

# EDA de Cancelaciones de Hotel

## Proyecto

Cancelaciones en Hoteles Predecir cancelación de reservas en hoteles - AM 2021

## Descripción del problema

Con el fin de planear tarifas y actividades de ventas o promoción, los hoteles hacen estimaciones adelantadas de su ocupación en cada día. Una parte de estas estimaciones requiere predecir cuántas de las reservaciones que ya se tienen van a terminar en cancelaciones, lo cual libera inventario que afecta en la planeación.

## Objetivo

Predecir cuáles reservaciones son probables que terminen o no en cancelación.

## Fuente de datos

Los datos que se utilizaron para este proyecto fueron obtenidos del sitio

<https://www.kaggle.com/c/cancelaciones-en-hoteles/data> (<https://www.kaggle.com/c/cancelaciones-en-hoteles/data>)

Los datos originales provienen de Hotel booking demand datasets, Antonio, de Almeida, Nunes

(<https://www.sciencedirect.com/science/article/pii/S2352340918315191>

(<https://www.sciencedirect.com/science/article/pii/S2352340918315191>))

## Variables

Variable	Tipo	Descripción
ADR	Numeric	Tarifa diaria promedio definida por [5]
Adults	Integer	Número de Adultos
Agent	Categorical	DNI de la agencia de viajes que realizó la reservaa
ArrivalDateDayOfMonth	Integer	Día del mes de la fecha de llegada
ArrivalDateMonth	Categorical	Mes de la fecha de llegada con 12 categorías: "enero" a "diciembre"
ArrivalDateWeekNumber	Integer	Número de semana de la fecha de llegada
ArrivalDateYear	Integer	Año de la fecha de llegada
AssignedRoomType	Categorical	Código del tipo de habitación asignada a la reserva. A veces, el tipo de habitación asignada difiere del tipo de habitación reservada debido a razones de operación del hotel (por ejemplo, overbooking) o por solicitud del cliente. El código se presenta en lugar de la designación por razones de anonimato
Babies	Integer	Numero de bebes
BookingChanges	Integer	Número de cambios / modificaciones realizadas a la reserva desde el momento en que se ingresó la reserva en el PMS hasta el momento del check-in o la cancelación
Children	Integer	Numero de niños
Company	Categorical	DNI de la empresa / entidad que realizó la reserva o responsable del pago de la reserva. La identificación se presenta en lugar de la designación por razones de anonimato
Country	Categorical	País de origen. Las categorías están representadas en el formato ISO 3155-3: 2013 [6]
CustomerType	Categorical	Tipo de reserva, asumiendo una de cuatro categorías:
DaysInWaitingList	Integer	Número de días que la reserva estuvo en lista de espera antes de que fuera confirmada al cliente
DepositType	Categorical	Indicación sobre si el cliente realizó un depósito para garantizar la reserva. Esta variable puede asumir tres categorías:
DistributionChannel	Categorical	Canal de distribución de reservas. El término "TA" significa "Agentes de viajes" y "TO" significa "Operadores turísticos"
IsCanceled	Categorical	Valor que indica si la reserva fue cancelada (1) o no (0)
IsRepeatedGuest	Categorical	Valor que indica si el nombre de la reserva fue de un huésped repetido (1) o no (0)
LeadTime	Integer	Número de días transcurridos entre la fecha de entrada de la reserva en el PMS y la fecha de llegada
MarketSegment	Categorical	Designación de segmento de mercado. En las categorías, el término "TA" significa "Agentes de viajes" y "TO" significa "Operadores turísticos"
Meal	Categorical	Tipo de comida reservada. Las categorías se presentan en paquetes de comidas de hospitalidad estándar:
PreviousBookingsNotCanceled	Integer	Número de reservas anteriores no canceladas por el cliente antes de la reserva actual
PreviousCancellations	Integer	Número de reservas anteriores que fueron canceladas por el cliente antes de la reserva actual

1/12/21 18:56

EDA de Cancelaciones de Hotel

RequiredCardParkingSpaces | Integer | Número de plazas de aparcamiento requeridas por el cliente |

ReservationStatus | Categorical | Último estado de la reserva, asumiendo una de tres categorías: |

ReservationStatusDate | Date | Fecha en la que se estableció el último estado. Esta variable se puede utilizar junto con ReservationStatus para comprender cuándo se canceló la reserva o cuándo se registró el cliente en el hotel. |

ReservedRoomType | Categorical | Código del tipo de habitación reservado. El código se presenta en lugar de la designación por razones de anonimato |

StaysInWeekendNights | Integer | Número de noches de fin de semana (sábado o domingo) que el huésped se hospedó o reservó para alojarse en el hotel |

StaysInWeekNights | Integer | Número de noches de la semana (de lunes a viernes) que el huésped se hospedó o reservó para alojarse en el hotel |

TotalOfSpecialRequests | Integer | Número de solicitudes especiales realizadas por el cliente (por ejemplo, dos camas individuales o piso alto) |

# Analisis Exploratorio de Datos

Con el fin de entender los datos realizamos una revisión general de estos y tratamos de identificar aquellas variables que pudieran ser interesantes para nuestro estudio

## Se leen los datos de Entrenamiento

```
data<-read.csv("entrena.csv",stringsAsFactors = F)
```

Se dividen los datos en tres conjuntos para usarse en los procesos de entrenamiento, prueba y calibración. con una proporción de 70%, 20% y 10%

```
knitr::opts_chunk$set( message = FALSE, results='hide')
idx <- sample(seq(1, 3), size = nrow(data), replace = TRUE, prob = c(.7, 2, .1))
train <- data[idx == 1,]
test <- data[idx == 2,]
cal <- data[idx == 3,]
```

Usamos la función skim en la base de datos de entrenamiento para conocer las características generales de cada variable

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, results = 'show')
skim(train)
```

### Data summary

Name	train
Number of rows	23014
Number of columns	30
Column type frequency:	
character	13
numeric	17
Group variables	
None	

### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
---------------	-----------	---------------	-----	-----	-------	----------	------------

**Variable type:** numeric

Podemos observar que tenemos 13 variables categorías, de las cuales podemos destacar que 3 tienen un número alto de categorías (country, agent, company)

Tenemos 17 variables numéricas.

En este primer acercamiento, podemos identificar que las variables corresponden a:

variables de tiempo: tiempo previo de reservación, fechas de llegada, duración de la reservación.

características de reservación: agencia, país, canal de distribución, segmento de mercado, tipo de depósito, tarifa diaria

características de los clientes y sus preferencias: adultos, bebés, tipo de hotel, tipo de habitación

## Revisamos la proporción de la variable objetivo

```
a <- table(train$is_canceled)
prop <- prop.table(a)
prop
```

```
##
##   cancelado no_cancelado
##   0.3569566   0.6430434
```

Extraemos el subconjunto de cancelados para hacer una revisión de todas las variables con respecto a las reservaciones canceladas.

```
sub_cancelados <- subset(train, is_canceled == "cancelado")
```

Iniciamos con la revisión de los histogramas de cada variable para ver si podemos identificar algun compartamiento interesante

Creamos una funcion para que grafique el histograma en forma ordenada de mayora menor para todas las variables

```
analisis_cancelaciones_ord <- function(datos,nombre) {
  x <- table(datos)
  cancel <- x[order(x,decreasing = TRUE)]
  par(cex=0.5) #control size of labels
  g_1 <- barplot(cancel, main = nombre)
  return(g_1)
}
```

Extraemos el nombre de todas las variables del subconjunto cancelado y se la pasamos a la funcion para su graficación

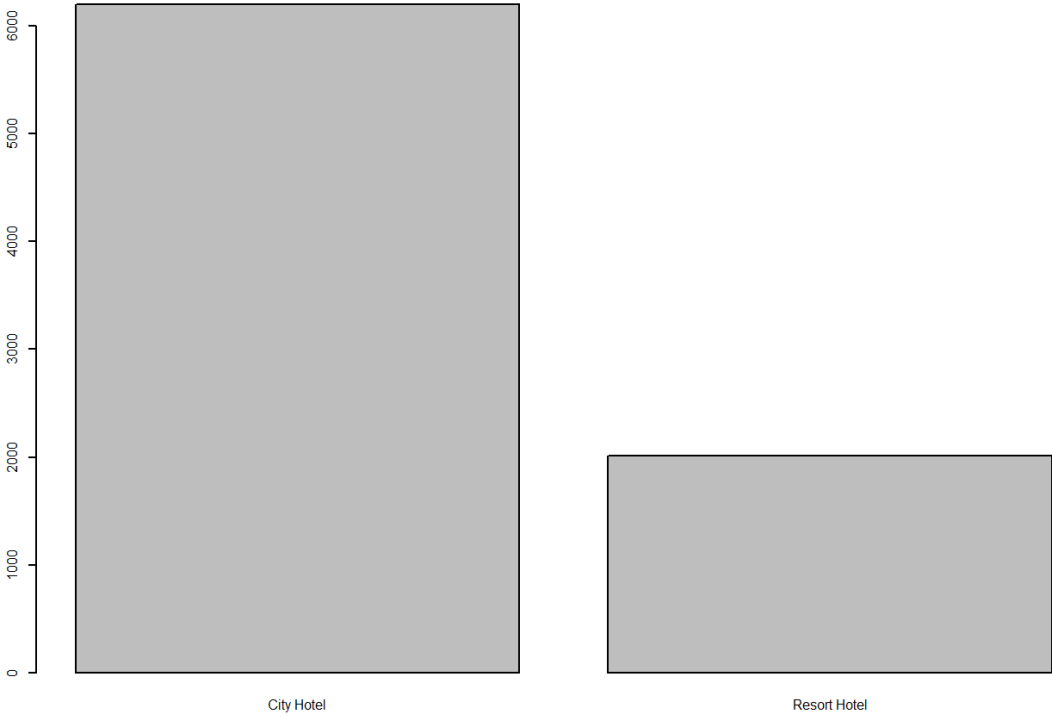
```
ncol(sub_cancelados)
```

