

# VD-PRAC04

JUAN ANTONIO VERA

2025-01-16

```
## Cargando paquete requerido: dplyr
##
## Adjuntando el paquete: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
## Cargando paquete requerido: ggplot2
## Cargando paquete requerido: fmsb
## Cargando paquete requerido: corrplot
## corrplot 0.95 loaded
## Cargando paquete requerido: reshape2
## Cargando paquete requerido: knitr
##
## Adjuntando el paquete: 'tidyr'
## The following object is masked from 'package:reshape2':
##
##   smiths
```

## 1. Introducció

En aquest document es realitza una exploració dels resultats de l'informe PISA durant aquests darrers anys a Catalunya.

L'objectiu serà trobar tendències sobre l'evolució dels resultats i possibles diferències entre col·lectius

#2. Anàlisi Exploratri de Dades (EDA)

L'objectiu és realitzar una anàlisi exploratòria de dades (EDA) del dataset «Avaluació de quart d'Educació Secundària Obligatoria» disponible a <https://datos.gob.es/gl/catalogo/a09002970-evaluacion-de-cuarto-de-educacion-secundaria-obligatoria>. Farem un primer estudi sobre la normalitat de les dades que es presenten.

Per simplificar, renombrarem l'arxiu descarregat com a notes.csv

#Càrrega i organització de les dades

A través de les següents ordres carreguem les dades de l'arxiu notes.csv al data frame df, crearem algunes noves columnes categòriques que ens permetin una millor visualització dels resultats

Fem una exploració de possibles valors nuls

Table 1: NAs de cada variable

	x
PCAT_CL	30453
PCAT_EE	30453
PCAST	39157
PCAST_CL	40751
PCAST_EE	40751
PANG	33478
PANG_CO	33478
PANG_CL	33478
PANG_EE	33478
PFRAN	712826
PFRAN_CO	713599
PFRAN_CL	713599
PFRAN_EE	713599
PMAT	24895
PMAT_EFM	85427
PMAT_CR	85427
PMAT_EST	85427
PMAT_NC	85427
PALE	777838
PALE_CO	777838
PALE_CL	777838
PALE_EE	777838
PCIEN	253623
PCIEN_COMP1	253623
PCIEN_COMP2	253623
PCIEN_COMP3	253623
PCIEN_COMP4	655599
GENERE	0
MES_NAIXEMENT	7805
ANY_NAIXEMENT	1971
NATURALES	0
ÀREA.TERRITORIAL	0
HÀBITAT	0

El dataset conté el resultat de tots els alumnes de 4 de la ESO. Trobem que hi han moltes columnes amb NAs perquè es corresponen a matèries que l'alumne no s'ha avaluat perquè no li corresponen avaluació.

Per tot això, es consideren normal, però no es tindran en compte per extreure informació estadística

## Resumen y visualización de la composición los datos

A continuació es mostren algunes ordres que ens ajuden a entendre la composició de la nostra base de dades. Amb les ordres `head` i `tail` ens permeten veure les primeres i últimes 6 files del data frame. L'ordre `names` observem els noms de les columnes de la matriu. L'ordre `str` llista el conjunt de variables que es troben a la base de dades juntament amb algunes dades particulars que destaquen el nombre i la naturalesa de les variables. Finalment, la comanda `summary` vindria a ser la més útil i ens ofereix un resum de les variables tant numèriques com no numèriques, les seves freqüències, quantitat, mitjana, mitjana i valors màxims i mínims.

##	ANY	CODI_ALUMNE	PCAT	PCAT_CL	PCAT_EE	PCAST	PCAST_CL	PCAST_EE	PANG	PANG_CO
## 1	2012	NA	85.33	92	78.65	94.96	96.17	93.75	90.00	80.00
## 2	2012	NA	68.25	76	60.50	67.40	69.17	65.63	NA	NA
## 3	2012	NA	69.76	76	63.53	72.94	61.50	84.38	74.78	93.33
## 4	2012	NA	88.35	92	84.70	91.48	92.33	90.63	92.00	100.00
## 5	2012	NA	90.35	96	84.70	94.17	88.33	100.00	92.00	100.00
## 6	2012	NA	92.40	88	96.80	91.04	88.33	93.75	92.44	93.33
##	PANG_CL	PANG_EE	PFRAN	PFRAN_CO	PFRAN_CL	PFRAN_EE	PMAT	PMAT_EFM	PMAT_CR	
## 1	100	90	NA	NA	NA	NA	82.6	71.54	75.0	
## 2	NA	NA	NA	NA	NA	NA	52.5	50.77	0.0	
## 3	76	55	NA	NA	NA	NA	47.3	89.62	32.5	
## 4	96	80	NA	NA	NA	NA	91.8	89.62	82.5	
## 5	96	80	NA	NA	NA	NA	96.0	100.00	95.0	
## 6	84	100	NA	NA	NA	NA	82.5	100.00	82.5	
##	PMAT_EST	PMAT_NC	PALE	PALE_CO	PALE_CL	PALE_EE	PCIEN	PCIEN_COMP1	PCIEN_COMP2	
## 1	87.50	93.33	NA	NA	NA	NA	NA	NA	NA	
## 2	100.00	51.00	NA	NA	NA	NA	NA	NA	NA	
## 3	0.00	58.33	NA	NA	NA	NA	NA	NA	NA	
## 4	100.00	93.33	NA	NA	NA	NA	NA	NA	NA	
## 5	87.50	100.00	NA	NA	NA	NA	NA	NA	NA	
## 6	66.67	80.00	NA	NA	NA	NA	NA	NA	NA	
##	PCIEN_COMP3	PCIEN_COMP4	GENERE	MES_NAIXEMENT	ANY_NAIXEMENT	NATURALES				
## 1	NA	NA	D	10	1996	Privada				
## 2	NA	NA	H	9	1996	Privada				
## 3	NA	NA	H	10	1996	Privada				
## 4	NA	NA	H	11	1996	Privada				
## 5	NA	NA	D	3	1996	Privada				
## 6	NA	NA	H	7	1996	Privada				
##	ÀREA.TERRITORIAL		HÀBITAT							
## 1	Maresme - Vallès Oriental	De 10001 a 100000								
## 2	Maresme - Vallès Oriental	De 10001 a 100000								
## 3	Maresme - Vallès Oriental	De 10001 a 100000								
## 4	Maresme - Vallès Oriental	De 10001 a 100000								
## 5	Maresme - Vallès Oriental	De 10001 a 100000								
## 6	Maresme - Vallès Oriental	De 10001 a 100000								
##	ANY	CODI_ALUMNE	PCAT	PCAT_CL	PCAT_EE	PCAST	PCAST_CL	PCAST_EE	PANG	
## 778548	2020	3259050501	80.35	86	71.88	84.31	92.59	71.88	85.6	
## 778549	2020	3258431796	NA	NA	NA	65.42	61.11	71.88	NA	
## 778550	2020	3257868411	79.15	84	71.88	78.19	74.07	84.38	81.2	
## 778551	2020	3258228375	74.15	84	59.38	89.44	90.74	87.50	73.2	
## 778552	2020	3258487126	85.20	92	75.00	86.94	90.74	81.25	85.8	
## 778553	2020	3258127160	91.25	100	78.13	85.83	88.89	81.25	95.4	
##	PANG_CO	PANG_CL	PANG_EE	PFRAN	PFRAN_CO	PFRAN_CL	PFRAN_EE	PMAT	PMAT_EFM	
## 778548	93.33	84	80	NA	NA	NA	NA	71.19	80	
## 778549	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 778550	93.33	88	60	NA	NA	NA	NA	55.36	50	
## 778551	86.67	88	40	NA	NA	NA	NA	83.37	70	
## 778552	93.33	92	70	NA	NA	NA	NA	91.67	100	
## 778553	100.00	96	90	NA	NA	NA	NA	100.00	100	
##	PMAT_CR	PMAT_EST	PMAT_NC	PALE	PALE_CO	PALE_CL	PALE_EE	PCIEN	PCIEN_COMP1	
## 778548	88.89	44.44	71.43	NA	NA	NA	NA	81.6	59.09	
## 778549	NA	NA	NA	NA	NA	NA	NA	64.1	68.18	
## 778550	44.44	55.56	71.43	NA	NA	NA	NA	61.7	59.09	

```

## 778551 88.89 88.89 85.71 NA NA NA NA 69.7 59.09
## 778552 77.78 88.89 100.00 NA NA NA NA 82.6 81.82
## 778553 100.00 100.00 100.00 NA NA NA NA 89.6 95.45
## PCIEN_COMP2 PCIEN_COMP3 PCIEN_COMP4 GERE MES_NAIXEMENT ANY_NAIXEMENT
## 778548 85.71 100.00 NA D NA 2004
## 778549 42.86 81.25 NA H NA 2003
## 778550 57.14 68.75 NA H NA 2004
## 778551 50.00 100.00 NA D NA 2004
## 778552 78.57 87.50 NA D NA 2004
## 778553 85.71 87.50 NA D NA 2004
## NATURALES ÀREA.TERRITORIAL HÀBITAT
## 778548 Privada Tarragona Fins a 10000
## 778549 Privada Tarragona Fins a 10000
## 778550 Privada Tarragona Fins a 10000
## 778551 Privada Tarragona Fins a 10000
## 778552 Privada Tarragona Fins a 10000
## 778553 Privada Tarragona Fins a 10000

## [1] "ANY" "CODI_ALUMNE" "PCAT" "PCAT_CL"
## [5] "PCAT_EE" "PCAST" "PCAST_CL" "PCAST_EE"
## [9] "PANG" "PANG_CO" "PANG_CL" "PANG_EE"
## [13] "PFRAN" "PFRAN_CO" "PFRAN_CL" "PFRAN_EE"
## [17] "PMAT" "PMAT_EFM" "PMAT_CR" "PMAT_EST"
## [21] "PMAT_NC" "PALE" "PALE_CO" "PALE_CL"
## [25] "PALE_EE" "PCIEN" "PCIEN_COMP1" "PCIEN_COMP2"
## [29] "PCIEN_COMP3" "PCIEN_COMP4" "GERE" "MES_NAIXEMENT"
## [33] "ANY_NAIXEMENT" "NATURALES" "ÀREA.TERRITORIAL" "HÀBITAT"

## 'data.frame': 778553 obs. of 36 variables:
## $ ANY : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...
## $ CODI_ALUMNE : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ PCAT : num 85.3 68.2 69.8 88.3 90.3 ...
## $ PCAT_CL : num 92 76 76 92 96 88 88 84 76 96 ...
## $ PCAT_EE : num 78.7 60.5 63.5 84.7 84.7 ...
## $ PCAST : num 95 67.4 72.9 91.5 94.2 ...
## $ PCAST_CL : num 96.2 69.2 61.5 92.3 88.3 ...
## $ PCAST_EE : num 93.8 65.6 84.4 90.6 100 ...
## $ PANG : num 90 NA 74.8 92 92 ...
## $ PANG_CO : num 80 NA 93.3 100 100 ...
## $ PANG_CL : num 100 NA 76 96 96 84 64 76 64 96 ...
## $ PANG_EE : num 90 NA 55 80 80 100 70 30 55 60 ...
## $ PFRAN : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ PFRAN_CO : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ PFRAN_CL : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ PFRAN_EE : int NA NA NA NA NA NA NA NA NA NA NA ...
## $ PMAT : num 82.6 52.5 47.3 91.8 96 82.5 41.5 84.8 45.7 34.9 ...
## $ PMAT_EFM : num 71.5 50.8 89.6 89.6 100 ...
## $ PMAT_CR : num 75 0 32.5 82.5 95 82.5 0 95 30 12.5 ...
## $ PMAT_EST : num 87.5 100 0 100 87.5 ...
## $ PMAT_NC : num 93.3 51 58.3 93.3 100 ...
## $ PALE : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ PALE_CO : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ PALE_CL : int NA NA NA NA NA NA NA NA NA NA NA ...
## $ PALE_EE : int NA NA NA NA NA NA NA NA NA NA NA ...
## $ PCIEN : num NA NA NA NA NA NA NA NA NA NA NA ...

