

# Trabajo individual

David Cardona Duque

8/20/2021

## Introducción

El presente trabajo desarrolla un análisis gráfico de coeficientes relevantes respecto a los contagios por 100000 habitantes en Colombia delimitado por variables de población, edad, regiones y periodos de tiempo, que servirán como materia prima para generar un acercamiento al análisis respecto a los efectos del déficit fiscal y el posible manejo de la pandemia por parte de los municipios teniendo en cuenta el aumento o no de contagios en las respectivas regiones; en otro punto los contagios diferenciados por sexos y edades en los periodos de tiempo donde se presentaran planteamientos con base en implicaciones diferenciadas con base en el género como las actividades y edades productivas, teniendo por ultimo en cuenta la diferencia de contagio según los periodos de tiempo que permiten un contraste entre la etapa de implementación de medidas restrictivas de aislamiento preventivo y su finalización.

## Marco Teórico

Debido a la pandemia ocasionada por el COVID- 19 los Estados a nivel mundial debieron aplicar diversas medidas de contención respecto al virus por sus implicaciones en la salud pública y desarrollo de actividades económicas. Considerando lo anterior fue necesario un registro de contagios, muertes, recuperaciones, entre otras con el fin de tomar decisiones pertinentes frente a la crisis, a partir de esto para el desarrollo de este analisis se utilizaran los datos del Instituto Nacional de Salud sobre los casos positivos en Colombia separados principalmente por sexo, edad, fecha y municipio donde se consideran los contagios según el Ministerio de Salud (2020) “a través de pruebas moleculares; sin embargo, primero se descarta o confirma la presencia de los virus más frecuentes así como el antecedente de viaje del paciente a países del exterior de Colombia o su contacto con un caso confirmado” (p.2). También servirán de referencia los datos del Índice de Desempeño Fiscal (IDF) entendido por el Departamento Nacional de Planeación (s.f) como “una medición del desempeño de la gestión financiera de las entidades territoriales que da cuenta de su sostenibilidad financiera”, que complementará el análisis en los periodos de tiempo especificados para el desarrollo del trabajo.

Entendiendo la pertinencia del análisis de los contagios con base en la pandemia se hace beneficioso el uso del lenguaje de programación R en especial por su enfoque estadístico óptimo para el análisis de datos, en conjunto los paquetes de R que contienen funciones útiles con códigos abreviados que mejoran la eficiencia; herramientas integradas a lecturas propias con elementos de estudio exploratorios y descriptivos examinando el tema abordado.

## Objetivos

Comparar paquetes de análisis exploratorio automático.

Examinar los contagios por 100000 habitantes en dos regiones de Colombia (Caribe y Pacífica) con base en periodos de tiempo, sexo y edades.

Observar desde una perspectiva exploratoria y descriptiva los resultados obtenidos.

Presentar hipótesis de acercamiento a los análisis realizados.

## Asignación correspondiente de base de datos de acuerdo al número de documento

Documento: 100.11.333.24

### Regiones:

Región Caribe - Región Pacífica. Último número 4

### Rangos de edad.

17-23 vs. 24-30. Penúltimo número 2

### Periodos

Periodo 5: 25/mayo/2021 - 18/agosto/2021 vs. Periodo 1: 2/marzo/2020 - 22/sep/2020. Antepenúltimo número 3

### Paquetes del R

4.funModeling Vs.5.arsenal. Tras antepenúltimo número 3

## Depuraciones

### Sobre la base de datos “covid\_19”

Se hace un cambio al formato RData para facilitar el trabajo con la misma.

Se cambia el tipo de variable de “Fecha.de.notificación” a fecha puesto que se acerca más a la naturaleza real de esta variable.

Se crea una nueva variable con base a la variable “Edad” puesto que los valores de esta dependían de si se expresaba en meses, días, etc. Se generaliza a años para un mayor entendimiento de la misma.

Se crean rangos de edades de [17,23] y de (23,30], además que se agrupan las edades que entran en estos rangos, puesto que son la población objetiva de este análisis. Se guardan en una nueva variable llamada “edad\_gr\_et”

Se hace un cambio en el tipo de la variable “Código.DIVIPOLA.municipio” a carácter puesto que es el tipo real de esta variable, además se le guarda como “Codigo\_munc” para un mejor entendimiento y acceso a futuro.

Se crea una nueva base de datos llamada “covid\_19\_contagios” donde se seleccionan las variables “ID.de.caso, Sexo, Edad\_R, edad\_gr\_et, Codigo\_munc, Fecha.de.notificación”

### **Sobre la base “covid\_19\_contagios”**

Se cambia el formato de la variable sexo, se pasa a llamar genero y las personas F pasan a ser mujeres, por otro lado las M pasan a ser Hombres.

Se cambia el tipo de la variable “mes” a date puesto que es el tipo real de la misma

Se agrupan las variables “mes, Codigo\_munc, edad\_gr\_et, genero” y se hace una suma de los contagios guardándose en la variable “contagios”.

Todo lo anterior se guarda en una nueva base de datos llamada “covid\_19\_contagios\_x\_mes”

### **Sobre “covid\_19\_contagios\_x\_mes”**

Se cambia el tipo de la variable “Codigo\_munc” a numerico para una estandarizacion de la misma, más específicamente la eliminación del cero a la izquierda en el código para una posterior unión con otra base de datos.

Se une la base de datos con “pobla\_total\_munc\_x\_gr\_et” para un análisis más completo a posteriori, puesto que tienen información útil, como por ejemplo, el número de la población estimada por grupo de edad, género y periodo.

La anterior unión se guarda en “covid\_19\_contagios\_x\_mes”.

Se crea una nueva variable “contagios\_x\_cien\_mil” que tiene en cuenta los contagios por 100000 en cuanto a población para un análisis posterior en el impacto real de los contagios en el territorio, puesto que si un territorio tiene muchos habitantes es probable que tenga más contagios y no es necesariamente grave vs. un territorio con menores contagios pero con una mayor tasa por 100000 de los mismos.

## **Sobre la base de datos “pobla\_munc”**

Se seleccionan las variables de interés para el análisis posterior, excluyendo “Total, Total Hombres, Total Mujeres” y se guarda en una nueva base de datos pobla\_munc1.

### **Sobre la base pobla\_munc1.**

Se transforman las variables “Hombre\_x” a una manera larga almacenándolas en la variable genero, luego se crea la variable edad con la parte luego del guion bajo, con la intención de un mayor entendimiento de las mismas y un mejor uso a futuro. Todo lo anterior se guarda en una nueva base llamada “pobla\_munc\_largo”.

### **Sobre la base de datos “pobla\_munc\_largo”.**

Se filtra la base de datos de acuerdo a que la variable género sea diferente de total y el ÁREA GEOGRÁFICA sea la completa, esto con el objetivo de tener una información completa a la hora de hacer análisis y no se generen errores al tomar la variable género.

Se seleccionan las variables “DPMP, MPIO, genero, edad, poblacion” y se hace un cambio de tipo a la variable edad a entero con el objetivo de un mejor uso de la misma, los valores de “100 y mas” se cambian solamente a 100 para realizar este cambio, puesto que esta no es la población objetiva de este análisis.

Se crean rangos de edades de [17,23] y de (23,30], además que se agrupan las edades que entran en estos rangos, puesto que son la población objetiva de este análisis. Se guardan en una nueva variable llamada “edad\_gr\_et”

Todo lo anterior se guarda en una nueva base llamada “pobla\_total\_munc\_2020”.

### **Sobre “pobla\_total\_munc\_2020”**

Se agrupan las variables “edad\_gr\_et, DPMP, MPIO, genero” puesto que son las variables de interés y se crea una variable “poblacion” que contenga la totalidad de la población por cada grupo pasado.

Lo anterior se guarda en una nueva base llamada “pobla\_total\_munc\_x\_gr\_et”

### **Sobre “pobla\_total\_munc\_x\_gr\_et”**

Se cambia el tipo de la variable “DPMP” a numérico para una estandarización de la misma, más específicamente la eliminación del cero a la izquierda en el código para una posterior unión con otra base de datos.

Se une la base de datos con “covid\_19\_contagios\_x\_mes” para un análisis más completo a posteriori, puesto que tienen información útil, como por ejemplo, el número de contagios estimados por grupo de edad, género y periodo.

La anterior unión se guarda en “covid\_19\_contagios\_x\_mes”.

## **Sobre la base de datos “regiones”**

Se cambia el tipo de la variable “CÓDIGO.DANE.DEL.MUNICIPIO” a numérico para una estandarización de la misma, más específicamente la eliminación del cero a la izquierda en el código para una posterior unión con otra base de datos.

Se une la base de datos con “covid\_19\_contagios\_x\_mes” para un análisis más completo a posteriori, puesto que tienen información útil, como por ejemplo, la región a la que pertenece cada municipio.

La unión se convierte en una nueva base de datos llamada “covid\_19\_contagios\_x\_mes\_regiones”.

## **Sobre la base de datos “fiscal”**

Se seleccionan las variables “Código, Indicador de desempeño Fiscal 7/” puesto que son las variables de interés en este análisis. Se guarda en una nueva base de datos llamada “fiscal1”

### **Sobre “fiscal1”**

Se cambia el tipo de la variable “Código” a numérico para una estandarización de la misma, más específicamente la eliminación del cero a la izquierda en el código para una posterior unión con otra base de datos.

Se cambia el nombre de la variable “Indicador de desempeño Fiscal 7/” a “IDF” para un mayor entendimiento de la misma en

análisis futuros.

Se seleccionan las variables “Código,IDF” puesto que son las variables de interés para este análisis. Se guardan en una nueva base de datos llamada “fiscal2”.

### Sobre “fiscal2”

Se une la base de datos con “covid\_19\_contagios\_x\_mes\_regiones” para un análisis más completo a posteriori, puesto que tienen información útil, como por ejemplo, el indicador de desempeño fiscal (IDF) de cada municipio. Se guarda en una nueva base llamada “covid\_19\_contagios\_x\_mes\_regiones\_IDF”.

### Sobre la base de datos “covid\_19\_contagios\_x\_mes\_regiones\_IDF”

Se cambia el tipo de la variable “Codigo\_munc” a carácter puesto que es el tipo real de la misma y ya se hicieron las operaciones que necesitaban su uso como numérico.

Se seleccionan todas las variables excluyendo “CÓDIGO.DANE.DEL.DEPARTAMENTO,MUNICIPIO” puesto que ya están repetidas con otro nombre dentro de la base de datos. Se guarda en una nueva base llamada “base\_final\_sin\_filtrar\_rangos\_contagios”

### Sobre la base de datos “base\_final\_sin\_filtrar\_rangos\_contagios”

Se filtra la variable “edad\_gr\_et” para todos los datos que no sean NA, puesto que estos grupos de edad son los que nos interesan en este análisis. Se guarda en una nueva base llamada “base\_final\_filtrada\_rangos\_contagios”.

### Sobre “base\_final\_filtrada\_rangos\_contagios”

Se cambia el tipo de la variable “mes” a fecha puesto que es el tipo real de esta variable.

Se filtra la variable “mes” entre las fechas 2020-03-02 y 2020-09-22 o 2021-05-25 en adelante, puesto que son los intervalos de tiempo de interés en este análisis.

Se cambia el nombre de la variable “mes” por “Fecha\_de\_notificacion” puesto que este nombre se acerca más al origen de la variable, ya que, si se observa esta no solo contiene el mes si no además el día, también muestra de mejor manera de donde proviene y cuál es su significado para análisis futuros.