```
## [1] 30
```

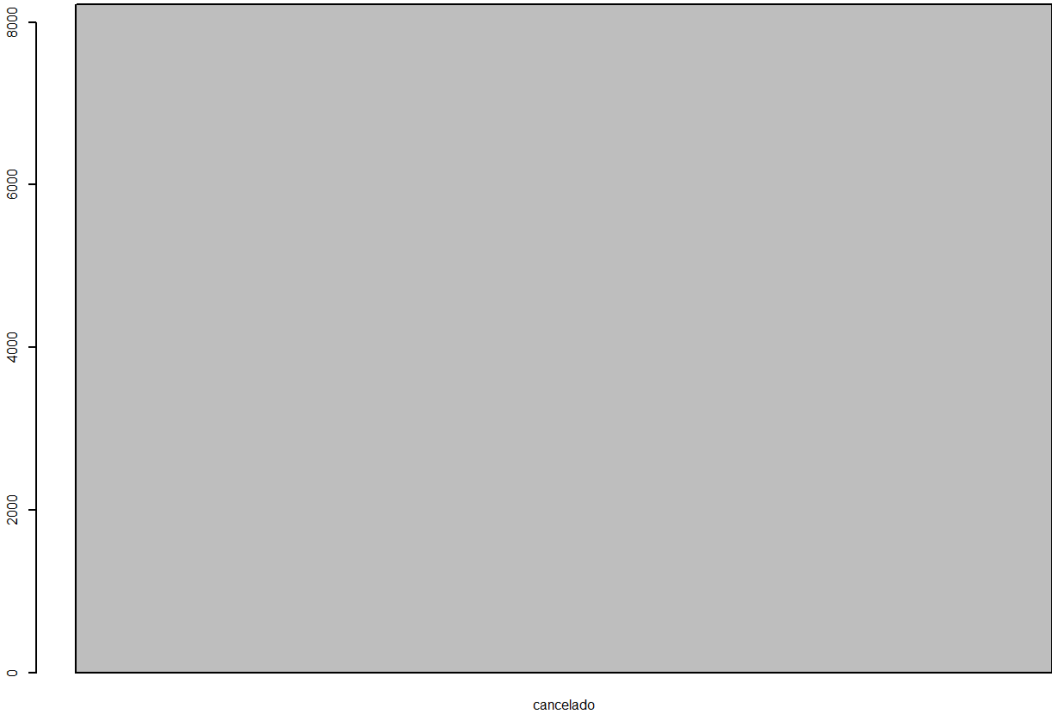
```
nombres <- names(sub_cancelados)
for (i in c(1:ncol(sub_cancelados))){
  analisis_cancelaciones_ord(sub_cancelados[,i],nombres[i])
}
```



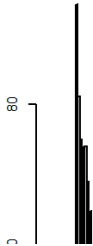
hotel

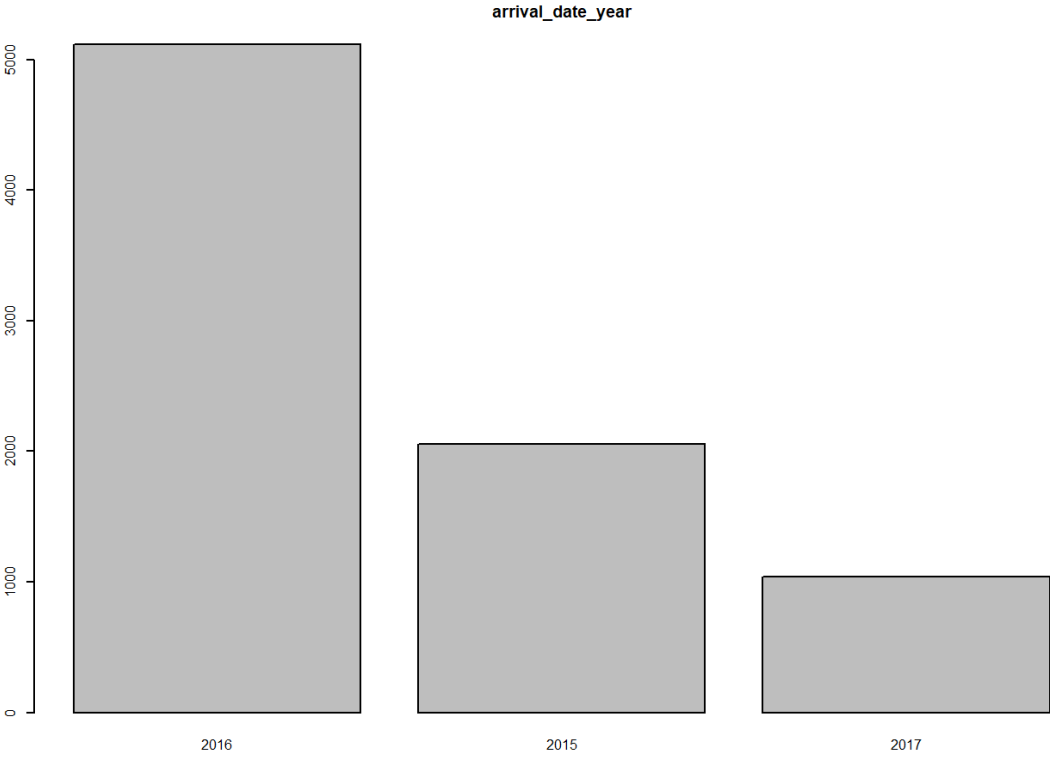
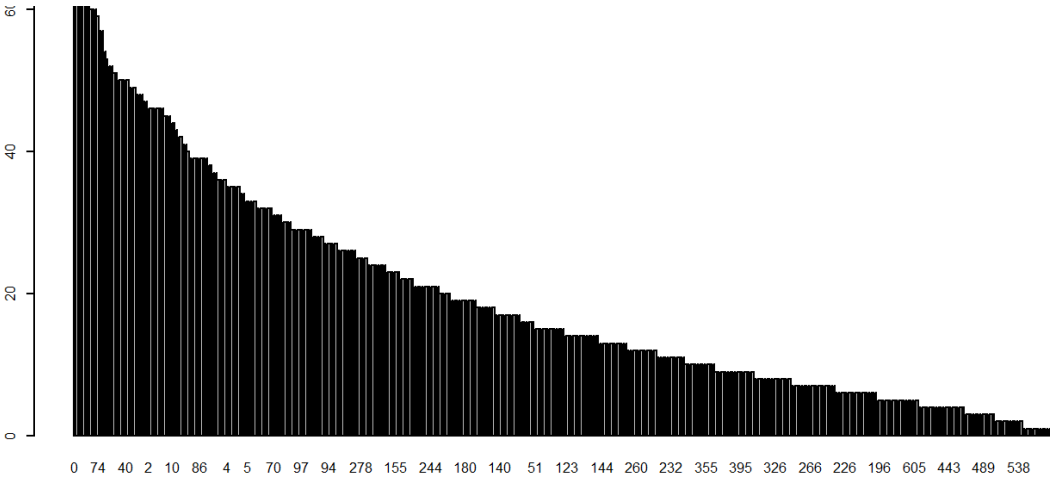


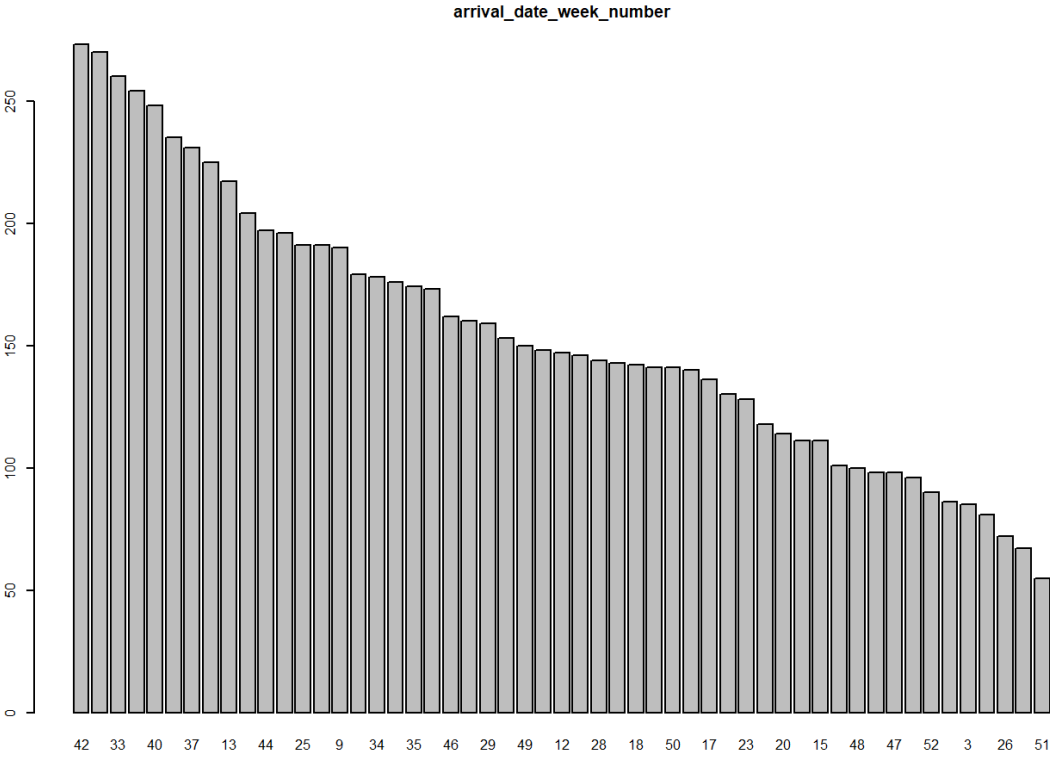
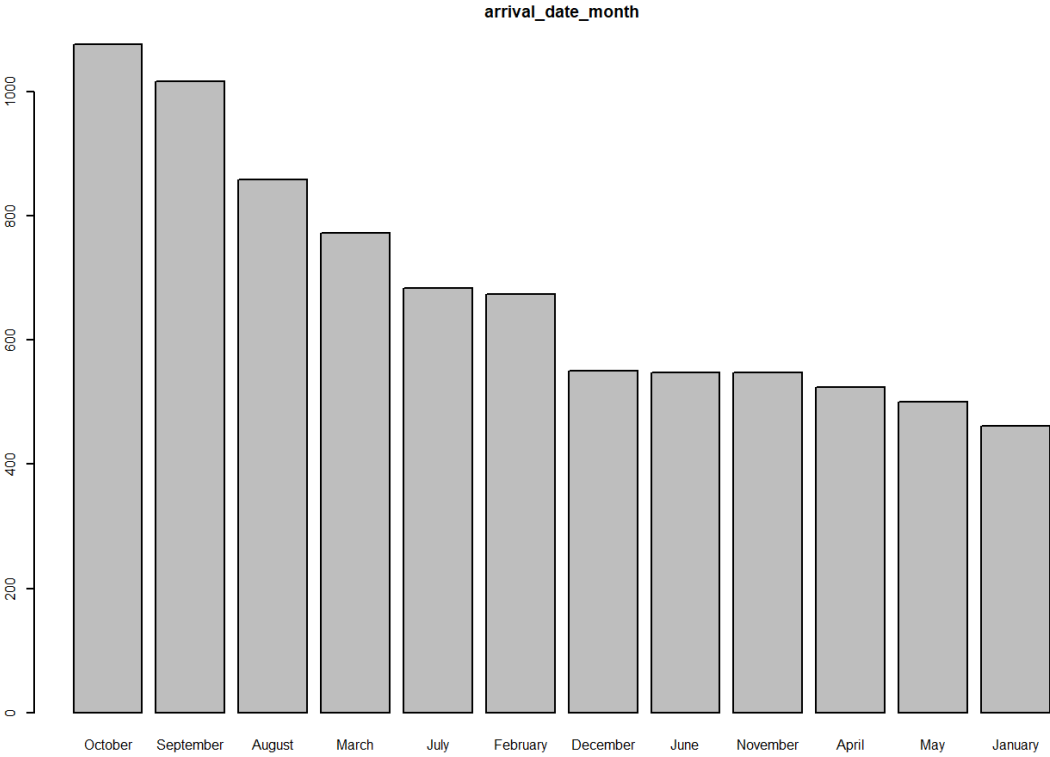
is\_canceled



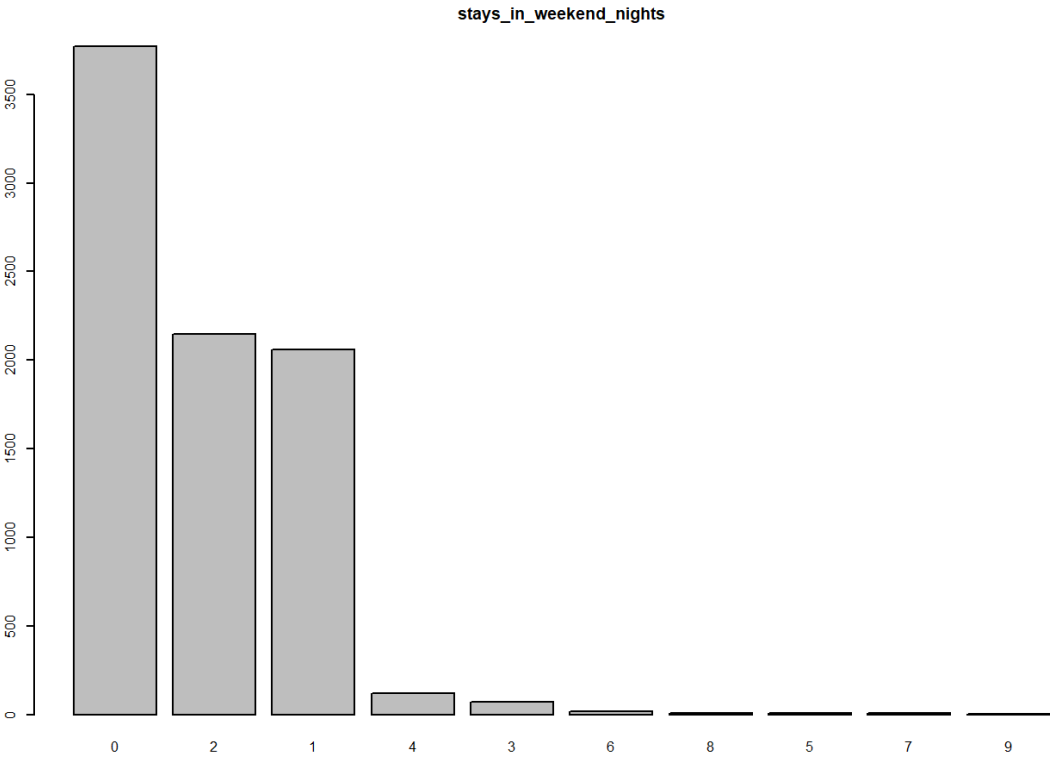
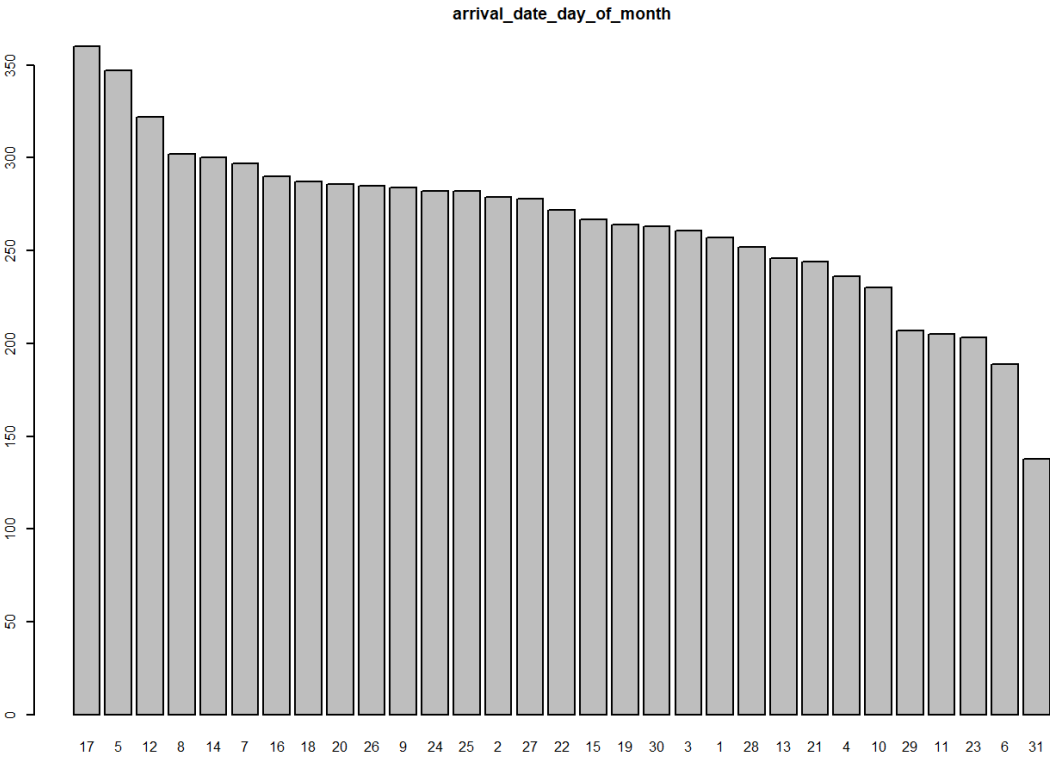
lead\_time

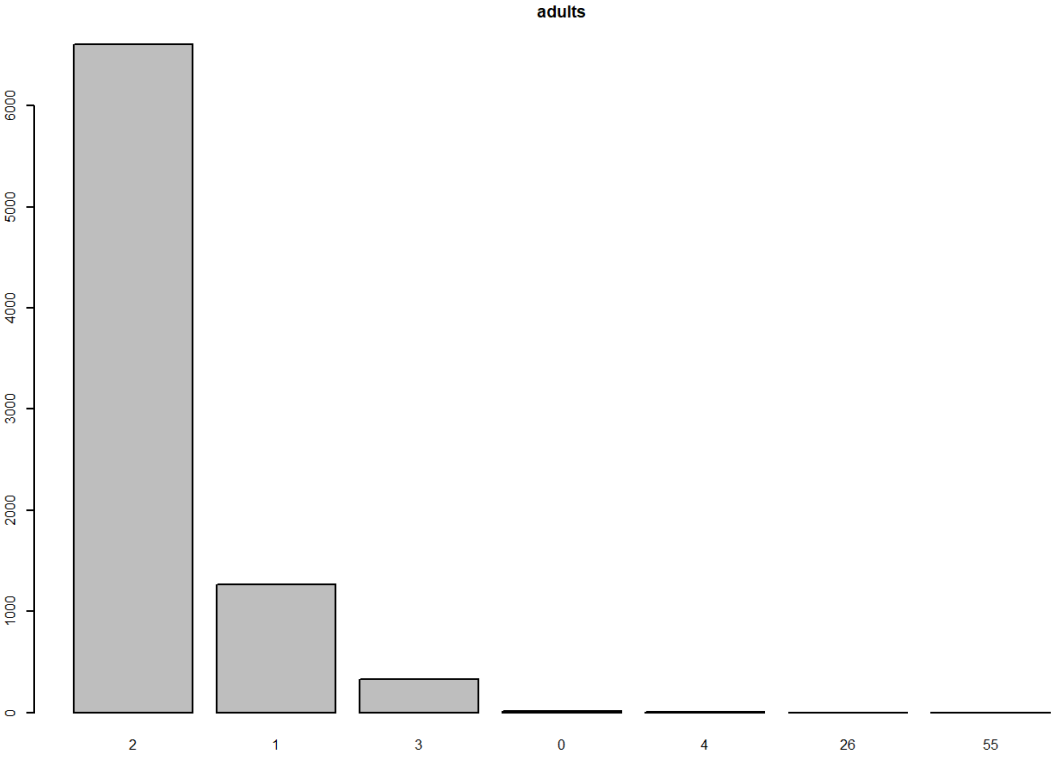
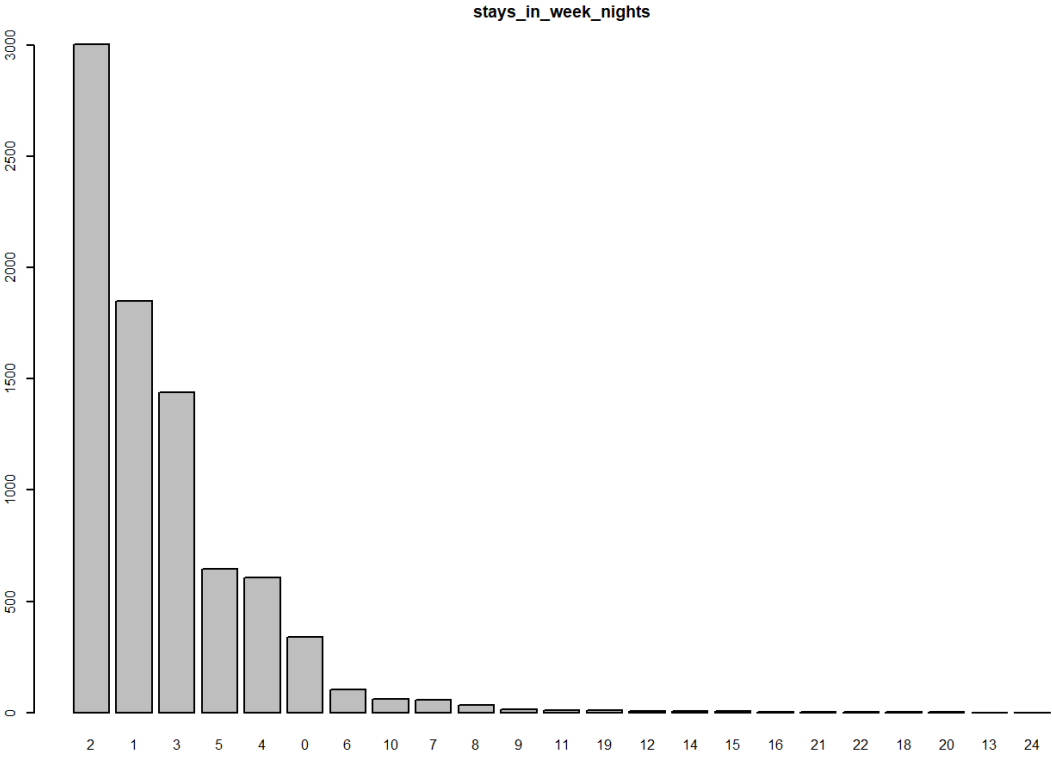




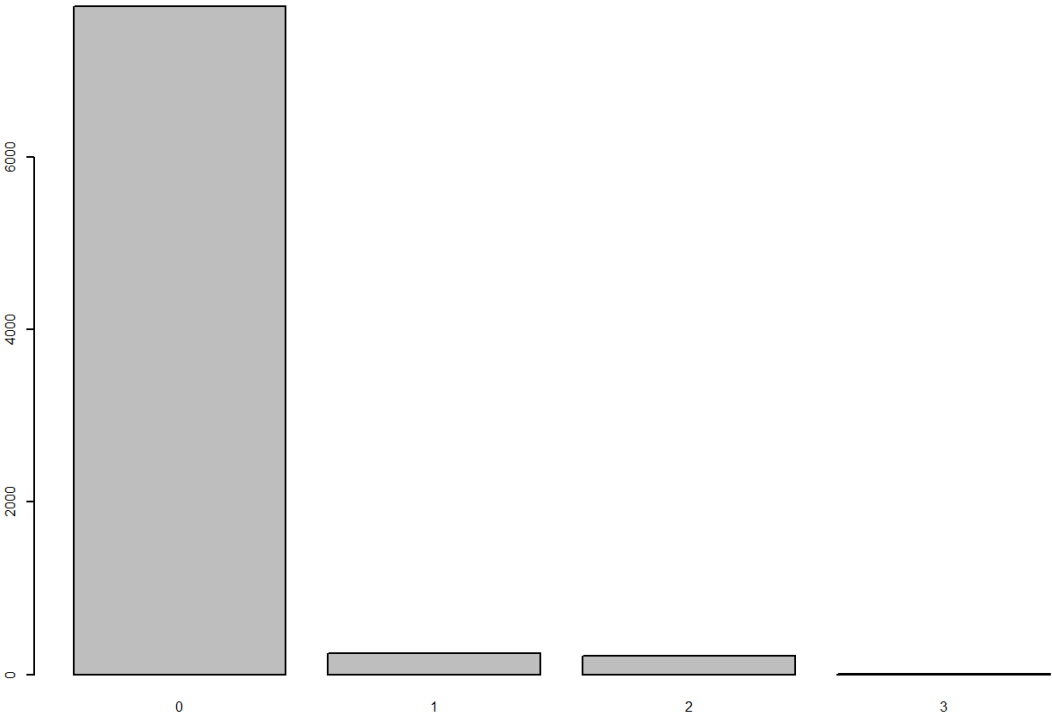




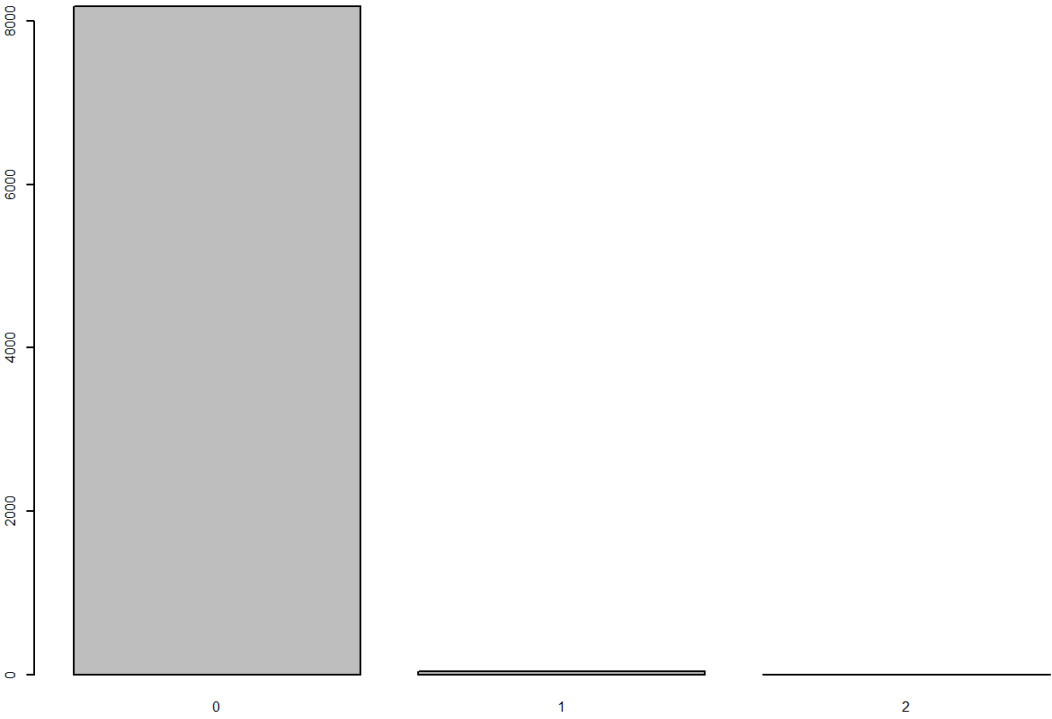




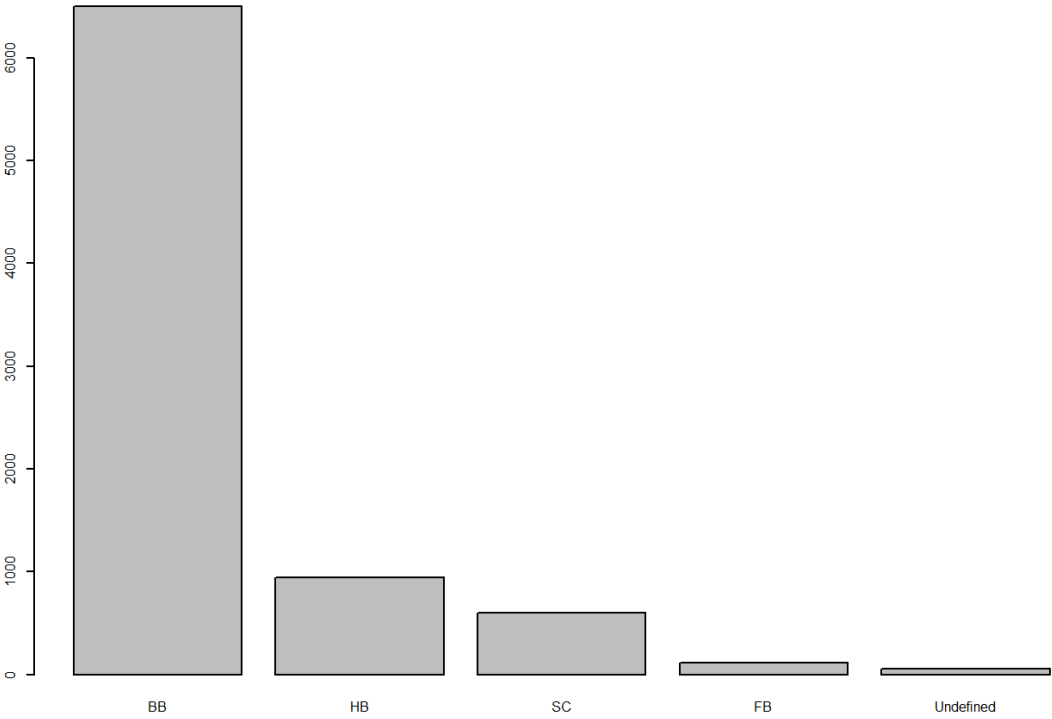
children



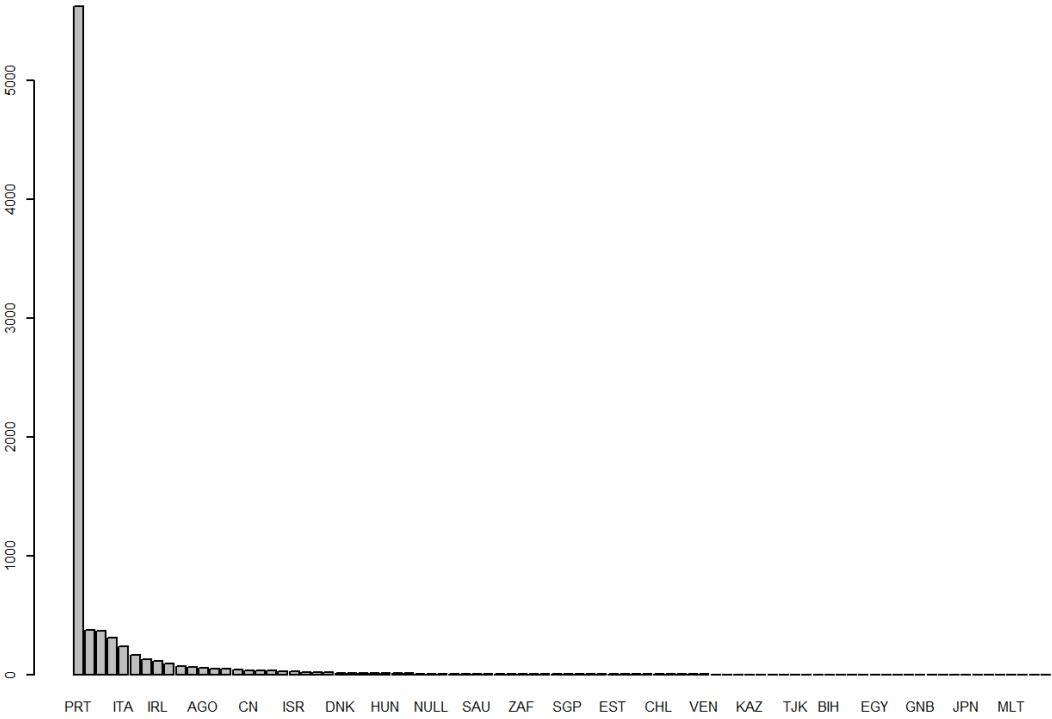
babies

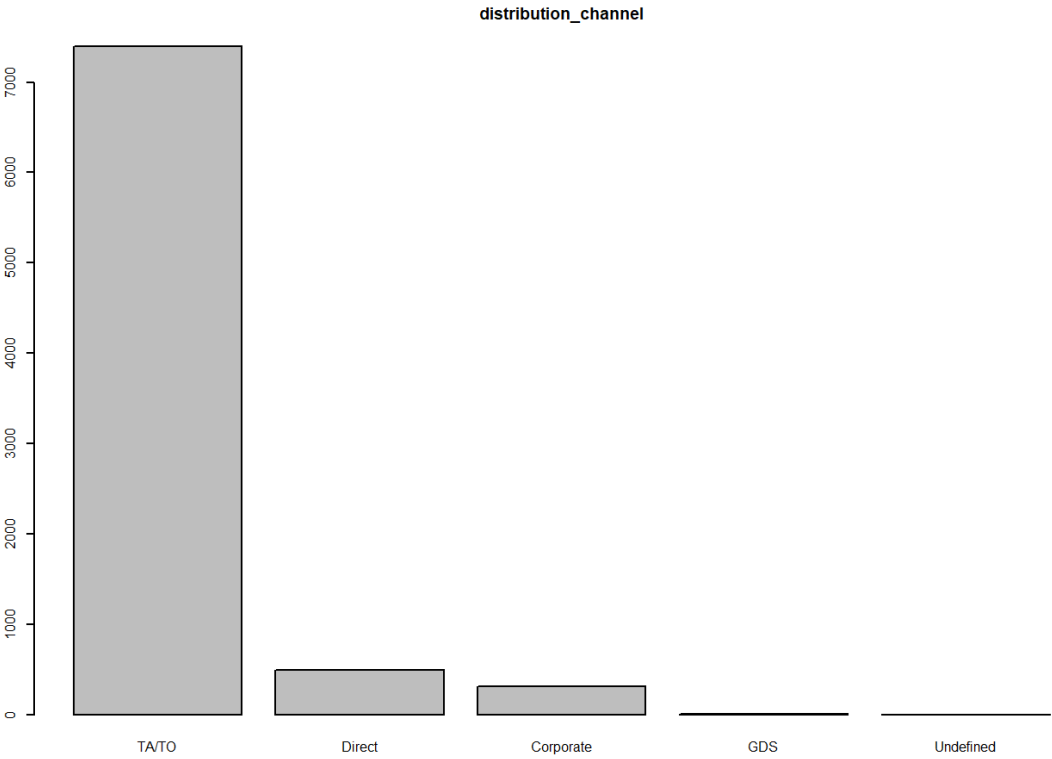
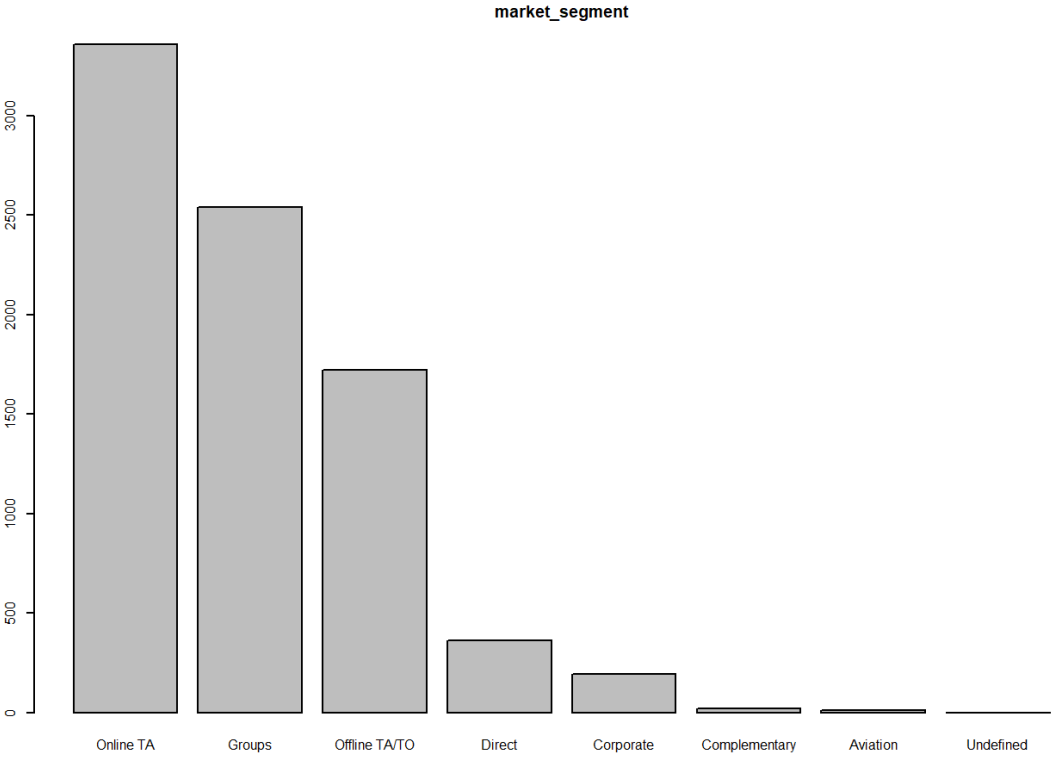


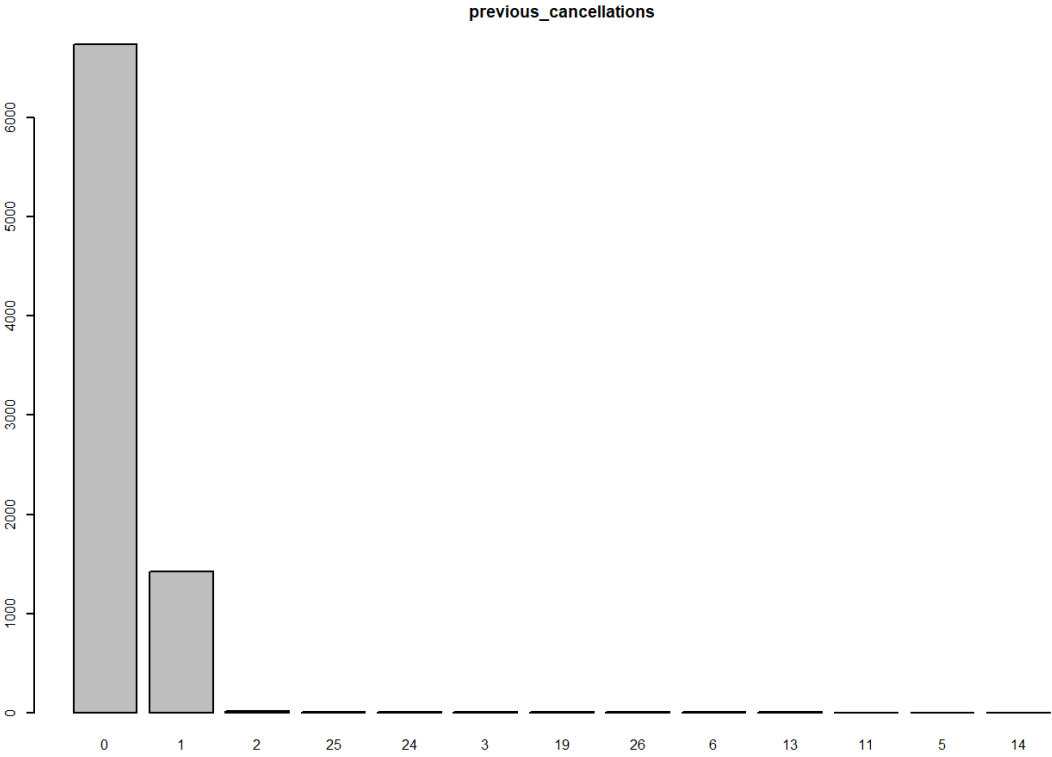
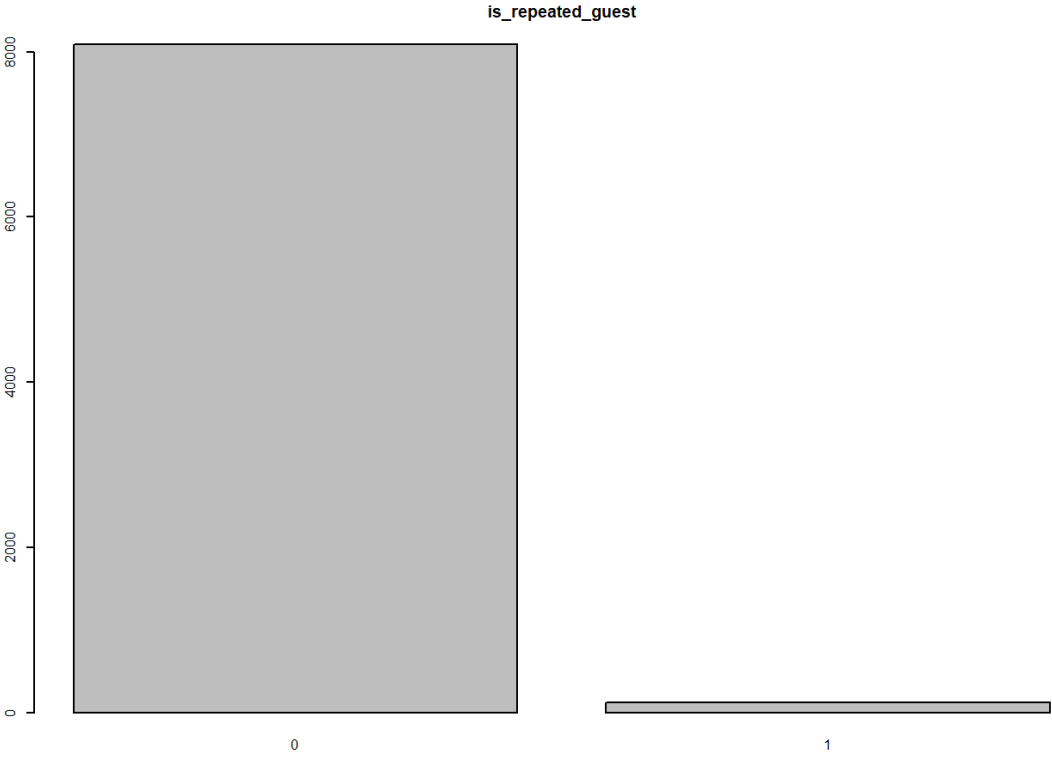
meal

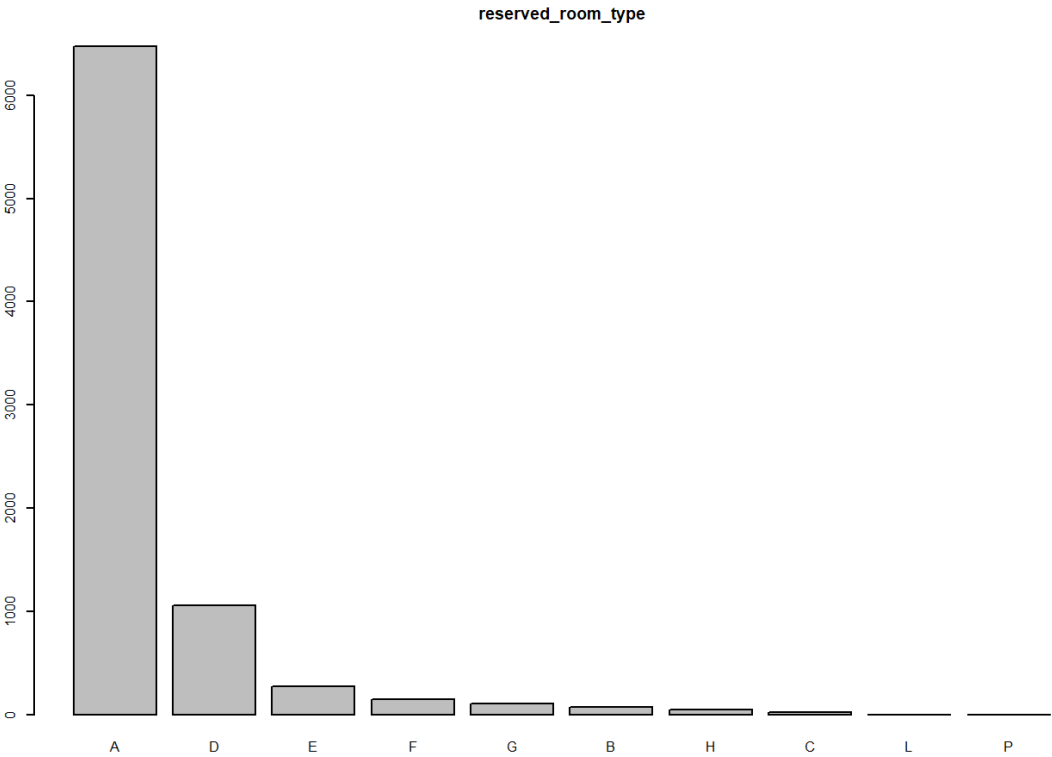
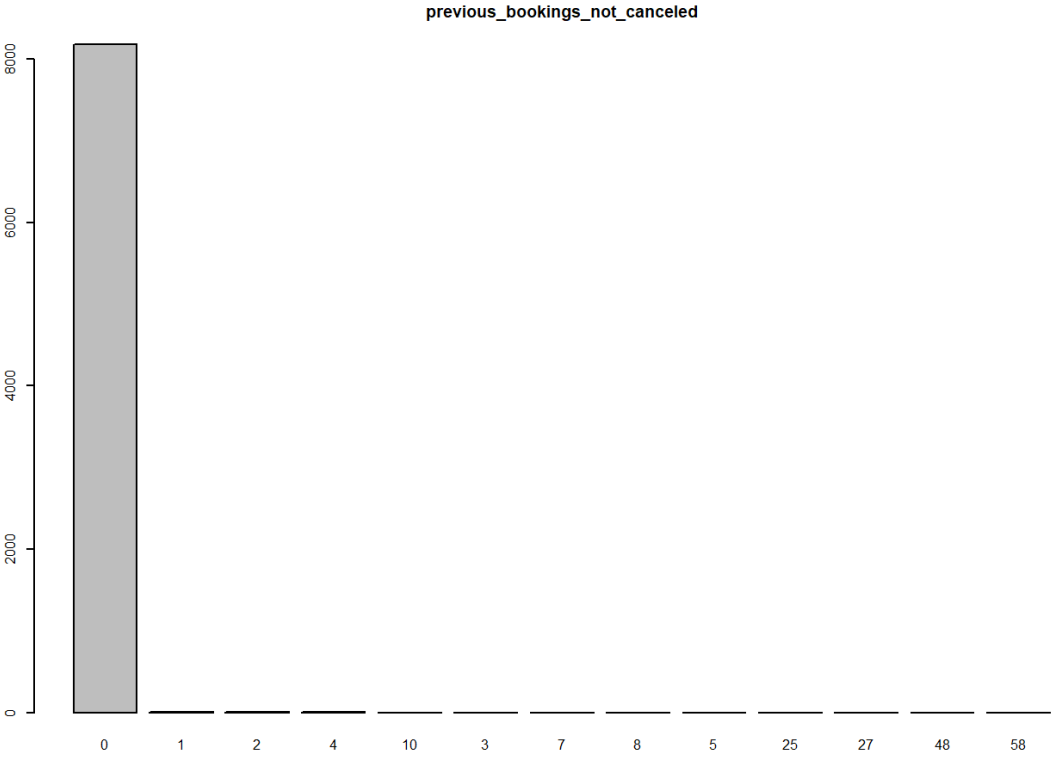


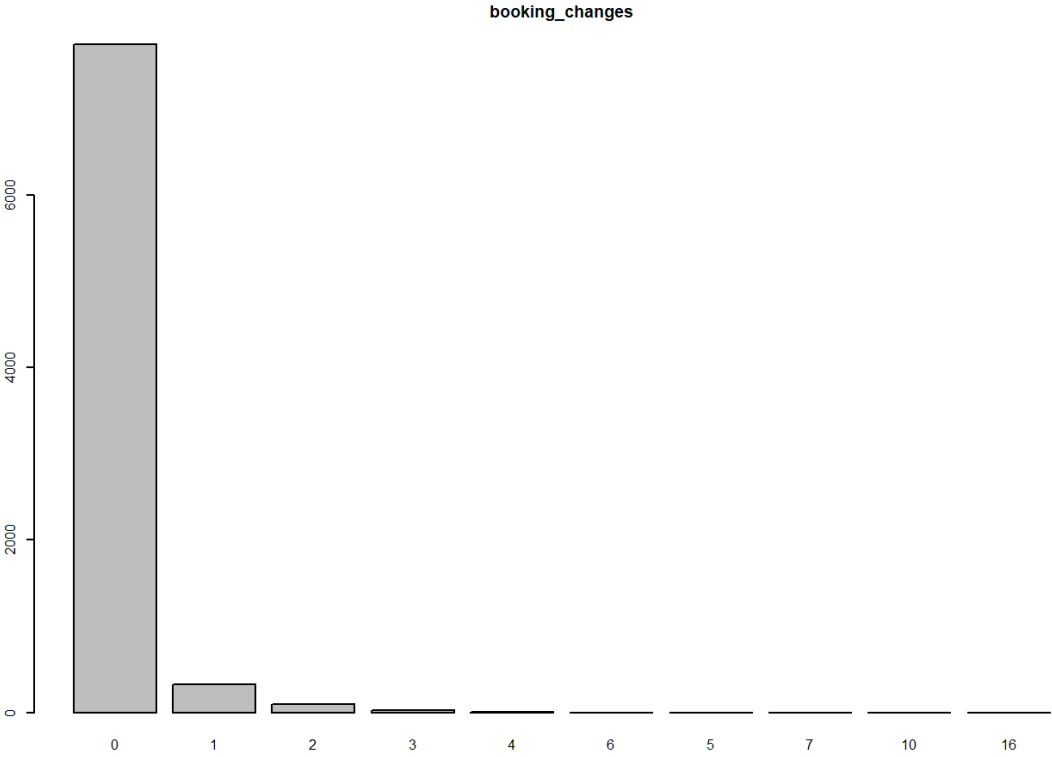
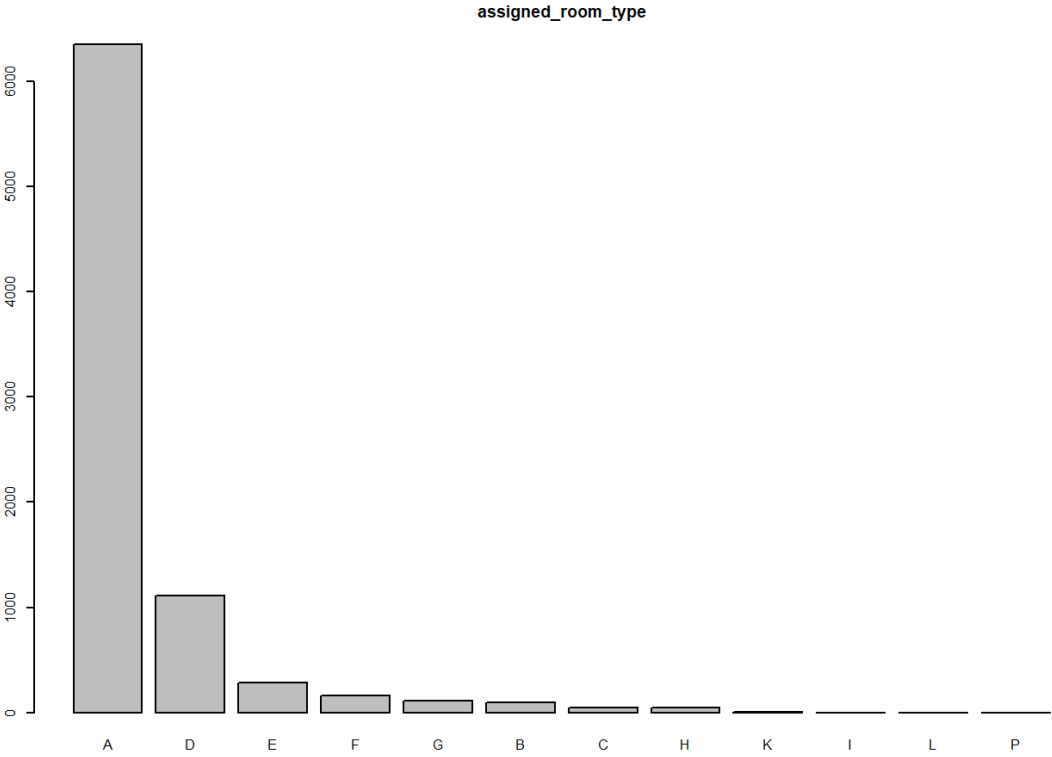
country



