Projecte R

Àngel Pérez Beumala
Carles Cano Casablanca
Oriol Campderròs Arís
Cristian Torres Barrantes

Master of Cybersecurity Management
Universitat Politècnica de Catalunya

Desembre 2016

Índex

1. Definició de la pregunta

Si bé els primers ciberatacs que es van produir tenien una finalitat lúdica i de satisfacció personal, ràpidament han evolucionat cercant altres propòsits molt més preocupants que afecten tant l'administració pública, les empreses i els ciutadans en general.

En el 2015, el nombre de 0-days descoberts i explotats va incrementar un 125 % més que l'any anterior. En aquest mateix any es van produir una mitja d'un milió d'atacs a diari degut a què s'estima que el 75 % de les pàgines web tenen vulnerabilitats no fixades.

Dins d'aquest context, ens trobem en la situació en la qual els equips de resposta d'incidents (CERTS) han d'anar darrera dels ciberatacs un cop ja han succeït; sense saber quan i on serà el següent. Per tant, considerem que seria interessant proposar un estudi que pogués aportar informació als CERTS sobre un potencial atac a un país.

Per fer-ho, és necessari estudiar les dades registrades dels ciberatacs a nivell mundial dels últims anys i intentar buscar una relació que ens permeti ubicar el pròxim esdeveniment en una zona i un interval de temps concret.

Degut a què un alt percentatge dels atacs estan lligats a fins econòmics, la nostra atenció l'enfocarem a saber si hi ha una relació entre els països més atacats amb el seu nivell de riquesa. Tot i així, aquest estudi ens portaria a saber el *target* però encara tindríem una finestra de 365 dies per esbrinar quan es podria produir.

En aquest punt, introduïm una hipòtesis que participarà com a variable en la nostra investigació: creiem que els dies més apropiats per realitzar un atac són en els quals els països objectiu celebren un dia festiu a nivell nacional; ja que és probable que els organismes i empreses no comptin amb tot el personal de seguretat com en un dia ordinari.

Per tant, suposant que els recursos destinats a protegir els actius en dies festius es veuen reduïts, relacionarem aquest fet amb el nivell de riquesa del país. En conseqüència, la pregunta resultant que ens plantegem en aquest treball és la següent:

Són els països amb un nivell de riquesa més elevat atacats en dies estratègics?

2. Definició i obtenció de les dades

En funció de la pregunta plantejada, contemplem buscar la relació entre les següents dades: ciberatacs, nivell de riquesa d'un país i dies festius d'aquest.

Ciberatacs

Pel que fa als ciberatacs ens interessa recopilar informació que registri quan es va produir l'atac, qui el va produir i el país objetiu. No obstant, hem hagut de descartar obtenir l'autor dels atacs ja que una gran quantitat dels registres no poden definir amb exactitud la procedència dels mateixos.

Les dades han estat extretes d'una reconeguda font[1] que registra els ciberatacs a nivell mundial des de fa anys.

PPP

Quan volem comparar la riquesa entre diversos països el primer que se'ns pot ocórrer és comparar els seus productes interiors bruts. A major PIB, més riquesa en un país. No obstant això, comparar simplement el PIB realment no diu molt sobre l'economia d'un país, ja que no és el mateix tenir un país de cent mil habitants que un de mil milions. Per tant, s'ha d'utilitzar un altre indicador: el PPP (*Purchasing Power Parity* o Paritat de poder adquisitiu), el qual registra el PIB per càpita.

Aquesta informació ha estat obtinguda del WorldBank [2], on hi figura el PPP de tots els països dels últims anys.

Dies festius

En els dies festius d'un país el més normal és que els equips de seguretat, encarregats de monitoritzar els events dels servidors, es vegin reduïts i hi hagi poc personal de guàrdia. Aquest pot ser un bon al·licient perquè els atacants duguin a terme els seus plans. Per altra banda, també poden aprofitar conéixer els dies festius per realitzar atacs de denegació de servei dies abans (p. ex. atacar Amazon una setmana abans de Nadal).

[1] http://www.hackmageddon.com
[2] http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD

Les dades dels dies festius han estat extretes de la pàgina *Officeholidays* [3].