Se filtra la variable “REGION” en las regiones “Región Pacífico, Región Caribe” puesto que son las regiones de interés en este análisis.

## Paquetes automáticos de análisis exploratorio y descriptivo

### Arsenal

Algunas funciones de arsenal son:

#### “mdy.Date”

Convierte fechas numéricas a objetos del tipo Date, y viceversa.

```
## [1] "2021-08-08"
```

```
## $month
## [1] 8
##
## $day
## [1] 8
##
## $year
## [1] 2021
```

#### “comparedf”

Compara dos data.frames y reporta cualquier diferencia entre ellos.

```
##
##
## Table: Summary of data.frames
##
## version  arg      ncol  nrow
## -----  -
## x      fiscal     17   1101
## y      regiones    5   1123
##
##
## Table: Summary of overall comparison
##
## statistic                                value
## -----
## Number of by-variables                    1
## Number of non-by variables in common      0
## Number of variables compared              0
## Number of variables in x but not y        16
## Number of variables in y but not x         4
## Number of variables compared with some values unequal  0
## Number of variables compared with all values equal    0
```

```

## Number of observations in common      851
## Number of observations in x but not y    400
## Number of observations in y but not x    422
## Number of observations with some compared variables unequal    0
## Number of observations with all compared variables equal      851
## Number of values unequal              0
##
##
##
## Table: Variables not shared
##
## version  variable                                position class
## -----
## x      Código                                1 character
## x      Departamento                          2 character
## x      1/ Autofinanciación de los gastos de funcionamiento      4 numeric
## x      2/ Respaldo del servicio de la deuda                    5 numeric
## x      3/ Dependencia de las transferencias de la Nación y las Regalías 6 numeric
## x      4/ Generación de recursos propios                      7 numeric
## x      5/ Magnitud de la inversión                          8 numeric
## x      6/ Capacidad de ahorro                            9 numeric
## x      Indicador de desempeño Fiscal 7/                     10 numeric
## x      Posición 2019 a nivel Nacional                      11 numeric
## x      Rango Clasificación                                12 character
## x      Categoría Ley 617 de 2000                          13 character
## x      Sistema de Ciudades                                14 numeric
## x      Area Metropolitana                                15 character
## x      Ciudades Capitales                                16 character
## x      Región PND 2018-2022                               17 character
## y      REGION                                             1 character
## y      CÓDIGO.DANE.DEL.DEPARTAMENTO                      2 integer
## y      DEPARTAMENTO                                       3 character
## y      CÓDIGO.DANE.DEL.MUNICIPIO                         4 integer
##
##
##
## Table: Other variables not compared
##
##
## -----
## No other variables not compared
## -----
##
##
##
## Table: Observations not shared
##
## version  Municipio                observation
## -----
## x      Abriaqui                    770
## x      Acandi                      782
## x      Achi                        904
## x      Agua De Dios                162
## x      Agustin Codazzi             1079
## x      Alban                       211
## x      Alban                       1031
## x      Alcala                      603
## x      Alejandria                  642
## x      Alto Baudó (Pie De Pato)     689
## x      Altos Del Rosario            1094
## x      Amaga                       152
## x      Ancuya                      1036
## x      Andalucia                   348
## x      Angelopolis                 349
## x      Anori                       584
## x      Antioquia                   283
## x      Anzoategui                  1027
## x      Apartado                    197
## x      Apia                        493
## x      Ariguani                    615
## x      Armero (Guayabal)            229
## x      Bagado                      485
## x      Bahia Solano (Mutis)         517
## x      Bajo Baudó (Pizarro)         130
## x      Barranca De Upia             14
## x      Barranco De Loba             1084
## x      Belalcazar                  217
## x      Belen                       892
## x      Belen                       692
## x      Belen Andaquies              818
## x      Belen De Umbria              401
## x      Beltran                     304

```

## x	Beteitiva	840
## x	Bogota D.C.	23
## x	Bojaca	143
## x	Bojaya (Bellavista)	958
## x	Bolivar	729
## x	Bolivar	464
## x	Bolivar	318
## x	Bolivar	798
## x	Boyaca	450
## x	Buga	441
## x	Buritica	188
## x	Busbanza	643
## x	Caceres	1091
## x	Cachira	788
## x	Cacota	1019
## x	Cajibio	515
## x	Cajica	8
## x	Calarca	321
## x	Calima-Darien	439
## x	Campo De La Cruz	1034
## x	Canton De San Pablo	1090
## x	Caparrapi	533
## x	Caracoli	691
## x	Carcasi	967
## x	Carmen De Apicala	281
## x	Carmen De Carupa	407
## x	Carmen De Viboral	124
## x	Carmen Del Darien	165
## x	Cartagena Del Chaira	347
## x	Casabianca**	927
## x	Castilla La Nueva	49
## x	Cepita	866
## x	Cerete	274
## x	Certegui	968
## x	Chachagui	362
## x	Chaguani	856
## x	Chalan	1061
## x	Chameza	482
## x	Charala	778
## x	Chia	19
## x	Chigorodo	350
## x	Chima	1087
## x	Chima	830
## x	Chinacota	475
## x	Chinchina	381
## x	Chinu	310
## x	Chipata	986
## x	Chiquinquira	221
## x	Chiquiza	820
## x	Chiriguana	929
## x	Chitaga	632
## x	Chivata	863
## x	Choachi	107
## x	Choconta	91
## x	Cienaga	335
## x	Ciénaga De Oro	109
## x	Cienega	846
## x	Cocorna	185
## x	Colon	541
## x	Colon-Genova	902
## x	Combita	591
## x	Concepcion	572
## x	Concepcion	926
## x	Contratacion	727
## x	Convencion	825
## x	Cordoba	634
## x	Cordoba	656
## x	Cordoba	668
## x	Covarachia	905
## x	Cuaspué-Carlosama	993
## x	Cubara	748
## x	Cucunuba	213
## x	Cucuta	156
## x	Cuitiva	894
## x	Curiti	402
## x	Curumani	425
## x	Dibulla	262
## x	Distraccion	952
## x	Don Matias	196
## x	Ebejico	857
## x	El Aguila	915
## x	El Carmen De Bolivar	556

## x	El Peñon	478
## x	El Peñon	1007
## x	El Peñon	912
## x	El Playon	879
## x	El Reten	673
## x	El Tablon	801
## x	Elias	708
## x	Facatativa	175
## x	Florian	46
## x	Fuente De Oro	665
## x	Fundacion	412
## x	Fuquene	148
## x	Fusagasuga	140
## x	Gachantiva	599
## x	Gacheta	535
## x	Galan	384
## x	Gámbita	961
## x	Garzon	451
## x	Genova	864
## x	Gomez Plata	504
## x	Gonzalez	900
## x	Guacari	422
## x	Guacheta	131
## x	Gualmatan	937
## x	Guapota	636
## x	Guatape	86
## x	Guataqui	837
## x	Guatica	721
## x	Guavata	685
## x	Guayabal De Siquima	557
## x	Guayata	794
## x	Guepsa	619
## x	Guican	753
## x	Gutierrez	985
## x	Hacari	728
## x	Hatillo De Loba	1101
## x	Herran	869
## x	Ibague	184
## x	Imues	921
## x	Inza	919
## x	Jambalo	458
## x	Jamundi	253
## x	Jardin	440
## x	Jerico	183
## x	Jerico	907
## x	Jerusalen	461
## x	Jesus Maria	57
## x	Jordan	686
## x	Juan De Acosta	428
## x	Junin	822
## x	La Jagua De Ibirico	138
## x	La Jagua Del Pilar	263
## x	La Union	51
## x	La Union	624
## x	La Union	489
## x	La Union	526
## x	La Uribe	827
## x	Landazuri	416
## x	Lebrija	81
## x	Lejanias	756
## x	Lerida	358
## x	Libano	291
## x	Litoral Del San Juan	29
## x	Lloro	956
## x	Lopez	102
## x	Los Cordobas	501
## x	Magangue	265
## x	Magui-Payan	831
## x	Malaga	363
## x	Manati	723
## x	Mani	155
## x	Mapiripan	751
## x	María La Baja	625
## x	Maripi	590
## x	Medellin	48
## x	Medio Baudó	90
## x	Milan	890
## x	Mistrato	581
## x	Mitu	332
## x	Mompos	920
## x	Mongui	844
## x	Moniquira	129

## x	Montelibano	239
## x	Monteria	258
## x	Murindo	1026
## x	Mutata	190
## x	Nataga	931
## x	Nechi	1062
## x	Necocli	245
## x	Nemocon	144
## x	Norosi	1069
## x	Novita	317
## x	Nuevo Colon	352
## x	Nunchia	435
## x	Nuqui	815
## x	Oicata	750
## x	Orocue	59
## x	Pacora	537
## x	Paez	40
## x	Paez	206
## x	Palmar De Varela	696
## x	Palmas Del Socorro	891
## x	Paramo	671
## x	Patia (El Bordo)	595
## x	Paz De Ariporo	177
## x	Paz Del Rio	295
## x	Piendamo	511
## x	Pijiño Del Carmen	843
## x	Piojo	545
## x	Polo Nuevo	949
## x	Popayan	108
## x	Potosi	936
## x	Puebloviejo	651
## x	Puerto Asis	724
## x	Puerto Berrio	181
## x	Puerto Boyaca	70
## x	Puerto Caycedo	506
## x	Puerto Gaitan	71
## x	Puerto Guzman	1049
## x	Puerto Inirida	261
## x	Puerto Leguizamo	776
## x	Puerto Lopez	21
## x	Puerto Rondon	562
## x	Puli	771
## x	Purace	910
## x	Purificacion	446
## x	Purisima	466
## x	Quibdo	193
## x	Quinchia	377
## x	Quipama	694
## x	Ramiriqui	216
## x	Raquira	546
## x	Repelon	1045
## x	Rio De Oro	380
## x	Rio Iro	979
## x	Rio Quito	704
## x	Rio Viejo	1054
## x	Rioblanco	828
## x	Riofrio	311
## x	Roberto Payan	951
## x	Robles (La Paz)	339
## x	Rondon	1051
## x	Sabana De Torres	218
## x	Sabanas De San Angel	1013
## x	Saboya	514
## x	Sacama	965
## x	Sachica	67
## x	Sahagun	36
## x	Samaca	39
## x	Samana	592
## x	Sampues	341
## x	San Antonio Del Tequendama	393
## x	San Vicente Del Caguan	275
## x	San Agustin	471
## x	San Andres	389
## x	San Andres	523
## x	San Andres Sotavento	408
## x	San Bernardo Viento	351
## x	San Carlos De Guaroa	215
## x	San Cristobal	1064
## x	San Jacinto Del Cauca	1017
## x	San Jeronimo	139
## x	San Joaquin	971
## x	San Jose	503

