ev02\_a01285371

Pedro Sotelo Arce

2023-05-06

Segun la OMS Los 20 paises con mas casos reportados de COVID son

Estados Unidos MN985325  
China MN908947  
India MT012098  
Francia MT320538  
Alemania MT394864  
Brazil MT126808  
Japon LC529905  
Korea MT039890  
Italia MT066156  
Reino unido LR813996  
Russia MT510643  
Turkia MT327745  
España MT198651  
Vietnam MT192772  
Australia MT007544  
Argentina MW553294  
Paises Bajos MT396266  
Iran MT281530  
Mexico MT810752  
Indonesia MZ026853

virus <- c( "MN985325", "MN908947", "MT012098", "MT320538", "MT394864" , "MT126808" , "LC529905", "MT039890" , "MT066156" , "LR813996" , "MT510643" , "MT327745" , "MT198651" , "MT192772" , "MT007544" , "MW553294" , "MT396266" , "MT281530" , "MT810752" , "MZ026853")

1. Carga las librerías necesarias:

library(Biostrings)

## Loading required package: BiocGenerics

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
## colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
## get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
## match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
## Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
## table, tapply, union, unique, unsplit, which.max, which.min

## Loading required package: S4Vectors

## Loading required package: stats4

##   
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:base':  
##   
## expand.grid, I, unname

## Loading required package: IRanges

##   
## Attaching package: 'IRanges'

## The following object is masked from 'package:grDevices':  
##   
## windows

## Loading required package: XVector

## Loading required package: GenomeInfoDb

##   
## Attaching package: 'Biostrings'

## The following object is masked from 'package:base':  
##   
## strsplit

library(adegenet)

## Loading required package: ade4

##   
## Attaching package: 'ade4'

## The following object is masked from 'package:Biostrings':  
##   
## score

## The following object is masked from 'package:BiocGenerics':  
##   
## score

##   
## /// adegenet 2.1.10 is loaded ////////////  
##   
## > overview: '?adegenet'  
## > tutorials/doc/questions: 'adegenetWeb()'   
## > bug reports/feature requests: adegenetIssues()

library(ape)

##   
## Attaching package: 'ape'

## The following object is masked from 'package:Biostrings':  
##   
## complement

library(ggtree)

## ggtree v3.6.2 For help: https://yulab-smu.top/treedata-book/  
##   
## If you use the ggtree package suite in published research, please cite  
## the appropriate paper(s):  
##   
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.  
## ggtree: an R package for visualization and annotation of phylogenetic  
## trees with their covariates and other associated data. Methods in  
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628  
##   
## LG Wang, TTY Lam, S Xu, Z Dai, L Zhou, T Feng, P Guo, CW Dunn, BR  
## Jones, T Bradley, H Zhu, Y Guan, Y Jiang, G Yu. treeio: an R package  
## for phylogenetic tree input and output with richly annotated and  
## associated data. Molecular Biology and Evolution. 2020, 37(2):599-603.  
## doi: 10.1093/molbev/msz240  
##   
## Shuangbin Xu, Lin Li, Xiao Luo, Meijun Chen, Wenli Tang, Li Zhan, Zehan  
## Dai, Tommy T. Lam, Yi Guan, Guangchuang Yu. Ggtree: A serialized data  
## object for visualization of a phylogenetic tree and annotation data.  
## iMeta 2022, 4(1):e56. doi:10.1002/imt2.56

##   
## Attaching package: 'ggtree'

## The following object is masked from 'package:ape':  
##   
## rotate

## The following object is masked from 'package:Biostrings':  
##   
## collapse

## The following object is masked from 'package:IRanges':  
##   
## collapse

## The following object is masked from 'package:S4Vectors':  
##   
## expand

library(DECIPHER)

## Loading required package: RSQLite

## Loading required package: parallel

library(viridis)

## Loading required package: viridisLite

library(ggplot2)  
library(seqinr)

##   
## Attaching package: 'seqinr'

## The following objects are masked from 'package:ape':  
##   
## as.alignment, consensus

## The following object is masked from 'package:Biostrings':  
##   
## translate

1. Obtén las secuencias:

virus\_sequences <- read.GenBank(virus)

