Table of Contents

Para este ejercicio vamos a utilizar datos reales provistos por el Ayuntamiento de Madrid, hablamos de los llamados datos abiertos u Open Data, en este caso, vamos a trabajar con las activaciones realizadas al SAMUR, es decir, cuando un ciudadano llama a la Protección Civil para una urgencia se realiza una activación al SAMUR, cuando eso sucede, se recopilan los siguientes datos:

Año Mes Hora de solicitud Hora de intervención Código de la emergencia Distrito en el que se produce la emergencia Hospital al que se traslada al ciudadano (si procede)

Concretamente, se utilizarán datos de activación al SAMUR con los años de 2017 a 2021, se dispondrá de los siguientes archivos.csv

activaciones\_samur\_2017.csv activaciones\_samur\_2018.csv activaciones\_samur\_2019.csv activaciones\_samur\_2020.csv activaciones\_samur\_2021.csv

La fuente original de los datos puede encontrarse en el siguiente enlace: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=50d7d35982d6f510VgnVCM1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default>

Teniendo en cuenta la siguiente información se pide lo siguiente:

1.Lee todos los archivos como dataframes

Primero importamos todas las librerias que vamos a necesitar

library(tidyr)

library(dplyr)

##   
## 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

library(ggplot2)

Ahora leeos los archivos csv (al descargarlos en W los edité para ver si estaban todos en el mismo código y condicones guardados, tuve que cambiar algunos para que se leyesen corectamente)

samur1 <- read.csv("activaciones\_samur\_2017 .csv", header = TRUE, sep = ";",encoding = "Latin-1")

head(samur1,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2017 ENERO 0:23:19 0:28:59 Heridas  
## 2 2017 ENERO 0:27:35 0:35:44 Agresión sin especificar  
## 3 2017 ENERO 0:47:26 0:55:49 Intoxicación etílica  
## 4 2017 ENERO 0:55:13 1:02:23 Agresión sin especificar  
## 5 2017 ENERO 1:07:11 1:19:44 Accidente menos de 3 victimas  
## 6 2017 ENERO 1:21:14 1:25:14 Inconsciente sin filiar causa  
## 7 2017 ENERO 1:22:50 1:24:52 Intoxicación etílica  
## 8 2017 ENERO 1:37:16 1:43:57 Casual: caída, etc  
## 9 2017 ENERO 1:47:34 Apertura de puerta  
## 10 2017 ENERO 1:49:08 1:57:38 Intoxicación etílica  
## Distrito Hospital  
## 1 CENTRO Concepción (Fund. J. Díaz)  
## 2 CARABANCHEL   
## 3 SALAMANCA   
## 4 CENTRO Doce de Octubre  
## 5 VILLA DE VALLECAS   
## 6 CENTRO   
## 7 CENTRO Clínico San Carlos  
## 8 VILLA DE VALLECAS INFANTA LEONOR  
## 9 TETUAN   
## 10 CENTRO

samur2 <- read.csv("activaciones\_samur\_2018.csv", header = TRUE, sep = ";",encoding ="Latin-1")

head(samur2,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2018 ENERO 0:07:43 0:42:53 Violencia de genero  
## 2 2018 ENERO 0:11:15 0:13:55 Casual: caída, etc  
## 3 2018 ENERO 0:13:16 0:13:16 Patología cardiovascular  
## 4 2018 ENERO 0:16:48 0:21:52 Certificación psiquiátrico  
## 5 2018 ENERO 0:17:11 0:31:31 Patología cardiovascular  
## 6 2018 ENERO 0:24:05 0:29:40 Patología cardiovascular  
## 7 2018 ENERO 0:25:37 0:31:54 Incendio  
## 8 2018 ENERO 0:27:43 0:34:59 Diabético  
## 9 2018 ENERO 0:32:27 0:42:58 Casual: caída, etc  
## 10 2018 ENERO 0:35:23 0:37:51 Otros  
## Distrito Hospital  
## 1 SALAMANCA   
## 2 CARABANCHEL   
## 3 CENTRO   
## 4 MORATALAZ   
## 5 HORTALEZA La Paz  
## 6 CENTRO   
## 7 CARABANCHEL   
## 8 CENTRO   
## 9 CIUDAD LINEAL Ramón y Cajal  
## 10 CENTRO Doce de Octubre

samur3 <- read.csv("activaciones\_samur\_2019.csv", header = TRUE, sep = ";",encoding ="Latin-1")

head(samur3,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2019 ENERO 0:08:09 0:08:15 Intoxicación etílica  
## 2 2019 ENERO 0:09:13 Violencia de genero  
## 3 2019 ENERO 0:24:12 0:24:22 Otros  
## 4 2019 ENERO 0:28:55 Intoxicación etílica  
## 5 2019 ENERO 0:29:11 0:34:02 Casual: caída, etc  
## 6 2019 ENERO 0:32:55 0:38:36 Patología respiratoria  
## 7 2019 ENERO 0:34:23 0:34:31 Heridas  
## 8 2019 ENERO 0:35:36 0:42:10 Accidente de motocicleta  
## 9 2019 ENERO 0:37:43 0:43:59 Accidente de motocicleta  
## 10 2019 ENERO 0:40:41 0:48:33 Accidente de motocicleta  
## Distrito Hospital  
## 1 CENTRO   
## 2 VALLECAS PTE.   
## 3 CENTRO   
## 4 CENTRO   
## 5 CENTRO   
## 6 RETIRO Gregorio Marañón  
## 7 CENTRO   
## 8 RETIRO Doce de Octubre  
## 9 RETIRO Doce de Octubre  
## 10 RETIRO Doce de Octubre

