

Mirem el món amb els ulls de l'Estadística

Pedro Delicado

Departament d'Estadística i Investigació Operativa
Universitat Politècnica de Catalunya

Cicle de conferències
L'Estadística és present en el nostre món

Càtedra Lluís Santaló, Fundació Casa de Cultura de Girona

Girona, 21 d'octubre de 2010

Introducció i esquema de la xerrada

- ▶ Una mateixa realitat pot ser interpretada de diverses maneres per diferents observadors.
- ▶ L'Estadística pot ajudar a valorar objectivament aspectes d'aquesta realitat de forma que les diferents interpretacions es vegin reforçades o refutades.
- ▶ Fixarem la nostra atenció en tres exemples del món social que ens envolta.

Dades electorals

Sondeigs electorals

Abstenció i orientació del vot

Piràmides de població dels països del món

Gapminder. Un observatori del desenvolupament humà

Tendències del desenvolupament humà 2005

Altres obres de Gapminder

Dades electorals

Sondeigs electorals

Abstenció i orientació del vot

Piràmides de població dels països del món

Gapminder. Un observatori del desenvolupament humà

Tendències del desenvolupament humà 2005

Altres obres de Gapminder

La mala fama dels sondeigs electorals, és merescuda?

(Treball conjunt amb Frederic Udina, UPF)

- ▶ Les setmanes prèvies a unes eleccions apareixen multitud de sondeigs electorals als mitjans de comunicació.
- ▶ Els dies posteriors el que sovint hi apareix és la discussió de per què els sondeigs es van equivocar.
- ▶ Analitzem els sondeigs de **6 eleccions** celebrades al llarg de **10 anys**:
 - ▶ Parlament'99, Parlament'03, Parlament'06,
 - ▶ Congrés'00, Congrés'04, Congrés'08.

Eleccions al Parlament '99. Resultats i sondeigs

	CiU	PSC	PP	IC-V	ERC
Resultat	56	52	12	3	12
CIS	57..59	45..47	14..15	4..5	11..13
El Mundo	58..62	48..50	12..14	3	10
La Vanguardia	56..58	51..52	13	4	9..10
El País	58..60	48..50	13..14	3	10
El Periódico	57..63	40..46	14..15	5..6	12..15
ABC	53..55	45..47	15..16	7..8	12..13
Diario 16	53..56	48..50	14..16	6..7	10..12

- ▶ Aquesta va ser la primera vegada que en Pasqual Maragall es va presentar com a candidat a President de la Generalitat.
- ▶ El PSC va guanyar en vots però no pas en escons.
- ▶ Els sondeigs van predir millors resultats per CiU i pitjors pel PSC.

Eleccions al Parlament '03. Resultats i sondeigs

	CiU	PSC	ERC	PPC	ICV
Resultats	46	42	23	15	9
ABC	49..49	48..49	18..19	13..13	6..6
CIS	48..48	47..47	18..18	14..14	8..8
El Periódico	45..47	50..51	19..20	12..13	7..8
La Vanguardia	48..49	49..50	18..19	12..12	6..7
El País	47..49	48..50	17..19	13..13	7..7
La Razón	48..49	48..51	17..18	13..15	5..6
El Mundo	46..49	46..49	18..20	15..16	5..6

- ▶ Primera victòria del tripartit.
- ▶ Tots els sondeigs van predir molt millors resultats pel PSC, millors resultats per CiU i pitjors pels tres partits petits.

- ▶ Donats aquests resultats tan poc satisfactoris, ens vam preguntar:

Les diferències entre les prediccions dels sondeigs i els resultats reals, es poden atribuir exclusivament a l'atzar?

- ▶ Comparem els parlaments pronosticats pels sondeigs amb els que s'obtindrien amb un mostreig aleatori ideal.
- ▶ Com comparar **sondeigs reals** i **sondeigs simulats** per ordinador?.
- ▶ Busquem una forma **gràfica** de comparar-los.

Si només hi hagués dos partits importants ... (I)

Exemple: PP i PSOE (i altres).

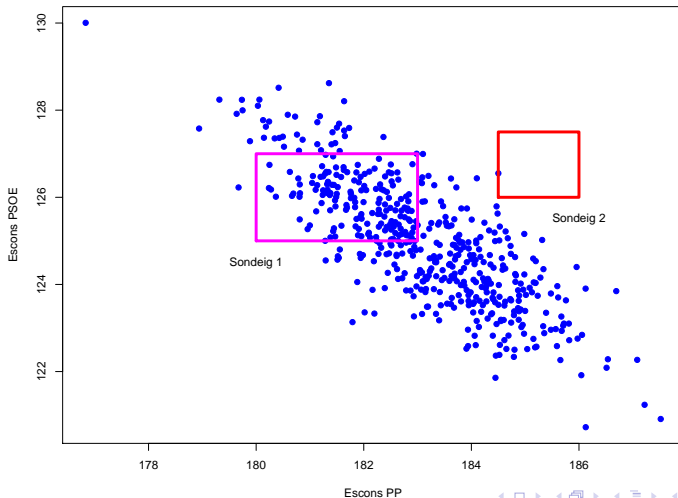
- ▶ Després de les eleccions, tenim els resultats reals.
- ▶ Simulem una mostra d'intenció de vot a cada circumscripció i calculem la corresponent assignació d'escons.
- ▶ La suma dels escons obtinguts per circumscripcions configura un **parlament simulat**.
- ▶ Resultat del sondeig simulat: (PP,PSOE), un punt al pla coordinat.
- ▶ Repetim $B=500$ vegades. Tindrem 500 punts al pla.
- ▶ Representem en uns eixos de coordenades els 500 punts:

Eix X: PP, Eix Y: PSOE

- ▶ També representem els sondeigs electorals publicats.
- ▶ S'assemblen els sondeigs reals als produïts per ordinador?

