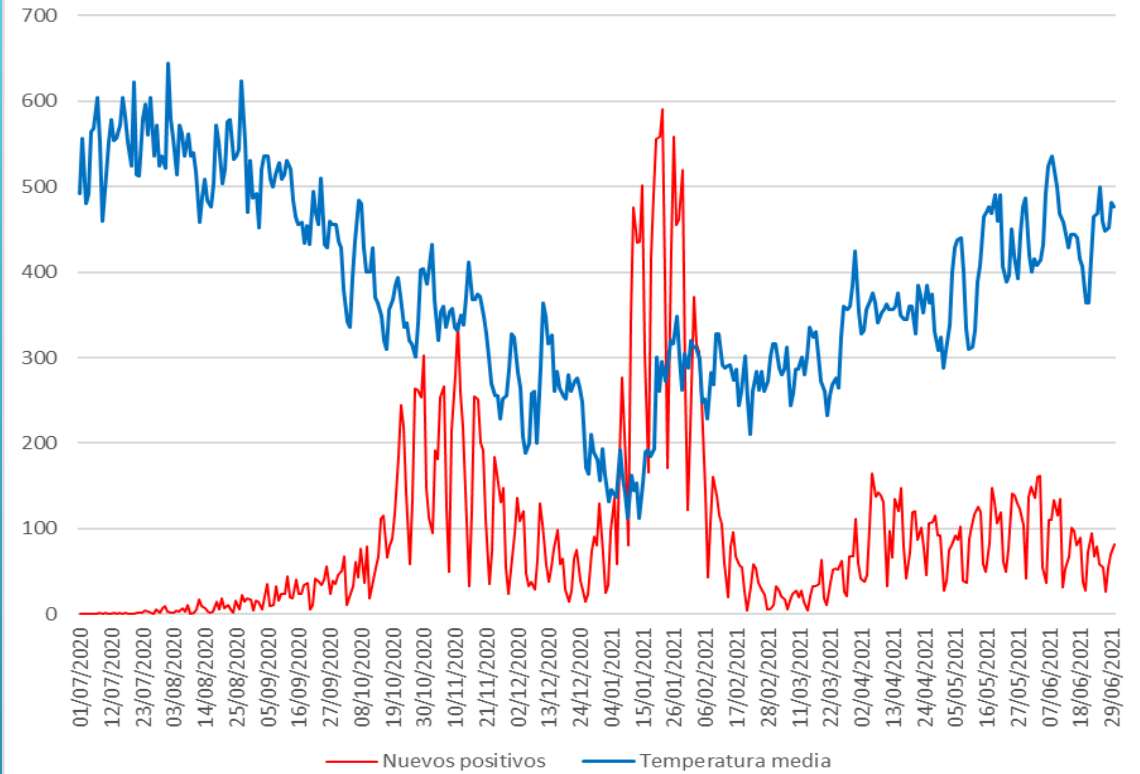
## Casos nuevos Covid Huelva



## Temperatura media



## Huelva -> Comparativa Temperatura - Casos Covid



### TEMPERATURA Y CLIMA

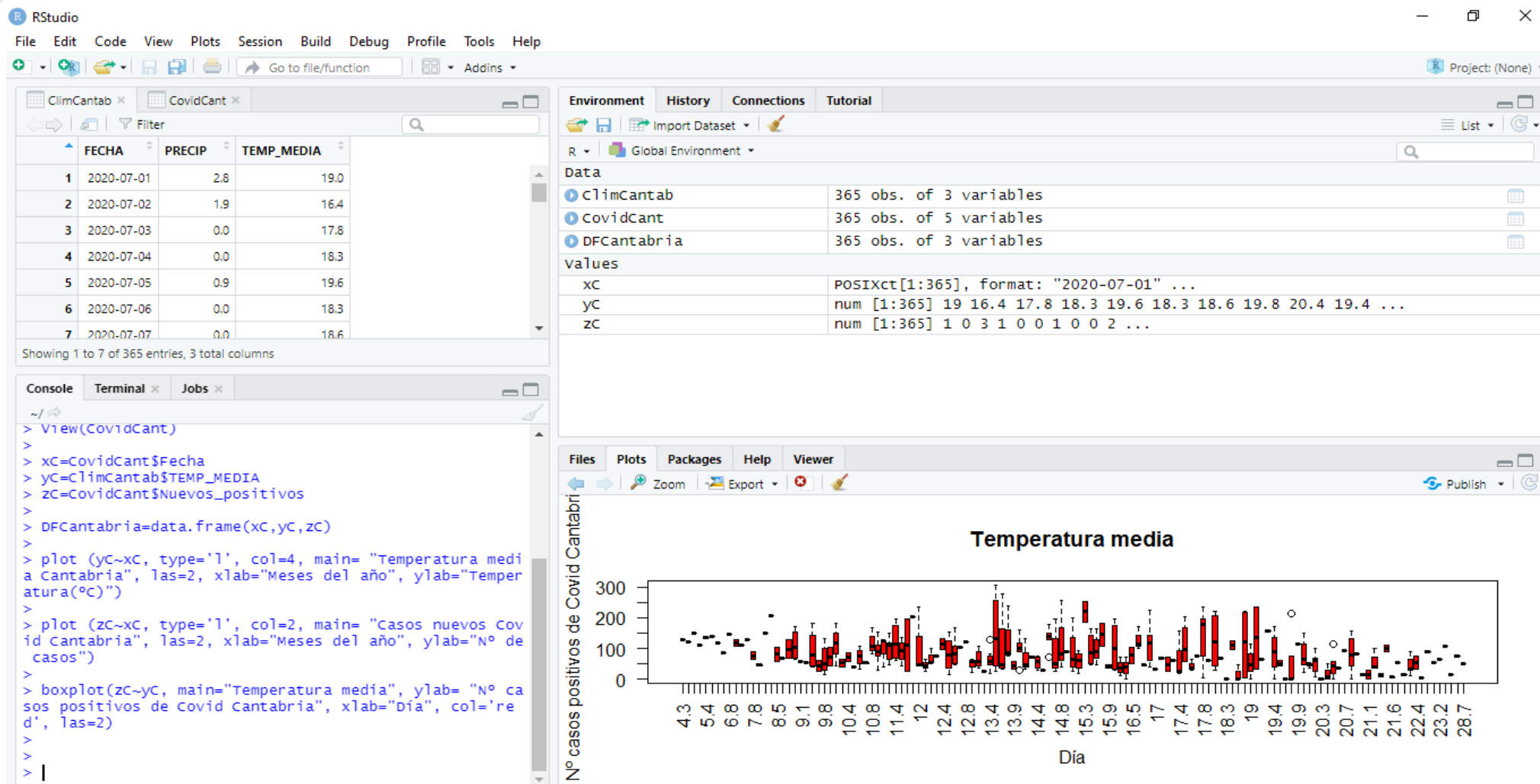
- ▶ Clima Mediterráneo Costero
- ▶ No hay tantas variaciones entre máximas y mínimas
- ▶ Veranos no demasiado calurosos por la influencia del mar.
- ▶ Inviernos suaves

