

Capítol 4

Data Frames

Objectius: Introduir l'estructura `dataframe` de R i el format i la seva utilització en R.

4.1 Exercicis

1. Dissenyeu una funció que rebí un `dataframe` i el nom d'una de les seves variables i retorni la mitjana d'aquesta variable (amb arrodoniment a 4 xifres decimals).
2. Dissenyeu una funció que rebí un `dataframe` i el nom d'una de les seves variables i retorni la mediana d'aquesta variable.
3. Tenim un `dataframe` amb dades dels estudiants de Programació del Grau d'Estadística. El `dataframe` `D` té, per a cada estudiant, les dades següents:

```
D: nom niub nota_prog nota_mitjana
```

Fes una funció que, donat aquest `dataframe` `D` i la nota mitjana històrica de programació a l'escola `MeanProg`, retorni un vector amb els noms dels estudiants que compleixin que la seva nota de programació sigui més gran que la seva pròpia mitjana i també que la mitjana històrica de programació.

4. Siguin X i Y dues variables estadístiques amb el mateix nombre d'observacions corresponents n . Si \bar{X} denota la mitjana d'una variable X , el coeficient de *correlació de Pearson* entre X i Y es defineix com:

$$r_{XY} = \frac{\sum((X_i - \bar{X}) * (Y_i - \bar{Y}))}{\sqrt{\sum(X_i - \bar{X})^2 * \sum(Y_i - \bar{Y})^2}}$$

Dissenyeu una funció en R tal que donat un `data frame` i dues variables, X, Y , que corresponen a dues columnes del `data frame`, calculi i retorni el valor de la covariància r_{XY} .

Podeu usar la funció `mean()` de R.

5. Disposem de dades sobre els matrimonis entre dones, entre homes i heterosexuals en cada mes de l'any 2012 (fitxer matrimonis.txt). Es demana:
 - (a) Dissenyeu una funció que retorni si en algun mes el nombre de matrimonis entre dones va ser més gran que el de matrimonis entre homes.
 - (b) Dissenyeu una funció que retorni quin mes la proporció de matrimonis entre parelles del mateix sexe va ser més gran respecte el total de matrimonis.
 - (c) Dissenyeu una funció que retorni quantes persones es van casar el mes que es va casar més gent.

6. Tenim un data frame amb els resultats de les eleccions al Parlament de Catalunya des del 1999 al 2012 (5 eleccions) on la primera columna conté els anys en que hi va haver eleccions i les columnes següents són els resultats obtinguts pels 5 primers partits (per aquest ordre, CIU, ERC, PSC, PP i ICV) en cada un dels anys. Sempre rebent com a entrada el dataframe `eleccions`, dissenyeu:
 - (a) Una funció que retorni quin any el PP va obtenir el seu pitjor resultat i quin resultat va ser. Podeu retornar les dues dades en un vector.
 - (b) Una funció que retorni quin any la diferència d'escons entre CIU i PSC va ser més gran i quina va ser aquesta diferència. Podeu retornar les dues dades en un vector.
 - (c) Una funció que retorni quin partit va obtenir més escons l'any 2003 i quants escons van ser. Podeu retornar les dues dades en un vector.
 - (d) Una funció que retorni un nou dataframe amb els anys de les eleccions i el partit que va obtenir més escons en cada una.

7. *Segons mitjana (Examen de recuperació 13/14)*. Feu una funció que donats un data frame `dades` i el nom de dues variables (columnes) `var1` i `var2` modifiqui el data frame `dades` afegint una columna `Result` de forma que el valor corresponent en aquesta nova columna es calculi de la següent manera:

Si en la fila corresponent el valor de `var1` és superior a la mitjana d'aquesta mateixa variable (mitjana de la columna `var1`), el resultat serà el valor de `var1` + el valor de `var2`. Si per contra, en la fila corresponent el valor de `var1` no és superior a la mitjana d'aquesta mateixa variable (mitjana de la columna `var1`), el resultat serà el valor de `var1` – el valor de `var2`.

Podeu usar qualsevol de les funcions que ofereix R que necessiteu.