3. Dades elegants

Les dades obtingudes de les fonts descrites a l'apartat anterior tenen un format que requereix d'un processat a mida; ja que hi ha valors buits o no vàlids.

A continuació, es mostraran evidències sobre l'estat inicial dels diferents tipus de dades i els *data frames* finals obtinguts (dades elegants) després d'aplicar un conjunt de transformacions. Aquests seran amb els quals es treballarà en les següents seccions.

Ciberatacs

En la següent figura es veuen les dades dels ciberatacs produïts a nivell mundial. Com es pot apreciar, hi ha cel·les que tenen valors que no ens serveixen per treballar amb elles. Per tant, cal triar les files i columnes adients fent servir diferents llibreries i expressions regulars.

ID	Date	Author	Target	Description	Attack	Target Class	Attack Class	Country	Link	Tags
1	16/10/2016	?	Road Signs	A number of people at the Chicago's Grand Avenue and	Unknown	Road Signs	CC	US	https://www.hackread.com/construction-sign-hacked-	Chicago, Rahm Emanuel
2	17/10/2016	Guccifer 2.0	Democratic National Committee	Guccifer 2.0 is back and leaks new fresh documents relating	Unknown	Org: Political Party	CC	US	http://www.ibtimes.co.uk/hacker-guccifer-2	Guccifer 2.0, Democratic National
3	17/10/2016	SCUWatch	University of Santa Clara Office of	A hacker dubbed SCUWatch leaks a trove of internal	Unknown	Education	CC	US	https://www.databreaches.net/omc-targeted-in-	SCUWatch, University of Santa Clara,
4	18/10/2016	Revolver AKA 1x0123 Peace	AdultFriend Finder	A hacker known as Revolver or 1x0123 claims to have	Local File Inclusion	Adult Site	CC	US	http://motherboard.vice.com/read/hookup-service-adult-	Revolver, 1x0123, Peace,
5	18/10/2016	?	RedBus	Online travel giant Ibibo Group-owned	Unknown	Industry: Tourism	CC	IN	http://www.business-standard.com/	Redbus, Ibibo Group
6	19/10/2016	?	Axis Bank	Axis Bank, India's third-largest private bank, announces	Account Hijacking	Finance	CC	IN	http://news.softpedia.com/news/hacker-infiltrates-the-	Axis Bank
10	20/10/2016	APT28	Several Targets	The cyber gang called Sednit, also known by the names Fancy Bear, APT28,	Targeted Attack	>1	CE	>1	http://www.welivesecurity.com/2016/10/20/new-eset-research-	Fancy Bear, APT28, Pawn Storm, Sofacy
7	20/10/2016	Unknown Criminals from China?	Several Top Indian Banks	Details of more than 3.2 million cash cards of customers	Malware	Finance	CC	IN	http://economictimes.indiatimes.com/industry/banking/fina	India, Visa, Mastercard, RuPay

FIGURA 1: Atacs realitzats a nivell mundial

[3] http://www.officeholidays.com/countries/country/year.php

Finalment, el resultat net final té els camps estrictament necessaris amb valors correctes, tal i com es pot apreciar en la següent imatge:

	Date	Attack	Category	Country
1	2016-05-06	Unknown	CC	AE
2	2016-05-15	DDoS	H	AE
3	2016-05-22	Unknown	CC	AE
4	2016-05-29	Targeted Attack	CE	AE
5	2016-07-30	Account Hijacking	CC	AF
6	2016-09-01	Defacement	H	AF
7	2016-09-23	Unknown	CW	AF
8	2016-05-12	SQLi	CC	AL
9	2016-06-17	SQLi	CC	AL
10	2016-04-01	Defacement	CW	AM
11	2016-04-07	Account Hijacking	H	AM
12	2016-09-02	Unknown	H	AM
13	2016-03-29	Defacement	H	AO
14	2016-05-22	SQLi	CC	AR
15	2016-01-19	Account Hijacking	CC	AT

FIGURA 2: Data frame resultant després del processat

PPP

En les dades obtingudes del .csv que proporciona WorldBank, es pot apreciar com hi ha columnes sense valors així com també columnes que no ens interessa tractar (p.ex: les columnes d'anys anteriors a 2014).