```

```

## $ PCIE_COMP1      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ PCIE_COMP2      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ PCIE_COMP3      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ PCIE_COMP4      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ GERE            : chr   "D" "H" "H" "H" ...
## $ MES_NAIXEMENT   : int   10 9 10 11 3 7 5 1 1 7 ...
## $ ANY_NAIXEMENT    : int   1996 1996 1996 1996 1996 1996 1995 1995 1996 1995 ...
## $ NATURALES      : chr   "Privada" "Privada" "Privada" "Privada" ...
## $ ÀREA.TERRITORIAL: chr   "Maresme - Vallès Oriental" "Maresme - Vallès Oriental" "Maresme - Vallès Oriental" ...
## $ HÀBITAT         : chr   "De 10001 a 100000" "De 10001 a 100000" "De 10001 a 100000" "De 10001 a 100000" ...

```

ANY	CODI_ALUMNE	PCAT	PCAT_CL
Min. :2012	Min. :5.465e+05	Min. : 0.00	Min. : 0.00
1st Qu.:2015	1st Qu.:2.946e+09	1st Qu.: 68.00	1st Qu.: 74.08
Median :2018	Median :7.057e+09	Median : 77.75	Median : 84.00
Mean :2018	Mean :7.138e+09	Mean : 75.11	Mean : 81.06
3rd Qu.:2021	3rd Qu.:1.088e+10	3rd Qu.: 85.24	3rd Qu.: 92.00
Max. :2023	Max. :2.361e+10	Max. :100.00	Max. :100.00
	NA's :234993	NA's :30453	NA's :30453

PCAT_EE	PCAST	PCAST_CL	PCAST_EE
Min. : 0.00	Min. : 0.00	Min. : 0.00	Min. : 0.00
1st Qu.: 56.25	1st Qu.: 69.15	1st Qu.: 74.00	1st Qu.: 62.28
Median : 69.58	Median : 78.15	Median : 84.00	Median : 71.88
Mean : 67.23	Mean : 76.00	Mean : 80.62	Mean : 70.27
3rd Qu.: 81.25	3rd Qu.: 85.40	3rd Qu.: 90.00	3rd Qu.: 84.38
Max. :100.00	Max. :100.00	Max. :100.00	Max. :100.00
NA's :30453	NA's :39157	NA's :40751	NA's :40751

PANG	PANG_CO	PANG_CL	PANG_EE
Min. : 0.00	Min. : 0.00	Min. : 0.00	Min. : 0.00
1st Qu.: 56.80	1st Qu.: 66.67	1st Qu.: 64.00	1st Qu.: 40.00
Median : 75.40	Median : 86.67	Median : 82.80	Median : 60.00
Mean : 71.09	Mean : 78.19	Mean : 75.66	Mean : 57.57
3rd Qu.: 88.00	3rd Qu.: 93.33	3rd Qu.: 92.00	3rd Qu.: 85.00
Max. :100.00	Max. :100.00	Max. :100.00	Max. :100.00
NA's :33478	NA's :33478	NA's :33478	NA's :33478

PFRAN	PFRAN_CO	PFRAN_CL	PFRAN_EE
Min. : 0.0	Min. : 0.0	Min. : 0.0	Min. : 0.0
1st Qu.: 57.1	1st Qu.: 59.2	1st Qu.: 66.7	1st Qu.: 68.0
Median : 71.4	Median : 76.6	Median : 80.0	Median : 84.0
Mean : 68.3	Mean : 71.7	Mean : 76.5	Mean : 78.1
3rd Qu.: 85.7	3rd Qu.: 87.4	3rd Qu.: 93.3	3rd Qu.: 92.0
Max. :100.0	Max. :153.3	Max. :100.5	Max. :100.0
NA's :712826	NA's :713599	NA's :713599	NA's :713599

PMAT	PMAT_EFM	PMAT_CR	PMAT_EST
Min. : 0.00	Min. : 0.00	Min. : 0.00	Min. : 0.00
1st Qu.: 50.48	1st Qu.: 41.70	1st Qu.: 50.00	1st Qu.: 51.67
Median : 68.73	Median : 63.91	Median : 68.55	Median : 75.00
Mean : 66.02	Mean : 62.19	Mean : 65.96	Mean : 69.92
3rd Qu.: 83.96	3rd Qu.: 87.31	3rd Qu.: 87.50	3rd Qu.: 88.89
Max. :100.00	Max. :100.00	Max. :100.00	Max. :100.00
NA's :24895	NA's :85427	NA's :85427	NA's :85427

PMAT_NC	PALE	PALE_CO	PALE_CL
Min. : 0.00	Min. : 25.4	Min. : 33.3	Min. : 16.0
1st Qu.: 50.00	1st Qu.: 64.0	1st Qu.: 70.0	1st Qu.: 64.0

```
## Median : 71.43      Median : 75.6      Median : 80.0      Median : 76.0
## Mean : 67.40      Mean : 74.3      Mean : 79.8      Mean : 74.4
## 3rd Qu.: 87.50      3rd Qu.: 86.0      3rd Qu.: 93.3      3rd Qu.: 88.0
## Max. :100.00      Max. :100.0      Max. :100.0      Max. :100.0
## NA's :85427      NA's :777838      NA's :777838      NA's :777838
## PALE_EE      PCIEEN      PCIEEN_COMP1      PCIEEN_COMP2
## Min. : 0.0      Min. : 0.00      Min. : 0.00      Min. : 0.00
## 1st Qu.: 50.0      1st Qu.: 53.70      1st Qu.: 50.00      1st Qu.: 50.00
## Median : 70.0      Median : 67.51      Median : 66.67      Median : 71.43
## Mean : 68.5      Mean : 65.60      Mean : 64.57      Mean : 67.84
## 3rd Qu.: 90.0      3rd Qu.: 79.22      3rd Qu.: 80.00      3rd Qu.: 85.71
## Max. :100.0      Max. :100.00      Max. :100.00      Max. :100.00
## NA's :777838      NA's :253623      NA's :253623      NA's :253623
## PCIEEN_COMP3      PCIEEN_COMP4      GERE      MES_NAIXEMENT
## Min. : 0.00      Min. : 0.0      Length:778553      Min. : 1.000
## 1st Qu.: 50.00      1st Qu.: 37.5      Class :character      1st Qu.: 4.000
## Median : 68.75      Median : 62.5      Mode :character      Median : 7.000
## Mean : 67.11      Mean : 59.5      Mean : 6.548
## 3rd Qu.: 87.50      3rd Qu.: 75.0      3rd Qu.:10.000
## Max. :100.00      Max. :100.0      Max. :12.000
## NA's :253623      NA's :655599      NA's :7805
## ANY_NAIXEMENT      NATURALES      ÀREA.TERRITORIAL      HÀBITAT
## Min. :1955      Length:778553      Length:778553      Length:778553
## 1st Qu.:1999      Class :character      Class :character      Class :character
## Median :2002      Mode :character      Mode :character      Mode :character
## Mean :2002
## 3rd Qu.:2005
## Max. :2009
## NA's :1971
```

També podem obtenir una taula resum amb les notes per tipus de centre i en funció del gènere

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

```
## # A tibble: 4 x 9
## # Groups:   TIPUS_CENTRE [2]
## TIPUS_CENTRE GERE PCAT PCAST PANG PCIEEN PFRAN PALE PMAT
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Privada D 79.8 80.5 78.9 68.4 69.6 93.9 69.4
## 2 Privada H 76.8 77.2 75.7 70.2 74.2 95.5 73.3
## 3 Pública D 75.0 76.3 68.9 62.6 64.2 74.6 60.8
## 4 Pública H 71.2 72.1 65.4 64.1 68.1 70.9 64.7
```

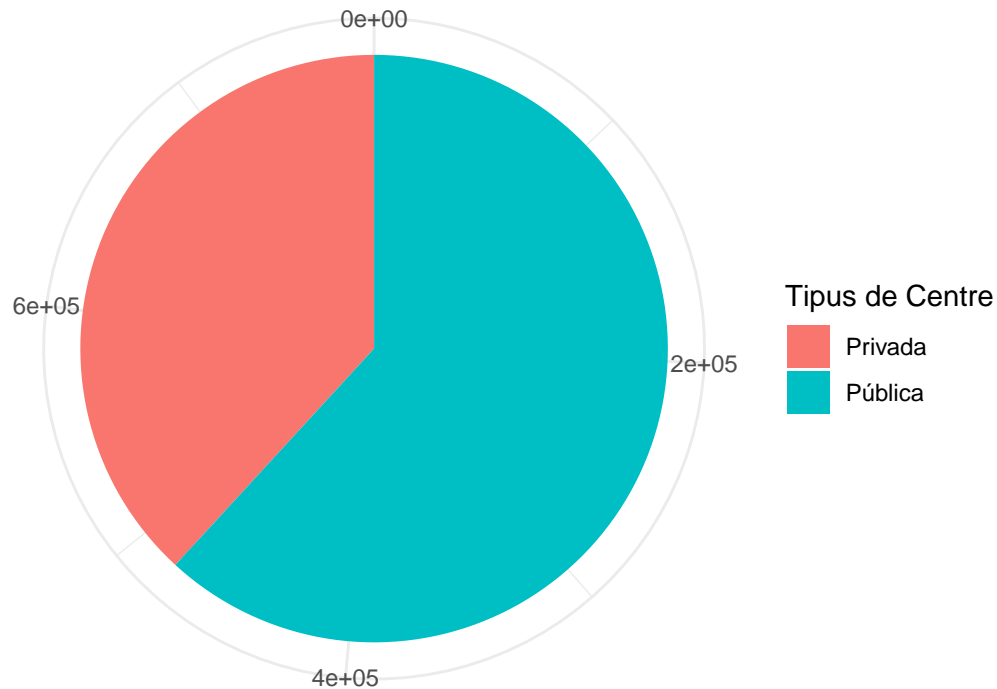
Tot i ser útil, farem una exploració més visual de les dades

## Tipus d'estudiants avaluats

El resultat de l'informe PISA son fruit de les proves realitzades als diferents centres de 4 ESO de Catalunya.