Si només hi hagués dos partits importants ... (II)

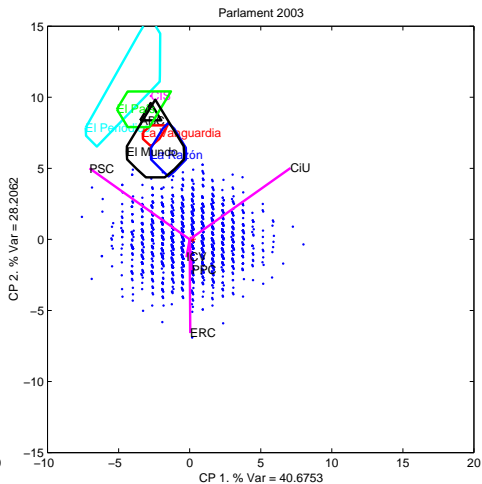
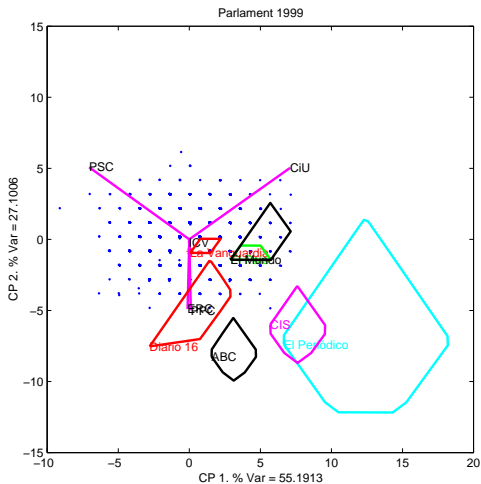
Sondeigs simulats i dos sondeigs ficticis



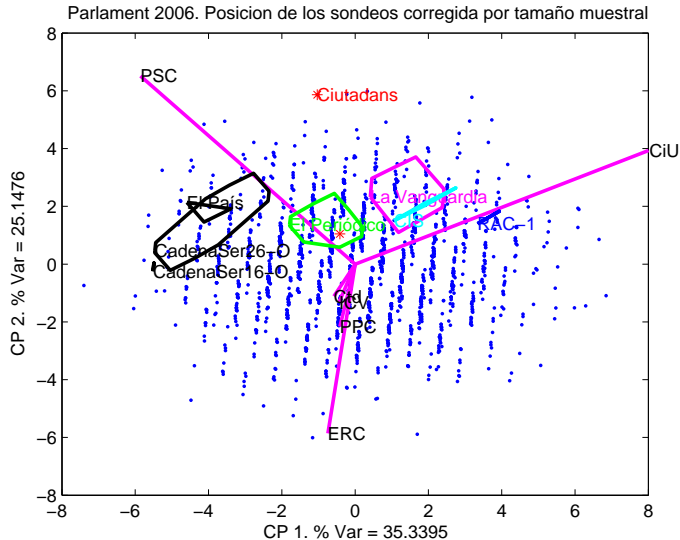
Però hi ha més de dos partits: La millor foto dels sondeigs.

- ▶ Hi ha molts partits polítics:
5 o 6 al Parlament, més de 10 al Congrés.
- ▶ Cada sondeig simulat és un vector de **dimensió 10** (o més):
(PP, PSOE, CiU, IU, PNV,..., CHA).
- ▶ Com representar els sondeigs en un paper o una pantalla?
- ▶ La millor foto dels sondeigs: **COMPONENTS PRINCIPALS**.

