



MÁSTER DE ESTADÍSTICA APLICADA PARA LA
CIENCIA DE DATOS CON R SOFTWARE
NOMBRE
DEL MÁSTER

**SIMILITUD CLIMATICA DE DIFERENTES
ZONAS PARA EL OPTIMO DESARROLLO
DEL CULTIVO DEL ARANDANO**

AUTOR: Jose Rufo Rodriguez

DIRECTOR: Juan Luis Lopez

FECHA: 01/06/2024

ENTIDAD COLABORADORA



RESUMEN

El cultivo del arándano se encuentra inmerso en una increíble expansión, por ello conocer la adaptabilidad de distintas especies de arándanos a nuevas localizaciones es fundamental.

Un factor importantísimo es conocer como afectaran las condiciones meteorológicas de las nuevas zonas de expansión.

Siguiendo este protocolo de actuación hemos logrado buscar similitudes climáticas entre distintas localizaciones a fin de poder predecir el éxito de la implantación del cultivo en localizaciones nuevas.

Usando una base datos generada a raíz de la recolección de datos histórico y con la ayuda de RStudio, hemos conseguido crear clústeres que relacionan parámetros meteorológicos como temperatura promedio, temperatura max, temperatura min, humedad relativa, punto de rocío e insolación.

De esta forma solo necesitamos buscar los datos climáticos de una zona nueva, incluirlos en nuestra base de datos y con RStudio veremos a que zonas conocidas se asemeja.

De la información conocida de esa zona sabremos que variedades podremos implantar.

AGRADECIMIENTOS

A todo el equipo de Máxima Formación por la ayuda prestada durante el desarrollo del máster y durante la realización del trabajo, en especial a Juan Luis Lopez, profesor del área Data Science con R por su tiempo, paciencia y disponibilidad.

ÍNDICE

Introducción	5
Material y métodos	7
Fuente de datos	7
Preparación de datos	8
Modelos	8
Resultados	9
Análisis descriptivos	9
Análisis de clúster jerárquicos	13
Introducimos la nueva localización	15
Estudio detallado mensual	18
Discusión	30
Conclusiones.	32
Bibliografía.	31
Anexo códigos R	32

INTRODUCCIÓN

El cultivo del arándano ha experimentado un crecimiento significativo en el comercio agrícola internacional en los últimos años, siendo en la actualidad un alimento cada vez más esencial para la salud humana y es cada vez es más importante para la dieta mundial.

Países como EEUU, Perú, Chile, México y Europa, cuentan con regiones que son importantes productores y exportadores, aprovechando sus ventajas climáticas y capacidad de producción a gran escala.

El cultivo de arándano tiene una gran importancia en términos sociales y económico, siendo una industria lucrativa. La creciente demanda en los mercados nacionales e internacionales ha generado oportunidades de exportación y ha impulsado el crecimiento de la agroindustria del arándano en muchos países.

El arándano es un cultivo que tiene sus orígenes en Estados Unidos (s. XIX). Posteriormente, su cultivo se expandió a otras partes del mundo, como Europa, Sudamérica y Oceanía. Hoy en día, el arándano es cultivado en diversos países, adaptándose a diferentes condiciones climáticas.

El arándano pertenece al género *Vaccinium*. Este género comprende unas 30 especies, siendo un grupo muy reducido las empleadas comercialmente. Se trata de un arbusto caduco, que puede ser de porte erecto o rastrero y de altura variable según la especie que se trate.

Clasificación Botánica	
Reino	Vegetal
Orden	Ericales
Familia	Ericáceas
Subgénero	Cyanococcus
Género	<i>Vaccinium</i>
Especies	5 grupos

Imagen 1: Clasificación botánica cultivo del arándano.

Es una planta que necesita de unos requerimientos climáticos especiales para su correcto desarrollo:

- Para el desarrollo del cultivo del arándano, el rango óptimo de temperatura oscila entre 16-25°C.
- El cultivo del arándano requiere de humedad relativa alta.
- Su requerimiento de GDA (Acumulado grados día) es de 850 a 1300 hrs, y su fotoperiodo es de días largos y días cortos
- Tiene un requerimiento de horas-frío (HF = Nº horas con temperatura inferior a 7°C para salir de la latencia) , que depende de la especie y nos permite clasificarlos entre:
 - o HighChill. Alto requerimiento de HF: Requieren más de 800 HF.

- MidChill. Medio requerimiento de HF: Requieren un rango de entre 400 y 800 HF.
- LowChill. Bajo requerimiento de HF: Requieren entre 200-400 HF.
- ZeroChill. Sin requerimientos de horas de frío.

CATEGORÍAS		
Especies	Nombre Común	Requerimiento Frío
V. Corymbosum L.	Arándanos Altos del Norte(Highbush)	800-1000 H.F.
V. Corymbosum L. y V.darrowi	Arándanos Altos del Sur	200-400 H.F. (a veces 600 H.F.)
V. angustifolium y V.myrtilloides	Arándanos Bajos	Requerimientos menores en H.F.
V. vigratum y Vashei Reade	Arándanos Ojo de Conejo(Rabbiteye)	400-800 H.F.
V.corimbosum y V.Angustifolium	Arándanos Medios-Altos	Requerimiento menores de frío

Imagen 2: Categorías desde un punto de vista botánico.

La intención del siguiente proyecto es tratar de buscar una herramienta que permita buscar similitudes climáticas en distintas zonas de cultivo lo que permitirá que se puedan desarrollar nuevas zonas de cultivo. A partir de las coordenadas de una nueva ubicación podremos relacionar con localizaciones conocidas y esto nos permitirá saber que variedades utilizar, o que producciones o manejos esperar.

MATERIAL Y MÉTODOS

Fuente de datos.

La base de nuestro análisis se centra en la recopilación de datos diarios climáticos de los últimos 4 años (2021-2022), para las principales zonas productoras de arándanos en el mundo.

Estos datos han sido obtenidos de la página web de la NASA (<https://power.larc.nasa.gov/data-access-viewer/>) donde entrando con las coordenadas te permite descargar información acerca de:

- Temperatura promedio (°C).
- Temperatura máxima(°C).
- Temperatura mínima(°C).
- Humedad relativa (%)
- Temperatura punto de Rocío (°C)
- Velocidad del viento (m/s)
- Dirección del viento (°)
- Precipitación (mm)
- Radiación (W/m²)

Esta base de datos esta realizada en Excel (csv) y cuenta con información climática de 4.700.000 datos, de 72 localizaciones y 23 países diferentes.

Descripción de los datos.

Nombre	Tipo	Descripción
Country	Carácter	País donde se encuentra el punto de estudio
Located	Carácter	Lugar de importancia productiva
Year	Numérico	Año de toma de datos
Month	Numérico	Mes de toma de datos
Week	Numérico	Mes de toma de datos
Date	Numérico	Fecha en la que se han tomado los registros
T_2m	Numérico	Temperatura en grados °C promedio de cada día a dos m de altura
T_2m_max	Numérico	Temperatura máxima en grados °C de cada día a dos m de altura
T_2m_min	Numérico	Temperatura mínima en grados °C de cada día a dos m de altura.
RH	Numérico	Humedad relativa diaria de cada día en % a dos m del suelo
Dew Point	Numérico	Temperatura a la que se alcanza el punto de rocío en °C de cada día
Wind_Speed	Numérico	Velocidad del viento medidos a dos m del suelo
Wind_Direction	Numérico	Dirección en grados del viento medido a dos m del suelo.
Insolation	Numérico	Radiación del sol medidos en W/m ²

Tabla 1: Descripción datos utilizados en Data_Weather.

Preparación de datos.

Una vez repasados los datos de la base de datos y dada la gran cantidad de datos existentes trabajaremos con datos promedio de cada una de las localizaciones. Así tendremos una observación por zona y evitaremos que haya dependencias temporales entre los datos.

Por eso transformamos la tabla **Data_Weather** en **Data_act** y comenzamos el estudio.

Modelos

Nos basaremos en el análisis de clúster jerárquico para agrupar en función de su similitud y obtener estructuras asociadas dando como resultado final un dendograma.

Para ello realizaremos tres pasos:

- 1.- Realizaremos una comparación entre todos los datos histórico-almacenados basados en los datos promedios de las variables enumeradas.
- 2.- Introducimos la nueva localización.
- 3.- Una vez identificada la localización similar climatología procederemos a realizar un estudio comparativo por meses que nos permitirá afinar más.

RESULTADOS

Análisis descriptivos.

Vamos a realizar un primer análisis descriptivo de la tabla **Data_Weather**.

Comenzamos viendo un resumen de los datos:

```
tibble [231,463 x 17] (S3: tbl_df/tbl/data.frame)
 $ Country          : chr [1:231463] "Italy" "Italy" "Italy" "Italy"
 $ Localted        : chr [1:231463] "Sicilia" "Sicilia" "Sicilia" "Sicilia" ...
 $ Year            : num [1:231463] 2021 2021 2021 2021 2021 ...
 $ Day            : num [1:231463] 1 2 3 4 5 6 7 8 9 10 ...
 $ Date           : POSIXct[1:231463], format: "2021-01-01" "2021-01-02" "2021-01-03" "2021-01-04" ...
 $ Week          : num [1:231463] 1 1 2 2 2 2 2 2 3 ...
 $ Month         : num [1:231463] 1 1 1 1 1 1 1 1 1 ...
 $ T_2M         : num [1:231463] 11.29 12.01 10.95 9.38 9.94 ...
 $ T_2M_max     : num [1:231463] 13.3 14.6 14.1 11.5 13 ...
 $ T_2M_min     : num [1:231463] 9.34 8.52 8.42 7.69 6.62 ...
 $ RH           : num [1:231463] 79.1 75.2 73.5 80.6 71.6 ...
 $ Dew_Point    : num [1:231463] 7.76 7.86 6.3 6.2 4.94 ...
 $ Wind_Speed   : num [1:231463] 3.17 3.57 3.34 3.54 3.88 3.73 3.66 3.08 2.93 2.77 ...
 $ Wind_Direction : num [1:231463] 211 224 230 266 227 ...
 $ Precipitation Corrected (mm/day) : num [1:231463] 1.76 3.96 3.51 8.94 0 0 0 0.85 1.46 0.08 ...
 $ All Sky Surface Shortwave Downward Irradiance (MJ/m^2/day): num [1:231463] 9.74 9.95 10.25 9.87 10.09 ...
 $ Insolation   : num [1:231463] 113 115 119 114 117 ...
```

Tenemos 231.463 filas y 17 variables numéricas y caracteres.

Country	Localted	Year	Day	Date	
Length:231463	Length:231463	Min.:2021	Min.: 1.0	Min.: 2021-01-01 00:00:00.00	
Class :character	Class :character	1st Qu.:2022	1st Qu.: 73.0	1st Qu.:2022-08-03 00:00:00.00	
Mode :character	Mode :character	Median :2023	Median :168.0	Median :2023-01-19 00:00:00.00	
		Mean :2022	Mean :171.4	Mean :2022-12-17 19:09:55.92	
		3rd Qu.:2023	3rd Qu.:265.0	3rd Qu.:2023-07-02 00:00:00.00	
		Max.:2024	Max.:365.0	Max.:2024-03-20 00:00:00.00	
			NA's :169723		
Week	Month	T_2M	T_2M_max	T_2M_min	RH
Min.: 1.00	Min.: 1.000	Min.: -32.77	Min.: -29.59	Min.: -38.82	Min.: -2.52
1st Qu.:16.00	1st Qu.: 4.000	1st Qu.: 13.96	1st Qu.: 16.36	1st Qu.: 6.57	1st Qu.: 48.53
Median :30.00	Median : 7.000	Median : 19.14	Median : 22.94	Median : 12.40	Median : 69.31
Mean :28.42	Mean : 6.867	Mean : 19.18	Mean : 21.88	Mean : 11.42	Mean : 65.32
3rd Qu.:41.00	3rd Qu.:10.000	3rd Qu.: 24.55	3rd Qu.: 28.53	3rd Qu.: 17.29	3rd Qu.: 84.44
Max.:53.00	Max.:12.000	Max.: 46.13	Max.: 47.19	Max.: 31.98	Max.:100.00
			NA's :151831	NA's :151831	NA's :24
Dew_Point	Wind_Speed	Wind_Direction	Precipitation Corrected (mm/day)		
Min.: -32.85	Min.: 0.00	Min.: 0.00	Min.: -2.52		
1st Qu.: 7.88	1st Qu.: 1.46	1st Qu.: 33.44	1st Qu.: 0.00		
Median :12.38	Median : 2.24	Median :149.88	Median : 0.11		
Mean :12.00	Mean : 5.86	Mean :143.66	Mean : 5.01		
3rd Qu.:15.67	3rd Qu.: 3.26	3rd Qu.:233.06	3rd Qu.: 1.88		
Max.: 95.50	Max.:316.12	Max.:354.12	Max.:519.28		
NA's :3258	NA's :151855	NA's :214326	NA's :155089		
All Sky Surface Shortwave Downward Irradiance (MJ/m^2/day)	Insolation				
Min.: 0.00	Min.: 0.00				
1st Qu.: 7.28	1st Qu.: 0.00				
Median :14.87	Median : 0.00				
Mean :15.20	Mean : 59.60				
3rd Qu.:22.29	3rd Qu.: 87.62				
Max.:35.28	Max.:408.33				
NA's :153028					