samur4 <- read.csv("activaciones\_samur\_2020.csv", header = TRUE, sep = ";",encoding ="Latin-1")

head(samur4,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2020 ENERO 0:03:21 0:08:49 Accidente de monopatin  
## 2 2020 ENERO 0:04:03 0:14:44 Patología obstétrica  
## 3 2020 ENERO 0:04:05 0:10:12 Inconsciente sin filiar causa  
## 4 2020 ENERO 0:05:09 0:13:15 Incendio  
## 5 2020 ENERO 0:07:43 0:16:19 Agresión sin especificar  
## 6 2020 ENERO 0:17:50 0:23:58 Intoxicación etílica  
## 7 2020 ENERO 0:24:11 0:28:35 Agresión sin especificar  
## 8 2020 ENERO 0:26:28 0:35:09 Incendio  
## 9 2020 ENERO 0:37:46 Otros  
## 10 2020 ENERO 0:38:53 0:46:15 Intoxicación etílica  
## Distrito Hospital  
## 1 ARGANZUELA   
## 2 SAN BLAS   
## 3 VILLA DE VALLECAS   
## 4 VALLECAS PTE.   
## 5 VALLECAS PTE.   
## 6 MONCLOA Clínico San Carlos  
## 7 CENTRO   
## 8 RETIRO   
## 9 CENTRO   
## 10 CHAMARTIN La Princesa

samur5 <- read.csv("activaciones\_samur\_2021.csv", header = TRUE, sep = ";",encoding ="Latin-1")

head(samur5,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2021 ENERO 0:04:31 0:18:46 Accidente menos de 3 victimas  
## 2 2021 ENERO 0:05:47 0:17:08 Sobredosis  
## 3 2021 ENERO 0:11:30 0:18:01 Accidente menos de 3 victimas  
## 4 2021 ENERO 0:11:47 0:16:46 Incendio  
## 5 2021 ENERO 0:14:23 0:22:30 Patología digestiva  
## 6 2021 ENERO 0:16:25 Incendio  
## 7 2021 ENERO 0:16:54 0:21:49 Intoxicación etílica  
## 8 2021 ENERO 0:23:24 0:26:35 Accidente menos de 3 victimas  
## 9 2021 ENERO 0:23:25 0:31:06 Heridas  
## 10 2021 ENERO 0:30:37 0:33:06 Intoxicación etílica  
## Distrito Hospital  
## 1 MORATALAZ Gregorio Marañón  
## 2 FUENCARRAL   
## 3 VALLECAS PTE.   
## 4 LATINA   
## 5 VILLAVERDE Doce de Octubre  
## 6 BARAJAS   
## 7 CIUDAD LINEAL   
## 8 MORATALAZ Gregorio Marañón  
## 9 VICALVARO La Paz  
## 10 CIUDAD LINEAL

samur6 <- read.csv("activaciones\_samur\_2022.csv", header = TRUE, sep = ";",encoding ="Latin-1")

head(samur6,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2022 ENERO 0:06:15 0:20:32 Patología psicosomática  
## 2 2022 ENERO 0:06:15 0:20:32 Patología psicosomática  
## 3 2022 ENERO 0:07:17 0:15:07 Intoxicación etílica  
## 4 2022 ENERO 0:07:23 0:12:50 Intoxicación etílica  
## 5 2022 ENERO 0:07:44 0:13:31 Intoxicación etílica  
## 6 2022 ENERO 0:08:13 0:13:24 Intoxicación etílica  
## 7 2022 ENERO 0:09:07 Casual: caída, etc  
## 8 2022 ENERO 0:09:47 0:09:49 Casual: caída, etc  
## 9 2022 ENERO 0:10:00 0:18:22 Casual: caída, etc  
## 10 2022 ENERO 0:26:33 0:28:55 Intoxicación etílica  
## Distrito Hospital  
## 1 CENTRO   
## 2 CENTRO   
## 3 CIUDAD LINEAL   
## 4 VALLECAS PTE.   
## 5 TETUAN   
## 6 CENTRO   
## 7 VILLA DE VALLECAS   
## 8 CENTRO   
## 9 VILLA DE VALLECAS   
## 10 CHAMARTIN

Una vez leídos concatenamos todos los archivo en uno solo

myfiles= list.files(path = "C:/Users/Usuario/Desktop/Máster\_IMF/Ejercicios\_R", pattern= ".csv", full.names= TRUE)

myfiles

## character(0)

samurtotal <- data.frame()  
for (i in 1:6) {  
 tab <- get(paste0("samur",i))  
 samurtotal <- rbind(samurtotal, tab)  
}

head(samurtotal,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2017 ENERO 0:23:19 0:28:59 Heridas  
## 2 2017 ENERO 0:27:35 0:35:44 Agresión sin especificar  
## 3 2017 ENERO 0:47:26 0:55:49 Intoxicación etílica  
## 4 2017 ENERO 0:55:13 1:02:23 Agresión sin especificar  
## 5 2017 ENERO 1:07:11 1:19:44 Accidente menos de 3 victimas  
## 6 2017 ENERO 1:21:14 1:25:14 Inconsciente sin filiar causa  
## 7 2017 ENERO 1:22:50 1:24:52 Intoxicación etílica  
## 8 2017 ENERO 1:37:16 1:43:57 Casual: caída, etc  
## 9 2017 ENERO 1:47:34 Apertura de puerta  
## 10 2017 ENERO 1:49:08 1:57:38 Intoxicación etílica  
## Distrito Hospital  
## 1 CENTRO Concepción (Fund. J. Díaz)  
## 2 CARABANCHEL   
## 3 SALAMANCA   
## 4 CENTRO Doce de Octubre  
## 5 VILLA DE VALLECAS   
## 6 CENTRO   
## 7 CENTRO Clínico San Carlos  
## 8 VILLA DE VALLECAS INFANTA LEONOR  
## 9 TETUAN   
## 10 CENTRO

summary(samurtotal)