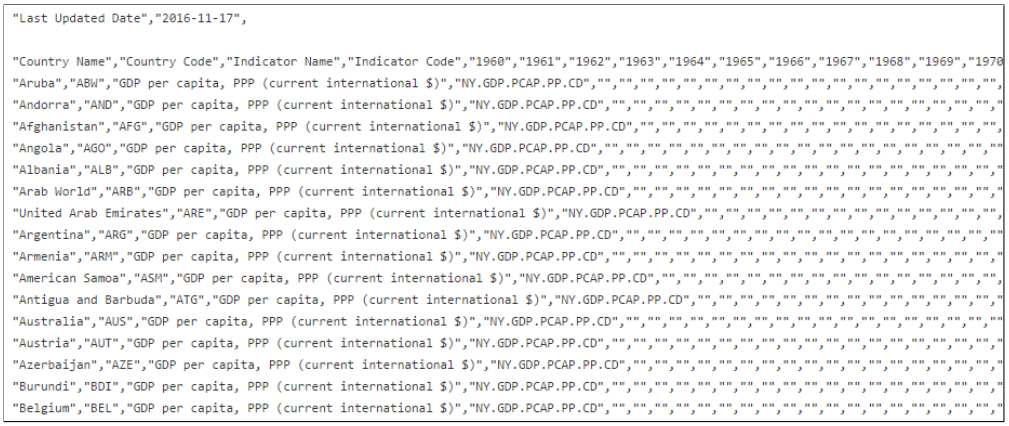

FIGURA 3: Dades sobre l'economia de tots els països

Un cop extretes les dades d'interès, el *data frame* obtingut que ens servirà per relacionar-lo amb les altres dades té el següent aspecte:

	Country	Freq	Country.Name	iso3c	PIB	Region	IncomeGroup
81	QA	1	Qatar	QAT	143788.242	Middle East & North Africa	High income
61	LU	2	Luxembourg	LUX	101926.424	Europe & Central Asia	High income
87	SG	5	Singapore	SGP	85208.811	East Asia & Pacific	High income
56	KW	2	Kuwait	KWT	71311.994	Middle East & North Africa	High income
1	AE	4	United Arab Emirates	ARE	70237.948	Middle East & North Africa	High income
72	NO	6	Norway	NOR	61471.574	Europe & Central Asia	High income
19	CH	8	Switzerland	CHE	60535.159	Europe & Central Asia	High income
42	HK	7	Hong Kong SAR, China	HKG	56719.498	East Asia & Pacific	High income
98	US	528	United States	USA	55836.793	North America	High income
45	IE	8	Ireland	IRL	54654.397	Europe & Central Asia	High income
85	SA	7	Saudi Arabia	SAU	53430.045	Middle East & North Africa	High income
71	NL	11	Netherlands	NLD	48458.940	Europe & Central Asia	High income
7	AT	3	Austria	AUT	47824.188	Europe & Central Asia	High income

FIGURA 4: Data frame resultant després del processat

Dies festius

Finalment, l'últim grup de dades a tractar són els dies festius de tots els països. A diferència dels casos anteriors, on les dades s'extreien d'excels o fitxers .csv i es convertien en *data frames*, els festius es troben en diferents taules HTML dins d'una pàgina web i s'han de representar en una llista degut a què cada país té un número diferent. En la següent figura es poden apreciar les dades en brut:

List of National Public holidays of Spain in 2016

Day	Date	Holiday	Comments
Friday	January 01	New Years Day	
Wednesday	January 06	Epiphany	Technically not a national holiday, but has been declared a regional holiday in all regions
Monday	February 29	Andalucía Day	Andalucía only. Became an autonomous community of Spain on 28 February 1980
Tuesday	March 01	Balearic Islands	Marks the Statute of Autonomy of March 1, 1983
Saturday	March 19	Father's Day	Celebrated on St. Joseph' s day
Saturday	March 19	St Josephs Day	Murcia, Valenciana only
Thursday	March 24	Maundy Thursday	except Catalonia
Friday	March 25	Good Friday	Friday before Easter Sunday
Monday	March 28	Easter Monday	Balearic Islands, Basque Country, Catalonia, La Rioja, Navarra, Valenciana
Monday	April 04	Feast of San Vincent Ferrer	Andalucía, Valenciana