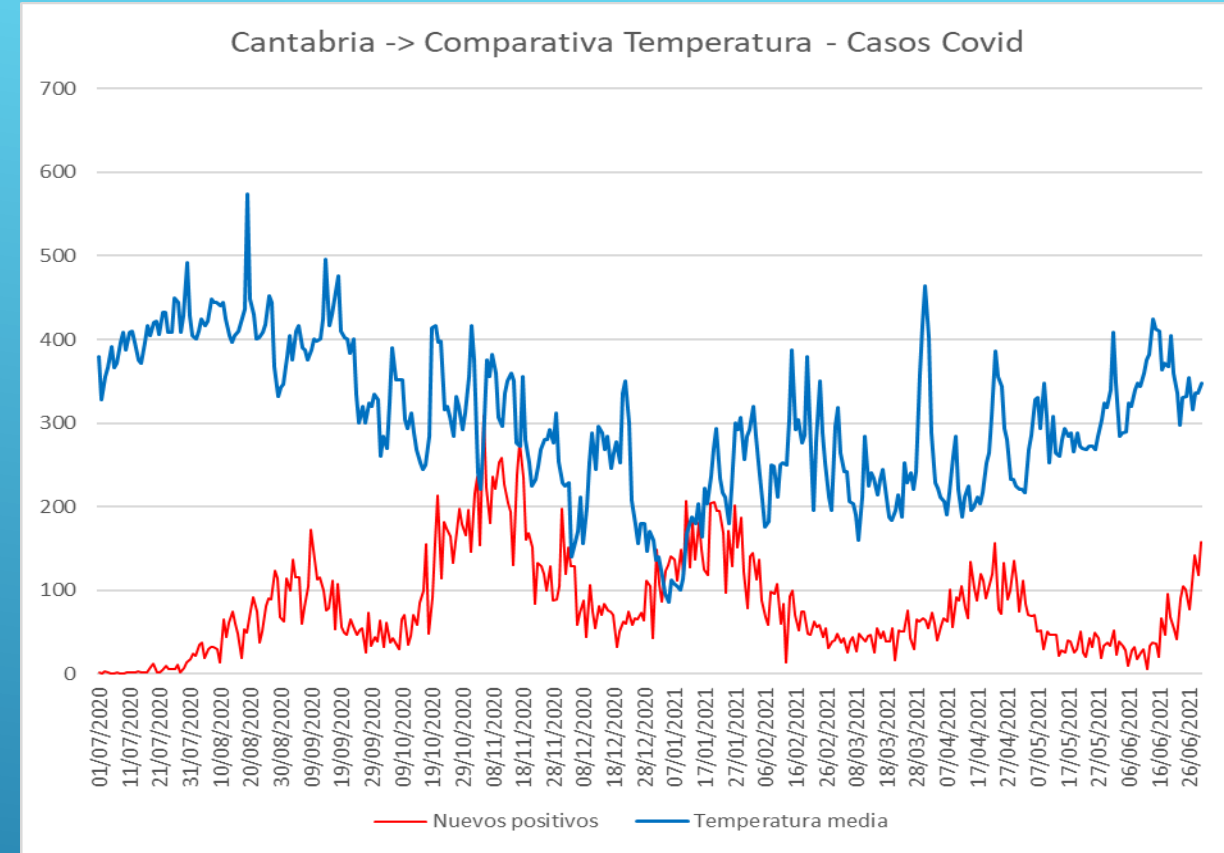
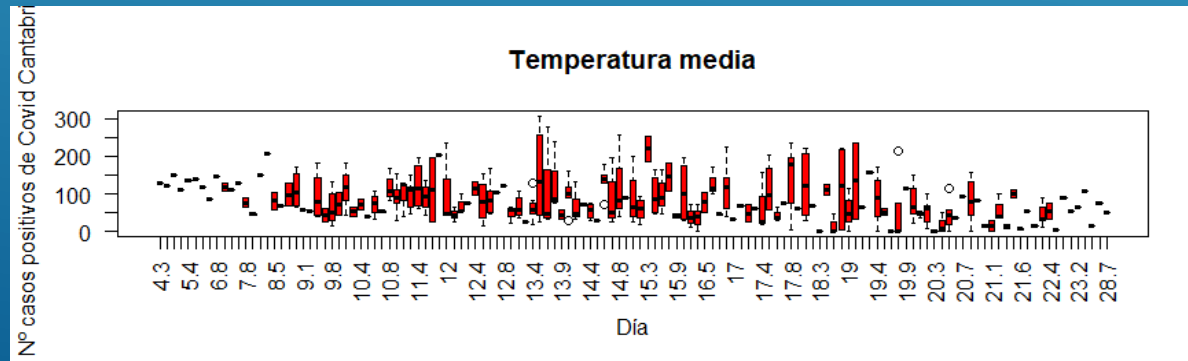
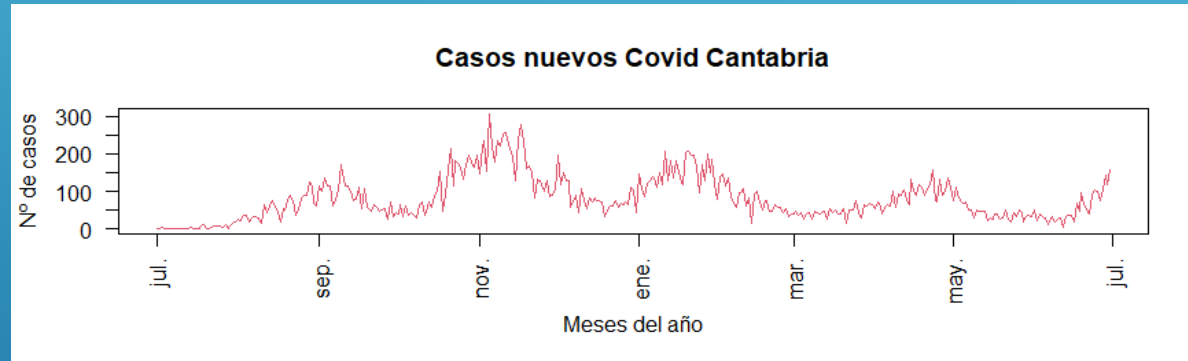
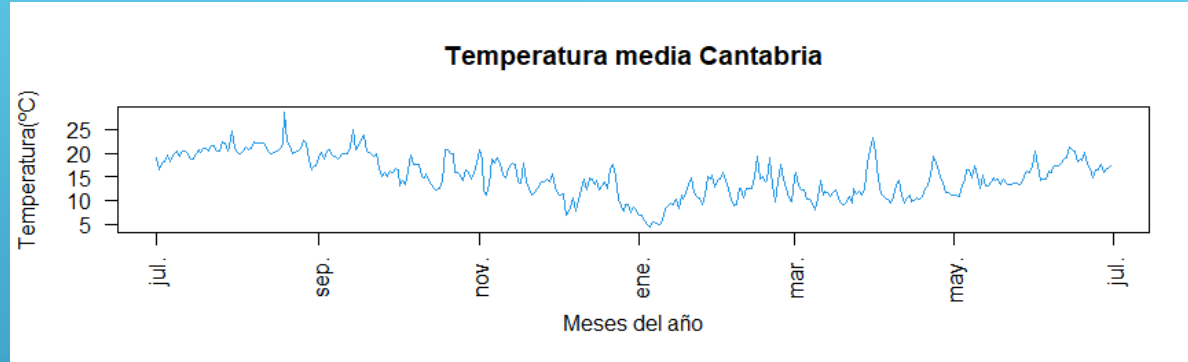
### DATOS COVID

- ▶ Tres episodios, dos de alta influencia en noviembre de 2020 y enero de 2021, así como uno de baja incidencia en el mes de mayo de 2021.

### Observaciones

- ▶ De julio a septiembre de 2020 y entre febrero y junio de 2021, se observa que con las temperaturas altas, la incidencia de Covid se mantiene en niveles bajos.
- ▶ Los días con temperatura media por debajo de los 17°C tienen una mayor incidencia de casos.







## TEMPERATURA Y CLIMA

- ▶ Clima Oceánico
- ▶ Temperaturas suaves por la cercanía al mar y mayor humedad.
- ▶ En invierno la media se sitúa entre los 12 a 15°C y en verano entre los 20-25°C.

## DATOS COVID

- ▶ Cuatro episodios, en septiembre y noviembre de 2020 y en enero y mayo de 2021.
- ▶ El de septiembre y mayo han sido de menor intensidad.

## Observaciones

- ▶ Al tener una temperatura más constante a lo largo del año, y no tener tanta amplitud térmica, se producen episodios de contagio de menor intensidad
- ▶ Los días con temperatura media por debajo de los 20°C tienen una mayor incidencia de casos.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Environment History Connections Tutorial

Import Dataset

R Global Environment

Data

- ClimaLleida 365 obs. of 3 variables
- CovidLleida 365 obs. of 4 variables
- DFLleida 365 obs. of 3 variables

Values

- xLL POSIXct[1:365], format: "2020-07-01" ...
- yLL num [1:365] 27.7 27 23.9 21.6 25.3 27.2 22.6 26.2 28.3 27.8 ...
- zLL num [1:365] 143 118 134 81 76 170 203 229 209 ...

Files Plots Packages Help Viewer

Zoom Export Publish

**Casos nuevos Covid Lleida**

Nº de casos

Meses del año

Jul. sep. nov. ene. mar. may. jul.

Console

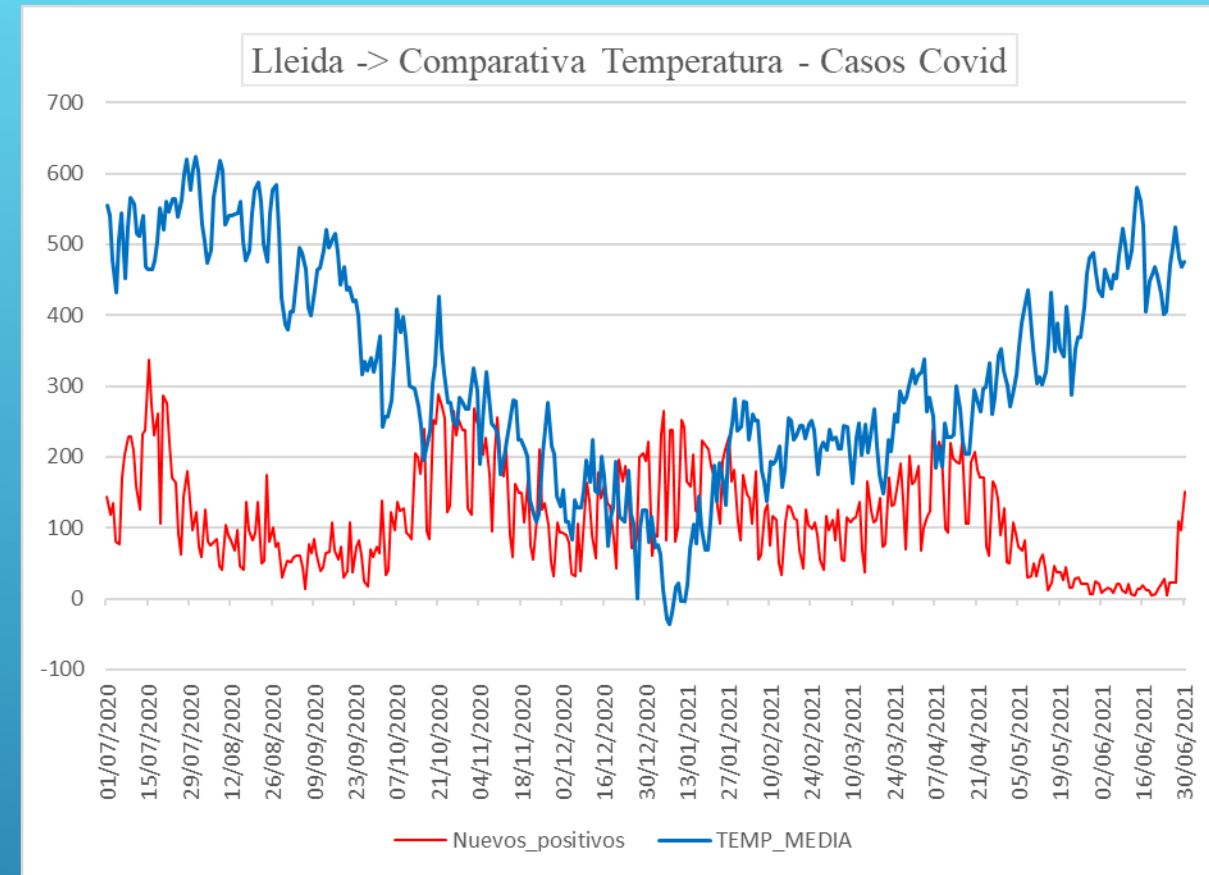
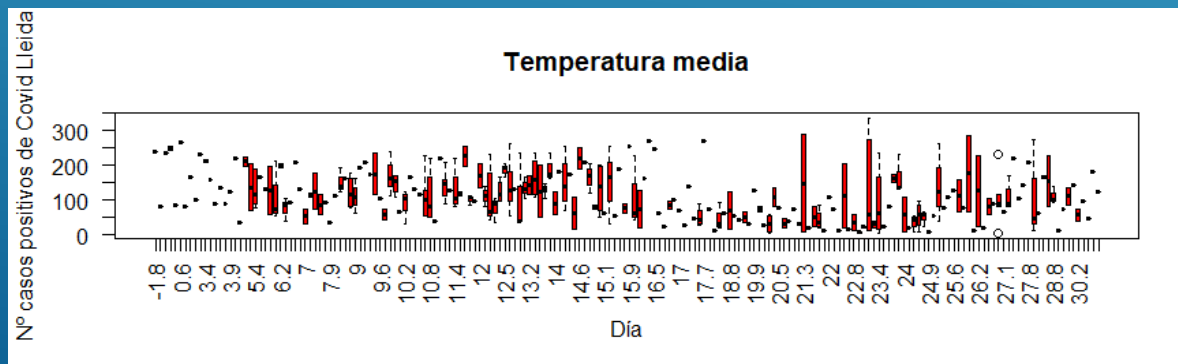
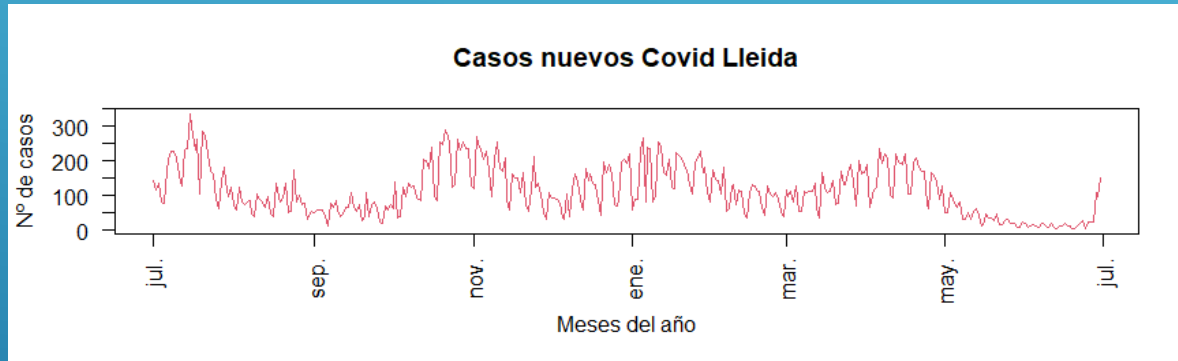
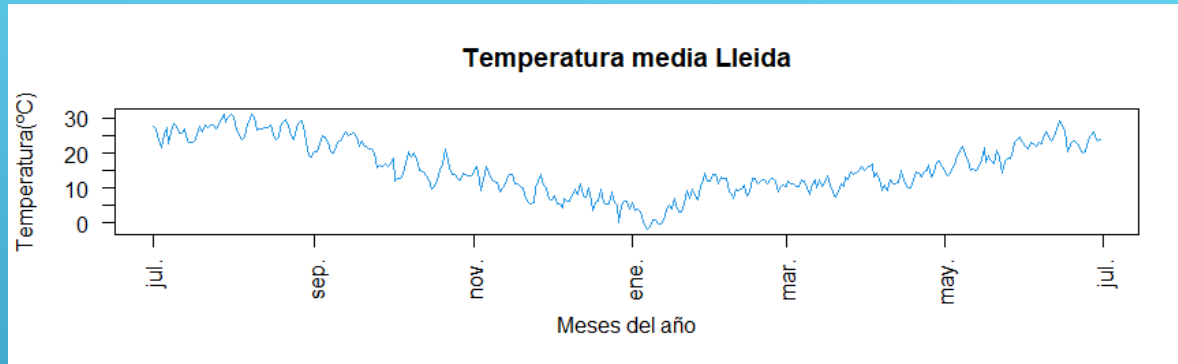
```

>
> yLL=ClimLleida$TEMP_MEDIA
Error: objeto 'ClimLleida' no encontrado
> yLL=ClimaLleida$TEMP_MEDIA
>
> DFLleida=data.frame(xLL,yLL,zLL)
>
>
> plot(yLL~xLL, type='l', col=4, main= "Temperatura me
dia Lleida", las=2, xlab="Meses del año", ylab="Tempera
tura(°C)")
>
>
> plot(zLL~xLL, type='l', col=2, main= "Casos nuevos C
ovid Lleida", las=2, xlab="Meses del año", ylab="Nº de
casos")
>

