

Ejercicio R

Oscar Andrei Zempoalteca Ramírez

3/9/2023

Introducción

Se proporcionaron 5 bases de información de siniestros, cada una corresponde a los años desde 2016 a 2021. Las 5 bases tienen la misma estructura de columnas que a continuación, se describe: 1) Nombre.- Nombre del Asegurado 2) Fecha de Emisión.- Fecha en que se ha emitido la póliza 3) Inicio de Vigencia.- Fecha en la cual entro en vigor la póliza 4) Sexo.- Sexo del Asegurado 5) Fecha de Nacimiento.- Fecha de Nacimiento del Asegurado 6) Fecha de Ocurrido.- Fecha en la cual Ocurrio el Siniestro 7) Pagos.- Monto Pagado al Asegurado

Para poder realizar un análisis de la información, en primer lugar se importaron las librerías necesarias para manipulación de los datos. Cabe destacar que se utilizó el software R para este realizar este reporte.

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

Una vez hecho esto, se procedió a importar cada una de las bases para así crear una base de siniestros acumulada a la que llamaremos “siniestrosacumulado”

```
siniestros1<- read.table("C:/Users/HP/Downloads/gen re/Sini_2016.txt", header = TRUE, sep = "\t", fill = TRUE)
siniestros2<- read.table("C:/Users/HP/Downloads/gen re/Sini_2017.txt", header = TRUE, sep = "\t", fill = TRUE)
siniestros3<- read.table("C:/Users/HP/Downloads/gen re/Sini_2018.txt", header = TRUE, sep = "\t", fill = TRUE)
siniestros4<- read.table("C:/Users/HP/Downloads/gen re/Sini_2019.txt", header = TRUE, sep = "\t", fill = TRUE)
siniestros5<- read.table("C:/Users/HP/Downloads/gen re/Sini_2020.txt", header = TRUE, sep = "\t", fill = TRUE)
siniestros6<- read.table("C:/Users/HP/Downloads/gen re/Sini_2021.txt", header = TRUE, sep = "\t", fill = TRUE)
```

```
siniestrosacumulado<- rbind(siniestros1, siniestros2, siniestros3, siniestros4, siniestros5, siniestros6)
```

Esta base de siniestros acumulada por año desde 2016 hasta 2021 tiene el siguiente número de filas o registros:

```
dim(siniestrosacumulado)[1:1]
```

```
## [1] 1620560
```

Y las columnas que la componen tienen las siguientes características:

```
glimpse(siniestrosacumulado)
```

```
## Rows: 1,620,560
## Columns: 8
## $ Nombre      <chr> "RACAWI", "GOCAPA", "VIDIMA", "VIDIMA", "GOMEJO...
## $ Fecha.de.Emission <chr> "01/02/2001", "03/03/2002", "01/06/1998", "01/0...
## $ Inicio.de.Vigencia <chr> "01/02/2003", "03/03/2004", "01/06/2004", "01/0...
## $ Sexo        <chr> "F", "M", "M", "M", "F", "F", "F", "F", "M", "M...
## $ Fecha.de.Nacimiento <chr> "16/01/1978", "21/09/1952", "05/05/1947", "05/0...
## $ Fecha.de.Ocurrido <chr> "02/12/2003", "01/05/2004", "20/12/2004", "20/1...
## $ Pagos       <dbl> 527.00, 8329.62, 1490.69, 1683.74, 500.00, 648....
## $ X           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
```

Desarrollo

Las bases de información ahora se encuentran cargadas y listas para analizarlas. Notemos que las fechas no tienen el respectivo formato de fecha y la variable “Sexo” nos conviene convertirla en variable categórica para poder analizar el comportamiento de los hombres y mujeres como categorías. También al importar la información, se cargó erróneamente una columna sin datos “X”, eliminémosla:

```
siniestrosacumulado<-subset(siniestrosacumulado, select = -X)
```

```
siniestrosacumulado$Fecha.de.Ocurrido<-as.Date(siniestrosacumulado$Fecha.de.Ocurrido,format = "%d/%m/%Y")
siniestrosacumulado$Fecha.de.Nacimiento<-as.Date(siniestrosacumulado$Fecha.de.Nacimiento,format = "%d/%m/%Y")
siniestrosacumulado$Fecha.de.Emission<-as.Date(siniestrosacumulado$Fecha.de.Emission,format = "%d/%m/%Y")
siniestrosacumulado$Inicio.de.Vigencia<-as.Date(siniestrosacumulado$Inicio.de.Vigencia,format = "%d/%m/%Y")
siniestrosacumulado$Sexo<-as.factor(siniestrosacumulado$Sexo)
```

Ahora veamos que la información ya tiene las características que buscamos:

```
glimpse(siniestrosacumulado)
```

```
## Rows: 1,620,560
## Columns: 7
## $ Nombre      <chr> "RACAWI", "GOCAPA", "VIDIMA", "VIDIMA", "GOMEJO...
## $ Fecha.de.Emission <date> 2001-02-01, 2002-03-03, 1998-06-01, 1998-06-01...
## $ Inicio.de.Vigencia <date> 2003-02-01, 2004-03-03, 2004-06-01, 2004-06-01...
## $ Sexo        <fct> F, M, M, M, F, F, F, F, M, M, M, M, F, M, M, M,...
## $ Fecha.de.Nacimiento <date> 1978-01-16, 1952-09-21, 1947-05-05, 1947-05-05...
## $ Fecha.de.Ocurrido <date> 2003-12-02, 2004-05-01, 2004-12-20, 2004-12-20...
## $ Pagos       <dbl> 527.00, 8329.62, 1490.69, 1683.74, 500.00, 648....
```

