

# Análisis de los Sets de Datos

Luciano Hernández

2025-12-30

## Contents

|  |          |
|--|----------|
| <b>Extracción y caracterización de datos</b>           | <b>4</b> |
| Siniestros buses interurbanos(1).xlsx . . . . .        | 4        |
| Granularidad temporal . . . . .                        | 4        |
| Distribución Geográfica . . . . .                      | 6        |
| Características de los Siniestros . . . . .            | 7        |
| Factores Causales . . . . .                            | 8        |
| Consecuencias y Severidad . . . . .                    | 10       |
| Optimizaciones Realizadas . . . . .                    | 14       |
| GH026929.xlsx . . . . .                                | 14       |
| Características Temporales . . . . .                   | 14       |
| Granularidad temporal . . . . .                        | 15       |
| Parámetros de Velocidad . . . . .                      | 16       |
| Calidad de Señal GPS . . . . .                         | 17       |
| Aspectos Requieren Clarificación . . . . .             | 18       |
| Procesamiento de Datos Realizado . . . . .             | 18       |
| DTPR . . . . .   | 19       |
| Metadatos de Procesamiento . . . . .                   | 19       |
| Período de Monitoreo . . . . .                         | 19       |
| Granularidad temporal (separado por EVT_PPU) . . . . . | 19       |
| Características Operacionales del Sistema . . . . .    | 20       |
| Procesamiento y Normalización de Datos . . . . .       | 25       |
| Análisis por Unidad Operativa . . . . .                | 26       |
| Data accidentes de carabineros.xlsx . . . . .          | 26       |
| Período de Análisis . . . . .                          | 26       |
| Granularidad Temporal . . . . .                        | 27       |
| Distribución Geográfica . . . . .                      | 27       |

|  |    |
|--|----|
| Clasificación de Accidentes . . . . .                    | 31 |
| Procesamiento y Normalización de Datos . . . . .         | 33 |
| Incidentes de tráfico radio.xlsx . . . . .               | 33 |
| Período de Registro . . . . .                            | 34 |
| Granularidad Temporal . . . . .                          | 34 |
| Clasificación de Incidentes . . . . .                    | 34 |
| Tipología de Incidentes . . . . .                        | 35 |
| Direccionalidad del Tráfico Afectado . . . . .           | 36 |
| Procesamiento y Normalización de Datos . . . . .         | 37 |
| Med velo CHIGUAYANTE.xlsx . . . . .                      | 38 |
| Contexto Operacional . . . . .                           | 38 |
| Granularidad Temporal . . . . .                          | 39 |
| Análisis de Velocidades . . . . .                        | 39 |
| Composición Vehicular . . . . .                          | 40 |
| Patrones de Movimiento . . . . .                         | 41 |
| Comportamientos Específicos . . . . .                    | 42 |
| Procesamiento y Normalización de Datos . . . . .         | 43 |
| Med velo LA VEGA.xlsx . . . . .                          | 44 |
| Contexto Operacional . . . . .                           | 44 |
| Granularidad Temporal . . . . .                          | 44 |
| Análisis de Velocidades . . . . .                        | 44 |
| Patrones de Movimiento . . . . .                         | 45 |
| Fenómenos de Congestión Documentados . . . . .           | 46 |
| Procesamiento y Normalización de Datos . . . . .         | 46 |
| tb_gps_historial_eventos_202509161626(1).csv . . . . .   | 47 |
| Período de Monitoreo . . . . .                           | 47 |
| Granularidad Temporal . . . . .                          | 47 |
| Distribución de Estados Vehiculares . . . . .            | 48 |
| Patrones de Direccionalidad . . . . .                    | 48 |
| Aspectos Requieren Investigación . . . . .               | 49 |
| Procesamiento y Normalización de Datos . . . . .         | 49 |
| Inventario CCTV Biobío(1).xlsx . . . . .                 | 50 |
| Distribución Geográfica . . . . .                        | 50 |
| Estado Operacional del Sistema . . . . .                 | 50 |
| Características Técnicas de la Infraestructura . . . . . | 50 |
| Evolución Temporal del Sistema . . . . .                 | 52 |

|   |           |
|---|-----------|
| Procesamiento y Normalización de Datos . . . . .  | 53        |
| Alertas de Tráfico.csv . . . . .  | 53        |
| Período de Análisis . . . . .   | 53        |
| Granularidad Temporal . . . . .   | 54        |
| Contexto Geográfico . . . . .   | 54        |
| Clasificación de Alertas . . . . .  | 55        |
| Procesamiento y Normalización de Datos . . . . .  | 56        |
| Copia de Accidentes.csv . . . . .   | 56        |
| Período de Monitoreo . . . . .  | 57        |
| Granularidad Temporal . . . . .   | 57        |
| Contexto Geográfico . . . . .   | 57        |
| Métricas de Confiabilidad . . . . .   | 58        |
| Procesamiento y Normalización de Datos . . . . .  | 59        |
| Waze for Cities Data Key Alerts Dashboard_Traffic Irregularities_Tabla(1).csv . . . . . | 59        |
| Período de Análisis . . . . .   | 59        |
| Granularidad Temporal . . . . .   | 60        |
| Distribución Geográfica . . . . .   | 60        |
| Clasificación por Causas Identificadas . . . . .  | 61        |
| Métricas de Impacto de Tráfico . . . . .  | 62        |
| Procesamiento y Normalización de Datos . . . . .  | 65        |
| Red de Waze . . . . .   | 65        |
| Metadatos de Procesamiento . . . . .  | 65        |
| Período de Monitoreo . . . . .  | 66        |
| Granularidad Temporal . . . . .   | 66        |
| Características del Dataset Consolidado . . . . .                                       | 66        |
| Procesamiento y Normalización de Datos . . . . .  | 73        |
| <b>Discretización de datos</b>  | <b>74</b> |
| Descripción Matemática . . . . .  | 74        |
| Parámetros de Entrada . . . . .   | 74        |
| Proceso de Discretización . . . . .   | 75        |
| 1. Constantes de Conversión Geográfica . . . . .  | 75        |
| 2. Cálculo de la Dirección de la Calle (Heading) . . . . .                              | 75        |
| 3. Transformación al Sistema de Coordenadas de la Calle . . . . .                       | 75        |
| 4. Discretización Temporal . . . . .  | 78        |
| 5. Discretización Espacial Orientada a Calles . . . . .                                 | 78        |
| 6. Agregación por Celdas 3D . . . . .   | 80        |

|   |    |
|---|----|
| 7. Construcción de la Salida . . . . .                            | 81 |
| 8. Propiedades Matemáticas Globales . . . . .                     | 81 |
| 9. Ventajas sobre Discretización Cartesiana Tradicional . . . . . | 82 |

## Extracción y caracterización de datos

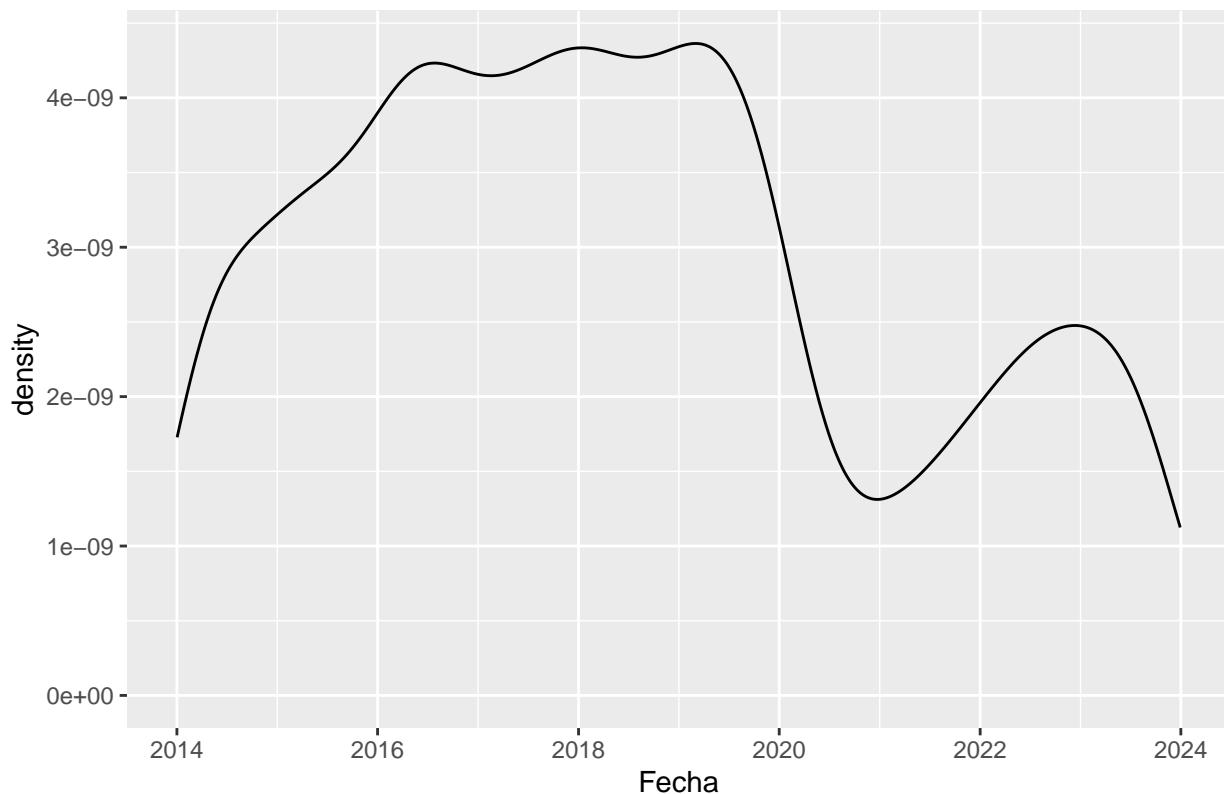
### Siniestros buses interurbanos(1).xlsx

El análisis comprende 6268 registros de siniestros de buses interurbanos ocurridos entre el 1 de enero de 2014 y el 29 de diciembre de 2023.

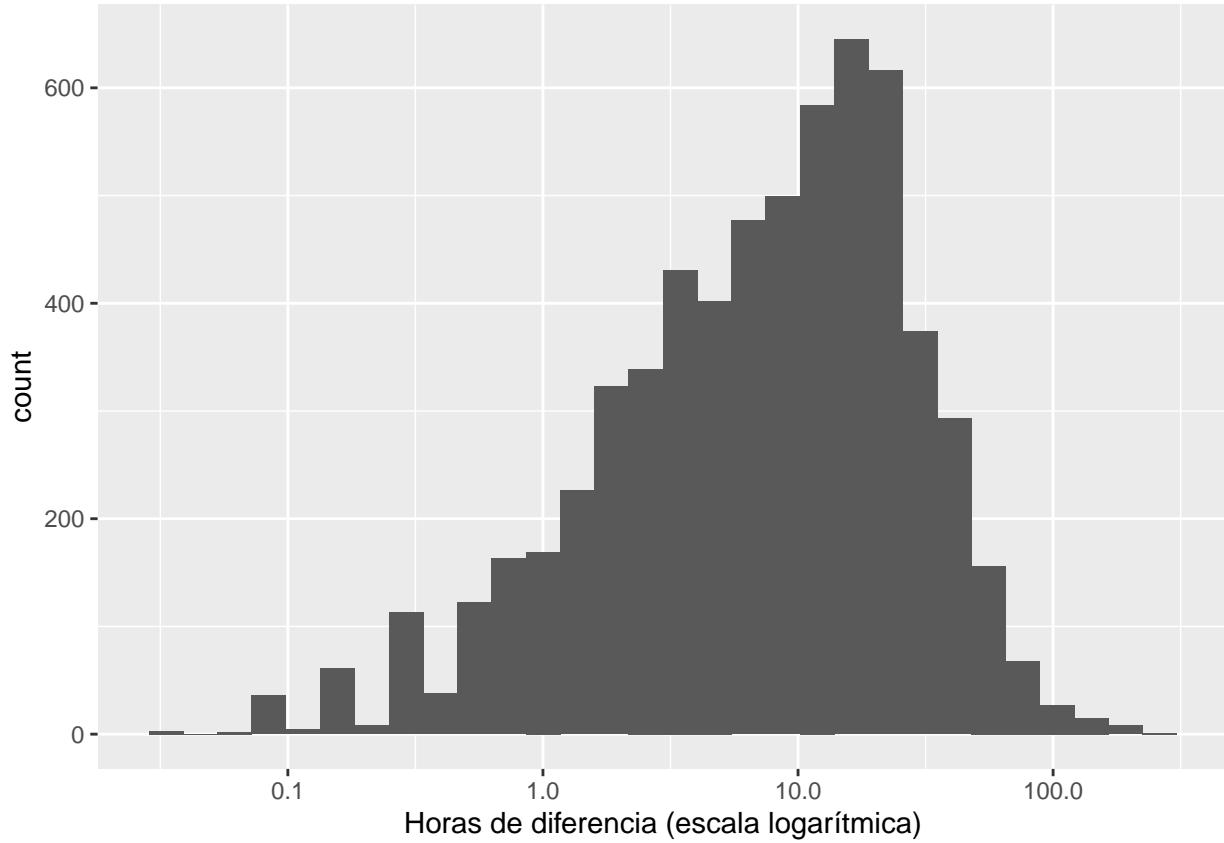
#### Granularidad temporal

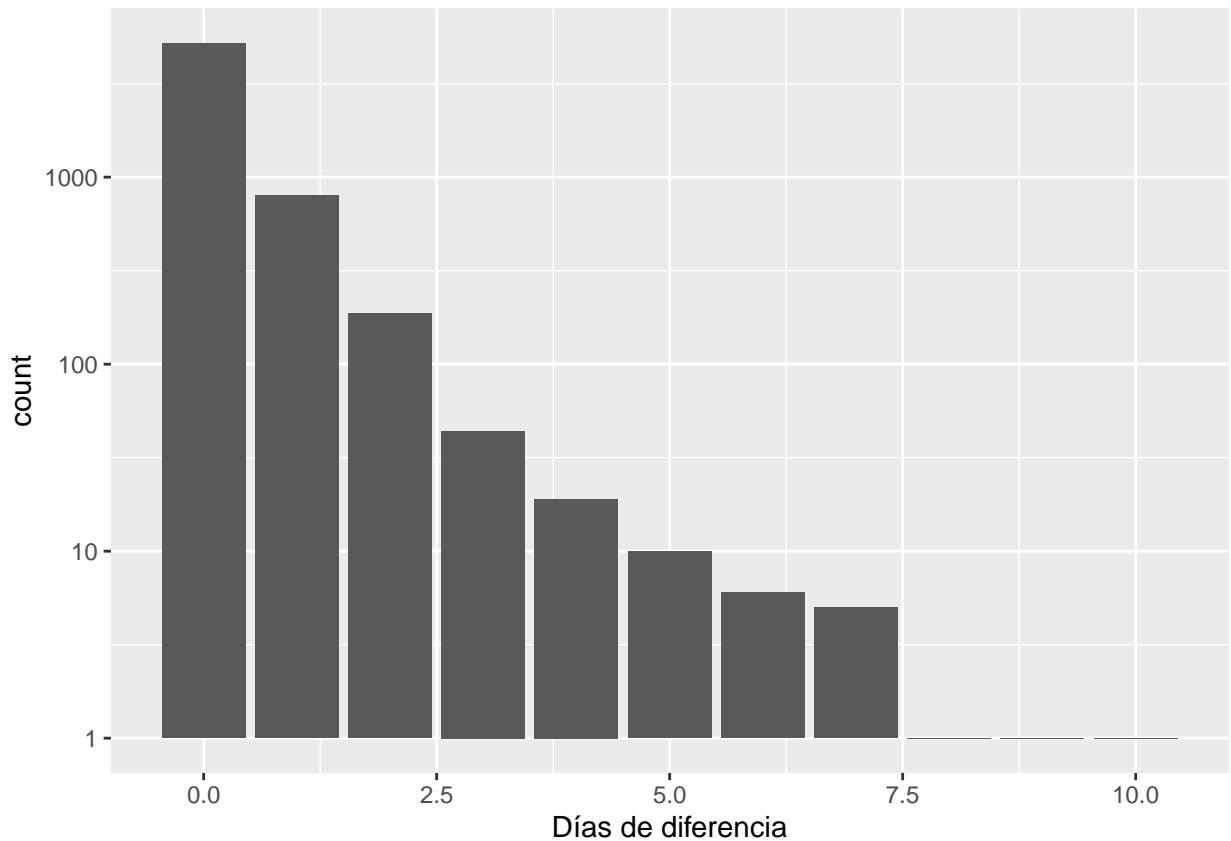
- Diferencia Mínima: 0 (ocurren al mismo tiempo)
- Diferencia Máxima: 10 días, 22 horas y 5 minutos
- Diferencia Media: 13 horas, 58 minutos y 27.13 segundos
- Diferencia Mediana: 8 horas y 25 minutos
- Desviación Standard: 17 horas, 58 minutos, 46.71 segundos

Frecuencia por fecha



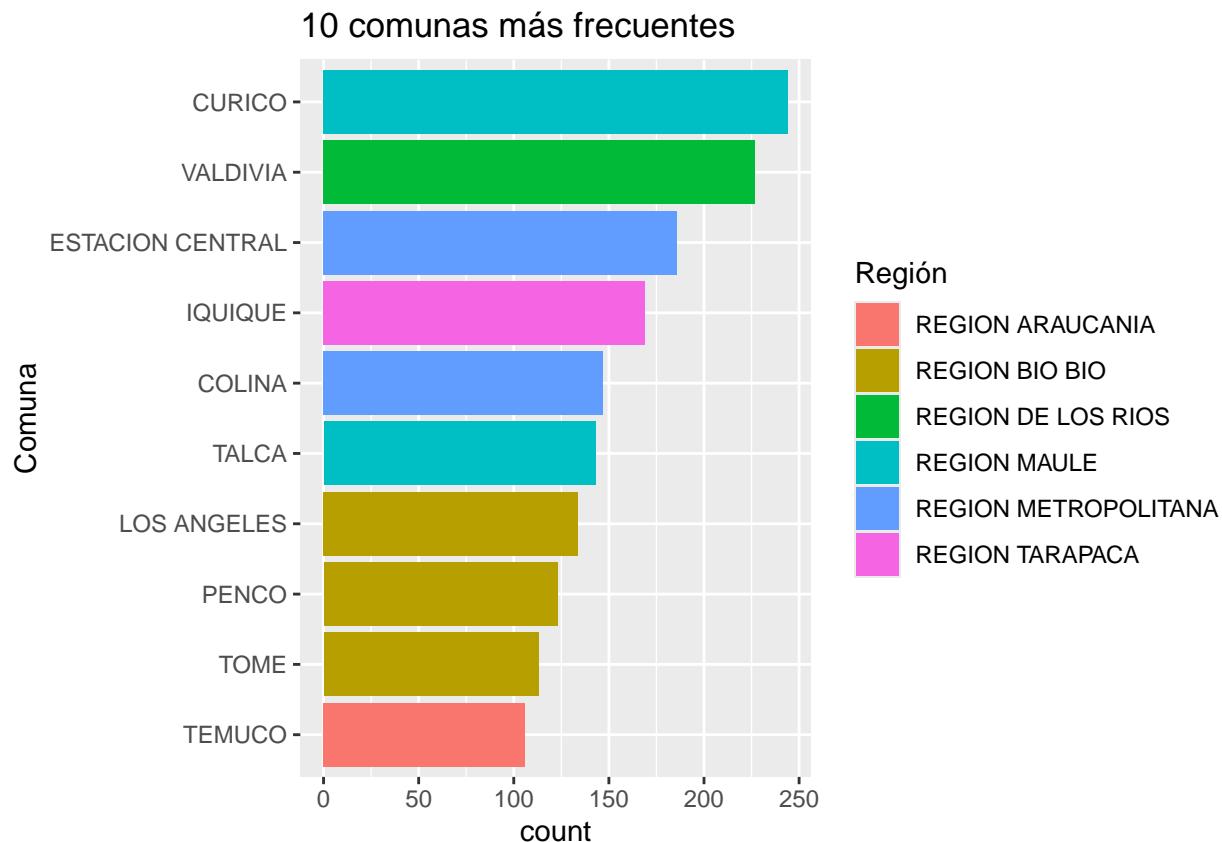
```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```





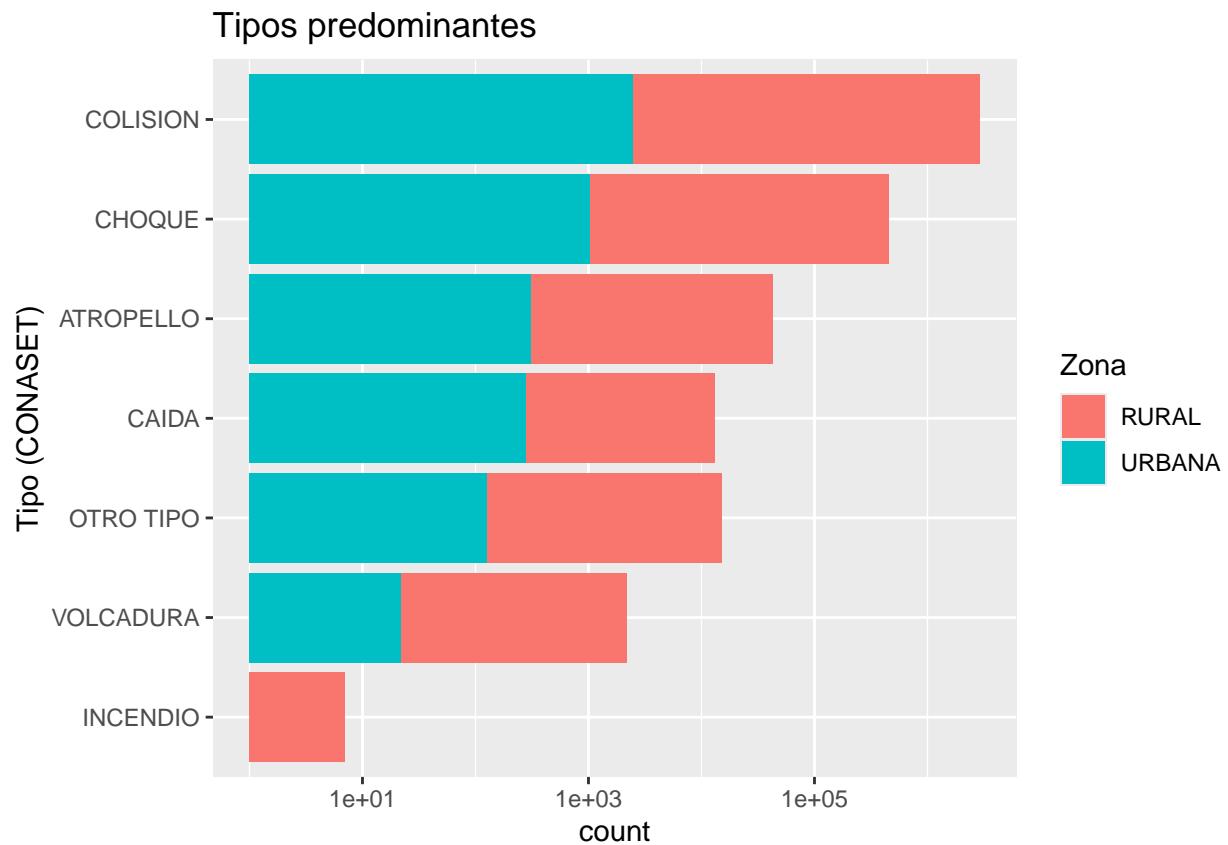
### Distribución Geográfica

- **Regiones más afectadas:** Región Metropolitana (1285 casos), Región del Maule (835 casos) y Región del Biobío (767 casos)
- **Comunas con mayor incidencia:** Curicó (244 siniestros), Valdivia (227) y Estación Central (186)



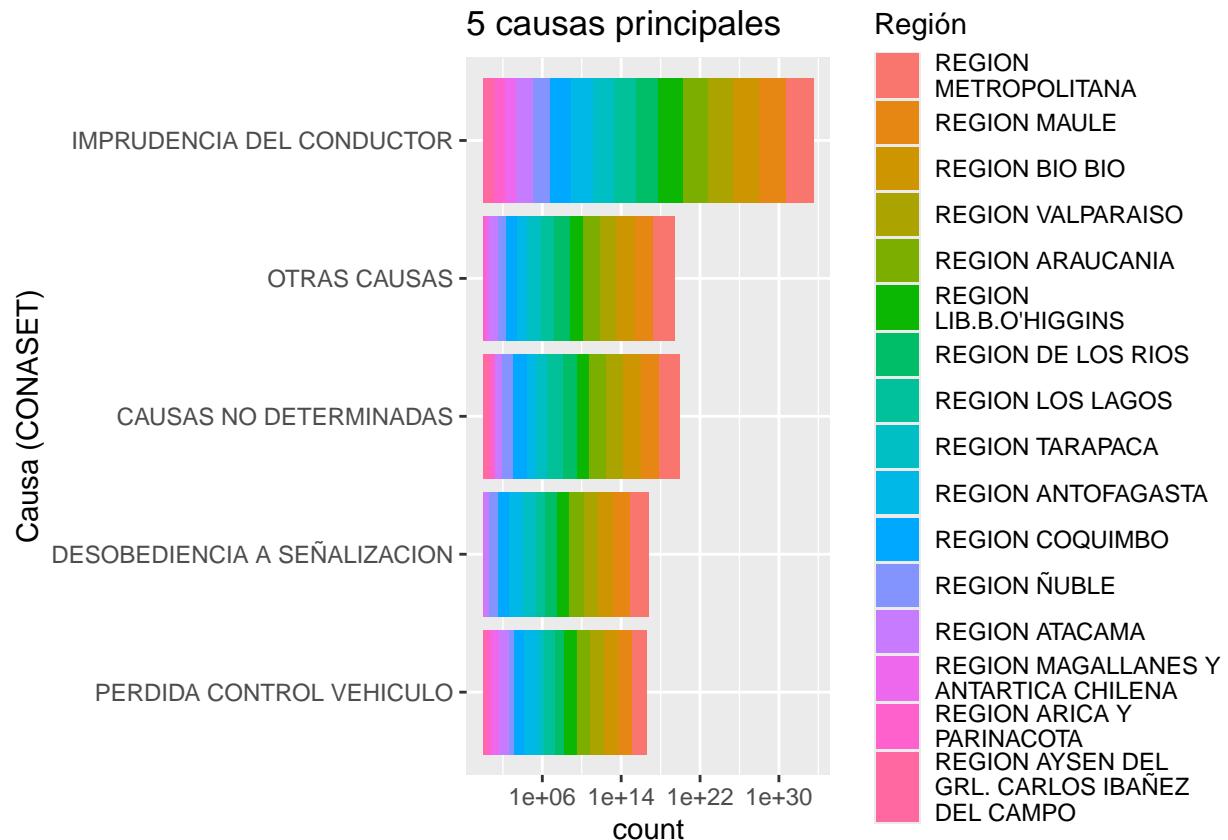
#### Características de los Siniestros

- **Tipos predominantes:** Colisión (3655 casos), choque (1464) y atropello (448)
- **Distribución por zona:** 4251 accidentes en zonas urbanas versus 2,017 en zonas rurales

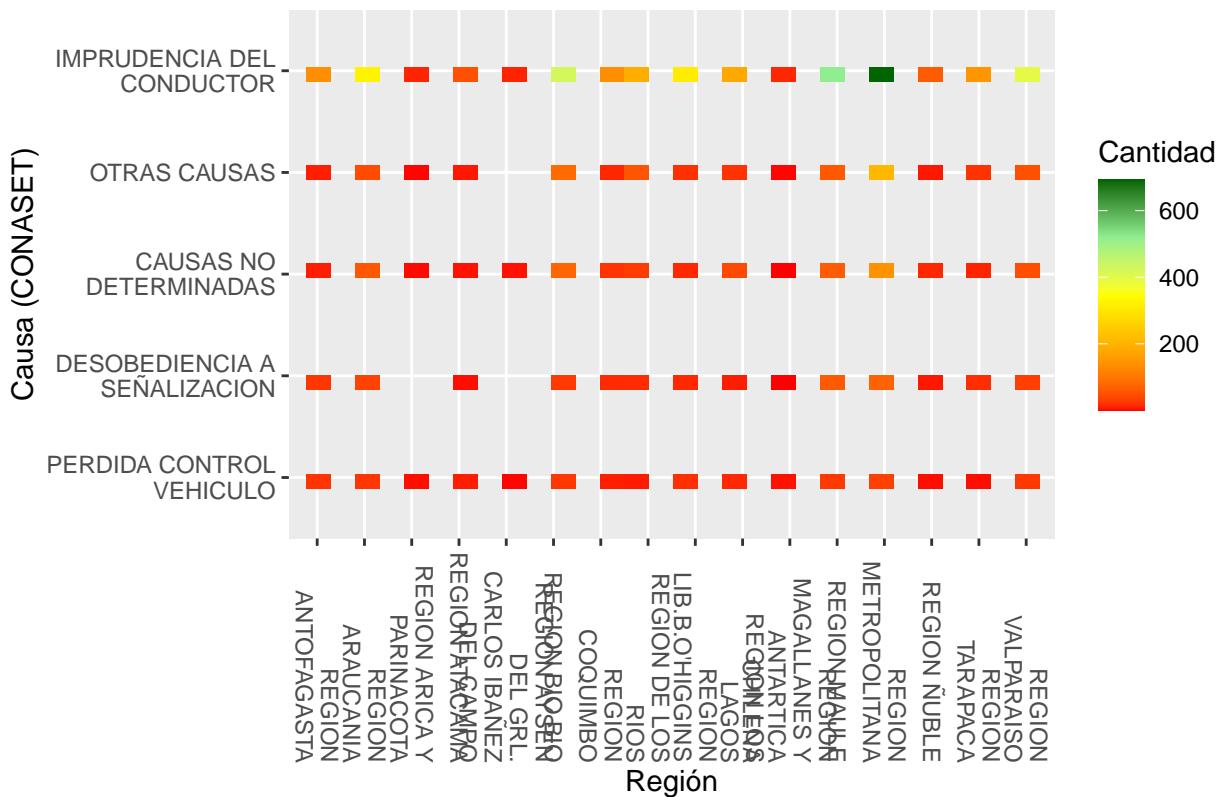


#### Factores Causales

- **Principales causas:** Imprudencia del conductor (3597 casos), otras causas (599) y causas no determinadas (540)
- *Nota:* La alta frecuencia de la categoría “otras causas” sugiere la necesidad de ampliar la clasificación causal existente



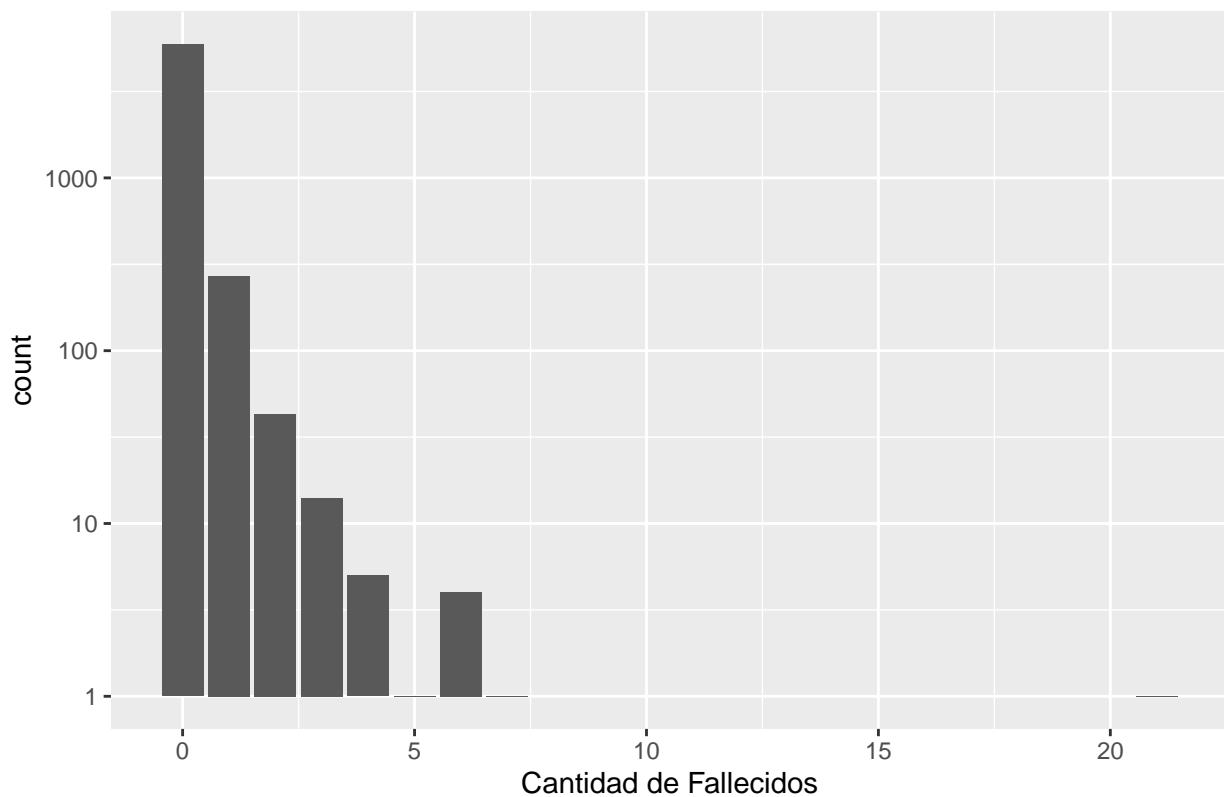
## 5 causas principales



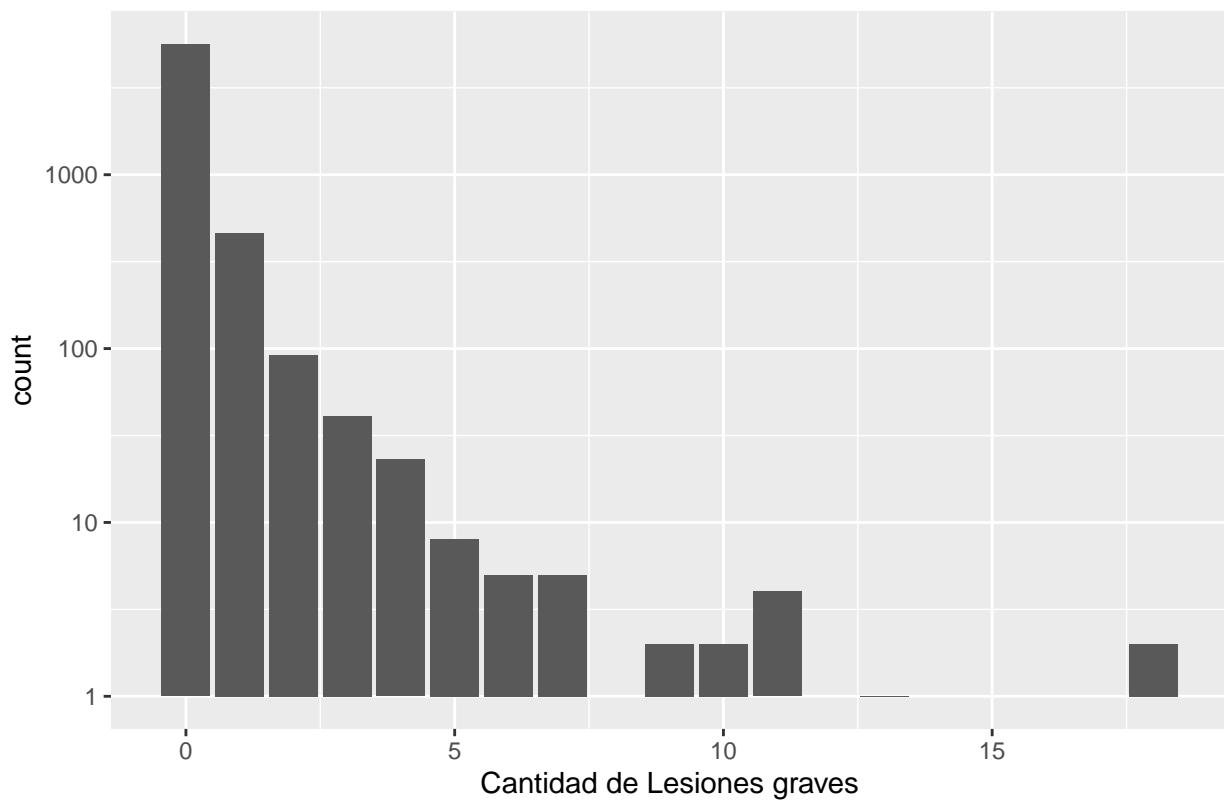
## Consecuencias y Severidad

- Fallecidos: 94.58% de los siniestros no registran víctimas mortales
- Lesiones graves: Ausentes en 89.69% de los casos
- Lesiones menos graves: No ocurren en 94.16% de los accidentes
- Lesiones leves: No se presentan en 63.10% de los incidentes

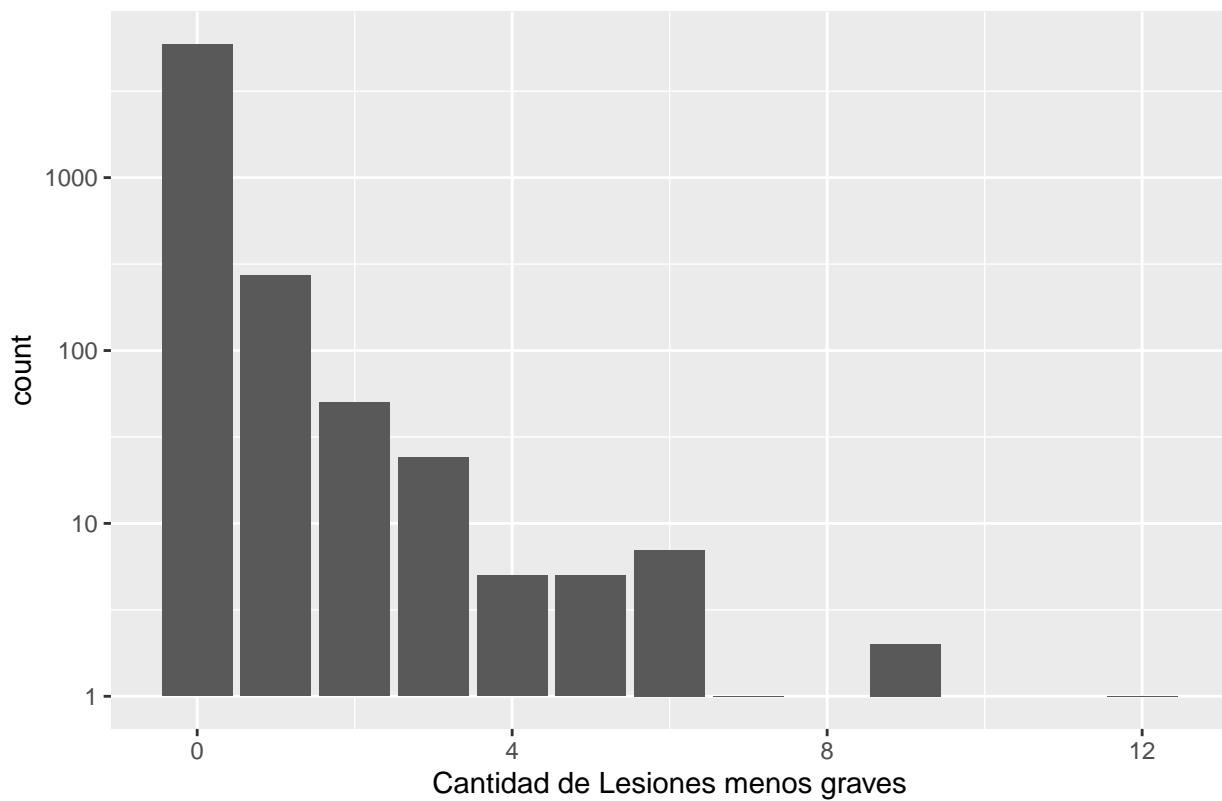
## Distribución de cantidad de Fallecidos por accidente



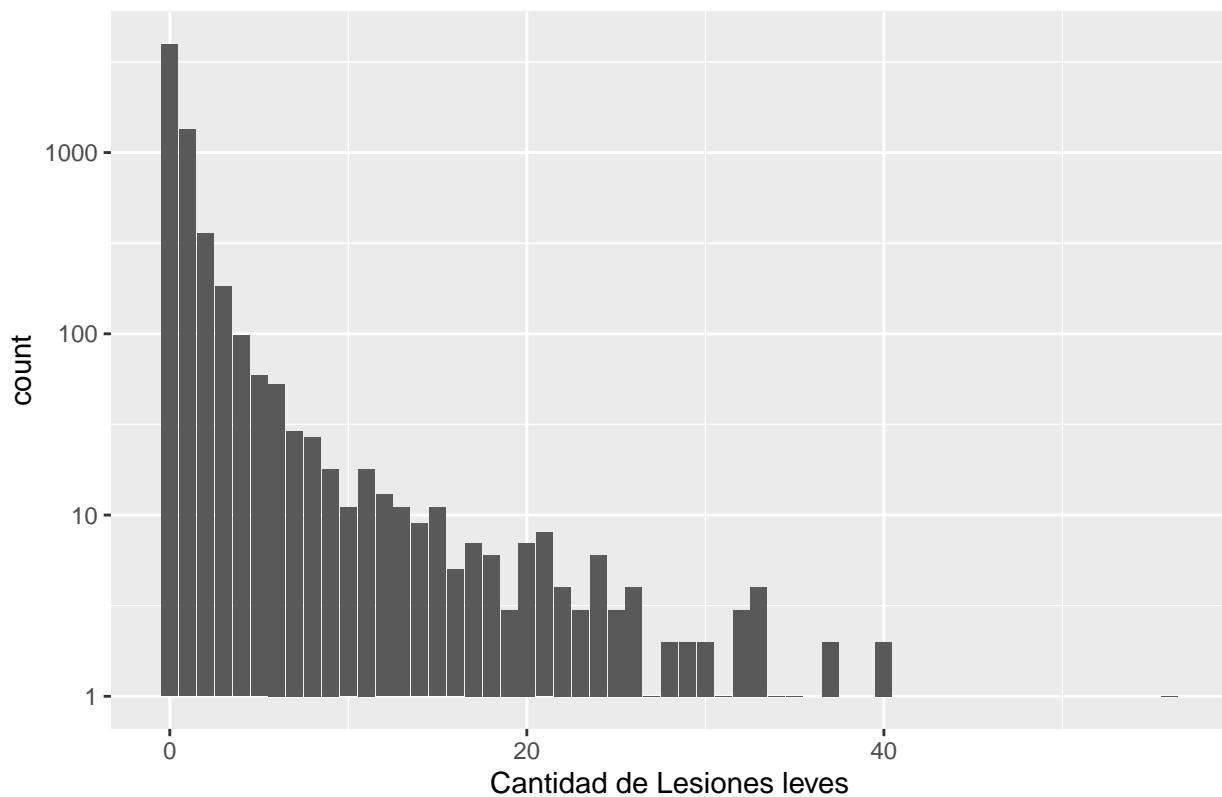
## Distribución de cantidad de Lesiones graves por accidente



Distribución de cantidad de Lesiones menos graves por accidente



## Distribución de cantidad de Lesiones leves por accidente



## Optimizaciones Realizadas

### Paso a paso

- “Fecha” se convirtió en la concatenación de “Fecha” y “Hora” y se convierte en POSIXct con formato “%d/%m/%Y %H:%M:%S”
- Se elimina la columna “Hora”
- “Región” se convierte en factor
- “Comuna” se convierte en factor
- “Tipo Accidente” se convierte en factor
- “Tipo (CONASET)” se convierte en factor
- “Zona” se convierte en factor
- “Ubicación Relativa” se convierte en factor
- “Causa (CONASET)” se convierte en factor
- “Causa Accidente” se convierte en factor

## GH026929.xlsx

El dataset contiene 1726 registros de telemetría GPS capturados durante un período específico de operación vehicular, .

### Características Temporales

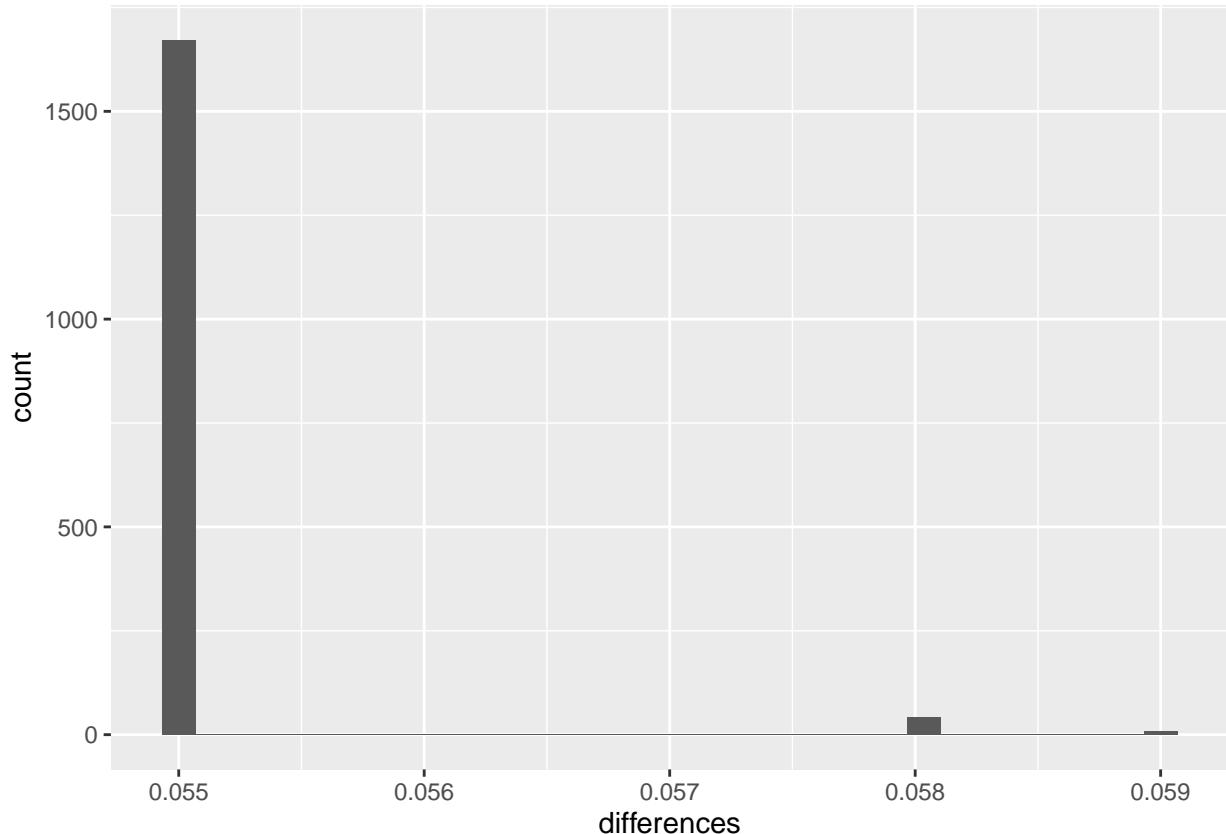
- **Período de registro:** 29 de agosto de 2024, entre las 01:32 y 01:34 horas

- **Duración total:** Aproximadamente 2 minutos de monitoreo continuo

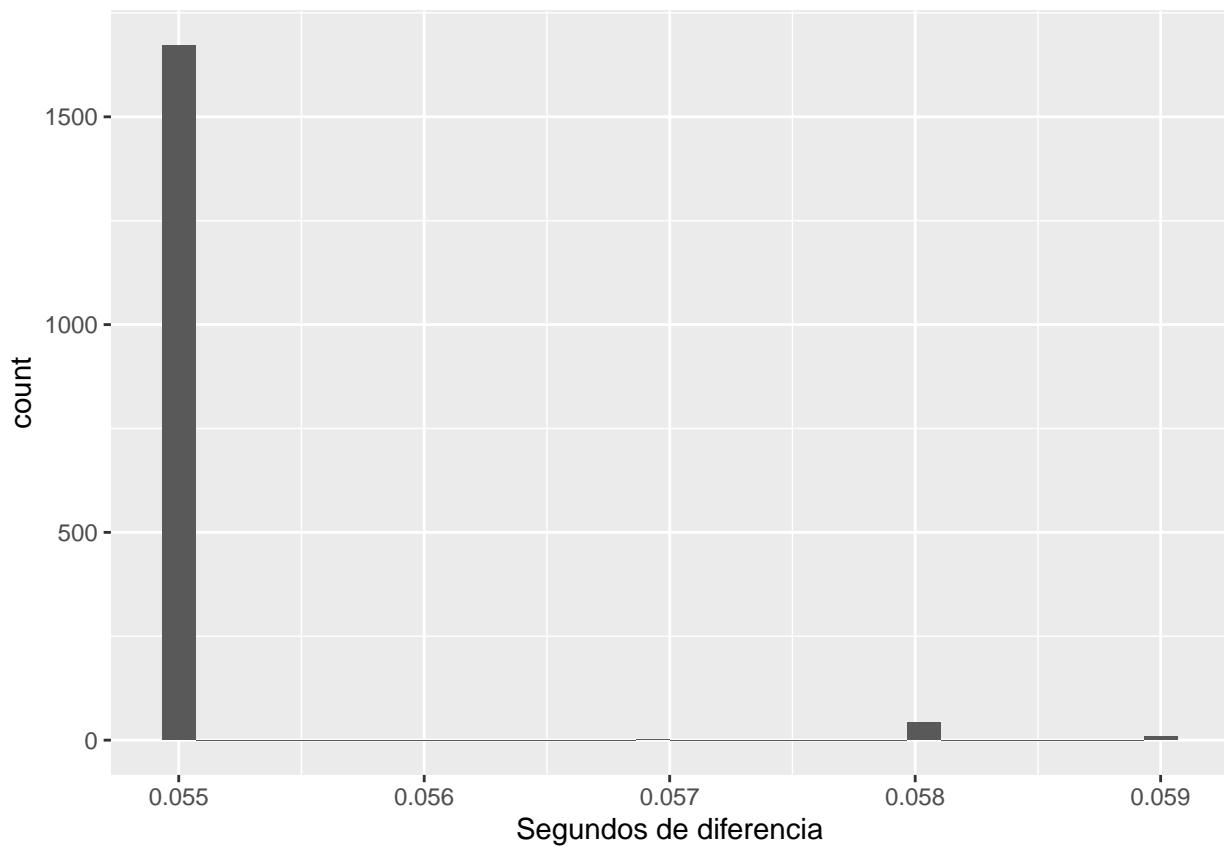
### Granularidad temporal

- Diferencia Mínima: 0.55 segundos
- Diferencia Máxima: 0.59 segundos
- Diferencia Media: 0.55 segundos
- Diferencia Mediana: 0.55 segundos
- Desviación Standard:  $5.42 \times 10^{-4}$  segundos

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```



