

PRÀCTICA: ANÀLISI DELS TRETOS EFECTUATS PER LA
POLICIA CONTRA HOMES I DONES ALS EUA

ALUMNE: RAFAEL TOBÍAS MIRAS

DATA: 15/01/2023

ASSIGNATURA: PROGRAMACIÓ EN SCRIPTING

Índex

1. Objectius del projecte	3
2. <i>Scripts</i>	3
2.1. <i>Script A</i>	3
2.2. <i>Scripts B</i>	4
2.2.1. <i>Script b1.sh</i>	4
2.2.2. <i>Script b.sh</i>	4
2.2.3. <i>Script b.awk</i>	5
2.3. <i>Script c.sh</i>	6
3. Taula de transformacions dels <i>scripts b</i>	8
4. Taula d'agrupacions i categorització de l' <i>script c</i>	8
5. Diagrama de flux.	9
6. Conclusions.	10
7. Bibliografia	10

1. Objectius del projecte

L'objectiu del projecte és generar un informe, en format HTML, que analitzi el número de trets efectuats per la policia als Estat Units d'Amèrica a homes i dones entre els anys 2015 i 2020. S'intentarà tenir una visió general de la situació sense entrar massa en detall. Es diferenciarà en alguns apartats entre sexe i en altres s'agruparà la informació a nivell general.

Un altre dels objectius és poder comprovar com l'ús de l'*scripting* pot aconseguir estalviar temps en la manipulació i poder analitzar les dades d'una manera més eficient. D'aquesta manera s'evitaran manualitats que poden conduir a errors que arribin a afectar l'anàlisi final. S'ha de tenir en compte, però, que el joc de dades inicial hauria de tenir la mateixa estructura per no haver de retocar el codi emprat.

La memòria intentarà explicar els objectius de cadascun dels scripts implementats, la transformació de les diferents tipus de dades o camps, així com un diagrama de flux que mostri el procés. S'inclourà, també captures de pantalla que permetin fer més entenedor el funcionament dels diferents *scripts*.

Finalment es resumirà breument les tasques apreses i les dificultats que s'han anat trobat en la realització de la pràctica així com les millores que es poden aplicar al projecte.

2. Scripts

2.1. Script A

L'*script* a.sh és el primer de tots. L'ús que té en el projecte és el següent:

- Guardar la *url* d'on prové el joc de dades en format *csv*.
- Eliminar aquells fitxers que s'hagin generat anteriorment per llençar d'una manera neta l'anàlisi.
- Càlcul del total de files, columnes que té l'arxiu.
- Mostra de les dades que es demanen: nombre de registres, nombre de columnes, *url* de descarrega, format del fitxer, i tipus de dades de cadascuna de les columnes.

```
psccrtobias@psccrtobias: ~/Documents/PRA
$ ./a.sh -v
Inici de l'execució de l'script a.sh
-----
Descarrega del fitxer shootings.csv
--2023-01-15 11:10:14-- https://drive.google.com/uc?export=download&id=1V_Kk0CKZ1zRa4QTFf0MS0HXZsE3uG-X6
S'està resolent drive.google.com (drive.google.com)... 142.250.185.14, 2a00:1450:4003:80c::200e
S'està connectant a drive.google.com (drive.google.com)[142.250.185.14]:443. connectat.
HTTP: s'ha enviat la petició, s'està esperant una resposta. 393 See Other
Ubicació: https://doc-0s-7k-docs.googleusercontent.com/docs/securesc/ha0937gcuc7L7deffksulhgsh7mbp1/euHl01ah7nu68g6d49j9brC9082t1jg/1673777400000/14593360732861647835/*1V_Kk0CKZ1zRa4QTFf0MS0HXZsE3uG-X6?e=download&uiid=0bb1f65f-688c-4875-bd52-7b0a3b061c39 [es segueix]
Perill: les vldcards no estan suportades en HTTP.
--2023-01-15 11:10:15-- https://doc-0s-7k-docs.googleusercontent.com/docs/securesc/ha0937gcuc7L7deffksulhgsh7mbp1/euHl01ah7nu68g6d49j9brC9082t1jg/1673777400000/14593360732861647835/*1V_Kk0CKZ1zRa4QTFf0MS0HXZsE3uG-X6?e=download&uiid=0bb1f65f-688c-4875-bd52-7b0a3b061c39
S'està resolent doc-0s-7k-docs.googleusercontent.com (doc-0s-7k-docs.googleusercontent.com)... 142.250.184.1, 2a00:1450:4003:808::2001
S'està connectant a doc-0s-7k-docs.googleusercontent.com (doc-0s-7k-docs.googleusercontent.com)[142.250.184.1]:443. connectat.
HTTP: s'ha enviat la petició, s'està esperant una resposta. 200 OK
Mida: 527909 (510K) [text/csv]
S'està desant a: «shootings.csv»
shootings.csv 100K[=====] 515,54K 2,85MB/s en 0,2s
2023-01-15 11:10:16 (2,85 MB/s) - s'ha desat «shootings.csv» [527909/527909]
URL de descarrega: https://drive.google.com/file/d/1V_Kk0CKZ1zRa4QTFf0MS0HXZsE3uG-X6/view?usp=sharing
Format del fitxer: csv
Nombre de columnes: 15
Nombre de registres: 4895
Tipus de dades per columna:
1. id: Number
2. name: Text
3. date: Date
4. manner_of_death: Text
5. armed: Text
6. age: Number
7. gender: Text
8. race: Text
9. city: Text
10. state: Text
11. signs_of_mental_illness: Boolean
12. threat_level: Text
13. flee: Text
14. body_camera: Boolean
15. arms_category: Text
Final de l'execució de l'script a.sh.
-----
```

