



A Digital Observatory of Vehicle Theft in Chile

Denis Parra
Pontificia Universidad Católica de Chile

SoAPS Workshop Gremoble, March 26th 2018

Research Team

- Richard Weber (Project Director)
Industrial Engineering, U. of Chile



- Barbara Poblete
Computer Science, U. of Chile



- Angel Jimenez
Industrial Engineering, U. of Chile



- Denis Parra
Computer Science, PUC Chile



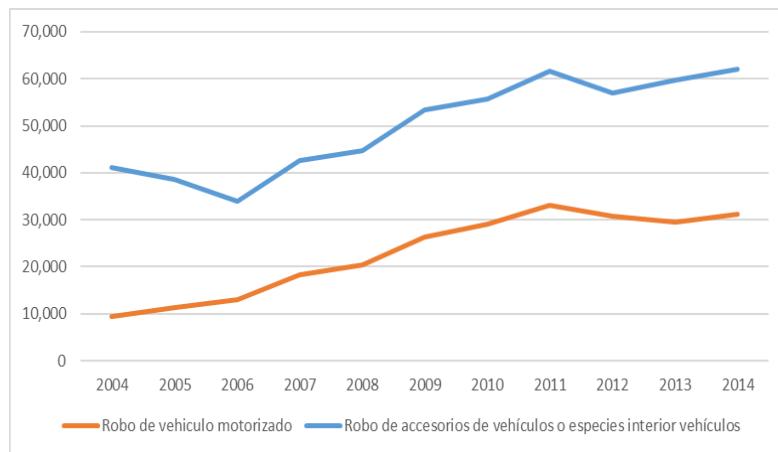
Problem

- Theft of vehicles and accessories is a problem which affects Chile and the rest of the world.
- In the latest years, there has been an increase in car theft with violence in Chile.



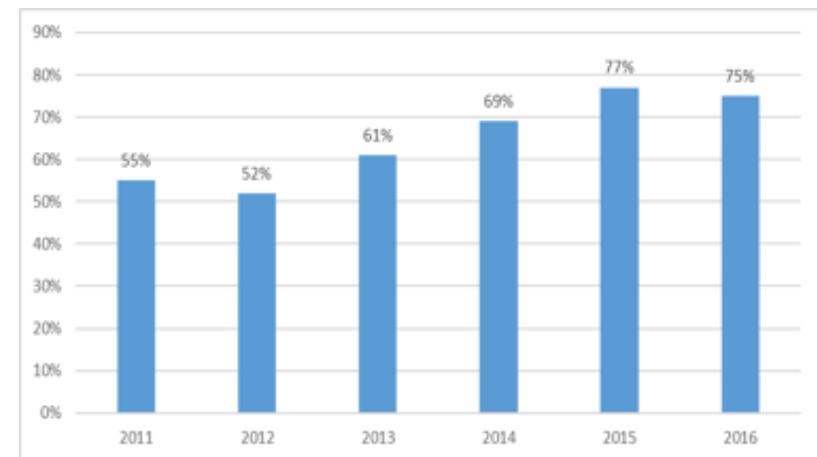
Context

- Chile's vehicle fleet is still increasing, but the same is happening with car-related crimes



Fuente: Informe anual Carabineros, 2004-2014. INE.

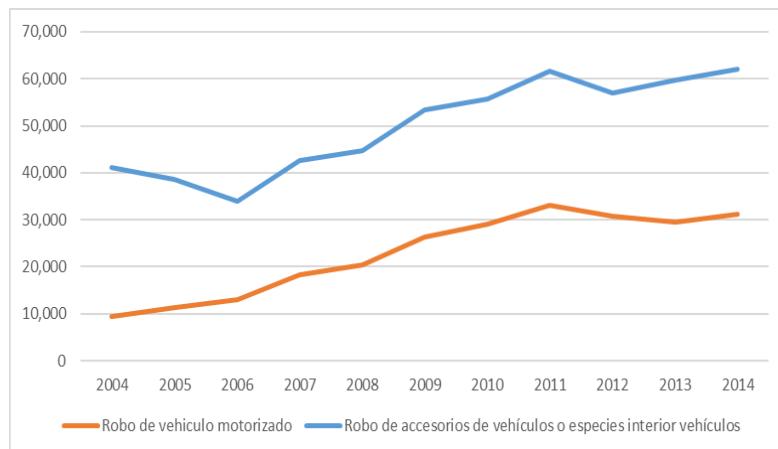
Cantidad de robos de vehículos y robos de accesorios de vehículos anuales en Chile (2004- 2014)



Tasa de robos con violencia del total de robo de autos de lujo (2011- 2016)

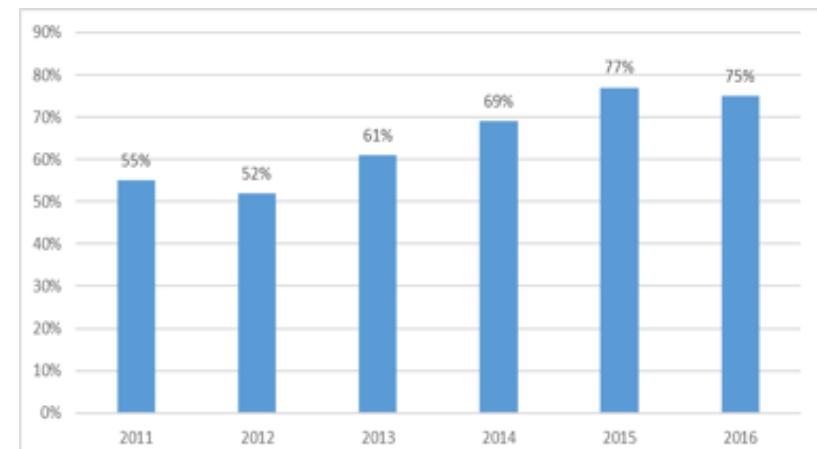
Context

- Chile's vehicle fleet is still increasing, but the same is happening with car-related crimes



Fuente: Informe anual Carabineros, 2004-2014. INE.

Cantidad de robos de vehículos y robos de accesorios de vehículos anuales en Chile (2004- 2014)



Tasa de robos con violencia del total de robo de autos de lujo (2011- 2016)

Average cost of a car insured and never recovered:
CLP\$5.000.000. (~ US\$ 8,000)



Total loss in 2015 in non-recovered cars:
CLP\$15.720 millones, (~ US\$ 25,309,200)

Statistics from other countries

- In 2011, 65,000 vehicles were stolen and never recovered in the United Kingdom. The average value of a stolen car was 4,600 British pounds (€ 5,474), totaling € 355 million annually.
- Germany 2011: total economic damage on stolen motor vehicles was almost 260 million Euros.
- Australia: the community is set back over 610 million Australian Dollars (€ 422 million) per year in higher insurance premiums and demands on the justice system.

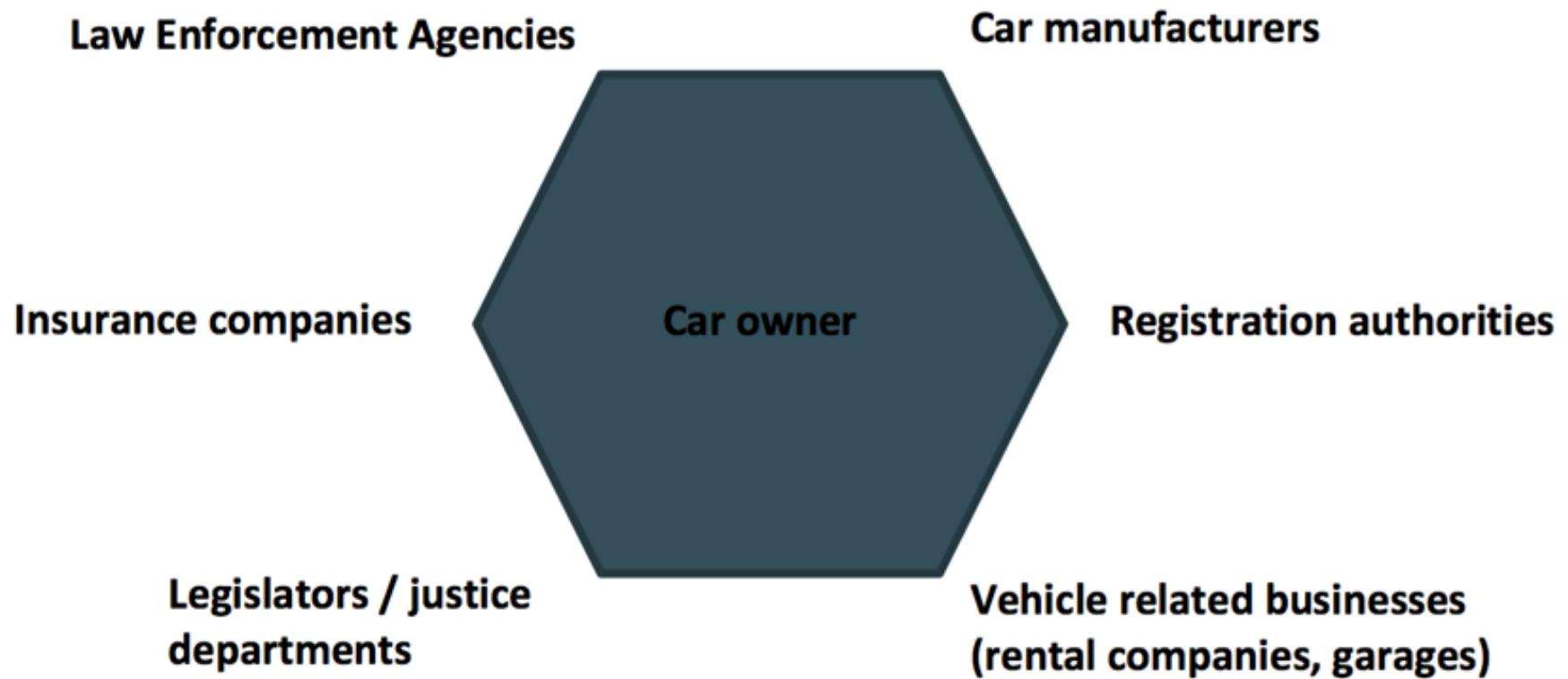
Car theft and other crimes

- We might think that car theft affects only the people who suffers a robbery
- However, an Interpol report* identified car vehicles crimes with other more serious crimes:
 - Terrorism
 - Trafficking of goods
 - Trafficking of humans

* <https://www.interpol.int/Crime-areas/Vehicle-crime/Vehicle-crime>

Stakeholders

- From Interpol report:



Opportunity

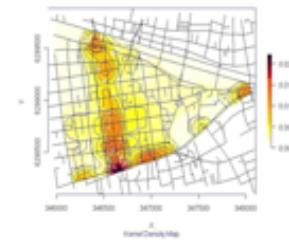
- There are new technologies and data sources which could help in understanding and preventing this type of crime.



