

DATUEN BISTARATZEA ESTATISTIKA OFIZIALETAN

Enara Galbete Ahechu



EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

Donostia - San Sebastián, 1
01010 VITORIA - GASTEIZ
Tel.: 945 01 75 00
Fax.: 945 01 75 01
E-mail: eustat@eustat.es
www.eustat.es

Aurkezpena

Eustat-en helburuetako bat informazio estatistikoa era argi eta ulergarrian zabaltzea da. Horretarako, datu estatistikoen bistaratzearen ikerkuntzaren inguruan esfortzu handia egiten ari da, erabiltzaileek datuok ere dinamiko eta interaktiboan aztertu eta erabili ditzaten.

2010ean XXIII Nazioarteko Estatistika Mintegia antolatu zuen, "Datuak grafiko interaktibo dinamikoez aztertzen" izenburupean eta 2 urteko iraupeneko prestakuntza eta ikerketarako beka deitu zuen, zehazki datu estatistikoen bistaratzearen gaiaren inguruan.

Argitalpen honek bekan zehar egindako ikerketa lana ezagutzera emateaz gain, datuak bistartzeko era ezberdinak ezagutu eta erabiltzean interesatuta dauden erabiltzaileei tresna erabilgarriak eskaintzen dizkie.

Dokumentu honek hiru zati ezberdin dauzka. Lehen zatian, grafikoak egin aurretik kontuan izan beharreko zenbait kontzeptu eta irizpide azaltzen dira, bigarrenean egindako grafiko estatiko zein dinamikoen katalogoa aurkezten da eta hirugarrenean produkzio-prozesuetarako zein Eustateko web orrian argitaratzeko aukeraturiko grafikoaren aplikazioak eta prozedurak azaltzen dira.

Vitoria-Gasteiz, 2013ko Urtarrila.

Josu Iradi Arrieta

EUSTATEko Zuzendari Nagusia

Aurkibidea

AURKEZPENA	2
AURKIBIDEA	3
SARRERA	4
METODOLOGIA	5
DATU-MOTA EZBERDINAK	5
KONTZEPTUALIZAZIOA	6
POPULAZIOA	6
BEHAKETA ESPAZIOA	7
ALDAGAIK	9
GRAFIKOA IRUDIKATU AURRETIK KONTUAN HARTZEKOAK	9
GRAFIKOEN SAILKAPENA	11
ALDAGAI BAKARREKO DATUAK	11
ALDAGAI BIKO DATUAK	13
ALDAGAI ANITZEKO DATUAK	15
DENBORAZKO DATUAK	18
DATU ESPAZIALAK	20
APLIKAZIOAK	21
PRODUKZIORAKO GRAFIKOAK	21
1. Dotplot edo puntu diagrama	21
2. Barra-diagramen grafiko konposatua	23
3. Korrespondentzia-analisiaren irudikapen grafiko	24
4. Zatikako erregresio lineala, sektoreka	26
5. Scatterplot matrix grafiko	28
ZABALKUNDERAKO GRAFIKOAK	30
1. Motion Chart grafiko	32
2. Time Line motako grafiko	38
3. Google Earth bidezko bistaratzea	41
ONDORIOAK	46
BIBLIOGRAFIA	47

Sarrera

Koaderno Tekniko honetan bilduriko edukia, Euskal Estatistika Erakundeak 2010. urtean metodologia estatistiko-matematikoen esparruko prestakuntza eta ikerketarako bekaren baitan, Datuen Bistaratzearen ingurukoan hain zuzen, egindako lanaren fruitua da.

Estatistika-ekoizpen ofizialaren metodologia estatistiko-matematikoen esparruan 2 urteko iraupeneko prestakuntza eta ikerketarako beka deitu zen, zehazki datu estatistikoen bistaratzearen gaiaren inguruan.

Dokumentu hau hurrengo ataletan banatuta dago:

Lehen kapituluan sarrera bat eta Koaderno Tekniko hau egiteko ezarritako helburuak aipatzen dira.

Bigarren kapituluan, datu-multzo batetik abiatuz grafiko bat egin aurretik ezagutu beharreko zenbait kontzeptu eta kontuan hartu beharreko zenbait irizpide definitzen dira.

Hirugarren kapituluan datu-mota ezberdinetarako diseinatu diren grafikoaren sailkapen bat ematen da.

Laugarren kapituluan, EUSTAT-ek erabiltzeko automatizatu diren grafikoak aipatzen dira, batzuk produkziarako grafiko estatikoak, eta besteak zabalkunderako grafiko dinamiko eta interaktiboak.

Azkenik datuak bistaratzeko metodo eta teknikekin erlazionaturiko zenbait ondorio aipatzen dira.

Bereziki eskertu nahiko nuke Eustat-eko Metodologia, Berrikuntza eta I+G Arloko kide guztien eta bistaratzeko taldeko kide guztien lana, EHUKo Informatika fakultateko Konputazio Zientziak eta Adimen Artifizialaren Saileko Yosu Yurramendi Mendizabalen aholkularitza eta laguntza eta, orokorrean, Eustateko langile guztien adeitasuna .

HITZ KLABEAK: grafiko estatikoak, grafiko dinamikoak, grafiko interaktiboak.

Metodologia

EUSTATEk izaera eta jatorri ezberdineko datu-mota asko dauzka. Datuak aztertu eta aurkezteko era ezberdinak daude, hala nola, grafikoen erabilera. Datu bistaratzearen helburu nagusia datuak begi-bistaz analizatu, sintetizatu eta irudikatzea da, zailtasun eta mota ezberdineko grafikoak erabiliz.

Datu-mota bakoitzari bistaratze ezberdin bat egokitzen zaio, era grafiko eta nahiko intuitibo (datuen eta grafikoaren zailtasunaren arabera) batean informazioa ateratzeko aukera ematen duena.

Datu-mota ezberdinak

EUSTATEk bi motatako datuak ditu; alde batetik, web-ean PC-AXIS formatuan dituenak, eta bestalde mikrodaturen fitxategiak.

Grafikoki irudikatu aurretik datuak sailkatu behar dira, bi irizpideren arabera:

1. DATUEN IZAERA:

Datuen izaeraren arabera, hau da, duten aldagai kopuruaren arabera, 5 mota ezberdinetan sailka daitezke:

- i. **Aldagai bakarreko datuak.**
- ii. **Aldagai biko datuak.**
- iii. **Aldagai anitzeko datuak.** Hiru edo aldagai gehiago dituzten datuak.
- iv. **Denborazko datuak.** Datu, behaketa edo balio sekuentziak dira, denboraren momentu jakinetan neurtuak, kronologikoki ordenatuak eta normalean elkarren artean uniformeki banatuak.
- v. **Datu espazialak.** Espazioko kokapen zehatz batzuei lotutako aldagaiak.

2. DATUEN JATORRIA:

Beren jatorriaren arabera, datuak bi motatan sailka daitezke:

- i. **Maiztasun-taulak.** Datu bakoitzari dagokion maiztasuna esleitzen diote.
- ii. **Mikrodatuak.** Indibiduo bakoitzari aldagai zerrenda bat dagokio.

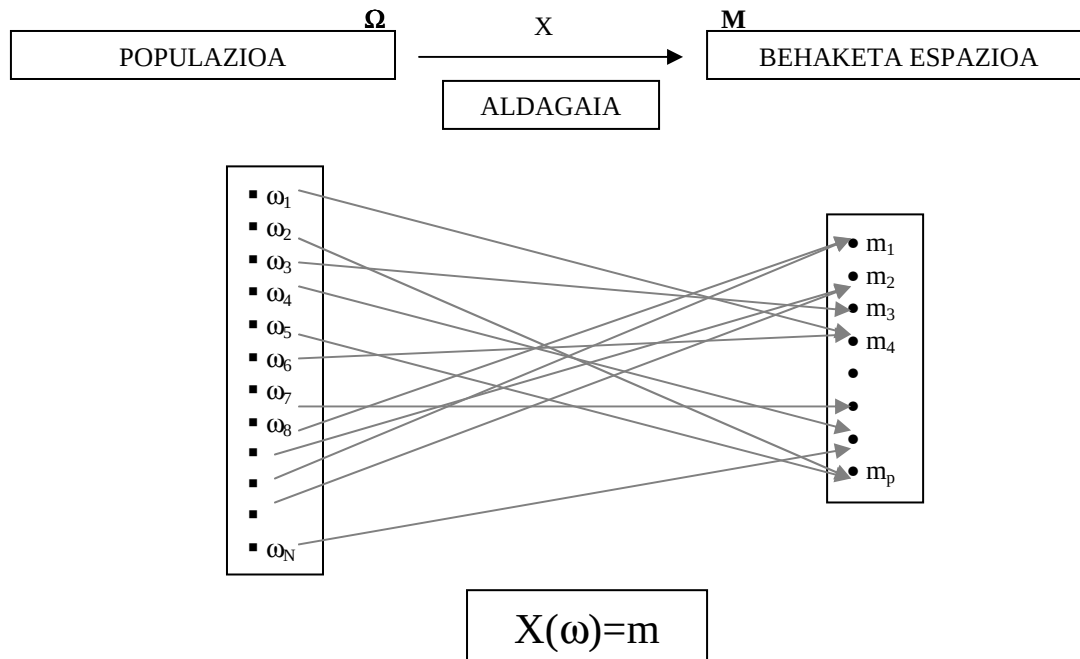
Kontzeptualizazioa

Oinarrizko kontzeptuak hiru dira: **populazioa**, **behaketa** espazioa eta **aldagaia**.

Populazioa: Estatistikoki ikertu nahi diren elementu guztien multzoa da.

Behaketa espazioa: Zorizko esperimentu baten emaitza posible guztien multzoa da.

Aldagaia: Populazio eta behaketa espazioaren arteko aplikazioa da.



Bai populazioa, bai behaketa espazioa mota ezberdinetakoak izan daitezke eta kontuan hartu behar dira grafiko bat eraiki nahi denean. Haien araberakoa da aldagaien izaera baita ere.

Populazioa

Populazioa mota ezberdinetakoa izan daiteke, **izaera**, **tamaina** eta **egituraren** arabera.

1. **Izaeraren** arabera sailkapena:
 - i. **Errolda edo Erregistroak.** Populazio estatistiko bat osatzen duten indibiduen zenbatzeari deritze. Funtsean, populazio estatistiko baten neurketak lortzean datzate, zenbaketa teknika ezberdinak erabiliz edo informazio administratibo bidez.
 - ii. **Lagina.** Populazioaren azpimultzo bat da, laginketa-tekniken bidez ateratzen dena, eta populazio osoaren adierazgarria dena.

2. **Tamainaren** araberako sailkapena:

- i. **Diskretua.** Elementu kopuru finitu edo infinitu zenbakigarria duen populazioari deritzo.
 - **Finitua.** Elementu kopuru finitua duen populazioa.
 - **Infinitua.** Elementu kopuru infinitu zenbakigarria duen populazioa.
- ii. **Jarraitua.** Zorizko aldagaiak ezaugarri jarraitu bat neurtzen duen populazioa.

3. **Egituraren** araberako sailkapena:

Populazioaren elementuak, multzoaren unitateak, erlazionatuta egon daitezke ala ez. Erlazio motaren arabera populazioak egitura bat edo beste du.

- i. **Egiturarik gabea.** Populazioaren elementuak erlazionatuta ez daudenean ematen da. Hau da estatistika klasikoan gehienetan ematen edo suposatzen den kasua. (Adibidez, *laginketa bat egiten denean elementuek, hau da pertsonak, elkarren artean erlazioirik ez dutela suposatzen da*).
- ii. **Sekuentziala.** Populazioaren elementuen artean erabateko ordena definituta dagoenean. Prozesu batean elementu jakin bati zenbait neurri hartzen zaizkionean eta neurrien artean zerbait eragiten zaionean ematen den kasua. (Adibidez, *paziente bati urratsez-urrats eginiko terapia, ikasle baten ikasketa metodoaren urratsak...*).
- iii. **Egitura lineala.** Kasurik arruntena denbora-segida batena da, unitateak bata bestearen atzetik datoz, eta gainera, kasu honetan, unitate bat dago definituta.
- iv. **Mapa.** Bi dimentsiotan (edo gehiagotan) grafo edo mapa batek definitzen du egitura bat (erpina da populazioa, eta ertzek erlazioa definitzen dute).
- v. **Planoa.** Bi dimentsiotan eta unitate bat definituta dagoenean populazioa R^2 planoan kokatzen dugu. (Adibidez, *argazki baten pixelek grafo bat osatzen dute eta gainera neurri bat dago definituta pixelen artean, R^2*).

Behaketa espazioa

Behaketa espazioa ere mota ezberdinetakoa izan daiteke, izaera, tamaina eta egituraren arabera.

1. **Izaeraren** araberako sailkapena:

- i. **Kuantitatiboa.** Oinarrizko gertaera kuantitatiboak soilik dituen behaketa espazioa.

- ii. **Kualitatiboa.** Oinarrizko gertaera kualitatiboak soilik dituen behaketa espazioa.
2. **Tamainaren** araberako sailkapena:
- i. **Diskretua.** Oinarrizko gertaera kopuru finitua edo infinitu zenbakigarria duen behaketa espazioa.
- **Finitua.** Behaketa kopuru finitua duen behaketa-espazioa.
 - **Infinitua.** Behaketa kopuru infinitu zenbakigarria duen behaketa-espazioa.
- ii. **Jarraitua.** Oinarrizko gertaera kopuru infinitu ez-zenbakigarria duen behaketa espazioa.
3. **Egituraren** araberako sailkapena:
- i. **Izenezkoa.** Haien artean inongo ordenarik ez hierarkiarik jarraitzen ez duten behaketez osaturiko behaketa-espazioa.
- ii. **Ordenazkoa.** Aurrez finkaturiko eskala, sekuentzia edo progresio natural bateko balio ordenatu ezberdinak har ditzaketen behaketez osaturiko behaketa-espazioa.
- **Osoa.** Behaketa-espazioa multzo guztiz ordenatua denean. Hau da, multzoko edozein elementu bikote elkarren artean konpara daitezkeenean.
 - **Partziala.** Behaketa-espazioa multzo partzialki ordenatua denean. Hau da, multzoko zenbait elementu bikotetan elementu bat bestearen ondoren doanean. Baina multzoko elementu bikote guztiek ez dute zertan erlazionatuta egon.
- iii. **Zenbakizkoa.** Zenbakizko behaketez osaturiko behaketa-espazioa.

Egitura	Izaera	Kualitatiboa	Kuantitatiboa
Izenezkoa		√	
Ordena partziala		√	
Ordena osoa		√	√
Zenbakizkoa		*	√

Tamaina finitua

* $\{m_1, m_2\} \leftrightarrow \{0,1\}$

Aldagaiak

Aldagai baten izaera dagokion **behaketa espazioaren** izaeraren, tamainaren eta egituraren arabera finkatzen da.

Behaketa-espazioaren arabera, hurrengo aldagai motak bereiz daitezke:

1. **Kualitatiboak.** Kualitate, ezaugarri edo modalitate ezberdinak adierazten dituzten aldagaiak dira. Adierazten den modalitate bakoitzari atributu edo kategoria deritzo eta neurketa aipatutako atributuen sailkapenean datza.
 - i. **Izenezkoak.** Haien artean inolako ordenarik ez hierarkiarik ez duten balioak hartzen dituzten aldagaiak (adibidez, *jaiotze-lekua*).
 - ii. **Ordenazkoak.** Aurrez finkaturiko eskala, sekuentzia edo progresio natural bateko balio ordenatu ezberdinak har ditzaketenak (adibidez, *txikia, ertaina, handia modalitadedun aldagaia*).
2. **Kuantitatiboak.** Zenbaki bidez adieraz daitezkeenak dira (adibidez, *adina, familiaren tamaina, etxebizitzaren azalera...*). Esperimentuetako emaitzak kuantifika daitezke.
 - i. **Diskretuak.** Bakarrik balio osoak hartzen dituzte (*adina, familiaren tamaina*, adibidez).
 - ii. **Jarraituak.** Tarte jakin bateko edozein balio har ditzakete, beraz, dezimalak izan ditzakete (*etxebizitzaren azalera*, adibidez).