## Año Mes Hora.Solicitud Hora.Intervención   
## Min. :2017 Length:701818 Length:701818 Length:701818   
## 1st Qu.:2018 Class :character Class :character Class :character   
## Median :2019 Mode :character Mode :character Mode :character   
## Mean :2019   
## 3rd Qu.:2020   
## Max. :2022   
## Código Distrito Hospital   
## Length:701818 Length:701818 Length:701818   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

1. Modifica los tipos de las columnas:

* Año: Entero (ya es entero) Mes: Factor Hora.Solicitud: Carater (ya es character) Hora.Intervencion: Caracter (ya es character) Codigo: Caracter(ya es character) Distrito: Factor Hospital: Factor

samurtotal$Mes <- as.factor(samurtotal$Mes)  
samurtotal$Distrito <- as.factor(samurtotal$Distrito)  
samurtotal$Hospital <- as.factor (samurtotal$Hospital)

str(samurtotal)

## 'data.frame': 701818 obs. of 7 variables:  
## $ Año : int 2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...  
## $ Mes : Factor w/ 12 levels "ABRIL","AGOSTO",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Hora.Solicitud : chr "0:23:19" "0:27:35" "0:47:26" "0:55:13" ...  
## $ Hora.Intervención: chr "0:28:59" "0:35:44" "0:55:49" "1:02:23" ...  
## $ Código : chr "Heridas" "Agresión sin especificar" "Intoxicación etílica" "Agresión sin especificar" ...  
## $ Distrito : Factor w/ 27 levels "","ARGANZUELA",..: 7 5 20 7 26 7 7 26 22 7 ...  
## $ Hospital : Factor w/ 24 levels "","Alcalá de Henares (Ppe. de Asturias)",..: 7 1 1 8 1 1 6 12 1 1 ...

1. Agrupa el dataframe por el código de la urgencia, ¿Cuáles son los tres códigos de urgencia más habituales? ¿Cuáles son los tres menos habituales?

samurtotal %>% group\_by(Código)

## # A tibble: 701,818 × 7  
## # Groups: Código [98]  
## Año Mes Hora.Solicitud Hora.Intervención Código Distrito Hospital  
## <int> <fct> <chr> <chr> <chr> <fct> <fct>   
## 1 2017 ENERO 0:23:19 "0:28:59" Heridas CENTRO "Concep…  
## 2 2017 ENERO 0:27:35 "0:35:44" Agresión sin … CARABAN… ""   
## 3 2017 ENERO 0:47:26 "0:55:49" Intoxicación … SALAMAN… ""   
## 4 2017 ENERO 0:55:13 "1:02:23" Agresión sin … CENTRO "Doce d…  
## 5 2017 ENERO 1:07:11 "1:19:44" Accidente men… VILLA D… ""   
## 6 2017 ENERO 1:21:14 "1:25:14" Inconsciente … CENTRO ""   
## 7 2017 ENERO 1:22:50 "1:24:52" Intoxicación … CENTRO "Clínic…  
## 8 2017 ENERO 1:37:16 "1:43:57" Casual: caída… VILLA D… "INFANT…  
## 9 2017 ENERO 1:47:34 "" Apertura de p… TETUAN ""   
## 10 2017 ENERO 1:49:08 "1:57:38" Intoxicación … CENTRO ""   
## # … with 701,808 more rows

samurtotal %>% count(Código, sort = TRUE)