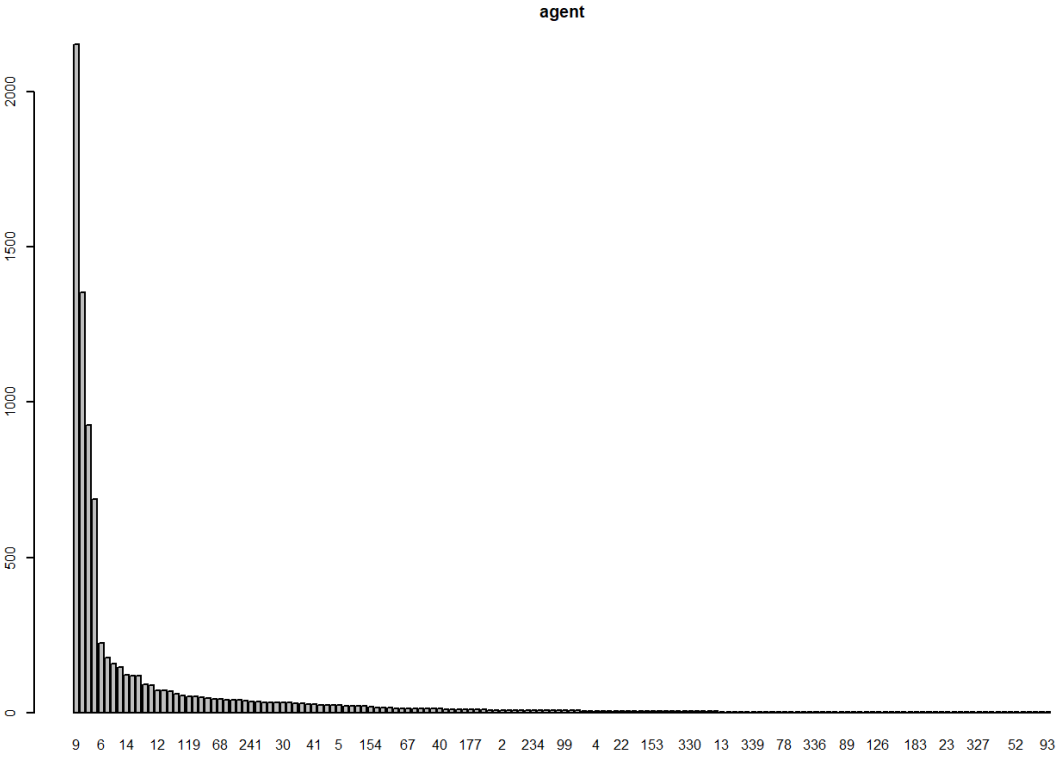
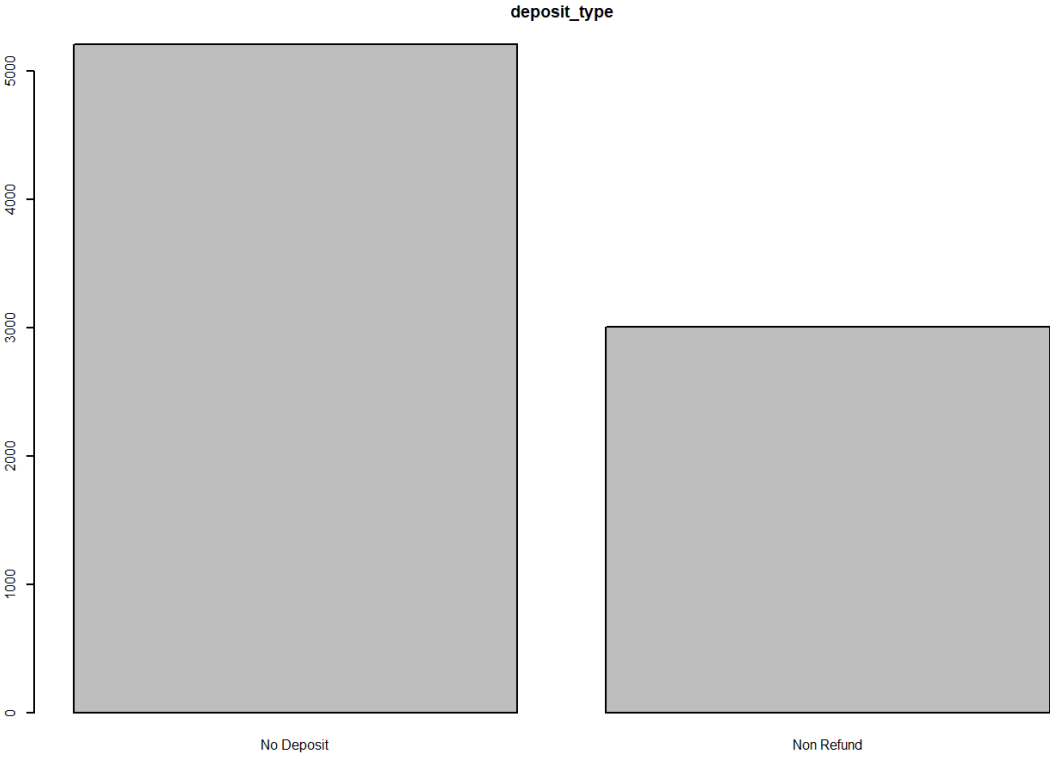


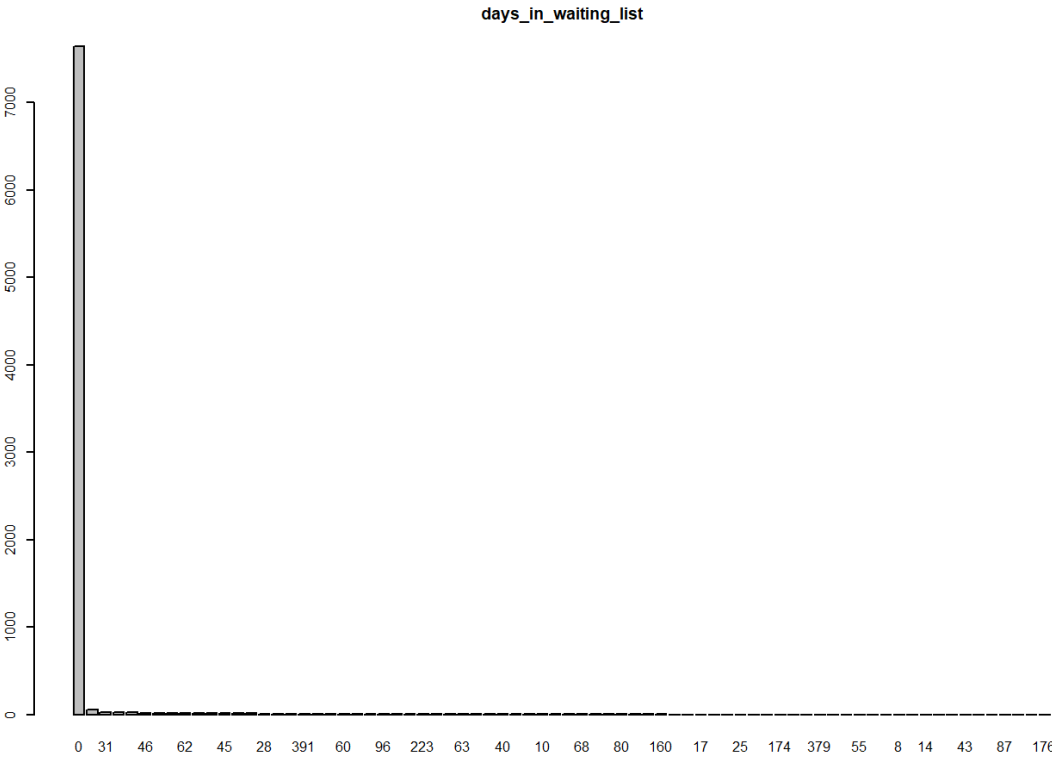
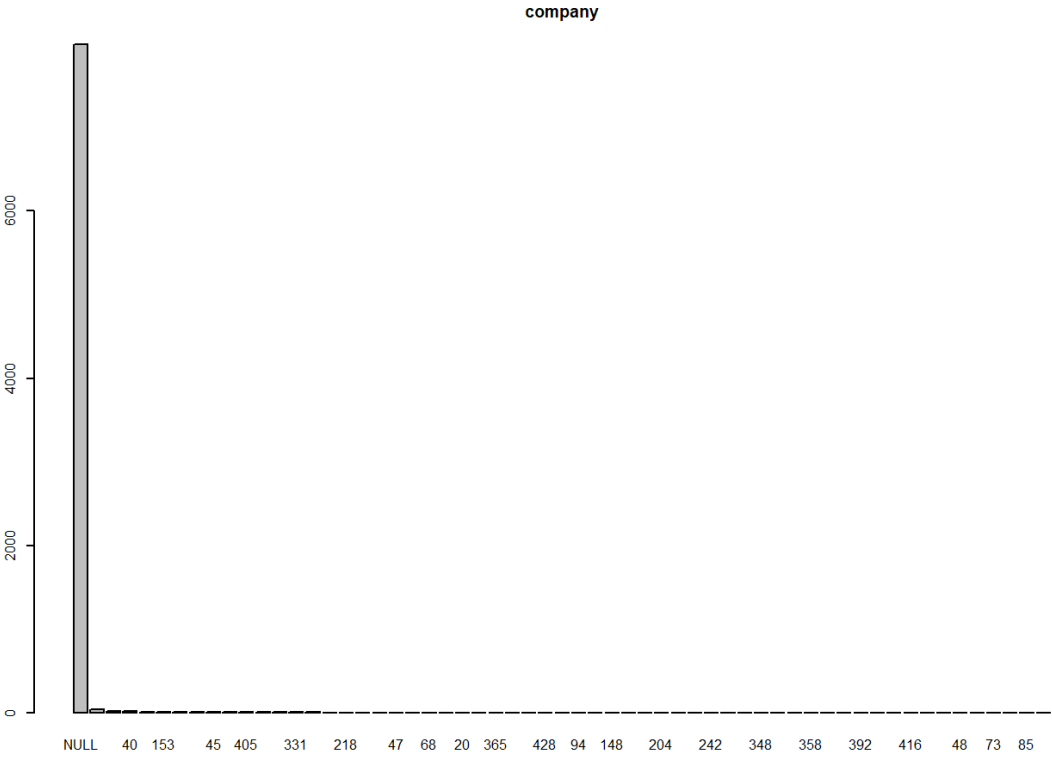


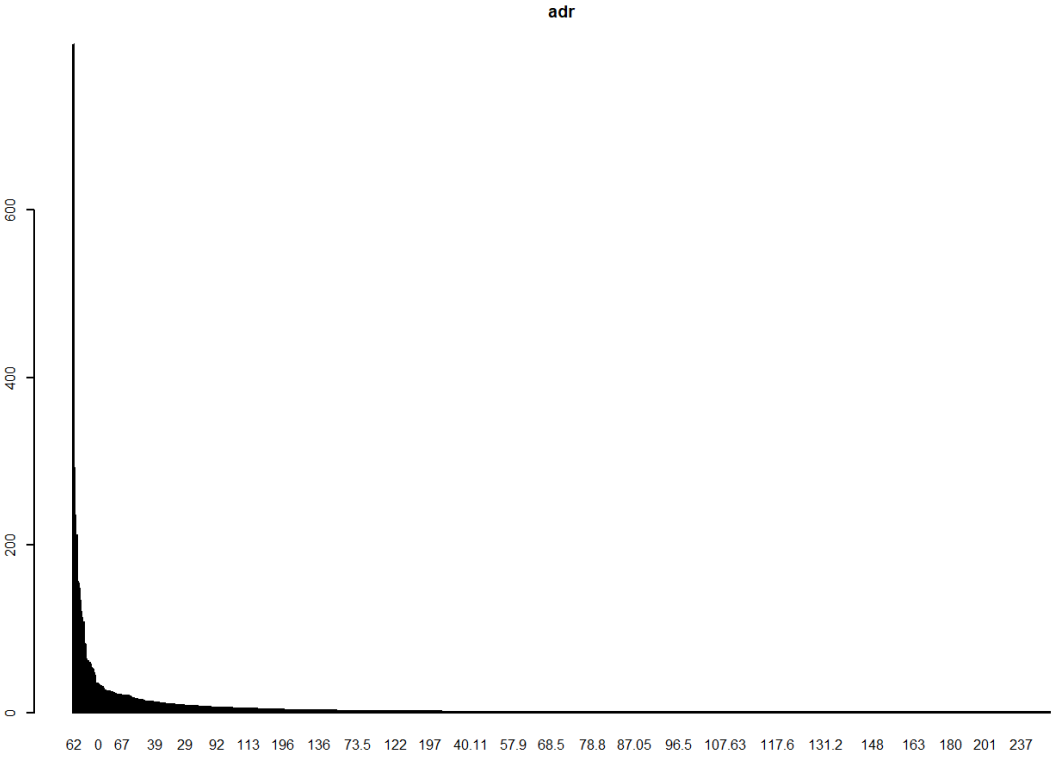
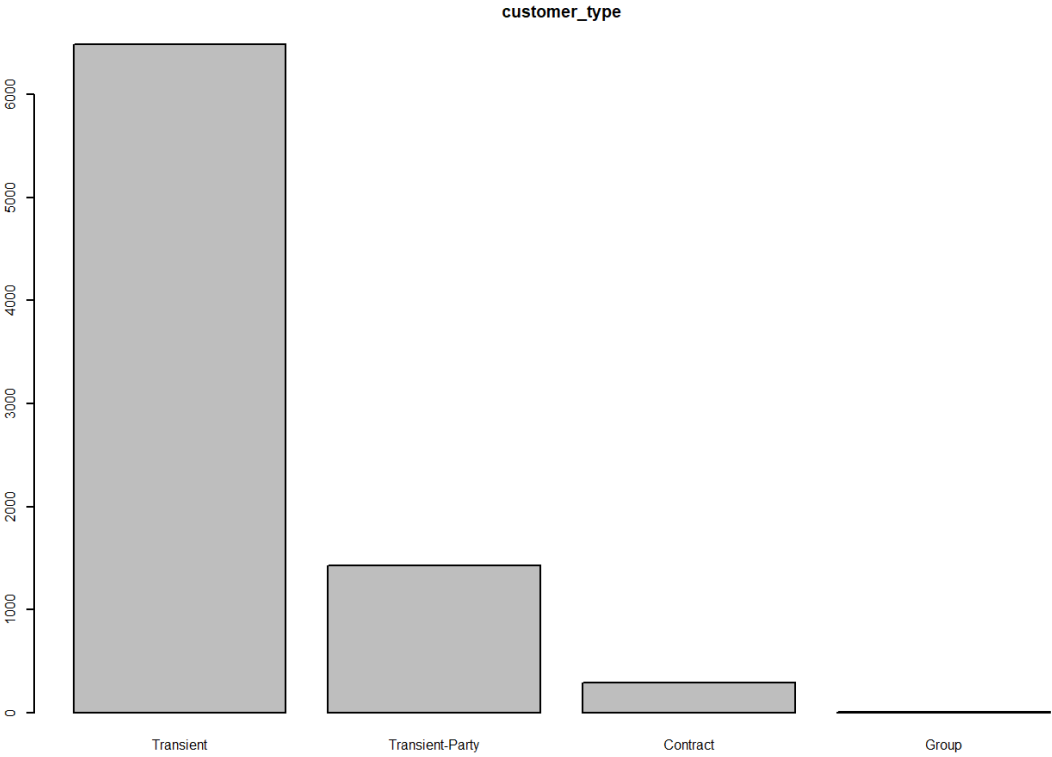


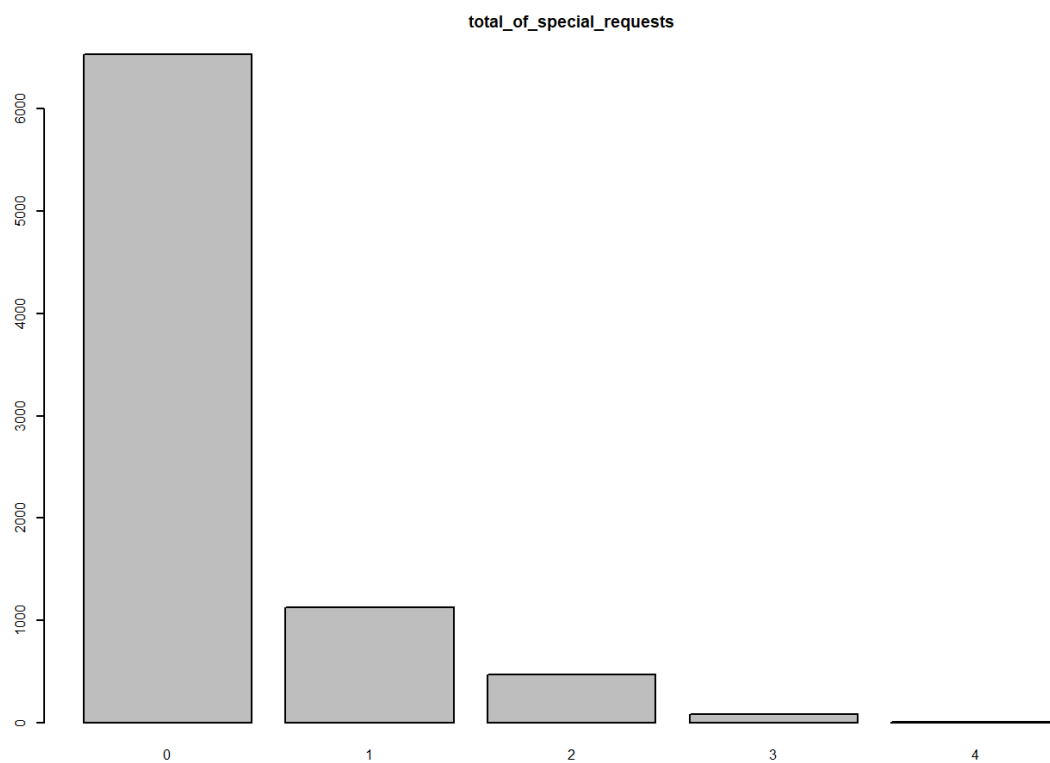
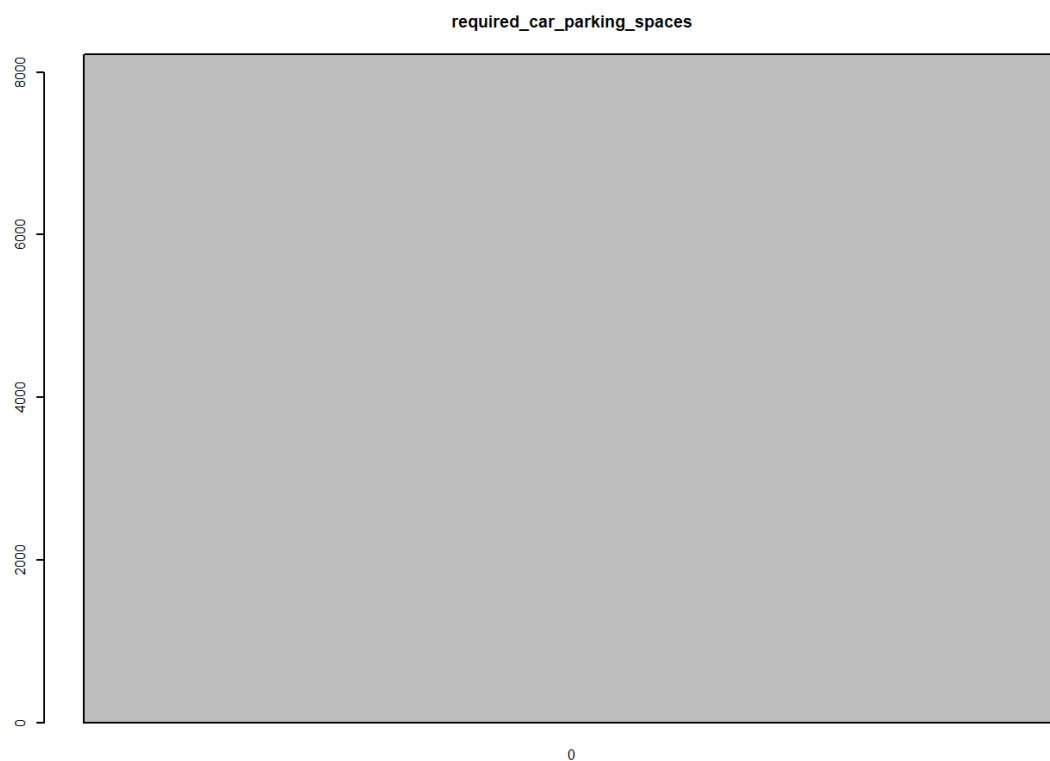












## Resultado de histogramas

Después de revisar cada uno de los histogramas encontramos que:

**Hotel:** hay una proporción considerablemente mayor de cancelaciones en la categoría City

**Lead\_time:** la distribución de sus datos no tiene un comportamiento lógico, porque el mayor número de cancelaciones proviene de 0 días previos de reservación, pero luego se mueve a valores de 90 días, 40 días y luego regresa a 2 días. será importante ver si existe algún patrón en esta variable

**arrival\_date:** observamos que la mayoría de los datos provienen de 2016 y muy pocos de 2017. nos hace pensar que la muestra fue tomada a partir de un mes intermedio de 2015 y cortada en un mes intermedio de 2017

**arrival\_data\_month:** los meses de mayores cancelaciones fueron en octubre, septiembre, marzo, febrero que coinciden