Grafikoa irudikatu aurretik kontuan hartzekoak

Grafikoak “onak” izan daitezen eta haien interpretaziotik ahalik eta informazio gehien ateratzeko zenbait irizpide kontuan hartzea komenigarria da. Esaterako, W.S.Cleveland-ek hurrengoak aipatzen ditu, 1985eko “The Elements of Graphing Data” liburuan:

- Ikuspen edo ikuskera garbia:

Datuak. Beharrezkoa ez dena baztertu behar da.

Datuak erakusteko elementu grafiko ikusgarri eta nabariak erabili.

Aldagai bakoitzeko eskala zuzen pare bat erabili. Datu-eremua eskala-zuzenek osatzen duten angelu zuzenaren barruko aldea izango da. Ezarri eskaletako markak datu-eremutik kanpo.

Ez nahaspilatu datu eremua.

Ez jarri marka gehiegi eskaletan.

Erabili erreferentzia-zuzen bat grafikoa zeharkatu behar duen balio bat dagoen aldiro, baina datuekin nahastu gabe.

Ez nahastu datuen etiketak datu kuantitatiboen datu-eremuan.

Datu-eremuan oharrak, giltzak eta seinaleak jartzea ekidin. Giltzak eta seinaleak datu-eremutik kanpo ezarri, eta oharrak legendan edo testuan.

Marraztutako ikurrek, gainjarritz gero, garbiki begiztagarriak izan behar dute.

Datu-multzoek, gainjarritz gero, begiz bereizgarriak izan behar dute.

Ikuste garbitasuna ez da galdu behar grafikoa laburtu eta erreproduzitzuz gero.

- Ulertze garbia.

Ondorio nagusiak era grafikoan jarri. Informazio-legenda ulergarriak egin.

Datuen aldagarritasuna (desbiderapen estandarra, errore estandarra, konfiantza-tartea) adierazten duten errore-barren esanahia garbiki azaldu.

Aldagai baten logaritmoak marrazten direnean eskalen etiketek bat etorri behar dute marken etiketekin.

Grafikoen akatsak zuzendu.

Garbitasuna zaindu.

- Eskalak.

Marken tartea datu guztien edo ia guztien tartea harrapatzeko moduan aukeratu.

Eskalak dituen murriztapenei eutsi eta datu-multzo osoa betetzeko moduan aukeratu.

Batzuetan lagungarri izaten da aldagai batentzat bi eskala-zuzen erabiltzea bi eskala ezberdin erakusteko.

Grafikoak elkarren artean alderatu behar direnean eskala egokiak aukeratu.

Ez tematu magnitude bat adierazten duen eskala batean zeroa beti erakustearekin.

Eskala logaritmikoa erabili faktore biderkatzailearen portzentaje aldakuntza ulertzea inportantea denean edo balio handien alborapena zuzentzeko.

Datuak eskala logaritmoan erakustek, datuak hobeto bereizten lagun dezake.

Eskala bat behar beharrezkoa denean bakarrik eten. Etendura ezin bada saihestu, erabateko eskala-etendura erabili. Ez lotu etenduraren bi aldeetako zenbakizko balioak.

- Estrategia orokorra:

Informazio kuantitatibo asko sar daiteke eremu txiki batean.

Grafikoak marraztea iterazio- edo esperimendu-prozesu bat izan daiteke.

Datuen grafikoa behin baino gehiagotan egin, behar denean.

Grafiko erabilgarri askok azterketa xehe eta arduratsu bat behar dute.

Grafikoen sailkapena

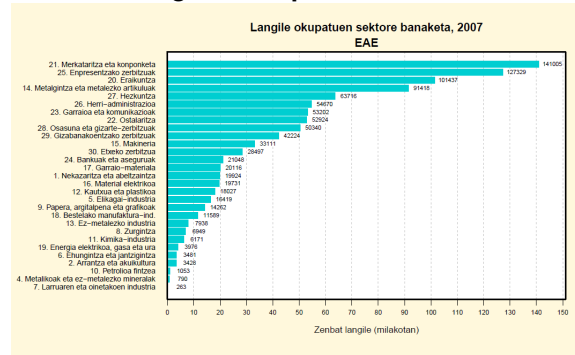
Datuen izaera eta jatorriaren arabera sailkapena hurrengoa da:

Aldagai bakarreko datuak

- Taulak:

i. Barra-diagrama

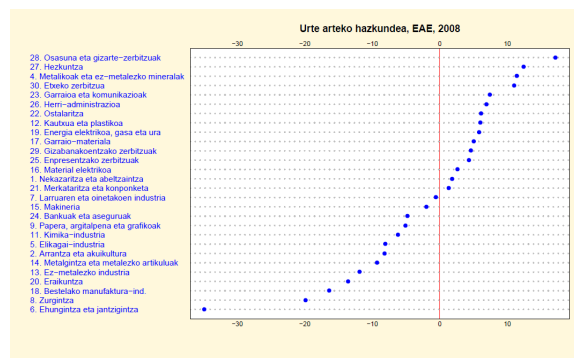
a. Barra-diagrama sinplea.



Aldagai baten modalitateen maiztasunak bistaratu nahi direnean erabiltzen da. Aipaturiko maiztasunak barra paraleloen bidez adierazten dira, bata bestearen gainean eta handienetik txikienera ordenatuta. Interpretazioa errazteko, barra

bakoitzaren eskuinean dagokion modalitatearen maiztasuna gehitzen da, eta baita zuzen bertikal batzuk ere. Posible da barren kolorea zein erabilirik ez dagoen aldatzea.

ii. Puntu-diagrama edo dotplot.

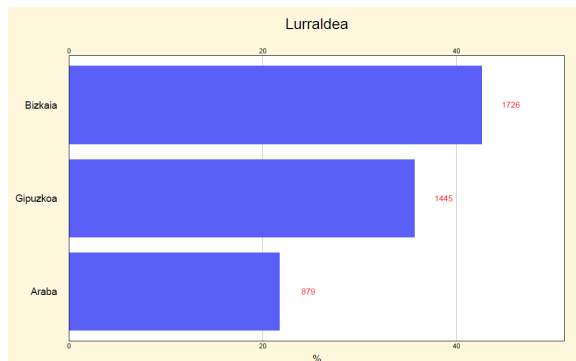


Aldagaiaren modalitateen aldakuntzak bistaratu nahi direnean erabiltzen da. Modalitateak aldakuntzaren arabera handienetik txikienera ordenatzen dira eta puntu bidez adierazi (barra oso bat margotu beharrean puntua erabiltzen da). 0-an kokatzen den zuzen gorriari esker azkar ikus daiteke aldakuntza positibo ala negatiboa den.

Gainera zuzen eten horizontalek interpretazioa errazten dute.

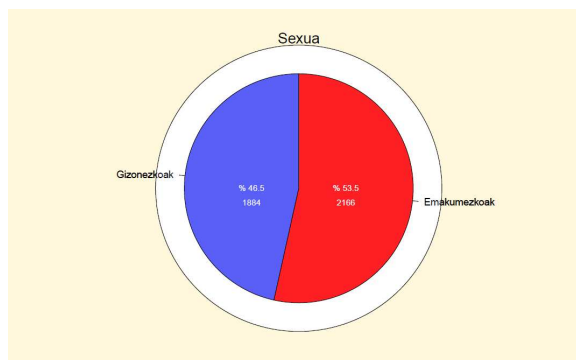
- Mikrodatuak:

- i. Barra-diagrama sinplea.



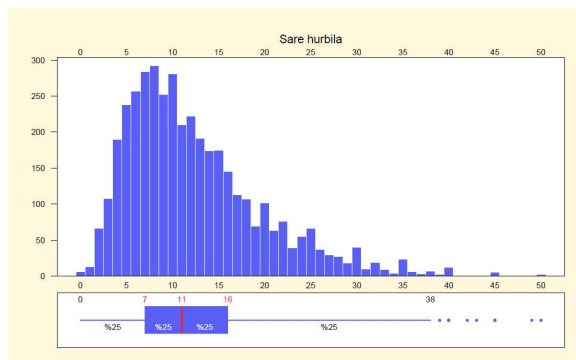
Taulen kasuan erabilitako barra-diagramaren antzekoa da. Kasu honetan, aldagaiaren modalitateek berezko ordena dutenean mantendu egiten da eta ez dutenean handienetik txikienera ordenatzen dira. Aldagaia jarraitua denean barren arteko hutsuneak ezabatzen dira.

- ii. Sektore-diagrama.



Modalitate gutxi dituzten aldagai diskretuak bistartzeko erabiltzen da. Sektore (eta beraz, modalitate) bakoitzari kolore bat ematen zaio, dagokion maiztasuna eta ehunekoak zehaztuz.

- iii. Histograma eta kutxa diagrama.



Aldagai kuantitatibo jarraituak bistartzeko erabiltzen dira. Kutxa diagrama horizontal bat marrazten da, zuzen gorri lodiago batek mediana adierazten duelarik. Gainera, kuartilei dagozkien datuak ere gorri irudikatzen dira eta interpretazioa errazteko kutxa eta whisker edo adarren esanahiari buruzko informazioa gehitzen da.

Egokia da balio arraroen presentzia bistartu eta aldagaiaren banaketa eta simetria aztertzeko ere.

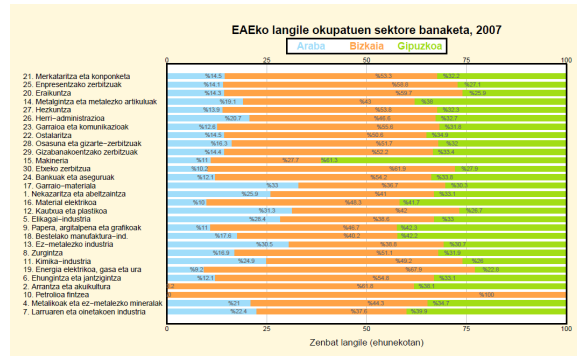
Barra-diagramaren gainean, datuei dagokien histograma irudikatzen da, urdinez; honek aztergai den aldagaiaren banaketaren ideia bat egiten laguntzen du.

Aldagai biko datuak

- Taulak:

- i. Barra-diagrama

- a. Ehunekoen edo maiztasun erlatiboen barra-diagrama.

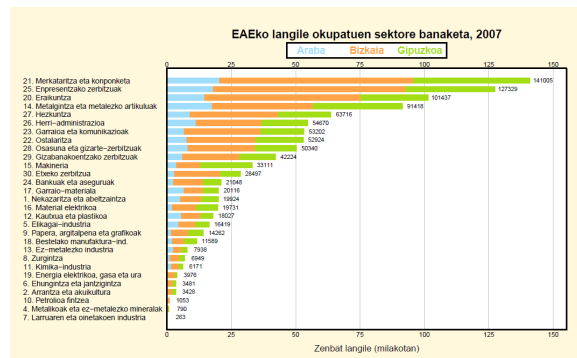


Aldagai kualitatibo baten modalitateen maiztasunak hiru lurren historikoen arabera irudikatzeko erabiltzen da. Lehenik modalitateak handienetik txikienera ordenatzen dira, euren EAEko maiztasunen arabera. Ordena hori mantenduz, hiru lurraldeen ehunekoa

kalkulatu eta barra moduan marrazten dira, bata bestearen ondoan eta kolore ezberdinekin, maiztasun totalaren %100a adierazten duen barra sortuz. Barra-diagrama sinpleei bezala, interpretazioa errazteko zuzen bertikalak gehitzen zaizkie.

Posible da lurralde historikoak erabili beharrean lau modalitate izan eta horietako bat besteetan desglosa daitekeen edozein aldagai kualitatibo erabiltzea.

- b. Barra-diagrama kategorizatua edo maiztasun absolutuen barra-diagrama.

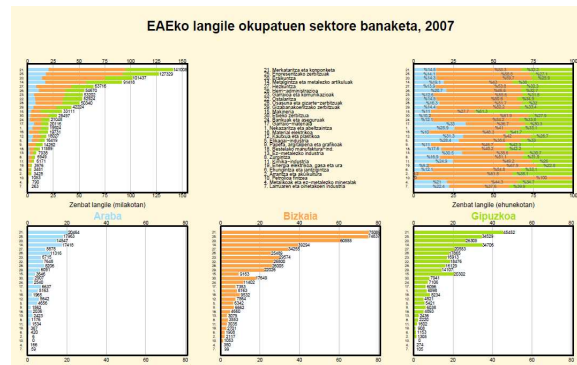


Hau ere, aurrekoa bezala, aldagai kuantitatibo baten modalitateen maiztasunak hiru lurren historikoen arabera irudikatzeko erabiltzen da. Baina kasu honetan barren tamaina dagokion modalitateak EAEn duen maiztasunaren proportzionala da.

Aurreko kasuan bezala, datuak dagokion modalitateak EAEn duen maiztasunen arabera ordenatu eta barra-diagrama sinple bat eratzen da. Ondoren, barra bakoitza gainontzeko modalitateen (lurraldeen) maiztasunekiko proportzionalak diren zatietan banatzen da, dagokion kolorea emanaz. Aurrekoetan bezala, interpretazioa errazten duten zuzen bertikalak gehitzen dira.

Kasu honetan ere posible da lurralde historikoak erabili beharrean lau modalitate izan eta horietako bat besteetan desglosa daitekeen edozein aldagai erabiltzea.

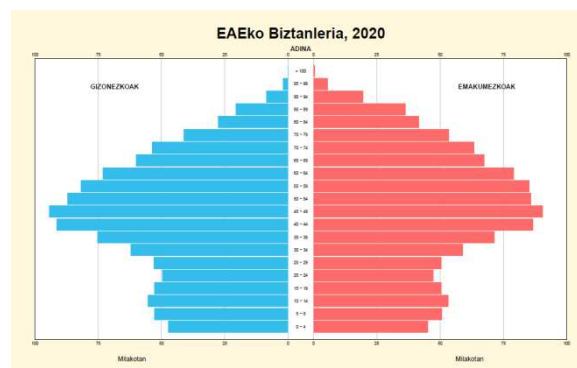
c. Barra-diagramen grafiko konposatua.



Euskal AE-ko barra-diagrama kategorizatuak eta ehunekoen barra-diagramak eta lurralde historiko bakoitzeko barra-diagrama sinpleek osaturiko irudikapen grafikoa da.

ii. Piramidea

a. Piramide sinplea.

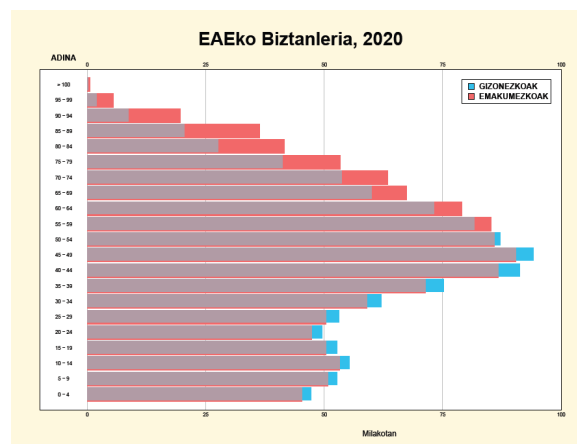


Grafikoki maiztasunen histograma bikoitz batean datza. Histograma bikoitzaren barrak horizontalki irudikatzen dira, hau da, abszisa-lerroaren gainean, eta bi histogramen artean adin-taldeak adierazten dira. Gizonezkoei dagozkien datuak urdinez irudikatzen dira,

histograma bikoitzaren ezker aldean eta emakumezkoei dagozkienak gorritz eta eskuin aldean. Interpretazioa errazteko zuzen bertikalak gehitzen dira. Posible da bai eskalak eta bai koloreak aldatzea

Tradizionalki populazio gaietan erabili den grafikoa bada ere, ekonomiako gaietarako ere erabil daiteke, hala nola, inportazio eta esportazioko saldo komertzialak irudikatzeko.