Como nos interesa analizar el monto pagado al asegurado por año, insertemos una columna que nos indique el año de ocurrido al que corresponde cada uno de los registros de la base. Esta columna la llamaremos “añooccurrido”. Una vez generada la convertiremos en factor y veremos sus niveles:

```
siniestrosacumulado$añooccurrido<-format(siniestrosacumulado$Fecha.de.Ocurrido,"%Y")
siniestrosacumulado$añooccurrido<-as.factor(siniestrosacumulado$añooccurrido)
levels(siniestrosacumulado$añooccurrido) #hay información de 2022 pero no tiene fecha de emisión
```

```
## [1] "1998" "1999" "2000" "2001" "2003" "2004" "2005" "2006" "2007" "2008"
## [11] "2009" "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017" "2018"
## [21] "2019" "2020" "2021" "2022"
```

De forma adicional, veamos si existen filas con valores nulos o NA en las dos principales columnas que nos interesan “Pagos” y “añooccurrido”:

```
#siniestrosacumulado[is.na(siniestrosacumulado$añooccurrido), ]
#eliminemos estas filas lascuales sólo muestran el nombre y sexo de los individuos pero el resto de inf
siniestrosacumulado <-siniestrosacumulado[!is.na(siniestrosacumulado$Pagos),]
siniestrosacumulado <- siniestrosacumulado[!is.na(siniestrosacumulado$añooccurrido),]
```

Ahora la base de información tiene el siguiente número de registros:

```
dim(siniestrosacumulado)[1:1]
```

```
## [1] 1414937
```

Ya que tenemos la estructura y variables en su forma deseada, comencemos analizando los montos pagados al asegurado por año de ocurrencia:

```
resumenxaño = siniestrosacumulado %>% group_by(añooccurrido) %>%
  summarise(pago_total = sum(Pagos),pago_promedio=mean(Pagos), número_pagos=n(),
            pago_mínimo=min(Pagos),pago_máximo=max(Pagos), )
print(knitr::kable(resumenxaño,n=27))
```

```
##
##
## |añooccurrido | pago_total| pago_promedio| número_pagos| pago_mínimo| pago_máximo|
## |:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|
## |1998 | 2.859680e+03| 1429.8400| 2| 960.97| 1898.71|
## |1999 | 5.375780e+03| 1343.9450| 4| 1033.94| 1710.94|
## |2000 | 4.530745e+04| 9061.4900| 5| 542.63| 38572.62|
## |2001 | 8.476238e+04| 3685.3209| 23| 261.00| 58086.93|
## |2003 | 4.446995e+04| 562.9108| 79| 15.00| 1241.67|
## |2004 | 2.269690e+05| 4728.5204| 48| 47.00| 9007.18|
## |2005 | 3.323161e+05| 1263.5592| 263| 1.00| 19196.00|
## |2006 | 4.906247e+05| 4341.8116| 113| 100.00| 177005.63|
## |2007 | 9.945041e+05| 1622.3558| 613| 2.00| 89784.07|
## |2008 | 3.609084e+06| 7614.1023| 474| -31513.06| 456206.83|
## |2009 | 2.330829e+06| 2014.5455| 1157| 1.00| 92825.35|
## |2010 | 7.611893e+06| 3344.4170| 2276| -40684.06| 374389.16|
## |2011 | 1.250773e+07| 2651.6285| 4717| -3555.71| 163976.35|
```

##	2012	3.244803e+07	2730.1662	11885	-32641.06	395536.07
##	2013	5.543583e+07	3501.2838	15833	-161422.00	1346443.70
##	2014	1.126500e+08	3142.5225	35847	-16626.28	765810.55
##	2015	2.389263e+08	3975.4790	60100	-171912.64	1826723.91
##	2016	1.490043e+09	7663.8469	194425	-282520.22	7436451.90
##	2017	3.109879e+09	9873.8554	314961	-158259.02	3109707.79
##	2018	1.353888e+09	10224.9675	132410	-1028716.91	2414647.46
##	2019	1.575312e+09	10102.7491	155929	-375819.53	3193192.17
##	2020	1.869458e+09	6828.0488	273791	-866454.00	4210202.46
##	2021	1.448145e+09	8022.1647	180518	-264418.00	4901277.00
##	2022	2.885284e+08	9792.5747	29464	-1267.00	1685912.00

En la tabla anteriormente presentada podemos observar por año de ocurrencia no sólo el monto pagado total sino también el pago promedio, número total de pagos y montos máximos y mínimos de pagos. De esta forma podemos observar que en 2016 se realizó el pago más alto de toda la cartera con un monto de 7,436,452 mientras que en 2017 fue el año en el que se resentó el mayor número de pagos realizados con 314,961. Por el lado del pago promedio, este fue el más alto para el año 2018 con un monto promedio de 10,225.

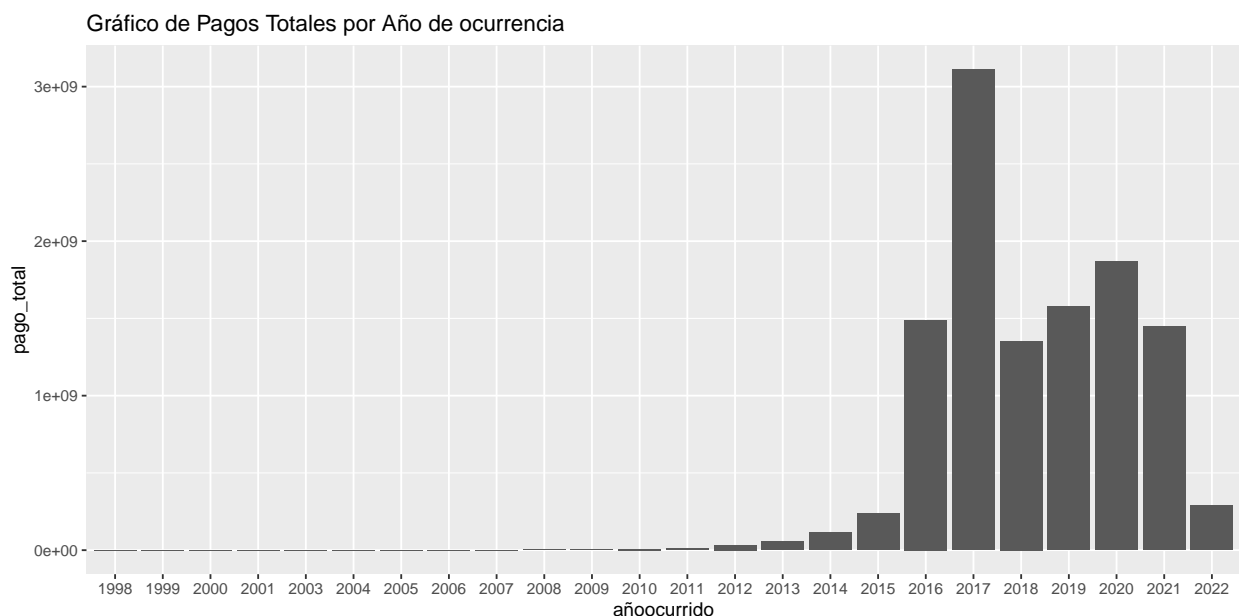
Ahora veamos graficamente el comportamiento del monto total pagado por año y del número de pagos realizados por año de ocurrencia:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

