1. Variables
   1. **Variables Identificatives**

Hem seleccionat, després de pensar en el passos a seguir, les següents variables que estan en ambdues data sets

* + 1. Dataset1 :

"ordre","POND\_GEO","BOP\_num","any","sexe","edat","gr\_edat"

* + 1. Dataset2:

"ORDRECINE","PONDERA","BOP\_NUM","ANY","SEXE","EDAT","GR\_EDAT"

* 1. Observacions sobre les variables identificatíves i compatibilitats.
     1. Variable ordre/ ORDRECINE . Variable que fa referencia a un tipus d’ordenació dels casos consecutivament.Totes dues son de tipus numèric, veiem que hi ha faltants en una edició de l’enquesta, a la BOP\_NUM==4 i la següent edició no esta ordenada sinó penjada de manera aleatòria.
     2. Variable POND\_GEO / PONDERA. Aquesta variable es la que pondera el pes de cada individu dins de la seva enquesta corresponent. Totes dues son del tipus numèric i es important per comprovar que no hem alterat les dades, nomes les hem ordenat.
     3. Variable any/ANY. Aquestes variables son de tipus diferents. “any” es del tipus “haven\_labelled”, que son variables etiquetades. I la variable “ANY” es de tipus numèric. Ho haurem de homogeneïtzar a un sol tipus, al mes complet en informació.
     4. Variable sexe/SEXE. Aquesta variables son del tipus “haven\_labelled”. I la seva codificació es 1 = “Home” , 0 = “Dona”.
     5. Variable gr\_edat / GR\_EDAT. Aquestes variables son totes dues del tipus “haven\_labelled”, però estan codificades diferents , perquè els intervals d’edat son diferents. Homogeneitzem. 1 grup entre 18-25 etc....
     6. Variable mes/Mes. Aquesta variables son totes dues de tipus “numeri”. Tot en ordre.
     7. Variable BOP\_num/BOP\_NUM. Aquestes variables son “haven\_labelled”.

1. **Variables Seleccionades :** 
   1. Context Sociopolític:
      1. Dintre de Principals Problemes:

**P1 : Principals problemes que té actualment Catalunya.**

Es una variable que des de jul.06 fins 2018 pregunta sobre els problemes que creu que té actualment Catalunya, i es recull el que espontàniament diu l’enquestat. Des de la immigració, la inseguretat ciutadana, l’atur i l’accés a l’habitatge... Tot en 23 variables codificades amb un numero concret que fa referencia al seu nom, exemple: cada columna te la codificació del enter de la seva etiqueta. Però a partir de la database2, la que va del 2014-18, esta codificada amb 1 i 0. Suposo que quan ajunti les dades, integraré totes les dades a 1 o 0. Consultar-ho.Degut a la dificultat de treballar amb 22 variables, intentaré convinarles totes en una sola variable que concentre la resposta però ho veig difícil ja que en una primera observació preliminar, un subjecte ha respost mes d’un problema....buscare informació de com treballar amb dades d’aquest tipus.

* + 1. Dintre de Situació econòmica :

**P4 : Valoració de la situació econòmica actual a Catalunya.**

Es una variable amb 7 possibles respostes, codificades numèricament amb un valor enter. Les etiquetes i codificació coincideixen en les 2 data sets. Dades des de març del 07 fins 2018.Nomès he de fer coincidir les etiquetes explicatives dins del nom de la variable en el Dataset. Es “haven\_labelled”

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Molt bona  2 Bona  3 Ni bona ni dolenta  4 Dolenta  5 Molt dolenta  98 No ho sap  99 No contesta | 1 Molt bona  2 Bona  3 Ni bona ni dolenta  4 Dolenta  5 Molt dolenta  98 No ho sap  99 No contesta |

* + 1. Dintre de Situació política:

**P9 : Valoració de la situació política actual a Catalunya.**

Es una variable amb 7 possibilitats també, amb una codificació compartida en dos Datasets. Dades des de març del 07 fins 2018. Observo que son del mateix tipus, estan codificades igual .Faig coincidir les etiquetes explicatives situades en el nom de la variable. Es “haven\_labelled”.

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Molt bona  2 Bona  3 Ni bona ni dolenta  4 Dolenta  5 Molt dolenta  98 No ho sap  99 No contesta | 1 Molt bona  2 Bona  3 Ni bona ni dolenta  4 Dolenta  5 Molt dolenta  98 No ho sap  99 No contesta |

* 1. Actituds davant la política:
     1. Extres:

**P18:Satisfacció amb el funcionament de la nostra democràcia.**

Dades des de jul. 06 fins 2018. La mateixa codificació en els dos datasets.

Modificació de l’etiqueta. Es “havent\_labelled”

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Molt bona  2 Bona  3 Ni bona ni dolenta  4 Dolenta  5 Molt dolenta  98 No ho sap  99 No contesta | 1 Molt bona  2 Bona  3 Ni bona ni dolenta  4 Dolenta  5 Molt dolenta  98 No ho sap  99 No contesta |

* 1. Valors Polítics:
     1. Extres:

**C700.Amb quina de les següents frases se sent mes identificat/da.**

Dades des de el juliol de 2006 fins al 2018. Hi ha 5 possibles respostes, però ha canviat la manera de construir aquesta variable entre les dues data sets, aquí cal trobar-la de manera indirecta, a traves d’altres preguntes, caldrà construir-la per l’ocasió. No la seleccionaré.

**P29/p27:Valoració del nivell D’autonomia de Catalunya.**

Variable única amb diverses categories, 5 en concret. Diferencies de etiquetes entre Dataset1 i Dataset2. Canvi de label i modificació en N/S , N/C per l’expressió escrita. Es “havent\_labelled”.