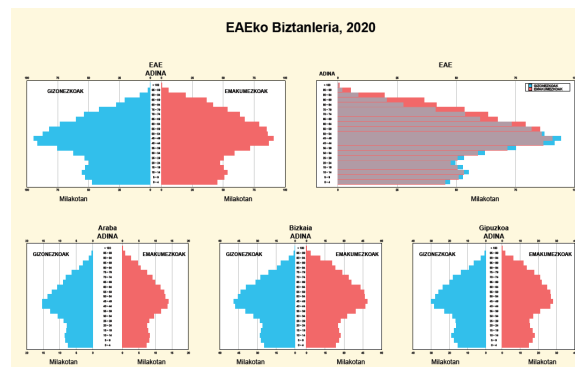
b. Piramide alternatiboa.



Piramide sinplearen antzeko irudikapen grafikoa da, baina kasu honetan bi sexuei dagozkien barrak bata bestearen gainean irudikatzen dira, honela haien arteko konparaketa era errazean egitea ahalbidetuz. Aurreko kasuan bezala, zuzen bertikalak gehitzen dira, eta posible da ekonomiako gaietarako

ere erabiltzea.

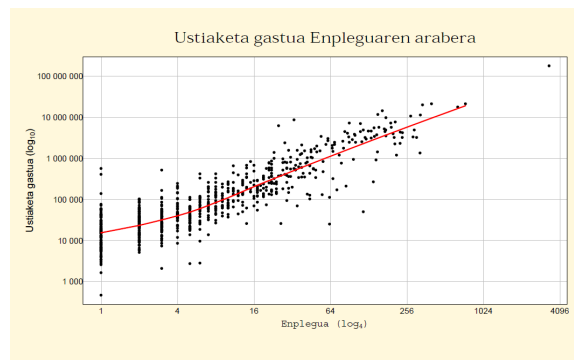
c. Piramideen grafiko konposatua.



Euskal AE-ko piramide sinple eta alternatibo banaz eta lurralde historiko bakoitzeko piramide sinple batez osatutako irudikapen grafikoa da.

• Mikrodatuak:

i. Zatikako erregresio lineala.



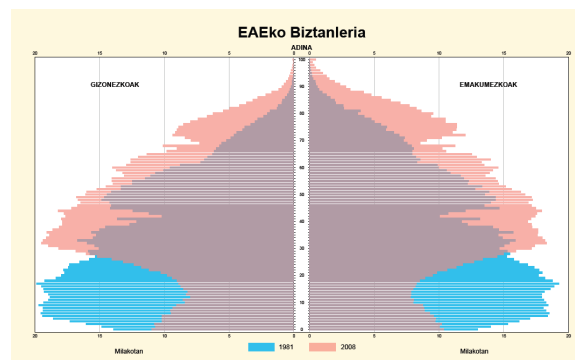
Bi aldagai kuantitatiboren arteko erlazioa zatika lineala den aztertzeke erabiltzen da. Datu bakoitza puntu batez irudikatzen da eta zatikako erregresio zuzen bat doitzen da (gorriz). Datuak oso asimetrikoak badira, aldez aurretik transformatu behar dira (esaterako, logaritmoak erabiliz).

Aldagai anitzeko datuak

• Taulak

i. Piramideak

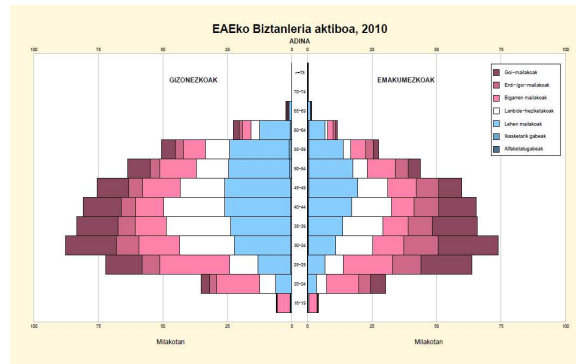
a. Bi-piramidea.



Bi urte ezberdinei dagozkien populazioen datuak irudikatzen ditu, adina eta sexuaren arabera, bi piramideren bitartez. Hain zuzen, urte bati dagokion piramidea beste urteari dagokionaren gainean irudikatzen da, bakoitza kolore batez.

Piramide bat bestearen gainean egoteak eta koloreen gardentasunak adin-tarte bakoitzerako eta sexu bakoitzerako bi urte ezberdinetako populazioak alderatzea ahalbidetzen dute.

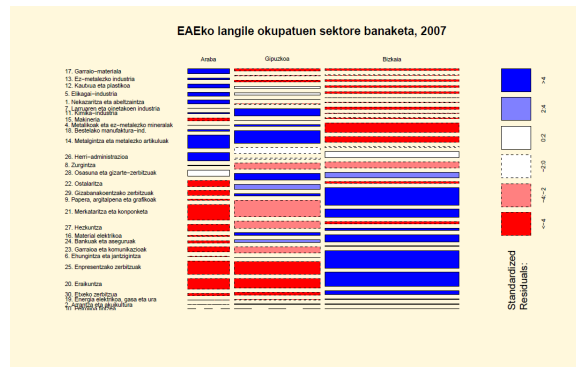
b. Piramide kategorizatuak.



Piramide simple baten antzekoa da, baina aldagai anitzeko datuen kasurako. Kasu honetan, adin-tarte eta sexu bakoitzeko populazioa irudikatzen da, baina aldagai kualitatibo baten modalitateen arabera. Hain zuzen, adin-tarte eta sexu bakoitzeko, populazioaren tamaina

adierazten duen barra irudikatzen da, aldagai kualitatiboaren modalitate bakoitzarekiko proportzionalak diren barratan zatituz eta kolore ezberdinez margotuz. Aurrekoetan bezala interpretazioa errazteko zuzen bertikalak gehitzen dira.

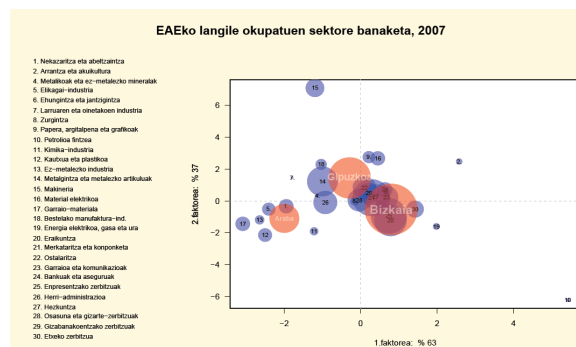
ii. Mosaiko diagrama bi-kategorikoa.



Bi aldagai kualitatiboren arteko erlazioa aztertzeko erabiltzen da. Laukizuzen bakoitzaren zabalerak adierazten duen aldagaiaren modalitatearen maiztasuna irudikatzen du, eta altuerak beste aldagaiaren modalitatearen maiztasuna. Laukizuzenen koloreek, hondarren balio eta zeinua adierazten dute. Hau da,

koloreak gelaxken maiztasunak emaniko ereduaren arabera esperotako maiztasunen azpitik edo gainetik dauden adierazten du.

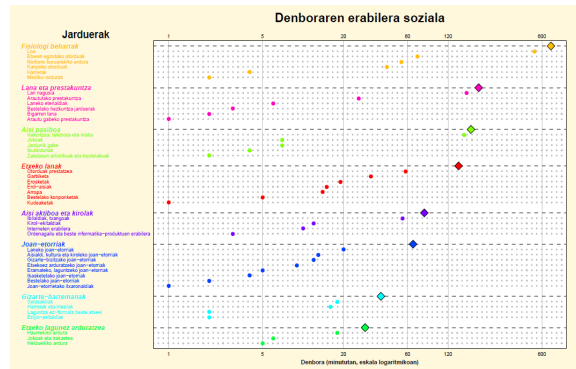
iii. Korrespondentzia analisiaren irudikapen grafikoa.



Korrespondentzia-analisi kontingentzia-taula baten aldagai kategorikoen arteko erlazioak aztertzeko metodo grafiko bat da. Puntuak ardatz bat edo biekiko duten hurbiltasunak (edo distantziak), ardatzak definitzen dituzten modalitateen batezbestekoarekiko

hurbiltasuna edo distantzia (positibo edo negatiboa) adierazten du. Horrela, puntuak antz gehiago (edo gutxiago) duten modalitateekiko gertuago (edo urrunago) egoteko joera izango dute. Bestalde, bolen tamainak haien maiztasuna adierazten du, eta bi kolore erabiltzen dira, bat errenkadetako modalitateak eta bestea zutabeetako modalitateak adierazteko.

iv. Cleveland-en dotplot grafikoa.



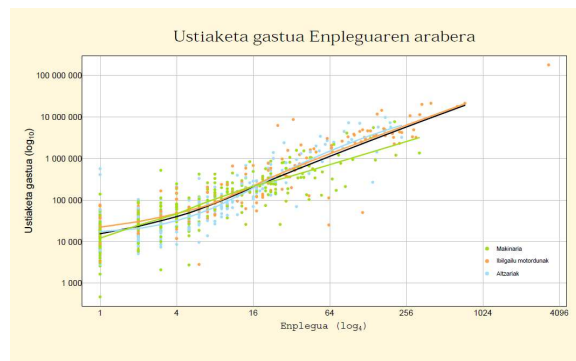
Jatorriz puntu diagrama bat da, baina aldagai askorekin, non aldagaien ordena partzial eta guztizko bat ikusten diren. Irudikatu nahi den aldagai kualitatiboak modalitate eta azpi-modalitateak dituen erabiltzen da. Alde batetik modalitateei dagokien puntu diagrama bat egiten da, eta bestalde, modalitate bakoitzeko azpi-modalitateei

dagokien beste bat.

Baliteke eskala logaritmikoa erabili behar izatea, modalitateen batek besteek baino askoz maiztasun handiagoa duenean, beste modalitateekiko ezberdintasunak argi ikus daitezkeen.

• Mikrodatuak

i. Zatikako erregresio lineala sektoreka.

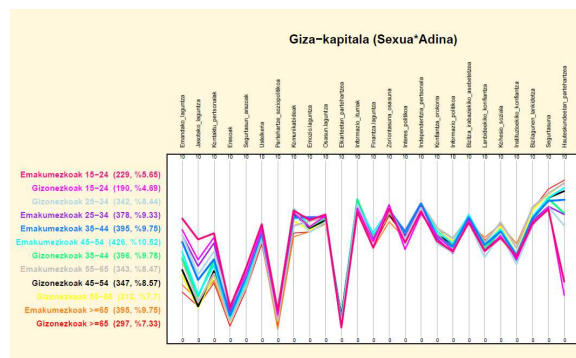


Bi aldagai kuantitatiboren arteko erlazioa, beste aldagai kualitatibo baten modalitateekiko zatika lineala den aztertzeke erabiltzen da. Datu bakoitza puntu batez irudikatzen da eta ondoren zatikako erregresio zuzen orokor bat (beltzez) eta aldagai kualitatiboaren modalitate bakoitzeko zatikako erregresio zuzen

banatzen (bakoitza dagokion koloreaz) doitzen dira.

Zenbaitetan, aldagaiak oso asimetrikoak direnean adibidez, komenigarria da hasi baino lehen transformatzea (esaterako, logaritmoak hartuz).

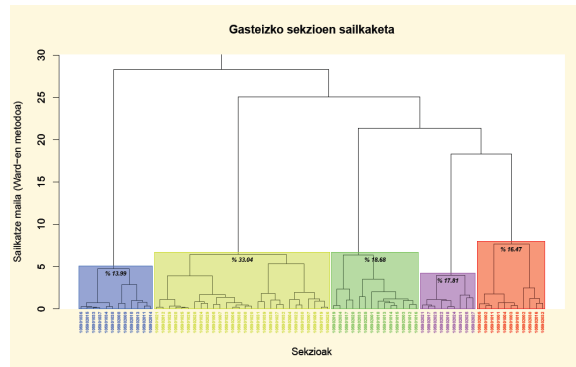
ii. Koordinatu paraleloen grafikoa.



Koordinatu paraleloen grafikoaren helburua aldagai anitzeko datu-multzoak irudikatzea da. Aldagai bakoitza beste guztiekiko paraleloa den eskala bertikal bezala irudikatzen da. Behaketak irudikatzeke, ardatz bakoitzean dagokion puntuak marrazten da, modalitate bakoitzeko kolore bat erabiliz. Ondoren kolore

bereko puntu guztiak zuzenen bidez elkartzen dira. Puntuak ardatzetan irudikatzeke ardatzean irudikatzen den aldagaian behaketak duen balioa hartzen da kontuan.

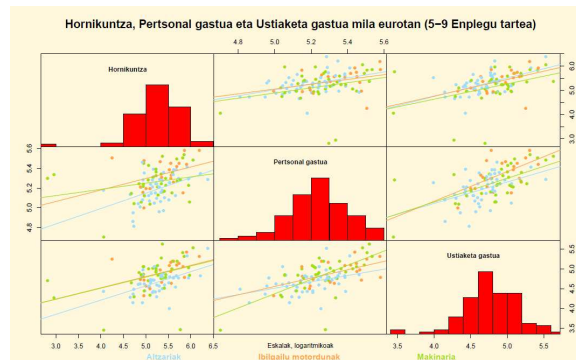
iii. Dendrograma.



Zuhaitz itxurako (dendro=zuhaitz) irudikapen grafiko edo diagrama honek datuak azpikategoriatan antolatzen ditu, zeinak gehiago zatitzen doazen nahi den xehetasun-mailara iritsi arte. Helburua behaketak taldetan ordenatzea da, kluster edo taldeen barruko behaketak ahalik eta antzekoenak eta beste

klusterretakoekiko ahalik eta ezberdinenak izan daitezke. Honek datuen arteko elkartzeko eta egiturak aditzera eman ditzake, hasiera batean ezezagunak zirenak baina erabilgarriak izan daitezkeenak. Ulergarriagoa egiteko, kluster bakoitzari kolore bat eman eta duen indibiduo kopuruaren portzentaje gehitzen zaio ondian.

iv. Scatter plot matrix grafikoa.



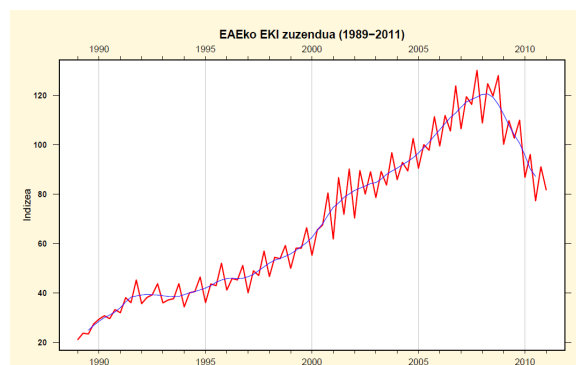
Aldagai biko diagramaz osaturiko $n \times n$ ($3 \leq n \leq 5$ izanik) dimentsioko matrize batean datza. Diagonaletan histogramak irudikatzen dira eta gainontzeko gelaxketan zatikako erregresio linealeko diagramak. Diagonalak behatuta, aldagaien banaketa azter daiteke, eta gainontzeko gelaxketan aldagai ezberdinak gurutzatzean

sortzen diren zatikako erregresio zuzen ezberdinak ikus daitezke.

Denborazko datuak

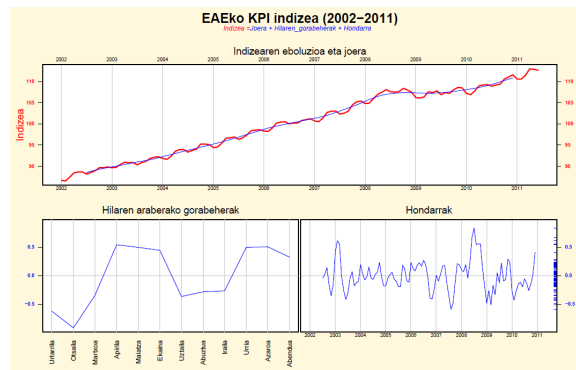
- Taulak

i. Indizeen grafikoak.