Per exemple, si tenim el data frame

```

  v1 v2 v3 v4
1  1  2  3  4
2  2  2  3  2
3  2  4  2  3
4  4  2  4  2

```

i les variables són "v4" i "v1", la mitjana de la columna amb nom "v4" és 2.75 i per tant el data frame resultant ha de ser:

```

  v1 v2 v3 v4 Result
1  1  2  3  4      5
2  2  2  3  2      0
3  2  4  2  3      5
4  4  2  4  2     -2

```

8. *Sobre mitjana (Examen final 13/14)*. Feu una funció que donats un data frame `dades` i el nom d'una variable (columna) `var` retorni un altra data frame que contingui totes les files de `dades` tals que el valor corresponent a la variable `var` sigui superior a la mitjana d'aquesta mateixa variable (columna).

Podeu usar qualsevol de les funcions que ofereix R que necessiteu.

Per exemple, si tenim el data frame

```

  v1 v2 v3 v4
1  1  2  3  4
2  2  2  3  2
3  2  4  2  3
4  4  2  4  2

```

i la variable és "v4", la mitjana de la columna amb nom "v4" és 2.75 i per tant el data frame resultant ha de ser:

```

  v1 v2 v3 v4
1  1  2  3  4
3  2  4  2  3

```

9. El dataframe `trees` conté dades sobre l'alçada, el diàmetre i el volum de 31 cirerers talats. Rebut com a entrada `trees`,

```

      Girth Height Volume
1     8.3     70  10.3
2     8.6     65  10.3
3     8.8     63  10.2
...
30    18.0     80  51.0
31    20.6     87  77.0

```

(a) Feu una funció que retorni quants arbres tenen alçada (Height) més gran que 85 i diàmetre (Girth) més gran que 10.

(b) Afegiu al dataframe les dades de 2 arbres nous amb variables:

```
# Arbre 1: Girth=9.2, Height=80 i Volume=11.2
# Arbre 2: Girth=14, Height=69 i Volume=33.6
```

(c) Feu una funció que retorni quin volum té l'arbre més alt. Utilitzeu-la per saber si és el volum més alt de tots els arbres.

10. *Analfabetisme vs. salari (Examen final única 13/14)*. Fes una funció que donat el data frame `estats`, descrit a continuació, retorni cert si i només si l'estat que té el mínim índex d'analfabetisme (variable `Analfabetisme`) és el mateix que té el màxim salari (variable `Salari`).

Podeu considerar que no hi han valors iguals.

El data frame `estats` té les dades en el format següent:

	Poblacio	Salari	Analfabetisme	Esp_vida	Regio
Alabama	3615	3624	2.1	69.05	South
Alaska	365	6315	1.5	69.31	West
.....					
Wisconsin	4589	4468	0.7	72.48	North_Central
Wyoming	376	4566	0.6	70.29	West

11. El dataframe `Traffic` conté dades d'accidents de trànsit en els anys 1961 i 1962 a Suècia.

L'heu de carregar de la llibreria `MASS` amb el codi següent:

```
library(MASS)
data(Traffic)
```

I les dades són com segueix:

	year	day	limit	y
1	1961	1	no	9
2	1961	2	no	11
3	1961	3	no	9
...				
183	1962	91	yes	15
184	1962	92	yes	9

- (a) Feu una funció que rebí el dataframe `Traffic` i retorni la diferència entre la mitjana dels dies en què hi ha límit de velocitat i la dels dies que no.

- (b) Feu una funció que rebi l'any (1961 o 1962) i retorni el màxim d'accidents d'aquest any i si el dia en que es va donar aquest màxim hi havia límit de velocitat o no.

12. Disposem de dos dataframes que contenen informació sobre els diferents estats dels Estats Units. Cada fila correspon a un estat i cada columna a una variable. Per més informació de les dades podeu consultar la help de R mitjançant

```
?state
```

Després de llegir les dades

```
estats1 <- read.table("estats1.txt", header=T, stringsAsFactors=FALSE)
estats2 <- read.table("estats2.txt", header=T, stringsAsFactors=FALSE)
```

es demana

- (a) Dissenyeu una funció que rebi com a entrada un dels dataframes i retorni quina és la regió de l'estat amb l'àrea més gran.
- (b) Amb la informació obtinguda en l'apartat anterior (i, per tant, utilitzant la funció que heu fet), dissenyeu una funció que rebi els dos dataframes i calculi la mitjana de població per als estats de la regió obtinguda en l'apartat anterior.