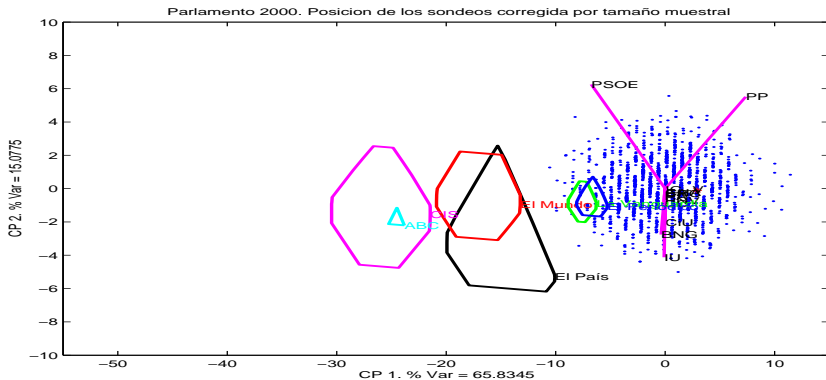
Parlament '99 i Parlament '03



Parlament '06

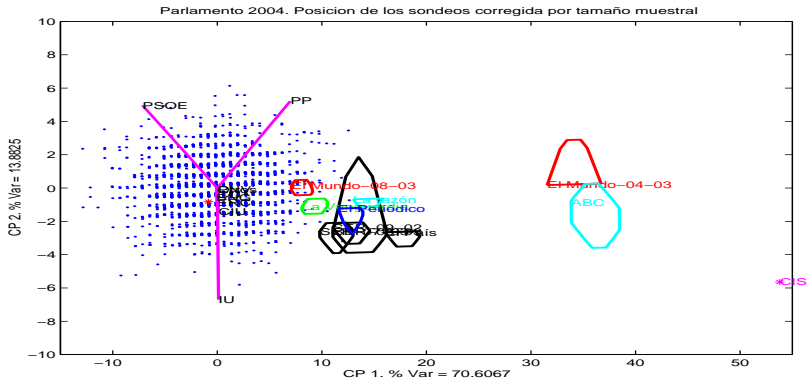


Congrés, 2000



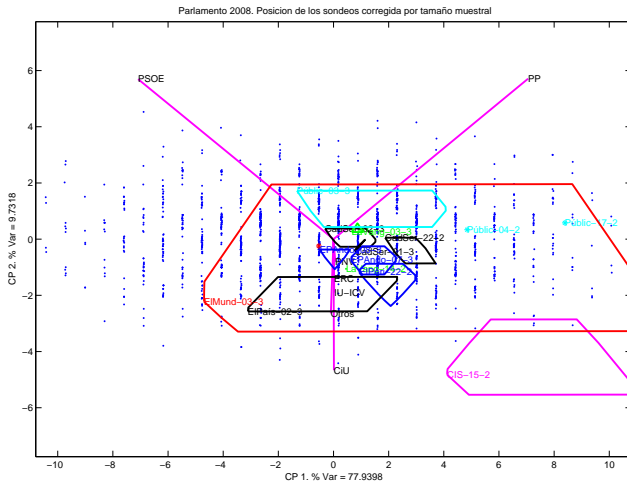
- ▶ El PP va guanyar per majoria absoluta.
- ▶ Cap sondeig va predir aquesta majoria absoluta.

Congrés, 2004



- ▶ Les eleccions van ser 3 dies després dels atemptants de l'11-M.
- ▶ Primera victòria de Zapatero al front del PSOE.
- ▶ Cap sondeig va predir la victòria del PSOE.

Congrés 2008: Els sondeigs ho van fer bastant bé.



Abstenció i orientació del vot

- ▶ És acceptat com a veritat per la majoria d'analistes que ...
 - ▶ a Espanya l'abstenció acostuma a perjudicar als partits d'esquerra,
 - ▶ l'electorat de dretes és més fidel i participatiu que el d'esquerres.
- ▶ Però alguns pensen que l'efecte de l'abstenció ...
 - ▶ és erràtic,
 - ▶ depèn de l'ocasió i el context polític.
- ▶ **Objectiu:** Quantificar la relació entre tendència política (dreta-esquerra) i abstenció.
- ▶ Analitzem els resultats de les Eleccions Generals al Congrés dels Diputats del 14 de març de 2004.
- ▶ Hi va haver una alta participació.
- ▶ Es va interpretar que va ser deguda a la mobilització de l'electorat d'esquerres, i això hauria ajudat a la victòria del PSOE.

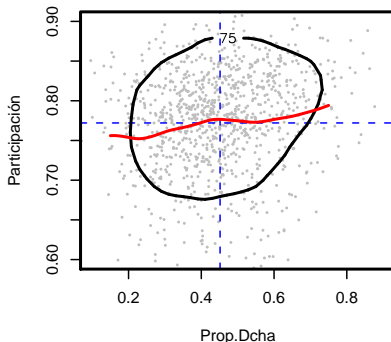
Eleccions Generals 2004. Dades disponibles

- ▶ Individus: les 56.585 meses electorals d'arreu de l'estat.
- ▶ Informació de cada mesa:
 - ▶ cens electoral,
 - ▶ nombre de vots emesos,
 - ▶ repartiment dels vots entre els diferents partits polítics,
 - ▶ vots en blanc i nuls.
- ▶ Construïm **dues noves variables**:
 - ▶ participació,
 - ▶ proporció de vots a partits de dreta,
- ▶ Si l'abstenció fos més freqüent entre l'electorat potencial d'esquerres, s'esperaria que ...
 - ▶ meses amb **alta prop. dreta** tingueren **alta participació**.
 - ▶ meses amb **baixa prop. dreta** tingueren **baixa participació**.

Anàlisi (I)

- ▶ **Distribució conjunta de prop. dreta i participació.**
 - ▶ Coeficient de correlació lineal
 - ▶ Estimació no-paramètrica de la densitat conjunta.
 - ▶ Estimació no-param. de la regressió de **participació** sobre **prop. dreta**.

España (Corr= 0.13)

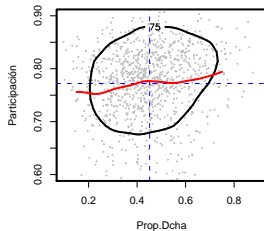


- ▶ Coeficient de correlació baix (significativament diferent de 0).
- ▶ Sembla que hi ha poca relació entre l'abstenció i l'orientació política.