Indizea (gorriz) eta egutegi efektua kendua duen indize desestazionalizatua (urdinez) alderatzen dituen grafikoa da. Bi indizeak alderatuz, aztergai den indizeak urtearen garai jakin batzuetan dituen gorakada eta beherakadak ikus daitezke.

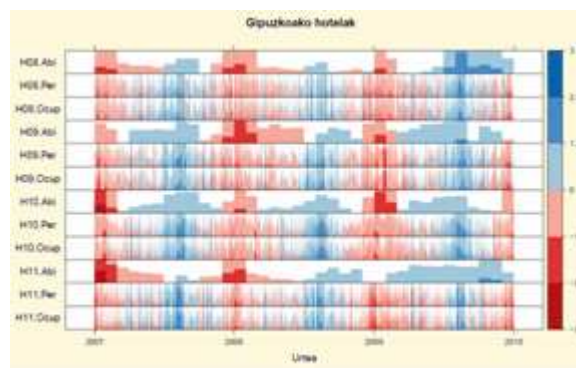
ii. Indizeen grafikoak. Ziklo eta joerak.



Grafiko honetan ere indizea eta indize desestazonalizatua alderatzen dira, baina kasu honetan datuak hilabeteka ematen direnez, hilabeteen araberrako grafikoa eta hondarren grafikoa gehitzen dira.

- **Mikrodatuak**

i. **Horizon grafikoak.**

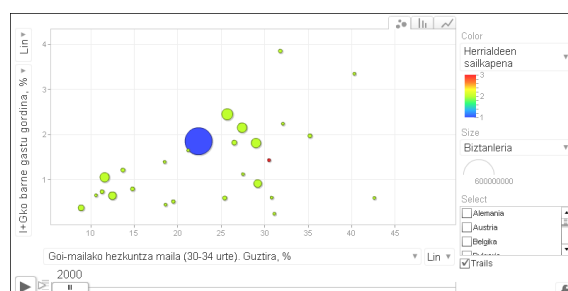


Horizon grafikoeak denborazko serie asko aldi berean eta espazio mugatu batean bistartzeko era eskaintzen dute. Azken finean, lerro-grafiko asko irudikatzean oinarritzen dira, baina espazioa aurrezteko eta serie guztiak aldi berean bistaratu ahal izateko koloreak erabiliz; gorria balio negatiboak adierazteko eta urdina

positiboetarako. Horizon grafikoek zenbait abantaila dituzte, hala nola:

- Jokaera bereziak eta nagusitzen diren joerak ikustea ahalbidetzen dute.
- Serie bakoitza besteeikiko independenteki ikustea ahalbidetzen dute.
- Serieen arteko konparaketa egitea ahalbidetzen dute.
- Aldaketak ikustea ahalbidetzen dute, sakonki aztertzea komeni den jabetzeko adinako zehaztasunarekin.

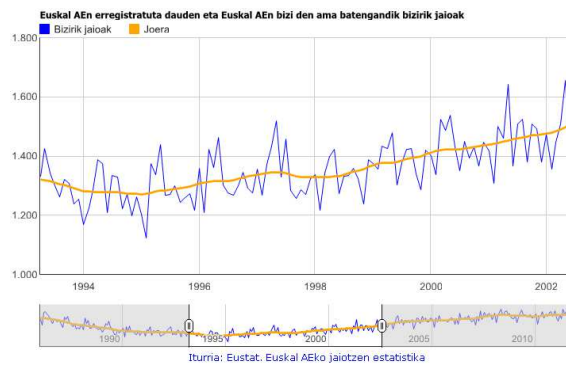
ii. **Gapminder Motion Chart** grafikoak.



aldagaiak, denbora neurtuko duen aldagaia (sarrerako fitxategian datorrena) eta burbuilen tamaina eta kolorea zehaztuko duten aldagaiak. Datuak denboran zehar aztertuz, grafiko honek beste grafiko tradizional batzuetan ikus ezin daitezkeen joera eta ereduak identifikatzeko aukera ematen du, interpretazioa erraztu eta grafikoarekin jolasteko beste zenbait tresna eskaintzeaz gainera.

Motion Chart delakoa grafiko dinamikoa da, datuak bost dimentsio ezberdinetan aztertzeko aukera ematen duena. Erabiltzaileak erabaki dezake zein bost aldagai aztertu nahi dituen, hauek direlarik egin beharreko aukerak: x eta y ardatzetako erako fitxategian datorrena) eta lugaia. Datuak denboran zehar zuetan ikus ezin daitezkeen joera du. interpretazioa erraztu eta

iii. Time Line grafikoak.



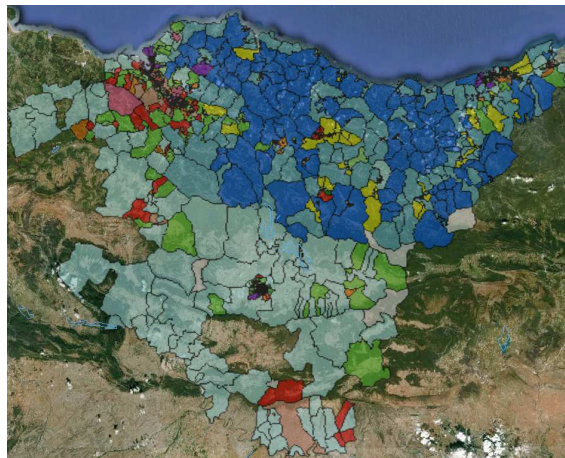
ematen du, y ardatzeko eskala automatikoki aukeratutako tartera egokituz.

Denborazko serieak bistaratzeko grafiko interaktibo batean datza. Kurtsorea bistaritzen den serie bakoitzaren gainean jarri, serieak puntu horretan duen balioa eta dagokion data azaltzen dituen koadro bat ateratzen da. Lerro-diagramaren azpian dagoen egiturak nahi den denbora-tartean zentratzeko aukera

Datu espazialak

- Taulak
- Mikrodatuak

i. Google Earth bidezko bistaratzea.



hurbildu ahala kaparen azpian dauden kaleak, eraikuntzak, parkeak... ikustea ahalbidetzen du. Bestalde, maparen gainean sakatuz, zonaren inguruko zenbait datu (herriaren izena; barrutia; zentsu-sekzioa; tipoa; 15 urtez beherakoen, 65 urtez gorakoen, euskaldunen eta goi-mailako titulazioa dutenen portzentajeak) eta EUSTATEko web-orrira apuntatzen duen linka dituen informazio-panela azaltzen da.

Google Earth-ek datu espazialen analisi bat egin eta kml fitxategi bat sortu ondoren, eraikitako kapa bistaratzeko tresna erabilgarri eta erraz bat eskaintzen du. Kasu honetan, Euskal AEko (aurrerantzean EAE) errolda-sekzioen tipologia analisi bat eginez 12 tipo edo sail sortu dira eta ondoren kml fitxategia. 12 tipo horien arabera margotu dira EAEko errolda-sekzio ezberdinak. Koloreen gardentasunak helbide batera

Aplikazioak

Behin grafikoen katalogoa egin ondoren, proposenak eta erabilgarrienak aukeratu dira; batzuk automatizatzeko eta web orrian erakusteko eta besteak Eustat-en barnean erabiltzeko, lanerako tresna bezala.

Grafiko hauek R programazio lengoaian inplementatu dira.

Produkzioarako grafikoak

Zenbait grafikoren R script-ak automatizatu dira, sarrerak estandarizatuz, Eustat-eko Estatistika Ekonomikoen taldeak lanerako tresna bezala erabil ditzan. Grafiko hauek erabiltzeko dagozkien funtzioak kargatu eta exekutatu besterik ez dago. Hauek dira automatizaturiko grafikoak:

1. Dotplot edo puntu diagrama.
2. Barra-diagramen grafiko konposatua.
3. Korrespondentzia-analisiaren irudikapen grafikoa.
4. Zatikako erregresio lineala, sektoreka.
5. Scatterplot matrix grafikoa.

1. Dotplot edo puntu diagrama

R-n DOTPLOT EDO PUNTU-DIAGRAMA SORTZEKO FUNTZIOA:
'PuntuDiagrama()'

DESKRIBAPENA

'PuntuDiagrama()' funtzioa R-ren berezko 'graphics' liburutegiko 'dotchart()' funtzioaren egokitza bat da.

Aldagaiaren modalitateen aldakuntzak bistaratzeko egokia da. Modalitateak aldakuntzaren arabera edo alfabetikoki ordenatu eta puntu bidez adierazten ditu, 0-an zuzen gorri bertikal bat ezarri, aldakuntzaren zeinua erraz ikusteko.

ERABILERA

PuntuDiagrama(datuak, main=FALSE, ordenalfabet=FALSE, cex=0.8)

...

ARGUMENTUAK

datuak:	Zenbakizko balioak eta beren etiketak.
main:	Defektuz FALSE izango da; hau da, grafikoaren izenburua ez da idatziko. Izenburua idatzi nahi bada, main="nahi den izenburua, kakotxoaren artean" zehaztu.
ordenalfabet:	Defektuz FALSE izango da; hau da, puntuak definitzen dituzten etiketak ez dira alfabetikoki ordenatuko. TRUE denean, etiketak alfabetikoki ordenatuta azalduko dira.
cex:	Puntu eta letren tamainak zehazteko parametroa. Defektuz, 0.8ko balioa du.

XEHETASUNAK

Puntu-diagramak barra-diagramen ordezkotako egokiak dira, batez ere, aldakuntzak (positibo/negatiboak) aztertu nahi direnean.

OHARRAK

Funtzio hau erabili aurretik, bai funtzioa bera eta baita irudikatu nahi den rda fitxategia ere karpeta berdinean gordeta izan behar dira eta ondoren R-n karpeta horren helbidea zehaztu behar da. Azken hau bi eratan egin daiteke; R-n hurrengo ikonoak "Archivo → Cambiar dir... → (dagokion karpeta)" sakatuz edo setwd("gordeta dagoen tokiaren helbide guztia, komatxo artean") komandoa jarritz.

Behin hau eginda, irudikatu nahi den fitxategia kargatu behar da, load("fitxategiaren izena luzapenarekin, komatxo artean") eta baita erabili nahi den funtzioa ere, source("funtzioaren izena, komatxo artean") aginduarekin.

SARRERAKO FITXATEGIAREN EGITURA:

Sarrerako fitxategiak bi zutabe izango ditu:

- Lehen zutabearen etiketa edo kategoriak joango dira (izenak, zenbakiak, edo zenbaki eta izenak izan daitezke).
- Bigarren zutabearen balioak zehaztuko dira (hazkundeak/beherapenak).

ADIBIDEA

```
setwd('C:/Produktzioarako grafikoak/Puntu_diagrama')
source('PuntuDiagrama')
load('RGrafico1_1.rda')
PuntuDiagrama(datuak=RGrafico1_1, main="HAZKUNDEA")
```

2. Barra-diagramen grafiko konposatua

R-n BARRA-DIAGRAMEN GRAFIKO KONPOSATUA SORTZEKO FUNTZIOA: 'MarraDiagramak()'

DESKRIBAPENA

Grafiko hau barra-diagrama ezberdinez osaturik dago: barra-diagrama kategorizatuaz, ehunekoen barra-diagramaz eta barra diagrama sinpleez, hain zuzen. 'MarraDiagramak()' funtzioak beste hiru funtzio erabiltzen ditu barnean, 'marradiagrama()', 'marradiagrama100()' eta 'MarraDiagramak.gr()' funtzioak, alegia. (Oharretan azalpena).

ERABILERA

MarraDiagramak(datuak, titulua="")

...

ARGUMENTUAK

datuak:	Zenbakizko balioak eta beren etiketak.
titulua:	Defektuz grafikoaren izenburua ez da idatziko. Izenburua idatzi nahi bada, titulua="nahi den izenburua, kakotxoaren artean" zehaztu.

XEHETASUNAK

Barra-diagrama ezberdinen bistaratze honek EAEko eta hiru lurralde historikoetako bakoitzeko datuak alderatzeko aukera ematen digu. Alde batetik, datu gordinen banaketa aldera dezakegu, eta bestalde, bai maiztasun absolutu zein erlatiboena ere.

Aldagai kuantitatibo baten modalitateen maiztasunak beste aldagai kualitatibo baten maiztasunen arabera irudikatzeko erabilgarria da, azken aldagai honek lau modalitate izan eta horietako bat gainontzekoetan desglosa daitekeenean.

OHARRAK

Funtzio hau erabili aurretik, bai funtzioa bera eta baita irudikatu nahi den rda fitxategia ere karpeta berdinean gordeta izan behar dira eta ondoren R-n karpeta horren helbidea zehaztu behar da. Azken hau bi eratan egin daiteke; R-n hurrengo ikonoak "Archivo → Cambiar dir... → (dagokion karpeta)" sakatuz edo setwd("gordeta dagoen tokiaren helbide guztia, komatxo artean") komandoa jarritz.

Behin hau eginda, irudikatu nahi den fitxategia kargatu behar da, load("fitxategiaren izena luzapenarekin, komatxo artean") eta baita erabili nahi den funtzioa ere, source("funtzioaren izena, komatxo artean") aginduarekin.

Funtzio honek hiru funtzio hauek erabiltzen ditu:

- 'marradiagrama()' funtzioa barra-diagrama sinpleak eta maiztasun absolutuen barra-diagrama sortzen dituen funtzioa da, R-ren berezkoa den 'barplot()' funtzioaren erabilpen berezi bat izanik.

- 'marradiagrama100()' funtzioa ehunekoen barra-diagrama edo maiztasun erlatiboen barra-diagrama sortzen duena da, eta hau ere R-ren berezkoa den 'barplot()' funtzioaren erabilpen berezi bat da.
- 'MarraDiagramak.gr()' funtzioa 'marradiagrama()' eta 'marradiagrama()' funtzioen konposaketa berezi bat da.

'MarraDiagramak()' funtzioak 'marradiagrama()' eta 'marradiagrama100()' funtzioak erabiltzen ditu barra-diagrama sinpleak eta maiztasun absolutu eta erlatiboen barra-diagramak sortzeko, eta sarrerako datu-matrizea moldatzen du 'MarraDiagramak.gr()' funtzioa erabiltzeko.

SARRERAKO FITXATEGIAREN EGITURA:

Sarrerako fitxategiak hiru zutabe izango ditu:

- Lehen zutabea etiketa edo kategoriak joango dira (izenak, zenbakiak, edo zenbaki eta izenak izan daitezke).
- Bigarren zutabea lurralde historikoak adieraziko dira (lurraldeen izenak edo kodeak izan daitezke).
- Hirugarren zutabea balioak zehaztuko dira (hazkundeak/beherapenak).

ADIBIDEA

```
setwd('C:/Produktzioarako grafikoak/Barra_diagramak')
source('MarraDiagramak')
load('RGrafico2_1.rda')
datuak <- RGrafico2_1[-c(1,1+nrow(datuak)/3,1+2*nrow(datuak)/3), ]
MarraDiagramak(datuak, titulu="Langile-banaketa / Distribución personal")
```

3. Korrespondentzia-analisiaren irudikapen grafikoa

R-n KORRESPONDENTZIA-ANALISIA IRUDIKATZEKO FUNTZIOA:
'korrespondentziak()'

DESKRIBAPENA

Bi aldagai kualitatiboren arteko korrespondentzi bakunaren analisia irudikatzeko funtzioa.

'ca' liburutegi edo paketeko 'ca()' funtzioa eraldatu eta erabiltzen da. (Oharretan azalpena).

ERABILERA

Korrespondentziak(datuak, main="izenburua", zeinux=0, zeinuy=0)

...