13. *Any estrena (Examen final 15/16).*

Tenim un **data frame** on s'emmagatzema informació de pel·lícules de manera que, per a cada títol de pel·lícula, tenim la durada en minuts i l'any en que es va estrenar:

```
D1: titol, durada, any_estrena
```

Implementa una funció que donat un data frame D1 amb les dades de les pel·lícules, retorni l'any en que es van estrenar més pel·lícules. En cas que hi hagi més d'un any que compleixin la condició, retornarà el primer en què es va complir (per exemple, si es compleix en 2008 i 2010, ha de retornar 2008).

Per exemple, si tenim el data frame:

```
D1:   titol  durada  any_estrena
     1  peli1    120      2012
     2  peli2    100      2010
     3  peli3     90      2010
     4  peli4     75      2009
     5  peli5    135      2011
```

la funció hauria de retornar:

```
2010
```

14. El dataset `co2` conté la concentració de CO2 mesurada mensualment des de l'any 1959 fins al 1997. Es vol comprobar si les distribucions anuals es mantenen o no constants al llarg dels anys. És a dir efectuar un contrast d'homogeneïtat:

H_0 : Les r poblacions tenen una distribució en comú

H_1 : Les r poblacions no tenen una distribució en comú

Sent r els 38 anys.

Cal dissenyar una funció que donat un `data.frame`, calculi l'estadístic χ_{exp}^2 i el compari amb el valor crític de referència. Ha de retornar TRUE en cas que H_0 sigui rebutjada. Considereu un nivell de significació del 5%.

Podeu carregar les dades executant `data("co2")`. Seguidament, executeu la següent comanda per a transformar les dades de classe `ts` a `matrix`:

```
co2=taapply(co2, list(year = floor(time(co2)),
                    month = month.abb[cycle(co2)]), c)
```

Sent r i c el nombre de files i columnes del `data.frame`. I n_{ij} l'element de la fila i i la columna j . L'estadístic χ_{exp}^2 ve definit:

$$\chi_{exp}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} \sim \chi_{(r-1)(c-1)}^2$$

On $n_{i\bullet} = \sum_{j=1}^c n_{ij}$ i $n_{\bullet j} = \sum_{i=1}^r n_{ij}$.

Rebutgeu H_0 si: $\chi_{exp}^2 > \chi_{(r-1)(c-1); 1-\alpha}^2$.

Per a calcular el valor crític $\chi_{(r-1)(c-1); 1-\alpha}^2$ amb R, podeu utilitzar la següent instrucció:

```
qchisq(1-alpha,df=(nrow(co2)-1)*(ncol(co2)-1))
```

15. Per tal d'estudiar l'efecte que causa el tabac sobre l'agregació de les plaquetes en la sang, Levine (1973) va dissenyar un estudi per mesurar el grau al qual les plaquetes es van agregar en individus abans i després de fumar. Les plaquetes estan implicades en la formació de coàguls de sang, i se sap que els fumadors pateixen més sovint de desordres que impliquen la seva formació que no pas les persones no fumadores.

Es van fer dos estudis de l'agregació de plaquetes, amb els mateixos individus, abans i després de fumar tabac i abans i després de fumar enciam. Les dades de totes dues taules les teniu als arxius `tabac.txt` i `enciam.txt` respectivament.

Les dades d'aquests dos arxius tenen el format següent (tots dos iguals):

Persona	Abans	Despres	Diferencia
1	25	24	-1
2	25	30	5
...
10	60	62	2
11	68	70	2

Si denotem per X i Y les variables aleatòries que mesuren el percentatge màxim d'agregació de plaquetes a la sang abans i després de fumar un cigarret d'enciam, respectivament columnes **Abans** i **Despres** en l'arxiu, es demana

- (a) Feu una funció que donat un dels data frames anteriors, comprovi si la mitjana de la diferència és la diferència de les mitjanes de les variables X i Y i retorni un list amb la següent informació: Un booleà que serà cert o fals responent a la pregunta anterior i la mitjana de la diferència que s'ha calculat. No podeu usar la funció `mean` de R.
- (b) Feu una funció que donat un dels data frames anteriors, calculi i retorni les variàncies de X , de Y i de la diferència. Podeu retornar els tres valors en un únic vector.

Recordeu que la fórmula per calcular la variància d'una variable X és

$$Var(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Es recomana que definiu una funció que us retorni la variància d'una certa variable donada (variància d'un vector de valors).

- (c) Feu una funció que donat un dels data frames anteriors i usant la crida a la funció de l'apartat b), comprovi la següent fórmula i retorni cert o fals depenent de si la fórmula es compleix o no.