```

Showing 1 to 7 of 365 entries, 4 total columns

	Fecha	Nuevos_positivos	Ingresos	Fallecimientos
1	2020-07-01	143	1	0
2	2020-07-02	118	0	0
3	2020-07-03	134	5	0
4	2020-07-04	81	6	1
5	2020-07-05	76	5	0
6	2020-07-06	170	5	0
7	2020-07-07	203	9	1



## TEMPERATURA Y CLIMA

- ▶ Clima de Montaña (norte), Clima Oceánico y Clima Mediterráneo Interior
- ▶ Las temperaturas no son similares en toda la provincia.

## DATOS COVID

- ▶ Cuatro episodios, en julio y noviembre de 2020 y en enero y abril de 2021.

## Observaciones

- ▶ La incidencia de nuevos casos disminuye en los meses en los que las temperaturas son más altas
- ▶ En la incidencia de casos según la temperatura media del día no existe un patrón claro, debido a los varios climas que afectan a la provincia.



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Environment History Connections Tutorial

Import Dataset Global Environment

Data

Variable	Observations	Variables
ClimaCR	365 obs.	3 variables
CovidCR	365 obs.	2 variables
DFCR	365 obs.	3 variables

Values

Variable	Values
xCR	POSIXct[1:365], format: "2020-07-01" ...
yCR	num [1:365] 29.4 27.6 26.6 27.4 29 29.6 30.9 29 27.4 27.4 ...
zCR	num [1:365] 23 18 35 12 12 13 22 14 30 20 ...

Console

```

~/CIENCE/10 - PROYECTO FIN DE MASTER/_Datos Ciudad Real/ClimaCR.xlsx
> view(ClimaCR)
> library(readxl)
> CovidCR <- read_excel("C:/_BEATRIZ/00.- MASTER DATA S
CIENCE/10 - PROYECTO FIN DE MASTER/_Datos Ciudad Real/CovidCR.xlsx")
> view(CovidCR)
> xCR=CovidCR$Fecha
> yCR=ClimaCR$TEMP_MEDIA
> zCR=CovidCR$Nuevos_positivos
> DFCR=data.frame(xCR,yCR,zCR)
> plot (yCR~xCR, type='l', col=4, main= "Temperatura me
dia Ciudad Real", las=2, xlab="Meses del año", ylab="Te
mperatura(°C)")
> plot (zCR~xCR, type='l', col=2, main= "Casos nuevos C
ovid Ciudad Real", las=2, xlab="Meses del año", ylab="N
º de casos")
> boxplot(zCR~yCR, main="Temperatura media", ylab= "Nº
casos positivos de Covid Ciudad Real", xlab="Día", col
='red', las=2)
>

```

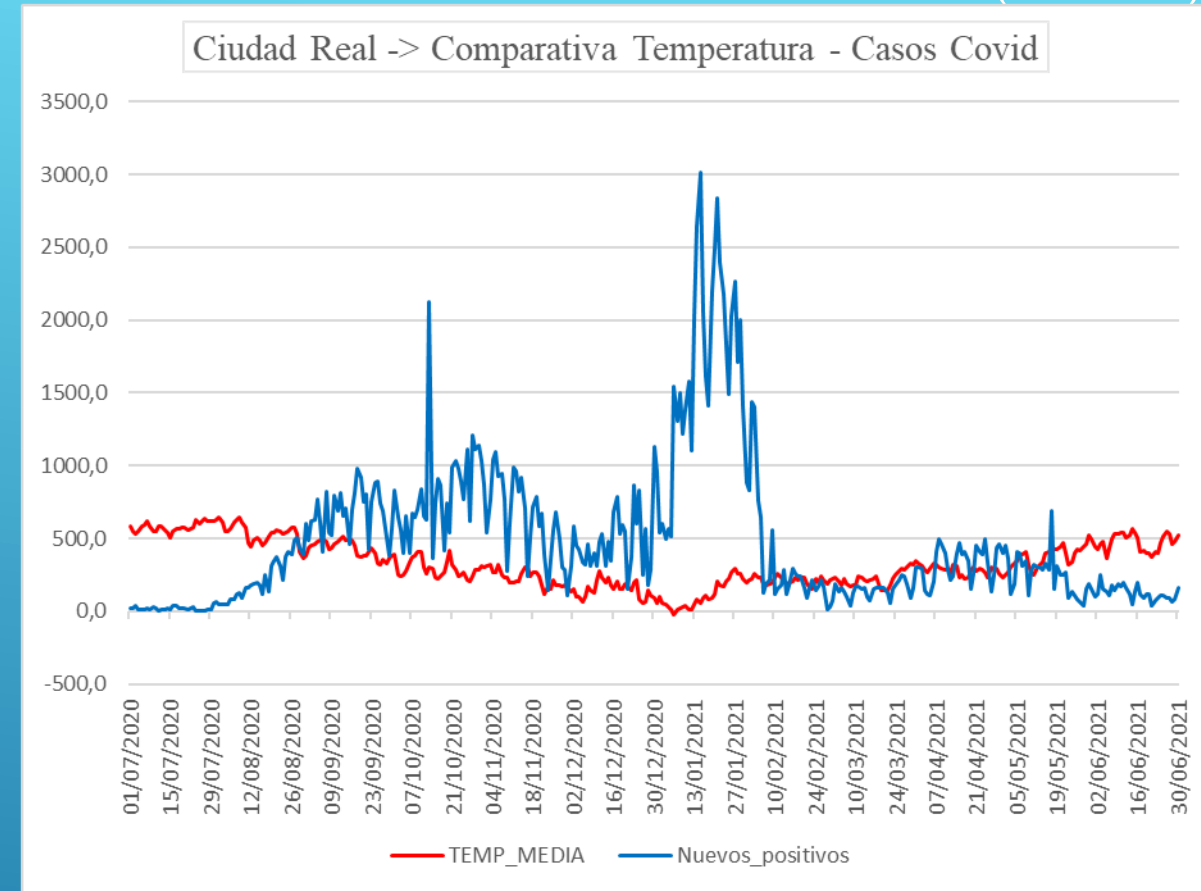
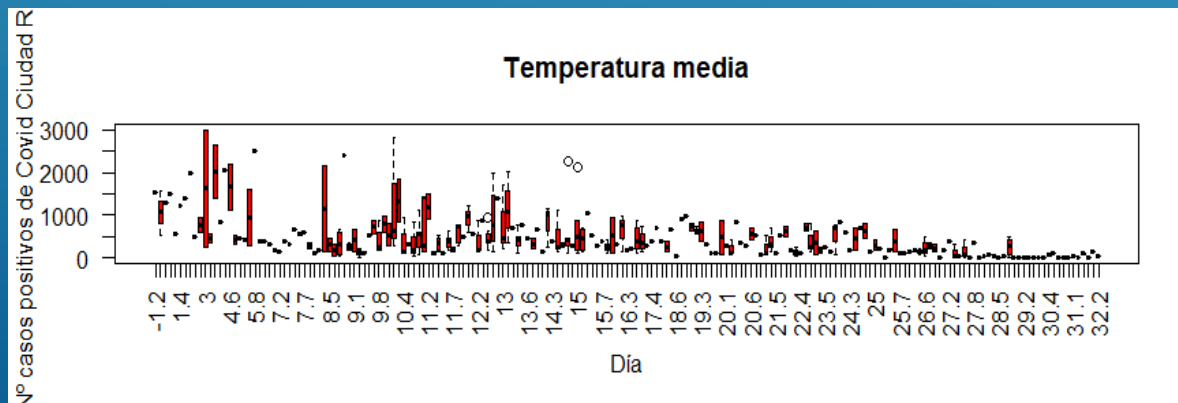
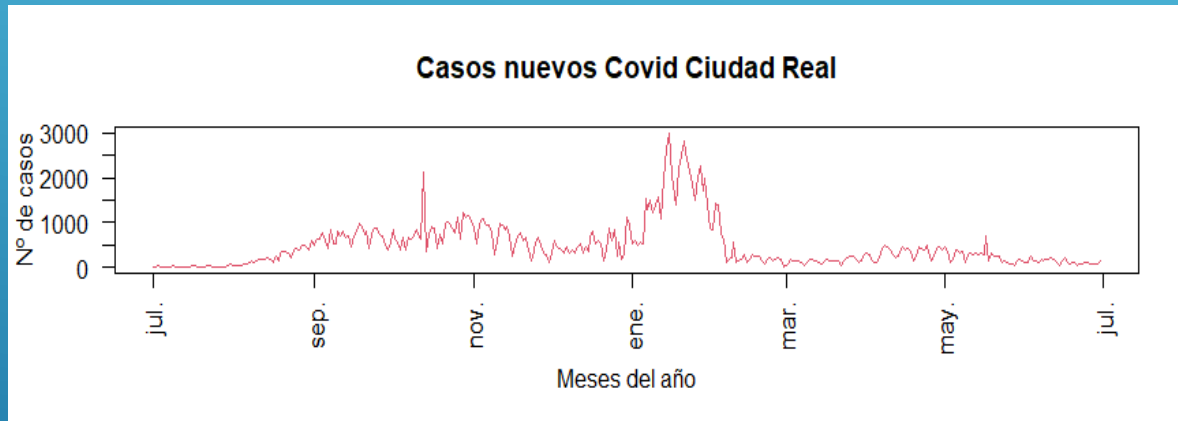
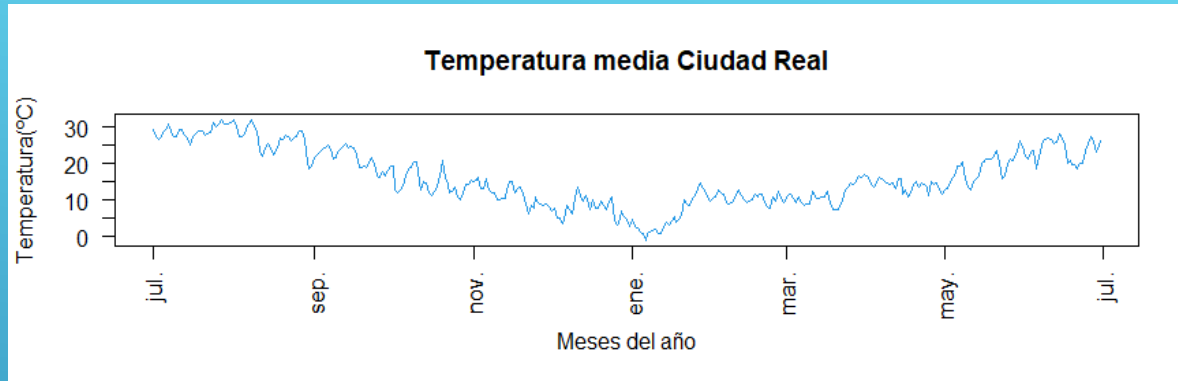
Files Plots Packages Help Viewer