ARGUMENTUAK

datuak:	Irudikatu nahi den fitxategia.
main:	Izenburua. Defektuz NULL izango da. Izenburua jarri nahi izanez gero, main="izenburuaren izena, komatxo artean" zehaztu.
zeinux:	Abzisen noranzkoa kontrolatzeko parametroa. Defektuz 0, hau da, x ardatzean biraketarik ez. 180º biratzeko, 1 balioa eman.
zeinuy:	Ordenatuen noranzkoa kontrolatzeko parametroa. Defektuz 0, hau da, y ardatzean biraketarik ez. 180º biratzeko, 1 balioa eman.

XEHETASUNAK

Korrespondentzia analisia kontingentzia-taula bateko aldagaien arteko erlazioak aztertzeko metodo grafikoa da. Korrespondentzia analisia ezagunagoa den adierazpen grafiko baten orokorpen bat da, sakabanatze diagramarena, hain zuzen.

Puntuak ardatz batekiko duten gertutasunak (edo distantziak), ardatza definitzen duten modalitateen batzbestekoarekiko gertutasuna edo distantzia (positiboa edo negatiboa) adierazten du. Honela, puntuak antzekotasun gehiago (edo gutxiago) duten modalitateekiko gertuago (edo urrutiago) egongo dira.

Bestalde, bolen tamainak beren maiztasuna adierazten du.

OHARRAK

Funtzio hau erabili aurretik, bai funtzioa bera eta baita irudikatu nahi den rda fitxategia ere karpeta berdinean gordeta izan behar dira eta ondoren R-n karpeta horren helbidea zehaztu behar da. Azken hau bi eratan egin daiteke; R-n hurrengo ikonoak "Archivo → Cambiar dir... → (dagokion karpeta)" sakatuz edo setwd("gordeta dagoen tokiaren helbide guztia, komatxo artean") komandoa jarritz.

Behin hau eginda, irudikatu nahi den fitxategia kargatu behar da, load("fitxategiaren izena luzapenarekin, komatxo artean") eta baita erabili nahi den funtzioa ere, source("funtzioaren izena, komatxo artean") aginduekin.

Funtzio honek hurrengo funtzioak erabiltzen ditu:

- 'ca()' funtzioak korrespondentzi analisi sinplea egiten duen 'ca' paketeko 'ca()' funtzioa moldatzen du. Abzisa zein ordenatu ardatzen noranzkoa kontrolatzeko aukera ematen du, hau da, 180º-ko biraketa egin nahi den edo ez zehazteko.
- 'korresp.gr()' funtzioak 'ca()' funtzioaren emaitzak erabiltzen ditu, korrespondentziak bi dimentsiotan adieraziz.

SARRERAKO FITXATEGIAREN EGITURA:

Sarrerako fitxategiak hiru zutabe izango ditu:

- Lehen zutabean etiketa edo kategoriak joango dira (izenak, zenbakiak, edo zenbaki eta izenak izan daitezke).
- Bigarren zutabean lurralde historikoak adieraziko dira (lurraldeen izenak edo kodeak izan daitezke).

- Hirugarren zutabean balioak zehaztuko dira.

ADIBIDEA

```
setwd('C:/Produktzioako grafikoak/Korrespondentzia')
source('korrespondentziak')
load('RGrafico3_1.rda')
datuak<-RGrafico3_1[-c(1,nrow(datuak)/3+1,2*nrow(datuak)/3+1), ]
korrespondentziak(datuak, main="Korrespondentziak")
```

4. Zatikako erregresio lineala, sektoreka

R-n ZATIKAKO ERREGRESIO LINEALA, SEKTOREKA EGIN ETA IRUDIKATZEKO FUNTZIOA: 'Erregresioa()'

DESKRIBAPENA

Bi aldagai kualitatiboren arteko erlazioa 3. aldagai kualitatibo baten modalitateekiko zatika lineala al den aztertzeako funtzioa.

'Erregresioa()' funtzioak beste hiru funtzio erabiltzen ditu barnean, 'bitxiizein()', 'transf()' eta 'LoessErregresio()' funtzioak, alegia (Oharretan, azalpena).

ERABILERA

Erregresioa(datuak, titulua=NULL, azpititulua=NULL, logbektorea=c(0,0))

...

ARGUMENTUAK

datuak:	Irudikatu nahi den fitxategia.
titulua:	Izenburua. Defektuz NULL izango da. Izenburua jarri nahi izanez gero, titulua="izenburuaren izena, komatxo artean" zehaztu.
azpititulua:	Azpi-izenburua. Hau ere, defektuz NULL izango da. Azpiizenburua jarri nahi izanez gero, azpitulua="azpiizenburua, komatxo artean" zehaztu.
logbektorea:	Eskala logaritmikoa erabiliko denean logaritmo kalkuluak egiteko bektorea. Besterik ezean, ez da logaritmo kalkulurik egingo (defektuz, c(0,0)). 1-ek eskala logaritmikoa adierazten du, eta 0-k dagoen eskala.

XEHETASUNAK

Erregresio lokala (ingelesez LOESS edo LOWESS) erregresio lineal edo ez-lineala bezalako metodo klasikoetan oinarritzen den eraikuntza metodo berria da. Erregresio metodo berri hauek prozedura klasikoak egokiak edo nahikoak ez diren kasuetarako diseinaturik daude.

Erregresio lokalaren helburua puntuz puntu datuen aldakuntzen zati determinista deskribatzen duen funtzioa sortzea da. Horretarako datuen azpimultzo lokalen gainean

eredu errazak doitzen ditu, minimo karratuen erregresio lineala eta erregresio ez-lineala konbinatuz.

OHARRAK

Funtzio hau erabili aurretik, bai funtzioa bera eta baita irudikatu nahi den rda fitxategia ere karpeta berdinean gordeta izan behar dira eta ondoren R-n karpeta horren helbidea zehaztu behar da. Azken hau bi eratan egin daiteke; R-n hurrengo ikonoak “Archivo → Cambiar dir... → (dagokion karpeta)” sakatuz edo setwd(“gordeta dagoen tokiaren helbide guztia, komatxo artean”) komandoa jarritz.

Behin hau eginda, irudikatu nahi den fitxategia kargatu behar da, load(“fitxategiaren izena luzapenarekin, komatxo artean”) eta baita erabili nahi den funtzioa ere, source(“funtzioaren izena, komatxo artean”) aginduarekin.

Funtzio honek hiru funtzio hauek erabiltzen ditu:

- ‘bitxiazein()’ funtzioa balio bitxiak salatzen duen funtzioa da, balioak arrarotzat jotzeko irizpide bat ezarriz. Hain zuzen, arrarotzat hartuko dira $(q1 - \text{zenbath} * (q3 - q1))$ baino handiagoen minimoa baino txikiagoak eta $(q3 + \text{zenbath} * (q3 - q1))$ baino txikiagoen maximoa baino handiagoak direnak. zenbath defektuz 2.75 balio duen parametroa da, eta q1 eta q3, lehen eta hirugarren kuartilak, hurrenez hurren.
- ‘transf()’ funtzioa datuen transformaziorik hoberena bilatzeko funtzioa da. Horretarako Shapiro-Wilks-en normaltasun testaren estatistikoa kalkulatzen du eta baliorik handiena duena izango da transformazio onena emango duena.
- ‘LoessErregresio()’ funtzioak lehenik translazioa egin eta eskala logaritmoa erabili behar ote den baloratzen du, gero balio bereko puntuak badaude mugitu egiten ditu, gainjar ez daitezkeen. Ondoren puntuak irudikatzen ditu, bakoitza dagokion kolorean eta erregresio lokalak doitzen ditu, guztirakoa eta partzialak (balio bitxiak kontuan hartu gabe), azkenik eredu nagusitik urruti dauden puntuak identifikatu eta etiketatuz.. Funtzio honek R-ren berezko ‘graphics’ liburutegiko ‘loess()’ funtzioa erabiltzen du, erregresio lokala kalkulatzen duena.

SARRERAKO FITXATEGIAREN EGITURA:

Sarrerako fitxategiak lau zutabe izango ditu:

- Lehen zutabearen puntuen identifikatzaileak joango dira (balio arraro, ezohiko balioak... identifikatzeko erabiliko ditugunak).
- Bigarren zutabearen etiketa edo kategoriak joango dira (izenak, zenbakiak, edo zenbaki eta izenak izan daitezke).
- Hirugarren zutabearen x ardatzean aztertu nahi den aldagaiaren balioak joango dira.
- Laugarren zutabearen y ardatzean aztertu nahi den aldagaiaren balioak joango dira.

ADIBIDEA

```
Setwd("C:/Produktzioarako grafikoa/ Zatikako erregresio lineala, sektoreka")
Source("Erregresioa")
load("RGrafico4_1.rda")
datuak <- RGrafico4_2
Erregresioa(datuak, titulua="Balioak Enpleguaren arabera", Azpitolua="Eskala
logaritmikoa, logbektorea=c(1,1))
```

5. Scatterplot matrix grafikoa

R-n DIAGONALEAN HISTOGRAMAK DITUEN SCATTERPLOT MATRIX, GRAFIKOA IRUDIKATZEKO FUNTZIOA: 'binakakoa()'

DESKRIBAPENA

n ($3 \leq n \leq 5$) aldagai kuantitatibo irudikatzeko aukera ematen digun irudikapen grafikoa da.

`binakakoa()` funtzioak beste lau funtzio erabiltzen ditu barnean, `'transf()'`, `'hist.puntuak()'`, `'panel.puntuak()'` eta `'binakako.pairs()'` funtzioak, alegia. (Oharretan, azalpena).

ERABILERA

```
binakakoa(datuak, titulua=NULL, azpitolua=NULL, logbektorea=rep(0,ncol(datuak)-2)
...
```

ARGUMENTUAK

<code>datuak:</code>	Irudikatu nahi den fitxategia.
<code>titulua:</code>	Izenburua. Defektuz NULL, izenbururik ez. Izenburua nahi izatekotan, <code>titulua="izenburuaren izena, komatxoaren artean"</code> zehaztu.
<code>azpitolua:</code>	Azpi-izenburua. Hau ere, defektuz NULL izango da. Azpi-izenburua nahi izatekotan, <code>azpitolua="azpi-izenburuaren izena, komatxoaren artean"</code> zehaztu.
<code>logbektorea:</code>	Logaritmo kalkulak egiteko bektorea. Besterik ezean, ez da kalkulurik egingo (defektuz, <code>rep(0, ncol(datuak)-2)</code>). 1-ek eskala logaritmikoa adierazten du, eta 0-k aldagaiaren berezko eskala.

XEHETASUNAK

Scatterplot deritzon grafikoa bi aldagaien arteko erlazioa nolakoa den azaltzen du. Zehatzago esanda, banaketaren itxura, joerak, aldagaien arteko erlazioa indartsua ala ahula den, klusterrak edota ereduak erakuts ditzake eta balio arraro edo ezohikoak agerian utzi ditzake.

Histogramak, ordea, laginketako banaketaren dentsitate erlatiboa adierazten du. Hasierako datu-multzoaren partiketa bat eginez sorturiko zabalera bereko tarte bakoitzean dagoen behaketa kopurua barren bidez adierazten da, beren altuera maiztasunarekiko proportzionala izanik.

Plot matrizea diagramadun gelaxkaz osaturiko matrizea da. Matrizearen diagrama-gelaxka bakoitzak gelaxkako ilara eta zutabeetako aldagaien arteko erlazioa adierazten duen diagrama bat du. Beraz, plot matrize batek datu-multzoko aldagai bikote posible guztien arteko erlazioak adierazten ditu.

Plot matrize bateko gelaxka guztietako grafikoek ez duten zertan berdinak izan. Kasu honetan scatterplot matrizeak egin dira, diagonaletan izan ezik, non histogramak irudikatu diren, aldagaien banaketei buruzko ideia bat egiteko.

OHARRAK

Funtzio hau erabili aurretik, bai funtzioa bera eta baita irudikatu nahi den rda fitxategia ere karpeta berdinean gordeta izan behar dira eta ondoren R-n karpeta horren helbidea zehaztu behar da. Azken hau bi eratan egin daiteke; R-n hurrengo ikonoak “Archivo → Cambiar dir... → (dagokion karpeta)” sakatuz edo setwd(“gordeta dagoen tokiaren helbide guztia, komatxo artean”) komandoa jarritz.

Behin hau eginda, irudikatu nahi den fitxategia kargatu behar da, load(“fitxategiaren izena luzapenarekin, komatxo artean”) eta baita erabili nahi den funtzioa ere, source(“funtzioaren izena, komatxo artean”) aginduarekin.

Funtzio honek hiru funtzio hauek erabiltzen ditu:

- ‘transf()’ funtzioa elementu bitxiak salatzeko funtzioa da.
- ‘panel.hist()’ funtzioa grafikoaren diagonaleko histogramak irudikatzeko funtzioa da.
- ‘panel.puntuak()’ funtzioa gainontzeko gelaxketako (diagonalak izan ezik) sakabanatze-diagramak eta erregresioak irudikatzen dituen funtzioa da.
- ‘binakako.pairs()’ funtzioa R-ko berezko ‘graphics’ liburutegiko ‘pairs()’ funtzioaren egokitzen bat da eta matrizearen diagonaletan histogramak eta gainontzeko gelaxkatan sakabanatze-diagramak eta erregresioak irudikatzen ditu.

SARRERAKO FITXATEGIAREN EGITURA:

Sarrerako fitxategiak lau zutabetik (n=3 kasurako) sei zutabera (n=5 kasurako) izango ditu (aztertu nahi den aldagai kopuruaren arabera):

- Lehen zutabearen etiketa edo kategoriak joango dira (izenak, zenbakiak, edo zenbaki eta izenak izan daitezke).
- Bigarren zutabetik aurrera aztertu nahi diren aldagaiak zehaztuko dira (3 eta 5 bitartean).

ADIBIDEA

```
setwd("C:/Produktzioarako grafikoak/ Scatterplot matrix grafikoa")
source("binakakoa")
load("RGrafico5_1.rda")
datuak <- RGrafico5_1
binakakoa(datuak, titulu="PB_sf, VAB_cf, C_P, eta EN_E (adin-geruzka)",
azpititulu="PB_sf, VAB_cf, C_P, eta EN_E logaritmo-eskalan", logbektorea=c(1,1,1,1))
```

Zabalkunderako grafikoak

Eustat-eko web-orrian publikatzeko 2 grafiko dinamiko-interaktibo eta Google Earth bidezko bistaratze bat eratu dira. Hauek erabiltzeko dagozkien funtzioak kargatu eta exekutatu besterik ez dago. Hauek dira automatizaturiko grafikoak:

1. Motion Chart grafikoa.
2. Time Line motako grafikoa.
3. Google Earth bidezko bistaratzea.

MOTION CHART ZEIN TIME LINE MOTAKO GRAFIKOAK BISTARATZEKO TRESNA. GOOGLE VISUALIZATION API

API Aplikazioak Programatzeko Interfazea (ingelesez Application Programming Interface) funtzio eta prozedura (edo metodo, objektuei zuzenduriko programazioaren kasuan) multzoa da, beste software batek erabiltzeko liburutegi jakin bat eskaintzen duena.

