

08 DE FEBRERO 2021

# ACCIDENTES VEHICULARES EN CDMX



# Integrantes

---

- Jesus Antonio Hernández Aguilera
- Angel Uriel Meléndez Rivera
- Adalberto Benitez Zapata
- Sergio Maldonado Rodriguez

# Introducción

---

De acuerdo con un reporte de la Organización Mundial de la Salud en 2017, los accidentes viales son causantes de más de 1.3 de millones de muertes al año, y un estimado de alrededor de 50 millones de personas con traumatismos no mortales siendo así la quinta causa de los fallecimientos mundiales anuales.

Por su parte México y de acuerdo con los datos del Instituto Nacional de Salud Pública (INSP), el país se encuentra en el séptimo lugar a nivel mundial en el rubro de siniestros automovilísticos mortales, y ocupa el tercer lugar en la región de Latinoamérica en este mismo rubro

# Justificación

---

Bajo este tenor, y con el fin de atenuar la situación descrita anteriormente, resulta útil identificar cuáles son los factores que tienen mayor influencia en los choques automovilísticos, cuáles son los perfiles más propensos a generar un choque, y simultáneamente diseñar estrategias que permitan atacar las causas raíz de los siniestros automovilísticos en México.

# Información en los datasets

- Estado
- Municipio
- Fecha
- Tipo de accidente
- Tipo de vehículo
- Causa presunta del accidente
- Sexo
- Aliento alcohólico
- Cinturón
- Edad
- Clase de accidente

# Realización del proyecto



# Importación del dataset

```
import pandas as pd
url="https://www.inegi.org.mx/contenidos/programas/accidentes/datosabiertos/atus_anual_csv.zip"
file_name="atus_anual_1997_2019"
```

```
def descarga(url,file_name,n_file=0):
    import requests, zipfile, io,os.path
    if not os.path.exists(file_name):
        r = requests.get(url)
        z = zipfile.ZipFile(io.BytesIO(r.content))
        if n_file==0:
            z.extractall("")
        else:
            z.extractall(file_name)
    return print("Descarga de archivos completa")
```

# Filtración del dataset

```
df_cdmx = df[df.ID_ENTIDAD==9]
df_cdmx = df_cdmx.reset_index(drop=True)
```

```
df_cdmx.shape
```

```
(136812, 45)
```

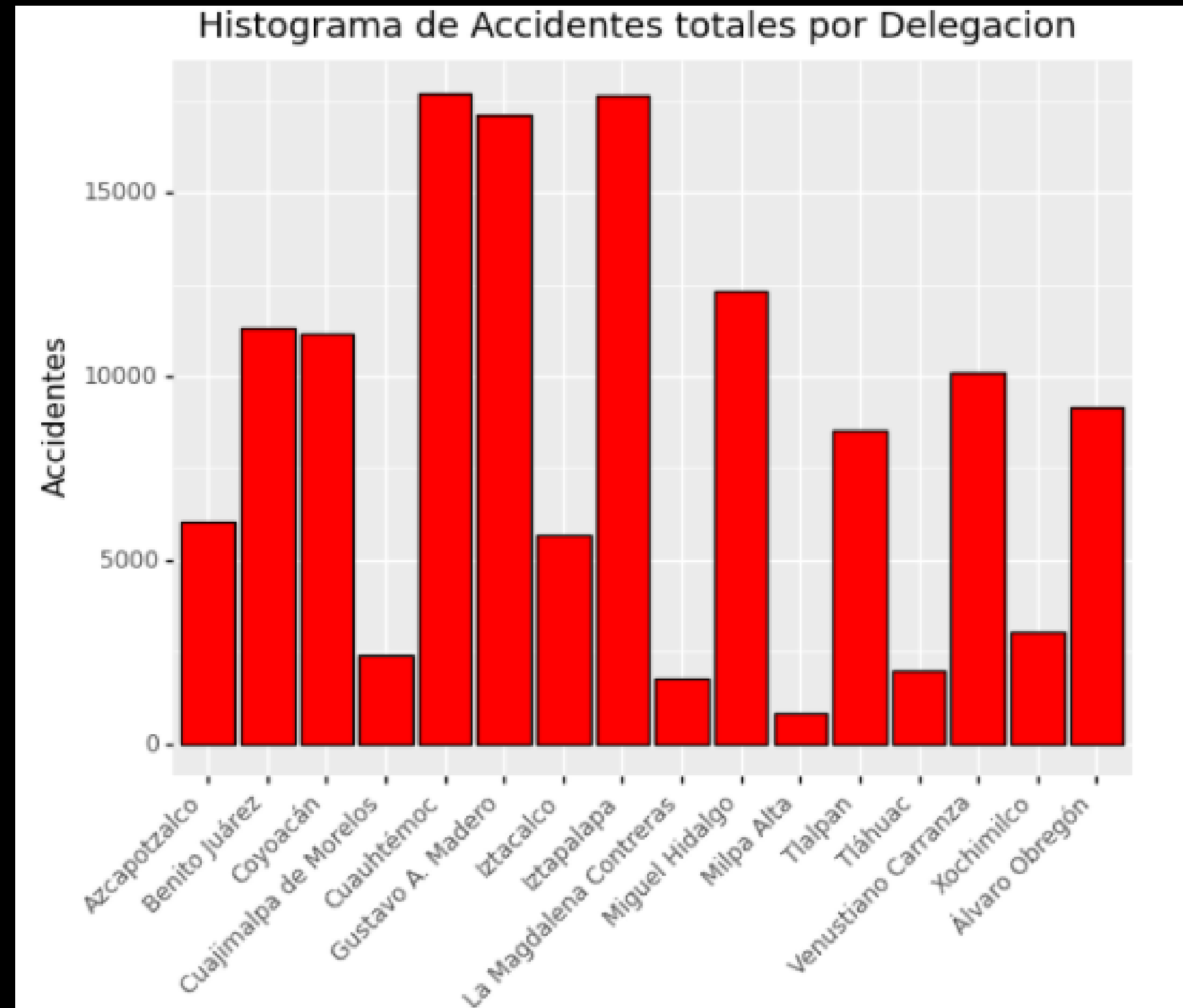
```
delegacion_nombre=delegacion_nombre[(delegacion_nombre.ID_ENTIDAD==9)&(delegacion_nombre.ID_MUNICIPIO!=9)]
delegacion_nombre.head()
```

	ID_ENTIDAD	ID_MUNICIPIO	NOM_MUNICIPIO
278	9	2	Azcapotzalco
279	9	3	Coyoacán
280	9	4	Cuajimalpa de Morelos
281	9	5	Gustavo A. Madero
282	9	6	Iztacalco