Imatge 1: Sortida de l'script a.sh. Elaboració pròpia.

2.2. Scripts B

Els *scripts* b son tres *scripts* en format *sh* o *awk* que s'encarreguen de manipular el fitxer per modificar les dades. Modificaran registres i crearan noves columnes.

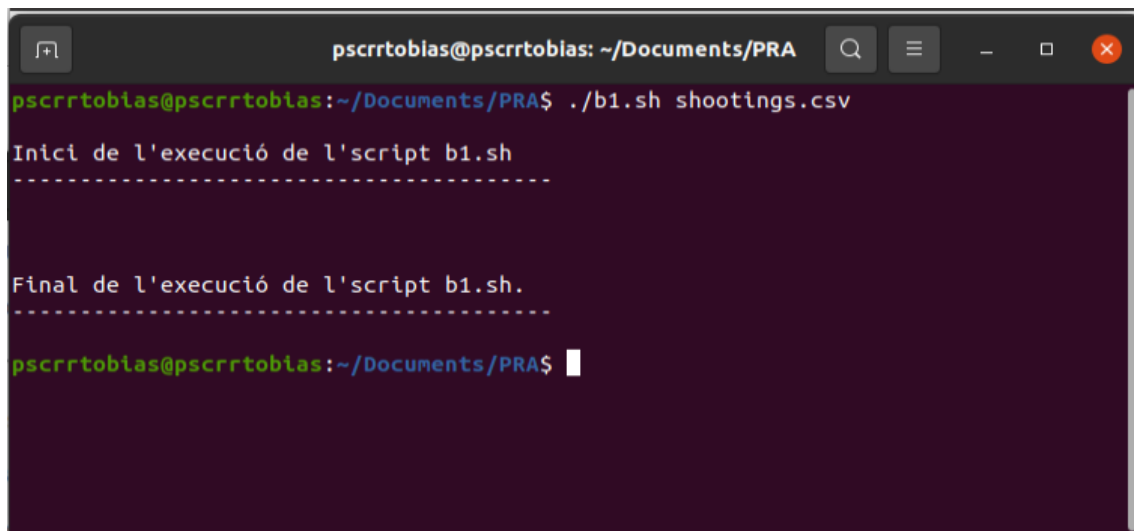
2.2.1. Script b1.sh

L'script b1.sh s'encarrega de:

- Modificar els valors de la columna *gender*, que apareixen com a M (de l'anglès *male*) per a homes i F (de l'anglès *female*) per a dones, per *male* i *female*.
- Eliminar aquells registres de la columna *age* que no segueixen el patró XX.0 ja que no son molts i no sabem si son correctes.
- Es revisarà que el nou fitxer no contingui valors buits i es comptaran els errors i a quines línies es troben.