Tenim aquesta distribució per tipus de centre

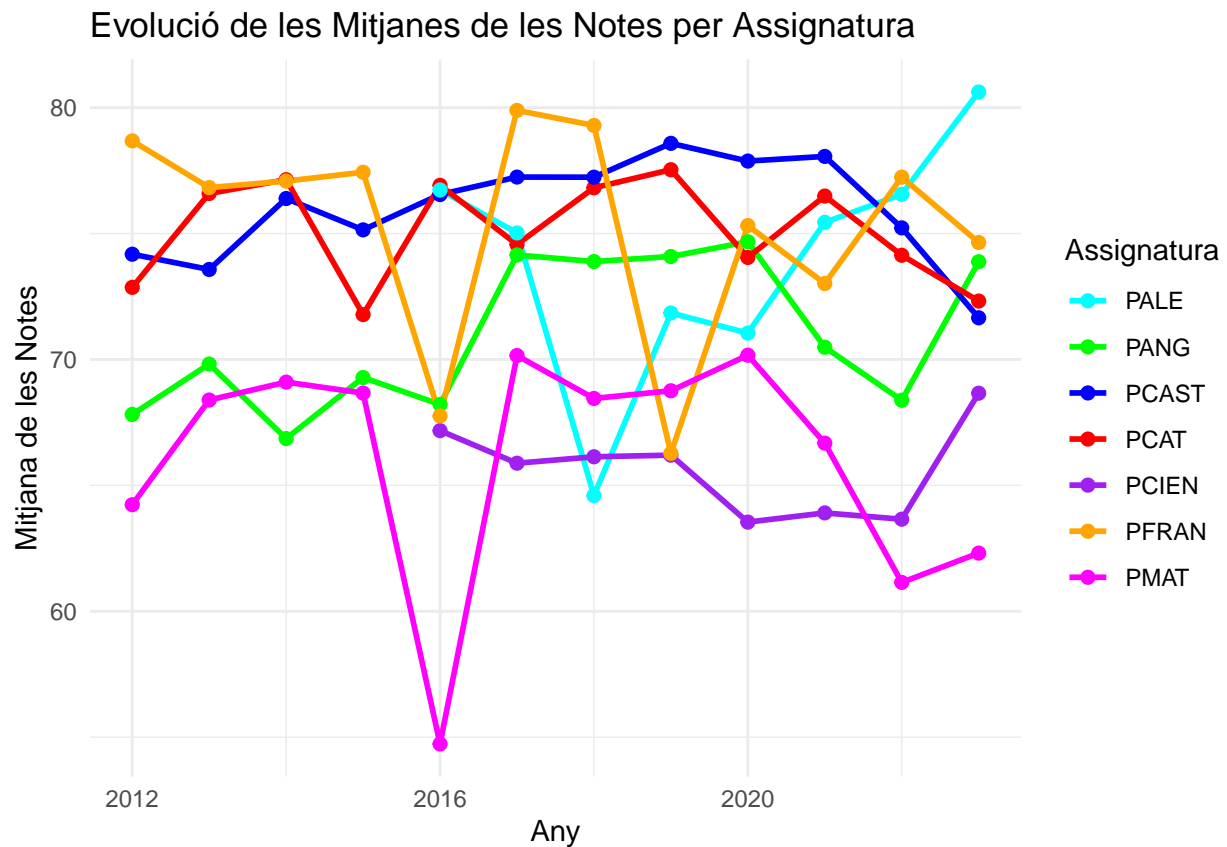
## Proporció d'Estudiants en Centres Públics i Privats



#Evolució dels resultats de les proves

A continuació podem veure l'evolució dels resultats de les diferents assignatures al llarg del temps:

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## i Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.  
  