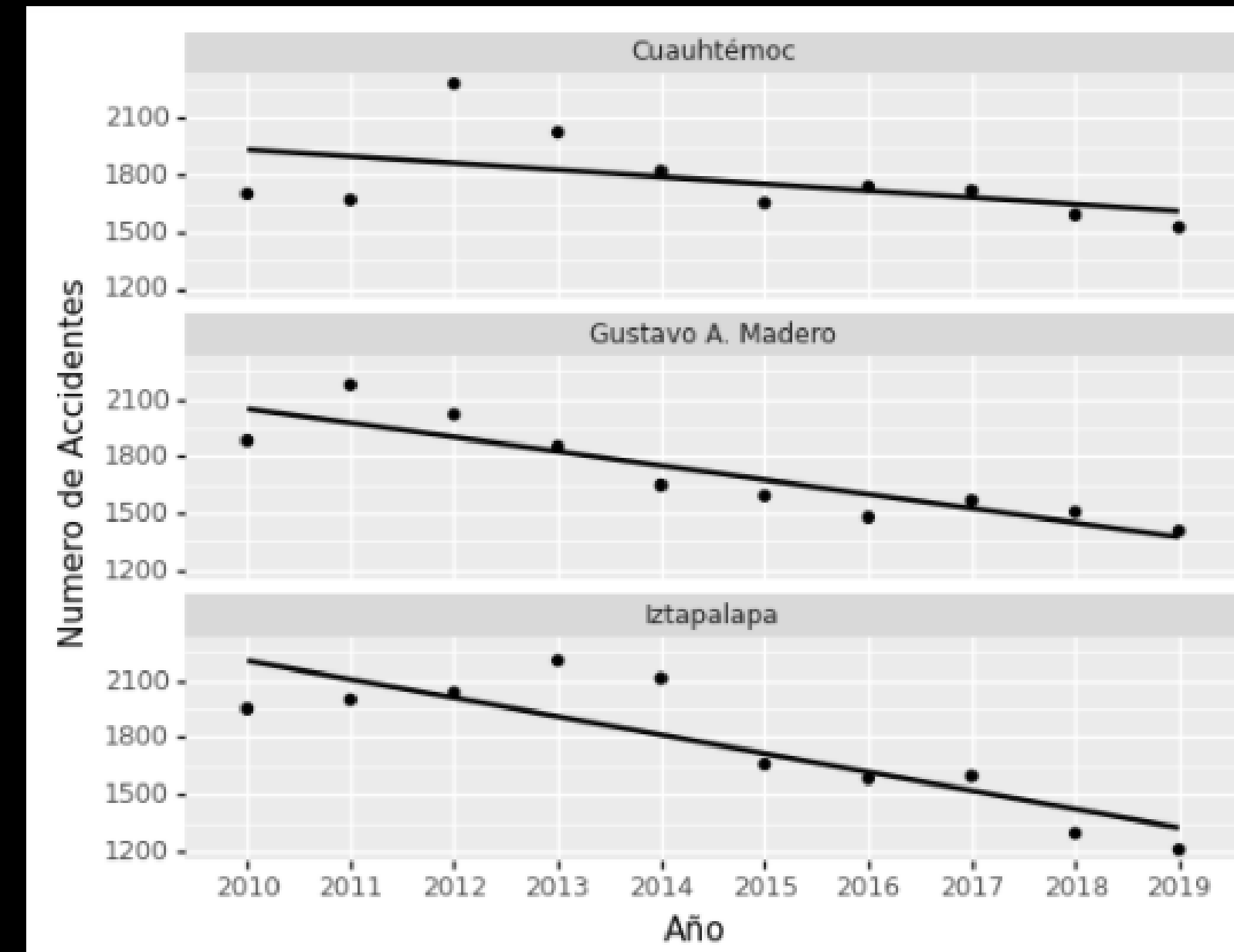
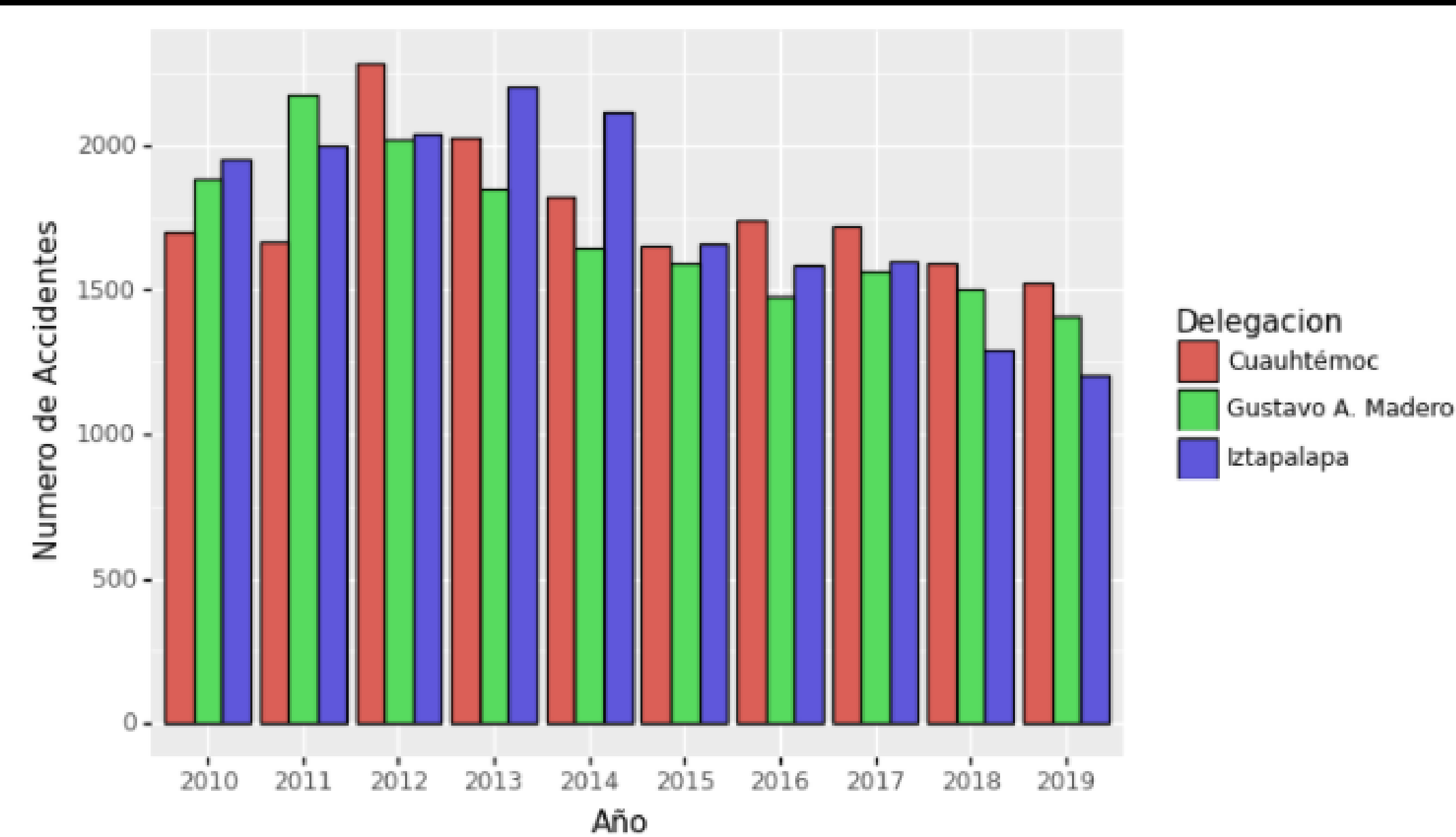


