

ANÁLISIS EXPLORATORIO DE DATOS CON



I ❤ TARJETAS “BLACK”
CLICKS.ES

WWW.ILOVECLICKS.ES

Importación de los datos de movimientos

- 1) Importar todos los datos que hay en los Excel de la carpeta datos en un único data frame llamado excel_data (<https://readxl.tidyverse.org/>)
- 2) Comprobar que tras la importación tenemos los movimientos de 85 titulares de tarjetas black
- 3) Calcular cuantos NA hay por variable y que porcentaje representa del total de observaciones.
- 4) Eliminar las filas en las que no hay datos
- 5) Eliminar las filas de totales
- 6) Crear una variable FECHA con formato HH-mm-dd HH:MM:SS y elimina la variable HORA_HOST
- 7) Crear las variables APELLIDOS, NOMBRE y NOMBRE_Y_APELLIDOS del titular
- 8) Poner en mayúsculas el nombre de todas las variables
- 9) Comprobar que los nombres de las variables del dataset son:

NIF TARJETA NOMBRE_Y_APELLIDOS APELLIDOS NOMBRE FECHA COD_OPERACION COD_OPE_DESC IMPORTE NOMBRE_COMERCIO DES_L_SEC_ACTIVI
EMPRESA

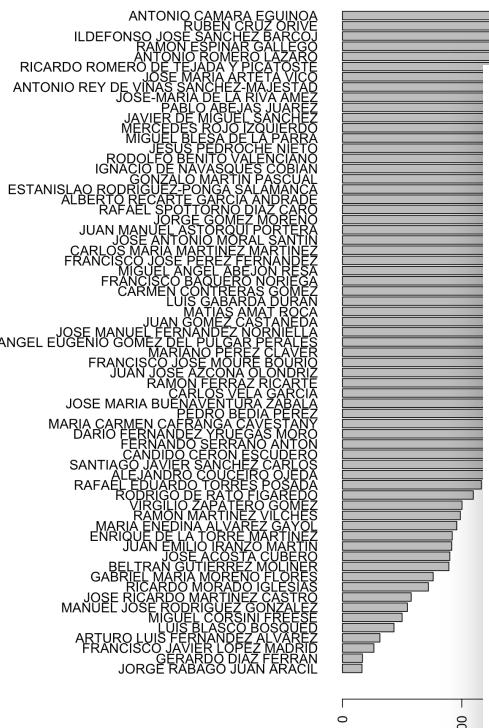
Importación de los nombres de los acusados

- 1) Importar el texto de la sentencia en un objeto "text" (pdftools::pdf_text())
- 2) Asignar al objeto "acusados" un data frame con el nombre de los 65 acusados que están en las páginas 1 a 11 de la sentencia. (stringr::str_match_all)
- 3) Limpieza de los nombres: Eliminar las tildes de los nombres (stringi::stri_trans_general())
- 4) Vamos a eliminar de excel_data los datos de aquellas transacciones cuyos titulares no están en el data set "acusados" proveniente de la sentencia.
 - 1) Realizar comprobaciones para ver las personas que tienen en común ambos datasets:
 - 2) Los que están en la sentencia y de los que no tenemos en las transacciones
 - 3) Los que están en las transacciones y de los que no tenemos en la sentencia
 - 4) Corregir en "acusados" los nombres de las personas para hacerlos iguales a los de excel_data (case_when)
 - 5) excel_data debe ser finalmente un dataset con todas las transacciones de los acusados. Comprobad que los nombres de los acusados en "acusados" es idéntico a la nombres de los titulares de tarjetas en "excel_data"
- 5) Incorporar la información del archivo miembros.csv al dataset "excel_data" para obtener la función que desempeñaba cada acusado y la organización a la que pertenecía
- 6) Comprobar que el número de apuntes y el dinero gastado de, por ejemplo, MIGUEL BLESA DE LA PARRA, coincide con el disponible en <https://elpais.com/especiales/2014/tarjetas-opacas-caja-madrid/>