1. Longitudes de los virus

USA = read.fasta("USA.fasta")  
cat("Tamaño de la secuencia de USA: ",length(USA[[1]]))

## Tamaño de la secuencia de USA: 29882

CHINA = read.fasta("CHINA.fasta")  
cat("Tamaño de la secuencia de CHINA: ",length(CHINA[[1]]))

## Tamaño de la secuencia de CHINA: 29903

INDIA = read.fasta("INDIA.fasta")  
cat("Tamaño de la secuencia de INDIA: ",length(INDIA[[1]]))

## Tamaño de la secuencia de INDIA: 29854

FRANCIA = read.fasta("FRANCIA.fasta")  
cat("Tamaño de la secuencia de FRANCIA: ",length(FRANCIA[[1]]))

## Tamaño de la secuencia de FRANCIA: 29882

ALEMANIA = read.fasta("ALEMANIA.fasta")  
cat("Tamaño de la secuencia de ALEMANIA: ",length(ALEMANIA[[1]]))

## Tamaño de la secuencia de ALEMANIA: 29782

BRAZIL = read.fasta("BRAZIL.fasta")  
cat("Tamaño de la secuencia de BRAZIL: ",length(BRAZIL[[1]]))

## Tamaño de la secuencia de BRAZIL: 29876

JAPON = read.fasta("JAPON.fasta")  
cat("Tamaño de la secuencia de JAPON: ", length(JAPON[[1]]))

## Tamaño de la secuencia de JAPON: 29903

S\_KOREA = read.fasta("S\_KOREA.fasta")  
cat("Tamaño de la secuencia de KOREA: ", length(S\_KOREA[[1]]))

## Tamaño de la secuencia de KOREA: 29903

ITALIA = read.fasta("ITALIA.fasta")  
cat("Tamaño de la secuencia de ITALIA: ", length(ITALIA[[1]]))

## Tamaño de la secuencia de ITALIA: 29867

UK = read.fasta("UK.fasta")  
cat("Tamaño de la secuencia del REINO UNIDO: ", length(UK[[1]]))

## Tamaño de la secuencia del REINO UNIDO: 29865

RUSSIA = read.fasta("RUSSIA.fasta")  
cat("Tamaño de la secuencia de RUSSIA: ", length(RUSSIA[[1]]))

## Tamaño de la secuencia de RUSSIA: 29544

TURQUIA = read.fasta("TURQUIA.fasta")  
cat("Tamaño de la secuencia de TURQUIA: ", length(TURQUIA[[1]]))

## Tamaño de la secuencia de TURQUIA: 29832

ESPANA = read.fasta("ESPANA.fasta")  
cat("Tamaño de la secuencia de ESPÁÑA: ", length(ESPANA[[1]]))

## Tamaño de la secuencia de ESPÁÑA: 29611

VIETNAM = read.fasta("VIETNAM.fasta")  
cat("Tamaño de la secuencia de VIET NAM: ", length(VIETNAM[[1]]))

## Tamaño de la secuencia de VIET NAM: 29891

AUSTRALIA = read.fasta("AUSTRALIA.fasta")  
cat("Tamaño de la secuencia de AUSTRALIA: ", length(AUSTRALIA[[1]]))

## Tamaño de la secuencia de AUSTRALIA: 29893

ARGENTINA = read.fasta("ARGENTINA.fasta")  
cat("Tamaño de la secuencia de ARGENTINA: ", length(ARGENTINA[[1]]))

## Tamaño de la secuencia de ARGENTINA: 29723

NETHERLANDS = read.fasta("NETHERLANDS.fasta")  
cat("Tamaño de la secuencia de PAISES BAJOS: ", length(NETHERLANDS[[1]]))

## Tamaño de la secuencia de PAISES BAJOS: 29880

IRAN = read.fasta("IRAN.fasta")  
cat("Tamaño de la secuencia de IRAN: ", length(IRAN[[1]]))

## Tamaño de la secuencia de IRAN: 29822

MEXICO = read.fasta("MEXICO.fasta")  
cat("Tamaño de la secuencia de MEXICO: ", length(MEXICO[[1]]))