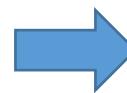
Social
networks



Data/text
Mining



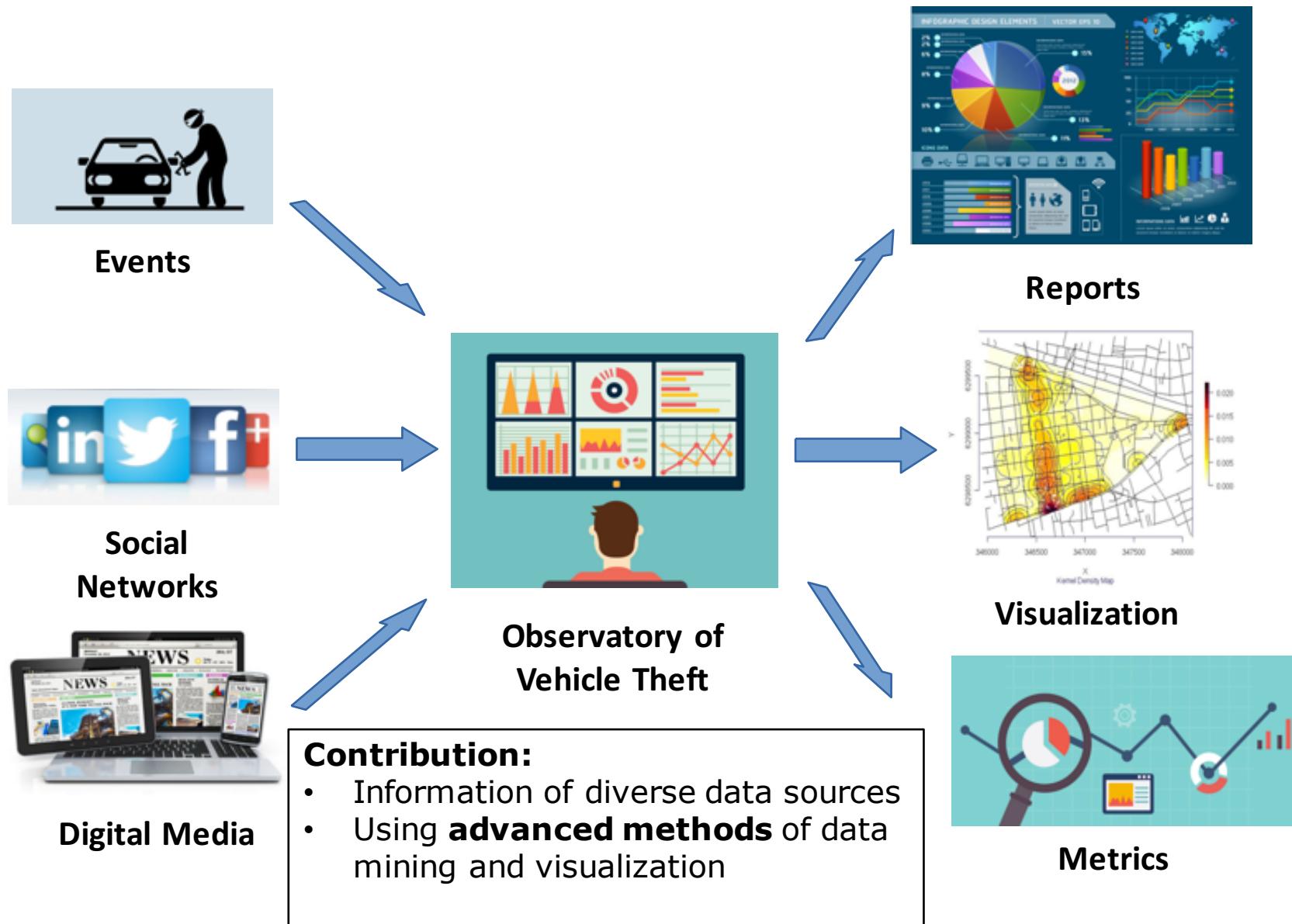
Visualization



Potential for
Technology transfer

World average: 68,5 vehicles
stolen per 100,000 inhabitants

Our Solution: Digital Observatory



Main Objective

- Help to achieve a 5% reduction of Vehicle Theft in Chile

Hypotheses and Project Goals

Hypotheses:

- **H1:** Theft of vehicles and accessories **is not random**. It follows **patterns** which are detectable through data mining techniques.
- **H2:** Methods of data mining and visualization reveal effectively the patterns, improving user experience.

General Goal:

To develop a digital observatory of vehicle and accessory theft in Chile.

Specific Goals:

- To characterise *modi operandi* of criminals.
- To detect theft patterns.
- To develop a visualization tool theft patterns and relevant indicators for the users.

Actors in this Project

Users:

- **AACh**: A group of 11 insurance companies which share information helpful for the whole car insurance industry.

Beneficiaries:

- **Insurance Companies**: The direct beneficiaries are car insurance companies which will see a decrease in vehicle, hence, decreases in loss revenue.
- **Insured People**: People who insure their cars will experience a decrease in their insurance prime (cost of insurance due to previous events).
- **Chile's Police (Carabineros)**: The information of the digital observatory could help police to develop effective and efficient plans for monitoring and preventing crime.
- **Society**: A decrease in vehicle theft will help society in reducing all the related crimes, and hence will also decrease the perception of insecurity.

Methodology

Data Sources



Events



Social Networks



Digital Media

Preprocessing

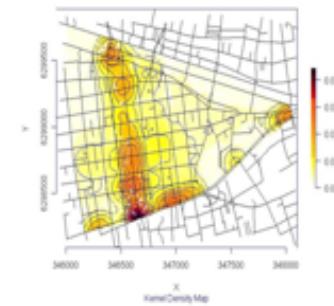


- ❑ Data cleaning
 - ❑ Transforming variables
 - ❑ *Feature selection and dimensionality reduction*
 - ❑ Data enrichment

Text Mining



Pattern Recognition & Visualization



Observatory



Insurance Companies' Data

- Data from 11 car insurance companies
- The cleanest among all data sources, but only from last year.
- Several fields (74) but large amounts of missing data

Nombre	Tipo de dato	Filas no null	Descripción
id_prose	ordinal	55626	ID única transacción.
sin_patente	nominal	55626	Patente vehículo siniestrado.
sin_fecha_siniestro	fecha	55626	Fecha (día, mes y año) en donde ocurrió el siniestro de robo.
ase_abreviatura	nominal	55626	Abreviatura de aseguradora en la cual se registra el robo.
obs_ult_estado	nominal	3179	Validación interna.

...

sin_relato	nominal	51451	Relato del siniestro.
------------	---------	-------	-----------------------

Insurance Companies' Data (2)

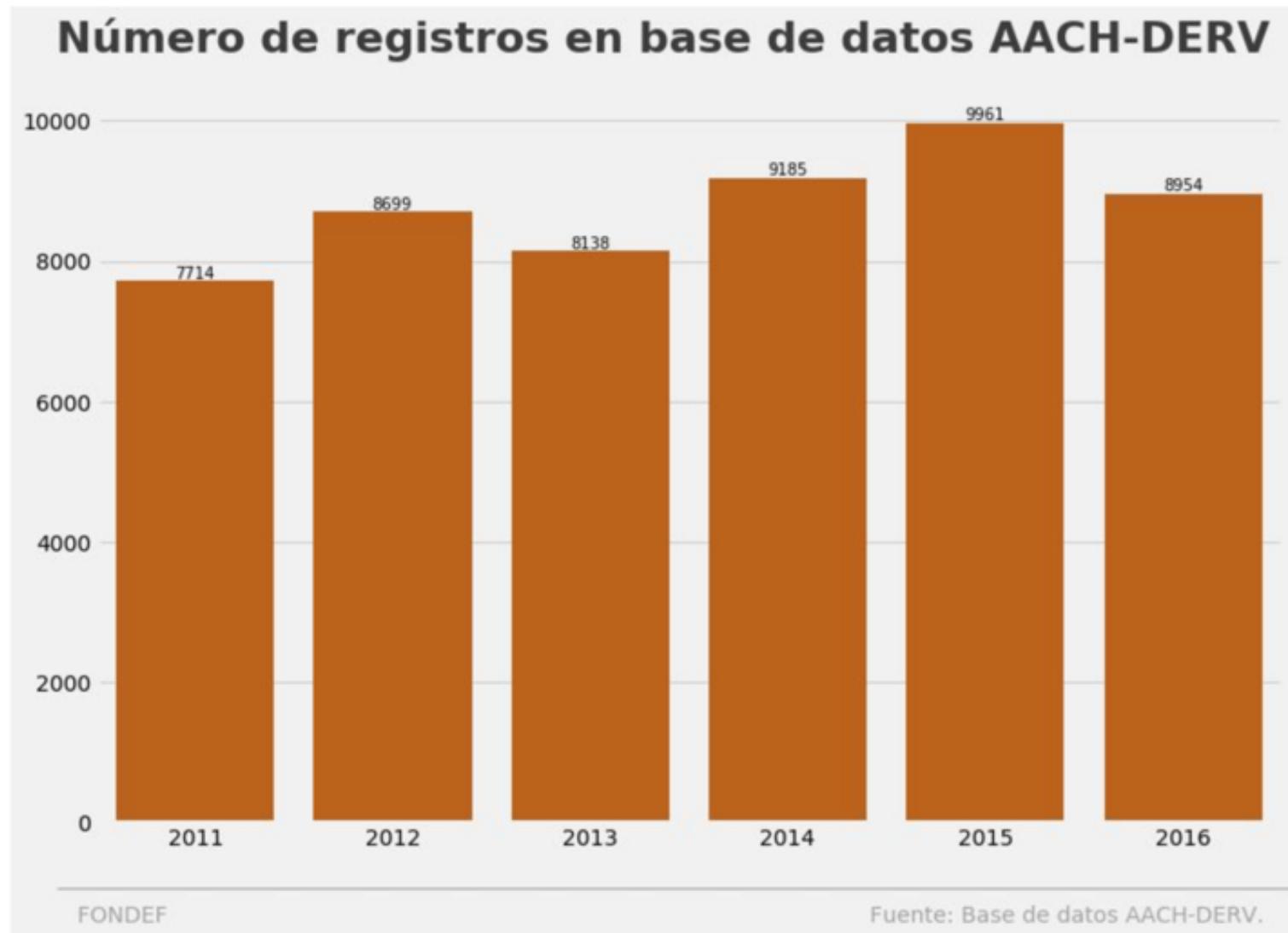


Figura 2. Cantidad de robos registrados por año en base de datos AACH-DERV

Insurance Companies' Data (3)

