



Data matching between road crash and vital registration datasets in São Paulo

Buenos Aires, April 2019



Road Safety in São Paulo 2017

ESTATÍSTICAS BASE 2017

15.455

VÍTIMAS FERIDAS

797

VÍTIMAS FATAIS

1 PESSOA MORRE A CADA
17 ACIDENTES REGISTRADOS

331 Pedestrians

311 Motorcyclists

188 4-wheeled vehicle users

37 Cyclists

Estimated total cost

US\$ 195 Million

MORTALIDADE NO TRÂNSITO E A META DA ONU

2011

12 MORTOS
/ 100 MIL HABITANTES

2017

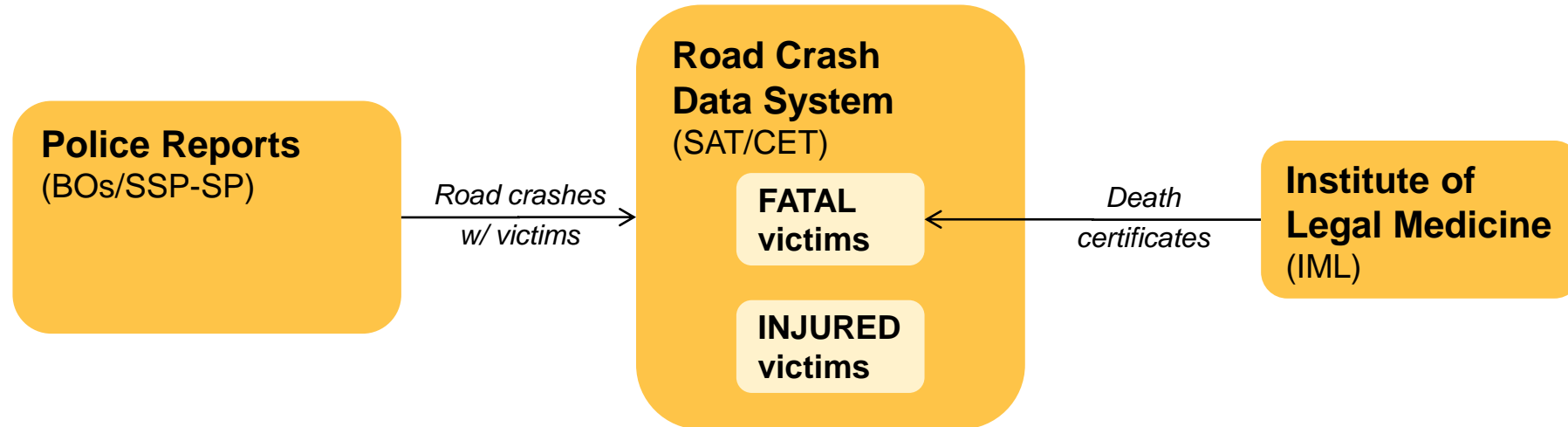
6,56 MORTOS
/ 100 MIL HABITANTES
(REDUÇÃO DE 45,3%)



Institutional background

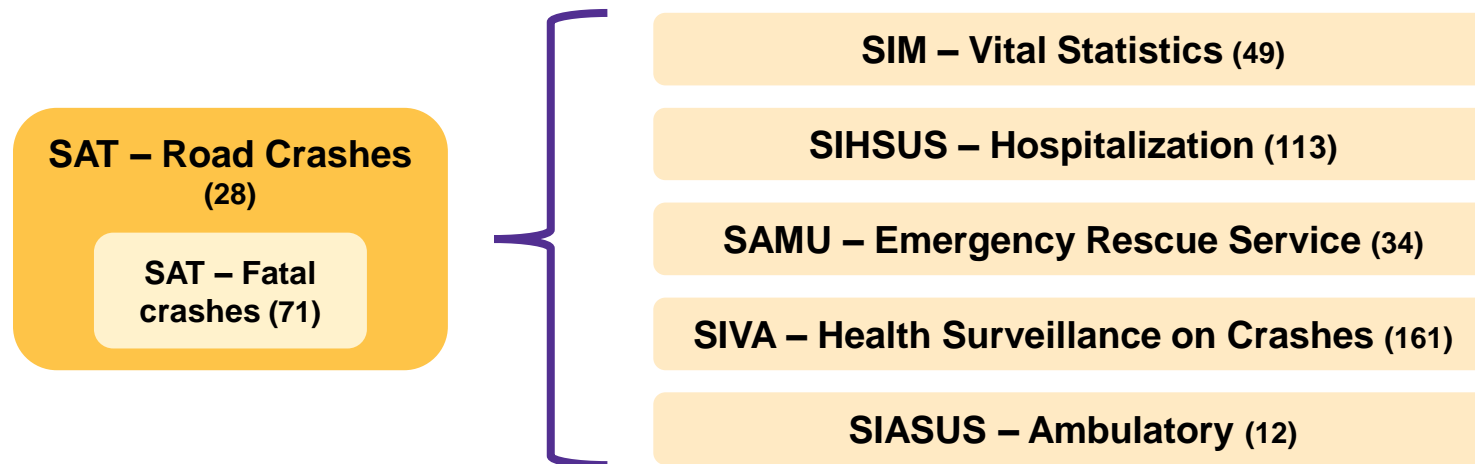
- Results from previous Bloomberg Initiative work (RS in 10 countries) – 2015
 - Inter-sectoral working group for Data Matching (dissolved)
- BIGRS: Project presentation to Health and Transportation Secretaries
 - City officials from RS-10 group
 - Data sharing agreements (NDA for health data)
 - 2017-2018: Development of R scripts and testing on 2014-2015 data
 - 2019: Working Group formalized and part of the Road Safety Plan Committee