## x	San José de Fragua	792
## x	San José de La Montaña	1041
## x	San Jose De Miranda	570
## x	San Jose De Pare	396
## x	San Jose De Ure	336
## x	San Jose Del Guaviare	432
## x	San Jose Del Palmar	738
## x	San Juan De Arama	611
## x	San Juan De Betulia	817
## x	San Juan De Nepomuceno	566
## x	San Juan De Rioseco	567
## x	San Juan De Uraba	375
## x	San Juan Del Cesar	273
## x	San Luis De Gaceno	521
## x	San Luis De Palenque	205
## x	San Martin	373
## x	San Martin	112
## x	San Martin De Loba	1083
## x	San Miguel De Sema	608
## x	San Pablo De Borbur	588
## x	San Pedro De Cartago	875
## x	San Pedro De Uraba	552
## x	San Sebastian	1035
## x	San Sebastian De Buenavista	987
## x	San Vicente De Chucuri	161
## x	San Zenon	1060
## x	Sandona	655
## x	Santa Barbara	513
## x	Santa Barbara	859
## x	Santa Barbara	884
## x	Santa Barbara De Pinto	1029
## x	Santa Helena	953
## x	Santa Lucia	1056
## x	Santa Maria	192
## x	Santa Maria	700
## x	Santa Rosa De Cabal	234
## x	Santa Rosa De Osos	158
## x	Santa Rosa De Viterbo	510
## x	Santa Rosa Del Sur	415
## x	Santa Rosalia	870
## x	Santa Sofia	20
## x	Santander De Quilichao	391
## x	Santo Tomas	228
## x	Sesquile	11
## x	Sibate	63
## x	Simiti	490
## x	Since	167
## x	Sipi	42
## x	Soata	718
## x	Socota	789
## x	Sonson	89
## x	Sopetran	126
## x	Sopo	7
## x	Soraca	457
## x	Sotaquira	209
## x	Suarez	507
## x	Suarez	835
## x	Supata	564
## x	Supia	419
## x	Surata	621
## x	Susacon	1025
## x	Sutamarchan	83
## x	Tado	834
## x	Tamara	944
## x	Tamesis	242
## x	Taraza	1098
## x	Tibana	433
## x	Timana	620
## x	Timbio	806
## x	Timbiqui	359
## x	Tinjaca	780
## x	Titiribi	392
## x	Tocancipa	5
## x	Togui	850
## x	Tolu	1067
## x	Toluviejo	664
## x	Topaga	755
## x	Topaipi	682
## x	Totoro	338
## x	Tubara	235
## x	Tuchin	607
## x	Tulu	141



## x	Tumaco	659	
## x	Tunungua	448	
## x	Tuquerres	637	
## x	Turbana	560	
## x	Turmeque	784	
## x	Tutasa	658	
## x	Ubala	492	
## x	Ubate	125	
## x	Unguia	1100	
## x	Union Panamericana		1093
## x	Usiacuri	786	
## x	Utica	400	
## x	Valle De S Juan	888	
## x	Valle Del Guamuez		244
## x	Valle San Jose	539	
## x	Valparaiso	479	
## x	Valparaiso	862	
## x	Vegachi	237	
## x	Velez	813	
## x	Venecia (Ospina Perez)		861
## x	Viani	631	
## x	Vigia Del Fuerte	735	
## x	Villa De Leyva	31	
## x	Villa Del Rosario	146	
## x	Villacaro	758	
## x	Villagarzon	360	
## x	Villagomez	745	
## x	Villamaria	69	
## x	Villapinzon	225	
## x	Viota	468	
## x	Viracacha	995	
## x	Yacopi	540	
## x	Yaguara	312	
## x	Yali	616	
## x	Yolombo	272	
## x	Yondo	93	
## x	Zapayan	1066	
## x	Zipacon	429	
## x	Zipaquira	94	
## y	Abriaquí	3	
## y	Acandí	498	
## y	Achí	138	
## y	Agua de Dios	944	
## y	Agustín Codazzi	355	
## y	Albán	1053	
## y	Albán	617	
## y	Alcalá	1036	
## y	Alejandro	4	
## y	Alto Baudó	499	
## y	Altos del Rosario	932	
## y	Amagá	5	
## y	Ancuyá	619	
## y	Andalucía	1026	
## y	Angelópolis	8	
## y	Anorí	10	
## y	Anzoátegui	1068	
## y	Apartadó	13	
## y	Apía	683	
## y	Ariguaní	571	
## y	Armero	791	
## y	Bagadó	501	
## y	Bahía Solano	502	
## y	Bajo Baudó	503	
## y	Barranca de Upía	883	
## y	Barranco de Loba	891	
## y	Barranco Minas	865	
## y	Belalcázar	282	
## y	Belén	1113	
## y	Belén	1080	
## y	Belén de Bajira	961	
## y	Belén de Los Andaquies		918
## y	Belén de Umbría	960	
## y	Beltrán	401	
## y	Betétiva	175	
## y	Bogotá D.C.	953	
## y	Bojacá	403	
## y	Bojaya	504	
## y	Bolívar	318	
## y	Bolívar	702	
## y	Bolívar	1014	
## y	Boyacá	177	
## y	Buenaventura	170	

## y	Buena Vista	179
## y	Buriticá	23
## y	Busbanzá	180
## y	Cacahual	870
## y	Cáceres	24
## y	Cachirá	1039
## y	Cácota	976
## y	Cajibío	320
## y	Cajicá	406
## y	Calarcá	1100
## y	Calima	1024
## y	Campo de La Cruz	939
## y	Caparrapí	407
## y	Caracolí	29
## y	Carcasí	705
## y	Carmen de Apicala	933
## y	Carmen de Carupa	954
## y	Carmen del Darien	1082
## y	Cartagena del Chairá	942
## y	Casabianca	1062
## y	Castilla la Nueva	1112
## y	Cepitá	706
## y	Cereté	380
## y	Cértegui	505
## y	Chachagüí	1085
## y	Chaguaní	409
## y	Chalán	772
## y	Chámeza	834
## y	Charalá	708
## y	Chía	1064
## y	Chigorodó	35
## y	Chimá	381
## y	Chimá	1049
## y	Chinácota	1019
## y	Chinchiná	283
## y	Chinú	382
## y	Chipatá	710
## y	Chiquinquirá	185
## y	Chíquiza	197
## y	Chiriguaná	360
## y	Chitagá	1023
## y	Chivatá	189
## y	Choachí	411
## y	Chocontá	412
## y	Ciénaga	1122
## y	Ciénaga de Oro	955
## y	Ciénega	88
## y	Ciudad Bolívar	21
## y	Cocomá	37
## y	Colón	621
## y	Colón	846
## y	Cómbita	190
## y	Concepción	38
## y	Concepción	712
## y	Contratación	714
## y	Convención	1004
## y	Córdoba	145
## y	Córdoba	675
## y	Córdoba	624
## y	Covarachía	193
## y	Cuaspud	625
## y	Cubará	194
## y	Cucunubá	415
## y	Cúcuta	1086
## y	Cuítiva	196
## y	Curití	716
## y	Curumaní	361
## y	Dibula	558
## y	Distracción	559
## y	Don Matías	42
## y	Ebéjico	43
## y	El Águila	1042
## y	El Cantón del San Pablo	894
## y	El Carmen de Atrato	927
## y	El Carmen de Bolívar	926
## y	El Carmen de Chucurí	958
## y	El Carmen de Viboral	959
## y	El Encanto	856
## y	El Litoral del San Juan	889
## y	El Peñón	1072
## y	El Peñón	1093
## y	El Peñón	1060

## y	El Playón	718
## y	El Retén	577
## y	El Santuario	101
## y	El Tablón de Gómez	947
## y	Elías	528
## y	Facatativá	1117
## y	Florián	721
## y	Fuente de Oro	887
## y	Fundación	578
## y	Fúquene	421
## y	Fusagasugá	1106
## y	Gachantivá	204
## y	Gachetá	424
## y	Galán	723
## y	Gambita	724
## y	Garzón	529
## y	Génova	1110
## y	Gómez Plata	51
## y	González	365
## y	Guacarí	999
## y	Guachetá	427
## y	Guadalajara de Buga	1018
## y	Gualmatán	635
## y	Guapotá	728
## y	Guatapé	55
## y	Guataquí	430
## y	Guática	686
## y	Guavatá	729
## y	Guayabal de Siquima	917
## y	Guayatá	209
## y	Güepso	730
## y	Güicán	210
## y	Gutiérrez	433
## y	Hacarí	1005
## y	Hatillo de Loba	898
## y	Herrán	1007
## y	Ibagué	1078
## y	Imués	637
## y	Inírida	864
## y	Inzá	328
## y	Jambaló	329
## y	Jamundí	1095
## y	Jardín	1094
## y	Jericó	61
## y	Jericó	213
## y	Jerusalén	434
## y	Jesús María	731
## y	Jordán	732
## y	Juan de Acosta	886
## y	Junín	435
## y	La Chorrera	857
## y	La Guadalupe	869
## y	La Jagua de Ibirico	923
## y	La Jagua del Pilar	922
## y	La Pedrera	858
## y	La Unión	65
## y	La Unión	643
## y	La Unión	776
## y	La Unión	994
## y	Landázuri	734
## y	Lebríja	736
## y	Leguízamo	850
## y	Lejanías	606
## y	Lérida	1103
## y	Líbano	1098
## y	Lloró	508
## y	López	332
## y	Los Córdoba	385
## y	Magangué	148
## y	Magüí	647
## y	Málaga	739
## y	Manatí	127
## y	Maní	1059
## y	Mapiripán	602
## y	Mapiripana	866
## y	María la Baja	1090
## y	Maripí	218
## y	Medellín	1
## y	Medio Baudó	510
## y	Milán	1066
## y	Miriti Paraná	1074
## y	Mistrató	690

## y	Mitú	873
## y	Mompós	152
## y	Monguí	221
## y	Moniquirá	222
## y	Montelíbano	1055
## y	Montería	376
## y	Morichal	872
## y	Murindó	70
## y	Mutatá	71
## y	Nátaga	537
## y	Nechí	74
## y	Necoclí	73
## y	Nemocón	448
## y	Norosí	154
## y	Nóvita	512
## y	Nuevo Colón	225
## y	Nunchía	1051
## y	Nuquí	513
## y	Oicatá	226
## y	Orocué	1097
## y	Pacoa	875
## y	Pácora	292
## y	Páez	229
## y	Páez	1077
## y	Palmar de Varela	909
## y	Palmas del Socorro	884
## y	Pana Pana	871
## y	Papunaua	877
## y	Páramo	747
## y	Patía	337
## y	Paz de Ariporo	919
## y	Paz de Río	897
## y	Piendamó	339
## y	Pijiño del Carmen	929
## y	Piojó	128
## y	Polonuevo	129
## y	Popayán	314
## y	Potosí	655
## y	Pueblo Viejo	1115
## y	Puerto Alegría	969
## y	Puerto Arica	860
## y	Puerto Asís	1056
## y	Puerto Berrío	79
## y	Puerto Boyacá	237
## y	Puerto Caicedo	848
## y	Puerto Gaitán	608
## y	Puerto Guzmán	849
## y	Puerto López	609
## y	Puerto Rondón	828
## y	Pulí	459
## y	Puracé	341
## y	Purificación	816
## y	Purísima	391
## y	Quibdó	497
## y	Quinchía	692
## y	Quípama	238
## y	Ramiriquí	239
## y	Ráquira	240
## y	Repelón	1070
## y	Rio Blanco	817
## y	Río de Oro	901
## y	Río Iro	514
## y	Río Quito	515
## y	Río Viejo	157
## y	Riofrío	1034
## y	Roberto Payán	660
## y	Rondón	241
## y	Sabana de Torres	935
## y	Sabanas de San Angel	899
## y	Saboyá	242
## y	Sácama	841
## y	Sáchica	243
## y	Sahagún	392
## y	Samacá	244
## y	Samaná	298
## y	Sampués	1050
## y	San Agustín	946
## y	San Andrés	948
## y	San Andrés	752
## y	San Andrés de Cuerquía	911
## y	San Andrés de Tumaco	1065
## y	San Andrés Sotavento	393