```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```

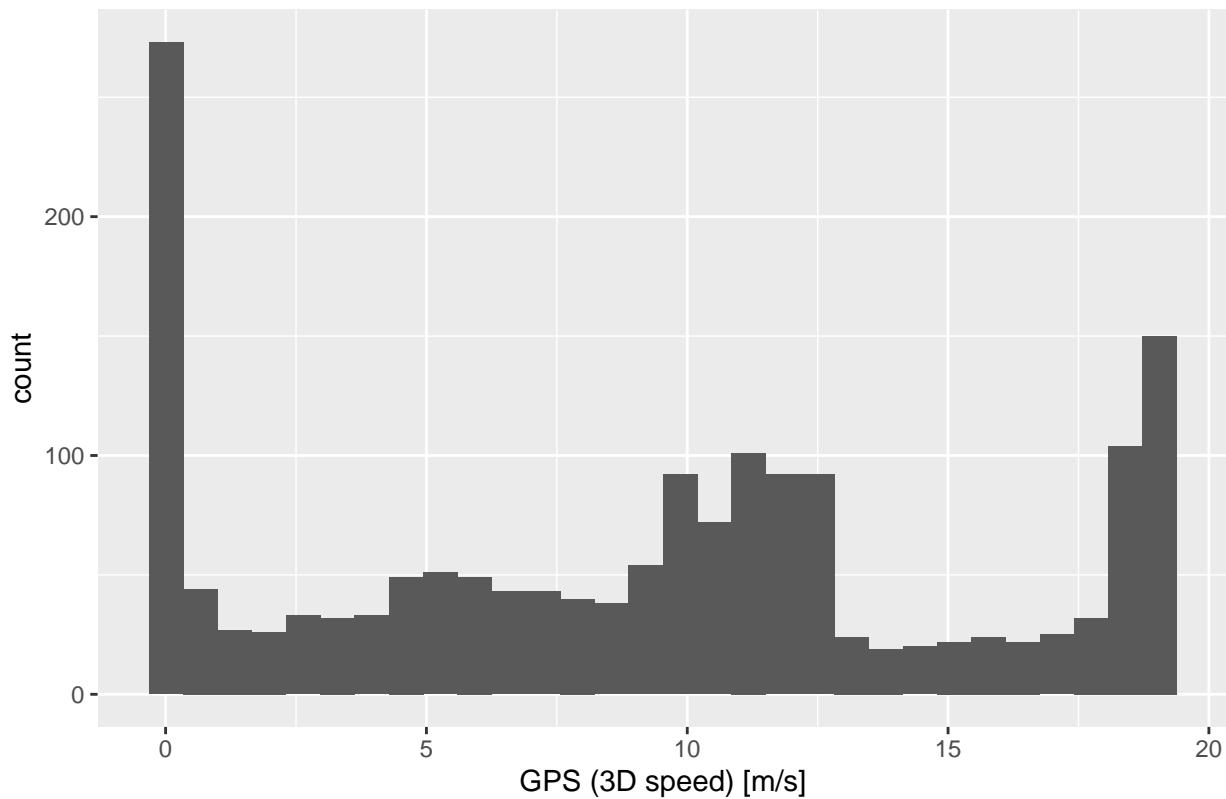


### Parámetros de Velocidad

- **Velocidad 3D:** Rango de 0.01 a 19.06 m/s (0.036 a 68.62 km/h)
- **Velocidad promedio:** 9.1 m/s (32.76 km/h)
- **Velocidad mediana:** 9.81 m/s (35.32 km/h), indicando una distribución sesgada hacia velocidades moderadas-altas

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```

## Distribución de velocidades

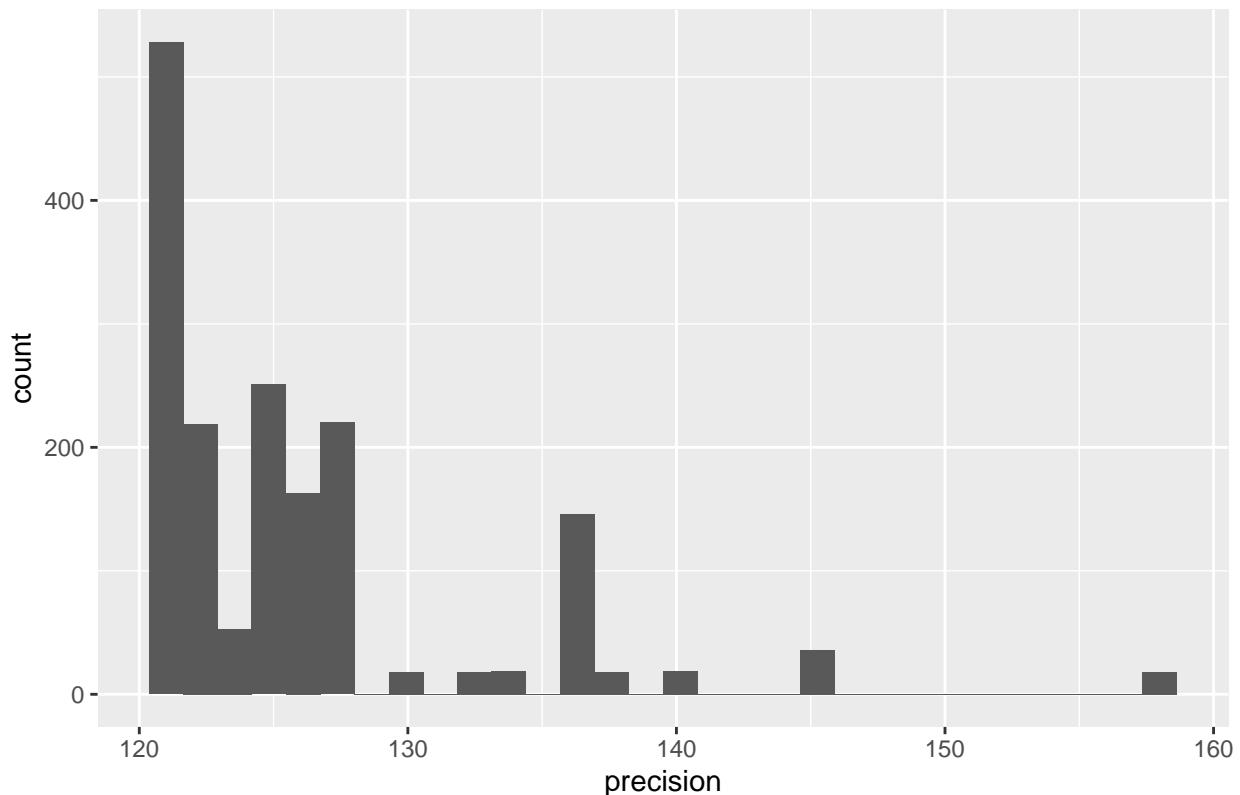


## Calidad de Señal GPS

- **Valor fix:** Constante en 3 para todos los registros, indicando que calculó su posición a partir de 3 satélites
- **Precisión:** Oscila entre 121 y 158 m (media: 125.9 m, mediana: 125 m)

```
## 'stat_bin()' using 'bins = 30'. Pick better value 'binwidth'.
```

## Distribución de precisiones



## Aspectos Requieren Clarificación

### Variables por Definir

- **cts:** Propósito y unidades no especificadas
- **Rango de alturas:** Valores de altitud GPS no reportados en el resumen

### Problemas de Formato Identificados

- Coordenadas GPS (Lat. y Long.) sin formato decimal apropiado
- Velocidad 2D con inconsistencias en la ubicación del punto decimal
- Múltiples campos numéricos presentados sin separador decimal

## Procesamiento de Datos Realizado

### Paso a paso

- “date” se convirtió en la concatenación de “date” y “hora” y se convierte en POSIXct con formato “%d/%m/%Y %H:%M”
- Se elimina la columna “hora”
- “GPS (2D speed) [m/s]” se trató de poner al mismo orden de magnitud de “GPS (3D speed) [m/s]” mediante la siguiente fórmula:

$$\text{reduce.2d.speed}(n, m) \equiv n \times 10^{\lfloor \log_{10} m+1 \rfloor - \lfloor \log_{10} n+1 \rfloor}$$

- “GPS (Lat.) [deg]” y “GPS (Long.) [deg]” se ajustaron para tener 2 dígitos enteros mediante la siguiente fórmula:

$$\text{reduce.coords}(n) \equiv n \times 10^{1 - \lfloor \log_{10} |n| \rfloor}$$

## DTPR

El dataset consolidado comprende 21683856 registros de telemetría GPS provenientes del sistema de monitoreo de transporte público DTPR, capturados durante el mes de agosto de 2025 a través de 31 unidades operativas diferentes.

### Metadatos de Procesamiento

- **Tiempo total de lectura:** 7 minutos 8.22 segundos
- **Tiempo de consolidación:** 5 minutos 0.63 segundos
- **Tiempo de transformación:** 30 minutos 7.08 segundos
- **Total de procesamiento:** 42 minutos 15.93 segundos

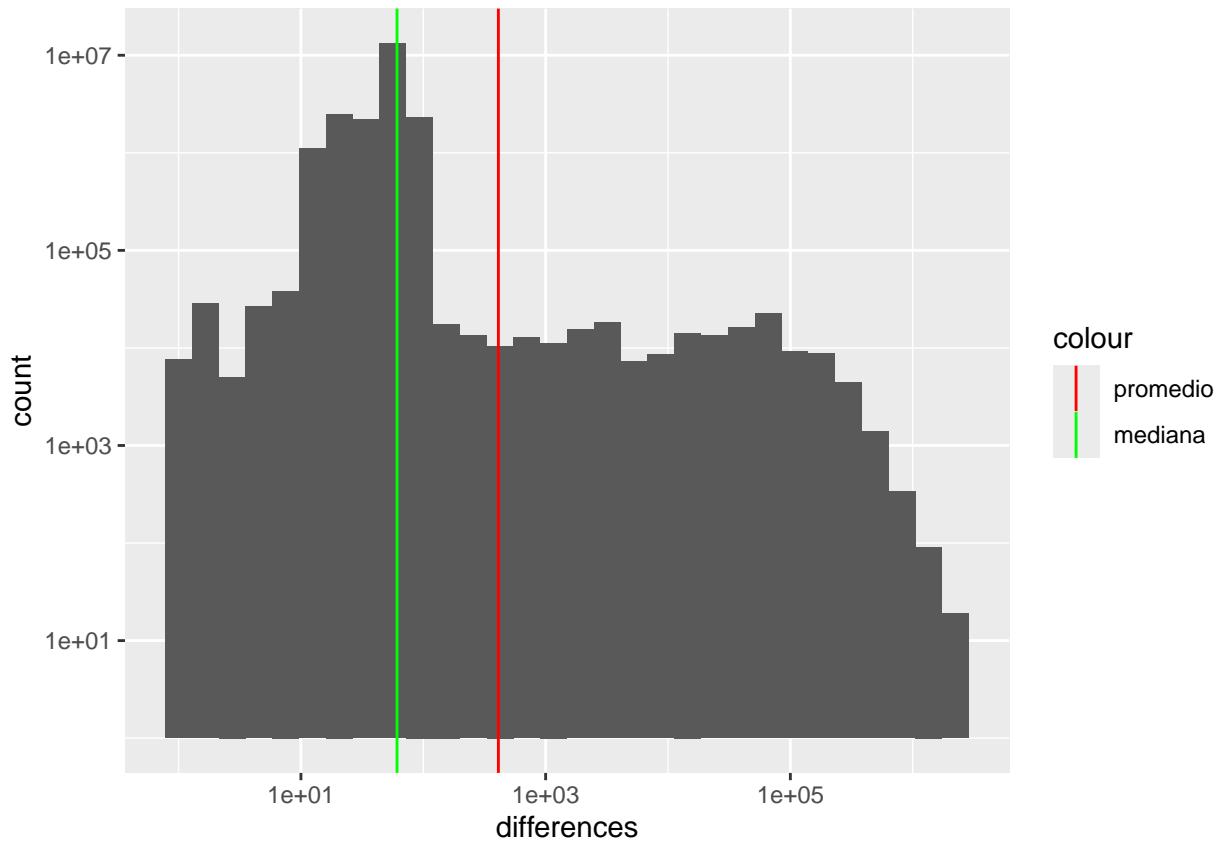
### Período de Monitoreo

- **Fecha inicial:** 1 de agosto de 2025, 00:00:00 (hora Chile)
- **Fecha final:** 1 de septiembre de 2025, 00:05:27 (hora Chile)
- **Duración:** 31 días completos + 5 minutos
- **Cobertura temporal:** Mes completo de agosto 2025

### Granularidad temporal (separado por EVT\_PPU)

- Diferencia Mínima: 0 (al mismo tiempo)
- Diferencia Máxima: 25 días, 22 horas, 42 minutos, 36 segundos
- Diferencia Media: 6 minutos, 51 segundos
- Diferencia Mediana: 1 minuto, 1 segundo
- Desviación Standard: 2 horas, 24 minutos, 17.19 segundos

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```

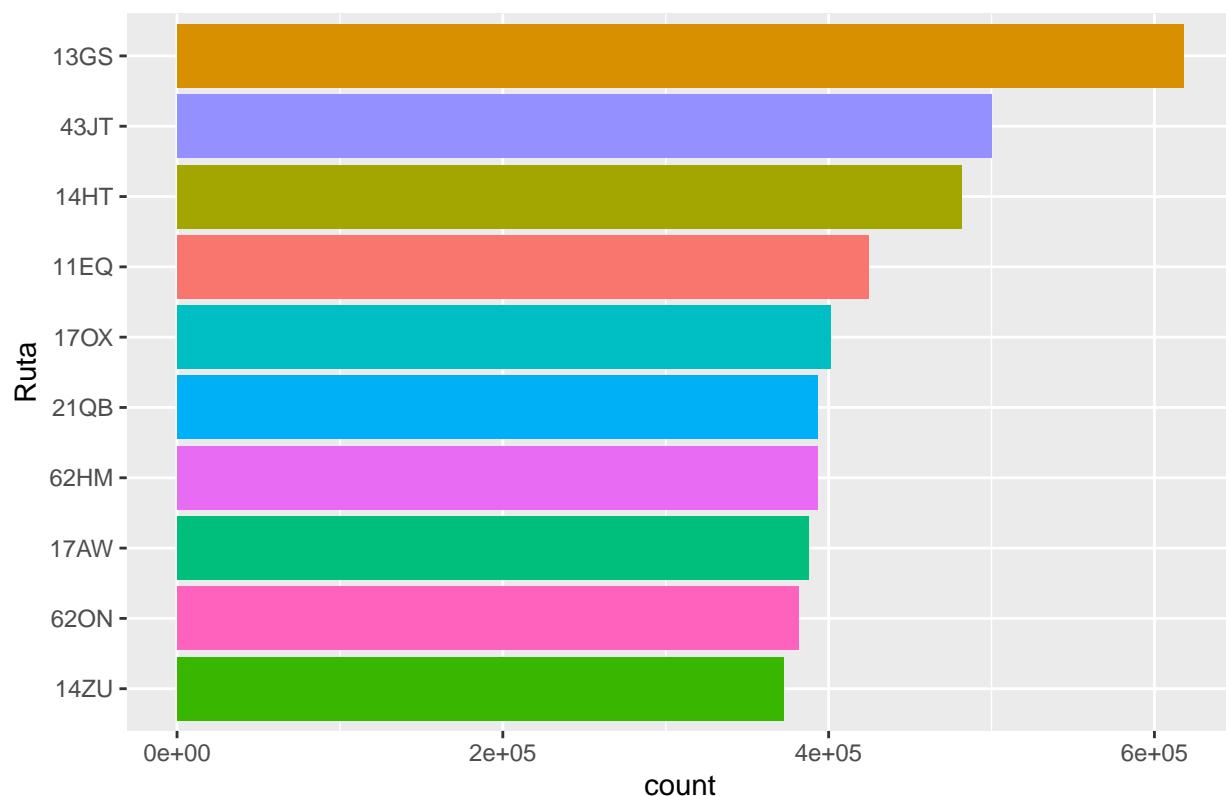


## Características Operacionales del Sistema

### Flota y Rutas Monitoreadas

- **Rutas más frecuentes:**
  - “13GS”: 617591 registros
  - “43JT”: 500230 registros
  - “14HT”: 481486 registros
- **Vehículos más activos:**
  - “FSLC62”: 27847 registros
  - “TTHL10”: 25901 registros
  - “SZDZ86”: 25483 registros

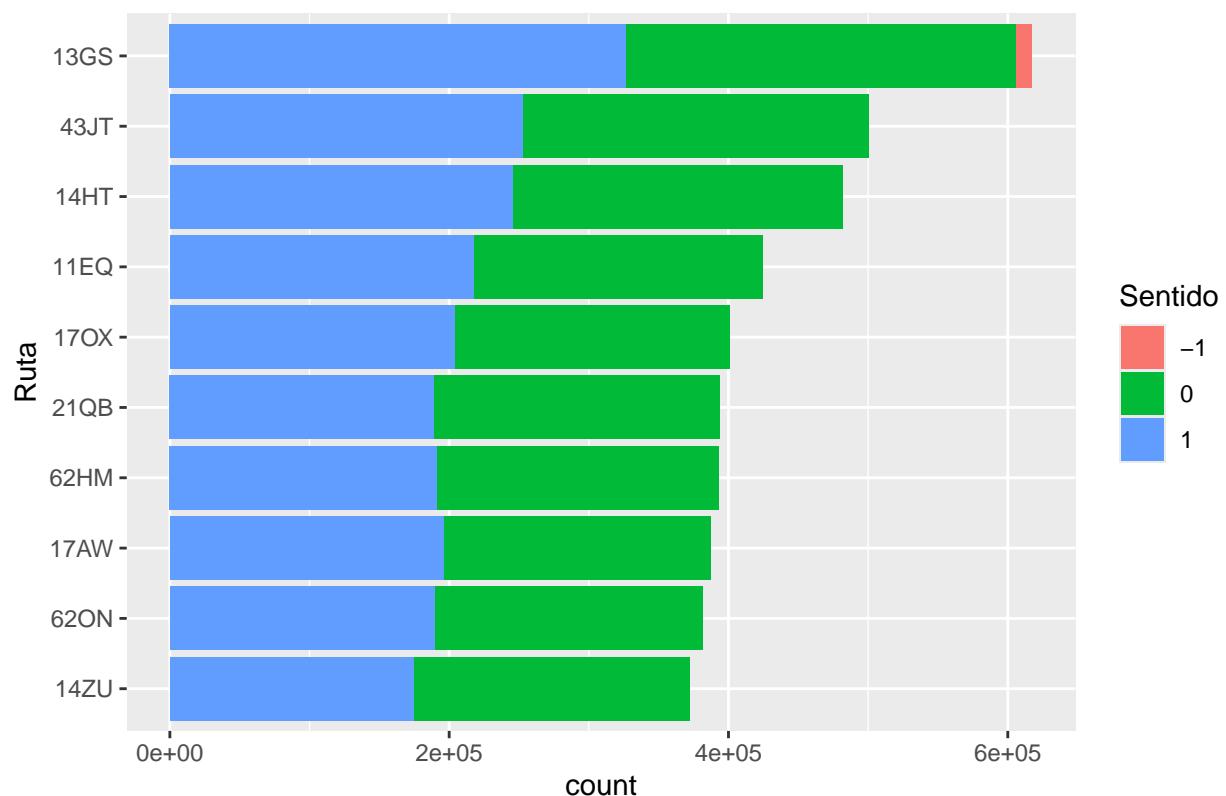
## 10 rutas más frecuentes



## Distribución Direccional

- **Sentido ida (0):** 10809559 registros
- **Sentido vuelta (1):** 10671146 registros
- **Sentido indeterminado (-1):** 203151 registros

