

Treball de Fi de Grau

Títol

Anàlisi i Visualització de Dades: Exploració
Demogràfica i Sanitària a Espanya

Autoria

Dídac González Muñoz

Professorat tutor

José Luis Valero Sancho

Grau

Comunicació Audiovisual	
Periodisme	
Publicitat i Relacions Públiques	
Comunicació Interactiva	x
Comunicació de les Organitzacions	

Tipus de TFG

Projecte	
Recerca	x

Data

Del 3 al 7 de juny de 2024	x
26 de juliol de 2024	

Full resum del TFG

Títol del Treball Fi de Grau:

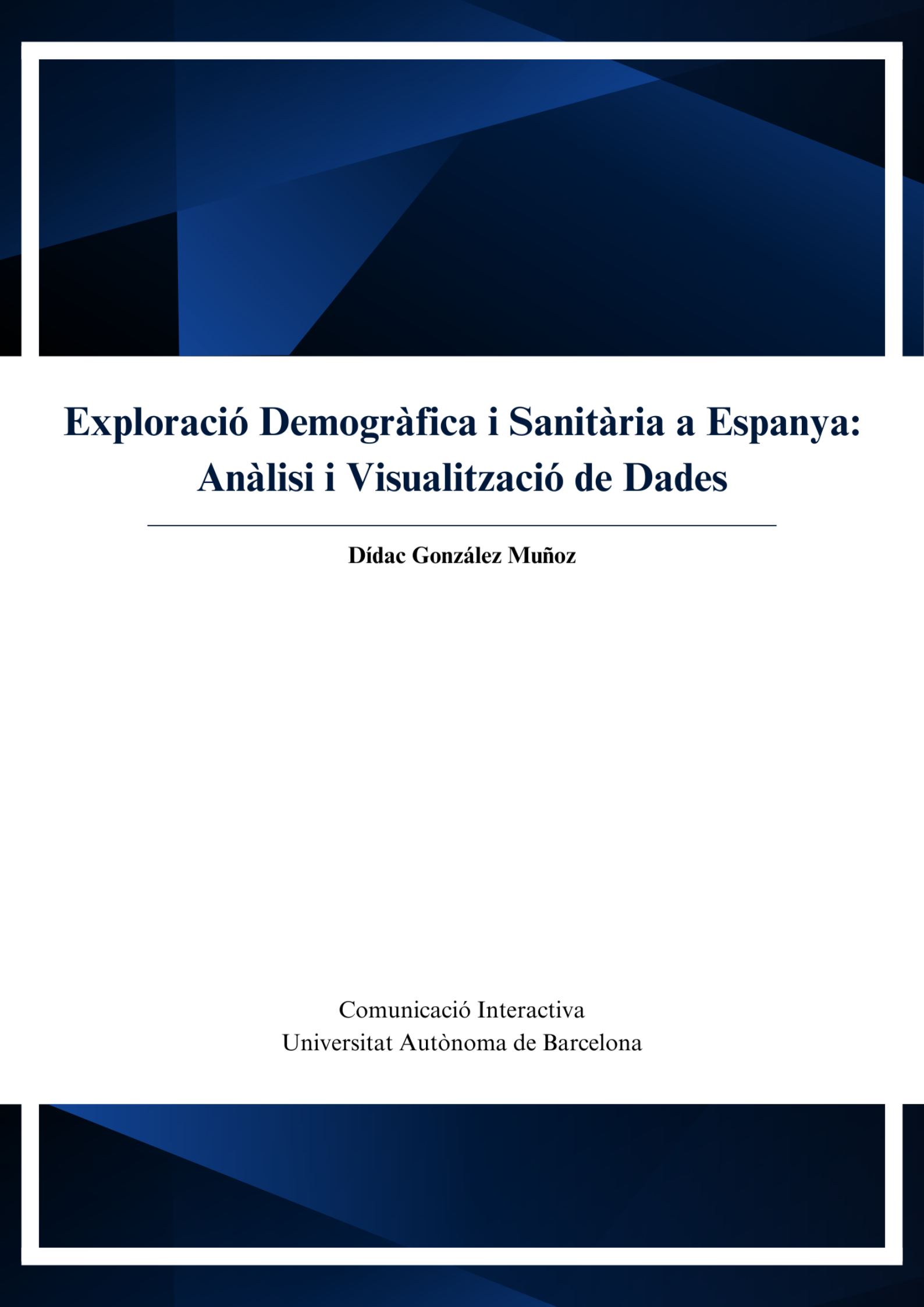
Català:	Anàlisi i Visualització de Dades: Exploració Demogràfica i Sanitària a Espanya			
Castellà:	Análisis y Visualización de Datos: Exploración Demográfica y Sanitaria en España			
Anglès:	Data Analysis and Visualization: Demographic and Health Exploration in Spain			
Autoria:		Dídac González Muñoz		
Professorat tutor:		José Luis Valero Sancho		
Curs:	2023/24	Grau:	Comunicació Audiovisual	
			Periodisme	
			Publicitat i Relacions Públiques	
			Comunicació Interactiva	x
			Comunicació de les Organitzacions	

Paraules clau (mínim 3)

Català:	Anàlisi de dades, Visualització de dades, Demografia, Salut pública, Indicadors sociodemogràfics
Castellà:	Análisis de datos, Visualización de datos, Demografía, Salud pública, Indicadores sociodemográficos
Anglès:	Data analysis, Data visualization, Demographics, Public health, Sociodemographic indicators

Resum del Treball Fi de Grau (extensió màxima 100 paraules)

Català:	Aquest treball explora el panorama demogràfic i sanitari d'Espanya a través de l'anàlisi i la visualització de dades. Es busca realitzar un estudi dels fonaments i la teoria tant del procés de l'anàlisi de les dades com de la seva representació visual, per tal de poder examinar els principals indicadors demogràfics i sanitaris del país. A través d'aquesta anàlisi, es busca identificar patrons i tendències significatives que puguin servir tant per comprendre l'escenari actual com per ser utilitzades com a base en el disseny i implementació de polítiques i estratègies efectives en l'àmbit demogràfic i sanitari.
Castellà:	Este trabajo explora el panorama demográfico y sanitario de España a través del análisis y la visualización de datos. Se busca realizar un estudio de los fundamentos y la teoría tanto del proceso del análisis de los datos como de su representación visual, para poder examinar los principales indicadores demográficos y sanitarios del país. A través de este análisis, se busca identificar patrones y tendencias significativas que puedan servir tanto para entender el escenario actual como para ser utilizadas como base al diseñar e implementar políticas y estrategias eficaces en el ámbito demográfico y sanitario.
Anglès:	This thesis explores the demographic and health scene of Spain through data analysis and visualization. The aim is to carry out a study of the foundations and theory of both the data analysis process and its visual representation, in order to examine the main demographic and health indicators of the country. Through this analysis, we seek to identify significant patterns and trends that can serve both to understand the current scenario and to be used as a basis when designing and implementing effective policies and strategies in the demographic and health field.



Exploració Demogràfica i Sanitària a Espanya: Anàlisi i Visualització de Dades

Dídac González Muñoz

Comunicació Interactiva
Universitat Autònoma de Barcelona

Índex de continguts

Índex d'il·lustracions	3
1. Introducció.....	7
1.1. Presentació del tema.....	7
1.2. Objectius del treball	9
1.3. Preguntes d'investigació	10
2. Marc teòric.....	12
2.1. Teoria i fases de l'anàlisi de dades	12
2.1.1. El pensament analític i el pensament sistèmic	13
2.1.2. Les sis fases de l'anàlisi de dades	14
2.1.3. El mètode SMART	16
2.1.4. La formulació de preguntes d'investigació, hipòtesis i objectius.	16
2.1.5. Dades quantitatives i qualitatives	17
2.1.6. Limitacions de les dades	18
2.1.7. Recopilació de dades	19
2.1.8. Organitzar les dades	20
2.1.9. Garantir la integritat de les dades	20
2.1.10. Analitzar les dades	22
2.1.11. Compartir les troballes	22
2.1.12. La importància de la visualització de dades	23
2.1.13. Implementació de decisions data-driven	24
2.1.14. El cicle de vida de les dades	25

2.2.	Teoria de la visualització de dades	28
2.2.1.	Semiologia gràfica.....	30
2.2.2.	Percepció gràfica	31
2.2.3.	Principals visualitzacions de dades	32
2.2.4.	Erroros comuns i com evitar-los	35
2.2.5.	El mètode McCandless	40
2.3.	Glossari de termes especialitzats	42
3.	Metodologia	45
3.1.	Disseny de la investigació	45
3.2.	Consideracions ètiques	47
3.3.	Selecció de dades i fonts d'informació	48
4.	Anàlisi de resultats	51
4.1.	Demografia i projeccions de població.....	51
4.2.	Salut i esperança de vida	61
4.3.	Mortalitat i problemes de salut.....	66
4.4.	Hàbits de salut i conductes de risc	70
4.5.	Salut mental.....	77
5.	Conclusions	81
5.1.	Conclusions de l'anàlisi	81
5.2.	Conclusions generals.....	84
6.	Bibliografia	87

Índex d'il·lustracions

Il·lustració 1: " 6 Phases Of Data Analysis", Font: Google Career Certificates. (2021).	14
Il·lustració 2: "Planetary movements shown as cyclic inclinations over time, by an unknown astronomer, appearing in a 10th century appendix to commentaries by A. T. Macrobius on Cicero's In Somnium Scipionus", Font: Funkhouser, H. G., 1937.....	28
Il·lustració 3: "The Da Ming Hunyi Tu", Font: Evan Andrews, 2023.	29
Il·lustració 4: "Charles Minard's Flow Map of Napoleon's Russian Campaign of 1812", Font: Michael Sandberg, 2013.....	29
Il·lustració 5: "Exemple d'un gràfic de distribució", Font: Ferdio, 2014.....	33
Il·lustració 6: "Exemple d'un gràfic de relació", Font: Ferdio, 2014.	33
Il·lustració 7: "Exemple d'un gràfic de composició", Font: Ferdio, 2014.....	34
Il·lustració 8: "Exemple d'un gràfic de comparació", Font: Ferdio, 2014.	34
Il·lustració 9: "Anomalies de la Temperatura Mitjana Mensual de la Superfície Terrestre Global en °C (2016-2024)", Font: Elaboració Pròpia a partir de dades extretes de National Centers for Environmental Information (2024).	35
Il·lustració 10: "Anomalies de la Temperatura Mitjana Mensual de la Superfície Terrestre Global en °C (1850-2024)", Font: Elaboració Pròpia a partir de dades extretes de National Centers for Environmental Information (2024).	36
Il·lustració 11: "Jocs Paralímpics de Rio 2016 – Natació – 100 Metres Braça Femení SB5", Font: Elaboració Pròpia a partir de dades extretes de International Paralympic Committee, Women's 100m breaststroke SB5, Rio 2016 (2024).	37
Il·lustració 12: "Jocs Paralímpics de Rio 2016 – Natació – 100 Metres Braça Femení SB5", Font: Elaboració Pròpia a partir de dades extretes de International Paralympic Committee, Women's 100m breaststroke SB5, Rio 2016 (2024).	37

Il·lustració 13: “Representació fictícia dels incidents per ofegament en piscines i les unitats de gelats venudes en un any”, Font: Elaboració Pròpia, 2024.	39
Il·lustració 14: “The McCandless Method – What makes a Good Visualization?”, Font: David McCandless, 2024.	41
Il·lustració 15: “Població a Espanya (1960-2024)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).	51
Il·lustració 16: “Projecció de Població a Espanya (2024-2100)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1st January by age, sex and type of projection (2024).	52
Il·lustració 17: “Piràmide de Població d’Espanya 1983”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).	53
Il·lustració 18: “Piràmide de Població d’Espanya 2023”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).	55
Il·lustració 19: “Piràmides de Població a Espanya (1983-2023)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).	56
Il·lustració 20: “Piràmide de Població Projectada d’Espanya 2083”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1st January by age, sex and type of projection (2024).	57
Il·lustració 21: “Evolució de les Taxes de Natalitat i Mortalitat a Espanya (1960-2022)”, Font: Elaboració Pròpia a partir de dades extretes de Datosmacro.com, Natalidad España 2022 i Mortalidad España 2022. (2024).	59
Il·lustració 22: “Evolució de l’Esperança de Vida a Espanya (1975-2023)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Life expectancy by age and sex (2024).	61

Il·lustració 23: “Esperança de Vida a Europa el 2024”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Life expectancy by age and sex (2024).	62
Il·lustració 24: “Esperança de Vida Projectada a Europa el 2100”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Projected life expectancy by age (in completed years), sex and type of projection (2024).	63
Il·lustració 25: “Anys de Vida Saludables a Espanya (2004-2021)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Healthy life years by sex (from 2004 onwards) (2024).	64
Il·lustració 26: “Morts a Espanya per Malaltia el 2021”, Font: Elaboració Pròpia a partir de dades extretes de Instituto Nacional de Estadística, Defunciones según la causa de muerte (2024).	66
Il·lustració 27: “Morts a Espanya per Malaltia el 2022”, Font: Elaboració Pròpia a partir de dades extretes de Instituto Nacional de Estadística, Defunciones según la causa de muerte (2024).	67
Il·lustració 28: “Percentatge de Població Afectada per Problemes Crònics de Salut a Espanya”, Font: Elaboració Pròpia a partir de dades extretes de Ministerio de Sanidad de España, Principales problemas crónicos de salud (2024).	68
Il·lustració 29: “Percentatge de Fumadors Diaris per País a Europa”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Smoking of tobacco products by sex, age and degree of urbanisation (2024).	71
Il·lustració 30: “Percentatge de Fumadors Diaris per Gènere a Espanya (1993-2020)”, Font: Elaboració Pròpia a partir de dades extretes de Indicadores Clave Sistema Nacional de Salud, Tabaquismo (2024).	72

Il·lustració 31: “Percentatge de Consumidors Diaris d'Alcohol per País a Europa”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Frequency of alcohol consumption by sex, age and educational attainment level (2024).	73
Il·lustració 32: “Percentatge de Consumidors Diaris d'Alcohol per Gènere i Franja d'Edat a Espanya”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Frequency of alcohol consumption by sex, age and educational attainment level (2024).	74
Il·lustració 33: “Percentatge de Població amb Sobrepès o Obesitat per País a Europa”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Obesity rate by body mass index (BMI) (2024).	75
Il·lustració 34: “Percentatge de Població amb Síntomes Depressius a Espanya”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Current depressive symptoms by sex, age and income quintile (2024).	77
Il·lustració 35: “Percentatge de Població que Acudeix Voluntàriament a Professionals de la Salut Mental”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Self-reported consultation of mental healthcare or rehabilitative care professionals (2024).	78
Il·lustració 36: “Suïcidis a Espanya (1980-2022)”, Font: Elaboració Pròpia a partir de dades extretes de Instituto Nacional de Estadística, Defunciones según la causa de muerte (2024).	79

1. Introducció

1.1. Presentació del tema

Aquest projecte d'investigació sobre el panorama demogràfic i sanitari a Espanya, mitjançant l'anàlisi i la visualització de dades, està motivat principalment per un recent i creixent interès en tot el procés d'una anàlisi de dades i la seva posterior visualització. I quina millor oportunitat que aquest treball per poder investigar, analitzar i visualitzar dades d'algun tema que personalment susciti interès? Aquest és el cas de les dades demogràfiques del país, necessàries per entendre les tendències i els patrons que caracteritzen i conformen la nostra població.

Des del punt de vista d'un jove estudiant universitari de Comunicació Interactiva, considero que comprendre la situació demogràfica i sanitària del país en què vivim és de gran interès i rellevància per diverses raons: des d'entendre la composició poblacional, incloent-hi les seves característiques i necessitats específiques, fins a estudiar les tendències demogràfiques i sanitàries emergents que tenen un impacte significatiu en el context social i cultural. Com a comunicadors, tenim la responsabilitat d'informar sobre temes crítics de salut pública i demogràfics per contribuir a una societat més informada i preparada per afrontar desafiaments col·lectius.

Conèixer aquest panorama demogràfic i sanitari resulta especialment interessant en el context actual, després de la recent pandèmia de la COVID-19 que va sacsejar el món fa només quatre anys. Aquest període ha alterat significativament les dinàmiques poblacionals, sanitàries i socials, oferint una oportunitat única per analitzar com aquests factors han evolucionat i com s'han adaptat els sistemes de salut i la societat a nous desafiaments.

Per això, l'anàlisi i la visualització de dades es presenten com a eines poderoses per desentranyar la complexitat d'aquests temes i exposar-los al públic en general. Aquest treball busca, per tant, contribuir de forma significativa a la comprensió dels desafiaments i oportunitats que presenta el panorama sociodemogràfic actual d'Espanya, a través de l'anàlisi i la visualització de grans conjunts de dades demogràfiques i sanitàries, proporcionant informació valuosa per a la presa de decisions informades en matèria de salut pública i benestar social.

Selecció de temes i indicadors

En primer lloc, es durà a terme una curiosa selecció de temes i indicadors demogràfics i de salut, buscant capturar la complexitat del panorama sociodemogràfic espanyol. La inclusió de diferents elements i indicadors permetrà obtenir una visió integral i representativa de la situació sanitària i demogràfica del país, proporcionant així una base sòlida per a l'anàlisi i la interpretació de les dades.

Obtenció de dades

La recopilació de dades es durà a terme a partir de fonts fiables i reconegudes, com l'Eurostat, l'Institut Nacional d'Estadística (INE) i el Ministeri de Sanitat d'Espanya, entre altres portals de dades tant primàries com secundàries, però sempre d'alta qualitat i tan recents com sigui possible. D'aquesta manera es garantirà la integritat i la fiabilitat de la informació utilitzada en l'anàlisi, proporcionant una base sòlida per a les conclusions i recomanacions derivades de l'estudi.

Neteja de dades

Els conjunts de dades obtinguts se sotmetran a un rigorós procés de neteja i preparació per garantir la coherència i la precisió de les dades utilitzades en l'anàlisi. Cada conjunt de dades serà revisat individualment en un procés manual, on es verificarà el format de les dades, s'eliminaran valors atípics i es corregiran possibles errors, en un pas crucial per assegurar que les dades s'alineïn amb els objectius específics de l'estudi i que siguin aptes per a la seva posterior anàlisi i visualització.

Visualització de dades

Quan les dades hagin estat netejades i preparades, es farà ús d'un programa referent especialitzat en la visualització de dades dins el món empresarial anomenat Tableau per representar els resultats de manera efectiva i comprensible (Tableau, 2023). D'aquesta manera, es facilitarà la interpretació de les troballes i es potenciarà la comunicació dels resultats de manera accessible per a diversos públics.

Tipologia d'estudi

L'anàlisi de les dades es durà a terme sota un enfocament quantitatiu i objectiu, emprant tècniques estadístiques apropiades per analitzar les dades recopilades, assegurant la

imparcialitat en la interpretació dels resultats. S'evitarà la introducció de biaixos en l'anàlisi de la informació obtinguda, cosa que garantirà l'objectivitat i la fiabilitat dels resultats obtinguts. L'extracció de conclusions es durà a terme sota el rigor i el respecte per les dades utilitzades en tot moment.

Consideracions ètiques

Dins aquest estudi, es prioritzarà el respecte a les consideracions ètiques a totes les fases de l'anàlisi de dades. Es vetllarà per la privacitat i confidencialitat de la informació, assegurant el compliment de les normatives i regulacions pertinents quant a la protecció de dades. Aquest enfocament ètic garantirà un tractament adequat i responsable de la informació utilitzada a l'estudi, enfortint la integritat i credibilitat dels resultats obtinguts.

Procés teòric de l'anàlisi de dades

Es posarà en pràctica de manera rigorosa i des d'un punt de vista acadèmic la teoria del procés d'anàlisi de dades i les diferents fases. Això implica seguir un enfocament metodològic adequat, que inclogui des de l'exploració inicial de les dades fins a la interpretació dels resultats obtinguts, fent especial èmfasi en les fases més rellevants dins d'aquest treball ja esmentades, com l'obtenció, la neteja, l'anàlisi i la visualització de les dades.

Marc teòric de la visualització de dades

De la mateixa manera que amb la teoria de l'anàlisi de dades, es posaran en pràctica els coneixements adquirits en relació amb la teoria treballada sobre la visualització de dades. S'utilitzaran principis i tècniques reconegudes per representar de manera clara i efectiva els resultats de l'anàlisi i se seleccionaran els tipus de gràfics, mapes i altres recursos visuals més adequats per comunicar de manera òptima les tendències i els patrons identificats en les dades, així com les conclusions extretes.

1.2. Objectius del treball

L'objectiu principal d'aquest treball és poder analitzar de manera integral el panorama demogràfic i sanitari d'Espanya actual i dels darrers anys mitjançant l'anàlisi i la visualització de dades. Es vol comprendre l'evolució de diferents indicadors clau demogràfics per tal de proporcionar informació valuosa per a la presa de decisions informades en matèria de salut pública i polítiques socials.

Per a aconseguir aquest propòsit, s'estableixen els següents objectius:

1. Presentar de manera clara i concisa el panorama demogràfic i sanitari d'Espanya actual i dels darrers anys, identificant canvis significatius i possibles factors subjacents.
2. Explorar les principals tendències en les causes de mortalitat a Espanya,
3. Seleccionar i recopilar dades rellevants de fonts fiables i reconegudes, assegurant la qualitat i la fiabilitat de la informació utilitzada en l'anàlisi.
4. Realitzar un procés exhaustiu de neteja i preparació dels conjunts de dades per garantir-ne la coherència i precisió.
5. Diferenciar i aplicar correctament les tècniques estadístiques i eines de visualització més apropiades per analitzar i representar les dades demogràfiques i de salut de manera efectiva.
6. Assegurar l'objectivitat en la interpretació de cadascun dels resultats.
7. Promoure el respecte a les consideracions ètiques a totes les etapes de l'estudi, assegurant el compliment de les normatives i regulacions en matèria de protecció de dades.
8. Sintetitzar de manera coherent i concisa les troballes i conclusions derivades de l'anàlisi de dades.

1.3. Preguntes d'investigació

I per guiar la investigació per un recorregut coherent, analitzant temes anant de més general a més concret, per extreure la màxima informació i de la millor qualitat possible, es plantegen una sèrie de preguntes d'investigació:

1. Com ha evolucionat el panorama demogràfic a Espanya en els últims anys i quins factors han pogut influir en aquests canvis?
2. Quina esperança de vida té la població espanyola actualment i com d'alta és comparada amb altres països europeus?
3. Va afectar la pandèmia de la COVID-19 pel que fa a l'esperança de vida i els anys de vida saludables?
4. Quines van ser les principals causes de mort a Espanya després de la pandèmia i quins canvis hi va haver en anys posteriors?

5. Quins són els principals factors de risc per a la salut de les persones que poden esdevenir en problemes crònics de salut i com es compara Espanya amb la resta de països d'Europa respecte a aquests?

2. Marc teòric

2.1. Teoria i fases de l'anàlisi de dades

En plena era digital, caracteritzada per la proliferació exponencial de dades generades d'una àmplia varietat de fonts i la hiperconnexió de la societat a internet i mitjans digitals, les dades són considerades com un dels actius més valuosos i amb més potencial per organitzacions, empreses i institucions. És per això que l'anàlisi de dades s'ha convertit en un pilar fonamental en diverses àrees, des de la investigació científica fins a la presa de decisions en entorns empresarials i la formulació de polítiques públiques.

L'anàlisi de dades és una ciència interdisciplinària que engloba una àmplia gamma de tècniques, mètodes i eines utilitzades per transformar quantitats massives de dades crues en informació precisa i coneixement significatiu mitjançant diferents tècniques estadístiques, matemàtiques i informàtiques (Crabtree, & Nehme, 2024). L'objectiu principal de l'anàlisi de dades és descobrir patrons, tendències i relacions aparentment ocultes dins de diferents conjunts de dades que puguin proporcionar *insights* valuosos per comprendre diferents fenòmens, prendre decisions i resoldre problemes sobre la base d'aquestes dades.

Per a portar a terme qualsevol anàlisi de dades de manera efectiva, se segueix un procés estandaritzat que consisteix en 6 fases: **preguntar**, **preparar**, **processar**, **analitzar**, **compartir** i **actuar** (Google Career Certificates, 2021). Aquestes fases treballen de manera conjunta i cadascuna esdevé un paper crucial en el procés efectiu d'una anàlisi de dades, des de la formulació de preguntes i la recopilació de dades, fins a la interpretació dels resultats i la implementació d'accions *data-driven*. A través d'aquestes fases es busca tancar el cicle de l'anàlisi de dades i aconseguir un impacte significatiu en la presa de decisions informades i la resolució de problemes.

No obstant això, en aquest treball es farà especial èmfasi en les fases més necessàries i més alineades amb els objectius d'estudi, com són les fases de **Preguntar** (establir objectius i enfocar l'anàlisi), **Processar** (netejar i preparar a mida els grans conjunts de dades per facilitar l'anàlisi), **Analitzar** (explorar i comprendre les dades en profunditat) i **Compartir** (comunicar els resultats de manera efectiva a través de visualitzacions de dades).

2.1.1. El pensament analític i el pensament sistèmic

Abans d'abordar una anàlisi de dades, cal posar unes bases adequades en la forma d'enfocar un procés tan llarg i complex. El pensament analític, també conegut com a *analytical thinking*, és un procés cognitiu que implica identificar i delinear meticulosament un problema per resoldre'l de manera metòdica mitjançant un enfocament estructurat i basat en dades, desenvolupant la capacitat per a pensar de manera reflexiva i discernir, resoldre problemes i analitzar dades utilitzant tota la informació a l'abast (Amer, 2005).

En primer lloc, l'estratègia és essencial per mantenir-nos enfocats i encaminats en mig de grans quantitats de dades al nostre abast. Estar orientat al problema ens ajuda a identificar, descriure i resoldre els reptes que afrontem. La capacitat per identificar correlacions entre diferents conjunts de dades ens permet descobrir relacions significatives entre elles, tot i que és important tenir en compte que correlació no sempre implicarà una causalitat. D'altra banda, pensar tant en termes generals com en els detalls ens permetrà obtenir una comprensió encara més completa de la situació. I finalment, la visualització farà possible la representació gràfica i visual de les troballes, facilitant la seva comprensió i explicació.

En comparació a aquest enfocament, està el pensament sistèmic o *systemic thinking* en anglès. Aquest se centra en entendre i gestionar les interrelacions i la interdependència entre les parts d'un sistema. Mentre el pensament analític descompon problemes complexos en components individuals per analitzar-los per separat, el pensament sistèmic considera el sistema en la seva totalitat, prestant atenció als patrons emergents i les dinàmiques que sorgeixen de les interaccions entre les parts. Aquest enfocament holístic permet reconèixer una sèrie de causalitats en sistemes més complexos, oferint una comprensió més profunda, i és especialment valuós en contextos en què les interaccions i les relacions entre elements són fonamentals, fomentant l'adaptabilitat i la capacitat per gestionar situacions dinàmiques i canviants (Bartlett, 2001).

La capacitat de poder integrar ambdós enfocaments en funció del context pot proporcionar una perspectiva més robusta i eficaç en la resolució de problemes. En l'anàlisi de dades, la capacitat per a formular les preguntes correctes i pensar tant de forma crítica com creativa són habilitats indispensables per a trobar solucions efectives i revelar informació significativa enmig de la complexitat de les dades disponibles.