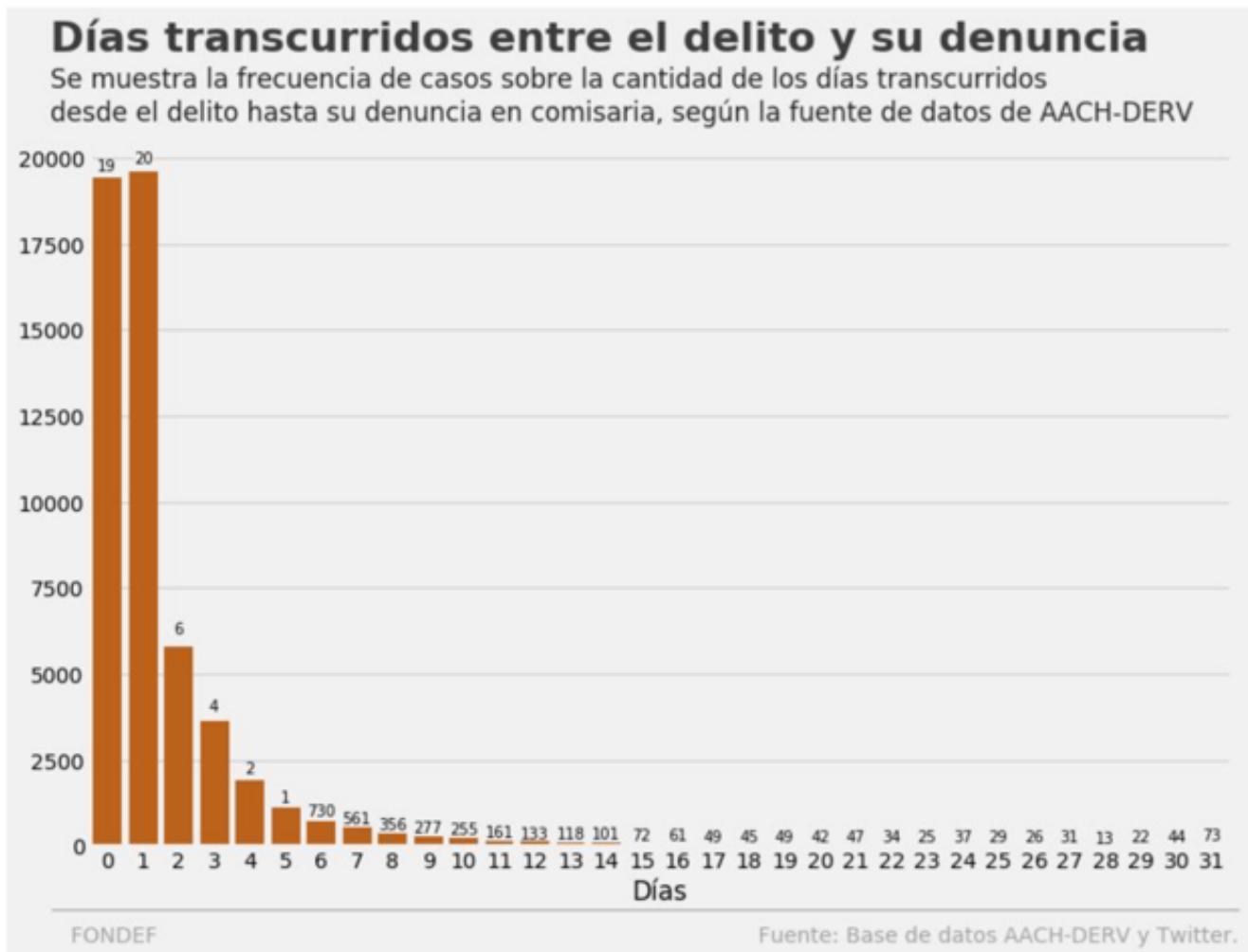


Figura 4. Tiempo entre el delito y su denuncia

La Figura 5 presenta un resumen de los robos de vehículos que se observan en la base de datos de AACH-DERV según la región en la que esta se registra el delito.

Insurance Companies' Data (4)

Cantidad de robos históricos por región

Se muestra la cantidad y proporción de robos desde 2011 en la fuente de datos de AACH-DERV por región el % es sobre el total del período

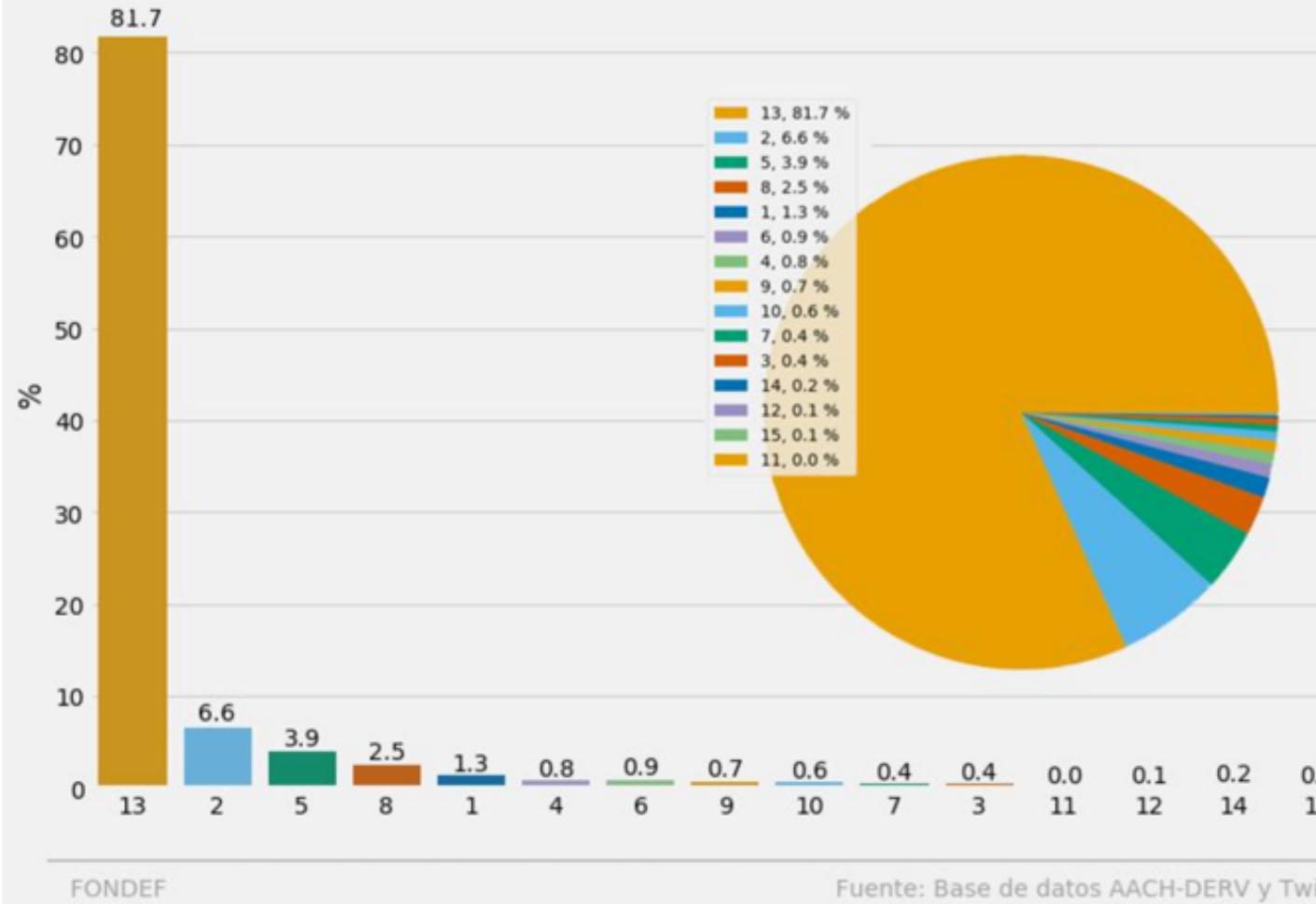
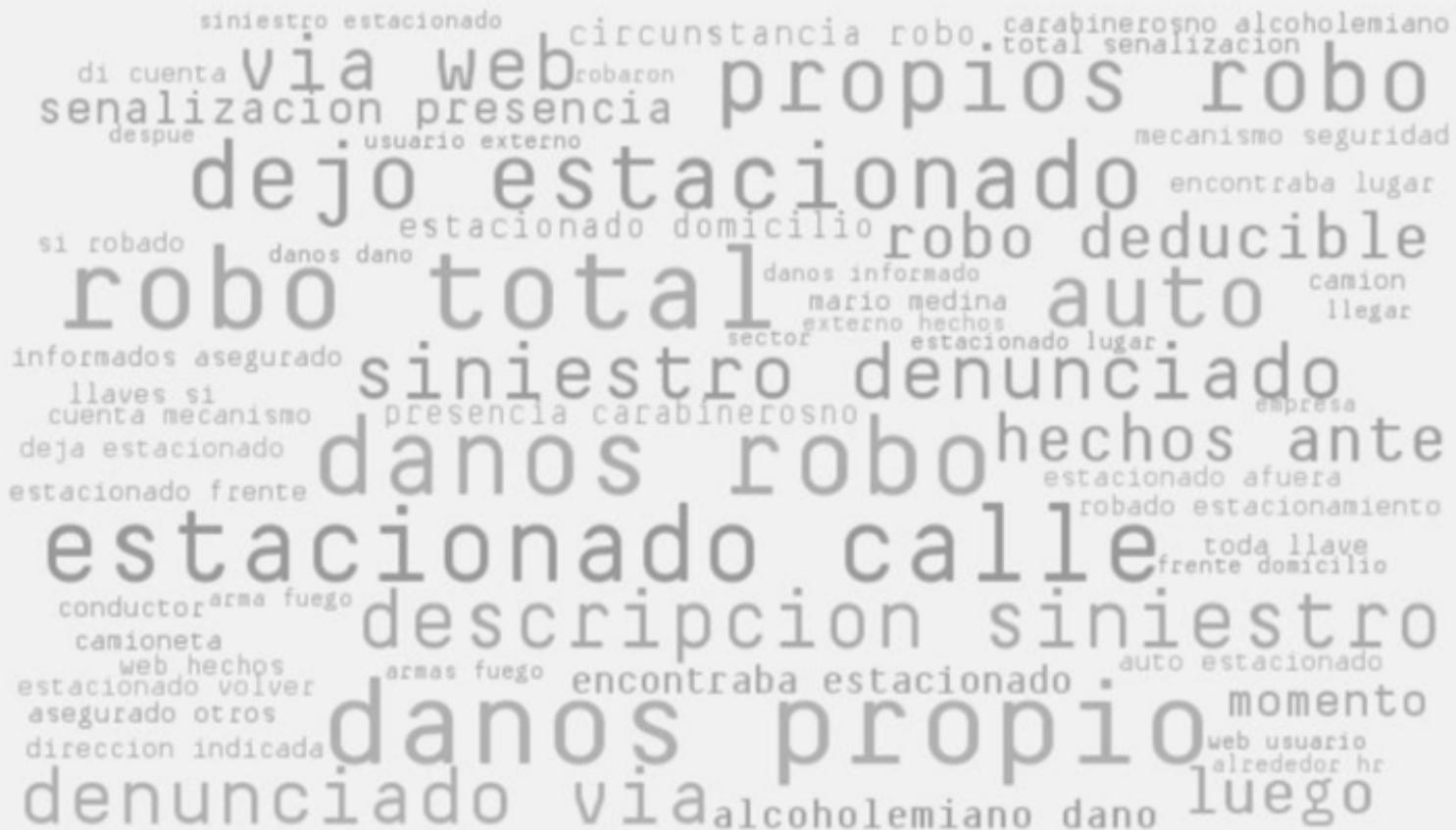


Figura 5. Distribución del robo de vehículos por región

Insurance Companies' Data (5)

Nube de palabras más frecuentes en relatos

En base a los siniestros históricos ocurridos en el período de 1990 hasta Enero de 2017, se muestran las palabras más comunes en los relatos del siniestro



FONDEF

Fuente: Base de datos aseguradoras.