## 10 rutas más frecuentes

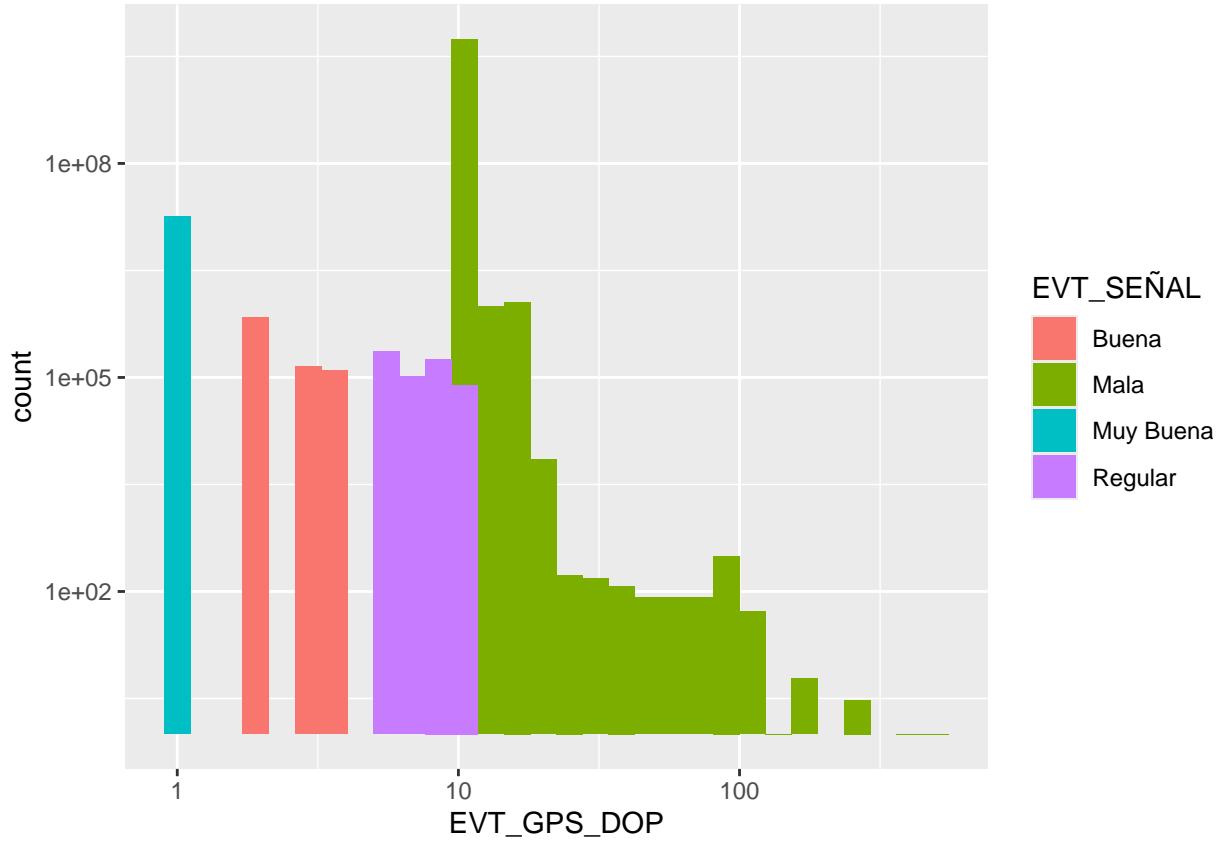


## Métricas de Desempeño Técnico

### *Calidad de Señal GPS*

- Muy Buena: 17924666 registros
- Buena: 966491 registros
- Regular: 592320 registros
- Mala: 2200379 registros

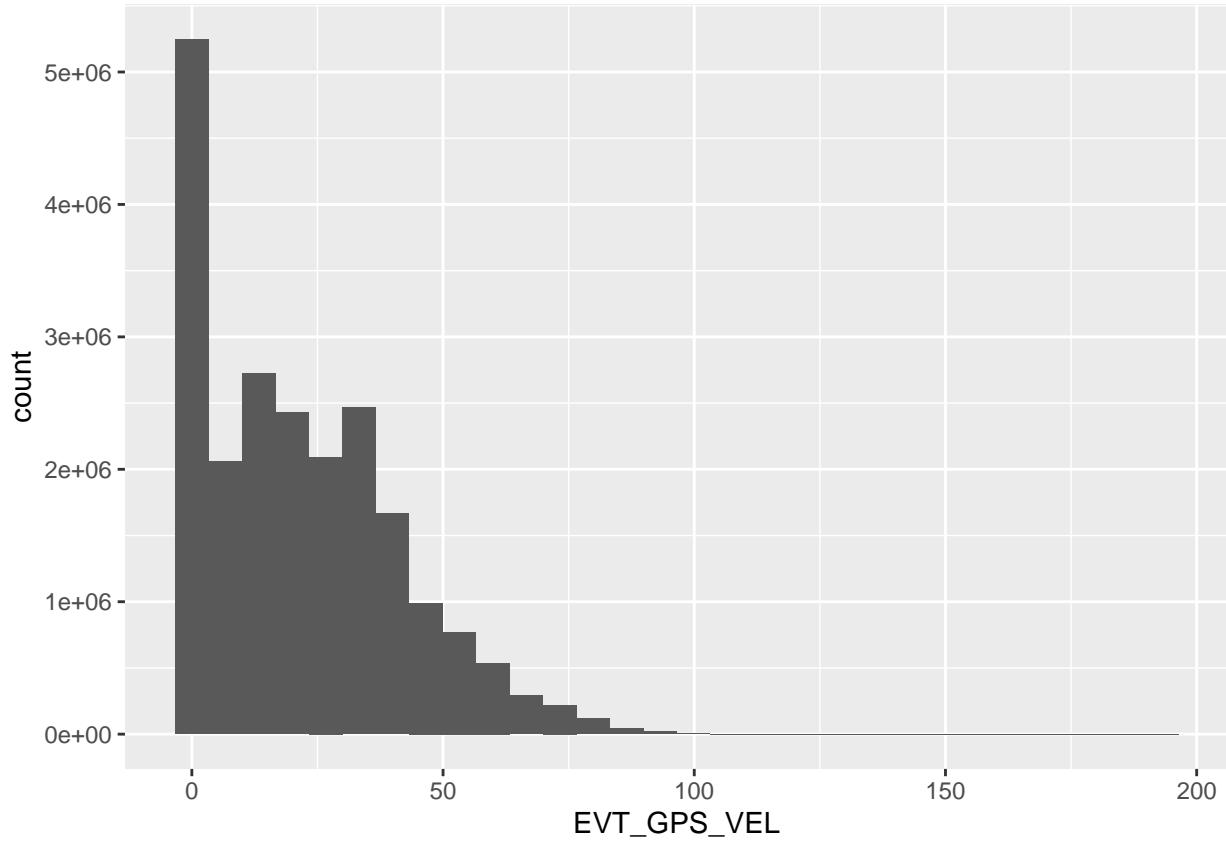
```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```



### *Velocidades Operacionales*

- Rango: 0 - 193 km/h
- Velocidad promedio: 21.69 km/h
- Velocidad mediana: 19 km/h

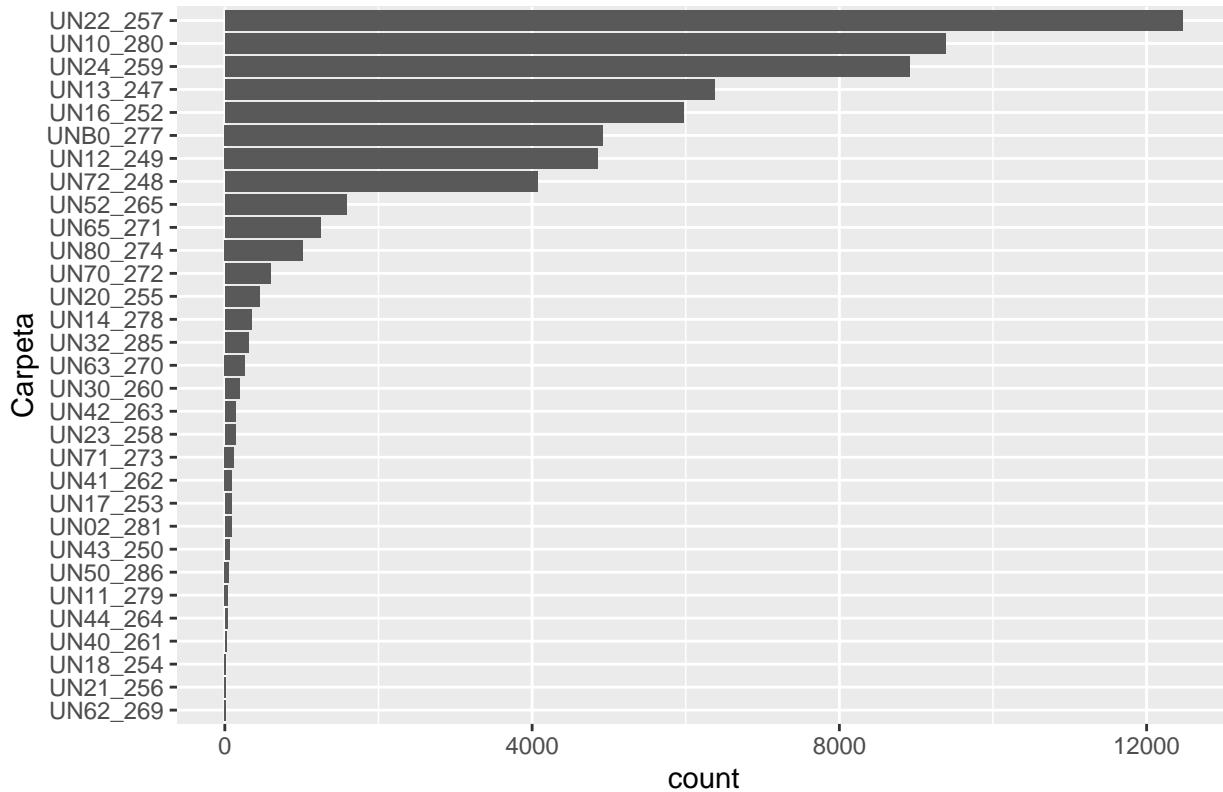
```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```



#### *Estados del Sistema*

- Motor encendido (1): 21619957 registros
- Motor apagado (0): 63899 registros

## Registros con motor apagado



## Procesamiento y Normalización de Datos

### Paso a paso

- Se lee la carpeta
- Se crea la columna “EVT\_CARPETA” con el nombre de la carpeta
- Se unifican los datasets en uno
- “EVT\_SERVICIO\_ID” se convierte en factor
- “EVT\_PPU” se convierte en factor
- “EVT\_GPS\_TIME\_CHILE\_STR” se convierte en POSIXct con formato “%d/%m/%Y %H:%M:%S” y zona horaria “America/Santiago”
- “EVT\_GPS\_TIME\_UTC\_0” se convierte en POSIXct con formato “%d/%m/%Y %H:%M:%S”
- En “Y\_EVT\_GPS\_LAT”, se reemplaza “,” por “.” y se convierte en numérico
- En “X\_EVT\_GPS\_LON”, se reemplaza “,” por “.” y se convierte en numérico
- En “EVT\_DISTANCIA\_RECORRIDA (Km)”, se reemplaza “,” por “.” y se convierte en numérico
- En “EVT\_DISTANCIA\_A\_SERVICIO”, se reemplaza “,” por “.” y se convierte en numérico
- “EVT\_CARPETA” se convierte en factor
- Se crea la columna “EVT\_SEÑAL” a partir de los valores de la columna “EVT\_GPS\_DOP”:
  - Excelente:  $EVTGPSDOP < 1$
  - Muy Buena:  $1 \leq EVTGPSDOP < 2$
  - Buena:  $2 \leq EVTGPSDOP < 5$
  - Regular:  $5 \leq EVTGPSDOP \leq 10$
  - Mala:  $EVTGPSDOP > 10$

## **Análisis por Unidad Operativa**

### **Patrones de Operación por Unidad**

#### *Frecuencia de Reporte*

- **Típico:** 1:00-1:03 minutos entre registros por vehículo (mediana, separado por patente)
- **Variaciones:** UN52\_265 muestra intervalos más irregulares (3.5s - 1min) (mediana, separado por patente)
- **Consistencia:** Mayoría mantiene intervalos estables

#### *Distribución por Unidad*

- **Mayor volumen:** UN80\_274 (1459884 registros)
- **Menor volumen:** UN52\_265 (118026 registros)
- **Promedio por unidad:** ~699479 registros

#### *Características Específicas Destacables*

- **UN13\_247:** Operación exclusiva de ruta “13GS”
- **UN18\_254:** Frecuencia muy consistente (1:03-1:03.5 min) (mediana, separado por patente)
- **UN22\_257:** Mayor variabilidad en intervalos (59s - 22:31 min) (mediana, separado por patente)
- **UNB0\_277:** Única unidad con datos en septiembre

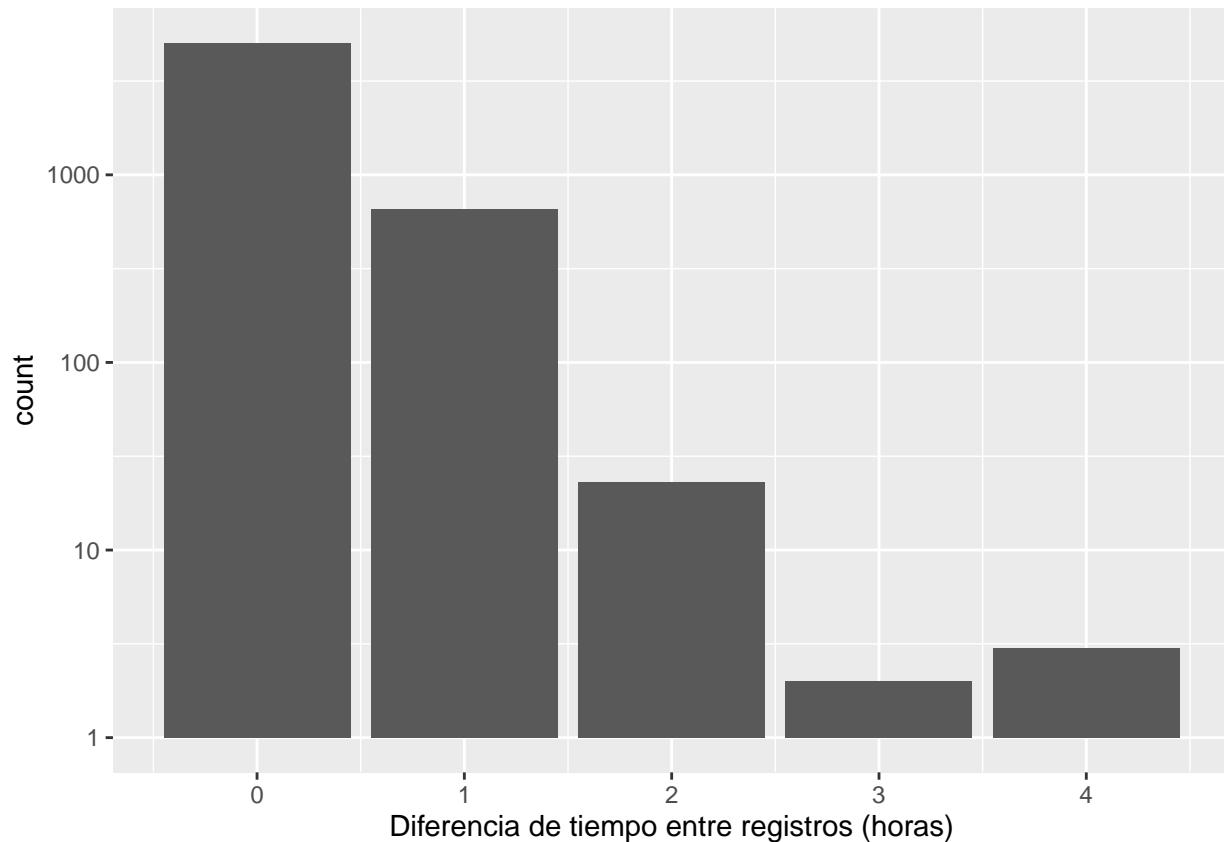
## **Data accidentes de carabineros.xlsx**

El dataset comprende 5739 registros de accidentes de tránsito documentados por Carabineros de Chile durante un período concentrado de 30 días.

### **Período de Análisis**

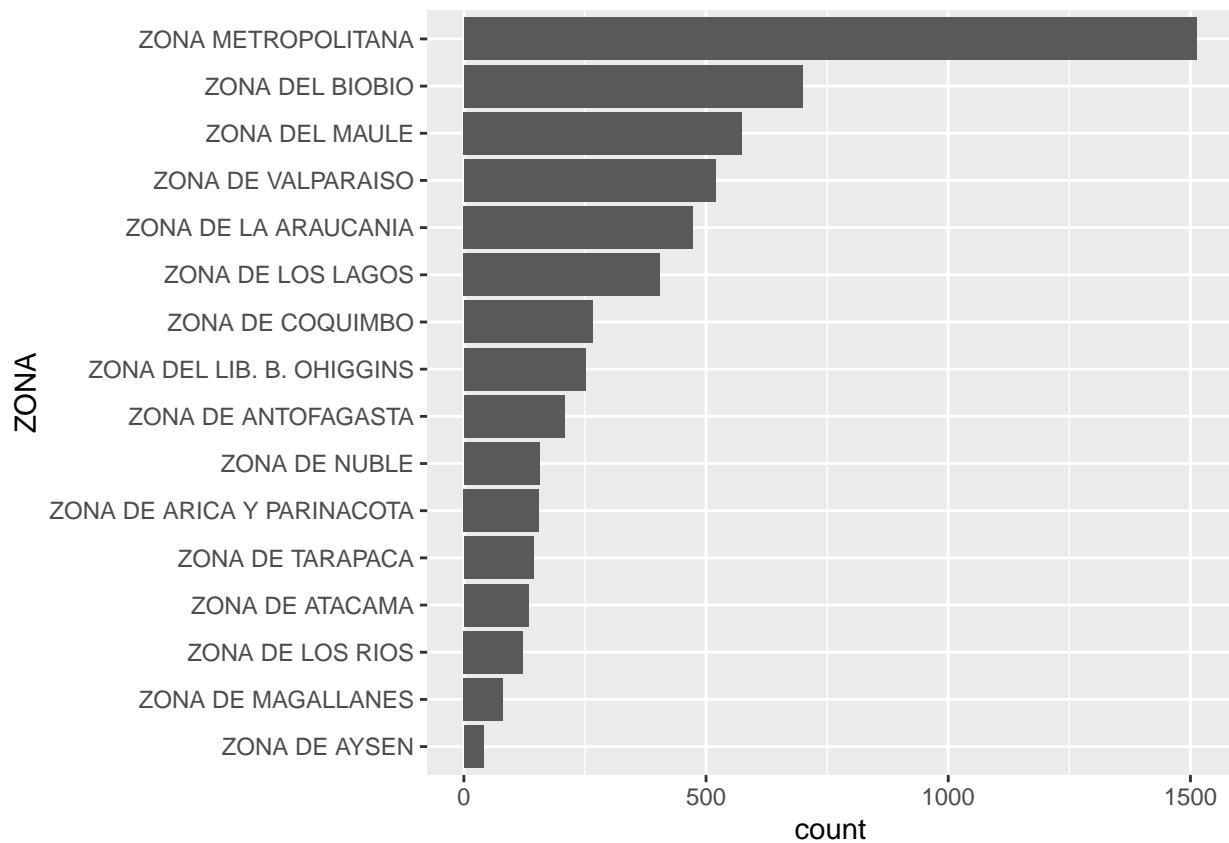
- **Cobertura temporal:** 1 al 30 de marzo de 2025 (30 días)
- **Registros diarios promedio:** aproximadamente 191 accidentes

## Granularidad Temporal



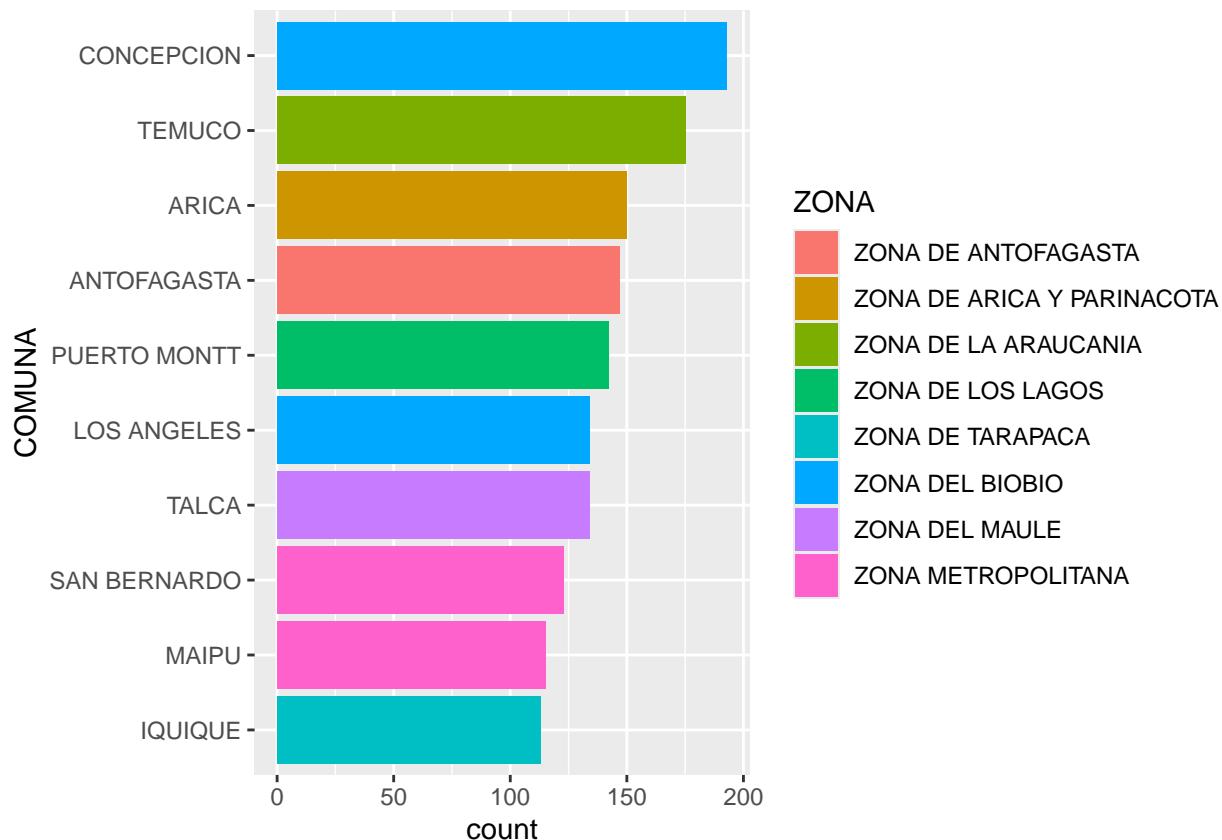
## Distribución Geográfica

- **Regiones con mayor siniestralidad:**
  - Región Metropolitana (1513 casos)
  - Región del Biobío (699 casos)
  - Región del Maule (574 casos)

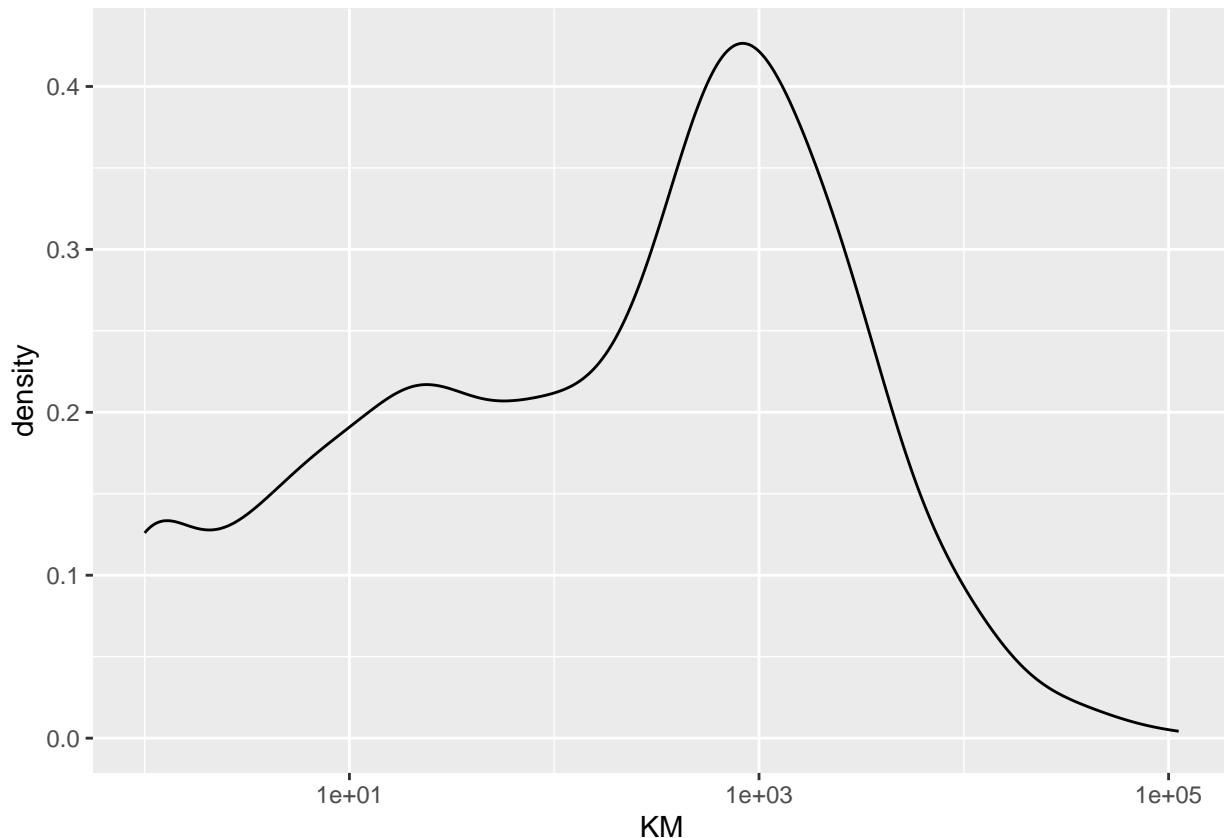


- Comunas más afectadas:

- Concepción (193 accidentes)
- Temuco (175 accidentes)
- Arica (150 accidentes)

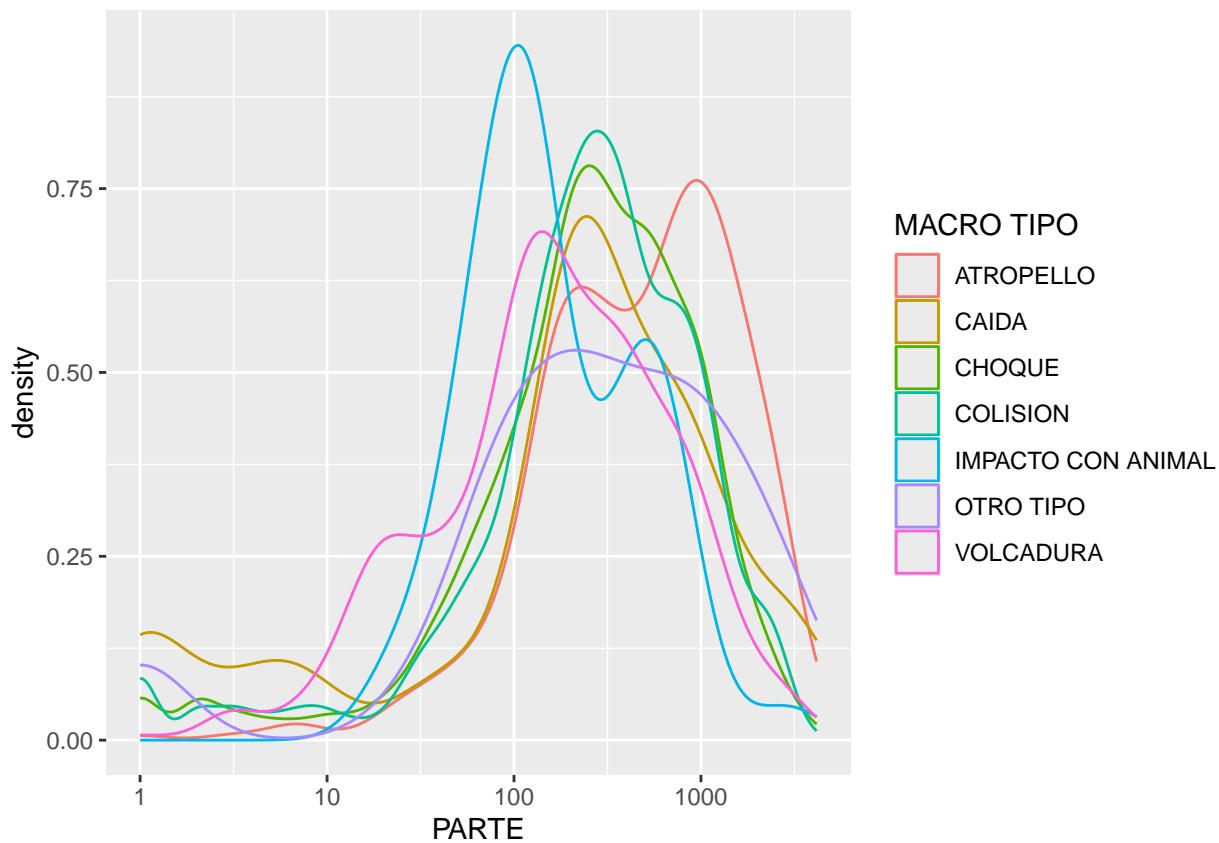


- **KM:**
  - **Rango:** Entre 0 y 112000 km
  - **Media:** 1629 km
  - **Mediana:** 312 km
  - **Nulos:** 3210 registros



- **PARTE:**

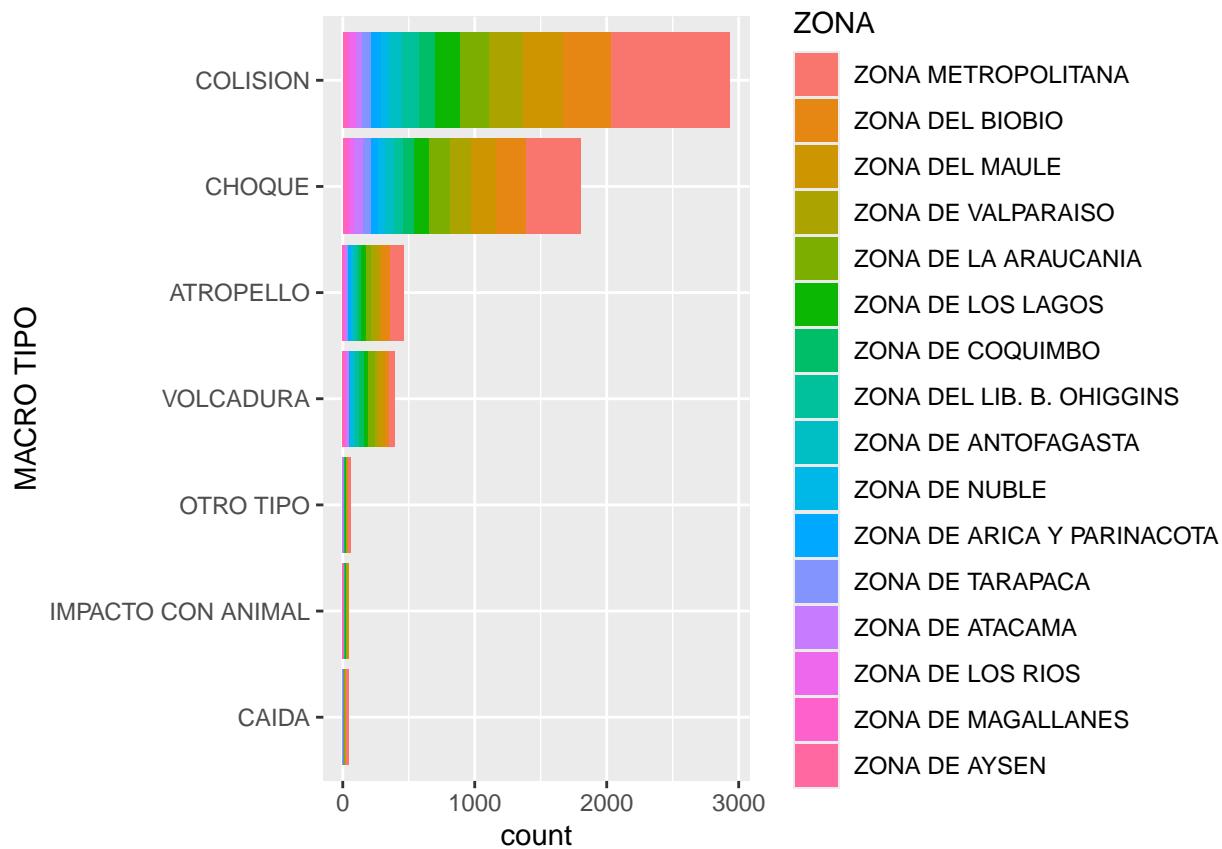
- **Rango:** Entre \$1 y 4157
- **Media:** \$491.4
- **Mediana:** \$282