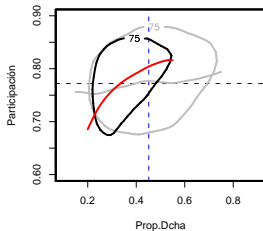
Anàlisi (II)

- ▶ Però, què passa si ens limitem a estudiar meses electorals del mateix municipi?
- ▶ Reduïm heterogeneïtat i potser podrem veure quelcom diferent.
- ▶ Només estudiem els 5 municipis amb nombre de meses electorals més gran.

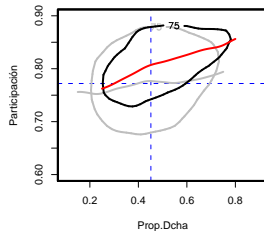
Espanya (Corr= 0.13)



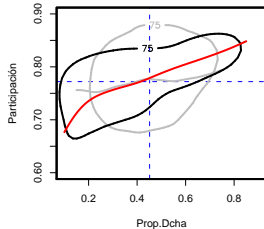
BARCELONA (Corr= 0.46)



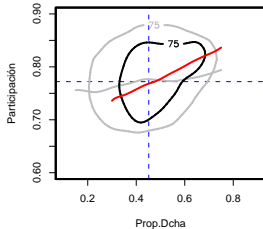
MADRID (Corr= 0.46)



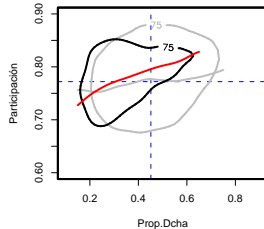
SEVILLA (Corr= 0.62)



VALENCIA (Corr= 0.47)



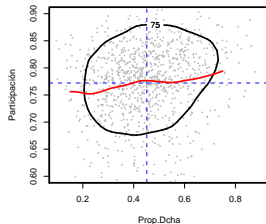
ZARAGOZA (Corr= 0.42)



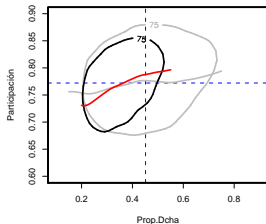
Conclusions

- ▶ A les 5 grans ciutats hi ha una relació positiva forta entre la proporció de vots a la dreta i la participació.
- ▶ Aquesta relació passa pràcticament inadvertida si s'analitzen les dades a nivell nacional degut a l'heterogeneïtat dels votants que comparteixen mesa electoral, que es redueix significativament quan ens limitem a meses electorals situades en una mateixa gran ciutat.
- ▶ Limitar l'estudi a una gran ciutat també permet controlar l'efecte d'altres variables influents, que prenen el mateix valor a totes les meses d'una mateixa ciutat.
- ▶ Concloem que l'abstenció perjudica als partits d'esquerra.
- ▶ Aquestes conclusions guanyen força si tenim en compte que les dades analitzades corresponen a les Eleccions Generals de 2004, quan l'electorat d'esquerra va estar especialment mobilitzat.

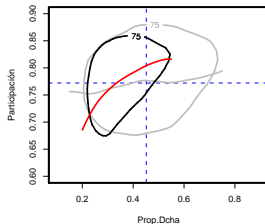
Espanya (Corr= 0.13)



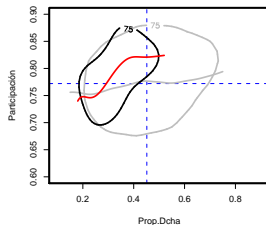
Catalunya (Corr= 0.34)



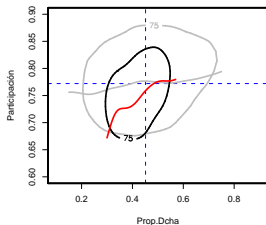
BARCELONA (Corr= 0.46)



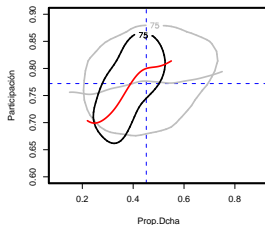
GIRONA (Corr= 0.52)



LLEIDA (Corr= 0.43)



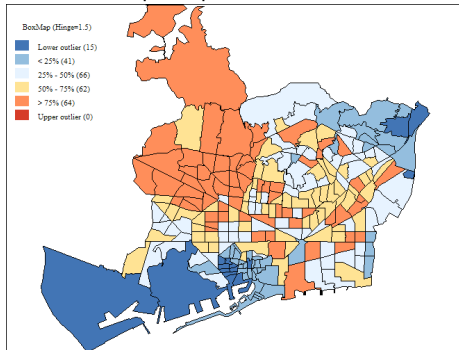
TARRAGONA (Corr= 0.57)



Municipi de Barcelona. Participació electoral en 248 ZRP. Eleccions Generals 2004

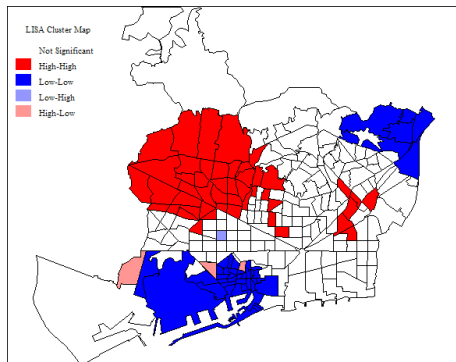
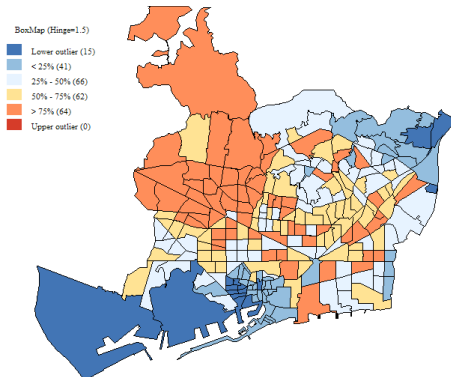
(Treball conjunt amb Sonia Broner, UPC)