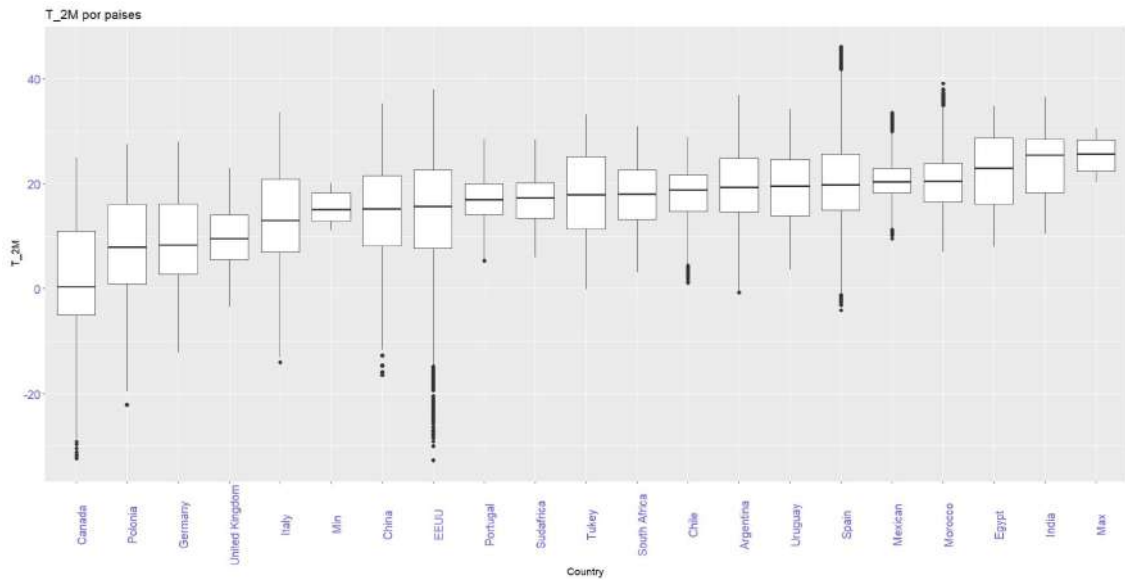
Todos los datos tienen coherencia. De todas las variables nos quedamos con la:

- T_2M
- T_2M_max
- T_2M_min
- RH
- Dew_Point
- Insolation

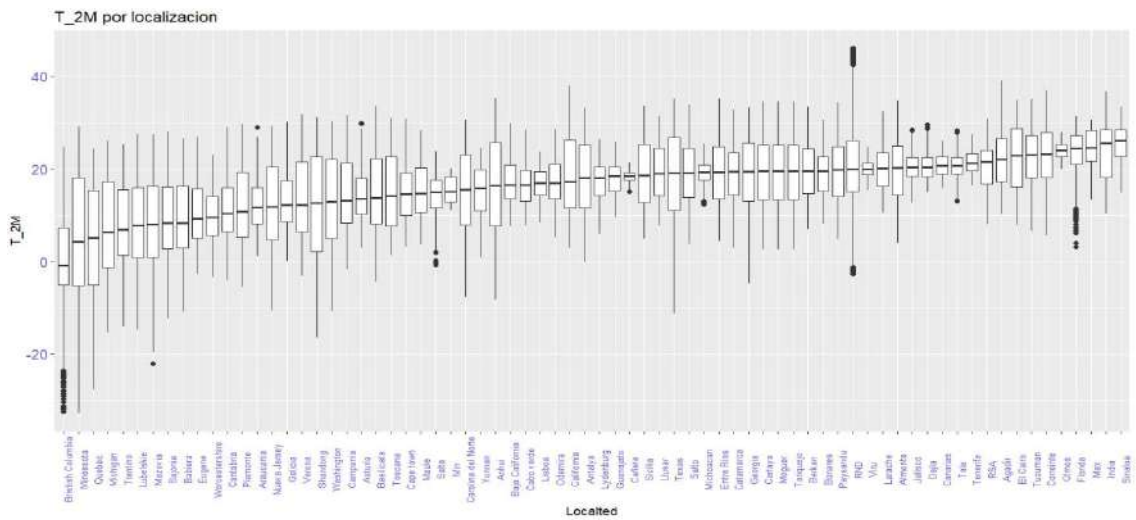
El resto de las variables: Velocidad del viento, Dirección del viento y Precipitación las descartamos por considerar que en la mayoría de las localizaciones el cultivo del arándano es bajo plástico y estas variables no van a interferir en el éxito del cultivo.

Hacemos ahora un análisis exploratorio de las variables para ver relaciones o información discordante:

a. Estudio de las variables independientemente:



Grafica 1: Promedio de temperaturas por Países.

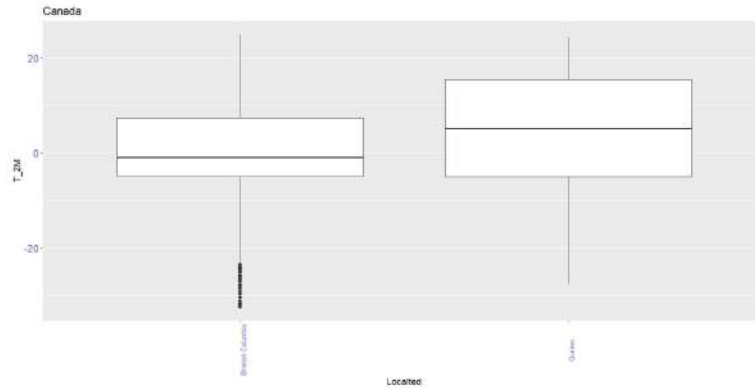


Grafica 2: Promedio de temperaturas por Localizaciones.

Podemos ver con algunos países presentan una serie de valores atípicos.

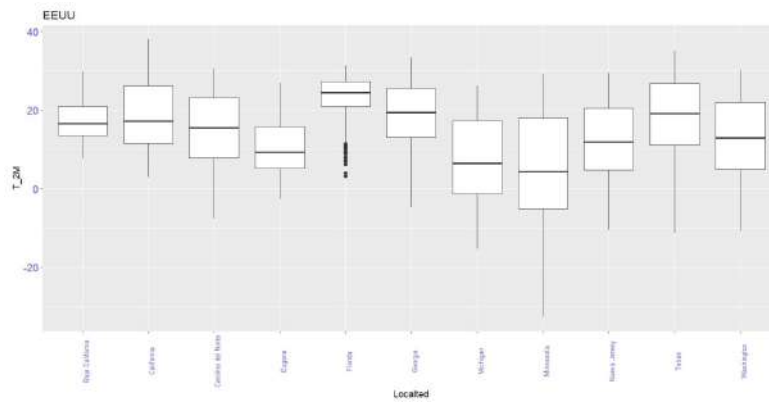
Vamos a ver si esos valores guardan alguna relación más en detalle para lo que estudiaremos independientemente las localizaciones afectadas.

i. Canadá: Se repasan los datos de British Cumbia.



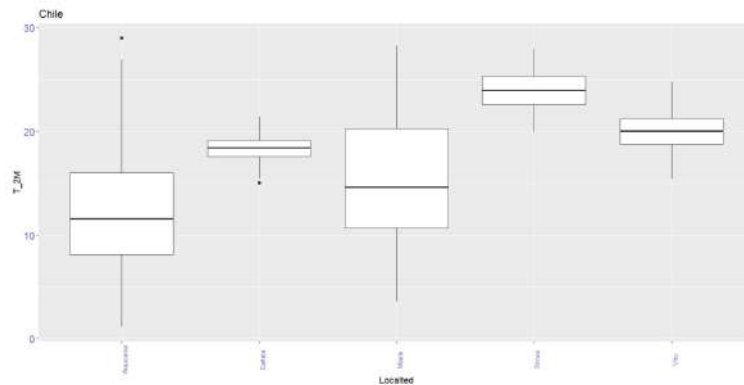
Grafica 3: Promedio de temperaturas Canada.

ii. EEUU: Se repasan los datos de Florida



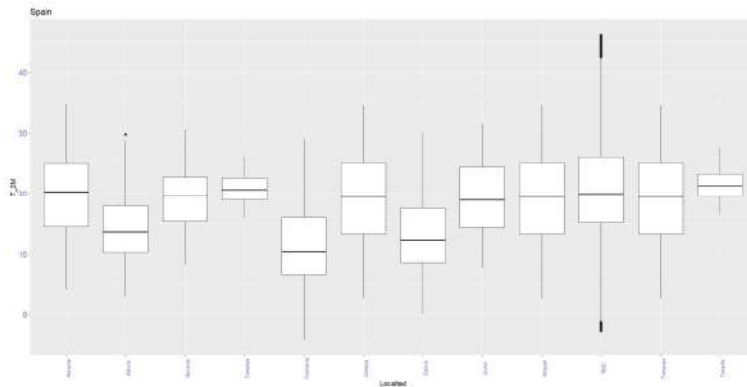
Grafica 4: Promedio de temperaturas Florida..

iii. Chile: No observamos datos variables y si variabilidad climática en el país entre las diferentes zonas.



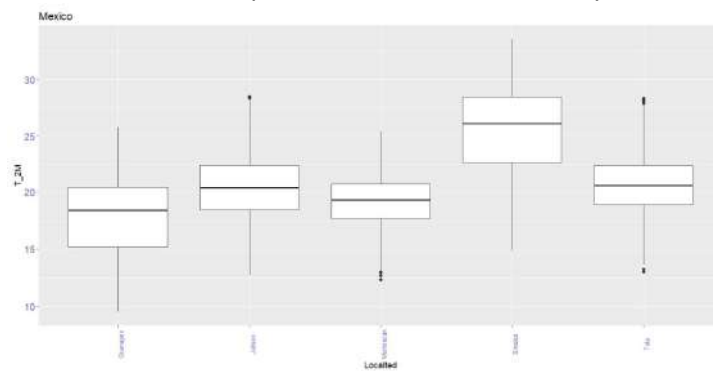
Grafica 5: Promedio de temperaturas Chile

iv. Spain. Se repasan los datos de RND



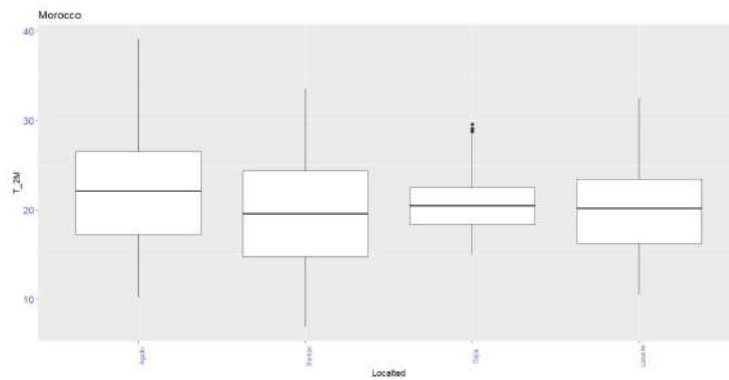
Grafica 6: Promedio de temperaturas Spain.

v. México: Se repasan los datos de Michican y Tala.



Grafica 7: Promedio de temperaturas Mexico..

vi. Marruecos: Se repasan los datos de Dajla.



Grafica 8: Promedio de temperaturas por Países.

Análisis de clúster jerárquico.