## Warning: Removed 8 rows containing missing values or values outside the scale range  
## (`geom_line()`).  
  
## Warning: Removed 8 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```



Tot i complicar una mica la visualització, afegirem la perspectiva de gènere

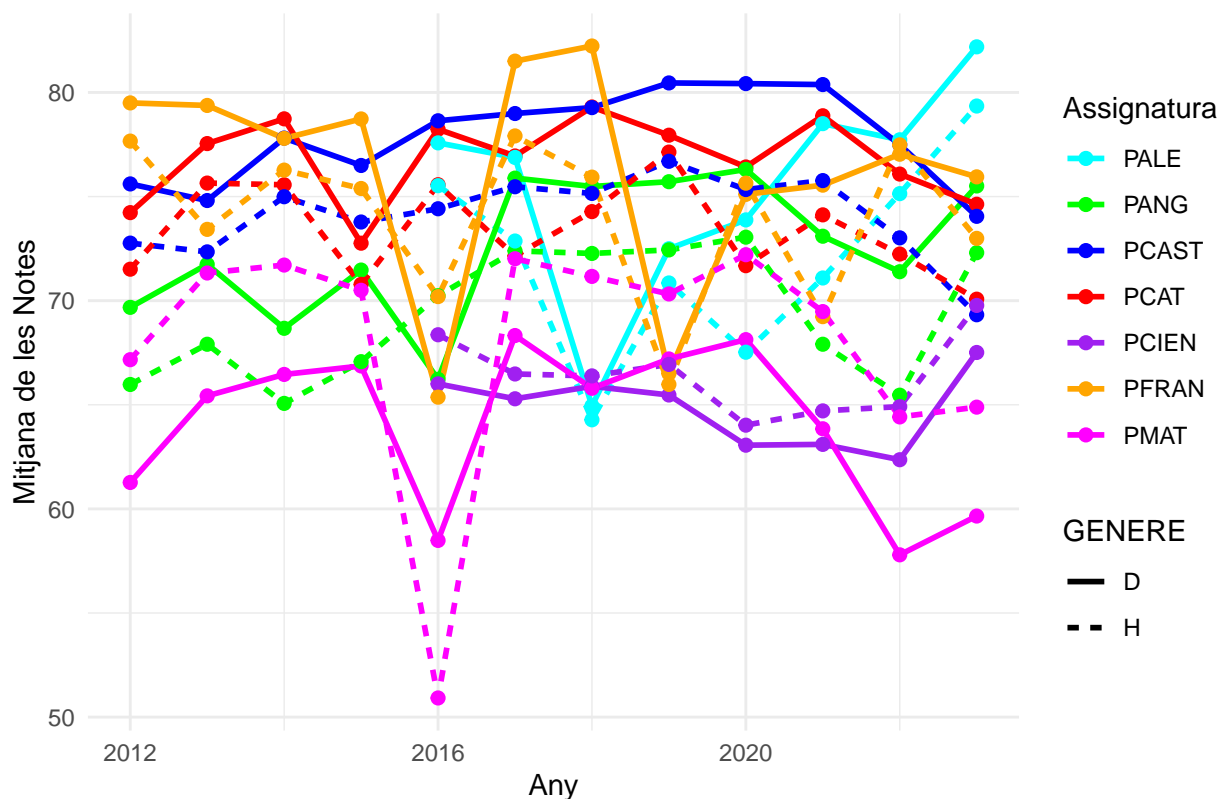
```
## `summarise()` has grouped output by 'ANY'. You can override using the `.groups`
## argument.

## Warning: Removed 16 rows containing missing values or values outside the scale range
## (`geom_line()`).

## Warning: Removed 16 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



## Evolució de les Mitjanes de les Notes per Assignatura i Gènere

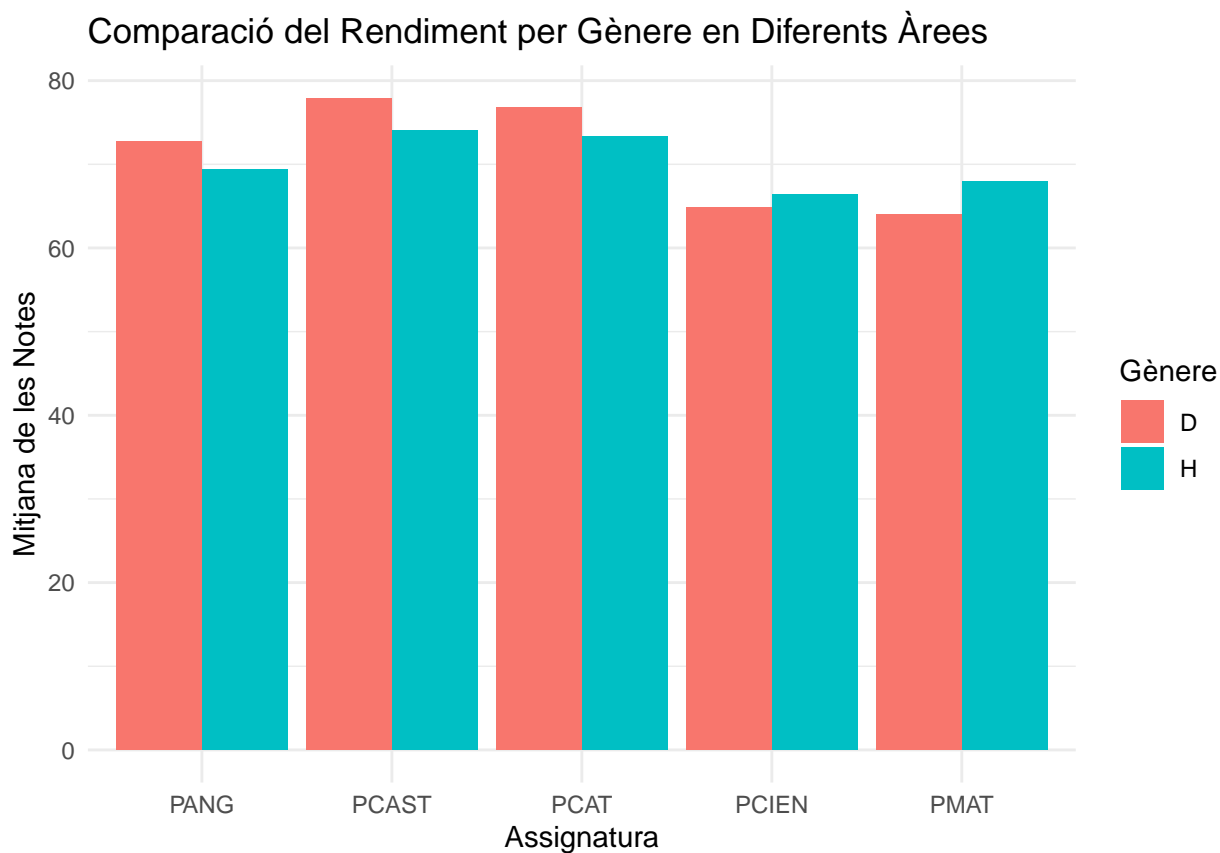


Podem observar l'evolució de les puntuacions per cada assignatura al llarg dels diferenciant el comportament per gènere. La tendència en el resultat pels dos gèneres és la mateixa, que no la competència en cada una d'elles perquè, com ja es pot veure amb més detall, hi ha una "blocs" d'assignatures que on les noies treuen millor nota que els nois, i al revès.

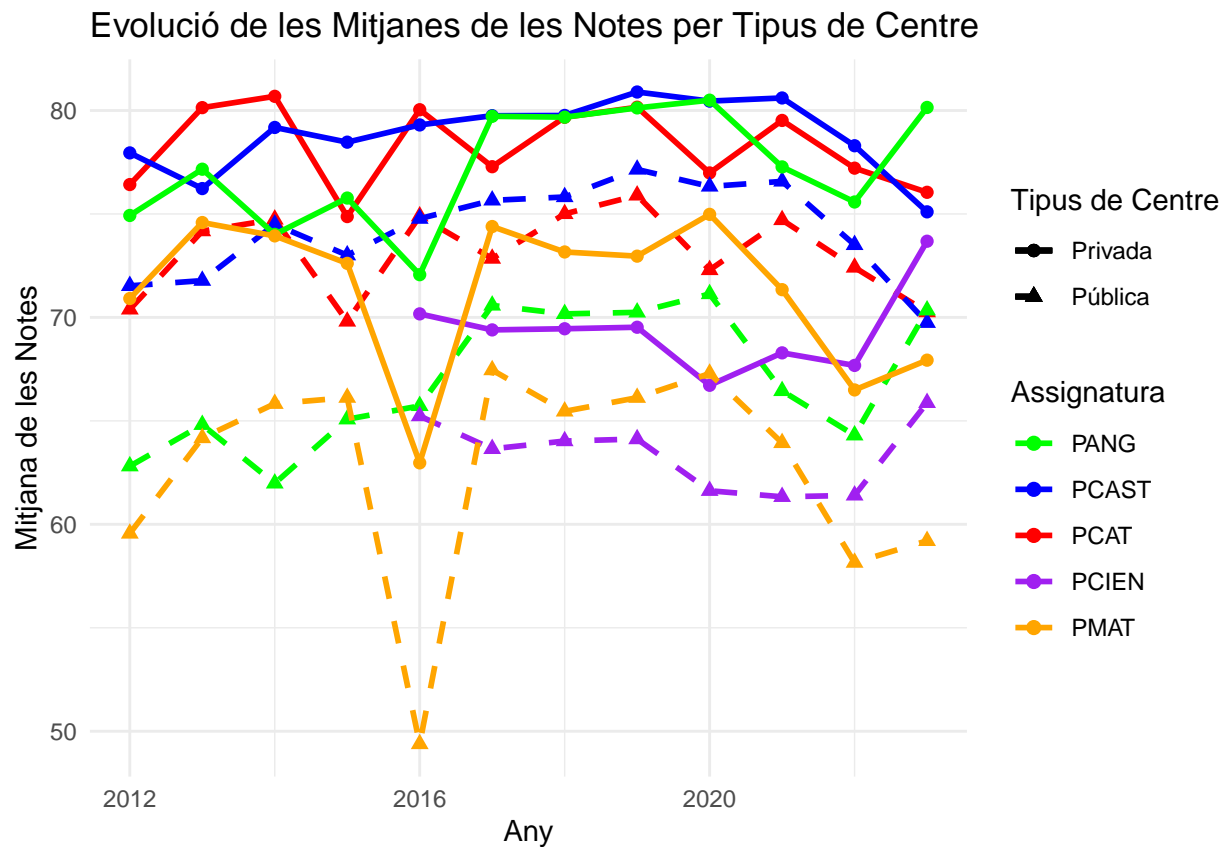
Podem destacar:

- Mínima recuperació de MATEMÀTIQUES, tot i això amb resultats bastant dolents
- Creixement continuat del resultat en ALEMANY.
- Recuperació de la tendència de millora en el coneixement d'ANGLÈS
- Forta davallada en la tendència de coneixement de CASTELLÀ i CATALÀ
- Millors resultats en assignatures d'idiomes per les noies i a les de ciències als nois

## Evolució de les qualificacions en funció del tipus de centre



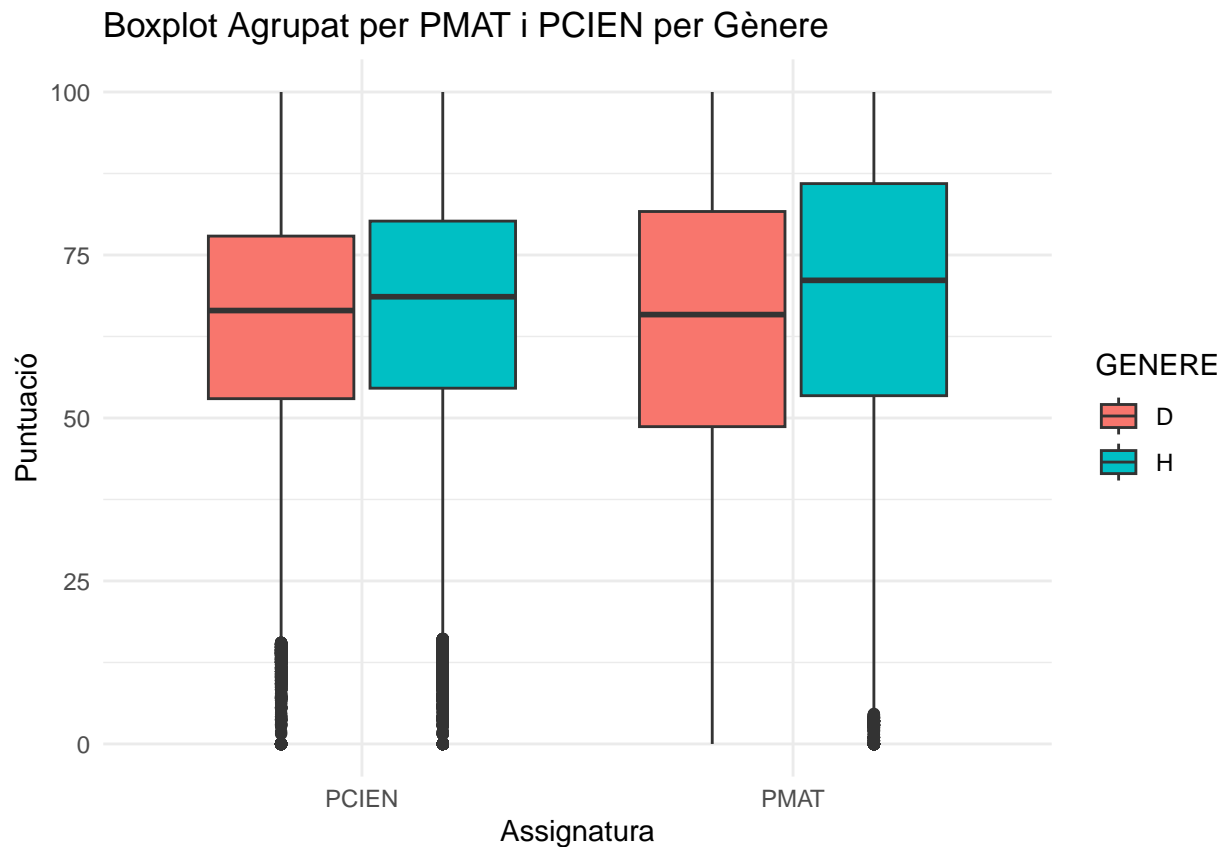
```
## `summarise()` has grouped output by 'ANY'. You can override using the `.groups`  
## argument.  
  
## Warning: Removed 8 rows containing missing values or values outside the scale range  
## (`geom_line()`).  
  
## Warning: Removed 8 rows containing missing values or values outside the scale range  
## (`geom_point()`).
```



Podem veure que l'evolució és semblant en les notes tant en els centres privats com en els centres públics, tot i que els resultats són millors en els centres privats.

### Situació de les ciències

```
## Warning: Removed 278518 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



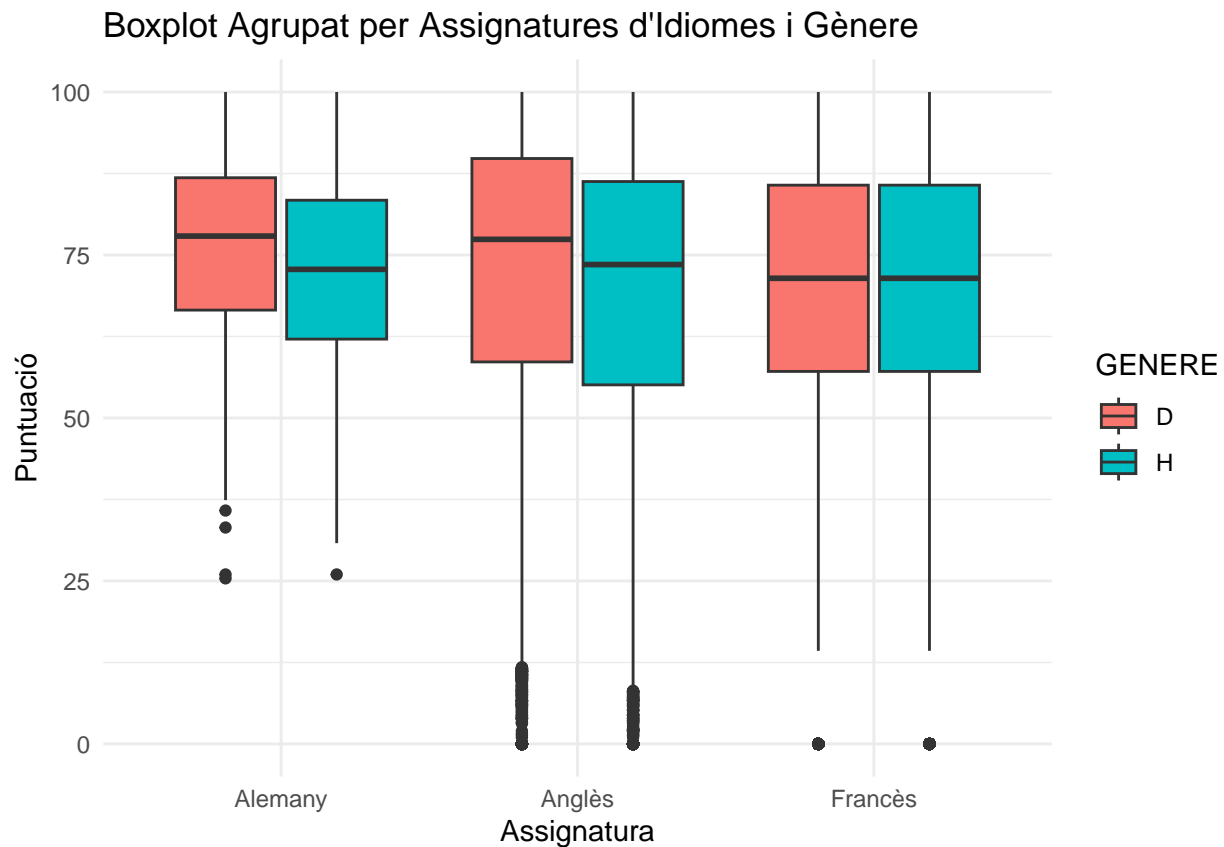
#Interpretació de la visualització: - Observem una concentració més gran dels resultats en les notes de CIÈNCIES, no observat en el cas de MATEMÀTIQUES. - Aquesta tendència s'observa en els dos gèneres - Millor resultat en el cas dels nois que de les noies