Initial structure for crash data

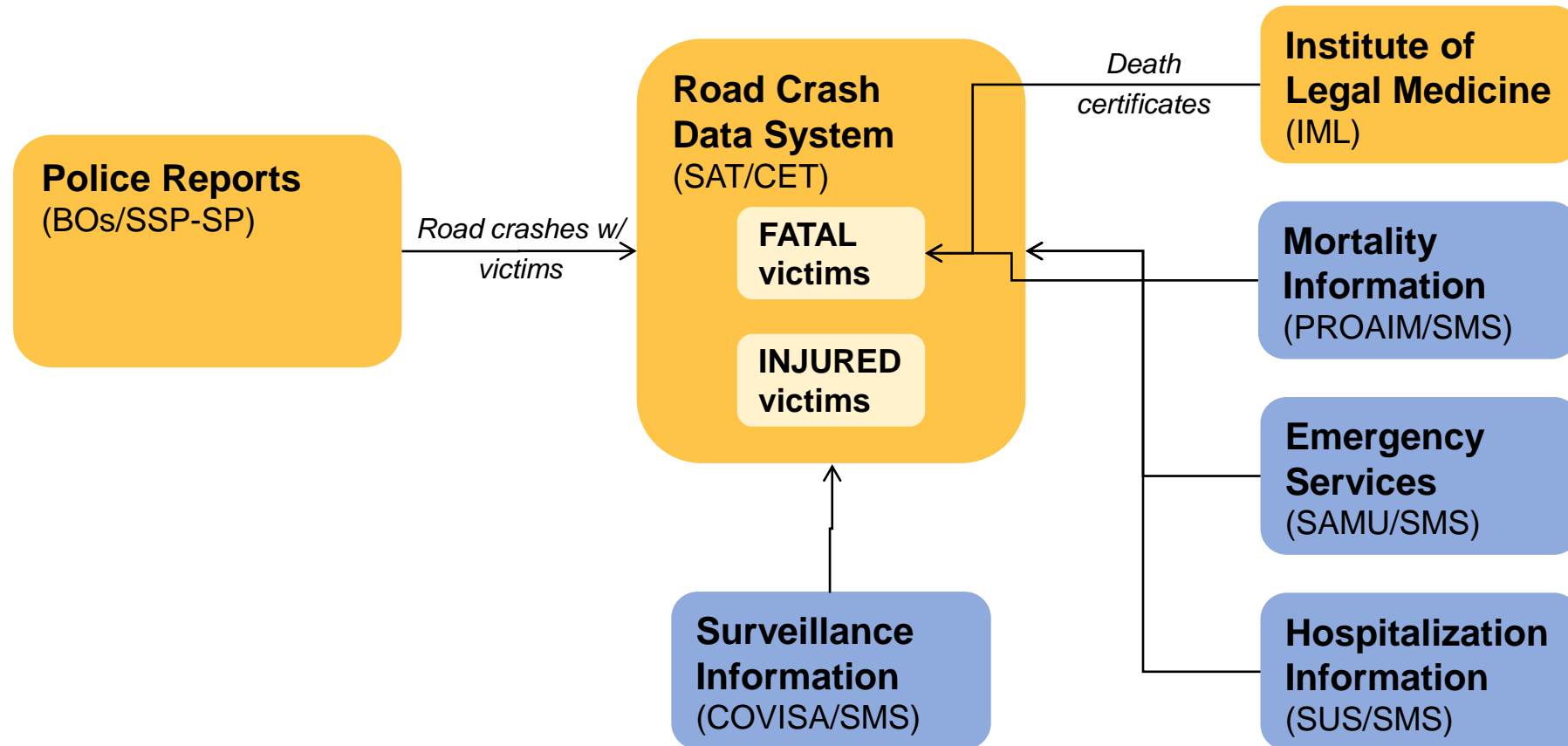


Goal and datasets

- Complement information on road crashes collected by Police/CET with information from health information systems
 - Generate more specific indicators that contribute to public policies aimed at RTI reduction

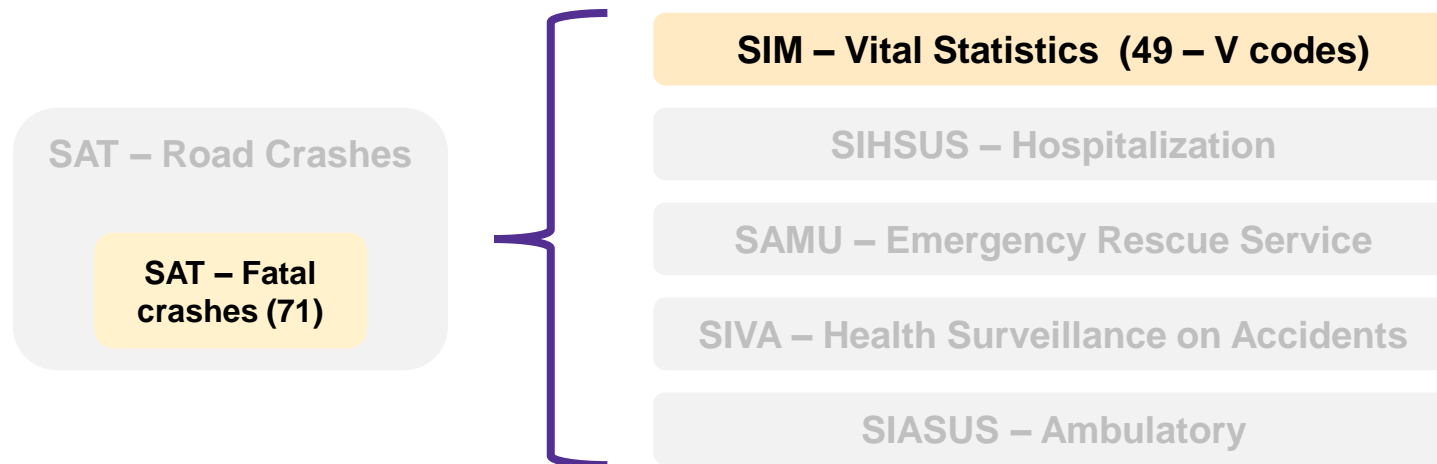


Proposed structure for crash data



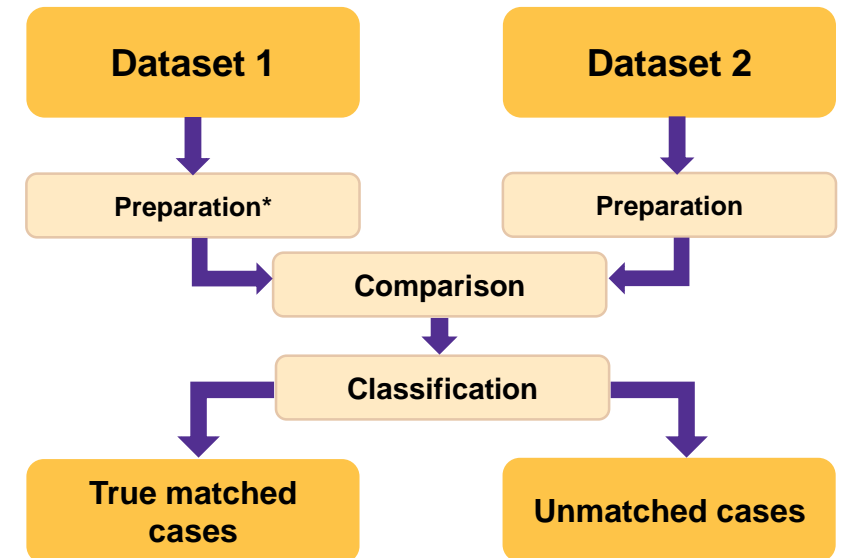
Goal and datasets

- Complement information on road crashes collected by Police/CET with information from health information systems
 - Generate more specific indicators that contribute to public policies aimed at road crash reduction



Probabilistic Matching

- Estimating the probability of 2 individuals from different databases being the same person
- Used when no common ID field is available
- Selected variables for matching:
 - Name
 - Birth date
 - Death date
 - Gender
- Observations:
 - phonetic algorithm (SOUNDEXBR)
 - gender as blocking



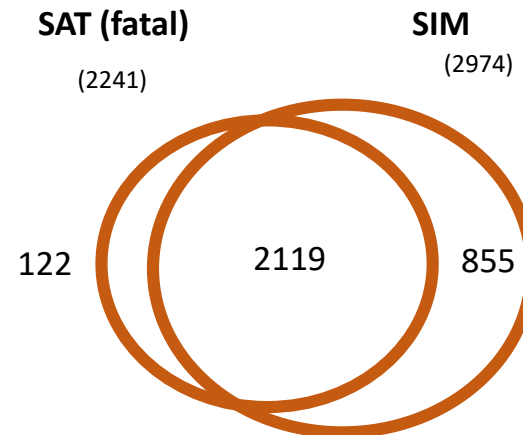
Matching fatal crash victims (SAT) x Vital Statistics (SIM)

id	FirstName	LastName	data.morte	DATA.NASC	Sexo	Weight
1146	ABILIO		22/11/2014	12/05/1925	M	
1463	ABILIO		22/11/2014	12/05/1925	M	41,229
1402	ADALBERTO		23/02/2015	27/03/1964	M	
1841	ADALBERTO		23/02/2015	27/03/1964	M	41,229
2184	ADALGISA		13/12/2015	20/02/1998	F	
2896	ADALGISA		13/12/2015	20/02/1998	F	41,229
2058	ADAO		06/10/2015	27/09/1971	M	
2684	ADAO		06/10/2015	27/09/1971	M	41,229