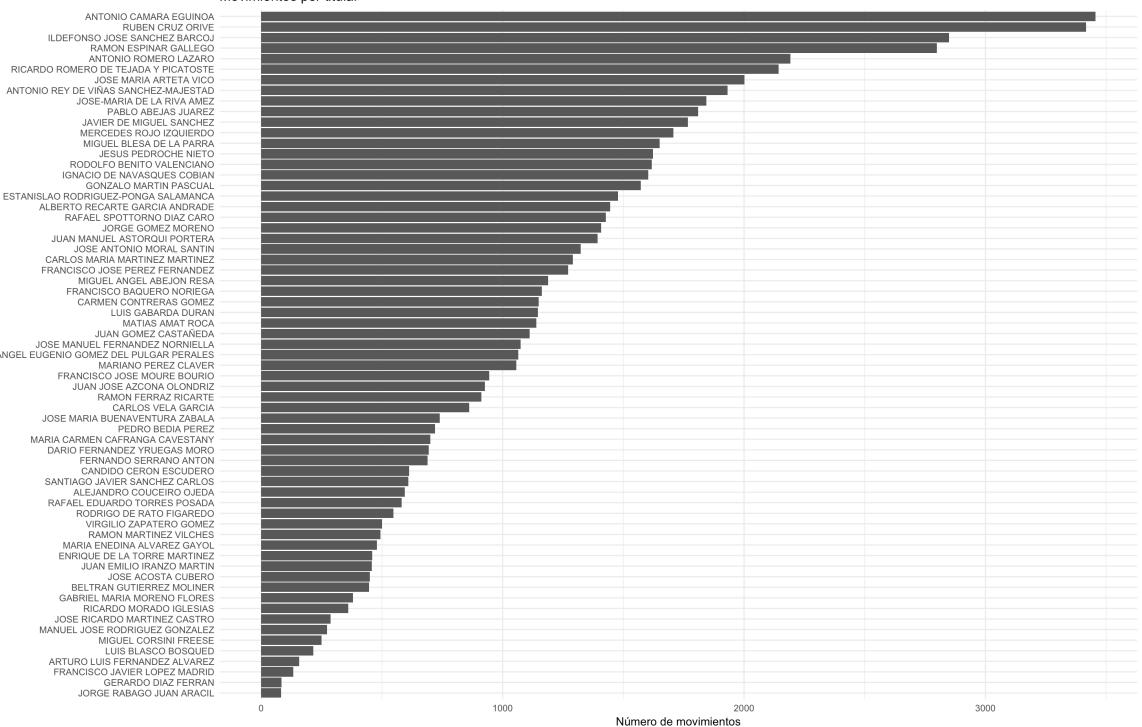
Ej. #1

Número de observaciones por titular de la tarjeta

Movimientos por titular



Movimientos por titular



Ej. #2

Cada anotación tiene un código de operación ¿Qué significa cada uno de ellos?

200 ANUL.COMPRA,MARCAR/DESMARCAR FOTO(NMT)
201 ANULACION DE COMPRA
310 ANULACION REINTEGRO CAJ.PROPIO
320 ANULACION REINTEGRO CAJ.AJENO

Calcula la media y la mediana de todos los importes exceptos de los que son abonos.

	importe_media	importe_mediana
	<dbl>	<dbl>
1	156.	68.5

Ej. #3

Quiénes son los titulares "extremos" que destacan por haber comprado y a la vez retirado más dinero en efectivo

1. Primero deberás agrupar las operaciones de reintegro en cajeros y disposiciones en efectivo dentro de una misma categoría