con los periodos vacacionales de muchos países y con semana santa o "spring break", esto nos sugiere hacer un análisis de serie de tiempo para determinar si existe algún componente estacional.

**Adults:** la mayor cancelación proviene de 2 y 1 adulto, lo que puede hacernos pensar que pueden ser reservaciones para viajes de trabajo. también es importante considerar que en muchos hoteles las habitaciones son de máximo 3 ocupantes.

**children y babies:** la mayor concentración es de cero y puede reforzar el supuesto de adultos.

**reserver\_room\_type y Assigned\_room\_type:** los valores de las categorías están reservados por privacidad por lo que no se puede interpretar claramente esta variable, sin embargo, ambas están asociadas a la categoría A. podríamos pensar que serían las estándar de 2 personas siguiendo el supuesto que las cancelaciones son mayormente por tipo de viaje de trabajo.

**Country:** esta variable presenta un dato totalmente atípico en la categoría PRT por lo que es importante considerarla ya que podría explicar una porción importante de las cancelaciones.

**market\_segment y distribution\_channel:** estas variables están relacionadas y tiene un sesgo a las categorías asociadas a las agencias de viajes y hechas online

**Deposit\_type:** aquí hay otro caso ilógico, ya que la categoría de no reembolsable está muy por arriba de los reembolsable, uno pensaría que debería ser menos frecuente la cancelación si no te van a devolver tu dinero. por lo que es otra variable importante.

**agent:** podría ser otra variable interesante porque el agente 9 tiene muchas cancelaciones

**customer\_type.** la categoría con más reservaciones es Transient que es cuando la reserva no es parte de un grupo o contrato, y no asociado a otras reservas transitorias

**Adr.** es otra variable que puede ser interesante porque tiene una frecuencia muy alta en el valor 62. se puede hacer el supuesto que es promedio de la mayoría de las habitaciones dobles. siguen el supuesto anterior.

las demás variables no proporcionan información de mayor interés en esta primera revisión.

## Primera revisión de variables interesantes

Primero revisamos si hay alguna relación entre las variables interesantes

Analizando la variable **deposit\_type**

se extrae el subset de **deposit\_type** cancelados