Zoom Export Publish

Temperatura media

Nº casos positivos de Covid Ciudad Real

Día



## TEMPERATURA Y CLIMA

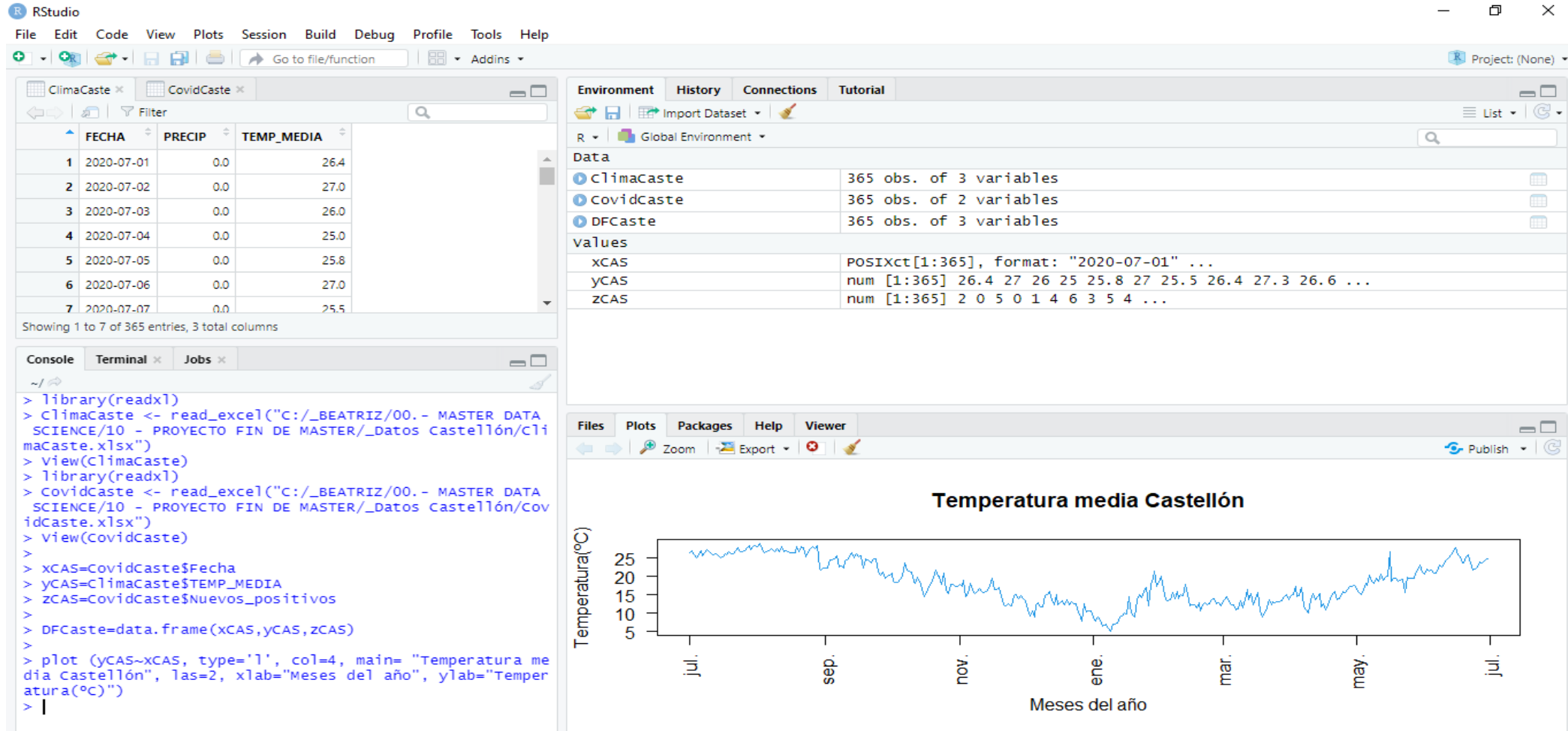
- ▶ Clima Mediterráneo de Interior.
- ▶ Los veranos son frescos y los inviernos bastante fríos.