2.1.2. Les sis fases de l'anàlisi de dades

L'anàlisi de dades és un procés crític per obtenir coneixements amb valor i proporciona una base sòlida per la presa de decisions informades i la resolució de problemes. Si bé no existeix un marc rígid i estandarditzat, generalment existeixen fases reconegudes i acceptades dins el món de les dades que serveixen per guiar aquest procés i proporcionar un enfocament estructurat per l'extracció d'*insights* útils i rellevants.

És per això que la multinacional tecnològica estatunidenca Google ha volgut definir un procés estàndard detallant sis fases claus dins aquest procediment (**Preguntar, Preparar, Processar, Analitzar, Compartir i Actuar**), fent que el seu seguiment pugui millorar significativament l'eficàcia dels resultats extrets. Tot i que la importància de cada fase pot variar en funció dels objectius de cada anàlisi, totes i cadascuna de les fases juguen un paper crucial en guiar als analistes a través del viatge de la transformació de les dades, facilitant la implementació de resultats basats en elles (Quantum Analytics, 2022).



Il·lustració 1: "6 Phases Of Data Analysis", Font: Google Career Certificates. (2021).

Començant amb la fase de **Preguntar**, els analistes busquen l'arrel del problema o el motiu de la investigació formulant els dubtes pertinents per conèixer les qüestions que necessiten una resolució. Entendre i delinear el propòsit darrere de l'anàlisi és fonamental per una correcta execució. Una comunicació i col·laboració efectiva amb les parts interessades, sovint també anomenades *stakeholders* en entorns empresarials, garanteix l'alineació dels objectius d'investigació, preparant i contextualitzant l'escenari.

Passant a la fase de **Preparació**, preparar les dades vol dir recopilar, organitzar i fer servir totes aquelles fonts de dades rellevants per una correcta anàlisi. Determinar l'origen de les dades,

utilitzar diferents formats i estructures de dades, assegurar-se que siguin imparcials i creïbles, organitzar les dades i protegir-les són processos amb una importància majúscula en aquesta etapa.

Un cop recollides les dades, l'enfocament es desplaça al tractament de dades. La fase de **Processar** té com a objectiu la correcta neteja i transformació de les dades en brut en dades organitzades i coherents per poder ser analitzades adientment. Aquí s'identifiquen i es corregeixen els errors, es busca la consistència i coherència entre diferents conjunts de dades i es prevé que aquestes continguin qualsevol biaix. Garantir que les dades siguin coherents i lliures de discrepàncies és vital per a obtenir informació valuosa i creïble.

Un cop tenim les dades processades, el següent pas és analitzar-les. La fase d'**Analitzar** és possiblement la fase més important i trivial en tot aquest procés, i es focalitza en descobrir patrons, relacions, tendències i correlacions dins dels *datasets*, servint-se de tècniques analítiques i eines informàtiques complexes per així poder descobrir informació aparentment oculta i revelant oportunitats per posar en pràctica. A través d'una anàlisi i interpretació rigorosa, els analistes sintetitzen dades complexes en coneixements pràctics i permet a les parts interessades prendre decisions *data-driven* amb confiança i claredat.

Un cop analitzades les dades és necessari disseminar les troballes efectivament. Aquest és el propòsit de la fase de **Compartir**, fer que el coneixement extret sigui accessible per tothom. Els analistes apropen les dades a les parts interessades, traduint la informació en visualitzacions de dades i narratives pertinents per fomentar la comprensió i posar a l'abast de l'audiència els descobriments. Les tècniques de visualització de dades com ara gràfics i infografies, són els instruments per transmetre aquest contingut complex de manera fàcil i accessible.

En últim lloc, cal actuar en conseqüència. La fase d'**Actuar**, les parts interessades aprofiten els coneixements obtinguts de l'anàlisi per a impulsar l'acció estratègica i la transformació. Informades per recomanacions i coneixements basats en dades, s'implementen iniciatives específiques dissenyades per a abordar desafiaments clau i capitalitzar oportunitats emergents.

Tot aquest procés d'estudi de les dades opera com un mecanisme d'engranatges on les peces o fases treballen de manera complementària i conjunta per garantir un correcte producte final, des de la formulació inicial de preguntes i objectius fins a l'acció resultant de tot aquest procés.

Aquesta sincronització i mirada holística assegura una anàlisi efectiva. A continuació, veurem detalladament cadascuna d'aquestes fases.

2.1.3. El mètode SMART

George T. Doran és conegut per desenvolupar el mètode SMART en la dècada de 1980, un enfocament popularment utilitzat per a establir objectius i metes específiques i assolibles en diferents àrees, com la gestió empresarial, l'educació i la recerca.

SMART és un acrònim que representa cinc característiques clau que han de tenir els objectius, en anglès *Specific, Measurable, Assignable, Realistic* i *Time-related*. Primer, els objectius han de ser específics, clars i definits, responen a preguntes com qui, quines, quan, on, per què i com. Segon, han de ser mesurables, incloent-hi indicadors concrets que permetin avaluar el progrés cap al seu assoliment. Tercer, els objectius han de ser assignables a persones, responsables de portar-los a terme. Quart, han de ser rellevants, alineats amb els objectius generals i prioritats estratègiques de l'organització o projecte. Finalment, els objectius han de tenir un límit de temps definit per al seu assoliment, establint una urgència i responsabilitat per a complir-los en un període específic (Doran, 1981).

Avui dia, el mètode SMART en la gestió es considera comunament l'estàndard per a desenvolupar metes i objectius eficaços i mesurables perquè proporciona un marc clar i pràctic per a la fixació d'objectius efectiva, garantint que siguin específics, mesurables, assolibles, rellevants i limitats en el temps, la qual cosa augmenta les possibilitats d'èxit i facilita el seguiment i l'avaluació del progrés cap al seu assoliment.

2.1.4. La formulació de preguntes d'investigació, hipòtesis i objectius.

La formulació de preguntes de recerca constitueix el punt de partida fonamental en qualsevol procés d'investigació. Un adequat marc conceptual per a aquest propòsit implica la comprensió de diversos elements teòrics i metodològics que guien la construcció d'interrogants pertinents i significatives (Codina, 2023).

El desenvolupament d'aquestes preguntes de recerca, incloent-hi les hipòtesis de suport i objectius, és un pas inicial, clau i necessari per a produir resultats clínicament rellevants que s'utilitzaran en la pràctica basada en evidència. Una pregunta de recerca ben definida i

específica és més probable que ens guiï en la presa de decisions sobre el disseny de l'estudi i la població i, posteriorment, en quines dades es recopilaran i analitzaran.

L'interès en un tema en particular generalment comença el procés de recerca, però és la familiaritat amb el tema el que ajuda a definir una pregunta de recerca apropiada per a un estudi. Les preguntes sorgeixen d'un dèficit de coneixement percebut dins d'una àrea o camp d'estudi. És important conèixer on es troba el límit entre el coneixement actual i la ignorància. El desenvolupament d'una pregunta de recerca apropiada implica determinar quines incerteses concretes podrien o haurien d'estudiar-se i també racionalitzar la necessitat de la seva recerca (Farrugia, Petrisor, Farrokhyar & Bhandari, 2010).

2.1.5. Dades quantitatives i qualitatives

Un dels principals objectius de l'anàlisi de dades és convertir la informació crua en coneixement. En aquest sentit, és crucial comprendre la distinció entre dades quantitatives i qualitatives, ja que tots dos exerceixen rols diferents en el procés d'anàlisi. Per això, la pràctica anomenada *Mixed Methods Research*, combina la utilització de dades quantitatives i dades qualitatives per la consecució de millors resultats d'anàlisi (Shorten & Smith, 2017).

Les dades quantitatives se centren en mesures específiques i objectives de fets numèrics. Això pot incloure informació sobre el quin, quants i amb quina freqüència ocorren uns certs fenòmens. D'altra banda, les dades qualitatives descriuen mesures subjectives o explicatives de qualitats i característiques, o coses que no poden ser mesures amb dades numèriques.

Els analistes de dades solen utilitzar tots dos tipus de dades en el seu treball. Mentre que les dades quantitatives ens diuen què està succeint, les dades qualitatives poden proporcionar una comprensió més profunda i contextual de les dades quantitatives en explicar les raons darrere d'uns certs resultats numèrics. El treball d'un bon analista de dades és mantenir un correcte equilibri treballant amb tots dos i transformar els números en informació útil i coneixement.

La integració de dades quantitatives i qualitatives en la recerca de mètodes mixtos és efectivament la millor solució per a l'extracció de coneixement dins de qualsevol anàlisi. Segons Almalki, la conclusió és que, sempre que la investigació de mètodes mixtos sigui apropiada per a un projecte donat, el seu ús generarà beneficis positius. Això es deu al fet que l'ús d'enfocaments diferents té el potencial de proporcionar una major profunditat i amplitud

d'informació que no és possible utilitzant enfocaments individuals de forma aïllada (Almalki, 2016).

2.1.6. Limitacions de les dades

Les dades són una eina poderosa per analitzar i conèixer millor la nostra realitat, però com tot, també tenen les seves limitacions inherents. A *The Limits of Data*, C. Thi Nguyen investiga i reflexiona sobre quines són aquestes principals limitacions inherents a l'anàlisi de dades, llançant llum sobre concretament cinc factors crítics que poden afectar a la seva qualitat i eficàcia.

En primer lloc, Nguyen destaca l'anomenada representació incompleta, la qual es dona quan els mètodes que fem servir per obtenir dades filtren informació personal o específica d'aquell mateix context per fer-la més fàcil d'entendre i accessible a tothom, fent que perdem detalls necessaris de camí. Connectat amb el següent punt, la descontextualització, necessària a la recol·lecció de dades per trobar l'equilibri entre fer les dades fàcilment transferibles i capturar tota la complexitat del context, perdent d'altre forma detalls importants a canvi de fer que les dades siguin més fàcils de fer servir.

Un altre aspecte assenyalat és la possible existència de biaix als sistemes de classificació utilitzats, ja que aquests sistemes poden agrupar coses molt diferents dins les mateixes categories, sovint relacionat amb biaixos polítics o socials, lligat amb el següent punt, els interessos polítics. Sovint, aquests influeixen en com es recol·lecten i classifiquen les dades, la qual cosa pot fer que es passin per alt certes dades o se'ls doni més importància a altres.

Finalment, l'estandardització també pot ser un problema donat a una simplificació exagerada de la realitat, deixant de banda la seva complexitat. A més, quan ens enfoquem massa a mesurar les coses objectivament, podem ignorar aspectes subjectius o qualitatius importants de l'experiència humana. Això s'ha ampliat a l'era digital amb els algorismes i l'aprenentatge automàtic (Nguyen, 2024).

Per aquesta raó és important conèixer, reconèixer i abordar aquestes limitacions per a tenir una comprensió més completa i precisa dels fenòmens a analitzar, equilibrar mètodes quantitius i qualitatius i tenint en compte el context en el qual es recullen les dades.

“Data is powerful because it’s universal. The cost is context.” – C. Thi Nguyen, 2024.

2.1.7. Recopilació de dades

En qualsevol procés d'investigació, la recopilació de dades serveix com a base sobre la qual es construeix l'anàlisi. No només és important determinar quines dades es requereixen per abordar les preguntes de recerca, sinó també organitzar-les de manera que condueixin a l'anàlisi.

Identificar i seleccionar fonts de dades apropiades és el primer pas del procés de recopilació de dades. Aquestes fonts inclouen bases de dades, fulls de càlcul, enquestes, interfícies de programació d'aplicacions i diversos altres repositoris de dades que contenen dades pertinents per analitzar-les.

És important la diferenciació entre dades pròpies i dades de tercers, ja que seran els dos tipus de fonts de dades que farem servir, al mateix temps que tenir en compte els avantatges i desavantatges de cadascuna.

Les dades pròpies són informació recopilada directament per una empresa o organització a partir de les interaccions amb clients, usuaris o qualsevol altre tipus d'audiència. Aquestes dades s'obtenen a través de canals controlats per l'empresa, com ara el vostre lloc web, aplicacions mòbils, xarxes socials, correus electrònics o interaccions cara a cara. Les dades de primera mà són considerades altament fiables i precises, ja que reflecteixen les accions i els comportaments reals de l'audiència específica de l'empresa, però també tenen les seves limitacions pròpies com ara un abast limitat, uns costos de recopilació més elevats i un procés purament manual.

D'altra banda, les dades de tercers provenen de fonts externes a l'empresa o organització que les utilitza. Aquestes dades són recopilades per entitats independents com agències governamentals, institucions acadèmiques, empreses de recerca de mercat i proveïdors de dades. A diferència de les dades de primera mà, que s'aconsegueixen directament de les interaccions amb l'audiència, les dades de tercers són adquirides a través de compra, llicència o intercanvi amb altres organitzacions. Sovint, aquestes dades estan disponibles a bases de dades comercials, enquestes públiques, registres governamentals o informes de recerca.

Encara que les dades de tercers poden proporcionar una visió més àmplia del mercat i dels consumidors, també poden presentar desafiaments com a preocupacions sobre la qualitat, la fiabilitat i la privadesa, així com la necessitat de validar i verificar la informació abans del seu ús (Neumann, Tucker, Subramanyam & Marshall, 2023).

Una vegada que s'identifiquen les possibles fonts de dades, el següent pas implica determinar els conjunts de dades o variables específics dins d'aquestes fonts que siguin pertinents per als objectius de la recerca. Això requereix un enfocament perspicaç per a filtrar dades supèrflues o irrelevantes i concentrar-se únicament en elements de dades que contribueixen significativament a la recerca o al plantejament del problema (Creswell & Creswell, 2017).

2.1.8. Organitzar les dades

Un cop tenim totes aquelles dades necessàries per al nostre procés d'anàlisi, cal organitzar-les de manera neta i ordenada, seguint principis clars i consistents per tal de proporcionar una estructura sòlida que faciliti la posterior anàlisi i interpretació de les dades.

Es treballa amb una estructura de dades rectangular per garantir que les dades estiguin ordenades coherentment i siguin fàcilment accessibles. Amb aquest format, cada fila representa una observació individual, mentre que cada columna representa una variable diferent relacionada amb aquestes observacions, fent que cada cel·la contingui un valor únic (Wickham, 2014).

La uniformitat en l'estructura de les dades és essencial per a garantir la coherència i la fiabilitat de la informació. Això implica que totes les observacions en el conjunt de dades han de seguir el mateix format i estar representades de manera consistent.

Una estructura d'un conjunt gran de dades també facilita la comparació i combinació de dades de diferents fonts, així com la realització d'anàlisis estadístiques i la generació d'informes precisos. A més, ajuda a evitar errors i biaixos que podrien sorgir a causa de la falta de coherència en les dades.

2.1.9. Garantir la integritat de les dades

La integritat de les dades és un aspecte fonamental en qualsevol procés d'anàlisi, ja que exerceix una influència directa sobre la qualitat i fiabilitat dels resultats obtinguts. Després de la recopilació de les dades, es torna imperatiu sotmetre-les a un procés de neteja i transformació concorde als objectius de l'anàlisi. La importància d'aquest pas radica en la necessitat de garantir que les dades estiguin lliures d'errors i siguin coherents, evitant així la generació de dades brutes que podrien comprometre la validesa de les troballes.

La integritat de les dades es veu amenaçada en múltiples fronts, com ho són la replicació, la transferència o la manipulació d'aquestes, podent implicar diverses conseqüències, des de la pèrdua d'informació fins a la presa de decisions incorrectes. Resulta essencial implementar mesures que assegurin la precisió, integritat i coherència de les dades al llarg del seu cicle de vida, amb la finalitat de fomentar una anàlisi sòlida i de confiança (Ethan & Alexander, 2022)

Diversos estudis han destacat aquesta importància de garantir la integritat de les dades per a evitar biaixos i errors en l'anàlisi, com el cas de Ziad Obermeyer en el seu estudi *Dissecting racial bias in an algorithm used to manage the health of populations*.

En aquest estudi, els investigadors van examinar un algorisme utilitzat per a assignar recursos d'atenció mèdica en pacients amb malalties cròniques. Es va descobrir que l'algorisme estava esbiaixat racialment, assignant de manera errònia recursos d'atenció mèdica a pacients blancs en lloc de pacients negres amb malalties cròniques similars.

Aquest biaix va ser donat per la inconsistent integritat de les dades emprades per a entrenar l'algorisme, que reflectia els prejudicis històrics i sistemàtics en l'atenció mèdica (Obermeyer, Powers, Vogeli, & Mullainathan, 2019).

Un altre cas semblant és l'estudi realitzat per Buolamwini i Gebru a *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, el qual va investigar el rendiment dels sistemes de reconeixement facial en funció del gènere i la raça.

Aquests sistemes mostraven biaixos significatius, amb taxes d'error més altes per a les dones i les persones negres, especialment les dones negres, un problema novament originat per una baixa integritat de les dades de les quals els algorismes aprenien a treure conclusions, ja que només s'havien tingut en compte a homes blancs.

Com a resultat, els sistemes de reconeixement facial tenien dificultats per a reconèixer amb precisió a persones de gènere femení i fosques de pell. Aquest estudi ressalta com la integritat de les dades influeix en l'equitat i precisió dels sistemes d'intel·ligència artificial, i com els biaixos en les dades poden perpetuar la discriminació en les tecnologies d'anàlisi (Buolamwini & Gebru, 2018)

És per això que tant la integritat de les dades com la neteja d'aquestes són aspectes fonamentals en el procés d'anàlisi de dades, ja que influeixen directament en la fiabilitat i utilitat dels resultats obtinguts.

2.1.10. Analitzar les dades

Una vegada que les dades han estat processades i netejades, el següent pas clau dins del procés d'anàlisi no és un altre que analitzar aquestes dades. Aquesta etapa se centra en la identificació de patrons, relacions, tendències i correlacions dins dels conjunts de dades.

Durant aquesta fase s'apliquen tècniques analítiques avançades i eines informàtiques per a realitzar càlculs complexos i combinar dades de múltiples fonts, amb eines com Excel i SQL o llenguatges de programació com Python o R.

Les dades estructurades són analitzades mitjançant dues tècniques; l'anàlisi quantitativa i l'anàlisi qualitativa. Mitjançant l'anàlisi quantitativa, es poden identificar tendències numèriques i relacions estadístiques entre variables, mentre que l'anàlisi qualitativa permet explorar el context i les implicacions més àmplies de les dades, capturant matisos que van més enllà de les xifres en brut. La combinació de tots dos enfocaments proporciona una visió més completa i detallada de les dades, permetent una presa de decisions més informada i precisa (Talabis & Martin, 2012).

L'anàlisi i la interpretació minuciosa permet als analistes convertir dades complexes en coneixements pràctics, informant a les parts interessades per poder prendre decisions basades en dades amb confiança i claredat.

2.1.11. Compartir les troballes

Compartir les troballes de l'anàlisi mitjançant la visualització de dades és d'alguna manera la culminació del procés d'anàlisi de dades. En aquesta etapa, es busquen mètodes per a presentar els resultats de manera clara, concisa i comprensible per a diversos públics, ja siguin experts en la matèria, directius d'una organització o públic en general.

La comunicació dels resultats de l'anàlisi de dades és fonamental per a garantir la seva utilitat. Si bé les dades proporcionen informació valuosa, el seu veritable valor es materialitza quan poden transformar-se en informació comprensible i coneixement aplicable.

A més, una comunicació efectiva de les troballes pot proporcionar informació valuosa per a la presa de decisions estratègiques així com accelerar el procés de recerca científica. En presentar els resultats de manera clara i comprensible, es facilita que les parts interessades compreguin les implicacions de les dades i prenguin mesures informades basades en ells, evitant esforços innecessaris, permetent actuar mirant noves perspectives i donant suport a avanços de recerca en camps concrets (Figueiredo, 2017).

Existeixen diverses eines per a facilitar aquesta transformació a través de la visualització de dades, però les dues eines principals són els informes i els panells de control.

Els informes ofereixen una representació estàtica de les dades, que normalment es proporcionen periòdicament a les parts interessades. Es destaquen per presentar dades històriques de manera concisa, oferint instantànies de les tendències analitzades. No obstant això, requereixen un manteniment regular, manquen d'atractiu visual i no reflecteixen dades en viu o en evolució.

D'altra banda, els panells de control, àmpliament coneguts com a *dashboards*, proporcionen una plataforma dinàmica per a monitorar les dades entrants a temps real. Ofereixen actualitzacions automàtiques i funcions interactives, la qual cosa permet a les parts interessades interactuar amb les dades a través de filtres. Malgrat el seu valor a llarg termini i un major accés de les parts interessades, els panells requereixen més temps per al seu disseny i poden resultar confusos per als usuaris poc freqüents.

És imperatiu dissenyar aquestes eines de visualització de dades d'acord amb les necessitats i expectatives dels usuaris finals o dels objectius buscats en compartir aquesta informació. Comprendre les fortaleces i limitacions de cada eina és crucial per a comunicar de manera efectiva les troballes de l'anàlisi de dades a les parts interessades (Few, 2006).

2.1.12. La importància de la visualització de dades

En el context de l'anàlisi de dades, la visualització exerceix un paper fonamental a causa de la seva capacitat per a transformar informació complexa en representacions gràfiques comprensibles. Aquesta eina facilita la interpretació i comunicació de dades, la qual cosa resulta crucial per a diversos aspectes, des de la presa de decisions fins a la presentació de troballes a les parts interessades.

La història de la visualització de dades es remunta a segles enrere, quan les persones utilitzaven pictogrames i gràfics rudimentaris per a representar informació. Aquestes primeres visualitzacions incloïen mapes, taules i gràfics simples que ajudaven a comunicar informació de manera visual. Amb el temps, la visualització s'ha tornat més sofisticada i accessible, i amb l'adveniment de l'era digital, el seu abast ha crescut exponencialment, juntament amb la quantitat de dades que es poden representar (Friendly, 2006).

La clau per a una visualització efectiva es troba en la seva capacitat per a transmetre informació de manera clara i ràpida. Els primers segons d'exposició a una visualització són crucials, l'audiència ha de comprendre la informació presentada i les conclusions a les quals s'arriba gairebé immediatament. Aquest enfocament permet que la visualització sigui accessible i útil fins i tot per a aquells que no estan familiaritzats amb els detalls de la recerca (Heer & Bostock, 2010).

Per a crear visualitzacions efectives, és essencial estructurar i organitzar les dades de manera coherent, identificant patrons i conclusions clau. Una visualització exitosa no sols presenta les dades, sinó que també guia a l'audiència cap a les mateixes conclusions que l'analista, de manera clara, simple i íntegra (Tufte, 2001).

En última instància, la visualització de dades no sols serveix com una eina per a presentar informació, sinó també com un mitjà per a contar històries convincent que poden influir en l'opinió pública. No obstant això, és important recordar la responsabilitat de tenir en compte les perspectives dels altres en dissenyar aquestes visualitzacions, garantint així una comunicació efectiva i comprensió mútua.

Aquesta capacitat de comunicar de manera eficaç informació complexa fa que la visualització de dades sigui una eina indispensable en l'arsenal de qualsevol analista de dades, proporcionant un mitjà poderós per a explorar, entendre i compartir coneixements.

2.1.13. Implementació de decisions data-driven

Després d'haver completat el procés de preguntar, preparar, processar, analitzar i compartir les dades de l'anàlisi, ja només queda actuar basant-se en aquestes. La fase d'actuar representa la culminació de tot procés d'anàlisi de dades, on es duen a terme accions concretes i es prenen decisions informades basades en els resultats obtinguts. Com explica Thomas H. Davenport en la seva obra *Competing on Analytics*, la fase d'actuar es pot resumir en 3 directrius clau

aplicables per a qualsevol individu o organització; implementar les decisions, seguiment i retroalimentació i alineament organitzacional.

Implementar decisions implica com no pot ser d'una altra forma, traduir les troballes de l'anàlisi i prendre una sèrie de mesures i accions per a traslladar-les a les activitats individuals o organitzacionals. Això pot implicar des de llançar nous serveis o productes fins a ajustar processos operatius o reassignar recursos en funció dels coneixements extrets de l'anàlisi.

La segona pauta és la realització d'un seguiment continu i exhaustiu de les estratègies aplicades derivades de l'anàlisi i una retroalimentació o feedback que permeti a les parts compartir informació entre elles per a periòdicament alinear els seus enfocaments. Els indicadors clau de rendiment (KPI) i les mètriques exerceixen un paper vital en l'avaluació de l'impacte de les accions preses durant aquesta fase, facilitant en gran manera aquesta tasca.

En últim lloc, encara que potser el més important, és l'alineament organitzacional, sinònim de crear una cultura basada en les dades en la qual qualsevol persona dins d'una organització estigui capacitada per a aprofitar els *insights* de qualsevol anàlisi i traduir-los en decisions *data-driven*. La implementació exitosa d'estratègies i mesures basades en dades passa per la correcta alineació dels objectius inicials i les iniciatives de l'anàlisi (Davenport, 2006).

En adoptar aquesta sèrie de mesures i activitats, els individus i organitzacions poden aprofitar el poder de l'anàlisi de dades per a innovar, optimitzar les operacions i prosperar en un panorama empresarial cada vegada més competitiu.

“Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.” – Thomas H. Davenport, 2006.

2.1.14. El cicle de vida de les dades

Comprendre els matisos i diferències entre les fases de l'anàlisi de dades i el cicle de vida de les dades és també important per comprendre els diferents processos i tractaments de les dades des d'una perspectiva holística de gestionar i fer un ús eficaç de les dades.

Si bé les fases del procés d'anàlisi de dades se centren principalment en la recol·lecció, exploració i processament d'aquestes per una posterior anàlisi per obtenir coneixements, el cicle

de vida de les dades comprèn una sèrie d'activitats de manera més general que governen tota la seva trajectòria des del moment que obtenim les dades fins que ens desfem d'elles.

També conegut com a *Data lifecycle management (DLM)*, la multinacional tecnològica estatunidenca IBM proposa aquest enfocament per gestionar les dades al llarg del seu cicle de vida, des de l'entrada de les dades fins a la seva destrucció. Un bon procés de DLM aporta estructura i organització a les dades d'una empresa, la qual cosa permet assolir objectius clau com la seguretat i la disponibilitat de les dades. Això es tradueix en una major confiança en la informació, una presa de decisions més informada i un avantatge competitiu en el mercat (IBM, 2024).

De la mateixa manera que el procés de l'anàlisi de dades, el cicle de vida de dades també té unes fases estandarditzades tot i que poden variar en funció dels objectius empresarials o d'investigació.

Un dels cicles més estàndard proposat per Kristin A. Briney a la seva obra *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success*, consta de diverses etapes (Creació, Processament i Emmagatzematge, Ús, Arxivament, Destrucció), totes interconnectades i cadascuna crucial per garantir la integritat, l'accessibilitat i la usabilitat de les dades per a la investigació, l'anàlisi i la presa de decisions (Briney, 2015).

La primera fase del cicle de vida de les dades és la **Creació** de dades. Aquesta fase implica l'adquisició o generació de dades en brut a través de diversos mitjans com observacions i enquestes. Es destaca la importància de capturar metadades precises i detallades durant el procés de creació de dades, donat que inclouen informació sobre les dades com la seva font, format i qualitat, facilitant l'organització i recuperació de les dades en etapes posteriors.