En aquest *script* s'han utilitzat expressions regulars tant per modificar com per eliminar registres així com altres instruccions de *bash*

Es crearan fins a 4 fitxers nous que s'aniran enniant fins el fitxer crear el fitxer *shootings4.csv* que serà l'entrada del següent *script*.



```
psscrrtobias@psscrrtobias: ~/Documents/PRA
psscrrtobias@psscrrtobias:~/Documents/PRA$ ./b1.sh shootings.csv

Inici de l'execució de l'script b1.sh
-----

Final de l'execució de l'script b1.sh.
-----

psscrrtobias@psscrrtobias:~/Documents/PRA$
```

Imatge 2: Sortida de l'script b1.sh. Elaboració pròpia.

2.2.2. Script b.sh

A l'script b.sh es crearà una nova columna, *age_groups* que pretén agrupar a les persones per edat. D'aquesta manera, per no tenir molts grups s'ha decidit fer-ne quatre:

- Menor de 21 anys.
- De 21 a 40 anys.
- De 41 a 60 anys.

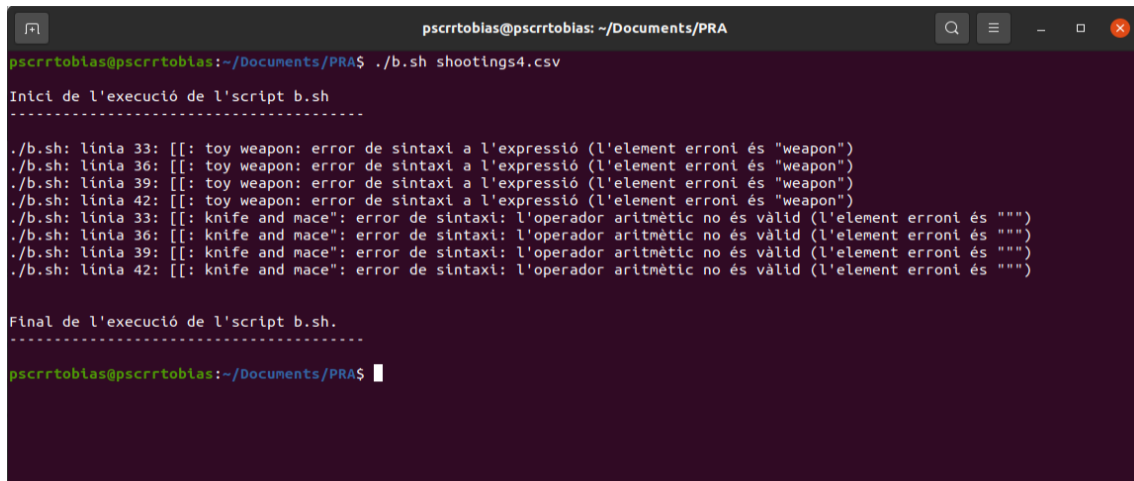
- Majors de 60 anys.

Per la categorització dels registres en grups d'edat s'ha utilitzat una instrucció *while* seguida de instruccions *if* que permetran filtrar els valors adequats d'edats.

Es podria haver fet una divisió més precisa però s'ha decidit fer aquesta per no tenir masses grups i complicar l'anàlisi.

Finalment s'obindrà un fitxer *shootings5.csv* que serà l'entrada del següent *script*.