### Clasificación de Accidentes

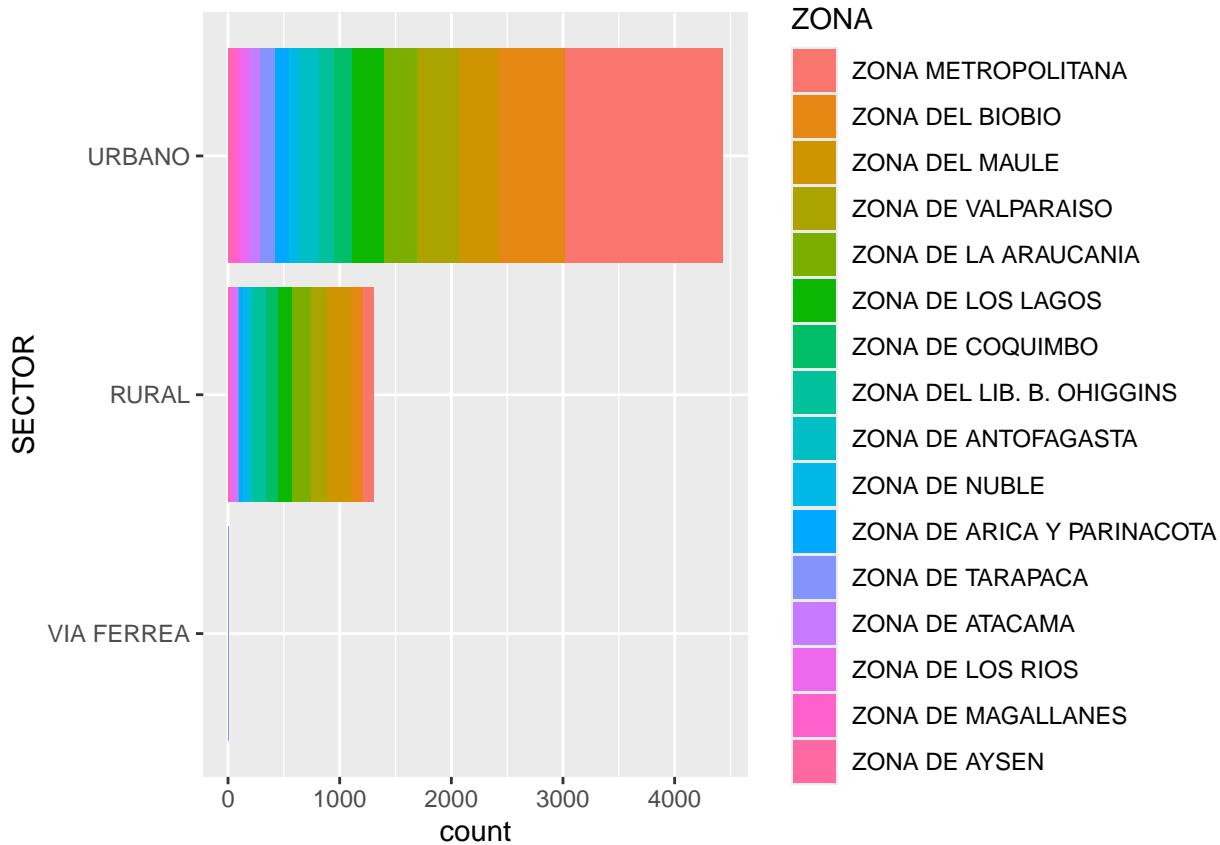
- Tipos predominantes:

- Colisión: 2,936 casos
- Choque: 1,801 casos
- Atropello: 461 casos



- Distribución por sector:

- Urbano: 4,432 casos
- Rural: 1,304 casos
- Vía férrea: 3 casos



## Procesamiento y Normalización de Datos

### Paso a paso

- “FECHA” se convirtió en la concatenación de “FECHA” y “HORA” y se convierte en POSIXct con formato “%d/%m/%Y %H:%M”
- Se elimina la columna “HORA”
- “ZONA” se convierte en factor
- “COMUNA” se convierte en factor
- “TIPO” se convierte en factor
- “CAUSA” se convierte en factor
- “SECTOR” se convierte en factor
- “KM” se convierte en numérico
- “PARTE” se convierte en numérico
- “TRIBUNAL” se convierte en factor
- Se crea la columna “MACRO TIPO” con el valor de “TIPO”, pero unificando los valores de “CHOQUE”, “CHOQUE FRONTAL”, “CHOQUE LATERAL”y “CHOQUE POSTERIOR” como “CHOQUE” y “COLISION”, “COLISION FRONTAL”, “COLISION LATERAL”, “COLISION PERPENDICULAR”, “COLISION POR ALCANCE” como “COLISION”
- “MACRO TIPO” se convierte en factor

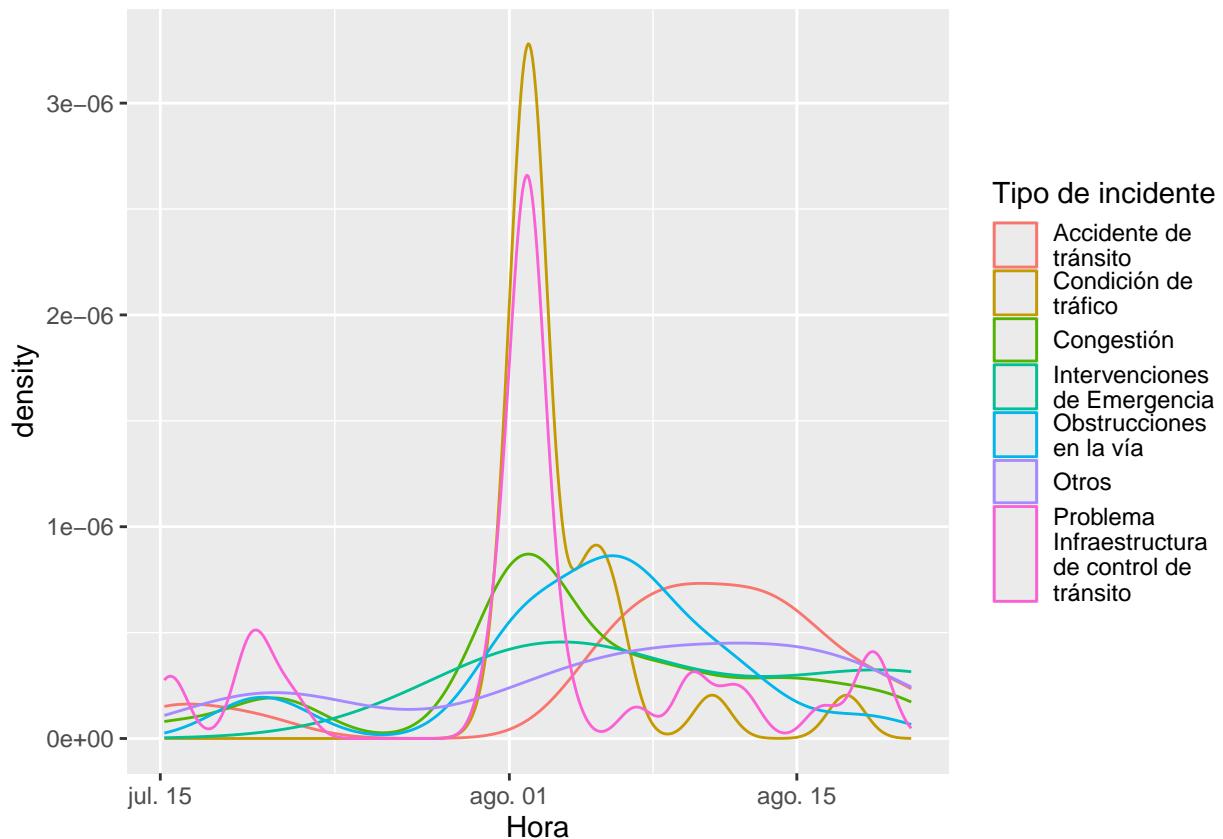
## Incidentes de tráfico radio.xlsx

El dataset contiene 312 registros de incidentes de tráfico reportados a través de sistemas de radio durante un período de 37 días.

## Período de Registro

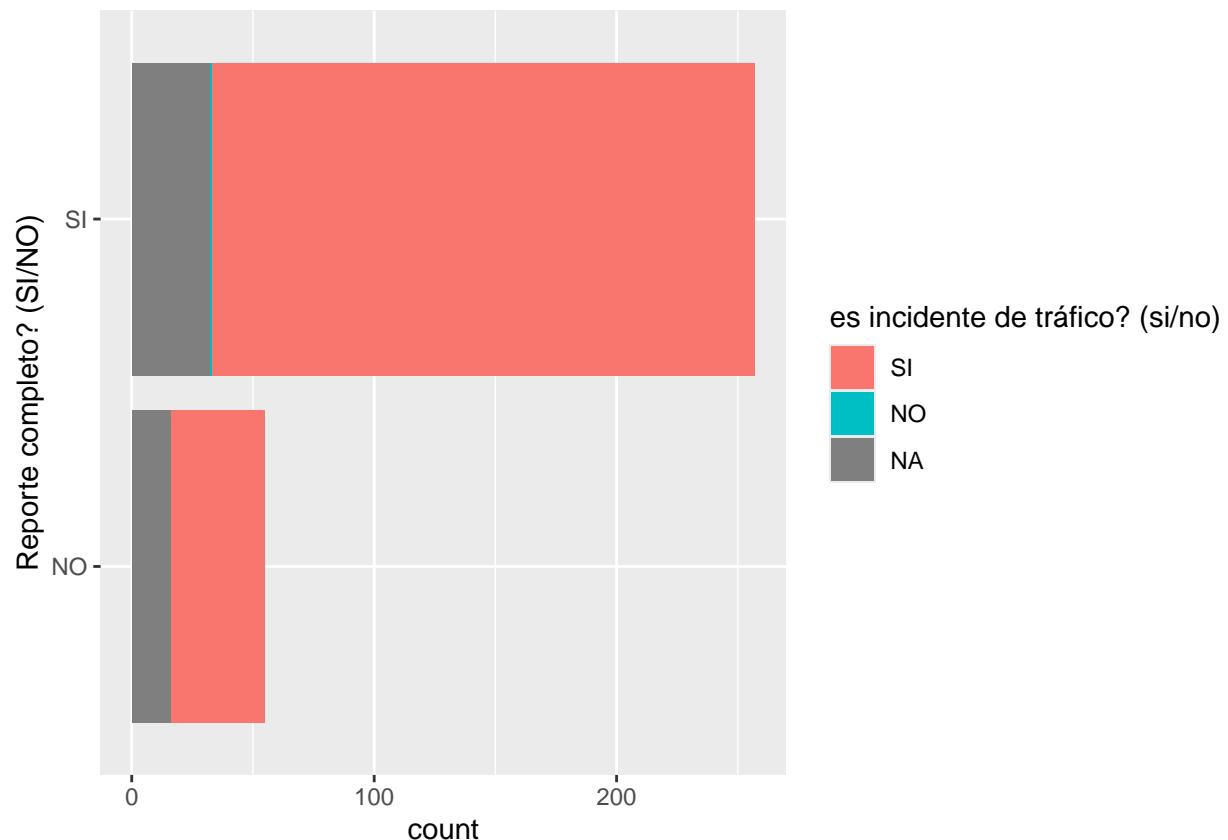
- **Duración:** 15 de julio al 20 de agosto de 2024
- **Cobertura temporal:** 37 días de monitoreo continuo

## Granularidad Temporal



## Clasificación de Incidentes

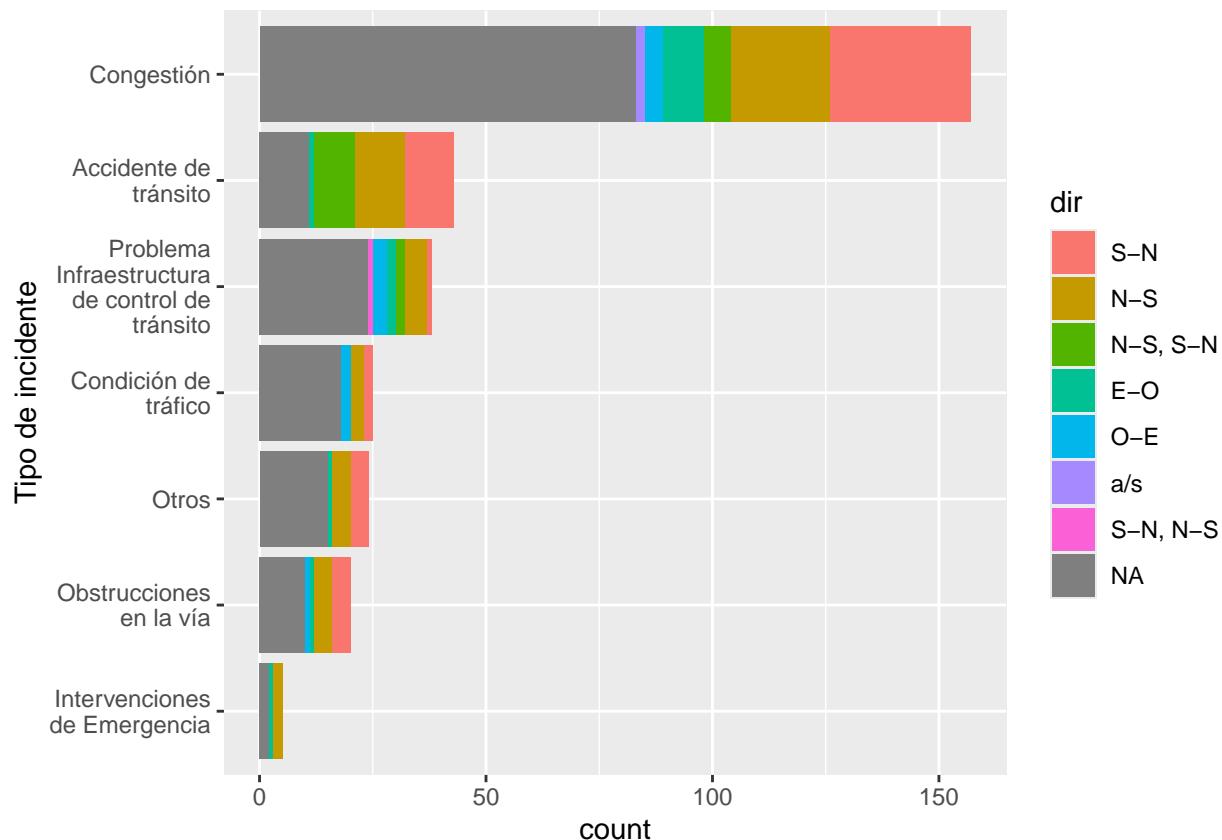
- **Naturaleza del incidente:**
  - **Incidentes de tráfico:** 263 casos
  - **No incidentes de tráfico:** 1 caso
  - **No especificado:** 48 casos
- **Completitud de reportes:**
  - **Reportes completos:** 257 casos
  - **Reportes incompletos:** 55 casos



## Tipología de Incidentes

*Distribución por categoría:*

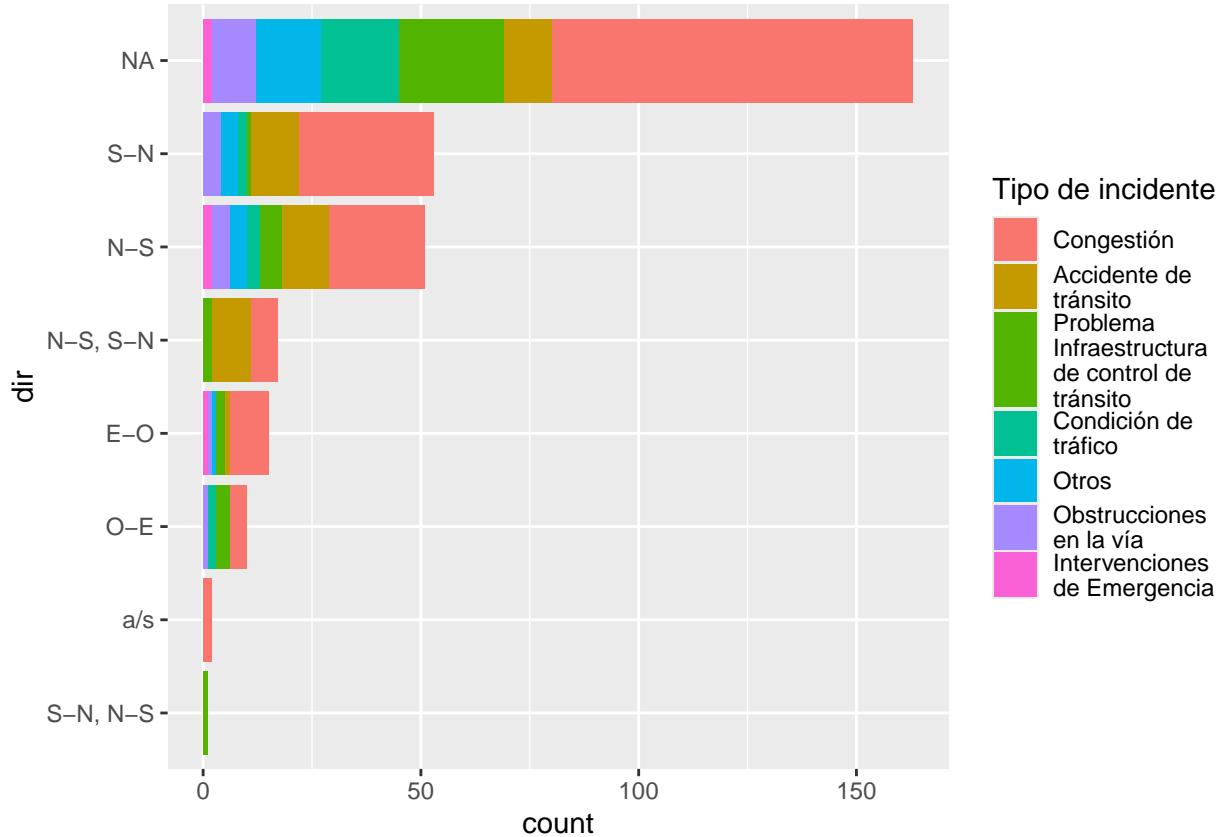
- Congestión: 157 casos
- Accidentes de tránsito: 43 casos
- Problema Infraestructura de control de tránsito: 38 casos
- Condición de tráfico: 25 casos
- Obstrucciones en la vía: 20 casos
- Intervenciones de emergencia: 5 casos
- Otros: 24 casos



### Direccionalidad del Tráfico Afectado

#### Patrones de flujo vehicular impactado:

- Sur a Norte (S-N): 53 casos
- Norte a Sur (N-S): 51 casos
- Bidireccional (N-S, S-N): 17 casos
- Este a Oeste (E-O): 15 casos
- Oeste a Este (O-E): 10 casos
- Ambos sentidos (a/s): 2 casos
- Bidireccional (S-N, N-S): 1 caso
- No especificado: 163 casos



## Procesamiento y Normalización de Datos

### Paso a paso

- Se lee la hoja
- Se lee el color de la columna “tipo de incidente”
- Se crea la columna “fecha” con el nombre de la hoja
- “#ID” se convierte en numérico
- “Hora” se convierte en POSIXct con formato “%H:%M”
- Se unen los datasets en uno
- Se crea la columna “sheet.name” con el nombre de las hojas
- Se crea la columna “original.row” con el número de la fila original
- Se crea la columna “color” con el código del color de la columna “tipo de incidente”
- Se eliminan las filas en las que “coord” o “Reporte completo? (SI/NO)” es nulo
- Se eliminan las columnas “lat”, “tipo de incidente”, “...7”, “...13”, “sheet.name”, “original.row”
- “dir” toma el valor de “...12” cuando “...12” no es nulo y “dir” es nulo
- Se elimina la columna “...12”
- En “Reporte completo? (SI/NO)”, los valores “is”, “si”, “SI” y “su” se unifican como “SI”, los valores “no” y “NO”, como “NO” y “-” se considera nulo
- En “es incidente de tráfico? (si/no)”, los valores “si” y “SI” se unifican como “SI” y los valores “no” y “NO”, como “NO”
- En “dir”, los valores “e-o” y “E-O” se unifican como “E-O”, “o-e” y “O-E”, como “O-E”, “n-s”, “n.s” y “N-S”, como “N-S”, “s-n” y “S-N” como “S-N”, “n-s”, “n-s.s-n” y “N-S, S-N”, como “N-S, S-N”, “s-n.n-s” y “S-N, N-S”, como “S-N, N-S” y “-”, “n/a” y “n/A” se consideran nulos
- “Reporte completo? (SI/NO)” se convierte en factor

- “es incidente de tráfico? (si/no)” se convierte en factor
- “dir” se convierte en factor
- “color” se convierte en factor
- Se crea la columna “Tipo de incidente” a partir de “color”:
  - “Congestión” cuando es nulo
  - “Intervenciones de Emergencia” cuando es FF00B0F0
  - “Congestión” cuando es FF00FFFF
  - “Accidente de tránsito” cuando es FFFF0000
  - “Obstrucciones en la vía” cuando es FFFF00FF
  - “Problema Infraestructura de control de tránsito” cuando es FFFF6600
  - “Condición de tráfico” cuando es FFFFC000
  - “Congestión” cuando es FFFFFFFF00
  - “Otros” cuando es FF00B050
- Se elimina la columna “color”
- Cuando “Hora” es nulo, toma el valor de la fila anterior
- “coord” se separa en “lat” y “lng”, separado por “,\s\*” y se convierten en numéricos
- Se eliminan las filas en las que “lat” es nulo
- “Hora” se convirtió en la concatenación de “fecha” y “hora” y se convierte en POSIXct con el formato “%Y-%m-%d %H:%M”

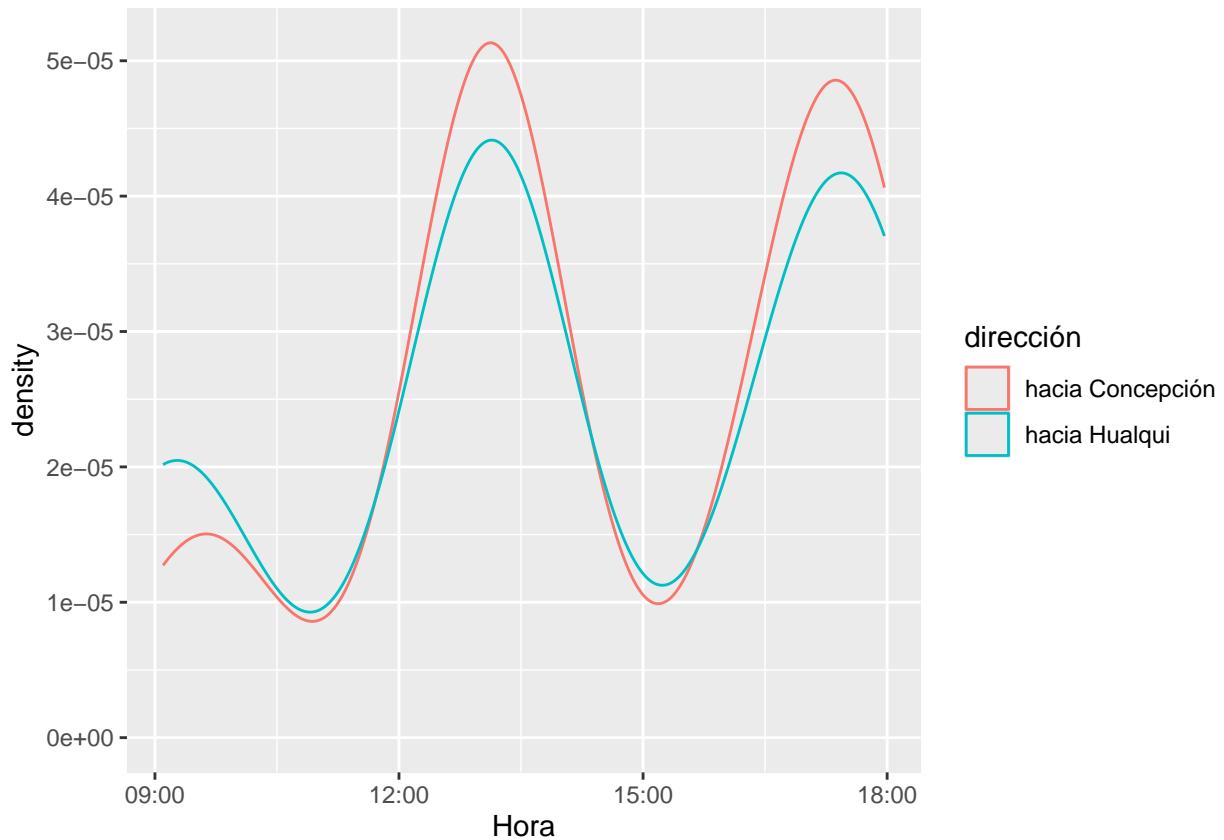
## Med velo CHIGUAYANTE.xlsx

El dataset contiene 338 mediciones de velocidad vehicular capturadas en el paradero “Bdo O’Higgins - Arauco” de Chiguayante, monitoreando el flujo entre Concepción y Hualqui.

### Contexto Operacional

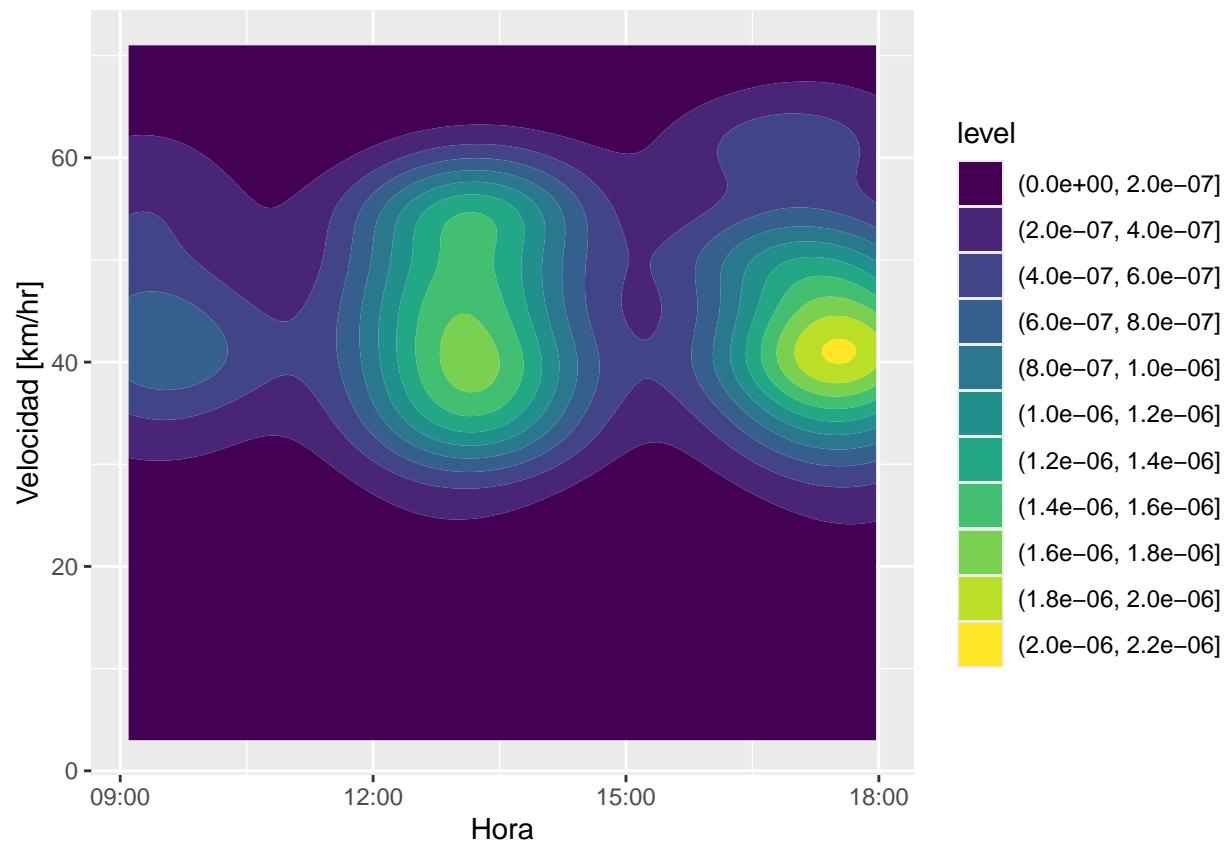
- **Ubicación:** Paradero “Bdo O’Higgins - Arauco”, Chiguayante
- **Corredor vial:** Concepción - Hualqui
- **Período de medición:** 26 de junio de 2024 entre las 09:06 y las 17:58 horas
- **Duración:** Aproximadamente 9 horas de monitoreo continuo

## Granularidad Temporal



## Análisis de Velocidades

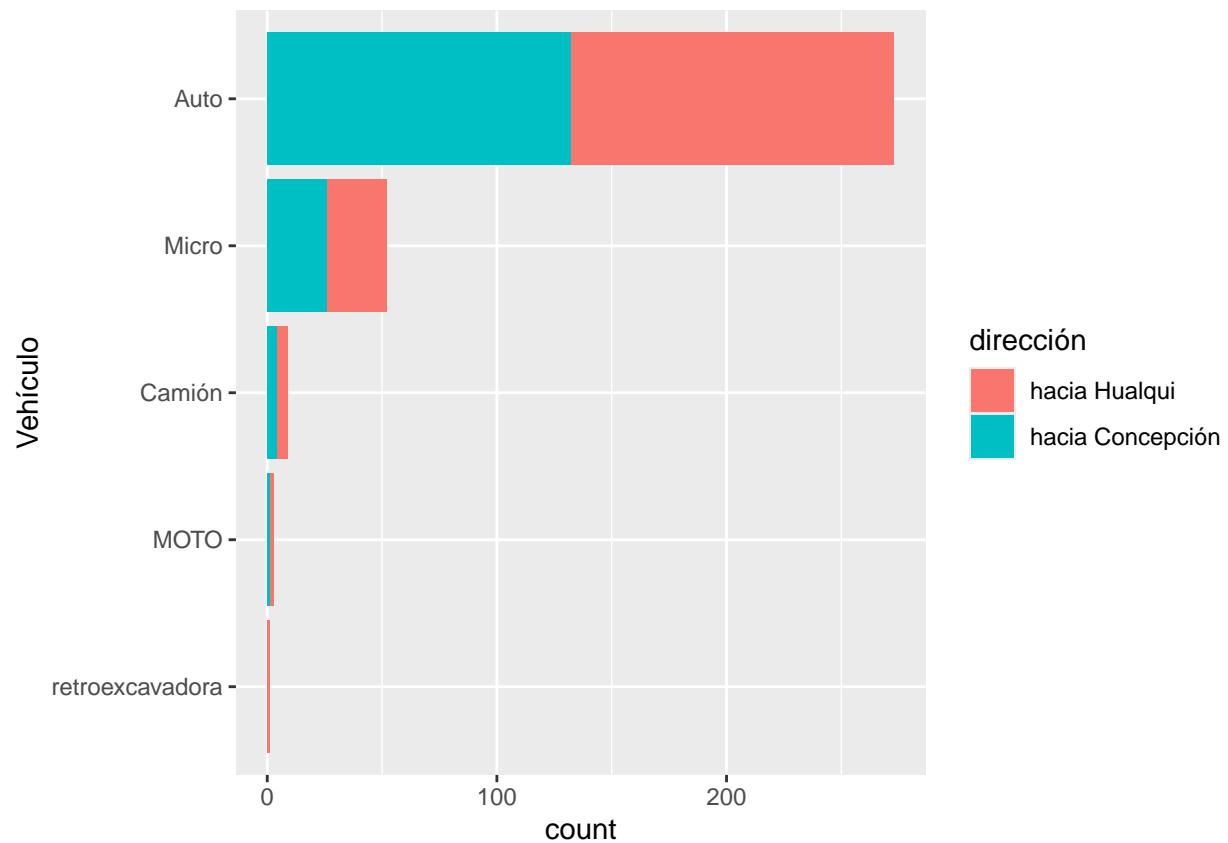
- Rango de velocidades: 3 a 71 km/h
- Velocidad promedio: 44.27 km/h
- Velocidad mediana: 43 km/h
- Distribución: Relativamente simétrica alrededor de la mediana



### Composición Vehicular

**Tipos de vehículos identificados:**

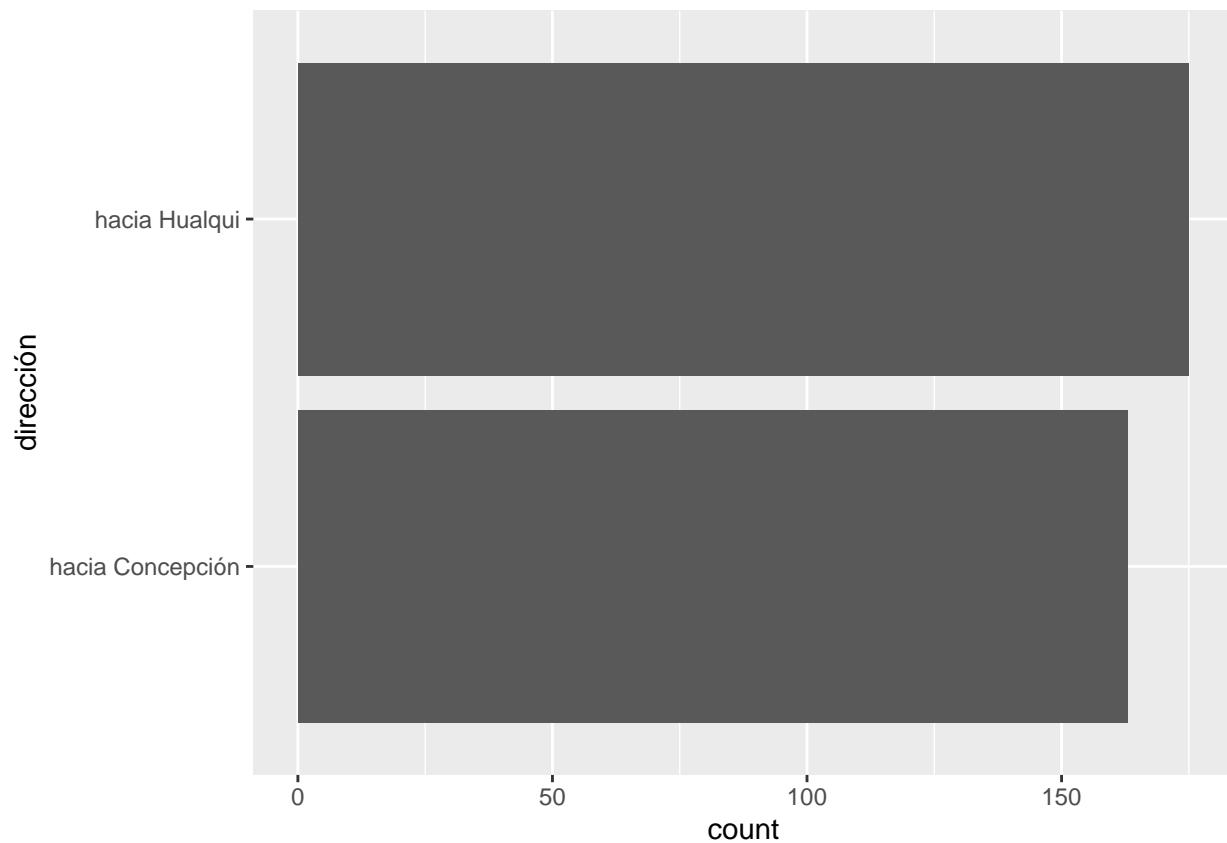
- **Automóviles:** 273 registros
- **Microbuses:** 52 registros
- **Camiones:** 9 registros
- **Motos:** 3 registros
- **Retroexcavadora:** 1 registro