## y	San Antonio del Tequendama	934
## y	San Bernardo del Viento	48
## y	San Carlos de Guaroa	908
## y	San Cristóbal	1105
## y	San Felipe	867
## y	San Jacinto del Cauca	945
## y	San Jerónimo	91
## y	San Joaquín	754
## y	San José	299
## y	San José de La Montaña	941
## y	San José de Miranda	914
## y	San José de Pare	949
## y	San José de Uré	937
## y	San José del Fragua	882
## y	San José del Guaviare	903
## y	San José del Palmar	943
## y	San Juan de Arama	940
## y	San Juan de Betulia	928
## y	San Juan de Río Seco	885
## y	San Juan de Urabá	956
## y	San Juan del Cesar	957
## y	San Juan Nepomuceno	160
## y	San Luis de Gaceno	888
## y	San Luis de Gaceno	925
## y	San Luis de Sincé	924
## y	San Martín	614
## y	San Martín	374
## y	San Martín de Loba	931
## y	San Miguel de Sema	964
## y	San Pablo de Borbur	951
## y	San Pablo de Borbur	921
## y	San Pedro de Cartago	938
## y	San Pedro de Uraba	902
## y	San Sebastián	972
## y	San Sebastián de Buenavista	896
## y	San Vicente de Chucurí	913
## y	San Vicente del Caguán	1092
## y	San Zenón	586
## y	Sandoná	662
## y	Santa Bárbara	99
## y	Santa Bárbara	666
## y	Santa Bárbara	756
## y	Santa Bárbara de Pinto	1089
## y	Santa Helena del Opón	920
## y	Santa Lucía	132
## y	Santa María	546
## y	Santa María	248
## y	Santa Rosa de Cabal	916
## y	Santa Rosa de Osos	910
## y	Santa Rosa de Viterbo	904
## y	Santa Rosa del Sur	893
## y	Santa Rosalía	880
## y	Santa Sofía	249
## y	Santafé de Antioquia	907
## y	Santander de Quilichao	905
## y	Santiago de Tolú	952
## y	Santo Tomás	133
## y	Sesquilé	468
## y	Sibaté	469
## y	Simití	163
## y	Sipí	517
## y	Soatá	253
## y	Socotá	254
## y	Sonsón	1101
## y	Sopetrán	103
## y	Sopó	1081
## y	Soracá	260
## y	Sotaquirá	259
## y	Suárez	346
## y	Suárez	1111
## y	Supatá	475
## y	Supía	300
## y	Suratá	761
## y	Susacón	261
## y	Sutamarchán	262
## y	Tadó	1096
## y	Támara	1075
## y	Támesis	104
## y	Tarapacá	863
## y	Tarazá	105
## y	Tibaná	266
## y	Timaná	552

## y	Timbío	348	
## y	Timbiquí	349	
## y	Tinjacá	267	
## y	Titiribí	107	
## y	Tocancipá	485	
## y	Togúí	892	
## y	Tolú Viejo	787	
## y	Tópaga	270	
## y	Topaipí	486	
## y	Totoró	351	
## y	Tubará	136	
## y	Tuchín	397	
## y	Tuluá	1061	
## y	Tununguá	11	
## y	Túquerres	670	
## y	Turbaná	168	
## y	Turmequé	272	
## y	Tutazá	273	
## y	Ubalá	487	
## y	Unguía	518	
## y	Unión Panamericana		1114
## y	Uribe	605	
## y	Usiacurí	137	
## y	Útica	490	
## y	Valle de Guamez		950
## y	Valle de San José		962
## y	Valle de San Juan		912
## y	Valparaíso	113	
## y	Valparaíso	313	
## y	Vegachí	114	
## y	Vélez	763	
## y	Vianí	491	
## y	Vigía del Fuerte		930
## y	Villa Caro	1040	
## y	Villa de Leyva	895	
## y	Villa de San Diego de Ubate		890
## y	Villa del Rosario	1022	
## y	Villagarzón	1116	
## y	Villagómez	492	
## y	Villamaría	302	
## y	Villapinzón	493	
## y	Viotá	495	
## y	Viracachá	276	
## y	Yacopí	1099	
## y	Yaguará	554	
## y	Yalí	116	
## y	Yavaraté	1054	
## y	Yolombó	118	
## y	Yondó	119	
## y	Zapayán	590	
## y	Zipacón	496	
## y	Zipaquirá	1109	

```
##
##
##
## Table: Differences detected by variable
##
##
## -----
## No differences detected by variable
## -----
##
##
##
## Table: Differences detected
##
##
## -----
## No differences detected
## -----
##
##
##
## Table: Non-identical attributes
##
##
## -----
## No non-identical attributes
## -----
```



## Conclusiones

Este paquete de análisis automático posee muy pocas funciones para un análisis exploratorio, su fuerte general está en sus funciones de depuración que dan información útil sobre, por ejemplo, las diferencias entre 2 bases de datos, pero más allá de eso el paquete no ofrece muchas ventajas que R y Rstudio no tengan de manera nativa o en su defecto otros paquetes como tidyverse lo hacen de una manera más eficiente.

## funModeling

Algunas funciones de funModeling son:

### “correlation\_table”

Obtiene una tabla de correlación para todas las variables contra una variable objetivo. Solo variables numéricas son analizadas.

```
##                               Variable
## 1      1/ Autofinanciación de los gastos de funcionamiento
## 2                Posición 2019 a nivel Nacional
## 3 3/ Dependencia de las transferencias de la Nación y las Regalías
## 4      2/ Respaldo del servicio de la deuda
## 5      5/ Magnitud de la inversión
## 6      4/ Generación de recursos propios
## 7      Indicador de desempeño Fiscal 7/
## 8      6/ Capacidad de ahorro
## 1/ Autofinanciación de los gastos de funcionamiento
## 1      1.00
## 2      0.40
## 3      0.26
## 4      0.00
## 5     -0.03
## 6     -0.08
## 7     -0.46
## 8     -0.65
```

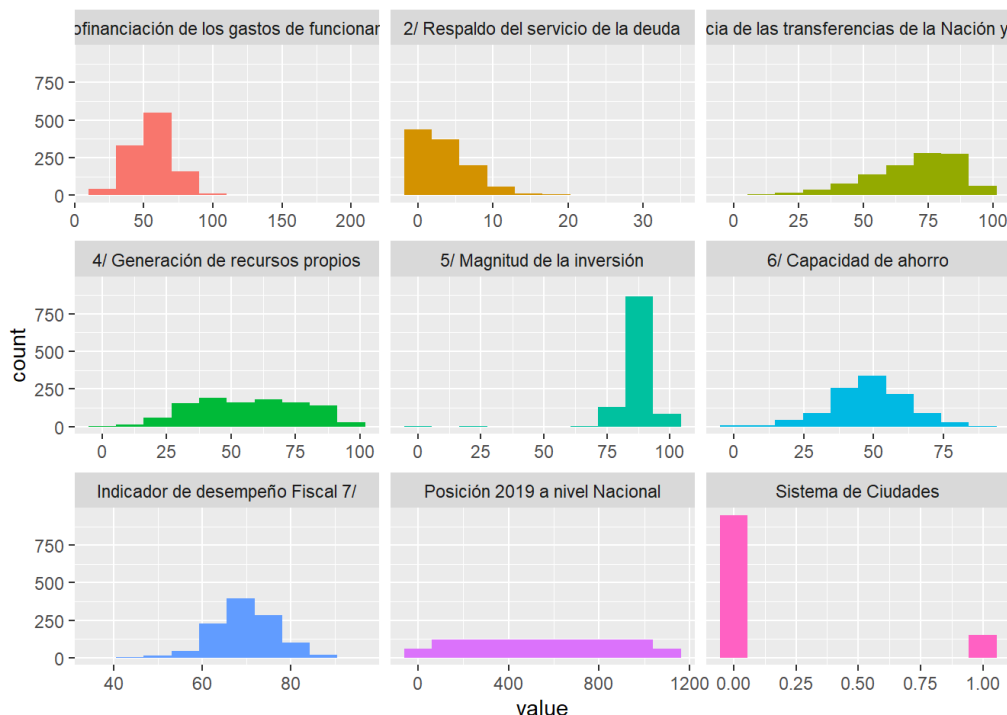
### “df\_status”

Para cada variable devuelve: Cantidad y porcentaje de ceros (q\_zeros y p\_zeros respectivamente). Mismos indicadores para valores NA (q\_NA/p\_na), y valores infinitos (q\_inf/p\_inf). Las últimas 2 columnas indican tipos de datos y cantidad de valores únicos. Esta función imprime y devuelve los resultados.

```
## variable q_zeros p_zeros q_na p_na q_inf p_inf type unique
## 1 var 1074 97.55 0 0 0 0 character 6
```

### “plot\_num”

Devuelve un gráfico que contiene todos los histogramas de las variables numéricas. Los valores NA no serán mostrados.



## Conclusiones

En general, el paquete funModeling tiene una gran cantidad de funciones interesantes que permite ahorrar trabajo a la hora de hacer un análisis exploratorio, como por ejemplo, permite de una manera más completa y eficiente que las funciones nativas desarrollar gráficas de interés además de resumir características relevantes de las bases de datos. Además la sintaxis para el uso de las mismas es sencilla y permite una fácil comprensión lógica de su uso.

## Comparación entre paquetes

### Sobre la sintaxis



La sintaxis de las funciones del paquete `arsenal` poseen un menor grado de interpretación en cuanto a quien las use, puesto que, sus palabras reservadas en algunos casos no terminan teniendo una relación de manera clara con su accionar, por ejemplo, en la función `selectall` se utiliza el argumento `"x"` para hacer referencia al data frame al cual se le aplicara, el argumento con nombre `"x"` es sumamente genérico y no representa el objetivo del argumento y si se lee sin una explicación termina siendo poco claro. En comparación el paquete `funModeling` tiene una sintaxis clara que si refleja el objetivo de la mayoría de sus argumentos, por ejemplo, en la función `"discretize_df"` el argumento `"stringsAsFactors"` muestra claramente desde su sintaxis que el objetivo del mismo es convertir variables de tipo `string` a `factor`, esto facilita un mayor entendimiento a posterior del código y una mayor efectividad debido a la recordación de las palabras que hace por lo tanto que su uso sea más fluido por parte de la persona programadora.

## Sobre la utilidad

En términos de la utilidad general, el paquete `funModeling` es mejor en muchos aspectos que `arsenal`, empezando por las funciones encaminadas a depurar, el paquete `funModeling` brinda una gama de funciones útiles y eficientes, pero lo más importante es que cumplen objetivos que no se podrían cumplir con las funciones nativas, en cambio, `arsenal` posee una cantidad de funciones que cumplen objetivos bastantes similares a funciones nativas, pero no lo hacen de una mejor manera, por el contrario, sus argumentos y su proceso lógico es en varias ocasiones menos claro, por ejemplo, la función `"mdy.Date"` cumple un fin parecido a la función `"as.date"` nativa pero el argumento `"year.cut"` de la primera no funciona correctamente además que no se le puede pasar una variable como argumento sino una fecha numérica. Ahora refiriéndonos a las funciones con uso en el análisis exploratorio `arsenal` tiene muy pocas funciones con este propósito, por otro lado, `funModeling` tiene una amplia gama que va desde histogramas hasta matrices de entropía que nos pueden ayudar en este objetivo, de una manera más eficiente que las nativas gracias a que se puede efectuar de manera automática a todas las variables con un solo comando y no uno a uno.