## DATOS COVID

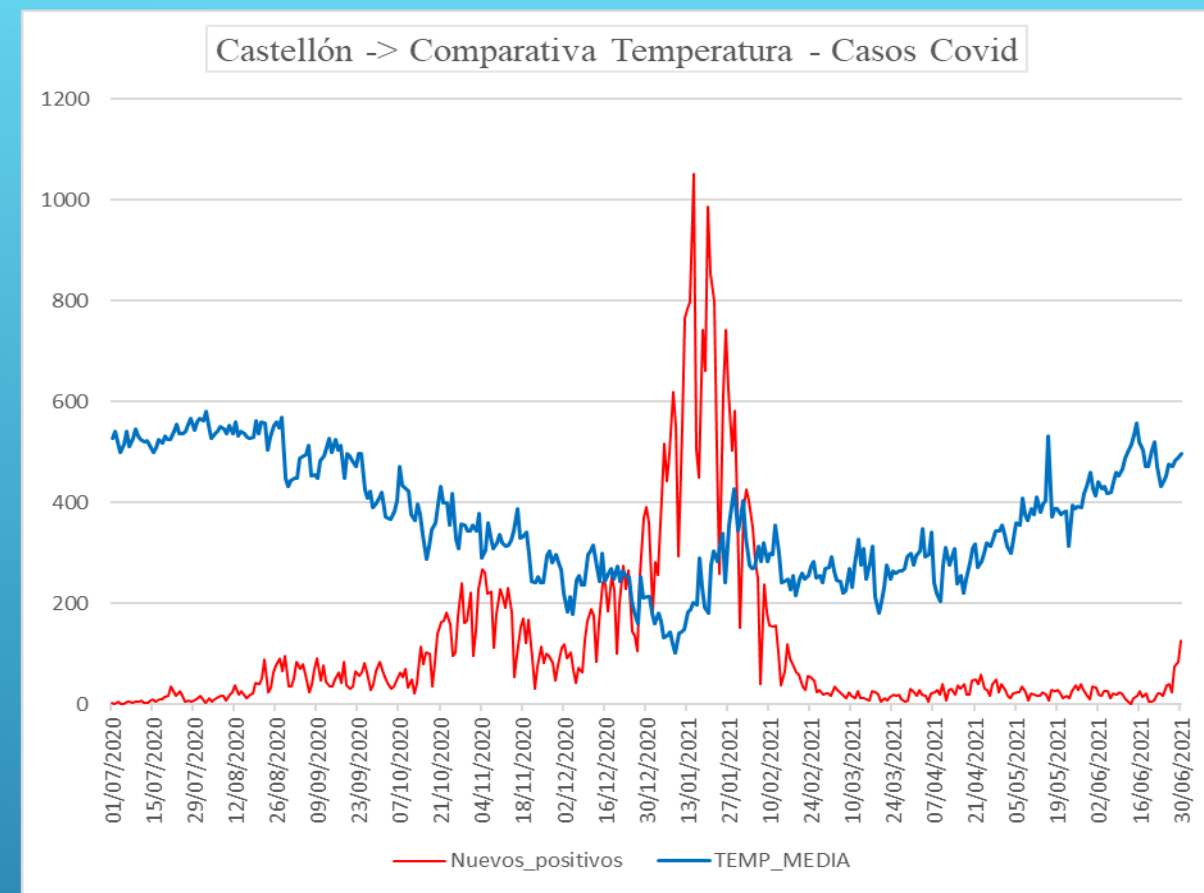
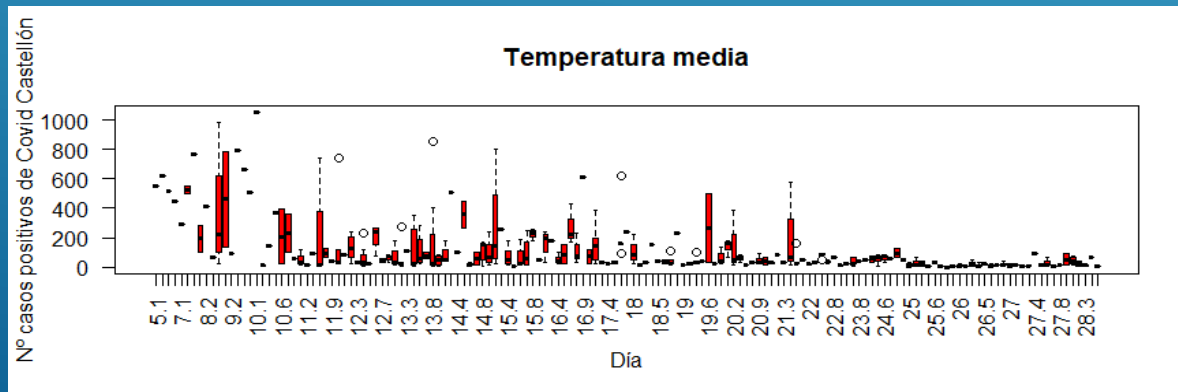
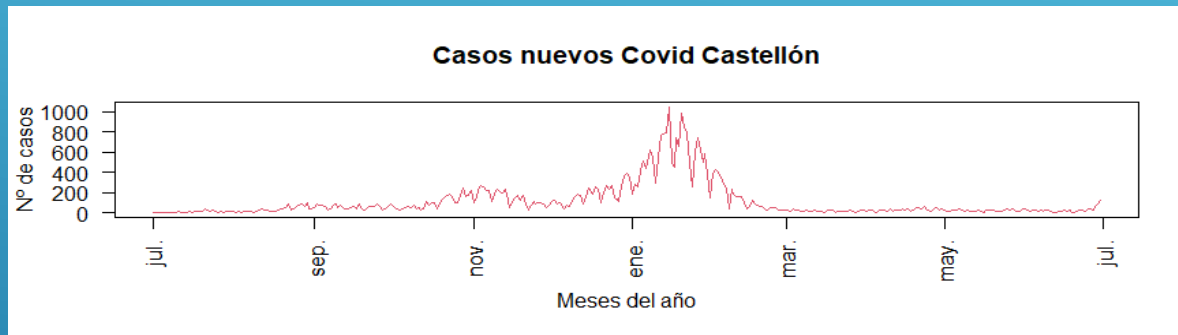
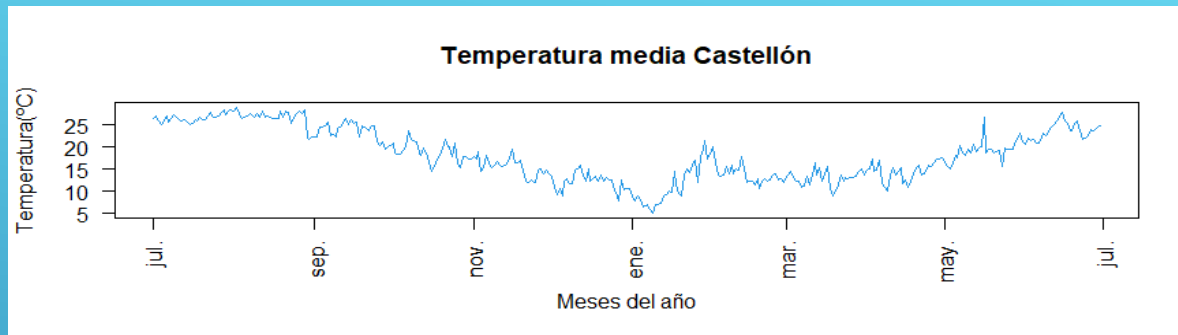
- ▶ Pequeño episodio en incidencia aunque prolongado en el tiempo, entre los meses de Septiembre a Diciembre de 2020.
- ▶ Episodio de gran intensidad en Enero y Febrero de 2021.

## Observaciones

- ▶ La incidencia disminuye con la temperatura en los meses de máximo calor como son junio, julio y agosto.
- ▶ Los días con temperatura media por debajo de los 13°C tienen una mayor incidencia de casos.







## TEMPERATURA Y CLIMA

- ▶ Clima Mediterráneo de Costa.
- ▶ Reducidas variaciones entre máximas y mínimas.
- ▶ Los veranos no son excesivamente calurosos, y los inviernos no son extremos.

## DATOS COVID

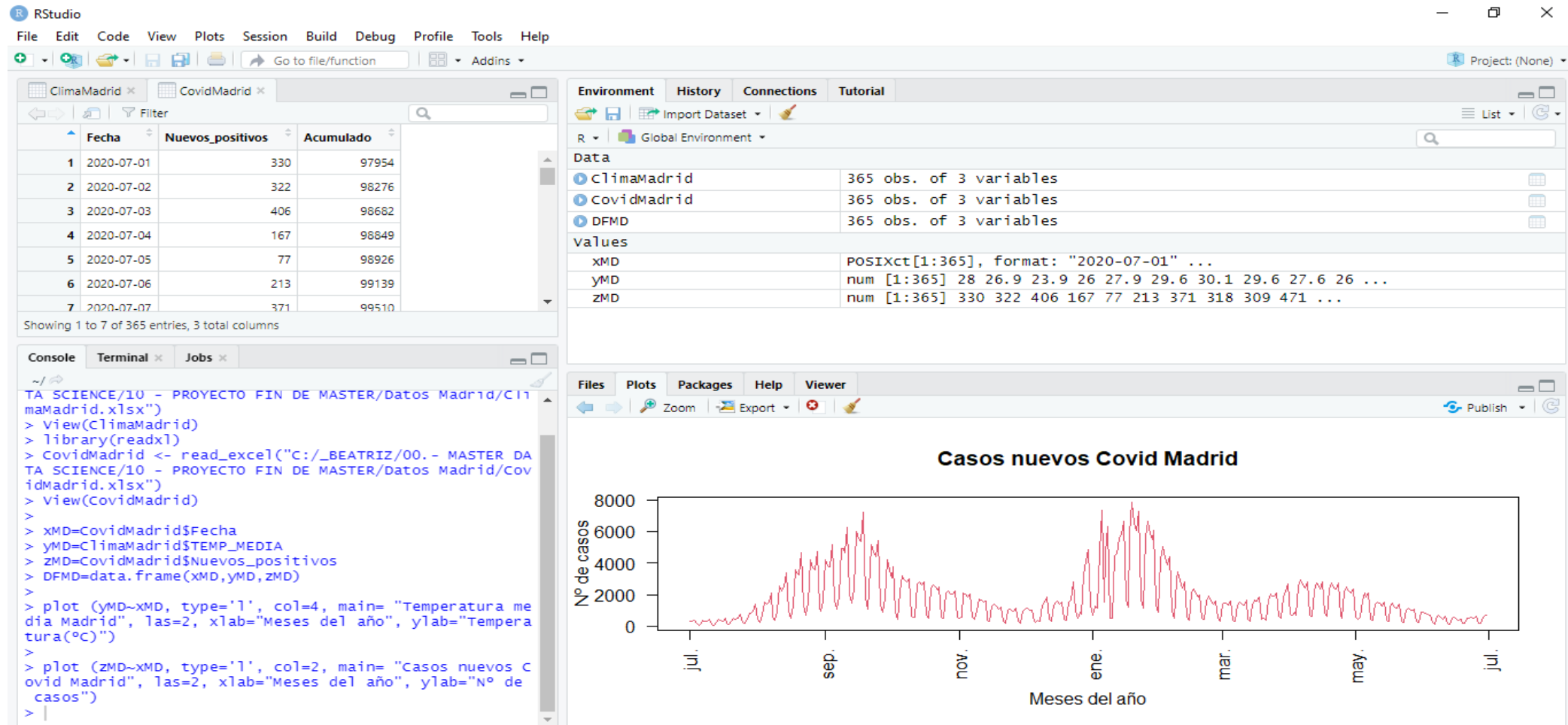
- ▶ Dos “olas”, una de baja intensidad en noviembre de 2020.
- ▶ Y otra “ola” de gran incidencia en enero de 2021.

## Observaciones

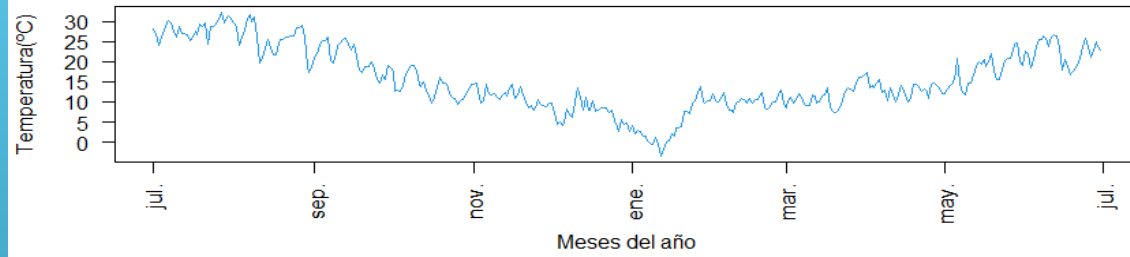
- ▶ Solo existe relación directa entre la temperatura media y los nuevos casos de covid, desde mediados de octubre de 2020 hasta mediados del mes de febrero de 2021.
- ▶ Los días con temperatura por debajo de los 17°C de media tienen una mayor incidencia de casos.