Comenzamos realizando un estudio de los datos

```
## {r}
Data_act
summary(Data_act)
head(Data_act)
str(Data_act)
##
```

Country	Located	T_2M	T_2M_max	T_2M_min	RH
Length:71	Length:71	Min. : 0.04957	Min. : 3.389	Min. : -3.570	Min. :20.51
Class :character	Class :character	1st Qu.:13.29090	1st Qu.:18.550	1st Qu.: 8.304	1st Qu.:62.22
Mode :character	Mode :character	Median :17.90797	Median :22.567	Median :12.658	Median :69.78
		Mean :16.36123	Mean :22.234	Mean :11.333	Mean :68.79
		3rd Qu.:19.80857	3rd Qu.:26.375	3rd Qu.:15.002	3rd Qu.:75.61
		Max. :25.51623	Max. :46.120	Max. :20.331	Max. :91.88

Dew_Point	Insolation
Min. : -1.492	Min. : 17.35
1st Qu.: 6.911	1st Qu.:146.19
Median : 9.433	Median :192.65
Mean : 9.382	Mean :175.90
3rd Qu.:12.158	3rd Qu.:218.44
Max. :18.281	Max. :254.09

```
'data.frame': 71 obs. of 8 variables:
 $ Country : chr "Chile" "Chile" "Chile" "Chile" ...
 $ Located : chr "Cañete" "Olmos" "Viru" "Araucania" ...
 $ T_2M : num 18.4 24 20 12.1 15.3 ...
 $ T_2M_max : num 22.9 31.8 24.4 19.2 24.1 ...
 $ T_2M_min : num 15.19 19.14 17.2 6.32 8.51 ...
 $ RH : num 64.5 65.8 75.6 75.3 62.9 ...
 $ Dew_Point : num 10.61 16.32 15.27 6.69 5.99 ...
 $ Insolation: num 64.3 64.4 64.2 204.3 228 ...
```

Agrupamos y vemos los valores medios de las diferentes localizaciones

```
## {r}
Data_act %>%
  group_by(Located) %>%
  get_summary_stats(T_2M, T_2M_max, T_2M_min, RH, Dew_Point, Insolation, type = "mean_sd")
##
```

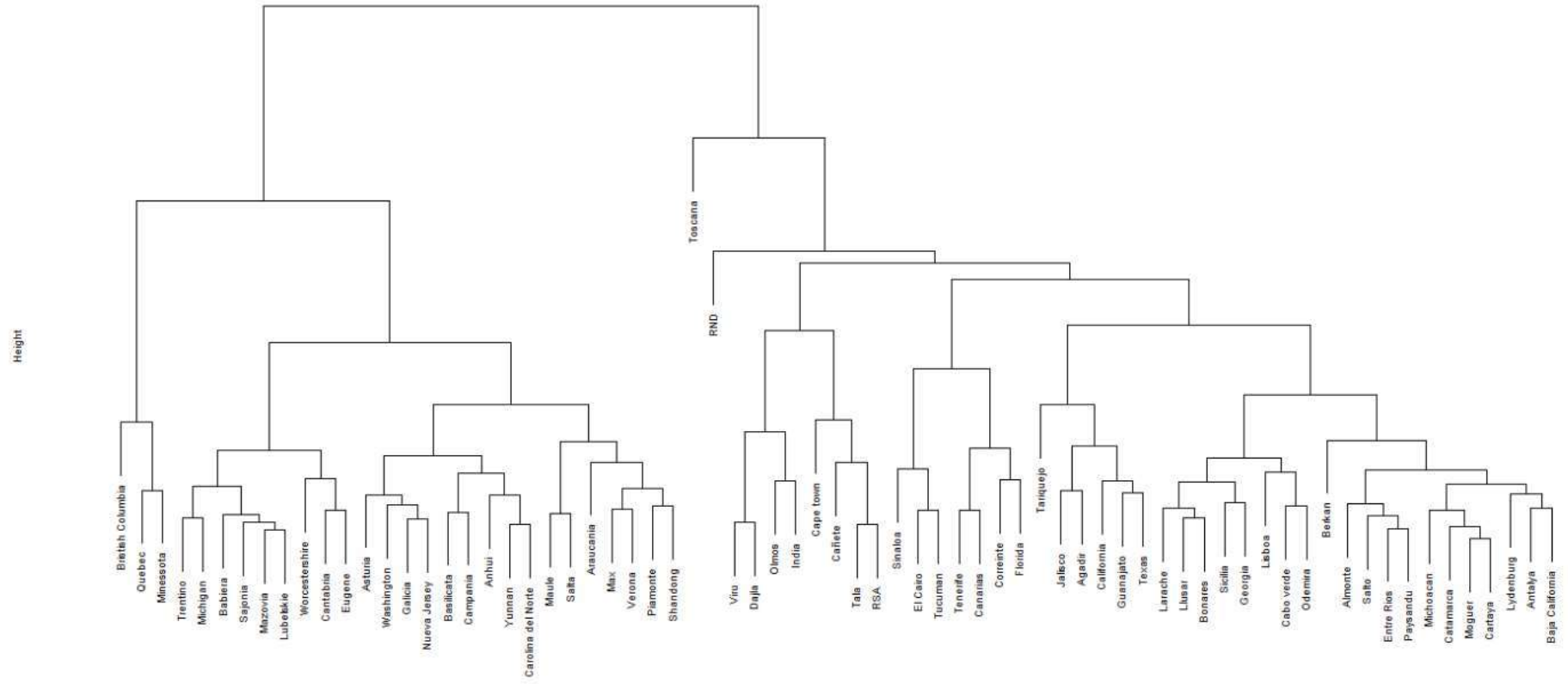
Al tener datos diferente y pretender comparar, previamente debemos escalar, calcular las distancias y hacer el análisis

```
## {r}
Data_act_stdTotal <- scale(Data_act [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_distTotal <- dist(Data_act_std)
Data_act_hcTotal <- hclust(Data_act_dist, method = "complete")
##
```

Graficamos

```
## {r}
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_hcTotal, axes=FALSE)
##
```

Cluster Dendrogram



Data_scd_distTotal
hclust(*, "complete")

Aquí podemos ver las similitudes entre localizaciones.

Introducimos la nueva localización.

Es el momento de introducir una nueva localización. Para realizar este ejemplo hemos decidido seleccionar la siguiente localización.



Imagen 3: Localización ejemplo.

Grospic en Croacia.

Conocida las coordenadas de esta ubicación descargamos los datos climáticos, y calculamos los promedios de las variables a estudiar. Una vez obtenidos los datos lo incluimos en la base de datos **Data_act** con el comando:

```
## {r}
Data_act
New_Located <- c("Croatia", "Grospic", "11.09", "15.87", "6.82", "77.81", "6.60", "160.21")
Croatia_Grospic <- rbind(Data_act, New_Located)
```

```
## {r}
Croatia_Grospic
summary(Croatia_Grospic)
head(Croatia_Grospic)
str(Croatia_Grospic)
```

A tibble: 72 × 8

Country <chr>	Located <chr>	T_2M <chr>	T_2M_max <chr>	T_2M_min <chr>
EEUU	Minessota	4.97074893617021	10.272085106383	-0.297625531914894
Croatia	Grospic	11.09	15.87	6.82

A tibble: 72 × 8

T_2M_max <chr>	T_2M_min <chr>	RH <chr>	Dew_Point <chr>	Insolation <chr>
10.272085106383	-0.297625531914894	77.9554893617021	0.636102127659574	159.797281323877
15.87	6.82	77.81	6.60	160.21

Y procedemos a realizar el análisis de clúster jerárquico:

```
{r}
Croatia_std <- scale(Data_act_Croatia [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Croatia_dist <- dist(Croatia_std)
Croatia_hc <- hclust(Croatia_dist, method = "complete")
---

{r}
par(cex=0.5, mar=c(5,8,4,1))
plot(Croatia_hc, axes=FALSE)
---

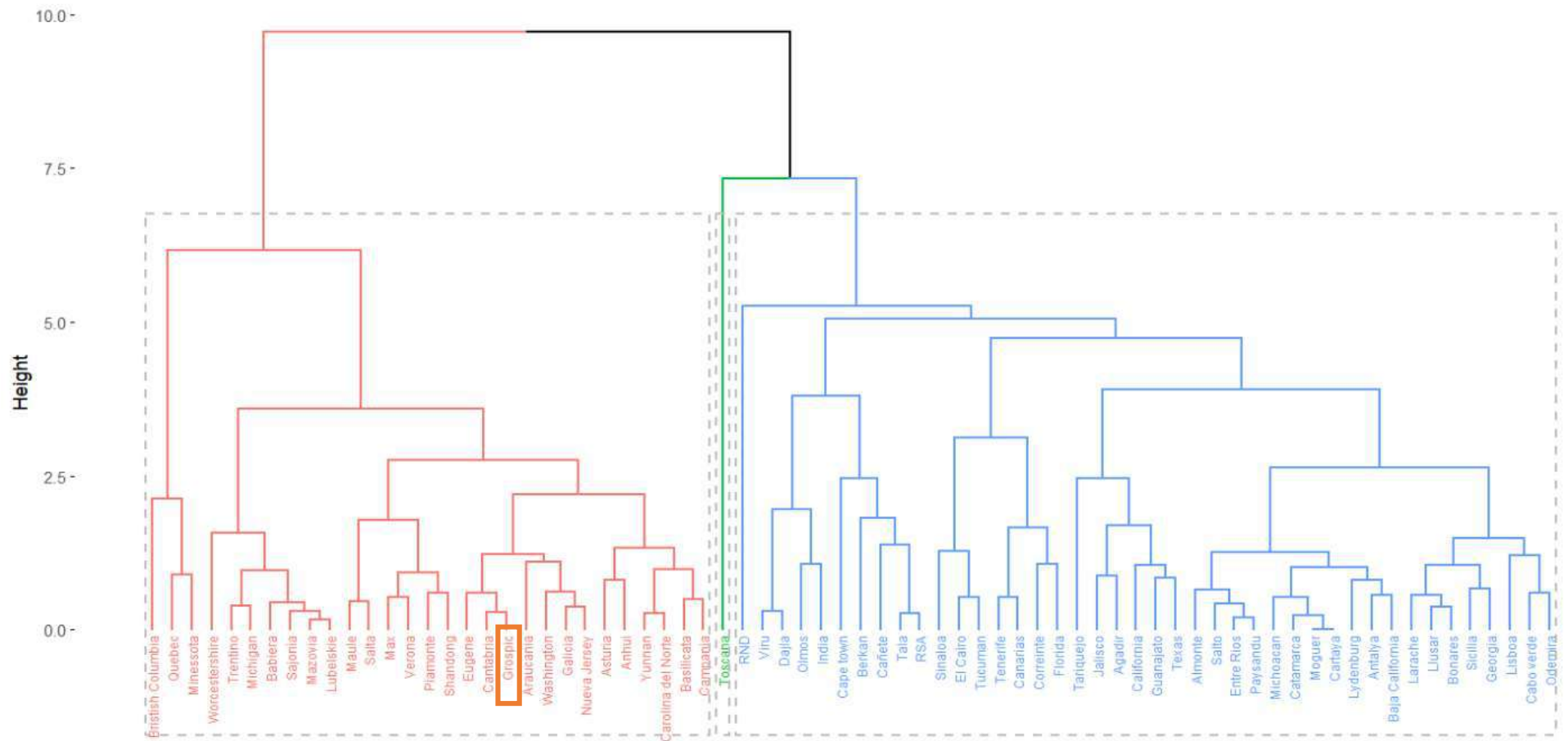
{r}
fviz_dend(Croatia_hc, k=3, rect = TRUE, cex = 0.5)
```

Podemos observar cómo pertenece al grupo rojo y es muy similar a la localización de Cantabria (España), bastante próxima a Eugene (EEUU) y Worcestershire (UK).



Imagen 4: Localización similitud climática con localización de estudio.