## Tamaño de la secuencia de MEXICO: 29409

INDONESIA = read.fasta("INDONESIA.fasta")  
cat("Tamaño de la secuencia de INDONESIA: ", length(INDONESIA[[1]]))

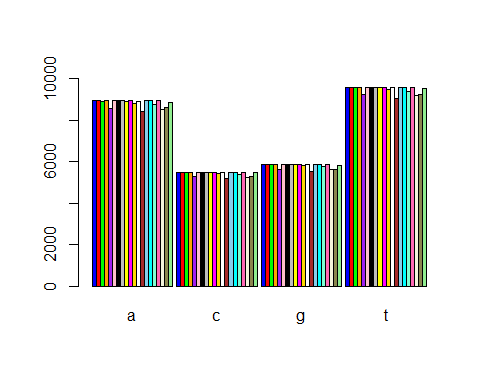
## Tamaño de la secuencia de INDONESIA: 29709

1. Grafica de TODAS las variantes y sus bases de ADN Al ser 20, es dificil integrar los nombres o las labels dentro de la misma grafica.

USA - blue China - red India - green Francia - orange Alemania - purple Brazil - pink Japon - black Corea del Sur - gray Italia - yellow Gran Bretaña - magenta Russia - gold Turquia - white España - brown Viet Nam - skyblue Australia - cyan Argentina - aquamarine Paises Bajos - hotpink Iran - beige Mexico - khaki4 Indonesia - palegreen2

Aqui esta un link de pdf si tiene algun inconveniente identificando colores <http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>

usa\_count <- count(USA[[1]], 1)  
china\_count <- count(CHINA[[1]], 1)  
india\_count <- count(INDIA[[1]], 1)  
francia\_count <- count(FRANCIA[[1]], 1)  
alemania\_count <- count(ALEMANIA[[1]], 1)  
brazil\_count <- count(BRAZIL[[1]], 1)  
japon\_count <- count(JAPON[[1]], 1)  
corea\_count <- count(S\_KOREA[[1]], 1)  
italia\_count <- count(ITALIA[[1]], 1)  
uk\_count <- count(UK[[1]], 1)  
russia\_count <- count(RUSSIA[[1]], 1)  
turquia\_count <- count(TURQUIA[[1]], 1)  
espana\_count <- count(ESPANA[[1]], 1)  
vietnam\_count <- count(VIETNAM[[1]], 1)  
australia\_count <- count(AUSTRALIA[[1]], 1)  
argentina\_count <- count(ARGENTINA[[1]], 1)  
netherlands\_count <- count(NETHERLANDS[[1]], 1)  
iran\_count <- count(IRAN[[1]], 1)  
mexico\_count <- count(MEXICO[[1]], 1)  
indonesia\_count <- count(INDONESIA[[1]], 1)  
  
counts\_matrix <- rbind(usa\_count, china\_count, india\_count, francia\_count, alemania\_count, brazil\_count, japon\_count, corea\_count, italia\_count, uk\_count, russia\_count, turquia\_count, espana\_count, vietnam\_count, australia\_count, argentina\_count, netherlands\_count, iran\_count, mexico\_count, indonesia\_count )  
barplot(counts\_matrix, beside = TRUE, col = c("blue", "red", "green", "orange", "purple", "pink", "black", "grey", "yellow", "magenta","gold","white","brown","skyblue","cyan","aquamarine","hotpink","beige","khaki4","palegreen2"),  
 ylim = c(0, 10000))



1. Concentraremos en un archivo todas las secuencias:

write.dna(virus\_sequences, file ="virus\_seqs.fasta", format = "fasta", append =  
FALSE, nbcol = 6, colsep = " ", colw = 10)

1. Cargamos la secuencias:

virus\_seq\_no\_alineadas <- readDNAStringSet("virus\_seqs.fasta", format = "fasta")

## Warning in .Call2("fasta\_index", filexp\_list, nrec, skip, seek.first.rec, :  
## reading FASTA file virus\_seqs.fasta: ignored 49669 invalid one-letter sequence  
## codes

1. Alineamiento de las secuencias:

virus\_seq\_no\_alineadas <- OrientNucleotides(virus\_seq\_no\_alineadas)