Un cop creades les dades, són sotmeses a un **Processament i Emmagatzematge** per millorar-ne la qualitat i l'accessibilitat. El processament de dades implica tasques com la neteja, la transformació i l'anàlisi per extreure informació significativa i preparar les dades per al consum. S'empren pràctiques sòlides de gestió de dades durant el processament, inclòs el control de versions i la documentació per fer un seguiment dels canvis i assegurar la reproductibilitat. L'emmagatzematge és un altre aspecte crític d'aquesta fase, on les dades s'emmagatzemen en repositoris o bases de dades per al seu ús futur, fent ús de solucions d'emmagatzematge

apropiades basades en factors com el volum de dades, els requisits d'accés i les consideracions de seguretat.

La fase sota el nom **Ús** de les dades, implica l'exploració i la utilització d'aquestes per a diversos propòsits, inclosa la recerca, la presa de decisions o la innovació. Briney destaca la importància de promoure l'intercanvi de dades i la col·laboració per maximitzar el valor derivat dels actius de dades. S'encoratja els investigadors a adherir-se a les directrius ètiques i legals que regeixen l'ús de les dades, inclosa l'obtenció del consentiment per a l'intercanvi de dades i el respecte dels drets de propietat intel·lectual.

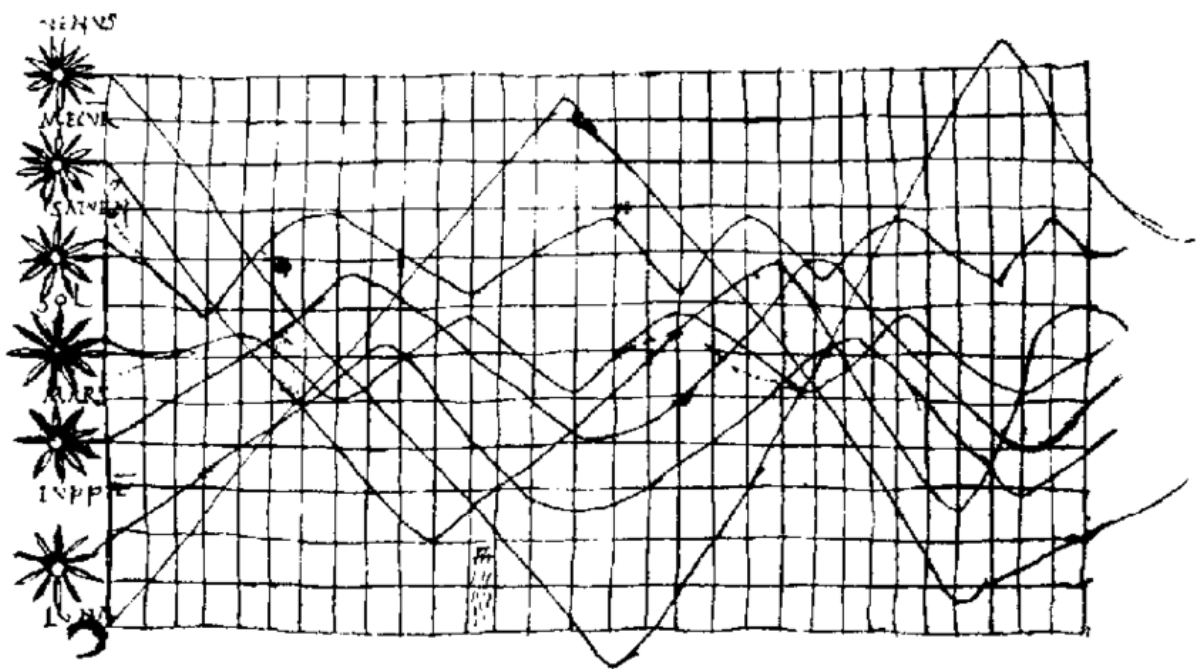
A mesura que les dades envelleixen o s'accedeix amb menys freqüència, es passa a la fase d'**Arxivament**. L'arxiu implica emmagatzemar dades en repositoris o arxius a llarg termini per preservar la seva integritat i accessibilitat al llarg del temps. Se subratlla la necessitat d'establir polítiques i procediments clars per a l'arxivat de dades, inclosos els períodes de retenció i els controls d'accés. L'arxivat eficaç assegura que les dades valuoses romanen disponibles per a futures referències o anàlisis, alhora que optimitza els recursos d'emmagatzematge.

La fase final del cicle de vida de les dades és la seva **Destrucció**. Com el seu nom indica, un cop les dades que ja no són necessàries o pertinents, s'eliminen de manera segura i irreversible dels sistemes d'emmagatzematge. Aquesta fase és crucial pel compliment de polítiques legals i reguladores, per la gestió pertinent de diversos riscos de seguretat, alhora que per l'optimització de recursos d'emmagatzematge. Mètodes segurs de destrucció de dades com ara l'encriptació, el triturat o la denegació ajuden a evitar l'accés no autoritzat o vulneracions de la privacitat de les persones.

2.2. Teoria de la visualització de dades

Com ja s'ha vist abans, la visualització de dades es va originar per la necessitat de compartir dades de manera visual i clara, mitjançant suports visuals com gràfics, mapes i pictogrames. Aquestes visualitzacions permetien compartir informació i poder prendre millors decisions en tenir la facilitat de detectar patrons o tendències amb un simple cop d'ull.

Una de les primeres visualitzacions de dades de la que es té constància apareix en un apèndix del segle X als comentaris d'A.T. Macrobius a *In Somnium Scipionis* de Cicero és la representació dels moviments planetaris mostrats com a inclinacions cícliques al llarg d'un període de temps. En un període marcat per avenços en camps tan complexos com l'astronomia i les matemàtiques, va sorgir la necessitat de deixar constància i estudiar els moviments celestes i els seus patrons mitjançant la visualització d'aquestes dades estudiades.



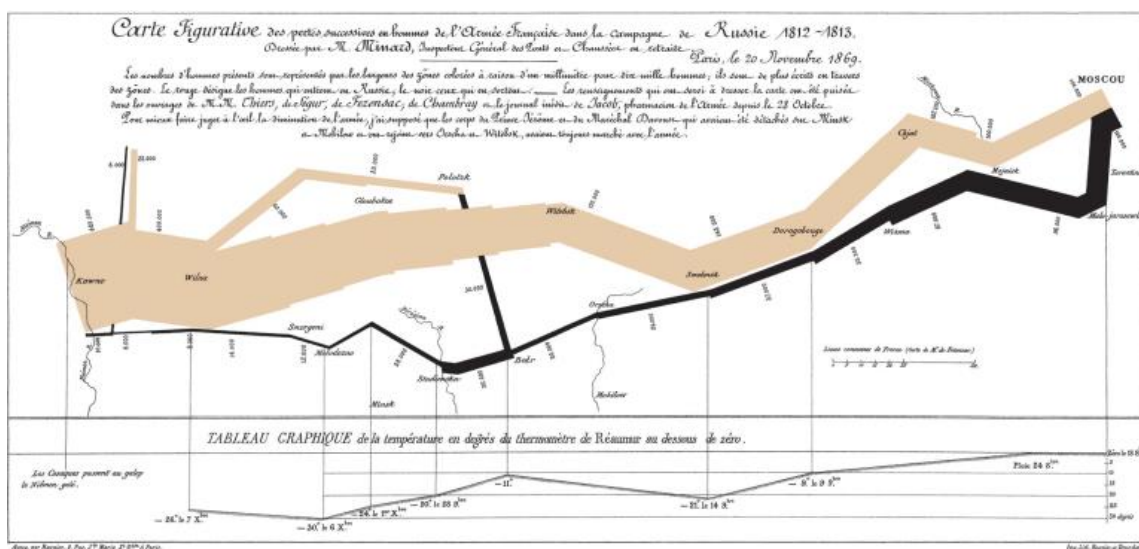
Il·lustració 2: "Planetary movements shown as cyclic inclinations over time, by an unknown astronomer, appearing in a 10th century appendix to commentaries by A. T. Macrobius on Cicero's *In Somnium Scipionis*", Font: Funkhouser, H. G., 1937.

Els primers mapes no van trigar a arribar. Tot i que les primeres representacions geogràfiques utilitzaven suports com la pedra, les primeres obres cartogràfiques sobre paper o tela daten d'entre els segles XII i XV. Un bon exemple és el mapa *Da Ming Hunyi Tu*, un mapa xinès del segle XV que representa el món conegut en aquest concret moment. Aquest mostra tant territoris xinesos com a regions estrangeres i és notable pel seu nivell de detall i precisió.



Il·lustració 3: “The Da Ming Hunyi Tu”, Font: Evan Andrews, 2023.

Durant el segle XIX, impulsat pels avenços en el disseny i la tècnica, es va produir un augment notable en el desenvolupament de la gràfica moderna, especialment en els camps de la gràfica estadística i la cartografia. Aquesta època va veure l'aparició de figures influents com William Playfair, John Snow i Charles Minard, que són àmpliament reconeguts com a pioners en el camp de la visualització de dades. En particular, Minard va ser l'autor de la icònica visualització *Napoleon's Russian campaign of 1812*, elogiada per Edward Tufte com el millor gràfic estadístic de la història. El mateix Tufte és conegut per la seva obra fonamental, *The Visual Display of Quantitative Information*, considerada una referència definitiva en l'estudi de la ciència estadística.



Il·lustració 4: “Charles Minard’s Flow Map of Napoleon’s Russian Campaign of 1812”, Font: Michael Sandberg, 2013.

Durant l'últim quart del segle XX, els pioners de la visualització de dades van establir les bases per al desenvolupament d'aquest camp. Aquest creixement va ser facilitat per la disponibilitat de grans volums de dades en l'era digital i l'àmplia gamma d'eines tecnològiques i programari informàtic disponibles per a la investigació en diverses disciplines. La visualització de dades ha experimentat un ràpid desenvolupament en múltiples camps a causa de la seva aplicabilitat generalitzada.

Els fonaments establerts pels pioners de la visualització de dades van possibilitar l'aparició del camp de la visualització de dades com la coneixem avui. Durant la segona meitat del segle XX es va establir una maduresa, alimentada també per l'abundant disponibilitat de dades en l'era digital i l'extens repertori d'eines tecnològiques i programari informàtic que faciliten la investigació a través de diverses disciplines. El ràpid desenvolupament de la visualització de dades s'ha produït a un ritme notable, abastant múltiples camps d'aplicació a causa de la seva versatilitat i utilitat.

2.2.1. Semiologia gràfica

En la seva obra *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*, el cartògraf francès Jaques Bertin va establir una base fonamental per al disseny efectiu de visualitzacions de dades en introduir el concepte de "variables visuals", obrint i explorant un nou camp de coneixement.

Aquestes variables representen les propietats perceptibles dels elements visuals que s'utilitzen per a codificar la informació en una visualització, que inclouen posició, grandària, forma, valor, color i orientació.

La **posició** permet situar dades en un espai bidimensional o tridimensional, la **grandària** facilita la representació de magnituds numèriques, la **forma** ajuda a distingir entre diferents tipus de dades, el **valor** de lluentor pot utilitzar-se en visualitzacions en blanc i negre per a representar magnituds, el **color** ofereix una codificació rica i expressiva, i l'orientació ajuda a transmetre informació sobre la relació espacial entre els elements (Bertin, 1967).

Cadascuna d'aquestes variables proporciona una eina poderosa per a representar dades de manera clara i comprensible, permetent la creació de visualitzacions expressives i efectives.

Respecte al títol de l'obra de Bertin, la semiologia és l'estudi dels signes i símbols en la vida social, així com de la seva interpretació i significat en diferents contextos. I en l'àmbit de la visualització de dades, la semiòtica juga un paper fonamental en analitzar com els elements visuals utilitzats actuen com a signes que representen informació.

Encara que tots dos camps comparteixen l'interès pels signes i el seu significat, la semiologia té un enfocament més limitat i específic dins de la lingüística i la cultura, mentre que la semiòtica comprèn un camp més ampli i divers d'estudis sobre els signes en general. Per tant, entendre la semiòtica és crucial per a dissenyar visualitzacions efectives que comuniquin clarament la informació continguda en les dades.

Aquest enfocament de Bertin en les variables visuals ofereix una guia sòlida per al disseny de visualitzacions que faciliten la interpretació de la informació i ressalten patrons i relacions importants en les dades.

2.2.2. Percepció gràfica

Anys després de les aproximacions de Bertin dins el camp de la semiologia respecte a les visualitzacions de dades, els estadístics nord-americans Cleveland i McGill van anar més enllà del concepte de les variables visuals, argumentant que els elements gràfics en una visualització no eren suficients per comprendre la informació codificada, que aquesta havia d'anar acompanyada de la percepció de l'espectador.

A l'article *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods* explicaren que la metodologia per l'anàlisi i visualització de dades necessitava una base científica sòlida, encunyant el terme percepció gràfica. El seu enfocament se centra en la capacitat humana per a percebre i comprendre visualment la informació presentada, establint una teoria basada en experimentacions que permetia establir un seguit de principis per a la descodificació visual eficaç d'informació en gràfics.

Van proposar deu tasques perceptives elementals, cadascuna referint-se a un mètode de representació de dades i ordenades segons la precisió amb què el cervell humà pot detectar diferències i fer comparacions entre elles. Aquestes són; posició sobre un eix comú, posició sobre eixos no-alineats, longitud, direcció, angle, àrea, volum, curvatura, ombreig i saturació de color (Cleveland & McGill, 1984).

Aquests principis, basats en la percepció visual i la psicologia cognitiva, han servit de model per diferents estudis dins d'aquest àmbit i serveixen com a pedra angular per al disseny de gràfics clars i efectius en el camp de la informació i la visualització de dades.

2.2.3. Principals visualitzacions de dades

Escollir el tipus de visualització de dades més adient pot no ser feina fàcil. Començant tenint en compte que el tipus de visualització que s'utilitzarà depèn principalment de la naturalesa de les dades i la informació que es vol comunicar és el primer pas. La primera gran distinció es fa a partir de l'objectiu en si de la visualització, diferenciant les visualitzacions explicatives de les exploratòries.

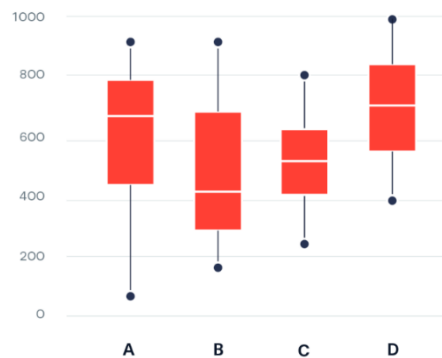
Les visualitzacions explicatives estan pensades i dissenyades per explicar qualsevol mena d'història i transmetre un missatge o una idea clara i concisa una audiència determinada. Aquestes visualitzacions sovint requereixen un alt grau de precisió i detall en la seva elaboració, ja que estan destinades a recolzar arguments, presentar troballes de recerca o comunicar resultats. És ideal que incloguin elements visuals que incentivin l'interès de l'espectador i ajudin a il·lustrar i donar suport a la narrativa central. A més, solen ser estàtiques o tenen una interactivitat limitada, perquè el seu objectiu principal és presentar informació de manera clara i efectiva.

En canvi, les visualitzacions exploratòries funcionen com a eines de recerca visual. Estan dissenyades per permetre als usuaris explorar i analitzar dades de manera interactiva, descobrint patrons, tendències o relacions ocultes que poden no ser evidents a primera vista. Aquestes visualitzacions solen ser dinàmiques i flexibles perquè els usuaris puguin interactuar amb la visualització, ajustant paràmetres i explorant diferents perspectives per generar hipòtesis, validar suposicions o extreure'n *insights*.

D'altra banda, Alberto Cairo va proposar en la seva obra *The Functional Art: An Introduction to Information Graphics and Visualization* quatre pilars fonamentals que guien la creació de gràfics efectius, abastant una àmplia varietat de contextos i necessitats analítiques, oferint a dissenyadors i analistes les eines necessàries per a comunicar de manera clara i efectiva la informació continguda en les dades. Aquests 4 pilars són: distribució, relació, composició i comparació (Cairo, 2012).

Distribució d'una sola variable

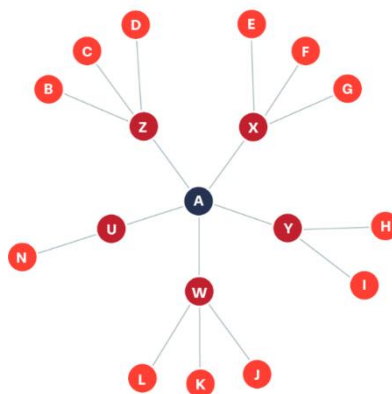
Els gràfics de distribució tenen com a objectiu representar com es distribueixen els punts de dades en diferents rangs o categories, per obtenir informació sobre la seva tendència central, variabilitat i potencials valors atípics. Els histogrames, els diagrames de caixa i els diagrames de densitat són eines habituals que s'utilitzen per visualitzar distribucions.



Il·lustració 5: "Exemple d'un gràfic de distribució", Font: Ferdio, 2014.

Relació entre dues variables

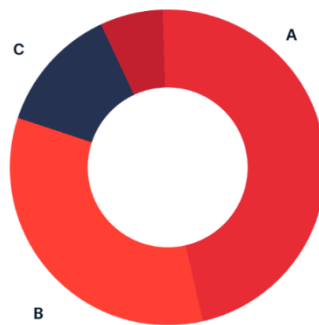
Les visualitzacions d'aquesta categoria ajuden a descobrir patrons, tendències i dependències entre diferents atributs de dades en visualitzar la correlació o associació entre variables d'un conjunt de dades. Identificar aquestes relacions pot ser útil per entendre com els canvis en una variable poden afectar una altra. Els diagrames de dispersió, els gràfics de línies i les matrius de correlació són exemples de les visualitzacions de relació més comuns.



Il·lustració 6: "Exemple d'un gràfic de relació", Font: Ferdio, 2014.

Composició d'una o més variables

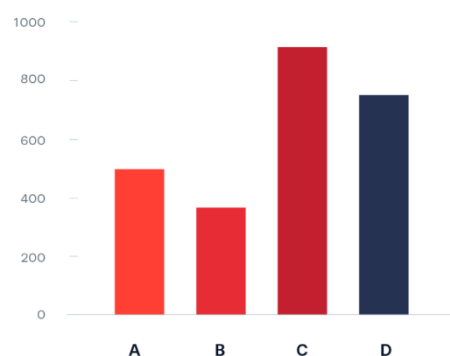
Aquest grup de visualitzacions s'encarrega de desglossar les dades complexes en les seves parts facilitant la comprensió i la comparació d'elements individuals i la seva proporció dins d'un conjunt de dades. Els gràfics circulars, els gràfics de barres apilades i els mapes d'arbre són eines habituals per visualitzar la composició d'una o més variables.



Il·lustració 7: "Exemple d'un gràfic de composició", Font: Ferdio, 2014.

Comparació entre diferents variables

L'últim pilar consisteix a visualitzar diferències o similituds entre diversos conjunts de dades, grups o categories, permetent als usuaris comparar valors, distribucions o tendències entre diferents entitats. En comparar variables es poden identificar de manera més senzilla les disparitats, les tendències i els coneixements que poden no ser evidents quan s'examinen conjunts de dades individuals de manera aïllada. Els gràfics de barres, els gràfics de línies i els diagrames de quadres junts són efectius per comparar dades.



Il·lustració 8: "Exemple d'un gràfic de comparació", Font: Ferdio, 2014.

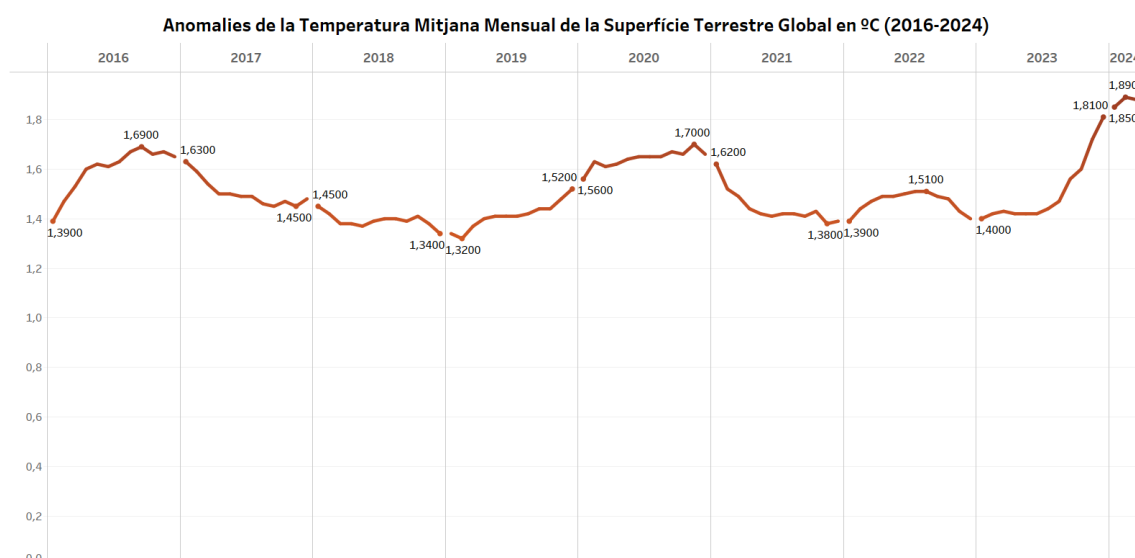
2.2.4. Errors comuns i com evitar-los

En l'àmbit de la visualització de dades, la creació de recursos visuals clars i reveladors dins les visualitzacions és essencial per a una comunicació eficaç. No obstant això, és necessari tenir en compte una sèrie d'errors habituals que poden perjudicar la claredat i la precisió la informació que es vol transmetre, provocant malinterpretacions i confusions. Aquests són els errors més comuns que habitualment es troben a l'hora de visualitzar dades:

Ocultar o ometre dades rellevants

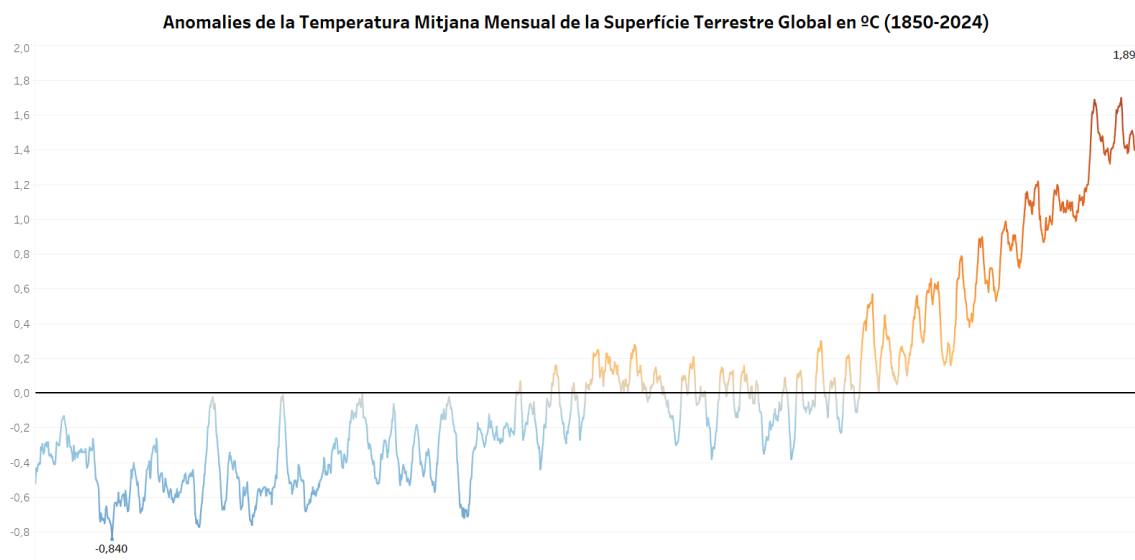
Un dels errors més greus en la visualització de dades és presentar de manera selectiva només certes parts de les dades mentre se n'ometen d'altres. Hi ha una tècnica àmpliament coneguda dins la comunitat científica anomenada *Cherry Picking*, que no és més que la pràctica de triar o destacar de manera selectiva només certes peces d'informació o dades que donen suport a un punt de vista determinat, ignorant o minimitzant la resta d'informació rellevant que la pugui contradir. Això, òbviament, condueix a interpretacions esbiaixades i conclusions errònies.

Considerem dos gràfics que representen variacions de temperatura registrades mensualment al llarg del temps. En el primer, només es mostren les dades des de l'any 2016, cosa que podria suggerir que les fluctuacions de temperatura són mínimes, fins i tot implicant una disminució de les temperatures a causa de l'anomalia d'un sol any.



Il·lustració 9: “Anomalies de la Temperatura Mitjana Mensual de la Superfície Terrestre Global en °C (2016-2024)”, Font: Elaboració Pròpia a partir de dades extretes de National Centers for Environmental Information (2024).

Ara bé, com veiem en aquest segon gràfic en el qual s'inclou el conjunt de dades complet mes a mes des de l'any 1850, la imatge és diferent. Amb més context, es revela una tendència significativa a l'alça de les anomalies de temperatura al llarg del temps.



Il·lustració 10: “Anomalies de la Temperatura Mitjana Mensual de la Superfície Terrestre Global en °C (1850-2024)”, Font: Elaboració Pròpia a partir de dades extretes de National Centers for Environmental Information (2024).

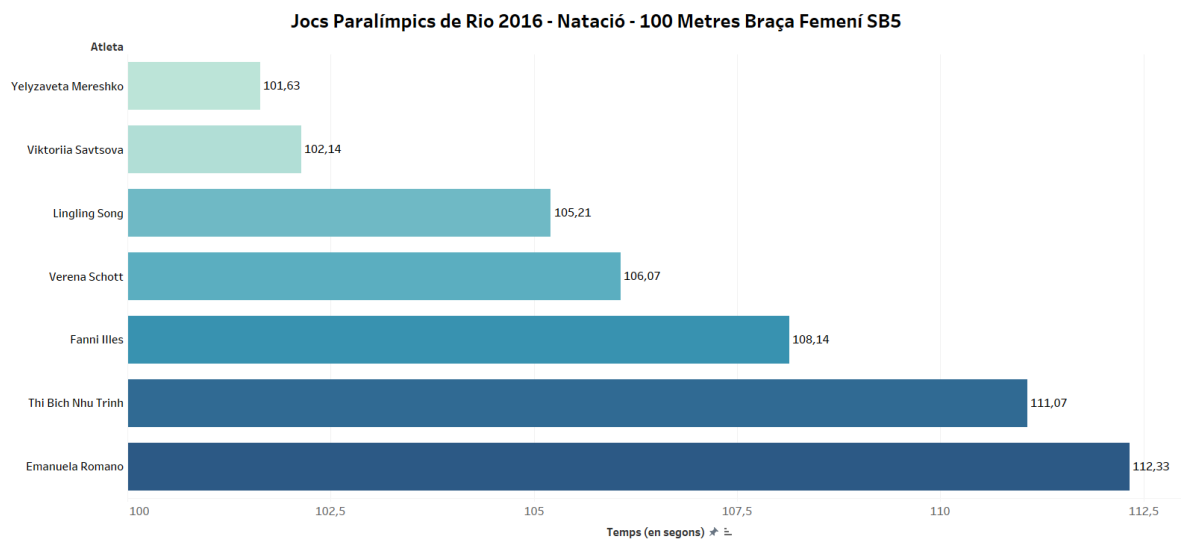
El fet de ressaltar selectivament els punts de dades favorables condueix a interpretacions esbiaixades i conclusions errònies, perjudicant la transparència i la integritat de la presentació de les dades. Aquesta transparència i honestat en presentar dades o informació és primordial i s'ha de garantir sempre, fent que aquest error es pugui evitar fàcilment presentant totes les dades rellevants i indicant clarament les exclusions o limitacions d'aquestes.