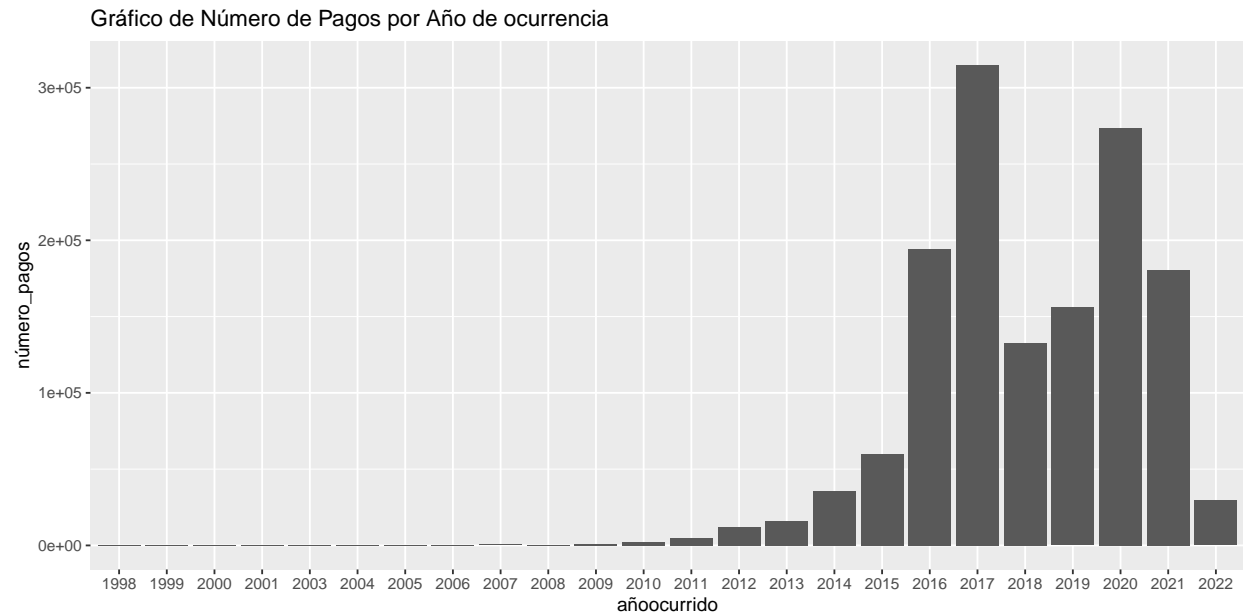
```
# monto total pagado por año de ocurrencia
```

```
ggplot(resumenx año, aes(x=añooccurrido, y=pago_total))+
  geom_bar(stat="identity") +
  labs(title = "Gráfico de Pagos Totales por Año de ocurrencia")
```



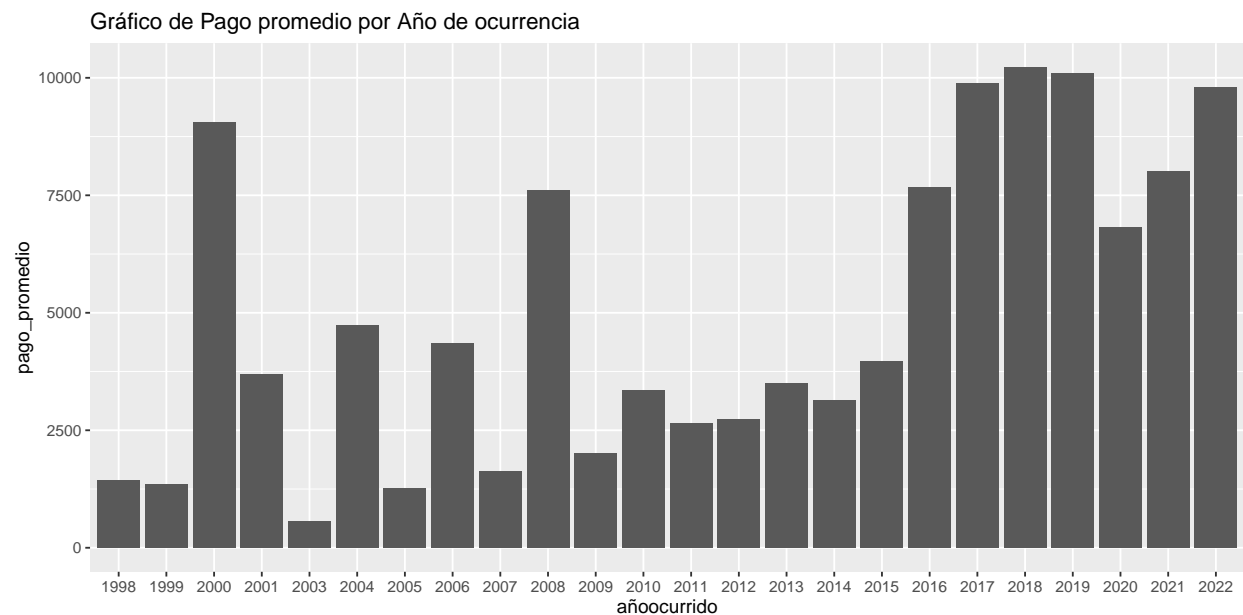
```
# número de pagos por año de ocurrencia
```

```
ggplot(resumenx año, aes(x=añooccurrido, y=número_pagos))+  
  geom_bar(stat="identity") +  
  labs(title = "Gráfico de Número de Pagos por Año de ocurrencia")
```



```
# pago promedio por año de ocurrencia
```