COD_OPE_DESC_AGRUPADO

<chr>

RETIRADA EFECTIVO

RECARGA TELEFONO GSM - CARGA EN FIRME

IMPOSICION EN TARJETA DE CREDITO DESDE CUENTA PERSONAL

IMPOSICION EN TARJETA DE CREDITO

DEVOLUCION CUOTA POR BMP

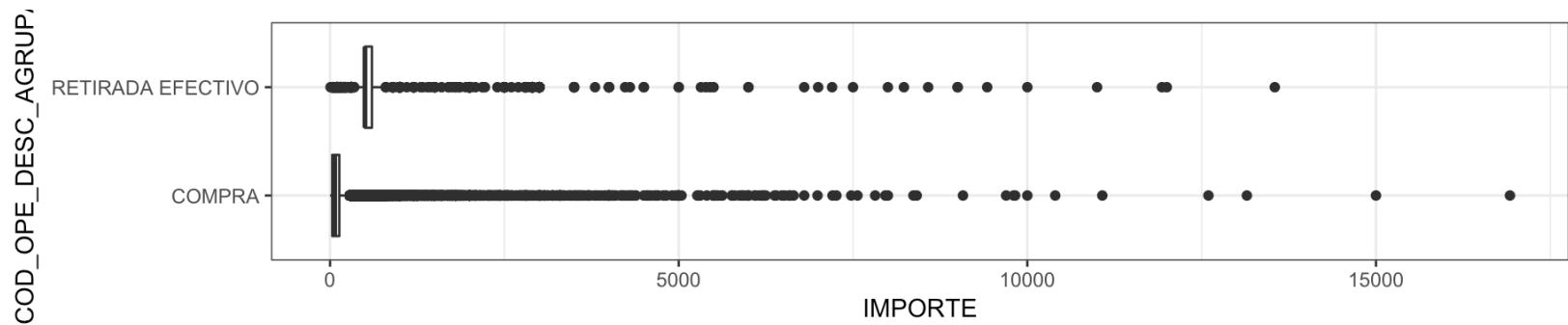
DEVOLUCION COMPRA (ABONO Y CARGO)

DEVOLUCION COMPRA

Ej. #3 (cont)

Quiénes son los titulares "extremos" que destacan por haber comprado y a la vez retirado más dinero en efectivo

2. Hacer un boxplot de los importes de los movimientos que nos permita comparar las disposiciones en efectivo Vs las compras



3. Repetir lo mismo pero eliminando los valores extremos siguiendo el criterio de Tukey:
https://en.wikipedia.org/wiki/Outlier#Tukey's_fences

Ej. #3 (cont)

Quiénes son los titulares "extremos" que destacan por haber comprado y a la vez retirado más dinero en efectivo

4. Presentar los titulares "extremos" en forma de tabla pivotada comparando el total gastado en compras con el total retirado en efectivo.

NOMBRE_Y_APELLIDOS <i><chr></i>	COMPRA `RETIRADA EFECTIVO` <i><dbl></i>	<i><dbl></i>
ILDEFONSO JOSE SANCHEZ BARCOJ	<u>165388.</u>	<u>53680.</u>
JUAN MANUEL ASTORQUI PORTERA	<u>109650.</u>	<u>4000.</u>
CARMEN CONTRERAS GOMEZ	<u>107341.</u>	<u>4231.</u>

Ej. #4

¿En qué han gastado más el dinero?

1. Dibuja un diagrama de barras que muestre las 10 actividades (DES_L_SEC_ACTIVI) en las que más se ha gastado con el importe total del gasto.

2. Obtén el gasto medio por actividad de todos los titulares

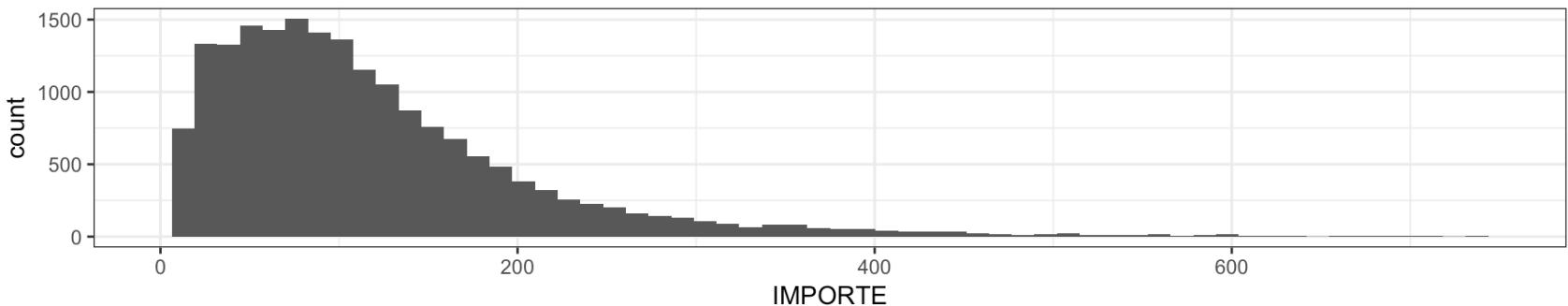
DES_L_SEC_ACTIVI <i><chr></i>	IMPORTE_MEDIO <i><dbl></i>
FUNERARIAS	<u>4293.</u>
STEIGENBERGER HOTELS	<u>3815.</u>
JOYERIA Y RELOJERIA NIVEL 4	<u>2730.</u>
VIAJES MARSANS-INTERNACIONAL EXPRESSO	<u>2538.</u>
FINNAIR	<u>2386.</u>
GRUPO RACC	<u>2227.</u>