Figura 8. Nube de palabras de los relatos prestados por las víctimas

Insurance Companies' Data (6)

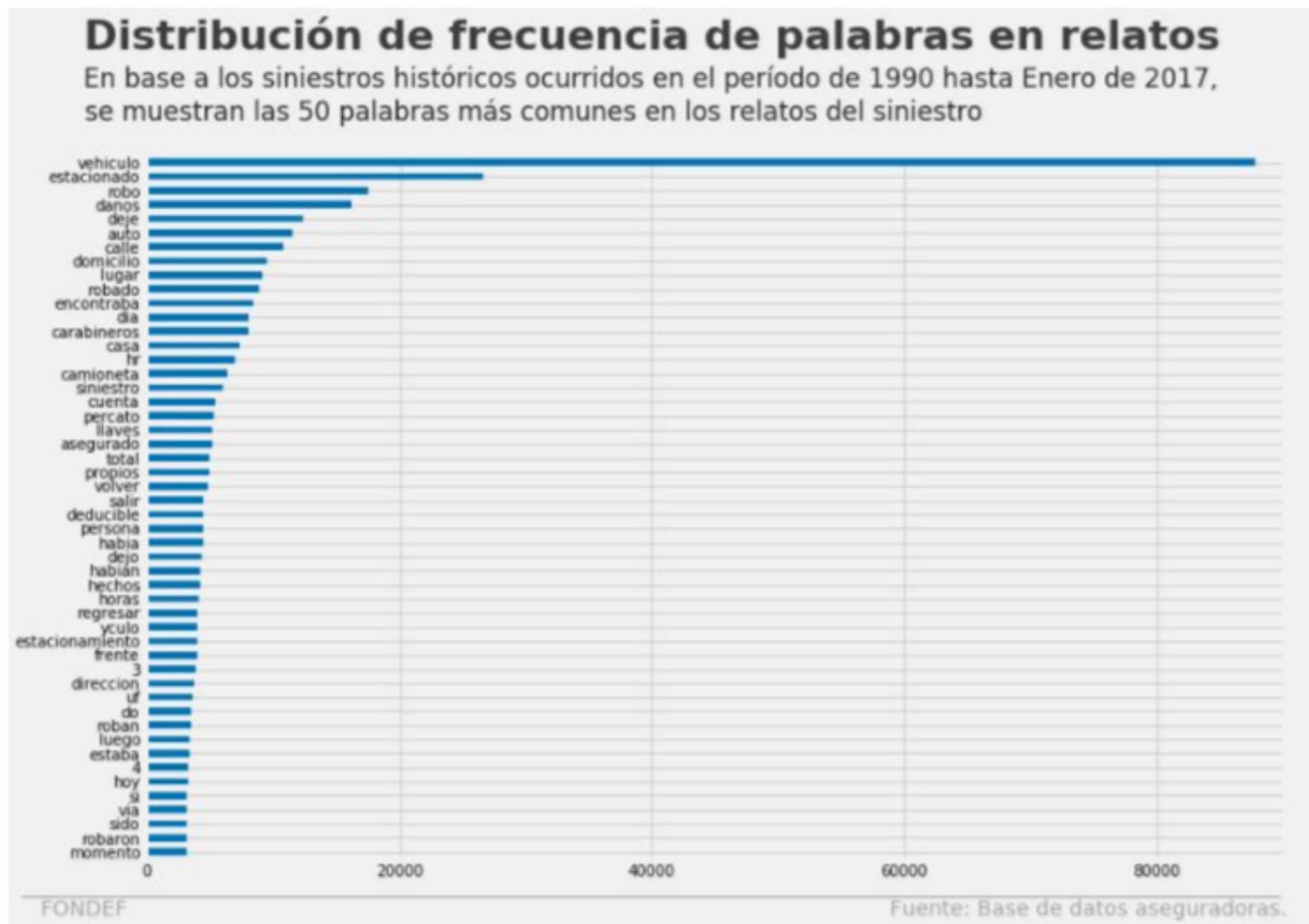


Figura 9. Frecuencia de las palabras en los relatos de las víctimas

Twitter Data

- How to collect it ? Search terms:
 - Stolen car (auto robado) **X**

Twitter Data

- How to collect it ? Search terms:
 - Stolen car (auto robado) **X**
 - Plaque (Patente) **👍**

 **Autos Robados Chile** @robados_chile 1d
Vehículos recuperados por personal de la 49 com de Quilicura. Patentes DJGY78 y HWHB71. Felicitaciones a @Carabdechile por su labor 🙌



Patente DJGY78
INDAI NEW TUC
Patente HWHB71
SUBARU IMPREZA

...

 **Autos Robados Chile** @robados_chile 1d
Vehículos recuperados por personal de la 49 com de Quilicura. PATENTES DRWZ24, BSWW33 y JSGF99. Felicitaciones a @Carabdechile por su labor 🙌



Twitter Data (2)

Tipos de vehículos más reportados en ambas fuentes

Ambas fuentes presentan la misma tendencia. En el caso de Twitter la tendencia al robo de tipo AUTOMOVILES es aún mayor.

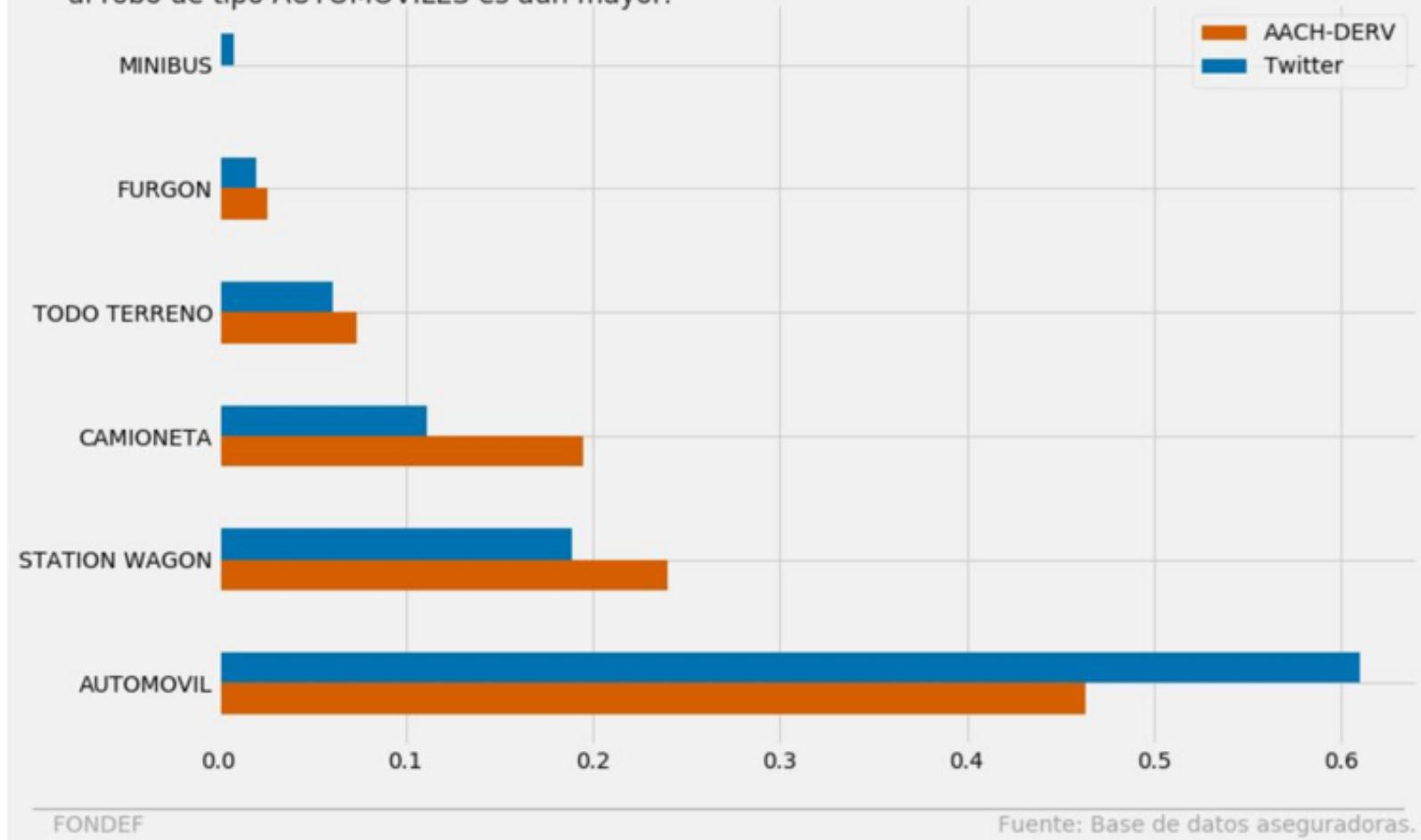
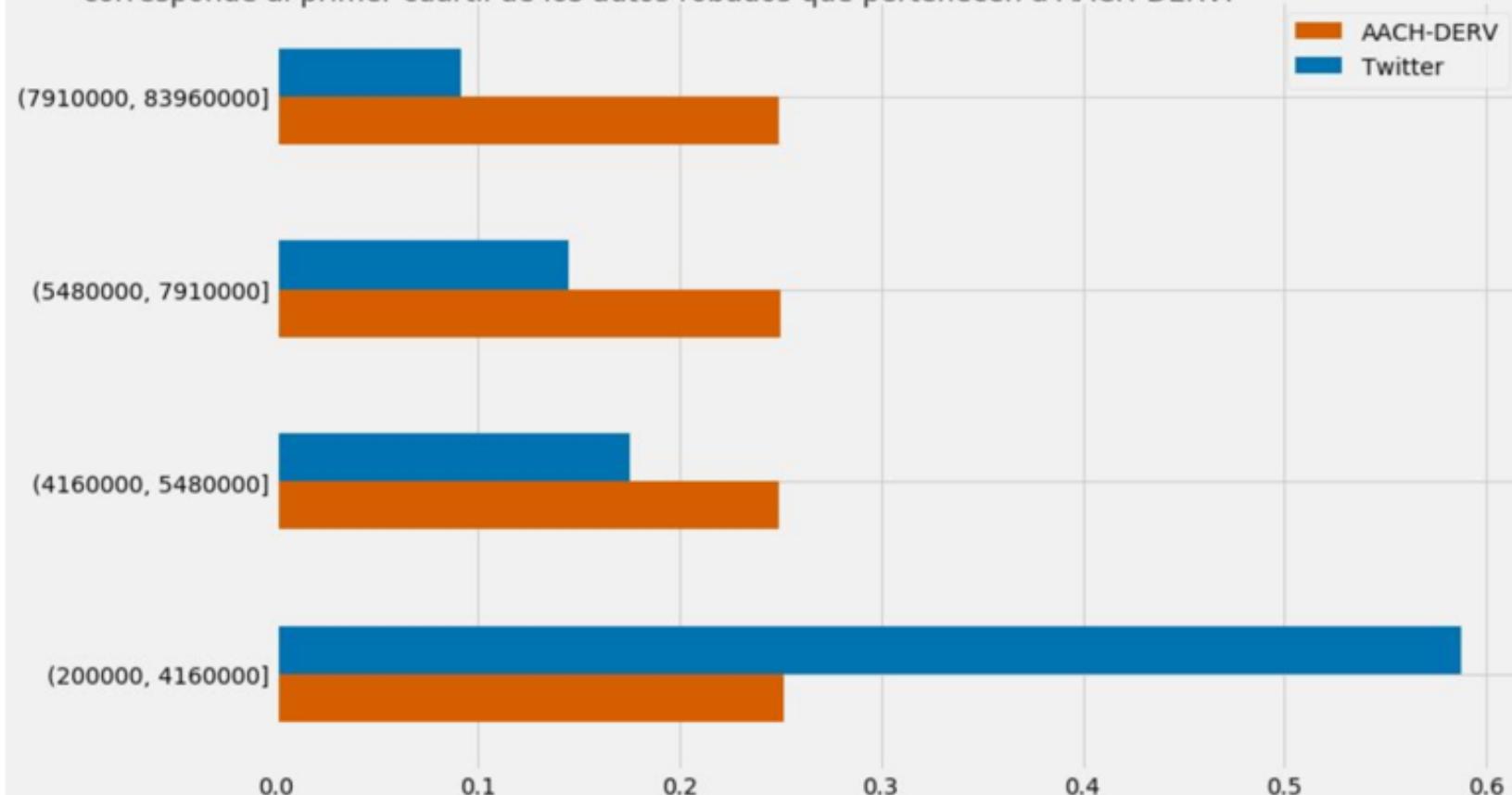