## Conclusiones

En conclusión, el paquete `funModeling` es ampliamente mejor que `arsenal`, desde la utilidad de sus funciones, sean para depurar o generar análisis, o desde su sintaxis y legibilidad. Siendo `funModeling` un paquete fácil de usar y con gran utilidad a la hora de explorar datos de manera eficaz.

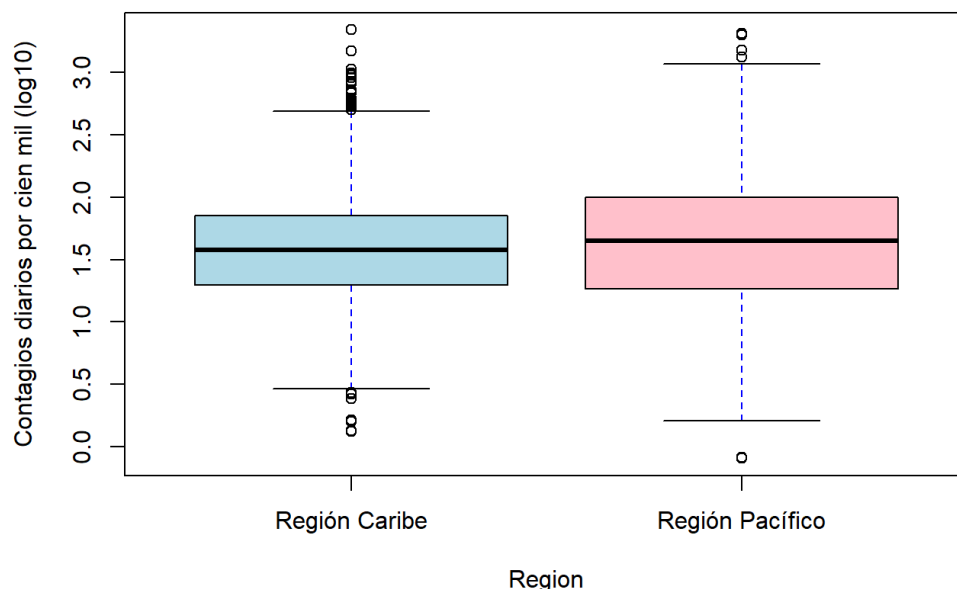
# Análisis descriptivo y exploratorio con "contagios\_x\_cien\_mil" como variable de respuesta.

## Sobre las regiones.

### Sexo Hombres.

Se efectúa una transformación logarítmica con base 10 a la variable `"contagios_x_cien_mil"` para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

### Contagios por día de hombres por cien mil habitantes log10



### Summary contagios diarios hombres por cien mil habitantes - Región Pacífico.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.808	18.460	44.850	78.026	99.399	2054.794

### Summary contagios diarios hombres por cien mil habitantes - Región Caribe.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.323	19.685	37.893	56.853	71.685	2197.802

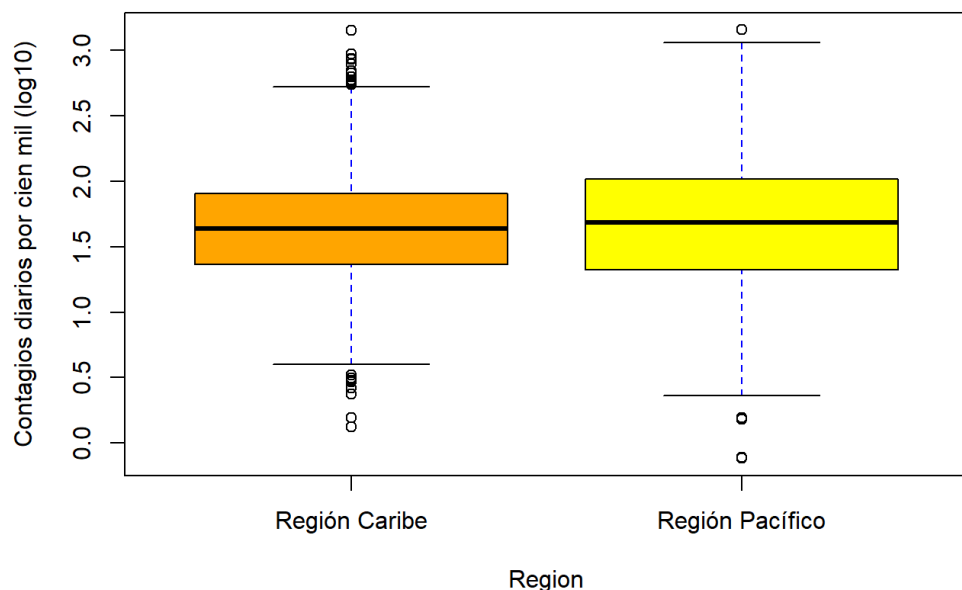
Se observan mediante el diagrama de cajas y los resúmenes de las variables que la región pacifico tiene ligeramente más contagios hombres por cien mil habitantes que la región caribe, aunque las diferencias mediante la gráfica puede no parecer alarmante, se debe tener en cuenta que cada aumento aunque sea marginal en las personas contagiadas es altamente peligroso gracias a la difusión del COVID-19. El promedio de contagiados hombres por día y cienmilaje en la región pacifico es de 78.026 vs. el promedio en la región caribe que es de 56.853, esto de la mano del modelo de análisis, que ayuda a ver en perspectiva el número de contagios gracias a su cienmilaje demuestra una peor gestión por parte de la región pacifica en

cuanto a los hombres contagiados diariamente.

## Sexo Mujeres.

Se efectúa una transformación logarítmica con base 10 a la variable “contagios\_x\_cien\_mil” para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

### Contagios por día de mujeres por cien mil habitantes log10



#### Summary contagios diarios mujeres por cien mil habitantes - Región Pacífico.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.7656	21.0493	48.4027	80.9971	104.7120	1433.6918

#### Summary contagios diarios mujeres por cien mil habitantes - Región Caribe.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.324	23.103	43.745	62.031	80.775	1415.094

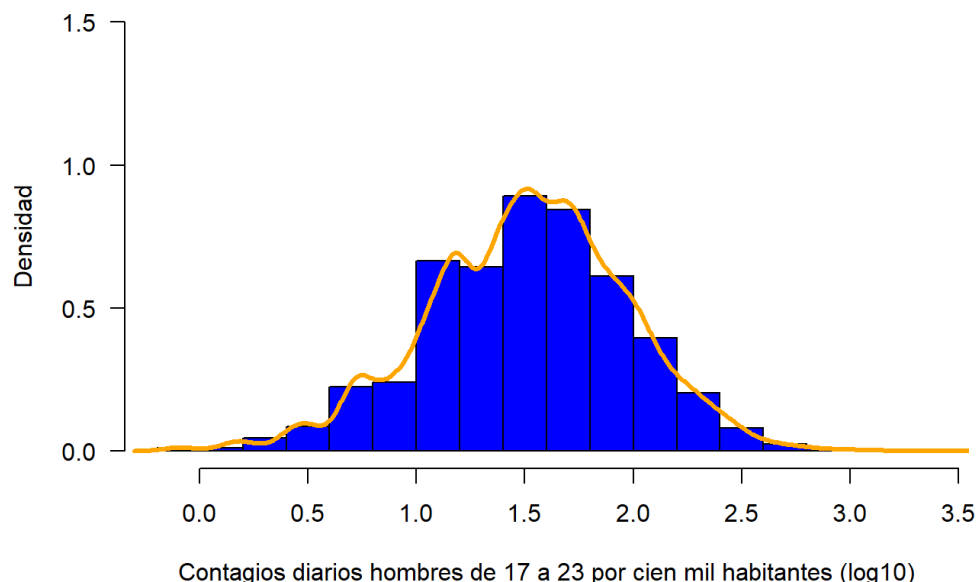
Se observan mediante el diagrama de cajas y los resúmenes de las variables que la región pacifico de nuevo tiene más contagios por cien mil habitantes, en este caso mujeres, algo de destacar es que hay una mayor tasa de contagio general en las mujeres con base a los promedios de cada análisis, siendo las de este 80.9971 para la región pacifica y 62.031 en la región caribe. Esto debido posiblemente al impacto de la pandemia en las mujeres y en sus necesidades económicas, provocando que se tengan que arriesgar más que los hombres a la hora de salir a trabajar o en general.

## Sobre los rangos de edad

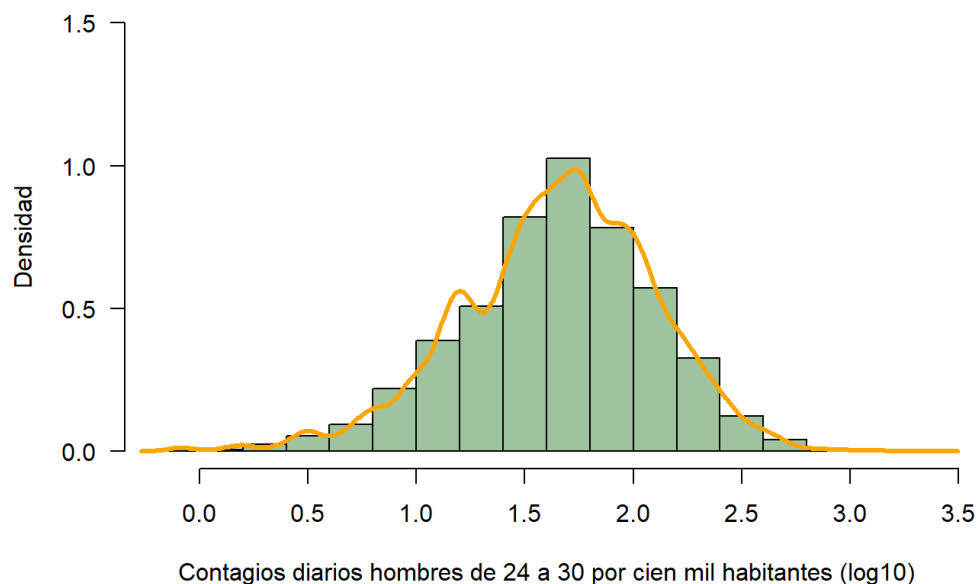
### Sexo Hombres.