FIGURA 5: Dades sobre els dies festius de tots els països

Tot i així, no suposa una gran variació en el procés d'extracció i posterior tractament de les dades. En la següent captura es mostren les dades filtrades en format *List*:

```
CL: Date[1:28], format: "2016-01-01" "2016-03-25" "2016-05-01" "2016-05-21"
CN: Date[1:24], format: "2016-01-01" "2016-02-07" "2016-02-08" "2016-02-09"
CO: Date[1:36], format: "2016-01-01" "2016-01-11" "2016-03-21" "2016-03-24"
CR: Date[1:22], format: "2016-01-01" "2016-03-24" "2016-03-25" "2016-04-11"
CY: Date[1:28], format: "2016-01-01" "2016-01-06" "2016-03-14" "2016-03-25"
CZ: Date[1:25], format: "2016-01-01" "2016-03-25" "2016-03-28" "2016-05-01"
DE: Date[1:31], format: "2016-01-01" "2016-01-06" "2016-03-25" "2016-03-28"
DK: Date[1:22], format: "2016-01-01" "2016-03-24" "2016-03-25" "2016-03-28"
DO: Date[1:26], format: "2016-01-01" "2016-01-04" "2016-01-21" "2016-01-25"
EC: Date[1:27], format: "2016-01-01" "2016-02-08" "2016-02-09" "2016-03-25"
EG: Date[1:36], format: "2016-01-07" "2016-01-25" "2016-04-25" "2016-05-01"
ES: Date[1:84], format: "2016-01-01" "2016-01-06" "2016-02-29" "2016-03-01"
```

FIGURA 6: Llista resultant després del processat

4. Interpretació dels resultats

Abans d'unir tota la informació per respondre directament la pregunta, analitzarem resultats intermitjos del processat de les dades per saber el context que l'envolta i poder obtenir unes millors conclusions.

En primer lloc, estudiarem les dades extretes del WorldBank. En concret, podem començar per agrupar els diferents països segons el seu nivell d'ingresos i veure quina mitja de PIB per càpita obtenim. Tal i com es pot apreciar en la figura que es mostra a continuació, hi ha quatre tipus d'economia: *Lower income*, *Lower middle income*, *Upper midle income* i *High income*.

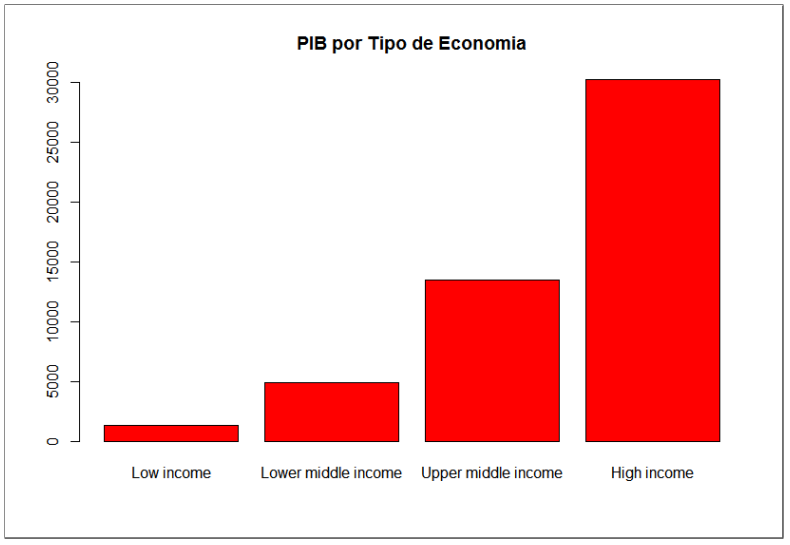

FIGURA 7: Tipus d'economies amb el seu PIB per càpita mig

Amb aquestes dades podem observar el PIB per càpita mig a partir del qual ens centrarem per seleccionar els països posteriorment.