Ej. #5

¿Cómo eran los gastos en restauración?

1. Dibuja un histograma de frecuencias para ver como se distribuía el gasto para toda las categorías de restaurantes.

```
grep([EXPRESION REGULAR], DES_L_SEC_ACTIVI))
```



2. ¿Existen diferencias significativas según el tipo de restaurante? Segmenta el histograma anterior por DES_L_SEC_ACTIVI

Ej. #6

Hoteles Vs Restauración

1. Calcula el total gastado por los titulares en hoteles y en restaurantes así como el número de operaciones anotadas para cada uno de ellos

ACTIVIDAD <i><chr></i>	TOTAL <i><dbl></i>	OPERACIONES <i><int></i>
Hoteles	<u>732836.</u>	<u>3084</u>
Restaurantes	<u>2424092.</u>	<u>19097</u>

Ej. #7

Hoteles Vs Restauración

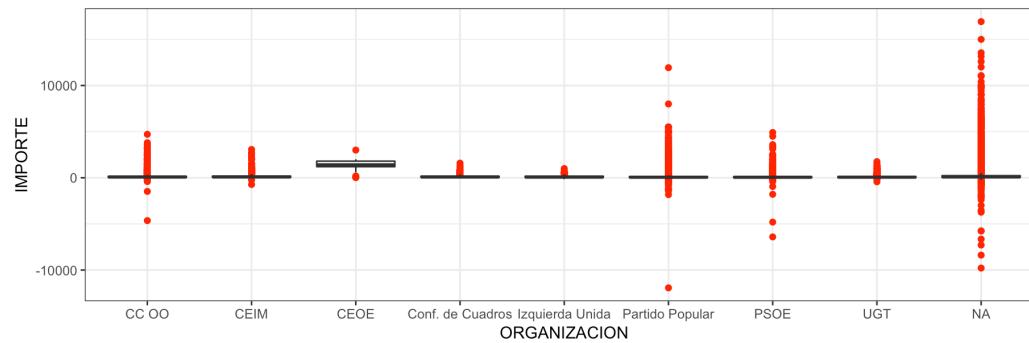
1. Calcula el total gastado por los titulares en hoteles y en restaurantes así como el número de operaciones anotadas para cada uno de ellos

ACTIVIDAD <i><chr></i>	TOTAL <i><dbl></i>	OPERACIONES <i><int></i>
Hoteles	<u>732836.</u>	<u>3084</u>
Restaurantes	<u>2424092.</u>	<u>19097</u>

Ej. #8

Sobre los puestos en Bankia y las organizaciones de pertenencia de los acusados

1. ¿Cuántos acusados hay por puesto en Bankia? ¿Y por organización de pertenencia?
2. ¿Cuál es el gasto total, el gasto medio y número de operaciones por de cada puesto?
3. Realiza un boxplot para comparar los importes de las operaciones por organización de pertenencia de cada acusado.