# ¿Que delegacion/municipio resgistró mas accidentes?

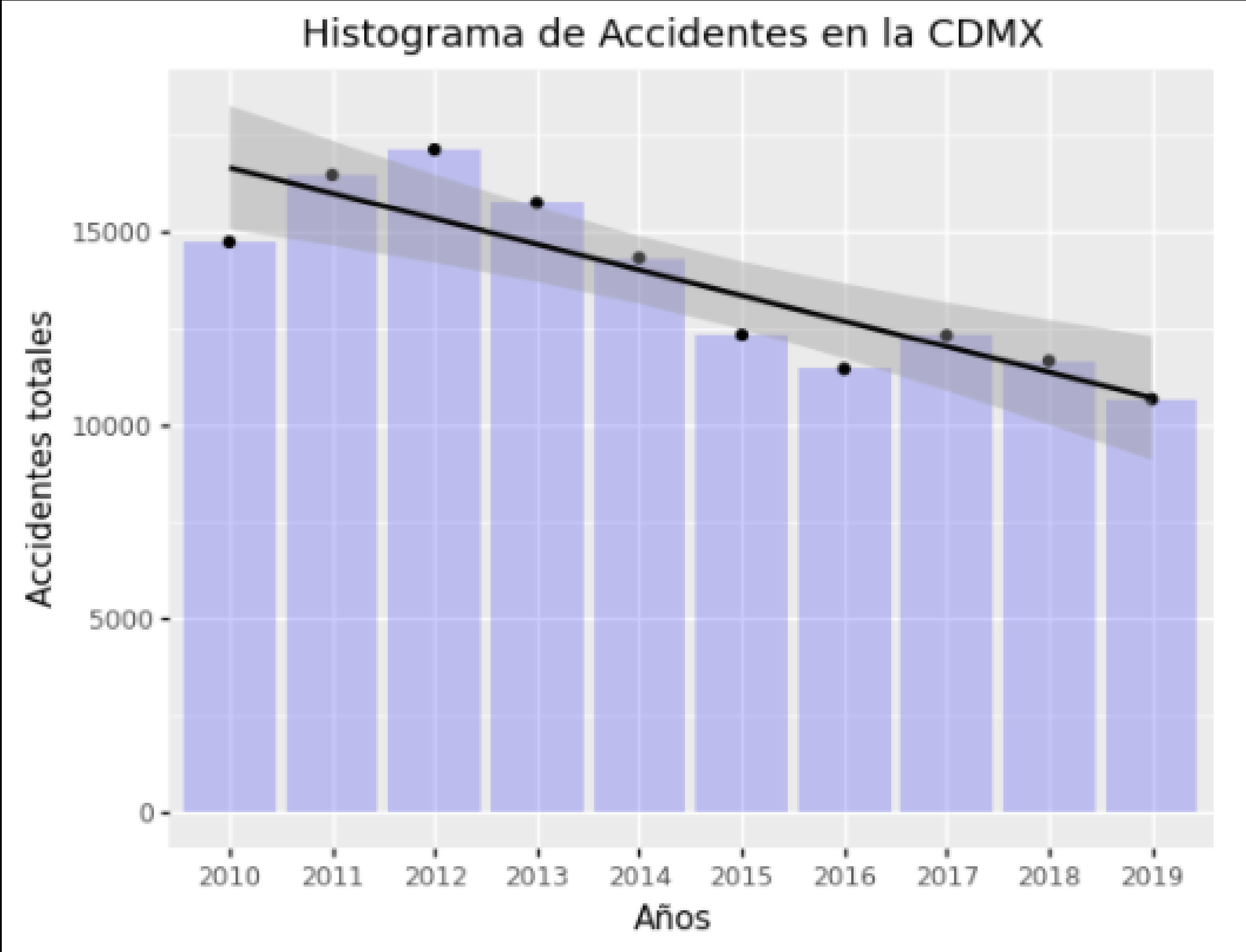
	NOM_MUNICIPIO	ANIO	Num_Acc
0	Cuauhtémoc	2010	1701
1	Cuauhtémoc	2011	1669
2	Cuauhtémoc	2012	2283
3	Cuauhtémoc	2013	2025
4	Cuauhtémoc	2014	1820
5	Cuauhtémoc	2015	1652
6	Cuauhtémoc	2016	1740
7	Cuauhtémoc	2017	1719
8	Cuauhtémoc	2018	1590
9	Cuauhtémoc	2019	1524
10	Iztapalapa	2010	1952
11	Iztapalapa	2011	1999
12	Iztapalapa	2012	2037
13	Iztapalapa	2013	2208
14	Iztapalapa	2014	2113
15	Iztapalapa	2015	1659
16	Iztapalapa	2016	1583
17	Iztapalapa	2017	1596
18	Iztapalapa	2018	1295
19	Iztapalapa	2019	1207



# ¿Cuales son las tres delegaciones con mas accidentes y como es su tendencia?



# ¿En que año se han registrado mas accidentes en la CDMX y cual es su tendencia?

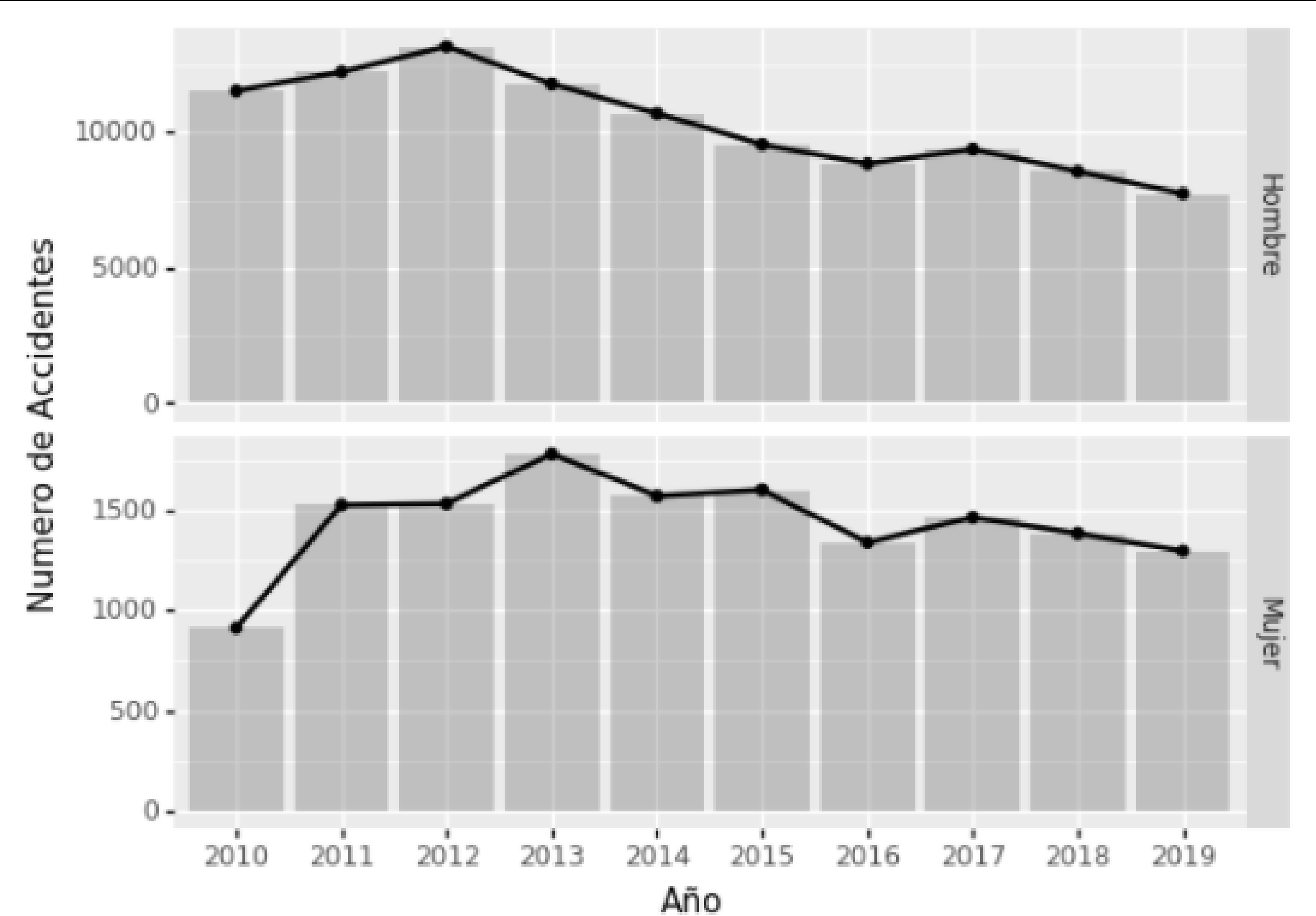


```
años_accidentes.corr(method="pearson")
```