## Código n  
## 1 Casual: caída, etc 118922  
## 2 Patología cardiovascular 114552  
## 3 Intoxicación etílica 56897  
## 4 Accidente menos de 3 victimas 36595  
## 5 Agresión sin especificar 36505  
## 6 Heridas 32116  
## 7 Otros 31978  
## 8 Accidente de motocicleta 27407  
## 9 Patología psicosomática 24252  
## 10 Certificación psiquiátrico 21103  
## 11 Incendio 17661  
## 12 Apertura de puerta 14755  
## 13 Convulsión y/o epilepsia 13733  
## 14 Patología digestiva 10523  
## 15 Atropello 10411  
## 16 Colaboración del FOXTROP en una actuación 10099  
## 17 Patología respiratoria 10086  
## 18 Inconsciente sin filiar causa 9063  
## 19 1 SVB 8919  
## 20 Violencia de genero 7708  
## 21 Parada Cardiorespiratoria 6631  
## 22 Accidente de bicicleta 6366  
## 23 Diabético 5458  
## 24 Patología neurológica 5273  
## 25 Orden medica 4838  
## 26 Servicios de seguimiento de riesgos 4130  
## 27 Sobredosis 3787  
## 28 Patología infecciosa 3677  
## 29 Preventivos 3525  
## 30 Serv. Preventivo desde oper. Ordinario 3074  
## 31 Precipitado 2916  
## 32 Patología anafiláctica 2901  
## 33 Arma blanca 2629  
## 34 Intoxicación por fármacos y/o productos tóxicos 2355  
## 35 Servicios especiales 2266  
## 36 Maltrato a menores 2256  
## 37 Infarto o patología coronaria 2115  
## 38 Patología urológica 1852  
## 39 Hasta 3 SVB 1534  
## 40 Accidente de monopatin 1427  
## 41 Agresión sexual 1423  
## 42 Policía nacional 1327  
## 43 Patología obstétrica 1233  
## 44 Patología ginecológica 1086  
## 45 Policía municipal 1068  
## 46 1 o + SVB y 1 o + SVA y PMA 1008  
## 47 Autolisis traumática 1007  
## 48 + de 3 SVB y eq. Sanit. 856  
## 49 Accidente con 3 o más víctimas confirmadas 811  
## 50 Agresión con spray de defensa 717  
## 51 1 o + SVB y 1 o + SVA 604  
## 52 Orden judicial 572  
## 53 Autolisis no traumática 540  
## 54 Determinación de PCR (covid 19) en casos excepcionales 521  
## 55 Incendio con Victimas confirmadas 517  
## 56 Accidente de autobús / autocar 495  
## 57 SAMUR 474  
## 58 Arrollamiento metro y/o tren 473  
## 59 Atención psicológica a Familiares y allegados 447  
## 60 Accidente sin especificar 353  
## 61 Atrapamientos/enterramientos 348  
## 62 NO APARECE EN CODIGO INICIAL 340  
## 63 Riesgo químico 326  
## 64 Bombero de Madrid 313  
## 65 Problema social 307  
## 66 Quemado 297  
## 67 Otras instituciones 255  
## 68 Arma de fuego 195  
## 69 Explosión 188  
## 70 Maltrato a ancianos 186  
## 71 Ahogado 173  
## 72 Accidente con vehículo pesado 154  
## 73 Paciente en huelga de hambre 142  
## 74 Electrocutado 111  
## 75 84  
## 76 Orden médica urgente /agresivo 76  
## 77 Servicios de análisis de riesgos 73  
## 78 Atención psicológica a la Víctima primaria 69  
## 79 Sumisión Química 55  
## 80 Vía publica 46  
## 81 Derrumbamientos 44  
## 82 Heridas complicadas 41  
## 83 Domicilio 32  
## 84 Accidente de avión 22  
## 85 Atención psicológica a Implicados. 16  
## 86 Todos los medios de SAMUR-PC 15  
## 87 Accidente de tren 14  
## 88 Riesgo biológico 14  
## 89 Local público (centro sanita.) 13  
## 90 Atención psicológica a intervinientes de SAMUR-PC. 12  
## 91 Riesgo público 9  
## 92 Atención psicológica a Testigos. 5  
## 93 Persona “sin techo” fallecida en vía pública 5  
## 94 No suceso 4  
## 95 Acc. Trasp. Merc. Peligrosas 3  
## 96 Atención psicológica a intervinientes de Cuerpos de Seguridad 2  
## 97 Atención psicológica a intervinientes de Bomberos. 1  
## 98 Pacientes en RCP prolongada 1

``

Los tres primeros son Casual: caída, etc.= 118.922 Patalogía cardiovascular = 114.552 Intoxicación etítica =56.897

Los tres últimos son Atención psicológica a intervinientes de Cuerpos de Seguridad = 2 Atención psicológica a intervinientes de Bomberos = 1 Pacientes en RCP prolongada = 1

head (samurtotal %>% count(Código, sort = TRUE),3)

## Código n  
## 1 Casual: caída, etc 118922  
## 2 Patología cardiovascular 114552  
## 3 Intoxicación etílica 56897

tail (samurtotal %>% count(Código, sort = TRUE),3)

## Código n  
## 96 Atención psicológica a intervinientes de Cuerpos de Seguridad 2  
## 97 Atención psicológica a intervinientes de Bomberos. 1  
## 98 Pacientes en RCP prolongada 1

5.Observa los valores únicos de la variable Distrito, parece que hay un nivel (categoría) que es una cadena de texto vacía, asigna valor nulo a todas las cadenas de texto vacías para esta variable, ¿cuántos nulos aparecen? Elimina todos los valores nulos que hayan aparecido a raíz de esta transformación.

unique(samurtotal$Distrito)

## [1] CENTRO CARABANCHEL   
## [3] SALAMANCA VILLA DE VALLECAS   
## [5] TETUAN VALLECAS PTE.   
## [7] LATINA MORATALAZ   
## [9] VILLAVERDE SAN BLAS   
## [11] CIUDAD LINEAL MONCLOA   
## [13] CHAMBERI CHAMARTIN   
## [15] USERA HORTALEZA   
## [17] BARAJAS ARGANZUELA   
## [19] RETIRO FUENCARRAL   
## [21] VICALVARO   
## [23] POZUELO FUERA TERMINO MUNICIPAL   
## [25] CARRETERAS Y CIRCUNVALACIONES C.A.M.   
## [27] LEGANES   
## 27 Levels: ARGANZUELA BARAJAS C.A.M. ... VILLAVERDE

summary(samurtotal)