## Patrones de Movimiento

### Distribución direccional:

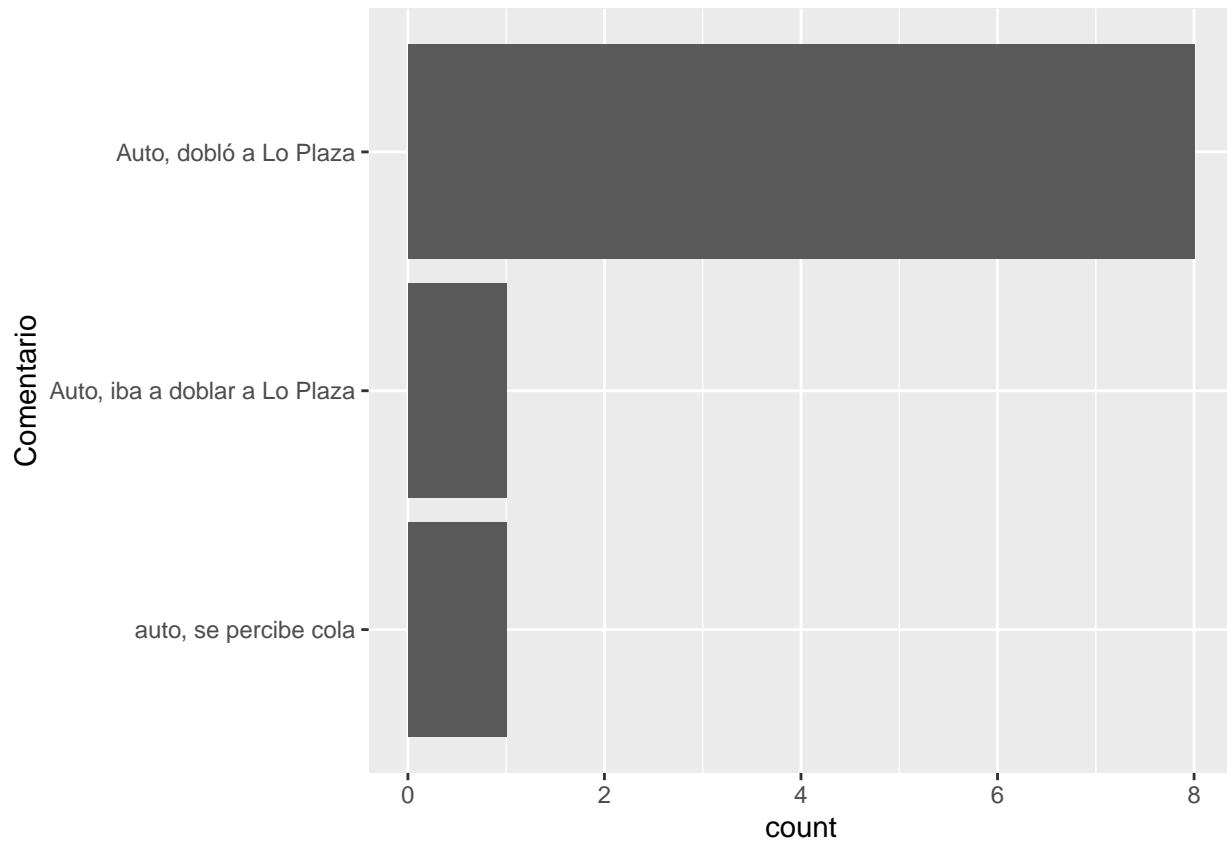
- **Hacia Concepción:** 163 mediciones
- **Hacia Hualqui:** 175 mediciones
- **Balance:** Ligero predominio del flujo hacia Hualqui



### Comportamientos Específicos

#### Maniobras documentadas:

- “Auto, dobló a Lo Plaza”: 8 casos
- “Auto, iba a doblar a Lo Plaza”: 1 caso
- “auto, se percibe cola”: 1 caso



## Procesamiento y Normalización de Datos

### Paso a paso

- Se lee la hoja
- Se crea la columna “dirección” con el nombre de la hoja
- La columna “Lugar (dirección)” se unen en un único valor que se arrastra hacia todas las filas de la hoja
- Se unen los datasets
- Se eliminan las columnas “...5”, “...6”, “...7”, “...8”, “...9”, “...10”, “...11”, “...12” y “...13”
- Se eliminan las filas en las que “Velocidad [km/hr]” es nulo
- Las filas donde “Hora” es nulo toman el valor de la fila anterior
- La columna “Hora” se convierte en la concatenación entre “2024-06-25” y “Hora” y se convierte en un POSIXct con formato “%Y-%m-%d %H:%M:%S”
- En la columna “Comentario”, los valores “auto, doble a Lo Plaza”, “Auto, doble a Lo Plaza” y “Auto, gira a Lo Plaza” se unifican como “Auto, dobló a Lo Plaza”, los valores “camion” y “Camion”, como “Camión” y “micro” y “Micro”, como “Micro”
- “Comentario” se convierte en factor
- “dirección” se convierte en factor
- “Lugar (dirección)” se convierte en factor
- Se crea la columna “Vehículo” que toma el valor de la columna “Comentario” si su valor es “Camión”, “Micro”, “MOTO” o “retroexcavadora” y “Auto” en otro caso
- “Vehículo” se convierte en factor

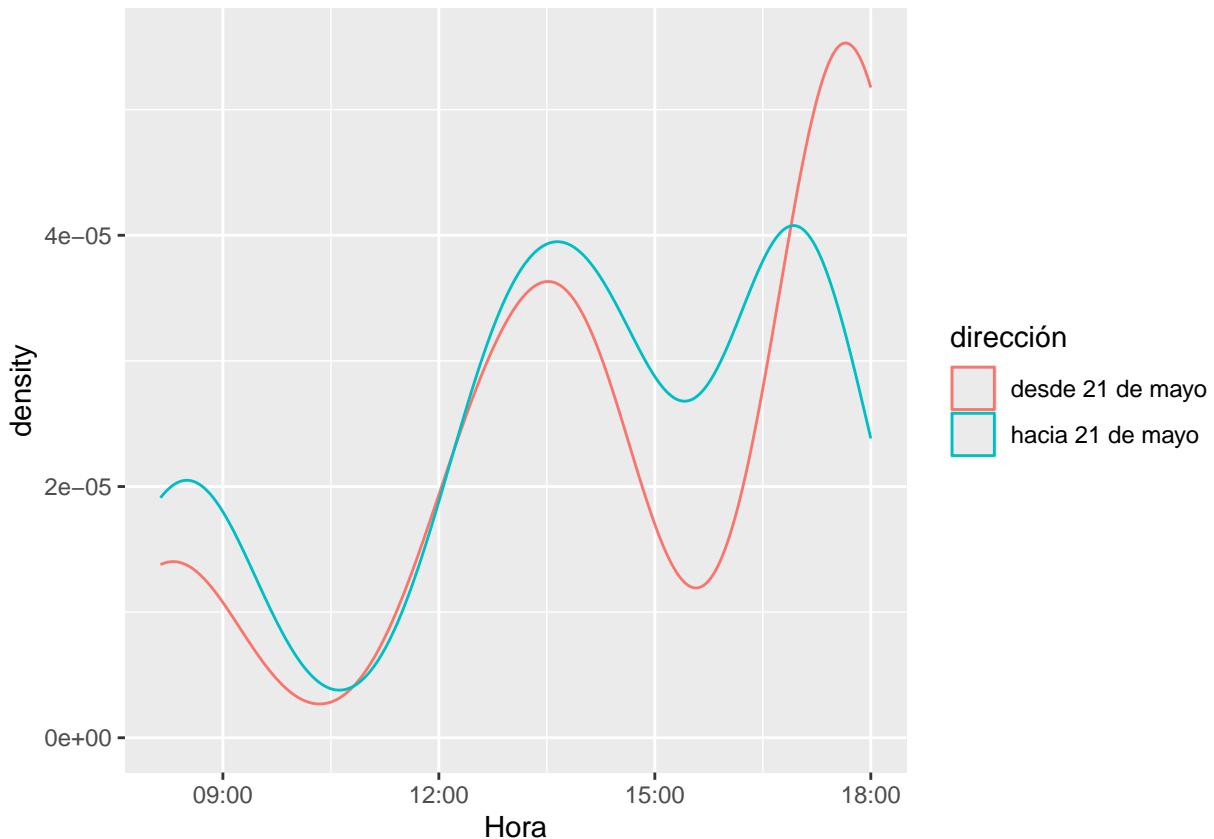
## Med velo LA VEGA.xlsx

El dataset contiene 432 mediciones de velocidad vehicular capturadas en el cruce de Briceños con Miraflores, monitoreando el flujo entre 21 de Mayo y Avenida Costanera a través del sector Miraflores.

### Contexto Operacional

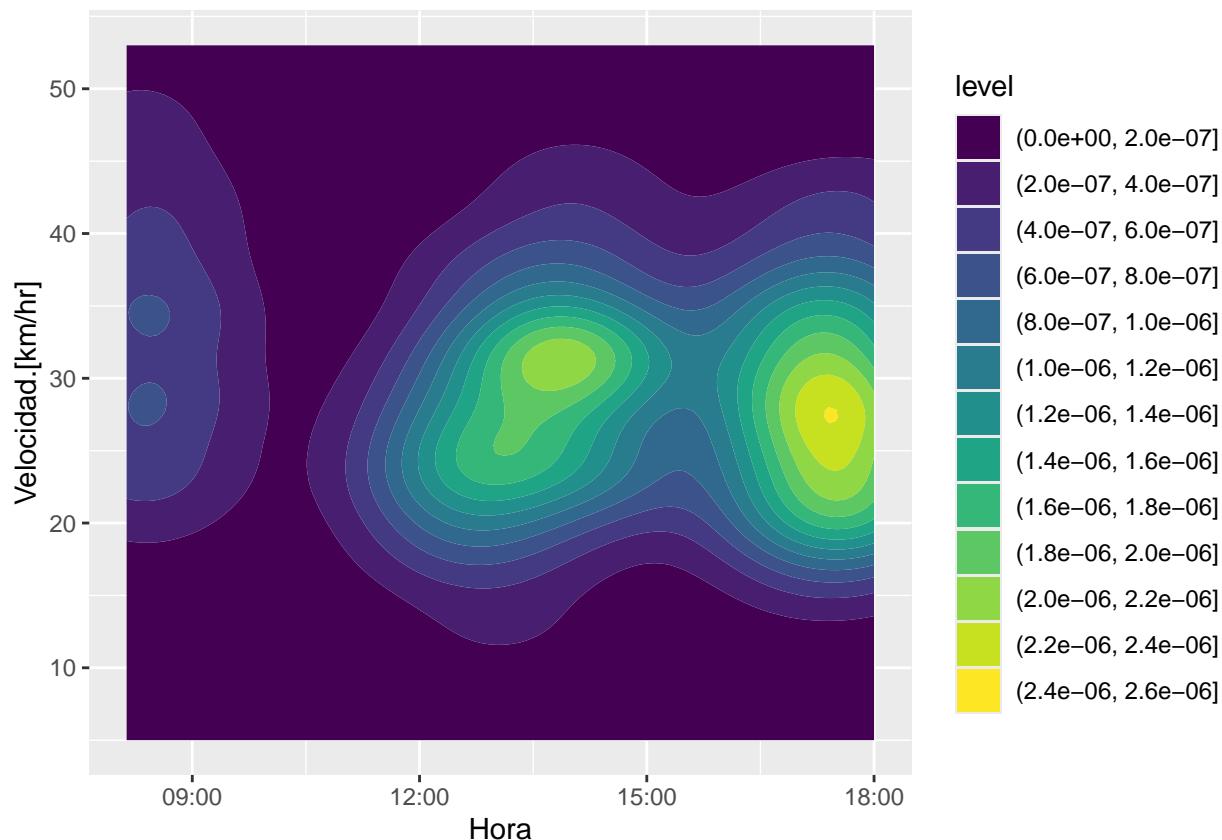
- **Ubicación:** Esquina Briceños con Miraflores, La Vega
- **Corredores viales:**
  - 21 de Mayo -> Briceños -> Miraflores
  - Av. Costanera -> Miraflores -> 21 de Mayo
- **Período de medición:** 25 de junio de 2024 entre las 08:08 y las 17:58 horas
- **Duración:** Aproximadamente 10 horas de monitoreo continuo

### Granularidad Temporal



### Análisis de Velocidades

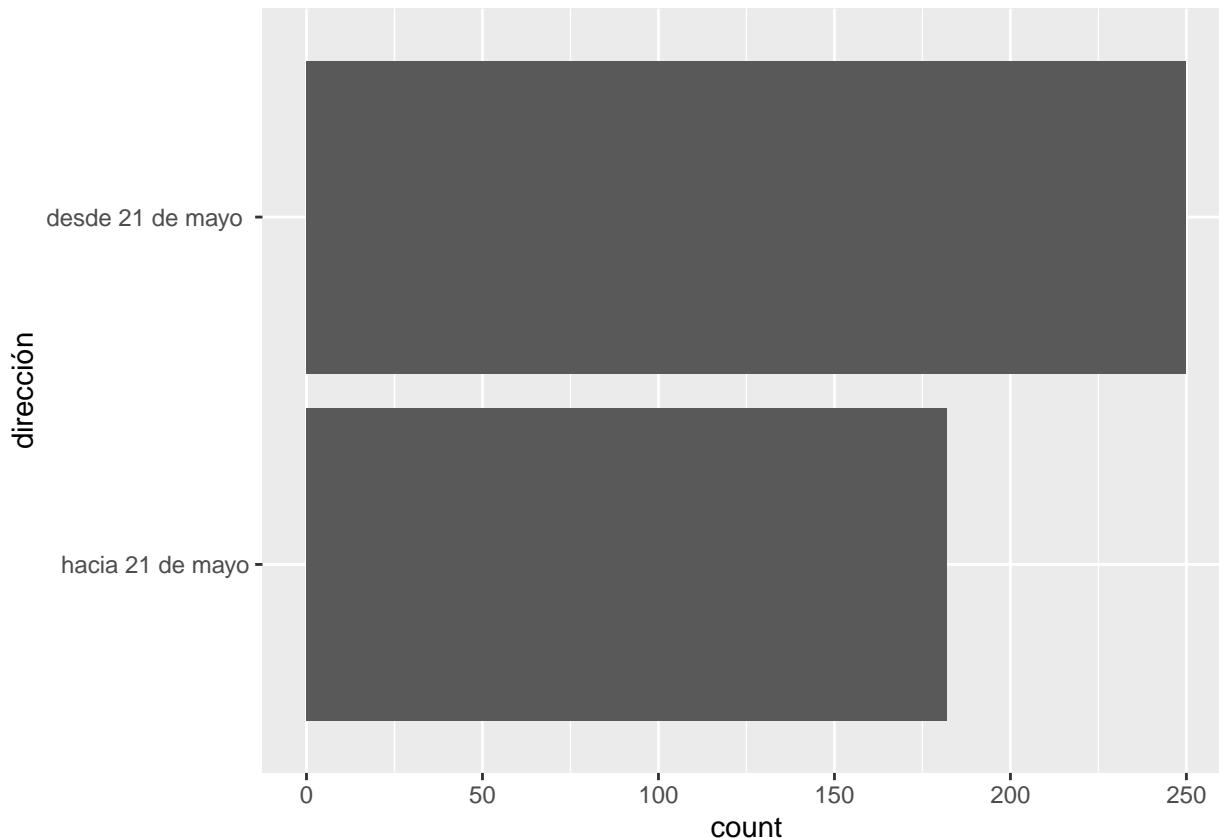
- **Rango de velocidades:** 5 a 53 km/h
- **Velocidad promedio:** 28.87 km/h
- **Velocidad mediana:** 29 km/h
- **Distribución:** Concentrada en velocidades bajas-medias, típicas de tráfico urbano congestionado



## Patrones de Movimiento

### Distribución direccional:

- Desde **21 de Mayo**: 250 mediciones
- Hacia **21 de Mayo**: 182 mediciones
- Balance: Predominio del flujo desde 21 de Mayo



## Fenómenos de Congestión Documentados

### Patrones de formación de colas:

- “Llegan juntos”: 49 casos - agrupamiento vehicular sincronizado
- “Llega a cola”: 22 casos - incorporación a colas existentes
- “Forman cola”: 14 casos - generación de nuevas colas

### Composición vehicular adicional:

- **Microbuses:** Registros específicamente identificados
- **Motocicletas:** Registros categorizados separadamente

## Procesamiento y Normalización de Datos

### Paso a paso

- Se lee la hoja
- En la columna “Hora”, se cambian las “,” por “.” (ej: “8,15” -> “8:15”)
- Las filas donde “Hora” es nulo toman el valor de la fila anterior
- La columna “Hora” se convierte en la concatenación entre “2024-06-25” y “Hora” y se convierte en un POSIXct con formato “%Y-%m-%d %H:%M:%S”
- Se crea la columna “dirección” con el nombre de la hoja

- La columna “Lugar (dirección)” se unen en un único valor que se arrastra hacia todas las filas de la hoja
- Se unen los datasets
- Se eliminan las filas en las que “Velocidad.[km/hr]” es nulo
- En la columna “Comentario”, los valores “llega a cola”, “llegan a cola”, “llegan a cola, aprox 4 vehiculos” y “llegan a cola formada antes de medir” se unifican como “Llega a cola”, los valores “llegan juntos”, “llegan juntos”, “llegan juntos, cola aprox 8 a 10 autos” y “llegan juntos, generan cola”, como “Llegan juntos”, los valores “micro” y “Micro”, como “Micro” y los valores “moto” y “Moto”, como “Moto”
- “Comentario” se convierte en factor
- “Lugar (dirección)” se convierte en factor
- “dirección” se convierte en factor

## tb\_gps\_historial\_eventos\_202509161626(1).csv

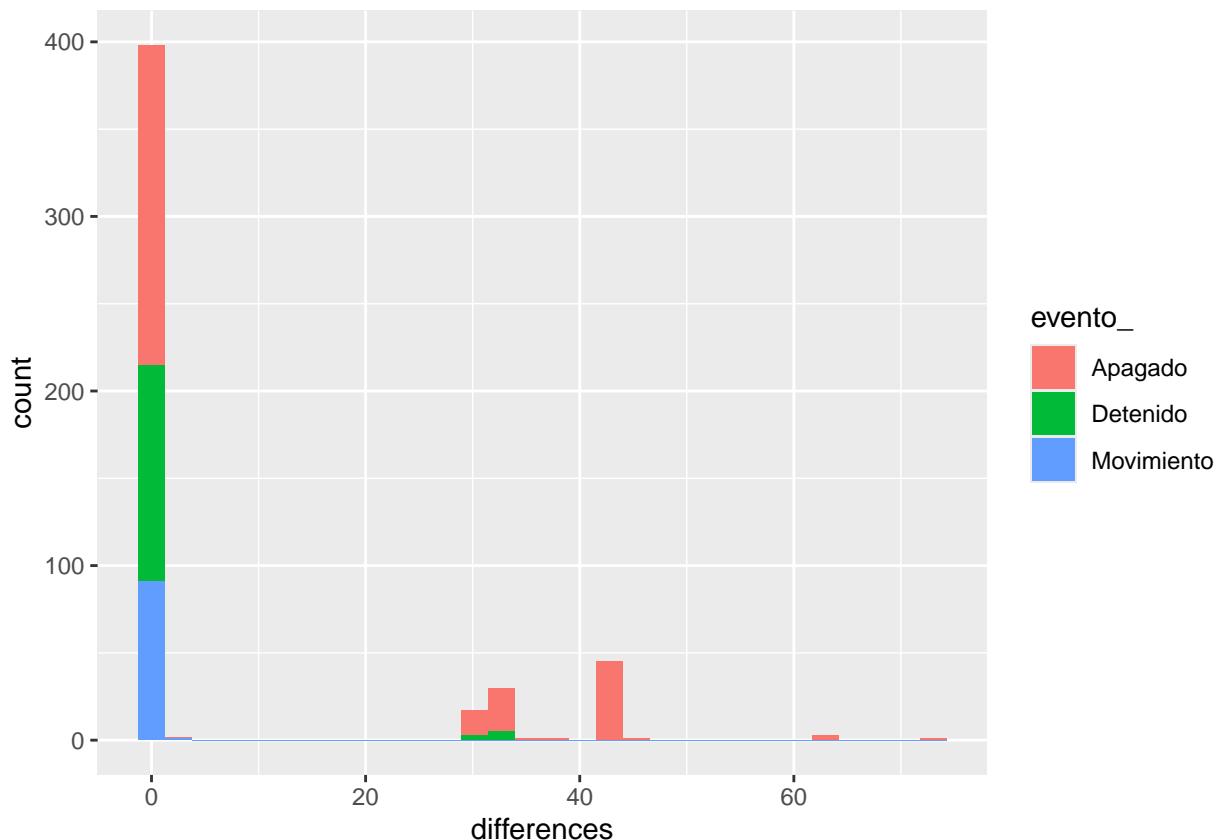
El dataset contiene 1000 eventos registrados por sistemas GPS vehiculares, capturando transiciones de estado en una ventana temporal específica.

### Período de Monitoreo

- **Fecha de referencia:** Transición año nuevo 2024-2025
- **Según timestamp del auto:** 1 de enero de 2025, 00:00:06 - 01:06:29 (1 hora, 6 minutos)
- **Según timestamp del GPS:** 31 de diciembre de 2024, 23:59:22 - 1 de enero de 2025, 01:06:49 (1 hora, 7 minutos)
- **Discrepancia temporal:** Diferencia de sincronización entre sistemas de aproximadamente 1 minuto