Representació enganyosa de les dades

Es poden produir representacions enganyoses quan es manipulen els eixos, les escales es distorsionen o les proporcions es tergiversen. Per exemple, alterar l'interval o el punt inicial d'un eix pot fer l'efecte de diferències més grans entre els punts de dades de les que realment existeixen, mentre que distorsionar l'escala o tergiversar les proporcions pot exagerar o disminuir la importància de determinades dades, donant lloc a interpretacions inexactes.

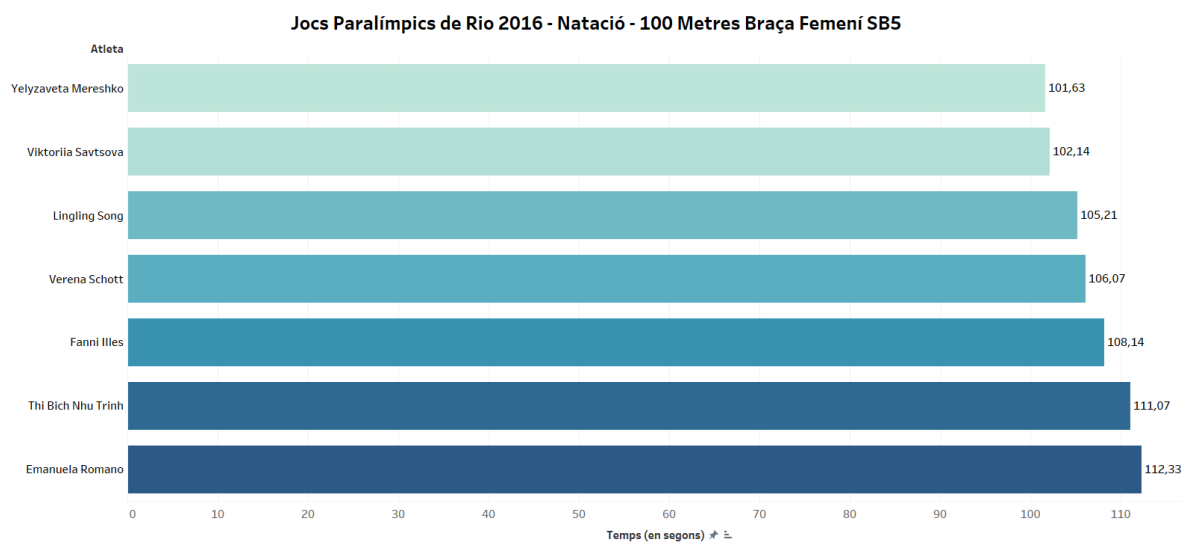
Per il·lustrar com l'elecció de l'escala pot influir en la percepció de les dades, considerem dos gràfics que representen el mateix conjunt de dades de 7 nadadores paralímpiques i els temps assolits per cadascuna a les olimpíades de Rio de Janeiro l'any 2016 en la modalitat de 100 m braça. Com il·lustra el primer gràfic, l'atleta Yelyzaveta Mereshko va guanyar la cursa en

finalitzar-la en 101,63 segons, a diferència de l'atleta Emanuela Romano que va marcar 112 segons.



Il·lustració 11: “Jocs Paralímpics de Rio 2016 – Natació – 100 Metres Braça Femení SB5”, Font: Elaboració Pròpia a partir de dades extretes de International Paralympic Committee, Women’s 100m breaststroke SB5, Rio 2016 (2024).

A primera vista, pot semblar que la guanyadora va ser sis cops més ràpida que l’última en acabar, però això no és més que el fruit de l’alteració visual que produeix que l’escala comenci pel valor 100. Simplement canviant un petit detall com és que l’escala comenci pel valor 0, es fa visible que no hi ha una diferència tan significativa entre nadadores, i s’observa com els mateixos valors es veuen molt més junts.



Il·lustració 12: “Jocs Paralímpics de Rio 2016 – Natació – 100 Metres Braça Femení SB5”, Font: Elaboració Pròpia a partir de dades extretes de International Paralympic Committee, Women’s 100m breaststroke SB5, Rio 2016 (2024).

Això es deu al fet que en establir un eix que no comenci per zero fa que les diferències entre els altres valors augmentin, de manera que en assegurar que l'eix comença a 0, s'assegura que les longituds de les barres són proporcionals al seu valor i que les diferències entre els valors no són tan grans.

Triar la forma per sobre de la funció

Escollir el gràfic adequat per a visualitzar un conjunt de dades pot ser feina difícil, especialment quan diversos tipus de gràfics són visualment atractius, però com hem vist abans, hi ha diferents grups de gràfics que s'utilitzen per a funcions específiques i concretes.

En cas que vulguem comparar ràpidament la magnitud de diferents categories, escolliríem un gràfic de barres simple. De la mateixa forma que si volem visualitzar la proporció de diferents valors dins d'un mateix conjunt de dades, l'opció més adequada seria la tria d'un diagrama de sectors o circular. I si el que es busca és visualitzar tendències o patrons al llarg d'un període de temps, l'ideal seria visualitzar-ho mitjançant un gràfic de línies.

La funcionalitat d'una visualització de dades ha de prevaldre sempre sobre de la forma i l'estètica, i el primer objectiu hauria de ser visualitzar les dades en el format més adient, no el més atractiu visualment. Preguntar-se: "Què estic intentant mostrar a l'audiència?" hauria d'ajudar a triar el gràfic adequat i comunicar de manera efectiva les idees extretes.

Sobrecarregar les visualitzacions d'informació innecessària

Tot i que és important presentar informació amb cert grau de complexitat, un error habitual és sobrecarregar i saturar les visualitzacions de dades amb informació o dades excessives o fins i tot innecessàries. Això pot fer que les dades que són realment rellevants o els *insights* extrets de la visualització costin més d'arribar a l'espectador o que es perdin per complet. El missatge que es vol transmetre mitjançant un gràfic ha de poder ser desxifrat en poc temps i de manera clara, i per evitar que passi el contrari, es poden fer servir tècniques com l'agregació, el filtratge o l'agrupació per reduir la complexitat sense sacrificar el significat.

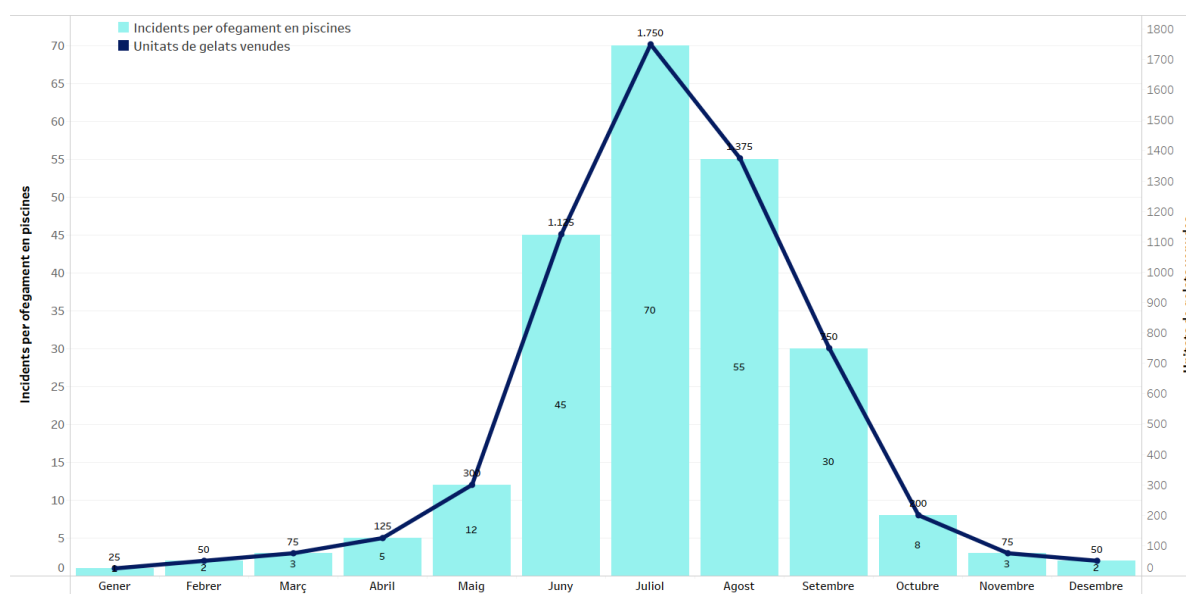
En cas que molta informació hagi de ser presentada i per evitar aquesta saturació en un sol gràfic, una bona solució és recórrer a la utilització de *dashboards* que permetin presentar múltiples gràfics en un mateix suport. A més, en cas que sigui possible, un recurs útil són les

visualitzacions interactives, que poden visualitzar diferents gràfics en un breu període de temps per mostrar per exemple tendències al llarg del temps.

Confondre correlació i causalitat

L'últim error contemplat i possiblement el més freqüent alhora que el més important a tenir en compte és confondre correlació i causalitat. Una correlació no implica una causalitat, però aquest principi fonamental sovint es passa per alt en l'anàlisi i visualització de dades. Aquest error de confondre correlació amb causalitat es produeix quan hi ha un malentès entre dos conceptes estadístics: la correlació, que mesura el grau d'associació entre dues variables, i la causalitat, que implica una relació causa-efecte entre elles. Aquesta confusió sovint porta a interpretacions enganyoses, cosa que suggereix una relació entre variables quan realment no n'hi ha cap.

Per exemple, si considerem la correlació entre les vendes de gelats i els incidents d'ofegament en piscines, entenem que tots dos poden assolir el seu màxim durant els mesos d'estiu, donant lloc a una correlació. No obstant això, és completament incorrecte concloure que consumir gelats provoca ofegaments o viceversa, és del tot necessari i primordial evitar fer afirmacions causals injustificades basades únicament en correlacions observades en una visualització de dades.



*Il·lustració 13: "Representació fictícia dels incidents per ofegament en piscines i les unitats de gelats venudes en un any",
Font: Elaboració Pròpia, 2024.*

Per evitar aquest freqüent error és crucial analitzar a fons les dades i considerar explicacions alternatives, proves addicionals, així com dur a terme anàlisis estadístiques rigoroses i tenir precaució a l'hora d'interpretar les relacions entre variables.

Tots aquests errors o manipulacions poden ser no intencionades o deliberades, però independentment de la intenció, perjudiquen la integritat de la visualització i la fiabilitat en la informació presentada. És essencial mantenir la fidelitat a les dades i garantir que les representacions visuals reflecteixin amb precisió la informació subjacent sense biaix ni distorsió (Treesak, 2024).

2.2.5. El mètode McCandless

Entre la diversitat d'enfocaments per visualitzar dades, un mètode que destaca és el mètode McCandless, popularitzat pel periodista i pioner de la visualització de dades David McCandless. Aquest mètode és un enfocament per crear visualitzacions de dades que prioritzen la claredat, la simplicitat i la comunicació eficaç d'informació complexa, posant l'accent en la transformació de dades en brut en representacions visualment atractives que siguin fàcils d'entendre i atractives per al públic. Subratlla quatre elements clau dins una visualització de dades:

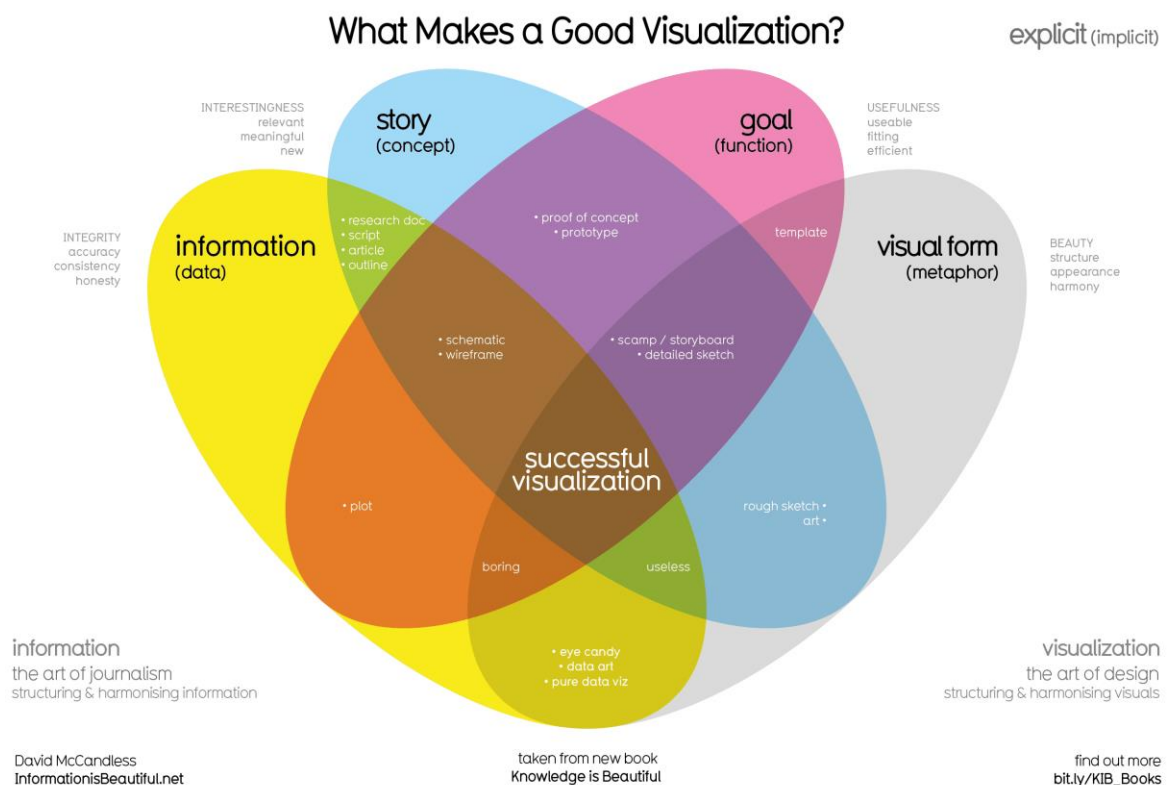
Dades: al nucli de qualsevol visualització es troben les dades en si, que serveixen com a matèria primera de la qual s'obtenen els coneixements i es teixeixen les narracions. L'estructuració i l'organització de les dades estableixen les bases per als esforços de visualització posteriors, guiant la trajectòria cap a visuals significatius i informatius.

Història: més enllà dels números i xifres, la narració o la història dona vida a les dades, donant context i significat. En establir els objectius els patrons i les conclusions descobertes durant l'exploració de les dades, la història dona a la visualització una coherència i un propòsit, incentivant la immersió de l'audiència.

Objectiu: la raó de ser d'una visualització està en la seva capacitat per assolir un objectiu o una funció específica, sigui per informar, persuadir o inspirar una acció concreta. Aclarir el propòsit previst de la visualització garanteix l'alineació amb les necessitats i els objectius del públic, fomentant la utilitat i la usabilitat en la presentació d'informació basada en dades.

Forma visual: la forma visual serveix per impactar visualment al públic i posar un marc sobre el qual es desenvolupa la narrativa de dades, sigui en forma de gràfics, infografies o pantalles interactives. L'elecció d'elements visuals configura la percepció i la comprensió de l'espectador de la informació presentada, combinant estètica amb funcionalitat.

La sinergia d'aquests quatre elements constitueix l'essència del mètode McCandless, que amb un enfocament holístic fa que la visualització de dades transcendeixi de la mera representació visual a un pla més complex en el que es busca evocar una sèrie de reaccions en l'espectador. A una era on abunda la informació i escasseja el temps d'atenció, la capacitat d'elaborar històries de dades convincents i concises és alhora un privilegi i una responsabilitat (McCandless, 2014).



Il·lustració 14: "The McCandless Method – What makes a Good Visualization?", Font: David McCandless, 2024.

2.3. Glossari de termes especialitzats

.csv: Format de fitxer de text amb valors separats per comes, utilitzat àmpliament per l'intercanvi de dades tabulars entre diferents aplicacions i plataformes.

.tsv: Format de fitxer de text amb valors separats per tabulacions, utilitzat per emmagatzemar dades tabulars de manera simple i compatible amb diversos sistemes.

.txt: Format de fitxer de text pla que conté dades sense cap format especial, útil per a l'emmagatzematge de text senzill i dades sense estructures complexes.

.xlsx: Format de fitxer de Microsoft Excel que permet emmagatzemar dades en fulls de càlcul amb capacitats avançades de format, càlcul i gràfics.

Algorismes: Seqüències de passos lògics i matemàtics que permeten la resolució de problemes o l'execució de tasques específiques en un procés computacional automatitzat.

Anàlisi qualitativa: Procés d'examinar dades no numèriques per comprendre fenòmens, comportaments i experiències a través de la interpretació de contingut textual o visual.

Anàlisi quantitativa: Procés d'examinar dades numèriques per identificar patrons, tendències i relacions mitjançant mètodes estadístics i matemàtics.

Baby-Boom: Període d'alta natalitat que es va produir després de la Segona Guerra Mundial caracteritzat per un augment significatiu del nombre de naixements.

Biaix: Tendència o preferència sistemàtica que distorsiona la interpretació o anàlisi de dades, compromentent la imparcialitat i objectivitat dels resultats.

Cohorts: Subgrups d'individus que comparteixen una característica o experiència comuna en un període determinat, utilitzats sovint per a l'anàlisi demogràfica en les piràmides de població.

Dades crues: Informació en el seu format original i sense processar, recollida directament de la font abans de ser analitzada o transformada.

Dades de primera font: Informació recollida directament per l'investigador o l'organització de la font original sense intermediaris, sovint gratuïtes i de caràcter públic.

Dades de tercera font: Informació recopilada i proporcionada per entitats diferents de la font original, sovint de pagament o de caràcter privat.

Dades estructurades: Informació organitzada en un format predefinit, com ara taules amb files i columnes, que facilita la cerca i l'anàlisi automàtica.

Dades qualitatives: Informació no numèrica que descriu qualitats o característiques, sovint recollida a través d'entrevistes, observacions o qüestionaris oberts.

Dades quantitatives: Informació numèrica que es pot mesurar i analitzar estadísticament, sovint recollida a través d'enquestes, experiments o registres administratius.

Dashboard: Panell interactiu que agrega i visualitza dades clau en un únic lloc de manera integrada, facilitant el seguiment de mètriques i indicadors de rendiment, així com la seva presentació.

Data-driven: Enfocament en la presa de decisions i estratègies basades en l'anàlisi de dades i evidències objectives en lloc de la intuïció o experiència personal.

Datasets: Conjunts estructurats de dades relacionades, sovint emmagatzemats en taules o bases de dades, utilitzats per a l'anàlisi i la investigació.

Ètica de dades: Conjunt de principis i pràctiques que guien l'ús responsable i respectuós de les dades, protegint la privacitat, la seguretat i els drets dels individus i col·lectius.

Hiperconnexió: Estat de connexió constant a les xarxes digitals a través de dispositius electrònics, afectant la manera com les persones interactuen i accedeixen a la informació.

Insights: Comprensions profundes derivades de l'anàlisi de dades que ofereixen una visió valuosa per a la presa de decisions.

Integritat de dades: Estat de les dades que assegura la seva precisió, coherència i fiabilitat al llarg del seu cicle de vida, evitant errors i corrupcions durant el seu intercanvi.

Metadades: Dades que descriuen altres dades, proporcionant informació sobre el context, la qualitat, la condició i altres aspectes rellevants dels conjunts de dades.

Mètriques: Indicadors quantificables utilitzats per avaluar el rendiment, l'eficàcia o altres aspectes d'un procés, projecte o sistema.

Repositori de dades: Espai centralitzat on es recullen, emmagatzemen i gestionen conjunts de dades, facilitant l'accés, la compartició i la reutilització d'informació.

Stakeholders: Persones o grups amb un interès directe en els resultats d'un projecte, organització o empresa, i que poden influir o ser influenciats per les seves decisions.

Tableau: Software de visualització de dades que permet crear gràfics interactius i dashboards per a l'anàlisi i la presentació de dades.

Tendència: Patró o direcció general en què es mou una sèrie de dades o esdeveniments al llarg del temps, sovint identificat a través de l'anàlisi estadística.

Treemap: Visualització jeràrquica de dades en forma de rectangles aniuats, on l'àrea de cada rectangle és proporcional a una dimensió de les dades, utilitzada per mostrar proporcions dins d'una jerarquia.

Visualitzacions explicatives: Representacions gràfiques de dades dissenyades per comunicar troballes específiques de manera clara i concisa al públic destinatari.

Visualitzacions exploratòries: Representacions gràfiques de dades utilitzades durant l'anàlisi de dades per identificar patrons, tendències i anomalies sense un objectiu predeterminat.

3. Metodologia

3.1. Disseny de la investigació

El disseny d'aquesta recerca es basa en una sèrie de passos metodològics estructurats per complir amb els objectius establerts i respondre a les preguntes de recerca prèviament definides. Els passos a seguir són els següents:

Definir els objectius d'estudi i les preguntes de recerca

Els objectius de l'estudi són explorar i analitzar les tendències demogràfiques i sanitàries a Espanya, identificant patrons i relacions entre diferents indicadors. Les preguntes de recerca giren al voltant de com han evolucionat aquests indicadors i quines són les implicacions per a les polítiques de salut i benestar.

Identificar una sèrie de variables i indicadors

Es seleccionaran variables i indicadors rellevants com esperança de vida, taxa de mortalitat, distribució per edats, incidència de malalties cròniques, hàbits de salut, conductes de risc i dades sobre salut mental. Aquests indicadors estan seleccionats per intentar donar resposta a les preguntes d'investigació i permetran una avaluació exhaustiva del panorama sociodemogràfic i sanitari.

Decidir el tipus d'estudi

S'optarà per un estudi quantitatiu, utilitzant mètodes estadístics per analitzar les dades recopilades. Aquest enfocament permetrà obtenir resultats objectius i precisos, adequats per a l'anàlisi de grans conjunts de dades.

Dividir en subapartats els temes d'anàlisi

Per a una organització clara i efectiva de l'anàlisi, es dividirà l'estudi en els següents subapartats:

- Demografia i projeccions de població
- Salut i esperança de vida
- Mortalitat i problemes de salut
- Hàbits de salut i conductes de risc
- Salut mental

Buscar en portals de dades de renom i fiables

Es recopilaran dades de fonts reconegudes com Eurostat, l'Institut Nacional d'Estadística (INE) i el Ministeri de Sanitat d'Espanya entre d'altres, buscant sempre conjunts de dades confiables que estiguin alineats amb els objectius del treball.

Fer el procés de selecció i neteja de les dades

Les dades recopilades seran sotmeses a un rigorós procés de neteja i preparació. Això inclou la verificació del format de les dades, l'eliminació de valors atípics i la correcció d'errors per assegurar que les dades siguin coherents i precisos. El procés serà descarregar les dades en brut en formats .xlsx, .tsv o .csv i acabar transformant-los en formats .txt, format menys pesat i millor suportat per altres aplicacions. A més, els documents finals es prepararan de tal manera que facilitin la seva posterior obertura i integració en programari per visualitzar-les.

Analitzar les dades recopilades

Utilitzant tècniques estadístiques adequades, s'analitzaran les dades per identificar patrons i tendències. Aquesta anàlisi quantitativa proporcionarà una base sòlida per a la interpretació dels resultats.

Visualitzar les dades analitzades

S'empraran diverses tècniques de visualització de dades, barrejant visualitzacions explicatives i exploratòries. S'utilitzaran gràfics de distribució, relació, composició i comparació per representar de manera clara i efectiva les troballes de l'anàlisi.

Interpretar les visualitzacions i compartir les conclusions

Les visualitzacions generades seran interpretades per extreure conclusions significatives. Aquestes conclusions es compartiran de manera comprensible, destacant les implicacions per a les polítiques de salut i benestar social.

Aquest disseny de recerca permetrà una exploració detallada i precisa del panorama demogràfic i sanitari a Espanya, proporcionant informació valuosa per a la presa de decisions informades en l'àmbit de la salut pública.

3.2. Consideracions ètiques

Aquest estudi es durà a terme amb un ferm compromís cap a l'ètica de dades, garantint la transparència i l'honestedat en la manipulació de la informació i la responsabilitat de transmetre-la correctament, avaluant al mateix temps l'impacte social dels resultats per contribuir de manera positiva a la societat. Per això, en aquest apartat es detallen les consideracions ètiques fonamentals que guiaran tot el procés de recerca, des de la recopilació de dades fins a la interpretació i presentació dels resultats. La integritat i l'ètica són pilars essencials per assegurar la validesa i la credibilitat de l'estudi.

Transparència en la metodologia

Es garantirà que totes les etapes del procés de recerca, incloent-hi la selecció de dades, la neteja i l'anàlisi, siguin documentades de manera clara i detallada. Les fonts d'informació seran detallades minuciosament amb els seus respectius enllaços i referències per permetre la replicació de l'estudi per altres investigadors i facilitar la revisió del treball.

Honestedat i responsabilitat en la interpretació de resultats

L'honestedat en la interpretació dels resultats és crucial per mantenir la integritat científica. S'evitaran les interpretacions esbiaixades o manipulades de les dades, i es presentaran els resultats de manera objectiva i precisa. A més, es reconeixeran les limitacions de l'estudi i s'indicaran clarament els possibles marges d'error i les incerteses inherents a les dades i a l'anàlisi. Els resultats es comunicaran de manera responsable, sense exagerar les implicacions dels resultats i proporcionant un context adequat per a la seva interpretació.

Impacte social i ètic dels resultats

S'avaluarà com els resultats de l'estudi poden afectar diferents grups de la societat, sigui per diferències de gènere o edat, o per involucrar col·lectius vulnerables. Es procurarà que les conclusions i recomanacions derivades de l'estudi promoguin el benestar social i contribueixin positivament a la formulació de polítiques públiques. A més, es respectarà la privacitat i la confidencialitat de les dades utilitzades, complint amb les normatives i regulacions pertinents en matèria de protecció de dades.