## ========================================================================================================================================================================================================  
##   
## Time difference of 0.19 secs

virus\_align\_seqs <- AlignSeqs(virus\_seq\_no\_alineadas)

## Determining distance matrix based on shared 11-mers:  
## ================================================================================  
##   
## Time difference of 0.06 secs  
##   
## Clustering into groups by similarity:  
## ================================================================================  
##   
## Time difference of 0 secs  
##   
## Aligning Sequences:  
## ================================================================================  
##   
## Time difference of 0.56 secs  
##   
## Iteration 1 of 2:  
##   
## Determining distance matrix based on alignment:  
## ================================================================================  
##   
## Time difference of 0.01 secs  
##   
## Reclustering into groups by similarity:  
## ================================================================================  
##   
## Time difference of 0 secs  
##   
## Realigning Sequences:  
## ================================================================================  
##   
## Time difference of 0.91 secs  
##   
## Alignment converged - skipping remaining iteration.

1. Guardar el resultado:

writeXStringSet(virus\_align\_seqs, file = "virus\_align\_seq.fasta")

1. Obtener el nuevo archivo:

virus\_aligned <- read.alignment("virus\_align\_seq.fasta", format = "fasta")

1. Crear una matriz de distancia:

matriz\_distancia <- dist.alignment(virus\_aligned, matrix = "similarity")

1. Visualiza la matriz de distancia: donde sombras más oscuras de gris significan una mayor distancia

temp <- as.data.frame(as.matrix(matriz\_distancia))

1. Creación del árbol con el paquete ape:

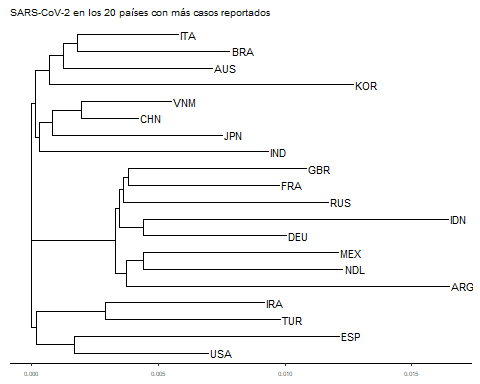
virus\_filogenetico <- nj(matriz\_distancia)  
class(virus\_filogenetico)

## [1] "phylo"

new\_names = c("USA","CHN","IND","FRA","DEU","BRA","JPN","KOR","ITA","GBR","RUS","TUR","ESP","VNM","AUS","ARG","NDL","IRA","MEX","IDN")  
virus\_filogenetico$tip.label = new\_names

1. Plot del árbol:

virus\_plot\_filogenetico <- ladderize(virus\_filogenetico)  
ggtree(virus\_plot\_filogenetico) +  
 geom\_tiplab(size = 3) +  
 theme\_tree2() +  
 theme(legend.position = "none", text = element\_text(size = 6)) +  
 ggtitle("SARS-CoV-2 en los 20 países con más casos reportados")



1. Referencias

citation("Biostrings")

##   
## To cite package 'Biostrings' in publications use:  
##   
## Pagès H, Aboyoun P, Gentleman R, DebRoy S (2022). \_Biostrings:  
## Efficient manipulation of biological strings\_. R package version  
## 2.66.0, <https://bioconductor.org/packages/Biostrings>.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {Biostrings: Efficient manipulation of biological strings},  
## author = {H. Pagès and P. Aboyoun and R. Gentleman and S. DebRoy},  
## year = {2022},  
## note = {R package version 2.66.0},  
## url = {https://bioconductor.org/packages/Biostrings},  
## }  
##   
## ATTENTION: This citation information has been auto-generated from the  
## package DESCRIPTION file and may need manual editing, see  
## 'help("citation")'.

citation("adegenet")

##   
## To cite the adegenet package:  
##   
## Jombart, T. (2008) adegenet: a R package for the multivariate  
## analysis of genetic markers. Bioinformatics 24: 1403-1405. doi:  
## 10.1093/bioinformatics/btn129  
##   
## Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the  
## analysis of genome-wide SNP data. Bioinformatics. doi:  
## 10.1093/bioinformatics/btr521  
##   
## To see these entries in BibTeX format, use 'print(<citation>,  
## bibtex=TRUE)', 'toBibtex(.)', or set  
## 'options(citation.bibtex.max=999)'.