Figura 10. Reporte de robos más frecuentes en Twitter y AACH-DERV

Twitter Data (3)

Tasaciones de vehículos reportados en ambas fuentes

Podemos ver que la tasación de la mayoría de los autos robados reportados en Twitter corresponde al primer cuartil de los autos robados que pertenecen a AACH-DERV.



FONDEF

Fuente: Base de datos aseguradoras.

Figura 11. Relación con la tasación de los vehículos en ambas fuentes de datos

Twitter Data (4)

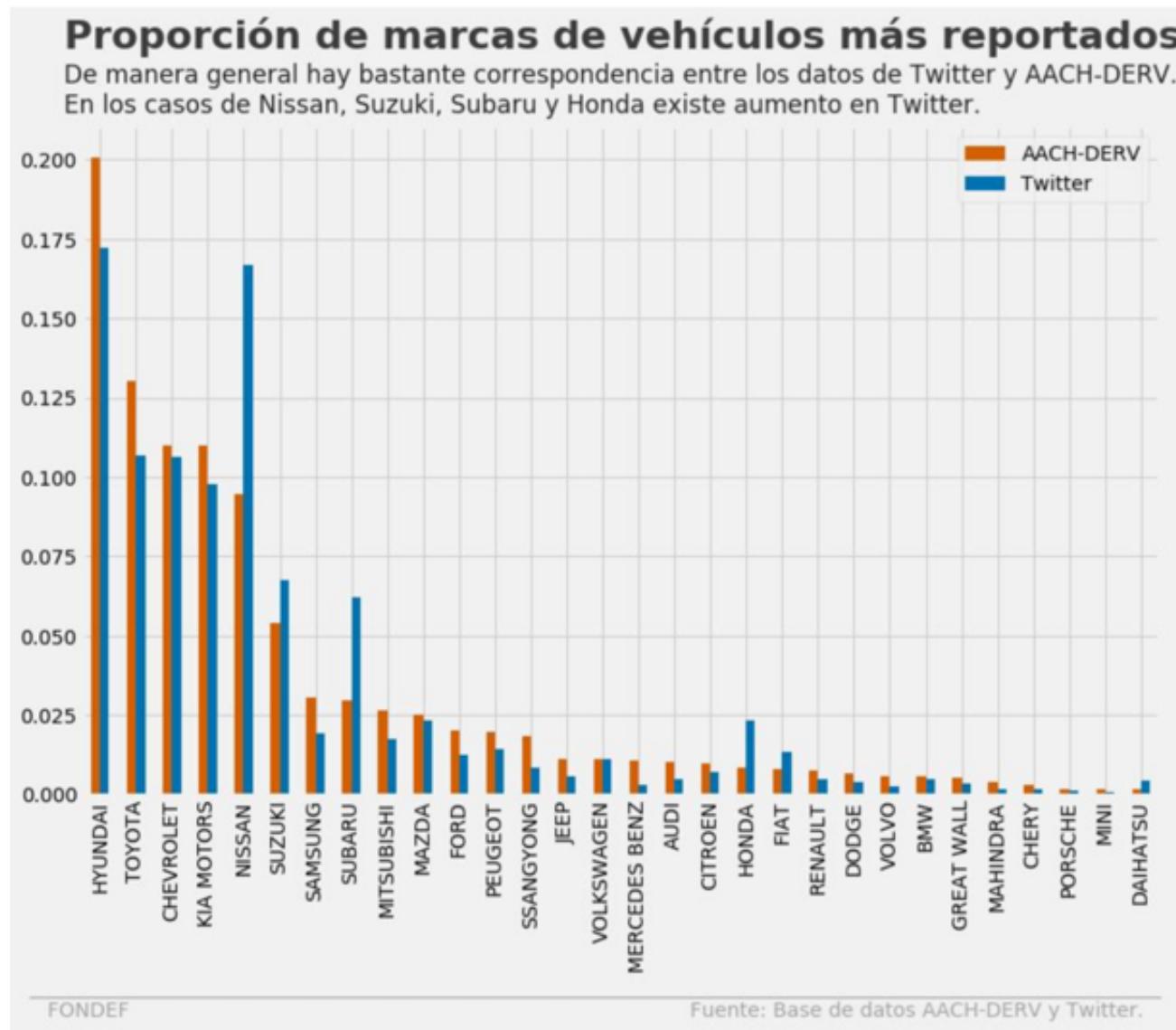
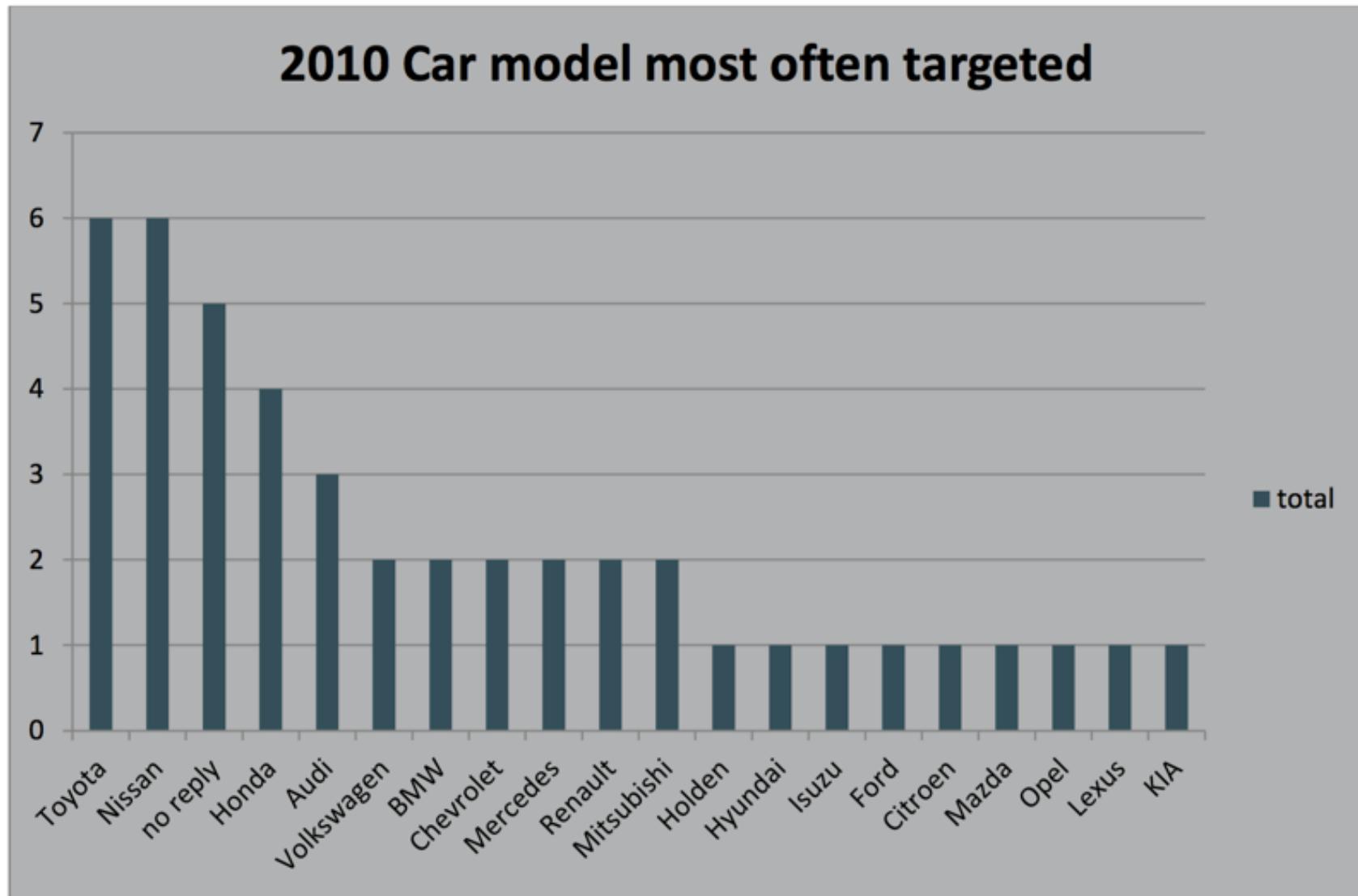


Figura 12. Marcas de autos más frecuentes en los reportes en ambas fuentes de datos

Interpol Data (2010)



Twitter Data (5)