3.3. Selecció de dades i fonts d'informació

Per dur a terme aquest estudi sobre la dinàmica demogràfica i sanitària a Espanya es seleccionaran fonts de dades d'alta qualitat i fiabilitat. També es busca posar-les a disposició per altres investigadors i deixar constància de la metodologia emprada en l'anàlisi. Aquestes fonts proporcionaran la base necessària per a una anàlisi exhaustiva que permetrà respondre adequadament a les preguntes de recerca plantejades i són detallades a continuació:

Font: Eurostat

Population on 1 January by age and sex

https://ec.europa.eu/eurostat/databrowser/view/demo_pjan/default/table?lang=en&category=demo.demo_pop

Population on 1st January by age, sex and type of projection

https://ec.europa.eu/eurostat/databrowser/view/proj_23np/default/table?lang=en&category=proj.proj_23n

Life expectancy by age and sex

https://ec.europa.eu/eurostat/databrowser/view/demo_mlexpec/default/table?lang=en&category=demo.demo_mor

Projected life expectancy by age (in completed years), sex and type of projection

https://ec.europa.eu/eurostat/databrowser/view/proj_23nalexp/default/table?lang=en&category=proj.proj_23n

Healthy life years by sex (from 2004 onwards)

https://ec.europa.eu/eurostat/databrowser/view/hlth_hlye/default/table?lang=en&category=hlt.h.hlth_state.hlth_hly

Smoking of tobacco products by sex, age and degree of urbanisation

https://ec.europa.eu/eurostat/databrowser/view/hlth_egis_sk1u/default/table?lang=en

Frequency of alcohol consumption by sex, age and educational attainment level

https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_all_e/default/table?lang=en&category=hlth.hlth_det.hlth_alc

Obesity rate by body mass index (BMI)

https://ec.europa.eu/eurostat/databrowser/view/sdg_02_10/default/table?lang=en

Current depressive symptoms by sex, age and income quintile

https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_mh1i/default/table?lang=en&category=hlth.hlth_state.hlth_sph

Self-reported consultation of mental healthcare or rehabilitative care professionals by sex, age and educational attainment level

https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_am6e/default/table?lang=en&category=hlth.hlth_care.hlth_consult

Font: **Instituto Nacional de Estadística (INE)**

Defunciones según la causa de muerte

<https://www.ine.es/jaxiT3/Tabla.htm?t=7947>

Font: **Ministerio de Sanidad de España**

Principales problemas crónicos de salud, porcentaje de población de 15 y más años que padece determinados problemas crónicos, registrado en atención primaria, por comunidad autónoma

<https://www.sanidad.gob.es/estadEstudios/sanidadDatos/tablas/tabla5.htm>

Font: **Indicadores Clave Sistema Nacional de Salud**

Prevalencia de consumo de tabaco en población adulta, tabaquismo

<https://inclasns.sanidad.gob.es/main.html>

Font: **Datosmacro.com**

Espanya – Mortalidad

<https://datosmacro.expansion.com/demografia/mortalidad/espana>

España – Natalidad

<https://datosmacro.expansion.com/demografia/natalidad/espana>

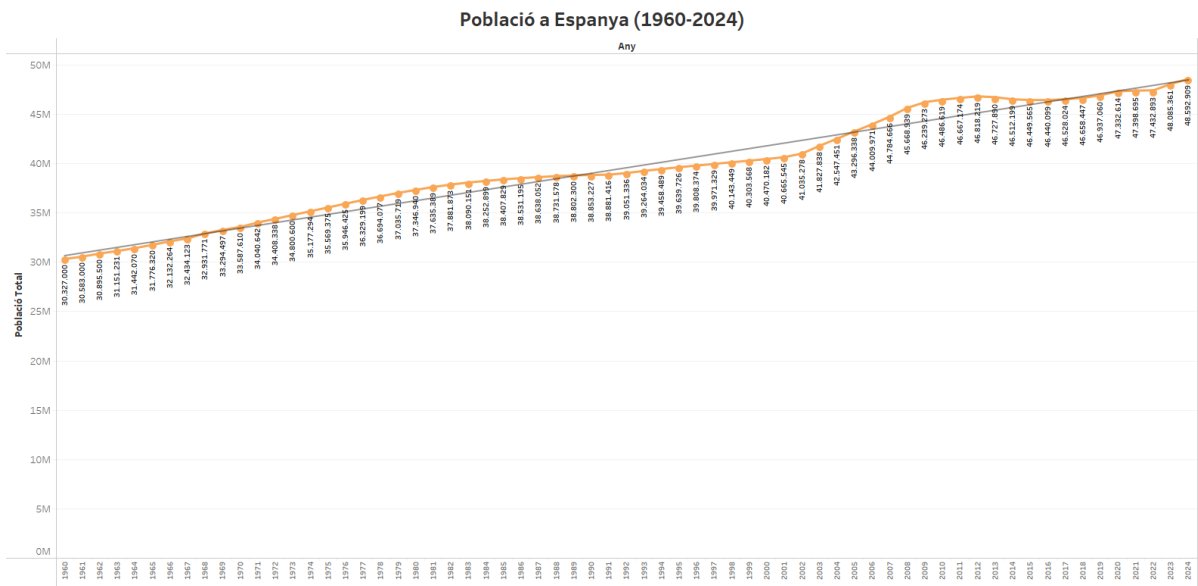
4. Anàlisi de resultats

En aquesta secció del treball es presenta una anàlisi integral de la salut demogràfica i sanitària d'Espanya a través d'una sèrie de visualitzacions de dades. La narrativa d'aquests gràfics s'estructura en cinc àrees clau: demografia i projeccions de població, salut i esperança de vida, mortalitat i problemes de salut, hàbits de salut i conductes de risc, i salut mental.

Cada subsecció proporciona una visió detallada i comparativa de les dades rellevants, destacant tendències històriques, comparacions amb altres països, projeccions futures i diferències d'edat i gènere, entre altres.

Cada gràfic d'aquesta anàlisi contribueix a una comprensió més profunda dels desafiaments actuals i emergents en la salut pública espanyola, facilitant la identificació d'àrees crítiques per a la intervenció i la formulació de polítiques efectives.

4.1. Demografia i projeccions de població



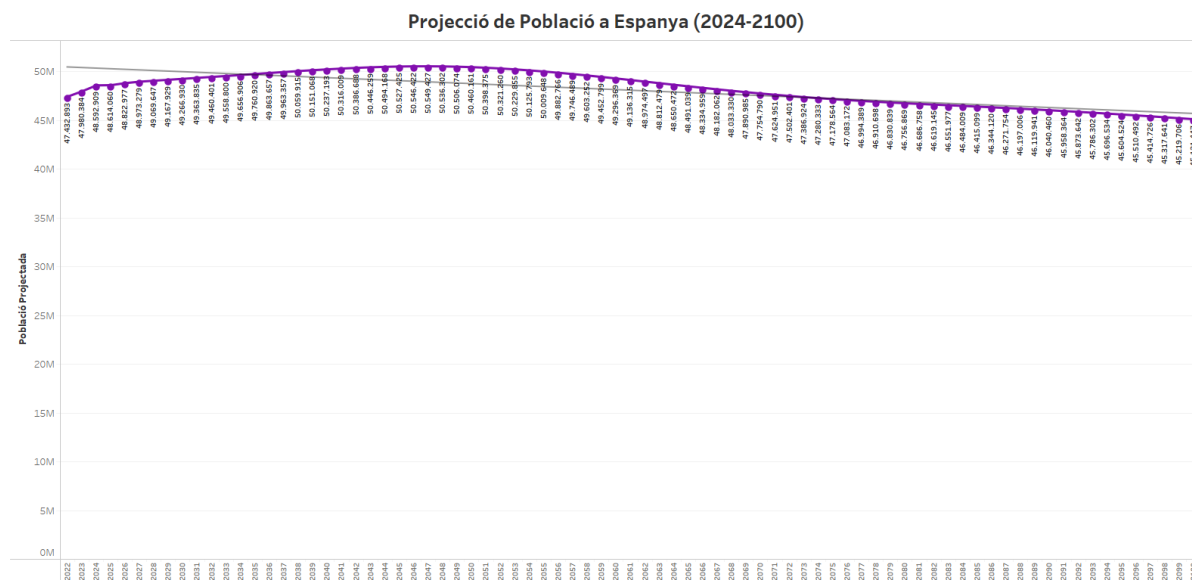
Il·lustració 15: “Població a Espanya (1960-2024)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).

En aquest primer gràfic que mostra l'evolució de la població a Espanya des de 1960 fins al 2024, es revela un augment significatiu en el nombre d'habitants. El 1960, la població era de 30,3 milions de persones i a 1 de gener de 2024, aquesta xifra ha ascendit a 48.592.909

habitants, cosa que converteix Espanya en el setè país més poblat d'Europa i el quart en extensió territorial amb 505.990 km² (Wikipedia, 2024).

Aquest increment de més de 20 milions de persones en poc més de seixanta anys reflecteix un creixement constant fins al 2010. L'any 2008 es destaca com el de major creixement percentual amb un increment de l'1,974%, mentre que el 2016 va registrar el menor creixement amb un decreixement del -0,020%. Un augment notable es va produir el 2003, quan la població va créixer un 1,932% en comparació amb el 0,909% de l'any anterior. No obstant això, entre 2013 i 2016, es va observar un creixement negatiu de la població, reflectint una disminució en el nombre d'habitants durant aquest període.

El creixement poblacional a Espanya ha estat influenciat per diversos factors, com la immigració i les variacions en les taxes de natalitat i mortalitat que més endavant seran analitzades. Aquest creixement constant fins al 2010 es pot atribuir a una economia en expansió i un flux migratori positiu. D'altra banda, els anys de creixement negatiu entre 2013 i 2016 podrien estar relacionats amb la crisi econòmica del 2008 i els seus efectes retardats en la demografia.

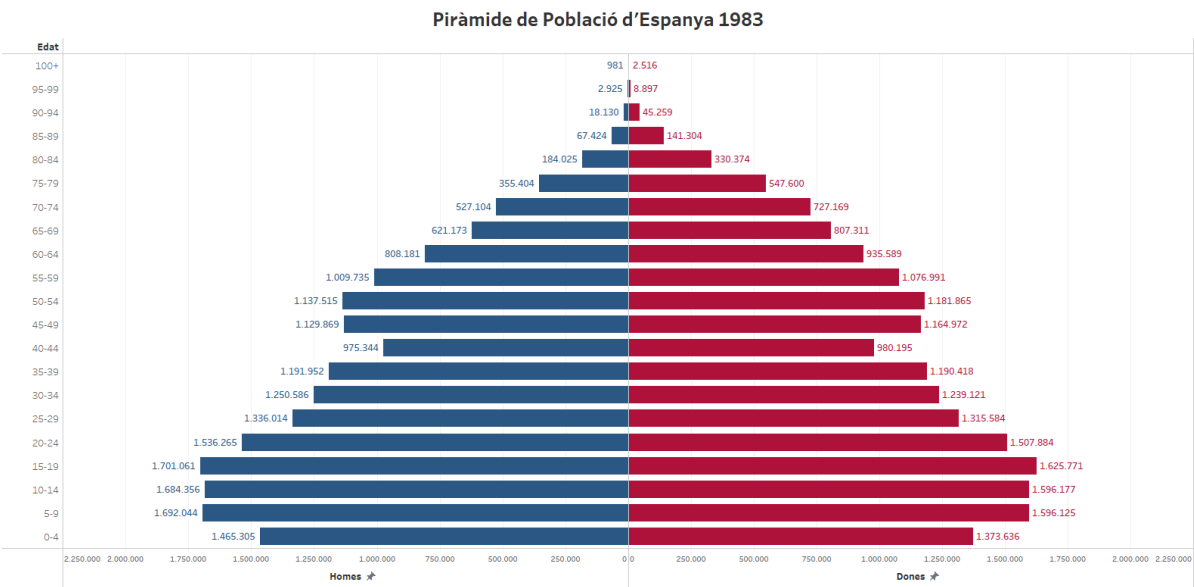


Il·lustració 16: "Projecció de Població a Espanya (2024-2100)", Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1st January by age, sex and type of projection (2024).

Ara bé, si tenim en compte les projeccions de població d'Eurostat, s'anticipa que a partir del 2048, el creixement de la població a Espanya serà negatiu, amb una disminució contínua fins a

assolir valors mínims l'any 2100. Les projeccions de població ofereixen una perspectiva futura basada en escenaris hipotètics que consideren supòsits de fertilitat, mortalitat i migració neta, entre altres.

Aquestes projeccions subratllen la importància de preparar-se per als desafiaments demogràfics futurs d'una zona determinada, especialment pel que fa a la disminució prevista de la població i, en conseqüència, a l'envelliment d'aquesta població.



Il·lustració 17: “Piràmide de Població d'Espanya 1983”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).

Situant-nos cinquanta anys enrere, tenim la fotografia de la població l'any 1983 en forma de piràmide de població. Una piràmide de població és una eina essencial en demografia, ja que ofereix una instantània visual de l'estructura d'edat i de gènere d'una població en un moment concret. Representada visualment, la població es divideix en grups d'edat, generalment en intervals de cinc anys i amb barres horitzontals anomenades cohorts; les barres de l'esquerra corresponen als homes i les de la dreta a les dones.

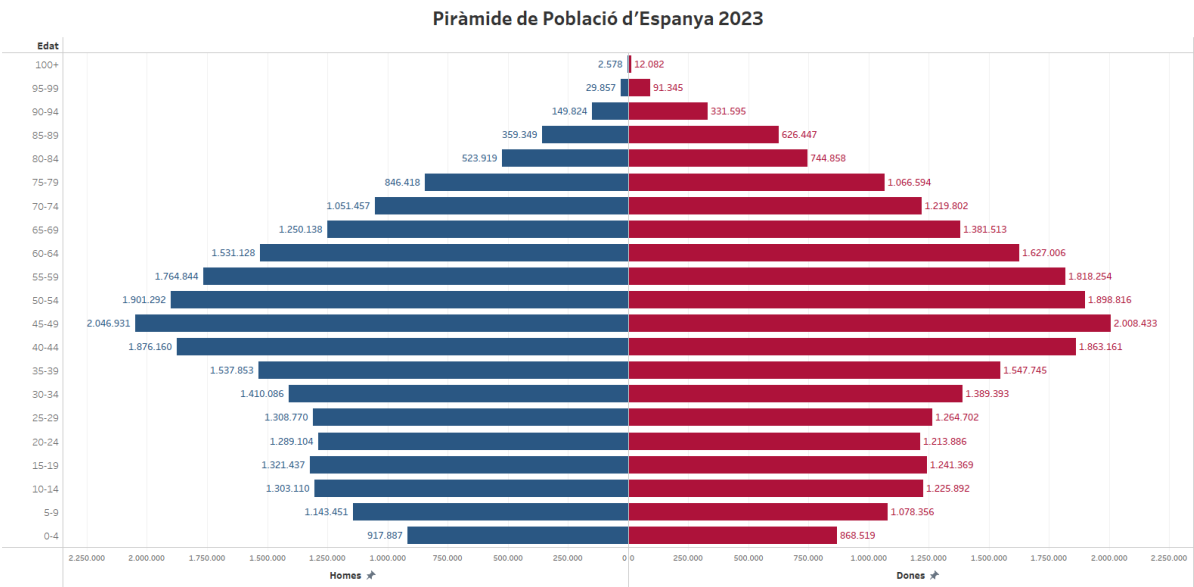
El 1983, Espanya es trobava en un període de transició després de la dictadura franquista, amb importants canvis socials, polítics i econòmics que influïen en la seva demografia. Tenint en compte això, la piràmide de població d'Espanya el 1983 presenta una estructura típica d'un país en creixement demogràfic, coneguda com a piràmide progressiva o expansiva.

La base de la piràmide és notablement ampla, indicant una alta proporció de població jove. Els infants i adolescents fins a 14 anys constitueixen el 24,7% de la població total. Aquesta alta taxa de natalitat és un indicador d'un creixement demogràfic robust, reflectint les polítiques pronatalistes de les dècades anteriors (Gobierno de España, 1960) i la recuperació econòmica després de les crisis sofertes als anys setanta (Alonso & Furio Blasco, 2007), donant com a resultat el fenomen del *baby-boom*.

La població de 15-64 anys constitueix el 63,79%, indicant una gran majoria en edat laboral, cosa que és beneficiosa per a l'economia. Tot i això, la franja d'edat de 15 a 19 anys és la més poblada, amb 1,7 milions d'homes i 1,6 milions de dones, representant el 8,68% de la població total i suggerint un elevat nombre de naixements en els anys previs. La franja d'edat de 40 a 44 anys en canvi, és significativament més estreta en comparació amb les seves franges immediates.

Finalment, la població de 65 anys o més representa l'11,52%, un percentatge relativament baix comparat amb la població jove, cosa que s'explica per una menor esperança de vida en aquells anys, ja sigui per les condicions socials i sanitàries o per diferents esdeveniments bèl·lics com la Guerra Civil Espanyola (1936-1939) i la Segona Guerra Mundial (1939-1945), a més de les dificultats econòmiques de la postguerra.

La piràmide de població d'Espanya el 1983 reflecteix una societat jove i en expansió, amb una alta taxa de natalitat i un gran segment de la població en edat productiva. L'estructura demogràfica suggereix que en aquell moment, Espanya estava en una fase de creixement poblacional significatiu, la qual cosa generaria una àmplia base de treballadors per al futur.



Il·lustració 18: “Piràmide de Població d'Espanya 2023”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).

D'altra banda, veiem la piràmide de població actual d'Espanya, l'any 2023. Entre 1983 i 2023, Espanya ha experimentat importants canvis demogràfics, augmentant 10 milions d'habitants en aquest període. La transició cap a una economia desenvolupada, la millora en les condicions de vida i les polítiques de benestar social han contribuït a un augment en l'esperança de vida i, conseqüentment, a un envelliment de la població. Tanmateix, aquests canvis també han anat acompanyats d'una disminució en les taxes de natalitat, fet comú en molts països europeus.

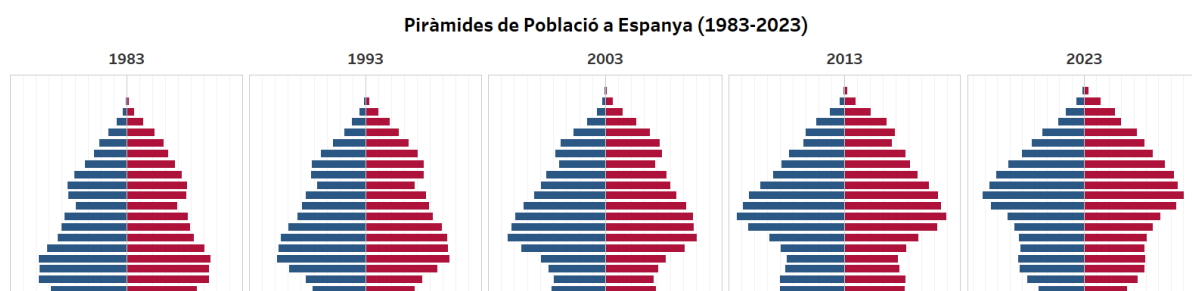
Aquesta piràmide presenta una estructura típica d'un país amb creixement demogràfic lent o en retrocés, coneguda com a piràmide regressiva o en forma de bulb. La base de la piràmide és notablement estreta, indicant una baixa proporció de població jove, en què els infants de fins a 14 anys constitueixen només el 13,6% de la població total. Aquesta baixa taxa de natalitat reflecteix canvis socials i econòmics significatius, com la incorporació de la dona al mercat laboral, la legalització de l'avortament el 1985, l'accés generalitzat a mètodes anticonceptius i, per descomptat, una millora en el sistema de sanitat pública nacional.

La franja de 15-64 anys representa el 66,26% de la població, mantenint-se com la majoria de la població, però amb una tendència cap a l'envelliment. Dins d'aquesta, l'interval que comprèn les edats de 45 a 49 anys és el més ample de tots, amb 2 milions d'homes i 2 milions de dones respectivament, representant el 8,43% de la població total. Això, sumat a les persones en la

cinquantena d'edat, les quals sumen gairebé 7 milions i mig de persones, s'explica pel *baby-boom* de principis de la dècada dels seixanta i finals de la dels setanta.

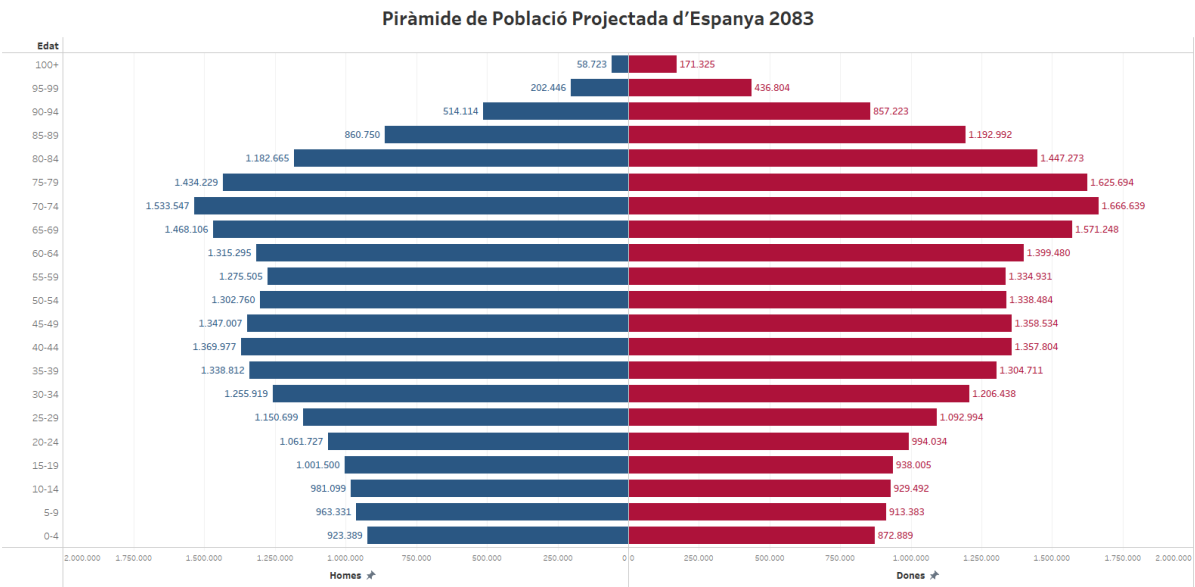
I ja a la punta de la piràmide, la població de 65 anys o més representa el 20,14% del total d'habitants espanyols, gairebé el doble que el 1983, reflectint aquest augment significatiu en l'esperança de vida i les millores en les condicions sanitàries i de vida.

A diferència de la piràmide de fa cinquanta anys, la piràmide de població d'Espanya el 2023 reflecteix una societat en procés d'envelliment ràpid, amb una baixa taxa de natalitat i mortalitat. Aquest panorama demogràfic planteja reptes significatius per al futur, com la sostenibilitat del sistema de pensions, la necessitat de serveis d'atenció per a una població major i la possible escassetat de mà d'obra jove. La creixent proporció de persones grans i la disminució de joves subratlla la importància de polítiques que abordin l'envelliment de la població i fomentin la natalitat per equilibrar l'estructura demogràfica.



Il·lustració 19: "Piràmides de Població a Espanya (1983-2023)", Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1 January by age and sex (2024).

Com a resultat s'observen les piràmides de població cada deu anys des de 1983 fins al 2023, on es pot veure aquesta distribució en les franges d'edat resultants dels motius explicats amb anterioritat. S'aprecia com ha canviat la forma de la piràmide de població en cinquanta anys, passant d'una forma progressiva a una forma regressiva en expandir-se la part mitjana, reduint-se la base i apareixent de forma residual una petita barra a la part més alta.



Il·lustració 20: “Piràmide de Població Projectada d'Espanya 2083”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Population on 1st January by age, sex and type of projection (2024).

Per contrastar les piràmides vistes, podem utilitzar projeccions de població, en aquest cas d'Eurostat, per predir la distribució d'edats futura, a cinquanta anys vista. La piràmide de població projectada per a Espanya el 2083 mostra una estructura estancada o contractiva. Aquest tipus de piràmide es caracteritza per una base estreta que indica baixes taxes de natalitat, una proporció més equilibrada de les cohorts d'edat mitjana i una punta que s'eixampla lleugerament a causa d'una major esperança de vida, explicada per un envelliment de la població i una reducció en el nombre de naixements.

Així és. La piràmide projectada reflecteix una baixa taxa de natalitat, una molt baixa taxa de mortalitat i un augment considerable en l'esperança de vida. Això implica que la població està envellint ràpidament i no es produeix un reemplaçament generacional adequat.

Encara que aquesta vegada les franges d'edat estiguin més equilibrades, això és típic d'una piràmide estancada. El grup d'edat de 70 a 74 anys és el més gran de tota la piràmide, amb 1,53 milions d'homes i 1,66 milions de dones, representant un 6,87% de la població total.

Una altra cosa que crida l'atenció és com a partir dels vuitanta anys, hi ha una notable disparitat de gènere, amb 4,1 milions de dones majors de vuitanta anys comparades amb 2,73 milions d'homes. Aquesta diferència es pot atribuir a una major esperança de vida de les dones en comparació amb els homes.

Pel que fa a l'estructura poblacional, la població jove continua disminuint amb el pas del temps, representant només l'11,99% de la població total, un punt i mig menys que actualment, cosa que indica una disminució contínua en la taxa de natalitat.

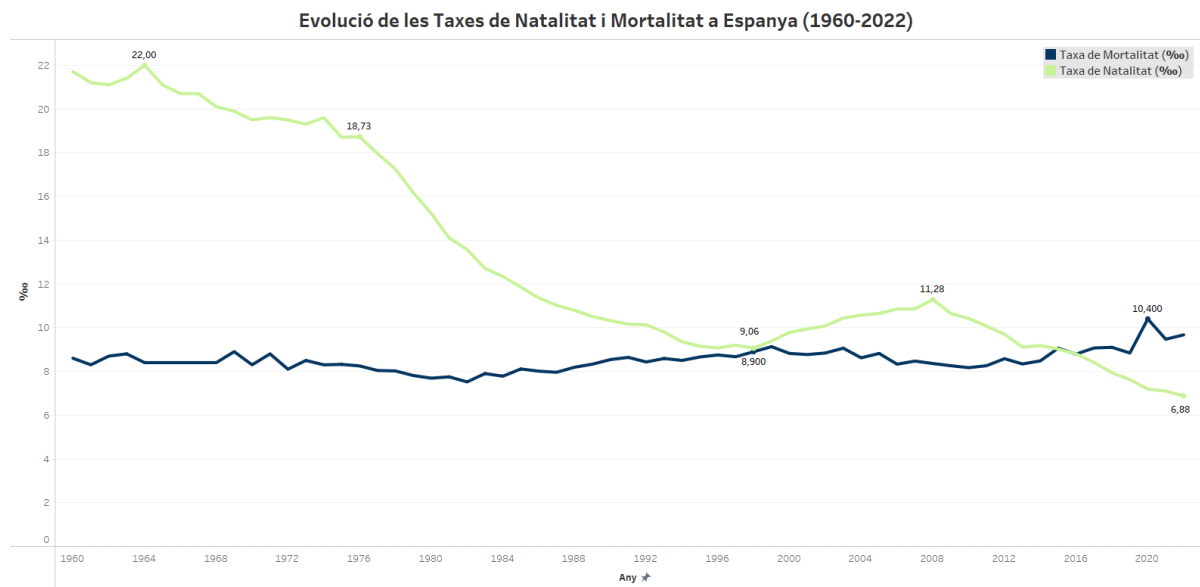
La població en edat laboral, per altra banda, constitueix el 53,15%, una reducció significativa comparada amb el 66,26% del 2023, canvi que implica una menor proporció de persones en edat de treballar.

D'altra banda, la població major de 65 anys representa un 34,86%, un augment alarmant del 15% respecte en l'actualitat, cosa que suposarà, segons les prediccions, un increment en la càrrega per als sistemes de salut i pensions, a la vegada que una població cada cop més envellida.

Encara que aquestes dades són projeccions, es basen en tendències i realitats actuals, per la qual cosa la piràmide projectada suggereix un panorama futur preocupant. Amb un terç de la població major de 65 anys, hi haurà una creixent necessitat de serveis de salut i assistència social, sense oblidar les malalties específiques en aquestes edats. Això, sumat a la disminució de la població en edat laboral, pot resultar en una menor productivitat econòmica, majors càrregues fiscals per sostenir la població envellida i un sistema públic de pensions en perill. Finalment, la contínua baixa taxa de natalitat implica que no hi haurà prou joves per reemplaçar la població envellida que augmenta exponencialment amb el pas dels anys.