Per altra banda, havent vist els diferents tipus d'economies i els països als quals engloben, ens interessa saber el nombre d'atacs que es produeixen a aquests països. Per aquest motiu, hem generat el següent gràfic; el qual mostra la relació dels atacs amb el PIB per càpita de cada país atacat.

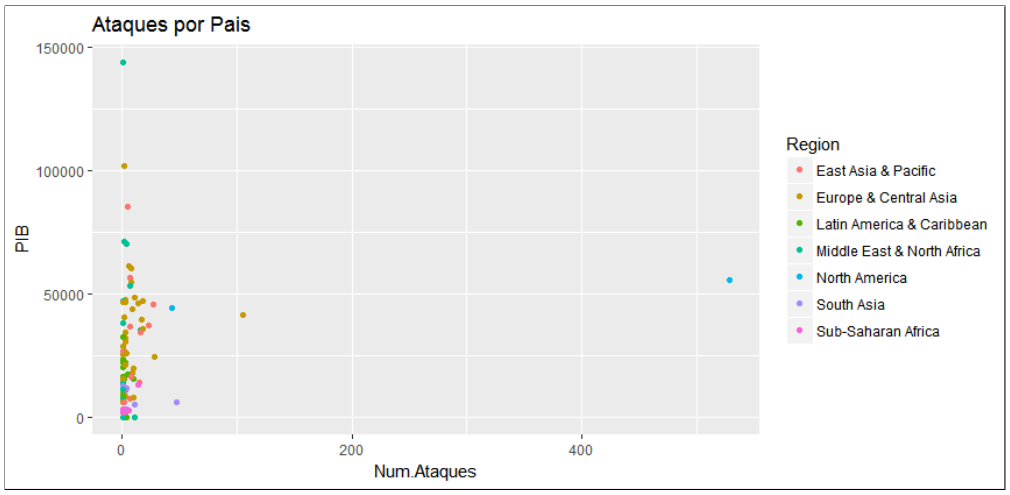

FIGURA 8: Número d'atacs per país segons el PIB per càpita

Analitzant aquestes dades podem extreure diverses conclusions. En primer lloc, podem veure com Estats Units (l'*outlier* horitzontal) és el país més atacat amb gran diferència respecte els altres països. És curiós apreciar com hi ha altres potències, com Qatar i Luxemburg, que tot i tenir un PIB per càpita clarament més elevat que EEUU no reben ni deu vegades menys atacs.

Per altra banda, es pot observar com no hi ha una gran, o com a mínim clara, distinció entre els diferents tipus d'economia. És a dir, regions amb un PIB per càpita al voltant de 50.000 tenen el mateix número d'atacs que altres considerades de baixos ingresos (*Lower income*).

Per poder seguir analitzant aquesta situació, cal representar aquesta gràfica d'una manera que ens faciliti la comprensió de les dades. Degut a aspectes de disseny, no hem pogut mostrar el nom de cada país a sobre de cada punt de la gràfica. A més, fer-ho per tonalitats de colors i referenciar-los amb una llegenda no solucionava el problema, més bé tot el contrari. Per tant, una millor representació visual de les dades la podem apreciar en el següent mapa mundi:

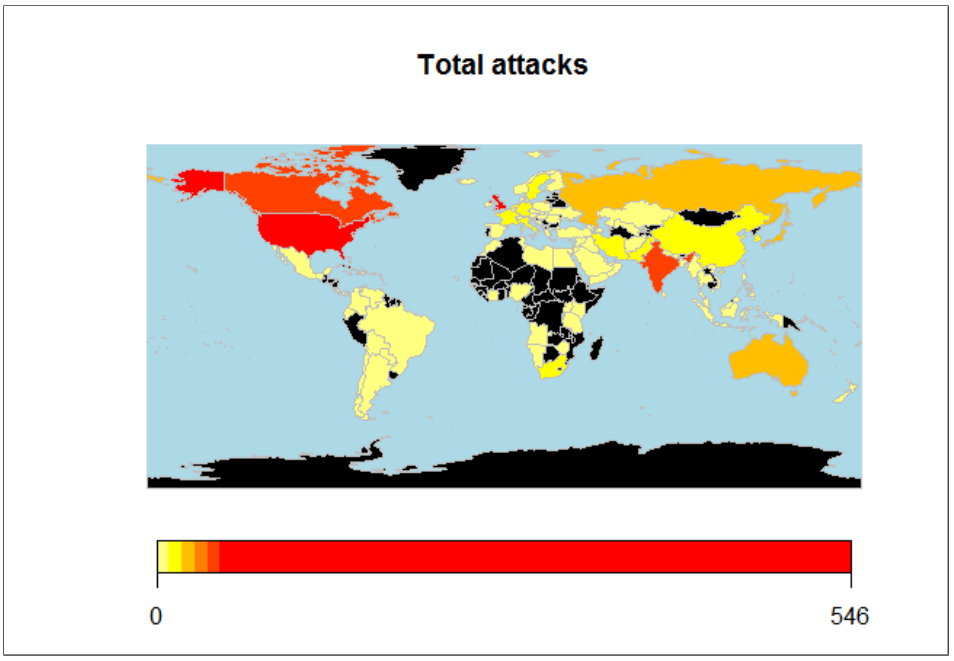

FIGURA 9: Mapa del número d'atacs per país segons el PIB per càpita

Ara ja podem distingir a primera vista els països que des de 2014 han reportat més atacats. Això vol dir que, amb les dades actuals que tenim, podem veure clarament que països com EEUU, Gran Bretanya, Canadà i Índia han rebut la majoria d'aquests atacs. No obstant, també hem d'assumir que hi ha incidents els quals no s'han reportat i que en conseqüència podria fer que aquesta distribució de la figura anterior variés.

Per observar millor la situació dels altres països, farem *zoom* a la gràfica de punts anterior sense tenir en compte EEUU ni GB, els quals distorsionen el resultat i ens impedeixen realitzar un millor anàlisi.

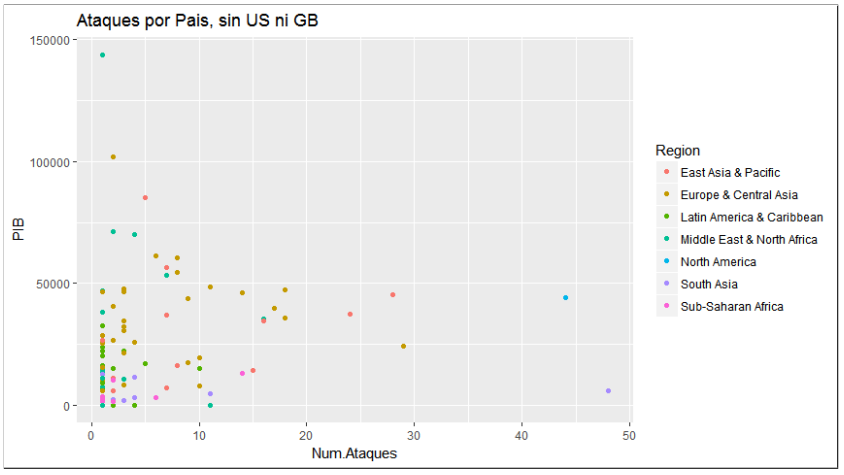

FIGURA 10: Número d'atacs per país segons el PIB per càpita (II)

Tal i com s'aprecia en la **Figura 10**, els països amb un major PIB per càpita (categoritzats anteriorment com *High income*) es mostren dispersos entre ells. És a dir, no s'observa una tendència que indiqui que aquests països oscil·len entre un rang concret d'atacs i que per tant es diferencien dels que tenen menor poder adquisitiu.

En aquest punt, podem començar a pensar si relacionar el PIB per càpita amb el número d'atacs no ha estat el millor indicador per aconseguir fer prediccions sobre futurs atacs. Tot i així, cal seguir analitzant la resta de dades per extreure resultats concloents.