Se efectúa una transformación logarítmica con base 10 a la variable “contagios\_x\_cien\_mil” para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

### Contagios diarios hombres de 17 a 23 por cien mil habitantes



## Contagios diarios hombres de 24 a 30 por cien mil habitantes



### Summary contagios diarios hombres por cien mil habitantes - [17,23].

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.808	15.743	33.129	56.375	65.920	2197.802

### Summary contagios diarios hombres por cien mil habitantes - (23,30]

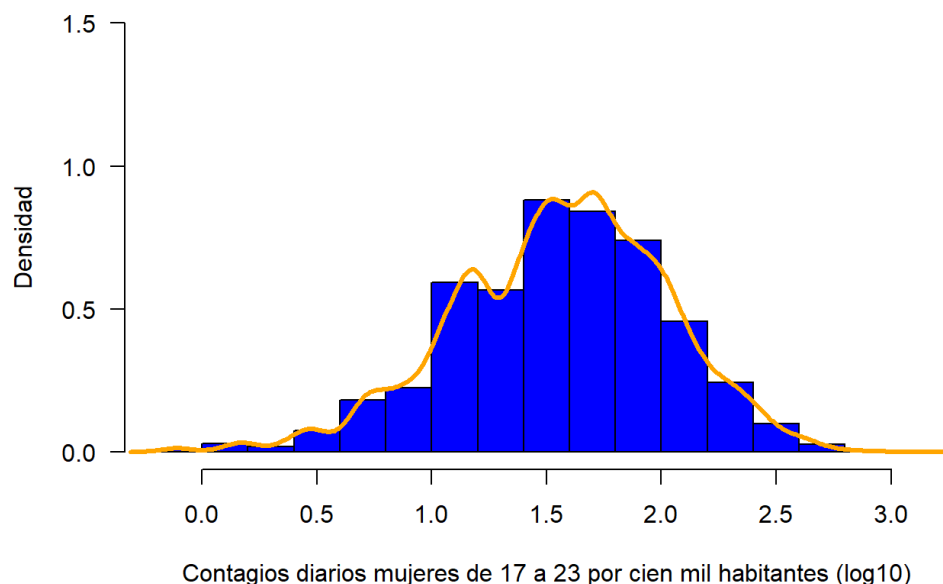
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.8232	23.5793	47.1587	71.7311	90.4704	2054.7945

De acuerdo a ambas gráficas se puede observar una tendencia a la simetría en ambos rangos de edades, debido posiblemente a que no son tan diferentes entre sí y que el comportamiento de ambos grupos en la pandemia fue parecido, aunque cabe resaltar que existe mayor tendencia a la simetría en las edades de 24 a 30 lo que se puede generar por una mayor consistencia en el comportamiento de estas personas, ya que la mayoría está en etapa productiva y se ven obligadas a exponerse a los contagios vs. el grupo de 17 a 23 donde algunas están en etapa productiva y otras no. Por otro lado, analizando las tablas de resumen se puede apreciar que quien tiene mayor tasa de contagio es el grupo de 24 a 30 con un promedio de 71.7311 contagiados por cien mil personas, esto debido a la explicación anterior que muestra como es más probable que se vean expuestos. Aunque se logra ver que el mayor valor pertenece al grupo de 17 a 23 con un contagio de 2197.802 por cien mil habitantes, esto debido posiblemente a un comportamiento anormal de este grupo en tiempos donde las medidas de bioseguridad se aminoraron.

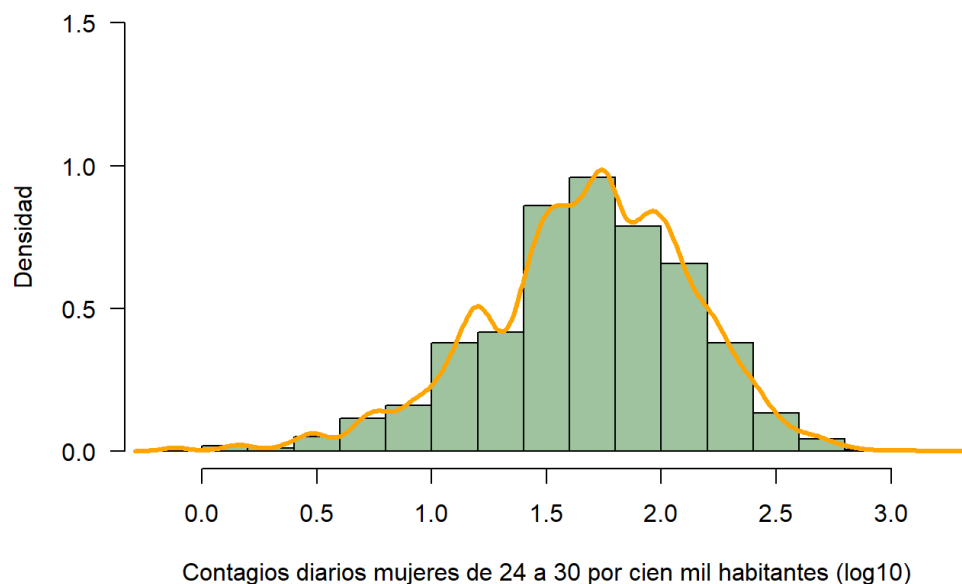
### Sexo Mujeres.

Se efectúa una transformación logarítmica con base 10 a la variable "contagios\_x\_cien\_mil" para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

## Contagios diarios mujeres de 17 a 23 por cien mil habitantes



## Contagios diarios mujeres de 24 a 30 por cien mil habitantes



### Summary contagios diarios mujeres por cien mil habitantes - [17,23].

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.7772	16.9549	38.1814	60.5867	76.4526	1063.8298

### Summary contagios diarios mujeres por cien mil habitantes - (23,30]

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.7656	26.2192	51.7866	76.0978	98.5222	1433.6918

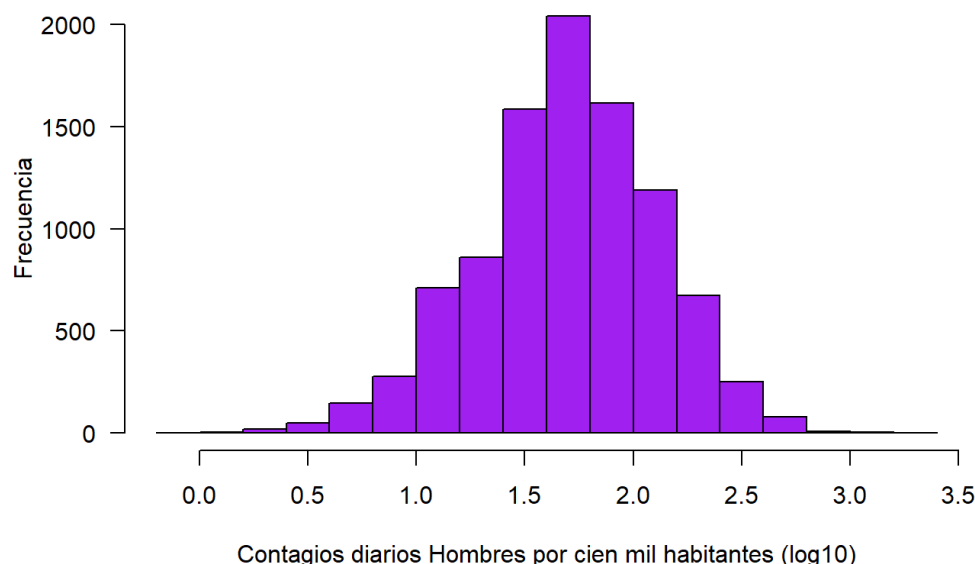
En estas gráficas, a diferencia de las anteriores, se puede ver una mayor acumulación de los datos hacia la derecha, lo que implica que en las mujeres hay más contagios por cien mil habitantes, en específico la gráfica del grupo de 24 a 30 muestra una acumulación más grande en datos mayores, esto debido probablemente a que en estos rangos de edad se está en etapa productiva, por el contrario, aunque en el grupo de 17 a 23 si se observa una acumulación a la izquierda, esta no es tan absoluta, esto debido posiblemente a que parte de las mujeres en esta edad estén estudiando y se les migrara a la virtualidad, lo que reduce el riesgo de contagio. Observando la tabla de resumen podemos confirmar que con base a los promedios, que son respectivamente 60.5867 y 76.0978, que si existe un mayor contagio por parte de en general las mujeres, pero más específicamente de las que pertenecen al grupo de 24 a 30, de nuevo posiblemente a los impactos de la pandemia y su obligación económica como personas que impide un correcto cuidado en la pandemia.

## Sobre los periodos considerados

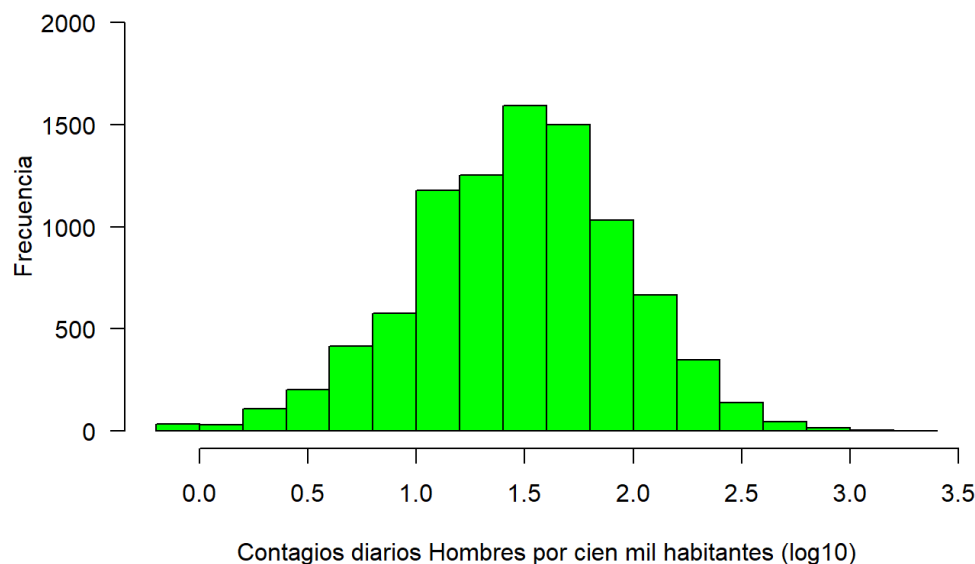
### Sexo Hombres.

Se efectúa una transformación logarítmica con base 10 a la variable "contagios\_x\_cien\_mil" para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

## Contagios diarios Hombres, entre 2021-05-25 y 2021-08-18



## Contagios diarios Hombres, entre 2020-03-2 y 2020-09-22



### Summary contagios diarios Hombres por cien mil habitantes - [2021-05-25 y 2021-08-18].

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.808	27.863	51.967	75.713	96.061	1993.355

### Summary contagios diarios Hombres por cien mil habitantes - [2020-03-2 y 2020-09-22].

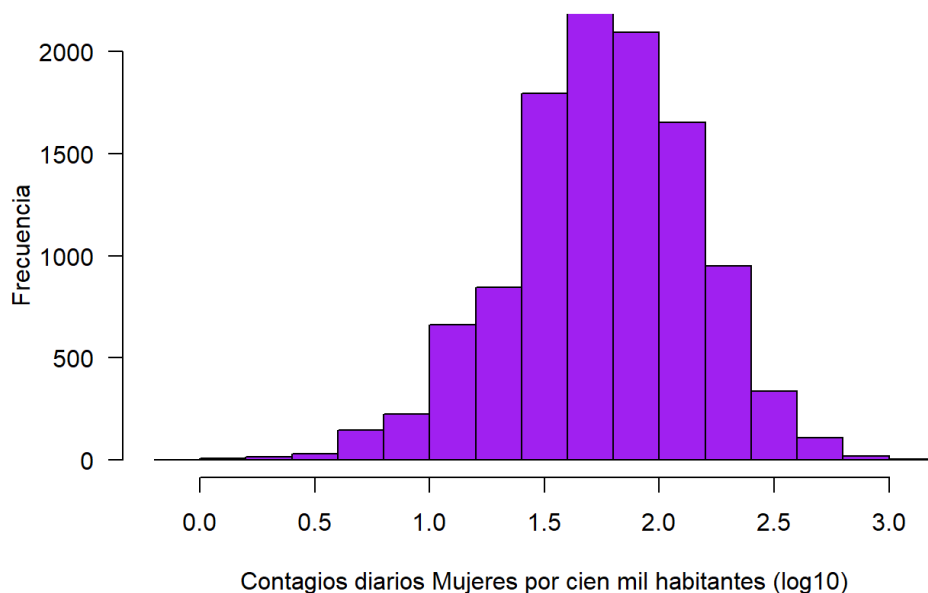
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.808	14.680	31.726	54.206	62.933	2197.802