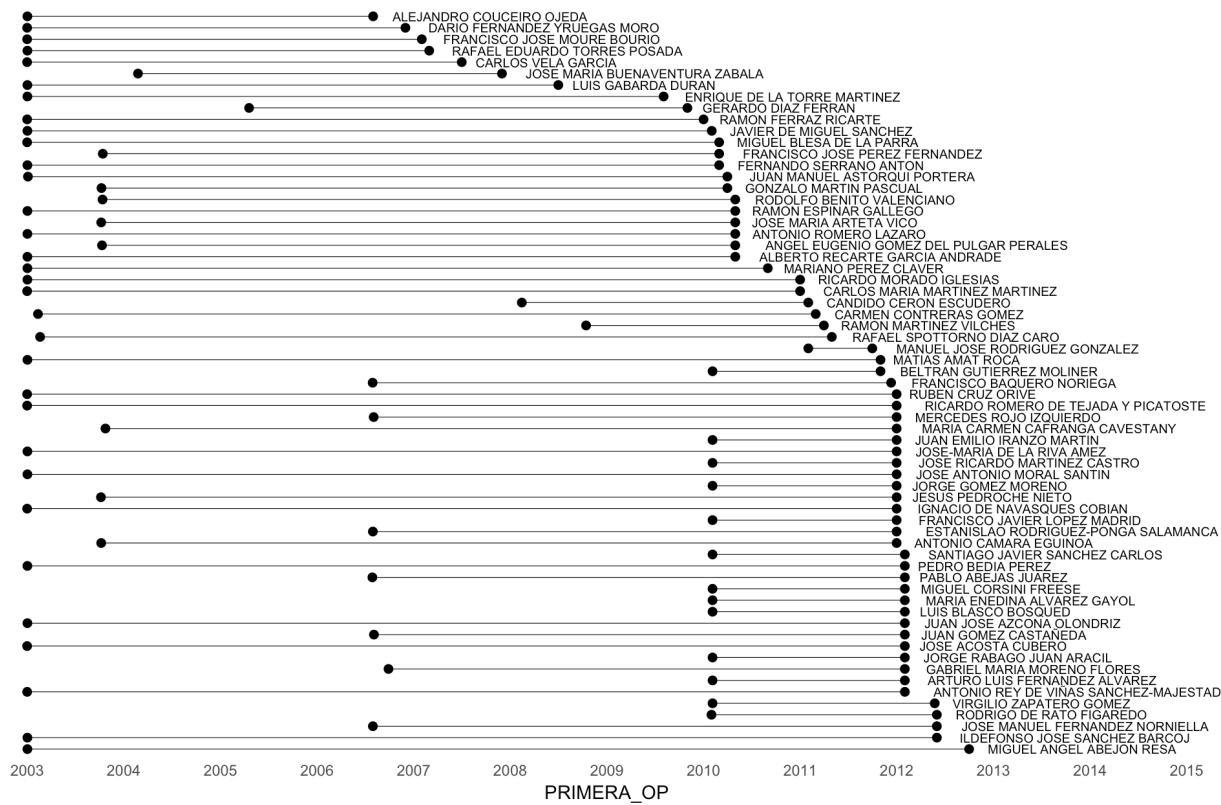


4. Realizar un boxplot en el cada punto en vez de una operación sea el importe total gastado por cada titular, de esta forma podremos comparar sus principales estadísticas.
5. ¿Quién es el acusado que más ha gastado dentro de cada organización?

Ej. #9

Actividad de los acusados

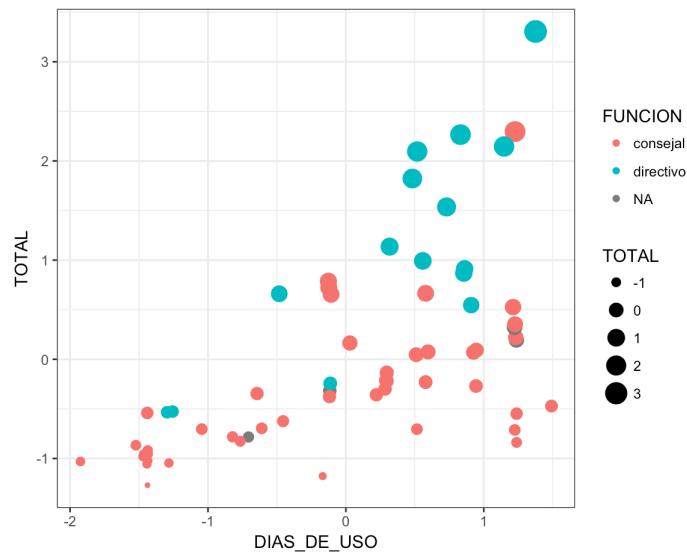
1. Obtener la fecha de la primera y última operación de cada titular y representarla gráficamente con el fin de visualizar su actividad