2223	HUGO		22/12/2015	02/02/1996	M	
4	JORGE		01/01/2014	02/02/1996	M	5,834
2232	AUDILEZ		18/12/2015	17/06/1986	M	
2007	JOAO		05/04/2015	17/06/1986	M	5,834
2236	VINICIUS		27/12/2015	05/12/1995	M	
446	LUIZ		09/04/2014	05/12/1955	M	5,834
2238	FLAVIO		02/12/2015	26/09/1994	M	
1366	GEAN		30/10/2014	26/09/1994	M	5,834

2014-2015



Matching SAT (fatal) x SIM

Results and further investigation

Matched cases

- % distribution of main body part injured stratified by road user type
- Define distance from home to site of crash

Unmatched cases

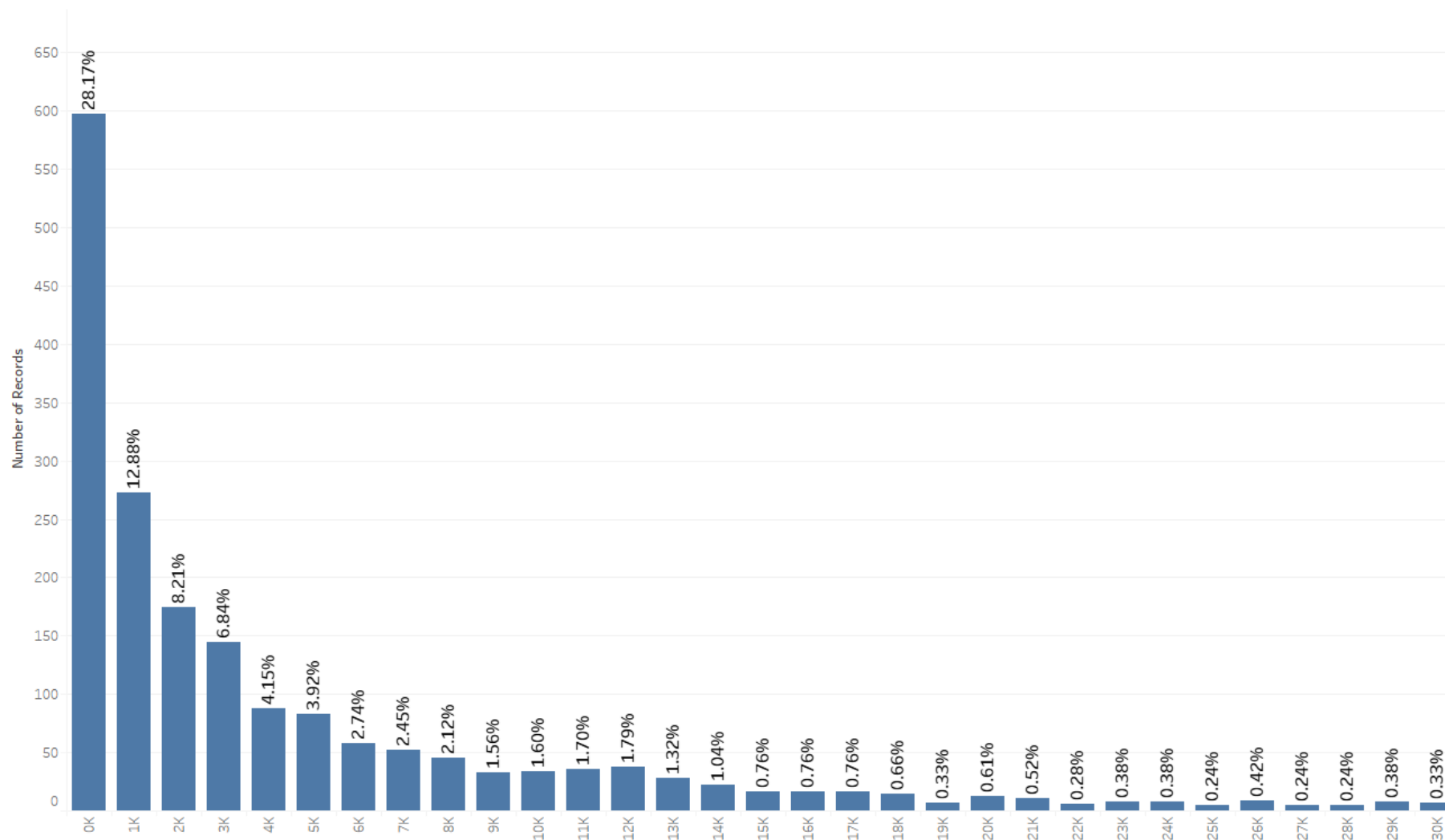
- Unmatched SAT cases: hypothesis that the death occurred outside the city or problem with ICD code.
- Unmatched SIM cases: Confirm whether all cases were due to crashes outside São Paulo (as hypothesized).

Distribution of body part injuries stratified by road user type

- Percentages are in relation to the number of fatal victims of specific road user types
- Victims often are coded with more than one ICD-10 code
- Not all victims have secondary ICDs (only the V-group coding)
 - Pedestrians:** high percentage of multiple and head injuries
 - Cyclists:** relatively few cases of multiple injuries, but highest percentage of head injury
 - Motorcyclists:** highest percentage of multiple injuries
 - Auto users:** high percentage of multiple injuries

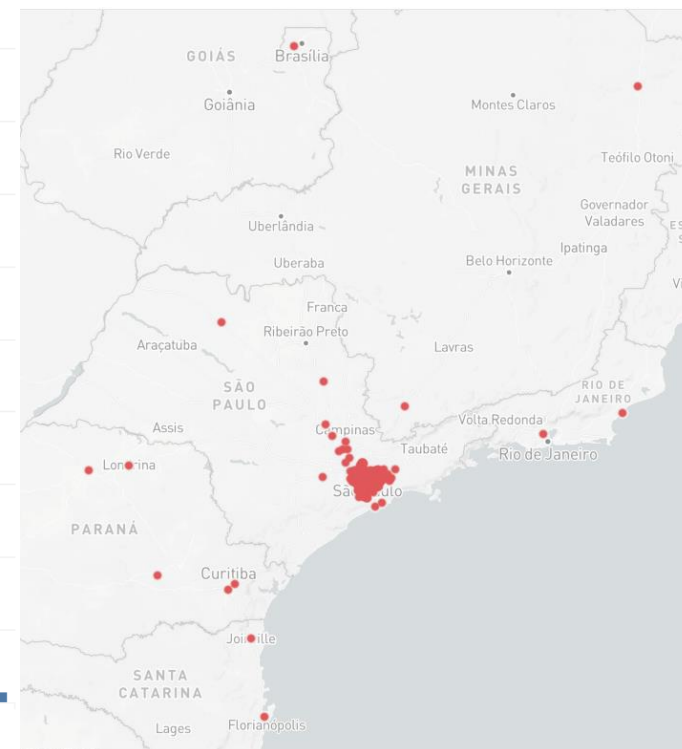
% notificação de ferimentos	Tipo de Usuário da via						
	Pedestre	Moto	Auto	Ciclista	Caminhão	Ônibus	Total
S0 - cabeça	40.04	34.32	33.04	52.56	29.17	37.50	39.36
S1 - pescoço	0.21	0.62	1.18	2.56	0.00	0.00	0.61
S2 - tórax	3.49	6.17	4.72	1.28	16.67	6.25	5.00
S3 - abdômen	2.46	5.06	3.83	1.28	12.50	0.00	3.87
S4 - ombro/braço	0.21	0.00	0.00	0.00	0.00	0.00	0.09
S5 - antebraço	0.00	0.00	0.00	1.28	0.00	0.00	0.05
S6 - pulso/mão	0.00	0.00	0.29	1.28	0.00	0.00	0.09
S7 - quadril/coxa	1.44	1.23	1.18	1.28	0.00	6.25	1.42
S8 - perna	0.62	0.37	0.29	0.00	0.00	0.00	0.47
S9 - pé	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T07 - múltipla	33.57	38.89	35.40	21.79	37.50	25.00	37.38
# vítimas fatais	974	810	339	78	24	16	2119