La concentració de les notes ens indica una assoliment més homogeni de coneixements en CIÈNCIES que en MATEMÀTIQUES

##Situació dels idiomes

A continuació podem veure una visualització de la concentració dels resultats dels diferents idiomes.

```
## Warning: Removed 1524142 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



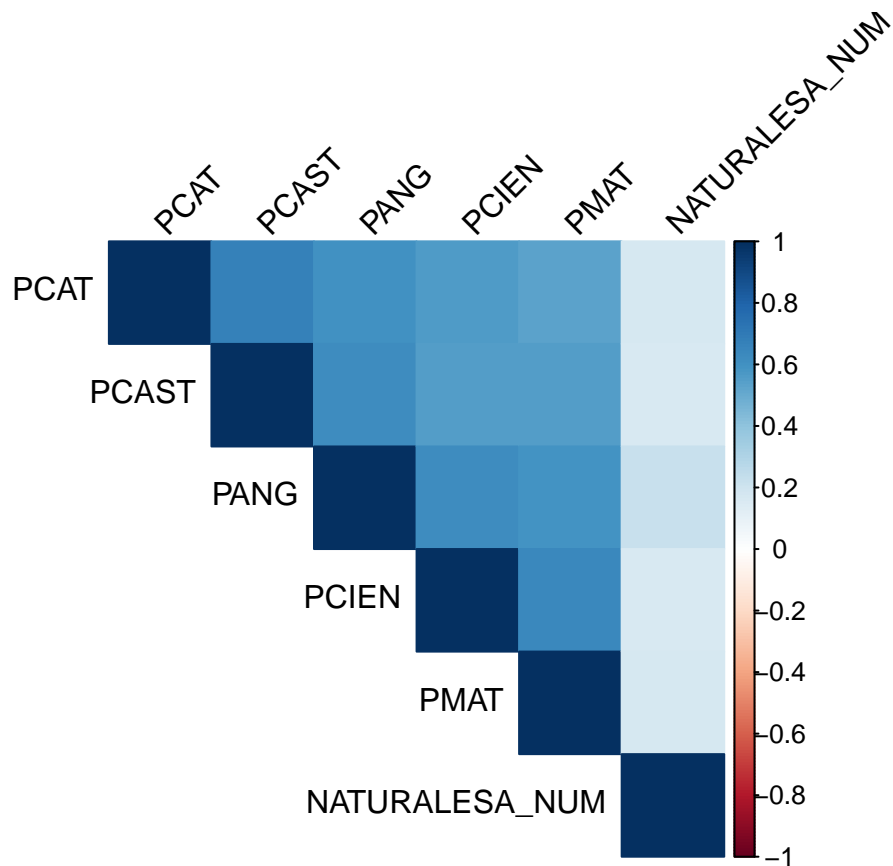
Per l'exploració visual inicial: 1. Trobem una concentració de notes més importants a Alemany. 2. Trobem també que, en general, es treuen millors notes a Alemany 3. A Francès trobem una disparitat més gran de notes 4. S'obtenen millors notes en general a Anglès que d'altres assignatures

##Relació entre assignatures

En aquest apartat volem conèixer si existeix alguna relació entre els resultats obtinguts en una matèria influeixen en un altre.

Per fer-ho, faren una matriu de correlació.

##	PCAT	PCAST	PANG	PCIEN	PMAT	NATURALES_NUM
## PCAT	1.0000000	0.6742487	0.6082529	0.5676034	0.5357889	0.1704597
## PCAST	0.6742487	1.0000000	0.6206784	0.5572713	0.5508126	0.1605624
## PANG	0.6082529	0.6206784	1.0000000	0.6281150	0.5993508	0.2294559
## PCIEN	0.5676034	0.5572713	0.6281150	1.0000000	0.6410284	0.1640471
## PMAT	0.5357889	0.5508126	0.5993508	0.6410284	1.0000000	0.1794525
## NATURALES_NUM	0.1704597	0.1605624	0.2294559	0.1640471	0.1794525	1.0000000



Podem observar que:

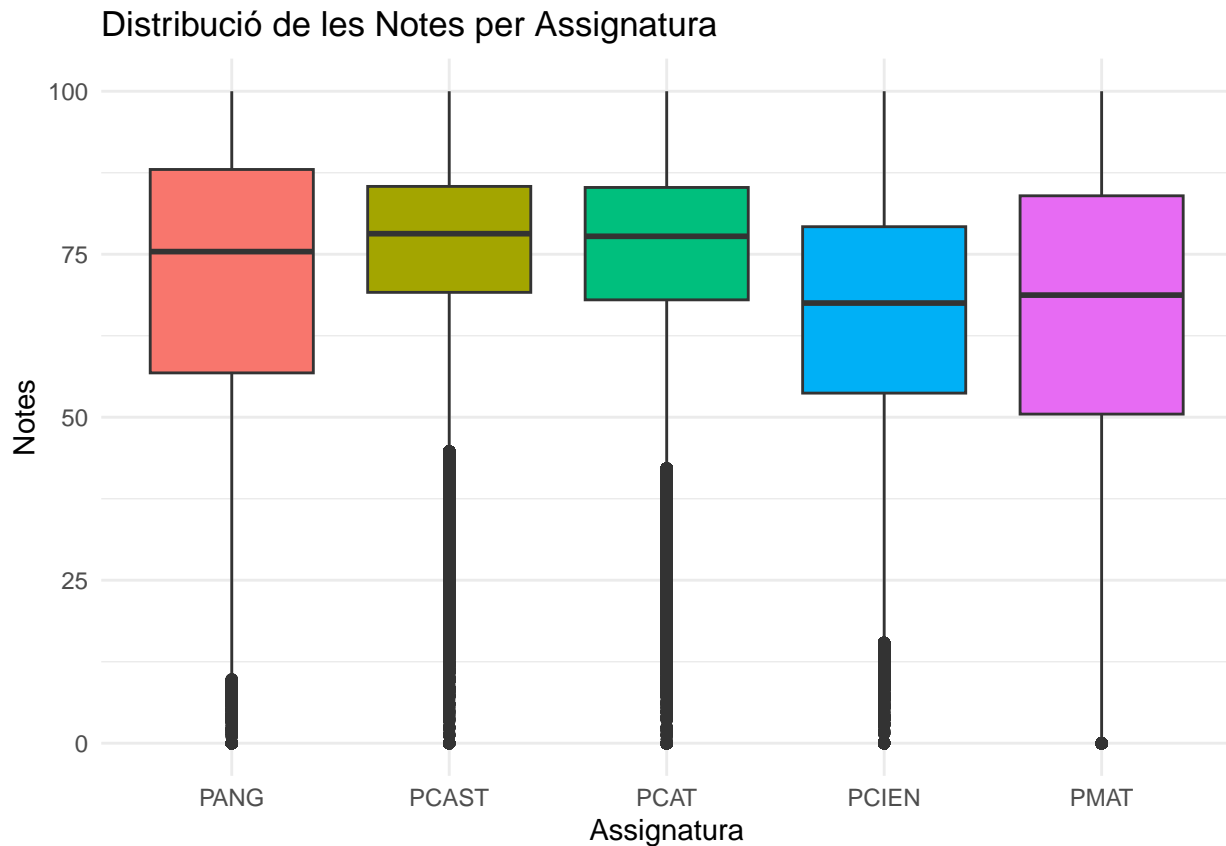
1. **PCAT:**
  - Correlació forta amb **PCAST** (0.674).
  - Correlació moderada amb **PANG** (0.608).
  - Correlació moderada amb **PCIEN** (0.568).
  - Correlació moderada amb **PMAT** (0.536).
2. **PCAST:**
  - Correlació forta amb **PCAT** (0.674).
  - Correlació forta amb **PANG** (0.621).
  - Correlació moderada amb **PCIEN** (0.557).
  - Correlació moderada amb **PMAT** (0.551).
3. **PANG:**
  - Correlació moderada amb **PCAT** (0.608).
  - Correlació forta amb **PCAST** (0.621).
  - Correlació forta amb **PCIEN** (0.628).
  - Correlació moderada amb **PMAT** (0.599).
4. **PCIEN:**
  - Correlació moderada amb **PCAT** (0.568).
  - Correlació moderada amb **PCAST** (0.557).
  - Correlació forta amb **PANG** (0.628).
  - Correlació forta amb **PMAT** (0.641).
5. **PMAT:**
  - Correlació moderada amb **PCAT** (0.536).
  - Correlació moderada amb **PCAST** (0.551).
  - Correlació moderada amb **PANG** (0.599).
  - Correlació forta amb **PCIEN** (0.641).

Podem observar que hi ha correlacions fortes i moderades entre les diferents assignatures. Aquestes correlacions suggereixen que les notes en una assignatura poden influir en les notes d'altres assignatures. Per exemple, les notes en ciències (**PCIEN**) tenen una correlació forta amb les notes en matemàtiques (**PMAT**), el que indica que els alumnes que obtenen bones notes en ciències també tendeixen a obtenir bones notes en matemàtiques. Les correlacions dèbils entre les notes de les assignatures i el tipus de centre suggereixen que el tipus de centre (Privada o Pública) té una influència menor en les notes dels alumnes en les assignatures analitzades.

Tot i això, les notes mitjanes de les assignatures en el centres privats son més grans.

## Distribució de les qualificacions

```
## Warning: Removed 381606 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