Ej. #10

Actividad de los acusados

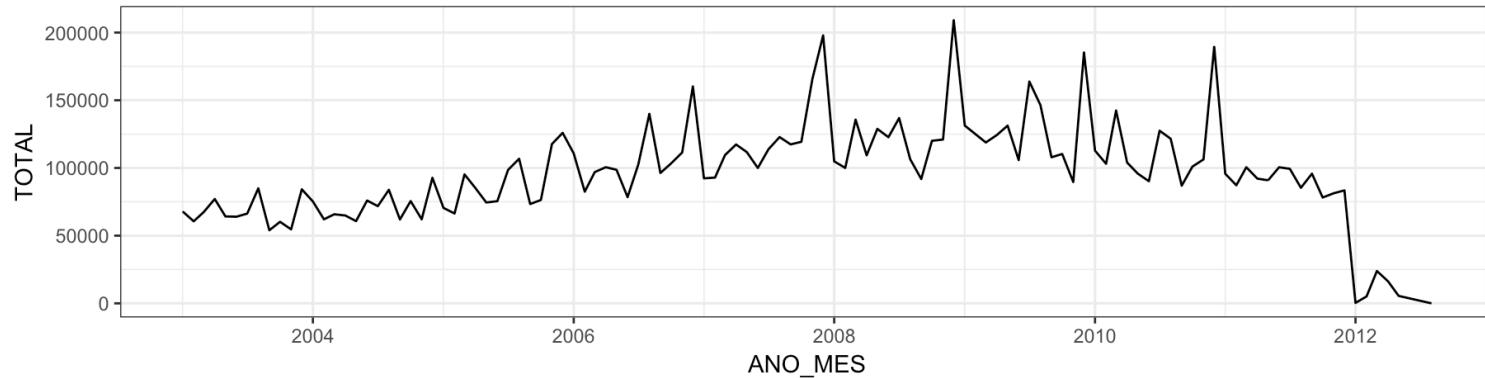
1. Vamos a comparar el tiempo que ha estado activo cada usuario con el dinero gastado mostrando a la vez su función dentro de Bankia



Ej. #11

Evolución temporal del gasto

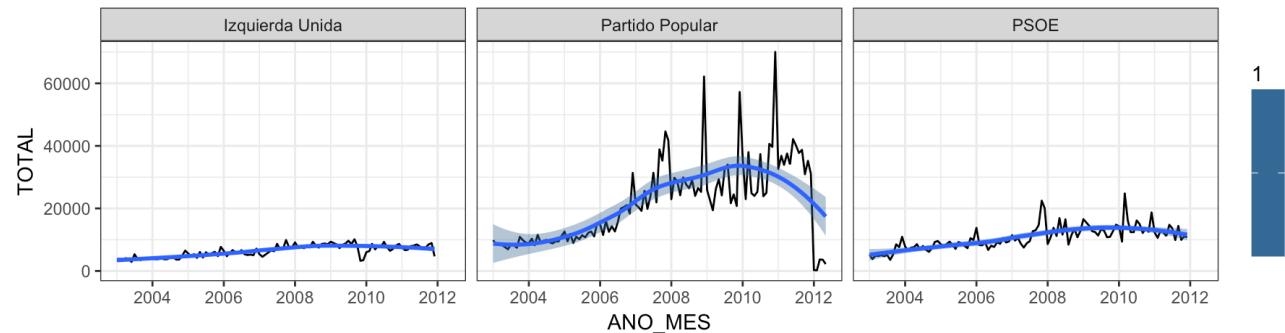
1. Muestra gráficamente la evolución mensual del gasto durante todo el periodo