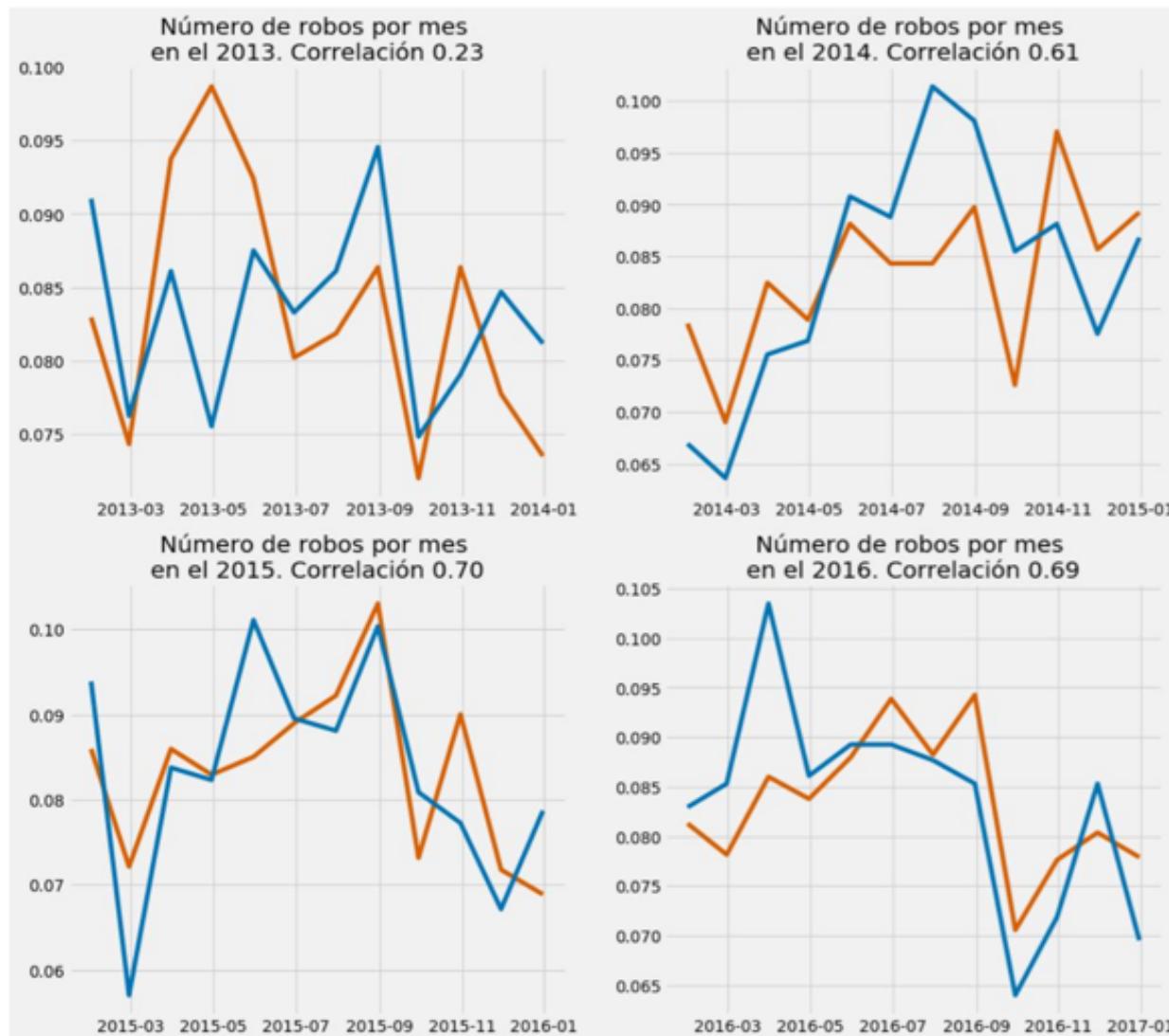


Figura 16. Número de robos mensuales por año en ambas fuentes de datos

Twitter Data (6)

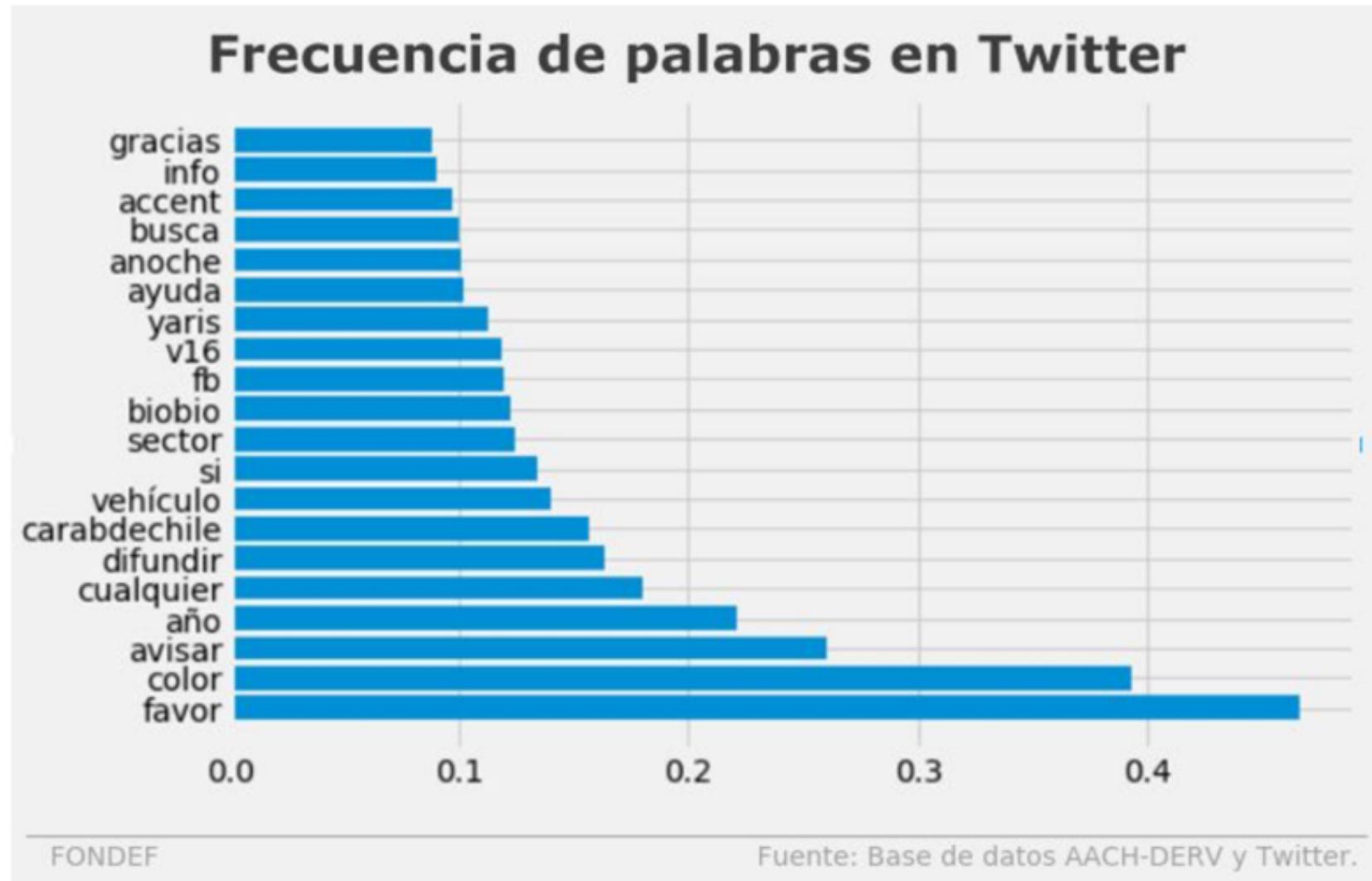


Figura 22. Frecuencia de las palabras que aparecen en los tweets consultados

Twitter Data (7)

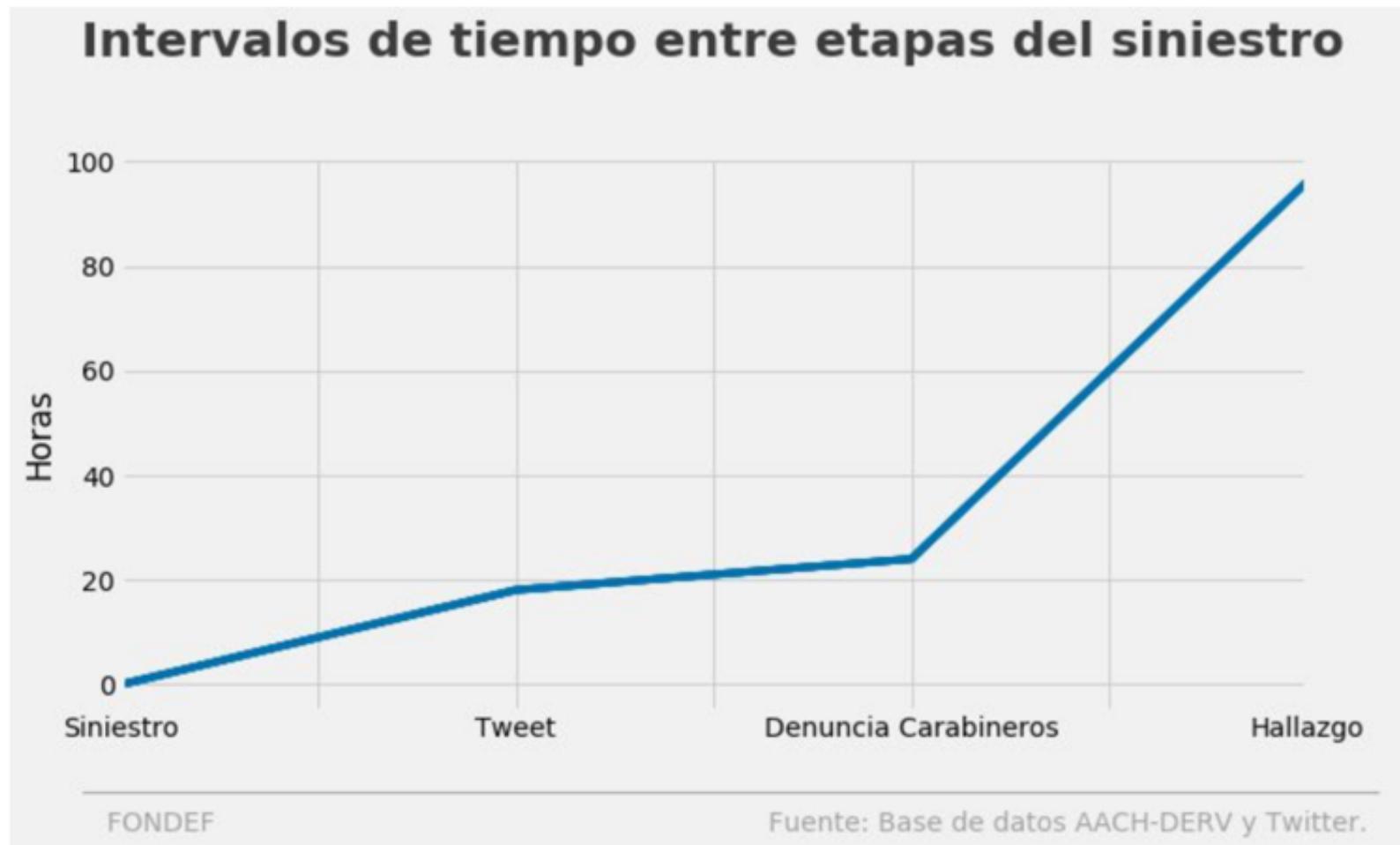
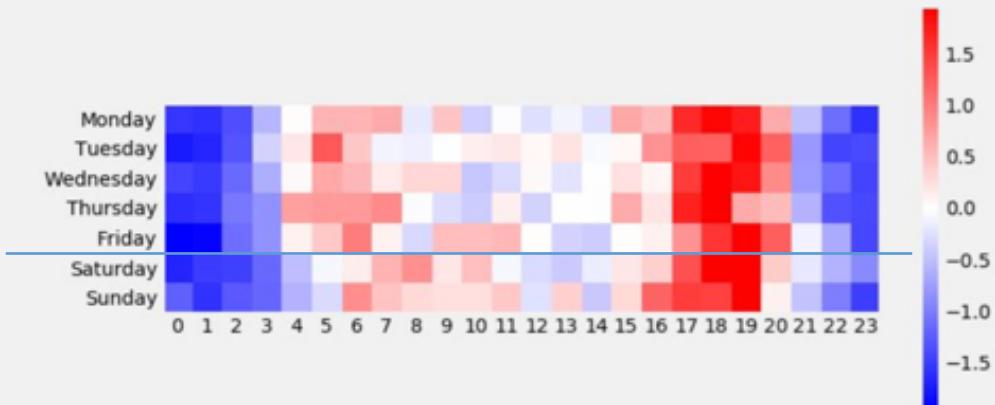