Tasa de participación a nivel de ZRP



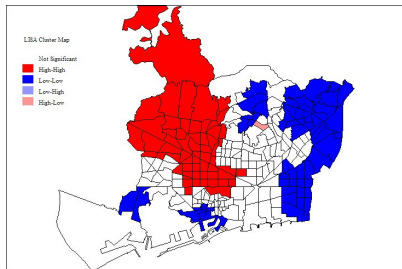
Variable	Min.	Q1	Mediana	Media	Q3	Max.
Participación	0.33	0.74	0.79	0.76	0.81	0.87

Mapa LISA: Participació en 248 ZRP de Barcelona. Eleccions Generals 2004

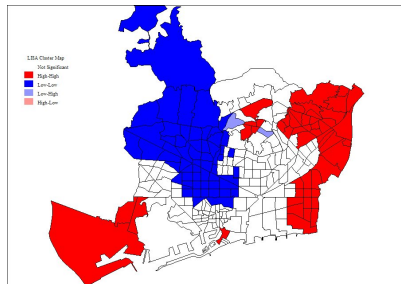


Mapes LISA pels resultats de CiU i PSC

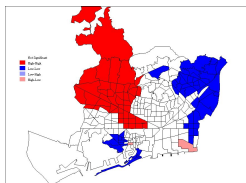
CIU: Proporción de votos



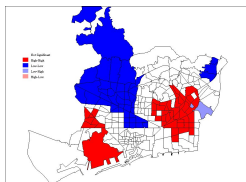
PSC: Proporción de votos



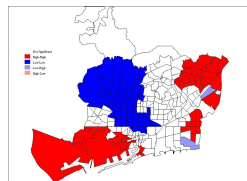
Mapes LISA per altres variables relacionades amb la participació



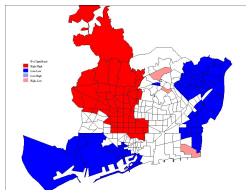
Renta.



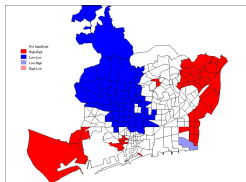
Técnicos.



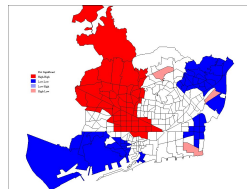
Primario.



Directivos y profesionales.

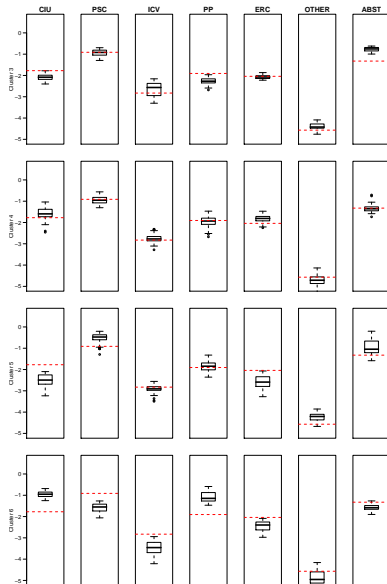
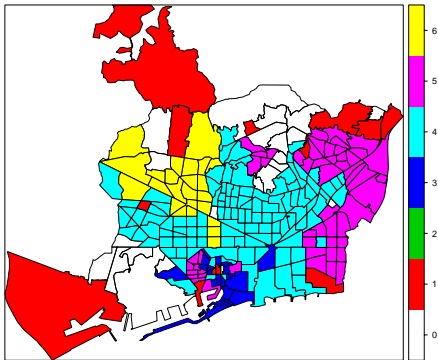


Trabajadores manuales.



Secundario y superior.

Orientació del vot i abstenció. Generals 2004.



Dades electorals

Sondeigs electorals

Abstenció i orientació del vot

Piràmides de població dels països del món

Gapminder. Un observatori del desenvolupament humà

Tendències del desenvolupament humà 2005

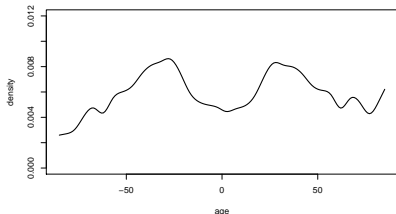
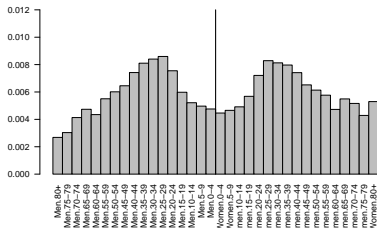
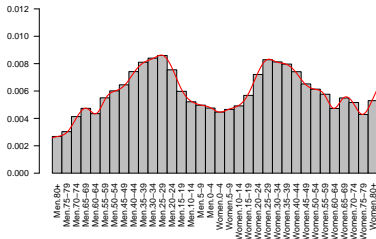
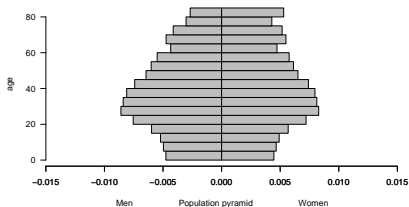
Altres obres de Gapminder