Aquestes projeccions plantegen una sèrie de seriosos reptes socials, econòmics i sanitaris a tenir molt en compte en un futur no tan llunyà. És crucial abordar aquestes realitats demogràfiques mitjançant polítiques que fomentin la natalitat, millorin les condicions per a les famílies joves i assegurin la sostenibilitat del sistema de pensions i salut dels nostres majors.

Per tancar l'apartat de demografia i projeccions de població, tenim aquest gràfic relatiu a l'evolució de les taxes de mortalitat i natalitat del període entre 1960 i 2022 que s'han anat veient fins ara, revelant canvis significatius en la dinàmica demogràfica del país al llarg d'aquestes sis dècades.



Il·lustració 21: “Evolució de les Taxes de Natalitat i Mortalitat a Espanya (1960-2022)”, Font: Elaboració Pròpia a partir de dades extretes de Datosmacro.com, Natalidad España 2022 i Mortalidad España 2022. (2024).

Entre 1960 i 1980, la taxa de natalitat a Espanya era significativament més gran que la taxa de mortalitat, gairebé triplicant-la amb valors entre 18‰ i 22‰. Durant aquests anys, Espanya va experimentar el *baby-boom* ja esmentat, reflectit en una alta taxa de naixements i contribuint a un creixement poblacional robust.

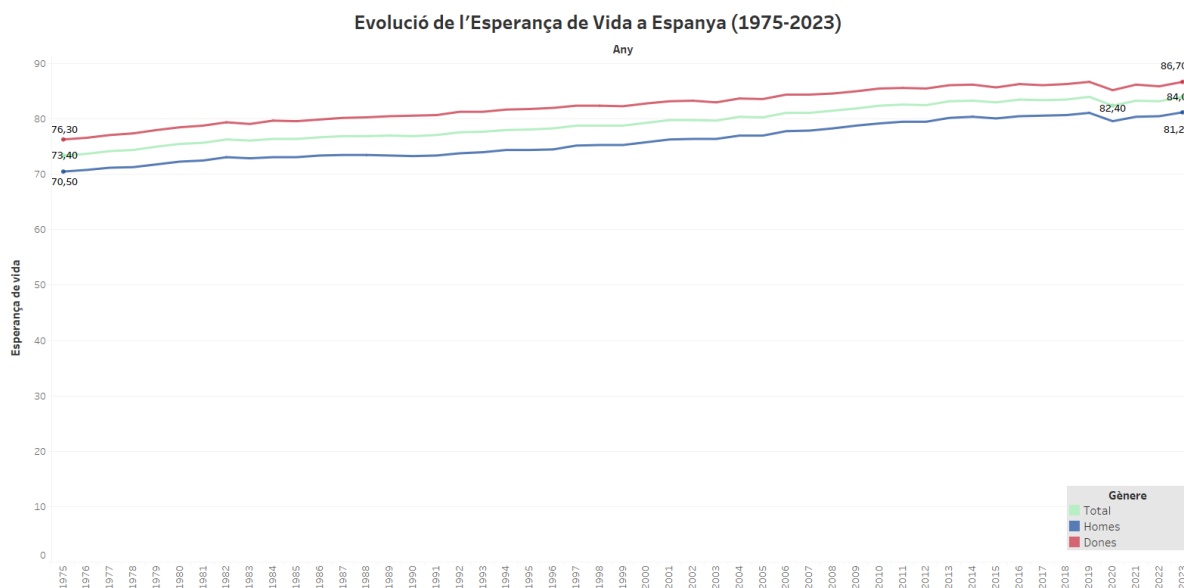
Però la dècada de 1980 marca un punt d'inflexió, concretament l'any 1976. La taxa de natalitat va començar a disminuir dràsticament, mentre que la taxa de mortalitat va mostrar un lleuger augment. Aquesta tendència va reduir considerablement la diferència entre naixements i morts, acostant-los cap a l'any 2000, fenomen que s'atribueix a canvis socials i econòmics fruit de la transició, incloent-hi una major incorporació de la dona al mercat laboral i canvis en els valors familiars i socials.

A partir de l'any 2000, la taxa de natalitat va mostrar una recuperació, assolint un pic l'any 2008 amb un valor d'11,28‰. Aquest repunt pot estar relacionat amb el creixement econòmic i l'augment de la immigració. No obstant això, la crisi econòmica que va començar el 2008 va tenir un impacte negatiu significatiu, resultant en una nova caiguda de la taxa de natalitat. Des de llavors, la taxa de natalitat ha disminuït de manera constant, assolint un mínim de 6,88‰ el 2022.

La taxa de mortalitat, per altra banda, s'ha mantingut relativament estable al llarg dels anys, amb una lleugera tendència creixent a causa de l'envelliment de la població. Com a dada rellevant es destaca l'any 2020, que a causa de la pandèmia de la COVID-19 la taxa de mortalitat es va disparar fins a 10,4‰, reflectint la crisi sanitària viscuda recentment.

4.2. Salut i esperança de vida

A banda d'examinar les estructures poblacionals així com les taxes de mortalitat i natalitat, és fonamental connectar aquests aspectes demogràfics amb indicadors de salut com l'esperança de vida i els anys saludables, per garantir una visió integral de les necessitats i els desafiaments en matèria de salut pública a Espanya.



Il·lustració 22: "Evolució de l'Esperança de Vida a Espanya (1975-2023)", Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Life expectancy by age and sex (2024).

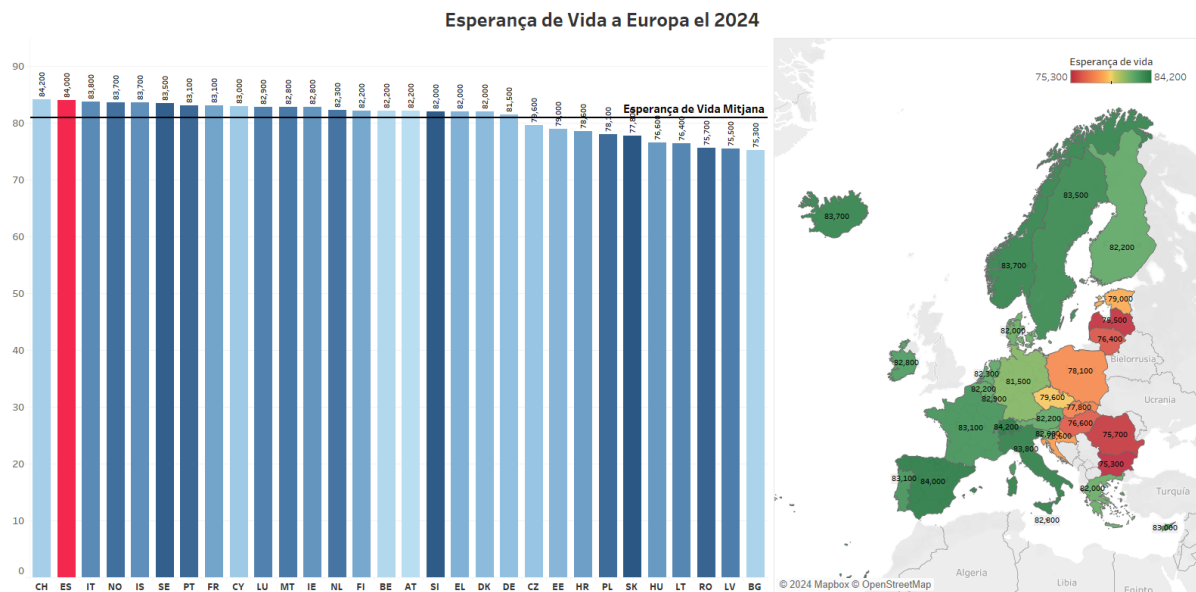
El gràfic de l'esperança de vida a Espanya entre el 1975 i el 2023 reflecteix el progrés notable en la salut i les condicions de vida del país al llarg d'aquestes dècades. El 1975, l'esperança de vida per als homes era de 70,5 anys i per a les dones de 76,3 anys. Aquestes xifres indiquen les diferències de gènere en la salut i les condicions de vida en aquella època, quan les dones ja vivien uns sis anys més que els homes.

Des de 1975, l'esperança de vida ha anat augmentant de manera constant, arribant el 2023 a 81,2 anys per als homes i 86,7 anys per a les dones. Aquest increment es deu a múltiples factors, incloent-hi millores en la sanitat, l'alimentació i la higiene, així com avanços en la medicina preventiva i el tractament de malalties cròniques. Les polítiques de salut pública i el desenvolupament socioeconòmic també han contribuït a aquest increment.

Un punt notable en el gràfic és la caiguda en l'esperança de vida l'any 2020, per culpa de la pandèmia de la COVID-19, que va impactar la salut pública d'arreu del món. Tot i això, l'any

següent l'esperança de vida va recuperar els seus nivells previs, demostrant la fortalesa del sistema sanitari espanyol i l'eficàcia de les mesures implementades per combatre la pandèmia.

Actualment, les dones a Espanya viuen de mitjana 5,5 anys més que els homes, una diferència que tot i reduir-se lleugerament al llarg dels anys, continua sent significativa.



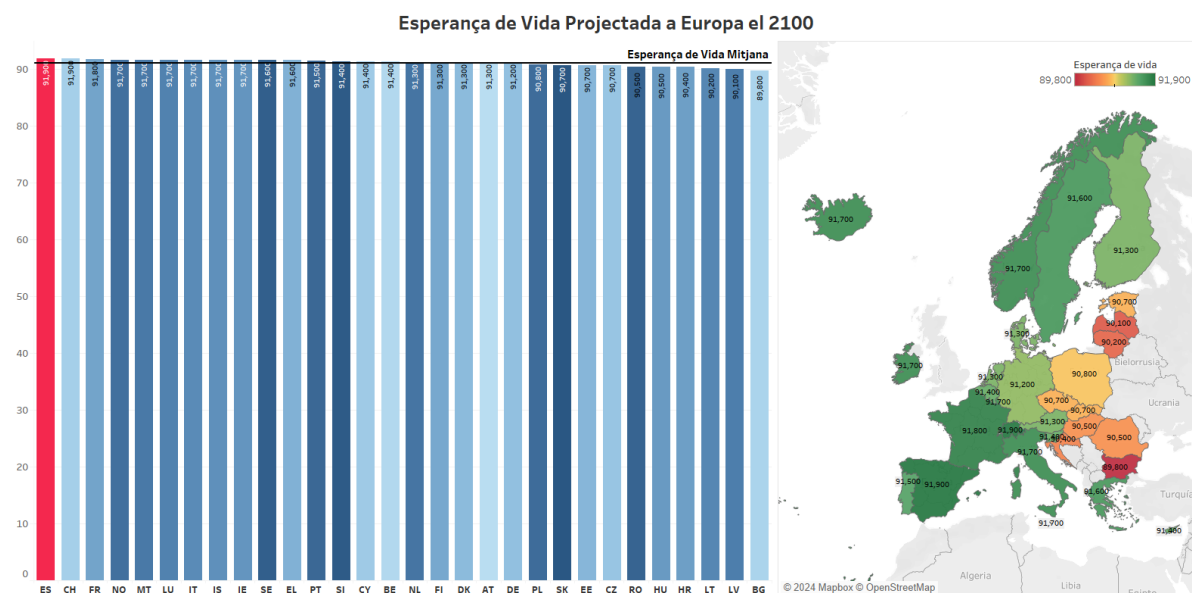
Il·lustració 23: “Esperança de Vida a Europa el 2024”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Life expectancy by age and sex (2024).

En comparar l'esperança de vida d'Espanya amb la de la resta de països de la Unió Europea l'any 2024, s'observa que Espanya, amb una esperança de vida de 84 anys, és el segon país amb més esperança de vida, només per darrere de Suïssa, que té una esperança de vida de 84,2 anys.

Els països que segueixen Espanya en termes de major esperança de vida són Itàlia (83,8 anys), Noruega i Islàndia (tots dos amb 83,7 anys) i Suècia (83,5 anys). Aquests països, situats tant al sud com al nord d'Europa, destaquen pels seus alts estàndards de vida, sistemes de salut avançats i polítiques socials robustes que promouen el benestar i la longevitat dels seus ciutadans.

D'altra banda, els països del centre d'Europa i les repúbliques exsoviètiques, com Estònia, Letònia, Lituània, Polònia, Hongria, Bulgària i Romania, mostren les esperances de vida més baixes, totes per sota de la mitjana europea de 81 anys. Aquesta diferència significativa pot explicar-se per diversos factors històrics i socioeconòmics, ja que aquests països han passat recentment per períodes d'inestabilitat econòmica i política, especialment durant i després de

l'era soviètica, cosa que ha afectat negativament els seus sistemes de salut i les condicions de vida de les seves poblacions, processos llargs i complexos que han impactat la qualitat de vida i l'accés a serveis de salut.



Il·lustració 24: “Esperança de Vida Projectada a Europa el 2100”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Projected life expectancy by age (in completed years), sex and type of projection (2024).

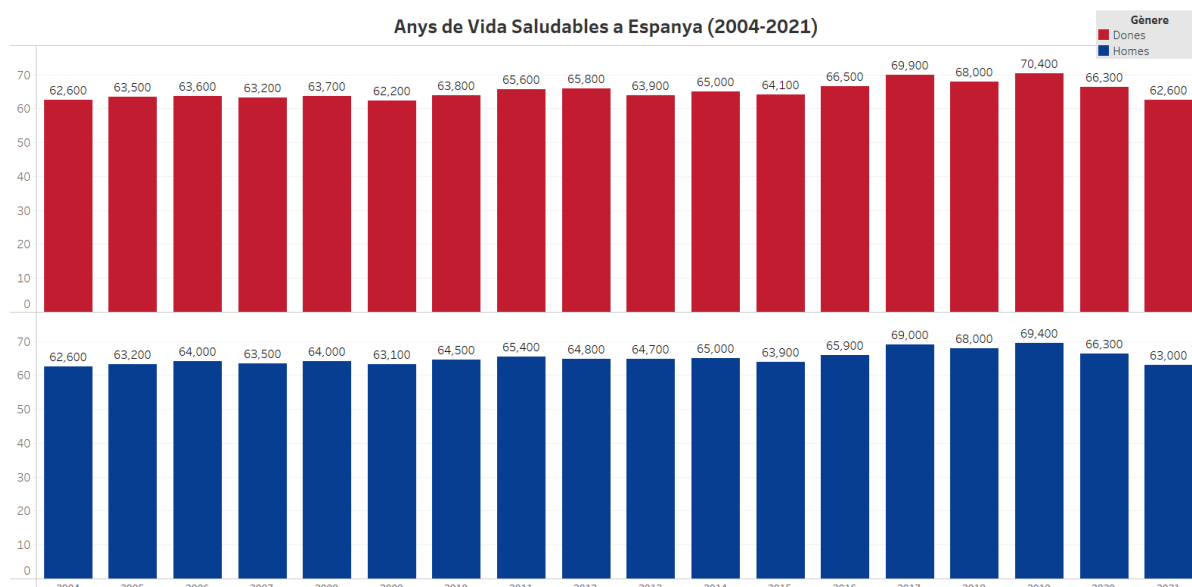
Mirant cap al futur, les projeccions de l'esperança de vida a Europa per a l'any 2100 segons les prediccions d'Eurostat indiquen que la tendència actual en la distribució de valors es mantindrà, i els països que actualment tenen la major esperança de vida continuaran liderant el 2100.

Ara bé, l'esperança de vida mitjana projectada a Europa per al 2100 és de 91,3 anys, cosa que representa un augment de 10 anys en comparació amb els valors actuals i una característica notable d'aquestes projeccions és que tots els països estaran més a prop de la mitjana, indicant una reducció en les disparitats d'esperança de vida entre els diferents països europeus.

Espanya, juntament amb Suïssa, es projecta com el país amb major esperança de vida el 2100, amb 91,9 anys, consolidant la seva posició com un dels líders en esperança de vida a Europa. Aquest fet destaca la qualitat dels sistemes de salut i el bon estil de vida en aquests països, que continuaran brindant beneficis als seus ciutadans en termes de longevitat.

De la mateixa manera, és crucial no només considerar la longevitat, sinó també la qualitat de vida durant aquests anys. Aquest gràfic il·lustra els anys de vida saludables a Espanya des de 2004 fins a 2021, que es defineixen com aquells anys de vida sense limitacions funcionals o

discapacitats significatives. Amb l'envelliment, augmenten les malalties cròniques, els problemes mentals i les discapacitats físiques, afectant la qualitat de vida.



Il·lustració 25: “Anys de Vida Saludables a Espanya (2004-2021)”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Healthy life years by sex (from 2004 onwards) (2024).

En aquests quinze anys, s'observa una variació en els anys de vida saludables. El 2004, la mitjana era de 62,7 anys, valor que va augmentar gradualment, assolint un màxim el 2019 amb 69,9 anys saludables de mitjana. No obstant això, la pandèmia de la COVID-19 va tenir un impacte negatiu, reduint aquest valor a 66,3 anys el 2021, similar al nivell de fa quinze anys. La caiguda en els anys de vida saludables el 2020 i 2021 subratlla l'impacte significatiu de la pandèmia de COVID-19, reflectint tant l'efecte directe del virus en la salut com les conseqüències indirectes, ja siguin el deteriorament de la salut mental, el retard en el tractament de malalties cròniques i la disminució de l'activitat física per culpa de les restriccions, entre altres.

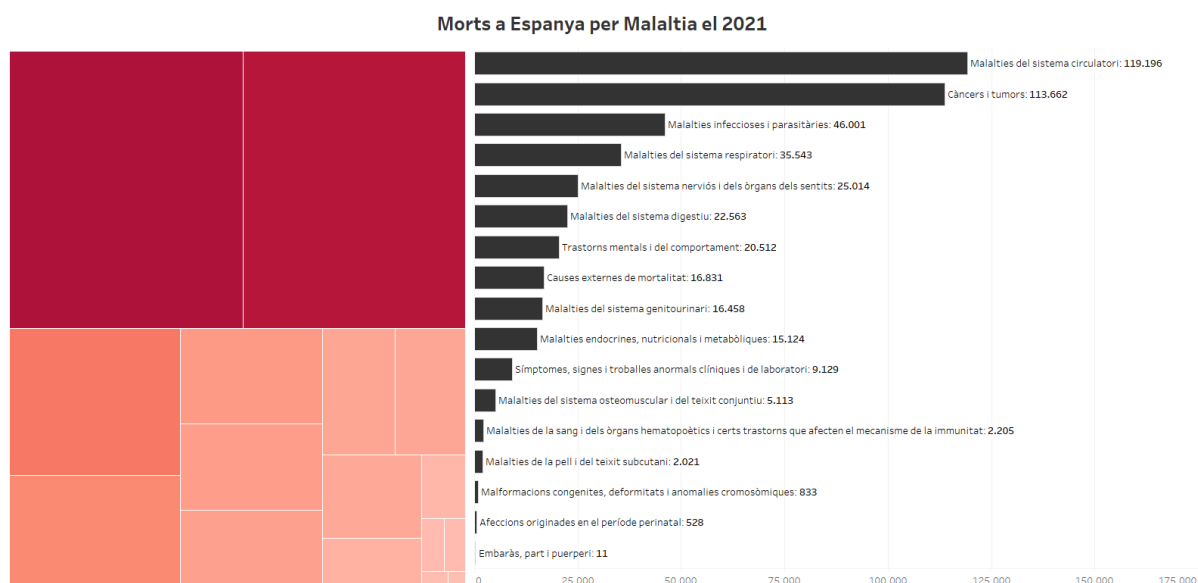
Pel que fa a diferències de gènere, s'aprecia que les dones a Espanya no només tenen una esperança de vida major que els homes, sinó que també gaudeixen d'un nombre lleugerament superior d'anys de vida saludables.

En general, hi ha hagut un increment en els anys de qualitat de vida de la població espanyola, anant de la mà de l'increment de l'esperança de vida així com dels avenços mèdics i socials. Aquesta perspectiva dual de l'esperança de vida i els anys saludables destaca la necessitat de

polítiques de salut pública que no només se centrin en la longevitat, sinó també en la qualitat de vida, especialment important en un context d'envelliment poblacional.

4.3. Mortalitat i problemes de salut

Després d'analitzar els principals indicadors demogràfics i vitals del país, aquest apartat cerca aprofundir sobre les causes de mortalitat dins del total de defuncions i els principals problemes crònics de salut a Espanya.

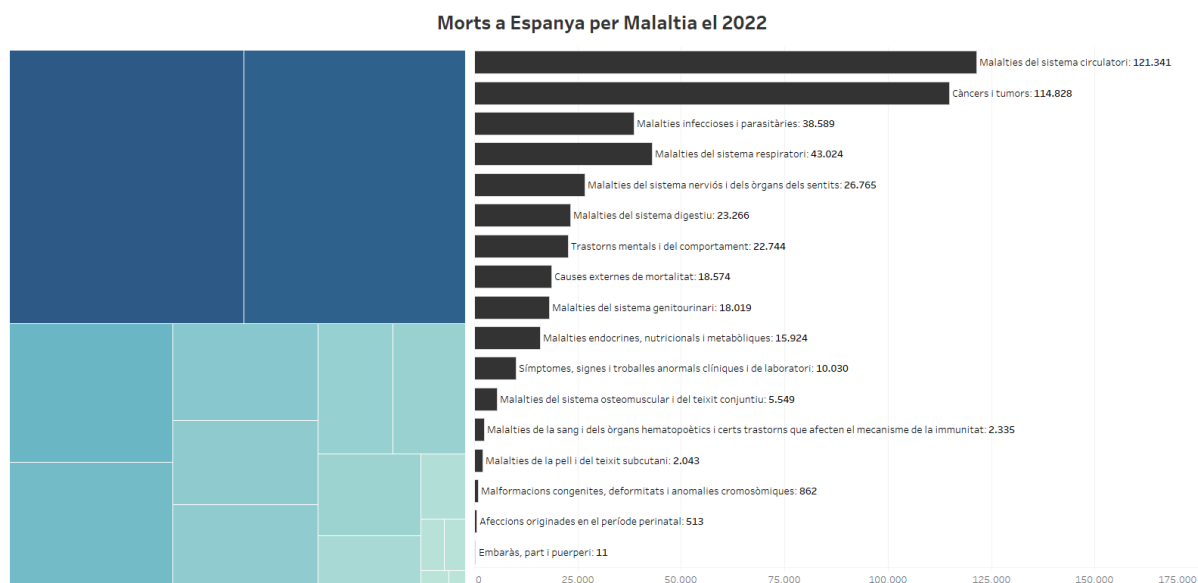


Il·lustració 26: “Morts a Espanya per Malaltia el 2021”, Font: Elaboració Pròpia a partir de dades extretes de Instituto Nacional de Estadística, Defunciones según la causa de muerte (2024).

En aquest panell, el gràfic de barres mostra els valors de cada categoria ordenats de més gran a més petit, mentre que el mapa d'arbre o *treemap* el complementa mostrant la proporció de cada categoria dins del total.

De les 450.744 persones que van morir l'any 2021, les tres causes principals de mort van ser: malalties de l'aparell circulatori, amb 119.196 defuncions; tumors i càncers, amb 113.662 defuncions; i malalties infeccioses i parasitàries, incloent-hi la COVID-19, amb 46.001 defuncions. Aquestes tres categories destaquen significativament sobre les altres.

Per contrastar el *dashboard* anterior, tenim aquest referent a l'any 2022. De les 464.417 persones que van morir l'any 2022, les tres causes principals de mort van ser: malalties de l'aparell circulatori, amb 121.341 defuncions; tumors i càncers, amb 114.828 defuncions; i malalties de l'aparell respiratori, amb 43.024 defuncions.



Il·lustració 27: “Morts a Espanya per Malaltia el 2022”, Font: Elaboració Pròpia a partir de dades extretes de Instituto Nacional de Estadística, Defunciones según la causa de muerte (2024).

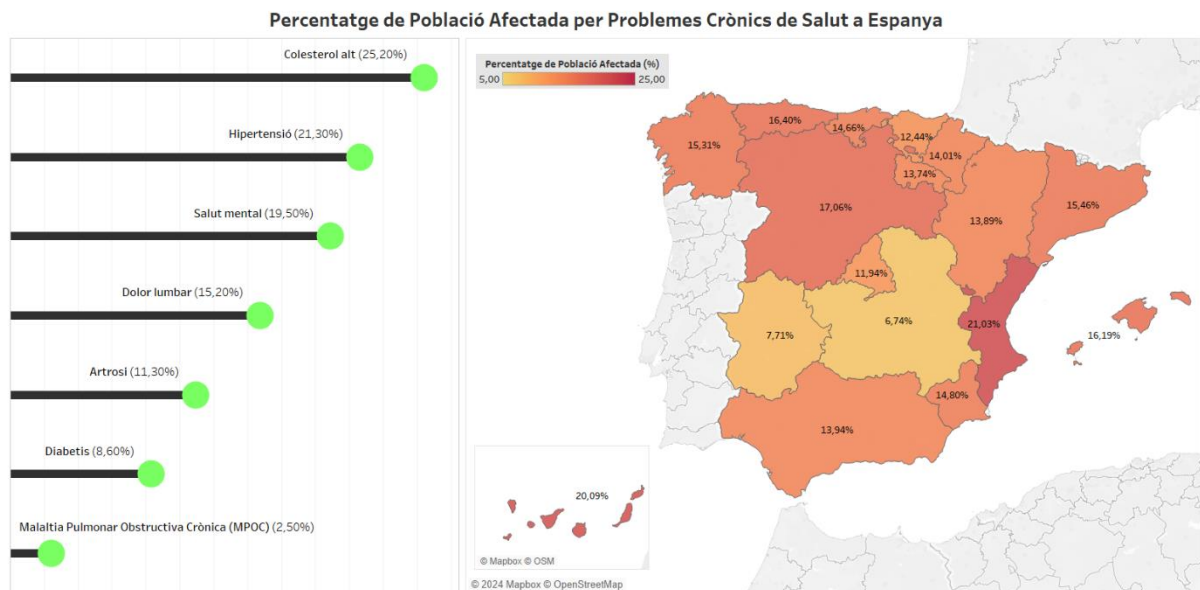
El grup de malalties del sistema circulatori i els tumors van ser responsables de la meitat de les defuncions ocorregudes el 2022, situant-se com les principals causes. Les malalties del sistema respiratori van representar la tercera causa de mort, responsables del 9,3% de les defuncions.

El nombre de morts per les tres principals causes va augmentar respecte al 2021. A més, la tercera causa de mortalitat més gran va canviar de les malalties infeccioses i parasitàries a les malalties del sistema respiratori. Això es pot explicar pel fet que, atès que la COVID-19 està inclòs en el grup de malalties infeccioses i parasitàries, les persones que no van morir a causa de la COVID-19 el 2021 podrien haver mort el 2022 a causa de les seves seqüeles.

La variació percentual del 2022 respecte al 2021 va ser negativa en les categories de malalties infeccioses i parasitàries (-16,1%) i afeccions originades en el període perinatal (-2,8%). D'altra banda, les categories que van tenir més variació van ser les malalties del sistema respiratori (21%) i els trastorns mentals i del comportament (10,9%).