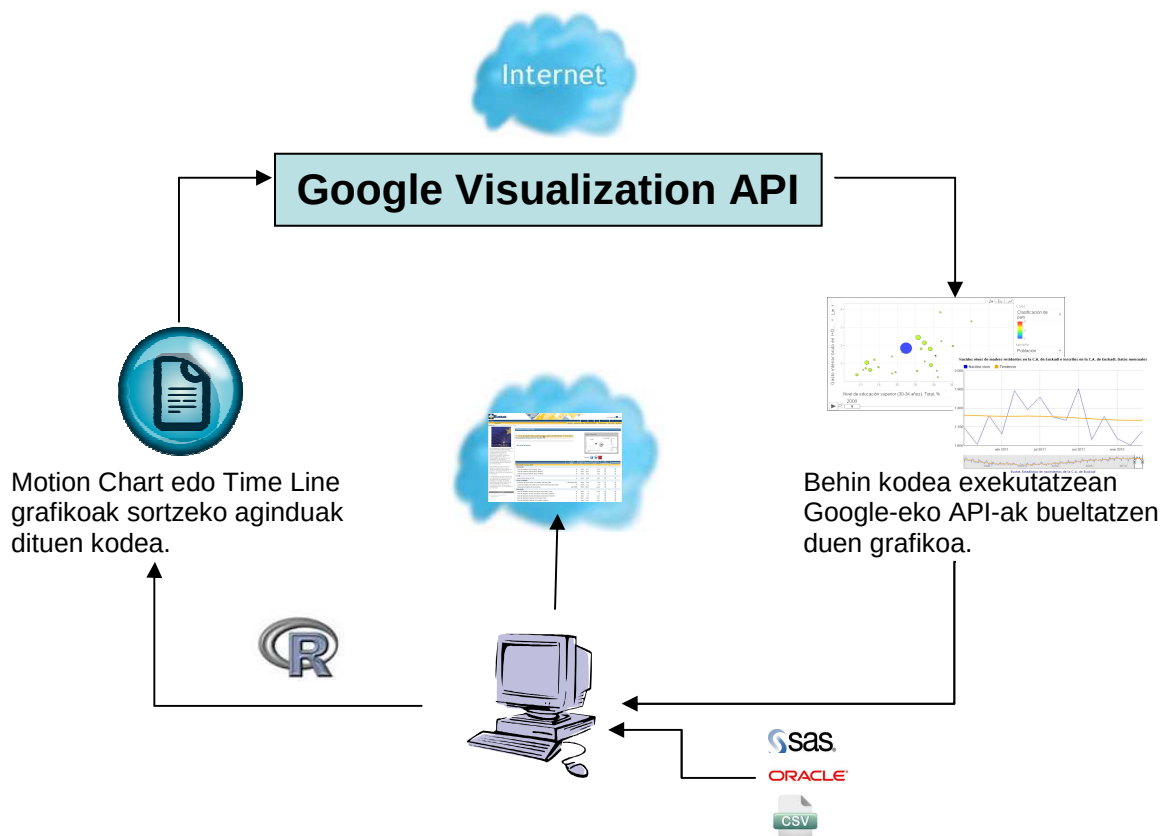
API-aren helburu nagusietako bat erabilera orokorreko funtzio multzo bat eskaintzea da, adibidez, pantaila leiho edo ikonoak marrazteko. Honela, programatzaileak APIaren abantailez baliatzen dira, dena hasieratik programatzea ekidinez.

Google Visualization API delakoa erabiltzaileei bistaratze teknologia aurreratuaren erabilpena errazteko diseinatua dagoen JavaScript API-a da. Gainera erraza eta azkarra da bistaratze berriekin integratzea. Hain zuzen, datuak sarean ezartzeko zein bistaratzeetarako datuak eskuratzeko interfaze arrunt bat da.

Bi elementu klabetan oinarritzen da: erraztasunean eta nonahikotasunean. Google Visualization API-ko galeriako edozein grafiko aukeratu ondoren, berau sortzeko kodea programatu (tresnak eskatzen dituen baldintzak eta egitura errespetatuz) eta API-ari deituz, honek online exekutatu eta aukeraturiko grafikoa bueltatzen du.

Grafikoa bistaratu ahal izateko ezinbestekoa da ordenagailuan Flash Player-a instalatuta izatea.

Hau da aukeraturiko grafikoa (eta APIko beste edozein) sortzeko prozeduraren eskema:



ZENTSU-SEKZIOEN TIPOLOGIEN ANALISIA BISTARATZEKO TRESNA. GOOGLE EARTH

Google Earth Informazio Geografiko Sistema (ingelesez GIS, Geographic Information System) baten antzekoa den programa informatikoa da. Planetako irudiak bistaratzea ahalbidetzen du, sateliteek azken hiru urteetan ateratako argazkiak, mapak eta datu-base oso on bat konbinatuz.

Google Earth-eko bilatzailea gauza ezberdinetarako erabil daiteke, esaterako:

- Herrialde bat, hiri bat edota hiri bateko helbide jakin bat bilatzeko.
- Lur Planetako edozein lurralde aurkitzeko, bere koordinatuak abiatuz.
- Herrialde bateko bi hiriren arteko edo hiri bateko bi helbideren arteko ibilbideak markatzeko.
- Bi hiri edo herrialde ezberdinen arteko ibilbide egokiena markatzeko.

Bestalde, informazioa duten “Layers” edo kapa batzuk dauzka, aldi berean edo independenteki aktiba eta desaktiba daitezkeenak, espazio geografikoko datu mota ezberdinak bistaratzeko.

Kapa hauei esker erabiltzaileak, besteak beste:

- Hiri bateko izen eta kaleak ikus ditzake.

- Ikastetxe, ospitale, hotel, jatetxe, parke, leku interesgarri eta abarren kokapena zehaztu dezake.
- Muga, errepide eta trenbideak bistara ditzake.
- Sumendiak, lurrikaren epizentroak, lakuak, urmael eta ibaiak... bistara ditzake.
- Leku historiko eta kultural garrantzitsuenen kokapena identifika dezake.
- Muino edo mendiak bezalako lur-eremu garaiak hiru dimentsiotan bistara ditzake.

1. Motion Chart grafikoa

1. MOTION CHART MOTAKO GRAFIKOAK SORTZEKO EGITURA EGOKIA DUEN FITXATEGIA PRESTATZEN DUEN FUNTZIOA R-n, EUSTATEKO INDIKADOREEN DATU-FORMATUTIK ABIATUTA: 'Elkartu_gapminder()'

DESKRIBAPENA

Funtzio honek adierazleen csv fitxategiak hartuta (adierazle bakoitzeko taula bat) R-ko 'googleVis' paketeko 'gvisMotionChart()' funtzioaren moldaketa bat erabiltzeko beharrezkoa den fitxategia bueltatzen du.

ERABILERA

Elkartu_gapminder(lista1, fitxategia_izena, euskara=TRUE, populazioa)

ARGUMENTUAK

lista1:	txt motako fitxategia da, non erabili nahi diren adierazleen csv fitxategien izenak (luzapenik gabe) zehaztuko diren, hutsunez bereiziak. Hurrengo ordena izan behar du: lehenik x ardatzeko adierazlea, ondoren y ardatzekoa, jarraian populazioa zehazten duena (egotekotan) eta ondoren gainontzeko adierazleak. Azkenik erabilitako adierazle / fitxategi kopurua zehaztu (zenbakia).
fitxategia_izena:	Irteerako txt motako fitxategiaren izena.
gazelania, euskara, ingelesa:	Sarrerako zein irteerako fitxategien hizkuntza. Beti aukeratu behar da bat eta bakarra. Nahi den hizkuntzari TRUE esleituko zaio.
populazioa:	Defektuz FALSE. Populazioa zehazten duen adierazlea egotekotan TRUE esleitu.

XEHETASUNAK

R-ko 'googleVis' packageak Google APIko Motion Chart motako grafikoak sortzeko erreminta erraz eta erabilgarria eskaintzen du. Mota honetako grafikoak sortzeko 'Motion_chart()' funtzioa sortu da, 'googleVis' paketeko 'gvisMotionChart()' eta 'gvis()' funtzioetatik abiatuz. Hurrengo atalean aztertuko den 'Motion_chart()' funtzioa exekutatu ahal izateko sarrerako fitxategiak egitura berezi bat izan behar du, eta hori da, hain zuzen, 'Elkartu_gapminder()' funtzioa exekutatzean lortuko den irteerako fitxategiaren egitura.

'Motion_chart()' funtzioaren sarrerako fitxategiak (eta 'Elkartu_gapminder()' funtzioaren irteerako fitxategiak) hurrengo egitura izan behar du:

- 1.zutabea: Karaktere motako zutabe honek entitateen izenak (kasu honetan herrialdeak) izan behar ditu.
- 2.zutabea: Zutabe honetan aztertu beharreko urteak azalduko dira, ordenan.
- 3.zutabea: X ardatzean aztertu nahi den aldagaia.
- 4.zutabea: Y ardatzean aztertu nahi den aldagaia.
- 5.zutabea: Burbuilen kolorea zehaztuko duen aldagaia.
- 6.zutabea: Burbuilen tamaina zehaztuko duen aldagaia.
- 7.zutabetik aurrera: Aztertu nahi diren gainontzeko aldagaiak.

OHARRA: 6. zutabean galdutako balioak baldin badaude, galdutako baliorik ez duen hurrengo zutabeak finkatuko du burbuilen tamaina. Beraz, populazioak dituzten fitxategiek ez dute galdutako baliorik izan behar!

'Elkartu_gapminder()' funtzioak aipaturiko egitura duen txt motako fitxategia bueltatzen du. Fitxategia sortzen duenean, R-n bertan editore bat ateratzen da, eta aldagaien gainean sakatuz izenak aldatzeko aukera ematen du, egindako aldaketak automatikoki gordetzen direlarik.

OHARRAK

Lehenik eta behin erabili nahi diren adierazleen fitxategiak karpeta berean gorde behar dira. Adierazleen fitxategien artean populazioa duen fitxategia ez izatekotan, EU 27-ko populazioa (Eurostat-eko web orritik jaitzia) eta EAEkoa (Eustat-eko web orritik jaitzia) biltzen dituzten fitxategiak ere egon beharko lirateke karpetan. Fitxategi guztiek csv formatuan egon behar dute.

Funtzioa exekutatu aurretik R-n adierazleen fitxategiak dauden direktorioa zehaztu behar da. Hau bi eratan egin daiteke; R-n hurrengo ikonoak "Archivo → Cambiar dir...→ (dagokion karpeta)" sakatuz edo setwd("gordeta dagoen tokiaren helbide guztia, komatxo artean") komandoa jarritz. Funtzioa exekutatu ondoren sortuko den fitxategia ere helbide horretan gordeko da.

SARRERAKO FITXATEGIEN EGITURA:

Adierazle bakoitzeko fitxategiek hurrengo egitura izan behar dute (populazioarenak izan ezik, adierazleetako bat ez bada):

- Lehen zutabean herrialde guztien izenak joango dira (ez utzi lehen gelaxka hutsik, izenburu bat eman). Kontuan izan herrialdeen izenak era berean idatzita (azentu-markak orain dauden bezala jarrita, parentesirik gabe...) egon behar direla fitxategi guztietan.
- Gainontzeko zutabeetan urteak egongo dira. Zutabe hauetako izenburuak urteak izango dira.

Horrez gain, hasierako fitxategi guztietan ikur bera erabili beharko da galdutako balioak adierazteko (":", hain zuzen ere) eta baita balio dezimalak adierazteko (".", alegia) ere.

ADIBIDEA

Elkartu_gapminder(lista1="indicadores.txt", fitxategia_izena="Euskadi_en_EU27",
gaztelania=FALSE, euskara=FALSE, ingelesa=FALSE, populazioa=TRUE)

2. MOTION CHART MOTAKO GRAFIKOA SORTZEKO KODEA SORTZEN DUEN FUNTZIOA R-n: 'Motion_chart()'

DESKRIBAPENA

Datuak bost dimentsio ezberdinetan aztertzeako aukera ematen duen Motion Chart delako grafiko dinamikoa eraikitzeako kodea sortzen duen funtzioa. Testu formatuan ematen duen kode horrek Google Visualization API-ari erreferentzia egiten dio eta web-orri batean txerta daiteke. 'Elkartu_gapminder()' funtzioa exekutatu ondoren lorturiko fitxategia erabiltzen du sarrerako fitxategi moduan.

'Motion_chart()' funtzioak beste lau funtzio erabiltzen ditu barnean, 'my.gvis()', 'my.gvisMotionChart()', 'my.gvis_web()' eta 'my.gvisMotionChart_web()' funtzioak, alegia. (Oharretan, azalpena).

ERABILERA

Motion_chart(lista1, fitxategia_izena, izenb_ind, euskara=TRUE, irteerako_fitxategia,
populazioa=FALSE, chartid, grafikoa=FALSE, kodea=FALSE)

...

ARGUMENTUAK

lista1:	txt motako fitxategia da, non erabili nahi diren adierazleen csv fitxategien izenak (luzapenik gabe) zehaztuko diren, hutsunez bereiziak. 'Elkartu_gapminder()' funtzioaren 'lista1' argumentuan erabiltzen den txt berdina izan behar da.
fitxategia_izena:	Bistaratu nahi den fitxategiaren izena. 'Elkartu_gapminder()' funtzioan 'fitxategia_izena' argumentuari emandako izen bera izan behar da.
izenb_ind:	NULL bada, adierazleen aldagaiei 'lista1' parametroari esleitutako fitxategian azaltzen diren izenak emango zaizkie. Kontrako kasuan, karaktere-bektore bat zehaztu daiteke, adierazleei eman nahi zaizkien izenekin.
gaztelania, euskara, ingelesa:	Sarrerako zein irteerako fitxategien hizkuntza. Beti aukeratu behar da bat eta bakarra. Nahi den hizkuntzari TRUE esleituko zaio.
irteera_fitxategia:	Irteerako fitxategiari emango zaion izena.
populazioa:	Defektuz FALSE. Populazioa zehazten duen adierazlea egotekotan TRUE esleitu.
chartid:	Grafikoa identifikatzeko izen bat.

grafikoa, kodea: Beti aukeratu behar da bietako bat eta bakarra. Sarean grafikoa ikusteko, grafikoa=TRUE, kodea soilik sortzeko, kodea=TRUE.

XEHETASUNAK

Grafiko hau dispersio- edo sakabanatze-diagrama baten hedapena dela esan daiteke, aldagai anitzeko kasurako. Izan ere, behin x eta y ardatzetako aldagaiak aukeratuta, beste aukera batzuk egin daitezke, hala nola, burbuilen tamaina zein kolorea zehaztu edo burbuilen bilakaera denboran zehar ikusi. Bost aldagai horiek erabiltzailearen aukeraren menpe daude, beraz dinamikoa izateaz gain, interaktiboa ere bada.

Hain zuzen, hauek dira egin daitezkeen aukerak:

- X ardatzeko aldagaia.
- Y ardatzeko aldagaia.
- Denbora neurtuko duen aldagaia. Egia esan, hau ezin da aukeratu, sarrerako fitxategian denbora neurtzen duen aldagaia izango da, normalean urtea.
- Burbuilen kolorea zehaztuko duen aldagaia.
- Burbuilen tamaina zehaztuko duen aldagaia.

Datuak denboran zehar aztertzeko aukerari esker, grafiko honek beste grafiko tradizional batzuetan ikusi ezin diren eredu eta joerak identifikatzeko aukera ematen du.

Datuak hautatzeko aukera emateaz gain, hiru grafiko ezberdinetan ikus daitezke:



Motion Chart motako grafikoa. (Aipaturikoa).



Barra-diagrama dinamikoa. Barra-diagrama tradizional bati denbora neurtzen duen barra dinamiko bat gehitzen dio. Honela barren bilakaera denboran zehar azter daiteke.

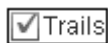


Lerro-diagrama estatikoa. Burbuila bakoitzak denboran zehar duen bilakaera estatikoki adierazten du, lerro baten bidez.

Horrez gain, lagungarriak diren beste zenbait tresna eskaintzen ditu, hala nola:



Eskala aldaketa. Eskala lineal eta logaritmikoaren artean hautatzeko aukera ematen du. Batzuetan aldagai batek besteek baino maiztasun askoz handiagoa duenean, aldagaien arteko aldeak ez dira ondo ikusten. Eskala logaritmikoa izan daiteke hau konpontzeko modu bat.



Ibilbideak. Ibilbideen erabilerak, denboran zehar aldagaiek dituzten joerak ikustea ahalbidetzen du.