	ANIO	Num_Acc
ANIO	1.000000	-0.874329
Num_Acc	-0.874329	1.000000

	ANIO	Num_Acc
0	2010	14729
1	2011	16466
2	2012	17120
3	2013	15742
4	2014	14319
5	2015	12337
6	2016	11449
7	2017	12321
8	2018	11656
9	2019	10873

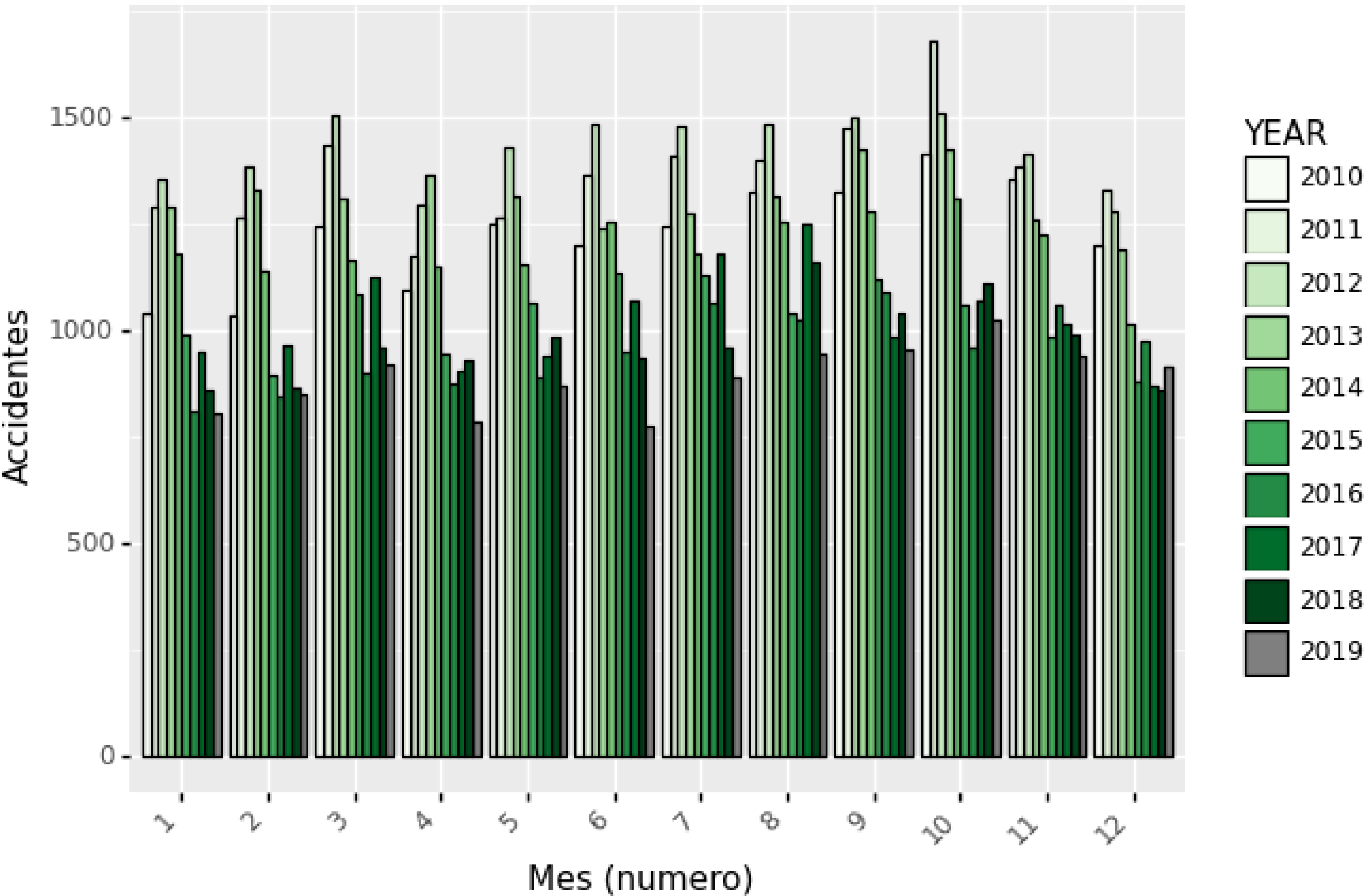
# ¿Quiénes son mas propensos a sufrir accidentes, los hombres o las mujeres y como es la tendencia?



	SEXO	ANIO	Num_Acc
0	Hombre	2012	13158
1	Hombre	2011	12228
2	Hombre	2013	11778
3	Hombre	2010	11518
4	Hombre	2014	10691
5	Hombre	2015	9546
6	Hombre	2017	9375
7	Hombre	2016	8829
8	Hombre	2018	8546
9	Hombre	2019	7735
10	Mujer	2013	1782
11	Mujer	2015	1602
12	Mujer	2014	1572
13	Mujer	2012	1535
14	Mujer	2011	1528
15	Mujer	2017	1465
16	Mujer	2018	1385
17	Mujer	2016	1340
18	Mujer	2019	1298
19	Mujer	2010	913