### Granularidad Temporal

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##  0.000  0.000  1.000   7.982  1.000  73.000       1
## 'stat_bin()' using 'bins = 30'. Pick better value 'binwidth'.
```



## Distribución de Estados Vehiculares

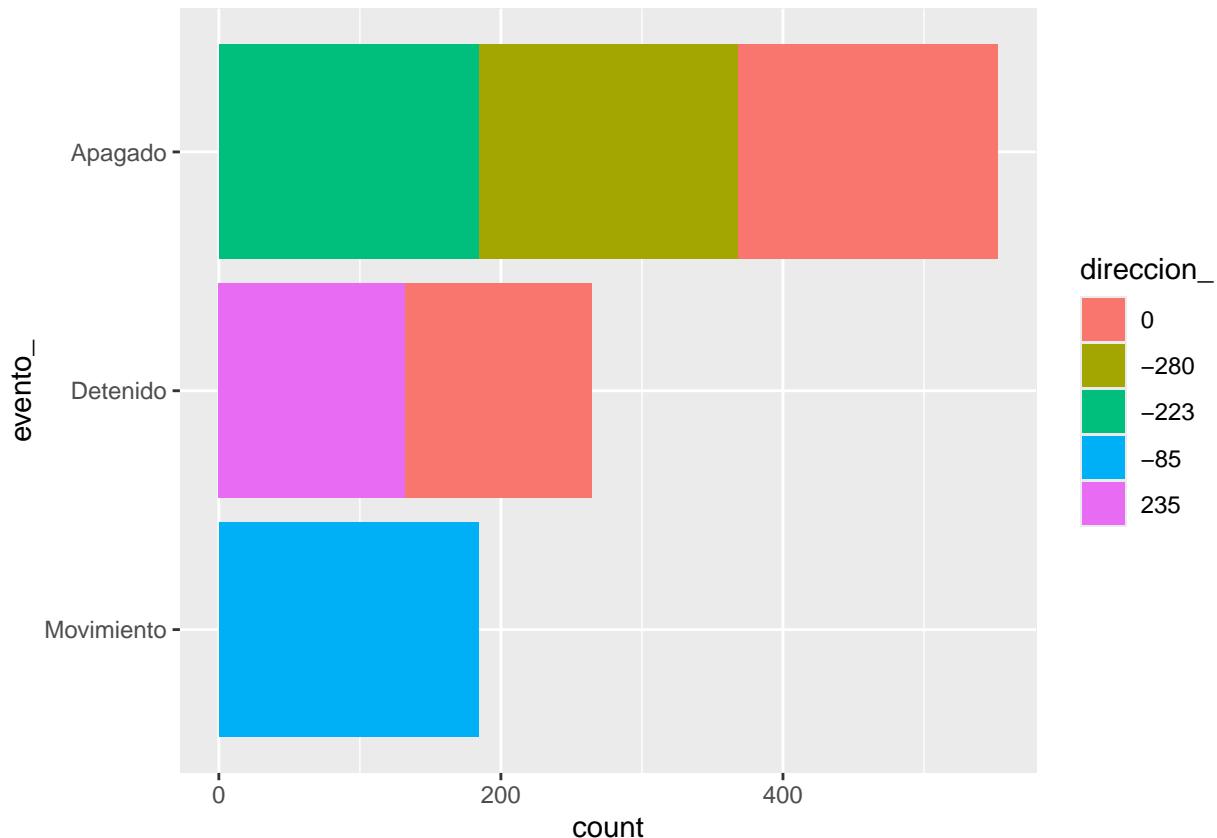
### Tipos de eventos registrados:

- **Apagado:** 552 eventos - vehículo sin operación
- **Detenido:** 264 eventos - vehículo inmovilizado pero encendido
- **Movimiento:** 184 eventos - vehículo en desplazamiento

### Patrones de Direccionalidad

### Valores de dirección registrados:

- 0: 316 casos
- -280: 184 casos
- -223: 184 casos
- -85: 184 casos
- 235: 132 casos



**Observación:** Los valores negativos y positivos sugieren un sistema de coordenadas o ángulos específico.

### Aspectos Requieren Investigación

#### Variables Críticas por Clarificar

- `id_evento`: Valor constante 2 en todos los registros
- `direccion_`:
  - Sistema de representación angular (grados) o coordenadas relativas
  - Significado de valores negativos vs. positivos
  - Relación con puntos cardinales o sistema de referencia

### Procesamiento y Normalización de Datos

#### Paso a paso

- Se eliminan las columnas “ubicacion\_” y “velocidad\_” debido a que todos sus valores son nulos
- “fecha\_auto” se convierte en un POSIXct con formato “%Y-%m-%d %H:%M:%S”
- “fecha\_gps\_” se convierte en un POSIXct con formato “%Y-%m-%d %H:%M:%S”
- “evento\_” se convierte en factor
- “direccion\_” se convierte en factor

## Inventario CCTV Biobío(1).xlsx

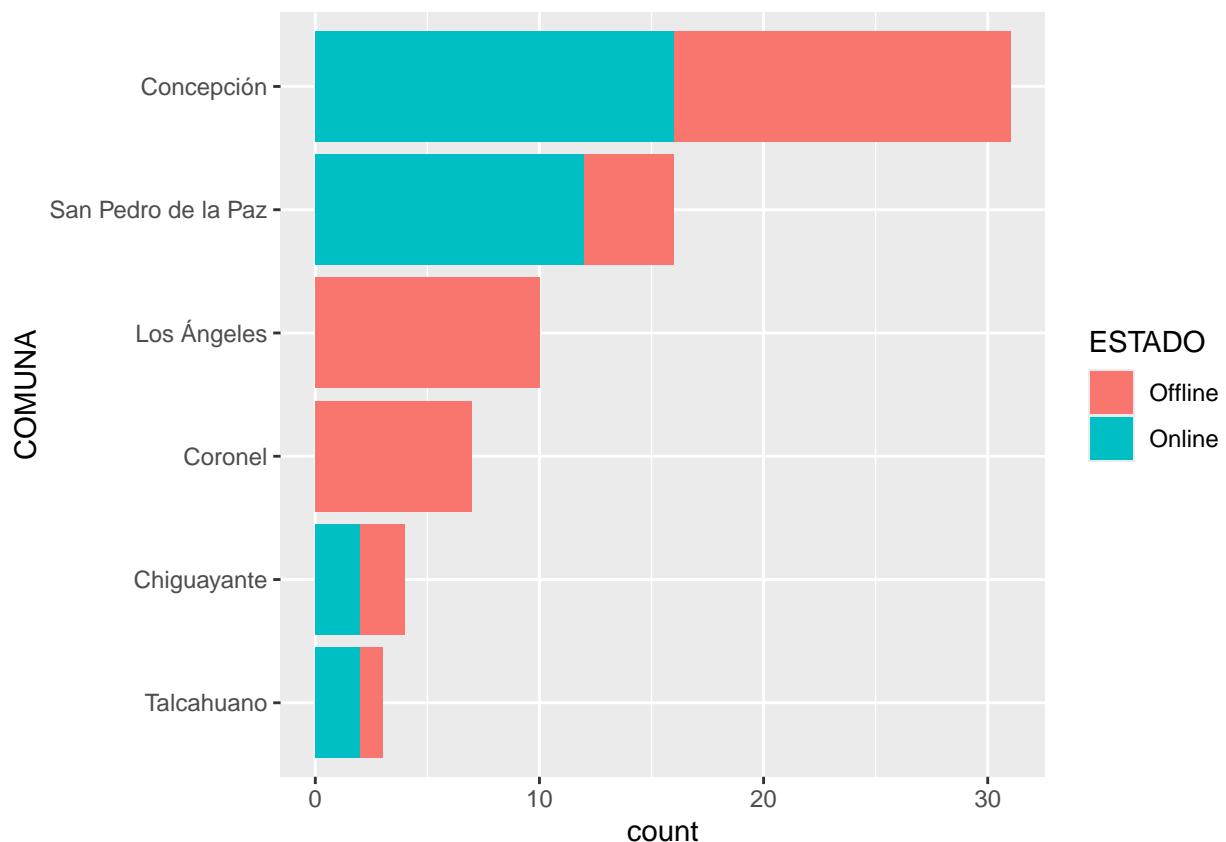
El inventario documenta 71 ubicaciones de cámaras de vigilancia y sus respectivos sistemas de codificación de video desplegados en la Región del Biobío, representando la infraestructura de monitoreo vial regional.

### Distribución Geográfica

- Comunas con mayor cobertura:
  - Concepción: 31 cámaras
  - San Pedro de la Paz: 16 cámaras
  - Los Ángeles: 10 cámaras
- Cobertura regional: Múltiples comunas del Biobío con sistemas CCTV

### Estado Operacional del Sistema

- Cámaras online: 32 unidades
- Cámaras offline: 39 unidades
- Disponibilidad general: Sistema operando con 45% de disponibilidad inmediata



### Características Técnicas de la Infraestructura

#### Conectividad y Proveedores

- **Tipo de enlace:** 100% digital (71/71 cámaras)
- **Proveedor de servicios:** Exclusivamente “Red Comunicaciones Propia” (71/71)

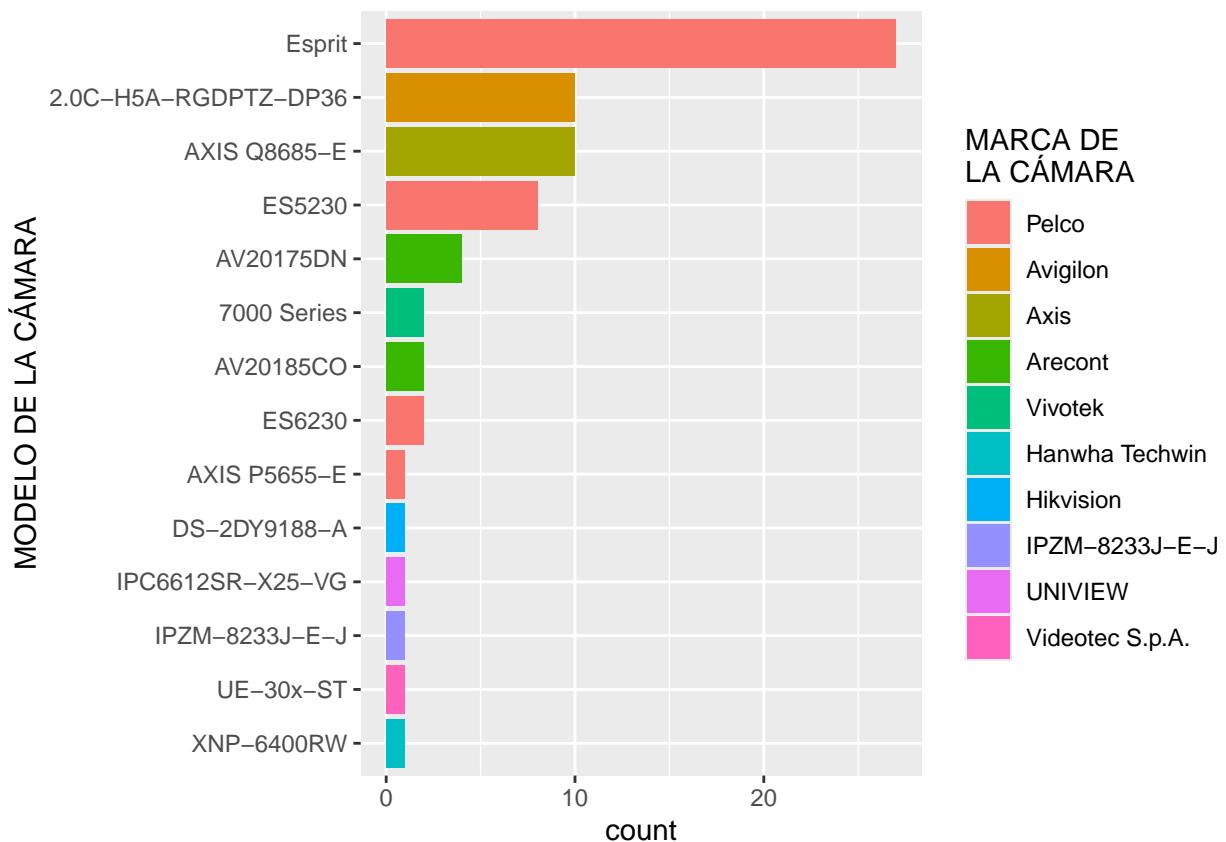
## Especificaciones de Cámaras

*Marcas predominantes:*

- Pelco: 38 cámaras
- Avigilon: 10 cámaras
- Axis: 10 cámaras

*Modelos más frecuentes:*

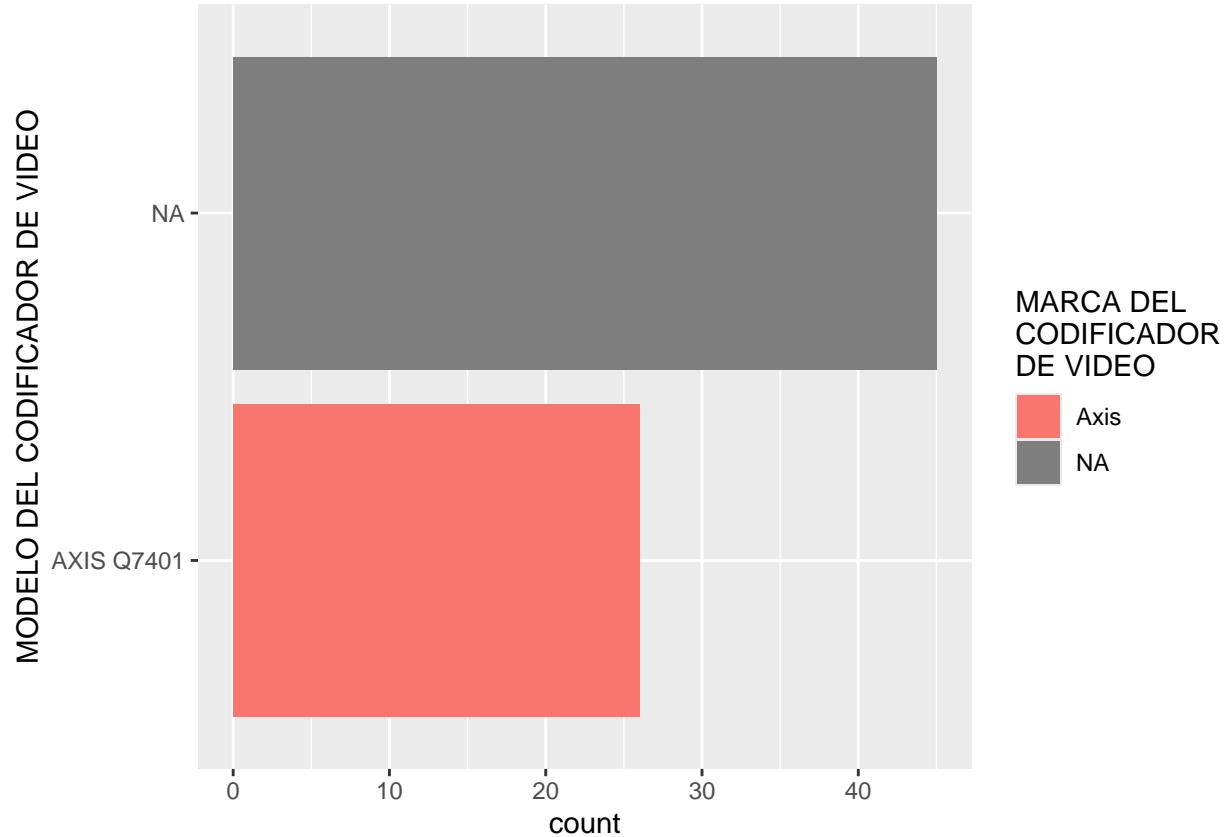
- Pelco Esprit: 27 unidades
- Avigilon 2.0C-H5A-RGDPTZ-DP36: 10 unidades
- Axis AXIS Q8685-E: 10 unidades



## Sistema de Codificación de Video

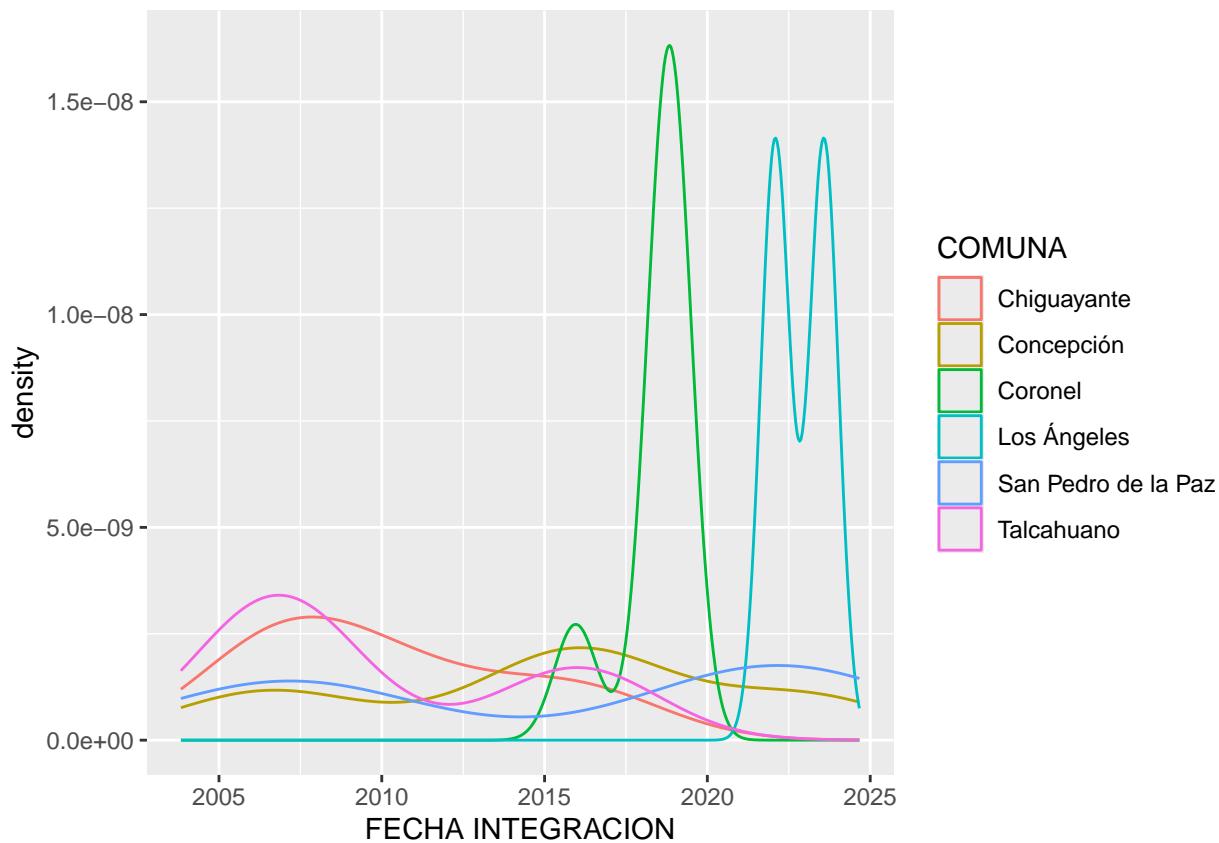
*Codificadores identificados:*

- Axis AXIS Q7401: 26 unidades
- Sin especificar: 45 unidades - posiblemente integrados en cámaras o no documentados



#### Evolución Temporal del Sistema

- **Rango de integración:** 1 de noviembre de 2003 - 1 de septiembre de 2024
- **Vida útil del sistema:** Más de 20 años de despliegue progresivo
- **Actualizaciones recientes:** Integraciones hasta septiembre 2024



## Procesamiento y Normalización de Datos

### Paso a paso

- Las columnas “MARCA”, “MODELO” y “NÚMERO DE SERIE” estaban duplicadas, por lo que se les agregó “DE LA CÁMARA” y “DEL CODIFICADOR DE VIDEO” para distinguirlos
- “COMUNA” se convierte en factor
- “ESTADO” se convierte en factor
- “TIPO DE ENLACE” se convierte en factor
- “PROVEEDOR” se convierte en factor
- “MARCA DE LA CÁMARA” se convierte en factor
- “MODELO DE LA CÁMARA” se convierte en factor
- “MARCA DEL CODIFICADOR DE VIDEO” se convierte en factor
- “MODELO DEL CODIFICADOR DE VIDEO” se convierte en factor
- “REGIÓN” se convierte en factor

## Alertas de Tráfico.csv

El dataset comprende 281759 alertas de tráfico generadas por usuarios de Waze, representando un año completo de reportes ciudadanos sobre condiciones viales en el Gran Concepción.

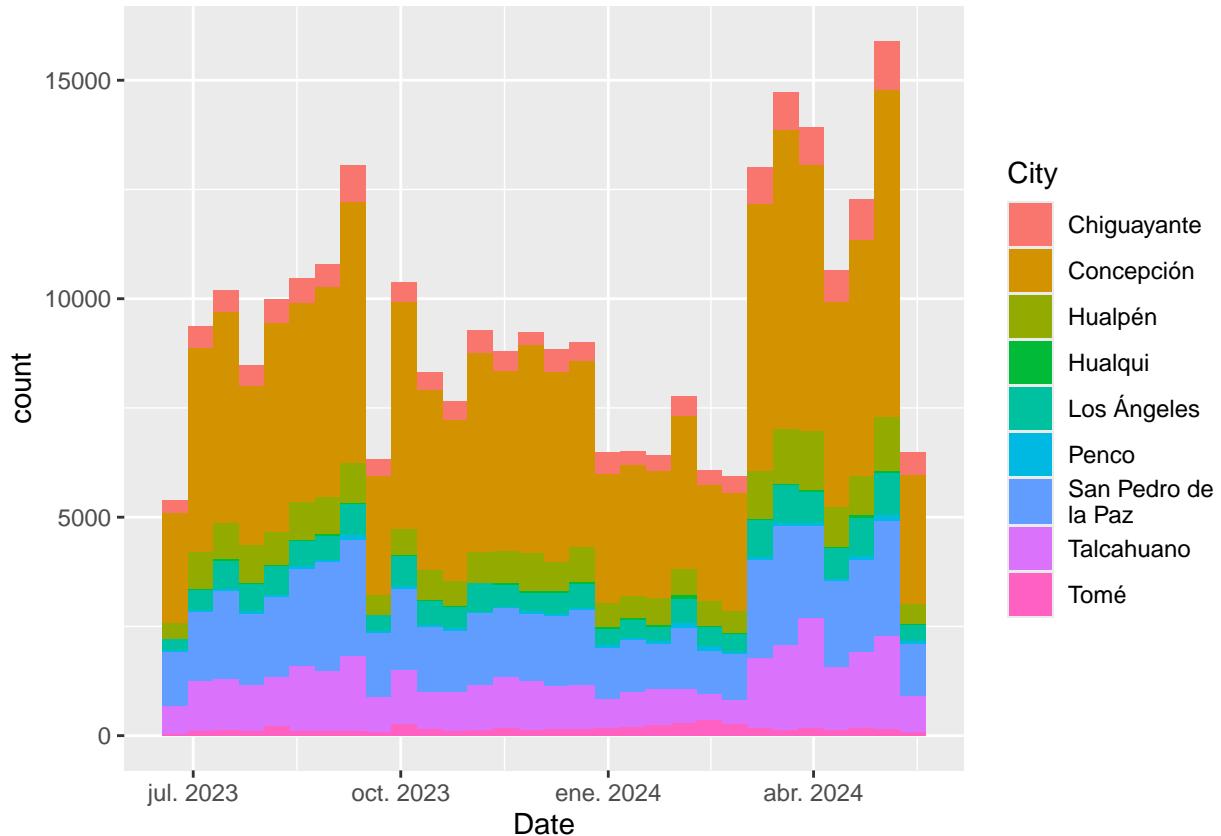
### Período de Análisis

- **Cobertura temporal:** 23 de junio de 2023 - 15 de mayo de 2024

- **Duración:** 327 de datos continuos
- **Registros diarios promedio:** 861.65 alertas por día

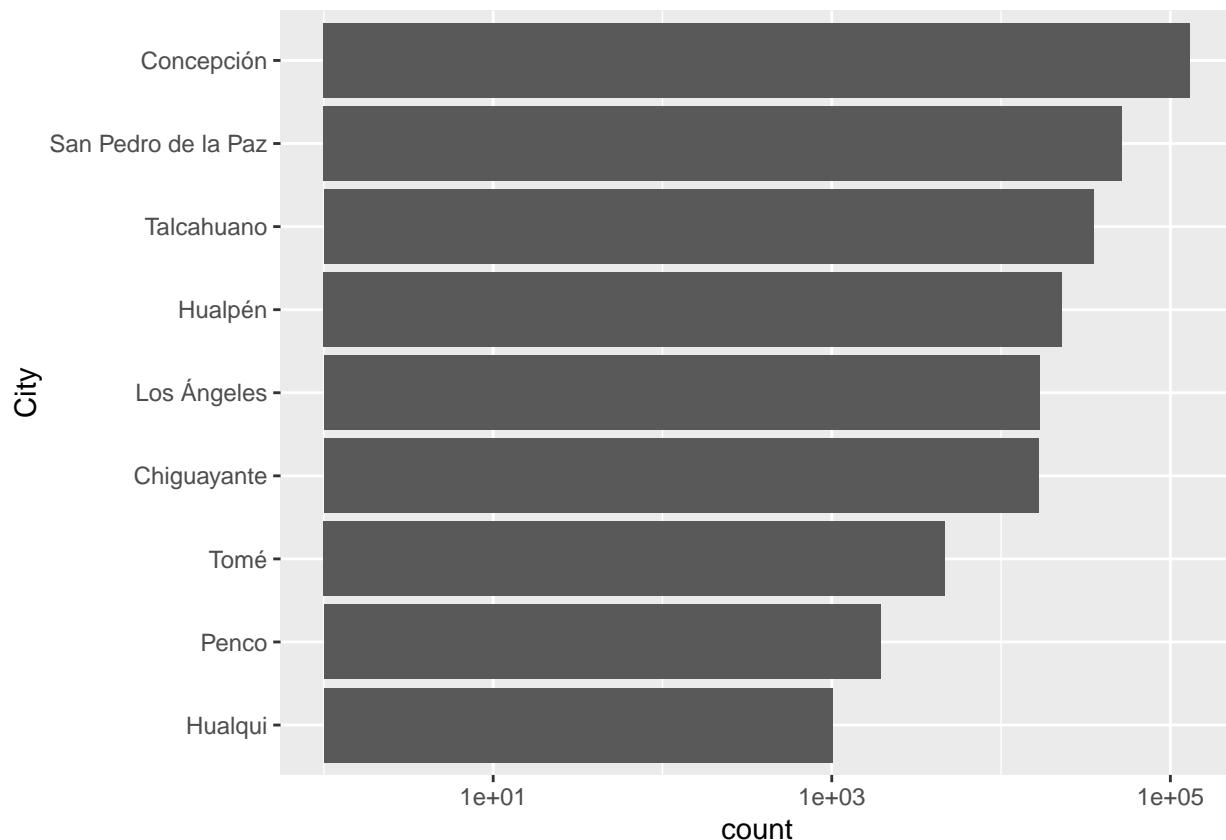
### Granularidad Temporal

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```



### Contexto Geográfico

- **Distribución por ciudad:**
  - Concepción: 130747 alertas
  - San Pedro de la Paz: 51816 alertas
  - Talcahuano: 35122 alertas
  - Otras comunas: 64074 alertas



## Clasificación de Alertas

### Categorías Principales (Type)

- Embottellamiento (JAM): 178785 casos
- Peligro (HAZARD): 53457 casos
- Peligro meteorológico (WEATHERHAZARD): 35984 casos
- Otras categorías: 13533 casos

### Especificaciones por Subcategoría (Subtype)

*Embottellamientos (JAM):*

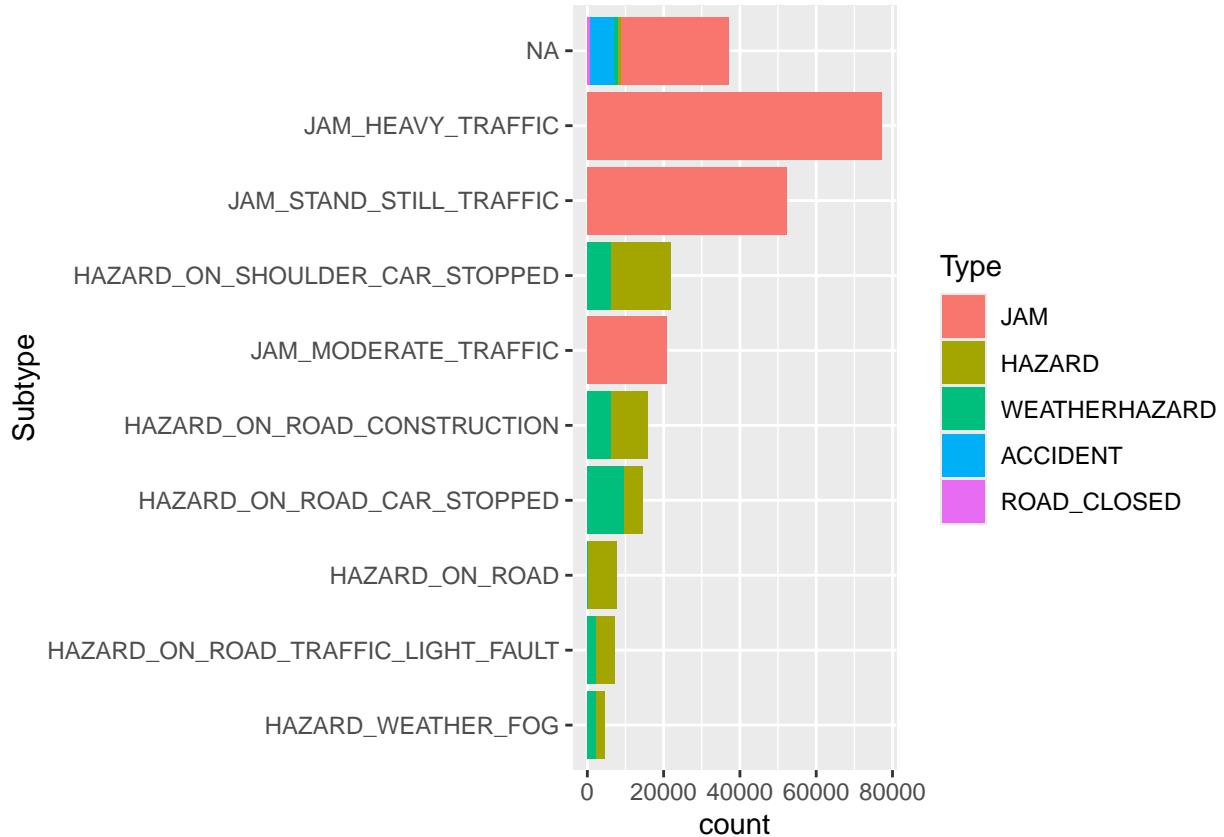
- Tráfico intenso (JAM\_HEAVY\_TRAFFIC): 77179 casos
- Tráfico paralizado (JAM\_STAND\_STILL\_TRAFFIC): 52298 casos
- Tráfico moderado (JAM\_MODERATE\_TRAFFIC): 20947 casos

*Peligros (HAZARD):*

- Vehículo detenido en franja lateral (HAZARD\_ON\_SHOULDER\_CAR\_STOPPED): 15869 casos
- Construcción en la vía (HAZARD\_ON\_ROAD\_CONSTRUCTION): 9458 casos
- Peligro en la vía (HAZARD\_ON\_ROAD): 7494 casos
- Otros peligros: 20636 casos

*Peligros meteorológicos (WEATHERHAZARD):*

- Vehículo detenido (HAZARD\_ON\_ROAD\_CAR\_STOPPED): 9467 casos
- Construcción en la vía (HAZARD\_ON\_ROAD\_CONSTRUCTION): 6287 casos
- Vehículo detenido en franja lateral (HAZARD\_ON\_SHOULDER\_CAR\_STOPPED): 6097 casos
- Otros peligros meteorológicos: 14133 casos



## Procesamiento y Normalización de Datos

### Paso a paso

- La columna “Location” se separa en “lng” y “lat” mediante los regex “(?<=Point\\() -?\\d+\\\\.?\\d\*” y “-?\\d+\\\\.?\\d\*(?=\\\$)”, respectivamente
- Se elimina la columna “Location”
- En la columna “Date”, se reemplazan los valores “ene”, “feb”, “mar”, “abr”, “may”, “jun”, “jul”, “ago”, “sep”, “oct”, “nov” y “dic” por “jan”, “feb”, “mar”, “apr”, “may”, “jun”, “jul”, “aug”, “sep”, “oct”, “nov” y “dec”, respectivamente
- “Country” se convierte en factor
- “City” se convierte en factor
- “Type” se convierte en factor
- “Subtype” se convierte en factor

## Copia de Accidentes.csv

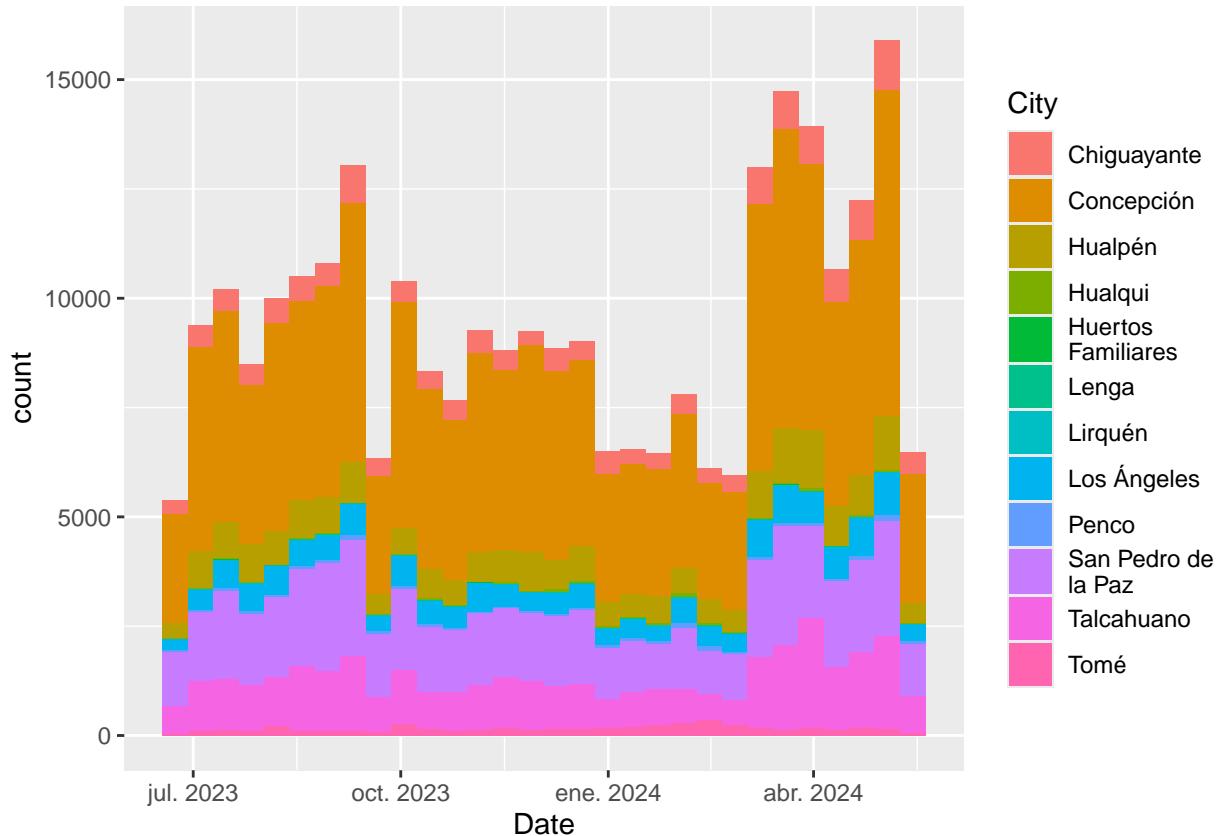
El dataset contiene 281981 registros de accidentes de tránsito, proporcionando una visión detallada de la siniestralidad vial en el Gran Concepción durante un período de 11 meses.

## Período de Monitoreo

- Cobertura temporal: 23 de junio de 2023 - 15 de mayo de 2024
- Duración: 327 días de reportes continuos
- Densidad de datos: ~862.33 accidentes reportados por día en promedio

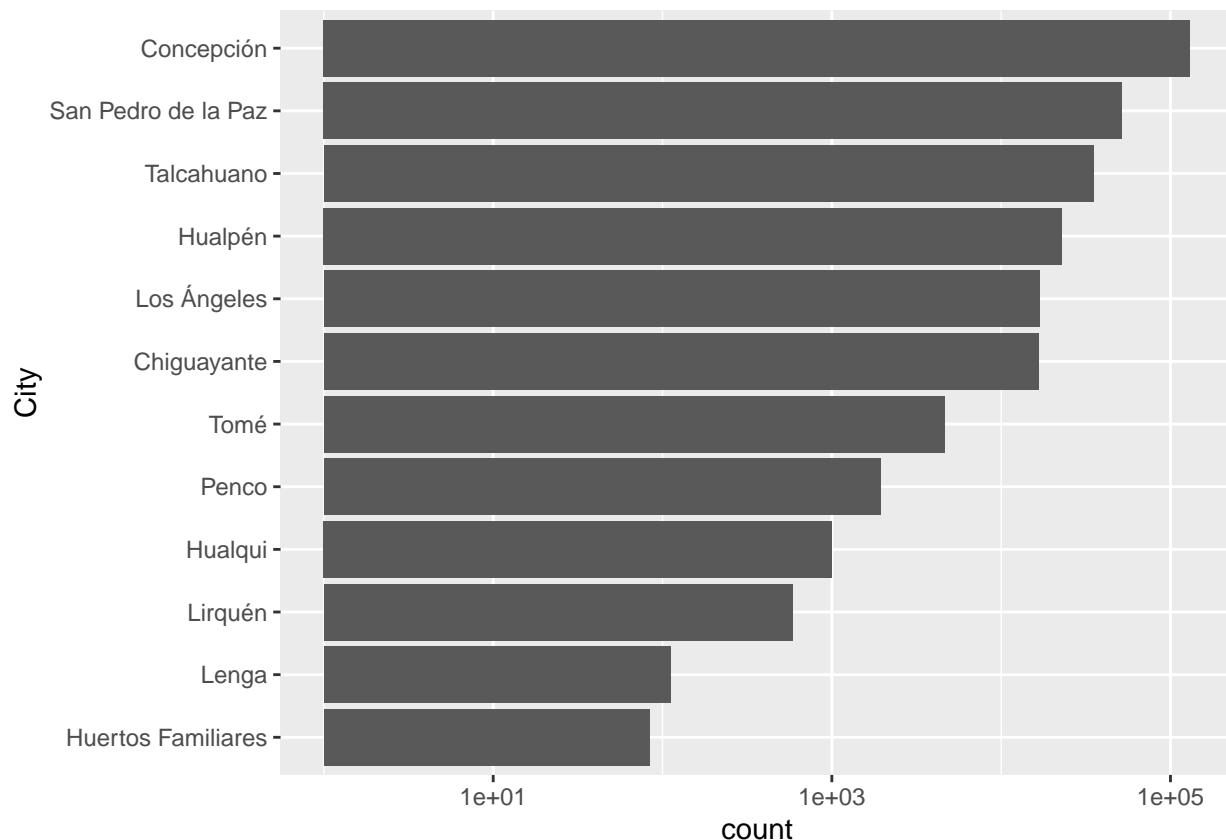
## Granularidad Temporal

```
## 'stat_bin()' using 'bins = 30'. Pick better value 'binwidth'.
```



## Contexto Geográfico

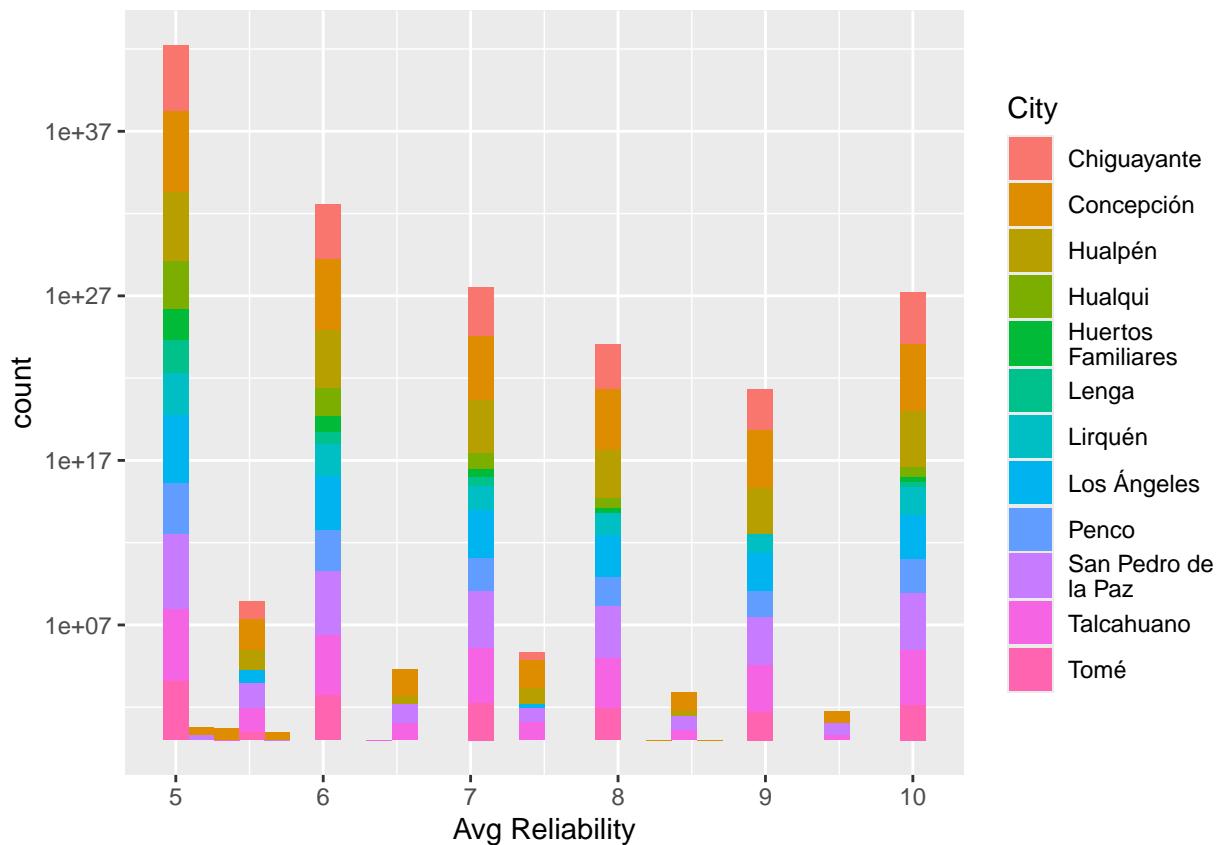
- Distribución por ciudad:
  - Concepción: 130516 alertas
  - San Pedro de la Paz: 51704 alertas
  - Talcahuano: 35059 alertas
  - Otras comunas: 64702 alertas



## Métricas de Confiabilidad

- **Rango de fiabilidad:** 5 a 10 puntos (escala no especificada)
- **Fiabilidad promedio:** 5.82 puntos
- **Fiabilidad mediana:** 5 puntos
- **Distribución:** Sesgada hacia valores bajos-medios de la escala

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```



## Procesamiento y Normalización de Datos

### Paso a paso

- La columna “Location” se separa en “lng” y “lat” mediante los regex “(?<=Point\\()-.?\\d+\\.?.?\\d\*” y “-?\\d+\\.?.?\\d\*(?=\\)\$”, respectivamente
- Se elimina la columna “Location”
- En la columna “Date”, se reemplazan los valores “ene”, “feb”, “mar”, “abr”, “may”, “jun”, “jul”, “ago”, “sep”, “oct”, “nov” y “dic” por “jan”, “feb”, “mar”, “apr”, “may”, “jun”, “jul”, “aug”, “sep”, “oct”, “nov” y “dec”, respectivamente
- “Country” se convierte en factor
- “City” se convierte en factor

## Waze for Cities Data Key Alerts Dashboard\_Traffic Irregularities\_Tabla(1).csv

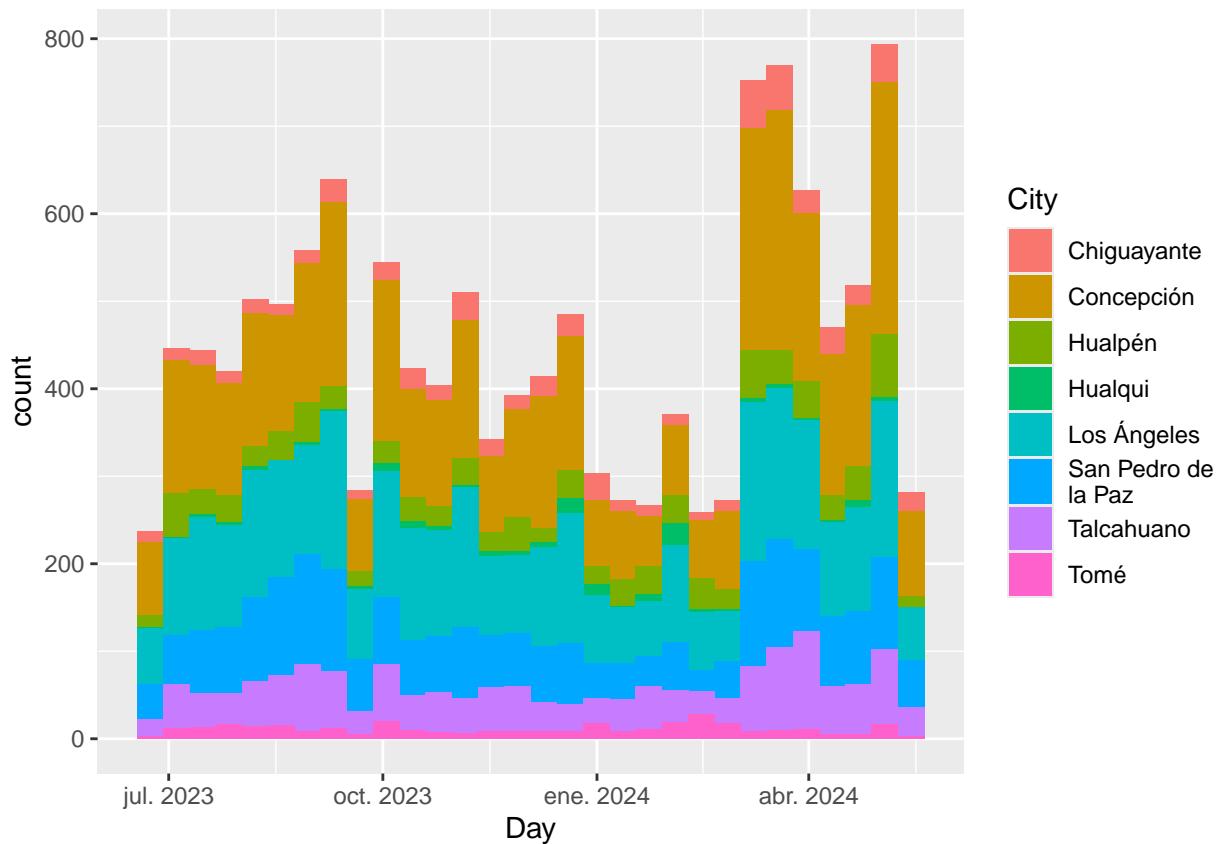
El dataset contiene 13502 registros de irregularidades de tráfico documentadas a través de la plataforma Waze for Cities, proporcionando métricas detalladas sobre el impacto de incidentes viales en la Región del Biobío durante un período de 11 meses.

### Período de Análisis

- Cobertura temporal:** 23 de junio de 2023 - 15 de mayo de 2024
- Duración:** 327 días de monitoreo continuo
- Frecuencia promedio:** 41.29 irregularidades reportadas por día

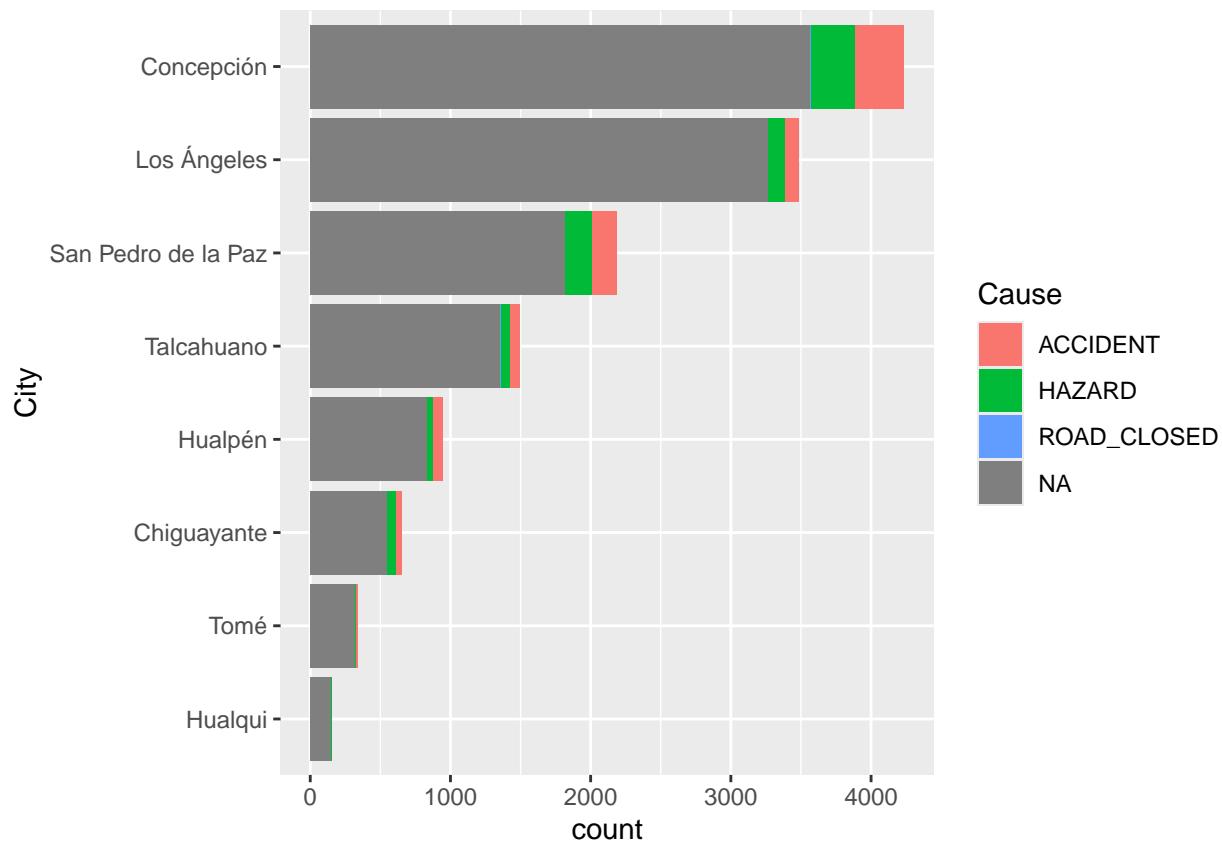
## Granularidad Temporal