2. Modifica la visualización siguiente para mostrar una comparativa entre PP, PSOE y Izquierda Unida

Ej. #12

Evolución temporal del gasto



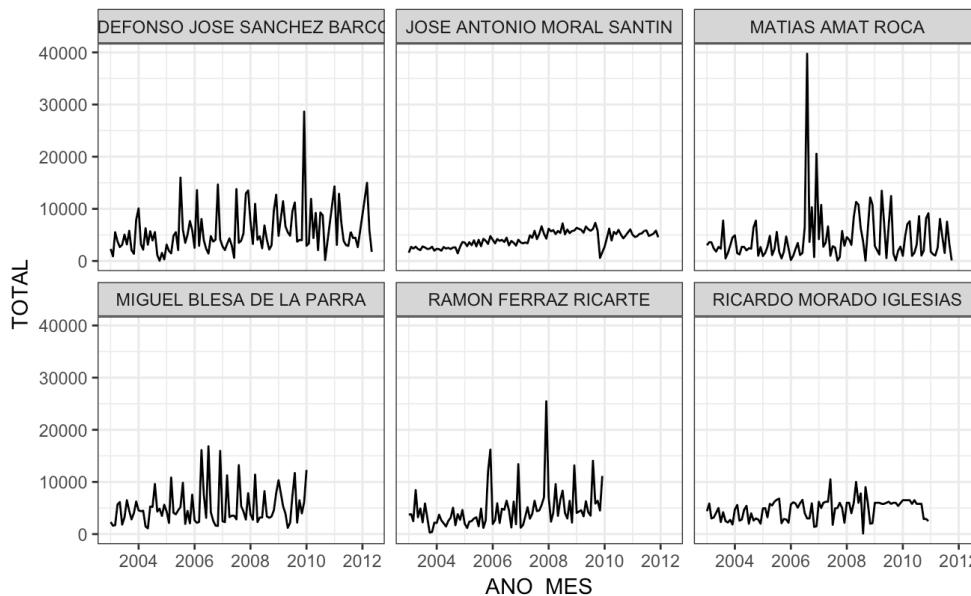
La anterior visualización no refleja el peso de cada partido en función del número de miembros de cada uno de ellos en el consejo

1. Visualiza primero la evolución temporal del número de consejeros de los tres partidos políticos
2. Visualiza ahora el gasto relativo dividiendo el gasto total entre el número de miembros de cada partido

Ej. #13

Los que más han gastado

1. Obtén el histórico de movimientos mensuales de los 6 que más han gastado. Para ello deberás primero averiguar cuáles son los seis acusados que más dinero han gastado.
2. Luego deberás obtener un gráfico como éste.



Ej. #14

Segmentación de acusados por comportamiento

1. Primero deberemos obtener un data set con el número de operaciones por acusado y por tipo de operación:

NOMBRE_Y_APELLIDOS <i><chr></i>	COD_OPE_DESC_AGRUPADO <i><chr></i>	NUM_OPERACIONES <i><int></i>
ALBERTO RECARTE GARCIA ANDRADE	COMPRA	<u>1307</u>
ALBERTO RECARTE GARCIA ANDRADE	DEVOLUCION COMPRA	3
ALBERTO RECARTE GARCIA ANDRADE	DEVOLUCION COMPRA (ABONO Y CARGO)	46
ALEJANDRO COUCEIRO OJEDA	ANUL.COMPRA,MARCAR/DESMARCAR FOTO(NMT)	1
ALEJANDRO COUCEIRO OJEDA	COMPRA	548

2. Deberás pivotar (`spread()`) la tabla por COD_OPE_DESC_AGRUPADO y reemplazar los NA con 0:

NOMBRE_Y_APELLIDOS <i><chr></i>	`AJUSTE DEBE SALD... <i><dbl></i>	`ANUL.COMPRA,MARC... <i><dbl></i>	`ANULACION DE C... <i><dbl></i>
ALBERTO RECARTE G...	0.	0.	0.
ALEJANDRO COUCEIR...	0.	1.	0.
ANGEL EUGENIO GOM...	0.	0.	0.
ANTONIO CAMARA EG...	0.	3.	0.
ANTONIO REY DE VI...	0.	0.	1.

Ej. #14 (cont)

Segmentación de acusados por comportamiento

Agrupa los titulares en 8 grupos para responder a: ¿Quién ha destacado por realizar retiradas de efectivo?