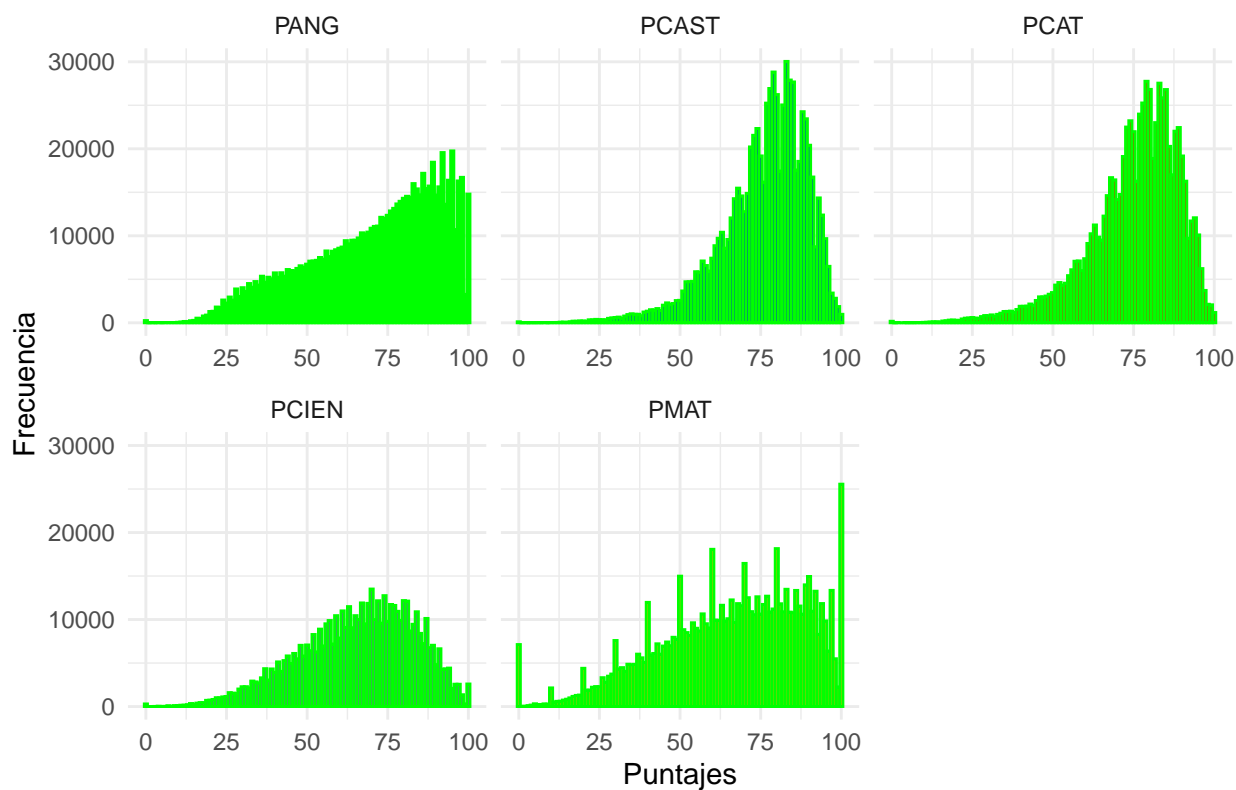


#Distribució de les notes

En aquest apartat volem examinar la distriució de les notes per les diferents assignatures.

```
## Warning: Removed 381606 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

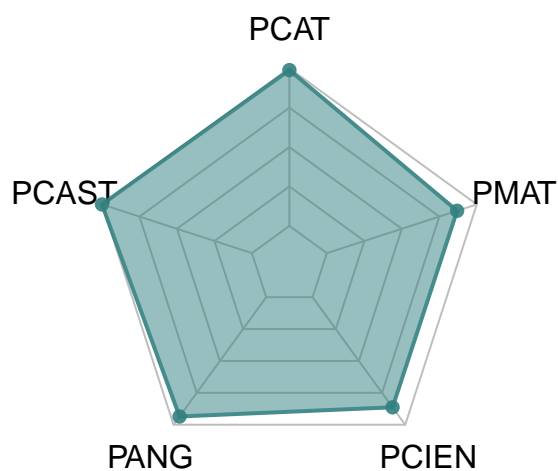
## Distribución de los Puntajes por Asignatura



Com a comentaris:

- Mateix tipus de distribució i concentració en PCAT i PCAST, a on és més fàcil arribar al notable

## Gràfic de Radar del Rendiment en Assignatures

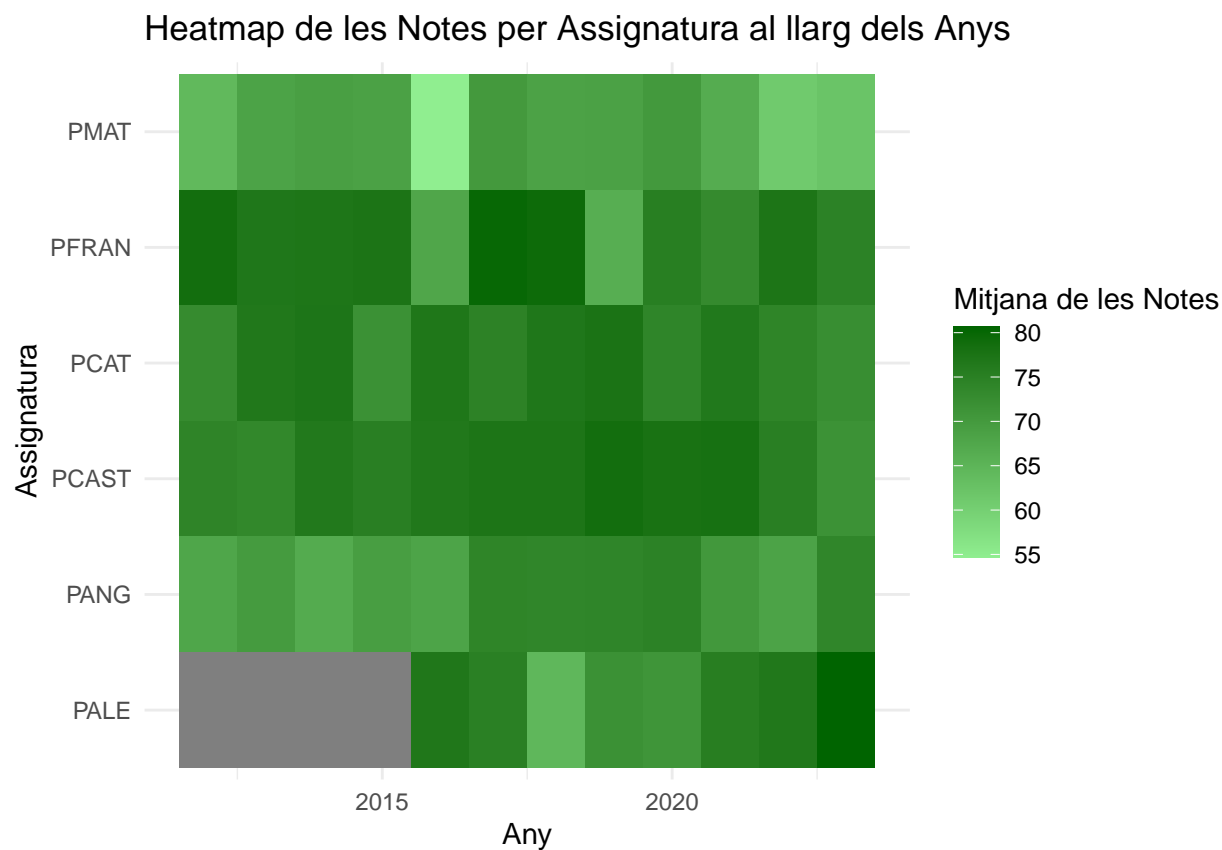
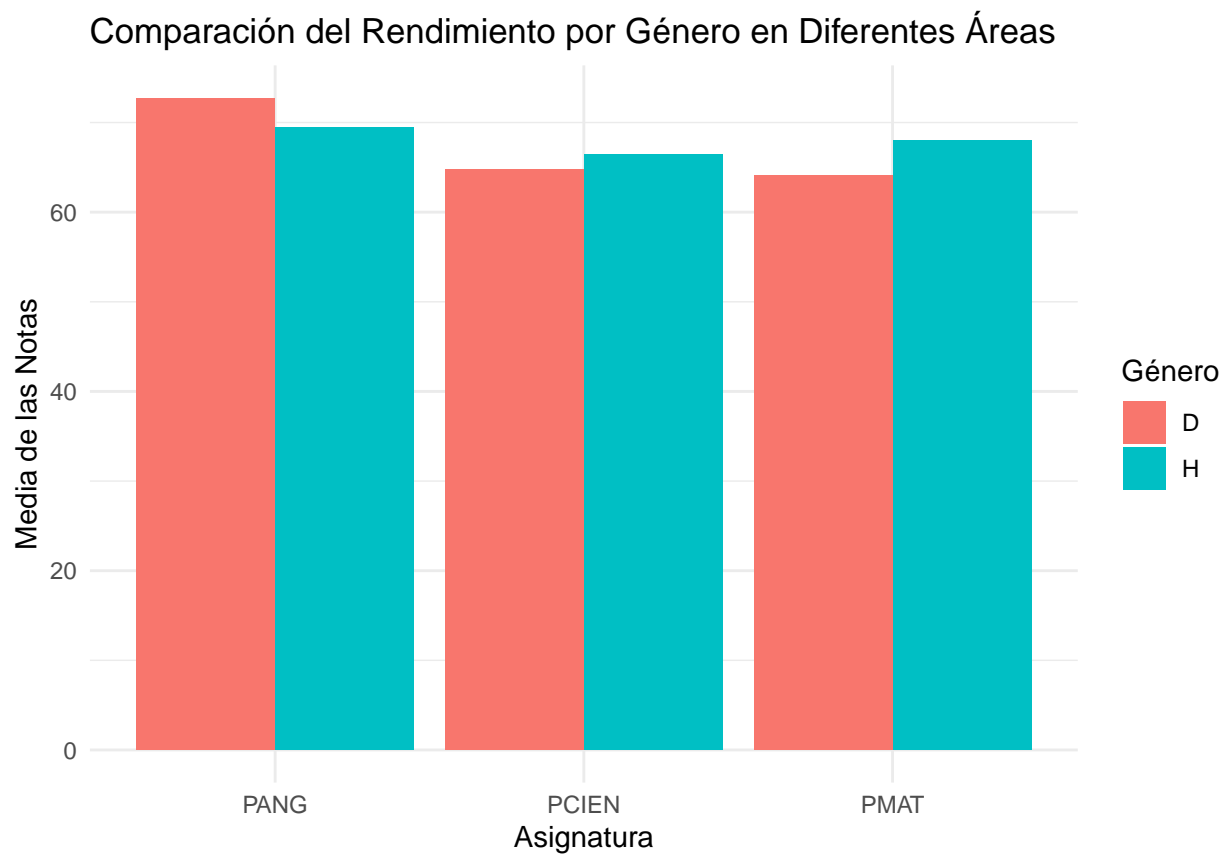


por Género en Diferentes Áreas

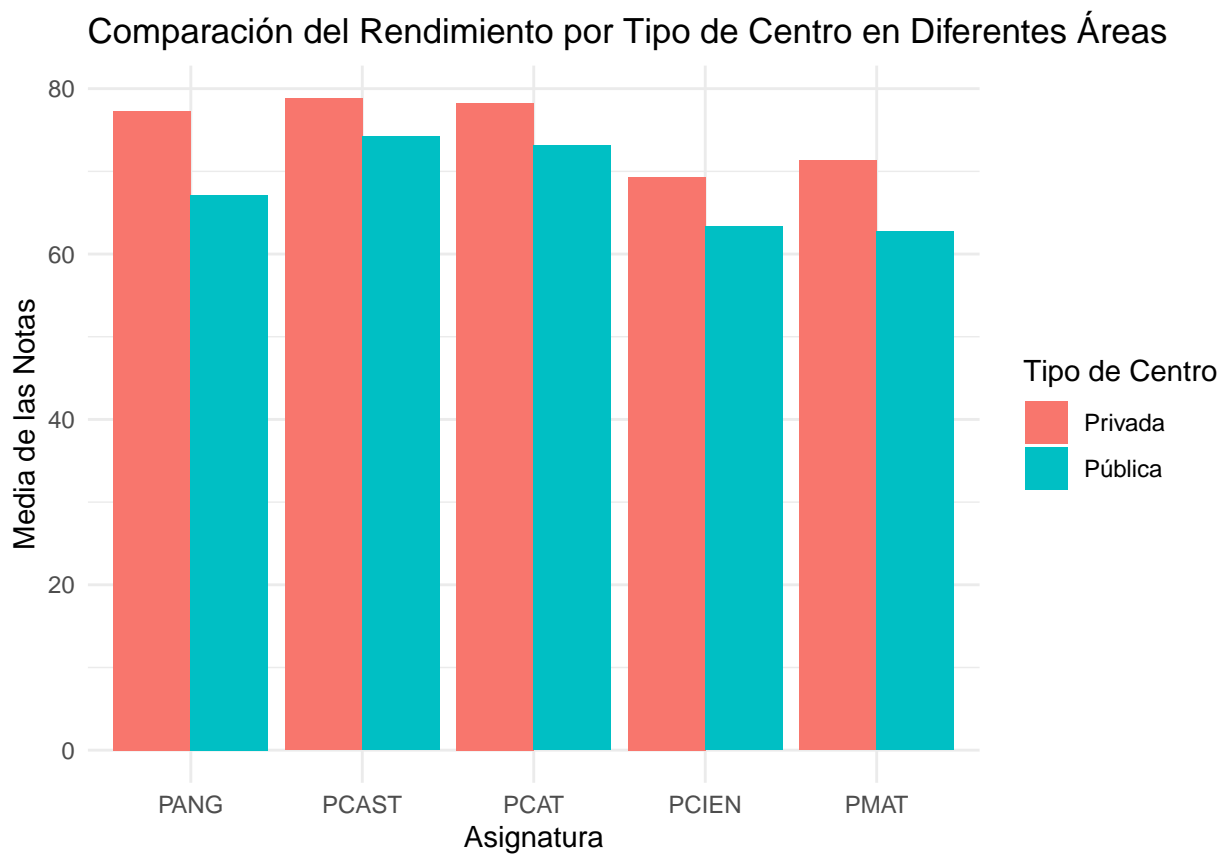
#Comparación del Rendimiento

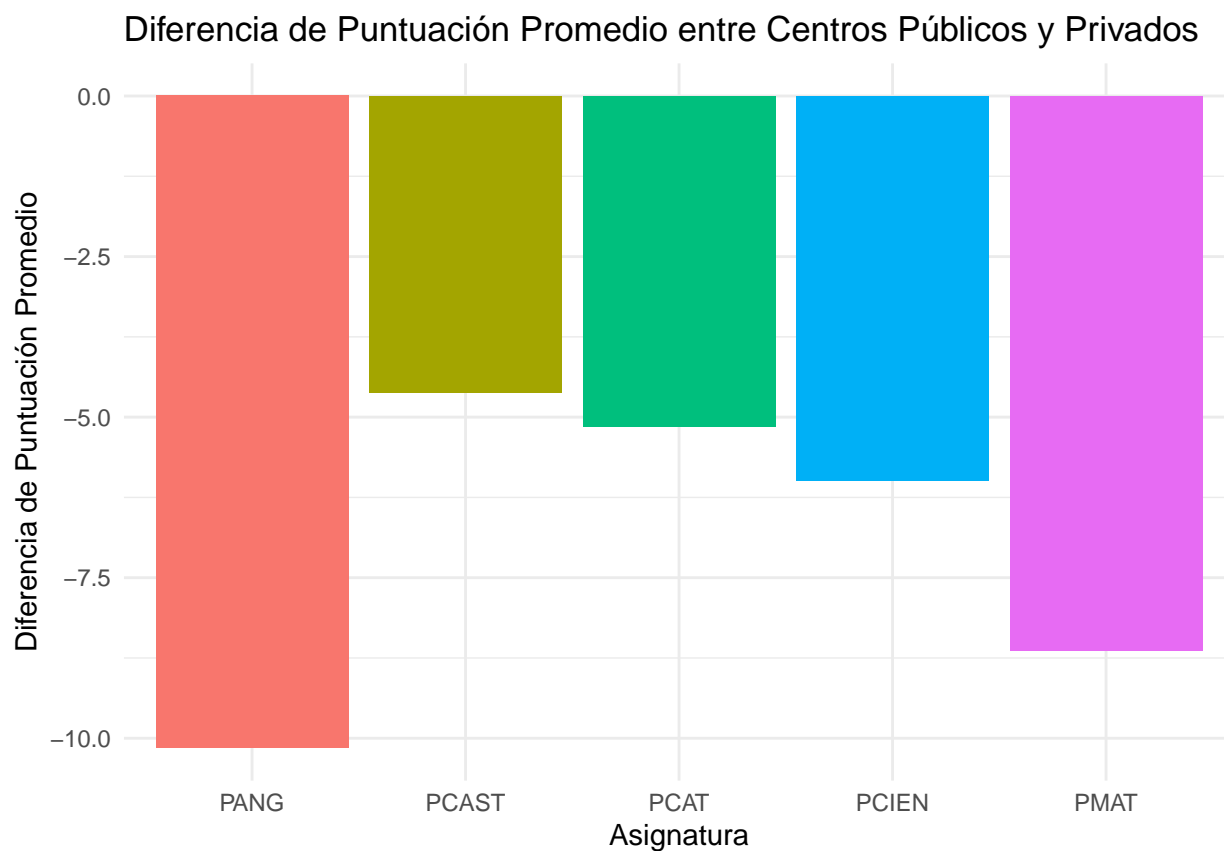
Podem veure amb la següent gràfica el rendiment per gènere en les diferents assignatures