```
sub_rembolso <- subset(sub_cancelados, deposit_type == "Non Refund")
```

revisamos los porcentajes de cada categoría en las otras variables

```
a <- table(sub_rembolso$country)
prop <- prop.table(a)
prop
```

```
##
##          BEL          CHE          CN          ESP          GBR          NULL
## 0.0003323363 0.0003323363 0.0006646726 0.0086407444 0.0112994350 0.0006646726
##          POL          PRT
## 0.0046527085 0.9734130941
```

## Hallazgo 1 observamos que el 97% de las cancelaciones sin reembolso pertenecen al país PRT

```
a <- table(sub_rembolso$agent)
prop <- prop.table(a)
prop
```

```
##
##          1          10          11          119          12          132
## 0.3512794948 0.0009970090 0.0063143902 0.0156198072 0.0199401795 0.0009970090
##          133          134          147          15          154          155
## 0.0029910269 0.0049850449 0.0003323363 0.0029910269 0.0053173812 0.0029910269
##          162          170          171          182          183          19
## 0.0029910269 0.0016616816 0.0069790628 0.0003323363 0.0003323363 0.0451977401
##          191          192          20          208          21          235
## 0.0009970090 0.0016616816 0.0106347624 0.0013293453 0.0285809239 0.0016616816
##          236          240          245          248          252          253
## 0.0186108342 0.0019940179 0.0029910269 0.0033233632 0.0036556996 0.0006646726
##          257          26          273          281          287          29
## 0.0013293453 0.0109670987 0.0076437355 0.0006646726 0.0003323363 0.0355599867
##          298          3          30          305          31          310
## 0.0009970090 0.0475240944 0.0019940179 0.0026586906 0.0109670987 0.0013293453
##          326          332          34          35          37          375
## 0.0132934530 0.0016616816 0.0129611167 0.0029910269 0.0442007311 0.0003323363
##          378          38          39          4          41          42
## 0.0029910269 0.0059820538 0.0026586906 0.0006646726 0.0086407444 0.0006646726
##          44          495          5          50          53          55
## 0.0149551346 0.0046527085 0.0036556996 0.0009970090 0.0013293453 0.0009970090
##          56          58          6          60          64          67
## 0.0066467265 0.0053173812 0.0518444666 0.0019940179 0.0016616816 0.0016616816
##          68          71          78          82          86          96
## 0.0132934530 0.0033233632 0.0009970090 0.0026586906 0.0056497175 0.0083084081
##          98          NULL
## 0.0039880359 0.0987038883
```

observamos que el 34% de las cancelaciones sin reembolso son del agente 1

## Analizando la variable agent

Revisamos el subset de agentes cancelados