# COMPROBACIÓN DEL MODELO

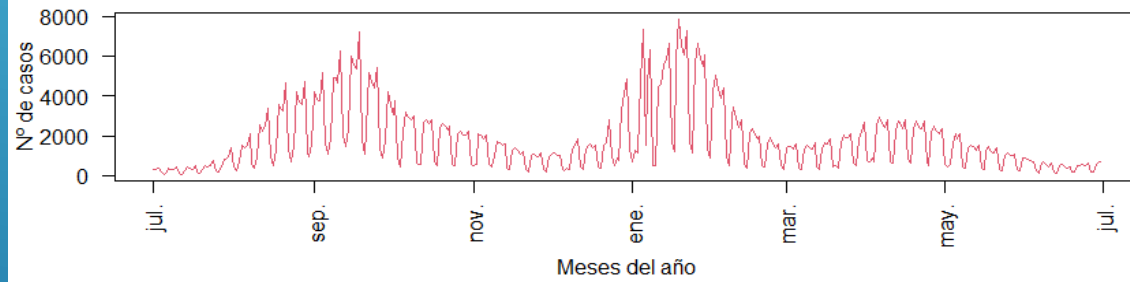
- Comprobamos el modelo diseñado para el estudio en Madrid y Barcelona:
  - Muy pobladas
  - Ciudades grandes
  - Características climáticas diferentes



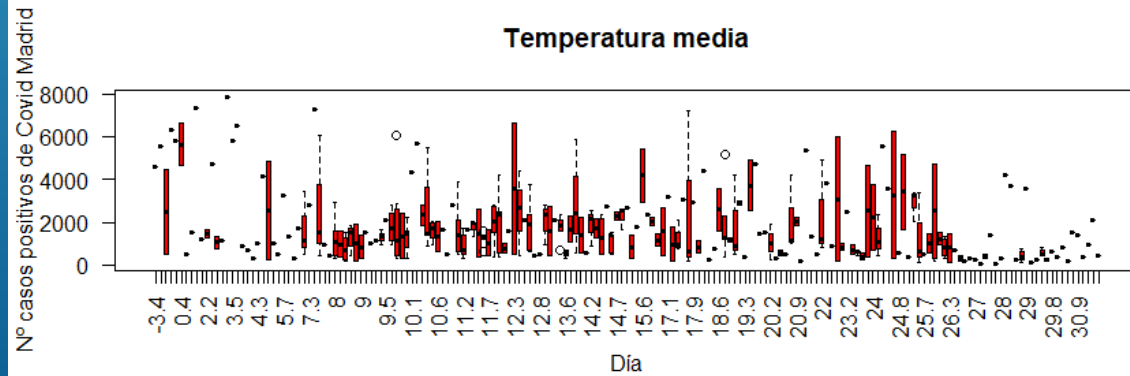
Temperatura media Madrid



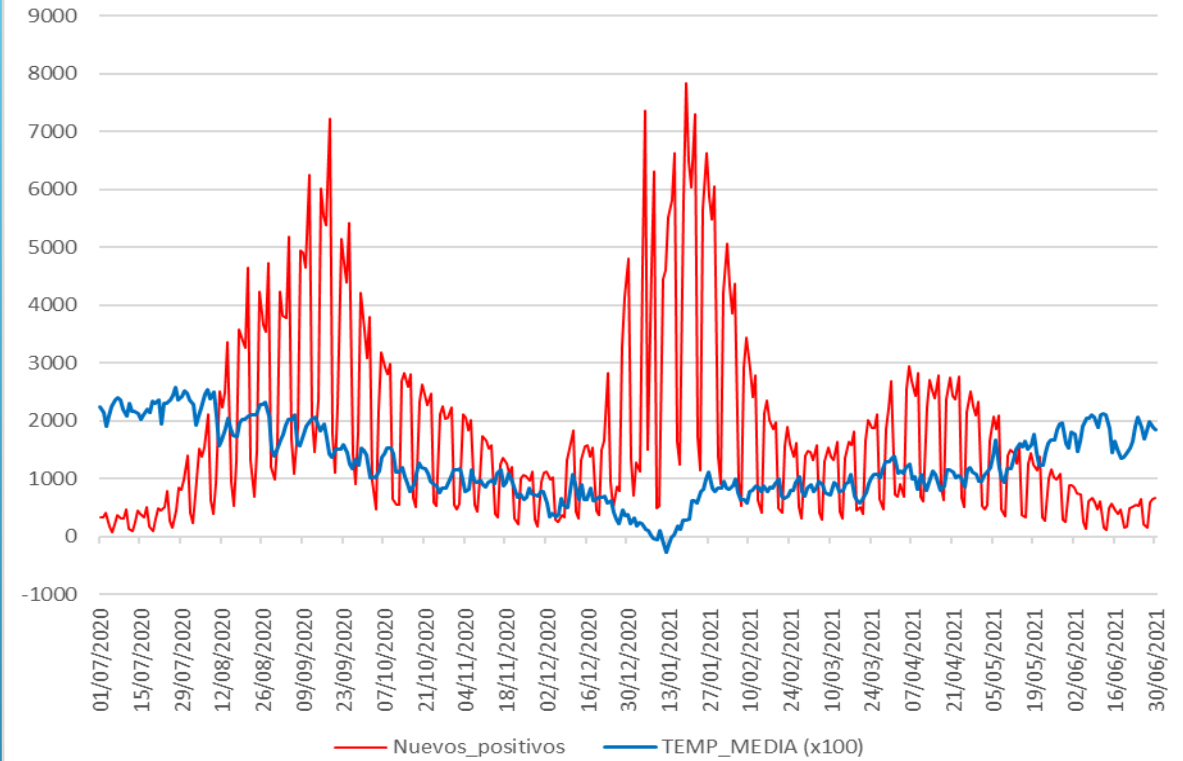
Casos nuevos Covid Madrid



Temperatura media



Madrid -&gt; Comparativa Temperatura - Casos Covid



## TEMPERATURA Y CLIMA

- ▶ Clima Mediterráneo Interior.
- ▶ Inviernos fríos y veranos con temperaturas altas.
- ▶ Amplitud térmica muy amplia, con temperaturas extremas tanto en verano como en invierno.