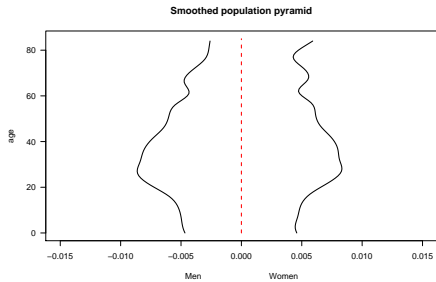
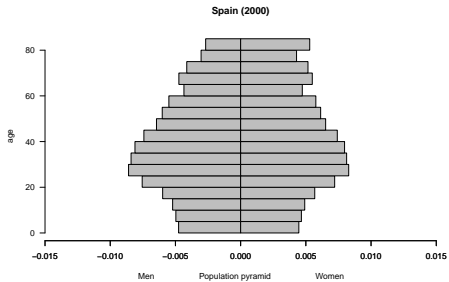
Piràmides de població

- ▶ La piràmide de població d'un territori (municipi, país) representa la distribució conjunta de les variables (**sexe**, **edat**) dels seus habitants.
- ▶ És un doble histograma de la variable **edat**, un per cada **sexe**.
- ▶ Podem considerar les piràmides de població com funcions suaus.
- ▶ Seran funcions de densitat: són no negatives i integren 1.
- ▶ **Anàlisi de dades funcionals.**

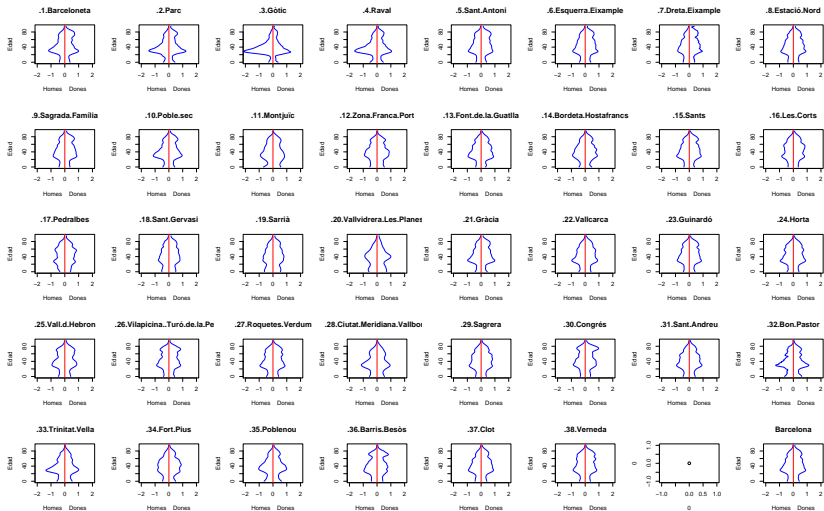
Exemple. Piràmide de població d'Espanya, any 2000

Spain (2000)

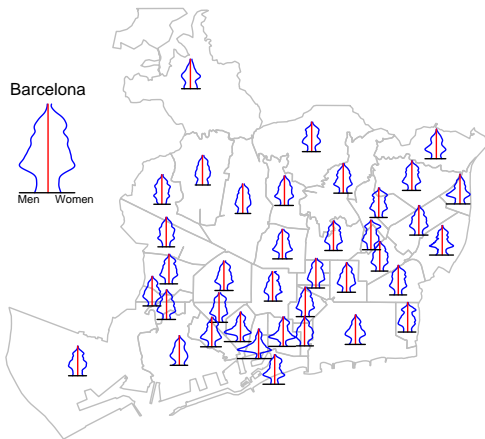




Piràmides de població, Barcelona (38 ZE, 1-1-2005)

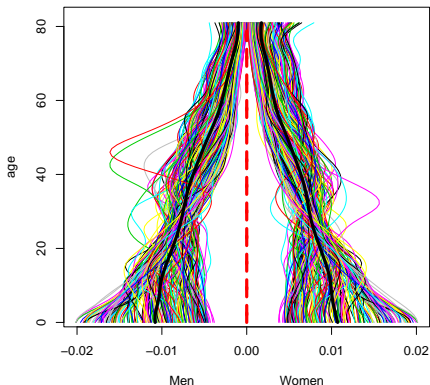


Exemple 2: Piràmides de població en Barcelona



Piràmides de Població, 223 països, any 2000

Smoothed population pyramids for 223 countries (2000)



- ▶ International Data Base (IDB), U.S. Census Bureau.
- ▶ Cada país: freqüències creuades de sexe i edat, en intervals de 5 anys.
- ▶ Les funcions s'obtenen fent interpolació per splines cúbics.