# ¿Cuáles son los meses con mayores accidentes a los largo de los años?

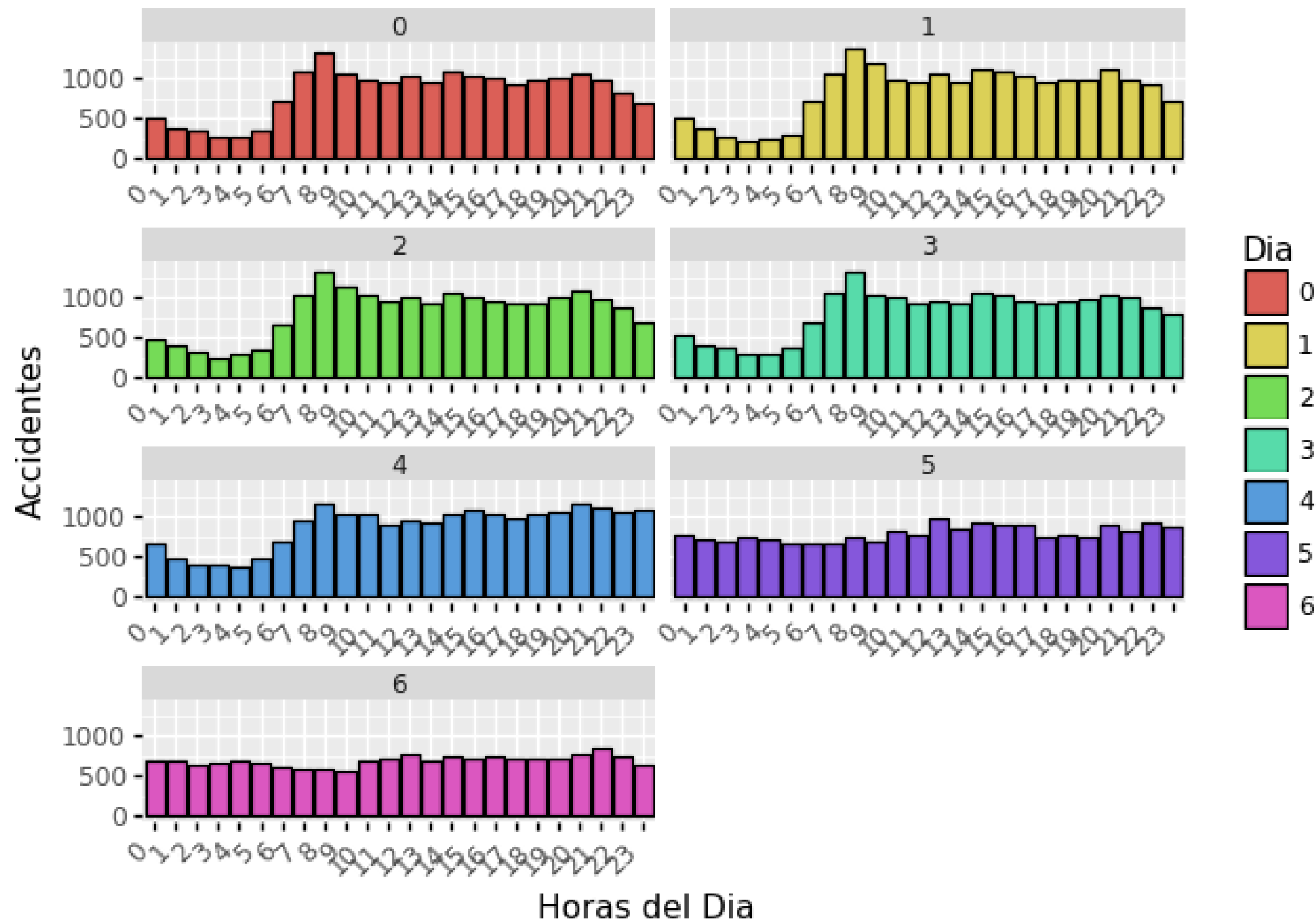
Histograma de Accidentes por mes y año



	Max_acci	MES
ANIO		
2019	1027.0	Octubre
2018	1160.0	Agosto
2017	1249.0	Agosto
2016	1089.0	Septiembre
2015	1135.0	Junio
2014	1311.0	Octubre
2013	1426.0	Octubre
2012	1510.0	Octubre
2011	1679.0	Octubre
2010	1414.0	Octubre

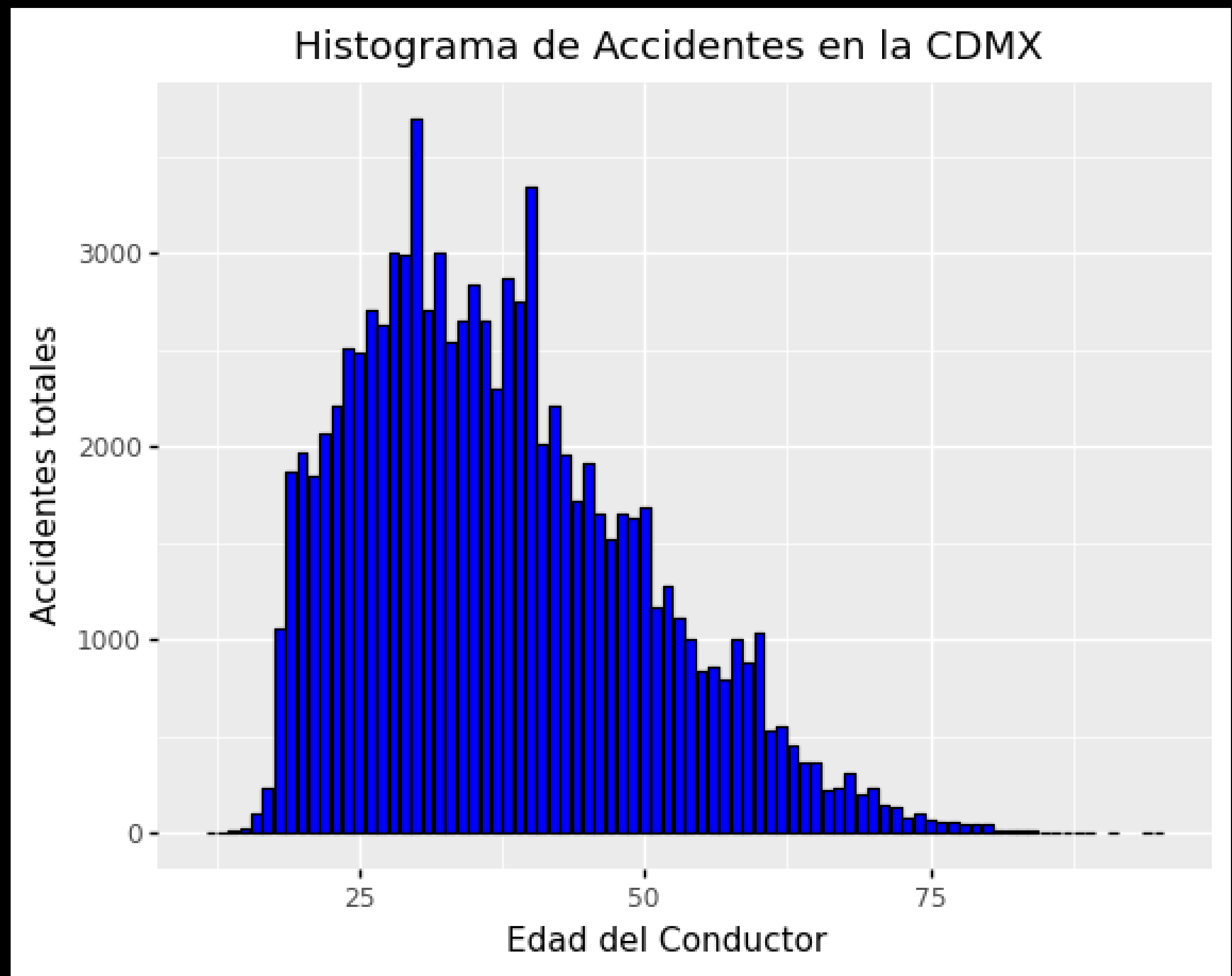
# ¿Cuáles son los días de la semana con mayores accidentes a los largo de los años?