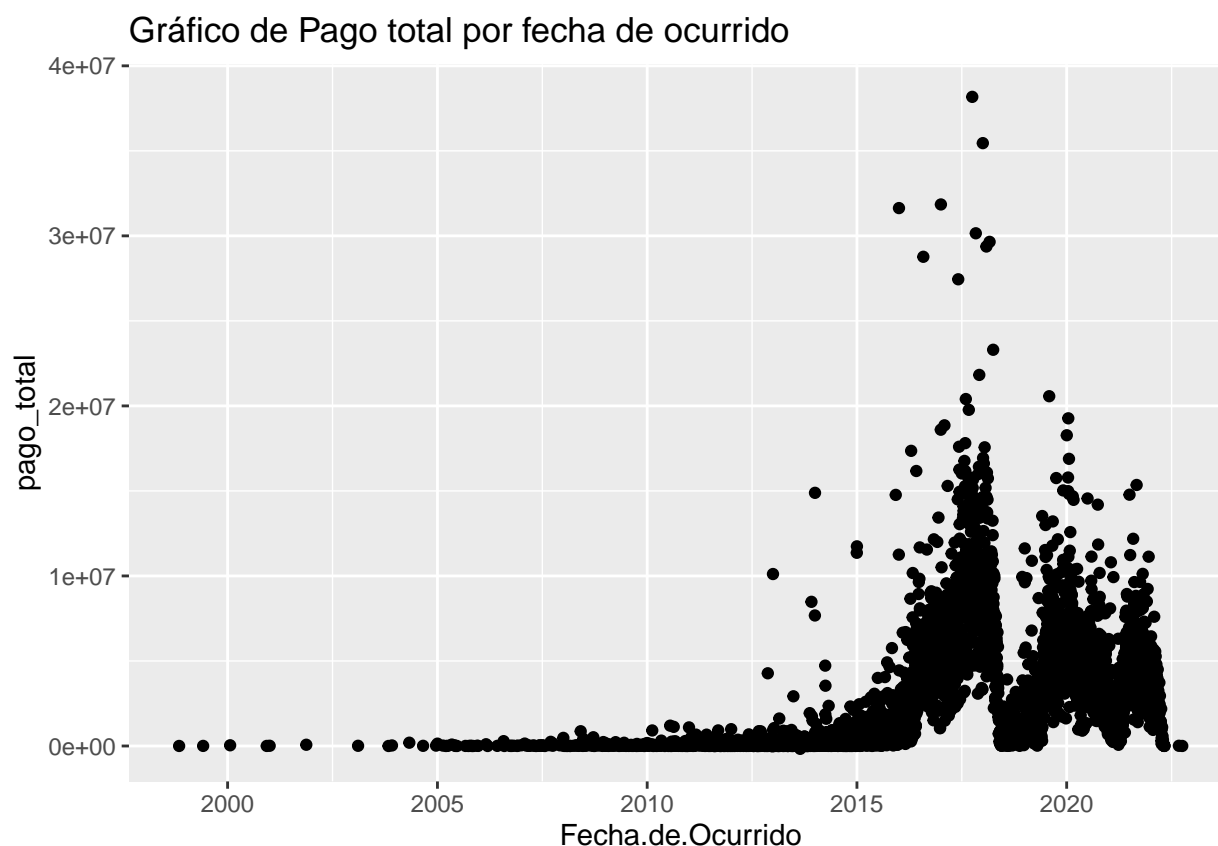
```
ggplot(resumenx año, aes(x=añooccurrido, y=pago_promedio))+  
  geom_bar(stat="identity") +  
  labs(title = "Gráfico de Pago promedio por Año de ocurrencia")
```



A partir de estas gráficas podemos observar que el año 2000 fue atípico en comparación con el periodo de 15 años por arriba y por abajo pues en este año se presenta un número bajo de pagos realizados con 5 pero el monto promedio de estos pagos es superior a cualquier otro realizado entre los años 1998 y 2016.

También podemos observar que el número de pagos realizados ha ido creciendo conforme más nos acercamos a 2021 siendo los años entre 2016 y 2021 los que presentan un mayor número de pagos realizados y esto directamente genera que este mismo periodo de años tenga los montos de pago total por año más altos. Veamos este comportamiento en el siguiente gráfico en donde registro tras registro de fecha de ocurrido vemos el comportamiento de los montos totales pagados

```
resumen3=siniestrosacumulado %>% group_by(Fecha.de.Ocurrido) %>%
  summarise(pago_total = sum(Pagos), número_pagos=n())
ggplot(resumen3,aes(x=Fecha.de.Ocurrido,y=pago_total))+
  geom_point() +
  labs(title = "Gráfico de Pago total por fecha de ocurrido")
```



Ahora veamos el comportamiento poniendo atención en el sexo del asegurado:

```
resumenxaño1 = siniestrosacumulado %>% group_by(añooccurrido, Sexo) %>%
  summarise(pago_total = sum(Pagos), pago_promedio=mean(Pagos), número_pagos=n(),
    pago_mínimo=min(Pagos), pago_máximo=max(Pagos), )
```

'summarise()' has grouped output by 'añooccurrido'. You can override using the '.groups' argument.