Abiadura. Grafikoaren erreprodukzio-abiadura aukera daiteke.

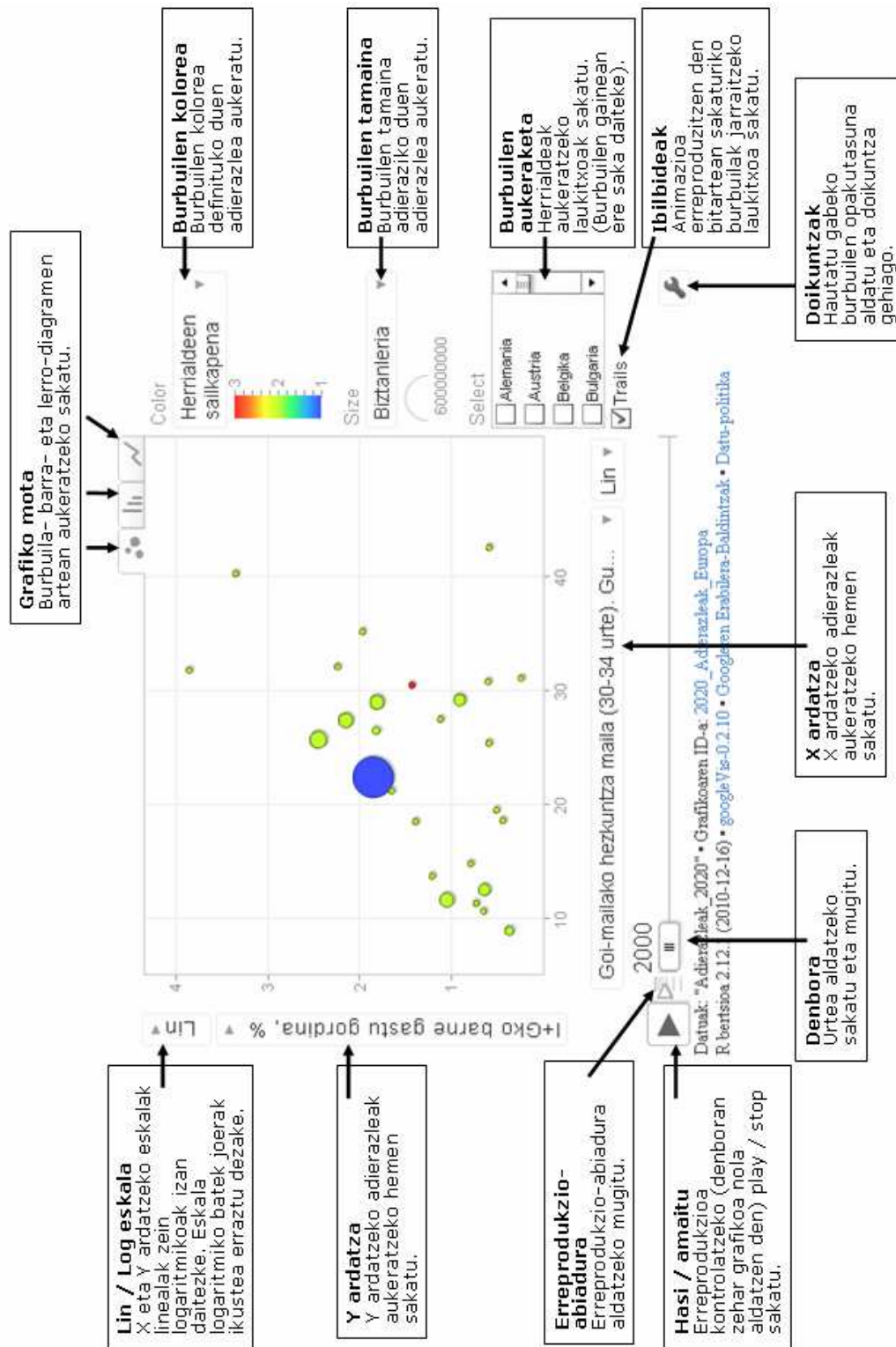


Zoom-a. Irudia hurbiltzeko aukera ematen du, grafikoaren zati jakin batean zentratzeko.



Konfigurazio-panela. Kontrol mugikorraren bidez hautatu gabeko burbuilen (edo barren edo lerroen, kasuan kasu) gardentasuna doitzeko aukera ematen du.

Jarraian grafikoaren egitura eta funtzionamendua grafikoki azaltzen dituen irudi bat gehitzen da:



www.gapminder.org
orritik moldatuta

OHARRAK

Funtzio hau erabili ahal izateko R-n 'googleVis' liburutegia instalatuta egon behar da. Instalatuta ez badago, hurrengo pausoak jarraitu: Paquetes → Instalar paquete(s) → "CRAN mirror bat aukeratu" → "googleVis paketea aukeratu". Behin instalatuta dagoenean, ez da hau berrir egin beharko.

Lehenik 'Elkartu_gapminder()' funtzioa exekutatu behar da. Ondoren, 'lista1' eta 'fitxategia_izena' argumentuei 'Elkartu_gapminder()' funtzioan esleitutako izen berberak esleitu beharko zaizkie.

Kasu honetan ere, erabili nahi diren fitxategi eta R-ko programa guztiak karpeta berean egon behar dira, eta aurreko funtzioan egin den bezalaxe, R-k direktorio horretara apuntatu behar du.

SARRERAKO FITXATEGIAREN EGITURA:

Fitxategiak izango duen egitura 'Elkartu_gapminder()' funtzioa exekutatzean lorturikoa fitxategiarena izango da.

ADIBIDEA

```
Motion_chart(lista1="indicadores.txt", fitxategia_izena="Euskadi_en_EU27",
izenb_ind=nombres_indicadores_castellano, euskara=TRUE,
irteerako_fitxategia="codigo_web", populazioa=TRUE, chartid="Euskadi_en_la_UE_27",
kodea=TRUE)
```

2. Time Line motako grafikoa

TIME LINE MOTAKO GRAFIKOA SORTZEKO KODEA SORTZEN DUEN FUNTZIOA R-n: 'TimeLine()'

DESKRIBAPENA

Denborazko serie bat duen fitxategi bat hartuta, Google Visualization APIko grafiko eta "control" deritzan egiturak erabiliz TimeLine motako grafikoa eraikitzeke kodea sortzen duen funtzioa.

Grafiko hau lerro diagrama batez eta kontrol deritzon egitura batez eratuta dago. Kontrol mugikorak denbora-tarte jakin batean zentratzeko aukera ematen digu.

ERABILERA

```
TimeLine(fitxategia_izena, chart.type, chartid, controlid, chart.options, control.options,
control.state, euskara=TRUE)
```

ARGUMENTUAK

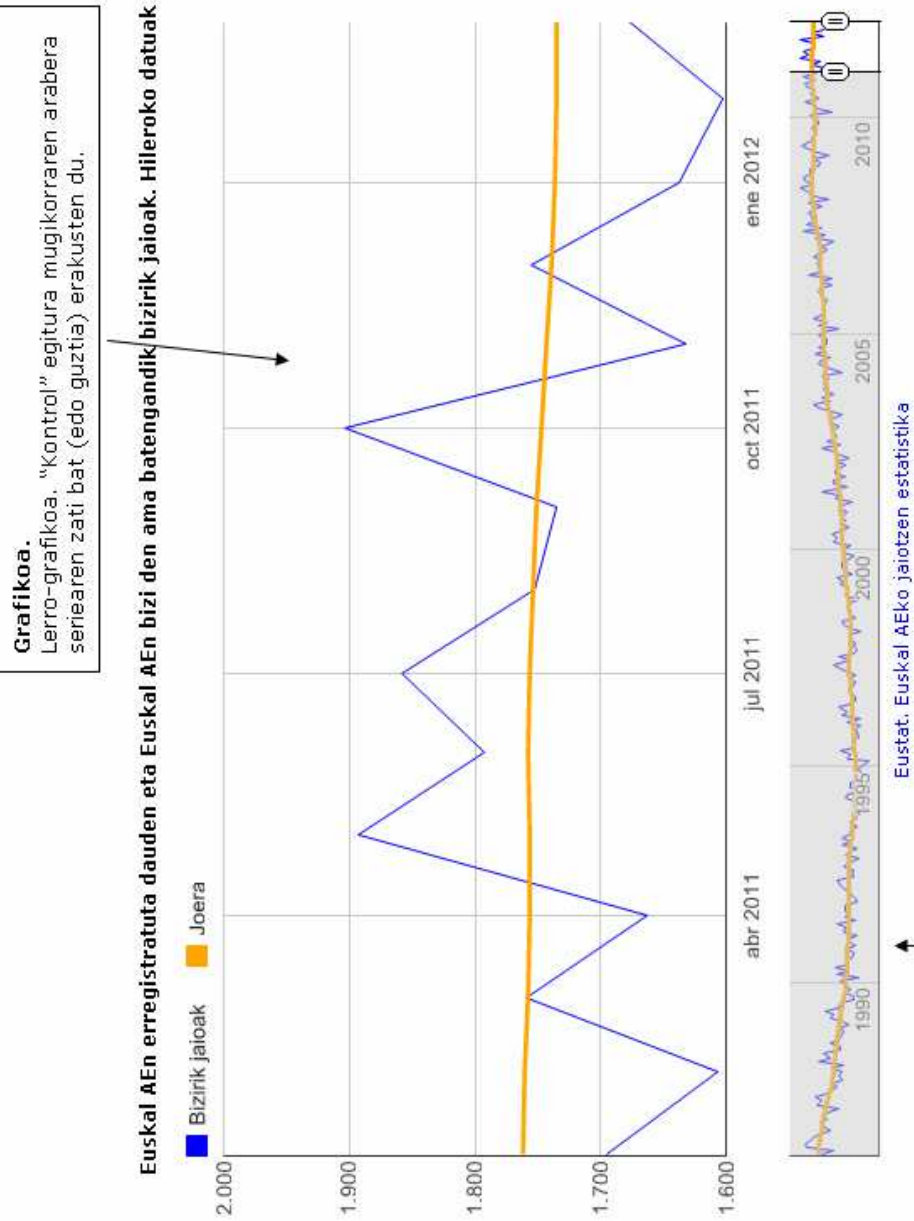
fitxategia_izena:	Sarrerako csv formatuko fitxategiaren izena.
chart_type:	Irudikatu nahi den grafiko mota. Kasu honetan "LineChart" (lerro-diagrama) da baina programa berak Google Visualization API-ko 'AreaChart', 'ComboChart' edo 'ScatterChart' kasuetarako ere funtzionatuko luke.
chartid:	Grafikoa identifikatzeko izen bat.
controlid:	Kontrol deritzon egitura mugikorra identifikatzeko izen bat.
chart.options:	Grafikoak izan ditzakeen aukera guztiak. Bai 'LineChart' grafikoak eta bai aipaturiko gainontzekoek ere konfigurazio aukera asko eskaintzen dituzte, hala nola, serieen kolorea eta zabalera, x eta y ardatzen itxura eta tamaina, testu-mota zein tamaina, edo grafikoaren altuera eta zabalera hauta daitezke. Funtzioa definitzean, horietako batzuk aukeratu dira eta besteak tresnak defektuz ematen dituenak dira.
control.options:	Kontrol egitura mugikorak izan ditzakeen aukera guztiak. Kasu honetan ere dauden aukera guztietatik batzuk aukeratu dira eta beteak tresnak defektuz ematen dituenak dira.
control.state:	Hasieran kontrol egitura (eta beraz grafikoa ere) zentratu nahi dugun denbora-tartea. Bi osagaiko bektore bat izango da; lehenengoak hasierako data eta bigarrenak bukaerako data zehaztuko dituzte, komatxo artean eta "uuuu-hh-ee" formatuan. Besterik ezean, grafikoa denbora-tarte osoan zentratuko da.
azpitolua:	Grafikoaren azpitolua.
gaztelania, euskara, ingelesa:	Sarrerako zein irteerako fitxategien hizkuntza. Beti aukeratu behar da bat eta bakarra. Nahi den hizkuntzari TRUE esleituko zaio.

XEHETASUNAK

TimeLine grafikoa lerro diagrama interaktibo bat da, aldi berean denborazko serie bat baino gehiago irudika ditzakeena. Arratoia denborazko serieetako bat irudikatzen duen lerroa hurbiltzean, aukeraturiko datan serieak duen balioari buruzko informazioa erakusten duen elkarriketa-koadro bat azaltzen da.

Lerro-grafikoaren azpian, denbora tarte batean zentratzeko aukera eskaintzen duen tresna gehitzen da. Gainera, denbora tarte zehatz batean zentratzean, grafikoa automatikoki bir-eskalatzen da.

Jarraian grafikoaren egitura eta funtzionamendua grafikoki azaltzen dituen irudi bat gehitzen da:



"Kontrol" egitura mugikorra.
Nahi dugun denbora-tartean zentratzeko aukera ematen digu. Hau mugitzean grafikoa automatikoki bir-eskalatzen da.

OHARRAK

Funtzioa exekutatu aurretik R-n erabiliko den csv formatuko fitxategia dagoen direktorioa zehaztu behar da. Hau bi eratan egin daiteke, R-n hurrengo ikonoak “Archivo → Cambiar dir...→ (dagokion karpeta)” sakatuz edo setwd(“gordeta dagoen tokiaren helbide guztia, komatxo artean”) komandoa jarritz. Funtzioa exekutatu ondoren sortuko den fitxategia ere helbide horretan gordeko da.

Behin hau eginda, erabiliko den fitxategia kargatu behar da, load(“fitxategiaren izena, komatxo artean”) eta baita erabiliko den funtzioa ere, source(“funtzioaren izena, komatxo artean”) aginduarekin.

SARRERAKO FITXATEGIAREN EGITURA:

Fitxategiak hurrengo egitura izan behar du:

- Lehen zutabearen datak agertuko dira (ez utzi lehen gelaxka hutsik, izenburu bat eman), “ee/hh/uuuu” formatuan.
- Bigarren zutabearen urdinez margotu nahi den serieari dagozkion datuak sartuko dira (ez utzi lehen gelaxka hutsik, izenburu bat eman), kasu honetan jaiotzen seriearen datu gordinei dagozkionak.
- Hirugarren zutabearen laranja margotu nahi den serieari dagozkion datuak sartuko dira (ez utzi lehen gelaxka hutsik, izenburu bat eman), kasu honetan jaiotzen seriearen joerari dagozkionak.

ADIBIDEA

```
TimeLine(fitxategia_izena="IPI_euskara", chart.type="LineChart", chartid= "chart_id",
controlid= "control_id", chart.options= list(chartArea=list(height="80%", width="90%"),
legend=list(position="top"), series=list(seriea1=list(color='blue', lineWidth=1),
seriea2=list(color='orange', lineWidth=1))), control.options= list(filterColumnIndex=0,
ui=list(chartType="LineChart", chartOptions=list(chartArea=list(width="90%",
height="50%"), series=list(seriea1=list(color="blue"), seriea2=list(color="orange"))),
minRangeSize=2592000000)), control.state=list("2011-01-01", "2011-12-01"),
euskara=TRUE)
```

3. Google Earth bidezko bistaratzea

1991an, 1996an eta 2001ean egin zen bezala, 2012an ere EAE-ko zentsu-sekzioen tipologiak sailkapena egin da. Azterketa honen helburua zentsu-sekzioak ezaugarri berdinen arabera biltzea da; ezaugarri komunak dituzten sekzioak bilduz. Honela, tipo edo sail bereko sekzioak haien artean ahalik eta homogeneoenak izango dira eta gainontzekoekiko ahalik eta ezberdinenak.

Analisi mota hau oso erabilgarria da arlo ezberdinetan, hala nola: laginketa geruzatuak diseinatzeko, hiri plangintzarekin loturiko azterketetan edota gizartearen egoera sozio-ekonomikoa zehazkiago ezagutzeko.