Aquests canvis en les causes de mort el 2022 reflecteixen tant les seqüeles de la pandèmia de COVID-19 com les tendències en matèria de salut pública. La disminució de morts per malalties infeccioses i parasitàries suggereix un maneig més efectiu de la pandèmia, mentre que l'augment de les malalties respiratòries i mentals destaca noves àrees de preocupació que requereixen atenció mèdica i polítiques de salut específiques.

El següent gràfic sobre els principals problemes crònics de salut a Espanya ofereix una perspectiva detallada de les condicions de salut més prevalents i la seva distribució en la població. A més de conèixer les causes de mort entre els ciutadans espanyols, també és interessant saber quins són els principals problemes crònics que pateix la població i veure si poden aportar informació valuosa sobre les seves causes.



Il·lustració 28: “Percentatge de Població Afectada per Problemes Crònics de Salut a Espanya”, Font: Elaboració Pròpia a partir de dades extretes de Ministerio de Sanidad de España, Principales problemas crónicos de salud (2024).

Segons dades registrades pel ministeri de sanitat, aquests set principals problemes crònics de salut i el percentatge de població afectada són:

1. Colesterol alt (25,2%)
2. Hipertensió (21,3%)
3. Problemes de salut mental (19,5%)
4. Mal d'esquena lumbar (15,2%)
5. Artrosi (11,3%)
6. Diabetis (8,6%)
7. Malaltia pulmonar obstructiva crònica (MPOC) (2,5%)

Les comunitats autònomes amb un major percentatge de població afectada són la Comunitat Valenciana (21,03%) i Canàries (20,09%). En canvi, les comunitats amb un menor percentatge d'afectats són Extremadura (7,71%) i Castella-la Manxa (6,74%). Catalunya se situa en sisè lloc

i per sobre de la mitjana nacional, amb un 15,31% de la seva població afectada per problemes crònics de salut.

D'entre tots aquests problemes de salut, cal destacar el fet que els problemes de salut mental ocupen el tercer lloc en la llista de problemes crònics, afectant ja una cinquena part de la població espanyola (19,5%). Això destaca la creixent preocupació per la salut mental, un problema que, encara que relativament nou en termes de reconeixement, té un impacte significatiu en la societat.

Analitzar aquestes dades és fonamental per entendre les causes subjacents de les malalties cròniques en la població espanyola i poder abordar-les amb contundència. A continuació, l'anàlisi posarà el focus sobre diferents hàbits de salut i conductes de risc que poden anar acompanyats dels problemes de salut vistos.

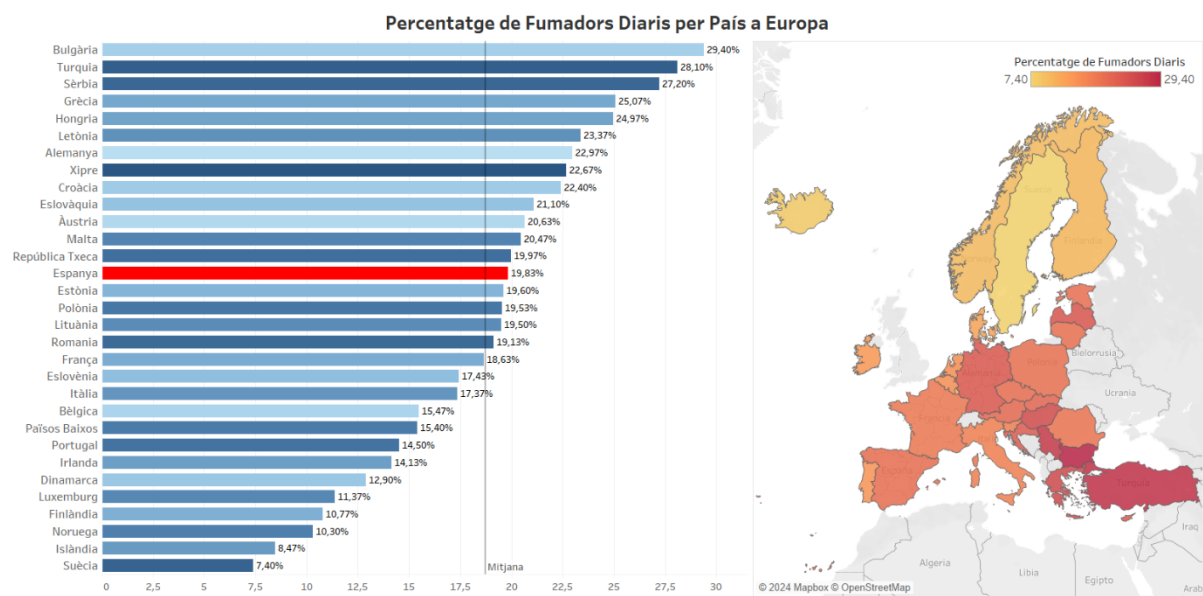
4.4. Hàbits de salut i conductes de risc

En aquesta quarta secció de l'anàlisi es presenten els principals factors de risc per a la salut i la vida de les persones. Segons l'Organització Mundial de la Salut, la pressió arterial alta, el consum de tabac, la glucosa en sang elevada, la inactivitat física i el sobrepès/obesitat són els principals factors de risc de mortalitat per malalties cardiovasculars a escala mundial (WHO, 2009).

És per això, que la selecció d'aquestes conductes i problemàtiques de salut queden resumides en les tres següents: el tabaquisme, l'alcoholisme i l'obesitat. Aquestes conductes de risc no només tenen un impacte directe en la salut individual, sinó que també representen un considerable desafiament per al sistema sanitari a causa de les seves conseqüències a llarg termini.

El **tabaquisme** és una de les principals causes de la Malaltia Pulmonar Obstructiva Crònica (MPOC) i un factor de risc considerable per a malalties cardiovasculars i diversos tipus de càncer, mantenint una prevalença notable malgrat les campanyes de conscienciació i les regulacions antitabac. L'**alcoholisme** contribueix a una àmplia gamma de problemes de salut, incloent-hi malalties hepàtiques com les més comunes, però també altres que poden passar més desapercebudes com trastorns mentals, a més d'influir indirectament en el desenvolupament d'altres malalties cròniques. Finalment, l'**obesitat** s'ha convertit en una epidèmia silenciosa que afecta un gran percentatge de la població espanyola i està estretament relacionada amb malalties cardiovasculars, la diabetis i diversos problemes ortopèdics i metabòlics.

Referent a la primera de les tres problemàtiques, el tabaquisme, el següent gràfic mostra el percentatge de persones fumadores diàries als diferents països europeus, revelant una considerable variabilitat entre nacions.



Il·lustració 29: “Percentatge de Fumadors Diaris per País a Europa”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Smoking of tobacco products by sex, age and degree of urbanisation (2024).

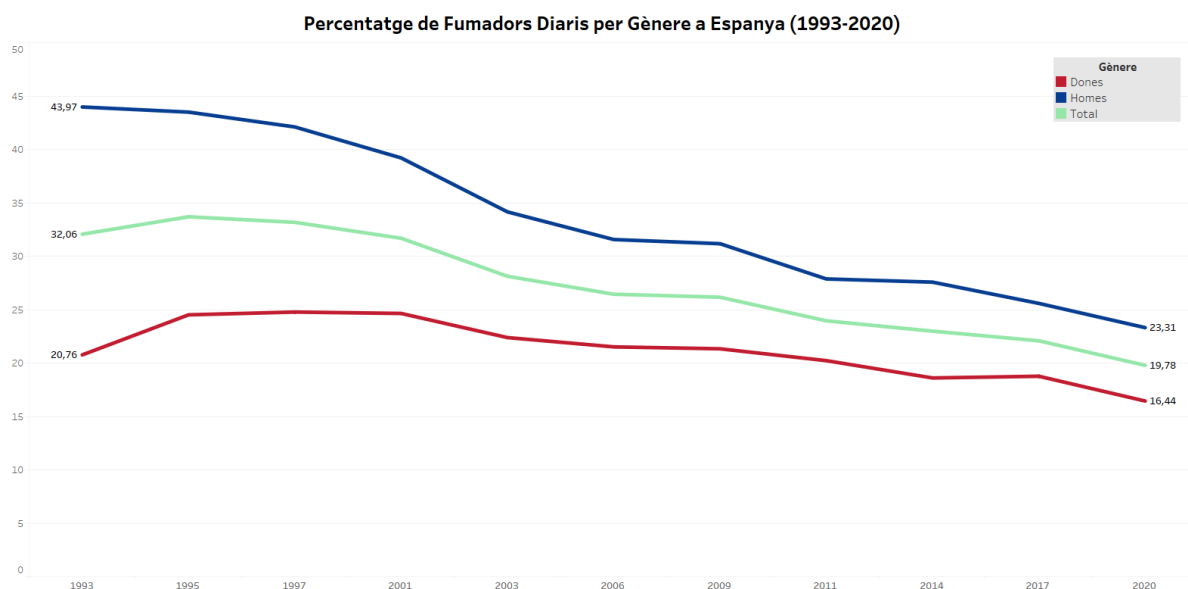
Bulgària presenta el percentatge més alt de fumadors diaris amb un 29,1%, Turquia és segona amb un 28,0% i Sèrbia està just darrere amb un 27,1% de la població total sent fumadora diària, taxes relativament altes i preocupants.

D'altra banda, Suècia destaca amb el percentatge més baix de fumadors diaris amb un 7,4%. Islàndia és el segon país amb un 8,4% i Noruega just per sobre amb un 10,3%.

El mapa de coropletes mostra una tendència en què els països més propers al Mediterrani tendeixen a tenir un percentatge més alt de població fumadora diària a diferència dels països escandinaus. Aquesta tendència es pot observar en països com Grècia (24,9%), Turquia (28,0%) i Espanya (19,8%), en contrast amb Suècia (7,4%), Islàndia (8,4%), Noruega (10,3%), Finlàndia (10,7%) i Dinamarca (12,9%). Aquestes dades suggereixen que la prevalença del tabaquisme diari és més alta a les regions mediterrànies en comparació amb altres parts d'Europa, podent estar relacionat amb factors culturals, socials i econòmics que influeixen en els hàbits de fumar en aquestes àrees.

Amb un percentatge del 19,8%, Espanya es troba en una posició intermèdia-alta en comparació amb altres països europeus. Això indica que, encara que existeixen mesures antitabac, encara hi ha marge per millorar l'eficàcia d'aquestes polítiques per reduir el tabaquisme.

Pel que fa a les dades a Espanya, el gràfic de línies que presenta l'evolució del percentatge de fumadors diaris per gènere entre 1993 i 2020 revela una disminució constant en el percentatge de fumadors diaris tant en homes com en dones al llarg dels anys.



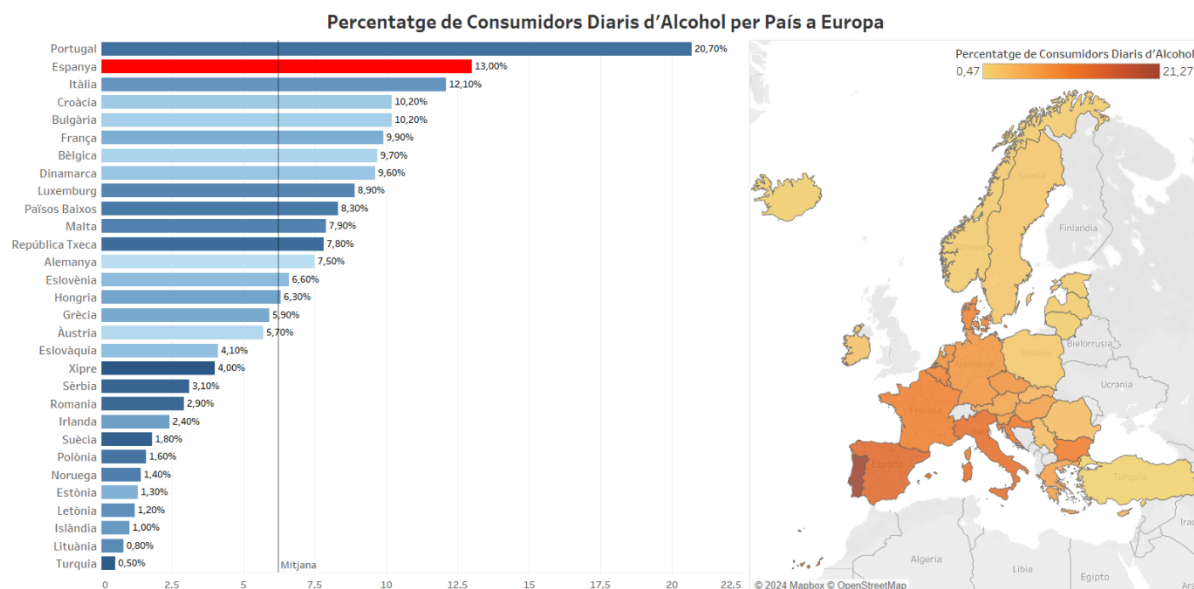
Il·lustració 30: “Percentatge de Fumadors Diaris per Gènere a Espanya (1993-2020)”, Font: Elaboració Pròpia a partir de dades extretes de Indicadores Clave Sistema Nacional de Salud, Tabaquismo (2024).

Fa més de trenta anys, el percentatge d'homes que fumaven diàriament era del 43,97%, valor que ha anat disminuint amb el pas dels anys fins a un 23,31% el 2020. D'altra banda, el percentatge de dones fumadores diàries era del 20,76%, i encara que no hi hagi hagut una reducció considerable, el valor també ha baixat fins al 16,44% el 2020 després d'un repunt a principis de segle.

Encara que ambdós gèneres han experimentat una tendència descendent en el tabaquisme diari, els homes sempre han registrat un percentatge més alt de fumadors diaris en comparació amb les dones. No obstant això, la bretxa de gènere s'ha reduït lleugerament amb el temps.

Per tant, el percentatge total de fumadors diaris en la població segueix la tendència decreixent observada fins a assolir el 19,78% de fumadors diaris el 2020, un 10% menys que fa tres dècades. Aquesta tendència pot ser causada per diverses intervencions de salut pública, com les constants campanyes de conscienciació sobre els perills del tabac, la prohibició de fumar en llocs públics i l'augment d'impostos sobre els productes de tabac.

Respecte a les dades relacionades amb l'alcoholisme, el següent panell mostra el percentatge de consumidors diaris d'alcohol en els diferents països europeus.

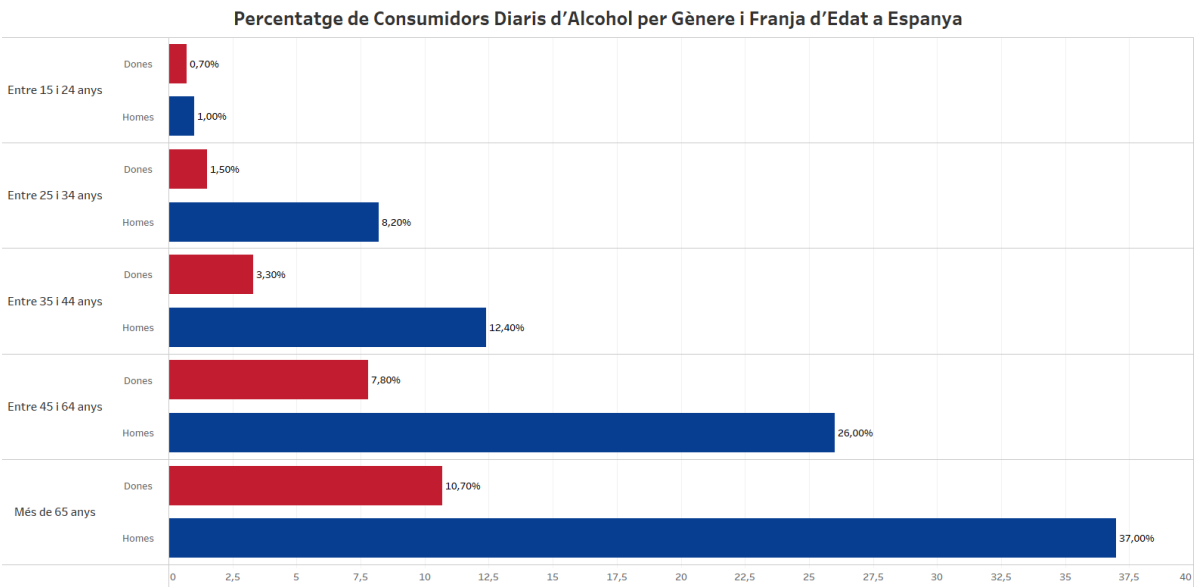


Il·lustració 31: “Percentatge de Consumidors Diaris d'Alcohol per País a Europa”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Frequency of alcohol consumption by sex, age and educational attainment level (2024).

Els països que registren els valors més baixos són Turquia com el que menys amb tan sols un 0,5% de la població total, seguit de Lituània amb un 0,8% i d'Islàndia amb un 1%. Aquests baixos percentatges poden deure's a diversos factors com polítiques estrictes sobre el consum d'alcohol, influències culturals o religioses que rebutgen el consum d'alcohol com a Turquia, o accés més difícil a begudes alcohòliques com països nòrdics com Islàndia o Noruega.

Els països amb els percentatges més alts de consumidors diaris d'alcohol, d'altra banda, són Portugal com el que més amb un 20,7%, seguit d'Espanya amb un 13% i, en tercer lloc, Itàlia amb un 12,1%. Aquests alts percentatges poden ser influenciats per la cultura de consum d'alcohol en aquests països, on beure vi o cervesa pot ser part de la vida quotidiana i de la dieta habitual. En països mediterranis com Espanya, Itàlia i Portugal, el consum d'alcohol és una tradició social profundament arrelada en la cultura.

Tenint en compte que Espanya és el segon país europeu amb més percentatge de consumidors diaris d'alcohol a Europa, el següent gràfic mostra aquesta dada desglossada per franges d'edat i gènere.

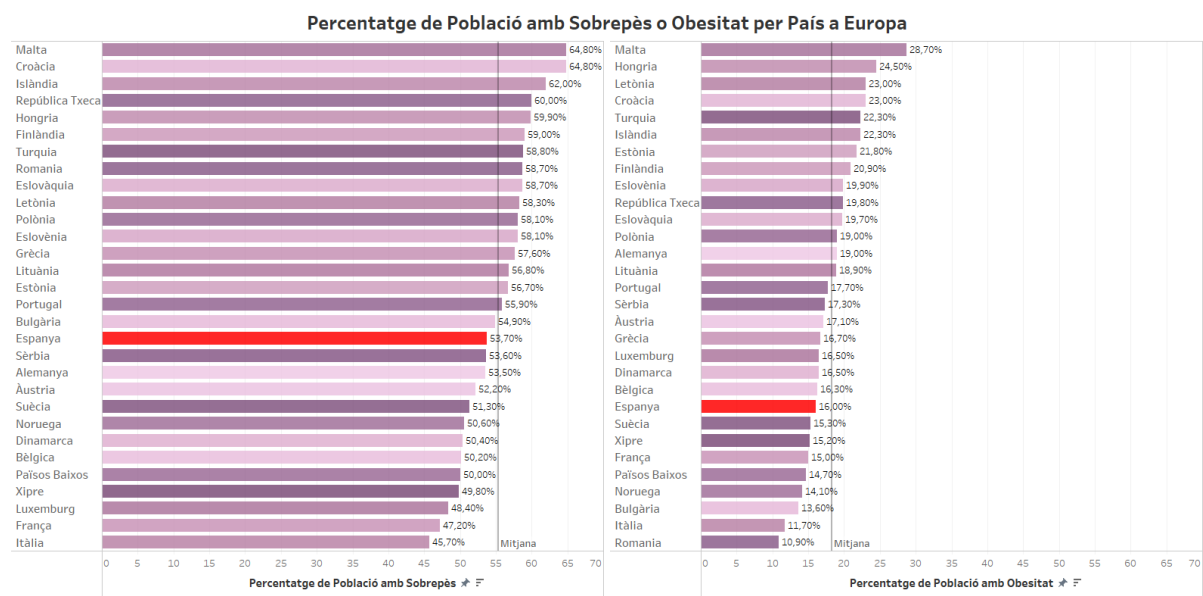


Il·lustració 32: “Percentatge de Consumidors Diaris d'Alcohol per Gènere i Franja d'Edat a Espanya”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Frequency of alcohol consumption by sex, age and educational attainment level (2024).

A grans trets, la tendència general mostra un augment de la població consumidora d'alcohol a mesura que augmenta l'edat, anant del 0,8% en joves de 15 a 24 anys, fins al 22,3% en els majors de 65 anys. A més, els homes presenten valors molt per sobre de les dones, sent aproximadament quatre vegades superiors en 4 de les 5 franges d'edat, a excepció dels més joves.

Aquests valors es deuen a una gran varietat de factors, però com ja s'ha mencionat anteriorment, no és estrany que a Espanya, un país on la cultura de l'alcohol està tan arrelada i en el qual és comú el consum regular de vi o cervesa amb els àpats, aquests valors siguin superiors. Quant a les diferències de gènere, els homes històricament han tingut més permissivitat social per consumir alcohol que les dones, reflectint-se així en els percentatges totals.

I per tancar la secció, es troben les dades referents a la tercera problemàtica en matèria de salut, l'obesitat. Els gràfics a continuació mostren el percentatge de població amb sobrepès o Obesitat a cada país d'Europa segons l'índex de massa corporal de les persones. L'índex de massa corporal (IMC) mesura el contingut de greix corporal en relació amb l'alçada i el pes i es calcula dividint el pes en quilos dividit pel quadrat de l'alçada en metres, on es consideren amb sobrepès les persones amb un índex de massa corporal igual o superior a 25 i obeses aquelles amb un índex de massa corporal igual o superior a 30.



Il·lustració 33: “Percentatge de Població amb Sobrepès o Obesitat per País a Europa”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Obesity rate by body mass index (BMI) (2024).

En termes de sobrepès, Malta i Croàcia presenten els percentatges més alts de població, amb un 64,8%. D'altra banda, Itàlia té el percentatge més baix, amb un 45,7%, seguida per França, amb un 47,2%. Espanya, amb un 53,7%, es troba lleugerament per sota de la mitjana europea en aquest aspecte.

Pel que fa a l'obesitat, Malta té el percentatge més alt, amb un 28,7%, seguida per Hongria, amb un 24,5%, i tant Croàcia com Letònia, amb un 23,0%. Romania té el percentatge més baix d'obesitat, amb un 10,9%, seguida per Itàlia, amb un 11,7%. Espanya, amb un 16,0%, també està per sota de la mitjana europea en obesitat.

Aquestes dades, juntament amb les que mostra el gràfic, porten a una sèrie d'interpretacions. En primer lloc, països com Espanya, Itàlia i França, que presenten els índexs més baixos de sobrepès i obesitat, són coneguts per les seves dietes mediterrànies i estils de vida actius, la qual cosa contribueix significativament a aquests resultats. Així mateix, els països amb millor educació i major consciència sobre la salut, així com economies més fortes que faciliten un major accés a aliments saludables i activitats físiques, com Noruega, Dinamarca i els Països Baixos, tendeixen a tenir menors percentatges d'obesitat.

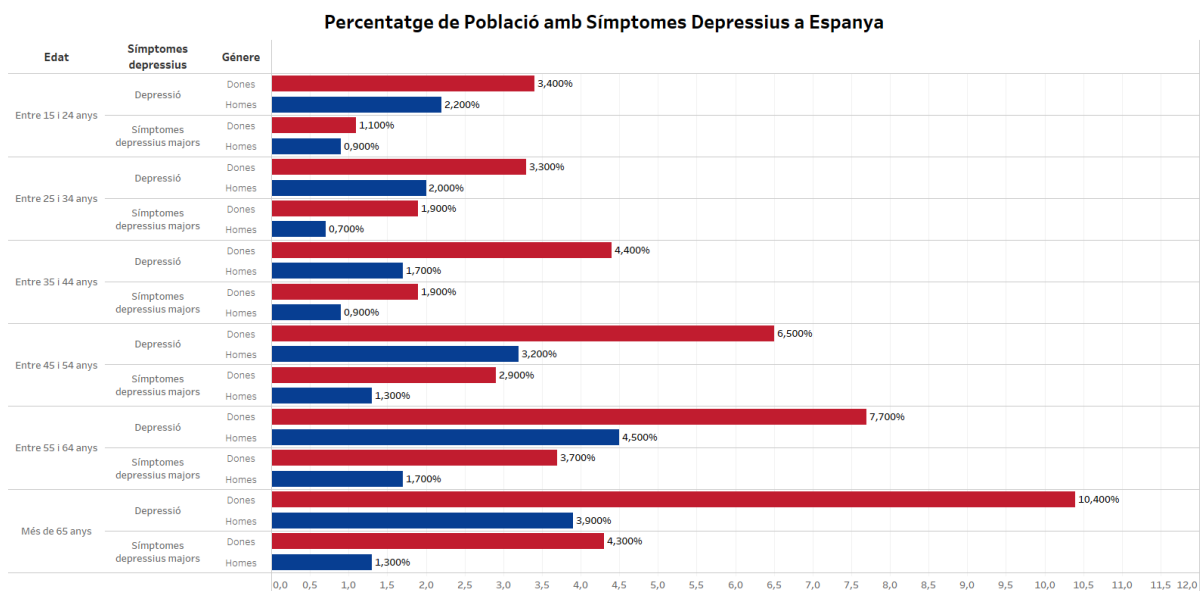
Espanya, per la seva banda, es beneficia de la seva dieta tradicional i estil de vida, cosa que li permet mantenir els seus valors per sota de la mitjana europea.

És evident que l'adopció d'hàbits de salut perjudicials com el tabaquisme, l'alcoholisme i l'obesitat té un impacte directament proporcional a l'alta incidència de malalties cròniques a Espanya, sigui causant-les o agreujant-les. Aquests comportaments de risc no només afecten la qualitat de vida individual sinó que també imposen una càrrega significativa sobre el sistema de salut pública, per la qual cosa cal promoure estils de vida saludables per reduir la incidència d'aquestes malalties i millorar el benestar general de la població.

4.5. Salut mental

Recuperant les dades vistes amb anterioritat al gràfic referent als problemes crònics de salut als quals s'enfronta la societat espanyola, on gairebé una cinquena part de la població pateix problemes de salut mental, aquesta secció es centra a abordar aquesta realitat presentant les principals dades relatives a la salut mental i les seves conseqüències amb tres gràfics.

El primer gràfic mostra les dades pel que fa a símptomes depressius a Espanya, indicant el percentatge de població que pateix tant depressió com símptomes depressius majors, separat per gènere i per franja d'edat.



Il·lustració 34: “Percentatge de Població amb Símtomes Depressius a Espanya”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Current depressive symptoms by sex, age and income quintile (2024).