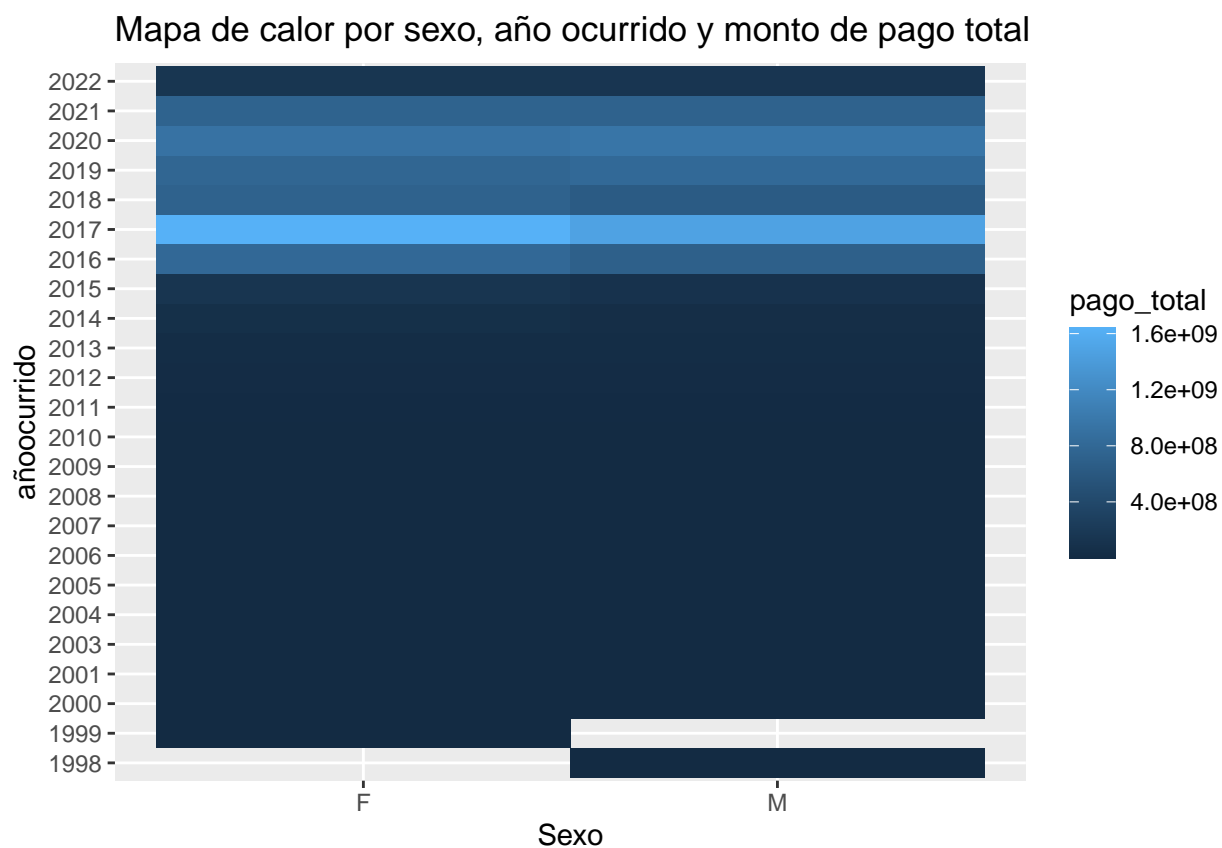
```
print(knitr::kable(resumenx año1, n=52))
```

```
##
##
## | añooccurrido | Sexo | pago_total | pago_promedio | número_pagos | pago_mínimo | pago_máximo |
## | :----- | :--- | :----- | :----- | :----- | :----- | :----- |
## | 1998 | M | 2.859680e+03 | 1429.8400 | 2 | 960.97 | 1898.71 |
## | 1999 | F | 5.375780e+03 | 1343.9450 | 4 | 1033.94 | 1710.94 |
## | 2000 | F | 4.052640e+03 | 1350.8800 | 3 | 542.63 | 2177.07 |
## | 2000 | M | 4.125481e+04 | 20627.4050 | 2 | 2682.19 | 38572.62 |
## | 2001 | F | 1.249768e+04 | 1249.7680 | 10 | 730.04 | 1922.27 |
## | 2001 | M | 7.226470e+04 | 5558.8231 | 13 | 261.00 | 58086.93 |
## | 2003 | F | 4.363335e+04 | 559.4019 | 78 | 15.00 | 1241.67 |
## | 2003 | M | 8.366000e+02 | 836.6000 | 1 | 836.60 | 836.60 |
## | 2004 | F | 1.042118e+04 | 5210.5900 | 2 | 2297.96 | 8123.22 |
## | 2004 | M | 2.165478e+05 | 4707.5609 | 46 | 47.00 | 9007.18 |
## | 2005 | F | 1.610525e+05 | 988.0523 | 163 | 1.00 | 11590.18 |
## | 2005 | M | 1.712635e+05 | 1712.6353 | 100 | 3.00 | 19196.00 |
## | 2006 | F | 3.015117e+05 | 13109.2061 | 23 | 168.00 | 177005.63 |
## | 2006 | M | 1.891130e+05 | 2101.2552 | 90 | 100.00 | 25620.18 |
## | 2007 | F | 5.798777e+05 | 1563.0128 | 371 | 2.00 | 80867.77 |
## | 2007 | M | 4.146264e+05 | 1713.3321 | 242 | 14.00 | 89784.07 |
## | 2008 | F | 1.914059e+06 | 8431.9772 | 227 | -31513.06 | 456206.83 |
## | 2008 | M | 1.695026e+06 | 6862.4520 | 247 | 8.00 | 107217.26 |
## | 2009 | F | 1.000652e+06 | 2021.5198 | 495 | 3.00 | 17112.98 |
## | 2009 | M | 1.330177e+06 | 2009.3306 | 662 | 1.00 | 92825.35 |
## | 2010 | F | 3.494693e+06 | 3007.4810 | 1162 | -1926.98 | 124472.18 |
## | 2010 | M | 4.117200e+06 | 3695.8709 | 1114 | -40684.06 | 374389.16 |
## | 2011 | F | 6.554830e+06 | 2653.7775 | 2470 | 1.00 | 163976.35 |
## | 2011 | M | 5.952901e+06 | 2649.2663 | 2247 | -3555.71 | 160672.53 |
## | 2012 | F | 1.592147e+07 | 2490.0639 | 6394 | -1033.03 | 395536.07 |
## | 2012 | M | 1.652656e+07 | 3009.7536 | 5491 | -32641.06 | 311654.18 |
## | 2013 | F | 2.709955e+07 | 3358.0605 | 8070 | -28461.00 | 777978.06 |
## | 2013 | M | 2.833628e+07 | 3650.1710 | 7763 | -161422.00 | 1346443.70 |
## | 2014 | F | 6.693702e+07 | 3390.0743 | 19745 | -10419.00 | 765810.55 |
## | 2014 | M | 4.571299e+07 | 2838.9633 | 16102 | -16626.28 | 355552.12 |
## | 2015 | F | 1.348508e+08 | 4109.1755 | 32817 | -6253.15 | 1826723.91 |
## | 2015 | M | 1.040755e+08 | 3814.6640 | 27283 | -171912.64 | 1039876.69 |
## | 2016 | F | 7.969761e+08 | 7395.6379 | 107763 | -282520.22 | 3320344.70 |
## | 2016 | M | 6.930673e+08 | 7997.3612 | 86662 | -188719.03 | 7436451.90 |
## | 2017 | F | 1.637625e+09 | 9230.1132 | 177422 | -158259.02 | 3109707.79 |
## | 2017 | M | 1.472254e+09 | 10704.2674 | 137539 | -124796.12 | 2632299.79 |
## | 2018 | F | 7.180148e+08 | 9716.5584 | 73896 | -319620.21 | 2414647.46 |
## | 2018 | M | 6.358731e+08 | 10867.0258 | 58514 | -1028716.91 | 2286621.10 |
## | 2019 | F | 7.642476e+08 | 8765.5137 | 87188 | -190386.82 | 3058543.03 |
## | 2019 | M | 8.110640e+08 | 11798.8385 | 68741 | -375819.53 | 3193192.17 |
## | 2020 | F | 9.196282e+08 | 5950.4376 | 154548 | -866454.00 | 4210202.46 |
## | 2020 | M | 9.498301e+08 | 7965.4995 | 119243 | -412920.76 | 3017210.62 |
## | 2021 | F | 7.318626e+08 | 7170.6947 | 102063 | -117704.00 | 2534985.00 |
## | 2021 | M | 7.162825e+08 | 9129.8516 | 78455 | -264418.00 | 4901277.00 |
## | 2022 | F | 1.472253e+08 | 8873.8063 | 16591 | -1267.00 | 1147544.00 |
## | 2022 | M | 1.413031e+08 | 10976.7033 | 12873 | 1.00 | 1685912.00 |
```