Cluster Dendrogram



Estudio detallado mensual

Es el momento de proceder a realizar el estudio comparado por meses para esas 3 localizaciones, procederemos a realizar el análisis de clúster para cada uno de los meses:

ENERO

```

####r
Data_act_Jan <- data.frame(Data_act_Jan)
row.names(Data_act_Jan) <- Data_act_Jan$Located

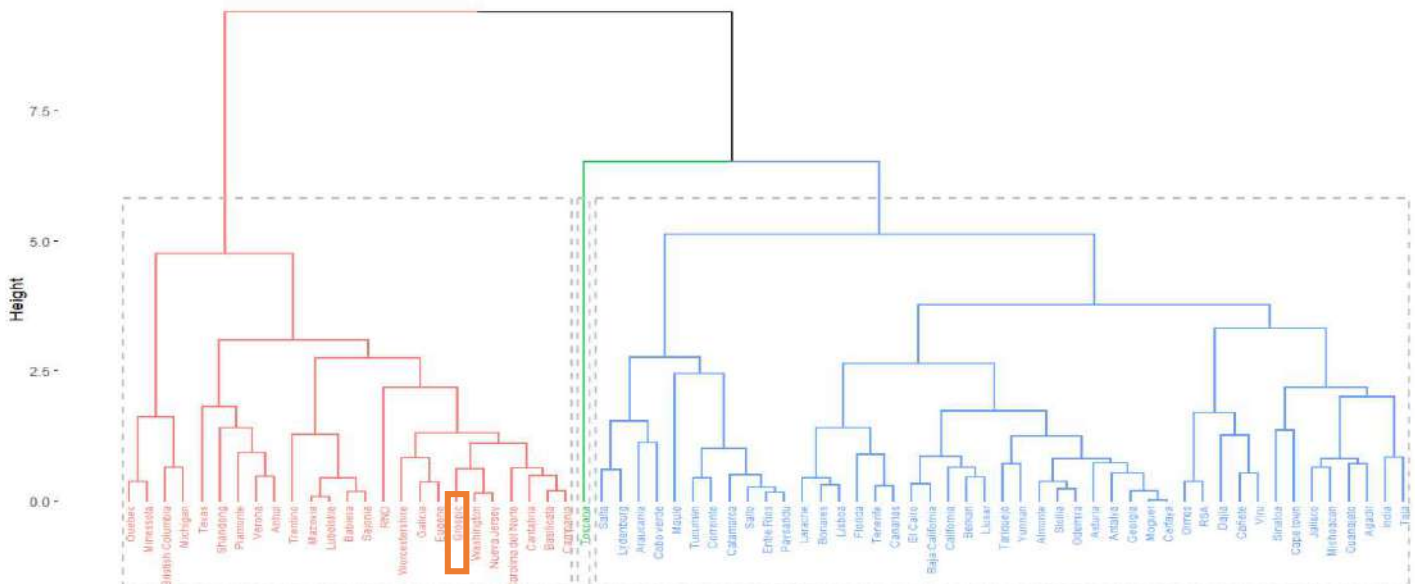
Data_act_Jan_std <- scale(Data_act_Jan [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Jan_dist <- dist(Data_act_Jan_std)
Data_act_Jan_hc <- hclust(Data_act_Jan_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Jan_hc, axes=FALSE)

fviz_dend(Data_act_Jan_hc, k=3, rect = TRUE, cex = 0.5)
####

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	14.5216365	20.050727	10.294365	69.378561	7.949835	118.29294
2	23.9798021	31.124135	17.419216	60.488292	14.137595	300.21884
3	0.8823207	NA	NA	85.481048	-1.653413	59.94524
4	5.9845161	5.112581	3.164597	9.584355	1.688226	23.61671

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	4.676075	5.576406	4.653532	11.398967	4.075511	44.78520
2	4.434485	4.435105	4.477902	9.575855	3.405757	36.35675
3	5.467007	NA	NA	9.751554	4.910551	35.56324
4	NA	NA	NA	NA	NA	NA

FEBRERO

```

####{r}
Data_act_Feb <- data.frame(Data_act_Feb)
row.names(Data_act_Feb) <- Data_act_Feb$located

Data_act_Feb_std <- scale(Data_act_Feb [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Feb_dist <- dist(Data_act_Feb_std)
Data_act_Feb_hc <- hclust(Data_act_Feb_dist, method = "complete")

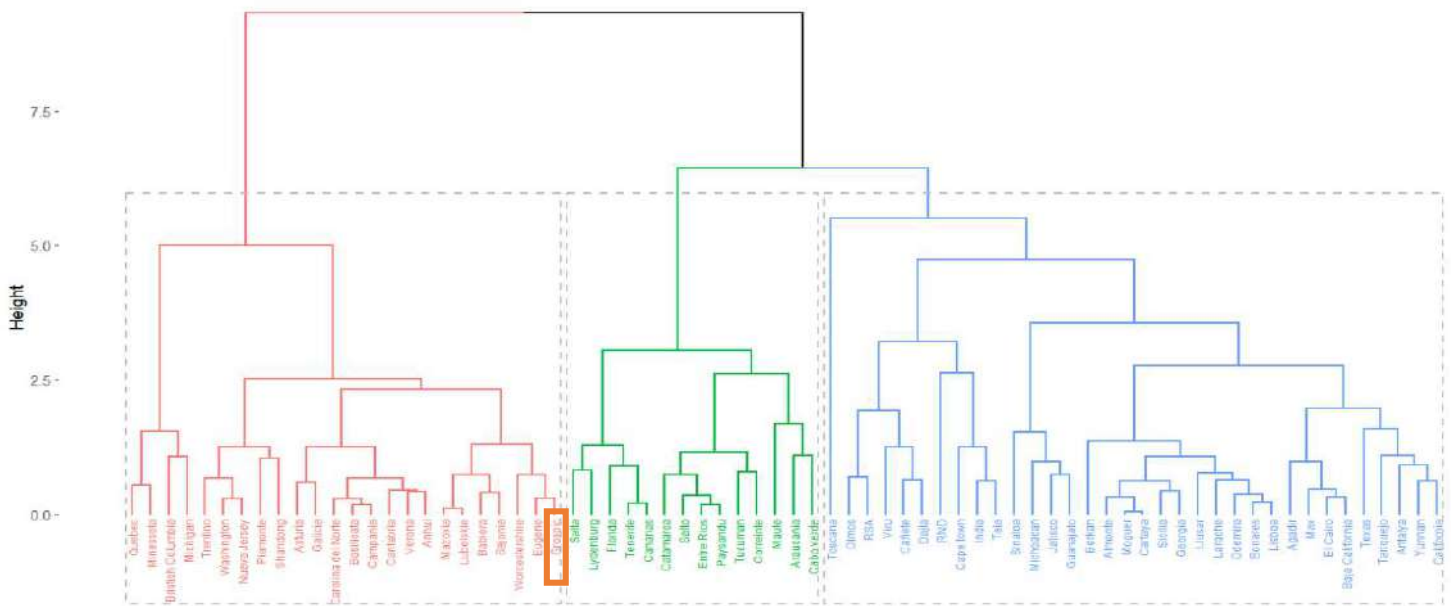
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Feb_hc, axes=FALSE)

fviz_dend(Data_act_Feb_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Feb_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
###

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	15.001447	NA	NA	66.54717	7.559726	138.86624
2	22.170493	28.467918	16.667296	64.92057	13.907870	255.48481
3	2.197198	6.777459	-1.728606	84.69459	-0.458149	99.92566
4	9.088053	7.127345	4.367699	12.49018	3.158142	16.28155

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	4.391386	NA	NA	12.545074	4.342333	59.90956
2	3.866778	5.302409	3.502265	10.138109	3.117064	43.07912
3	5.617189	5.866464	6.071530	7.496559	5.106497	24.32391
4	NA	NA	NA	NA	NA	NA

MARZO

```

---{r}
Data_act_Mar <- data.frame(Data_act_Mar)
row.names(Data_act_Mar) <- Data_act_Mar$Located

Data_act_Mar_std <- scale(Data_act_Mar[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Mar_dist <- dist(Data_act_Mar_std)
Data_act_Mar_hc <- hclust(Data_act_Mar_dist, method = "complete")

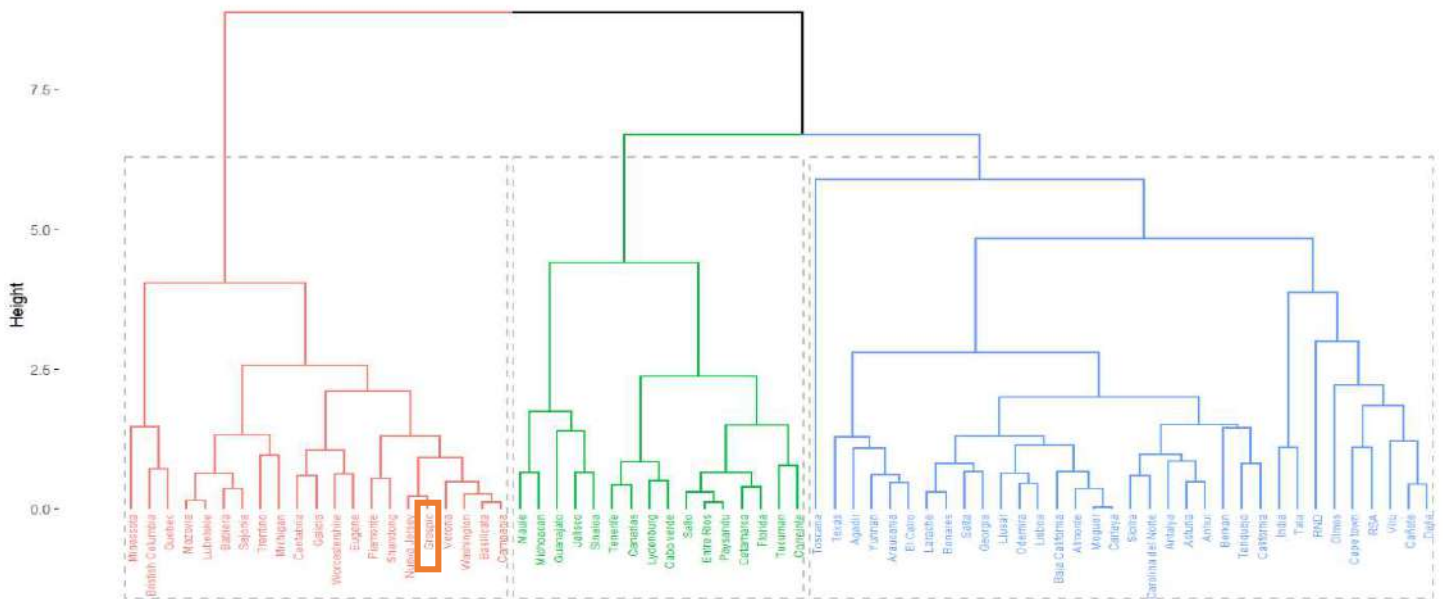
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Mar_hc, axes=FALSE)

fviz_dend(Data_act_Mar_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Mar_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
---

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	15.79736	NA	NA	68.70440	8.8830654	159.50715
2	21.53728	28.367167	15.8288061864	60.24219	11.4792750	245.68446
3	4.39772	9.364158	-0.0001793215	80.37600	0.8377143	145.39694
4	14.47354	8.305752	4.1072566372	14.27690	3.3438053	12.14868

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	3.912800	NA	NA	11.501910	3.698593	66.93243
2	2.601913	3.911064	3.205433	16.892394	6.643308	33.31229
3	4.199546	4.765427	4.137315	7.635044	3.536493	22.36891
4	NA	NA	NA	NA	NA	NA

ABRIL

```

...{r}
Data_act_Abr <- data.frame(Data_act_Abr)
row.names(Data_act_Abr) <- Data_act_Abr$Located

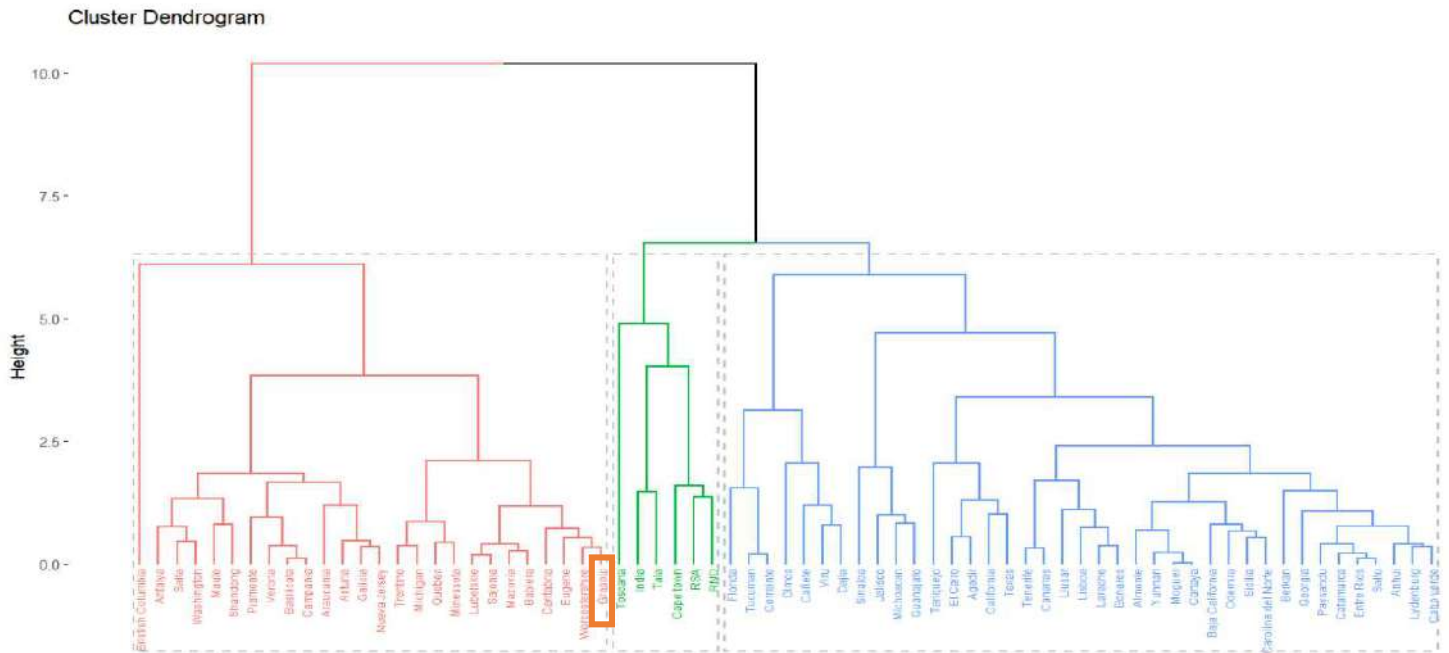
Data_act_Abr_std <- scale(Data_act_Abr[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Abr_dist <- dist(Data_act_Abr_std)
Data_act_Abr_hc <- hclust(Data_act_Abr_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Abr_hc, axes=FALSE)

fviz_dend(Data_act_Abr_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Abr_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
...

```



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	18.773111	24.7359259	13.666390	66.08418	10.688764	225.90131
2	9.427191	14.9259111	4.433200	75.19751	4.615827	190.72680
3	21.228815	NA	NA	40.99065	5.897164	48.90839
4	-3.113333	-0.3421111	-6.985111	95.28267	-4.040111	184.22197

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	2.724999	4.094324	2.924294	12.964498	4.127313	68.45642
2	3.339699	4.130333	3.067153	5.983474	2.508050	22.95661
3	4.262004	NA	NA	18.034842	1.913909	34.27187
4	NA	NA	NA	NA	NA	NA

MAYO