citation("ape")

##   
## To cite ape in a publication please use:  
##   
## Paradis E, Schliep K (2019). "ape 5.0: an environment for modern  
## phylogenetics and evolutionary analyses in R." \_Bioinformatics\_,  
## \*35\*, 526-528. doi:10.1093/bioinformatics/bty633  
## <https://doi.org/10.1093/bioinformatics/bty633>.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Article{,  
## title = {ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}},  
## author = {Emmanuel Paradis and Klaus Schliep},  
## journal = {Bioinformatics},  
## year = {2019},  
## volume = {35},  
## pages = {526-528},  
## doi = {10.1093/bioinformatics/bty633},  
## }  
##   
## ape is evolving quickly, so you may want to cite its version number  
## (found with 'library(help = ape)' or 'packageVersion("ape")').

citation("ggtree")

##   
## To cite ggtree in publications use:  
##   
## Guangchuang Yu. (2022). Data Integration, Manipulation and  
## Visualization of Phylogenetic Trees (1st edition). Chapman and  
## Hall/CRC.  
##   
## Guangchuang Yu. Using ggtree to visualize data on tree-like  
## structures. Current Protocols in Bioinformatics, 2020, 69:e96. doi:  
## 10.1002/cpbi.96  
##   
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods  
## for mapping and visualizing associated data on phylogeny using  
## ggtree. Molecular Biology and Evolution 2018, 35(2):3041-3043. doi:  
## 10.1093/molbev/msy194  
##   
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk  
## Lam. ggtree: an R package for visualization and annotation of  
## phylogenetic trees with their covariates and other associated data.  
## Methods in Ecology and Evolution 2017, 8(1):28-36.  
## doi:10.1111/2041-210X.12628  
##   
## To see these entries in BibTeX format, use 'print(<citation>,  
## bibtex=TRUE)', 'toBibtex(.)', or set  
## 'options(citation.bibtex.max=999)'.

citation("DECIPHER")

##   
## To cite package 'DECIPHER' in publications use:  
##   
## Wright ES (2016). "Using DECIPHER v2.0 to Analyze Big Biological  
## Sequence Data in R." \_The R Journal\_, \*8\*(1), 352-359.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Article{,  
## title = {Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R},  
## author = {Erik S. Wright},  
## journal = {The R Journal},  
## year = {2016},  
## volume = {8},  
## number = {1},  
## pages = {352-359},  
## }

citation("viridis")

##   
## To cite viridis/viridisLite in publications use:  
##   
## Simon Garnier, Noam Ross, Robert Rudis, Antônio P. Camargo, Marco  
## Sciaini, and Cédric Scherer (2023). viridis(Lite) -  
## Colorblind-Friendly Color Maps for R. viridis package version 0.6.3.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {{viridis(Lite)} - Colorblind-Friendly Color Maps for R},  
## author = {{Garnier} and {Simon} and {Ross} and {Noam} and {Rudis} and {Robert} and {Camargo} and Antônio Pedro and {Sciaini} and {Marco} and {Scherer} and {Cédric}},  
## year = {2023},  
## note = {viridis package version 0.6.3},  
## url = {https://sjmgarnier.github.io/viridis/},  
## doi = {10.5281/zenodo.4679424},  
## }

citation("ggplot2")

##   
## To cite ggplot2 in publications, please use  
##   
## H. Wickham. ggplot2: Elegant Graphics for Data Analysis.  
## Springer-Verlag New York, 2016.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Book{,  
## author = {Hadley Wickham},  
## title = {ggplot2: Elegant Graphics for Data Analysis},  
## publisher = {Springer-Verlag New York},  
## year = {2016},  
## isbn = {978-3-319-24277-4},  
## url = {https://ggplot2.tidyverse.org},  
## }

citation("seqinr")