Al ser una tabla bastante grande, veamos graficamente si el sexo tiene alguna influencia en el número de pagos realizados por año y en el monto total de pagos realizados por año.

Mapa de calor- Impacto del sexo en el monto total de pago por año de ocurrencia

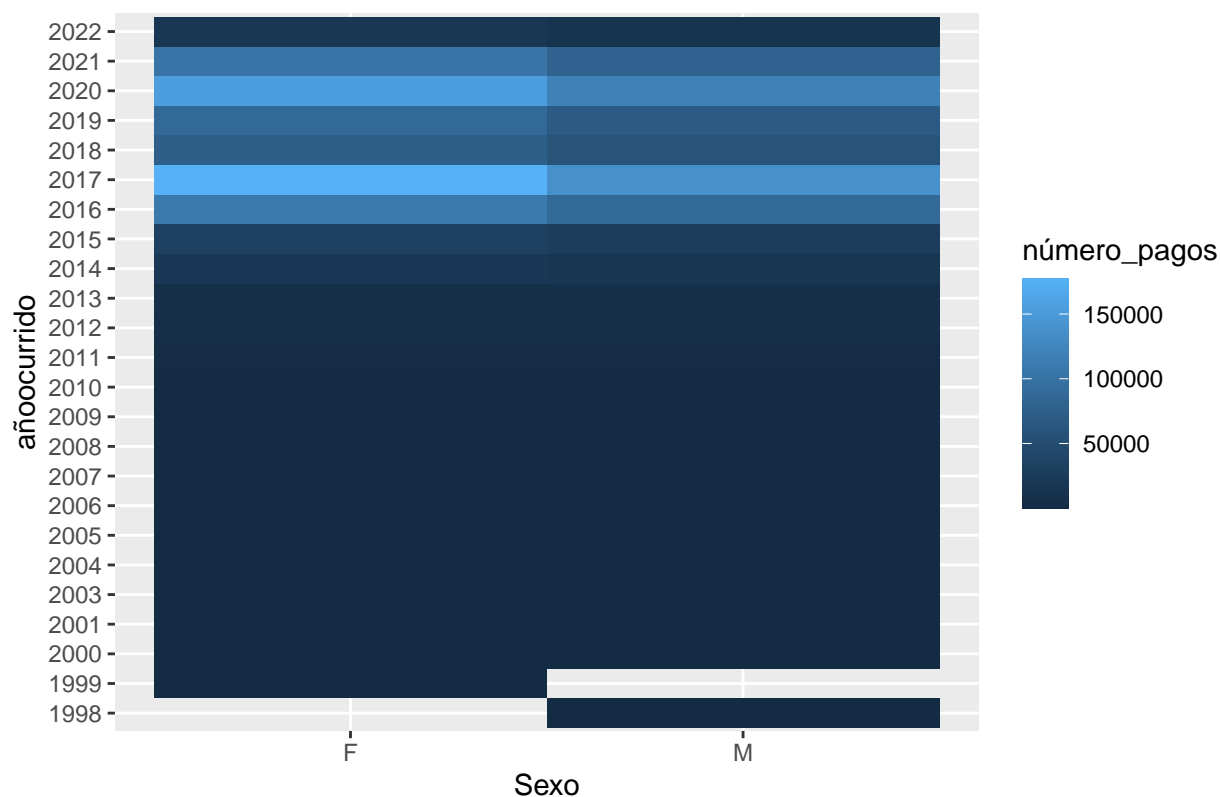
```
ggplot(resumenx año1, aes(x=Sexo, y=añooccurrido, fill=pago_total)) +  
  geom_tile() +  
  labs(title = "Mapa de calor por sexo, año ocurrido y monto de pago total")
```