```
a <- table(sub_cancelados$agent)
prop <- prop.table(a)
prop
```

```
##
##      1      10      104      105      11      110
## 0.1646987219 0.0012172855 0.0001217285 0.0001217285 0.0025562995 0.0001217285
##      115      118      119      12      126      127
## 0.0004869142 0.0009738284 0.0064516129 0.0086427267 0.0001217285 0.0001217285
##      129      13      132      133      134      138
## 0.0001217285 0.0003651856 0.0008520998 0.0012172855 0.0086427267 0.0009738284
##      14      142      143      146      147      15
## 0.0147291540 0.0015824711 0.0004869142 0.0001217285 0.0012172855 0.0029214851
##      151      152      153      154      155      156
## 0.0003651856 0.0001217285 0.0004869142 0.0024345709 0.0014607425 0.0006086427
##      157      16      162      17      170      171
## 0.0002434571 0.0002434571 0.0010955569 0.0034083993 0.0006086427 0.0047474133
##      175      177      181      182      183      184
## 0.0010955569 0.0012172855 0.0002434571 0.0001217285 0.0001217285 0.0001217285
##      185      187      19      191      192      195
## 0.0003651856 0.0001217285 0.0178940962 0.0038953135 0.0007303713 0.0007303713
##      196      2      20      205      208      21
## 0.0014607425 0.0010955569 0.0074254413 0.0001217285 0.0025562995 0.0143639684
##      215      22      23      234      235      236
## 0.0001217285 0.0006086427 0.0001217285 0.0009738284 0.0006086427 0.0068167985
##      240      241      242      243      245      248
## 0.1125989044 0.0043822276 0.0037735849 0.0004869142 0.0012172855 0.0018259282
##      249      250      251      252      253      254
## 0.0004869142 0.0107121120 0.0004869142 0.0013390140 0.0003651856 0.0001217285
##      257      26      261      262      27      273
## 0.0004869142 0.0052343274 0.0002434571 0.0001217285 0.0010955569 0.0043822276
##      28      281      287      29      296      298
## 0.0019476567 0.0019476567 0.0001217285 0.0143639684 0.0001217285 0.0035301278
##      3      30      305      307      31      310
## 0.0216676811 0.0038953135 0.0009738284 0.0003651856 0.0040170420 0.0004869142
##      314      315      326      327      330      331
## 0.0027997565 0.0030432136 0.0048691418 0.0001217285 0.0004869142 0.0001217285
##      332      336      339      34      35      36
## 0.0009738284 0.0002434571 0.0003651856 0.0049908704 0.0014607425 0.0003651856
##      368      37      375      378      38      387
## 0.0004869142 0.0192331102 0.0002434571 0.0010955569 0.0040170420 0.0001217285
##      39      4      40      41      42      436
## 0.0009738284 0.0007303713 0.0014607425 0.0031649422 0.0003651856 0.0001217285
##      44      440      45      468      481      495
## 0.0057212416 0.0008520998 0.0002434571 0.0001217285 0.0001217285 0.0017041996
##      5      50      52      53      55      56
## 0.0029214851 0.0003651856 0.0001217285 0.0004869142 0.0003651856 0.0063298844
##      57      58      6      60      64      66
## 0.0001217285 0.0030432136 0.0271454656 0.0007303713 0.0006086427 0.0001217285
##      67      68      69      7      71      75
## 0.0015824711 0.0052343274 0.0001217285 0.0109555691 0.0012172855 0.0002434571
##      78      79      8      82      83      85
## 0.0003651856 0.0002434571 0.0085209982 0.0009738284 0.0048691418 0.0027997565
##      86      87      89      9      91      93
## 0.0038953135 0.0004869142 0.0002434571 0.2618381010 0.0001217285 0.0001217285
##      94      95      96      98      99      NULL
## 0.0002434571 0.0004869142 0.0059646987 0.0015824711 0.0009738284 0.0835057821
```

obsevamos que el 16% de las cancelaciones sin reembolso son del agente 1

se extrae el subset del agente 1 cancelados

```
sub_agent <- subset(sub_cancelados, agent == "1")
```

```
a <- table(sub_agent$country)
prop <- prop.table(a)
prop
```

```
##
##          ESP          PRT
## 0.01256467 0.98743533
```

tambien observamos una relacion entre el agente 1 y el pais PRT

Analizando la variable county

se extrae el subset de county cancelados

```
sub_country <- subset(sub_cancelados, country == "PRT")
```

revisamos los porcentajes de cada catagoria en las otras variables

```
a <- table(sub_country$agent)
prop <- prop.table(a)
prop
```



```

##
##      1      10      105      11      110      115
## 0.2376801281 0.0017790429 0.0001779043 0.0037359900 0.0001779043 0.0007116171
##      118      119      12      126      127      129
## 0.0014232343 0.0092510229 0.0120974915 0.0001779043 0.0001779043 0.0001779043
##      13      132      133      134      138      14
## 0.0005337129 0.0012453300 0.0017790429 0.0120974915 0.0014232343 0.0058708415
##      142      143      146      147      15      151
## 0.0014232343 0.0007116171 0.0001779043 0.0017790429 0.0035580857 0.0005337129
##      152      153      154      155      156      157
## 0.0001779043 0.0001779043 0.0035580857 0.0019569472 0.0008895214 0.0003558086
##      16      162      170      171      175      177
## 0.0003558086 0.0016011386 0.0008895214 0.0069382672 0.0016011386 0.0017790429
##      181      182      184      185      187      19
## 0.0003558086 0.0001779043 0.0001779043 0.0005337129 0.0001779043 0.0261519303
##      191      192      195      196      2      20
## 0.0056929372 0.0010674257 0.0001779043 0.0021348514 0.0016011386 0.0108521615
##      205      208      21      215      22      23
## 0.0001779043 0.0037359900 0.0209927059 0.0001779043 0.0008895214 0.0001779043
##      234      235      236      240      241      242
## 0.0014232343 0.0008895214 0.0099626401 0.0917986123 0.0062266501 0.0055150329
##      243      245      248      249      250      251
## 0.0007116171 0.0017790429 0.0024906600 0.0007116171 0.0122753958 0.0007116171
##      252      253      254      257      26      261
## 0.0019569472 0.0005337129 0.0001779043 0.0007116171 0.0076498844 0.0003558086
##      262      27      273      28      281      287
## 0.0001779043 0.0016011386 0.0035580857 0.0026685643 0.0028464686 0.0001779043
##      29      296      298      3      30      305
## 0.0209927059 0.0001779043 0.0051592243 0.0316669632 0.0056929372 0.0014232343
##      307      31      310      314      315      326
## 0.0005337129 0.0058708415 0.0007116171 0.0037359900 0.0044476072 0.0071161715
##      327      330      331      332      336      339
## 0.0001779043 0.0007116171 0.0001779043 0.0014232343 0.0003558086 0.0005337129
##      34      35      36      368      37      375
## 0.0072940758 0.0021348514 0.0005337129 0.0007116171 0.0254403131 0.0003558086
##      378      38      387      39      4      40
## 0.0016011386 0.0058708415 0.0001779043 0.0014232343 0.0010674257 0.0021348514
##      41      42      436      44      440      45
## 0.0046255115 0.0005337129 0.0001779043 0.0083615015 0.0012453300 0.0003558086
##      468      481      495      5      50      52
## 0.0001779043 0.0001779043 0.0024906600 0.0042697029 0.0005337129 0.0001779043
##      53      55      56      57      58      6
## 0.0007116171 0.0005337129 0.0092510229 0.0001779043 0.0044476072 0.0396726561
##      60      64      66      67      68      69
## 0.0010674257 0.0008895214 0.0001779043 0.0023127557 0.0067603629 0.0001779043
##      7      71      75      78      79      8
## 0.0151218644 0.0017790429 0.0003558086 0.0005337129 0.0003558086 0.0119195873
##      82      83      85      86      87      89
## 0.0014232343 0.0071161715 0.0040917986 0.0056929372 0.0007116171 0.0003558086
##      9      91      93      94      95      96
## 0.0441202633 0.0001779043 0.0001779043 0.0003558086 0.0007116171 0.0065824586
##      98      99      NULL
## 0.0023127557 0.0014232343 0.1145703611

```

el 23 por ciento de las reservaciones del país PRT fueron realizadas con el agente 1

## Primeras conclusiones

las variables deposit\_type, county y agent tienen una relacion directa entre el pais PRT, el agente 1 y reservacion sin reembolso

## Analisis de serie de tiempo

```
library(tidyverse)
library(Hmisc)
library(skimr)
library(naniar)
library(patchwork)
library(plotly)
library(ggplot2)
library(DataExplorer)
library(knitr)
library(reshape2)
library(lubridate)
```

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, results='hide')
if(!require("PerformanceAnalytics")) install.packages("PerformanceAnalytics")
if(!require("corrplot")) install.packages("corrplot")
```

## leo datos

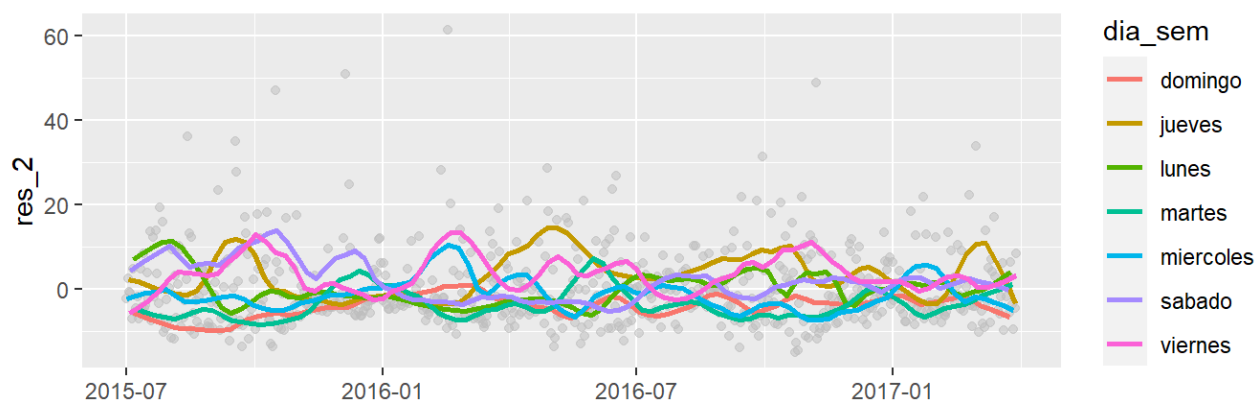
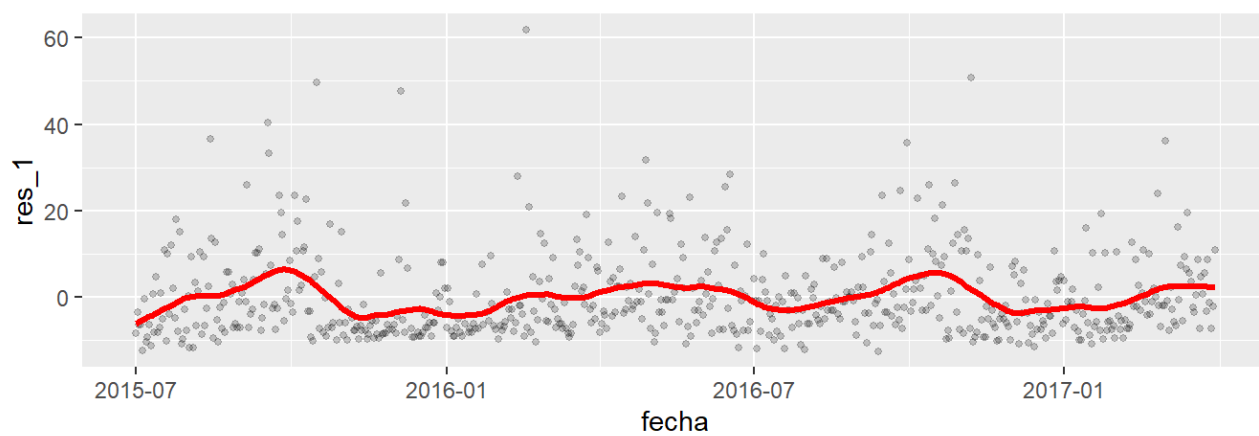
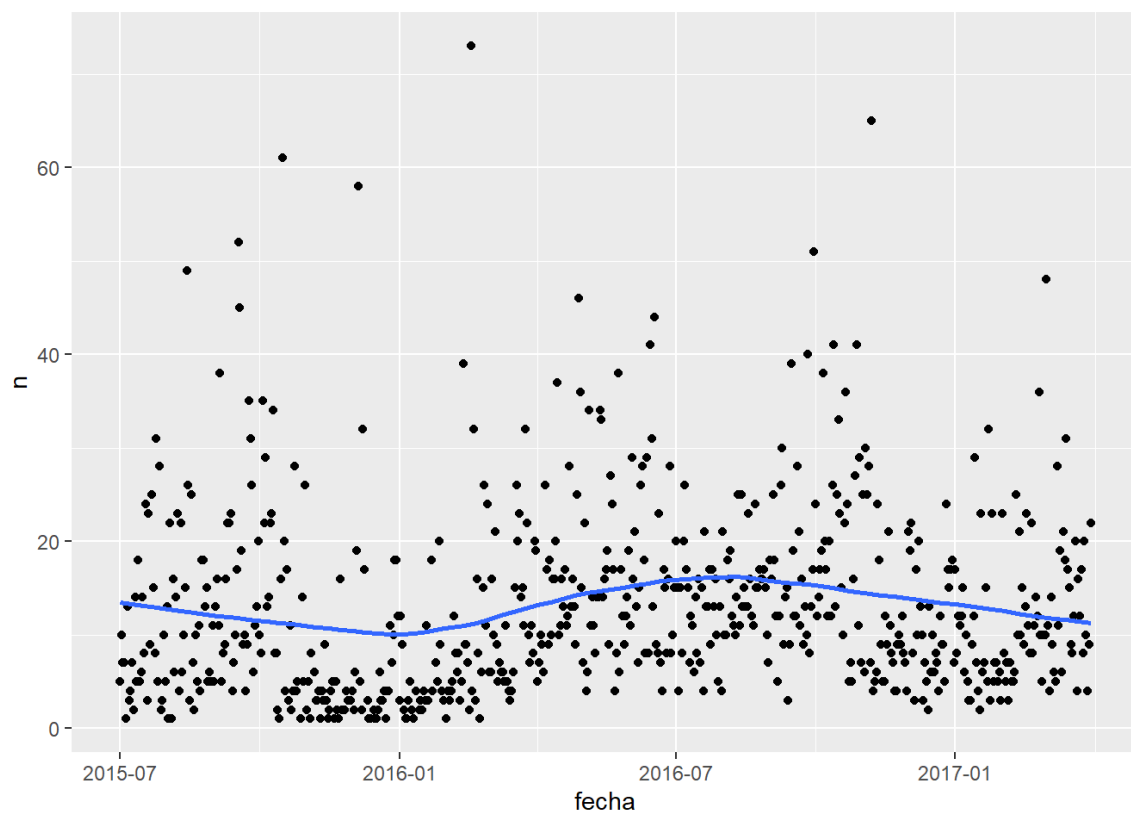
```
data<-read.csv("entrena.csv",stringsAsFactors = F)
```