Altrament, s'eliminaran els fitxers creats anteriorment i que no es necessiten per seguir el procés: *shootings1.csv*, *shootings2.csv* i *shootings3.csv*



```

pscrtrtobias@pscrtrtobias: ~/Documents/PRA
pscrtrtobias@pscrtrtobias:~/Documents/PRA$ ./b.sh shootings4.csv
Inici de l'execució de l'script b.sh
-----
./b.sh: línia 33: [: toy weapon: error de sintaxi a l'expressió (l'element erroni és "weapon")
./b.sh: línia 36: [: toy weapon: error de sintaxi a l'expressió (l'element erroni és "weapon")
./b.sh: línia 39: [: toy weapon: error de sintaxi a l'expressió (l'element erroni és "weapon")
./b.sh: línia 42: [: toy weapon: error de sintaxi a l'expressió (l'element erroni és "weapon")
./b.sh: línia 33: [: knife and mace": error de sintaxi: l'operador aritmètic no és vàlid (l'element erroni és "")
./b.sh: línia 36: [: knife and mace": error de sintaxi: l'operador aritmètic no és vàlid (l'element erroni és "")
./b.sh: línia 39: [: knife and mace": error de sintaxi: l'operador aritmètic no és vàlid (l'element erroni és "")
./b.sh: línia 42: [: knife and mace": error de sintaxi: l'operador aritmètic no és vàlid (l'element erroni és "")
Final de l'execució de l'script b.sh.
-----
pscrtrtobias@pscrtrtobias:~/Documents/PRA$

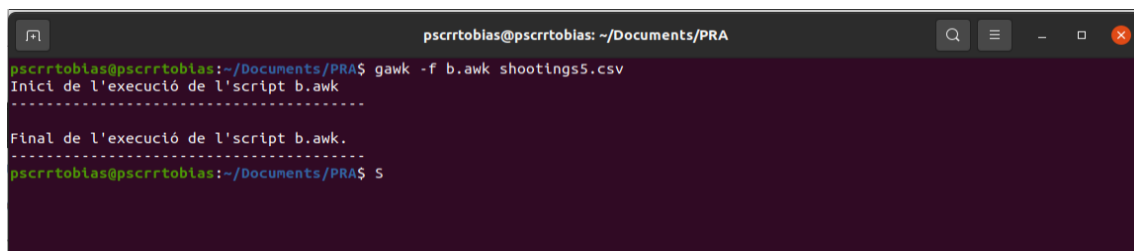
```

Imatge 3: Sortida de l'script b.sh. S'observen 8 errors que no he pogut solucionar però no he vist que afectin al resultat final. Elaboració pròpia.

2.2.3. Script b.awk

A l'*script* b.awk s'ha utilitzat per crear una nova columna anomenada *year* de tipus enter que extreu l'any de la columna *date*. La sortida es el fitxer *shootings6.csv* que es el que s'utilitzarà per fer l'anàlisi final en l'*script* c.sh

L'*script* utilitza una instrucció *split* per poder separar els camp *date* en diferents caps i la instrucció *if* per guardar els valors en una nova columna, registre a registre. Finalment s'utilitza un bucle *for* per generar el camp *llera* en cada registre.



```

pscrtrtobias@pscrtrtobias: ~/Documents/PRA
pscrtrtobias@pscrtrtobias:~/Documents/PRA$ gawk -f b.awk shootings5.csv
Inici de l'execució de l'script b.awk
-----
Final de l'execució de l'script b.awk.
-----
pscrtrtobias@pscrtrtobias:~/Documents/PRA$ S

```

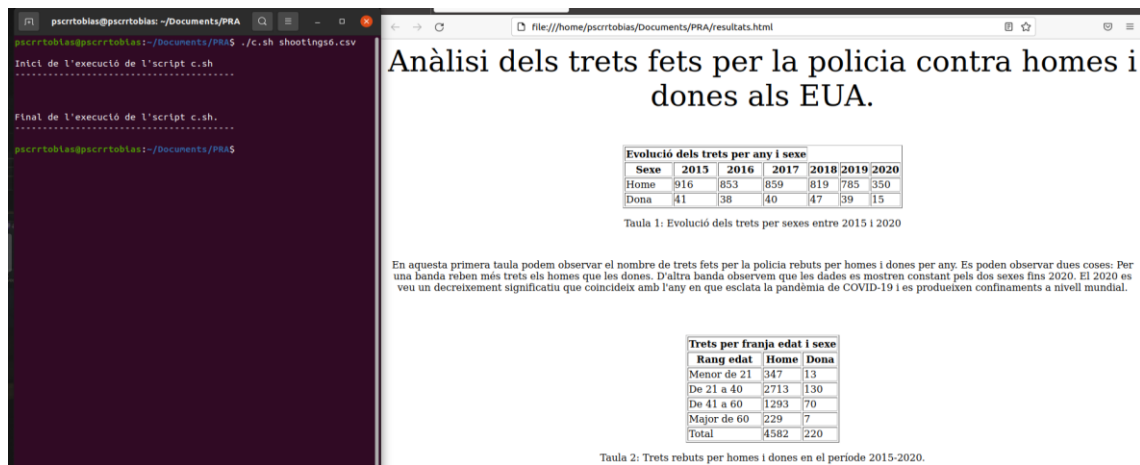
Imatge 4: Sortida de l'script b.awk. Elaboració pròpia.