Mapa de calor- Impacto del sexo en el número total de pagos por año de ocurrencia

```
ggplot(resumenx año1, aes(x=Sexo, y=añooccurrido, fill=número_pagos)) +  
  geom_tile() +  
  labs(title = "Mapa de calor por sexo, año ocurrido y número de pagos")
```


Mapa de calor por sexo, año ocurrido y número de pagos



A partir de estos gráficos observamos que el sexo no juega un papel importante en como se comporta el número de pagos realizados ni para el monto total pagado. Sólo podemos observar que del año 2017 en adelante, el sexo femenino generó un mayor número de pagos que el masculino en la gran mayoría de los años en especial en 2017.

Ahora veamos el ranking de asegurados con el pago más alto para cada uno de los años de ocurrencia:

```
resumenxañoaseg = siniestrosacumulado %>% group_by(añooccurido,Nombre) %>%
  summarise(pago_total = sum(Pagos),pago_promedio=mean(Pagos), número_pagos=n())
```

'summarise()' has grouped output by 'añooccurido'. You can override using the '.groups' argument.

```
resumen_aseg<-resumenxañoaseg%>%
  group_by (añooccurido)%>%
  filter (pago_total == max (pago_total, na.rm = TRUE ))
print(knitr::kable(resumen_aseg, n=30))
```

```
##
##
## |añooccurido |Nombre | pago_total| pago_promedio| número_pagos|
## |:-----:|:-----:|:-----:|:-----:|:-----:|
## |1998      |LOJIRU | 2859.68| 1429.8400| 2|
## |1999      |GOSTYU | 2630.90| 1315.4500| 2|
## |2000      |GALOBK | 41254.81| 20627.4050| 2|
## |2001      |LUCERA | 60614.49| 15153.6225| 4|
```

##	2003	DEDAAL	4423.42	737.2367	6
##	2004	GOCANA	38354.64	6392.4400	6
##	2005	R.PEOS	22997.00	11498.5000	2
##	2006	ROMALA	179178.63	35835.7260	5
##	2007	PAPESO	89784.07	89784.0700	1
##	2008	LIAHHE	456206.83	456206.8300	1
##	2009	CENOJO	191160.08	47790.0200	4
##	2010	DEY CL	748103.32	374051.6600	2
##	2011	ESTRPA	351559.14	43944.8925	8
##	2012	PRTEJO	676729.78	112788.2967	6
##	2013	RIJIIV	2797303.56	466217.2600	6
##	2014	JIGANA	1393672.80	46455.7600	30
##	2015	SARIMI	4309042.16	143634.7387	30
##	2016	CAMAJU	14984366.69	881433.3347	17
##	2017	ESDIAL	6291919.72	629191.9720	10
##	2018	CASOMA	5363216.82	893869.4700	6
##	2019	MAGABR	3196085.17	532680.8617	6
##	2020	SAJACE	4376227.68	437622.7680	10
##	2021	MIBRAI	5017603.00	385969.4615	13
##	2022	HEROIN	1687785.00	337557.0000	5

Notemos que en los años 2014 y 2015 existieron asegurados a los cuales les hicieron hasta 30 pagos. Adicionalmente no se encuentran asegurados que hayan repetido en generar los montos de pagos más altos por año. Por el lado del comparativo de montos, una referencia más precisa se generaría si todos los valores son llevados a alguna fecha en particular.

Proyección año 2022

Considerando una inflación del 4% anual crearemos una base idéntica que nos servirá de soporte para no modificar la original de siniestros acumulada. A esta base le llamaremos “base_soporte”.

```
base_soporte<-siniestrosacumulado
#glimpse(base_soporte)
```

En esta base crearemos una nueva columna llamada “Difvs_2022” la cual nos indicará cuántos años de diferencia contra 2022 tiene cada uno de los registros tomando como base su año de ocurrencia. Utilizaremos esta columna para generar la variable llamada “pagos_a2022” la cual presenta el monto de pago a año 2022 de cada uno de los registros tomando como referencia la inflación del 4% anual. La fórmula para calcular el monto de pago a año 2022 es la siguiente:

```
# pagos_2022 i = pagos i *(1.04)^(2022 - año ocurrido i) donde i corresponde a la i-ésima fila

base_soporte$añooccurrido<-as.numeric(as.character(base_soporte$añooccurrido))
base_soporte$Difvs_2022<-2022 - base_soporte$añooccurrido
base_soporte$pagos_a2022<-base_soporte$Pagos*(1.04)^base_soporte$Difvs_2022
glimpse(base_soporte)
```

```
## Rows: 1,414,937
## Columns: 10
## $ Nombre      <chr> "RACAWI", "GOCAPA", "VIDIMA", "VIDIMA", "GOMEJO..."
## $ Fecha.de.Emission <date> 2001-02-01, 2002-03-03, 1998-06-01, 1998-06-01...
## $ Inicio.de.Vigencia <date> 2003-02-01, 2004-03-03, 2004-06-01, 2004-06-01...
```

```
## $ Sexo <fct> F, M, M, M, F, F, F, F, M, M, M, M, F, M, M, M,...
## $ Fecha.de.Nacimiento <date> 1978-01-16, 1952-09-21, 1947-05-05, 1947-05-05...
## $ Fecha.de.Ocurrido <date> 2003-12-02, 2004-05-01, 2004-12-20, 2004-12-20...
## $ Pagos <dbl> 527.00, 8329.62, 1490.69, 1683.74, 500.00, 648....
## $ añoocurrido <dbl> 2003, 2004, 2004, 2004, 2005, 2005, 2005, 2005,...
## $ Difvs_2022 <dbl> 19, 18, 18, 18, 17, 17, 17, 17, 17, 17, 17, 16,...
## $ pagos_a2022 <dbl> 1110.3095, 16874.2818, 3019.8644, 3410.9483, 97...
```