Distance between residence and place of crash



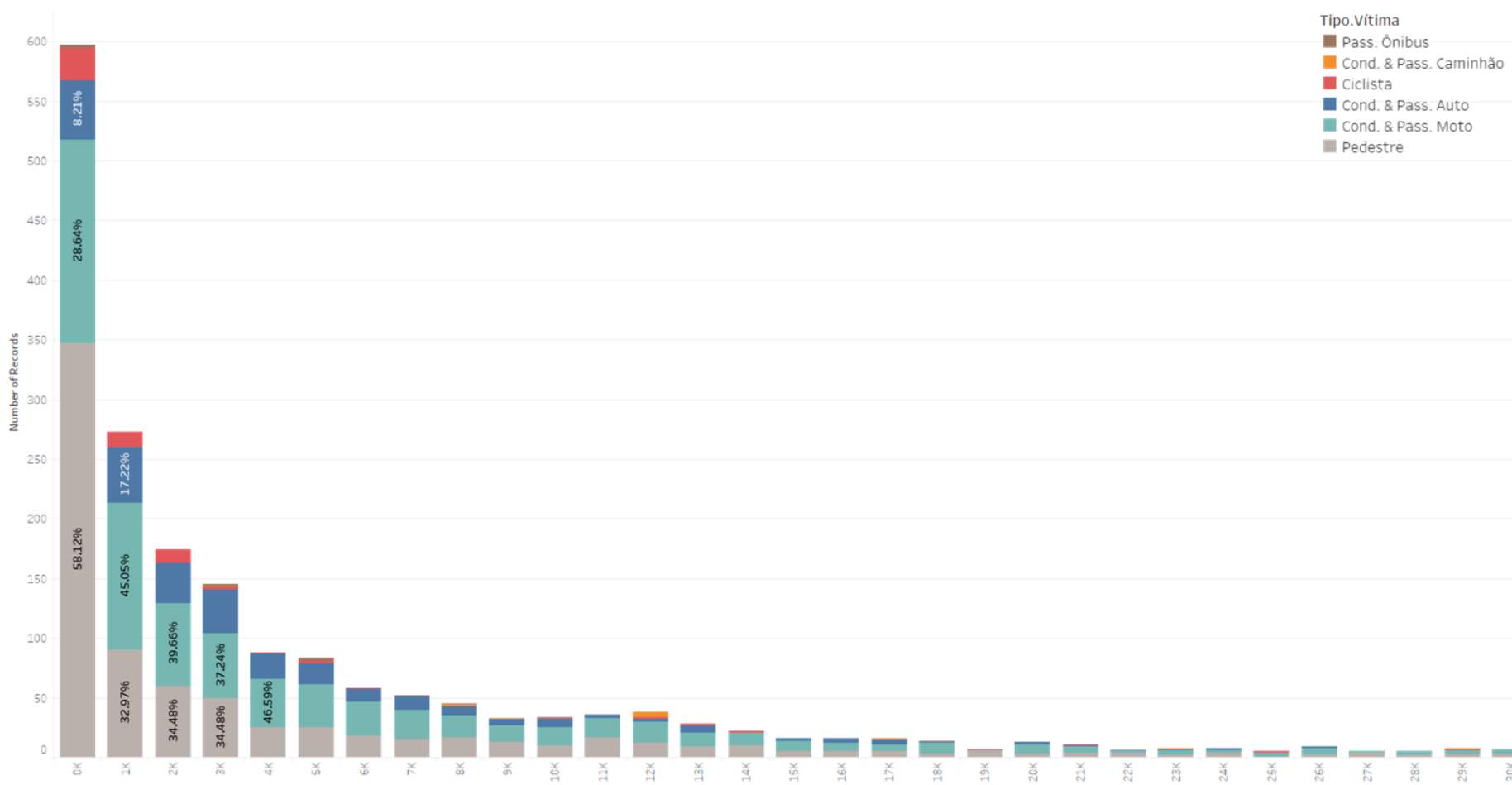
64,2% less than **5km** away (1360)

28,2% less than **1km** away (597)



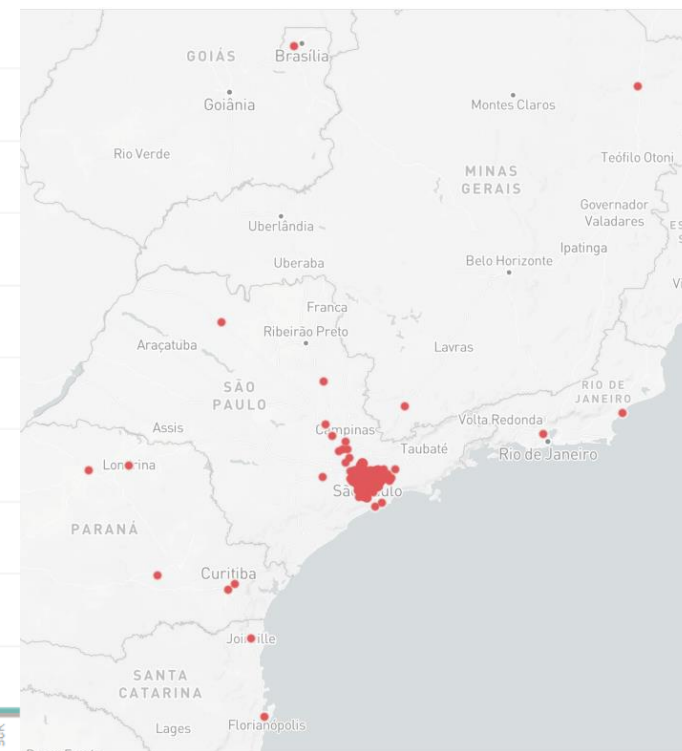
Distance between residence and place of crash