```

...{r}
Data_act_May <- data.frame(Data_act_May)
row.names(Data_act_May) <- Data_act_May$located

Data_act_May_std <- scale(Data_act_May[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_May_dist <- dist(Data_act_May_std)
Data_act_May_hc <- hclust(Data_act_May_dist, method = "complete")

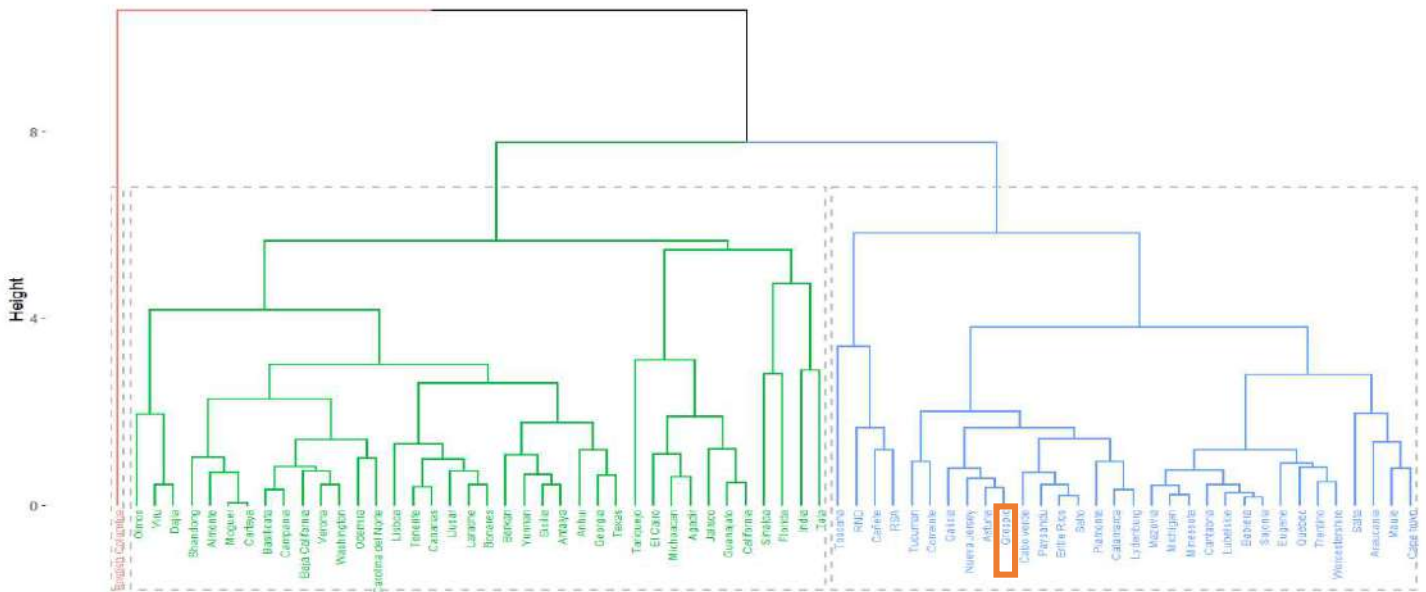
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_May_hc, axes=FALSE)

fviz_dend(Data_act_May_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_May_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
...

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	12.17559	18.30629	7.325379	73.80261	6.774869	129.96770
2	25.90829	NA	NA	52.33666	15.611411	40.54552
3	22.48100	27.84832	17.385982	65.92473	14.619822	264.75720
4	26.19106	33.99261	18.362056	38.91961	3.351611	324.61098

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	2.787801	3.152916	2.8097680	10.562820	3.398024	72.06470
2	3.300279	NA	NA	13.595756	2.017724	33.60426
3	3.314366	4.329791	3.1688370	10.116755	2.713256	53.59498
4	1.167748	1.444776	0.8724912	9.845991	4.234705	20.35933

JUNIO

```

---{r}
Data_act_Jun <- data.frame(Data_act_Jun)
row.names(Data_act_Jun) <- Data_act_Jun$Located

Data_act_Jun_std <- scale(data_act_Jun[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Jun_dist <- dist(Data_act_Jun_std)
Data_act_Jun_hc <- hclust(Data_act_Jun_dist, method = "complete")

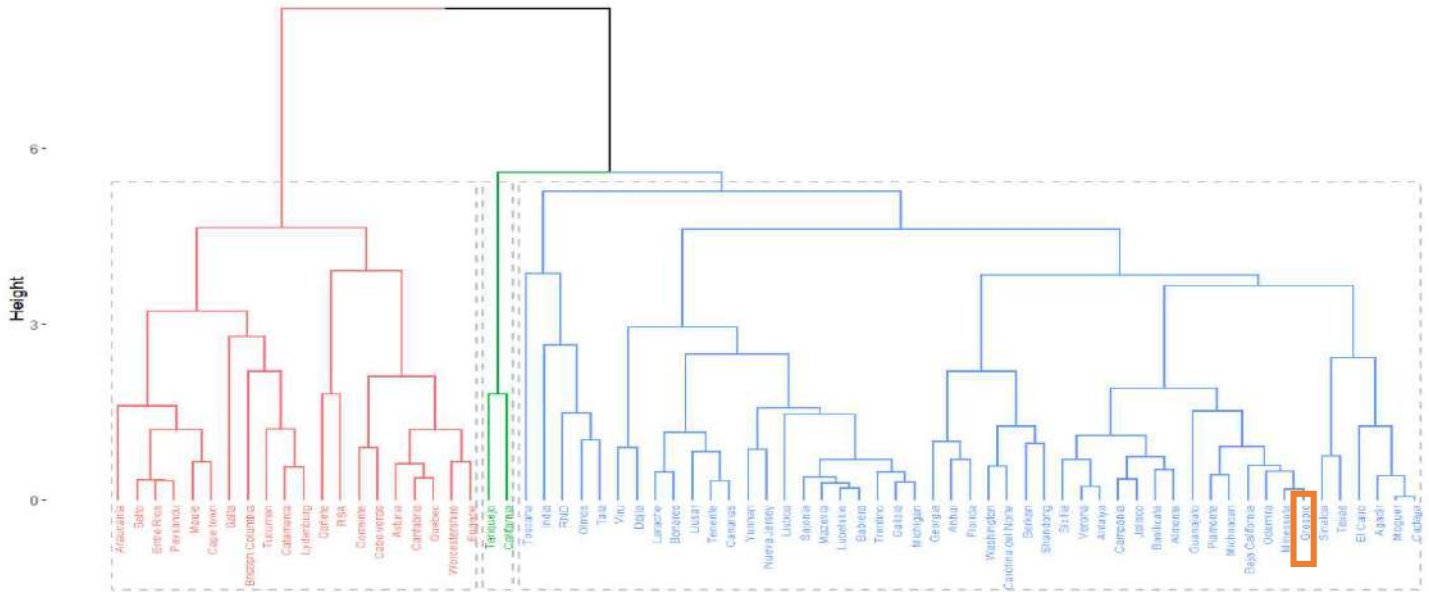
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Jun_hc, axes=FALSE)

fviz_dend(Data_act_Jun_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Jun_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
---

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	12.88860	18.78199	8.11250	74.71273	7.720156	144.58841
2	25.90829	NA	NA	52.33666	15.611411	40.54552
3	22.72560	28.11804	17.58845	65.29497	14.684508	266.44904
4	26.19106	33.99261	18.36206	38.91961	3.351611	324.61098

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	3.103739	3.133061	3.2397846	9.957718	3.886348	75.16462
2	3.300279	NA	NA	13.595756	2.017724	33.60426
3	3.176525	4.233332	3.0819799	9.922899	2.753988	54.21997
4	1.167748	1.444776	0.8724912	9.845991	4.234705	20.35933

JULIO

```

#### {r}
Data_act_Jul <- data.frame(Data_act_Jul)
row.names(Data_act_Jul) <- Data_act_Jul$Located

Data_act_Jul_std <- scale(Data_act_Jul [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Jul_dist <- dist(Data_act_Jul_std)
Data_act_Jul_hc <- hclust(Data_act_Jul_dist, method = "complete")

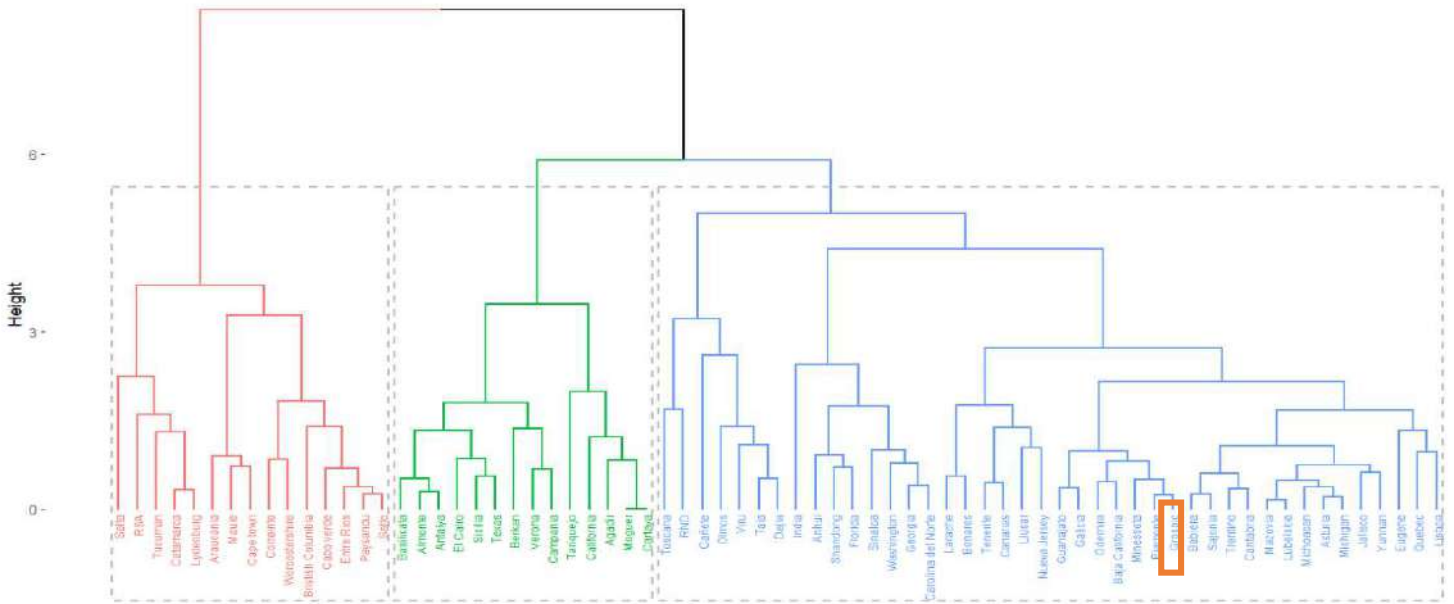
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Jul_hc, axes=FALSE)

fviz_dend(Data_act_Jul_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Jul_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	21.13891	NA	NA	46.57893	11.245453	32.5916
2	23.55399	28.95481	18.909285	68.45073	16.349124	222.1736
3	12.49937	19.34696	6.953871	69.01061	5.600761	167.2062
4	29.27957	37.43284	21.616559	35.36065	7.829118	293.7819