##   
## To cite seqinr in publications use:  
##   
## Charif, D. and Lobry, J.R. (2007)  
##   
## A BibTeX entry for LaTeX users is  
##   
## @InCollection{,  
## author = {D. Charif and J.R. Lobry},  
## title = {Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis.},  
## booktitle = {Structural approaches to sequence evolution: Molecules, networks, populations},  
## year = {2007},  
## editor = {U. Bastolla and M. Porto and H.E. Roman and M. Vendruscolo},  
## series = {Biological and Medical Physics, Biomedical Engineering},  
## pages = {207-232},  
## address = {New York},  
## publisher = {Springer Verlag},  
## note = {{ISBN :} 978-3-540-35305-8},  
## }

Interpretación de las gráficas y conclusiones:

La gráfica de las bases de ADN es un poco complicada de analizar debido a la gran concentración de datos en un espacio reducido, pero funciona para comprobar las similitudes de las 20 variantes de COVID en los 20 países con más casos reportados. Todas tienen casi la misma composición, es decir, no hay ninguna sola que varíe notablemente, todas van más o menos a la par con una tendencia grande a “(A y T)”, llegando la mayoría a los 10,000, componiendo 1/3 o en este caso 2/3 del total del ADN. En promedio, las secuencias tenían una longitud de 29,000, lo que implica que este tipo de cadenas son ricas en pares adenina-timina. Lo mencioné en mi evidencia 01 y, aun con más datos y variantes, la tendencia a los grupos de adenina-timina es notable y es similar a lo largo de todas las variantes.

Podemos observar que en longitud todas las secuencias son de tamaño similar, rondando los 29,500 más o menos. Hay algunas que casi llegan a los 30 y otras más cerca de los 28, pero es normal ya que estas son mutaciones de un virus, por lo que en esencia son el mismo virus, pero con pequeños cambios de nucleótidos dentro de sus genomas. Esto no necesariamente indica que provoquen diferentes síntomas ni nada, pero tienen diferentes cualidades que a simple vista no se ven. A la hora de analizar, nos podemos dar cuenta de que sí son diferentes, aunque no sea la cosa más notable del mundo.

Esto se puede ver en el árbol filogenético ordenado por países. Podemos ver que todos son similares, aunque hay ciertos grupos que se parecen más entre ellos. Hay que tener en cuenta que estos datos son los primeros genomas completos de variantes de COVID descubiertos en cada país de los 20 de la lista. Por ejemplo, podemos observar que las variantes de Vietnam, China y Japón son bastante similares. Esto nos indica que puede que la variante de China haya mutado y se haya esparcido por Asia. Al mutar, se generaron estas variantes tanto en Vietnam como en Japón con composiciones MUY similares. Otra cosa a notar es que la variante de Corea del Sur es una de las que más diferencias presenta con respecto a las demás, y eso que está dentro del continente asiático al igual que las mencionadas previamente. Las variantes de Indonesia y Argentina también se separan notablemente de los grupos respectivos a los que pertenecen. Esto nos puede decir una variedad de cosas, pero principalmente que son los que presentan más diferencias dentro de sus genomas. Realmente es muy interesante observar un árbol filogenético. Es una visual realmente simple, solo son líneas con nombres, pero la información que puedes obtener es muchísima. Es muy interesante ver las similitudes, las diferencias y pensar y analizar más a profundidad el porqué de las mismas. Eso no es algo que vaya a realizar en esta evidencia, es algo que me gustaría realizar en el futuro. Me gustó mucho trabajar con R a lo largo de este periodo y siento que aprendí muchísimo

Referencias de la investigacion <https://covid19.who.int/table> <https://pubmed.ncbi.nlm.nih.gov/> <https://www.datosmundial.com/codigos-de-pais.php> Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7.

Wang M, Yan M, Xu H, et al. SARS-CoV infection in a restaurant from palm civet. Emerg Infect Dis. 2005 Apr;11(4):1860-5. doi: 10.3201/eid1112.041293.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., … & Chen, H. D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798), 270-273

Lam, T. T., Shum, M. H., Zhu, H. C., Tong, Y. G., Ni, X. B., Liao, Y. S., … & Cheng, S. H. (2020). Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. BioRxiv, 2020.