Figura 20. Intervalos de tiempo que transcurren entre las etapas del proceso que comprende el robo de un vehículo y su eventual hallazgo

Twitter Data (8)

- Thieves need rest too: they start later on weekends (or people report later the events)

Robos por hora y día de la semana en AACH-DEV

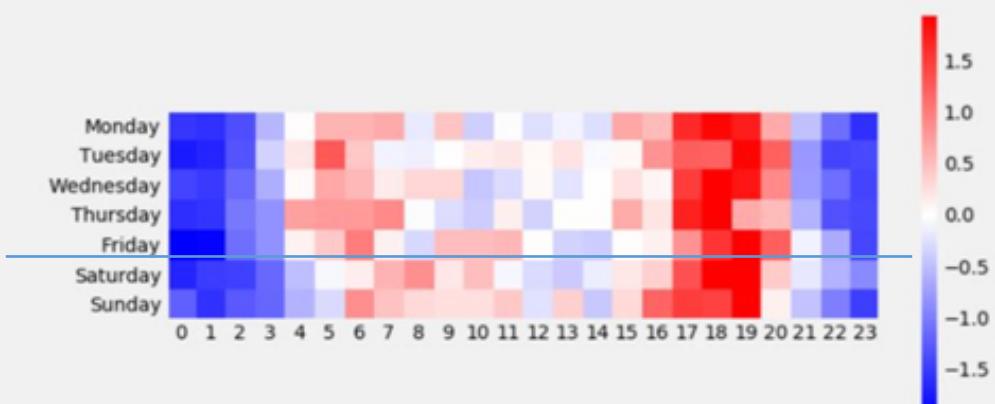


FONDEF

Fuente: Base de datos AACH-DEV y Twitter.

Figura 18. Robos por hora del día y día de la semana en AACH-DEV

Robos por hora y día de la semana en Twitter



FONDEF

Fuente: Base de datos AACH-DEV y Twitter.

Digital Media Data



Figura 24. Distribución de noticias por mes

Digital Media Data (2)

Tópicos	Palabras relacionadas
Autoridades-Política	Alcalde – Alessandri – Municipio – Autoridades – Ministros – Bachelet – (Pre) Candidato – Candidatura – Primarias – Lavín – Concejales – Intendencia – Comisión – Diputado – Senador – Piñera – Seremi
Educación	Liceo – Apoderado – Estudiantes – Universitarios – Colegio – Universidad – Marcha – Secundario – Estudiantil – Confech - Rector
Policial-Delitos	Acoso – Allanamiento – Querella – Demanda – Maltrato – Sumario – Abuso – Apuñalan – Asaltante – Asalto – Delincuente – Balacera – Baleado – Disparo – Banda – Ataque – Carabineros – Tribunal – Encapuchados – Víctima – Condenan – Culpable – Prisión – Presidio - PDI
Economía-Negocios	Inversión – Déficit – Deuda – Crédito – Empresa – Concesión – Quiebra – Financiero – Pesos – Millones – Cobro – Morosos - Evasión
Accidentes vehiculares	Vehículo – Automóvil – Atropello – Colisión –

Digital Media Data (3)

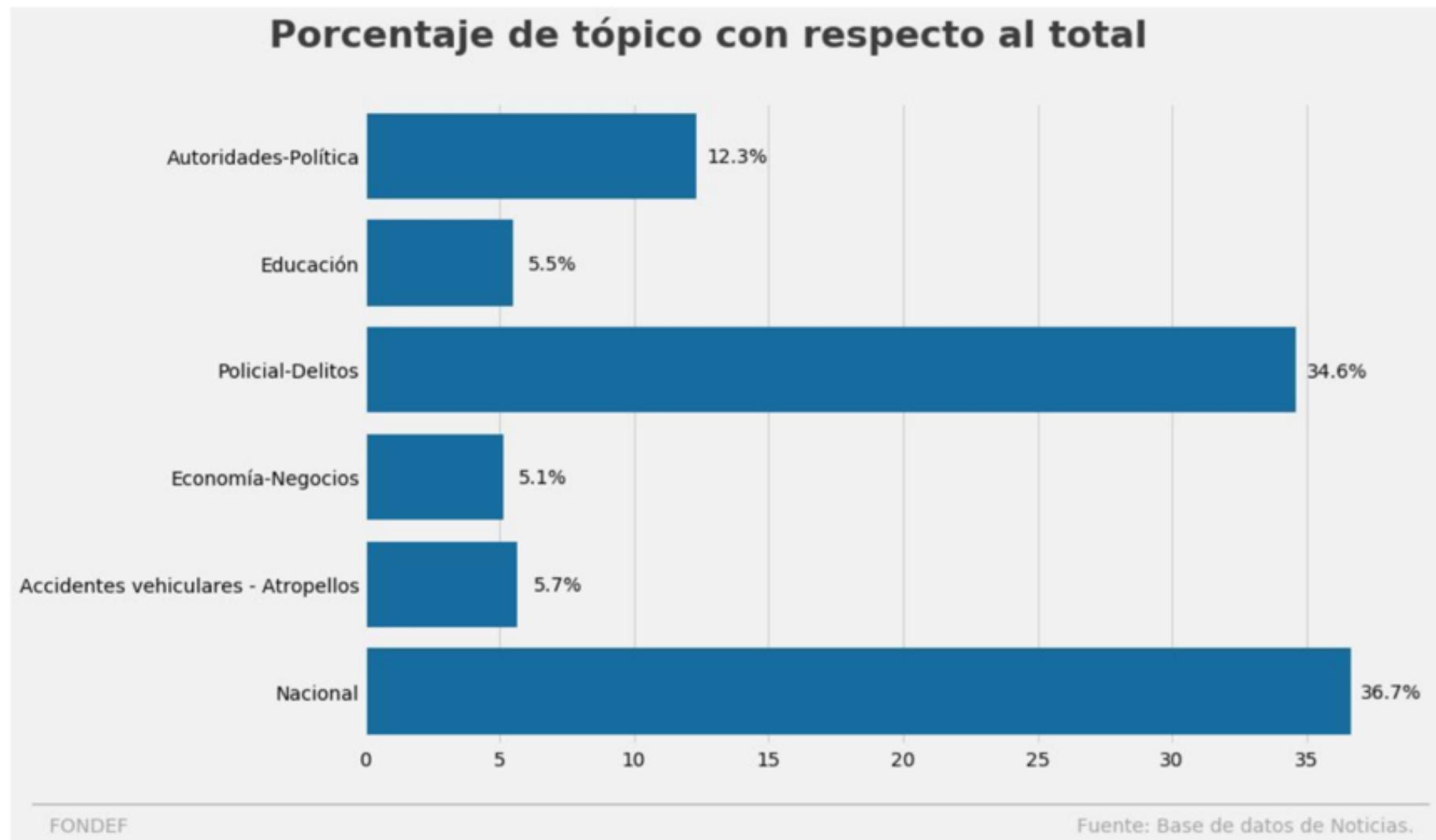


Figura 25. Porcentaje de noticias por cada tópico

Data Analysis (Comparison)