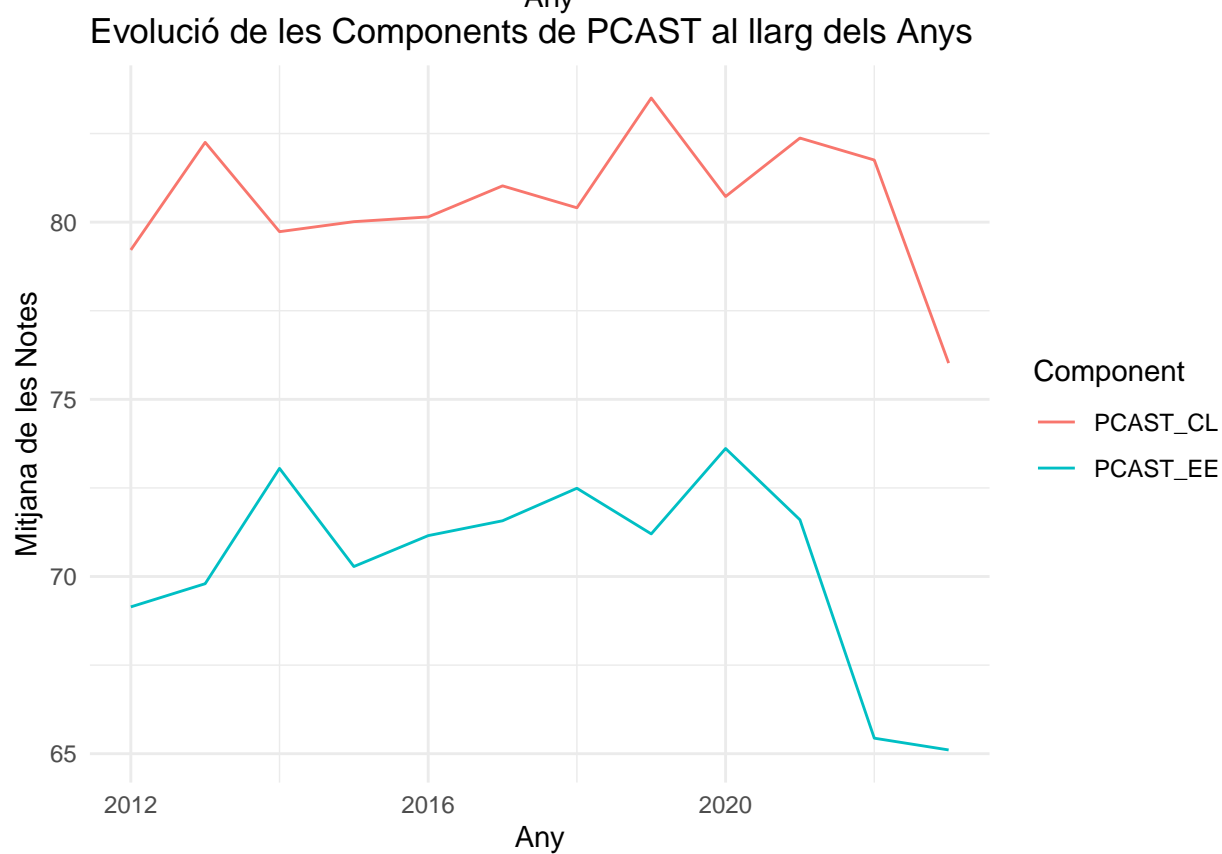
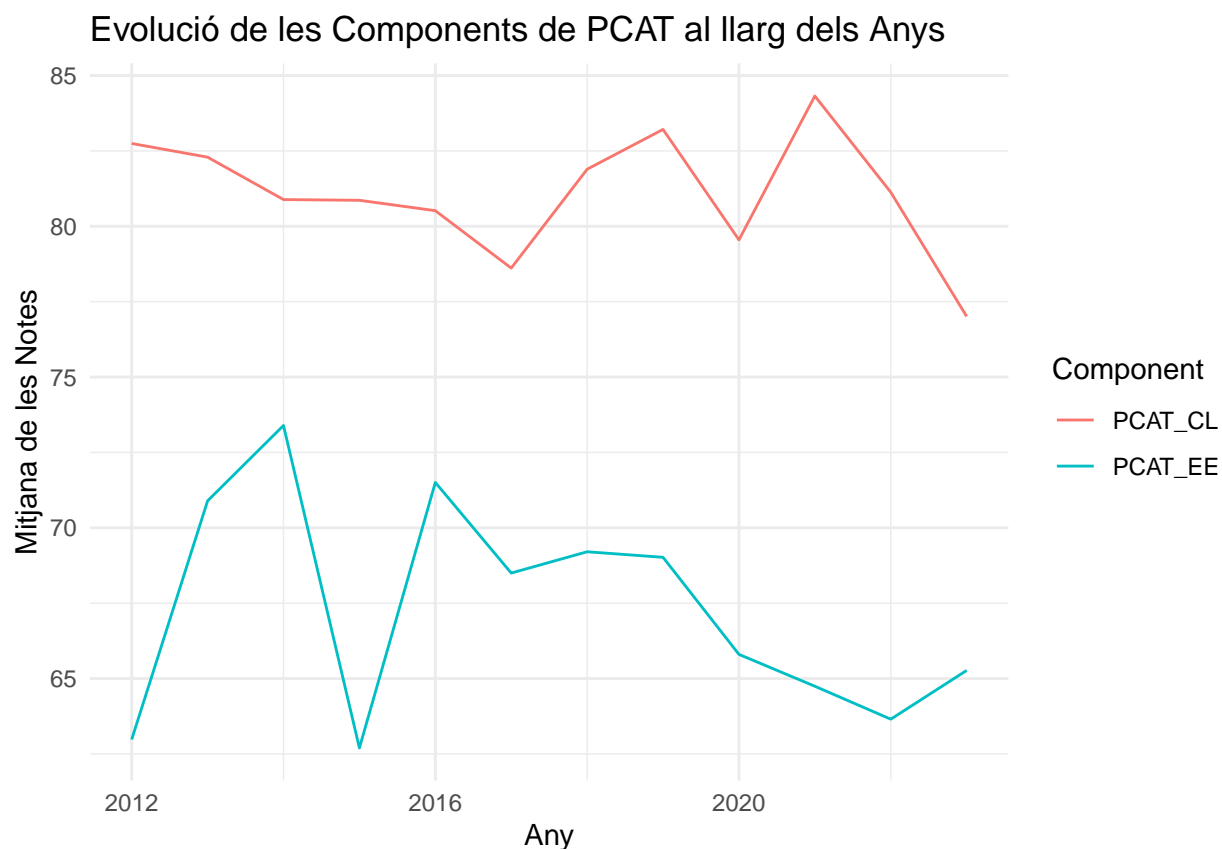
Aquest mapa de calor ens indica clarament l'evolució que han tingut les diferents puntuacions al llarg del temps. Podem veure la davallada important que va haver-hi en matemàtiques l'any 2016.



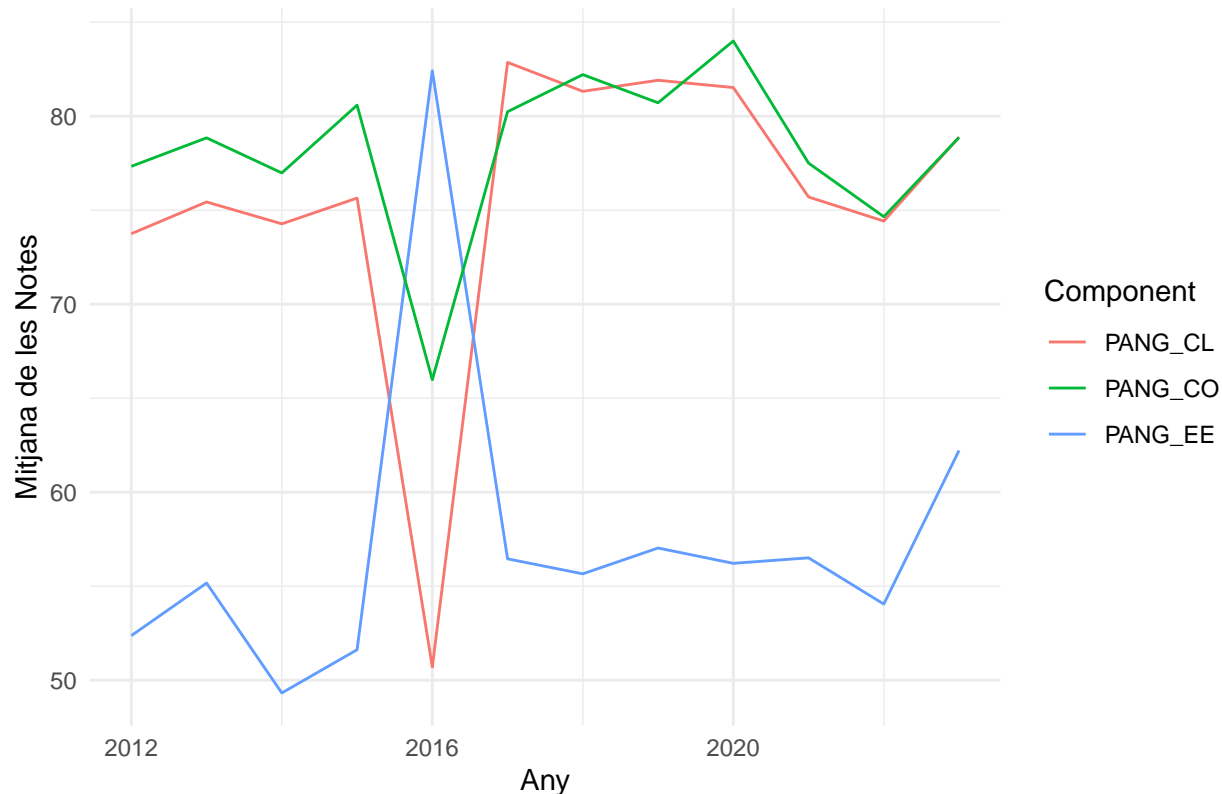


#Avaluació de cada component de les notes

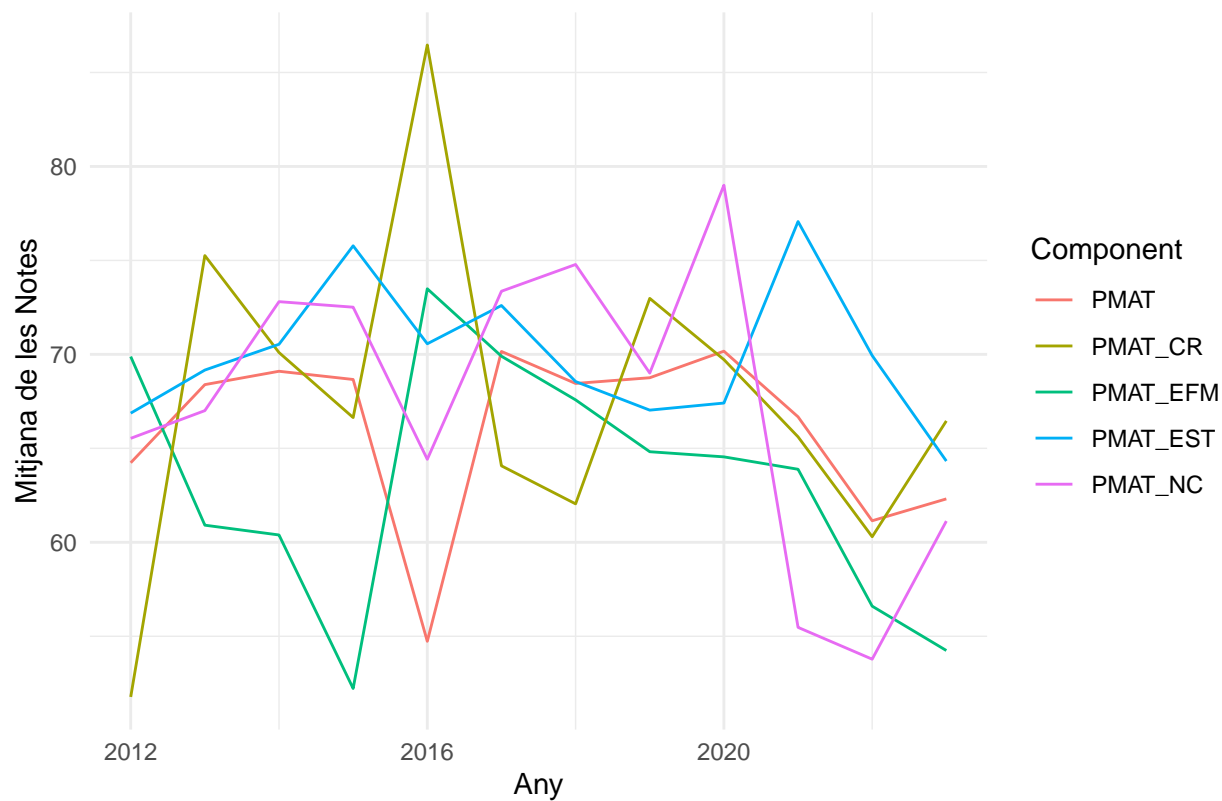
Cada nota està formada per diferents components. Podrem trobar una explicació en el resultat global de l'assignatura i avaluem cada component per separat.



### Evolució de les Components de PANG al llarg dels Anys



### Evolució de les Components de PMAT al llarg dels Anys



Per exemple, es pot veure que en el cas d'ANGLÈS, la component que te pitjor comportament és PANG\_EE

(expressió escrita) i, per tant, s'hauria de posar més recursos en treballar aquesta component

## Conclusió

1. Millor rendiment en centres privats que en públics
2. Mateixa tendència en evolució dels resultats tant en públics com en privats
3. Continua havent un millor rendiment en general en assignatures de “lletres” per a les noies, i de “ciències” per als nois

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

```
## # A tibble: 4 x 9
## # Groups:   TIPUS_CENTRE [2]
##   TIPUS_CENTRE GENERE PCAT PCAST PANG PCIEN PFRAN PALE PMAT
##   <chr>         <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Privada      D      79.8  80.5  78.9  68.4  69.6  93.9  69.4
## 2 Privada      H      76.8  77.2  75.7  70.2  74.2  95.5  73.3
## 3 Pública     D      75.0  76.3  68.9  62.6  64.2  74.6  60.8
## 4 Pública     H      71.2  72.1  65.4  64.1  68.1  70.9  64.7
```