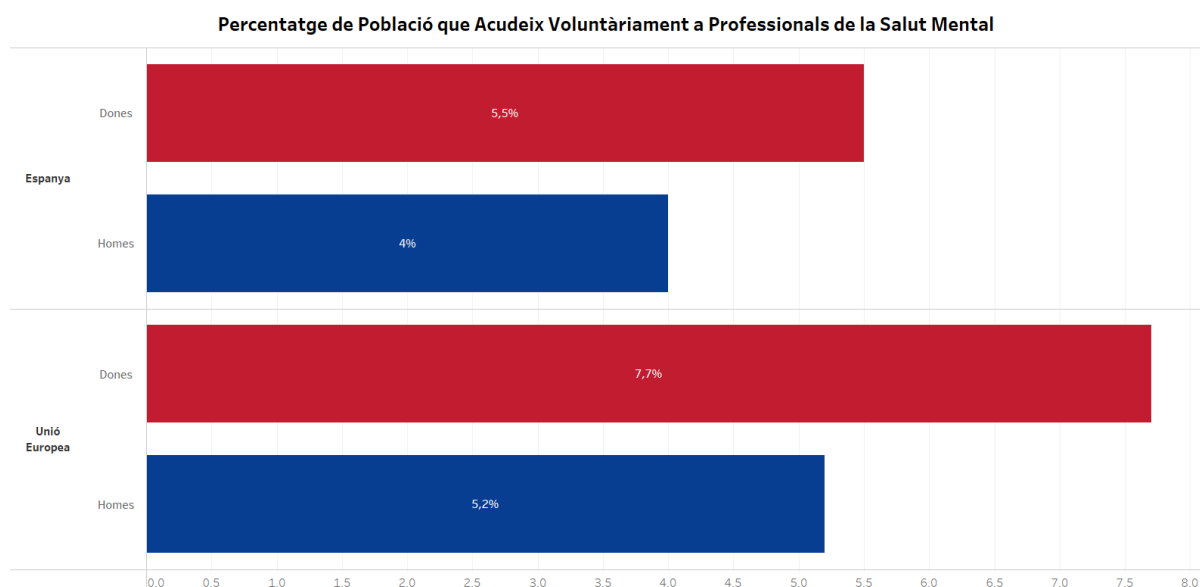
L'anàlisi de les dades relatives a la prevalença de símptomes depressius a Espanya revela diferències significatives entre gèneres i grups d'edat. En general, les dones tendeixen a patir símptomes depressius amb més freqüència que els homes. La mitjana d'incidència és del 5,95% per a dones i del 2,9% per a homes, fet que indica que gairebé el doble de dones es veuen afectades per aquests símptomes en comparació amb els homes.

A més, s'observa una tendència creixent de símptomes depressius a mesura que les persones envelleixen. Aquesta tendència és especialment pronunciada en les dones majors de 65 anys, on un alarmant 10,4% d'elles pateix depressió. Aquest grup demogràfic no només mostra una major prevalença de símptomes depressius, sinó també una major gravetat d'aquests símptomes.

En concret, el 4,3% de les dones majors de 65 anys i el 3,7% de les dones entre 55 i 64 anys presenten símptomes depressius majors, fet que probablement implica la necessitat de medicació.

Aquesta informació subratlla la importància de centrar els esforços de salut mental en la població femenina, especialment en les dones majors, que semblen estar en una situació de major vulnerabilitat.

El segon gràfic proporciona dades aparentment simples, però realment rellevants sobre la proporció de persones que acudeixen voluntàriament a professionals de la salut mental, com psicòlegs, psicoterapeutes o psiquiatres, tant a Espanya com a la Unió Europea, desglossat per gènere.



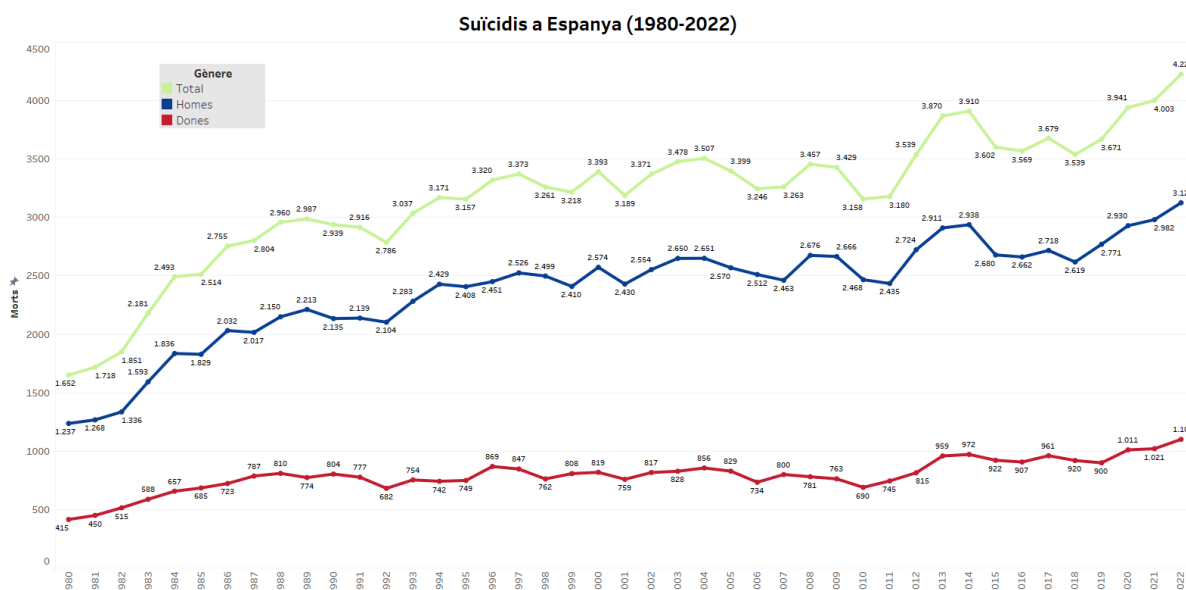
Il·lustració 35: “Percentatge de Població que Acudeix Voluntàriament a Professionals de la Salut Mental”, Font: Elaboració Pròpia a partir de dades extretes de Eurostat, Self-reported consultation of mental healthcare or rehabilitative care professionals (2024).

A Espanya, un 5,5% de la població femenina busca ajuda professional en comparació amb un 4,0% de la població masculina. En l'àmbit europeu, aquests percentatges són encara més grans, ja que un 7,7% de les dones i un 5,2% dels homes acudeixen a professionals de la salut mental de manera voluntària. En ambdós contextos, s'observa que les dones tenen una major propensió a buscar ajuda professional que els homes; tanmateix, la bretxa de gènere és més gran a la Unió Europea que a Espanya.

El fet que més dones a Espanya busquin assistència en salut mental en comparació amb els homes reflecteix una resposta comprensible davant la seva major vulnerabilitat als símptomes depressius. Aquest patró també s'observa a la Unió Europea, encara que amb percentatges més alts i una bretxa de gènere més pronunciada, fet que podria indicar una major consciència i disponibilitat de serveis de salut mental en l'àmbit europeu.

El fet que la cerca d'ajuda a la Unió Europea sigui més gran que a Espanya, suggereix que hi ha una major acceptació i menor estigmatització dels problemes de salut mental, la qual cosa podria ser un factor important per fomentar una major utilització de serveis de salut mental a Espanya. És fonamental que es continuïn promovent polítiques i campanyes de conscienciació per millorar l'accessibilitat i l'acceptació de la salut mental com a component essencial del benestar general.

I finalment, el tercer gràfic mostra els suïcidis anuals a Espanya entre 1980 i 2022, separats per homes, dones i el nombre total d'ambdós gèneres sumats.



Il·lustració 36: "Suïcidis a Espanya (1980-2022)", Font: Elaboració Pròpia a partir de dades extretes de Instituto Nacional de Estadística, Defunciones según la causa de muerte (2024).

Les dades mostren un augment gradual en el nombre total de suïcidis durant aquests anys, passant de 1.652 morts el 1980 a 4.227 morts el 2022. Encara que s'observen fluctuacions al llarg dels anys, destaca un increment notable a partir de l'any 2010, registrant-se un augment constant en els suïcidis totals entre 2018 i 2022, passant de 3.539 el 2018 a 4.227 el 2022.

Una observació important és la marcada diferència de gènere en aquestes estadístiques. Els homes registren valors tres vegades superiors als de les dones. El 1980 es van reportar 1.237 suïcidis entre homes enfront de 415 entre dones, i el 2022, aquests nombres van augmentar a 3.126 suïcidis entre homes i 1.101 entre dones.

La conclusió és clara, i és que el nombre de suïcidis ha augmentat significativament en les últimes dècades, especialment entre els homes, mostrant que els problemes de salut mental estan en ascens aquests últims anys i que els mecanismes actuals de suport i prevenció no són suficients.

Tenint en compte la relació entre els tres gràfics, les dades revelen una situació alarmant i complexa de la salut mental a Espanya. Encara que els homes mostrin una major propensió a patir símptomes depressius, són els que més posen fi a la seva vida, mostrant una major taxa de suïcidis.

Aquestes dades analitzades en aquest apartat de salut mental subratllen la importància crítica d'aquesta relativament recent realitat. L'alta prevalença de símptomes depressius juntament amb la creixent taxa de suïcidis entre la població espanyola destaca la urgent necessitat de millorar els serveis de salut mental i les estratègies de prevenció del suïcidi. És essencial implementar polítiques que promoguin la intervenció psicològica en etapes inicials i reduir l'estigmatització associada amb la cerca d'ajuda per a problemes de salut mental. Només a través d'un enfocament integral i proactiu es podrà abordar adequadament aquesta creixent crisi de salut mental a Espanya.

5. Conclusions

Aquest treball ha tingut com a objectiu principal analitzar diversos indicadors demogràfics i sanitaris a Espanya, per tal de comprendre millor els reptes que enfronta la població espanyola en termes de població, salut i benestar. I aquest apartat busca donar resposta a les preguntes d'investigació formulades al principi del projecte, així com aportar una sèrie de conclusions i reflexions de tot el procés. És per això, que les conclusions es dividiran en dos; les conclusions extretes a l'anàlisi per tal de respondre a les preguntes d'investigació de manera clara i concisa i les conclusions generals de tot el treball.

5.1. Conclusions de l'anàlisi

A continuació es recuperen les preguntes d'investigació inicials i s'exposen les principals conclusions de cadascuna.

Com ha evolucionat el panorama demogràfic a Espanya en els últims anys i quins factors han pogut influir en aquests canvis?

Des de 1960, Espanya ha experimentat un creixement poblacional positiu constant, passant de 30,3 milions d'habitants a 48,5 milions el 2024. Aquest creixement es va viure en major mesura fins al 2010, motivat per diferents factors segons el període, entre ells una sèrie de polítiques pronatalistes, una transició socioeconòmica i una creixent immigració de població jove i activa.

Ara bé, aquest creixement de la població a Espanya ha mostrat una clara tendència cap a l'envelliment. Aquest fenomen ha estat influït per diversos factors clau, incloent-hi una baixa taxa de natalitat i una alta esperança de vida.

La taxa de natalitat a Espanya és insuficient per al reemplaçament generacional, la qual cosa subratlla la necessitat de polítiques que fomentin la natalitat, així com la correcta integració d'immigrants joves.

Per contrarestar l'envelliment de la població, és necessari adaptar els sistemes de salut i pensions per gestionar eficaçment una població major, assegurant sostenibilitat i accessibilitat.

Aquestes polítiques públiques han de centrar-se en la planificació a llarg termini per abordar aquests desafiaments demogràfics i garantir el benestar d'una població cada vegada més envellida.

Quina esperança de vida té la població espanyola actualment i com d'alta és comparada amb altres països europeus?

Pel que fa a l'esperança de vida a Espanya, hem vist com ha seguit una tendència positiva durant les últimes dècades, passant de 73,4 anys el 1975 a 84 el 2023. Aquesta té una sèrie de diferències segons el gènere, sent de 70,5 anys per als homes i 76,3 per a les dones el 1975, i actualment de 81,2 per als homes i 86,7 per a les dones, una diferència significativa de més de cinc anys.

Comparada amb la resta d'Europa, Espanya surt molt ben parada. Se situa com el segon país amb més esperança de vida a la Unió Europea, només darrere de Suïssa, i segons les projeccions, el 2100 serà el primer en la llista. Aquesta xifra tan alta d'esperança de vida és un reflex dels avenços mèdics i el gran sistema de salut pública amb el que compta Espanya, així com les bones condicions de vida i l'alimentació al país.

Espanya destaca per la seva longevitat respecte altres països europeus, encara que aquest assoliment també ve acompanyat de reptes relacionats amb l'envelliment de la població i la necessitat d'adaptar les polítiques de salut pública per mantenir i millorar la qualitat de vida en la tercera edat, sobretot en un context d'envelliment poblacional.

Va afectar la pandèmia de la COVID-19 pel que fa a l'esperança de vida i els anys de vida saludables?

La pandèmia de COVID-19 va tenir un impacte negatiu significatiu en l'esperança de vida i els anys de vida saludables a Espanya. El 2020, l'esperança de vida va caure notablement a causa de l'alta mortalitat causada per la pandèmia. No obstant això, el 2021, l'esperança de vida va recuperar els seus nivells previs a la pandèmia, demostrant així la robustesa del sistema sanitari i l'eficàcia de les mesures implementades per combatre el virus.

Pel que fa als anys de vida saludables, aquests també es van veure afectats. El 2019, la mitjana d'anys de vida saludables a Espanya va arribar a un màxim de 69,9 anys, valor que va baixar l'any següent a la pandèmia a 66,3 anys el 2021, similar al nivell de fa quinze anys. Això demostra tant l'impacte directe del virus com les seves conseqüències indirectes, com el deteriorament de la salut mental, els retards en la detecció i el tractament de malalties cròniques i la disminució de l'activitat física a causa de les restriccions.

Quines van ser les principals causes de mort a Espanya després de la pandèmia i quins canvis hi va haver en anys posteriors?

L'any després de la pandèmia, el 2021, les principals causes de mort van ser malalties de l'aparell circulatori (119.196 defuncions), tumors i càncers (113.662 defuncions), i malalties infeccioses i parasitàries, incloent-hi la COVID-19 (46.001 defuncions). El 2022, les causes principals van ser malalties de l'aparell circulatori (121.341 defuncions), tumors i càncers (114.828 defuncions), i malalties de l'aparell respiratori (43.024 defuncions).

Els canvis més significatius en aquests anys van ser la disminució de morts per malalties infeccioses i parasitàries, amb un -16,1% de variació negativa, deixant en tercera posició les morts per malalties de l'aparell respiratori, que van registrar un 21% de variació positiva, i un augment en els trastorns mentals i del comportament com a causa de mort, amb una variació del 10,9%.

Aquestes dues últimes tendències suggereixen diverses coses. En primer lloc, que d'un any per l'altre es pot afirmar que es va aconseguir neutralitzar el virus, reafirmant que a Espanya comptem amb un gran sistema sanitari i que les mesures per contrarestar la pandèmia, com les campanyes de vacunació i les restriccions, van funcionar. Ara bé, aquesta baixada en la categoria de malalties infeccioses i parasitàries va implicar directament la pujada de la mortalitat en la categoria de malalties de l'aparell respiratori, suggerint que els efectes de la pandèmia han tingut un efecte retardat en el temps i s'han manifestat posteriorment, de manera que molta gent que va vèncer el virus ha pogut morir per les seves seqüeles.

Finalment, l'augment del gairebé 11% de les morts a causa de trastorns mentals i del comportament ressalta més que mai la necessitat de centrar-se en la salut mental de les persones. Aquest augment reflecteix possiblement l'impacte prolongat dels efectes de la pandèmia en la salut mental de la gent, per factors com l'aïllament social i la falta d'activitat física. Aquesta impactant xifra de gairebé 23.000 morts, juntament amb el 19,5% de població afectada per problemes crònics de salut mental, ha motivat una petita investigació referent al tema que no estava plantejada des d'un principi. Avui dia, els joves sobretot estan enfrontant nivells sense precedents d'estrès i ansietat a causa de factors com la sobreexposició a les xarxes socials, la manca de suport emocional i la incertesa econòmica. Factors que no va portar la pandèmia, però que sí que va ser responsable d'incrementar.

Quins són els principals factors de risc per a la salut de les persones que poden esdevenir en problemes crònics de salut i com es compara Espanya amb la resta de països d'Europa respecte a aquests?

Els principals factors de risc per a la salut a Espanya que poden esdevenir en problemes crònics de salut són el tabaquisme, l'alcoholisme i l'obesitat.

En primer lloc, el tabaquisme és una de les principals causes de la malaltia pulmonar obstructiva crònica (MPOC) i un factor de risc significatiu en contribuir a l'empitjorament de malalties cardiovasculars, així com a la causa de càncers i tumors en el sistema respiratori. A Espanya, malgrat la disminució d'aquesta pràctica nociva amb el pas del temps i les polítiques constants per reduir el tabaquisme, el 19,8% de la població continua fumant diàriament, situant-se per sobre de la mitjana en comparació a la resta de països europeus.

D'altra banda, l'alcoholisme té un major impacte en la població respecte a la resta d'Europa, sent Espanya el segon país amb un major percentatge de consumidors diaris d'alcohol, amb un 13,8% del total de la població, valor que incrementa amb l'edat i que té un impacte significativament major en homes que en dones. Aquest factor es pot atribuir a l'arrelada que està la consumició d'alcohol en la nostra cultura i dieta, i que si no es té precaució, aquesta tendència a l'alcoholisme pot contribuir a patir malalties hepàtiques, així com trastorns del comportament.

El tercer d'aquests factors de risc és l'obesitat, una malaltia que afecta sense fer gaire soroll i que està estretament relacionada amb malalties cardiovasculars, la diabetis i diversos problemes ortopèdics i metabòlics. Espanya té un 16% de la seva població total considerada obesa, valor lleugerament inferior a la mitjana europea, beneficiant-se de la seva dieta mediterrània i estil de vida. Tot i això, aquesta és una malaltia complexa que cal abordar de manera integral promovent un estil de vida saludable i actiu, ja que les seves conseqüències van molt més enllà de l'aspecte físic de les persones.

5.2. Conclusions generals

Finalment, aquest segon apartat serveix per fer una reflexió general de totes les conclusions extretes del treball de fi de grau. Respecte als objectius marcats inicialment, personalment crec que he aconseguit fer una bona i exhaustiva exploració del panorama demogràfic i sanitari

espanyol. Al llarg d'aquesta investigació, he aplicat de manera efectiva els coneixements teòrics adquirits sobre anàlisi i visualització de dades, desenvolupant una investigació ordenada i coherent. He respectat en tot moment les consideracions ètiques pertinents i he promogut l'objectivitat i la qualitat de les dades analitzades, cosa que ha permès obtenir resultats precisos i valuosos.

Un dels aspectes més destacats i reveladors del procés d'anàlisi ha estat la fase de neteja de les dades. He pogut comprovar de primera mà que aquesta etapa és crucial i requereix una inversió significativa de temps i esforç. Una neteja minuciosa i a mida de les dades per molt grans que puguin arribar a ser en segons quines ocasions, és fonamental per evitar problemes en fases posteriors de l'anàlisi, assegurant així la precisió i la fiabilitat dels resultats obtinguts.

Al llarg de la investigació també m'he trobat amb diverses limitacions a l'hora d'accedir a determinades dades. En primer lloc, he constatat en primera persona la dificultat que sovint té accedir a dades de qualitat i actualitzades, donat que la qualitat i l'actualitat d'aquestes en certs portals de dades públiques i gratuïtes a vegades és limitada. Aquest és un desafiament particular per als estudiants universitaris, els quals ens hem d'adaptar a les dades disponibles a Internet, ja que no disposem d'un pressupost elevat. Ara bé, tot i aquestes restriccions, he comprovat que amb dedicació, temps i la intenció de realitzar una bona feina, és possible obtenir resultats interessants i significatius.

Una altra troballa important d'aquesta investigació és la necessitat de promoure mesures per abordar l'onada creixent de problemes de salut mental que afecta la població arreu d'Espanya i del món. La investigació ha posat de manifest la importància de desenvolupar i aplicar polítiques i estratègies eficaces per millorar la salut mental dels ciutadans, un aspecte que mereix atenció i acció urgent per part de les autoritats sanitàries i governamentals.

Per acabar, aquest treball ha representat una aportació valuosa al meu desenvolupament acadèmic i professional. M'ha permès aprofundir en els meus interessos acadèmics i ampliar els meus coneixements previs sobre el procés d'anàlisi i de visualització de dades. Al llarg del projecte, m'he submergit en un procés real de recerca, enfrontant-me a diverses limitacions i obstacles. Tot i això, aquestes dificultats m'han proporcionat una valuosa experiència pràctica, ensenyant-me no només una gran quantitat de coneixements teòrics, sinó també com aplicar-los de manera efectiva de cara a un futur professional no molt llunyà.

A títol personal, aquest projecte d'investigació ha reforçat la meva vocació i interès per les dades i el panorama demogràfic i sanitari espanyol i m'ha permès desenvolupar habilitats crucials per presentar informació de manera clara i entenedora al públic. L'experiència adquirida m'ha mostrat la importància de la precisió i l'ètica a la recerca, així com la necessitat d'adaptar-se i perseverar davant de les dificultats.

6. Bibliografia

Almalki, S. (2016). *Integrating Quantitative and Qualitative Data in Mixed Methods Research--Challenges and Benefits*. Journal of education and learning, 5(3), 288-296. <https://doi.org/10.5539/jel.v5n3p288>

Alonso, M., & Furio Blasco, E. (2007). *La modernización de la economía española en el siglo XX: De la autarquía a la adopción del euro*. HAL Archives - Sciences sociales. <https://shs.hal.science/halshs-00137878>

Amer, A. (2005). *Analytical thinking*. Pathways to Higher Education.

Andrews, E. (2023). 8 Remarkable Early Maps. History. <https://www.history.com/news/8-remarkable-early-maps>

Bartlett, G. (2001). *Systemic thinking: a simple thinking technique for gaining systemic focus*. In The International Conference on Thinking" Breakthroughs (pp. 1-14).

Bertin, J. (1967). *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*. Paris: Gauthier-Villars.

Briney, K. (2015). *Data Management for Researchers: Organize, maintain and share your data for research success*. Pelagic Publishing Ltd

Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 77-91). Proceedings of Machine Learning Research, 81. Recuperat de <https://proceedings.mlr.press/v81/buolamwini18a.html>

Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.

Cleveland, W. S., & McGill, R. (1984). *Graphical perception: Theory, experimentation, and application to the development of graphical methods*. Journal of the American Statistical Association, 79(387), 531-554. <https://doi.org/10.2307/2288400>

Codina, L. (2023). *Preguntas de investigación en tesis doctorales y trabajos académicos*. Lluís Codina · Comunicación y Documentación. <https://www.lluiscodina.com/preguntas-de-investigacion-tesis-doctorales/>

Crabtree, M., & Nehme, A. (2024). *¿Qué es la ciencia de datos? Definición, ejemplos, herramientas y más*. DataCamp. <https://www.datacamp.com/es/blog/what-is-data-science-the-definitive-guide>

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Davenport, T. (2006). *Competing on Analytics*. Harvard Business Review, 84(2), 98-107, 134.

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Journal of Management Review*, 70, 35-36.

Ethan, A., & Alexander, D. (2022). *Ensuring Data Integrity in Big Data Environments*. International Journal of Computer Science and Technology, 5(2). <https://doi.org/10.5281/zenodo.8415841>

Farrugia, P., Petrisor, B. A., Farrokhyar, F., & Bhandari, M. (2010). *Research questions, hypotheses and objectives*. Canadian journal of surgery, 53(4), 278.

Ferdio. (2014). *Data Viz Project by Ferdio*. Data Viz Project <https://datavizproject.com/>

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc..

Figueiredo, A. S. (2017). *Data sharing: Convert challenges into opportunities*. Frontiers in Public Health, 5, 327. <https://doi.org/10.3389/fpubh.2017.00327>

Friendly, M. (2006). A brief history of data visualization. *Handbook of Computational Statistics: Data Visualization*. Springer-Verlag.

Funkhouser, H. G. (1937). *Historical Development of the Graphical Representation of Statistical Data*. Osiris, 3, 269–404. <http://www.jstor.org/stable/301591>

Gobierno de España (1960). *Ley 45/1960, de 21 de julio, por la que se crean determinados Fondos Nacionales para la aplicación social del Impuesto y del Ahorro*. Boletín Oficial del Estado, núm. 176, 10285-10289. <https://www.boe.es/datos/pdfs/BOE/1960/176/A10285-10289.pdf>

Google Career Certificates. (2021). *Análisis de Datos de Google Professional Certificate*. Coursera. <https://www.coursera.org/professional-certificates/analisis-de-datos-de-google>

Heer, J., & Bostock, M. (2010). *Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design*. Conference on Human Factors in Computing Systems - Proceedings, 1, 203-212. <https://doi.org/10.1145/1753326.1753357>

IBM. (2024). *What is data lifecycle management (DLM)?* IBM. <https://www.ibm.com/topics/data-lifecycle-management>

International Paralympic Committee. (2016). *Results Archive - Rio 2016 - Swimming - Womens 100 M Breaststroke Sb5*. International Paralympic Committee <https://www.paralympic.org/rio-2016/results/swimming/womens-100-m-breaststroke-sb5>

McCandless, D. (2014). *Knowledge is beautiful*. London: William Collins.

McCandless, D. (2024). *The McCandless method – What makes a good visualization?* Information is Beautiful. <https://informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/>

National Centers for Environmental Information, National Oceanic and Atmospheric Administration. (2024). *Climate at a Glance: Global Time Series*. National Centers for Environmental Information, National Oceanic and Atmospheric Administration. <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/global/time-series>

Neumann, N., Tucker, C. E., Subramanyam, K., & Marshall, J. (2023). *Is first-or third-party audience data more effective for reaching the 'right' customers? The case of IT decision-makers*. Quantitative Marketing and Economics, 21(4), 519-571. <https://doi.org/10.1007/s11129-023-09268-7>

Nguyen, C. T. (2024). *The Limits of Data*. Issues in Science and Technology, 40(2), 94–101. <https://doi.org/10.58875/LUXD6515>

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 366(6464), 447-453. <https://doi.org/10.1126/science.aax2342>

Quantum Analytics NG. (2022). *6 Phases of Data Analysis According to Google*. LinkedIn. <https://www.linkedin.com/pulse/6-phases-data-analysis-according-google-quantum-analytics-ng/>

Sandberg, M. (2013). *Dataviz history: Charles Minard's flow map of Napoleon's Russian campaign of 1812*. Michael Sandberg's Data Visualization Blog. <https://datavizblog.com/2013/05/26/dataviz-history-charles-minards-flow-map-of-napoleons-russian-campaign-of-1812-part-5/>

Shorten, A., & Smith, J. (2017). *Mixed methods research: Expanding the evidence base*. Evidence Based Nursing, 20(3), 74-75. <https://doi.org/10.1136/eb-2017-102699>

Tableau. (2023). *What Is Tableau?* Tableau.com. <https://www.tableau.com/why-tableau/what-istableau>

Talabis, M., & Martin, J. (2012). *Information Security Risk Assessment Toolkit: Practical Assessments through Data Collection and Data Analysis*. Syngress. <https://doi.org/10.1016/B978-1-59-749735-0.00004-X>

Treesak, Y. (2024). *10 common data visualization mistakes and how to avoid them*. Agoda Engineering & Design. <https://medium.com/agoda-engineering/10-common-data-visualization-mistakes-and-how-to-avoid-them-e3896fe8e104>

Tufte, E. (2001). *The Visual Display of Quantitative Information*. American Journal of Physics, 31.

Wickham, H. (2014). *Tidy data*. Journal of statistical software, 59, 1-23. <https://doi.org/10.18637/jss.v059.i10>

Wikipedia. (2024). *Europe*. Wikipedia, The Free Encyclopedia. <https://en.wikipedia.org/wiki/Europe>

World Health Organization. (2009). *Global health risks: mortality and burden of disease attributable to selected major risks*. World Health Organization.
<https://iris.who.int/handle/10665/44203>