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Massa Autonomia  2 Un nivell suficient d’autonomia  3 Un nivell insuficient d’autonomia  98 No ho sap  99 No contesta | 1 Massa Autonomia  2 Un nivell suficient d’autonomia  3 Un nivell insuficient d’autonomia  98 No ho sap  99 No contesta |

**P30/p28:Relació entre Catalunya y Espanya.**

Variable única on canvien les etiquetes de les categories en les 2 datasets. Veiem que son totes dues “havent\_labelled”. Canvio l’explicació de la variable i, “NS” i “NC” per “No ho sap”, “No contesta”.

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Una regió d’Espanya  2Una comunitat autònoma d’Espanya  3 Un estat dins d’una Espanya federal  4 Un estat independent  98 No ho sap  99 No contesta | 1 Una regió d’Espanya  2Una comunitat autònoma d’Espanya  3 Un estat dins d’una Espanya federal  4 Un estat independent  98 No ho sap  99 No contesta |

* 1. Comportament electoral:
     1. Parlament de Catalunya:

**P38/Cens.Record de vot sobre cens al Parlament de Catalunya.**

Aquesta variable te noms diferents segons en quin data set estiguem treballant. Però ve a demanar els partits que l’enquestat a votat i fa una relació sobre el cens.Variables de la classe “havent\_labelled”.

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 PPC  2 CIU  3 ERC  4 PSC  5 ICV-EUIA  6 C's. Ciutadans. Partido de la ciutadania  7 Reagrupament.cat  8 SI. Solidaritat Catalana  9 PxC. Plataforma per Catalunya  10 CUP  11 UPyD  80 Altres  93 Nul  94 En blanc  96 No va votar  97 No podia votar  98 No ho sap  99 No contesta | 1 PPC  2 CiU  3 ERC  4 PSC  5 ICV-EUiA  6 C's  7 Reagrupament.cat  8 SI  9 PxC  10 CUP  11 UPyD  12 Podemos  13 Barcelona En Comú  14 CDC  15 Junts pel Sí  16 Catalunya sí que es pot  17 Democràcia i Llibertat  18 En Comú Podem  19 PACMA  20 PDeCAT  21 Junts per Catalunya  22 Catalunya en Comú Podem  80 Altres partits  93 Nul  94 En blanc  96 No va votar  97 No podia votar  98 No ho sap  99 No contesta |

OBSERVEM com hi ha un aument en els partits que es presenten, alguns que han desaparegut i altres de nova creació.

* 1. Dades sociodemogràfiques:
     1. Lloc de naixement:

**C100:Lloc de naixement del enquestat.**

Aquestes dues variables, tenen la mateixa codificació i etiquetes, i fan referencia al lloc geogràfic de naixement del enquestat@. Variables de la classe “havent\_labelled”.

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Catalunya  2 Altres comunitats autònomes  3 Unió Europea  4 Resta del mon  98 NS  99 NC | 1 Catalunya  2 Altres comunitats autònomes  3 Unió Europea  4 Resta del món  98 No ho sap  99 No contesta |

* + 1. Situació laboral:

**C400:Situació laboral del enquestat.**

Aquesta variable, te diferencia de criteris segons la base de dades que tractem, es una variable que ha anat evolucionant molt clarament amb el pas del temps, directament dels data set te aquests nivells:

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Treballa per compte propi  2 Treballa per compte d'altri  3 No treballa  98 NS  99 NC | 1 Treballa  2 No treballa  3 Està temporalment de baixa  98 No ho sap  99 No contesta |

Veiem que hem de fer els canvis usuals de modificar NS i NC i també ajuntar els dos valors que donen treballa del dataset1 i posteriorment trobar la informació amagada al dataset1 , sobre si està o no temporalment de baixa i traspassar aquesta informació a aquesta variable c400 del dataset1 per aconseguir que s’assembli a la variable homologa C400 del dataset2. Hem queda així:

|  |  |
| --- | --- |
| Value Label Dataset1 | Value Label Dataset2 |
| 1 Treballa  2 No treballa  3 Està temporalment de baixa  98 No ho sap  99 No contesta | 1 Treballa  2 No treballa  3 Està temporalment de baixa  98 No ho sap  99 No contesta |

* + 1. Classe social:

**C800:Autoubicació en la classe social.**

* + 1. Extres:

**C600:Estat civil.**

Metodologia

Al correguir les etiquetes que expliquen les variables , he decidit suprimir el numero de la variable, exemple, p28 , al devant de la frase explicativa, ja que a vegades el numero de aquesta variable no coincideix amb el numero de la pregunta en el Questionari i porta a errors i busqueda infructuosa.

Canvio el diversos nivells de possibles respostes al criteri mes actual. Si el CEO ha decidit , per exemple, no registrar “NS” i possar “No sabe” deu ser per algun motiu.

The purpose of the **labelled** package is to provide functions to manipulate metadata as variable labels, value labels and defined missing values using the ***haven\_labelled*** and ***haven\_labelled\_spss*** classes introduced in haven package.

Vale...

Bibliografia:

1. Per treballar amb variables de la classe “haven\_ lavelled” : <https://cran.r-project.org/web/packages/labelled/vignettes/intro_labelled.html>
2. Per ordenar variables : <http://vita.had.co.nz/papers/tidy-data.pdf>
3. Per ordenar variables: <https://tidyr.tidyverse.org/>
4. Per saber mes sobre les dades categoriques el llibre : agresti-introduction-to-categorical-data
5. Per saber com