by road user type



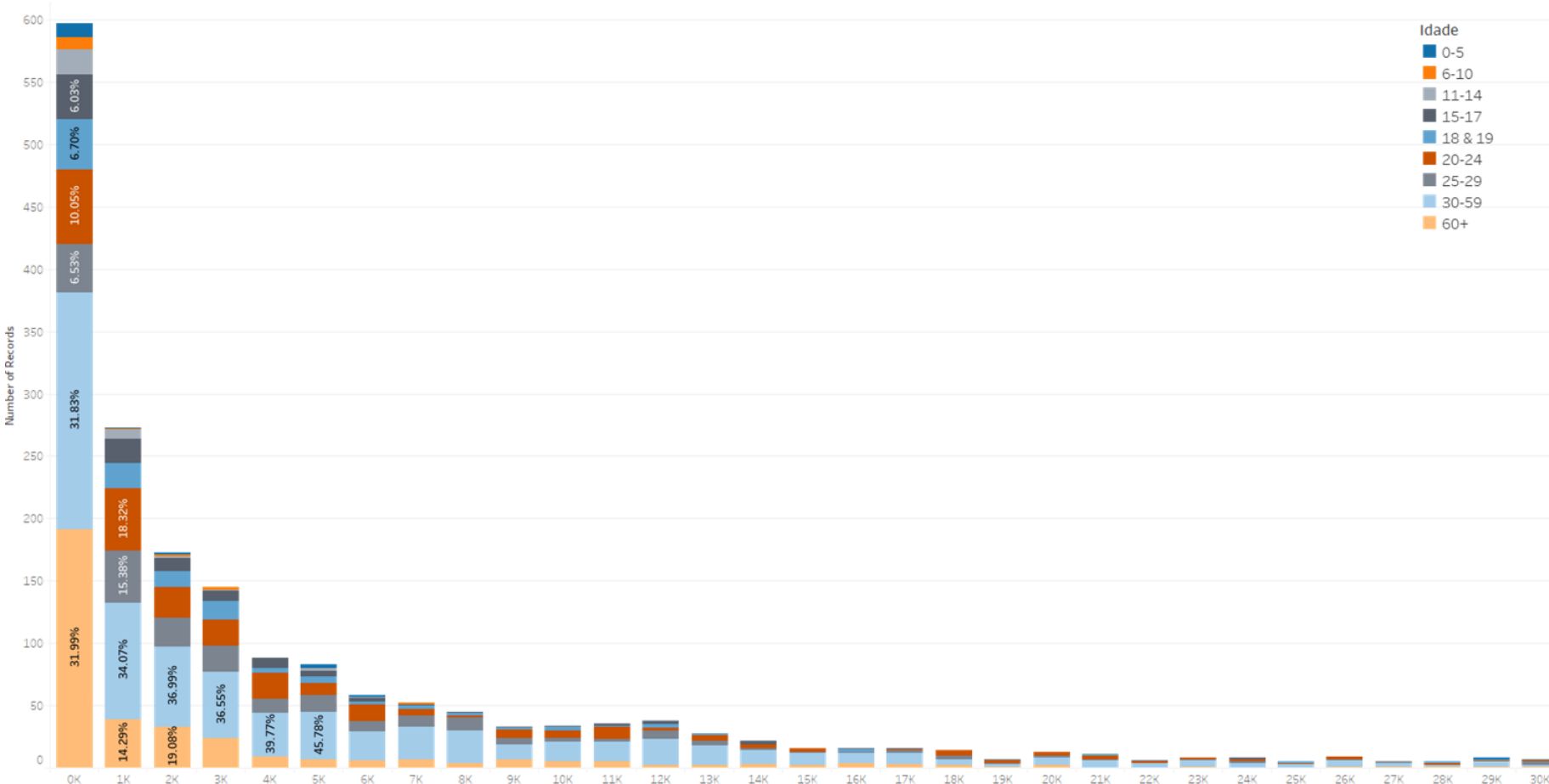
< 1km: 58% Pedestrians
29% Motorcyclists

< 5km: 45% Pedestrians
36% Motorcyclists



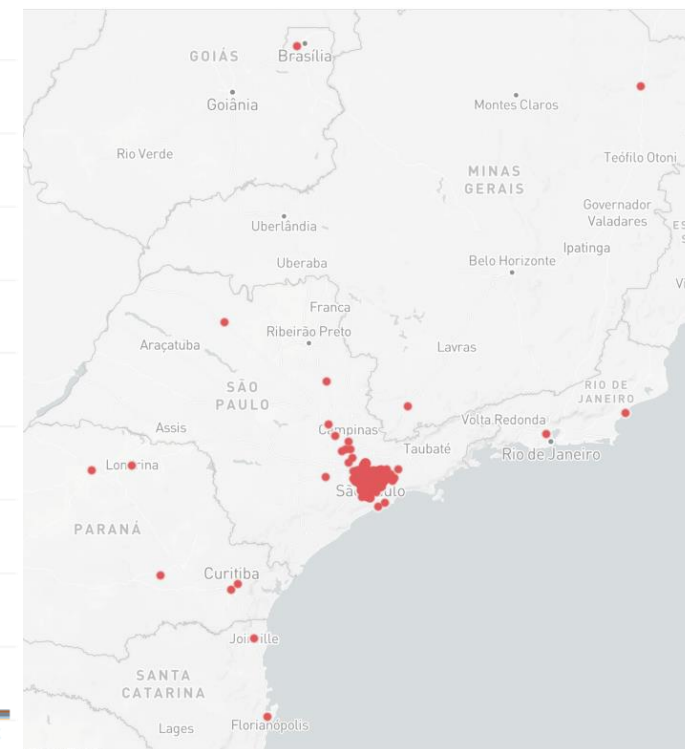
Distance between residence and place of crash

by age group



< 1km: **32%** 60+
3.5% 0-10

< 5km: **23%** 60+
2.2% 0-10



Unmatched SAT cases

“hypothesis that the death occurred outside the city or problem with ICD code.”

- **Verifying deaths outside SP**

- would require asking each of the other 38 cities in the metropolitan region or state level data (currently with little political support)

- **Problems with ICD code**

- worked with subset from identified road crash cases (**V** group)
- Use full SIM dataset for 2016-2017 period (182k entries)