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

Recordeu que la fórmula per calcular la covariància entre dues variables X i Y és

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

- (d) Feu una funció que donat un dels data frames anteriors i usant les funcions anteriors que us siguin necessàries, retorni el valor de la mesura de discrepància.

$$d = \frac{\bar{D}}{S_D/\sqrt{n}}$$

$$S_D = \sqrt{Var(D)}$$

16. Basant-nos en els mateixos arxius que l'exercici anterior, `tabac` i `enciam`, es demana fer una funció que donats tots dos data frames i usant totes les funcions necessàries de l'exercici anterior, calculi i retorni el valor de la mesura de discrepància entre la diferència d'abans i després de fumar tabac i la diferència d'abans i després de fumar enciam (cal calcular la diferència d'interès $D = D_{Tabac} - D_{Enciam}$).

17. *Què fem?* (Examen final 15/16).

La funció `queFa` `<- function (D1, D2)` està implementada com:

```
queFa <- function(D1,D2){
  DM <- merge (D1, D2)
  a <- vector()
  for (i in 1:nrow(DM)){
    a <- c(a, DM[i,"ingressos"]-DM[i,"despeses"])
  }
  DB <- cbind (DM, S=a)
  DB <- DB[order(DB$nom),]
  return(DB$$S)
}
```

El data frame `D1` conté el camp "nom" i el camp "ingressos" mentre que el data frame `D2` conté el camp "nom" i el camp "despeses".

Si els data frames d'entrada de la funció són:

```
D1 <- data.frame (nom=c("Pep","Jana","Robert","Maria","Anna","Albert","Kim"),
  ingressos=c(5500,7000,7200,6400,6700,7500,6000),
  stringsAsFactors=FALSE)

D2 <- data.frame (nom=c("Jana","Robert","Maria","Albert","Kim","Pep","Isabel"),
  ingressos=c(5500,6500,6000,8000,5500,5000,4700),
  stringsAsFactors=FALSE)
```

quina serà la sortida de la funció `queFa (D1, D2)`?

18. *Despeses de la compra* (Examen de recuperació 14/15) Tenim dos dataframes amb informació dels productes que ha comprat cada persona i el preu d'aquests productes. En el primer dataframe `D1` tenim informació de la persona i el que ha comprat (quantitat):

```
D1:  nom producte Kg
```

I en el segon dataframe `D2` tenim la informació del que costa cada producte per Kg:

```
D2:  producte preuKg
```

Es demana implementar una funció que, donats els dos dataframes D1 i D2 generi un tercer dataframe D3 que contingui la despesa de la compra, de manera que per a cada persona tindrem la suma de la despesa de tots els productes que ha comprat. Aquest dataframe resultat ha d'estar ordenat per nom.

D3: nom despesa

Per exemple, si tenim:

D1:	nom	producte	Kg	D2:	producte	preuKg
1	Pep	pomes	5	1	pomes	2.5
2	Kim	peres	3	2	peres	3.0
3	Pep	peres	1.5	3	platans	4.5
4	Maria	pomes	3	4	cireres	6.0
5	Maria	platans	2	5	figues	5.5
6	Albert	cireres	3.5			
7	Pep	figues	5			

El resultat hauria de ser:

D3:	nom	despesa
1	Albert	21
2	Kim	9
3	Maria	16.5
4	Pep	44.5

19. Tenim un dataframe amb dades sobre els desplaçaments d'estudiants universitaris per comarques (fitxer desplaçaments_uni.txt). Tenint en compte que el total d'universitaris atrets per una comarca és la suma dels interns, dels des de fora i dels que no consta i que el total d'estudiants generats són els interns i els que marxen a fora, es demana
- Feu una funció que retorni quina comarca atreu més estudiants.
 - Feu una funció que retorni quina comarca genera menys estudiants.
 - Dissenyeu una funció que retorni quants desplaçaments interns té la comarca que genera més estudiants.
 - Dissenyeu una funció que retorni quina és la diferència més gran entre universitaris atrets i generats.
 - Feu una funció que retorni quantes comarques hi ha que no atreguin cap estudiant.