```
## 'stat_bin()' using 'bins = 30'. Pick better value 'binwidth'.
```



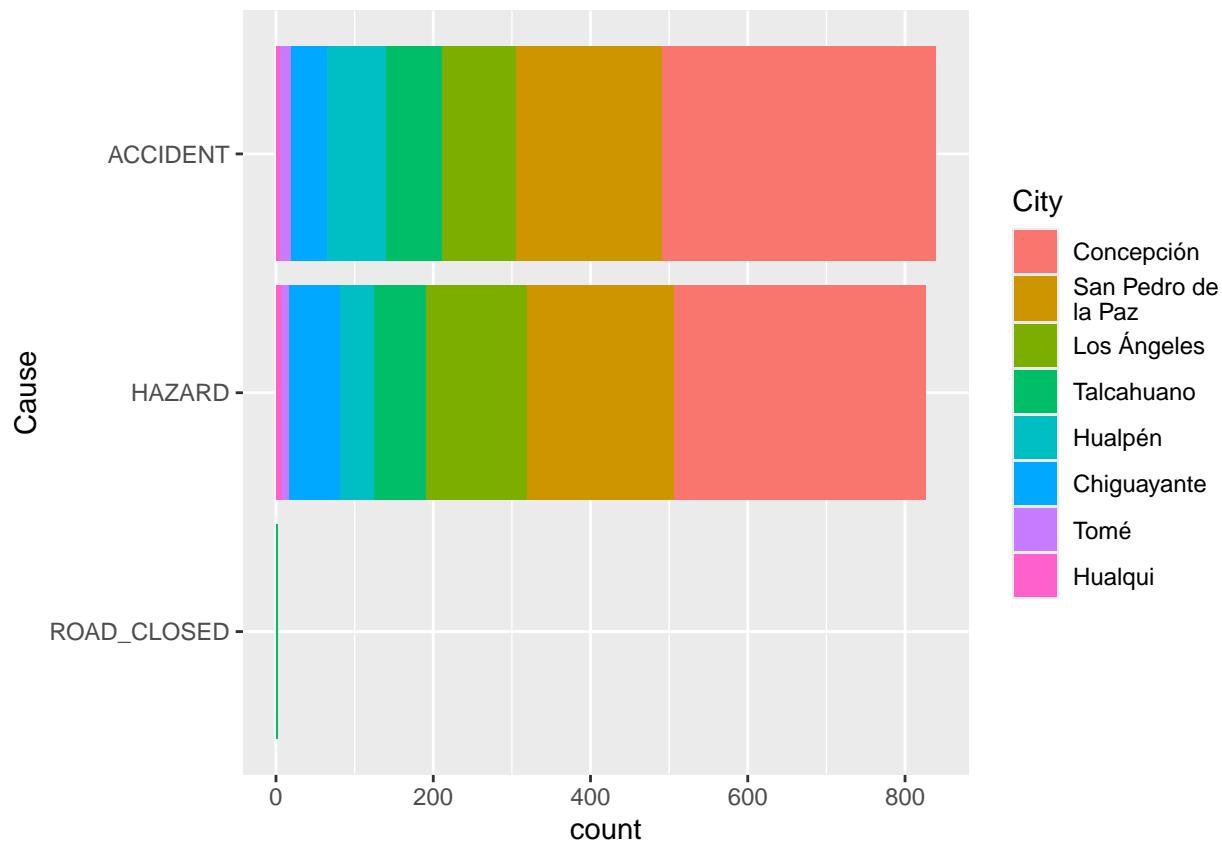
## Distribución Geográfica

- Concentración por ciudad:
  - Concepción: 4236 registros
  - Los Ángeles: 3484 registros
  - San Pedro de la Paz: 2190 registros
  - Otras localidades: 3592 registros



#### Clasificación por Causas Identificadas

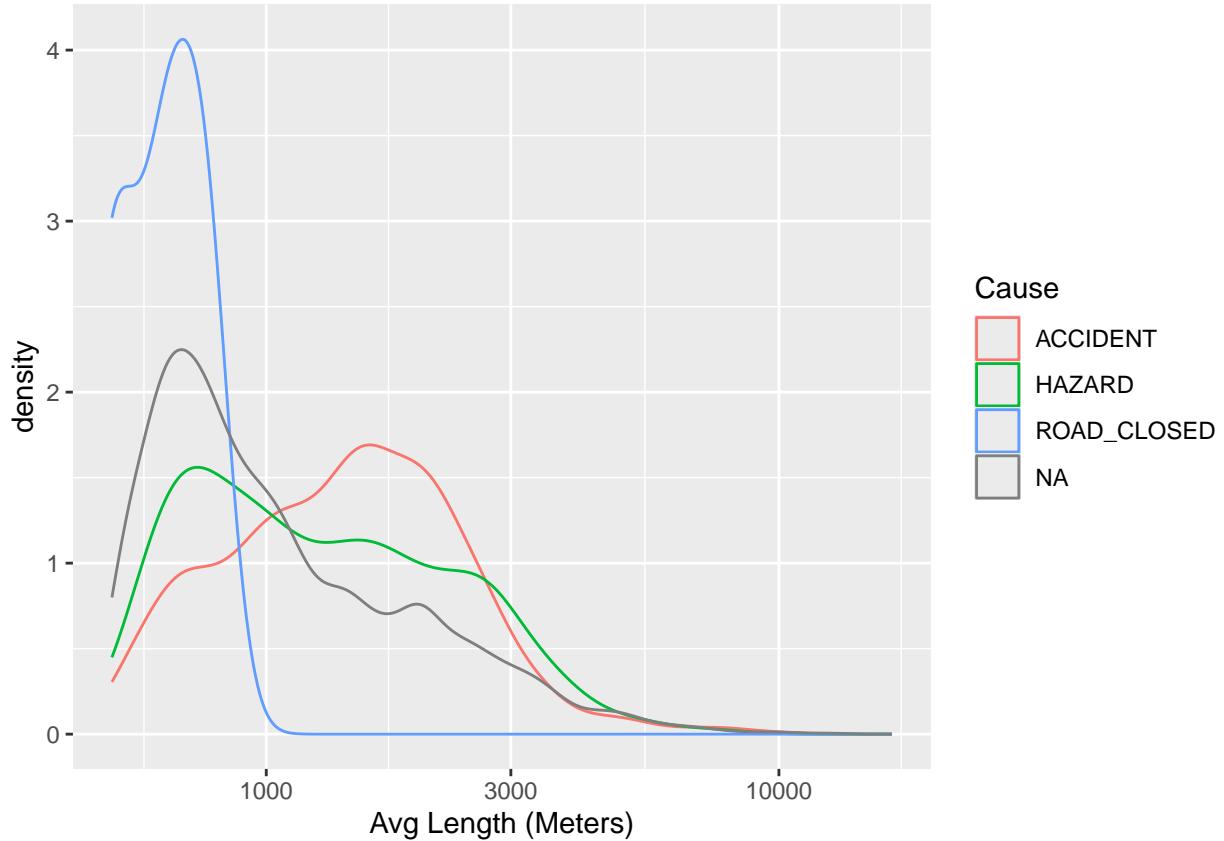
- Accidentes (ACCIDENT): 839 casos
- Peligros en la vía (HAZARD): 826 casos
- Carreteras cerradas (ROAD\_CLOSED): 3 casos
- Causa no especificada: 11834 casos



## Métricas de Impacto de Tráfico

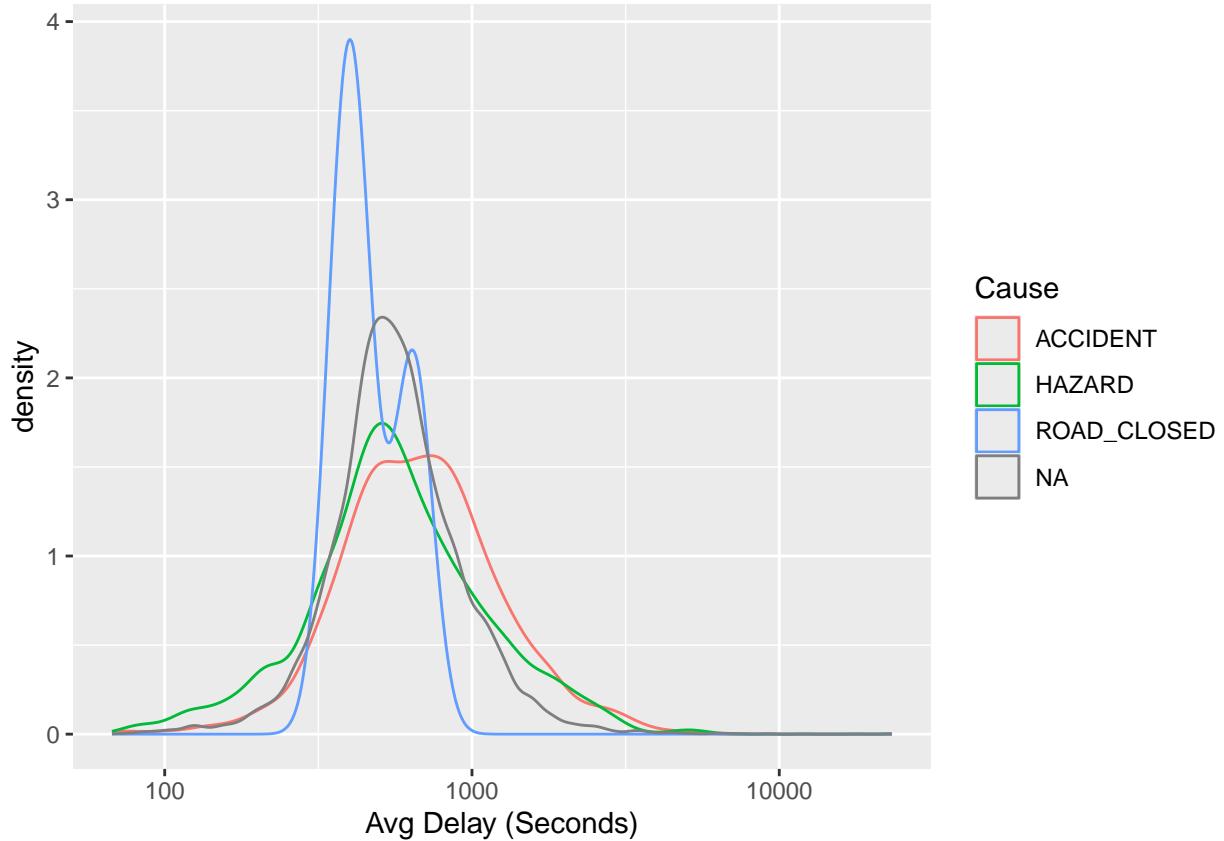
### Longitud de Afectación Vial

- **Rango:** 500 m - 16.61 km
- **Longitud promedio:** 1.34 km
- **Mediana:** 947 m
- **Distribución:** Sesgo hacia afectaciones de mediana extensión



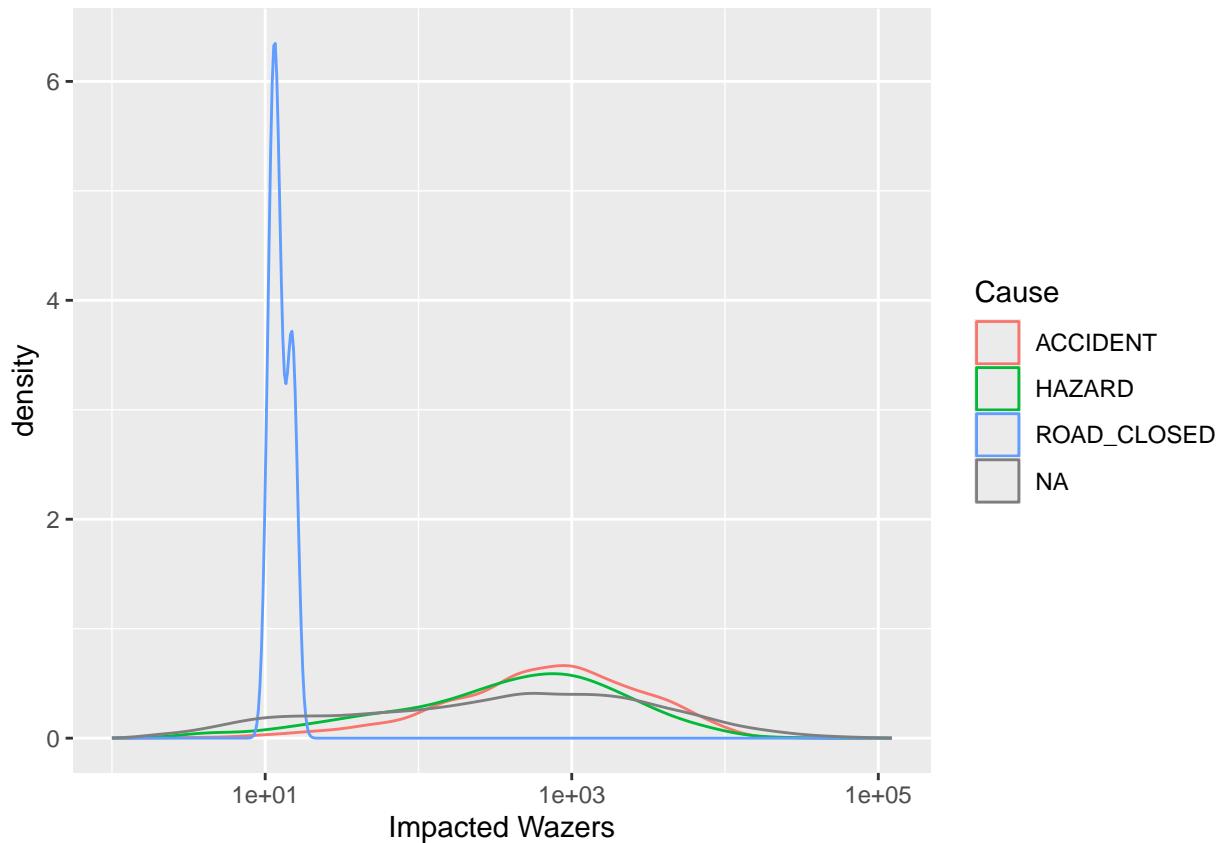
### Tiempos de Desfase

- **Rango:** 1 minuto 7.33 segundos - 6 horas 27 minutos 26 segundos
- **Desfase promedio:** 11 minutos 4.97 segundos
- **Mediana:** 9 minutos 18.25 segundos
- **Significado:** Retrasos significativos en la movilidad urbana



### Impacto en Usuarios

- **Usuarios afectados (Impacted Wazers):** 1 - 122563 personas
- **Promedio de afectados:** 2033 usuarios por incidente
- **Mediana:** 465 usuarios
- **Distribución:** Alta variabilidad con eventos masivos ocasionales



## Procesamiento y Normalización de Datos

### Paso a paso

- En la columna “Date”, se reemplazan los valores “ene”, “feb”, “mar”, “abr”, “may”, “jun”, “jul”, “ago”, “sep”, “oct”, “nov” y “dic” por “jan”, “feb”, “mar”, “apr”, “may”, “jun”, “jul”, “aug”, “sep”, “oct”, “nov” y “dec”, respectivamente
- “Country” se convierte en factor
- “City” se convierte en factor
- “Street” se convierte en factor
- “Cause” se convierte en factor

## Red de Waze

El sistema procesó 766668 registros de datos de rutas provenientes de la API de Waze, consolidando información de navegación y alertas de tráfico durante un período de 21 días.

### Metadatos de Procesamiento

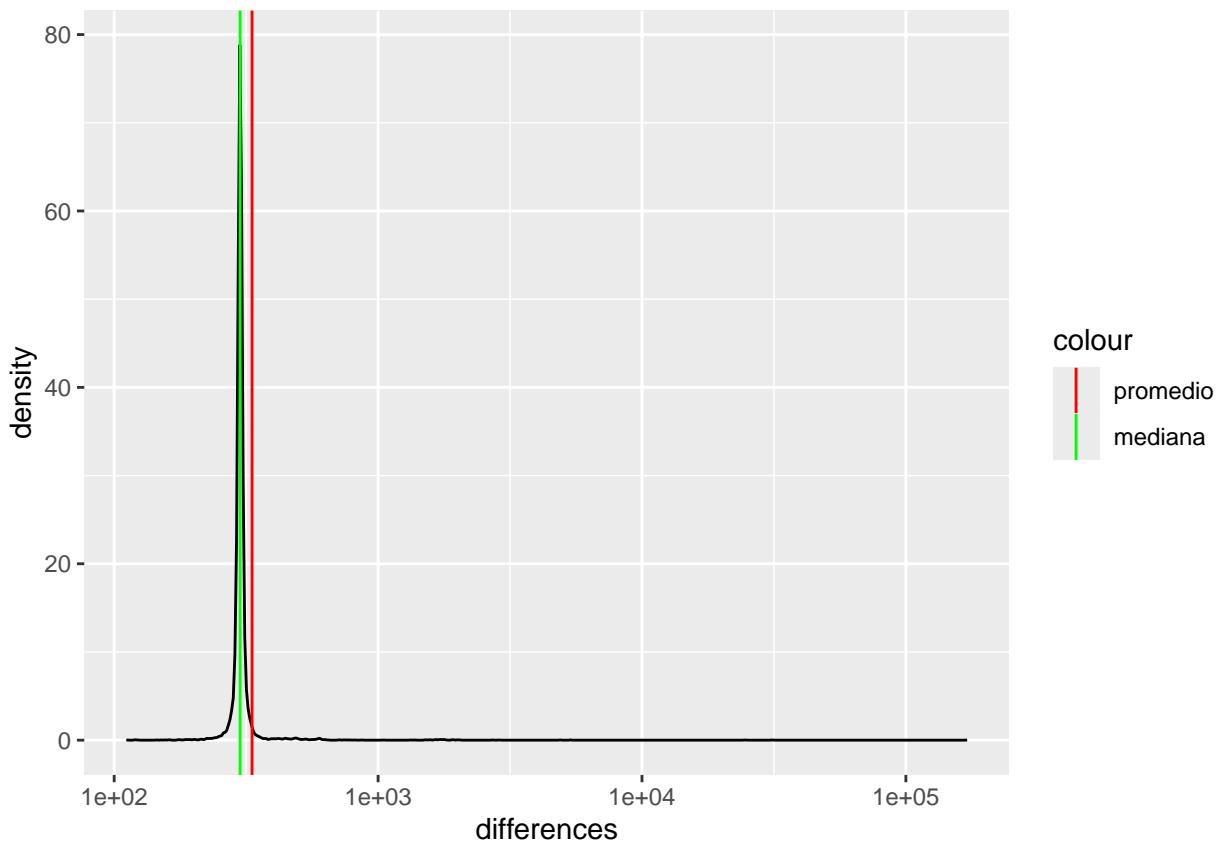
- **Tiempo de lectura:** 4 horas, 13 minutos, 36.01 segundos
- **Tiempo de unión:** 15.76 segundos
- **Tiempo de transformación:** 4 minutos, 35.99 segundos
- **Total de procesamiento:** 4 horas, 18 minutos, 27.76 segundos

## Período de Monitoreo

- **Fecha inicial:** 24 de noviembre de 2025 a las 15:31:51
- **Fecha final:** 29 de diciembre de 2025 a las 19:46:46
- **Duración:** 35 días, 4 horas, 14 minutos, 55 segundos
- **Estructura temporal:** Datos organizados por fecha en estructura de carpetas

## Granularidad Temporal

- Diferencia Mínima: 1 minuto, 51.2 segundos
- Diferencia máxima: 1 día, 23 horas, 25 minutos, 2.3 segundos
- Diferencia Media: 5 minutos, 33 segundos
- Diferencia Mediana: 5 minutos
- Desviación Standard: 30 minutos, 34.06 segundos



## Características del Dataset Consolidado

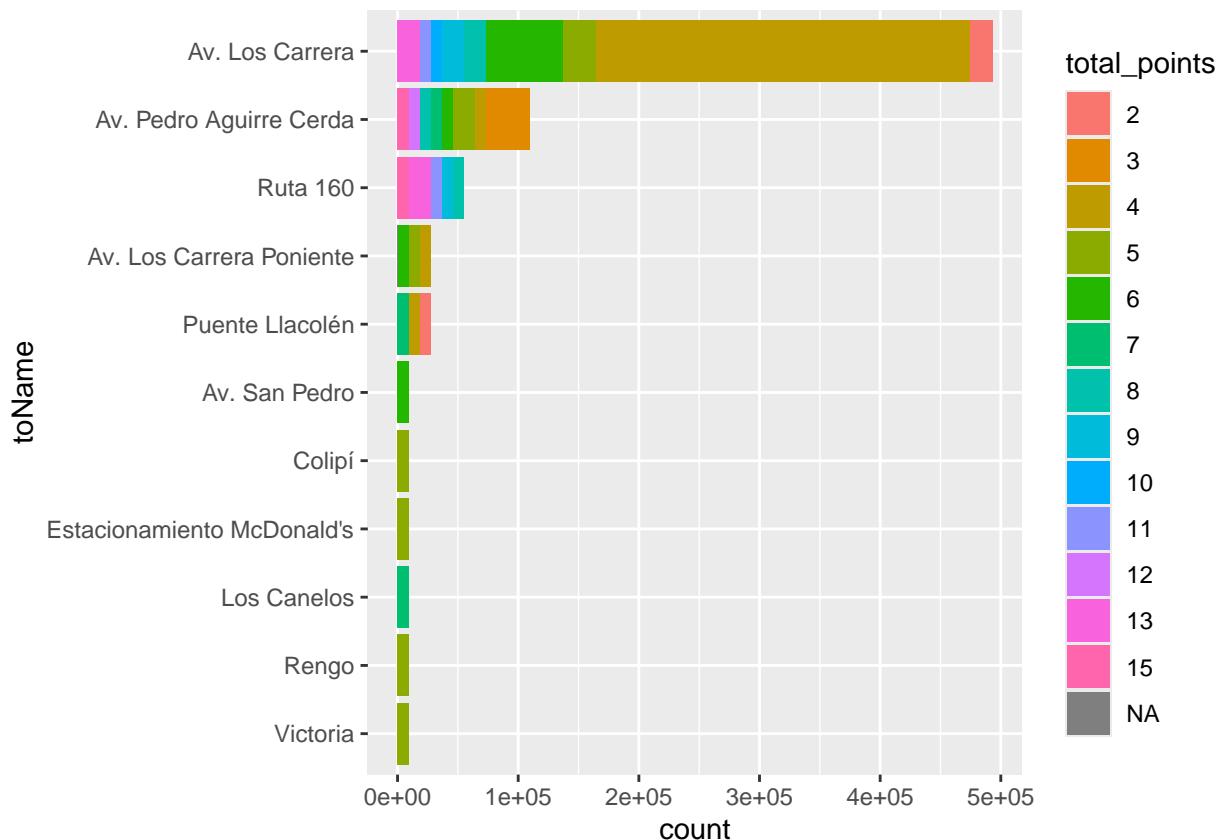
### Volumen y Estructura

- **Registros totales:** 4399211 puntos de ruta
- **Rutas únicas (agrupadas):** 766668 rutas
- **Nivel de detalle:** Desagregación punto por punto de trayectos

## Análisis de Orígenes y Destinos

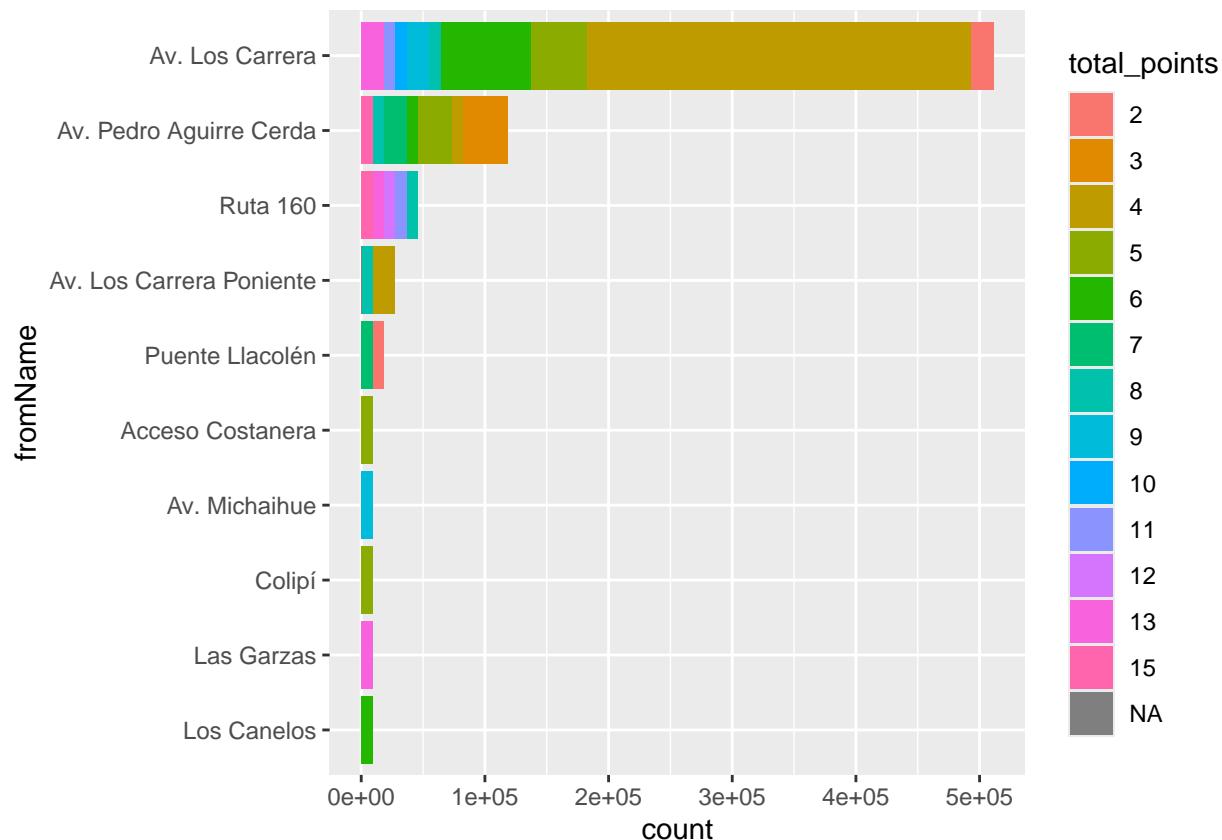
### Principales Destinos (toName)

- “Av. Los Carrera”: 492858 referencias
- “Av. Pedro Aguirre Cerda”: 109524 referencias
- “Ruta 160”: 54762 referencias



### Principales Orígenes (fromName)

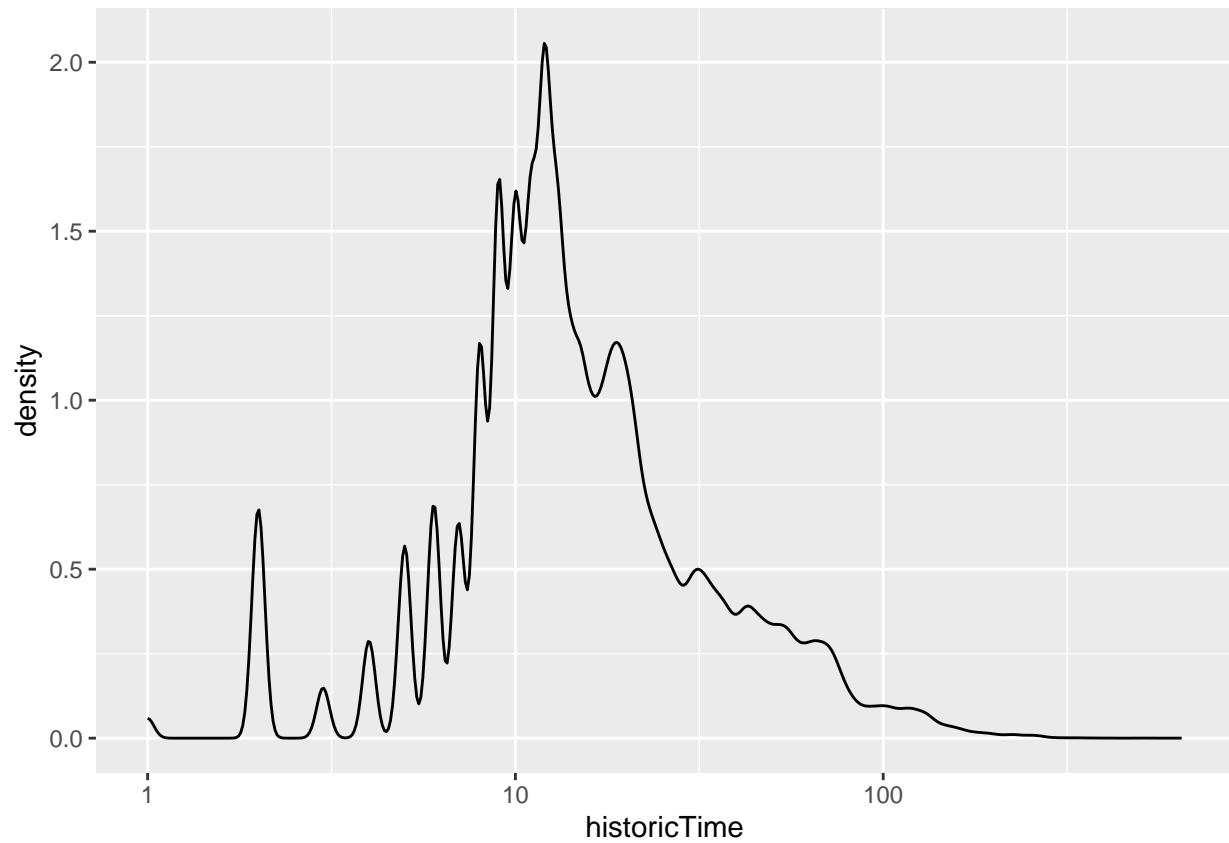
- “Av. Los Carrera”: 511112 referencias
- “Av. Pedro Aguirre Cerda”: 118651 referencias
- “Ruta 160”: 45635 referencias



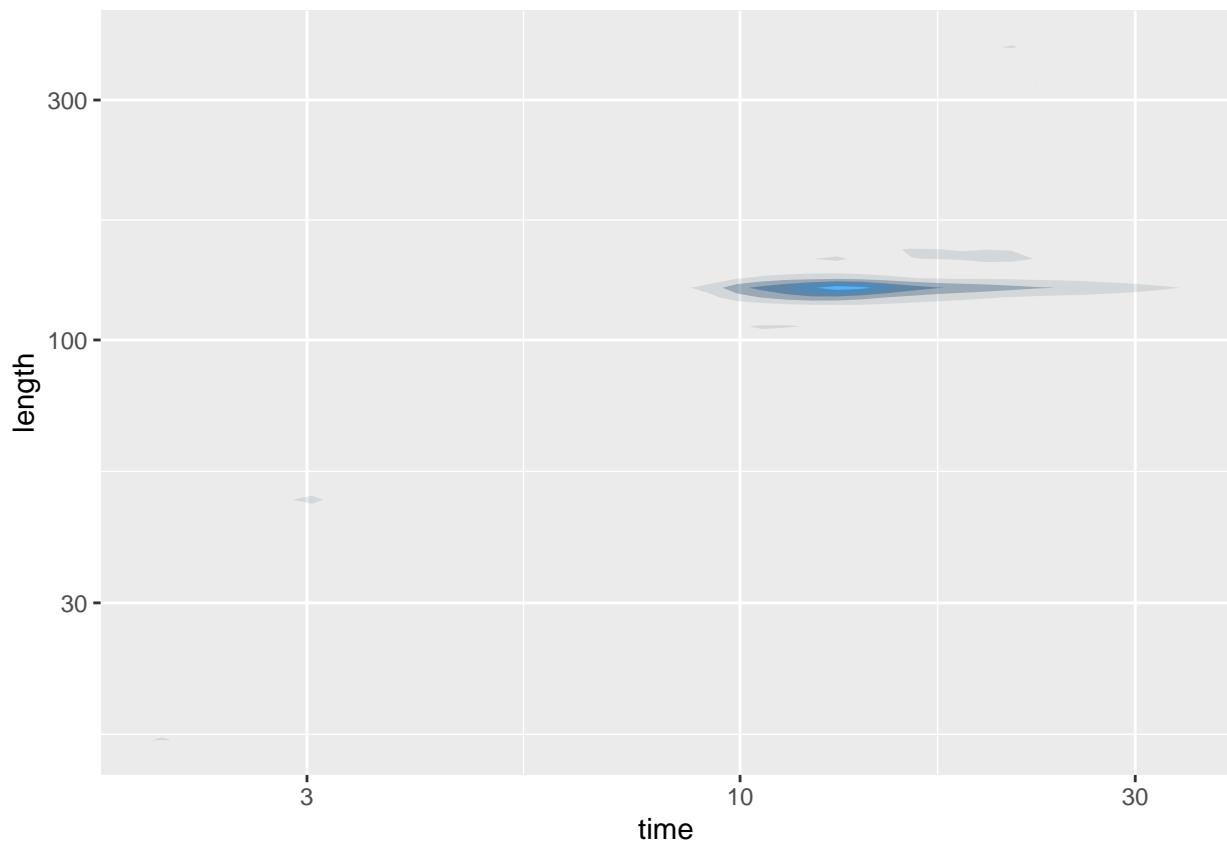
## Métricas de Desempeño de Rutas

### Tiempos de Viaje

- **Tiempo histórico (historicTime)**: 1-344 unidades (media: 20.74, mediana: 13)
- **Tiempo estimado (time)**: 0-952 unidades (media: 26.96, mediana: 16)
- **Unidades por confirmar**: Minutos o segundos (requiere validación)



*Distancias y Velocidades - Longitud de ruta (length): 0-1257 unidades (media: 175.4, mediana: 126) - Velocidad calculada (speed): 0.52-24 unidades (media: 8.06, mediana: 7.89) - Relación distancia/tiempo: Consistente con tráfico urbano congestionado*



## Sistema de Alertas Waze

*Distribución de Alertas*

- Total de alertas registradas: 301

*Tipología de Alertas (leadAlert.type)*

**Accidentes (ACCIDENT) - 159 casos:**

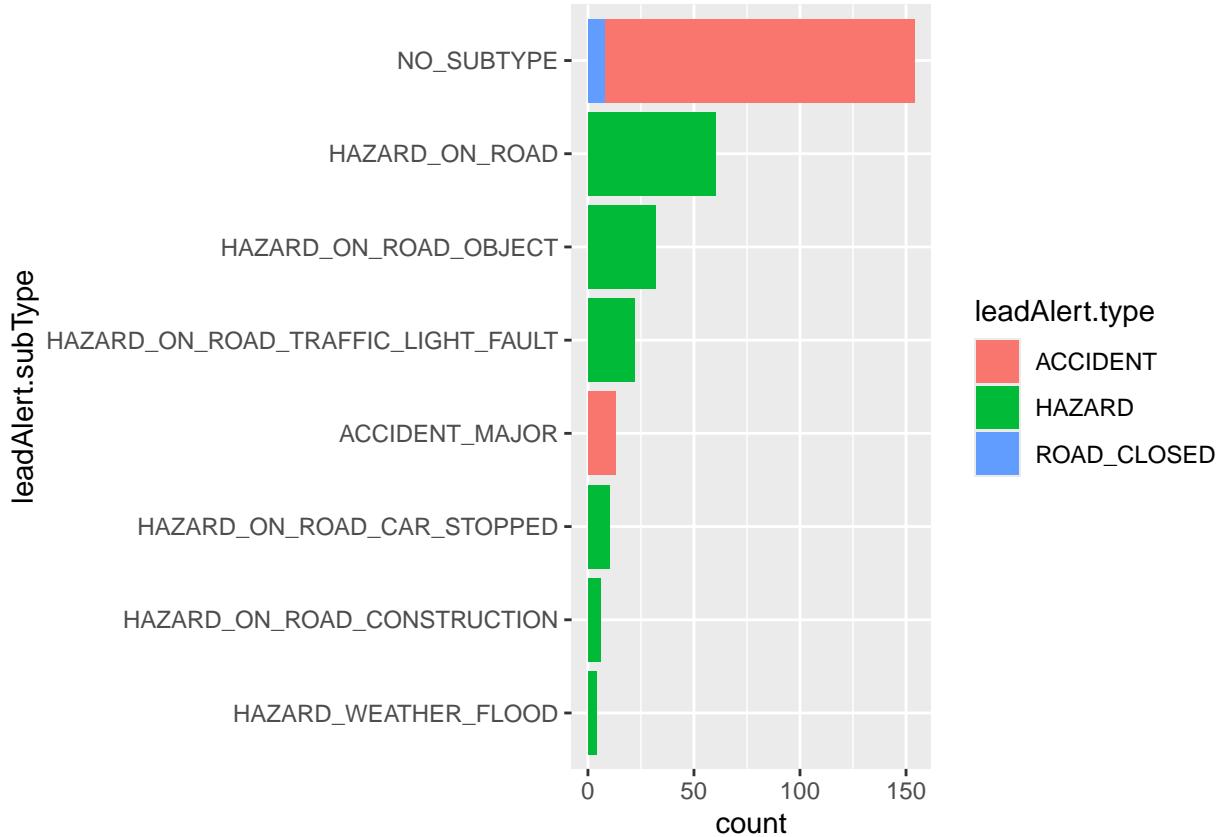
- NO\_SUBTYPE: 146 casos
- ACCIDENT\_MAJOR: 13 casos

**Peligros (HAZARD) - 134 casos:**

- HAZARD\_ON\_ROAD: 60 casos
- HAZARD\_ON\_ROAD\_OBJECT: 32 casos
- HAZARD\_ON\_ROAD\_TRAFFIC\_LIGHT\_FAULT: 22 casos
- HAZARD\_ON\_ROAD\_CAR\_STOPPED: 10 casos
- HAZARD\_ON\_ROAD\_CONSTRUCTION: 6 casos
- HAZARD\_WEATHER\_FLOOD: 4 casos

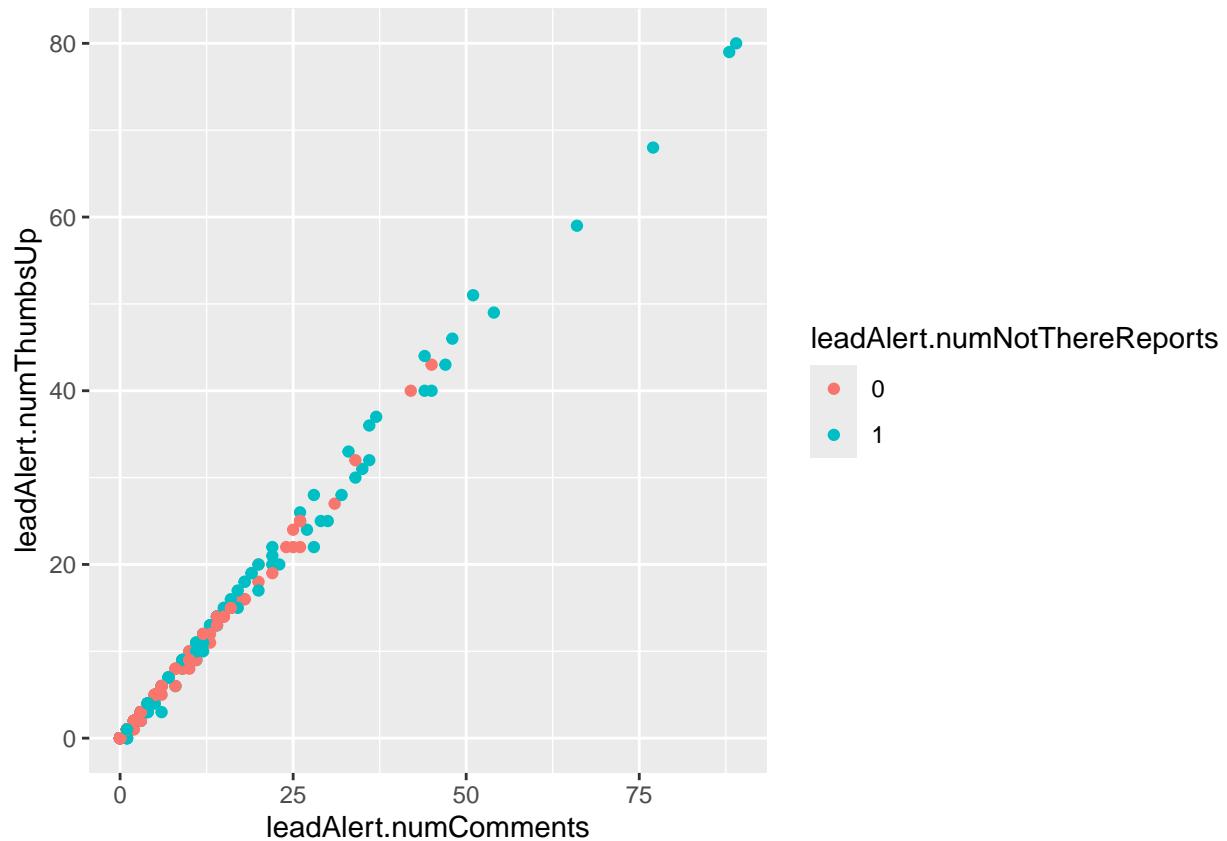
**Carreteras cerradas (ROAD\_CLOSED) - 8 casos**

- NO\_SUBTYPE: 8 casos



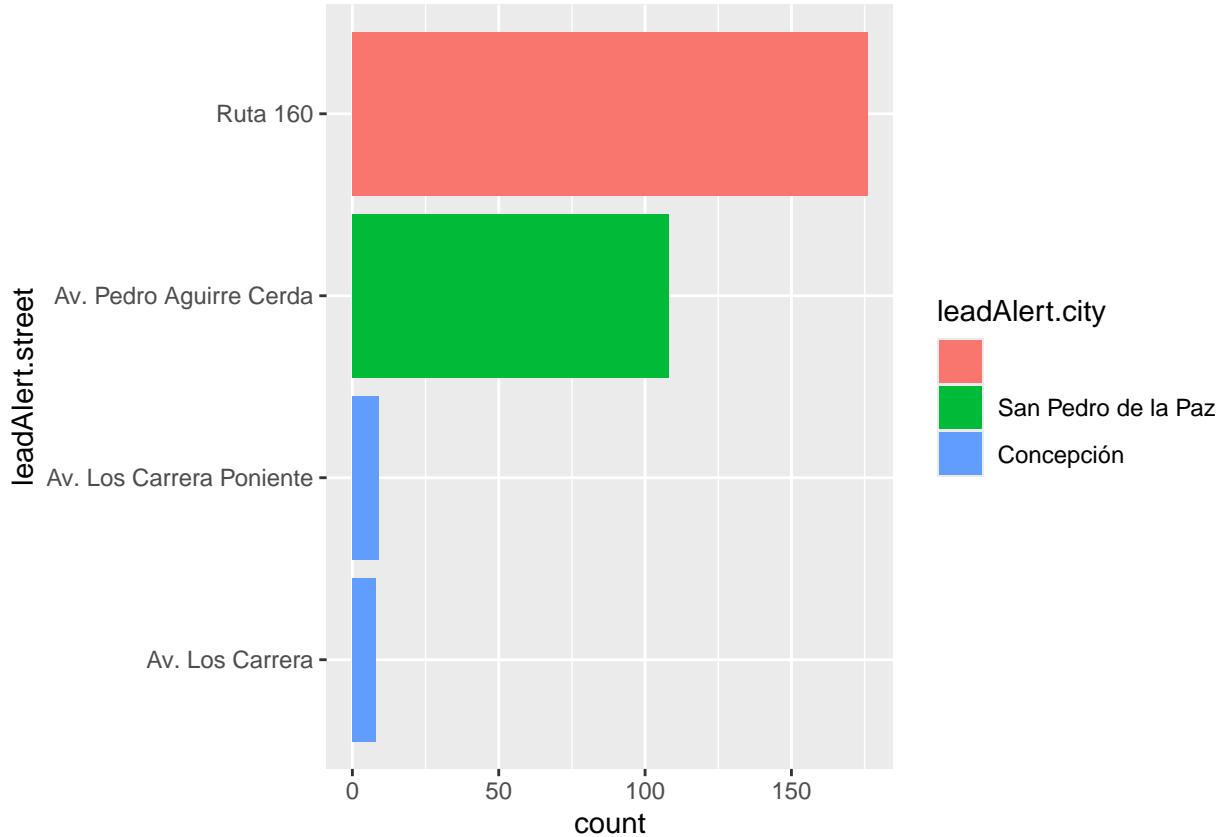
### *Interacción de Usuarios con Alertas*

- **Comentarios** (numComments): 0-89 (media: 9.2, mediana: 4)
- **Me gusta** (numThumbsUp): 0-46 (media: 7.5, mediana: 3)
- **Reportes de inexistencia** (numNotThereReports): 0 (120 casos) y 1 (89 casos)



#### *Contexto Geográfico de Alertas*

- **Calles más reportadas:**
  - “Ruta 160”: 176 alertas
  - “Av. Pedro Aguirre Cerda”: 108 alertas
  - “Av. Los Carrera Poniente”: 9 alertas
  - “Av. Los Carrera”: 8 alertas
- **Ciudades involucradas:**
  - “San Pedro de la Paz”: 108 alertas
  - “Concepción”: 17 alertas
  - Sin especificar: 176 alertas



## Procesamiento y Normalización de Datos

### Paso a paso

- Se leen las carpetas
- En cada carpeta, se leen las sub-carpetas
- En cada sub-carpeta, se leen los archivos .json
- En cada archivo, se leen las rutas
- Por cada ruta, se crea la fila con todos sus valores, excepto “line”, “bbox”, “subRoutes”, “leadAlert” y “alternateRoute”
- Si “leadAlert” no es nulo, las columnas “leadAlert.numComments”, “leadAlert.city”, “leadAlert.externalImageId”, “leadAlert.numThumbsUp”, “leadAlert.street”, “leadAlert.subType”, “leadAlert.id”, “leadAlert.position”, “leadAlert.type” y “leadAlert.numNotThereReports” toman los valores de “numComments”, “city”, “externalImageId”, “numThumbsUp”, “street”, “subType”, “id”, “position”, “type” y “numNotThereReports”, respectivamente de “leadAlert”
- Si “bbox” no es nulo, las columnas “minY”, “minX”, “maxY” y “maxX” toman los valores de “minY”, “minX”, “maxY” y “maxX” de “bbox”, respectivamente
- Si “line” no es nulo, se clona la fila una vez por cada elemento en “line”, por cada punto, “x” e “y” toman el valor de “x” e “y”, respectivamente del elemento de “line”, “point\_index”, la posición del punto en “line” y “total\_points”, la cantidad de puntos en “line”
- Se unen las filas en un data.frame
- Se crea la columna “speed” con el valor de “length” / “time”
- Se crea la columna “broadcasterId” con el valor de “broadcasterId” del archivo
- Se crea la columna “areaName” con el valor de “areaName” del archivo
- Se crea la columna “isMetric” con el valor de “isMetric” del archivo

- Se crea la columna “updateTime” con el valor de “updateTime” del archivo
- Se crea la columna “date” con el nombre de la carpeta
- Se crea la columna “subfolder” con el nombre de la sub-carpeta
- Se crea la columna “file” con el nombre del archivo quitando la cadena “.json”, separándolo en una lista de cadenas con el carácter “\_” y tomando en último elemento (ej: “1763409379301\_2025-12-01\_000004.267298.json” -> “000004.267298”)
- Se unen las filas en un solo data.frame
- “toName” se convierte en factor
- “name” se convierte en factor
- “fromName” se convierte en factor
- “type” se convierte en factor
- “leadAlert.city” se convierte en factor
- “leadAlert.externalImageId” se convierte en factor
- “leadAlert.street” se convierte en factor
- “leadAlert.subType” se convierte en factor
- “leadAlert.position” se divide en “leadAlert.position.lat” y “leadAlert.position.lng” separándolo por el carácter ” ”
- Se elimina la columna “leadAlert.position”
- “leadAlert.type” se convierte en factor
- “broadcasterId” se convierte en factor
- “areaName” se convierte en factor
- “date” se convierte en POSIXct con formato “%d-%m-%Y”
- “subfolder” se convierte en factor
- “file” se convierte en factor
- “updateTime” se divide por 1000 y se convierte en POSIXct con origen en “1970-01-01” y zona horaria “UTC”
- Se crea la columna “weekday” siendo el día de la semana de “updateTime” y se convierte en factor

## Discretización de datos

### Descripción Matemática

La función implementa un proceso de cuantización vectorial adaptativa para datos espacio-temporales, donde la discretización espacial se alinea con la dirección de las calles. Formalmente:

$$quantum.data : \mathbb{R}^{4 \times n} \times \mathbb{R}^2 \times (\mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}) \rightarrow \mathbb{R}^{6 \times m}$$

Donde:

- **Dominio:** 4 vectores de dimensión  $n$  (tiempo, longitud, latitud y campo), parámetros de discretización y función de agregación
- **Codominio:** Matriz de datos discretizados con 6 atributos y  $m \leq n$  puntos

### Parámetros de Entrada

- time.data  $\in \mathbb{R}^n$ : Vector de tiempos en segundos
- lng.data  $\in \mathbb{R}^n$ : Vector de longitudes en grados decimales
- lat.data  $\in \mathbb{R}^n$ : Vector de latitudes en grados decimales
- field.data  $\in \mathbb{R}^n$ : Vector de valores del campo de interés
- time.gap  $\in \mathbb{R}^+$ : Intervalo temporal de discretización en segundos
- space.gap  $\in \mathbb{R}^+$ : Intervalo espacial a lo largo de la calle en metros

- street.aligned = TRUE: Activa discretización alineada a calles
- street.angle.tolerance = 30: Tolerancia angular para agrupación transversal en grados
- compile.function :  $\mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$ : Función de agregación (media, mediana, moda, máximo, mínimo, etc.)

## Proceso de Discretización

### 1. Constantes de Conversión Geográfica

Para convertir coordenadas geográficas a distancias métricas:

$$C_{lat} \equiv 111320m/\ddot{r}$$

$$C_{lng}(\phi) \equiv C_{lat} \times \cos\left(\phi \times \frac{\pi}{180}\right)$$

Donde  $\phi$  es la latitud en grados.

### 2. Cálculo de la Dirección de la Calle (Heading)

Para cada punto  $P_i = (t_i, \lambda_i, \phi_i, v_i)$  donde:

- $t_i$ : tiempo
- $\lambda_i$ : longitud
- $\phi_i$ : latitud
- $v_i$ : valor del campo

Calculamos el ángulo de dirección  $\theta_i$  usando diferencias centrales:

$$\Delta x_i^m = (\lambda_{i+1} - \lambda_{i-1}) \times C_{lng}(\phi_i)$$

$$\Delta y_i^m = (\phi_{i+1} - \phi_{i-1}) \times C_{lat}$$

$$\theta_i = \text{atan2}(\Delta y_i^m, \Delta x_i^m)$$

Donde  $\text{atan2}(y, x)$  es la función arcotangente de dos argumentos que devuelve el ángulo en radianes entre el eje X positivo y el vector  $(x, y)$ .

### 3. Transformación al Sistema de Coordenadas de la Calle

Para cada punto  $P_i = (\lambda_i, \phi_i, \theta_i)$ , donde:

- $\lambda_i$ : longitud en grados
- $\phi_i$ : latitud en grados
- $\theta_i$ : ángulo de dirección de la calle en radianes (calculado en el paso anterior)

Realizamos una transformación de coordenadas afín que alinea el sistema de referencia local con la dirección de la calle:

### 3.1 Conversión a Coordenadas Métricas Cartesianas

Primero, convertimos las coordenadas geográficas a un sistema métrico local:

$$\begin{aligned}x_i^G &= \lambda_i \times C_{lng}(\phi_i) \\y_i^G &= \phi_i \times C_{lat}\end{aligned}$$

Donde:

- $x_i^G$ : coordenada este-oeste en metros
- $y_i^G$ : coordenada norte-sur en metros
- $C_{lng}(\phi_i) = 111320 \times \cos(\phi_i \times \frac{\pi}{180})$ : metros por grado de longitud a latitud  $\phi_i$
- $C_{lat} = 111320$ : metros por grado de latitud

### 3.2 Construcción del Sistema de Coordenadas Local

Definimos una base ortonormal  $\mathcal{B}_i = \vec{e}_{\parallel}, \vec{e}_{\perp}$  donde:

- $\vec{e}_{\parallel} = (\cos \theta_i, \sin \theta_i)$ : vector unitario paralelo a la calle
- $\vec{e}_{\perp} = (-\sin \theta_i, \cos \theta_i)$ : vector unitario perpendicular a la calle

Esta base satisface:

- $|\vec{e}_{\parallel}| = |\vec{e}_{\perp}| = 1$
- $\vec{e}_{\parallel} \cdot \vec{e}_{\perp} = 0$  (ortogonalidad)
- $\det(\vec{e}_{\parallel}, \vec{e}_{\perp}) = 1$  (orientación positiva)

### 3.3 Transformación de Coordenadas

La transformación del sistema global de coordenadas geográficas al sistema local alineado con la calle se realiza mediante una rotación ortogonal que depende del ángulo de dirección  $\theta_i$ .

*Definición Formal*

Dado un punto  $P_i$  con:

- Coordenadas geográficas originales:  $(\lambda_i, \phi_i)$  en grados
- Coordenadas métricas globales:  $(x_i^G, y_i^G)$  en metros
- Ángulo de dirección de la calle:  $\theta_i$  en radianes

La transformación al sistema local es:

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = R(\theta_i)^{\top} \cdot \begin{bmatrix} x_i^G \\ y_i^G \end{bmatrix}$$

donde  $R(\theta_i)$  es la matriz de rotación:

$$R(\theta_i) = \begin{bmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{bmatrix}$$

y por lo tanto:

$$R(\theta_i)^\top = R(-\theta_i) = \begin{bmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{bmatrix}$$

*Desarrollo Paso a Paso*

1. Vector de posición en sistema global:

$$\mathbf{r}_i^G = \begin{bmatrix} x_i^G \\ y_i^G \end{bmatrix}$$

2. Bases de los sistemas de coordenadas: Sistema global  $\mathcal{G}$ :

$$\mathcal{G} = \left\{ \vec{e}_x^G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{e}_y^G = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

Sistema local  $\mathcal{L}_i$ :

$$\mathcal{L}_i = \left\{ \vec{e}_{\parallel} = \begin{bmatrix} \cos \theta_i \\ \sin \theta_i \end{bmatrix}, \vec{e}_{\perp} = \begin{bmatrix} -\sin \theta_i \\ \cos \theta_i \end{bmatrix} \right\}$$

3. Cambio de base: Las coordenadas  $(x'_i, y'_i)$  en el sistema local satisfacen:

$$\bar{r}_i^G = x'_i \vec{e}_{\parallel} + y'_i \vec{e}_{\perp}$$

En forma matricial:

$$\begin{bmatrix} x_i^G \\ y_i^G \end{bmatrix} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \end{bmatrix}$$

4. Solución para coordenadas locales: Despejando  $(x'_i, y'_i)$ :

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix}^{-1} \begin{bmatrix} x_i^G \\ y_i^G \end{bmatrix}$$

*Interpretación Geométrica*

1. Sistema global:

- Eje X: dirección este (positivo hacia el este)
- Eje Y: dirección norte (positivo hacia el norte)

2. Sistema local:

- Eje  $x'$  (longitudinal): paralelo a la dirección de la calle  $\theta_i$
- Eje  $y'$  (transversal): perpendicular a la calle  $(\theta_i + \frac{\pi}{2})$

3. Transformación como rotación activa vs. pasiva: Esta transformación puede interpretarse de dos formas equivalentes:

- a) Rotación pasiva del sistema de referencia: Rotamos los ejes de coordenadas en un ángulo  $-\theta_i$ , manteniendo fijo el vector  $\bar{r}_i^G$ .
- b) Rotación activa del vector: Rotamos el vector  $\bar{r}_i^G$  en un ángulo  $\theta_i$  en sentido contrario a las agujas del reloj, manteniendo fijos los ejes.

## 4. Discretización Temporal

Sea  $T = t_1, t_2, \dots, t_n$  el conjunto ordenado de tiempos, con  $t_1 < t_2 < \dots < t_n$ .

### 4.1 Construcción de Bins Temporales

Definimos el ancho de bin temporal como  $\Delta t = \text{time.gap}$ . Los límites temporales son:

- $t_{\min} = \min(T)$
- $t_{\max} = \max(T)$

El número de bins temporales  $N_t$  se calcula como:

$$N_t = \left\lceil \frac{t_{\max} - t_{\min}}{\Delta t} \right\rceil + 1$$

Los bins temporales son intervalos semiabiertos:

$$B_t[k] = [t_{\min} + (k - 1)\Delta t, t_{\min} + k\Delta t), \quad k = 1, 2, \dots, N_t - 1$$

El último bin es cerrado para incluir  $t_{\max}$ :

$$B_t[N_t] = [t_{\min} + (N_t - 1)\Delta t, t_{\max}]$$

### 4.2 Asignación de Puntos a Bins Temporales

Para cada tiempo  $t_i$ , determinamos su bin temporal:

$$b_t(i) = \left\lfloor \frac{t_i - t_{\min}}{\Delta t} \right\rfloor + 1$$

Con la restricción:  $1 \leq b_t(i) \leq N_t$

## 5. Discretización Espacial Orientada a Calles

### 5.1 Discretización Longitudinal (a lo largo de $x'$ )

Sea  $X' = \{x'_1, x'_2, \dots, x'_n\}$  el conjunto de coordenadas longitudinales transformadas.

*Paso 1:* Determinación del rango:

$$x'_{\min} = \min(X'), \quad x'_{\max} = \max(X')$$

*Paso 2:* Construcción de bins longitudinales con ancho  $\Delta s = \text{space.gap}$ :

$$N_x = \left\lceil \frac{x'_{\max} - x'_{\min}}{\Delta s} \right\rceil + 1$$

Los bins longitudinales son:

$$B_x[\ell] = [x'_{\min} + (\ell - 1)\Delta s, x'_{\min} + \ell\Delta s), \quad \ell = 1, 2, \dots, N_x - 1$$

El último bin incluye el límite superior:

$$B_x[N_x] = [x'_{\min} + (N_x - 1)\Delta s, x'_{\max}]$$

*Paso 3:* Asignación de puntos a bins:

$$b_x(i) = \left\lfloor \frac{x'_i - x'_{\min}}{\Delta s} \right\rfloor + 1$$

## 5.2 Discretización Transversal (a lo largo de $y'$ )

*Paso 1:* Cálculo del ancho de bin transversal:

$$\Delta s_y = \Delta s \times \sin \left( \alpha \times \frac{\pi}{180} \right)$$

donde  $\alpha = \text{street.angle.tolerance} = 30^\circ$ .

Para  $\alpha = 30^\circ$ :

$$\sin(30^\circ) = \frac{1}{2} \Rightarrow \Delta s_y = \frac{\Delta s}{2}$$

**Interpretación:** Esta elección crea bins con relación de aspecto 2:1, lo que significa que los bins son el doble de largos que de anchos. Esto refleja la intuición de que la variación a lo largo de una calle es más importante que la variación transversal para análisis de tráfico.

*Paso 2:* Determinación del rango transversal:

$$y'_{\min} = \min(Y'), \quad y'_{\max} = \max(Y')$$

donde  $Y' = \{y'_1, y'_2, \dots, y'_n\}$  es el conjunto de coordenadas transversales transformadas.

*Paso 3:* Construcción de bins transversales:

$$N_y = \left\lceil \frac{y'_{\max} - y'_{\min}}{\Delta s_y} \right\rceil + 1$$

Los bins transversales son:

$$B_y[m] = [y'_{\min} + (m - 1)\Delta s_y, y'_{\min} + m\Delta s_y], \quad m = 1, 2, \dots, N_y - 1$$

El último bin incluye el límite superior:

$$B_y[N_y] = [y'_{\min} + (N_y - 1)\Delta s_y, y'_{\max}]$$

*Paso 4:* Asignación de puntos a bins:

$$b_y(i) = \left\lfloor \frac{y'_i - y'_{\min}}{\Delta s_y} \right\rfloor + 1$$

## 6. Agregación por Celdas 3D

### 6.1 Definición de Celdas

Una **celda**  $C_{k,\ell,m}$  se define como la intersección de: - Bin temporal  $B_t[k]$  - Bin longitudinal  $B_x[\ell]$  - Bin transversal  $B_y[m]$

Formalmente:

$$C_{k,\ell,m} = \{i \in \{1, \dots, n\} : b_t(i) = k, b_x(i) = \ell, b_y(i) = m\}$$

### 6.2 Estadísticos por Celda

Para cada celda no vacía ( $|C_{k,\ell,m}| > 0$ ):

1. **Tiempo representativo** (promedio temporal):

$$\bar{t}_{k,\ell,m} = \frac{1}{|C|} \sum_{i \in C} t_i$$

2. **Posición representativa** (centroide en coordenadas originales):

- Longitud promedio:  $\bar{\lambda}_{k,\ell,m} = \frac{1}{|C|} \sum_{i \in C} \lambda_i$
- Latitud promedio:  $\bar{\phi}_{k,\ell,m} = \frac{1}{|C|} \sum_{i \in C} \phi_i$

3. **Valor del campo agregado:**

$$\bar{v}_{k,\ell,m} = \text{compile.function}(\{v_i : i \in C\})$$

Donde `compile.function` puede ser:

- Media:  $\frac{1}{|C|} \sum_{i \in C} v_i$
- Máximo:  $\max\{v_i : i \in C\}$
- Mínimo:  $\min\{v_i : i \in C\}$
- Suma:  $\sum_{i \in C} v_i$

4. **Número de puntos:**

$$n_{k,\ell,m} = |C| = \text{cardinalidad de } C$$

5. **Dirección promedio de la calle** (promedio circular): Para ángulos  $\theta_i$ , el promedio circular se calcula como:

- Componente coseno:  $C_\theta = \frac{1}{|C|} \sum_{i \in C} \cos \theta_i$
- Componente seno:  $S_\theta = \frac{1}{|C|} \sum_{i \in C} \sin \theta_i$
- Dirección promedio:  $\bar{\theta}_{k,\ell,m} = \text{atan2}(S_\theta, C_\theta)$

**Nota:** Este método evita problemas en el promedio de ángulos cerca de  $0^\circ/360^\circ$ .

### 6.3 Propiedades de la Agregación

1. **Conservación de masa** (para funciones de suma):

$$\sum_{k,\ell,m} \bar{v}_{k,\ell,m} = \sum_{i=1}^n v_i$$

2. **Monotonicidad:** Si `compile.function` es monótona, el orden se preserva.

3. **Idempotencia:** Si todos los puntos caen en la misma celda y `compile.function` = identidad, entonces  $\bar{v} = v_i$  para todo  $i$ .

## 7. Construcción de la Salida

La función retorna un conjunto de  $m$  puntos discretizados:

$$D = \{(\bar{t}_j, \bar{\lambda}_j, \bar{\phi}_j, \bar{v}_j, n_j, \bar{\theta}_j)\}_{j=1}^m$$

donde  $j$  indexa las celdas no vacías ordenadas por:

1. Tiempo  $\bar{t}_j$  (creciente)
2. Coordenada longitudinal transformada  $x'_j$  (creciente)
3. Coordenada transversal transformada  $y'_j$  (creciente)

### 7.1 Transformación Inversa para Visualización

Para visualizar o analizar los resultados en coordenadas originales, cada punto discretizado representa una **celda rectangular** en el sistema de coordenadas de la calle:

- Centro:  $(\bar{\lambda}_j, \bar{\phi}_j)$
- Dimensiones:  $\Delta s$  (largo)  $\times \Delta s_y$  (ancho)
- Orientación:  $\bar{\theta}_j$

Los vértices de la celda en coordenadas globales son:

$$\begin{bmatrix} \lambda_{\text{vértice}} \\ \phi_{\text{vértice}} \end{bmatrix} = \begin{bmatrix} \bar{\lambda}_j \\ \bar{\phi}_j \end{bmatrix} + \frac{1}{C_{ln}( \bar{\phi}_j)} R(\bar{\theta}_j) \begin{bmatrix} \pm \frac{\Delta s}{2} \\ \pm \frac{\Delta s_y}{2} \end{bmatrix}$$

## 8. Propiedades Matemáticas Globales

### 8.1 Reducción Dimensional

El número de puntos de salida  $m$  satisface:

$$m \leq \min(n, N_t \times N_x \times N_y)$$

En la práctica, típicamente:

$$m \approx \frac{n}{\left\lfloor \frac{\Delta t}{\Delta t_{\text{original}}} \right\rfloor \times \left\lfloor \frac{\Delta s}{\Delta s_{\text{original}}} \right\rfloor}$$

donde  $\Delta t_{\text{original}}$  y  $\Delta s_{\text{original}}$  son las resoluciones originales de los datos.

### 8.2 Error de Discretización

Definimos el **error de representación** para una celda  $C$ :

$$\epsilon_C = \frac{1}{|C|} \sum_{i \in C} \|P_i - \bar{P}_C\|$$

donde  $P_i = (\lambda_i, \phi_i)$  y  $\bar{P}_C = (\bar{\lambda}_C, \bar{\phi}_C)$ .

Bajo condiciones suaves:

$$\mathbb{E}[\epsilon_C] \approx \frac{\Delta s}{2\sqrt{3}} \text{ (longitudinal)} + \frac{\Delta s_y}{2\sqrt{3}} \text{ (transversal)}$$

### 8.3 Conservación de Información

La transformación preserva:

- Orden temporal:** Si  $t_i < t_j$ , entonces  $\bar{t}_{C(i)} \leq \bar{t}_{C(j)}$
- Vecindad espacial:** Puntos cercanos en la misma dirección se mantienen cercanos
- Propiedades estadísticas:** Dependiendo de compile.function

## 9. Ventajas sobre Discretización Cartesiana Tradicional

- Adaptabilidad direccional:** Se ajusta a la geometría de la red vial
- Mejor agrupación semántica:** Agrupa puntos que están en la misma calle
- Reducción de error en curvas:** Menor distorsión en tramos curvos
- Separación natural de direcciones:** Diferentes sentidos de circulación se discretizan por separado
- Interpretabilidad física:** Las dimensiones de las celdas tienen significado físico (largo/ancho de calle)