Un altre aspecte a considerar, descrit en la definició de la pregunta, són els dies festius del país atacat. La hipòtesi es basa en una intuïció que guarda relació amb el que ocorre en els dies festius d'un país: el personal de seguretat i mitigació d'incidents no és al complet. Amb aquest raonament, tot fa pensar que aquesta variable pot ser un bon indicador per ubicar en el temps els moments més propensos de què es produeixi un atac.

Fent un còmput global de les dades dels festius extretes anterioment, en la següent gràfica podem veure el percentatge de dies festius en els quals es van realitzar atacs.

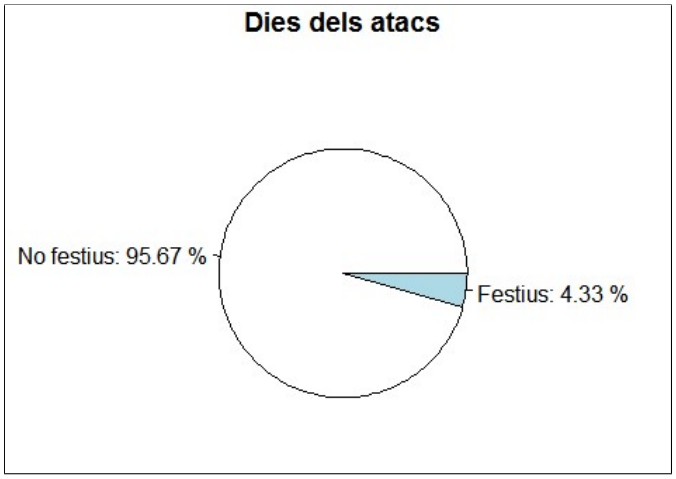

FIGURA 11: Atacs segons dies festius i no festius

Veient aquesta gràfica podem pensar que la hipòtesi sobre els dies festius que vam prendre com a partida inicial no és un aspecte que els atacants tenen en compte a l'hora d'assetjar un país. Podria ser que sí es produís però a l'inversa, és a dir, els països que realitzen els atacs ho fan quan és un dia festiu en el seu país. No obstant, degut a què en molts casos no ha estat possible registrar l'origen d'un atac, hem decidit no optar per aquesta hipòtesi.

En la següent secció recollirem les conclusions extretes de les diferents gràfiques mostrades al llarg d'aquest projecte per determinar si països amb un PIB per càpita elevat són l'objectiu dels atacs i específicament en dies estratègics.

5. Resposta a la pregunta

Tal i com hem anat veient en els resultats intermitjos anteriors, en els quals la relació entre el PIB d'un país, el número d'atacs que rep i els dies festius no guarden la relació esperada, ja podem respondre a la pregunta plantejada.

Tot sembla indicar que no hi ha una relació directa entre les tres variables tractades. Prenent com exemple EEUU, es pot apreciar com és un tipus d'economia *High income* que rep molts atacs, en concret 528, però que d'aquests només 17 cauen en festius. És a dir, un percentatge gens significatiu.

En aquest exemple, la relació del PIB per càpita amb els dies festius no té rellevància. Si agaféssim altres països veuríem que la tèndencia és similar.

Per tant, hem escollit malament la pregunta? Creiem que no és el cas ja que el raonament individual de cadascuna de les parts té un fonament lògic i realístic, però com hem pogut comprobar no guarden una relació entre elles.

Els resultats extrets poden ser útils de cara a futurs treballs, ja que les variables treballades en aquest projecte poden ser descartades i es pot centrar l'atenció en buscar altres per trobar la relació i poder predir quan i on es produirà el següent atac.

Com a reflexió final i en vista dels resultats de les gràfiques, pensem que potser seria interessant, de cara a un futur estudi, plantejar si en comptes de mirar si les economies amb més poder adquisitiu són les més atacades, intentar relacionar-ho amb el nivell de desenvolupament tecnològic, és a dir, els països punters en I+D, ja que potser una de les intencionalitats dels atacants és retrassar tot el procés de creació i innovació de la competència i guanyar terreny en aquest sentit. Això podria explicar, per exemple, per què EEUU és molt més atacat que altres països amb major riquesa però que alhora no són punters en I+D.