20. *Guanyos esportius (Examen final 14/15)*. Tenim dos dataframes amb informació dels tornejos guanyats enguany. En el primer dataframe D1 tenim informació de l'equip que ha guanyat cada torneig:

```
D1: equip torneig
```

I en el segon dataframe D2 tenim la informació dels diners que es guanyen amb cada torneig (el campió del torneig):

```
D2: torneig recompensa
```

Es demana implementar una funció que, donats els dos dataframes D1 i D2 generi un tercer dataframe D3 que contingui la classificació dels equips pels seus guanyos, de manera que per a cada equip tindrem la suma de tots els seus guanyos (segons els diferents tornejos que hagi guanyat) i a més els tindrem ordenats de més guanyos a menys guanyos.

```
D3: equip guanyos
```

Per exemple, si tenim:

D1:	equip	torneig	D2:	torneig	recompensa
1	Barça	Lliga-1a	1	Lliga-1a	1500
2	Madrid	Lliga-ACB	2	Lliga-ACB	450
3	Girona	Lliga-2a	3	CopaRey	500
4	Girona	Gamper	4	Lliga-2a	150
5	Barça	CopaRey	5	Champions	3000
6	Barça	Champions	6	FinalFour	1000
7	Juventud	FinalFour	7	Gamper	200

El resultat hauria de ser:

```
D3:      equip  guanyos
1   Barça    5000
2  Juventud  1000
3   Madrid   450
4   Girona   350
```

21. *Examen de recuperació 15/16*

Tenim dos dataframes amb informació dels ingressos i despeses d'un conjunt d'empreses. Els ingressos estan en el primer dataframe (D1) i aquest està ordenat per nom de l'empresa:

```
D1: NIF nom ingressos
```

les dades dels ingressos estan en milers d'Euros i poden haver diferents ingressos per a una mateixa empresa.

I en el segon dataframe D2 tenim la informació de les despeses de cada empresa, però en aquest data frame només tenim el NIF de l'empresa no el nom. En el cas de les despeses només hi ha una per empresa. Aquest data frame està ordenat per NIF:

```
D2: NIF despeses
```

les dades de les despeses també estan en milers d'Euros.

Es demana implementar una funció que, donats els dos dataframes D1 i D2 generi un tercer dataframe D3 que contingui només el conjunt d'empreses que han hagut de cotitzar a hisenda. Les empreses que han hagut de cotitzar a hisenda són aquelles que han tingut guanys superiors als 500 mil Euros (500 en milers d'Euros). En aquest tercer dataframe (D3), a més de les dades de l'empresa (NIF i nom), han d'aparèixer les columnes **ingressos**, **despeses** i **guanys**, on aquesta última es calcula fent la diferència entre els ingressos totals (que són la suma dels diferents ingressos de l'empresa) i les despeses.

Aquest dataframe resultat ha d'estar ordenat per NIF.

```
D3: NIF nom ingressos despeses guanys
```

Per exemple, si tenim:

D1:	nom	NIF	ingressos	D2:	NIF	despeses
1	Abacus	J33244098	50700	1	A33232411	250050
2	Amazon	Q44488877	200000	2	B25880945	700500
3	Amazon	Q44488877	100000	3	F66787231	380000
4	Apple	G28282828	200000	4	G28282828	580500
5	Apple	G28282828	200000	5	H10098477	74700
6	Apple	G28282828	400000	6	J33244098	50500
7	Ikea	F66787231	550000	7	Q44488877	230000
8	Nike	A33232411	250500			
9	Samsung	B25880945	500500			
10	Samsung	B25880945	250200			
11	Watch	H10098477	75050			

El resultat hauria de ser:

D3:	NIF	nom	ingressos	despeses	guanys
1	B25880945	Samsung	750700	700500	50200
2	F66787231	Ikea	550000	380000	170000
3	G28282828	Apple	800000	580500	219500
4	Q44488877	Amazon	300000	230000	70000

22. El dataframe `iris` conté dades sobre les dimensions (longitud i amplada) dels sèpals i pètals, i l'espècie de 150 flors. Rebent com a entrada `iris`:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
...					
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
...					
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3.0	5.1	1.8	virginica

- (a) Dissenyeu una funció que rebi el dataframe `iris`, el nom d'una de les seves variables i el nom d'una de les seves espècies i retorni la mitjana de la variable per a l'espècie indicada.
- (b) Dissenyeu una funció que rebi el dataframe `iris` i retorni un altre dataframe amb les mitjanes de cada variable de flor en cada espècie, és a dir es busca un dataframe amb les dades següents: (pots considerar sabut que només hi ha 3 espècies diferents)