Histograma de Accidentes por Dia y hora del 2010-2019



	Max_acci	DIAS
ANIO		
2019	1612.0	Viernes
2018	1807.0	Viernes
2017	1909.0	Viernes
2016	1740.0	Viernes
2015	1939.0	Jueves
2014	2292.0	Viernes
2013	2571.0	Viernes
2012	2620.0	Viernes
2011	2523.0	Viernes
2010	2265.0	Viernes

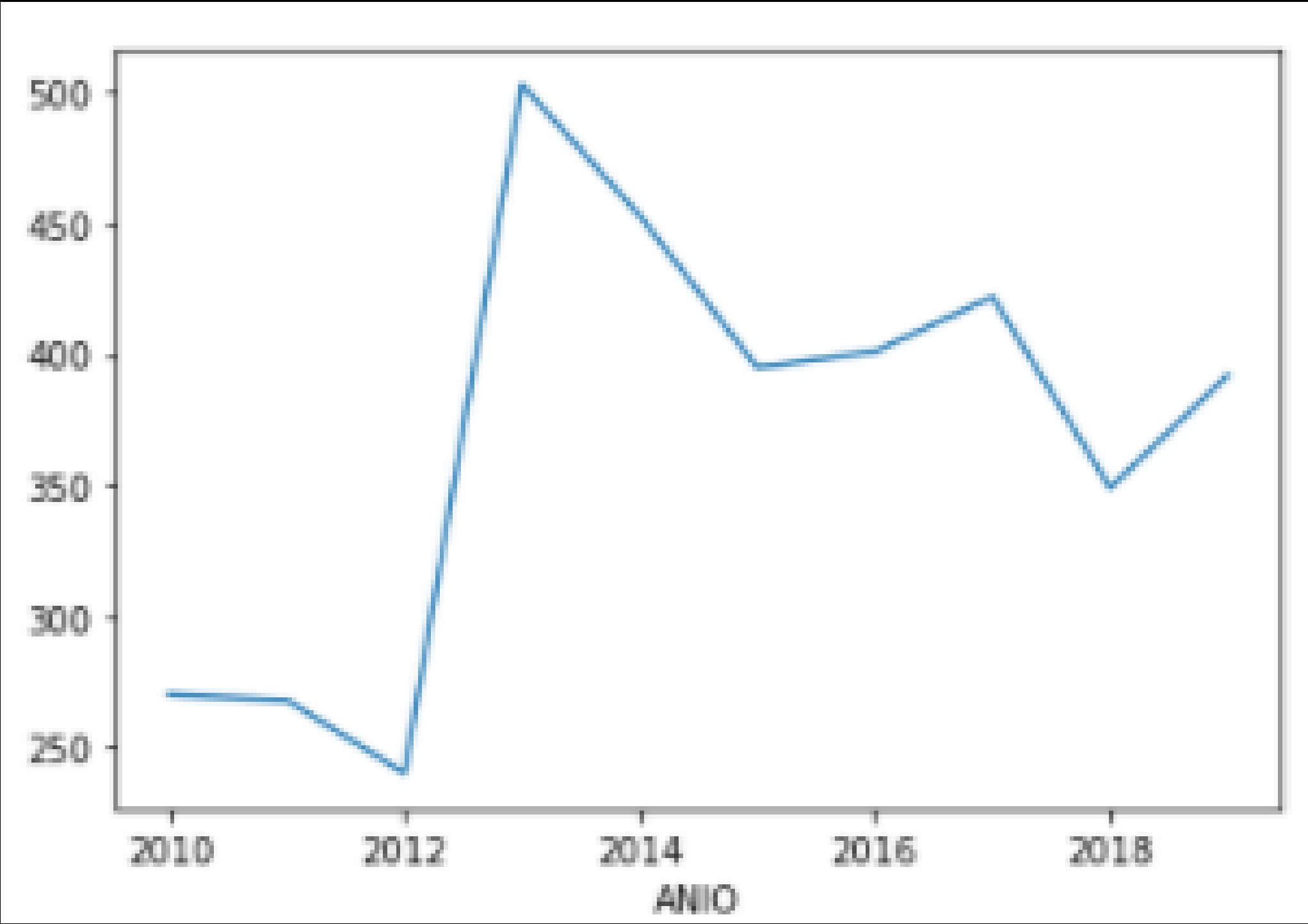
# ¿Cual es la edad mas propensa a sufrir accidentes?



ID_EDAD	
30	3693
40	3340
28	3000
32	2998
29	2989

# ¿Cual es la Tedencia de las edades mas propensas a sufrir accidentes del 2010 al 2019?

Para los 30 años



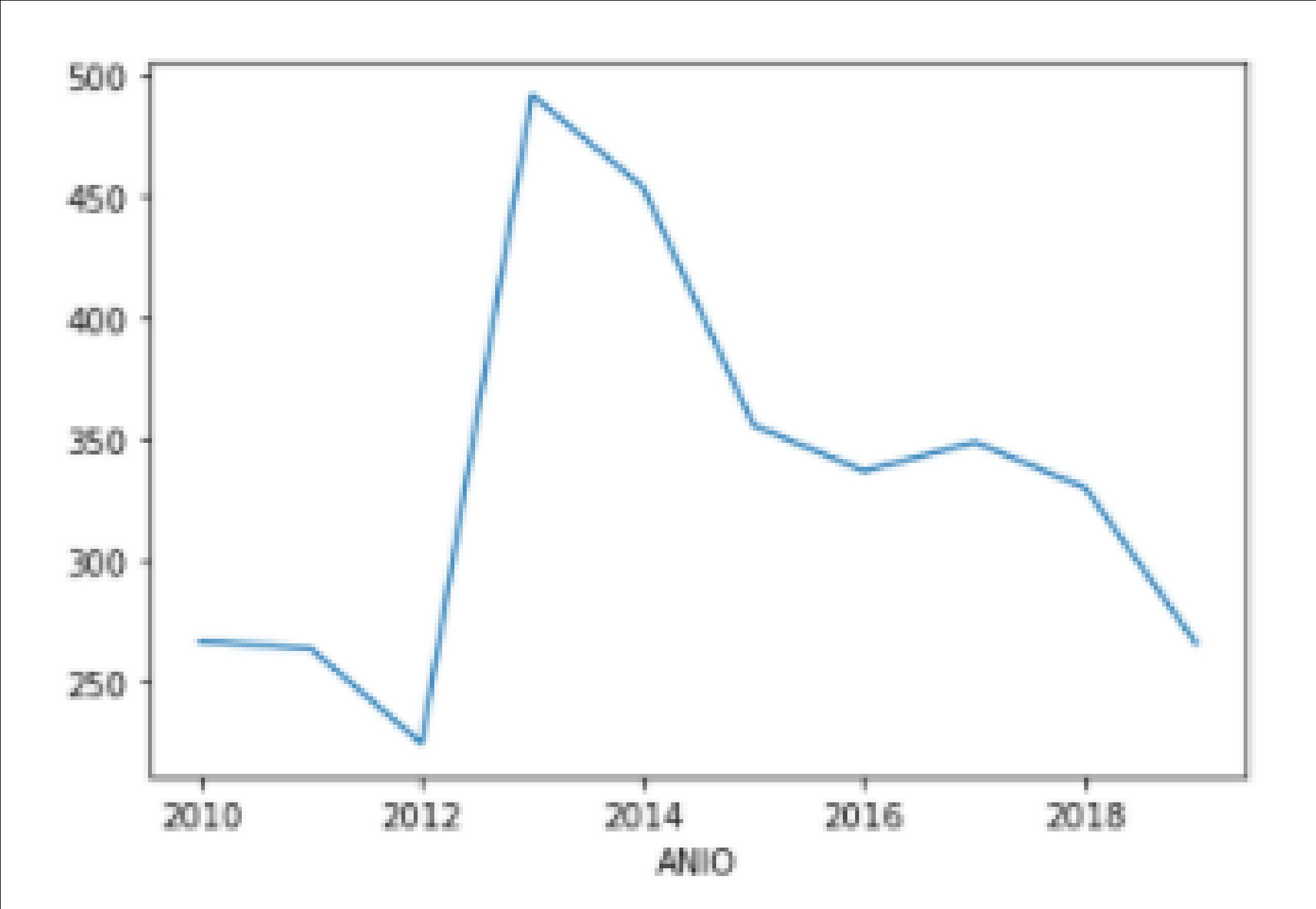
	ANIO	NUM_ACC
ANIO	1.00000	0.48989
NUM_ACC	0.48989	1.00000