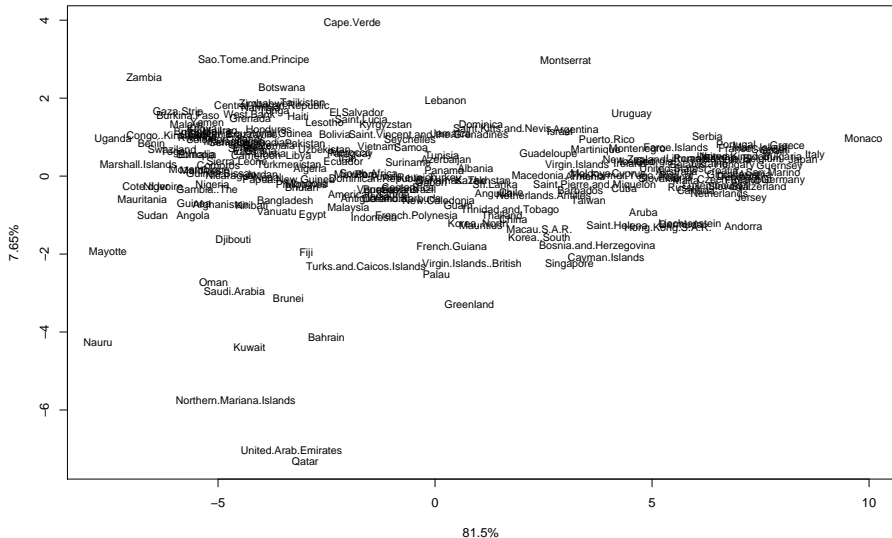
FPCA i MDS per piràmides de població

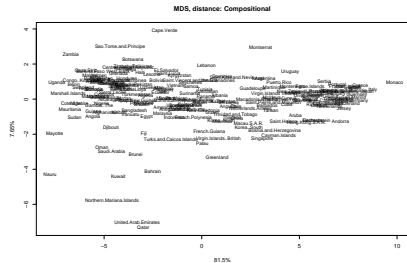
- ▶ Hi ha versions funcionals de moltes tècniques estadístiques.
- ▶ En particular, són interessants les tècniques de reducció de la dimensionalitat per a dades funcionals:
 - ▶ Anàlisi de components principals funcionals (FPCA)
 - ▶ Escalat multidimensional (MDS) per a dades funcionals
- ▶ Les piràmides de població són funcions de densitat.
- ▶ Les components principals funcionals no són adequades.
- ▶ Les densitats són dades composicionals de dimensió infinita: Egozcue, Díaz-Barrero, and Pawlowsky-Glahn (2006).
- ▶ MDS partint de la distància de Aitchison entre densitats:

$$d_A(f, g)^2 \propto \int_I \int_I \left(\log \frac{f(x)}{f(y)} - \log \frac{g(x)}{g(y)} \right)^2 dx dy.$$

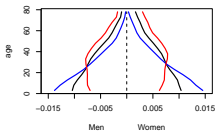
- ▶ Més detalls a Delicado (CSDA, 2011).

MDS, distance: Compositional

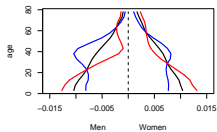




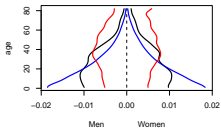
mean +/- regr.coef. 1 by MDS: Compositional



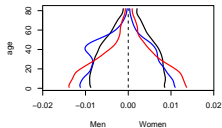
mean +/- regr.coef. 2 by MDS: Compositional