	Specie	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	setosa	meanSL-set	meanSW-set	meanPL-set	meanPW-set
2	versicolor	meanSL-ver	meanSW-ver	meanPL-ver	meanPW-ver
3	virginica	meanSL-vir	meanSW-vir	meanPL-vir	meanPW-vir

23. Disposem d'un dataframe amb dades sobre el nombre d'adopcions a Catalunya segons el país d'origen.

- (a) Dissenyeu una funció que rebi el dataframe i retorni de quin país hi va haver més adopcions entre els anys 2005 i 2010.
- (b) Dissenyeu una funció que tingui com a entrada el dataframe, dos anys (entre 2005 i 2012) i un país i retorni la diferència entre les adopcions en aquell país entre els dos anys en valor absolut.
- (c) Dissenyeu una funció que tingui com a entrada el dataframe i un any i retorni quantes adopcions d'Àfrica hi va haver aquell any.
- (d) Feu una funció que rebi com a entrada el dataframe i retorni un altre dataframe on cada fila sigui un any i cada columna un continent, obtenint així quantes adopcions hi va haver en cada any de cada continent.

- (e) Utilitzeu la funció definida en l'apartat anterior per construir-ne una altra que rebí el dataframe original i un any i retorni de quin continent va haver-hi més adopcions aquell any.

24. Disposem de tres dataframes. En el fitxer `estudiants.txt` hi ha informació sobre estudiants i el curs al qual estan matriculats, en `cap_aules.txt` hi trobem informació sobre aules i la seva capacitat i `aules_curs.txt` relaciona els cursos amb les aules on s'imparteixen.

- (a) Feu una funció que retorni els noms dels estudiants matriculats en els cursos que es fan a l'aula A1.
- (b) Tenint en compte el nombre d'estudiants matriculats a cada curs, feu una funció que retorni quins cursos estan programats en aules que no tenen prou capacitat.

Tenim ara un altre dataframe `curs_coord` amb la informació sobre els coordinadors de cada curs i la seva nacionalitat.

- (c) Dissenyeu una funció que retorni quina és la nacionalitat del professor del curs amb menys alumnes matriculats.
- (d) Feu una funció que donada una aula retorni els noms dels professors que fan classe en aquella aula.
- (e) Utilitzant la funció de l'apartat b), feu una funció que retorni els noms dels professors que hauran de buscar aules amb més capacitat pels seus cursos.

25. Disposem de dos dataframes amb informació sobre clients en un supermercat. El primer dataframe, que està en el fitxer `cl_prod.txt` conté informació sobre els productes de cada client. El segon, `caixes`, conté les caixes, el nombre de productes que té assignat cada caixa en aquest moment i el màxim de productes que pot tenir.

- (a) Donat un número de client, feu una funció que retorni a quines caixes podria anar.
- (b) Tenint en compte que les caixes sempre s'han d'omplir per ordre (és a dir, un client no pot anar a la caixa 3 si pot anar a la caixa 2), feu una funció que actualitzi el dataframe `caixes` amb els clients que tenim. La funció ha de retornar el dataframe actualitzat i un vector amb els clients que no han pogut ser col·locats en cap caixa.

26. Es té un `dataframe` amb els resultats de la Lliga d'un cert any. A cada fila hi ha els resultats d'un equip (amb els equips ordenats per ordre alfabètic) i la informació que tenim per a cada equip és: Partits Jugats (PJ), Partits Guanyats (PG), Partits Empatats (PE), Gols a Favor (GF) i Gols en Contra (GC). Fent sempre funcions que rebin com a entrada el `dataframe` amb les dades de la lliga, feu
- (a) Dissenyeu una funció que retorni el nom de l'equip que ha guanyat la Lliga.
 - (b) Dissenyeu una funció que retorni quins equips han baixat a Segona Divisió. Recordeu que els equips que baixen a Segona Divisió són els tres últims de la classificació.
 - (c) Dissenyeu una funció que retorni quin equip té la diferència més petita entre gols marcats i gols rebuts.
 - (d) Dissenyeu una funció que retorni `TRUE` o `FALSE` segons si en la classificació final hi ha hagut equips empatats a punts o no.
 - (e) Dissenyeu una funció que rebi els noms de dos equips i retorni la diferència de punts entre ells.
 - (f) Dissenyeu una funció que rebi els noms de dos equips i retorni quantes posicions els van separar.