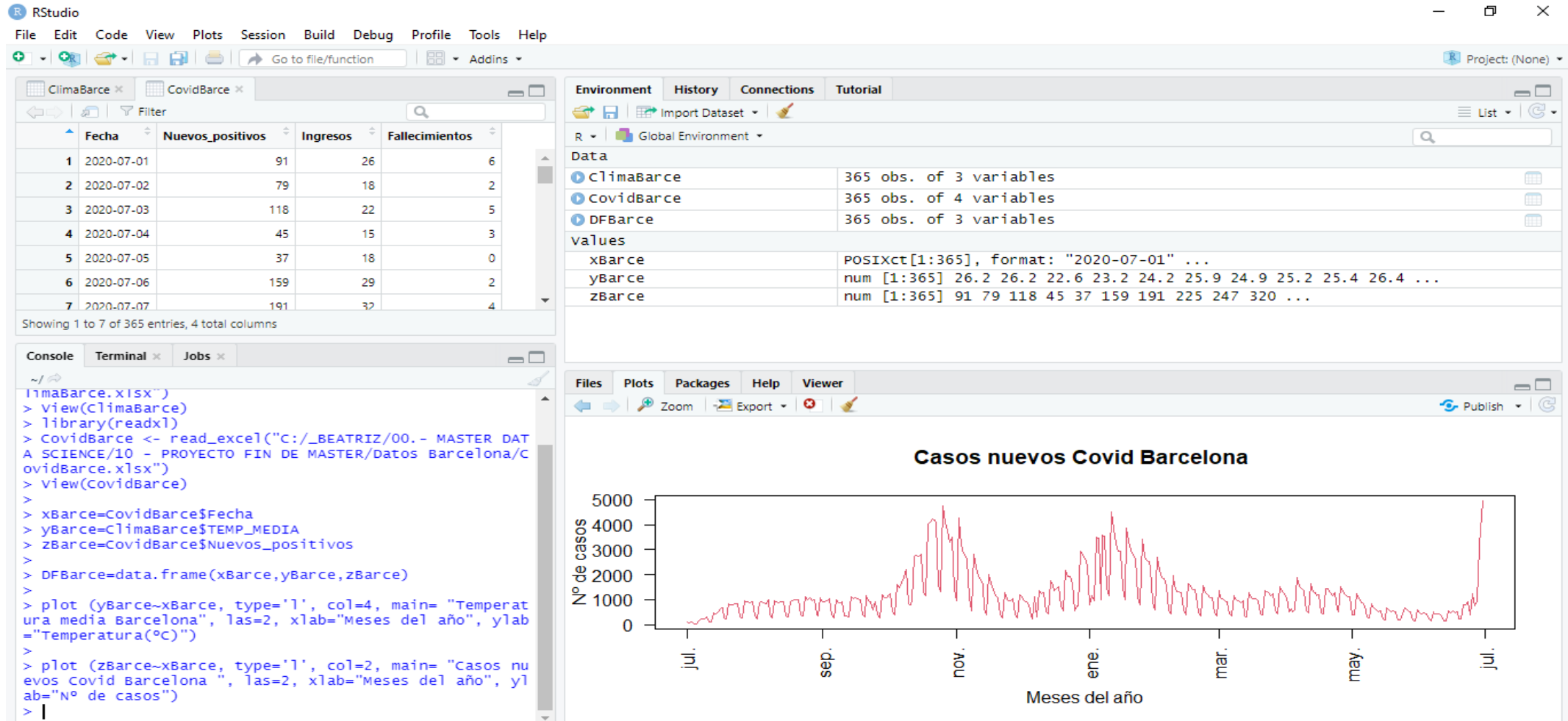
## DATOS COVID

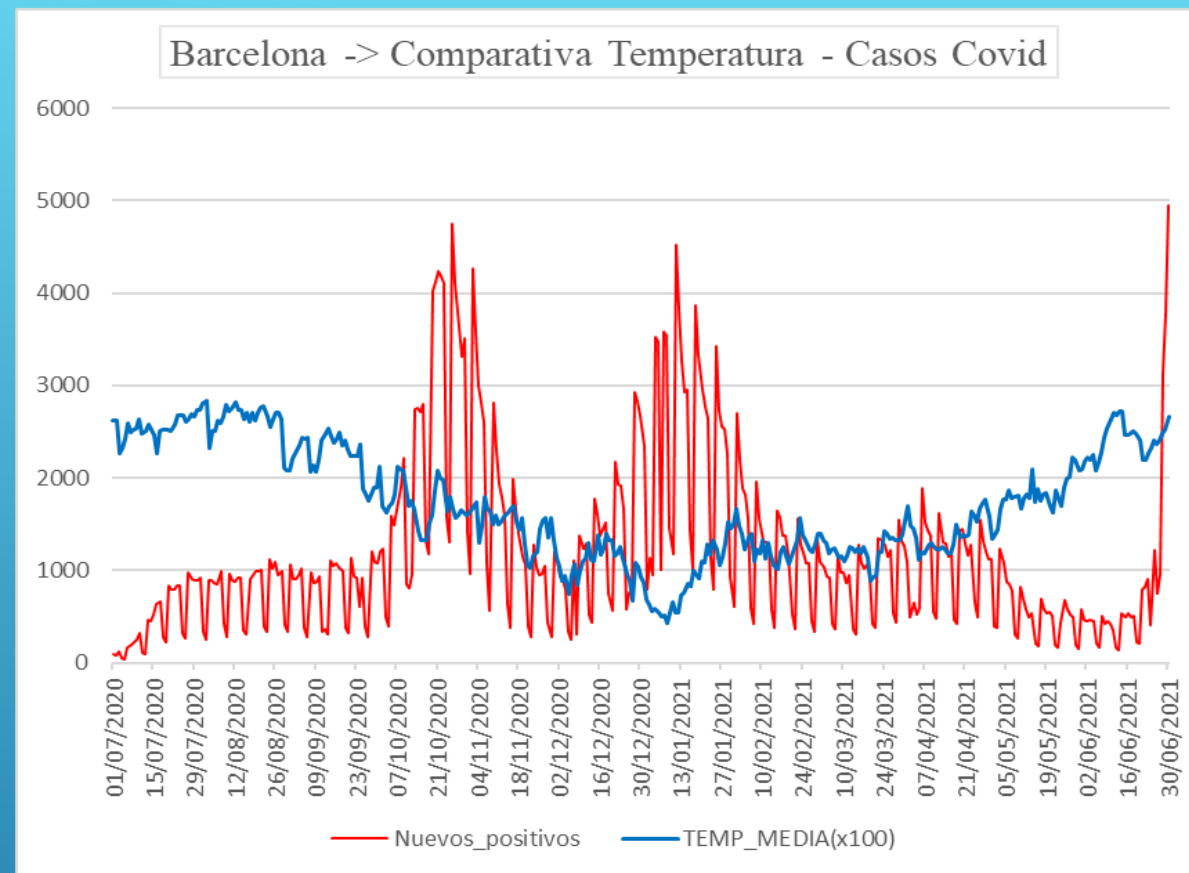
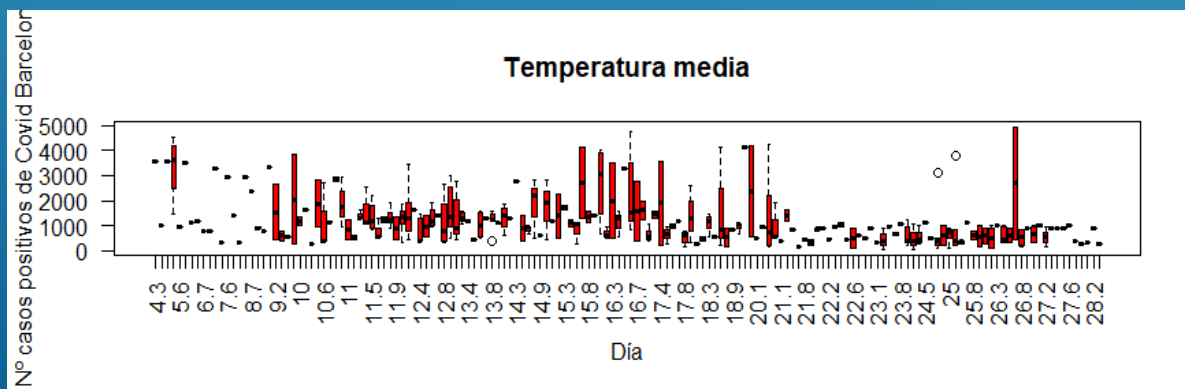
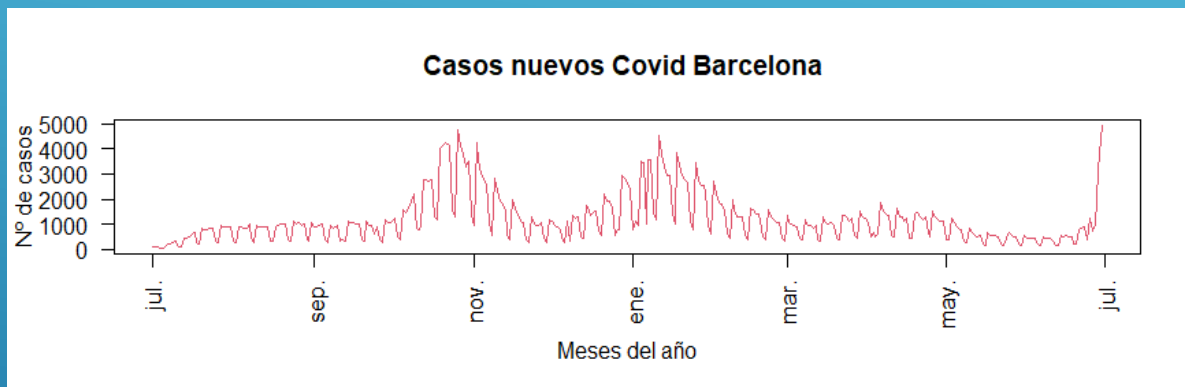
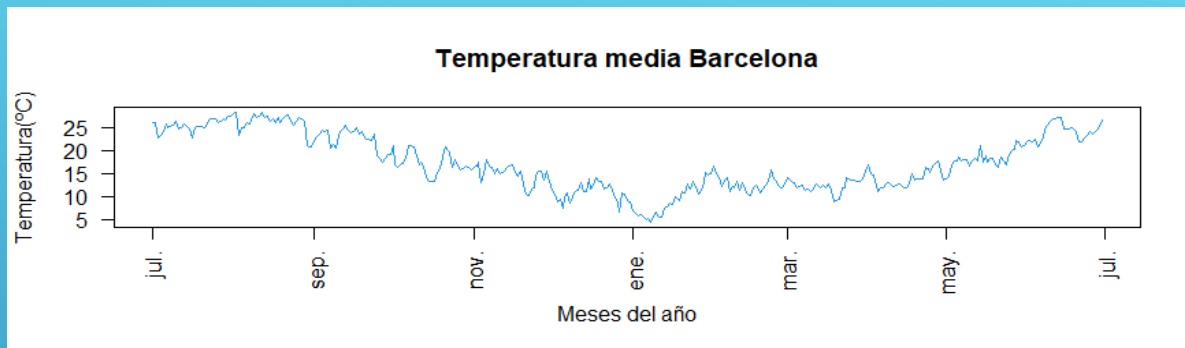
- ▶ Dos episodios con alta incidencia, en septiembre de 2020 y en febrero de 2021.
- ▶ Y otro episodio de baja incidencia en abril de 2021.

## Observaciones

- ▶ En los meses de temperaturas mas altas como son junio y julio se puede concluir que la elevada temperatura media hace que la incidencia de Covid disminuya.
- ▶ Los días con temperatura media por debajo de los 26°C tienen una mayor incidencia de casos.







## TEMPERATURA Y CLIMA

- ▶ Clima Mediterráneo Costero.
- ▶ Amplitud térmica pequeña.
- ▶ Pocos días al año la temperatura baja de los 10°C, ni sube de 25°C.