2.3. Script c.sh

Aquest *script* és l'encarregat de dissenyar l'anàlisi del fitxer HTML. A partir del fitxer `shootings6.csv` obtingut en l'últim *script* es crearan tres taules que mostraran els següents valors:

- Taula 1: Evolució dels trets per sexe entre els anys 2015 i 2020.
- Taula 2: Trets rebuts per homes i dones en el període 2015-2020 per grups d'edats.
- Taula 3: Trets rebuts per sexe i raça en el període 2015-2020. S'inclouen percentatges sobre el total de cada binomi raça-sexe per arribar a una conclusió de manera molt més ràpida.

Bàsicament el codi construeix les taules en llenguatge HTML i a partir de varies estructures *if* calcula dades a nivell agregat per any, sexe o per raça. També es calculen percentatges com en l'END de la tercera taula.



Anàlisi dels trets fets per la policia contra homes i dones als EUA.

Sexe	2015	2016	2017	2018	2019	2020
Home	916	853	859	819	785	350
Dona	41	38	40	47	39	15

Taula 1: Evolució dels trets per sexes entre 2015 i 2020

En aquesta primera taula podem observar el nombre de trets fets per la policia rebuts per homes i dones per any. Es poden observar dues coses: Per una banda ruben més trets els homes que les dones. D'altra banda observem que les dades es mostren constant pels dos sexes fins 2020. El 2020 es veu un decreixement significatiu que coincideix amb l'any en que esclata la pandèmia de COVID-19 i es produeixen confinaments a nivell mundial.

Rang edat	Home	Dona
Menor de 21	347	13
De 21 a 40	2713	130
De 41 a 60	1293	70
Major de 60	229	7
Total	4582	220

Taula 2: Trets rebuts per homes i dones en el període 2015-2020.

Imatge 5: Sortida de l'script c.sh. Elaboració pròpia.

Evolució dels trets per any i sexe						
Sexe	2015	2016	2017	2018	2019	2020
Home	916	853	859	819	785	350
Dona	41	38	40	47	39	15

Taula 1: Evolució dels trets per sexes entre 2015 i 2020

Imatge 6: Anàlisi Taula 1. Elaboració pròpia.

En aquesta primera taula podem observar el nombre d'incidents amb trets registrats per any i sexe. Es veu una evolució bastant similar tant per homes i per dones que es trenca quan arribem a l'any 2020. Això, possiblement, es dona pel fet que esclata la pandèmia de COVID-19 a nivell mundial i es decreten els primers confinaments. Per ampliar informació podríem revisar en quin mes comença a haver un descens del número d'incidents per contrastar que es doni en el període de confinaments als EUA.

El que sobta, encara que era esperat, és que el número d'incidents amb dones es bastant inferior al nombre d'incidents amb homes.

Trets per franja edat i sexe		
Rang edat	Home	Dona
Menor de 21	347	13
De 21 a 40	2713	130
De 41 a 60	1293	70
Major de 60	229	7
Total	4582	220

Taula 2: Trets rebuts per homes i dones en el període 2015-2020.

Imatge 7: Anàlisi Taula 2. Elaboració pròpia.

La següent taula ens mostra el número d'incidents registrats per rang d'edat i per sexe en el període 2015-2020. Es pot veure clarament com la majoria del incidents es troba en la franja d'edats entre el 21 anys i el 60 anys i més concretament entre en la franja d'edat dels 21 als 40 anys. Aquestes dades podrien entrar en la lògica que els incidents estan provocats per gent jove. Un altre exercici interessant seria augmentar el número de grups d'edat per veure en quines franges d'edat hi ha més incidents.

Aquí tornem a veure com el número d'incidents amb dones és molt més menor que el protagonitzat per homes, encara que la franja d'edat amb més incidents torna a repetir-se, de 21 a 40 anys.

Trets per sexe i raça				
Raça	Home	% Home	Dona	% Dona
Asiatic	87	1.90	4	1.54
Negre	1223	26.69	47	18.15
Hispa	852	18.59	25	9.65
Nadiu	71	1.55	5	1.93
Altres	43	0.94	4	1.54
Blanc	2306	50.33	135	52.12
total	4582	100.00	259	100.00

Taula 3: Trets rebuts per sexe i raça en el període 2015-2020.