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	4.7220838	NA	NA	9.754710	7.232912	31.066828
2	3.6762498	4.1306523	3.915235	10.651789	3.513854	64.642899
3	3.0546183	4.1795623	2.501044	17.143037	4.180249	50.355555
4	0.8630161	0.7448212	1.463815	4.655888	4.585082	4.446982

AGOSTO

```

#### (r)
Data_act_Ago <- data.frame(Data_act_Ago)
row.names(Data_act_Ago) <- Data_act_Ago$Located

Data_act_Ago_std <- scale(Data_act_Ago [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Ago_dist <- dist(Data_act_Ago_std)
Data_act_Ago_hc <- hclust(Data_act_Ago_dist, method = "complete")

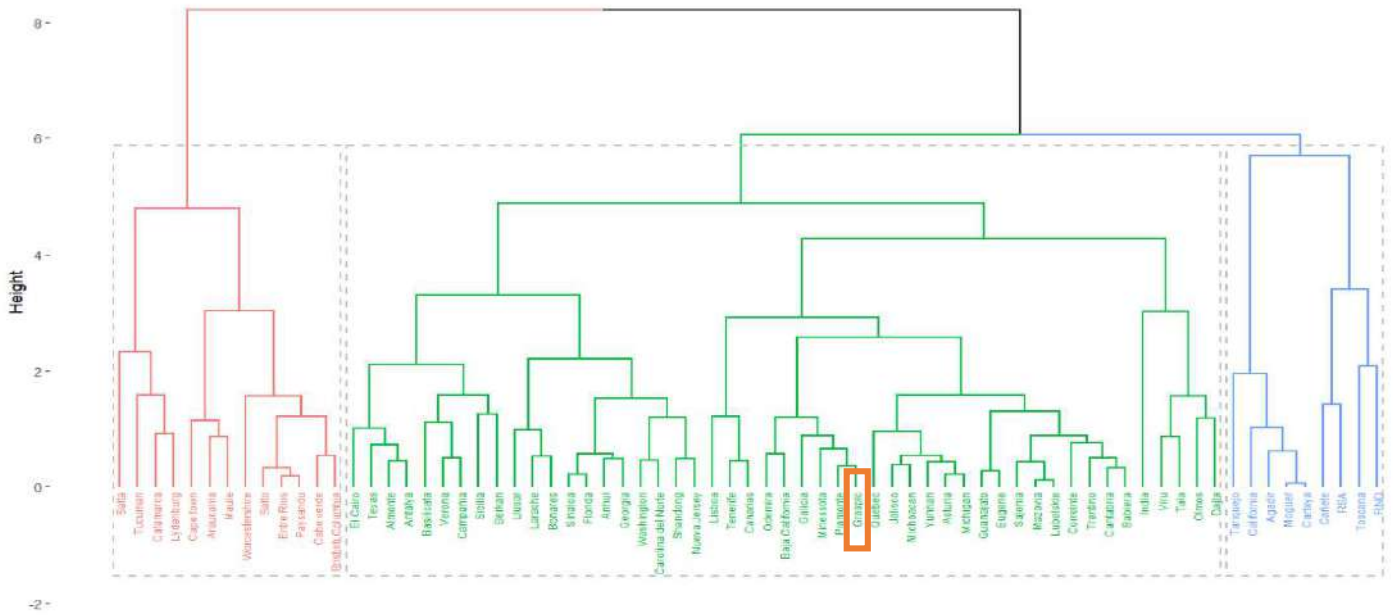
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Ago_hc, axes=FALSE)

fviz_dend(Data_act_Ago_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Ago_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
####

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	21.13891	NA	NA	46.57893	11.245453	32.5916
2	23.55399	28.95481	18.909285	68.45073	16.349124	222.1736
3	12.49937	19.34696	6.953871	69.01061	5.600761	167.2062
4	29.27957	37.43284	21.616559	35.36065	7.829118	293.7819

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	4.7220838	NA	NA	9.754710	7.232912	31.066828
2	3.6762498	4.1306523	3.915235	10.651789	3.513854	64.642899
3	3.0546183	4.1795623	2.501044	17.143037	4.180249	50.355555
4	0.8630161	0.7448212	1.463815	4.655888	4.585082	4.446982

SEPTIEMBRE

```

...[r]
Data_act_Sep <- data.frame(Data_act_Sep)
row.names(Data_act_Sep) <- Data_act_Sep$Located

Data_act_Sep_std <- scale(Data_act_Sep[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Sep_dist <- dist(Data_act_Sep_std)
Data_act_Sep_hc <- hclust(Data_act_Sep_dist, method = "complete")

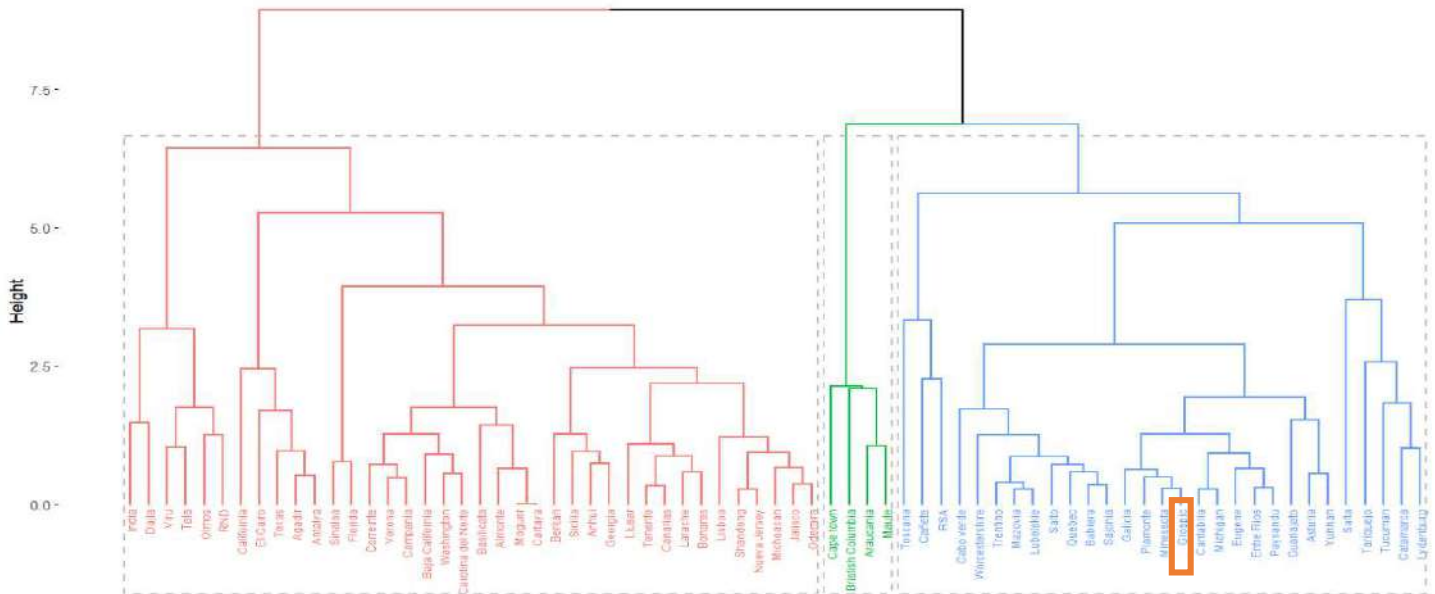
par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Sep_hc, axes=FALSE)

fviz_dend(Data_act_Sep_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Sep_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
...

```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	16.929958	23.24207	11.524375	66.03597	9.359510	173.50269
2	22.635268	NA	NA	73.76730	17.130919	52.99618
3	9.092889	14.92186	4.595472	81.65192	5.556417	138.63844
4	23.524094	28.83269	19.055014	66.86214	15.852351	217.59119

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	2.510269	3.472319	2.3787790	14.661004	3.7506189	53.79990
2	2.487152	NA	NA	6.508459	3.2991076	26.72611
3	1.754088	3.184957	0.8269862	5.493464	0.5131354	58.48435
4	2.428998	3.330065	2.7536658	11.837643	3.4649512	26.92020

OCTUBRE

```
[r]
Data_act_Oct <- data.frame(Data_act_Oct)
row.names(Data_act_Oct) <- Data_act_Oct$Located

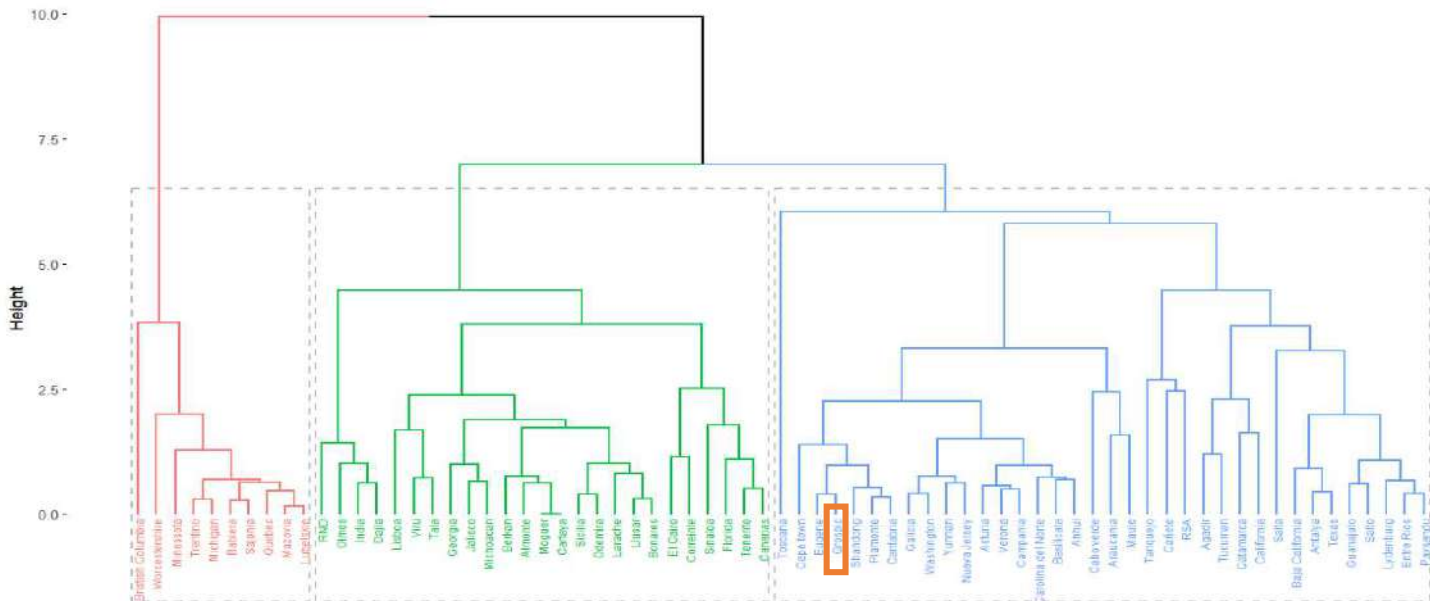
Data_act_Oct_std <- scale(Data_act_Oct [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Oct_dist <- dist(Data_act_Oct_std)
Data_act_Oct_hc <- hclust(Data_act_Oct_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Oct_hc, axes=FALSE)

fviz_dend(Data_act_Oct_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Oct_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
>>>
```

Cluster Dendrogram



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	18.071755	24.166614	13.105217	66.31584	10.381321	166.34679
2	7.441416	NA	NA	83.91035	4.610246	65.61201
3	6.640000	10.638222	7.644000	15.58256	6.948667	36.07767
4	-2.021889	1.384667	-5.348741	87.73774	-4.115926	53.94805

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	3.807990	4.352131	3.978989	8.998033	3.813919	79.89326
2	3.539592	NA	NA	6.787212	3.340795	28.10006
3	NA	NA	NA	NA	NA	NA
4	2.249439	3.048225	2.013475	8.473252	1.411073	18.73776

NOVIEMBRE

```

####[r]
Data_act_Nov <- data.frame(Data_act_Nov)
row.names(Data_act_Nov) <- Data_act_Nov$located