Al interesarnos encontrar una proyección para 2022 podemos usar desde metodologías que usen algoritmos de machine learning, ajustar distribuciones al número de pagos y monto de los mismos o como se muestra a continuación, para el número de pagos tomaremos el promedio de los últimos 5 años (2017 a 2021) y para el monto de los mismos los extraeremos con una muestra aleatoria de la columna “pagos_a2022”

```
#primero veamos el promedio del número de pagos
promedio<- mean(resumenx año$ número_pagos[19:23])
promedio
```

```
## [1] 211521.8
```

En promedio para los años 2017 a 2021 se generaron 211522 pagos por año por lo que ahora generemos una muestra aleatoria de 211522 montos de pago:

```
muestra_pagos2022 <- sample(base_soporte$pagos_a2022, 211522)
```

De la muestra obtenida podemos observar lo siguiente:

```
#El monto total que se espera pagar en 2022 es:
sum(muestra_pagos2022)
```

```
## [1] 2029643906
```

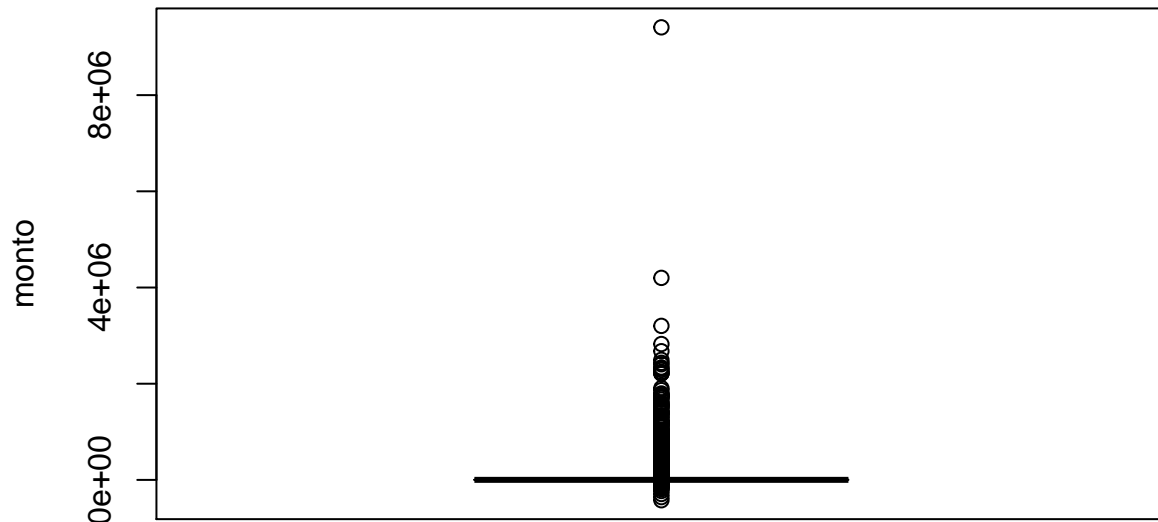
```
#El pago promedio para 2022 es:
mean(muestra_pagos2022)
```

```
## [1] 9595.427
```

Y por el lado del comportamiento de estos montos de pagos para 2022 observamos en el siguiente boxplot que se presentan colas pesadas y datos que parecieran atípicos pero al tratarse de pagos a siniestros, estos valores no pueden ser eliminados pues contienen información que pudiera servir para futuros análisis

```
boxplot(muestra_pagos2022,main = "Boxplot del monto de pagos para 2022",
        xlab = "Pagos en 2022", ylab="monto")
```

Boxplot del monto de pagos para 2022



Pagos en 2022

Conclusiones

El pago promedio de la base de siniestros ha ido incremenando año tras año aún sin considerar el impacto de la inflación por lo que esto puede deberse a la inflación médica. Por el lado del sexo de los reclamantes, no existe una relación clara en cómo impacta el sexo al monto del pago reclamado, sólo para el año 2017 se observa una ligera relación entre el sexo femenino y la generación de montos y números de pagos totales más altos.

Por el lado de los asegurados, se identificaron aquellos que generaron el monto de pago más alto para cada uno de los años de ocurrencia. No se encontraron asegurados que hayan aparecido dos veces en el ranking por lo que un análisis más profundo y sin tomar en cuenta los años de ocurrido puede realizarse para identificar aquellos asegurados que generaron los montos más grandes de la base.

Tomando en cuenta el comportamiento de los años 2017 a 2021 se predijo el número de pagos y el monto de los mismos para 2022 obteniendo una distribución de pagos con colas pesadas y alta variabilidad en los montos más altos de pago por lo que nos da indicios de que para 2022 pudieran existir reclamaciones de siniestros catastróficos. El comportamiento de 2022 mantiene un monto de pago promedio similar a los 5 años anteriores y se espera que el número de pagos sea superior al año 2021.