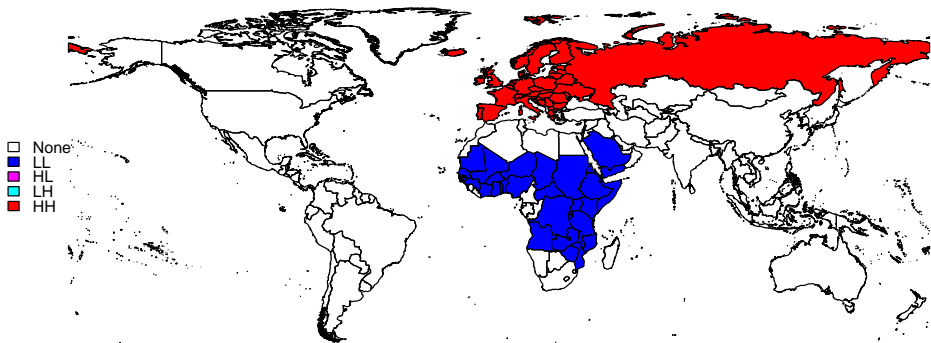
1st ppal. funct. by MDS: Compositional



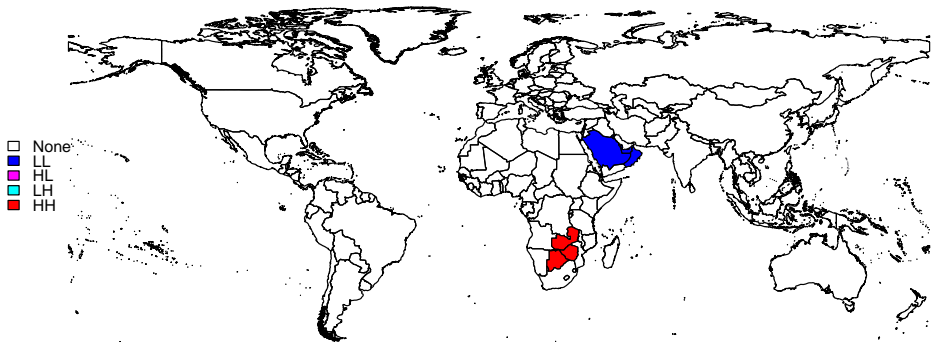
2nd ppal. funct. by MDS: Compositional



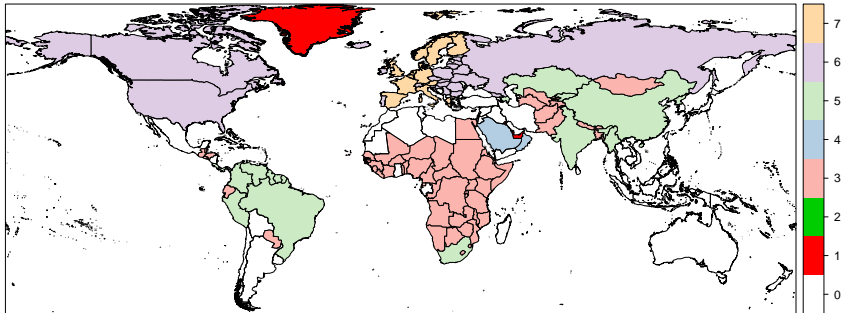
Mapa LISA. Primera coordenada principal MDS. Piràmides de població, països del món



Mapa LISA. Segona coordenada principal MDS. Piràmides de població, països del món

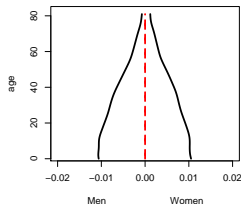


Piràmides de població, països del món. Mapa LISA-bd

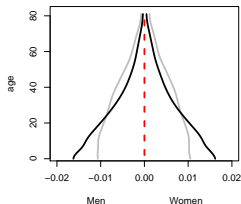


Piràmides de població, països del món. Interpretació

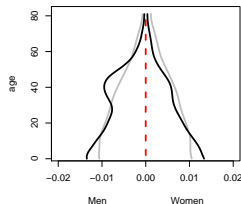
Average population pyramid



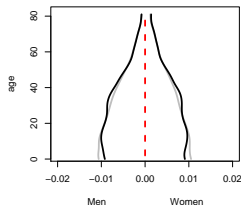
Cluster 3



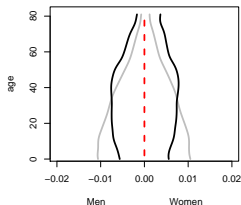
Cluster 4



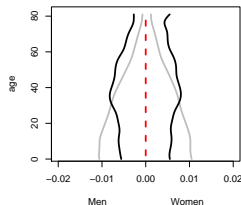
Cluster 5



Cluster 6



Cluster 7



Dades electorals

Sondeigs electorals

Abstenció i orientació del vot

Piràmides de població dels països del món

Gapminder. Un observatori del desenvolupament humà

Tendències del desenvolupament humà 2005

Altres obres de Gapminder

Gapminder. Un observatori del desenvolupament humà

- ▶ **Gapminder:**
Unveiling the beauty of statistics for a fact based world view.
- ▶ *Gapminder promotes a fact based world view by converting the international statistics into moving, interactive, understandable and enjoyable graphics.*
- ▶ Gapminder és una fundació privada sense ànim de lucre.
- ▶ Va ser fundada l'any 2005 a Estocolm (Suècia).
- ▶ El seu Director, i un dels seus fundadors, és **Hans Rosling**.
- ▶ Mireu l'article de Champkin (2009), a Significance, sobre ell.
- ▶ www.gapminder.org

Tendències del desenvolupament humà 2005

- ▶ Va ser una de les primeres feines de Gapminder.
- ▶ És una aplicació feta per a les Nacions Unides per presentar l'Informe de Desenvolupament Humà de 2005
- ▶ Accessible a:
www.gapminder.org/downloads/human-development-trends-2005

Altres obres de Gapminder

- ▶ El producte estrella: [Gapminder World](#).
- ▶ Molt recomenable també: [Dollar Street](#).
- ▶ Les 6 conferències que Hans Rosling ha donat a TED. Accessibles al web de Gapminder, han estat visionades milions de vegades.
- ▶ **TED** (Tecnologia, Entreteniment i Disseny) organitza cicles anuals de conferències des de 1984. Reuneixen els pensadors i emprenedors més fascinants del món, que tenen el repte de donar la xerrada de la seva vida (en 18 minuts).
- ▶ Hans Rosling ha participat des de 2006, amb títols molt suggerents:
 - ▶ Debunking myths about the "third world" (2006).
 - ▶ The seemingly impossible is possible (2007).
 - ▶ Let my dataset change your mindset (2009).
 - ▶ Why ending poverty is crucial to stop population growth (2010).

Moltes gràcies!

Champkin, J. (2009, December).

Hans rosling.

Significance 6, 179–181.

DOI 10.1111/j.1740-9713.2009.00394.x.

Delicado, P. (2011).

Dimensionality reduction when data are density functions.

Computational Statistics and Data Analysis 55, 401–420.

Egozcue, J. J., J. L. Díaz-Barrero, and V. Pawlowsky-Glahn (2006,
July).

Hilbert space of probability density functions based on aitchison
geometry.

Acta Mathematica Sinica 22, 1175–1182.