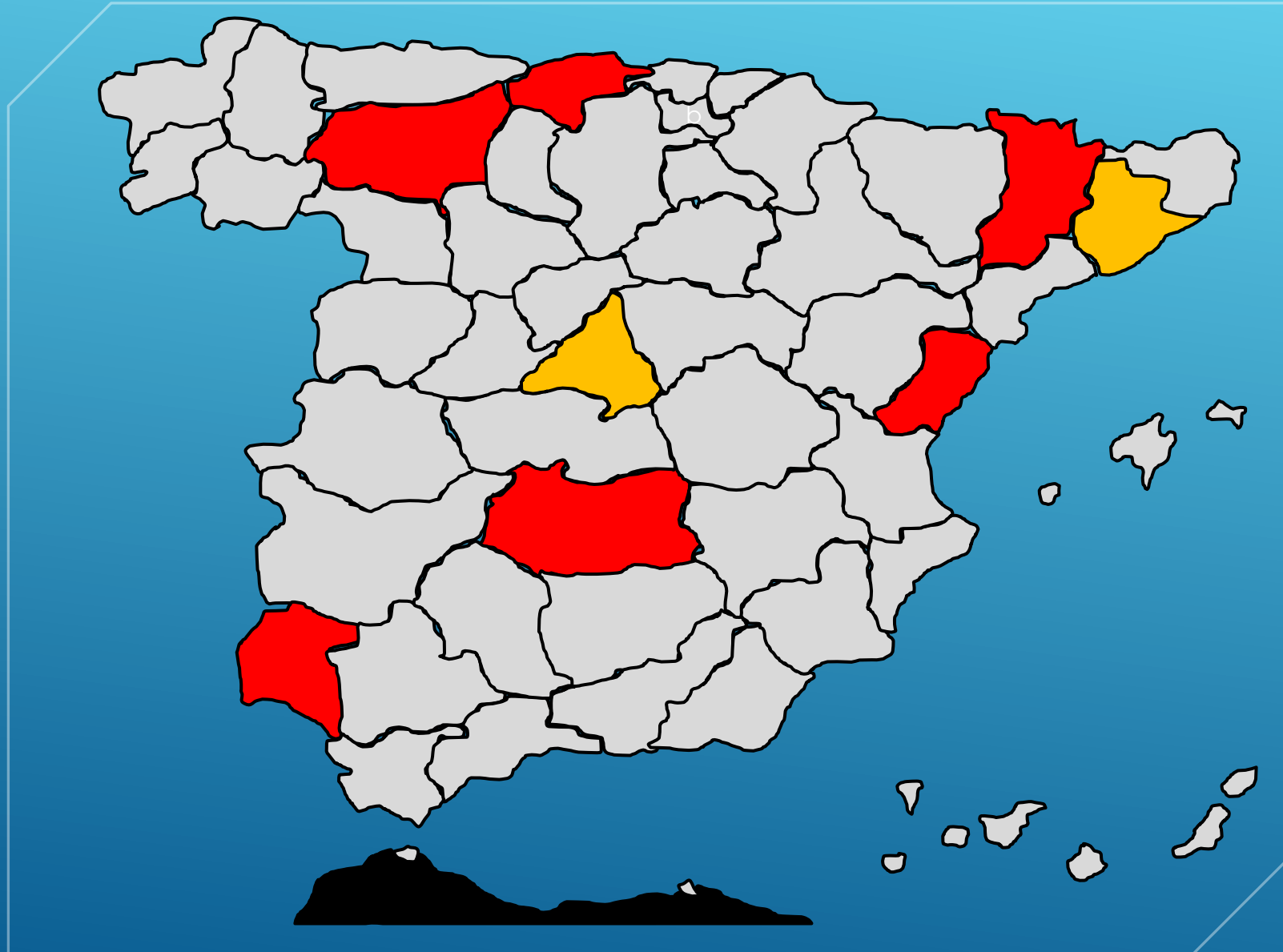
## DATOS COVID

- ▶ Dos episodios de contagio masivo en noviembre de 2020 y enero de 2021.
- ▶ Un episodio casi sin consideración en abril de 2021.

## Observaciones

- ▶ En los meses de temperaturas mas altas, que son de julio a septiembre de 2020 y de mayo a junio de 2021 la incidencia de covid se mantiene en niveles bajos.
- ▶ Los días con temperatura media por encima de los 20°C desciende mucho la probabilidad de contagio.

# CONCLUSIONES

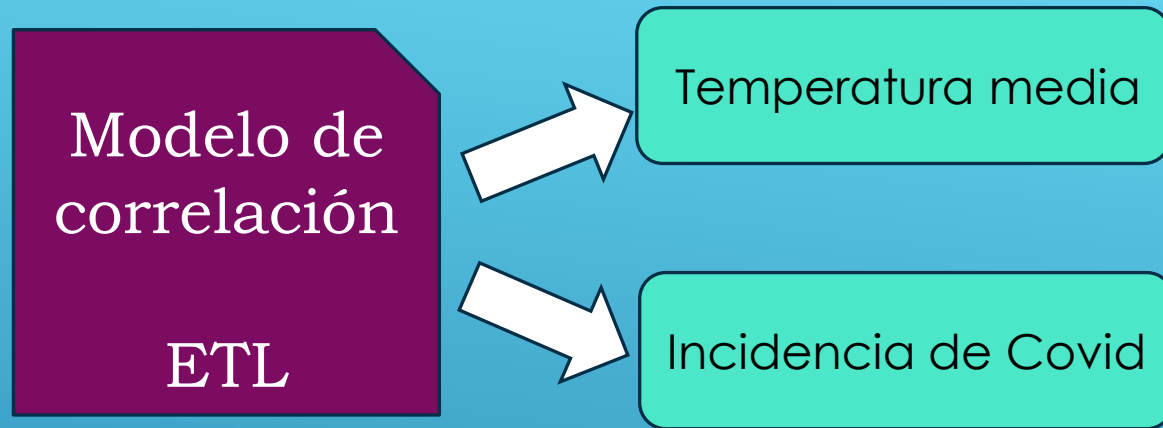


Provincias de  
estudio

Provincias de  
comprobación.

# RESUMEN DE LOS DATOS OBTENIDOS

PROVINCIA	Población provincial	Tipo de Clima	Media temp. Máx.	Media temp. Mín.	Amplitud térmica	Temp. umbral del estudio	Días por encima de la temperatura umbral
LEÓN	455.066	Med. interior	22,2	-0,7	22,9	12	163
HUELVA	530.386	Med. costero	28,4	7,4	21	17	217
CANTABRIA	583.904	Oceánico	21,5	3,8	17,7	20	70
LLEIDA	436.911	Oceánico + Med. interior + Montaña	28,3	0,3	28	- -	
CIUDAD REAL	492.065	Med. interior	30,2	1,2	29	13	215
CASTELLÓN	576.941	Med. costero	26,7	6,2	20.5	17	193
MADRID	6.752.763	Med. interior	29,8	1,9	27,9	26	51
BARCELONA	5.638.094	Med. costero	27,8	4,5	23,3	21	129



**TEMPERATURA**



**INDICENCIA DE  
NUEVOS CASOS**

- ✓ **LEON Y CIUDAD REAL:** Mayor amplitud térmica.
- ✓ **HUELVA, CASTELLÓN Y BARCELONA:** Días con temperatura media por encima de la temperatura umbral muy elevado.
- ✓ **CANTABRIA:** Su temperatura media anual se mantiene más constante y es más difícil determinar la influencia de la temperatura en el número de casos de contagio nuevos
- ✓ **LLEIDA:** el gráfico de los contagios en función de la temperatura media del día no nos puede arrojar un valor umbral.
- ✓ **MADRID:** Mucha población, gran interacción y contaminación.



# DESTACAR

## GRADO DE INTERACCIÓN SOCIAL

### INVIERNO

- Menos contactos sociales
- Menor número de personas
- Espacios interiores

### VERANO

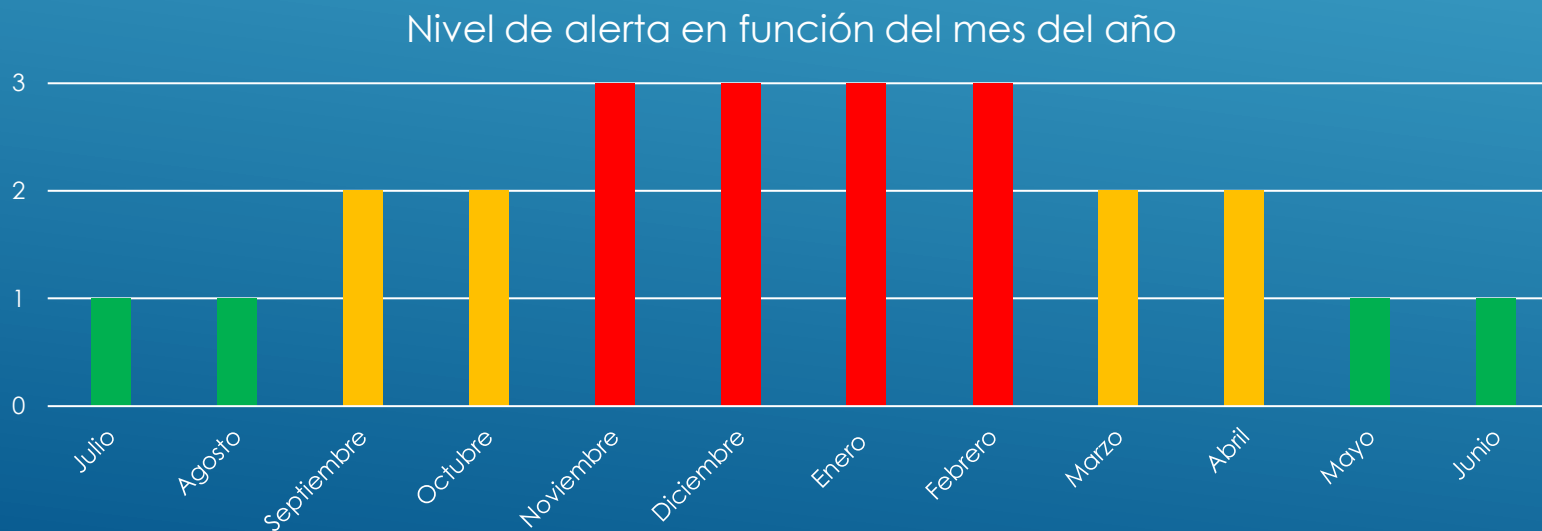
- Reuniones con más frecuencia
- Multitudinarias
- Espacios exteriores y abiertos

Menos  
contagios

# OBJETIVO DEL PROYECTO:

## ELABORAR UNA SERIE DE NORMAS Y RESTRICCIONES

- ♦ En los meses, de noviembre a febrero: se limitará al nivel máximo la interacción social
- ♦ En los meses de septiembre, octubre, marzo y abril: las medidas de control y restricción de movilidad e interacción podrían ser de intensidad media
- ♦ Finalmente en los meses de mayo, junio, julio y agosto: se puede permitir más interacción social, sin limitación de horario, y se promoverán las actuaciones al aire libre.



# PLANIFICACIÓN

## FASES DEL PROYECTO


- Fase inicial
  - Fase de análisis
  - Fase de desarrollo (Proceso ETL)
  - Fase de redacción de la memoria
- 
- Several white lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

DIAGRAMA DE GANTT			JULIO				AGOSTO			
Descripción	Fecha inicio	Días	1ª s	2ª s	3ª s	4ª s	1ª s	2ª s	3ª s	4ª s
Fase Inicial										
Investigación preliminar	01/07/2021	5								
Estudio inicial	08/07/2021	5								
Fase de Análisis										
Especificación de requisitos	08/07/2021	5								
Fase de Desarrollo (proceso ETL)										
Extracción	15/07/2021	7								
Transformación	26/07/2021	10								
Carga	09/08/2021	10								
Fase de redacción de la Memoria										
Desarrollo de la memoria	19/07/2021	30								

# PRESUPUESTO

- **Recursos humanos**

2.284,00 €

- **Recursos materiales**

00,00 €

- **Costes indirectos**

274,08 €

- **Beneficio**

409,29 €

- **IVA**

623,14 €

**COSTE TOTAL DEL PROYECTO**



## Personal interviniente

- Director del proyecto:

26 euros / hora

Horas asignadas: 14 horas

- Data Scientist:

16 euros / hora

Horas asignadas: 120 horas

**3.590,51 €**

# MEJORAS FUTURAS DEL PROYECTO

- Estudio de más provincias, e incluso regiones dentro de la misma provincia
- Estudiar otras variables climáticas como:
  - Humedad de la región
  - Cantidad de precipitación caída
  - Horas de sol diarias
  - Velocidad del viento
  - Altitud media de la provincia





GRACIAS POR SU ATENCIÓN