ANIO	
2010	270
2011	268
2012	240
2013	503
2014	453
2015	395
2016	401
2017	422
2018	349
2019	392



# ¿Cual es la Tedencia de las edades mas propensas a sufrir accidentes del 2010 al 2019?

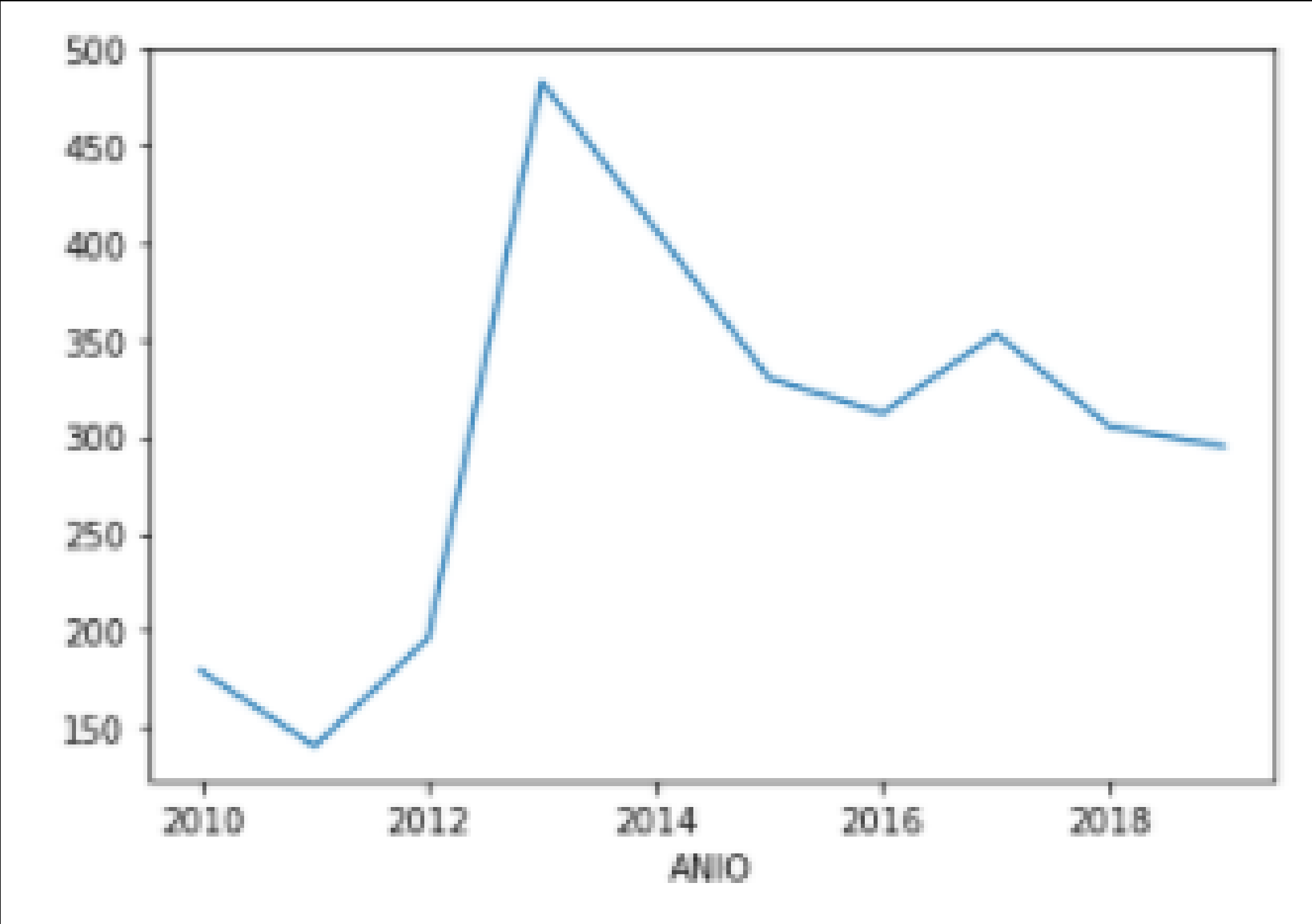
Para los 40 años



	ANIO	NUM_ACC
ANIO	1.000000	0.109287
NUM_ACC	0.109287	1.000000

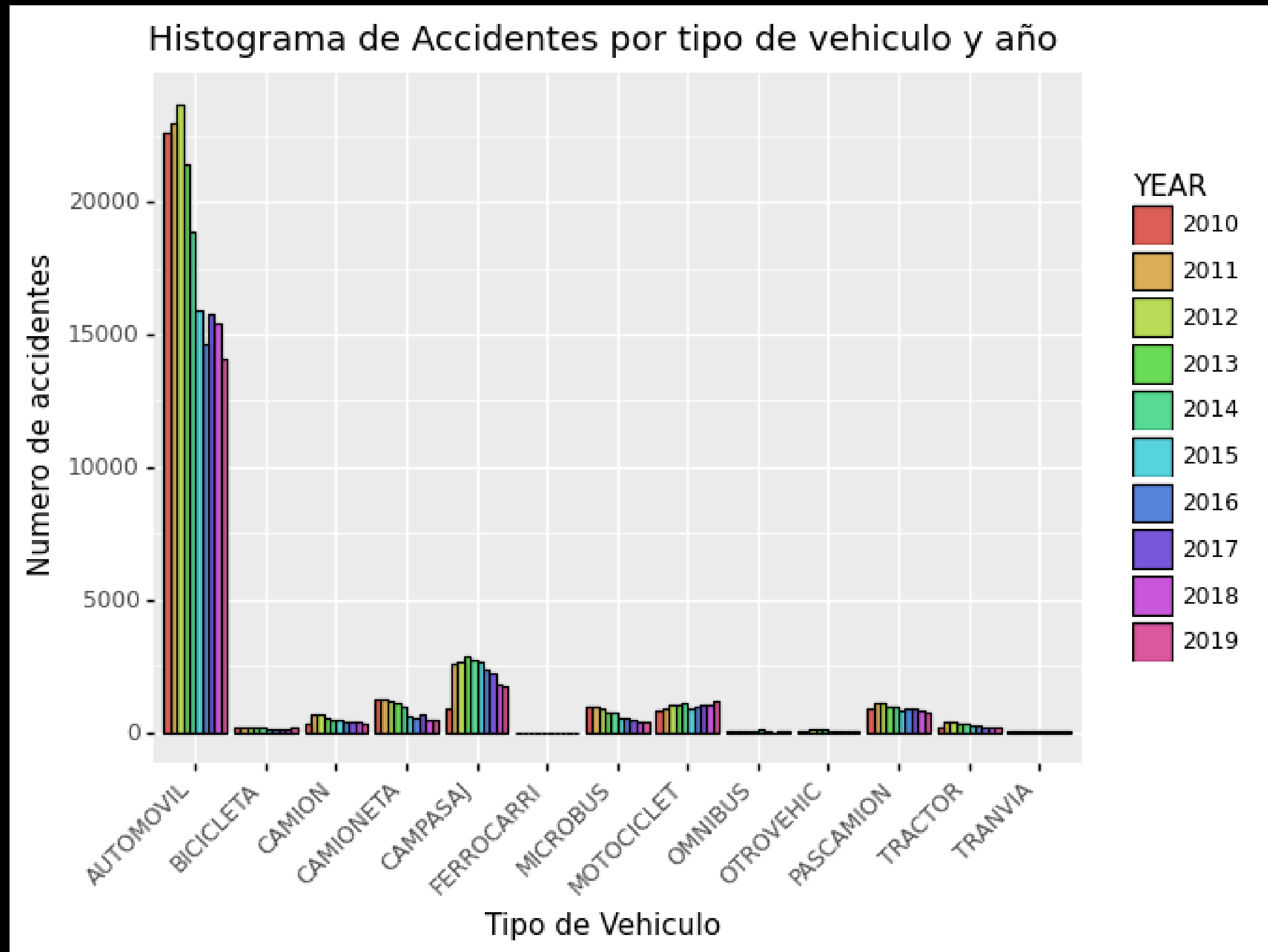
# ¿Cual es la Tedencia de las edades mas propensas a sufrir accidentes del 2010 al 2019?