## dias de la semana

## pascuas

## Creo las 3 bases

```
knitr::opts_chunk$set( message = FALSE, results='hide')
idx <- sample(seq(1, 3), size = nrow(data), replace = TRUE, prob = c(.7, 2, .1))
train <- data[idx == 1,]
test <- data[idx == 2,]
cal <- data[idx == 3,]
```



```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

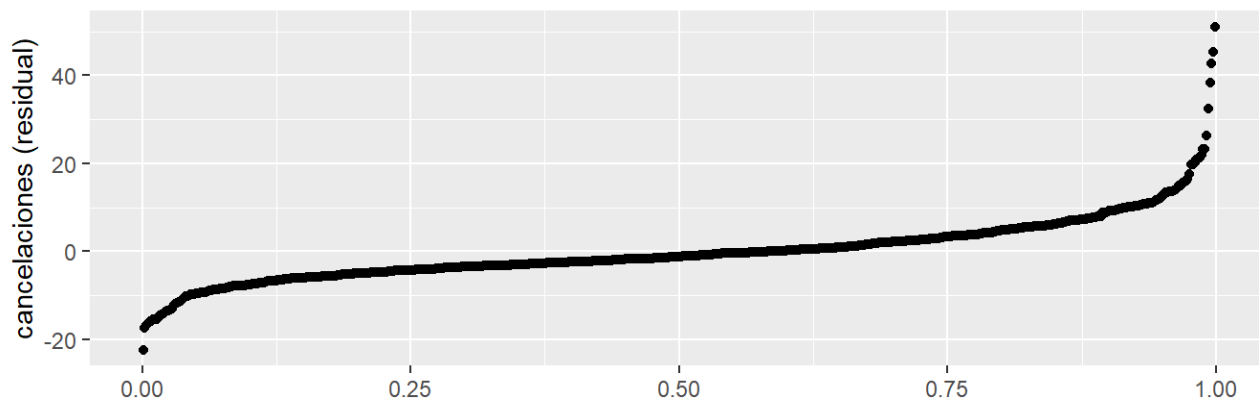
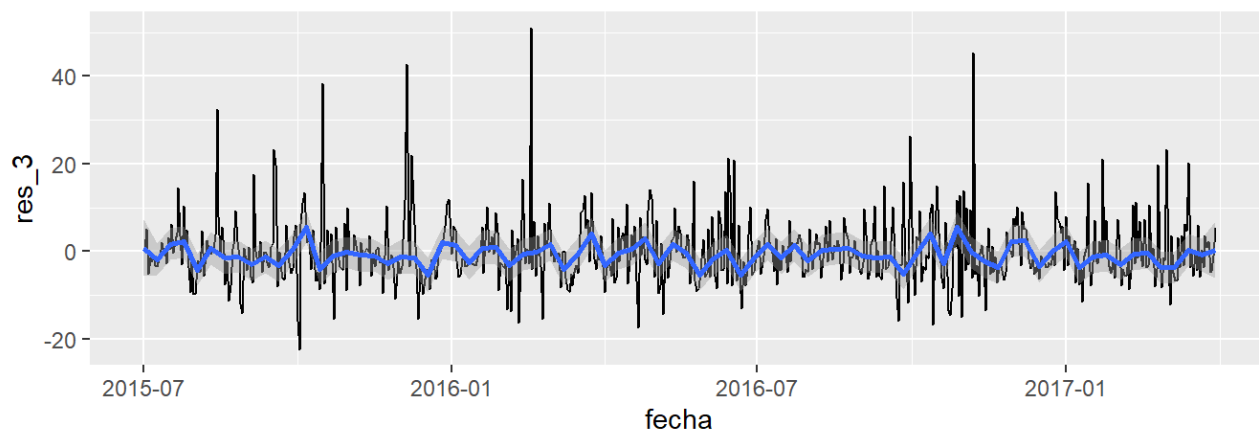
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object))), : k-d tree limited by
## memory. ncmax= 630
```



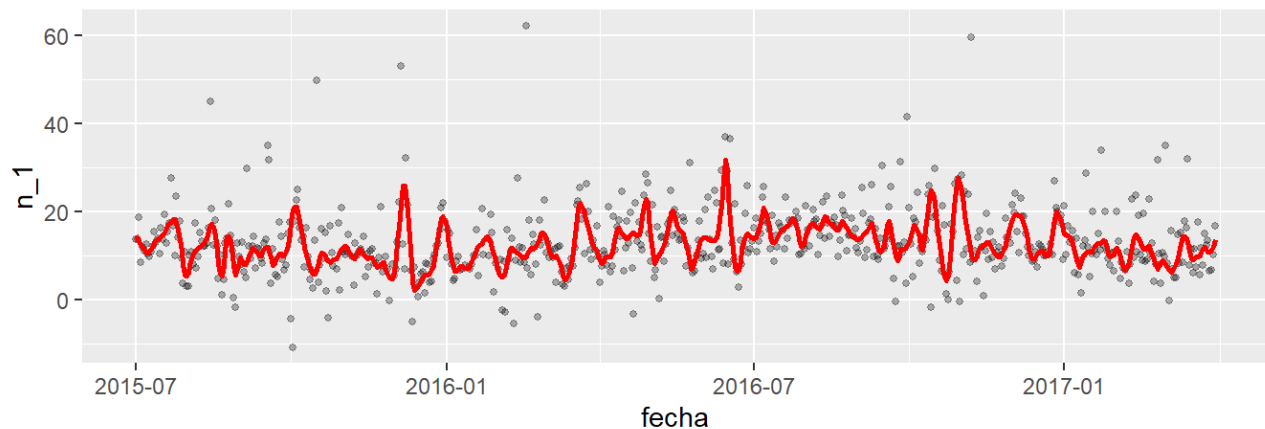
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```



```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 16849
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 3
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0
```

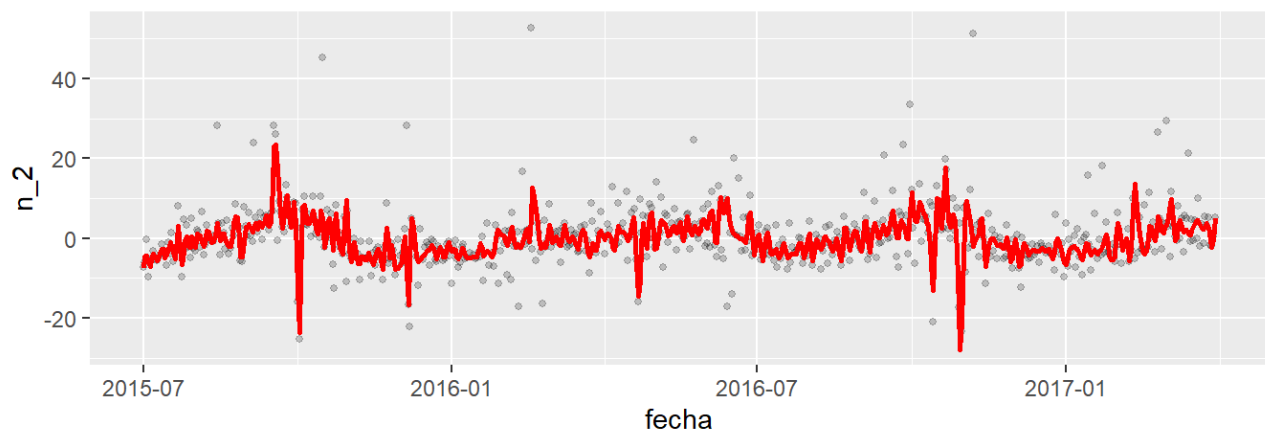
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 9
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

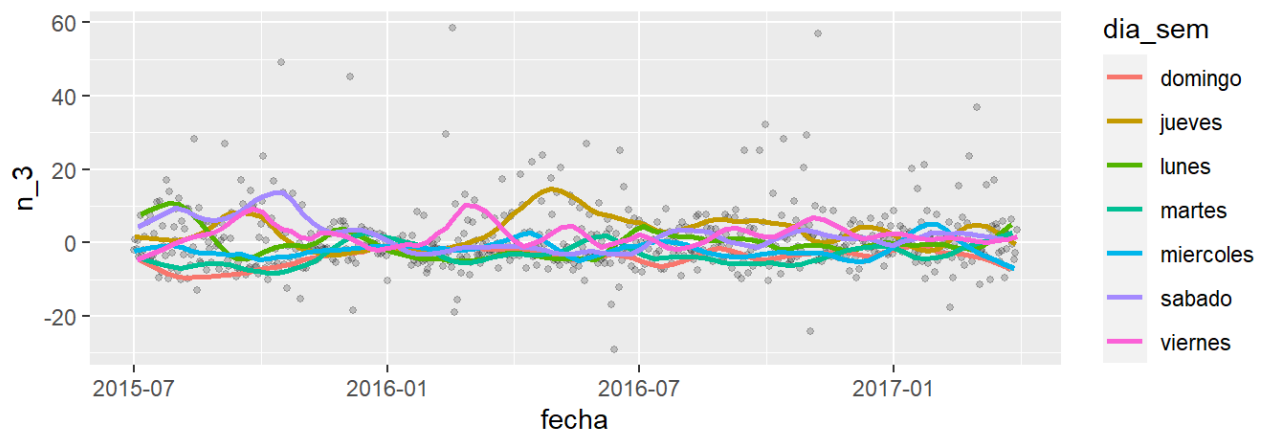
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : k-d tree limited by memory. ncmax= 630
```

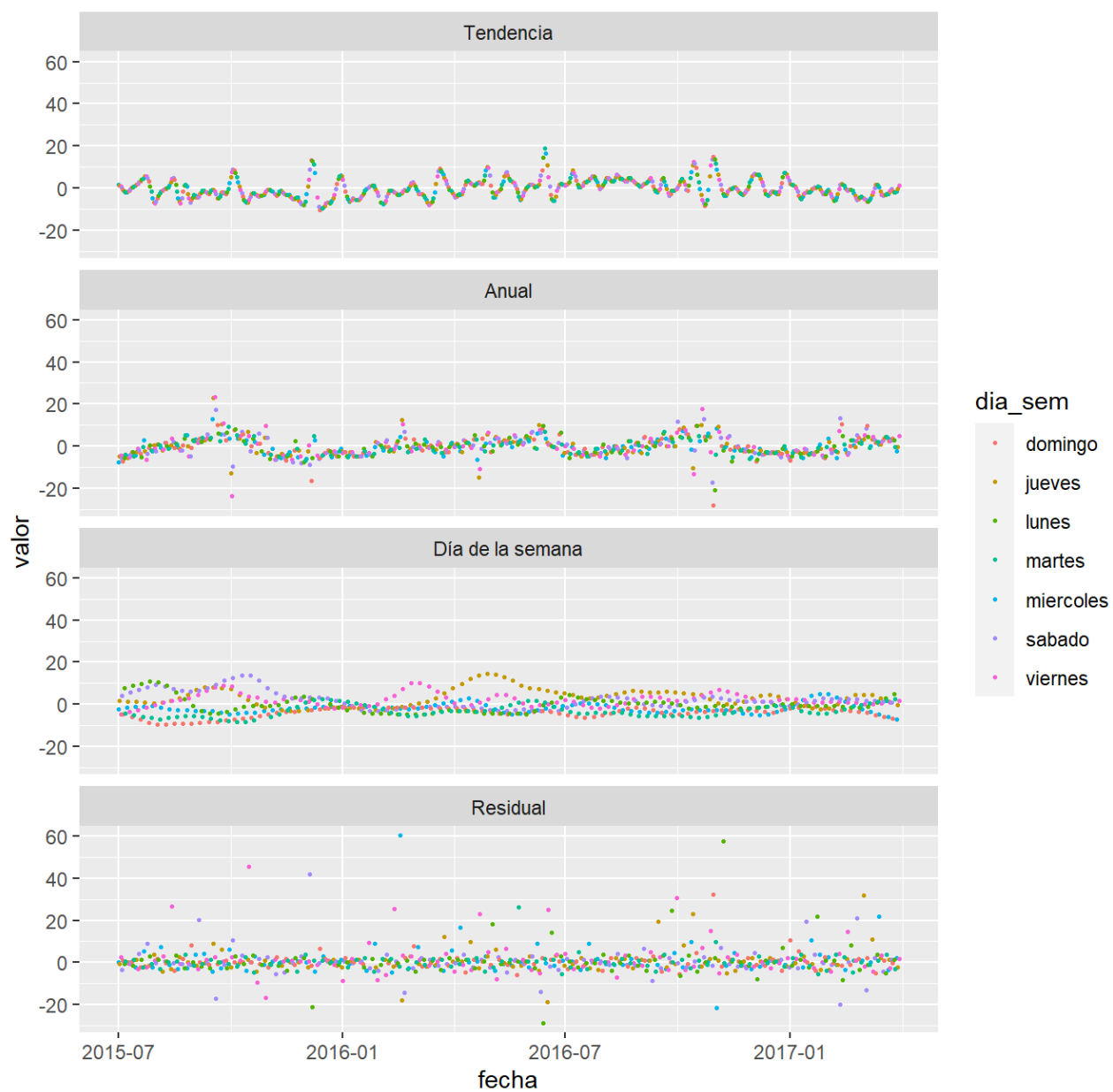


Y ahora repetimos con la componente de día de la semana:



## Análisis de componentes

Ahora comparamos las componentes estimadas y los residuales en una misma gráfica. Por definición, la suma de todas estas componentes da los datos originales.



Media total: 6435