Unmatched SIM cases

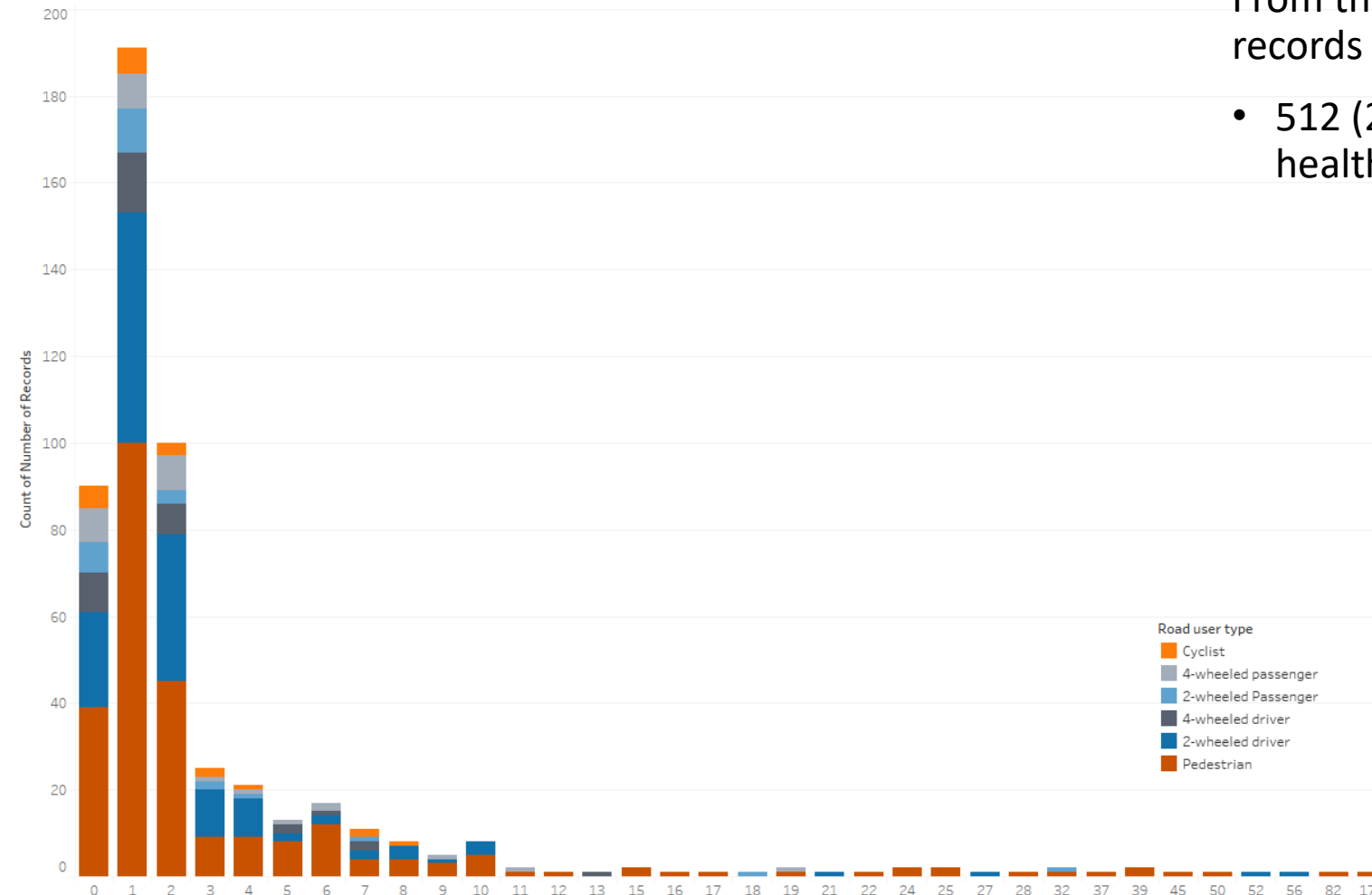
“Consider a follow up investigation to confirm whether all cases were due to crashes outside Sao Paulo”

- SIM for the 2014-2015 period has 2974 entries, 855 unmatched with SAT.
- 685 entries have no *external cause address* (SIM field)
 - Among the 2119 pairs SAT-SIM, only 21 cases lack address (and 1 didn't have an address on SAT)
- The remaining 2289 entries with road crash address on SIM are less precise than SAT, which frustrates geocoding efforts



Hospitalization records of fatal victims

Days of Hospitalization



From the 2119 matched cases between crash and death records in the 2014-2015

- 512 (24%) had hospitalization records on the public health system (SIH)

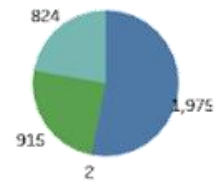
Average cost per road user type (in R\$)

	Average Cost SIH (A)	Matched cases (B)	Total cost (A*B) R\$
Pedestrian	4,140	259	1,072,191
2-wheeled driver	3,843	146	561,020
4-wheeled driver	5,810	36	209,160
2-wheeled Passenger	6,565	26	170,695
4-wheeled passenger	4,938	32	158,012
Cyclist	4,662	20	93,249

Crash and Emergency Rescue Service records

- SAMU (Serviço de Atendimento Médico de Urgência) data was made available for matching
- Low-quality victim information; helpful for emergency resource and hospital allocation
- Different matching procedure (time of day and geolocation)

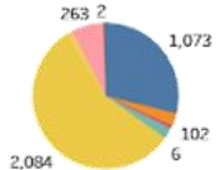
Tipo de Vítimas



Tipo Vítima



Veículos envolvidos



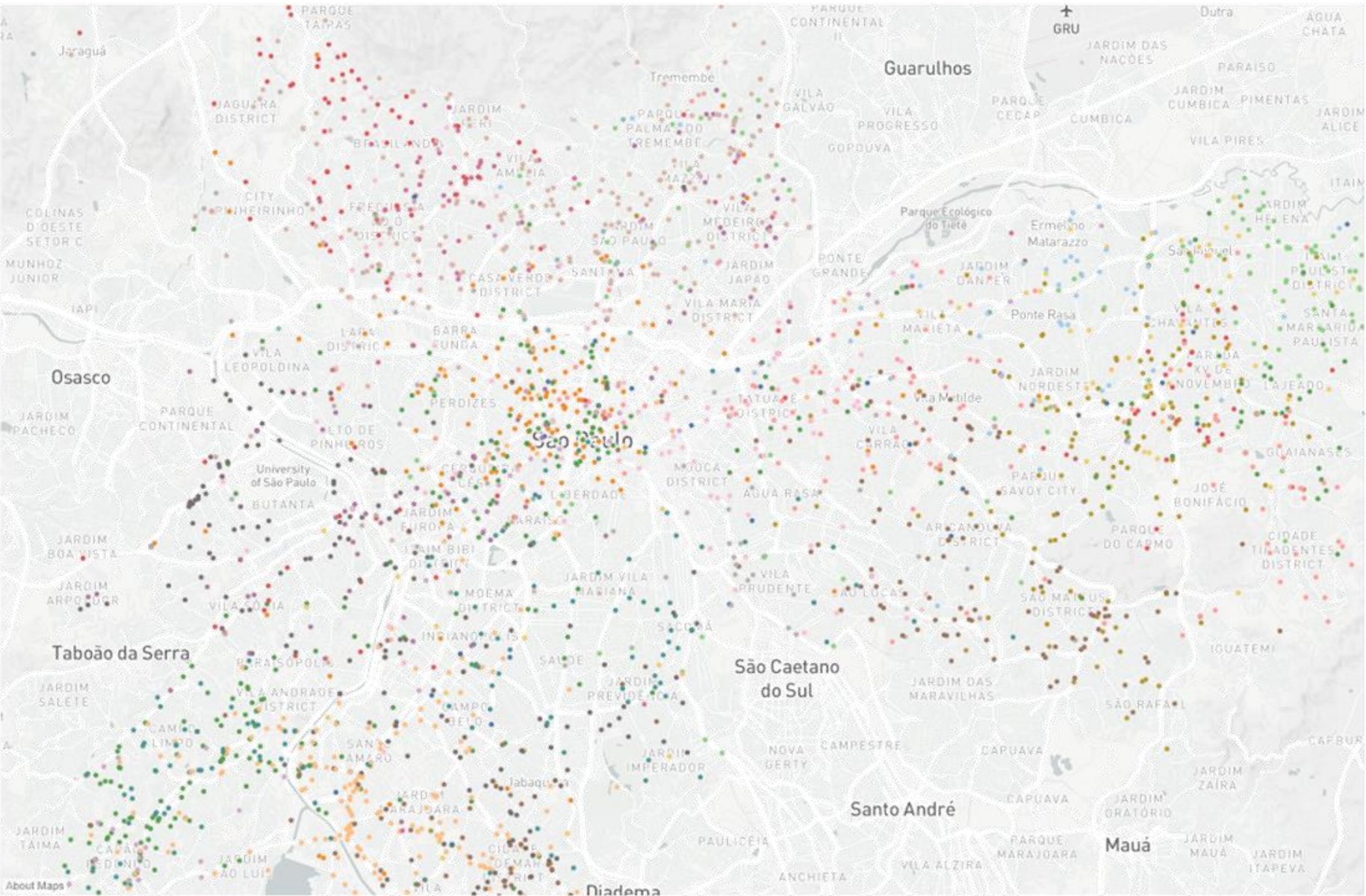
Tipo Veículo (group)