Se observa en las gráficas como en el periodo de 2021-05-25 y 2021-08-18 existe una mayor acumulación de los datos a la derecha a comparación de la que pertenece al periodo 2020-03-2 y 2020-09-22, además que las frecuencias de manera general son más alta en la primera, esto es causado posiblemente por las distintas dinámicas pandémicas, entre ellas, las cuarentenas obligatorias que en el segundo periodo estaban en vigor vs. la aminoración paulatina en el primer periodo, también existe la posibilidad de que esto se combinara con el surgimiento de nuevas variantes y la necesidad imperante de salir a trabajar de manera presencial, lo que produce claramente un aumento en los contagios, puesto que existe mayor transmisibilidad del virus. Ahora bien observando las tablas de resumen, se ve una gran diferencia entre los promedios de ambos periodos, siendo 75.713 y 54.206 respectivamente, lo cual respalda la hipótesis anterior, si observamos los valores máximos de ambas variables podemos ver que sorpresivamente el valor más alto es el del segundo periodo siendo este de 2197.802 contagiados por cien mil habitantes vs. el del primero que es de 1993.355 contagiados por cien mil habitantes, una hipótesis para esto es que al principio de la pandemia (periodo 2) tomo un tiempo considerable adaptarnos a la misma y tomar medidas de bioseguridad, por lo cual al principio eran probables contagios por día altos, por otro lado, si observamos el primer valor máximo, se ve como aunque los contagios en general son más altos por las necesidades de las personas, hay más herramientas para disminuirlos y que no tomen valores por día tan altos, entre estas herramientas podemos observar la vacunación.

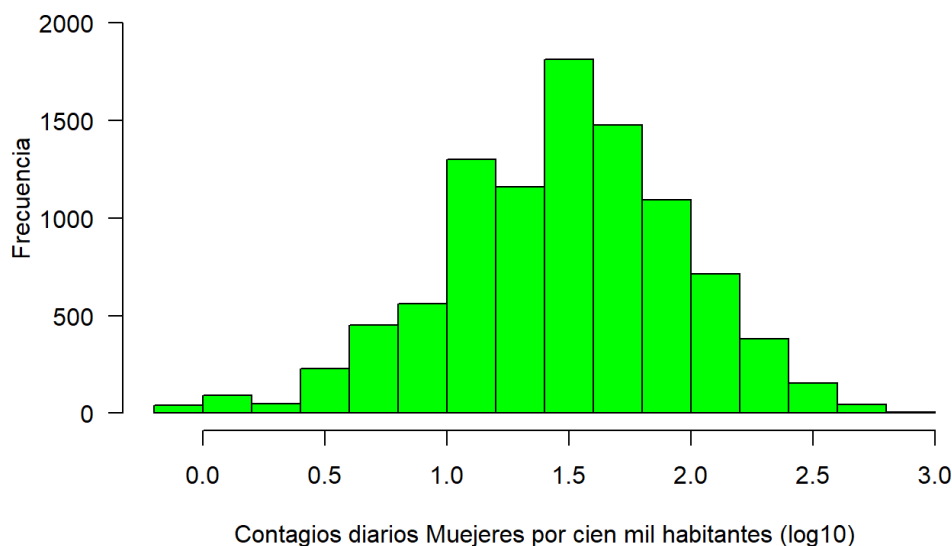
### Sexo Mujeres

Se efectúa una transformación logarítmica con base 10 a la variable "contagios\_x\_cien\_mil" para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

### Contagios diarios Mujeres, entre 2021-05-25 y 2021-08-18



### Contagios diarios Mujeres, entre 2020-03-2 y 2020-09-22



#### Summary contagios diarios Mujeres por cien mil habitantes - [2021-05-25 y 2021-08-18].

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.7656	32.2139	57.3477	83.5898	106.0165	1433.6918

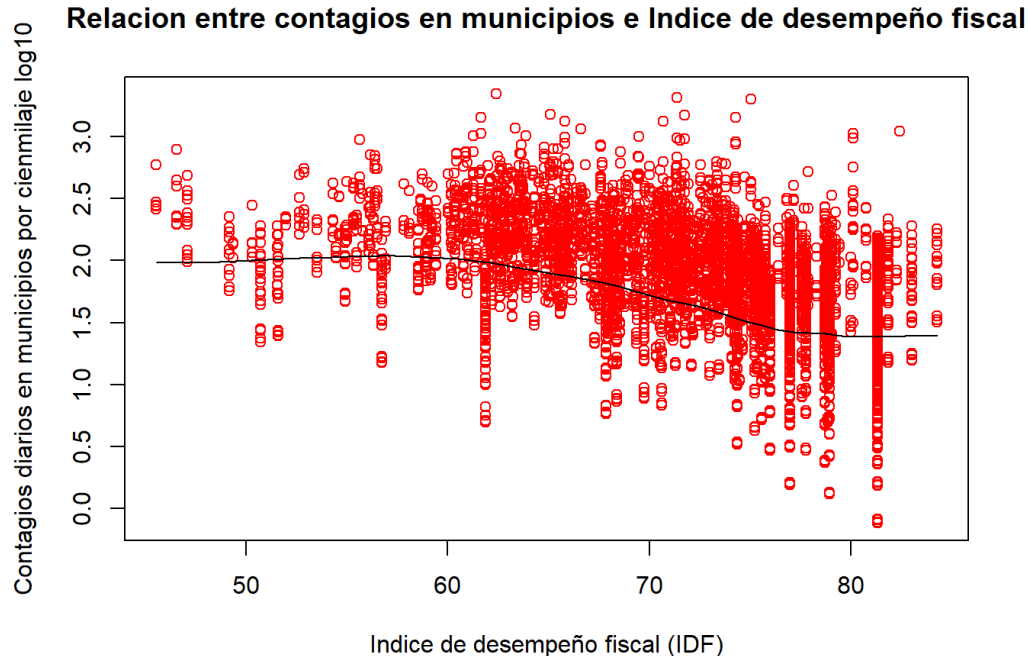
#### Summary contagios diarios Hombres por cien mil habitantes - [2020-03-2 y 2020-09-22].

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.7656	14.7798	31.8878	52.6181	63.2511	985.2217

En estas gráficas se muestra una diferencia considerablemente grande entre ambos periodos, siendo el que tiene mayor cantidad de casos por día por cien mil habitantes el de 2021-05-25 y 2021-08-18, aunque en el análisis pasado se dieron varias hipótesis para esto, en este caso es aún más sorprendente que sea tan marcada la diferencia, además que dentro de todos los análisis exploratorios anteriores se ha mostrado como las mujeres tienen una mayor tasa de contagios que los hombres. Si observamos las tablas resúmenes se confirma lo anterior, donde el promedio del primer periodo es de 83.5898 casos por día por cien mil habitantes y el del segundo periodo es de 52.6181 casos diarios por cien mil habitantes. Una diferencia de casi 30 casos lo cual es ampliamente preocupante, puesto que en este primer periodo ya hay mecanismos como la vacunación, analizando los valores máximos observamos como a diferencia que con los hombres, el mayor valor es perteneciente al primer periodo, una posible hipótesis para esto y para en general la mayor tasa de contagio en las mujeres es que las dinámicas de la pandemia les han forzado en mayor medida a arriesgarse que a los hombres, debido a la falta de apoyo estatal y que se ven mediadas por una sociedad patriarcal.

## Análisis considerando el desempeño fiscal de cada municipio y su relación con en nivel de contagios.

Se efectúa una transformación logarítmica con base 10 a la variable "contagios\_x\_cien\_mil" para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.



#### Summary sobre las variables contagios por cien mil habitantes e Indice de desempeño fiscal

```
## contagios_x_cien_mil   IDF
## Min. : 0.7656   Min. :45.54
## 1st Qu.: 20.6807   1st Qu.:68.20
## Median : 43.6491   Median :73.33
## Mean : 67.3594   Mean :71.75
## 3rd Qu.: 85.5432   3rd Qu.:75.97
## Max. :2197.8022   Max. :84.25
##      NA's :419
```

#### Coefficiente de Correlación de Pearson

```
## [1] -0.3553127
```

Desde la gráfica se puede observar que la relación entre el IDF y el nivel de contagios diarios por municipio no es realmente alto, pero que si existe cierta relación negativa aunque sea leve, se ve como la nube de puntos va con cierta consistencia en declive, esto significa que de cierta manera los municipios con mayor desempeño fiscal tendrán una tendencia, aunque sea baja, a tener menores contagios. Cuando observamos el coeficiente de correlación de Pearson se confirma lo anterior, existe una relación negativa, pero no muy fuerte entre ambas variables, una hipótesis para esto es que los municipios con mejor IDF tienen una mejor gestión de sus recursos y por lo tanto un mejor manejo del dinero para afrontar la pandemia, o sea, una mejor compra de insumos, vacunas, entre otras cosas, también lo que nos lleva a pensar es que aunque el IDF si influye, esta no es tan alta debido a que los municipios que tienen una mayor capacidad fiscal son los mismos que tienen actividades productivas fuertes, como por ejemplo Cali, por lo tanto es más difícil mantener medidas que frenen el comercio como las cuarentenas, ya que se tiene grandes maquinarias empresariales que se verían afectadas además de la economía del país, así es que podemos intuir que estos lugares tienen una obligación de continuar sus objetivos económicos a pesar de la pandemia. En conclusión estos municipios con mayor IDF tienen mayores capacidades organizacionales para afrontar la pandemia, pero también tienen una actividad económica esencial que evita un manejo perfecto frente a los contagios. Perciba que en la tabla resumen el valor máximo es de 2197.8022 contagios por cien mil habitantes, este valor pertenece al municipio de Distracción con un IDF de 62.39166 el cual es considerablemente bajo, ejemplificando así la hipótesis anterior.

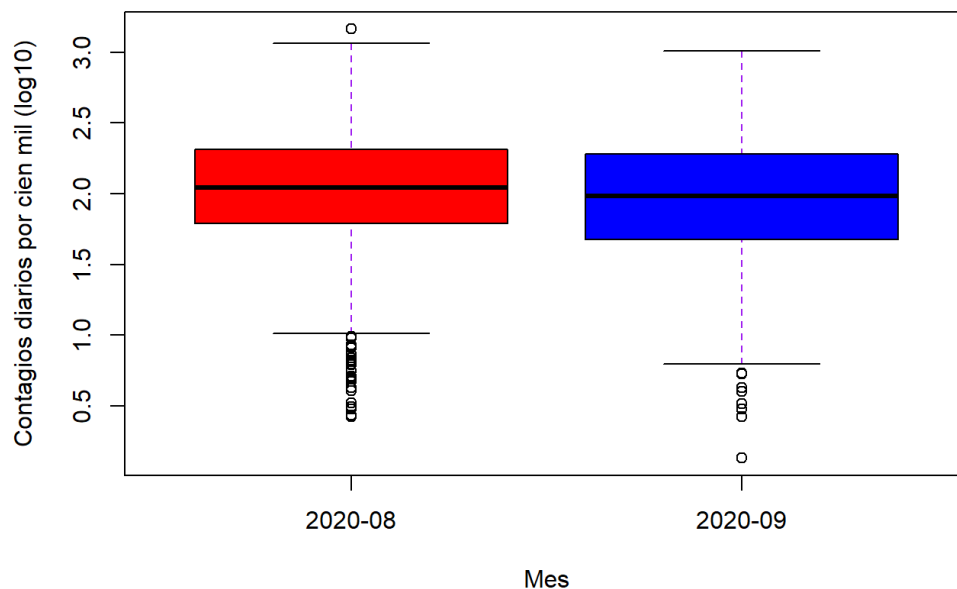
## Preguntas propias.