Imatge 8: Anàlisi Taula 2. Elaboració pròpia.

Per tal d'incloure el component racial s'ha realitzat una taula que agrupa els incidents per sexe però també per raça. S'han llegit infinitats de notícies sobre incidents entre policia estatunidenca i gent de raça negra i esperava torbar un percentatge més elevat d'incidents contra persones d'aquesta raça però el 50% d'incidents son produïts per homes i dones de

raça blanca. S'hauria d'analitzar més en detall i potser amb *datasets* oficials per poder arribar a una conclusió més ferma.

3. Taula de transformacions dels *scripts* b.

A la següent taula es mostren les diferents transformacions generades pels *scripts* tipus b en el joc de dades inicial *shootings.csv*.

SCRIPT	DADES A TRANSFORMAR	LÍNIES	TRANSFORMACIÓ
b1.sh	gender (text)	21-22	Modificar valors F i M per <i>male</i> i <i>female</i> .
b1.sh	age (float)	25	Eliminar els registres on la variable <i>age</i> no tingui el format XX.0
b1.sh	age (enter)	28	Eliminar els caràcters .0
b.sh	age_groups (text)	29-47	Creació de una nova columna que agrupa per edats
b.awk	year (enter)	20-48	Creació d'una nova columna amb l'any del registre a partir de la columna <i>date</i> .

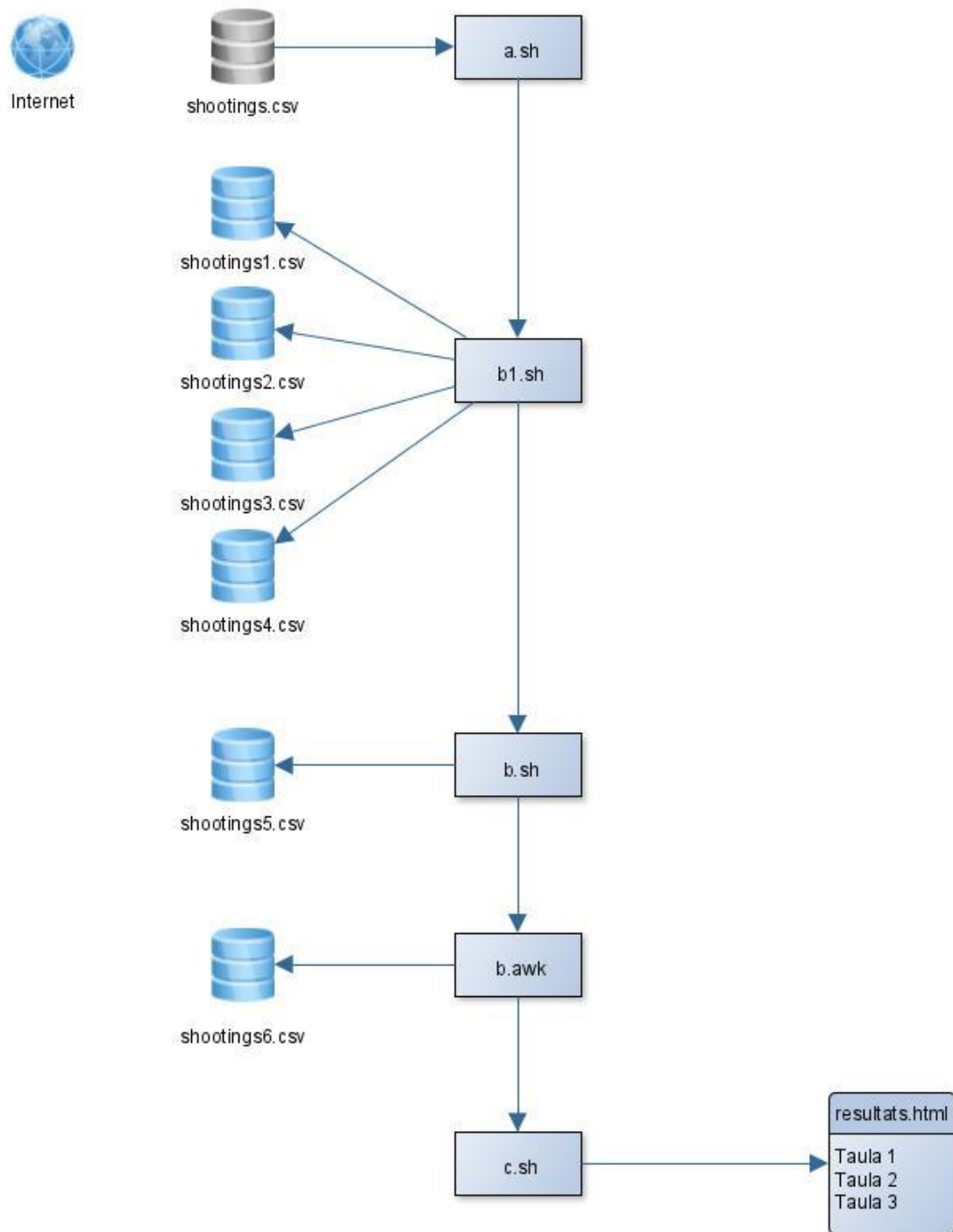
Taula 1: Transformacions generades pels scripts. Elaboració pròpia.