## Año Mes Hora.Solicitud Hora.Intervención   
## Min. :2017 MARZO : 65580 Length:701818 Length:701818   
## 1st Qu.:2018 ENERO : 65302 Class :character Class :character   
## Median :2019 FEBRERO : 62902 Mode :character Mode :character   
## Mean :2019 JUNIO : 60983   
## 3rd Qu.:2020 OCTUBRE : 60559   
## Max. :2022 DICIEMBRE: 59609   
## (Other) :326883   
## Código Distrito Hospital   
## Length:701818 CENTRO :100939 :484090   
## Class :character CARABANCHEL : 47955 Gregorio Marañón : 35434   
## Mode :character VALLECAS PTE.: 43499 Doce de Octubre : 32966   
## MONCLOA : 43109 Concepción (Fund. J. Díaz): 31587   
## LATINA : 40434 La Paz : 26528   
## TETUAN : 38418 Clínico San Carlos : 26184   
## (Other) :387464 (Other) : 65029

samurtotal$Distrito[samurtotal$Distrito == ""] <- NA

sum(is.na(samurtotal$Distrito))

## [1] 5720

Aparecen 5720

samurtotal <- samurtotal %>% drop\_na(Distrito)

Cuando los elimino, tenemos 696098 líneas, antes teníamos 701818 (696.098+5.720 = 701.818)

summary(samurtotal)

## Año Mes Hora.Solicitud Hora.Intervención   
## Min. :2017 MARZO : 65295 Length:696098 Length:696098   
## 1st Qu.:2018 ENERO : 64988 Class :character Class :character   
## Median :2019 FEBRERO : 62644 Mode :character Mode :character   
## Mean :2019 JUNIO : 60497   
## 3rd Qu.:2020 OCTUBRE : 60199   
## Max. :2022 DICIEMBRE: 59017   
## (Other) :323458   
## Código Distrito Hospital   
## Length:696098 CENTRO :100939 :479103   
## Class :character CARABANCHEL : 47955 Gregorio Marañón : 35286   
## Mode :character VALLECAS PTE.: 43499 Doce de Octubre : 32867   
## MONCLOA : 43109 Concepción (Fund. J. Díaz): 31481   
## LATINA : 40434 La Paz : 26445   
## TETUAN : 38418 Clínico San Carlos : 26105   
## (Other) :381744 (Other) : 64811

1. Muestra los valores únicos de la variable Hospital, si hay algún factor que este vacío o que indique valor nulo, esto querrá decir que esa urgencia no requirió de hospitalización, para todos estos casos crea un nuevo factor o categoría que sea NO\_HOSPITAL ¿cuántos casos hay en el dataframe que no requirieron de hospitalización?

unique(samurtotal$Hospital)

## [1] Concepción (Fund. J. Díaz)   
## [2]   
## [3] Doce de Octubre   
## [4] Clínico San Carlos   
## [5] INFANTA LEONOR   
## [6] Gregorio Marañón   
## [7] Central de la Defensa   
## [8] Ramón y Cajal   
## [9] La Paz   
## [10] La Princesa   
## [11] Niño Jesús (Infantil)   
## [12] Getafe   
## [13] Doctor Rodriguez Lafora   
## [14] Centros Municipales (Asist. y Albergues)  
## [15] SANTA CRISTINA   
## [16] Otros Centros Públicos   
## [17] Puerta de Hierro   
## [18] VIRGEN DE LA TORRE   
## [19] Militares (Aire,Polic. Naval, Otros...)   
## [20] Alcalá de Henares (Ppe. de Asturias)   
## [21] Centros Municipales (Asist. y Albergues)   
## [22] Moncloa (ASISA)   
## [23] Móstoles   
## [24] Leganés (Severo Ochoa)   
## 24 Levels: Alcalá de Henares (Ppe. de Asturias) ... VIRGEN DE LA TORRE

head(samurtotal,10)

## Año Mes Hora.Solicitud Hora.Intervención Código  
## 1 2017 ENERO 0:23:19 0:28:59 Heridas  
## 2 2017 ENERO 0:27:35 0:35:44 Agresión sin especificar  
## 3 2017 ENERO 0:47:26 0:55:49 Intoxicación etílica  
## 4 2017 ENERO 0:55:13 1:02:23 Agresión sin especificar  
## 5 2017 ENERO 1:07:11 1:19:44 Accidente menos de 3 victimas  
## 6 2017 ENERO 1:21:14 1:25:14 Inconsciente sin filiar causa  
## 7 2017 ENERO 1:22:50 1:24:52 Intoxicación etílica  
## 8 2017 ENERO 1:37:16 1:43:57 Casual: caída, etc  
## 9 2017 ENERO 1:47:34 Apertura de puerta  
## 10 2017 ENERO 1:49:08 1:57:38 Intoxicación etílica  
## Distrito Hospital  
## 1 CENTRO Concepción (Fund. J. Díaz)  
## 2 CARABANCHEL   
## 3 SALAMANCA   
## 4 CENTRO Doce de Octubre  
## 5 VILLA DE VALLECAS   
## 6 CENTRO   
## 7 CENTRO Clínico San Carlos  
## 8 VILLA DE VALLECAS INFANTA LEONOR  
## 9 TETUAN   
## 10 CENTRO

summary(samurtotal$Hospital)

##   
## 479103   
## Alcalá de Henares (Ppe. de Asturias)   
## 8   
## Central de la Defensa   
## 9980   
## Centros Municipales (Asist. y Albergues)   
## 58   
## Centros Municipales (Asist. y Albergues)   
## 22   
## Clínico San Carlos   
## 26105   
## Concepción (Fund. J. Díaz)   
## 31481   
## Doce de Octubre   
## 32867   
## Doctor Rodriguez Lafora   
## 67   
## Getafe   
## 29   
## Gregorio Marañón   
## 35286   
## INFANTA LEONOR   
## 13131   
## La Paz   
## 26445   
## La Princesa   
## 17205   
## Leganés (Severo Ochoa)   
## 5   
## Militares (Aire,Polic. Naval, Otros...)   
## 6   
## Moncloa (ASISA)   
## 55   
## Móstoles   
## 2   
## Niño Jesús (Infantil)   
## 2018   
## Otros Centros Públicos   
## 42   
## Puerta de Hierro   
## 40   
## Ramón y Cajal   
## 22123   
## SANTA CRISTINA   
## 16   
## VIRGEN DE LA TORRE   
## 4