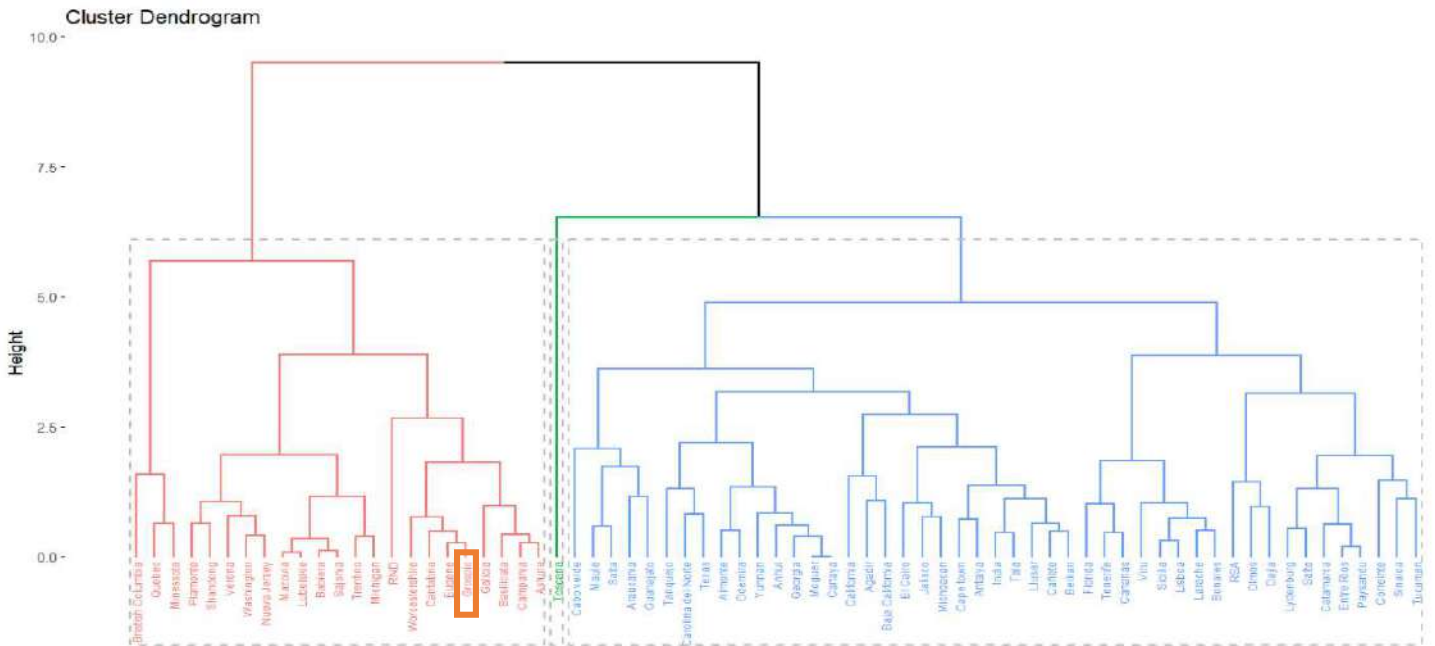
Data_act_Nov_std <- scale(Data_act_Nov[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Nov_dist <- dist(Data_act_Nov_std)
Data_act_Nov_hc <- hclust(Data_act_Nov_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Nov_hc, axes=FALSE)

fviz_dend(Data_act_Nov_hc, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Nov_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
####

```



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	14.5325957	NA	NA	74.95528	9.264343	96.30233
2	22.8388082	29.983378	16.313226	59.98464	12.951326	286.83024
3	0.1786559	3.774677	-2.966822	87.60354	-1.903035	48.07825
4	4.5541935	7.324731	5.878495	11.73645	4.022688	31.52255

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	4.337976	NA	NA	10.473179	3.785821	41.18143
2	3.821006	3.795208	3.974197	8.826407	3.719614	52.70591
3	4.700137	5.010753	5.035080	8.894884	4.491357	25.57020
4	NA	NA	NA	NA	NA	NA

DICIEMBRE

```

---{r}
Data_act_Dec <- data.frame(Data_act_Dec)
row.names(Data_act_Dec) <- Data_act_Dec$located

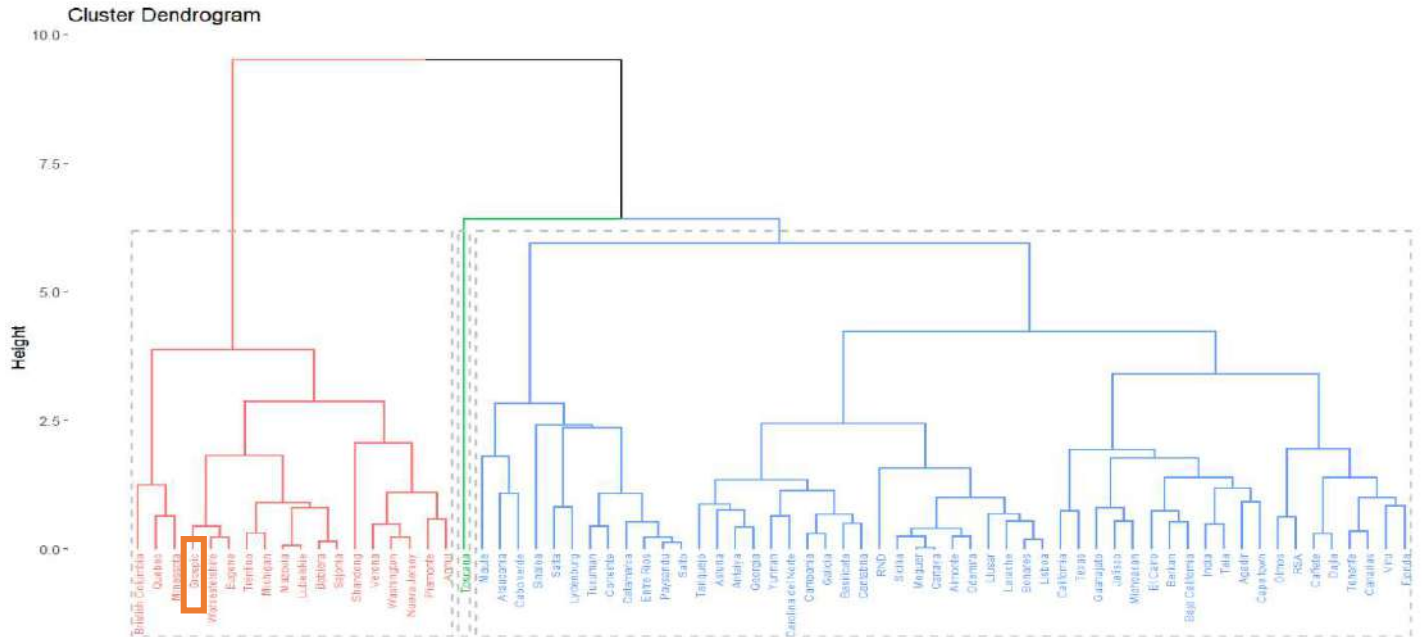
Data_act_Dec_std <- scale(Data_act_Dec[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Dec_dist <- dist(Data_act_Dec_std)
Data_act_Dec_hc <- hclust(Data_act_Dec_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Dec_hc, axes=FALSE)

fviz_dend(Data_act_Dec_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Dec_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )
---

```



Incluimos la media de la variable para los distintos grupos y la desviación típica.

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	14.5325957	NA	NA	74.95528	9.264343	96.30233
2	22.8388082	29.983378	16.313226	59.98464	12.951326	286.83024
3	0.1786559	3.774677	-2.966822	87.60354	-1.903035	48.07825
4	4.5541935	7.324731	5.878495	11.73645	4.022688	31.52255

Group.1 <int>	T_2M <dbl>	T_2M_max <dbl>	T_2M_min <dbl>	RH <dbl>	Dew_Point <dbl>	Insolation <dbl>
1	4.337976	NA	NA	10.473179	3.785821	41.18143
2	3.821006	3.795208	3.974197	8.826407	3.719614	52.70591
3	4.700137	5.010753	5.035080	8.894884	4.491357	25.57020
4	NA	NA	NA	NA	NA	NA

DISCUSIÓN

De este detalle podemos concluir que nuestra localización ejemplo objeto de estudio **Grospic** en **Croatia** es bastante similar a Eugene (EEUU), Cantabria (Spain) y Worcestershire (UK) en casi todos los meses y de forma global, para las variables estudiadas.

Mes	1º Similitud	2º Similitud	3º Similitud
Enero	Washington (EEUU)	Eugene (EEUU)	Nueva Jersey (EEUU)
Febrero	Eugene (EEUU)	Worcestershire (UK)	
Marzo	Nueva Jersey (EEUU)	Washington (EEUU)	Verona (Italia)
Abril	Worcestershire (UK)	Eugene (EEUU)	Cantabria (Spain)
Mayo	Asturia (Spain)	Nueva Jersey (EEUU)	
Junio	Minesota (EEUU)	Odemira	
Julio	Piamonte	Minesota (EEUU)	
Agosto	Piamonte	Minesota (EEUU)	
Septiembre	Cantabria (Spain)	Michigan (EEUU)	
Octubre	Eugene (EEUU)	Cantabria (Spain)	Worcestershire (UK)
Noviembre	Eugene (EEUU)	Cantabria (Spain)	Worcestershire (UK)
Diciembre	Worcestershire (UK)	Eugene (EEUU)	Minesota (EEUU)

Una vez conocida la relación podemos replicar los modelos que mejor se han adaptado a las zonas conocidas.

En concreto para el ejemplo estudiado, podemos replicar las variedades cosechadas en Cantabria, Eugene y Worcestershire, variedades de Fall Creek como *Sekoya Crunch*® 'FC13-063' o *Sekoya Fiesta*® 'FC13-122', irían perfectamente en Grospic (Croacia).



Por el contrario otras variedades como *Sekoya Beauty*® 'FCM12-097' o *Sekoya POP*® 'FCM14-052', no tienen ningún sentido implantarlas, ya que sería un fracaso.



Igual podemos proceder con otras localizaciones, solo tenemos que introducir los datos climáticos promedios en la base de datos y volver a calcular el análisis de clúster con ayuda de RStudio.

CONCLUSIONES

Con el estudio de los cluster, ayudamos a la selección de variedades en el avance de la búsqueda de nuevas localizaciones en el cultivo del arándano.

Si bien tenemos que reconocer que con lo expuesto en el proyecto nos podemos guiar para la selección varietal. Una de las muchas decisiones que hay que tomar a la hora de hacer una inversión agrícola en este cultivo.

El ver las similitudes climáticas nos es de gran ayuda pero tenemos que reconocer que este es solo un apartado de los muchos que hay que estudiar previamente a la inversión como son: comercio, distribución, mano de obra, calidad de suelo, disponibilidad y calidad de agua,...

La base de los cálculos se basa en una amplia base de datos climatológicos. Mientras mayor sea esta base de datos mas fiable y mejor serán las similitudes estudiadas.

Para el proyecto conocer las relaciones promedias de las variables estudiadas nos da una información vital, pero existe una dependencia temporal que tratamos de superar con estudio mensual paralelo.

Dentro del avance y para que este sirva de base seria interesante prepararnos para estudiar modelos de clustering mediante series temporales y afinar aún más nuestras relaciones.

BIBLIOGRAFIA.

Stackhouse, P. (n.d.). *POWER*. Data Access Viewer. Retrieved March 27, 2024, from
<https://power.larc.nasa.gov/data-access-viewer/>

Material Didáctico Maxima Formación

ANEXO CODIGOS R.

```
---
title: "Meses_estudio"
author: "JRufo"
date: "2024-05-01"
output: html_document
---
```

```
```{r}
library(carData)
library(ggplot2)
library(rstatix)
library(ggpubr)
library(dplyr)
library(gmodels)
library(dataset)
library(RVAideMemoire)
library(vcd)
library(MASS)
library(tidyr)
library(Lock5Data)
library(carData)
library(ggfortify)
library(car)
library(asbio)
library(emmeans)
library(ISLR)
library(summarytools)
library(caret)
library(effects)
library(sjPlot)
library(Epi)
library(performance)
library(lme4)
library(lmerTest)
library(MUMIn)
library(report)
library(scales)
library(GGally)
library(cluster)
library(factoextra)
library(NbClust)
library(flexclust)
library(FeatureImpCluster)
library(ggrepel)
library(cluster)
library(fpc)
library(mvnormtest)
library(faraway)
library(GGally)
library(klaR)
library(FactoMiner)
library(factoextra)
library(psych)
library(factoextra)
library(heatmap)
```
```

```
# 0.- CARGAR LOS DATOS.
```

```
```{r}
Data_Weather
summary(Data_Weather)
head(Data_Weather)
str(Data_Weather)
```
```

a) Dibujar grafico.

a.1) Estudio de las variables independientemente.

Tras estudiarlos datos vemos coherencia en las graficas en todas las variables y posibilidad de valores outliers.

```
```{r}
ggplot(Data_Weather, aes(x=Date, y=T_2M))+geom_point()+ ggtitle("Temperaura promedio")
ggplot(Data_Weather, aes(x=Date, y=T_2M_max))+geom_point()+ggtitle("Temperaura max")
ggplot(Data_Weather, aes(x=Date, y=T_2M_min))+geom_point()+ggtitle("Temperaura min")
ggplot(Data_Weather, aes(x=Date, y=RH))+geom_point()+ ggtitle("Humedad relativa")
ggplot(Data_Weather, aes(x=Date, y=Dew_Point))+geom_point()+ggtitle("Dew Point")
ggplot(Data_Weather, aes(x=Date, y=Insolation))+geom_point()+ggtitle("Insolacion")
```
```

a.2) Estudio variables por paises.