4. Taula d'agrupacions i categorització de l'*script* c.

CAMP	AGRAGACIÓ/CATEGORITZACIÓ	RESULTATS
gender	Suma (agregació)	Nombre de trets rebuts per homes i dones per any.
year		
age_groups	Suma (agregació)	Nombre de trets rebuts per homes i dones per rang d'edat.
gender		
gender	Suma (agregació)	Nombre de trets rebuts per home i dones per raça.
race	Percentatge	Percentatge sobre el total segons sexe.

Taula 2: Taula d'agrupacions. Elaboració pròpia.

5. Diagrama de flux.



Imatge 9: Diagrama de flux. Elaboració pròpia.

6. Conclusions.

Amb aquest projecte s'ha pogut posar en pràctica molts dels conceptes apresos al llarg del curs. També s'ha pogut observar la potencia de l'ús de l'*scripting* per crear solucions a pràctiques que poden ser molt tedioses i provocar errors si es fan manualment, per exemple, mitjançant l'ús de fulls de càlcul.

Hem observat com es pot fer ús d'altres llenguatges de programació com l'HTML per fer petites presentacions. Recordem que hem estat veient els resultats sempre per la pantalla de la terminal i l'ús de la codificació amb HTML provoca que la sortida sigui més amigable per l'usuari que no està acostumat a la programació.

La pràctica ens ha donat l'opció de construir un petit projecte per l'anàlisi de dades amb *bash* i *awk* que fins l'inici del curs no coneixia de la seva existència. Actualment és fàcil llegir articles o veure vídeos sobre llenguatges més àmpliament utilitzats en la ciència de dades com poden ser Python, SQL o R. però no tant d'aquest dos llenguatges emprats durant el curs.

Com a aspectes de millora s'haurien de corregir els 8 errors que mostra l'*script* b.sh. Es podrien fer més transformacions com per exemple extreure el mes del camp *date* per poder fer un estudi més profund sobre els incidents en època COVID-19. Un altre exercici interessant seria mostrar gràfiques que permetessin més claredat a l'hora d'analitzar les dades.

Com a conclusió l'exercici ha servit per ordenar tots els conceptes estudiats durant el curs, veure la potencia de l'*scripting* i veure una aplicació pràctica senzilla que un cop acabat el curs podem millorar.

7. Bibliografia

US Police Shootings: <https://www.kaggle.com/datasets/ahsen1330/us-police-shootings>

CSVSTAT: <https://csvkit.readthedocs.io/en/1.0.2/scripts/csvstat.html>

Extract filename and extension in Bash: <https://stackoverflow.com/questions/965053/extract-filename-and-extension-in-bash>

Make ``rm`` move to trash: <https://unix.stackexchange.com/questions/42757/make-rm-move-to-trash>

How to Compare Numbers in a Bash Script: <https://linuxways.net/centos/compare-numbers-bash-script/>

Guillem Lluch Moll (2020) *Expresions Regulars*

Guillem Lluch Moll (2020) *Introducció a AWK*

Gerard Farràs Ballabriga (2020) Cas Pràctic: *Manipulació de fitxers de text*