Tenemos 479.103 casos de no hospitalización

samurtotal$Hospital <- replace(samurtotal$Hospital,samurtotal$Hospital == "", "NO\_HOSITAL")

## Warning in `[<-.factor`(`\*tmp\*`, list, value = "NO\_HOSITAL"): invalid factor  
## level, NA generated

summary(samurtotal)

## Año Mes Hora.Solicitud Hora.Intervención   
## Min. :2017 MARZO : 65295 Length:696098 Length:696098   
## 1st Qu.:2018 ENERO : 64988 Class :character Class :character   
## Median :2019 FEBRERO : 62644 Mode :character Mode :character   
## Mean :2019 JUNIO : 60497   
## 3rd Qu.:2020 OCTUBRE : 60199   
## Max. :2022 DICIEMBRE: 59017   
## (Other) :323458   
## Código Distrito Hospital   
## Length:696098 CENTRO :100939 Gregorio Marañón : 35286   
## Class :character CARABANCHEL : 47955 Doce de Octubre : 32867   
## Mode :character VALLECAS PTE.: 43499 Concepción (Fund. J. Díaz): 31481   
## MONCLOA : 43109 La Paz : 26445   
## LATINA : 40434 Clínico San Carlos : 26105   
## TETUAN : 38418 (Other) : 64811   
## (Other) :381744 NA's :479103

Me salen ya como NA, no me cambian a NO\_HOSÌTAL, no sé por qué, ya que justo en la línea anterior si que estaba el espacio vacío y debería funcionar.

7.Filtra el dataframe por todos los niveles distintos de NO\_HOSPITAL, muestra los tres hospitales a los que más ciudadanos hayan sido trasladados

head (samurtotal %>% count(Hospital, sort = TRUE),4)

## Hospital n  
## 1 <NA> 479103  
## 2 Gregorio Marañón 35286  
## 3 Doce de Octubre 32867  
## 4 Concepción (Fund. J. Díaz) 31481

Los de mayor afluencia son el Gregorio Marañón, seguido del 12 de Octubre y tercero la Fundación J. Díaz (Concepción)

1. Agrupa el dataframe por la media de casos que se reciben al mes para cada año ¿En qué año se realizaron más activaciones del SAMUR?

samurtotal %>% group\_by(Mes,Año) %>% count(Código, sort = TRUE) %>% summarise (mean (n))

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

## # A tibble: 63 × 3  
## # Groups: Mes [12]  
## Mes Año `mean(n)`  
## <fct> <int> <dbl>  
## 1 ABRIL 2017 151.   
## 2 ABRIL 2018 156.   
## 3 ABRIL 2019 154.   
## 4 ABRIL 2020 70.9  
## 5 ABRIL 2021 117.   
## 6 AGOSTO 2017 122.   
## 7 AGOSTO 2018 137.   
## 8 AGOSTO 2019 147.   
## 9 AGOSTO 2020 120.   
## 10 AGOSTO 2021 134.   
## # … with 53 more rows

samurtotal %>% group\_by(Año) %>% count(Código) %>% summarise(mean(n))

## # A tibble: 6 × 2  
## Año `mean(n)`  
## <int> <dbl>  
## 1 2017 1562.  
## 2 2018 1630.  
## 3 2019 1686.  
## 4 2020 1286.  
## 5 2021 1415.  
## 6 2022 394.

El año con más avisos ha sido el 2019

1. Transforma la columna Hora.Solicitud en horas, minutos y segundos de solicitud, si aparece algún valor nulo elimínalo.

samurtotal <- samurtotal %>% separate(Hora.Solicitud,c("horas", "minutos", "segundos"), sep= ":", remove = TRUE)

## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 7315 rows [433,  
## 434, 435, 436, 437, 1052, 1434, 1435, 1436, 1806, 1807, 2080, 2367, 2368, 2369,  
## 2683, 2684, 3016, 3017, 3018, ...].

Vemos que si que aparecen valores nulos, los identificamos y eliminamos.

head(samurtotal,10)

## Año Mes horas minutos segundos Hora.Intervención  
## 1 2017 ENERO 0 23 19 0:28:59  
## 2 2017 ENERO 0 27 35 0:35:44  
## 3 2017 ENERO 0 47 26 0:55:49  
## 4 2017 ENERO 0 55 13 1:02:23  
## 5 2017 ENERO 1 07 11 1:19:44  
## 6 2017 ENERO 1 21 14 1:25:14  
## 7 2017 ENERO 1 22 50 1:24:52  
## 8 2017 ENERO 1 37 16 1:43:57  
## 9 2017 ENERO 1 47 34   
## 10 2017 ENERO 1 49 08 1:57:38  
## Código Distrito Hospital  
## 1 Heridas CENTRO Concepción (Fund. J. Díaz)  
## 2 Agresión sin especificar CARABANCHEL <NA>  
## 3 Intoxicación etílica SALAMANCA <NA>  
## 4 Agresión sin especificar CENTRO Doce de Octubre  
## 5 Accidente menos de 3 victimas VILLA DE VALLECAS <NA>  
## 6 Inconsciente sin filiar causa CENTRO <NA>  
## 7 Intoxicación etílica CENTRO Clínico San Carlos  
## 8 Casual: caída, etc VILLA DE VALLECAS INFANTA LEONOR  
## 9 Apertura de puerta TETUAN <NA>  
## 10 Intoxicación etílica CENTRO <NA>

samurtotal$horas<- as.factor(samurtotal$horas)  
samurtotal$minutos <- as.factor(samurtotal$minutos)  
samurtotal$segundos <- as.factor (samurtotal$segundos)

summary(samurtotal)