```
{r}
Data_Weather %>%
  dplyr::select(T_2M) %>%
  get_summary_stats(type = "mean_sd")

Data_Weather %>%
  dplyr::select(T_2M_max) %>%
  get_summary_stats(type = "mean_sd")

Data_Weather %>%
  dplyr::select(T_2M_min) %>%
  get_summary_stats(type = "mean_sd")

Data_Weather %>%
  dplyr::select(RH) %>%
  get_summary_stats(type = "mean_sd")

Data_Weather %>%
  dplyr::select(Dew_Point) %>%
  get_summary_stats(type = "mean_sd")

Data_Weather %>%
  dplyr::select(Insolation) %>%
  get_summary_stats(type = "mean_sd")
---
```

```
{r}

group_orderedDW <- with(Data_Weather, reorder(Country, T_2M, median))
data_orderDW <- Data_Weather
data_orderDW$Country <- factor(data_orderDW$Country, levels = levels(group_orderedDW))

ggplot(data_orderDW, aes(x= Country, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 12,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("T_2M por paises")+
geom_boxplot()
```

```
{r}

Data_Weather %>%
  dplyr::select(T_2M) %>%
  get_summary_stats(type = "mean_sd")

group_orderedLocal <- with(Data_Weather, reorder(Localited, T_2M, median))
data_orderLocal <- Data_Weather
data_orderLocal$Localited <- factor(data_orderLocal$Localited, levels = levels(group_orderedLocal))

ggplot(data_orderLocal, aes(x= Localited, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("T_2M por
localizacion")+ geom_boxplot()
```

Estudio problemas en Canada

```
{r}

datosCanada <- filter(Data_Weather, Country == "Canada")

ggplot(datosCanada, aes(x= Localited, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("Canada")+
geom_boxplot()
```

Estudio problemas en EEUU

```
{r}

datosEEUU <- filter(Data_Weather, Country == "EEUU")

ggplot(datosEEUU, aes(x= Localited, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("EEUU")+
geom_boxplot()
```

Estudio problemas en Chile

```
{r}

datosChile <- filter(Data_Weather, Country == "Chile")

ggplot(datosChile, aes(x= Localited, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("Chile")+
geom_boxplot()
```

```
Estudio problemas en Spain
{r}

datosSpain <- filter(Data_weather, Country == "Spain")

ggplot(datosSpain, aes(x= Localted, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("Spain")+
geom_boxplot()
```

```
Estudio problemas en Mexico
{r}

datosMexican <- filter(Data_weather, country == "Mexican")

ggplot(datosMexican, aes(x= Localted, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("Mexico")+
geom_boxplot()
```

```
Estudio problemas en Morocco
{r}

datosMorocco <- filter(Data_weather, Country == "Morocco")

ggplot(datosMorocco, aes(x= Localted, y= T_2M)) + theme(axis.text.x = element_text(color = "slateblue", size = 8,
angle = 90), axis.text.y = element_text(color = "slateblue", size = 12, angle = 0)) + ggtitle("Morocco")+
geom_boxplot()
```

2.- EVALUAR LOS SUPUESTOS:
a) outliers

```
{r}
Data_weather %>%
  identify_outliers(T_2M)
---
```

```
{r}
Data_weather %>%
  identify_outliers(T_2M_max)
---
```

```
{r}
Data_weather %>%
  identify_outliers(T_2M_min)
---
```

```
{r}
Data_weather %>%
  identify_outliers(Dew_Point)
---
```

```
{r}
Data_weather %>%
  identify_outliers(RH)
---
```

```
{r}
Data_weather %>%
  identify_outliers(Insolation)
---
```

b) Normalidad

```
{r}
Data_weather %>%
  shapiro_test(T_2M)
---
```

```
c) Homogeneidad
{r}
plot(Data_weather$T_2M)
plot(Data_weather$T_2M_max)
plot(Data_weather$T_2M_min)
plot(Data_weather$RH)
plot(Data_weather$Dew_Point)
plot(Data_weather$Insolation)
...

```

```
# 1.- Meter nueva localizacion en Data_Act
{r}
Data_act
New_Located <- c("Croatia", "Grosopic", "11.09", "15.87", "6.82", "77.81", "6.60", "160.21")
Croatia_Grospic <- rbind(Data_act, New_Located)
...

```

```
# 2.- Comprobamos que esta medida la nueva localizacion.
{r}
Croatia_Grospic
summary(Croatia_Grospic)
head(Croatia_Grospic)
str(Croatia_Grospic)
...

```

```
{r}
Croatia_Grospic <- data.frame(Croatia_Grospic)
row.names(Croatia_Grospic) <- Croatia_Grospic$Located
...

```

```
{r}
Data_act_Croatia
Data_act_Jan
Data_act_Feb
Data_act_Mar
Data_act_Abr
Data_act_May
Data_act_Jun
Data_act_Jul
Data_act_Ago
Data_act_Sep
Data_act_Oct
Data_act_Nov
Data_act_Dec
...

```

```
{r}
Data_act_Croatia <- data.frame(Data_act_Croatia)
row.names(Data_act_Croatia) <- Data_act_Croatia$Located

Data_act_Croatia_std <- scale(Data_act_Croatia[,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point",
"Insolation")])
Data_act_Croatia_dist <- dist(Data_act_Croatia_std)
Data_act_Croatia_hc <- hclust(Data_act_Croatia_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Croatia_hc, axes=FALSE)

fviz_dend(Data_act_Croatia_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Croatia_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )

grupos <- cutree(Data_act_Croatia_hc, 4)

aggregate(Data_act_Croatia[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos),
mean)

aggregate(Data_act_Croatia[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)
...

```

```

ENERO
```{r}

Data_act_Jan <- data.frame(Data_act_Jan)
row.names(Data_act_Jan) <- Data_act_Jan$Located

Data_act_Jan_std <- scale(Data_act_Jan [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Jan_dist <- dist(Data_act_Jan_std)
Data_act_Jan_hc <- hclust(Data_act_Jan_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Jan_hc, axes=FALSE)

fviz_dend(Data_act_Jan_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Jan_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Jan_hc, 4)

grupos

aggregate(Data_act_Jan[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Jan[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)
```

```

```

FEBRERO
```{r}

Data_act_Feb <- data.frame(Data_act_Feb)
row.names(Data_act_Feb) <- Data_act_Feb$Located

Data_act_Feb_std <- scale(Data_act_Feb [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Feb_dist <- dist(Data_act_Feb_std)
Data_act_Feb_hc <- hclust(Data_act_Feb_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Feb_hc, axes=FALSE)

fviz_dend(Data_act_Feb_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Feb_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Feb_hc, 4)

grupos

aggregate(Data_act_Feb[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Feb[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)
```

```

```

MARZO
```{r}

Data_act_Mar <- data.frame(Data_act_Mar)
row.names(Data_act_Mar) <- Data_act_Mar$Located

Data_act_Mar_std <- scale(Data_act_Mar [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Mar_dist <- dist(Data_act_Mar_std)
Data_act_Mar_hc <- hclust(Data_act_Mar_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Mar_hc, axes=FALSE)

fviz_dend(Data_act_Mar_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Mar_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Mar_hc, 4)

grupos

aggregate(Data_act_Mar[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Mar[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)
```

```

```

ABRIL
## {r}

Data_act_Abr <- data.frame(Data_act_Abr)
row.names(Data_act_Abr) <- Data_act_Abr$Located

Data_act_Abr_std <- scale(Data_act_Abr [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Abr_dist <- dist(Data_act_Abr_std)
Data_act_Abr_hc <- hclust(Data_act_Abr_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Abr_hc, axes=FALSE)

fviz_dend(Data_act_Abr_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Abr_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )

grupos <- cutree(Data_act_Abr_hc, 4)

grupos

aggregate(Data_act_Abr[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Abr[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

MAYO
## {r}

Data_act_May <- data.frame(Data_act_May)
row.names(Data_act_May) <- Data_act_May$Located

Data_act_May_std <- scale(Data_act_May [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_May_dist <- dist(Data_act_May_std)
Data_act_May_hc <- hclust(Data_act_May_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_May_hc, axes=FALSE)

fviz_dend(Data_act_May_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_May_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )

grupos <- cutree(Data_act_May_hc, 4)

grupos

aggregate(Data_act_May[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_May[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

JUNIO
## {r}

Data_act_Jun <- data.frame(Data_act_Jun)
row.names(Data_act_Jun) <- Data_act_Jun$Located

Data_act_Jun_std <- scale(Data_act_Jun [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Jun_dist <- dist(Data_act_Jun_std)
Data_act_Jun_hc <- hclust(Data_act_Jun_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Jun_hc, axes=FALSE)

fviz_dend(Data_act_Jun_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Jun_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )

grupos <- cutree(Data_act_Jun_hc, 4)

grupos

aggregate(Data_act_Jun[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Jun[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

JULIO
```{r}

Data_act_Jul <- data.frame(Data_act_Jul)
row.names(Data_act_Jul) <- Data_act_Jul$Located

Data_act_Jul_std <- scale(Data_act_Jul [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Jul_dist <- dist(Data_act_Jul_std)
Data_act_Jul_hc <- hclust(Data_act_Jul_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Jul_hc, axes=FALSE)

fviz_dend(Data_act_Jul_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Jul_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Jul_hc, 4)

grupos

aggregate(Data_act_Jul[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Jul[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

```

AGOSTO
```{r}

Data_act_Ago <- data.frame(Data_act_Ago)
row.names(Data_act_Ago) <- Data_act_Ago$Located

Data_act_Ago_std <- scale(Data_act_Ago [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Ago_dist <- dist(Data_act_Ago_std)
Data_act_Ago_hc <- hclust(Data_act_Ago_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Ago_hc, axes=FALSE)

fviz_dend(Data_act_Ago_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Ago_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Ago_hc, 4)

grupos

aggregate(Data_act_Ago[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Ago[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

```

SEPTIEMBRE
```{r}

Data_act_Sep <- data.frame(Data_act_Sep)
row.names(Data_act_Sep) <- Data_act_Sep$Located

Data_act_Sep_std <- scale(Data_act_Sep [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Sep_dist <- dist(Data_act_Sep_std)
Data_act_Sep_hc <- hclust(Data_act_Sep_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Sep_hc, axes=FALSE)

fviz_dend(Data_act_Sep_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Sep_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Sep_hc, 4)

grupos

aggregate(Data_act_Sep[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Sep[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

```

OCTUBRE
```{r}

Data_act_Oct <- data.frame(Data_act_Oct)
row.names(Data_act_Oct) <- Data_act_Oct$Located

Data_act_Oct_std <- scale(Data_act_Oct [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Oct_dist <- dist(Data_act_Oct_std)
Data_act_Oct_hc <- hclust(Data_act_Oct_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Oct_hc, axes=FALSE)

fviz_dend(Data_act_Oct_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Oct_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Oct_hc, 4)

grupos

aggregate(Data_act_Oct[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Oct[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

NOVIEMBRE
```{r}

Data_act_Nov <- data.frame(Data_act_Nov)
row.names(Data_act_Nov) <- Data_act_Nov$Located

Data_act_Nov_std <- scale(Data_act_Nov [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Nov_dist <- dist(Data_act_Nov_std)
Data_act_Nov_hc <- hclust(Data_act_Nov_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Nov_hc, axes=FALSE)

fviz_dend(Data_act_Nov_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Nov_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100) )

grupos <- cutree(Data_act_Nov_hc, 4)

grupos

aggregate(Data_act_Nov[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Nov[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```

```

DICIEMBRE
```{r}

Data_act_Dec <- data.frame(Data_act_Dec)
row.names(Data_act_Dec) <- Data_act_Dec$Located

Data_act_Dec_std <- scale(Data_act_Dec [,c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")])
Data_act_Dec_dist <- dist(Data_act_Dec_std)
Data_act_Dec_hc <- hclust(Data_act_Dec_dist, method = "complete")

par(cex=0.5, mar=c(5,8,4,1))
plot(Data_act_Dec_hc, axes=FALSE)

fviz_dend(Data_act_Dec_hc, k=3, rect = TRUE, cex = 0.5)

pheatmap(Data_act_Dec_std, cutree_rows = 4, row_names_gp = gpar(fontsize = 100))

grupos <- cutree(Data_act_Dec_hc, 4)

grupos

aggregate(Data_act_Dec[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), mean)
aggregate(Data_act_Dec[, c("T_2M", "T_2M_max", "T_2M_min", "RH", "Dew_Point", "Insolation")], list(grupos), sd)

```