**¿Existe alguna diferencia entre los contagios por cien mil habitantes en el grupo de edad de 24 a 30 entre el mes de agosto de 2020 donde se tenía aislamiento preventivo estricto y el mes de septiembre del 2020 (Decreto 1168,2020) donde se acabó esta medida y se pasó a otras menos contundentes como el pico y cédula?**

**Variables utilizadas: Fecha de notificación, contagios por cien mil habitantes, edad**

Se efectúa una transformación logarítmica con base 10 a la variable "contagios\_x\_cien\_mil" para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.

## Contagios diarios de personas entre 24 y 30 por cien mil habitantes



### Summary contagios diarios por cien mil habitantes de edad 24 a 30 - 2020-08

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2.648	61.569	110.558	157.590	204.499	1459.854

### Summary contagios diarios por cien mil habitantes de edad 24 a 30 - 2020-09

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.351	47.184	96.363	141.948	190.206	1020.408

Sorprendentemente se observa en las gráficas una ligera disminución en la tasa de contagios en el periodo 2020-09, una hipótesis para esto es que en último mes de medidas restrictas de confinamiento como fue agosto, la población objetiva de este análisis la cual es personas de 24 a 30, ya había salido bajo excepciones o por necesidad a trabajar y por lo tanto los contagios se vieron altos, esto fundamentado en que esta población está en su gran mayoría en etapa productiva y normalmente son las personas que mantienen económicamente los núcleos familiares. Cuando vemos la tabla resumen se aprecia lo anteriormente dicho, una ligera disminución en el promedio en el segundo periodo siendo estos 157.590 y 141.948 respectivamente, la disminución de contagios entre periodos se puede tener una explicación en que antes de que se levantaran las restricciones la mayoría de lugares en general no tenía medidas de bioseguridad estrictas, puesto que se seguía con confinamiento y estas suponen un costo económico que deberían asumir las personas que son encargadas de los lugares, pero luego de que se levantaran se vieron obligados a implementar y hacer cumplir las medidas por presión seguramente del gobierno para la reactivación, haciendo así que nuestra población objetiva tuviera un poco más de garantías para evitar los contagios. En conclusión los contagios disminuyeron un poco comparativamente entre ambos periodos presumiblemente porque la población objetiva ya estaba fuera de las medidas estrictas y los lugares donde se esparcen los contagios tenían medidas más estrictas.

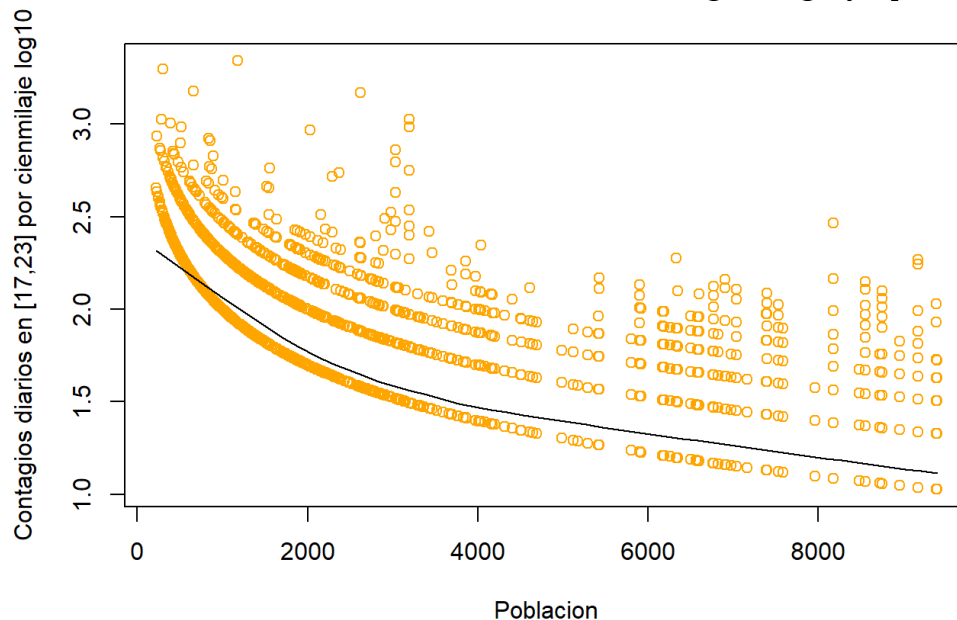
**¿Existe evidencia de una relación entre un número de población bajo (hasta de 10.000 habitantes) y los contagios diarios por cien mil habitantes en la edad de 17 a 23?**

**Variables utilizadas: Población, contagios por cien mil habitantes, edad**

Se efectúa una transformación logarítmica con base 10 a la variable "contagios\_x\_cien\_mil" para una mejor apreciación gráfica, esto solo se efectúa para la realización de estas.



## Relacion Poblacion menor a 10.000 - Contagio en grupo [17,23]



### Summary sobre las variables contagios por cien mil habitantes y población

```
## poblacion contagios_x_cien_mil
## Min. :220 Min. : 10.65
## 1st Qu.:1382 1st Qu.: 28.42
## Median :2634 Median : 50.02
## Mean :3418 Mean : 74.59
## 3rd Qu.:5293 3rd Qu.: 91.91
## Max. :9394 Max. :2197.80
```

### Coefficiente de Correlación de Pearson

```
## [1] -0.5002498
```

Cuando observamos la gráfica vemos una curva extremadamente decreciente a medida que se aumenta la población, una hipótesis para este comportamiento es que cuando se trata de municipios con población menor a 10.000 se debe tener en cuenta que entre menos población se tenga es probable que se tenga menos dinero y recursos para afrontar la pandemia, esto en el sentido de lógicas de producción donde no se tiene una industria conformada ni se tiene una fuerza de trabajo grande, por lo tanto es probable que sus recursos antes de la pandemia fueran significativamente cortos a comparación de los municipios grandes y con la pandemia se vieron tremendamente reducidos, evitando así un correcto afrontamiento de la situación por la incapacidad de comprar los suministros, como por ejemplo, tapabocas, camas UCI, entre otros. Además cuando observamos el coeficiente de Correlación de Pearson se confirma lo anterior, existe una fuerte relación negativa entre ambas variables bajo las condiciones dadas, otro análisis para lo mismo es que al tener menor población un solo contagio como unidad marginal tiene muchísima más afectación en los contagios por cien mil habitantes, esto es correcto porque nos ayuda a medir el impacto real de ese contagio en el municipio, puesto que, solo esa persona enferma implica una disminución en la capacidad del municipio y en general los contagios se pueden viralizar de una mayor medida. Observando la tabla resumen salta a la vista que el valor máximo de contagios por cien mil bajo las condiciones de esta pregunta, es el mismo valor máximo de la base de datos general que incluye municipios grandes, lo que nos lleva a una mejor comprobación de la hipótesis, puesto que el valor pertenece al municipio de Distracción con una población de 1183 habitantes, lo cual es significativamente poco a comparación por ejemplo del municipio de Cali con una población de 130623 habitantes. En conclusión en los municipios con población de hasta 10.000 entre más aumente esta es más probable que se disminuyan los contagios por cien mil habitantes, puesto que es probable que se tenga un poco más de recursos para afrontar la pandemia y que un contagio no impacte tanto al municipio.

## Bibliografía y Cibergrafía

### Cibergrafía

Departamento Nacional de Planeación.(2019).Desempeño Fiscal.*Resultados de desempeño fiscal de los departamentos y municipios de la vigencia 2019.* [https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Territorial/Anexos%20Desempe%C3%B1o%20Fiscal%202019\\_Resultado%20Oficial](https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Territorial/Anexos%20Desempe%C3%B1o%20Fiscal%202019_Resultado%20Oficial).

Departamento Administrativo Nacional de Estadística.(2020, 9 octubre).Proyecciones de población.*Serie municipal de población por área, sexo y edad, para el periodo 2018 -2026.* [https://www.dane.gov.co/files/censo2018/proyecciones-de-poblacion/Municipal/anexo-proyecciones-poblacion-Municipal\\_2018-2026.xlsx](https://www.dane.gov.co/files/censo2018/proyecciones-de-poblacion/Municipal/anexo-proyecciones-poblacion-Municipal_2018-2026.xlsx)

Departamento Nacional de Planeación. (s. f.).*Nuevo Índice de Desempeño Fiscal. Nuevo Índice de Desempeño Fiscal.* <<https://www.dnp.gov.co/DNPN/Paginas/Nuevo-índice-de-Desempeno-Fiscal.aspx>>

Instituto Nacional de Salud. (2021, 18 agosto). Casos positivos de COVID-19 en Colombia, Datos Abiertos Colombia.*Casos positivos de COVID-19 en Colombia.* <https://www.datos.gov.co/Salud-y-Proteccion-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr>

Ministerio de Tecnologías de la Información y las Comunicaciones.(2020, 20 mayo).Departamentos y municipios de Colombia,Datos Abiertos Colombia.*Departamentos y municipios de Colombia.* <https://www.datos.gov.co/Mapas-Nacionales/Departamentos-y-municipios-de-Colombia/xdk5-pm3f>

## Bibliografía

Casas, P.(2020).*funModeling: Exploratory Data Analysis and Data Preparation Tool-Box [R package funModeling version 1.9.4]*. Comprehensive R Archive Network (CRAN). Retrieved from <https://cran.r-project.org/package=funModeling>

Dirección de Impuestos y Aduanas Nacionales. (2018). *Código municipios de Colombia*. [https://www.dian.gov.co/atencionciudadano/formulariosinstructivos/Formularios/2006/Codigo\\_municipios\\_2006.pdf](https://www.dian.gov.co/atencionciudadano/formulariosinstructivos/Formularios/2006/Codigo_municipios_2006.pdf)

Heinzen, E., Sinnwell, J., Atkinson, E., Gunderson, T., & Dougherty, G.(2021).*An Arsenal of “R” Functions for Large-Scale Statistical Summaries [R package arsenal version 3.6.3]*. Comprehensive R Archive Network (CRAN). Retrieved from <https://cran.r-project.org/package=arsenal>

Ministerio de Salud y Protección Social de Colombia. (2020).*ABECÉ. NUEVO CORONAVIRUS (COVID-19)*. <<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/ET/abece-coronavirus.pdf>>

Presidencia de la República.(25 de agosto del 2020).Por el cual se imparten instrucciones en virtud de la emergencia sanitaria generada por la pandemia del Coronavirus COVID-19, y el mantenimiento del orden público y se decreta el aislamiento selectivo con distanciamiento individual responsable.[Decreto 1168].Recuperado de <https://dapre.presidencia.gov.co/normativa/normativa/DECRETO%201168%20DEL%2025%20DE%20AGOSTO%20DE%202020.pdf>