## Año Mes horas minutos   
## Min. :2017 MARZO : 65295 12 : 46586 26 : 11883   
## 1st Qu.:2018 ENERO : 64988 13 : 45980 29 : 11822   
## Median :2019 FEBRERO : 62644 11 : 41910 28 : 11801   
## Mean :2019 JUNIO : 60497 19 : 41708 32 : 11772   
## 3rd Qu.:2020 OCTUBRE : 60199 14 : 41149 27 : 11761   
## Max. :2022 DICIEMBRE: 59017 20 : 40518 (Other):629744   
## (Other) :323458 (Other):438247 NA's : 7315   
## segundos Hora.Intervención Código Distrito   
## 07 : 11722 Length:696098 Length:696098 CENTRO :100939   
## 21 : 11701 Class :character Class :character CARABANCHEL : 47955   
## 42 : 11696 Mode :character Mode :character VALLECAS PTE.: 43499   
## 51 : 11695 MONCLOA : 43109   
## 53 : 11688 LATINA : 40434   
## (Other):630281 TETUAN : 38418   
## NA's : 7315 (Other) :381744   
## Hospital   
## Gregorio Marañón : 35286   
## Doce de Octubre : 32867   
## Concepción (Fund. J. Díaz): 31481   
## La Paz : 26445   
## Clínico San Carlos : 26105   
## (Other) : 64811   
## NA's :479103

samurtotal$horas[samurtotal$horas == ""] <- NA

samurtotal$minutos[samurtotal$minutos == ""] <- NA

samurtotal$segundos[samurtotal$segundos== ""] <- NA

summary(samurtotal)

## Año Mes horas minutos   
## Min. :2017 MARZO : 65295 12 : 46586 26 : 11883   
## 1st Qu.:2018 ENERO : 64988 13 : 45980 29 : 11822   
## Median :2019 FEBRERO : 62644 11 : 41910 28 : 11801   
## Mean :2019 JUNIO : 60497 19 : 41708 32 : 11772   
## 3rd Qu.:2020 OCTUBRE : 60199 14 : 41149 27 : 11761   
## Max. :2022 DICIEMBRE: 59017 (Other):471450 (Other):629744   
## (Other) :323458 NA's : 7315 NA's : 7315   
## segundos Hora.Intervención Código Distrito   
## 07 : 11722 Length:696098 Length:696098 CENTRO :100939   
## 21 : 11701 Class :character Class :character CARABANCHEL : 47955   
## 42 : 11696 Mode :character Mode :character VALLECAS PTE.: 43499   
## 51 : 11695 MONCLOA : 43109   
## 53 : 11688 LATINA : 40434   
## (Other):630281 TETUAN : 38418   
## NA's : 7315 (Other) :381744   
## Hospital   
## Gregorio Marañón : 35286   
## Doce de Octubre : 32867   
## Concepción (Fund. J. Díaz): 31481   
## La Paz : 26445   
## Clínico San Carlos : 26105   
## (Other) : 64811   
## NA's :479103

samurtotal <- samurtotal %>% drop\_na(segundos)

samurtotal <- samurtotal %>% drop\_na(minutos)

samurtotal <- samurtotal %>% drop\_na(horas)

summary(samurtotal)

## Año Mes horas minutos   
## Min. :2017 MARZO : 64579 12 : 46586 26 : 11883   
## 1st Qu.:2018 ENERO : 64318 13 : 45980 29 : 11822   
## Median :2019 FEBRERO : 62039 11 : 41910 28 : 11801   
## Mean :2019 JUNIO : 59806 19 : 41708 32 : 11772   
## 3rd Qu.:2020 OCTUBRE : 59575 14 : 41149 27 : 11761   
## Max. :2022 DICIEMBRE: 58395 20 : 40518 30 : 11732   
## (Other) :320071 (Other):430932 (Other):618012   
## segundos Hora.Intervención Código Distrito   
## 07 : 11722 Length:688783 Length:688783 CENTRO : 99750   
## 21 : 11701 Class :character Class :character CARABANCHEL : 47457   
## 42 : 11696 Mode :character Mode :character VALLECAS PTE.: 43102   
## 51 : 11695 MONCLOA : 42493   
## 53 : 11688 LATINA : 40000   
## 29 : 11687 TETUAN : 38084   
## (Other):618594 (Other) :377897   
## Hospital   
## Gregorio Marañón : 35286   
## Doce de Octubre : 32866   
## Concepción (Fund. J. Díaz): 31481   
## La Paz : 26443   
## Clínico San Carlos : 26102   
## (Other) : 64810   
## NA's :471795

Una forma opción hubiese sido, antes de eliminar la columna Horas.solicitud, identificar en ella los valores vacíos y ponerlos como NA y eliminarlos. Y después hacer el separate.Vemos que se han elinado porque pasamos además de tener 696098 filas a 688783. Había 7315 nulos dentro de Horas.Solicitud

10.¿En qué hora se han producido más solicitudes al SAMUR?