ZENTSU-SEKZIOEN ANALISIA. METODOLOGIA

Lehenik Estatistika sozio-demografikoen datu-basetik (BSD) familia etxebizitza nagusiak eta bertan bizi diren biztanleak hautatu dira. Ondoren, aztertu nahi diren aldagaiak hautatu eta gehienak (errenta izan ezik) modalitatetan zatitu dira (ez dago zatiketa mugarik, izan ere erabili den Analisi Faktorial Anizkoitzak ez baitu modalitate kopurua mugatzen).

Hautaturiko aldagai gehienek norbanakoen ezaugarri nagusiak islatzen dituzte, hala nola, sexua, adina, titulazioa, errenta, lan-munduarekiko erlazioa, egoera soziolinguistikoa eta jaioterria. Horiez gain, bukaeran etxebizitzarekin (tamaina eta eraikitze-urtea) eta familiarekin loturikoak gehitzen dira.

Behin aztergai izango diren aldagaiak hautatuta, zentsu-sekzio bakoitzari modalitate bakoitzean duen maiztasuna esleitu zaio, balio erlatibotan, aldagai bakoitzeko modalitate guztien batura 100ekoa izanik (errentaren kasuan izan ezik). Familia-errenta ere modalitatetan sailkatu da, nahiz eta analisisan parte hartu ez, sailak deskribatzean lagungarria suertatu delako.

Pausu hauek jarraituz sorturiko datu-matritzetik abiatuta baina errentak eta ama-hizkuntza kontuan hartu gabe Analisi Faktorial Anizkoitza egin eta lehen 7 faktoreak gorde dira, hasierako bariantzaren (informazio-kantitatearen) %70.56a adierazten baitute. Ondoren sailkapen hierarkiko bat egin da 7 faktoreekin. Honela 15 sail lortu dira, eta egonkortasun prozesu bat aplikatu da. Baina gehiegi zirelakoan, sail batzuk elkartu dira gertutasun eta urruntasun irizpideak erabiliz, honela 12 sailera pasaz. Azkenik 12 sail horien erdiguneak hartu eta hazi gisa erabili dira partiketa berri bat osatzeko, kmeans edo nuées dynamiques prozesuaren bidez.

EAEko TIPOLOGIEN ANALISIA GOOGLE EARTH-EN BIDEZ BISTARATZEKO FUNTZIOA R-n: 'KmlSortu()'**DESKRIBAPENA**

Funtzio honek shp fitxategi batetik abiatuz eta Rko *maptools* liburutegiko *kmlPolygon()* funtzioa moldatuz, Google Earth programaren bidez bistara daitekeen kml fitxategia sortzen du.

Google Earth datu espazialak begi bistaz eta era erraz batean bistaratu eta aztertzeko aukera ematen duen dohaineko tresna erraz eta ezaguna da.

ERABILERA

KmlSortu(n_tipol, fitxategia, gaztelania=TRUE, alpha)

ARGUMENTUAK

n_tipol:	Behin tipologiaren analisia egin ondoren lortutako tipo kopurua.
fitxategia:	Behar den informazio guztia duen shp formatudun fitxategia.
gaztelania, euskara, ingelese:	Sortuko den bistaratzearen hizkuntza. Beti aukeratu behar da bat eta bakarra. Nahi den hizkuntzari TRUE esleituko zaio.
	Google Earth-eko sekzioak koloreztatze erabiliko diren koloreen

alpha:	gardentasuna zehazten duen parametroa. Defektuz 120 izango da.
legenda:	Defektuz FALSE izango da eta ez da legenda erakutsiko. TRUE denean pantailaren ezker aldean tipologiaren deskribapena duen legenda gehituko da.

XEHETASUNAK

Erabiltzailea Google Earth tresnak eskaintzen dituen aukera guztiez baliatu daiteke EAEko zentsu sekzioak tipo edo sail ezberdinetan sailkatzen dituen mapa koloreztatua erabiltzen duenean.

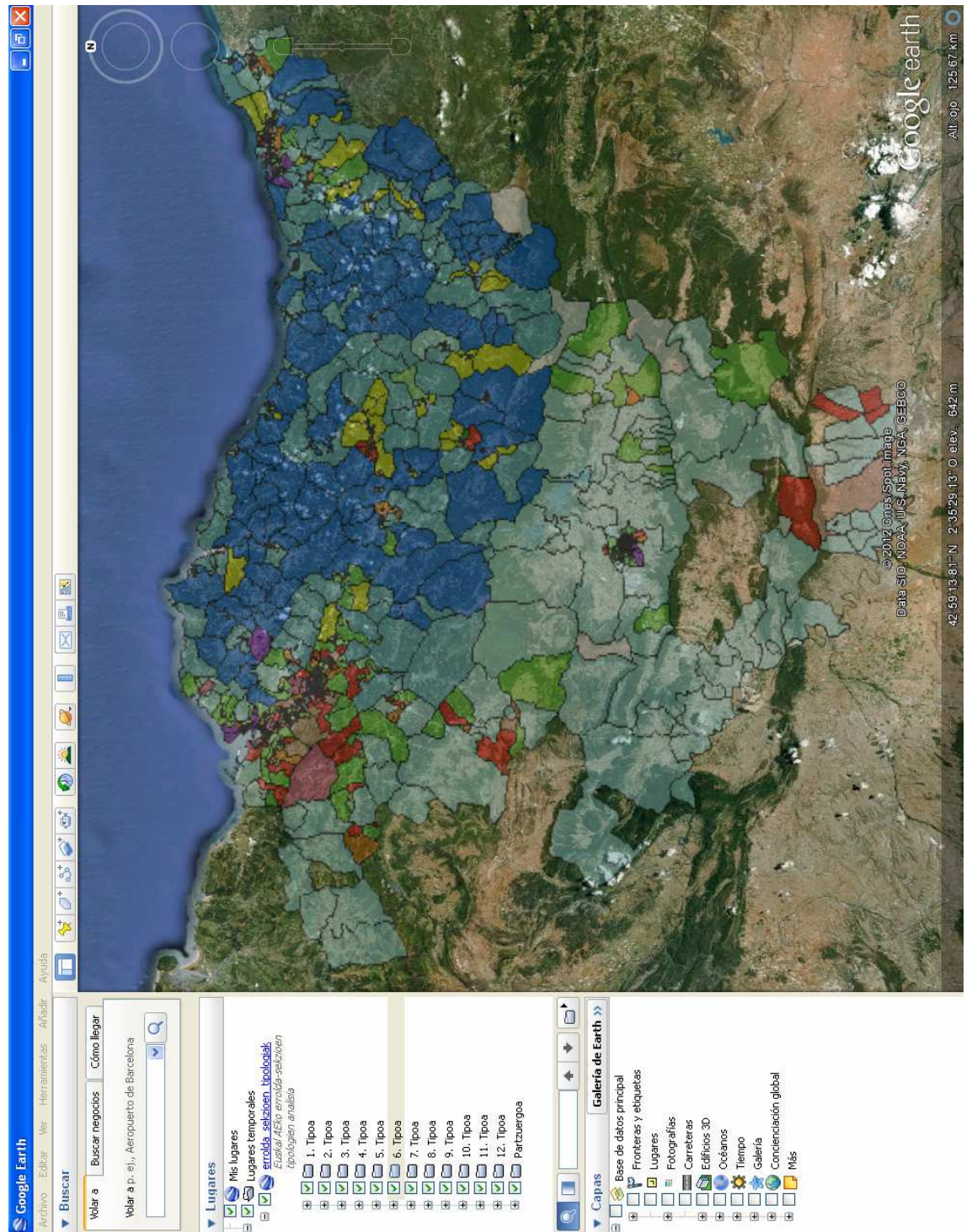
Sorturiko tipo edo sailen kml fitxategiak baditu beste aukera batzuk ere, tresnak eskaintzen dituenek gain.

Ezkerreko goitibehera-koadroa sakatuz, erabiltzaileak interesatzen zaizkion tipo edo sailak soilik aukera ditzake. Honela, mapan aukeraturiko tipoei dagozkien zentsu-sekzioak soilik koloreztatuko dira, eta goitibehera-koadroan aipaturiko tipoak zabaltzean haietan sailkatutako sekzioen zerrenda agertuko da.

Helbide batera hurbildu ahala, kapa gero eta gardenagoa dela ikus daiteke, bere azpiko kaleak, eraikuntzak, parkeak... ikustea ahalbidetuz.

Maparen gainean sakatuz zentsu-sekzioaren informazioa (udalerriaren izena; barrutia; sekzioa; tipo edo saila; 16 urtez beherakoen, 65 urtez gorakoen, euskaldunen eta goimailako tituludunen portzentajea; Eustat-eko web orrian dagoen EAEko tipologiaren analisiaren dokumentaziorako link-a) duen elkarriketa-koadro bat agertzen da.

Jarraian EAEko zentsu-sekzioen tipologiaren analisiaren bistaratzeak Google Earth-en duen itxura azaltzen duen irudi bat gehitzen da.



OHARRAK

Kasu honetan ez da beharrezkoa funtzioa kargatu aurretik shp fitxategia gorderik dagoen direktorioa zehaztea, ondoren egin daiteke, funtzioa exekutatu baino lehen.

Direktorioa bi eratan zehaztu daiteke, R-n hurrengo ikonoak “Archivo → Cambiar dir...→ (dagokion karpeta)” sakatuz edo setwd(“gordeta dagoen tokiaren helbide guztia, komatxo artean”) komandoa jarritz. Funtzioa exekutatu ondoren sortuko den kml fitxategia ere helbide horretan gordeko da.

SARRERAKO FITXATEGIAREN EGITURA:

EAEko zentsu-sekzioak tipo edo sail ezberdinen arabera koloreztatzen dituen fitxategia sortu ahal izateko shapefile deritzon shp fitxategia behar da.

Shapefile delakoa biltegitratze digitaleko formatu bektoriala da, non elementu geografikoen eta haiei lotutako atributuen kokapena gordetzen den. Hainbat fitxategi informatikoz sortua dago. Haiek dira, hain zuzen, hemen erabili den shp-aren osagaiak:

- .shp. Objektuen entitate geometrikoak dituen fitxategia da.
- .shx. Entitate geometrikoen indizea duen fitxategia da.
- .dbf. Objektuen atributuen informazioa duen datu-basea da, dBASE formatuan. Hau da fitxategi honek duen egitura:
 - o Sekzioaren kodea, bai agregatuta eta bai desagregatuta (probintzia, udalerria, barrutia eta sekzioa).
 - o Sekzioko guztizko indibiduo kopurua.
 - o Sekzioari dagokion tipologia.
 - o Sekzioa dagoen udalerriaren izena.
 - o Elkarrizketa-koadroan joango diren aldagaiak: 16 urtez beherakoen, 64 urtez gorakoen, euskaldunen eta goi-mailako titulaziodunen portzentajeak.

OHARRA: Erabilitako shp fitxategia 2009ko sekzionatua biltzen duena da, partzuergoak barne.

ADIBIDEA

KmlSortu(n_tipol=12, fitxategia="Seccion_regionT.shp", gaztelania=TRUE, alpha=150)

Ondorioak

Datuen bistaratzeak taula eta fitxategietako datuak behatuz soilik ikus ezin daitezkeen ereduak, datuen egitura eta joerak ikustea ahalbidetzen du. Grafikoei esker datuei doitutako eredu konplexuak ikus daitezke eta baita beren baliagarritasuna ebaluatu ere. Hain zuzen, datuak bistaratu eta sintetizatzeak haien erabilpena indartzen du eta datu estatistikoak ezagutza bihurtzen ditu.

Bistaratzeak eskaintzen dituen abantailez baliatzeko, beharrezkoa da metodologia zehatz bat eta bigarren atalean zerrendatutako oinarritzko printzipioak kontuan hartzea.

Grafiko bat irudikatzen denean, informazio kuantitatibo eta kategorikoa kodifikatu egiten da, normalean, posizioaren, tamainaren, ikurren eta koloreen bitartez. Grafikoa aztertzean, informazio hori begi-bistaz deskodetzen da. Honela, deskodetze prozesua eraginkorra bada, metodo grafiko bat ona dela deritzogu.

Begi-bista bidezko deskodetze prozesu ez eraginkor batera eramaten duen bistaratze metodo batek datuen ezaugarri garrantzitsuenak ez detektatzea edo informazioaren pertzepzioa distorsionatzea eragin dezake.

Askotan, datuak bistaratze tresnekin aztertuz, datuak jasotzean sor daitezkeen ezustekak eta gaizkiulertuak ekidin daitezke.

Egindako lanaren helburua Eustat erakundeak dituen datu-mota ezberdinei grafiko egokienak esleitzea zen. Batzuetan zaila izan da R-script orokor eta parametrizatu bat egitea, fitxategiko datuen ezaugarri oso lotuta gertatzen delako.

Zenbaitetan script-ak exekutatu aurretik fitxategiari zenbait aldaketa egin behar zaizkio eta R programazio-inguruneak aldaketa horiek egiteko tresna malgua eskaintzen du.

Dagoeneko Eustat-eko web orrian txertatuta dauden Google Visualization APIko grafikoaren kasuan, asmoa da berauek sortzen dituzten kodea erabiltzaileentzat eskura uztea, beraien blog eta orrietara grafikoak txertatzeko aukera izan dezaten.

Gaur egungo teknologiak orain dela oso gutxi arte ezinezkoa zena egitea erraztu du, hala ere metodoz egiteko lana da eta estatistika ofizialak teknika hauek datuen zabalkuntza eta analisi arloetan eskaintzen dituzten aukerez baliatzen hasi besterik ez dira egin.

Bistaratze arloan egindakoak eta egiteke dagoenak ondorio eta eragin zuzena izango du estatistika ofizialaren erabilpenean eta erabiltzaile ororengan.

Bibliografia

- [1] YURRAMENDI, Y. (2011)

Estatistika ofizialak sintetizatzeko eta ikustarazteko metodo grafikoen eta teknologia berrien azterketa. Konputazio Zientziak eta Adimen Artifiziala saila.

- [2] CLEVELAND, W.S (1985)

The Elements of Graphing Data. Monterey, CA: Wadsworth.

- [3] YOUNG, F.W.; VALERO-MORA, P.M. AND FRIENDLY, M. (2006)

Visual Statistics. Seeing Data with Dynamic Interactive Graphics. Wiley, New Jersey.

- [4] GALBETE, E.; ADIN, A.; ARAMENDI, J.; IZTUETA, A. AND YURRAMENDI, J. (2012)

Análisis de tipologías de las secciones censales de la CAE. Visualización en Google Earth. IX Congreso Vasco de Sociología y Ciencia Política.

- [5] ESCOFIER B. AND PAGÈS, J. (1992)

Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación. Euskal Herriko Unibertsitatea, Argitalpen zerbitzua.

- [6] LÊ, S., JOSSE, J. AND HUSSON, F.. (March 2008)

FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, Volume 25, Issue 1.

- [7] EUSTAT (1996)

Análisis de tipologías de las secciones censales de la C.A. de Euskadi. Vitoria-Gasteiz: Instituto Vasco de Estadística.

- [8] EUSTAT (2001)

Análisis de tipologías de las secciones censales de la C.A. de Euskadi. Vitoria-Gasteiz: Instituto Vasco de Estadística.

- [9] GREENACRE (2008)

La práctica de análisis de correspondencias. Vitoria-Gasteiz: Instituto Vasco de Estadística. Fundación BBVA, Bilbao. Rubes Editorial.

- [10] **R Graph Gallery:** <http://gallery.r-enthusiasts.com/>
- [11] **R-Project:** <http://www.r-project.org/>
- [12] **FactoMineR:** <http://cran.r-project.org/web/packages/FactoMineR/index.html>
- [13] **Google Visualization API:** <https://developers.google.com/chart/>
- [14] **API. Wikipedia:**
http://es.wikipedia.org/wiki/Interfaz_de_programaci%C3%B3n_de_aplicaciones
- [15] **Google Earth. Eduteka:** <http://www.eduteka.org/GoogleEarth.php>