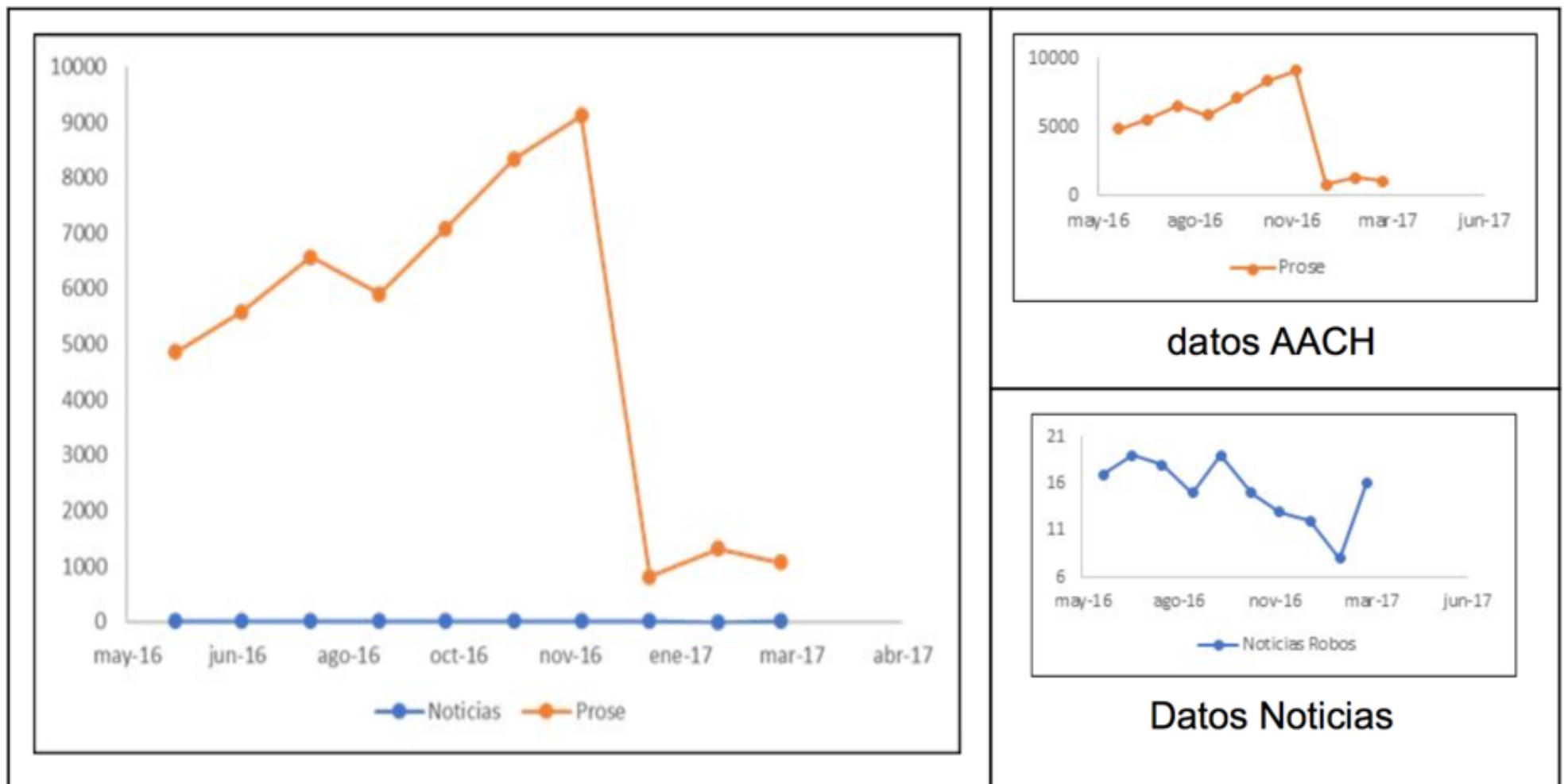
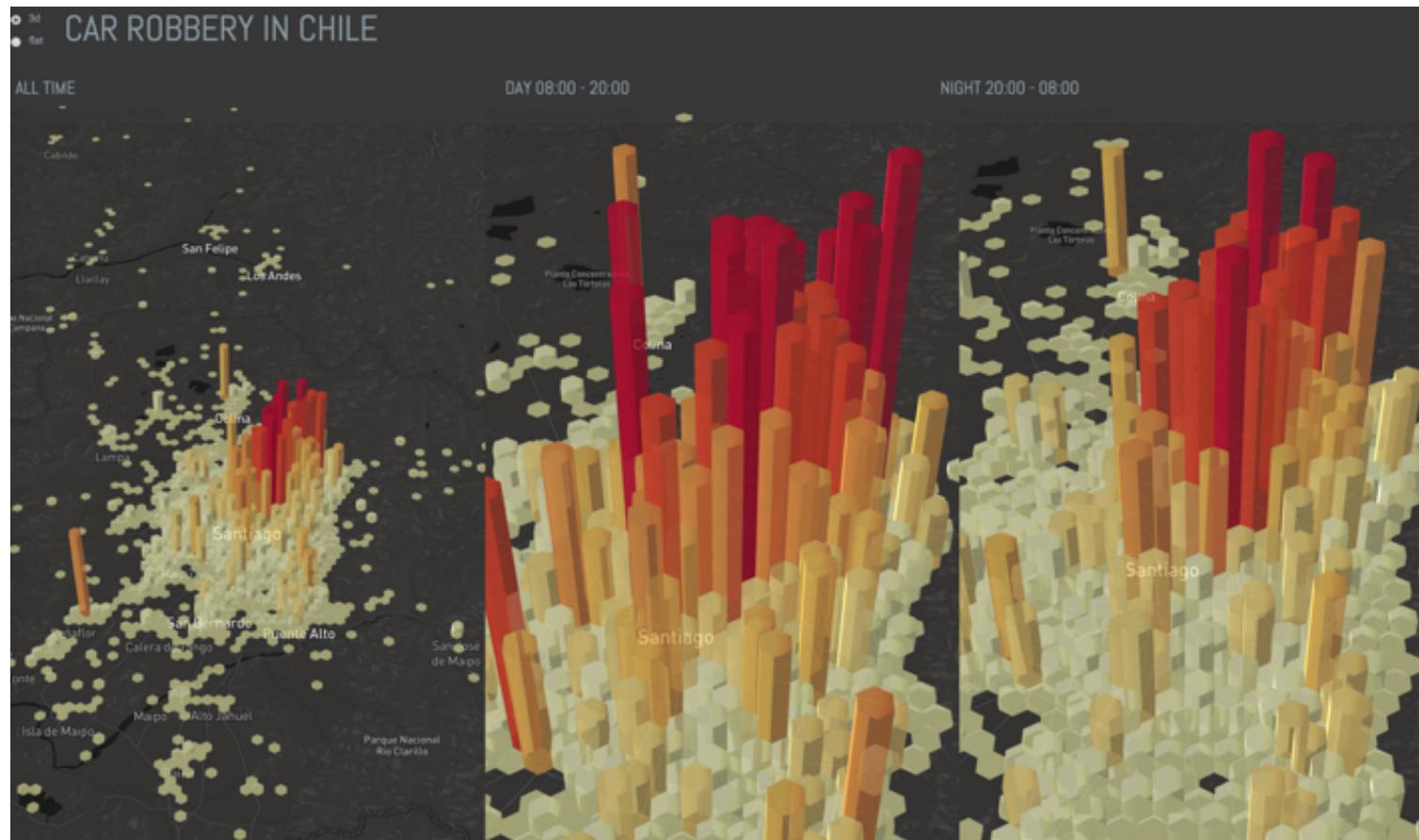


Figura 26. Serie de tiempo comparando volumen de datos de AACH con datos de noticias

Prototype I

- http://niebla.ing.puc.cl/car_robbery_map/



Twitter Data + Insurance Data

Nombre	Tipo de dato	Filas no null	Descripción
id_prose	ordinal	55626	ID única transacción.
sin_patente	nominal	55626	Patente vehículo siniestrado.
sin_fecha_siniestro	fecha	55626	Fecha (día, mes y año) en donde ocurrió el siniestro de robo.
ase_abreviatura	nominal	55626	Abreviatura de aseguradora en la cual se registra el robo.
obs_ult_estado	nominal	3179	Validación interna.

sin_relato	nominal	51451	Relato del siniestro.
------------	---------	-------	-----------------------

 **Autos Robados Chile** @robados_chile 1d
Vehículos recuperados por personal de la 49 com de Quilicura. Patentes DJGY78 y HWHB71. Felicitaciones a @Carabdechile por su labor 🙌



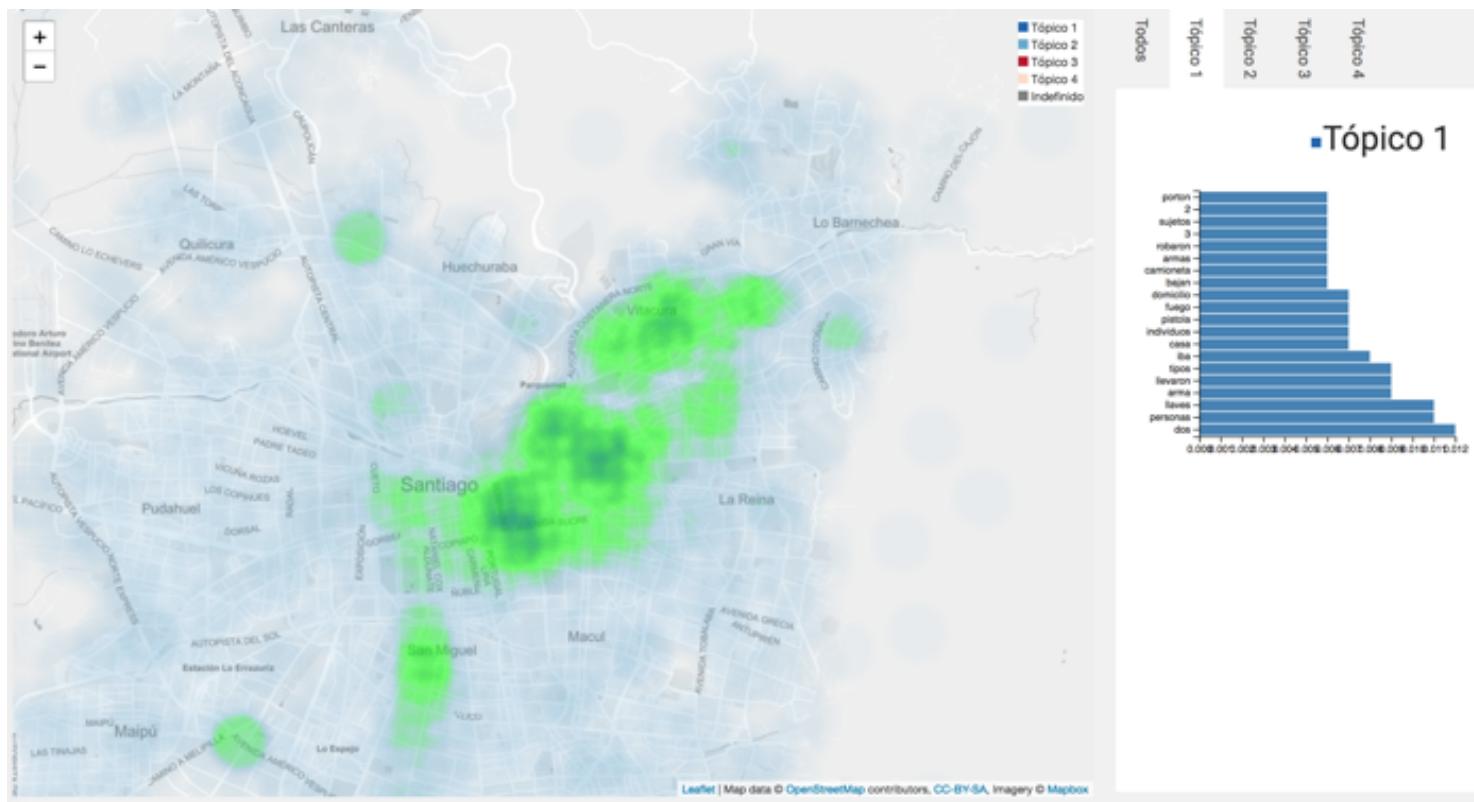
Comment Reply Like More

 **Autos Robados Chile** @robados_chile 1d
Vehículos recuperados por personal de la 49 com de Quilicura. PATENTES DRWZ24, BSWW33 y JSGF99. Felicitaciones a @Carabdechile por su labor 🙌



Prototype II: ITACaT

- Topic modeling over description of the criminal events (Twitter & Car Insurance Data)



Thanks!

- dparra@ing.puc.cl