El mayor número de avisos es a la una (13 horas)

samurtotal %>% group\_by(Mes,horas) %>% count (Código)

## # A tibble: 16,732 × 4  
## # Groups: Mes, horas [288]  
## Mes horas Código n  
## <fct> <fct> <chr> <int>  
## 1 ABRIL 0 + de 3 SVB y eq. Sanit. 6  
## 2 ABRIL 0 1 o + SVB y 1 o + SVA y PMA 1  
## 3 ABRIL 0 1 SVB 3  
## 4 ABRIL 0 Accidente de bicicleta 10  
## 5 ABRIL 0 Accidente de monopatin 3  
## 6 ABRIL 0 Accidente de motocicleta 40  
## 7 ABRIL 0 Accidente menos de 3 victimas 50  
## 8 ABRIL 0 Accidente sin especificar 1  
## 9 ABRIL 0 Agresión con spray de defensa 2  
## 10 ABRIL 0 Agresión sexual 1  
## # … with 16,722 more rows

samurtotal %>% group\_by(horas) %>% count(Código) %>% summarise(mean(n))

## # A tibble: 24 × 2  
## horas `mean(n)`  
## <fct> <dbl>  
## 1 0 255.  
## 2 1 213.  
## 3 10 394.  
## 4 11 461.  
## 5 12 523.  
## 6 13 535.  
## 7 14 484.  
## 8 15 425.  
## 9 16 385.  
## 10 17 411.  
## # … with 14 more rows

El mayor número de incidencias se produce a la una de la tarde

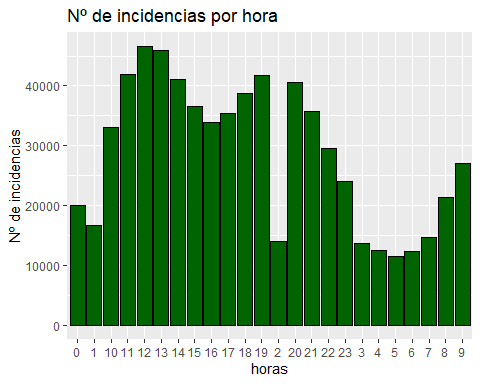
1. Muestra de forma gráfica el número de avisos por hora (Promedio mes).

colnames(samurtotal)

## [1] "Año" "Mes" "horas"   
## [4] "minutos" "segundos" "Hora.Intervención"  
## [7] "Código" "Distrito" "Hospital"

ggplot(samurtotal, aes(x=horas))+geom\_histogram(bins = 24,  
 colour="black", fill="darkgreen",   
 stat="count")+  
 labs(title = "Nº de incidencias por hora")+  
 ylab("Nº de incidencias")

## Warning: Ignoring unknown parameters: binwidth, bins, pad



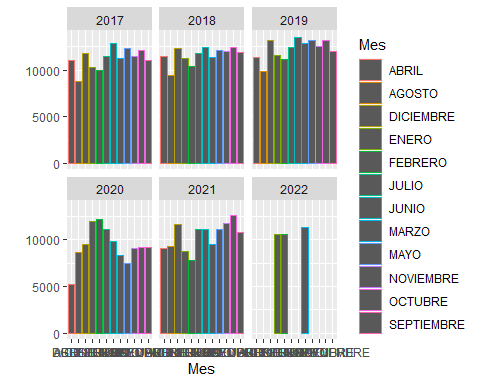
1. Muestra gráficamente los avisos recibidos para cada año en función de cada mes ¿Se aprecian las consecuencias de la pandemia del COVID-19 en las activaciones realizadas a lo largo de los años? (Recuerda que el confinamiento en España comenzó en marzo de 2020)

summary(samurtotal)

## Año Mes horas minutos   
## Min. :2017 MARZO : 64579 12 : 46586 26 : 11883   
## 1st Qu.:2018 ENERO : 64318 13 : 45980 29 : 11822   
## Median :2019 FEBRERO : 62039 11 : 41910 28 : 11801   
## Mean :2019 JUNIO : 59806 19 : 41708 32 : 11772   
## 3rd Qu.:2020 OCTUBRE : 59575 14 : 41149 27 : 11761   
## Max. :2022 DICIEMBRE: 58395 20 : 40518 30 : 11732   
## (Other) :320071 (Other):430932 (Other):618012   
## segundos Hora.Intervención Código Distrito   
## 07 : 11722 Length:688783 Length:688783 CENTRO : 99750   
## 21 : 11701 Class :character Class :character CARABANCHEL : 47457   
## 42 : 11696 Mode :character Mode :character VALLECAS PTE.: 43102   
## 51 : 11695 MONCLOA : 42493   
## 53 : 11688 LATINA : 40000   
## 29 : 11687 TETUAN : 38084   
## (Other):618594 (Other) :377897   
## Hospital   
## Gregorio Marañón : 35286   
## Doce de Octubre : 32866   
## Concepción (Fund. J. Díaz): 31481   
## La Paz : 26443   
## Clínico San Carlos : 26102   
## (Other) : 64810   
## NA's :471795

qplot(data = samurtotal, x = Mes,colour = Mes, stat = "count", facets = ~Año )

## Warning: `stat` is deprecated



Vemos que antes de la pandemia más o menos el número de avisos era similar en el tiempo. A raiz del Covid esto cambió. Disminuyó, sobre todo en el 2020, año del confinamiento en casa. Del 2022 solo tenemos datos hasta marzo, aunque se puede ver que vuelven a aumentar los casos.