Data Hora, Cet

01/01/2016 00:20:00

Destinos



Finalizacão
(Multiple values)

Tipo Veículo
(Multiple values)

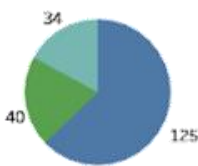
Tipo Vítima

- ☐ (All)
- ☒ Conductor
- ☐ NA
- ☒ Outros
- ☒ Passageiro
- ☒ Pedestre

Destino

- H STA IZABEL - STA CASA ...
- H STA MAGGIORE
- H STA MARCELINA ITAIM ...
- H STA MARCELINA ITAQU ...
- H STA PAULA
- H TAIPAS - GERAL EST
- H TATUAPE - MUN - CARM...
- H UNIVERSITARIO - HU
- H V ALPINA - GERAL EST
- H V N CACHOERINHA - GE...
- H V PENTEADO - GERAL EST
- H VILA MARIA - MUN - JOS...
- H VILLA LOBOS
- HOSPITAL MONTEMAGNO
- HOSPITAL VITORIA
- HSPE - SERV PUBL ESTAD...
- HSPM - DO SERV PUB MU...
- PS AUGUSTO GOMES DE ...
- PS BALNEARIO SAO JOSE ...
- PS BANDEIRANTES-MUN...
- PS BARRA FUNDA- MUN-A...
- PS DONA MARIA ANTONIE...
- PS FREGUESIA DO O - 21 D...
- PS LAPA - MUN - JOAO CA...
- PS SANTANA - MUN - LAU...
- PS SAO MATEUS II - MUN...
- PS STO AMARO - MUN - JO...
- PS V FORMOSA
- SANTA CASA DE SAO PAU...
- SANTA CASA DE STO AMA...
- UPA CAMPO LIMPO
- UPA VILA SANTA CATARINA

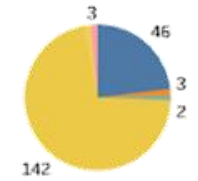
Tipo de Vítimas



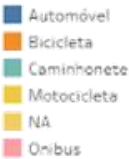
Tipo Vítima



Veículos envolvidos



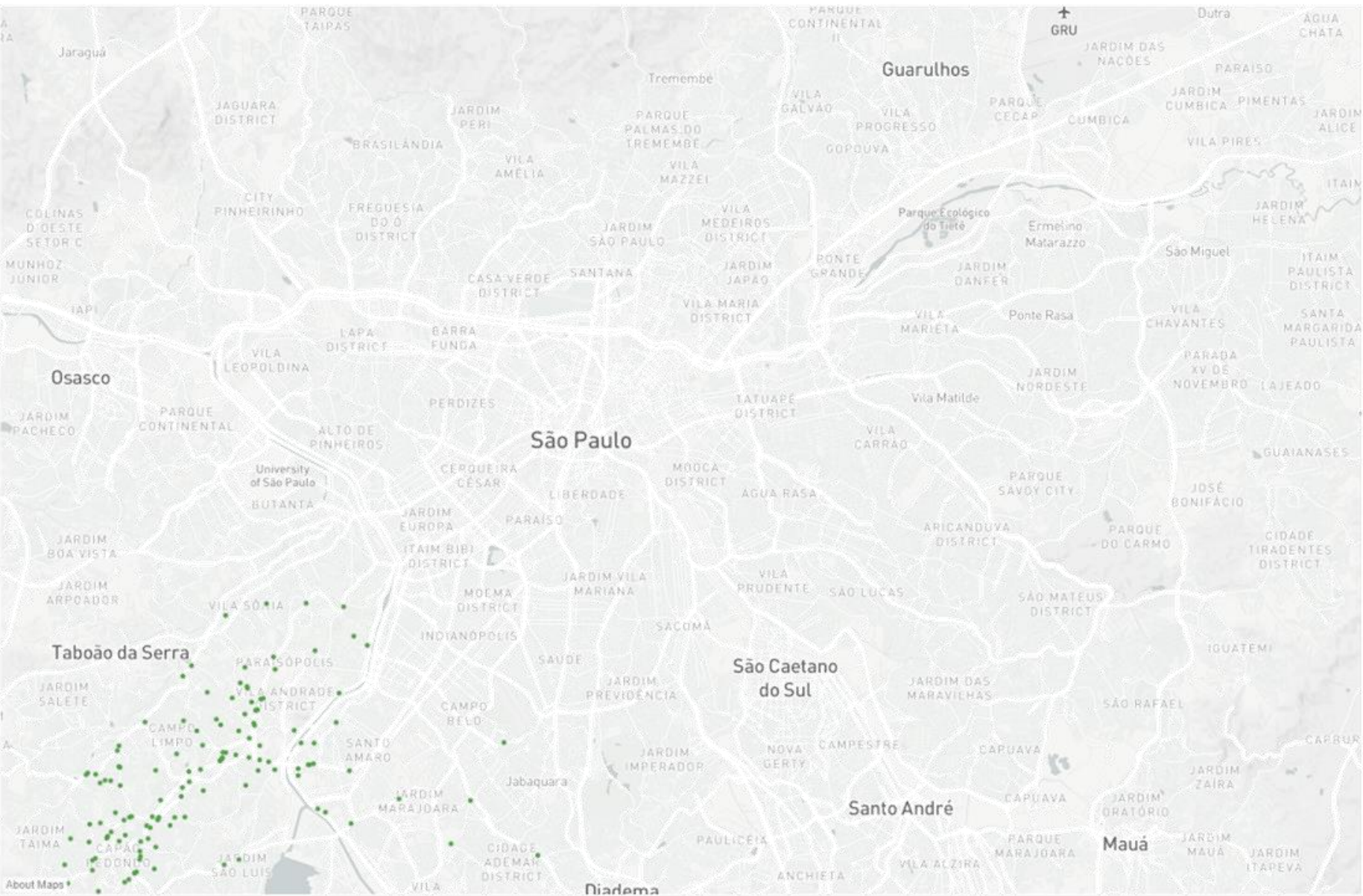
Tipo Veículo (group)



Data Hora.Cet

01/01/2016 00:20:00

Destinos



Finalizacão

(Multiple values)

Tipo Veículo

(Multiple values)

Tipo Vítima

- ☐ (All)
- ☒ Conductor
- ☐ NA
- ☒ Outros
- ☒ Passageiro
- ☒ Pedestre

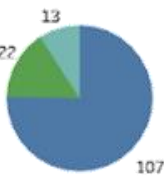
Destino

- H STA IZABEL - STA CASA ..
- H STA MAGGIORE
- H STA MARCELINA ITAIM ..
- H STA MARCELINA ITAQU..
- H STA PAULA
- H TAIPAS - GERAL EST
- H TATUAPE - MUN - CARM..
- H UNIVERSITARIO - HU
- H V ALPINA - GERAL EST
- H V N CACHOERINHA - GE..
- H V PENTEADO - GERAL EST
- H VILA MARIA - MUN - JOS..
- H VILLA LOBOS
- HOSPITAL MONTEMAGNO
- HOSPITAL VITORIA
- HSPE - SERV PUBL ESTAD..
- HSPM - DO SERV PUB MU..
- PS AUGUSTO GOMES DE ..
- PS BALNEARIO SAO JOSE ..
- PS BANDEIRANTES-MUN..
- PS BARRA FUNDA- MUN-A..
- PS DONA MARIA ANTONIE..
- PS FREGUESIA DO O - 21 D..
- PS LAPA - MUN - JOAO CA..
- PS SANTANA - MUN - LAU..
- PS SAO MATEUS II - MUN ..
- PS STO AMARO - MUN - JO..
- PS V FORMOSA
- SANTA CASA DE SAO PAU..
- SANTA CASA DE STO AMA..
- UPA CAMPO LIMPO
- UPA VILA SANTA CATADINA