Para los 28 años

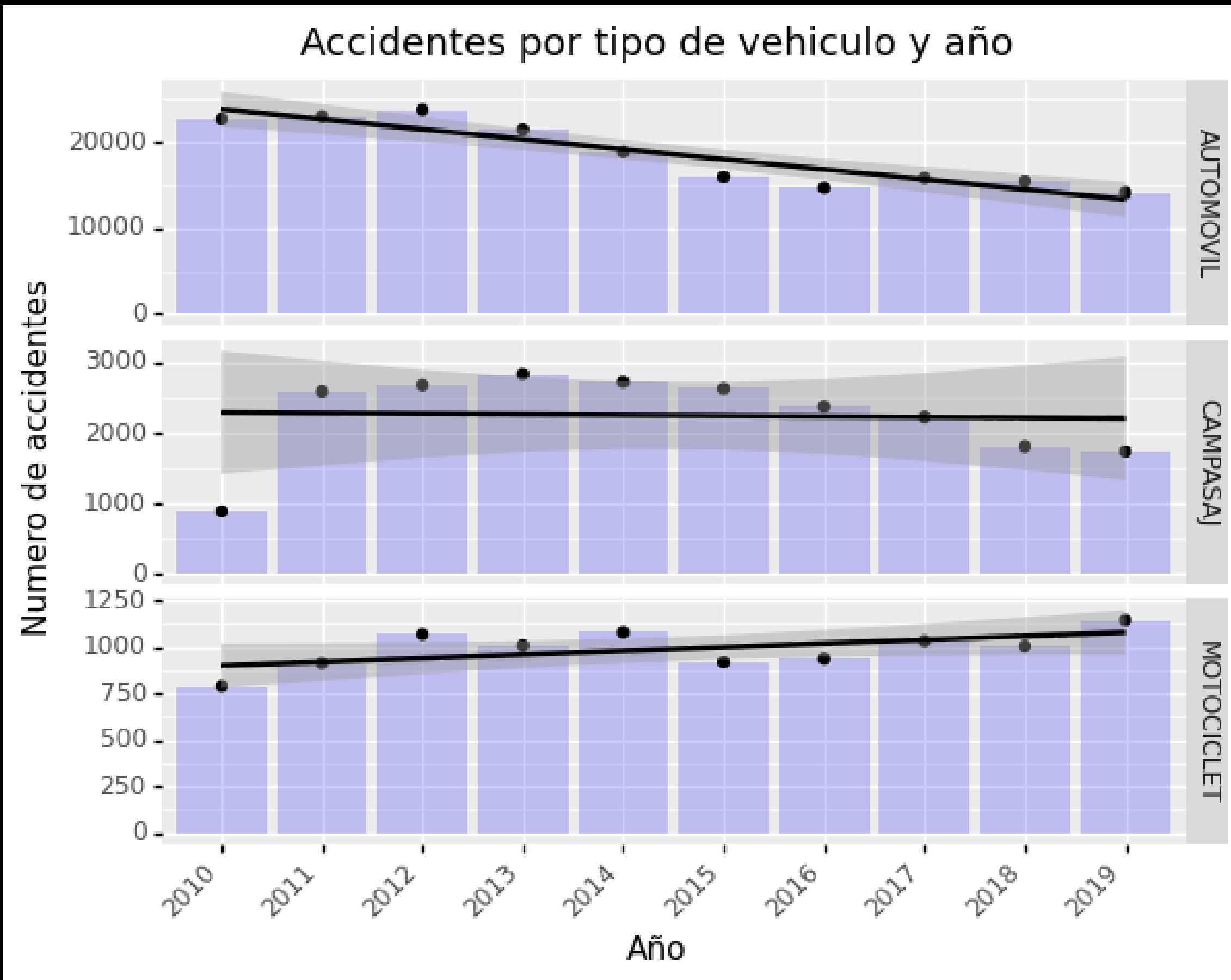


	ANIO	NUM_ACC
ANIO	1.000000	0.418855
NUM_ACC	0.418855	1.000000

# ¿Que vehiculo es el que causa mas accidentes?



# ¿Que vehiculo es el que causa mas accidentes?



# Problemas al realizar el proyecto

## FALTA DE DATOS Y SESGO

---

Dificultad al momento de establecer de manera correcta la relación entre genero y accidentes cometidos. Debido a falta de información obtenida.

## INFORMACIÓN COMPLICADA DE OBTENER

---

No se logró establecer de manera correcta el factor de aliento alcohólico debido a la necesidad de mas información para probar correctamente la hipótesis establecida

## CAPTURA DE DATOS DEFICIENTE

---

Problema en el agrupamiento de los datos debido a una captura deficiente de la información contenida en los datasets.

## SESGO EN LOS DATASETS

---

un sesgo en la información referente a los automóviles y motocicletas ocasionado por falta de datos.

# Proyección del Análisis Inferencial



## ANALISIS DE FACTORES SOBRE CHOQUES.

---

para esto será necesario llevar a cabo análisis de regresión multivariable en los cuáles se puedan evaluar los coeficientes, así como las interacciones de estos factores. Dentro de las herramientas estadísticas que podremos utilizar para este análisis destacan ANOVA, K - Means, Random Forests, entre otros.

## ANÁLISIS COMPARATIVO CON OTROS ARCHIVOS DE DATOS

---

parque vehicular, la cantidad de hombres y mujeres registrados, persona registradas por delegación, personas registradas por grupos de edad, entre otros. Los cuáles serán eficaces para eliminar ciertos tipos de sesgo que por ahora se tienen y que impiden un análisis más profundo de los factores determinantes para la generación de un choque.

## PREDICCIÓN DE CHOQUES

---

efectuar métodos como las series de tiempo. asimismo y complementando lo anterior estas predicciones podrán ayudar a también generar posibles análisis econométricos de la situación.

# Información adicional del proyecto

---

REPOSITORIO

[https://github.com/Bedu-Equipo13/Proyecto\\_Python\\_Accidentes](https://github.com/Bedu-Equipo13/Proyecto_Python_Accidentes)