Destino: UPA CAMPO LIMPO

31/05/2018 16:50:00

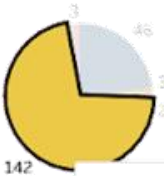
Tipo de Vítimas



Tipo Vítima

- Conductor
- Passageiro
- Pedestre

Veículos envolvidos

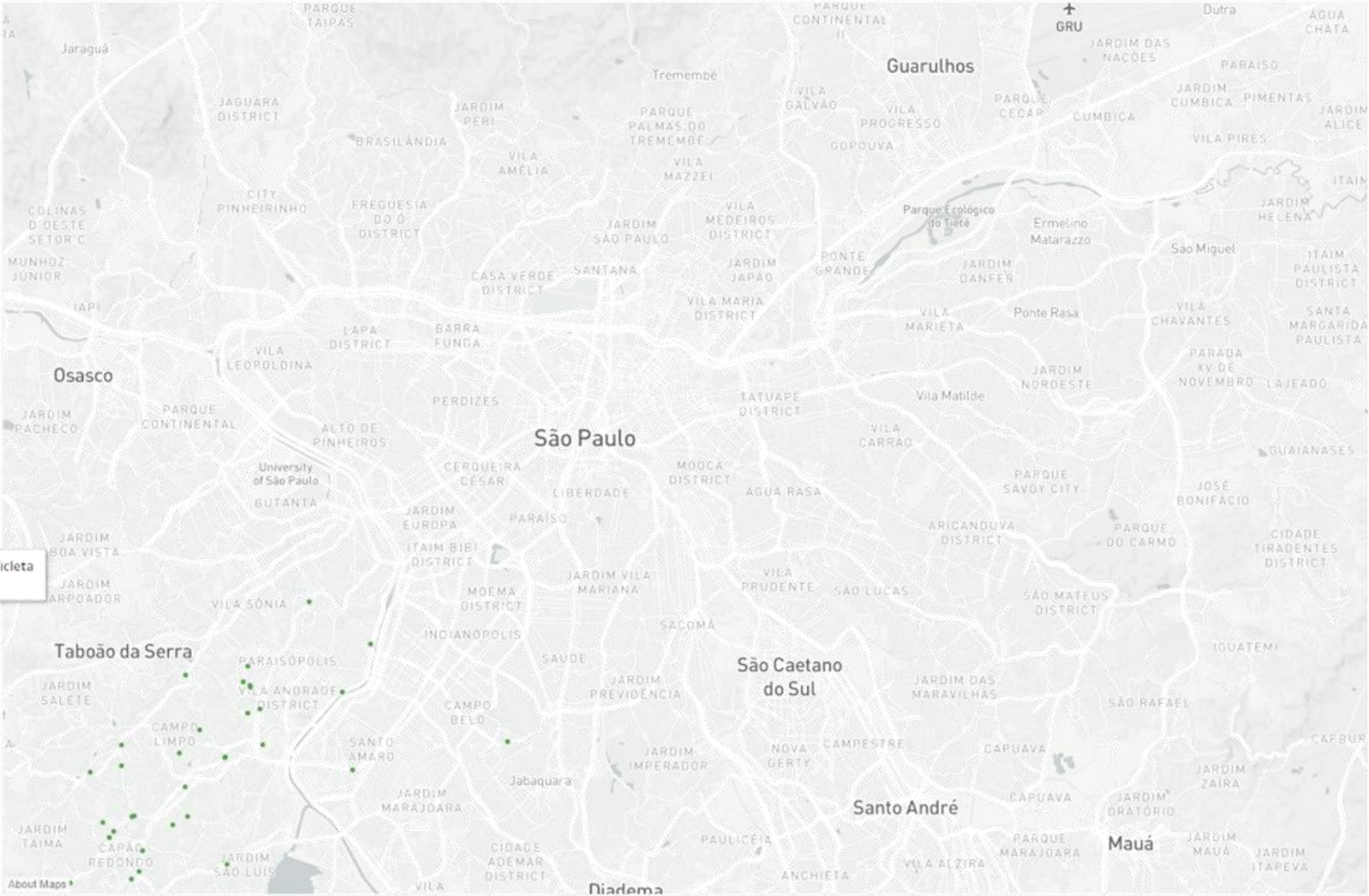


Tipo Veículo (group): **Motocicleta**
Number of Records: **142**

Tipo Veículo (group)

- Automóvel
- Bicicleta
- Caminhonete
- Motocicleta
- NA
- Onibus

Destinos



Finalizacao
(Multiple values)

Tipo Veiculo
(Multiple values)

Tipo Vítima

- ☐ (All)
- ☒ Conductor
- ☐ NA
- ☒ Outros
- ☒ Passageiro
- ☒ Pedestre

Destino

- H SAPOEMBIA - GERAL EST
- H STA CECILIA
- H STA IZABEL - STA CASA
- H STA MAGGIORE
- H STA MARCELINA ITAIM
- H STA MARCELINA ITAQU
- H STA PAULA
- H TAIPAS - GERAL EST
- H TATUAPE - MUN - CARM
- H UNIVERSITARIO - HU
- H V ALPINA - GERAL EST
- H V N CACHOERINHA - GE
- H V PENTEADO - GERAL EST
- H VILA MARIA - MUN - JOS
- H VILLA LOBOS
- HOSPITAL MONTEMAGNO
- HOSPITAL VITORIA
- HSPE - SERV PUBL ESTAD
- HSPM - DO SERV PUB MU
- PS AUGUSTO GOMES DE
- PS BALNEARIO SAO JOSE
- PS BANDEIRANTES-MUN
- PS BARRA FUNDA - MUN-A
- PS DONA MARIA ANTONIE
- PS FREGUESIA DO O - 21 D
- PS LAPA - MUN - JOAO CA
- PS SANTANA - MUN - LAU
- PS STO AMARO - MUN - JO
- SANTA CASA DE SAO PAU
- SANTA CASA DE STO AMA
- UPA CAMPO LIMPO**
- UPA VILA SANTA CATARINA

Current status

- **Advancements on matching**

- More recent and complete datasets: 2016 – 2018
- Improving understanding and coverage on hospital data
- Beginning application development to implement a sustainable procedure on city government

- **Institutional arrangements**

- High city-level engagement
- Lack of collaboration with state-level road safety initiative
- Local data integration solution under development by federal govt., focused on small and medium-sized cities.

Thank you

Rafael Godoy Bueno da Silva

Data Coordinator | Coordenador de Dados

Bloomberg Initiative for Global Road Safety

[+55 11 99510 8754](tel:+5511995108754)

rafaelgodoy@bigrs.org



**Bloomberg
Philanthropies**

INITIATIVE FOR GLOBAL ROAD SAFETY